Mechanisms and Measurements of Sialidase in Pneumonia

Prepared by:
Carla Coetzee
CTZCAR020

Scientific Computing Research Unit
Department of Chemistry
University of Cape Town

Declaration

I declare that this dissertation titled, ***Mechanisms and Measurements of Sialidase in Pneumonia***, is a presentation of my original research work done at the Scientific Computing Research Unit, Department of Chemistry, University of Cape Town, South Africa. No part of this thesis has been submitted elsewhere for any other degree or qualification. Whenever contributions of others are involved, every effort is made to indicate this clearly, with due reference to the literature, and acknowledgment of collaborative research and discussions.

Signed by candidate

Carla Coetzee

Acknowledgements

I would first like to sincerely thank all my funders for providing me the opportunity to pursue this career, including the South African Research Chair Initiative (SAChI) and the UCT Post-Grad funding office.

I would like to thank my partner in life and in love, Nicholas Bowden, for his endless support.

I would also like to thank my future supervisor, Prof. Kuttel for her support and guidance throughout this process.

I also want to acknowledge my fellow students who supported me endlessly during this time. A special thank you to Mia Truter, Kyllen Dilsook, Lateef Nashed and Emre Kaya for their support and advice. I would particularly like to thank a fellow student, Tomas Bruce-Chwatt, for always being willing to help and offer support.

Thank you to Ms. Terblanché, who inspired me to pursue science in the first place.

Lastly, I would love to thank my parents, Fantie and Carlu, and my sibling Luc. This would not have been possible without them.

Abstract

Streptococcus pneumonia is a major human pathogen that resides in the upper airway and nasopharynx. It colonizes asymptomatically but can spread to other regions in the host body causing diseases such as bacteremia, otitis media and meningitis. This pathogen has evolved to evade the host-immune response and make individuals susceptible to reinfection by encoding three enzymes: NanA, NanB and NanC where NanA is the focus of this study. NanA cleaves 3'-sialyllactose or 6'-sialyllactosefound on the host cell walls to release N-acetylneuraminic acid (sialic acid) and lactose. This aids the bacterial adhesion which contributes to the overall virulence of the pathogen. Thus, one aim of the pharmaceutical industry is to design and develop inhibitors for these enzymes. However, for any inhibitor screening (to evaluate the potency of a potential drug) certain kinetic constants are required that can only be obtained doing biological enzymatic assays. A prerequisite of these assays is to either have a detectable substrate or product via a spectroscopic method (e.g. fluorescence) to monitor the rate of the reaction. Neither 3'-sialyllactose,6'-sialyllactose nor their products display intrinsic fluorescence and only have absorbance. Unfortunately, the signals are often weak and broad and so it is not a sensible method for detection. An artificial substrate, 2'-(4-Methylumbelliferyl)-α-D-N-acetylneuraminic acid (MUANA) is cleaved by NanA to produce sialic acid and 4-methylumbelliferone (4-MU). MUANA was introduced to mimic the natural 3'-sialyllactose and 6'-sialyllactose as 4-MU (unlike sialic acid or lactose) is fluorescent. However, there is still the question as to how valid MUANA is in replacing either 3'-sialyllactose or 6'-sialyllactose. To answer this question, an enzyme kinetic experiment was conducted using MUANA to obtain kinetic constants. These constants were then compared to those of a modified assay using 3'-sialyllactose and 6'-sialyllactose as substrates. The constants differed across the two assays; thus, an inhibition screening was performed to test the potency of four commercial drugs against NanA using both the MUANA and 3'-sialyllactose. The inhibition constants agreed extremely well for the two most potent inhibitors but deviated slightly for the lowest ranked two. The inhibitors were ranked from most potent to least potent

in the order of Oseltamivir, DANA, Zanamivir and Peramivir. Two *in-silico* methods were used to verify the experimental results: molecular modelling and molecular docking. Molecular modelling was used to calculate the free energy of the reaction between NanA and MUANA and to compare to the free energy obtained by the experimental data. However, only one of the two required mechanistic steps could be simulated. Molecular docking is a virtual tool used to assess the binding affinity (potency) of a molecule, and thus was used to calculate a score to rank the inhibitors' potency. The *in-silico* approach yielded the same order as the experimental results. Thus, MUANA is a suitable replacement for the 3'-sialyllactose and 6'-sialyllactose.

Abbreviations

3’-sialyllactose	3’-sialyllactose
4-MU	4-methylumbelliferone
6’-sialyllactose	6’-sialyllactose
CHARMM	Chemistry at Harvard Macromolecular Mechanics
DFT	Density functional theory
ESF	Empirical scoring function
FEARCF	Free energies from adaptive reaction coordinate forces
HA	Hemagglutinin
HTS	High throughput screening
IgA	Immunoglobulin A
IMAC	Immobilized metal affinity chromatography
LOD	Limit of detection
LOQ	Limit of quantification
MD	Molecular dynamics
MEP	Minimum energy pathway
MM	Michaelis-Menten
MUANA	2’-(4-Methylumbelliferyl)-α-D-N-acetylneuraminic acid
NA	Neuraminidase
NanA	(Sialidase) Neuraminidase A
NanB	(Sialidase) Neuraminidase B
NanC	(Sialidase) Neuraminidase C
NETs	Neutrophil extracellular traps
Neu5Ac	N-acetyl-neuraminic acid
DANA	2-deoxy-2,3-didehydro-N-acetylneuraminic acid
NR	Newton-Rhapson
OC	Oseltamivir carboxylate
PBC	Periodic boundary conditions

PDB	Protein data bank
RFU	Relative fluorescence units
SCCDFTB	Self-consistent charge density functional tight binding
TBA	Thiobarbituric acid

Contents

List of Figures

List of Tables

Section 1

Chapter 1

Introduction

Streptococcus pneumonia is a major human pathogen that resides in the upper airway and nasopharynx. It colonizes asymptomatically but can spread to other regions in the host body, causing diseases such as bacteremia, otitis media and meningitis. Currently, it is the leading cause of pneumonia and upper respiratory tract infections and is a major contributing factor in preventable deaths for pre-school aged children, where it accounts for 11% of the total mortality[1]. Pneumococcus evades the host immune system in multiple ways. First, surface glycosides (carbohydrate molecules) contribute to virulence by aiding adhesion to the host cell. Second, it produces sialidases (a class of enzymes) that sequentially deglycosylase host glycans present on the cell surface (or extracellular matrix) as glycoproteins/lipids. The deglycosylation releases carbohydrates, providing a carbon source for the bacteria. Essentially, these enzymes modify the host immune system effectors to decrease the efficacy of the host immune response[2].

Sialic acids are major constituents of the host cell surface. Removal of these carboxylated, nine-carbon sugars promote infection, signaling and apoptosis. *S. pneumoniae* secretes sialidases that cleave terminal sialic acids from long chain carbohydrates (glycoconjugates) to aid bacterial adhesion[3]. These enzymes are known as NanA, NanB and NanC[4].

Enzymes are globular proteins. Polypeptide chains fold to form a three-dimensional structure, often with small pockets (active site) for specific substrate binding[5]. The lock and key model first described enzyme binding by the complementary nature of the substrate. Implying that only a substrate (key) of the correct geometry can fit into the enzyme's active site (lock)[6]. X-ray crystallography led to the 'induced fit' model where the enzyme changes its shape to

accommodate binding of the substrate[7]. Both models focus solely on the active site, implying that the remainder of the protein is inactive. However, studies show it stabilizes the active site and provides an environment for the substrate to interact with it[8].

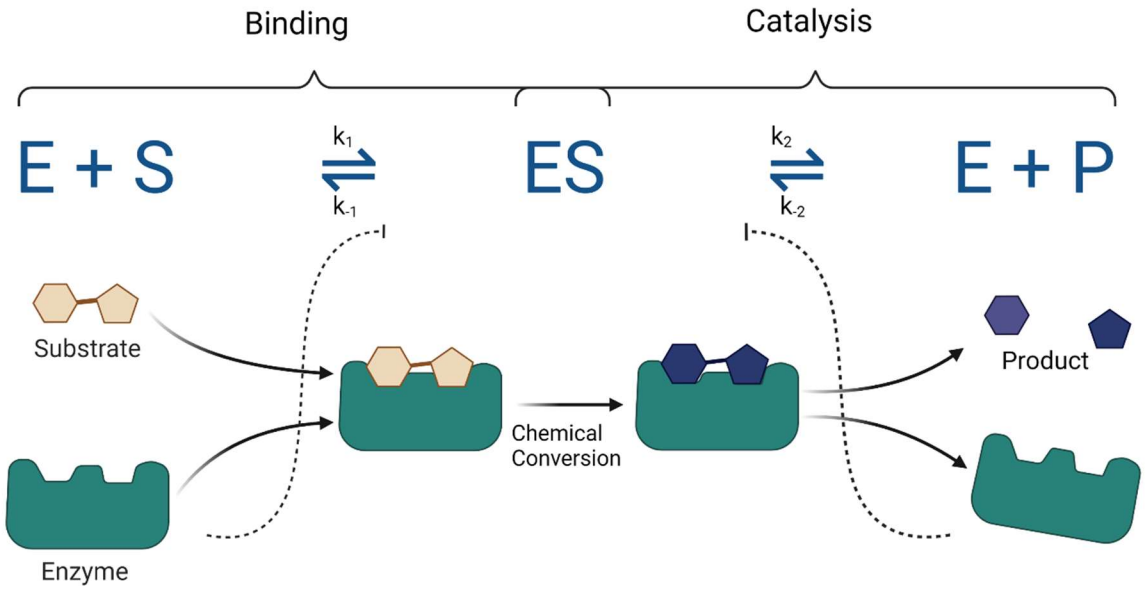


Figure 1. 1 The three main phases of a reaction catalysed by an enzyme.

.

An enzymatic reaction progresses in three stages, with two distinct phases (Figure 1.1). **E** and **S** represent the enzyme and substrate, respectively. The **ES** complex is where the substrate is bound to the enzyme and the reaction is catalyzed. The substrate is then converted into an intermediate form (transition state) and then into the final product. Once completed, product **P** dissociates from the enzyme. The free enzyme is recycled back into the reaction cycle where it can bind to another substrate. Enzymes take part in chemical reactions without undergoing change; they also lower the activation energy for the formation of the transition state but do not alter those of the reactant or product[9].

Enzymes are biocatalysts that facilitate chemical reactions. Many drugs inhibit enzymes that mediate disease phenotypes. Understanding the underlying biochemical and kinetic principles

facilitating enzyme actions allows for the design and validation of robust enzymatic assays (Figure 1.2). These assays can be used for high throughput screening (HTS) applications[10].

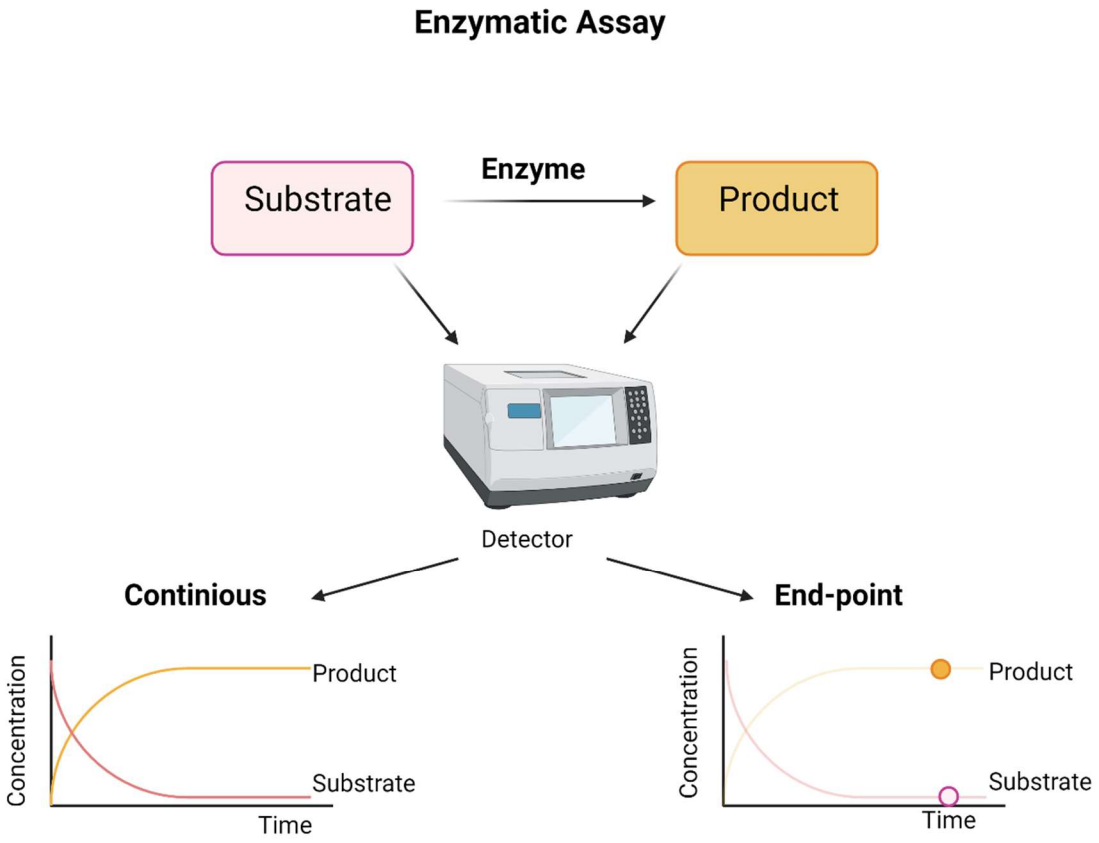


Figure 1.2 The two types of enzymatic assays: continious and endpoint.

Simply stated, an enzyme will convert one compound (substrate) into another (product)as seen in Figure 1.2[11]. The simplest enzyme assay only requires the enzyme and substrate. Often though, they will require multiple substrates, cofactors, activators, donors/acceptors, and other stabilizing molecules. Identification of the enzyme target, purity and availability is vital for the validation of an assay.

Identification and characterization of an enzyme requires the reaction to be measurable. Thus, it must be possible to quantitatively detect either substrate or product and find a clear, distinguishing signal between the two. The compound with the more pronounced signal will be

used and is irrespective of whether substrate depletion or product formation is detected. However, the latter is preferred as the concentration (and signal) at the start of the reaction equates to zero, and any increase is due to the progression of the reaction [12]. There are two methods of detection (Figure 1.2); continuous and endpoint. Spectroscopic methods used for continuous detection include fluorescence, absorption, and optical rotatory dispersion (ORD) and allow either substrate or product to be monitored in real time (Figure 1.2 bottom left). Absorption ultraviolet and visible spectroscope (UV-vis) is used to quantify biological substances. However, the limitation is the ability to separate two overlapping signals, as most biological molecules have similar absorption features. This is especially true for glycoconjugates, as after enzymatic cleavage they produce neuraminic acid and lactose, which both have absorption in the same region. Thus, glycoconjugates cannot be monitored with a continuous detection method[13].

Alternatively, the end-point assay (Figure 1.2 bottom right) can be used to quantify the product formation. Prior to detection, the product released is modified using a series of chemical reactions to produce a 'chromophore'. Chromophores are detectable with UV-vis and produce a strong, distinguished signal. However, the product must be modified first to quantify it and so cannot not be monitored in real time[14].

Fluorescence is the ideal detection method as it is 150 times more sensitive than absorbance. Unfortunately, few compounds have intrinsic fluorescence, glycoconjugates included. A continuous detection method allows for the product formation to be monitored in real-time and gives the most accurate estimate of the reaction rate needed for enzyme kinetics[15]. An artificial substrate was introduced to mimic the natural glycoconjugates with the advantage of being fluorescent. Glyconjugates are only detectable with an end-point assay approach and thus the rate of product formation is an estimate, rather than an exact value[16].

As *S. pneumoniae* encodes three key enzymes., there are two main approaches to combat infections: enzyme inhibitors and carbohydrate binding inhibitors[17, 18]. The challenges associated with carbohydrate derived drug molecules include poor bioavailability, the need for active transporters, short plasma half-life and poor metabolic stability causing rapid excretion[19]. However, current research aims to improve the pharmokinetic profiles by replacing some

moieties, introducing hydrophobic moieties, or focusing on a pro-drug strategy. The aim is to develop a strategy to design compounds like sialic acid, known as 'glycomimetics'[18, 20]. Inhibition screenings are used to evaluate the potency of these potential inhibitors against a protein target. However, these screenings require kinetic constants derived from enzymatic assays.

Simulations can also be used to investigate the efficacy of inhibitor binding at lower cost than experimental approaches. One method would be to use molecular dynamics to simulate an enzymatic reaction to calculate the free energy. One of the kinetic constants derived from an enzyme assay can be converted to calculate free energy. Thus, using molecular modeling one can compare the free energy of the simulation to that of the experimental results to verify the mechanism.

Docking is a vital tool for drug design and occurs when a small molecule is allowed to bind to a larger molecule. Many X-ray crystal structures of the enzymes contain native (bound) ligands that may be removed. Other ligands can then be docked into the empty active site to evaluate the binding affinity relative to the native ligand. Essentially, the potential inhibitor is virtually docked into the enzyme and a score is generated ranking the affinity of the inhibitor.

Problem Statement

S. pneumoniae is a pathogen that encodes three sialidases that cleave terminal sialic acid from glycoconjugates on mammalian cell surfaces. Inhibiting these enzymes could effectively treat *S.pneumoniae* infections. To estimate the efficacy of an inhibitor, by inhibition screening would require experimental kinetic constants. 3'-sialyllactose and 6'-sialyllactose do not fluoresce, preventing accurate determination of the kinetic constants. An artificial substrate, MUANA, was introduced to mitigate this problem as it is detectable with fluorescence. It is currently unclear whether the artificial substrate is a suitable replacement for the natural glycoconjugate.

Aims

The aim of this work was to assess whether artificial substrate (MUANA) can replace the natural glycoconjugates found on mammalian cell walls. We approached this by comparing kinetic constants from enzyme assays using both the artificial and natural substrate with NanA. Next, we determined the potency of four commercial inhibitors against NanA using both substrates. Molecular modeling was used to evaluate the free energy of the reaction of NanA and the artificial substrate. Finally, molecular allowed the four commercial inhibitors to be ranked from most to least potent.

Research Questions

1. **Can we compare substrates on the basis of their kinetic constants?** For any inhibition screening there are certain parameters required from enzyme kinetic experiments to assess the efficacy more accurately. The natural substrate (glycoconjugate) has weak absorbance and no fluorescence thus an end-point assay approach is used. The main question is can we determine rate considering the product formation cannot be determined in real time? And how does it compare to the kinetic parameters Xu *et al* (2009)[4] obtained using this technique? Additionally, how does it compare to the kinetic constants of the artificial substrate as the product can be monitored in real time with fluorescence? The next step is to evaluate the potency of four commercial inhibitors against the same protein using two different substrates. Can we compare an inhibition screening across two different assays? Will the potency of the four drugs agree across the two different substrates?
2. **How do the *in-silico* results compare with those of the *in vitro* approach?** Molecular modelling is used to simulate the reaction between NanA and the artificial substrate. Can the free energy be determined from MD simulations? And will it agree with experimental free energy? Molecular docking is used to evaluate the potency of the four inhibitors

against NanA, ultimately ranking them. Will the ranked order be in the same order as determined with the experimental inhibition screening?

Approach

The work explores two different approaches to determine if an artificial substrate is a suitable replacement for the natural substrate*: in silico* and *in vitro*. Firstly, the *in vitro* component involves optimization of a fluorescence-based assay to obtain kinetic constants needed for inhibition screening. The artificial substrate 2'-(4-Methylumbelliferyl)-α-D-N-acetylneuraminic acid (MUANA) is cleaved by NanA into α-D-N-acetylneuraminic acid (Neu5Ac) and 4-methylumbelliferone (4-MU). Both the substrate MUANA and product 4-MU have intrinsic fluorescence. Thus, the instrumentation has to be optimized to detect the release of 4-MU. This is challenging as the instrumentation must be sensitive enough to quantify small amounts of 4-MU released, but not saturate the signal should the concentration of the product increases. The rates of different substrate concentrations are determined to give the kinetic constants.

Alternatively, the natural substrate has no intrinsic fluorescence and weak absorbance. A novel approach was taken to chemically alter the product released (Neu5Ac) to form a chromophore that gives a clear signal using absorbance. Thus, the natural substrate uses an end-point assay to determine the reaction rate. The main drawback of this method being that the rate is an estimation rather than an exact value determined by a tangent.

An inhibition screening was done across the two different substrates using four commercial inhibitors to determine the potency (measured by the IC_{50}) against NanA. The potency/efficacy was compared between the two substrates as this tells you whether the drug binds better to the enzyme than the substrate. Thus, should the IC_{50} scores be similar for the inhibitors it means that both substrates display the same affinity towards to the enzyme.

Lastly, the *in-silico* approach involves molecular modelling and molecular docking. Molecular modelling is used to simulate the enzyme (NanA) reaction with the substrate MUANA. The crystal was obtained from the protein data bank (PDB) whereafter it followed a series of minimizations

to obtain the correct starting structure. From there, the active site of the enzyme was isolated along with the substate to quantum mechanically simulate the rection, including bond breaking and bond formation. The MEP determined the reaction path, and the free energy can be read from the corresponding contour plot displaying regions of bond formations and breakage. Free energy can be compared to experimental free energy.

Molecular docking was used to virtually dock ligands (four inhibitors) onto a protein target (NanA) to determine the potency. The free binding energy can be determined using MM/GBSA and the score using a built-in empirical scoring function (ESF) in Schrodinger.

Contribution

The work evaluates the validity of replacing the natural glycoconjugate with an artificial substrate for enzyme assays. This is important as most neuraminidase inhibitors are tested using this artificial substrate. Also, we demonstrate that an end-point assay is less accurate than a continuous assay. We further compare *in-vitro* and in-silico results to see whether both rank inhibitors similarly and hence whether an artificial substrate can accurately mimic the natural human substrate.

Thesis Overview

This dissertation consists of two sections. Section 1 is the experimental, *in-vitro* approach while Section 2 is the *in-silico* approach. The work is as arranged:

Chapter 1 introduces core concepts, the scope, purpose, and scientific significance of this work, and the broad methodology involved.

Chapter 2 will provide the reader with background on the concepts relevant to the structure of the substrate (sialic acid), their applications in the drug discovery industry and the operation of *S. pneumoniae* enzymes.

Chapter 3 describes the methodology involved in expressing and purifying proteins (enzymes) to ultimately express and purify NanA for enzyme kinetics.

Chapter 4 starts with a theoretical overview of the continuous assay principle as well as the artificial substrate known as MUANA. This chapter then briefly describes the experimental methods used in this assay, followed by the results.

Chapter 5 starts with a theoretical overview of the end point assay principle. The chapter explores the success of different methods used to generate a final protocol for the assay.

Chapter 6 describes the methodology used for inhibition screening to determine the efficacy of four commercial inhibitors using both MUANA and 3'-sialyllactose as substrates.

Section2, chapter 7 introduces the theory describing molecular simulation techniques, including molecular dynamics and molecular docking. It will focus on the different strategies used for enzyme-substrate simulations as well as how docking can be used to fast-track drug screening.

Chapter 8 describes the methodology and shows the results of the molecular dynamic simulation of NanA with MUANA, as well as the docking result of the four commercial inhibitors.

Chapter 2

Streptococcus Pneumoniae

2.1 Systems Biology

Throughout history, humans have relied on herbal remedies to treat common afflictions. While initially such remedies could have arisen by chance, the accumulation of knowledge over generations led to extensive traditional medicine systems. In the modern world this process has been standardized using directed approaches to identify novel but highly effective drugs[21], giving rise to the term 'systems biology'.

Multiple drug targets can now be virtually screened against known; validated targets due to the advancement in technology. However, the number of FDA approved drugs still has not increased significantly over the past decade and is potentially due to screening against targets that are less understood from a biological viewpoint[22]. Thus, simply identifying a target is not sufficient as it does not explain the functionality nor interaction with the inhibitor, potentially leading to serious and unknown adverse effects. The modern approach to systems biology can be utilized to prevent this and potentially improve the subpar approval rate of new drug discoveries[23].

One aim of systems biology is to fully understand how a disease operates from the molecular level to the organism. This can be achieved by developing different models and an approach to reduce the complexity of biological systems. Developing adequate computational and bioinformatical tools will allow for the integration of mass amounts of reductionist data into the regulatory pathways to ultimately model cell behavior. The complexity of living organisms arise from the cross-interaction of many biological systems[21]. In order to mitigate this, relevant information from the molecular pathways must be linked to the whole organism[23]. The three key principles are illustrated in Figure 2.1:

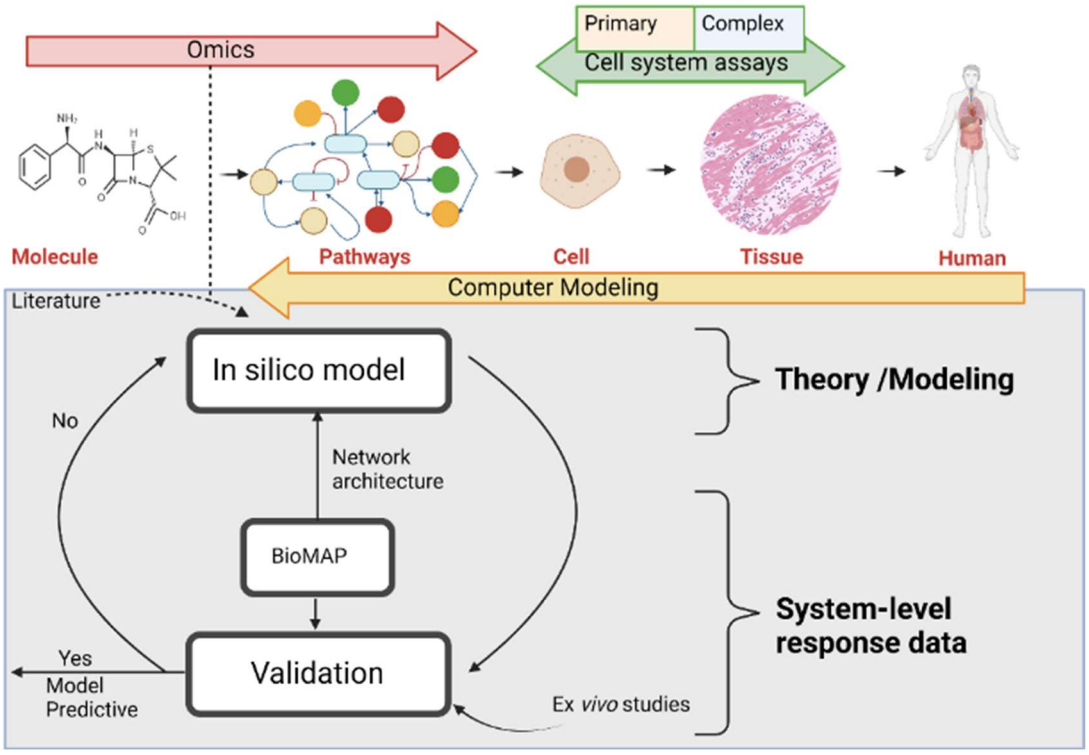


Figure 2.1 Systematic approaches to drug development.

The omics integration (Figure 2.1) is known as a bottom-up approach and can be equated to understanding a building's architecture by tracing it to the individual building blocks. Thus, reducing the complexity to identify possible primary components responsible for a diseased state. These components include genes, proteins and metabolites which can be expressed or upregulated to study the effect on the organism. Many pharmacal companies integrate protein measurements (proteomics), gene measurements(genomics) and metabolism measurements (metabolosim)to accelerate the study of complex biological processes. These measurements are vital as it can eliminate some pathways in the top-down approach, however, is limited to predicting the properties of the full system hence both approaches are necessary[1, 4].

Computational modelling (Figure 2.1) aims to predict or simulate human disease from the molecular pathway to the disease response using in silico methods. However, there are challenges associated with integrated models of simulated cell behaviour[21, 24].

Computational modelling aims to simulate the disease phenotype on either a tissue, cell or organ level response. This requires detailed input to identify potential drug targets and to predict the effect on an organ level. The difficulty lies in developing models to include and integrate molecular pathways, cellular pathways into organ level complexity and is limited due to the insufficient knowledge regarding the bottom-up approach and cross-interactions of different pathways. A possible solution would be to increase the complexity of the framework and to identify key nodes or regulatory points in pathway networks or how they interact, causing computational modelling to be one of multiple inputs[21, 25].

It is evident that both the bottom-up and top-down approaches have their advantages as well as limitations. The omics integration is too disconnected from the systematic disease as it only accounts up until the molecular pathway whereas the computational modelling can account for the cell, organ, and organism response. However, the latter is heavily reliant on cell response data. Cell system assays are used to bridge the gap between these two approaches.

These assays integrate the complex and emergent characteristics of the cell regulatory systems and are designed to be reproducible to be shared across databases and quantitively analyzed. High throughput screening utilizes assays to screen or profile drugs to investigate their effect on pathways or how they interact with the target. Despite constant development of HTS methods, individual pathways are mostly studied and ultimately limits the emergent properties of the cellular sytems[21, 26].

The effects of a drug must be studied in various environments to ensure accurate integration of pathway data. Primary and complex cell systems are two approaches used to achieve this. Briefly, primary cells are combined and exposed to stimuli thought to be linked to the disease. Proteins, genes, and other cellular constituents relevant to the disease serve as the output, giving insight as to how the system discriminates against certain drug and/or gene effects. Whereas with complex cell the systems are interrogated with drugs or genes and selected readouts are then determined to generate a profile showcasing how the test agent functions. A good example of this approach is the map clustering of gene interactions.

Thus, the best method for drug development and discovery is the systems biology approach as it provides fast interpretation and categorization of disease relevant cell and cellular system response. It is crucial for all three approaches (Figure 2.1) to be integrated to achieve its full potential.

2.2 Sialic Acid

Biomolecules play significant roles in the growth, development, and optimal function of the human body[17]. Glycoproteins and glycolipids consist of carbohydrates bound to proteins and lipids, respectively and play key roles in signaling, host-parasite interaction and the immune system. Altered glycosylation and sialylation are associated with diseases in humans such as cancer and related auto immune disorders[27]. Therefore, it is essential to decipher their biological role in finer detail.

Sialic or N-acetylneuraminic acids (Neu5Ac) are carboxylated monosaccharides located at the end of N-linked and *O*-linked carbohydrate chains (Figure 2.2). Some bacterial species synthesize sialic acids *de novo* while others scavenge sialic acid from the host[20, 28].

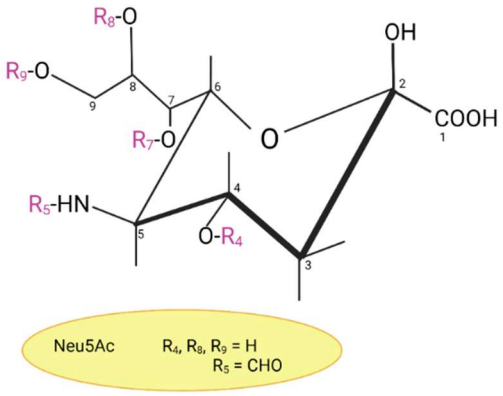


Figure 2.2 Core structure of sialic acid.

Sialic acids can provide a binding site for pathogens and toxins, by recognising specific linkages by means of extrinsic receptor. Acetylated and sulphated sialic acids are frequently targeted whereas methylated sialic acids are mostly not[29, 30]. Some bacteria engage in 'molecular mimicry', decorating their exterior with sialic acids to evade the host immune system.

Terminal glycans affect the structure and function of glycoconjugates and act as ligands for lectins, antibodies, and enzymes. These glycans also mediate cell-cell recognition, communication, development as well as regulating the lifetimes of glycoconjugates[31]. These terminal sugars are also vital in mediating bacterial and viral infections, tumor growth and metastasis. Due to their negative charge and hydrophilicity, they prevent blood component aggregation and are vital to red blood cell stabilization[32].

There are many diverse forms of sialic acid, mainly differing in position 5 of the amino group (Figure 2.2) or a hydroxyl group. The NH_2 moiety can be acetylated while the hydroxyl groups can be used to add phosphate, sulfate, methyl, or an acetyl group. More than 50 variations of sialic acid have been recorded with the two most common forms being Neu5Ac and N-glycolneuraminic acid (Neu5Gc). The core structure is modified by enzymatically adding acetyls, methyl or lactyl groups to positions 4,7,8 and 9[33]. Most common modifications include *O*-acetylation of the hydroxyl groups on C_4, C_7, C_8 and C_9. The monosaccharide can link to other sugars in either a α2,3, α2,6 or α2,8 configuration, further contributing to the diversity[34].

The N-acetyl group of Neu5Ac at position 5 is acetylated by acetyl coenzyme A (AcCOA) during the conversion of GlcNH2-6-P to GlcNAc-6-P (the precursor of UDP-GlcNAc). Bacteria can acquire sialic acid to metabolize them to use them for colonization, owing to the disease-causing property[17].

Strategies to combat these pathogens include: enzyme or carbohydrate binding receptor inhibitors. Examples include the design of sialyltransferase (ST) inhibitors to reduce cell surface sialylation[19]. *S. pneumonaie* leads to septicemia, meningitis and community acquired pneumonia in humans. Prophylactic vaccines are also used as a disease targeting strategy. However, there are challenges associated with carbohydrate derived drug molecules such as poor bioavailability, the need for active transporters, short plasma half-life and poor metabolic stability causing rapid

excretion. Current research aims to improve the pharmokinetic profiles by replacing some moieties, introducing hydrophobic moieties, or focusing on a prodrug strategy[27]. The aim is to design compounds mimicking the biological activity of the carbohydrate precursor, otherwise known as 'glycomimetics'. Zanamivir and Oseltamivir were designed on this principle and are FDA approved antivirals.

GNE is one of sixty genes involved in sialic acid biology and is primarily responsible for the synthesis of sialic acid. Mutations of this gene led to hereditary inclusion body myopathy and other related disorders depending on if sialic acid was over- or under expressed. It is thus clear that sialic acids are vital in maintaining a controlled cellular environment[28].

2.3 Lethal Synergism Between Influenza and *S. pneumonaie*

In 1918 an influenza pandemic occurred that resulted in almost 50 million deaths worldwide with the number almost doubling in the following years. Approximately 95% of these deaths were due to bacterial co-infections[35]. *S. pneumoniae* was the major contributing pathogen in this pandemic[35, 36].

S. pneumoniae is a gram-positive diplococcus that colonizes in the upper respiratory tract. Due to its diversity, it can be difficult to classify. Bacterial serotypes are linked to 90 capsules, complicating vaccine development and therapy[37].

Organisms develop hyper inflammatory responses to influenza often linked to decreased pulmonary clearance (congestion). While the infection mechanism remains unclear, it damages the lung alveoli epithelium.

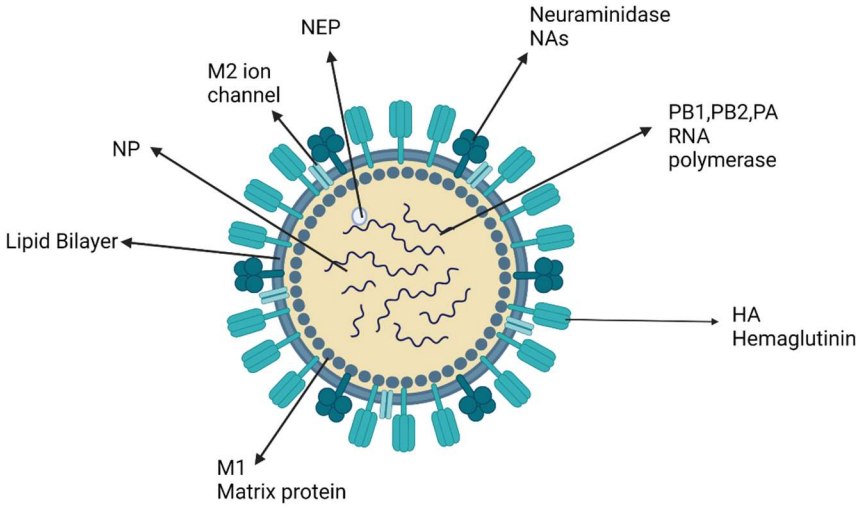


Figure 2.3 Structure of influenza virus

The influenza virion (Figure 2.3) is surrounded by a host-derived lipid membrane. The outside is covered in glycoproteins called hemagglutinin (HA) and neuraminidase (NA). They determine the virus subtype. As the virus enters the host cell, an immune response is launched to produce antibodies to combat infection. NA and M2 are both drug targets. The M1 matrix provides strength and rigidity to the lipid envelope. Inside the virion are viral RNAs that represent the genetic material of the virus and can encode up to two proteins. Each RNA segment is joined by several proteins (PB1, PB2 and PA). Together, these segments form the genes of the influenza virus[35, 36, 38]. Infection occurs when the virus reproduces in healthy cells (Figure 2.4) and mainly occurs in the lung epithelium.

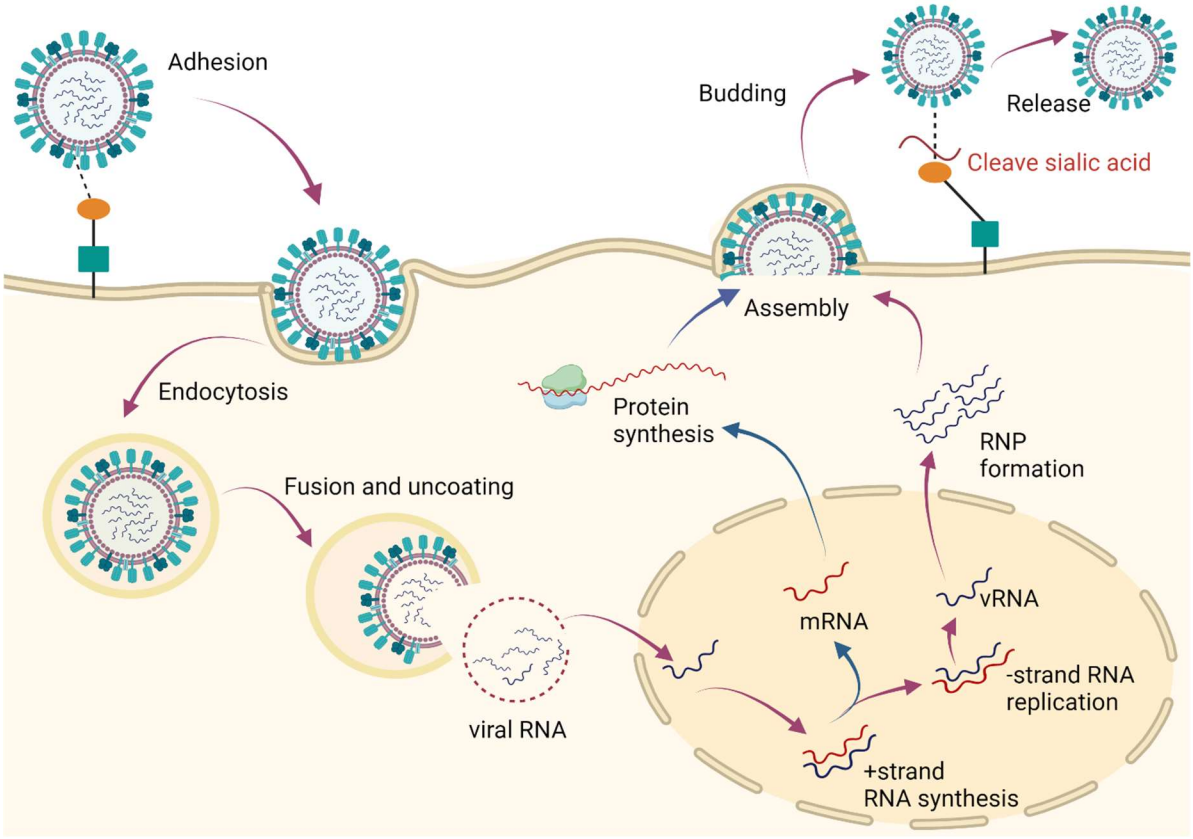


Figure 2.4 Cellular mechanism of Influenza virus.

Initially, HA binds to the sialic acid on the cell surface. The virus then enters the host cell via endocytosis. The acidification of the lysosome activates the M2 protein inducing a membrane fusion and dissociation of the viral RNP core[38]. The viral RNA is transported to the nucleus where it will be replicated. The replicated RNA along with HA and NA will aggregate near lipid layers in the plasma membrane. Eventually, the plasma membrane will bud and form a viral progeny linked by interactions between the sialic acid and HA. NA will release the virion by cleaving the sialic acid, rendering the virion free to interact with other healthy cells[39].

Desialylation of the host cell epithelial aids pathogen adherence, causing individuals with influenza to be more susceptible to secondary pneumonia. The virion causes damage to the epithelial cells leaving exposed glycans. Pneumococcal strains have virulence factors that allow for adhesion to the receptors of these damaged cells[40].

Another contributing factor to the lethal synergism is the exaggerated immune response[41]. Neutrophils are terminally differentiated cells with a short lifespan and are mainly involved in clearing the body of bacterial cells via phagocytosis. Ingested bacteria are destroyed by either oxidant from NADPH or lytic enzymes. When pneumonia infects the body, neutrophils will increase exponentially in the span of a few hours while dendritic cells and macrophages (part of the immune response) do not follow the same rapid progression. The influx in neutrophils result in fewer of them having the same antibacterial properties; along with a decrease in phagocytosis the outcome can be fatal[42, 43]. Additionally, *S. pneumoniae* can have mutations that protect it from the reactive oxygen species produced by the neutrophils and macrophages (Figure 2.5).

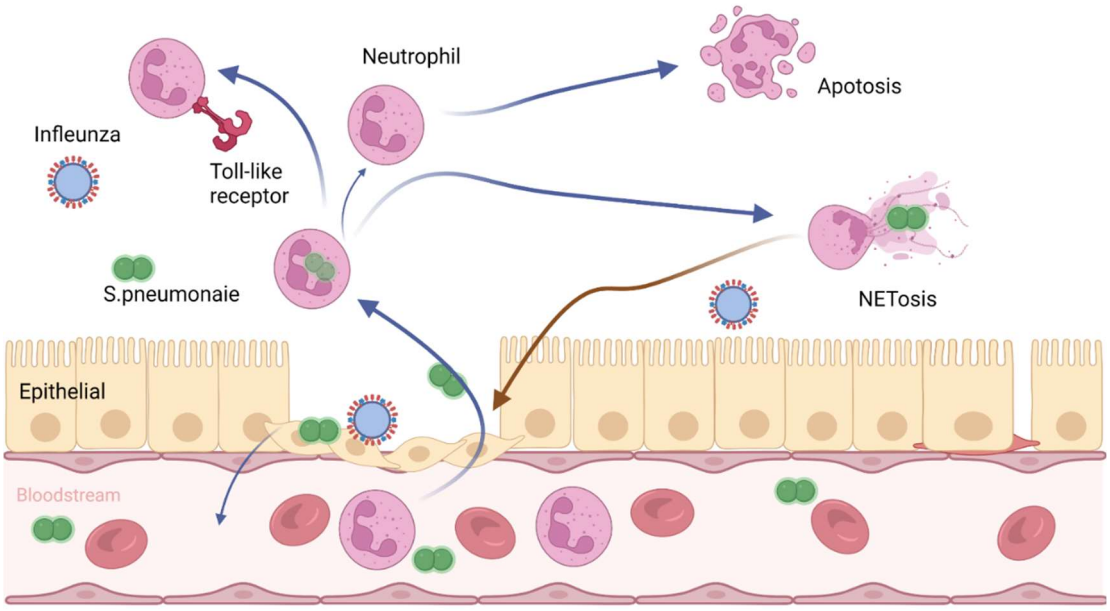


Figure 2.5 Influenza co-infection with S.pneumoniae.

Neutrophils release neutrophil extracellular traps (NETs) due to an induced enzyme (α-enolase) from pneumonia. Accumulation of these NETs in lung tissue can cause capillary damage[44]. The toll like receptors (Figure 2.5) form an essential part of the innate immune response as they recognize conserved patterns in various pathogens. For *S. pneumoniae* the receptors recognize

cellular wall parts of this gram-positive bacteria whereas for Influenza's RNA and DNA serve as targets. Although the innate immune response is essential for early mitigation of a disease, it can worsen pathology if elevated for long periods of time[45].

2.4 Pneumococcal Sialidases: NanA, NanB and NanC

Streptococcus pneumonia is the leading cause of pneumonia and upper respiratory tract infections. It is also a major contributing factor in preventable deaths, especially pre-school aged children where it accounts for 11% of the total mortality. The bacteria colonize in the human nasopharynx but can spread to different areas of the host and cause other diseases such as bacteremia, otitis, and meningitis. The mechanisms explaining the transition of the asymptomatic colonization into the pathogenic manifestation is still unknown[46].

Pneumococcus has evolved to employ different mechanisms to evade host immune responses. Firstly, it contains surface glycosides that contribute to the virulence factor by aiding with nutrient collection and adhesion to the host cell. Secondly, it encodes glycosides that sequentially deglycosylate host cell surface glycans. Deglycosylation releases carbohydrates, providing a carbon source for the bacteria. Essentially, these enzymes modify the host immune system effectors to decrease the efficacy[2].

Sialic acids are the main glycans found on the host cell surface. The removal of these carboxylated, nine-carbon sugars promote infection, signaling and apoptosis. *S. pneumoniae* secretes sialidases that will hydrolyze sialoglycosides (long chain sialic acid) to release Neu5Ac, whereas trans-sialidases transfer 2-deoxy-2,3-didehydro-N-acetylneuraminic acid (DANA) from α2,3-sialosides to other sugars. The leech intramolecular (IT) trans-sialidase produces a completely different product known as 2,7-anhydro-Neu5Ac instead of the sialidase product (Neu5Ac)[47].

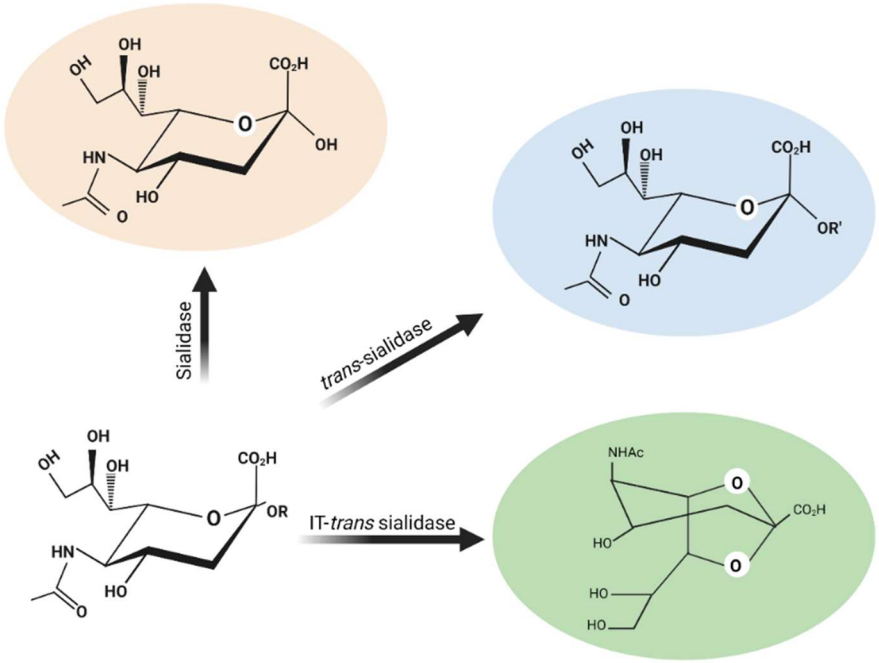


Figure 2.6 The three possible products formed by the sialidase enzymes.

Sialidases are either exo- or endoglycosidases. Exoglycosidases remove terminal sugars whereas endoglycosidases cleave non-terminal glycosidic bonds[48]. Different strains of pneumococcus encode up to three neuraminidases (sialidases) known as NanA, NanB and NanC. NanB is present in 96% of all pneumococcal genomes, NanC is only encoded in 51% while NanA is present in all the known pneumococcal genomes[46, 49].

NanA can cleave α2,3-, α2,6- and α2,8 linked sialic acids and acts as an exosialidase. It hydrolyses linked sialic acids to release α-Neu5Ac which will then convert into a more stable isoform known as β-Neu5Ac. NanB operates as an IT *trans*-sialidase producing 2,7-anhydro-Neu5Ac from α2,3- or α2,6 sialosides while NanC produces 2-deoxy-2,3-didehydro-N-acetylneuraminic acid (DANA) only from α2,3 linked glycosides[4, 47]. The products of the respective sialidases can be seen in Figure 2.7.

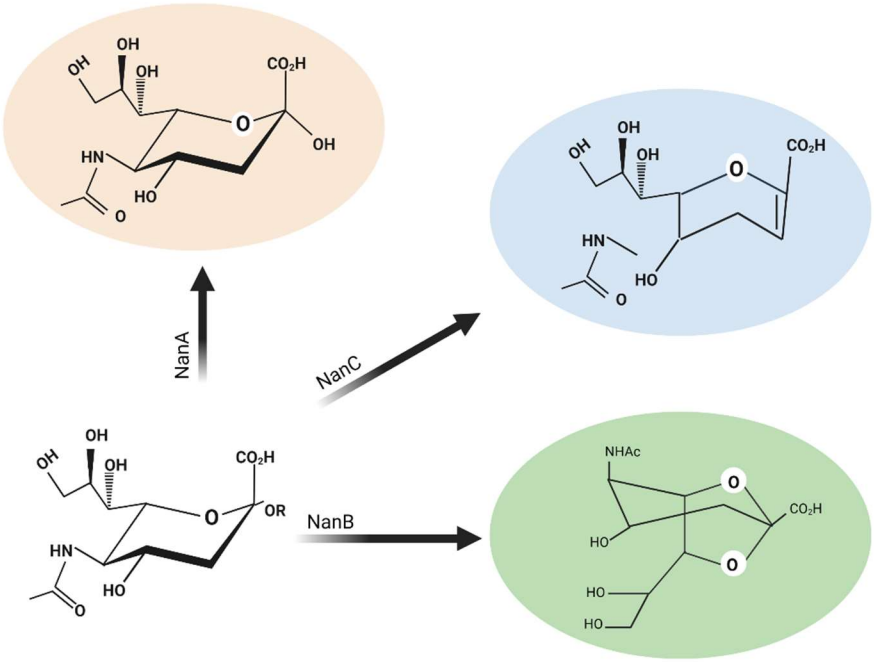


Figure 2.7 The three products formed the three pneumococcal sialidases.

2.5 Importance of Sialidase NanA

Influenza neuraminidase activity is better understood and documented than that of pneumonia[32]. NanA is the most frequently produced pneumococcal sialidases. The encoding gene is found on the Nan*AB* locus. It modifies the host glycoconjugates and exposes receptors used for binding.

Bacterial biofilms are highly organized community structures found on surfaces. They are encased in extracellular polymeric substances. NanA is thought to contribute to biofilm formations. Biofilms appear to favor re-infections[50, 51]. Glycosylation of immune proteins modify their properties, helping them evade the host immune response (Figure 2.8).

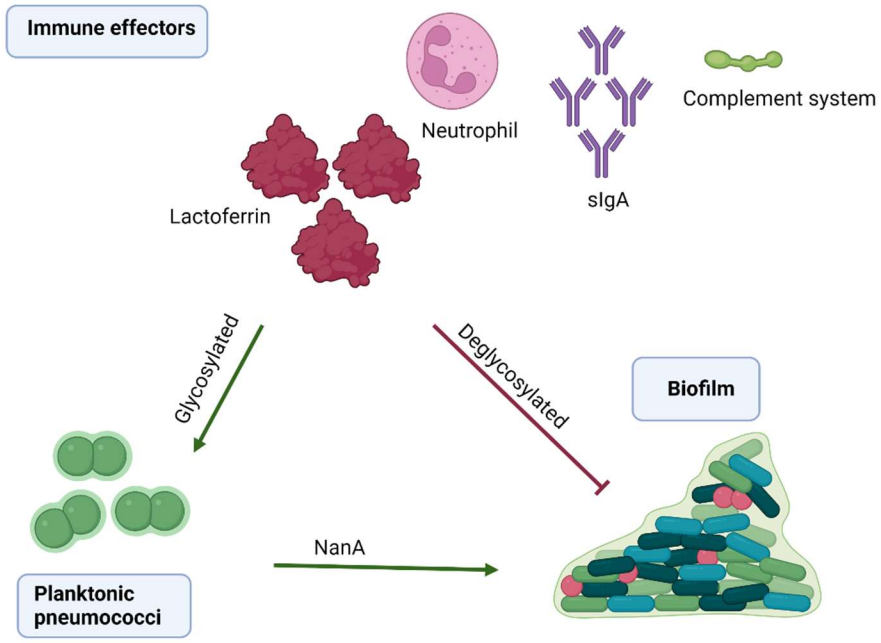


Figure 2.8 Glycosylated immune effectors prevent biofilm formation by attacking pneumococci.

Tagging foreign pathogens with opsonins will signal phagocytes to eliminate them from the body. C3b is a human receptor that regulates the complement system and binds to itself on the bacterial surface, inducing phagocytosis. NanA is an exosialidase and deglycosylates the complement system, preventing pneumococcal opsonization (tagging)[52]. An alternative theory is that deglycosylation of the complement system renders it more vulnerable to serum proteases, reducing the activity on *S. pneumoniae*.

Lactoferrin on the mucosal membrane of the human airway[2, 53] has bactericidal properties. By binding to pneumococcal surface protein A (PspA) it prevents adhesion to the host cells as well as biofilm formation. It also prevents planktonic cocci inhibiting further infection. Deglycosylated lactoferrin has reduced affinity for PspA[54, 55].

Immunoglobulin A (IgA) maintains homeostasis in the upper and lower airway. It recognizes surface antigens in pathogens, promoting phagocytosis. Subclass IgA2 contains sialated glycans that are cleaved by NanA, preventing bacterial recognition and hence an immune response[52, 56].

Sialic acids are a diverse group of molecules. Most commonly are Neu5Ac and Neu5Gc. Most animals encode cytidine-monophosphate-N-acetylneuraminic acid hydroxylase (CMAH) that converts Neu5Ac to Neu5Gc. Humans lack the CMAH gene ($CMAH^-$) such that only Neu5Ac is found on the cell surface. Infecting $CMAH^-$ mice with TIGR4 (strain of pneumonia) increases the infection rate relative to $CMAH^+$. Clearly the pathogen prefers Neu5Ac over the glycolate form. It was also shown that transcription of NanA increased when incubated with Neu5Ac compared to Neu5Gc.

NanA cleaves host sialic acid and is thus a virulence factor. Bacteria adapt to environmental changes by a two-component system. The first component, the cytoplasmic response regulator (RR) is activated by phosphorylation causing a conformational change and acts as a transcriptional regulator. *S. pneumonaie* has one RR, known as CiaR, that controls various promoters and genes. One of these genes enhances the resistance to oxidative stress, increasing the viability of this pathogen in the host and serves as the second component. It was determined that sialic acid uptake and metabolism is dependent on NanA, CiaR and sialic acid transporters. Thus, making this sialidase an ideal drug target for this study.

Chapter 3

Sialidase Expression, Characterization and Purification

3.1 Theory Overview

3.1.1 Plasmid Design

During bacterial colonization, *S. pneumoniae* cleaves receptors on the human respiratory epithelial surface. NanA's virulence is due to having both a lectin binding and catalytic domain and is the main target for antivirals. Molecular dynamics and visualization of the three-dimensional structure can offer insight as to why both domains exist. To fully characterize and obtain a kinetic profile for an enzyme, it must first be produced (Figure 3.1).

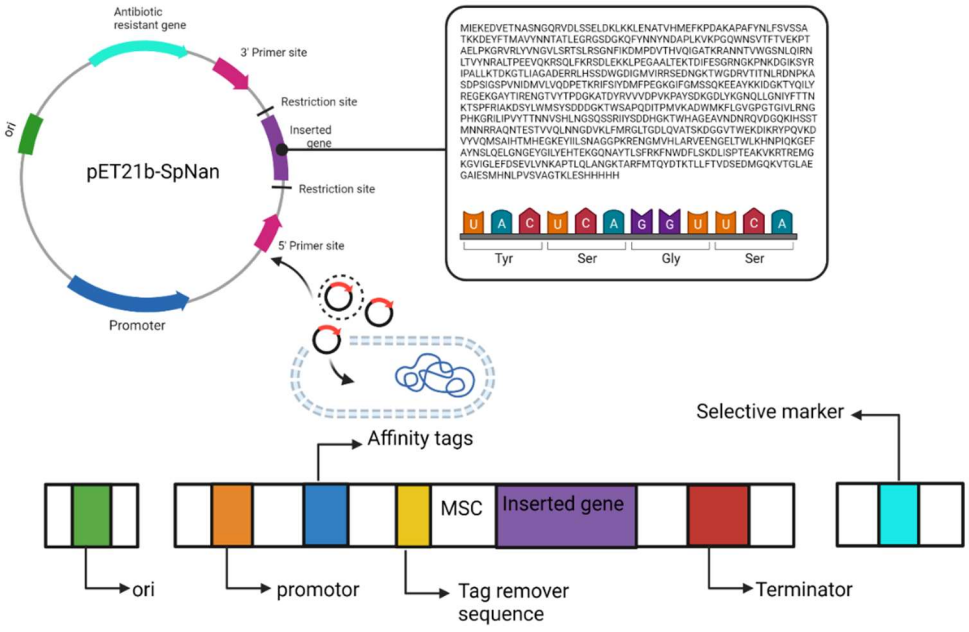


Figure 3.1 The structure of the expression vector with all major features included.

The production of recombinant proteins in microbial systems is a revolutionary discovery. Proteins can be produced that would not occur in nature, allowing for easier biochemical characterization.

Escherichia coli (E.coli) was chosen for its rapid growth, with a doubling time of 20 minutes under all optimum conditions. It also accommodates high cell densities of up to 1×10^{13} viable cells per ml.

The gene of interest is thus inserted into the expression vector and the vector used to transform *E. coli*. Induction and purification of the protein then follows[72].

Key elements allow expression vectors to facilitate protein production (Figure 3.1). The replicon (*ori*) contains DNA repeats and motifs that allow the DNA replication machinery to initiate DNA replication. In general, a higher plasmid equates to a higher recombinant protein yield. However, it sometimes imposes a high metabolic burden resulting in a decrease in bacterial growth. The most popular vector used is the pET series with a pMB1 origin (15-60 copies per cell) [59].

For prokaryotic organisms the most frequent promoter is the *lac* promotor, a component of the *lac* operon. Lactose is a sugar and will induce the system, resulting in protein production. Induction is hindered if alternative sources of carbon are used (e.g. glucose in rich media). If both lactose and glucose are present, the promoter will not be fully induced until all glucose have been metabolized. Cyclic adenosine monophosphate (cAMP) is produced as glucose levels drop, which will activate the lac operon[60]. Lactose permease facilitates lactose into the cell and is inactive in the presence of glucose. To minimize this effect, a mutated promotor (*lacI*) is used to reduce the sensitivity to catabolite regulation and increase the expression levels. The T7 promotor is mostly used for the pET vectors to ensure that the majority of the total cell protein consists of the target protein[59].

To inhibit the growth of plasmid void cells (that contain no target protein) a resistance marker is added. Specifically for *E. coli* an antibiotic resistance gene is chosen[61]. The *bla* gene is responsible for the resistance to ampicillin and encodes an enzyme that deactivates the β-lactam ring of the anti-biotics. It remains a popular choice and was used in our expression system[57].

The aim of any plasmid design is to obtain a pure, soluble, and active protein. The protein should be detectable throughout the expression and purification process. To achieve this an affinity tag (poly-His tag) is inserted on either the C-or-N terminal end of the protein. This allows the recombinant protein to be detected by western blot since it contains antibodies for the specific tags[61].

3.2 Methodology

3.2.1 Plasmid Expression

The bacterial host cells are transformed to induce uptake of the target plasmid. From there, the gene will be expressed. The protein will then be purified and quantified (Figure 3.2) [62].

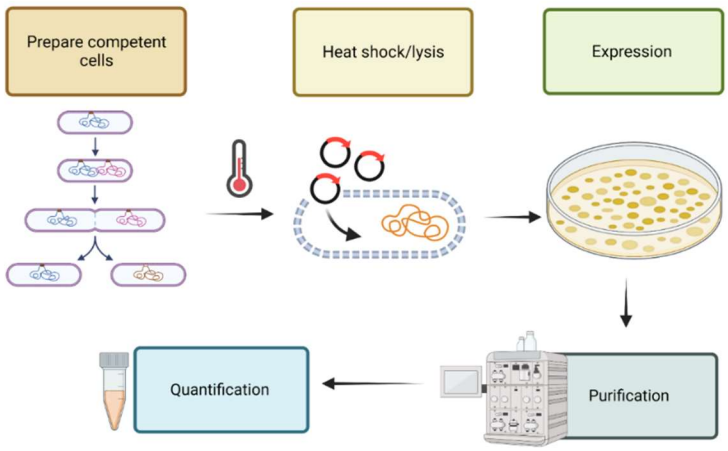


Figure 3.2 The general scheme for expressing a target gene in bacterial host.

3.2.2 Preparing Competent Cells.

From Figure 3.2 the first step of plasmid insertion involves the preparing competent cells of *E.coli* bacteria (BL21) and is performed in four steps as in Figure 3.3:

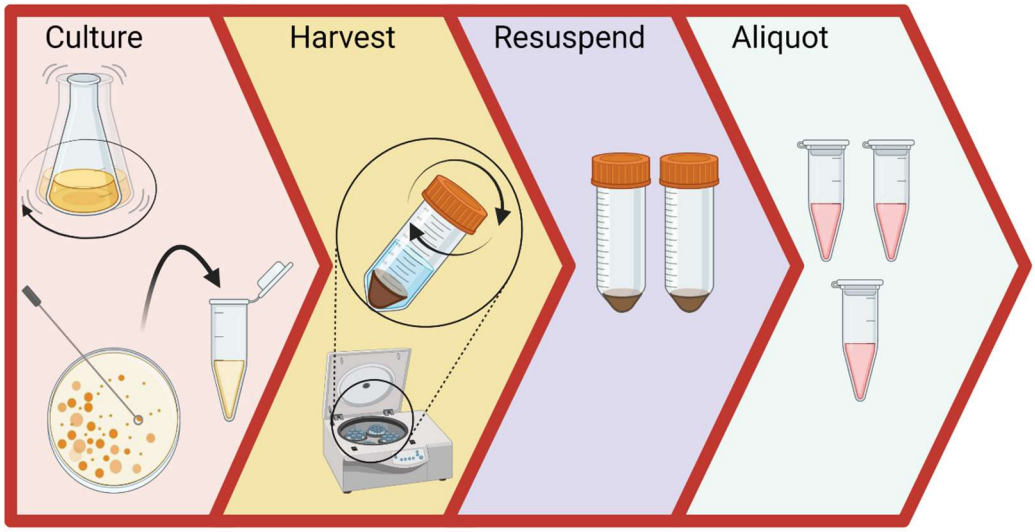


Figure 3.3 Preparing competent cells for target gene expression.

E.coli cells were streaked onto plain LB (nutrient rich medium) agar plates and left to grow overnight at 37°C. The medium is suitable for non-selective cultivation of *E.coli* strains and allows anti-biotics (ampicillin) to be added for selective breeding. After 24 hours the single colonies were picked and inoculated into 5mL LB media and then incubated at 37°C until the solution appeared slightly cloudy. It was then poured into a Falcon tube and centrifuged at 4500 rpm for 5 minutes. The supernatant (aqueous solution above cell pellet) is discarded, and the pellet resuspended in 3mL of cold 50mM $CaCl_2$.

The resuspended solution is left on ice for 30 minutes, then centrifuged for another 5 minutes at 4500 rpm. The supernatants discarded and the pellet resuspended in 1 mL of the $CaCl_2$ solution. The cell suspension is aliquoted into 1mL tubes and stored at 4°C.

3.2.3 Heat Shock Transformation

The second step is to induce the cells to take up the plasmid (Figure 3.2) by heat shock transformation. 0.5 µL of the plasmid stock is added to 80 µL of the competent cells in a 1 ml tube and placed on ice for 30 minutes. The temperature is increased to 42°C for 45 seconds. The sample is returned to ice for 5 minutes after which it is added to 900 µL of LB media and then

incubated for 1 hour at 37°C. The cells are then centrifuged at 8000 rpm for 4 minutes and 850 µL of supernatant is removed afterwards. The cells are resuspend ended in the remaining 50 µL supernatant.

3.2.4 Expression

The resuspended cells are spread onto a plate containing LB agar and antibiotics (ampicillin). The plates should not exceed 24 hours incubation as this can cause satellite colonies to form, containing no plasmid but are resistant to antibiotics. The colonies containing the target protein are then inoculated into 500 mL LB media (with ampicillin). Batch cultures are grown until an optical density between 0.6 to 0.8 is reached, before being induced with Isopropyl β-D-1-thiogalactopyranoside (IPTG). IPTG is a molecular analog of allolactose induces gene expression. After 24 hours of incubation the cells are centrifuged to pellet the cells (Figure 3.2).

3.2.5. Plasmid Purification

3.2.5.1 Affinity Chromatography

The pelleted cells (section 3.2.4) must be lysed and purified prior to protein quantification. Affinity chromatography is a separation method based on specific binding interactions, offering high resolution and selectivity. The specificity of the interactions determines the degree of protein purification.

Immobilized metal affinity chromatography (IMAC) was used for sample purification as it is selective for many macromolecules such as DNA, proteins, and polypeptides. IMAC can be performed at high salt concentrations. The high binding affinity of metal-ligand interactions allows for separation of proteins differing with only one amino acid in the sequence. Immobilized metal ions on the column interact with lone pairs of electrons on either the oxygen or nitrogen

containing amino acid side chain. For this experiment, Nickel was used as it shows preference to the nitrogen containing side chains.

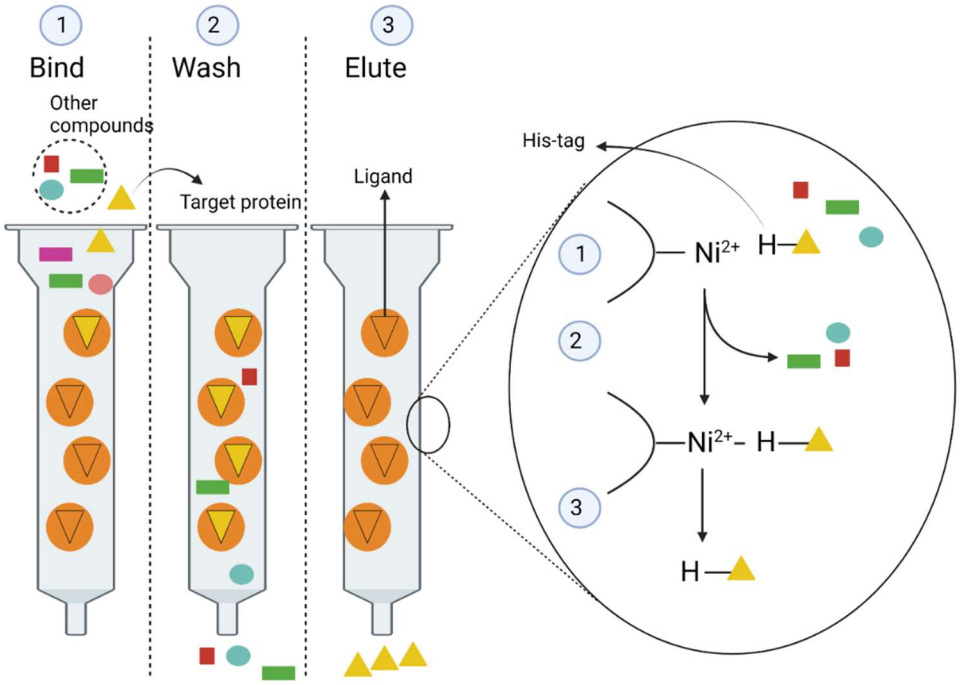


Figure 3.4 The three stages of IMAC purification.

There are three steps in the IMAC purification process (Figure 3.4). Firstly, the His-tagged protein will bind to the immobilized Nickel and cause salt and other components to elute first. All the his-tagged proteins are bound during the wash state. Lastly, a step-gradient of increasing concentration of imidazole is added, displacing his tagged proteins.

Briefly, the pelleted cells were thawed and resuspended in lysis buffer (10mL IMAC buffer,20 mg/ml lysozyme and a PI cocktail). The suspension was sonicated in bursts of 3 seconds for 1 minute and repeated 5 times with a resting period between each repetition. The lysed cells are centrifuged at 15000 rpm for 40 minutes, where the lysate is then passed through a sterile filter of 0.4 µM. The His trap chelating column (1 mL) is prepared by preloading it with Ni^{2+} from a $NiSO_4$ solution. The sample is loaded onto the column and the imidazole concentration is slowly increased to 500 mM.

3.2.5.2 Desalting

Desalting columns are a form of size exclusion chromatography[63]. The columns are filled with porous matrix beads. Small molecules can enter the pores or move along with the mobile phase[64].Thus, smaller molecules will elute later than the target proteins as there is large discrepancy in the molecular size (Figure 3.5)[65, 66].

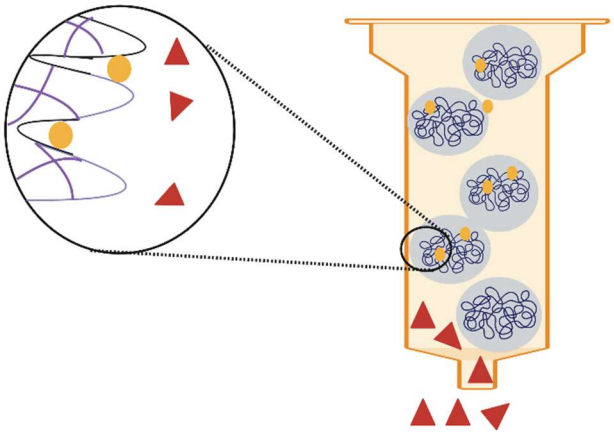


Figure 3.5 The concept of desalting columns used in protein purification. The pore size is selected to exclude salt (yellow). The target proteins (red triangle) are larger than the salt and will elute first after the void volume.

2 mL of the IMAC sample was loaded onto a His-trap desalting column. The buffer consisted of 25 mM Tris and $MgCl_2$.

3.2.6 Protein Quantification

Unidentified proteins of unknown extinction coefficient are often quantified using the Bradford assay. Serial dilutions of bovine serum albumin (BSA) are used as standards. The Bradford 'dye' (Coomassie G-250) is then added to each standard[67] resulting in a blue color. The readings are taken at an absorbance of 595nm.

3.3 Results and Discussion

3.3.1 Purification

3.3.1.1 Affinity Chromatography

From section 3.2.5.1, the sample was loaded onto the 1mL column, and the elution gradient was slowly increased up to 500mM imidazole, resulting in the first chromatogram in Figure 3.6:

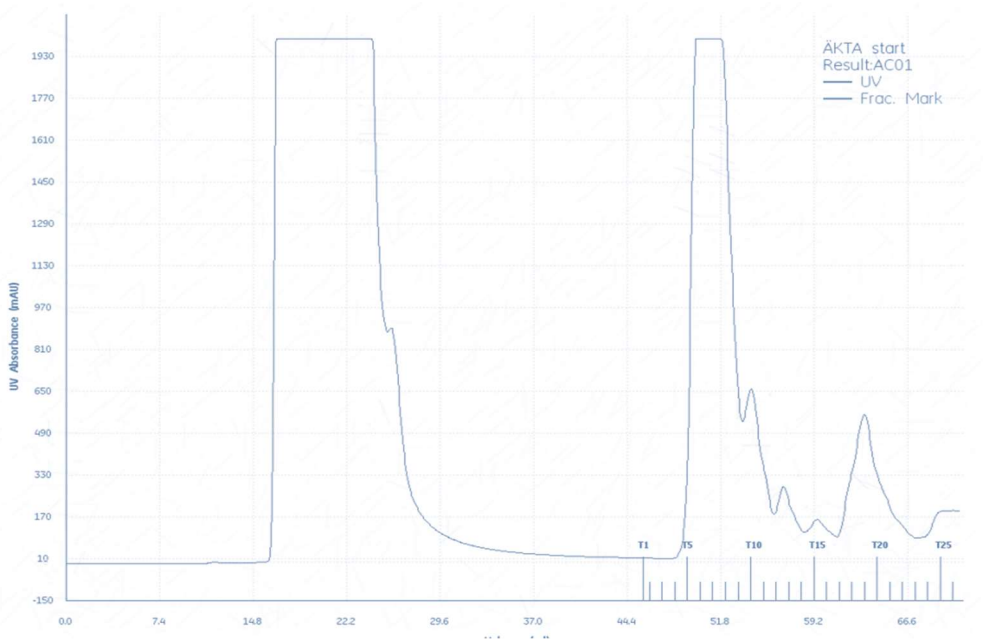


Figure 3.6 The resulting chromatogram after the first IMAC run.

The tubes corresponding to the peaks were loaded onto the acrylamide gel, where electrophoresis was used to separate compounds based on molecular size[43]. The wells were labelled as in Table 3.1, with the resulting SDS-page in Figure 3.7.

Table 3.1 Different Fraction used on the SDS-page.

Label	Tubes used from chromatogram (Figure 3.5)
Unmarked	Molecular marker
E1	T5-6
E2	T7
E3	T8
E4	T9
E5	T11-12
E6	T17-22

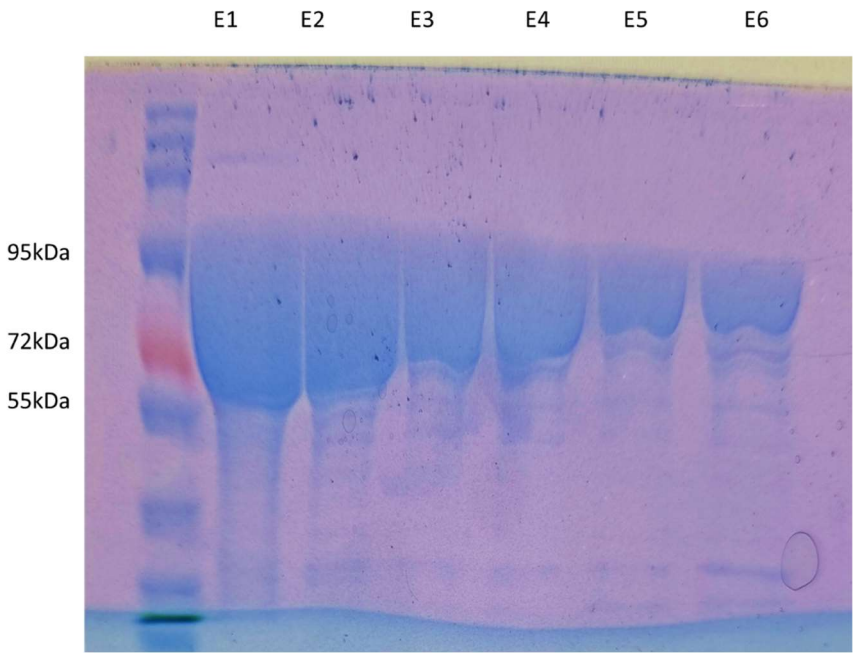


Figure 3.7 Gel fractions that were collected during IMAC.

The first lane (E1) represents the largest eluent from Figure 3.6. E3 and E4 seem to have similar impurities. E5 and E6 are perceived to be the least pure as they contain the least amount of target protein. All fractions (E1 to E4) were combined and diluted 2x to undergo a second round of IMAC, producing the chromogram in Figure 3.8:

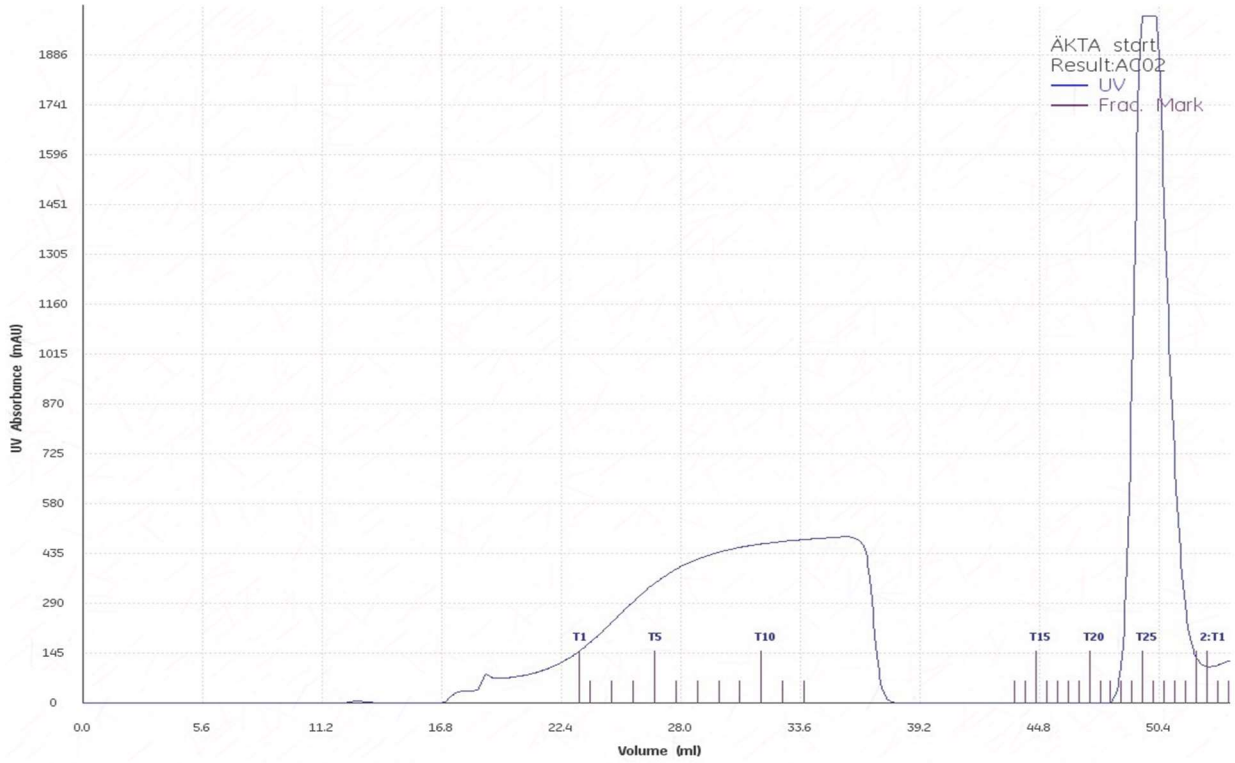


Figure 3.8 The resulting chromatogram after the second round of IMAC.

The first, broad peak represents the flow through and was isolated as it may contain unspecific binding (contaminants) or incomplete binding of the target protein. The second, narrow peak represents the His-tagged proteins after the elution gradient. Some of the sample(prior to second IMAC) was kept aside for comparison. All three samples were loaded onto the gel, resulting in the SDS page in Figure 3.9.

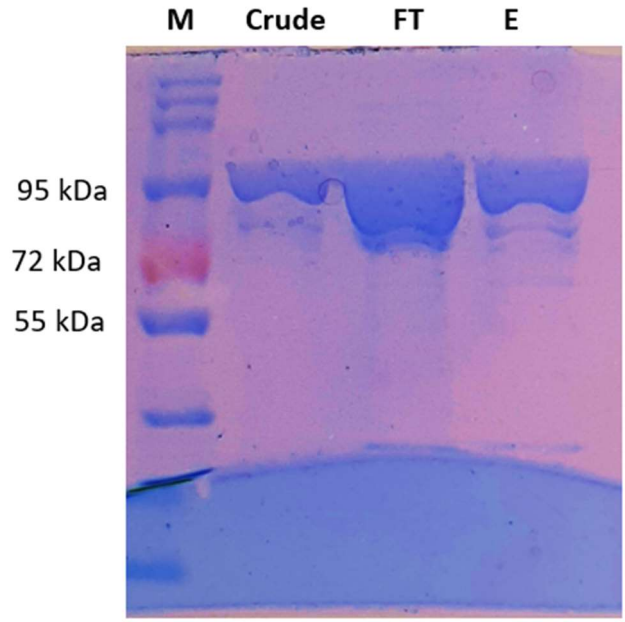


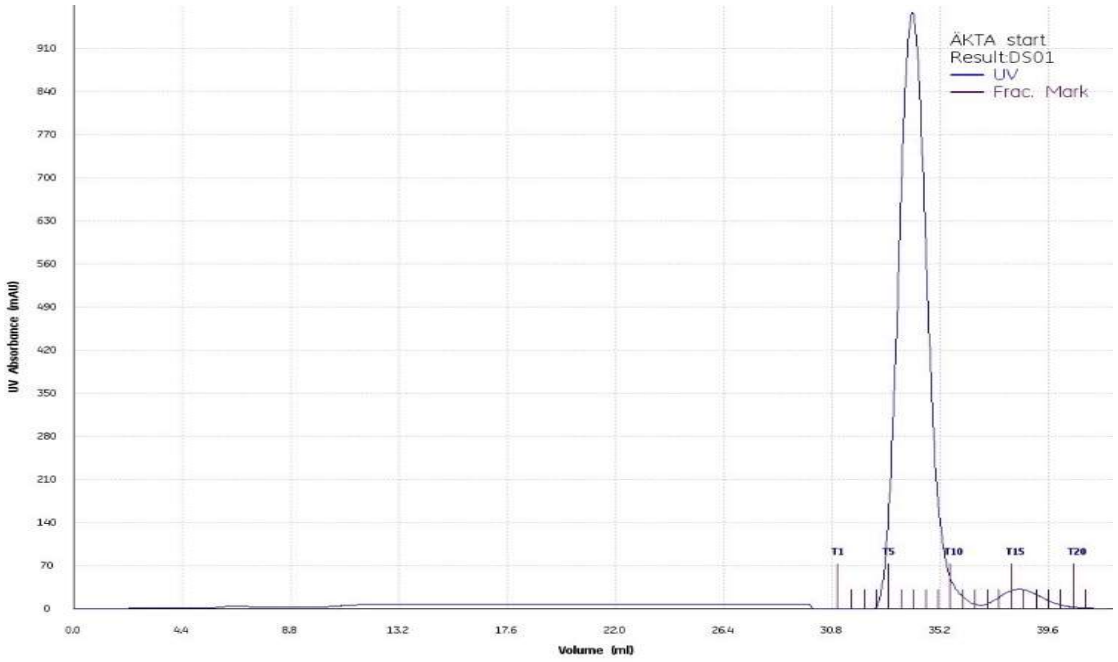
Figure 3.9 The resulting gel after 2 hours and multiple staining. FT represents the flow through (T6-12). Crude represents the crude sample after the first IMAC. E represents the eluent (T23-29).

The purity of the eluent (Figure 3.9) is higher than the crude (before second round of IMAC) and is much more concentrated. It also matches the theoretical molecular weight of 86 kDa, confirming the expression of the target protein. The eluent fraction was desalted prior to further purification.

3.3.1.2 Desalting

The eluent from section 3.3.1.1 were loaded onto the desalting column. Two runs were performed as the maximum loading size is 2 mL (Figure 3.10 a and b) [66].

(a)



(b)

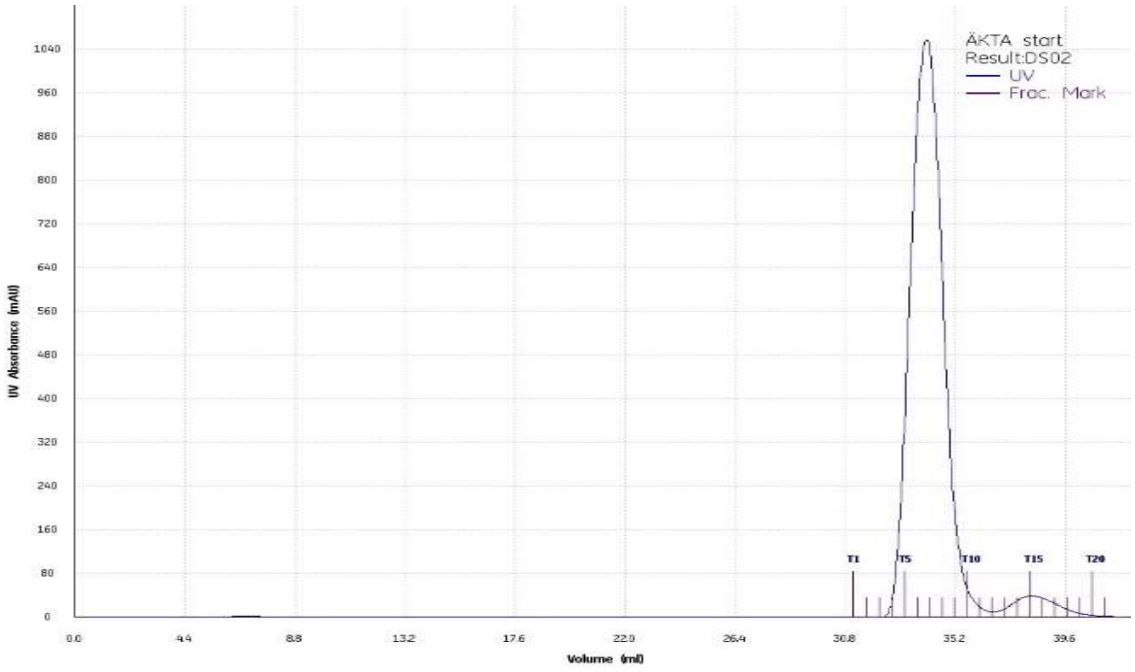


Figure 3.10 The resulting chromatogram after (a) the first attempt at desalting (b) the second attempt.

The peaks from both chromatograms correspond to the sample in tubes 4 to 9 (x-axis, Figure 3.10). Both fractions were combined and added to the gel. The purity was determined to be 91%, eliminating the need for ion-exchange chromatography[64, 67].

3.3.2 Protein Quantification

The total protein concentration of an uncharacterized sample can be inferred from a standard graph, such as in Figure 3.11.

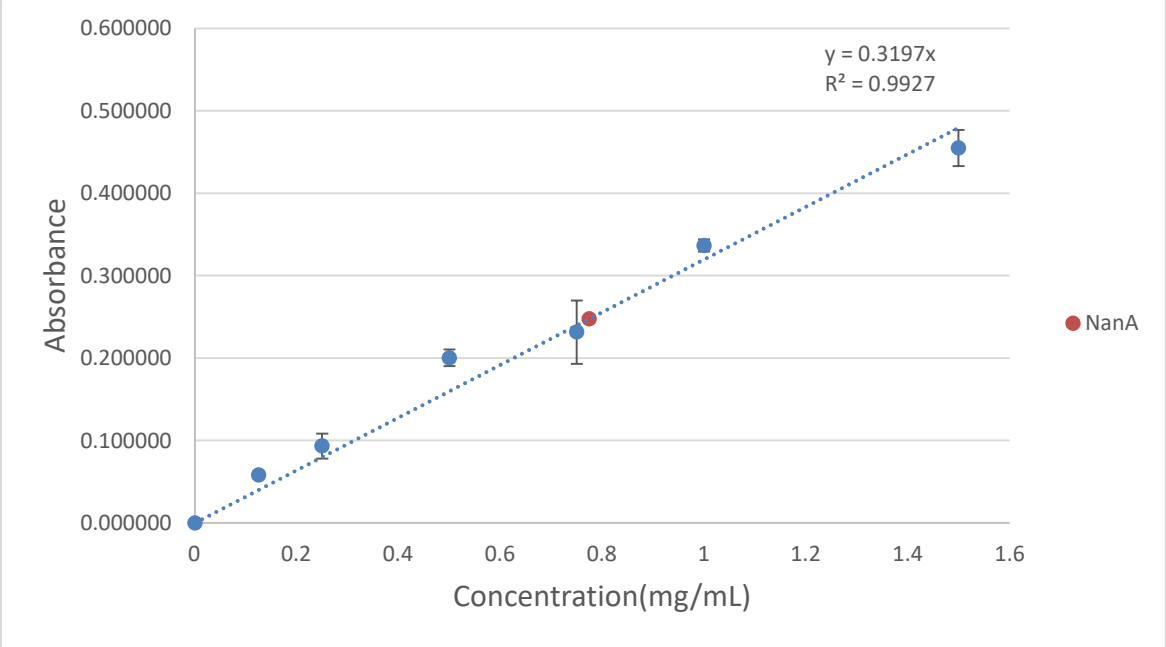


Figure 3.11 BSA standard to determine the unknown protein (NanA) concentration.

The theoretical molecular mass of NanA (sequence based) is 86200 Da or 86.2 kDa. The final concentration determined was 17.94 ± 0.35 µM. NanA was thus successfully produced and partly purified, permitting kinetic experiments to be conducted.

Chapter 4

Fluorescence Based Neuraminidase Assay

4.1 Theory Overview

4.1.1 Progress Curves and Determination of Reaction Velocity

An enzymatic reaction has three distinct phases: pre-steady-state, linear steady-state and the non-linear phase (Figure 4.1 left). The linearity of the progress curve is a prerequisite for the steady-state condition. The Michaelis-Menten equation is valid when [S]>>[E][78]. It is best to estimate the velocity using a tangent to the initial linear range (Figure 4.1 right)[79].

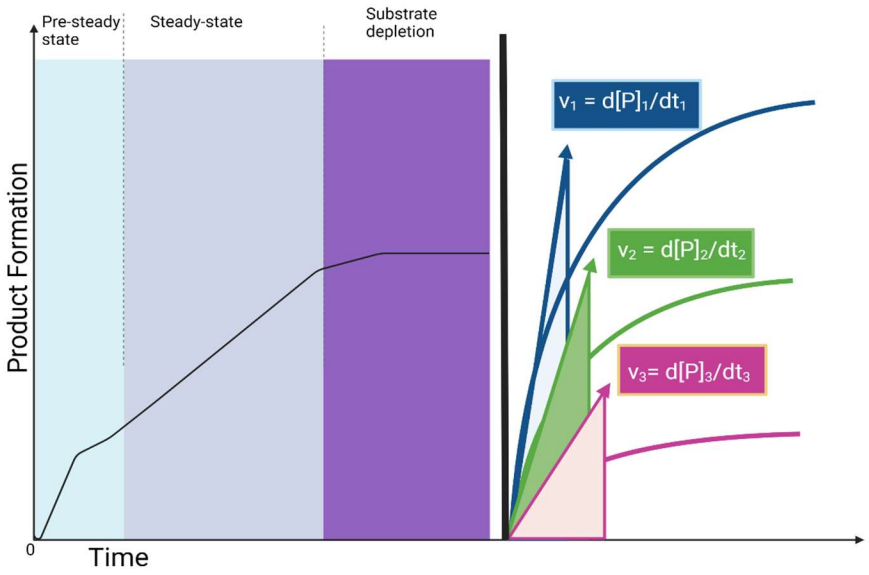


Figure 4.1 The three distinct phases in any enzymatic reactions are shown (left); as well as the progress curves of different substrate concentrations (right).

The tangent is the change in concentration (of either substrate or product) as a function of time and represents the actual rection velocity. The linear region will start to deviate and curve with depletion of the substrate. It is essential to detect this region to ensure that the steady-state condition holds[80].

The 'dead time' is one of the most prevalent challenges in determining the initial velocities. It occurs when there is a delay from the start of the reaction to the detection and is a common limitation in instruments. The initial velocity has passed if the progress curve is non-linear and is due to depletion of the substrate. The solution is to decrease the enzyme concentration to reduce substrate conversion during dead time. By decreasing the enzyme concentration, the linear range is extended for better observation of the steady state. Enzyme titration is used to determine the optimal concentration needed for an assay (Figure 4.2).

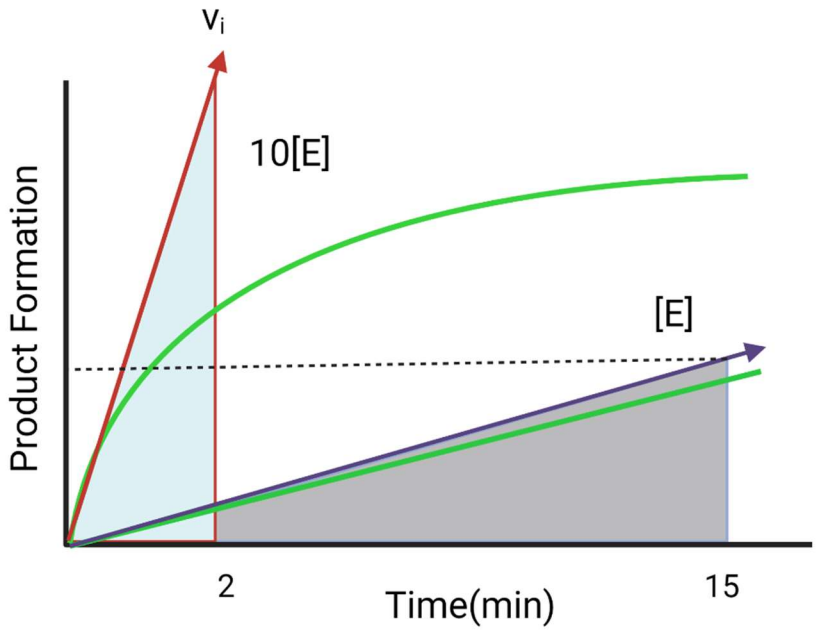


Figure 4.2 The effect of the enzyme concentration on the observable progress curve.

A dilute enzyme concentration results in a slower, extended linear range whereas the more concentrated (10[E]) results in a non-linear progress curve, violating the steady-state assumption.

Different types of enzymes, substrates and conditions all impact the time required to perform an assay. Short time periods are preferred over extended ones. However, some reactions can take hours or days due to low enzymatic activity. Optimizing conditions for continuous assays occur much faster as adjustments can be made in real time. The initial velocity can be determined directly from the progress curve generated[14, 80].

4.1.2 Artificial Substrate "MUANA"

Currently, neuraminidase inhibitors are the only class of antivirals approved to effectively treat influenza. It is necessary for health authorities to monitor the susceptibility of a pathogen to an inhibitor. Using various assays or methods of detection. Fluorescence is one of the most sensitive methods and is preferred over absorbance[11].

This chapter uses a florescence-based assay to assess the rate at which pneumococcal neuraminidase (NanA) cleaves 2'-(4-Methylumbelliferyl)-α-D-N-acetylneuraminic acid (MUANA) to release 4-methylumbelliferone (4-MU) and can be seen in Figure 4.3.

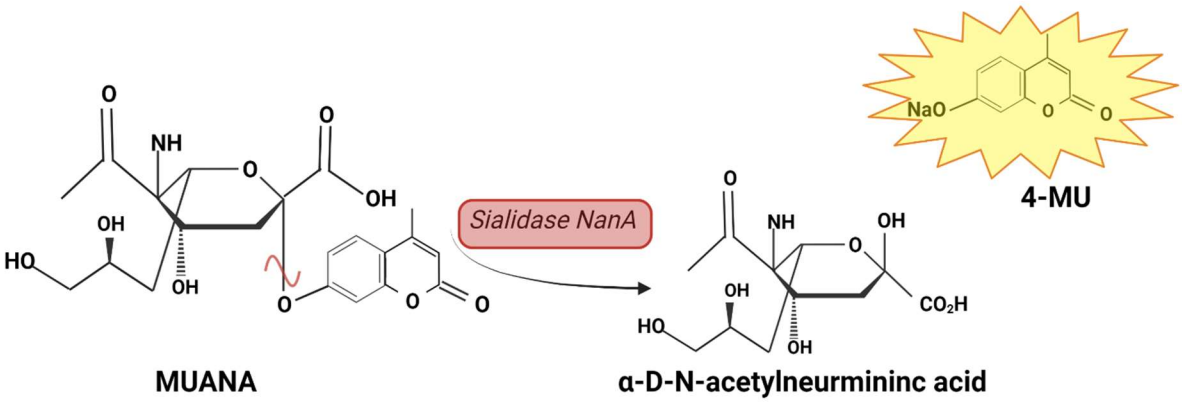


Figure 4.3 The enzymatic cleavage of 2'-(4-Methylumbelliferyl)-α-D-N-acetylneuraminic acid to release 4-methylumbelliferone.

The main limitation of fluorescence-based assays is that either the substrate or product must be a fluorophore. Mostly only heterocyclic, aromatic compounds fluoresce sufficiently well for detection. In the reaction in Figure 4.3 both substrate (MUANA) and product (4-MU) are fluorescable[81].

Annually, more than 2000 influenza virus strains are tested for antiviral susceptibility, implying this artificial substrate to be suitable for this type of assay and extending to high virus or bacterial titre. Fluorescence offers some advantages such as the linear correlation between the signal and concentration, increased specificity, and sensitivity[82]. It is also robust against environmental influences. The main disadvantage being that the output signal is relative to the machine and not universal[16].

All three pneumococcal neuraminidases function as monomers and fold after secretion from the cell. These enzymes function in an optimum pH between 5 and 6.5. From the three enzymes, NanA has the highest affinity for sialic acid. It forms a covalent sialyl enzyme intermediate, followed by hydrolysis via an oxocarbenium transition state[70]. Briefly, NanA cleaves the glycosidic bond of the substrate, releasing 4-MU. The aim of this chapter is to adapt the MUANA-based assay to assess the efficacy of inhibitors[83].

4.2 Methodology (Protocol)

4-MU is released when NanA cleaves MUANA. The approach involves two objectives:

1) The linear range of 4-MU must be determined.
2) To assess the neuraminidase activity of *S. pneumoniae* NanA in terms of kinetic parameters such as K_m, V_{max} and K_{cat}.

4.2.1 4-MU Standard

The concentration range of the 4-MU will be determined for which the fluorescence signal is proportional to the 4-MU concentration. The standard curve will allow for the conversion of the signal into concentration.

1. Assay buffer with 33.3 mM MES pH 6.5, 4mM $CaCl_2$ was sterile filtered.
2. Preparation of 4-MU stock solution

   10 mL of 5000 µM 4-Methylumbelliferone sodium salt (Sigma-Aldrich M1508) was prepared by dissolving 9.91 mg in 5mL absolute ethanol. 5 mL of 0.9 % NaCl (w/v) was added to the stock solution to make 10 mL. A 2500 µM working solution was prepared by adding 5 mL of the 5 mM 4-MU to 5 mL 1x assay buffer.

The following procedure was used to obtain the standard curve:

1. A range of 4-MU concentrations were prepared through a two-fold serial dilution of the 2500 µM 4-MU, all in 1x assay buffer, starting at 500 µM.
2. 30 µL of each of the 4-MU concentrations were added to a 384-well black microplate and 30 µL of the 1x assay buffer was added to each well for a total working volume of 60 µL. One row was only filled with 60 µL of 1x assay buffer to serve as blanks.
3. Each concentration of 4-MU was prepared in triplicate.
4. Samples were incubated in 37°C for 30 minutes and 3 readings were taken every 10 minutes to ensure product stability.
5. The plate was then removed, and the final readings were taken 5 times per well (every minute for 5 minutes) on a fluorometer with an excitation wavelength of 355 nm and emission wavelength of 460nm.
6. The average signal per concentration was calculated as relative fluorescent units (RFU).
7. The readings were then normalized by subtracting the average background RFU signal.
8. The plot of RFU against 4-MU concentration (µM) is used to determine the linear range and optimal target signal (Figure 4.4 a).

4.2.2 Substrate and Product Stability

It is essential for the substrate and product to be stable in solution over time as it ensures accurate kinetic data. A two-fold serial dilution of both substrate MUANA and product (4-MU) were made in 1x assay buffer. MUANA ranged from 0 to 500 µM and 4-MU from 0 to 6.25 µM, where each concentration was added in triplicate with one blank row per experiment.

4.2.3 Enzyme Titration

Selecting the appropriate enzyme concentration forms one of the three key components for assay design. It is vital for the steady-state condition as the initial velocity needs to be captured accurately. If the enzyme concentration is too high, the higher-end concentration substrates will saturate much quicker and vice versa[84].

The theorized K_m is approximately 50 µM. The lowest substrate concentration used is $\frac{1}{5}$ of the K_m and the highest 5-fold[84, 85]. Thus, 200 µM MUANA is the highest concentration used in the kinetic assay, so for the titration a 100 µM was chosen.

Two rounds of enzyme titration are performed, all substrates and enzymes are in 1x buffer solutions. The protocol established is as follow:

1. Prepare serial dilution of the enzyme in 1x assay buffer and include a blank (only buffer).
2. Distribute 30 µL of each enzyme concentration to the wells in triplicate.
3. Prepare 200 µM of MUANA in 1x assay buffer.
4. Initiate the reaction by adding 30µL of MUANA to each well and immediately read the plate using the optimized protocol for 30 minutes.
5. Convert the signal to 4-MU concentration using the standard curve from Figure 4.7.
6. From the linear range of each reaction progress curve, calculate the slopes (rates).
7. Select the concentration of enzyme that produces the best linear range.

4.2.4 Progress Curves

Progress curves are generated by the release of product as a function of time. From section 4.1.1, there are three phases in enzymatic reactions. The pre-steady-state is too short to detect in real time, thus the steady-state is most often observed until the substrate depletes (plateau region). It is more accurate to estimate the velocity directly from the progress curve by a tangent or slope as this represents the change in product concentration over time.

The convenience of the MUANA assay is due to the luxury of monitoring the product release in real time. The linear regions can be isolated from the progress curves. The signal can be converted into the release of 4-MU over time i.e., the rate. The initial velocity (rate) of each concentration of MUANA will be used to plot the Michaelis Menten graph. The protocol for the kinetic assay is as follow:

1. Prepare a serial dilution of MUANA: 500, 250, 125, 62, 31, 15, 7, 3 µM.
2. Add 30 µL of each concentration in triplicate.
3. Initiate the reaction by adding 30µL of 10 nM NanA to each well.
4. Shake plate for 3 seconds before readings take place and normalize the signal. This is to ensure that the substrate mixes adequately with the enzyme.
5. Plot the signal as a function of time.

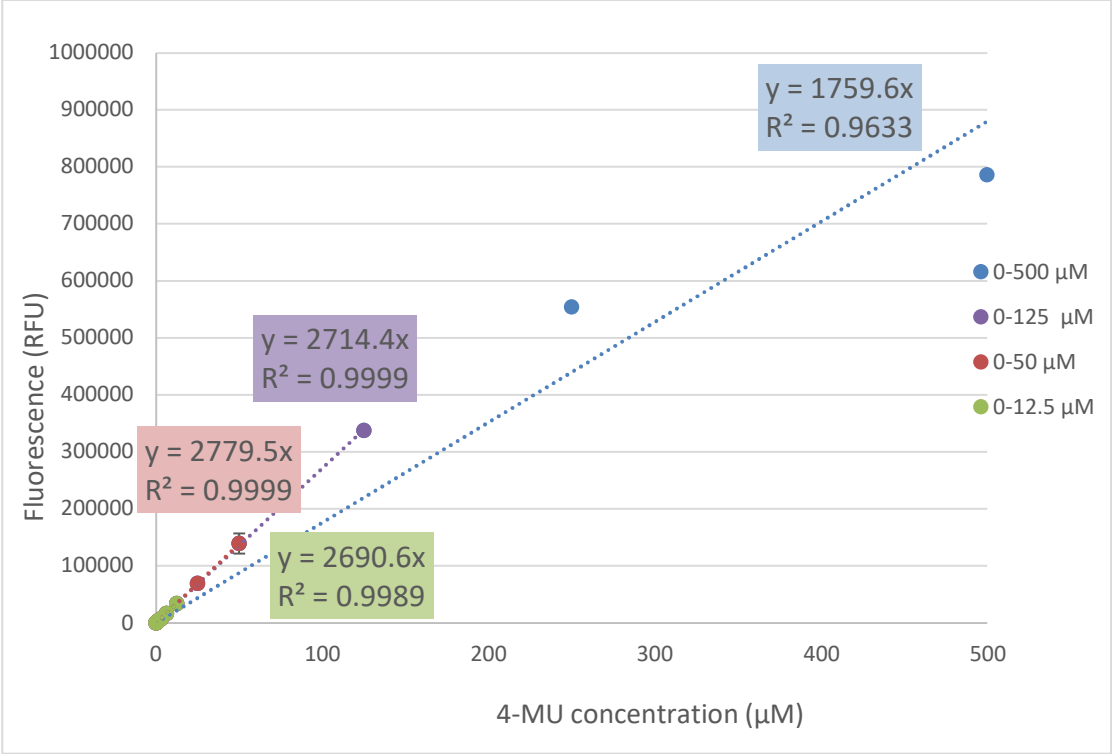
Convert the signal to 4-MU formation by using the standard 4-MU graph.

4.3 Results and Discussions

4.3.1 4-MU Standards

Following the protocol as in section 4.2.1, the resulting standard graphs are seen in Figure 4.4.

(a)



(b)

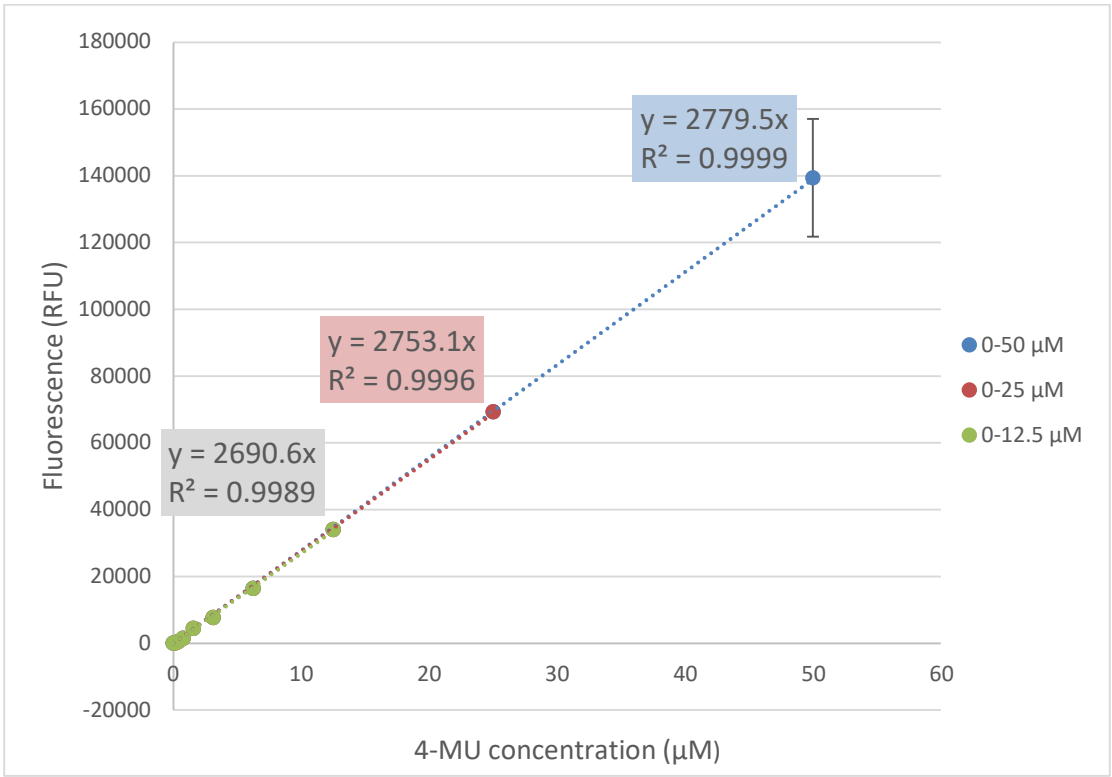
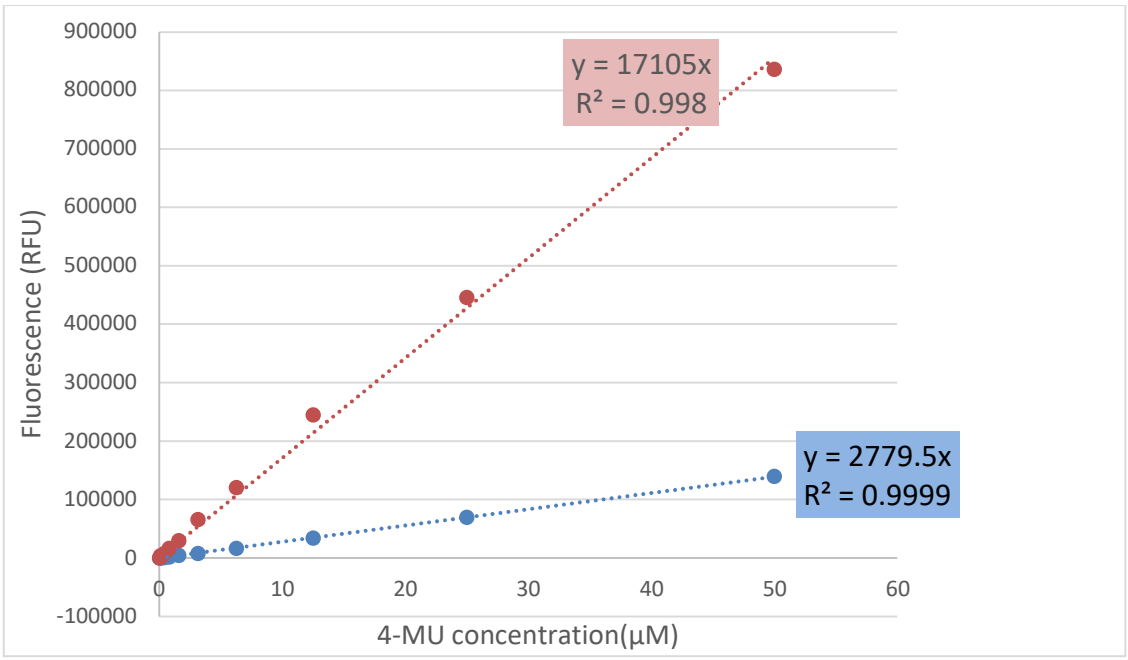


Figure 4.4 The standard curve of 4-methylumbelliferone in (a) the full concentration range; as well as (b) the lower concentration range.

The slope decreases as the concentration increases. The non-linear (exponential) region is indicative of signal saturation. Thus, only the linear region isolated, and the slope used to convert the signal into 4-MU. The linearity diminishes from 125 µM onwards (Figure 4.4 a). The theoretical K_m of NanA is around 50 µM and a maximum of 15% of the substrate gets converted in any enzyme reaction. Thus, further optimization of the 4-MU standard is needed to increase the sensitivity for the lower concentration regions (Figure 4.4 b).

The fluorometer was re-calibrated to optimize for the lower concentrations, resulting in a new standard curve (Figure 4.5 a). The new signal increased with a factor of 5 and at 50 µM is comparable to the pre-optimized one at 500 µM (around 800 000 RFU). Saturation of the instrument occurs when the threshold for signal detection is reached, making the higher concentrations indistinguishable. This is not of concern as the maximum conversion of the theoretical K_m (50 µM) will equate to 7.5 µM thus still in the linear region. The improvement of the sensitivity is evident by the 10-fold increase in the S/N ratio (Figure 4.5 b)

(a)



(b)

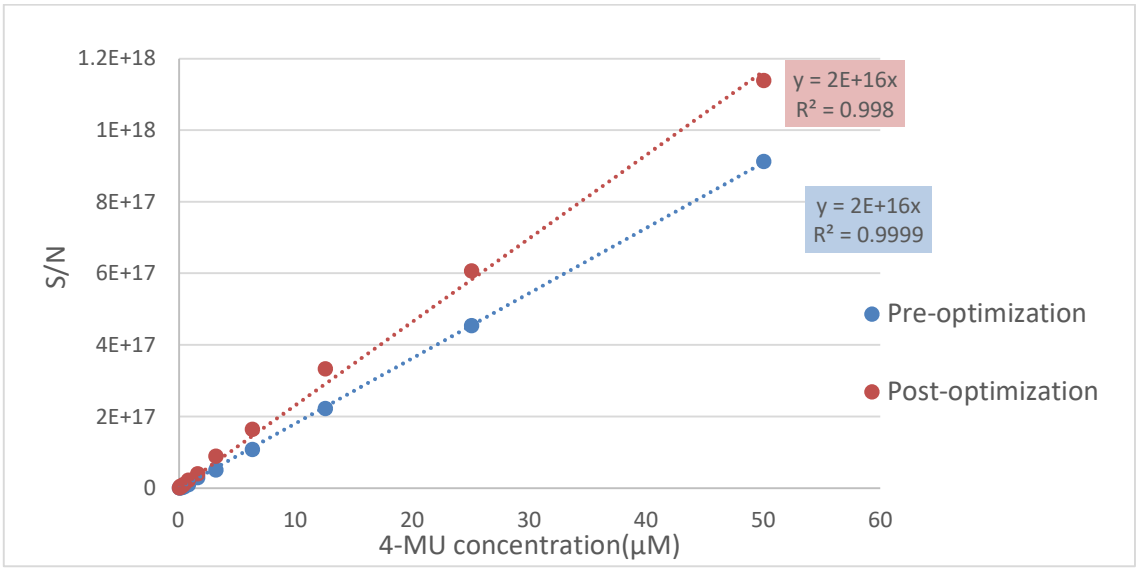
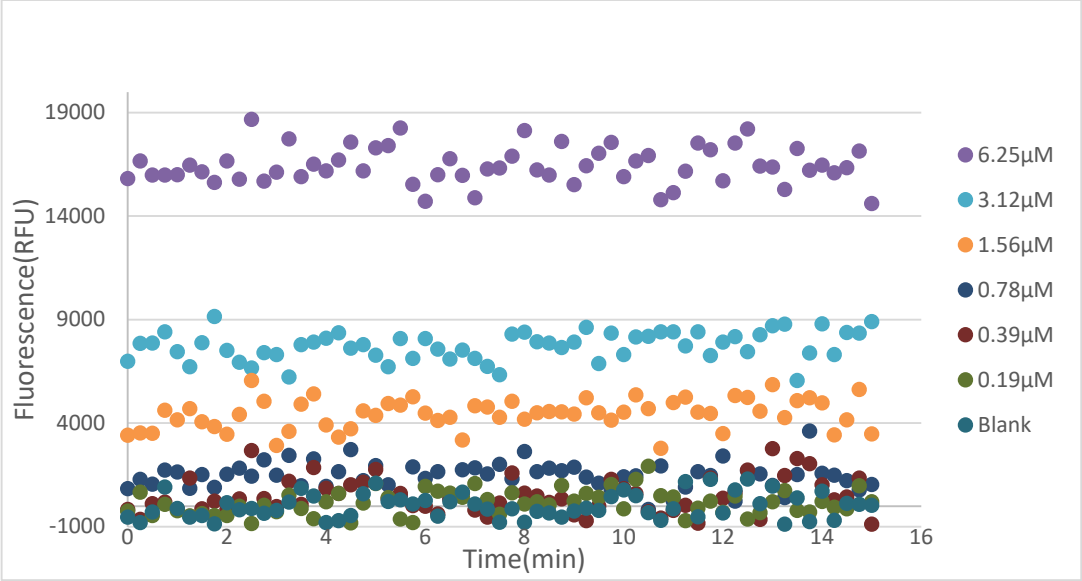


Figure 4.5 The standard curve (a) before and after optimization and (b) the S/N ratio before and after optimization.

It is essential for the substrate and product to be stable in solution over time as it ensures accurate kinetic data. The signal for 4-MU was monitored over the course of 15 minutes for both pre- and post-calibrated datasets. The aim of the experiment is twofold. The first being to eliminate spontaneous hydrolysis or degradation of the product, the second to demonstrate the increase in sensitivity (Figure 4.6 a and b).

(a)



(b)

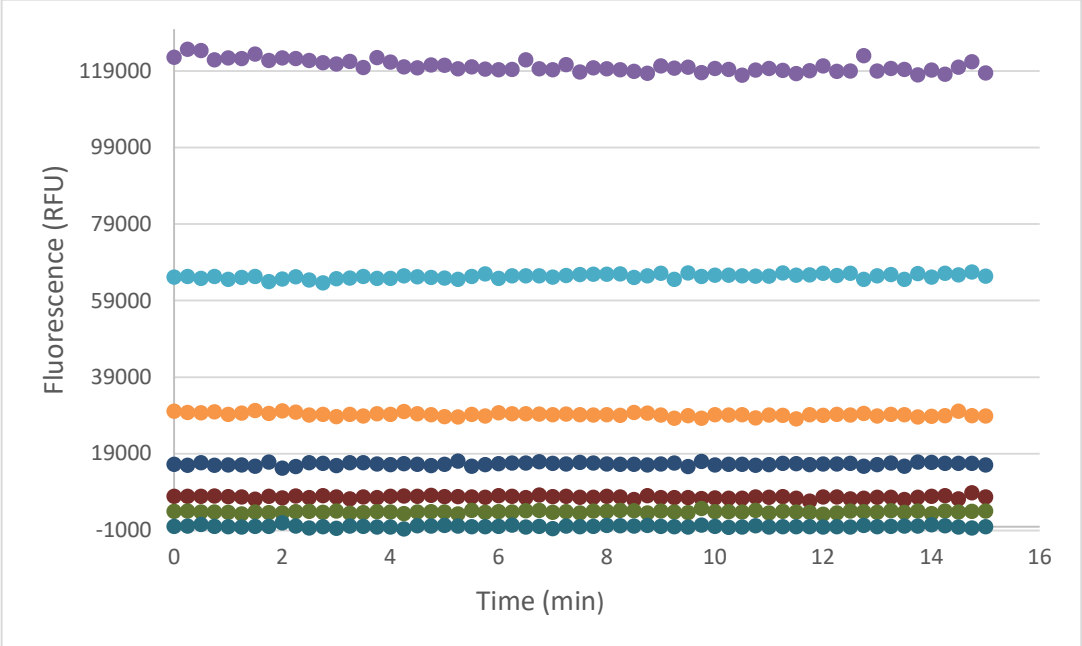


Figure 4.6 The 4-methylumbelliferone concentrations as a function of time for the (a) pre-optimized dataset and the (b) post-optimized dataset. The concentrations are shown on the right-hand side.

In the pre-optimized dataset (Figure 4.6 a), the separation of concentrations below 3.12 µM show an erratic pattern similar to the blank (background noise), whereas in the post-optimized dataset (Figure 4.6 b) they are clearly distinguishable. To further illustrate this, a Breusch-Pagan statistical test was performed on both datasets to determine homo- or heteroscedastic behaviour[86]. The

absolute variance (Equation 4.1) of each concentration was calculated (Table 4.1) and graphically illustrated (Figure 4.7).

$$S^2 = \frac{\sum(x_i - \bar{x})^2}{n-1} \quad [4.1]$$

Table 4.1 Results of the Breusch-Pagan statistical test as well as the absolute variance for each concentration range of 4-MU.

4-MU (μM)	Breusch-Pagan test		Variance	
	Pre-optimization	Post-optimization	Pre-optimization	Post-optimization
6.25	Homoscedastic	Homoscedastic	785040.2	400370.5
3.13	Homoscedastic	Homoscedastic	469798.9	331347.1
1.56	Homoscedastic	Homoscedastic	519114.2	217943.7
0.78	Homoscedastic	Homoscedastic	418938.7	127241.9
0.39	Homoscedastic	Homoscedastic	725874.6	111474.0
0.20	Homoscedastic	Homoscedastic	337059.8	69448.0
Blank	Homoscedastic	Homoscedastic	342114.2	59684.5

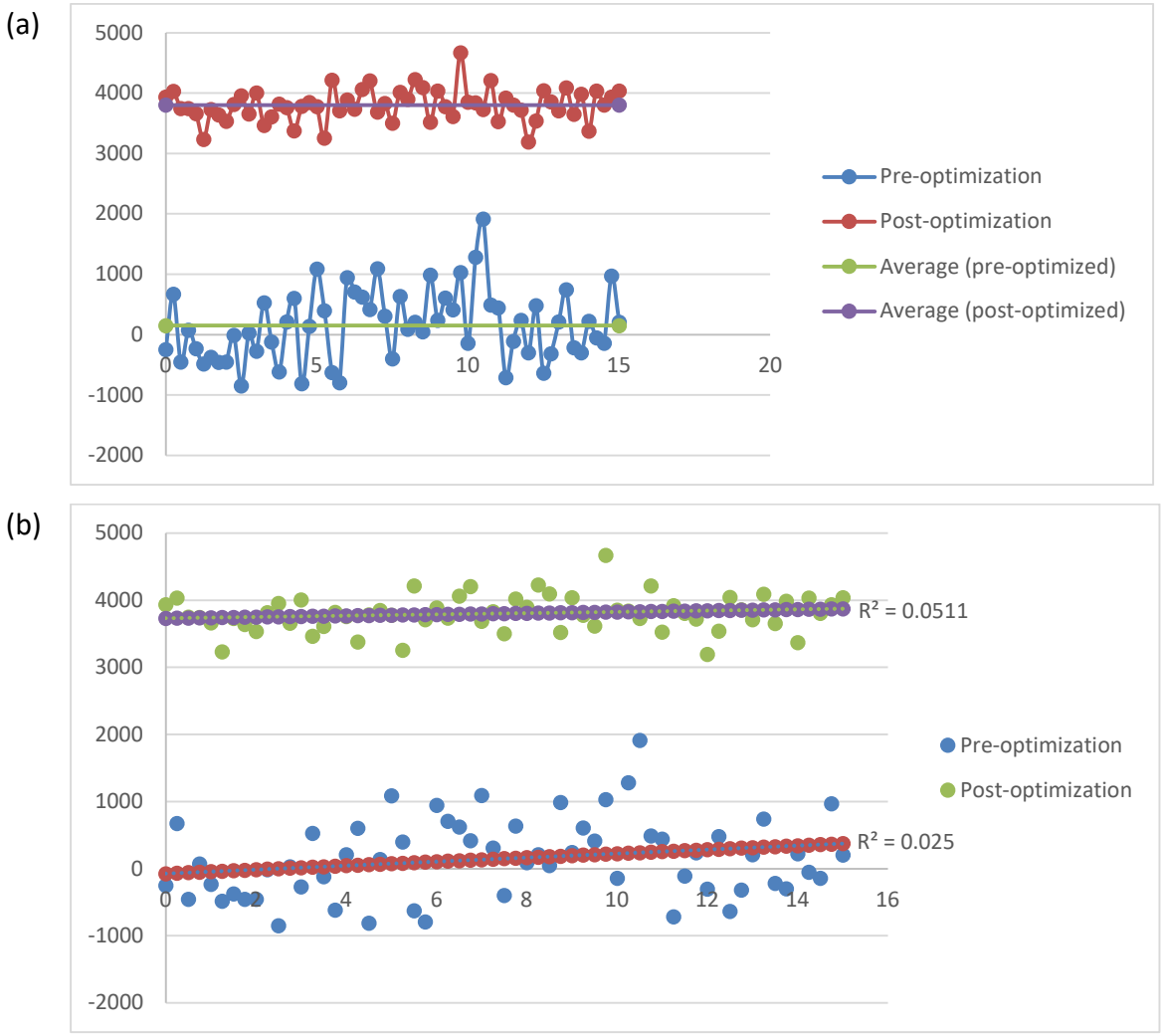


Figure 4.7 The signal comparison for 0.39 µM 4-methylumbelliferone (a) before and after optimization and (b) the Breusch-Pagan statistical test result.

Both pre- and post-optimized datasets are homoscedastic (Table 4.1), indicating that the residuals (concentration) are stable over time. It is further confirmed in the graphical Bruesch-Pagan statistical test (Figure 4.7 b) where the predicted trendlines for both datasets are shown in blue and green, respectively. The difference between the predicted and actual data points are referred to as squared residuals. These are used to determine the R^2 value using a linear regression model. Thus, it is evident for the post optimized dataset that the correlation is greater due to the increased R^2 value.

Variance is a statistical measurement of the spread between numbers in a data set i.e., how much each number deviates from the mean and each other. The variance decreased after optimization (Table 4.1), confirming the increase in the sensitivity. It is also evident that the data points reside closer to the average for the post-optimized dataset than for the pre-optimized (Figure 4.7 a); given that these data sets were compared for the 0.20 µM 4-MU standard, it shows excellent sensitivity as it is the lowest concentration other than the blank.

4.3.2 Substrate and Product Stability

A standard curve for MUANA was generated (Figure 4.8) using the protocol from section 4.2.2. and the signal converted to 4-MU (using Figure 4.5 a). The results are shown in Table 4.2.

Table 4.2 The average fluorescence of different substrate concentrations.

MUANA (μM)	Average Fluorescence (RFU)	Converted to 4-MU(μM)	
500.000	117042.78	6.84 ± 0.03	Detectable
250.000	66707.07	3.90 ± 0.03	
125.000	36493.17	2.13 ± 0.02	
62.500	18480.10	1.09 ± 0.02	
31.250	10926.38	0.63 ± 0.01	
15.625	3592.73	0.21 ± 0.01	
7.813	368.31	0.02 ± 0.01	
3.906	-898.09	-0.05 ± 0.01	Non-detectable
1.953	-1003.57	-0.06 ± 0.08	
0.977	-1974.82	-0.11 ± 0.01	
0.488	-2122.55	-0.12 ± 0.01	
0.244	-872.47	-0.05 ± 0.02	
0.122	-2339.28	-0.13 ± 0.01	
0.061	-2370.06	-0.13 ± 0.01	
0.031	-2332.53	-0.13 ± 0.01	
0.000	0.00	0.00 ± 0.02	

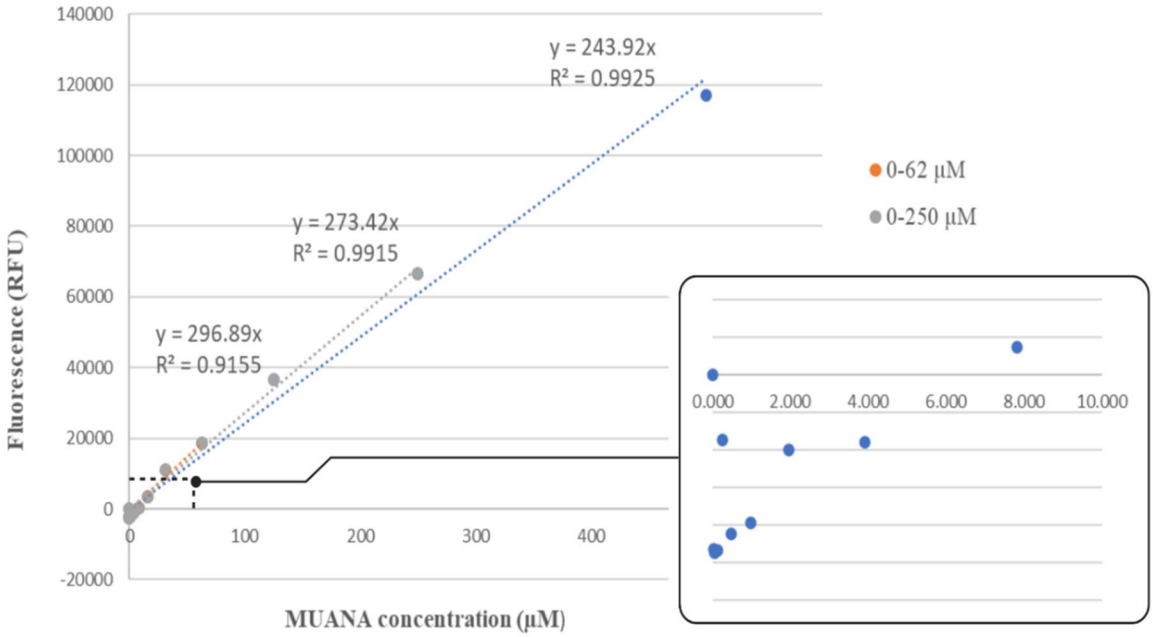


Figure 4.8 The standard graph of 2'-(4-Methylumbelliferyl)-α-D-N-acetylneuraminic acid.

As the concentration range is increased, the slope decreases (Figure 4.8). The R^2 values for concentrations exceeding 125 µM are all above 0.99, indicating excellent linearity. The lower concentration range seem to exhibit an erratic pattern, like background noise. It can be explained by considering the maximum conversion of 15% for any enzymatic reaction, thus 6 µM MUANA would correspond to a 4-MU signal for 0.9 µM. From Table 4.2 it is evident that any substrate concentration below 3.9 µM falls outside of the detectable linear region (LOD) [85]. To exclude substrate degradation, a serial dilution of MUANA was monitored a time course of 15 minutes (Figure 4.9). The substrate is stable over time as all concentrations show heteroscedasticity.

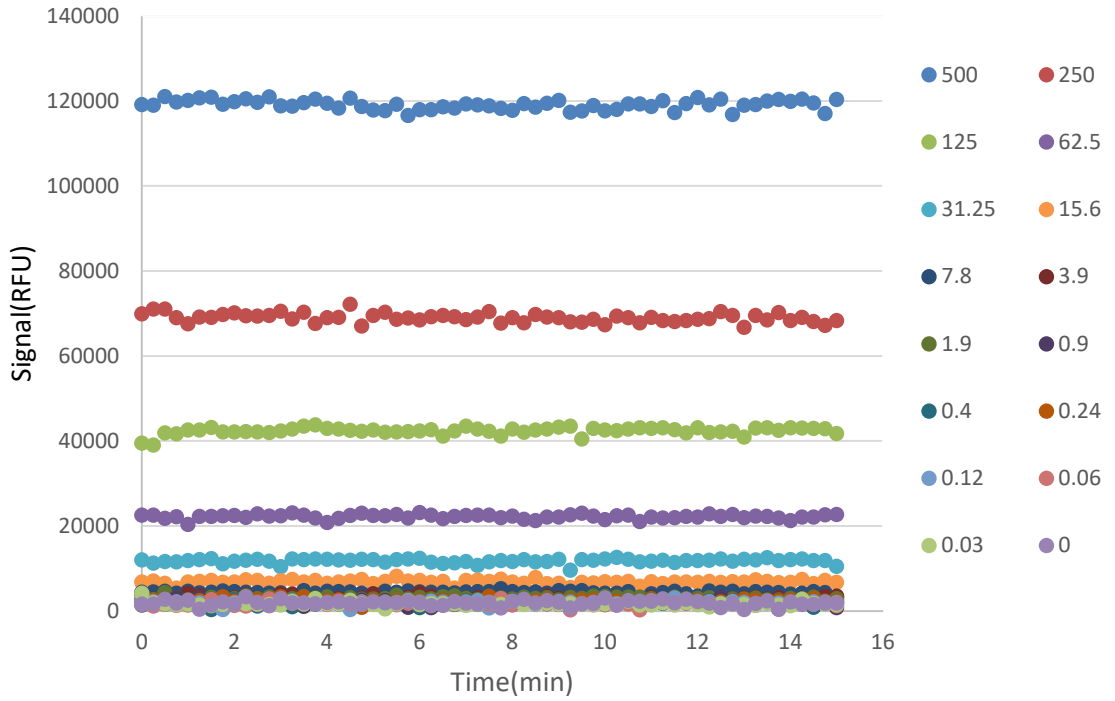


Figure 4.9 The 2'-(4-Methylumbelliferyl)-α-D-N-acetylneuraminic acid signal as a function of time. The concentrations are shown in µM.

The substrate is stable over time (Figure 4.9) as all concentrations show heteroscedasticity.

Thus, the product (4-MU) and substrate (MUANA) are stable over time. An optimized standard curve was generated for 4-MU to allow signal conversion. With all these factors in place, the kinetic parameters can now be obtained for NanA.

4.3.3 Enzyme Titration

Following the protocol from section 4.2.3, the first result after the first round of enzyme titration is shown in Figure 4.12.

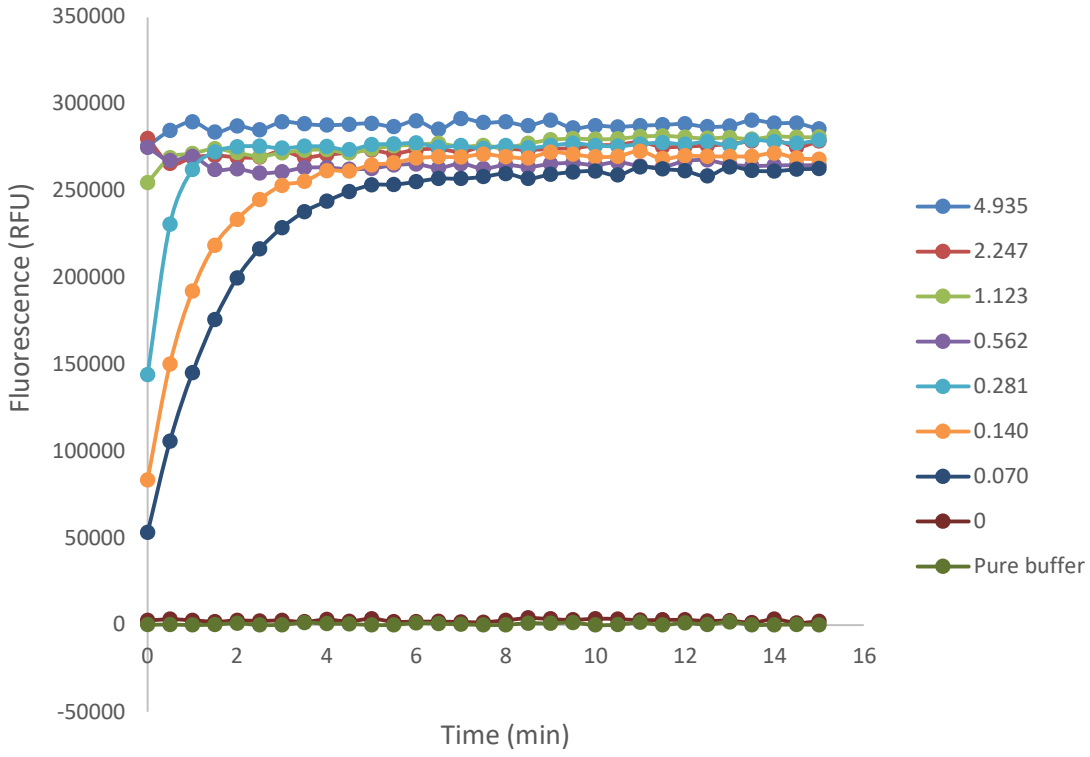


Figure 4.10 The resulting progress curves after the first round of enzyme titration. The concentrations are shown in µM.

Enzyme concentrations from 0.56 to 4.93 µM are plateaud,indicating that the reaction is complete and the signal saturated. In fact,any concentration exceeding 0.07 µM showed no linearity. Thus, a second round of titration was performed (Figure 4.11).

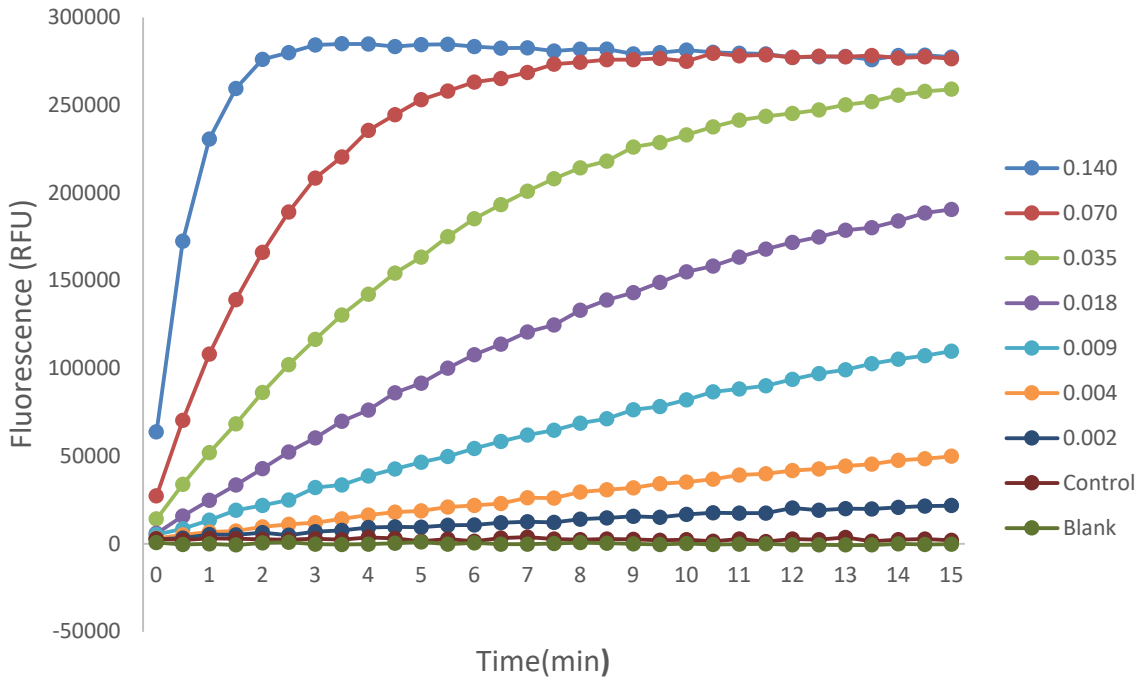


Figure 4.11 The resulting progress curves for the second round of enzyme titration. The concentrations are shown in µM.

None of the progress curves saturate before 2 minutes (Figure 4.11). It should be noted that lower substrate concentrations saturate much quicker, hence the 0.004 µM was chosen as the 1x enzyme concentration.

4.3.4 Progress Curves and Determination of Reaction Velocity

After the reaction has been initiated with the enzyme (NanA) as per protocol in section 4.2.4, the first set of progress curves are seen in Figure 4.12:

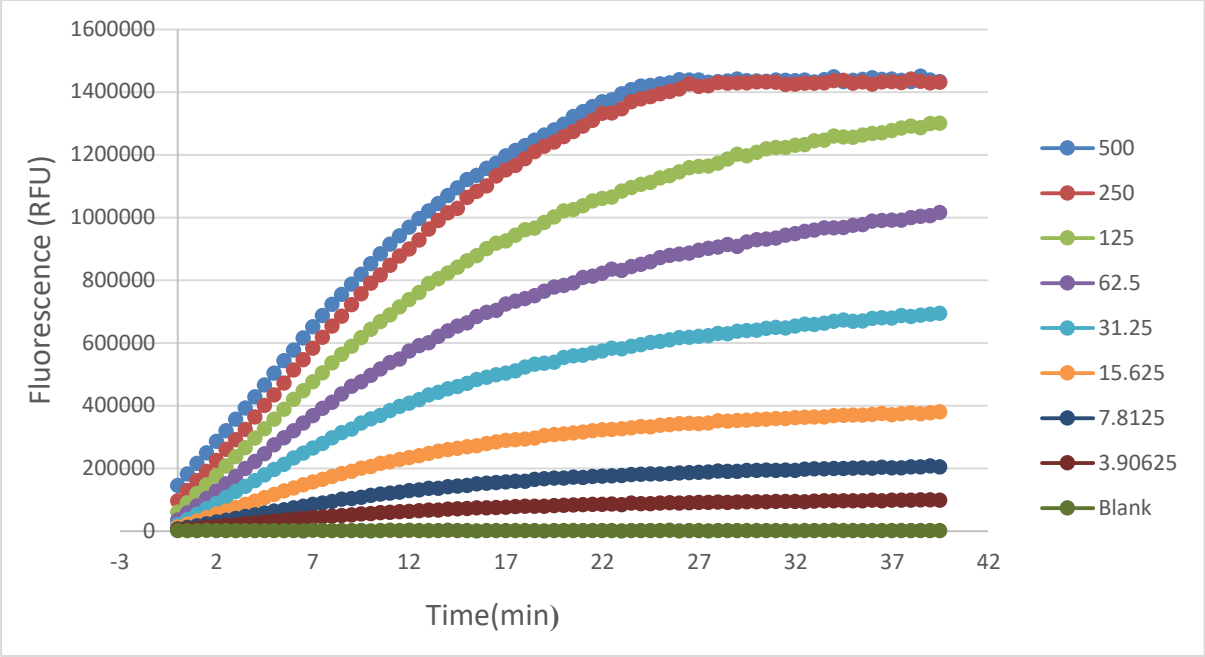


Figure 4.12 The overlaid progress curves showing the formation of the product as a function of time. The concentrations are shown on the right in µM.

The lower concentration ranges saturate much faster than the higher ranges. Linearity is a pre-requisite for the steady-state and was isolated for each concentration. The slope corresponds to the initial velocity or rate of the rection. An example of this process is demonstrated in Figure 4.13.

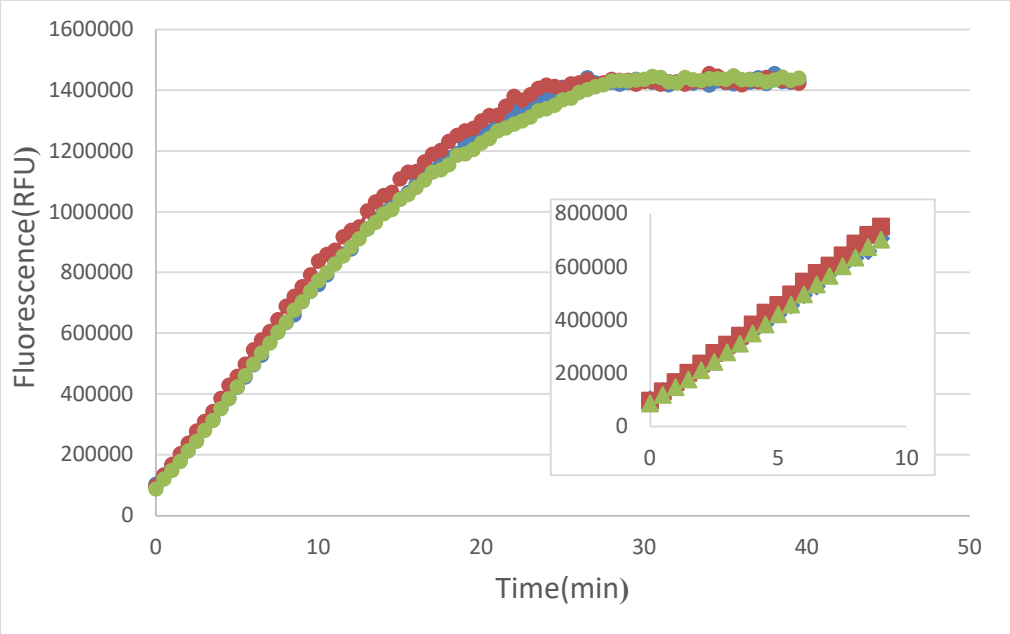


Figure 4.13 Three progress curves of 250 µM of the substrate is shown as experiment is performed in triplicate.

Figure 4.13 shows the progress curve for 250 µM MUANA. Each concentration was done in triplicate to account for the well-to-well variation and hence there are three different curves. The linear region for each curve was selected as demonstrated by the zoom-in on the right-hand side. Using Figure 4.5, the slopes can be converted to represent the release of 4-MU over time i.e., the reaction rate.

4.3.5 Michaelis Menten Graph

The Michaelis-Menten equation relies on the steady-state assumption as this is when the enzyme concentration is a constant. Thus, the reaction rate *v* is determined solely on the Michaelis complex, EA. The initial velocity of each substrate concentration is one data point in the Michaelis Menten graph (Table 4.3). After plotting all the initial velocities, the graph will resemble a hyperbolic function (Figure 4.14). All kinetic parameters were determined by nonlinear regression using GraphPad™.

Table 4.3 Initial velocities of each substrate concentration.

	Initial velocity (µM.min^{-1})				
MUANA (µM)	1	2	3	Average	Standard deviation
500.00	3.21	3.49	3.41	3.37	0.12
250.00	3.17	3.47	3.28	3.31	0.12
125.00	2.43	2.91	3.07	2.80	0.27
62.50	1.97	2.33	2.46	2.25	0.21
31.25	1.56	1.73	1.72	1.67	0.08
15.63	0.91	1.01	1.03	0.98	0.05
7.81	0.51	0.55	0.54	0.53	0.02
3.91	0.22	0.28	0.29	0.26	0.03

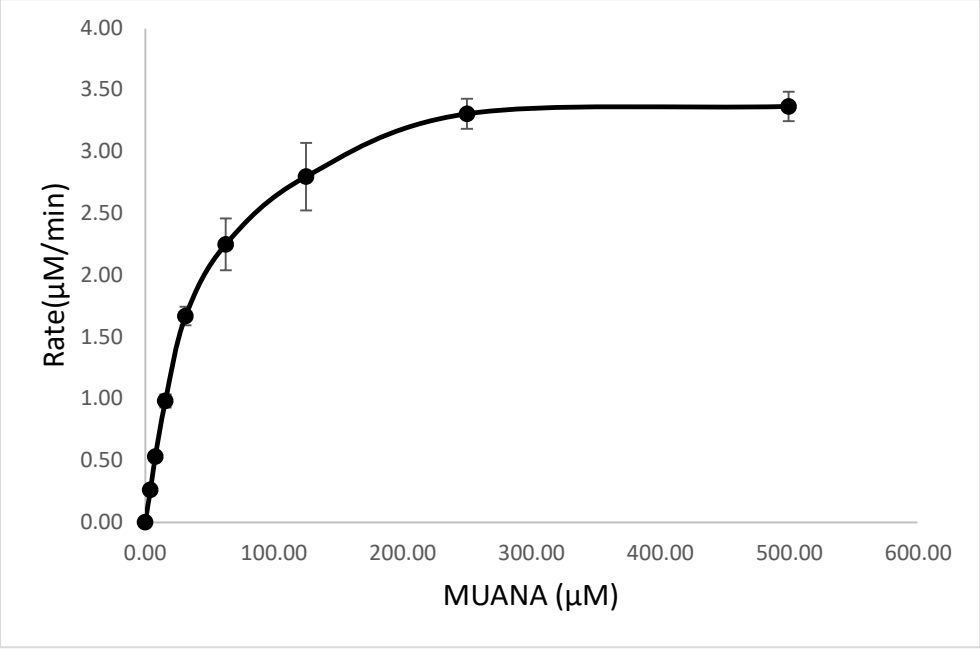


Figure 4.14 The resulting Michaelis Menten graph of NanA with the 2'-(4-Methylumbelliferyl)-α-D-N-acetylneuraminic acid as a substrate.

The Michaelis-Menten graph was successfully plotted for NanA with the MUANA substrate. The plateaued region indicates no change in the velocity above 200 µM, thus confirming the correct concentration range was chosen.

4.3.4 Kinetic Parameters

This section aims to conclude the application of all the protocols followed. In order to accurately obtain the kinetic profile of NanA, a detailed investigation is necessary. The two kinetic constants (K_m and V_{max}) can be obtained from the Michaelis-Menten graph. V_{max} is the maximum velocity of the enzyme, where Km is the substrate concentration at half V_{max}. The kinetic constants of the fluorometric assay are summarized in Table 4.4.

Table 4.4 Summary of kinetic parameters for NanA with 2'-(4-Methylumbelliferyl)-α-D-N-acetylneuraminic acid.

Kinetic constant	Value
K_m (µM)	41.56 ± 3.12
V_{max} (µM.min^{-1})	3.75 ± 0.87
k_{cat} (min^{-1})	851.19 ± 20.91

The Michaelis constant of 41.56 µM compares relatively well with the literature value of 45 µM for a pH of 6.5[81]. It is relatively low to the other two enzymes as NanB has a K_m of around 300 µM and NanC 900 µM .The K_{cat} for this experiment yielded 851 min^{-1}, whereas Klenow et al(2023)[27] reported K_{cat} values up to 12 000 min^{-1}. The discrepancy can be due to the enzymes differing in purity. The turnover rate also assumes that 100% of the enzyme is active, which is not the case. Thus, the turnover rate is relative to other enzymes and used for comparison.

This chapter fully investigated the MUANA assay as well as optimized the standard protocols and calibrations. The sensitivity of product detection (4-MU) improved and the Michaelis-Menten

graph obtained for NanA. The kinetic constants for this fluorometric based assay will be compared to an alternative assay in chapter 5.

Chapter 5

Assay Development for Natural Substrate

5.1 Theory Overview

5.1.1 End-Point Assay Principle

Throughout the previous chapter the benefits of using a continuous assay were highlighted. Alternatively, kinetic information can be extrapolated through a stopped (end point) assay. The enzymatic reaction is stopped after a specific time and the product formation is then quantified. It is assumed that the product formation is a function of time, thus representing velocity.

Product formation cannot be monitored in real time when using end point assays. The velocity could be underestimated should the time interval be too long (Figure 5.1). To mitigate this, appropriate time intervals should be selected and serial measuring points taken throughout the course of the reaction[14]. However, this is a tedious task and does not guarantee linearity. One can plot the time points taken during the reaction course against a control point to ensure reproducibility. However, this is only valid for the same set of conditions[14, 80].

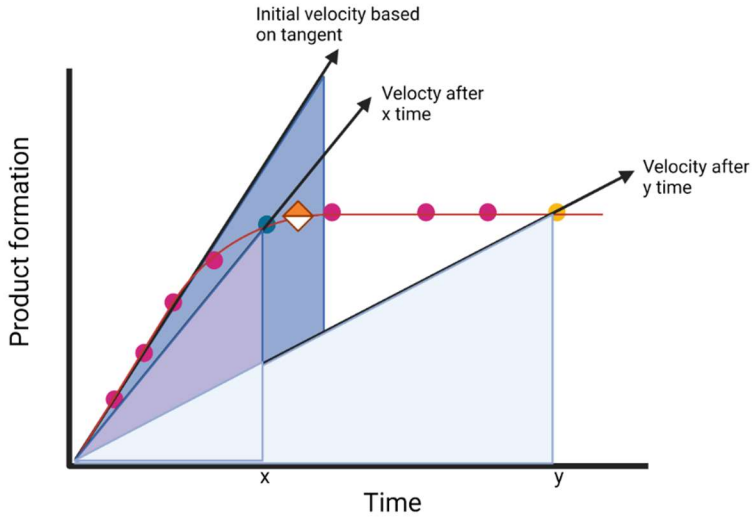


Figure 5.1 Time points taken during a stop (endpoint) assay to form a curve.

Alternatively, an integrated Michaelis Menten equation can be used (Equation 5.1):

$$v = \frac{-d[A]}{dt} = \frac{V[A]}{K_m + [A]} \qquad [5.\ 1]$$

Integrating with respect to time yields Equation 5.2:

$$\mathrm{K_m ln}\frac{[\mathrm{A}]_0}{[\mathrm{A}]} + [\mathrm{A}]_0 - [\mathrm{A}] = \mathrm{Vt} \qquad [5.\ 2]$$

The entire progress curve is represented, disregarding the need to isolate the linear region for the steady-state condition. The linear representation is obtained by rearranging Equation 5.2 into 5.3:

$$\frac{[\mathrm{A}]_0 - [\mathrm{A}]}{\mathrm{t}} = \mathrm{V} - \mathrm{K_m}\left(\frac{\ln\frac{[\mathrm{A}]_0}{[\mathrm{A}]}}{\mathrm{t}}\right) \qquad [5.\ 3]$$

V_{max} and K_m can now be obtained from the intercept and slope, respectively. Although more convenient, the method is only valid and reliable if the progress curve obeys the Michaelis

Menten equation. However, most of the time the product formation inhibits the enzyme and favors the backward reaction. Thus, to ensure accuracy it is encouraged to take time points over the course of the reaction[74].

5.1.2 Natural Substrate

An enzyme's main objective is to render feasible biological reactions and its activity is characterized by product formation. The previous chapter used fluorescence as a detection method[14]. *S. pneumoniae* encodes three neuraminidases that cleave terminal neuraminic acid (sialic acid) from host cell carbohydrate moieties to facilitate colonization of the pathogen[4]. NanA acts as an exosialidase, hydrolyzing sialyllactose into neuraminic acid (α-Neu5Ac)[87]. Substrate specificity assays revealed NanA can remove sialic acid from 3'-sialyllactose or 6-sialyllactose as shown in Figure 5.2.

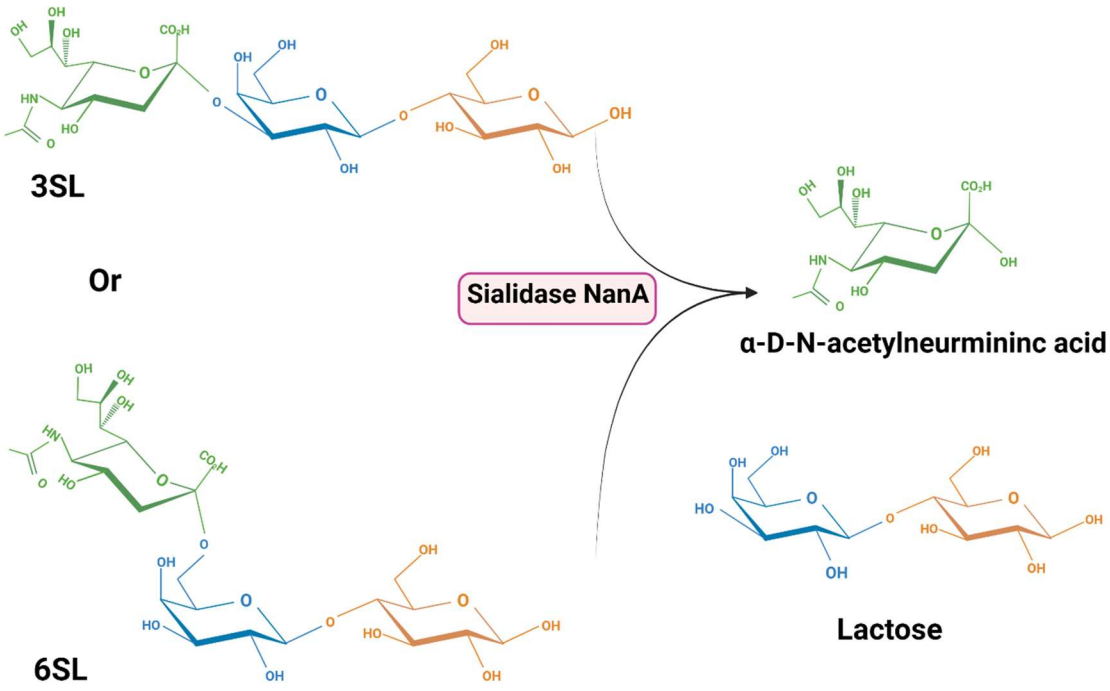


Figure 5.2 Sialidase NanA acts as exosialidase, hydrolysing 3'-sialyllactose or 6'-sialyllactose into lactose and α-Neu5Ac.

NanA cleaves the glycosidic bond to form sialic acid and lactose (glucose and galactose). The difference between the two substrates (3'-sialyllactose and 6'-sialyllactose) is the glycosidic linkages that reside on position 3 and 6 respectively. Unfortunately, sialyloligosaccharides do not have natural fluorescence. There have been multiple methods developed to modify natural substrates to be detectable via fluorescence. The main concern of these methods involves derivatizing or "tagging" the substrate with a fluorescent additive[87, 88].

Another prerequisite for the detection is a non-overlapping signal and is of concern when analyzing enzymatic reactions with artificial MUANA substrate as it was proven that it overlaps significantly with its respective product 4-MU. Neither 3'-sialyllactose nor 6'-sialyllactose have intrinsic fluorescent properties. Alternatively, biological substances show absorbance in at least one region. However, in this case the substrate and product have similar absorption features. Both the intact substrate (3'-sialyllactose or 6'-sialyllactose) and sialic acid (Neu5Ac) show absorbance in the 295 nm region, and signals are too weak to distinguish or fall outside of most commercial detectors' range. This means that some method must be developed to make one of the products detectable[14, 16].

5.1.3 Thiobarbituric Acid (TBA) Assay

Warren *et al* (1959)[89] was the first to publish an assay development for the detection of sialic acids. This includes naturally occurring substituted neuraminic acid and derivatives found in the human body. The unsubstituted acid is referred to as N-acetyl neuraminic acid (Neu5Ac). Many attempts have been made to detect and quantify sialic acid with the use of orcinol, resorcinol, acetic acid, or strong hydrochloric acid quenching[87, 88]. All these methods lack sensitivity; specificity and cannot be used in samples with unpurified matter. Warren *et al* performed the first quantification of a sample containing sialic acid by chemically altering the product. Absorption was used for detection of the new derivatized product, and the amount could be quantified using the formula in Equation 5.4:

$$\mu\text{Moles of neurmainic acid} = \frac{V \times O.D_{549}}{57} \quad [5.4]$$

Where *V* refers to the final volume of the test or extraction volume.

It was also discovered that there were possible interfering substances, such as deoxyribose, that will need to be accounted for by using Equation 5.5:

$$\mu\text{Moles of neurmainic acid} = 0.138\ O.D_{549} - 0.009O.D_{532} \quad [5.5]$$

Aminoff (1961)[90] published an improved version of the assay by Warren *et al*. One must consider that in both these assays, the main objective was to quantify sialic acid in a tissue sample or dissolved in water[91, 92]. The key principles of both methods as well as the integration into an enzymatic assay will be discussed in the following sections.

5.1.4 Mechanism of Reaction

In all variations of the TBA assay, three distinct phases are present: oxidation, reduction and then chromophore formation. A schematic representation of the process is portrayed in Figure 5.3.

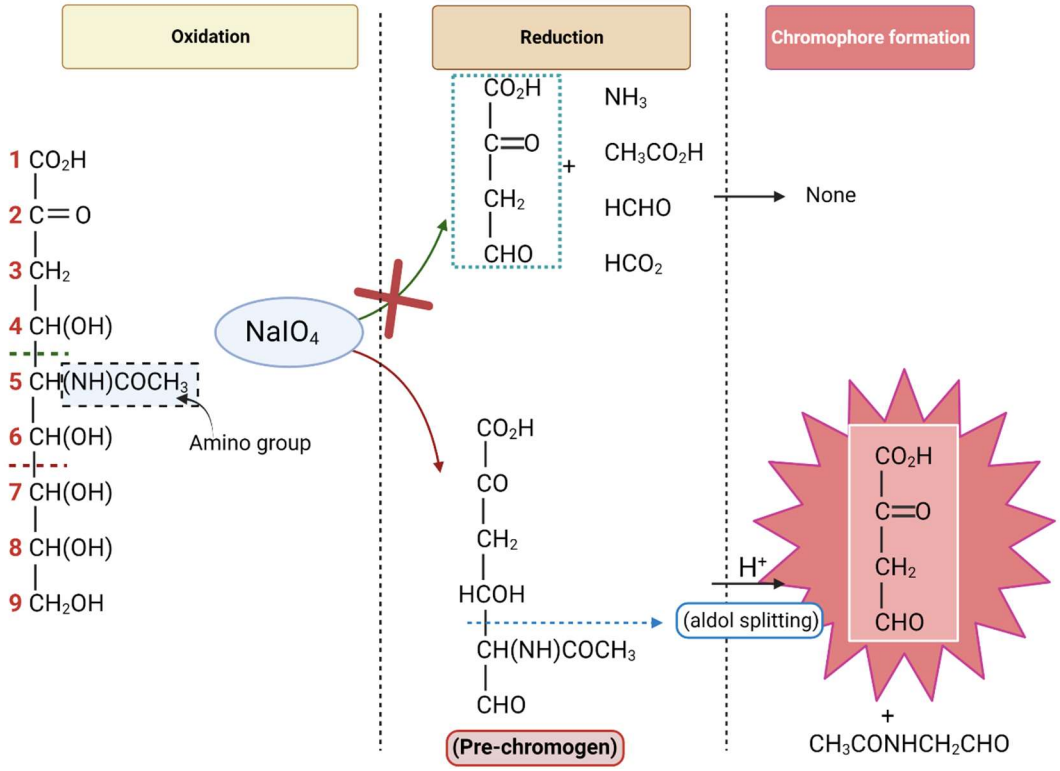


Figure 5.3 The two proposed mechanisms of the Thiobarbituric acid assay.

Warren *et al* (1959)[89] used 0.2 M sodium periodate in 9M phosphoric acid for the oxidation step. They claimed that a strong acidic environment is needed for maximum color development (molecular extinction coefficient) later in the process for the chromophore formation. It is also hypothesized to be required to remove the glycolyl or acetyl group (substituting group) from the amino group prior to oxidation (indicated in Figure 5.3). Color intensity is also directly proportional to the amount of periodate used, requiring 20 to 40 µmoles per sample. The mechanism of this reaction was still under investigation at the time of publication.

Aminoff (1961)[90] investigated the detection of free sialic acids from sialomucoids via mild acid hydrolyses. Questions about the method developed by Warren *et al* were raised by highlighting the use of strong acidic conditions in the oxidation step. This infers that the bound sialic acid (sialosides) would be cleaved without the aid of any hydrolysis or enzyme. This implied that no distinction can be made between sialosides and free sialic acid, which is essential as enzymatic release of the product needs to be monitored.

To increase the specificity for free sialic acid Aminoff used a lower concentration of 0.025 M periodic acid in 0.125M sulphuric acid (H_2SO_4) with a pH of 1.2. They also reported that the sialosides used did not react in the assay. In conclusion, the high acidity of the periodate is not essential for color development. The pH can also be adjusted to distinguish between sialic acid derivatives e.g. acid labile compounds will have less color development at a higher pH in comparison to free sialic acid. This leads to the first hypothesis of the chromophore formation and the role of oxidation.

Glycosides can be hydrolyzed to release sialic acid. As mentioned, sialosides do not react in this assay. This implies that during oxidation, the glycosidic bond is not broken. What is also interesting is that the acylation of the amino group (Figure 5.3, indicated group) does not protect against oxidation and in the pyranose ring it is rather the oxidation of the C_3-C_4 bond (and not C_4-C_5) that results in an activated methylene group on C_5. This is impossible for sialic acids, hence the formation of β-formylpyruvate cannot be explained by the proposed mechanism as indicated in Figure 5.3.

If a cleavage of the C_4-C_5 bond was to occur, the resulting product would be a ketal form of the β-formylpyruvate. This would be readily hydrolysed in acidic conditions into β-formylpyruvic acid. Methoxyneuraminic acid does not have an acylated amino group and should theoretically react to give the same product. Studies showed it consumed 3 moles of the periodate/mole and released 70% of the calculated ammonia. Given these postulates, it can be theorised that both methoxyneuraminic acid and C_7-methoxyneuraminc acid (C_7-methoxyNeu5Ac) should react in the TBA assay. This was confirmed as maximal color development for C_7-methoxyNeu5Ac occurred at 1 mol consumption. In contrast to this, neither of the above-mentioned compounds reacted in the Warren or Aminoff assays. This is due to the extreme acidic conditions which will protonate the amino group, preventing oxidation. This was confirmed as no color developed nor periodate consumed during the reaction. MethoxyNeu5Ac showed a very slow rate of periodate consumption and this consumption decreased as the pH did.

From the above information, it can be deduced that glycosylation alone does not inhibit oxidation but also the acylation of the amino group. This is important as the assay relies on the selectivity

of the free sialic acid, therefore from a mechanistic point of view the sialosides are viewed as inactive.

Assuming the acylation of the amino group contributes to the protection against oxidation, the next problem is specificity towards free sialic acid. It was reported that 3 moles of periodate are required for every mol of substrate to reach a maximum coefficient. Thus, β -formyl pyruvate is not a direct product of the oxidation but rather a pre-chromogen intermediate is formed. This unstable compound is then converted into the chromophore via aldol splitting, resulting in a pink-colored compound that is detectable via absorbance (Figure 5.3)[89, 93].

5.2 Methodology for TBA Assay

The main objective of this assay is to detect and quantify Neu5Ac, the product of the cleavage of sialyllactose. Three key stages of the assay have been identified namely: oxidation, reduction, and chromophore formation (extraction) as demonstrated in Figure 5.3.

It is thus essential to choose the most suitable reagent for each stage. The main concerns of the assay is to detect only free sialic acid and to obtain kinetic parameters even though this procedure was only used for quantification in tissues.

This colorimetric assay is used to detect the chromogen formation of Neu5Ac. When adding the color solution 2-TBA, the pre-chromogen will complex with the acidified solution to form a bright pink product. This chromogen-complex has an absorbance at 549 nm. Although not as sensitive as fluorescence, it can still be used if no overlapping wavelengths occur.

5.2.1 Reagents

Most commercial kits for any TBA assay are used for lipid oxidation. The contents of these kits are unknown. The only commercially available set specifically excludes *S. pneumonaie* as an enzyme.

The first obstacle is to choose a suitable oxidant. Based on Warren *et al*, 200 mM sodium periodate in 9M phosphoric acid was used. The results yielded an orange precipitation. The

extraction was carried out using 2% TBA and a mild acidic reducing agent. Different combinations of reagents were used, as can be seen by reaction 1 to 6 in Table 5.1. The conclusion of the first three reactions is the concentration of the TBA solution as well as the optimal ratios of reagents used as confirmed by the faint color development of reaction 3. One then wanted to establish if DMSO really helped facilitate chromogen development. Reaction 4 and 5 were performed with the optimal ratios, while 6 and 7 were performed with the previous incorrect ratio. The only other variable was the use of DMSO, which was included to possibly facilitate color development. Only reaction 4 and 5 was successful, with 5 showing more intense color. Hence it can be concluded that DMSO helps facilitate color development.

Table 5.1 Summary of different reagents used for fluorescence TBA assay.

O1 = 0.2 M sodium periodate in 9M phosphoric acid

R1 = 2% (w/v) sodium arsenite in 0.5 M HCl

T1 = 1% TBA w/v (0.4g in 40mL), 100mM sodium acetate buffer, pH 9

Ratios (volume in µM) = sample mix: oxidant: reductant: TBA solution: (extracting agent)

	Sample (µM Neu5Ac)	Stop solution	Oxidant	Reductant	2-TBA	Extracting agent	Chromophore	Ratios
1	2000	0.25M HCL	O1	R1	T2	None	Precipitate	2:1:3:1
2	2000	None	O1	R1	T2	None	Precipitate	2:1:3:1
3	2000	None	O1	R1	T2	None	Precipitate	1:1:1:10
4	2000	0.25M HCl	O1	R1	T2	None	Faint pink	2:1:1:10

5	2000	0.25M HCl	O1	R1	T2	DMSO	Pink	1:2:1:1:10
6	2000	0.25M HCl	O1	R1	T2	None	Precipitate	1:2:1
7	2000	0.25M HCl	O1	R1	T2	DMSO	Precipitate	1:1:2:1:3

The next objective was to establish if the product Neu5Ac could be monitored and quantified in a sample with an enzyme and substrate. To ensure that we are only monitoring the product, one must control for every sample. Another concern is the concentration of enzyme used, if too low the reaction will not proceed until saturation, if too high the steady-state region is missed. Table 5.2 summarizes the control used to ensure that only the product formed because of enzyme activity is monitored.

Table 5.2 Experimental scheme to ensure enzymatic activity in the fluorescent TBA assay.

Reaction	Sample	Stop solution	Oxidant	Reductant	2-TBA	Extracting agent	Chromophore
1	Neu5Ac	0.25M HCl	O1	R1	T2	DMSO	Pink
2	NanA + 3'-sialyllactose	0.25M HCl	O1	R1	T2	DMSO	Pink
3	Deactivated NanA + 3'-sialyllactose	0.25M HCl	O1	R1	T2	DMSO	No colour
4	3'-sialyllactose	0.25M HCl	O1	R1	T2	DMSO	Very faint colour

Reaction 3 (Table 5.2) demonstrates that no product is produced by an inactive enzyme, while reaction 4 shows a very subtle color development. This can infer to spontaneous hydrolysis of the substrate and must be accounted for. What is important is reaction 1 and 2 yields a pink chromogen, indicating the detection of the Neu5Ac product.

The groundwork done narrows down the search for the ideal reagents. One of these 'elimination' processes indicated that the oxidant used in Aminoff is ideal, which further justifies the use of a milder acidic solution.

5.2.2 Detection: Absorbance

The color solution using 2-TBA turns pink instead of red (as TBA is not used). The choice of extracting agent is also vital in the color development process as Warren *et al* reported that some organic solvents would interfere with this. They reported that the ideal solvent is cyclohexanone. Color formation resonance is also less prominent in polar solvents (e.g. DMSO) than non-polar. Aminoff reported that the complex is more stable in acidified butan-1-ol. Preliminary testing with acidified propan-1-ol showed an increase in absorbance at 549nm. The ideal ratios of reagents, based off previous work is confirmed and will be used for the final method of this assay.

5.2.3 End-Point Limitations

As discussed previously, two main assays exist: continuous and endpoint. TBA falls in the latter category (section 5.1.1), meaning that you cannot monitor the product formation or depletion in real time.

Figure 5.4 (left) represents a singular progress curve of an arbitrary enzyme kinetic reaction:

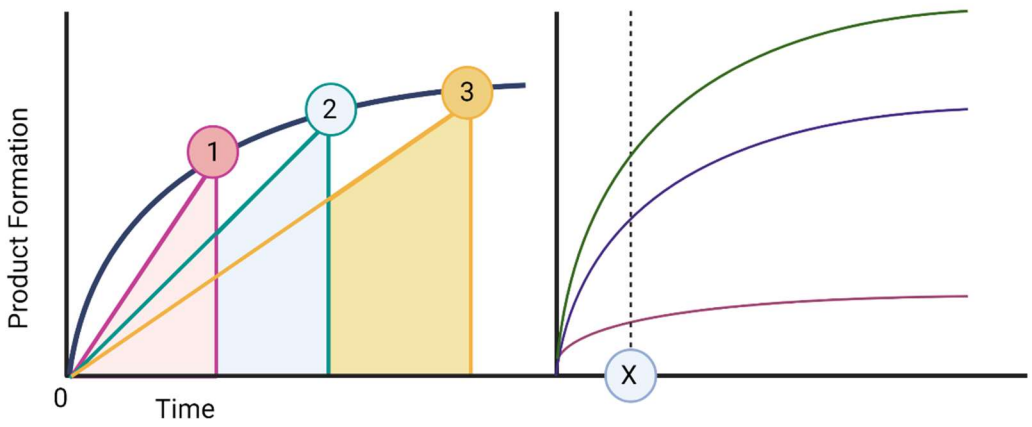


Figure 5.4 A singular progress curve shown with (a) three different slopes and (b) multiple progress curves of varying substrate concentrations at an arbitrary timepoint.

The initial velocity of any enzymatic reaction is represented by the linear portion of the progress curve. This is the ideal region and is referred to as coinciding with the 'steady state'[82]. In Figure 5.3 (left) this is represented by any point before **1**. This is very easy to plot or monitor in real time, however for the TBA assay this is not possible as the chromogen is extracted after almost 40 minutes. Ideally, the reaction will proceed for a defined amount of time, the product quantified and then an estimated average rate will be determined. The major drawback of this is 'going in blind'. The time for the steady-state region can be overestimated, which will correspond to point **2** in Figure 5.3. The steady state is no longer valid, and the average rate at this point will deviate. As the reaction proceeds to the plateau region, it will saturate and is indicated by point **3**. This is most likely the method used by Xu *et al* (2009)[4] since the enzyme and substrate was incubated for 1 hour. Again, by doing this the slope will deviate significantly from the steady state, in fact it will decrease as it approaches saturation[11, 16].

To account for this, a sample will need to be taken at small time intervals for at least 20 minutes. Table 5.3 summarizes the theoretical product formation of an enzymatic reaction since we expect 10-15% conversion. The conversion equates to the final concentration product of the mixture as the final volume has increased 10-fold. When the reaction progresses, the concentration will

increase. In other words, the final amount of product will represent point **3** in Figure 5.3 (left). Any time point before then, it will have a lower concentration. This is essential as the sensitivity needs to be increased for the lower concentrations. The limit of detection (LOD) is the upper range of the product that can be detected, while the limit of quantification is the lowest possible concentration of product[94]. Thus, if the LOQ is 2 µM then any final concentration below that will be undetectable in the assay (Table 5.3).

Table 5.3 Theoretical product conversion of different substrate concentrations.

Substrate			Product		
Concentration (µM)	10 % Conversion	Final Concentration (µM)	15% Conversion	Final Concentration (µM)	
3000	300	5.41	450	8.11	Detectable
1500	150	2.70	225	4.05	
750	75	1.35	112.5	2.03	
375	37.5	0.68	56.25	1.01	Undetectable
187.5	18.75	0.34	28.12	0.51	
93.75	9.375	0.17	14.062	0.25	

The most prevalent issue with timing can be seen in Figure 5.4 (right), where progress curves are shown in decreasing substrate concentration (from top to bottom). If an arbitrary timepoint is selected (indicated by point **x**) the steady state might be observed for the higher concentrations but missed for the lower.

5.2.4 Final Protocol

All of the experiments were performed using this protocol:

- Add 50 µL of sample to 2mL eppi.
- Add 100 µL of the oxidant and incubate for 30 minutes at 37 ℃.
- Add 50 µL of the reducing agent and wait until the tubes become clear after turning brown.
- Add 750 µL of the 2-TBA solution.
- Put the tubes in the heatblock for 10 minutes at 99 ℃.
- Remove tubes and put on ice to cool down
- Add 900 µL of the acidified butanol
- Centrifuge for 3 minutes at 1500 rpm
- Extract 200 µL of the top butanol layer (pink) and read at 549 nm

5.3 Results

5.3.1 N-acetylneuraminic Acid Standards

With all the key reagents identified and the method optimized, the first step in any enzymatic reaction is to obtain a standard curve of the product. For this, we make a serial dilution of Neu5Ac (product) and plot it as a function of absorbance observed at 549nm.

To increase the sensitivity, all reactions must be normalized by removing any background interference. The blank used most frequently is water as most of the neuraminic acid standards were dissolved in that. However, in this case a MES/$CaCl_2$ buffer is used to facilitate enzymatic reactions. The blank must be subtracted from every sample to account for any background signal. Figure 5.5 shows it is stable and will not contribute to the noise.

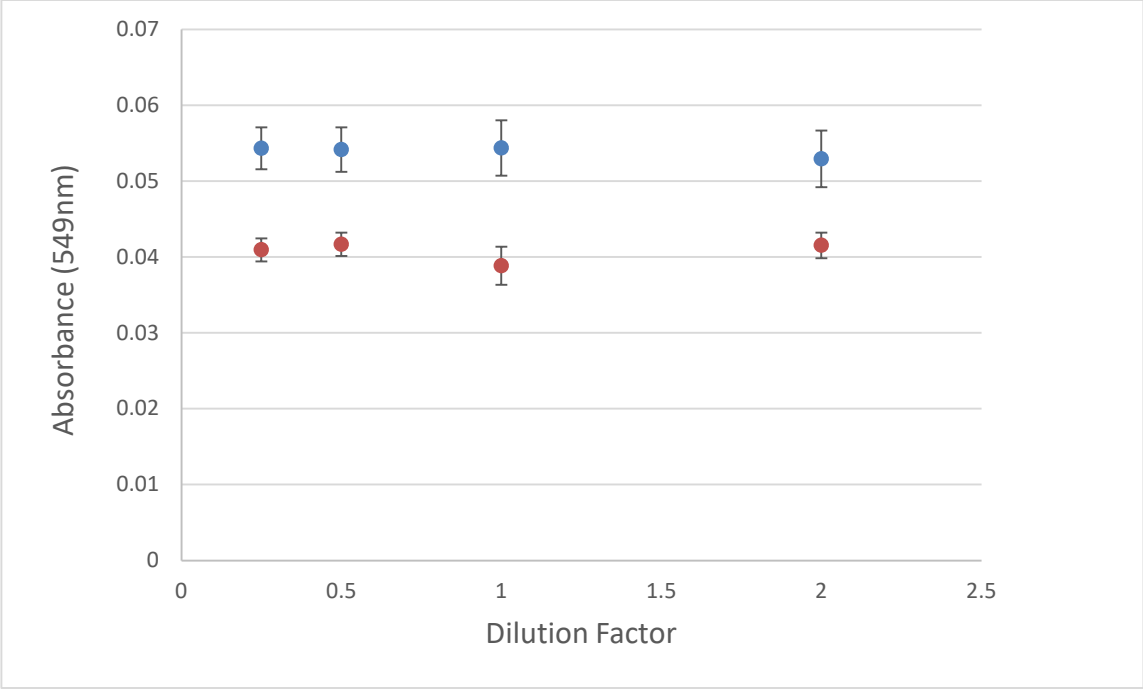


Figure 5.5 Graphical representation of blank serial dilutions and their absorbance. The blue and red dataset represents a volume of 350 µL and 200 µL respectively.

Another important variable apart from the choice extracting agent, is the volume used. Even with ideal ratios, the amount of acidified butanol can be modified. The main reason is that in the original papers, the final volume exceeded 4 mL total. Standard eppis (tubes) used in the methodology have a maximum volume of 2 mL. These can be heated more adequately and evenly in the heat block. Whereas if you use larger tubes (15mL or above) you need to transfer into a smaller volume to heat with the coloring solution (step 3 in methodology). This leads to loss of sample as well as transfer errors.

To optimize the standards, three separate experiments were conducted, with decreasing volumes of butanol used so that it will all fit into one standard 2 mL tube. The volume of butanol used decreased from 3 mL (method A) to 2 mL (method B) and finally to 900µL (method C). All three standards are plotted in Figure 5.6. They all showed excellent linearity (R^2 values all exceed 0.99).

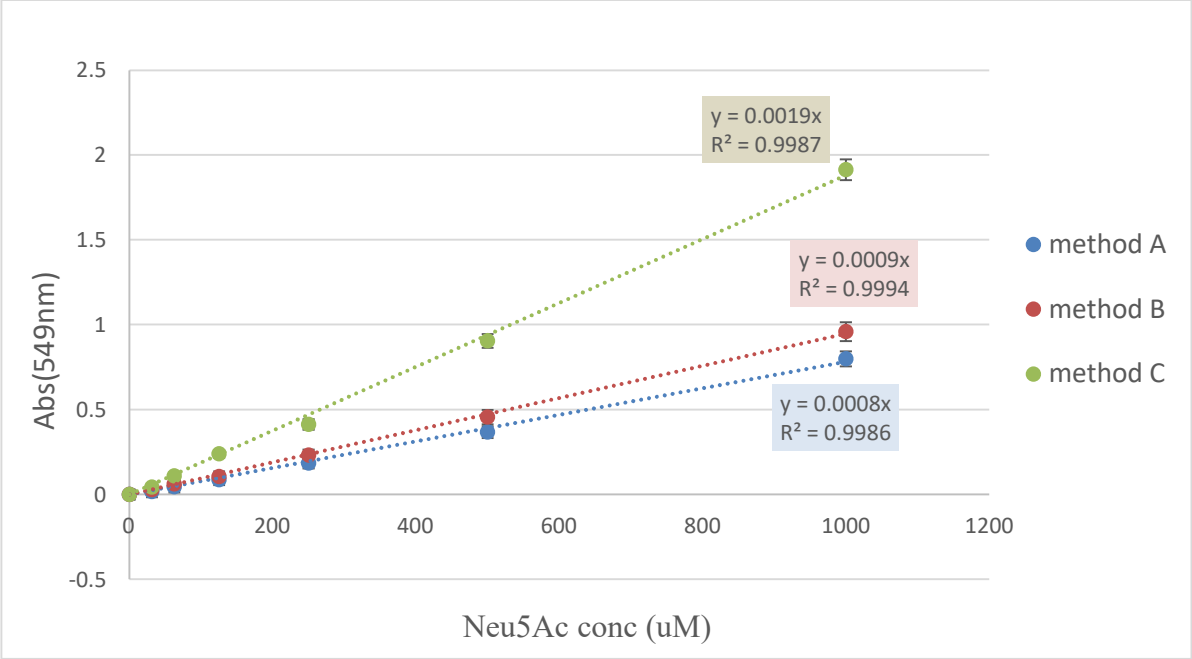


Figure 5.6 Neuraminic acid standards with different volumes of extracting agent used.

Any quantity that exceeds the LOD or falls short of the LOQ is outside of the linear range. The LOD usually corresponds to a S/N ratio of 10, while LOQ is 3. These are a set of loose criteria for any linear regression and can be modified depending on the assay. The S/N and corresponding concentrations are summarized in Table 5.4 and plotted in Figure 5.7.

Table 5. 4 S/N ratios of three different methods used for extraction.

Method	LOD (µM)	S/N	LOQ (µM)	S/N	Range
A	664.80	14.81	157.55	1	507.25
B	490.30	19.67	113.96	1	376.34

C	248.59	38.90	63.41	1	185.19

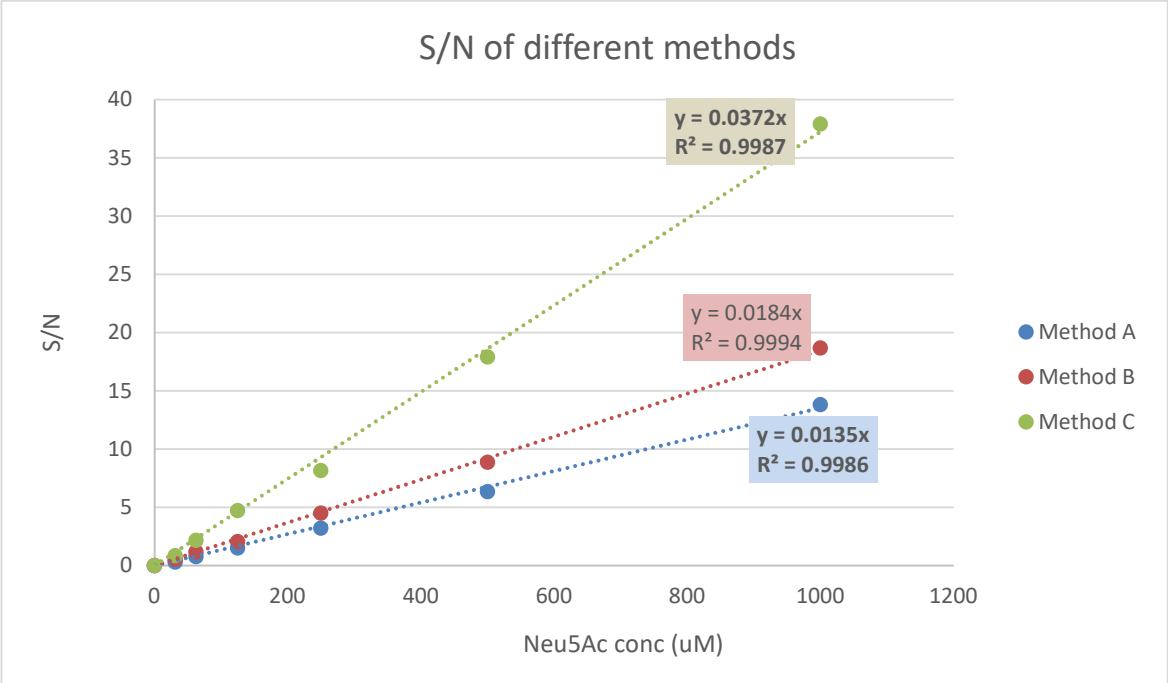


Figure 5.7 The S/N ratio of the three different extraction methods used for Neuraminic acid.

The S/N ratio of the highest concentration of neuraminic acid (Table 5.4 column 2) increases as the extracting volume decreases, which is ideal as background noise needs to be minimized. Reducing the extracting volume by 50% increases the S/N by more than 100%. The detectable range (Table 5.4 column 6) reduces and the LOD to LOQ window shifts. A larger extraction volume results in a larger LOD, while the reduced volume results in a much lower LOD. One might consider this observation as a limiting factor but referring to Table 5.3; enzymatic reactions have a maximum of 15% conversion. Hence, shifting the linear range to a lower LOQ is beneficial as low product formation is still quantifiable. On the contrary, the LOD should also not be too low, as higher concentrations will saturate in this region and sensitivity is lost[95].

By reducing the extraction volume, we minimize the transfer error (loss of sample of product) and improve the S/N ratio, thus increasing the sensitivity for the lower concentration regions.

5.3.2 Substrate Hydrolysis

Reviews of colorimetric assays, specifically TBA, indicate that sialosides (such as 3'-sialyllactose or 6'-sialyllactose) should not react in this assay, especially since harsh acidic conditions are not used. However, in Cheeseman *et al* (2021)[96] they reported that 3'-sialyllactose cleaves 0.92 ± 0.15 % and 6'-sialyllactose 1.5 + 0.14%. Thus, the main question is how much of the intact substrate will hydrolyze in the new buffer system, and should it be accounted for in terms of significance?

A serial dilution of 3'-sialyllactose was made and the procedure carried out using Method C. The different substrate concentrations were plotted as a function of the corresponding absorbance (Figure 5.8) and converted using the standard graph from Figure 5.6, method C to render Figure 5.9.

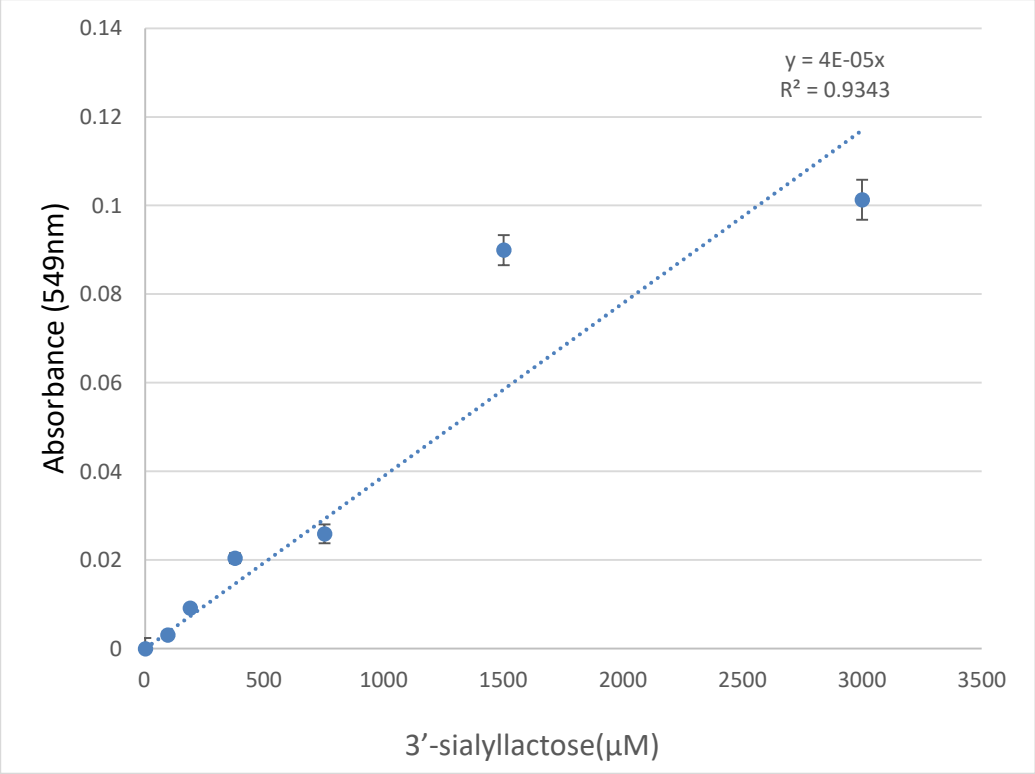


Figure 5.8 Different concentrations of 3'-sialyllactose and their corresponding absorbance.

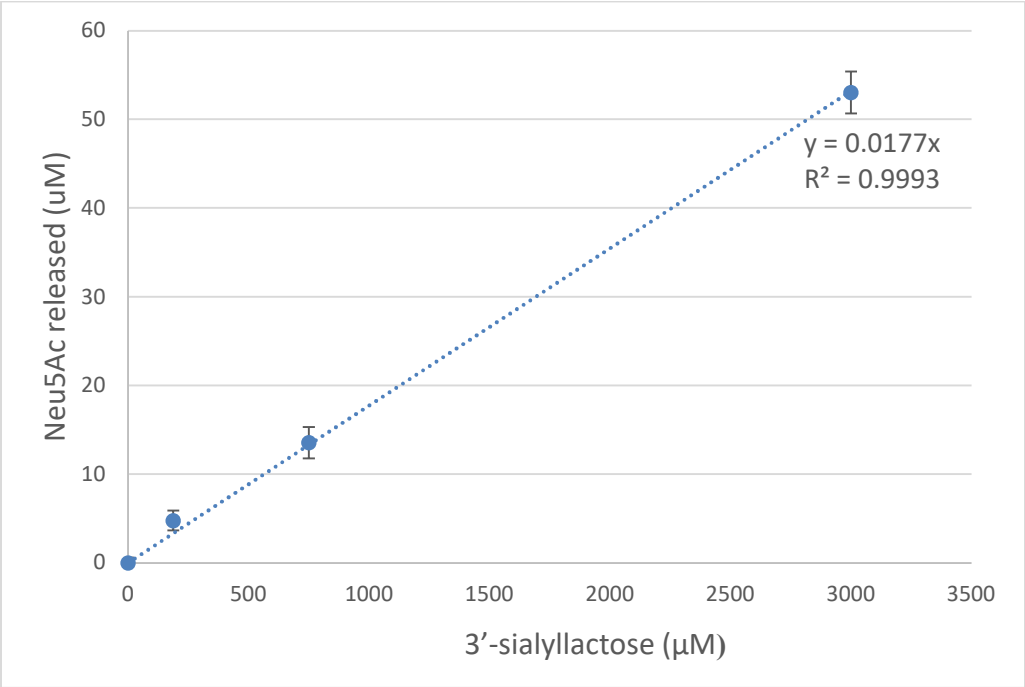


Figure 5.9 Free neuraminic acid as a function of 3'-sialyllactose concentration.

From Figure 5.9 we can see that for every 1µM of 3'-sialyllactose we release 0.017 µM of sialic acid. The highest concentration hydrolyzed 1.7 ± 0.01 % after 30 minutes. The S/N of the substrate concentration is plotted in Figure 5.10. Note the S/N of the highest concentration is below 3, meaning that the signal falls below the LOD and thus varies in the percentage hydrolysis. The product release is linear, meaning we can estimate how much sialic acid is released at a given time.30 Minutes was the default time for oxidation chosen as this is the incubation time of the samples.[85].

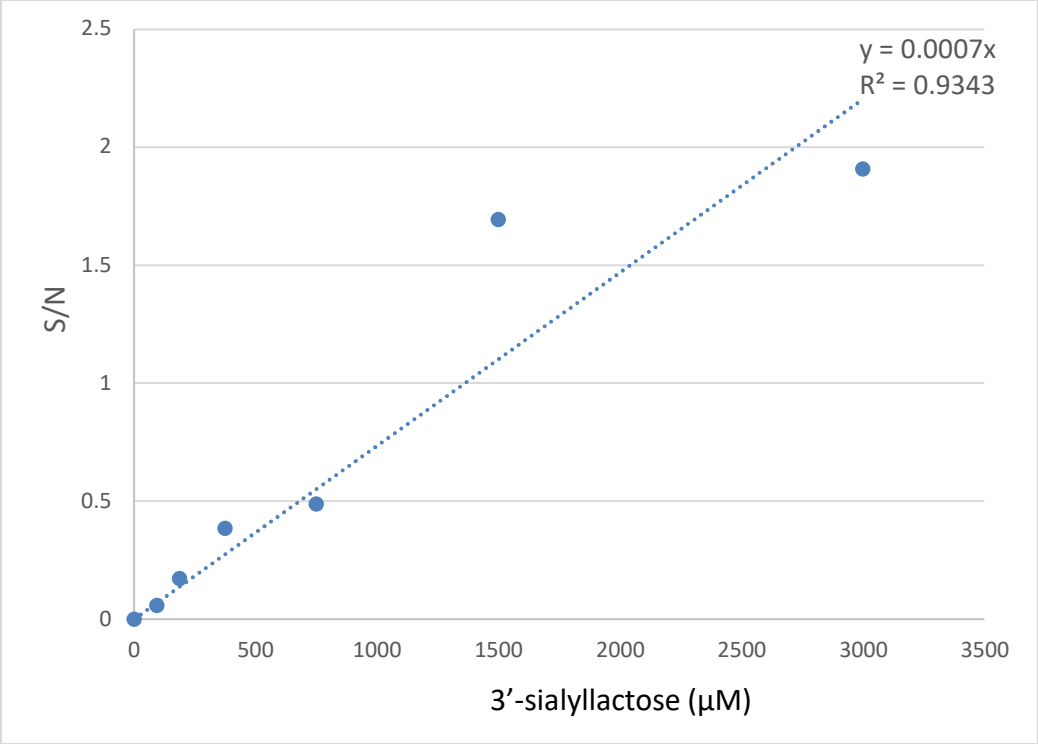


Figure 5.10 S/N ratio of 3’-sialyllactose hydrolysis

To determine whether the hydrolysis was significant enough to be accounted for, a two-tailed t test was performed on 2 sets of data, one being the blank and the other a set concentration of 3’-sialyllactose. The ‘overlap’ was then determined to be significant or not. The results are summarized in Table 5.5.

Table 5.5 Results of two-tailed t test to determine if hydrolysis of substrate is significant or not.

3'-sialyllactose concentration (μM)	P (T<=t) two-tail	Significance
0	160×10^{-3}	Insignificant
187.5	6.50×10^{-3}	Significant
375	4.11×10^{-4}	Significant
750	3.40×10^{-4}	Significant
1500	6.75×10^{-4}	Significant
3000	9.87×10^{-4}	Significant

From the substrate concentration of 187 μM onwards, the hydrolysis rate is significant enough to account for in terms of making corrections, this can be done by using Figure 5.9 to estimate the sialic acid released by spontaneous hydrolysis.

5.3.3 Time Dependency of Oxidation

The set incubation time of 30 minutes is essential for this assay. Without this crucial step, oxidation cannot occur in full, and no sialic acid is detected. To test if incubation time has a significant influence on the oxidation rate or hydrolysis, the highest substrate concentration of 3000 μM was used as reference. The sialic acid released was quantified using different time intervals up to 60 minutes. The results are summarized in Table 5.6.

Table 5.6 Free sialic acid released as a function of time for 3000 µM of 3'-sialyllactose.

Time	Sialic acid released (µM)	%
0	11.85 ± 0.12	0.39 ± 0.09
15	19.56 ± 0.23	0.65 ± 0.12
30	41.79 ± 0.45	1.40 ± 0.31
60	96.53 ± 0.31	3.21 ± 0.52

The data from Table 5.6 is consistent with the previous section's results in Figure 5.9. After 30 minutes incubation the 3000 µM substrate hydrolyzed 1.7 %. According to Table 5.6, the same concentration hydrolyzed 1.4 ± 0.31 %. This means that the hydrolyses can be corrected as a function of time.

5.4 Kinetic Parameters

5.4.1 3'-sialyllactose

Thus far we have established suitable reagents, an optimized method for absorbance detection, as well as accounting for any substrate hydrolysis as a function of time. As previously discussed, for any kinetic experiment a product output or depletion of a substrate needs to be monitored. For an end point assay, the most frequently used method is discussed in section 5.1.1, Xu *et al* (2009)[4] performed a kinetic experiment using *Pneumonia* NanA with both 3'-sialyllactose and 6'-sialyllactose.

From their research, they reported that the samples were incubated with the enzyme for 1 hour. This principle was used along with the designed method to obtain a Michaelis Menten plot and constant. The raw data for the product formation is summarized in Table 5.7 and plotted in Figure 5.11.

Table 5.7 N-Acetylneuramini**c** acid formation due to enzymatic cleavage for 3'-sialyllactose as substrate.

Substrate (µM)	Neu5Ac released (µM)-corrected for hydrolysis	Rate(µM/min)
3000	412.58 ± 3.21	6.87
1500	390.39 ± 2.19	6.50
750	266.15 ± 1.91	4.43
375	185.91 ± 3.14	3.098
187.5	117.91 ± 4.12	1.96
93.75	58.61 ± 2.22	0.97
0	0 ± 1.11	0

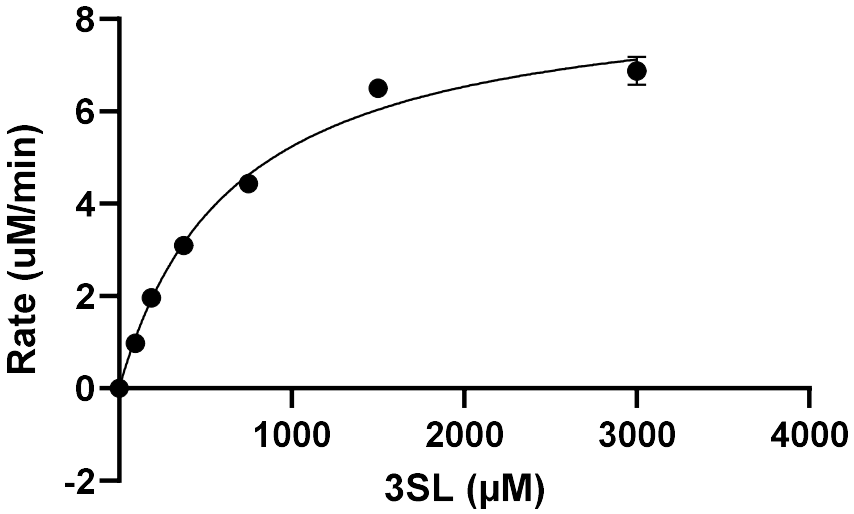


Figure 5.11 The resulting Michaelis-Menten curve of NanA with 3'-sialyllactose after 60 minutes of incubation.

The K_m obtained was 659.73 ± 91.12 µM, while the K_{cat} was 827.81 min^{-1}. This compares well with the literature values (all kinetic parameters summarized in Table 5.8). However, this is assuming

the enzyme cleaves the substrate continuously for an hour as the rate was determined by dividing the net product formation by 60 minutes. To improve this, time points were taken to establish a progress curve for each concentration. The rate then equates to the slope of each curve, giving a better estimate of the initial velocity. This is represented in Figure 5.12.

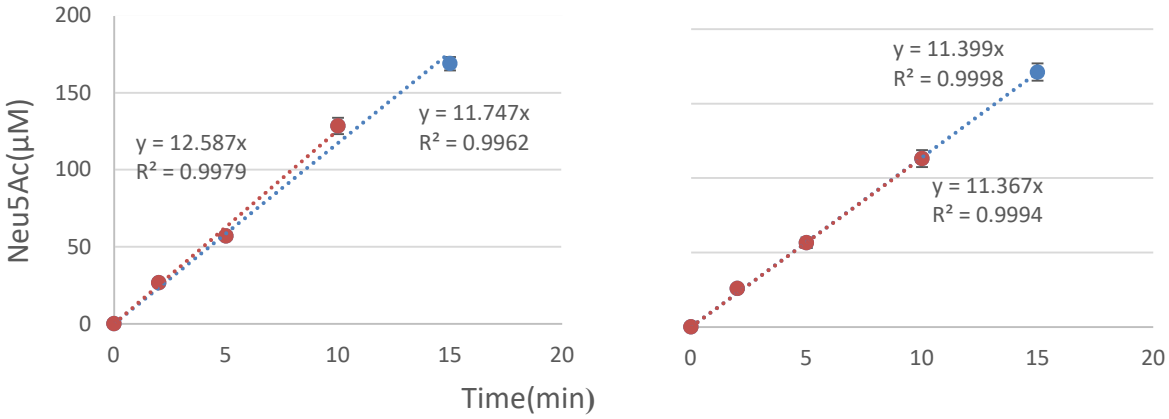


Figure 5.12 Example of two progress curves of 3'-sialyllactose to determine the initial velocity for a substrate concentration of (left) 3000µM and (right) 1500 µM.

By plotting the initial velocity of the various substrate concentrations, a Michaelis-Menten curve with a much lower K_m is yielded (Figure 5.13), but a similar K_{cat} to the reference.

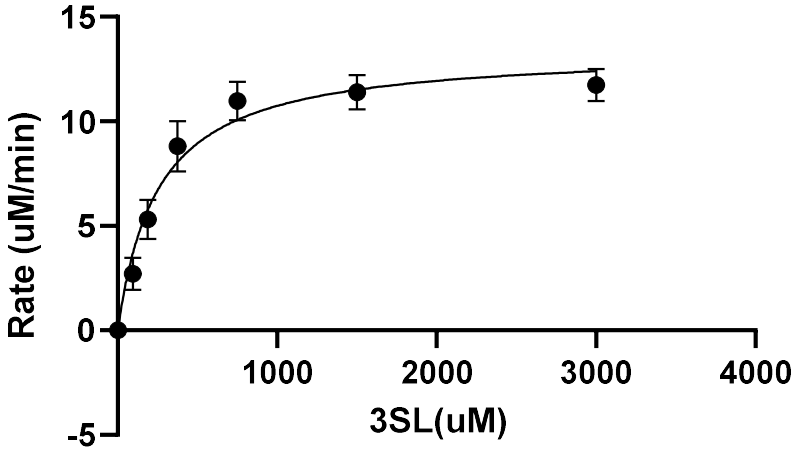


Figure 5.13 The resulting Michaelis-Menten graph of 3'-sialyllactose determined by the initial velocity.

5.4.2 6'-sialyllactose

For NanA and 6'-sialyllactose, the reported K_m value was significantly higher than 3'-sialyllactose (2170 ± 140 µM). This indicates the affinity for the enzyme is much lower.

The color development was not ideal, even though the enzyme was left to incubate with the sample for 5 minutes. It was almost the same as the hydrolyses rate i.e. the enzyme had negligible effect. It was then decided to incubate the substrate and enzyme for one hour. After correcting for hydrolysis, the final Michaelis-Menten curve yielded a result like that reported (Figure 5.14).

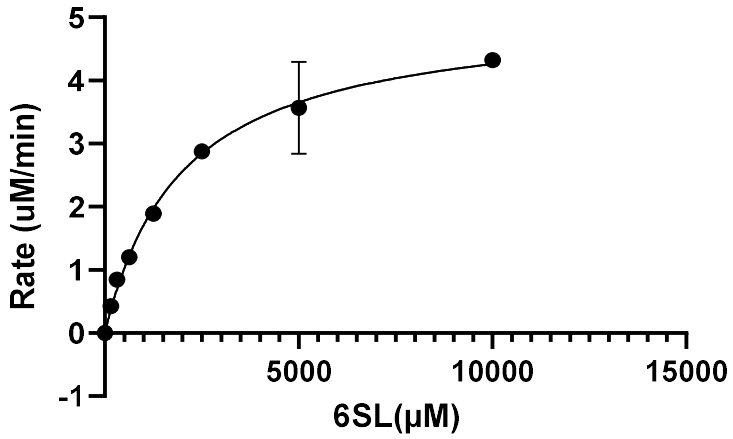


Figure 5.14 The resulting Michaelis-Menten graph of NanA with 6-sialyllactose as substrate after 60 minutes incubation.

The result yields kinetic parameters similar to that of Xu *et al* (2009)[4, 97] and is shown in Table 5.8. This confirms that the rate of the product formation is an average taken over the time course of 60 minutes. Unfortunately, NanA has a much lower affinity for 6'-sialyllactose than 3'-sialyllactose, hence the turnover rate is slower, and no progress curves can be taken. The kinetic parameters can be obtained from all three substrates analyzed so far and will be discussed in the next section.

5.5 Discussion

5.5.1 Comparison Between 3'-sialyllactose and 6'-sialyllactose

The kinetic parameters for both natural substrates are summarized in Table 5.8.

Table 5.8 Kinetic parameters for both 3'-sialyllactose and 6-sialyllactose.

	3'-sialyllactose			6'-sialyllactose	
	Experimental (end point)	Experimental (initial velocity)	Xu *et al*	Experimental	Xu *et al*
Km (μM)	659.71	250.89	640.00	1989.41	2170
Vmax (μM)	6.87	10.91		5.10	
K_{cat} (min^{-1})	827.83	1221.00	892.00	1021	916
ΔA‡ (kcal/mol)	-15.89	-15.66	-15.85	-15.88	-16.12
K_{cat}/Km ($\mu M^{-1}.min^{-1}$)	1.25	4.86	1.39	0.51	0.42

K_m and K_{cat} are two fundamental kinetic parameters. K_{cat}/K_m is the second order kinetic constant and referred to as the specificity number as it varies with different substrates. It is common practice to monitor the change of this ratio with varying conditions or substrates to gain better insight to the mechanism of action.

The K_m is simply the concentration of substrate at which the enzyme is half saturated. From Table 5.8 the end point method used for both substrates compares well with the literature values. K_{cat}/K_m can give insight to the effects of certain modifications. If any of these alter the energy of the highest transition state, a change in this specificity ratio will be evident. This is why enzymologists consider K_{cat}/K_m along with K_{cat} as a fundamental kinetic parameter in enzyme kinetics and consider K_M as simply the ratio of the two fundamental kinetic parameters. 3'-sialyllactose has the same catalytic efficiency for both methods used, as well as 6'-sialyllactose. The free energy conversion of K_{cat} agrees well with the literature. The K_{cat}/K_m ratio for the end

point method for both assays agrees with the literature values. Of course, using the modified method (initial velocity) results in a higher ratio. In general, the increased ratio indicates that enzyme is more specific for that substrate. In conclusion, NanA shows a higher specificity for 3'-sialyllactose than 6'-sialyllactose.

5.5.2 Evaluation of 2'-(4-methylumbelliferyl)-α-D-N-acetylneuraminic acid and 3'-sialyllactose

The aim of chapters 4 and 5 is to evaluate MUANA as a substrate against the natural substrate. Table 5.9 summarizes all the important kinetic parameters.

Table 5.9 Summary of all kinetic parameters for the two natural substrates and artificial substrate.

Substrate	K_m	V_{max}	K_{cat}
MUANA	41.56	3.75	851.19
3'-sialyllactose	659.75	6.87	827.82
6'-sialyllactose	1989.41	5.10	1021.45

From Table 5.9 the Michaelis constant K_m is vastly different across all three substrates. NanA has a much higher affinity towards the artificial substrate due to lower Km. The K_m for 3'-sialyllactose is more than 10-fold higher than that of MUANA, with 6'-sialyllactose having a K_m 50-fold that of MUANA. Although the affinity is much different across all three substrates, the K_{cat} or turnover rate is relatively fast for each substrate. Thus, even if the enzyme is 'slow' with the initial binding to the enzyme, the conversion from substrate to product proceeds at a similar rate for all three substrates. This conclusion is intriguing and will be investigated further by conducting inhibition studies with NanA to determine if MUANA is a suitable replacement for the natural substrate.

Chapter 6

Enzyme inhibition

6.1 Theory Overview

6.1.1 Different Modes of Enzyme Inhibition

The aim of this section is to characterize NanA produced by *S. pneumoniae*. In general, K*cat* and Km are sufficient, however inhibition will reveal if MUANA is a suitable replacement for the natural substrate.

An inhibitor is a molecular agent that binds specifically to an enzyme to interfere with the catalytic process. There are other factors such as pH, temperature or ionic strength that can influence overall enzymatic activity. However, for true enzyme inhibition a molecule must bind to a specific enzymatic site[16, 98].

An enzyme reaction proceeds as in Equation 6.1:

$$E + S \leftrightarrow EP \rightarrow E + P \qquad [6.\ 1]$$

When inhibitors are introduced, they will bind to the enzyme and interfere with the catalytic process as shown in Equation 6.2:

$$I + E \leftrightarrow IE \rightarrow \boldsymbol{X}P \qquad [6.\ 2]$$

There are two main categories of enzyme inhibition: irreversible and reversible inhibition (Figure 6.1). For the former, the inhibitor will attach itself to the enzyme via a covalent bond and cannot be removed. Reversible inhibition occurs when the inhibitor binds to the enzyme's active site through a weak bond instead of a covalent bond and can be removed by dialysis or dilution.

Furthermore, reversible inhibition can be categorized into either competitive, uncompetitive or non-competitive inhibition (Figure 6.1).

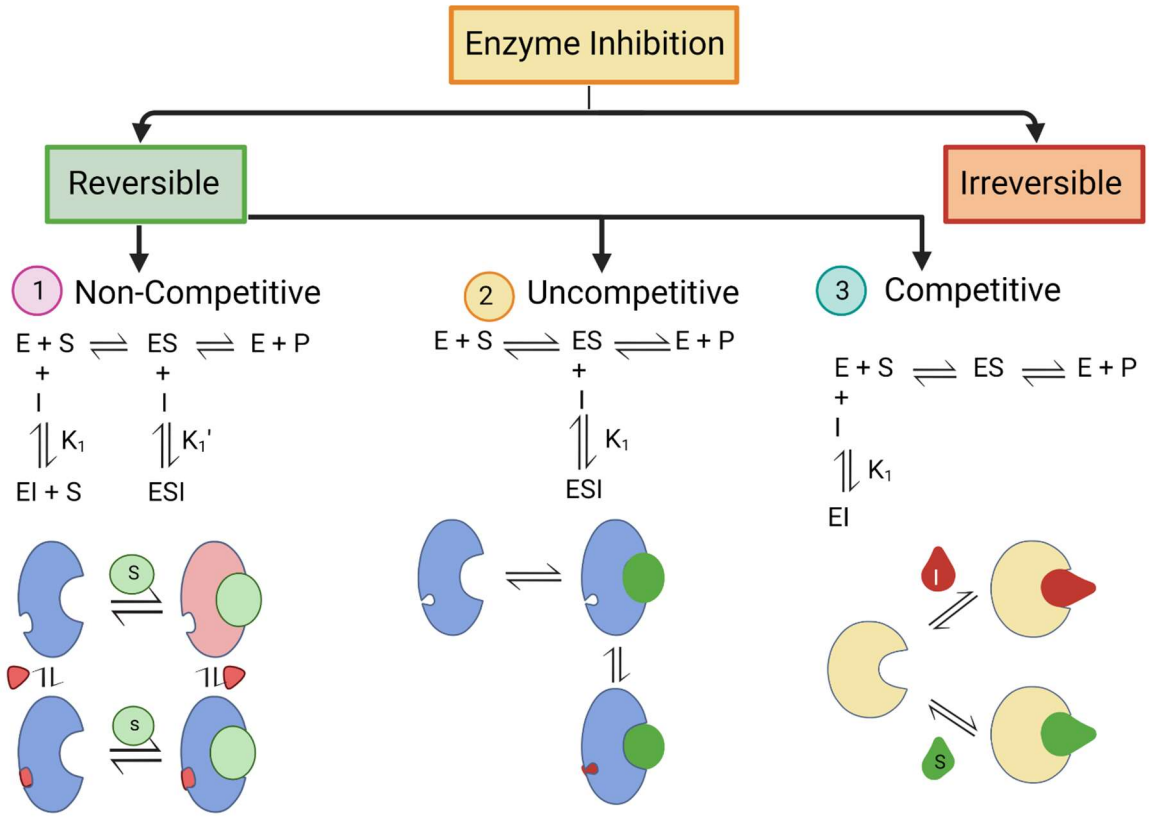
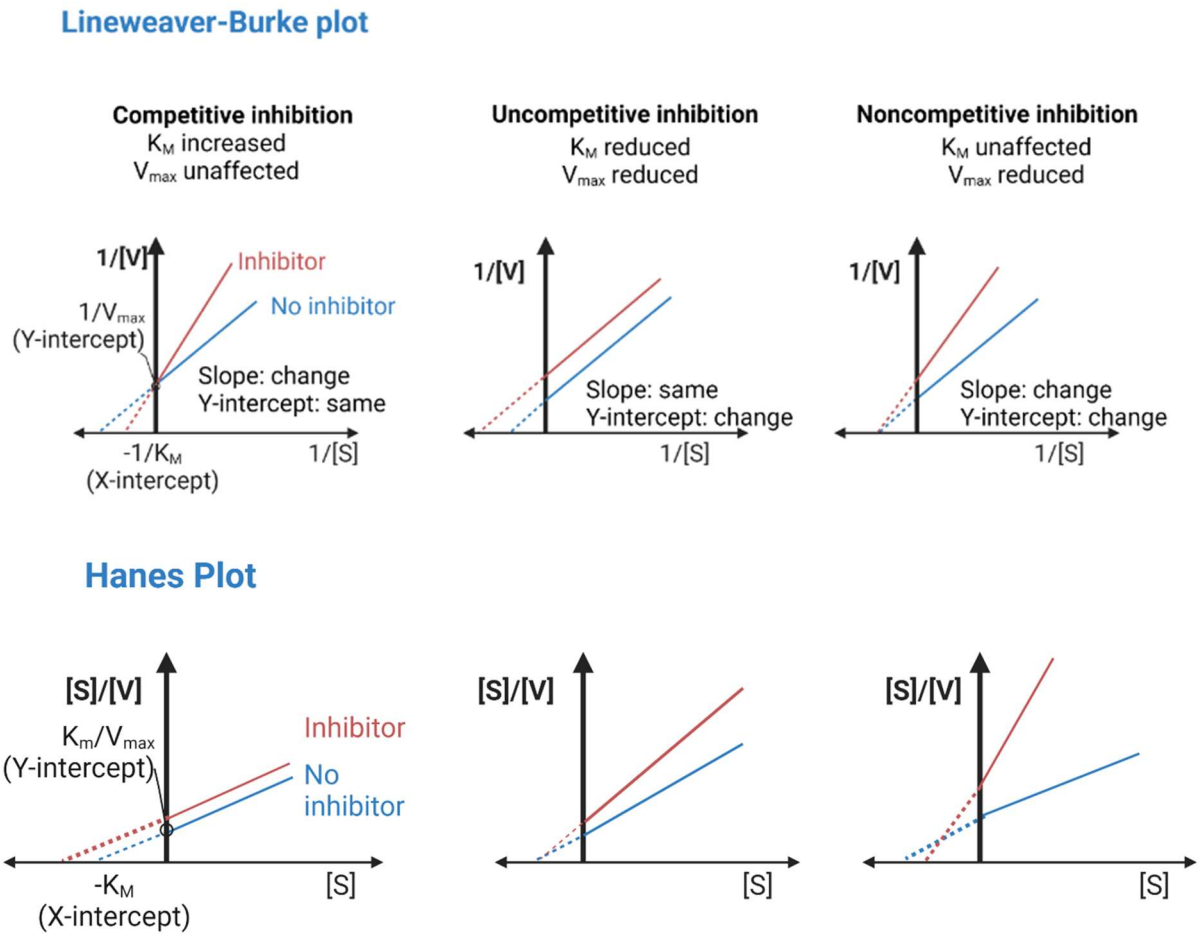


Figure 6.1 The three different modes of reversible enzyme inhibition (from left to right) are non-competitive, uncompetitive, and competitive.

Firstly, there is non-competitive inhibition where the inhibitor binds to both enzyme and enzyme-substrate complexes to inhibit the enzymatic function. The reaction rates K_1 or K_1' are affected by these inhibitors[99].Secondly, there is uncompetitive inhibition where the inhibitor exclusively binds to the enzyme-substrate complex. Lastly, there are competitive inhibitors where the inhibitor binds exclusively to the enzyme and competes with the substrate for the active site. When bound to the enzyme, an enzyme-inhibitor complex forms and the substrate is unable to bind[100] .

The efficacy of an inhibitor must be established; from Equation 6.2 the first step is the binding of the inhibitor to the enzyme. The association/dissociation constant (K_i) characterizes the intensity of this binding and represents the amount of inhibitor required to achieve half-saturation. However, to determine K_i both substrate and varying inhibitor concentrations are required[101]. In any inhibition assay, there are two controls used, the first being the 0% (fully inhibited) and the second being the 100% (uninhibited) reaction. The ultimate test series is where both substrate and inhibitor concentration are varied where the reciprocal of the initial velocity is plotted as a function of increasing substrate concentration[79]. The resulting lines form distinct patterns, revealing the different modes of inhibition (Figure 6.2).



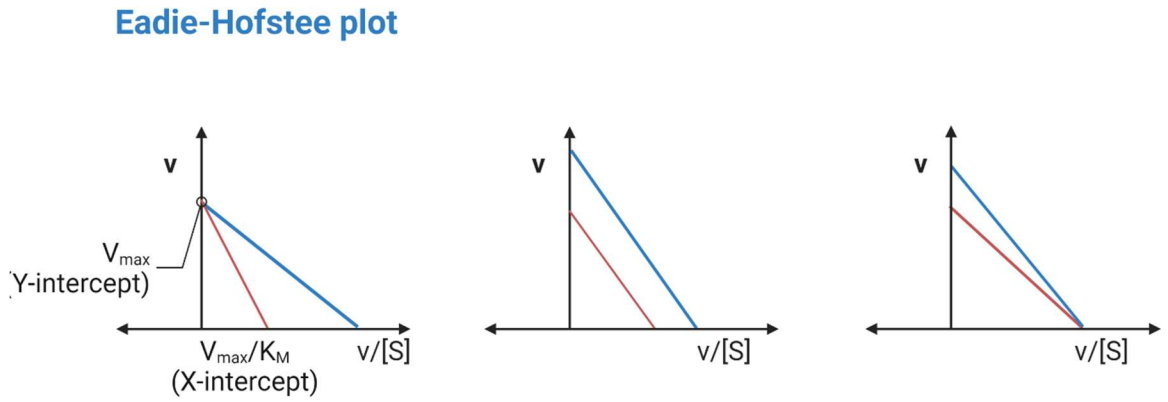


Figure 6.2 The three alternative ways to plot enzyme inhibition are (from top to bottom): Lineweaver-Burke plot, Hanes plot and Eadie-Hofstee plot.

K_{ic} refers to the binding constant of the inhibitor and is defined in Equation 6.3:

$$K_{ic} = \frac{[E][I]}{[EI]} \qquad [\,6.\,1\,]$$

The most common way to determine this constant is to plot the Lineweaker-Burke as a Dixon plot as shown in Figure 6.3. The inhibitor concentration is plotted against the initial velocities of each concentration of substrate. The lines will intersect at $-K_{ic}$[70].

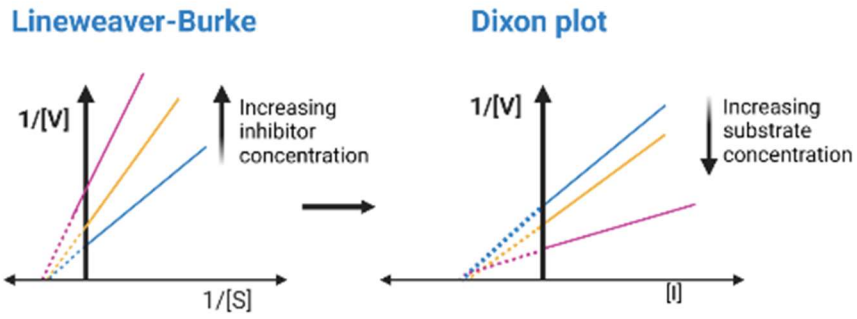


Figure 6. 3 The transformation of a Lineweaver-Burke plot into a Dixon plot to determine the K_i value.

6.1.2 IC_{50} and K_i

Pharmaceutical companies strive to discover the top inhibitors in the shortest amount of time. It is a balancing act between finding the lead compounds and number of assays performed, thus introducing the inhibitory constant (IC_{50}) and effective concentrations (EC_{50}). These two constants evaluate the suitability of inhibitors early in the drug discovery process with an added advantage being that they allow for a quantitative comparison of the inhibitory effect across different target compounds. This enables standardisation of inhibition measurements as the potency of compounds can be compared against each other[102].

The two most common constants used to illustrate the inhibitory effect of different compounds are the IC_{50} and K_i. However, before establishing the relationship between the two a more in-depth understanding is required for the IC_{50} model used[103, 104]. It is assumed that there is a monotonic relationship between the dose (of inhibitor) and response. The general scheme for an IC_{50} graph is given in Figure 6.4:

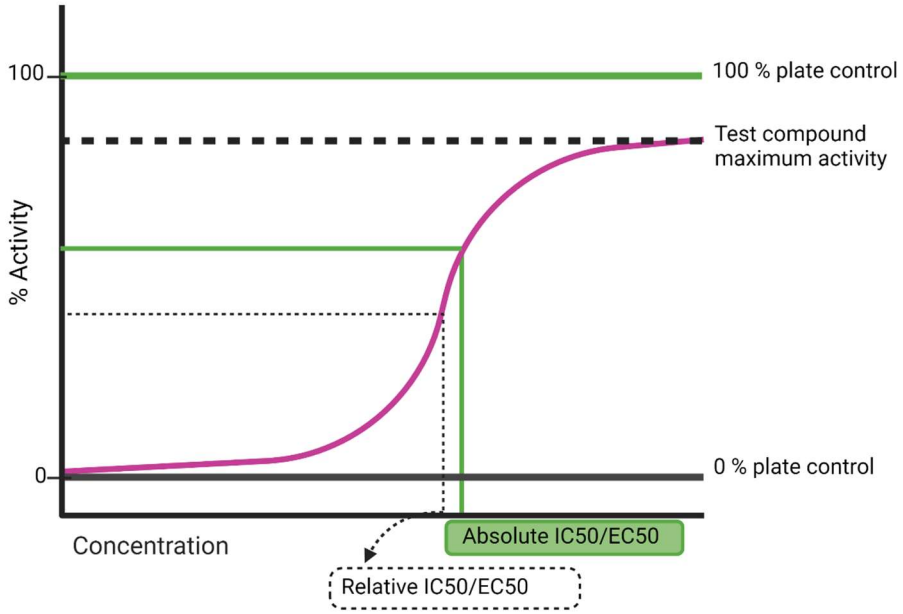


Figure 6.4 The two alternative ways to represent the IC_{50} /EC_{50}.

In Figure 6.4 there are two values for the IC_{50}/EC_{50} due to the two controls implemented during an inhibition screening. The baseline corresponds to the 0% plate control containing no enzyme and serves as the zero reference point[105]. There can however be many options to determine the 100% plate control such as low inhibitor concentration, exclusion of co-factors or simply having no inhibitor present to produce the maximum effect. The problem is evident as the relative IC_{50}/EC_{50} does not coincide with the absolute IC_{50}/EC_{50} and can be due to the 100% plate control differing from the maximum test activity. The latter may even vary among all test compounds. Thus, it is good practise to consistently have a measure of high activity that can serve as the standard 100% control[106].

The sigmoidal-shaped response curve can be described by a mathematical model with four parameters:

$$Y = \frac{a - d}{1 + \left(\frac{X}{c}\right)^{b}} + d \qquad [6.4]$$

Where **X** represents the concentration of the inhibitor and **Y** the response. The lowest asymptote is represented by a and the upper by d. The slope factor b describes the steepness of the linear region of the curve. The concentration midway between a and d is represented by c, the relative IC_{50}/EC_{50}. The mean of the 0% and 100% control is the absolute IC_{50}/EC_{50}. The data is fit using a nonlinear regression followed by an inverse regression to solve Equation 6.5 by setting **Y** to 50%. Sebaugh (2011)[105] concludes that at least two concentrations beyond the lower and upper bend points are required to estimate the relative EC_{50}/IC_{50} . Equation 6.5 represents the final equation used to determine the IC_{50}/EC_{50}:

$$EC50|IC50 = c\left(\frac{a - 50\%response}{50\%response - d}\right)^{\frac{1}{b}} \quad [6.5]$$

With guidelines in place for an accurate analysis, the relationship between the IC_{50} and Ki needs to be understood. The smaller the Ki value, the tighter the binding to the enzyme and the more potent the inhibitory effect. A similar concept applies for the IC_{50}, as a low value indicates that less inhibitor is needed to reduce the enzyme activity by 50%[106]. The main difference in obtaining the two constants is that for K_i one must vary both the inhibitor and substrate concentration to determine the rate of the enzymatic reaction whereas for the IC_{50} one can use one substrate concentration over a range of inhibitor concentrations. The convenience of the IC_{50} is why it is primarily used in drug screening; however, the value is relative to the substrate concentration whereas Ki is a constant value for any inhibitor with a certain substrate and enzyme[102].

For competitive inhibitors the relationship can be expressed in Equation 6.6:

$$IC_{50} = K_i\left(1 + \frac{[S]}{K_m}\right) \quad [6.6]$$

Where [S] refers to the substrate concentration used in the assay and K_m refers to the Michaelis-constant. The relation between IC_{50} and K_i depends on the mode of inhibition. For uncompetitive inhibitors the relationship is expressed in Equation 6.7:

$$IC_{50} = K_i\left(1 + \frac{K_m}{[S]}\right) \quad [6.7]$$

Mixed inhibition is a combination between competitive and uncompetitive inhibition. This class of inhibitors bind to both free enzyme and the enzyme-substrate complex with K_{ie} and K_{ies} (K_1 and K_1' in Figure 6.1 respectively) as inhibition constants and can be expressed in equation 6.8:

$$IC_{50} = \frac{[S] + K_M}{\frac{[S]}{K_{ies}} + \frac{K_M}{K_{ie}}} \quad [6.8]$$

This concludes that there is a way to compare IC_{50} and K_i where the only requirement is knowing the mode of inhibition and substrate concentration used in the assay[107].

Considering equation 6.7, when [S] << K_m, then the second term in the bracket approaches zero and the IC_{50} will approach K_i. Thus, for competitive inhibition the IC_{50} will always be higher than K_i. A similar concept applies for uncompetitive inhibition where the IC_{50} will always be higher than the K_i, but for a different reason. This will only occur when [S] >> K_m and the term approaches zero.

Burlingham *et al* (2023)[107] developed a method to extrapolate IC_{50} off a Dixon plot. However, if the mode of inhibition is known the above equations will suffice.

6.2 Methodology: Drug Screening Protocol.

6.2.1 MUANA Substrate

As discussed in section 6.1.2, most drug screenings are performed using an end-point approach. The inhibitor concentration is varied and is plotted as a percentage inhibition against a 'blank' or 'uninhibited' reaction. However, in the majority of published papers the substrate concentration used in the assay is never stated. This is a vital factor as the ratio of substrate to inhibitor directly affects the potency. Should the substrate concentration be much lower than the inhibitor concentration, the inhibitory effect could be exaggerated as there is little to no substrate to compete with. Should the substrate concentration exceed that of the inhibitor by a large margin, the inhibitory effect could be underestimated. Thus, it is recommended to use the Michaelis constant, K_m as the substrate concentration. In the case of the 2'-(4-METHYLUMBELLIFERYL)-A-D-N-ACETYLNEURAMINIC ACID substrate, the Km was approximately 50 µM was this was the final substrate concentration used in the assay. The inhibition protocol was as follows:

1. Prepare a 3x concentration of 150 µM MUANA and keep aside.
2. Prepare a 3x concentration of 15 nM NanA and aliquot 20 µL it into a 384 well plate in triplicate.
3. Prepare 3x concentration of each inhibitor and aliquot 20 µL into each well that contains NanA.
4. Cover and incubate the plate at 37°C for 30 minutes.
5. Take plate out and add 20 µL of the 3x MUANA concentration.
6. Immediately insert into fluorometer to monitor product formation over time.
7. Analyze all progress curves. There should be two control curves, one representing the zero percent reaction (completely inhibited) and the other the 100% reaction (completely uninhibited).

8. Choose a timepoint where all curves are linear and not yet saturated or curved. Take the slopes of each inhibitor concentration and normalize it to the inhibited. Then calculate the slopes as a percentage of the uninhibited reaction.
9. Plot the percentages as a function of the inhibitor concentration to obtain $IC_{50.}$

6.2.2 Natural Substrate

As discussed in section 6.3.1 the substrate concentration should be equal to the Michaelis constant. In the case of the natural substrate, the K_m was around 650 µM, thus also the final concentration of the substrate used in the assay. The protocol used is as follow:

1. Prepare a 3x concentration of 150 µM MUANA and keep aside.
2. Prepare a 3x concentration of 15 nM NanA and aliquot 20uL it into a 2000uL tube in triplicate.
3. Prepare 3x concentration of each inhibitor and aliquot 20 µL into each tube that contains NanA.
4. Cover and incubate the tubes at 37°C for 30 minutes.
5. Take tubes out and add 20 µL of the 3x MUANA concentration.
6. Let the reaction continue for 20 minutes.
7. Take each tube and add all chemicals for oxidation, reduction, and chromophore formation (methodology used in chapter 5).
8. Determine the amount of product formed for each inhibitor concentration.
9. Plot the amount of product formed as a function of inhibitor concentration to obtain IC_{50} (including normalization to the uninhibited and inhibited reaction).

6.3 Screening Results and Discussion of Drug Candidates

6.3.1 Oseltamivir

Oseltamivir phosphate (OP) is an inactive prodrug used to prevent and treat influenza viral infection. It is commercially available as Tamiflu and administered orally. The drug is absorbed into the circulating system and then hydrolyzed by carboxylesterase enzymes (found on liver tissues) into oseltamivir carboxylate (OC), the active form (Figure 6.5)[108].

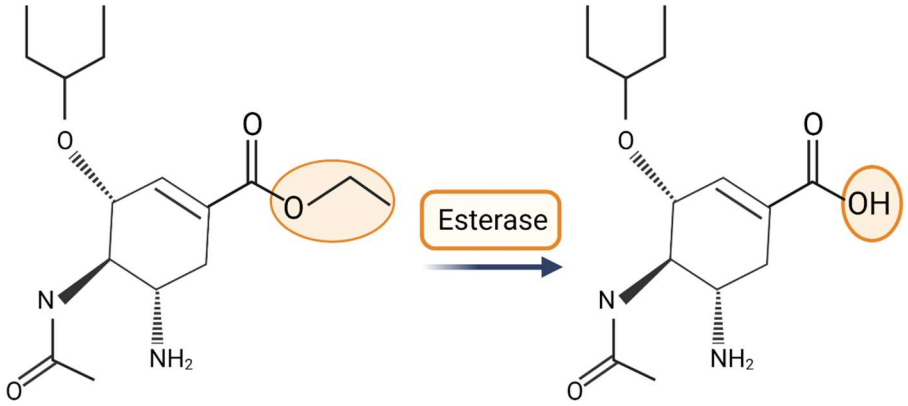


Figure 6.5 Oseltamivir phosphate is hydrolysed via esterase into the active metabolite known as oseltamivir carboxylate.

Influenza relies on neuraminidase to grow and duplicate. OC can reduce the severity and duration of such viral infection and is deemed a safe drug with minor side effects given it is administered in the appropriate dose. *S. pneumoniae* produces sialidase NanA to facilitate pathogenesis. However, these effects are suppressed using neuraminidase inhibitors such as OC. There are many studies focusing on the inhibition of these inhibitors on influenza, but not pneumococcal bacteria[47].

Referring to section 6.1.1 the main concern is what constitutes a valid response point on the dose-response curve, in other words the values on the y-axis? As discussed in chapter1 there are two main methods of assay detection: continuous and stopped (endpoint). If MUANA is used as a substrate, the output can be measured in real time. The versatility of the IC_{50} is what makes it universal. You can choose what your response point is given it is consistent[108].

Thus, the screening for all inhibitors was conducted using both MUANA and 3'-sialyllactose. For MUANA, the initial velocity or linear region portion of the progress curve served as the

corresponding point on the y-axis. While for the 3'-sialyllactose the amount of Neu5Ac (product) produced after one hour served as the data point.

The result of the IC_{50} screening of Oseltamivir with MUANA is in Figure 6.6.

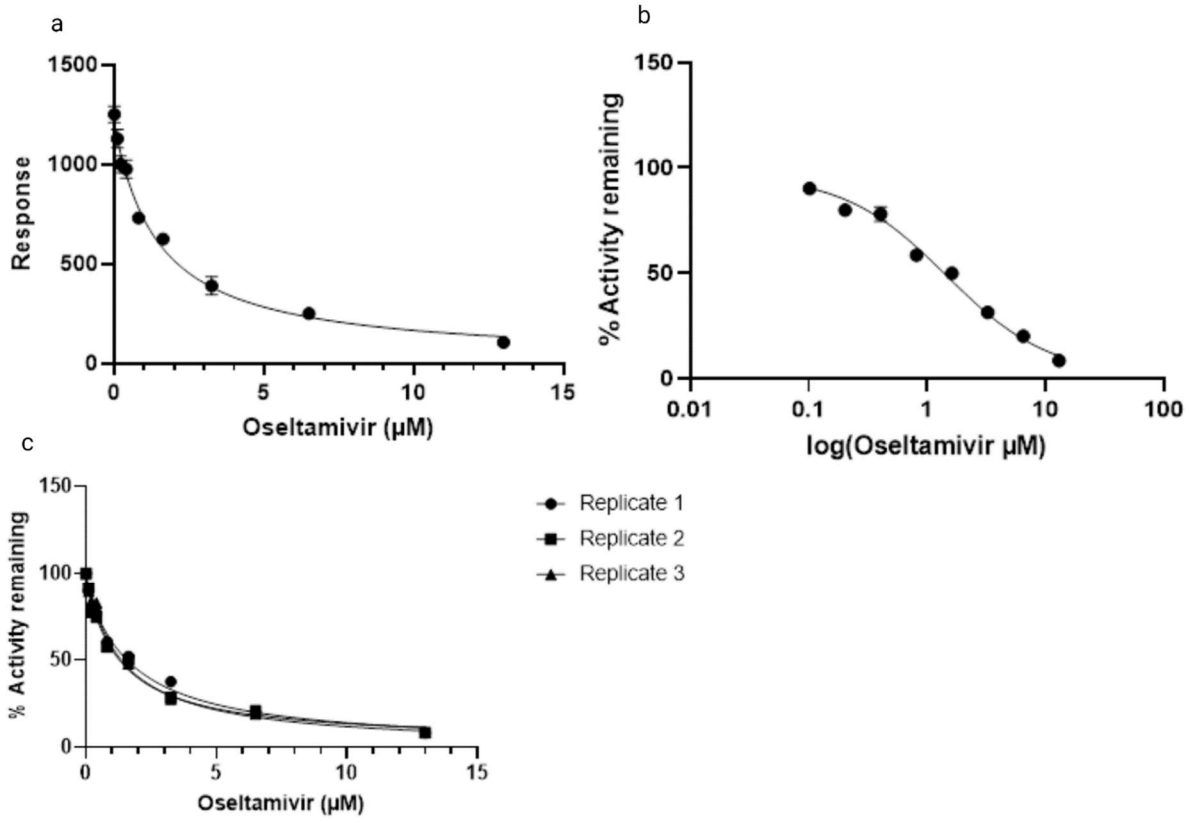


Figure 6.6 Response curves for oseltamivir carboxylate for NanA with 2'-(4-Methylumbelliferyl)-α-D-N-acetylneuraminic acid as substrate plotted in three alternative ways. (a) As a direct response (b) % activity remaining on a log scale and (c) % activity remaining.

Figure 6.6 a plots the response directly as a function of inhibitor concentration and will correspond to the relative IC_{50} /EC_{50} as the maximum inhibitor concentration far exceeds the theorized value (10x higher) and will act as the 0% correspondence. Whereas Figure 6.6 c the response is converted into a percentage activity remaining and is relative to the control plate where no inhibitor is used vs the maximum amount. The multiple data points represent replicates

(triplicates). Figure 6.6 b is probably the most common way to portray $IC_{50.}$ The percentage activity is plotted as a function of inhibitor concentration on a logarithmic scale.

While differences are subtle, it does influence the value of the calculated IC_{50} done with non-linear regression using the model as described in Equation 6.5. A summary of the IC_{50} of the three methods are Table 6.1.

Table 6.1 IC_{50} values for the three methods using NanA with 2'-(4-methylumbelliferyl)-α-D-N-acetylneuraminic acid.

	a	b		c	
			1	2	3
IC_{50} (μM)	1.51 ± 0.45	1.64 ± 0.72	2.20 ± 1.21	1.516 ± 0.77	1.56 ± 0.66

As can be seen, the three methods reflect different IC_{50} values. This demonstrates the validity of the absolute vs relative IC_{50} concept discussed in section 6.1.1, as Table 6.1 a and c represent the relative approach whereas b the absolute.

From Table 6.1 the IC_{50} value is within the range of 1.51 to 2.64 μM. Based on early principles of 'hit discovery', a molecule is considered potent when it inhibits an enzyme with a IC_{50} value of less than 10 μM. This makes OC a potential drug candidate for the inhibition of pneumococcal sialidase NanA.

The same concept was applied when using 3'-sialyllactose as the substrate. After one hour incubation the amount of sialic acid released is calculated using methods as described in detail in chapter 5. This is used as the corresponding data point on the y-axis. It is important to reiterate that the substrate concentration used in the IC_{50} screening is kept as close to the K_m value as possible. Figure 6.7 summarizes the screening results of the natural substrate.

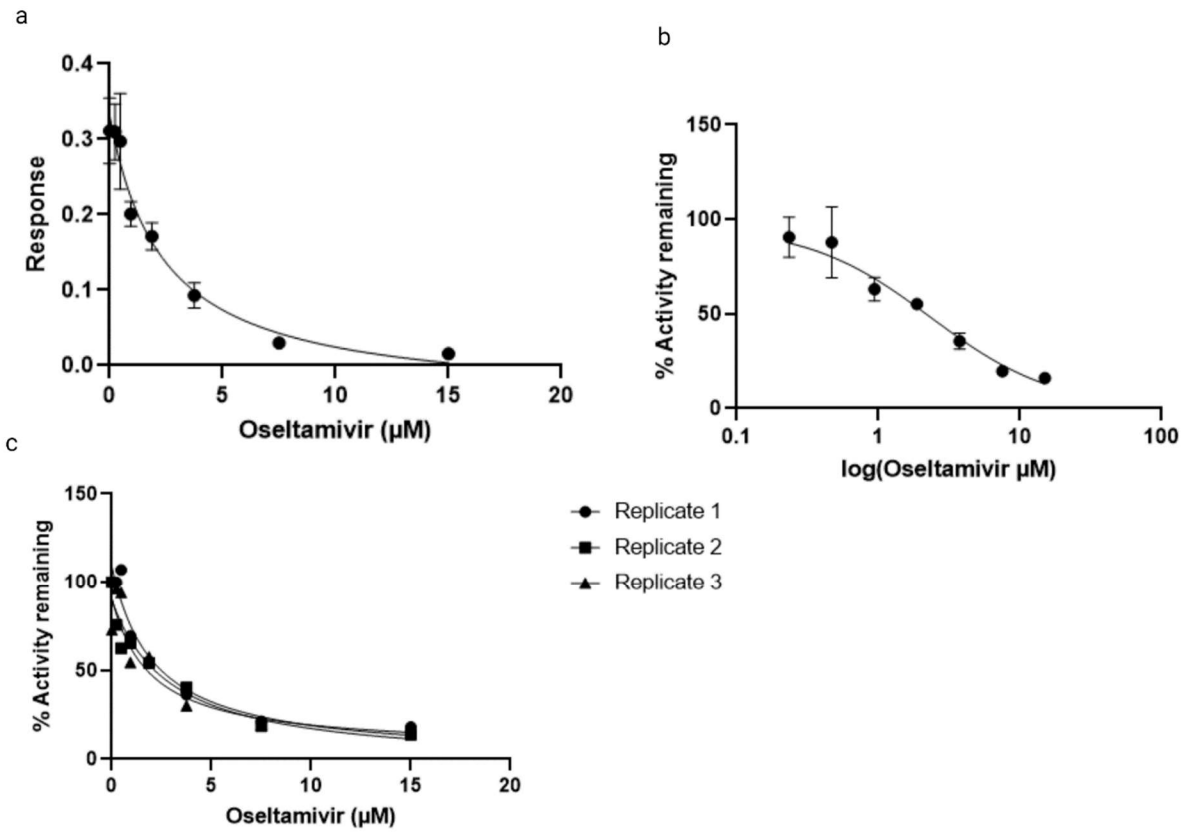


Figure 6.7 Response curve of oseltamivir carboxylate for NanA with 3'-sialyllactose as substrate plotted in three alternative ways. (a) As a direct response (b) % activity remaining on a log scale and (c) % activity remaining.

Figure 6.7 represents the three alternative methods to plot the IC_{50} screening results for the natural substrate. Again, the limitation is the fact that product formation can only be detected after an incubation period of one hour, whereas for the continuous assay (MUANA) any point on the linear region can be isolated and compared at any arbitrary time point. The results of the IC_{50} screening for 3'-sialyllactose is summarized in Table 6.2.

Table 6.2 **IC_{50}** values for the three methods using NanA with 3'-sialyllactose.

	a	b		c	
			1	2	3
IC_{50} (μM)	3.15 ± 1.95	4.32 ± 3.14	2.19 ± 1.31	3.98 ± 1.21	2.24 ± 0.91

Table 6.2 shows that the IC_{50} values range between 2.19 to 4.32 μM. Although it still fulfils the criteria to be a potential inhibitor, the uncertainties are much larger than those in Table 6.1 (MUANA substrate). This is due to the well-to-well variation and the detection method not being as sensitive as fluorescence. Again, the product output cannot be monitored in real time, and it is assumed all substrate converts into product within one hour.

Although the two methods differ in principle, the IC_{50} values are still below 10 μM, meaning that either screening would classify OC as a potential potent inhibitor.

6.3.2 Peramivir

Peramivir is a sialic acid analogue and antiviral agent developed by Biocryst Pharmaceuticals. It is a selective inhibitor to the neuraminidase enzymes produced by influenza A/B viruses. The development of this drug has been supported by the US department of health as it prevents new viruses emerging from infected cells. Unfortunately, it has poor oral bioavailability and is injected intravenously as set by the FDA in September 2017[109].

Influenza is a respiratory tract disease that infects 10% of the population annually. It can lead to serious health complications in individuals with chronic pulmonary disease as they have an increased probability of developing acute respiratory failure. Peramivir shows potent antiviral activity as well as a strong affinity towards the viral neuraminidase. Thus, great emphasis has been placed on studying the efficiency of this inhibitor against influenza, with IC_{50} values ranging between 1.11 and 2.81 nM. The structure is shown in Figure 6.8.

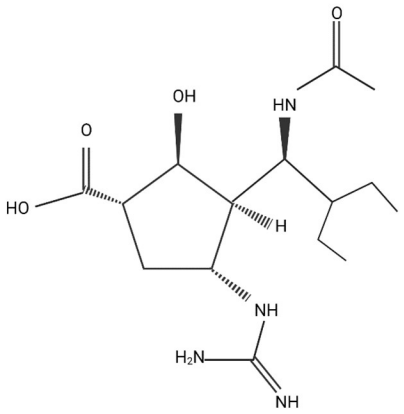


Figure 6.8 2D structure of Peramivir.

This inhibitor has been tested against multiple influenza strains, but not against pneumococcal sialidase. To evaluate if it will potentially serve as an inhibitor, an IC_{50} screening was performed. No value has been reported so far, so to cover an adequate range, the first batch had eight different inhibitor concentrations ranging from 20µM- 1333µM. This is to ensure that the high concentration regions are covered. Should a plateau region form after plotting (horizontal line) then the range will shift to a lower concentration. This is only because there is no reference point for the hypothetical $IC_{50.}$

The first screening with MUANA yielded the results plotted in Figure 6.9.

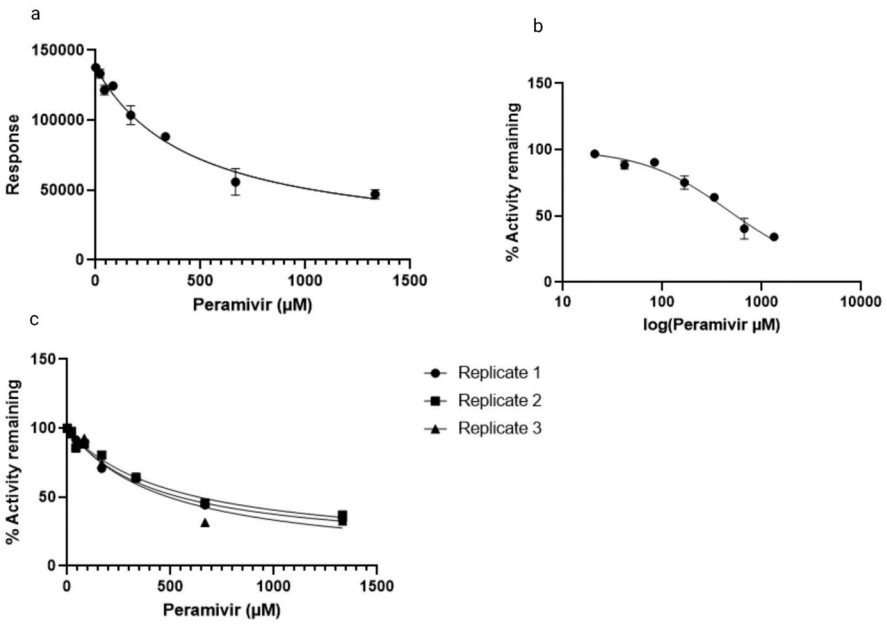


Figure 6.9 Response curve of Peramivir for NanA with 2'-(4-Methylumbelliferyl)-α-D-N-acetylneuraminic acid as a substrate plotted in three alternative ways. (a) As a direct response (b) % activity remaining on a log scale and (c) % activity remaining.

From the initial screening it seems that although effective for influenza A/B,it is not as potent for pneumococcal sialidase. The experiment is repeated for 3'-sialyllactose, and the results shown in Figure 6.10:

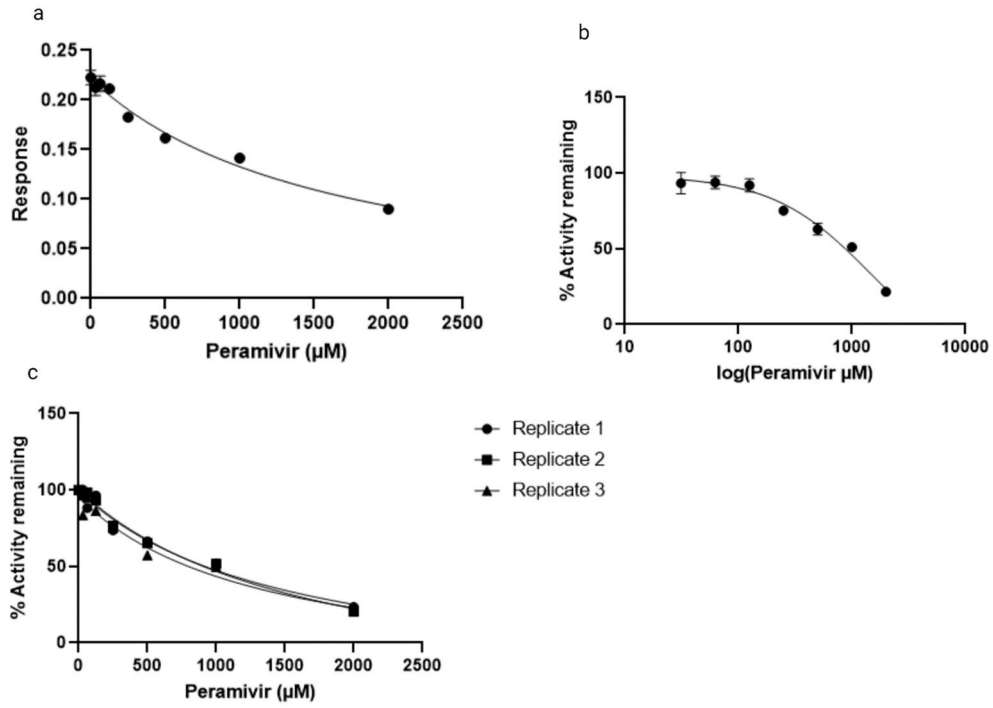


Figure 6.10 Response curve of Peramivir for NanA with 3'-sialyllactose as substrate plotted in three alternative ways. (a) As a direct response (b) % activity remaining on a log scale and (c) % activity remaining.

Figure 6.10 b shows that the concentration range chosen (0-2000μM) is adequate to witness both bend points in the dose-response, albeit not as clear as in Figure 6.9 b. The IC_{50} values for both substrates and methods are summarized in Table 6.3.

Table 6.3 IC_{50} values for the three methods using NanA with 3'-sialyllactose and 2'-(4-methylumbelliferyl)-α-D-N-acetylneuraminic acid for Peramivir.

	(a)	(b)	(c)		
IC_{50} (μM)			1	2	3
MUANA	462.21± 80.12	461.23 ± 77.53	417.78 ± 30.56	508.13 ± 37.96	465.24 ± 44.58
3SL	1644.01 ± 140.21	1394.51 ± 130.13	1500.98 ± 150.67	1686.78 ± 230.13	1312.57 ± 245.21

From Table 6.3 it is evident that the IC_{50} value is much different depending on the method used. The IC_{50} on average for 3'-sialyllactose is almost three times as high as when 2'-(4-METHYLUMBELLIFERYL)-A-D-N-ACETYLNEURAMINIC ACID is used. This is because for the 3'-sialyllactose, the bend point when approaching zero is not adequately defined. It is also assumed that all product formed is due to the 'active' enzyme that has not been inhibited, which is not always true. This is where endpoint fails and does not always serve as a fair comparison. The sensitivity of this assay is also much lower than for fluorescence, hence the uncertainties are so much higher.

Even though the IC_{50} values differ, Peramivir is not a potential inhibitor for pneumococcal sialidase. This is because even the lowest IC_{50} value is in the range of 400-500 µM, almost 50 times higher than the 'cut-off' criteria.

6.3.3 N-acetyl-2,3-dehydro-2-deoxyneuraminic acid

N-acetyl-2,3-dehydro-2-deoxyneuraminic acid (DANA) is a molecule produced by NanC and is classified as a broadband neuraminidase inhibitor as it suppresses both in vitro and in vivo effects[47, 97]. The structure is shown in Figure 6.11:

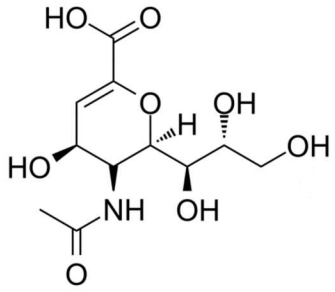


Figure 6. 11 Structure of N-acetyl-2,3-dehydro-2-deoxyneuraminic acid.

The reported K_i for DANA for influenza is extremely low, ranging between 1.5 and 1.8 µM[110]. As discussed, K_i and IC_{50} values are only comparable if the mechanism of inhibition is known. Since the substrate concentration used was kept close to the K_m value, the approximation of $\frac{K_m}{[S]}$ in equation 6.7 for competitive inhibition approaches 1. The same is true for the $\frac{[S]}{K_M}$ term in equation 6.8 for uncompetitive inhibition. That means an approximation for both modes can be made to determine the $IC_{50,\ given}$ the K_i is 1.8 µM. The approximate IC_{50} value is calculated in Equation 6.9-6.11:

$$IC50 = K_i\left(1 + \frac{[S]}{K_M} or \frac{K_M}{[S]}\right) \quad [6.9]$$

$$IC50 = K_i(1 + 1) \quad [6.10]$$

$$IC50 = 1.8(2) \sim 3.6 \mu M \quad [6.11]$$

This is still an extremely low IC_{50} value and indicates potential as an inhibitor, however, has largely been used to test the potency against influenza. For both experiments inhibitor concentrations ranged from 0.2 µM (minimum)to 16 µM (maximum). The results of the screening for MUANA and 3'-sialyllactose are plotted in Figures 6.12 and 6.13 respectively.

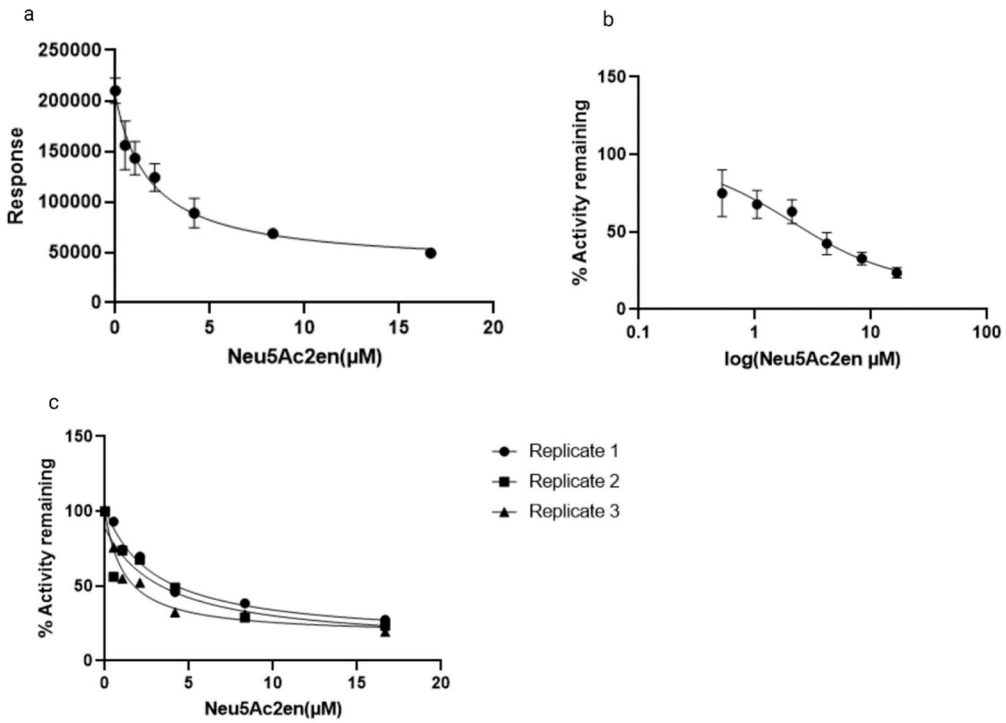


Figure 6.12 Response (dose) curves of N-acetyl-2,3-dehydro-2-Deoxyneuraminic Acid for NanA with MUANA as substrate. (a) As a direct response (b) % activity remaining on a log scale and (c) % activity remaining.

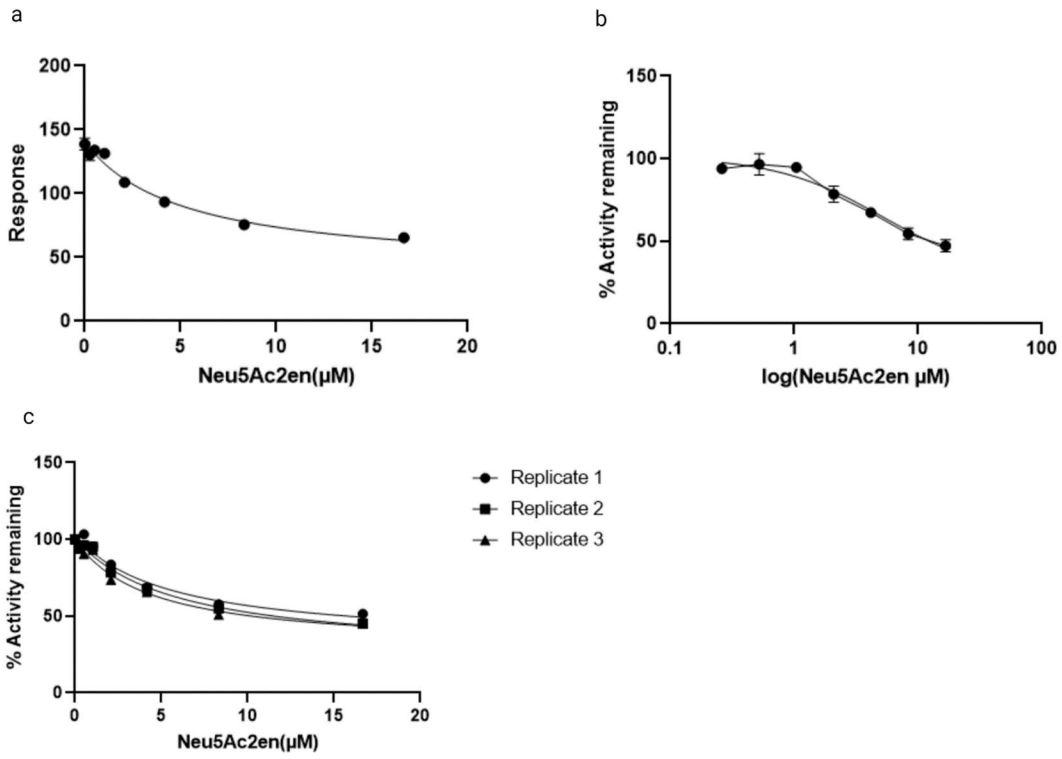


Figure 6.13 Response curves of N-acetyl-2,3-dehydro-2-Deoxyneuraminic Acid for NanA with 3-siallylactose as substrate. (a) As a direct response (b) % activity remaining on a log scale and (c) % activity remaining.

The results of both screenings are summarized in Table 6.4.

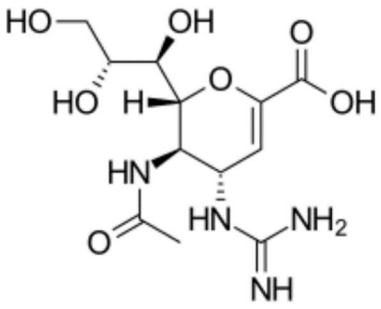
Table 6.4 **IC_{50}** values for the three methods using NanA with 3'-sialyllactose and 2'-(4-methylumbelliferyl)-α-D-N-acetylneuraminic acid for DANA.

	(a)	(b)	(c)		
IC_{50} (μM)			1	2	3
MUANA	1.88 ± 0.59	2.10 ± 1.27	2.95 ± 0.99	3.47 ± 1.32	1.53 ± 0.87
3SL	5.16 ± 2.65	5.17 ± 1.99	5.506 ± 2.76	5.82 ± 2.34	4.32 ± 2.21

From Table 6.4 there is a discrepancy between the IC_{50} values of the two methods. The results obtained from MUANA are much lower than for 3SL (almost 2-fold). Again, there is no way to monitor product formation in real time for the TBA assay, hence it is much more accurate to compare points on the progress curve. Although different, both screenings still do not exceed the cut-off criteria of 10µM. This implies that DANA may serve as a potential inhibitor for sialidase NanA.

6.3.4 Zanamivir

Zanamivir prevents the growth and spread of the influenza virus and shortens the duration of flu symptoms[47]. It has an extremely low Ki for influenza with a reported value of around 1.8 nM. For pneumococcal sialidase A, it has a reported Ki value of around 0.72 mM (~720µM). The structure is shown in Figure 6.14 and screening results for MUANA and 3'-sialyllactose are shown in Figure 6.15 and Figure 6.16 respectively. Table 6.5 summarizes the inhibition constants for both substrates.



.5

Figure 6.14 Structure of Zanamivir.

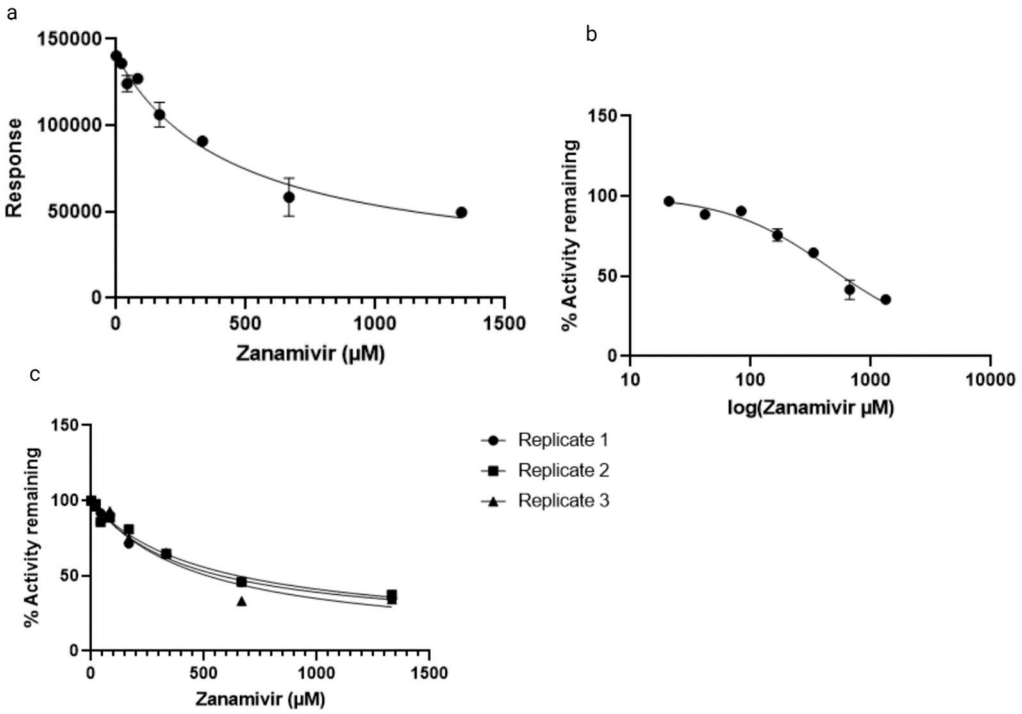


Figure 6.15 Response curve of Zanamivir for NanA with 2'-(4-Methylumbelliferyl)-α-D-N-acetylneuraminic acid as substrate. (a) As a direct response (b) % activity remaining on a log scale and (c) % activity remaining.

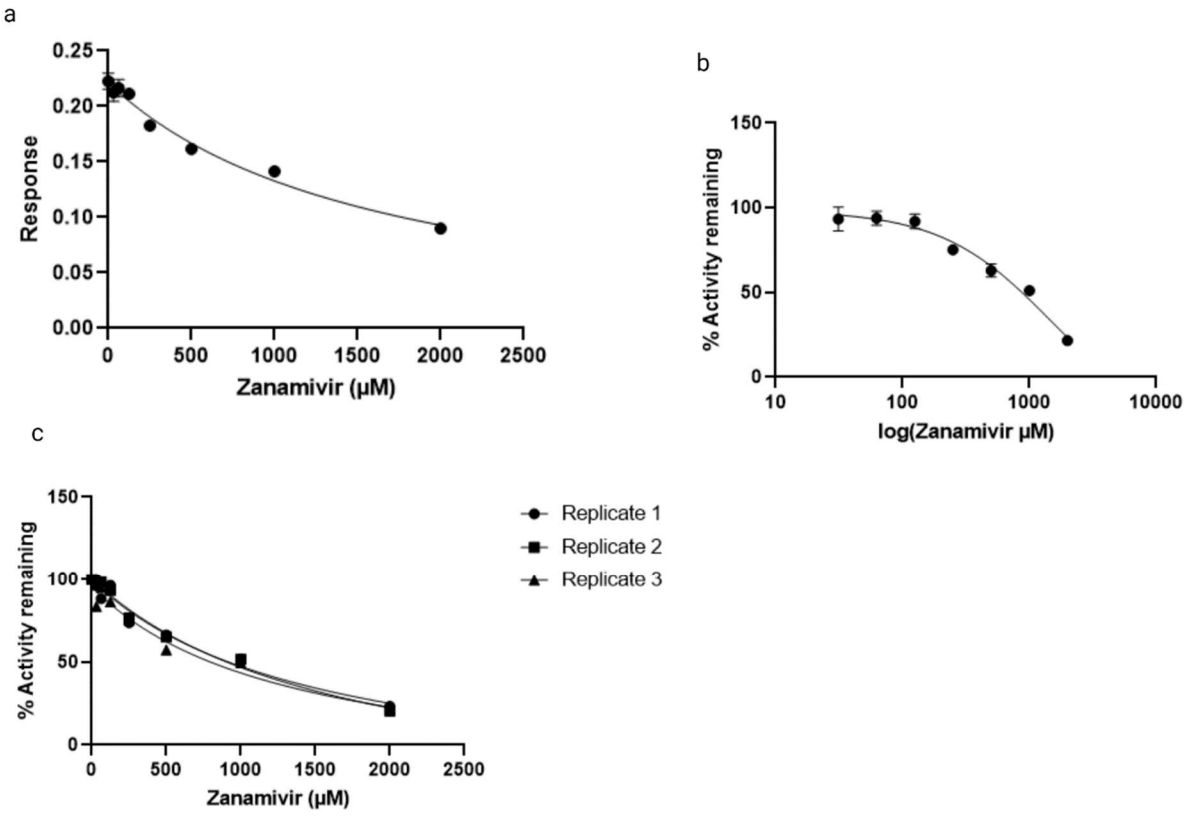


Figure 6.16 Response curve of Zanamivir for NanA with 3'-sialyllactose as substrate. (a) As a direct response (b) % activity remaining on a log scale and (c) % activity remaining.

Table 6.5 IC_{50} values for the three methods using NanA with 3'-sialyllactose and 2'-(4-METHYLUMBELLIFERYL)-A-D-N-ACETYLNEURAMINIC ACID for DANA.

	(a)	(b)	(c)		
IC_{50} (µM)			1	2	3
MUANA	462.12 ± 98.12	465.15 ± 88.98	417.12 ± 80.12	508.76 ± 102.45	465.34 ± 90.45
3SL	1200.10 ± 213	1345.15 ± 198.12	1110.98 ± 140.56	1560.95 ± 197.12	1237.45 ± 187.67

Again, from Table 6.5 there is a large discrepancy between the MUANA and 3'-sialyllactose inhibition screening. Both methods indicate that Zanamivir is not a suitable inhibitor for

pneumococcal sialidase as IC_{50} values all exceed 500 µM, despite being an ideal candidate for influenza.

6.4 Summary of Inhibition Constants

For all four commercial inhibitors, two methods were used to determine the IC_{50}. It should be reiterated that IC_{50} provides no information on the mechanism of inhibition, substrate concentration or Michaelis constant unless specified. IC_{50} is an excellent quick screening tool to separate potential drug candidates from non-candidates. The absolute IC_{50} was chosen for comparison across all inhibitors and substrates. A summary of all inhibition constants is shown in Table 6.6.

Table 6.6 Summary of inhibition constants for all four commercial inhibitors.

	Commercial inhibitor	IC_{50} MUANA	IC_{50} (3'-sialyllactose)
Efficacy Increases 	Zanamivir	465.45 ± 88.98	1345.15 ± 198.12
	Peramivir	461.23 ± 77.53	1394.51 ± 130.13
	DANA	2.10 ± 1.27	5.17 ± 1.99
	Oseltamivir	1.64 ± 0.72	4.32 ± 3.14

The screening results from both MUANA and 3'-sialyllactose assays agree on the effectiveness of the inhibitors. Although Zanamivir and Peramivir are excellent inhibitors for influenza A/B, it fails to show the same trend for pneumococcal NanA. DANA and Oseltamivir show excellent potential for inhibition as both IC_{50} values do not exceed 10 µM. OC is chosen for further investigation of the mechanism of inhibition and how it compares to *in silico* methods, further analyzed in chapter 8.

6.5 In Depth: Oseltamivir Mode of Inhibition

As confirmed by the literature and results from the previous section, Oseltamivir is one of the best drug candidates as it shows great inhibition against influenza. Thus, to confirm it the potency it was decided to investigate the mechanism of inhibition and determine the K_i for this drug. The reported value is approximately 1.7 µM[47].

6.5.1 Methodology (Protocol)

From Figure 6.2 a good initial step is to plot the Lineweaver-Burke to see how the lines intersect. The most prevalent challenge is choosing the appropriate substrate range. Thus, the protocol for the Lineweaver-Burke plot is:

1. Choose an appropriate substrate range so that the linear region can be isolated for all curves. After some preliminary screenings, the substrate range was chosen to be 2 to 133 µM.
2. For a Lineweaver-Burke, the substrate and inhibitor concentration should be varied. Thus, to ensure that the K_i is accurately determined, the inhibition range was 1 to 6.5µM.
3. Three experiments are performed in triplicate, one experiment for each inhibitor concentration (1.6 µM, 3.3 µM, 6.6 µM). An example of the plate layout is in Table 6.7.
4. Each well contains a set inhibitor concentration per experiment whereby the substrate is added and incubated for 30 minutes at 37℃.
5. NanA (5nM 1x) is added to each well to initiate the reaction and the progress curves generated.
6. After saturation, each progress curve was isolated and the linear region used to calculate the slope, which represents one data point (the initial velocity).
7. Plot the reciprocal of the initial velocity as a function of substrate concentration.

Table 6.7 Example of plate layout for inhibition experiment.

	1	2	3	4	5
A	3x concentration MUANA (µM)	3x concentration MUANA (µM)	3x concentration MUANA (µM)	3x Inhibitor concentration (e.g. 6.6 µM)	3x NanA (nM)
H					

6.5.2 Results and Discussion

Following the protocol from section 6.5.1, the Lineweaver-Burke is shown in Figure 6.17:

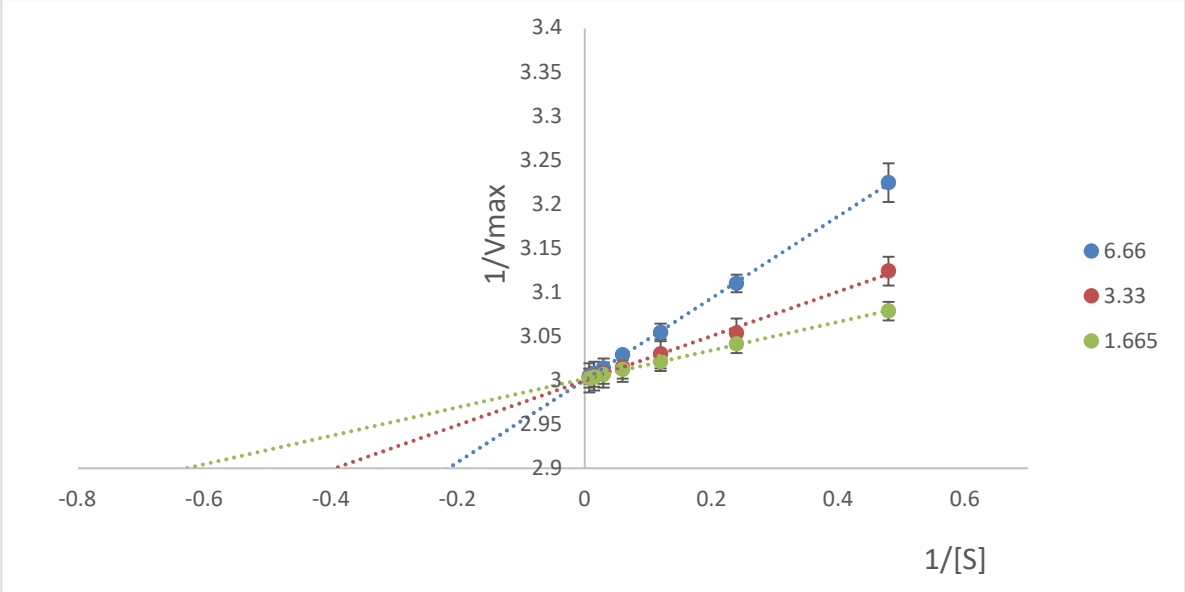


Figure 6.17 Lineweaver-Burke plot of NanA with 2'-(4-Methylumbelliferyl)-α-D-N-acetylneuraminic acid with Oseltamivir as inhibitor. The inhibitor concentrations are shown in µM. The x-axis is scaled with a factor of 10^3.

From Figure 6.17 the lines intercept on the x-axis. From Figure 6.2 the V_{max} is unchanged and the K_m increased, this appears to be competitive inhibition. Alternatively, to confirm this a Hanes plot can also be used. If the lines (slopes) are parallel, then it is a sign of competitive inhibition.

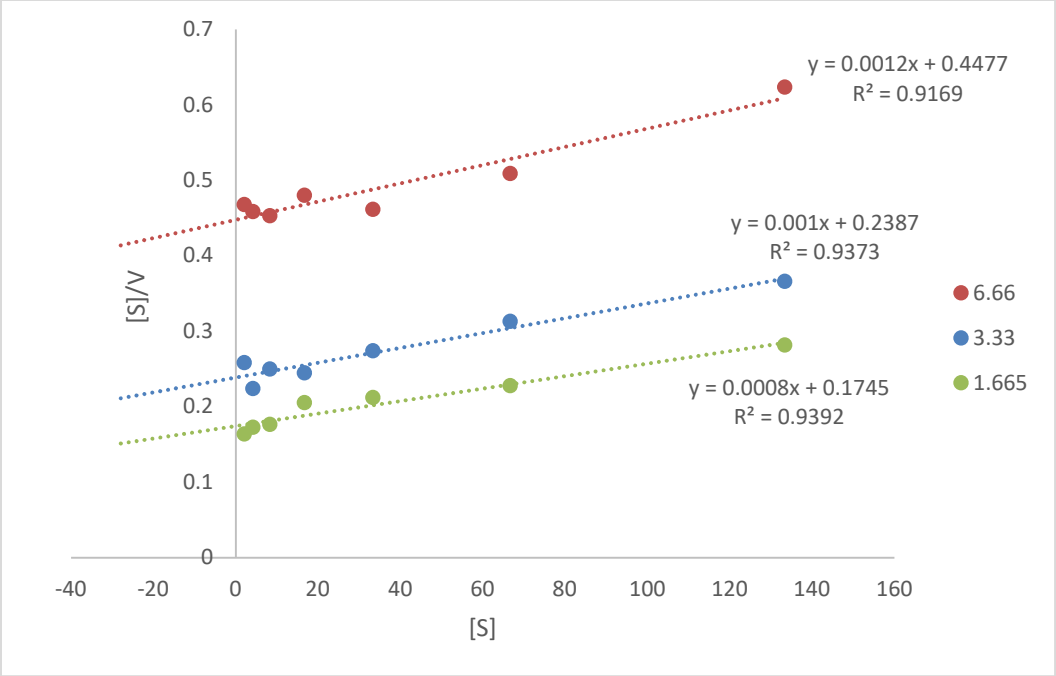


Figure 6.18 The Hanes plot of NanA with 2'-(4-Methylumbelliferyl)-α-D-N-acetylneuraminic acid and Oseltamivir as inhibitor. The inhibitor concentrations are shown in µM.

From Figure 6.18 the lines (slopes) are parallel to each other. This further confirms that it is competitive inhibition taking place for Oseltamivir. There are multiple methods to extrapolate the inhibition constant, K_i from experimental data. The easiest is to use a Dixon plot. This is where the inhibitor concentration is varied as a function of the substrate. For competitive inhibition the slopes should intersect above the x-axis, and this value corresponds to the K_i.

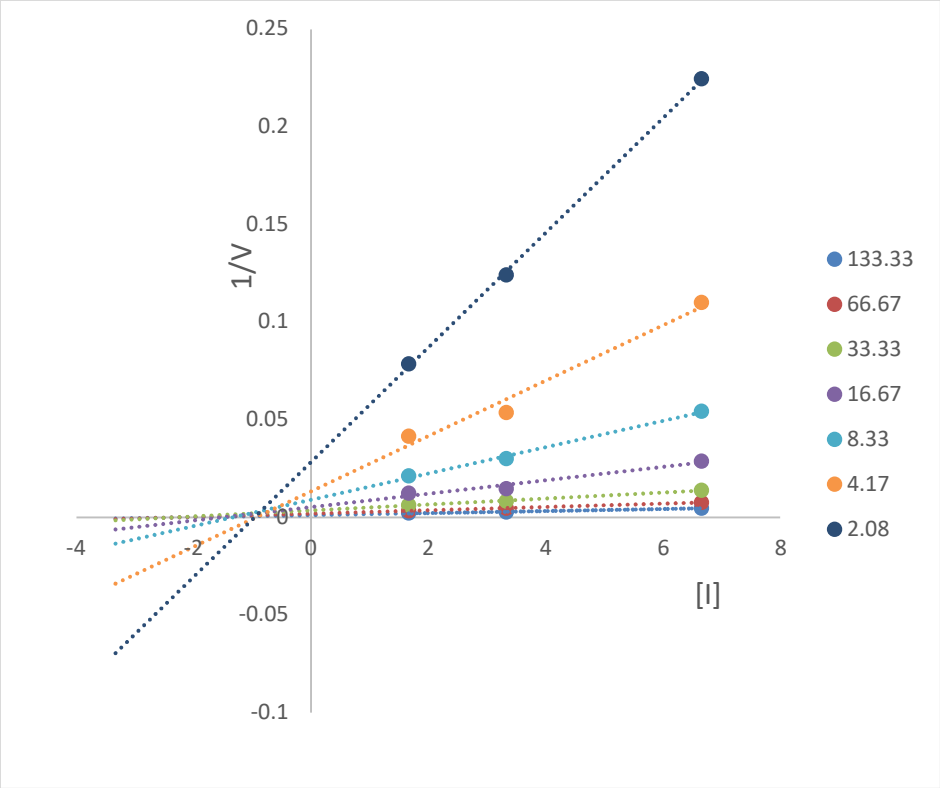


Figure 6.19 The Dixon plot of NanA with 2'-(4-Methylumbelliferyl)-α-D-N-acetylneuraminic acid and Oseltamivir. The inhibitor concentrations are shown in µM.

It was then calculated where all lines intercept above the x-axis. All linear lines can be expressed with $y = mx + c$. Since they all have the same y coordinates, each line can be equated to each other to solve the x coordinate. The result yielded a $-K_i$ of -1.19 ± 0.09 µM. This compares extremely well with the literature as the K_i is below 2µM. This not only proves that Oseltamivir is a potent inhibitor for pneumococcal neuraminidase, but also the mode of inhibition that occurs.

Section 2

Chapter 7

Computational enzymology

7.1 Modelling Starting Structures.

The accuracy of biomolecular simulations is dependent on the choice of starting structure as all topology and parameter files containing the active site are modelled from this. Most proteins (enzymes) are formatted in high-resolution crystal structures found on the Protein Data Bank (PDB). It is important that the substrate is in the correct position in the active site to best represent the Michalis complex[111]. However, most PDB structures do not contain the protein with the substrate as enzyme catalysis is a rapid process and the enzyme-substrate complex cannot be isolated. One of the two approaches used to obtain the correct binding position is known as site directed mutagenesis. Changes are made to a DNA sequence of a gene involved in the enzyme production, inhibiting the reaction from proceeding. The substrate is now in a static position and can be structurally isolated. Many of these mutated residue containing proteins are found in the PDB database, with the prerequisite being that these residues need to be mutated back before launching any simulation. Alternatively, the substrate can be mutated to identify the active site and binding poses[112].

However, it is still challenging to model these enzymes as the binding poses of most substrates are unknown. A common approach is to find a structurally similar substrate in the catalytic site of the enzyme and overlap with the desired substrate. The latter is used in our simulation as the protein (NanA) does not contain MUANA or 3'-sialyllactose in the active site. The starting structures are modelled using alignment of the crystal structure of the Michaelis complex of NanA

(PDB ID 2vvz) with DAN. The MUANA substrate is from the complex of Trypanosoma *cruzi* trans-sialidase (PDB ID 1S0J).

Alternatively, one can use molecular docking to model the Michaelis complexes to determine the correct substrate position. This allows for large datasets of molecules to be screened against the same active site as to determine to best scores leading to potential inhibitors. The docking process samples several conformations of the substrate/inhibitor in the active site to determine the best binding pose set by a predefined scoring function. However, most docking simulations are performed in the gas phase rendering the enzyme in a static or rigid mode. Thus, it does not account for the conformational changes or the effects from the solvent[113]. Assigning protonation states of titratable amino acids is a critical component as crystal structures do not include this information.

Computational modelling is initiated after finding a suitable starting structure to then undergo energy minimizations needed to calculate the free energy of the reaction (Figure 7.1).

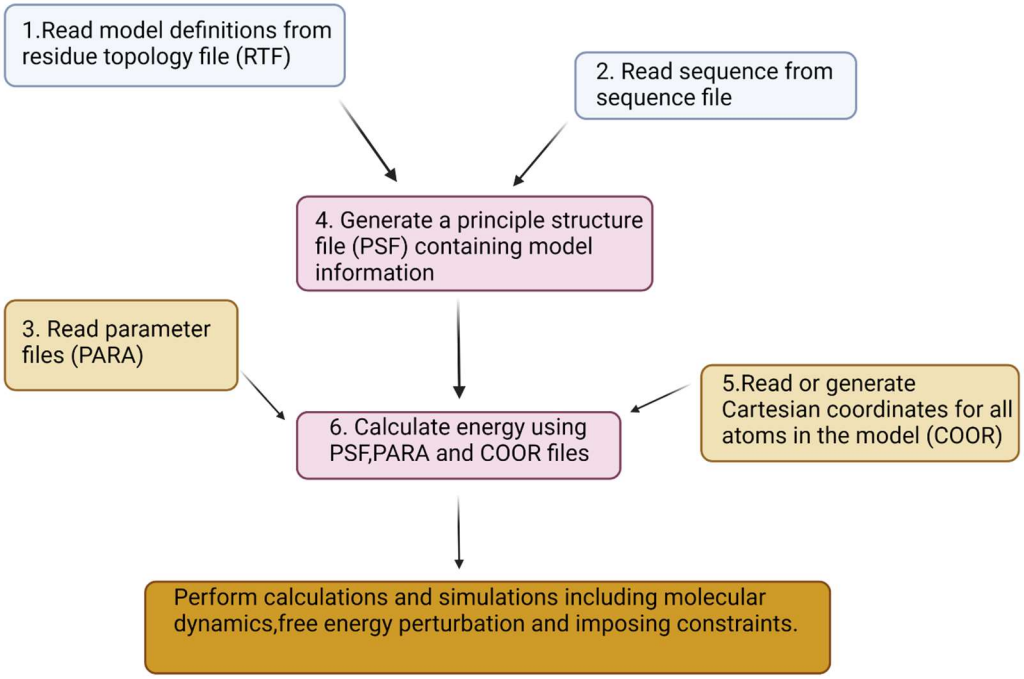


Figure 7.1 Activity flow to perform free energy simulations and molecular dynamics from a starting structure.

The residue topology file (RTF) includes information about the residues such as atoms, bonds, bond angle, torsion angles and hydrogen bond acceptors or donors. The sequence file must then be supplied prior to the simulation. However, after generating a starting structure the energy can only be evaluated if parameters exist for all internal, external, and special energy terms. Thus, the parameter (PARA) file specifies equilibrium geometries, van der Waals radii and data necessary to calculate these energies. The principal structure file (PSF) contains information on the entire structure, consisting of a hierarchy where atoms are placed into groups that will eventually form segments. The coordinate (COOR) file contains the relative positions of atoms within a structure. All these files need to be properly generated to serve as input for a potential energy evaluation of a macromolecular system.

7.2 Computational Enzyme Simulation

After modelling the starting structure, computational simulation methods can be used to obtain details about the structure and function of this system. The interest lies in simulating a complex enzymatic system; thus, this section aims to give a brief theoretical overview of some of the methods used.

Form the CHARMM workflow it is recommended to minimize the potential energy of the system to remove faulty and overlapping atom positions. There are different methods and algorithms available in the CHARMM source code of which the Newton Raphson was selected. It is a powerful technique based on linear approximation that is used for solving equations numerically and uses gradients to optimize the Hessian matrix[114].

Minimization involves introducing one variable at a time until the function reaches a minimum. However, most variables are not independent, and several cycles are required to achieve a minimum. Thus, many computational chemistry methods assume that the gradient $\boldsymbol{g}$ and the Hessian $\boldsymbol{H}$ can be calculated analytically. To achieve convergence either the gradient must be below a chosen cutoff value or the change between two iterations needs to be miniscule.

The Newton Raphson method expands the Taylor function to second order as in Equation 7.1:

$$f(x) \approx f(x_0) + g^t(x - x_0) + \frac{1}{2}(x - x_0)^t H(x - x_0) \quad [7.1]$$

Where x_0 represents the current point. The gradient in Equation 7.1 is required to be zero to produce Equation 7.2:

$$(x - x_0) = \boldsymbol{H}^{-1} g \quad [7.2]$$

The Newton Raphson can then be written as Equation 7.3:

$$\Delta x' = (\Delta x_1', \Delta x_2', \Delta x_3' \ldots \Delta x_N') \quad [7.3]$$

$$\Delta \Delta x_i' = -\frac{f_i}{\varepsilon_i} \quad [7.4]$$

The assumption is that in the coordinate system x' the Hessian is diagonal. From Equation 7.4, f_i is the projection of the gradient with a corresponding eigenvalue ε_i. All Hessian eigenvalues are positive near a minimum and the step direction is opposite to that of the gradient. Thus, should the Hessian eigenvalue be negative the step will be in the same direction as the gradient. The main advantage is that convergence is second order with respect to a stationary point[114]. However, the function will reduce quadratically with a factor of two between any two iterations should it contain higher order terms. The Newton Raphson is often used with the conjugate gradient method to minimize starting structures[115].

7.3 Molecular Dynamics

Molecular dynamics (MD) methods solve Newton's equation of motion for atoms. The total energy for each molecule is distributed between the potential and kinetic energy. The molecules can overcome this energy barrier if the height does not exceed the difference between the total

and potential energy. If the energy is sufficient, the dynamics will sample the whole surface[116, 117]. Unfortunately, this often requires a long and impractical solution time as Newton's equation requires a short time variable. Additionally, very high temperatures are required to overcome the energy barriers, which cannot currently be simulated. Thus, only small energy barriers can be overcome. MD is reliant on the inherent dynamics of the system to locate low-energy modes which can be used to sample the conformational space of a larger system. These MD methods are mainly used when the starting structure's geometry is based on experimental information e.g. crystal structures[113].

Langrangian mechanics is a formulation of classical mechanics used to initiate MD[116] and is defined in Equation 7.5:

$$L = T - V \qquad [7.5]$$

T represents the energy of the system's motion i.e. the kinetic energy. ***V*** is the potential energy and reflects the energy of interaction between particles in the system. The Langrangian is time-dependent due to external influences and time varying constraints (Equation 7.6):

$$L(q, \dot{q}) = T(\dot{q}) - V(q) \qquad [7.6]$$

Where $\boldsymbol{q}$ and $\boldsymbol{\dot{q}}$ represent the generalised coordinates and time derivative of the generalised coordinate. Equation 7.6 can be converted to the Hamiltonian H(*p*,*q*) and is given in Equation 7.7:

$$H(p, q) = \sum_{i}^{N} q_1 p_1 - L(q, \dot{q}) \qquad [7.7]$$

N represents the number of particles and ***p*** the momentum. The potential energy is independent of the velocity (and momentum) of a system, thus only dependant on the energy q. Equation 7.8 is used:

$$H(p,q) = \sum_{i=1}^{N} \frac{{p_i}^2}{2m_i} + U(q_1, q_2, \ldots . q_N) \qquad [\,7.8\,]$$

The Langrangian numerically solves Newton's equation of motion[118]. The time-evaluation of the cartesian coordinates can then be expressed in Equation 7.9:

$$F_i = m\ddot{r} = \frac{\partial}{\partial r_i} U \qquad [\,7.9\,]$$

$\boldsymbol{F_i}$ represents the force of the particle, ***m*** the mass and ***r*** the position vector on the cartesian coordinates. MD simulations calculate the potential energy and force, where the latter updates the atom position at each timestep and generates a trajectory[111]. The verlet algorithm is one such method used in the CHARMM source code to generate traceries[119]. Consider a set of particles with a position vector $\boldsymbol{r_i}$ at a given time $\boldsymbol{t_i}$. The positions can be calculated from the velocities and accelerations at a small-time increment $\boldsymbol{\Delta t}$ and corresponds to the Taylor expansion where time is the variable (Equation 7.10).

$$r_{i+1} = r_i + v_i(\Delta t) + \frac{1}{2} a_i(\Delta t)^2 + \frac{1}{6} a_i(\Delta t)^3 + \cdots \qquad [\,7.10\,]$$

The positions from the previous time step is defined in Equation 7.11:

$$r_{i+1} = r_i - v_i(\Delta t) + \frac{1}{2} a_i(\Delta t)^2 - \frac{1}{6} a_i(\Delta t)^3 + \cdots \qquad [\,7.11\,]$$

The positions can be predicted for the next time step by adding the current and previous positions. The verlet algorithm is presented in Equation 7.12:

$$r_{i+1} = (2r_i - r_{i-1}) + a_i(\Delta t)^2 + \cdots \quad [7.12]$$

Where substituting $a = \frac{F}{m}$ yields Equation 7.13:

$$r_{i+1} = (2r_i - r_{i-1}) + \frac{F_i}{m}(\Delta t)^2 + \cdots \quad [7.13]$$

$$v_i = \frac{1}{2\Delta t}\{r_{i+1} - r_{i-1}\}$$

Thus, the current velocity ($\boldsymbol{v_i}$) and position ($\boldsymbol{r_i}$) can be calculated as well as the previous and next step of the simulation ($\boldsymbol{r_{i+1}}$ and $\boldsymbol{r_{i-1}}$). Alternatively, the leapfrog algorithm can be used where the velocities are defined for half a time step behind and in front of the current step.

7.3.1 Boundary Conditions

Several hundred solvent molecules are required to simulate a realistic system. The outer solvent molecules experience different forces than the molecules in bulk in a simulation box[120]. To account for this, the solvent molecules are placed in a box which is duplicated in all directions. An example for an enzymatic system is illustrated in Figure 7.2:

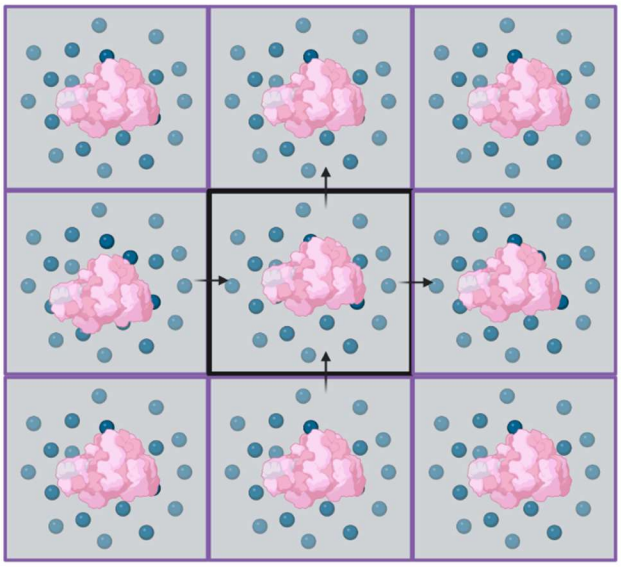


Figure 7.2 Illustration of periodic boundary conditions on an enzymatic system.

If a solvent molecule leaves the central box, the image will enter an adjacent box (Figure 7.2). The electrostatic interaction is long-range and will extend past the boundary of the box[121].

A simulation box is defined in three dimensions: nxL, nyL and nzL where ***n*** is position in the cell and ***L*** the length of the cube. The Coulomb potential of two points ***C*** and ***D*** can be given as:

:

$$U_{coloumb=} \frac{1}{2}\sum_{|n|}^{\infty}\sum_{C}^{N}\sum_{D}^{N}\frac{Q_C Q_D}{4\pi\epsilon|\mathbf{r_{CD}}+\mathbf{n}|} \qquad [7.14]$$

r_{CD} is the minimum distance between the two-point charges (***Qc*** and Q_D) and ϵ the dielectric constant. Convergence is accelerated by using Ewald's modification of the coulomb potential:

$$U_{coloumb=}U_{fourier} + U_{real} + U_{self} \qquad [7.15]$$

U_{real} is the real space summation, where an opposite charge is added to each point charge in a Gaussian function. The Gaussian is then recalculated in the reciprocal space, $U_{fourier}$. U_{self} is then used to cancel interactions between the introduced countercharges[121].

Stochastic boundary conditions (SBC) use elements of both Langevin and Newtonian dynamics but are computationally more expensive[114]. The SBC use water spheres where the sphere is centered around the active site of the enzyme. The system is divided into three regions and each region operates differently as portrayed in Figure 7.3:

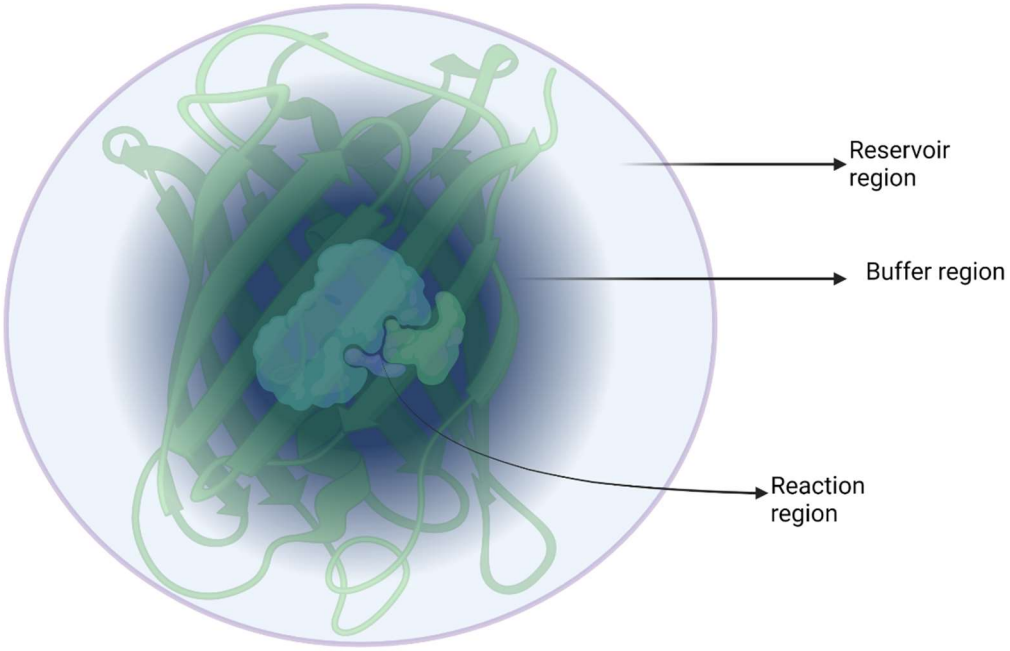


Figure 7.3 An illustration showing stochastic boundary conditions.

The size of the sphere is dependent on the system where the large reaction region is chosen to accommodate the active site of interest. MD simulations are performed using either Leapfrog or Verlet integrators. The buffer region is defined by the Langevin equations as it accommodates changes in energy, density, and structure in the reaction region[120, 122].

7.3.2 Force Fields

CHARMM utilizes a comprehensive and flexible empirical energy function and is the summation of many individual energy terms (Equation 7.16)[123].

$$
\begin{aligned}
&U(\vec{r}) \\
&= \sum_{Bond} k_b(r - r_0)^2 \\
&+ \sum_{Angle} k_\theta(\theta - \theta_0)^2 \sum_{Dihedral\ Angle} \left|k_\varphi\right| \\
&- k_\varphi \cos(n\varphi) \sum_{Improper\ torsions} k_\omega(\omega \\
&- \omega_0)^2 \sum_{Electrostatic} \frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}} \sum_{Van\ der\ Waals} \left(\frac{A_{ij}}{r_{ij}}\frac{B_{ij}}{r_{ij}}\right) sw\left({r_{ij}}^2, {r_{on}}^2, {r_{off}}^2\right)
\end{aligned}
\qquad [\ 7.\ 16\ ]
$$

Force fields are a collection of functions that give the overall potential energy needed for rapid MD simulations[124]. The first four terms in Equation 7.16 correspond to internal energy, whereas the last two represent external energy. *r* denotes the bond length, θ the valence angle,φ the dihedral angle and *n* the multiplicity of the dihedral. $k_b, k_\theta, k_\omega, k_\varphi$ all represent the force constants of bonds, angles, improper torsions, and dihedral angles[113, 125].

7.3.3 DFT and SCCDFTB

Self-consistent tight binding density function theory (SCCDFTB) is a second order expansion of the density functional theory (DFT) total energy with respect to charge density (Equation 7.17)[126].

$$\delta p = \sum_a \delta p_a \qquad [\ 7.\ 17\ ]$$

The charge density variation is represented by δp and is relative to the reference density which is a superposition of atomic contributions p_a. SCCDFTB is different to other tight binding methods as charge variations are introduced at the monopole level through a self-consistent treatment.

The DFTB is a semi-empirical method with three types: DFTB1, DFTB2 and DFTB3 where DFTB3 is the most accurate for molecules. There are several parameters for each set[127].

For a closed shell DFTB1 its energy is calculated as in Equation 7.18:

$$E^{DFTB1} = \sum_{i}^{N/2} \sum_{u}^{K} \sum_{v}^{K} c_{ui} c_{vi} H_{uv}{}^{0} + \sum_{a} \sum_{b>a} E_{repulsive}{}_{ab} \qquad [\ 7.\ 18\ ]$$

Where c_{ui} is the molecular orbital coefficient[128] , ***i*** with basis set ***u*** e.g., the Kohn-Sahm orbital coefficients[129]. Unfortunately, DFTB1 is limited to molecules with ground-state energy close to the reference state. It cannot be used for systems that undergo charge transfer as these changes the electron density of the system. Thus, DFTB3 is used for biological systems (Equation 7.19):

$$E^{DFTB3} = E^{DFTB2} + \sum_{a} \sum_{b>a} \sum_{c>b>a} \Gamma \mathrm{ab} \Delta \mathrm{q}^{2} a \Delta q b \qquad [\ 7.\ 19\ ]$$

Γ is derived from a term in the DFTB2 formula and used to account for charge fluctuation[129]. All reaction dynamics in this dissertation are performed using DFTB3 due to the charge fluctuations in the quantum mechanical (QM) region[130].

7.4 QM/MM Methods

Chemical accuracy is limited to systems with less than 10 heavy atoms. However, for biological simulations only a small region can be treated quantum mechanically and requires DFTB3 or *ab initio* methods. Thus, combined quantum mechanical/molecular mechanical (QM/MM) methods are used[118].

The approach involves choosing a small region where the electronic structural changes are of interest. For enzymatic reactions, this small region usually corresponds to the catalytic region while the rest of the system is treated with a MM force field. The combination of the two theories

hosts its own challenges as electronic and steric effects must be incorporated for each region[128]. The simplest representation for the QM/MM describes the Hamiltonian in Equation 7.20:

$$H_{complete} = H_{QM} + H_{MM} + H_{QM/MM} \qquad [\,7.\,20\,]$$

The last term in Equation 7.20 can be simplified when the two regions do not share a covalent bond and can be described with the van der Waals and electrostatic interactions[128] as in Equation 7.21:

$$H_{QM/MM} = -\sum_{i,C} \frac{q_m}{r_{ic}} + \sum_{d,C} \frac{Z_d q_m}{r_{dC}} + \sum \left[4\epsilon_{dC} \left(\frac{\sigma_{dC}{}^{12}}{r_{dC}{}^{12}} \right) - \left(\frac{\sigma_{dC}{}^{6}}{r_{dC}{}^{6}} \right) \right] \qquad [\,7.\,21\,]$$

Where ***i*** is through the QM electrons, *C* the classical point charges and *d* the QM nuclei. The QM system is treated as a vacuum where the electron density is used as point charges interacting with point charges in the MM region[131]. The electrostatic interactions between these two systems are calculated accurately using the Hamiltonian (***$H_{QM/MM}$***) [129].

7.5 Molecular Docking

Docking is a vital tool for drug design and occurs when a small molecule is fitted into a larger molecule. In most X-ray crystal structures, the enzyme contains a native (bound) ligand that can be removed. Other ligands can then be docked into the empty active site to evaluate the binding affinity relative to the native ligand.

Rigid docking has six degrees of freedom where three of them are rotational and the other translational[114]. Each ligand has a set of intermolecular rotational and translational variables where a specific combination of these renders a ligand pose. However, many of these poses are rejected due to the atom pair distance between the ligand and receptor being too close or

overlapping. Docking is mainly used to virtually screen thousands of ligands against a receptor (target). Thus, methods need to be developed for the rapid estimation of binding energy.

Free energy is not calculated by enthalpic reactions as is computationally too expensive. Rather, entropy and solvent effects are used in conjunction with experimental data to render a scoring function. Structural aspects such as torsion, degrees of freedom and hydrogen donors and/or receptors are used to describe entropy. The torsion angles are constrained or fixed when a ligand binds to an enzyme, thus decreasing the entropy[132]. The binding free energy is given in Equation 7.22:

$$\Delta G_{score} = a1\Delta E_{vdw} + a2\Delta E_{el} + a3\Delta G_{rot} + a4G_{Hbond} + a5\Delta G_{solv} \qquad [\,7.\,22\,]$$

Where $\boldsymbol{a_i}$ represents the weighing factors than can be fitted more accordingly by using binding data for specific ligand-protein interactions. There are many scoring functions that utilize only the first two or last three terms of Equation 7.22.

Docking's main purpose is to rank multiple poses where from the top results the ligands can then be further analyzed by more sophisticated methods to accurately estimate binding energies[133]. One main drawback is that the empty space in the crystal structure resembles the native ligand that was removed. Thus, high scoring candidates may just be similar in structure to the native ligand and no novel discoveries can be made. Most docking is also rigid and thus will ignore any flexibility of the protein (enzyme). The two main approaches to obtain the scoring function are the empirical scoring function (ESF) and the MM/GBSA approach.

7.5.1 Empirical Scoring Function (ESF)

The empirical scoring function (ESF) maximizes the separation of compounds with strong binding affinities to those with minimal binding affinities. ESF estimates the binding affinity of a protein-ligand complex by adding energy terms such as in Equation 7.23:

$$\Delta E_{bind} = a_0 + a2\Delta E_{vdw} + +a3\Delta G_{rot} + a4G_{Hbond} + a5\Delta G_{solv} \qquad [\,7.\,23\,]$$

Where ai represents the energy factor coefficients that can be optimized using linear regression analysis. The energy terms are simplified and are accurate in predicting binding affinity and ligand poses. Additionally, the computational cost is lower compared to other scoring functions such as machine learning, random forest, and support vector machine[134, 135]. A simplified scheme for ESF is shown in Figure 7.4 where the ligand is complexed to a protein displaying different configurations or poses. The algorithm then chooses the most favorable docking pose and displays a docking score.

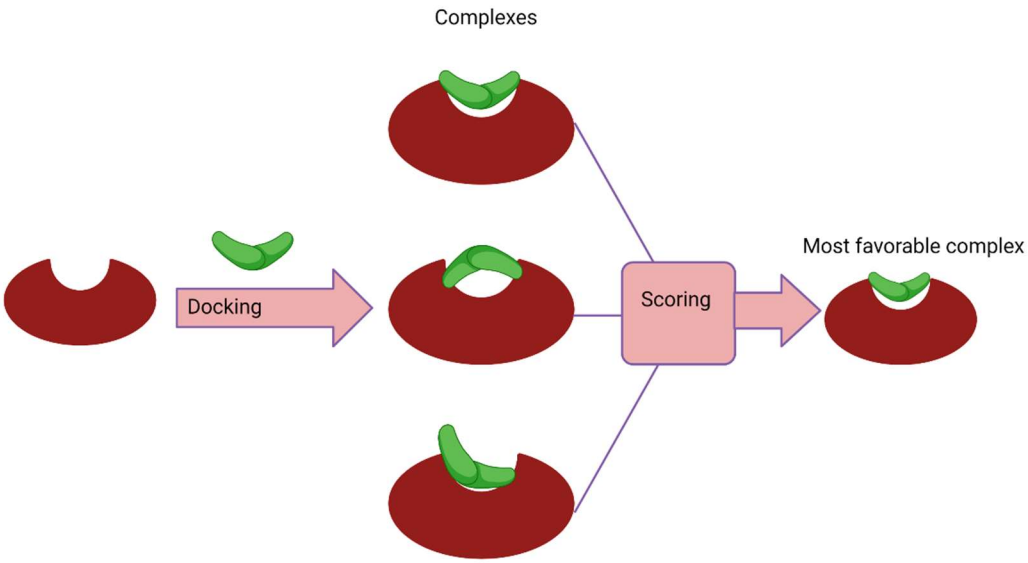


Figure 7.4 The docking process for empirical scoring functions.

7.5.2 MM/GBSA

The binding energy of a protein-ligand is calculated computationally using a thermodynamic cycle as in Figure 7.5:

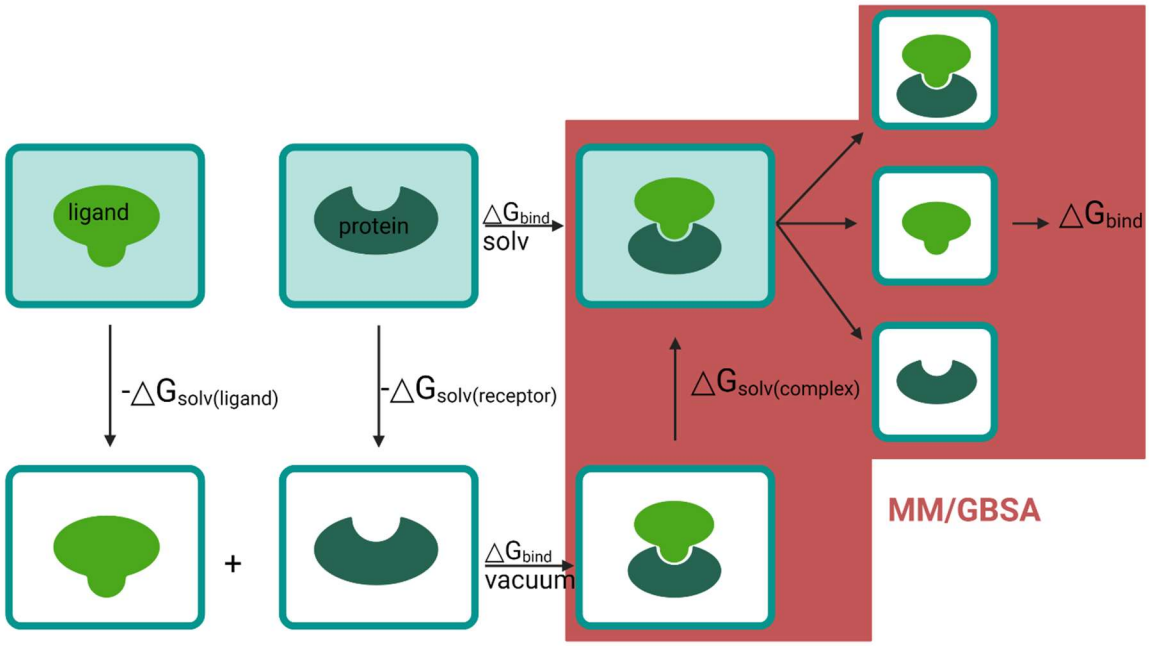


Figure 7.5 The thermodynamic cycle for free binding energy of a protein-ligand complex is shown. The MM/GBSA approach utilizes the average ΔG_{bind} as the final energy.

The binding free energy is defined as in equation 7.24

$$\Delta G_{bind(solv)} = \Delta G_{bind(vacuum)} + \Delta G_{solv(complex)} - \Delta G_{solv(ligand)} - \Delta G_{solv(receptor)} \quad [7.24]$$

Where the solvation energy is split into two components:

$$\Delta G_{solv} = \Delta G_{polar} + \Delta G_{non-polar} \quad [7.25]$$

The free energy in vacuum consists of two terms:

$$\Delta G_{vacuum} = E_{(g)} - T\Delta S \quad [7.26]$$

Where $\boldsymbol{E_{(g)}}$ represents the gas phase energy. Thus, the overall enthalpy can be represented as

$$\Delta H = \Delta E_{(g)} + \Delta G_{solv} \quad [7.27]$$

MM/GBSA is an end-point free energy method and is used to estimate $\Delta G_{solv(complex)}$ by using the Poisson-Boltzman (PB) or generalized Börn (GB) implicit solvent mode along with estimating the gas-phase energy by using classical mechanics force field. By using this method, the average of $\Delta G_{solv(complex)}$ is calculated[136].

Initially the protein-ligand complex is in vacuum and then solvated in either an implicit or explicit solvent method[136]. Multiple snapshots from a MD simulation are then generated to resemble the structure. After removal of the explicit solvent molecules the binding free energies of these structures are generated and the average of these are then considered as the final ΔG_{bind}, as denoted in Equation 7.28.

$$\Delta G_{bind} = \langle G_{complex} - (i)G_{protein} - (i)G_{ligand}(i)\rangle_i \qquad [\ 7.\ 28\ ]$$

where ***i*** represents the average over ***i*** MD trajectory snapshots[137, 138].

The two approaches discussed here were used to score the docking of the ligands in chapter 8, along with calculating the binding energy using MM/GBSA.

Chapter 8

Reaction Mechanisms of Pneumococcal Sialidase NanA with The Artificial Substrate

8.1 Theory Overview

8.1.1 Previously Proposed Mechanisms

S. pneumoniae sialidase enzymes have similar active sites but three different mechanistic pathways[4]. The general reaction mechanism involves two steps: glycosylation followed by deglycosylation of the substrate. However, only the first catalytic step is simulated. NanA can cleave either 3'-sialyllactose,6'-sialyllactose or MUANA as they only differ in the **R** group as in Figure 8.1:

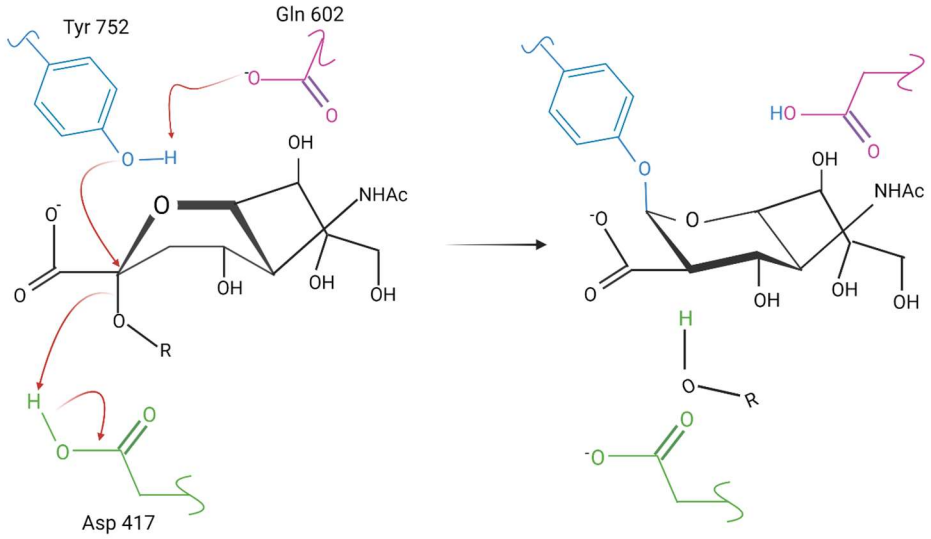


Figure 8.1 Glycosylation of NanA with either 3'-sialyllactose or 2'-(4-methylumbelliferyl)-α-D-N-acetylneuraminic acid. The R group represents lactose (with 3'-sialyllactose) or 4-MU (2'-(4-methylumbelliferyl)-α-D-N-acetylneuraminic acid).

The first mechanistic step involves breakage between the glycosidic **OR** bond and anomeric carbon of sialic acid. The Tyrosine (Figure 8.2 shown in blue) is deprotonated by the Glutamate (Figure 8.1 shown in pink) residue resulting in a Tyrosine anion. The tyrosine then undertakes a nucleophilic attack on the anomeric carbon of the sialic acid, forming a covalent glycosidic bond. The anionic **OR** group then deprotonates the aspartic acid residue (shown in green). The most likely mechanism for this step is thought to occur as an SN2 reaction.

. All three Nans undergo the first glycosylation step. Xu *et al*[4]studied and reported the kinetics of these enzymes using the TBA assay. The relationship between K_{cat} and the activation free energy is described with Equation 8.1[139]:

$$\Delta A^{\ddagger} = RTln\left(K_{cat} \times \frac{h}{K_B T}\right) \quad [8.1]$$

The results of the theoretical calculations are seen in Table 8.1.

Table 8.1 The experimental kinetic data for the *S. Pneumonia* sialidase catalysis and free energy barrier as derived from experimental data.

Enzyme	K_{cat} (min^{-1})	$\Delta \boldsymbol{A}^{\ddagger}$(kcal/mol)
NanA	894.71 ± 14.57	15.92

8.2 QM/MM Setup

8.2.1 Methodology

The aim of this section is to use QM/MM free energy simulation to construct free energy surfaces and calculate the minimum energy pathway. The enzyme-substrate complexes were used as starting structures for the QM/MM simulations after five iterations of minimizations [128].

The first step involves defining the SBC, where a water sphere is centered around the anomeric carbon with a 35 Å radius. A 5 Å was used as a buffer region (see Figure 7.3 for reference). The QM region of NanA with MUANA is shown in Figure 8.2. Hydrogen link atoms ('dummy atoms') were used to separate the classical and quantum regions.

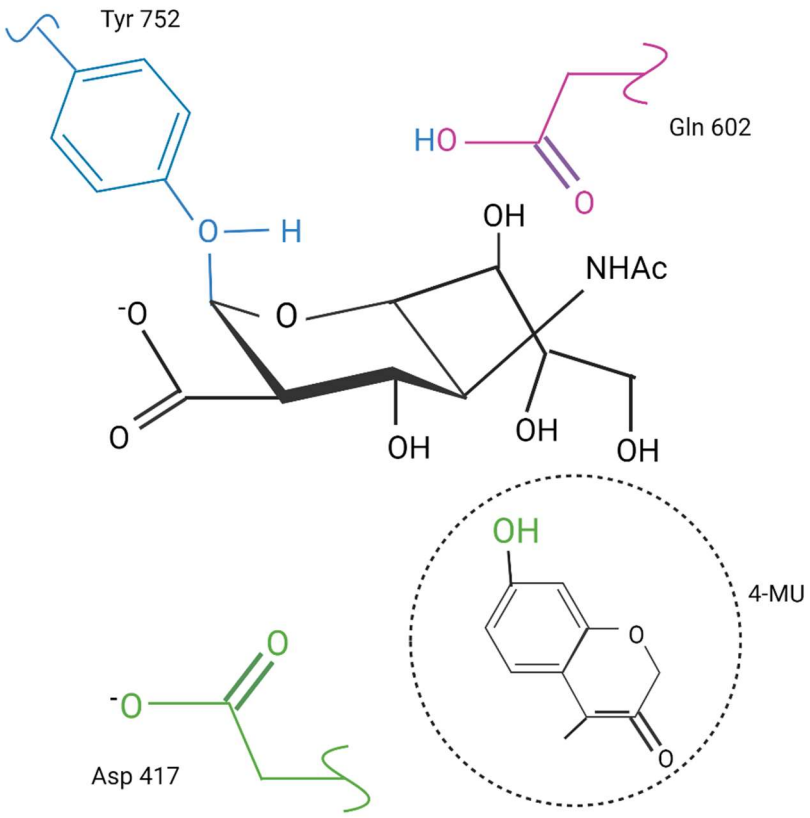


Figure 8.2 The QM/MM setup for NanA with 2'-(4-methylumbelliferyl)-α-D-N-acetylneuraminic acid. Hydrogen link atoms were used as boundaries to separate classical and quantum regions. The catalytic residues are indicated along with the respective numbering.

A 2D ligand interaction was then generated as in Figure 8.3. The arginine residues (ARG) were shown to form strong hydrogen bonds with the carboxylate groups after classical MD simulations. Other key ligand interactions can also be seen in Figure 8.3.

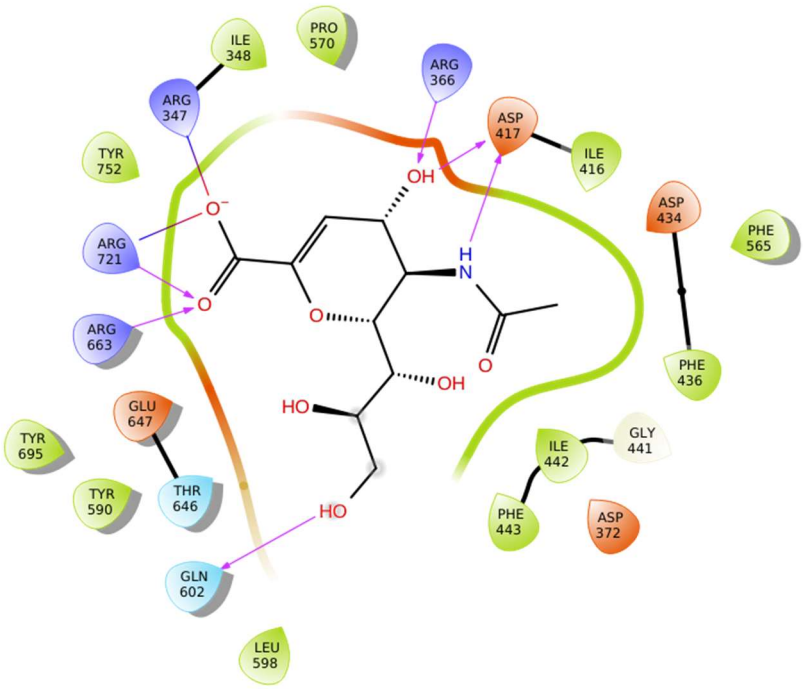


Figure 8.3 Ligand interaction diagram of NanA and α-D-N-acetylneuraminic acid after classical MD simulation.

The aspartic acid (ASP) residues were protonated while the glutamate (GLU) residues were deprotonated. The water molecule in closest proximity to the anomeric carbon was also included as it is involved in the second mechanistic step (not shown). The classic MM regions were modelled using CHARMM 36 force field and for the QM region SCC-DFTB3. The 3OB parameter set is used due to its improvement for hydrogen bond energies and proton transfers. Dispersion correction for DFTB is also used as well as the Hubbard derivatives for oxygen (-0.157), nitrogen (-0.154), carbon (-0.149) and hydrogen (0.186). A 1.2 ns equilibration was carried out prior to any QM/MM simulations.

8.3 Free energy simulations with FEARCF

8.3.1 Methodology

The reaction free energy simulations were carried out using Free Energies from Adaptive Reaction Coordinate Forces (FEARCF)[140]. Two reaction coordinates ξ1 and ξ2 are used to present a bond formation and a bond breakage in all the free energy systems. Each reaction coordinate is divided into 110 bins and sampled from 0.5 to 6 Å. In total, 15 FEARCF iterations were run on 120 cores using SCCDFTB3/MM. All the trajectories and histograms were collected after each iteration (30 ps) and the histograms were reweighed using the weighted histogram analysis method (WHAM). The free energy surfaces were then constructed to analyze further.

8.3.2 Results and Discussion

NanA glycosylation reaction with MUANA converged in the 4th FEARCF iteration. The one-dimensional free energy surface for the glycosylation step is shown in Figure 8.4.

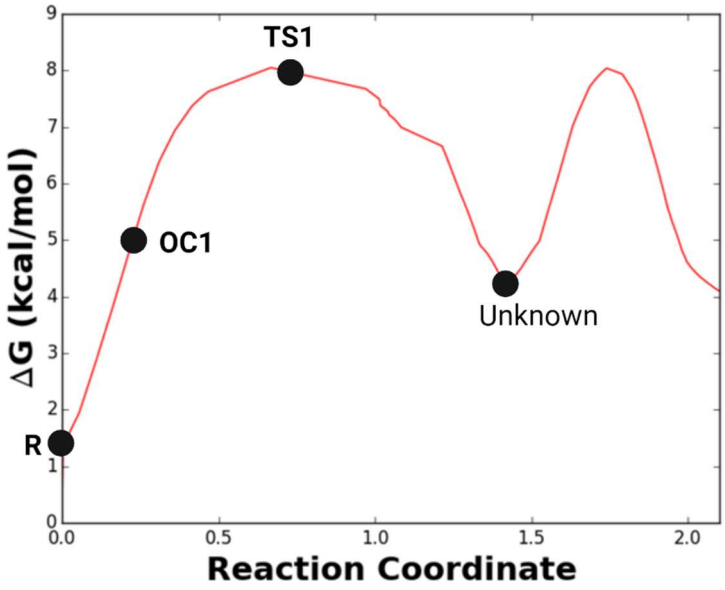


Figure 8.4 One-dimensional free energy profile of the MEP given as a function of a linear combination of the two reaction coordinates. The Reaction coordinates are in Å and the free energy in kcal/mol.

The reaction proceeds from the starting structure (Figure 8.4 R) to oxocarbenium intermediate where the tyrosine residue and 4-MU group is simultaneously bound to the anomeric carbon

(OC1). Destabilization of the OC1 orbitals leads to the first transition state (TS1). However, the free energy along the reaction coordinate reveals a minimum (unknown) that implies the formation of an intermediate. Thus, the MUANA glycosylation mechanism is different than that of 3'-sialyllactose. This unexpected result requires a detailed QM/MM reaction dynamics investigation of the deglycosylation (second mechanistic step) which is beyond the scope of this dissertation.

8.4 Molecular Docking

8.4.1 Methodology

In chapter 6 the efficacy of four commercial inhibitors were determined experimentally. The focus of this section is to study the binding of the four inhibitors to validate the inhibition constants.

Molecular docking experiments were performed using the GALAXY BRIDGE platform. The NanA protein structure was prepared by removing solvents and the default ligand (DAN) from the crystal structure (PDB ID 2vvz). The protein preparation wizard in Schrodinger software was used for structure minimization. The structures for all four inhibitor ligands were obtained from the PDB.

The ligands and NanA structures were prepared for docking using AutoDock Vina, an open sourced and user-friendly program. The program has been developed to optimize docking for carbohydrates, halogen bonds and proteins. However, limitations include the lack of macrocyclic flexibility, some metal coordination models, explicit water molecules, integration of coarse-grained ligand models and irreversible binding of the ligand. AutoDock Vina has an improved search algorithm consisting of Monte Carlo (MC) iterated search integration with a BFGS17 gradient based optimizer. The energy of each ligand pose was evaluated using VINA scoring function[141]. The ligands were then scored using Schrodinger's GLIDE[142] package and the absolute binding energy calculated using the MM/GBSA extension[138].

8.4.2 Results and Discussion

The results of three scoring functions are in Table 8.2 along with the experimental inhibition constant.

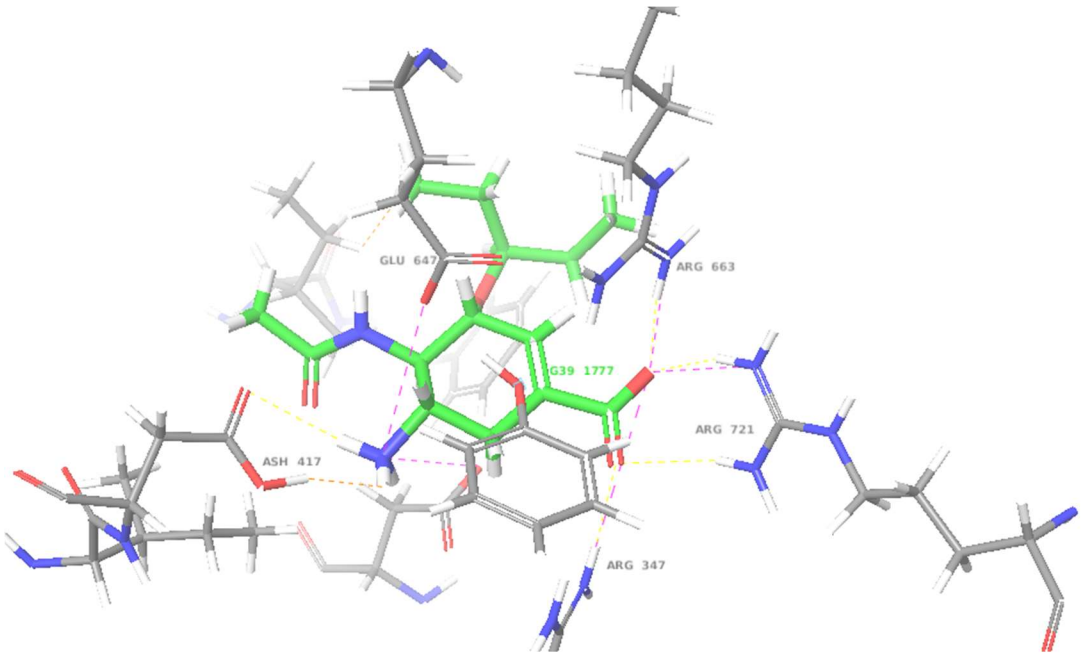
Table 8.2 The predicted binding energies of known sialidase inhibitors to *S.pneumoniae* using molecular docking and free energy simulations. The values are compared to the experimental values obtained in chapter 6.

Inhibitor	Docking score (GLIDE)	ΔG_{Bind}(kcal.mol^{-1}) (MM/GBSA)	IC_{50} value (μM)	VINA (kcal.mol^{-1})
Oseltamivir	-7.115	-31.67 ± 1.20	1.64 ± 0.72	-5.45
DANA	-6.572	- 29.56 ± 0.89	2.10 ± 1.27	-6.12
Zanamvir	-3.211	-17.34 ± 0.56	465.82 ± 88.98	-2.19
Peramivir	-4.345	-23.78 ± 0.98	461.23 ± 77.53	-2.98

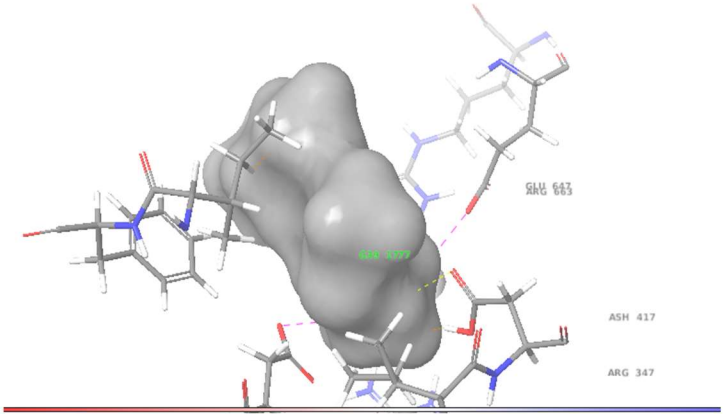
From Table 8.2 Oseltamivir and DANA are the top two inhibitors due to having the most negative docking score as calculated by the GLIDE extension, followed by Peramivir and Zanamivir. The absolute binding free energy from using MM/GBSA also ranked the inhibitors in the same order, as the more negative the binding energy is, the more favorable the enzyme-ligand complex. However, VINA ranks DANA as the most potent inhibitor followed by Oseltamivir, Peramivir and Zanamivir. There is also a clear discrepancy in the magnitude of the binding energies e.g. for Zanamivir the MM/GBSA yielded -17.34 ± 0.56 kcal.mol^{-1} while VINA produced -2.19 kcal.mol^{-1}. Thus, Autodock VINA is limited by the force fields and rigid docking as no flexible regions in the protein is considered. Overall, the docking scores (GLIDE) and binding free energy (MM/GBSA) rank Oseltamivir as the most potent inhibitor, followed by DANA, Peramivir and then Zanamivir. The experiment IC_{50} results also rank these inhibitors in the same order. Thus, Oseltamivir and

DANA show great affinity for NanA while Zanamivir and Peramivir are not considered potential inhibitors due to their poor affinity for the enzyme target. The best pose for each ligand is shown in Figure 8.5 to Figure 8.8. Table 8.3 summarizes all the key residues from NanA that interact with the four inhibitors as well as Neu5Ac.

(a)



(b)



(c)

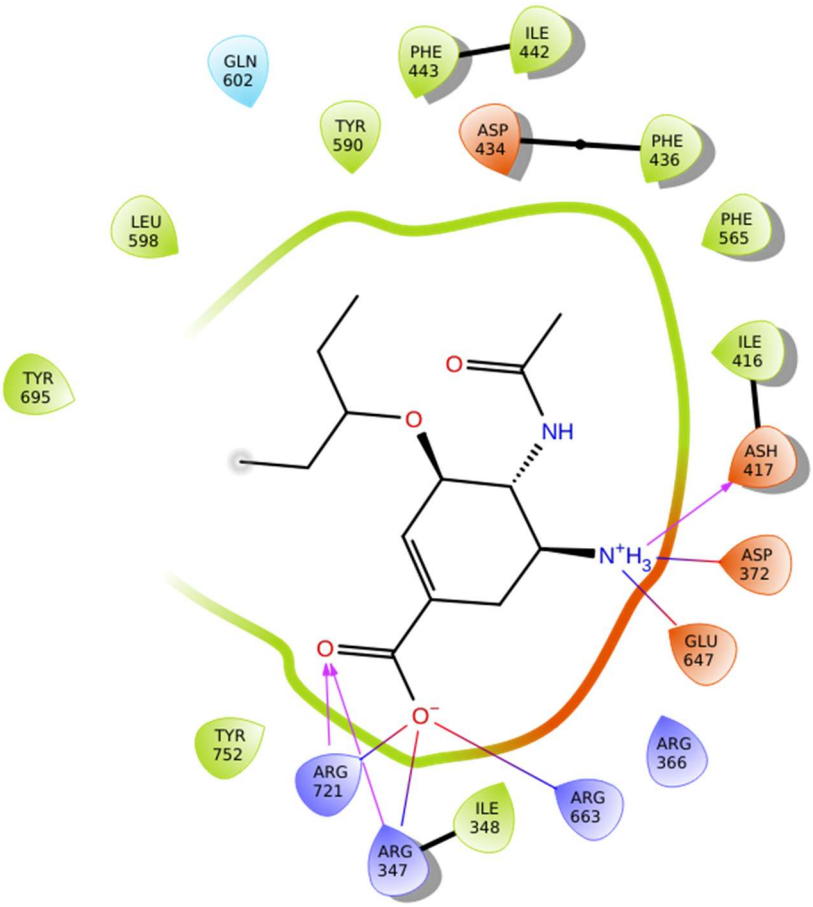
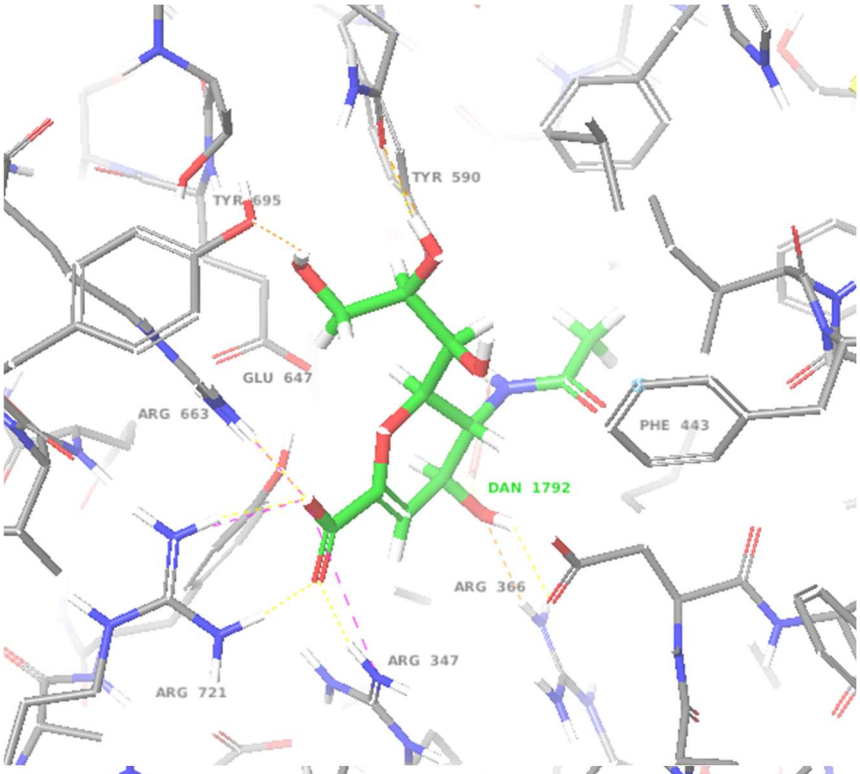
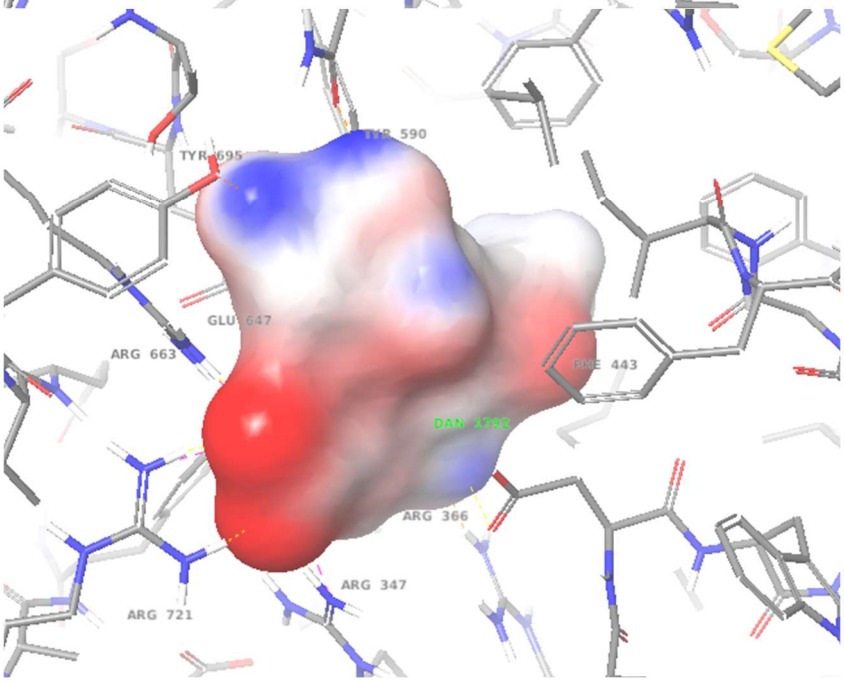


Figure 8.5 Best ligand pose for Oseltamivir carboxylate where (a) shows the ligand in green interacting with surrounding residues. (b) Shows the electrostatic potential. (c) Shows the 2D representation of the ligand interacting with residues on NanA.

(a)


TYR 590
TYR 695
GLU 647
ARG 663
PHE 443
DAN 1792
ARG 366
ARG 347
ARG 721

(b)


TYR 590
TYR 695
GLU 647
ARG 663
PHE 443
DAN 1792
ARG 366
ARG 347
ARG 721

(c)

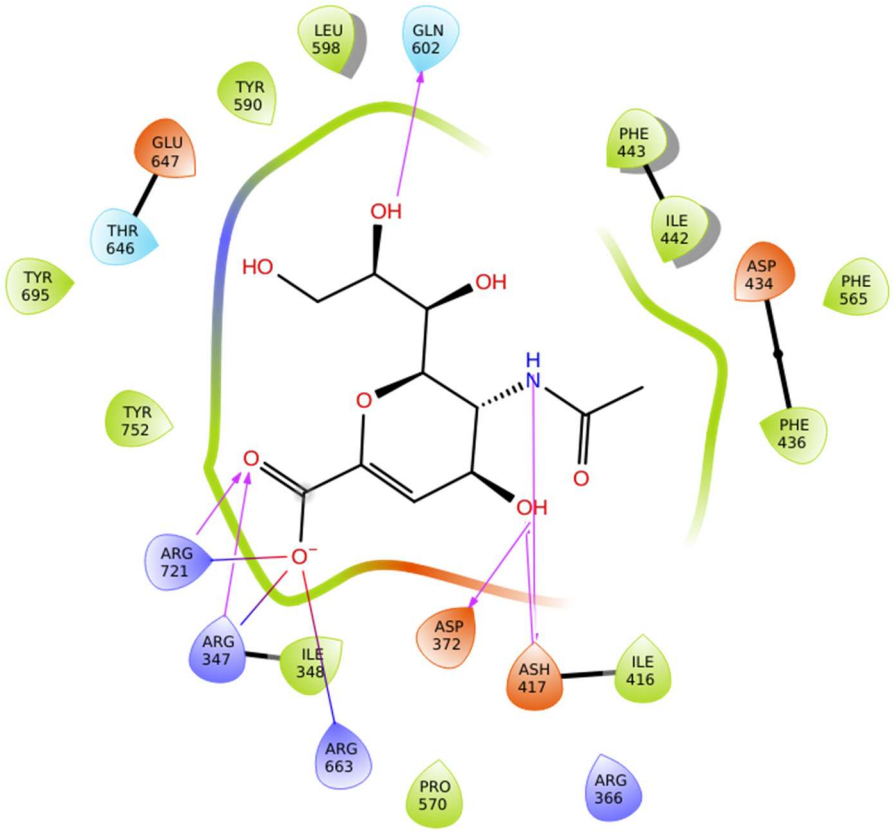
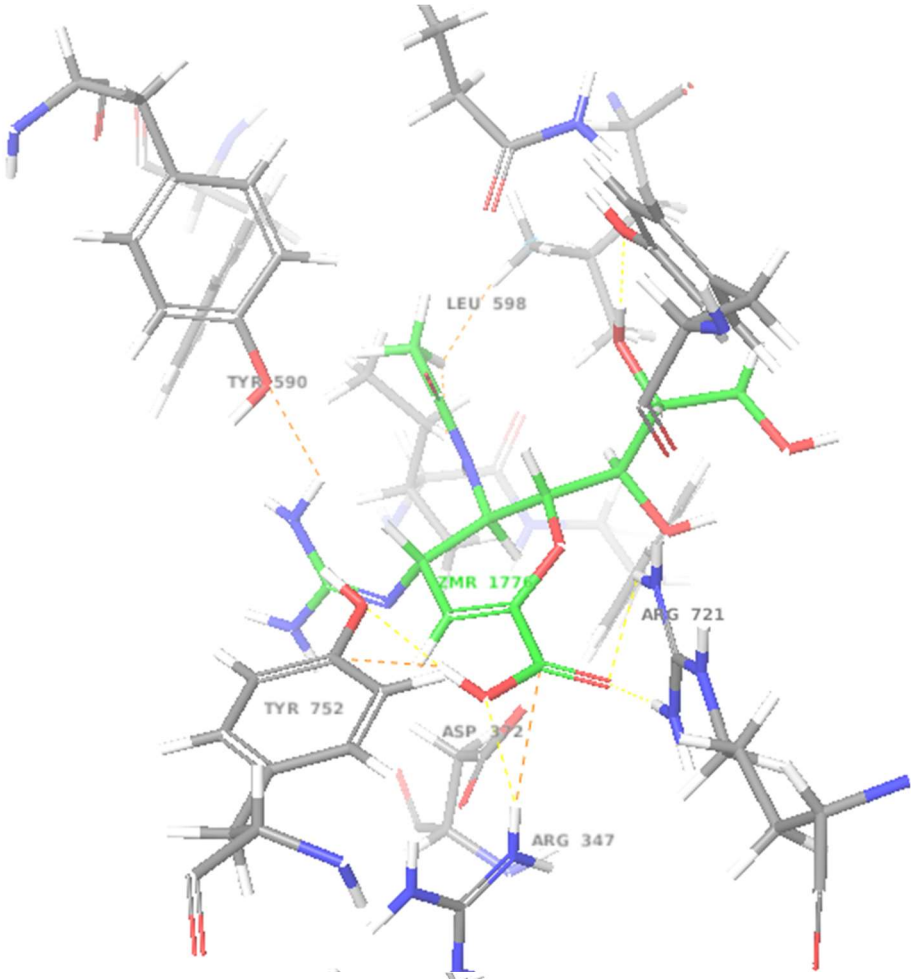
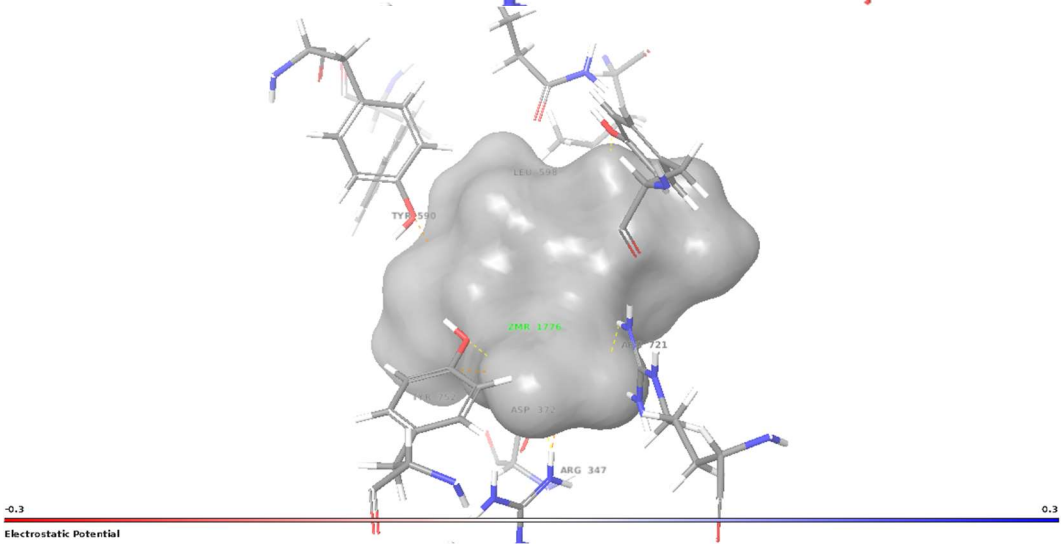


Figure 8.6 Best ligand pose for DANA. (a) Shows the ligand in green interacting with surrounding residues. (b) Shows the electrostatic potential. (c) Shows the 2D representation of the ligand interacting with residues on NanA.

(a)


LEU 598
TYR 590
ZMR 1776
ARG 721
TYR 752
ASP 372
ARG 347

(b)


-0.3
0.3
Electrostatic Potential

(c)

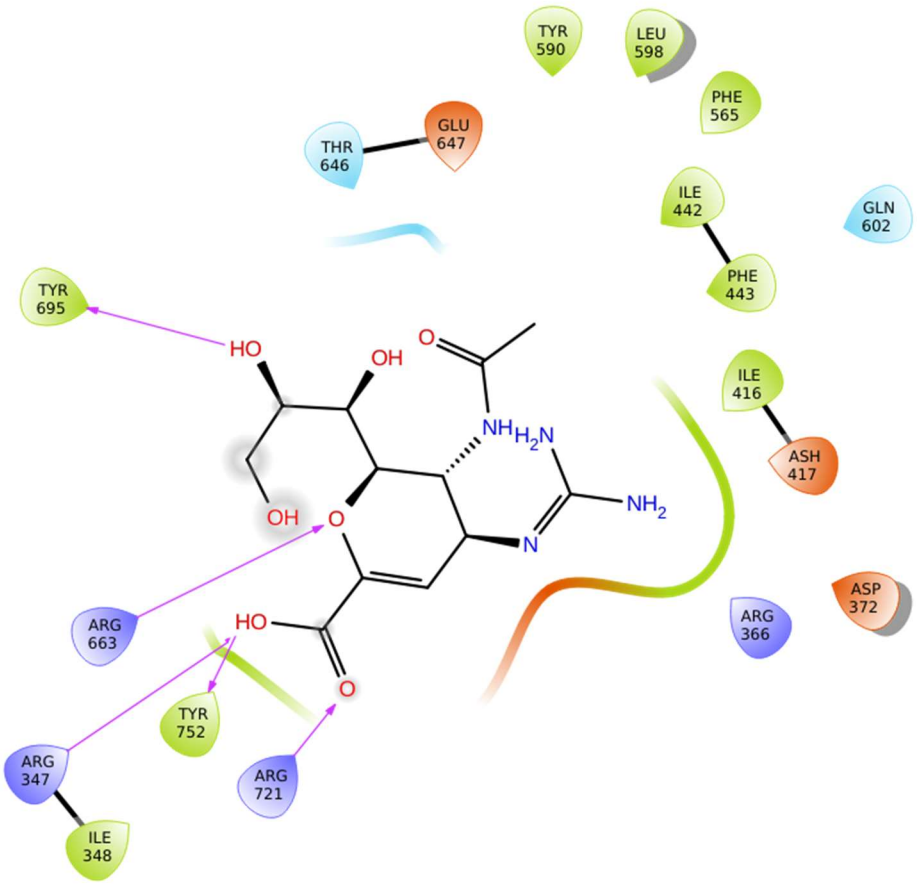
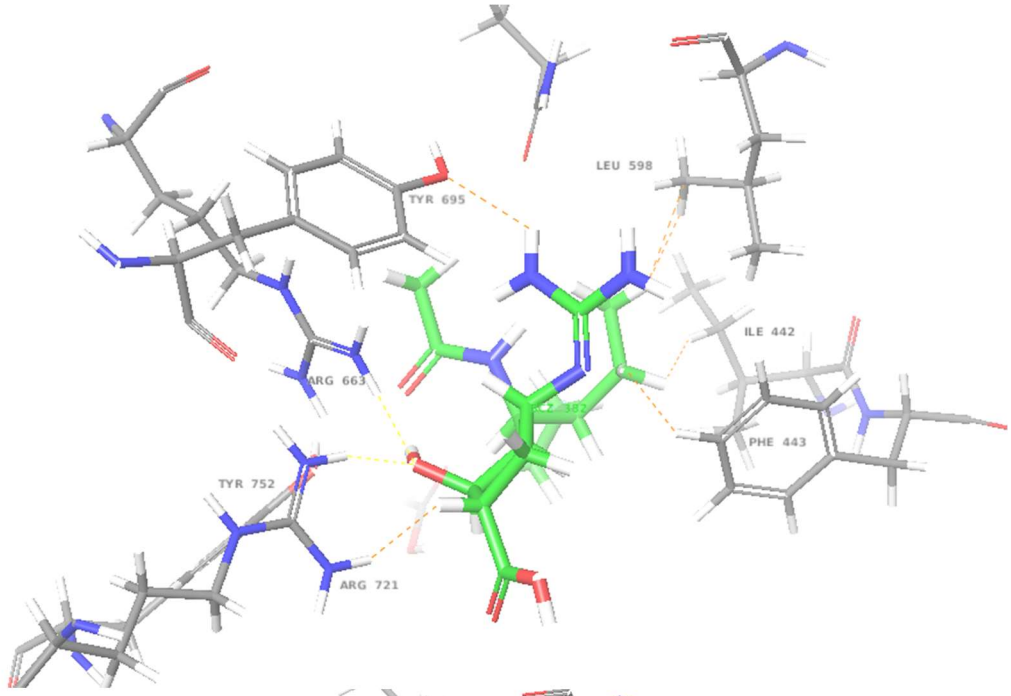
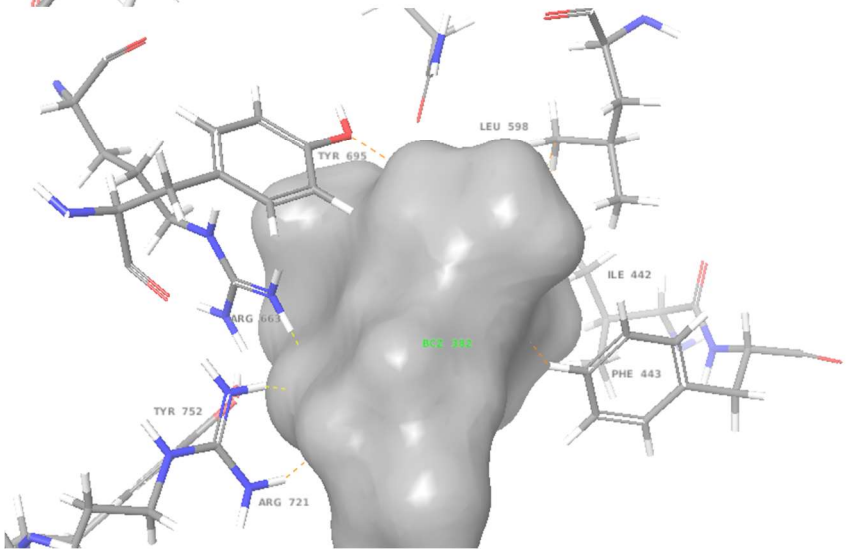


Figure 8.7 Best ligand pose for Zanamivir. (a) Shows the ligand in green interacting with surrounding residues. (b) Shows the electrostatic potential. (c) shows the 2D representation of the ligand interacting with residues on NanA.

(a)


LEU 598
TYR 695
ILE 442
ARG 663
PHE 443
TYR 752
ARG 721

(b)


LEU 598
TYR 695
ILE 442
PHE 443
TYR 752
ARG 721

(c)

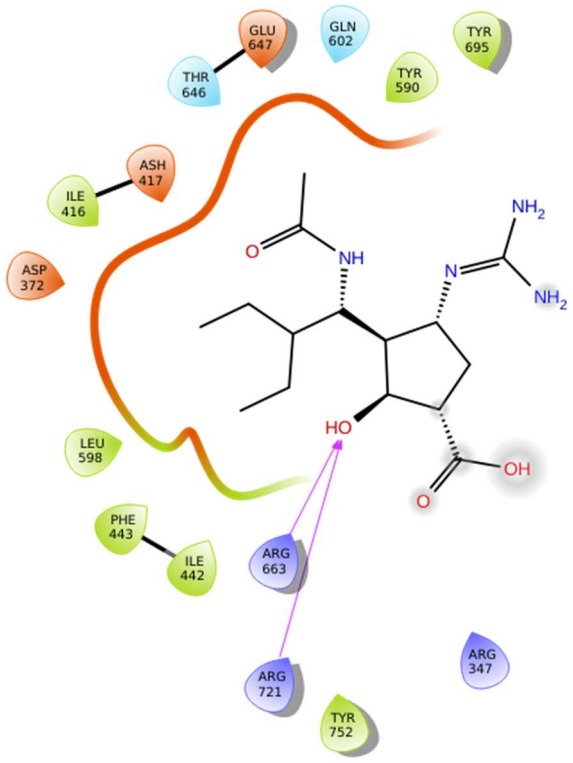


Figure 8.8 Best ligand pose for Peramivir. (a) Shows the ligand in green interacting with surrounding residues. (b) Shows the electrostatic potential. (c) Shows the 2D representation of the ligand interacting with residues on NanA.

Table 8.3 Key amino acids affecting binding in Neu5Ac and inhibitors.

	Interacting residues	Hydrophobic pocket	Positively charged residues
Oseltamivir	ARG 721 ARG 347 ARG 663 ASH 417 ASP 372 GLU 647	TYR 695 ,590,752 LEU 598 PHE 565 ILE 416,348	ASH 417 ASP 372 GLU 647
Neu5Ac	ARG 721,347, 663,366 ASP 417 GLN 602 TYR 752	TYR 752 ILE 416	GLU 647 ASP 417 ASP 372
DANA	GLN 602 ASH 417 ASP 372 ARG 721 ARG 347 ARG 663	TYR 590,695,752 LEU 598 PHE 443 ILE 348,442	ASP 372 ASH 417
Zanamivir	TYR 695 ARG 721, 347,663 TYR 752	TYR 695,750,752	GLU 647 ASH 417 ASP 372
Peramivir	ARG 721,663 LEU 598 ASP 372 ASH 417	TYR 590,695,752 ILE 416,	ASP 372

From Table 8.3 the arginine triad consists of ARG 721, 347 and 663 and is conserved in each ligand interaction. It seems to stabilize the carboxylic acid groups in Oseltamivir carboxylate, DANA and Zanamivir (Figure 8.5 to 8.7 a) and interacts with the hydroxyl group in the Peramivir scaffold

(Figure 8.8 a). The hydrophobic pocket for all ligands contains TYR 590, 695 or 752 or a combination of all three. The positively charged residues GLU 647 and ASP 372 is present in all binding pockets of the different ligands. For Neu5Ac, ARG 366 interacts with the hydroxyl group on the scaffold while ASP 427 interacts with nitrogen on the amide group. The positively charged residues stabilize the amine group for Oseltamivir carboxylate whereas for DANA ASP 372 and ASH 417 interact with the hydroxyl and amino group. For Zanamivir TYR 752 interacts with one hydroxyl group in the diol (Figure 8.7 a).

Oseltamivir is the most potent inhibitor amongst the candidates whereas Zanamivir ranks the lowest. It is important to evaluate the cause as Oseltamivir competes competitively for the biding site. Comparing the 2D structure, Oseltamivir carboxylate consists of a cyclohexene scaffold while Zanamivir consists of a dihydropyran (oxygen containing cyclohexene ring). The C_2-C_7 double bond represents sp^2 hybridized carbon atoms that mimic the oxocarbenium intermediate of the enzyme reaction mechanism (Figure 8.1 left). Oseltamivir contains two different function groups when compared to DANA and analogues. An example is the 3-pentyl side chain on C_6 (Oseltamivir) compared to the glycerol group on the same carbon for DANA. The C_4 amino group for Oseltamivir is a hydroxyl group in DANA. The 3-pentyl side chain on Oseltamivir allows high binding affinity while also lowering the overall polarity of the inhibitor.

Zanamivir is a DANA derivative where the C_4 hydroxyl group is replaced by a guanidium group that optimizes the interaction with the negatively charged GLU 119 and GLU 227 residues found in influenza neuraminidase. However, pneumococcal neuraminidase i.e. NanA differs significantly from influenza neuraminidase in the catalytic region. ARG 663 in NanA extends further than ARG 156 from influenza and reduces the available space for the bulky guanidium group. The negatively charged residues in influenza are absent in NanA, thus there is not stabilizing interaction between the guanidium and amino acid residues. However, both NanA and influenza have residues that interact similar with the carboxylate groups of Zanamivir. Zanamivir has an increased inhibitory potential to influenza neuraminidase with an IC_{50} of 1.8 nM, while for the pneumococcal neuraminidase the K_i is reported as 0.72 mM (720 000 nM). There is severe steric clash between the guanidium group and the ARG 663 residue resulting in a weak inhibitor potency [47].

Conclusions and Future Work

We have shown though our work that the artificial substrate MUANA is a valid substitute for the natural glycoconjugates 3'-sialyllactose and 6'-sialyllactose. Moreover, we have used two approaches, *in vitro* and *in-silico,* to validate results.

The MUANA assay used fluorescence as a detection method and the reaction rate of different concentrations were used to determine the initial velocity. The continuous assay allowed for a more accurate determination of the initial velocity as the formation of 4-MU was monitored in real-time. The Michalis constant K_m calculated was 41.56 ± 3.12 µM for NanA with MUANA, which matched the literature value of ~ 45 µM, validating our experiment. Thus, the optimizations and calibration of the instrument were sufficient in detecting the lower quantities of 4-MU. The catalytic efficiency K_{cat} of 851.19 ± 20.91 min^{-1} was much lower than the reported value of 12 000 min^{-1} and is possibly due to the batch-to-batch variation during enzyme expression and relies on 100% of the enzyme being active, which is not the case.

The MUANA assay constants were then compared to the TBA assay constants. However, a method had to be developed to detect the formation of Neu5Ac when 3'-sialyllactose is cleaved. The Neu5Ac product was successfully transformed into a chromophore to display a strong signal using absorbance. Unfortunately, the end-point assay approach led to the initial velocity being an estimate instead of an exact value. We chose an arbitrary incubation time of 60 minutes and used this to determine the rate. The Michaelis constant was determined to be 659.70 µM, which matched the reported value by Xu *et al* (659.70 µM). However, when time-points were taken to mimic a continuous reaction curve, the K_m decreased to 250.8 µM. This significant reduction confirmed that using an end-point assay has severe limitations and will underestimate the initial velocity.

The K_m values for the MUANA and 3'-sialyllactose substrate highlight the difference in binding affinities between the two substrates. NanA has a much higher affinity towards MUANA as the Km is more than 10 times lower than that of 3'-sialyllactose. The catalytic efficiencies, however,

were quite similar, indicating a similar turnover rate despite the affinities being so different. This suggests that 2'-(4-METHYLUMBELLIFERYL)-A-D-N-ACETYLNEURAMINIC ACID might be a suitable replacement for 3'-sialyllactose but needed further confirmation.

The inhibition study determined the potency of four commercial inhibitors by calculating the IC_{50} for each drug with NanA. Two separate screenings were performed using either MUANA or 3'-sialyllactose as the substrate. The IC_{50} values for Oseltamivir and DANA across the two substrates yielded similar results. Oseltamivir had an IC_{50} of 1.64 ± 0.72 µM for MUANA and 5.17± 1.99 µM for 3'-sialyllactose. However, for Peramivir the IC_{50} differed quite significantly between MUANA and 3'-sialyllactose, with a value of 461.23 ± 77.53 µM and 1394.51 ± 130.13 µM respectively. The TBA assay is not sensitive enough to evaluate the potency of weak inhibitors with high IC_{50} values. However, both inhibition screenings ranked the inhibitors from most to least potent as Oseltamivir, DANA, Peramivir and Zanamivir. Thus, even if NanA shows a higher affinity towards MUANA, the inhibition screening proved that for potential potent inhibitors the enzyme will prefer the inhibitor over the substrate. This further validates that MUANA is a suitable replacement for 3'-sialyllactose.

Molecular modelling was used to determine the free energy of the reaction of NanA with MUANA to compare to the binding free energy calculated from the experimental K_{cat}. However, both mechanistic steps need to be simulated to obtain the overall free energy. The first glycosylation step revealed an unknown intermediate in the one-dimensional free energy graph. Thus, a detailed QM/MM analysis must be performed for this step as well as the subsequent deglycosylation step to determine the overall free energy.

We then investigated the binding affinities and potency of the four inhibitors by virtually docking them to our protein target, NanA. The binding free energy of the ligands resulted in the same ranking from most least favorable using both MM/GBSA and VINA autodock. However, VINA autodock had much higher energies which will not result in any reaction occurring. Thus, VINA autodock is not an accurate estimation of binding affinities due to incorrect forcefields and the restrictions of rigid docking. However, the GLIDE docking score, MM/GBSA, VINA autodock and IC50 results all ranked Oseltamivir and DANA as the most potent inhibitors, followed by Peramivir

and Zanamivir. The agreement on both experimental and in-silico results strongly indicates that MUANA is a suitable replacement for 3'-sialyllactose.

Future work includes a full QM/MM analysis of NanA with MUANA for the deglycosylation step to determine the overall energy barrier to compare to the experimental K_{cat}. A full QM//MM simulation of both the glycosylation and deglycosylation step for 3'-sialyllactose will allow comparison of the free energy path and transition states between the two substrates.

References

(1) Cedrone, F.; Montagna, V.; Del Duca, L.; Camplone, L.; Mazzocca, R.; Carfagnini, F.; Fortunato, V.; Di Martino, G. The Burden of Streptococcus pneumoniae-Related Admissions and In-Hospital Mortality: A Retrospective Observational Study between the Years 2015 and 2022 from a Southern Italian Province. *Vaccines (Basel)* **2023**, *11* (8). DOI: 10.3390/vaccines11081324 From NLM PubMed-not-MEDLINE.

(2) Manco, S.; Hernon, F.; Yesilkaya, H.; Paton, J. C.; Andrew, P. W.; Kadioglu, A. Pneumococcal neuraminidases A and B both have essential roles during infection of the respiratory tract and sepsis. *Infect Immun* **2006**, *74* (7), 4014-4020. DOI: 10.1128/IAI.01237-05.

(3) Neu, U.; Bauer, J.; Stehle, T. Viruses and sialic acids: rules of engagement. *Curr Opin Struct Biol* **2011**, *21* (5), 610-618. DOI: 10.1016/j.sbi.2011.08.009 From NLM Medline.

(4) Xu, G.; Kiefel, M. J.; Wilson, J. C.; Andrew, P. W.; Oggioni, M. R.; Taylor, G. L. Three Streptococcus pneumoniae Sialidases: Three Different Products. *Journal of the American Chemical Society* **2011**, *133* (6), 1718-1721. DOI: 10.1021/ja110733q.

(5) Gupta, M. N.; Uversky, V. N. Chapter 1 - Enzymology: early insights. In *Structure and Intrinsic Disorder in Enzymology*, Gupta, M. N., Uversky, V. N. Eds.; Academic Press, 2023; pp 1-29.

(6) Lemieux, R. U.; Spohr, U. How Emil Fischer was led to the lock and key concept for enzyme specificity. *Adv Carbohydr Chem Biochem* **1994**, *50*, 1-20. From NLM Medline.

(7) Yan, C.; Zou, X. 3.13 - Modeling Protein Flexibility in Molecular Docking. In *Comprehensive Medicinal Chemistry III*, Chackalamannil, S., Rotella, D., Ward, S. E. Eds.; Elsevier, 2017; pp 319-328.

(8) Barnsley, K. K.; Ondrechen, M. J. Enzyme active sites: Identification and prediction of function using computational chemistry. *Current Opinion in Structural Biology* **2022**, *74*, 102384. DOI: https://doi.org/10.1016/j.sbi.2022.102384.

(9) Koshland, D. E. Application of a Theory of Enzyme Specificity to Protein Synthesis. *Proc Natl Acad Sci U S A* **1958**, *44* (2), 98-104. DOI: 10.1073/pnas.44.2.98.

(10) Acker, M. G.; Auld, D. S. Considerations for the design and reporting of enzyme assays in high-throughput screening applications. *Perspectives in Science* **2014**, *1* (1), 56-73. DOI: https://doi.org/10.1016/j.pisc.2013.12.001.

(11) Robinson, P. K. Enzymes: principles and biotechnological applications. *Essays Biochem* **2015**, *59*, 1-41. DOI: 10.1042/bse0590001.

(12) Blank, K.; De Cremer, G.; Hofkens, J. Fluorescence-based analysis of enzymes at the single-molecule level. *Biotechnol J* **2009**, *4* (4), 465-479. DOI: 10.1002/biot.200800262.

(13) Shanker, N.; Bane, S. L. Basic Aspects of Absorption and Fluorescence Spectroscopy and Resonance Energy Transfer Methods. In *Methods in Cell Biology*, Vol. 84; Academic Press, 2008; pp 213-242.

(14) Hans, B. *Practical Enzymology*; 2008.

(15) Lai, C. J.; Wu, J. C. A simple kinetic method for rapid mechanistic analysis of reversible enzyme inhibitors. *Assay Drug Dev Technol* **2003**, *1* (4), 527-535. DOI: 10.1089/154065803322302781.

(16) Brooks, H. B.; Geeganage, S.; Kahl, S. D.; Montrose, C.; Sittampalam, S.; Smith, M. C.; Weidner, J. R. Basics of Enzymatic Assays for HTS. In *Assay Guidance Manual*, Markossian, S., Grossman, A., Brimacombe, K., Arkin, M., Auld, D., Austin, C., Baell, J., Chung, T. D. Y., Coussens, N. P., Dahlin, J. L., et al. Eds.; 2004.

(17) Varki, A. Sialic acids in human health and disease. *Trends Mol Med* **2008**, *14* (8), 351-360. DOI: 10.1016/j.molmed.2008.06.002 From NLM Medline.

(18) Hsu, P. H.; Chiu, D. C.; Wu, K. L.; Lee, P. S.; Jan, J. T.; Cheng, Y. E.; Tsai, K. C.; Cheng, T. J.; Fang, J. M. Acylguanidine derivatives of zanamivir and oseltamivir: Potential orally available prodrugs against influenza viruses. *Eur J Med Chem* **2018**, *154*, 314-323. DOI: 10.1016/j.ejmech.2018.05.030 From NLM Medline.

(19) Hentrich, K.; Lofling, J.; Pathak, A.; Nizet, V.; Varki, A.; Henriques-Normark, B. Streptococcus pneumoniae Senses a Human-like Sialic Acid Profile via the Response Regulator CiaR. *Cell Host Microbe* **2016**, *20* (3), 307-317. DOI: 10.1016/j.chom.2016.07.019 From NLM Medline.

(20) Ghosh, S. Chapter 1 - Sialic acid and biology of life: An introduction. In *Sialic Acids and Sialoglycoconjugates in the Biology of Life, Health and Disease*, Ghosh, S. Ed.; Academic Press, 2020; pp 1-61.
(21) Butcher, E. C.; Berg, E. L.; Kunkel, E. J. Systems biology in drug discovery. *Nat Biotechnol* **2004**, *22* (10), 1253-1259. DOI: 10.1038/nbt1017.
(22) DiMasi, J. A.; Hansen, R. W.; Grabowski, H. G. The price of innovation: new estimates of drug development costs. *J Health Econ* **2003**, *22* (2), 151-185. DOI: 10.1016/S0167-6296(02)00126-1.
(23) Hunter, P. J.; Borg, T. K. Integration from proteins to organs: the Physiome Project. *Nat Rev Mol Cell Biol* **2003**, *4* (3), 237-243. DOI: 10.1038/nrm1054.
(24) Bhalla, U. S.; Iyengar, R. Emergent properties of networks of biological signaling pathways. *Science* **1999**, *283* (5400), 381-387. DOI: 10.1126/science.283.5400.381.
(25) Ideker, T.; Thorsson, V.; Ranish, J. A.; Christmas, R.; Buhler, J.; Eng, J. K.; Bumgarner, R.; Goodlett, D. R.; Aebersold, R.; Hood, L. Integrated genomic and proteomic analyses of a systematically perturbed metabolic network. *Science* **2001**, *292* (5518), 929-934. DOI: 10.1126/science.292.5518.929.
(26) Weston, A. D.; Hood, L. Systems biology, proteomics, and the future of health care: toward predictive, preventative, and personalized medicine. *J Proteome Res* **2004**, *3* (2), 179-196. DOI: 10.1021/pr0499693.
(27) Klenow, L.; Elfageih, R.; Gao, J.; Wan, H.; Withers, S. G.; de Gier, J. W.; Daniels, R. Influenza virus and pneumococcal neuraminidases enhance catalysis by similar yet distinct sialic acid-binding strategies. *J Biol Chem* **2023**, *299* (2), 102891. DOI: 10.1016/j.jbc.2023.102891.
(28) Uchiyama, S.; Carlin, A. F.; Khosravi, A.; Weiman, S.; Banerjee, A.; Quach, D.; Hightower, G.; Mitchell, T. J.; Doran, K. S.; Nizet, V. The surface-anchored NanA protein promotes pneumococcal brain endothelial cell invasion. *J Exp Med* **2009**, *206* (9), 1845-1852. DOI: 10.1084/jem.20090386 From NLM Medline.
(29) Corfield, A. P.; Veh, R. W.; Wember, M.; Michalski, J. C.; Schauer, R. The release of N-acetyl- and N-glycolloyl-neuraminic acid from soluble complex carbohydrates and erythrocytes by

bacterial, viral and mammalian sialidases. *Biochem J* **1981**, *197* (2), 293-299. DOI: 10.1042/bj1970293 From NLM Medline.

(30) Coccimiglio, M.; Chiodo, F.; van Kooyk, Y. The sialic acid-Siglec immune checkpoint: an opportunity to enhance immune responses and therapy effectiveness in melanoma. *Br J Dermatol* **2024**. DOI: 10.1093/bjd/ljad517 From NLM Publisher.

(31) Ghosh, M.; Hazarika, P.; Dhanya, S. J.; Pooja, D.; Kulhari, H. Exploration of sialic acid receptors as a potential target for cancer treatment: A comprehensive review. *Int J Biol Macromol* **2023**, *257* (Pt 1), 128415. DOI: 10.1016/j.ijbiomac.2023.128415 From NLM Publisher.

(32) Tolentino, M. J.; Tolentino, A. J.; Tolentino, E. M.; Krishnan, A.; Genead, M. A. Sialic Acid Mimetic Microglial Sialic Acid-Binding Immunoglobulin-like Lectin Agonism: Potential to Restore Retinal Homeostasis and Regain Visual Function in Age-Related Macular Degeneration. *Pharmaceuticals (Basel)* **2023**, *16* (12). DOI: 10.3390/ph16121735 From NLM PubMed-not-MEDLINE.

(33) Awofiranye, A. E.; Dhar, C.; He, P.; Varki, A.; Koffas, M. A. G.; Linhardt, R. J. N-glycolylated carbohydrates in nature. *Glycobiology* **2022**, *32* (11), 921-932. DOI: 10.1093/glycob/cwac048 From NLM Medline.

(34) Yu, H.; Zheng, Z.; Zhang, L.; Yang, X.; Varki, A.; Chen, X. Chemoenzymatic Synthesis of N-Acetyl Analogues of 9-O-Acetylated b-Series Gangliosides. *Tetrahedron* **2023**, *142*. DOI: 10.1016/j.tet.2023.133522 From NLM PubMed-not-MEDLINE.

(35) Rudd, J. M.; Ashar, H. K.; Chow, V. T.; Teluguakula, N. Lethal Synergism between Influenza and Streptococcus pneumoniae. *J Infect Pulm Dis* **2016**, *2* (2). DOI: 10.16966/2470-3176.114.

(36) McCullers, J. A. The co-pathogenesis of influenza viruses with bacteria in the lung. *Nat Rev Microbiol* **2014**, *12* (4), 252-262. DOI: 10.1038/nrmicro3231.

(37) Kash, J. C.; Taubenberger, J. K. The role of viral, host, and secondary bacterial factors in influenza pathogenesis. *Am J Pathol* **2015**, *185* (6), 1528-1536. DOI: 10.1016/j.ajpath.2014.08.030.

(38) Viboud, C.; Simonsen, L.; Fuentes, R.; Flores, J.; Miller, M. A.; Chowell, G. Global Mortality Impact of the 1957-1959 Influenza Pandemic. *J Infect Dis* **2016**, *213* (5), 738-745. DOI: 10.1093/infdis/jiv534.

(39) Wang, X. Y.; Kilgore, P. E.; Lim, K. A.; Wang, S. M.; Lee, J.; Deng, W.; Mo, M. Q.; Nyambat, B.; Ma, J. C.; Favorov, M. O.; Clemens, J. D. Influenza and bacterial pathogen coinfections in the 20th century. *Interdiscip Perspect Infect Dis* **2011**, *2011*, 146376. DOI: 10.1155/2011/146376.
(40) McCullers, J. A. Insights into the interaction between influenza virus and pneumococcus. *Clin Microbiol Rev* **2006**, *19* (3), 571-582. DOI: 10.1128/CMR.00058-05.
(41) Narasaraju, T.; Yang, E.; Samy, R. P.; Ng, H. H.; Poh, W. P.; Liew, A. A.; Phoon, M. C.; van Rooijen, N.; Chow, V. T. Excessive neutrophils and neutrophil extracellular traps contribute to acute lung injury of influenza pneumonitis. *Am J Pathol* **2011**, *179* (1), 199-210. DOI: 10.1016/j.ajpath.2011.03.013.
(42) Brinkmann, V.; Reichard, U.; Goosmann, C.; Fauler, B.; Uhlemann, Y.; Weiss, D. S.; Weinrauch, Y.; Zychlinsky, A. Neutrophil extracellular traps kill bacteria. *Science* **2004**, *303* (5663), 1532-1535. DOI: 10.1126/science.1092385.
(43) Narayana Moorthy, A.; Narasaraju, T.; Rai, P.; Perumalsamy, R.; Tan, K. B.; Wang, S.; Engelward, B.; Chow, V. T. In vivo and in vitro studies on the roles of neutrophil extracellular traps during secondary pneumococcal pneumonia after primary pulmonary influenza infection. *Front Immunol* **2013**, *4*, 56. DOI: 10.3389/fimmu.2013.00056.
(44) Spelmink, L.; Sender, V.; Hentrich, K.; Kuri, T.; Plant, L.; Henriques-Normark, B. Toll-Like Receptor 3/TRIF-Dependent IL-12p70 Secretion Mediated by Streptococcus pneumoniae RNA and Its Priming by Influenza A Virus Coinfection in Human Dendritic Cells. *mBio* **2016**, *7* (2), e00168-00116. DOI: 10.1128/mBio.00168-16.
(45) Mathew, B. J.; Gupta, P.; Naaz, T.; Rai, R.; Gupta, S.; Gupta, S.; Chaurasiya, S. K.; Purwar, S.; Biswas, D.; Vyas, A. K.; Singh, A. K. Role of Streptococcus pneumoniae extracellular glycosidases in immune evasion. *Front Cell Infect Microbiol* **2023**, *13*, 1109449. DOI: 10.3389/fcimb.2023.1109449.
(46) Taylor, G. Sialidases: structures, biological significance and therapeutic potential. *Curr Opin Struct Biol* **1996**, *6* (6), 830-837. DOI: 10.1016/s0959-440x(96)80014-5.
(47) Gut, H.; Xu, G.; Taylor, G. L.; Walsh, M. A. Structural basis for Streptococcus pneumoniae NanA inhibition by influenza antivirals zanamivir and oseltamivir carboxylate. *J Mol Biol* **2011**, *409* (4), 496-503. DOI: 10.1016/j.jmb.2011.04.016.

(48) Juge, N.; Tailford, L.; Owen, C. D. Sialidases from gut bacteria: a mini-review. *Biochem Soc Trans* **2016**, *44* (1), 166-175. DOI: 10.1042/BST20150226.
(49) Pettigrew, M. M.; Fennie, K. P.; York, M. P.; Daniels, J.; Ghaffar, F. Variation in the presence of neuraminidase genes among Streptococcus pneumoniae isolates with identical sequence types. *Infect Immun* **2006**, *74* (6), 3360-3365. DOI: 10.1128/IAI.01442-05.
(50) Bjarnsholt, T. The role of bacterial biofilms in chronic infections. *APMIS Suppl* **2013**, (136), 1-51. DOI: 10.1111/apm.12099.
(51) Chen, L.; Wen, Y. M. The role of bacterial biofilm in persistent infections and control strategies. *Int J Oral Sci* **2011**, *3* (2), 66-73. DOI: 10.4248/IJOS11022.
(52) Janoff, E. N.; Rubins, J. B.; Fasching, C.; Charboneau, D.; Rahkola, J. T.; Plaut, A. G.; Weiser, J. N. Pneumococcal IgA1 protease subverts specific protection by human IgA1. *Mucosal Immunol* **2014**, *7* (2), 249-256. DOI: 10.1038/mi.2013.41.
(53) King, S. J. Pneumococcal modification of host sugars: a major contributor to colonization of the human airway? *Mol Oral Microbiol* **2010**, *25* (1), 15-24. DOI: 10.1111/j.2041-1014.2009.00564.x.
(54) Parker, D.; Soong, G.; Planet, P.; Brower, J.; Ratner, A. J.; Prince, A. The NanA neuraminidase of Streptococcus pneumoniae is involved in biofilm formation. *Infect Immun* **2009**, *77* (9), 3722-3730. DOI: 10.1128/IAI.00228-09.
(55) Hakansson, A.; Roche, H.; Mirza, S.; McDaniel, L. S.; Brooks-Walter, A.; Briles, D. E. Characterization of binding of human lactoferrin to pneumococcal surface protein A. *Infect Immun* **2001**, *69* (5), 3372-3381. DOI: 10.1128/IAI.69.5.3372-3381.2001.
(56) Skattum, L.; van Deuren, M.; van der Poll, T.; Truedsson, L. Complement deficiency states and associated infections. *Mol Immunol* **2011**, *48* (14), 1643-1655. DOI: 10.1016/j.molimm.2011.05.001.
(57) Gill, D. R.; Pringle, I. A.; Hyde, S. C. Progress and prospects: the design and production of plasmid vectors. *Gene Ther* **2009**, *16* (2), 165-171. DOI: 10.1038/gt.2008.183.
(58) Bok, J. W.; Keller, N. P. Fast and easy method for construction of plasmid vectors using modified quick-change mutagenesis. *Methods Mol Biol* **2012**, *944*, 163-174. DOI: 10.1007/978-1-62703-122-6_11.

(59) Green, M. R.; Sambrook, J. Cloning and Transformation with Plasmid Vectors. *Cold Spring Harb Protoc* **2021**, *2021* (11). DOI: 10.1101/pdb.top101170.
(60) Bolivar, F.; Backman, K. [16] Plasmids of Escherichia coli as cloning vectors. In *Methods in Enzymology*, Vol. 68; Academic Press, 1979; pp 245-267.
(61) Isfort, R. J. Enzymatic purification of plasmid DNA. *Biotechniques* **1992**, *12* (6), 798-800, 802, 804.
(62) Xu, Z.; von Grafenstein, S.; Walther, E.; Fuchs, J. E.; Liedl, K. R.; Sauerbrei, A.; Schmidtke, M. Sequence diversity of NanA manifests in distinct enzyme kinetics and inhibitor susceptibility. *Sci Rep* **2016**, *6*, 25169. DOI: 10.1038/srep25169.
(63) Wang, F.; Liu, Y.; Du, C.; Gao, R. Current Strategies for Real-Time Enzyme Activation. *Biomolecules* **2022**, *12* (5). DOI: 10.3390/biom12050599.
(64) Jiang, P. L.; Wang, C.; Diehl, A.; Viner, R.; Etienne, C.; Nandhikonda, P.; Foster, L.; Bomgarden, R. D.; Liu, F. A Membrane-Permeable and Immobilized Metal Affinity Chromatography (IMAC) Enrichable Cross-Linking Reagent to Advance In Vivo Cross-Linking Mass Spectrometry. *Angew Chem Int Ed Engl* **2022**, *61* (12), e202113937. DOI: 10.1002/anie.202113937.
(65) Tolmachov, O. Designing plasmid vectors. *Methods Mol Biol* **2009**, *542*, 117-129. DOI: 10.1007/978-1-59745-561-9_6.
(66) Valles, M.; Pujals, S.; Albertazzi, L.; Sanchez, S. Enzyme Purification Improves the Enzyme Loading, Self-Propulsion, and Endurance Performance of Micromotors. *ACS Nano* **2022**, *16* (4), 5615-5626. DOI: 10.1021/acsnano.1c10520.
(67) Johansen, T.; Birgisdottir, Å. B.; Huber, J.; Kniss, A.; Dötsch, V.; Kirkin, V.; Rogov, V. V. Chapter Nine - Methods for Studying Interactions Between Atg8/LC3/GABARAP and LIR-Containing Proteins. In *Methods in Enzymology*, Galluzzi, L., Bravo-San Pedro, J. M., Kroemer, G. Eds.; Vol. 587; Academic Press, 2017; pp 143-169.
(68) Boeker, E. A. Integrated rate equations for irreversible enzyme-catalysed first-order and second-order reactions. *Biochem J* **1985**, *226* (1), 29-35. DOI: 10.1042/bj2260029.
(69) Imoto, T. Formulation of a universal first-order rate constant for enzymatic reactions. *Biosci Biotechnol Biochem* **2013**, *77* (8), 1703-1708. DOI: 10.1271/bbb.130270.

(70) Copeland, R. *Enzymes: A Practical Introduction to Structure, Mechanism, and Data Analysis*; 2000.
(71) Schnell, S.; Mendoza, C. The condition for pseudo-first-order kinetics in enzymatic reactions is independent of the initial enzyme concentration. *Biophys Chem* **2004**, *107* (2), 165-174. DOI: 10.1016/j.bpc.2003.09.003.
(72) Peletier, L. A.; Gabrielsson, J. Impact of enzyme turnover on the dynamics of the Michaelis-Menten model. *Math Biosci* **2022**, *346*, 108795. DOI: 10.1016/j.mbs.2022.108795.
(73) Seibert, E.; Tracy, T. S. Fundamentals of Enzyme Kinetics: Michaelis-Menten and Non-Michaelis-Type (Atypical) Enzyme Kinetics. *Methods Mol Biol* **2021**, *2342*, 3-27. DOI: 10.1007/978-1-0716-1554-6_1.
(74) Choi, B.; Rempala, G. A.; Kim, J. K. Beyond the Michaelis-Menten equation: Accurate and efficient estimation of enzyme kinetic parameters. *Sci Rep* **2017**, *7* (1), 17018. DOI: 10.1038/s41598-017-17072-z.
(75) Srinivasan, B. A guide to the Michaelis-Menten equation: steady state and beyond. *FEBS J* **2022**, *289* (20), 6086-6098. DOI: 10.1111/febs.16124.
(76) Schnell, S. Validity of the Michaelis-Menten equation--steady-state or reactant stationary assumption: that is the question. *FEBS J* **2014**, *281* (2), 464-472. DOI: 10.1111/febs.12564.
(77) Saganuwan, S. A. Application of modified Michaelis - Menten equations for determination of enzyme inducing and inhibiting drugs. *BMC Pharmacol Toxicol* **2021**, *22* (1), 57. DOI: 10.1186/s40360-021-00521-x.
(78) Buchholz, P. C. F.; Ohs, R.; Spiess, A. C.; Pleiss, J. Progress Curve Analysis Within BioCatNet: Comparing Kinetic Models for Enzyme-Catalyzed Self-Ligation. *Biotechnol J* **2019**, *14* (3), e1800183. DOI: 10.1002/biot.201800183.
(79) Ring, B.; Wrighton, S. A.; Mohutsky, M. Reversible mechanisms of enzyme inhibition and resulting clinical significance. *Methods Mol Biol* **2014**, *1113*, 37-56. DOI: 10.1007/978-1-62703-758-7_4.
(80) Briggs, G. E.; Haldane, J. B. A Note on the Kinetics of Enzyme Action. *Biochem J* **1925**, *19* (2), 338-339. DOI: 10.1042/bj0190338.

(81) Leang, S. K.; Hurt, A. C. Fluorescence-based Neuraminidase Inhibition Assay to Assess the Susceptibility of Influenza Viruses to The Neuraminidase Inhibitor Class of Antivirals. *J Vis Exp* **2017**, (122). DOI: 10.3791/55570.

(82) Bayarri-Olmos, R.; Idorn, M.; Rosbjerg, A.; Perez-Alos, L.; Hansen, C. B.; Johnsen, L. B.; Helgstrand, C.; Zosel, F.; Bjelke, J. R.; Oberg, F. K.; et al. SARS-CoV-2 Neutralizing Antibody Responses towards Full-Length Spike Protein and the Receptor-Binding Domain. *J Immunol* **2021**, *207* (3), 878-887. DOI: 10.4049/jimmunol.2100272.

(83) Endo, S.; Morita, M.; Ueno, M.; Maeda, T.; Terabayashi, T. Fluorescent labeling of a carboxyl group of sialic acid for MALDI-MS analysis of sialyloligosaccharides and ganglioside. *Biochem Biophys Res Commun* **2009**, *378* (4), 890-894. DOI: 10.1016/j.bbrc.2008.12.011.

(84) Roth, M.; Selz, L. Molecular titration as a means of calibrating enzyme reference materials. *Clin Chim Acta* **1988**, *173* (1), 27-34. DOI: 10.1016/0009-8981(88)90354-3.

(85) Beal, J. Signal-to-Noise Ratio Measures Efficacy of Biological Computing Devices and Circuits. *Front Bioeng Biotechnol* **2015**, *3*, 93. DOI: 10.3389/fbioe.2015.00093.

(86) Barker, L. E.; Shaw, K. M. Best (but oft-forgotten) practices: checking assumptions concerning regression residuals. *Am J Clin Nutr* **2015**, *102* (3), 533-539. DOI: 10.3945/ajcn.115.113498.

(87) Peters, B. P.; Aronson, N. N. Reactivity of the sialic acid derivative 5-acetamido-3,5-dideoxy-L-arabino-heptulosonic acid in the resorcinol and thiobarbituric acid assays. *Carbohydrate Research* **1976**, *47* (2), 345-353. DOI: https://doi.org/10.1016/S0008-6215(00)84204-4.

(88) Cao, C.; Wang, W. J.; Huang, Y. Y.; Yao, H. L.; Conway, L. P.; Liu, L.; Voglmeir, J. Determination of Sialic Acids in Liver and Milk Samples of Wild-type and CMAH Knock-out Mice. *J Vis Exp* **2017**, (125). DOI: 10.3791/56030.

(89) Warren, L. The thiobarbituric acid assay of sialic acids. *J Biol Chem* **1959**, *234* (8), 1971-1975.

(90) Aminoff, D. Methods for the quantitative estimation of N-acetylneuraminic acid and their application to hydrolysates of sialomucoids. *Biochem J* **1961**, *81* (2), 384-392. DOI: 10.1042/bj0810384.

(91) Sobenin, I. A.; Tertov, V. V.; Orekhov, A. N. Optimization of the assay for sialic acid determination in low density lipoprotein. *J Lipid Res* **1998**, *39* (11), 2293-2299.
(92) Svennerholm, L. Quantitive estimation of sialic acids: II. A colorimetric resorcinol-hydrochloric acid method. *Biochimica et Biophysica Acta* **1957**, *24*, 604-611. DOI: https://doi.org/10.1016/0006-3002(57)90254-8.
(93) Paerels, G. B.; Schut, J. The mechanism of the periodate-thiobarbituric acid reaction of sialic acids. *Biochem J* **1965**, *96* (3), 787-792. DOI: 10.1042/bj0960787.
(94) Armbruster, D. A.; Pry, T. Limit of blank, limit of detection and limit of quantitation. *Clin Biochem Rev* **2008**, *29 Suppl 1* (Suppl 1), S49-52.
(95) Brunetti, B. About Estimating the Limit of Detection by the Signal to Noise Approach. *Pharmaceutica Analytica Acta* **2015**, *06*. DOI: 10.4172/2153-2435.1000355.
(96) Cheeseman, J.; Kuhnle, G.; Spencer, D. I. R.; Osborn, H. M. I. Assays for the identification and quantification of sialic acids: Challenges, opportunities and future perspectives. *Bioorg Med Chem* **2021**, *30*, 115882. DOI: 10.1016/j.bmc.2020.115882.
(97) Xu, G.; Potter, J. A.; Russell, R. J.; Oggioni, M. R.; Andrew, P. W.; Taylor, G. L. Crystal structure of the NanB sialidase from Streptococcus pneumoniae. *J Mol Biol* **2008**, *384* (2), 436-449. DOI: 10.1016/j.jmb.2008.09.032.
(98) Boulton, S.; Selvaratnam, R.; Blondeau, J. P.; Lezoualc'h, F.; Melacini, G. Mechanism of Selective Enzyme Inhibition through Uncompetitive Regulation of an Allosteric Agonist. *J Am Chem Soc* **2018**, *140* (30), 9624-9637. DOI: 10.1021/jacs.8b05044.
(99) Ramsay, R. R.; Tipton, K. F. Assessment of Enzyme Inhibition: A Review with Examples from the Development of Monoamine Oxidase and Cholinesterase Inhibitory Drugs. *Molecules* **2017**, *22* (7). DOI: 10.3390/molecules22071192.
(100) Kedderis, G. L. 1.07 - Biotransformation of Toxicants. In *Comprehensive Toxicology (Second Edition)*, McQueen, C. A. Ed.; Elsevier, 2010; pp 137-151.
(101) Copeland, R. A.; Pompliano, D. L.; Meek, T. D. Drug-target residence time and its implications for lead optimization. *Nat Rev Drug Discov* **2006**, *5* (9), 730-739. DOI: 10.1038/nrd2082.

(102) Park, A.; Joo, M.; Kim, K.; Son, W. J.; Lim, G.; Lee, J.; Kim, J. H.; Lee, D. H.; Nam, S. A comprehensive evaluation of regression-based drug responsiveness prediction models, using cell viability inhibitory concentrations (IC50 values). *Bioinformatics* **2022**, *38* (10), 2810-2817. DOI: 10.1093/bioinformatics/btac177.
(103) Kalliokoski, T.; Kramer, C.; Vulpetti, A.; Gedeck, P. Comparability of mixed IC(5)(0) data - a statistical analysis. *PLoS One* **2013**, *8* (4), e61007. DOI: 10.1371/journal.pone.0061007.
(104) Swinney, D. C. Biochemical mechanisms of drug action: what does it take for success? *Nat Rev Drug Discov* **2004**, *3* (9), 801-808. DOI: 10.1038/nrd1500.
(105) Sebaugh, J. L. Guidelines for accurate EC50/IC50 estimation. *Pharm Stat* **2011**, *10* (2), 128-134. DOI: 10.1002/pst.426.
(106) Caldwell, G. W.; Yan, Z.; Lang, W.; Masucci, J. A. The IC(50) concept revisited. *Curr Top Med Chem* **2012**, *12* (11), 1282-1290. DOI: 10.2174/156802612800672844.
(107) Burlingham, B. T.; Widlanski, T. S. An Intuitive Look at the Relationship of Ki and IC50: A More General Use for the Dixon Plot. *Journal of Chemical Education* **2003**, *80* (2), 214. DOI: 10.1021/ed080p214.
(108) Tagarro, A.; Cruz-Canete, M.; Otheo, E.; Launes, C.; Couceiro, J. A.; Perez, C.; Alfayate, S.; en nombre del Grupo de trabajo de Infecciones Respiratorias de la Sociedad Espanola de Infectologia, P.; Miembros del Grupo de Trabajo de Infecciones Respiratorias de la, S. q. h. p. e. l. r. d. m. [Oseltamivir for the treatment of influenza in children and adolescents]. *An Pediatr (Engl Ed)* **2019**, *90* (5), 317 e311-317 e318. DOI: 10.1016/j.anpedi.2019.01.009.
(109) Peramivir for influenza. *Aust Prescr* **2019**, *42* (4), 143. DOI: 10.18773/austprescr.2019.047.
(110) Shi, Y.; Si, L.; Han, X.; Fan, Z.; Wang, S.; Li, M.; Sun, J.; Zhang, Y.; Zhou, D.; Xiao, S. Synthesis of novel pentacyclic triterpene-Neu5Ac2en derivatives and investigation of their in vitro anti-influenza entry activity. *Medchemcomm* **2017**, *8* (7), 1531-1541. DOI: 10.1039/c7md00245a.
(111) Braun, E.; Gilmer, J.; Mayes, H. B.; Mobley, D. L.; Monroe, J. I.; Prasad, S.; Zuckerman, D. M. Best Practices for Foundations in Molecular Simulations [Article v1.0]. *Living J Comput Mol Sci* **2019**, *1* (1). DOI: 10.33011/livecoms.1.1.5957.
(112) Smith, M. Nobel lecture. Synthetic DNA and biology. *Biosci Rep* **1994**, *14* (2), 51-66. DOI: 10.1007/BF01210301.

(113) Lonsdale, R.; Ranaghan, K. E.; Mulholland, A. J. Computational enzymology. *Chem Commun (Camb)* **2010**, *46* (14), 2354-2372. DOI: 10.1039/b925647d.
(114) Jensen, F. *Introduction to Computational Chemistry*; 2016.
(115) MacDonald, I. L. Does Newton-Raphson really fail? *Stat Methods Med Res* **2014**, *23* (3), 308-311. DOI: 10.1177/0962280213497329.
(116) Hollingsworth, S. A.; Dror, R. O. Molecular Dynamics Simulation for All. *Neuron* **2018**, *99* (6), 1129-1143. DOI: 10.1016/j.neuron.2018.08.011.
(117) Lamberti, V. E.; Fosdick, L. D.; Jessup, E. R.; Schauble, C. J. C. A Hands-On Introduction to Molecular Dynamics. *Journal of Chemical Education* **2002**, *79* (5), 601. DOI: 10.1021/ed079p601.
(118) Colombo, L. Molecular Dynamics Calculations. In *Encyclopedia of Condensed Matter Physics*, Bassani, F., Liedl, G. L., Wyder, P. Eds.; Elsevier, 2005; pp 1-8.
(119) Rozmanov, D.; Kusalik, P. G. Robust rotational-velocity-Verlet integration methods. *Phys Rev E Stat Nonlin Soft Matter Phys* **2010**, *81* (5 Pt 2), 056706. DOI: 10.1103/PhysRevE.81.056706.
(120) Nam, K. Acceleration of Ab Initio QM/MM Calculations under Periodic Boundary Conditions by Multiscale and Multiple Time Step Approaches. *Journal of Chemical Theory and Computation* **2014**, *10* (10), 4175-4183. DOI: 10.1021/ct5005643.
(121) Vasilevskaya, T.; Thiel, W. Periodic Boundary Conditions in QM/MM Calculations: Implementation and Tests. *J Chem Theory Comput* **2016**, *12* (8), 3561-3570. DOI: 10.1021/acs.jctc.6b00269.
(122) Rebolini, E.; Baardsen, G.; Hansen, A. S.; Leikanger, K. R.; Pedersen, T. B. Divide–Expand–Consolidate Second-Order Møller–Plesset Theory with Periodic Boundary Conditions. *Journal of Chemical Theory and Computation* **2018**, *14* (5), 2427-2438. DOI: 10.1021/acs.jctc.8b00021.
(123) Vanommeslaeghe, K.; MacKerell, A. D. CHARMM additive and polarizable force fields for biophysics and computer-aided drug design. *Biochimica et Biophysica Acta (BBA) - General Subjects* **2015**, *1850* (5), 861-871. DOI: https://doi.org/10.1016/j.bbagen.2014.08.004.
(124) Vanommeslaeghe, K.; Hatcher, E.; Acharya, C.; Kundu, S.; Zhong, S.; Shim, J.; Darian, E.; Guvench, O.; Lopes, P.; Vorobyov, I.; Mackerell, A. D., Jr. CHARMM general force field: A force

field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. *J Comput Chem* **2010**, *31* (4), 671-690. DOI: 10.1002/jcc.21367.
(125) Bjelkmar, P.; Larsson, P.; Cuendet, M. A.; Hess, B.; Lindahl, E. Implementation of the CHARMM Force Field in GROMACS: Analysis of Protein Stability Effects from Correction Maps, Virtual Interaction Sites, and Water Models. *Journal of Chemical Theory and Computation* **2010**, *6* (2), 459-466. DOI: 10.1021/ct900549r.
(126) Kurth, S.; Marques, M. A. L.; Gross, E. K. U. Density-Functional Theory. In *Encyclopedia of Condensed Matter Physics*, Bassani, F., Liedl, G. L., Wyder, P. Eds.; Elsevier, 2005; pp 395-402.
(127) Gaus, M.; Cui, Q.; Elstner, M. DFTB3: Extension of the self-consistent-charge density-functional tight-binding method (SCC-DFTB). *J Chem Theory Comput* **2012**, *7* (4), 931-948. DOI: 10.1021/ct100684s.
(128) van der Kamp, M. W.; Mulholland, A. J. Combined quantum mechanics/molecular mechanics (QM/MM) methods in computational enzymology. *Biochemistry* **2013**, *52* (16), 2708-2728. DOI: 10.1021/bi400215w.
(129) de, M. S. G.; Walker, R. C.; Elstner, M.; Case, D. A.; Roitberg, A. E. Implementation of the SCC-DFTB method for hybrid QM/MM simulations within the amber molecular dynamics package. *J Phys Chem A* **2007**, *111* (26), 5655-5664. DOI: 10.1021/jp070071l.
(130) Baseden, K. A.; Tye, J. W. Introduction to Density Functional Theory: Calculations by Hand on the Helium Atom. *Journal of Chemical Education* **2014**, *91* (12), 2116-2123. DOI: 10.1021/ed5004788.
(131) Nam, K.; Gao, J.; York, D. M. An Efficient Linear-Scaling Ewald Method for Long-Range Electrostatic Interactions in Combined QM/MM Calculations. *J Chem Theory Comput* **2005**, *1* (1), 2-13. DOI: 10.1021/ct049941i.
(132) Stanzione, F.; Giangreco, I.; Cole, J. C. Use of molecular docking computational tools in drug discovery. *Prog Med Chem* **2021**, *60*, 273-343. DOI: 10.1016/bs.pmch.2021.01.004.
(133) Onodera, K.; Satou, K.; Hirota, H. Evaluations of molecular docking programs for virtual screening. *J Chem Inf Model* **2007**, *47* (4), 1609-1618. DOI: 10.1021/ci7000378.
(134) Brooijmans, N.; Kuntz, I. D. Molecular recognition and docking algorithms. *Annu Rev Biophys Biomol Struct* **2003**, *32*, 335-373. DOI: 10.1146/annurev.biophys.32.110601.142532.

(135) Friesner, R. A.; Murphy, R. B.; Repasky, M. P.; Frye, L. L.; Greenwood, J. R.; Halgren, T. A.; Sanschagrin, P. C.; Mainz, D. T. Extra precision glide: docking and scoring incorporating a model of hydrophobic enclosure for protein-ligand complexes. *J Med Chem* **2006**, *49* (21), 6177-6196. DOI: 10.1021/jm051256o.
(136) Forouzesh, N.; Mishra, N. An Effective MM/GBSA Protocol for Absolute Binding Free Energy Calculations: A Case Study on SARS-CoV-2 Spike Protein and the Human ACE2 Receptor. *Molecules* **2021**, *26* (8). DOI: 10.3390/molecules26082383.
(137) Schneider, G.; Bohm, H. J. Virtual screening and fast automated docking methods. *Drug Discov Today* **2002**, *7* (1), 64-70. DOI: 10.1016/s1359-6446(01)02091-8.
(138) Ylilauri, M.; Pentikäinen, O. T. MMGBSA As a Tool To Understand the Binding Affinities of Filamin–Peptide Interactions. *Journal of Chemical Information and Modeling* **2013**, *53* (10), 2626-2633. DOI: 10.1021/ci4002475.
(139) Schramm, V. L. Enzymatic Transition States and Drug Design. *Chem Rev* **2018**, *118* (22), 11194-11258. DOI: 10.1021/acs.chemrev.8b00369.
(140) Barnett, C. B.; Naidoo, K. J. Free Energies from Adaptive Reaction Coordinate Forces (FEARCF): an application to ring puckering. *Molecular Physics* **2009**, *107* (8-12), 1243-1250. DOI: 10.1080/00268970902852608.
(141) Eberhardt, J.; Santos-Martins, D.; Tillack, A. F.; Forli, S. AutoDock Vina 1.2.0: New Docking Methods, Expanded Force Field, and Python Bindings. *J Chem Inf Model* **2021**, *61* (8), 3891-3898. DOI: 10.1021/acs.jcim.1c00203.
(142) Friesner, R. A.; Banks, J. L.; Murphy, R. B.; Halgren, T. A.; Klicic, J. J.; Mainz, D. T.; Repasky, M. P.; Knoll, E. H.; Shelley, M.; Perry, J. K.; et al. Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *J Med Chem* **2004**, *47* (7), 1739-1749. DOI: 10.1021/jm0306430.