# Dynamics and selectivity of the NleB and SseK effectors studied by NMR spectroscopy and MD simulation


Thomas L. Hicks

*Thesis submitted in fulfilment of the requirement for the degree of*

*Doctor of Philosophy*

*December 2022*

# To Keisha,

My only reason.

# Abstract

The NleB and SseK effectors are a unique series of enzymes which act as glycosyltransferases (GT), transferring a single GlcNAc ring from donor UDP-GlcNAc to arginine residues. This activity is unique as GT activity is typically limited to serine and threonine, in the case of O-glycosyltransferases (O-GT), and asparagine, in the case of N-Glycosyltransferases (N-GT). These effectors are used by *E. coli* and *S. enterica,* affording the ability to disrupt the innate immune system response by inhibiting key protein-protein interactions in the NF-κB signalling pathway within host cells. To date a complete understanding of these effectors has not been found, with their selectivity, mechanism of action and dynamics not fully understood.

There are two main perspectives this thesis looks to explore. How do these effectors discern between acceptor substrates and how do they recognise their donor substrate? These two areas encompass the protein-protein and protein-ligand interactions that these enzymes have when performing catalysis. With an understanding of these interactions, it would be possible to begin to unravel the complexity of these enzymes and take steps towards disrupting their activity in the development of medicines to combat the bacteria that utilise them.

With this thesis, MD simulation, two-dimensional NMR and STD NMR were used to understand the dynamics and selectivity of these enzymes. Through a mutagenesis study it was possible to begin to understand the acceptor interface of NleB1, providing a structural rationale for a single tyrosine residue in the recognition of the acceptor substrate FADD. With the study of the binding of UDP-GlcNAc and other Leloir donors it is possible to see how they are recognised and discern different recognition profiles between the SseK effectors, SseK1 and SseK2. With these results, this thesis contributes to an understanding of the Dynamics and selectivity of the NleB and SseK effectors.

# List of Contents

# List of Figures and Tables

**CHAPTER 1 INTRODUCTION**

**CHAPTER 4 MOLECULAR BASIS FOR THE DONOR SUBSTRATE RECOGNITION OF BACTERIAL VIRULENCE EFFECTORS**



**CHAPTER 5 CONTRIBUTIONS TO OTHER WORKS**

**CHAPTER 6 APPENDIX**

# Acknowledgements

Three primary supervisors, a global pandemic and four years later I am finally finishing my PhD. While this process has been challenging, there has never been a time where I have been alone. Science is not just the single pursuit of knowledge; it is a community. This thesis is a testament to that community and all the people that I have had the privilege of working with. It is to them I owe a tremendous debt and with this work I thank them all.

Jesús, you gave me my first opportunity to work in science. I can still remember very vividly my disappointment of being rejected by every summer project I applied for in the second year of my pharmacy degree. Now as I look back on that time, I think of myself as lucky. This brought me to ask if I could work with you and through that first summer project I was first introduced to NMR, a field that I now greatly enjoy. Another summer studentship and a masters project later, the work conducted by your research group inspired me to pursue a PhD. Again, you were there to take me on. I cannot properly thank you enough for the support and opportunities you have given me, I can only strive to pay them forward in the future.

Chris, you were a tremendous help and I thank you greatly for your support and patience. I am told it is not common for there to be a change in supervision during a PhD, but with you and Jesús that was never going to be an issue. You took responsibility as my primary supervisor and welcomed me into your research group. Under your tenure and with your help I became a more rounded scientist. With you, I worked through some difficult challenges, from the expression and purification of proteins to the writing of this very thesis. I was always in good hands; I could not be more grateful for all that you did.

Iñaki, you are a good man. Even in the short time you have been my supervisor you have done so much for me. As is the same with Jesús and Chris, I struggle to properly come up with the words to say how fortunate I am to have had your help. Since you joined our research group, we have worked on many projects together and I have learned so much from you. You took on the role of my supervisor at a difficult time for me. There was a lot of stress around the writing of this thesis but with your guidance I became more confident. I hope to have produced a work worthy of your effort.

Now, I must take the time to thank my research group. All of those that work in the NMR department at UEA have been instrumental over the course of my PhD. I would like to thank Alex, Sam, Serena, and Trey. I can safely say that before meeting the four of you I knew nothing about how to use a spectrometer. You were all important when I was just getting started and I am lucky to have had you there to help with anything I needed. To Haider, Hassan, Jake, James, Joe, Juan Carlos, Liam, Matthew, Shannon, Tharin, Valeria, Vani, and Yaroslav. I thank you for your company through my time at UEA, you were all brilliant people to work with. I would also like to thank Colin and the technical team in Chemistry and Pharmacy, your hard work was not unnoticed. I am tremendously grateful for all that you do to support us in our research work at UEA.

To my mother, Amanda. In a sense you are the one that it truly responsible for this work. For everything you have done for me, there are no words that can properly thank you. In their place I can only do whatever I can to make you proud and to make every sacrifice you have made for me matter. You have asked me before if you are a good parent, I think it is obvious you are. I am beyond fortunate to be your son, you made me the person that I am today and every one of my achievements are yours.

Finally, I want to thank my partner, Keisha. This thesis is an achievement I want to share with you. With you in my mind I had the will to work through the hardest challenges and most stressful times. You are my only reason for doing any of this. When I began my PhD, it was something I undertook selfishly, something I could enjoy for myself. Now, with you, I have someone to share this with. I never took on this PhD for the title or the prestige, those things do not matter. The real meaning of all this effort is the opportunities it will give me to build a perfect life together with you. With this part of my life is coming to an end, though it is not possible to know what the future will hold, with you by my side I am certain it will be something to look forward to.

They say a person is only as good as the company they keep. I am not sure I am as good as all the people I have around me, but if I even come close then I must be doing something right.

# CHAPTER 1

## INTRODUCTION

# 1.1 Carbohydrates

A carbohydrate is defined as a molecule composed of carbon, hydrogen and oxygen and originally considered to have the empirical formula $C(H_2O)_n$. Though this general rule is useful, as with many others, there are exceptions. In truth, all carbohydrates are part of a diverse family of molecules with far reaching implications; from simple sugars needed for our diet to the fundamental biological processes essential to life (1). Because of this, the study of carbohydrates is compelling to any scientist, be that an organic chemist synthesising new pharmaceutics or an enzymologist uncovering the details of signalling pathways. For this thesis, the latter is of most interest.

## 1.1.1 The Importance of Carbohydrates in Biology

Despite their simplicity, carbohydrates are often involved in highly specific interactions. Carbohydrates often find themselves part of many roles in the body. Important to this thesis, they play a crucial role in the fine biomolecular machinery that underpins life, cell signalling (1). Even though this topic encompasses an enormous amount of biology, carbohydrates can be found all over. On the cell surface, complex glycans are responsible for cells directly interfacing with one another (2). These molecules interact with glycan-recognising proteins named lectins (3). Interactions of this type are prevalent in the innate immune system, allowing the body to discern bodily cells from those of invading pathogens. An important family of lectins that undertakes this responsibility are the mannose-binding lectins (MBL) (4). These proteins can discern between host and pathogen through Ca(II)-dependent binding to different glycan motifs. MBL binding sites selectively bind to the terminal monosaccharide units of glycans often present on the surface of invading pathogens but not to those often present on human cells. The specificity of these lectins arises from a conserved Glu-Pro-Asn motif that selectively binds to monosaccharide units that contain equatorial hydroxyl groups at C3 and C4 (2, 4). This motif selectively recognises D-mannose, L-fucose, and N-acetyl-D-glucosamine (GlcNAc) but do not bind to sialic acid or D-galactose, which are often the terminal units of human cell surface glycans (4).

The role of carbohydrates in cell signalling also extends to within the cell. Many pathways require specific protein-protein interactions. These interactions can be modulated by the addition of single or multiple monosaccharide units through a process called glycosylation (5). One such example of this is the very early transcription factors to have been found to be a target of glycosylation, c-Myc (6). Glycosylation modulates c-Myc turnover by inhibiting its degradation (Figure 1.1). Within the cell, c-Myc is an unstable protein, only existing within the cell for around 30 minutes (6). The reason for this short life span is that c-Myc is selectively phosphorylated at Thr58 by glycogen synthase kinase 3 beta (GSK3β) leading to ubiquitin degradation (6, 7). The addition of a single GlcNAc ring at Thr58 by protein O-GlcNAc transferase (OGT) prevents this phosphorylation, extending the lifespan of c-Myc and allowing it to participate in downstream signalling. Once c-Myc has participated in signalling, the removal of the GlcNAc ring can occur, allowing phosphorylation of Thr58 and subsequent ubiquitination (6, 7). Through just the addition and removal of a single glycan ring, it can be seen here that carbohydrates can play a pivotal role in crucial cellular activity.

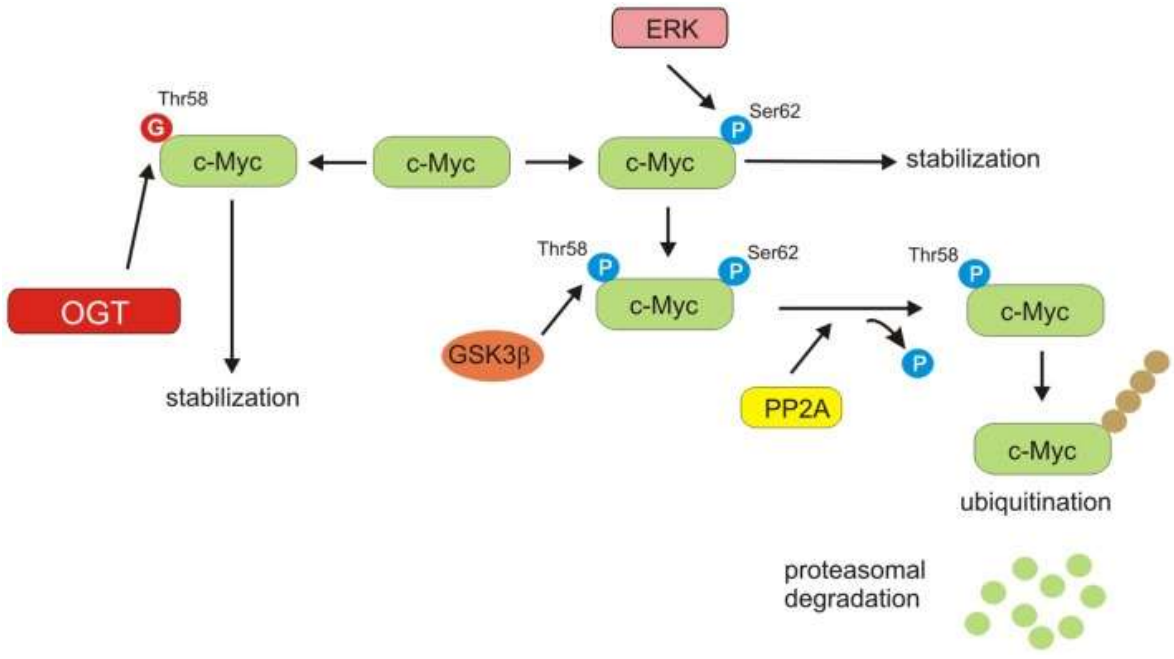



*Figure 1.1 - The c-Myc degradation pathway. c-Myc is stabilised by OGT by glycosylation of Thr58. This is achieved by blocking phosphorylation by GSK3β on Thr58 which would lead to polyubiquitination and subsequent degradation. Taken from (7).*

Many signalling pathways that involve carbohydrates can be responsible for disease, due to environmental stress or genetics (8, 9). Because of this, carbohydrates often provide platform for the development of small molecule drugs. Many carbohydrates and carbohydrate-conjugate molecules act as agonists or antagonists to many cell receptors as well as acceptor or donor substrates to many enzymes. The study of such systems has led to the development of medical interventions that are in use today, spanning cardiovascular therapies to anticancer medicines as well as being used to treat invasive pathogens and the diseases they can cause (10).

## 1.1.2 Carbohydrates and Their Rich Chemical Diversity

Despite their simple chemical composition, carbohydrates are a complex series of structures that have varied chemical properties (11, 12). In nature, carbohydrates are often the sum of individual units, called monosaccharides. These carbohydrate units are typically aldoses and ketoses (Figure 1.2), which are chains of carbons bearing hydroxyl groups with a single carbonyl group. In the case that the carbonyl group is on the end of the chain the molecule will have an aldehyde functional group and therefore be an aldose and if carbonyl group is not at the end of the chain, the molecule will have a ketone functional group and will be a ketose.

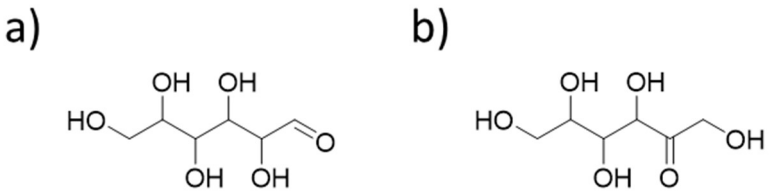



*Figure 1.2 - Example of an aldose and ketose, showing the generic structure of an aldohexose (a) and a ketohexose (b).*

For each carbon in an aldose and ketose, excluding the carbonyl carbon, there is a stereocenter. Because of this, aldoses and ketoses can have their hydroxy groups arranged in different ways to form stereoisomers. This means that there can be a lot of variation for a particular aldose or ketose, which becomes even more diverse if hydroxyl group substitutions are considered. Some of these stereoisomers will be mirror images of each other or, enantiomers. Such molecules will possess the same name but will be distinguished as D- or L-enantiomers, such as D-Glucose and L-Glucose.

Because they often possess many polar groups, unsubstituted carbohydrates are typically soluble in water (13). In aqueous solution, aldoses and ketoses do not stay as linear hydrocarbons. Instead, they undergo an intramolecular ring closing reaction to form a hemi acetal. In the case of 6-carbon chain aldoses and ketoses, the ring closure produces a 6-member pyranose or a 5-member furanose ring, respectively, to form monosaccharides. Much of this discussion will focus on pyranoses though a lot of the same principles do apply to furanoses. The position where the ring closure occurs on is called the anomeric carbon. This position is important in monosaccharides as it has many special properties. The hydroxyl group attached to the anomeric carbon can either be axial or equatorial relative to the ring and conformations are denoted the α- the or β- conformation. This is determined by comparing the chirality of the anomeric carbon to the chirality of the carbon on the opposite side of the endocyclic oxygen (12). If the chirality is the same, it is the α-anomer and if they are opposite, the β-anomer.

The anomeric conformation of the monosaccharide is not stable in solution. A monosaccharide with an unconjugated anomeric hydroxyl will be in equilibrium between its ring open and ring closed state (Figure 1.3). During the ring closing reaction the hydroxyl group can attack from either plane of the monosaccharide forming either an α- or β-anomer. This results in their being an equilibrium between the two anomeric conformations as the monosaccharide undergoes subsequent ring opening and ring closure reactions. The equilibria between the α- or β-anomer is often not equal and can be influenced greatly by many different factors, including the nature of the solution and the substituents on the ring. Intuitively for a pyranose it could be expected that the β-anomer would be preferred as the hydroxyl is directed away from other bulky functional groups, whereas the α-anomer would clash with diaxial functional groups. However, there are cases where the α-anomer is preferred such as mannose (14).


*Figure 1.3 - Equilibrium between the ring open (right) and ring closed (left) states of glucose. During the forward reaction, the endocyclic oxygen (blue) forms a hydroxyl group and the anomeric hydroxyl group (red) forms the aldehyde carbonyl. During the reverse reaction the hydroxyl group can attack either face of the carbonyl to form the α- or β-anomer.*

Pyranoses can also adopt different ring conformations with the chair conformer (C) being usually preferred (12). The C conformer can be denoted $^{4}C_{1}$ and $^{1}C_{4}$, with the $^{4}C_{1}$ conformer being most common. This notation denotes the carbon that is above and below the reference plane for the ring defined by C2, C3, C5 and the endocyclic oxygen. The reason the chair conformer is preferred is because it minimises steric clashes within the ring. As is true for the anomeric conformation, this is not always the case and ring substitutions can play a role in determining the ring conformation of a sugar. Some examples of other ring conformations for a pyranose ring (Figure 1.4) include the boat (B), envelope (E), half-chair (H) and skew (S).

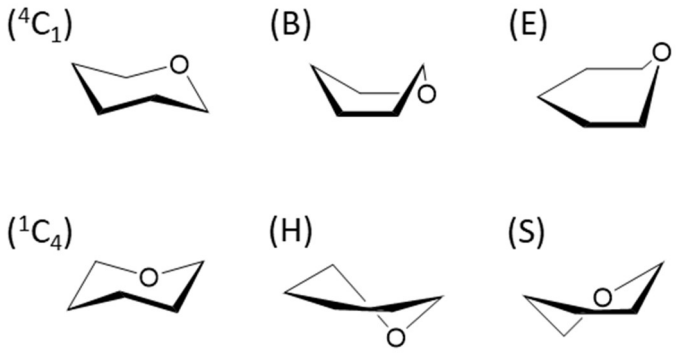



*Figure 1.4 - A sample of ring conformations for pyranose including the $^{4}C_{1}$ Chair ($^{4}C_{1}$), $^{1}C_{4}$ Chair, ($^{1}C_{4}$), boat (B), envelope (E), half-chair (H) and skew (S).*

There are many ways that pyranose monosaccharides can exist, providing a high amount of structural and chemical variety through the different substitutions, isomerisation, and three-dimensional conformation. Despite this potential for variety, nature has narrowed its selection to almost exclusively the D-enantiomeric configuration of these molecules in mostly the $^4C_1$ conformation. Furthermore, this selection is narrowed to only nine monosaccharides which are commonly observed in vertebrates (Figure 1.5) (12).

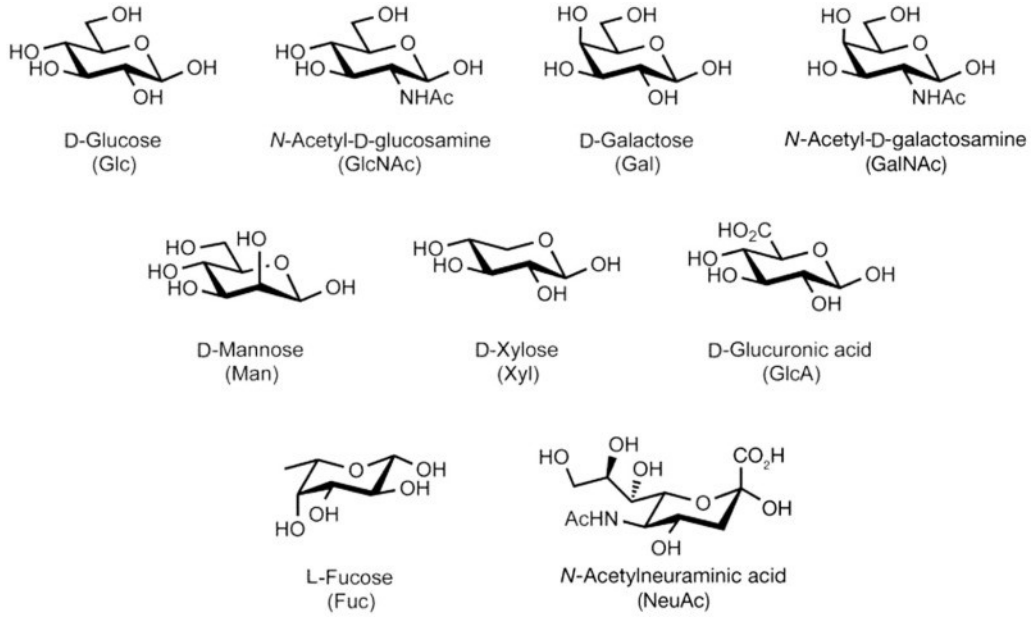



*Figure 1.5 - The nine naturally occurring monosaccharides found in vertebrate. Of the nine only one is an example of an L sugar (Fuc) and one is an example of a ketose (NeuAc). Taken from (12).*

Despite this limited selection, the families of molecules that are made using these monosaccharides possess incredibly varied chemical properties. When combined into chains monosaccharides form glycans, with longer chains being known as polysaccharides (11). Shorter glycans include disaccharides composed of two monosaccharides and oligosaccharides which are shorter chains composed of three to twenty monosaccharides. In the formation of polysaccharides there is a condensation of the anomeric hydroxyl group to the non-anomeric hydroxyl groups of another monosaccharide. Through this condensation, a glycosidic linkage is formed. Via this reaction long chains can be built sequentially or by the combination of many chains. Unlike DNA or proteins where linkages

occur on single specific positions, monosaccharides bear many sites which condensation can occur. This means that polysaccharides are not always linear and can branch (11, 12). The linkages between monosaccharides are noted with their α or β conformation and numbered from the anomeric carbon to the position of the ring that it is attached to. An example of this is maltose (Figure 1.6) in which there is a α-1,4 linkage between two glucose units.

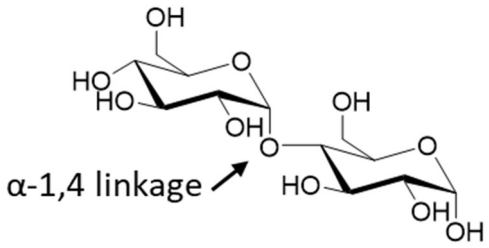



*Figure 1.6 - Linkage between the two glucose units of maltose. The α-1,4 linkage is indicated by an arrow and numbered from the anomeric carbon (position 1) of the non-reducing unit to position 4 of the reducing end unit.*

Monosaccharide units do not only form glycosidic linkages to other monosaccharides but also to other chemical structures such as amino acids in proteins and can be seen to form bonds to phosphorous, sulphur, nitrogen, and oxygen in P-, S,- N-, and O- linkages (15). Nucleobases are one such example of this in which there is a bond between the furan ribose with purine and pyrimidines via an N-linkage. In these cases, the numbering follows a similar convention as for polysaccharide linkages.

### 1.1.3 The Modification of Proteins with Carbohydrates

Carbohydrates are not only conjugated together or with simple organic molecules but are also incorporated into proteins in a post translational modification called glycosylation (15). These modifications can vary in form and their effect on proteins, including the modulation of their action, quality control and even acting as molecular anchors (15-17). In proteins, glycosylation takes two main forms, O-glycosylation, on oxygen-bearing amino acids including serine and threonine (18) and N-glycosylation which historically has been limited to asparagine (19). In recent years however, the N-Glycosylation of proteins arginine residues has been observed and will be discussed more later.

Glycosylation is performed in the body by enzymes called glycosyltransferases (GT). Enzymes of this class use sugar nucleotides as donors. These molecules are single monosaccharide units conjugated to typically diphospho- and triphospho- nucleobases such as in UDP or GTP. GTs can be classified as either an O-glycosyltransferase (O-GT), or an N- glycosyltransferase (N-GT) depending on the residue that they can act on. They are also further classified by whether they retain or invert the configuration at the anomeric centre of the monosaccharide from the sugar nucleotide donor. Inversion is considered the more straightforward of the two mechanisms, being a nucleophilic attack, in a reaction akin to $S_N2$-substitution (Figure 1.7). Retention though, is still to this day the subject of debate as the exact details of the rection intermediates and how they are stabilised by the enzyme are much more complex (17). Despite this complexity there are some proposed mechanisms for retaining GTs (Figure 1.7) (20).

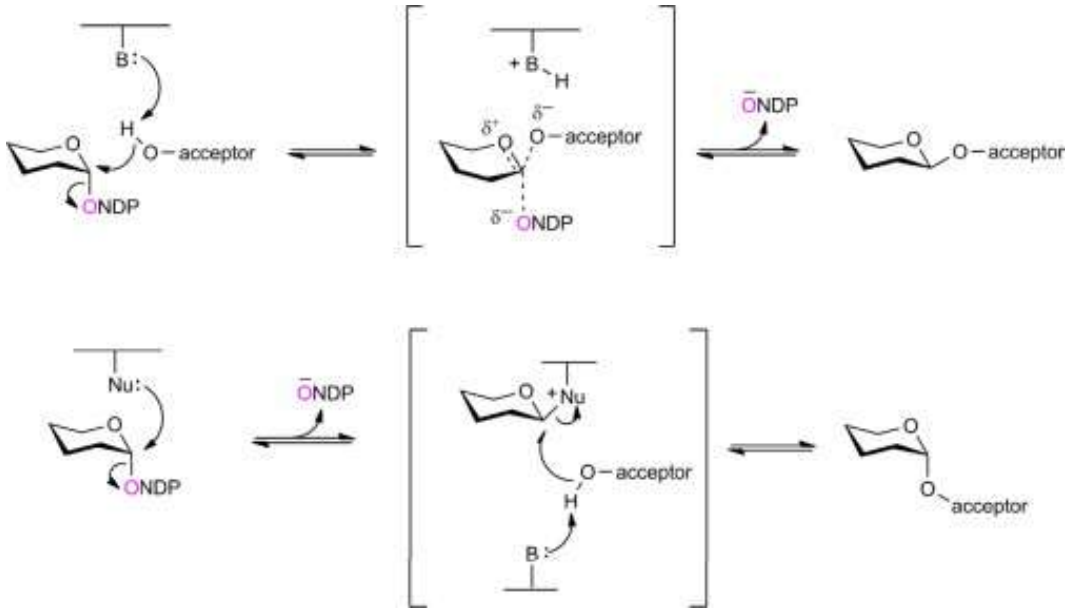



*Figure 1.7 - The inverting (top) and retaining (bottom) mechanisms for GTs. In the inverting mechanism a base catalysed $S_N2$ type reaction is shown. The inverting mechanism proposes a double inversion mechanism to retain the anomeric configuration. This mechanism proposes an inverted, covalently bound, intermediate though a consensus on exact inversion mechanisms has not been found. Taken from (21).*

There are many families of GTs based on their sequence, encompassing both inverting and retaining mechanisms. However, there are only a handful of GT superfamilies based on their folding structure. These include GT-A through to GT-E (Figure 1.8) (22), though until recently GTs were only found to show GT-A or GT-B folding. The GT-A fold consists of two pairs of closes associated, alternating, α-helices and β-sheets called Rossman folds containing nucleotide binding domains. GT-B folded GTs contain two Rossman folds which face one-another, connected by a flexible linker (22).

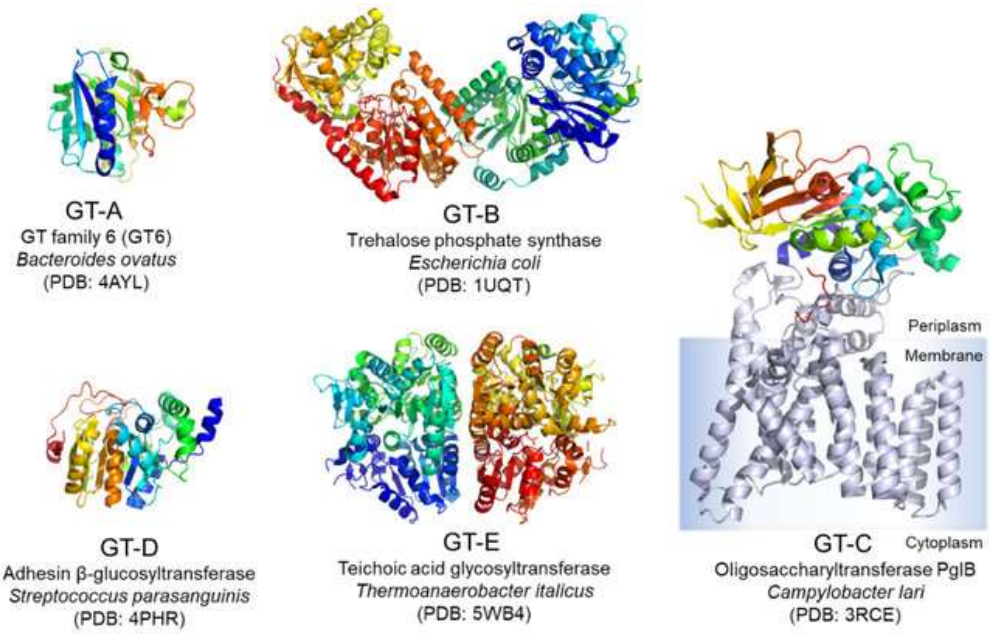



*Figure 1.8 - The five classes of glycosyltransferases based on folding, GT-A through to GT-E. The most observed classes of glycosyltransferases are GT-A and GT-B. Taken from (22).*

GT-A GTs typically bind the divalent cations Mn(II) or Mg(II). These ions are bound to a highly conserved Asp-X-Asp (DXD) motif present in the sugar nucleotide binding domain (23). The divalent cations bound to the DXD motif co-ordinate the negatively charged phosphate group of sugar nucleotides and can assist in inducing donor conformational changes necessary for catalysis. For inverting and retaining GTs there are trends in how the DXD motif is used to co-ordinate the divalent cation (24). Both aspartic acid residues co-ordinate the divalent cation in the case of inverting GTs whereas, in retaining GTs typically only one aspartic acid is used. GT-B GTs tend to be metal-ion-independent, instead favouring direct sugar nucleotide-protein contacts that are typically located in the cleft between the two opposing Rossman folds. In comparison to GT-A GTs the binding of sugar nucleotides to GT-B GTs is much more dynamic. When no sugar nucleotide is bound, GAT-B GTs are in an open conformation with the two Rossman folds less tightly associated. On binding of a sugar nucleotide GT-B GTs the protein structure adopts a closed conformation (25).

## 1.1.4 Arginine Glycosyl Transferases

Arginine GTs have, in recent years, been the subject of much interest in the field of enzyme catalysis. The classical understanding is that GTs target nucleophilic residues such as serine and threonine in the case of O-GTs (18) and asparagine in the case of N-GTs (19). Arginine however is a poor nucleophile owing to the resonance delocalisation in the guanidinium group of the residue side chain (Figure 1.9). This leaves arginine N-GTs rare with only very few known examples in nature (26). The exact catalytic mechanism of action and dynamics of these GTs is currently not completely understood, owing to their recent discovery and rarity.

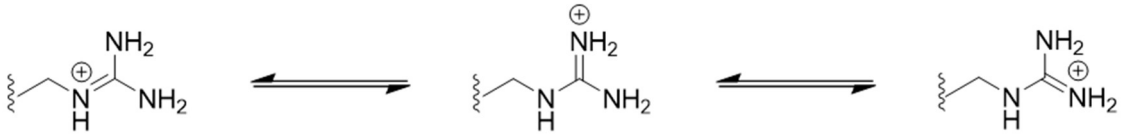


*Figure 1.9 - Resonance delocalisation of guanidinium. At physiological pH arginine side chains will be protonated and there will be resonance delocalisation of the guanidinium function group making for a weaker nucleophile.*

Bacterial arginine GTs have been observed to be expressed by clinically relevant gram-negative bacteria including *P. aeruginosa*, *E. coli, C. rodentium* and *S. enterica (27-30)*. The pathogens *P. aeruginosa*. *E. coli* and *S. enterica* all display pathogenicity in humans while *C. rodentium* is typically limited to murine hosts. *P. aeruginosa* is an opportunistic pathogen that has been identified to be a risk for hospitalised patients and those that suffer with cystic fibrosis (31). Virulent strains of *E. coli* are often food or, waterborne and cause infections within the gut resulting in severe diarrhoea. Though considered a lower risk in the developed world, in developing countries and areas with limited access to medical care, *E. coli* poses a serious public health concern (32). *S. enterica* is alongside *E. coli* as a major contributor to diarrhoeal disease by foodborne infection. Antibiotic resistance in these pathogens is a serious and rising danger with resistant strains being highlighted by the World Health Organisation as in critical need for the development of new antibiotics. In the microorganisms that express them arginine GTs support pathogenicity which opens a

potential avenue for the development of antibiotic and antibiotic-supportive therapies to treat the diseases that they cause. It is because of this that the study of these enzymes is important (30, 33).

*P. aeruginosa*, expresses the GT-B arginine N-GT EarP (30). This enzyme catalyses the transfer of a rhamnose monosaccharide from dTDP-beta-L-rhamnose (TDP-Rha) and is used to modulate the translation of protein essential for motility and antibiotic resistance (30). In the case of *E. coli*, *C. rodentium*, *S. enterica,* arginine GTs play a pivotal role acting as effectors in the evasion of the host innate immune system by suppressing crucial signalling pathways (29, 33-35). These immune system-supressing arginine N-GTs are all GT-A folded containing a Mn(II) co-ordinating DXD motif that catalyses the transfer of a GlcNAc monosaccharide from uridine diphosphate N-acetylglucosamine (UDP-GlcNAc) (36, 37). These effectors include two series of enzymes named NleB and SseK. There are two NleB effectors expressed by *E. coli* (both EPEC and EHEC) including NleB1 and NleB2 as well as one encoded by *C. rodentium*, named *Cr*NleB (28, 38). *S. enterica* expresses three SseK effectors including SseK1, SseK2 and SseK3 (29, 34, 38). Within both series of effectors there is a good degree of sequence identity (Table 1.1) and even between NleB and SseK effectors there is similarity (38).

*Table 1.1 - Sequence identity of the NleB and SseK effectors found by clustalW, including NleB effectors expressed by C. rodentium, E. coli, E. albertii, E. marmotae, S. enterica and Y. hibernica. Taken from (38).*

| | **NleB *C. rodentium*** | **NleB1 EPEC** | **NleB1 EHEC** | **NleB *E. albertii*** | **NleB *E. marmotae*** | **NleB2** | **SseK1** | **SseK2** | **SseK3** |
|---|---|---|---|---|---|---|---|---|---|
| NleB *C. rodentium* | | | | | | | | | |
| NleB1 EPEC | 88.75 | | | | | | | | |
| NleB1 EHEC | 89.06 | 97.87 | | | | | | | |
| NleB1 *E. albertii* | 88.45 | 97.57 | 99.09 | | | | | | |
| NleB *E. marmotae* | 71.42 | 71.73 | 71.73 | 71.43 | | | | | |
| NleB2 EPEC/EHEC | 59.82 | 61.04 | 61.35 | 60.43 | 60.43 | | | | |
| SseK1 *S. enterica* | 57.45 | 58.05 | 59.57 | 58.97 | 58.97 | 53.68 | | | |
| SseK2 *S. enterica* | 52.58 | 52.28 | 53.19 | 52.58 | 51.06 | 46.93 | 55.36 | | |
| SseK3 *S. enterica* | 51.98 | 53.19 | 53.50 | 52.89 | 54.10 | 47.55 | 54.93 | 75.22 | |
| NleB *Y. hibernica* | 32.52 | 32.83 | 34.04 | 34.04 | 34.04 | 33.13 | 35.12 | 33.14 | 33.43 |

### 1.1.4.1 The Enzymatic Behaviour of the NleB/SseK Effectors

The NleB and SseK effectors are introduced by their respective bacteria directly into host cells using the type 3 secretion system (T3SS) (38) (Figure 1.10). The T3SS is a complex structure used by gram negative bacteria that allows for the targeted delivery of virulence factors by bypassing transmembrane domains. This mechanism allows these effectors to directly disrupt host innate immune response from within the cell.

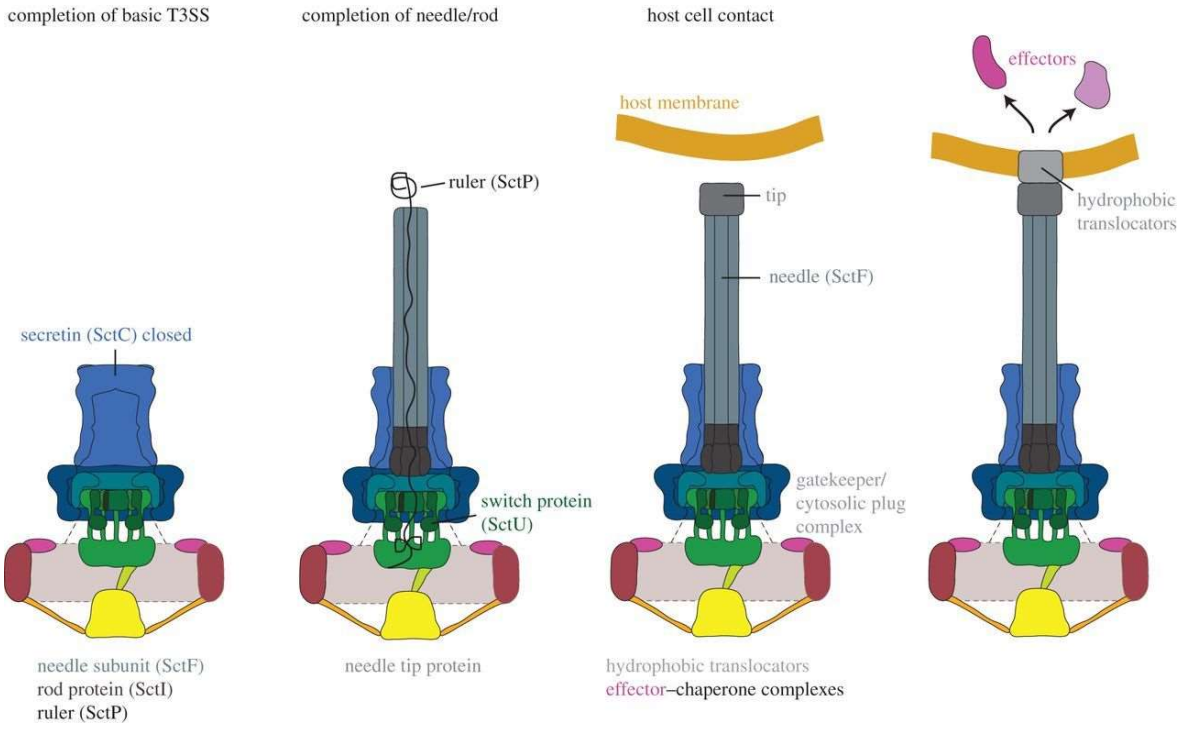



*Figure 1.10 - Schematic representation of the T3SS used by gram negative bacteria to bypass the host cell membrane. The T3SS is assembled stepwise until it nears the host cell membrane, with the distance judged by a ruler subunit. Once contact is made, hydrophobic translocators form a complex with the host cell membrane allowing for the passage of effector enzymes. Taken from (39).*

*E. coli* utilises one T3SS that acts to bypass the outer cellular membrane (40). In *S. enterica* there are two T3SSs, T3SS1 and T3SS2. T3SS1 is expressed once *S. enterica* contacts host cells and acts in an equivalent manner to the T3SS used by *E. Coli,* bypassing the outer cell membrane. Once Salmonella has invaded host cells T3SS2 is expressed which acts to bypass the membrane of cytosolic vesicles and hijack trafficking in the cell (26).

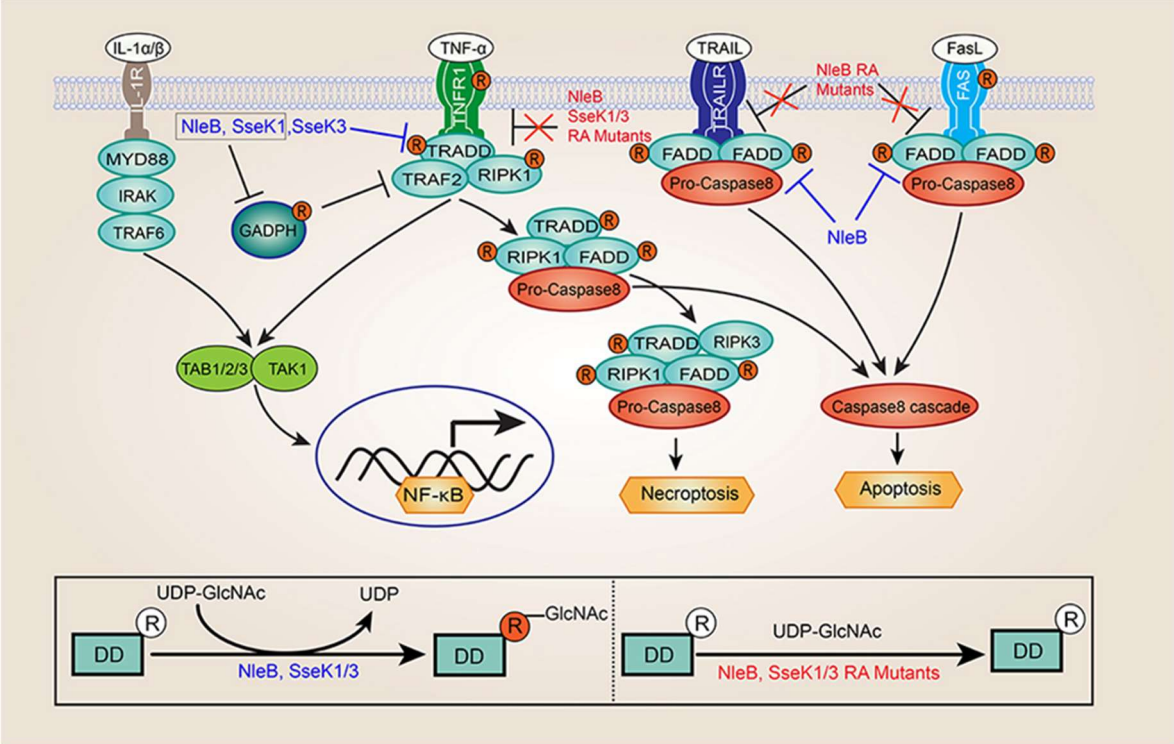



*Figure 1.11 - The inhibition pathways (top) of the NleB and SseK effectors. Once introduced to the intracellular matrix by the T3SS, the effectors can selectively glycosylate downstream signalling proteins in the TNFR-α mediated immune response. Glycosylation (bottom) inhibits the formation of complexes necessary for induction of Necroptosis, Apoptosis and NF-κB recruitment. Taken from (26).*

The signalling pathways targeted by the NleB/SseK effectors are the tumour-necrosis factor receptor (TNFR) and nuclear factor κB (NF-κB) mediated innate immune response (Figure 1.11) (26). The NleB and SseK effectors have been shown to have N-GT activity on members of these pathways important for the formations of complexes needed for the initiation of necroptosis, apoptosis, and recruitment of NF-κB. Despite their similarities, the NleB and SseK effectors all display varying levels of activity and differences in substrate specificity (29). The NleB effectors are the most promiscuous and most active displaying activity towards Fas-associated death domain (FADD), Tumour necrosis factor receptor type 1-associated death domain (TRADD), receptor interacting serine/threonine kinase 1 (RIPK1) and glyceraldehyde 3-phosphate dehydrogenase (GAPDH). NleB2 possesses much lower overall activity compared to NleB1. The SseK effectors have a greater disparity in selectivity and activity compared to the NleB effectors. SseK1 is active towards GAPDH and TRADD (29, 34), while SseK3 shows activity towards TRADD, TNFR1 and TRAILR (41). The role of SseK2 is difficult to discern but, there are some reports that SseK2 displays

very weak activity towards FADD (29). Moreover, *S. enterica* mutants with SseK1/3 deletions show little detectable arginine glycosylation activity further clouding the role that SseK2 plays (34). Recent studies have also indicated that some of the NleB and SseK enzymes may play roles outside of the suppression of the innate immune response. CrNleB has been identified to show N-glycosylation activity towards Hypoxia-inducible factor 1-alpha (HIF-1α) (42) and bacterial glutathione synthetase (GshB) (43) both of which play important roles in bacterial metabolism and management of oxidative stress.

Though varied in their selectivity the NleB and SseK effectors show common sites of glycosylation in their respective target proteins. NleB1 glycosylates TRADD Arg235 (27, 29) whereas SseK1 and SseK3 can glycosylate both Arg235/245 (34). FADD is glycosylated on Arg117 by NleB1 (28, 29). For RIPK1 the target of glycosylation by NleB1 is Arg602 (35). These residues are all positioned on the death domain (DD) of these proteins and glycosylation is thought to prevent inter-DD interactions important for signalling. GAPDH is glycosylated in two positions by NleB1, Arg 197/200 (29). Glycosylation at these positions prevents GAPDH from performing polyubiquitination on the tumour necrosis factor receptor-associated factor (TRAF), TRAF2 and TRAF3 supressing NF-κB signalling (29). In addition to glycosyltransferase activity the NleB and SseK effectors can perform arginine N-GT on themselves when over expressed (26, 36, 41) However, the rationale for this is unknown.

Between the different effectors there are varied reports on their glycosyl transfer mechanisms. One early study on the activity of the SseK effectors found that the SseK1 followed a retaining mechanism when glycosylating peptide substrates consisting of short sequences from FADD, TRADD and GAPDH (36). This was evidenced by an NMR study on the glycosidic linkage between the GlcNAc ring and arginine in the modified peptides indicated that there was a α-linkage formed (36). Another study found that SseK3 could behave as a retaining GT also (37). Given the high sequence similarity between the SseK effectors it could be concluded that they all would be retaining GTs however this has not been confirmed. For the NleB effectors one study successfully crystallised structures of the NleB1-N-glycosylated FADD, TRADD and RIPK1 death domains. In these structures the glycosidic linkage observed is not in the α-configuration but is instead of the β-configuration. This indicates that, in contrast to the SseK effectors, NleB1 behaves as an inverting glycosyltransferase (35).

### 1.1.4.2 The Structure and Dynamics of the NleB/SseK Effectors

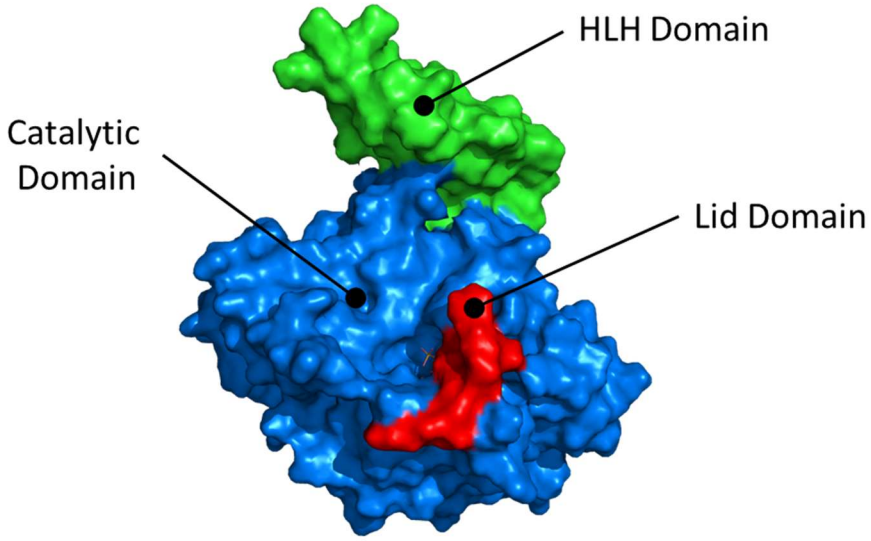



*Figure 1.12 - Surface structure of the NleB and SseK effectors, showing the HLH Domain (green), Catalytic Domain (blue) and Lid Domain (Red). The structure of SseK1 (PDB entry 5H60) was used for illustrative purposes.*

The crystal structure of many members of the NleB and SseK family effectors have been published, providing key insights into the structure of these effectors. These structures show how the NleB and SseK effectors can recognise their donor and acceptor substrates. To date the structures available include all SseK effectors including UDP-bound SseK1, UDP- and UDP-GlcNAc-bound SseK2 (36), and SseK3 in complex with UDP-GlcNAc hydrolysed to UDP and GlcNAc (37). The structure of NleB1 has also been published, including the only example to date of an NleB/SseK effector in complex with an acceptor substrate, in this case FADD (37). From these published structures it has been observed that the effectors all possess similar concave shape and three conserved domains including helix-loop-helix (HLH), catalytic and lid domain (Figure 1.12) (35-37).

The HLH domain is unique to the NleB and SseK effectors. This domain consists of a pair of antiparallel α-helices connected by a short loop that protrudes from the catalytic domain of the effector. This structure has been found by molecular dynamics (MD) simulations to be dynamic, able to move towards the substrate binding site and close around bound proteins (36). The HLH domains of the NleB and SseK effectors are proposed

to provide the unique interactions that contribute to differential recognition patterns that are observed for these effectors. The HLH domain in NleB1 can be seen to contact FADD forming complementary salt bridge interactions in line with the predicted behaviour of this domain by MD.

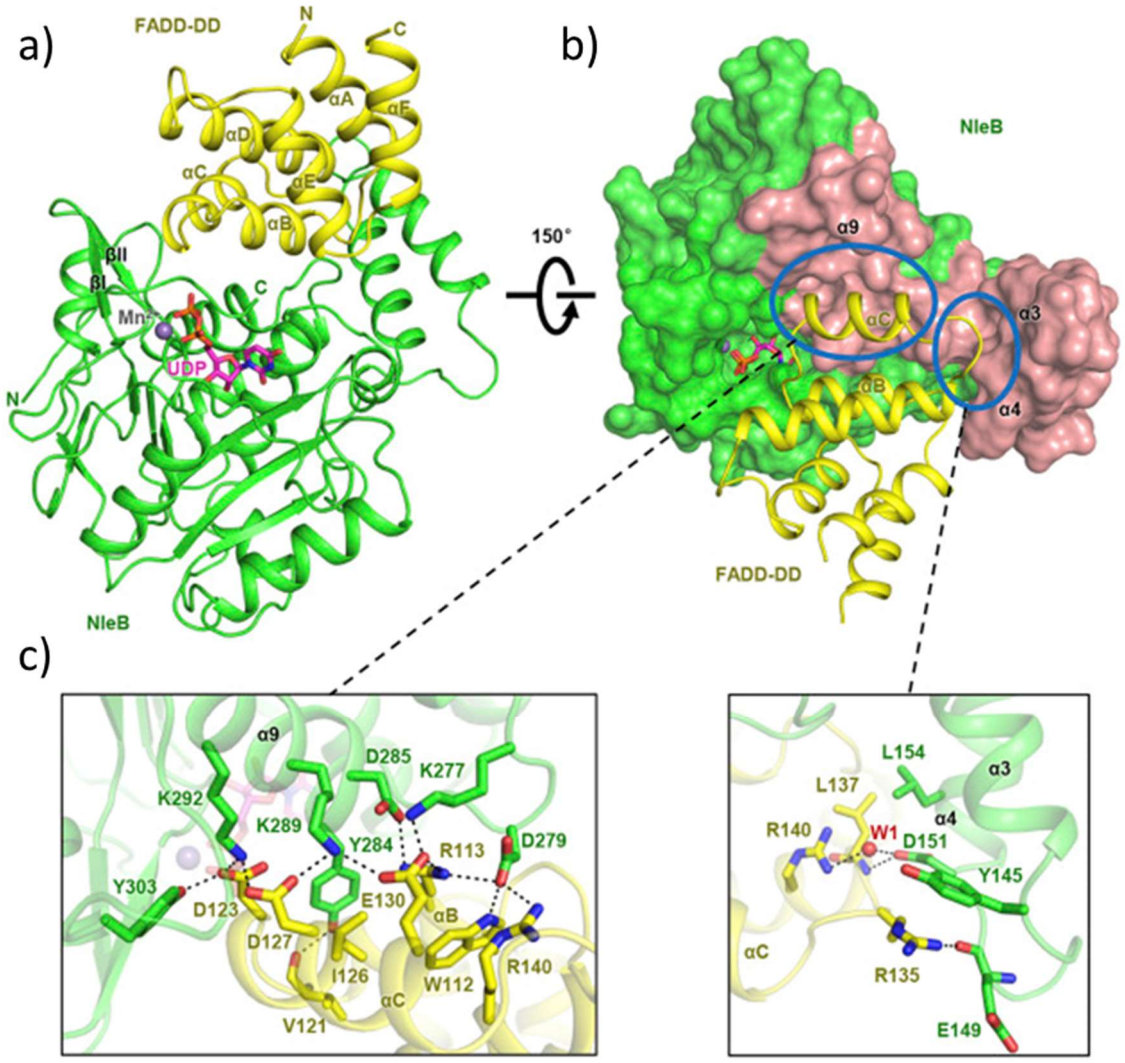



*Figure 1.13 - Complete structure of NleB1 in complex with FADD and UDP-GlcNAc (a) and the interaction (pink surface) between NleB1 and FADD (b). Detailed interactions (c) show a complex series of hydrogen bonding and salt bridges between the HLH domain and FADD. Taken from (35).*

All the effectors have a C-shape which can accommodate the acceptor substrate, formed by the HLH and catalytic domains (36, 37). Between these domains, there is the proposed acceptor recognition surface. To date a detailed understanding of this binding

surface for all the NleB and SseK effectors has not been described. This owes to the limited information around of the enzyme-acceptor complex with, as mentioned, only one example crystal structure having been published (35). Despite this though, it is possible to still draw important conclusion around how these effectors may be able to recognise their substrate. In NleB1 α9 forms the main binding surface, forming close contacts with αB and αC of FADD (Figure 1.13) (35). In this contact there are two sets of hydrogen binding interactions. In the α9 helix a triad of positively charged residues including NleB1$^{K292}$, NleB1$^{K289}$ and NleB1$^{K277}$ form a complex hydrogen bonding network with a complementary triad of negatively charged residues, FADD$^{D123/D127/E130}$. These interactions are further supported by a complementary hydrogen bonding interaction between NleB1$^{Y303}$ and FADD$^{D123}$. Adjacent to these interactions a pair of negatively charged residues NleB1$^{D279/D285}$ form another hydrogen bonding network with FADD$^{R113/R140/W112}$. This region is further supported by an interaction between NleB1$^{K277}$ and FADD$^{E130}$. In the HLH domain, α3 and α4 helices, there are three specific contacts with FADD (Figure 1.13) including a hydrophobic interaction between NleB1$^{L154}$ and FADD$^{L137}$, a water-mediated hydrogen bond between NleB1$^{D151}$ and FADD$^{R140}$ with a π-cation interaction between NleB1$^{Y145}$ and FADD$^{R137}$. Completing interactions in this region, a single tyrosine reside, NleB1$^{Y284}$, is inserted between αB and αC of FADD into a hydrophobic pocket formed by NleB1$^{V121/I126}$. It is proposed that the makeup of this region is important for the selectivity between acceptor substrates (35). Though crystallographic structures of other enzyme-acceptor complexes have not been published to date, there has been a mutation study to explore the features of this binding surface. This work identified that the tyrosine residue, NleB1$^{Y284}$, that is inserted between αB and αC of FADD, is important for FADD recognition (44).

The catalytic domain of all NleB and SseK effectors contain conserved features that are important for catalysis. Being GT-A GTs all the NleB and SseK effectors contain a DXD motif that co-ordinates a single Mn(II) ion with coordination facilitated by a single aspartic acid residue (35-37). In addition to this there is a conserved His-Glu-Asn (HEN) motif that is important for catalysis. Point mutations of any of these amino acids in NleB1, NleB2, SseK1 and SseK2 has reported to abolish their activities towards FADD, TRADD and GAPDH. This triad of residues therefore play a crucial role in the catalytic mechanism of action of these enzymes. In the crystal structure of NleB1 with FADD, the glutamic acid of the HEN motif

makes close contacts with the FADD glycosylation site, $FADD^{R117}$, forming a favourable salt bridge with the guanidinium group (35). This observation is consistent with docking studies of a FADD peptide with SseK2 where the glutamic acid in that effector forms a similar interaction with $FADD^{R117}$. This interaction appears to support catalysis by providing a stabilising platform for the arginine residue within the catalytic machinery. The asparagine residue of the HEN motif is positioned close to the anomeric carbon of the GlcNAc ring and can support recognition and catalysis, while the histidine is expected to behave as a catalytic base.

In addition to interactions that support catalysis there are also key features in the catalytic domain that facilitate substrate binding and recognition of the donor ligand. As mentioned, there are examples of crystal structures with UDP bound to the effector. To date the only structure to report full electron density of the GlcNAc ring and complete UDP-GlcNAc structure is SseK2 (36). While the crystal structure of SseK3 also reports full electron density for the GlcNAc ring, it is presented in its hydrolysed form (37). The complex of UDP-GlcNAc bound to SseK2 provides an important insight into how these effectors can recognise their substrate. UDP-GlcNAc contains three main substructures that form interactions with the catalytic domain: a uridine nucleobase, a pyrophosphate and a GlcNAc ring. The donor binding site of the NleB and SseK effectors can be also broken down into three corresponding regions which contain interactions that complement each substructure of UDP-GlcNAc.

The uridine nucleobase is sandwiched between a phenylalanine and tryptophan residues in a π-π stacking interactions in SseK2, SseK3 and NleB1 (Figure 1.14) (35-37). Though the residues are conserved in SseK1, the π-π stacking interaction from phenylalanine is replaced by a lid-domain tryptophan (Trp331) (36). In addition to this hydrophobic interaction the uridine moiety is further supported by a water-mediated hydrogen nitrogen backbone interaction. This residue is a conserved arginine residues in the SseK effectors and NleB this is an alanine residue (36). Completing the uridine hydrogen bonding network there are a pair of hydrogen bonding interactions with a conserved phenylalanine and glutamic acid.

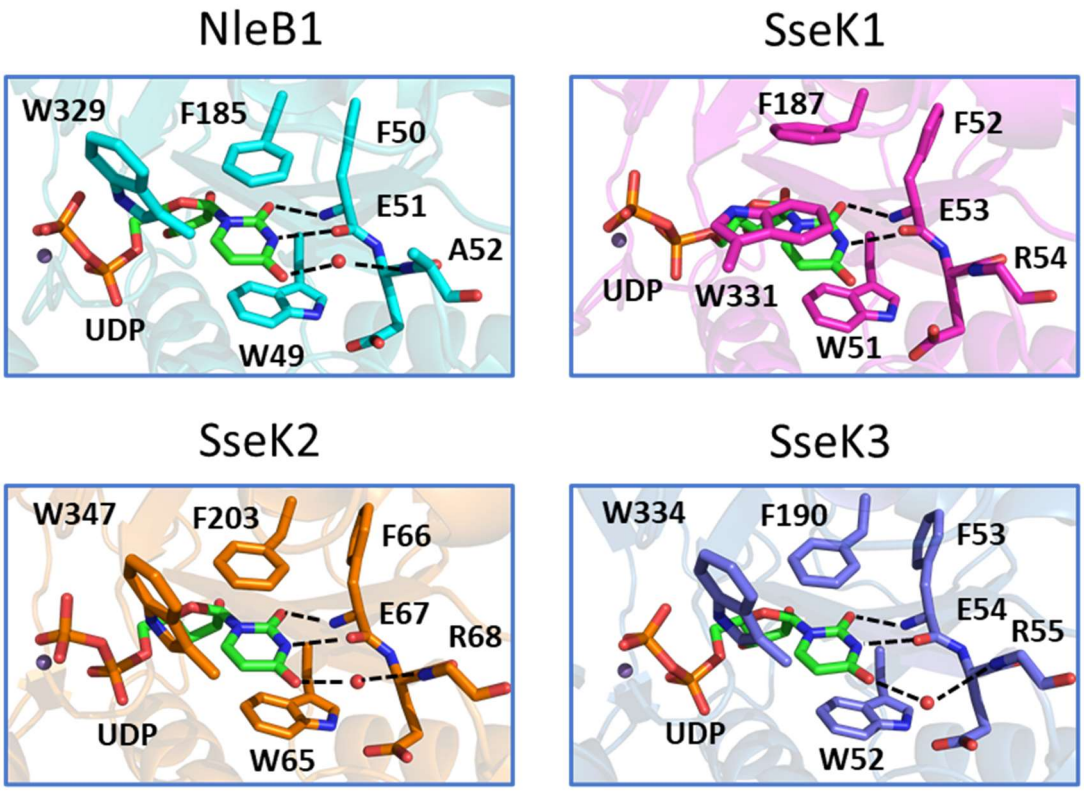



*Figure 1.14 - NleB and SseK effector interactions with the uracil moiety of UDP. Uracil forms a highly conserved π-π stacking interaction with phenylalanine and tryptophan is NleB1 (PDB entry 6ACI) (a), SseK2 (PDB entry 5H62) (b) and SseK3 (PDB entry 6EHY) (c). However, in the case of SseK1 (PDB entry 5H30) (b) the π-π stacking interaction comes from SseK1$^{W331}$ of the lid domain in place of phenylalanine. In addition to the aromatic interactions, the uracil moiety forms complementary hydrogen bonding interactions (black dashed lines) with the backbone amines of a conserved phenylalanine-glutamic acid pair.*

The pyrophosphate of UDP-GlcNAc is coordinated by a Mn(II) ion in the catalytic domain. The Mn(II) ion itself is coordinated in an octahedral geometry with contribution interaction from the DXD motif using a single aspartic residue (35-37). The crystal structure of UDP-GlcNAc-bound SseK2 shows the role this ion plays in stabilising the bound sugar nucleotide (Figure 1.15) (36). In this structure SseK2$^{D241}$ (DXD) co-ordinates the manganese ion alongside SseK2$^{N338}$ and SseK2$^{S340}$. There is additionally a single water molecule that interacts with the Mn(II) ion, SseK2$^{D239}$ (DXD) and hydrogen bonds to the carbonyl oxygen of the GlcNAc amide. The coordination of the Mn(II) ion in the catalytic domain and the interactions that it forms with the sugar nucleotide substrate are particularly important to catalysis with the omission of the metal ion leaving non-complementary repulsive interactions between the DXD motif and the pyrophosphate. The importance of Mn(II) coordination in the active site is further echoed by the highly conserved nature of these residues on other NleB and SseK GTs.

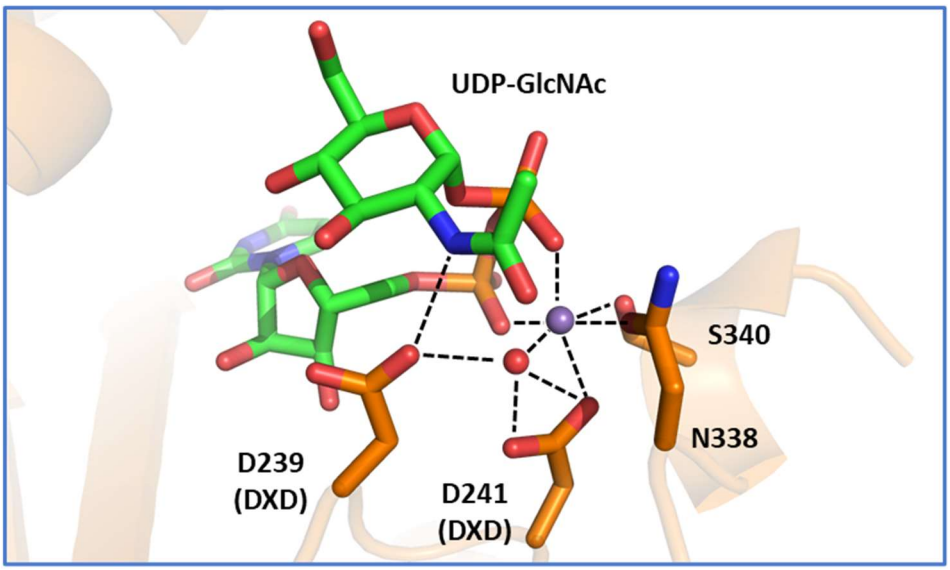



*Figure 1.15 - Coordination of Mn(II) by SseK2 in complex with UDP-GlcNAc. The Mn(II) ion is coordinated by hydrogen bonds and salt bridges (black dashed lines) from SseK2$^{S340}$ and SseK2$^{N338}$ of the lid domain and SseK2$^{D241}$ of the DXD motif.*

From the structure of SseK2 in complex with UDP-GlcNAc there is a clear insight into how the GlcNAc ring can be recognised within the catalytic domain (Figure 1.16). In SseK2, four residues form complementary interactions with the GlcNAc ring. In addition to the indirect, water-mediated, interaction mention above SseK2$^{D239}$ (DXD) interacts with the nitrogen of the amide as well as hydrogen of hydroxyl at position 3 (36). Alongside the interaction from the DXD motif SseK2$^{D204}$, SseK2$^{R207}$ and SseK2$^{R348}$ of the lid domain provide additional hydrogen bonding interactions with the hydroxyl groups at positions 3, 4 and 6 as well as the endocyclic oxygen. There is evidence to suggest that the spatial arrangement of these hydrogen bonding interactions contributes to the donor substrate specificity of the NleB and SseK effectors. ITC measurements of the binding of UDP Glucose (UDP-Glc), shows that omission of the N-acetyl group leads to an approximately 17-fold increase in $K_d$ compared to UDP-GlcNAc (36). Furthermore, altered stereochemistry of the ring in addition to the lack of the N-acetyl group, such as UDP Galactose (UDP-Gal), leads to an approximately 70-fold increase in $K_d$ (36).

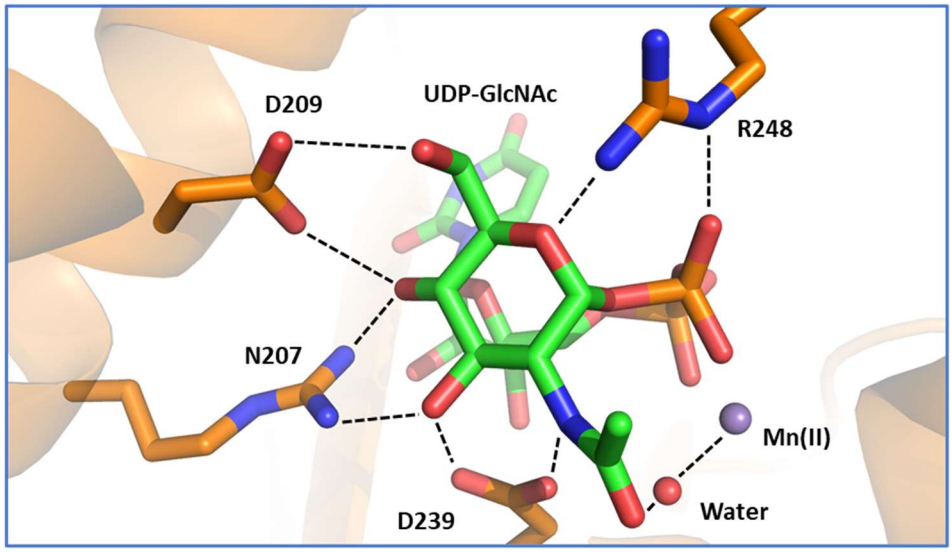



*Figure 1.16 - Recognition of the GlcNAc moiety of UDP-GlcNAc by SseK2 (PDB entry 5H63). Recognition is facilitated through hydrogen bonding (dashed lines) and includes a water-mediated coordination of the NAc carbonyl group by Mn(II) (purple sphere).*

The Lid domain of the NleB and SseK effectors is a highly dynamic loop at the C-terminus. In the presence of UDP or UDP-GlcNAc the lid domain becomes highly ordered and favours the closed conformation (Figure 1.17) (36). The interactions from this domain play a crucial role in donor recognition with truncation of this domain leading to a significant reduction in ligand affinity. The C-terminal domain shared conserved residues between the NleB and SseK effectors however there are some features unique to different effectors (Figure 1.17). Important to hydrogen bonding interactions with UDP-GlcNAc there is a conserved tryptophan-arginine (WR) motif in the SseK effectors. In SseK3 it has been demonstrated that this pair of residues are important for catalytic mechanism of action, with alanine point mutation of the tryptophan residue leading to a loss of activity (37).

This motif, though conserved in SseK2, is slightly altered in SseK1 in which has an alanine between the two residues (36). In addition, the lid domain of SseK1 is unique compared to the other effectors in that it is extended with an additional His-Val-Gln (HVQ) sequence (36). The addition of the alanine residue between the tryptophan and arginine of the WR motif and the three-residue extension allows the SseK1 lid domain to be more dynamic (36). Given that the lid domain SseK1 uniquely contributes to π-π stacking interactions with the uracil aromatic ring of bound UDP-GlcNAc it is possible that potential for increased dynamics is what allows that interaction to occur. Though the WR motif can be seen to be important in the SseK effectors, the NleB effectors only retain a terminal tryptophan reside and lack the C-terminal arginine (35). This, in combination with the alanine point mutation study in SseK3 (37) positions the tryptophan residue as being much more crucial than the arginine in the WR motif.

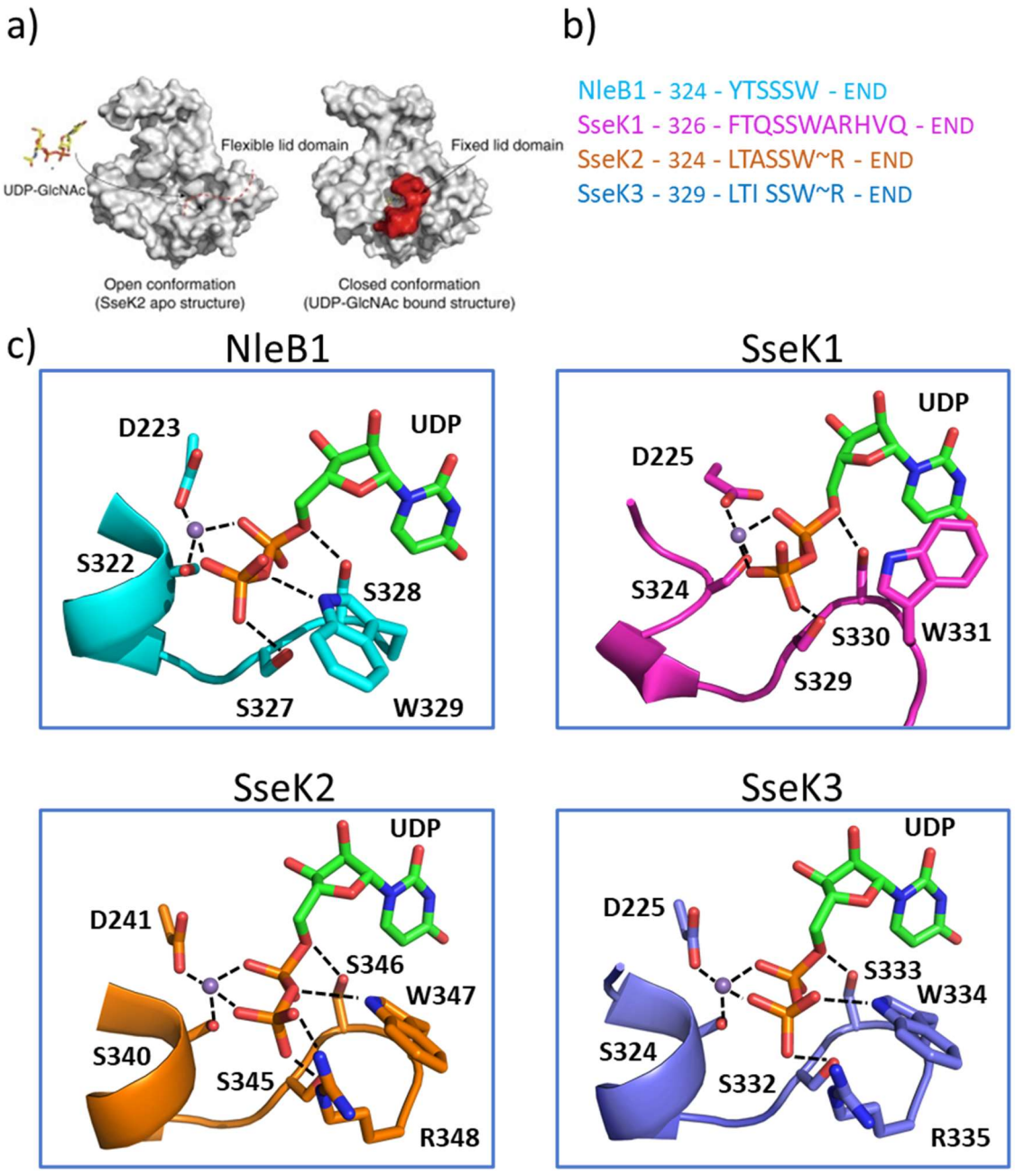



*Figure 1.17 - The lid domain of the NleB and SseK effectors. The open and closed conformations of SseK2 (a) (Adapted from (36)) and (b) the lid domain sequence s of NleB1, SseK1, SseK2 and SseK3. The lid domain interactions of UDP-bound NleB1 (PDB entry 6ACI), SseK1 (PDB entry 5H60), SseK2 (PDB entry 5H62) and SseK3 (PDB entry 6EHY) (c). In NleB1, SseK2 and SseK3 a conserved SSW sequence contributes to coordination of the pyrophosphate group. In both SseK2 and SseK3, an addition arginine residue that is not present in NleB1 also contributes to coordination. In SseK1 as the lid domain tryptophan contributions to π-π stacking interactions with uracil, it does not co-ordinate the pyrophosphate however the SS residues do in a similar manner to the other enzymes.*

# 1.2 Nuclear Magnetic Resonance

Nuclei possess an intrinsic quality called spin angular momentum or just ‘spin'. Spin can take on half-integer values denoted *I* for a given nucleus. Nuclear magnetic resonance (NMR) is a phenomenon in which nuclei that have non-zero spin can interact with an applied magnetic field. Since its discovery, NMR has been developed into a remarkably versatile tool finding use in different fields ranging from the study of fundamental quantum properties of matter to understanding the mechanical details of biological systems. For this thesis, the study of biological systems is of most interest, particularly the study of protein-protein and protein-ligand interactions.

## 1.2.1 Nuclei Within a Magnetic Field

The consequence of the spin of a nucleus is that it can allow for the existence multiple nuclear spin states in the presence of an applied magnetic field. Using the example of $^{1}H$, with a spin of ½, there are two states the nucleus can adopt. This can be generalised for any value for *I*, allowing for there to be 2*I*+1 spin states which are bound between +*I* and –*I* in integer intervals. In the absence of an applied magnetic field the nuclear spin states represent degenerate energy levels. However, when a magnetic field is applied, the spin states adopt energy levels with the energy difference (ΔE) (Equation 1).

(1) $$\Delta E = \gamma \hbar B_0$$

Where γ is the gyromagnetic ratio of the nucleus in rad $s^{-1}$ $T^{-1}$, $\hbar$ is Planck’s constant over 2π and $B_0$ is the applied magnetic field strength in tesla (T). The resultant separation into different energy levels in the presence of a magnetic field produces and effect called Zeeman splitting (Figure 1.18). Following the 2*I*+1 rule, a spin 1/2 nuclei will have two energy levels. For $^{1}H$ nuclei which has a positive gyromagnetic ratio, the +1/2 spin state is lower energy, denoted α, and -1/2 spin state is higher energy, denoted β. However, there are many more states for nuclei with higher values for spin described as possessing higher

multiplicity. As can be seen in the equation, the difference in energy levels between two populations is directly proportional to the gyromagnetic ratio. The gyromagnetic ratio is another intrinsic property of a nucleus defined as the ratio between the magnetic moment of a nucleus and its angular momentum. Nuclei that possess higher gyromagnetic ratio such as $^{1}H$ are more sensitive and are easier to measure by NMR than nuclei that have lower gyromagnetic ratios such as $^{15}N$.

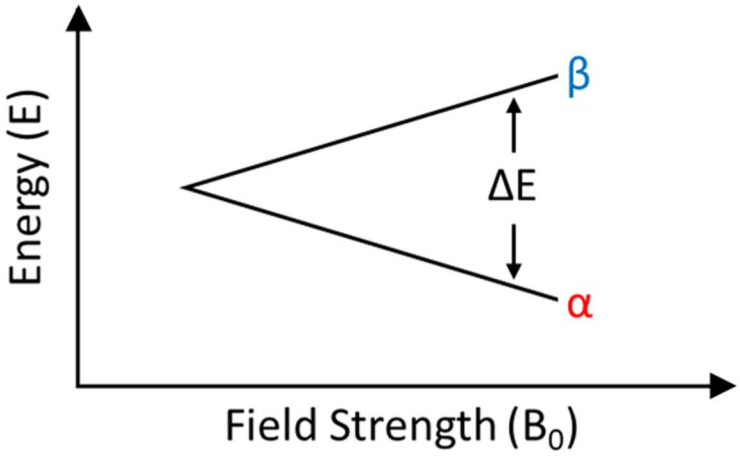



*Figure 1.18 - The nuclear Zeeman effect. Showing that the energy difference between the α and β state of a spin increases with field strength.*

When aligned, nuclei precess around the applied magnetic field. The frequency of this precession is called the Larmor frequency ($\omega_0$ when expressed in rad $s^{-1}$ or $\nu_0$ when expressed in Hz). This value is subtly influenced by many factors for a given spin within a particular molecule, however, it is most dependent on the strength of the magnetic field (Equation 2). To express the Larmor frequency ratio in Hz, $\omega_0$ can be multiplied by $2\pi$. Transitions between energy levels for a particular spin can be induced by a radiofrequency signal that matches the Larmor frequency (45).

(2) $$\nu_0 = 2\pi\omega_0 = \frac{-\gamma B_0}{2\pi}$$

## 1.2.2 Chemical Shift

The nucleus is not the only part of an atom that possesses spin. Electrons also possess spin and can interact with magnetic field applied by the spectrometer. Because of how they are arranged into bonding orbitals and the effects of different functional groups, electrons are not uniformly distributed in molecules. This means that each nucleus within a molecule can experience a magnetic field different to $B_0$ and therefore have their own frequency of Larmor precession. The effect of nuclei possessing their own Larmor frequencies that are different from that of the 'expected' frequency based solely on gyromagnetic ratio and applied magnetic field is called chemical shift (δ). Chemical shift can be incorporated in to the Larmor frequency expression (Equation 3). The units for chemical shift are parts per million (ppm) and presented as the offset from a reference frequency ($\nu_{ref}$) which is arbitrarily set to be 0 ppm, normalised against $\nu_{ref}$ (Equation 4) (45, 46).

(3) $$\nu_0 = \frac{-\gamma B_0[1+(\delta \times 10^{-6})}{2\pi}$$

(4) $$\delta = \left(\frac{\nu_0 - \nu_{ref}}{\nu_{ref}}\right) 10^6$$

The magnetic field generated by an electron is very weak and opposes the field applied by the spectrometer. This means that a nucleus surrounded by electrons will experience a magnetic field slightly weaker than $B_0$. The term used to describe this effect is shielding, as the electrons can be thought of as acting as a 'shield' from the applied magnetic field. The more electrons around a nucleus or, the higher the electron density, the more shielding is experience. Some functional groups, such as aromatic and electron withdrawing groups, can reduce the electron density around neighbouring nuclei. The term used to describe this effect is deshielding. A more deshielded nucleus will experience a stronger magnetic field compared to one that is shielded. An important note is that shielding and deshielding are relative terms used when comparing different nuclei. Electrons themselves can only generate a magnetic field that opposes the applied magnetic field so can only truly shield a nucleus.

Nearby nuclei can also contribute to the shielding and deshielding of a nucleus. When a nucleus is in the α state its magnetic field aligns with $B_0$, producing a deshielding effect. If a nucleus is in the β state, the opposite is true and there will be a shielding effect. When neighbouring nuclei influence the chemical shift of one-another they are described as coupled (45). For organic molecules in solution coupling is typically carried through bonds and observed between hydrogen nuclei, which is an example of homonuclear scalar coupling. This effect can also occur between any neighbouring nuclei that have non-zero spin through heteronuclear scalar coupling. As for most nuclei the populations in the α and β state are approximately equal, the signal from a nucleus coupled to a single nucleus will be split equally into two separate signals in a splitting pattern called a doublet. The more nuclei that are coupled to a particular nucleus the more complicated its splitting pattern will be. The separation between signals in is called a coupling constant, denoted J, measured in Hz. Coupling constants can be used to identify nuclei that are coupled to one another as their splitting pattern will share the same value for J.

In organic molecules, functional groups can influence chemical shift in predictable ways. Conjugation between bonds and the difference in electronegativity between neighbouring atoms are two such examples. There are also indirect influences such as the composition of the solvent that a molecule is in that will affect chemical shift. For example, a polar molecule within water will experience hydrogen bonding which can change the distribution of electrons around nuclei. If a molecule contains pH sensitive functional groups such as carboxylic acids and amines the pH of a sample will affect chemical shift by changing the protonation states of these functional groups. These effects alongside the shielding from electrons and coupling are used to inform the assignment of NMR spectra.

## 1.2.3 Bulk Magnetisation

Within a magnetic field all of spins align to $B_0$ and bear their own magnetic moment. These moments are vectors, and they can be broken down into three components $\mu_z$ and the transverse components $\mu_y$ and $\mu_x$. By convention $\mu_z$ is aligned with $B_0$. The sign of the spin state is contained within the $\mu_z$ component and, with respect to $B_0$, is either positive, in the case of α spins or, negative in the case of β spins. When aligned to $B_0$ the $\mu_y$ and $\mu_x$ components cancel as the magnetic moments all point in random directions. This leaves just the positive and negative $\mu_z$ values. These do not cancel as the populations of α and β spins follow a Boltzmann distribution (Equation 5) (47).

(5) $$\frac{N_\alpha}{N_\beta} = e^{\frac{\Delta E}{RT}}$$

Where $N_\alpha$ and $N_\beta$ are the populations of α and β spins, ΔE is the energy difference between the two states, R is the gas constant and T is temperature. The result is that there are more positive α magnetic moments than negative magnetic moments leaving a net positive vector in the Z axis from the sum of all magnetic moments. This vector is called the bulk magnetisation of a system, $M_z$ (Figure 1.19) (45).

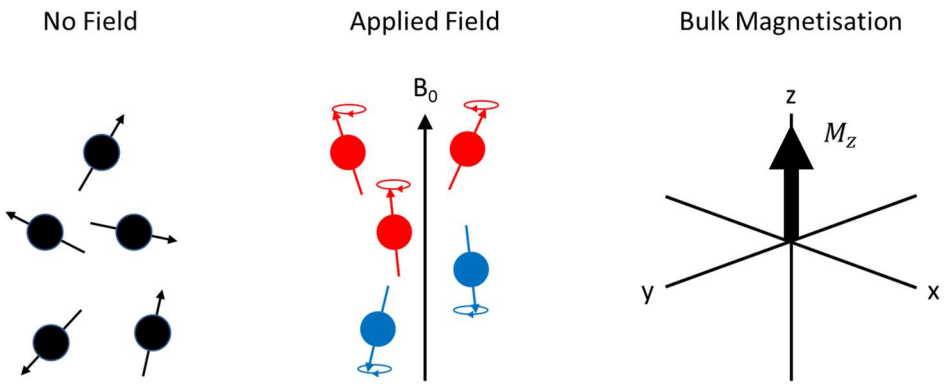



*Figure 1.19 - The alignment of spin 1/2 nuclei. On application of a magnetic field ($B_0$) spins will align with the direction of the magnetic field; at equilibrium there will be a slight preference for the α spin state (red) over the β spin state (blue). When spin populations are at equilibrium, there will be a net magnetic moment in the Z axis called the bulk magnetisation ($M_z$).*

Bulk magnetisation will remain at equilibria within a fixed magnetic field until acted upon. In NMR, the bulk magnetic vector can be manipulated through the application of radiofrequency pulses. In the case of a simple pulse-acquisition experiment, the radiofrequency pulse is applied along the x axis. This has the effect of there being a temporary, strong magnetic field ($B_1$) which makes all resonant spins point towards the -y axis. At this time, the spins are described as coherent once their transverse components align. Once the spins are coherent, the bulk magnetic vector will then precess around the x axis until the pulse is no longer applied. The angle that the vector moves away from the z axis (θ°) can be controlled by the duration of the radiofrequency pulse ($t_p$) is applied and its power. The angle is dependent on the power and duration of the radiofrequency pulse and is dependent on gyromagnetic ratio (Equation 6) (47).

(6) $$\theta° = 360\gamma B_1 t_p$$

From this equation the, longer a pulse of fixed power (fixed $B_1$) is applied, the longer the bulk magnetic vector can precess about the x axis and therefore the larger the angle that it moves from the z axis. The higher the power of a pulse the stronger the magnetic field it creates on the x axis. For a given pulse length, higher power pulses will allow the bulk magnetic vector to move further away from the z axis as the Larmor frequency about the z axis will increase proportional to the value of its gyromagnetic ratio.

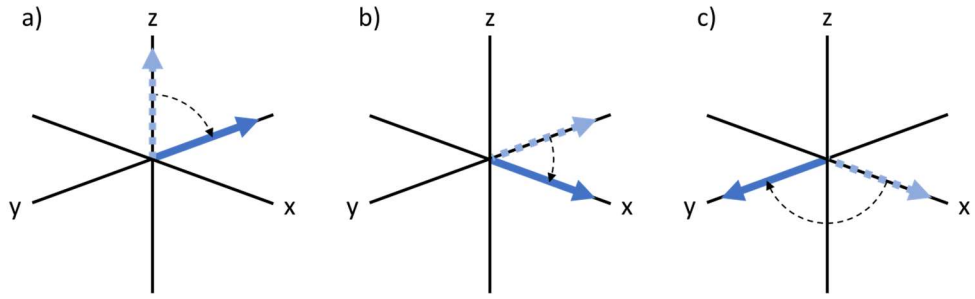



*Figure 1.20 - Movement of the bulk magnetic vector after application of a 90° pulse on the x axis a). Once in the transverse plane, for nuclei with a positive gyromagnetic ratio, the bulk magnetic vector moves from the -y axis to the x axis b) and then to from the x axis to the y axis c) and so on.*

When the radiofrequency pulse stops, the bulk magnetic vector then returns to precession about $B_0$ in the x axis (Figure 1.20) but now will have $M_y$ and $M_x$ components. While undergoing precession in the transverse plane the $M_y$ and $M_x$ components produce a signal that can be detected during acquisition.

The bulk magnetic vector does not remain in the xy plane and over time returns to equilibrium along the z axis, losing its $M_y$ and $M_x$ components. While returning to the z axis the intensity of the detected signal reduces to zero as it moves to equilibria. The intensity detected can also be lost through a process called dephasing. Over time while in the xy plane, the individual vectors that make up the bulk magnetic vector separate and lose coherency. This leads to a loss in intensity of the bulk magnetic vector in the xy plane as the separating vectors start cancelling. The loss of signal intensity due to dephasing and the bulk magnetic returning to its equilibrium position is called relaxation.

## 1.2.4 Relaxation

When a radiofrequency pulse is applied that matches the Larmor frequency of a nucleus it will absorb energy and become excited from the ground state. This spin will not remain in the excited state indefinitely and will release the energy after some time in a process called relaxation. Understanding the relaxation properties of a sample can be useful to guide experimental design and the selection of the specific NMR experiment performed. As mentioned, after a pulse is applied, bulk magnetisation does not remain at a constant in the xy plane and undergoes relaxation. This relaxation process takes two forms, longitudinal ($T_1$) and transverse ($T_2$), with each having a distinct mechanism (47). $T_1$ relaxation is the loss of energy leading to the return of the bulk magnetic vector to the z axis. $T_2$ relaxation involves is a loss of coherency between the spins that form the bulk magnetic vector. Both processes come about due to subtle fluctuations in magnetic field around every nucleus within a magnetic field. In the case of spin ½ nuclei the dominant factors that affect local magnetisation come from the interactions that nearby nuclei have on one another (Figure 1.21) and the uneven distribution of electrons around a molecule (45). The magnetic interaction between two uncoupled nuclei is called a dipole-dipole or, dipolar interaction.

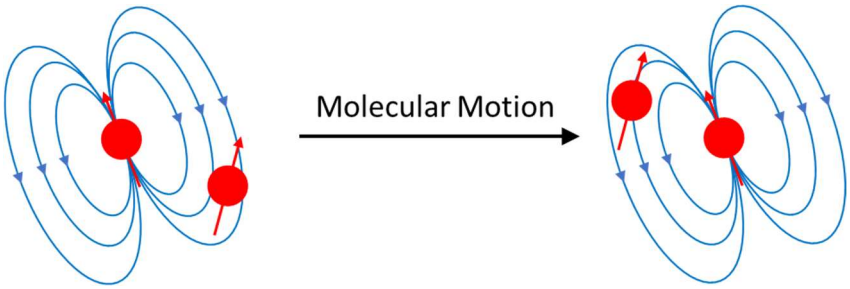



*Figure 1.21 - The dipole-dipole interaction between two spins. Their mutual interaction will depend on their position in space. For molecules that can tumble randomly at different time points the dipole-dipole interaction will be different at different time points. In fast motion this effect will average out however in slower tumbling molecules this will induce relaxation.*

The dipolar interaction is distance dependent, with the strength having an inverse-cube relationship with the distance between the two nuclei. Because of this, the typical range of these interaction is only a few angstroms (45).The dipolar interaction is also dependent on the gyromagnetic ratio of the atoms involved and its orientation compared to the applied magnetic field (45).

The second dominant factor, the uneven distribution of electrons within a molecule, has been discussed previously in chemical shift. Important to relaxation, each nucleus will have a different level of shielding from the magnetic field because of the opposing magnetic field produced by electrons. This means that the magnetic field vector will not necessarily align exactly with the applied magnetic field. This variation in local field around a nucleus can induce relaxation and is called chemical shift anisotropy (CSA) (45). This effect is less dominant for molecules in solution but plays a significant role in solids. The reason for this is because in solution a molecule can tumble at a high enough speed to average out the impact that these variations in magnetic field have on each nucleus.

In both the dipolar interaction and in CSA molecular motion plays a key role in determining how much each mechanism contributes to relaxation. Specifically, for molecules in solution that are free to move this is tumbling. This motion is described by correlation time ($\tau_c$), which is the average time for a molecule to rotate through one radian. There are multiple factors that influence the correlation time of a molecule ranging from physical factors such as the temperature and viscosity of the solution to the relative size and shape of a molecule. In general, the correlation time can be estimated from molecular weight. That is to say, the heavier a molecule is, the larger it will tend to be and therefore the slower it will tumble in solution.

As mentioned, $T_1$ relaxation is the return of bulk magnetisation to equilibrium at the z axis. This process is the relaxation of individual spins through loss of the energy they gained from the radiofrequency pulse. As a molecule rotates the magnetic field around each nucleus will fluctuate at a frequency dependent on its correlation time. If this fluctuation in magnetic field matches the Larmor frequency of a nucleus it can induce relaxation from the excited to the ground state. However, this motion is random and therefore for molecules in solution there will be a distribution in the sample of different correlation times. For small molecules with a short correlation time, this distribution will be such that it will be likely a given molecule will rotate at a frequency high enough to induce $T_1$ relaxation. For larger molecules with longer correlation times this distribution will extend lower rotational frequencies meaning there is a lower chance that there will be $T_1$ relaxation.

During $T_2$ relaxation, the spins lose coherency over time. Up until this point magnetic vectors have been described as coherent meaning that all the spins align to produce distinct vectors. Through $T_2$ relaxation the spins will begin to spread out and if given enough time would point in random directions leading there to be a no net magnetisation in the XY plane. In a real system this often does not typically happen completely as the signal may relax longitudinally, or the signal would be acquired before completely relaxing transversely. Due to the random nature loss of coherency has, $T_2$ contributes to line broadening and the Gaussian shape of signals in an NMR spectrum. The linewidth at half height ($\Delta\nu_{1/2}$) (Equation 7) can be used to find apparent $T_2$ (48).

(7) $$\Delta\nu_{1/2} = \frac{1}{\pi T_2}$$

From this equation we can see that the linewidth at half height or, broadness, of a peak has an inverse relation with $T_2$. The reason for the dephasing during $T_2$ relaxation is that the spins in a magnetic field will not all experience the same magnetic field. Though after a 90° pulse the spins are coherent, these differences in magnetic field will mean as they precess around $B_0$ they will spread out and lose coherency. As coherency is lost, the magnitude of the vector will decrease as the transverse components of faster spins cancel those of slower spins.

## 1.2.5 The Nuclear Overhauser Effect (NOE)

The NOE is a process in which there is a transfer of population differences (or polarisation) through space from a polarised nucleus (i.e. with perturbed population differences) to neighbouring nuclei which are at equilibrium. If a low-power selective pulse is applied to a sample which matches the Larmor precession of one spin in a system and a pulse-acquisition spectrum is recorded, the intensity of the selected peak can be reduced to zero (Figure 1.22). This pulse is described as a saturating pulse and its effect is called saturation. When a saturating pulse is applied there is a change in populations between the α and β states of targeted spins and they are perturbed from equilibrium as defined by the normal Boltzmann distribution. In the example of a saturating pulse which reduces signal intensity zero, α spins are promoted from their ground state to β spins such that the population difference will be zero. This would mean that the net magnetic moment from these spins will be zero, and therefore the observed signal will be zero.

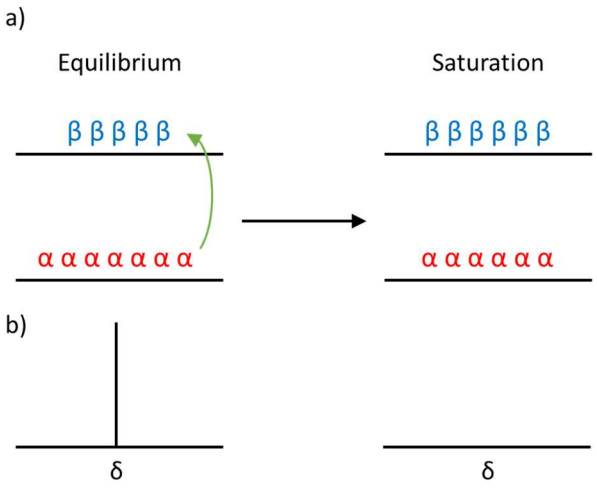



*Figure 1.22 - Effect of saturation on a spin transition and the observed spectra. In a), on application of a saturating pulse spins are promoted to the β state such that spin populations are equalised. In the resulting spectrum, as there is no difference in spin populations there will be no net magnetic moment, so that there will be no observable signal.*

Now, if we consider three homonuclear spins in space I, S and U. I and S are very close in space and U is far from I and S. If a saturating pulse is applied on spin I then there will be an effect on the intensity of signals from spin S, but not spin U in a following pulse-

acquisition experiment. The change in signal intensity of spin S because of the saturation of spin I is the NOE. The reason U is unaffected is that the NOE is very distance dependent, inversely proportional to $r^6$, where r is the distance between two nuclei. The sign of the NOE is dependent on a variety of factors including the physical parameters of the sample, the strength of the applied magnetic field and the gyromagnetic ratio of the observed nucleus. When there is a reduction of peak intensity there is a negative NOE and when there is an increase in peak intensity there is a positive NOE. By convention, an NOE is typically given as the relative change in intensity between the intensity of signal from a spin at equilibrium compared to when influenced by an NOE. The NOE is expressed as a % ($\eta_S$) (Equation 8) (46). Where $I_0$ *is* the intensity of the peak with no saturating pulse applied and *I* is the intensity of the peak when a saturating pulse is applied.

(8) $$\eta_S = \frac{I - I_0}{I_0} \cdot 100$$

The NOE arises due to through-space dipolar interactions that nearby spins have with each other. For two spins, I and S, there are four energy states (Figure 1.23). Between these energy states there are six possible transitions including four single-quantum transitions ($W_1$), one double-quantum transition ($W_2$) and one zero quantum transition ($W_0$). The four single quantum transitions are those related to a single spin relaxing. The double-quantum and zero-quantum transitions can describe relaxation where the relaxation of one spin influences the spin state of another. Depending on the motion of a molecule, one double quantum transition can be dominant over the other. Large molecules which slowly tumble with a long $\tau_c$ relaxation will be $W_0$ dominant. If spin I and S are in a large molecule, during the $W_0$ transition $\beta_I\alpha_S$ will flip to $\alpha_I\beta_S$. This will mean that for spin S some α spins will become β spins. When this happens the difference in population for the α and β spins becomes smaller. Therefore, during a pulse acquisition experiment the signal for spin S will have a lower intensity than when spin I is not saturated which is a negative NOE. For small molecules that tumble very fast with a short $\tau_c$ the dominant relaxation pathway will be $W_2$. If I and S are in a small molecule, during this transition $\alpha_I\alpha_S$ will flip to $\beta_I\beta_S$. This will mean that there is an increase in the number of β spins for spin S and the population difference between α and β spins will increase. In this case if a pulse acquisition

experiment is performed the signal for spin S will have a higher intensity than when spin I is not saturated which is a positive NOE.

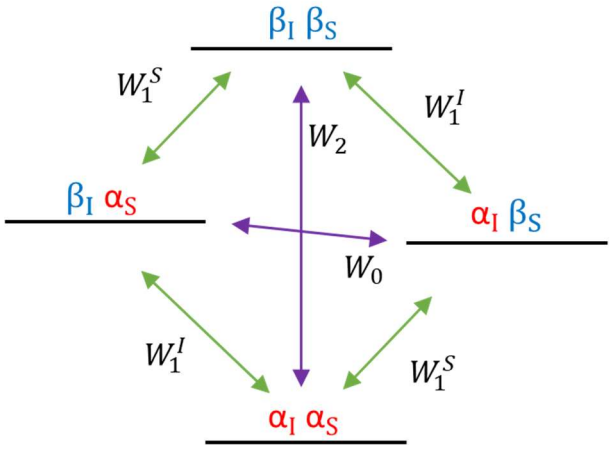



*Figure 1.23 - The available transitions for an IS spin system. The single quantum transitions (green arrows) a single spin flips of either the I or S spin, these transitions are associated with relaxation. The double quantum transitions (purple arrows) represent double spin flips which occur via the NOE.*

The two cases that have been presented are the extremes of motion, a plot of the dependence of NOE on $\tau_c$ (Figure 1.24) shows that there is a crossover point where the NOE is 0 as motion changes from fast to slow. This point is called the extreme narrowing limit and is where neither double quantum transition is dominant.

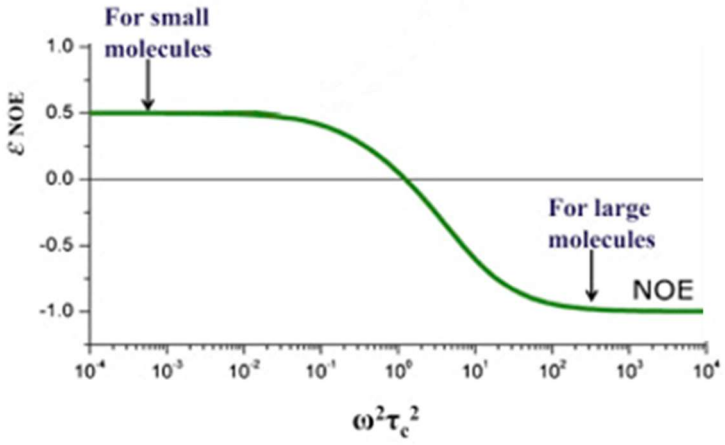



*Figure 1.24 - The dependence of NOE on correlation time showing that the NOE is positive for small molecules with a short correlation time and negative for large molecules with a long correlation time. Adapted from (49).*

### 1.2.5.1 The Transient NOE

One method of observing the NOE is the transient NOE (T-NOE) experiment (Figure 1.25) (45). In this experiment, two spectra are recorded. In the first spectrum a selective 180° pulse is applied to invert a particular population of spins. Then, over a set mixing time ($\tau_m$) the spins are allowed to relax through both self- and cross-relaxation.

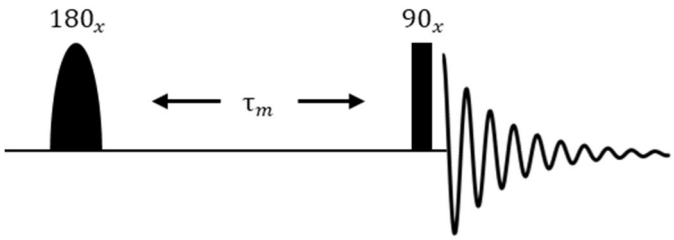



*Figure 1.25 - The pulse sequence for the Transient NOE experiment. The sequence begins with a selective 180° pulse on the x axis followed by a mixing time ($\tau_m$). Finally, a non-selective 90° pulse is applied on the x axis prior to acquisition.*

After $\tau_m$ has passed, a 90° pulse is applied to move magnetisation into the transverse plane for detection. During $\tau_m$ the cross-relaxation of the inverted spins with nearby spins will perturb the population differences of the latter. The second spectrum acts as a reference in which only a 90° pulse is applied prior to acquisition. The reference spectra can then be subtracted from the first spectra to produce a difference spectrum. In the difference spectrum only, peaks will be seen coming from cross relaxation. Dependent on the dominant transition pathway during cross relaxation (i.e. dependent on the molecular tumbling in solution) the sign of signals in the spectra will be either positive or negative. Due to the distance-dependent nature of the NOE the specific distance between two spins can therefore be found, given proper intensity/distance calibration is carried out. This feature is useful particularly when studying the 3D structure and conformation of molecules.

### 1.2.5.2 The Steady State NOE

The steady state NOE (SS-NOE) experiment (45) uses a similar scheme to the T- NOE experiment except that instead of a selective 180° pulse a continuous, low-power, selective pulse is applied (Figure 1.26).

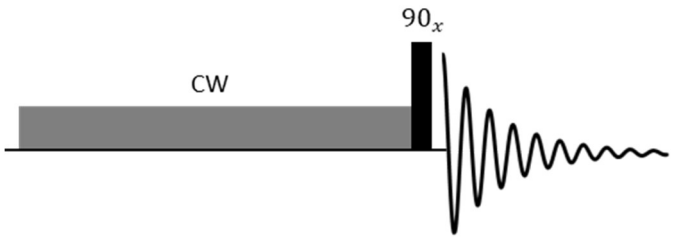



*Figure 1.26 - The pulse sequence for the SS-NOE experiment. In place of the selective 180° pulse a selective low-power pulse (CW) is used.*

If applied to spin *I*, instead of an inversion of spin populations there will be an equalisation. As there will be a continuous pulse equalising spin populations the system will not be able to relax via transverse relaxation leaving only the cross-relaxation pathway. This method however does not discretely describe the NOE as the T-NOE does. Here the NOE enhancement is not only a product of cross relaxation but also self-relaxation. However, this can be more sensitive as theoretically it should produce the maximum NOE, larger than the T-NOE, as a spin can only relax through cross-relaxation.

## 1.2.6 Saturation Transfer Difference NMR

Saturation Transfer Difference NMR (STD NMR) (50) can provide detailed information on the recognition of a ligand by a protein. Using STD NMR, it is possible to generate a map of contacts that a ligand makes with a protein during binding. As the name implies, this experiment can be thought of as taking place in three steps. Initially a protein is selectively saturated and then the saturation is then transferred to a ligand. This saturation transfer can then be measured as differences in the resulting spectrum when compared to a control spectrum (off-resonance spectrum minus on-resonance spectrum).

STD NMR is an attractive technique for the study of ligand binding as it does not require an expensive isotopically labelled protein and only requires a low protein concentration in the range of 10's of μM (51). Additionally, once an appropriate sample has been prepared, the technique is easy to implement in most NMR laboratories. Despite the potential for STD NMR to be a versatile tool for probing ligand-protein interactions, there are limitations that should be considered when deciding if it is appropriate for a given study. The most important of which is, due to sensitivity to binding kinetics, STD NMR suffers from low sensitivity to ligands with very strong binding to the target. These are typically ligands with dissociation constants ($K_d$) in the nM range (51).

### 1.2.6.1 The STD NMR Experiment

In STD NMR, like other difference experiments, two spectra are recorded. These spectra are called the on-resonance spectrum and the off-resonance spectrum (50, 51). For the on-resonance spectrum there is a selective radiofrequency pulse that directly irradiates the signals of the protein. This pulse selectively saturates the protein only over a narrow band of signals. However, for larger molecules there is an efficient cross relaxation which spreads saturation directly to nearby spins (Figure 1.27). The saturation can then be propagated further by spins that have been saturated by cross relaxation indirectly to other spins by a process called the relayed NOE. This process in which saturation is propagated from the site of irradiation through the protein is called spin diffusion. The off-resonance spectrum does not have a selective saturating pulse and the protein is not saturated. This spectrum is used as a reference spectrum.

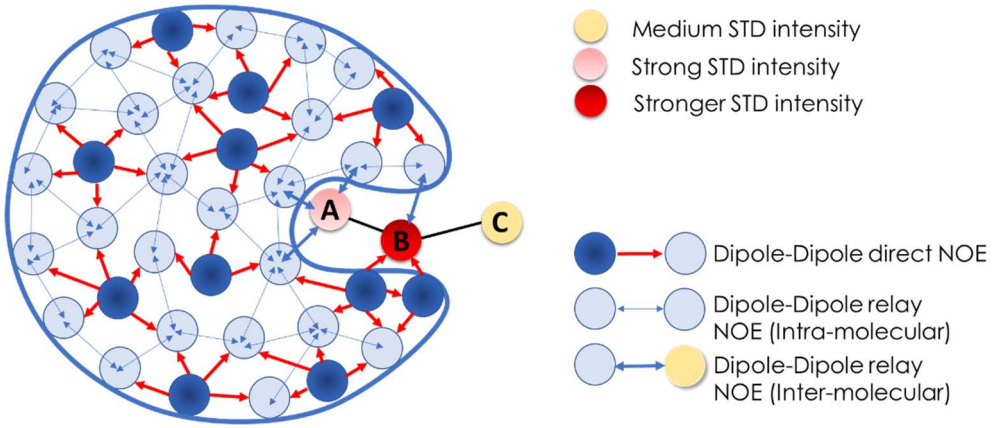



*Figure 1.27 - Schematic representation of spin diffusion. The saturated spins (dark blue circles) saturation to neighbouring spins through a direct intramolecular NOE. This then perturbs neighbouring spins (light blue circles). This saturation can then be transferred to a bound ligand through intermolecular relay NOE. Used with the permission of Dr J. Angulo.*

When a ligand binds to a saturated protein there is a transfer of saturation by intermolecular spin diffusion. While a ligand is bound to the protein it will temporarily take on the qualities of the macromolecule. This means that during binding the ligand will behave more so like a large molecule allowing for efficient spin diffusion to the ligand (Figure 1.28).

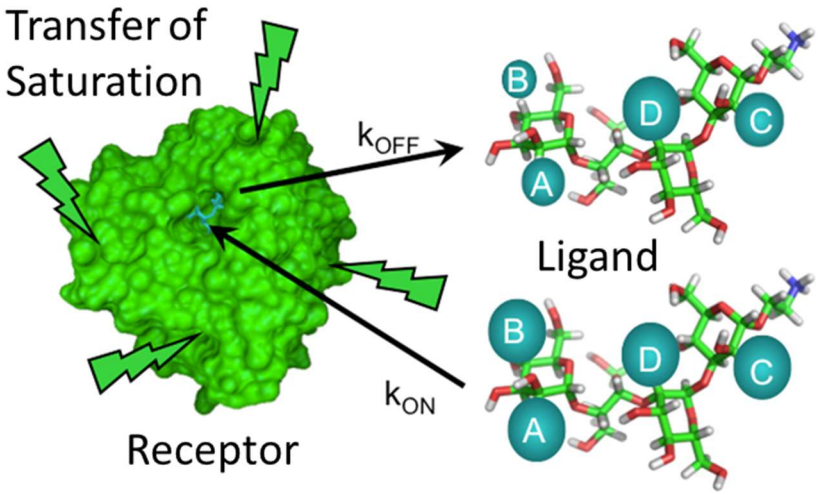



*Figure 1.28 - Transfer of saturation from the receptor during ligand in STD NMR. Prior to binding (bottom ligand) the spins are not perturbed. On binding the saturation transfer perturbs the signals of the ligand. The perturbations will be reflective of the closeness of the contacts the ligand makes with the protein. Adapted from (52).*

After a certain time, dependent on the binding affinity of the ligand, the ligand will then dissociate and move into the solvent. The free ligand will then be 'labelled' with saturation. Over the duration of the saturating pulse the process of binding, intermolecular spin diffusion and unbinding leads to a build-up of labelled free ligand. It is important that there is a build-up of ligand in solution as while bound to the protein saturation transfer is not readily observable due to spin diffusion. This is because while bound to the protein it will behave more like a large molecule and will be subject to efficient $T_2$ relaxation.

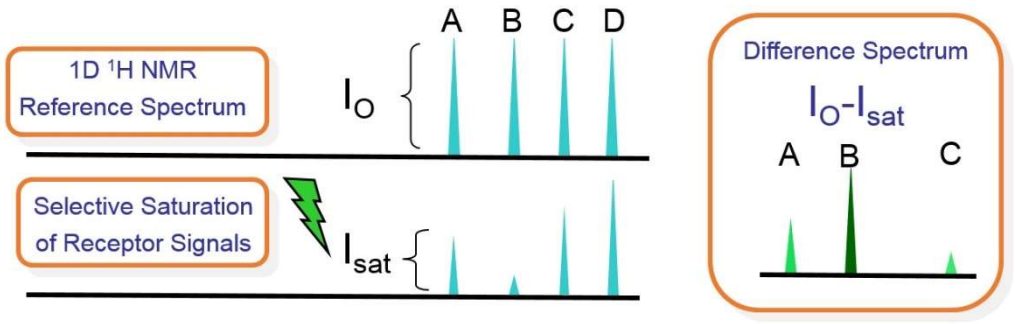



*Figure 1.29 - Cartoon schematic of off- (top) and on- (bottom) resonance spectra. Subtracting the on-resonance spectrum from the off-resonance (reference) spectrum produces the difference spectrum (right). The intensities of the signals in the difference spectrum approximately correlate to the proximity of a hydrogen from the ligand to the surface of the protein. Taken from (52).*

After the saturating pulse, a 90° pulse is then applied tilting the magnetisation into the transverse plane. During acquisition, the resulting spectra will show modulation of the signals of the ligand. This on-resonance spectrum ($I_{sat}$) can then be subtracted from the off-resonance reference spectrum ($I_0$) to produce a difference spectrum ($I_{diff}$) (Figure 1.29). The signals in $I_{diff}$ (Equation 9) can then be normalised and expressed as percentages (STD%) (Equation 10) (50, 52).

(9) $$I_{diff} = I_o - I_{sat}$$

(10) $$STD\% = \frac{I_o - I_{sat}}{I_o} \cdot 100$$

The intensities for each signal in the difference spectrum will indicate how close a particular hydrogen atom within the molecule contacts the protein in the bound state. This information, as it is strongly distance-dependent, can be used to find structural information about the ligand in the bound state. Finding this information is fundamental in the application of STD NMR.

For the optimal design of STD NMR experimental conditions, it is important for the ligand to be in a large excess. This is important as during fast exchange with a protein, the large excess averages the NMR properties of the ligand to be the weighted sum of the bound and free state. This ensures that the ligand maintains the properties of a small ligand, presenting sharp peaks and limited spin diffusion. Additionally, with a large ligand excess the protein binding sites will be saturated, maximising the amount of intensity in the difference spectrum. Furthermore, the saturation of binding sites in the protein will prevent re-binding. This will make it more likely that ligands which have undergone no spin diffusion will bind increasing the intensity of the difference spectrum further.

## 1.2.6.2 Binding Kinetics for STD NMR

Since its development STD NMR has become a popular technique for drug development and the study of biomolecular interactions, especially those involving carbohydrates. This is because these early hits in drug discovery and carbohydrate ligands can typically have $K_d$ values in the mM to µM range. As has been touched upon the sensitivity of STD NMR to the binding kinetics of ligands being studied is important to the technique.

$$P + L \underset{k_{off}}{\overset{k_{on}}{\rightleftharpoons}} PL$$

In a simplified on-site model, representing a protein with a single binding site for a particular ligand, the system can be represented as three components: the Protein (P), the ligand (L) and the complex (PL). The rate of the forward binding process ($k_{on}$) represents the probability there will be a contact between protein and ligand resulting in binding. For many systems it is possible to assume that this will be dependent on diffusion and in the range of $10^8$ to $10^9$ $M^{-1}$ $s^{-1}$. The rate of the backward unbinding process ($k_{off}$) is inversely proportional to the half-life of the complex. This means that $k_{off}$ can be used to describe the amount of time that the ligand spends bound to the protein or the residence time.

At equilibrium, the relative concentration of unbound ligand protein to the concentration of the complex will be determined by the values for $k_{on}$ and $k_{off}$ (Equation 11).

(11) $$k_{on}[P][L] = k_{off}[PL]$$

The ratio of $k_{off}$ to $k_{on}$ gives the dissociation constant ($K_d$) of the system (Equation 12).

(12) $$K_d = \frac{[P][L]}{[PL]} = \frac{k_{off}}{k_{on}}$$

Here it can be seen that for ligands with high affinity, or lower values for $K_d$, at equilibrium the complex will be favoured and vice versa for ligands with low affinity. For some systems it is not possible to assume that $k_{on}$ will be dependent on diffusion. Such systems typically involve more dynamic binding or conformational re-arrangements to accommodate a ligand. Given a value of $k_{on}$, $K_d$ is proportional to $k_{off}$ which is related to the residence time of the ligand. Instead, the binding kinetics can be considered from the perspective of the fraction of bound protein ($f_B^{PL}$) at equilibrium which can be derived from Equation 11 and 12 (Equation 13).

(13) $$f_B^{PL} = \frac{[L]}{[L]+K_d}$$

If plotted, Equation 13 describes a Langmuir Isotherm (Figure 1.30). This plot is linear for values of [L] which are significantly lower than $K_d$ and become asymptotic as [L] increases and becomes significantly larger than Kd. For large values of [L] where $f_B^{PL}$= 1 the protein is saturated. When the 50% of the protein is occupied ($f_B^{PL}$=0.5), [L] will be equal to $K_d$ as at that concentration of [L] the protein is at half saturation. This is the definition of $K_d$.

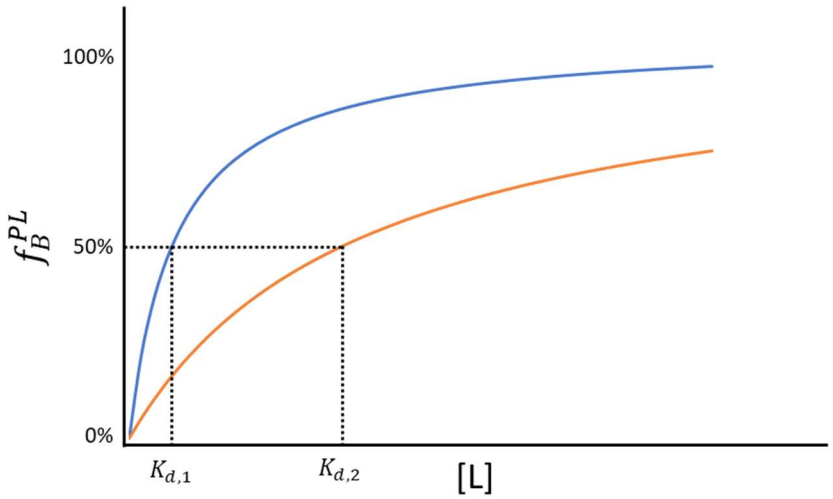



*Figure 1.30 - Langmuir isotherms for Ligand binding. Shown are a strongly binding ligand (blue) ($K_{d,1}$) and weakly binding ligand (orange) ($K_{d,2}$). The dissociation constant can be found from the isotherm by finding the ligand concentration where the fraction of bound protein is 50%.*

It is important to emphasise how crucial the binding kinetics of the ligand-protein system being studied is to STD NMR. As mentioned, $k_{on}$ is assumed to be diffusion dependent, leaving just the kinetics dependence on $k_{off}$. For an experiment to yield the best results the ligand must be in 'fast exchange' between the free and protein-bound states and therefore must spend a short amount of time bound to the protein. This would represent ligands with higher values for $k_{off}$ or ligands with low residence times in the bound state. The reason for this is that ligands binding tightly to the protein have long residence times, so that self-relaxation will occur to such a degree in the bound state that there is not a significant build-up of labelled, dissociated ligand in the solvent to be measured. As while in the bound state the ligand will take on qualities of a large molecule its residence time should be lower than 1 ms to 10 ms, the typical time for full $T_2$ relaxation. This corresponds to ligands with $K_d$ values in the low millimolar to low micromolar range, which would correspond to residence times less than 1 ms.

### 1.2.6.3 Binding Epitope Mapping

For the screening of small molecules STD NMR is useful as peaks in a difference spectrum indicate binding. As STD NMR is sensitive to molecules with low binding affinities, it can detect the binding of ligands which otherwise would be missed by other methods during screening (51, 52). This means that it is particularly useful in the screening of low-affinity small fragments and early-stage ligands in drug discovery campaigns. Additionally, STD NMR can also provide some qualitative structural information for the ligand in the bound state. The reason this information is qualitative is like that of the SS-NOE. While in the bound state, there will be some self-relaxation of the ligand. In addition, rebinding can disrupt the saturation transferred from the protein. Despite this, the intensities are still useful. The most intense peak in the difference spectrum will correspond to a hydrogen that is placed closest to the protein. This can be inferred as the most important interaction the ligand has with the protein. The rest of the intensities can then be ranked in order of STD%. This means that when discussing STD% they must be thought of as relative intensities indicating relative distances from the protein in the bound state, not explicit descriptions of intermolecular distances.

While it could be attractive to perform STD NMR at a fixed saturation time (e.g long enough for there to be strong enough intensities in the difference spectra), such as 2 s, this approach is limited. The reason for this is that protons within a ligand will relax at different rates in the bulk and the ligand can re-bind to the protein within these longer saturation times (52). This will mean that the level of saturation measured from the ligand will be compromised and therefore may not be representative of contacts made with the proteins. To overcome this a single experiment could be performed at a short saturation time, such as 0.5 s. However, the intensities in the difference spectra may not be intense enough to accurately measure. Where possible, STD NMR experiments should be performed using the initial growth rates method (51, 52).

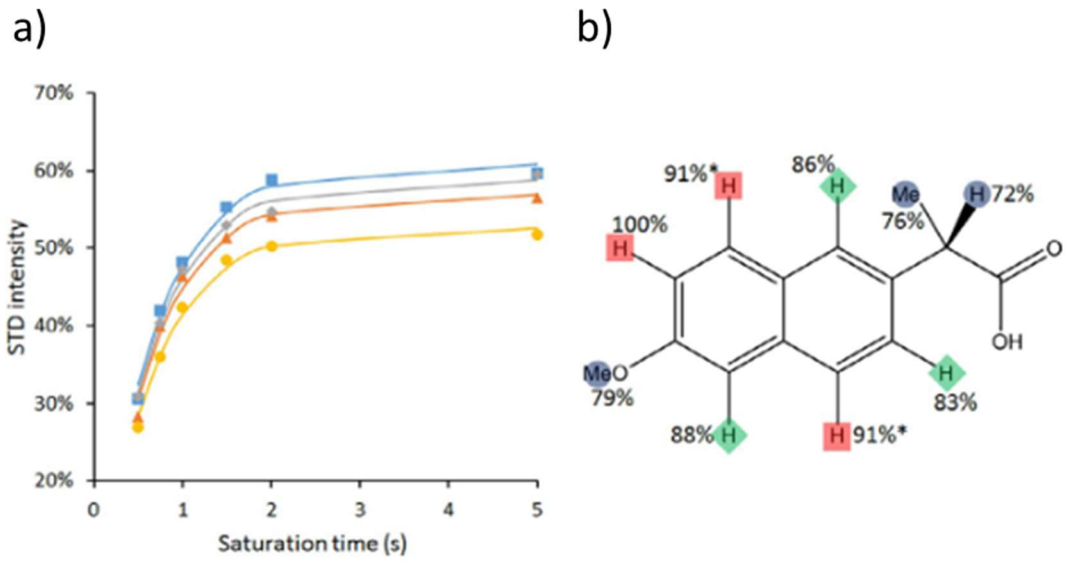



*Figure 1.31 - Build-up curves a) for the binding of naproxen b) with human serum albumin. STD percentages for the naproxen molecule are calculated from fitting to the measured STDs at different time points, normalised against the highest STD signal. Adapted from (51).*

In the initial growth rates method, multiple STD NMR experiments are performed where saturation time ($t_{sat}$) is typically varied incrementally from 0.5 s to 5 s, with more data points being collected between 0.5 s and 2 s. The STD% for each signal can then be plotted against $t_{sat}$ to produce a 'build-up' curve (Figure 1.31). These curves will all be monoexponential and plateau at the maximum STD% ($STD_{max}$) (Equation 14) (52, 53).

(14) $$STD_{t_{sat}} = STD_{max} - STD_{max} e^{(-k_{sat} t_{sat})}$$

Where $k_{sat}$ is the saturation transfer rate constant. For increasing saturation times there will be a loss of saturation for unbound ligands due to relaxation in the bulk. The reason that the build-up curves plateau is because for longer saturation times there will be a steady state as the saturation lost from relaxation is replaced at the same rate by the unbinding of saturated ligands from the protein. From monoexponential fitting, using equation 14, $k_{sat}$ and $STD_{max}$ can be derived. From the equation, initially, i.e. for very low saturation times, the values for $STD_{t_{sat}}$ will increase with a linear relationship with $t_{sat}$ dependent on $k_{sat}$ (Equation 15) (52).

(15) $$STD_0 = k_{sat} \mathrm{STD}_{max}$$

This early linear relationship is called the initial growth rate. For each hydrogen nuclei the initial growth rate can be found ($STD_0$). These values are then arbitrarily normalised against one value of $STD_0$, typically the largest, to give a percentage which can then be mapped onto the ligand. Once this mapping has been performed the result is a binding epitope for a ligand (Figure 1.31). This result will be a semi-quantitative description of the contacts the ligand makes with the protein on binding, with higher percentages indicating a closer or more important interactions

## 1.2.7 Two-Dimensional NMR

Until now NMR experiments have been discussed from the perspective of observing signals from a single nucleus along one frequency dimension. Two-dimensional NMR adds a second frequency dimension from which additional information can be found. Using this method, cross peaks between the two frequency dimensions allow for correlations to be found between signals in either dimension. This approach to NMR is particularly useful when studying large, structurally complex molecules such as proteins or branched carbohydrates. This is because the one-dimensional spectra for $^{1}H$ can be too crowded for reliable interpretation. However, the use of two-dimensional NMR is not limited to the study of large molecules and is often used as part of the routine structural determination of small molecules.

An important note on the nomenclature of one- and two-dimensional spectra is that they both have more dimensions than the name suggests. A one-dimensional spectrum is in fact two-dimensional, including frequency and intensity axes. The same idea is true for two-dimensional spectra except there are two frequency axes alongside one intensity axis. This is the reason despite being called a two-dimensional spectrum they can have a three-dimensional appearance. In practice, two-dimensional spectra are presented as a contour plot like a map of a hilled landscape.

## 1.2.7.1 Generation of the Second Dimension

In two-dimensional experiments, the second dimension is not directly observed. Instead, the second dimension is constructed though processing, so it is synthetic. Though there are important differences between homo and heteronuclear experiments, they follow the same general scheme to generate the second dimension. The general scheme of two-dimensional experiments involves four important steps: preparation, evolution time ($t_1$), mixing and detection ($t_2$) (47) (Figure 1.32).

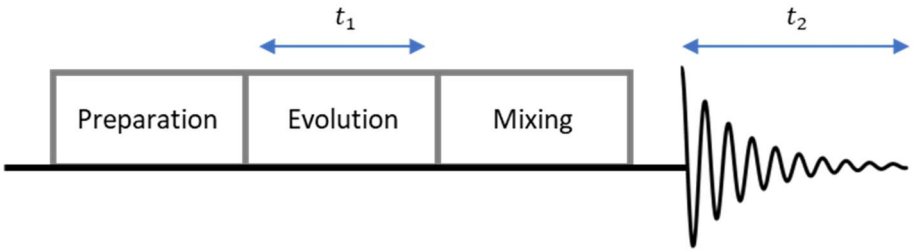



*Figure 1.32 - Generalised scheme of a two-dimensional NMR experiment. Shown are the blocks for preparation, evolution ($t_1$) mixing and acquisition ($t_2$).*

Preparation is where coherence between two spins is generated, which is followed by a variable evolution period. During the mixing period the coherence generated and evolved in chemical shift during $t_1$ is exchanged or 'mixed', after which it is then converted into a signal that is observable during acquisition. As the second dimension will need to be a frequency dimension, there needs to be a time-dependent change measured. This is achieved by performing multiple acquisitions, incrementally varying the evolution time $t_1$. To illustrate this process, the pulse sequence for Correlation Spectroscopy (COSY) can be used. In this experiment, 90° pulses are used in the places of preparation and mixing (Figure 1.33) (47).

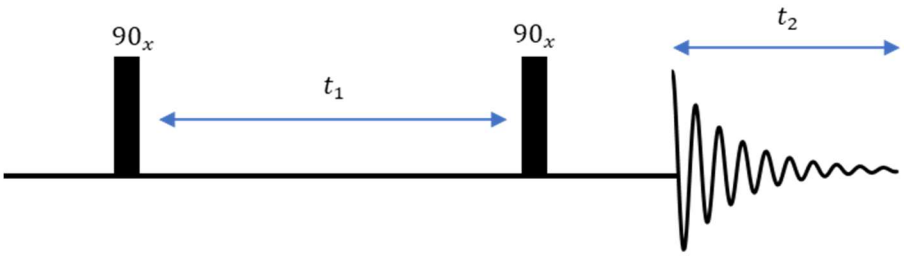



*Figure 1.33 – Pulse sequency for COSY, in which 90° pulses are used in place of preparation and mixing and a delay ($t_1$) is used for evolution.*

If we consider an isolated, uncoupled spin with a Larmor frequency of $\nu_0$, and apply two 90° pulses one directly after with a $t_1$ of 0 its magnetisation will be just inverted. This is because in this scheme the two 90° pulses are the equivalent of a 180° pulse. During acquisition, there would then be no signal to detect as there is no magnetisation in the

transverse plane. When evolution time is not 0, the x component of the magnetic vector ($I_x$) is able to evolve dependent on its Larmor precession (Equation 16) (47).

(16) $$I_x = I_0 \sin(v_0 t_1)$$

From the equation, as $t_1$ increases from 0, $I_x$ will be modulated in a sinusoidal manner at a frequency that matches it's Larmor precession. This modulation decays over time for extended values of $t_1$ as the nuclei within the system undergo relaxation. If the intensities of the peak at the Larmor frequency of the spin are plotted with increasing $t_1$ the result would be a free induction decay (FID) like the one acquired during direct acquisition, called an interferogram. If this interferogram is then Fourier transformed a single peak will be found at the Larmor frequency of the spin. These interferograms can be found for every frequency in detection and each Fourier transformed. If then plotted as a function of $t_1$, acquisition ($t_2$) and intensity the result will be a plot with two frequency dimensions, indirect ($f_1$) and the direct ($f_2$), containing a single cross peak at the Larmor frequency of the spin (Figure 1.34).

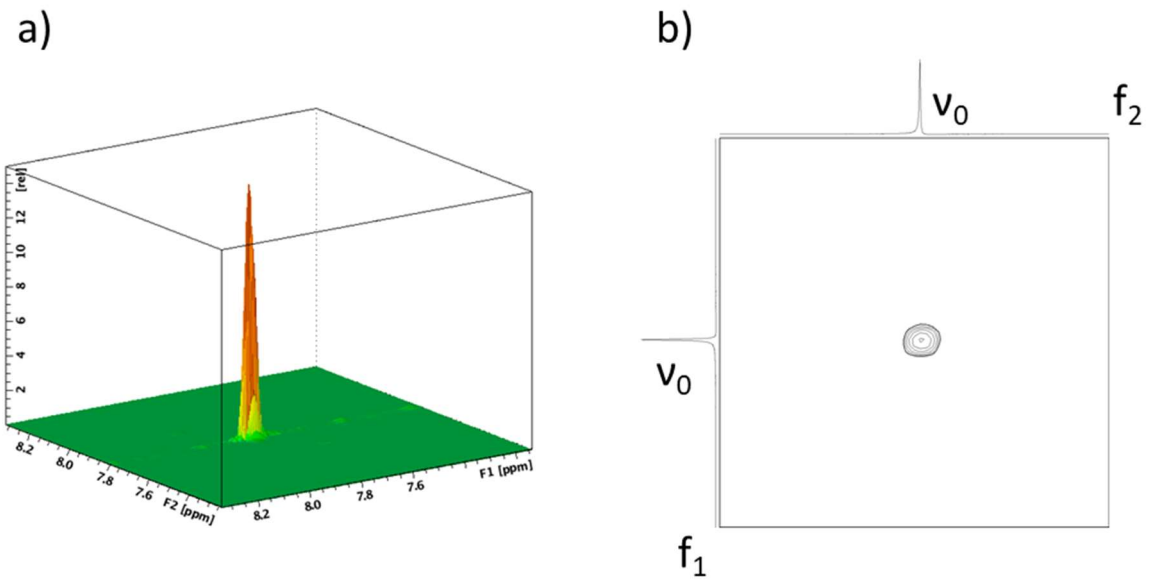



*Figure 1.34 - A cross peak in a two-dimensional spectrum, presented as a) a 3D peak showing the $f_1$, $f_2$ and intensity axes and b) the typical presentation of a two-dimensional spectra as a contour plot. Used with permission from Dr I. Delso.*

The resulting spectrum from this example only shows what a one-dimensional experiment would have shown that there is a peak at $v_0$ for the isolated spin. This is because in this example the frequency in $f_1$ is the same as that in $f_2$ ($v_1 = v_2$) (47). For there to be more information to extract, there would need to be coupled spins. Then, during the processing of the FIDs there is information where the frequencies found in $f_1$ are different from those in $f_2$ ($v_1 \neq v_2$). This means that there would now be magnetisation which evolves during $t_1$ with a frequency of $v_1$ which is altered to evolve during $t_2$ with a different frequency of $v_2$.

## 1.2.7.2 Homonuclear Experiments

Two-dimensional homonuclear experiments are designed to observe the relationship between two of the same nuclei. In practice such experiments are commonly used to observe correlations between two hydrogen nuclei. COSY and Total Correlation Spectroscopy (TOCSY) are important experiments, typically used in the assignment of small to medium sized molecules. Nuclear Overhauser Effect Spectroscopy (NOESY) is another example that can be used to assist assignment but also has application in studying the conformation of molecules, as, for example, ligands bound to a protein.

### 1.2.7.2.1 Correlation Spectroscopy (COSY)

$^1$H-$^1$H COSY experiments produce two $^1$H chemical shift axes. An important feature of COSY spectra is that there will be a diagonal set of cross peaks from matching shifts, which are flanked symmetrically by other cross peaks (47). In the COSY experiment magnetisation is allowed to evolve with chemical shift during the evolution step. During mixing, the magnetisation of one spin then is then transferred to coupled spins. The cross peaks flanking either side of the diagonal are correlations that arise from this coupling which is typical a range three bonds. This information can be used to identify which signals in the direct spectra come from coupled hydrogen nuclei (Figure 1.35). Though COSY can be used on larger organic molecules such as peptides and branched carbohydrates, $^1$H-$^1$H homonuclear experiments are limited by the resolution of hydrogen. This means that the challenges of interpreting the crowded one-dimensional spectra often translate to two-dimensional spectra.

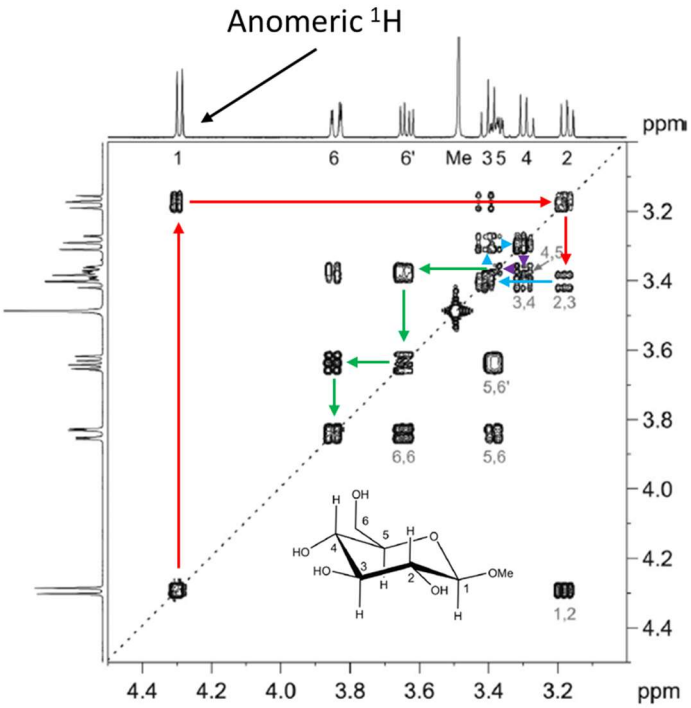



*Figure 1.35 - COSY spectra of β-Gluc-OMe. Shown are the steps taken to assign the signals. Using the anomeric hydrogen (highest chemical shift) as a starting point the signals can be assigned stepwise, as shown with the coloured arrows. Adapted from (47).*

## 1.2.7.2.2 Total Correlation Spectroscopy (TOCSY)

TOCSY, though it does show the same information as a COSY, it is not limited to just scalar coupling between neighbouring spins (Figure 1.36). Instead TOCSY experiments can show correlations within a particular spin system at longer range. A spin system in this context is a sequential chain of coupled $^{1}H$ nuclei, for example those within a sugar ring. The TOCSY experiments initially follows the same sequence as COSY however during mixing a continuous pulse called a spin-lock is used. This has the effect of relaying magnetisation along the chain of coupled spins in a spin system (47). The resulting spectra from a TOCSY experiment will therefore contain the same diagonal and cross peaks in a COSY, but there will be additional correlations which come from the relay of magnetisation from the spinlock. In the TOCSY spectra longer range correlations will present with reduced intensity. This is because the relay is not perfectly efficient, and decays over longer ranges. This is useful information as it can provide an estimate of the relative distance between correlation spins within a spin system.

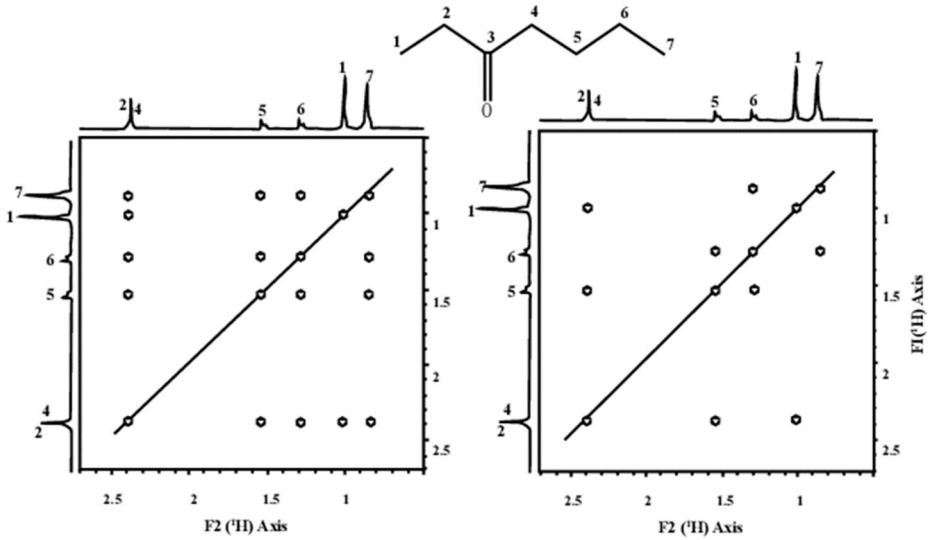



*Figure 1.36 - Comparison of the COSY (left) and TOCSY (right) spectra for 3-heptanone. The TOCSY spectra can be seen to contain the same information as the COSY but, with additional correlations which correspond to couplings within the spin system. Taken from (54).*

## 1.2.7.2.3 Nuclear Overhauser Effect Spectroscopy (NOESY)

In a NOESY spectrum correlations are seen for nuclei that are close together in space, with magnetisation being transferred by the NOE. This is achieved by inverting the populations of spins in a similar manner as the T-NOE and then allowing NOEs to evolve during mixing (47). The range of the NOE detected by a NOESY experiment can be controlled by the mixing time. In practice, NOESY experiments are performed using different mixing times to provide a full picture of long and short range through space correlations. The resultant spectra from a NOESY experiment will again present similarly to a COSY experiment, though with many more correlations (Figure 1.37). These correlations now, include those spins that are close in space.

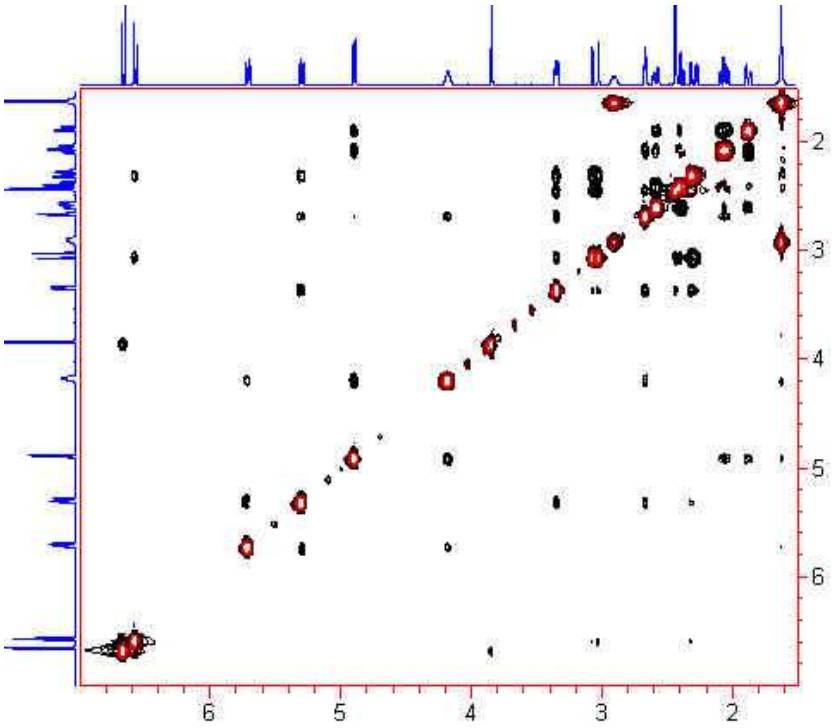



*Figure 1.37 - NOESY spectra for codeine. Used under permission of the Wikimedia Commons. (https://commons.wikimedia.org/wiki/File:Noesy.jpg)*

### 1.2.7.3 Heteronuclear Experiments

Two-dimensional heteronuclear nuclear experiments observe the coupling between two different types of nuclei. Notable example is the Heteronuclear Single Quantum Correlation (HSQC), and the version for large biomolecules: Transverse Relaxation Optimised Spectroscopy (TROSY). The most common application of these experiments is for observing the coupling between $^{1}H$ and $^{13}C$ or $^{15}N$. Though these correlations will be the focus of this discussion on heteronuclear experiments, it is possible to perform these experiments on other pairs of NMR-active nuclei. Heteronuclear experiments are often used alongside homonuclear experiments for the structural elucidation of small molecules. However, due to increased resolution in the indirect dimension from the larger distribution of shifts that $^{13}C$ and $^{15}N$ nuclei display, it can be used to study much larger molecules such as proteins. Particularly the version optimised for transverse relaxation (TROSY), is very efficient for large proteins in high magnetic fields. Important to this thesis, in addition to structural data, heteronuclear experiments are also able to find important information on protein-protein binding.

## 1.2.7.3.1 Heteronuclear Single Quantum Correlation (HSQC)

HSQC spectra provide information on pairs of nuclei of different nature that are covalently linked to one-another. The most used HSQC experiment focusses on single bond correlations however, Heteronuclear Multiple Bond Correlation (HMBC) experiments can allow for much longer-ranged correlations to be observed (47). In HSQC experiments preparation is used to transfer magnetisation from $^{1}H$ nuclei to a different type of nuclei, such as $^{13}C$ and $^{15}N$. This is achieved with the use of a series of pulses called insensitive nuclei enhanced by polarization transfer (INEPT) during preparation. Once this has been transferred magnetisation is allowed to evolve and is then converted to a detectable signal by the inverse of INEPT, called retro-INEPT.

The correlations in a HSQC experiment will be between coupled nuclei, presented as cross peaks (Figure 1.38). For the structural determination of small molecules these experiments are useful as they can correlate assigned signals from the one-dimensional $^{1}H$ spectra to those in the in the indirect spectra. As there is a larger distribution of chemical shifts in the indirect spectra, HSQC can also be particularly helpful for assigning a crowded one-dimensional $^{1}H$ spectrum.

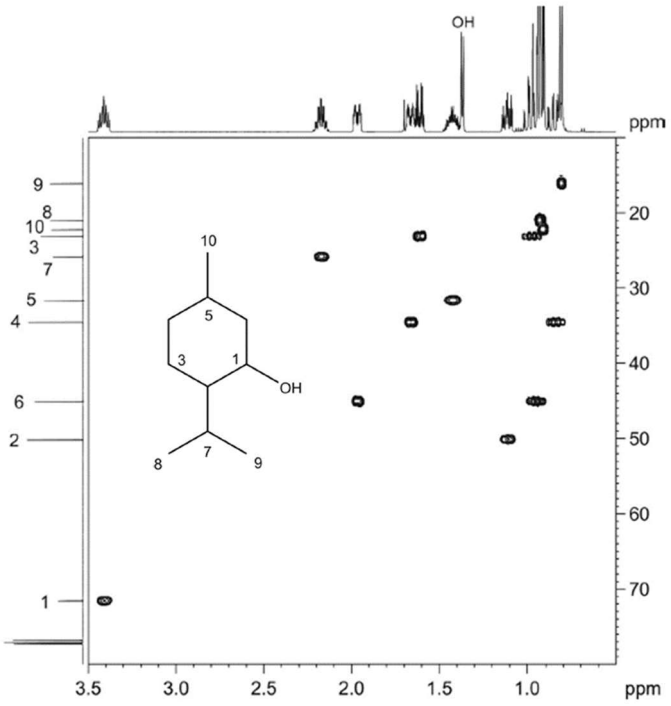



*Figure 1.38 - $^{1}H$-$^{13}C$ HSQC spectrum for menthol. Peaks in the $^{13}C$ indirect spectrum have been assigned to the molecule, denoted with numbers corresponding to the structure. Adapted from (47).*

## 1.2.7.3.2 Two-Dimensional NMR to Study Protein Interactions

An important application of HSQC is the study of protein-ligand and protein-protein interactions. To study protein interactions by HSQC, it is important to have an assigned $^{1}H$-$^{15}N$ HSQC protein spectrum. Assignment experiments are performed using dual $^{13}C$-$^{15}N$-labelled protein samples, where, in some cases, also deuteration is needed, particularly for very large proteins. Using three-dimensional NMR experiments it is possible to assign the backbone of proteins, producing an assigned $^{1}H$-$^{15}N$ HSQC protein spectrum (Figure 1.39).

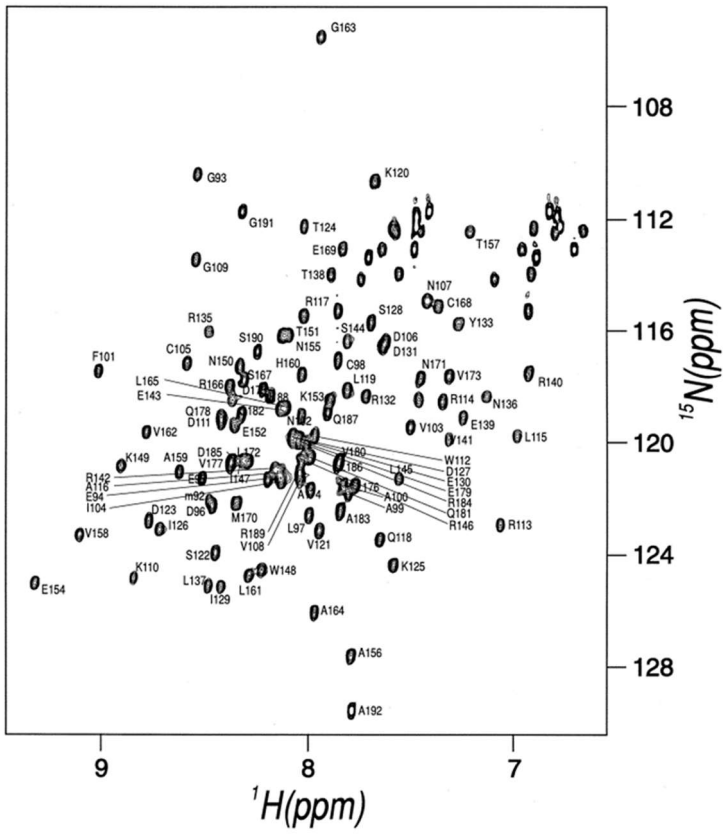



*Figure 1.39 - $^{1}H$-$^{15}N$ HSQC spectrum for the death domain of FADD. Taken from (55).*

To study protein-protein and protein-ligand interactions using HSQC, titrations are performed. During these experiments the concentration of the isotopically labelled protein is typically fixed while the concentration of a ligand or protein titrant of interest is varied. During such an experiment the chemical shift of cross peaks for the isotopically labelled protein will be perturbed. These perturbations can be followed over the course of the titration and be used to map the contacts the labelled protein has with the titrant.

### 1.2.5.3.3 Transverse Relaxation-Optimised Spectroscopy (TROSY)

A challenge for the study of proteins by HSQC is that they can often be large macromolecules or form large complexes. The trouble with such systems is that they will be subject to very efficient $T_2$ relaxation, and their spectra will present with extremely broadened lines (Figure 1.40). To overcome this a variation of the HSQC experiment was developed, called transverse relaxation-optimised spectroscopy (TROSY) (56).

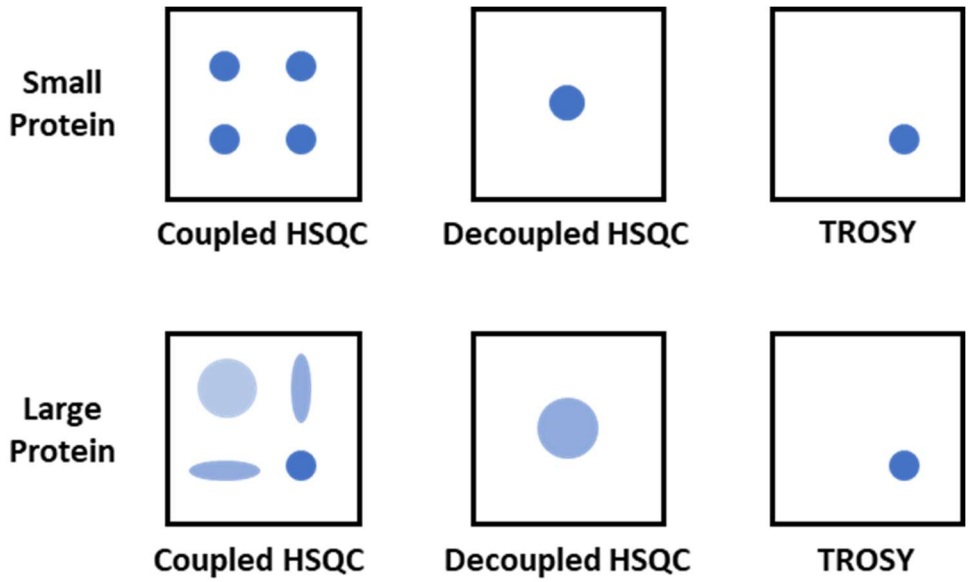



*Figure 1.40 - Illustration of the use of TROSY compared to HSQC. For small proteins (top row) that do not show pronounced $T_2$ relaxation the coupled HSQC spectra will show similar peaks. In a decoupled HSQC spectra these peaks collapse into a single sharp peak. As TROSY is selective for only part of the signal, the resultant spectra will only contain signals from that pathway. This is useful for large proteins (bottom row) as in their coupled HSQC some components will relax at different rates which results in broadening in the decoupled HSQC. As TROSY is selective for only the slowly relaxing components they present in the TROSY as a single sharp peak. Recreated from http://u-of-o-nmr-facility.blogspot.com/search?q=TROSY.*

In a standard HSQC experiment, as discussed earlier, during acquisition broadband decoupling is used. This has the effect of making cross peaks appear as single peaks. Instead, if an HSQC experiment was performed without decoupling, scalar $^{1}$H-S coupling would evolve during acquisition and the cross peaks would appear as multiplets. For $^{15}$N isotopically labelled proteins, the coupling observed in a spectrum would come from the

$^{1}H$-$^{15}N$ bond of backbone amines. The multiplet peaks from a coupled HSQC experiment present different line widths in large proteins, implying that the different components relax by different mechanisms. In a decoupled experiment the equivalent cross peak would be an average of the two cross peaks into one, averaging relaxation rates. When one of the peaks relaxes very quickly, this will reduce the overall intensity of a coupled cross peak. For particularly large protein or protein-protein complexes, this becomes a problem as peak intensity can fall to the level of noise making the spectrum very challenging to interpret.

For proteins above 20 kDa, at fields of 700MHz or higher, there are two major mechanisms which contribute to relaxation. These are the CSA of the nitrogen atoms and the dipolar interaction between the hydrogen and the nitrogen of the amide itself. The interference between these two mechanisms produces different relaxation rates for each component of a multiplet, resulting in the differing line widths. The TROSY experiment is designed such that the slow-relaxing components of the signal producing the sharper peaks are retained, and the fast-relaxing components are removed. To do this TROSY selects only for the components of the signal in which CSA and dipolar relaxation partially cancel out (Figure 1.40). This enhances the resolution and intensity of a $^{1}H$-$^{15}N$ spectrum despite removing half of the possible signal.

## 1.2.5.3.4 The Impact of Chemical Exchange on Protein NMR

In the study of protein interactions by HSQC, chemical shift perturbations are not all the same. Over the course of a titration, cross peaks can be seen to smoothly move from one position to another or even completely disappear only to reappear in a new location (47). The fundamental process that explains these behaviours is chemical exchange. In general terms, there are three main regimes for chemical exchange, fast, intermediate, and slow. These regimes refer to the rate of changes in the system over time, for example in the study of small molecules this could refer to an exchange between conformers. For the study of protein interactions, exchange typically refers to bound and free states. Exchange between these states will impact how a spectrum will appear for a particular system during titration. To explain how exchange impacts a protein spectrum, consider a single cross peak in position A when there is no titrant present. At the end of the titration, the protein is then saturated and then the peak will be at position B (Figure 1.41). For simplicity the one-dimensional projection of the cross peaks will be used to discuss the impact of chemical exchange.

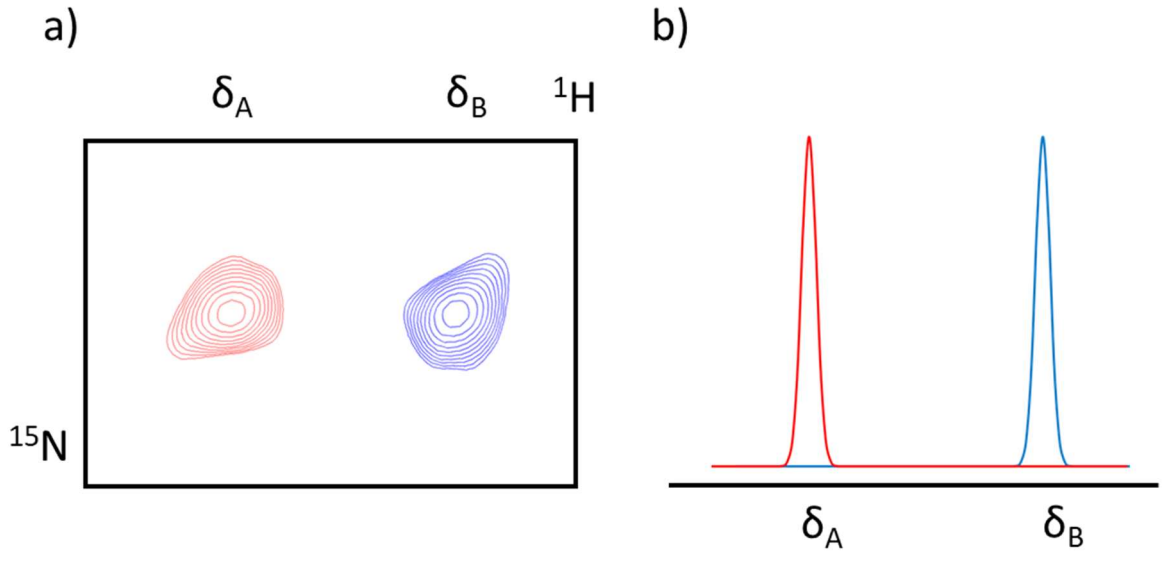



*Figure 1.41 - Comparison of the appearance of peaks in a two-dimensional and one-dimensional spectrum. Peak A ($\delta_A$) in presence of no titrant and Peak B ($\delta_B$) when the system is saturated with titrant. The two dimensional (a) and one dimensional (b) projection of peaks A and B.*

If titrant concentration is such that half of the protein is bound to the titrant, in the fast exchange regime there will just one cross peak in the middle of peaks A and B (Figure 1.42). This is because the binding and unbinding is so fast that there is an averaging of the observed frequency (chemical shift) of the signal. In the slow exchange regime, the bound and unbound states will be so long lived that it will be possible to observe both states at the same time and therefore there will be two equal peaks at positions A and B in the spectra (Figure 1.42). The intermediate exchange regime though is more complex. If a system were able to move from slow to fast regimes, for example by increasing temperature, the two peaks of the slow exchange would broaden with increasing temperature before coalescing into a single peak and then sharpening. At and around the point of coalescence, is the intermediate regime. Cross peaks from systems undergoing this type of exchange are difficult to observe as they are broadened such that they can be hard to distinguish from noise (Figure 1.42).

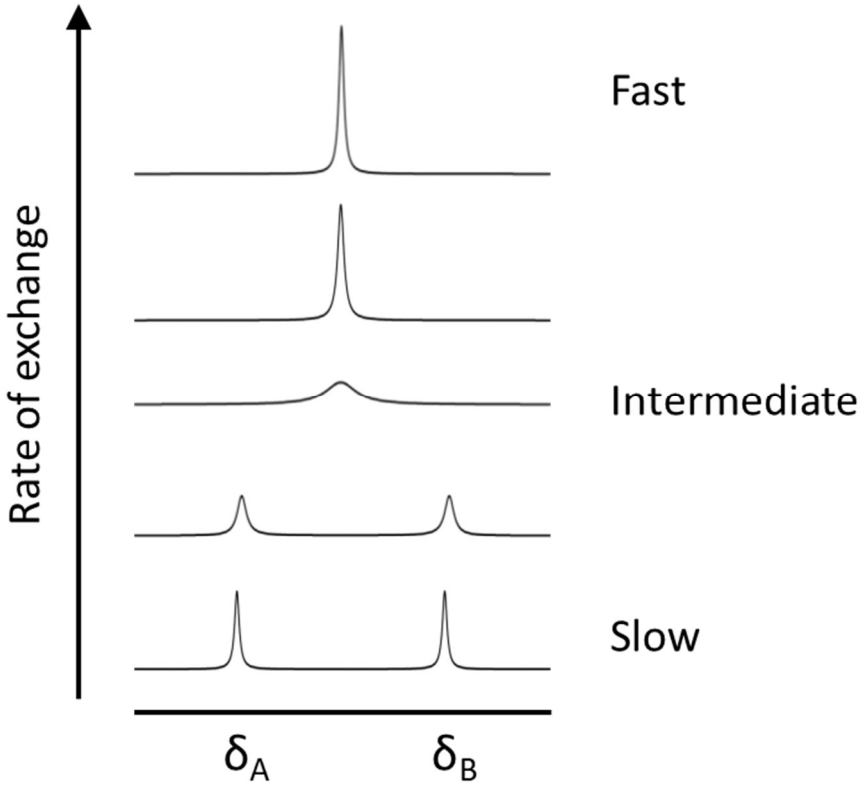



*Figure 1.42 - Impact of exchange regime on one-dimensional peaks. In the fast exchange regime, there is an averaging of the chemical shifts of the two signals seen in the slow exchange regime. In the intermediate regime signals can become so broad that they cannot easily be seen. Adapted from (47) .*

During a titration, the bound and unbound populations are not always equal. This means that over the course of a titration the exchange regimes instead must be considered from the perspective of an unequal distributions of bound and unbound protein. For the fast regime there will still be an averaging of position of the two peaks however, this will not be in the midpoint. Instead, the cross peak will move from $\delta_A$ to $\delta_B$ along the titration as the protein becomes saturated (Figure 1.43). In the slow regime two peaks will still be observed but their intensity ratio will be reflective of the bound and unbound populations. Along the titration as the protein becomes saturated the intensity of $\delta_B$ will increase while $\delta_A$ decreases until only the cross peak at $\delta_B$ is observed (Figure 1.43). Intermediately exchanging systems present a challenge during titrations. In the early stages of the titration there can be some perturbation of the cross peaks, however as the protein becomes more saturated the signal intensity will reduce until it is indistinguishable from noise.

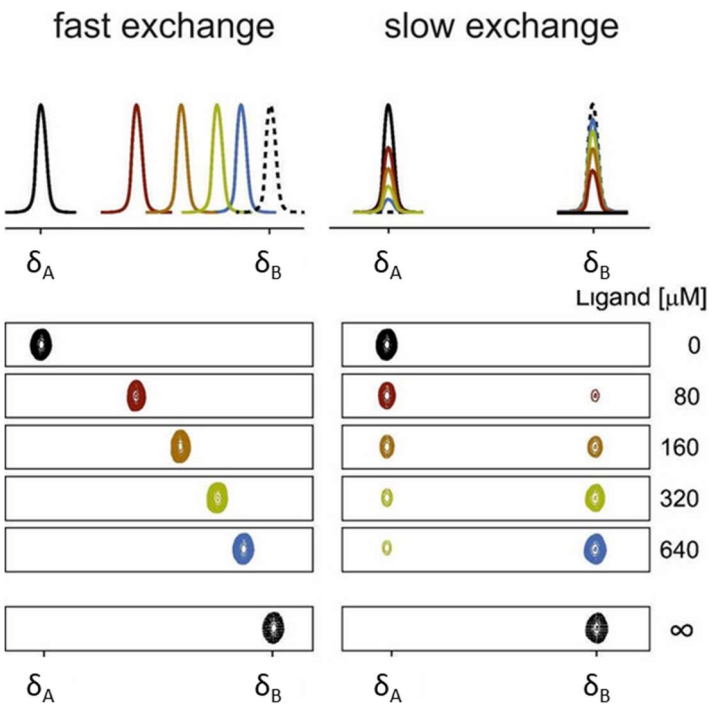



*Figure 1.43 - The effect of exchange on perturbations during titrations. Changes in spectra during titrations for fast exchanging systems (left) and slow exchanging systems (right), showing the one-dimensional projection of a peak (top) and the two-dimensional projection (bottom) are shown. Adapted from (57).*

# 1.3 Molecular Modelling

Molecular modelling encompasses a group of computer simulation methods that attempt to recreate the behaviour of real-world systems at atomic resolution. In many different experimental approaches, the details of molecular motion can be hard to observe directly. Molecular modelling provides an opportunity to gain this information but becomes especially powerful when validated by experiment. In the study of protein-protein and protein-ligand interactions molecular modelling is a popular structural prediction tool which is often paired with NMR.

## 1.3.1 Molecular Mechanics

Molecular mechanics (MM) is a method of molecular modelling that uses classical mechanics to simulate chemical systems. This means that the behaviour of the system is not considered at the level of the electron or nucleus. Instead, atoms are just considered to be particles with mass and the bonds between them like springs. The focus of MM is optimising geometry, such as bond angles and torsions, while forming sensible non-bonding interactions. Important to keep in mind is that MM only considers the ground state of the system, with thermodynamic and kinetic parameters not within the scope of such simulations (58). While it is possible to argue that MM is a more simplistic method of modelling, it is an essential tool for structural prediction and flexible enough to be used in the study of many different chemical systems. Using MM, it is possible to explore the conformational freedom of molecules, the binding of small molecules to proteins and even approximate protein-protein interactions.

Within the field of biochemistry, the features offered by MM has positioned it to be an excellent tool for screening potential small molecules or ligands, model the binding of known ligands to proteins. The latter use of MM often finds its niche when paired with protein crystallography. It can be the case that it is possible to crystallise a protein, but not gain crystallographic information on a bound ligand. Computational modelling using MM can allow for the prediction of molecules within binding sites and fill the gap between protein crystal structure and the structure in complex with a ligand of interest.

## 1.3.2 Force Fields

A force field is a set of terms that approximate the potential energy ($E_{pot}$) of the system and is used to evaluate the simulation of a chemical system. Within a force field the terms are separated into bonded ($E_{bonded}$) and non-bonded ($E_{non\text{-}bonded}$) interactions. The sum of bonded and non-bonded terms yields the potential energy ($E_{pot}$) of a system (Equation 17) (59).

(17) $$E_{pot} = \Sigma E_{bonded} + \Sigma E_{non-bonded}$$

Bonded terms consider the geometry features resulting from covalent bonds within a molecule. This includes stretches, bond angles, torsions and improper dihedrals and is often compared to Newtonian physics in its use of force constants. Non-bonded terms describe how atoms and molecules interact through non-covalent bonds. These include chemical interactions related to the partial charges of atoms within a molecule, including examples such as hydrogen bonding, van der Waals and π-π stacking. With these terms the forces between atoms that are within molecules and the forces between different molecules can be estimated.

Different chemical systems however can demand more terms to accurately model them. AMBER (Assisted Model Building with Energy Refinement) (60) is a popular example of a general, all-purpose force field that has many applications. Though able to simulate many different chemical systems, AMBER is best suited to the simulation of proteins and nucleic acids (61). For the study of small organic molecules, the AMBER-compatible force field called the generalised AMBER force field (GAFF) was developed (62). This force field contains the parameters for H, C, N, O, S, P and the halogens, which are atoms commonly seen in drug molecules and was originally designed for applications to pharmaceuticals. Depending on the application, systems that include carbohydrates can often demand the inclusion of quantum mechanics terms as the conformation of rings is strongly dictated by the anomeric effect (63). GLYCAM is an example of a force field that does include such terms, though with specificity can come sacrifice. Specialised force fields often focus on

specific terms and compromise on others. In the case of GLYCAM, while being a field that accurately simulates carbohydrates, it might not accurately simulate other systems.

Though AMBER, GAFF and GLYCAM are popular force fields which can be used when simulating many different systems, depending on application other force fields can be more appropriate to use. If an experiment is to simulate the binding of a ligand to a protein, then including terms that consider solvation would be useful. OPLS3 (Optimised Potentials for Liquid Simulation) (64) considers the thermodynamic properties of water, in addition to having general terms. Force fields can also be designed to take different approaches to modelling systems. CHARMM (Chemistry at Harvard Molecular Mechanics) (65) is a method that converts molecules into fragments which are then given parameters based on the force field terms. These fragments can them be combined with the assumption that the whole is just the sum of the parts.

For MM, as they have different limitations and specialities, the choice in force field can be important. While many systems can be simulated well by general force fields, some systems do demand more to be simulated properly. Experimental design and the purpose of an experiment can also determine which force field should be of choice. While it can be optimal to use a more specialised force field, in ligand screening or more preliminary experiments general force field with lower computational demand can be more appropriate.

### 1.3.3 Conformational Sampling

An important application of MM force fields is in the conformational sampling of molecules. Conformational sampling is often helpful to find low energy conformers of molecules. These conformers can then be used as starting points for other computational simulations such as docking and molecular dynamics. There are many different algorithms that can sample the conformation of molecules. These algorithms work by adjusting a chemical structure and recalculating the overall energy of a molecule each time, using a force field. The energy of a molecule with changes in structure can be plotted as a potential energy space (PES) (Figure 1.44) (66).

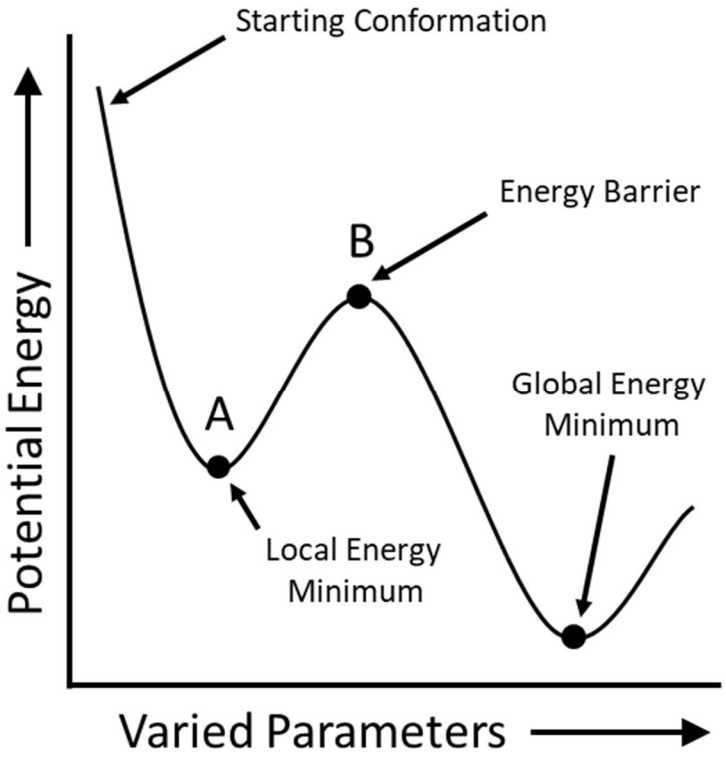



*Figure 1.44 - An example of a potential energy surface for arbitrary parameters. Over the course of a simulation, an algorithm can reach a local energy minimum (A) which would not be representative of the global energy minimum. To reach a global energy minimum, for some systems the sampling algorithm will have to overcome an energy barrier (B). Recreated from (66).*

There is one 'true' minimum in a PES called the global energy minimum but, within the PES are many peaks and troughs. There are often points in the PES that represent low energy states that are higher in energy than the global energy minimum. These points are called local minima and can be bound between energy barriers representing the energy that would be required to undergo a conformational transition within a molecule. When performing energy minimisation, some algorithms may not favour increasing the energy of a molecules to overcome these barriers and may not find the global energy minima of a molecule. Algorithms that perform conformational sampling are designed to be able to cross these barriers in order to explore the PES more thoroughly.

### 1.3.4 Molecular Docking

Molecular docking, or docking, is a technique often used to study the binding of molecules to proteins. As with many computational techniques, docking becomes a powerful predictive tool when paired with experimental data such as that from NMR, X-ray crystallography and more recently developed cryo-electron microscopy (Cryo-EM). Docking protocols encompass much of what has been discussed for MM but incorporates methods for finding the orientations of input structures, fitting molecules to an active site and assessing end structures.

To model a receptor active site for docking, a 'Grid' must be generated (67), which is a defined space that contains the receptor site in which ligands are docked. This step is important as it acts to restrict the amount of protein space that is explored during docking. A grid is designed by the user with boundaries around a desired binding site and can be used to restrict the position of a ligand during binding. Ideally, of course, the whole protein structure would be completely sampled however this is too computationally demanding for most practical applications of docking. In addition to defining the desired binding site, the grid can contain information on how flexible the protein structure can be when accommodating a ligand. With limited movement, a receptor grid is described as rigid whereas a more mobile receptor grid would be considered flexible. While the more modern understanding of ligand binding considers induced-fit binding, rigid docking is still a valid method of modelling systems. For example, in the case of X-ray crystallography a protein can be crystallised in the presence of a ligand, but the exact binding mode of the ligand can be difficult to find due to low electron density. In this case, the protein active site may already be in the correct conformational state to accept a ligand and therefore a more rigid model would be more appropriate to preserve this conformational state during docking.

With a defined site of interest, that has been parameterised by a user, docking then takes input ligand structures and, depending on the algorithm, attempts to explore the conformations they can adopt. As is the case for a protein the degrees of rotational, translational, and conformational freedom of a ligand can be too computationally demanding to completely explore. To simplify this, a similar strategy to handing the protein structure can be used, ligands structures can be treated as rigid or flexible. There are

applications for both methods but as ligand molecules are significantly simpler than an active site it is quite common for ligands to be treated as flexible rather than rigid.

In truth, it is rare for a protein or a ligand to be treated as completely rigid or completely flexible in docking. It is better to think of there being a spectrum and the degree of rigidity and flexibility of both protein and ligand being tuned to the requirements of the application and the limitations of hardware. With defined degrees of flexibility there are two common methods of exploring conformations. These are broken into systematic and stochastic methods. In systematic methods an input structure will have defined parameters varied within the set bounds of flexibility to find energetically minimised structures. This is of course done systematically and will therefore heavily depend on the input starting structure as the systematic method can often be trapped in local minima. In the stochastic method, randomised structures are generated within certain bounds of flexibility and their energies are found. In some stochastic approaches the low energy structures can be grouped and iterated on in an evolutionary manner with optimal structures being refined every generation.

### 1.3.4.1 Docking Algorithms

Docking algorithms vary in their approach but all attempt to solve the problem of fitting a structure to a defined binding site and can be divided into systematic and stochastic approaches. Two popular systematic methods include Matching Algorithms (MA) and Incremental Construction Algorithms (ICA) (67). MAs look to compare the shape of a ligand and functional groups present to a binding site. Each ligand can be fitted to the binding site and can be excluded from binding if a certain conformation does not fit. For ligands that do fit to a binding site the distance between complementary interactions such as hydrogen bonds and hydrophobic interactions can be found. Due to their more simplistic approach MAs require lower computing power and can be used to sample large sets of structures. ICAs take an input ligand structure and breaks it down into smaller fragments by breaking rotatable bonds. The largest fragment is then be fitted to an active site to act a starting point from which the rest of the molecule is built. ICAs have the advantage of considering the flexibility of a ligand during fitting, each fragment that is added the ligand is built to fit the receptor and can have a different conformation to the input structure.

Monte Carlo Algorithms (MCA) and Genetic Algorithms (GA) are the two popular stochastic docking algorithms (67). MCAs generate many different structures based on different rotations of bonds within a molecule, with the various structures filtered by their energy. This method is useful as it can sample more of a given PES and find local minima with high energy barriers. GA docking algorithms use a Darwinian approach to find bound structures. Variations on bond rotations and other parameterise encoded into 'genes' and 'chromosomes' which describe different families of structures. More successful, lower energy binding conformations are favoured, and their chromosomes are used to generate the new set of structures. This process is iterated upon in until an optimal conformation is found.

### 1.3.4.2 Scoring Functions

Important to understanding the results of molecular docking are Scoring Functions (SF) (67, 68). These are fundamentally applications of the force field which act to assess the binding conformations which are produced by docking. SFs are classified into four main categories: Empirical, Physics-based, Knowledge-based, and Machine Learning-based. Empirical SFs represent the more basic goal of scoring binding structures, summing up the energetics of non-bonding interactions to estimate the accuracy of the predictions from docking. The other SF methods build upon this concept trading computational time for enhanced accuracy.

Physics-based SFs are like Empirical SFs as they attempt to calculate the binding energies of a ligand to a protein using the force field. This method considers all the non-bonding interactions such as hydrogen bonds and van der Waals interactions that a ligand has with a protein and itself as well as the energetics of the ligand conformation in the active site. The force field contributions are limited to the enthalpic contributions to binding energies. To find the entropic contributions, Physics-based SFs also calculate the effect of solvation and the torsions of the ligand because of the binding conformation. In more recent years quantum mechanics (QM) has been incorporated into Physics-based SFs. Though more accurate, considering the contributions of charge and covalent bonding within a ligand, QM calculations come with a significant increase in the amount of computing power needed. Knowledge-based SFs are based on insights from libraries of

known complexes of ligands bound to proteins. These SFs measure the frequency of and the distances between sets of atom pairs and converts them into a potential mean force (PMF). PMF can be used to rank different interactions assuming that interactions that occur often are more favourable than ones that are rarer.

In more recent years, advances in machine learning algorithms have been taking advantage of in the field of computational chemistry. In contrast to older methods of scoring, Machine Learning-based SFs do not use discrete mathematical operations that can easily expressed. Instead, they take advantage of technologies such as Neural Networks and Deep Learning to train an algorithm to 'understand' ligand binding. Though it is possible for a trained Machine Learning-based SF to outperform other SF methods, it is typically used after docking to rescore output structures.

## 1.3.5 Molecular Dynamics

While MM is an excellent tool for structural prediction, one principal limitation it has is that any given result can only represent the system at a single time point. With MM as it has been described up until this point it is not possible to observe the evolution of a system over time. To observe this, molecular dynamics (MD) can be used. For a given protein-protein or protein-ligand system the influence that quantum mechanics has on how a system can change over time would be incredibly demanding to simulate. MD approximates this by modelling a system with molecular mechanics using Newtonian equations of motion (Equation 18 and 19).

(18) $$f_i = mx_i$$

(19) $$f_i = -\frac{dV}{dx_i}$$

Where, $f_i$ is the force acting on a particle, m is its mass, $x_i$ is its acceleration and V is the potential energy. MD simulates motion by calculating the forces acting on each atom within

a system and then moving them in keeping with the equations of motion. The forces on each atom are then re-calculated and the cycle repeats. In MD, motion is calculated stepwise because there is no solution for the differential equations of motion in complex systems. While it could on the surface appear to be a promising idea to adjust this time step to be long to save on computing time there are serious issues with this approach. Primarily if this is done the accuracy of the simulation will suffer but, for small atoms such as hydrogen this becomes a significant issue. This is because smaller atoms experience a higher acceleration for a particular force than other atoms due to their lower mass. This leads to them having much higher accelerations which for longer time steps can crash the simulation by producing extremely large bond lengths. In practice, this can be managed with the use of algorithms such as SHAKE (69) and RATTLE (70) which constrain hydrogen bonds lengths. With the use of hydrogen retraining algorithms typically the time steps used are in the order of femtoseconds with 1 fs and 2 fs being common.

A crucial feature of MD is that it simulates the system over a certain temperature in contrast to energy minimisation. While solving the equations of motion, for a molecules total energy there will be a kinetic energy contribution alongside potential energy. By incorporating kinetic energy, the dynamic motion of a system can be simulated allowing for time-dependent changes in conformation to be observed. At every time step the bulk physical properties of a system can be calculated allowing for the time-weighted distribution of different states to be found. If given enough time and energy, MD simulations can sample the entire PES of a system in accordance with the ergodic hypothesis (71). The time-weighted distribution of states would be equal to the average of the complete set of states the system can adopt.

### 1.3.5.1 Statistical Ensembles

A statistical ensemble is the set of microstates a system can adopt with a set constant thermodynamic parameter. One such statistical ensemble is the microcanonical NVE ensemble in which the number of molecules, volume and total system energy remains constant. This would represent an isolated system where no energy or additional particles are able to flow in or out. Using this ensemble, it can be difficult to explore the PES of a system. This is because as the energy remains constant for any given change in kinetic

energy there will be an equal and opposite change in potential energy. The result of this is that the total energy of a system never changes, meaning that energy barriers which exceed the total potential energy of the system cannot be crossed. For the simulation of protein systems this can be a problem as conformational changes would not occur leaving the system stuck in local minima.

Other ensembles exist which fix other parameters such as the canonical NVT and the isobaric/isothermal NPT. The NVT ensemble represents a closed system with a fixed number of molecules, volume, and temperature. In the case of the NVT system heat is applied from an external source. MD simulations, however, often use the NPT system in which the number of molecules, pressure, and temperature are kept constant. This model represents a closed system which allows contraction and expansion dependent on kinetic energy. The reason this ensemble is preferred is that it more closely imitates a real-world system. In the NPT ensemble MD simulations often utilise a Berendsen barostat and Langevin thermostat to measure pressure and temperature, respectively.

### 1.3.5.2 Solvation in MD simulation

Protein-protein and protein-ligand interactions do not occur in a vacuum in nature, taking place in solution with water. The influence that water has on these systems is especially important to consider when performing MD simulations. Many different interactions are directly influenced by water, such as hydrogen bonding, hydrophobic interactions, and salt bridges. Occasionally water also participates in these interactions through hydrogen bonds or coordination to metal ions. In addition, the physical characteristics of a water, such as viscosity, also has an impact on the dynamic motion and diffusion of system. Depending on application, MD takes two different approaches to simulating water implicit and explicit solvation. Implicit solvation models the solvent as a continuum, introducing a general interaction term representing the solvent. Explicit solvation simulates all molecules within a solvent. This solvation model uses different ‘sites’ to model water. TIP3P is an example of a 3-site model which includes two hydrogen atoms and one oxygen atom (72). The 3-site model is a simple approximation for an explicit solvent, though more complex models exist. TIP4P, TIP5P and TIP6P are examples which include sites that model the oxygen lone pairs and a site that improves the polarity of the water molecule (73).

### 1.3.5.3 Periodic Boundary Conditions

A molecule in solution, such as a protein, will often be surrounded by an enormous number of water molecules. The computational demand of simulating so much solvent explicitly would be too high to be considered practical. MD simulations instead simulate molecules within a small, bounded box with a minimal number of solvent molecules. Outside of the boundary of the box there will be a vacuum as no matter is simulated. Though this method reduces computational demand problems with accuracy arise when simulated molecules cross the boundary and into the vacuum. To overcome this, periodic boundary conditions (PBC) are used (Figure 1.45).

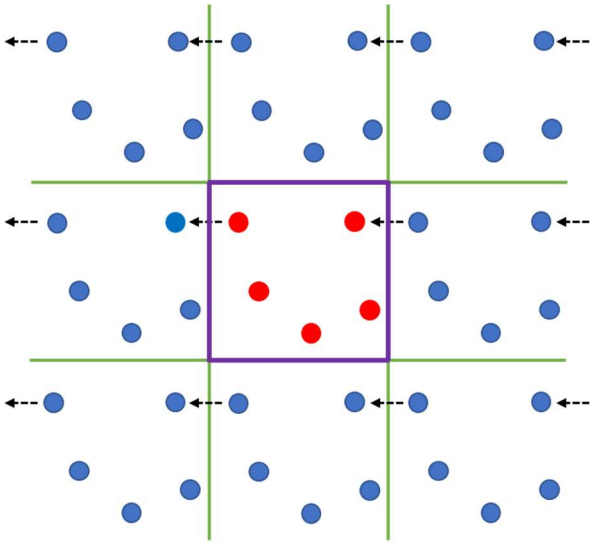


*Figure 1.45 - Illustration of periodic boundary conditions. The bounded box (purple box) surrounded by image cells. As particles cross the borders of the bounded box, they are replaced by particles from a unit cell.*

In this model the bounded box that contains the molecule and solvent is repeated around the simulation. These repeats are called images. Within each image, the simulation proceeds in the same way as in the simulation box with each atom moving in the manner at the same time. The atoms in the images are used to apply forces to the molecules in the simulation box creating the simulation of a continuous system. In the case that a molecule does cross the boundary of the simulation box, the same atoms from the opposite image will take their place and be considered part of the simulation box.

### 1.3.5.4 Enhanced Sampling by GaMD

Though there have been great strides in the development of MD software there is still a practical limitation on how long simulations can be. Much has been discussed on the trade-offs that improved accuracy has on computational demand, but even simple MD simulations can still require a lot of computing power. This often means that simulations of protein-protein and protein-ligand interactions are only performed over a short timescale. The issue with this is that MD simulations can only sample a small amount of a system's PES and can be sensitive to the chosen starting structure. To attempt to overcome this while keeping the computational demand low, so-called enhanced sampling methods have been developed to explore more of the PES. One such method is Gaussian-accelerated molecular dynamics (GaMD) (74).

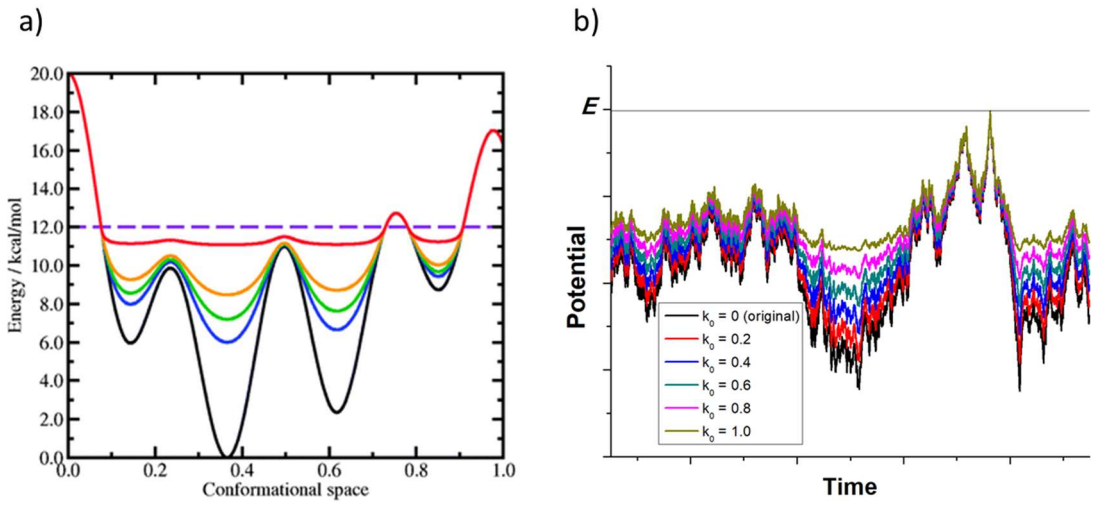



*Figure 1.46 - The impact of accelerated molecular dynamics on the PES of a molecule. (a) aMD is used which has the effect of smoothing the PES. The dashed purple line in a) is the energy threshold with the boost potential being increased from the black to the red line. Taken from (75). GaMD is used to smooth the surface (b). Using GaMD the potential energy surface is still flattened but the shape is maintained. Taken from (76).*

One challenge for sampling methods is overcoming energy barriers in the PES. Accelerated molecular dynamics (aMD) attempts to overcome this issue by increasing the likelihood that energy barriers in the PES are overcome during simulation. This is achieved by applying a boost potential to the system if the overall potential energy falls below a

defined energy threshold (77). To control the boost potential aMD uses an acceleration factor which acts to reduce energy barriers by flattening the PES (Figure 1.46) (75, 77).

With lower energy barriers the PES can be sampled much more efficiently with shorter simulation times. Any given observation from an aMD simulation, such as the potential energy, can be re-weighted to represent its real state more accurately. The problem with this, however, is that there can be a large amount of statistical noise leading to inaccurate results. This problem is exacerbated in more complex systems containing larger molecules such as proteins. To overcome this, GaMD builds on the principle of aMD by using a Gaussian harmonic boost potential (74, 76). Using this approach, the PES is smoothed rather than flattened (Figure 1.46), preserving the original shape of the energy landscape whilst reducing energy barriers. This approach leads to more accurate structures by reducing the noise during re-weighting of the PES (74, 76).

# 1.4 Objectives of This Thesis

This thesis presents the use of NMR spectroscopy and MD simulation to study the behaviour of the NleB and SseK effectors. These experimental approaches were used to achieve the following objectives:

***1 - Provide insights on the molecular basis of the fine tuning for arginine glycosylation of bacterial virulence effectors***

1.1 Identify, through the least number of mutations, a mutant of SseK1 that triggers FADD activity using the NleB1:FADD-DD complex as a template.

1.2 Understand the structural rationale for the impact this mutation has with respect to recognition of FADD.

1.3 Validate the 3D molecular models resulting from MD simulation using NMR spectroscopy.

***2 - Provide insights on the molecular basis for the donor substrate recognition of bacterial virulence effectors***

2.1 Characterise the interaction between UDP, UDP-GlcNAc and other donors and the NleB and SseK effectors.

2.2 Produce a structural model for these interactions and rationalise the differences in measured binding epitopes.

# CHAPTER 2

## MATERIALS AND METHODS

# 2.1 Chemicals and Ligands

## 2.1.1 Raw Chemicals, Ligands, and Their Sources

Deuterium oxide (99% $^{2}H$), tris-(hydroxymethyl-$d_3$ ) amino-$d_2$–methane (Tris-$d_{11}$ , 98% $^{2}H$), potassium phosphate monobasic ($KH_2PO_4$), potassium phosphate dibasic trihydrate ($K_2HPO_4 \cdot 3H_2O$), Bis-Tris, Tris HCl (Trizma hydrochloride), Ammonium-$^{15}N$ chloride (98% $^{15}N$), calcium chloride ($CaCl_2$), magnesium sulphate ($MgSO_4$), glucose ($C_6H_{12}O_6$), ampicillin sodium salt, uridine diphosphate disodium (UDP · 2Na), uridine diphosphoglucose disodium (UDP-Glc · 2Na), uridine diphospho-N-acetylglucosamine disodium (UDP-GlcNAc · 2Na), uridine diphosphogalactose disodium (UDP-Gal · 2Na), uridine diphospho-N-acetylgalactosamine disodium (UDP-GalNAc · 2Na) were all purchased from Sigma Aldrich (Merck). Isopropyl-β-D-thiogalactopyranoside (IPTG) and 1,4-Dithiothreitol (DTT) were both purchased from Formedium. Sodium Chloride (NaCl) and Gibco 100x minimal expression media (MEM) vitamin solution was purchased from Thermo Fisher.

## 2.1.2 Prepared Solutions and Their Compositions

*Table 2.1 - Prepared solutions used in protein expression (section 2.2).*

| Prepared Solution | Composition |
|---|---|
| Wash Buffer | 25 mM TRIS, 150 mM NaCl, 10 mM Imidazole, pH 7.5 |
| Lysis Buffer | 25 mM TRIS, 150 mM NaCl, 50 mM Imidazole, pH 7.5 |
| Elution Buffer | 25 mM TRIS, 150 mM NaCl, 300 mM Imidazole, pH 7.5 |
| Minimal Expression Media (78) | Per litre of media: 1 g $^{15}NH_4Cl$, 6 g $Na_2HPO_4$, 3 g $KH_2PO_4$, 0.5 g NaCl, 1 mL 0.1 M $CaCl_2$, 1 mL 1 M $MgSO_4$, 10 mL 40% glucose, 10 mL 100x MEM vitamin solution. |

# 2.2 Expression of SseK1, SseK1-S286Y, SseK2 and $^{15}$N-FADD-DD

## 2.2.1 SDS-PAGE

SDS-PAGE was performed using pre-cast 4-12% BIS-TRIS 15-well mini gels (Invitrogen) with a running buffer prepared from Bolt 20x MES SDS (Invitrogen). 5 µL of PageRuler Prestained Protein Ladder (Thermo Fisher) was used as a reference. Protein samples were prepared in NuPAGE 4x LDS Sample buffer (Invitrogen) and NuPAGE 10x reducing agent (Invitrogen), diluted to an appropriate volume using deionised water then heated at 70 °C for 10 minutes. 15 µL of the prepared protein mixture was loaded into each well. The gels were run at 200 V for 35 minutes or until the lowest marker on the ladder neared the bottom of the gel. Quick Coomassie (Generon) was used to stain the bands with shaking until bands developed. The gel was then counter stained using de-ionised water with shaking overnight.

## 2.2.2 Bacterial Growth and Cell Harvesting

For SseK1-WT, SseK1-S286Y and SseK2, transformed BL21-DE3 cells (Dr Hurtado-Guerrero lab, Zaragoza, Spain) were plated onto LB-Agar containing 100 µg ml$^{-1}$ ampicillin and grown overnight at 37 °C. A single colony was used to produce a starter culture of 100 mL LB containing 100 µg ml$^{-1}$ ampicillin. The starter culture grew overnight at 37 °C with shaking at 250 rpm. 5 mL of starter culture was added to 1 L of LB broth containing 100 µg ml$^{-1}$ ampicillin. The culture grew at 37 °C with shaking at 250 rpm until an $OD_{600}$ of 0.6-0.8 was reached. The culture was then induced with 200 µL of 1M IPTG and incubated overnight at 20 °C with shaking at 250 rpm. The pellet was then collected by centrifugation using a Backman-Coulter Avanti J-20 XPI with a JLA-8.1000 rotor at 5000 rpm for 20 minutes at 4 °C.

## 2.2.3 Cell Lysis and Protein Purification

The pellet was resuspended in a buffer containing 25 mM Tris and 150 mM NaCl, pH 7.5, at a ratio of 2 mL of buffer to 1 g of pellet. Cells were lysed using Constant Systems continuous flow cell disruptor at 30,000 psi and then centrifuged using a Beckman-Coulter Avanti J-26 XP with a JA-25.50 rotor at 20,000 rpm for 20 minutes at 4 °C. The lysate was purified manually by loading onto a 5 mL HisTrap HP Ni-NTA column (GE Healthcare). The column was washed with 50 mL of lysis buffer and then 50 mL of wash buffer. The protein was then eluted from the column using the elution buffer. SDS-PAGE, as described by 3.2.1, was then performed on samples of the lysis, wash, and elution steps as well as the raw lysate (Figure 2.1). The Ni-NTA purified protein was then loaded on to an ÄKTA for further purification by size exclusion chromatography using a HiLoad 16/600 Superdex 200pg column at a flowrate of 1 mL $min^{-1}$ with a running buffer containing 25 mM TRIS, 150 mM NaCl, pH 7.5 and monitored by UV output measured at 280 nm. 1.5 mL fractions were collected from the column until the purified protein had eluted. SDS PAGE, as described by 3.2.1, was performed on fractions from the major peak from the UV output as well as any satellite peaks (Figure 2.2). Fractions containing the purified protein were collected and concentrated by centrifugation with a Multifuge X3R swing bucket centrifuge (Thermo Fisher) at 4000 rpm using a Vivaspin 20 concentration unit (Sartorius) with a 10 kDa cut-off. The samples were then exchanged into a 100% $D_2O$ buffer containing 25 mM Tris-$d_{11}$ and 150 mM NaCl, pH 7.5 to a concentration of 500 µM for storage at -80 °C.

## 2.2.4 Expression of Uniformly Labelled $^{15}$N FADD-DD

Uniformly $^{15}$N-labelled FADD-DD was expressed using transformed BL21-DE3 cells (Dr Hurtado-Guerrero lab, Zaragoza, Spain) with the same method outlined in 2.2.2 except minimal media, as described in 2.1.2 (78), was used in place of LB broth after a starter culture was grown. Uniformly $^{15}$N-labelled FADD-DD was concentrated by centrifugation with a multifuge X3R swing bucket centrifuge (Thermo Fisher) at 4000 rpm using an Amicon Ultra 15 concentration unit (Merck) with a 3 kDa cutoff. The sample was then exchanged to a concentration of 1.5 mM into a 90% $H_2O$/10% $D_2O$ buffer containing 25 mM TRIS and 150 mM NaCl, pH 7.5 for storage at -80 °C.

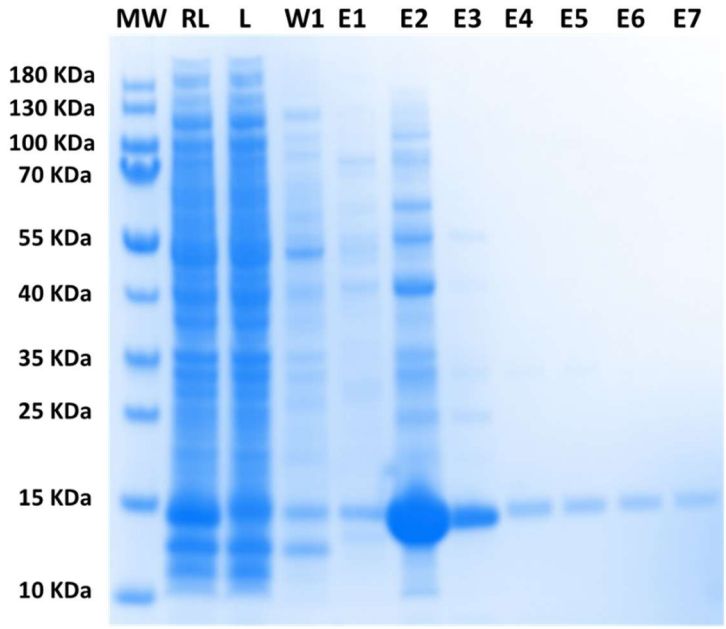



*Figure 2.1 - Representative SDS-PAGE gel of purification by His-Trap Column. Purification of $^{15}$N-FADD-DD as an example. The gel shows the soluble Raw Lysate (RL), Lysis (L), Wash (W) and seven Elution fractions (E1-E7). $^{15}$N-FADD-DD can be seen at approximately 14 kDa.*

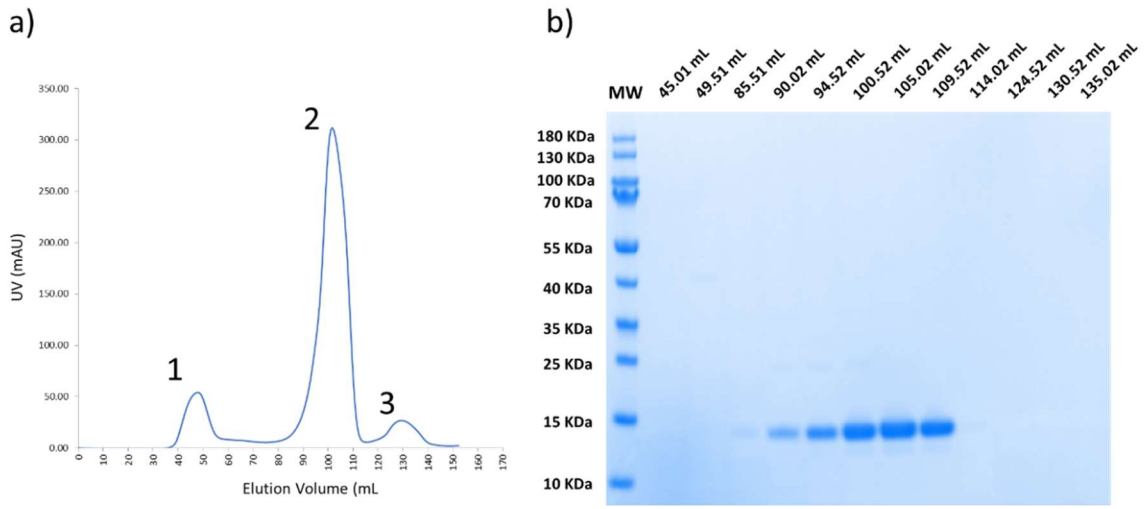



*Figure 2.2 - Representative SDS-PAGE gel of purification by AKTA. The UV trace (a) shows a major peak and two satellite peaks and (b) SDS-PAGE gel of purification by Size-Exclusion using the expression of $^{15}$N-FADD-DD as an example. The gel shows the soluble sampling of the three peaks that appear on the UV output. Peak 1 has been sampled at 45.01 mL and 49.51 mL showing no discernible protein. The largest peak, peak 2, has been sampled from 85.51 mL to 109.52mL where $^{15}$N-FADD-DD can be seen to appear at approximately 14 kDa. Peak 3 has been sampled at 124.52 mL to 135.02 mL showing no discernible protein.*

# 2.3 2D NMR to Study the Binding of FADD-DD with SseK1-S286Y

## 2.3.1 HSQC Experiments

All experiments were performed using a Bruker AVANCE-III 800 MHz spectrometer with a 5 mm TXI triple resonance probe with z axis pulse field gradients (5 mm PATXI $^{1}H$-$^{13}C$/$^{15}N$/D Z-GRD). Apo experiments were performed using a $^{15}N$-FADD-DD concentration of 910 µM in a 90% $H_20$/10% $D_2O$ buffer with 25 mM, and 150 mM NaCl, pH 7.5. Apo spectra were recorded with and without 10 mM UDP and 10 mM $MgCl_2$. A titration with three concentration points was performed with a fixed $^{15}N$-FADD-DD concentration of 200 µM with SseK1-S286Y (Dr Hurtado-Guerrero lab, Zaragoza, Spain) concentrations of 400 µM, 300 µM and 200 µM. All experiments were performed at a temperature of 303 K. The HSQC (*fhsqcf3gpph*) experiments were performed with 32 dummy scans (ds), 32 scans (ns) and a coupling constant (cnst4) of 90 Hz. In the F2 and F1 frequency axes, 2048 and 64 data points were collected respectively. The spectral width was 11.7 ppm to -2.3 ppm for the $^{1}H$ (F2) axes and 132 ppm to 102 ppm for the $^{15}N$ (F1) axes.

## 2.3.2 TROSY Experiments

All experiments were performed using a Bruker AVANCE-III 800 MHz spectrometer with a 5 mm TXI triple resonance probe with z axis pulse field gradients (5 mm PATXI $^{1}H$-$^{13}C$/$^{15}N$/D Z-GRD). All experiments were performed at 298 K. Apo experiments were performed in different buffer conditions as per table 2.2, including phosphate buffered saline, TRIS, and BIS-TRIS with and without 10 mM UDP and 10 mM $MgCl_2$. A titration with four concentration points was performed with a fixed $^{15}N$-FADD-DD concentration of 200 µM with SseK1-S286Y concentrations of 200 µM, 150 µM, 100 µM and 50 µM using a BIS-TRIS buffer as per table 2.2. The TROSY (*trosyf3gpph19*) (56, 79) experiments were performed with 32 dummy scans (ds), 64 scans (ns) and a coupling constant (cnst4) of 90 Hz. In the $^{1}H$ (F2) and $^{15}N$ (F1) frequency axes, 1382 and 64 data points were collected respectively. The

spectral width was 12.7 ppm to -3.3 ppm for the $^{1}H$ (F2) axes and 135 ppm to 100 ppm for the $^{15}N$ (F1) axes. To suppress water signal, the watergate pulse sequence was used.

*Table 2.2 - Buffer compositions used in TROSY titration experiments.*

| Buffer | Composition |
|---|---|
| Phosphate | 50 mM potassium phosphate, 150 mM NaCl, 1 mM DTT, 10 mM UDP, 10 mM $MgCl_2$ pH 6.2 |
| TRIS | 25 mM TRIS, 150 mM NaCl, 1 mM DTT, 10 mM UDP, 10 mM $MgCl_2$ pH 6.2 |
| BIS-TRIS | 25 mM BIS-TRIS, 150 mM NaCl, 1 mM DTT, 10 mM UDP, 10 mM $MgCl_2$ pH 6.2 |

# 2.4 MD simulations of FADD-DD in complex with NleB1, SseK1 and mutants of SseK1

## 2.4.1 Generation of MD Starting Structures (44)

The starting complexes were prepared using Schrödinger Maestro. Cartesian co-ordinates for the NleB1 EPEC-UDP-Mn(II)-FADD-DD complex were taken from the protein data bank (PDB entry 6ACI) (35). To generate the NleB1 EPEC-UDP-GlcNAc-Mn(II)-FADD-DD complex, UDP-GlcNAc was added manually but based on the UDP coordinates of the above complex. The system was then prepared using the protein preparation wizard. Hydrogen atoms were added, and any ligands/non-structurally important waters were removed. The ionisation state of side chains was predicted with PROPKA (80). The resulting structure was minimised with OPLS3 (64) using a threshold of 0.3 Å to converge heavy atoms. From the crystal structure of SseK1-UDP-Mn(II) (PDB entry 5H60) (36), all three single mutants were built (SseK1-S286Y, SseK1-S286N and SseK1-S286I). Missing lid domain residues were added using the build tool. The systems were then prepared using the protein preparation wizard. Hydrogen atoms were all added, and any ligands and non-structurally important waters were removed. The ionization state of side chains was predicted with PROPKA. The

resulting structures were minimised with OPLS3 using a threshold of 0.3 Å to converge heavy atoms. For each mutant complex, UDP was removed and UDP-GlcNAc was docked into the donor site. UDP-GlcNAc first prepared with LigPrep using the OPLS3 force field and default parameters to ensure correct atom types compatible with Glide. Low energy conformers were generated with MacroModel, using the OPLS3 force field with water as the implicit solvent. Structures were minimised using the Polak-Ribier Conjugate Gradient method with converging structures by gradient with a threshold of 0.05 Kcal $mol^{-1}$ with a maximum of 2500 iterations. Conformers were retained if they had an energy < 5 Kcal $mol^{-1}$ from the global energy minimum and a maximum atom deviation of > 0.5 Å from other conformers. Conformers were then docked into a grid with a 23 Å x 23 Å x 23 Å outer box and 10 Å x 10 Å x 10 Å inner box. The centroid of the box was placed in the middle of the UDP-GlcNAc binding site. UDP-GlcNAc was then docked using Glide (81) with standard precision and a 2-times sampling without ring canonicalization and sampling ring conformations. Docking calculations were performed with enhanced sampling and a distance-dependent dielectric constant of 2. Poses were scored using GlideScore. One structure was selected from the docking and used to build all the SseK1 complexes. The FADD-DD acceptor complexes were constructed by structural alignment of the enzyme with NleB1 EPEC in the crystal structure of the NleB1 EPEC-UDP-Mn(II)-FADD-DD complex (PDB entry 6ACI) (35) followed by energy optimization with OPLS3.

## 2.4.2 MD Simulation: Parameterisation

The partial charges for UDP and UDP-GlcNAc were derived with the AM1-BCC charge model with a net charge of -3 and -2, respectively, using antechamber. Molecular dynamics simulations of the NleB1 EPEC-UDP-GlcNAc/UDP-Mn(II)-FADD-DD, SseK1-WT-UDP-GlcNAc/UDP-Mn(II)-FADD-DD, and SseK1-S286Y/N/I-UDP-GlcNAc/UDP-Mn(II)-FADD-DD complexes were performed using Amber 16 PMEMD (82). The Amber ff14SB force field was used to parameterise all protein atoms with the Mn(II) ion modelled using a 12-6-4 LJ-type parameters derived from Amber ions234lm_1264_tip3p. GLYCAM 06j (63) and GAFF (62) was used to parameterise UDP-GlcNAc. GAFF alone was used to parameterise UDP. All the systems were solvated using a truncated octahedral box using TIP3P water with a buffer set to 10 Å. The systems were neutralised using $Na^+$ ions.

### 2.4.3 MD Simulation: Equilibration

The conjugate gradient algorithm was used to minimise the systems with convergence threshold set to $10^{-4}$ kcal $mol^{-1}$ $Å^{-1}$, first using a 20 kcal $mol^{-1}$ $Å^{-2}$ restrain on solute atoms after which minimisation was repeated without any restraints. Each system was then heated to a temperature of 310 K over 500 ps prior to pressure equilibration to 1 atm over the course of another 500 ps. During both heating and pressure equilibration restraints of 20 kcal $mol^{-1}$ $Å^{-2}$ were placed on solute atoms. Restraints were then released from the system in four stages, each of 200 ps length.

### 2.4.4 MD Simulation: Production

Gaussian accelerated Molecular Dynamics (GaMD) was used to simulate each system for 1 µs. Amber 16 was allowed to calculate the boost potential with a 2 ns MD simulation. This boost potential was then used to simulate the first 400 ps of the simulation prior to allowing for adaptation for 5.6 ns. The final boost potential was then used to perform the molecular dynamics simulation for 1 µs. The SHAKE (69) algorithm was used in all simulations to restrain bonds to hydrogen atoms, with the time step being set to 2 fs. To maintain temperature a Langevin thermostat was implemented using a collision frequency of 5 $ps^{-1}$. To maintain pressure an isotropic Berendsen algorithm was used as a barostat with relaxation time set to 1 ps. Periodic boundary conditions were used with particle mesh Ewald for calculating electrostatics.

## 2.5 STD NMR Experiments for the Binding of Sugar Nucleotides to SseK effectors

### 2.5.1 Ligand Sample Preparation and Assignment

UDP, UDP-Gal, UDP-GalNAc, UDP-Glc and UDP-GlcNAc had their labile hydrogens exchanged with deuterium by dissolving in $D_2O$ and freeze-drying, repeated three times for each ligand. All experiments were performed using a Bruker AVANCE-III 800 MHz

spectrometer with a 5 mm TXI triple resonance probe with z axis pulse field gradients (5 mm PATXI $^{1}$H-$^{13}$C/$^{15}$N/D Z-GRD) at a temperature of 288 K. $^{1}$H and $^{13}$C signals were assigned using $^{1}$H-$^{13}$C HSQC (*hsqcedetgp*) (83), $^{1}$H-$^{1}$H TOCSY (*mlevphpp*) (84) and $^{1}$H-$^{1}$H COSY (*cosygpmfqf*) (85, 86).

## 2.5.2 STD NMR Experiments

All experiments were performed using a Bruker AVANCE-III 800 MHz spectrometer with a 5 mm TXI triple resonance probe with z axis pulse field gradients (5 mm PATXI $^{1}$H-$^{13}$C/$^{15}$N/D Z-GRD). STD NMR experiments were performed at a ligand concentration of 1 mM and protein concentration of 25 µM in a 100% $D_2O$ buffer containing 25 mM Tris-$d_{11,}$ 250 µM $MgCl_2$ and 150 mM NaCl, pH 7.5. Experiments for SseK1 were also performed in the absence of $MgCl_2$. A train of 50 ms Gaussian pulses was used at 0.1 mW applied on the F2 channel at 0 ppm (on resonance) and 40 ppm (off resonance) to obtain the STD spectra. To remove any unwanted xy magnetization from the previous scan, a spoil sequence of two trim pulses of 2.5 and 5 ms with a 40% z-gradient applied for 3 ms at the start of the experiment was used. To suppress protein signals, a spinlock of 1.55 W and 40 ms was used (*stddiff.3*) (50). Build-up curve experiments were performed at a temperature of 288 K with a recycle delay (D1) of 5 s, saturation times (D20) of 0.5 s, 1 s and 1.5 s were performed using 512 scans (ns) and D20 of 2 s, 3 s, 4 s and 5s were performed using 256 scans (ns). For the variable temperature STD NMR experiment a spectrum was acquired at 298 K with a D20 of 2 s and 256 scans (ns). STD intensities were measured and then fitted to equation 14 using QTIplot. $STD_0$ values were calculated for each signal and normalised against the largest value. These values were then used to plot binding epitopes.

# 2.6 MD Simulations of Donors in Complex with NleB1, SseK1, SseK1-S286Y SseK2

## 2.6.1 Generation of MD Starting Structures

The starting complexes were prepared using Schrödinger Maestro. Cartesian co-ordinates for NleB1 EPEC (PDB entry 6ACI), SseK2 (PDB entry 5H63) were taken from the protein data bank. The enzymes were then prepared using the protein preparation wizard. Hydrogen atoms were all added, and any ligands and non-structurally important waters were removed. The ionisation state of side chains was predicted with PROPKA (80). The resulting structure was minimised with OPLS3 (64) using a threshold of 0.3 Å to converge heavy atoms. The Cartesian coordinates for SseK1-WT were taken from the protein data bank (PDB entry 5H60) (36), from which SseK1-S286Y was built. Missing lid domain residues were added using the build tool. The system was then prepared using protein preparation wizard. Hydrogen atoms were added, and any ligands/non-structurally important waters were removed. The ionisation state of side chains was predicted with PROPKA. The resulting structure was minimised with OPLS3 using a threshold of 0.3 Å to converge heavy atoms. For each enzyme UDP, UDP-GlcNAc, UDP-Glc, UDP-GalNAc and UDP-Gal were docked to the active site. All ligands were first prepared with LigPrep using the OPLS3 force field and default parameters to ensure correct atom types compatible with Glide. Low energy conformers were generated with MacroModel, using the OPLS3 force field with water as the solvent. Structures were minimised using the Polak-Ribier Conjugate Gradient method with converging structures by gradient with a threshold of 0.05 Kcal $mol^{-1}$ with a maximum of 2500 iterations. Conformers were retained if they had an energy < 5 Kcal $mol^{-1}$ from the global energy minimum and a maximum atom deviation of > 0.5 Å from other conformers. Conformers were then docked into a grid with a 23 Å x 23 Å x 23 Å outer box and 10 Å x 10 Å x 10 Å inner box. The centroid of the box was placed in the middle of the UDP-GlcNAc binding site. UDP-GlcNAc was then docked using Glide (81) with standard precision and a 2-times sampling without ring canonicalization and sampling ring conformations. Docking calculations were performed with enhanced sampling and a distance-dependent dielectric constant of 2. Poses were scored using GlideScore. For each

ligand-protein complex, one result was chosen based on proper coordination of the pyrophosphate group in the active site.

## 2.6.2 MD Simulation: Parameterisation

The partial charges for UDP, UDP-GlcNAc, UDP-Glc, UDP-GalNAc and UDP-Gal were derived with the AM1-BCC charge model with a charge of -3 for UDP and -2 for the other ligands, using antechamber. Molecular dynamics simulations for all the Enzyme-Mn(II)-Ligand complexes were performed using Amber 22 PMEMD (82). The Amber ff14SB force field was used to parameterise all protein atoms with the Mn(II) ion modelled using a 12-6-4 LJ-type parameters derived from Amber ions234lm_1264_tip3p. GAFF was used to parameterise all ligands. All the systems were solvated using a truncated octahedral box using TIP3P water with a buffer set to 10 Å. The system was neutralised using $Na^+$ ions.

## 2.6.3 MD Simulation: Equilibration

The conjugate gradient algorithm was used to minimised using the steepest descent method for 4000 cycles, then using the conjugate gradient method for 6000 cycles. No restraint weights were used. The complexes were then simulated for 10,000 cycles under no restraints. Each complex was then heated to a temperature of 300 K over 50 ps prior to pressure equilibration to 1 bar over the course of another 50 ps. The complexes were then allowed to equilibrate of 300 ns under no restraints.

## 2.5.4 MD Simulation: Production

Complexes were simulated for 500 ns over 5, 100 ns increments using standard MD. The SHAKE (69) algorithm was used in all simulations to restrain bonds to hydrogen atoms, with the time step being set to 2 fs. To maintain temperature a Langevin thermostat was implemented using a collision frequency of 5 $ps^{-1}$. To maintain pressure an isotropic Berendsen algorithm was used as a barostat with relaxation time set to 1 ps. Periodic boundary conditions were used with particle mesh Ewald for calculating electrostatics.

# CHAPTER 3

## MOLECULAR BASIS OF THE FINE TUNING FOR ARGININE GLYCOSYLATION OF BACTERIAL VIRULENCE EFFECTORS

# 3.1 Introduction

Protein-protein interactions often involve many diffuse, low strength, interactions between two protein surfaces. This poses a challenge when understanding how proteins recognise one another as it can be difficult to pinpoint essential interactions. The challenge in understanding the interactions formed by the NleB and SseK effectors is no different, but with the work of the scientific community there has been some light shed on how these enzymes recognise their acceptors (35). For the study of these unique arginine N-GTs, untangling how these GTs can distinguish between different members of the TNF-α-mediated immune response is of particular interest. With this knowledge it would be possible to begin understanding how to prevent these interactions and design pharmaceutics which disrupt their ability to GlcNAcylate their targets.

The crystal structure of NleB1 in complex with FADD-DD (35) provides a great insight into the contacts made between enzyme and acceptor, highlighting regions of interest for both members. Despite this information, the challenge remains to understand the significance of each interaction highlighted by that complex. This chapter presents a mutagenesis study which sought to isolate specific residues and begin to develop a thorough structural understanding of the interaction interface. This was achieved by using the NleB1:FADD-DD crystal structure and its interactions as a basis for mutating the FADD-inactive enzyme SseK1 to trigger FADD activity with the least mutations. With this information it was possible to find a single point mutation that satisfy this, SseK1-S286Y. This mutant highlighted the importance of this single tyrosine residue in FADD recognition. The results of this chapter lay out the use of MD and NMR with the aim of describing the structural rationale for the importance of this residue.

### 3.1.1 Goals

- Identify a set of potential sites of mutation in SseK1 based on the published NleB-FADD crystal structure that make contacts with FADD-DD.

- Use multiple sequence alignment to narrow down which of these sites could be mutated to trigger FADD activity in SseK.

- Identify the least number of mutations that would trigger FADD-activity in SseK1.

- Use GaMD to simulate the mutants in solution and provide a structural rationale of the impact the point mutations have on protein dynamics and recognition.

- Verify the conclusions from computer simulation using two-dimensional NMR experiments to observe interactions made by $^{15}$N-labelled FADD-DD.

# 3.2 Results and Discussion

## 3.2.1 Mutagenesis of SseK1 to Trigger FADD Activity

*The multiple sequence alignment of NleB1 and the SseK effectors was performed in collaboration with the Dr Hurtado-Guerrero Lab, University of Zaragoza.*

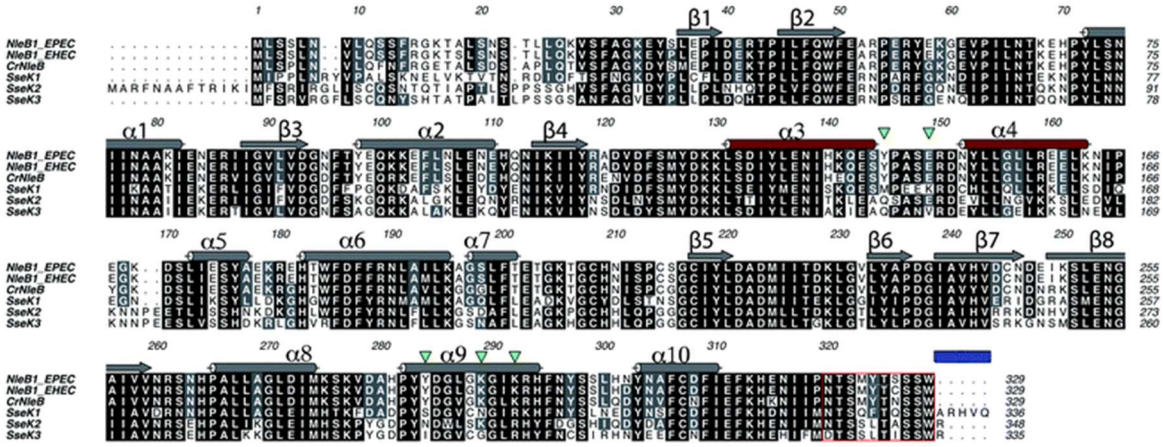



*Figure 3.1 - Multiple sequence alignment of the NleB and SseK effectors. Included are: NleB1_EPEC, NleB1_EHEC, CrNleB, SseK1, SseK2 and SseK3. Sequences are colour-coded with conserved residues coloured black, grey and white denoting similarity and dissimilarity, respectively. Secondary structures are indicated above the NleB1_EPEC sequence showing the catalytic domain (grey) and the HLH domain (brown) based on the structure of NleB1 in complex with FADD-DD and UDP (PDB entry 6ACI). The green arrows indicate positions selected for mutation. Lid domain residues are highlighted by the red box with the extended lid domain sequence of SseK1 highlighted by the blue bar. Taken from (44).*

To Identify potential sites of mutation based on the regions of direct contacts identified in the NleB1:FADD-DD crystal structure, multiple sequence alignment between the NleB and SseK effectors was first performed (Figure 3.1). Two primary areas were focussed on, the α3 and α4 helices of the HLH domain and the catalytic domain-adjacent α9 helix. These features were highlighted as close contacts from the crystal structure and found to have conserved interactions in the FADD active effectors that were not present in SseK1. Within these two areas there are a total of fourteen close contacts made between NleB1 and

FADD-DD. Most of these interactions are salt bridges with five coming from the HLH domain (Tyr145, Glu149, Ala151, Tyr153 and Lys153) and nine present near the catalytic domain of NleB1 (Glu253, Lys277, Ala279, Tyr283, Tyr284, Ala285, Tyr289, Tyr292 and Tyr303) (Figure 3.2). Of these residues five were estimated to be of the most importance based on their alignment with SseK1 and the other effectors, two in the HLH domain and three adjacent to the catalytic site.

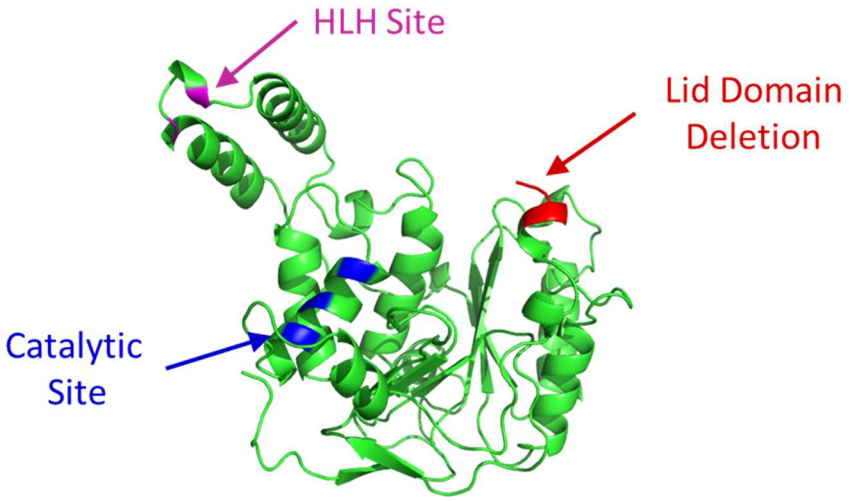



*Figure 3.2 - Structure of SseK1 with extended lid domain (built from PDB entry 5H60). Shown are the sites of mutation including the HLH Site (pink), Catalytic Site (blue) and Lid Domain Deletion (red).*

The two residues from the HLH domain include NleB1$^{Y145}$ and NleB1$^{E149}$, located at the loop connecting the α3 and α4 helices. NleB1$^{Y145}$ forms a pi-cation stack with FADD$^{R136}$ with NleB1$^{E149}$ contributing to this interaction though a backbone interaction. NleB1$^{Y145}$ is conserved in all the NleB effectors but not within the SseK effectors. In the SseK effectors the aligned residues include SseK1$^{M147}$, SseK2$^{Q161}$ and SseK3$^{Q148}$. Though not able to form a pi-cation stacking interaction the glutamine residues present in SseK2 and SseK3 could form some complementary interactions with FADD$^{R136}$ through hydrogen bonding. The methionine present in SseK1 however, cannot form similar interactions as tyrosine. In addition, it cannot form other complementary interactions with FADD$^{R136}$ as it has a hydrophobic sulphide group. Because of this it was rationalised that the SseK1-M147Y mutant would be best to probe the impact of this interaction given the conserved tyrosine

residues are present in the most active NleB1 effectors and the methionine being unable to form similar interactions.

NleB1$^{E149}$ is conserved in all the NleB effectors and the FADD-active SseK effector, SseK2. In the SseK effectors the aligned residues include SseK1$^{K151}$ and SseK3$^{V152}$. Though the glutamic acid of NleB1 is observed to only contribute via a backbone interaction with FADD$^{R136}$, the relative size and different interactions offered by the residues present in SseK1 and SseK3 could have an impact on the efficiency of that backbone interaction. The lysine present in SseK1 is longer and has a positively charged primary amine functional group which differs from the conserved glutamic acid of the FADD-active effectors. This significant difference in residue properties could prevent the backbone being in the proper orientation for an efficient interaction with FADD$^{R136}$. To explore this, the SseK1-K151E mutant was chosen.

Adjacent to the catalytic site of NleB1, the three residues of interest that contact with FADD-DD from the α9 helix are NleB1$^{Y284}$, NleB1$^{K289}$ and NleB1$^{K292}$. NleB1$^{Y284}$ is projected into a hydrophobic pocket formed by FADD$^{A116}$, FADD$^{V121}$ and FADD$^{I126}$. Alongside this hydrophobic interaction the hydroxyl hydrogen of the phenol group forms a hydrogen bond with the backbone of FADD$^{R117}$. In the SseK effectors the aligned residues include SseK1$^{S286}$, SseK2$^{N302}$ and SseK3$^{I289}$. Given the intimate nature of this interaction in the NleB1:FADD-DD complex, mutation at this site to form the SseK1-S286Y mutant was proposed. As this interaction is close to FADD$^{R117}$, the target residue for GlcNAcylation, the residues in the position of both the FADD-active SseK2 and FADD-inactive SseK3 effectors were chosen to further explore the significance of this interaction. This produces the mutants SseK1-S286N and SseK1-S286I.

NleB1$^{K289}$ and NleB1$^{K292}$ are conserved in all the NleB effectors, with only NleB1$^{K289}$ conserved in the FADD-active SseK effector, SseK2. In the SseK effectors the NleB1$^{K289}$ aligned residues include SseK1$^{N291}$ and SseK3$^{G294}$. Arginine is present all the SseK effectors in place of NleB1$^{K292}$. NleB1$^{K289}$ and NleB1$^{K292}$ both form salt bridges with FADD$^{E130}$ and FADD$^{D123}$ respectively, which are both near the hydrophobic pocket NleB1$^{Y284}$ interacts with. While SseK1$^{N291}$ can offer similar interactions as the NleB1$^{K289}$ both the difference in functional group and size of the residue leaves space for replacement with lysine to probe its importance in FADD recognition. For similar reasons, SseK1$^{R294}$ is an interesting residue

to mutate to lysine. These two mutations provide the final pair of single point mutants SseK1-N291K and SseK1-R294K.

In addition to the seven point mutations described, the multiple sequence alignment highlights an additional chain of five amino acids present on the C-terminus of SseK1, ARHVQ. This C-terminus is the lid domain, which has been shown through structural studies to be highly dynamic, with potential roles in both substrate binding and acceptor recognition. Interestingly, in the crystal structure the NleB1:FADD-DD complex no contacts with the lid domain and FADD-DD are seen. Coupled with the fact that the lid domain is highly conserved in across the other NleB and SseK effectors the deviation observed in SseK1 is curious. Given the role and typically conserved nature of this domain it could be that the extended SseK1 lid domain is involved in its selectivity profile. To probe whether this region could be responsible for SseK1 being FADD-inactive deletion of the ARHVQ chain was included in the mutagenesis study.

The mutagenesis study produced eight mutations of SseK1 (Table 3.1). These mutations encompass features of SseK1 that that are analogous to sites of acceptor recognition in NleB1 (Figure 3.2). None of the mutations involve residues proposed to be a direct member of catalytic machinery or substrate recognition, except for the deletion of the ARVHQ C-terminus residues. Therefore, this mutagenesis study and its findings can only reliably inform the impact on specific interactions involved in acceptor recognition. Further mutagenesis studies of these effectors will need to be done to tease out whether differential substrate recognition or catalytic behaviour is responsible for acceptor selectivity.

*Table 3.1 - The selected point mutations for SseK1.*

| Region | SseK1 Mutation |
|---|---|
| α3 and α4 Helices (HLH Site) | M147Y<br>K151E |
| α9 helix (Catalytic Site) | S286Y/S286N/S286I<br>N291K<br>R294K |
| Lid Domain | ARVHQ-del |

## 3.2.2 Finding the FADD-Active SseK1 Mutant

*The following work was conducted by Ana Garcia-Garcia in the Dr Hurtado-Guerrero Lab of the university of Zaragoza. A description of these works has been included as it provides necessary context for later sections.*

A series of experiments were performed on the WT effectors and mutants to measure the impact of mutation on the FADD activity of SseK1 and determine their kinetic properties. To find a FADD-active mutant of SseK1, single and multiple mutants were expressed, with their affinity and GT activity measured. To simplify the number of combinations, the mutants were clustered into groups base on how close they are to other sites of mutation and their location within SseK1. The first two groups were the pairs of residues in the HLH domain and the pair adjacent to the catalytic domain giving the SseK1-M147Y-K151E and SseK1-N291K-R294K mutants, respectively. As the residue is isolated and forms its own interaction separate from the other sites of mutation the SseK1-S286X mutants were not grouped in this initial simplification. These sites were then secondarily grouped to form the SseK1-M147Y-K151E-N291K-R294K quadruple and SseK1-M147Y-K151E-S286X-N291K-R294K quintuple mutants. Finally, the inclusion of the ARVHQ deletion was included in the quintuple mutant to give the quintuple-del mutant.

The Michaelis-Menten ($K_M$) and Turnover Number ($k_{cat}$) kinetic parameters of the NleB1-WT, SseK1-WT SseK2-WT and the SseK1 mutant enzymes were first measured (Table 3.2). This pair of parameters allow for the measurement of effector-substrate affinity and the rate at which the enzymes can perform GlcNAc transfer to FADD, respectively. These parameters were then used to determine the catalytic efficiency of the enzymes, given by $k_{cat}/K_M$. Catalytic efficiency is a useful measure in this case as it can differentiate the preference of the different enzymes from FADD with higher catalytic efficiency indicating that an enzyme is more able to effectively perform GlcNAc transfer. Using the strongly FADD active NleB1-WT and weakly active SseK2-WT as a benchmark catalytic efficiency in these experiments allow the assessment of the impact single and multiple mutations have on SseK1.

*Table 3.2 - The kinetic parameters for the SseK1 constructs, SseK2 and NleB1. Cells marked with **W/A** indicates that there was activity but too weak to measure reliably and cells marked N/A were not FADD-active (44).*

| | $K_M$ (μM) | $R^2$ | $v_{max}$ (μmol $min^{-1}$ $mg^{-1}$) | $k_{cat}$ ($min^{-1}$) | $k_{cat}/K_M$ ($min^{-1}$ μM-1) |
|---|---|---|---|---|---|
| **NleB1** | 13 ± 2.5 | 0.94 | 2.92± 0.20 | 107 ± 7.6 | 8.2 |
| **SseK1** | N/A | N/A | N/A | N/A | N/A |
| **SseK2** | 1226 ± 574 | 0.96 | 0.43 ± 0.13 | 16 ± 4.9 | 0.013 |
| **K151E-M147Y** | N/A | N/A | N/A | N/A | N/A |
| **N291K-R294K** | N/A | N/A | N/A | N/A | N/A |
| **Quadruple** | 79 ± 36.7 | 0.90 | 0.67 ± 0.21 | 25 ± 7.8 | 0.3 |
| **S286Y** | 57 ± 12.0 | 0.95 | 1.86 ± 0.10 | 68 ± 3.9 | 1.2 |
| **Quintuple** | 31 ± 2.8 | 0.99 | 1.23 ± 0.04 | 45 ± 1.6 | 1.45 |
| **Quintuple-Del** | 27 ± 4.7 | 0.97 | 1.35 ± 0.14 | 49 ± 5.2 | 1.8 |
| **S286N** | **W/A** | **W/A** | **W/A** | **W/A** | **W/A** |
| **S286I** | N/A | N/A | N/A | N/A | N/A |

The $K_M$ for NleB1-WT with respect to FADD was measured to be 13 μM ±2.5 and the $K_M$ for SseK2-WT was measured to be 1226 μM ±574. The $K_M$ for SseK1-WT and SseK1-M147Y-K151E/ SseK1-N291K-R294K mutants with respect to FADD was not measurable. The only mutants of SseK1 with a comparable $K_M$ to NleB1-WT were those that contained the S286Y mutation (SseK1-S286Y, quintuple and quintuple-del mutants) and the quadruple mutant. The lowest $K_M$ recorded was for the quintuple-del mutant at 27 μM ±4.7. Interestingly however, the single mutant SseK1-S286Y managed to achieve a near NleB1 $K_M$ of 57 μM ±12.0 even without the contribution of the other mutations of SseK1. This result is quite impressive as it shows that this single point mutation is enough to meaningfully change the selectivity of SseK1.

The $k_{cat}$ of NleB1-WT with respect to GlcNAcylation of FADD was measured as 107 $min^{-1}$ ± 7.6, for SseK2-WT this was measured to be 16 $min^{-1}$ ±4.9. It was not possible to

measure the $k_{cat}$ of SseK1 and SseK1-M147Y-K151E/SseK1-N291K-R294K mutants with respect to GlcNAcylation of FADD. Like $K_M$ the only mutants of SseK1 with comparable $k_{cat}$ to NleB1-WT were those that contained the S286Y mutation (SseK1-S286Y, quintuple and quintuple-del mutants) and the quadruple mutant. The SseK1-S286Y mutant showed the best turnover rate achieving a near NleB1 $k_{cat}$ of 68 $min^{-1}$ ± 3.9. Once again, the single point mutation SseK1-S286Y stands out as significantly altering the catalytic behaviour of SseK1. Given the impact on FADD activity mutation has at SseK1 position 286 there was promise for interesting observations for the SseK1-S286N and SseK1-S286I mutants from alignment with SseK2 and SseK3, respectively. Unfortunately, neither $K_M$ of $k_{cat}$ was measurable for SseK1-S286N and SseK1-S286I was inactive.

The Catalytic Efficiency of NleB1-WT with respect to GlcNAcylation of FADD was calculated to be 8.2 $min^{-1}$ $\mu M^{-1}$, for SseK2-WT this was calculated to be 0.013 $min^{-1}$ $\mu M^{-1}$. The single point mutant SseK1-S286Y was calculated to have a catalytic efficiency of 1.2 $min^{-1}$ $\mu M^{-1}$, a value like the quintuple and quintuple-del mutants achieved catalytic efficiencies of 1.45 $min^{-1}$ $\mu M^{-1}$ and 1.8 $min^{-1}$ $\mu M^{-1}$, respectively.

The standout result from this analysis was of course the dramatic effect the single SseK1-S286Y mutant has on the activity of SseK1 with respect to FADD. Curiously however, this single mutation appears to support the interactions of the other sites of mutation and thus improve recognition. Without the S286Y mutation both the double mutants (SseK1-M147Y-K151E/ SseK1-N291K-R294K) could not have their $K_M$ values measured with respect to FADD. Additionally, when combined into the quadruple mutant these mutations achieve a $K_M$ of 79 µM ±36.7, which is significantly higher than the $K_M$ of the single mutant (57 µM ±12.0). When combined to form the quintuple mutant however, the $K_M$ is improved reaching a value of 31 µM ±2.8. This result shows an additive effect these mutations have with respect to FADD. Furthermore, this trend continues when the deletion is included in the quintuple-del mutant reaching the best $K_M$ of any of the other mutations at 27 µM ±4.7.

This additive trend however is not extended to the $k_{cat}$ of the mutants. Again, the double mutants could not have their $k_{cat}$ measured, and the quadruple mutant could, achieving a $k_{cat}$ of 79 $min^{-1}$ ± 36.7. This value was worse than the single mutant SseK1-S286Y which achieved a $k_{cat}$ of 68 $min^{-1}$ ± 3.9. Unlike $K_M$ however, when the mutations were combined into a quintuple mutant the $k_{cat}$ worsened significantly to a value of 31 $min^{-1}$ ±

2.8. The $k_{cat}$ was further worsened for the quintuple-del mutant only achieving a $k_{cat}$ of 27 $min^{-1}$ ± 4.7. These results show that curiously the mutations had a negative effect on $k_{cat}$ when combined with the S286Y mutation.

The consequence of these two observations is that this mutagenesis study provides two separate methods that mutation can improve the catalytic efficiency of SseK1 with respect to FADD. The first method is to alter the apparent binding surface of SseK1 multiple mutation. This method leads to a reduction the $K_M$ by improving the affinity of SseK1 by selectively introducing complementary interactions. Secondly it is possible to improve turnover rate with the single S286Y mutation. This suggests that this point mutation is important in providing an interaction platform for GlcNAcylation. Being a single point mutation that was able to trigger FADD activity in SseK1, this mutant was selected as a candidate for additional study.

### 3.2.3 Rationalising the Impact of the S286Y Mutation by MD

To provide a structural explanation of the impact of the single S286Y mutation has on SseK1 and its ability to interact with FADD, a series of MD simulations were performed. The method chosen was GaMD as it allows for efficient sampling and fast simulation over the timescale chosen for these simulations. In line with the crystal structure a truncated FADD construct (FADD-DD), containing the death domain, was used. To explore the importance of this position with respect to FADD recognition, 5 systems were simulated both UDP-bound and UDP-GlcNAc-bound. The systems were complexes of FADD-DD with NleB1-WT, SseK1-WT, SseK1-S286Y, SseK1-S286N and SseK1-S286I. These systems were chosen to provide a thorough overview of the impact of the S286Y mutation encompassing both active and inactive enzymes. The goal being to provide a comprehensive picture of how these complexes behave in solution and evolve over time to narrow down the exact impact of mutation with respect to FADD binding.

#### 3.2.3.1 Monitoring the Enzyme:FADD-DD Complexes

To gain an initial insight into the behaviour of the simulated systems, root-mean-square deviation (RMSD) and root-mean-square fluctuation (RMSF) calculations were performed. RMSD calculations were performed on the backbone atoms of both the enzyme and FADD-DD in the complexes, compared to the first frame. By doing this, differences in conformation can be monitored along the length of the simulation. These calculations are used here highlight motion or, dynamic events, during the simulations which can be observed more closely. RMSD calculations were also used to guide at which time the simulations equilibrated. RMSF calculations are like RMSD calculations but consider the amount of motion individual residues undergo within a particular timeframe. This measure is useful as while RMSD gives an indication of the overall motion of a system relative to its starting point, RMSF will show which individual residues and structural features are dynamic.

### 3.2.3.1.1 System Equilibration

The RMSD calculations show that, NleB1 aside, the systems all took some time to reach equilibrium. Comparing the RMSD calculations of UDP-GlcNAc-bound systems (Figure 3.3) the complexes including SseK1-S286Y and SseK1-S286N took the longest time to stabilise, reaching stability at 0.6 μs. UDP-bound complexes reached equilibrium sooner, with the complex including SseK1-WT taking the longest time to stabilise at 0.4 μs. For consistency between the simulations, it was chosen that analyses of all simulations were performed from 0.6 μs to 1 μs.

A comparison of the UDP-GlcNAc-bound systems shows there are some relative differences in equilibration. The NleB1 complex is much more stable over the whole simulation compared to the SseK1-WT and SseK1 mutant complexes. This is a product of the method used to build the systems. As the NleB1:FADD-DD complex comes from the crystal structure, the complementary interactions at the protein interface will be more optimal than the other systems. This is because while FADD-DD is positioned in a similar position on the acceptor interface of SseK1-WT and SseK1 mutants, many of the interactions may not be optimal. As a result of this the systems are likely to have to undergo some re-arrangement relative to their initial starting state over the length of the simulation, which is reflected in the overall higher RMSD values for these systems. A similar pattern can be seen for the RMSD calculations of the UDP-bound systems. The reason for this being the same as for the UDP-GlcNAc-bound systems.

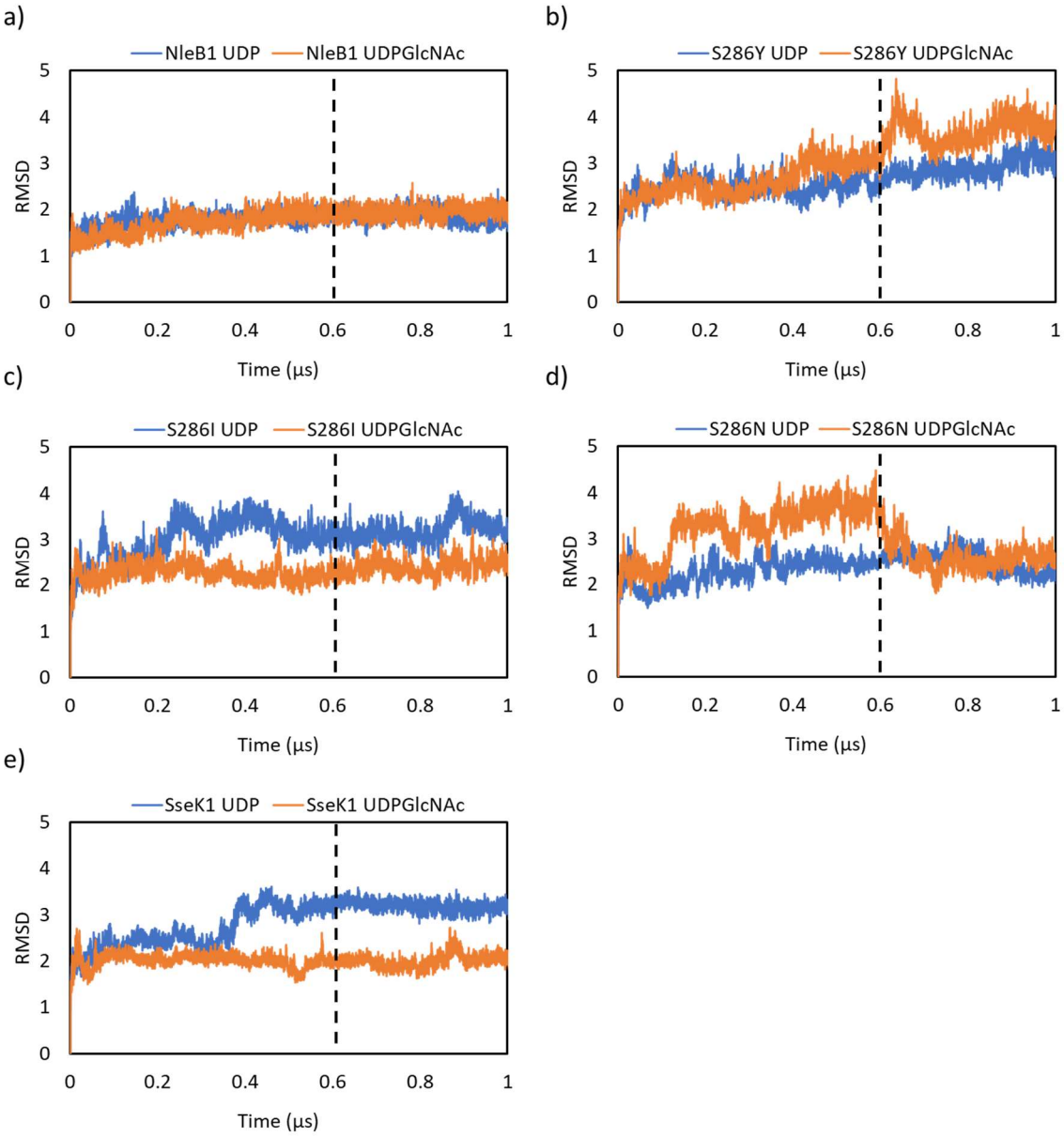



*Figure 3.3 – RMSD (Å) of the backbone atoms along the 1 µs simulations. Shown are the UDP and UDP-GlcNAc-bound complexes of FADD-DD with NleB1 (a), SseK1-S286Y (b), SseK1-S286I (c), SseK1-S286N (d) and SseK1-WT (e). All complexes are overall stable along the entire simulation time, with all members remaining associated. The SseK1 complexes all show periods of equilibration which resolves at 0.6 µs. Indicated by the black dashed line is the point which analysis of these systems was performed. SseK1-S286Y does show RMSD variability past the 0.6 µs cut-off, though examination of the simulation shows this is due to motion of the HLH domain.*

### 3.2.3.1.2 Dynamic Motion of the HLH Domain

The SseK1-S286I:FADD-DD complexes are stable over the length of the simulation. The UDP-GlcNAc bound complex shows a similar lack of structural deviations as for NleB1. When UDP is bound however there appears to be an increase in the difference in the C-terminal lid domain, N-terminus and between a loop formed by residues 295 to 305. The dynamics of these three regions are part responsible for the RMSD values for the respective simulations. Comparing the structure to 0.8 µs and 1.0 µs there is the largest difference in RMSD values. At 0.8 µs the HLH domain has disengaged from FADD-DD, returning to contact FADD-DD at 1.0 µs (Figure 3.4). When engaged with the FADD-DD, the HLH domain interacts with FADD-DD$^{K110}$ and FADD-DD$^{L137}$ through a series of hydrophobic interactions from the α4 helix, supported by a salt bridge with SseK1-S286I$^{E163}$ from the α3 helix. In addition to this interaction FADD-DD$^{R135}$ forms interactions with the backbone of a pair of glutamic acid residues between the α3 and α4 helices. While disengaged, the distance between the HLH domain and FADD-DD is such that these contacts are not made.

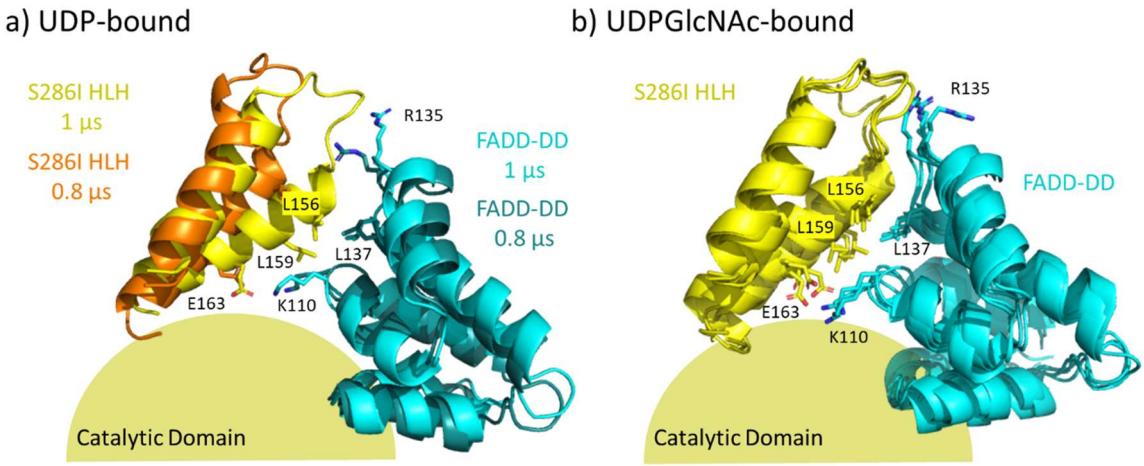



*Figure 3.4 - Molecular dynamics simulations of SseK1-S286I:FADD-DD complexes. Shown are the cartoon representations of the HLH domain and its contacts with FADD-DD. The superposition of the UDP-bound complex at 0.8 µs and 1 µs (a) (enzyme HLH domain shown in in yellow/orange and FADD-DD shown in cyan/teal). The superposition of the UDP-GlcNAc-bound complex at 0.7 µs, 0.8 µs, 0.9 µs and 1 µs (b) (enzyme in yellow and FADD-DD in cyan). Residues that make close contacts have been represented as sticks including SseK1-S286I$^{L156}$, SseK1-S286I$^{L159}$, SseK1-S286I$^{E163}$, FADD-DD$^{K110}$, FADD-DD$^{K137}$ and FADD-DD$^{R135}$.*

The active mutant, SseK1-S286Y shows the highest overall RMSD values along the length of the simulation in both the UPD-bound and UDP-GlcNAc-bound complexes. The RMSD calculations highlight that there is a dynamic event that occurs late in the simulation. Unlike in other simulations, the HLH domain remains disengaged from the FADD-DD forming minimal contacts (Figure 3.5). While these dynamic events do occur after the cut-off it is reasoned that as the HLH domain is known to undergo motion when it does not contact an acceptor it is reasonable to still interpret this section of the simulation. At 0.63 µs there is a dynamic event that increases RMSD comparing the UDP-bound and UDP-GlcNAc bound complexes. The HLH domain of the UDP-GlcNAc-bound complex is completely disengaged from FADD-DD (Figure 3.5). At 0.8 µs the HLH has moved close to FADD-DD and reforms minimal contacts. This observation is not the same for the UDP-bound complex in which the HLH domain does contact FADD-DD over the course of the simulation though there has been a repositioning of FADD-DD. In this complex there is not the same series of complementary interactions between FADD-DD$^{K110}$ and the HLH domain as seen in the other mutants. The closest contact that remains however is between FADD-DD$^{R135}$ and the loop between the α3 and α4 helices of the HLH domain.

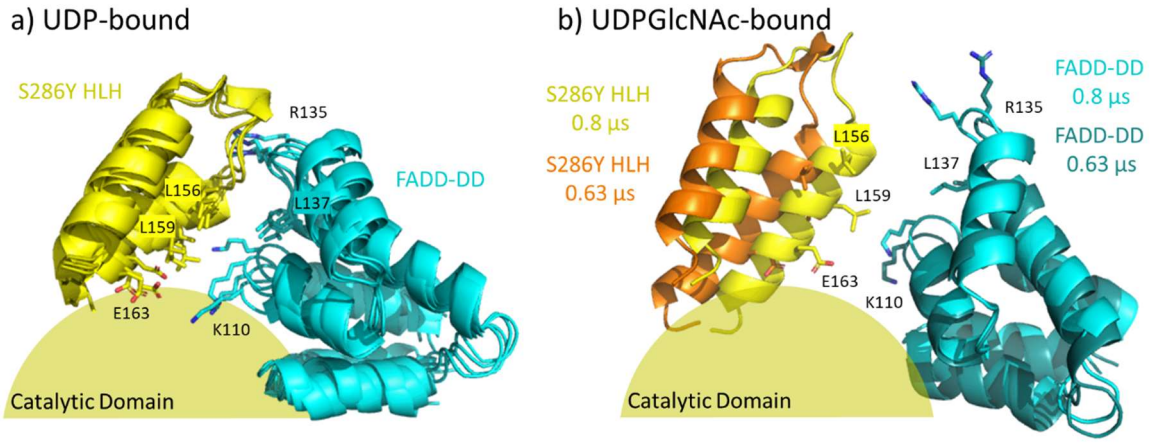



*Figure 3.5 - Molecular dynamics simulations of SseK1-S286Y:FADD-DD complexes. Shown are the cartoon representations of the HLH domain and its contacts with FADD-DD. The superposition of the UDP-bound complex at 0.7 µs, 0.8 µs, 0.9 µs and 1 µs (a) (enzyme in yellow and FADD-DD in cyan. The superposition of the UDP-GlcNAc-bound complex at 0.63 µs and 0.8 µs (b) (enzyme HLH domain shown in in yellow/orange and FADD-DD shown in cyan/teal). Residues that make close contacts in the SseK1-S286I and SseK1-S286N complexes have been represented as sticks including SseK1-S286Y$^{L156}$, SseK1-S286Y$^{L159}$, SseK1-S286Y $^{E163}$, FADD-DD$^{K110}$, FADD-DD$^{K137}$ and FADD-DD$^{R135}$.*

The largest difference in behaviour between UDP-bound and UDP-GlcNAc-bound can be seen for the SseK1-WT complex. The systems have both similar RMSD values until approximately 0.4 µs. At this point there is a large deviation followed by a stabilisation for the UDP-bound complex. For the UDP-GlcNAc-bound complex there is limited change after an initial period of instability and overall, the system is comparatively as stable as the NleB1:FADD-DD complex. In the UDP-GlcNAc-bound complex, comparing the structures for 0.6 µs and 1 µs there is little structural change other than that from the HLH domain between 0.8 µs and 1µs (Figure 3.6) consistent with the SseK1-S286I and SseK1-S286N complexes. In the UDP-bound complex however, at 0.9 µs 1 µs the HLH domain has repositioned and forms different contacts with FADD-DD. Where the loop between the HLH helices in NleB1 forms most interactions with FADD-DD. In the simulation of SseK1-WT contacts are made between the α4 helix and FADD-DD. In this orientation there is a salt bridge between FADD-DD$^{R135}$ and SseK1-WT$^{D153}$ and a hydrophobic interaction between FADD-DD$^{K110}$ and SseK1-WT$^{L159}$ supported by a salt bridge with SseK1-WT$^{D166}$.

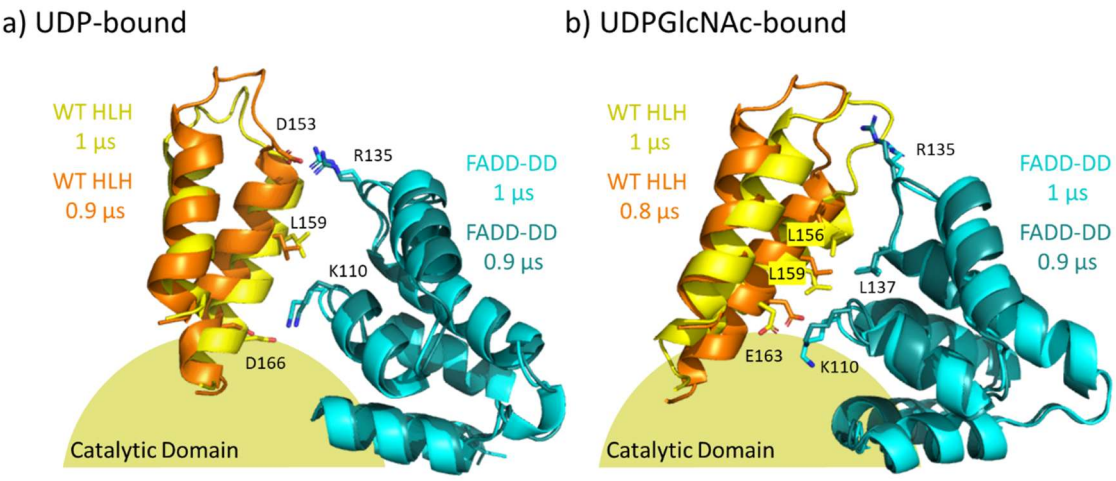



*Figure 3.6 - Molecular dynamics simulations of SseK1-WT:FADD-DD complexes. Shown are the cartoon representations of the HLH domain and its contacts with FADD-DD. The superposition of the UDP-bound complex at 0.9 µs and 1 µs (a) (enzyme HLH domain shown in in yellow/orange and FADD-DD shown in cyan/teal). The superposition of the UDP-bound complex at 0.8 µs and 1 µs (b) (enzyme HLH domain shown in in yellow/orange and FADD-DD shown in cyan/teal). Residues that make close contacts have been represented as sticks including SseK1-WT$^{D153}$, SseK1-WT$^{L156}$, SseK1-WT$^{L159}$, SseK1-WT$^{E163}$, SseK1-WT$^{D166}$, FADD-DD$^{K110}$, FADD-DD$^{K137}$ and FADD-DD$^{R135}$.*

The SseK1-S286N mutant shows a large phase of initial instability in the UDP-GlcNAc-bound complex which resolves at 0.6 µs. This is not seen in the case of the UDP-bound complex, where there are some changes in RMSD past 0.6µs. In both cases similar contacts are seen between the HLH domain and FADD-DD as in the SseK1-S286I complex. Comparing the 0.7 µs and 0.9 µs structures for the UDP-GlcNAc structure show dynamics in the HLH domain as with the other complexes (Figure 3.7). By contrast, the UDP-bound complex is much more stable showing a limited amount of HLH domain movement along the length of the simulation.

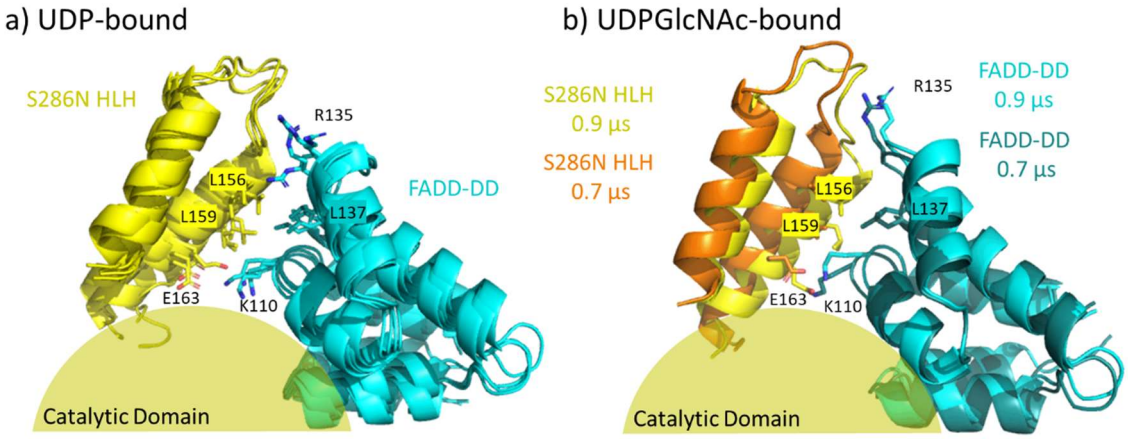



*Figure 3.7 - Molecular dynamics simulations of SseK1-S286N:FADD-DD complexes. Shown are the cartoon representations of the HLH domain and its contacts with FADD-DD. The superposition of the UDP-bound complex at 0.7 µs, 0.8 µs, 0.9 µs and 1 µs (a) (enzyme in yellow and FADD-DD in cyan. The superposition of the UDP-GlcNAc-bound complex at 0.7 µs and 0.9 µs (b) (enzyme HLH domain shown in in yellow/orange and FADD-DD shown in cyan/teal). Residues that make close contacts have been represented as sticks including SseK1-S286N$^{L156}$, SseK1-S286N$^{L159}$, SseK1-S286N$^{E163}$, FADD-DD$^{K110}$, FADD-DD$^{K137}$ and FADD-DD$^{R135}$.*

### 3.2.3.1.3 RMSF Calculations Highlight Dynamic Loops

RMSF calculations of both UDP-GlcNAc-bound and UDP-bound complexes of NleB1, SseK1-WT and SseK1 mutants show a series of domains that are dynamic along the length of the simulation. The largest sequence of residues that shows motion is from 140 to 170 (Figure 3.8), representing the HLH domain. In all simulations of UDP-GlcNAc-bound complexes the HLH domains all show similar levels of fluctuation aside from the complex including SseK1-S286I. This agrees with the observations from the RMSD calculations which presents the HLH domain of SseK1-S286I showing less movement overall than the other simulations (Figure 3.8).

In addition to the residues of the HLH domain, there are three additional dynamic sequences present in the enzyme near to the FADD-DD binding domain (Figure 3.8 and 3.9). Residues 240-255 and 295-305 both contain loops within the structure of the enzyme and residues 321 to 330/336 is the C-terminal lid domain (Figure 3.9). Residues 240-255 and 295-305 are adjacent to the FADD binding site over the course of simulation they are very dynamic and able to form contacts with FADD-DD. Residues 240-255 are all similarly dynamic in all enzymes either UDP-GlcNAc-bound or UDP-bound. This is not the case however for the loop formed by resides 295-305 for which there are differences in dynamics for both UDP-GlcNAc-bound and UDP-bound complexes (Figure 3.8). Given the proximity of these loops to FADD-DD in these simulations, changes in the binding orientation of FADD-DD can impact the dynamics of these loops.

The lid domain of all enzymes is similarly dynamic in the case of UDP-GlcNAc-bound and UDP-bound complexes. However, there is an increase in dynamics for some UDP-bound complexes. In these MD simulations the presence of UDP-GlcNAc can be seen to stabilise the lid domain keeping RMSF values between 1-2 Å aside from the final terminal residues of the SseK1 enzymes (Figure 3.8). This behaviour agrees with a previous study that proposed the presence of a ligand in the catalytic domain orders the lid domain (36). In that same study it was proposed that some residues from the lid domain can participate in hydrogen bonding interactions with the GlcNAc ring of UDP-GlcNAc. This is again in agreement with the RMSF calculations.

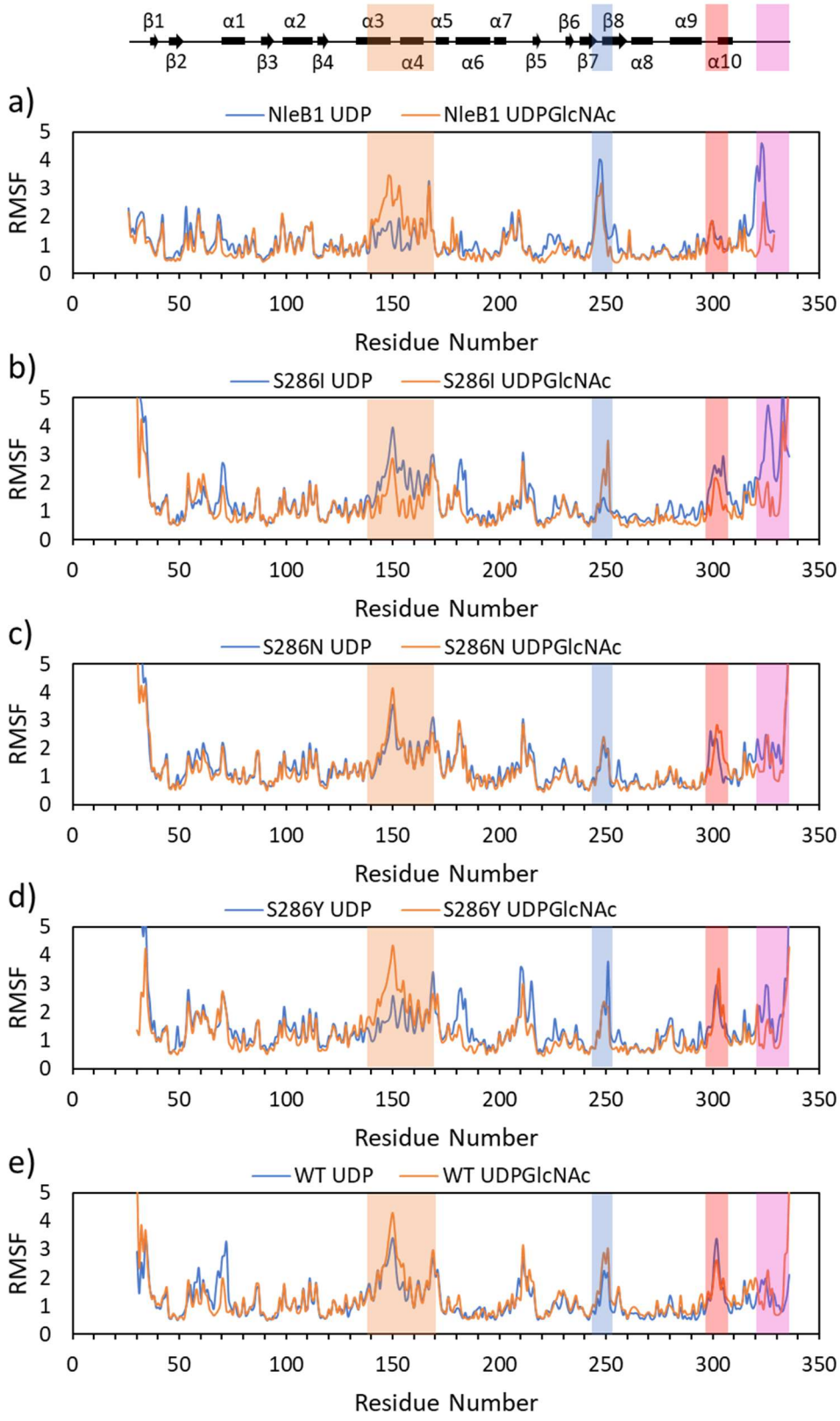



*Figure 3.8 - RMSF (Å) of side chain non-hydrogen atoms. Shown are both UDP and UDP-GlcNAc-bound complexes of FADD-DD with NleB1 (a), SseK1-S286I (b), SseK1-S286N (c), SseK1-S286Y (d) and SseK1-WT (e). Analysis was performed from 0.6 µs to 1µs. Highlighted are dynamic regions that are close to FADD-DD including the HLH domain (orange), two loops (blue, red) and the Lid domain (pink).*

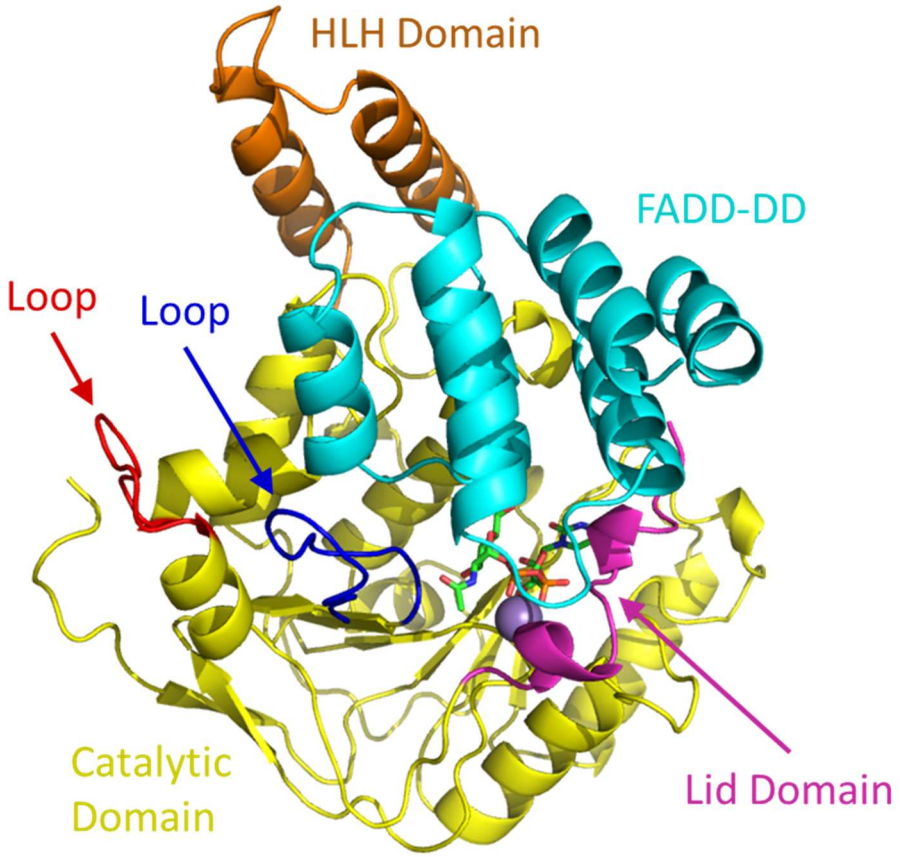



*Figure 3.9 - The UDP-GlcNAc-bound SseK1-S286Y:FADD-DD complex. Highlighted are the dynamic domains and loops near FADD-DD (cyan). These domains include the catalytic domain (yellow), HLH domain (orange), lid domain (pink). Included alongside these domains are the loops formed by residues 240-255 (blue) and 295-315 (red).*

### 3.2.3.2 The S286Y Mutation Influences Interactions with FADD-DD

From the kinetics study of the SseK1 mutants it was clear that the single point mutation, S286Y, triggers FADD-DD recognition. On the surface it would be possible to assume that it is this residue and its interactions alone that allows FADD activity, behaving in a similar manner as in NleB1. The results of the MD study of these systems, however, showed that the presence of this residue allowed for other interactions that stabilised the target acceptor arginine residue, $FADD^{R117}$. In addition to this finding, simulations of the other SseK1 mutants outlined the importance of the tyrosine residue itself in these interactions. The following sections will outline these findings from the MD experiments and will discuss the structural rationale for why the S286Y mutation significantly changes the behaviour of SseK1.

### 3.2.3.2.1 Observations of Known Enzyme:FADD-DD Interactions

More focused observations were initially placed on residues that are known to interact with acceptor FADD$^{R117}$. These include NleB1$^{E253}$ of the HEN motif (36) and NleB1$^{Y282}$ which provides a pi-cation stack with FADD$^{R117}$ alongside the respective aligned residues within SseK1-WT and its mutants. In addition, a spotlight was placed on the mutated site of SseK1-S286Y, and its aligned position in NleB1-WT and the other SseK1 constructs. To measure the stability of specific interactions distance measurements were made between residues and plotted as a histogram. To provide a consistent reference position for different residues, the centre of mass of each residue was used to measure these distances. These plots show different populations of interactions which indicate relative stability of individual interactions and show the presence of multiple interaction modes. NleB1$^{Y283}$ remained in a stable interaction with FADD$^{R117}$ along the length of the simulation (Figure 3.10). This interaction was well conserved in the other complexes, including those with enzymes that were not FADD-active. To measure how well NleB1$^{Y284}$, SseK1-WT$^{S286}$ and equivalent residue in the mutants remained buried within the hydrophobic pocket in FADD-DD, measurements were taken from the centre of mass of the residues to the centre of mass of FADD-DD$^{I126}$ (Figure 3.11).

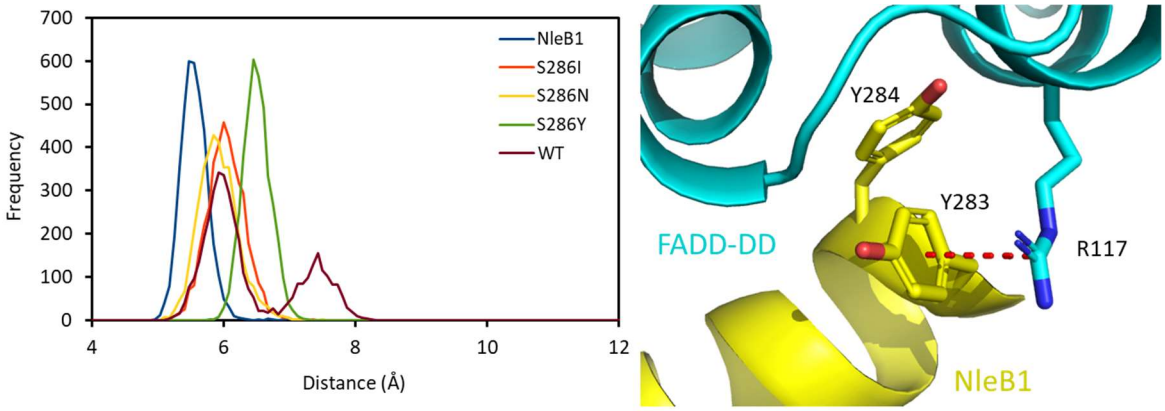



*Figure 3.10 - Plot of the π-cation stacking interaction of FADD-DD$^{R117}$. Distance distribution plot of inter-residue distances between the centre of mass of NleB1$^{Y283}$, SseK1-WT$^{Y285}$, SseK1-S286I$^{Y285}$, SseK1-S286N$^{Y285}$ and SseK1-S286Y$^{Y285}$ to FADD-DD$^{R117}$ (left) in the UDP-GlcNAc-bound complexes. Cartoon representation of the measurement using the example of NleB1 (enzyme shown in yellow and FAD-DD shown in cyan) with an example of the inter-residue distance measurement (red dashed line) (right).*

Over the course of the simulation the tyrosine residues NleB1$^{Y284}$ and SseK1-S286Y$^{Y286}$ remained in place, deeply buried within the hydrophobic pocket, stabilised by a hydrogen bonding interaction with the backbone of FADD-DD$^{V121}$ (Figure 3.11). The distance-based analyses then indicated that Ssek1-S286N$^{N286}$, Ssek1-S286I$^{I286}$ and SseK1-WT$^{S286}$ were placed much further from FADD$^{I126}$. Inspection of these specific interactions showed that neither SseK1-WT$^{S286}$ or SseK1-S286N$^{N286}$ could be placed within the hydrophobic pocket, unable to form complementary interactions with the residues or backbones. Instead, these residues favoured interactions with the hydrogen bond accepting FADD-DD$^{D123}$ adjacent to the hydrophobic binding site (Figure 3.12). The isoleucine residue, SseK1-S286I$^{I286}$, though able to from hydrophobic interactions with the pocket in FADD is not observed to occupy this space curiously being measured to be the furthest away from FADD$^{I126}$. This inability to form an interaction is due to the relative size of the isoleucine side chain compared to that of tyrosine, being far too bulky to occupy the hydrophobic pocket and that the pocket is not dynamic enough to accommodate the increased size of the residue. These observations indicate that the pocket within FADD is selective by two mechanisms: the first is that a residue must form complementary interactions and the second being that the residue must be of the correct size.

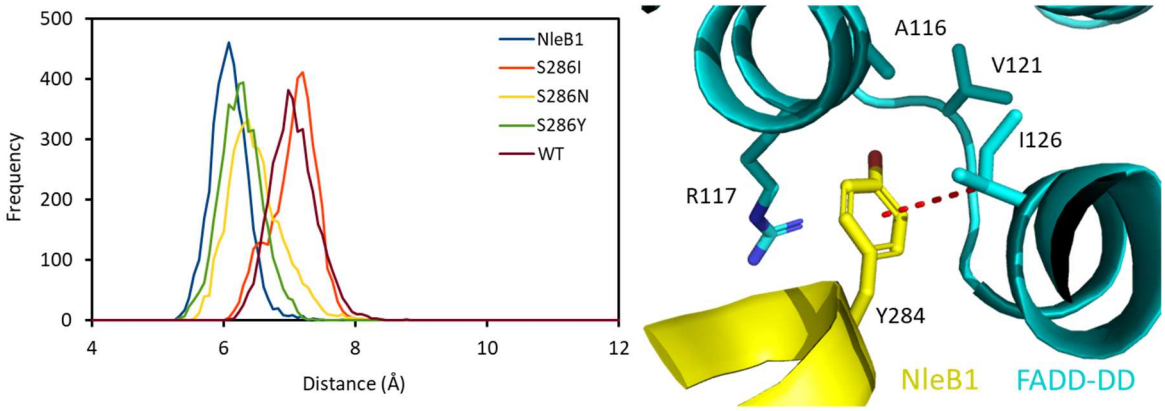



*Figure 3.11 - Plot of the FADD-DD hydrophobic pocket anchor interaction. Distance distribution plot of inter-residue distances between the centre of mass of NleB1$^{Y284}$, SseK1-WT$^{S286}$ and SseK1-S286$^{X286}$ to FADD-DD$^{I126}$ (left) in the UDP-GlcNAc-bound complexes. Cartoon representation of the measurement using the example of NleB1 (enzyme shown in yellow and FAD-DD shown in cyan) with an example of the inter-residue distance measurement (red dashed line).*

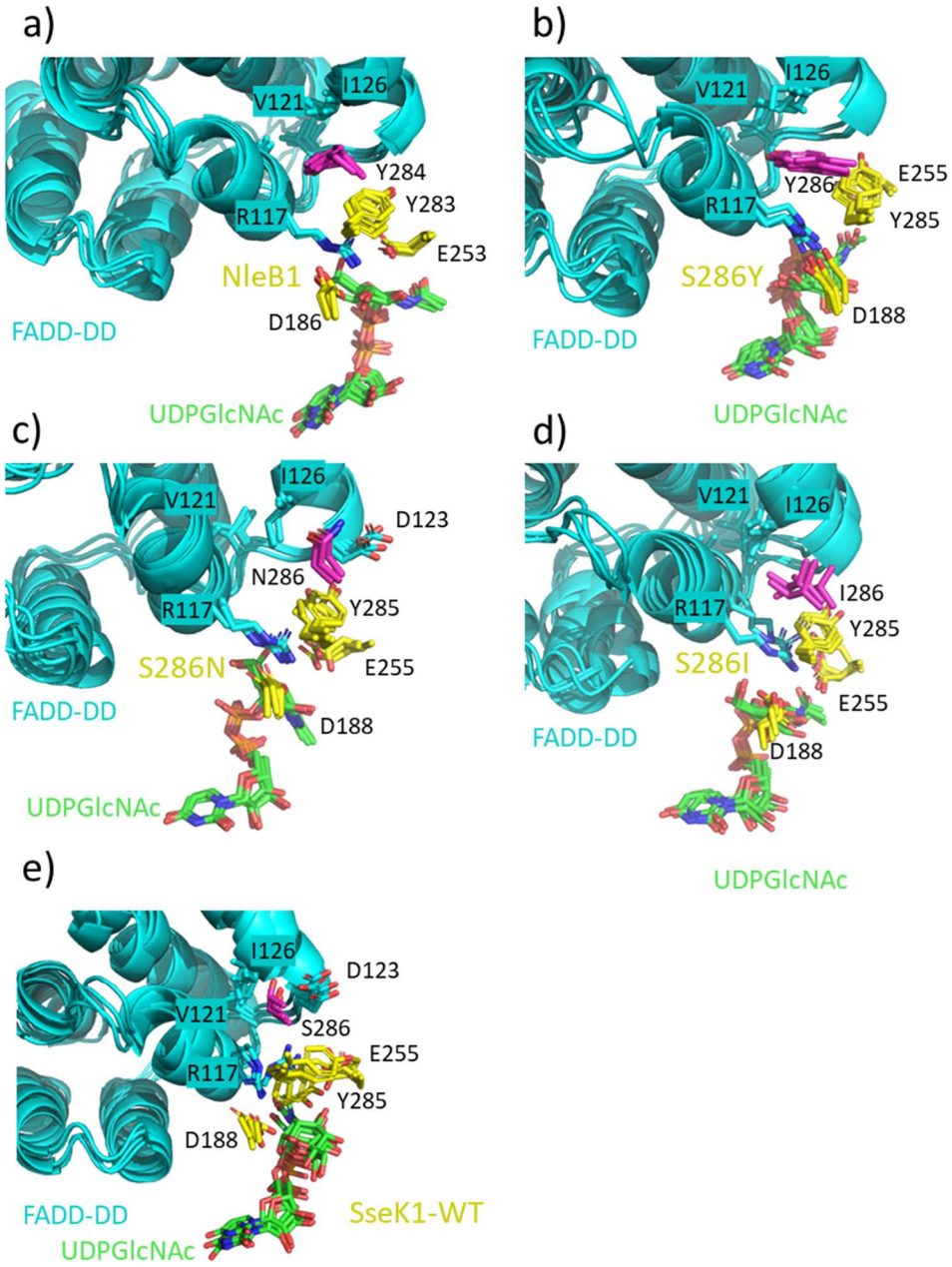



*Figure 3.12 - MD simulations of UDP-GlcNAc-bound complexes. Shown are NleB1, SseK1-WT and SseK1-S286X mutants in complex with FADD-DD. Super position of frames at 0.6 µs, 0.7 µs, 0.8 µs and 1 µs show that the interactions are stable along the length of the simulation. For NleB1 and SseK1-S286Y the tyrosine at position 284 and 286 respectively remains in place in the hydrophobic pocket formed by FADD-DD$^{V121}$/FADD-DD$^{I126}$ supported by a hydrogen bonding interaction between the phenyl hydroxyl group of the tyrosine side chain and the backbone carbonyl of FADD-DD$^{V121}$. In the case of SseK1-S286N and SseK1-WT the asparagine and serine residues respectively are unable to form complementary interactions with the hydrophobic pocket, instead forming hydrogen bonding interactions with the neighbouring FADD-DD$^{D123}$. Despite being a hydrophobic residue, the isoleucine of SseK1-S286I is unable to fit in the hydrophobic pocket due to a steric clash.*

### 3.2.3.2.2 An Aspartic Acid Stabilises FADD$^{R117}$

NleB1$^{E253}$ forms an efficient, stable salt bridge with FADD$^{R117}$ over the course of the simulation with little deviation. This interaction is also conserved in the FADD complexes with SseK1-WT, SseK1-S286I and SseK1-S286N. The behaviour of this glutamic acid agrees with the literature which highlights this residue as important in recognition of the acceptor arginine residue (36). However, this interaction is not conserved in SseK1-S286Y with SseK1-S286Y$^{E255}$ held far from FADD$^{R117}$ (Figure 3.13). The lack of a stable glutamic acid interaction in SseK1-S286Y was unexpected. Until now it was thought that all residues of the HEN motif were important for enabling catalysis. The glutamic acid being needed to act as a catalytic base, 'activating' the acceptor FADD$^{R117}$ residue. Instead, these simulations find that for the FADD-active enzymes under dynamic conditions an additional residue can interact with FADD-DD$^{R117}$. This residue is an aspartic acid, which for SseK1-S286Y is the only residue that stably co-ordinates the acceptor arginine. SseK1-S286Y$^{D188}$ assumes the role of catalytic base forming an efficient, stable salt bridge with FADD$^{R117}$ via the carboxylic acid in its side chain (Figure 3.13). Distance-based measurements were performed, finding that this interaction is conserved in the active enzymes (Figure 3.13).

Moreover, this aspartic acid can form interactions with only one eta nitrogen of arginine (Figure 3.13). This is potentially beneficial as it leaves exposed a single eta nitrogen, the site of GlcNAc transfer in FADD$^{R117}$. Having this nitrogen exposed allows for catalysis to occur more easily by permitting the nitrogen to become close to the anomeric carbon of the GlcNAc ring. On the other hand, glutamic acid NleB1$^{E253}$ forms interactions with both eta nitrogens. The consequence of this is that the glutamic acid will need to move out the way for GlcNAc transfer to occur.

One study supports the importance of the aspartic acid residue identified here, finding that bacteria that expressed NleB1-D186A displayed impaired bacterial colonisation (35). This same study also reported the NleB1:FADD-DD crystal structure (PDB entry 6ACI) but this does not contain the same aspartic acid-FADD-DD$^{R117}$ interaction. The reason for this may be that the use of MD simulation allows the interaction to evolve under dynamic conditions, placing the two proteins in the proper orientation for this interaction to take place. Crystal structures can be limited as they can only represent a single state of a system which is able to crystallise, which may not be entirely representative of a system in solution.

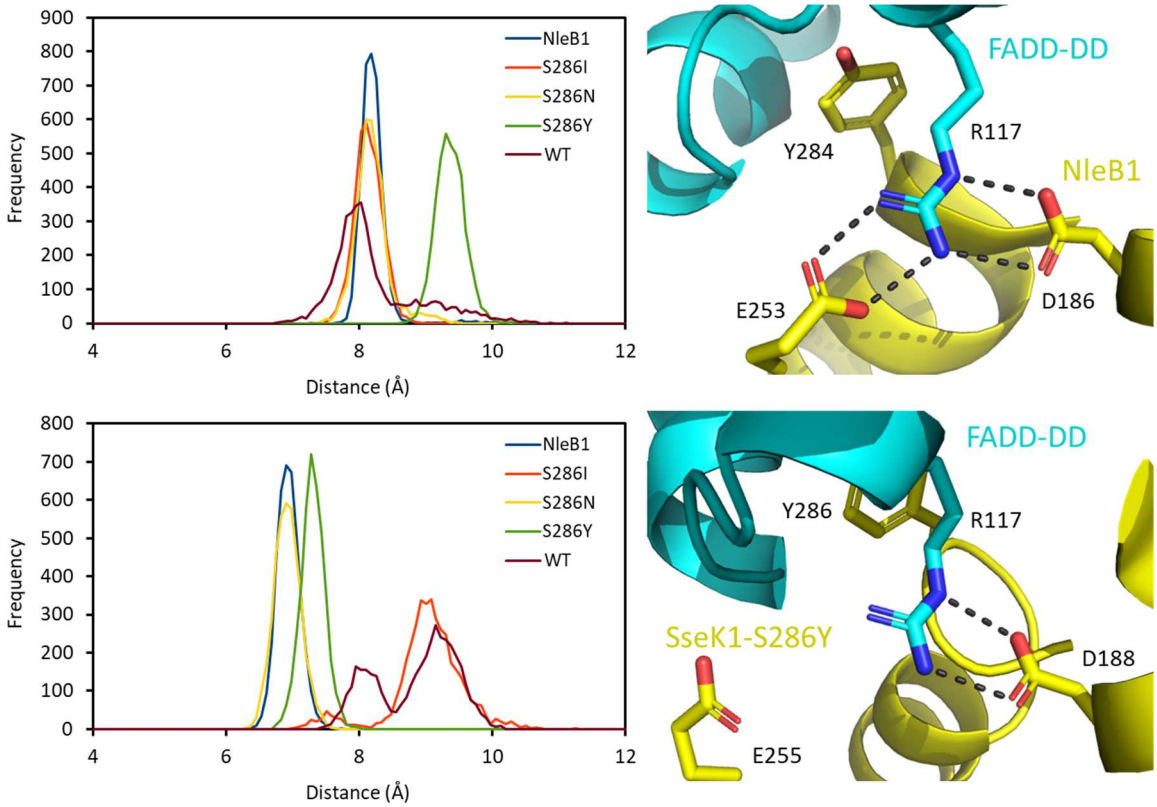



*Figure 3.13 - Plots of the FADD-DD$^{R117}$ salt bridge interactions with the NleB1 or SseK1 enzymes. Distance distribution plots (left) showing the difference in the inter-residue distances for coordination of FADD-DD$^{R117}$ by the HEN motif glutamic acid (top) and the identified aspartic acid (bottom) in the UDP-GlcNAc-bound complexes. In NleB1 (top right), and in other enzymes that use the glutamic acid to stabilise FADD-DD$^{R117}$, both eta nitrogen atoms of the guanidine group are coordinated whereas in SseK1-S286Y (bottom right) as the glutamic acid is not used to co-ordinate FADD-DD$^{R117}$, but only the aspartic acid is used, one eta nitrogen remains exposed (enzyme shown in yellow and FAD-DD shown in cyan).*

The inclusion of the aspartic acid interaction impacts the overall dynamics of FADD$^{R117}$. To quantify this, RMSF calculations were used to gauge the overall mobility of FADD$^{R117}$ along the length of the simulations (Figure 3.14). The result from these calculations is that the active complexes (NleB1, SseK1-S286Y and SseK1-S286N) showed FADD$^{R117}$ at the lowest level of mobility with RMSF values of less than 1Å. This indicates that a contributing reason for inactivity of the SseK1-S286I and SseK1-WT is not able to effectively stabilise FADD$^{R117}$ and maintain it in the proper orientation for GlcNAc transfer.

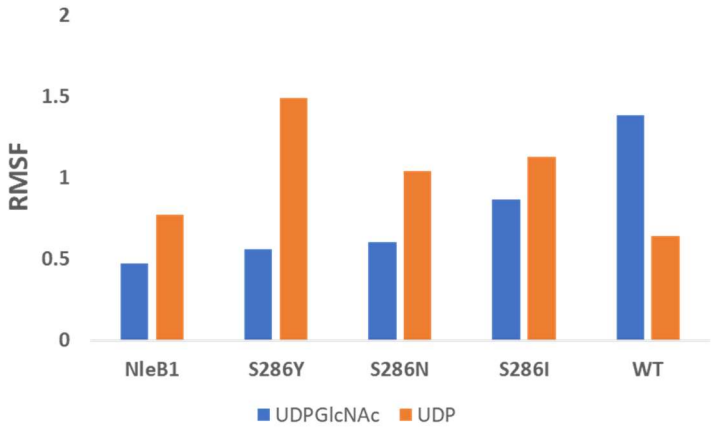



*Figure 3.14 - RMSF (Å) of the FADD-DD$^{R117}$ side chain from 0.6 µs. Show are results for simulations of both UDP-GlcNAc-bound (blue bars) and UDP-bound (orange bars) complexes.*

A comparison of the RMSF measurements for UDP-bound complexes shows in all cases, except for SseK1-WT, there is more motion for FADD-DD$^{R117}$. A reason for this increase in motion is that the GlcNAc ring of UDP-GlcNAc participates in interactions with FADD-DD$^{R117}$ alongside enzyme residues (Figure 3.14). Looking at the active enzyme complexes, NleB1:FADD-DD and SseK1-S286Y:FADD-DD with UDP-GlcNAc present, the guanidinium group is coordinated by the hydroxyl groups of the GlcNAc ring by hydrogen bonding interactions. Between the two enzymes the coordination environment for the GlcNAc ring is different, in part owing to the ability of NleB1 to use the glutamic acid residue of the HEN motif to coordinate FADD-DD$^{R117}$ alongside NleB1$^{D186}$ (Figure 3.15). In NleB1, both eta hydrogens are coordinated by hydroxyl 4 and 6 of the GlcNAc ring. In the case of SseK1-S286Y, only one eta nitrogen is coordinated by hydroxyl 5 of the GlcNAc ring. This hydrogen bonding network is not only important for positioning the GlcNAc ring in the active site but also contributes to stabilisation of FADD-DD$^{R117}$. Of course, as UDP lacks the GlcNAc ring it is unable to provide such interactions and therefore other interactions around FADD-DD$^{R117}$ are destabilised.

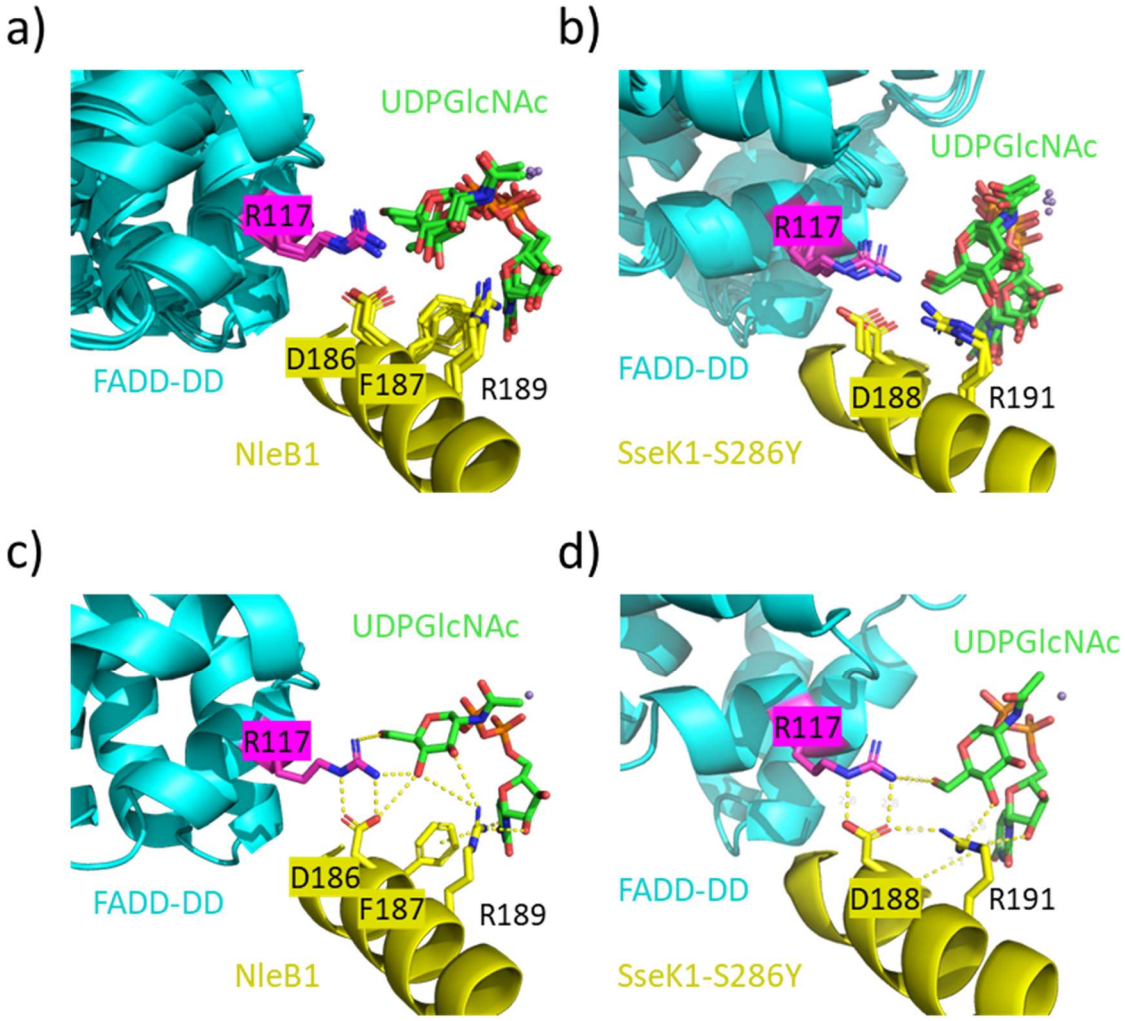



*Figure 3.15 - GlcNAc ring stabilisation in NleB1 and SseK1-S286Y MD simulations. Interactions with aspartic acid-arginine dyad (D186/R189 in NleB1 and D188/R191 in SseK1-S286Y) support the GlcNAc ring in the active site via a hydrogen bonding network. Alongside SseK1-S286Y$^{D188}$ and both NleB1$^{E253}$/NleB1$^{D186}$ the GlcNAc ring also contributes to interactions with FADD-DD$^{R117}$ (enzyme shown in yellow, FAD-DD shown in cyan and FADD-DD$^{R117}$ highlighted in pink). Superposition of frames at 0.6 µs, 0.7 µs, 0.8 µs and 1 µs show that this network is stable in both NleB1 (a) and SseK1-S286Y (b). The networks are represented by a single frame from 1 µs for NleB1 (c) and SseK1-S286Y (d) with hydrogen bonds shown (yellow dashed lines).*

In the simulations of the SseK1-WT:FADD-DD complex, FADD-DD$^{R117}$ undergoes less fluctuations when UDP is bound. This is unique as in all other systems, the lack of the GlcNAc ring leads to destabilisation of FADD-DD$^{R117}$ and more fluctuations. However, in the SseK1-WT:FADD-DD complex FADD-DD$^{R117}$ forms a salt bridge with SseK1-WT$^{D283}$ (Figure 3.16).

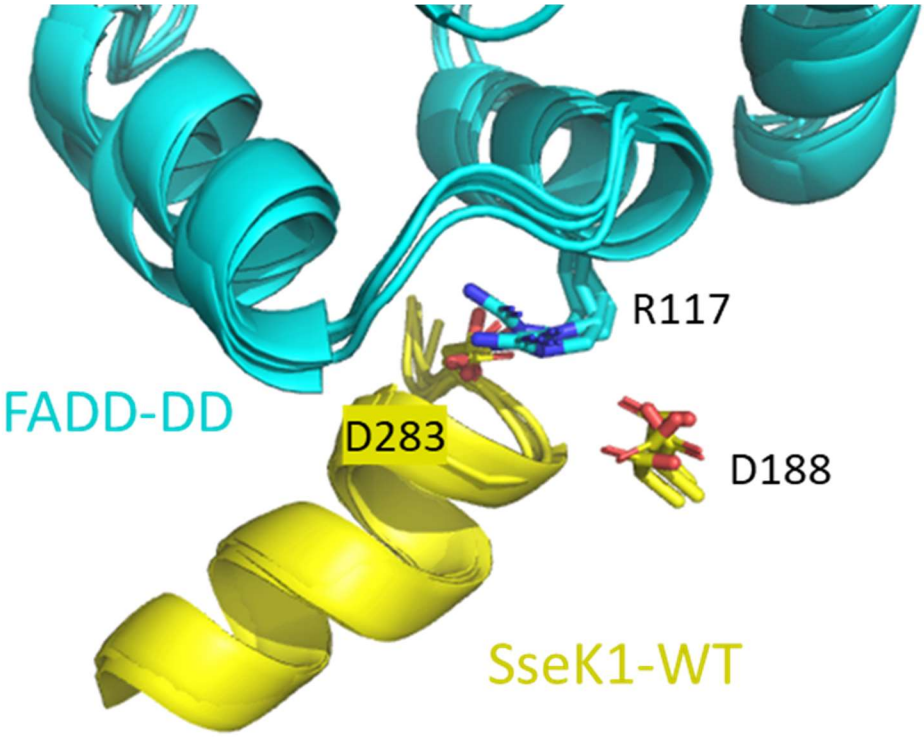



*Figure 3.16 - MD simulation of the UDP-bound SseK1-WT:FADD-DD complex. FADD-DD$^{R117}$ is stabilised in this complex by an interaction with SseK1-WT$^{D283}$ which results in the decreased level in flux shown in the RMSF calculations. Super position of frames at 0.6 µs, 0.7 µs, 0.8 µs and 1 µs show that the interaction is stable along the length of the simulation.*

### 3.2.3.2.3 Interactions are Different in the UDP Complexes

RMSD and RMSF calculations indicate that there are differences in the global and local behaviour of the UDP-bound complexes. Many structures such as the HLH domain, lid domain and loops in the enzymes have different dynamics. Measurements of distances, the same as those performed in 3.2.3.2.1 and 3.2.3.2.2, show different distance distributions for these interactions (Figure 3.17).

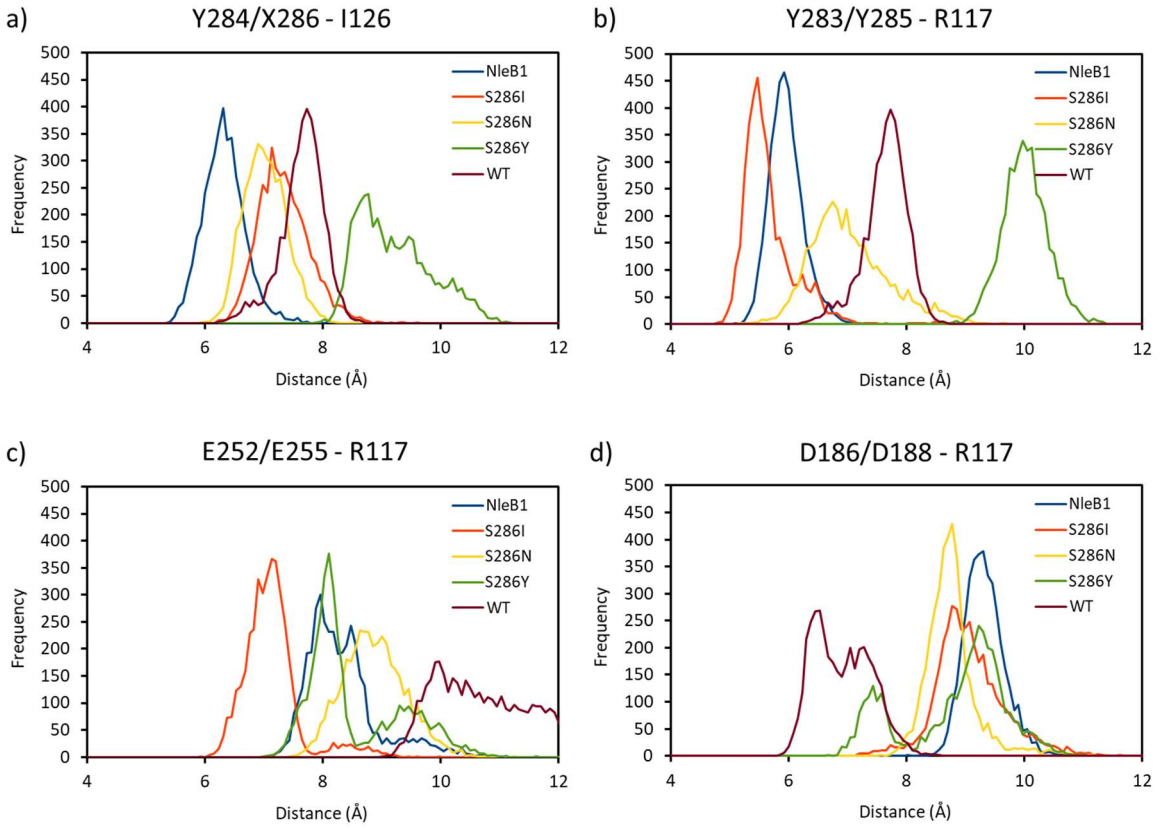



*Figure 3.17 - Distance plots of interactions seen in the UDP-bound complexes. Distances between the centre of mass of NleB1$^{Y284}$/SseK1-WT$^{S286}$/SseK1-S286I$^{I28}$/SseK1-S286N$^{N28}$/SseK1-S286Y$^{Y286}$ to FADD-DD$^{I126}$ (a). The distances of NleB1$^{D186}$/SseK1-WT$^{D188}$/SseK1-S286X$^{D188}$ (b), NleB1$^{E252}$/SseK1-WT$^{E255}$/SseK1-S286X$^{E255}$ (c) and NleB1$^{Y283}$/SseK1-WT$^{Y285}$/SseK1-S286X$^{Y285}$ to FADD-DD$^{R117}$ (d).*

In the presence of UDP-GlcNAc, SseK1-S286Y$^{Y286}$ forms a stable interaction with a hydrophobic pocket within FADD. In the presence of UDP however, this interaction appears lost (Figure 3.17a). Examination of the simulation of the SseK1-S286Y:FADD-DD complex in the presence of UDP shows that the residue does begin inserted into the pocket however

shifts out of place and cannot easily return (Figure 3.18). For NleB1, the equivalent tyrosine residue remains stably inserted into the pocket with a distance distribution like that of the UDP-GlcNAc-bound complex. SseK1-WT, SseK1-S286I and SseK1-S286N have their residues at position 286 in a similar position to that of the UDP-GlcNAc-bound complex also. This change in placement of SseK1-S286Y$^{Y286}$ in the UDP-bound complex can be owed to a shift in the placement of FADD-DD in SseK1-S286Y. comparing the UDP-GlcNAc and UDP complexes the α4 helix of FADD-DD has shifted towards SseK1-S286Y$^{Y286}$ precluding it from entering the hydrophobic pocket between the α3 and α4 helices of FADD-DD.

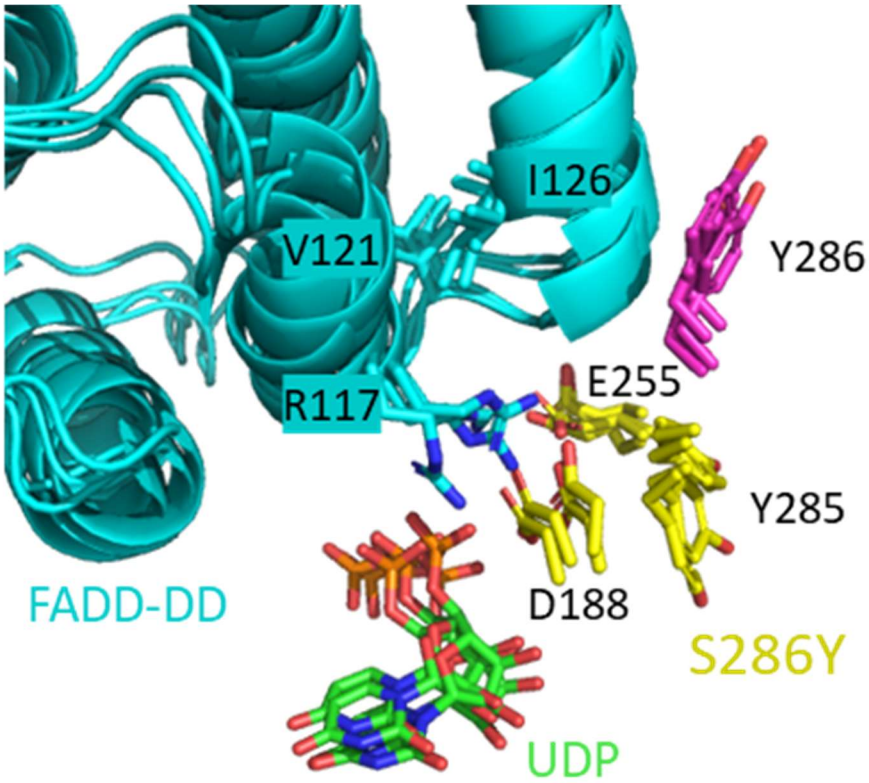



*Figure 3.18 - MD simulation of the UDP-bound SseK1-S286Y:FADD-DD complex. SseK1-S286Y$^{Y286}$ does not interact with the hydrophobic pocket the same way that it can in the UDP-GlcNAc-bound complex. Superposition of frames at 0.6 µs, 0.7 µs, 0.8 µs and 1 µs show that the interaction is stable along the length of the simulation.*

UDP-bound NleB1 keeps a similar interaction between NleB1$^{Y283}$ and FADD-DD$^{R117}$ (3.17b). The largest change in distance distribution for this interaction comes from SseK1-S286Y. The distance distribution shows not only that the residue has shifted to a larger distance from FADD-DD$^{R117}$ but also that the residue is mobile, given the broadened distribution of the interaction. The other mutants can be seen to also have a smaller increase in distances compared to the UDP-GlcNAc-bound system.

UDP-GlcNAc-bound complexes all show a stable FADD-DD$^{R117}$ interaction with the HEN motif glutamic acid apart from SseK1-S286Y. For the UDP-bound complexes this interaction is improved for SseK1-S286Y, but the distribution of distances is bimodal (Figure 3.17c). The major distribution of distances around 8 Å is narrow suggesting a stable interaction. However, there is also a second minor distribution around 9.5 Å indicating that the residue does spend some time disengaged from FADD$^{R117}$. There is a similar distribution of distances seen in the NleB1:FADD-DD complex. For SseK1-S286I, the distance between the two residues is shortened however for SseK1-S286N and SseK1-WT the interaction is compromised.

From the analysis of the UDP-GlcNAc-bound SseK1-S286Y:FADD-DD complex, FADD-DD$^{R117}$ is not stabilised by the accepted catalytic base of the HEN motif, SseK1-S286Y$^{E255}$. Instead, this interaction is replaced by SseK1-S286Y$^{D188}$. In the presence of UDP however, this interaction is lost appearing to have been replaced by the HEN motif glutamic acid (Figure 3.17c and 3.17d). The interaction is additionally lost in the case of NleB1 and SseK1-S286N. For SseK1-S286I the interaction remains the same however, in SseK1-WT the aspartic acid in much closer to FADD$^{R117}$. However, the distribution is broad, suggesting some instability in this interaction.

## 3.2.4 The SseK1-S286Y:FADD-DD Interaction, Studied by 2D NMR

The MD simulations provided a prediction of the molecular basis by which the single S286Y point mutation triggers FADD activity in SseK1. Given the similarity in tertiary structure between SseK1 and NleB1 it can be assumed that in solution the SseK1-S286Y mutant would be able to interact with FADD-DD in a similar manner as shown in the NleB1:FADD-DD crystal structure. In the MD study of SseK1-S286Y in complex with FADD-DD, this assumption was made when the systems were built. Though the simulations showed that the SseK1-S286Y:FADD-DD complex could equilibrate and form sensible contacts; computer simulation of chemical systems are limited by their starting structures. To provide experimental evidence for these predictions, two-dimensional NMR experiments were used, specifically $^{1}H$-$^{15}N$ HSQC and $^{1}H$-$^{15}N$ HSQC TROSY. For these experiments, FADD-DD was isotopically labelled with $^{15}N$ as there was already an assigned $^{15}N$ spectrum available in the literature (55). The following sections will outline the optimisation of experiments used to observe enzyme:FADD-DD interactions what the titration shows about the contacts made between these two proteins in solution.

### 3.2.4.1 Optimisation of Experimental Conditions

Initially, experiments were performed using conditions that NleB1 was known to be stable in, as it is prone to precipitation. In place of the paramagnetic Mn(II) ion that the NleB/SseK enzymes use, Mg(II) was used to minimise any additional source of (paramagnetic) relaxation. In addition, Mg(II) has the same charge and an ionic radius similar to Mn(II), 0.72 Å and 0.67 Å respectively (octahedral coordination) (87). Using these conditions, it was possible to recover a high-quality spectrum of $^{15}N$-labelled FADD-DD in solution (Figure 3.19). Titrations were then performed with an excess of enzyme. However, it was found that there was significantly low acquired signal in the titration spectra (Figure 3.19). It was reasoned that main issue of this approach was that the overall size of the NleB1:FADD-DD complex was large enough to suffer significant $T_2$ relaxation leading to extreme signal broadening.

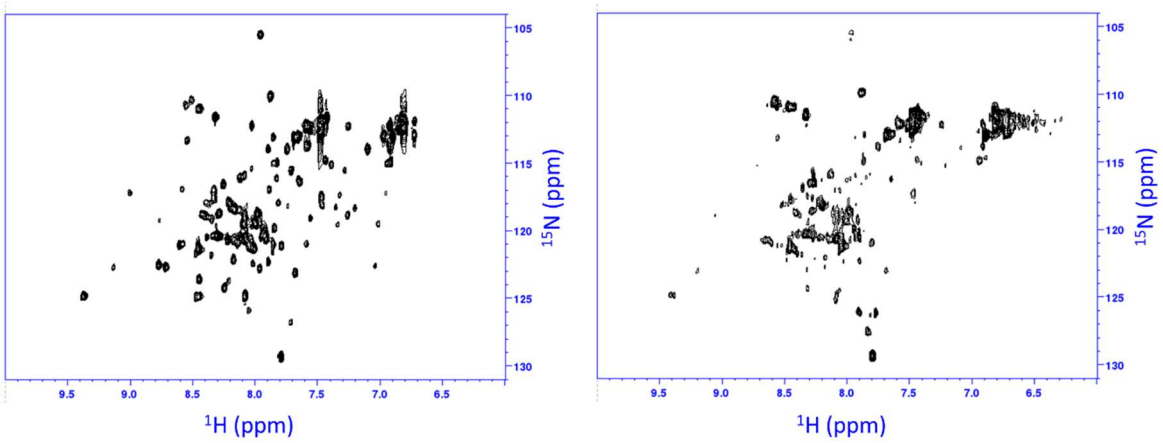



*Figure 3.19 - $^{1}H$-$^{15}N$ HSQC spectra for $^{15}N$-FADD-DD in TRIS buffer. The apo spectra (left) was recorded with a FADD-DD concentration of 910 µM at 800 MHz in a buffer composed of 25 mM TRIS, 150 mM NaCl, 10% $D_2O$, pH 7.5. Shown also is a single point in the titration (right) with a 1:1 ratio of FADD-DD to SseK1-S286Y at a concentration of 200 µM.*

Because of the issues with observed signal intensity, the experimental approach had to be redesigned. To that end the first change was to move to using a TROSY version of HSQC. This is because the TROSY element in the pulse sequence selects for the component of the signal where the relaxation mechanisms cancel (56). Secondly, future titrations were to be performed with FADD-DD in excess rather than the enzyme. This approach was taken due to practical concerns around the amount of available protein but also offers the benefit of reducing the fraction of bound protein and therefore increasing amount of acquired signal. In the early TROSY experiments, experimental conditions were chosen to match those in the publication that assigned FADD-DD signals (55). These experiments produced high-quality spectra of FADD-DD (Figure 3.20). There were concerns however, that a phosphate buffer would be incompatible with study of the enzyme:FADD-DD complex. This is because phosphate ions are known to sequester the Mg(II) ions in solution. This spectrum though is still very useful as it can be compared to different experimental conditions that are more compatible with this system and check for any potential buffer component or, pH related effects on the spectra.

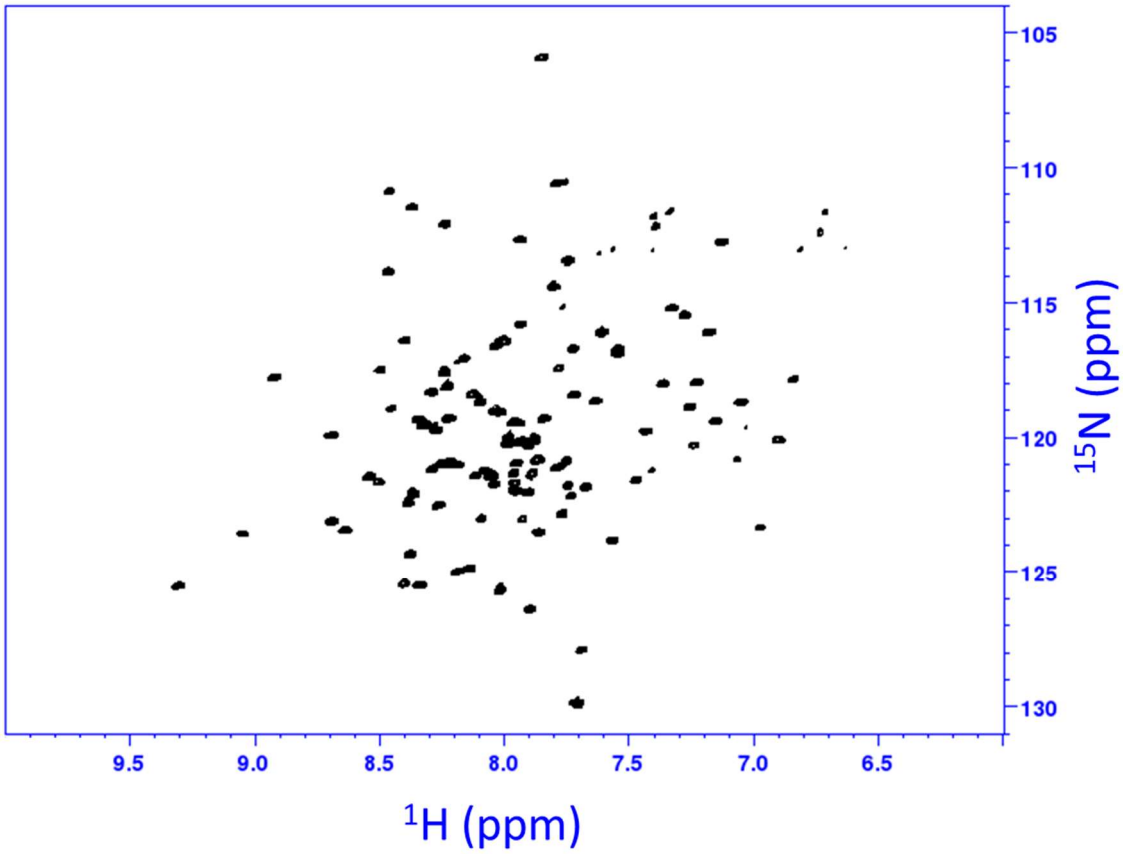



*Figure 3.20 - $^{1}H$-$^{15}N$ TROSY spectrum for $^{15}N$-FADD-DD in potassium phosphate buffer. The spectrum was recorded with a FADD-DD concentration of 200 µM at 800 MHz in a buffer composed of 50 mM potassium phosphate, 150 mM NaCl, 1 mM DTT, 10% $D_2O$, pH 6.2.*

To replace the phosphate buffer, TRIS buffer was chosen as it is known to be compatible with the system from previous experiments. Using these conditions, it was possible to record a high-quality spectrum of FADD-DD which, when compared to the spectrum in phosphate buffer, contained very similar chemical shifts (Figure 3.21). The use of TRIS buffer, however, was found to not be optimal for these experiments as the desired pH of 6.2 was too far outside the buffering capacity of TRIS (pH 7 to 9). This was found to be an issue as on addition of $MgCl_2$ and UDP to the system there were significant differences in chemical shifts (Figure 3.21) which could be related to changes in pH. Because of this the decision was made to move to a buffer which was chemically similar, to have the best possible chance of system stability, which could reliably buffer pH 6.2. The buffer chosen to fulfil this role was BIS-TRIS.

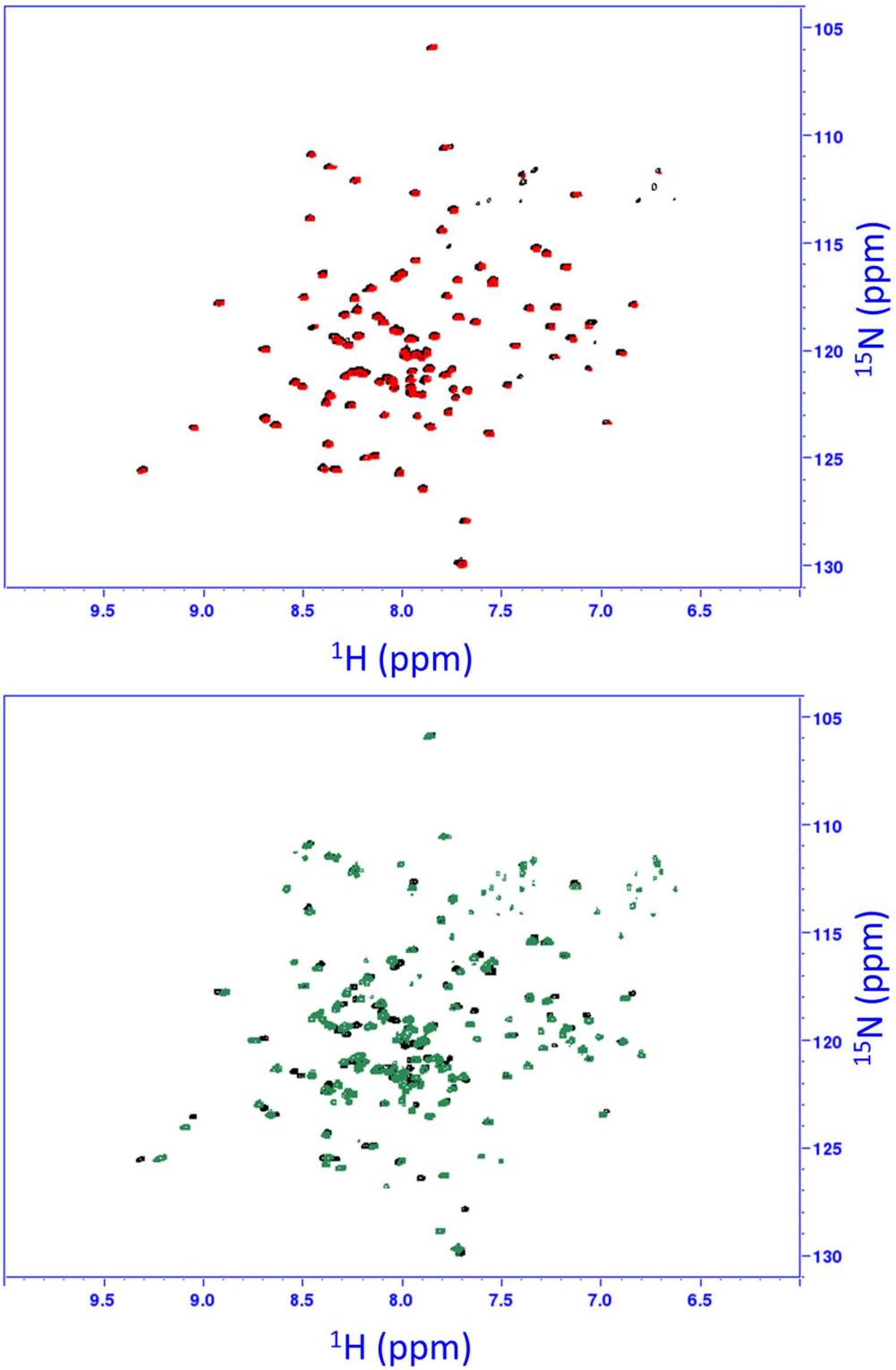



*Figure 3.21 - Comparison of TRIS and potassium phosphate buffer. (Top) Overlay of 200 µM $^{15}$N-FADD-DD spectrum in 50 mM Potassium Phosphate buffer, 150 mM NaCl, pH 6.2 (black) and 200 µM FADD-DD spectrum in 25 mM TRIS buffer, 150 mM NaCl, pH 6.2 (red) showing only very subtle chemical shift difference between the two buffers. (Bottom) Overlay of 200 µM FADD-DD spectrum in 25 mM TRIS buffer, 150 mM NaCl, pH 6.2 (black) and 200 µM FADD-DD spectrum in 25 mM TRIS buffer, 150 mM NaCl with added 10 mM UDP and 10 mM $MgCl_2$, pH 6.2 (green). Spectra recorded at 800 MHz, 298 K.*

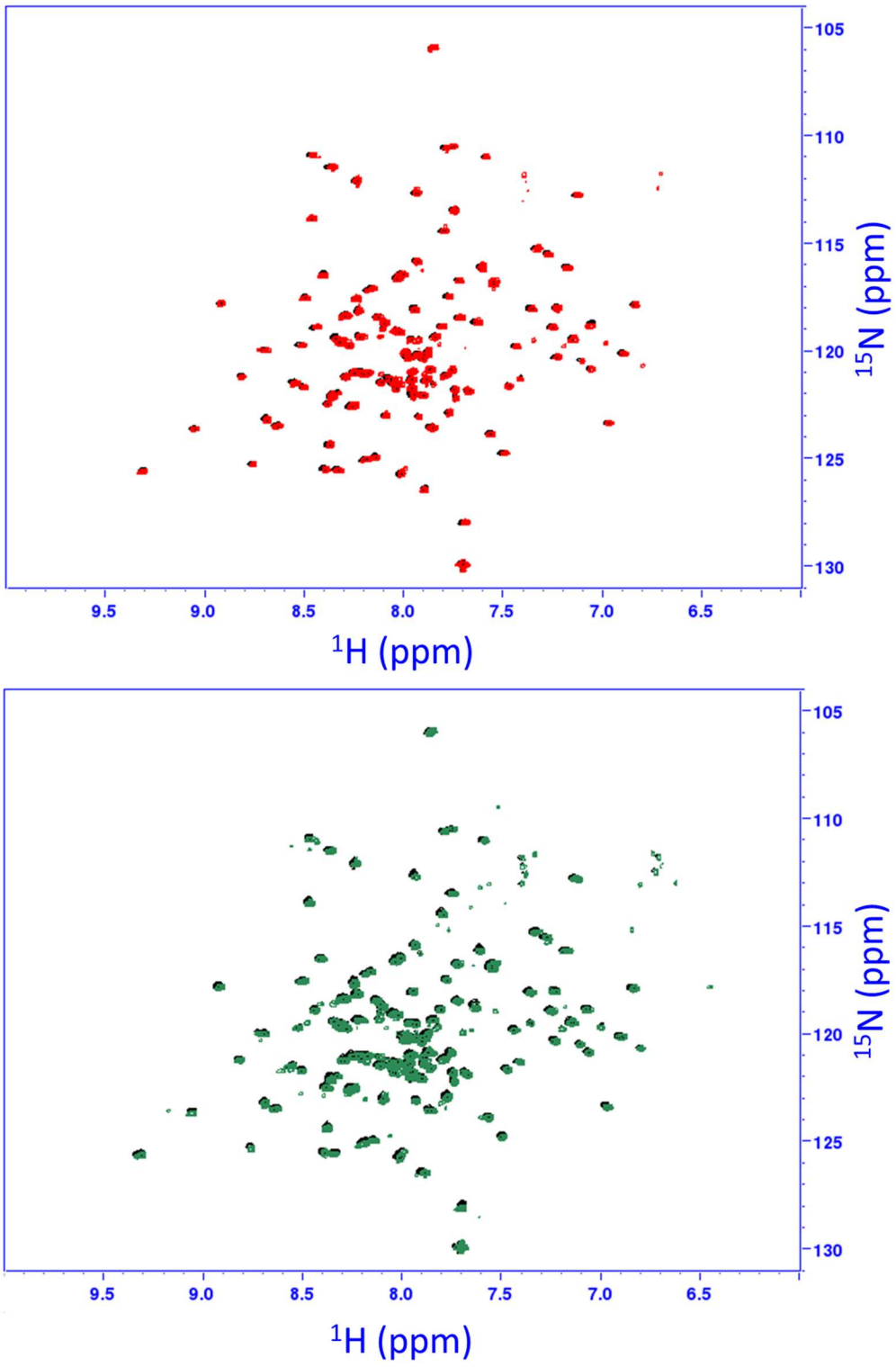



*Figure 3.22 - Comparison of BIS-TRIS and potassium phosphate buffer. (Top) Overlay of 200 µM $^{15}$N-FADD-DD spectrum in 50 mM potassium phosphate buffer, 150 mM NaCl, pH 6.2 (black) and 200 µM FADD-DD spectrum in 25 mM BIS-TRIS buffer, 150 mM NaCl, pH 6.2 (red) showing only very subtle chemical shift difference between the two buffers. (Bottom). Overlay of 200 µM FADD-DD spectrum in 25 mM BIS-TRIS buffer, 150 mM NaCl, pH 6.2 (black) and 200 µM FADD-DD spectrum in 25 mM BIS-TRIS buffer, 150 mM NaCl with added 10 mM UDP and 10 mM $MgCl_2$, pH 6.2 (dark green). Spectra recorded at 800 MHz, 298 K.*

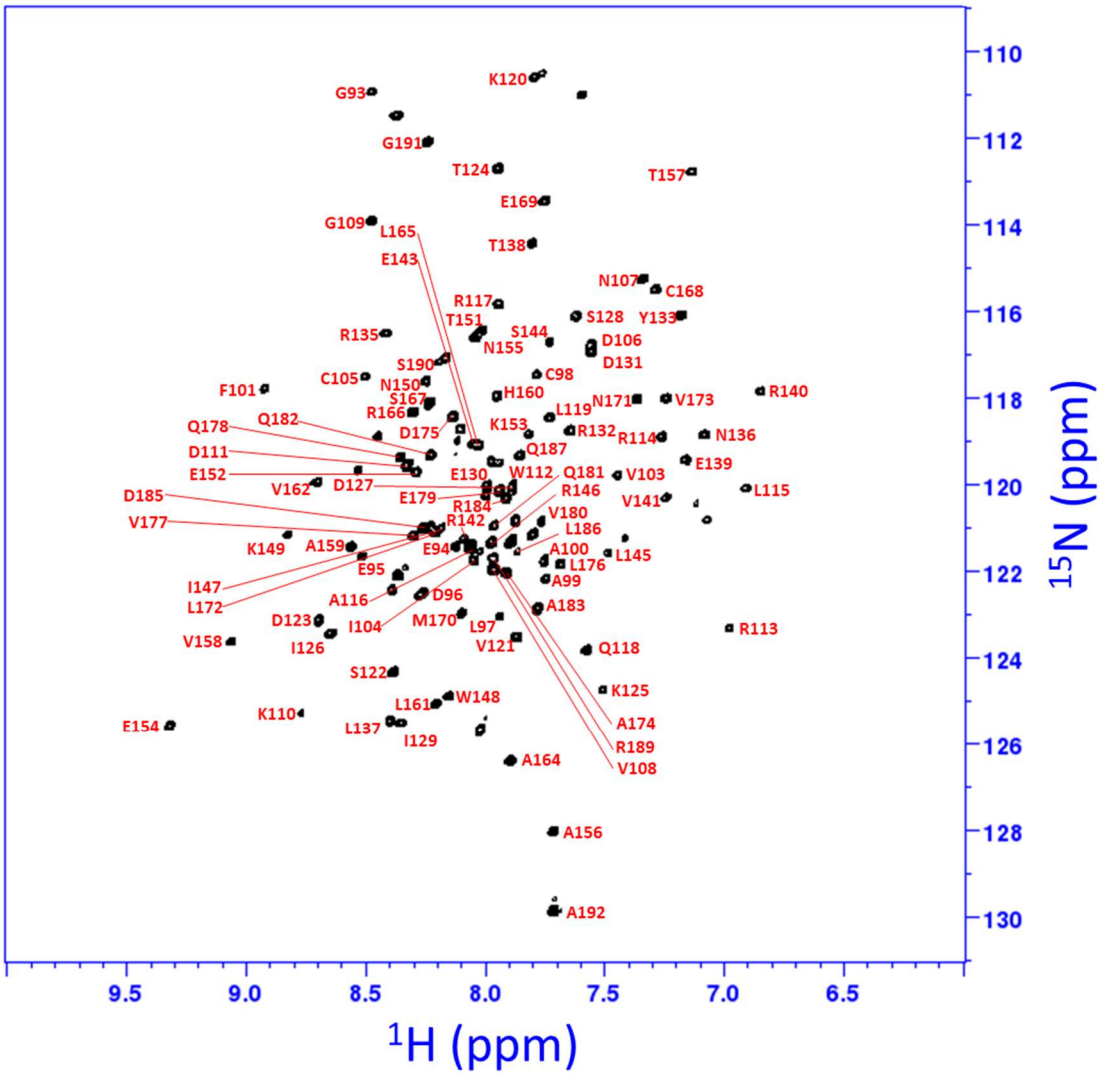



*Figure 3.23 - Assigned $^{1}H$-$^{15}N$ TROSY spectrum of $^{15}N$-FADD-DD in 25 mM BIS-TRIS buffer, 150 mM NaCl, pH 6.2, recorded at 800 MHz with a temperature of 298 K. The spectrum was recorded with a $^{15}N$-FADD-DD concentration of 200 µM at 800 MHz in a buffer composed of 25 mM BIS-TRIS, 150 mM NaCl, 1 mM DTT, 10% $D_2O$, pH 6.2.*

Using the BIS-TRIS buffer it was possible to record a high-quality spectrum of $^{15}N$-FADD-DD. Using the BIS-TRIS buffer there were also no changes in shifts on addition on $MgCl_2$ and UDP (Figure 3.22). Importantly, the recorded spectrum of $^{15}N$-FADD-DD in BIS-TRIS buffer is like the one recorded in phosphate buffer (Figure 3.22). This means that despite conditions being different, it is still possible to use the assigned literature reference spectrum recorded in phosphate buffer (55) to assign the peaks. With manual assignment, 97 of the 100 residues of the $^{15}N$-FADD-DD spectra were identified (Figure 3.23). These conditions were chosen for the titration experiments.

### 3.2.4.2 Titrations of SseK1-S286Y in the presence of FADD-DD

MD simulation showed that FADD-DD could interact with SseK1-S286Y in a similar manner as NleB1 with limited re-orientation of FADD-DD over the course of the simulation. There were though some differences in the complexes in the presence of UDP-GlcNAc and UDP. It is important to note however, that these results show how SseK1-S286Y can recognise FADD-DD in silico only. These NMR experiments seek to provide experimental evidence for these predictions, corroborating the proposed SseK1-S286Y:FADD-DD structures provided by MD simulation and determine which more accurately reflects the complex in solution.

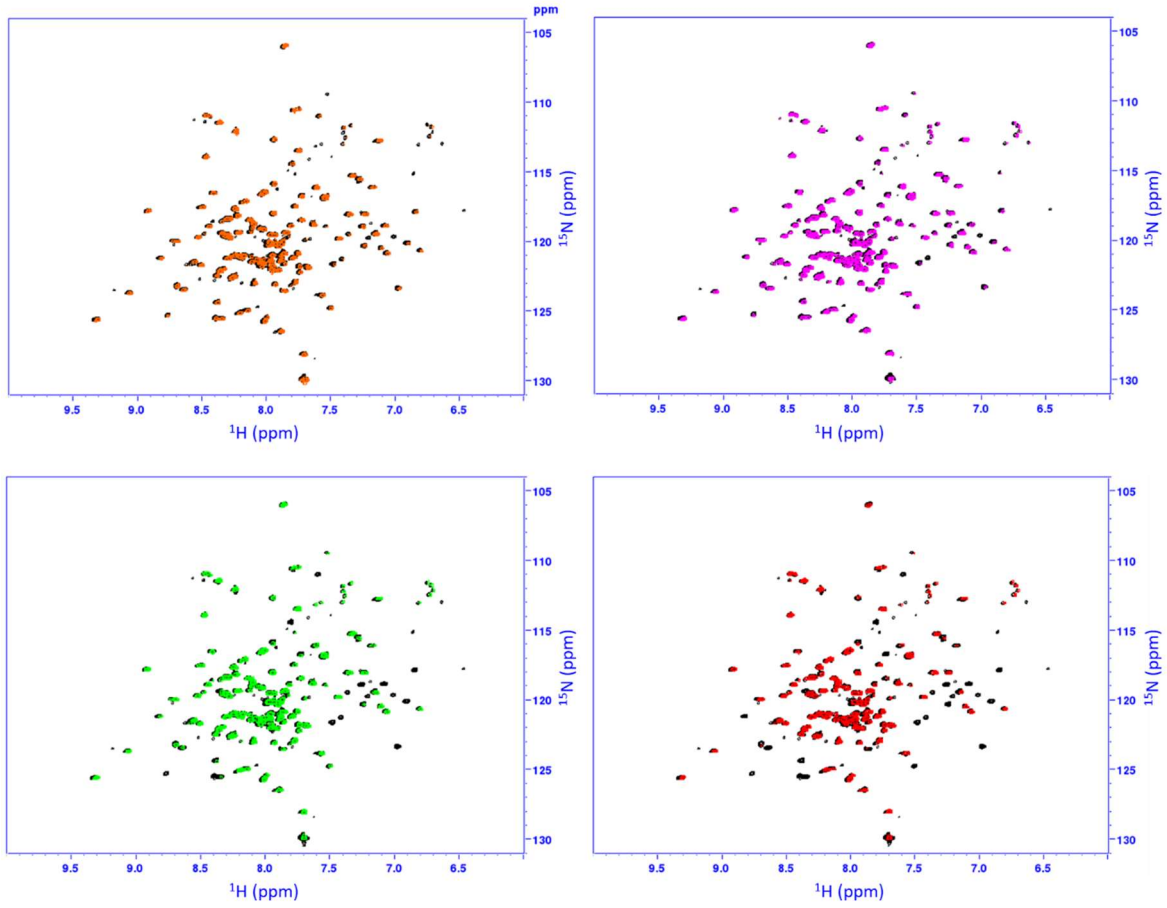



*Figure 3.24 - $^{1}$H-$^{15}$N TROSY titration spectra for SseK1-S286Y. Shown are spectra of 200 µM $^{15}$N-FADD-DD in 25 mM BIS-TRIS buffer, 150 mM NaCl, pH 6.2, recorded at 800 MHz with a temperature of 298 K, overlaying the different points along the titration. Overlays show 200 µM $^{15}$N-FADD-DD apo spectra (black) with 50 µM SseK1-S286Y (orange, top left), 100 µM SseK1-S286Y (pink, top right), 150 µM SseK1-S286Y (green, bottom left), 200 µM SseK1-S286Y (red, bottom right).*

Over the course of the titration, there is an observable pattern of perturbed signals in the TROSY spectra (Figure 3.24). As was the case with the HSQC experiments described during experimental optimisation, section 3.2.4.1, perturbations take the form of lowered signal intensity. There are very limited changes in peak chemical shifts, with the reductions in intensity correlating with increasing SseK1-S286Y concentration. One possible explanation for this could be that the macromolecule is large enough for there to be slow tumbling and more efficient $T_2$ relaxation (88). There is some evidence that this is partly responsible for the losses in signal intensity as there is a global loss in intensity with increasing SseK1-S286Y concentration. It is also possible that the exchange regime could be responsible for the loss in signal intensity. The system is unlikely to be in fast or fast-intermediate exchange as there is limited movement in peaks over the course of the titration. This leaves intermediate and slow exchange regimes, both of which could explain the signal loss over the titration (88). In slow exchange, loss in peak intensity typically would also happen alongside the appearance of a peak at a new chemical shift. In these experiments though there are no new peaks at later points in the titration. However, it is possible that these titrations are not sufficient to see those new peaks, with only the loss in peak intensity observed. During intermediate exchange there is a loss in signal intensity due to signal coalescence, which can mean that the signal is not seen in a new chemical shift even at later points in a titration. Within the context of these experiments however, it is not possible to claim which specific regime is responsible for the observations. Despite this the perturbations seen in these titrations are a strong indication of which residues are involved in contacts with FADD-DD and SseK1-S286Y. Using these perturbations, it is possible to map the contacts that the FADD-DD makes with SseK1-S286Y and provide experimental evidence for the proposed complex structures found by MD. Looking to the titration point with SseK1-S286Y concentration at 50 µM there are 35 FADD-DD resonances that are significantly perturbed (Figure 3.25). Resonances were selected from this titration point as they would represent those that form the closest contacts with SseK1-S286Y. These resonances mostly correlate to those found in the α2/α3/α4 helices of FADD-DD.

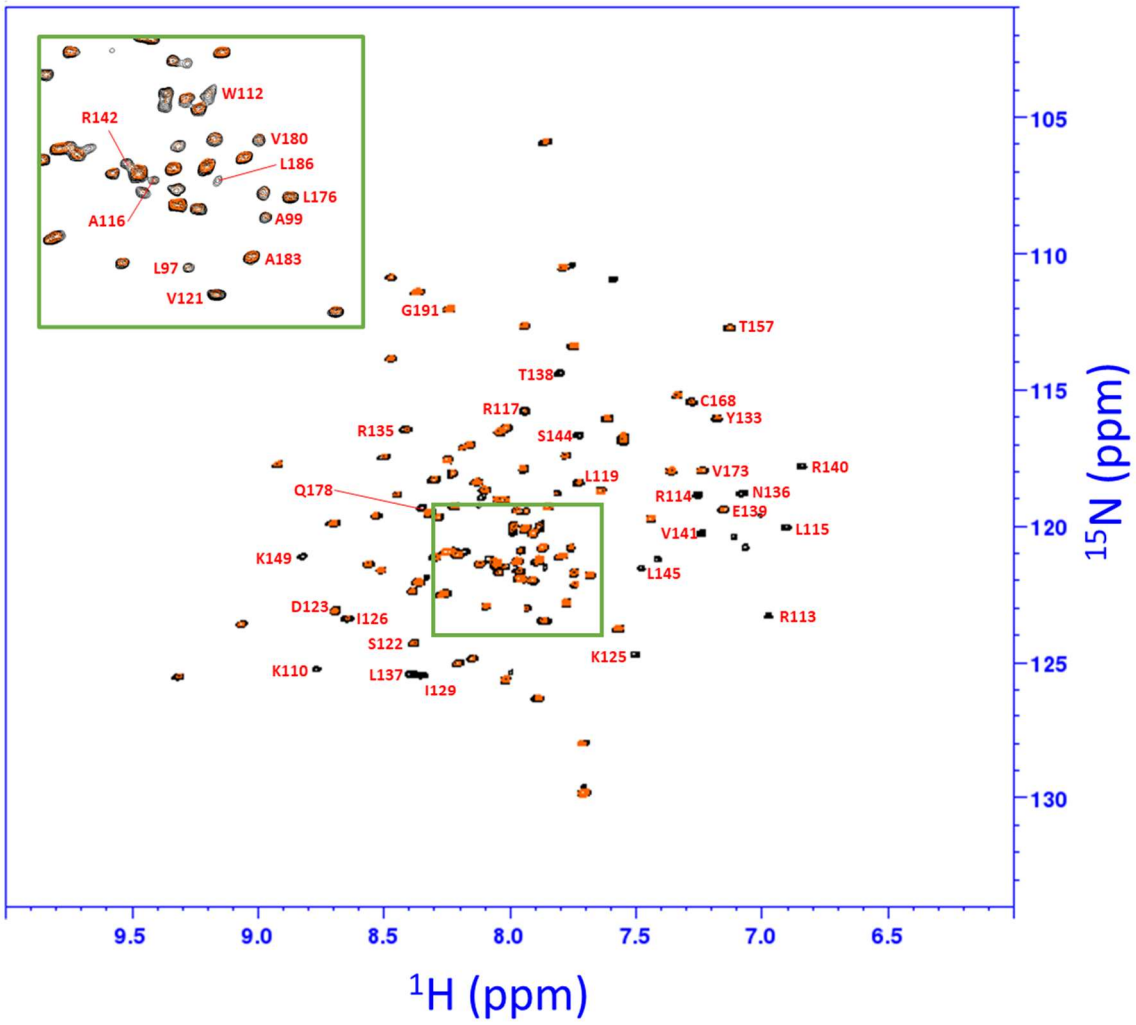



*Figure 3.25 - TROSY spectrum perturbations at 50 µM SseK1-S286Y. Overlay of apo $^{15}$N-FADD-DD spectrum (black) and the 50 µM SseK1-S286Y (orange) titration point. Both spectra measured in 25 mM BIS-TRIS buffer, 150 mM NaCl, pH 6.2, recorded at 800 MHz with a temperature of 298 K. Residues found to be significantly perturbed have been labelled, for clarity some resonances in the spectrum have been zoomed in (green box).*

Signal integrals were normalised against a resonance that is isolated and does not significantly lose intensity over the course of the titration (FADD-DD$^{G191}$) (Figure 3.26). This residue is far from the interaction interface between FADD-DD and SseK1-S286Y, near the C-terminus. For each point in the titration all the normalised signal integrals were then expressed as a percentage of their intensity in the apo spectrum. This was done to correct for the global reduction in intensity due to some residual T2 relaxation not cancelled out by the TROSY pulse sequence. For the 35 FADD-DD resonances selected it was possible to determine their concentration dependence and rank them by their reduction in signal intensity over the course of titration.

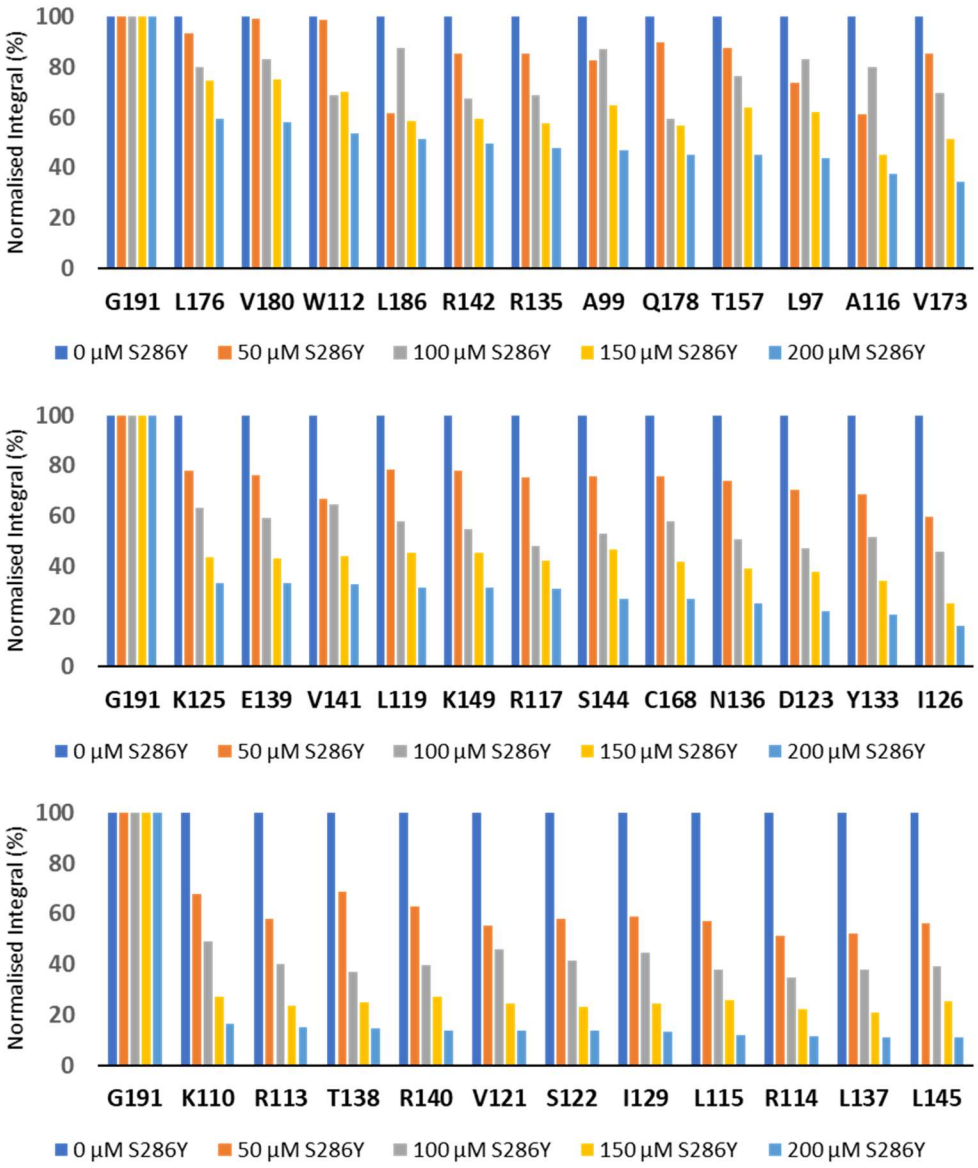



*Figure 3.26 - Normalised peak integrals for most perturbed TROSY peaks. Peaks were measured and all normalised against FADD-DD$^{G191}$, then expressed at a percentage of their intensity in the apo spectrum.*

These perturbations were then mapped onto the FADD-DD structure in complex with SseK1-S286Y. A global view of the FADD-DD shows strongly perturbed residues positioned in sites which correlate with predicted interactions. These sets of perturbed residues agree with interactions in the published NleB1:FADD-DD crystal structure, providing strong experimental evidence that the FADD-DD interacts in a comparable orientation to SseK1-S286Y as with NleB1. From the MD simulation, these observations are in more agreement with the UDP-GlcNAc-bound SseK1-S286Y:FADD-DD complex (Figure 3.27) rather than the simulation of the UDP-bound complex.

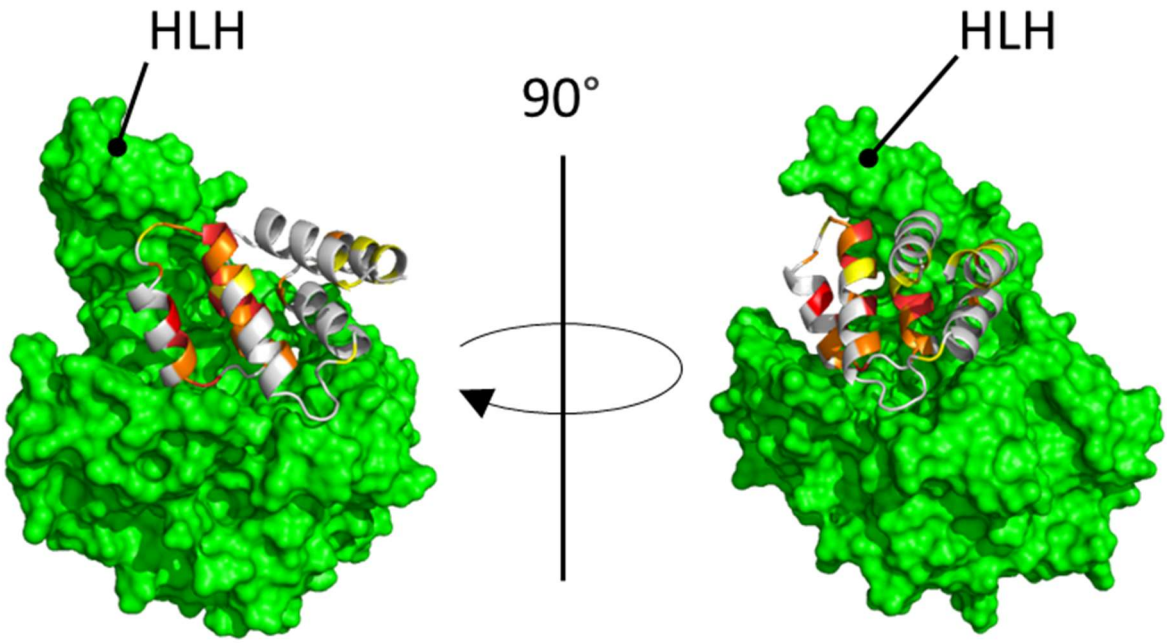



*Figure 3.27 - MD predicted UDP-GlcNAc-SseK1-S286Y-FADD-DD structure at 0.7 µs. In the cartoon representation of FADD-DD (grey) are residues which are identified to be perturbed in the 50 µM SseK1-S286Y titration point is highlighted. Shown in the structure are strong (red, >20% normalised integral at 200 µM SseK1-S286Y), medium (orange, 20% to 40% normalised integral at 200 µM SseK1-S286Y) and low (>40% normalised integral at 200 µM SseK1-S286Y).*

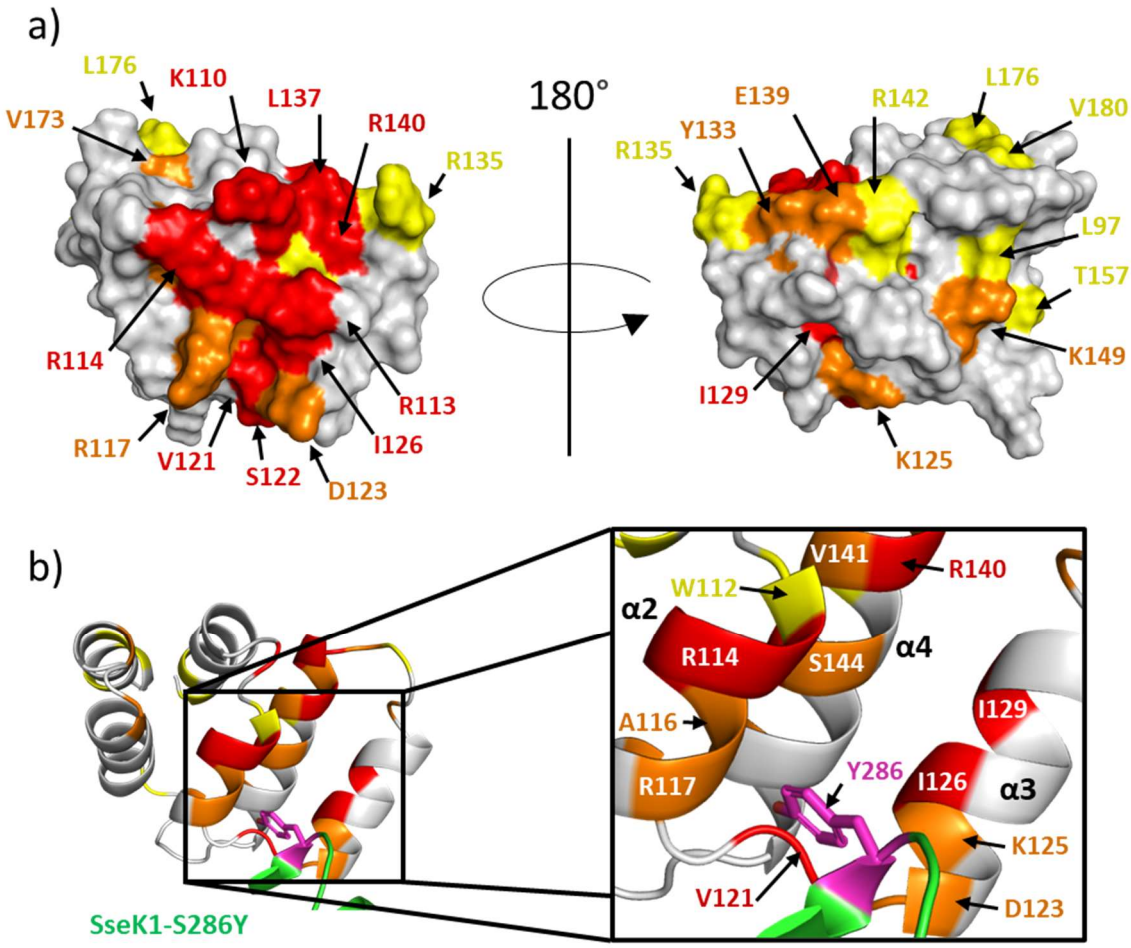



*Figure 3.28 - FADD-DD contacts highlighted by TROSY experiments. Surface representation (a) of FADD (grey), Perturbations are shown on the surface as strong (red, >20% normalised integral at 200 µM SseK1-S286Y), medium (orange, 20% to 40% normalised integral at 200 µM SseK1-S286Y) and weak (>40% normalised integral at 200 µM SseK1-S286Y). Molecular dynamics predicted structure of UDP-GlcNAc-bound SseK1-S286Y (green) in complex with FADD-DD (grey) at 0.7 µs highlighting the contacts SseK1-S286Y$^{Y286}$ (pink) make with the hydrophobic pocket of FADD-DD formed between the α2 and α3 helices (b).*

The main site where most perturbed residues are clustered around the hydrophobic pocket between the α2 and α3, where the mutated tyrosine SseK1-S286Y$^{Y286}$ is predicted to interact (Figure 3.28). Focussing on the triad of hydrophobic residues that were identified by MD simulation to form the hydrophobic pocket between α2 and α3 we can see that FADD-DD$^{V121}$ and FADD-DD$^{I126}$ are both strongly perturbed. In contrast the perturbation of FADD-DD$^{A116}$ is smaller (Figure 3.28). These observations agree with the MD simulations as the tyrosine residue interacts most closely with the isoleucine residue rather than the alanine and forms a backbone hydrogen bonding interaction the valine

residue. This, in combination with the overall distribution of strong perturbations, strongly suggests that interactions in this site are also important for recognition by SseK1-S286Y. Surrounding these three residues there are other strongly perturbed residues on the α2, α3 and α4 helices (Figure 3.28). These perturbations support the notion that the interactions between SseK1-S286Y$^{Y286}$ and this hydrophobic pocket are important for anchoring this SseK1-S286Y:FADD-DD complex. To accommodate the tyrosine residue there would have to be some subtle rearrangement of these helices, leading to the observed perturbations. In addition, there are perturbations seen in the FADD-DD structure, which are not explained by contacts with SseK1-S286Y, far from the interactions interface (Figure 3.28). These perturbations, however, are much weaker and can be due to some subtle global rearrangement of the FADD-DD structure on contact with SseK1-S286Y.

# CHAPTER 4

# MOLECULAR BASIS FOR THE DONOR SUBSTRATE RECOGNITION OF BACTERIAL VIRULENCE EFFECTORS

# 4.1 Introduction

Carbohydrates tend to form weak but highly specific interactions with proteins, greatly impacted by small changes in stereochemistry or functional groups. Sugar nucleotides carry these same characteristics. The NleB/SseK effectors use the sugar nucleotide UDP-GlcNAc to catalyse the transfer of a single GlcNAc ring to arginine residues in the acceptor protein. Importantly, UDP-GlcNAc is the main sugar nucleotide donor used by these effectors. The only exception to this is the unique case of NleB2 which can use UDP-Glc and UDP-Gal in addition to UDP-GlcNAc as donors (89). Despite this selectivity for catalysis, there is some evidence that suggest that the NleB/SseK effectors can bind other UDP sugar nucleotides but, with much lower binding affinity and no catalytic activity (36). This poses the important question of how these effectors can discern between different UDP sugar nucleotides and how this impacts their ability to participate in catalysis as donors. The scientific community has contributed to answering this question with the crystallisation of the NleB/SseK effectors (35-37). Though to date only one example with the complete UDP-GlcNAc donor has been published. The challenge for such studies is the weak interactions the GlcNAc ring has with the active site. This means it can be difficult to acquire sufficient electron density from the sugar ring to include it in a crystal structure of the enzyme-donor complex.

This chapter outlines the use of STD NMR and MD to study the binding of different sugar nucleotides to the NleB/SseK effectors. As the $K_d$ of the sugar nucleotides fall within the μM and mM ranges, STD NMR is a particularly useful tool in the study of these systems. To complement the STD NMR results and provide a structural rationale for the observations, MD simulations were used. Both methods have advantages over the X-ray crystallographic approaches used in the study of these systems as they observe the systems in solution, in a state closer to the natural one. As a crystal structure will typically represent the structure that can crystallise, they can often not account for dynamics which may impact ligand binding which would be observed in the binding epitope of the ligand from STD NMR. Furthermore, crystal structures can only represent a structure at a single point in time. MD simulations, on the other hand, with their interpretation informed by experimental data can show how a system develops over time. The use of both methods show that the effectors can bind other sugar nucleotide donors and lays out a rationale for why they may not participate in catalysis.

### 4.1.1 Goals

- To measure binding epitopes for UDP and the donor UDP-GlcNAc, as well as the donor-like ligands UDP-Glc, UDP-GalNAc and UDP-Gal in the presence of the SseK effectors.

- To determine if the single point mutation in SseK1 that triggered FADD activity impacts donor recognition.

- Evaluate the binding affinities of the different donors using VT STD NMR.

- Assess whether the use of Mg(II) is a suitable replacement for Mn(II) as the divalent cation for the NleB and SseK effectors.

- Generate a structural model for the binding of the different donors using MD simulations.

# 4.2 Results and Discussion

## 4.2.1 The Binding of UDP, UDP-GlcNAc and Donor-Like Ligands

Other than UDP-GlcNAc and UDP, the ligands used in this study included UDP-GalNAc, UDP-Gal and UDP-Glc. UDP-GalNAc was chosen to explore the importance of the stereochemistry of the sugar ring when binding to the effectors. UDP-Glc was chosen to explore the significance of the NAc group in binding and UDP-Gal combines both the change in stereochemistry and the lack of the NAc group. The following section will present the binding epitopes of these ligands to SseK1-WT, SseK1-S286Y and SseK2. In addition, to compare the affinities the enzymes have for these ligands a method called Variable Temperature STD NMR (VT STD NMR) will be presented and the importance of the divalent cation in recognition will be explored.

### 4.2.1.1 Epitope Mapping of the binding of Sugar Nucleotides

Epitopes were possible to measure for all sugar nucleotides in the presence of SseK1, SseK2 and SseK1-S286Y with a ligand to protein ratio of 20:1. In all the epitopes strongest STD intensities are seen for the uridine nucleoside, with the sugar rings showing weaker intensities. This is typical, as in previous STD NMR studies of sugar nucleotides in the context of other glycosyltransferases the most strongly recognised structure is the nucleoside (52). Focusing first on the natural donor of these effectors, UDP-GlcNAc, there is a clear pattern of interactions (Figure 4.1). As mentioned, the nucleoside is strongly recognised with the strongest intensities for UDP-GlcNAc being the H5 of the uracil and the anomeric proton of the ribose. In addition, there are strong contacts of H2 and H3 of the ribose. This pattern of interactions is consistent along all enzymes, though in the presence of SseK1-S286Y the strongest intensity is from the anomeric proton of the ribose rather that H5 of uracil. This difference in intensity is relatively small and is unlikely indicative of a different binding mode. These results agree with other structural studies which show the uracil is tightly associated with the active site, with the uracil sandwiched in a conserved π-π stacking interaction.

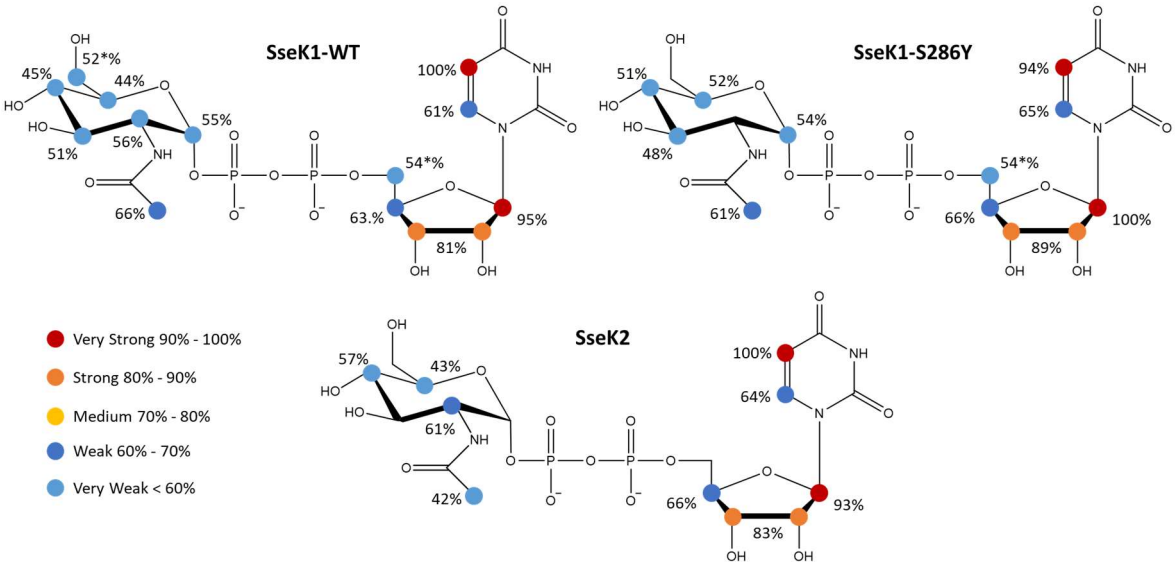



*Figure 4.1 - The binding epitopes for UDP-GlcNAc, measured in the presence of SseK1-WT, SseK1-S286Y and SseK2. STD intensities have been represented by coloured circles with STDs of 90% - 100% (Very Strong, red), 80% - 90% (Strong, orange), 70% - 80% (Medium, yellow), 60% - 70% (Weak, dark blue) and <60% (Very Weak, light blue).*

Moving to the sugar ring of UDP-GlcNAc, for the binding to both SseK1-WT and SseK1-S286Y the STD intensities are similar with there being only subtle differences in epitope. For these enzymes, the most important observation is that the acetyl group shows the highest STD intensity indicating that it does form an interaction that is important in recognition of the sugar ring. This observation is not the same in the case of UDP-GlcNAc binding to SseK2, where the epitope indicates weaker recognition of the acetyl group with the hydrogen at position 2 showing the strongest STD intensity.

For UDP a similar pattern can be seen for the nucleotide moiety as seen in UDP-GlcNAc. SseK1-WT and SseK1-S286Y show strong STD intensities for the H5 of the uracil and the ribose both of which having very similar overall epitopes (Figure 4.2). The epitope of UDP binding to SseK2, however, again is different when compared the epitopes for SseK1-WT and SseK1-S286Y (Figure 4.2). This epitope places somewhat weaker intensities for the uracil with a much stronger emphasis placed on the ribose of the nucleoside. These could be indicative of slightly different overall binding mode, suggesting that UDP is differently recognised by SseK2.

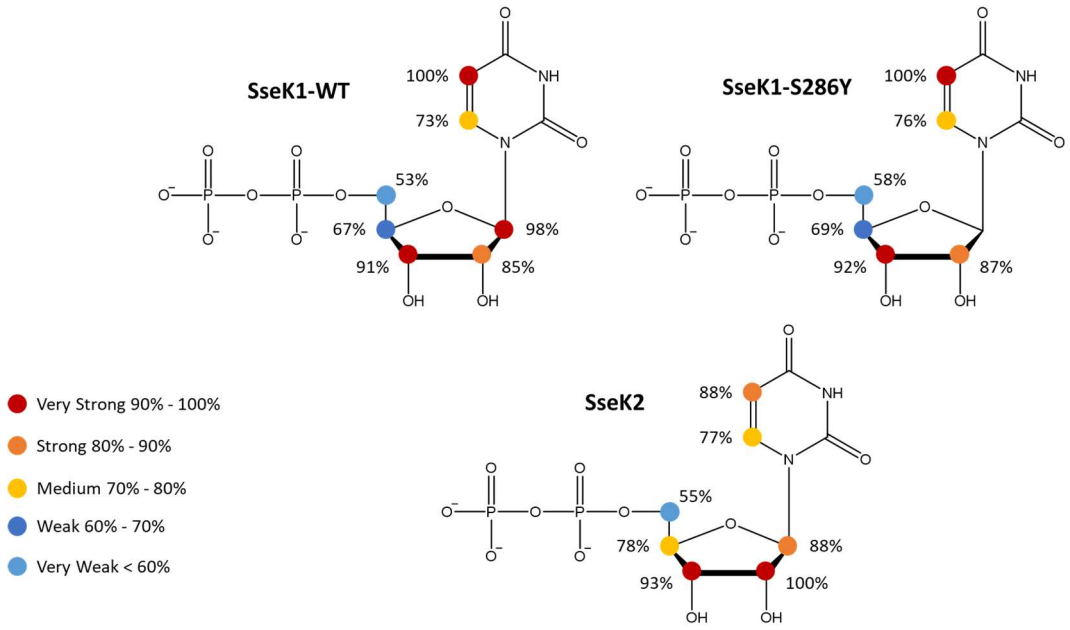



*Figure 4.2 - The binding epitopes for UDP, measured in the presence of SseK1-WT, SseK1-S286Y and SseK2. STD intensities have been represented by coloured circles with STDs of 90% - 100% (Very Strong, red), 80% - 90% (Strong, orange), 70% - 80% (Medium, yellow), 60% - 70% (Weak, dark blue) and <60% (Very Weak, light blue).*

The results for UDP-GlcNAc and UDP highlight two important points related to the SseK effectors. The first is that there is some difference in the interactions for both UDP and UDP-GlcNAc between the SseK1 effectors and SseK2. This observation could explain why there has been success in producing a UDP-GlcNAc-bound crystal structure for SseK2 and not SseK1-WT. As SseK2-WT shows a somewhat stronger recognition of the GlcNAc ring itself compared to SseK1-WT, perhaps because it is more stable in the active site. Secondly, the inclusion of the S286Y mutation does not significantly impact the recognition of either UDP-GlcNAc or UDP. This implies that though the mutation can trigger FADD activity in SseK1 the mechanism underpinning this change is likely related to changes in FADD recognition only and does not alter the recognition of UDP-GlcNAc such that catalysis more readily takes place.

To explore how these enzymes can discern between different donors, STD NMR experiments were performed on three other sugar nucleotide donors including UDP-Glc, UDP-Gal and UDP-GalNAc. The first important result of these experiments is that the three enzymes used in this study were all able to recognise each of these donors. One study measured the $K_d$ of UDP-Glc and UDP-Gal to SseK1-WT, finding values of 38 µM and 159

µM, respectively, where the $K_d$ of UDP-GlcNAc was measured to be 2.3 µM for SseK1-WT and 1.2 µM for SseK2-WT. These relatively low affinities however are ideal for measurement by STD NMR. In this work it was possible to measure STD epitopes for these two ligands as well as UDP-GalNAc.

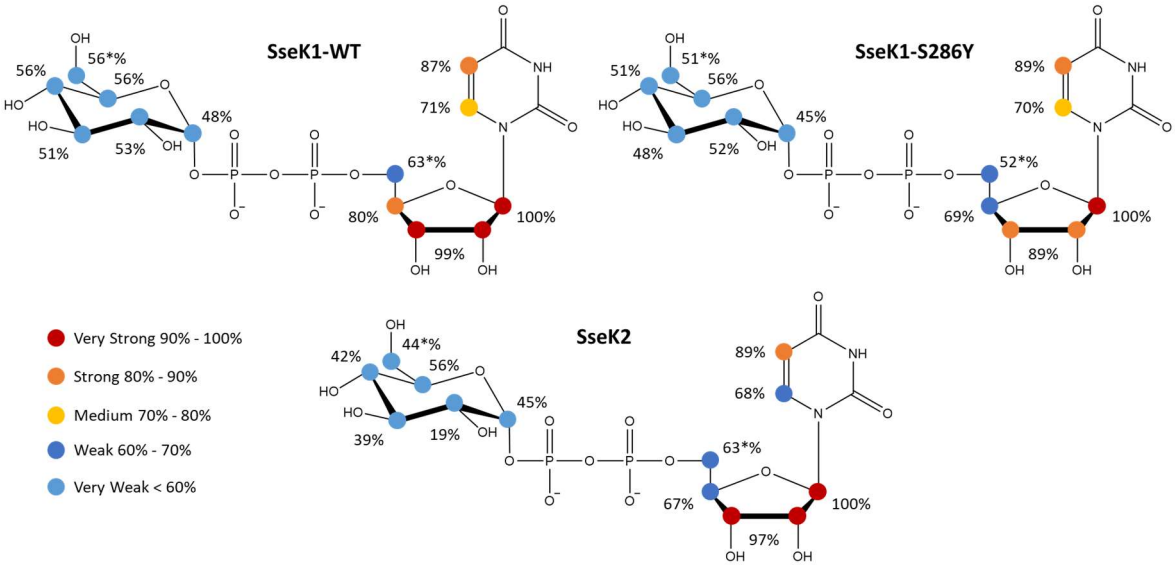



*Figure 4.3 - The binding epitopes for UDP-Glc, measured in the presence of SseK1-WT, SseK1-S286Y and SseK2. STD intensities have been represented by coloured circles with STDs of 90% - 100% (Very Strong, red), 80% - 90% (Strong, orange), 70% - 80% (Medium, yellow), 60% - 70% (Weak, dark blue) and <60% (Very Weak, light blue). Asterisks (*) indicate average STD intensities.*

For UDP-Glc, the loss of the NAc group does appear to significantly impact the binding epitope of the ligand compared to UDP-GlcNAc (Figure 4.3). An important note however is that there is a reduced STD intensity present on H5 of the uridine. Furthermore, this observation is consistent in all three of the enzymes studied. Moving to the Glc ring, for both SseK1-WT and SseK1-S286Y, STD intensities are comparable to those seen in UDP-GlcNAc. This implies that he Glc ring can bind in a similar manner as GlcNAc, despite the lack of the NAc group. This is not true, however, in the case of SseK2-WT, where the high intensity seen on position 2 of the GlcNAc ring is not carried over to UDP-Glc. This observation is interesting as while the SseK2-WT-bound UDP-GlcNAc epitope suggest a reduced role for NAc group in recognition, compared to the SseK1-WT and SseK1-S286Y epitopes, here its loss has impacted the epitope and potential binding mode of the ring.

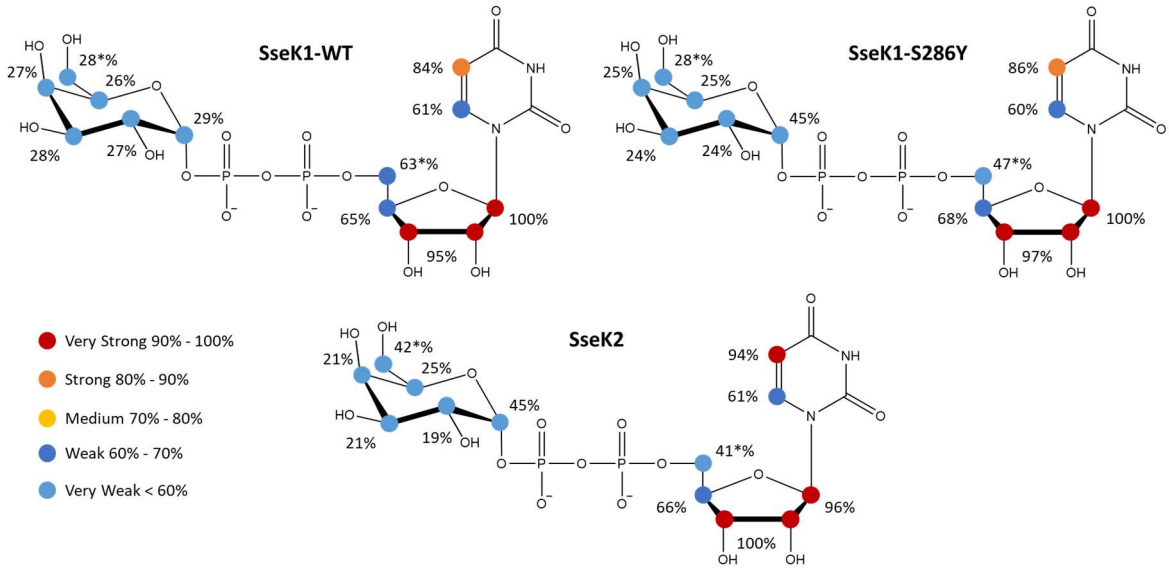



*Figure 4.4 - The binding epitopes for UDP-Gal, measured in the presence of SseK1-WT, SseK1-S286Y and SseK2. STD intensities have been represented by coloured circles with STDs of 90% - 100% (Very Strong, red), 80% - 90% (Strong, orange), 70% - 80% (Medium, yellow), 60% - 70% (Weak, dark blue) and <60% (Very Weak, light blue). Asterisks (*) indicate average STD intensities.*

UDP-Gal is more different than UDP-GlcNAc, lacking the NAc group and having the alternate axial stereochemistry at position 4 of the sugar ring. These two changes both can be seen to impact the binding epitopes of the sugar ring (Figure 4.4). In all cases there were significantly lower STD intensities compared to those seen in UDP-GlcNAc and UDP-Glc. This indicates that the change in stereochemistry has changed the binding mode of the sugar ring. In the published crystal structure of SseK2 in complex with UDP-GlcNAc, the hydroxyl groups form specific spatial interactions in the active site. The change in stereochemistry in position 4 has two effects on recognition. There can be steric clash which precludes the ring being properly placed in the active site and the lack of an equatorial hydrogen bonding interaction can reduce the number of interactions that can stabilize the ring. In the case of SseK1-WT and SseK1-S286Y, similarly to UDP-Glc, the presence of the Gal ring appears to have impacted the recognition of the uridine moiety with a lower STD intensity on the H5 of the uracil moiety. This is not the same as in SseK2-WT where the typical strong uracil H5 STD% is shown.

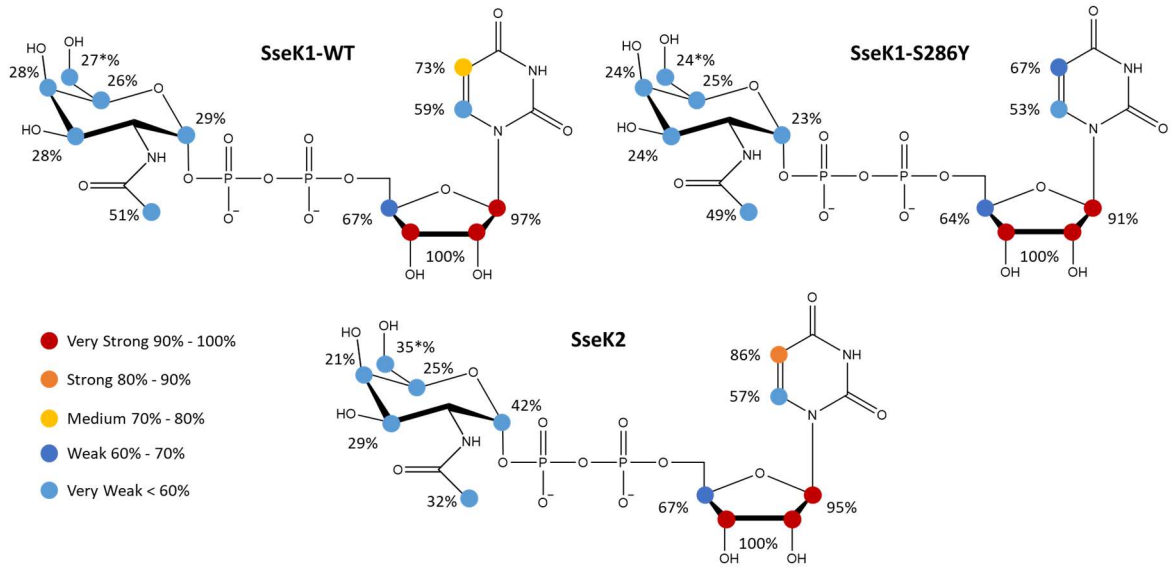



*Figure 4.5 - The binding epitopes for UDP-GalNAc, measured in the presence of SseK1-WT, SseK1-S286Y and SseK2. STD intensities have been represented by coloured circles with STDs of 90% - 100% (Very Strong, red), 80% - 90% (Strong, orange), 70% - 80% (Medium, yellow), 60% - 70% (Weak, dark blue) and <60% (Very Weak, light blue). Asterisks (*) indicate average STD intensities.*

The binding epitopes of UDP-GalNAc show a similar pattern to those of UDP-Gal. Again, there is a change in the STD epitope for the uracil which is lessened in the case of binding to SseK2-WT (Figure 4.5). This alongside the epitopes for UDP-Gal could suggest that the uridine binding site of SseK2-WT is more tolerant to changes in the sugar ring compared to SseK1-WT and SseK1-S286Y. There are differences seen in the π-stacking interactions between SseK1-WT and SseK2, with SseK1-WT using a tryptophan residue from the lid domain in place of a phenylalanine from the catalytic domain used in SseK2. These results may imply that the π-π stacking interactions offered by SseK2 are more robust and tolerant of substitution at the GlcNAc ring. Moving to the GalNAc ring, for all enzymes it is poorly recognised. The NAc group is most strongly recognised in the case of SseK1-WT and SseK1-S286Y, but less so in the case of SseK2 which is consistent with the observations from the epitopes for UDP-GlcNAc.

### 4.2.1.2 Variable Temperature STD NMR

In the study of protein-ligand binding, STD NMR can be a helpful tool in understanding the affinity of ligand to a protein. In fact it is possible with STD NMR to measure $K_d$ (52) . STD NMR is particularly sensitive to the $K_d$ of a ligand which in some contexts can be considered a limitation of the approach. This sensitivity, however, can be used to probe the $K_d$ of ligands in particular systems. The dissociation constants of the uridine sugar nucleotides to SseK1-WT, SseK1-S286Y and SseK2 have not all been measured. From those that have been studied (36, 89) it is possible to estimate that they cover the low µM to low mM range. This places them firmly within the optimal range for study by STD NMR. Being a thermodynamic equilibrium constant, $K_d$ is temperature dependent, so it is possible to modulate it by adjusting sample temperature. In STD NMR experiments to measure binding epitopes, this feature is often taken advantage of during experimental optimization to enhance the intensity of measured STD percentages. For a given dissociation constant, for example one that is very low µM to high nM, just outside the optimal range for STD NMR, increasing temperature can increase $K_d$. Though there will be some changes to $k_{on}$ in STD NMR this is assumed to be diffusion dependent. When temperature is increased the residence time a ligand can have in the active site is reduced and therefore $k_{off}$ is increased. This will mean that the ligand will undergo less $T_2$ relaxation as part of the macromolecule, as well as more saturated ligand will be accumulated during the saturation time so therefore the saturation observed in the bulk will be increased. For ligands at the upper limit of the optimal $K_d$ range, reduction in residence time will mean that less saturation will be transferred to the ligand by spin diffusion and therefore less saturation will be observed in the bulk. The method used in this section, Variable-Temperature STD NMR (VT STD NMR), takes advantage of this effect and here is used to assess the affinities of UDP, UDP-GlcNAc, UDP-Glc, UDP-GalNAc and UDP-Gal to SseK1-WT, SseK1-S286Y and SseK2.

For UDP-GlcNAc and UDP, when temperature is increased from 288 K to 298 K there is an increase in STD% (Figure 4.6). This increase, however, is not uniform between the different effectors (Figure 4.6). Though both SseK1-WT and SseK1-S286Y have similar values for STD% at 288 K, at 298 K there are subtle differences. For both UDP-GlcNAc UDP the increase is most pronounced in SseK1-WT compared to SseK1-S286Y. SseK2 shows overall much lower STD% values at 288 K and 298 K compared to SseK1-WT and SseK1-

S286Y. This data implies that there may be subtle differences in the affinity of UDP-GlcNAc to SseK1-WT and SseK1-S286Y, while there is a larger difference affinity for SseK2. For UDP-GlcNAc the latter is corroborated by the literature which presents the $K_d$ of UDP-GlcNAc to SseK1-WT and SseK2 as 2.3 µM and 1.2 µM respectively. The binding affinities for UDP to SseK1-WT and SseK2-WT have not been reported but this data indicates that they may bind in a similar $K_d$ range as UDP-GlcNAc, with the highest affinity being for SseK2.

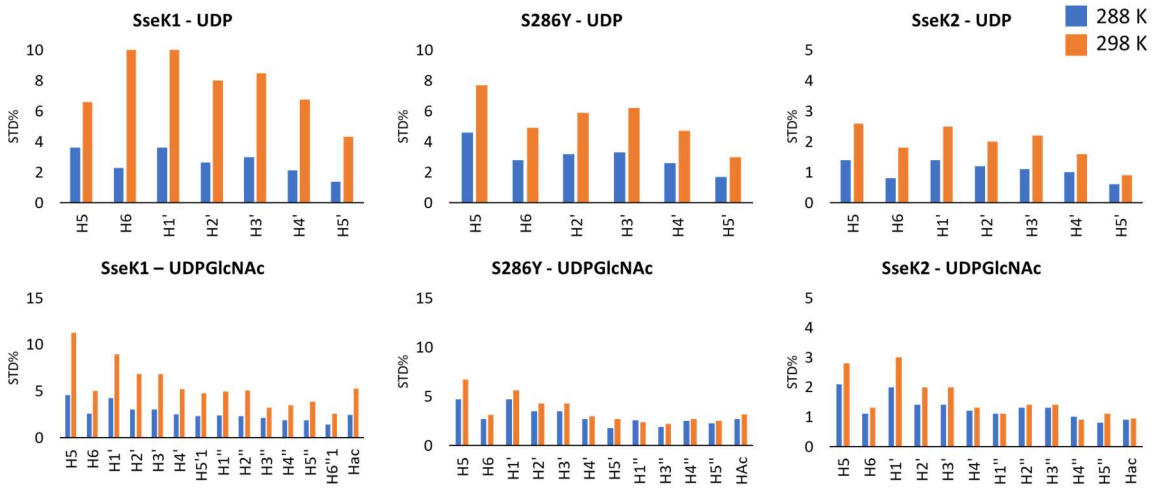



*Figure 4.6 - VT STD intensities for UDP and UDP-GlcNAc. Intensities measured in the presence of SseK1-WT, SseK1-S286Y and SseK2 using 2s saturation time. STD intensities were measured at 288K (blue bars) and 298K (orange bars). For both these ligands, when temperature is increased an increase in measured STD intensity is observed.*

In the case of UDP-Glc, UDP-Gal and UDP-GalNAc there is a decrease in STD% when the temperature is increased from 288K to 298 K (Figure 4.7). At 288 K some of these donors do show similar STD% values to UDP and UDP-GlcNAc. However, the decrease in these values at 298 K implies that their $K_d$ values are more towards the upper optimal bound for STD NMR. This means that these donors have lower affinity than UDP and UDP-GlcNAc. This observation agrees in the case of SseK1-WT where UDP-Glc and UDP-Gal both have much lower values for $K_d$ than UDP-GlcNAc (36). Comparing values for SseK1-WT and SseK1-S286Y, they are similar except in the case of UDP-Glc where values for SseK1-S286Y are more comparable to SseK2 than SseK1-WT.

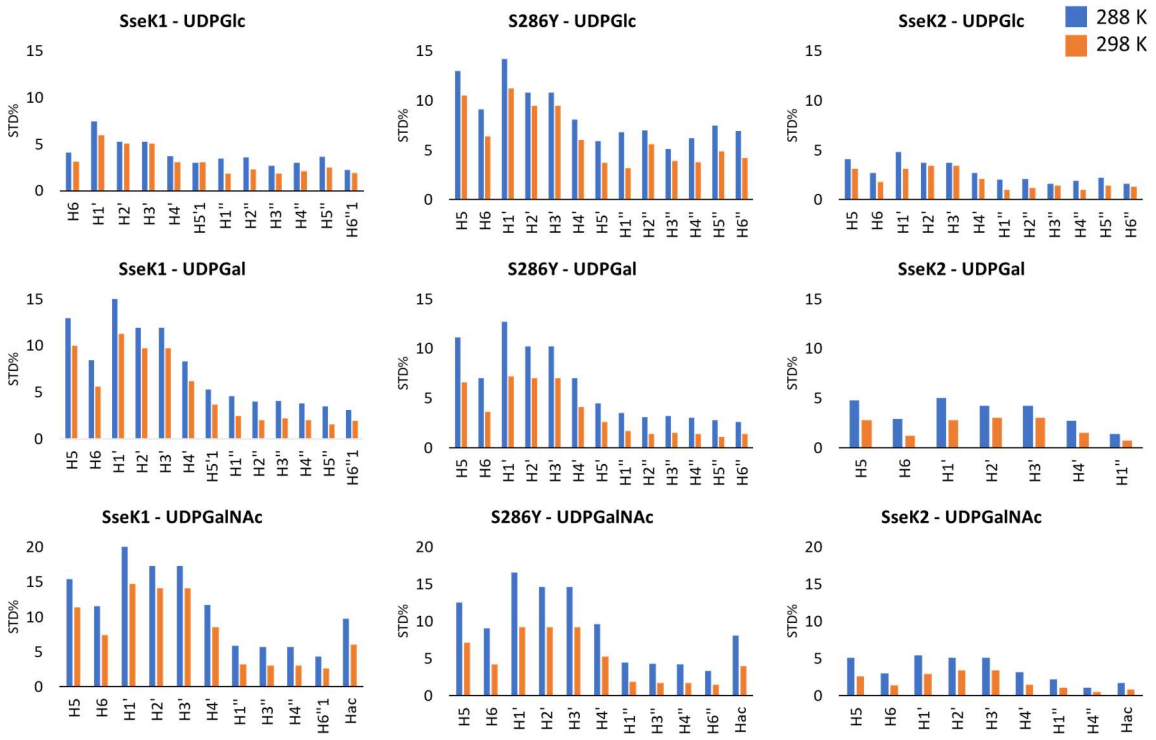



*Figure 4.7 - VT STD intensities for non-UDP-GlcNAc sugar nucleotides. STD intensities measured for UDP-Glc, UDP-Gal and UDP-GalNAc in the presence of SseK1-WT, SseK1-S286Y and SseK2 using 2s saturation time. STD intensities were measured at 288K (blue bars) and 298 K (orange bars). For both these ligands, when temperature is increased a decrease in measured STD intensity is observed.*

For UDP-Glc, UDP-Gal and UDP-GalNAc there is a consistent change in STD% intensities when the temperature is increased to 298 K. Looking to the STD intensities for UDP-Glc there are much lower relative changes in STD% compared to the UDP-Gal and UDP-GalNAc. The $K_d$ of UDP-Glc for SseK1-WT is lower than UDP-Gal and it would be reasonable to estimate that the same is true for UDP-GalNAc. This data suggests that these effectors generally have an affinity for UDP-Glc worse than for UDP of UDP-GlcNAc but much better than for UDP-Gal and UDP-GalNAc.

### 4.2.1.3 The effect of Mg(II) concentration on Binding Epitopes

In this STD NMR study of SseK1-WT, SseK1-S286Y and SseK2 the decision was made to use Mg(II) as the divalent ion for these effectors, in place of the natural ion Mn(II). Mg(II) was chosen as it is known that GT-A folded glycosyltransferases that include the DXD motif can accommodate this divalent ion (23). The reason that Mn(II) was avoided is that it is paramagnetic and would therefore hinder the recorded STD signal by inducing paramagnetic $T_2$ relaxation. To assess this decision and determine whether the presence of the Mg(II) could impact the binding of the sugar nucleotides used in this study a series of STD NMR experiments were performed with varied Mg(II) concentration. The purpose of these experiments was to establish if this ion could be taken up by these effectors and whether the presence of the ion impacted the binding of these ligands. The three concentrations chosen for this study included 0 µM, 25 µM (1:1 enzyme: Mg(II)) and 250 µM (1:10 enzyme:Mg(II)).

The data generated by these experiments indicate that the enzymes studies can utilize Mg(II). There is a reduction in $STD_{max}$ with increasing Mg(II) concentration (Appendix 6.4.3.1, Table 6.7 to 6.21). This indicates that there may be an increase in binding affinity with an increase in Mg(II) concentration. With the Mg(II) ion present there will be more favourable interactions offered in the active site for the pyrophosphate group, improving ligand binding. This will mean that the ligand will have an increased residence time in the active site and therefore the STD effect would be reduced due to $T_2$ relaxation as the ligand is part of the macromolecule. The result of this therefore being a reduction in $STD_{max}$ values as there will be less saturation built up at the steady state. The observations outlined by these experiments are also in agreement with the literature in which it was indicated that the affinity of the sugar nucleotides is metal ion dependent (37).

Looking to the normalized STD% values for the epitopes (Figure 4.8) contrary to the $STD_{max}$ values the epitopes do not significantly change with increasing Mg(II) concentration. This implies that the mode of recognition in the binding site is the same as when a divalent ion is or, is not present. This could be an unexpected result as in the active site it could be expected that there would be repulsive forces between the DXD and the phosphate group which could affect binding mode of the sugar nucleotides. These results alone could also indicate that due to the lack in change of binding epitope with increasing Mg(II)

concentration that this ion is not taken up by these enzymes and therefore will not impact binding epitope. However, as has been described there is a change in $STD_{max}$ which suggest that the ion is taken up. A likely explanation for this is that there are positively charged ions accommodated in the active site when Mg(II) is not present. This could be hydronium ions, though given that the NaCl is in a large excess from the buffer (150 mM) it could be more likely $Na^+$ ions are being coordinated in the active site.

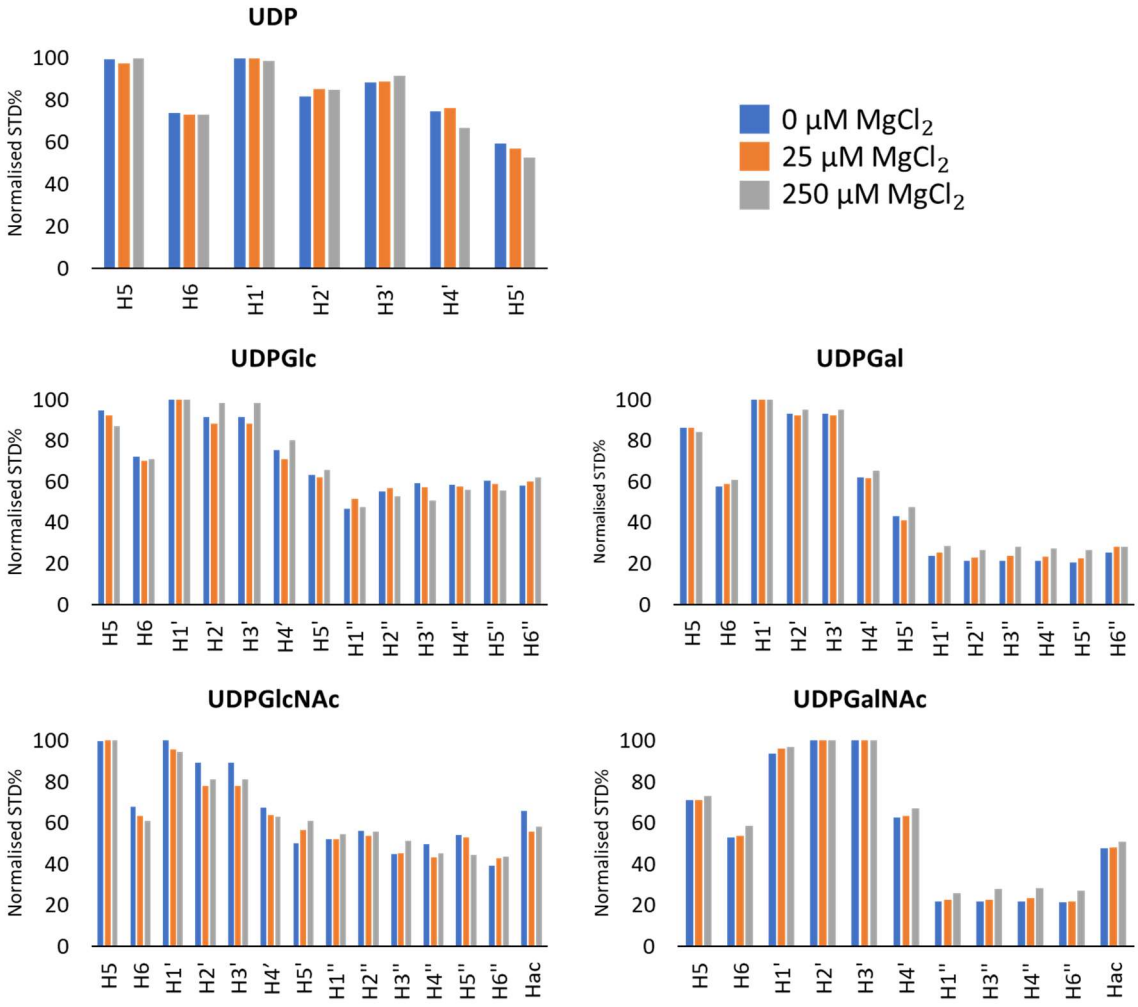



*Figure 4.8 - Impact of Mg(II) concentration on binding epitope. Normalised STD intensities of UDP, UDP-Glc, UDP-Gal, UDP-GlcNAc and UDP-GalNAc in the presence of SseK1-WT at concentrations of $MgCl_2$ including 0 µM (blue bars), 25 µM (orange bars) and 250 µM (grey bars).*

### 4.2.2 MD Simulation of the Binding of Sugar Nucleotide Ligands

To provide a structural model of how UDP, UDP-GlcNAc, UDP-Glc, UDP-GalNAc and UDP-Gal are recognised by the NleB and SseK effectors, 500 ns MD simulations were performed. The STD NMR results show for SseK1-S286Y, SseK1-WT and SseK2, excluding subtle differences in epitopes, that these ligands are largely recognised in a similar manner. This poses the question of whether these subtle differences in epitope are related to different binding modes or are related to the influence that affinity has on the STD intensities measured for these ligands. These simulations aim to answer this question and explore if UDP-Glc, UDP-Gal and UDP-GalNAc can be accommodated in the active site in a similar manner to UDP-GlcNAc.

An important point to keep in mind is that while the nucleosides of the sugar nucleotides are strongly recognised by GTs, the sugar ring often only participates in weak interactions in the active site. It is known that, in the case of SseK1-WT, the affinity of UDP-GlcNAc is highest compared to UDP-Glc and UDP-Gal (36). As these molecules all have similar structure there will be limited differences in entropic contributions for binding, leaving just enthalpic contributions. The only difference between UDP-GlcNAc and the other donors included in this study are the sugar rings. This means that these enthalpic differences are due to interactions the sugar rings have in the active site. Given that the GlcNAc ring likely interacts weakly with the active site, that would mean that the sugar rings of the other donors would interact even more weakly. The consequence of these weakened interactions is that the sugar rings of UDP-Glc, UDP-Gal and UDP-GalNAc can be quite mobile. Within the 500 ns simulation it is possible for this mobility to be properly sampled. However, this simulation time would not be long enough to allow for there to be large changes in protein conformation. This means that within these simulations only small local protein conformations can be sampled, such as those for flexible loops.

### 4.2.2.1 RMSD and RMSF calculations

To gain an insight into the overall stability and guide analysis of the simulations, RMSD and RMSF calculations were performed. RMSD calculations were performed on the backbone atoms of both the enzyme and the ligand in the complexes, compared to the first frame, to measure system stability. To understand how stable each ligand is in the active site, RMSF calculations of non-hydrogen atoms were performed. These calculations provide an understanding of how much motion each ligand undergoes over the length of the simulation.

#### 4.2.2.1.1 RMSD calculations

Overall, the RMSD calculations show that the systems were reasonably stable over the length of the simulations, though with some notable exceptions. In contrast to the simulations presented in Chapter 3, these simulations were performed in the absence of acceptor substrate. Because of this, there is more freedom for the dynamic regions such as the HLH and lid domain to undergo motion. The motion of these regions can in part explain the disparity in RMSD calculations between the different systems. Importantly, despite some systems showing increased RMSD values, these were not related to unfolding of the protein structures.

For the UDP-bound systems, the RMSD calculations for SseK1-WT show that after 0.35 μs there is an increase in instability. Close inspection of the simulation from this time point shows that there is improper coordination of the pyrophosphate group leading to a change of conformation in the active site. Because of this, analysis of this system was performed from 0 μs to 0.35 μs. The UDP-GlcNAc systems were stable over the length of the simulation. This can be explained as this ligand is the natural donor for these enzymes. In the active site, the complementary interactions offered by UDP-GlcNAc stabilises the lid domain which without such interactions can move more freely (36).

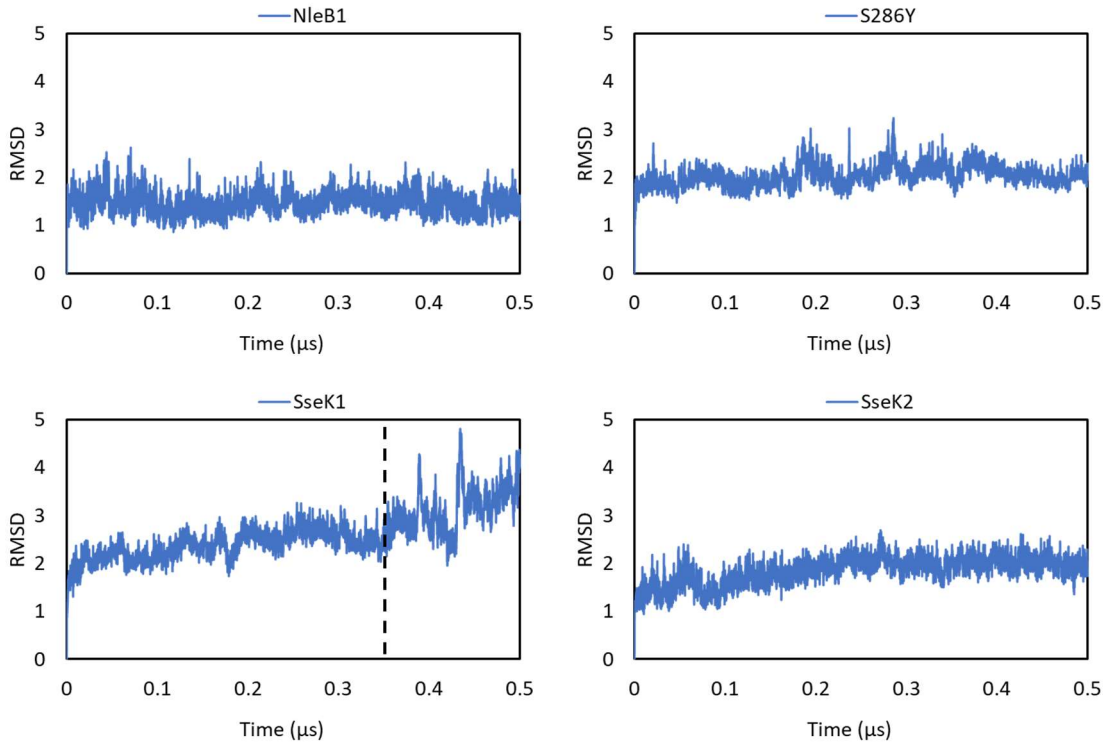



*Figure 4.9 - RMSD (Å) calculations for UDP-bound complexes. RMSD plots shown for NleB1, SseK1-S286Y, SseK1-WT and SseK2. Due to improper pyrophosphate coordination the SseK1-WT system was not analysed past 0.35 µs (black dashed line).*

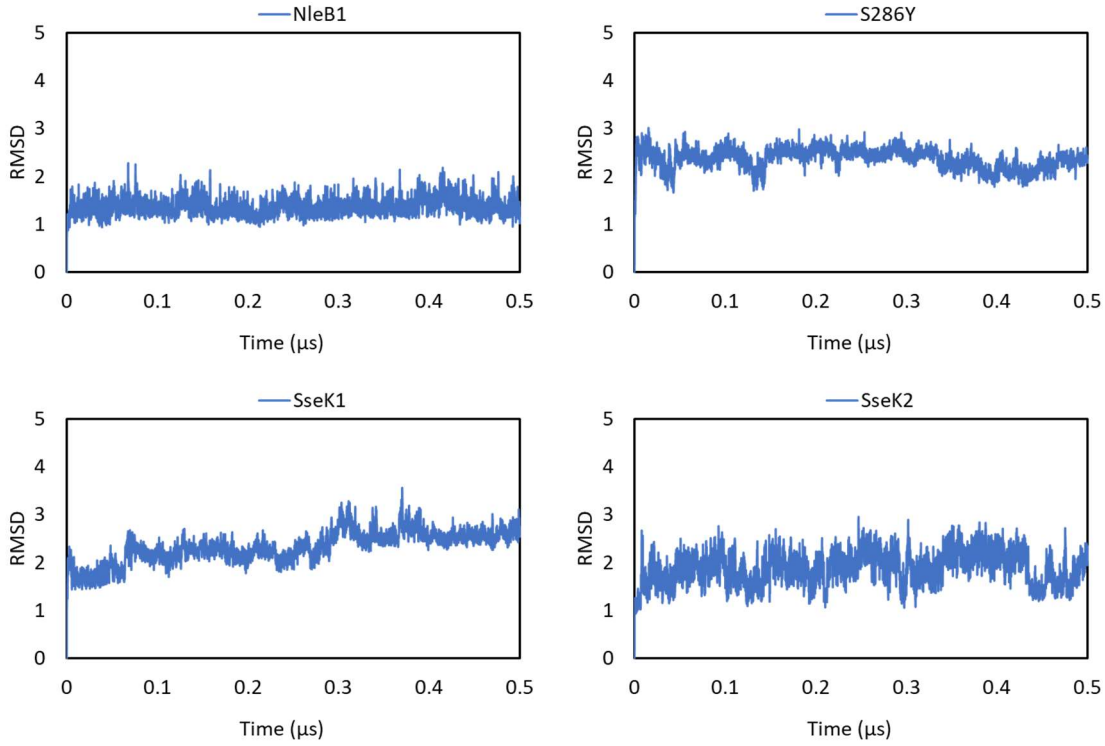



*Figure 4.10 - RMSD (Å) calculations for UDP-GlcNAc-bound complexes. RMSD plots shown for NleB1, SseK1-S286Y, SseK1-WT and SseK2.*

The UDP-Glc-bound systems were all stable over 0.5 µs except for SseK1-WT which showed an increase in dynamics after 0.35 µs. In this simulation, the lid domain was again responsible for the increased dynamics as it became disordered at this time point. Because of this increase in disorder at the end of the simulation, analysis of this system was performed from 0 µs to 0.35 µs.

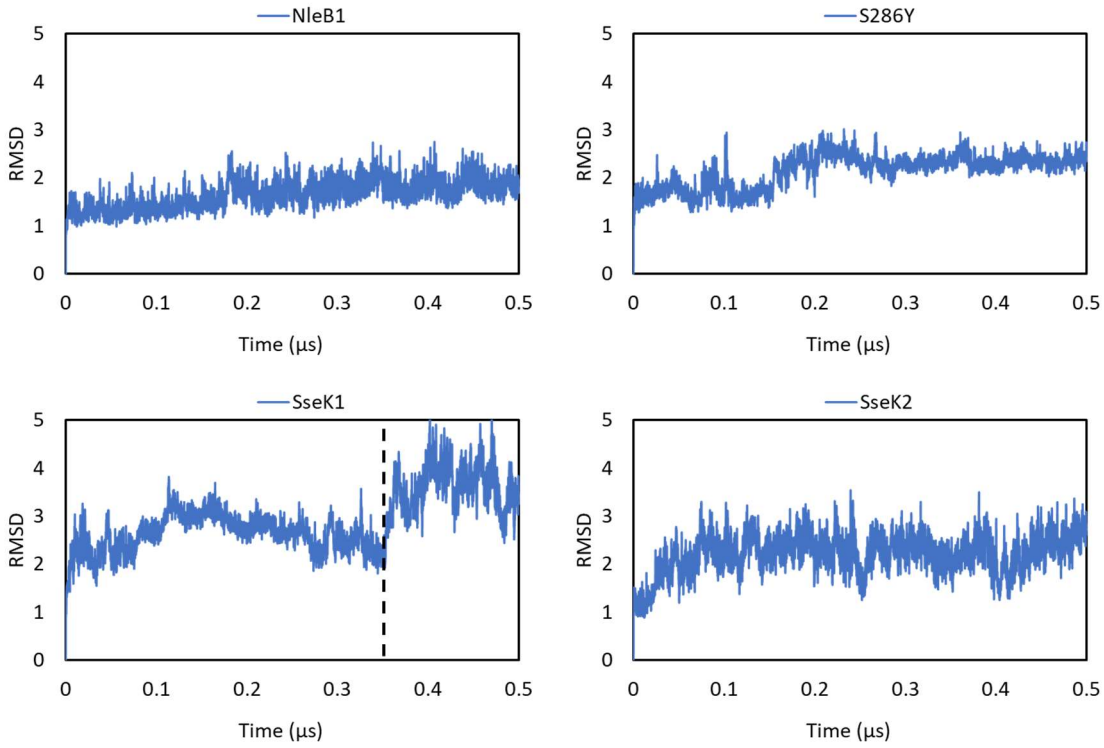



*Figure 4.11 - RMSD (Å) calculations for UDP-Glc-bound complexes. RMSD plots shown for NleB1, SseK1-S286Y, SseK1-WT and SseK2.*

For the simulation of UDP-GalNAc, there is less stability when compared to the other systems. For NleB1, there is a change in RMSD between 0.2 µs and 0.25 µs. Within this time, the lid domain can be seen to disengage from the ligand leaving the catalytic site solvent exposed. Because of this, analysis was performed after 0.25 µs. RMSD calculations for SseK1-WT show that there are dynamic events throughout the duration of the simulation. In this simulation the lid domain undergoes much more motion than in NleB1 but does not disengage from the ligand. Because the simulation returns to a similar position based on the RMSD calculations the whole simulation was used in analysis.

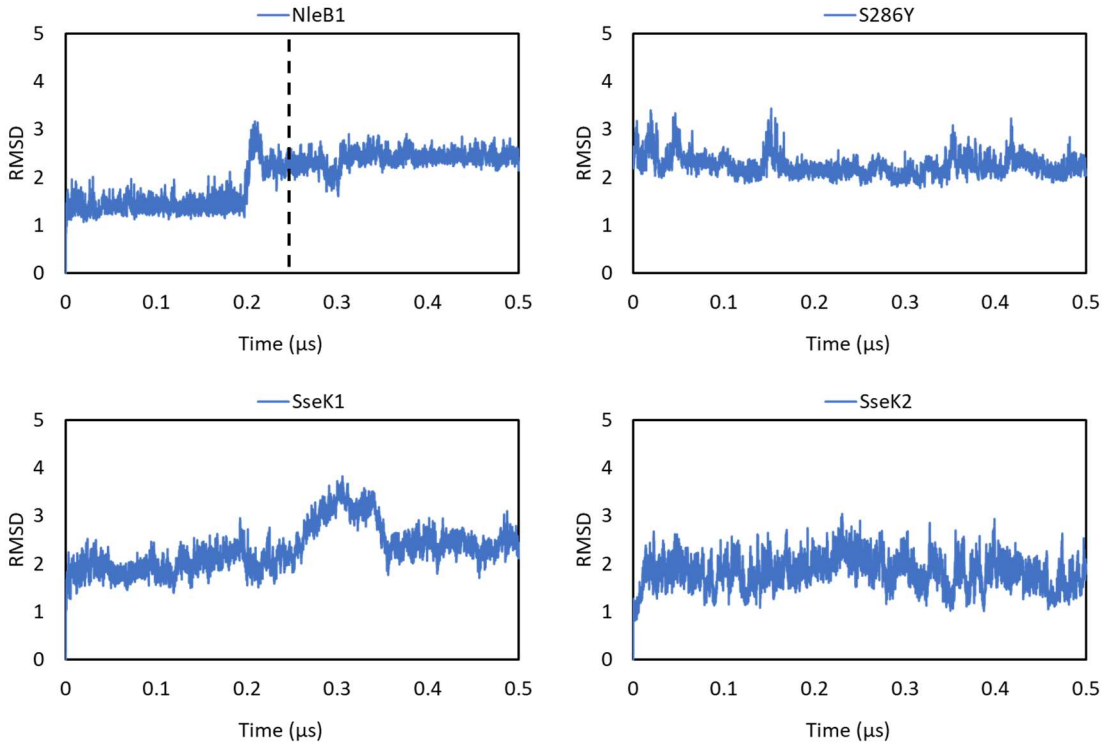



*Figure 4.12 - RMSD (Å) calculations for UDP-GalNAc-bound complexes. RMSD plots shown for NleB1, SseK1-S286Y, SseK1-WT and SseK2. The NleB1 system undergoes a period of instability starting at 0.2 µs, which resolves by 0.25 µs (black dashed line). Because of this, the NleB1 system was analysed from this point forward.*

Simulations of UDP-Gal-bound systems were also similarly stable except for SseK1-S286Y which showed a change from 0 µs to 0.15 µs. In this period there is a repositioning of the lid domain which can explain the sharp increase seen in the RMSD plot. The starting point of the simulation, after equilibration, also places the uridine group antiparallel to SseK1-S286Y$^{F187}$, not being properly sandwiched in the π-π stacking interactions characteristic of these enzymes. During this period, this is also corrected for leading to some changes in the placement of residues around the uridine, also contributing to the increase seen in the RMSD plot. Because of this, analysis of this system was performed from 0.15 µs.

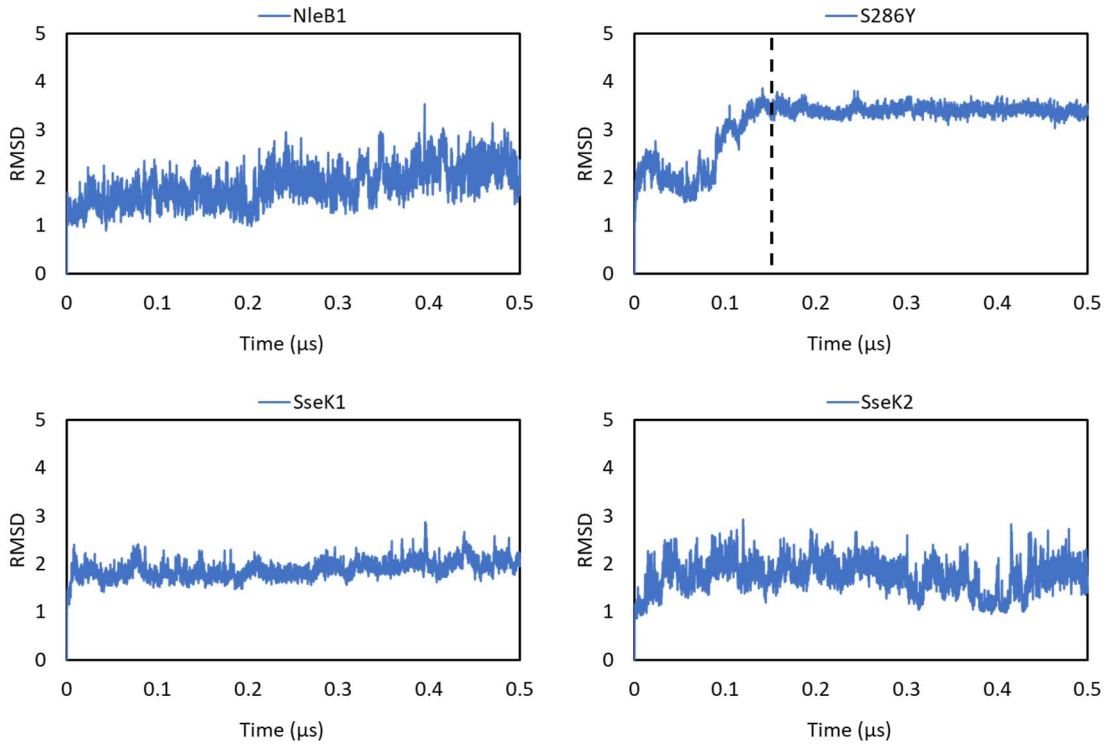



*Figure 4.13 - RMSD (Å) calculations for UDP-Gal-bound complexes. RMSD plots shown for NleB1, SseK1-S286Y, SseK1-WT and SseK2. The SseK1-S286Y system undergoes a period of instability which stabilises at 0.15 µs (black dashed line). Because of this, the SseK1-S286Y system was analysed from this point forward.*

### 4.2.2.1.2 RMSF calculations

RMSF calculations for UDP-bound systems showed that the ligands all undergo similar amounts of motion in the active site (Figure 4.14). For UDP-GlcNAc-bound systems, this is not the case, and the ligand is more mobile in SseK1-WT and SseK1-S286Y than NleB1/SseK2. An important difference between SseK1 constructs compared to NleB1/SseK2 is that they have an extended lid domain. As this domain is involved in contacts with UDP-GlcNAc, motion in this domain will have an impact on its stability in the active site.

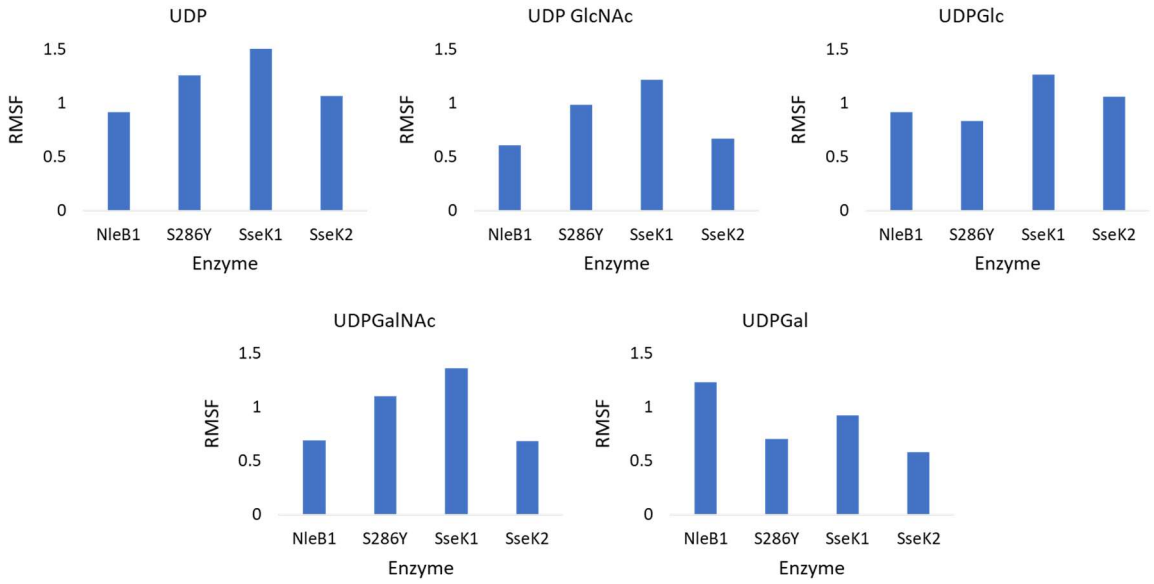



*Figure 4.14 - RMSF (Å) calculations for ligands. RMSF values shown for UDP, UDP-GlcNAc, UDP-Glc, UDP-GalNAc and UDP-Gal bound to NleB1, SseK1-S286Y, SseK1-WT and SseK2.*

These trends are continued to the other donors, except for the UDP-Gal-bound NleB1 system. In this system the ligand undergoes more motion in the active site. It could be expected if the ligand was stably accommodated that the RMSF values for the other donors would be much higher when compared to UDP-GlcNAc. This is because the sugar ring could be more solvent exposed and thus able to undergo more motion. As the RMSF values are comparable on an enzyme-to-enzyme basis, these results suggest that it is possible for these enzymes to accommodate donors other than UDP-GlcNAc in the active site.

### 4.2.2.2 The Binding Mode of UDP, UDP-GlcNAc and Other Donors

In order to provide a representative binding mode of each ligand within the active site of the enzymes, MD simulations were clustered by *k*-means. This method measured the RMSD of the ligand, compared to the first frame, and groups similar structures into populations called clusters. From these groups, the frames are then compared and the one that is least different from the other frames is chosen as the representative frame. *K*-means was chosen as the simulations are dynamic and it is difficult to manually select a conformation from thousands of frames. The calculations, performed on each ligand-enzyme system, were set to find 10 clusters, from which 10 representative frames were found. Each of these frames will belong to different populations of frames from the simulation, with the most populated cluster producing the overall representative frame. However, it can be the case that other clusters can have similar population sizes to the one with the overall most representative frame. In this discussion, the binding mode and overall conformation of each enzyme will be considered when choosing a representative frame from the clusters that are most populated.

With the knowledge that it is possible for NleB1, SseK1-WT, SseK1-S286Y and SseK2 to accommodate the ligands stably in the active site from the RMSF calculations, this section explores whether this is with comparable binding modes. With an understanding of the possible binding modes of these ligands based on this clustering analysis it would be possible to suggest if the subtle difference in binding epitopes seen in the STD NMR epitopes are due to differences in binding mode. An important note is that it was not possible to perform STD NMR experiments involving the use of NleB1 due to the stability of the protein in solution. NleB1 however, was included in this study as it is possible to gain additional information on the potential binding modes of the ligands studied to this protein and whether they may differ to those of SseK1-WT, SseK1-S286Y and SseK2.

#### 4.2.2.2.1 Recognition of UDP and UDP-GlcNAc

The representative frames for UDP-bound and UDP-GlcNAc-bound systems suggest that the binding mode for these ligands are similar in all systems (Figure 4.15). There is one exception to this which is the representative frame for UDP-bound SseK1-S286Y. In this

complex UDP is not accommodated in the same manner as seen in the other enzymes. However, as the RMSD calculations suggest, this simulation is more dynamic than the others. It can be seen for the SseK1-S286Y simulation that the lid domain undergoes significant motion over the course of the simulation. As this domain is important in both interactions with the bound ligand and coordination of the divalent cation, Mn(II), motion of this domain will impact the ligand binding. The first three representative frames for this simulation all have similar populations. The third most populated representative frame was chosen as the lid domain is ordered, with the uracil group in a proper π-π stacking interaction.

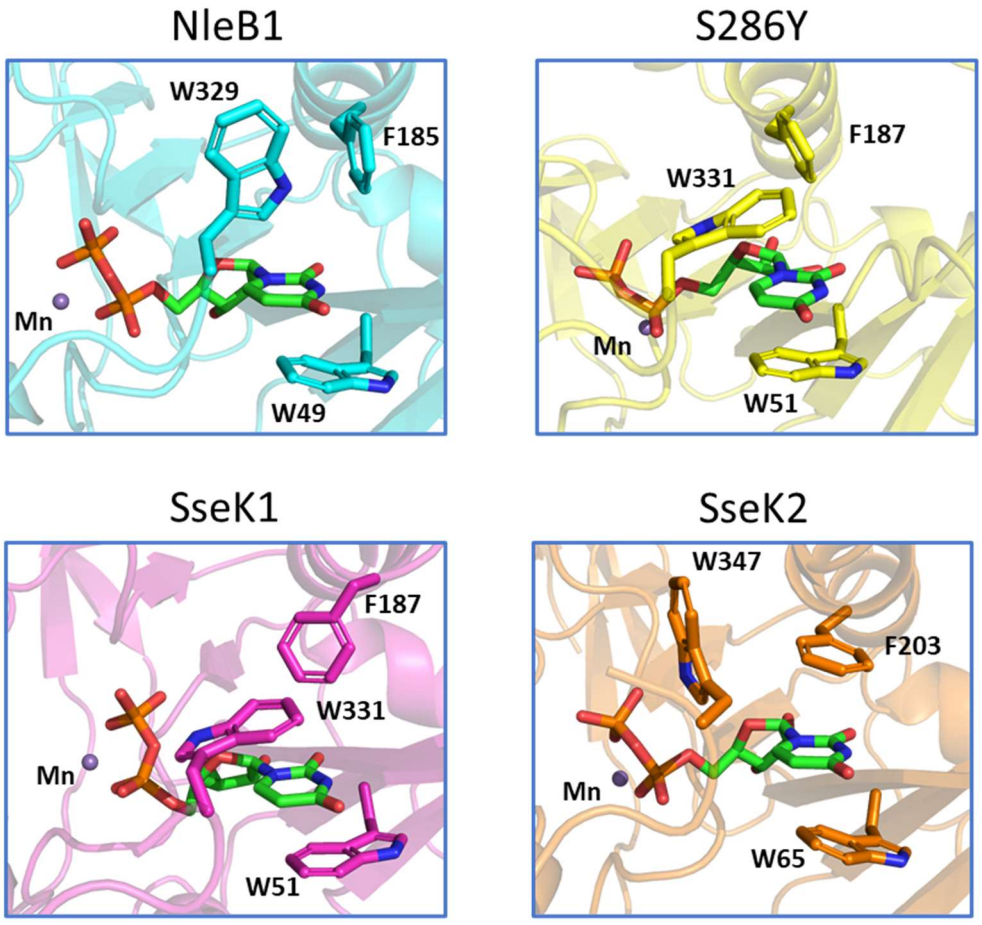



*Figure 4.15 - Representative frames from MD simulation of UDP (green carbons) complexes. Shown are NleB1 (light blue), SseK1-S286Y (yellow), SseK1-WT (pink) and SseK2 (orange). The third most representative frame was chosen for SseK1-S286Y as it was similar in population to the most representative frame and presented the proper π-π stacking interaction for the uracil group.*

In these simulations the uridine ring remains within a π-π stacking interaction in agreement with crystal structure data (36), except for NleB1 (Figure 4.14). In NleB1, the π-π stacking appears to be less stable than in the other simulations with the representative frame showing only an interaction from NleB1$^{W49}$. For SseK2, the π-π stacking comes from a tryptophan-phenylalanine pair. For SseK1-WT and SseK1-S286Y, these interactions use an equivalent tryptophan however the phenylalanine is replaced by the lid domain tryptophan SseK1-WT/S286Y$^{W331}$. In all simulations, both H5 and H6 of the uridine ring are within proximity of the residues contributing to the π-π stacking interaction. This is in good agreement with the STD NMR data which places a strong emphasis on interaction with the uridine ring as being important for recognition. The same is true for the ribose where H1' is positioned near NleB1$^{F185}$/SseK1$^{F187}$/SseK2$^{F203}$ and H2'/H3' near NleB1$^{W49}$/ SseK1$^{W51}$/ SseK2$^{W65}$. For H4', this hydrogen is not placed near the π-π stacking interaction, however, is directed close to the DXD motif glutamic acid NleB1$^{D221}$/SseK1$^{D223}$/SseK2$^{D239}$. In a similar sense to H5 and H6 these proximities would correlate to strong STD contacts, which agrees with the STD NMR data for SseK2. The H5' protons however are not as close to residues with the nearest being the other DXD glutamic acid NleB1$^{D223}$/SseK1$^{D225}$/SseK2$^{D241}$. Though the absolute difference in distance is small as inter-molecular spin diffusion is highly distance dependent this could result in significant differences in the epitope. This observation agrees with the STD NMR epitope which shows contacts for H5' to have lower STD% values compared to the other ribose hydrogens.

The binding modes of UDP-GlcNAc with NleB1, SseK1-WT, SseK1-S286Y and SseK2 in the representative frames are all very similar. For binding to SseK1-WT and SseK1-S286Y the difference in binding mode is very minimal. As is seen with UDP, the uridine moiety of UDP-GlcNAc is contained within the same residues, with the uracil ring sandwiched between Trp51 and Trp331 in the SseK1 constructs (Figure 4.16). For NleB1 and SseK2, the π-π stacking interactions for uracil are consistent with their crystal structures, showing instead the favoured use of a phenylalanine in place of the lid domain tryptophan seen in SseK1. The patterns of proximities for protons for uridine are the same as for UDP, in agreement with the STD NMR epitopes.

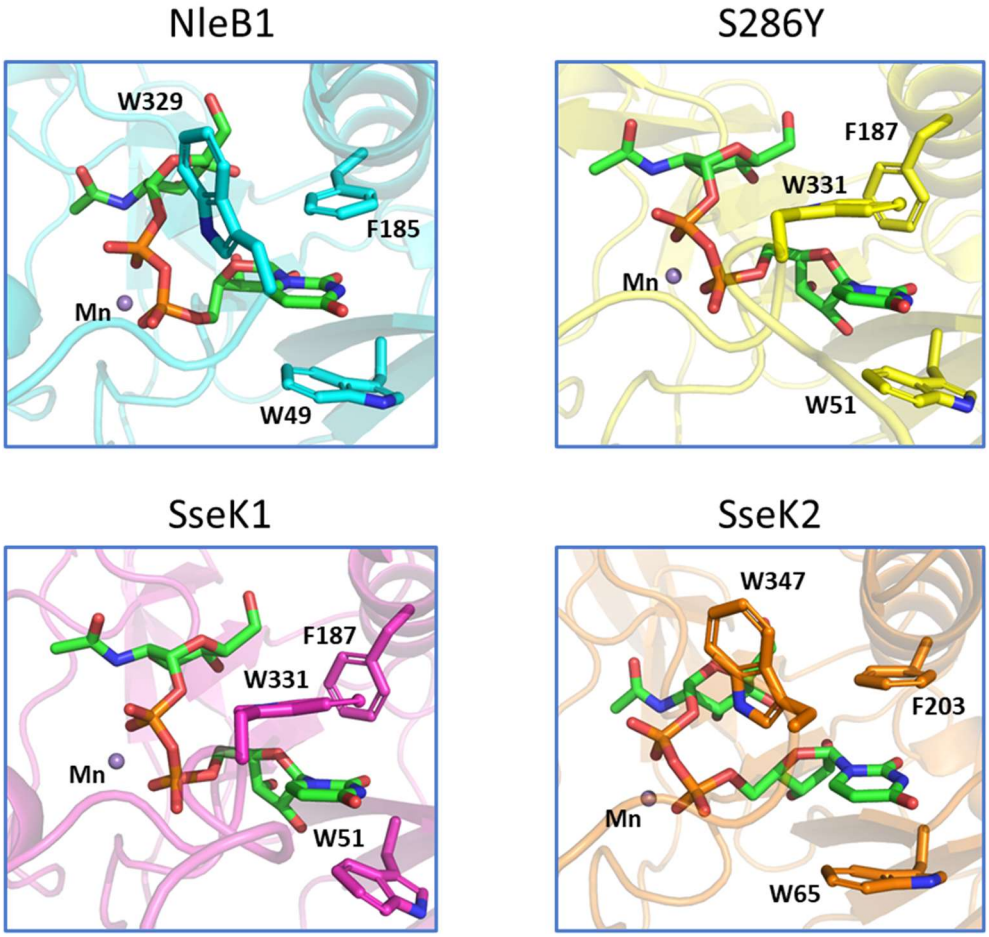



*Figure 4.16 - UDP-GlcNAc uracil π-π stacking interactions. Representative frames from MD simulation shown for UDP-GlcNAc-bound (green carbons) NleB1 (light blue), SseK1-S286Y (yellow), SseK1-WT (pink) and SseK2 (orange).*

In the SseK1 constructs, a hydrogen bonding network formed by Asp188/Arg191/Glu255 stabilising the GlcNAc ring (Figure 4.16). In SseK2 a similar hydrogen bonding network is seen involving Asp204/Arg207/Asn272 (Figure 4.17). The aspartic acid interaction is not conserved in this hydrogen bonding network as the sugar ring is placed in a different position with the SseK2 active site when compared to the SseK1 constructs. In the conformation seen in SseK2, there is an addition intramolecular hydrogen bonding interaction between the hydroxyl groups on position 4 of the GlcNAc ring and position 2 of the ribose ring.

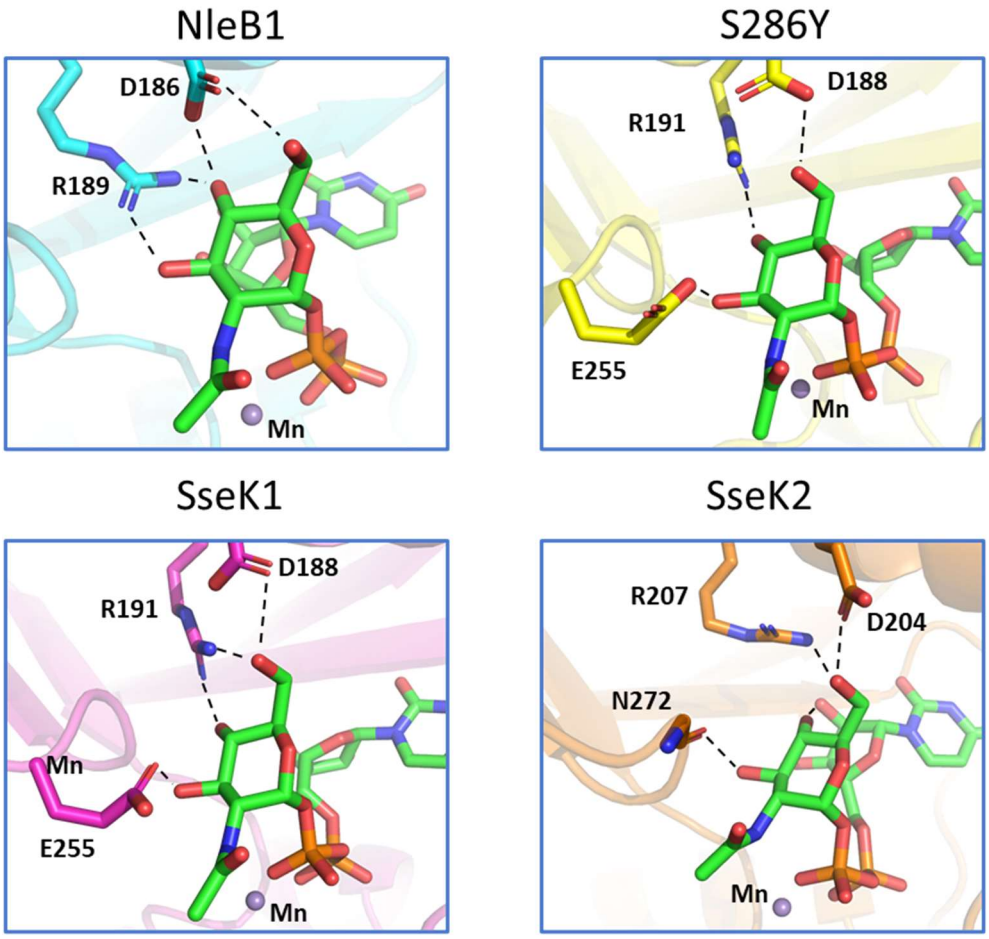



*Figure 4.17 - UDP-GlcNAc sugar ring hydrogen bonding interactions. Representative frames from MD simulation shown for UDP-GlcNAc-bound (green carbons) NleB1 (light blue), SseK1-S286Y (yellow), SseK1-WT (pink) and SseK2 (orange).*

Though stabilised by hydrogen bonding interactions, the protons of the GlcNAc ring are not situated as close to residues as those of the uridine moiety. Coupled with the relatively weak interactions that hydrogen bonding offers there is also potential for more motion for sugar ring within the hydrogen bonding network. These two factors would lead to the overall lower STD% values seen in the epitope for the GlcNAc ring. Comparing the binding epitopes for UDP-GlcNAc when bound to SseK1-WT, SseK1-S286Y or SseK2, there are subtle differences on the ring that may be attributed to different binding modes for the ring. It is difficult to precisely correlate which interactions are responsible for the disparity as the GlcNAc ring is close to the highly dynamic lid domain and itself is involved in weak interactions in the active site. However, the difference in binding mode shown in MD simulation and the subtle differences in epitope provide evidence that there may be a

difference in GlcNAc recognition between the SseK1 and SseK2 effectors. From the representative frame, NleB1 recognizes the GlcNAc ring in a similar manner to SseK2, however lacking the intramolecular hydrogen bond (Figure 4.17).

In the crystal structure of UDP-GlcNAc-bound SseK2, the acetyl group of the GlcNAc ring can be seen to form hydrophobic interactions Met242/His260 (36). In the representative frame from the MD simulation the acetyl group is directed towards this pocket in a similar manner (Figure 4.18). In the SseK1 constructs, the acetyl group is not positioned within this pocket, instead being placed close to the aliphatic hydrogens of Glu246/Glu255 (Figure 4.18). The STD NMR data suggests that in the SseK1 constructs the acetyl group of the GlcNAc may be more closely recognised when compared to SseK2. It is difficult, however, to determine if these differences in recognition would explain the differences in epitope for UDP-GlcNAc binding to SseK1-WT, SseK1-S286Y and SseK2. The reason for this is that many factors can influence the measurements from the STD NMR experiment, namely the dynamic nature of these systems and the weak interactions sugar rings have with proteins.

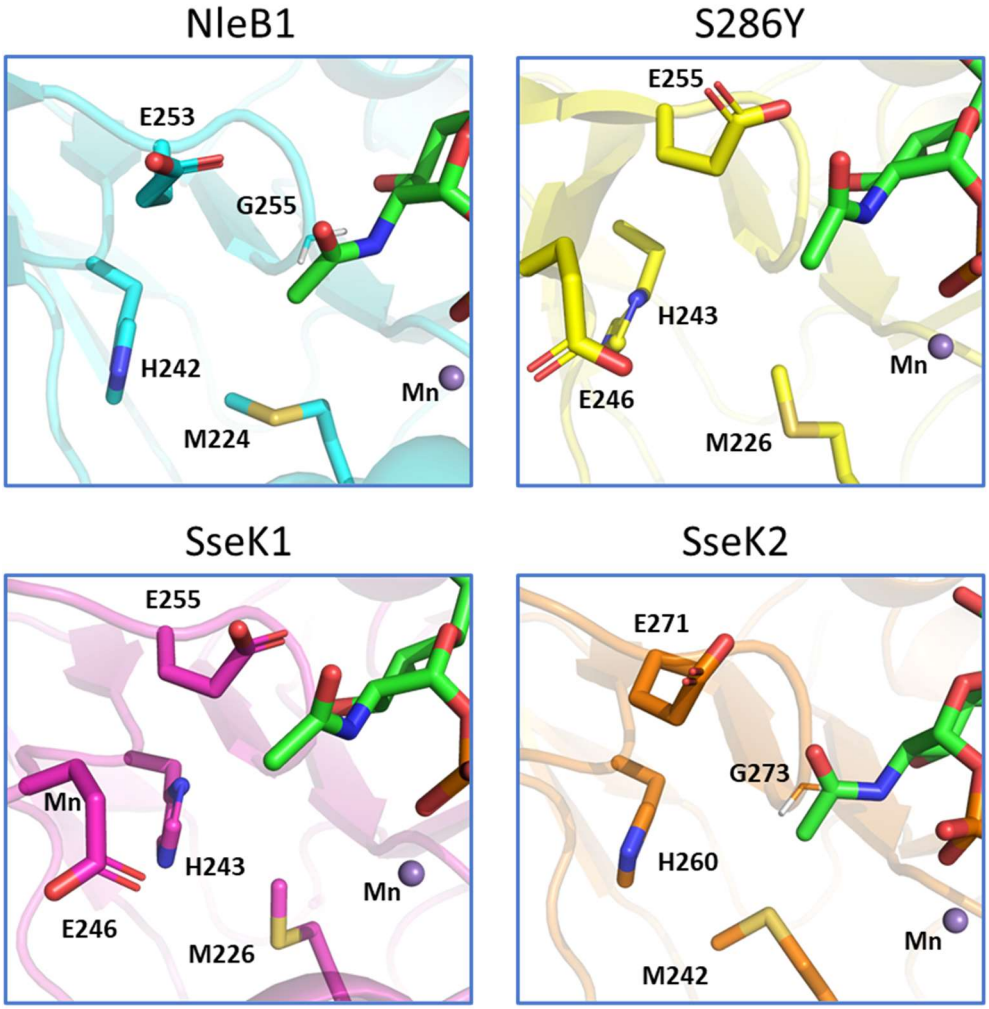



*Figure 4.18 - UDP-GlcNAc N-acetyl hydrophobic interactions. Representative frames from MD simulation shown for UDP-GlcNAc-bound (green carbons) NleB1 (light blue), SseK1-S286Y (yellow), SseK1-WT (pink) and SseK2 (orange).*

#### 4.2.2.2.2 Recognition of Non-UDP-GlcNAc Donors-Like Ligands

UDP-Glc is recognised in a similar manner to UDP-GlcNAc within the active site. In the representative frame for SseK1-WT and SseK1-S286Y, there is a slight difference in conformation of the lid domain which moderately effects the binding mode. Like UDP and UDP-GlcNAc, the uridine moiety is recognised with the same tryptophan pair as in other simulations of SseK1-WT and SseK1-S286Y (Figure 4.19). Also, the sugar ring can form similar hydrogen bonding interactions despite lacking the NAc group at position. For both NleB1 and SseK2, UDP-Glc takes on a similar binding mode though the hydroxyl at position 2 coordinated by the Mn(II) ion. In both SseK1 constructs the hydroxyl at position 2 is not coordinated by Mn(II) and is instead placed close to Glu255, forming a hydrogen bonding interaction in the representative frame for SseK1-S286Y but not SseK1-WT.

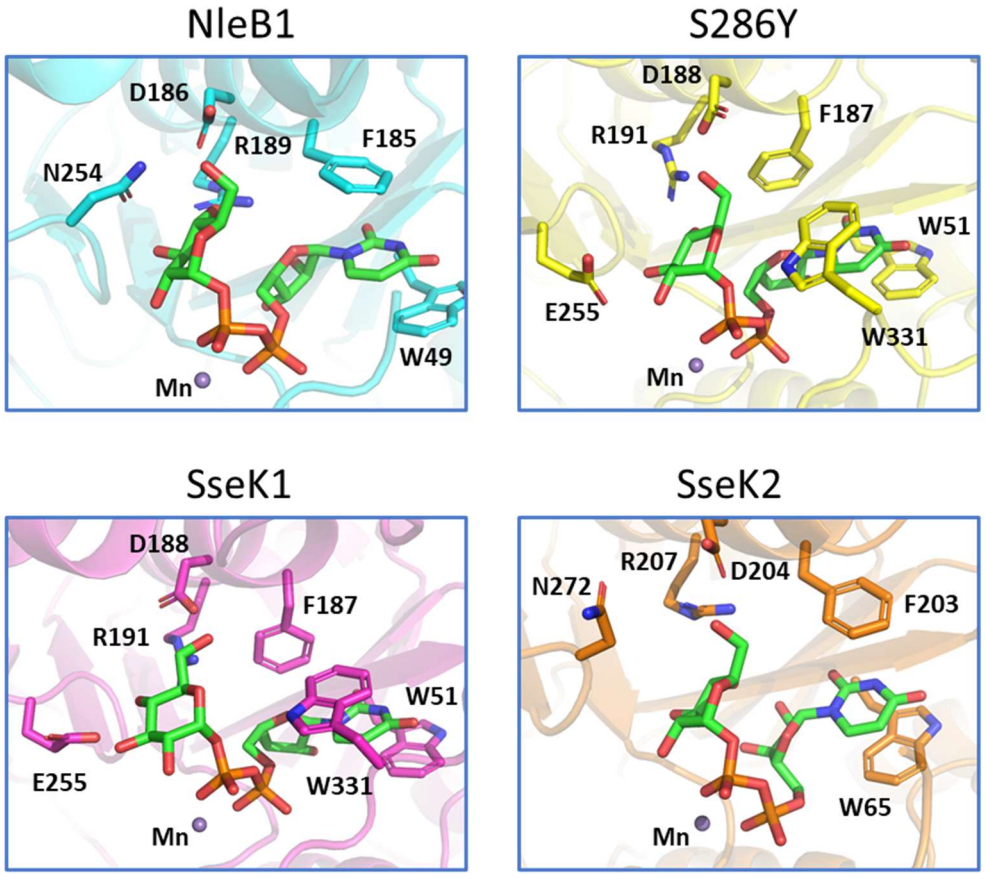



*Figure 4.19 - Representative frames from MD simulation of UDP-Glc (green carbons) complexes. Shown are NleB1 (light blue), SseK1-S286Y (yellow), SseK1-WT (pink) and SseK2 (orange).*

Considering UDP-Gal, given the epimerization at position 4, it would be reasonable to assume that the Gal ring would not be able to form the same interactions that GlcNAc or Glc can. This is because the axial hydroxyl group at position 4 would not be directed towards complementary residues and could experience a steric clash in the active site. Despite this and lacking the NAc group at position 2, all the enzymes could accommodate UDP-Gal. Compared to the binding modes of the ligands up until now UDP-Gal, however, displays less agreement in binding mode between the enzymes (Figure 4.20). In the representative frames for the UDP-Gal-systems there are also differences in the π-π stacking interactions. In NleB1 the lid domain tryptophan NleB1$^{W329}$ takes the place of NleB1$^{F185}$ forming a more SseK1-like interaction with uracil. In SseK1-WT/S286Y the lid domain has repositioned, the result of this is that rather than the lid domain tryptophan forming a π-π stacking interaction with uracil SseK1-WT/S286Y$^{F187}$ takes its place. As it is expected that the lid domain is ordered by interactions with UDP-GlcNAc (36), it is possible that UDP-Gal does not offer the same stabilizing interactions, leading to a change in lid domain dynamics and thus the change in π-π stacking interactions.

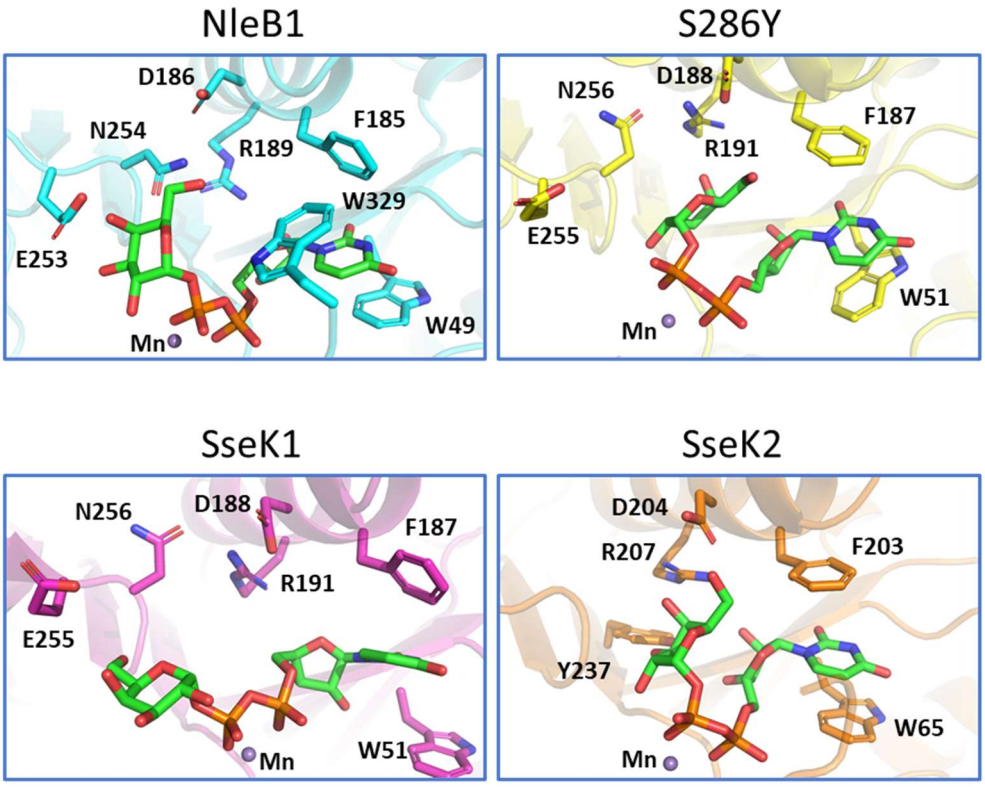
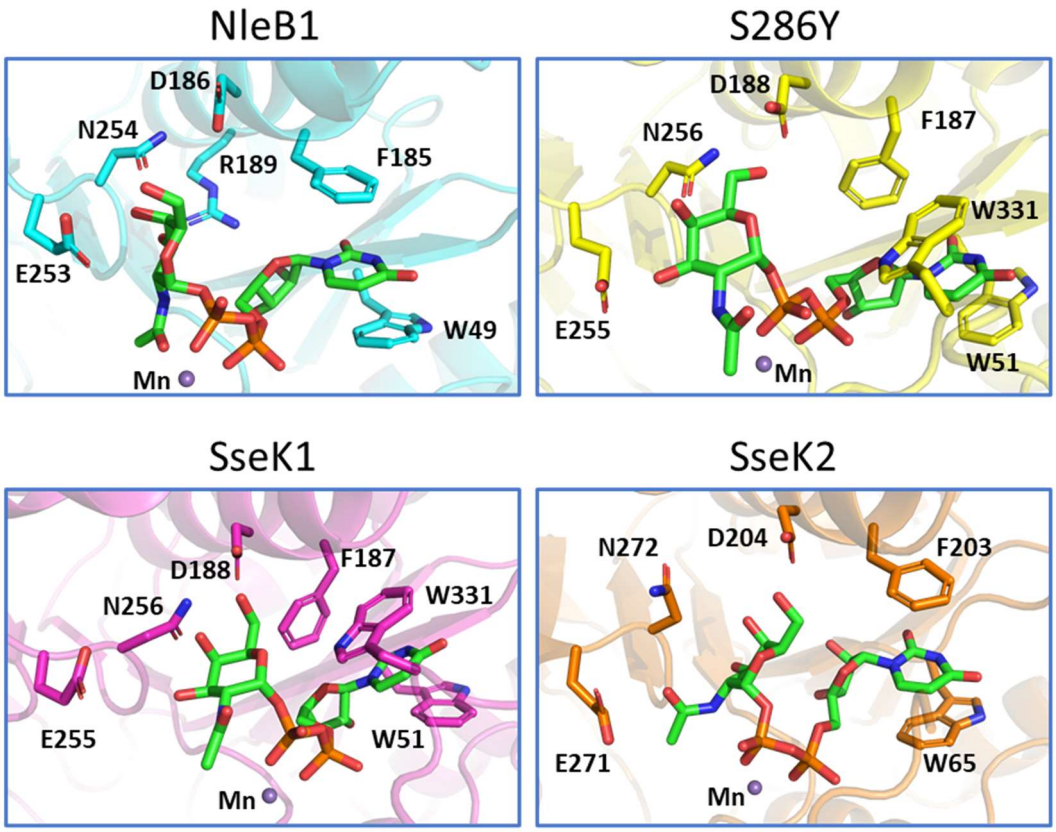



*Figure 4.20 - Representative frames from MD simulation of UDP-Gal (green carbons) complexes. Shown are NleB1 (light blue), SseK1-S286Y (yellow), SseK1-WT (pink) and SseK2 (orange).*

In the representative frames for SseK1-S286Y and SseK2, the Gal sugar ring is placed in an analogous position to that of Glc and GlcNAc (Figure 4.20). In SseK1-S286Y, the hydroxyl group at position 2 of Gal is instead coordinated by Mn(II). For SseK2, the hydroxyl at position 2 remains coordinated by Mn(II) as with UDP-Glc. Notably despite being in a similar position to Glc and GlcNAc in the active site, Gal cannot form the same efficient hydrogen bonding interactions. Despite being very similar structures, in the representative frame for SseK1-WT the Gal ring is positioned differently. In this simulation the ring does not fully explore the active, instead the hydroxyl at position 2 forms a stable hydrogen bond with the pyrophosphate group. In the representative frame the hydroxyl at position 2 remains coordinated to Mn(II) as for SseK2. In this simulation the Gal ring is not placed in the GlcNAc site and instead placed in a more solvent exposed position (Figure 4.20).

The binding of UDP-GalNAc shows similar disparities as UDP-Gal (Figure 4.21). Focusing first on NleB1, RMSD calculations show that there is a conformation change between 0.2 μs and 0.25 μs. In this period, the lid domain disengages and moves to the lid-open conformation. Clustering of the period before and after this dynamic event show that there is limited impact of the binding conformation of UDP-GalNAc. As is typical for this enzyme, the uracil ring remains within a π-π stacking with NleB1$^{F185}$ and NleB1$^{W49}$. In the representative frame, the GalNAc ring can also be seen to be accommodated in a similar position to GlcNAc though in a different orientation. The acetyl group of the GalNAc ring can be seen to be coordinated by the Mn(II) ion directing the hydroxyl groups at positions 3 and 4 to NleB1$^{D221}$ and NleB1$^{E253}$, respectively. In the representative frame for SseK1-WT/S286Y UDP-GalNAc appears to be coordinated quite well in the active site (Figure 4.21). Despite there being a complete hydrogen bonding network in this representative frame the GalNAc ring is more solvent exposed compared to the binding of the GlcNAc ring. In the representative frame for SseK2, the GalNAc ring is positioned deeper in the active compared to SseK1-S286Y.




*Representative frames from MD simulation of UDP-GalNAc (green carbons) complexes. Shown are NleB1 (light blue), SseK1-S286Y (yellow), SseK1-WT (pink) and SseK2 (orange).*

While the donors could be accommodated in a manner similar UDP-GlcNAc there are compromised interactions with the Gal and GalNAc rings. Notably, despite the uridine nucleotide being accommodated in a similar manner, these sugar rings cannot form interactions with the same hydrogen bonding network. The result of this is that there is a weaker recognition of Gal and GalNAc in the active site. Factoring in that carbohydrate interactions with proteins can already be quite weak this will mean that the sugar rings can be more mobile and therefore experience fewer stable contacts. This would agree with the STD epitopes which show lower STD% values for the Gal and GalNAc ring of ~20% compared to the values of ~50% seen for Glc and GlcNAc. From these MD results it would be reasonable to conclude that the differences in epitopes between the sugar nucleotides is therefore related to the stabilization of the sugar ring in the active site.

# CHAPTER 5

## CONTRIBUTIONS TO OTHER WORKS

# 5.1 Introduction

Alongside the experimental work to achieve the general aims of this thesis, I made contributions to other publications through collaboration. These contributions involved the use of the main methods used in this thesis, including STD NMR and MD simulation. This chapter will present an introduction to each publication and detail the contribution I made. Over the course of my PhD studies, I have been listed as author on more publications than those listed here. However, the publications presented in this chapter are ones for which I made a meaningful intellectual contribution.

# 5.2 Self-Correcting Method for the Measurement of Free Calcium and Magnesium Concentrations by $^{1}$H NMR.

***Publication:*** *Wallace M,* ***Hicks T****, Khimyak YZ, Angulo J. Self-Correcting Method for the Measurement of Free Calcium and Magnesium Concentrations by $^{1}$H NMR. Anal Chem. 2019 Nov 19;91(22):14442-14450. doi: 10.1021/acs.analchem.9b03008.*

## 5.2.1 Introduction

For the study of many chemical and biological systems by NMR, it is important to have a reliable way to measure the presence of free ions in solution. In the presence of metal-ion binding materials there will be a difference in concentration between free and bound ions. Knowing this information becomes particularly useful in the study of the formation and physical properties of hydrogels, which are dependent on the concentration of ions in solution. There are many established methods of measuring free ion concentration in solution however, they can often disrupt gel structure or be incompatible with NMR study. To solve this problem this publication presents a method of measuring free-ion concentration by NMR. This is achieved by measuring the chemical shift of reporter ligands, glycolate and sulphoacetate, which are ion binding dependent.

## 5.2.2 Contribution

To validate the approach laid out in this publication the concentration of Ca(II) and Mg(II) ions (M(II)) in solution was measured in the presence of a complexing ligand. By doing this, the solution will contain a mixture of M(II)-ligand complex and free M(II) ions, the concentration of which can be measure by a reporter ligand. One such complexing ligand that was used during validation was maleate. Because it is possible to determine the gauche conformation (Figure 5.1) of maleate from the $^3J_{AX}$ and $^3J_{BX}$ couplings, the conformational distribution of maleate can be measured as a function of M(II) concentration. As part of this study, this investigation was performed using a concentration gradient within an NMR tube, spatially resolved using NMR imaging. The result of this study indicated that with increasing M(II) concentration the trans conformation of maleate decreased while the gauche conformation increased (Figure 5.1). To explore these conformation distributions further, my contribution to this work was to perform MD simulations of the M(II)-maleate complexes with Ca(II) and Mg(II) as well as the free maleate ligand. The simulations were 10 ns long and the dihedral angle (ψ) between the two carboxylic acid groups were plotted as a histogram to represent the conformational distribution (Figure 5.1).

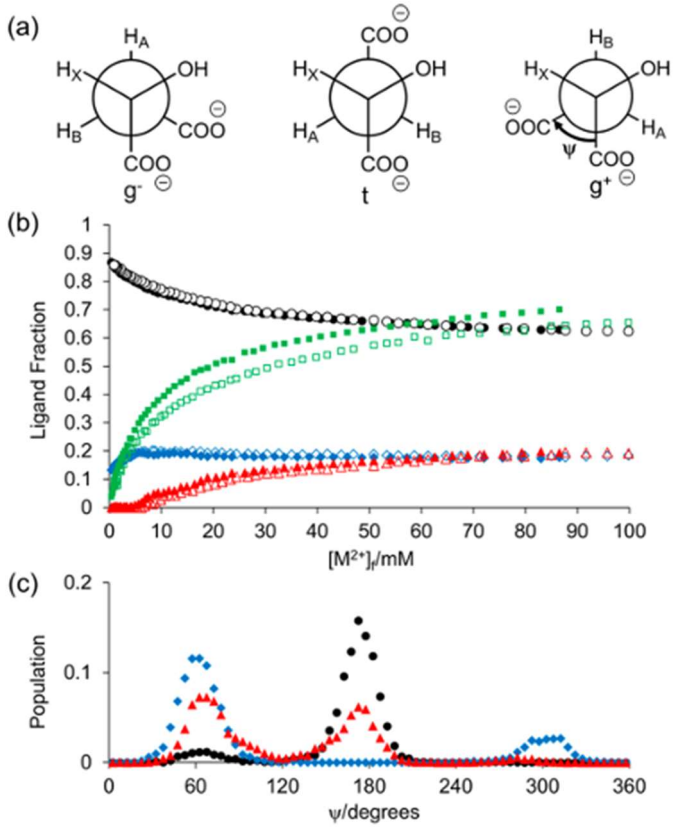



*Figure 5.1 - Agreement of NMR and MD conformational distributions. The Newman projections of maleate (a) showing the gauche negative ($g^-$), trans (t) and gauche positive ($g^+$) conformations. (b) The conformational distributions found by NMR showing $g^-$ (red triangle), trans (black circle) and $g^+$ (green squared). This plot includes both in the presence of Ca(II) (solid marks) and Mg(II) (hollow marks). Conformational distribution of maleate from MD simulation by dihedral angle (ψ) for the Ca(II)-maleate (blue diamond), Mg(II)-maleate (red triangle) and free maleate (black circle) (c). In this plot, the $g^-$, trans and $g^+$ conformations are seen at approximately 60°, 180° and 300°, respectively.*

The simulations agreed with the NMR data, in the presence of M(II) ions, the conformation of maleate tends towards gauche. Moreover, the $g^+$ conformation is typically preferred over the $g^-$ conformation, in agreement with the NMR data. This makes intuitive sense as the $g^+$ conformation minimises steric clashes. Moreover, the MD data predicted that the Ca(II)-maleate complex prefers the $g^+$ conformations to a greater degree than the Mg(II)-maleate complex, which is also in agreement with the NMR data.

# 5.3 Fucosidases from the human gut symbiont *Ruminococcus gnavus*

***Publication:*** *Wu H, Rebello O, Crost EH, Owen CD, Walpole S, Bennati-Granier C, Ndeh D, Monaco S,* ***Hicks T****, Colvile A, Urbanowicz PA, Walsh MA, Angulo J, Spencer DIR, Juge N. Fucosidases from the human gut symbiont Ruminococcus gnavus. Cell Mol Life Sci. 2021 Jan;78(2):675-693. doi: 10.1007/s00018-020-03514-x.*

## 5.3.1 Introduction

*Ruminococcus gnavus* is an important gut bacterium that uses fucosidases to digest complex carbohydrates such as human milk oligosaccharides (HMO) and mucins for use as a source of nutrition. The enzymes responsible for this being glycoside hydrolases (GH). There are numerous strains of *R. gnavus* that have been observed to inhabit the gut, with their ability to use HMO or mucins in their diet being strain dependent. These differences can be explained by the different distribution of GH families expressed by the different strains of *R. gnavus.* This study focussed on characterising the GH29 fucosidase, E1_10125, from the E1 strain of *R. gnavus*. This enzyme can recognise sialic acid-capped, fucosylated glycans and catalyses the hydrolysis of α-1,3/4-fucosyl linkages. In this study the binding of different oligosaccharides was studied, and part of this work was to characterise their interactions with E1_10125.

## 5.3.2 Contribution

E1_10125 was successfully crystallised in this study however, without any bound oligosaccharides. One of my contributions to this work was studying the binding of the oligosaccharide, αGal-Le$^X$ to E1_10125 by STD NMR. The STD epitope measured for αGal-Le$^X$ could not be reliably measured to atomic resolution due to a crowded $^1$H NMR spectrum. Despite this, it was possible to assign the monosaccharide units that some signals corresponded to. The overlap was due to anomeric exchange of the αGal-Le$^X$ reducing end monosaccharide and that both the galactose rings occupied a very similar chemical shift

range. Because of this, it was only possible to represent the epitope as an average of the STD intensities that could be associated with each monosaccharide (Figure 5.2). The average epitope found for αGal-Le$^X$ was consistent with the epitopes of other ligands studied, with fucose and GlcNAc displaying the strongest STD intensities.

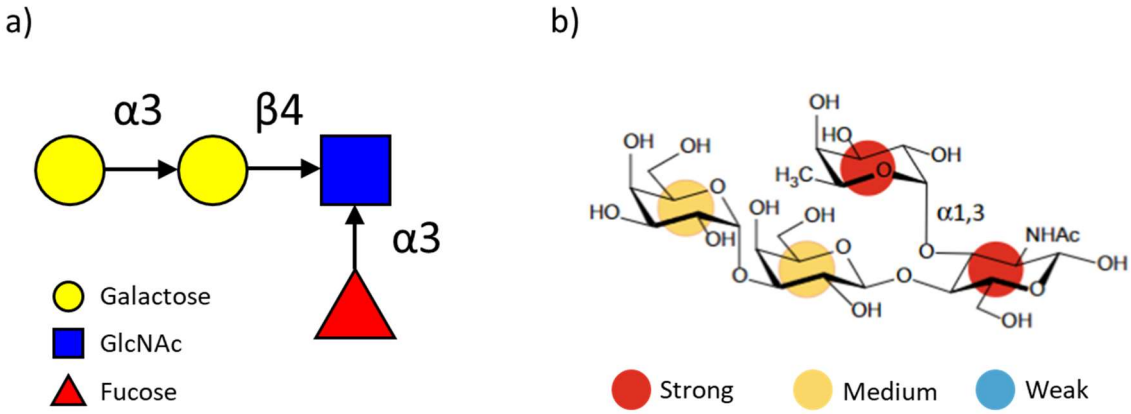



*Figure 5.2 - The binding epitope of αGal-Le$^X$. Schematic representation of αGal-Le$^X$ (a) following the SNFG glycan symbol nomenclature. The STD NMR epitope of αGal-Le$^X$ (b), showing average intensities for each monosaccharide. The STD NMR legend represents strong (51% - 100%), medium (25% - 50%) and weak (0% - 24%) intensities.*

In addition to the STD NMR study my contribution to this work also included the molecular modelling by protein-ligand docking of another oligosaccharide, sLe$^X$, to the fucosidase E1_10125. The crystal structure of E1_10125 had the oligosaccharide binding site in a closed conformation that was too small for the docking of sLe$^X$. To overcome this and find a more open conformation of E1_10125, a 500 ns simulation of the protein in solution was performed. Clustering of the conformations along the simulation length was then used to find representative conformations of E1_10125. This approach was chosen as while it is possible to select a frame from the simulation with an open binding site, that structure may be a transitional state and not an appropriate conformer of the protein. The docking study used many conformers of E1_10125 found from clustering, but the output structures were assessed by their match to the STD NMR data for sLe$^X$. To do this, structures were excluded if they lacked contacts that were highlighted as close contacts by the STD NMR epitope of sLe$^X$ (Figure 5.3).

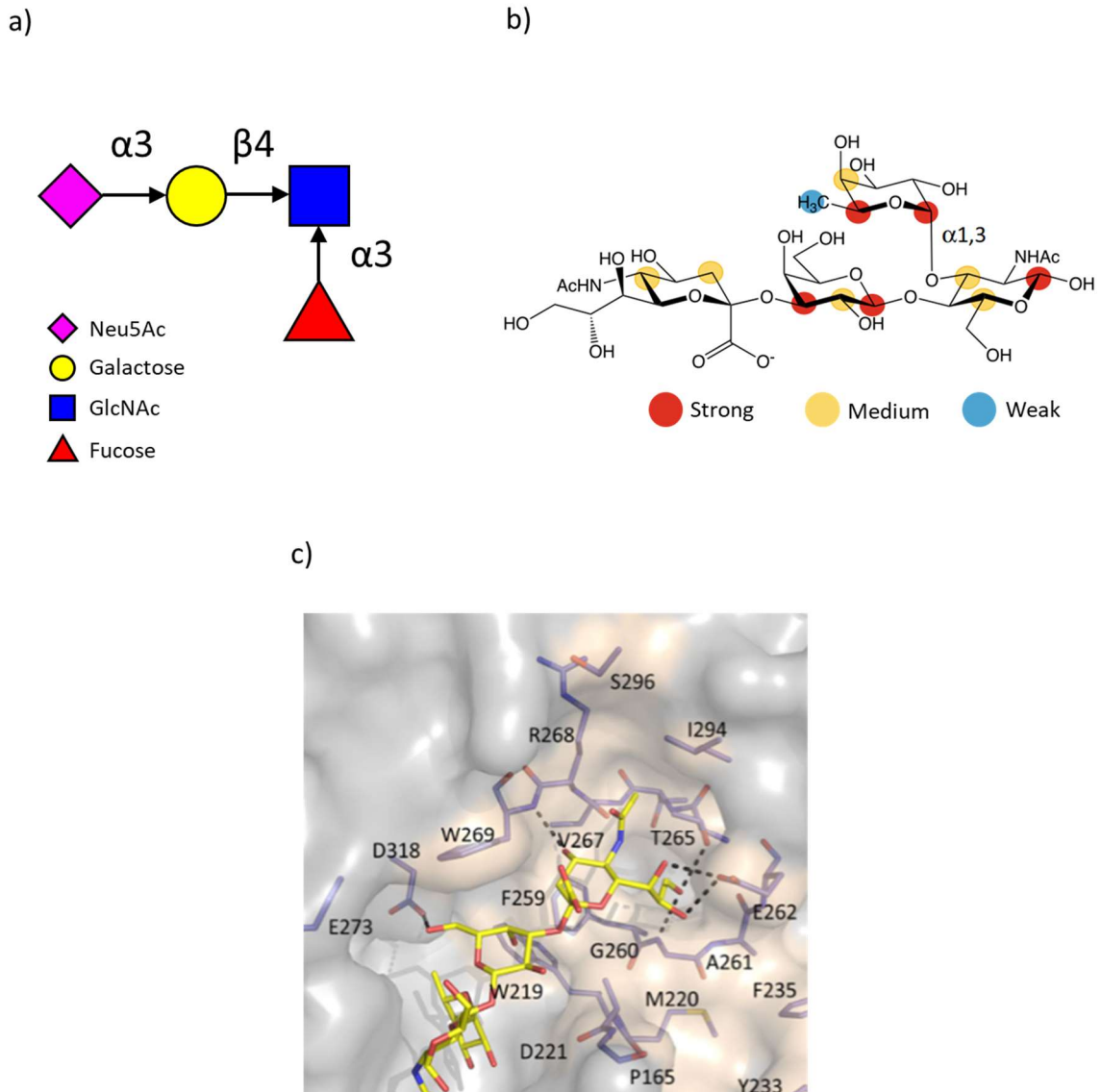



*Figure 5.3 - The binding epitope of sLe$^X$ and predicted binding complex. Schematic representation of sLe$^X$ (a) following the SNFG glycan symbol nomenclature. The STD NMR epitope of sLe$^X$ in the presence of E1_10125 (b), showing average intensities for each monosaccharide. The STD NMR legend represents strong (51% - 100%), medium (25% - 50%) and weak (0% - 24%) intensities. The predicted structure of sLe$^X$ binding to E1_10125 (c), chosen from docking calculations.*

# 5.4 Self-acetylation at the active site of phosphoenolpyruvate carboxykinase (PCK1) controls enzyme activity

***Publication:*** *Latorre-Muro P, Baeza J, Hurtado-Guerrero R,* ***Hicks T****, Delso I, Hernández-Ruiz C, Velázquez-Campoy A, Lawton AJ, Angulo J, Denu JM, Carrodeguas JA. Self-acetylation at the active site of phosphoenolpyruvate carboxykinase (PCK1) controls enzyme activity. J Biol Chem. 2021 Jan-Jun;296:100205. doi: 10.1074/jbc.RA120.015103.*

## 5.4.1 Introduction

Phosphoenolpyruvate carboxykinase (PCK1) is an important enzyme in gluconeogenesis that catalyses the conversion of oxaloacetic acid (OAA) to phosphoenolpyruvate (PEP). In response to changes in cellular energy demands, acetylation of PCK1 takes place leading to promotion of the reverse catalytic reaction (PEP to OAA). There are many acetyltransferases which are known modulate the activity of PCK1, but it is not known whether this is by direct or indirect acetylation. This work found that PCK1 was in fact able to modulate itself by self-acetylation, in the presence of high concentrations of acetyl-COA.

## 5.4.2 Contribution

My contribution to this work was to understand the interaction between acetyl-COA and PCK1 on a molecular level. To achieve this, STD NMR and DEEP-STD NMR was used. The results of this work initially indicated that acetyl-COA binds to PCK1 with stronger STD intensities compare to COA (Figure 5.4). Despite there being differences in binding affinities between acetyl-COA and COA the binding epitopes are comparable. This suggests that likely both ligands bind to the same binding site in a similar binding mode. The structure most important for recognition is the adenosine moiety which is typical of nucleotide-bearing ligands. The other contacts for both ligands tend to be much weaker other than

the two methyl groups on the pantothenic acid moiety. The weaker contacts can be explained by increased flexibility of this moiety compared to the adenosine.

DEEP-STD NMR analysis showed that both ligands bound in a similar binding mode, further supporting that both ligands bind to the same site. Importantly, there are differences in the epitopes between the two saturating frequencies for each ligand. For frequencies near the aliphatic region, in both ligands there are stronger intensities on the methyl groups on the pantothenic acid moiety which are not present when measured at aromatic frequencies. This provides evidence that these methyl groups interact with aliphatic groups within the binding site. This observation was in good agreement with docking results found as part of this study (Figure 5.5), providing supporting evidence for those structures.

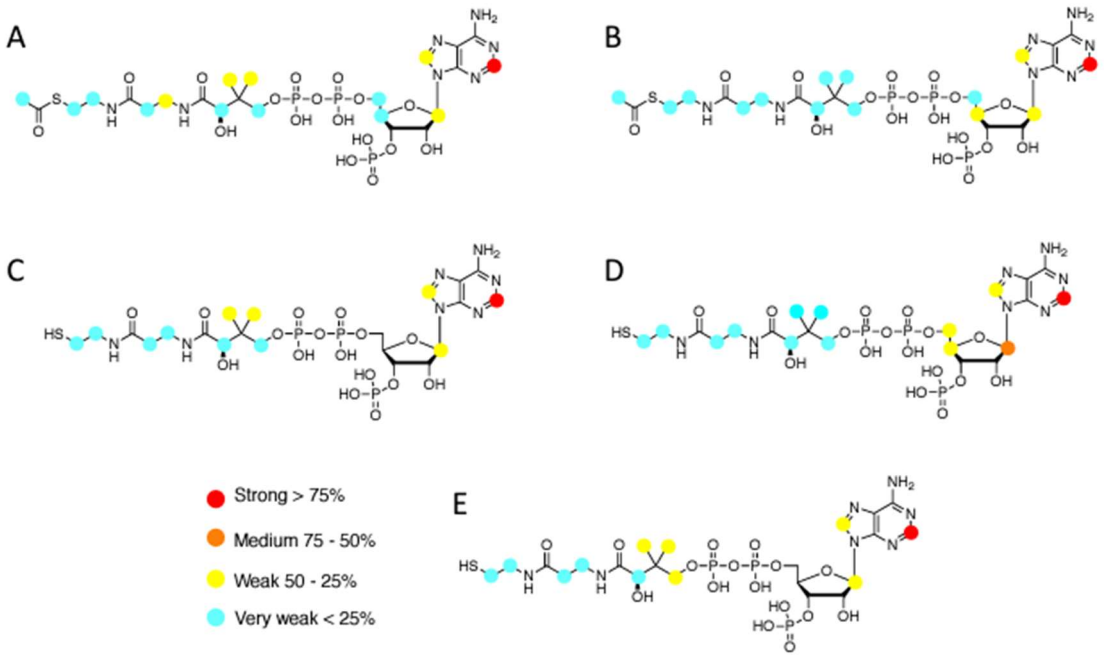



*Figure 5.4 - The binding epitopes of acetyl-COA and COA. Measured in the presence of PCK1, showing (A) acetyl-COA with a saturation frequency of -1 ppm, (B) acetyl-COA with a saturation frequency of 7.19 ppm, (C) COA with a saturation frequency of -1 ppm, (D) COA with a saturation frequency of 7.19 ppm and (E) the binding epitope of COA in the presence of 1 mM $MgCl_2$, and a saturation frequency of -1 ppm.*

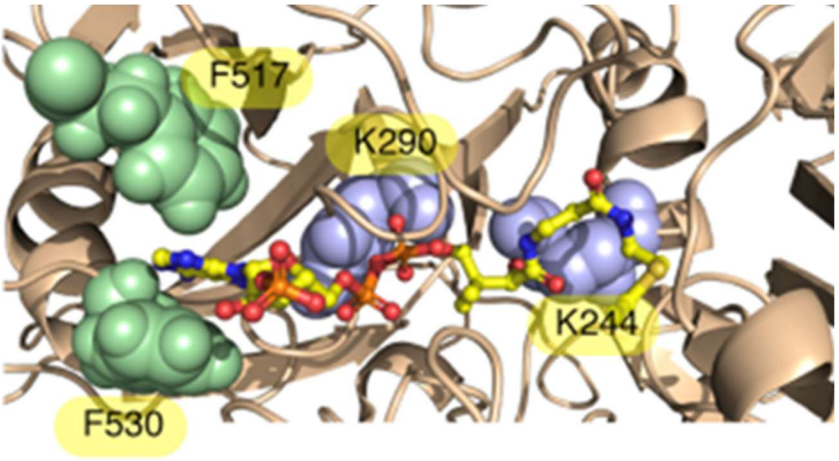



*Figure 5.5 - The docking pose of acetyl-COA in complex with PCK1. Aliphatic residues (purple spheres) and aromatic residues (green spheres) are shown to be in proximities that agree with DEEP-STD NMR analysis.*

# 5.5 Fucosyltransferase-specific inhibition via next generation of fucose mimetics

***Publication:*** *Martin KC, Tricomi J, Corzana F, García-García A, Ceballos-Laita L,* ***Hicks T****, Monaco S, Angulo J, Hurtado-Guerrero R, Richichi B, Sackstein R. Fucosyltransferase-specific inhibition via next generation of fucose mimetics. Chem Commun (Camb). 2021 Jan 28;57(9):1145-1148. doi: 10.1039/d0cc04847j.*

## 5.5.1 Introduction

Fucosyltransferases (FT) are a class of glycosyltransferase that catalyse the transfer of a fucose monosaccharide from donor guanosine diphosphate fucose (GDP-Fuc) to cell surface glycoconjugates. Being able to control the activity of these enzymes through non-toxic, non-gene therapies is an attractive goal for the development of therapeutics. A principal challenge of the discovery of such compounds is the lack of crystallographic data and an understanding of the catalytic mechanism for these enzymes. To overcome this challenge, ligands based on the structure of fucose were synthesised to develop fucose mimetics that may act as inhibitors, competing with GDP-Fuc for the active site (Figure 5.6).

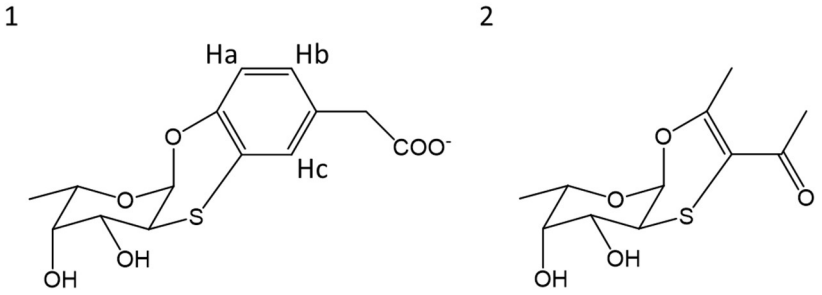



*Figure 5.6 - Fucose mimetics for the inhibition of FUT enzymes. Shown are mimetic 1 (left) and mimetic 2 (right)*

## 5.5.2 Contribution

Though the synthesised fucose mimetics were demonstrated to have activity towards three FTs (FUT6, FUT7 and FUT9), there was not a complete understanding of their potential binding modes. My contribution to this work was to use STD NMR to study the binding of the mimetics to FUT6 and determine if they could compete for the active site by competition STD NMR experiments.

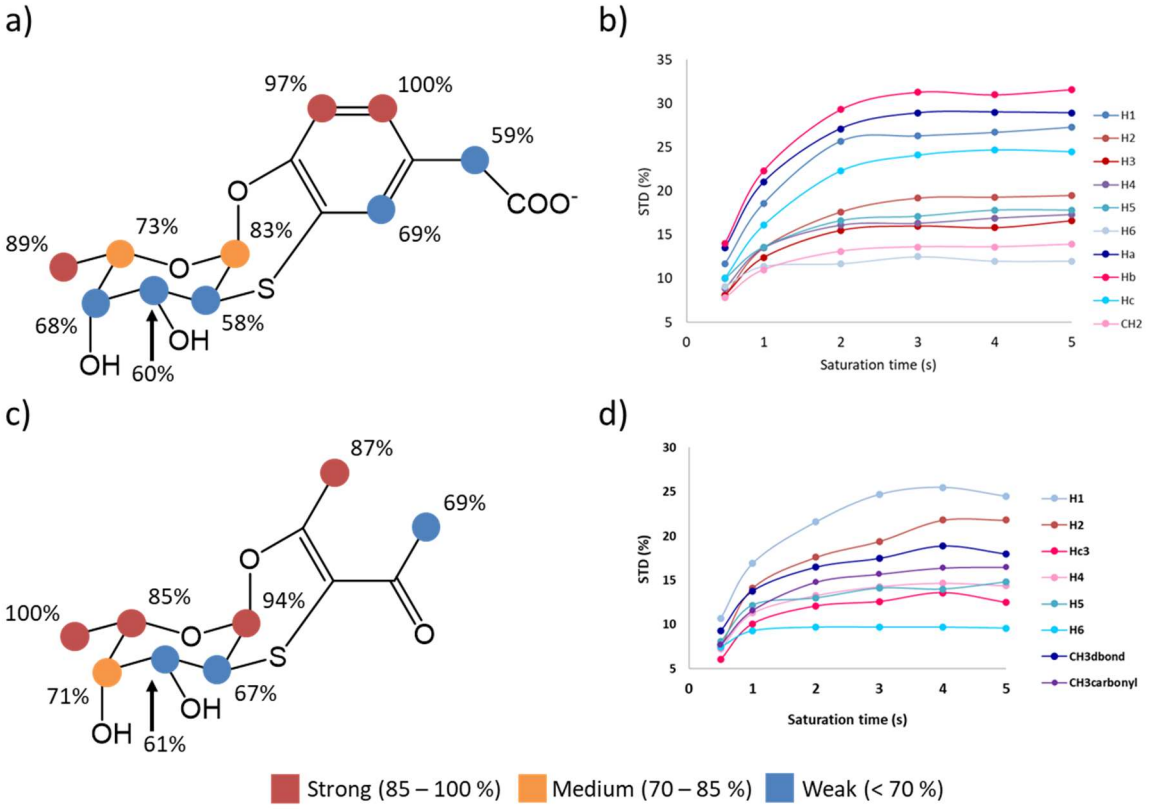



*Figure 5.7 - The binding epitopes of the fucose mimetics for their interactions with FUT6. Shown are the binding epitope (a) and build up curves for mimetic 1 (b) alongside the binding epitope (c) and the build-up curves for mimetic 2 (d). STD epitope intensities are represented by coloured circles with strong (red) medium (orange) and weak (blue) intensities shown.*

From STD NMR the binding epitopes for the two fucose mimetics was found (Figure 5.7). Looking at the STD percentages for both ligands, they were quite high, in the range of ~20%. STD percentages of this magnitude can be indicative of non-specific binding but, given that the $K_d$ of these ligands is in the low mM range, it is more likely that these ligands

are in fast exchange with FUT6. This exchange regime would likely be within the optimal range for a very strong STD effect and therefore larger absolute STD% values. The epitopes show that the modifications to the fucose ring were able to strongly interact with the FUT6. Moreover, the distribution of STD intensities suggested that interactions favoured one face of the ligands. For mimetic 1 the strongest intensities present on Ha and Hb of the aromatic ring as well as H6 and H1 of fucose. For Mimetic 2 there is a similar one-sided distribution of STD intensities with C6, C5 and C1 of the fucose ring and methyl group showing the strongest intensities.

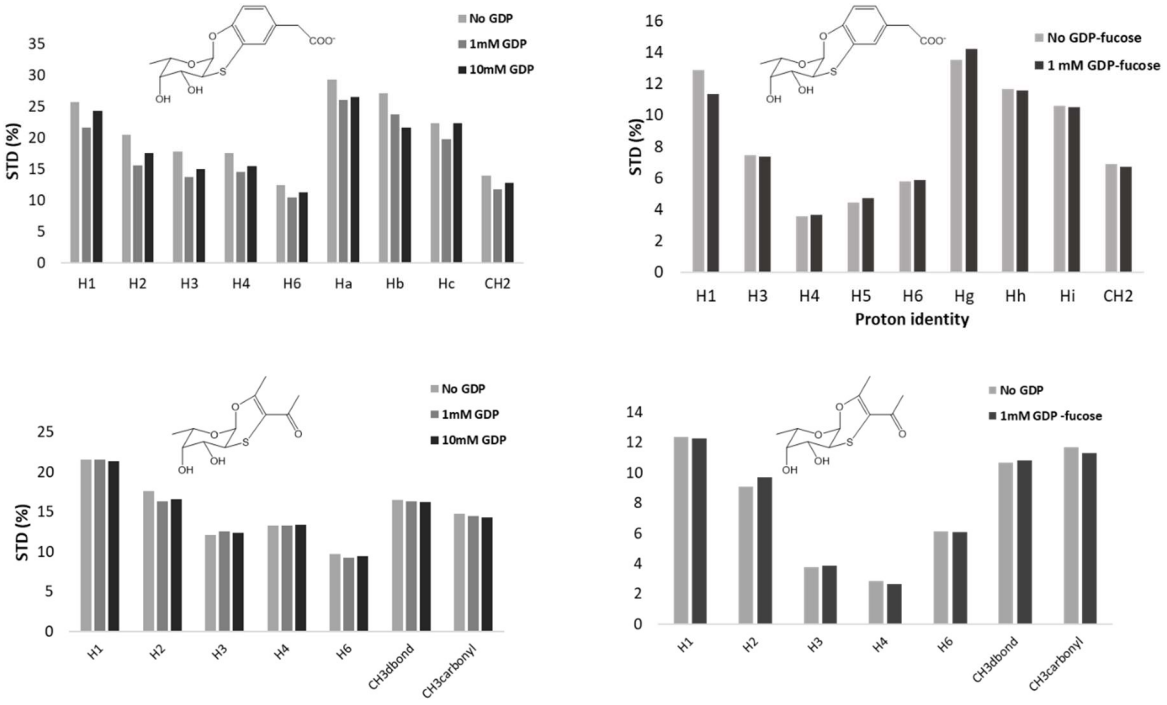



*Figure 5.8 - Fucose mimetic competition STD NMR results. Shown are measured STD intensities for the addition of GDP (left) and GDP-Fuc (right) to mimetic 1 (top) and mimetic 2 (bottom). Experiments were performed with a 2 s saturation time.*

To probe whether the ligands bound in the same site as donor GDP-Fucose, competition experiments were performed. These experiments first measured the STD intensities of the ligands alone and then again on addition of GDP or GDP-Fucose. The results of these experiments (Figure 5.8) showed that the inhibitors did not compete for the same binding site as GDP or GDP-Fucose.

# 5.6 FUT8-Directed Core Fucosylation of N-glycans Is Regulated by the Glycan Structure and Protein Environment

***Publication:*** *García-García A, Serna S, Yang Z, Delso I, Taleb V,* ***Hicks T****, Artschwager R, Vakhrushev SY, Clausen H, Angulo J, Corzana F, Reichardt NC, Hurtado-Guerrero R. FUT8-Directed Core Fucosylation of N-glycans Is Regulated by the Glycan Structure and Protein Environment. ACS Catal. 2021 Aug 6;11(15):9052-9065. doi: 10.1021/acscatal.1c01698.*

## 5.6.1 Introduction

Fucosyltransferase-8 (FUT8) is an enzyme that fucosylates biantennary N-glycans. The fucosylation performed by FUT8 targets the innermost GlcNAc residues of these N-Glycans, transferring a single fucose ring from donor GDP-Fucose and forming an α-1,6-linkage. This process in called core fucosylation. Mutations of this enzyme that disrupt its activity are associated with rare genetic diseases, making it an attractive target to study for the development of medicines. Preferentially, *in vivo*, FUT8 fucosylates the complex N-Glycan G0 (Figure 5.9) but can also fucosylate high-mannose N-Glycans lacking the terminal GlcNAc monosaccharides (Figure 5.9). Important to their recognition, the N-Glycan molecules include a short KVANKT peptide sequence, attached to the innermost GlcNAc monosaccharide by the Asn residue (Figure 5.9). This short peptide is important as without it, fucosylation of these molecules does not occur. This project focussed on the impact that the inclusion of this peptide sequence has on FUT8 recognition for these N-Glycans.

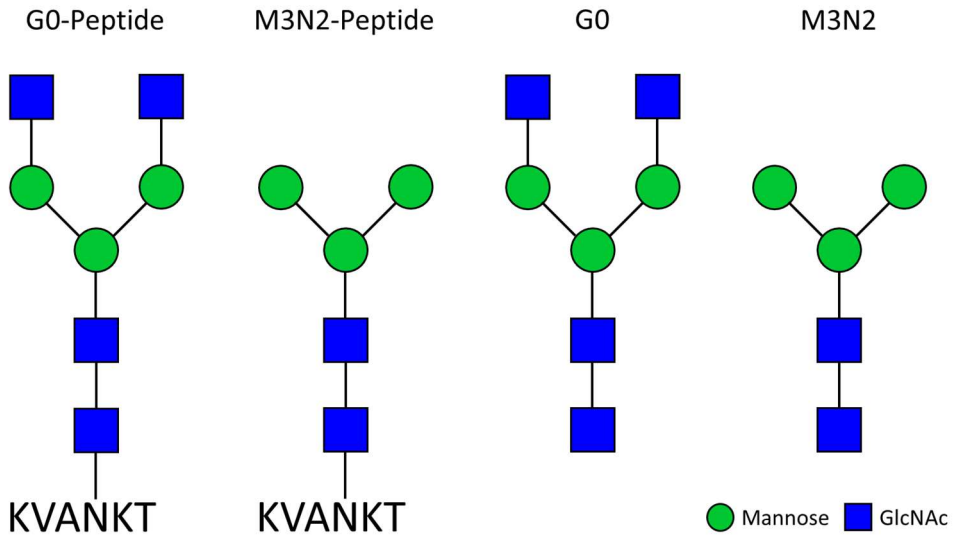



*Figure 5.9 - Schematic representation of the ligands used in this study. Representations follow the SNFG glycan symbol nomenclature. Left to right, G0-peptide, M3N2-peptide, G0 and M3N2.*

## 5.6.2 Contribution

To understand how these complex N-Glycans are recognised by FUT8, my contribution to this publication was to perform STD NMR to find binding epitopes for M3N2, M3N2-peptide, G0-pepide and the short KVANKT peptide alone. The purpose of this being to provide a structural rationale for the recognition preferences that FUT8 has. As FUT8 undergoes structural rearrangement on binding of GDP, STD NMR experiments were performed in both the presence and absence of GDP to determine if the presence of the donor impacted acceptor recognition.

From the STD NMR study it was found that the naked peptide, M3N2 and M3N2-peptide produced little to no STD response with FUT8 even in the presence of GDP. This observation was likely due to the poor affinity of these ligands being outside the optimal ranges for STD NMR. It was possible, however, to measure the full STD NMR epitope for G0-peptide both in the presence and absence of GDP (Figure 5.10). From the STD epitopes, the most strongly recognised part of the molecule is around the innermost GlcNAc monosaccharide. The largest shift in recognition is for the peptide indicating that it is better recognised after the structural rearrangement of the active site, triggered by GDP. Changes are not limited to the peptide with the glycan being overall more strongly recognised.

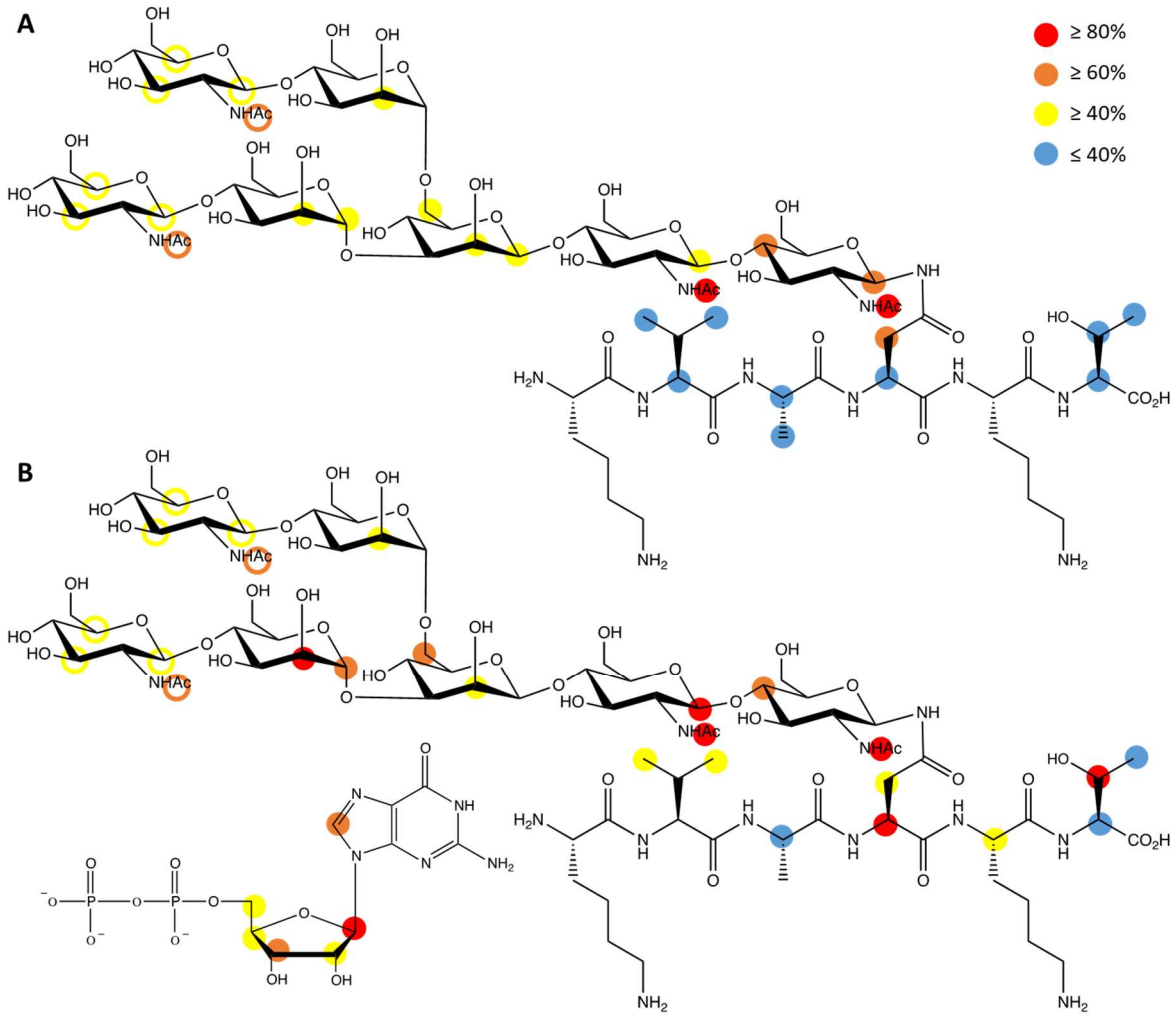



*Figure 5.10 - The binding epitopes of the G0-peptide. Shown are the binding epitopes for G0-peptide (A) and the binding of G0-peptide in the presence of GDP, including the STD NMR epitope for GDP (B). The hollow circles indicate average signals due to overlap in the $^{1}H$ spectra of the terminal GlcNAc rings.*

# 5.7 The human gut symbiont *Ruminococcus gnavus* shows specificity to blood group A antigen during mucin glycan foraging: Implication for niche colonisation in the gastrointestinal tract

***Publication:*** *Wu H, Crost EH, Owen CD, van Bakel W, Martínez Gascueña A, Latousakis D,* ***Hicks T****, Walpole S, Urbanowicz PA, Ndeh D, Monaco S, Sánchez Salom L, Griffiths R, Reynolds RS, Colvile A, Spencer DIR, Walsh M, Angulo J, Juge N. The human gut symbiont Ruminococcus gnavus shows specificity to blood group A antigen during mucin glycan foraging: Implication for niche colonisation in the gastrointestinal tract. PLoS Biol. 2021 Dec 22;19(12):e3001498. doi: 10.1371/journal.pbio.3001498.*

## 5.7.1 Introduction

As mentioned in 5.2, *Ruminococcus* gnavus expresses fucosidases that can recognise fucosylated oligosaccharides. These oligosaccharides that can be used as nutrition, but the molecules that can be digested are strain specific. A separate gene encoding study for a mucin-grown *R. gnavus* predicted that they expressed a blood-group endo-β-1,4-galactosidase, GH98. In response to this, it was hypothesised that a GH98 enzyme would be important for these strains of bacteria, allowing them to use mucins as a source of nutrition. The result of this study found that *Rg*GH98 allowed *R. gnavus* strains to survive in mucin rich environments. It was also found that *Rg*GH98 was able to recognise blood group antigen A (BgA) preferentially over blood group antigens B and H, using an inactive mutant, *Rg*GH98-E411A.

## 5.7.2 Contribution

To determine if the blood group antigens, BgB and BgH did show any binding to *Rg*GH98-E411A, my contribution was to perform competition STD NMR experiments. These experiments were used to determine whether there was any weak binding and if this

binding was in the same site as BgA. The results of these competition experiments found that BgH did not bind to *Rg*GH98-E411A, however it was found that BgB was recognised and did compete with the same site as BgA (Figure 5.11).

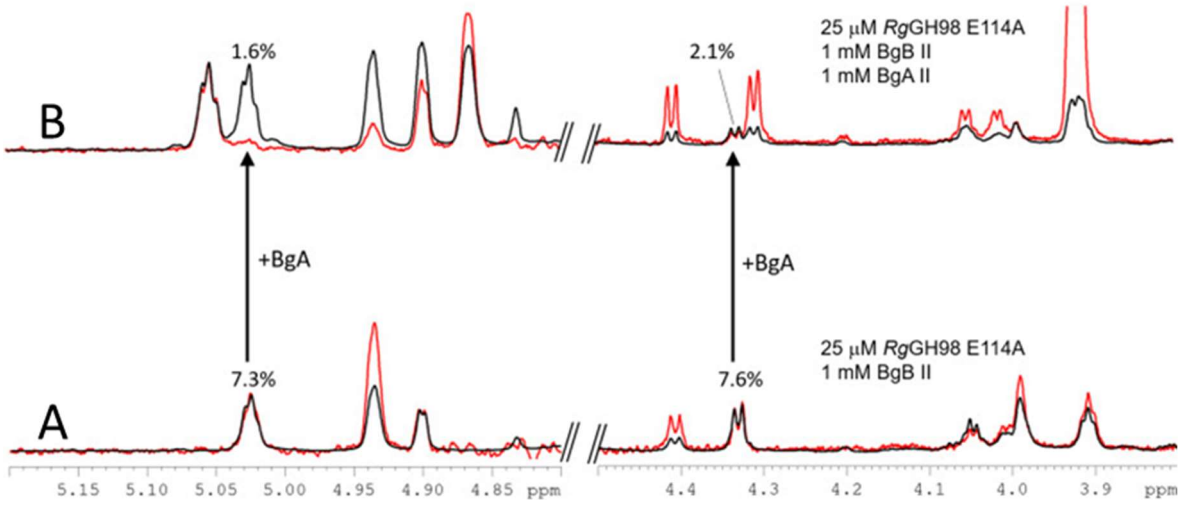



*Figure 5.11 - Competition STD NMR experiments for BgB and BgA in the presence of RgGH98 E411A. The STD intensities for BgB (A) were significantly reduced on addition of BgA (B), leading to a reduction of ~75%. This indicates that BgA can dislodge BgB from RgGH98 E411A and therefore compete for the same binding site.*

# CHAPTER 6

# APPENDIX

# 6.1 List of Abbreviations

| | |
|---|---|
| AMBER | Assisted Model Building with Energy Refinement |
| aMD | Accelerated Molecular Dynamics |
| CHARMM | Chemistry at Harvard Molecular Mechanics |
| COSY | Correlation Spectroscopy |
| Cryo-EM | Cryo-Electron Microscopy |
| CSA | Chemical Shift Anisotropy |
| DD | Death Domain |
| EHEC | Enterohemorrhagic *E. coli* |
| EPEC | Enteropathogenic *E. coli* |
| FADD | Fas-Associated Death Domain |
| GA | Genetic Algorithm |
| Gal | Galactose |
| GalNAc | N-Acetyl-Galactosamine |
| GaMD | Gaussian-accelerated Molecular Dynamics |
| GAPDH | Glyceraldehyde 3-Phosphate Dehydrogenase |
| Glc | Glucose |
| GlcNAc | N-Acetyl-Glucose |
| GshB | Bacterial Glutathione Synthetase |
| GT | Glycosyltransferase |
| GT-A | Glycosyltransferase Type A Fold |
| GT-B | Glycosyltransferase Type B Fold |
| GT-C | Glycosyltransferase Type C Fold |
| GT-D | Glycosyltransferase Type D Fold |

| | |
|---|---|
| GT-E | Glycosyltransferase Type E Fold |
| HIF-1α | Hypoxia-Inducible Factor 1-alpha |
| HLH | Helix Loop Helix |
| HMBC | Heteronuclear Multiple Bond Correlation |
| HSQC | Heteronuclear Single Quantum Correlation |
| ICA | Incremental Construction Algorithm |
| MA | Matching Algorithm |
| MBL | Mannose Binding Lectin |
| MCA | Monte Carlo Algorithm |
| MD | Molecular Dynamics |
| MEM | Minimal Expression Media |
| MM | Molecular Mechanics |
| NF-κB | Nuclear Factor κB |
| NGT | N-Glycosyltransferase |
| NOE | Nuclear Overhauser Effect |
| NMR | Nuclear Magnetic Resonance |
| NPT | Isobaric-Isothermal Ensemble |
| NVE | Microcanonical Ensemble |
| NVT | Canonical Ensemble |
| OGT | O-Glycosyltransferase |
| PBC | Periodic Boundary Condition |
| PES | Potential Energy Surface |
| PMF | Potential Mean Force |
| RIPK1 | Receptor Interacting Serine/Threonine Kinase 1 |

| | |
|---|---|
| RMSD | Root-Mean-Square Deviation |
| RMSF | Root-Mean-Square Fluctuation |
| SF | Scoring Function |
| SS-NOE | Steady-State Nuclear Overhauser Effect |
| STD | Saturation Transfer Difference |
| STD NMR | Saturation Transfer Difference Nuclear Magnetic Resonance |
| T3SS | Type-3 Secretion System |
| T-NOE | Transient Nuclear Overhauser Effect |
| TNFR | Tumour Necrosis Factor Receptor |
| TRADD | Tumour Necrosis Factor Receptor Type 1-Associated Death Domain |
| TRAF | Tumour Necrosis Factor Receptor-Associated Factor |
| TROSY | Transverse Relaxation Optimised Spectroscopy |
| UDP | Uridine Diphosphate |
| VT STD NMR | Variable Temperature Saturation Transfer Difference Nuclear Magnetic Resonance |

# 6.2 Full List of Publications

1) Wallace M, **Hicks T**, Khimyak YZ, Angulo J. Self-Correcting Method for the Measurement of Free Calcium and Magnesium Concentrations by $^{1}$H NMR. Anal Chem. 2019 Nov 19;91(22):14442-14450. doi: 10.1021/acs.analchem.9b03008.

2) Wu H, Rebello O, Crost EH, Owen CD, Walpole S, Bennati-Granier C, Ndeh D, Monaco S, **Hicks T**, Colvile A, Urbanowicz PA, Walsh MA, Angulo J, Spencer DIR, Juge N. Fucosidases from the human gut symbiont Ruminococcus gnavus. Cell Mol Life Sci. 2021 Jan;78(2):675-693. doi: 10.1007/s00018-020-03514-x.

3) Latorre-Muro P, Baeza J, Hurtado-Guerrero R, **Hicks T**, Delso I, Hernández-Ruiz C, Velázquez-Campoy A, Lawton AJ, Angulo J, Denu JM, Carrodeguas JA. Self-acetylation at the active site of phosphoenolpyruvate carboxykinase (PCK1) controls enzyme activity. J Biol Chem. 2021 Jan-Jun;296:100205. doi: 10.1074/jbc.RA120.015103.

4) Martin KC, Tricomi J, Corzana F, García-García A, Ceballos-Laita L, **Hicks T**, Monaco S, Angulo J, Hurtado-Guerrero R, Richichi B, Sackstein R. Fucosyltransferase-specific inhibition via next generation of fucose mimetics. Chem Commun (Camb). 2021 Jan 28;57(9):1145-1148. doi: 10.1039/d0cc04847j.

5) García-García A, Serna S, Yang Z, Delso I, Taleb V, **Hicks T**, Artschwager R, Vakhrushev SY, Clausen H, Angulo J, Corzana F, Reichardt NC, Hurtado-Guerrero R. FUT8-Directed Core Fucosylation of N-glycans Is Regulated by the Glycan Structure and Protein Environment. ACS Catal. 2021 Aug 6;11(15):9052-9065. doi: 10.1021/acscatal.1c01698.

6) García-García A, **Hicks T**, El Qaidi S, Zhu C, Hardwidge PR, Angulo J, Hurtado-Guerrero R. NleB/SseK-catalyzed arginine-glycosylation and enteropathogen virulence are finely tuned by a single variable position contiguous to the catalytic machinery. Chem Sci. 2021 Aug 19;12(36):12181-12191. doi: 10.1039/d1sc04065k.

7) Wu H, Crost EH, Owen CD, van Bakel W, Martínez Gascueña A, Latousakis D, **Hicks T**, Walpole S, Urbanowicz PA, Ndeh D, Monaco S, Sánchez Salom L, Griffiths R, Reynolds RS, Colvile A, Spencer DIR, Walsh M, Angulo J, Juge N. The human gut symbiont Ruminococcus gnavus shows specificity to blood group A antigen during mucin glycan foraging: Implication for niche colonisation in the gastrointestinal tract. PLoS Biol. 2021 Dec 22;19(12):e3001498. doi: 10.1371/journal.pbio.3001498.

8) Huettner I, Krumm SA, Serna S, Brzezicka K, Monaco S, Walpole S, van Diepen A, Allan F, **Hicks T**, Kimuda S, Emery AM; IAVI Protocol C Investigators & The IAVI African HIV Research Network, Landais E, Hokke CH, Angulo J, Reichardt N, Doores KJ. Cross-reactivity of glycan-reactive HIV-1 broadly neutralizing antibodies with parasite glycans. Cell Rep. 2022 Mar 29;38(13):110611. doi: 10.1016/j.celrep.2022.110611.

9) Nepravishta R, Walpole S, **Hicks T**, Muñoz-García JC, Angulo J. Fast Validation of Static and Dynamic 3D Models of Weak Protein Ligand Complexes from STD NMR Spectroscopy. [Preprint] 2022 Oct. doi: 10.26434/chemrxiv-2022-b7s0x

# 6.3 Supplementary Material for Chapter 3

## 6.3.1 SDS PAGE for Proteins used in Chapter 3

$^{15}$N-FADD-DD and was expressed from transformed cells (R. Hurtado-Guerrero Lab) and purified using a Ni-NTA nickel affinity column and then further purified by size exclusion using a HiLoad 16/600 Superdex 200pg column. NleB1 and SseK1-S286Y was provided by the R. Hurtado-Guerrero Lab. SDS page was then performed on the purified protein samples to assess purity (Figure 23).

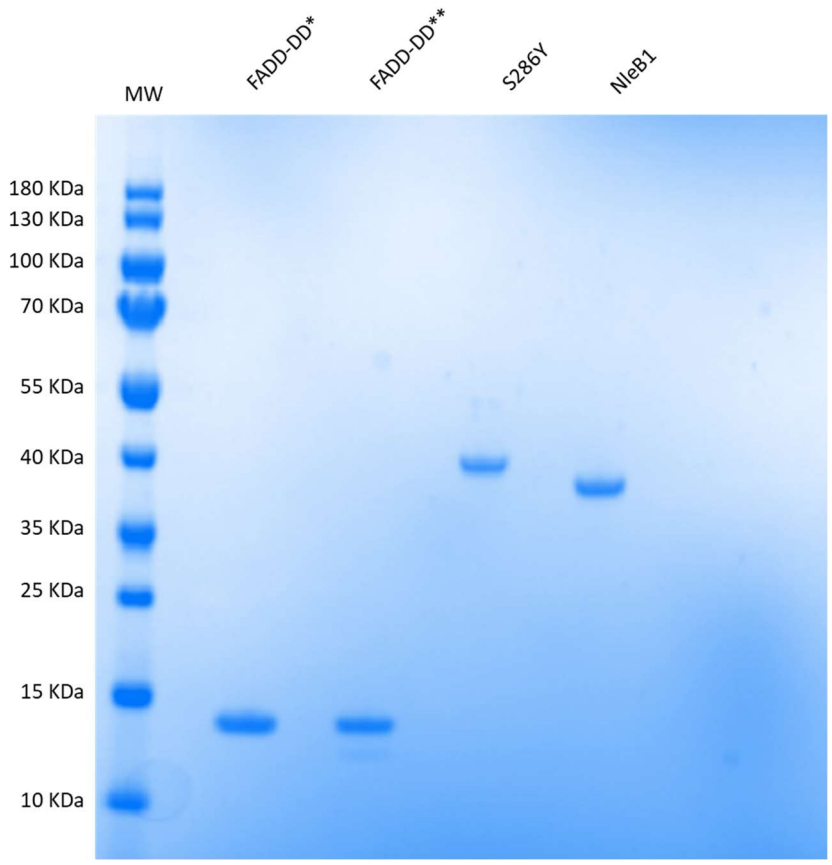



*Figure 6.1 - SDS PAGE of proteins used in chapter 3. Shown are the bands for $^{15}$N-FADD-DD, SseK1-S286Y and NleB1 proteins used in both optimization and final TROSY experiments. $^{15}$N-FADD-DD (FADD-DD* and FADD-DD**, respectively) can be seen at approximately 14 kDa. SseK1-S286Y (S286Y) and NleB1 can be seen at approximately 40 kDa and 38 kDa respectively. TROSY titration results for NleB1 were not recorded due to problems with solubility.*

## 6.3.2 $^{15}$N-FADD-DD $^{1}$H-$^{15}$N HSQC TROSY Assignment

*Table 6.1 - $^{1}$H-$^{15}$N HSQC TROSY assignment for $^{15}$N-FADD-DD. Residues marked with an asterisk (*) could not be assigned.*

| Residue | $^{1}$H Position (ppm) | $^{15}$N Position (ppm) |
|---|---|---|
| **G93** | 8.47 | 110.83 |
| **E94** | 8.11 | 121.31 |
| **E95** | 8.51 | 121.64 |
| **D96** | 8.26 | 122.48 |
| **L97** | 7.93 | 122.99 |
| **C98** | 7.78 | 117.40 |
| **A99** | 7.74 | 122.15 |
| **A100** | 7.74 | 121.73 |
| **F101** | 8.92 | 117.72 |
| **N102** | * | * |
| **V103** | 7.44 | 119.73 |
| **I104** | 8.04 | 121.73 |
| **C105** | 8.50 | 117.44 |
| **D106** | 7.55 | 116.65 |
| **N107** | 7.33 | 115.16 |
| **V108** | 7.96 | 121.87 |
| **G109** | 8.47 | 113.86 |
| **K110** | 8.76 | 125.22 |
| **D111** | 8.32 | 119.49 |
| **W112** | 7.88 | 119.91 |
| **R113** | 6.97 | 123.27 |
| **R114** | 7.25 | 118.84 |
| **L115** | 6.90 | 120.05 |
| **A116** | 8.06 | 121.40 |
| **R117** | 7.94 | 115.77 |
| **Q118** | 7.57 | 123.69 |
| **L119** | 7.72 | 118.33 |
| **K120** | 7.79 | 110.50 |
| **V121** | 7.86 | 123.45 |
| **S122** | 8.38 | 124.25 |
| **D123** | 8.69 | 123.13 |
| **T124** | 7.94 | 112.65 |
| **K125** | 7.50 | 124.80 |
| **I126** | 8.64 | 123.36 |
| **D127** | 7.89 | 120.05 |
| **S128** | 7.61 | 116.05 |

*Table 6.1 - $^{1}H$-$^{15}N$ HSQC TROSY assignment for $^{15}N$-FADD-DD. Residues marked with an asterisk (*) could not be assigned.*

| Residue | $^{1}H$ Position (ppm) | $^{15}N$ Position (ppm) |
|---|---|---|
| **I129** | 8.34 | 125.46 |
| **E130** | 7.99 | 120.15 |
| **D131** | 7.54 | 116.89 |
| **R132** | 7.63 | 118.70 |
| **Y133** | 7.17 | 116.09 |
| **P134** | * | * |
| **R135** | 8.41 | 116.62 |
| **N136** | 7.07 | 118.75 |
| **L137** | 8.39 | 125.41 |
| **T138** | 7.80 | 114.37 |
| **E139** | 7.15 | 119.35 |
| **R140** | 6.84 | 117.77 |
| **V141** | 7.24 | 120.19 |
| **R142** | 8.08 | 121.12 |
| **E143** | 8.11 | 118.89 |
| **S144** | 7.72 | 116.65 |
| **L145** | 7.48 | 121.50 |
| **R146** | 7.96 | 121.26 |
| **I147** | 8.28 | 119.63 |
| **W148** | 8.15 | 125.85 |
| **K149** | 8.82 | 121.08 |
| **N150** | 8.24 | 11.749 |
| **T151** | 8.01 | 116.37 |
| **E152** | 8.28 | 119.68 |
| **K153** | 7.81 | 118.84 |
| **E154** | 9.31 | 125.50 |
| **N155** | 8.03 | 116.51 |
| **A156** | 7.70 | 127.97 |
| **T157** | 7.13 | 112.69 |
| **V158** | 9.06 | 123.59 |
| **A159** | 8.56 | 121.36 |
| **H160** | 7.94 | 117.96 |
| **L161** | 8.29 | 124.99 |
| **V162** | 8.70 | 119.82 |
| **G163** | 7.86 | 105.84 |
| **A164** | 7.89 | 126.34 |
| **L165** | 8.02 | 119.03 |
| **R166** | 8.30 | 118.28 |
| **S167** | 8.23 | 118.10 |

*Table 6.1 - $^{1}H$-$^{15}N$ HSQC TROSY assignment for $^{15}N$-FADD-DD. Residues marked with an asterisk (*) could not be assigned.*

| Residue | $^{1}H$ Position (ppm) | $^{15}N$ Position (ppm) |
|---|---|---|
| **C168** | 7.28 | 115.44 |
| **Q169** | 7.74 | 113.39 |
| **M170** | 8.09 | 122.85 |
| **N171** | 7.36 | 117.96 |
| **L172** | * | * |
| **V173** | 7.24 | 117.96 |
| **A174** | 7.91 | 122.01 |
| **D175** | 8.13 | 118.38 |
| **L176** | 7.68 | 121.78 |
| **V177** | 8.35 | 119.26 |
| **Q178** | 8.35 | 119.31 |
| **E179** | 7.94 | 120.05 |
| **V180** | 7.75 | 120.66 |
| **Q181** | 7.95 | 120.89 |
| **Q182** | 8.22 | 119.21 |
| **A183** | 7.78 | 122.80 |
| **R184** | 7.90 | 120.24 |
| **D185** | 8.33 | 119.54 |
| **L186** | 7.85 | 121.50 |
| **Q187** | 7.93 | 119.49 |
| **N188** | * | * |
| **R189** | 7.96 | 121.64 |
| **S190** | 8.19 | 117.12 |
| **G191** | 8.24 | 111.99 |
| **A192** | 7.70 | 129.79 |

# 6.4 Supplementary Material for Chapter 4

## 6.4.1 SDS PAGE for Proteins used in Chapter 4

SseK1-WT, SseK1-S286Y and SseK2-WT were expressed from transformed cells (R. Hurtado-Guerrero Lab) and purified using a Ni-NTA nickel affinity column and then further purified by size exclusion using a HiLoad 16/600 Superdex 200pg column. SDS page was then performed on the purified protein samples to assess purity (Figure 23). NleB1 was provided by the R. Hurtado-Guerrero Lab.

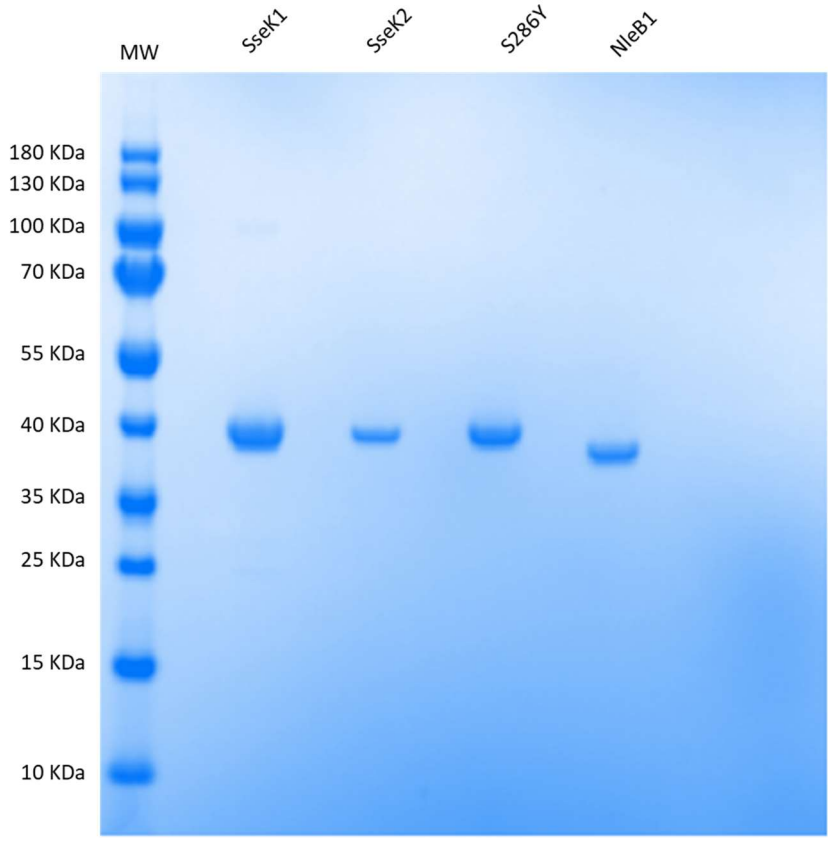



*Figure 6.2 - SDS PAGE of proteins used in chapter 4. Shown are the bands for SseK1, SseK2, mutant SseK1-S286Y and NleB1. SseK1, SseK2 and mutant SseK1-S286Y can all be seen at approximately 40 kDa and NleB1 can be seen at approximately 38 kDa. STD NMR results for NleB1 were not recorded due to problems with solubility.*

## 6.4.2 Assignment of UDP, UDP-GlcNAc and Other Donors

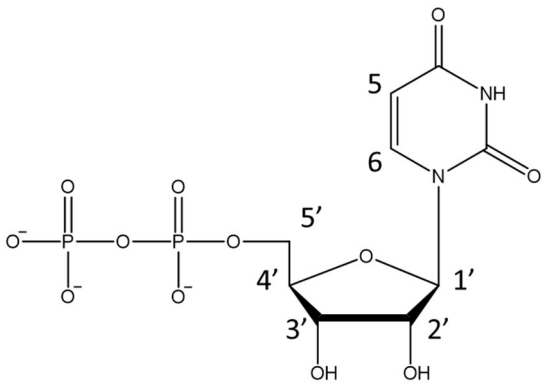



*Figure 6.3 - Structure of UDP showing the assigned positions. To discern between the different substructures the signals for the ribose are denoted with single apostrophe (').*

*Table 6.2 - Assignment for UDP recorded at 288 K. Measured using $^{1}H$-$^{1}H$ COSY and $^{1}H$-$^{13}C$ HSQC.*

| Proton | $^{1}H$ Position (ppm) | $^{13}C$ Position (ppm) |
|---|---|---|
| **H5** | 5.78 | 102.35 |
| **H6** | 7.81 | 141.68 |
| **H1'** | 5.81 | 88.16 |
| **H2'** | 4.22 | 69.38 |
| **H3'** | 4.20 | 73.83 |
| **H4'** | 4.10 | 82.97 |
| **H5'** | 4.04 | 64.41 |

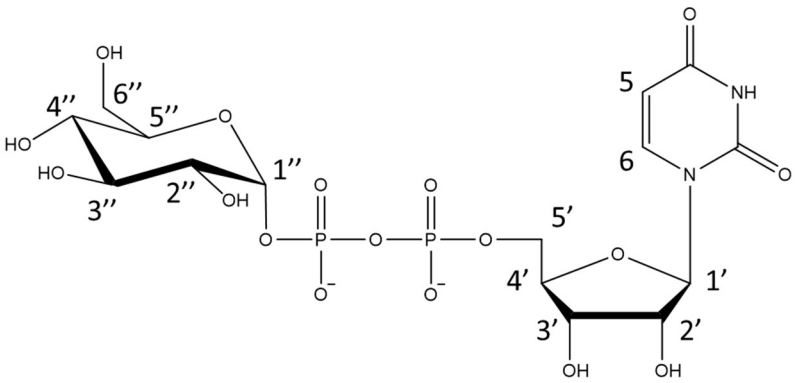


*Figure 6.4 - Structure of UDP-Glc showing the assigned positions. To discern between the different substructures the signals for the ribose are denoted with single apostrophe (‘) and the Glucose ring with double apostrophe (‘’).*

*Table 6.3 - Assignment for UDP-Glc recorded at 288 K. Measured using $^{1}$H-$^{1}$H COSY and $^{1}$H-$^{13}$C HSQC. Signals for H2’ and H3’ overlap and appear in the same $^{1}$H position.*

| Proton | $^{1}$H Position (ppm) | $^{13}$C Position (ppm) |
|---|---|---|
| **H5** | 5.79 | 102.66 |
| **H6** | 7.79 | 141.57 |
| **H1’** | 5.81 | 88.06 |
| **H2’** | 4.19 | 69.36 / 73.73 |
| **H3’** | 4.19 | 69.36 / 73.73 |
| **H4’** | 4.11 | 83.15 |
| **H5’a** | 4.07 | 64.93 |
| **H5’b** | 4.02 | 64.93 |
| **H1’’** | 5.42 | 95.49 |
| **H2’’** | 3.35 | 71.48 |
| **H3’’** | 3.69 | 73.00 |
| **H4’’** | 3.71 | 72.71 |
| **H5’’** | 3.28 | 69.27 |
| **H6’’a** | 3.67 | 60.23 |
| **H6’’b** | 3.58 | 60.23 |

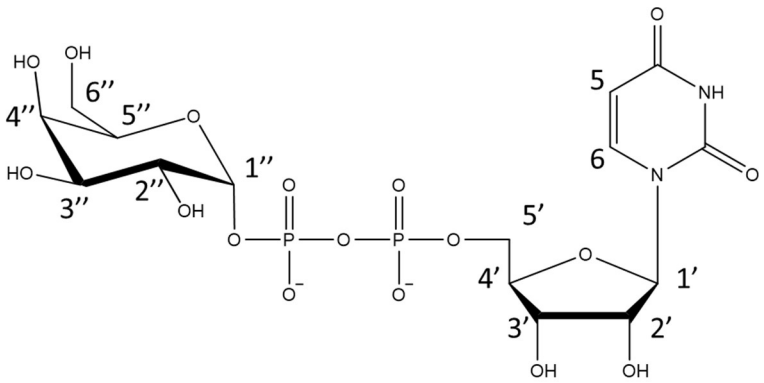



*Figure 6.5 - Structure of UDP-Gal showing the assigned positions. To discern between the different substructures the signals for the ribose are denoted with single apostrophe (') and the Galactose ring with double apostrophe ('').*

*Table 6.4 - Assignment for UDP-Gal recorded at 288 K. Measured using $^{1}$H-$^{1}$H COSY and $^{1}$H-$^{13}$C HSQC. Signals for H2' and H3' overlap and appear in the same $^{1}$H position.*

| Proton | $^{1}$H Position (ppm) | $^{13}$C Position (ppm) |
|---|---|---|
| **H5** | 5.79 | 102.44 |
| **H6** | 7.78 | 141.23 |
| **H1'** | 5.81 | 88.02 |
| **H2'** | 4.19 | 69.59 / 72.13 |
| **H3'** | 4.19 | 69.59 / 72.13 |
| **H4'** | 4.11 | 83.16 |
| **H5'a** | 4.07 | 64.48 |
| **H5'b** | 4.02 | 64.48 |
| **H1''** | 5.45 | 97.02 |
| **H2''** | 3.62 | 68.02 |
| **H3''** | 3.73 | 69.55 |
| **H4''** | 3.84 | 70.36 |
| **H5''** | 3.99 | 72.69 |
| **H6''a** | 3.57 | 60.62 |
| **H6''b** | 3.54 | 60.62 |

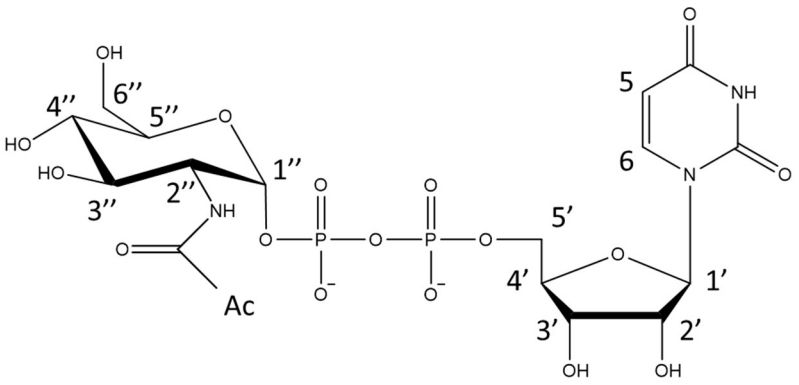



*Figure 6.6 - Structure of UDP-GlcNAc showing the assigned positions. To discern between the different substructures the signals for the ribose are denoted with single apostrophe (') and the N-Acetyl-Glucose ring with double apostrophe ('').*

*Table 6.5 - Assignment for UDP-GlcNAc recorded at 288 K. Measured using $^{1}$H-$^{1}$H COSY and $^{1}$H-$^{13}$C HSQC. Signals for H2' and H3' overlap and appear in the same $^{1}$H position.*

| Proton | $^{1}$H Position (ppm) | $^{13}$C Position (ppm) |
|---|---|---|
| **H5** | 5.78 | 102.66 |
| **H6** | 7.79 | 141.58 |
| **H1'** | 5.80 | 88.65 |
| **H2'** | 4.19 | 69.73 / 73.86 |
| **H3'** | 4.19 | 69.73 / 73.86 |
| **H4'** | 4.11 | 83.25 |
| **H5'a** | 4.07 | 65.25 |
| **H5'b** | 4.01 | 65.25 |
| **H1''** | 5.33 | 94.41 |
| **H2''** | 3.81 | 53.79 |
| **H3''** | 3.64 | 70.67 |
| **H4''** | 3.75 | 73.08 |
| **H5''** | 3.38 | 69.46 |
| **H6''a** | 3.69 | 60.09 |
| **H6''b** | 3.62 | 60.09 |
| **HAc** | 1.90 | 21.79 |

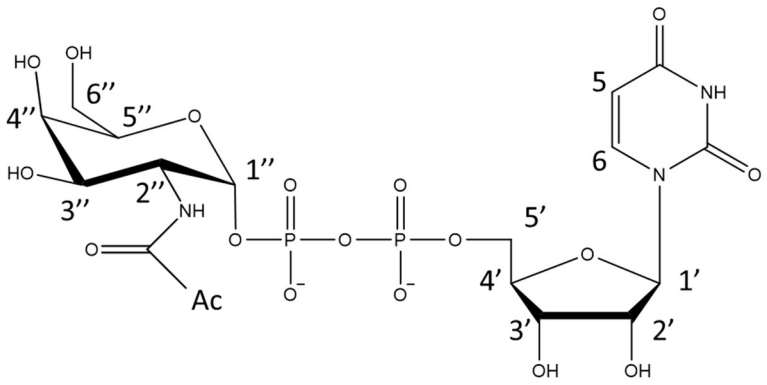



*Figure 6.7 - Structure of UDP-GalNAc showing the assigned positions. To discern between the different substructures the signals for the ribose are denoted with single apostrophe (') and the N-Acetyl-Galactose ring with double apostrophe ('').*

*Table 6.6 - Assignment for UDP-GalNAc recorded at 288 K. Measured using $^{1}H$-$^{1}H$ COSY and $^{1}H$-$^{13}C$ HSQC. Signals for H2' and H3' overlap and H5'b and H2'' overlap appearing in the same $^{1}H$ position.*

| Proton | $^{1}H$ Position (ppm) | $^{13}C$ Position (ppm) |
|---|---|---|
| **H5** | 5.781 | 102.28 |
| **H6** | 7.78 | 141.39 |
| **H1'** | 5.802 | 88.24 |
| **H2'** | 4.185 | 69.61 / 73.38 |
| **H3'** | 4.185 | 69.61 / 73.38 |
| **H4'** | 4.107 | 83.12 |
| **H5'a** | 4.07 | 65.01 |
| **H5'b** | 4.01 | 65.01 |
| **H1''** | 5.36 | 94.41 |
| **H2''** | 4.01 | 71.80 |
| **H3''** | 3.78 | 67.65 |
| **H4''** | 3.86 | 68.29 |
| **H5''** | 4.11 | 83.15 |
| **H6''a** | 3.59 | 60.81 |
| **H6''b** | 3.56 | 60.81 |
| **HAc** | 1.90 | 21.92 |

### 6.4.3 STD NMR Data

#### 6.4.3.1 SseK1-WT Build-Up Curves and Calculated Parameters

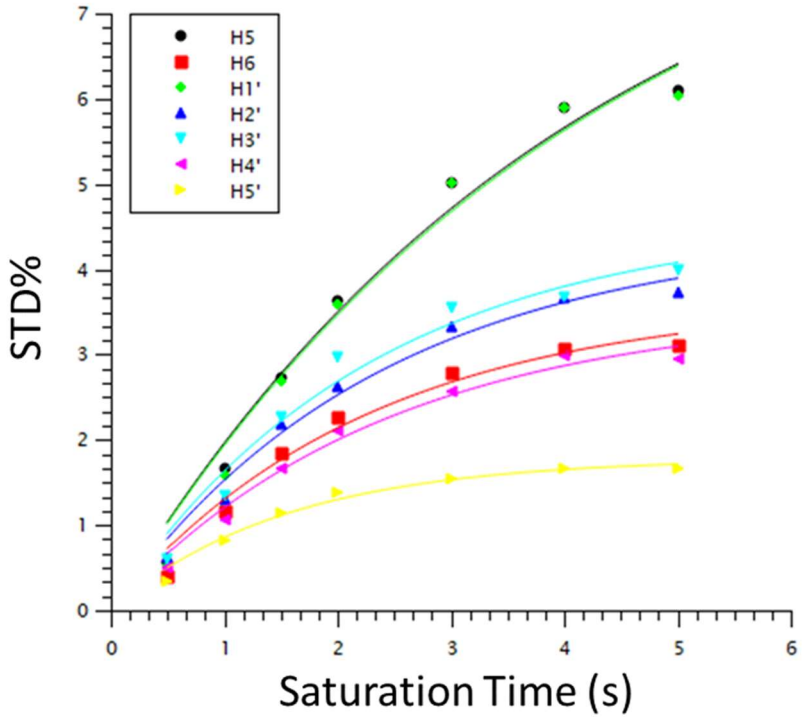



*Figure 6.8 - Experimental STD NMR build-up curves for 1 mM UDP in the presence of 25 µM SseK1-WT and 250 µM $MgCl_2$. STD NMR experiments were performed in a deuterated water buffer containing 25 mM TRIS-$D_{11}$, 150 mM NaCl, 250 µM $MgCl_2$, pH 7.5. Curves were fit to equation 14 using least squares.*

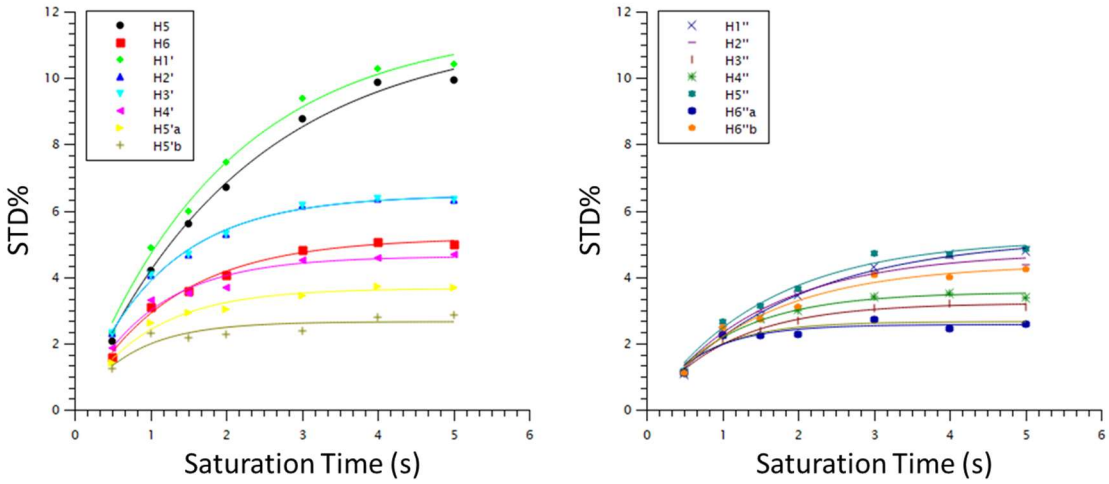



*Figure 6.9 - Experimental STD NMR build-up curves for 1 mM UDP-Glc in the presence of 25 µM SseK1-WT and 250 µM $MgCl_2$. STD NMR experiments were performed in a deuterated water buffer containing 25 mM TRIS-$D_{11}$, 150 mM NaCl, 250 µM $MgCl_2$, pH 7.5. Curves were fit to equation 14 using least squares. For clarity the build-up curves for uridine (left) and Glucose (right) are shown separately.*

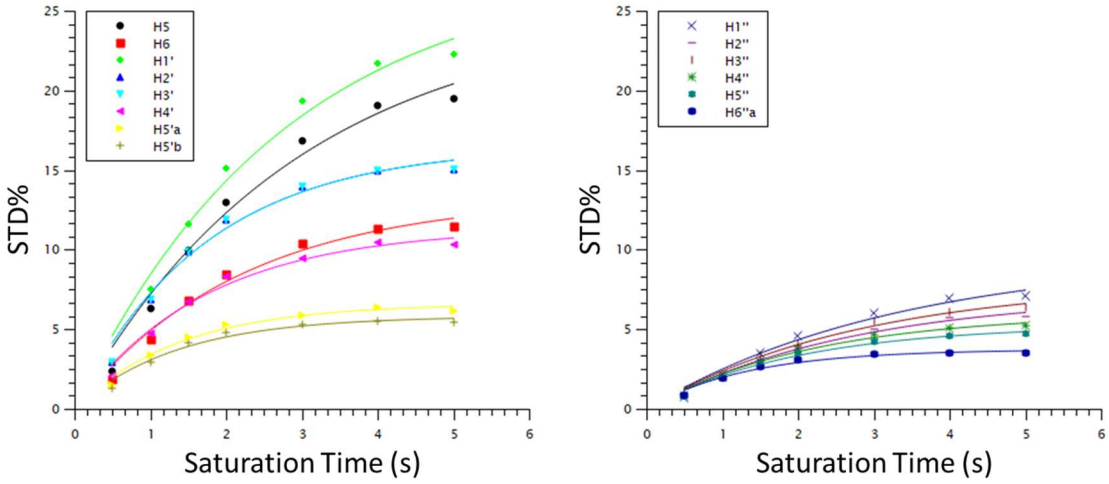



*Figure 6.10 - Experimental STD NMR build-up curves for 1 mM UDP-Gal in the presence of 25 µM SseK1-WT. STD NMR experiments were performed in a deuterated water buffer containing 25 mM TRIS-$D_{11}$, 150 mM NaCl, 250 µM $MgCl_2$, pH 7.5. Curves were fit to equation 14 using least squares. For clarity the build-up curves for uridine (left) and Glucose (right) are shown separately.*

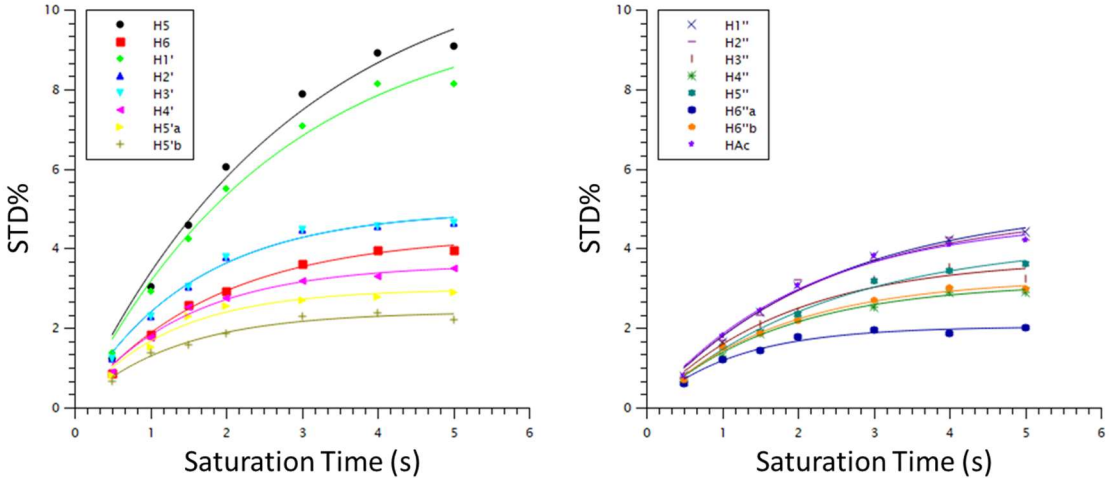



*Figure 6.11 - Experimental STD NMR build-up curves for 1 mM UDP-GlcNAc in the presence of 25 µM SseK1-WT and 250 µM $MgCl_2$. STD NMR experiments were performed in a deuterated water buffer containing 25 mM TRIS-$D_{11}$, 150 mM NaCl, 250 µM $MgCl_2$, pH 7.5. Curves were fit to equation 14 using least squares. For clarity the build-up curves for uridine (left) and Glucose (right) are shown separately.*

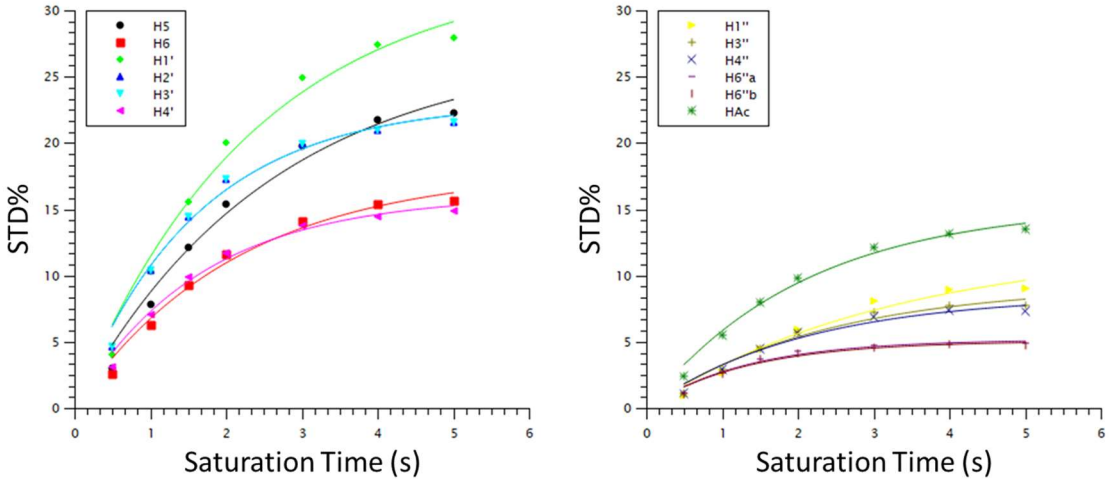



*Figure 6.12 - Experimental STD NMR build-up curves for 1 mM UDP-GalNAc in the presence of 25 µM SseK1-WT and 250 µM $MgCl_2$. STD NMR experiments were performed in a deuterated water buffer containing 25 mM TRIS-$D_{11}$, 150 mM NaCl, 250 µM MgCl2, pH 7.5. Curves were fit to equation 14 using least squares. For clarity the build-up curves for uridine (left) and Glucose (right) are shown separately.*

*Table 6.7 - Calculated parameters from fitting of STD NMR data for binding of UDP to SseK1-WT in the presence of 250 µM $MgCl_2$. $STD_{max}$ and $k_{sat}$ were derived from binding epitope fitting, $STD_0$ was found using equation 14. Normalised intensities (Norm) were calculated from the highest calculated $STD_0$ (bold). STD NMR experiments were performed in a deuterated water buffer containing 25 mM TRIS-$D_{11}$, 150 mM NaCl, 250 µM $MgCl_2$, pH 7.5.*

| Proton | $STD_{max}$ (%) | $k_{sat}$ ($s^{-1}$) | $STD_0$ (% $s^{-1}$) | Norm (%) |
|---|---|---|---|---|
| **H5** | 9.15 | 0.24 | **2.20** | 100 |
| **H6** | 3.64 | 0.44 | 1.61 | 73 |
| **H1'** | 9.23 | 0.24 | 2.17 | 98.5 |
| **H2'** | 4.43 | 0.42 | 1.87 | 85 |
| **H3'** | 4.59 | 0.44 | 2.02 | 91.5 |
| **H4'** | 3.54 | 0.42 | 1.47 | 67 |
| **H5'** | 1.78 | 0.65 | 1.16 | 53 |

*Table 6.8 - Calculated parameters from fitting of STD NMR data for binding of UDP-Glc to SseK1-WT in the presence of 250 µM $MgCl_2$. $STD_{max}$ and $k_{sat}$ were derived from binding epitope fitting, $STD_0$ was found using equation 14. Normalised intensities (Norm) were calculated from the highest calculated $STD_0$ (bold). STD NMR experiments were performed in a deuterated water buffer containing 25 mM TRIS-$D_{11}$, 150 mM NaCl, 250 µM $MgCl_2$, pH 7.5.*

| Proton | $STD_{max}$ (%) | $k_{sat}$ ($s^{-1}$) | $STD_0$ (% $s^{-1}$) | Norm (%) |
|---|---|---|---|---|
| **H5** | 11.44 | 0.45 | 5.20 | 87 |
| **H6** | 5.18 | 0.82 | 4.24 | 71 |
| **H1'** | 11.58 | 0.52 | **5.97** | 100 |
| **H2'** | 6.47 | 0.91 | 5.89 | 99 |
| **H3'** | 6.47 | 0.91 | 5.89 | 99 |
| **H4'** | 4.62 | 1.04 | 4.79 | 80 |
| **H5'1** | 3.66 | 1.07 | 3.92 | 66 |
| **H5'2** | 2.65 | 1.37 | 3.63 | 61 |
| **H1''** | 5.20 | 0.55 | 2.84 | 48 |
| **H2''** | 4.74 | 0.67 | 3.16 | 53 |
| **H3''** | 3.21 | 0.94 | 3.03 | 51 |
| **H4''** | 3.54 | 0.95 | 3.35 | 56 |
| **H5''** | 5.15 | 0.65 | 3.33 | 56 |
| **H6''1** | 2.56 | 1.45 | 3.71 | 62 |
| **H6''2** | 4.38 | 0.69 | 3.03 | 51 |

*Table 6.9 - Calculated parameters from fitting of STD NMR data for binding of UDP-Gal to SseK1-WT in the presence of 250 µM $MgCl_2$. $STD_{max}$ and $k_{sat}$ were derived from binding epitope fitting, $STD_0$ was found using equation 14. Normalised intensities (Norm) were calculated from the highest calculated $STD_0$ (bold). STD NMR experiments were performed in a deuterated water buffer containing 25 mM TRIS-$D_{11}$, 150 mM NaCl, 250 µM $MgCl_2$, pH 7.5.*

| Proton | $STD_{max}$ (%) | $k_{sat}$ ($s^{-1}$) | $STD_0$ (% $s^{-1}$) | Norm (%) |
|---|---|---|---|---|
| **H5** | 25.02 | 0.34 | 8.47 | 84 |
| **H6** | 13.32 | 0.46 | 6.10 | 61 |
| **H1'** | 27.82 | 0.36 | **10.04** | 100 |
| **H2'** | 16.53 | 0.58 | 9.56 | 95 |
| **H3'** | 16.53 | 0.58 | 9.56 | 95 |
| **H4'** | 11.37 | 0.58 | 6.55 | 65 |
| **H5'1** | 6.61 | 0.72 | 4.77 | 47 |
| **H5'2** | 5.82 | 0.75 | 4.35 | 43 |
| **H1''** | 9.57 | 0.30 | 2.89 | 29 |
| **H2''** | 7.18 | 0.37 | 2.66 | 26 |
| **H3''** | 7.98 | 0.35 | 2.81 | 28 |
| **H4''** | 6.01 | 0.45 | 2.73 | 27 |
| **H5''** | 5.27 | 0.50 | 2.66 | 26 |
| **H6''1** | 3.73 | 0.76 | 2.82 | 28 |
| **H6''2** | * | * | * | * |

*Table 6.10 - Calculated parameters from fitting of STD NMR data for binding of UDP-GlcNAc to SseK1-WT in the presence of 250 µM $MgCl_2$. $STD_{max}$ and $k_{sat}$ were derived from binding epitope fitting, $STD_0$ was found using equation 14. Normalised intensities (Norm) were calculated from the highest calculated $STD_0$ (bold). STD NMR experiments were performed in a deuterated water buffer containing 25 mM TRIS-$D_{11}$, 150 mM NaCl, 250 µM $MgCl_2$, pH 7.5.*

| Proton | $STD_{max}$ (%) | $k_{sat}$ ($s^{-1}$) | $STD_0$ (% $s^{-1}$) | Norm (%) |
|---|---|---|---|---|
| **H5** | 11.55 | 0.35 | **4.00** | 100 |
| **H6** | 4.34 | 0.56 | 2.44 | 61 |
| **H1'** | 10.09 | 0.38 | 3.79 | 95 |
| **H2'** | 4.96 | 0.66 | 3.25 | 81 |
| **H3'** | 4.96 | 0.66 | 3.25 | 81 |
| **H4'** | 3.60 | 0.70 | 2.52 | 63 |
| **H5'1** | 2.98 | 0.82 | 2.44 | 61 |
| **H5'2** | 2.40 | 0.77 | 1.84 | 46 |
| **H1''** | 5.12 | 0.43 | 2.19 | 55 |
| **H2''** | 4.93 | 0.45 | 2.23 | 55 |
| **H3''** | 3.73 | 0.55 | 2.06 | 51 |
| **H4''** | 3.13 | 0.58 | 1.82 | 45 |
| **H5''** | 4.20 | 0.42 | 1.78 | 45 |
| **H6''1** | 2.03 | 0.86 | 1.75 | 44 |
| **H6''2** | 3.24 | 0.58 | 1.87 | 47 |
| **HAc** | 4.75 | 0.49 | 2.33 | 58 |

*Table 6.11 - Calculated parameters from fitting of STD NMR data for binding of UDP-GalNAc to SseK1-WT in the presence of 250 µM $MgCl_2$. $STD_{max}$ and $k_{sat}$ were derived from binding epitope fitting, $STD_0$ was found using equation 14. Normalised intensities (Norm) were calculated from the highest calculated $STD_0$ (bold). STD NMR experiments were performed in a deuterated water buffer containing 25 mM TRIS-$D_{11}$, 150 mM NaCl, 250 µM $MgCl_2$, pH 7.5.*

| Proton | $STD_{max}$ (%) | $k_{sat}$ ($s^{-1}$) | $STD_0$ (% $s^{-1}$) | Norm (%) |
|---|---|---|---|---|
| **H5** | 27.19 | 0.39 | 10.52 | 73 |
| **H6** | 17.94 | 0.47 | 8.45 | 59 |
| **H1'** | 33.19 | 0.42 | 13.97 | 97 |
| **H2'** | 23.10 | 0.62 | **14.39** | 100 |
| **H3'** | 23.10 | 0.62 | **14.39** | 100 |
| **H4'** | 16.03 | 0.60 | 9.68 | 67 |
| **H5'1** | * | * | * | * |
| **H5'2** | * | * | * | * |
| **H1''** | 12.29 | 0.31 | 3.75 | 26 |
| **H2''** | * | * | * | * |
| **H3''** | 9.29 | 0.43 | 4.01 | 2 |
| **H4''** | 8.51 | 0.48 | 4.09 | 28 |
| **H5''** | * | * | * | * |
| **H6''1** | 5.17 | 0.75 | 3.89 | 27 |
| **H6''2** | 5.05 | 0.75 | 3.79 | 26 |
| **HAc** | 15.35 | 0.48 | 7.32 | 51 |

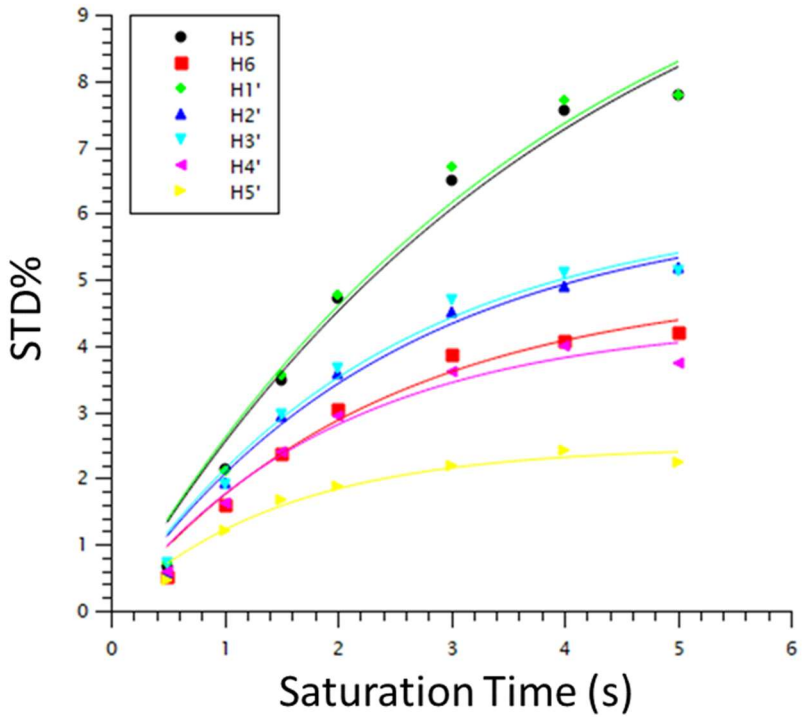



*Figure 6.13 - Experimental STD NMR build-up curves for 1 mM UDP in the presence of 25 µM SseK1-WT and 25 µM $MgCl_2$. STD NMR experiments were performed in a deuterated water buffer containing 25 mM TRIS-$D_{11}$, 150 mM NaCl, 25 µM $MgCl_2$, pH 7.5. Curves were fit to equation 14 using least squares.*

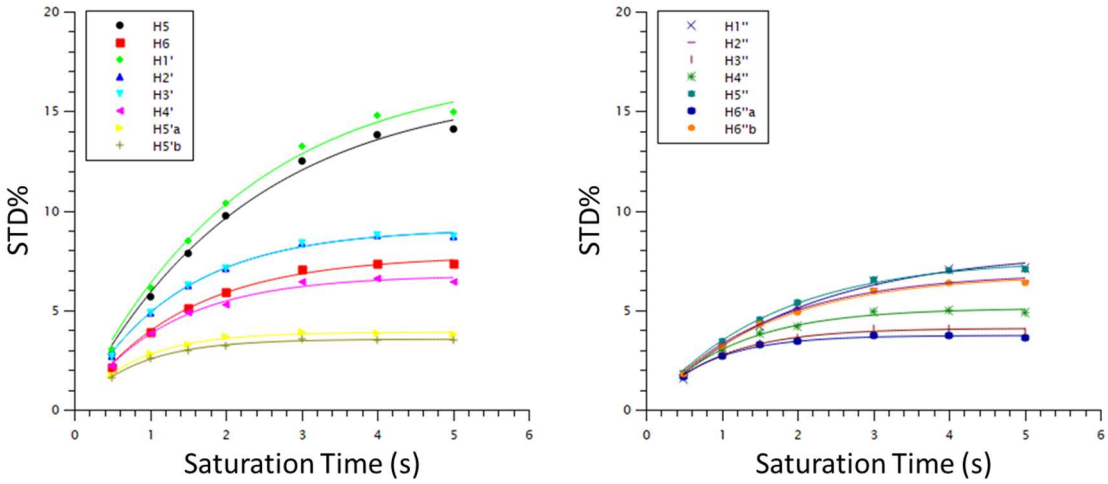



*Figure 6.14 - Experimental STD NMR build-up curves for 1 mM UDP-Glc in the presence of 25 µM SseK1-WT and 25 µM $MgCl_2$. STD NMR experiments were performed in a deuterated water buffer containing 25 mM TRIS-$D_{11}$, 150 mM NaCl, 25 µM $MgCl_2$, pH 7.5. Curves were fit to equation 14 using least squares. For clarity the build-up curves for uridine (left) and Glucose (right) are shown separately.*

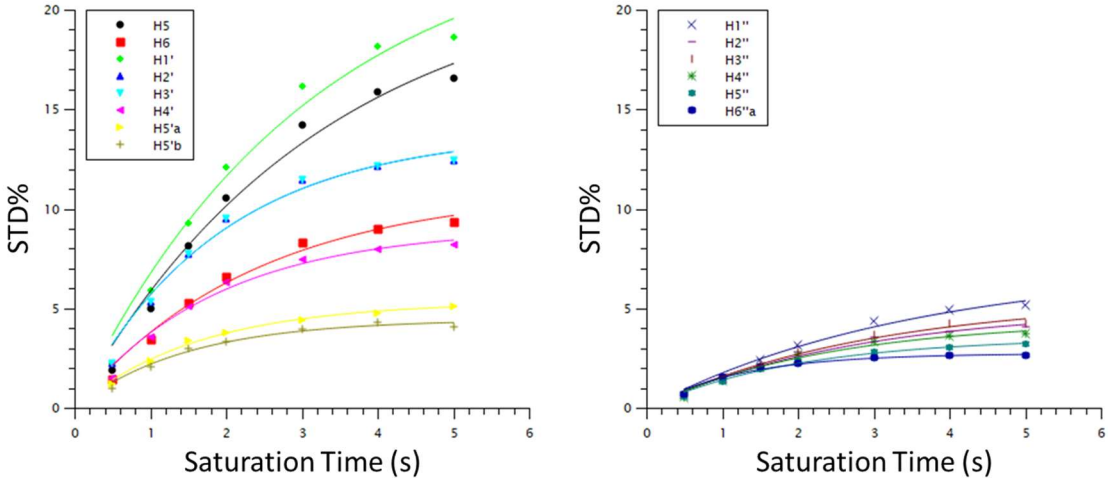



*Figure 6.15 - Experimental STD NMR build-up curves for 1 mM UDP-Gal in the presence of 25 µM SseK1-WT and 25 µM $MgCl_2$. STD NMR experiments were performed in a deuterated water buffer containing 25 mM TRIS-$D_{11}$, 150 mM NaCl, 25 µM $MgCl_2$, pH 7.5. Curves were fit to equation 14 using least squares. For clarity the build-up curves for uridine (left) and Glucose (right)* ***are shown separately.***

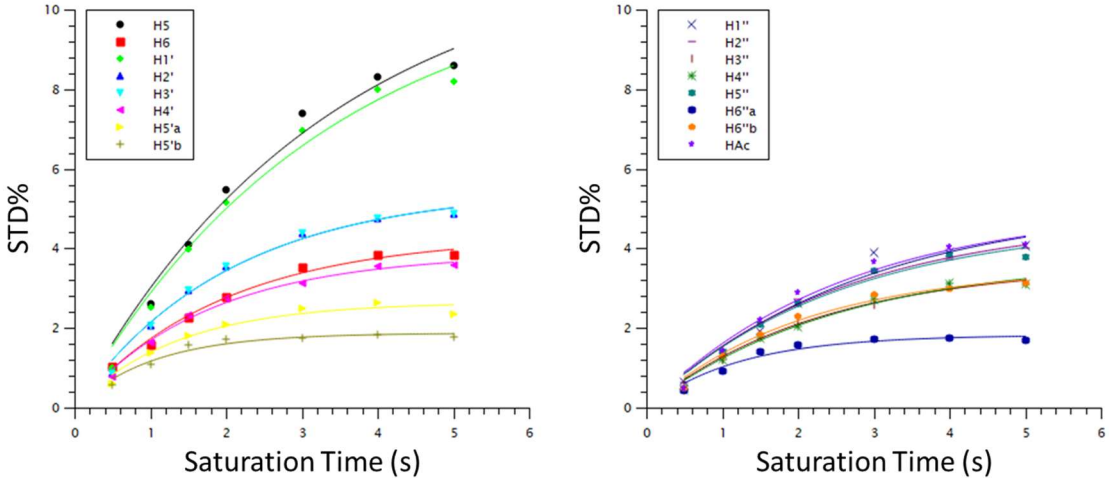



*Figure 6.16 - Experimental STD NMR build-up curves for 1 mM UDP-GlcNAc in the presence of 25 µM SseK1-WT and 25 µM $MgCl_2$. STD NMR experiments were performed in a deuterated water buffer containing 25 mM TRIS-$D_{11}$, 150 mM NaCl, 25 µM $MgCl_2$, pH 7.5. Curves were fit to equation 14 using least squares. For clarity the build-up curves for uridine (left) and Glucose (right) are shown separately.*

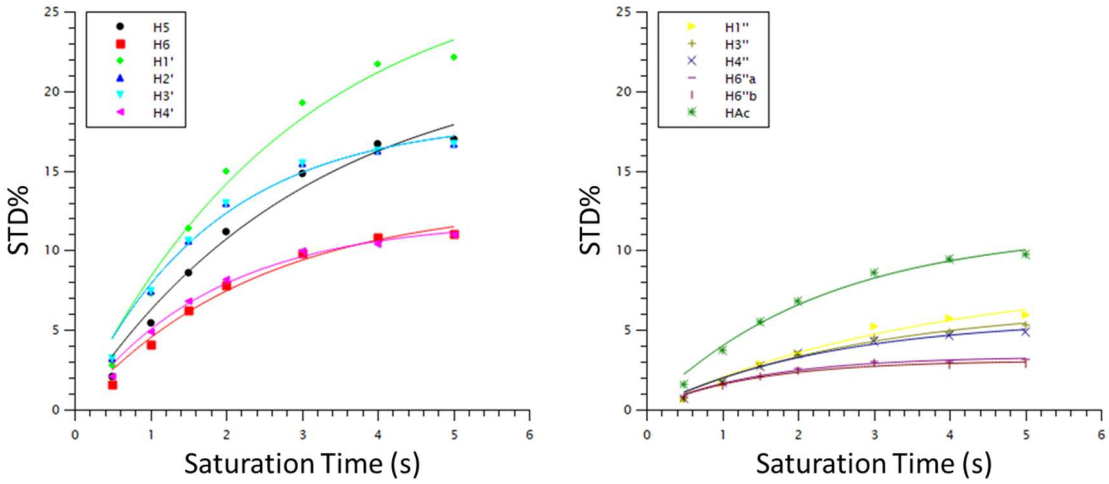



*Figure 6.17 - Experimental STD NMR build-up curves for 1 mM UDP-GalNAc in the presence of 25 µM SseK1-WT and 25 µM $MgCl_2$. STD NMR experiments were performed in a deuterated water buffer containing 25 mM TRIS-$D_{11}$, 150 mM NaCl, 25 µM $MgCl_2$, pH 7.5. Curves were fit to equation 14 using least squares. For clarity the build-up curves for uridine (left) and Glucose (right) are shown separately.*

*Table 6.12 - Calculated parameters from fitting of STD NMR data for binding of UDP to SseK1-WT in the presence of 25 µM $MgCl_2$. $STD_{max}$ and $k_{sat}$ were derived from binding epitope fitting, $STD_0$ was found using equation 14. Normalised intensities (Norm) were calculated from the highest calculated $STD_0$ (bold). STD NMR experiments were performed in a deuterated water buffer containing 25 mM TRIS-$D_{11}$, 150 mM NaCl, 25 µM $MgCl_2$, pH 7.5.*

| Proton | $STD_{max}$ (%) | $k_{sat}$ ($s^{-1}$) | $STD_0$ (% $s^{-1}$) | Norm (%) |
|---|---|---|---|---|
| **H5** | 11.57 | 0.25 | 2.86 | 97 |
| **H6** | 4.95 | 0.43 | 2.15 | 73 |
| **H1'** | 11.49 | 0.26 | **2.94** | 100 |
| **H2'** | 6.11 | 0.41 | 2.51 | 85.5 |
| **H3'** | 6.12 | 0.43 | 2.61 | 89 |
| **H4'** | 4.39 | 0.51 | 2.24 | 76 |
| **H5'** | 2.48 | 0.67 | 1.68 | 57 |

*Table 6.13 - Calculated parameters from fitting of STD NMR data for binding of UDP-Glc to SseK1-WT in the presence of 25 µM $MgCl_2$. $STD_{max}$ and $k_{sat}$ were derived from binding epitope fitting, $STD_0$ was found using equation 14. Normalised intensities (Norm) were calculated from the highest calculated $STD_0$ (bold). STD NMR experiments were performed in a deuterated water buffer containing 25 mM TRIS-$D_{11}$, 150 mM NaCl, 25 µM $MgCl_2$, pH 7.5.*

| Proton | $STD_{max}$ (%) | $k_{sat}$ ($s^{-1}$) | $STD_0$ (% $s^{-1}$) | Norm (%) |
|---|---|---|---|---|
| **H5** | 16.31 | 0.45 | 7.26 | 92 |
| **H6** | 7.73 | 0.71 | 5.51 | 70 |
| **H1'** | 17.21 | 0.46 | **7.86** | 100 |
| **H2'** | 9.10 | 0.76 | 6.95 | 88 |
| **H3'** | 9.10 | 0.76 | 6.95 | 88 |
| **H4'** | 6.75 | 0.83 | 5.57 | 71 |
| **H5'1** | 3.89 | 1.25 | 4.87 | 62 |
| **H5'2** | 3.54 | 1.26 | 4.45 | 57 |
| **H1''** | 7.99 | 0.51 | 4.06 | 52 |
| **H2''** | 6.87 | 0.65 | 4.48 | 57 |
| **H3''** | 4.08 | 1.10 | 4.49 | 57 |
| **H4''** | 5.10 | 0.89 | 4.55 | 58 |
| **H5''** | 7.57 | 0.61 | 4.62 | 59 |
| **H6''1** | 3.73 | 1.27 | 4.74 | 60 |
| **H6''2** | 6.80 | 0.64 | 4.36 | 55 |

*Table 6.14 - Calculated parameters from fitting of STD NMR data for binding of UDP-Gal to SseK1-WT in the presence of 25 µM $MgCl_2$. $STD_{max}$ and $k_{sat}$ were derived from binding epitope fitting, $STD_0$ was found using equation 14. Normalised intensities (Norm) were calculated from the highest calculated $STD_0$ (bold). STD NMR experiments were performed in a deuterated water buffer containing 25 mM TRIS-$D_{11}$, 150 mM NaCl, 25 µM $MgCl_2$, pH 7.5.*

| Proton | $STD_{max}$ (%) | $k_{sat}$ ($s^{-1}$) | $STD_0$ (% $s^{-1}$) | Norm (%) |
|---|---|---|---|---|
| **H5** | 21.97 | 0.31 | 6.79 | 86 |
| **H6** | 10.99 | 0.42 | 4.65 | 59 |
| **H1'** | 24.40 | 0.32 | **7.89** | 100 |
| **H2'** | 13.87 | 0.53 | 7.29 | 92 |
| **H3'** | 13.87 | 0.53 | 7.29 | 92 |
| **H4'** | 9.04 | 0.54 | 4.86 | 62 |
| **H5'1** | 5.32 | 0.61 | 3.24 | 41 |
| **H5'2** | 4.43 | 0.68 | 3.01 | 38 |
| **H1''** | 7.11 | 0.28 | 2.01 | 25 |
| **H2''** | 4.99 | 0.36 | 1.82 | 23 |
| **H3''** | 5.48 | 0.34 | 1.86 | 24 |
| **H4''** | 4.39 | 0.42 | 1.85 | 23.5 |
| **H5''** | 3.53 | 0.50 | 1.76 | 22 |
| **H6''1** | 2.73 | 0.81 | 2.21 | 28 |
| **H6''2** | 2.22 | 0.84 | 1.85 | 23.5 |

*Table 6.15 - Calculated parameters from fitting of STD NMR data for binding of UDP-GlcNAc to SseK1-WT in the presence of 25 µM $MgCl_2$. $STD_{max}$ and $k_{sat}$ were derived from binding epitope fitting, $STD_0$ was found using equation 14. Normalised intensities (Norm) were calculated from the highest calculated $STD_0$ (bold). STD NMR experiments were performed in a deuterated water buffer containing 25 mM TRIS-$D_{11}$, 150 mM NaCl, 25 µM $MgCl_2$, pH 7.5.*

| Proton | $STD_{max}$ (%) | $k_{sat}$ ($s^{-1}$) | $STD_0$ (% $s^{-1}$) | Norm (%) |
|---|---|---|---|---|
| **H5** | 11.62 | 0.30 | **3.47** | 100 |
| **H6** | 4.31 | 0.51 | 2.21 | 64 |
| **H1'** | 11.03 | 0.30 | 3.33 | 96 |
| **H2'** | 5.49 | 0.49 | 2.71 | 78 |
| **H3'** | 5.49 | 0.49 | 2.71 | 78 |
| **H4'** | 3.87 | 0.57 | 2.21 | 64 |
| **H5'1** | 2.65 | 0.74 | 1.97 | 57 |
| **H5'2** | 1.87 | 0.97 | 1.81 | 52 |
| **H1''** | 5.19 | 0.35 | 1.82 | 52 |
| **H2''** | 4.76 | 0.39 | 1.87 | 54 |
| **H3''** | 3.60 | 0.44 | 1.58 | 45 |
| **H4''** | 3.75 | 0.40 | 1.50 | 43 |
| **H5''** | 4.68 | 0.39 | 1.84 | 53 |
| **H6''1** | 1.82 | 0.82 | 1.49 | 43 |
| **H6''2** | 3.52 | 0.48 | 1.70 | 49 |
| **HAc** | 5.05 | 0.38 | 1.93 | 56 |

*Table 6.16 - Calculated parameters from fitting of STD NMR data for binding of UDP-GalNAc to SseK1-WT in the presence of 25 µM $MgCl_2$. $STD_{max}$ and $k_{sat}$ were derived from binding epitope fitting, $STD_0$ was found using equation 14. Normalised intensities (Norm) were calculated from the highest calculated $STD_0$ (bold). STD NMR experiments were performed in a deuterated water buffer containing 25 mM TRIS-$D_{11}$, 150 mM NaCl, 25 µM $MgCl_2$, pH 7.5.*

| Proton | $STD_{max}$ (%) | $k_{sat}$ ($s^{-1}$) | $STD_0$ (% $s^{-1}$) | Norm (%) |
|---|---|---|---|---|
| **H5** | 22.12 | 0.33 | 7.30 | 71 |
| **H6** | 13.07 | 0.42 | 5.50 | 54 |
| **H1'** | 28.08 | 0.35 | 9.84 | 96 |
| **H2'** | 18.28 | 0.56 | **10.25** | 100 |
| **H3'** | 18.28 | 0.56 | **10.25** | 100 |
| **H4'** | 11.91 | 0.55 | 6.50 | 63 |
| **H5'1** | * | * | * | * |
| **H5'2** | * | * | * | * |
| **H1''** | 8.40 | 0.27 | 2.31 | 23 |
| **H2''** | * | * | * | * |
| **H3''** | 6.54 | 0.35 | 2.32 | 23 |
| **H4''** | 5.71 | 0.42 | 2.42 | 24 |
| **H5''** | * | * | * | * |
| **H6''1** | 3.33 | 0.67 | 2.24 | 22 |
| **H6''2** | 3.07 | 0.72 | 2.20 | 22 |
| **HAc** | 11.32 | 0.43 | 4.91 | 48 |

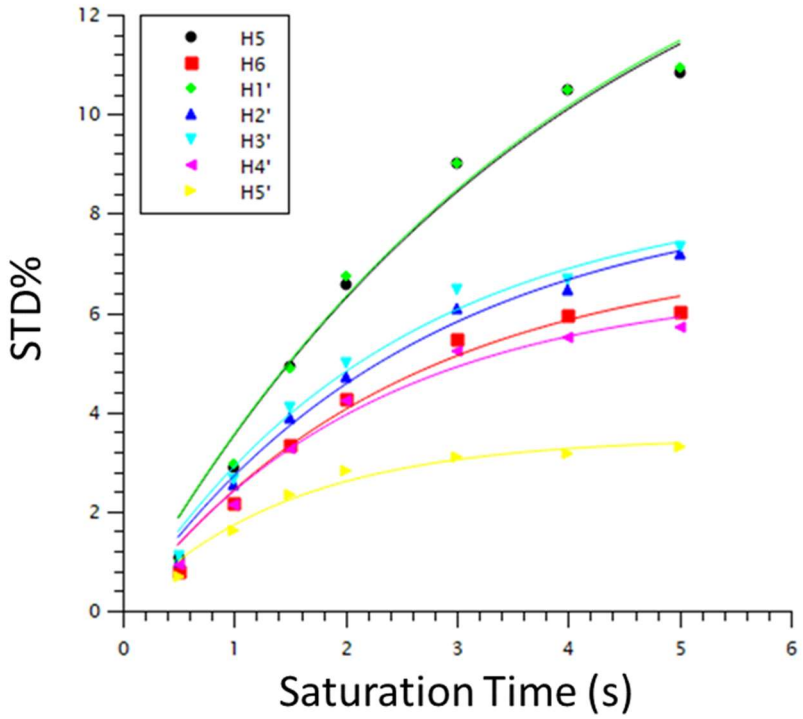



*Figure 6.18 - Experimental STD NMR build-up curves for 1 mM UDP in the presence of 25 µM SseK1-WT and 0 µM $MgCl_2$. STD NMR experiments were performed in a deuterated water buffer containing 25 mM TRIS-$D_{11}$, 150 mM NaCl, pH 7.5. Curves were fit to equation 14 using least squares.*

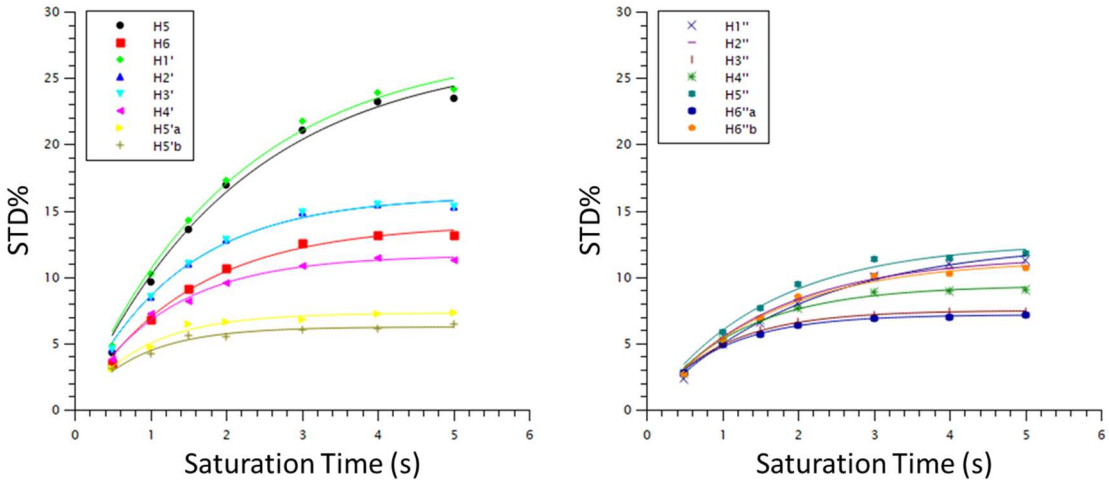



*Figure 6.19 - Experimental STD NMR build-up curves for 1 mM UDP-Glc in the presence of 25 µM SseK1-WT and 0 µM $MgCl_2$. STD NMR experiments were performed in a deuterated water buffer containing 25 mM TRIS-$D_{11}$, 150 mM NaCl, pH 7.5. Curves were fit to equation 14 using least squares. For clarity the build-up curves for uridine (left) and Glucose (right) are shown separately.*

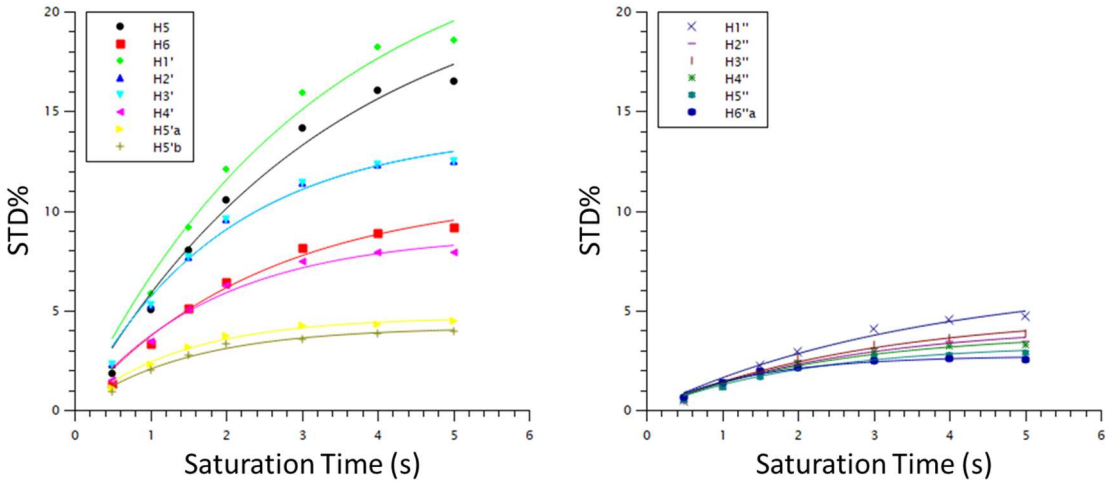



*Figure 6.20 - Experimental STD NMR build-up curves for 1 mM UDP-Gal in the presence of 25 µM SseK1-WT and 0 µM $MgCl_2$. STD NMR experiments were performed in a deuterated water buffer containing 25 mM TRIS-$D_{11}$, 150 mM NaCl, pH 7.5. Curves were fit to equation 14 using least squares. For clarity the build-up curves for uridine (left) and Glucose (right) are shown separately.*

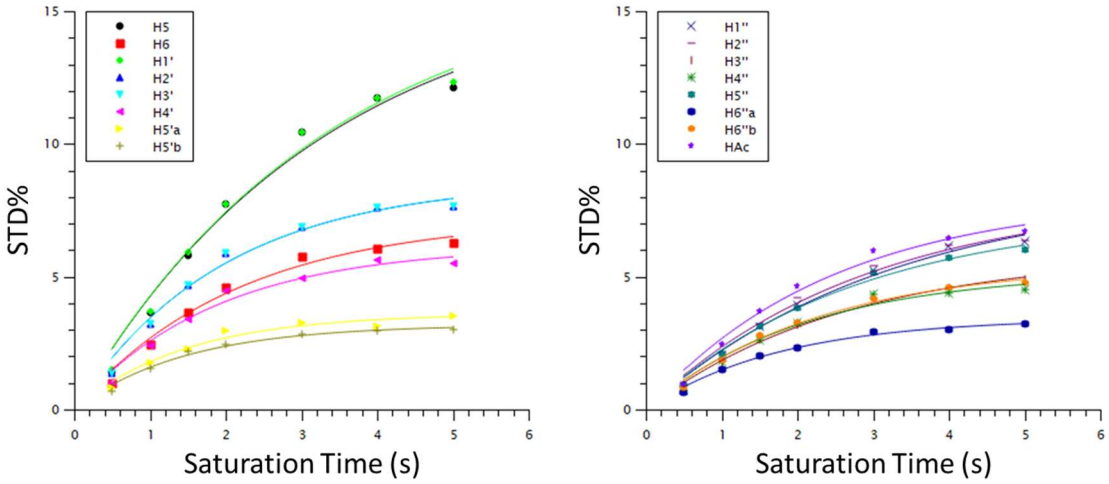



*Figure 6.21 - Experimental STD NMR build-up curves for 1 mM UDP-GlcNAc in the presence of 25 µM SseK1-WT and 0 µM $MgCl_2$. STD NMR experiments were performed in a deuterated water buffer containing 25 mM TRIS-$D_{11}$, 150 mM NaCl, pH 7.5. Curves were fit to equation 14 using least squares. For clarity the build-up curves for uridine (left) and Glucose (right) are shown separately.*

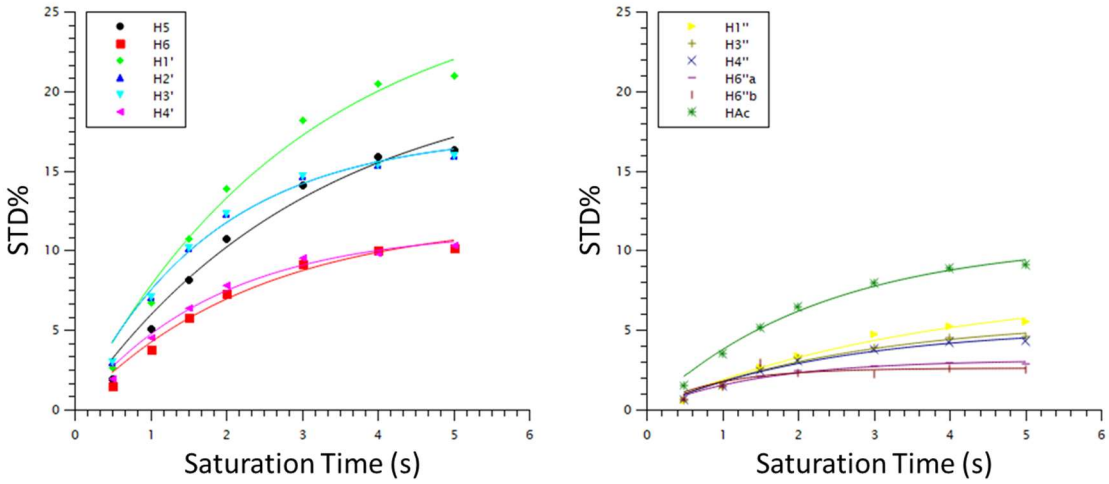



*Figure 6.22 - Experimental STD NMR build-up curves for 1 mM UDP-GalNAc in the presence of 25 µM SseK1-WT and 0 µM $MgCl_2$. STD NMR experiments were performed in a deuterated water buffer containing 25 mM TRIS-$D_{11}$, 150 mM NaCl, pH 7.5. Curves were fit to equation 14 using least squares. For clarity the build-up curves for uridine (left) and Glucose (right) are shown separately.*

*Table 6.17 - Calculated parameters from fitting of STD NMR data for binding of UDP to SseK1-WT in the presence of 0 µM $MgCl_2$. $STD_{max}$ and $k_{sat}$ were derived from binding epitope fitting, $STD_0$ was found using equation 14. Normalised intensities (Norm) were calculated from the highest calculated $STD_0$ (bold). STD NMR experiments were performed in a deuterated water buffer containing 25 mM TRIS-$D_{11}$, 150 mM NaCl, pH 7.5.*

| Proton | $STD_{max}$ (%) | $k_{sat}$ ($s^{-1}$) | $STD_0$ (% $s^{-1}$) | Norm (%) |
|---|---|---|---|---|
| H5 | 16.03 | 0.25 | 3.99 | 99 |
| H6 | 7.29 | 0.41 | 2.96 | 74 |
| H1' | 16.14 | 0.25 | **4.01** | 100 |
| H2' | 8.47 | 0.39 | 3.27 | 82 |
| H3’ | 8.47 | 0.42 | 3.55 | 89 |
| H4’ | 6.59 | 0.46 | 3.00 | 75 |
| H5’ | 3.49 | 0.68 | 2.38 | 59 |

*Table 6.18 - Calculated parameters from fitting of STD NMR data for binding of UDP-Glc to SseK1-WT in the presence of 0 µM $MgCl_2$. $STD_{max}$ and $k_{sat}$ were derived from binding epitope fitting, $STD_0$ was found using equation 14. Normalised intensities (Norm) were calculated from the highest calculated $STD_0$ (bold). STD NMR experiments were performed in a deuterated water buffer containing 25 mM TRIS-$D_{11}$, 150 mM NaCl, pH 7.5.*

| Proton | $STD_{max}$ (%) | $k_{sat}$ (s-1) | $STD_0$ (% $s^{-1}$) | Norm (%) |
|---|---|---|---|---|
| **H5** | 26.98 | 0.47 | 12.60 | 95 |
| **H6** | 13.98 | 0.68 | 9.57 | 72 |
| **H1'** | 27.44 | 0.48 | **13.29** | 100 |
| **H2'** | 16.13 | 0.76 | 12.18 | 92 |
| **H3’** | 16.13 | 0.76 | 12.18 | 92 |
| **H4’** | 11.64 | 0.86 | 10.01 | 75 |
| **H5’1** | 7.28 | 1.15 | 8.40 | 63 |
| **H5’2** | 6.23 | 1.24 | 7.74 | 58 |
| **H1’’** | 12.68 | 0.49 | 6.23 | 47 |
| **H2’’** | 11.57 | 0.63 | 7.34 | 55 |
| **H3’’** | 7.47 | 1.05 | 7.87 | 59 |
| **H4’’** | 9.36 | 0.83 | 7.77 | 58 |
| **H5’’** | 12.59 | 0.64 | 8.06 | 61 |
| **H6’’1** | 7.15 | 1.08 | 7.70 | 58 |
| **H6’’2** | 11.31 | 0.64 | 7.24 | 54 |

*Table 6.19 - Calculated parameters from fitting of STD NMR data for binding of UDP-Gal to SseK1-WT in the presence of 0 µM $MgCl_2$. $STD_{max}$ and $k_{sat}$ were derived from binding epitope fitting, $STD_0$ was found using equation 14. Normalised intensities (Norm) were calculated from the highest calculated $STD_0$ (bold). STD NMR experiments were performed in a deuterated water buffer containing 25 mM TRIS-$D_{11}$, 150 mM NaCl, pH 7.5.*

| Proton | $STD_{max}$ (%) | $k_{sat}$ ($s^{-1}$) | $STD_0$ (% $s^{-1}$) | Norm (%) |
|---|---|---|---|---|
| **H5** | 22.25 | 0.30 | 6.73 | 86 |
| **H6** | 10.91 | 0.41 | 4.49 | 58 |
| **H1'** | 24.54 | 0.32 | **7.78** | 100 |
| **H2'** | 14.03 | 0.52 | 7.25 | 93 |
| **H3'** | 14.03 | 0.52 | 7.25 | 93 |
| **H4'** | 8.83 | 0.55 | 4.83 | 62 |
| **H5'1** | 4.66 | 0.72 | 3.35 | 43 |
| **H5'2** | 4.15 | 0.68 | 2.82 | 36 |
| **H1''** | 6.59 | 0.28 | 1.84 | 24 |
| **H2''** | 4.26 | 0.39 | 1.65 | 21 |
| **H3''** | 4.83 | 0.35 | 1.67 | 21 |
| **H4''** | 3.85 | 0.43 | 1.65 | 21 |
| **H5''** | 3.28 | 0.48 | 1.58 | 20 |
| **H6''1** | 2.72 | 0.73 | 1.97 | 25 |
| **H6''2** | 2.27 | 0.72 | 1.63 | 21 |

*Table 6.20 - Calculated parameters from fitting of STD NMR data for binding of UDP-GlcNAc to SseK1-WT in the presence of 0 µM $MgCl_2$. $STD_{max}$ and $k_{sat}$ were derived from binding epitope fitting, $STD_0$ was found using equation 14. Normalised intensities (Norm) were calculated from the highest calculated $STD_0$ (bold). STD NMR experiments were performed in a deuterated water buffer containing 25 mM TRIS-$D_{11}$, 150 mM NaCl, pH 7.5.*

| Proton | $STD_{max}$ (%) | $k_{sat}$ ($s^{-1}$) | $STD_0$ (% $s^{-1}$) | Norm (%) |
|---|---|---|---|---|
| **H5** | 16.30 | 0.30 | 4.93 | 100 |
| **H6** | 7.23 | 0.46 | 3.35 | 68 |
| **H1'** | 16.57 | 0.30 | **4.94** | 100 |
| **H2'** | 8.61 | 0.51 | 4.40 | 89 |
| **H3'** | 8.61 | 0.51 | 4.40 | 89 |
| **H4'** | 6.16 | 0.54 | 3.33 | 67 |
| **H5'1** | 3.61 | 0.69 | 2.48 | 50 |
| **H5'2** | 3.19 | 0.69 | 2.19 | 44 |
| **H1''** | 8.38 | 0.31 | 2.57 | 52 |
| **H2''** | 8.05 | 0.34 | 2.78 | 56 |
| **H3''** | 5.88 | 0.38 | 2.22 | 45 |
| **H4''** | 5.20 | 0.47 | 2.46 | 50 |
| **H5''** | 7.40 | 0.36 | 2.68 | 54 |
| **H6''1** | 3.44 | 0.57 | 1.95 | 39 |
| **H6''2** | 5.51 | 0.44 | 2.43 | 49 |
| **HAc** | 8.01 | 0.41 | 3.25 | 66 |

*Table 6.21 - Calculated parameters from fitting of STD NMR data for binding of UDP-GalNAc to SseK1-WT in the presence of 0 µM $MgCl_2$. $STD_{max}$ and $k_{sat}$ were derived from binding epitope fitting, $STD_0$ was found using equation 14. Normalised intensities (Norm) were calculated from the highest calculated $STD_0$ (bold). STD NMR experiments were performed in a deuterated water buffer containing 25 mM TRIS-$D_{11}$, 150 mM NaCl, pH 7.5.*

| Proton | $STD_{max}$ (%) | $k_{sat}$ ($s^{-1}$) | $STD_0$ (% $s^{-1}$) | Norm (%) |
|---|---|---|---|---|
| **H5** | 21.28 | 0.33 | 6.92 | 71 |
| **H6** | 12.05 | 0.43 | 5.16 | 53 |
| **H1'** | 26.91 | 0.34 | 9.13 | 94 |
| **H2'** | 17.42 | 0.56 | **9.74** | 100 |
| **H3'** | 17.42 | 0.56 | **9.74** | 100 |
| **H4'** | 11.30 | 0.54 | 6.09 | 63 |
| **H5'1** | * | * | * | * |
| **H5'2** | * | * | * | * |
| **H1''** | 7.66 | 0.28 | 2.12 | 22 |
| **H2''** | * | * | * | * |
| **H3''** | 5.67 | 0.37 | 2.12 | 22 |
| **H4''** | 5.15 | 0.41 | 2.12 | 22 |
| **H5''** | * | * | * | * |
| **H6''1** | 3.11 | 0.67 | 2.10 | 22 |
| **H6''2** | 2.60 | 1.10 | 2.85 | 29 |
| **HAc** | 10.59 | 0.44 | 4.63 | 48 |

### 6.4.3.2 SseK1-S286Y Build-Up Curves and Calculated Parameters

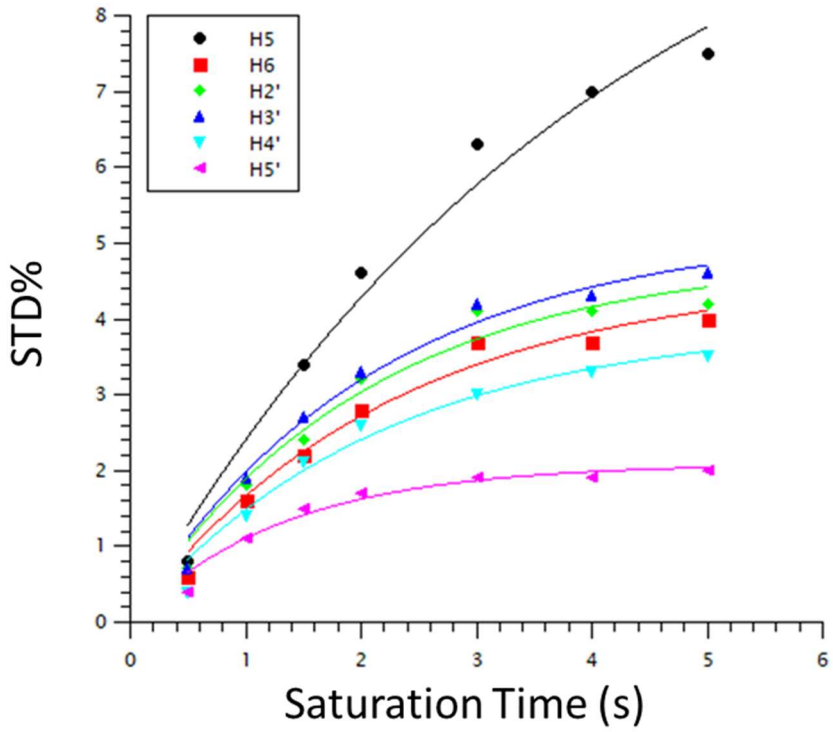



*Figure 6.23 - Experimental STD NMR build-up curves for 1 mM UDP in the presence of 25 µM SseK1-S286Y. STD NMR experiments were performed in a deuterated water buffer containing 25 mM TRIS-$D_{11}$, 150 mM NaCl, 250 µM $MgCl_2$, pH 7.5. Curves were fit to equation 14 using least squares.*

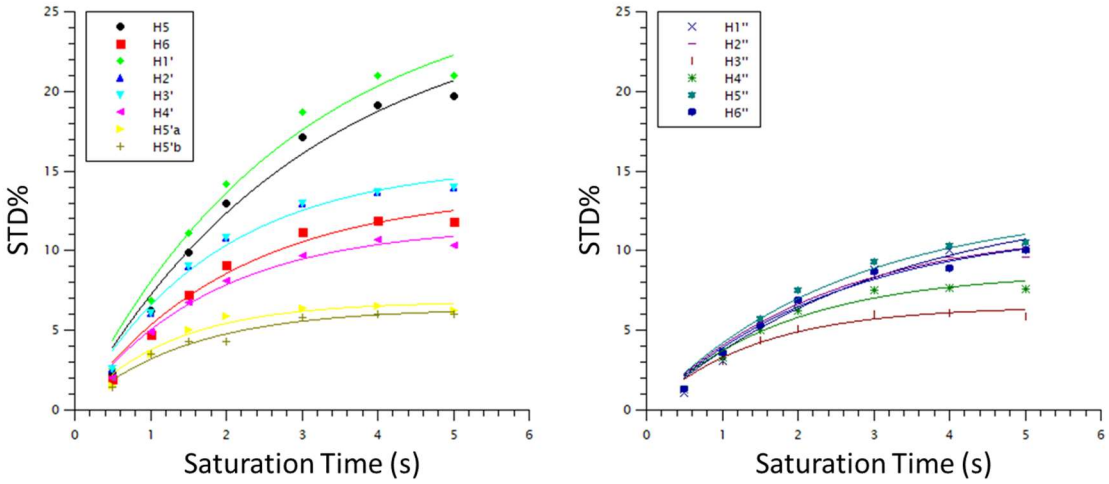



*Figure 6.24 - Experimental STD NMR build-up curves for 1 mM UDP-Glc in the presence of 25 µM SseK1-S286Y. STD NMR experiments were performed in a deuterated water buffer containing 25 mM TRIS-$D_{11}$, 150 mM NaCl, 250 µM $MgCl_2$, pH 7.5. Curves were fit to equation 14 using least squares. For clarity the build-up curves for uridine (left) and Glucose (right) are shown separately.*

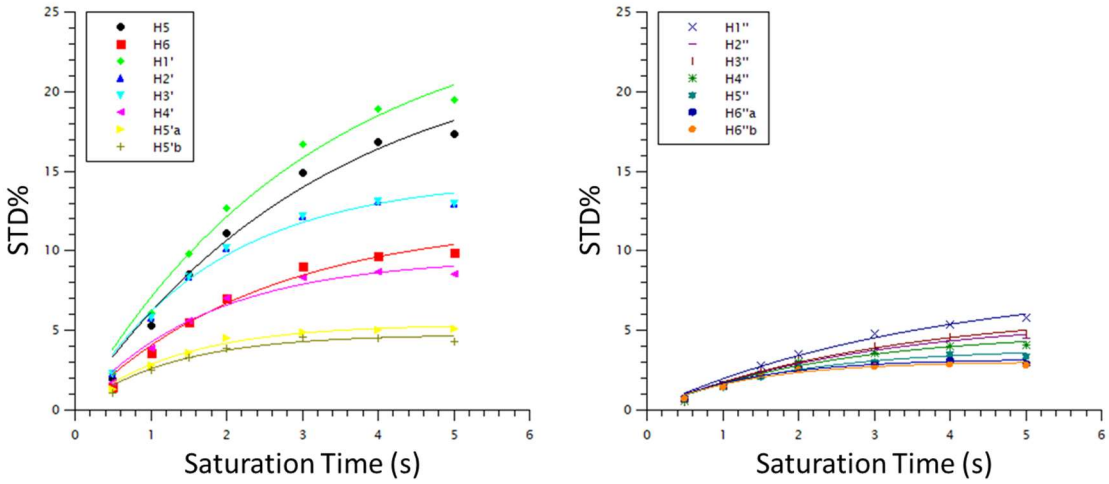



*Figure 6.25 - Experimental STD NMR build-up curves for 1 mM UDP-Gal in the presence of 25 µM SseK1-S286Y. STD NMR experiments were performed in a deuterated water buffer containing 25 mM TRIS-$D_{11}$, 150 mM NaCl, 250 µM $MgCl_2$, pH 7.5. Curves were fit to equation 14 using least squares. For clarity the build-up curves for uridine (left) and Glucose (right) are shown separately.*

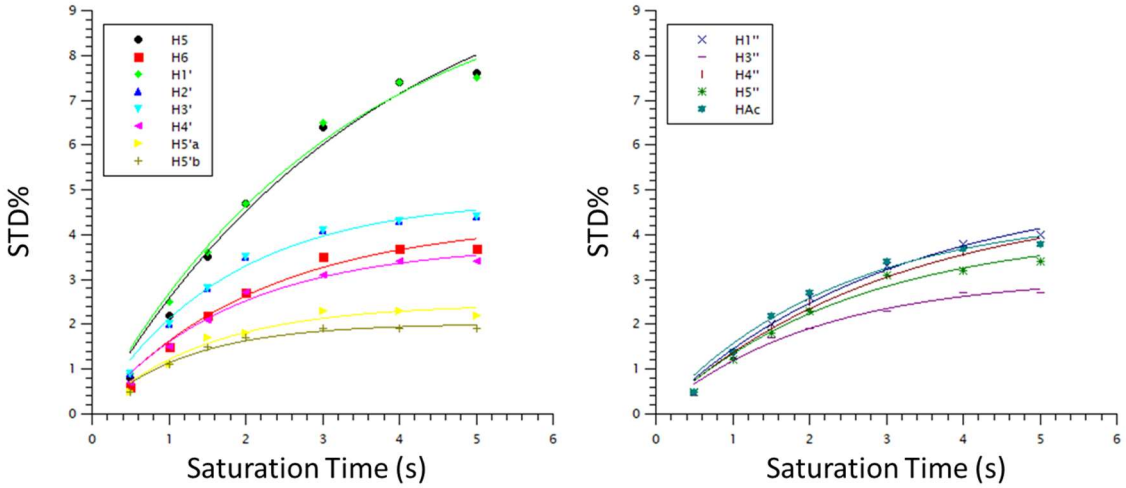



*Figure 6.26 - Experimental STD NMR build-up curves for 1 mM UDP-GlcNAc in the presence of 25 µM SseK1-S286Y. STD NMR experiments were performed in a deuterated water buffer containing 25 mM TRIS-$D_{11}$, 150 mM NaCl, 250 µM $MgCl_2$, pH 7.5. Curves were fit to equation 14 using least squares. For clarity the build-up curves for uridine (left) and Glucose (right) are shown separately.*

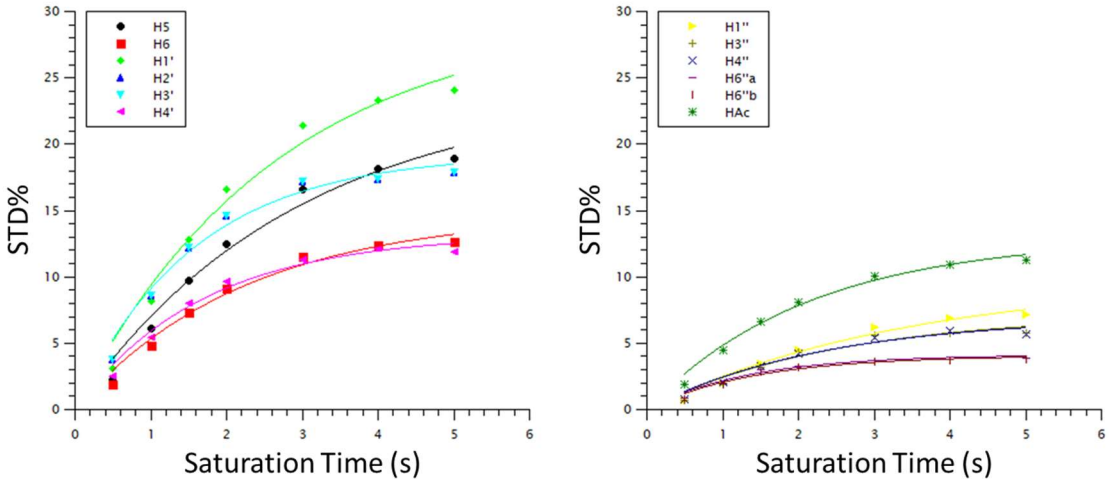



*Figure 6.27 - Experimental STD NMR build-up curves for 1 mM UDP-GalNAc in the presence of 25 µM SseK1-S286Y. STD NMR experiments were performed in a deuterated water buffer containing 25 mM TRIS-$D_{11}$, 150 mM NaCl, 250 µM $MgCl_2$, pH 7.5. Curves were fit to equation 14 using least squares. For clarity the build-up curves for uridine (left) and Glucose (right) are shown separately.*

*Table 6.22 - Calculated parameters from fitting of STD NMR data for binding of UDP to SseK1-S286Y. $STD_{max}$ and $k_{sat}$ were derived from binding epitope fitting, $STD_0$ was found using equation 14. Normalised intensities (Norm) were calculated from the highest calculated $STD_0$ (bold). STD NMR experiments were performed in a deuterated water buffer containing 25 mM TRIS-$D_{11}$, 150 mM NaCl, 250 µM $MgCl_2$, pH 7.5.*

| Proton | $STD_{max}$ (%) | $k_{sat}$ ($s^{-1}$) | $STD_0$ (% $s^{-1}$) | Norm (%) |
|---|---|---|---|---|
| **H5** | 0.24 | 11.24 | **2.69** | 100 |
| **H6** | 0.45 | 4.59 | 2.05 | 76 |
| **H1'** | * | * | * | * |
| **H2'** | 0.50 | 4.81 | 2.38 | 89 |
| **H3'** | 0.48 | 5.17 | 2.48 | 92 |
| **H4'** | 0.47 | 3.94 | 1.84 | 69 |
| **H5'** | 0.75 | 2.08 | 1.55 | 58 |

*Table 6.23 - Calculated parameters from fitting of STD NMR data for binding of UDP-Glc to SseK1-S286Y. $STD_{max}$ and $k_{sat}$ were derived from binding epitope fitting, $STD_0$ was found using equation 14. Normalised intensities (Norm) were calculated from the highest calculated $STD_0$ (bold). STD NMR experiments were performed in a deuterated water buffer containing 25 mM TRIS-$D_{11}$, 150 mM NaCl, 250 µM $MgCl_2$, pH 7.5.*

| Proton | $STD_{max}$ (%) | $k_{sat}$ ($s^{-1}$) | $STD_0$ (% $s^{-1}$) | Norm (%) |
|---|---|---|---|---|
| **H5** | 25.56 | 0.33 | 8.41 | 89 |
| **H6** | 13.67 | 0.49 | 6.68 | 70 |
| **H1'** | 26.80 | 0.35 | **9.49** | 100 |
| **H2'** | 15.47 | 0.55 | 8.49 | 89 |
| **H3'** | 15.47 | 0.55 | 8.49 | 89 |
| **H4'** | 11.55 | 0.57 | 6.53 | 69 |
| **H5'1** | 6.77 | 0.80 | 5.41 | 57 |
| **H5'2** | 6.34 | 0.69 | 4.36 | 46 |
| **H1''** | 13.37 | 0.32 | 4.29 | 45 |
| **H2''** | 11.47 | 0.43 | 4.92 | 52 |
| **H3''** | 6.47 | 0.70 | 4.52 | 48 |
| **H4''** | 8.60 | 0.56 | 4.79 | 51 |
| **H5''** | 12.77 | 0.39 | 5.03 | 53 |
| **H6''1** | 11.60 | 0.41 | 4.70 | 50 |
| **H6''2** | 25.56 | 0.33 | 8.41 | 89 |

*Table 6.24 - Calculated parameters from fitting of STD NMR data for binding of UDP-Gal to SseK1-S286Y. $STD_{max}$ and $k_{sat}$ were derived from binding epitope fitting, $STD_0$ was found using equation 14. Normalised intensities (Norm) were calculated from the highest calculated $STD_0$ (bold). STD NMR experiments were performed in a deuterated water buffer containing 25 mM TRIS-$D_{11}$, 150 mM NaCl, 250 µM $MgCl_2$, pH 7.5.*

| Proton | $STD_{max}$ (%) | $k_{sat}$ ($s^{-1}$) | $STD_0$ (% $s^{-1}$) | Norm (%) |
|---|---|---|---|---|
| **H5** | 23.11 | 0.31 | 7.12 | 87 |
| **H6** | 11.88 | 0.41 | 4.89 | 60 |
| **H1'** | 25.45 | 0.32 | **8.22** | 100 |
| **H2'** | 14.55 | 0.55 | 7.98 | 97 |
| **H3'** | 14.55 | 0.55 | 7.98 | 97 |
| **H4'** | 9.51 | 0.58 | 5.56 | 68 |
| **H5'1** | 5.36 | 0.75 | 4.02 | 49 |
| **H5'2** | 4.72 | 0.78 | 3.67 | 45 |
| **H1''** | 8.05 | 0.27 | 2.20 | 27 |
| **H2''** | 5.70 | 0.35 | 2.00 | 24 |
| **H3''** | 6.22 | 0.32 | 2.01 | 24 |
| **H4''** | 4.88 | 0.41 | 2.02 | 25 |
| **H5''** | 3.83 | 0.53 | 2.03 | 25 |
| **H6''1** | 3.18 | 0.73 | 2.34 | 28 |
| **H6''2** | 3.02 | 0.74 | 2.22 | 27 |

*Table 6.25 - Calculated parameters from fitting of STD NMR data for binding of UDP-GlcNAc to SseK1-S286Y. $STD_{max}$ and $k_{sat}$ were derived from binding epitope fitting, $STD_0$ was found using equation 14. Normalised intensities (Norm) were calculated from the highest calculated $STD_0$ (bold). STD NMR experiments were performed in a deuterated water buffer containing 25 mM TRIS-$D_{11}$, 150 mM NaCl, 250 µM $MgCl_2$, pH 7.5.*

| Proton | $STD_{max}$ (%) | $k_{sat}$ ($s^{-1}$) | $STD_0$ (% $s^{-1}$) | Norm (%) |
|---|---|---|---|---|
| **H5** | 10.80 | 0.27 | 2.92 | 94 |
| **H6** | 4.31 | 0.47 | 2.02 | 65 |
| **H1'** | 10.05 | 0.31 | **3.10** | 100 |
| **0H2'** | 4.80 | 0.58 | 2.77 | 89 |
| **H3'** | 4.80 | 0.58 | 2.77 | 89 |
| **H4'** | 3.77 | 0.55 | 2.07 | 67 |
| **H5'1** | 2.44 | 0.68 | 1.65 | 53 |
| **H5'2** | 2.01 | 0.82 | 1.65 | 53 |
| **H1''** | 5.11 | 0.33 | 1.68 | 54 |
| **H2''** | * | * | * | * |
| **H3''** | 3.04 | 0.49 | 1.48 | 48 |
| **H4''** | 4.84 | 0.33 | 1.59 | 51 |
| **H5''** | 4.09 | 0.39 | 1.61 | 52 |
| **H6''1** | * | * | * | * |
| **H6''2** | * | * | * | * |
| **HAc** | 4.49 | 0.42 | 1.91 | 61 |

*Table 6.26 - Calculated parameters from fitting of STD NMR data for binding of UDP-GalNAc to SseK1-S286Y. STDmax and ksat were derived from binding epitope fitting, $STD_0$ was found using equation 14. Normalised intensities (Norm) were calculated from the highest calculated $STD_0$ (bold). STD NMR experiments were performed in a deuterated water buffer containing 25 mM TRIS-D11, 150 mM NaCl, 250 µM MgCl2, pH 7.5.*

| Proton | $STD_{max}$ (%) | $k_{sat}$ ($s^{-1}$) | $STD_0$ (% $s^{-1}$) | Norm (%) |
|---|---|---|---|---|
| **H5** | 24.03 | 0.34 | 8.24 | 67 |
| **H6** | 14.78 | 0.44 | 6.54 | 53 |
| **H1'** | 29.67 | 0.38 | 11.15 | 91 |
| **H2'** | 19.25 | 0.64 | **12.24** | 100 |
| **H3'** | 19.25 | 0.64 | **12.24** | 100 |
| **H4'** | 13.17 | 0.59 | 7.76 | 63 |
| **H5'1** | * | * | * | * |
| **H5'2** | * | * | * | * |
| **H1''** | 9.85 | 0.29 | 2.82 | 23 |
| **H2''** | * | * | * | * |
| **H3''** | 7.18 | 0.40 | 2.88 | 23.5 |
| **H4''** | 6.94 | 0.43 | 2.97 | 24 |
| **H5''** | * | * | * | * |
| **H6''1** | 4.09 | 0.74 | 3.02 | 25 |
| **H6''2** | 4.02 | 0.69 | 2.79 | 23 |
| **HAc** | 12.90 | 0.46 | 5.98 | 49 |

### 6.4.3.3 SseK2 Build-Up Curves and Calculated Parameters

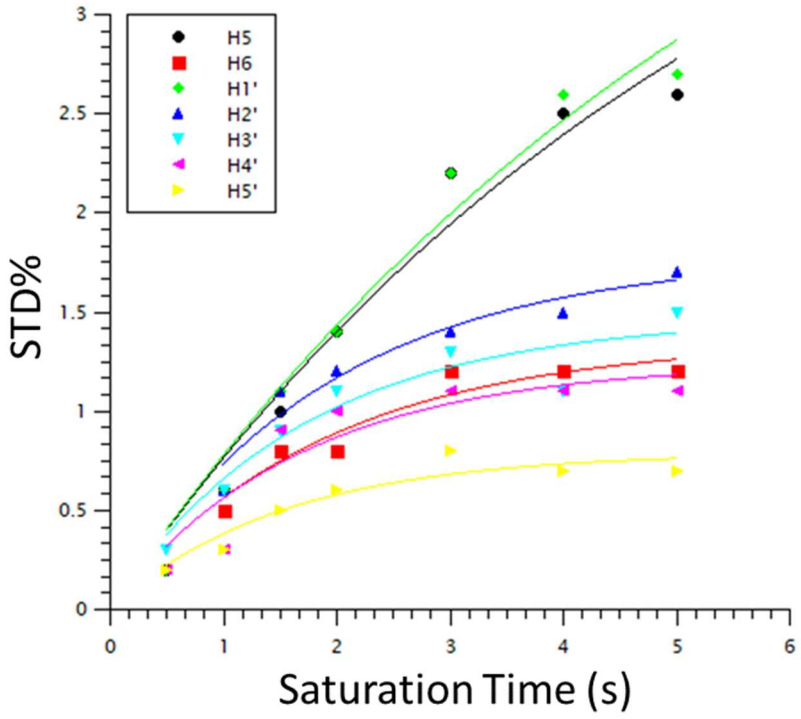



*Figure 6.28 - Experimental STD NMR build-up curves for 1 mM UDP in the presence of 25 µM SseK2-WT. STD NMR experiments were performed in a deuterated water buffer containing 25 mM TRIS-$D_{11}$, 150 mM NaCl, 25 µM $MgCl_2$, pH 7.5. Curves were fit to equation 14 using least squares.*

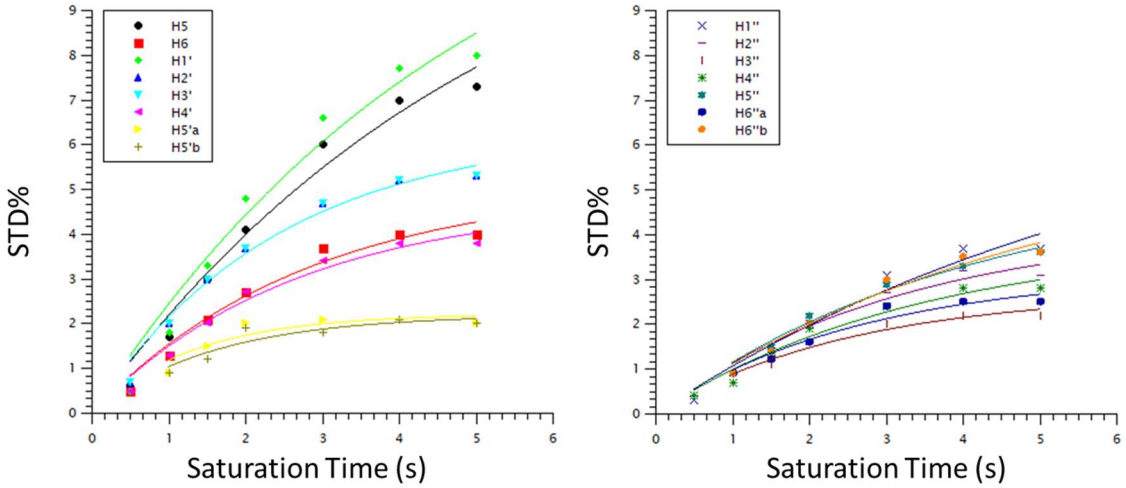



*Figure 6.29 - Experimental STD NMR build-up curves for 1 mM UDP-Glc in the presence of 25 µM SseK2-WT. STD NMR experiments were performed in a deuterated water buffer containing 25 mM TRIS-D11, 150 mM NaCl, 250 µM $MgCl_2$, pH 7.5. Curves were fit to equation 14 using least squares. For clarity the build-up curves for uridine (left) and Glucose (right) are shown separately.*

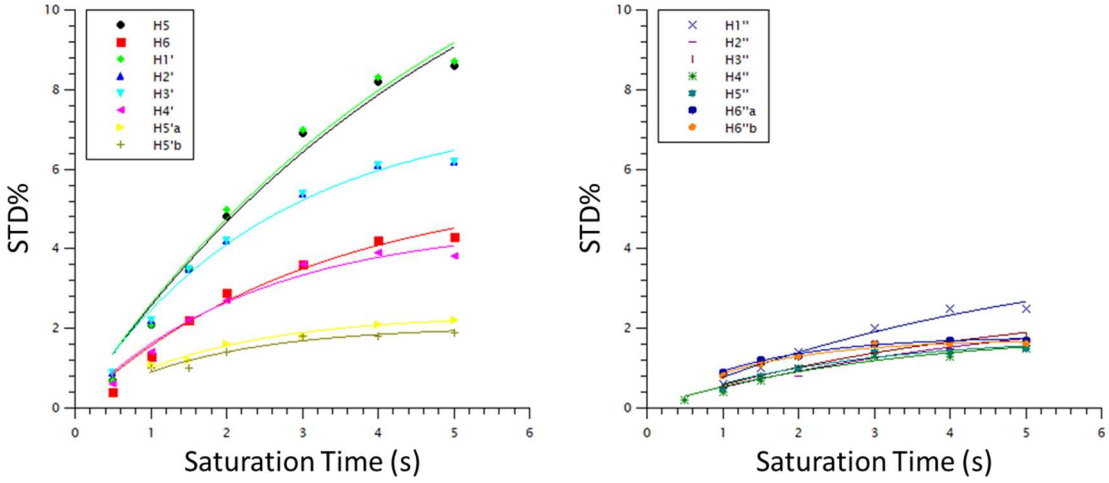



*Figure 6.30 - Experimental STD NMR build-up curves for 1 mM UDP-Gal in the presence of 25 µM SseK2-WT. STD NMR experiments were performed in a deuterated water buffer containing 25 mM TRIS-$D_{11}$, 150 mM NaCl, 250 µM $MgCl_2$, pH 7.5. Curves were fit to equation 14 using least squares. For clarity the build-up curves for uridine (left) and Glucose (right) are shown separately.*

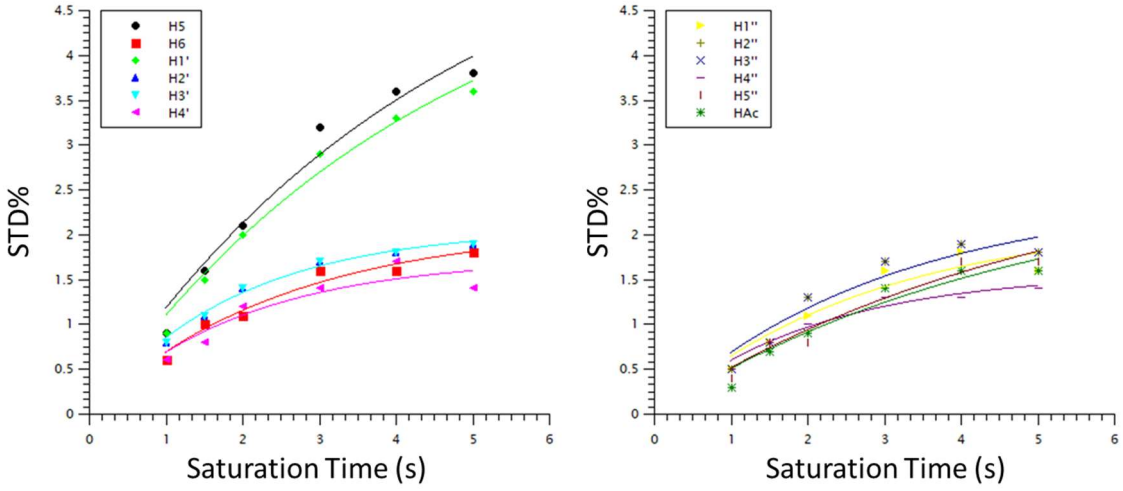



*Figure 6.31 - Experimental STD NMR build-up curves for 1 mM UDP-GlcNAc in the presence of 25 µM SseK2-WT. STD NMR experiments were performed in a deuterated water buffer containing 25 mM TRIS-D11, 150 mM NaCl, 250 µM $MgCl_2$, pH 7.5. Curves were fit to equation 14 using least squares. For clarity the build-up curves for uridine (left) and Glucose (right) are shown separately.*

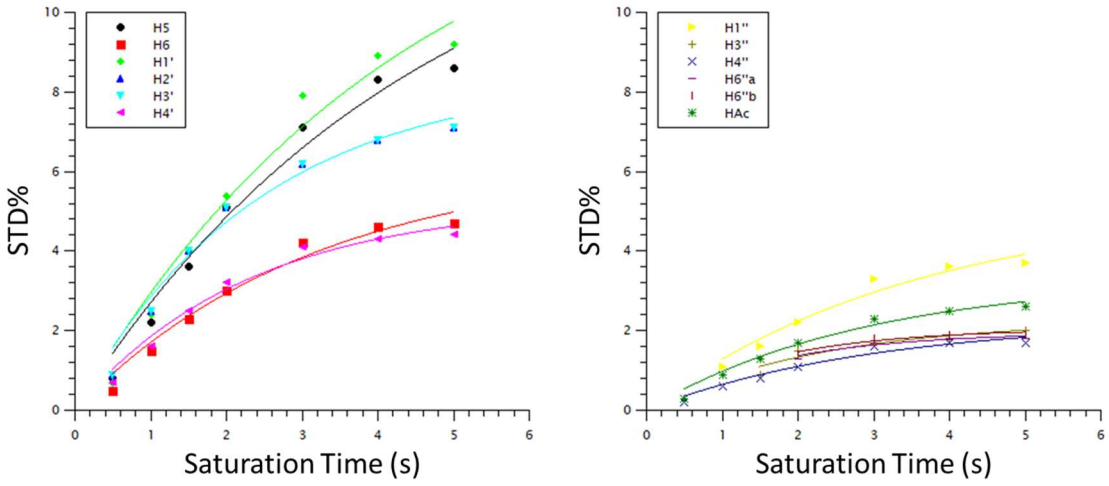



*Figure 6.32 - Experimental STD NMR build-up curves for 1 mM UDP-GalNAc in the presence of 25 µM SseK2-WT. STD NMR experiments were performed in a deuterated water buffer containing 25 mM TRIS-$D_{11}$, 150 mM NaCl, 250 µM $MgCl_2$, pH 7.5. Curves were fit to equation 14 using least squares. For clarity the build-up curves for uridine (left) and Glucose (right) are shown separately.*

*Table 6.27 - Calculated parameters from fitting of STD NMR data for binding of UDP to SseK2. $STD_{max}$ and $k_{sat}$ were derived from binding epitope fitting, $STD_0$ was found using equation 14. Normalised intensities (Norm) were calculated from the highest calculated $STD_0$ (bold). STD NMR experiments were performed in a deuterated water buffer containing 25 mM TRIS-$D_{11}$, 150 mM NaCl, 250 µM $MgCl_2$, pH 7.5.*

| Proton | $STD_{max}$ (%) | $k_{sat}$ ($s^{-1}$) | $STD_0$ (% $s^{-1}$) | Norm (%) |
|---|---|---|---|---|
| **H5** | 4.79 | 0.17 | 0.83 | 88 |
| **H6** | 1.35 | 0.53 | 0.72 | 77 |
| **H1'** | 5.16 | 0.16 | 0.84 | 89 |
| **H2'** | 1.78 | 0.53 | **0.94** | 100 |
| **H3'** | 1.47 | 0.59 | 0.87 | 92 |
| **H4'** | 1.24 | 0.60 | 0.74 | 78 |
| **H5'** | 0.79 | 0.66 | 0.52 | 55 |

*Table 6.28 - Calculated parameters from fitting of STD NMR data for binding of UDP-Glc to SseK2. $STD_{max}$ and $k_{sat}$ were derived from binding epitope fitting, $STD_0$ was found using equation 14. Normalised intensities (Norm) were calculated from the highest calculated $STD_0$ (bold). STD NMR experiments were performed in a deuterated water buffer containing 25 mM TRIS-$D_{11}$, 150 mM NaCl, 250 µM $MgCl_2$, pH 7.5.*

| Proton | $STD_{max}$ (%) | $k_{sat}$ ($s^{-1}$) | $STD_0$ (% $s^{-1}$) | Norm (%) |
|---|---|---|---|---|
| **H5** | 12.55 | 0.19 | 2.40 | 89 |
| **H6** | 5.13 | 0.36 | 1.82 | 68 |
| **H1'** | 13.41 | 0.20 | **2.69** | 100 |
| **H2'** | 6.32 | 0.41 | 2.62 | 97 |
| **H3'** | 6.32 | 0.41 | 2.62 | 97 |
| **H4'** | 4.74 | 0.38 | 1.79 | 67 |
| **H5'1** | 2.21 | 0.76 | 1.69 | 63 |
| **H5'2** | 2.19 | 0.64 | 1.39 | 52 |
| **H1''** | 7.59 | 0.15 | 1.14 | 42 |
| **H2''** | 4.21 | 0.31 | 1.31 | 49 |
| **H3''** | 2.71 | 0.39 | 1.06 | 39 |
| **H4''** | 3.90 | 0.29 | 1.13 | 42 |
| **H5''** | 5.19 | 0.25 | 1.30 | 48 |
| **H6''1** | 3.16 | 0.37 | 1.16 | 43 |
| **H6''2** | 5.93 | 0.21 | 1.22 | 45 |

*Table 6.29 - Calculated parameters from fitting of STD NMR data for binding of UDP-Gal to SseK2. $STD_{max}$ and $k_{sat}$ were derived from binding epitope fitting, $STD_0$ was found using equation 14. Normalised intensities (Norm) were calculated from the highest calculated $STD_0$ (bold). STD NMR experiments were performed in a deuterated water buffer containing 25 mM TRIS-$D_{11}$, 150 mM NaCl, 250 µM $MgCl_2$, pH 7.5.*

| Proton | $STD_{max}$ (%) | $k_{sat}$ ($s^{-1}$) | $STD_0$ (% $s^{-1}$) | Norm (%) |
|---|---|---|---|---|
| **H5** | 14.85 | 0.19 | 2.79 | 94 |
| **H6** | 5.64 | 0.32 | 1.81 | 61 |
| **H1'** | 14.83 | 0.19 | 2.85 | 96 |
| **H2'** | 7.50 | 0.39 | **2.96** | 100 |
| **H3'** | 7.50 | 0.39 | **2.96** | 100 |
| **H4'** | 4.61 | 0.42 | 1.96 | 66 |
| **H5'1** | 2.33 | 0.55 | 1.28 | 43 |
| **H5'2** | 2.04 | 0.57 | 1.17 | 40 |
| **H1''** | 4.25 | 0.20 | 0.84 | 28 |
| **H2''** | 2.69 | 0.21 | 0.56 | 19 |
| **H3''** | 2.78 | 0.23 | 0.63 | 21 |
| **H4''** | 1.90 | 0.32 | 0.61 | 21 |
| **H5''** | 1.77 | 0.41 | 0.73 | 25 |
| **H6''1** | 1.78 | 0.71 | 1.27 | 43 |
| **H6''2** | 1.71 | 0.70 | 1.19 | 40 |

*Table 6.30 - Calculated parameters from fitting of STD NMR data for binding of UDP-GlcNAc to SseK1-S286Y. $STD_{max}$ and $k_{sat}$ were derived from binding epitope fitting, $STD_0$ was found using equation 14. Normalised intensities (Norm) were calculated from the highest calculated $STD_0$ (bold). STD NMR experiments were performed in a deuterated water buffer containing 25 mM TRIS-$D_{11}$, 150 mM NaCl, 250 µM $MgCl_2$, pH 7.5.*

| Proton | $STD_{max}$ (%) | $k_{sat}$ ($s^{-1}$) | $STD_0$ (% $s^{-1}$) | Norm (%) |
|---|---|---|---|---|
| **H5** | 5.97 | 0.22 | **1.32** | 100 |
| **H6** | 2.09 | 0.40 | 0.84 | 64 |
| **H1'** | 5.54 | 0.22 | 1.23 | 93 |
| **H2'** | 2.07 | 0.53 | 1.09 | 83 |
| **H3'** | 2.07 | 0.53 | 1.09 | 83 |
| **H4'** | 1.74 | 0.50 | 0.87 | 66 |
| **H5'1** | * | * | * | * |
| **H5'2** | 2.16 | 0.35 | 0.77 | 58 |
| **H1''** | 2.43 | 0.33 | 0.81 | 61 |
| **H2''** | * | * | * | * |
| **H3''** | 1.57 | 0.48 | 0.75 | 57 |
| **H4''** | 2.93 | 0.19 | 0.57 | 43 |
| **H5''** | * | * | * | * |
| **H6''1** | * | * | * | * |
| **H6''2** | 2.64 | 0.21 | 0.56 | 42 |
| **HAc** | 5.97 | 0.22 | 1.32 | 100 |

*Table 6.31 - Calculated parameters from fitting of STD NMR data for binding of UDP-GalNAc to SseK2. STDmax and ksat were derived from binding epitope fitting, $STD_0$ was found using equation 14. Normalised intensities (Norm) were calculated from the highest calculated $STD_0$ (bold). STD NMR experiments were performed in a deuterated water buffer containing 25 mM TRIS-D11, 150 mM NaCl, 250 µM MgCl2, pH 7.5.*

| Proton | $STD_{max}$ (%) | $k_{sat}$ ($s^{-1}$) | $STD_0$ (% $s^{-1}$) | Norm (%) |
|---|---|---|---|---|
| **H5** | 13.61 | 0.22 | 2.99 | 86 |
| **H6** | 6.27 | 0.31 | 1.97 | 57 |
| **H1'** | 14.32 | 0.23 | 3.28 | 95 |
| **H2'** | * | * | * | * |
| **H3'** | 8.41 | 0.41 | **3.46** | 100 |
| **H4'** | 8.41 | 0.41 | **3.46** | 100 |
| **H5'1** | * | * | * | * |
| **H5'2** | * | * | * | * |
| **H1''** | 5.19 | 0.44 | 2.27 | 66 |
| **H2''** | 5.17 | 0.28 | 1.46 | 42 |
| **H3''** | 2.24 | 0.44 | 1.00 | 29 |
| **H4''** | 2.25 | 0.33 | 0.74 | 21 |
| **H5''** | * | * | * | * |
| **H6''1** | 1.95 | 0.60 | 1.16 | 34 |
| **H6''2** | 2.04 | 0.63 | 1.28 | 37 |
| **HAc** | 3.31 | 0.34 | 1.14 | 33 |

# Bibliography


1. Vraki A. Biological Functions of Glycans. In Essentials of glycobiology, 4th edition, 2022.

2. Gabius HJ. Cell surface glycans: the why and how of their functionality as biochemical signals in lectin-mediated information transfer. Crit Rev Immunol. 2006;26(1):43-79.

3. Vijayan M, Chandra N. Lectins. Current Opinion in Structural Biology. 1999;9(6):707-14.

4. Ip WK, Takahashi K, Ezekowitz RA, Stuart LM. Mannose-binding lectin and innate immunity. Immunol Rev. 2009 230(1):9-21.

5. Bektas M, Rubenstein DS. The role of intracellular protein O-glycosylation in cell adhesion and disease. 25. 2011;4.

6. Kamemura K, Hart GW. Dynamic interplay between O-glycosylation and O-phosphorylation of nucleocytoplasmic proteins: a new paradigm for metabolic control of signal transduction and transcription. Prog Nucleic Acid Res Mol Biol. 2003;73:107-36.

7. Jozwiak P, Forma E, Brys M, Krzeslak A. O-GlcNAcylation and Metabolic Reprograming in Cancer. Front Endocrinol (Lausanne). 2014;5:145.

8. Hennet T. Diseases of glycosylation beyond classical congenital disorders of glycosylation. Biochim Biophys Acta. 2012;1820(9):1306-17.

9. Zhou X, Motta F, Selmi C, Ridgway WM, Gershwin ME, Zhang W. Antibody glycosylation in autoimmune diseases. Autoimmun Rev. 2021;20(5):102804.

10. Ernst B, Magnani JL. From carbohydrate leads to glycomimetic drugs. Nat Rev Drug Discov. 2009;8(8):661-77.

11. Lebrilla CB, Liu J, Widmalm G, Prestegard JH. Oligosaccharides and Polysaccharides. In Essentials of Glycobiology 4th edition.

12. Seeberger PH. Monosaccharide Diversity. In Essetials of glycobiology, 4th edition, 2022.

13. Linhardt RJ, Bazin HG. Physical properties of carbohydrates. In Glycoscience: Chemistry and Chemical Biology I–III, 1st edition, 2001.

14. Takahashi K, Ono S. Calorimetric studies on the mutarotation of D-galactose and D-mannose. J Biochem. 1972;73(4):763-70.

15. Eichler J. Protein glycosylation. Curr Biol. 2019;29(7):R229-R31.

16. Moremen KW, Tiemeyer M, Nairn AV. Vertebrate protein glycosylation: diversity, synthesis and function. Nat Rev Mol Cell Biol. 2012;13(7):448-62.

17. Rini JM, Moremen KW, Davis BG, Esko JD. Glycosyltransferases and Glycan-Processing Enzymes. In Essentials of Glycobiology, 4th edition, 2022.

18. Van den Steen P, Rudd PM, Dwek RA, Opdenakker G. Concepts and principles of O-linked glycosylation. Crit Rev Biochem Mol Biol. 1998;33(3):151-208.

19. Mellquist JL, Kasturi L, Spitalnik SL, Shakin-Eshleman SH. The amino acid following an asn-X-Ser/Thr sequon is an important determinant of N-linked core glycosylation efficiency. Biochemistry. 1998;37(19):6833-7.

20. Ardevol A, Iglesias-Fernandez J, Rojas-Cervellera V, Rovira C. The reaction mechanism of retaining glycosyltransferases. Biochem Soc Trans. 2016;44(1):51-60.

21. Malik V, Black GW. Advances in Protein Chemistry and Structural Biology. Volume 87, Chapter 4 - Structural, Functional, and Mutagenesis Studies of UDP-Glycosyltransferases, 2012.

22. Mestrom L, Przypis M, Kowalczykiewicz D, Pollender A, Kumpf A, Marsden SR, et al. Leloir Glycosyltransferases in Applied Biocatalysis: A Multidisciplinary Approach. Int J Mol Sci. 2019;20(21).

23. Lairson LL, Henrissat B, Davies GJ, Withers SG. Glycosyltransferases Structures, Functions,and Mechanisms. Annu Rev Biochem. 2008;77;77:521-55.

24. Breton C, Snajdrova L, Jeanneau C, Koca J, Imberty A. Structures and mechanisms of glycosyltransferases. Glycobiology. 2006;16(2):29R-37R.

25. Liang DM, Liu JH, Wu H, Wang BB, Zhu HJ, Qiao JJ. Glycosyltransferases: mechanisms and applications in natural product development. Chem Soc Rev. 2015;44(22):8350-74.

26. Pan X, Luo J, Li S. Bacteria-Catalyzed Arginine Glycosylation in Pathogens and Host. Front Cell Infect Microbiol. 2020;10:185.

27. Li S, Zhang L, Yao Q, Li L, Dong N, Rong J, et al. Pathogen blocks host death receptor signalling by arginine GlcNAcylation of death domains. Nature. 2013;501(7466):242-6.

28. Pearson JS, Giogha C, Ong SY, Kennedy CL, Kelly M, Robinson KS, et al. A type III effector antagonizes death receptor signalling during bacterial gut infection. Nature. 2013;501(7466):247-51.

29. El Qaidi S, Chen K, Halim A, Siukstaite L, Rueter C, Hurtado-Guerrero R, et al. NleB/SseK effectors from Citrobacter rodentium, Escherichia coli, and Salmonella enterica display distinct differences in host substrate specificity. J Biol Chem. 2017;292(27):11423-30.

30. Lassak J, Keilhauer EC, Furst M, Wuichet K, Godeke J, Starosta AL, et al. Arginine-rhamnosylation as new strategy to activate translation elongation factor P. Nat Chem Biol. 2015;11(4):266-70.

31. Wagner VE, Iglewski BH. P. aeruginosa Biofilms in CF Infection. Clin Rev Allergy Immunol. 2008;35(3):124-34.

32. Qadri F, Svennerholm AM, Faruque AS, Sack RB. Enterotoxigenic Escherichia coli in developing countries: epidemiology, microbiology, clinical features, treatment, and prevention. Clin Microbiol Rev. 2005;18(3):465-83.

33. Xue J, Hu S, Huang Y, Zhang Q, Yi X, Pan X, et al. Arg-GlcNAcylation on TRADD by NleB and SseK1 Is Crucial for Bacterial Pathogenesis. Front Cell Dev Biol. 2020;8:641.

34. Gunster RA, Matthews SA, Holden DW, Thurston TLM. SseK1 and SseK3 Type III Secretion System Effectors Inhibit NF-kappaB Signaling and Necroptotic Cell Death in Salmonella-Infected Macrophages. Infect Immun. 2017;85(3).

35. Ding J, Pan X, Du L, Yao Q, Xue J, Yao H, et al. Structural and Functional Insights into Host Death Domains Inactivation by the Bacterial Arginine GlcNAcyltransferase Effector. Mol Cell. 2019;74(5):922-35 e6.

36. Park JB, Kim YH, Yoo Y, Kim J, Jun SH, Cho JW, et al. Structural basis for arginine glycosylation of host substrates by bacterial effector proteins. Nat Commun. 2018;9(1):4283.

37. Esposito D, Gunster RA, Martino L, El Omari K, Wagner A, Thurston TLM, et al. Structural basis for the glycosyltransferase activity of the Salmonella effector SseK3. J Biol Chem. 2018;293(14):5064-78.

38. Araujo-Garrido JL, Bernal-Bayard J, Ramos-Morales F. Type III Secretion Effectors with Arginine N-Glycosyltransferase Activity. Microorganisms. 2020;8(3).

39. Diepold A, Armitage JP. Type III secretion systems: the bacterial flagellum and the injectisome. Philos Trans R Soc Lond B Biol Sci. 2015;370(1679).

40. Gaytan MO, Martinez-Santos VI, Soto E, Gonzalez-Pedrajo B. Type Three Secretion System in Attaching and Effacing Pathogens. Front Cell Infect Microbiol. 2016;6:129.

41. Newson JPM, Scott NE, Yeuk Wah Chung I, Wong Fok Lung T, Giogha C, Gan J, et al. Salmonella Effectors SseK1 and SseK3 Target Death Domain Proteins in the TNF and TRAIL Signaling Pathways. Mol Cell Proteomics. 2019;18(6):1138-56.

42. Xu C, Liu X, Zha H, Fan S, Zhang D, Li S, et al. A pathogen-derived effector modulates host glucose metabolism by arginine GlcNAcylation of HIF-1alpha protein. PLoS Pathog. 2018;14(8):e1007259.

43. El Qaidi S, Scott NE, Hays MP, Geisbrecht BV, Watkins S, Hardwidge PR. An intra-bacterial activity for a T3SS effector. Sci Rep. 2020;10(1):1073.

44. Garcia-Garcia A, Hicks T, El Qaidi S, Zhu C, Hardwidge PR, Angulo J, et al. NleB/SseK-catalyzed arginine-glycosylation and enteropathogen virulence are finely tuned by a single variable position contiguous to the catalytic machinery. Chem Sci. 2021;12(36):12181-91.

45. Keeler J. Understanding NMR Spectroscopy, 2nd edition, 2010.

46. Hore PJ. Nuclear Magnetic Resonance, 2nd edition, 2015.

47. Claridge T. High-Resolution NMR Techniques in Organic Chemistry, 1st edition, 2016.

48. Atkins P, Paula J, Keeler J. Atkins Physical Chemistry, 11th edition, 2018.

49. Karunakaran C, Santharaman P, Balamurugan M. Spin Resonance Spectroscopy Chapter Two - 1H and 13C Nuclear Magnetic Resonance Spectroscopy, 1st edition, 2018.

50. Mayer M, Meyer B. Characterization of Ligand Binding by Saturation Transfer Difference NMR Spectroscopy. Angew Chem Int Ed Engl. 1999;38(12):1784-8.

51. Walpole S, Monaco S, Nepravishta R, Angulo J. STD NMR as a Technique for Ligand Screening and Structural Studies. In Methods in Enzymology Volume 615, 2019.

52. Angulo J, Nieto PM. STD-NMR: application to transient interactions between biomolecules-a quantitative approach. Eur Biophys J. 2011;40(12):1357-69.

53. Mayer M, Meyer B. Group epitope mapping by saturation transfer difference NMR to identify segments of a ligand in direct contact with a protein receptor. J Am Chem Soc. 2001;123(25):6108-17.

54. Fuloria NK, Fuloria S. Structural elucidation of small organic molecules by 1D, 2D and multi-dimensional-solution NMR spectroscopy. Anal Bioanal Tech. 2013;11:1-8.

55. Berglund H, Olerenshaw D, Sankar A, Federwisch M, McDonald NQ, Driscoll PC. The three-dimensional solution structure and dynamic properties of the human FADD death domain. J Mol Biol. 2000;302(1):171-88.

56. Pervushin K, Riek R, Wider G, Wüthrich K. Attenuated T2 relaxation by mutual cancellation of dipole-dipole coupling and chemical shift anisotropy indicates an avenue to NMR structures of very large biological macromolecules in solution. Proc Natl Acad Sci U S A. 1997 94(23):12366-71.

57. Becker W, Bhattiprolu KC, Gubensak N, Zangger K. Investigating Protein-Ligand Interactions by Solution Nuclear Magnetic Resonance Spectroscopy. Chemphyschem. 2018;19(8):895-906.

58. Ramachandran KI, Deepa G, Namboori K. Computational Chemistryand Molecular Modeling. Principles and Applications, 1st edition, 2008.

59. Becker O, MacKerell AD, Roux B, Watanabe M. Computational Biochemistry and Biophysics, 1st edition, 2001.

60. Tian C, Kasavajhala K, Belfon KAA, Raguette L, Huang H, Migues AN, et al. ff19SB: Amino-Acid-Specific Protein Backbone Parameters Trained against Quantum Mechanics Energy Surfaces in Solution. J Chem Theory Comput. 2020;16(1):528-52.

61. Cheatham TE, 3rd, Case DA. Twenty-five years of nucleic acid simulations. Biopolymers. 2013;99(12):969-77.

62. Wang J, Wolf RM, Caldwell JW, Kollman PA, Case DA. Development and testing of a general amber force field. J Comput Chem. 2005;25(9):1157-74.

63. Kirschner KN, Yongye AB, Tschampel SM, Gonzalez-Outeirino J, Daniels CR, Foley BL, et al. GLYCAM06: a generalizable biomolecular force field. Carbohydrates. J Comput Chem. 2008;29(4):622-55.

64. Harder E, Damm W, Maple J, Wu C, Reboul M, Xiang JY, et al. OPLS3: A Force Field Providing Broad Coverage of Drug-like Small Molecules and Proteins. J Chem Theory Comput. 2016;12(1):281-96.

65. Huang J, Rauscher S, Nawrocki G, Ran T, Feig M, de Groot BL, et al. CHARMM36m: an improved force field for folded and intrinsically disordered proteins. Nat Methods. 2017;14(1):71-3.

66. Roy K, Kar S, Das RN. Computational Chemistry. In Understanding the Basics of QSAR for Applications in Pharmaceutical Sciences and Risk Assessment. 2015.

67. Meng X, Zhang H, Mezei M, Cui M. Molecular Docking: A powerful approach for structure-baseddrug discovery. Curr Comput Aided Drug Des. 2011;7(2):146-57.

68. Li J, Fu A, Zhang L. An Overview of Scoring Functions Used for Protein-Ligand Interactions in Molecular Docking. Interdiscip Sci. 2019;11(2):320-8.

69. Ryckaert J, Ciccotti G, Berendsen JCH. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. Journal of Computational Physics. 1977;23(3):327-41.

70. Andersen HC. Rattle: A “velocity” version of the shake algorithm for molecular dynamics calculations. Journal of Computational Physics. 1983;52(1):24-34.

71. Moore CC. Ergodic theorem, ergodic theory, and statistical mechanics. Proc Natl Acad Sci U S A. 2015;112(7):1907-11.

72. Huggins DJ. Correlations in liquid water for the TIP3P-Ewald, TIP4P-2005, TIP5P-Ewald, and SWM4-NDP models. J Chem Phys 2012;136(6).

73. Dick TJ, Madura JD. Chapter 5 a review of the TIP4P, TIP4P-ew, TIP5P, and TIP5P-e water models. In Annual reports in computational chemistry. Elsevier. 2005;1:59-74.

74. Miao Y, McCammon JA. Gaussian Accelerated Molecular Dynamics: Theory, Implementation, and Applications. Annu Rep Comput Chem. 2017;13:231-78.

75. Markwick PR, McCammon JA. Studying functional dynamics in bio-molecules using accelerated molecular dynamics. Phys Chem Chem Phys. 2011;13(45):20053-65.

76. Wang J, Arantes PR, Bhattarai A, Hsu RV, Pawnikar S, Huang YM, et al. Gaussian accelerated molecular dynamics (GaMD): principles and applications. Wiley Interdiscip Rev Comput Mol Sci. 2021;11(5).

77. Hamelberg D, Mongan J, McCammon JA. Accelerated molecular dynamics: a promising and efficient simulation method for biomolecules. J Chem Phys. 2004;120(24):11919-29.

78. Oxenoid K, Kim HJ, Jacob J, Sönnichsen FD, Sanders CR. NMR assignments for a helical 40 kDa membrane protein. J Am Chem Soc. 2004 126(16).

79. Piotto M, Saudek V, Sklenář V. Gradient-tailored excitation for single-quantum NMR spectroscopy of aqueous solutions. J Biomol NMR. 1992;2:662-5.

80. Olsson MH, Søndergaard CR, Rostkowski M, Jensen JH. Consistent Treatment of Internal and Surface Residues in Empirical pKa Predictions. Journal of Chemical Theory and Computation. 2011;7(2):525-37.

81. Friesner RA, Banks JL, Murphy RB, Halgren TA, Klicic JJ, Mainz DT, 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).

82. Salomon-Ferrer R, Götz AW, Poole D, Le Grand S, Walker RC. Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 2. Explicit Solvent Particle Mesh Ewald. . J Chem Theory Comput. 2013;9(9):3878-88.

83. Wilker M, Leibfritz R, Kerssebaum R, Bermel W. Gradient selection in inverse heteronuclear correlation spectroscopy. Magn Resn Chem. 1993;31:287-92.

84. Bax A, Davis DG. MLEV-17-Based Two-Dimensional Homonuclear Magnetization Transfer Spectroscopy. J Magn Resn. 1985;65:355-60.

85. Shaw AA, Salaun C, Dauphin J, Ancian B. Artifact-Free PFG-Enhanced Double-Quantum-Filtered COSY Experiments. Journal of Magnetic Resonance. 1993;120(1):110-5.

86. Ancian B, Bourgeois I, Dauphin J, Shaw AA. Artifact-Free Pure Absorption PFG-Enhanced DQF-COSY Spectra Including a Gradient Pulse in the Evolution Period. Journal of Magnetic Resonance. 1997;125(2).

87. Shannon R. Revised Effective Ionic Radii and Systematic Study of Inter Atomic Distances in Halides and Chalcogenides. Acta Crystallographica Section A. 1976;32:751-67.

88. Williamson MP. Using chemical shift perturbation to characterise ligand binding. Prog Nucl Magn Reson Spectrosc. 2013;73:1-16.

89. Giogha C, Scott NE, Wong Fok Lung T, Pollock GL, Harper M, Goddard-Borger ED, et al. NleB2 from enteropathogenic Escherichia coli is a novel arginine-glucose transferase effector. PLoS Pathog. 2021;17(6):e1009658.