University of South Bohemia in České Budějovice
Faculty of Science

Elucidation of the retrograde transport of contact-dependent inhibition (CDI) toxins through bacterial SecYEG translocon complex

Bachelor's Thesis

Alfred Emiri

Supervisor: Tomas Fessl, Ph.D.

České Budějovice, 2024

Emiri, A., 2024: Elucidation of the retrograde transport of contact-dependent inhibition (CDI) toxins through bacterial SecYEG translocon complex. Bc. Thesis, in English. – 68 p., Faculty of Science, University of South Bohemia, České Budějovice, Czech Republic.

Annotation

This study investigates the retrograde translocation of CdiA toxins through the bacterial SecYEG translocon, focusing on the unfolding of the effector domain, proven to be essential for successful transport through the translocon's narrow channel. Using molecular dynamics simulations, the research highlights the pivotal role of pH in influencing the domain unfolding. Additionally, a novel smFRET-adapted import assay to study the retrograde transport mechanism of these toxins *in vitro* was developed. The findings pave the way for deeper insights into CDI toxins retrograde transport mechanisms.

I declare that I am the author of this qualification thesis and that in writing it I have used the sources and literature displayed in the list of used sources only.

České Budějovice, Czech Republic

09 May 2024

Acknowledgments

This was my first ever thesis, and I am excited to have successfully completed it. As a young curious scientist, during this project I learnt a lot of scientific knowledge about the research topic I worked on, learnt a lot of new biophysical and computational methods as well, but importantly to me, I got once again reaffirmed that, as in every other aspect in life, achieving great things in science also takes a lot of hard work and dedication.

I want to thank my supervisor Dr Tomas Fessl for having me as his student for this project and for guiding me to be a better young scientist. I liked the way he fed my curiosity, and I liked the freedom of learning-by-mistakes that he always offered me. Even though demanding, I also liked that we both always wanted to achieve the best of things out of what we did.

I want to thank my close friends that I made during this time in the lab or outside it. They have all contributed to my journey at USB, and definitely, have given colours to it.

Lastly, I want to thank my family. I am lucky to have two older brothers that continue to pave the way for me in every aspect of life. I am lucky to have a mom and dad that sacrifice continuously for me to have the opportunity to follow my dreams. Throughout all of this, their support was as strong as always, pushing me to overcome any difficulty encountered.

Abstract

Contact-dependent inhibition (CDI) systems are mechanisms that inhibit competing bacteria through direct cell-to-cell contact, utilizing toxins. The complex mechanism by which these toxins are transported through the inner membrane of the target cell remains unknown. This study investigated the reverse translocation of these toxins via the bacterial SecYEG translocon, focusing on the unfolding of the CdiA toxin's effector domain, a process shown to be a prerequisite for passing through the narrow channel. Molecular dynamics (MD) simulations were employed to analyse the domain's mechanical unfolding under varying pH conditions representative of the periplasmic and cytoplasmic environments. The results showed the unfolding to follow a sequential pattern that aligned with the proposed toxin entry mechanism. Additionally, simulations revealed that the periplasmic pH facilitated easier unfolding, with potential of mean force (PMF) calculations indicating lower energy requirements under these conditions. To further investigate the CDI toxin's shipment process, a novel import assay was developed and adapted to single-molecule Förster resonance energy transfer (smFRET), enabling real-time observation of translocation. The assay proved effective for monitoring these events, providing a reliable tool for studying CdiA toxin transport across bacterial membranes. Continuing the findings presented here, future research could further investigate the import mechanism by using the newly developed assay, by implementing constant pH MD simulations to better replicate *in vivo* conditions, or by utilizing atomic force microscopy (AFM) for *in vitro* pulling experiments.

Table of Contents

1. Introduction

1.1. Bacterial Warfare

Bacterial warfare, also known as microbial competition or antagonism, refers to the strategies bacteria employ to compete for resources and survival in various environments. These mechanisms are crucial for understanding bacterial ecology, evolution, and their potential applications in medicine and biotechnology.

One prominent form of bacterial warfare involves the production of antibacterial compounds known as bacteriocins. These proteinaceous toxins are produced by bacteria to inhibit the growth of closely related or competitive bacterial strains. Bacteriocins vary widely in their mode of action and specificity, acting against competitors by forming pores in the target cell membrane, degrading cellular nucleic acids, or disrupting essential enzymatic activities (Riley and Wertz, 2002). The diversity and specificity of bacteriocins reflect a complex evolutionary arms race, where both producers and targets continuously evolve mechanisms of resistance and counteraction (Riley, 1998).

Another critical pathway in bacterial warfare is the secretion of small molecules that disrupt cellular processes in competing bacteria. For instance, some species of Streptomyces produce antibiotics like streptomycin that can inhibit protein synthesis in other microbes, a strategy that enhances their competitive advantage in soil environments rich in microbial diversity (Watve *et al.*, 2001). This antibiotic production is not only a defensive mechanism but also a means of pre-emptive attack to secure resources and space.

Quorum sensing, a stimulus-response system linked to population density, is a more indirect strategy in bacterial warfare. It enables bacteria to adjust gene expression according to their population size, controlling virulence factor production, biofilm formation, and antibiotic-related gene expression (Miller and Bassler, 2001). This coordination influences bacterial competition, allowing bacteria to fine-tune their strategies depending on the density of their own and rival populations.

Biofilm formation is another strategy linked to bacterial competition. By forming biofilms — complex communities of bacteria embedded in a self-produced protective matrix — bacteria can resist physical removal and chemical attack, including antibiotics and

bacteriocins produced by competitors (Costerton, Stewart and Greenberg, 1999). These biofilms act as microbial fortresses, allowing their constituent cells to survive in hostile environments and outcompete single, free-floating bacterial cells.

The interaction between bacterial pathogens and the host immune system can also be viewed as a form of bacterial warfare. Pathogenic bacteria have developed various strategies to evade or manipulate the host immune response, such as modifying surface antigens to evade detection, releasing immune-modulatory proteins, and directly targeting immune cells (Finlay and McFadden, 2006). These interactions are crucial in human bacterial infections, where the outcome often hinges on the pathogen's capacity to resist or undermine the host's immune response.

In the intricate landscape of bacterial competition, another significant aspect of bacterial warfare are the Contact-Dependent Inhibition (CDI) systems. These systems add a layer of complexity to the microbial interactions by enabling direct, targeted attacks against neighbouring cells, thus extending the battlefield from secreted chemicals to direct cell-to-cell contact (Aoki *et al.*, 2005; Ruhe *et al.*, 2013). This transition from indirect to direct forms of bacterial combat underlines a continuum in microbial warfare strategies. The shift to CDI highlights the adaptability and tactical diversity of bacteria, reflecting a progression from employing environmental mediators of conflict to engaging in close-quarters combat via direct cell-to-cell interactions. This progression not only illustrates the versatility and depth of bacterial strategies to outcompete rivals but also emphasizes the evolutionary pressures that drive the sophistication of these microbial interactions. The study of CDI systems, therefore, not only enriches our understanding of bacterial ecology and evolutionary biology but also provides potential avenues for novel antimicrobial strategies and therapeutic applications, capitalizing on the very mechanisms bacteria use to dominate their environments.

1.1.1. CDI Toxins and their Effector Domain

CDI systems represent sophisticated means by which bacteria can directly interfere with the growth of neighbouring cells. These systems are particularly intriguing due to their reliance on direct cell-to-cell contact, facilitating the delivery of polymorphic toxic effector proteins that can induce cell death or dormancy in target cells (Aoki *et al.*, 2005).

CDI toxins are typically organized into three domains: the N-terminal domain, which is involved in binding to the target cell, the central domain which facilitates the toxin's translocation into the target cell, and the C-terminal effector domain, which harbours the toxic activity (Ruhe *et al.*, 2013). This modular organization allows for considerable diversity among CDI toxins, as different bacteria can possess different effector domains even when the other domains are conserved.

The effector domain is the key functional component of the CDI system. Upon delivery into the target cell, the effector domain interacts with specific intracellular targets to disrupt key cellular processes, leading to cell death or stasis. The mechanisms by which these effector domains induce cytotoxicity are diverse, including DNA/RNA degradation, inhibition of protein synthesis, and disruption of cell wall synthesis, among others (Zhang, Iyer and Aravind, 2011).

Focusing on the effector domain, its involvement in cell death is particularly crucial for understanding its potential applications and risks. In Escherichia coli, for instance, the effector domain of the CdiA protein has been shown to carry an RNase activity, which cleaves tRNA within the target cell, effectively shutting down protein synthesis and leading to cell death (Aoki *et al.*, 2005). This specific action not only highlights the lethal capabilities of CDI systems but also underscores their precision in targeting molecular processes within the cell.

Moreover, the diversity of effector domains across different bacterial strains suggests a complex evolutionary background where different bacteria have adapted unique mechanisms to combat competitors effectively. This diversity is indicative of the evolutionary arms race mentioned in the general context of bacterial warfare, where organisms continuously develop new methods to outcompete rivals and survive in competitive environments (Braun and Patzer, 2013).

1.2. CDI Toxins "*Shipment*" into the Target Cell

When infecting the target cell, CDI has to cross the outer and inner membrane, respectively. The passage through the outer membrane is better known, whereas mechanism by which CDI toxins cross periplasm and the inner membrane is not yet fully understood. The key components in this process include the CdiA and CdiB proteins. The CdiA protein on the CDI-utilizing bacteria is translocated into the periplasm via the Sec-dependent pathway and

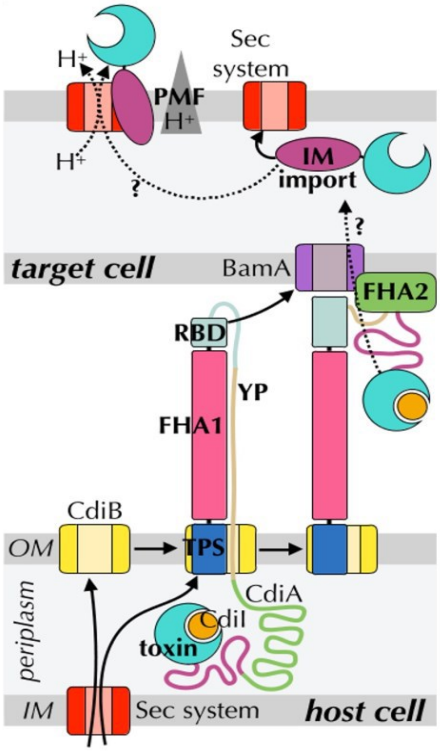

Figure 1: This diagram, provided by Dr William Allen (University of Bristol, UK), illustrates the mechanism of CDI in bacteria, depicting the secretion and targeting pathway of the toxins across bacterial membranes.

then through the outer membrane by CdiB (see Figure 1). Structurally, CdiA comprises N-terminal FHA-1 hemagglutinin peptide repeats which form an elongated β-helix, projecting the protein several hundred angstroms away from the bacterial surface. This extension is crucial as it positions the receptor-binding domain (RBD) of CdiA to effectively interact with specific receptors on the target cell. Upon successful binding to a receptor such as BamA (Aoki *et al.*, 2008), the C-terminal region of CdiA, which harbours the toxic domain (CdiA-CT), is transferred into the target cell (Ruhe *et al.*, 2018). Studies have shown that the transfer of CdiA-CT utilizes the SecYEG translocon of the target cell in a retrograde fashion (Jones *et al.*, 2021) and the proton motive force (Ruhe *et al.*, 2014), however, the mechanism of the transport remains still unclear. Critical to this whole process is also CdiI, an immunity protein produced by the CDI-utilizing bacteria to protect itself from its own toxin, thereby preventing self-inhibition and ensuring survival while inhibiting competitors (Aoki *et al.*, 2010).

1.3. The Sec-dependent Protein Translocation

Protein translocation across lipidic membranes is a fundamental process in all living cells, being crucial for the correct localization, proper function, and highly specific interaction of proteins. Consequently, proteins destined for secretion or insertion into membranes must navigate through or integrate into the hydrophobic core of the lipid bilayer, which in most cases imposes a large energetic barrier. This process is highly regulated and requires specific machinery to ensure specificity and efficiency. Throughout evolution, specialized and sophisticated mechanisms of transport have been developed for this (Verner and Schatz, 1988). Among the various systems facilitating protein translocation, the secretory (Sec) pathway is one of the most well-studied and universally conserved across most of the domains of life, specifically catalysing the transport of the majority of proteins in bacteria (Blobel *et al.*, 1979; Rapoport, 2007).

The Sec-dependent translocation pathway is vital for moving proteins across the bacterial cytoplasmic membrane, facilitating both protein secretion and membrane protein integration. This pathway centres on the Sec translocase, a protein-conducting channel formed by the SecYEG complex, which is driven by the ATPase activity of SecA (Mori and Ito, 2001). The newly synthesized preproteins are directed to the translocase commonly using a short signal sequence which binds to the SecYEG (de Keyzer, van der Does and Driessen, 2003). Once bound, the preproteins are driven across the membrane through the channel formed by the SecYEG complex, in a process that is tightly regulated and highly efficient (den Blaauwen and Driessen, 1996; Driessen and Nouwen, 2008; Fessl *et al.*, 2018; Komarudin Amalina Ghaisani and Driessen Arnold J. M., 2019).

The dynamics of the Sec machinery's interactions during the translocation process, which include protein insertion and membrane integration, are complex and multifaceted. Key interactions occur between SecA and the SecYEG integral membrane components, facilitating the transmembrane movement of preproteins. This pathway also interacts with other cellular components to achieve efficient protein translocation and integration, demonstrating its critical role in bacterial cell function (Mori and Ito, 2001).

Understanding the structural and mechanistic aspects of the Sec-dependent pathway, specifically in relation to the SecYEG complex and SecA, is essential for a comprehensive grasp of bacterial protein translocation. This understanding is further deepened by studies on the interaction of SecA with ribosomes, which facilitate posttranslational translocation in bacteria (Huber *et al.*, 2011), and research on the dynamic action of the Sec machinery during initiation, protein translocation, and termination (Fessl *et al.*, 2018).

1.3.1. Structure and Dynamics of SecYEG

The SecYEG translocon is a heterotrimeric complex comprising the subunits SecY, SecE, and SecG. The functionality of this complex is inherently linked to its structural properties (Figure 2 (Collinson, Corey and Allen, 2015)).

SecY, the largest subunit containing 10 transmembrane helices (TMHs), forms the core of the translocon channel (Berg *et al.*, 2004). Its structure is described as an hourglass-shaped pore, typically closed by a ring of six hydrophobic amino acids, predominantly isoleucine, located in the channel's interior (Cannon *et al.*, 2005). This arrangement keeps the lateral gate (LG) narrowly closed, preventing molecular

passage. The 'plug' domain, located between TMHs 1 and 2, is a crucial structural feature that blocks the channel entrance in its inactive state, playing a significant role in regulating translocon activity (Mori and Ito, 2001).

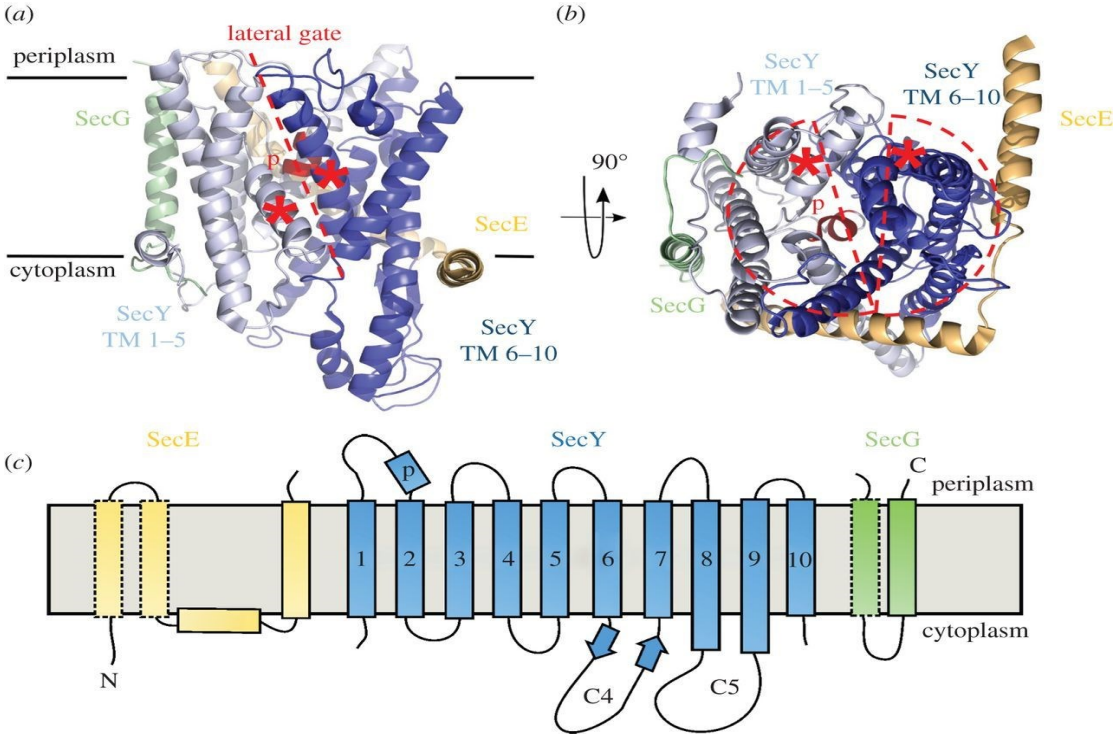

Figure 2: Structure of SecYEβ complex from *M. jannaschii*. (a) SecYEβ viewed from the side, in position in the lipid bilayer (black lines). TMHs 1–5 of SecY are coloured light blue, TMHs 6–10 dark blue, with the plug helix (labelled 'p') in red, SecE in wheat colour and SecG/β in green. The LG is indicated with a dashed red line, and the lateral gate (LG) helices are marked with asterisks. Structural data from (Berg et al., 2004). (b) As in panel (a) but viewed from the cytoplasm. Red semicircles have been superimposed to indicate the separate halves of SecY. (c) Schematic of E. coli SecYEG. SecE is in yellow, SecY in blue with the TMHs numbered and the primary cytoplasmic loops (C4 and C5) and plug (p) marked, and SecG is green. Conserved regions are shown in solid lines and the non-conserved in dashed lines. Figure and caption taken from (Collinson, Corey and Allen, 2015).

Activation of the SecYEG translocon involves significant structural rearrangements. The pore must widen, necessitating the disruption of the hydrophobic ring, and the TMH 2 and 7/8 helices must separate to expand the LG sufficiently for protein passage (Tsukazaki *et al.*, 2008).

SecE, with 1-3 TMHs, and the less conserved SecG, enhance the overall efficiency of protein translocation. SecE's C-terminal helix and a horizontally oriented amphipathic helix contribute to maintaining the proximity of SecY's domains, ensuring a narrow LG in the inactive state (Mori *et al.*, 2010). Notably, SecG's presence markedly increases the translocation process's efficiency, as evidenced both in vivo and in vitro (Gari *et al.*, 2013).

1.3.2. Mechanism of Protein Translocation by SecYEG

Proteins destined for transport through the SecYEG translocon complex are typically targeted to it via a short signal sequence located at their N-terminus. They may either be translocated completely across the membrane (secretion) or integrated into the membrane (insertion) (Collinson, 2019). Notably, during this process, the protein must remain in an unfolded conformation to facilitate passage through the translocon (Arkowitz, Joly and Wickner, 1993; Mori and Ito, 2001; Koch *et al.*, 2019).

The translocation can occur co-translationally or post-translationally. Co-translational translocation is a universal and well-understood process (Collinson, 2019). The signal recognition particle (SRP) identifies the signal sequence, binds to it, and guides the resultant complex to the SRP receptor in the membrane (Nilsson *et al.*, 2015). The ribosome-nascent chain complex (RNC) is then directed to the Sec machinery, where the transport of the growing polypeptide chain occurs co-translationally (Bange, Wild and Sinning, 2007). Generally, such polypeptides are inserted into the membrane, while secretion processes typically occur post-translationally (Fessl et al., 2018).

In contrast, the post-translational mechanism is more complex, as it is not coupled to protein synthesis. Bulkier polypeptide regions formed due to folding in the cytoplasm necessitate a distinct system for secretion through the SecYEG complex. The post-translational translocation mechanism involves the SecA ATPase interacting with the SecYEG integral membrane components, facilitating the transmembrane movement of newly synthesized preproteins (Mori and Ito, 2001). This intricate interplay between SecA and SecYEG, along with the necessity for proteins to remain unfolded, highlights the complexity and sophistication of the bacterial protein translocation system.

Single-molecule studies using Förster resonance energy transfer (FRET) have shed light on the dynamics of the SecY pore and the kinetics of the translocation process. These studies have revealed that SecA, the signal sequence, the pre-protein, and ATP hydrolysis each play specific roles in unlocking and opening the Sec channel, preparing it for transport (Fessl *et al.*, 2018). This preparation sets the stage for protein transport, which unfolds in two distinct phases. Initially, there is an initiation phase

that occurs independently of the substrate's length, where the basic setup and entry of the pre-protein into the channel are achieved. Following this, the process transitions into a length-dependent transport phase, wherein the actual movement of the substrate through the channel is influenced by its length.

1.4. Proposed Mechanism of Retrograde Translocation of CdiA Toxins by SecYEG

In an intriguing reversal of its typical function, the SecYEG translocon is not only crucial for exporting proteins across the cell membrane but also plays a pivotal role in the retrograde translocation of CDI toxins (Jones *et al.*, 2021). This process essentially operates in the opposite direction to the normal translocational flow, where proteins are exported out of the cell. In the context of CDI toxins, the SecYEG machinery is co-opted to facilitate their entry into target cells, marking a significant deviation from its usual role. This retrograde utilization of the SecYEG translocon allows the toxins to be transported from the periplasmic space to the cytoplasm of the cell, effectively using the cell's own machinery against itself. Studies have shown that mutations in SecY, the core component of the SecYEG complex, can drastically reduce the effectiveness of this toxin translocation, thereby confirming the critical role of the SecYEG translocon in the retrograde movement of these bacterial toxins (Jones *et al.*, 2021). Furthermore, unpublished data from Dr Tomas Fessl showed the CDI transport resistant mutant has altered dynamics of SecY plug domain, inferring SecYEG dynamics to be a key player in the initiation stage of toxin import.

The hypothesis studied here is that the transport involves several stages (see Figure 3). (i) Initiation - when the dynamic plug and lateral gate domains of the SecYEG complex undergo spontaneous opening, allowing the import domain of the CdiA toxin to bind within the periplasmic cavity. Subsequently, mimicking a signal peptide, the toxin's import domain intercalates into the lateral gate, which triggers further channel opening. (ii) & (iii) Transport stage - this opening, paves the way for an unfolded effector domain, connected by a flexible linker, to cross the membrane and enter the cytoplasm. (iv) Refolding – after crossing the membrane, the effector domain folds to its active conformation.

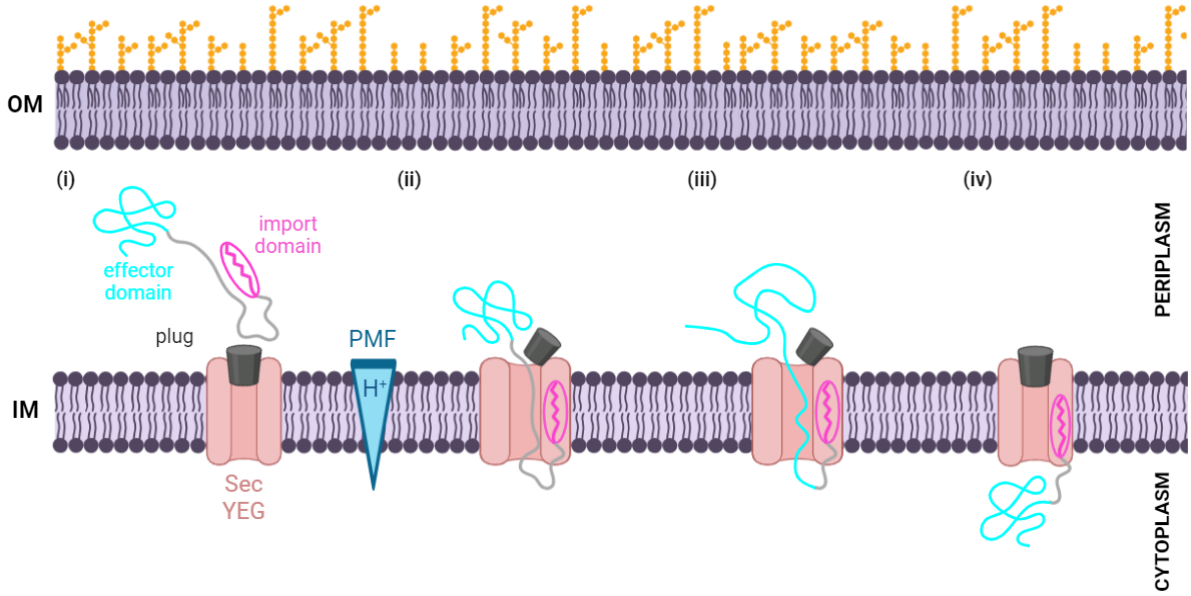

Figure 3: Sequential Stages of CdiA Toxin Retrograde Transport via the Sec Translocon. (i) Binding Initiation: The CdiA toxin's import domain binds to the SecYEG translocon at a moment when the plug domain spontaneously opens. (ii) Translocon Engagement: The import domain's alpha-helix penetrates the lateral gate of SecYEG, resembling a signal peptide, to facilitate the opening of the translocation channel. (iii) Effector Unfolding: The effector domain, tethered by a flexible linker, begins to unfold directionally due to the specialized anisotropic unfolding properties of the CdiA toxins. (iv) Cytoplasmic Entry: The fully unfolded effector domain is transported through the Sec channel into the cytoplasm where it refolds into a functional conformation, completing the retrograde transport process.

1.4.1. Necessity for Unfolding of the Effector Domain

The unfolding of toxins is known to be a critical prerequisite for their retrograde transport through the SecYEG translocon. The narrow pore of the SecYEG complex, approximately 1-2 nm in diameter (Bonardi *et al.*, 2011), necessitates that proteins must first unfold before they can be translocated (Jones *et al.*, 2021). Since it is assumed that the system has to be in an unfolded state prior to transport of the effector domain across the inner membrane, the identification of how this is achieved in periplasm, an environment without ATP, is needed. The usual culprits, such as periplasmic chaperones and pH are discussed in text below.

Genetic evidences have shown the PpiD and YfgM periplasmic chaperones play a crucial role in the structural stability of proteins during their translocation through the SecYEG translocon (Jones *et al.*, 2021). It is speculated that these chaperones assist in maintaining the unfolded state required for CDI toxin import across the inner membrane, acting as stabilators to maintain a transport-competent conformation of the CdiA toxin, thus facilitating the effective passage through the narrow SecYEG pore. (Jones *et al.*, 2021)

On the other end, pH is suspected to be another factor that might be involved in the preparing of the effector domain for the retrograde translocation. The stability of proteins, including toxins undergoing retrograde transport, are profoundly influenced by pH levels. Different pH environment can provoke instability and induce conformational changes leading to unfolding. Such changes are primarily driven by the altered ionization states of amino acid residues, affecting both intra- and intermolecular bonds critical for maintaining structural integrity (Tollinger *et al.*, 2003).

Due to these changes in charge distribution, the stability of the protein can be affected. At certain pH, attractive forces between oppositely charged residues can stabilize the protein structure by promoting favourable columbic interactions, however, at distinct pH conditions, excessive protonation or deprotonation may lead to repulsion between similarly charged residues, destabilizing the protein, and fostering conformational changes (Yang and Honig, 1993).

Given the differing pH levels in the periplasm and cytoplasm (Krulwich, Sachs and Padan, 2011), investigating the mechanical unfolding of the CdiA toxin's effector domain under periplasmic pH conditions, which mimics the environment where this process occurs, and comparing it to cytoplasmic pH, could reveal crucial insights into the role that it plays in retrograde transport through the SecYEG translocon. Grasping how this unfolding process occurs is vital for understanding the mechanism of CDI toxins transport.

1.5. Importance of the Study

In light of the escalating crisis of antibiotic resistance, which the World Health Organization (WHO) identifies as a formidable threat to global health (Lin *et al.*, 2015), the significance of studying bacterial warfare mechanisms becomes increasingly vital. Antibiotic resistance compromises the efficacy of conventional treatments, heralding a potential post-antibiotic era where even minor infections could prove lethal (Morehead and Scarbrough, 2018). In this context, the comprehensive understanding of CDI systems — especially the process of CDI toxins' retrograde translocation — presents a promising frontier in the development of novel antibacterial strategies. These bacterial-derived toxins, capable of precise and targeted bacterial cell inhibition, could serve as a blueprint for creating

sophisticated therapeutics that circumvent traditional resistance pathways. Moreover, elucidating the mechanisms by which these toxins interact with and penetrate bacterial cells could inform the design of drugs that mimic these natural processes, offering a strategic advantage in combating pathogenic bacteria.

Additionally, this research could profoundly influence the understanding and manipulation of the gut microbiome, where maintaining bacterial balance is crucial for health yet is often disrupted by conventional antibiotics (Hills Jr. *et al.*, 2022). By harnessing the intrinsic bacterial mechanisms like CDI toxins, it may be possible to selectively target pathogenic bacteria while preserving beneficial microbes, thus sustaining an essential component of human health (Prestinaci, Pezzotti and Pantosti, 2015). Therefore, the study of CDI toxin dynamics not only addresses an urgent public health need but also opens avenues for innovative treatments that could revolutionize our approach to infectious diseases and microbiome management.

2. Aims of the Thesis

The main aim of this thesis research is to analyse the structural stability and energetic landscape of the effector domain of CdiA toxins in cytoplasmic and periplasmic conditions using Molecular Dynamics (MD) simulations. The study seeks to understand the unfolding profile of the effector domain through pulling simulations, specifically investigating whether pH changes have a destabilizing effect on the domain. Additionally, by calculating the Potential of Mean Force (PMF) derived from the umbrella pulling simulations, the research aims to elucidate the energy landscape during the unfolding process, providing deeper insights into how the effector domain stability of these toxins adapts in response to periplasmic pH.

Another important goal of this project is to develop a reliable assay for studying CdiA toxin import via SecYEG suited to single-molecule Förster Resonance Energy Transfer (smFRET). This assay strives to enable future detailed investigations into the mechanism of retrograde translocation, aiming to improve the tools available for accurate and physiologically relevant future studies of CdiA toxin import *in vitro*.

3. Materials & Methods

3.1. CdiA Toxin Primary Sequence

This study specifically focused on the effector domain (residues 244-361) of the CdiA toxin, employing it as the primary subject for an in-depth analysis. The fundamental amino acid sequence of the toxin used in this thesis was provided by Dr Tomas Fessl and is shown in the figure below (Figure 4: produced by Protter (Omasits *et al.*, 2014)).

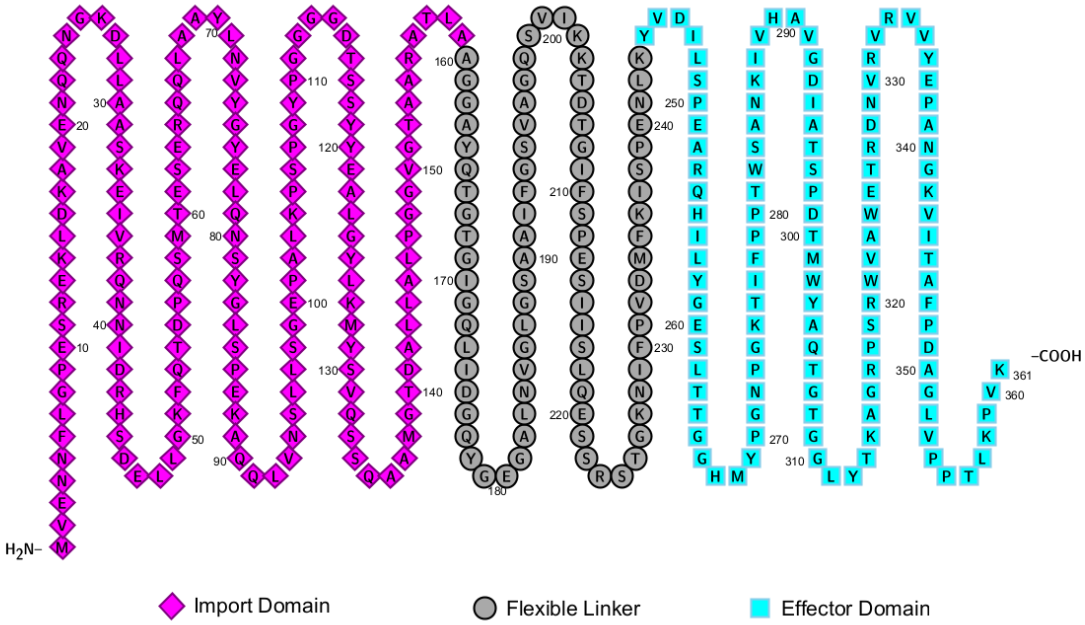

Figure 4: Amino acid sequence of the CdiA toxin. The import domain (res. 1-159) is shown in magenta, the flexible linker (res. 160-243) in grey, and the effector domain (res. 244-361) in cyan.

3.2. CdiA Toxin and CdiA — SecYEG Complex Structure Prediction with AlphaFold2

AlphaFold2, a neural network-based model, has revolutionized the field of protein structure prediction by providing atomic accuracy even when no similar structure is known (Jumper *et al.*, 2021). Its ability to incorporate physical and biological knowledge about protein structure, using multi-sequence alignments, enables accurate and efficient prediction of three-dimensional structures solely from amino acid sequences. Additionally, advancements in AlphaFold2 have demonstrated its utility in expanding the structural coverage of protein sequence space, greatly aiding in the structural bioinformatics domain (Varadi *et al.*, 2022).

Since no precise experimentally assigned 3D structure is known, this study utilized the AlphaFold2 model capabilities to predict the structure of CdiA Toxin, based on the protein sequence depicted in Figure 4, aiming to elucidate its structural stability by further analysis with MD simulations. Additionally, the study extended to predicting the structure of CdiA toxin in complex with SecYEG to investigate the interactions of the toxin's import domain with the translocon.

Following the prediction, using PyMol (Schrödinger, LLC, 2015), a 3D structure editing and visualization tool, a new *pdb* file was generated from AlphaFold2's best model of the CdiA toxin. This file specifically contained only the effector domain of the toxin and was used in the performed MD simulations and further analysed in the later phases of this study.

3.3. Ionization State Analysis of Effector Domain Residues via pK_a Prediction

The pKa prediction of the effector domain's residues was a critical component in understanding the biochemical properties and stability of the CdiA toxin under different pH conditions. This step was crucial because the protonation states of amino acids, such as Asp, Glu, Tyr, Arg, Lys, and His, can significantly influence the protein's structure and stability. In the cellular environment, these residues often undergo shifts in their pK_a values due to their interactions with nearby amino acids and the solvent, affecting the overall charge and conformation of the protein.

The pK_a prediction of the toxin’s effector domain predicted 3D structure in this study was conducted using PROPKA 3.0, a reliable and widely used tool for predicting the pK_a values of ionizable groups in proteins. PROPKA 3.0 employs an empirical approach to estimate the pK_a shifts, considering both internal and surface residues within proteins (Olsson *et al.*, 2011). The PROPKA 3.0 algorithm effectively interpolates between internal and surface residues to reduce the number of outliers and the sensitivity to minor structural changes, which can significantly affect the pK_a prediction of ionizable residues in proteins (Olsson *et al.*, 2011; Søndergaard *et al.*, 2011).

The pK_a values predicted for the effector domain by PROPKA 3.0 were subsequently employed in the later molecular dynamics simulations to accurately mimic the toxin's behaviour under the selected pH environments.

3.4. GROMACS Molecular Dynamics Simulations

GROMACS (GROningen MAchine for Chemical Simulations) is a widely used software suite for molecular dynamics simulation, originally developed at the University of Groningen in the early 1990s. GROMACS is compatible with several force fields like GROMOS, OPLS, and AMBER, and can handle various models and constraints (Van Der Spoel *et al.*, 2005), making it a very useful tool for simulation of different conditions. This allows researchers to perform simulations that are not only detailed and extensive but also highly customizable, catering to a broad spectrum of molecular dynamics research needs.

In this study, the structural stability and energetic landscape of the CdiA toxin effector domain were analysed under varying pH conditions using GROMACS MD simulations. Through pulling simulations, the general unfolding pathway was initially studied, followed by an examination of how different pH values induced changes to this pathway. This detailed analysis facilitated the identification of structural interactions crucial for the domain's integrity and allowed for the calculation of the PMF through umbrella sampling. Such an approach was vital for delineating the energy landscape of mechanical unfolding, offering insights into the stability of the effector domain of CdiA toxins in different pH conditions.

The entire all-atomistic MD simulations were performed in GROMACS MD software package (version 2023.2) supported by CUDA GPU accelerated libraries, which significantly enhanced the computational efficiency, enabling more complex systems and longer simulation timescales to be investigated with greater precision (Abraham *et al.*, 2015, 2024). The workflow depicted in Figure 5 outlines a structured approach to the steps performed for each MD simulation run.

3.4.1. Simulations Setup

In the simulations setup for this thesis, the GROMOS96 53a6 force field was employed to model the molecular interactions. The water model used was the simple point charge (SPC) type. Critical residues were protonated, as specified in Table 1 (see section 4.3), ensuring an accurate representation of their chemical states under the simulated conditions. The toxin was centred in a simulation box with dimensions of 6 x 6 x 85 nm. To achieve electrostatic neutrality and mimic physiological conditions, ions were added to the system to reach a concentration of 0.1 M NaCl. In this process,

a corresponding number of water molecules were replaced by sodium and chloride ions, maintaining the integrity of the simulation environment. The entire system at the end consisted of approximately 290,000 atoms.

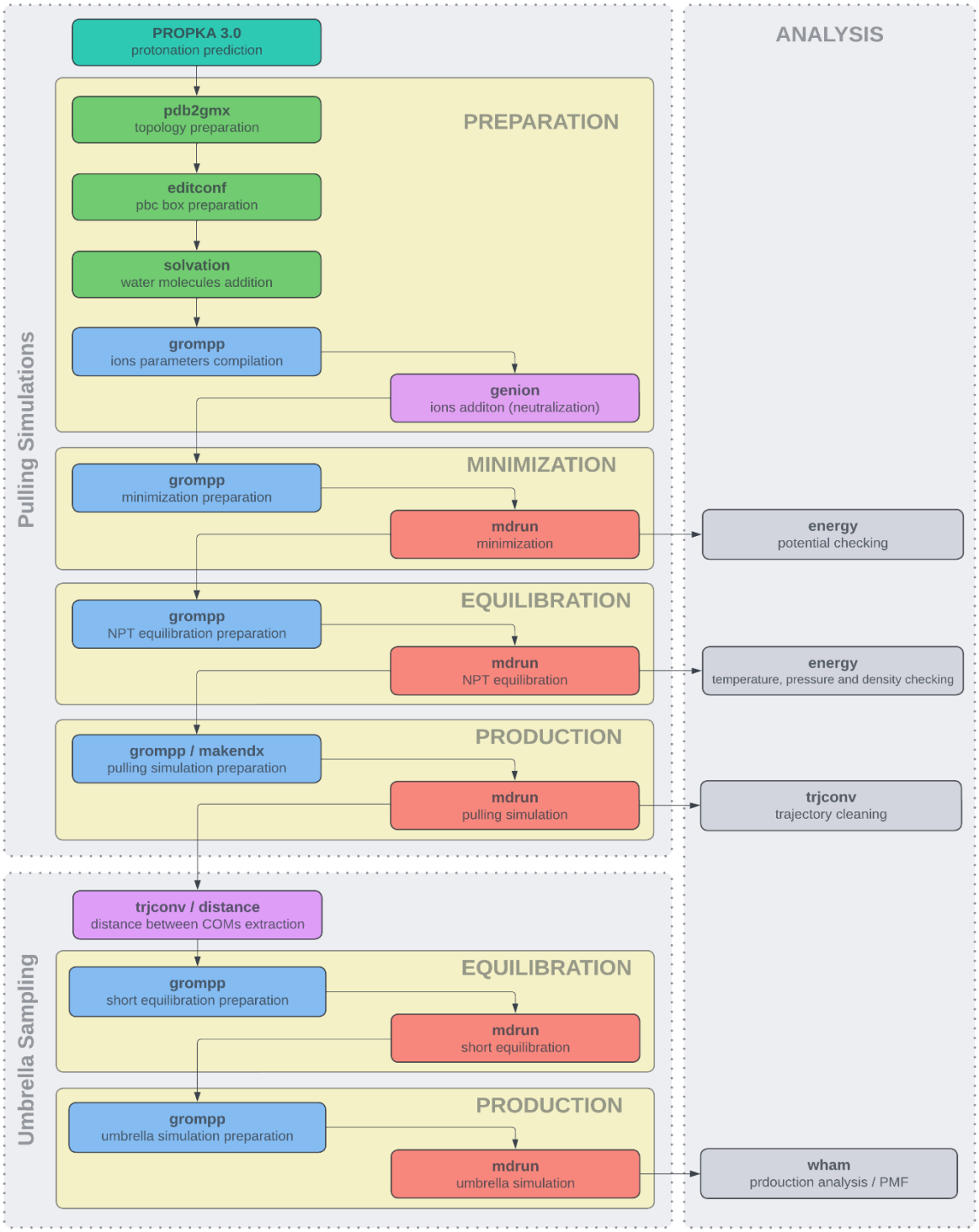

Figure 5: A detailed roadmap of the conducted MD simulations using GROMACS, outlining two workflows for the two different types of simulations run: pulling simulations and umbrella sampling. The processes included the initial preparation stage with protonation, solvation, and ion addition, followed by energy minimization and NPT equilibration. For pulling simulations, the production run led to trajectory cleaning and lastly to various production analysis. The umbrella sampling workflow entailed a short equilibration and the umbrella simulation run, concluding with the production analysis and the PMF calculations.

3.4.2. Minimization

The minimization of the system was a critical step in the simulation process. For the minimization phase, the *steepest descent* method was employed, using a maximum force threshold of 1000 kJ mol^{-1} nm^{-1} and an energy step size of 0.01. The process was limited to a maximum of 50,000 steps. Key parameters included a neighbour list update frequency of every step, a short-range neighbour list cutoff of 1.4 nm, and the use of the *Particle Mesh Ewald (PME)* method for treating long-range electrostatic interactions. *Periodic Boundary Conditions (PBC)* in all directions (xyz) were applied.

3.4.3. Equilibration

The equilibration phase involved the use of the *leap-frog* integrator for a total of 50,000 steps, corresponding to 100 ps with a time step of 2 fs. Position restraints were applied to the toxin. Output of coordinates, velocities, energies, and log updates were set to occur every 2 ps. *LINCS* algorithm was used for constraining all bonds, ensuring system stability. Short-range interaction cutoffs for electrostatics and van der Waals forces were set at 1.4 nm.

Temperature coupling was managed via the *v-rescale* method with a reference temperature of 310 K for both protein and non-protein groups. Pressure coupling was employed using an isotropic approach with a reference pressure of 1.0 bar and a time constant of 2.0 ps. Velocity generation was included to initiate the system at 310 K, with dispersion correction for energy and pressure and centre-of-mass (COM) motion removal implemented to maintain the system's stability.

3.4.4. Pulling Simulations

Following the confirmation of system stability in terms of temperature, pressure, potential, and density using the *gmx energy* module of GROMACS, pulling simulations were conducted (refer to Figure 19 in Appendix A section for an overview of conducted pulling simulations). This phase of the study aimed to explore the mechanical properties and responses of the toxin's effector domain under applied forces. In the pulling simulations, two outermost residues of the toxin's effector domain were selected as the target for the applied force. The method used was *umbrella*

pulling, a well-established technique in molecular dynamics simulations for studying the response of molecules to mechanical forces (Bartels and Karplus, 1997). This technique involves applying a controlled force to specific parts of the molecule and observing the resulting changes in its configuration.

The simulation was run using a *leap-frog* integrator with a time step of 2 fs, amounting to a total simulation time of 40 to 4000 ns, depending on the pulling rate. Coordinates were saved every 10 ps, with energy parameters being recorded at the same interval. The LINCS algorithm was employed to constrain all bonds, ensuring structural integrity under the force application.

The pulling force was applied along the Y-axis, with a pulling rate of 0.01, 0.001 or 0.0001 nm/ps and a force constant of 1000 kJ mol^{-1} nm^{-2}. This setup was designed to gently and progressively pull the specified residues apart, allowing for detailed observation of the structural changes occurring in the effector domain.

3.4.5. Umbrella Sampling Simulations

In this project, umbrella sampling was a key technique employed to estimate the free energy profiles of molecular interactions (refer to Figure 19 in Appendix A section for an overview of conducted umbrella sampling). This procedure involved the use of the GROMACS tool for extracting umbrella windows, where the COM distances between specific groups were calculated using *gmx distance* command.

In the short equilibration phase, *leap-frog* integrator was used with a time step of 2 fs, spanning 50,000 steps, equivalent to 100 ps. Notably, position restraints were applied to the protein, similar to the prior equilibration phase during setup. The LINCS algorithm was used for constraining all bonds. Short-range interaction cutoffs for electrostatics and van der Waals forces were maintained at 1.4 nm, with the *PME* method for long-range electrostatics. Temperature coupling was managed via the *v-rescale* method with a reference temperature of 310 K, and pressure coupling was executed using an isotropic approach with a reference pressure of 1.0 bar.

Following this, the production runs involved a longer simulation time of 3 ns with a consistent time step of 2 fs. The *Nose-Hoover* method was used for temperature coupling in two groups (Protein and Non-Protein), while pressure coupling employed

the *Parrinello-Rahman* method. The constraints and cutoff schemes remained consistent with the equilibration phase, ensuring continuity in simulation conditions.

These parameters facilitated the effective sampling of the energy landscape, enabling the precise calculation of free energy differences critical to the study's objectives.

3.4.6. Post-Simulation Data Analysis

In the post-simulation analysis of the molecular dynamics study, a comprehensive approach was employed to interpret the data obtained from GROMACS simulations. The analysis focused on understanding the structural behaviour and energy profiles of the molecules under two different pH conditions and three different pulling rates.

The first aspect of the analysis involved the utilization of the *gmx dssp* module to generate DSSP (Define Secondary Structure of Proteins) graphs. This module allowed for the detailed observation of the secondary structure elements, such as alpha-helices and beta-sheets, within the effector domain of the toxin during the pulling. To account for the variability in the simulations, the DSSP profiles were averaged across replicas for each pH condition and pulling rate. This averaging approach ensured a robust representation of secondary structure behaviour under different simulation conditions. The results were visualized using Jupyter Notebook, providing a clear graphical representation of the secondary structure evolution during the simulations.

A crucial part of the analysis involved calculating the rolling averages of the force required for unfolding. This was done for each set of replicas under the various pH conditions and pulling rates. The computation of standard deviation alongside the averages was instrumental in illustrating the variability and consistency of the unfolding forces across different simulation scenarios. These calculations provided a quantitative measure of the mechanical strength and stability of the toxin structures under study.

The PMF was calculated using the *gmx wham* module (Hub, De Groot and Van Der Spoel, 2010). This analysis was pivotal in understanding the energy landscape of the molecular systems. It gave insights into the stability and conformational

preferences of the molecules under study. The PMF calculations were complemented by displaying the distribution of counts from the umbrella sampling simulations.

Finally, the unfolding pathway of the toxin was studied and displayed. This was achieved by employing *trjconv* for trajectory processing, followed by visualization of key frames using PyMol. These visualizations were critical in providing a detailed and clear representation of the structural changes occurring during the unfolding process. This part of the analysis not only offered a visual understanding of the unfolding events but also complemented the quantitative analyses conducted earlier.

Each of these analytical approaches, employed in synergy, provided a thorough understanding of the molecular dynamics of the effector domain under various conditions. This comprehensive analysis not only illuminated the intricate details of molecular behaviour but also laid a robust foundation for understanding the mechanistic aspects of toxin's effector domain unfolding and stability.

3.5. Single-Molecule FRET Experiments

Single-Molecule FRET (smFRET) is an essential tool for understanding structural dynamics in biological systems, particularly in the context of protein research. smFRET works on the principle of energy transfer between two fluorophores - a donor and an acceptor - which are specifically attached to a biomolecule. The energy transfer, and thus the FRET efficiency, is highly sensitive to the distance between these fluorophores, typically in the range of 2-10 nm (Lerner *et al.*, 2021). This sensitivity makes smFRET a powerful *"spectroscopic ruler"* for studying conformational changes and interactions at the single-molecule level in a near-physiological environment.

smFRET has been instrumental in providing insights into the dynamics of macromolecules under biologically relevant conditions. Since its inception, it has been used to uncover fundamental biological mechanisms in DNA maintenance, replication, transcription, translation, protein un/folding, enzymatic function, and membrane transport (Lerner *et al.*, 2018; Huynh and Sengupta, 2022). This technique has confirmed existing hypotheses and revealed new ones, enhancing the understanding of these crucial processes.

In this research, the smFRET methodology was employed to devolped an assay that indirectly monitor the import of CdiA toxin via the SecYEG complex within a near-

physiological context. The assay capitalized on a double fluorescently labeled RNA hairpin as a proxy for toxin translocation across the proteoliposome membrane (see Figure 6). The left panel of the figure illustrates the intact encapsulated RNA hairpin, resulting in high FRET efficiency. Upon the introduction and successful retrograde translocation of the CdiA toxin, as shown in the right panel, the RNA hairpin is cleaved within the proteoliposome, leading to the absence of FRET signal. This drop in FRET efficiency is indicative of the hairpin's cleavage and, by extension, the toxin's passage, providing an indirect but powerful indication of successful translocation. The development and implementation of this assay marked a significant step forward in the study of toxin import, allowing for more accurate investigations of the transport mechanisms at the single-molecule level.

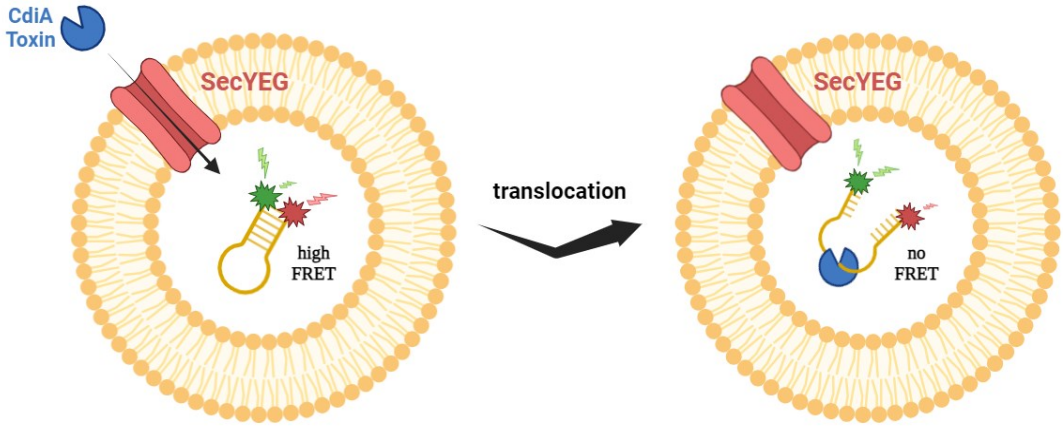

Figure 6: Schematic representation of the smFRET assay to monitor the entry of CdiA toxin into SecYEG-enriched proteoliposomes and subsequent RNA hairpin cleavage. The left panel illustrates the proteoliposome with high FRET efficiency due to the intact FRET-labelled RNA hairpin. The right panel shows the expected outcome after retrograde translocation of the CdiA toxin, resulting in the cleavage of RNA hairpins and a subsequent decrease in FRET efficiency, indicated by the separation of the FRET fluorophores.

3.5.1. Liposomes Preparation

In this study, liposomes encapsulating RNA hairpins were prepared using *E. coli* Polar Lipid Extract sourced from Avanti Lipids (Product #100600), which was stored in chloroform at -80°C to ensure lipid integrity. The lipid extract was vacuum rota-vaporized to effectively remove the chloroform, after which it was reconstituted in buffer solution. The base buffer for all liposome preparations was TKM (20 mM TRIS + 50 mM KCl + 2 mM $MgCl_2$) mixed with 2 mM TROLOX and 1% BME (β-mercaptoethanol). The final lipid concentration for each suspension was maintained at 3.125 mg mL^{-1}.

For the liposomes encapsulating RNA hairpins, the RNA was added to the lipid mixture to achieve a final concentration of 60 pM. The suspension was then subjected to 30 minutes of extensive vertexing to ensure homogeneous distribution of RNA hairpins within the lipid mixture. Liposome formation was carried out using the Mini-Extruder Extrusion Technique, as per the protocol provided by Avanti Lipids. This process involved an extrusion device consisting of two syringes connected by a holder with a 100 nm pore-sized polycarbonate membrane. The suspensions were each extruded through the membrane 11 times, a procedure crucial for achieving liposomes of reduced size and enhanced size uniformity.

This methodical extrusion process was pivotal for producing uniformly sized liposomes with a controlled size distribution, a feature of paramount importance for the study's focus.

3.5.2. RNA Hairpin Structure

For the conducted experiments of this study, a specifically designed RNA hairpin was utilized. The sequence of the RNA hairpin, synthesized by Integrated DNA Technologies (IDT), was 5' - /5ATTO565K/ CACGUGAGUGUUUUGCGACGUG /3ATTO488N/ - 3', with a length of 22 bases. The RNA hairpin was modified at the 5' end with ATTO 565 and at the 3' end with ATTO 488 (NHS Ester) for smFRET applications (refer to Figure 27 in Appendix B section for absorption and emission spectra overlay of these dyes). The liposome encapsulation method was carefully optimized to maintain the integrity of the RNA hairpin structure, with the goal of preserving its fluorescent properties for subsequent smFRET analysis.

3.5.3. Experimental Measurements

In the smFRET experiments for developing the import assay, two distinct setups were investigated: one involving free-in-solution RNA hairpins and the other involving RNA hairpins encapsulated within liposomes.

For the assessment of buffer effects on the hairpins itself, only the free-in-solution RNA hairpins samples were used. The buffers examined included TKM; TKM with 1% BME; TKM with 2mM TROLOX; and a combination of TKM with both BME and TROLOX. Each buffer condition was designed to probe the influence of its

components on the fluorescence signal and stability of the RNA hairpins. The fluorescence measurements to quantify these effects were extended over a duration of 3600 seconds (carried out on three samples prepared under identical conditions, with each sample measured for a duration of 1200 seconds), providing ample time to monitor and record the fluctuations in the fluorescence signal attributable to buffer properties.

The analysis for the encapsulated RNA hairpins differed from that of the free-in-solution ones. Specifically, for each measurement involving liposome-encapsulated RNA, a quencher, 6M potassium iodide (KI), was added immediately prior to the measurement. This step was crucial to selectively quench the fluorescence signal from any RNA hairpins that were not encapsulated, thus ensuring that the measured FRET signal was entirely from species within the liposomes.

Furthermore, each measurement setup was conducted with addition of both active (RNase activity) and inactive variant (H255A mutant) of the CdiA toxin (both kindly provided by Dr William Allen from the University of Bristol, UK). Similarly to RNA hairpin, the toxin was added while preparing the lipid mixture. The inclusion of the active and inactive forms of CdiA toxin was instrumental in discerning their respective effects on the FRET signal, allowing for an insightful comparison of toxin activity in relation to the RNA hairpins' conformational changes.

To further explore the dynamics of the RNA hairpin's structural alterations, the experimental design also included a key step where samples were incubated for 60 minutes at room temperature (RT) following the addition of the toxin. This allowed for a comprehensive analysis of the incubation effect on the RNA hairpin, with measurements taken at the 60-minute mark to assess any temporal changes induced by the toxin's presence.

Prior to each of the measurements, meticulous calibration of the microscopy equipment was carried out, ensuring the light intensity and power of the laser beams were optimized for optimal detection. The blue 488 nm laser beam was adjusted to a power of 78 μW, equating to 76% of the laser's full intensity. Simultaneously, the green 566 nm laser beam was set to a power of 31 μW, or 73% of its maximum intensity.

The system was then allowed a stabilization period of 10-15 minutes to ensure stable laser output and thermal equilibrium, which is essential for the reproducibility and reliability of the smFRET measurements.

3.5.4. Data Analysis

In the study of smFRET experiments involving RNA hairpins, both free-in-solution and encapsulated within liposomes, data analysis played a critical role in interpreting the outcomes and drawing meaningful conclusions. To this end, the FRETbursts package (Ingargiola *et al.*, 2016) on Jupyter Notebook was employed to analyse the fluorescence bursts and extract quantitative and qualitative information regarding FRET efficiencies. The critical stages of this analytical process are concisely shown in the workflow depicted in Figure 7 and described below. Additionally, a visual representation of the data acquired is shown in Figure 26 in Appendix B section.

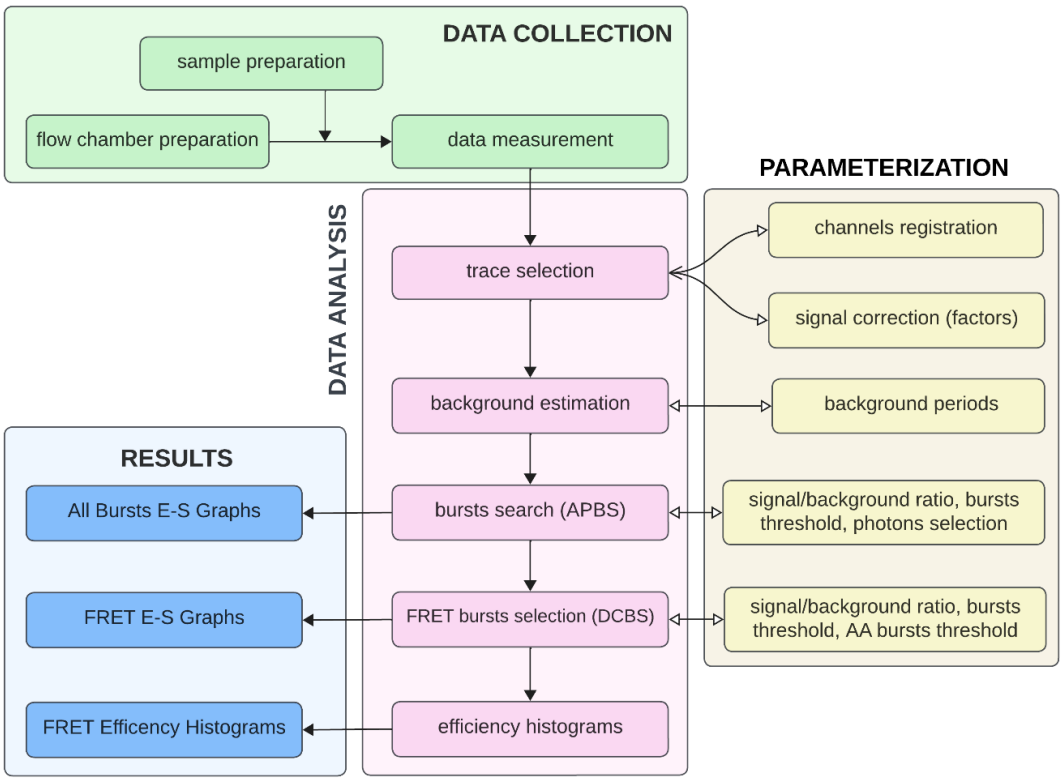

Figure 7: Workflow diagram representing the sequential methodology for smFRET data analysis and interpretation, including sample preparation, parameterization, data collection, and comprehensive analysis resulting in the production of E-S Graphs and efficiency histograms.

In the initial stage of data analysis, the primary focus was on the identification and enumeration of fluorescence bursts. A "burst" refers to a noticeable, brief rise in fluorescence intensity, which signals a probable FRET event between donor and

acceptor fluorophores on a solitary RNA hairpin molecule. The analysis utilized the Acceptance Photon Burst Search (APBS) algorithm, a method for detecting these bursts from the temporal fluorescence data. APBS operates by setting an intensity threshold of 30 — any fluorescence signal surpassing this threshold over a predetermined time frame is categorized as a burst, providing a primary selection of potential FRET events for further analysis.

Once the fluorescence bursts were detected using APBS, the next phase involved discerning the specific FRET events of interest from this broad population. This was carried out using Dual-Channel Burst Search (DCBS), which is a refined method that allows for the concurrent consideration of signals from both donor and acceptor channels. During this stage, a specific threshold for FRET AA burst selection of 15 was applied to ensure high-quality data for analysis. The resultant dataset, filtered through APBS and subsequently refined by DCBS, provided a high-fidelity pool of FRET events for subsequent quantitative and qualitative assessments.

After isolating the FRET populations, histograms of FRET efficiency were generated, and gaussian distribution were fitted to them. These histograms are visual representations of the distribution of FRET efficiency values across the selected burst populations, providing insights into the conformational states of the RNA hairpins.

The foundational parameters that guided both data collection and analysis, such as ALEX window, donor and acceptor excitation windows, and others, are summarized in Table 2 in the Appendix B section. These parameters establish the general framework for the experimental procedure and analytical approach. Additionally, specific correction factors tailored to the samples measured are detailed in Table 3 in the same section (see Figure 28 and Figure 29 for the derivation method of them).

4. Results

4.1. Predicted Structure of the CdiA Toxin

The very initial significant step toward achieving the aims of this thesis involved utilizing AlphaFold2's AI-driven capabilities to predict the intricate three-dimensional structure of the CdiA toxin. Out of the predctided structures, the best-ranked one is shown in Figure 8, where for better visualisation, the three main domains are color-coded.

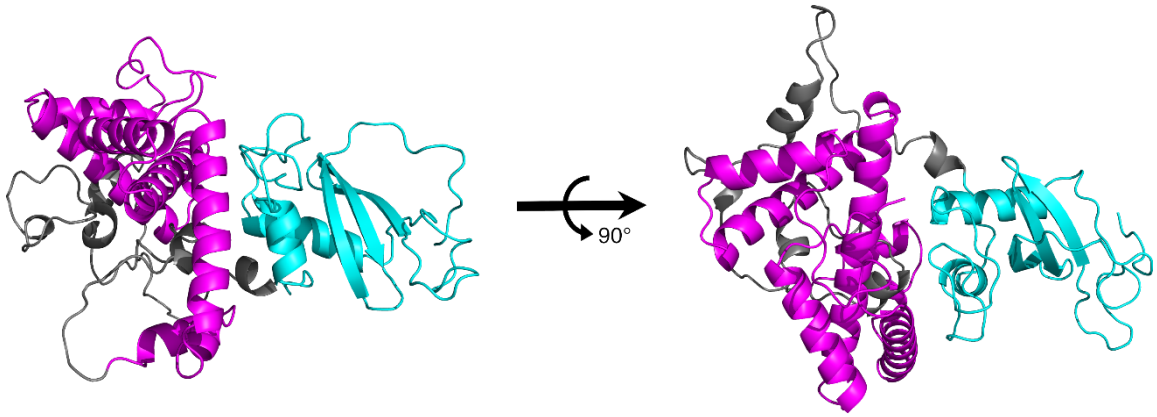

Figure 8: Structural prediction of the CdiA toxin with domain color-coding: the import domain (residues 1-159) in magenta, the flexible linker (residues 160-243) in grey, and the effector domain (residues 244-361) in cyan.

To substantiate the model's accuracy, Figure 9 presents two robust analytical visualizations: (a) illustrates the sequence coverage plot, showcasing the confidence levels of the AlphaFold2 model across the CdiA toxin's amino acid sequence and (b) a heat map displaying the expected positional error of amino acid residues within the toxin, demarcated by functional domains. These representations are crucial in confirming and validating the precision of the predictive model.

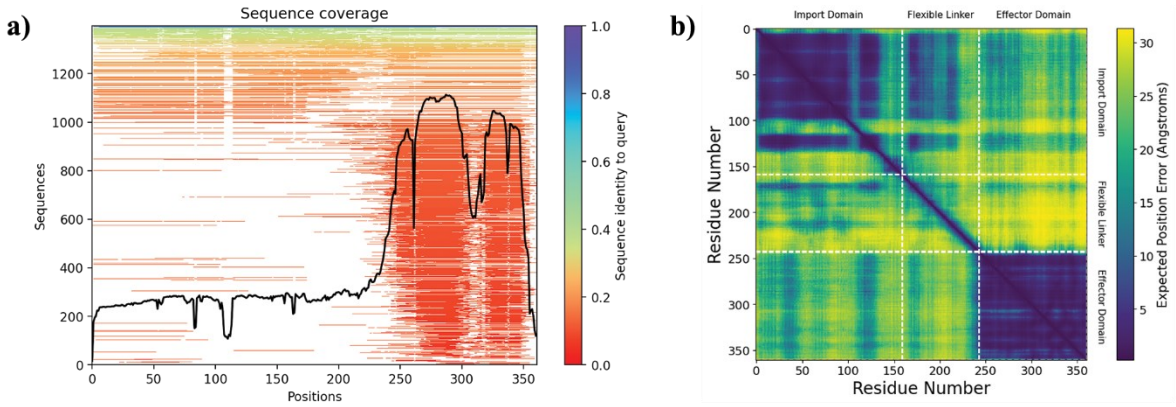

Figure 9: a) Coverage plot showing the confidence levels of the AlphaFold2 model across the CdiA toxin's amino acid sequence. b) Heat map of expected positional error in Angstroms for amino acid residues within the toxin, segmented by functional domains.

In specificity, it can be seen that the prediction above showcases a relatively reliable structure for some regions of the toxin. Especially, the whole effector domain and parts of the import one, show high confidence prediction, considering when cross-referenced with already established protein structure deposits in different databases.

Given the fact that the effector domain was of great interest in this study, only the structure of the region making up this domain (residues 244-361) was considered for further analysis. This part of the predicted structure consists of two α-helices and one three-stranded anti-parallel β-sheet (Figure 10).

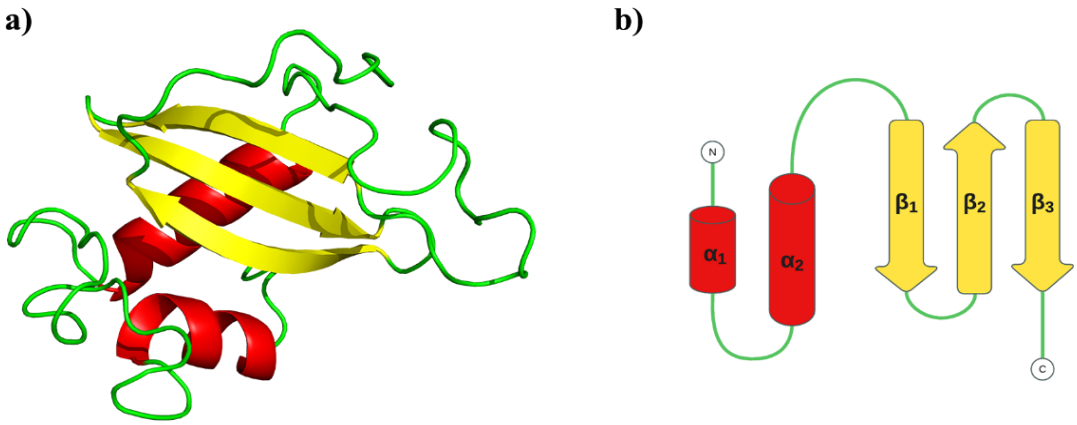

Figure 10: Structural overview of the CdiA toxin effector domain: (a) 3D conformation displaying alpha helices α_1 (residues 250-258) and α_2 (residues 284-296) in red, and beta strands β_1 (residues 321-327), β_2 (residues 330-337), and β_3 (residues 341-348) in yellow, (b) Schematic topology indicating the sequence locations of the alpha helices and beta strands within the toxin's effector domain.

4.2. Predicted Structure of the CdiA — SecYEG Complex

As described in Figure 3 (see Introduction section), the retrograde translocation process is thought to be initiated when the toxin's import domain binds and stimulates the lateral gate to open, subsequently facilitating the transit of the effector domain. Once within the cytoplasm, the effector domain is postulated to refold into its active conformation, completing the retrograde transport process.

Building on the foundational hypothesis of this mechanism, the predicted conformation between CdiA toxin and SecYEG furnishes evidence that substantiates the necessity for the effector domain of the CdiA toxin to unfold prior to translocation. The AlphaFold2 prediction offers a comprehensive visualization of the toxin in complex with the SecYEG translocon (Figure 11). Notably, the model illustrates a remarkably narrow opening

within the translocon, as observed in the enlarged segment. This passing channel, delineating a constricted pathway, seemingly just wide enough for single polypeptide chains, corroborates the hypothesis that no fully folded protein can traverse this channel. The structural prediction distinctly supports the claim of (Jones *et al.*, 2021) that the effector domain must undergo an unfolding process to navigate through the confined space provided by the translocon.

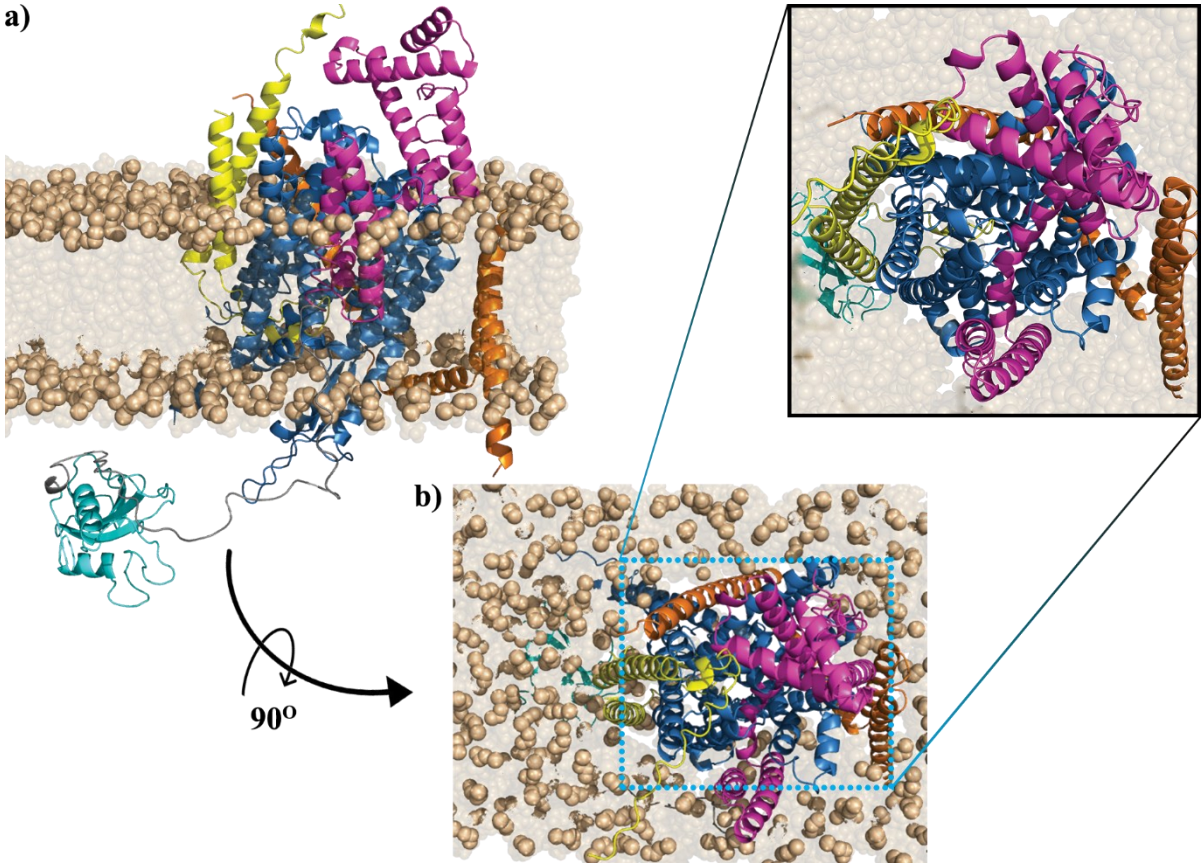

Figure 11: Structural AlphaFold2 prediction of the CdiA toxin interaction with the SecYEG translocon. Panel (a) illustrates the CdiA toxin domains in relation to the SecYEG complex: import domain in magenta, flexible linker in grey, and effector domain in cyan. The SecYEG components are color-coded with SecY in blue, SecE in orange, and SecG in yellow. The membrane bilayer is depicted as gold spheres. Panel (b) provides a view from the periplasmic side, demonstrating the orientation of the toxin and the SecYEG complex within the inner membrane environment. The channel section is zoomed to the upper-right.

4.3. Ionisation States of the Effector Domain

Transitioning from the macroscopic view of the CdiA toxin's navigation through the SecYEG translocon, the focus now narrows to the molecular intricacies that dictate this journey. The ionization states of the effector domain's ionizable residues are key, influencing both the necessary unfolding for translocation and the subsequent refolding within the host cell. The exploration sets the stage for correlating these microscopic properties with the macroscopic implications for the toxin's behaviour in the target cell. The pK_a predictions performed on the effector domain of the CdiA toxin using PROPKA 3.0 revealed significant

insights into the protonation behaviour of ionizable residues. These predicted pK_a values are summarized in Table 1.

Table 1: List showing the predicted pK_a values of amino acid residues within the effector domain of the CdiA toxin and providing insight into their protonation states across the studied pH values. The residues that are differently protonated between pH 6.5 and 7.5 are highlighted in orange.

Residue	Residue Index	pK_a	pH value	
			6.5	7.5
ASP	246	3.63	**deprotonated**	**deprotonated**
	293	2.88		
	299	3.9		
	328	2.74		
	349	2.45		
GLU	251	3.78	**deprotonated**	**deprotonated**
	260	3.6		
	325	4.95	**protonated**	**deprotonated**
	337	3.32	**deprotonated**	**deprotonated**
HIS	255	6.66	**protonated**	**deprotonated**
	267	4.98		
	289	6.06		
TYR	244	10.12	**protonated**	**protonated**
	258	10.22		
	269	10.13		
	303	10.52		
	312	10.22		
	336	10.71		
LYS	275	10.16	**protonated**	**protonated**
	286	11.36		
	314	10.45		
	342	10.53		
	358	10.41		
	361	10.41		
ARG	253	12.6	**protonated**	**protonated**
	317	12.41		
	320	13.14		
	327	13.32		
	331	12.26		
	333	11.92		

In this study, the pH values of 6.5 and 7.5 were specifically chosen for the MD simulations to accurately represent the intended biological contexts. The pH value of 6.5 was selected since it generally represents the periplasmic space of *E. coli* (Wilks and Slonczewski, 2007). Simulations at this pH helped to understand toxin behaviour in the environment where the unfolding takes place. pH 7.5 is akin to the intracellular pH (Wilks and Slonczewski, 2007), providing insights into how the toxin behaves inside the cell where it exerts its effect.

For each pH value, the protonation states of key residues alter, potentially affecting the stability of the effector domain. At pH 6.5, aspartic acid and glutamic acid residues (except GLU-325) were deprotonated, while histidine, tyrosine, lysine, and arginine residues were protonated. Conversely, at pH 7.5, which represents the intracellular pH, all aspartic acid, glutamic acid, and histidine residues were deprotonated, while tyrosine, lysine, and arginine remained protonated due to their naturally high pKa values.

This protonation-deprotonation pattern was vital for understanding how pH affects the toxin's structural stability. The structures presented in Figure 12 offer a visual exploration of the domain, pinpointing residues that are particularly susceptible to protonation shifts across pH values of 6.5 and 7.5.

The observed protonation dynamics of the effector domain residues suggested potential stability implications due to columbic interactions (Lee, Fitch and García-Moreno E., 2002), especially at pH 6.5. Specifically, the deprotonation of E325 at pH 7.5 fostered potential interactions with proximate positively charged residues, such as R327 and R331. This proximity potentially introduced additional stabilizing forces absent at pH 6.5. On the contrary, H289, which was positively charged at pH 6.5 and neutral at pH 7.5, likely was prone to repulsive interactions with R327 and K286. This could likely have led to a destabilization effect on the connection between the α_2 helix and the adjacent beta sheet. Parallel destabilizing consequences might be attributed to H267 and H255; both carrying a positive charge at pH 6.5 and shift to neutral at 7.5, potentially to have resulted in reduced structural stability due to repulsive forces.

An additional structural change at pH 7.5 included the emergence of a new short β-strand, denoted as β_0, which precedes the β_1 strand, expanding the beta sheet. This structural expansion, absent at pH 6.5 where the region remains a random coil, could substantially have enhanced the overall stability of the effector domain.

The alterations in the toxin's secondary structure were also mirrored in the hydrogen bonding profile, with a total count of 70 hydrogen bonds at pH 7.5 compared to a lower count of 65 at pH 6.5. Collectively, these modifications in protonation states and the secondary structure underscored a potential shift towards greater structural stability in the effector domain at pH 7.5, while likely compromising stability at the lower pH of 6.5.

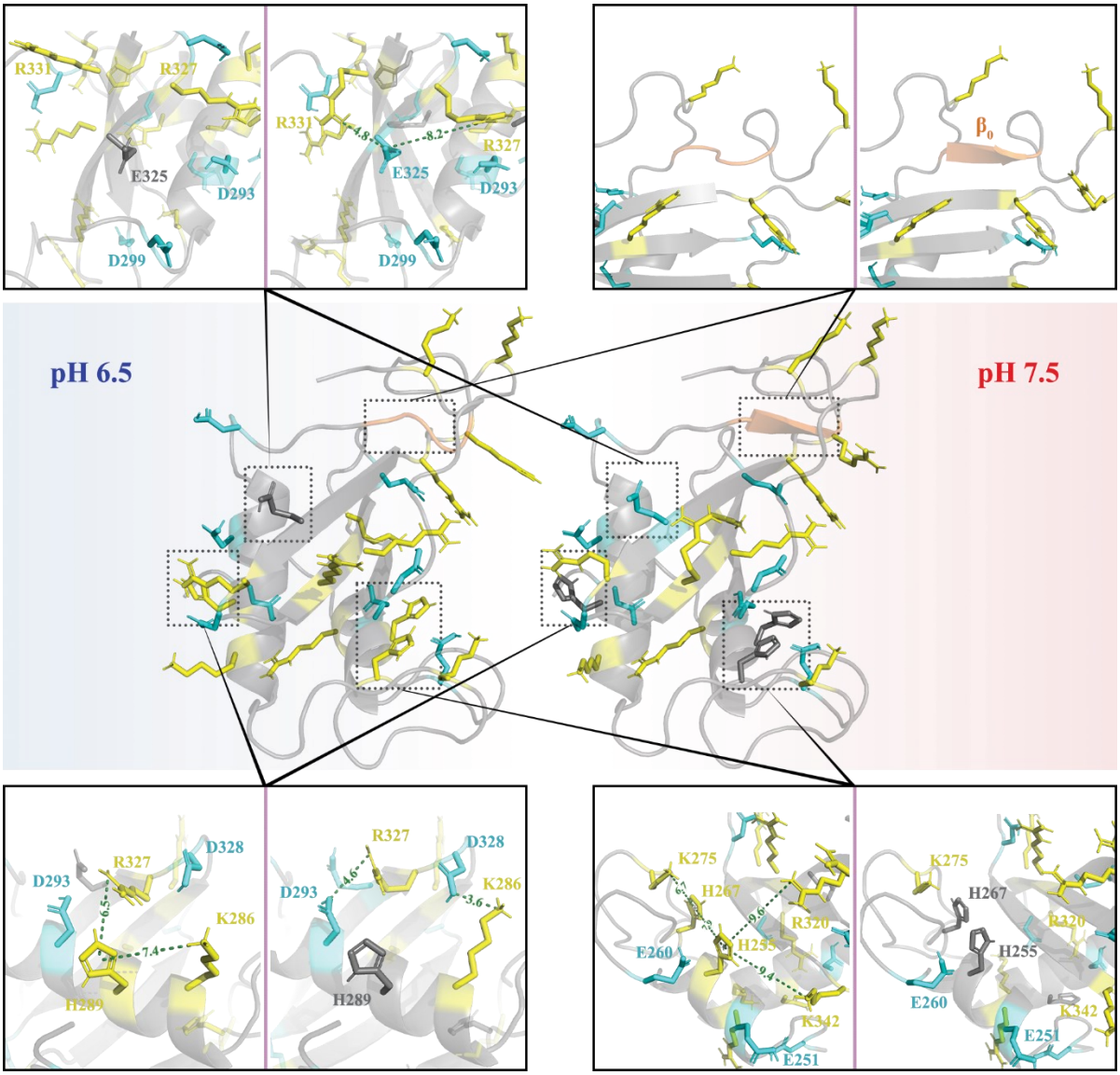

Figure 12: The structural conformation of the CdiA Toxin Effector Domain under two pH conditions: **pH 6.5** on the left and **pH 7.5** on the right. The effect of pH on the protonation states of specific residues is highlighted, showcasing the structural differences as a result. Positively charged residues are indicated in yellow, negatively charged ones in cyan, and neutral residues in grey, while with green dashed lines are shown the distances in angstroms (Å) between the charged atoms. Close-up views are provided to emphasize the four residues that differ in protonation between the two pH states, enhancing the visibility of these critical areas. Additionally, a notable secondary structure variation is displayed in the upper right corner, where a short β-strand, labelled as β_0, is present at pH 7.5 but absent at pH 6.5.

4.4. Effector Domain Unfolding

Following the examination of protonation dynamics and their influence on the effector domain's structural behavior, this subchapter turns to the practical implications of these biochemical properties during the process of toxin's effector domain mechaninal unfolding. Illustrated by the comprehensive set of molecular dynamics simulations, the unfolding pathway under the selected pH conditions is dissected in detail. The results, as elucidated through the rigor of computational analyses, reveal the nuanced interplay between the domain's structural integrity and the variable forces exerted upon it, underpinning the domain's remarkable capacity for conformational change in response to its immediate environment.

Located in the N-terminal of the effector domain of the CdiA toxin, the α-helices are the first in the sequence to be prone to enter the open SecYEG channel during the toxin redtrograde translocation process. This positioning can be correlated with why they were the first to unfold under pulling stress, aligning well with the hypothesis and supporting the idea that these elements are crucial in initiating the unfolding sequence. In specificiy, the pulling simulations of the CdiA toxin's effector domain, detailed in Appendix A, delineate a consistent unfolding sequence across the different experimental setups. Initiating with the disruption of the α_1 helix, the unfolding proceeded through α_2, and onto the β_3 strand, with the final disruption occurring at the β_1 and β_2 strands.

Within this unfolding paradigm, it becomes apparent that the α-helices not only serve as the initial structural elements to engage with the translocation channel but also play a protective role for the effector domain's hydrophobic core. Once the helices are unraveled during the initial stages of the pulling simulations, the inherent stability of the beta-sheet is compromised, something correlated with other studies (Parui and Jana, 2019). The α-helices act as a shield, and their disruption exposes the hydrophobic beta-strands, making them more susceptible to the unfolding process. This increased exposure likely accelerates the unfolding process, leading to the eventual separation of the β-strands under increased mechanical force.

Out of the different pulling rates, the moderate pulling (0.001 nm/ps) provided a particularly insightful view into the stability profile of the effector domain (Figure 13 and Figure 21 in Appendix A section). Within this dataset, it was observed that the forces required to make possible the unfolding at pH 6.5 were consistently lower than those at pH 7.5,

especially during the separation of the β_3 strand from the beta-sheet. This suggests a pH-dependent destabilization of the β_3 strand, implying that lower pH values may promote easier disruption of this structural element.

Furthermore, the introduction of the additional beta-strand β_0 at pH 7.5 relatively unfavoured the unfolding process, necessitating higher forces for the disruption of the beta-sheet, particularly evident at pulling distances of 26-30 nm. This higher force requirement indicates a more fortified structure at the higher pH, enhancing the stability of the beta-sheet complex. Moreover, despite similar forces being required to unfold the alpha helices across the physiological pH values, there was a notable increase in the force necessary to straighten the random coils at pH 7.5, around the 13-14 nm pulling distance, suggesting a more robust structure under this condition.

The effect of pulling rates on the unfolding process also offered valuable insights into the behavior of the effector domain under mechanical stress. While the general trend of unfolding did not differ significantly across different rates, it is evident from the data in Appendix A (Figure 23 & Figure 24) that at slower pulling rates, the structures unfolded more readily and required less force for disruption. This phenomenon can likely be attributed to the fact that during slower pulling, the structure has more time to adapt adequately to the applied forces (Sheridan, Gräter and Daday, 2019). This suggests that the unfolding process in its natural biological context is not rapid but occurs at a relatively slow pace, allowing the protein to adjust structurally in a more controlled manner.

Additionally, the PMF profiles (Figure 14 and Figure 25 in Appendix A section) provide a comprehensive view into the energetic landscape of the effector domain's unfolding under the studied pH conditions. Up to a pulling distance of approximately 12 nm, which corresponds with the completion of the α_1 helix disruption, the PMF curves for pH 6.5 and 7.5 are relatively identical. However, beyond this point, a noticeable divergence occurs. The energy required at pH 7.5 becomes significantly and consistently higher than that at pH 6.5, suggesting a more stable structure at the higher pH.

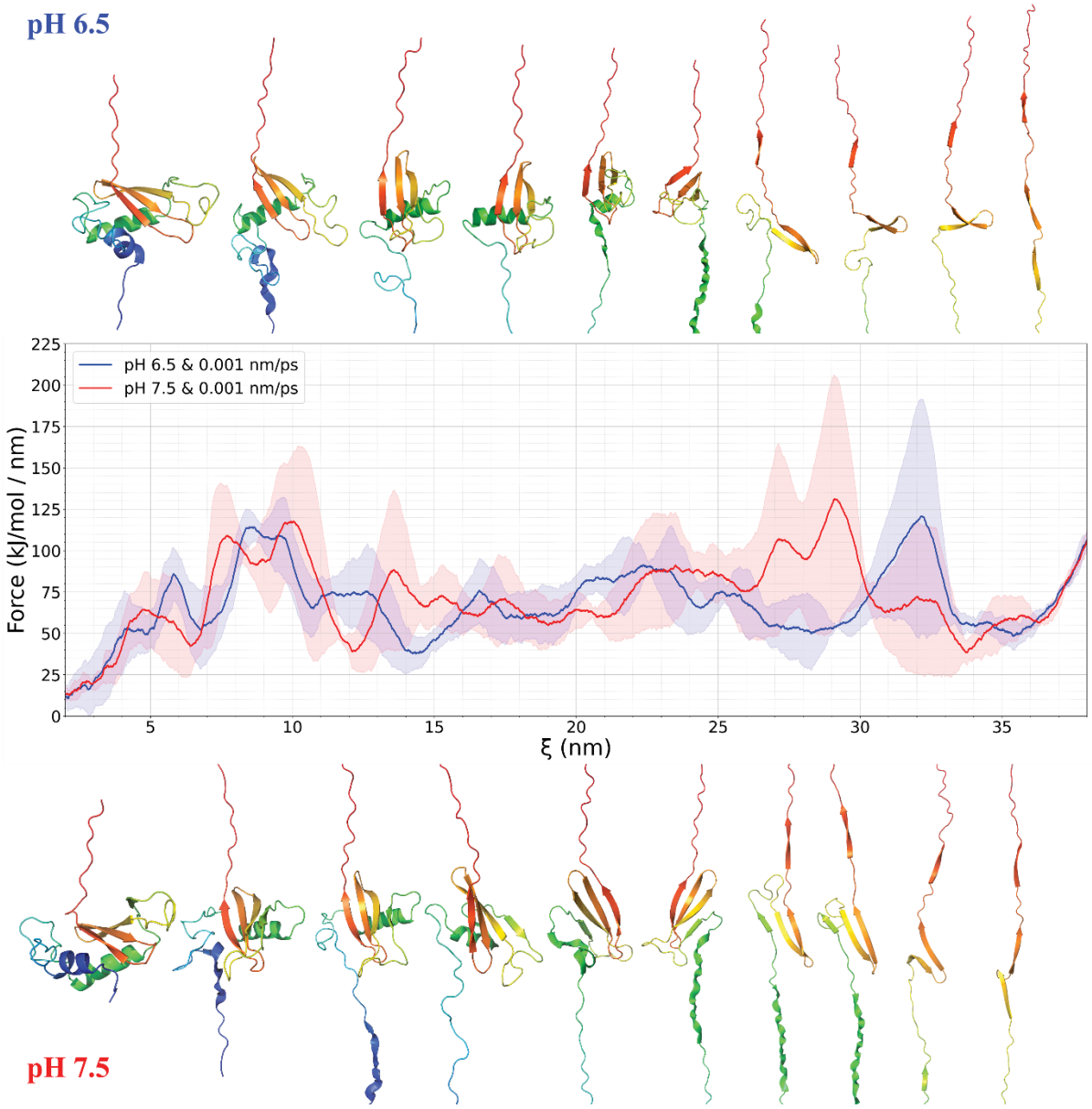

Figure 13: Comparative analysis of effector domain unfolding pathways at different pH values as demonstrated by force-extension curves and corresponding structural changes, highlighting the pH-dependent mechanical stability and unfolding mechanisms of the toxin. Unfolding pathways of the effector domain at **pH 6.5** (top unfolding series) and **pH 7.5** (bottom unfolding series) at a pulling rate of **0.001 nm/ps**. The force-extension graph in the centre, with a rolling average indicated in blue (pH 6.5) and red (pH 7.5) lines, measures the force as the toxin is extended. Shaded regions around each curve represent the standard deviation, indicating data variability.

This difference indicates that the overall energy landscape for unfolding at pH 6.5 is lower, supporting the hypothesis that unfolding is more favored under periplasmic conditions compared to cytoplasmic ones. The decreased stability and lower energy requirement at pH 6.5 reflect a structural fragility that likely influences the toxin's unfolding responses to environmental changes, aligning with the biological roles that these pH conditions simulate within the cellular context.

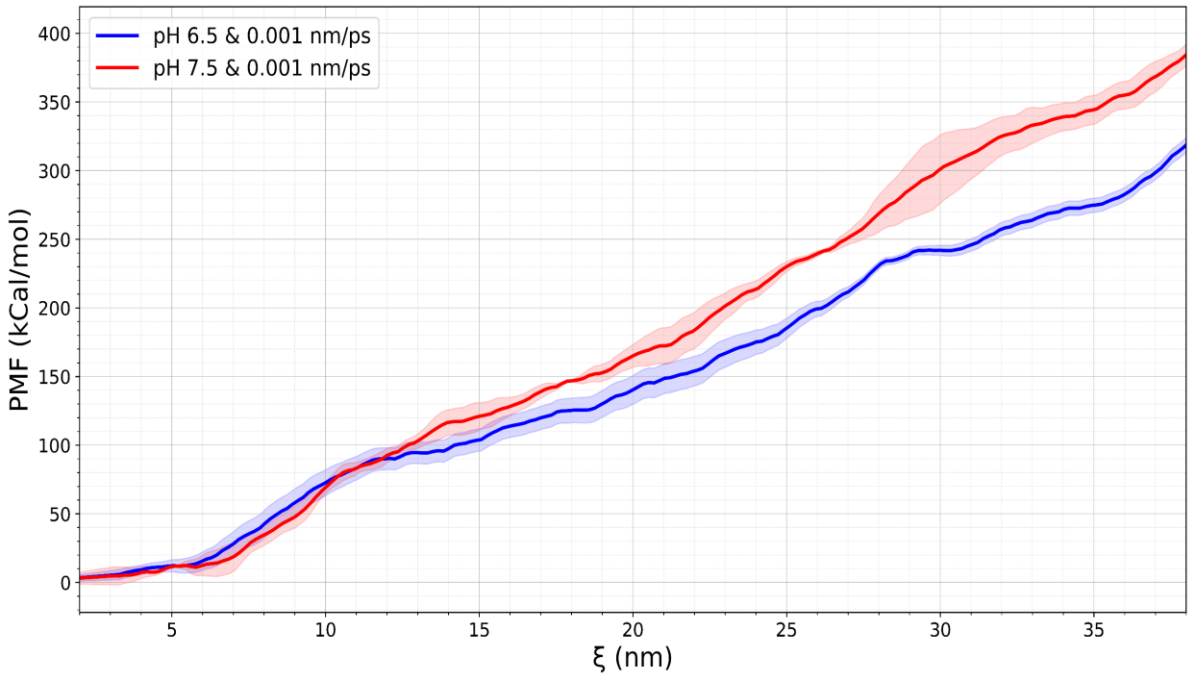

Figure 14: PMF profiles from umbrella sampling simulations. The graph depicts PMF as a function of the pulling distance, ξ (in nm), for the two different pH conditions. The blue line represents the rolling average PMF curve obtained at **pH 6.5** while the red line corresponds to **pH 7.5**, with a pulling rate of **0.001 nm/ps**. Shaded regions around each curve represent the standard deviation, indicating data variability.

4.5. Import Assay

In this section of the results, the data for the second main aim of the thesis, on the development of the import assay for investigating the import of CdiA toxin into liposomes using smFRET, are shown. This assay was crafted to facilitate future detailed investigations on the mechanism of retrograde translocation of these toxins through SecYEG channel.

4.5.1. Buffers Effects on RNA Hairpin

Firstly, the effects of various buffer compositions on the stability and FRET efficiency of the RNA hairpin were investigated. Employing smFRET as a sensitive tool for detecting molecular interactions, the impact of different combinations of TKM (Tris-KCl-$MgCl_2$), TROLOX, and BME on the conformation and dynamics of the RNA hairpin was studied.

The results (shown in Figure 15 and Figure 16) revealed that the buffer containing TKM, TROLOX, and BME was optimal for maintaining RNA hairpin stability, as evidenced by the largest population of high FRET efficiency. This suggests that the RNA hairpin structure was stable and properly folded, facilitating effective

FRET interactions. In contrast, the use of TKM alone, as well as combinations of just TKM with TROLOX or TKM with BME, resulted in significantly fewer bursts and lower counts of high FRET populations. These findings indicate suboptimal conditions for the stability and proper folding of the RNA hairpin under these buffer compositions.

Additionally, the lower burst counts and FRET populations observed with the simpler buffer mixtures could be attributed not only to the less optimal RNA hairpin conformations but also to potential interactions between the buffer components and the fluorescent dyes used in the smFRET experiments. Such interactions could alter the quantum efficiency and fluorescence properties of the dyes, thereby affecting the detectable FRET signal. For example, guanosine residues have been shown to quench fluorescence through dynamic interactions with dyes like rhodamine, which could significantly influence FRET efficiency (Heinlein *et al.*, 2003). Hydrophobic interactions within the buffer might also alter the stability and quantum yield of fluorescent dyes, further complicating the FRET measurement (Lu *et al.*, 2016).

Furthermore, the Gaussian distribution analysis of FRET efficiencies showed that the mixture of TKM, TROLOX, and BME produced a narrow and peaked distribution, characteristic of a well-defined and stable high FRET state. This contrasted with the broader distribution observed with TKM alone, and the combinations of TKM with TROLOX or BME, implying a more dynamic population with potentially unstable RNA hairpin conformations.

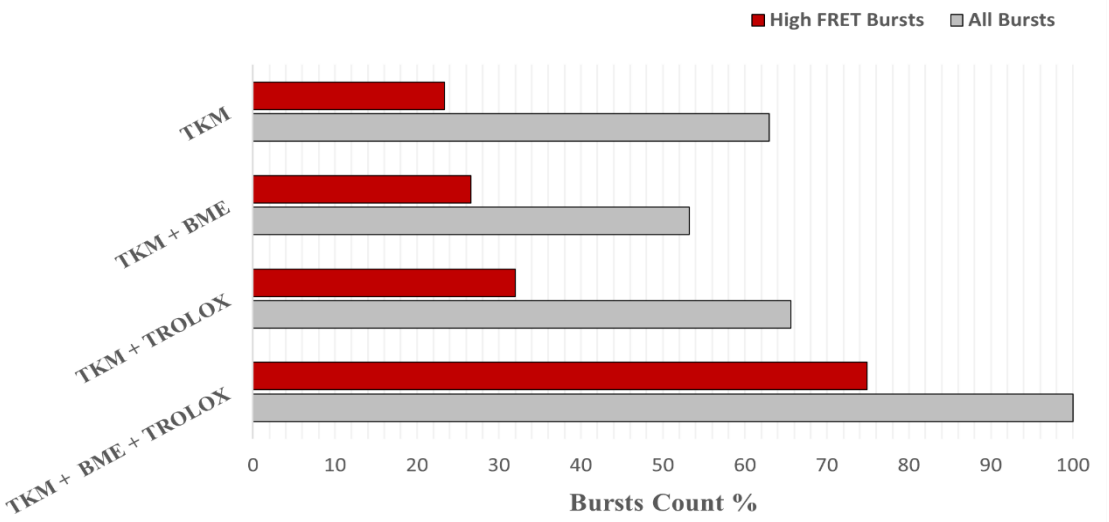

Figure 15: Bar graph displaying the percentage of FRET bursts counts of the RNA hairpins across different buffer compositions used in the study. The grey bars represent the total bursts observed, while the red bars indicate high FRET bursts specifically. The graph illustrates a significant increase in both total and high FRET bursts with the full mixture, highlighting the synergistic effect of the buffer components on RNA hairpin stability and FRET efficiency.

Collectively, these findings underscore the importance of the combined buffer components in promoting a consistent and high FRET population, essential for reliable experimental outcomes.

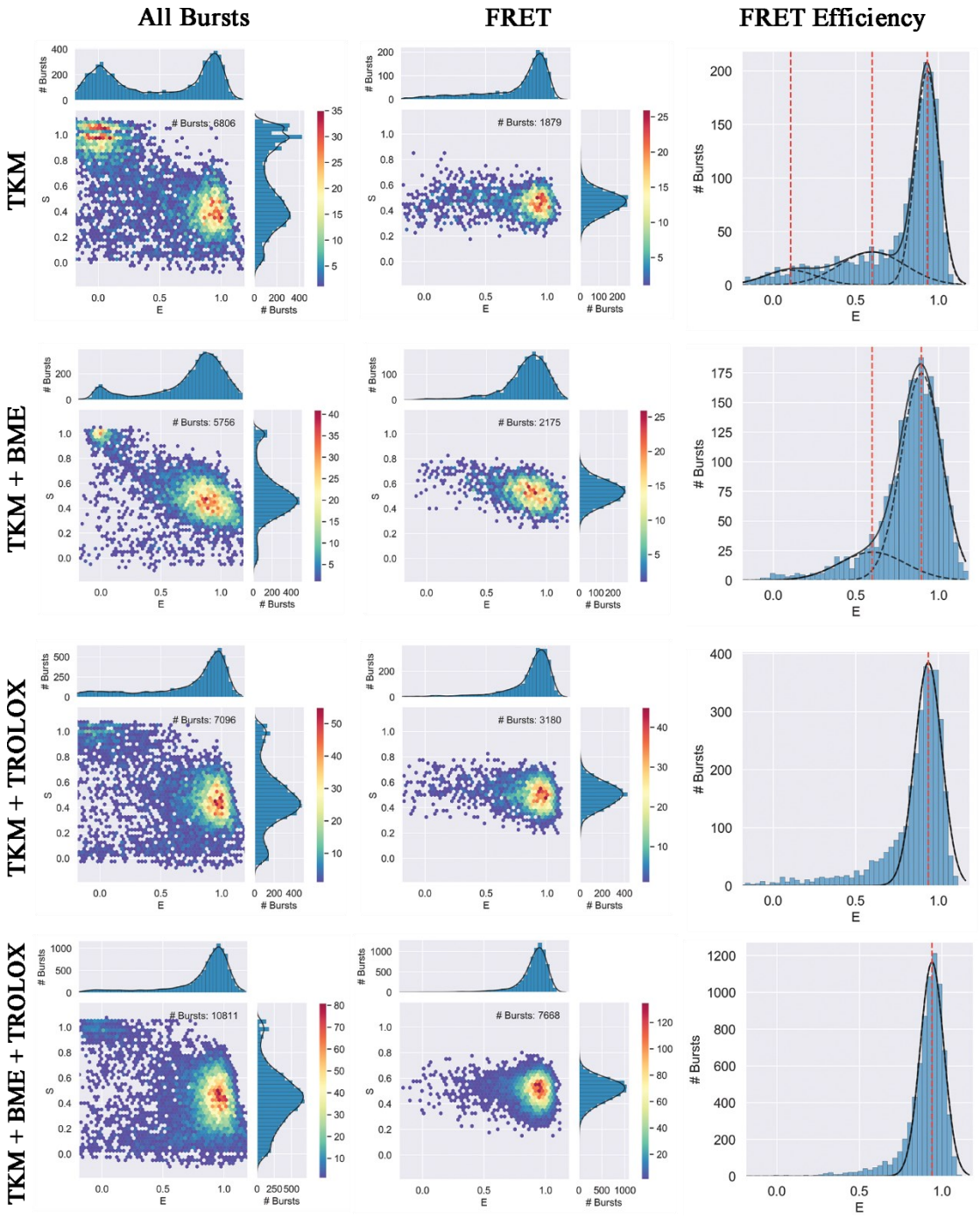

Figure 16: E-S (Efficiency vs. Stoichiometry) plots showcasing the FRET dynamics of the RNA hairpin across various buffer conditions. Each row represents a unique buffer mixture, with the first column showing the overall E-S plot for all observed bursts, the middle column focusing on FRET bursts, and the rightmost column depicting the FRET efficiency histograms. The density of points in E-S plots indicates the frequency of bursts with particular E and S values, while histograms illustrate the distribution of FRET efficiencies, with Gaussian fits overlaid to highlight the central tendencies under each condition.

4.5.2. Toxin Activity on RNA Hairpin

In this subsection, the direct effects of the active CdiA toxin possessing RNase activity, and its inactive variant, on the structural integrity of the free-in-solution RNA hairpins were assessed. Utilizing the sensitive technique of smFRET, the conformational dynamics of RNA hairpins were monitored upon exposure to these toxins. This experimental design aimed to elucidate the biochemical interactions at play, distinguishing between the catalytic action of the active toxin and the non-catalytic of the inactive form. The subsequent results (Figure 17 and Figure 18a) delve into the specific alterations observed in the FRET efficiency of the RNA hairpins, providing insights into toxin-induced changes in hairpin structure.

In the investigation of toxin effects on RNA hairpin stability, three different sample setups were analysed: RNA hairpins free-in-solution, hairpins with active toxin, and hairpins with inactive toxin. The introduction of the active toxin to the RNA hairpins free-in-solution resulted in a marked (almost complete) decrease in the high FRET population, indicating that the active toxin, with its RNase activity, effectively cleaved the majority of the RNA hairpins. Nevertheless, a residual moderate FRET population persisted, albeit in low numbers, suggesting that some hairpins were bound by the active toxin, altering their conformation but without complete cleaving them.

Conversely, the addition of the inactive toxin exhibited minimal impact on the high FRET population, indicating that the RNA hairpins largely remained stable and intact. However, a small moderate FRET population was observed, similar in profile to that seen with the active toxin, which may be attributed to the inactive toxin binding to the hairpins and causing minor conformational changes without executing cleavage. These observations suggest that while the active toxin was adept at cleaving the RNA hairpins, it also exhibited a propensity to bind and induce structural alterations in some cases. Similarly, the inactive toxin partially displayed a tendency to bind without subsequent cleavage, rightly reflecting its lack of RNase activity.

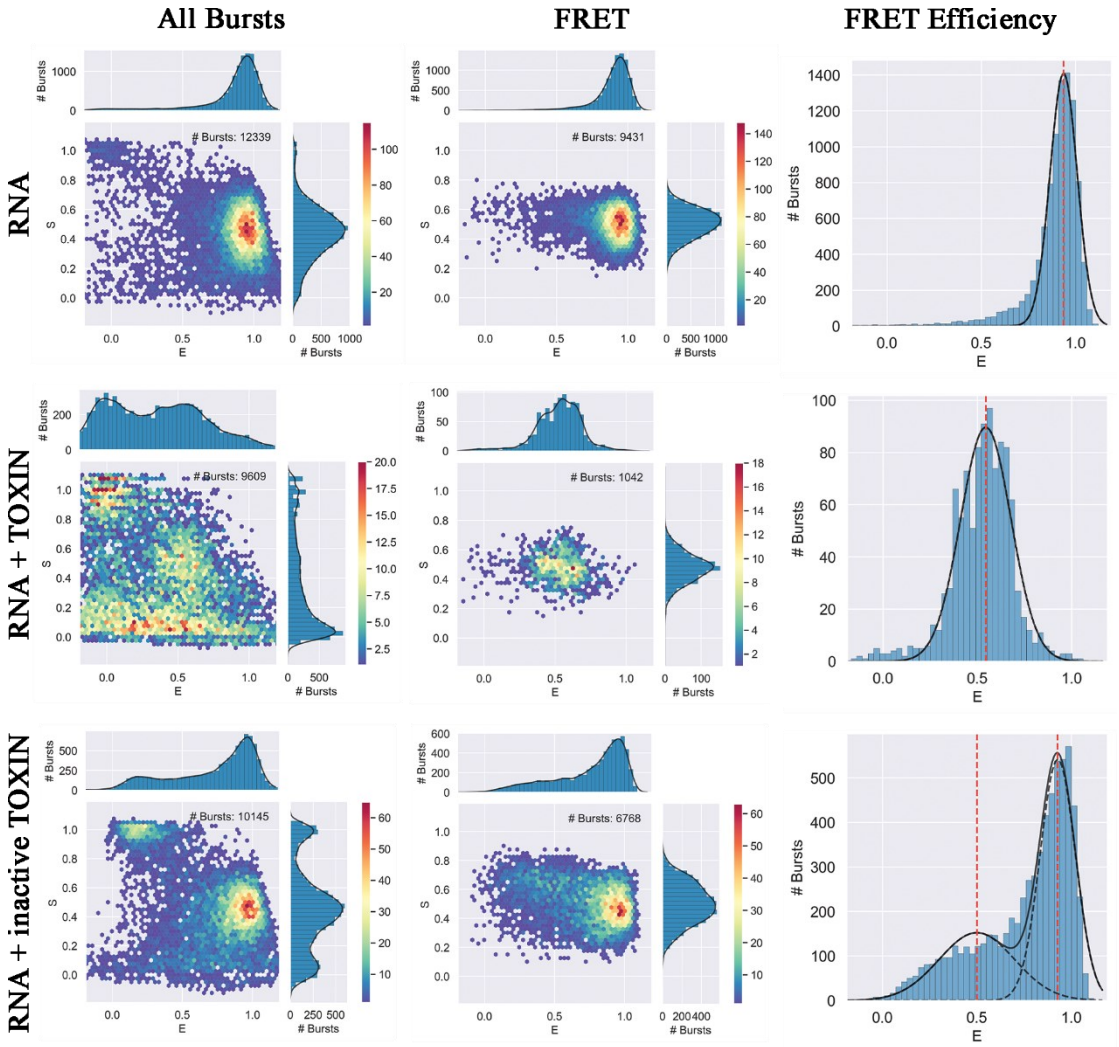

Figure 17: E-S plots and FRET efficiency histograms for the RNA hairpin to analyse the cleavage effect of the CdiA toxin. The top row serves as a control, showing the behaviour of the RNA hairpin without the toxin. The middle row illustrates the impact of the active CdiA toxin, and the bottom row depicts the effects of an inactive CdiA mutant. In each row, the left column represents the E-S plot for all bursts, the middle column isolates the FRET bursts, and the right column shows FRET efficiency histograms with Gaussian distributions fitted to the data.

While investigating the incubation of samples at RT for 60 minutes (Figure 30 in Appendix B section), a complete cleavage of the hairpin by active toxin was seen. However, while not cleaving it, the inactive mutant showed a subtler interaction with the RNA hairpin, shiftting the majority of the high FRET population to slightly lower FRET efficiency.

Overall, these results confirm the functionality of the active toxin in successfully cleaving the RNA hairpin and highlight the potential for slight conformational change upon binding (non-cleavage effect) by both active and inactive forms of the toxin.

4.5.3. Import Assay of the CdiA Toxin

As initially intended, a specialized import assay was developed to investigate the translocation of CdiA toxin into liposomes and its subsequent effects on RNA hairpin structure. By encapsulating RNA hairpins within the 100 nm liposomes, the experiments aimed to act as a control to distinguish the effects of active toxin on RNA structure inside these vesicles. Subsequently, Dr. Tomas Fessl conducted additional experiments at a later time using SecYEG enriched proteoliposomes (data not shown here). The comparison of these two setups was designed to verify the successful import and translocation of the toxin via the SecYEG complex and to observe the cleavage efficiency of the active toxin on the RNA hairpins, confirming the efficiency of the developed assay.

Although unpublished, the data from the experiments involving SecYEG proteoliposomes showed high similarities to the results from the simple liposomes setup shown here (Figure 18), validating the assay's efficacy. In the examination of the CdiA toxin's import, the introduction of the active toxin to the RNA hairpins encapsulated within 100nm liposomes was observed to significantly reduce the high FRET population. Specifically, upon the addition of the active toxin to the liposome-encapsulated RNA hairpins, the FRET efficiency diminished almost completely, transitioning mostly to low/moderate FRET events, albeit with a low count relative to the overall burst detected. This abrupt decrease indicated a substantial cleavage of the RNA hairpins.

Moreover, the similarity in FRET dynamics between the control liposomes (data shown here) and the proteoliposomes (data not shown here) under the simulated import conditions further validated the efficacy of the import assay. In both cases, the introduction of the active toxin resulted in an almost complete eradication of the high FRET population, mirroring the effect observed in free-in-solution RNA hairpins. This parallelism confirms the robustness of the assay in replicating the natural import process and detecting the consequential cleavage of the hairpin due to toxin translocation.

These findings demonstrate the import assay's capacity to accurately monitor the translocation and subsequent enzymatic action of the toxin. The drop in high FRET

populations aligns with expectations of the toxin's RNase activity, supporting the assay's potential as a reliable tool for studying the import mechanisms in a controlled experimental framework.

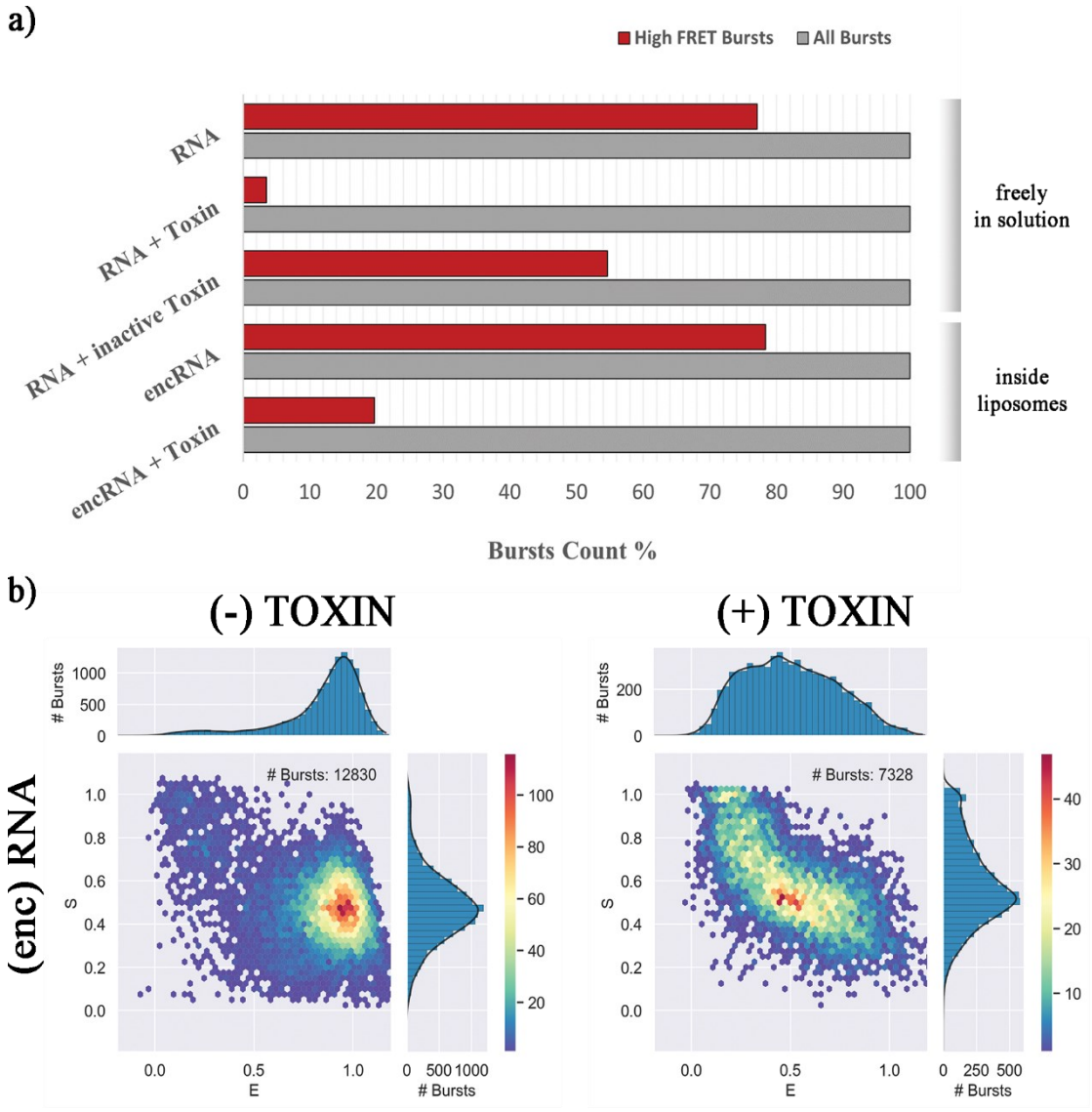

Figure 18: a) Bar graph depicting the FRET burst counts of RNA hairpins under diverse experimental conditions. The grey bars indicate the total bursts detected, while the red bars represent the subset of high FRET bursts. A noticeable decline in high FRET bursts is observed upon active toxin introduction, reflecting the toxin's cleavage efficiency. Conversely, the inactive toxin's presence slightly affects the high FRET population, suggesting low alterations in hairpin conformation and without substantial cleavage. b) E-S plots of RNA hairpin to analyse the successful cleavage effect of the CdiA toxin inside the formed liposomes.

5. Discussion

In the course of this study, the unfolding dynamics of the CdiA toxin's effector domain were thoroughly examined. Utilizing Alphafold2, the structural prediction of the CdiA – SecYEG complex supported the premise that the effector domain must be unfolded prior to translocation through the SecYEG channel (Jones *et al.*, 2021), attributed to the translocon's narrow opening. The structure of CdiA toxin was similarly predicted and subsequently adapted to pH conditions of 6.5 (periplasmic space) and 7.5 (cytoplasmic space) through targeted protonation and deprotonation of susceptible residues, as dictated by their predicted pKa values. Comparative analyses, at first, revealed a structural destabilization at pH 6.5 compared to pH 7.5, which was then seen through subsequent simulation studies.

Analysis of the simulation data revealed a sequential unfolding of secondary structure elements in the order of alpha helices α_1, α_2, followed by beta strands β_3 and β_1 & β_2. This order aligns well with the biological context where alpha helices, continuation of flexible linker and positioned proximally to the translocon, should initiate entry, succeeded by the unfolding of the remaining toxin structure. Comparative studies under different pH conditions elucidated a notable pH-induced destabilization during beta sheet unfolding, requiring diminished force for structural disruption. Furthermore, out of the different pulling simulations, the moderate pulling rate of 0.001 nm/ps was identified as optimal, optimizing the required force for extension of the toxin's effector domain, and facilitating a detailed examination of unfolding processes under the studied pH conditions.

The unfolding energy landscape, delineated by PMF calculations at the moderate pulling rate, demonstrated lower energy requirements at pH 6.5 than at 7.5, supporting the hypothesis that pH influences the unfolding process within the periplasmic space.

Nonetheless, it should be noted that the forces required for the unfolding of the domain might not entirely reflect real-world stability since the direction in which the force is applied also significantly influences the process, as suggested by (Kumar and Giri, 2007). Additionally, since the protonation states of the residues are only modified at the beginning, the interactions may not remain reliable towards the later stages of pulling, as the buried residues become more exposed and are likely to alter due to potential environmental adaptations of their pKa values.

To enhance the reliability and relevance of these findings, future work could focus on several experimental and computational strategies. Conducting molecular dynamics simulations using the constant pH methodology, as in other studies (Chen, Huang and Shen, 2016; Henderson *et al.*, 2022), throughout the pulling experiments could provide a more accurate representation of the toxin's behaviour *in vivo*, accounting for the effect of pH on residues initially not exposed to the solvent. Including cofactors known to facilitate the retrograde translocation process in these simulations, like PpiD-YfgM complex (Jones *et al.*, 2021), could offer insights into their potential role in lowering the energetic barriers of unfolding. Moreover, implementing Atomic Force Microscopy (AFM) for pulling experiments *in vitro* (Sheridan, Gräter and Daday, 2019), would allow for direct comparison with the *in silico* data, enhancing the validation of the computational results deducted.

Additionally, exploring the refolding processes of the unfolded effector domain within the cytoplasm through long-term coarse-grained molecular dynamics simulations (Maisuradze *et al.*, 2010) could aid in understanding its function as an RNase, crucial for its biological activity post-translocation. This investigation could contribute to exploring ways to control or manipulate this RNase activity.

Moreover, this thesis developed an import assay for investigating the retrograde transport of CdiA toxins using smFRET. The optimal assay conditions were established with a buffer composed of TKM, TROLOX, and BME, which produced the most robust signal with the studied RNA hairpin. Control experiments employing active and inactive mutants of the CdiA toxin validated the assay's efficacy in accurately assessing the translocation process.

Following the development of the import smFRET assay, additional experiments such as mutations in the SecYEG complex or the toxin itself or attempting to study *in vitro* the observed results from MD simulations, could shed light on the specific interactions and structural requirements essential for efficient retrograde transport. Furthermore, conducting experiments that mimic the pH conditions simulated in this study could provide crucial validation for the hypothesis that lower pH conditions favour the translocation process by facilitating the unfolding of the effector domain. This could significantly enhance our understanding of the physicochemical factors influencing toxin translocation and unfolding dynamics.

6. Conclusion

This study has extensively analysed the complex mechanisms underlying the unfolding of the effector domain in the verge of shedding light on retrograde translocation of CdiA toxins through the SecYEG translocon. By utilizing a combination of structural predictions with Alphafold2, GROMACS MD simulations, and single-molecule FRET experiments, the research has advanced the understanding of how the toxin's structural dynamics interplay with environmental conditions to influence its unfolding as a prerequisite for retrograde translocation.

The MD simulations revealed that the unfolding of the effector domain generally occurs sequentially, starting with the alpha helices followed by the beta strands, proving to be critical for the domain's passage through the narrow SecYEG channel. Notably, the periplasmic pH was found to affect the stability of the beta sheet structure, illuminating the role of environmental pH in toxin stability and unfolding dynamics.

The unfolding energy landscape, delineated by PMF calculations at the moderate pulling rate, demonstrated lower energy requirements at pH 6.5 than at 7.5, also supporting the hypothesis that pH influences the unfolding process within the periplasmic space.

Furthermore, this research pioneered the development of an import assay utilizing smFRET to investigate the retrograde transport of CdiA toxins. This innovative approach provides a vital experimental tool for directly observing the translocation mechanics of bacterial toxins in real-time.

In conclusion, this study not only advances the understanding of the structural dynamics governing the CdiA toxin's effector domain but also establishes a robust experimental framework for probing the mechanisms of toxin retrograde translocation through the SecYEG complex.

References

Abraham, M. *et al.* (2024) 'GROMACS 2024.1 Manual'. Zenodo. Available at: https://doi.org/10.5281/zenodo.10721192.

Abraham, M.J. *et al.* (2015) 'GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers', *SoftwareX*, 1–2, pp. 19–25. Available at: https://doi.org/10.1016/j.softx.2015.06.001.

Aoki, S.K. *et al.* (2005) 'Contact-Dependent Inhibition of Growth in *Escherichia coli*', *Science*, 309(5738), pp. 1245–1248. Available at: https://doi.org/10.1126/science.1115109.

Aoki, S.K. *et al.* (2008) 'Contact-dependent growth inhibition requires the essential outer membrane protein BamA (YaeT) as the receptor and the inner membrane transport protein AcrB', *Molecular Microbiology*, 70(2), pp. 323–340. Available at: https://doi.org/10.1111/j.1365-2958.2008.06404.x.

Aoki, S.K. *et al.* (2010) 'A widespread family of polymorphic contact-dependent toxin delivery systems in bacteria', *Nature*, 468(7322), pp. 439–442. Available at: https://doi.org/10.1038/nature09490.

Arkowitz, R.A., Joly, J.C. and Wickner, W. (1993) 'Translocation can drive the unfolding of a preprotein domain.', *The EMBO Journal*, 12(1), pp. 243–253. Available at: https://doi.org/10.1002/j.1460-2075.1993.tb05650.x.

Bange, G., Wild, K. and Sinning, I. (2007) 'Protein Translocation: Checkpoint Role for SRP GTPase Activation', *Current Biology*, 17(22), pp. R980–R982. Available at: https://doi.org/10.1016/j.cub.2007.09.041.

Bartels, C. and Karplus, M. (1997) 'Multidimensional adaptive umbrella sampling: Applications to main chain and side chain peptide conformations', *Journal of Computational Chemistry*, 18(12), pp. 1450–1462. Available at: https://doi.org/10.1002/(SICI)1096-987X(199709)18:12<1450::AID-JCC3>3.0.CO;2-I.

Berg, B. van den *et al.* (2004) 'X-ray structure of a protein-conducting channel', *Nature*, 427(6969), pp. 36–44. Available at: https://doi.org/10.1038/nature02218.

den Blaauwen, T. and Driessen, A.J.M. (1996) 'Sec-dependent preprotein translocation in bacteria', *Archives of Microbiology*, 165(1), pp. 1–8. Available at: https://doi.org/10.1007/s002030050289.

Blobel, G. *et al.* (1979) 'Translocation of proteins across membranes: the signal hypothesis and beyond.', *Symposia of the Society for Experimental Biology*, 33, pp. 9–36.

Bonardi, F. *et al.* (2011) 'Probing the SecYEG translocation pore size with preproteins conjugated with sizable rigid spherical molecules', *Proceedings of the National Academy of Sciences*, 108(19), pp. 7775–7780. Available at: https://doi.org/10.1073/pnas.1101705108.

Braun, V. and Patzer, S.I. (2013) 'Intercellular communication by related bacterial protein toxins: colicins, contact-dependent inhibitors, and proteins exported by the type VI secretion system', *FEMS Microbiology Letters*, 345(1), pp. 13–21. Available at: https://doi.org/10.1111/1574-6968.12180.

Cannon, K.S. *et al.* (2005) 'Disulfide bridge formation between SecY and a translocating polypeptide localizes the translocation pore to the center of SecY', *Journal of Cell Biology*, 169(2), pp. 219–225. Available at: https://doi.org/10.1083/jcb.200412019.

Chen, W., Huang, Y. and Shen, J. (2016) ‘Conformational Activation of a Transmembrane Proton Channel from Constant pH Molecular Dynamics’, *The Journal of Physical Chemistry Letters*, 7(19), pp. 3961–3966. Available at: https://doi.org/10.1021/acs.jpclett.6b01853.

Collinson, I. (2019) ‘The Dynamic ATP-Driven Mechanism of Bacterial Protein Translocation and the Critical Role of Phospholipids’, *Frontiers in Microbiology*, 10. Available at: https://doi.org/10.3389/fmicb.2019.01217.

Collinson, I., Corey, R.A. and Allen, W.J. (2015) ‘Channel crossing: how are proteins shipped across the bacterial plasma membrane?’, *Philosophical Transactions of the Royal Society B: Biological Sciences*, 370(1679), p. 20150025. Available at: https://doi.org/10.1098/rstb.2015.0025.

Costerton, J.W., Stewart, P.S. and Greenberg, E.P. (1999) ‘Bacterial Biofilms: A Common Cause of Persistent Infections’, *Science*, 284(5418), pp. 1318–1322. Available at: https://doi.org/10.1126/science.284.5418.1318.

Craggs Lab, U. of S. (2024) *smfBox*. Available at: https://craggslab.github.io/smfBox/smFRET.html (Accessed: 20 April 2024).

Driessen, A.J.M. and Nouwen, N. (2008) ‘Protein Translocation Across the Bacterial Cytoplasmic Membrane’, *Annual Review of Biochemistry*, 77(Volume 77, 2008), pp. 643–667. Available at: https://doi.org/10.1146/annurev.biochem.77.061606.160747.

Fessl, T. *et al.* (2018) ‘Dynamic action of the Sec machinery during initiation, protein translocation and termination’, *eLife*. Edited by T. Ha, 7, p. e35112. Available at: https://doi.org/10.7554/eLife.35112.

Finlay, B.B. and McFadden, G. (2006) ‘Anti-Immunology: Evasion of the Host Immune System by Bacterial and Viral Pathogens’, *Cell*, 124(4), pp. 767–782. Available at: https://doi.org/10.1016/j.cell.2006.01.034.

Gari, R.R.S. *et al.* (2013) ‘Dynamic Structure of the Translocon SecYEG in Membrane: DIRECT SINGLE MOLECULE OBSERVATIONS *’, *Journal of Biological Chemistry*, 288(23), pp. 16848–16854. Available at: https://doi.org/10.1074/jbc.M113.471870.

Heinlein, T. *et al.* (2003) ‘Photoinduced Electron Transfer between Fluorescent Dyes and Guanosine Residues in DNA-Hairpins’, *The Journal of Physical Chemistry B*, 107(31), pp. 7957–7964. Available at: https://doi.org/10.1021/jp0348068.

Hellenkamp, B. *et al.* (2018) ‘Precision and accuracy of single-molecule FRET measurements—a multi-laboratory benchmark study’, *Nature Methods*, 15(9), pp. 669–676. Available at: https://doi.org/10.1038/s41592-018-0085-0.

Henderson, J.A. *et al.* (2022) ‘A Guide to the Continuous Constant pH Molecular Dynamics Methods in Amber and CHARMM [Article v1.0]’, *Living Journal of Computational Molecular Science*, 4(1), pp. 1563–1563. Available at: https://doi.org/10.33011/livecoms.4.1.1563.

Hills Jr., R.D. *et al.* (2022) ‘Gut Microbiome: Profound Implications for Diet and Disease’, *Kompass Nutrition & Dietetics*, pp. 1–16. Available at: https://doi.org/10.1159/000523712.

Hub, J.S., De Groot, B.L. and Van Der Spoel, D. (2010) ‘g_wham—A Free Weighted Histogram Analysis Implementation Including Robust Error and Autocorrelation Estimates’, *Journal of Chemical Theory and Computation*, 6(12), pp. 3713–3720. Available at: https://doi.org/10.1021/ct100494z.

Huber, D. *et al.* (2011) 'SecA Interacts with Ribosomes in Order to Facilitate Posttranslational Translocation in Bacteria', *Molecular Cell*, 41(3), pp. 343–353. Available at: https://doi.org/10.1016/j.molcel.2010.12.028.

Huynh, M. and Sengupta, B. (2022) 'Analysis of Enzyme Conformation Dynamics Using Single-Molecule Förster Resonance Energy Transfer (smFRET)', *Biophysica*, 2(2), pp. 123–134. Available at: https://doi.org/10.3390/biophysica2020014.

Ingargiola, A. *et al.* (2016) 'FRETBursts: An Open Source Toolkit for Analysis of Freely-Diffusing Single-Molecule FRET', *PLOS ONE*, 11(8), p. e0160716. Available at: https://doi.org/10.1371/journal.pone.0160716.

Jones, A.M. *et al.* (2021) 'Genetic Evidence for SecY Translocon-Mediated Import of Two Contact-Dependent Growth Inhibition (CDI) Toxins', *mBio*. Edited by K.T. Hughes, 12(1), pp. e03367-20. Available at: https://doi.org/10.1128/mBio.03367-20.

Jumper, J. *et al.* (2021) 'Highly accurate protein structure prediction with AlphaFold', *Nature*, 596(7873), pp. 583–589. Available at: https://doi.org/10.1038/s41586-021-03819-2.

de Keyzer, J., van der Does, C. and Driessen, A.J.M. (2003) 'The bacterial translocase: a dynamic protein channel complex', *Cellular and Molecular Life Sciences CMLS*, 60(10), pp. 2034–2052. Available at: https://doi.org/10.1007/s00018-003-3006-y.

Koch, S. *et al.* (2019) 'Two distinct anionic phospholipid-dependent events involved in SecA-mediated protein translocation', *Biochimica et Biophysica Acta (BBA) - Biomembranes*, 1861(11), p. 183035. Available at: https://doi.org/10.1016/j.bbamem.2019.183035.

Komarudin Amalina Ghaisani and Driessen Arnold J. M. (2019) 'SecA-Mediated Protein Translocation through the SecYEG Channel', *Microbiology Spectrum*, 7(4), p. 10.1128/microbiolspec.psib-0028–2019. Available at: https://doi.org/10.1128/microbiolspec.psib-0028-2019.

Krulwich, T.A., Sachs, G. and Padan, E. (2011) 'Molecular aspects of bacterial pH sensing and homeostasis', *Nature Reviews Microbiology*, 9(5), pp. 330–343. Available at: https://doi.org/10.1038/nrmicro2549.

Kumar, S. and Giri, D. (2007) 'Does Changing the Pulling Direction Give Better Insight into Biomolecules?', *Physical Review Letters*, 98(4), p. 048101. Available at: https://doi.org/10.1103/PhysRevLett.98.048101.

Lee, K.K., Fitch, C.A. and García-Moreno E., B. (2002) 'Distance dependence and salt sensitivity of pairwise, coulombic interactions in a protein', *Protein Science*, 11(5), pp. 1004–1016. Available at: https://doi.org/10.1110/ps.4700102.

Lerner, E. *et al.* (2018) 'Toward dynamic structural biology: Two decades of single-molecule Förster resonance energy transfer', *Science*, 359(6373), p. eaan1133. Available at: https://doi.org/10.1126/science.aan1133.

Lerner, E. *et al.* (2021) 'FRET-based dynamic structural biology: Challenges, perspectives and an appeal for open-science practices', *eLife*. Edited by O. Boudker, 10, p. e60416. Available at: https://doi.org/10.7554/eLife.60416.

Lin, J. *et al.* (2015) 'Mechanisms of antibiotic resistance', *Frontiers in Microbiology*, 6. Available at: https://doi.org/10.3389/fmicb.2015.00034.

Lu, T. *et al.* (2016) 'The Non-Specific Binding of Fluorescent-Labeled MiRNAs on Cell Surface by Hydrophobic Interaction', *PLOS ONE*. Edited by G. Camussi, 11(3), p. e0149751. Available at: https://doi.org/10.1371/journal.pone.0149751.

Maisuradze, G.G. *et al.* (2010) 'Investigation of Protein Folding by Coarse-Grained Molecular Dynamics with the UNRES Force Field', *The Journal of Physical Chemistry A*, 114(13), pp. 4471–4485. Available at: https://doi.org/10.1021/jp9117776.

McCann, J.J. *et al.* (2010) 'Optimizing Methods to Recover Absolute FRET Efficiency from Immobilized Single Molecules', *Biophysical Journal*, 99(3), pp. 961–970. Available at: https://doi.org/10.1016/j.bpj.2010.04.063.

Miller, M.B. and Bassler, B.L. (2001) 'Quorum Sensing in Bacteria', *Annual Review of Microbiology*, 55(1), pp. 165–199. Available at: https://doi.org/10.1146/annurev.micro.55.1.165.

Morehead, M.S. and Scarbrough, C. (2018) 'Emergence of Global Antibiotic Resistance', *Primary Care: Clinics in Office Practice*, 45(3), pp. 467–484. Available at: https://doi.org/10.1016/j.pop.2018.05.006.

Mori, H. and Ito, K. (2001) 'The Sec protein-translocation pathway', *Trends in Microbiology*, 9(10), pp. 494–500. Available at: https://doi.org/10.1016/S0966-842X(01)02174-6.

Mori, T. *et al.* (2010) 'Molecular Mechanisms Underlying the Early Stage of Protein Translocation through the Sec Translocon', *Biochemistry*, 49(5), pp. 945–950. Available at: https://doi.org/10.1021/bi901594w.

Nilsson, I. *et al.* (2015) 'The Code for Directing Proteins for Translocation across ER Membrane: SRP Cotranslationally Recognizes Specific Features of a Signal Sequence', *Journal of Molecular Biology*, 427(6, Part A), pp. 1191–1201. Available at: https://doi.org/10.1016/j.jmb.2014.06.014.

Olsson, M.H.M. *et al.* (2011) 'PROPKA3: Consistent Treatment of Internal and Surface Residues in Empirical pKa Predictions', *Journal of Chemical Theory and Computation*, 7(2), pp. 525–537. Available at: https://doi.org/10.1021/ct100578z.

Omasits, U. *et al.* (2014) 'Protter: interactive protein feature visualization and integration with experimental proteomic data', *Bioinformatics*, 30(6), pp. 884–886. Available at: https://doi.org/10.1093/bioinformatics/btt607.

Parui, S. and Jana, B. (2019) 'Relative Solvent Exposure of the Alpha-Helix and Beta-Sheet in Water Determines the Initial Stages of Urea and Guanidinium Chloride-Induced Denaturation of Alpha/Beta Proteins', *The Journal of Physical Chemistry B*, 123(42), pp. 8889–8900. Available at: https://doi.org/10.1021/acs.jpcb.9b06859.

Prestinaci, F., Pezzotti, P. and Pantosti, A. (2015) 'Antimicrobial resistance: a global multifaceted phenomenon', *Pathogens and Global Health*, 109(7), pp. 309–318. Available at: https://doi.org/10.1179/2047773215Y.0000000030.

Rapoport, T.A. (2007) 'Protein translocation across the eukaryotic endoplasmic reticulum and bacterial plasma membranes', *Nature*, 450(7170), pp. 663–669. Available at: https://doi.org/10.1038/nature06384.

Riley, M.A. (1998) 'MOLECULAR MECHANISMS OF BACTERIOCIN EVOLUTION', *Annual Review of Genetics*, 32(1), pp. 255–278. Available at: https://doi.org/10.1146/annurev.genet.32.1.255.

Riley, M.A. and Wertz, J.E. (2002) 'Bacteriocins: Evolution, Ecology, and Application', *Annual Review of Microbiology*, 56(1), pp. 117–137. Available at: https://doi.org/10.1146/annurev.micro.56.012302.161024.

Ruhe, Z.C. *et al.* (2013) 'Receptor Polymorphism Restricts Contact-Dependent Growth Inhibition to Members of the Same Species', *mBio*. Edited by E.P. Greenberg, 4(4), pp. e00480-13. Available at: https://doi.org/10.1128/mBio.00480-13.

Ruhe, Z.C. *et al.* (2014) 'The proton-motive force is required for translocation of CDI toxins across the inner membrane of target bacteria', *Molecular Microbiology*, 94(2), pp. 466–481. Available at: https://doi.org/10.1111/mmi.12779.

Ruhe, Z.C. *et al.* (2018) 'Programmed Secretion Arrest and Receptor-Triggered Toxin Export during Antibacterial Contact-Dependent Growth Inhibition', *Cell*, 175(4), pp. 921-933.e14. Available at: https://doi.org/10.1016/j.cell.2018.10.033.

Schrödinger, LLC (2015) 'The PyMOL Molecular Graphics System, Version 1.8'.

Schwarz, J. *et al.* (2019) 'Macromolecular crowding effects on energy transfer efficiency and donor-acceptor distance of hetero-FRET sensors using time-resolved fluorescence', *Methods and Applications in Fluorescence*, 7(2), p. 025002. Available at: https://doi.org/10.1088/2050-6120/ab0242.

Sheridan, S., Gräter, F. and Daday, C. (2019) 'How Fast Is Too Fast in Force-Probe Molecular Dynamics Simulations?', *The Journal of Physical Chemistry B*, 123(17), pp. 3658–3664. Available at: https://doi.org/10.1021/acs.jpcb.9b01251.

Søndergaard, C.R. *et al.* (2011) 'Improved Treatment of Ligands and Coupling Effects in Empirical Calculation and Rationalization of pKa Values', *Journal of Chemical Theory and Computation*, 7(7), pp. 2284–2295. Available at: https://doi.org/10.1021/ct200133y.

Tollinger, M. *et al.* (2003) 'Site-specific contributions to the pH dependence of protein stability', *Proceedings of the National Academy of Sciences*, 100(8), pp. 4545–4550. Available at: https://doi.org/10.1073/pnas.0736600100.

Tsukazaki, T. *et al.* (2008) 'Conformational transition of Sec machinery inferred from bacterial SecYE structures', *Nature*, 455(7215), pp. 988–991. Available at: https://doi.org/10.1038/nature07421.

Van Der Spoel, D. *et al.* (2005) 'GROMACS: Fast, flexible, and free', *Journal of Computational Chemistry*, 26(16), pp. 1701–1718. Available at: https://doi.org/10.1002/jcc.20291.

Varadi, M. *et al.* (2022) 'AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models', *Nucleic Acids Research*, 50(D1), pp. D439–D444. Available at: https://doi.org/10.1093/nar/gkab1061.

Verner, K. and Schatz, G. (1988) 'Protein Translocation Across Membranes', *Science*, 241(4871), pp. 1307–1313. Available at: https://doi.org/10.1126/science.2842866.

Wallace, B. and Atzberger, P.J. (2017) 'Förster resonance energy transfer: Role of diffusion of fluorophore orientation and separation in observed shifts of FRET efficiency', *PLOS ONE*. Edited by S. D'Auria, 12(5), p. e0177122. Available at: https://doi.org/10.1371/journal.pone.0177122.

Watve, M. *et al.* (2001) 'How many antibiotics are produced by the genus Streptomyces ?', *Archives of Microbiology*, 176(5), pp. 386–390. Available at: https://doi.org/10.1007/s002030100345.

Wilks, J.C. and Slonczewski, J.L. (2007) ‘pH of the Cytoplasm and Periplasm of Escherichia coli: Rapid Measurement by Green Fluorescent Protein Fluorimetry’, *Journal of Bacteriology*, 189(15), pp. 5601–5607. Available at: https://doi.org/10.1128/JB.00615-07.

Yang, A.-S. and Honig, B. (1993) ‘On the pH Dependence of Protein Stability’, *Journal of Molecular Biology*, 231(2), pp. 459–474. Available at: https://doi.org/10.1006/jmbi.1993.1294.

Zhang, D., Iyer, L.M. and Aravind, L. (2011) ‘A novel immunity system for bacterial nucleic acid degrading toxins and its recruitment in various eukaryotic and DNA viral systems’, *Nucleic Acids Research*, 39(11), pp. 4532–4552. Available at: https://doi.org/10.1093/nar/gkr036.

Appendix A

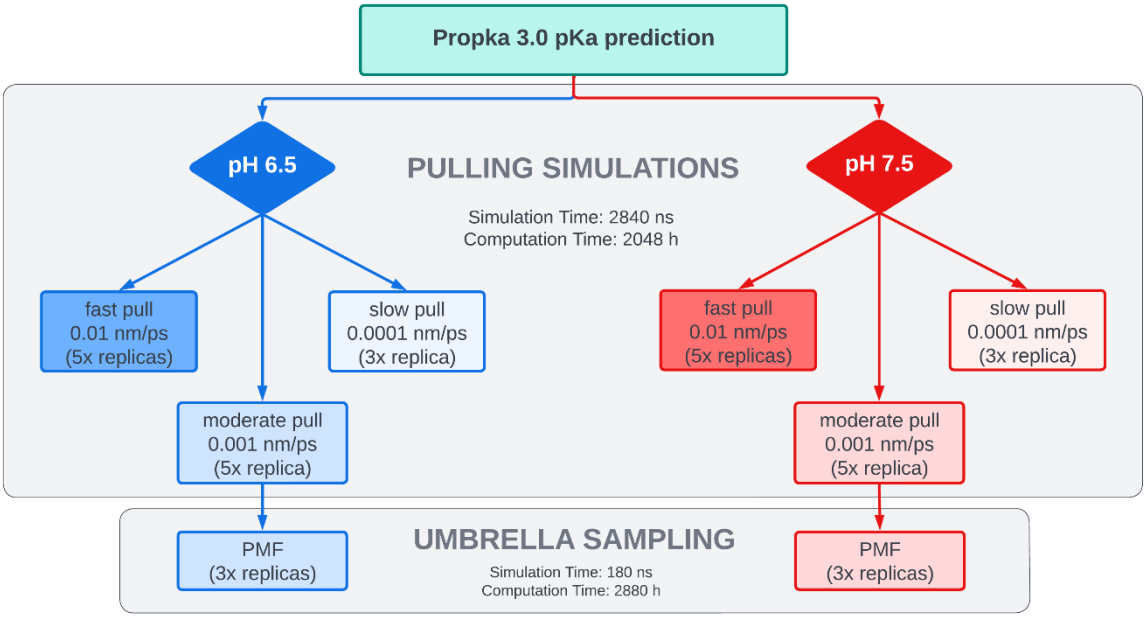

Figure 19: Experimental Setup for Molecular Dynamics Simulations. This flowchart delineates the procedural framework for Pulling and Umbrella Sampling Simulations conducted at two distinct conditions — **pH 6.5** and **pH 7.5**. The experiment entailed mechanical pulling simulations executed at varying pulling rates, specifically fast (0.01 nm/ps), moderate (0.001 nm/ps), and slow (0.0001 nm/ps). It also details the subsequent PMF calculations via umbrella sampling. The number of replications performed for each type of simulation is noted to ensure experimental reliability.

Figures below show the DSSP structural changes of the effector domain alongside the forces applied as it is pulled at different rates for the two studied pH values. It is worth noting that at the slow pulling (see Figure 20), the forces for unfolding the β_0 structure are distinguishable from the other elements of the beta sheet, something that cannot be seen in the other faster pulling rates. Additionally, it can be seen that slightly lower forces are required for the unfolding of the first alpha helix α_1 and straightening of the random coils preceding it (pulling distance of 3 – 8 nm) in the periplasmic pH conditions. Conversely, at pH of 6.5 the beta-sheet remained remarkably stable until subjected to significantly higher forces (pulling distance of 28 – 32 nm), causing simultaneous strands separation. The underlying mechanisms contributing to this anomalous stabilization at lower pH remain unclear.

On the other side, the fast pulling (see Figure 22) showed no significant difference between the two pHs, where the force profiles mostly overlap each other, indicating a lack of sensitivity under rapid pulling conditions. Additionally, the comparison of the pulling rates (Figure 23 & Figure 24), show lower forces for the slower pulling rates, but at the same time, nicely visualize the differences of when the main secondary structure components start unfolding in these different rates.

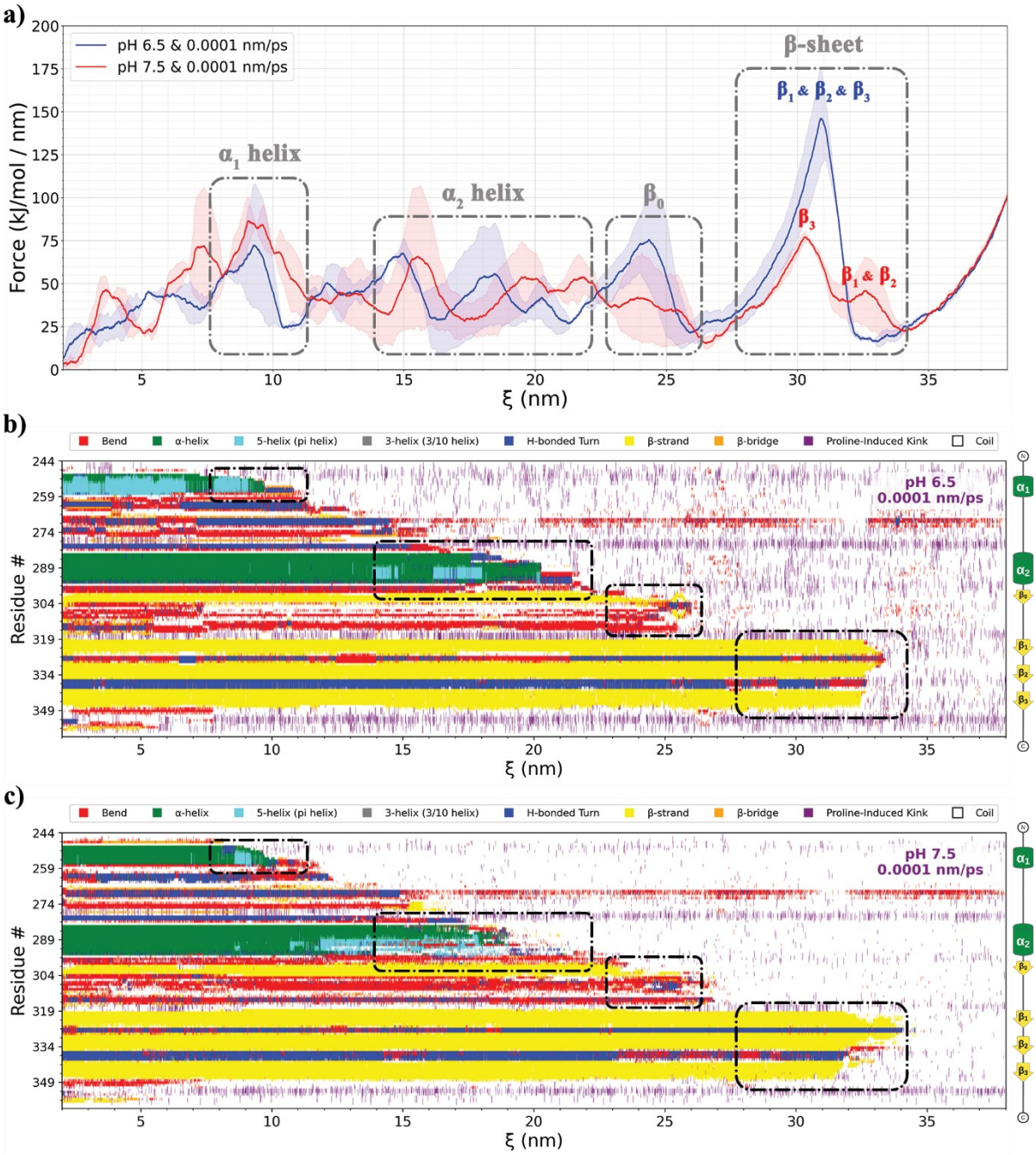

Figure 20: Analysis of Force and Structural Changes During Effector Domain Unfolding in Pulling Simulations. a) Force profiles as a function of the pulling distance, ξ (in nm), with the blue curve representing simulations at **pH 6.5** and the red curve at **pH 7.5**, both at a pulling rate of **0.0001 nm/ps**. The shaded regions around each curve indicate the standard deviation, representing the variability within the force measurements for each average profile. Regions corresponding to α_1 and α_2 helices, and β-sheet are highlighted. b) & c) Derived from DSSP analysis, these plots detail the evolution of secondary structures along the effector domain sequence (residues) as it is extended, for pH 6.5 and 7.5 respectively. Key secondary structures are indicated by colour coding: green for α-helices, yellow for β-strands, and other colours representing different structural motifs, illustrating the effector's domain conformational changes over the simulation.

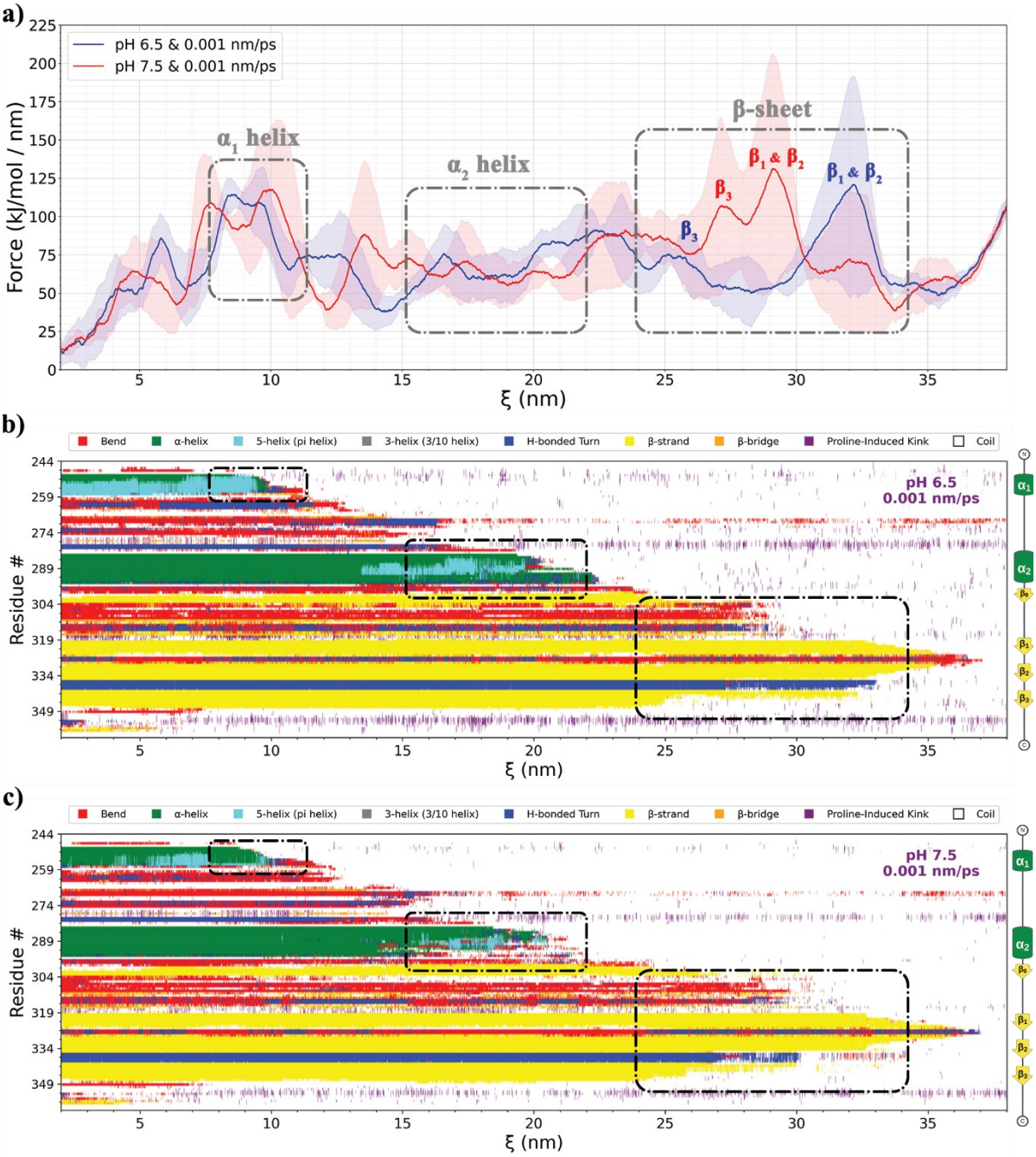

Figure 21: Analysis of Force and Structural Changes During Effector Domain Unfolding in Pulling Simulations. a) Force profiles as a function of the pulling distance, ξ (in nm), with the blue curve representing simulations at **pH 6.5** and the red curve at **pH 7.5**, both at a pulling rate of **0.001 nm/ps**. The shaded regions around each curve indicate the standard deviation, representing the variability within the force measurements for each average profile. Regions corresponding to α_1 and α_2 helices, and β-sheet are highlighted. b) & c) Derived from DSSP analysis, these plots detail the evolution of secondary structures along the effector domain sequence (residues) as it is extended, for pH 6.5 and 7.5 respectively. Key secondary structures are indicated by colour coding: green for α-helices, yellow for β-strands, and other colours representing different structural motifs, illustrating the effector's domain conformational changes over the simulation.

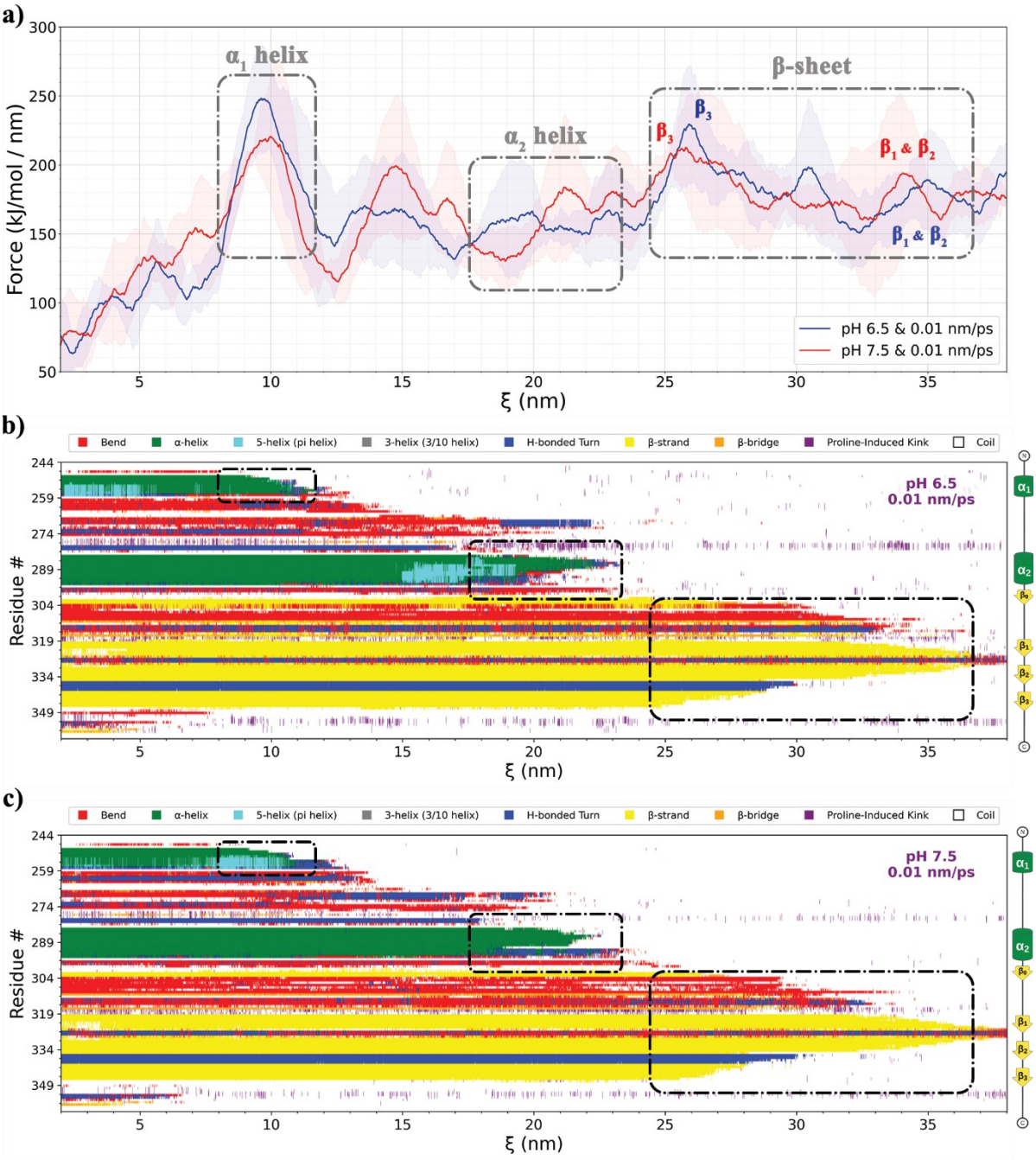

Figure 22: Analysis of Force and Structural Changes During Effector Domain Unfolding in Pulling Simulations. a) Force profiles as a function of the pulling distance, ξ (in nm), with the blue curve representing simulations at **pH 6.5** and the red curve at **pH 7.5**, both at a pulling rate of **0.01 nm/ps**. The shaded regions around each curve indicate the standard deviation, representing the variability within the force measurements for each average profile. Regions corresponding to α_1 and α_2 helices, and β-sheet are highlighted. b) & c) Derived from DSSP analysis, these plots detail the evolution of secondary structures along the effector domain sequence (residues) as it is extended, for pH 6.5 and 7.5 respectively. Key secondary structures are indicated by colour coding: green for α-helices, yellow for β-strands, and other colours representing different structural motifs, illustrating the toxin's conformational changes over the simulation.

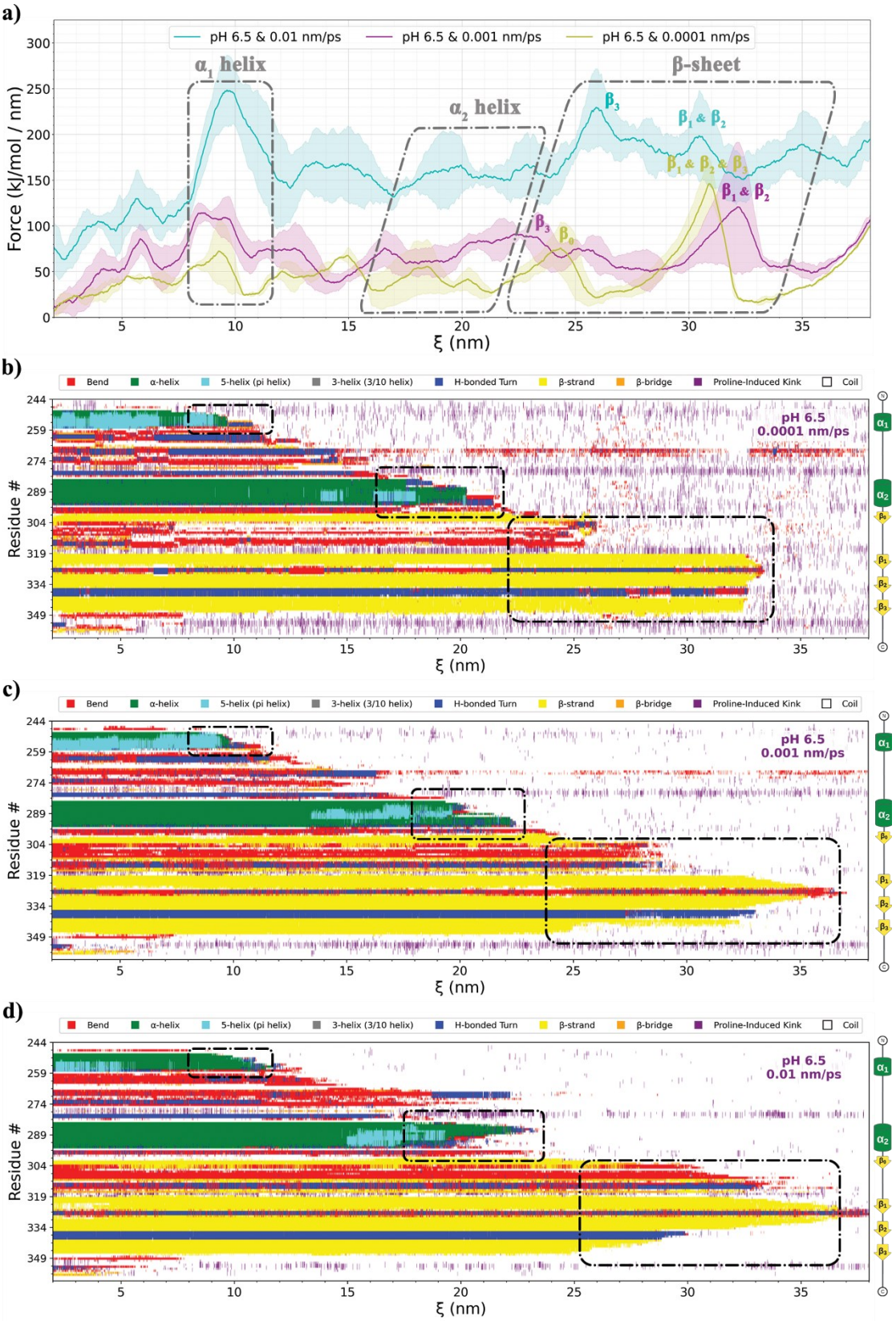

Figure 23: Analysis of Force and Structural Changes During Effector Domain Unfolding in Pulling Simulations. a) Force profiles as a function of the pulling distance, ξ (in nm), with the blue curve representing simulations at pulling rate **0.01 nm/ps**, the red curve at pulling rate **0.001 nm/ps** and the green curve at pulling rate **0.0001 nm/ps**, all at **pH 6.5**. The shaded regions around each curve indicate the standard deviation, representing the variability within the force measurements for each average profile. Regions corresponding to α_1 and α_2 helices, and β-sheet are highlighted. b) & c) & d) Derived from DSSP analysis, these plots detail the evolution of secondary structures along the effector domain sequence (residues) as it is extended, for pulling rates of 0.0001 nm/ps, 0.001 nm/ps and 0.01 nm/ps respectively. Key secondary structures are indicated by colour coding: green for α-helices, yellow for β-strands, and other colours representing different structural motifs, illustrating the effector's domain conformational changes over the simulation.

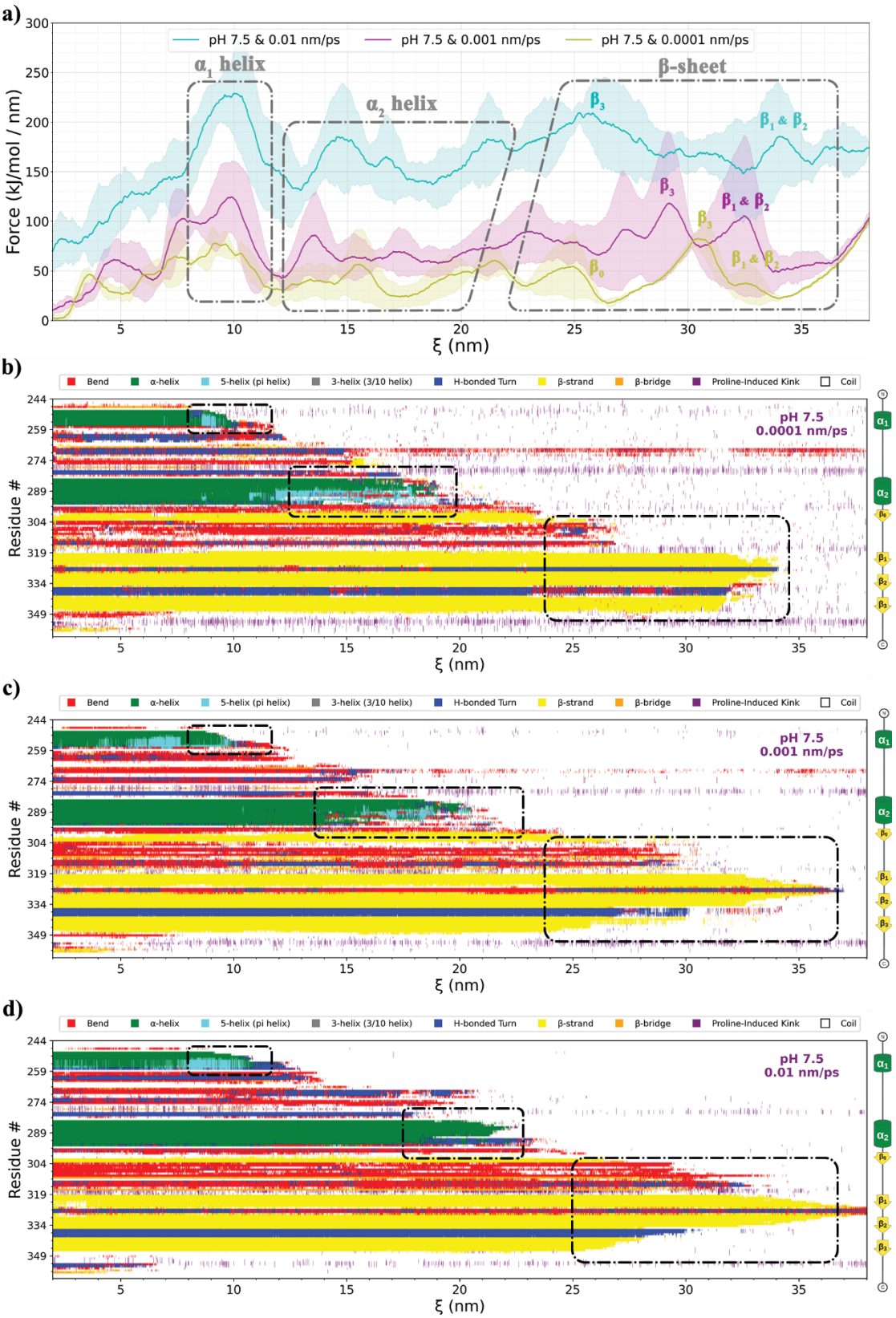

Figure 24: Analysis of Force and Structural Changes During Effector Domain Unfolding in Pulling Simulations. a) Force profiles as a function of the pulling distance, ξ (in nm), with the blue curve representing simulations at pulling rate **0.01 nm/ps**, the red curve at pulling rate **0.001 nm/ps** and the green curve at pulling rate **0.0001 nm/ps**, all at **pH 7.5**. The shaded regions around each curve indicate the standard deviation, representing the variability within the force measurements for each average profile. Regions corresponding to α_1 and α_2 helices, and β-sheet are highlighted. b) & c) & d) Derived from DSSP analysis, these plots detail the evolution of secondary structures along the effector domain sequence (residues) as it is extended, for pulling rates of 0.0001 nm/ps, 0.001 nm/ps and 0.01 nm/ps respectively. Key secondary structures are indicated by colour coding: green for α-helices, yellow for β-strands, and other colours representing different structural motifs, illustrating the t effector's domain conformational changes over the simulation.

Figure 25 displays the results from replica umbrella sampling experiments conducted at the two selected pH values, as analysed using the Weighted Histogram Analysis Method (WHAM). The bottom graphs show the histograms counts, visualizing the coverage of the umbrella windows for each simulation.

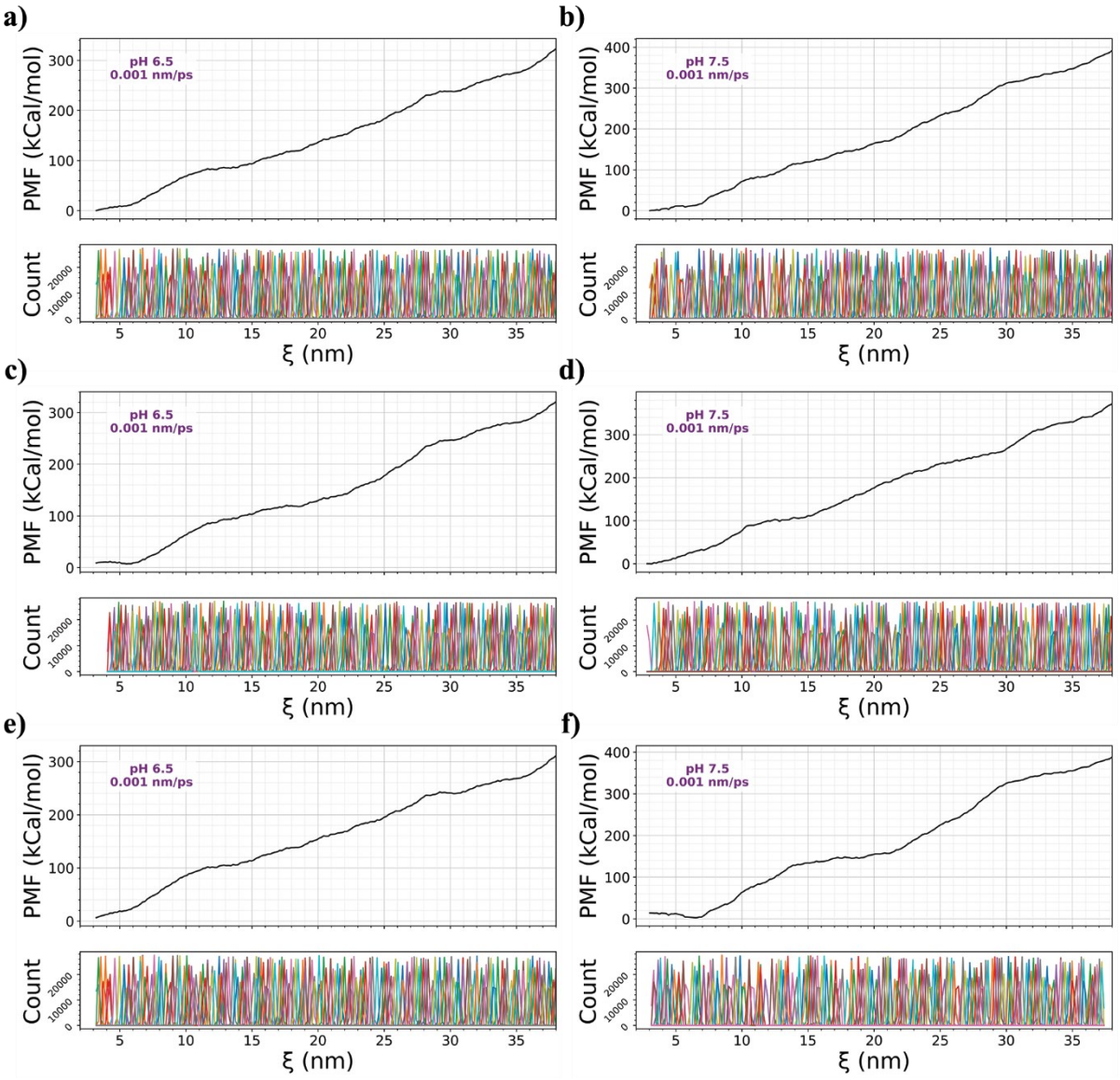

Figure 25: PMF and Histogram Distribution from replica Umbrella Sampling at the two **different pH** values. The top graph presents the PMF curve derived via the Weighted Histogram Analysis Method (WHAM), plotting PMF (kcal/mol) against the pulling distance, ξ (nm). The bottom graph shows the corresponding histogram counts for each umbrella sampling window, reflecting the distribution of sampled conformations.

Appendix B

The figure below (Figure 26) summarizes the methodology and results of a model smFRET experiment designed to observe molecular interactions at a nanoscopic scale. It combines a schematic of the experimental setup, a sample data trace, and the analytical framework for interpreting FRET efficiency and stoichiometry.

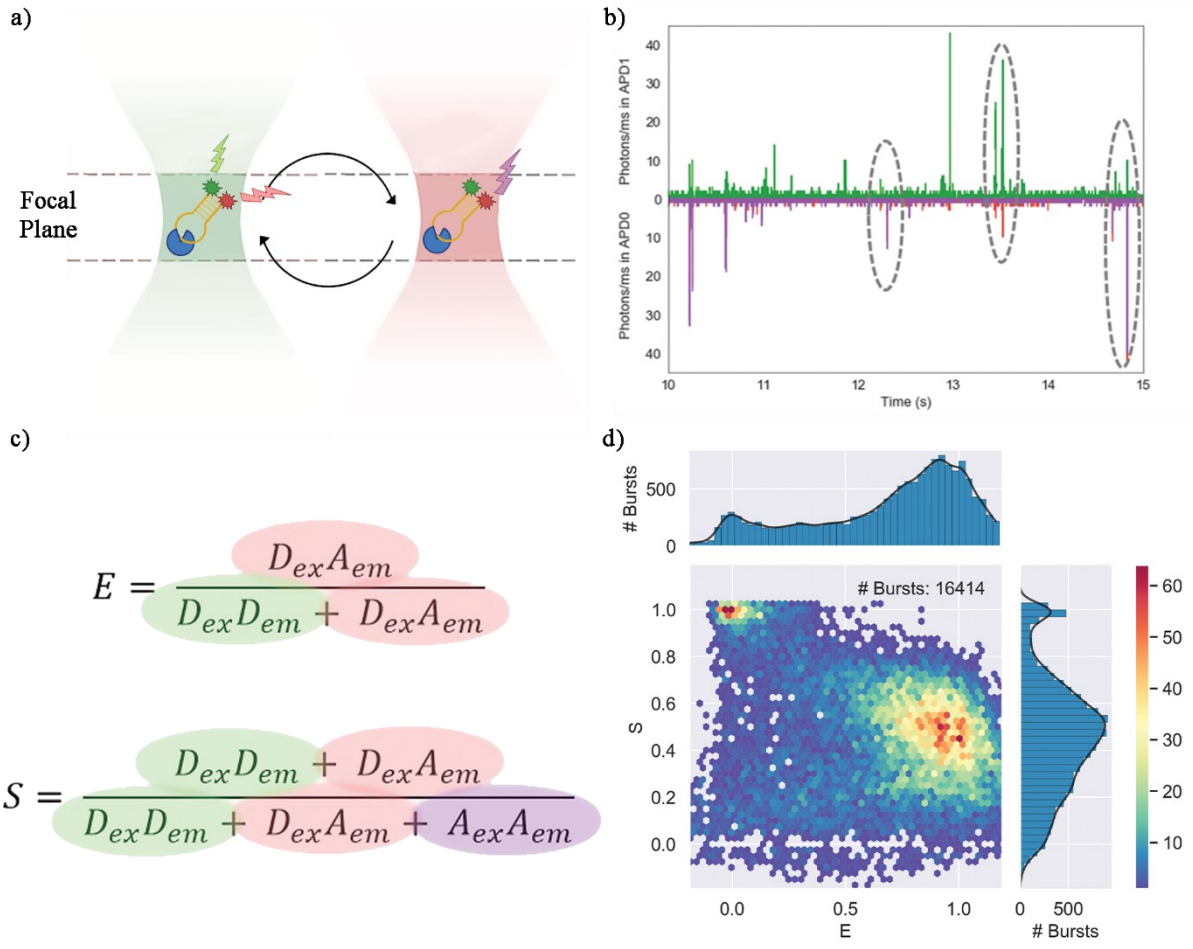

Figure 26: (a) Single molecules diffuse in and out of a confocal volume of approximately 1 μm^3 constructed by focusing the lasers into a diffraction limited spot and using lenses and a pinhole to select light in a thin focal plane. Alternating laser excitation (ALEX) between the donor (green) and acceptor (red) allows excitation of either dye multiple times per molecule as it transits the confocal volume. (b) A typical time trace for an smFRET experiment: green emission under green excitation is shown in green ($D_{ex}D_{em}$), red emission under green excitation shown in red, and red emission under red excitation is shown in purple. Encircled bursts from left to right are acceptor only, donor only, and doubly labelled. (c) Equations for E and S calculation. (d) A 2D ES histogram showing dye stoichiometry against FRET efficiency. Donor only molecules appear with low E but high S, and acceptor only molecules appear with low S. Doubly labelled molecules appear with intermediate S. Image and caption adapted from (Craggs Lab, 2024).

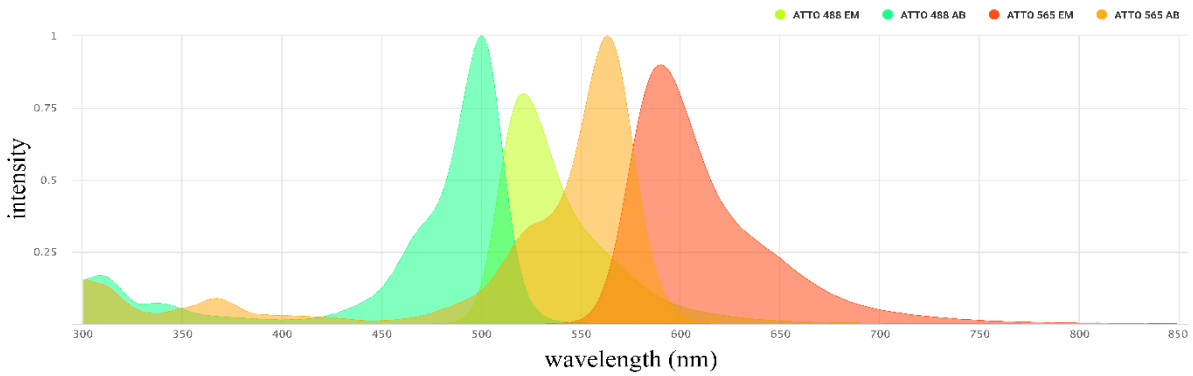

Figure 27: Absorption and emission spectra of fluorescent dyes used in the smFRET experiments, where ATTO 488 was attached to one end of the RNA hairpin, and ATTO 565 to the other end. The peaks indicate the wavelengths where maximum absorption and emission occur, critical for the FRET efficiency.

Table 2: Parameters used in data collection and analysis of the smFRET experiments.

Parameter	Value
ALEX window (alex_period)	4000 ns
donor exitation window (D_ON)	35-2150 ns
acceptor exitation window (A_ON)	2550-3900 ns
background period (time_s)	300 μs
minimum tail threshold of bg fit (tail_min_us)	auto
number of photons for burst (L)	10
background to signal ratio (F)	6
bursts selection threshold (.size th1)	30
FRET AA bursts selection (add.naa)	True
FRET AA bursts selection threshold (.naa th1)	15

To derive the correction factors necessary for adjusting the FRET (Figure 28 shows the physical problems of deriving accurate FRET (Craggs Lab, 2024), while Figure 29 show the stepwise workflow of the solution to accurate data (Hellenkamp *et al.*, 2018)), two distinct modelling approaches based on the nature of the data were utilized. Firstly, for alpha and delta correction factors, a Gaussian model was fitted to the histogram data. The alpha factor, which serves as a leakage correction, was calculated using the formula $\alpha = \frac{1}{S_{\text{only}}}$, where (S_{only}) is the centre value derived from the Gaussian fit. Conversely, the delta factor, which corrects for direct excitation, was calculated using $\delta = \frac{S_{\text{only}}}{1-S_{\text{only}}}$. Secondly, for the derivation of gamma and beta correction factors, an Expression Model was applied to fit the experimental data. The parameters β and γ from the fitted model represent beta and gamma, respectively. Gamma quantifies the fractional contribution of a secondary effect, whereas beta adjusts for the interaction between primary and secondary influences on the observed data.

The robustness of these methods ensures that the subsequent analysis was based on data corrected for systemic biases inherent in the experimental setup.

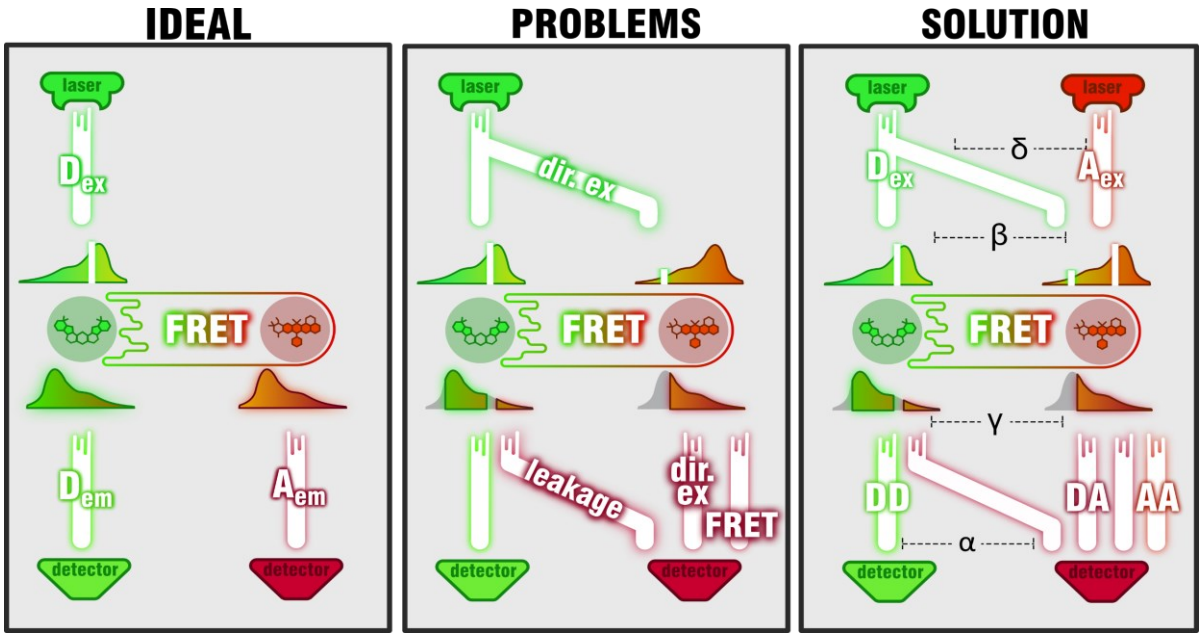

Figure 28: Schematic of how accurate FRET is achieved. (a) The ideal situation; the donor is excited by a laser and emission is separately detected from both the donor and acceptor with equal efficiency. (b) The problem; the acceptor can be directly excited by the laser due to the shorter wavelength tail of its absorption spectrum and the donor can leak emission into the acceptor detection channel due to the longer wavelength tail of its fluorescence spectrum. Furthermore, detection efficiencies and quantum yields are not equal. (c) The solution; in addition to exciting the donor, the acceptor is also excited (at a longer wavelength) to distinguish the presence of either dye or gain access to comparable quantities to calculate correction factors. The amount of direct excitation of the acceptor with the donor excitation laser is corrected for by Delta. Beta relates the quantity of excitation of each dye by its corresponding laser (dependent on the relative excitation efficiencies and laser powers). Gamma relates the relative emission and subsequent detection efficiencies of each dye. Alpha corrects for the emission of the donor into the acceptor channel. This caption and scheme were taken from (Craggs Lab, 2024).

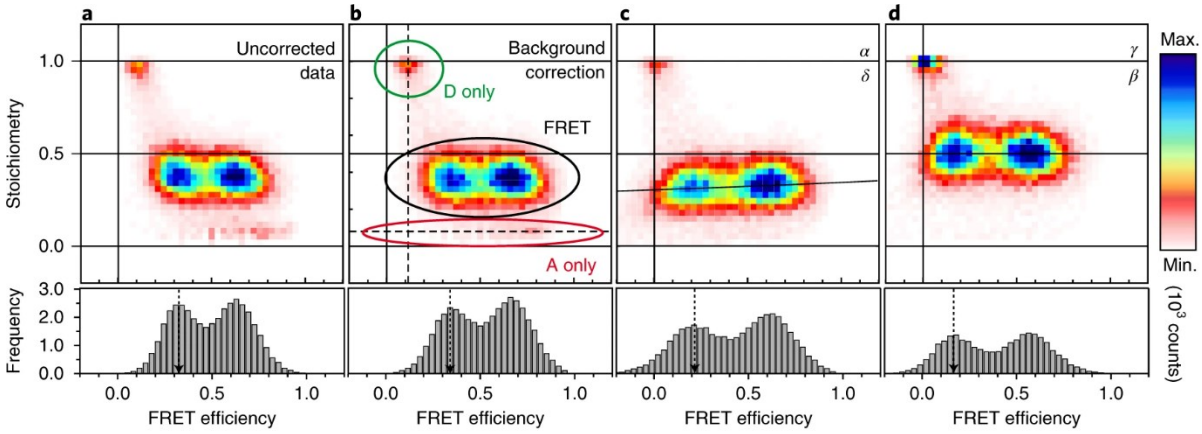

Figure 29: Stepwise data correction for smFRET efficiency and stoichiometry. Panels (a) through (d) illustrate the workflow for correcting confocal data for background (a to b), leakage (α), and direct excitation (δ) (b to c), as well as excitation and detection factors (β, γ) (c to d). This image was taken from (Hellenkamp *et al.*, 2018).

Table 3: Specific parameters used in data analysis of smFRET experiments for the studied samples.

Sample	leakage α *(leakage)*	direct excitation δ *(dir_ex)*	gamma-factor γ *(gamma)*	beta-factor β *(beta)*
RNA (TKM)	0.15	0.16	0.8	0.5
RNA (TKM + BME)	0.17	0.15	1.2	0.7
RNA (TKM + TROLOX)	0.21	0.16	1.1	0.5
RNA (TKM + TROLOX + BME)	0.19	0.16	1.3	0.6
RNA + Toxin	0.19	0.16	1.0	0.7
RNA + inactive Toxin	0.16	0.14	1.3	0.6
RNA + Toxin (incubated)	0.11	0.10	1.5	0.9
RNA + inactive Toxin (incubated)	0.15	0.17	0.9	0.8
encRNA	0.18	0.16	1.6	0.7
encRNA + Toxin	0.20	0.18	1.0	0.9

The smFRET analysis revealed that correction parameters such as leakage (α), direct excitation (δ), gamma-factor (γ), and beta-factor (β) varied across different experimental conditions. These variations can largely be attributed to the inherent properties of fluorophores, the presence of buffers and additives, and the experimental setup. For instance, the quantum yield and detection efficiencies of fluorophores, which are sensitive to the optical setup and sample environment, significantly impact the observed FRET efficiencies (McCann *et al.*, 2010). Furthermore, the addition of substances such as BME and TROLOX appeared to modify the microenvironment around the fluorophores, affecting their interactive dynamics and, consequently, the FRET efficiency (Schwarz *et al.*, 2019). Interactions with toxins, whether active or inactive, were also observed to influence the structural or conformational states of the samples, thereby altering the FRET measurements. This suggests that both the physical state of interacting molecules and the chemical composition of the solution play critical roles in determining FRET efficiency (Wallace and Atzberger, 2017). These insights are pivotal for interpreting FRET data accurately and underscore the necessity of considering environmental and experimental variables in FRET-based measurements.

The results depicted in Figure 30 distinctly showcase the effects of the CdiA toxin (both active and inactive forms) on RNA hairpin cleavage after 60 minutes incubation at room temperature. The most prominent effect is observed with the active form of the CdiA toxin. The E-S plots clearly show a complete cleavage of all RNA hairpins post-incubation with the active toxin. This is evident from the data as the samples predominantly transition to an acceptor-only population (bottom of the E-S plots). This drastic shift signifies that the active toxin efficiently cleaves all the RNA hairpins, leaving no moderate or low FRET populations.

In contrast, the inactive form of the toxin demonstrates a subtler interaction with the RNA hairpins. Particularly in the free-in-solution samples, there is a noticeable shift of the majority of the high FRET population to slightly lower FRET efficiency, mirroring the behaviour observed in unincubated samples. This observation similarly indicates that although the inactive toxin does not induce significant cleavage of the hairpins, it appears to bind to them, as evidenced by the remaining, yet altered, high FRET population.

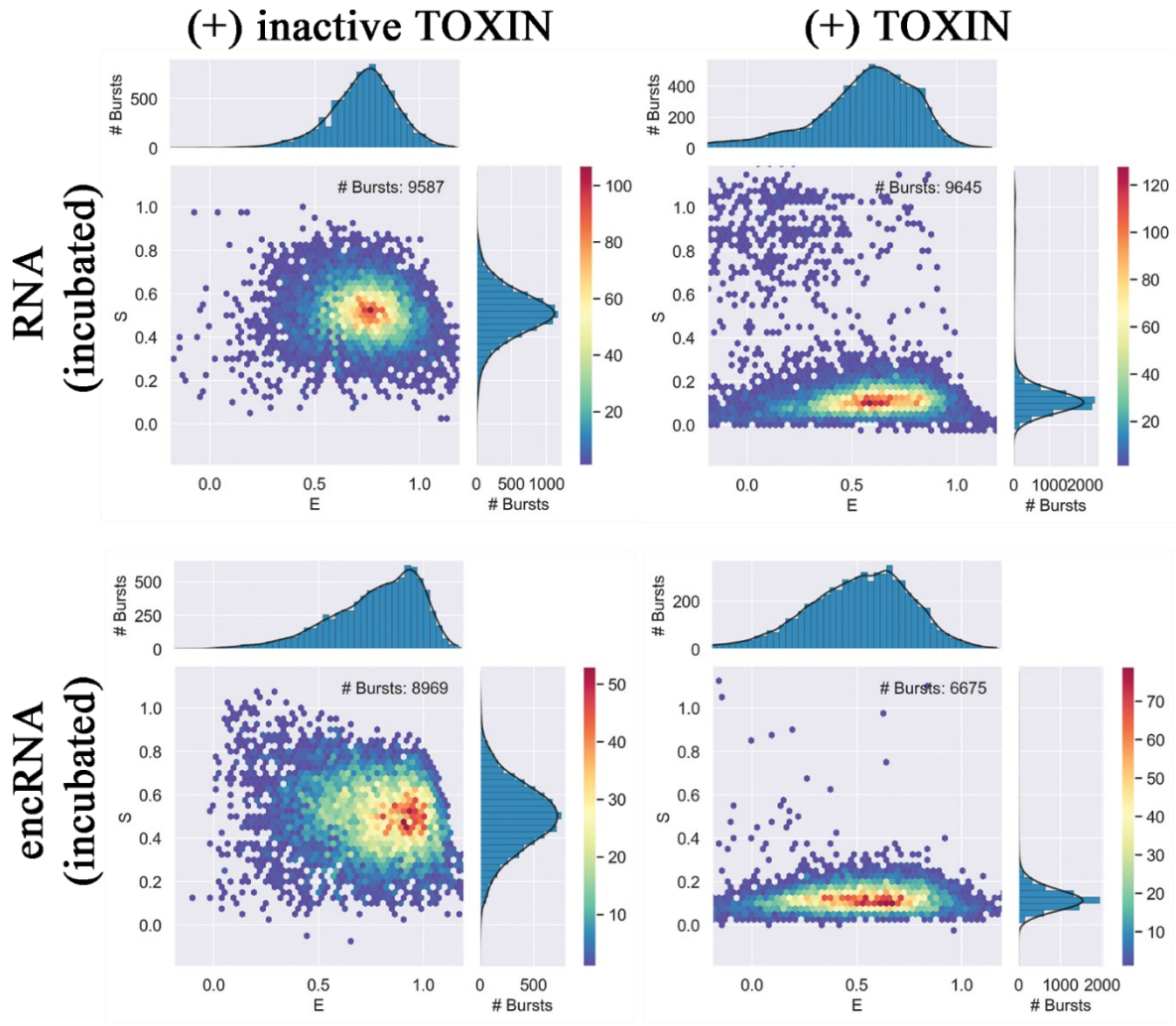

Figure 30: E-S plots showcasing the effect of the CdiA toxin (active & inactive) on RNA hairpin cleavage **after 60 minutes at RT**. The differences in burst densities across the plots reflect the varying influences of the active and inactive forms of CdiA on the RNA hairpin's structural dynamics and cleavage efficiency.