Application of Sialic Acid Specific Proteins for Sialic Acid Detection and Quantification

Thesis
presented for the Degree of
DOCTOR OF PHILOSOPHY
By

Jakub Szczepaniak, B.Sc., M.Sc.

Under the supervision of Dr. Brendan O'Connor and
Dr. Michael O'Connell

School of Biotechnology, Dublin City University,
Dublin 9, Ireland
September 2009

Declaration

I hereby certify that this material, which I now submit for assessment on the programme of study leading to the award of ..... (insert title of degree for which registered) is entirely my own work, that I have exercised reasonable care to ensure that the work is original, and does not to the best of my knowledge breach any law of copyright, and has not been taken from the work of others save and to the extent that such work has been cited and acknowledged within the text of my work.

Signed: _____ (Candidate) ID No.: _____ Date: _____

Acknowledgments

There were many people directly and indirectly supporting this project as well as helping and advising me in its realisation and I am grateful to all of them. In particular I would like to thank those who impacted the origin and development of this thesis mostly.

I express my gratitude to Dr. Michael O'Connell and Dr. Brendan O'Connor for giving me the opportunity to pursue this PhD study under their enthusiastic supervision. Their suggestions and constructive guidance encouraged me throughout the project.

My deepest thanks to Dr. Paul Clarke, who was my motivating mentor, for his time and expertise that he kindly shared. I enjoyed his interest in my research as well as the fruitful discussions.

I am grateful to my colleagues for the irreplaceable, supportive and friendly atmosphere in the lab that contributed essentially to the final outcome of my studies.

For support and unwavering belief in me, for endless patience and for encouragement when it was most needed I am indebted to my fiancée Magda.

Table of Contents

List of Figures

List of Tables

Abstract

Sialic Acid (Sia) plays a significant role in a number of essential processes in living organisms ranging from simple charged molecule interactions, stabilization of glycoconjugate conformations, recognition in pathogen invasion and determination of glycoprotein half life. Furthermore, the degree of sialylation of glycoproteins that are used as therapeutics is crucial for their activity and efficacy. Therefore, the requirement for a robust, rapid, specific and sensitive method for Sia quantification has become an important challenge. The goal of this study was to develop a novel lectin-based method for Sia detection and quantification that is: (i) more specific and less prone to interference by other substances than existing colorimetric and enzymatic assays and (ii) more rapid than existing HPLC methods.

This project describes the cloning, expression and purification of Sia binding proteins that were subsequently immobilised onto a number of analytical platforms. A prokaryotic expression system was used and optimized for production of these Sia binding proteins.

Firstly, a terminal sialic acid specific hemagglutinin from Influenza virus was successfully cloned and expressed in *Escherichia coli* and purified by affinity chromatography. Studies on this protein involved procedures aimed at solubilisation of aggregated proteins as well as various solubility-promoting genetic engineering techniques.

Another Sia specific protein, namely SiaP, from *Haemophilus influenzae* was His-tagged and expressed in the KRX strain of *E. coli* and purified using immobilized metal ion affinity chromatography. The activity of this protein was confirmed by Mass Spectrometry (MS) analysis. This MS platform was subsequently used for assay development for free Sia quantification. A positive correlation was found between free Sia concentration and the amount of SiaP molecules that were saturated with the sugar in the sample.

In an alternative approach the SiaP protein was tested using surface plasmon resonance (SPR). A method for stable and orientation-specific immobilization of the recombinant SiaP protein on a Biacore CM5 chip was established. A correlation between free Sia concentration and mass of the immobilized components was established and an operational range for the potential test was successfully identified. This indicates the possibility of application of the SiaP based method for rapid and specific sialic acid detection and quantification.

List of abbreviations

µ	micro
Ĺ	Angstrem
Ac	acetyl
ADOA	4-(acetylamino)-2,4-dideoxy-D-glycero-D-galacto-octonic acid
AMP	adenosine monophosphate
APS	ammonium persulfate
ATP	adenosine triphosphate
BMY-27709	4-amino-5-chloro-2-hydroxy-*N*-9AαH-octahydro-6β-methyl-2H-quinolizin-2α-benzamide
CFA	colonization factor antigens
cfu	colony forming units
CMP	cytidine monophosphate
D	aspartic acid
Da	Dalton
E	glutamic acid
EDC	ethyl-N-(3-diethylaminopropyl)carbodiimide
EDTA	ethylenediaminetetraacetic acid
EGF	epidermal growth factor
ELLA	enzyme linked lectin assay
EPO	erythropoietin
ESPDMA	*N*-ethyl-*N*-(3-sulfopropyl)-3,5-dimethoxy aniline
ESR	extracytoplasmic solute receptor
f	femto
FPLC	fast protein liquid chromatography
Fuc	fucose
Gal	galactose
GalNAc	*N*-Acetylgalactosamine
Gc	glycolyl
GDP	guanidine diphosphate

Glc	glucose
GlcNAc	*N*-Acetylglucosamine
GlcUA	uridine Diphosphoglucuronic Acid
H_2O_2	hydrogen peroxide
HA	hemagglutinin
HRP	horseradish peroxidase
k	kilo
K_d	dissociation constant
Kdn	2-Keto-3-Deoxynonic Acid (deaminated sialic acid)
l	litre
Le^a	sialyl Lewis a
Le^x	sialyl Lewis x
LOS	lipooligosaccharides
LPS	lipopolysaccharide
LSTa	lactoseries tetrasaccharide a
LSTc	lactoseries tetrasaccharide c
Lt	lactyl
M	molar
MAH	*Maackia amurensis* hemagglutinin
MAL	*Maackia amurensis* leukoagglutinin
Man	mannose
MDA	malondialdehyde
Me	methyl
mg	milligram
MS	mass spectrometry
N	Asparagine
n	nano
NA	neuraminidase
NAD	nicotinamide adenine dinucleotide
NADH	reduced form of NAD
NDV	Newcastle disease virus

NEP	nuclear export protein
Neu5,9$Ac_2$8Me	9-*O*-acetyl-8-*O*-methyl-*N*-acetylneuraminic acid
Neu5Ac	*N*-acetylneuraminic acid
Neu5Gc	*N*-glycolylneuraminic acid
Neu5Gc	*N*-glycolylneuraminic acid
Neu5Gc7,8,9Ac_3	7,8,9-tri-*O*-acetyl-*N*-glycolylneuraminic acid
NHS	*N*-hydroxysuccinimide
nm	manometer
NTHi	nontypeable *Haemophilus influenzae*
PCR	polymerase chain reaction
Q	glutamine
R	functional group
R^2	correlation coefficient
RNP	ribonucleoprotein
RU	resonance unit
S	sulphate
SCR	short concensus repeat
SDS	sodium dodecyl sulphate
Sia	sialic acid
SNA	*Sambucus nigra* bark lectin
TAE	Tris Acetate EDTA buffer
TBA	thiobarbituric acid
TRAP	tripartite ATP-independent periplasmic
UDP	uridine diphosphate
WGA	wheat germ agglutinin
Xyl	xylose

Chapter 1
Introduction

1 Introduction

1.1 Sialic acid, the unusual sugar

Sialic acid, also known as *N*-acetyl-neuraminic acid, was discovered in the 1930s by Klenk and Blix (Blix *et al.*, 1957). The terms “sialic acid” and “neuraminic acid” both bear the hallmarks of their original discoveries: Blix isolated sialic acid from submaxillary mucin (sialos = saliva in Greek) and Klenk isolated neuraminic acid derivative from brain glycolipids (neuro- + amine + acid). Some features of sialic acids cause them to be described as “unusual” molecules. Unlike most monosacharides in vertebrate glycoconjugates, which are five- or six-carbon residues, the sialic acids have a nine-carbon backbone. In nature, aside from sialic acid, only pseudaminic and legionaminic acids of bacteria are nine carbon keto-sugars (reviewed in Angata and Varki, 2002). They typically occupy the terminal position in the glycan chain, which makes them suitable for interaction with numerous agents (Varki, 1999). The sialic acids are not ubiquitous in nature and their distribution in various forms of life is irregular (Table 1.1). They are predominantly found in vertebrates, echinoderms and some bacteria but are almost completely absent in lower animals and plants, as opposed to other vertebrate monosacharides (Traving and Schauer, 1998). Their synthesis is also different from other monosacharides as they are produced via condensation of a neutral six-carbon unit with the three-carbon molecule pyruvate (Tanner, 2005). Finally, sialic acids are unusual in the nature of their high-energy nucleotide sugar donor form. While all the other vertebrate monosaccharides are activated in the form of guanine or uridine dinucleotides, e.g., GDP-Man and -Fuc, UDP-Glc, -Gal, -GlcNAc, -GalNAc, -GlcUA, and -Xyl, sialic acids are activated as cytidine mononucleotides, i.e., CMP-Sia molecules (reviewed in Angata and Varki, 2002).

1.1.1 Structure of sialic acid

Sialic acids are common name for a family of over 50 naturally occurring nine-carbon keto sugars derived from the parent compound 2-keto-3-deoxy-5-acetamido-D-glycero-D-galacto-nonulosonic acid (*N*-acetylneuraminic acid [Neu5Ac]), (Figure 1.1), (Vimr *et al.*, 2004; Lehmann *et al.*, 2006).

As the complete chemical names of sialic acids are too cumbersome for routine use a uniform and simple nomenclature system is being increasingly used, in which the abbreviation Neu denotes the core structure neuraminic acid, and Kdn denotes the core structure 2-keto-3-deoxynononic acid. Various substitutions are then designated by letter codes (Ac = acetyl, Gc = glycolyl, Me = methyl, Lt = lactyl, S = sulfate), and these are listed along with numbers indicating their location relative to the carbon positions. For example, *N*-glycolylneuraminic acid is Neu5Gc, 9-*O*-acetyl-8-*O*-methyl-*N*-acetylneuraminic acid is Neu5,9$Ac_2$8Me, and 7,8,9-tri-*O*-acetyl-*N*-glycolylneuraminic acid is Neu5Gc7,8,9Ac_3. If there is no certainty what type of sialic acid is present at a particular location, the generic abbreviation Sia should be used. If other partial information is available, this can be incorporated, for example, a Sia of otherwise unknown type with an acetyl substitution at the C9 position could be written as Sia9Ac (Varki, 1999).

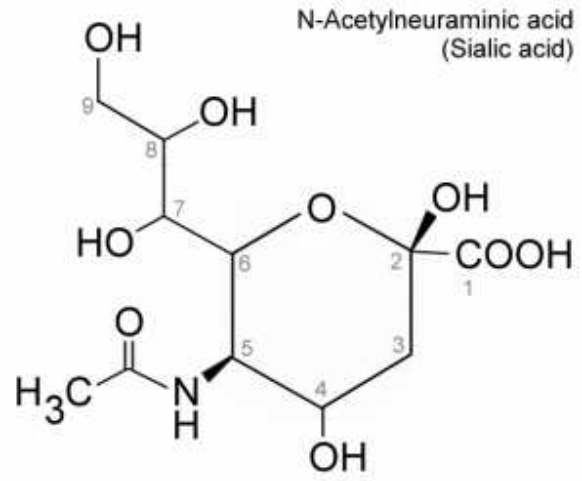


Figure 1.1. Structure of N-Acetylneuraminic acid (Sialic acid).

The Sias' specific structural features are the substitution of either an N-acetyl group (in Neu5Ac) or a hydroxyl group (in Kdn) at position 5 and the carboxyl group at position 1 that confers a negative charge on the molecule under physiological conditions and characterizes it as a strong organic acid (Angata and Varki, 2002). The unsubstituted form, neuraminic acid, does not exist in nature. Sia molecules can be substituted in more than one position with acetyl, methyl, hydroxyl and other groups resulting in their broad diversity (Table 1.1). The variability of Sia is further extended by their location on cells and

molecules. They usually represent the terminal sugar residue of a glycan chain, i.e. they are linked via C2 to position 3 or 6 of the penultimate sugar (most commonly galactose) or to position 8 of another Sia molecule, respectively. The α(2→8) linked homopolymers are known as polysialic acid (Lehmann *et al.*, 2006).

Table 1.1. Sialic Acids: Occurrence and Structural Divergence. Table from Angata and Varki (2002). [a] Abbreviations used: V, vertebrates; E, echinoderms; Ps, protostomes (insects and molluscs); Pz, protozoa; F, fungi; B, bacteria.[b] Present only as bound form.[c] Biosynthetic intermediate.[d] Present only as free form.

Compound name	**Abbreviation**	**Occurrence**[a]
neuraminic acid	Neu	V[b]
neuraminic acid 1,5-lactam	Neu1,5lactam	V
5-*N*-acetylneuraminic acid	Neu5Ac	V, E, Ps, Pz, F, B
5-*N*-acetyl-4-*O*-acetylneuraminic acid	Neu4,5Ac_2	V
5-*N*-acetyl-7-*O*-acetylneuraminic acid	Neu5,7Ac_2	V, Pz, B
5-*N*-acetyl-8-*O*-acetylneuraminic acid	Neu5,8Ac_2	V, B
5-*N*-acetyl-9-*O*-acetylneuraminic acid	Neu5,9Ac_2	V, E, Pz, F, B
5-*N*-acetyl-4,9-di-*O*-acetylneuraminic acid	Neu4,5,9Ac_3	V
5-*N*-acetyl-7,9-di-*O*-acetylneuraminic acid	Neu5,7,9Ac_3	V, B
5-*N*-acetyl-8,9-di-*O*-acetylneuraminic acid	Neu5,8,9Ac_3	V
5-*N*-acetyl-4,7,9-tri-*O*-acetylneuraminic acid	Neu4,5,7,9Ac_4	V
5-*N*-acetyl-7,8,9-tri-*O*-acetylneuraminic acid	Neu5,7,8,9Ac_4	V
5-*N*-acetyl-4,7,8,9-tetra-*O*-acetylneuraminic acid	Neu4,5,7,8,9Ac_5	V
5-*N*-acetyl-9-*O*-lactylneuraminic acid	Neu5Ac9Lt	V
5-*N*-acetyl-4-*O*-acetyl-9-*O*-lactylneuraminic acid	Neu4,5$Ac_2$9Lt	V
5-*N*-acetyl-7-*O*-acetyl-9-*O*-lactylneuraminic acid	Neu5,7$Ac_2$9Lt	V
5-*N*-acetyl-8-*O*-methylneuraminic acid	Neu5Ac8Me	V, E
5-*N*-acetyl-9-*O*-acetyl-8-*O*-methylneuraminic acid	Neu5,9$Ac_2$8Me	V, E
5-*N*-acetyl-8-*O*-sulfoneuraminic acid	Neu5Ac8S	V, E
5-*N*-acetyl-4-*O*-acetyl-8-*O*-sulfoneuraminic acid	Neu4,5$Ac_2$8S	V, E
5-*N*-acetyl-9-*O*-phosphoneuraminic acid	Neu5Ac9P	V[c,d]
5-*N*-acetyl-2-deoxy-2,3-didehydroneuraminic acid	Neu2en5Ac	V[d]
5-*N*-acetyl-9-*O*-acetyl-2-deoxy-2,3-didehydroneuraminic acid	Neu2en5,9Ac_2	V[d]
5-*N*-acetyl-2-deoxy-2,3-didehydro-9-*O*-lactylneuraminic acid	Neu2en5Ac9Lt	V[d]
5-*N*-acetyl-2,7-anhydroneuraminic acid	Neu2,7an5Ac	V[d]
5-*N*-acetylneuraminic acid 1,7-lactone	Neu5Ac1,7lactone	V
5-*N*-acetyl-9-*O*-acetylneuraminic acid 1,7-lactone	Neu5,9$Ac_2$1,7lactone	V
5-*N*-acetyl-4,9-di-*O*-acetylneuraminic acid 1,7-lactone	Neu4,5,9$Ac_3$1,7lactone	V
5-*N*-glycolylneuraminic acid	Neu5Gc	V, Pz, F
4-*O*-acetyl-5-*N*-glycolylneuraminic acid	Neu4Ac5Gc	V

Compound name	**Abbreviation**	**Occurrence**[a]
7-*O*-acetyl-5-*N*-glycolylneuraminic acid	Neu7Ac5Gc	V
8-*O*-acetyl-5-*N*-glycolylneuraminic acid	Neu8Ac5Gc	V
9-*O*-acetyl-5-*N*-glycolylneuraminic acid	Neu9Ac5Gc	V, E
4,7-di-*O*-acetyl-5-*N*-glycolylneuraminic acid	Neu4,7$Ac_2$5Gc	V
4,9-di-*O*-acetyl-5-*N*-glycolylneuraminic acid	Neu4,9$Ac_2$5Gc	V
7,9-di-*O*-acetyl-5-*N*-glycolylneuraminic acid	Neu7,9$Ac_2$5Gc	V
8,9-di-*O*-acetyl-5-*N*-glycolylneuraminic acid	Neu8,9$Ac_2$5Gc	V
7,8,9-tri-*O*-acetyl-5-*N*-glycolylneuraminic acid	Neu7,8,9$Ac_3$5Gc	V
5-*N*-glycolyl-9-*O*-lactylneuraminic acid	Neu5Gc9Lt	V
4-*O*-acetyl-5-*N*-glycolyl-9-*O*-lactylneuraminic acid	Neu4Ac5Gc9Lt	V
8-*O*-acetyl-5-*N*-glycolyl-9-*O*-lactylneuraminic acid	Neu8Ac5Gc9Lt	V
4,7-di-*O*-acetyl-5-*N*-glycolyl-9-*O*-lactylneuraminic acid	Neu4,7$Ac_2$5Gc9Lt	V
7,8-di-*O*-acetyl-5-*N*-glycolyl-9-*O*-lactylneuraminic acid	Neu7,8$Ac_2$5Gc9Lt	V
5-*N*-glycolyl-8-*O*-methylneuraminic acid	Neu5Gc8Me	E
9-*O*-acetyl-5-*N*-glycolyl-8-*O*-methylneuraminic acid	Neu9Ac5Gc8Me	E
7,9-di-*O*-acetyl-5-*N*-glycolyl-8-*O*-methylneuraminic acid	Neu7,9$Ac_2$5Gc8Me	E
5-*N*-glycolyl-8-*O*-sulfoneuraminic acid	Neu5Gc8S	V, E
5-*N*-glycolyl-9-*O*-sulfoneuraminic acid	Neu5Gc9S	E
5-*N*-(*O*-acetyl)glycolylneuraminic acid	Neu5GcAc	V
5-*N*-(*O*-methyl)glycolylneuraminic acid	Neu5GcMe	E
2-deoxy-2,3-didehydro-5-*N*-glycolylneuraminic acid	Neu2en5Gc	V[d]
9-*O*-acetyl-2-deoxy-2,3-didehydro-5-*N*-glycolylneuraminic acid	Neu2en9Ac5Gc	V[d]
2-deoxy-2,3-didehydro-5-*N*-glycolyl-9-*O*-lactylneuraminic acid	Neu2en5Gc9Lt	V[d]
2-deoxy-2,3-didehydro-5-*N*-glycolyl-8-*O*-methylneuraminic acid	Neu2en5Gc8Me	E[d]
2,7-anhydro-5-*N*-glycolylneuraminic acid	Neu2,7an5Gc	V[d]
2,7-anhydro-5-*N*-glycolyl-8-*O*-methylneuraminic acid	Neu2,7an5Gc8Me	E[d]
5-*N*-glycolylneuraminic acid 1,7-lactone	Neu5Gc1,7lactone	V
2-keto-3-deoxynononic acid	KDN	V, B
5-*O*-acetyl-2-keto-3-deoxynononic acid	KDN5Ac	V
7-*O*-acetyl-2-keto-3-deoxynononic acid	KDN7Ac	V
9-*O*-acetyl-2-keto-3-deoxynononic acid	KDN9Ac	V
4,5-di-*O*-acetyl-2-keto-3-deoxynononic acid	KDN4,5Ac_2	V
4,7-di-*O*-acetyl-2-keto-3-deoxynononic acid	KDN4,7Ac_2	V
5,9-di-*O*-acetyl-2-keto-3-deoxynononic acid	KDN5,9Ac_2	V
7,9-di-*O*-acetyl-2-keto-3-deoxynononic acid	KDN7,9Ac_2	V
8,9-di-*O*-acetyl-2-keto-3-deoxynononic acid	KDN8,9Ac_2	V
2-keto-3-deoxy-5-*O*-methylnononic acid	KDN5Me	B
2-keto-3-deoxy-9-*O*-phosphononononic acid	KDN9P	V[c,d]

Sias autonomously form a pyranose (six-membered) ring in solution via intramolecular hemiketal condensation and adopt a $^2C_5$ chair conformation (Figure 1.2). In natural glycoconjugates, Sias exist only in the α-configuration, except in the high-energy donor form CMP-Sias, where the anomeric carbon is in the β-configuration (α and β configurations refer to the C7 carbon and C1 carboxyl groups being in a trans and cis orientation, respectively) (Angata and Varki, 2002).

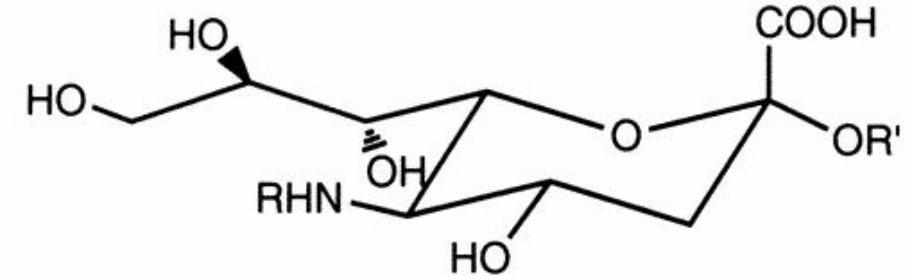


Figure 1.2. Structures of sialic acids: (A) Neuraminic acid (Neu, R = H), *N*-acetylneuraminic acid (Neu5Ac, R = CH_3CO-) and *N*-glycolylneuraminic acid (Neu5Gc, R = $HOCH_2CO-$). Figure from (Angata and Varki, 2002).

1.1.2 Natural functions of sialic acid

Sialic acid as a small, hydrophilic and negatively charged molecule has a simple physicochemical impact on its environment. Involvement of Sias in such processes as binding and transport of positively charged molecules is attributed to their negative charge (reviewed in Traving and Schauer, 1998). Similarly attraction and repulsion phenomena between cells and molecules are often driven by highly charged Sias (reviewed in Traving and Schauer, 1998 and in Vimr *et al.*, 2004). Furthermore, the negative charge of Sias present on glycoconjugates stabilizes their correct conformation (reviewed in Traving and Schauer, 1998; Vimr *et al.*, 2004 and in Matsuno *et al.*, 2008). However, its main function is that of specific phenomena related to cellular and molecular recognition.

As a very diverse group, Sias participate in various recognition processes between cells and molecules. They are antigenic determinants, for example, of blood groups and as such enable the immune system to distinguish between self and non-self structures according to their Sia pattern. They serve as ligands for lectins and antibodies and in this way play a regulatory role in many physiological processes. Nevertheless, many pathogenic agents

such as toxins (e.g. cholera toxin), viruses (e.g. influenza), bacteria (e.g. *Escherichia coli*, *Helicobacter pylori*) and protozoa (e.g. *Trypanosoma cruzi*) also bind to host cells via Sia-containing receptors (reviewed in Lehmann *et al.*, 2006).

Another important feature of Sias, that seems to be in direct contrast to their recognition function, is the masking of cells and molecules. The Sia rich mucous layer of the respiratory epithelium serves as a protective shield against pathogens. This thick glycan coating efficiently insulates receptors on the surface of the host cells from bacteria and viruses (reviewed in Traving and Schauer, 1998). The density of Sia structures [more than 10 million molecules per human erythrocyte, (Varki, 1999)] on the surface of erythrocytes determines their lifespan. Red blood cells that have lost certain numbers of terminal Sias, from their surface glycans, over time, have penultimate galactoses of the glycan chains exposed to the environment (Aminoff *et al.*, 1977). This serves as a signal for macrophages to phagocytose such cells as aged (Bratosin *et al.*, 1995) (Figure 1.3).

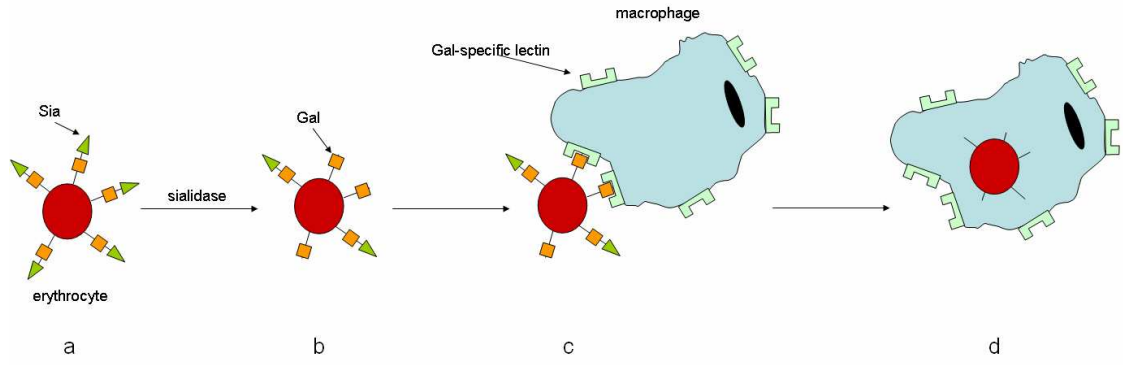


Figure 1.3. The masking function of sialic acids: mechanism of binding (b) and phagocytosis (c) of sialidase-treated erythrocytes (a) by macrophages. Sialylation enables these cells to have a long lifetime in circulation.

Likewise, various serum glycoproteins are removed from the bloodstream by hepatocytes after losing terminal Sias and exposure of subterminal galactoses (Egrie *et al.*, 2001; Fernandes *et al.*, 2001). Similarly, masking of endogenous structures by Sias protects cancer cells and some pathogens from being recognized by immune system cells (Traving *et al.*, 1998). In terms of cancer, the higher degree of sialylation has an immunosuppressive effect and corresponds to higher sialyltransferase activities often seen in tumour tissues.

Thus, terminal galactose residues that would otherwise inhibit further cell growth and spreading are masked. This might be one reason for the loss of growth and division inhibition of cancer cells (Wieser *et al.*, 1995; Traving and Schauer., 1998).

Sia was also found to be a potent defence molecule against oxidative damage being able to detoxify hydrogen peroxide. Equimolar amounts of Sia and H_2O_2 react with each other resulting in the production of non-toxic 4-(acetylamino)-2,4-dideoxy-*D*-glycero-*D*-galacto-octonic acid (ADOA) (Iijima *et al.*, 2004).

Thus, as Sias display such a diverse functionality, the changes in their abundance as well as the sialylation pattern of glycoconjugates is seen during many important physiological and pathological processes.

1.1.3 Sialic acids in human physiology and pathology

The very diverse group of Sias play significant roles in human physiology. The defined level of Sia expression and its quality in the body as well as the sialylation pattern of various entities are crucial for the correct functioning of the organism. Uncontrolled alterations in any of them results in serious disorders. Increased serum concentrations of Sias have been detected in a number of cancerous states, such as pancreatic cancer, skin squamous cell carcinoma, lung, prostate, breast, ovary, colon, and thyroid cancers (Diamantopoulou *et al.*, 1999; Paszkowska *et al.*, 2000; Wongkham *et al.*, 2003; Marzouk *et al.*, 2007). Sia levels have been correlated with the tumour size, positive lymph nodes metastasis, and advanced clinical stage in patients with head and neck squamous cell carcinoma. Moreover, Sia concentrations are strongly associated with the patient survival rate (Marzouk *et al.*, 2007). Enhanced expression of α(2→6)-linked Sias on N-glycans often correlates with human cancer progression, metastatic spread, and poor prognosis. The changes of sialylation pattern are often a reflection of gene expression aberrations. Up-regulation of the ST6GALI gene encoding the enzyme β-galactoside: α(2→6)-sialyltransferase (ST6Gal-I) leads to increased Sia α(2→3)Gal β(1→4)GlcNAc (Sia6LacNAc) production, that is found on the surface of many cancer cells (Hedlund *et al.*, 2008).

Sia increased expression is reported in carcinomas of the colon, breast and cervix,

choriocarcinomas, acute myeloid leukemia, and in some brain tumours (Marzouk *et al.*, 2007; Hedlund *et al.*, 2008). Elevated plasma Sia concentration is strongly related to the presence of microvascular complications in type I diabetes (Crook *et al.*, 2001) and cardiovascular morbidity in the general population (Lindberg *et al.*, 1991; Crook, 1993). Sia level of low-density lipoprotein LDL is also elevated in coronary atherosclerotic patients in comparison to healthy subjects (Orekhov *et al.*, 1989). The overproduction of Sia and its elevated excretion in the urine is observed in the patients with the rare genetic disease sialuria (Wang and Brand-Miller, 2003).

Changes in Sia level in the human organism can be observed during some physiological states such as pregnancy. An increase in concentration of Sia in maternal saliva and plasma is associated with pregnancy. Sia increases from about 50 mg/l at 10 weeks gestation to over 150 mg/l at 21–40 weeks gestation in saliva corresponding to the period of rapid Sia accumulation in the fetal brain (Salvolini *et al.*, 1998).

A definite sialylation level is also crucial for activity and stability of glycoprotein hormones and enzymes such as erythropoietin (EPO) (Egrie *et al.*, 2001) and asparaginase (Fernandes and Gregoriadis, 2001; Egrie and Browne, 2001). EPO is a glycoprotein hormone produced in kidneys with the main function of the stimulating red bloo cells production (Fried, 2009). Furthermore, EPO has been used as a drug since 1988 primarily for the clinical treatment of anemia, especially anemia caused by renal failure. The number of the Sia residues on the glycans' terminal positions was found to affect the speed of catabolism and biologic activity of EPO (Egrie and Browne, 2001; Marzouk *et al.*, 2007). It was found that the removal of terminal Sia from carbohydrate chains of EPO increased their in vitro activity but abolished the in vivo activity completely (Marzouk *et al.*, 2007).

Sia has become increasingly interesting for the biopharmaceutical industry producing protein-based drugs because of its unique features that increase activity and stability of sialylated drugs *in vivo*. Furthermore, sialyation and hypersialylation (Fernandes and Gregoriadis, 2001) of heterogeneous glycoproteins in humans efficiently decreases their immunogenic potential in comparison to non-sialylated forms Egrie and Browne, 2001). Therefore, a growing number of glycosylated pharmaceutical products are being studied with regards to their sialylation and the effects of it (Byrne *et al.*, 2007).

Sias have a high potential to serve as a diagnostic marker for many diseases and

physiological states. By monitoring their levels in the organism it may be possible to determine and monitor pathological disorders of such importance as cancer. Furthermore, accurate estimation of Sia level in glycoprotein-based drugs as well as determination of their sialylation pattern enables production of more efficient compounds. This lowers the costs of production as less substance is needed to achieve certain action, which is certainly beneficial for the pharma industry. The concentration of Sias in the bio-processors during the production is of great importance for the quality of the final product and should also be monitored.

Since the discovery of Sia almost a century ago many methods for its analysis have been developed. The requirement for reliable and sensitive assays for Sia detection and quantification has been addressed with various approaches applied in biochemistry, medicine and the pharma industry. Furthermore, many potential methods are under scrutiny and may be developed into useful tools in the near future.

1.1.4 Sialic acid detection and quantification methods

1.1.4.1 Colorimetric methods for sialic acid quantification

The first Sia detection methods were modifications of assays which had been previously used for the measurement of other carbohydrates. They involved orcinol, resorcinol, diphenylamine, direct Ehrlich, tryptophan-perchloric acid, hydrochloric acid, and sulfuric-acetic acid procedures (Warren, 1959).

The two most common colorimetric methods for the measurement of Sia levels are the thiobarbituric acid (TBA) assay of Warren, (1959) and the resorcinol method of Svennerholm (Svennerholm, 1956, 1957). These colorimetric methods are based on the conversion of Sia into chromophores that can be monitored spectrophotometrically.

1.1.4.2 Thiobarbituric acid (TBA assay, Warren's assay)

In this method the Sia is oxidized in strong acid solution by treatment with periodic acid. The periodate oxidation product is subsequently coupled with 2-thiobarbituric acid and

forms a red chromophore with λ_{max} = 549 nm (Warren, 1959). This method is suitable for free Sia only. Color production varies linearly with the concentration of N-acetylneuraminic acid over the range of 0.01 to 0.06 µmole. In terms of sensitivity, the molar extinction coefficient is 57,000 (Warren, 1959) and also described as 3-300 µM (Chaplin and Kennedy, 1994). The procedure takes about one hour and suffers from malondialdehyde (MDA) interference in the sample (Sobenin *et al.*, 1998).

1.1.4.3 Resorcinol method of Svennerholm

The resorcinol method described in 1957 by Svennerhold is similar to the above thiobarbituric acid method (Svennerhold, 1957). The Sia after periodate oxidation in acidic conditions is coupled with resorcinol to form a chromophore with λ_{max} = 630 nm (Jourdian *et al.*, 1971). The chromogens formed from free and glycosidically bound Sia by periodate oxidation show marked differences in stability. Therefore by suitable variation of time and temperature, it is possible to convert the assay procedure to a differential method for determining total, glycosidically bound and, by calculation, also free Sia (Jourdian *et al.*, 1971). This method is however less sensitive then the thiobarbituric acid method with the molar extinction coefficient of 27,900 (Jourdian *et al.*, 1971).

1.1.4.4 Enzymatic methods for sialic acid quantification

All the enzymatic reactions involve initial enzymatic release of Sia from sialylated glycans by neuraminidase. The free Sia is than subjected to *N*-acetylneuraminic acid aldolase to generate *N*-acetyl-*D*-mannosamine and pyruvic acid. One of the Sia cleavage products is subsequently enzymatically converted into a chromophore that can be accurately measured spectrophotometrically (Figure 1.4).

Sialylated glycan → free sialic acid + glycan

Free sialic acid → *N*-acetyl-*D*-mannosamine + pyruvic acid

Figure 1.4. Common reactions in the enzymatic measurement of sialic acid

Detection methods involving pyruvate

In the method by Sugahara (1980) the oxidation of pyruvate by pyruvate oxidase produces hydrogen peroxide which is used as a substrate for peroxidase to produce 4-amino-antipyrine, a colorimetry measured dye.

The pyruvic acid can be also reduced in the presence of reduced form of nicotinamide adenine dinucleotide (NADH) by lactate dehydrogenase. Then the concentration of nicotinamide adenine dinucleotide (NADH) is measured spectrophotometrically at 349 nm (Kolisis, 1986).

Both of these methods suffer from interference, mostly by pyruvic acid which is normally present e.g. in the serum (Teshima *et al.*, 1988).

Detection methods based on *N*-acetyl-*D*-mannosamine

The assay developed by Teshima in 1988 was based on conversion of *N*-acetyl-*D*-mannosamine into *N*-acetyl-*D*-glucosamine by acylglucosamine 2-epimerase. Then *N*-acetyl glucosaminic acid and hydrogen peroxide are produced by *N*-acetylhexosamine oxidase. Ultimately, hydrogen peroxide is detected using peroxidase, 4-aminoantipyrine and *N*-ethyl-*N*-(3-sulfopropyl)-3,5-dimethoxy aniline (Figure 1.5) (Teshima *et al.*, 1988). This method is affected by ascorbic acid; however, its interference could be prevented by adding ascorbate oxidase (Teshima *et al.*, 1988). The procedure takes about 10 minutes.

Free sialic acid → *N*-acetyl-*D*-mannosamine + pyruvic acid

N-acetyl-*D*-mannosamine →*N*-acetyl-*D*-glucosamine

N-acetyl-*D*-glucosamine → *N*-acetyl glucosaminic acid + H_2O_2

H_2O_2 + 4-aminoantipyrine + ESPDMA → quinine imine dye (λ_{max}, 590 nm)

Figure 1.5. Teshima method of sialic acid detection and quantification. ESPDMA = N-ethyl-N-(3-sulfopropyl)-3,5-dimethoxy aniline.

1.1.4.5 Chromatographic methods

A fluorometric detection following High-performance liquid chromatography (HPLC) separation of previously labelled Sia has been successfully applied to determine 13 members of the Sia family in tissues samples from rats and mice (Morimoto *et al.*, 2001).

The HPLC method is sensitive and specific with a detection limit of 25 fmol (Hara *et al.*, 1989; Matsuno and Suzuki, 2008). However it requires long analysis time, expensive equipment, columns and may be subjected to interference by contaminants in the sample (Matsuno and Suzuki, 2008).

1.1.4.6 Amperometric biosensor based method

The Sia amperometric biosensor operation is based on the sequential action of two enzymes (*N*-acetylneuraminic acid aldolase and pyruvate oxidase) that ultimately produce hydrogen peroxide, which is then detected by anodic amperometry at the platinum electrode. The biosensor presents several advantages such as simple preparation, low cost, high sensitivity and selectivity, reasonable operational stability (at least 8 days), and successful applicability in biological samples. The limit of detection is 10 µM, and the response is linear to 3.5 mM (Marzouk *et al.*, 2007). This type of assay is not commercially available nor has it has been widely used for Sia detection and hence it will not be further discussed.

1.1.4.7 Conclusions

Several methods for colorimetric analysis of Sias developed to date have been found to be not specific or sensitive enough to get an accurate determination of Sias in biological materials, which usually contain very low Sia concentrations and large amounts of substances that may interfere with the colorimetric tests (Romero *et al.*, 1997). These assays are not only relatively insensitive, but in general have a low specificity and cannot be applied to the direct measurement of Sias in tissues and in other unpurified biological materials (Warren, 1959). The increased specificity and sensitivity of chromatographic methods are at the price of extended time of analysis and requirement for expensive equipment and highly skilled personnel. This clearly indicates that the currently available methods of specific Sias detection and quantification are not rapid and easy to use while the more robous methods suffer from unspecificity and interference. Presently there are no universal methods that would be specific, precise, sensitive, reliable, quick, uncomplicated and, preferably, inexpensive. However, as it has been presented, the role of Sia in living organisms is of great significance and thus the regulation of its metabolism and precise

estimation of its level is also critical. Therefore, unsurprisingly, mechanisms with the capacity to maintain Sia exist in nature and Sia turnover is, indeed, closely and precisely monitored by living organisms. In these mechanisms crucial functions are played by proteins that can bind Sia with high specificity and high affinity. The aim of this study was to explore the potential of such proteins in order to develop a novel Sia detection and quantification method that would fulfil a niche left unfulfilled (specific and quick method) by the already existing methods. A method that would be specific, precise, sensitive, reliable, quick, uncomplicated and also inexpensive.

1.2 Sialic acid binding proteins

The great importance of Sia in various biological processes is reflected in the abundance of natural Sia binders present in living organisms. Sia binding proteins play a significant role in cell recognition (Lasky, 1995; Bevilacqua *et al.*, 1991), pathogen infection (Spackman, 2008; Guo *et al.*, 2009), Sia acquisition (Johnston *et al.*, 2007; Severi *et al.*, 2007) and detachment from the host cell and spread of viruses (Spackman, 2008; Lee and Saif, 2009). These proteins, although from distinct organisms often share some similarity (Johnston *et al.*, 2007; Lehmann *et al.*, 2006).

1.2.1 Viral sialic acid binding proteins

Many viruses utilize Sias to facilitate attachment to host cells (Angata and Varki, 2002; Lehmann *et al.*, 2006), although their degree of dependence on Sias for this purpose varies. Haemagglutinins of influenza viruses (A, B, and C) (Harms *et al.*, 1996; Xu *et al.*, 1996; Ito *et al.*, 1997), Newcastle disease virus [NDV, an avian pathogen (Suzuki *et al.*, 1985)], mouse polyoma virus (Cahan *et al.*, 1983), Sendai virus [a rodent pathogen (Suzuki *et al.*, 1985)], mouse hepatitis virus (Regl *et al.*, 1999), and many others (Angata and Varki, 2002) have been isolated and shown to bind Sias. Viruses often possess two Sia binding proteins of “contradictory” functions: haemagglutinins, that facilitate cell entry by recognizing and binding to specific receptors containing terminal Sia on the host cell surface and neuraminidases, that cleave off terminal Sia of the human receptors which enables the virus to detach from the cell surface at the end of one cycle (Figure 1.6) (Spackman, 2008).

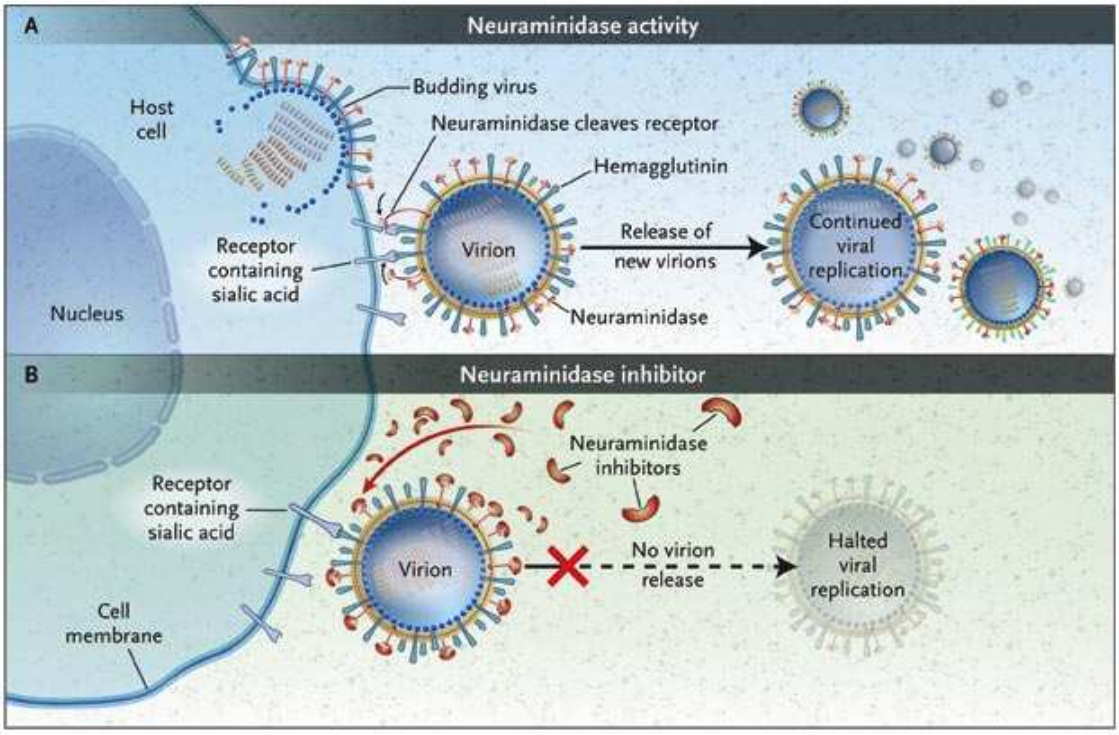


Figure 1.6. Mechanism of Action of Neuraminidase Inhibitors. Panel A shows the action of neuraminidase in the continued replication of virions in influenza infection. The replication is blocked by neuraminidase inhibitors (Panel B), which prevent virions from being released from the surface of infected cells. Figure from Moscona (2005).

1.2.2 Bacterial adhesins

Many bacterial pathogens also utilize Sia receptors on the surface of host cells. Prokaryotic proteins involved in cell attachment are collectively known as adhesins. Bacterial adhesins are often expressed in a strain-specific way and can influence the range of tissues the strain can infect or colonise (Ofek and Sharon, 1990; Kline *et al.*, 2009). Strains shown to use sialoglycoconjugates as attachment sites express either S-fimbriae, K99-fimbriae, the F41 adhesin or one of the colonization factor antigens (CFA) (Lehmann *et al.*, 2006). Such examples include *Helicobacter pylori* (Miller-Podraza *et al.*, 1996) and *E. coli* strain K99 (Ono *et al.*, 1989). The former is an etiological agent of peptic ulcers in humans, and the latter causes lethal dysentery among piglets and calves. *H. pylori* express two different adhesins (in an environment-dependent manner) which can recognize Sias (Miller-Podraza

et al., 1997). SabA recognizes all terminal α(2→3)-linked Sia regardless of the underlying glycan structure. The neutrophil-activating protein, HPNAP, binds solely Neu5Acα(2→3)Galβ(1→4)GlcNAcβ(1→3)Galβ(1→4)GlcNAc structures (Teneberg *et al.*, 1997). The recognition is mediated solely by the SabA adhesion (Unemo *et al.*, 2005).

The adhesins expressed by *E. coli* K99 strain shows high specificity toward Neu5Gcα(2→3)Galβ(1→4)Glc structure on glycolipids, which is abundantly expressed in the gastrointestinal tract of piglets (Yuyama *et al.*, 1993; Angata and Varki, 2002). The binding site of K99-fimbriae adhesin is found in the major subunit. The presence of a hydrophobic region close to the binding site seems to enhance Sia binding affinity, which favours Neu5Gc over Neu5Ac. The specific recognition of Neu5GcLacCer by K99-fimbriated *E. coli* might contribute to host specificity, since humans and animals that lack Neu5Gc cannot be infected (Lehmann *et al.*, 2006).

S-fimbriae of *E. coli* were found to preferentially bind to gangliosides carrying Neu5Gcα(2→3)Gal and Neu5Acα(2→8)Neu5Ac structures, with the C-8 and C-9 hydroxyl groups on Sia being required for recognition (Hanisch *et al.*, 1993). The adhesion protein, SFaS, is a minor component of the multi-subunit S-fimbriae (Hacker *et al.*, 1993).

Another example of Sia-specific adhesin found in *E. coli* is F41. This 28 kDA protein is specific for glycophorin A with a clear selectivity for the M blood type (Brooks *et al.*, 1989).

Of the CFA the most extensively studied are CFAI, CFAII and CFAIV. Whereas CFAI is a single fimbrial antigen, CFAII and CFAIV are composed of antigenically distinct structures called coli surface antigens (Lehmann *et al.*, 2006).

The Sia specific adhesisns are also utilized by protozoan pathogens. They were found in *Tritrichomonas foetus* (Babal *et al.*, 1999), *Entamoeba histolytica* (Feingold *et al.*, 1984) and *Plasmodium falciparum* (the protozoan which causes the most virulent form of malaria [Orlandi *et al.*, 1992]).

1.2.3 Bacterial toxins

A separate group of Sia-specific proteins found in bacteria are toxins. The toxicity of these soluble lectins results from their ability to catalytically modify macromolecules that are required for essential cellular functions such as vesicular trafficking, cytoskeletal assembly, signalling or protein synthesis. To reach their targets, these proteins bind specific surface receptors before endocytosis and translocation across the internal membrane can occur. These toxins classically bind to oligosaccharide receptors on host cell surfaces, and many of them show high specificity toward Sia, generally located on gangliosides (reviewed in Lehmann *et al.*, 2006). One of the best examples of a Sia-binding soluble lectin is cholera toxin, belonging to the AB5 family, produced by *Vibrio cholerae* (Richards *et al.*, 1979). It is composed of five B subunits and an A subunit (Figure 1.7). The B subunits show specific binding to a sialylated glycolipid (ganglioside GM1), delivering the A subunit to the cytosol. This in turn causes overactivation of an intracellular signaling pathway (adenylate cyclase, producing cyclic AMP) in gastrointestinal epithelial cells, causing severe diarrhea and electrolyte imbalance (Vanden Broeck *et al.*, 2007; Sanchez and Holmgren, 2008). Other notable examples of Sia-dependent toxins are those from *Clostridium tetani* (Bizzini, 1979; Eisel *et al.*, 1986) and *Clostridium botulinum* (Dolly and Aoki, 2006), causing tetanus and botulism, respectively (Angata and Varki, 2002).

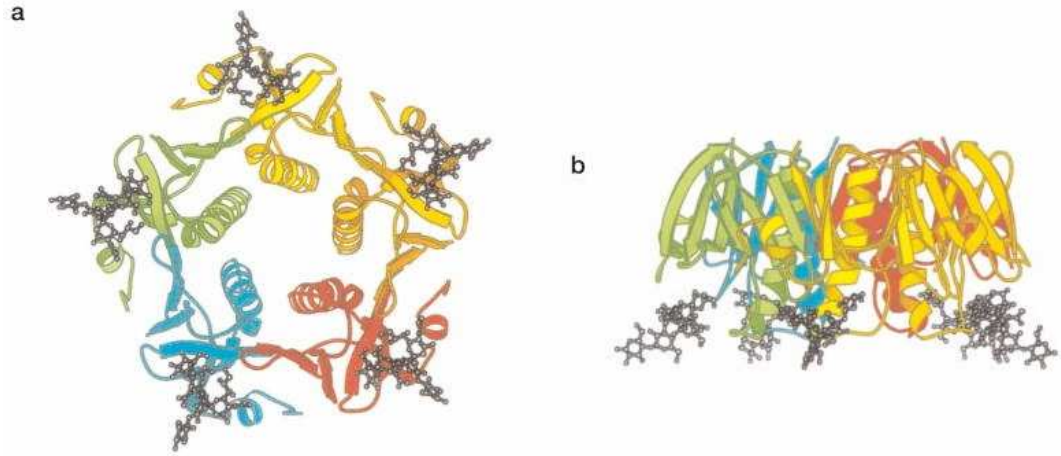


Figure 1.7. Crystal structure of cholera toxin B-subunit pentamer with bound GM1 pentasaccharide shown from the bottom (a) and from the side (b). Figure from Varki (1999).

1.2.4 Animal endogenous sialic acid-specific lectins

It is obvious that binding of Sia receptors on the surface of the cells by pathogens is often detrimental for the host. However, expression of such receptors and their specific recognition by "self" proteins is necessary for many physiological processes. The great variety of such proteins has already been identified and characterized (Angata and Varki, 2002) and others are yet to be found. The majority of the identified animal, endogenous, lectins were found in the immune system. The first reported example of a vertebrate Sia-recognizing lectin was a part of the alternative pathway of complement, one of the earliest response components of the innate immune system (Ram *et al.*, 1998). The protein consists of 20 short consensus repeats (approximately 60 amino acids per repeat unit, also called "sushi" domains). Some of the repeat units are involved in Sia recognition and require an intact unsubstituted Sia side chain (Ripoche *et al.*, 1988). Factor H is a regulatory (inhibitory) factor, and binding of it to cell surface glycoconjugates containing Sias prevents the alternative complement pathway from inadvertently attacking "self" cells (Nilsson and Mueller-Eberhard, 1965). Foreign cells which are not covered with Sias will not be protected by factor H and are hence exposed to the attack by complement (Ram *et al.*, 1998; Whaley *et al.*, 1976).

Particular carbohydrate structures present on human cells containing Sias and fucose called sialyl Lewis x (Le^x) and sialyl Lewis a (Le^a) are recognized by a group of lectins called selectins. The family of three C-type lectins is defined by a shared structural motif and calcium requirement for carbohydrate recognition (revied in Varki, 1999). These cell adhesion molecules are critically involved in the regulation of leukocyte traffic (Cambia and Figdor, 2003) . The selectins are typical type I proteins composed of a tandem array of discrete protein domains. These include an amino terminal C-type lectin domain, a single epidermal growth factor (EGF)-like domain, from two to nine short consensus repeat (SCR) domains, a single membrane spanning region, and a cytoplasmic tail (Kansas, 1996) (Figure 1.8).

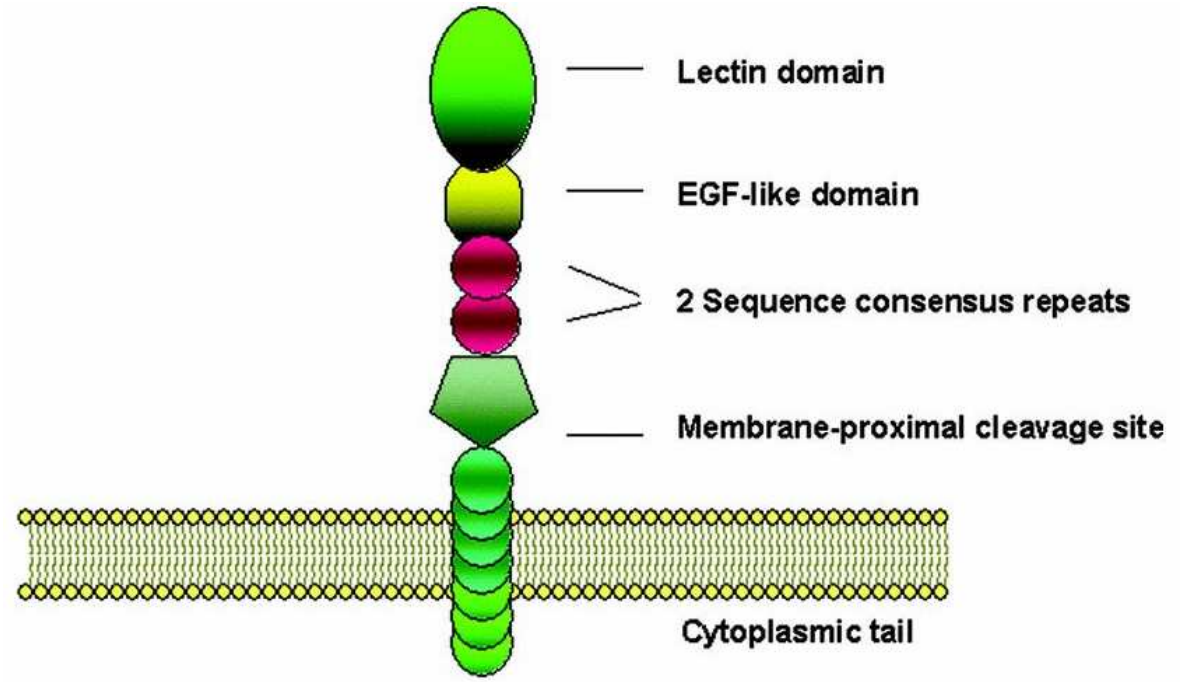


Figure 1.8. Structure of L-selectin. All the selectins have similar extracellular domains. L-selectin is the only selectin that carries a defined extracellular cleavage site that is close to the cell membrane. Leucocyte stimulation can often lead to rapid cleavage at this site. Figure from Ivetic *et al.* (2004).

Another group of mammalian Sia recognizing lectins are siglecs (Sia-binding Ig superfamily lectins) (Crocker and Varki, 2001). Siglecs share a common domain structure with an amino-terminal V-set Ig-like domain, variable numbers (1-16) of C2-set Ig-like domains, a single-pass transmembrane domain, and a cytosolic tail (Figure 1.9), (Angata *et al.*, 2004). An essential arginine (Arg97) on the F strand (conserved in the other siglecs) forms a salt bridge with the carboxylate of Sia and two tryptophans (Trp2 and Trp106) on the A and G strands form hydrophobic contacts with the *N*-acetyl and glycerol side groups of Neu5Ac respectively (Crocker *et al.*, 2001). Biological functions of the siglecs involve regulation of B-cell activation, siglec-2/ CD22 (O'Keefe *et al.*, 1996) and maintenance of myelin sheath structure, siglec-4a/MAG (Schachner *et al.*, 2000).

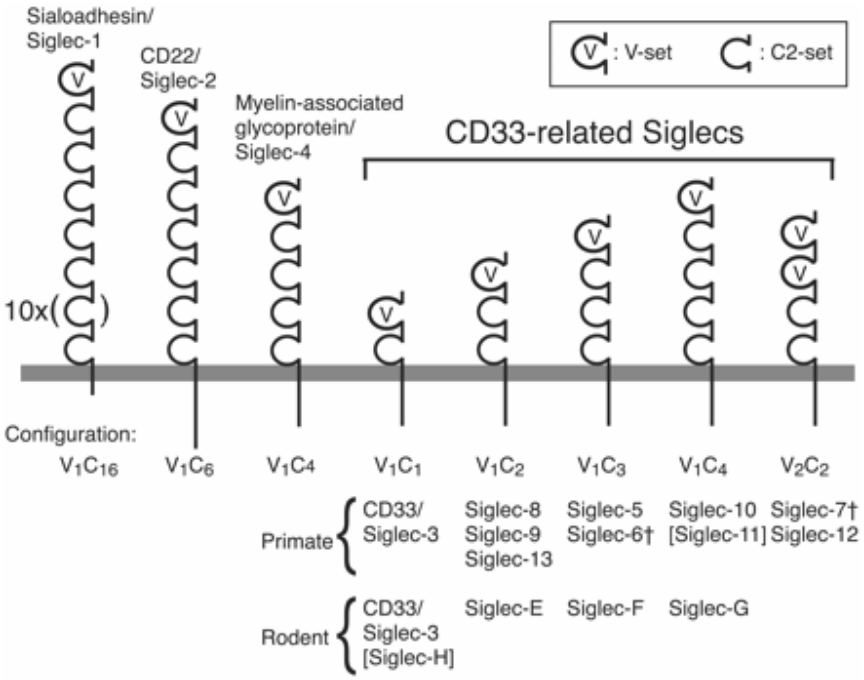


Figure 1.9. Schematic representation of Siglecs in primates and rodents. Siglecs have one V-set domain (a domain similar to Ig's variable region) and 1–16 C2-set domains (domains similar to Ig's constant region), followed by transmembrane and cytoplasmic domains. Genes for Sialoadhesin/Siglec-1, CD22/Siglec-2, and myelin-associated glycoprotein/Siglec-4 are located outside of the Siglec gene cluster in both primates and rodents. CD33rSiglecs are further classified into five subgroups (V_1C_1, V_1C_2, V_1C_3, V_1C_4, and V_2C_2), based on the number of V- and C2-set Ig-like domains. Figure from Angata *et al*. (2004).

1.2.5 Fungal lectins

Sia specific lectins have been isolated and characterized from the fruiting bodies of various mushroom species (reviewed in Lehmann *et al.*, 2006). Their natural function, in many cases, is not clearly understood. However, the identification and isolation of Sia-specific lectins from pathogenic fungi, particularly airborne species that cause severe infections in immunocompromized individuals, has raised the possibility that the initial stages of infection, particularly fungal spore (conidia) binding to the lung epithelial cells, may be mediated through Sia receptors. The first human pathogenic fungal species thought to possess a Sia-specific lectin were *Chrysosporium keratinophilum* and *Anixiopsis stercoraria* (synonym of *Aphanoascus fulvescens*) (Chabasse and Robert, 1986), which cause skin and nail infections in humans (Chabasse and Robert, 1986; Lehmann *et al.*, 2006).

Histoplasma capsulatum is the causative agent of histoplasmosis, a severe pulmonary infection that is most commonly found in tropical areas. Early studies showed that a 50-kDa cell wall protein from *H. capsulatum* yeast was able to bind laminin of extracellular matrix with high affinity, a process thought to be important in the initial stages of infection (McMahon *et al.*, 1995).

In developed countries, *Aspergillus fumigatus* is now regarded as the most important airborne fungal human pathogen. The inhaled conidia adhere and germinate in the lung. The preferential receptor for *A. fumigatus* lectin is Neu5Ac-α(2→6)-GalNAc (Lehmann *et al.*, 2006).

1.2.6 Plant lectins

Although only a few Sia specific plant lectins have been isolated, their historical importance in investigating the expression and biology of Sia is unquestioned. The function of Sia-specific lectin in these organisms is unclear as the presence of Sia in plants is questionable. The expression of Sia was reported (Shah *et al.*, 2003), however, this research was subsequently questioned (Seveno *et al.*, 2004). It was then concluded that it is not possible to demonstrate unequivocally that plants synthesize Sias because the amounts of these sugars detected in plant cells and tissues are so small that they may originate from extraneous contaminants (Zeleny *et al.*, 2006). According to the most popular theory the Sia-binding lectins are involved in defense mechanisms (Peumans *et al.*, 1995). The highly sialylated epithelial cells of animal digestive tracts provide numerous ligands for Sia specific lectins. Presumably, it is this binding of Sia-specific lectins from elderberry (*Sambucus nigra*) bark and wheat germ agglutinin that initiates the severe toxicity symptoms observed upon ingestion of plant lectins in higher organisms. The consequence of this is that elderberry, for example, is virtually never attacked in the wild. Moreover, Sia-specific plant lectins, like other plant lectins, are predominantly localized in regions of the plant that are most susceptible to attack, and thus require an adequate protection strategy (Peumans *et al.*, 1995; Lehmann *et al.*, 2006).

Sambucus nigra bark lectin (SNA) is a 140 kDa glycoprotein consisting of two different subunits 34.5 kDa and 37.5 kDa respectively, held together by intramolecular disulphide bridges (Broekaert *et al.*, 1984). SNA requires the presence of terminal Neu5Ac-Gal/GalNAc sequences for high affinity binding (Shibuya *et al.*, 1987)

Another plant lectin, widely regarded as Sia-specific, is wheat germ agglutinin (WGA). The 18 kDa WGA monomer is composed of four conserved 43 amino acid residue segments, which are folded into spatially distinct domains (A,B,C and D) of similar sequence. Monomers associate into dimers in a head-to-tail fashion, such that domains in contact across the dimer interface (A1-D2, B1-C2, C1-B2, D1-A2) are quasi-2-fold related. The domains are stabilized by four disulfide bonds and have an aromatic amino acid residue rich region for sugar binding (Wright *et al.*, 1993)

Although wheat germ lectin is known to bind Sia, this lectin binds more preferably to *N*-acetyl-*D*-glucosamine and its *p*-1,4-linked oligomers. The interaction of wheat germ

agglutinin with Sia is based on its structural similarity to N-acetyl-D-glucosamine, namely, the superimposable configuration of amino and hydroxyl groups at C-5 and C-4 of the pyranose ring of Sia with the C-2 and C-3 of *N*-acetyl-*D*-glucosamine (Shibuya *et al.*, 1987).

Maackia amurensis express two Sia-specific lectins, haemagglutinin (MAH) and leukoagglutinin (MAL) (Imberty *et al.*, 2000). The MAL lectin was shown to bind strongly to carbohydrate chains containing NeuAc-α(2→3)-Galβ-(1→4)GlcNAc/Glc structure whereas the MAH lectin has highest affinity and is directed toward the NeuAc-α(2→3) Gal-β(1→3)[αNeuAc(2→6)]αGalNAc tetrasaccharide (Knibbs *et al.*, 1991; Imberty *et al.*, 2000).

The 258 amino acid MAL lectin is composed of two asymmetric subunits named A and B. Each monomer folds into two large β-pleated sheets. Two monomers associate to form the so-called "canonical legume lectin dimer" characterised by a large 12-stranded β-sheet, resulting from the abutment of the two back sheets (Figure 1.10) (Imberty *et al.*, 2000).

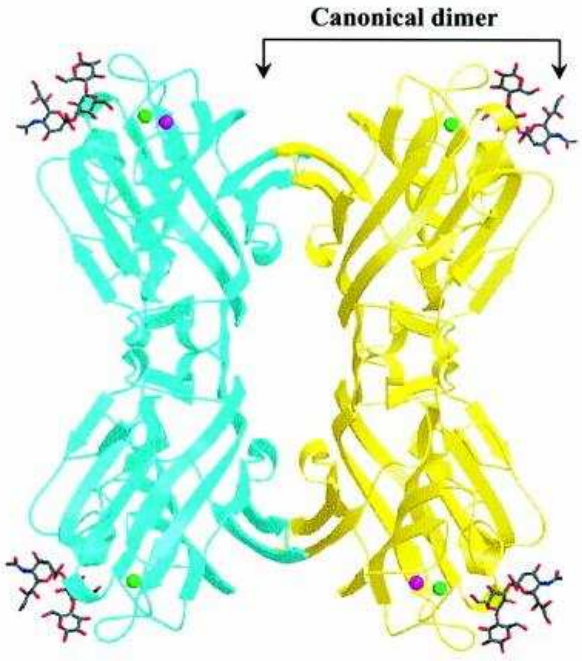


Figure 1.10. Crystal structure of the leukoagglutinin (MAL) from *Maackia amurensis* in complex with sialyllactose. The two independent monomers are shown with different colors. The Ca^{2+} and Mn^{2+} ions are displayed as *green* and *magenta spheres*, respectively. Figure form Imberty *et al.* (2000).

1.2.7 Current application of sialic acid specific lectins

In the light of the importance and versatility of Sia, the high specificity of Sia binding lectins has made them useful tools for a variety of applications. Initially the Sia specific lectins were used for detection of sialylated viruses, such as the plant rhabdoviruses potato yellow dwarf virus and eggplant mottled dwarf virus, propagated in plants of *Nicotiana rustica* (Slifkin and Doyle, 1990). They enabled development of rapid methods for pathogen identification in the field of medical microbiology (Wagner, 1982) as well as characterisation of sialylated viral envelopes (Slifkin and Doyle, 1990). Sia binding lectins were also applied in studying the specific Sia receptors on the surface of eukaryotic cells (Welty *et al.*, 2006) as well as sialylation of glycoproteins (Pinto da Silva *et al.*, 1981). In histology, lectins are used for detection of Sia receptors, e.g. on epithelial cells of the human airway (Nicholls *et al.*, 2007). Cancer research benefited from new cancer biomarkers' analysis methods based on lectins (Taylor *et al.*, 2009) and the biomedical industry in new purification and separation protocols based on Sia-specific lectins (Madera *et al.*, 2005).

1.2.8 Sia-specific proteins utilized in this project

Two Sia-specific lectins were chosen for the purpose of this project, namely haemagglutinin of influenza virus and SiaP of *Haemophilus influenzae*. They have different specificity for particular forms of Sia. One specifically binds glycans' terminal Sias and the second naturally targets free Sia. They come from different organisms and have different structures; however, both are relatively small proteins and uncomplicated molecules which should be advantageous in some aspects of the study, e.g. expression in a bacterial host or orientated immobilisation.

1.2.8.1 Haemagglutinin of Influenza Virus

Influenza virus, with thogovirus, belongs to the Orthomyxoviridae family. Three different types (A, B, and C) of influenza viruses are known and were classified on the basis of antigenic differences among their nucleocapsid and matrix proteins. Influenza A viruses were found to inhabit different animal species including humans, horses, swine, and a wide variety of domesticated and wild birds (Lee and Saif, 2009). Influenza type B viruses were isolated from humans and seals whereas influenza type C viruses have been isolated from humans and from swine (Ha et al., 2002). Influenza A viruses are further divided into subtypes based on the antigenic relationships of their haemagglutinin (HA) and neuraminidase (NA) surface glycoproteins (Lee and Saif, 2009). The HA and NA genes are extremely variable in sequence, and less than 30% of the amino acids are conserved among all the virus subtypes. A total number of 16 different HA subtypes (H1-H16) and 9 different NA subtypes (N1-N9) have been identified (Fouchier et al., 2005). Only three subtypes of HAs (H1, H2, H3) and two NAs (N1, N2) serotypes have adapted to humans, although some are very notable as the pandemic strains: H1N1 in 1918, H2N2 in 1957, H3N2 in 1968 (Stevens et al., 2007) and H5N1 in 2005 (Michaelis et al., 2009).

Ten viral proteins are encoded in eight segments of single stranded, negative sense RNA molecules, contained within the viral envelope in tight association with the nucleocapsid and the three polymerase proteins (PB2, PB1, PA), (Figure 1.11). Together they form the ribonucleoprotein (RNP), this complex being the minimal functional unit of replication (Nakajima, 1997). The matrix protein (M1), which lies inside the lipid envelope, is associated with both the RNP and the viral envelope. It is thought to play a fundamental role in virus assembly. The M2 protein decorates the surface of the envelope and acts as an ion channel, playing a role in triggering the viral uncoating in the endosome. The NS2 protein, also known as the nuclear export protein (NEP), mediates the export of newly synthesized RNPs from the nucleus (Lee and Saif, 2009). The NS1 protein is a regulatory protein with numerous functions that is not packed into the virion (Krug *et al.*, 2003). The most prominent feature of the virus envelope is tightly packed projections of two types of glycoproteins: rod-shaped trimers of HA and mushroom-shaped tetramers of NA. Virions can exhibit a variety of shapes and sizes ranging from spherical particles to elongated filamentous forms depending on the virus strain and passage history (Lee and Saif, 2009).

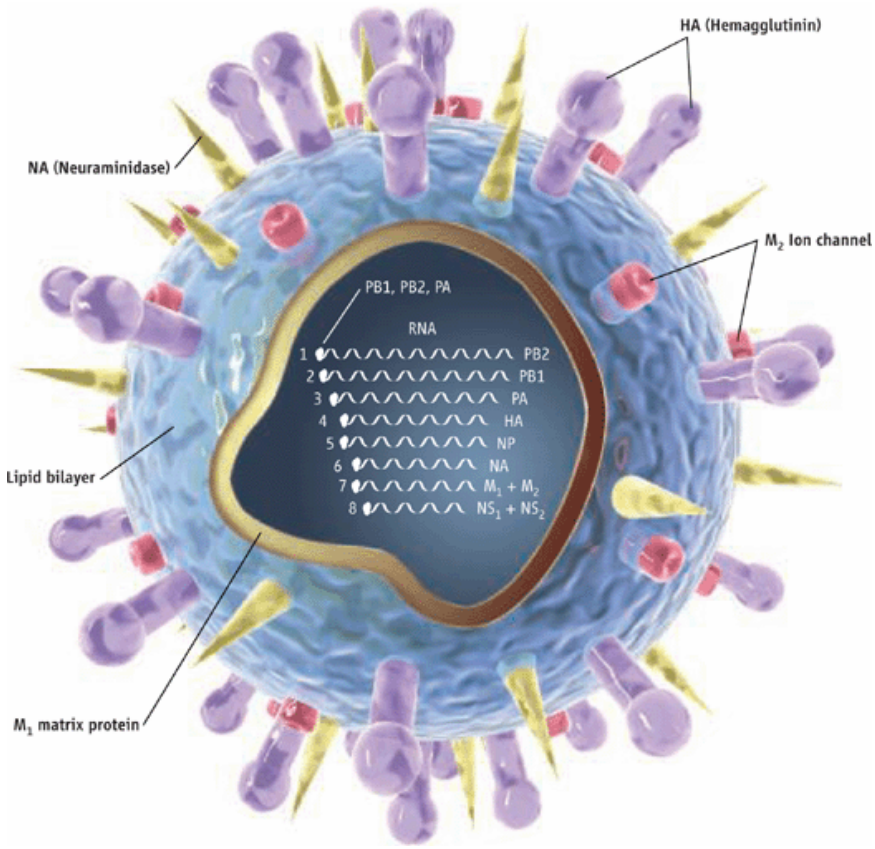


Figure 1.11. A schematic diagram of the structure of the influenza A virus. Figure from Kaiser (2006).

1.2.8.2 Structure of haemagglutinin

The haemagglutinin is a major component of the virus envelope and accounts for approximately 30% of the total virus protein (Wrigley, 1979). The trimeric, transmembrane glycoprotein has a molecular weight of about 210,000 Da. Each subunit comprises two domains, HA1 (58,000 Da) and HA2 (26,000 Da), linked together via two disulfide bonds (Skehel *et al.*, 1975; Doms *et al.*, 1986). The two subunits are formed during virus assembly at the plasma membrane of the infected cell as the result of proteolytic cleavage of a glycosylated precursor polypeptide HA0 (Hay, 1974). The precursor has a molecular mass of approximately 80,000 and is approximately 550 amino acids long with an *N*-terminal signal sequence, a transmembrane domain near the C-terminus and a short cytoplasmic tail (Skehel *et al.*, 1975; Stevens *et al.*, 2007).

The first X-ray structure of haemagglutinin from influenza virus was published by Wilson *et al*. (1981). Since that time over 25 structures have been reported for the HA, including three principal molecular conformations assumed by the protein during the virus life cycle: the precursor structure, the structure of the cleaved native HA present on the surface of infectious virions, and the structure of HA following the conformational changes involved in membrane fusion (Figure 1.12) (Stevens *et al.*, 2007).

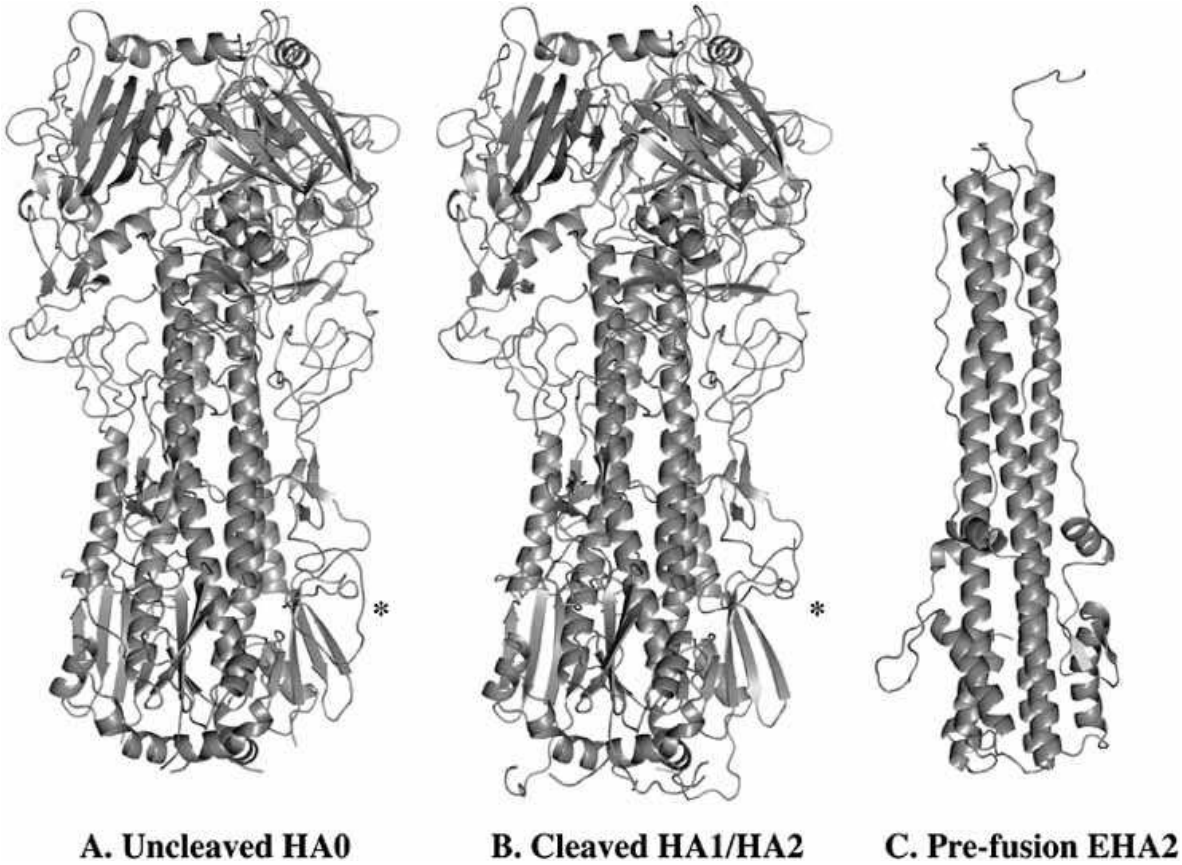


Figure 1.12. X-ray crystallographic models for the 3 principal molecular conformations assumed by the HA protein during the virus life cycle. The precursor structure (A), and the structure of the cleaved native HA (B) present on the surface of infectious virions are virtually identical to each other except for the HA1/HA2 cleavage site (marked by an *). After uptake of the virus through the endocytotic pathway, the HA undergoes a dramatic molecular rearrangement (C) that enable, the HA to expose the fusion peptide for subsequent membrane fusion. Figure from Stevens *et al.* (2007).

The mature HA homotrimer is ~135 Å in length with two distinct domains (Figure 1.13). The cylindrical protein is composed of a globular head, containing the receptor binding domain (HA1 residues 52 to 275 in H3 numbering), and an intertwined ‘stem’ domain at its membrane proximal base, composed of both HA1 residues (11 to 51 and 276 to 329) and

HA2 (1 to 176). From each monomer, long parallel α-helices of approximately 50 amino acids in length dominate this region and associate to form a triple stranded coiled-coil. Glycosylation is essential for the proper refolding and cellular processing of HA and can vary between strains with respect to the number and positioning of glycosylation sites. For example, the H3 haemagglutinin of the 1968 pandemic virus has seven glycosylation sites whereas the 1918 H1 HA has only four. Acquisition of glycosylation sites that change the antigenic properties of the protein can be directly related to escape from the host immune response (Stevens *et al.*, 2007).

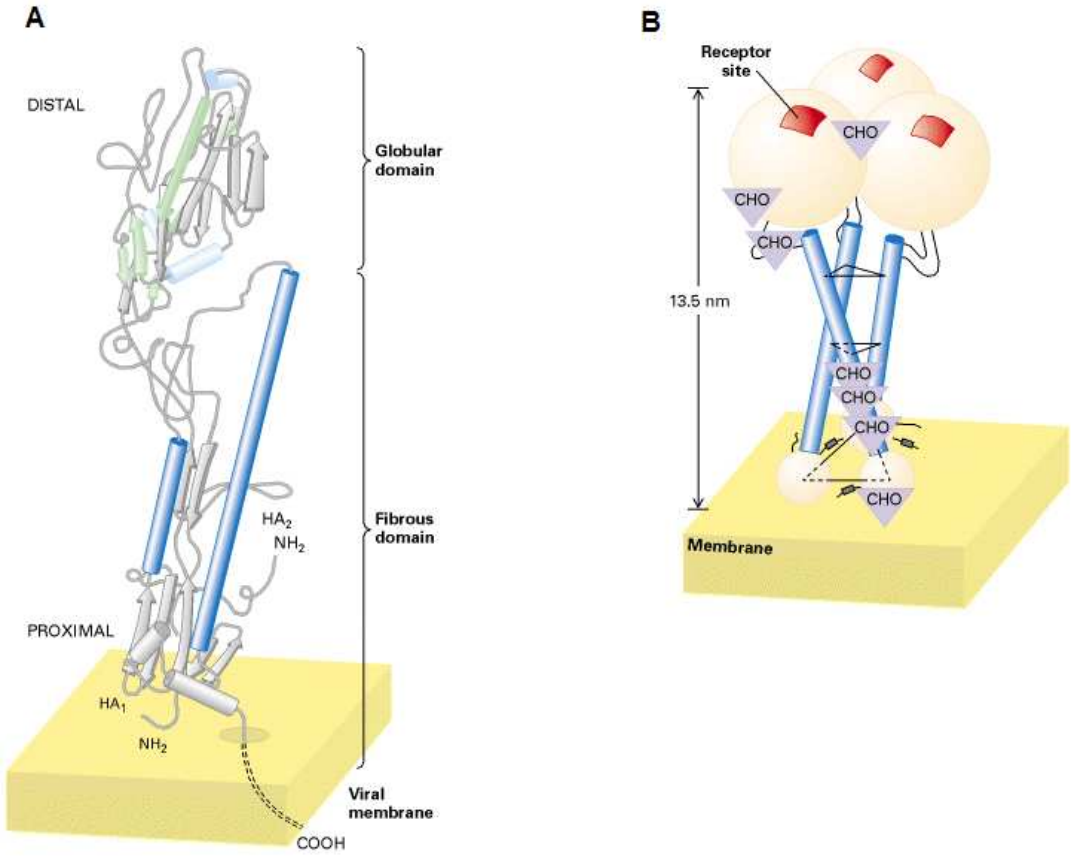


Figure 1.13. The structure of hemaglutinin. A - Globular and fibrous domain of a single subunit. B - Structure of the homotrimer. Figure from Lodish *et al*. (1999).

1.2.8.3 Priming of haemaggutinin

To gain full activity the HA protein requires a posttranslational alteration, consisting of HA0 cleavage, which takes place before the virion leaves the host cell. The cleavage results in the formation of the two disulfide-linked subunits HA1 and HA2, eliminating an arginine residue, R329, that separates them in the precursor. The cleavage takes place in the loop comprising residues 323–328 of HA1, and 1–12 of HA2. The loop projects the eight residues at the cleavage site, HA1 327 to HA2 5, away from the surface of the molecule (Figure 1.12). The newly created amino terminus (Kapatral *et al.*, 2002, 2003; Johnston *et al.*, 2007) of HA2 is a nonpolar sequence called the fusion peptide (Chen *et al.*, 1998). A negatively charged cavity is found adjacent to the loop, into which HA2 residues 1–10, the fusion peptide, inserts after HA0 cleavage, possibly guided by an electrostatic force generated when the positively charged amino terminus of HA2 is formed (Chen *et al.*, 1998).

1.2.8.4 Haemagglutinin's receptor-binding site (RBS) structure and function

One of the primary functions of HA protein, receptor binding, is mediated by the HA2 domain. A shallow depression at the top of the HA molecule binds terminal Sias of animal glycoproteins and glycolipids with millimolar affinity (Sauter *et al.*, 1992a). The receptor-binding site is composed of residues conserved in all subtypes of influenza (Skehel *et al.*, 2000; Ha *et al.*, 2001; Ha *et al.*, 2003) and comprises three structural elements, namely an -helix (190-helix, HA1 188 to 190) and two loops, 130-loop and 220-loop. The lower left and right edges of the site are formed by two surface loops; residues 225-228, the 220 loop and 135-138, the 130 loop, respectively (Figure 1.14) (Ha *et al.*, 2003). One side of the Sia's pyranose ring faces the base of the site, and the axial carboxylate, the acetamido nitrogen, and the 8- and 9-hydroxyl groups face into the site and form hydrogen bonds with conserved side-chain or main-chain polar atoms. Serine 136 forms a hydrogen bond with the carboxylate, which is also hydrogen bonded to the amide of the peptide bond 137; histidine 183 and glutamic acid 190 form hydrogen bonds with the 9-hydroxyl group, and tyrosine 98 forms hydrogen bonds with the 8-hydroxyl group. The 5-acetamido nitrogen forms a hydrogen bond with the carbonyl of peptide bond 135, and the methyl group of this

substituent is in van der Waals contact with the six-membered ring of tryptophan 153. The 7-hydroxyl group and acetamido carbonyl hydrogen bond to each other and the *N*-acetyl methyl group forms van der Waals contacts with leucine 194 (Skehel *et al.*, 2000).

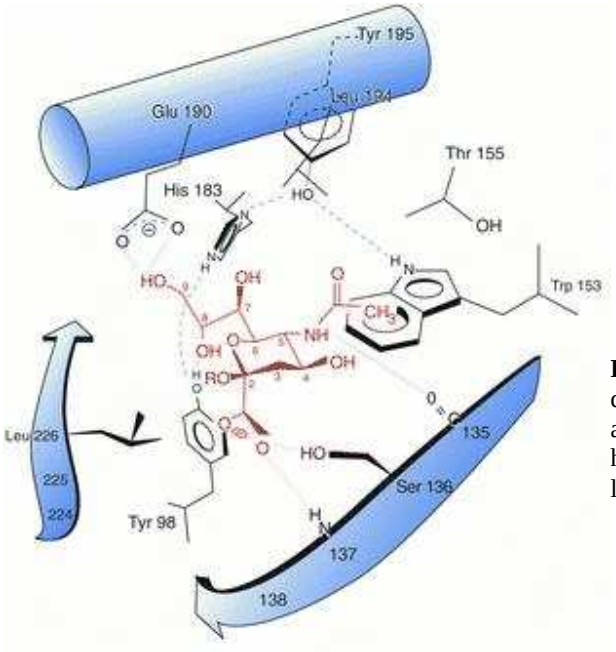


Figure 1.14. Haemagglutinin receptor binding. Schematic diagram of the receptor-binding site of a HA-receptor analog complex, showing the positions of the residues and hydrogen bonds that they form with sialic acid (dashed lines). Figure from Skehel *et al.* (2000).

Although the Sia receptor is bound in almost identical orientation in both avian and human H3 proteins there is a significant difference between these two binding sites at positions 226 and 228. The positions 226 and 228 are occupied by leucine (occasionally other non-polar residues) and serine in the human H3 protein or by glutamine and glycine in the avian H3 protein (Ha *et al.*, 2003). In human H3 the L226 side chain *D*-methyl group points downward into the binding site where it contacts conserved Y98 and A138, resulting in a slightly different geometry for adjacent residues. The 130 loop is shifted away from the 220 loop and thus the human H3 binding site is 0.5 Å wider than the avian H3 one. In the absence of this methyl group, the avian HA site can collapse inward, with the ring of Y98 rotating about 20° and the side chains of W153, H183, and the backbone of the 190 α-helix also changing slightly (Figure 1.15) (Ha *et al.*, 2003). The residue number 228 in human H3 is serine of which the OH-group contacts the Sia 9-OH whereas in avian H3 glycine hydrogen-bonds via a water molecule to the 8- and 9-OH of the Sia. This residue, along with residue 226, significantly influences the geometry of the binding site and, thus, the specificity of the protein (Ha *et al.*, 2001).

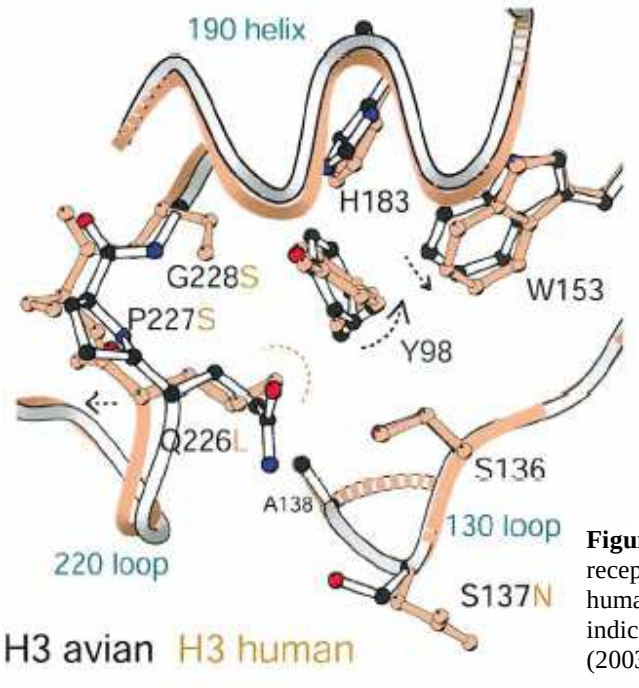


Figure 1.15. Conformational differences in the receptor-binding sites of the H3 avian and H3 human HAs. Side-chain and main-chain shifts are indicated by dashed arrows. Figure from Ha *et al.* (2003).

Interestingly, H2N2 viruses isolated from humans during the first year of the 1957 pandemic contain 228G (as do most avian viruses), whereas some avian H3 viruses contain "human" 228S (Matrosovich *et al.*, 2000). Similarly, in the H1 subtype haemagglutinin, isolated from humans, the avian specific Gln-226/Gly-228 pair was present (Ha *et al.*, 2003). It is possible that the mechanism directing the specificity of H1 subtype viruses is different than in H3 viruses. In H1 subtype the HA residues at positions 190 and 225 were found to dictate the specificity (Matrosovich *et al.*, 2000) while position 226 is occupied by glutamine in both human and avian viruses. The receptor-binding properties of such atypical avian and human viruses have not been characterized (Matrosovich *et al.*, 2000).

1.2.8.5 Specificity of the Receptor Binding Site

The sialylated gicans on the surface of human respiratory tract cells and avian intestinal cells have identical composition with regards to the three terminal monosaccharides: Sia, Gal, GlcAc. There is, however, a major difference between them with regards to the linkages between these monosaccharides. Human terminal glycans adopt a "folded" conformation imposed by the Siaα(2→6)Gal linkage and Galβ(1→4)GlcNAc linkage

which position the GlcNAc directly over the Sia. The folded conformation is mainly stabilized by hydrogen bonds between GlcNAc 3-OH and the ring oxygen of either the Sia or Gal. Alternatively, GlcNAc 3-OH forms a hydrogen bond with the Sia gliosidic oxygen. The avian glycans adopt an "extended" form imposed by the Siaα(2→3)Gal linkage and Galβ(1→3)GlcNAc linkage (Eisen *et al.*, 1997) (Figure 1.16 and Figure 1.17). The NeuAc group on the third saccharide appears to be crucial for stabilizing the folded conformation as α(2→6)-sialyllactose, identical to the first three saccharides of lactoseries pentasaccharide c (LSTc) but lacking the Neu5Ac group on the third Glc residue, and exerts an extended conformation (Skehel *et al.*, 2000). Accommodation of such different receptors by human and avian influenza viruses' haemaggluinins dictates the different structures of their binding sites. The receptor-binding site of the human HA is wider than avian to accommodate glycans in folded conformation.

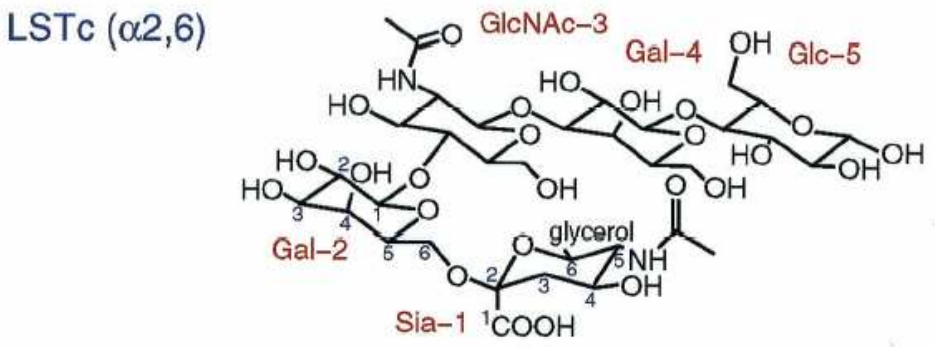


Figure 1.16. Lactoseries pentasaccharide c (LSTc). An oligosaccharide containing the three terminal monosaccharides present on the surface of human respiratory tract cells. Sequence of sugar residues and linkages are: NeuAcα (2-6)Galβ (1-4)GlcNAcβ (1-3)Galβ (1-4)Glc. Figure from Ha *et al.* (2001).

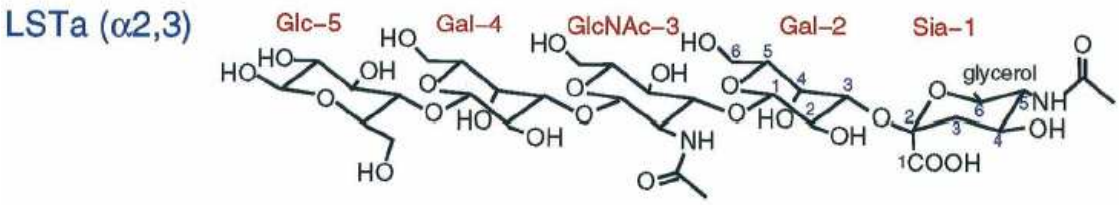


Figure 1.17. Lactoseries pentasaccharide a (LSTa). An oligosaccharide containing the three terminal monosaccharides present on the surface of human respiratory tract cells. Sequence of sugar residues and linkages are: NeuAcα (2-3)Galβ (1-3)GlcNAcβ (1-3)Galβ (1-4)Glc. Figure from Ha *et al.* (2001).

As the structures of the LSTc (human) and LSTa (avian) saccharides are different, they both exit the HA binding site in opposite directions. The human receptor exits the human HA binding site pointing outwards from the trimer threefold symmetry axis whereas the avian receptor exits the binding site near the interface between two HA monomers and the carbohydrate part of the HA glycoprotein (Figure 1.18) (Eisen *et al.*, 1997). This difference imposes the receptor linkage specificity, as for a given (avian or human) H3 one version of the saccharide is energetically preferred and thus the natural receptor for human influenza H3 virus is terminal Siaα(2→6)Gal which for avian H3 virus it is terminal Siaα(2→3)Gal (Connor *et al.*, 1994).

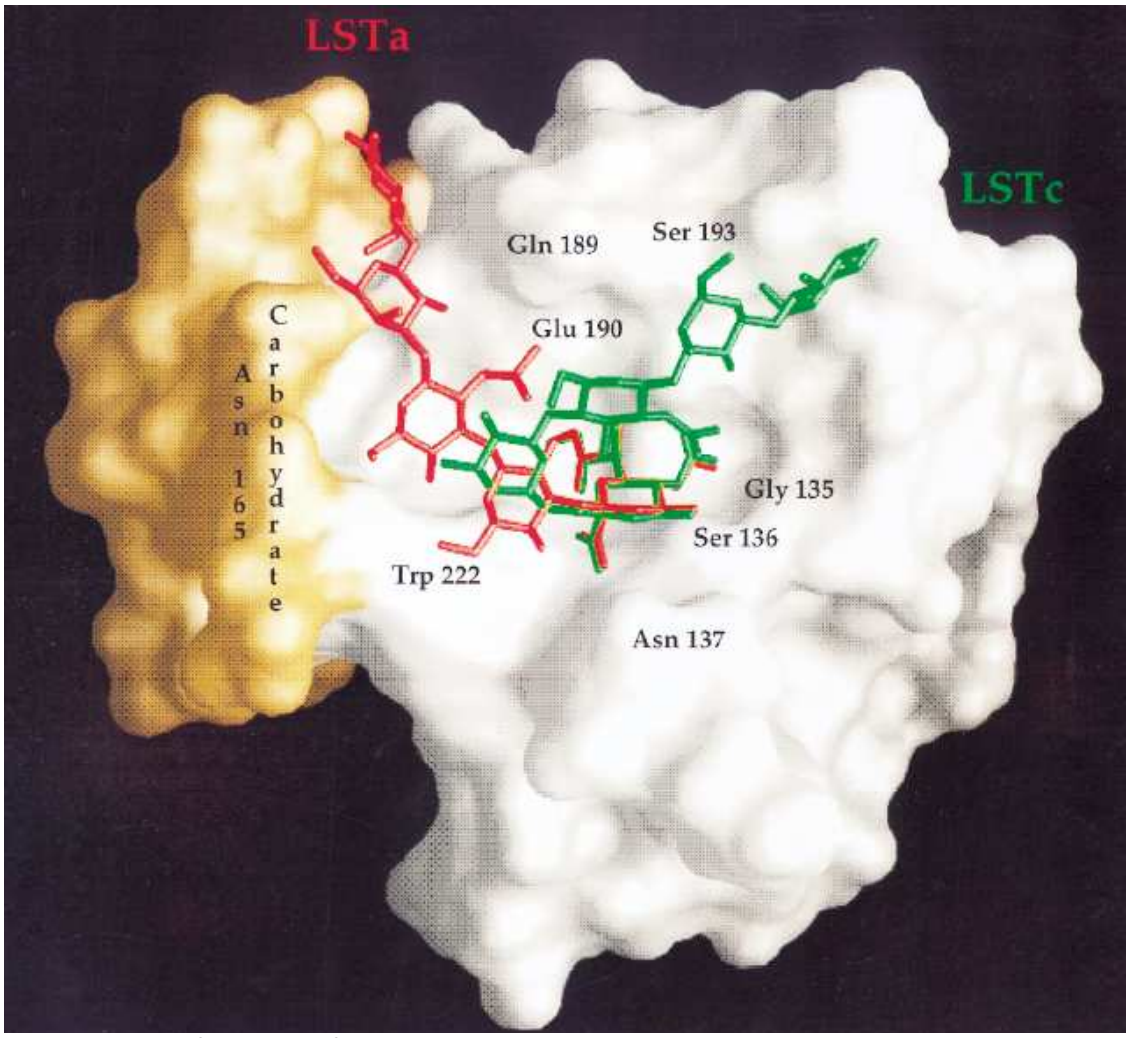


Figure 1.18. Conformations of LSTa (red) and LSTc (green) and HA monomers (white and gold). The molecular surface of HA viewed along the threefold molecular symmetry axis. Ligand models are from one binding site, least-squares fitted onto the other binding sites. Figure from Eisen *et al.* (1997).

1.2.8.6 Biological Significance of Receptor-Binding Specificity

The influenza viruses infect mainly respiratory tract cells of humans, horses and pigs as well as the intestinal epithelial cells of birds. The preference of influenza viruses for these cells is driven by the presence of specific receptors on the surface of the above cells, namely by an abundance of Siaα(2→6)Gal receptors on airway ephitelial cells in humans and Siaα(2→3)Gal receptors in avian intestines (Ha *et al.*, 2003). This explains why avian viruses are specific for birds and human viruses for humans. Furthermore the presence of Siaα(2→3)Gal rich mucins in the human lungs (Couceiro *et al.*, 1993) may impose additional selective pressure on the Siaα(2→6)Gal specific human influenza viruses and prevent binding of avian viruses. Similarly, Siaα(2→6)Gal rich equine fluids (Matrosovich *et al.*, 1998) inhibit the human viruses in birds. The nature of the Sia linkage on cells, the action of these soluble inhibitors, or both, appear to be involved in selecting HAs with different linkage recognition specificity.

Another factor dictating the influenza virus tissue tropism is the distribution of specific host proteases in the organism that are necessary for haemagglutin in activation. The protein is synthesized as a single-chain precursor (HA0) in the endoplasmic reticulum, where it is assembled as a trimer, glycosylated by the host glycosylation machinery in the Golgi apparatus and subsequently exported to the cell surface (Stevens *et al.*, 2006). The HA0 is then cleaved by specific host proteases into HA1 and HA2, required for membrane fusion activity and infectivity (Skehel *et al.*, 2000). The HA1 and HA2 polypeptide chains are usually separated by a single arginine residue, which is eliminated from the C terminus of HA1 by a virion-associated carboxypeptidase after cleavage (Garten *et al.*, 1983). The cleavage may be mediated by the serine protease, tryptase Clara, produced by Clara cells of the bronchiolar epithelium (Kido *et al.*, 1992). This enzyme shows recognition specificity for the sequence Q/E-X-R found at the cleavage sites of these HAs, which in conjunction with the narrow tissue distribution of the enzyme, restricts infection to the lung in mammals. In some HAs of the H5 and H7 subtypes there are polybasic sequences inserted at the cleavage site, separating the HA1 and HA2 polypeptide chains. In these cases, cleavage is intracellular and involves subtilisin-like enzymes that are active in the post-translational processing of hormone and growth factor precursors. The furin recognition sequence R-X-R/K-R is a common feature of the inserted polybasic sequences. The wide

tissue distribution of furin-like enzymes and the efficiency of intracellular compared with extracellular cleavage appear to be related to the widespread systemic and virulent infections caused by the H5 and H7 viruses in birds and the localized outbreak in Hong Kong, in 1997, of severe respiratory infection by an H5 virus in humans (Skehel *et al.*, 2000).

1.2.8.7 A Second Ligand Binding Site in HA

A second receptor binding site was identified between subunits of the HA trimer in an interface where two HA1 domains and an HA2 domain make close contact and are known to dissociate when the molecule changes conformation to affect membrane fusion (Sauter *et al.*, 1992b). A study of ligand binding to the second site suggested that a substituent that is larger than methyl is required at position 2 but that α(2→6)-sialyllactose does not fit. The affinity of α(2→3)-sialyllactose for the second site was estimated, by collecting X-ray diffraction data at various ligand concentrations, to be at least fourfold weaker than its affinity for the primary site. As the ligands from the second site can extend up toward the first site, it was suggested to construct a bridging ligand as an inhibitor of virus-cell binding that could take advantage of both sites. There is, however, no evidence that the second site is physiologically important (Sauter *et al.*, 1992b).

1.2.8.8 Membrane fusion as a second function of haemagglutinin

The second function of the HA protein after adhesion to the Sias containing receptors on the cell surface is the fusion of the membrane of the endocytosed virus particle with the endosomal membrane. This promotes the release of the viral nucleocapsids into the cytoplasm (Lee *et al.*, 2009; Stevens *et al.*, 2007). The membrane fusion and viral entry is mediated by the HA2 subunit. Comparison of the HA1 and HA2 sequences among influenza virus subtypes reveals that the HA2 sequence is well conserved, suggesting a highly conserved membrane fusion mechanism (Guo *et al.*, 2009).

Following interaction with a receptor at the cell surface the virus is rapidly internalized via endocytosis. Viruses are trafficked through the endocytic pathway and, ultimately, reach a low-pH compartment (approx pH 5.5). At this pH, the viral fusion machinery is triggered

and HA undergoes an irreversible conformational change (Stevens *et al.*, 2007). The change exposes the HA2 N-terminal fusion peptide, while preserving the structure of the HA1 receptor binding domain. The exposed fusion peptide inserts into the cellular membrane, anchoring HA into both prefusion membranes; by its N terminal in the endosomal membrane and by its C terminal in the virus membrane (Skehel *et al.*, 2000).

The refolding process involves three steps. Initially the fusion peptide, which at neutral pH is buried in the trimer interface only 30 Å from the virus membrane, is extruded at fusion pH to the N-terminal tip of the 100 Å trimeric coiled coil that is newly formed at this low pH; its formation involves relocation of a short α helix and refolding of a segment of the extended chain into an α helix (Bullough *et al.*, 1994). Then the middle of the long α helix of native HA2 unfolds to form a reverse turn, and the second half of the long α helix jackknifes back to lie antiparallel against the first half, also relocating C-terminal residues by >100 Å in relation to coiled-coil residues 76–105, which form a common structure in both conformations. Subsequently, residues 141–175, located C terminal to a small β sheet hairpin that accompanied the jackknifed α helix, appeared to be extruded from their compact association in HA2 to become a mostly extended structure that reached back up the central coiled-coil, packing antiparallel within the groove between adjacent α helices. The overall effect of this refolding is apparently to deliver the fusion peptide toward the target membrane and to bend the molecule in half so that the fusion peptide and the viral membrane anchor are near the same end of the rod-shaped molecule (Figure 1.19) (Skehel *et al.*, 2000).

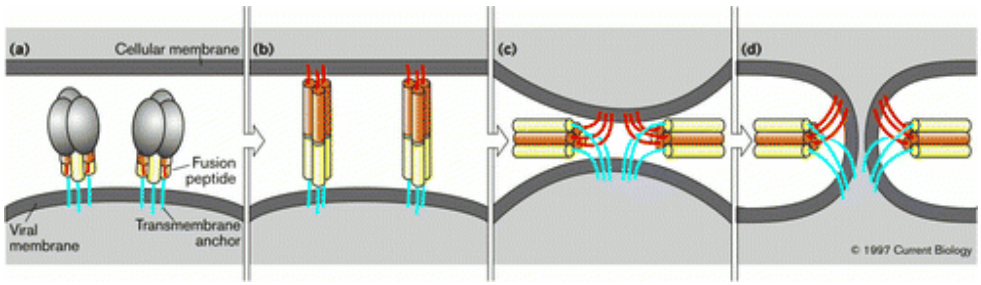


Figure 1.19. Hypothetical mechanism for membrane fusion by virus glycoproteins. (*a*) The entire HA molecule. (*b–d*) The HA2 subnunits alone, in which the termini of the proteins come into relative proximity, thereby inducing membrane fusion. Figure from Skehel *et al.*, (2000).

1.2.8.9 Haemagglutinin targeting influenza treatment approaches

Currently, a number of steps in the influenza virus replication cycle have been identified for therapeutic intervention and drug design such as viral uncoating that can be inhibited by amantadine or the release of newly formed virus particles that can be blocked by neuraminidase inhibitors (De Clercq, 2006). Their effect is to reduce the yield of viable virus progeny, thus reducing morbidity, allowing time for the host immune system to mount an effective response and decreasing community transmission.

Drugs oriented at haemaggutinin inhibition can truly block the entrance of the pathogen into the host cell and thus efficiently prevent the infection and/or spread of the virus. Monomeric Sias or methyl sialosides have been proposed as potential treatments (reviewed in Carlescu *et al.*, 2008). However, due to rapid clearance by enzymatic breakdown and low binding affinity, high concentrations are required for effective inhibition and this results in unwanted toxicity issues. The main obstacle from a drug development perspective associated with all receptor-haemagglutinin interactions is their relatively low binding affinity. An effective HA antiviral must have improved binding compared to the natural receptor, and ideally be independent of Sia-linkage. The most potent HA inhibitors reported are based on polyvalent constructs (reviewed in Carlescu *et al.*, 2008). Use of such polymers with multivalent ligands overcomes the problem of low affinity. The simplest form reported for influenza has been polymers such as polyglutamic acid with Sia bound through neuraminidase-resistant linkages (Stevens *et al.*, 2007). Such compounds by mimicking the natural sialylated animal cell receptors bind to both HA and NA on the virus surface. Tests of Sia-polyacrylamide conjugates, also termed sialylglycopolymers in mouse models resulted in 100% survival rate with lethal doses of mouse-adapted, different influenza A strains (H3N2, H1N1) (De Clercq, 2006). This suggests such compounds as potentially valuable broad spectrum reagents.

Another approach focused on the blocking of viral attachment has recently been studied. In the system investigated by Malakhov *et al.*, (2006) the haemagglutinin binding was inhibited by the removal of the receptor from the susceptible cells. A recombinant fusion protein composed of the sialidase (neuraminidase) catalytic domain derived from *Actinomyces viscosus* fused with a cell-surface-anchoring sequence was used to remove the influenza-virus receptors from the airway epithelium. A sialidase fusion construct, DAS181

(Fludase), was shown to effectively cleave the receptors used by both human and avian influenza viruses. Significant *in vivo* efficacy of the sialidase fusion construct was noted in both prophylactic and therapeutic approaches (Malakhov *et al.*, 2006).

It was recently demonstrated that a 20-amino-acid peptide (EB, for entry blocker) derived from the signal sequence of fibroblast growth factor 4 exhibits broad-spectrum antiviral activity against influenza viruses including the H5N1 subtype in vitro. The EB peptide was protective *in vivo*, even when administered postinfection. Mechanistically, the EB peptide inhibits the attachment to the cellular receptor, preventing infection (Jones *et al.*, 2006). Further studies demonstrated that the EB peptide specifically binds to the viral haemagglutinin protein. The mechanism of action of this peptide needs further investigation which may confer benefit by a completely new approach to antiviral therapy.

The entry of influenza virus to the host cell, driven by haemagglutinin molecules on the surface of the virion, is dependent on the conformational change of the HA. An inhibitor of viral growth, (Deshpande *et al.*, 2001) BMY-27709 (4-amino-5-chloro-2-hydroxy-N-9AαH-octahydro-6β-methyl-2H-quinolizin-2α-benzamide), was found to prevent the low pH-induced conformational rearrangement of HA into its fusogenic state (Luo *et al.*, 1996; Luo *et al.*, 1997; Deshpande *et al.*, 2001). This was found to be active against all H1 and H2 subtype viruses; it is however, inefficient as of yet against H3 subtype viruses (Luo *et al.*, 1997).

1.2.8.10 Summary of the haemagglutinin from influenza virus

Haemagglutin's HA1 domain containing the receptor binding site is relatively small and thus is useful for large scale production and application on various analytical platforms. The specificity of the HA towards either α(2→3) and α(2→6) linked Sia can be altered by changing only three amino acids (in H3 subtype) while the overall structure remains the same. This would be clearly advantageous for production of proteins with different specificities where the expression, purification and other procedures are the same. The specificity of such proteins towards terminal Sia may be explored for development of sialylated glycan analysis method.

1.2.8.11 SiaP protein of *Haemophilus influenzae*

The bacterial pathogen, *Haemophilus influenzae,* is exclusively adapted to infect or colonize humans. Strains can be encapsulated or nonencapsulated (nontypeable). Nontypeable *Haemophilus influenzae* (NTHi) is a frequent colonizer of the nasopharynx. When the airway is compromised, NTHi can cause local infections such as otitis media (Leibovitz *et al.*, 2003), chronic bronchitis and pneumonia (Murphy, 2003).

The bacterium evades the innate immune response of the host by sialylation of its lipooligosaccharides (LOS) (Mandrell *et al.*, 1992) and thus appears as self for the host's defence system. The acceptors for Sia are LOS terminal lactose, N-acetyllactosamine, and possibly N-acetylgalactosamine (Mansson *et al.*, 2001; Jones *et al.*, 2002). It was also found that Sia can be utilized by *H. influenzae* as a carbon and nitrogen source (Vimr *et al.*, 2000).

Interestingly, although the Sia is crucial for the pathogen survival inside the host, the NTHi is incapable of synthesizing Sia, and thus requires an exogenous source of Sia (Bouchet *et al.*, 2003). It was found that Sia acquisition is mediated via a tripartite ATP-independent periplasmic (TRAP) transporter driven by an electrochemical gradient (Muller *et al.*, 2006). TRAP transporters consist of three components: an extracytoplasmic solute receptor (ESR) and two distinct integral membrane components of unequal size which are sometimes fused (Kelly *et al.*, 2001). In NTHi the two integral membrane components are fused and known as SiaT or SiaQM and the extracytoplasmic solute receptor is SiaP (Johnston *et al.*, 2007) (Figure 1.20). The ESR protein (often also known as a periplasmic-binding protein) increases the uptake affinity of the transport system by binding Sia and delivering it to the inner membrane transporter (Allen *et al.*, 2005). Such a high-affinity Sia uptake system may be important for the bacteria in their normal physiological environment.

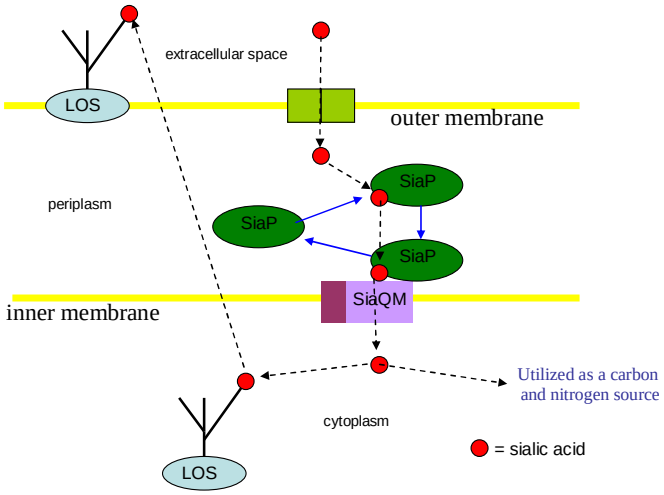


Figure 1.20. Overview of the Sia utilization in *Haemophilus influenzae*. Sia gets into the periplasm via NanC pores (Severi *et al.*, 2007) and is captured by SiaP which delivers it to the inner membrane SiaQM transporter. Sia is then transported into the cytoplasm and utilized either as a carbon and nitrogen source or for LOS sialylation. Sialylated LOS is later transported to the outer membrane. Dashed arrows indicate the path of Sia in the bacterium and solid blue lines indicate the cycle of the SiaP protein.

1.2.8.12 Structure of the SiaP

Structurally SiaP is a monomeric protein with a molecular weight of about 37 kDa. (Johnston *et al.*, 2007). The 329 amino acid polypeptide secreted to the periplasmic space and during translocation through the inner membrane a 23 amino acid leader peptide is cleaved off (Johnston *et al.*, 2007). SiaP has two α/β domains connected by three segments of the polypeptide and separated by a large cleft (Figure 1.21) (Muller *et al.*, 2006). Domain I, encompassing residues 1-124 and 213-252, contains a 5-stranded β-sheet against which are packed six α-helices. The strand order is β2-β1-β3-β10-β4, with strand β10 running anti-parallel to the other four strands (Figure 1.21 and Figure 1.22) (Muller *et al.*, 2006). Domain II contains residues 125-212 and 253-306 and has a 6-stranded β-sheet surrounded by 3 α-helices (Figure 1.21 and Figure 1.22) (Muller *et al.*, 2006). Here the sheet topology is β7-β6-β8-β5-β9-β11 with strand β5 running anti-parallel to the other strands. Residues 280-306 at the C terminus of the molecule form a pair of α-helices that fold across the base of the molecule and pack against both domains. A striking feature of the structure is the long helix, α9, which spans the breadth of the molecule (Muller *et al.*, 2006).

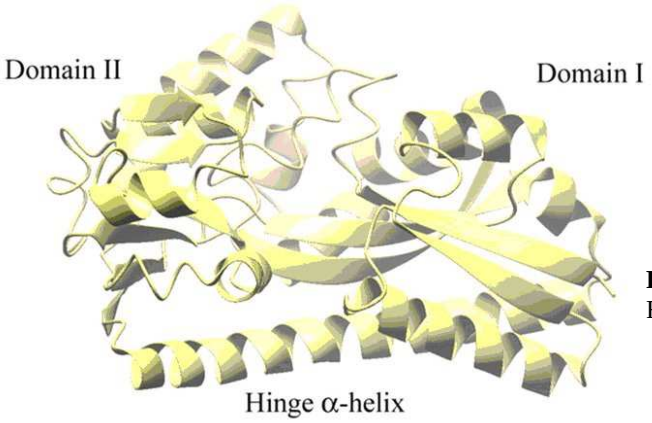


Figure 1.21. Overall structure of the SiaP. Figure from Muller *et al.* (2006).

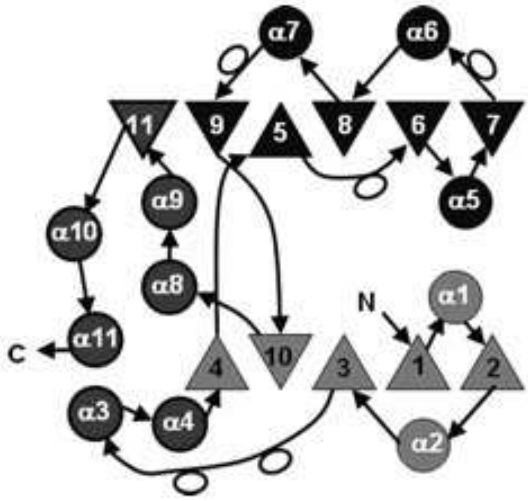


Figure 1.22. Schematic diagram of the topology of SiaP. The N-terminal domain is in light gray, and the C-terminal domain is in black. Features that distinguish SiaP are in dark gray. Figure from Muller *et al.* (2006).

1.2.8.13 SiaP ligand bindig site

The binding of Neu5Ac, a natural ligand, by the SiaP protein is initiated through direct interactions between the highly charged carboxylate at the C1 position of the sugar with positively charged residues R127 and R147 (Figure 1.23) (Johnston *et al.*, 2007). The primary role of R127 is electrostatic stabilization of the charge on the sugar. R127 δ-guanido interacts with both the C1 carboxyl and the C2 hydroxyl of the sugar (forming a salt bridge). E186 interacts with R127 and appears to optimize its position. The guanidino group of R147 seems to be key to the binding interaction, both in terms of its charge interaction with the C1-carboxyl group as well as in the orientation of the ligand (Johnston *et al.*, 2007). The orientation is also governed by an interaction between the glycol group hydroxyls and the carboxyl of E67. This interaction also induces the ligand specificity of

the binding site. N187 makes direct interactions with Neu5Ac at the C1 carboxyl and the C2 hydroxyl. D49 interacts with O7 and R70 probably stabilizing the position of the former. N10 makes two interactions with the acetyl oxygen, from the side chain and a water mediated interaction by the main chain amide. The N10 interactions are not crucial for the binding. Additionally Q214, interacts with the ligand through water molecules (Johnston *et al.*, 2007).

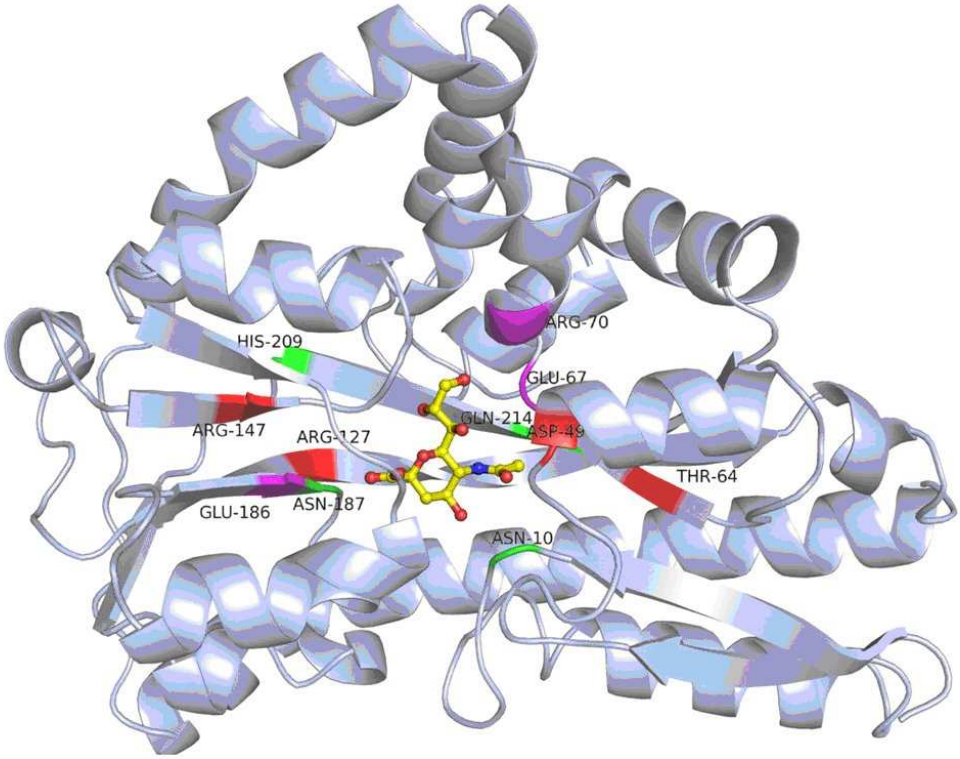


Figure 1.23. Structure of SiaP and the binding site. A ribbon diagram showing SiaP with Neu5Ac bound. The residues marked in red show no complementation; the ones in maroon show partial complementation, and those in green show complete complementation. Neu5Ac is shown as a ball-and-stick model. Figure from Johnston *et al*. (2007).

Quantitative comparison of the open and closed structures reveals a rotation of 28.7-34.7° about a hinge that runs close to the peptide bonds connecting residues Thr-127-Arg-128, Ile-211—Leu-212, and Glu-254—Lys-255 (Muller *et al.*, 2006; Johnston *et al.*, 2007). During the rotation about the hinge region the SiaP protein changes its conformation from "open" to "closed" (Figure 1.24).

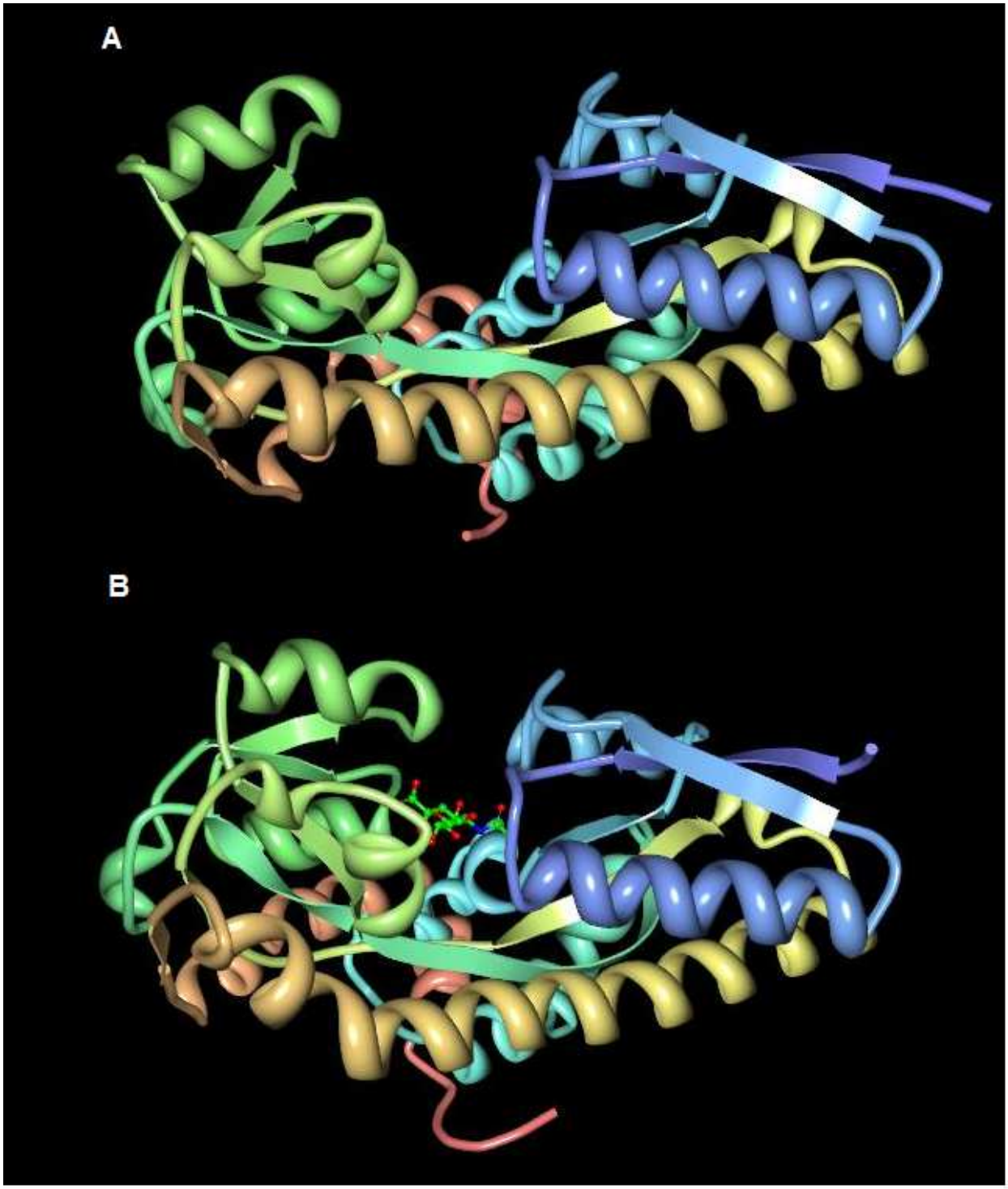


Figure 1.24. SiaP protein in an "open" (A) form before Neu5Ac accommodation and a "closed" (B) conformation after Neu5Ac accommodation. Image created using Protein Workshop software with PDB file 3b50 (Johnston *et al.* 2007).

The ligand is almost completely buried, with only 32 $Å^2$ of its 435-$Å^2$ surface area accessible to the solvent (Figure 1.25) (Muller *et al.*, 2006), which probably influences the affinity. The affinity of SiaP for Neu5Ac was determined using isothermal titration calorimetry (ITC). The K_d of 28 nM was reported, which indicates a high affinity interaction between SiaP and Neu5Ac (Johnston *et al.*, 2007).

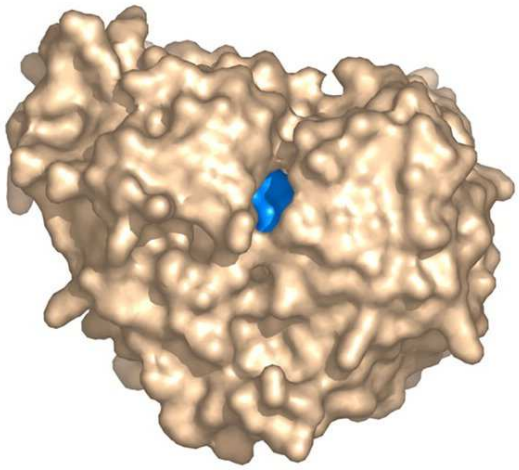


Figure 1.25. Surface representation of the Neu5Ac2en (close homolog of Neu5Ac) structure viewed looking down into the binding cleft. The ligand is in blue, and domain I is on the left, whereas domain II is on the right. Figure from Muller *et al.* (2006).

The carboxylate and N-acetyl groups of Neu5Ac are essential for high affinity binding to SiaP. A homologous ligand lacking the carboxylate group binds to SiaP with 2000-fold lower affinity then Neu5Ac (Muller *et al.*, 2006) which proves the importance of it. However, 4-acetylaminocyclohexane carboxylic acid, containing both a carboxylate and an N-acetyl group in analogous positions to the natural ligand Neu5Ac and also adopting a chair conformation similar to Neu5Ac, does not bind to SiaP (Muller *et al.*, 2006). This suggests that the carboxylate and N-acetyl groups are essential, but not sufficient, for the high affinity binding of Neu5Ac to SiaP.

1.2.8.14 SiaP homologues

The residues crucial for carboxylate coordination in SiaP, namely R127, R147 and N187 are present in homologous sequences found in operons with genes for sialometabolism (Muller *et al.*, 2006) (Figure 1.26). High conservation is also detected among residues D49 and E67, responsible for glycerol moiety binding. The *N*-acetyl binding N10 residue is not conserved, being replaced by glutamine, valine, or threonine in the homologs of SiaP.

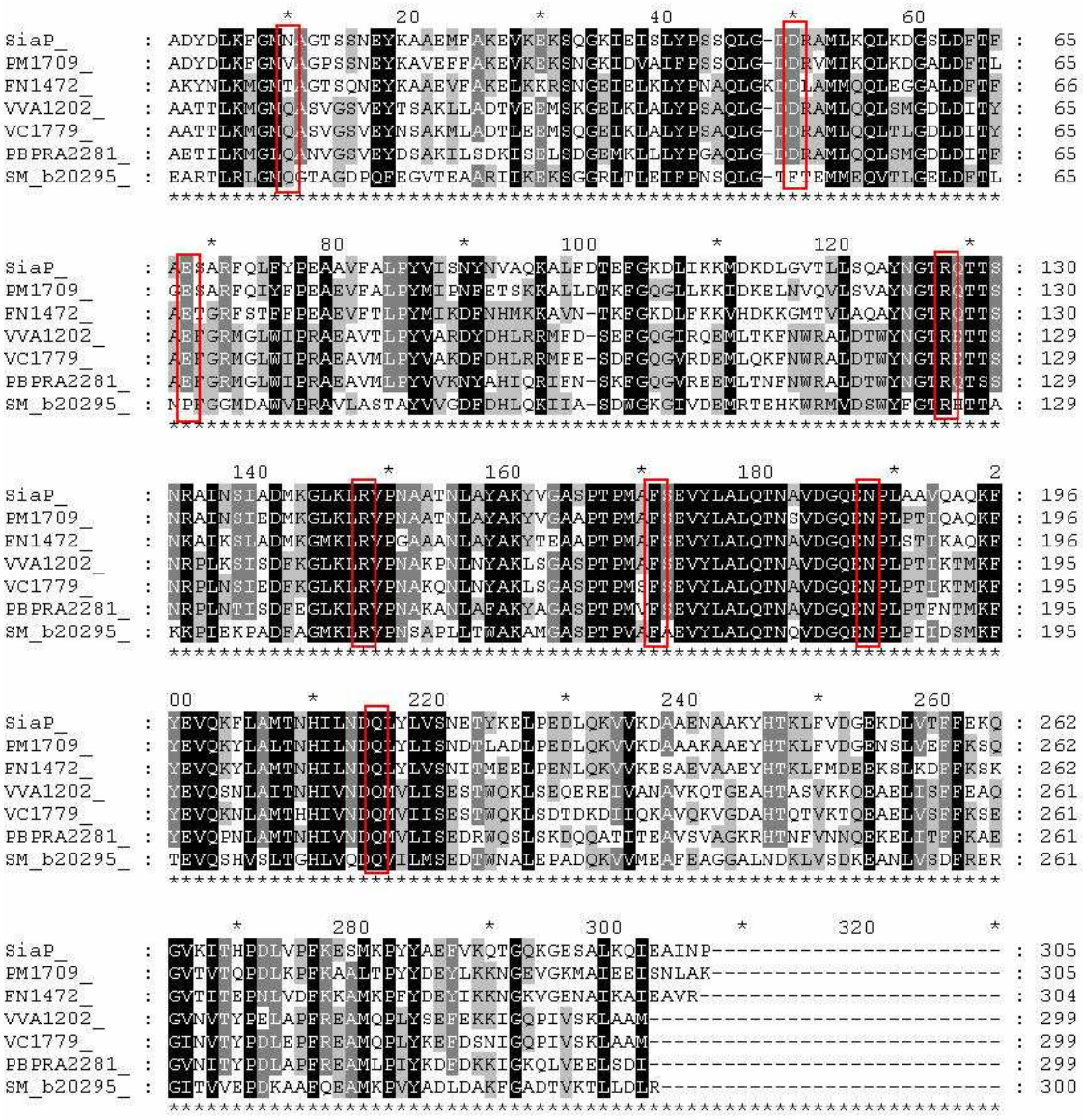


Figure 1.26. Multiple sequence alignment of amino acid sequences of *H. influenzae* SiaP and 6 related TRAP ESR proteins that are likely to form components of a Sia transporter. The genes for all seven of these sequences are encoded in operons encoding sialometabolic genes.

In most cases the Sia specificity is unconfirmed; however, initial studies on the SiaP homologue from *V. cholerae* indicates that it binds Neu5Ac with an affinity of about 300 nM (Johnston *et al.*, 2007). Furthermore, a TRAP ESR from *F. nucleatum* subsp. *vincentii* shares high homology with SiaP (Kapatral *et al.*, 2002, 2003; Johnston *et al.*, 2007). The Sia binding by the protein has not been characterized yet. However, the fact that the organism appears to lack a homolog of Neu5Ac synthase, yet incorporates Sia into its

lipopolysaccharide (LPS) (Kapatral *et al.*, 2002, 2003; Johnston *et al.*, 2007), indicates that it has a Sia acquisition system most probably homologous to the one in *H. influenzae*.

Not homologous in terms of overall amino acid sequence, although bearing a similar feature of the ligand binding site, are some other Sia-specific proteins. A conserved arginine is a common theme among various proteins of diverse structures and biological functions that bind Sia. For example, enzymes that cleave at the terminal Sia of the glycans, neuraminidases, contain a characteristic arginine triad that coordinates the carboxylate group (Crennell *et al.*, 1993; Taylor, 1996; Muller *et al.*, 2006). SiaP is similar in using a triad of residues to coordinate the carboxylate group but achieves this using two arginine residues and one asparagine, a conserved structural motif that appears to be important for high affinity binding (Muller *et al.*, 2006).

1.2.8.15 Summary of the SiaP protein

The relatively small size of the SiaP protein would be advantageous for its production and application for Sia detection. Free Sia specificity would enable development of analytical methods for free Sia analysis. The SiaP protein C-terminus is located on exactly the opposite side of the protein than the sugar binding site. This would be beneficial for applications requiring utilization of the protein in immobilised form, as a desired protein orientation would be achieved by specific C-terminal tagging.

1.3 Aims

The aim of this study was to develop a novel method for Sia detection and quantification based on Sia-binding proteins. A method, that would be specific, precise, sensitive, reliable, quick, uncomplicated and also inexpensive. As a cost effective way of the protein production prokaryotic expression system was selected. Additionally, the goals involved optimization of protein expression and purification, the methodology development for Sia-free protein production and orientated immobilisation of the protein on a surface of the analytical device. The produced Sia-specific protein was envisaged to be characterise using various analytical platforms including ELLA, MS and surface plasmon resonance, and subsequently applied for Sia detection and quantification using these platforms.

Chapter 2
Materials and Methods

2 Materials and Methods

2.1 Materials

2.1.1 Bacterial Strains, Primer Sequences and Plasmids

Bacterial strains, primer sequences and plasmids used in this thesis are presented in Table 2.1, Table 2.2 and Table 2.3 respectively. The *ha1* gene from pI.17 plasmid was cloned into the pCR2.1 derived pPCRHA1 plasmid and then subcloned into pJS102 expression vector (pLecB3 derived). In order to promote soluble production of the HA1 protein the *ha1* gene was subsequently subcloned into the pJS103 (pQE50-PelB derived) and the pJS104 (pMALp2E derived) plasmids. For the SiaP protein production the *siaP* gene was obtained on the pGTY2 plasmid and then the plasmid was modified by adding a 6xHis tag, creating the pJS201 plasmid. The pJS201 plasmid was subsequently modified by adding a C-terminal 6xLys tag (pJS202 and pJS2021), Strep2 tag (pJS203 and pJS2031) and a $(Gly\ Gly\ Gly\ Ser)_2$ linker (pJS2011, pJS2021 and pJS2031).

Table 2.1. *E. coli* strains used in this thesis.

Strain	Genotype	Features	Source
BL21	F′ *dcm ompT hsdS*(r_B- m_B-) *gal*	Protease deficient	Stratagene
JM109	F′ (traD36, *pro*AB+ *lac*Iq, D(*lac*Z)M15*) end*A1 *rec*A1 *hsd*R17(rk-, mk+) *mcr*A *sup*E44 l- *gyr*A96 *rel*A1 DΔ*(lacproAB) thi*-1	General cloning strain	Sigma
KRX	[F´, *tra*D36, Δ*omp*P, *pro*A+B+, *lacI*^{q}, Δ(*lac*Z)M15] Δ*omp*T, *end*A1, *rec*A1, *gyr*A96 (Nal^r), *thi*-1, *hsd*R17 (r_K–, m_K+), e14– (McrA–), *rel*A1, *sup*E44, Δ(*lac-pro*AB), Δ(*rha*BAD)::T7 RNA polymerase	Cloning and expression strain. Provides tight control over expression.	Promega
Rossetta	F′ *ompT hsdS*$_B$*(*R_B^- m_B^-*) gal dcm* pRARE2	Expression strain for eukaryotic proteins. Contains pRARE plasmid with genes for rare *E. coli* tRNAs. Chloramphenicol resistant.	Novagen
XL-10 Gold	Tet^r Δ*(mcrA)183* Δ*(mcrCB-hsdSMR-mrr)173 endA1 supE44 thi-1 recA1 gyrA96 relA1 lac* Hte [F′ *proAB lacI*^{q}*Z* Δ*M15 Tn10* (Tet^r) Amy Cam^r]	General cloning and expression strain.	Stratagene

Table 2.2. Primers used in this thesis.

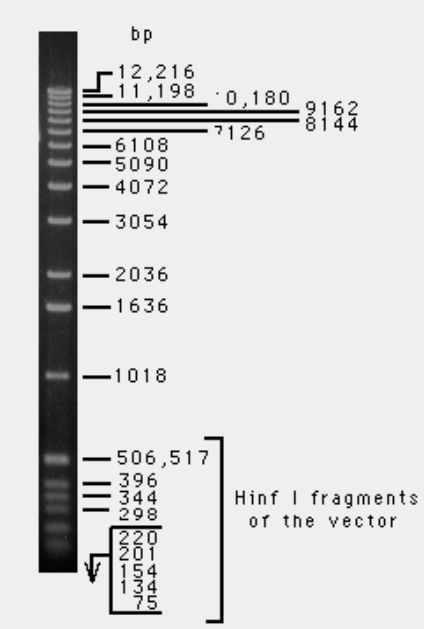
Name	Sequence (5′ – 3′)	Tm (C°)
HA1F1	CCATGGCAATGCAAAAACTTCCCGGAAATGACAACAGC	67
HA1R1	AGATCTAGTTTGTTTCTCTGGTACATTCCGC	60
1HA1NotI	ATCGGCGGCCGCCATGCAAAAACTTCCCGGAAATGACAACAGC	72
R1HA1BglII	GAAGATCTAGTTTGTTTCTCTGGTACATTCCGC	62
HA1KpnIF1	GACAGGTACCGATGCAAAAACTTCCCGGAAATGACAACAGC	69
HA1BamHIR	ACTGGATCCTTAAGTTTGTTTCTCTGGTACATTCCGC	64
F1SiaPNdeI	GGGAATTCCATATGATGAAATTGACAAAACTTTTCC	60
R1SiaPXhoI	CCGCTCGAGTGGATTGATTGCTTCAATTTGTTTTAAAGC	64
FSiaLYS	CTCCTCCTCTGAGATCCGGCTGCTAACAAAGCCC	69
RSiaPLYS	GAGGAGGAGGTGGTGGTGGTGGTGGTGCTCGAG	72
FSiaPCS	CCACAATTCGAGAAATGAGATCCGGCTGCTAACAAAGCCC	69
RSiaPCS	GTGAGACCAGTTACTGTGGTGGTGGTGGTGGTGCTCGAG	72
JS2011F	GGCGGAGGAAGTCTCGAGCACCACCACCACCACCAC	74
JS2012R	ACTTCCGCCACCTGGATTGATTGCTTCAATTTGTTTTA	63

Table 2.3. Plasmids used in this thesis

Name	Description	Source
pI.17	pUC-based plasmid carrying the gene for hemagglutinin.	Johnson et al., (2000)
pCR2.1	Linearized vector for TA cloning. AmpR, KanR, LacZα.	Invitrogen
pPCRHA1	pCR2.1 vector with *ha1* gene	This study
pLecB3	pKK233-3 (Amersham) derived plasmid for LecB expression with 6His-tag. Amp^R. Ptac.	Creavin *et al*, (in prep)
pJS102	pLecB3 derived, *lecB* replaced with *ha1*. Amp^R. Ptac	This study
pQE60-PelB	pQE60 (Qiagen) derived plasmid with PelB signal sequence for periplasmic expression of C-terminally 6His-tagged proteins. Amp^R. T5 promoter.	Ryan (Ph.D Thesis)
pJS103	pQE60-PelB derived for HA1 periplasmic expression with 6His-tag. Amp^R. T5 promoter.	This study
pMALp2E	Vector containing *malE* gene for periplasmic expression of N-terminally MBP tagged proteins. Amp^R. Ptac.	NEB
pMALp2Ef	pMALp2E (NEB) derived with random ~1kb DNA fragment cloned in EcoRI site. Amp^R. Ptac.	This study
pJS104	pMALp2E (NEB) derived for MBP fused HA1 periplasmic expression. Amp^R. Ptac.	This study
pGTY3	pET21b (Novagen) derived plasmid containing *siaP* gene.	Severi *et al*., (2005)
pJS201	pGTY3 derived plasmid for C-terminally 6His-tagged SiaP expression. T7 promoter. Amp^R.	This study
pJS201 derived plasmids:		
pJS202	Additional 6Lys-tag at C-terminus	This study
pJS203	Additional $Strep_2$-tag at C-terminus	This study
pJS2011	$(Gly\ Gly\ Gly\ Ser)_2$ linker introduced between SiaP and C-terminal 6His-tag	This study
pJS2021	$(Gly\ Gly\ Gly\ Ser)_2$ linker introduced between SiaP and C-terminal 6His-tag. Additional 6Lys-tag at C-terminus	This study
pJS2031	$(Gly\ Gly\ Gly\ Ser)_2$ linker introduced between SiaP and C-terminal 6His-tag. Additional Strep2-tag at C-terminus	This study

2.1.2 DNA molecular marker and protein molecular weight markers

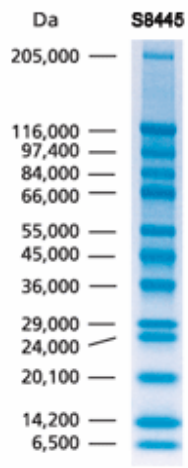
- 1 Kb Molecular Marker (Invitrogen).



- Protein molecular weight markers
    - Sigma Marker Wide Range (M.W. 6,500-200,000), S8445

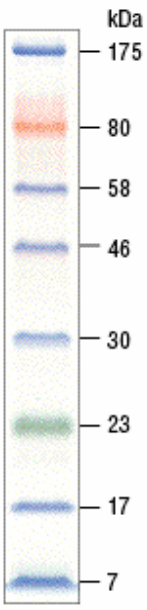
Constituent proteins:

Myosin (porcine heart) 205 kDa
β-Galactosidase (*E. coli)* 116 kDa
Phosphorylase B (rabbit muscle) 97 kDa
Glutamic dehydrogenase (bovine liver) 55 kDa
Albumin (bovine serum) 66 kDa
Ovalbumin (chicken egg) 45 kDa
Glyceraldehyde-3-phosphate dehydrogenase (rabbit muscle) 36 kDa
Carbonic anhydrase (bovine erythrocytes) 29 kDa
Trypsinogen (bovine pancreas) 24 kDa
Trypsin inhibitor (soybean) 20 kDa
α-Lactalbumin (bovine milk) 14.2 kDa
Aprotinin (bovine lung) 6.5 kDa



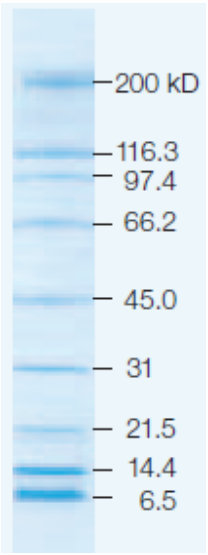
- NEB ColorPlus Prestained Protein Marker, Broad Range (7-175 kDa), P7709L

Constituent proteins:
MBP-β-galactosidase (*E. coli*) 175 kDa
MBP-truncated-β-galactosidase (*E. coli*) 80 kDa
MBP-CBDE (E. coli) 58 kDa
CBD-Mxe Intein-2CBDE (*E. coli*) 46 kDa
CBD-Mxe (*E. coli*) 30 kDa
CBD (*E. coli*) 23 kDa
Lysozyme (chicken egg white) 17 kDa
Aprotein (bovine lung) 7 kDa



- Bio-Rad Precision Plus Protein All Blue Standard (161-0373):

Constituent proteins:
Myosin (rabbit muscle) 200 kDa
β -Galactosidase (*E. coli*) 116 kDa
Phosphorylase b (rabbit muscle) 97 kDa
BSA 66 kDa
Ovoalbumin (hen egg white) 45 kDa
Carbonic anhydrase (bovine) 31 kDa
Tripsin inhibitor (soyabean) 21 kDa
Lysozyme (hen egg white) 14 kDa
Aprotein (bovine pancreas) 6.5 kDa



2.1.3 Media, Solutions and Buffers

Tryptone and Yeast extract were from Scharlau Chemie. Other chemicals were from Sigma Chemicals Co. Ltd and BDH Chemicals Ltd. All chemicals were Analar® grade.

- **Luria Bertani (LB) Broth**

Used for routine culturing of *E. coli*.

Tryptone	10 g
Yeast Extract	5 g
Sodium Chloride	10 g

NaOH was used to adjust to pH 7.0 and the volume brought to 1 L with distilled water (dH_2O). The solution was then sterilised by autoclaving at 121°C and 1.5 bar for 21 min. This medium was used for culturing of bacteria.

- **PBS Buffer**

NaCl	8 g
KCl	0.2 g
Na_2HPO_4	1.44 g
KH_2PO_4	0.24 g

Dissolved in 800 ml of distilled H_2O. Adjusted the pH to 7.4 with HCl. Added H_2O to 1 l. Sterilized by autoclaving.

- **TE Buffer**

Tris – HCl	10 mM
Na_2-EDTA	1 mM
pH	8.0

- **Solutions for the 1,2,3 Plasmid Preparation Method**

Solution 1:

Glucose (0.5 M)	1 ml
EDTA (0.1 M)	1 ml
Tris – HCl (1 M)	0.25 ml
dH_2O	7.75 ml

Solution 2:

NaOH (1 M)	2 ml
SDS (10%)	1 ml
dH_2O	7 ml

Solution 3:

Potassium Acetate	3 M
pH	4.8

To 60 ml of 5 M potassium acetate 11.5 ml of glacial acetic acid was added and 28.5 ml of dH_2O. The resulting solution was 3 M with respect to potassium and 5 M with respect to acetate.

- **50X Tris Acetate EDTA (TAE) Buffer**

EDTA (0.5 M)	100 ml
Glacial Acetic Acid	57.1 ml
Tris – HCl	242 g
pH	8.0
dH_2O	to 1 l

The solution was diluted to 1x with dH_2O before use.

- **Gel Loading Dye (6X)**

Bromophenol Blue	0.25%
Xylene Cyanol	0.25%
Ficoll (Type 400)	15%

Made in dH_2O and stored at 4°C.

- **TFB1 Buffer for Competent Cells**

RbCl	100 mM
$MnCl_2$	50 mM
Potassium Acetate	30 mM
$CaCl_2$	10 mM
Glycerol	15% v/v
pH	5.8

The pH of the solution was adjusted with HCl before the $MnCl_2$ was added. The solution was then sterilised through a 0.2 μm sterile filter and stored at 4^oC.

- **TFB2 Buffer for Competent Cells**

MOPS	10 mM
RbCl	10 mM
$CaCl_2$	75 mM
Glycerol	15% v/v
pH	6.8

The pH of the solution was adjusted with KOH. The solution was then sterilised through a 0.2 μm sterile filter and stored at 4^oC.

- **Lysis Buffer:**

Na_2HPO_4	6.9 g
NaCl	17.54 g
Imidazole	0.68 g

Adjusted to pH = 8.0 with NaOH and the volume brought to 1 L with dH_2O. The solution was then sterilised by autoclaving and stored at room temperature. Before use the solution was filtered through a 0.45 μm filter.

- **Running Buffer for SDS-PAGE (5x)**

Trizma Base	15 g
Glycine	72 g
SDS	5 g
dH_2O	to 1000 ml

The solution was stored at 4°C.

- **Sample Buffer for SDS-PAGE (5x)**

Tris-HCl pH 6.8 [0.5 M]	1.25 ml
Glycerol	5 ml
SDS [10% w/v]	2 ml
Bromophenol blue [0.5% w/v]	0.25 ml
Beta-Mercaptoethanol	0.5 ml

- **Coomassie Blue stain for SDS-PAGE**

Coomassie Blue	1 g
Methanol	180 ml
dH_2O	180 ml
Acetic Acid	40 ml

- **Destain Solution for SDS-PAGE**

Methanol	180 ml
dH_2O	180 ml
Acetic Acid	40 ml

- **Stop Solution:**

Trizma base	50 g
Acetic Acid	2.5% (v/v)
dH_2O	up to 1000 ml

- **Transfer Buffer for Western Blots**

Tris-HCl	25 mM
Glycine	192 mM
Methanol	20% (v/v)

- **Tris Buffered Saline (TBS) – Tween (10x)**

Tris-HCl	250 mM
NaCl	1.5 M
KCl	30 mM
Tween-20	1% (added from 20% (v/v) stock)
pH	8.0

- **Blocking Buffer for Western Blots**

Tris-HCl	25 mM
NaCl	150 mM
KCl	3 mM
Tween-20	0.1% (added from 20% (v/v) stock)
Milk powder	5% (w/v)
pH	8.0

- **TBST:**

Tris-HCl	20 mM
NaCl	150 mM
Triton X-100	0.1% (v/v)
$CaCl_2$	1 mM
$MnCl_2$	1 mM
$MgCl_2$	1 mM
pH = 7.6	

The buffer was prepared as a 5x stock (without Triton X-100 which was added directly before use) and kept at 4ºC. The $MnCl_2$ was added after the pH was brought to 7.6 with HCl.

- **TMB solution:**

Two mg of TMB (3,3´,5,5´-tetramethyl benzidine) were dissolved in 200 µl of DMSO and added to 9.8 ml of 50 mM citrate buffer pH 5.5. The solution was mixed thoroughly and 3 µl of H_2O_2 were added directly before use. The fresh solution was prepared before each use.

- **HBS buffer:**

HEPES	10 mM
EDTA	3 mM
NaCl	150 mM
Tween 20	0.005% (added from 20% (v/v) stock)
pH	7.4

Before Tween 20 addition the solution was filtered through a 0.2 µm filter. The fresh solution was prepared before each use.

2.2 Methods

2.2.1 Polymerase chain reaction (PCR)

This technique was used to amplify genes in order to clone them into plasmids.

- **Standard PCR Reaction Mixtures**
  - **With Phusion DNA Polymerase (Sigma):**

dNTP [10 mM] mix	1µl
5x Phusion Reaction Buffer containing 7.5 mM $MgCl_2$	10 µl
Template DNA (varying concentration)	1 µl
Forward primer [10 µM]	1.5 µl
Reverse primer [10 µM]	1.5 µl
Phusion DNA polymerase [2 u/ µl]	0.3 µl
Sterile dH_2O	to 50 µl

- **With Red Taq Polymerase (Sigma):**

dNTP [10mM] mix	1 µl
10x Red Taq PCR Buffer containing 500mM $MgCl_2$	5 µl
Forward primer [10 µM]	1 µl
Reverse primer [10 µM]	1µl
Red Taq DNA polymerase [5 u/µl]	1 µl
Sterile dH_2O	to 50 µl

- **PCR Program Cycles**
  - **with Phusion DNA Polymerase:**

Stage 1: Step 1: 95°C for 10 sec

Stage 2: Step 1: 95°C for 10 sec
Step 2: Annealing Temperature for 15 sec
Step 3: 72°C for 0.5 minute for every Kb to be synthesized

Stage 2 was usually repeated for 30 cycles

Stage 3: Step 1: 72°C for 10 minutes.

  - **with Red Taq DNA Polymerase:**

Stage 1: Step 1: 95°C for 10 minutes

Stage 2: Step 1: 95°C for 1 minute
Step 2: Annealing Temperature for 30 sec
Step 3: 72°C for 1 minute for every Kb to be synthesized.

Stage 2 was usually repeated for 30 cycles

Stage 3: Step 1: 72°C for 10 minutes.

2.2.2 Plasmid Preparation by the 1,2,3 Method

This method was used for the isolation of plasmids that were then used for analysis by agarose gel electrophoresis and restriction analysis. For ligation the gel extraction kit (RBC Bioscience). The principle of this method is selective alkaline denaturation of high molecular mass chromosomal DNA while covalently closed circular DNA remains double-stranded (Birnboim *et al.*, 1979). A 1.5 ml aliquot of a bacterial culture grown in selective media was pelleted by centrifugation at 12,000xg in a microfuge. The supernatant was

removed and the pellet was resuspended in 200 µl of solution 1 (section 2.3) by vortexing. Then 200 µl of solution 2 was added and mixed by inversion for several seconds. The tube was left at room temperature for 5 min. Then 200 µl of solution 3 was added and the tube was gently mixed by inversion for several seconds. The tube was then left for 10 min at room temperature. A clot of chromosomal DNA formed and was pelleted by centrifugation at 12,000xg for 10 min in a microfuge. The supernatant was transferred to a fresh tube and 450 µl of phenol: chloroform: isoamylalcohol (25:24:1) was added and mixed by vortexing. After centrifugation at 12,000xg for 5 min the upper, aqueous layer was carefully transferred to a fresh tube. An equal volume of isopropanol was added. After mixing by inversion for several seconds, the tube was centrifuged at 12,000xg for 30 min to pellet the plasmid DNA. The isopropanol was removed and the pellet washed with 750 µl 70% (v/v) ethanol and centrifuged at 12,000xg for 4 min. The ethanol was removed and the plasmid DNA was dried in a vacuum dryer for 1 minute. The plasmid DNA was resuspended in 50 µl of TE buffer and stored at 4^{o}C.

2.2.3 Agarose Gel Electrophoresis for DNA Characterization

DNA was analysed by separation on agarose gels in a horizontal gel apparatus. Gels were prepared by dissolving agarose in 1xTAE buffer to the required concentration (typically 0.7 – 1.2% [w/v]) and heating in the microwave oven until the agarose was fully melted. The 1xTAE buffer was also used as the running buffer. A tracker dye (1kb Molecular Marker [Invitrogen]) was incorporated into DNA samples (2 µl to 7 µl of the sample) to facilitate loading. Gels were frequently run at 140 V for 20 minutes. Gels were stained by immersing in a bath with ethidium bromide for 10-20 minutes. Gels were then visualized on a UV transilluminator and photographed using a UV image analyser.

2.2.4 Ethidium Bromide Stain

A 10 mg/ml stock solution in dH_2O was stored at 4ºC in the dark. For the staining of agarose gels, 100 µl of the stock solution was mixed with 1 l of dH_2O. The staining solution was kept in a plastic tray and covered to protect against light. Used ethidium bromide stain was collected and disposed of according to manufacturer instruction.

2.2.5 Isolation of DNA from Agarose Gels

The DNA sample was separated using agarose gel electrophoresis on a 0.7 % agarose gel containing SYBR Safe™ DNA gel stain (Invitrogen). HiYield Gel/PCR DNA Fragment Extraction Kit (RBC Bioscience) was used to extract relevant DNA fragments from the gel according to manufacturer's instructions.

2.2.6 DNA Restriction

This technique was used on DNA after PCR and plasmid DNA to clone the PCR product into a plasmid or in order to degrade the template DNA after site directed mutagenesis (use of KpnI enzyme restricting methylated DNA). The enzymes and buffers were from New England Biolabs (NEB) and were used according to manufacturer's instruction.

2.2.7 Ligation

This technique was used to ligate DNA fragments after DNA restriction with restriction enzymes. The DNA ligase and buffer were from NEB and were used according to manufacturer's instructions. Typical ligation mix:

Fresh PCR product/ DNA restriction fragment	2 µl
Vector DNA	2 µl
10X Ligation Buffer	1 µl
T4 DNA Ligase	1 µl
Sterile dH2O	to 10 µl

The reaction was performed at room temperature overnight.

2.2.8 TA Cloning of PCR Products

Polymerase chain reaction (PCR) products, after amplification with Red Taq polymerase (Sigma), were cloned using the TA Cloning Kit from Invitrogen into the pCR2.1 vector supplied. The diagram below shows the concept behind the TA Cloning method (Figure 2.1).

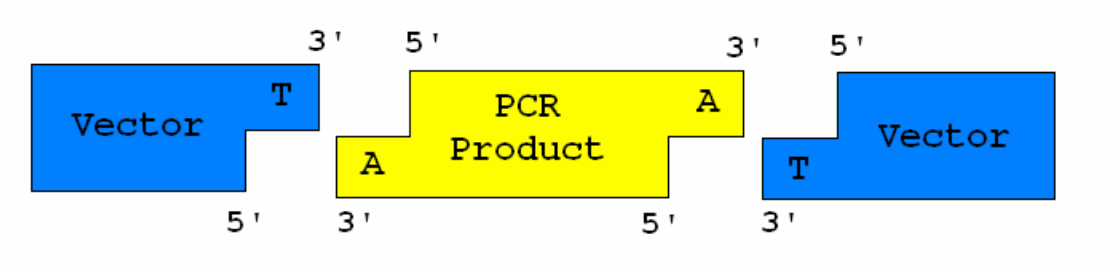


Figure 2.1. Principle of TA Cloning.

This method is based on the fact that thermostable polymerases like Red Taq lacks 3′ – 5′ exonuclease activity and because of that they leave 3′ A-overhangs. PCR products generated by Taq polymerases are subsequently ligated with the pCR2.1 vector which is linearized with 3′ T-overhangs. The ligation mix consisted of the following:

Fresh PCR product	1µl
pCR2.1 Vector (25 ng/ µl)	2 µl
10x Ligation Buffer	1 µl
Sterile dH_2O	5 µl
T4 DNA Ligase	1 µl

The reaction was performed at room temperature overnight. The reaction mix was then used to transform the competent cells.

2.2.9 Preparation of Competent Cells by RbCl Treatment

A glycerol stock of an *E. coli* strain was streaked on LB agar and incubated at 37^{o}C overnight. A 10 ml aliquot of LB broth was inoculated with a single colony from this fresh plate and incubated at 37^{o}C overnight. One ml of the overnight culture was added to 250 ml of LB broth in a 500 ml flask and incubated at 37^{o}C and 220 rpm until an OD_{600} of 0.5 was reached. The flask was then placed on ice for 10 min. The culture was transferred to a sterile, chilled 250 ml centrifuge tube and centrifuged at 3,500 rpm (Beckman TA-10-250 Rotor) at 4^{o}C for 5 min. The cell pellet was carefully resuspended in 75 ml of ice cold TFB1 buffer, placed on ice for 90 min and centrifuged as before. The cell pellet was gently resuspended in 10 ml of ice cold TFB2 buffer. The cell suspension was dispensed into 0.6 ml aliquots in microfuge tubes, flash frozen and stored at -80^{o}C.

2.2.10 Transformation of Competent Cells Prepared by RbCl Treatment

A 2 µl aliquot of a ligation mixture was added to 200 µl of competent cells that were thawed on ice and mixed by gentle flicking of the tube. The cells were then kept on ice for 30 min and then heat shocked at 42^{o}C for 60 s and immediately placed back on ice for 5 min. An 800 µl aliquot of LB broth was added to the cells and then incubated in a 37^{o}C water bath for 60 min. A 200 µl aliquot of the resulting transformation mixture was then aseptically spread on the LB agar plates containing an appropriate antibiotic and incubated at 37^{o}C overnight.

2.2.11 Determining cell efficiency

Competent cell efficiency is defined in terms of the number of colony forming units (cfu) obtained per µg of transforming plasmid DNA. A 0.25 µg/µl stock of pBR322 plasmid DNA was diluted to 250 pg/µl, 25 pg/µl and 2.5 pg/µl. Then 2 µl of each of the dilutions (and a negative control – no DNA) were used to transform competent cells as described above. The cell efficiency was calculated from the number of colonies obtained, taking into account the dilution factor and the fraction of culture transferred to the spread plate.

2.2.12 Bacterial storage

Bacterial strains were stored as 40 % (v/v) glycerol stocks. A 750 µl aliquot of an overnight culture (37ºC in LB broth) was mixed with 750 µl of sterile 80 % (v/v) glycerol in a microfuge tube. If the bacterial strains contained plasmids, the selective antibiotic was included in the culture. Duplicate stocks were stored at -20^{o}C and -70^{o}C. Working stocks streaked on LB agar plates, containing antibiotics where appropriate, were stored at 4^{o}C.

2.2.13 Antibiotic Preparation

Antibiotics used were from Sigma. Antibiotic stocks were prepared to a concentration of 100 mg/ml and stored at -20°C.

- **Ampicillin** was prepared in dH_2O and used at a final concentration of 100 µg/ml in both solid and liquid media.

- **Tetracycline** was prepared in 50% (v/v) ethanol and stored in the dark as it is light sensitive. It was used at a concentration of 10 μg/ml in both solid and liquid media.
- **Kanamycin** was prepared in dH_2O and used at a final concentration of 50 μg/ml in both solid and liquid media.

2.2.14 Soluble protein isolation

This technique was used for protein isolation from the soluble fraction of *E. coli* cells. A 100 ml culture of *E. coli* cells was centrifuged for 10 min at 6,000xg and the pellet was resuspended in 10 ml of Lysis Buffer. The cells were disrupted on ice with a 3 mm micro-tip sonicator (Sonics & Materials Inc) using 2.5 sec, 40 kHz pulses for 30 sec. The cell debris was removed by centrifugation at 13 000xg for 20 min at 4°C. The cell lysate was subsequently filtered through a 0.45 μm filter.

2.2.15 Inclusion bodies isolation and preparation

This technique was used for the isolation of inclusion bodies isolation from E. coli in order to analyse or solubilise HA1 protein. A 100 ml culture of E. coli cells was centrifuged for 10 minutes at 6,000xg and the pellet was resuspended in 10 ml of Lysis Buffer (section 2.1.3). The cells were disrupted on ice with a 3 mm micro-tip sonicator (Sonics & Materials Inc) using 2.5 sec, 40 kHz pulses for 30 sec. The sample was centrifuged at 13,000xg for 20 minutes at 4°C. The supernatant was removed and the pellet was resuspended in 5 ml of Lysis Buffer (section 2.1.3) containing 1 % (v/v) Triton X-100. The sample was mixed for 5 min at room temperature and centrifuged as previously. The washing step with 1 % (v/v) Triton X-100 was repeated once more and subsequently the inclusion bodies were washed with Lysis Buffer (section 2.1.3) but without Triton. The inclusion bodies were either solubilised straight after isolation or kept at -20°C.

2.2.16 Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis (SDS-PAGE)

SDS-PAGE was carried out as described by Laemmli (1970). Briefly, glass plates used for gel preparation were washed with detergent, rinsed with water and wiped with a tissue

soaked in 70% (v/v) ethanol. Two glass plates were put together, sealed with a gasket and kept together by two clamps. The matrix was prepared as follows:

	12% Resolving gel:	**5% Stacking gel:**
Acrylamide/bis-acrylamide (30/0.8%)	3.96 ml	0.425 ml
1 M Tris-HCl pH 8.8	1.945 ml	-
1 M Tris-HCl pH 6.8	-	0.312 ml
dH_2O	3.83 ml	1.71 ml
SDS	99 µl	25 µl
APS (10% [w/v])	49 µl	20 µl
TEMED	20 µl	10 µl

TEMED and freshly prepared APS were added straight before pouring the gels. The resolving gel was poured first up to 2 cm below the limit of the plates. It was layered with 200 µl of isopropanol and left for 10 minutes to polymerize. After polymerization an isopropanol layer was removed and the stacking gel was poured to the top of the cassette. A comb was inserted and the gel was left for 15 minutes to polymerize. The gels were fitted into the gel box at an angle so as to minimize the creation of air bubbles between the buffer and the bottom of the gel. The gel box was filled with 1x Running Buffer (section 2.1.3) for SDS-PAGE. Prepared samples were loaded into the gel by deposition into the gel wells. Typically the gels were run at 60 mA per two gels for 50 min or until the dyefront reached the end of the gel. Then the gels were immersed in Coomassie Blue stain (section 2.1.3) for SDS-PAGE and stained for 30 minutes with constant agitation at room temperature. Subsequently the gels were washed with dH_2O and immersed in Destain Solution for SDS-PAGE (section 2.1.3) for 1 h. The last step was repeated twice and the gels were placed between two sheets of acetate, scanned and saved in a suitable electronic format.

- **Sample preparation for SDS-PAGE**

Soluble proteins or soluble cellular extracts were prepared by mixing 20 µl of the sample with 5 µl of 5x Sample Buffer for SDS-PAGE. Then the samples were boiled for 5 minutes in a water bath, chilled at room temperature and applied on the gel.

Insoluble cell debris from 1ml of liquid culture after sonication and centrifugation were mixed with 60 µl of Sample Buffer and boiled for 10 minutes in a water bath. Typically a 5 µl aliquot of the boiled mixture was applied on the gel after cooling down at room temperature.

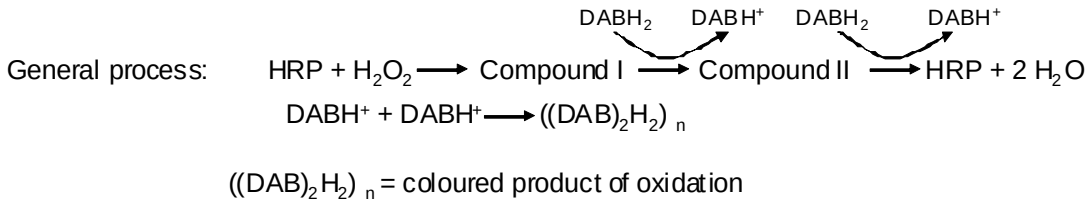
- **Silver Staining of SDS-PAGE Gels (Merril *et al.*, 1981)**

Silver staining methods are significantly more sensitive than various Coomassie Blue staining techniques (Rabilloud *et al.*, 1994). Consequently, they are the method of choice when very low amounts of protein have to be detected on electrophoresis gels. After SDS-PAGE the gel was stained and destained using Coomassie Blue stain in order to maximise sensitivity. The gel was then subjected to the fixation process – 1 h in 100 ml of 30% (v/v) ethanol and 10% (v/v) acetic acid. Then the gel was washed for 15 min in 100 ml of 20% (v/v) ethanol followed by 15 min wash in water. Then sensitization took place that involved 1 min wash in 100 ml of 0.1% sodium thiosulphate (222 µl of 4.5% (w/v) stock). The gel was then rinsed two times with water for 15 s and subjected to silver staining – 30 min incubation in 100 ml of 0.1% silver nitrate (1 ml of 10% (w/v) stock) and 70 µl of 37% (v/v) formaldehyde. The gel was then rinsed with water twice for 20 s. Then the development process took place – incubation in 100ml of 3% (w/v) Sodium Carbonate, 50 µl of 37% (v/v) formaldehyde and 4 µl of sodium thiosulphate 4.5% (w/v) stock. The reaction was stopped when the protein bands were fully stained in dark brown colour using 100 ml Stop Solution (section 2.1.3).

2.2.17 Western Blotting

An SDS-PAGE gel was run as described above. Protein samples were loaded along with prestain molecular weight marker. Four pieces of filter paper (Whatman) and one piece of nitrocellulose membrane were cut to the size of the gel. The pieces of paper and membrane were soaked in Transfer Buffer (section 2.1.3) for 10 min. The gel was rinsed with water. Two pieces of filter paper followed by the membrane, the gel and another two pieces of filter paper were placed on the cathode of the horizontal semi-dry electro-blotter, avoiding air bubbles. The stack was covered with anode and the transfer was carried out for 30 min at 10 V. As optional verification of successful transfer, the membrane was incubated for 1

minute in 0.1% Ponceau S (Sigma), 5% acetic acid staining solution with gentle agitation in order to visualize transferred proteins. The membrane was subsequently destained by several washes in dH_2O. The membrane was then blocked using 50 ml of Blocking Buffer (section 2.1.3) for 60 min (or overnight at 4°C) and incubated with 15 ml anti-His HRP-labeled antibodies [Sigma; 1:5000 in Blocking Buffer (section 2.1.3)] for 60 min. The membrane was then washed several times with TBS-Tween (section 2.1.3) and developed using 3, 3-diaminobenzidine tetrahydrochloride (DAB, product no. d5637, Sigma) tablets. The DAB tablets were applied as a chromogen substrate for horse radish peroxidase and in that way they enabled the staining procedure. DAB produces a brown colour during the enzymatic reaction catalysed by horse radish peroxidase. The tablet container was removed from the refrigerator and allowed to stand at room temperature for about 15 min. Tablets were dissolved in 15 ml of dH_2O. The solution was then poured on the top of the membrane and left to react for 5 min. The membrane was subsequently washed with water and photographed. It could also be dried and saved.



2.2.18 Solubilisation of inclusion bodies

- **Urea solubilisation:**

The inclusion bodies from a 100 ml overnight culture of *E. coli* cells were isolated. The inclusion bodies were mixed with 10 ml of 8 M Urea in 50 mM TRIS buffer pH 8.0, shaken vigorously for 5 min and split into 3 samples. Each sample was incubated for 1 h under mild agitation at different temperatures. The temperatures were respectively 4°C, 22°C and 37°C. Then, to verify the effectiveness of the solubilisation, the samples were split in two and one half was centrifuged for 10 min at 13,000xg. Both the supernatant as a soluble fraction, and the uncentrifuged sample as a control of protein presence, were subjected to SDS-PAGE analysis and Western blotting.

- **Guanidine Hydrochloride GnHCl solubilisation:**

The inclusion bodies from a 100 ml overnight culture of *E. coli* cells were isolated. The inclusion bodies were mixed with 10 ml of 6 M GnHCl in 50 mM TRIS buffer pH 8.0, shaken vigorously for 5 min and split into 3 samples. Each sample was incubated for 1 h under mild agitation at different temperatures. The temperatures were respectively 4°C, 22°C and 37°C. Then, to verify the effectiveness of the solublisation, the samples were split in two and one half was centrifuged for 10 min at 13,000xg. Both the supernatant as a soluble fraction and the uncentrifuged sample as a control of protein presence were subjected to SDS-PAGE analysis and Western blotting.

2.2.19 Protein purification

Affinity Chromatography was used to purify recombinant proteins containing 6xHis tag or/and Strep2 tag.

- **Standard Immobilised Metal Affinity Chromatography (IMAC) procedure**

One ml of nickel-nitrilotriacetic acid agarose resin (Ni-NTA, Qiagen) was poured into a column (“5" Polypropylene chromatography column” with a medium [45-90 µm] filter [Evergreen Scientific]). Once the resin settled it was washed with 5 ml of distilled water to wash off the ethanol and then pre-equilibrated with 5 ml of Lysis Buffer (section 2.1.3). Then 10 ml of the protein isolate was applied on the resin and washed with 5 ml of Lysis Buffer (section 2.1.3). The resin was subsequently washed with 5 ml of Lysis Buffer (section 2.1.3) with 20 mM – 120 mM imidazole (concentration needs to be optimized previously for a given protein) and eluted with 250 mM imidazole in Lysis Buffer (section 2.1.3). Fractions containing purified protein were pooled together and desalted using Amicon columns with 10,000 Da cut off point. The column after use was washed with 10 column volumes of water followed by 10 column volumes of 20% (v/v) ethanol and stored filled with 20% (v/v) ethanol at 4°C.

- **IMAC purification under denaturing conditions**

Proteins solubilised in either urea or guanidine hydrochloride were applied on pre-equilibrated columns as described above. Purification was performed similarly to the standard IMAC purification but denaturant (urea or guanidine hydrochloride) was present in the buffers.

- **IMAC purification using FPLC**

The purification was performed using HisTrap FF 1ml colum prepacked with precharged NiSepharose™ 6 Fast Flow (GE Healthcare). The flow was 0.5 – 1 ml/minute and the maximum bar pressure was 0.5 bar. After column attachment according to the manufacturer's instructions it was washed with 5 ml of water to remove the 20% (v/v) ethanol used for storage and then washed with 10 ml of Lysis Buffer (section 2.1.3) to equilibrate the resin. Then 50 ml of protein isolate were applied from the injection loop. The column was washed using buffers A and B (various volumes – in order to achieve various concentrations of imidazole) containing Lysis Buffer (section 2.1.3) and 250 mM imidazole in Lysis Buffer (section 2.1.3) respectively. The buffers were mixed automatically by the pumping and mixing machinery (ÄKTAxpress) in order to achieve desired imidazole concentration. The step gradient of imidazole was applied to remove unspecifically bound contaminants from the column and subsequently to elute the His-tagged protein of interest. Protein elutions were monitored at 280 nm and the resulting fractions were analyzed by SDS-PAGE. Fractions containing purified protein were pooled together and desalted using Amicon columns with 10,000 Da cut off point. After the run the column was washed with 10 ml of water followed by 10 ml of 20% ethanol and stored at room temperature, filled with 20% (v/v) ethanol.

- **Affinity chromatography using StrepTrap column and Strep2 tag**

StrepTrap HP 1 ml columns (GE Healthcare) containing StrepTactin Sepharose™ were used for Strep2-tagged protein purification. The purification was done according to the manufacturer's instructions. Briefly, the cell extract of *E. coli* after protein isolation in a

binding buffer was applied on the pre-equilibrated StrepTrap HP 1 ml column and washed with binding buffer to remove contaminants. The Strep2-tagged protein of interest was eluted using desthiobiotin. The protein was then washed with water using ultrafiltration columns (Amicon) with a 10,000 Da cut off point.

2.2.20 Enzyme Linked Lectin Assay (ELLA)

The assay was performed on NUNC MaxiSorp 96 well plates. Fifty µl of 1-10 µg/ml glycoprotein was applied in PBS buffer on the plate and incubated overnight at 4ºC to immobilise it. Then an excess of the glycoprotein was removed and 150 µl of 2.5% (w/v) BSA in TBST (section 2.1.3) was applied on the plate to block it. After one hour incubation at 4ºC 2.5% (w/v) BSA in TBST was removed and the well of the plate was rinsed three times with 300 µl of TBST (section 2.1.3). Then 50 µl of 1-10 µg/ml lectin was applied in 2.5% BSA-TBST and incubated for 1 hour at room temperature. Next the lectin was removed and the well of the plate was rinsed three times with 300 µl of TBST (section 2.1.3). A 50 µl aliquot of 1:10 000 diluted anti-lectin, HRP-linked antibodies in 2.5% BSA-TBST were applied to react for 1 hour. Then the antibody was removed and the well of the plate was rinsed three times with 300 µl of TBST (section 2.1.3). A 100 µl aliquot of TMB solution (section 2.1.3) was applied and left for 5 minutes at room temperature and the reaction was stopped by addition of 50 µl of 10% (v/v) HCl. The absorbance was read at 450 nm. Each sample was tested in triplicate.

2.2.21 Mass Spectrometry (MS)

Protein solutions were desalted (Amicon Ultra-15 Centrifugal Filter, 10,000 Da cut-off), reconstituted into MS grade water and diluted with 10 mM ammonium formate (pH 4) to obtain appropriate concentrations (5-20 pmol/µl). Buffered solutions were directly infused into the quadrupole time-of-flight QTOF mass spectrometer (Waters QToF Ultima Global) using a standard electrospray interface (ESI). The collision gas was Argon. Spectra were manually deconvoluted to obtain the molecular mass of the analytes. The software used for data analysis and processing was MassLynx version 4.1. The following parameters were used and were found to be critical for the analysis: accelerating voltage 3.25 kV, extraction

cone voltage (Vc) 40V declustering potential (DP) 30 V, collision energy (CE) 10 V, Pi 5.2 mbar, ST 65°C, and desolvation temperature (DT) 65°C.

2.2.22 Surface plasmon resonance using a Biacore biosensor

Binding analysis of the SiaP protein with Sia was undertaken by surface plasmon resonance using a biosensor (Biacore 3000; Biacore). Covalent linkage of the protein of interest to the dextran matrix was achieved by amine coupling (Amine Coupling kit, Biacore). Activation of the carboxyl groups on the matrix was performed by injection of 70 µl of the 1:1 mixture of *N*-hydroxysuccinimide (NHS, 11.5 mg/ml in deionized water) and ethyl-N-(3-diethylaminopropyl) carbodiimide (EDC, 75 mg/ml in deionized water) at 10 µl/min to form active esters that react spontaneously with amine groups on the protein. The protein of interest (usually at concentration of 100 µg/ml) was immobilised onto a CM5 sensor chip in 10 mM acetic acid. The pH was previously optimized during pre-concentration studies where 5 µl of the protein of interest was injected in a range of pH (3.6 - 5.0). All subsequent binding experiments were performed in HBS buffer (section 2.1.3). Samples were applied to the sensor chip at a flow rate of 10 µl/min. The remaining active esters were subsequently deactivated by injection of 80 µl of ethanolamine-HCl (1 M, pH 8.5). Unspecifically bound material was removed by regeneration with 10 mM NaOH (three 5 µl flushes). All monosaccharides tested were injected in HBS buffer (section 2.1.3).

2.2.23 *In silico* analysis of DNA and protein sequences

In silico analysis of DNA and protein was performed using programmes listed in Table 2.4.

Table 2.4. Programs used for DNA and protein analysis.

Name	Features	Application	Source
Basic Local Alignment Search Tool (*BLAST*)	Finds regions of local similarity between sequences deposited in GenBank.	Used to compare DNA sequences.	(Altschul *et al.*, 1990); available at http://blast.ncbi.nlm.nih.gov/Blast.cgi
ClustalW	General purpose multiple sequence alignment program for DNA or proteins.	Used to create and manipulate multiple sequence alignments of DNA and proteins.	(Thompson *et al.*, 1994); available at http://www.ebi.ac.uk/Tools/clustalw2/index.html

Name	Features	Application	Source
GeneDoc	Full Featured Multiple Sequence Alignment Editor, Analyser and Shading Utility	Used to edit multiple sequence alignments of DNA and proteins.	Karl Nicholas; available at http://www.nrbsc.org/gfx/genedoc
Webcutter 2.0	Restriction analysis	Used to perform *in silico* restriction analysis of DNA sequences.	Max Heiman; available at http://rna.lundberg.gu.se/cutter2/
Oligo Calc	Oligonucleotide Properties Calculator	Used during primer design.	(Kibbe, 2007); available at http://www.basic.northwestern.edu/biotools/oligocalc.html
Protein Workshop	3D protein structure viewer	Used to create 3D protein images.	(Moreland *et al.*, 2005); available at www.rscb.org
MassLynx version 4.1	MS data processing and analysis.	The software was used for data analysis and processing.	www.waters.com

Chapter 3
Results

3 Results

3.1 Overview

The application of Sia-binding proteins in various analytical platforms for Sia detection is inevitably dependent on production of such proteins by methods that are reproducible with respect to the protein properties, as well as being cost efficient. The utilization of a prokaryotic system for protein expression is generally considered to be the most effective in terms of both cost and time. Furthermore, the well defined genetics and physiology of *Escherichia coli* makes it the host of choice for protein production and also enables relatively easy and fast modifications of the expressed protein to be undertaken by application of different vectors, strains and conditions. Additionally, the production of functional, non-glycosylated mammalian protein in *E. coli* would be advantageous, resulting in homogenous product, in contrast to protein product obtained from mammalian systems (Castilho, 2008). Thus, the activity of the product would not be subjected to variations in the glycan component of the glycoprotein, which often happens with such proteins like mammalian enzymes. To avoid drawbacks of glycosylation and to achieve the benefit of cost and time effectiveness, the *E. coli* based system of expression was applied for production of proteins in this project. Subsequently, several analytical platforms were utilized for characterization of the produced protein and ultimately the applicability of produced Sia-binding protein was tested for Sia detection and quantification.

3.2 Production of haemagglutinin protein (HA) from Influenza A virus

The goal of this part of the project was to produce the globular domain (HA1) of the haemagglutinin protein (HA) from Influenza A virus in *E. coli* cells and to characterize it.

3.2.1 Initial cloning of the *ha1* gene

The Influenza A virus (A/Sichuan/2/87(H3N2)) gene for haemagglutinin (HA) was obtained on a plasmid from Dr. Patricia Johnson (DCU, Ireland). On the basis of the *ha* gene sequence (PubMed accession number D21173, Nerome *et al.*, 1995) primers were designed in order to amplify a *ha1* sequence for the globular domain of the

CAAAAACTTCCCGGAA**ATG**ACAACAGCACAGCAACGCTGTGCCTGGGACATCATGCAGTGCCAAACG
GAACGCTAGTGAAAACAATCACGAATGATCAGATTGAAGTGACTAATGCTACTGAGCTGGTTCAGAG
TTCCTCAACAGGTAGAATATGCGACAGTCCTCACCGAATCCTTGATGGAAAAAACTGCACACTGATA
GATGCTCTATTGGGAGACCCTCATTGTGATGGCTTCCAAAATGAGAAATGGGACCTTTTTGTTGAAC
GCAGCAAAGCTTACAGCAACTGTTACCCTTATGATGTGCCGGATTATGCCTCCCTTAGGTCACTAGT
TGCCTCATCAGGCACCCTGGAGTTTATCAATGAAGACTTCAATTGGATTGGAGTCACTCAGAGTGGG
GGAAGCTGTGCTTGCAAAAGGGGATCTGTTAACAGTTTCTTCAGTAGATTGAATTGGTTGCACAAAT
CAGAATACAAATATCCAGCGCTGAACGTGACTATGCCAAACAATGGCAAATTTGACAAATTGTACAT
TTGGGGGGTTCACCACCCGAGCACGGACAGAGAACAAACCAACCTATATGTTCGAGCATCAGGGAGA
GTCACAGTCTCTACCAAGAGAAGCCGGCAAACTGTAATCCCGAATATCGGGTCTAGACCCTGGGTAA
GGGGTCTGTCCAGTAGAATAAGCATCTATTGGACAATAGTAAAACCGGGAGACATACTGTTGATTAA
TAGCACCGGGAACCTAATTGCTCCTCGGGGTTACTTCAAAATACGCACTGGGAAAAGCTCAATAATG
AGGTCAGATGCACCTATTGGCACCTGCATTTCTGAATGCATCACTCCAAATGGAAGCATTCCCAATG
ACAAACCCTTTCAAAATGTAAACAAGATCACATATGGGGCATGCCCCAGGTATGTTAAGCAAAACAC
TCTGAAATTGGCAACAGGGATGCGGAATGTACCAGAGAAACAAACTAGAGGCATATTCGGCGCAATA
GCAGGTTTCATAGAAAATGGTTGGGAGGGAATGGTAGACGGTTGGTACGGTTTCAGGCATCAAAATT
CTGAGGGCACAGGACAAGCAGCAGATCTTAAAAGCACTCAAGCAGCAATCGACCAAATCAACGGGAA
ACTGAATAGGTTAATCGAGAAAACGAACGAGAAATTCCATCAAATCGAAAAGGAATTCTCAGAAGTA
GAAGGGAGAATTCAGGACCTCGAGAAATATGTTGAAGACACTAAAATAGATCTCTGGTCTTACAACG
CGGAGCTTCTTGTCGCCCTGGAGAACCAACATACAATTGATCTGACTGACTCAGAAATGAACAAACT
GTTTGAAAAAACAAGGAAGCAACTGAGGGAAAATGCTGAGGACATGGGCAATGGTTGCTTCAAGATA
TACCACAAATGTGACAATGCCTGCATAGGGTCACTCAGAAATGGAACTTATGACCATGATGTATACA
GAGACGAAGCATTAAACAACCGGTTCCAGATCAAAGGTGTTGAGCTGAAGTCAGGATACAAAGACTG
GATACTGTGGATTTCCATTGCCATATCATGCTTTTTGCTTTGTGTTGTTTTGCTGGGGTTCATCATG
TGGGCCTGCCAAAAAGGCAACATTAGGTGCAACATTTGCATTA

Figure 3.1. A **s**equence of the *ha* gene. The globular domain is underlined and primers for *ha1* amplification marked in red.

The primer's sequences (Table 3.1) contained an NcoI restriction site (forward primer) and a BglII restriction site (reverse primer) to enable further cloning of the gene into an expression vector. The *ha1* gene was initially cloned into the pPCR2.1 vector to avail of the TA cloning and enable the amplified sequence of the gene to be verified. PCR reactions were carried out using RedTaq DNA Polymerase with both HA1F1 and

HA1R1 primers. The optimised annealing temperature was 54°C, annealing time 1 min and extension time 2 min.

Table 3.1. PCR primers used in the amplification of the *ha1* gene of the Influenza A virus (A/Sichuan/2/87(H3N2)). Restrictions sites are highlighted and described

	NcoI
HA1F1:	5′ CC ATG GCA ATG CAA AAA CTT CCC GGA AAT GAC AAC AGC 3′
	Ala
	BglII
HA1R1:	5′ AGA TCT AGT TTG TTT CTC TGG TAC ATT CCG C 3′

Products of the PCR reaction were verified by agarose gel electrophoresis. DNA fragments of ~ 1kb in size, which correspond to the *ha1* gene (990 bp), were gel extracted and used for a ligation reaction with a linearized PCR2.1 vector from the TA cloning kit (Invitrogen). The ligation mixture was used to transform *E. coli* XL10-Gold competent cells by heat shock. The resulting plasmid was named pPCRHA1 (Figure 3.2). As the MCS of the PCR2.1 vector is situated inside the lacZα gene the initial screening of the transformants was done on the basis of lacZ α-complementation test. The insert is flanked on both sides by EcoRI sites in the PCR2.1 vector and there is no EcoRI site in the *ha1* gene (Nerome *et al.*, 1995). The successful cloning of the gene was further confirmed by restriction analysis with EcoR1 enzyme. A 1 kb fragment was generated which corresponds to the size of the *ha1* gene (Figure 3.2). Subsequently the *ha1* gene was sequenced (MWG-Biotech Ltd). The results were aligned with the expected sequence of the gene (Figure 3.3).

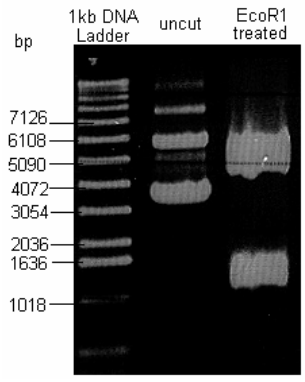


Figure 3.2. An EcoRI restriction analysis of the pPCRHA1 plasmid after an agarose gel electrophoresis.

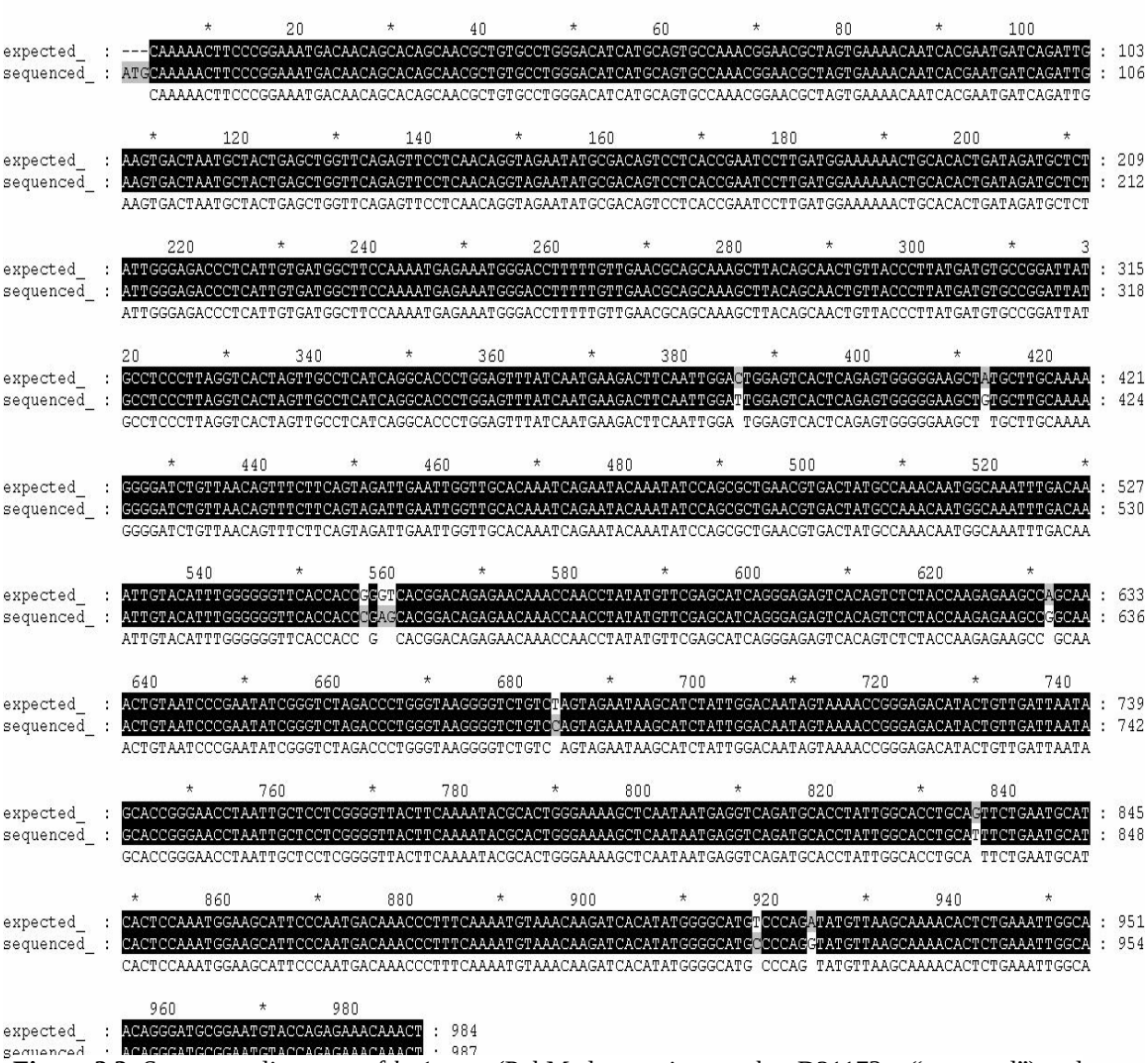


Figure 3.3. Sequence alignment of *ha1* gene (PubMed accession number D21173 = "expected") and the *ha1* gene in pPCRHA1 plasmid = "sequenced".

The sequencing confirmed the sequence of the *ha1* gene in the TA clone; however few mismatches were found. The changes in the gene sequence correspond to changes in the amino acid sequence of the HA1 protein. Six alterations were found in the amino acid sequence of the TA clone (Figure 3.4). In order to avoid the immune response of the host the surface of influenza virus changes rapidly which involve random mutations in both haemagglutinin and neuraminidase proteins (Peiris *et al.*, 2007). The difference between the expected and the obtained amino acid sequence of the HA1 protein may be attributed to such spontaneous changes. As the reading frame as well as amino acids involved in ligand binding were unchanged further work was undertaken using the pPCRHA1 plasmid.

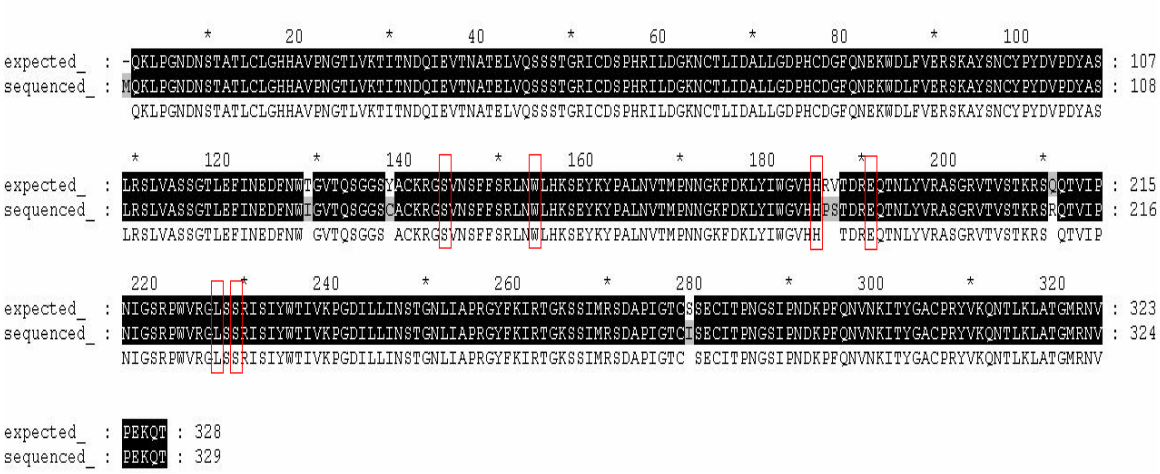


Figure 3.4. Amino acid sequence alignment of HA1 protein (PubMed accession number D21173 = "expected") and HA1 protein encoded in pPCRHA1 plasmid = "sequenced". The amino acids involved in ligand binding are marked with red rectangles.

In order to facilitate efficient expression of the *ha1* gene in *E. coli* the gene was subcloned into the pLecB3 (Figure 3.5) expression vector (pKK233 derived), replacing the *lecB* gene. To facilitate further purification of the HA1 protein the gene was cloned in frame with a vector borne 6x histidine tag. An ampicillin resistance gene encoded on the plasmid enabled plasmid selection. The expression of the *ha1* gene was facilitated by the vector borne Ptac promoter and replication by the ColE1 origin of replication. The *ha1* gene in the pPCRHA1 plasmid is flanked by NcoI and BglII sites introduced in the primer sequences during PCR (Table 3.1). Similarly the *lecB* gene in the pLecB3 plasmid is flanked by NcoI and BglII sites. This pPCRHA1 and pLecB3 plasmids were restricted with NcoI and BglII enzymes. DNA fragments were separated by agarose gel electrophoresis. A 1 kb DNA fragment, corresponding to the *ha1* gene, was extracted from the gel where the pPCRHA1 plasmid restriction mix was separated and a 4.6 kb DNA fragment corresponding to an empty pLecB3 vector was gel extracted from the gel where the pLecB3 plasmid restriction mix was separated. Subsequently, the *ha1* gene and the vector were ligated resulting in formation of the pJS102 plasmid (Figure 3.6). pJS102 has a relatively small size (5.6 kb) which promotes successful transformation of competent cells (Hanahan, 1983). The sequence of the plasmid was verified by restriction analysis and subsequent sequencing (MWG-Biotech Ltd) (not shown).

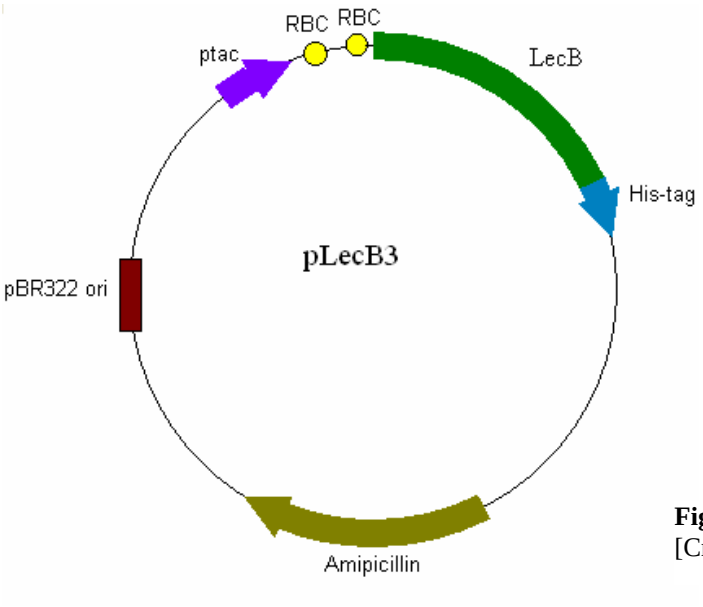


Figure 3.5. A map of the pLecB3 plasmid [Creavin *et al* (in prep)].

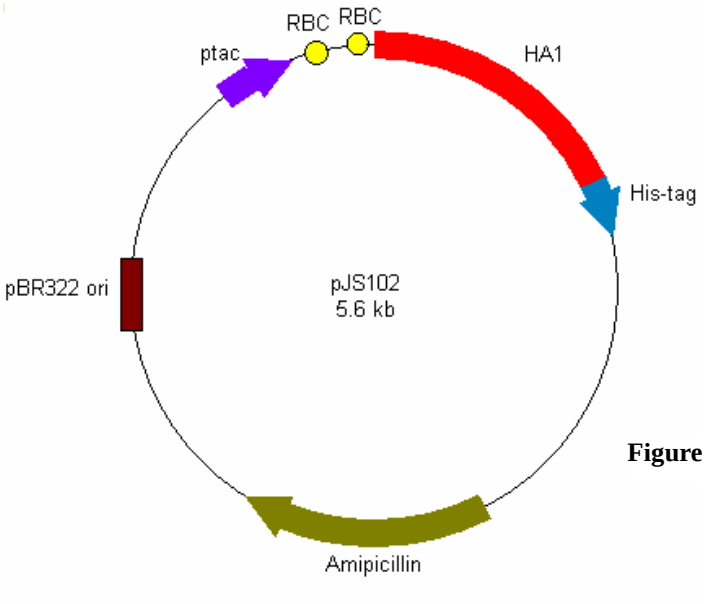


Figure 3.6. A map of the pJS102 plasmid.

3.2.2 Initial expression of the haemagglutinin globular subunit

Expression in BL 21 and Rosetta strains of *E. coli*

The heterogenic proteins expressed in *E. coli* cells are often subjected to proteolytic degradation by the host proteases, especially if the protein comes from an evolutionarily distinct organism. Taking into consideration the fact that the HA1 protein is native for mammalian expression systems the initial expression was performed using protease deficient strains of *E. coli* to prevent the possible protein degradation. The BL 21 *E. coli* strain and its derivatives harbour mutations of the protease genes that make them protease deficient. It is also widely known that expression of a mammalian protein in a prokaryotic system may result in lower yield or even lack of expression caused by different codon distribution in bacterial and mammalian genes (Bonekamp *et al.*, 1988). The bioinformatics analysis of the *ha1* gene sequence revealed 15 rare Arg codons, 5 rare Leu codons, 7 rare Ile codons and 5 rare Pro codons (Figure 3.7). These rare codons in *ha1* gene might interfere with expression in *E. coli*.

atg gca atg caa aaa ctt CCC gga aat gac aac agc aca gca acg ctg tgc ctg gga cat cat
gca gtg cca aac gga acg CTA gtg aaa aca atc acg aat gat cag att gaa gtg act aat gct
act gag ctg gtt cag agt tcc tca aca ggt AGA ATA tgc gac agt cct cac CGA atc ctt gat
gga aaa aac tgc aca ctg ATA gat gct CTA ttg gga gac cct cat tgt gat ggc ttc caa aat
gag aaa tgg gac ctt ttt gtt gaa cgc agc aaa gct tac agc aac tgt tac cct tat gat gtg
ccg gat tat gcc tcc ctt AGG tca CTA gtt gcc tca tca ggc acc ctg gag ttt atc aat gaa
gac ttc aat tgg att gga gtc act cag agt ggg gga agc tgt gct tgc aaa AGG gga tct gtt
aac agt ttc ttc agt AGA ttg aat tgg ttg cac aaa tca gaa tac aaa tat cca gcg ctg aac
gtg act atg cca aac aat ggc aaa ttt gac aaa ttg tac att tgg ggg gtt cac cac ccg agc
acg gac AGA gaa caa acc aac CTA tat gtt CGA gca tca ggg AGA gtc aca gtc tct acc aag
AGA agc cgg caa act gta atc ccg aat atc ggg tct AGA CCC tgg gta AGG ggt ctg tcc agt
AGA ATA agc atc tat tgg aca ATA gta aaa ccg gga gac ATA ctg ttg att aat agc acc ggg
aac CTA att gct cct cgg ggt tac ttc aaa ATA cgc act ggg aaa agc tca ATA atg AGG tca
gat gca cct att ggc acc tgc att tct gaa tgc atc act cca aat gga agc att CCC aat gac
aaa CCC ttt caa aat gta aac aag atc aca tat ggg gca tgc CCC AGG tat gtt aag caa aac
act ctg aaa ttg gca aca ggg atg cgg aat gta cca gag aaa caa act AGA tct cat cac cat
cac cat cac taa

Figure 3.7: Bioinformatic analysis of rare codon distribution in *ha1* gene by Rare Codon Calculator – RaCC. Red = rare Arg codons AGG, AGA, CGA, Green = rare Leu codon CTA, Blue = rare Ile codon ATA, Orange = rare Pro codon CCC.

The Rosetta strain of *E. coli* was utilized to express the HA1 protein, as this strain contains a pRARE plasmid harbouring tRNAs for mammalian codons that rarely occur in *E. coli* (Novagen). The HA1 protein was successfully expressed in both BL21 and Rosetta strains of *E. coli* (Figure 3.8). However the HA1 protein was found solely in an insoluble form (Figure 3.8 lanes 3 and 4). This indicates that the expression of HA1 is successful in this system, hovewer it is not optimal.

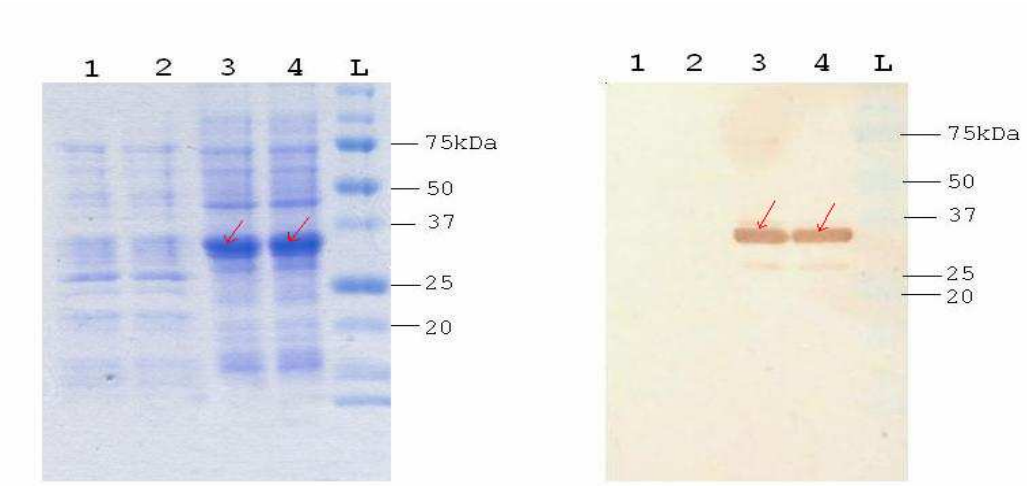


Figure 3.8: SDS-PAGE (left) analysis and the corresponding Western blot (right) analysis of the protein isolates from BL 21 and Rossetta strains of *E. coli* used to overexpress the His-tagged HA1 protein. Lanes were loaded as follows: (1) - soluble fraction, BL21; (2) - soluble fraction, Rosetta; (3) - insoluble fraction, BL21; (4) - insoluble fraction, Rosetta, (L) - Bio-Rad prestained protein ladder. Red arrows indicate His-tagged HA1 protein detected by Western Blotting using anti-His antibodies. The calculated molecular weight of the HA1 protein is 38918.89 Da. Samples were collected after overnight culturing at 37°C in 100ml LB amp broth.

Overexpression of the proteins often leads to deposition of extensively produced protein in the form of inclusion bodies (Vallejo *et al.*, 2004). One of the possible ways to prevent aggregation of the product is to lower the expression rate of the protein. Reduced rate of expression promotes correct folding of the protein and thus promotes its solubility (Dyson *et al.*, 2004). A widely applied method to slow down the expression is to lower the concentration of inducing agent (Winograd *et al.*, 1993). As the initial expression of the HA1 protein was performed in BL21 and Rossetta, which offer no induction regulation, other strains of *E. coli* were used.

Expression in XL10 Gold strain of *E. coli*

In order to obtain control over protein expression the XL10-Gold strain of *E. coli* was used. The $lacI^q$ mutation is responsible for overproduction of the *lac* repressor protein

(LacI) in this strain which facilitates efficient control over the expression of the genes that are under *lac* (or *lac* derived) promoter in the absence of the IPTG. The HA1 protein was successfully expressed in XL10-Gold and the expression was tightly regulated as there was no expression of the HA1 protein in uninduced cultures (Figure 3.9).

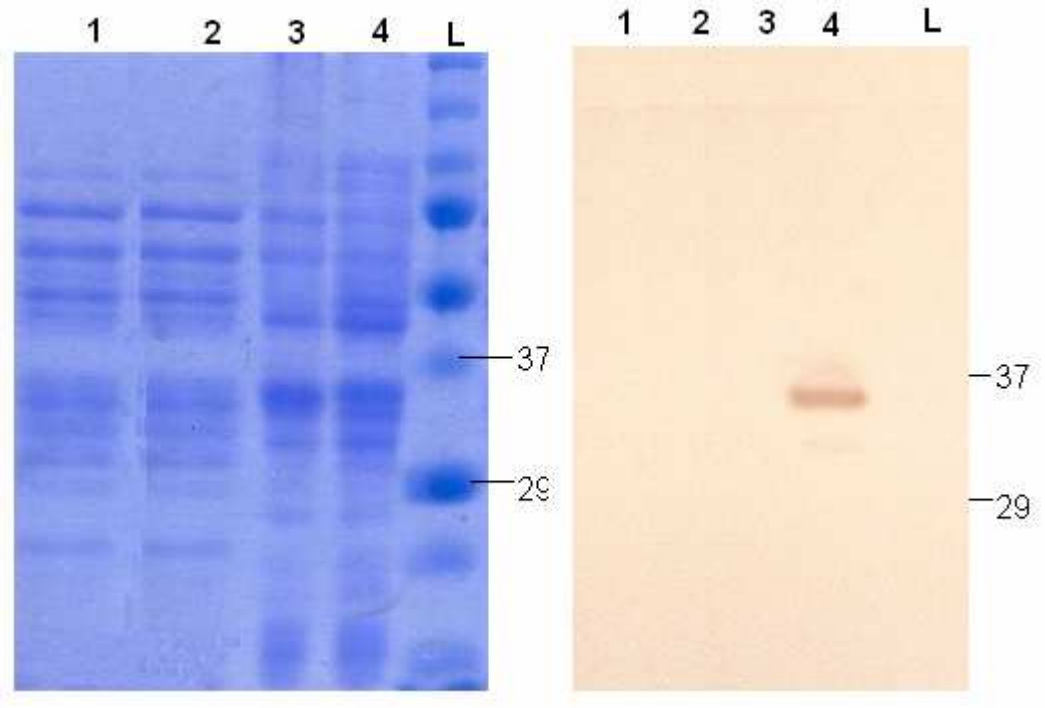


Figure 3.9. SDS-PAGE (left) analysis and the corresponding Western blot (right) analysis of the protein isolates from XL 10 Gold strain of *E. coli* used to overexpress the His-tagged HA1protein. Lanes were loaded as follows: (1) soluble fraction, uninduced culture, (2) soluble fraction, induced culture, (3) insoluble fraction, uninduced culture, (4) insoluble fraction, induced culture, (L) Bio-Rad prestained protein ladder. Cultures were induced with 50 μM IPTG at OD=0.6. Samples were collected after overnight culturing at 37°C in 100ml LB amp broth. The Western Blot was probed with anti-His antibodies

Expression conditions optimization

Utilization of the XL 10 Gold strain enabled controlled expression of the HA1 protein; however, the HA1 protein was exclusively expressed in the inclusion bodies. In order to down-regulate the expression of HA1 protein, and thus promote its solubility, lower concentrations of IPTG were used for induction. To investigate the influence of IPTG level on the expression of a soluble HA1 protein the XL10-Gold pJS102 cultures were induced with different concentrations of IPTG. The IPTG concentrations applied were as follows: 0.5 μM, 1 μM, 5 μM, 50 μM, 100 μM, 250 μM, 500 μM and 1000 μM. SDS-PAGE and Western blotting were used to analyse the samples. The HA1 protein was not

detected in the soluble fraction of the cells which indicates that the application of the lower concentrations of IPTG was not a successful method for soluble HA1 expression (Figure 3.10). The expression of an insoluble form of HA1 protein (Figure 3.11 A lane 5 and Figure 3.11 B lanes 1-5) indicates correct culturing conditions and successful induction with IPTG. The minimum level of IPTG required to induce the HA1 protein expression was 50 μM (lane 5 Figure 3.11 A).

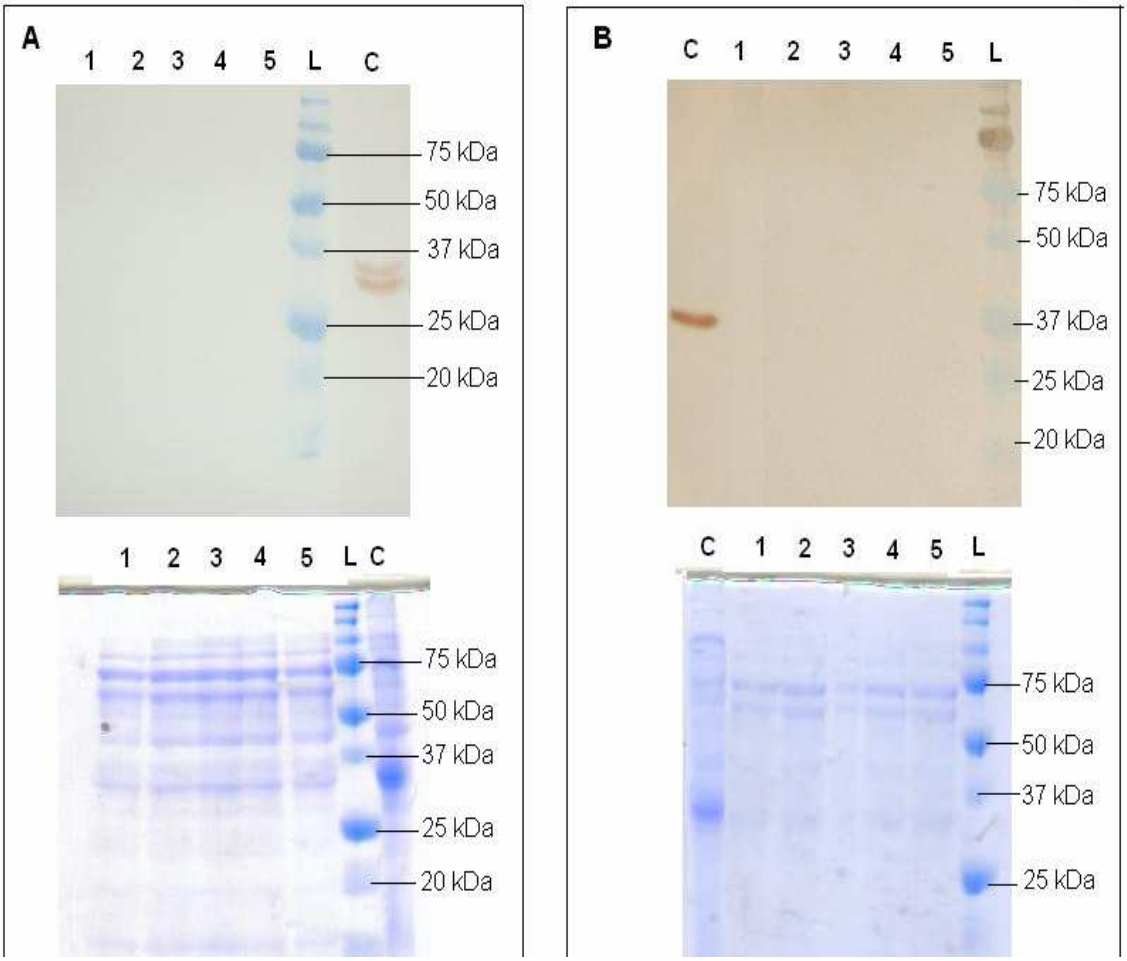


Figure 3.10. Expression of soluble HA1 protein using various concentrations of IPTG. SDS-PAGE analysis (bottom part of the images) and the corresponding Western blot analysis (top part of the images) of the soluble protein isolates. Cultures were induced with different concentration of IPTG at OD=0.6. The Western Blot was probed with anti-His antibodies. The concentration of IPTG used for induction was as follows:

A:
1 – 0
2 - 0.5 μM
3 - 1 μM
4 - 5 μM
5 - 50 μM
L - Bio-Rad prestained protein ladder.
C – positive control (insoluble fraction with HA1 protein)

B:
C – positive control (insoluble fraction with HA1 protein)
1 - 50 μM
2 - 100 μM
3 - 250 μM
4 - 500 μM
5 - 1000 μM
L - Bio-Rad prestained protein ladder

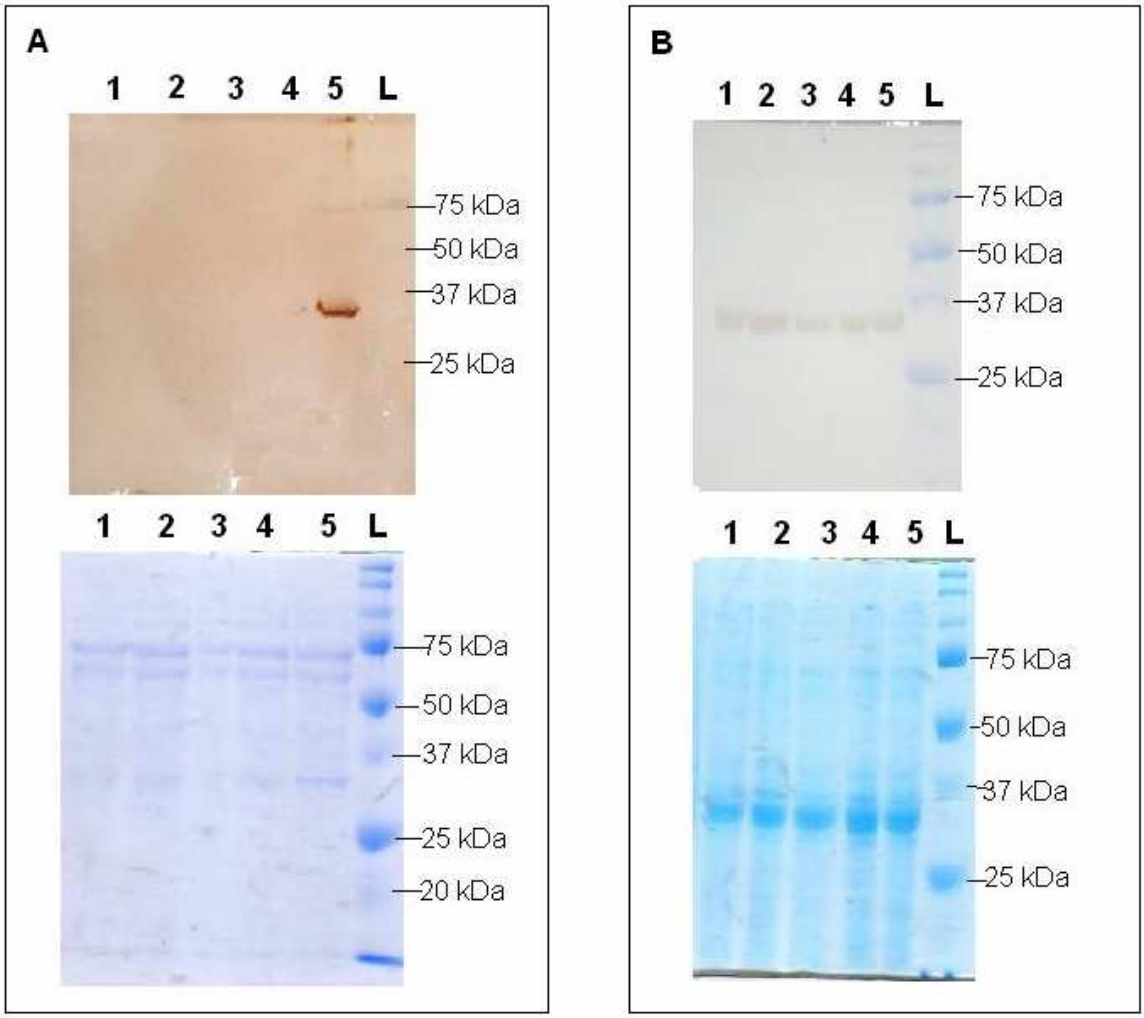


Figure 3.11. Expression of insoluble HA1 protein using various concentrations of IPTG.. SDS-PAGE analysis (bottom part of the images) and the corresponding Western blot analysis (top part of the images) of the insoluble protein isolates. Cultures were induced with different concentrations of IPTG at OD=0.6. The Western Blot was probed with anti-His antibodies. The concentrations of IPTG used for induction were as follows:

A:
1 – 0
2 - 0.5 μM
3 - 1 μM
4 - 5 μM
5 - 50 μM
L - Bio-Rad prestained protein ladder.

B:
1 - 50 μM
2 - 100 μM
3 - 250 μM
4 - 500 μM
5 - 1000 μM
L - Bio-Rad prestained protein ladder.

As the optimisation of IPTG concentration aimed at slowing down expression and promote soluble HA1 production failed, another method of expression down regulation was applied. Aggregation of heterologous proteins in recombinant systems is often minimized through the reduction of the growth temperature (Sørensen *et al.*, 2005) . In order to investigate the impact of lowered temperature on soluble HA1 expression the

cells were grown up to the density of OD = 0.6 at 37°C. Then, after 15 min adaptation period at 30°C, the expression was induced with 50 μM IPTG. The expression of HA1 protein at 30°C was monitored over a period of 5 hours. It has been found that the expression of HA1 is clearly detectable in all samples. There was no HA1 protein detected in the soluble fraction of the cells (Figure 3.12, part A); however, the HA1 protein was only present in the insoluble fraction of the cells (Figure 3.12, part B).

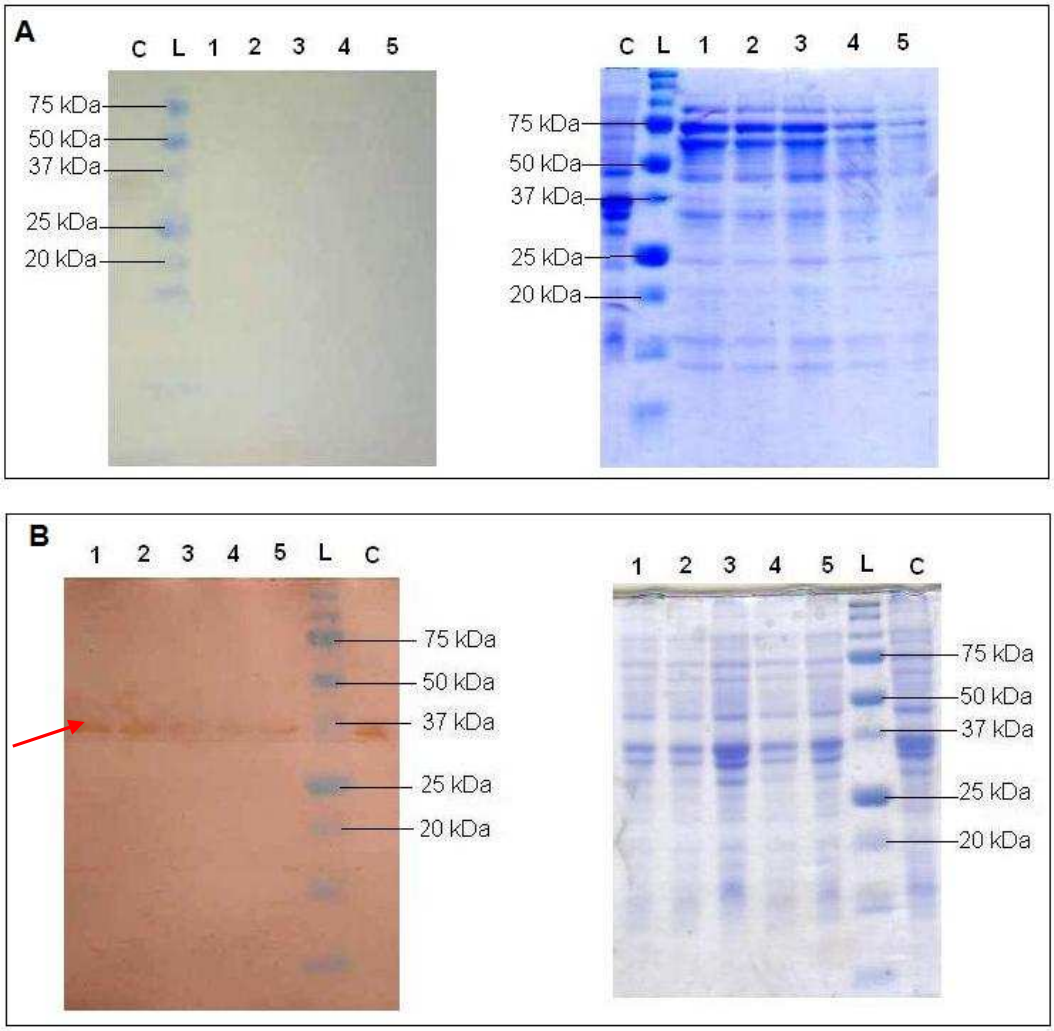


Figure 3.12. SDS-PAGE analysis and the corresponding Western blot analysis of the soluble (A) and insoluble (B) protein isolates from cultures grown at 30°C. Samples were collected over the period of 5 hours in one hour intervals. Hourly collected samples are presented in lanes 1-5. Number of the lane corresponds to the number of the hour after which the sample was collected. (L) Bio-Rad prestained protein ladder. (C) positive control (insoluble fraction with HA1 protein). The red arrow indicates the HA1 protein. Cultures were grown in 100ml LB amp broth. The Western Blot was probed with anti-His antibodies.

3.2.3 Two approaches to produce soluble HA1 protein

As the initial expression resulted in insoluble protein production, two separate approaches for the solubilisation of the expressed protein were undertaken. The first approach focused on cloning the *ha1* gene into various expression vectors in order to achieve the expression of the soluble HA1 protein. The second approach was aimed at extraction of the HA1 protein from the inclusion bodies, its further solubilisation by various denaturing agents followed by purification and finally, the refolding of the active HA1 protein.

3.2.3.1 Expression of the soluble HA1 protein in *E. coli*

There are several widely applied methods that facilitate expression of soluble heterologous proteins in *E. coli* (reviewed in Sørensen *et al.*, 2005). Several of them involve modification of the expression conditions such as the coexpression with molecular chaperones or the periplasmic targeting of the polypeptide (reviewed in Jonasson *et al.*, 2002). There are also methods that are aimed at the post- or co-translational alteration of the expressed protein, whereby the solubility is increased by fusion to another more soluble protein domain such as maltose binding protein (MBP), thioredoxin (Trx) and glutathione-S-transferase (GST) (Dyson *et al.*, 2004).

Periplasmic targeting of the HA1 protein

Secretion of heterologous proteins expressed in *E. coli* into the periplasm is a widely applied practice in molecular biology as it has several advantages over cytoplasmic expression (Soares *et al.*, 2003; Harvey *et al.*, 2004; Takayama and Akutsu, 2007). The most important feature of this periplasmic compartment**,** in terms of mammalian protein production, is its more oxidative environment that promotes disulphide bond formation and facilitates correct heterologous protein folding (Gasser *et al.*, 2008). Furthermore, due to the fact that the concentration of other proteins in the periplasm is much lower than that in the cytosol**,** the purification of overexpressed protein is simplified. There are three major pathways for the translocation of polypeptides across the cytoplasmic membrane into the periplasm of Gram-negative bacteria: (i) the Sec pathway (Fekkes *et al.*, 1999), (ii) the signal recognition particle (SRP) pathway (Koch *et al.*, 2003) and (iii) the Tat pathway (Fisher *et al.*, 2004). The Sec and SRP translocation systems are responsible for transmembrane transport of unfolded proteins whereas the Tat system is

used by folded proteins. The translocation of the protein through the inner membrane of *E. coli* cell can be either post-translational (Sec) or cotranslational (SRP) (Fekkes *et al.*, 1999; Koch *et al.*, 2003). Recognition of the protein targets by the Sec, SRP or Tat translocation system is facilitated by the specific N-terminal signal sequence of the nascently produced polypeptide (Fisher *et al.*, 2004). Whether a signal targets its protein to SRP or to SecA depends on the relative degree of hydrophobicity of the core region of the signal rather than its amino acid sequence (Steiner *et al.*, 2006). While the more hydrophobic protein signals associate with the SRP pathway, the less hydrophobic signals of secretory proteins utilize the SecA pathway (Fekkes *et al.*, 1999; Steiner *et al.*, 2006).

Utilization of the Sec systems for HA1 expression

The Sec based system was chosen to target the nascently produced, unfolded HA1 protein into the periplasmic space of *E. coli* cells. As the aggregation of HA1 protein expressed in *E. coli* is most likely caused by its misfolding, the periplasmic targeting has been applied in order to potentially produce the soluble, correctly folded HA1 protein. The *ha1* gene was cloned into the pQE-60-PelB expression vector in frame with an N-terminal PelB signal sequence (PelBss) recognized by the Sec system (Steiner *et al.*, 2006). The *ha1* gene was amplified by PCR. The primer's sequences contained a NotI restriction site (forward primer) and a BglII restriction site (reverse primer) to enable further directional cloning of the PCR product into the expression vector (Figure 3.13). Additional nucleotides were included at the 5' end of both primers to facilitate PCR product restriction with respective enzymes.

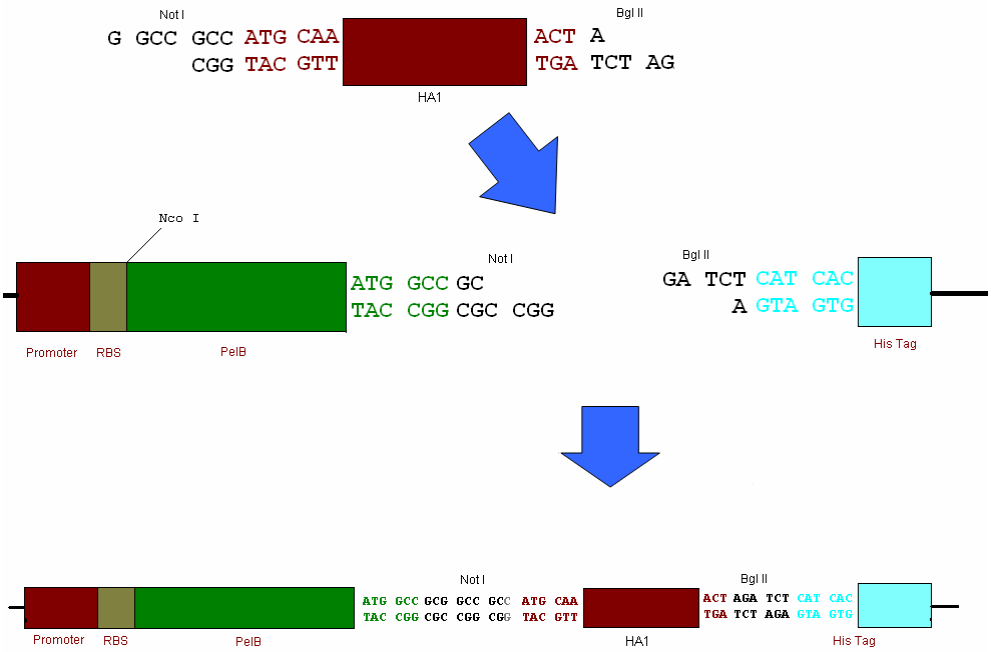


Figure 3.13. Ligation of *ha1* gene with the pQE-60-PelB plasmid. Relevant restriction sites are marked.

PCR reactions were optimized for the ha1 gene amplification with F1HA1NotI and R1HA1BglII primers

Table **3.2**) using Phusion DNA Polymerase. The optimised PCR annealing temperature was 60°C, annealing time 15 s and extension time 45 s.

	NotI
F1HA1NotI:	5′ ATCG GCG GCC GCC ATG CAA AAA CTT CCC GGA AAT GAC AAC AGC 3′
	BglII
R1HA1BglII:	5′ GAAGA TCT AGT TTG TTT CTC TGG TAC ATT CCG C 3′

Table 3.2. PCR primers used in the cloning of the *ha1* gene of the Influenza A virus A/Sichuan/2/87(H3N2)) into the pQE60-PelB vector. Restriction sites are highlighted.

The 3518 bp pQE60-PelB expression vector is derived from pQE-60 (Qiagen) which had been engineered to incorporate the PelB leader sequence (Figure 3.14) that directs expressed protein to the periplasm via the Sec system. The *ha1* gene was NotI and BglII restricted and ligated into the pQE60-PelB vector. To facilitate further purification of the HA1 protein the gene was cloned in frame with a vector borne 6x histidine tag. The selection of plasmids was facilitated by the β-lactamase gene and their replication by the ColE1 origin of replication. Protein expression was driven by the IPTG inducible T5 promoter. Ligation of the *ha1* gene with the pQE60-PelB vector resulted in formation of pJS103 plasmid (Figure 3.15). The sequence of the plasmid was verified by restriction

analysis and subsequent sequencing (MWG-Biotech Ltd) (not shown). Competent cells (*E. coli* XL10-Gold) were transformed with ligation mix and used to propagate the pJS103 plasmid which was subsequently isolated and verified by restriction analysis (not shown).

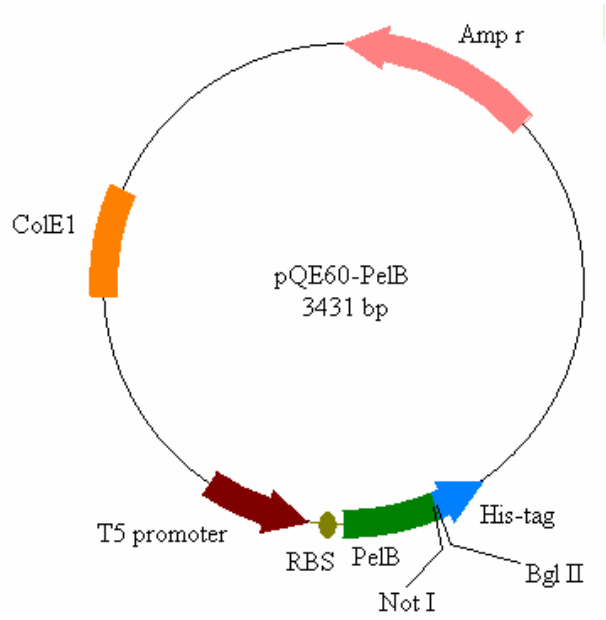


Figure 3.14. A map of the pQE60-PelB plasmid (Ryan Ph.D Thesis).

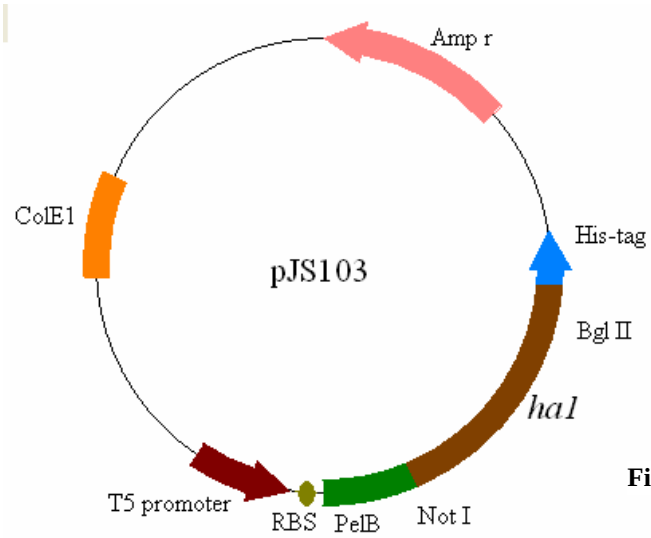


Figure 3.15. A map of the pJS103 plasmid.

Expression of the HA1 protein with the PelB signal sequence

The XL10-Gold strain of *E. coli* was chosen as a host for the HA1 protein expression, with the PelB signal sequence. The samples of both soluble and insoluble fractions isolated from the cells were analysed by SDS-PAGE and Western blotting (Figure 3.16 andFigure 3.17). No HA1 protein was detected in the soluble fraction of the cells (Figure 3.16 lane 2). However the PelB-HA1 fusion protein was detected in the insoluble fraction of the cells (Figure 3.17 lane 2). The shift in the position of PelB-HA1 fusion protein versus the control HA1 protein in the acrylamide gel can be attributed to the difference in calculated molecular weights of these two proteins: HA1 protein with PelB signal sequence - 40076.6 Da (although appears to be about 50 kDa on SDS-PAGE gels – Figure 3.17), HA1 without signal sequence - 38918.9 Da (although appears to be about 45 kDa on SDS-PAGE gels – Figure 3.17). This indicates that the HA1 protein was successfully expressed with a desired signal sequence. However it does not promote solubility of the protein.

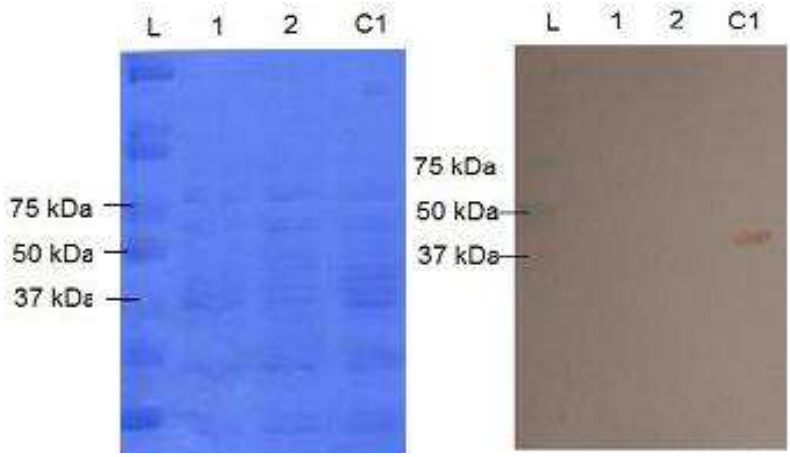


Figure 3.16. Soluble PelB-HA1 expression. SDS-PAGE analysis and (left) the corresponding Western blot analysis (right) of the soluble protein isolates. Lanes were loaded as follows: (1) – PelB-HA1 protein, uninduced culture; (2) – PelB-HA1 protein, induced culture; (C1) – control – HA1 protein from insoluble fraction; (L) Bio-Rad prestained protein ladder. Cultures were grown at 37°C in 100ml LB amp broth. Samples were collected after overnight culturing following induction with 50 μM IPTG. The Western Blot was probed with anti-His antibodies.

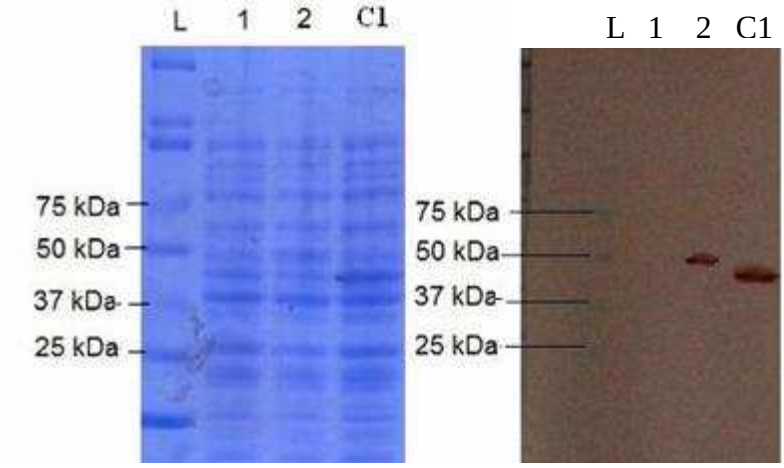


Figure 3.17. Insoluble PelB-HA1 expression. SDS-PAGE analysis (left) and the corresponding Western blot analysis (right) of the insoluble protein isolates. Lanes were loaded as follows: (1) – PelB-HA1 protein, uninduced culture; (2) – PelB-HA1 protein, induced culture; (C1) – control – HA1 protein from insoluble fraction; (L) Bio-Rad prestained protein ladder. Cultures were grown at 37°C in 100ml LB amp broth. Samples were collected after overnight culturing following induction with 50 μM IPTG. The Western Blot was probed with anti-His antibodies.

Utilization of the fusion proteins to enhance HA1 solubility

The application of fusion proteins for heterologous protein offers several advantages including simplified and efficient purification, increased expression rates and increased solubility of the product (Pryor *et al.*, 1997; Dyson *et al.*, 2004). Therefore, another approach for soluble HA1 protein expression represented co-expression with maltose binding protein (MBP). The tagging promotes solubility by the intrinsic soluble properties of the MBP. Furthermore, as the MBP is at the N-terminus of the construct it undergoes translation and correct folding before the HA1 protein is translated and thus additionally promotes proper folding of the HA1 fusion partner. Maltose binding protein was chosen as a fusion partner for the HA1 protein as it has consistently proven to be the most efficient solubility promoting tag for mammalian protein expression in *E. coli* (Dyson *et al.*, 2004).

Cloning of the HA1 gene into the pMALp2e expression vector

In order to express the MBP-fused HA1 protein, the *ha1* gene was cloned in frame with *malE* gene encoding MBP in the pMALp2E expression vector from NEB (Figure 3.18). The above vector was selected as it contains the *malE* gene and, furthermore, upstream from the malE gene there is a signal sequence encoded to direct the polypeptide to the periplasm. An ampicillin resistance gene encoded on the plasmid enables plasmid selection. The expression is facilitated by the vector-borne Ptac promoter. The replication of the vector is driven by the ColE1 origin of replication.

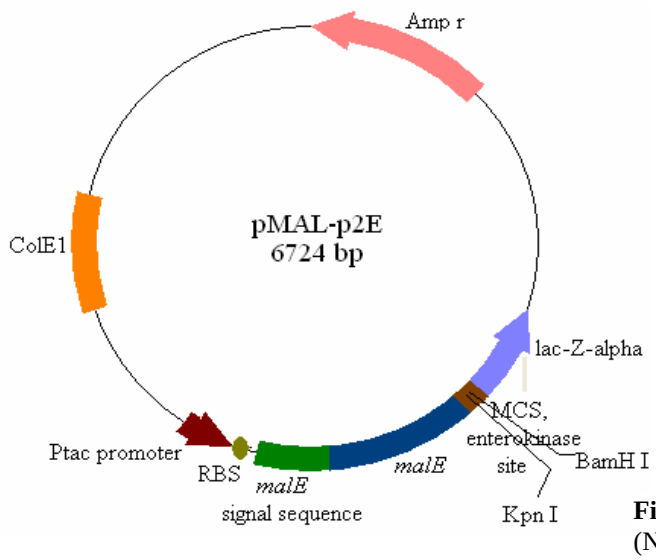


Figure 3.18. A map of the pMALp2E plasmid (NEB).

To clone the HA1 gene into the pMALp2E vector KpnI and BamHI sites were used. Primers that were used to amplify the *ha1* gene with KpnI and BamHI sites are presented in Figure 3.19. The PCR amplification with HA1KpnIF1 and HA1BamHIR1 primers and pJS102 as a template. The optimised annealing temperature was 65°C, annealing time 15 s and extension time 45 s.

KpnI

HA1KpnIF1: 5' GA CAG GTA CCG ATG CAA AAA CTT CCC GGA AAT GAC AAC AGC 3'

BamHI

HA1BamHIR1: 5' ACT GGA TCC TTA AGT TTG TTT CTC TGG TAC ATT CCG C 3'

Figure 3.19. PCR primers for the amplification of ha1 gene with KpnI and BamHI restriction sites. Restriction sites are highlighted and overhangs required by restriction enzymes are underlined.

The PCR product was extracted from the reaction mix and restricted with KpnI and BamHI enzymes. The pMALp2E vector was also restricted with KpnI and BamH1. The restricted fragments of the *ha1* gene and the vector were purified and subjected to overnight ligation. The ligation mix was later used to transform competent cells (*E. coli* XL10-Gold). No colonies were found on the plate where the transformation mix was spread which indicates that either ligation or transformation was unsuccessful. The experiment was repeated twice with same relt. Problems with ligation may be attributed to the distance between the KpnI site and the BamHI site in the pMALp2E vector which is only 7 bp and this may hamper the double restriction using these enzymes (Figure 3.20).

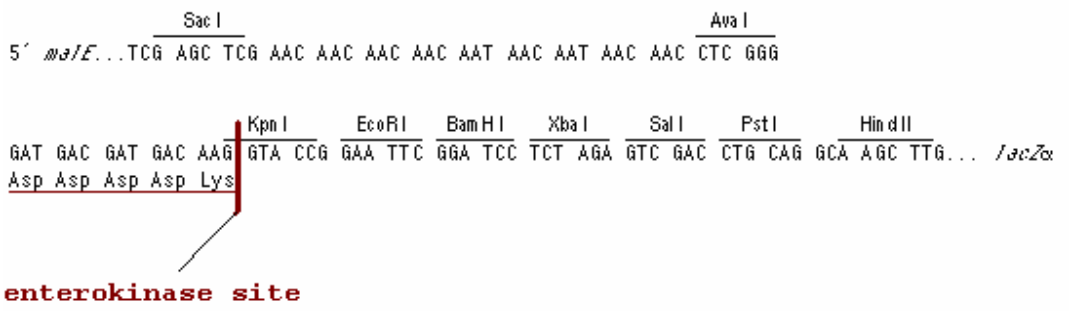


Figure 3.20. A fragment of pMALp2E (NEB) sequence containing the polylinker. Restriction sites are marked as well as an enterokinase site for further fusion tag cleavage.

Modification of the pMALp2E plasmid to facilitate KpnI/BglII double restriction

In order to facilitate successful restriction of the pMALp2E plasmid by both KpnI and BamHI enzymes the distance between KpnI and BamHI sites was increased by insertion of a DNA fragment in between them. The EcoRI site which is between KpnI and BamHI sites (Figure 3.20) was utilized. An EcoRI restricted random ~1kb DNA fragment was incorporated in the EcoRI site of the pMALp2E plasmid. The ligation mix was used to transform competent cells (*E. coli* XL10-Gold). The plasmid was subsequently isolated, verified by restriction analysis (results not shown) and named pMALp2Ef.

Cloning of the *ha1* gene into the modified pMALp2E vector

The *ha1* gene amplified by PCR with KpnI and BamH1 sites introduced in primer sequences as presented in Figure 3.19 was KpnI and BamH1 cloned into the pMALp2Ef vector. The DNA fragments of *ha1* gene and the vector were subjected to an overnight ligation and the ligation mix was used to transform competent cells (*E. coli* XL10-Gold). The plasmid was then isolated, verified by restriction analysis and sequencing (MWG) (results not shown) and named pJS104 (Figure 3.21).

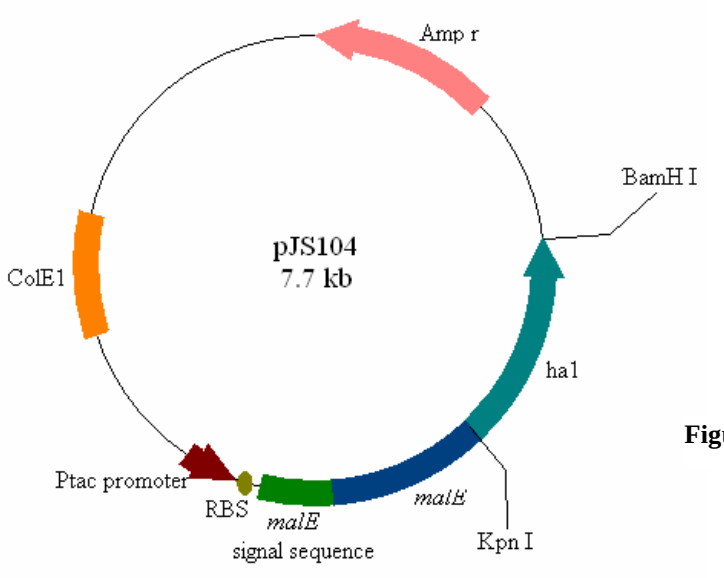


Figure 3.21. A map of the pJS104 plasmid.

Once the clone was confirmed by sequencing the expression was performed in the XL10-Gold strain of *E. coli* (using LB broth, induced with 0.1 mM IPTG). Samples of both soluble and insoluble fractions isolated from the cells were analysed by SDS-PAGE (Figure 3.22). Over-expression of the MBP-HA1 protein was not detected as there was no distinct band found on the SDS-PAGE in the lane representing induced culture at the expected hight of 60 kDa – 80 kDa (Figure 3.22 A lane 2).

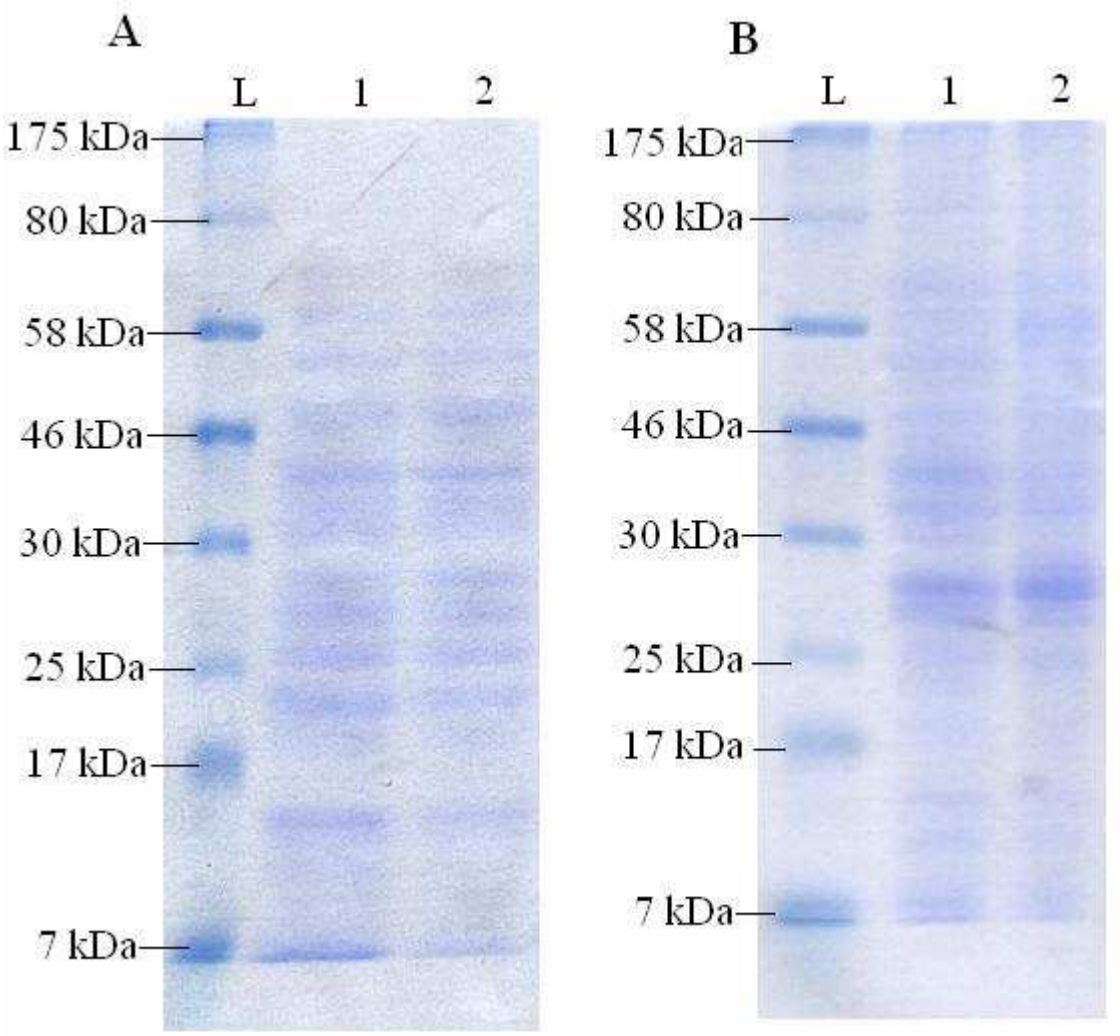


Figure 3.22. SDS PAGE analysis of the soluble (A) and insoluble (B) fractions of protein isolates from *E. coli* XL10-Gold pJS104 cultures. Lanes were loaded as follows: (L) – Wide Range Protein Marker (Sigma); (1) – uninduced culture; (2) – induced culture. Cultures were grown at 37°C in 100ml LB amp broth. Samples were collected after overnight culturing following induction with 50 μM IPTG.

3.2.3.2 Solublisation of the HA1 protein produced in the form of inclusion bodies

It was observed previously that the host cell had a great influence on the quality and quantity of the produced recombinant protein (reviewed in Vallejo *et al.* [2004]). The yields of soluble, active, mammalian (non-native) proteins expressed in *E. coli* are often highly affected by the lack of required post-translational modifications. Under such circumstances the protein of interest is often produced in the form of inclusion bodies. There are several advantages of inclusion bodies formation. Proteins that are normally toxic for bacterial cell; can be expressed and stored in an inactive form. Inclusion bodies deposition is also a way to avoid or reduce the rate of proteolytic degradation of the heterologous proteins. Furthermore, the concentration of an overexpressed protein in inclusion bodies exceeds significantly the concentration of protein that can be overexpressed in the cytosol. Finally, if it is possible to ultimately obtain an active product, the application of a prokaryotic expression system is much more cost effective as the microbial media are much cheaper than mammalian cell culture media, the time and equipment requirements are lower, and the volumetric yields are greatly higher (Vallejo *et al.*, 2004). However, the additional processing that is required to render such protein soluble and active is a drawback if the protein is produced in inclusion bodies. Nevertheless, successful solubilisation and refolding of such protein aggregates has been reported (Carrio and Villaverde, 2001). The initial step in inclusion bodies processing is solubilisation of the protein aggregates by use of chaotropic agents such as urea and guanidine HCl.

Urea treatment of inclusion bodies

The HA1 protein aggregates were treated with urea as this chaotropic agent is one of the most effective (Vallejo *et al.*, 2004) and is relatively cheap and available. The inclusion bodies from 100 ml culture of *E. coli* XL10-Gold pJS102 cells were isolated, washed, split into 3 samples and subjected to 8 M urea treatment for protein solubilisation (for one hour). Three different incubation temperatures were used, namely 4°C, 22°C and 37°C. The samples were then analysed by SDS-PAGE and Western blotting (Figure). The HA1 protein was detected after 8 M urea treatment using Western blotting (lanes 1, 3 and 5 in Figure). The position of the HA1 protein on the Western blot and the fact

that no fragmented proteins were detected indicates that the primary structure of the HA1 protein was intact during the urea treatment. However, the HA1 protein was not detected in the soluble fraction (Figure 3.23 lanes 2, 4 and 6). This indicates that solubilisation of HA1 with 8M urea under the applied conditions was unsuccessful.

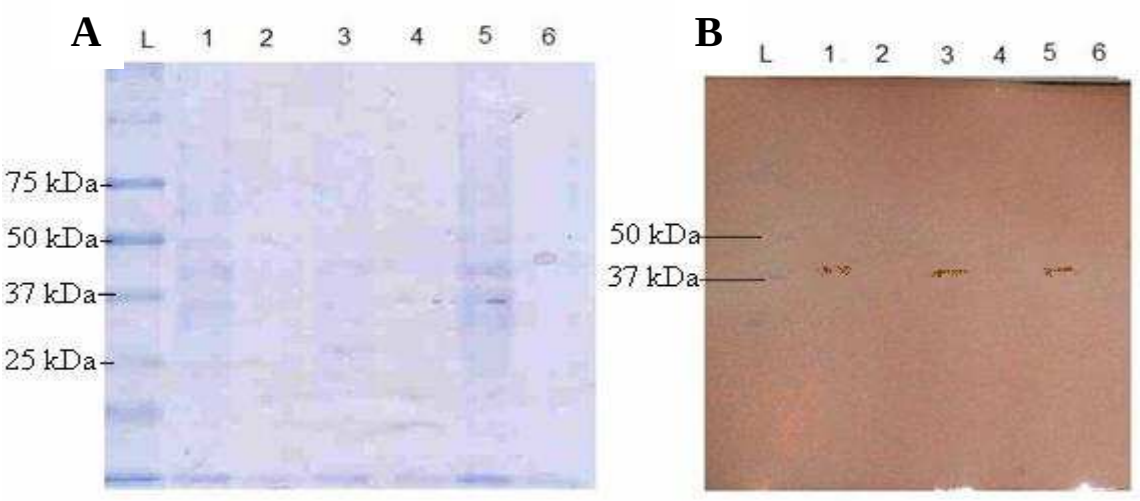


Figure 3.23. SDS-PAGE Commassie Blue stained (A) and corresponding Western blot (B) analysis of the HA1 protein samples subjected to solubilisation with 8 M urea at different temperatures. Lanes are as follows: (L) Bio-Rad prestained protein ladder; (1) - proteins incubated at 4°C, not centrifuged; (2) - soluble proteins after incubation at 4°C and centrifugation; (3) - proteins incubated at 22°C, not centrifuged; (4) - soluble proteins after incubation at 22°C and centrifugation; (5) - proteins incubated at 37°C, not centrifuged; (6) - soluble proteins after incubation at 37°C and centrifugation. The red arrow indicates the HA1 protein. The Western Blot was probed with anti-His antibodies.

Guanidine hydrochloride treatment of inclusion bodies

Following the unsuccessful solubilisation of the HA1 protein aggregates with urea, a more potent denaturant was applied. Although more expensive, guanidine hydrochloride (GnHCl) is superior to urea in terms of its chaotropic properties (Vallejo *et al.*, 2004). The inclusion bodies from 100 ml culture of pJS102 XL10-Gold culture of *E. coli* cells were isolated, washed and treated with 6 M GnHCl (for one hour). This concentration is the most widely used in such experiments (Vallejo *et al.*, 2004). Three different temperatures of incubation were used, namely 4°C, 22°C and 37°C. The samples were then analysed by SDS-PAGE and Western blotting (Figure 3.24). Although it was possible to detect the HA1 protein on the SDS-PAGE and Western blot (single bands on the gel indicates that the primary structure of the protein was intact and not trunkated) (Figure 3.24 lanes 1, 3 and 5) the HA1 protein was not detected in the soluble fraction (Figure 3.24 lanes 2, 4 and 6). This indicates that solubilisation of HA1 with 6 M GnHCl under the applied conditions was unsuccessful.

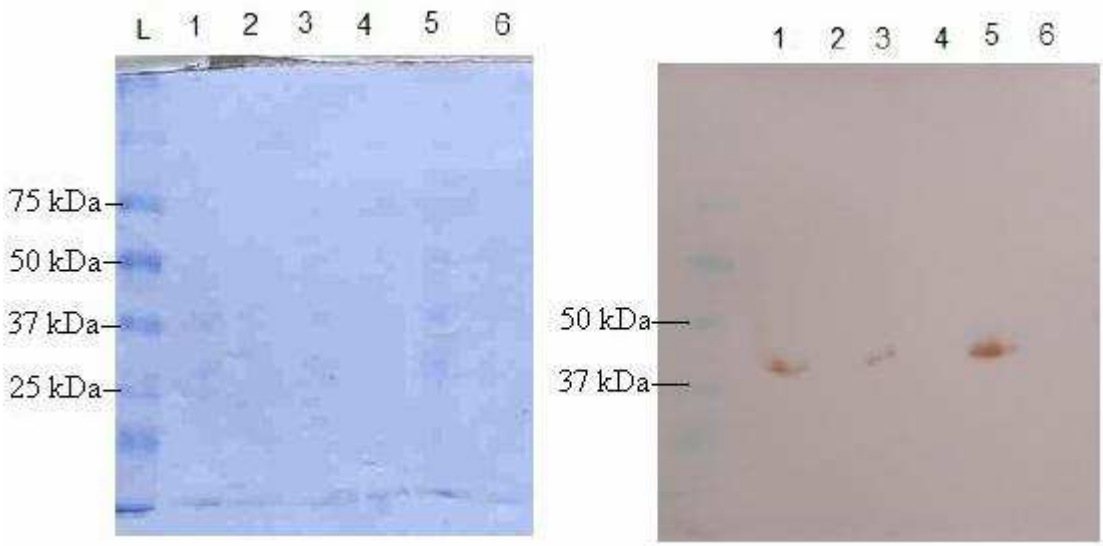


Figure 3.24. SDS-PAGE (left) and corresponding Western (right) blot analysis of the HA1 samples subjected to solubilisation with 6 M GnHCl at different temperatures. Lanes are as follows: (L) Bio-Rad prestained protein ladder; (1) - proteins incubated at 4°C, not centrifuged; (2) - soluble proteins after incubation at 4°C and centrifugation; (3) - proteins incubated at 22°C, not centrifuged; (4) - soluble proteins after incubation at 22°C and centrifugation;(5) - proteins incubated at 37°C, not centrifuged; (6) - soluble proteins after incubation at 37°C and centrifugation. The Western Blot was probed with anti-His antibodies.

In order to solubilise the HA1 protein a higher concentration of GnHCl was used, namely 8 M. After utilization of 8 M GnHCl the HA1 protein was detected by Western-blotting and SDS-PAGE Commassie Blue stained gel in the soluble fraction (Figure 3.25 lane 2).

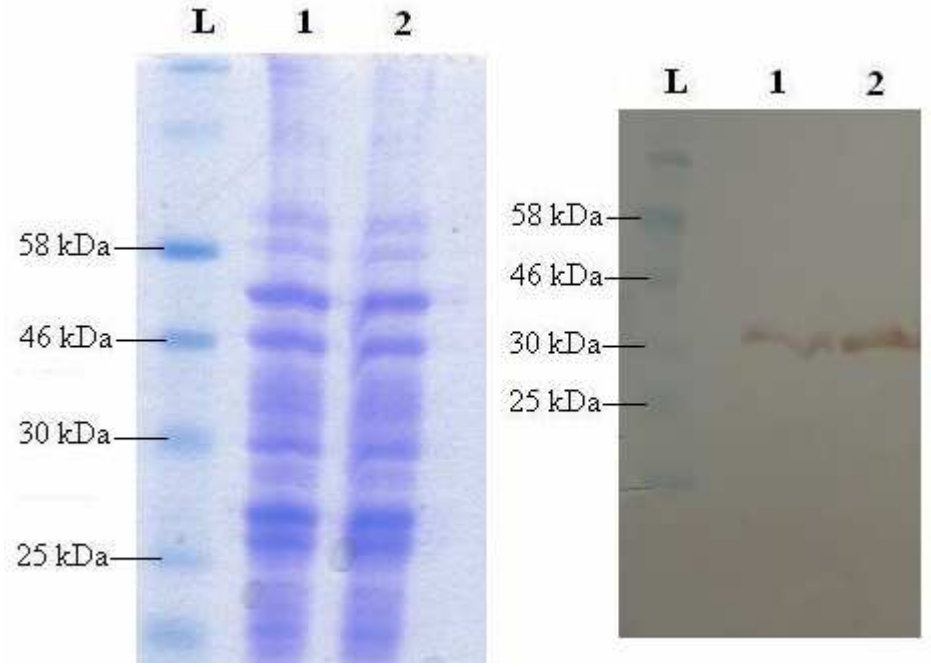


Figure 3.25. SDS-PAGE (left) and Westen-blot (right) analysis of the HA1 protein samples subjected to solubilisation with 8 M GnHCl at different temperatures. Lanes are as follows: (L) Bio-Rad prestained protein ladder;s (1) - protein sample not centrifuged; (2) - protein sample centrifuged. The Western Blot was probed with anti-His antibodies.

3.2.4 Purification of the HA1 protein under denaturing conditions

The ultimate goal of the utilization of inclusion bodies for protein production is to obtain a soluble and active final product. The protein aggregates are firstly solubilised and the protein of interest is finally refolded. However, it is essential that prior to the refolding process the polypeptide of interest is pure, as any contaminatants present in the sample often hamper proper protein folding (Middelberg, 2002). The HA1 protein (after solubilisation with 8 M GnHCl) was subjected to IMAC purification under denaturing conditions (8 M GnHCl). There was no HA1 protein detected in both wash and elution fractions which indicates that the HA1 protein did not bind the resin (Figure 3.26 lanes 3 and 4). It was concluded that the result of the HA1 protein purification under denaturing conditions (8 M GnHCl) does not work, which may be attributed to the fact that such a high concentration of GnHCl is not compatibile with the Ni-NTA resin (Qiagen).

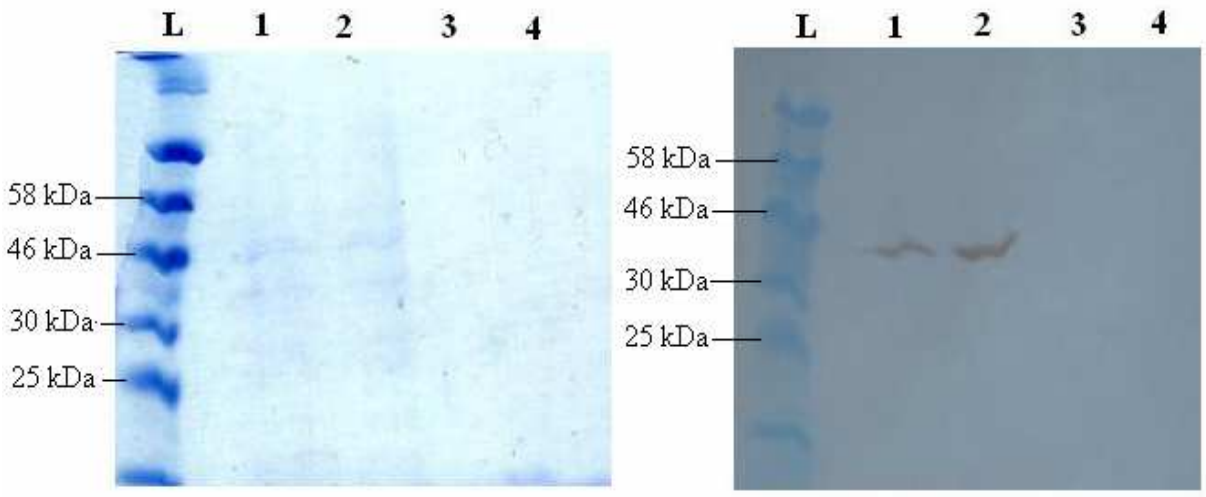


Figure 3.26. SDS-PAGE (left) and corresponding Western blotting (right) analysis of the HA1 sample subjected to purification under denaturing conditions (8 M GnHCl) at room temperature. Lanes were loaded as follows: (L) - Prestained Protein Marker, Broad Range (NEB); (1) – solubilised HA1 in 8 M GnHCl before application on the column; (2) – flow through; (3) – wash with 8 M GnHCl buffer; (4) – elution with 8 M GnHCl buffer containing 300 mM imidazole. Each sample comes from 10 ml fraction. The Western Blot was probed with anti-His antibodies.

3.3 Production, characterisation and application of SiaP protein for Sia detection and quantification

The goal of this section of the project was to produce the SiaP protein from *H. influenzae* utilizing an *E. coli*-based expression system and characterize the product with regards to molecular weight, purity and homogeneity. Results described in this section include also testing of the SiaP protein applicability for Sia detection and quantification.

3.3.1 Cloning of the *siaP* gene with a His-tag

The *siaP* gene was obtained on a plasmid from Dr. Emmanuele Severi (University of York, UK). The pET21b derived pGTY3 plasmid (Figure 3.27), has a STOP codon directly downstream from the *siaP* gene (Severi *et al.*, 2005) and a vector-borne STOP codon following the His-tag coding sequence.

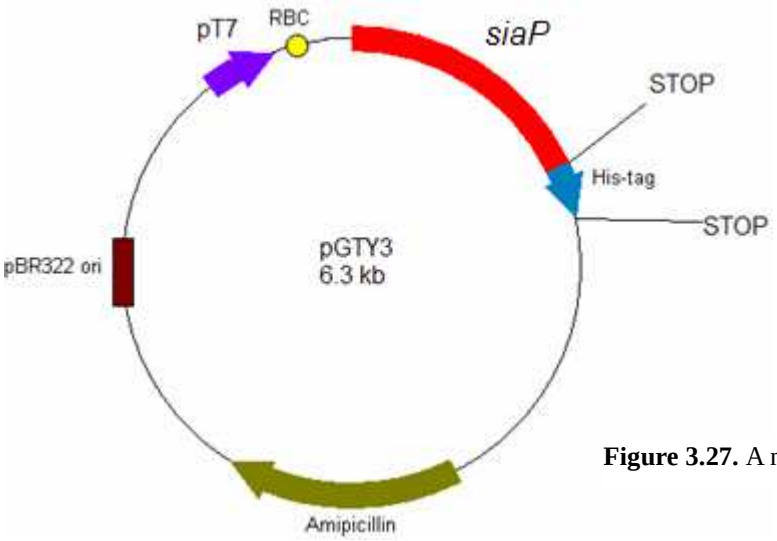


Figure 3.27. A map of the pGTY3 plasmid.

To facilitate further purification of the expressed SiaP protein the pGTY3 plamid was modified in order to utilize the 6xHis tag. The *siaP* native STOP codon was removed. Originally the *siaP* gene, including the signal sequence, was cloned into the pGTY3 vector using XhoI and NdeI restriction sites. Hence, to remove the STOP codon from the end of the *siaP* gene, this gene was PCR amplified without the STOP codon. Forward and reverse primers for PCR, containing NdeI and XhoI sites, respectively, are presented in Figure 3.28. The PCR amplification was performed using Phusion DNA

Polymerase. The optmised annealing temperature was 65°C, annealing time 15 s, extension time 45 s.

F1SiaPNdeI:	5′ GGGAATTCCATATGATGAAATTGACAAAACTTTTCC 3′ (NdeI)
R1SiaPXhoI:	5′ CCGCTCGAG TGG ATT GAT TGC TTC AAT TTG TTT TAA AGC 3′ (XhoI)

Figure 3.28. The primers used for PCR amplification of *siaP* gene with NdeI and XhoI sites.

After PCR the DNA fragments were extracted from the reaction mix (RBC gel extraction kit). Then the PCR products and the pGTY3 plasmid were restricted with NdeI and XhoI enzymes. The PCR products were extracted from the reaction mix (RBC gel extraction kit) and the vector was separated by gel electrophoresis in SYBR Safe agarose. Subsequently, the 5.3 kb DNA fragment corresponding to the empty pGTY3 vector was extracted from the electrophoresis gel (RBC gel extraction kit). Then the PCR products and the plasmid were ligated together. The ligation mix was used to transform *E. coli* XL10 Gold competent cells. Plasmids were then isolated and the sequence of the clone was confirmed by restriction analysis (results not shown) and sequencing (MWG; Figure 3.29).

Figure 3.29. Nucleotide sequence alignment of *siaP* gene from *Haemophilus influenzae* and the *siaP* gene present in the pJS201 plasmid sequenced by MWG. The sequence encoding the 6His-tag and STOP codon are marked in red.

The plasmid was named pJS201 (Figure 3.30). The *siaP* gene was placed in frame, directly upstream of the vector-borne His-tag which was followed by a vector borne STOP codon. The pJS201 plasmid includes a strong T7 promoter driving the expression of the *siaP* gene. An ampicillin resistance gene encoded on the plasmid enables plasmid selection and a pBR322-derived ColE1 origin of replication drives its replication.

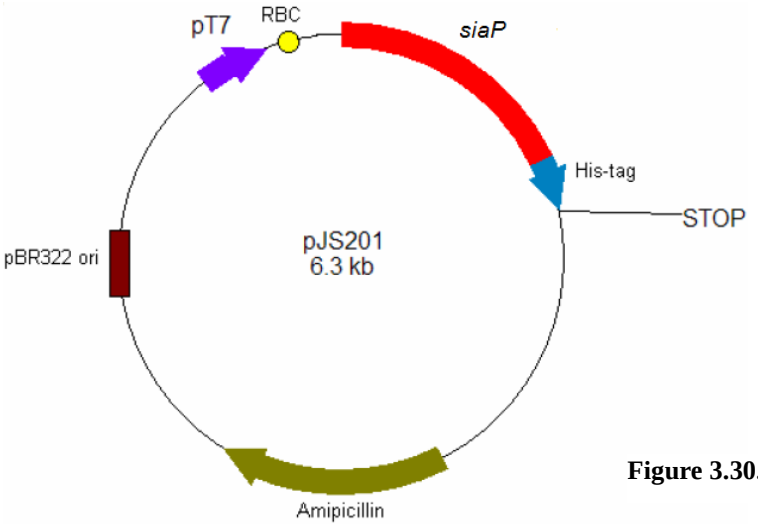


Figure 3.30. A map of the pJS201 plasmid.

3.3.2 Cloning of the *siaP* gene with a Strep2-tag

In order to facilitate orientated immobilization of the SiaP protein on various surfaces using amine coupling a C-terminal 6xLys-tagged derivative of SiaP was engineered. The string of six lysines (Figure 3.31) on the C-terminus increases the possibility for immobilization of the SiaP protein via lysine's amine group. C-terminal orientated immobilization is advantageous for SiaP application as a Sia sensor because the Sia-binding site is on the opposite side of the protein which increases the ability of immobilised protein to capture the ligand. Similarly, to facilitate further purification over streptactin-coated resins and orientated immobilisation of the SiaP protein on streptactin and analogous surfaces, a C-terminal Strep2-tagged derivative of SiaP was engineered. The Strep2-tag is a string of ten amino acids which shares high homology with biotin with regards to its recognition by streptavidin (Cai *et al.*, 2001) (Figure 3.32). Thus, the tag has high affinity towards streptavidin and other proteins with similar specificity. The tag sequence was introduced by PCR amplification of the entire pJS201 plasmid using phosphorylated primers. The optimised annealing temperature for the PCR was 60°C, annealing time 15 s and extension time 45 s. The primers used are presented in Figure 3.31 and Figure 3.32.

A 6Lys-tag:

Amino acid sequence: Lys Lys Lys Lys Lys Lys

Nuceotide sequence: 5′ CTC CTC CTC CTC CTC CTC 3′

B Primer sequences:

Forward primer (FSiaLYS): 5’ CTCCTCCTCTGAGATCCGGCTGCTAACAAAGCCC 3’

Reverse primer (RSiaPLYS): 5’ GAGGAGGAGGTGGTGGTGGTGGTGGTGCTCGAG 3’

Figure 3.31. Sequences of a 6Lys-tag and primers used to introduce the tag. (A)Amino acid and corresponding nucleotide sequences of the 6Lys-tag at the C-terminus of the SiaP protein. (B) Primer sequences used to amplify the pJS202 plasmid. The part of the 6Lys-tag introduced in the forward primer is marked in green and the part of the linker introduced in the reverse primer is marked in brown. The sequences shown in black are derived from the pJS201 plasmid used as a template.

A Strep2-tag:

Amino acid sequence: Ser Asn Trp Ser His Pro Gln Phe Glu Lys

Nuceotide sequence: 5’ AGT AAC TGG TCT CAC CCA CAA TTC GAG AAA 3’

B Primer sequences:

Forward primer (FSiaPCS): 5’ CCACAATTCGAGAAATGAGATCCGGCTGCTAACAAAGCCC 3’

Reverse primer (RSiaPCS): 5’ GTGAGACCAGTTACTGTGGTGGTGGTGGTGGTGCTCGAG 3’

Figure 3.32. Sequences of a Strep2-tag and primers used to introduce the tag. (A)Amino acid and corresponding nucleotide sequences of the Strep2-tag at the C-terminus of the SiaP protein. (B) Primer sequences used to amplify the pJS203 plasmid. The part of the Strep2-tag introduced in the forward primer is marked in green and the part of the linker introduced in the reverse primer is marked in brown. The sequences shown in black are derived from the pJS201 plasmid used as a template.

After PCR the DNA fragments were extracted from the reaction mix (RBC gel extraction kit) and subjected to DpnI digestion that cleaves only methylated DNA, and thus the original plasmid (pJS201 used as a template) was eliminated. Then the products were separated by gel electrophoresis in SYBR Safe agarose, gel extracted and subjected to ligation with T4 DNA ligase. Subsequently the ligation mixes were used to transform *E. coli* KRX competent cells. The plasmids were named pJS202 (Figure 3.33) and pJS203 (Figure 3.34). Isolated plasmids were confirmed by restriction analysis (not shown) and sequencing (MWG; Figure 3.35 and Figure 3.36).

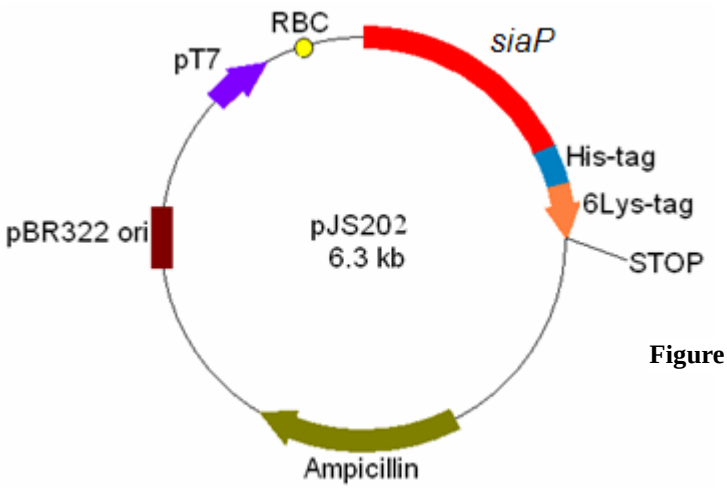


Figure 3.33. A map of the pJS202 plasmid.

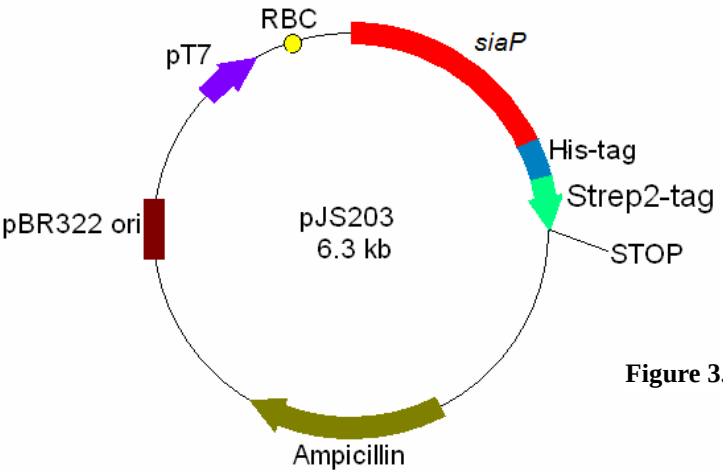


Figure 3.34. A map of the pJS203 plasmid.

Figure 3.35. Nucleotide sequence alignment of the *siaP* gene from pJS201 plasmid and the *siaP* gene present in the pJS202 plasmid sequenced by MWG. The sequence coding for 6Lys-tag and the STOP codon are marked in red.

Figure 3.36. Nucleotide sequence alignment of the *siaP* gene from pJS201 plasmid and the *siaP* gene present in the pJS203 plasmid sequenced by MWG. The sequence coding for the Strep2-tag and STOP codon are marked in red.

3.3.3 Expression of the His-tagged SiaP protein

The expression of the His-tagged SiaP protein (SiaP 201) was optimized in terms of expression time, concentration of inducing agent and culture cell density at the induction time. As the transcription of the *siaP* gene in the pJS201 plasmid is driven by the T7 promoter the vector was transformed into the KRX strain of *E.coli* (Promega). The genotype of the KRX strain contains a T7 polymerase gene under the control of a rhmnose promoter which prevents "leaky" expression of the target protein (Promega).

Optimisation of the harvesting time and cell concentration at the point of induction

In order to establish an optimal post-induction harvesting time as well as optimise the cell concentration during the induction that leads to highest yield of SiaP protein, the cultures of *E. coli* KRX pJS201 were induced at various stages of growth and sampled every hour afterwards. The cultures were incubated at 37°C and induced with 0.1% rhamnose. The optical densities when the induction took place were 0.5 (midlog phase), 1.0 and 1.2 (staionary phase) respectively. Both soluble (Figure 3.37, 3.38 and 3.39) and insoluble (Figure 3.40, 3.41 and 3.42) fractions of the cells were analysed over time. Overexpression of the SiaP protein was visualized by SDS-PAGE. The concentration of expressed SiaP protein was increasing over time in the soluble fraction and reached its maximum after 6 hours following induction when induced at OD_{600}=0.5 (Figure 3.37) and after 5 hours following induction when induced at OD_{600}=1.0 and OD_{600}=1.2 (Figure 3.38 and 3.39 respectively). The concentration of soluble SiaP protein in the cells after overnight culturing was lower then after 5 or 6 hours following induction (Figure 3.37, 3.38 and 3.39). The concentration of SiaP in insoluble fractions was also increasing over time, it was found to be highest in the samples taken after overnight culturing regardless of the OD during the induction and did not decrease overnight (Figure 3.40, 3.41 and 3.42). The concentration of soluble protein was found to be highest when subjected to induction at OD_{600}=1.2 and harvested after 6 hours (Figure 3.37, 3.38 and 3.39). Conversely, the concentration of insoluble SiaP protein was higher in cultures induced in the earlier stage of culturing. The highest concentration of insoluble SiaP protein was found in cultures induced at OD_{600}=0.5 and harvested after 6 hours (Figure 3.40, 3.41 and 3.42).

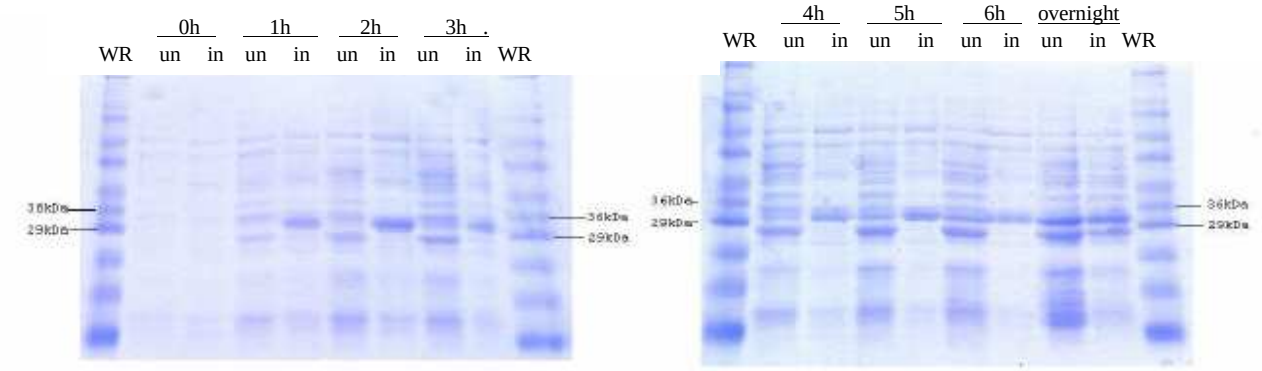


Figure 3.37. SDS-PAGE gels showing soluble fractions from cultures induced at OD=0.5. Samples from uninduced (un) cultures presented along with samples from induced (in) cultures. Samples were collected every hour for 6 hours (1h, 2h, … , 6h) and the last sample was collected after an overnight culturing. WR-Sigma Wide Range protein ladder. The predicted molecular weight of SiaP is 35.5 kDa.

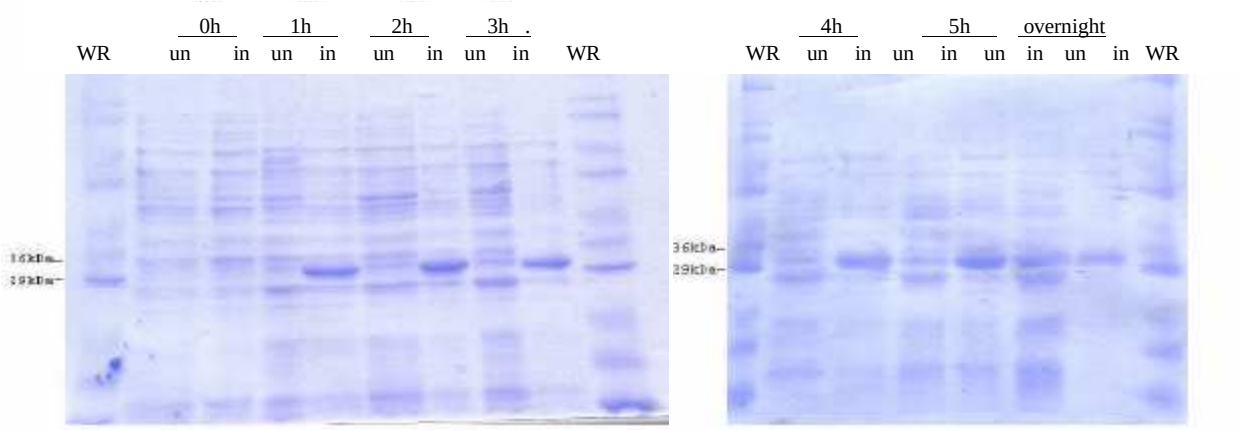


Figure 3.38. SDS-PAGE gels showing soluble fractions from cultures induced at OD=1.0. Samples from uninduced (un) cultures presented along with samples from induced (in) cultures. Samples were collected every hour for 5 hours (1h, 2h, … , 5h) and the last sample was collected after an overnight culturing. WR-Sigma Wide Range protein ladder. The predicted molecular weight of SiaP is 35.5 kDa.

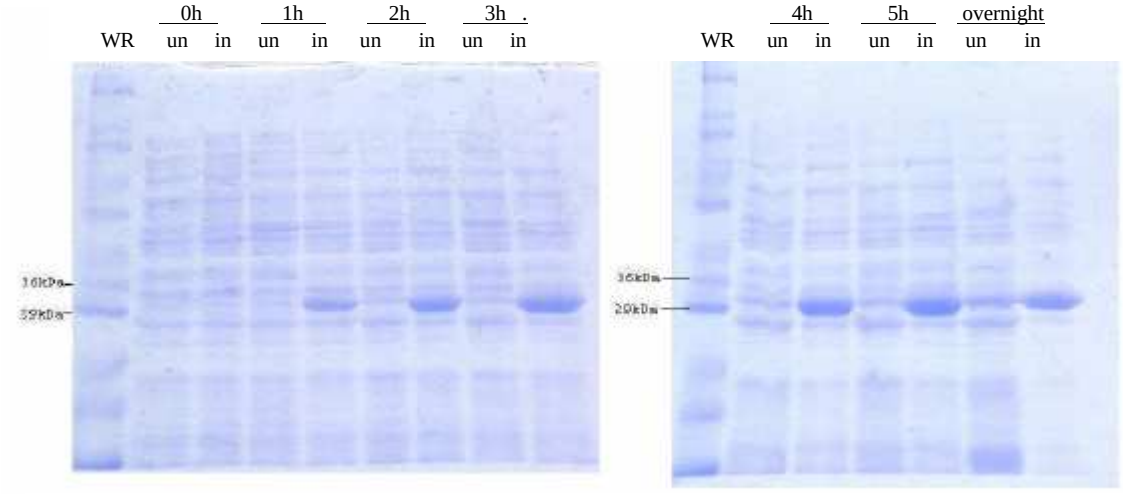


Figure 3.39. SDS-PAGE gels showing soluble fractions from cultures induced at OD=1.2. Samples from uninduced (un) cultures presented along with samples from induced (in) cultures. Samples were collected every hour for 5 hours (1h, 2h, … , 5h) and the last sample was collected after an overnight culturing. WR-Sigma Wide Range protein ladder. The predicted molecular weight of SiaP is 35.5 kDa.

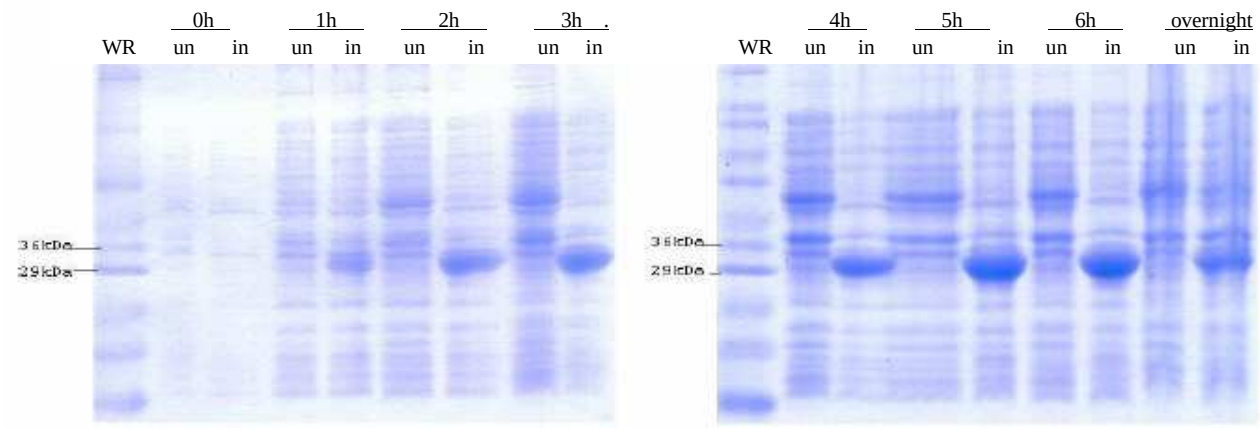


Figure 3.40. SDS-PAGE gels showing insoluble fractions from cultures induced at OD=0.5. Samples from uninduced (un) cultures presented along with samples from induced (in) cultures. Samples were collected every hour for 6 hours (1h, 2h, … , 6h) and the last sample was collected after an overnight culturing. WR-Sigma Wide Range protein ladder. The predicted molecular weight of SiaP is 35.5 kDa.

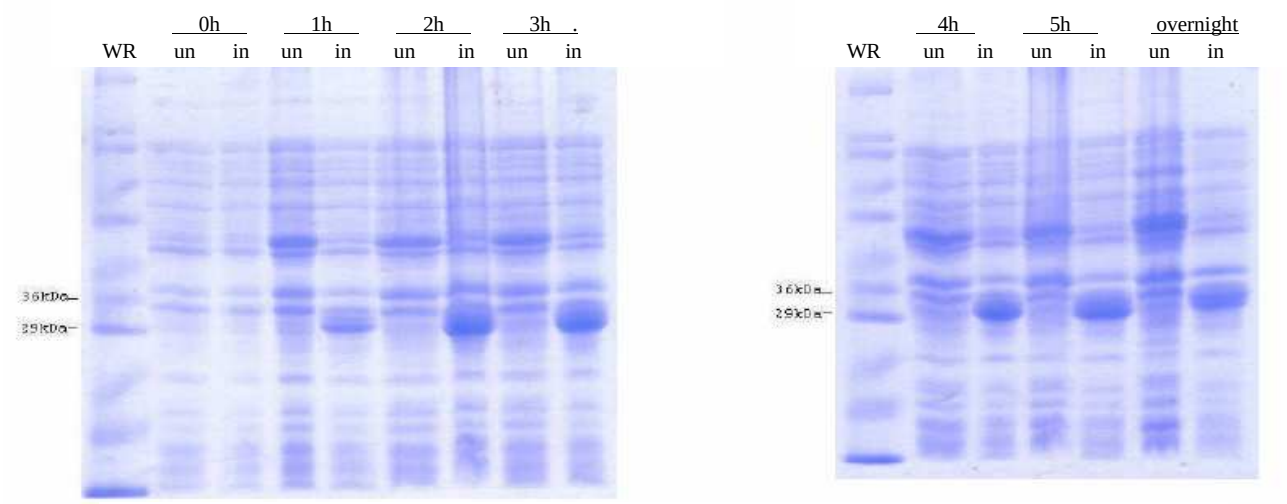


Figure 3.41. SDS-PAGE gels showing insoluble fractions from cultures induced at OD=1.0. Samples from uninduced (un) cultures presented along with samples from induced (in) cultures. Samples were collected every hour for 5 hours (1h, 2h, … , 5h) and the last sample was collected after an overnight culturing. WR-Sigma Wide Range protein ladder. The predicted molecular weight of SiaP is 35.5 kDa.

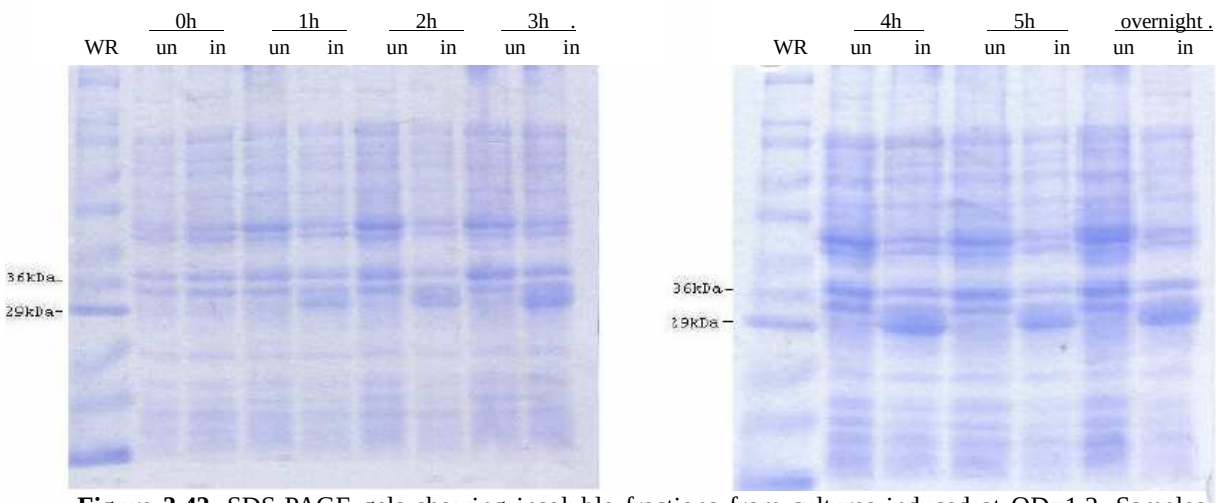


Figure 3.42. SDS-PAGE gels showing insoluble fractions from cultures induced at OD=1.2. Samples from uninduced (un) cultures presented along with samples from induced (in) cultures. Samples were collected every hour for 5 hours (1h, 2h, … , 5h) and the last sample was collected after an overnight culturing. WR-Sigma Wide Range protein ladder. The predicted molecular weight of SiaP is 35.5 kDa.

Investigation of the optimal concentration of the inducing agent

To investigate the impact of the inducing agent (rhamnose) concentration on the SiaP protein expression the cultures were incubated at 37°C and induced at OD=0.5. The final concentrations of rhamnose used to induce expression were 0.001%, 0.005%, 0.01%, 0.05%, 0.1%, 0.2% and 0.5% (w/v) respectively. Both soluble (Figure 3.43 - 3.48) and insoluble (Figure 3.49 – 3.54) fractions of the cells were analysed over time. Higher concentration of rhamnose resulted in similar levels of SiaP concentration as 0.01% rhamnose in samples taken one and two hours after induction (Figure 3.43 and 3.44). However, in samples taken in later stages of culturing the concentration of SiaP was lower if induced with more than 0.01% rhamnose (Figure 3.45 – 3.48). This indicates that the optimal concentration of rhamnose for the induction of soluble expression of the SiaP protein is 0.01%.

In the insoluble fractions the concentration of SiaP protein was constant if induced with rhamnose in the concentration range between 0.01% and 0.5% (Figure 3.49 - 3.54). The insoluble SiaP protein was also found in samples taken from cultures induced with 0.001% and 0.005%, albeit, at lower concentrations than in cultures induced with higher levels of rhamnose (Figure 3.53).

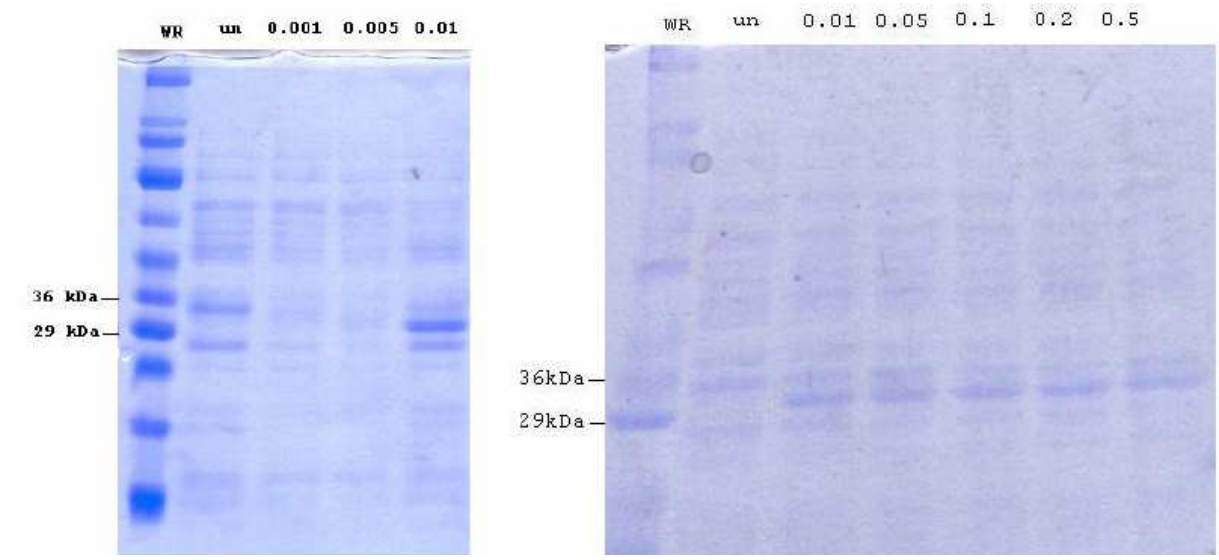


Figure 3.43. SDS-PAGE analysis of soluble protein fraction isolated from samples taken one hour after induction with different rhamnose concentrations. WR –Sigma Wide Range protein ladder; un – uninduced; 0.001-culture induced with 0.001% rhamnose; 0.005- culture induced with 0.005% rhamnose; 0.01-culture induced with 0.01% rhamnose; 0.05-culture induced with 0.05% rhamnose; 0.1-culture induced with 0.1% rhamnose; 0.2- culture induced with 0.2% rhamnose; 0.5-culture induced with 0.5% rhamnose. The predicted molecular weight of SiaP is 35.5 kDa.

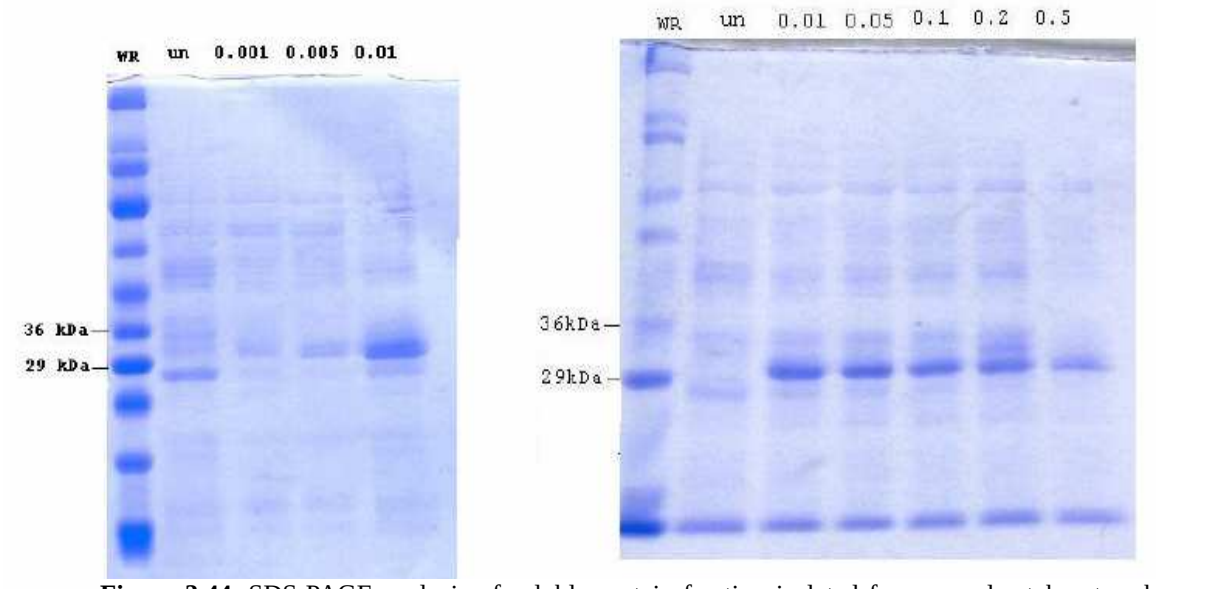


Figure 3.44. SDS-PAGE analysis of soluble protein fraction isolated from samples taken two hours after induction with different rhamnose concentrations. WR –Sigma Wide Range protein ladder; un – uninduced; 0.001-culture induced with 0.001% rhamnose; 0.005- culture induced with 0.005% rhamnose; 0.01-culture induced with 0.01% rhamnose; 0.05-culture induced with 0.05% rhamnose; 0.1-culture induced with 0.1% rhamnose; 0.2- culture induced with 0.2% rhamnose; 0.5-culture induced with 0.5% rhamnose. The predicted molecular weight of SiaP is 35.5 kDa.

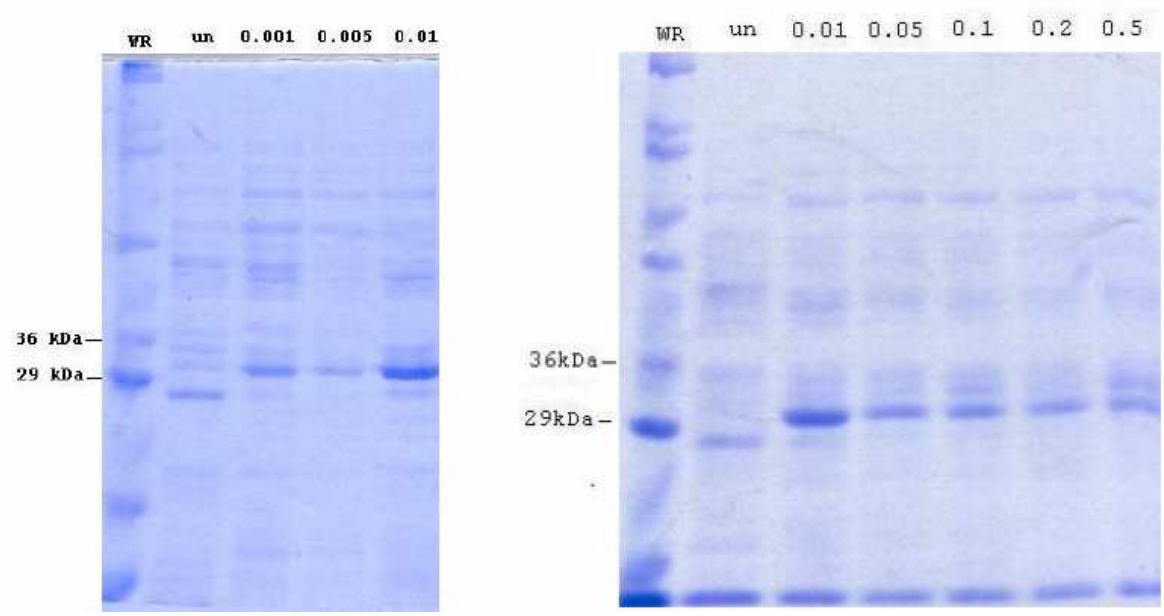


Figure 3.45. SDS-PAGE analysis of soluble protein fraction isolated from samples taken three hours after induction with different rhamnose concentrations. WR –Sigma Wide Range protein ladder; un – uninduced; 0.001-culture induced with 0.001% rhamnose; 0.005- culture induced with 0.005% rhamnose; 0.01-culture induced with 0.01% rhamnose; 0.05-culture induced with 0.05% rhamnose; 0.1-culture induced with 0.1% rhamnose; 0.2- culture induced with 0.2% rhamnose; 0.5-culture induced with 0.5% rhamnose. The predicted molecular weight of SiaP is 35.5 kDa.

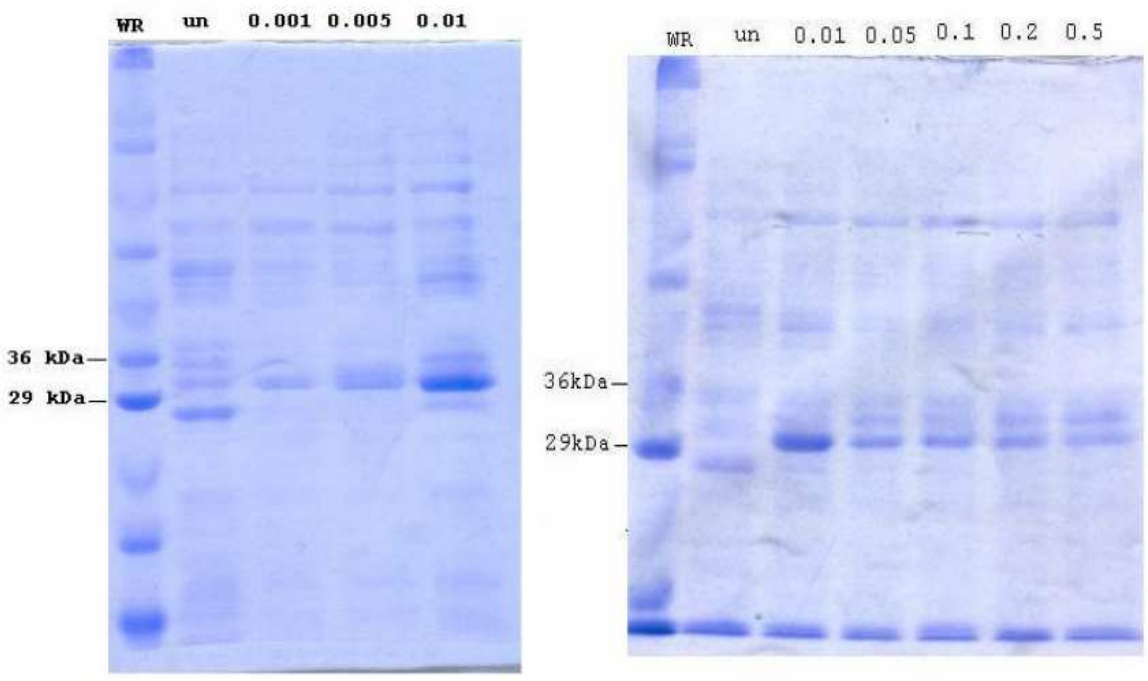


Figure 3.46. SDS-PAGE analysis of soluble protein fraction isolated from samples taken four hours after induction with different rhamnose concentrations. WR –Sigma Wide Range protein ladder; un – uninduced; 0.001-culture induced with 0.001% rhamnose; 0.005- culture induced with 0.005% rhamnose; 0.01-culture induced with 0.01% rhamnose; 0.05-culture induced with 0.05% rhamnose; 0.1-culture induced with 0.1% rhamnose; 0.2- culture induced with 0.2% rhamnose; 0.5-culture induced with 0.5% rhamnose. The predicted molecular weight of SiaP is 35.5 kDa.

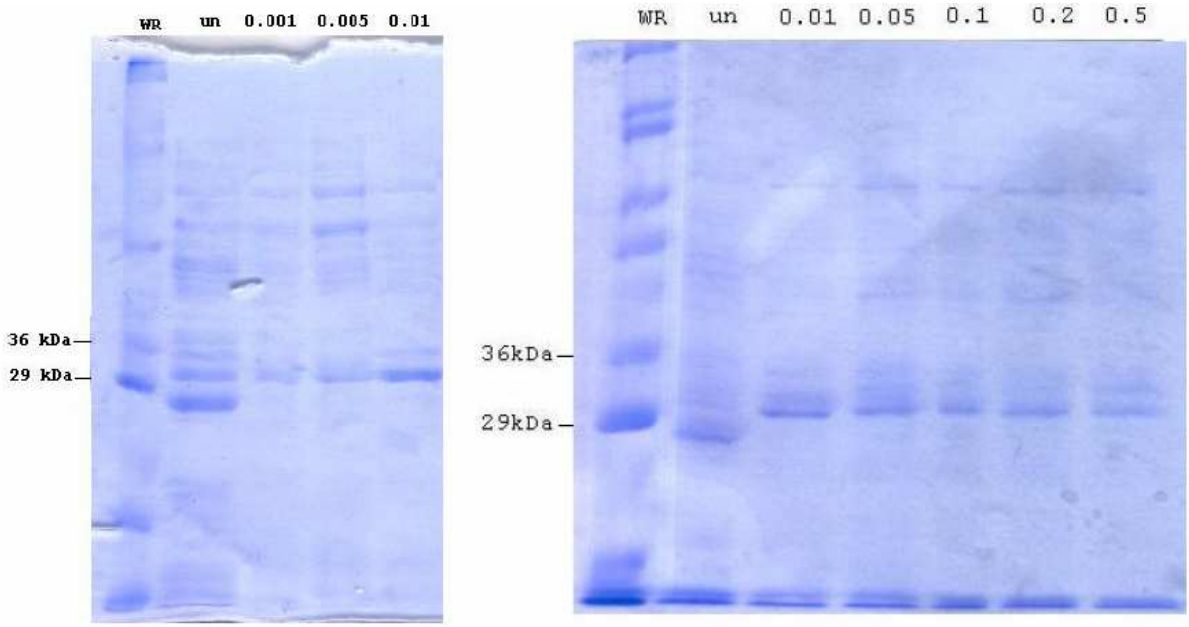


Figure 3.47. SDS-PAGE analysis of soluble protein fraction isolated from samples taken five hours after induction with different rhamnose concentrations. WR –Sigma Wide Range protein ladder; un – uninduced; 0.001-culture induced with 0.001% rhamnose; 0.005- culture induced with 0.005% rhamnose; 0.01-culture induced with 0.01% rhamnose; 0.05-culture induced with 0.05% rhamnose; 0.1-culture induced with 0.1% rhamnose; 0.2- culture induced with 0.2% rhamnose; 0.5-culture induced with 0.5% rhamnose. The predicted molecular weight of SiaP is 35.5 kDa.

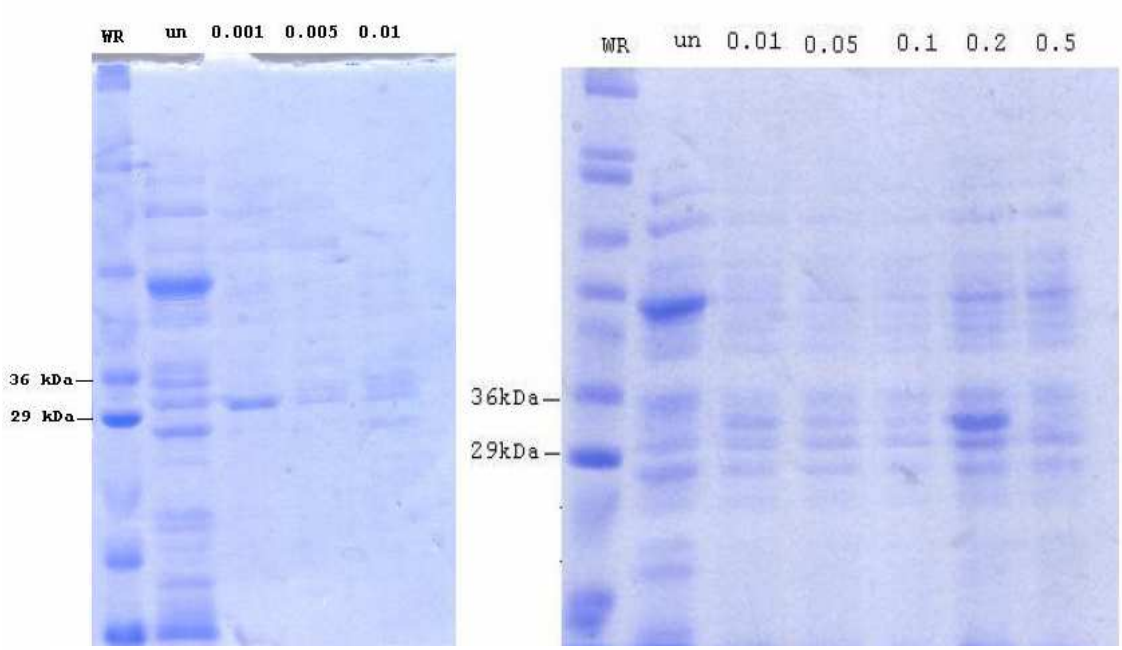


Figure 3.48. SDS-PAGE analysis of soluble protein fraction isolated from samples taken after overnight culturing following induction with different concentrations of rhamnose. WR –Sigma Wide Range protein ladder; un – uninduced; 0.001-culture induced with 0.001% rhamnose; 0.005-culture induced with 0.005% rhamnose; 0.01-culture induced with 0.01% rhamnose; 0.05-culture induced with 0.05% rhamnose; 0.1-culture induced with 0.1% rhamnose; 0.2- culture induced with 0.2% rhamnose; 0.5-culture induced with 0.5% rhamnose. The predicted molecular weight of SiaP is 35.5 kDa.

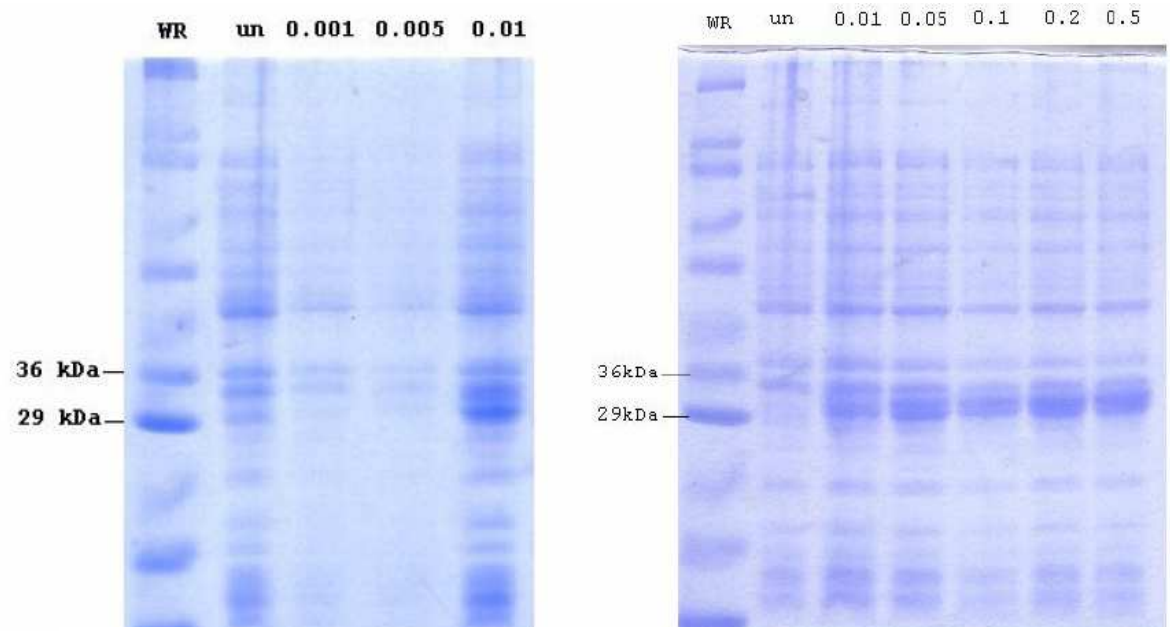


Figure 3.49. SDS-PAGE analysis of insoluble protein fraction isolated from samples taken one hour after induction with different rhamnose concentrations. WR –Sigma Wide Range protein ladder; un – uninduced; 0.001-culture induced with 0.001% rhamnose; 0.005- culture induced with 0.005% rhamnose; 0.01-culture induced with 0.01% rhamnose; 0.05-culture induced with 0.05% rhamnose; 0.1-culture induced with 0.1% rhamnose; 0.2- culture induced with 0.2% rhamnose; 0.5-culture induced with 0.5% rhamnose. The predicted molecular weight of SiaP is 35.5 kDa.

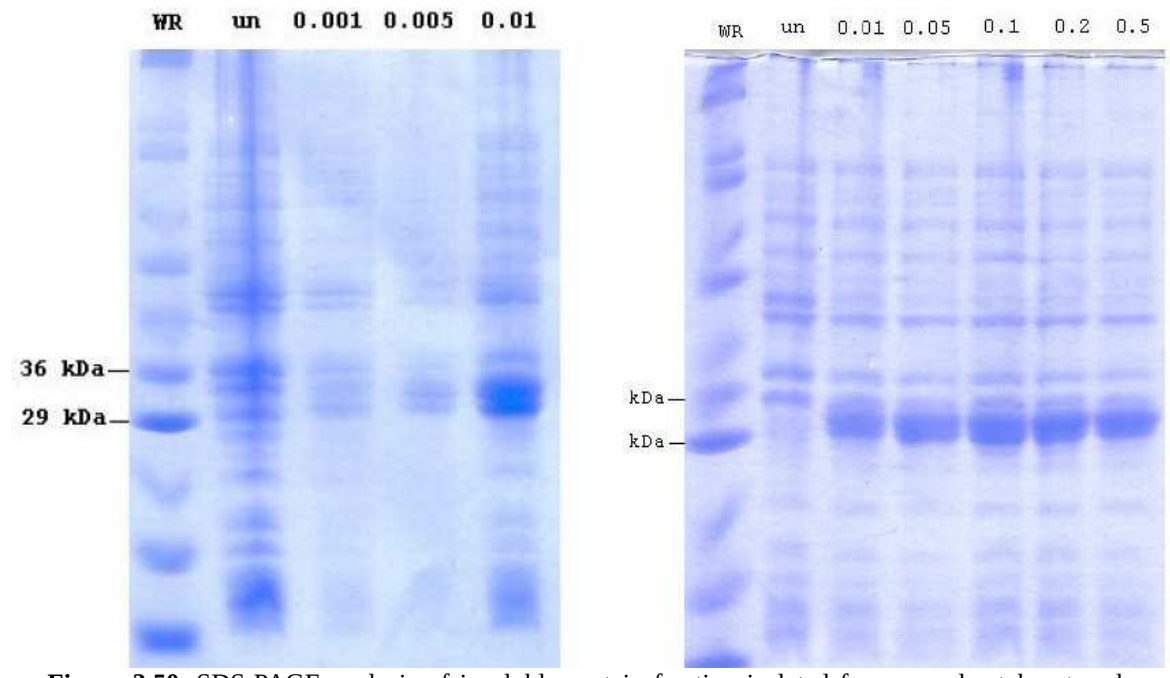


Figure 3.50. SDS-PAGE analysis of insoluble protein fraction isolated from samples taken two hours after induction with different rhamnose concentrations. WR –Sigma Wide Range protein ladder; un – uninduced; 0.001-culture induced with 0.001% rhamnose; 0.005- culture induced with 0.005% rhamnose; 0.01-culture induced with 0.01% rhamnose; 0.05-culture induced with 0.05% rhamnose; 0.1-culture induced with 0.1% rhamnose; 0.2- culture induced with 0.2% rhamnose; 0.5-culture induced with 0.5% rhamnose. The predicted molecular weight of SiaP is 35.5 kDa.

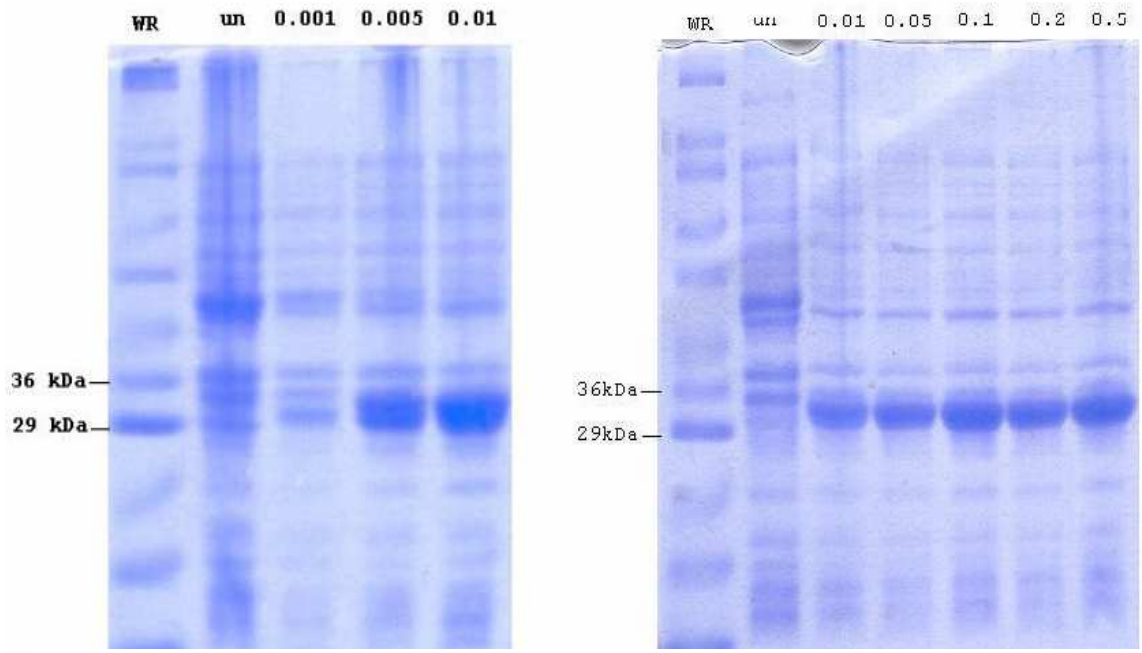


Figure 3.51. SDS-PAGE analysis of insoluble protein fraction isolated from samples taken three hours after induction with different rhamnose concentrations. WR –Sigma Wide Range protein ladder; un – uninduced; 0.001-culture induced with 0.001% rhamnose; 0.005- culture induced with 0.005% rhamnose; 0.01-culture induced with 0.01% rhamnose; 0.05-culture induced with 0.05% rhamnose; 0.1-culture induced with 0.1% rhamnose; 0.2- culture induced with 0.2% rhamnose; 0.5-culture induced with 0.5% rhamnose. The predicted molecular weight of SiaP is 35.5 kDa.

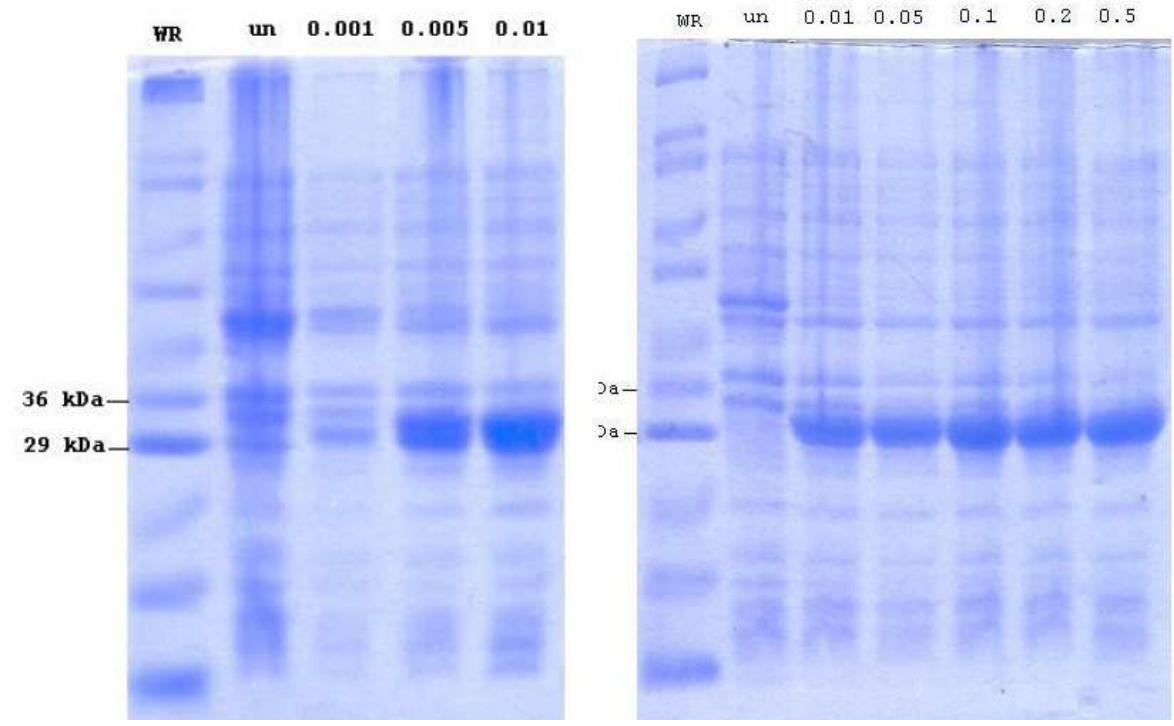


Figure 3.52. SDS-PAGE analysis of insoluble protein fraction isolated from samples taken four hours after induction with different rhamnose concentrations. WR –Sigma Wide Range protein ladder; un – uninduced; 0.001-culture induced with 0.001% rhamnose; 0.005- culture induced with 0.005% rhamnose; 0.01-culture induced with 0.01% rhamnose; 0.05-culture induced with 0.05% rhamnose; 0.1-culture induced with 0.1% rhamnose; 0.2- culture induced with 0.2% rhamnose; 0.5-culture induced with 0.5% rhamnose. The predicted molecular weight of SiaP is 35.5 kDa.

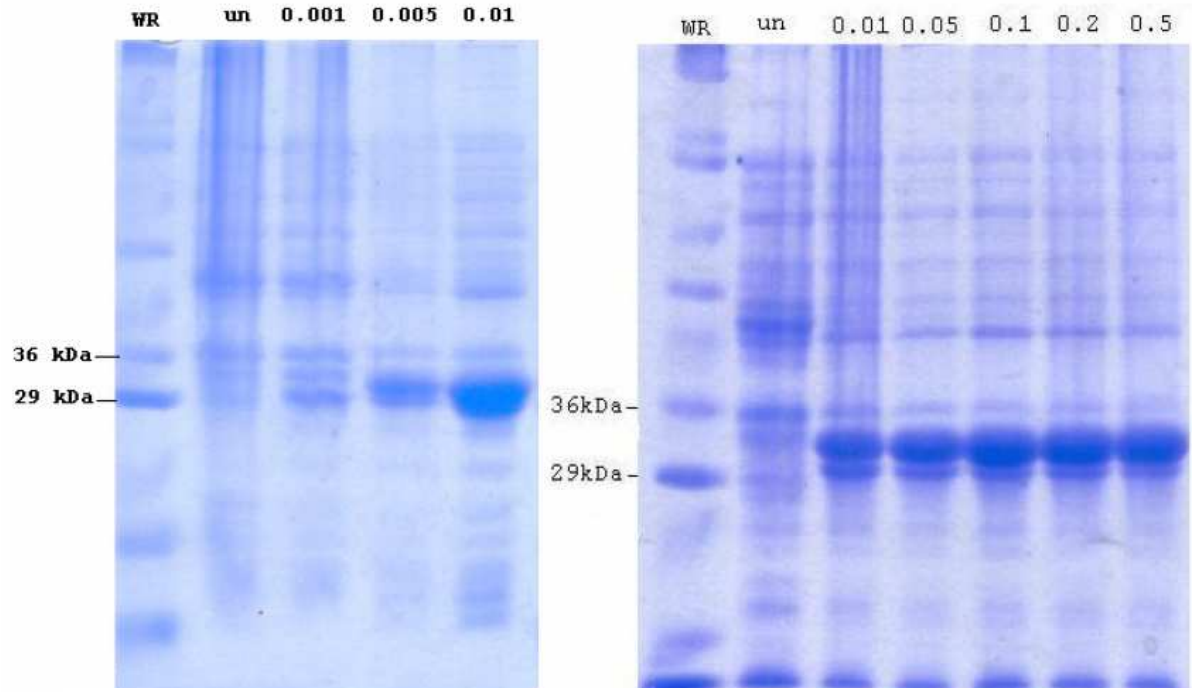


Figure 3.53. SDS-PAGE analysis of insoluble protein fraction isolated from samples taken five hours after induction with different rhamnose concentrations. WR –Sigma Wide Range protein ladder; un – uninduced; 0.001-culture induced with 0.001% rhamnose; 0.005- culture induced with 0.005% rhamnose; 0.01-culture induced with 0.01% rhamnose; 0.05-culture induced with 0.05% rhamnose; 0.1-culture induced with 0.1% rhamnose; 0.2- culture induced with 0.2% rhamnose; 0.5-culture induced with 0.5% rhamnose. The predicted molecular weight of SiaP is 35.5 kDa.

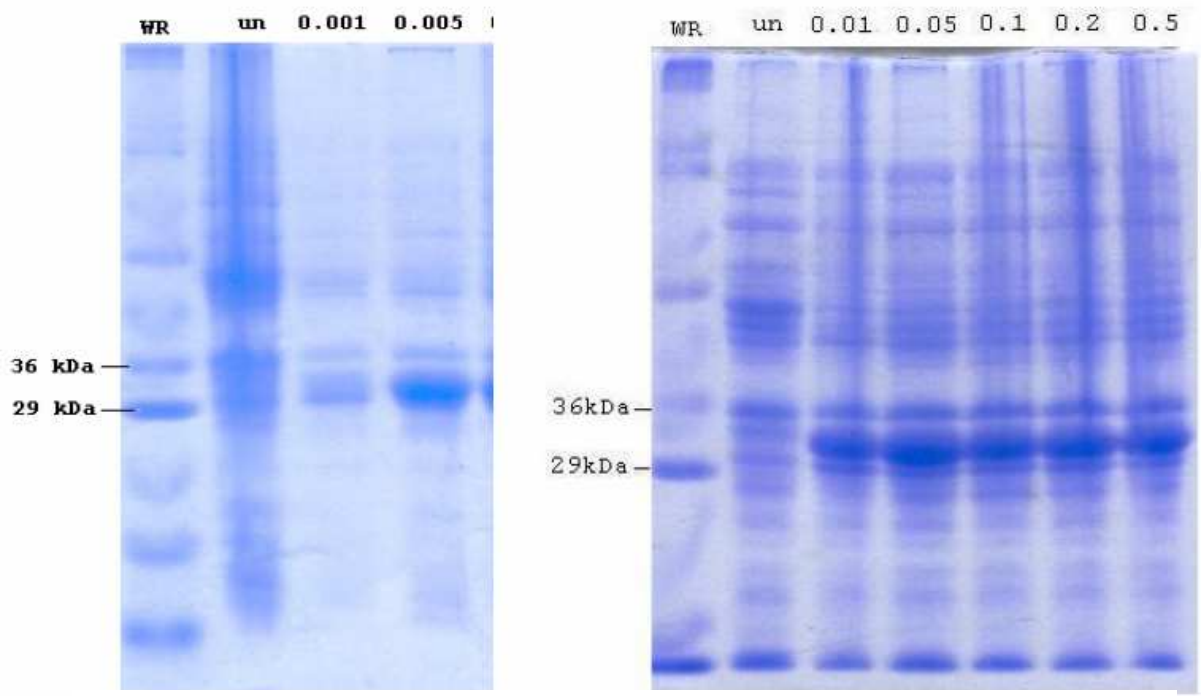


Figure 3.54 SDS-PAGE analysis of insoluble protein fraction isolated from samples taken after overnight culturing following induction with different concentrations of rhamnose. WR –Sigma Wide Range protein ladder; un – uninduced; 0.001-culture induced with 0.001% rhamnose; 0.005- culture induced with 0.005% rhamnose; 0.01-culture induced with 0.01% rhamnose; 0.05-culture induced with 0.05% rhamnose; 0.1-culture induced with 0.1% rhamnose; 0.2- culture induced with 0.2% rhamnose; 0.5-culture induced with 0.5% rhamnose. The predicted molecular weight of SiaP is 35.5 kDa.

Initial purification of the His-tagged SiaP protein over Ni-NTA agarose

Immobilised nitriloacetic acid nickel ion (Ni-NTA) resin was used to purify the expressed His-tagged SiaP protein expressed from pJS201. The stringency of imidazole washes was optimized. Soluble SiaP protein was isolated from 100 ml overnight culture induced with 0.01% (w/v) rhamnose at OD=1.2, grown at 37°C and harvested after six hours. The isolated protein was applied on pre equilibrated Ni-NTA column and washed with 20 ml of lysis buffer, followed by 4 ml washes with 20 mM, 40 mM, 60 mM, 80mM, 120mM and 300mM imidazole. Each of the 4 ml washes was collected in 1 ml aliquots and analysed by SDS-PAGE and Western blotting (Figure 5.56 -3.59). Most contaminants were removed during the washes with lysis buffer and the 20 mM imidazole (Figure 3.55). The lowest concentration of imidazole causing elution of the SiaP protein was found to be 40 mM (Figure 3.56 and 3.59), which indicates that the SiaP protein is not binding strongly to the resin. The most effective elution of the SiaP protein (assessed by comparison of lanes in SDS-PAGE gels) was observed after application of 60 mM imidazole (visible in Figure 3.56 and Figure 3.59 lane “II 60”). The elution of SiaP protein was also observed in 80 mM, 120 mM and 300 mM imidazole fractions although the concentration of SiaP protein in these fractions was lower (Figure 3.57, 3.58 and 3.59). However, the SiaP protein was also detected by Western blotting in the flow through fraction (Figure 3.59) which may indicate low affinity of the protein’s His tag towards the Ni-NTA resin. The SiaP protein yield using this purification was estimated to be 1.6 mg from 500 ml culture (purification fractions 40 mM – 300 mM imidazole were pooled together).

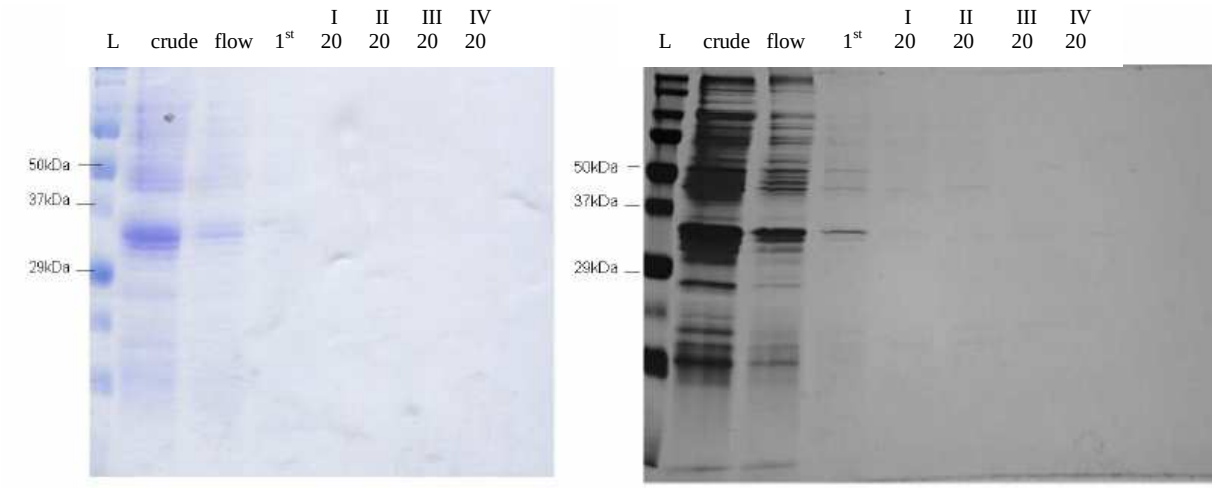


Figure 3.55. SiaP protein manual IMAC purification, part 1. Coomassie stained (left) and corresponding silver stained (right) SDS-PAGE gel with the samples of the following purification stages of the SiaP protein: crude – soluble fraction of the cell extract; flow – flow through; 1st - 10mM imidazole wash (lysis buffer); I 20mM – 1st ml of the 4ml fraction of 20mM imidazole wash; II 20mM – 2nd ml of the 4ml fraction of 20mM imidazole wash; III 20mM – 3rd ml of the 4ml fraction of 20mM imidazole wash; IV 20mM – 4th ml of the 4ml fraction of 20mM imidazole wash. L - Sigma Wide Range protein ladder. The predicted molecular weight of SiaP is 35.5 kDa

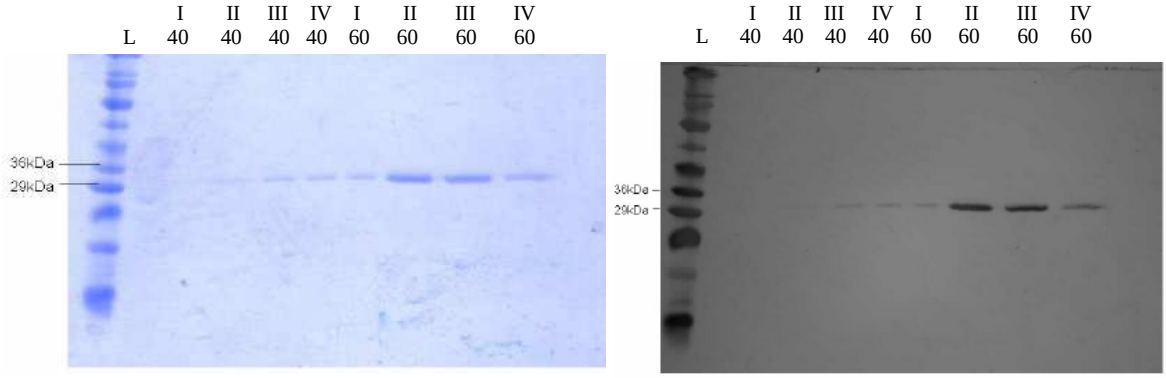


Figure 3.56. SiaP protein manual IMAC purification, part 2. Coomassie stained (left) and a corresponding silver stained (right) SDS-PAGE gel with the samples of the following purification stages of the SiaP protein: I 40mM – 1st ml of the 4ml fraction of 40mM imidazole wash; II 40mM – 2nd ml of the 4ml fraction of 40mM imidazole wash; III 40mM – 3rd ml of the 4ml fraction of 40mM imidazole wash; IV 40mM – 4th ml of the 4ml fraction of 40mM imidazole wash; I 60mM – 1st ml of the 4ml fraction of 60mM imidazole wash; II 60mM – 2nd ml of the 4ml fraction of 60mM imidazole wash; III 60mM – 3rd ml of the 4ml fraction of 60mM imidazole wash; IV 60mM – 4th ml of the 4ml fraction of 60mM imidazole wash. L - Sigma Wide Range protein ladder. The predicted molecular weight of SiaP is 35.5 kDa

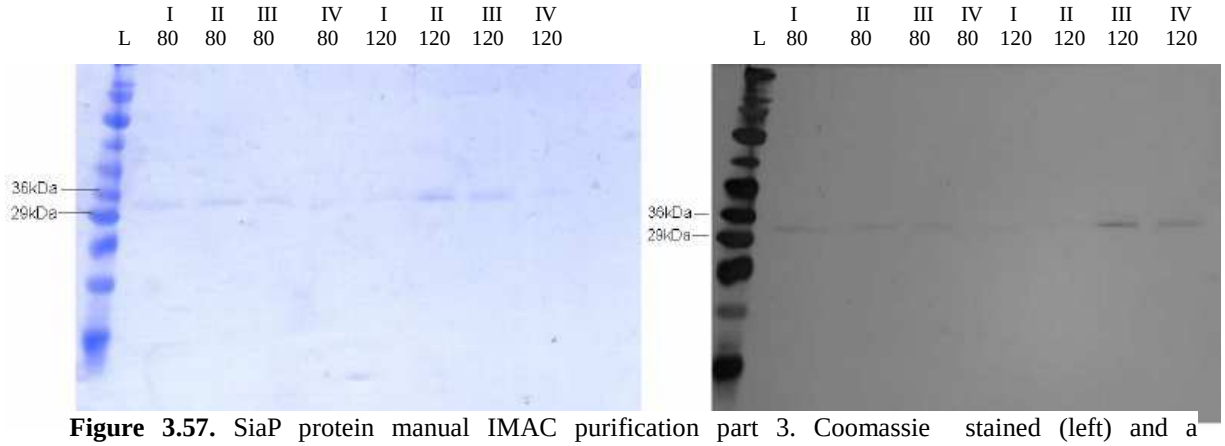


Figure 3.57. SiaP protein manual IMAC purification part 3. Coomassie stained (left) and a corresponding silver stained (right) SDS-PAGE gel with the samples of the following purification stages of the SiaP protein: I 80mM – 1st ml of the 4ml fraction of 80mM imidazole wash; II 80mM – 2nd ml of the 4ml fraction of 80mM imidazole wash; III 80mM - 3rd ml of the 4ml fraction of 80mM imidazole wash; IV 80mM – 4th ml of the 4ml fraction of 80mM imidazole wash; I 120mM – 1st ml of the 4ml fraction of 120mM imidazole wash; II 120mM – 2nd ml of the 4ml fraction of 120mM imidazole wash; III 120mM – 3rd ml of the 4ml fraction of 120mM imidazole wash; IV 120mM – 4th ml of the 4ml fraction of 120mM imidazole wash. L - Sigma Wide Range protein ladder. The predicted molecular weight of SiaP is 35.5 kDa

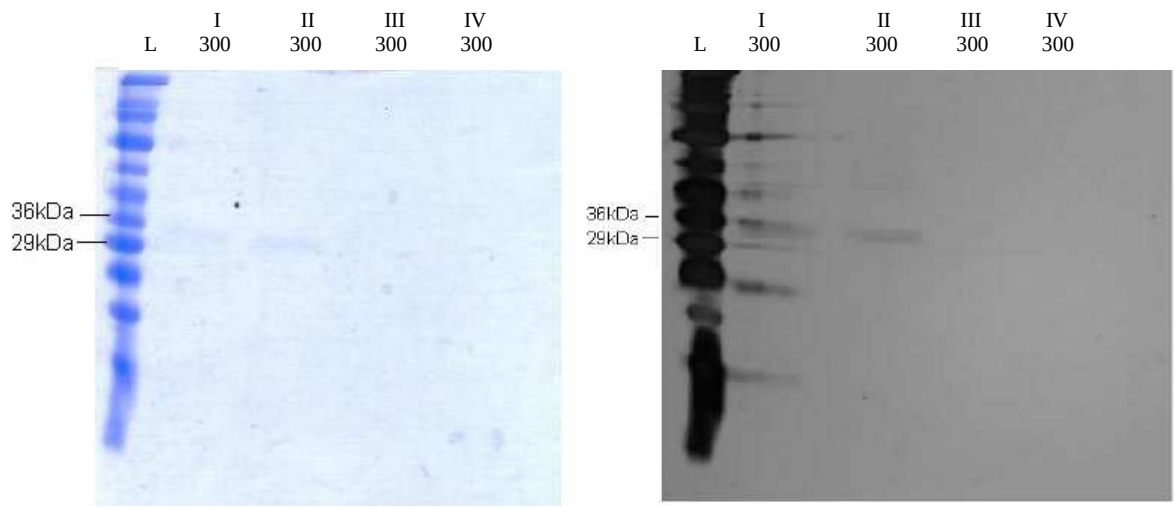


Figure 3.58. SiaP protein manual IMAC purification part 4. Coomassie stained (left) and a corresponding silver stained (right) SDS-PAGE gel with the samples of the following purification stages of the SiaP protein: I 300mM – 1st ml of the 4ml fraction of 300mM imidazole wash; II 300mM – 2nd ml of the 4ml fraction of 300mM imidazole wash; III 300mM – 3rd ml of the 4ml fraction of 300mM imidazole wash; IV 300mM – 4th ml of the 4ml fraction of 300mM imidazole wash. L - Sigma Wide Range protein ladder. The predicted molecular weight of SiaP is 35.5 kDa.

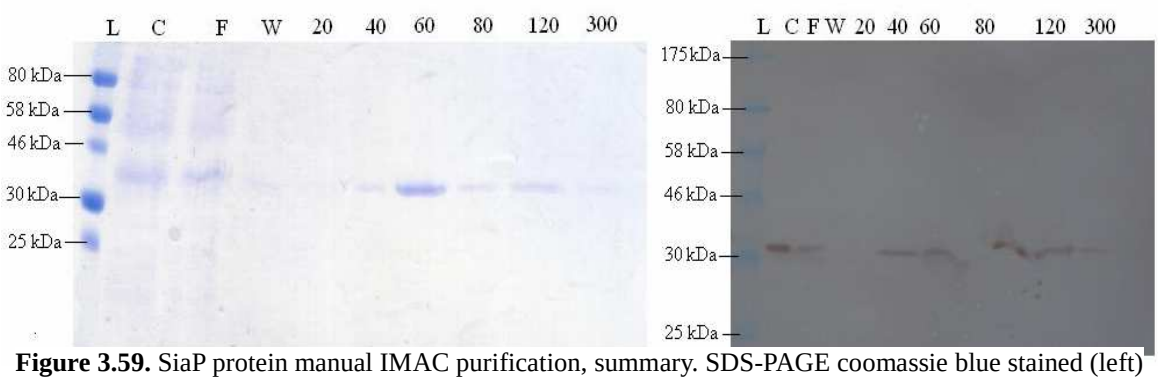


Figure 3.59. SiaP protein manual IMAC purification, summary. SDS-PAGE coomassie blue stained (left) and corresponding Western blot (right) analysis of the samples of the following purification stages of the SiaP protein: C – crude, soluble cell extract; F - flow through; 20 – 2nd ml of the 4ml fraction of 20 mM imidazole wash; 40 – 2nd ml of the 4ml fraction of 40 mM imidazole wash; 60 – 2nd ml of the 4ml fraction of 60 mM imidazole wash; 80 – 2nd ml of the 4ml fraction of 80 mM imidazole wash; 120 – 2nd ml of the 4ml fraction of 120 mM imidazole wash; 300 – 2nd ml of the 4ml fraction of 300 mM imidazole wash. L – NEB Prestained Protein Marker, Broad Range. The predicted molecular weight of SiaP is 35.5 kDa.

Linker introduction

The inefficient binding of the His-tagged SiaP protein expressed from pJS201 to the Ni-NTA resin (as the protein is present in a flow through fraction) may be attributed to low accessibility of the tag. In order to increase the accessibility of the tag the SiaP protein was modified. An eight amino acid linker was introduced to separate the protein and the 6His-tag. The $(Gly\ Gly\ Gly\ Ser)_2$ linker is a straight flexible string of amino acids that extends the His tag from the protein and thus makes the His-tag more accessible. The linker sequence (Figure A) was introduced by PCR amplification of the entire pJS201 plasmid using phosphorylated primers. The optimised PCR annealing temperature was 60°C, annealing time was 15 s and extension time 45 s.The primers containing liker sequence are presented in Figure 3.60.

A linker sequence:
Amino acid sequence: Gly Gly Gly Ser Gly Gly Gly Ser
Nucleotide sequence: GGT GGC GGA AGT GGC GGA GGA AGT

B primer sequences:
forward primer (JS2011F): 5' GGCGGAGGA AGTCTCGAGCACCACCACCACCACCAC 3'

reverse primer (JS2012R): 5' ACTTCCGCCACCTGGATTGATTGCTTCAATTTGTTTTA 3'

Figure 3.60. Primers for linker introduction into the SiaP protein. (A)Amino acid and corresponding nucleotide sequences of the linker introduced to separate the SiaP protein and the His tag. (B) primer sequences used to amplify the pJS2011 plasmid. The part of the linker introduced in the forward primer is marked in brown and the part of the linker introduced in the reverse primer is marked in green. The sequences in black correspond to the pJS201 plasmid used as a template. In order to avoid secondary structures formation of the primers various alternative codons for Gly were used: GGT, GGC and GGA.

After PCR the DNA fragments were extracted from the reaction mix (RBC gel extraction kit) and subjected to DpnI digestion that cleaves only methylated DNA (NEB) and thus the original plasmid (pJS201 used as a template) was eliminated. Then the product was separated by gel electrophoresis in SYBR Safe agarose, gel extracted and subjected to ligation with T4 DNA Ligase. Subsequently the ligation mix was used to transform *E. coli* KRX competent cells. The plasmid was isolated, analysed by restriction and sequencing (Figure 3.61) and named pJS2011 (Figure 3.62).

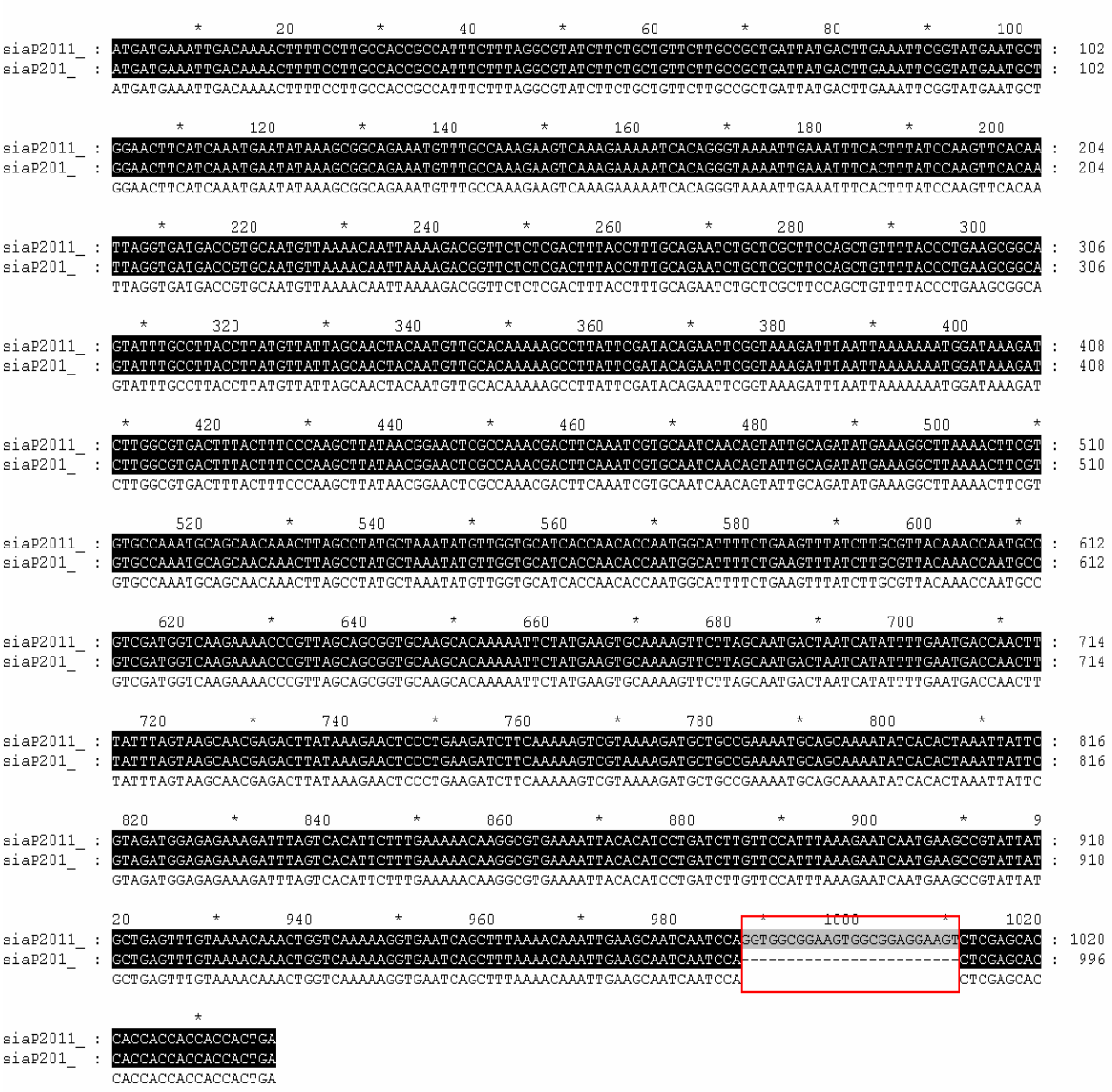


Figure 3.61. Nucleotide sequence alignment of *siaP* gene from pJS201 plasmid and *siaP* gene present in the pJS2011 plasmid sequenced by MWG. The sequence coding for the linker is marked in red.

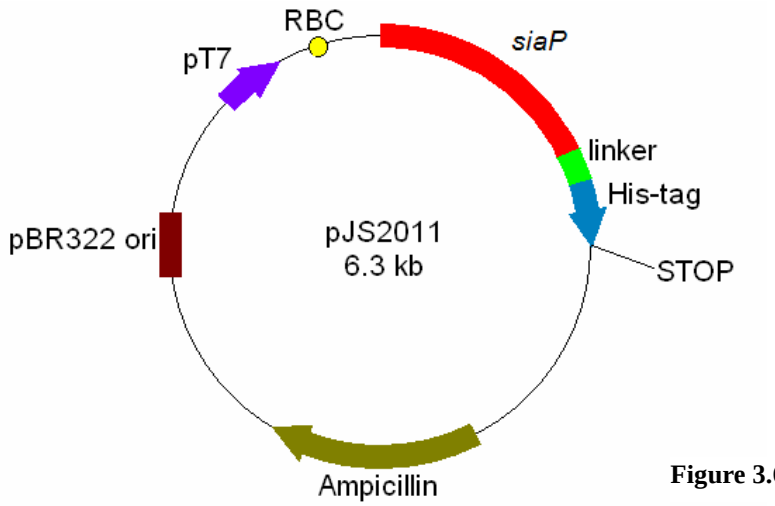


Figure 3.62. A map of the pJS2011 plasmid.

A similar procedure for linker introduction was applied for pJS202 and pJS203 plasmids. The linker was introduced between the protein and the His-tag using JS2011F and JS2011R primers (Figure 3.60). The PCR conditions as well as further cloning steps were identical to these described above with the exception of the template used. Plasmids pJS202 and pJS203 were used as templates in the PCR step. The sequences of newly created plasmids were confirmed by restriction analysis and sequencing (MWG) (data not shown) and the plasmids were named pJS2021 (Figure 3.63) and pJS2031 (Figure 3.64).

The nomenclature for modified SiaP proteins in subsequent sections is as follows:

- SiaP 201 – SiaP-6His, expressed from pJS201,
- SiaP 202 – SiaP-6Lys, expressed from pJS202,
- SiaP 203 – SiaP-Strep2, expressed from pJS203,
- SiaP 2011 – SiaP-linker-6His, expressed from pJS2011,
- SiaP 2021 – SiaP-linker-6His-6Lys, expressed from pJS2021,
- SiaP 2031 – SiaP-linker-6His-Strep2, expressed from pJS2031.

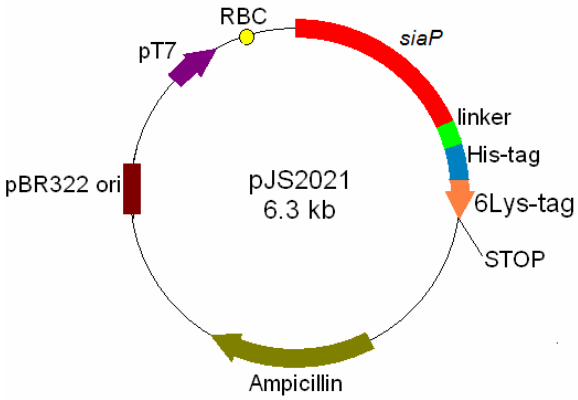


Figure 3.63. A map of the pJS2021 plasmid.

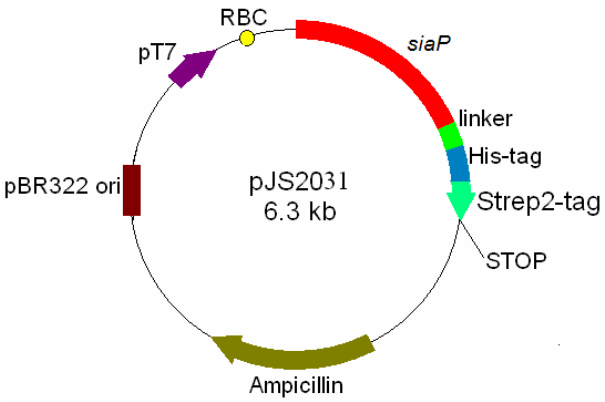


Figure 3.64. A map of the pJS2031 plasmid.

Purification of the His-tagged SiaP protein with the linker

The His-tagged SiaP protein with the linker (SiaP 2011) was expressed in the KRX strain of *E. coli* and purified over immobilised nitriloacetic acid nickel ion resin. The stringency of imidazole washes was optimized. Soluble SiaP 2011 was isolated from 100 ml overnight culture induced with 0.01% (w/v) rhamnose at OD=1.2 (stationary phase), grown at 37°C and harvested after six hours. The isolated protein was applied on pre equilibrated Ni-NTA column and purified in the same way as previously described for the SiaP 201 protein. It was found that in converse to the SiaP 201 protein the SiaP 2011 was not present in the flow through fraction (Figure) which sugests that introduction of the linker altered the behaviour of the protein increasing its affinity towards the Ni-NTA resin. Furthermore, the SiaP 2011 protein was not detected in the 40 mM imidazole wash (Figure 3.65), as the linker free SiaP had been (Figure 3.59), and the lowest concentration causing SiaP 2011 elution was 60 mM. The SiaP 2011 protein yield using this purification was estimated to be 6 mg/500 ml culture, which also indicates higher efficiency of the procedure (same as before).

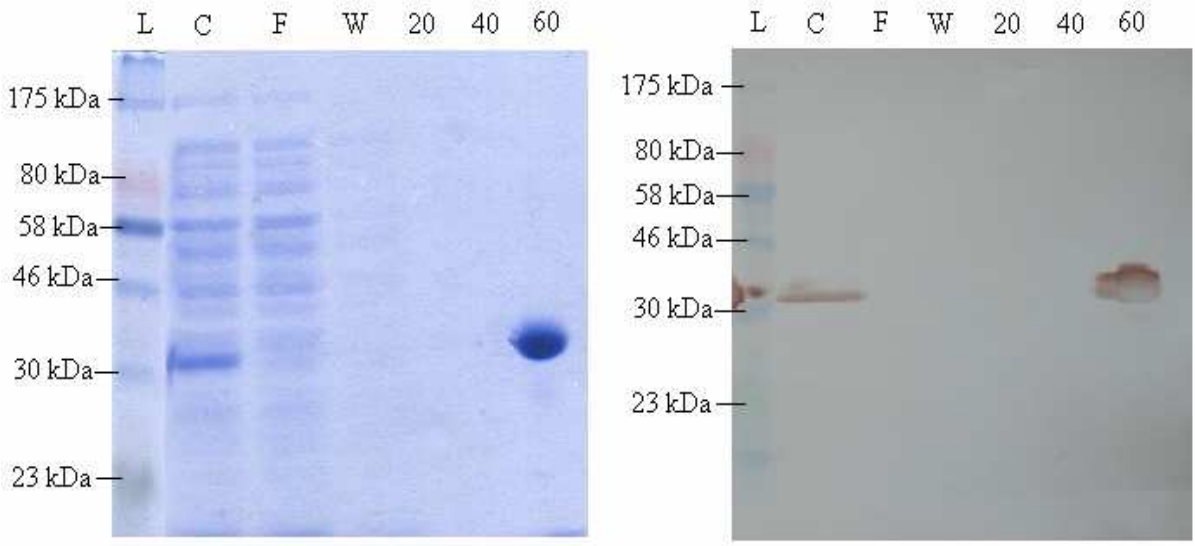


Figure 3.65 SDS-PAGE and corresponding Western blot analysis with the samples of SiaP 2011 protein purification stages. C – crude, soluble cell extract; F - flow through; W – wash with lysis buffer; 20 – 2nd ml of the 4ml fraction of 20mM imidazole wash; 40 – 2nd ml of the 4ml fraction of 40mM imidazole wash; 60 – 2nd ml of the 4ml fraction of 60mM imidazole wash; L – NEB Prestained Protein Marker, Broad Range. The predicted molecular weight of SiaP is 35.5 kDa.

Purification of SiaP 201 and SiaP 2011 proteins using IMAC on FPLC

The His-tagged proteins, both with the linker and without the linker, were purified by IMAC using FPLC. Step gradient of imidazole in the range between 10 mM (lysis buffer) and 250 mM was used for purification, 0.5 ml samples were collected and analysed by SDS-PAGE. During the purification of SiaP 201 protein a minor increase in absorbance at 280 nm was observed which correlated with the change (increase) in the imidazole concentration. The wash with 20 mM imidazole resulted in a slight, flat peak which may be attributed to contaminant elution from the column but also a SiaP elution as the protein was detected in this fraction by SDS-PAGE analysis (Figure 3.66 C lane "20"). This indicates that SiaP 201 binding to the resin is weak and this result supports the findings from the manual IMAC purification described in sections above. The peak at 40 mM imidazole may be a result of both contaminants and SiaP elution as both of them can be found in Figure 3.66 C lane "40". The sharpest peak was observed after 85 mM imidazole application and may be attributed solely to SiaP protein elution as no contaminants were detected in lane "85" in Figure 3.66 C. The second band on the SDS-PAGE gel visualisation of this fraction may be attributed to the SiaP protein with a partially degraded N-terminus; hovewer this was not investigated. The 250 mM imidazole application resulted in an increase in absorbance but no peak was formed which indicates lack or very low amounts of protein contribution in this fraction. Low amounts of SiaP 2011 protein present in the 250 mM fraction were confirmed by SDS-PAGE analysis (lane "250" in Figure 3.66 C).

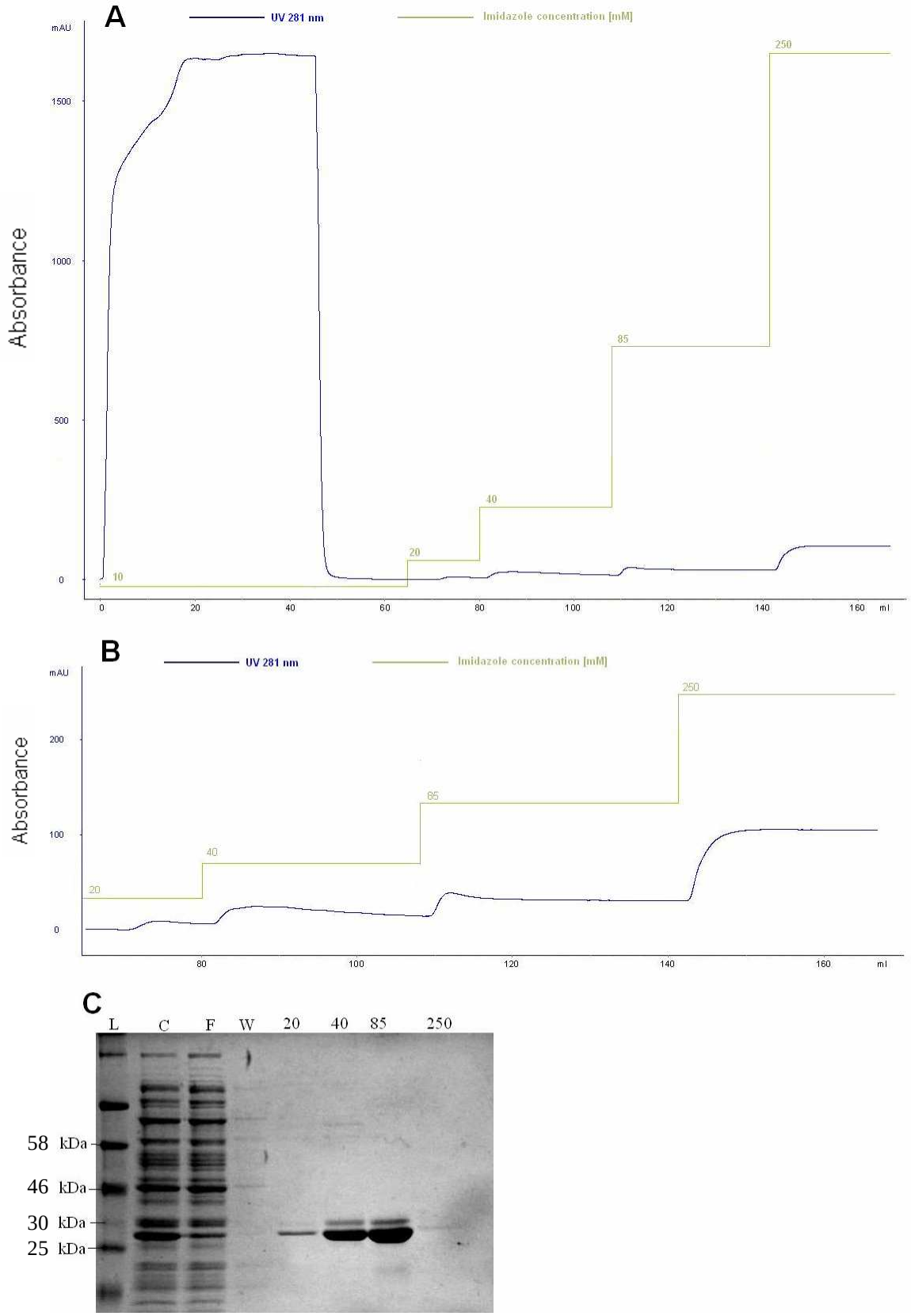


Figure 3.66. SiaP 201 protein purification by IMAC using FPLC. (A) – A_{280} reading of the overall purification including the initial sample injection and lysis buffer wash. (B) – A_{280} reading of the 20 mM – 250 mM imidazole washes. (C) – SDS-PAGE analysis of the SiaP 201 purification fractions. Lanes were loaded as follows: L – Prestained Protein Ladder, Broad Range (NEB); C – crude cell extract; F – flow through; W – wash with lysis buffer; 20 – 20 mM imidazole wash; 40 – 40 mM imidazole wash; 85 – 85 mM imidazole wash; 250 – 250 mM imidazole wash. The predicted molecular weight of SiaP is 35.5 kDa.

Similarly, during the IMAC FPLC purification of SiaP 2011 protein a minor increase in absorbance at 280 nm was observed which correlated with the change (increase) of the imidazole concentration (Figure 3.67). The 20 mM imidazole wash resulted in a slight, flat peak which may be attributed to contaminant elution from the column. The contaminants present in the 20 mM imidazole fraction are observed by SDS-PAGE analysis (Figure 3.67 C lane “20”). No SiaP was detected in this fraction by SDS-PAGE analysis which indicates that the modified version of the protein with the linker interacts with the resin with higher affinity. This result is also in concert with findings from manual IMAC purification studies of these two proteins. The peak at 40 mM imidazole may be a result of both contaminant and SiaP elution as both of them can be found in Figure 3.67 C lane “40”. Similarly to the SiaP 201, the sharpest peak was observed after 85 mM imidazole application and may be attributed solely to SiaP protein elution as no contaminants were detected in Figure 3.67 C lane “85”. The 250 mM imidazole application resulted in an increase of absorbance but no peak was formed which indicates lack or very low amounts of protein in this fraction. Lack of any protein in this fraction was also confirmed by SDS-PAGE analysis (Figure 3.67 C lane “250”).

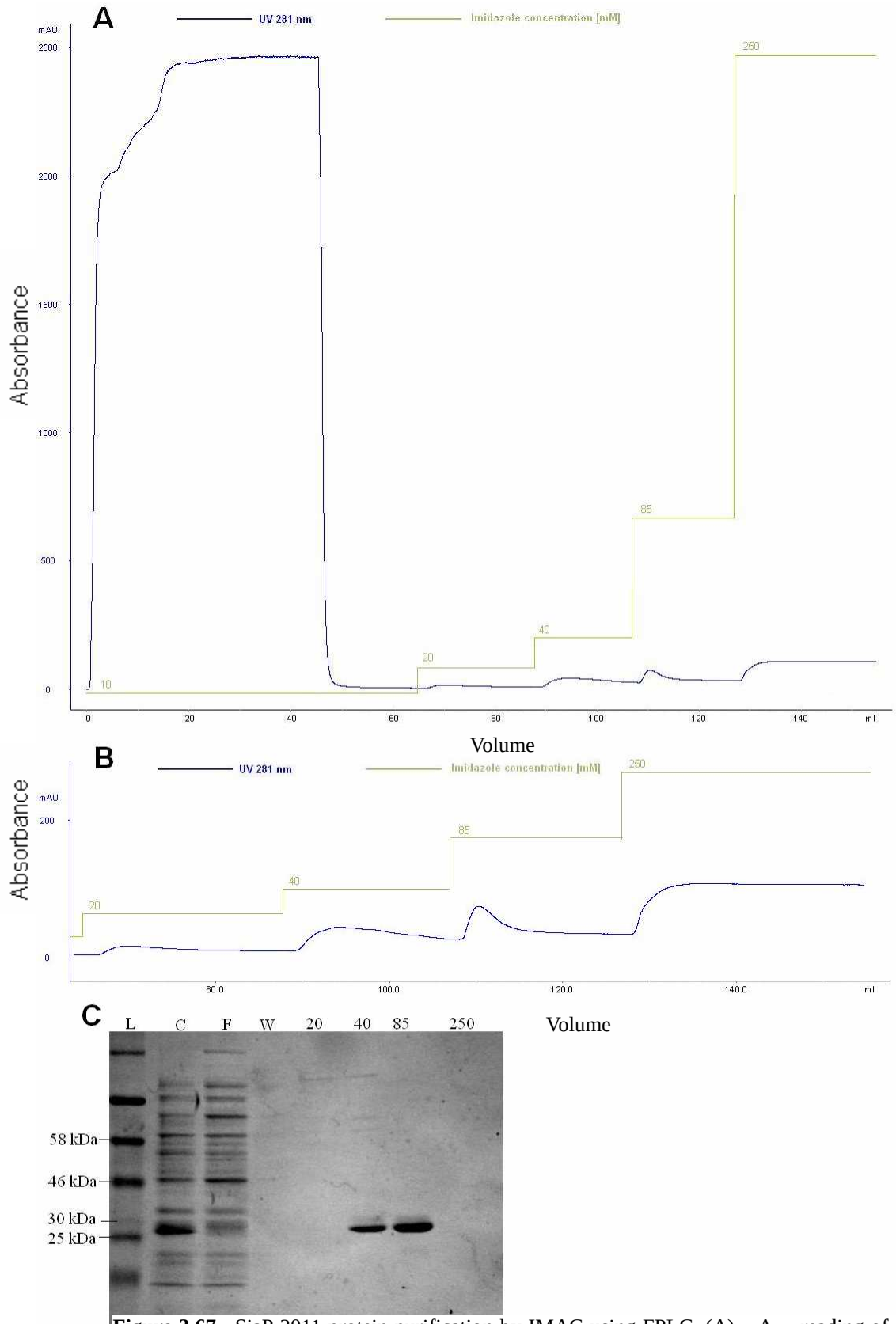


Figure 3.67. SiaP 2011 protein purification by IMAC using FPLC. (A) – A_{280} reading of the overall purification including the initial sample injection and lysis buffer wash. (B) – A_{280} reading of the 20 mM – 250 mM imidazole washes. (C) – SDS-PAGE analysis of the SiaP 2011 purification fractions. Lanes were loaded as follows: L – Prestained Protein Ladder, Broad Range (NEB); C – crude cell extract; F – flow through; W – wash with lysis buffer; 20 – 20 mM imidazole wash; 40 – 40 mM imidazole wash; 85 – 85 mM imidazole wash; 250 – 250 mM imidazole wash. The predicted molecular weight of SiaP is 35.5 kDa.

Purification of the Strep2-tagged SiaP protein using StrepTactin sepharose

StrepTactin is a modified version of straptavidin offering high affinity and high specificity binding to the Strep2-tag. The competitive elution of the target Strep2-tagged protein with desthiobiotin (reversibly binding biotin analogue) adds additional specificity to the system. In order to confirm the functionality of the introduced C-terminal, Strep2-tag the StrepTactin sepharose columns were used to purify the SiaP protein. Additionally, to study the impact of the linker addition on the protein features, two derivatives of Strep2-tag SiaP proteins were analysed, namely SiaP 203 (Strep2-tagged without the linker) and SiaP 2031 (Strep2-tagged with the linker). Both proteins were found in the elution fractions (Figure and Figure), which indicates that the Strep2-tag in these proteins is functional. The introduction of the linker was found to influence the protein affinity for the StrepTactin resin. The SiaP 203 protein, in converse to the SiaP 2031 protein, was detected in the wash fraction (Figure), which indicates low affinity of the protein for the resin. The lack of SiaP 2031 in the wash fraction (Figure) indicates that the linker introduction increased the affinity of the Strep2-tagged SiaP protein for the StrepTactin resin used in the purification protocol.

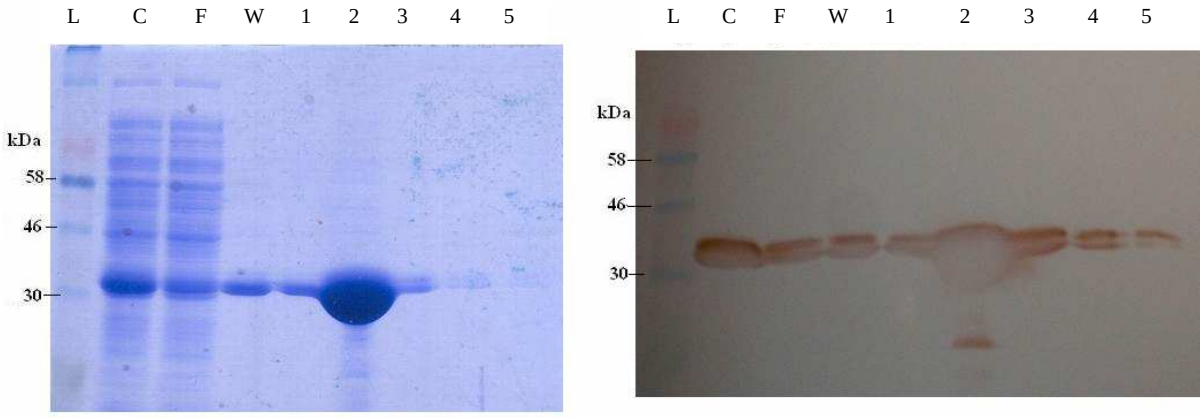


Figure 3.68. Purification of the Strep2-tagged SiaP 2031 protein over StrepTrap HP column. SDS-PAGE analysis (left) and corresponding Western blot (right) analysis. Samples colected in 1 ml fractions. Lanes: L – Prestained Protein Marker, NEB; C – Crude cell extract; F – flow through; W – wash; 1 – elution fraction, first ml; 2 – elution… The Western Blot was probed with anti-His antibodies. The predicted molecular weight of SiaP is 35.5 kDa

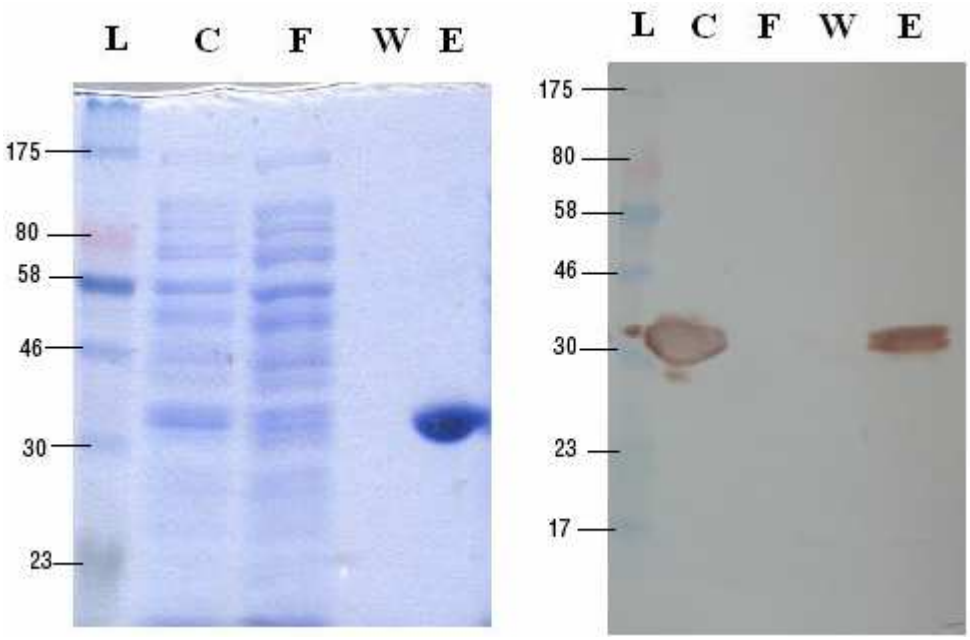


Figure 3.69. SDS-PAGE (left) analysis and corresponding Western blot (right) analysis of the purification of the Strep2-tagged SiaP 2031 protein over StrepTrap HP column. Samples colected in 1 ml fractions. Lanes: L – Prestained Protein Marker, NEB; C – Crude cell extract; F – flow through; W – wash; E - elution fraction. The Western Blot was probed with anti-His antibodies.

3.3.4 Characterisation of the SiaP protein expressed in *E. coli*

Determination of molecular weight and homogeneity of the His-tagged SiaP protein using MS

The homogeneity and molecular weight of the His-tagged SiaP 201 protein expressed in *E. coli* was determined using electrospray ionization (EI) mass spectrometry. The sample was prepared and analysed using standard protocol (see section 2.23). Two protein components were identified during this analysis. The revealed molecular mass of the major component was 35238.41 ± 0.66 which is almost identical to the predicted (35229.97) of the mature protein after cleavage of the 23 amino acid signal sequence (Figure 3.70). The revealed molecular weight of the second component was 35538.41 ± 2.35 which corresponds to the 35539.67 predicted mass of the mature protein substituted with Neu5Ac.

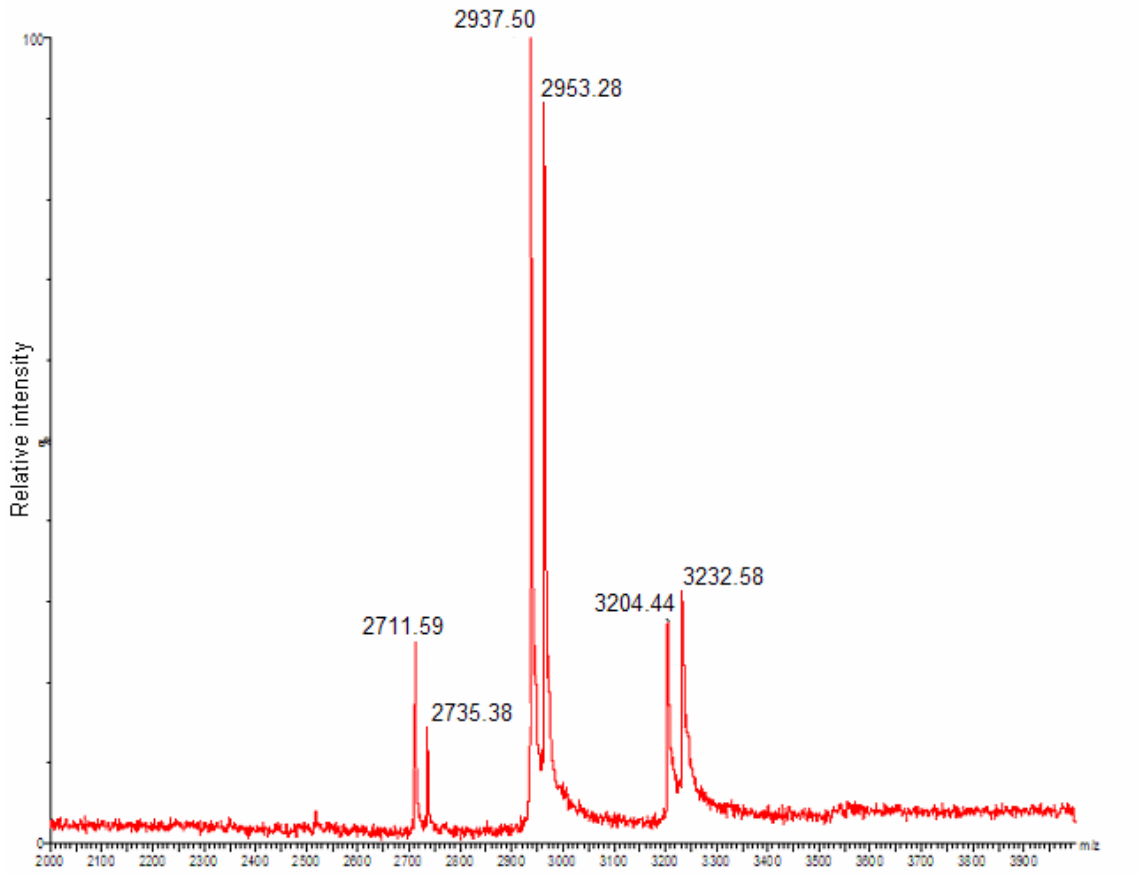


Figure 3.70. The EI-MS analysis of the His-tagged SiaP 201 protein expressed in *E. coli*. The molecular masses corresponding to peaks 2711.59, 2937.50 and 3204.44 corresponds to unsubstituted SiaP proteins ions (molecular mass 35238.41 ± 0.66 Da) and peaks 2735.38, 2953.28 and 3232.58 corresponds to substituted SiaP ions (molecular mass 35538.41 ± 2.35**)**. The concentration of the SiaP in the injected sample was 285 nM.

3.3.5 Production of the sialic acid free form of SiaP

The application of the SiaP protein as an analytical tool in various platforms requires a diverse, in-depth functional analysis. As presented in previous sections of this thesis the isolated and purified SiaP protein from *E. coli* cells was partially saturated with Sia which would limit further studies of the protein properties. Therefore, a production method of Sia-free SiaP protein had to be developed. In order to produce Sia-free SiaP protein two separate approaches were undertaken.

3.3.5.1 Post-induction culturing time extension

The first approach to produce Sia-free SiaP was based on the fact that *E.coli* cells metabolize Sia and are able to use it a sole source of carbon (Vimr and Troy, 1985). Culture of *E. coli* KRX pJS201 were induced at OD=0.5 (midlog phase) with rhamnose 0.01% (w/v). Post induction culturing time was extended in order to allow cells to metabolize Sia present in the expression medium (LB) that would otherwise be bound by the SiaP protein. The cultures were grown at 37°C for 4 and 24 hours after inductio. Isolated proteins from both cultures were analysed by MS. There were both Sia bound and Sia free fractions of SiaP detected in the sample isolated from the culture after 4 hours post induction (Figure 3.71). The SiaP sample isolated 24 hours after induction was composed solely of Sia free molecules (Figure 3.72). This indicates that production of Sia free SiaP protein is possible by extension of culturing time.

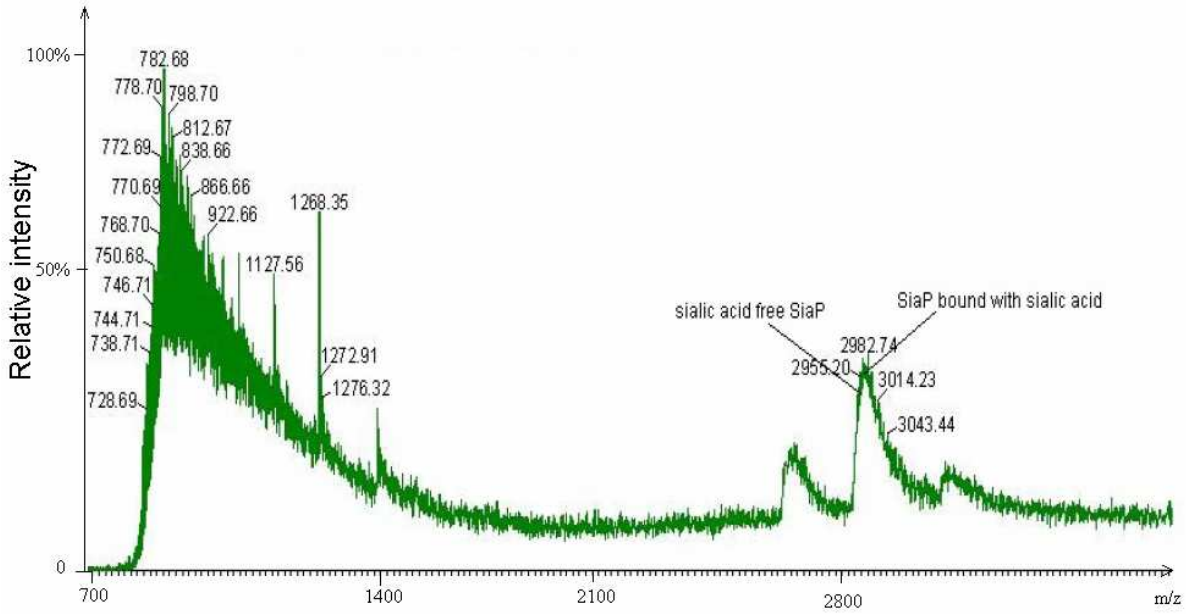


Figure 3.71. MS analysis of the SiaP protein isolated from *E. coli* KRX 4 hours following induction. The peaks corresponding to both Sia free and Sia bound forms of SiaP are marked. The concentration of the SiaP in the injected sample was 285 nM.

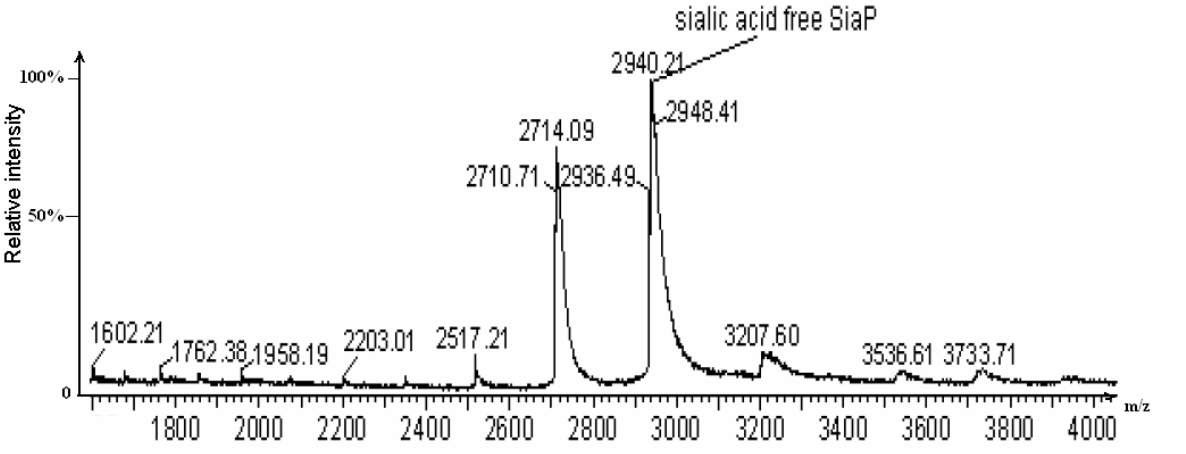


Figure 3.72. MS analysis of the SiaP protein isolated from *E. coli* KRX 24 hours following induction. The peak corresponding to Sia free SiaP is marked. The concentration of the SiaP in the injected sample was 285 nM.

3.3.5.2 Sialic acid removal by SiaP protein denaturation

In order to produce the Sia-free form of SiaP, the protein was immobilised on a NiNTA agarose column and subjected to strong chaotropic agents (urea or GnHCl) and after denaturation the Sia was washed off. As the native three-dimensional structure of ligand-binding pocket of the protein is required for Sia binding, the denaturation of the protein leads to Sia release. Initially, SiaP (2mg/ml = 567.7 μM in lysis buffer) was saturated with Neu5Ac (1135.4 μM) and after 10 min incubation the free Sia was removed by washing with molecular grade water (using 10 kDa cut off point Amicon columns). The saturation of SiaP with Sia was confirmed by MS analysis (Figure 3.73 A). Then the SiaP protein was immobilised via the His-tag on Ni-NTA agarose beads and the column was washed with 10 ml of 8 M urea or 6 M GnHCl (in lysis buffer) to remove the Sia. The protein was then washed with 2 ml of lysis buffer to remove the denaturants, eluted with 3 ml of 300 mM imidazole buffer and washed with water. The SiaP protein after denaturation was analysed by MS to confirm the effectiveness of the procedure. There was no Sia bound SiaP detected in samples treated with either GnHCl or urea (Figure 3.73 B and C respectively) which indicates that the Sia was removed from the protein during the denaturation step. In order to confirm that the removal of Sia is due to denaturant application, and not the impact of the wash, the control sample of SiaP was treated with lysis buffer instead of denaturant. After the procedure the SiaP protein was still saturated with Sia (Figure 3.74) which indicates that it was not removed by the lysis buffer wash.

The efficient removal of Sia from the SiaP protein binding site requires denaturation of the protein; conversely the binding of Sia occurs only if the SiaP protein is correctly folded. Hence, after Sia removal by denaturant application it is crucial to make certain that the protein refolds and regains its native conformation in order to be fully active. After denaturation the SiaP protein immobilised on the Ni-NTA resin was washed with lysis buffer to remove urea or GnHCl. To verify the activity of the SiaP protein the sample was re-saturated with Sia and analysed with MS. It was found that only Sia bound molecules were present (Figure 3.73 D) which indicates full activity of the protein.

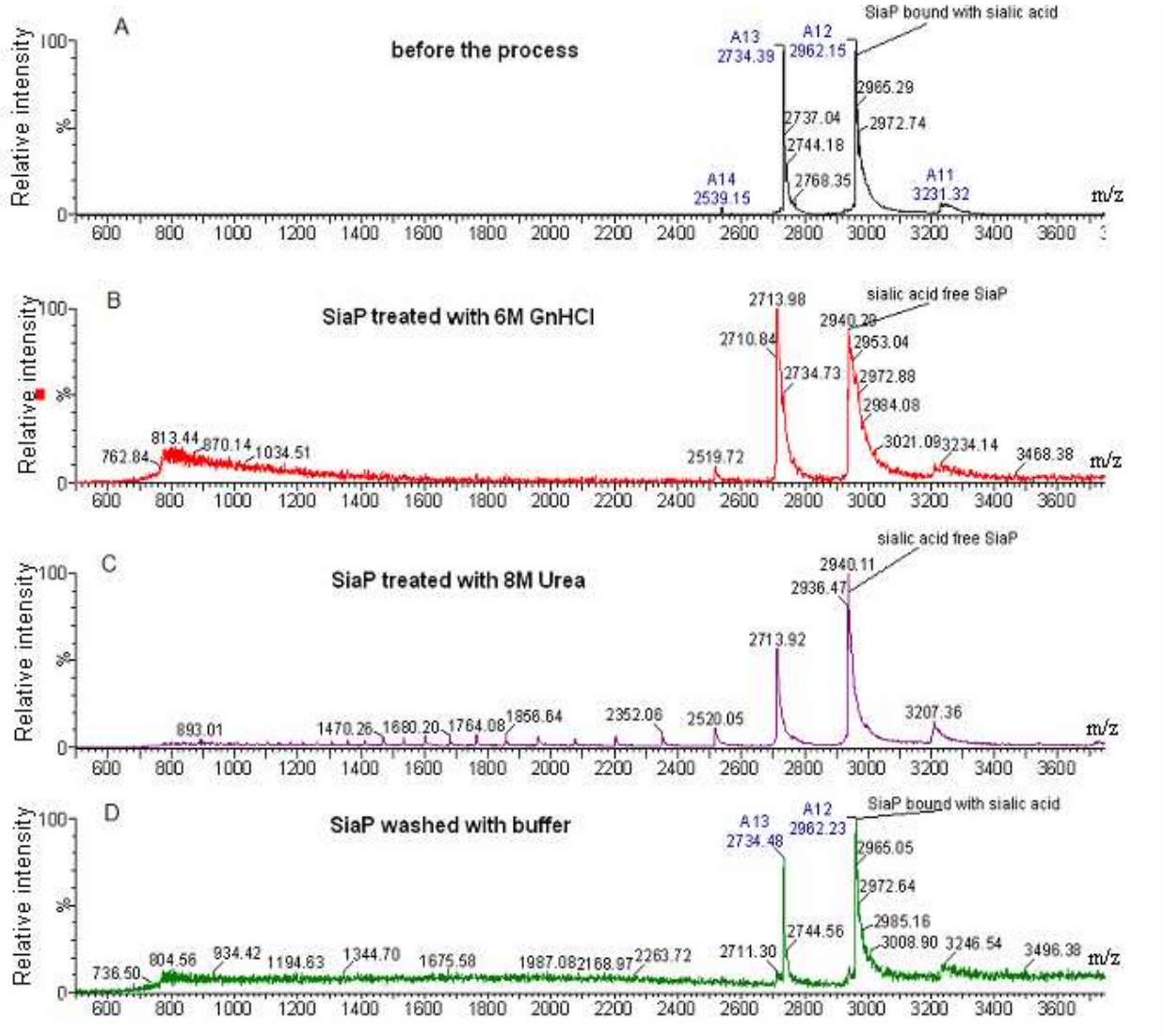


Figure 3.73. MS analysis of Sia elution from SiaP using GnHCl and urea. The concentration of SiaP in all samples was 285 nM. SiaP samples before (A) the process, after treatment with respectively GnHCl (B) and urea (C) and a control sample (washed with lysis buffer) (D) are presented. Both ligand bound and ligand free SiaP fractions are marked.

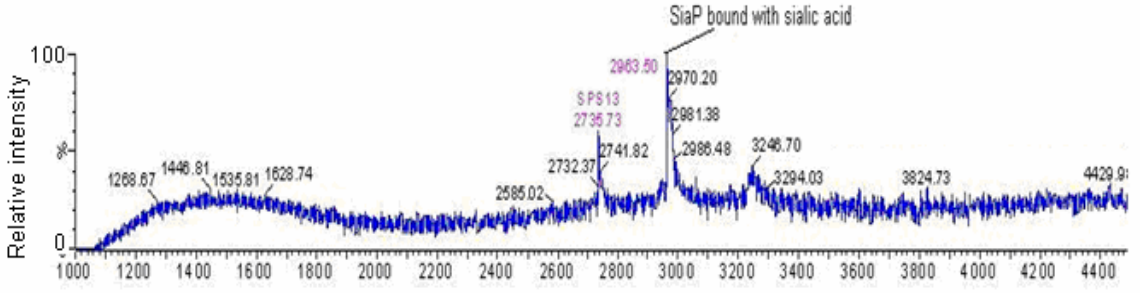


Figure 3.74. MS analysisof the SiaP activity after 8 M Urea treatment.

Investigation of SiaP protein specificity using ELLA

The specificity of the SiaP protein for sialylated glycans was investigated using the ELLA. Immobilised glycoproteins carrying terminal Sia were used as targets for SiaP and two commercially available Sia binders, wheat germ agglutinin (WGA) and *Sambuccus nigra* agglutinin (SNA) used as positive controls. Sialylated glycoproteins fetuin and transferin, along with asialofetuin used as a negative control, were probed with 6His-tagged SiaP protein and biotinylated WGA and SNA. It was found that SiaP did not bind to either of the sialylated glycoproteins tested (Figure 3.75). The presence of glycoproteins on the surface of the well was confirmed by positive results of WGA and SNA tests. The His-tagged SiaP immobilised directly on the surface as a positive control for anti-His antibodies gave a positive result in the test (anti-His Abs positive control, Figure 3.75) which confirms that SiaP can be detected in a given concentration (5µg/ml). The negative controls for both anti-His and anti-biotin antibodies were tested negatively which confirms the antibodies' specificities.

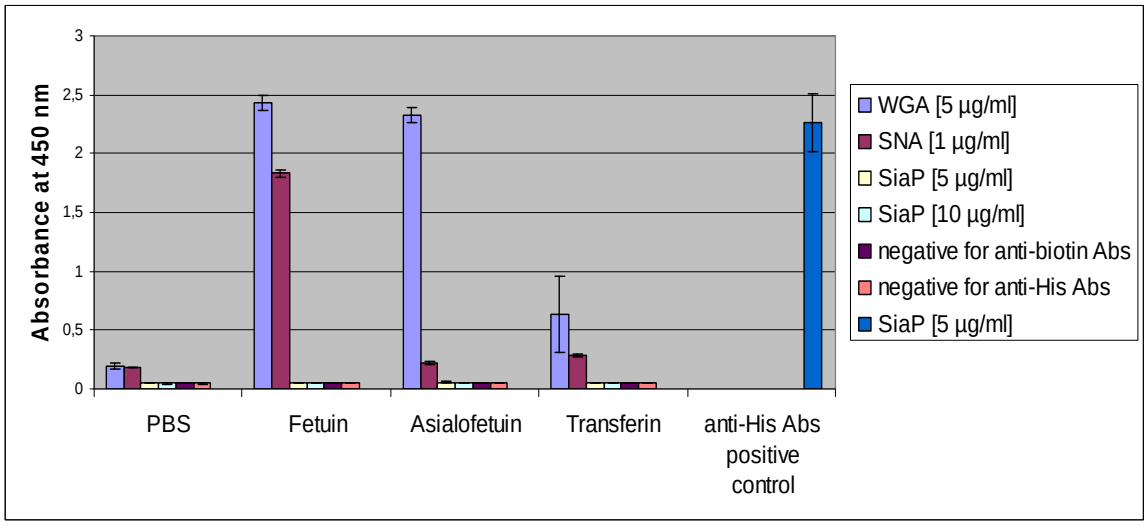


Figure 3.75. ELLA analysis of the SiaP specificity for sialylated glycoproteins. SiaP protein at two different concentrations was tested against Fetuin, Asialofetuin and Transferin immobilised directly on the plate. The PBS buffer was used as a negative control for Sia-binding proteins. The negative controls for antibodies consisted of glycoproteins (or PBS) untreated with Sia-binding proteins. The positive control for anti-His antibodies was a SiaP protein [5 µg/ml] immobilised directly on the plate. As controls two other Sia-binding proteins were tested: WGA and SNA.

The negative result of SiaP protein in the ELLA assay with glycoproteins may be attributed to either inability of the protein to bind the terminal Sia of the glycoprotein or inability to bind Sia in any conjugate. It is possible that the lack of interaction of SiaP with glycoproteins is due to steric constraints as the SiaPs conformational change during ligand accommodation may be prohibited by the glycan structure to which the Sia is attached.

The SiaP's deep binding pocket may limit the range of possible ligands to free Sia or Sia coupled with much less complex structure than a glycan tree. The ability of the SiaP protein to bind such structures was tested using a biotinylated form of Sia. A flexible linker (polyacrylamide [PAA]) between Sia and biotin should decrease the steric constraints and allow ligand accommodation. Immobilisation of various concentrations of biotinylated Sia was achieved by utilization of the biotin-specific neutravidin protein which was initially immobilised on NUNC MaxiSorp 96 well plates. Subsequently, the surface was incubated with 0.5 mM biotin to cap unsaturated neutravidin binding sites and finally tested with SiaP protein and commercial lectins (WGA and SNA). The result of this test for the SiaP protein was negative which indicates that SiaP does not bind to biotinylated Sia (Figure 3.76). No signal increase was also detected for WGA and SNA when applied on the surface treated with 1 µg/ml biotinylated Sia (Sia-biot [1 µg/ml] in Figure 3.76). The signal for WGA and SNA doubled in comparison to the negative control (PBS) when applied on the surface treated with 10 µg/ml biotinylated Sia. However, the same surface tested with anti-biotin antibodies yielded also twice as much signal as a reference surface (PBS). This indicates that the biotinylated Sia at the concentration of 10 µg/ml was successfully immobilised on the surface; however, SiaP showed no specificity for this conjugate. The results of ELLA experiments suggest that SiaP protein does not bind to conjugated Sia, neither terminal Sia of glycans nor PAA-linked Sia.

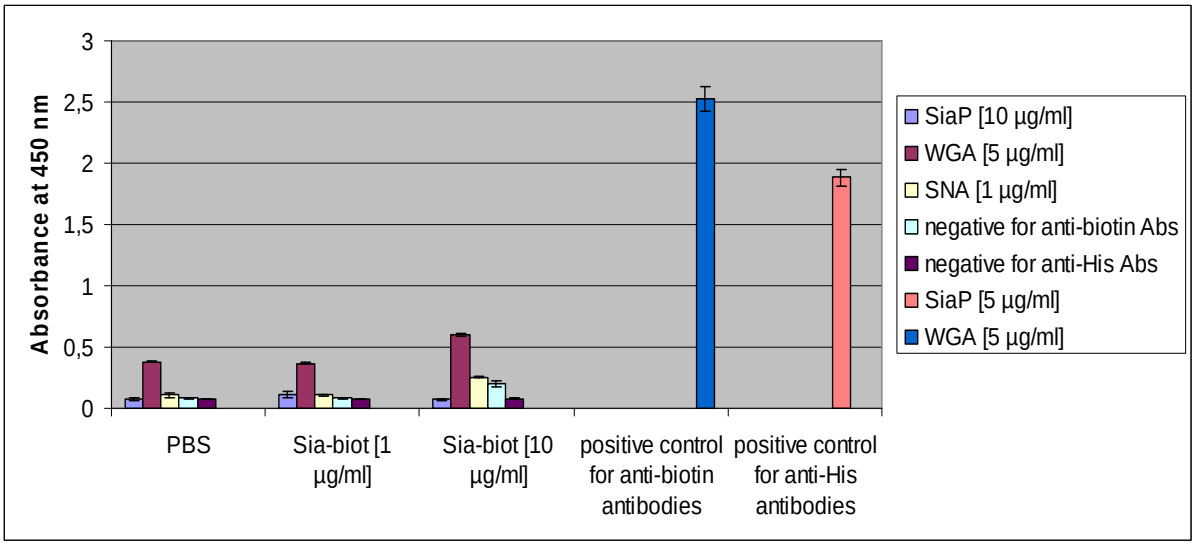


Figure 3.76. ELLA analysis of the SiaP specificity for sialylated conjugates. SiaP protein was tested against biotinylated Neu5Ac immobilised on the plate via neutravidin. The PBS buffer was used as a negative control for Sia binding proteins. The negative controls for antibodies consisted of neutravidin/biot-Neu5Ac (or PBS) untreated with Sia binding proteins. The positive controls for anti-His and anti-biotin antibodies were SiaP protein [5 µg/ml] and WGA [5 µg/ml] respectively immobilised directly on the plate. As controls, two other Sia binding proteins were tested: WGA and SNA.

3.3.6 Utilization of SiaP protein for sialic acid determination

In order to utilize the SiaP protein for detection and quantification of free Sia, the activity of the protein was tested with various analytical platforms. Results described in this section involve experiments performed using Mass Spectrometry and surface plasmon resonance.

3.3.6.1 Mass Spectrometry based free sialic acid quantification using SiaP

It was previously found (section 3.3.4) that the molecular mass of the SiaP protein can be precisely measured using MS. The Sia-free and Sia-bound species of SiaP can be distinguished on the basis of their mass to charge ratios. It was also found that after addition of the free Sia to the SiaP sample, the only species detected was the ligand-bound SiaP (Figure 3.75). The aim of this section of the study was to find a range of Sia concentrations in which the changes of Sia concentration linearly correlates with changes of ligand-free and ligand-bound concentrations of SiaP. This would subsequently lead to the development of Sia quantification method based on MS and SiaP protein.

The sample of SiaP was incubated with different concentrations of Sia and as the concentration of Sia was gradually increasing the decrease of Sia-free SiaP was observed using MS (Figure 3.77). At the same time the Sia-bound SiaP fraction was increasing.

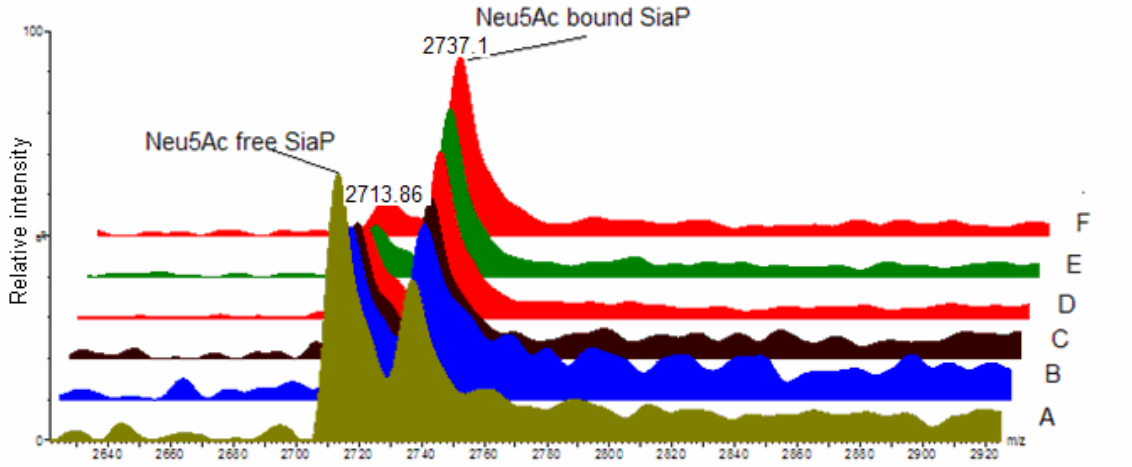


Figure 3.77. MS analysis of SiaP samples perincubated with different levels of Neu5Ac. The concentration of SiaP was the same in all samples (285 nM) and the concentration of Neu5Ac was as follows: A – 0, B – 10 nM , C – 100 nM, D – 285 nM, E- 400 nM, F - 1000 nM.

It was found that although the decrease in the amount of the ligand free protein and increase in the ligand bound protein correlates with the increase in the Neu5As concentration, the correlation is not linear. The concentration of SiaP in all samples was 285nM. The control sample with no external Sia added was composed of mostly Sia free SiaP protein. In the sample with a 10 nM concentration of Sia a significant drop in Sia free SiaP and increase in Sia bound SiaP was observed (Figure 3.77 B) in comparison to the control sample with no additional Sia added (Figure 3.77 A). However, in the sample with a ten times higher concentration of Sia no further changes were observed (Figure 3.77 C). The addition of an equimolar amount of Sia to the SiaP sample caused larger changes in Sia free and Sia bound fractions of SiaP; nonetheless, the Sia free compounds were still detectable. Further addition of Sia caused only slight changes to the ratio of ligandfree and ligand-bound SiaP protein concentrations. The lack of linearity in the SiaP response to various Sia levels may be attributed to the intrinsic features of the protein. The SiaP protein solution may interact in a non-linear way with increasing level of Sia. It is also possible that the insufficient resolution of the available MS equipment influenced the results as the difference in m/z values between unsubstituted and substituted forms of SiaP is minute (2937 and 2953 respecively – see Figure 3.70).

3.3.6.2 Surface plasmon resonance-based free sialic acid quantification using SiaP

Neutravidin and biotinylated sialic acid tests

In order to test usability of a surface plasmon resonance for free Sia detection and quantification using the SiaP protein various formats were used initially. One approach involved immobilisation of neutravidin, addition of biotinylated Sia and probing with SiaP (Figure 3.78). The SiaP protein was shown in above ELLA experiments not to bind conjugated Sia; hovewer, in the below presented tests a different analytical platform is used (surface plasmon resonance). This test was designed for development of a competitive assay for free Sia where the SiaP would be mixed with Sia and then applied on the chip. As the SiaP has a relatively large mass (35.5 kDa) its interaction with the surface would be clearly detectable, increasing the sensitivity of the assay.

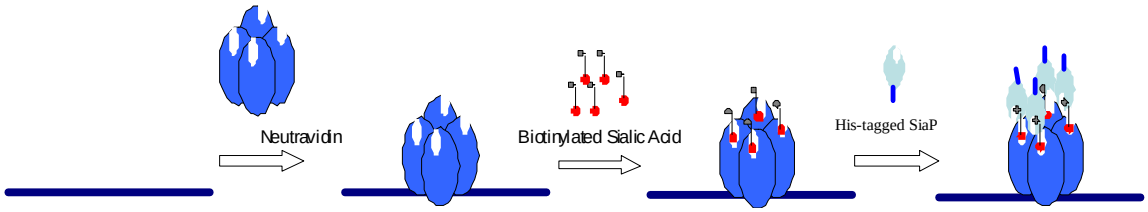


Figure 3.78. Immobilisation of biotinylated Sia on CM5 chip via Neutravidin and subsequent treatment with SiaP.

The immobilisation of neutravidin on the CM5 biacore chip was optimised. Preconcentration of 50 µg/ml protein was performed using a standard procedure (see section 2.24 above). It was found that the optimal pH of the sodium acetate buffer was pH 4.4, as the response indicating protein interaction with the dextran surface is the highest when sodium acetate buffer at this pH was selected Figure 3.79).

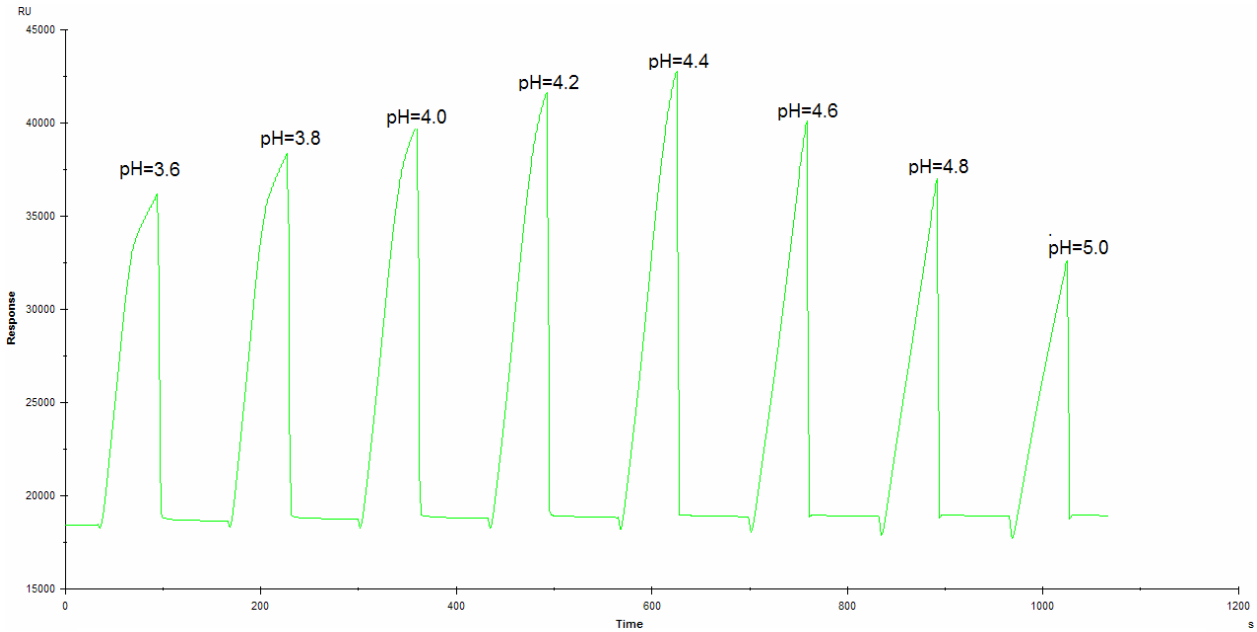


Figure 3.79. BIACORE 3000 sensogram representing the preconcentration of neutravidin protein [50 µg/ml] on the biacore CM5 chip. The pH of the sodium acetate buffers used for each injection are marked over each peak.

The neutravidin was then successfully immobilised on a CM5 chip via standard EDC/NHS coupling. After application of 200 µl of 50 µg/ml neutravidin the chip was capped with ethhanoane and unspecifically bound material was removed by four injections of sodium hydroxide. The baseline was found to be stable after NaOH injections (Figure 3.80).

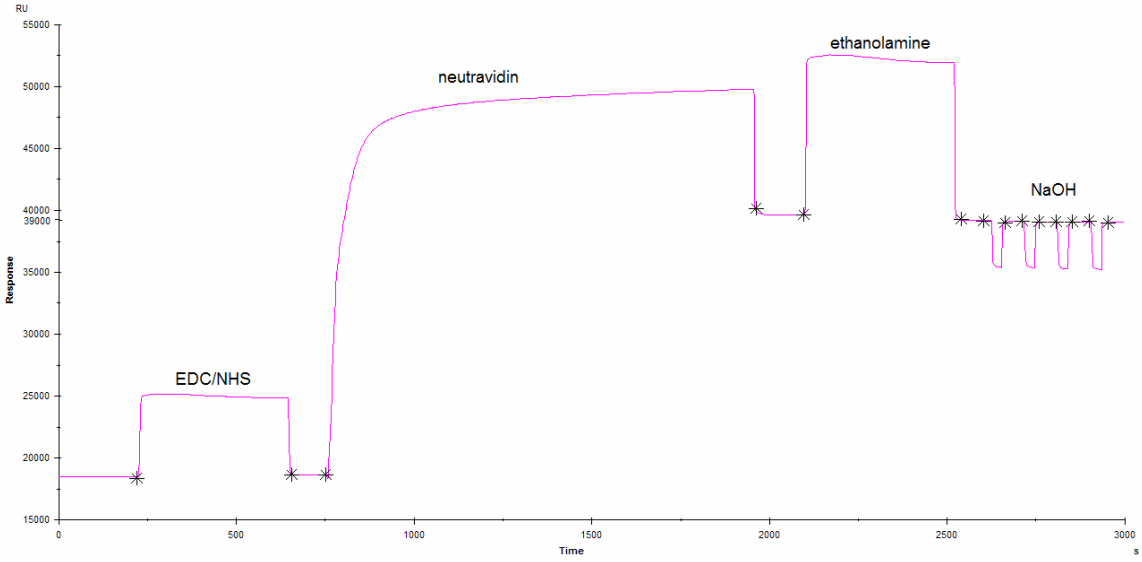


Figure 3.80. BIACORE 3000 sensogram representing neutravidin immobilisation on the CM5 chip. The asterisks indicate the beginning and end of injections. The injections of EDC/NHS, neutravidin, ethanolamine and NaOH are marked over the areas corresponding to each injection.

The neutravidin surface was subsequently used to capture biotinylated Sia. The biotinylated Sia was successfully captured on a neutravidin-coated CM5 chip (Figure 3.78). It was observed that 791 RU of biotinylated Sia were captured on this surface

after 150 µl of 50 µg/ml sample was injected (concentration previously optimised, data not shown, Biacore, 2003). Binding was represented by a typical increase in RU, as illustrated in Figure 3.81.

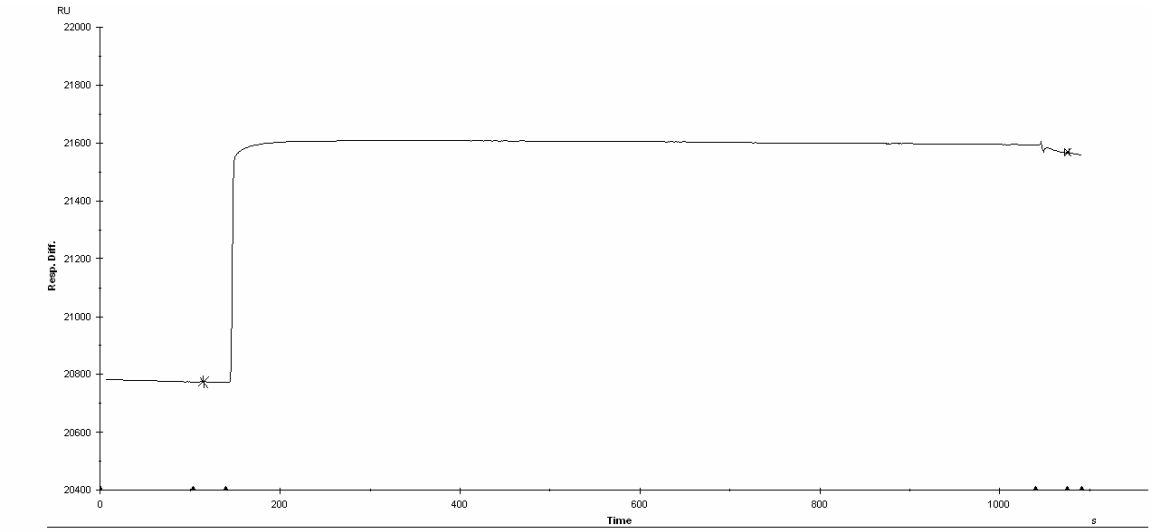


Figure 3.81. BIACORE 3000 sensorgram demonstrating binding of biotinylated Sia to a neutravidin coated CM5 chip. The asterisks indicate the beginning and end of injection of biot-Neu5Ac (150 µl of 50 µg/ml).

Finally, the SiaP protein was passed over the surface. A 50 µl sample of 50 µg/ml was selected and the response was monitored on the sensorgram presented in Figure . No stable change in response was observed in the sensogram. A rise of response straight after the start of the injection was detected but it declined afterwards resulting in overall baseline drift of 35 RU (Figure 3.82). The molecular mass of SiaP is 35.5 kDa and, if bound to the CM5 chip, it would clearly be detected with the biacore system. Lack of a stable signal on the sensorgam indicates that there is no true binding between the neutravidin/biot-Neu5Ac CM5 chip and the SiaP. This eliminates the possibility of utilisation of this particular format and confirms earlier ELLA-based analysis, which shows that SiaP can only bind to a free (and not conjugated) Sia.

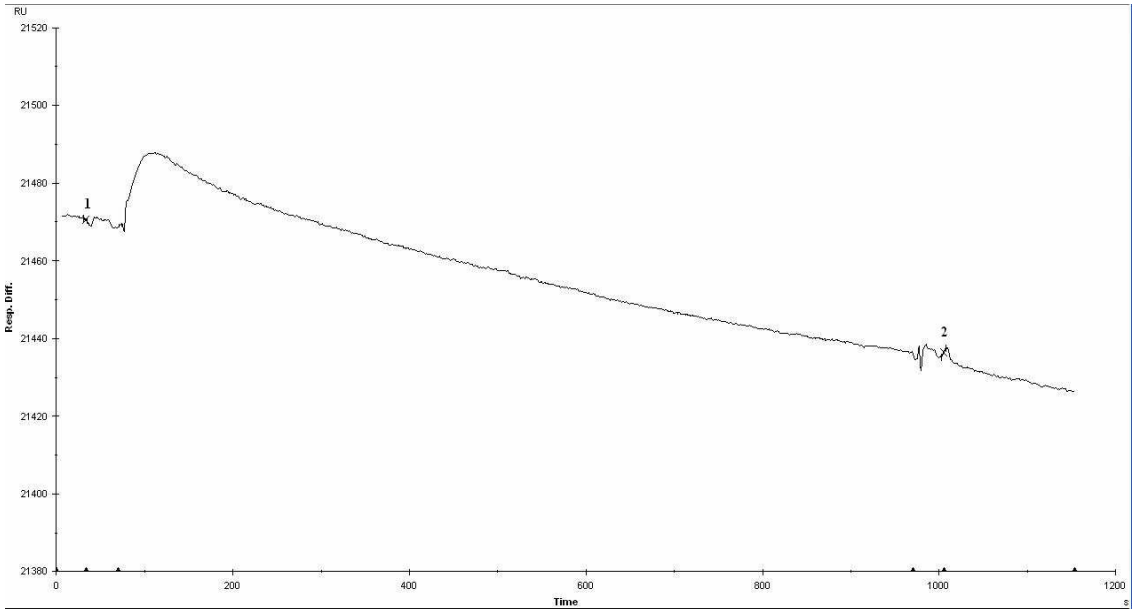


Figure 3.82 BIACORE 3000 sensorgram demonstrating injection of the SiaP protein on a neutravidin/biot-Neu5Ac-coated CM5 chip. Points 1 and 2 indicate respectively start and stop of the SiaP (50 µl of 50 µg/ml) injection.

Immobilisation of and Strep2-tagged SiaP protein via neutravidin

In the subsequently tested formats the SiaP protein was immobilised on CM5 biacore chips and used to detect free Sia. Two of the formats applied involved indirect immobilisation of SiaP on the biacore CM5 chip via either a deglycosylated version of avidin, neutravidin or antibodies. These would benefit from orientated immobilisation and enable regeneration of the surface as the neutravidin (or antibody)/SiaP interaction is not covalent. In the first format the SiaP 2031 (C-terminally Strep2 tagged with a linker) was immobilised on a neutravidin-coated CM5 chip via the Strep2 tag (Figure 3.83). Sia was subsequently passed over the sensor surface at varying concentrations and the changes in signal were monitored.

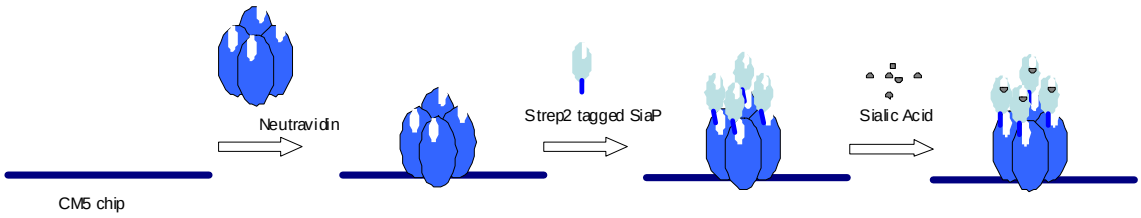


Figure 3.83. Immobilisation of Strep2-tagged SiaP via Neutravidin and subsequent treatment with free Sia.

The neutravidin was successfully immobilised on a CM5 chip via standard EDC/NHS coupling. Application of 200 µl of 50 µg/ml neutravidin yielded a stable baseline (baseline did not shift ater regeneration with NaOH) at 39.400 response units (RU) (Figure 3.84).

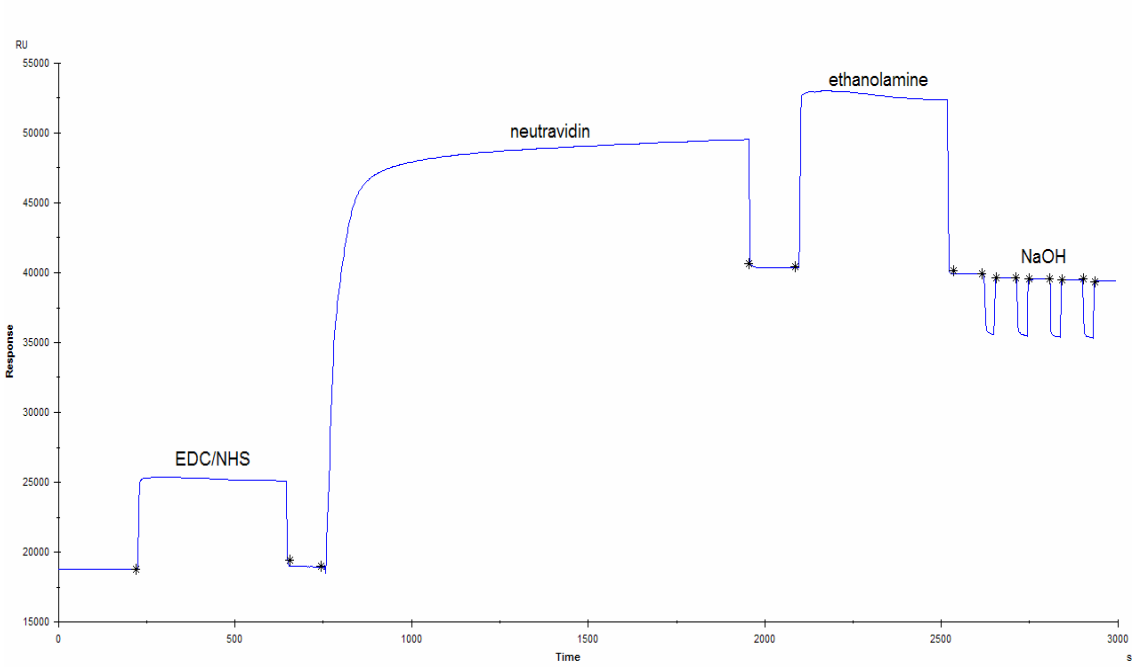


Figure 3.84. BIACORE 3000 sensogram representing neutravidin immobilisation on the CM5 chip. The astersks indicate the beginning and end of injections. The injections of EDC/NHS, neutravidin, ethanolamine and NaOH are marked over the areas corresponding to each injection.

The neutravidin surface was used to immoblise SiaP 2031 protein (Strep2-tagged with a linker). The injection of 50 µl of 50 µg/ml of SiaP protein resulted in an increase of 121 RU of the baseline (Figure 3.85). However, as the baseline was not stable after the injection and significant drift was observed, the injection of SiaP was repeated. An additional injection of 150 µl of 50 µg/ml of SiaP resulted in an overall increase of 275 RU. The baseline after SiaP immobilisation was found to be unstable and significant baseline drift was observed over time in the absence of any analyte (Figure 3.85). This baseline drift is attributed to the decrease of the mass of the molecules present on the surface of the chip. As the neutravidin was covalently linked to the dextran layer of the chip the mass change may be attributed to the leaching of the SiaP protein from the chip under the flow of the buffer (10 µl/min), which in turns indicates low affinity of the SiaP for the neutravidin surface.

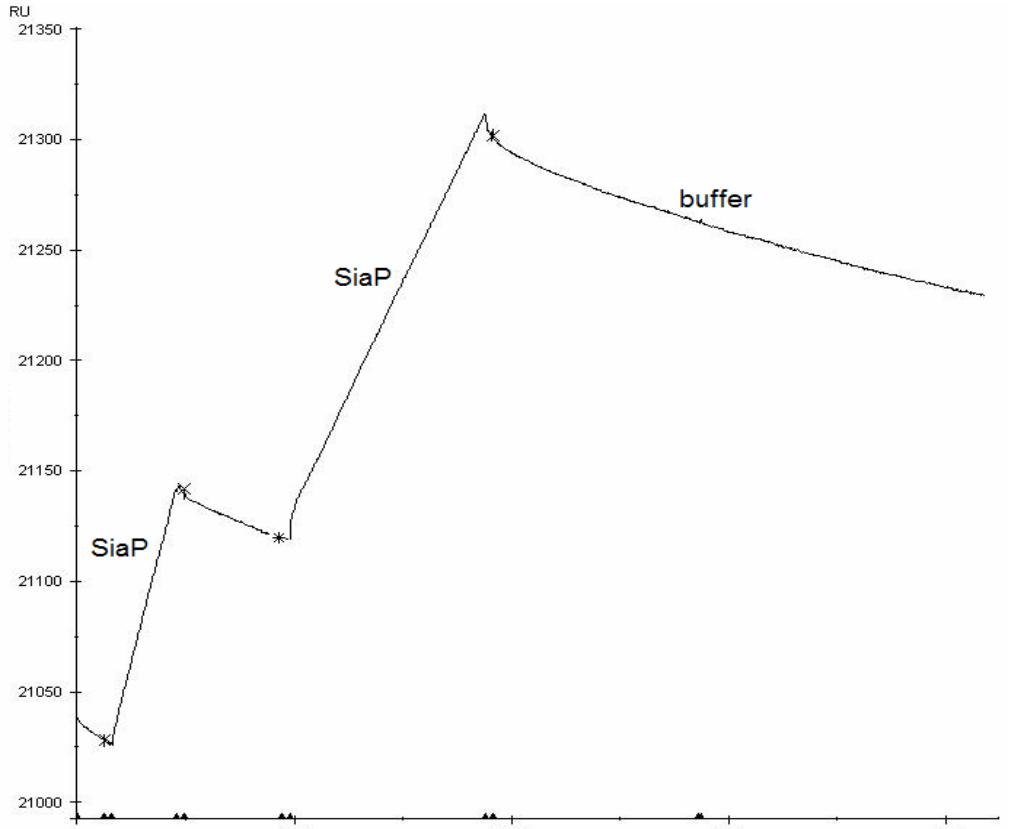


Figure 3.85. BIACORE 3000 sensorgram demonstrating binding of the SiaP protein to a neutravidin-coated CM5 chip. The astersks indicate the beginning and end of injections. During the first injection of SiaP 50 µl of 50 µg/ml protein was applied; during the second injection of SiaP 150 µl of 50 µg/ml protein was applied. The HBS buffer flow was monitored after each protein injection.

Application of various concentrations of Sia over neutravidin-SiaP surface

In order to test the surface response to Sia, various concentrations of Sia were selected. Application of various concentrations of Sia (1 nM – 1 mM) did not result in a noticeable increase in RU, or did not stabilise the baseline. The injection of a 5mM Sia sample resulted in a temporarily-increased signal (Figure 3.86). This phenomenon is caused by change of refractive index (RI) as the buffer injected had a different composition (higher Sia concentration in comparison to "pure" HBS buffer used between injections). The lack of a stable response after Sia injection may be attributed to insufficient quantities of SiaP on the sensor surface.

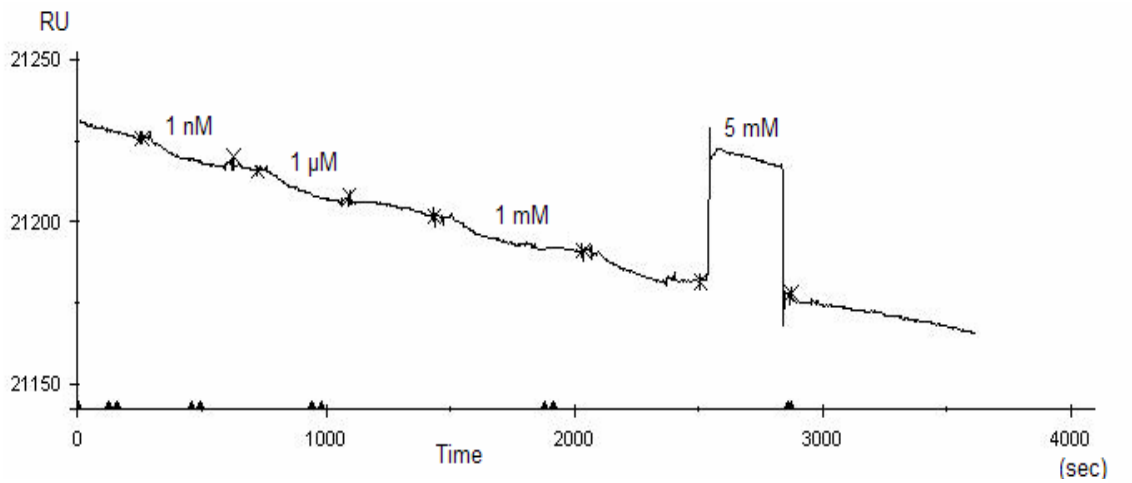


Figure 3.86. BIACORE 3000 sensorgram demonstrating injections of Sia at various concentrations on the SiaP coated surface. The astersks indicate the beginning and end of injections. The concentrations of Sia used are indicated above the parts of sensogram corresponding to them. Between the Sia injections the HBS buffer flow was monitored.

Immobilisation of the Strep2-tagged SiaP protein via Anti-Strep2 tag antibodies

In order to increase the SiaP protein concentration on the CM5 biacore chip the protein was immobilised via antibodies. Additionally, this format potentially allows for chip regeneration as the secondary antibodies could be stripped down. Goat anti-mouse IgG was immobilised directly on the CM5 chip and used to capture anti-Strep2 antibody (StrepMAB-Immo). Subsequently, the SiaP protein was immobilised on the chip via its Strep2 tag and used to capture free Sia Figure 3.87.

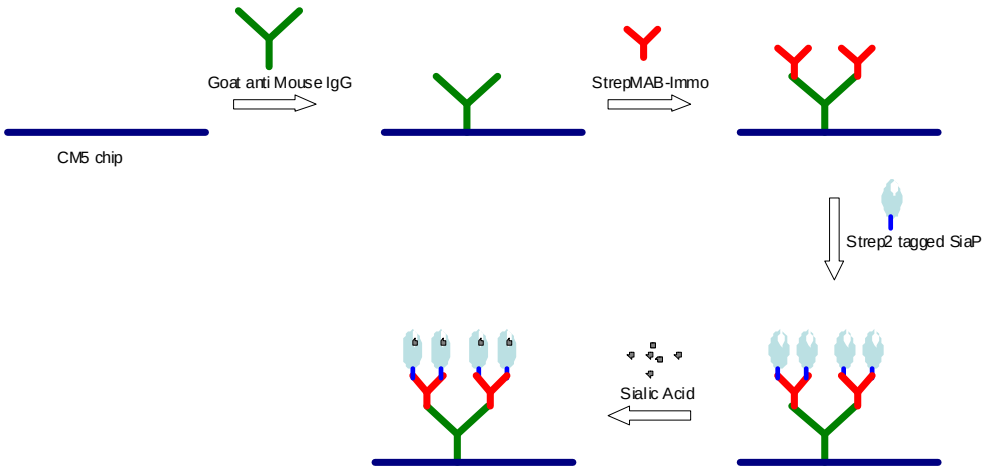


Figure 3.87. Immobilization of Strep2 tagged SiaP via a Goat anti-mouse IgG and a StrepMAB-Immo antibody, and subsequent treatment with free Sia.

The immobilisation of goat anti-mouse antibodies on a CM5 biacore chip was optimised. Preconcentration of 50 µg/ml protein (Biacore, 2003) was performed using standard procedure. It was found that the optimal pH of the sodium acetate buffer was 4.6 (Figure 3.88).

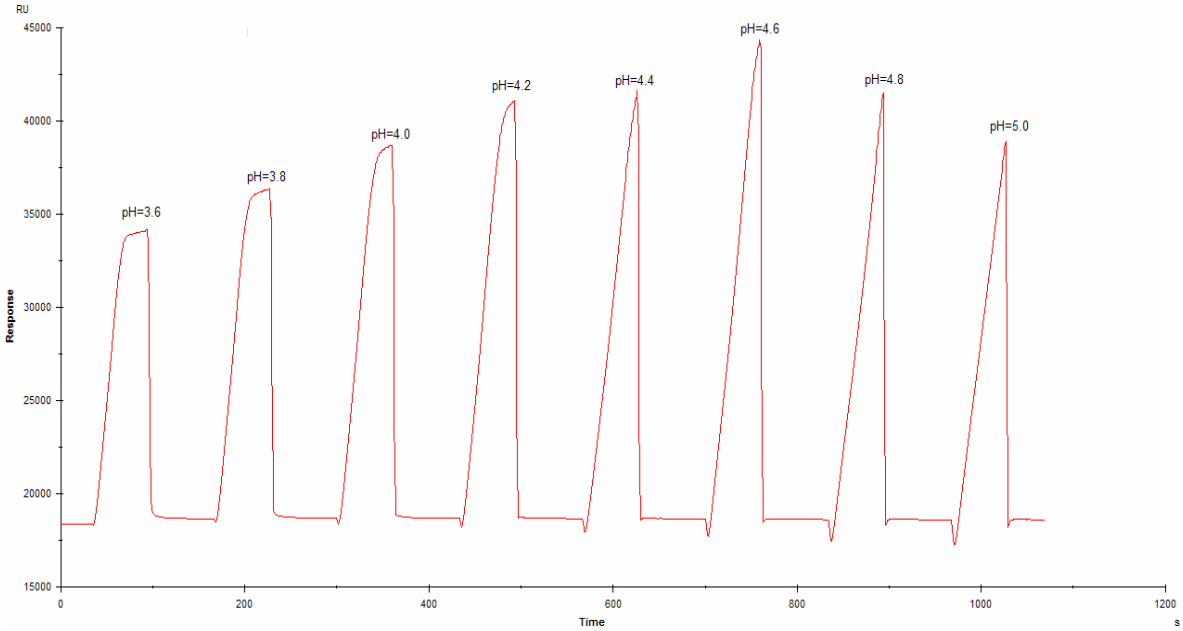


Figure 3.88. BIACORE 3000 sensogram representing the preconcentration of neutravidin protein [50 µg/ml] on the biacore CM5 chip. The pH of the sodium acetate buffer used for each injection is marked over each peak.

The goat anti-mouse antibody was then successfully immobilised on a CM5 chip via standard EDC/NHS coupling. Application of 200 µl of 50 µg/ml of the antibody yielded a stable baseline (baseline did not shift after regeneration with NaOH) at 36.400 response units (RU) (Figure 3.89).

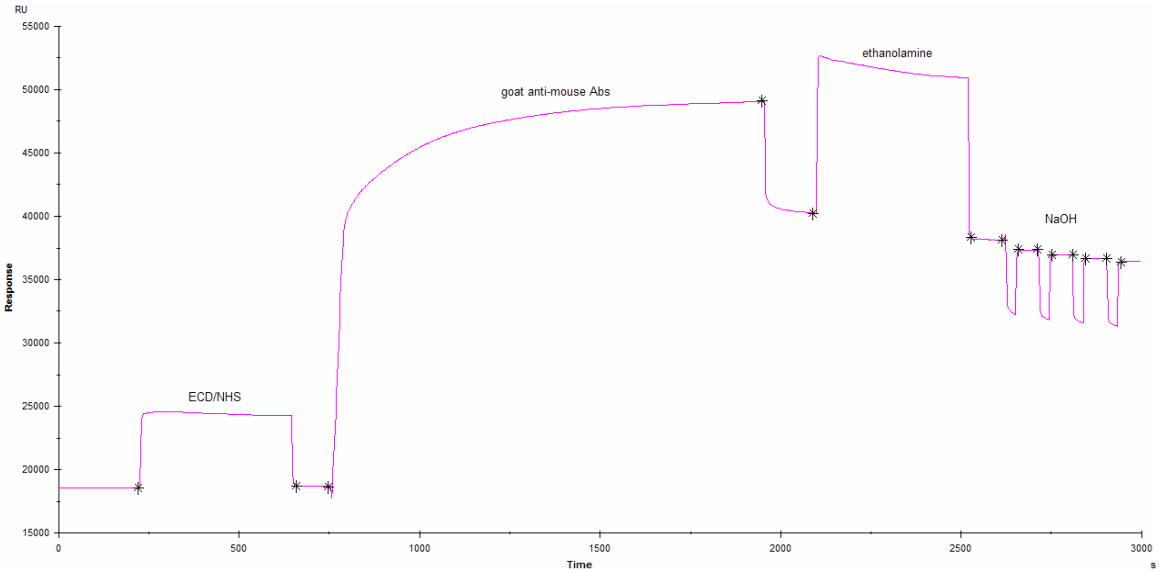


Figure 3.89. BIACORE 3000 sensogram representing goat anti-mouse antibodies immobilisation on the CM5 chip. The asterisks indicate the beginning and end of injections. The injections of EDC/NHS, goat anti-mouse antibodies, ethanolamine and NaOH are marked over the areas corresponding to each injection.

The goat anti-mouse antibody surface was used to immobilise mouse anti-Strep2 antibodies (StrepMAB-Immo). The injection of 200 µl of 50 µg/ml of StrepMAB-Immo resulted in an increase of 2391 RU over the baseline (Figure 3.90).

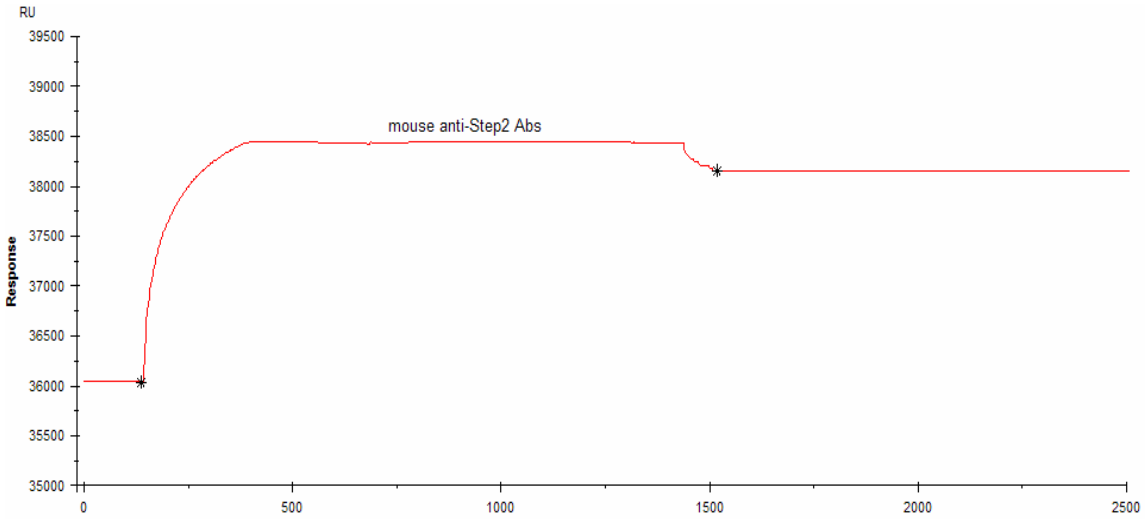


Figure 3.90. BIACORE 3000 sensogram representing mouse anti-Strep2 antibodies immobilisation on the goat anti-mouse antibodies coated CM5 chip. The asterisks indicate the beginning and end of injections. The injection of mouse anti-Strep2 antibodies is marked over the area corresponding to each injection.

Subsequently the SiaP protein was immobilised on the antibody surface. The injection of 200 µl of 300 µg/ml of SiaP 2031 protein resulted in an increase of 207 RU in the baseline (Figure 3.91). After the injection the baseline was not stable and a drift of the baseline was observed; however, the drift was not as sharp as when the SiaP was immobilised on the neutravidin surface, indicating that the antibody surface has a higher affinity for SiaP than neutravidin. The immediate drift of the baseline after the start of injection and sharp rise at the end of the injection is attributed to the buffer composition change. The sample of SiaP had a relatively high concentration (300 µg/ml of protein) in comparison to the HBS buffer that flows through the flowcell before and after the injection. The unstable interaction between SiaP protein and the surface was not optimal for the development of a quantitative test; however, the surface was tested with Sia in order to investigate the responsiveness of the equipement.

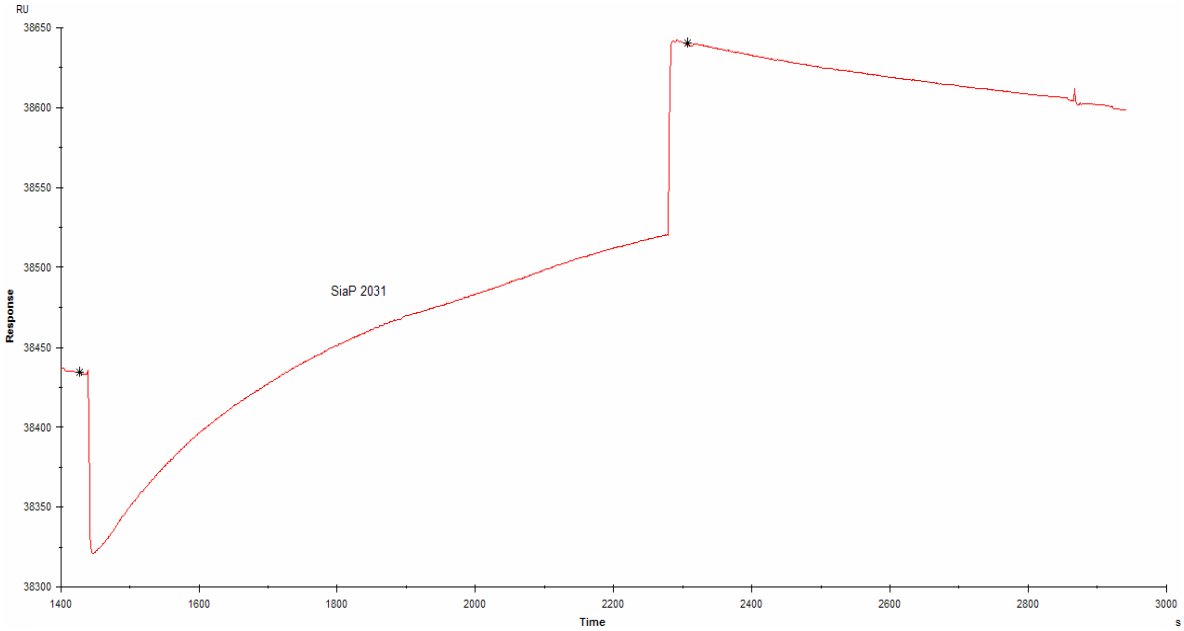


Figure 3.91. BIACORE 3000 sensogram representing SiaP immobilisation on the antibodies coated CM5 chip. The asterisks indicate the beginning and end of injections. The injection of SiaP is marked over the area corresponding to the injection.

Application of various concentrations of sialic acid over StrepMAB antibodies-SiaP surface

The antibody/SiaP surface was ultimately tested with Sia. The initial application of 5 mM Sia was found to cause a stable rise of an overall signal of 432 RU (Figure 3.92). The sharp increase of the signal after the start of the Sia injections and drift of the signal at the end of the injections were attributed to the change of the refractive index of the buffer caused by different buffer compositions during the injections. The active surface was subsequently treated with 10mM NaOH (regeneration step). An incline of the signal of 512 RU indicates that Sia was washed off the chip as the baseline returned almost to the state before Sia application. The sudden drift of the baseline at the beginning of the 10 mM NaOH application and sharp increase of the signal at the end on the injection were attributed to the buffer composition changes. The application of 5 mM Sia was repeated two more times. After each Sia application the increase in response units was observed which indicates that Sia was captured (Figure 3.92). However, injection of 1 mM Sia did not yield a stable increase of the signal. This indicates that the lower detection limit of this surface for Sia is between 1 mM and 5 mM. More importantly this indicates that it is possible to detect Sia using the SiaP protein and Biacore systems. Nonetheless, it is crucial for further tests to optimize SiaP immobilisation in order to have a stable baseline before Sia injections.

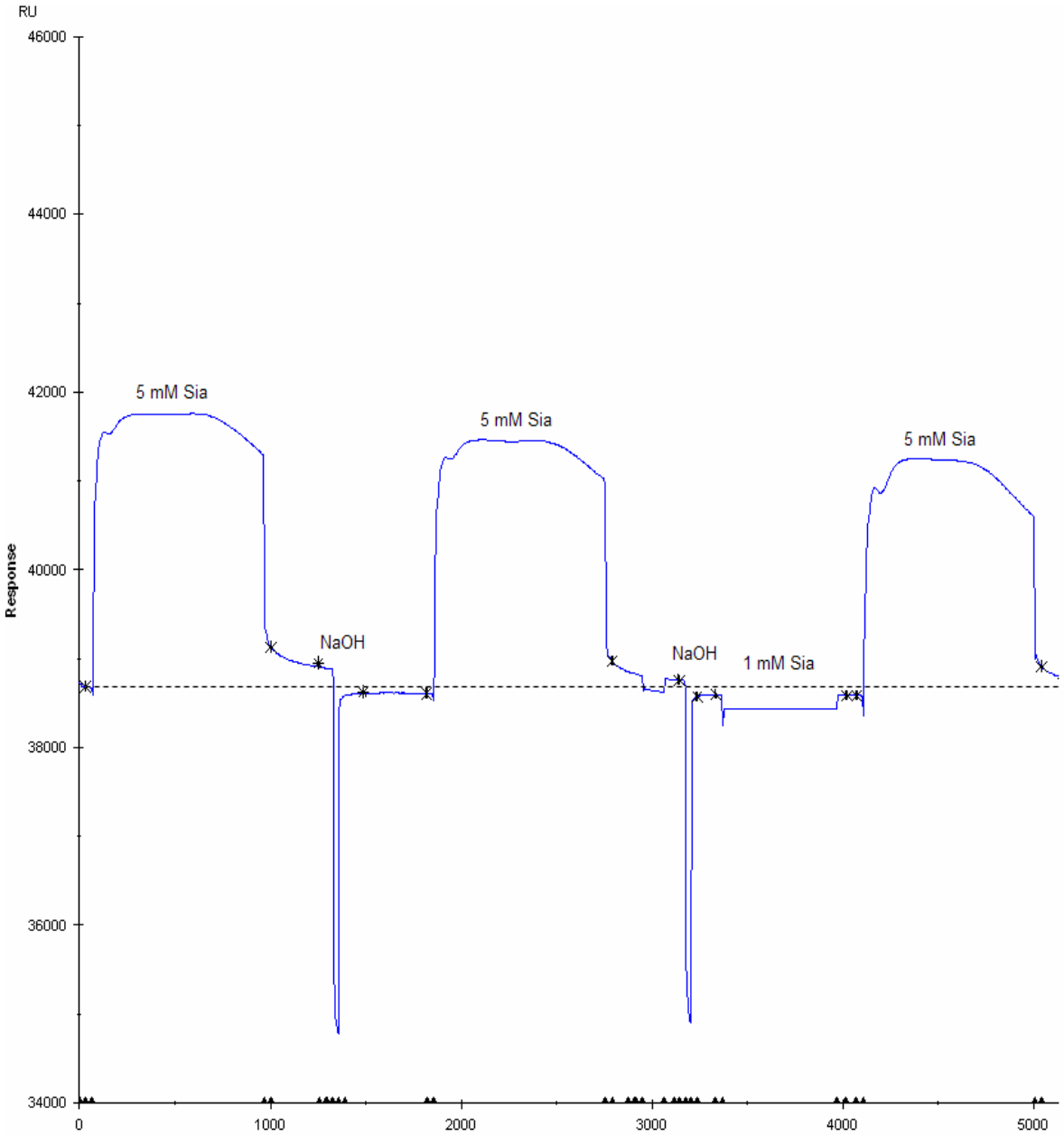


Figure 3.92. BIACORE 3000 sensorgram demonstrating injections of Sia at various concentrations on the SiaP coated surface. The asterisks indicate the beginning and end of injections. The concentrations of Sia used are indicated above the parts of sensogram corresponding to them. Between the Sia injections the HBS buffer flow was monitored.

Direct immobilisation of the 6xLys tagged SiaP protein on the CM5 chip

Another format tested in order to maximize the stability of SiaP on the CM5 biacore

chip surface involved direct immobilisation of the 6Lys-tagged SiaP (SiaP 2021) protein on the active surface. This simplified format offers higher stability of the immobilised SiaP protein as the protein is covalently immobilised on the dextran surface of the chip. The 6Lys-tagged SiaP protein was selected for this process as the C-terminally located lysine tag used for EDC/NHS standard amine coupling increases the possibility of orientated immobilisation (Allard *et al.*, 2001). Preconcentration of 100 µg/ml protein was performed using the standard procedure (see section 2.2.4 above). It was found that the SiaP protein unspecifically binds to the surface (Figure 3.93) as the signal was gradually increasing after each protein sample injection.

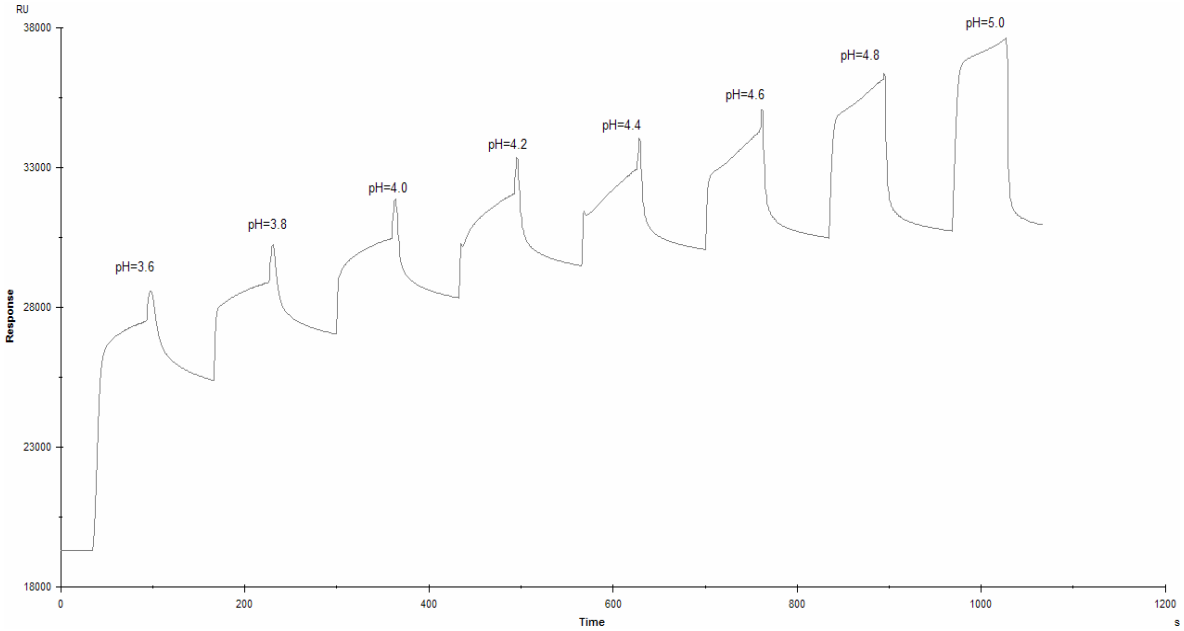


Figure 3.93. BIACORE 3000 sensogram representing the preconcentration of SiaP 2021 protein [100 µg/ml] on the biacore CM5 chip. The pH values of the sodium acetate buffer used for each injection are marked over each peak.

The preconcentraion procedure was modified by adding a NaOH regeneration step after each SiaP injection in order to remove the unspecifically bound SiaP molecules. The regeneration involved injection of 5 µl of 10 mM NaOH at 10 µl/min. Using this modified preconcentration procedure the immobilisation conditions for SiaP 2021 were optimised. It was found that the optimal pH of the sodium acetate buffer was 5.0 (Figure 3.94).

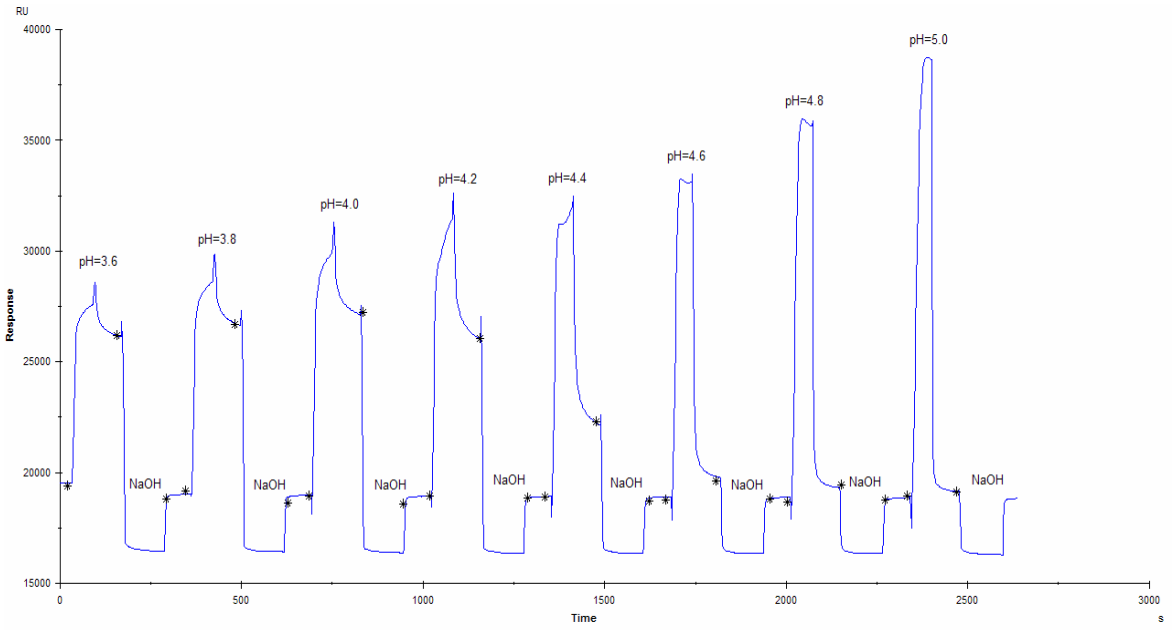


Figure 3.94. BIACORE 3000 sensogram representing the modified preconcentration of the SiaP 2021 protein [100 µg/ml] on the biacore CM5 chip. The asterisks indicate the beginning and end of injections. The pH of the sodium acetate buffer used for each injection are marked over each peak.

The SiaP 2021 protein was then successfully immobilised on a CM5 chip via standard EDC/NHS coupling. Application of 200 µl of 100 µg/ml of the protein yielded a stable baseline at 24,007 RU (Figure 3.95). After subtraction of the baseline before the process only 4657 RU of SiaP was immobilised.

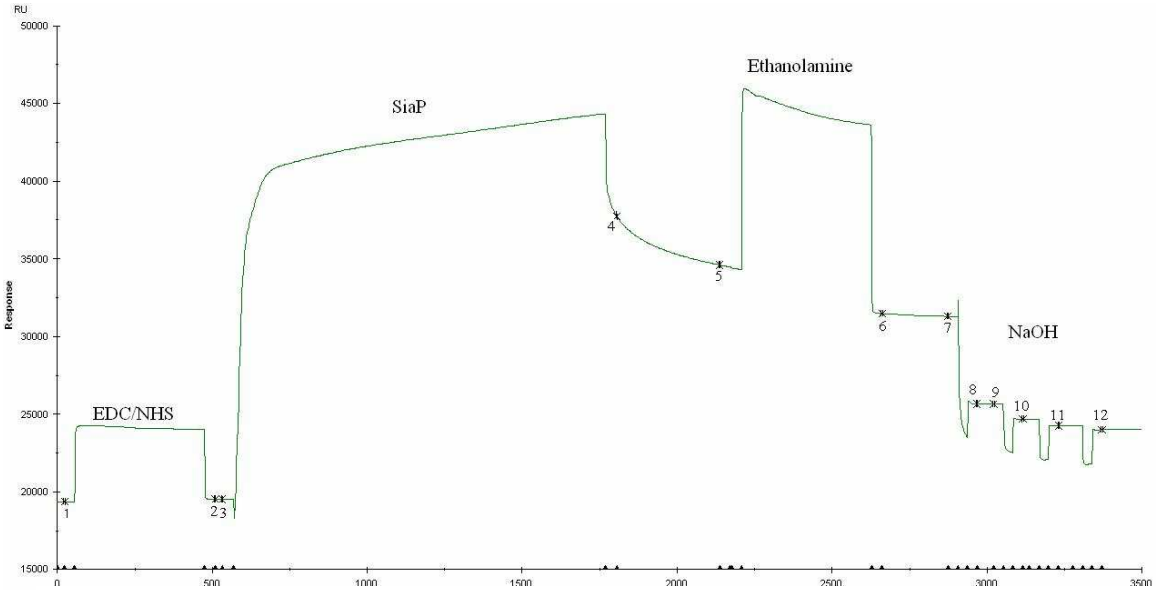


Figure 3.95. BIACORE 3000 sensogram representing SiaP 2021 immobilisation on the CM5 chip. The asterisks indicate the beginning and end of injections. The injections of EDC/NHS, SiaP, ethanolamine and NaOH are marked over the areas corresponding to each injection.

In order to confirm the activity of the immobilised SiaP in this format the chip was tested with 5 mM Sia. This concentration of Sia was previously found to be detectable with SiaP protein using biacore system (SiaP protein immobilised on StrepMAB antibodies). Injection of 5 mM Sia resulted in a drift of 5 RU (Figure 3.96) which indicates that there was no true binding between Sia and SiaP on the chip. This may be attributed to either the protein being inactive after direct immobilisation or the amount of protein immobilised on the chip being too low to detect the Sia.

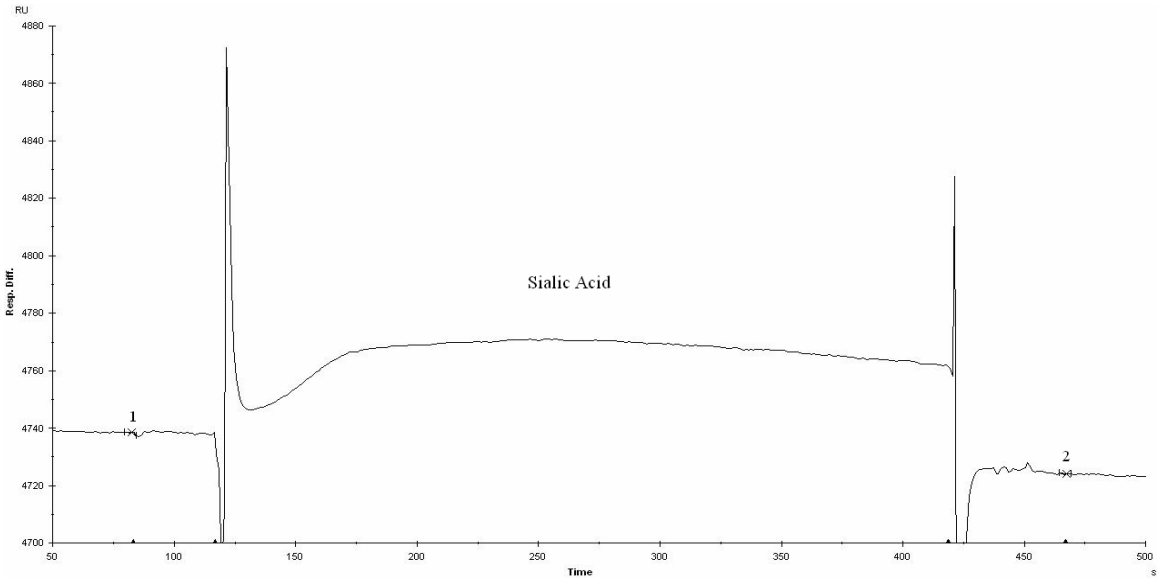


Figure 3.96. Activity test of SiaP directly immobilised on the CM5 chip with Sia. Points 1 to 2: injection of 50 µl of 5 mM Sia.

To increase the concentration of SiaP 2021 on the CM5 chip the immobilisation procedure was modified. A higher concentration of the protein was used (200 µg/ml) and the volume of the injection was increased to 300 µl. It was found that a significantly greater amount of SiaP was immobilised using these conditions: 12336 RU (Figure 3.97). Furthermore the baseline after SiaP immobilisation was stable. This was achieved by the application of amine coupling which generates stable covalent bonds between the dextran layer and the SiaP protein.

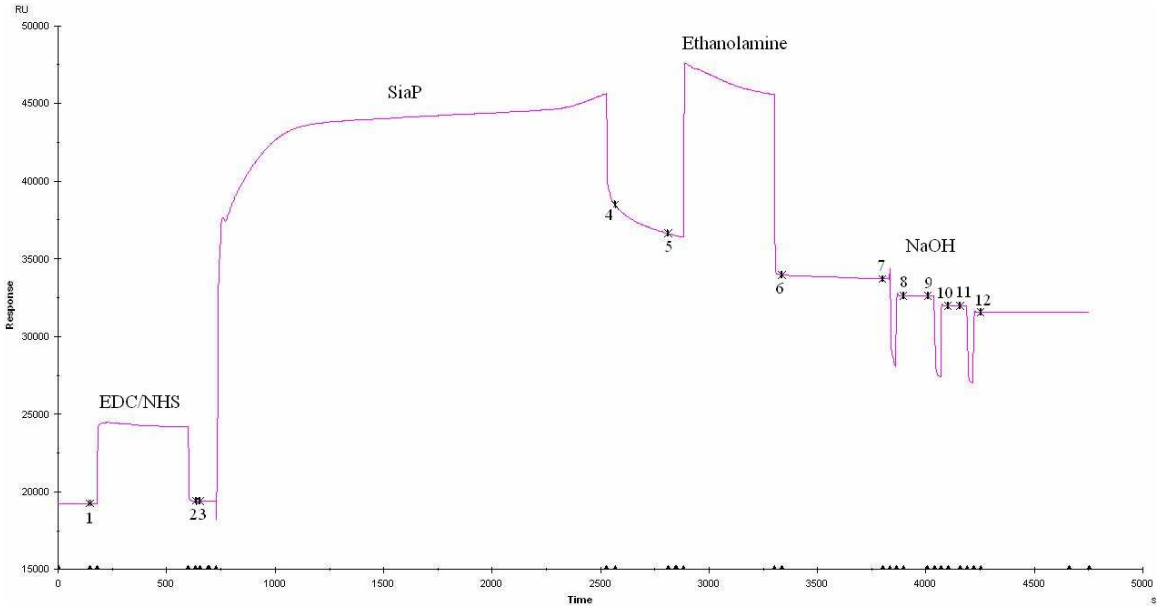


Figure 3.97. Immoblilisation of the SiaP protein on the Biacore CM5 chip. Points 1 to 2: surface activation with 70 µl of EDC/NHS; points 3 to 4 immobilisation of 300 µl of SiaP [200 µg/ml]; points 5 to 6: blocking with 70 µl of ethanolamine-HCl (pH 8.5, 1 M); points 7 to 12: regeneration with three 5 µl injections of NaOH [5 mM].

In order to confirm the activity of the immobilised SiaP the chip was tested with 5 mM Sia. Injection of the Sia resulted in an increase of 81.3 RU (Figure 3.98); however, on the reference surface (non activated carboxymethylated dextran) a similar increase of 79.4 RU (Figure 3.99) was observed.

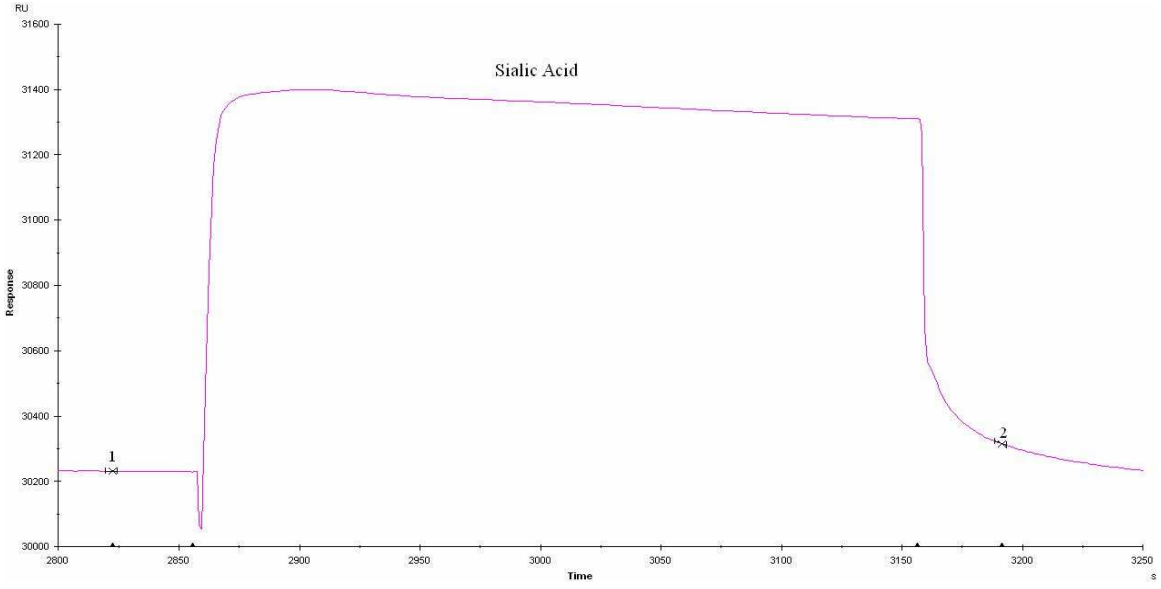


Figure 3.98. Activity test of SiaP directly immobilised on the CM5 chip with Sia. Points 1 to 2: injection of 50 µl of 5 mM Sia.

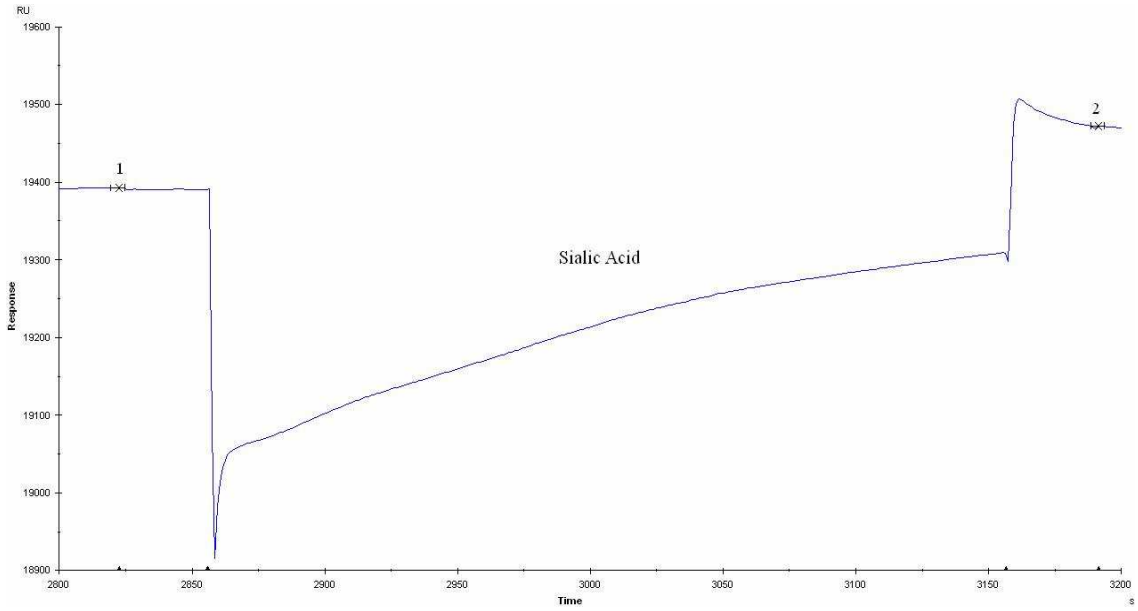


Figure 3.99. Sensorgam representing the injection of 50 µl of 5 mM Sia onto the reference surface during the activity test of immobilised SiaP protein.

To eliminate the Sia interaction with the reference surface, it was blocked by activation followed by capping. The dextran surface was activated by 140 µl injection of EDC/NHS followed by capping with 70 µl of ethanolamine and regeneration with NaOH (Figure 3.100).

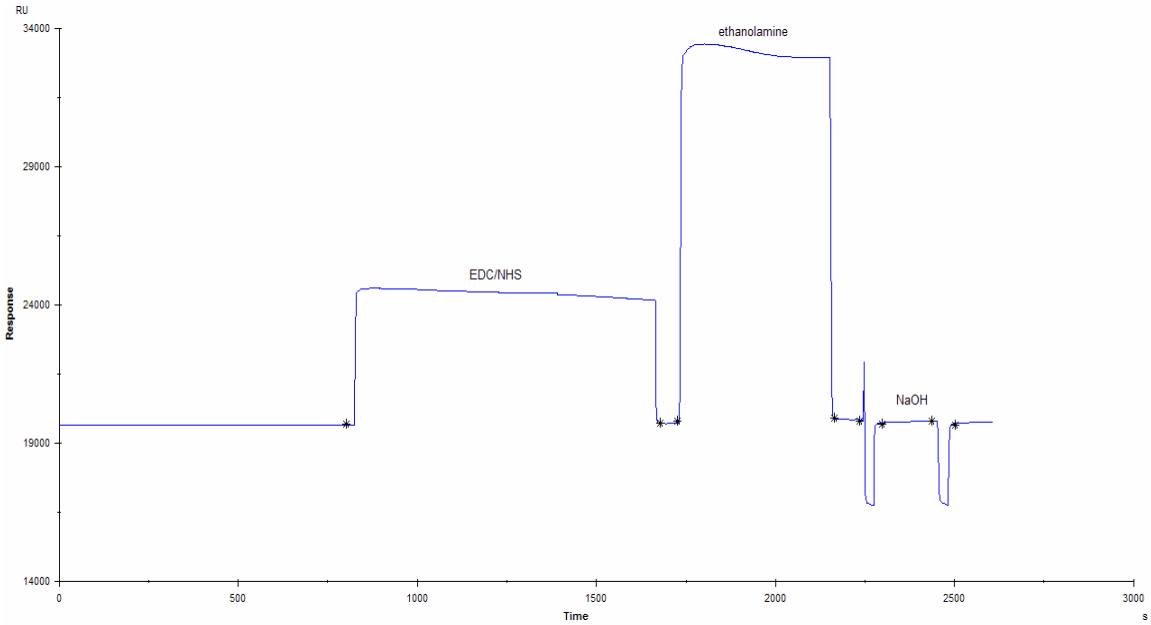


Figure 3.100. Blocking of the dextran surface by EDC/NHS activation and ethanolamine capping followed by NaOH removal of unspecifically bound material.

After blocking the reference surface, the activity test with 5 mM Sia was repeated. Injection of the Sia resulted in an increase of 203.1 RU (Figure 3.101). On the reference

surface an increase of 50.6 RU (Figure 3.102) was observed. The net increase after subtraction of the reference surface was 152.5 RU which indicates that true binding between Sia and SiaP immobilised directly on the chip takes place and that the protein is active in this format.

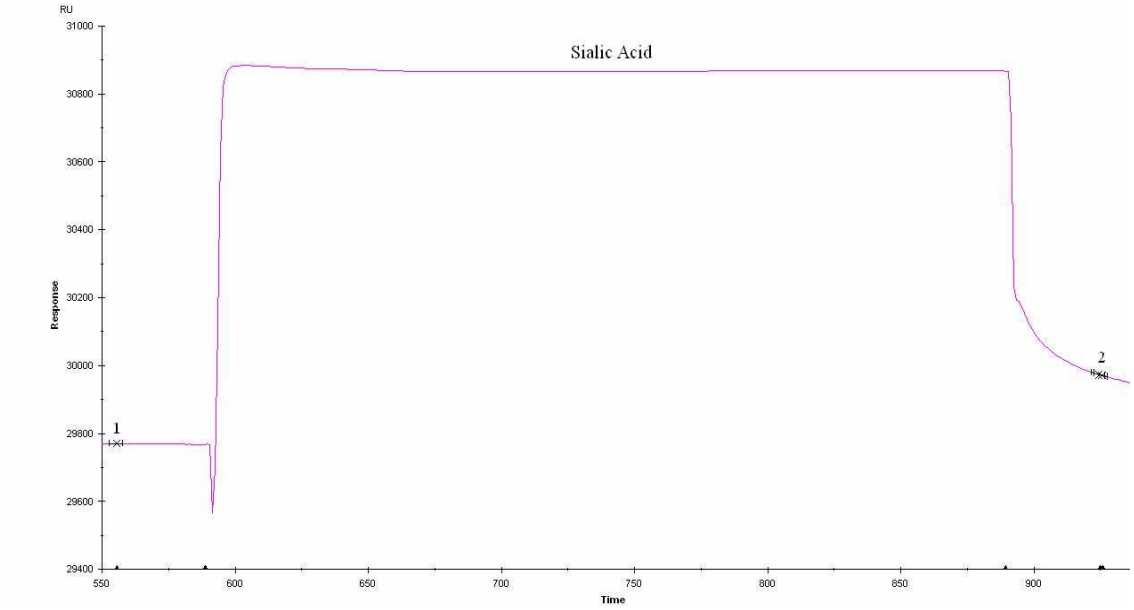


Figure 3.101. Activity test of SiaP directly immobilised on the CM5 chip with Sia. Points 1 to 2: injection of 50 µl of 5 mM SA.

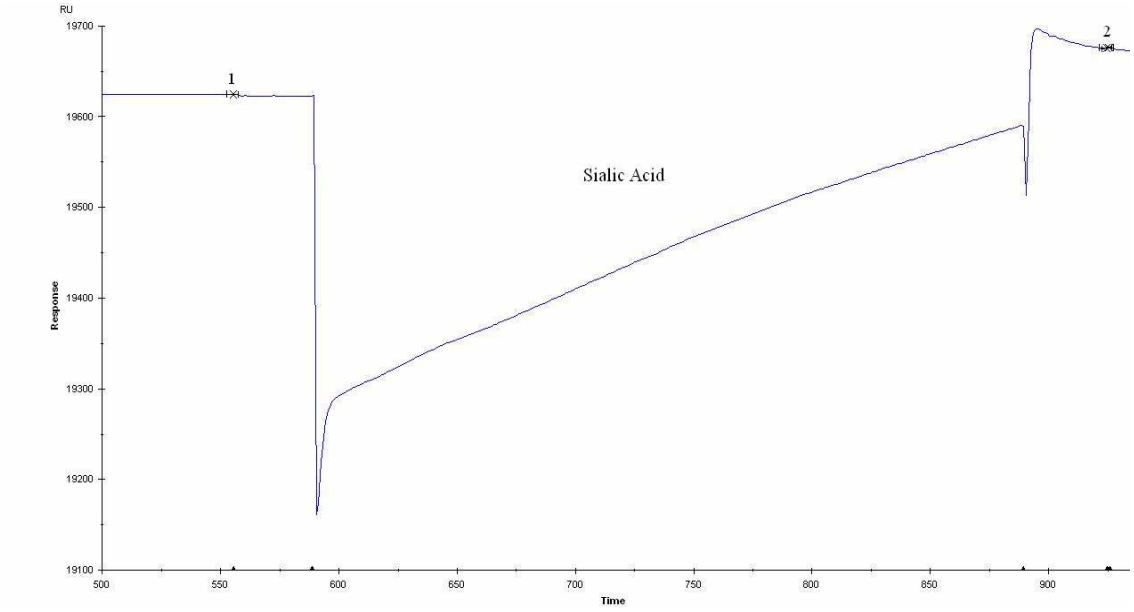


Figure 3.102. Sensorgam presenting the injection of 50 µl of 5 mM Sia onto the blocked reference surface during the activity test of immobilised SiaP protein.

Detection of various concentrations of sialic acid using SiaP on the Biacore chip

The response of SiaP directly immobilised on the CM5 chip was studied using various concentrations of Sia (in triplicates). Initially a range of Sia concentrations from 1 mM to 8 mM was studied. It was found that increasing concentrations of Sia resulted in the increasing response of the active surface. The injection of HBS buffer yielded 3 RU (Figure 3.103 and Appendix), 1mM Sia yielded 17 RU, 3 mM – 20 RU, 5 mM – 24 RU, 7 mM – 34 RU, 8 mM – 40 RU. The injection of 10 mM galactose used as a specificity control yielded 7 RU. Furthermore the response in a given range was close to linear (R^2 = 0,9283) (Figure 3.103 and Appendix).

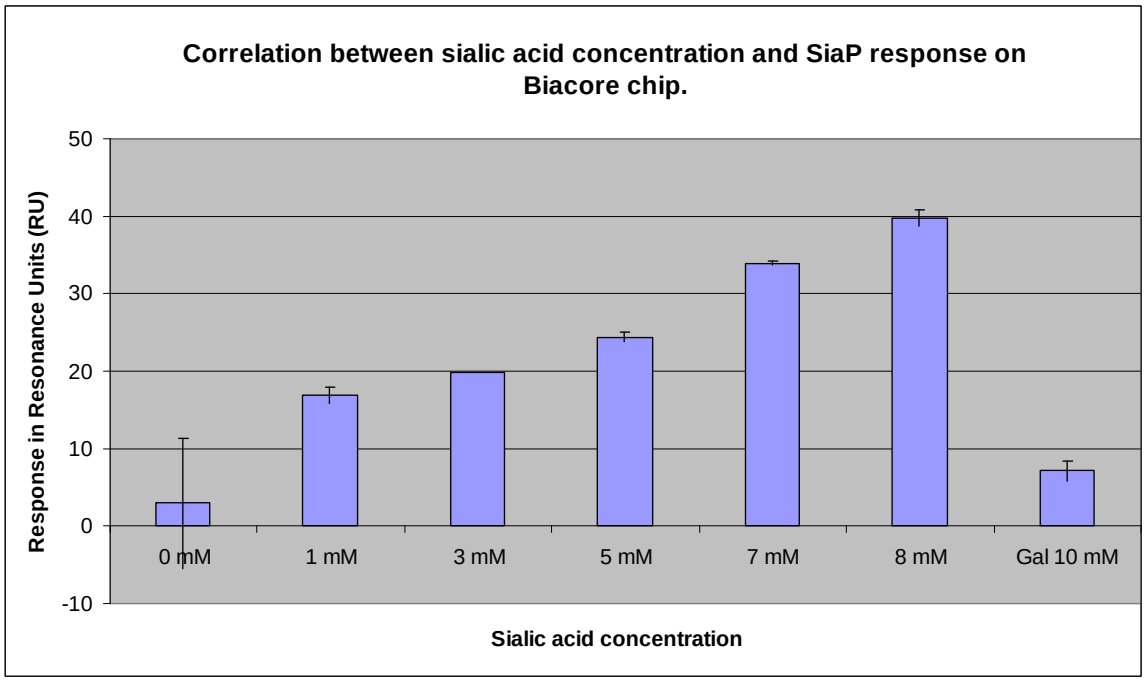


Figure 3.103. Correlation between Sia concentration in the range of 0 mM to 8 mM and SiaP protein response on the biacore chip.

In order to test the upper detection limit of the system higher concentrations of Sia were tested (in triplicates) and the results are plotted in Figure 3.104. It was found that the injection of 10 mM Sia yielded 166 RU (Appendix) which is unproportionally high in comparison to 8 mM result of 40 RU. The subsequent injection of 15 mM – resulted in 135 RU, which is a decline in comparison to 10 mM 166 RU. The result of injection of 20 mM was again lower than previous – 49 RU. The active surface was ultimately tested with 8 mM Sia to verifiy its activity. It was found that injection of 8 mM Sia resulted in 40 RU (Appendix), which is in concert with the first 8 mM test. This indicates that the active surface was fully functional. The specificity of the method was

tested with 10 mM galactose. It was found that the galactose injection resulted in the increase of the baseline of 7 RU (Appendix), which is similar to the HBS buffer value and thus indicates no interaction with the surface of the chip.

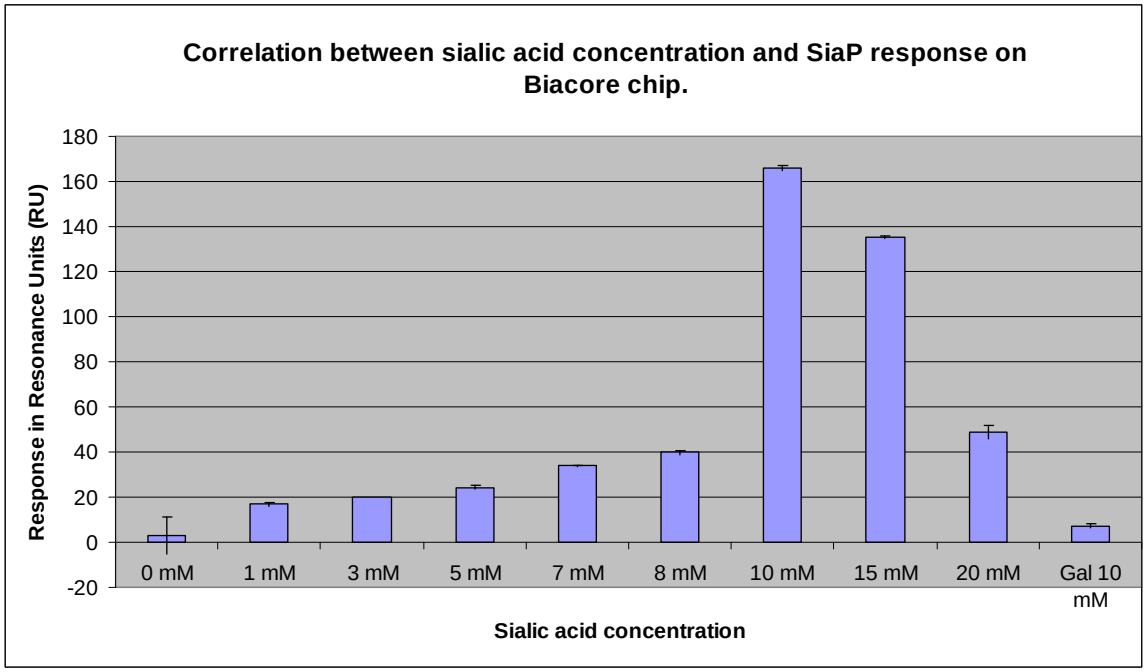


Figure 3.104. Correlation between Sia concentration in the range of 0 mM to 20 mM and SiaP protein response on the biacore chip.

Chapter 4
Discussion

4 Discussion

As the ultimate goal of the project was to apply Sia binding lectins as an analytical tool the goal of this section was to produce active HA1 protein expressed in *E. coli*. The *ha1* gene was initially cloned into the pKK233 derived pLecB3 expression vector containing an additional Ribosome Binding Site and His-tag sequence to facilitate efficient expression and convenient further purification of the HA1 protein. Another advantageous feature of the pLecB3 vector is the strong IPTG inducible Ptac promoter which allows for regulated expression of the *ha1* gene.

E. coli BL21 and Rosetta strains were initially chosen as hosts for expression. These are protease deficient strains that should limit heterologous protein degradation. Furthermore the Rosetta strain contains an additional plasmid with tRNAs for mammalian codons that rarely occur in *E. coli*. The initial expression of the HA1 protein in *E. coli* from the pJS102 plasmid confirmed the successful use of the expression system utilized. However, the successfully expressed HA1 protein was exclusively produced in an insoluble form in inclusion bodies. Expression of the protein in a number of *E. coli* strains was undertaken under a variety of expression conditions. However, the results of these experiments were not fully successful, regretfully, with respect to the production of a soluble HA1 protein. Furthermore, there was no control over the induction process caused by “leaky” expression at a substantial level. The “leaky” expression in the *E. coli* Bl21 strain is in accordance with previous reports (Gosse *et al.*, 1993; Tougu *et al.*, 1996). Therefore, the XL10-Gold strain was utilized to lower the pace of expression by regulated induction of the *ha1* gene. It is a common practice in protein expression studies to promote solubility by lowering the growth temperature and concentration of inducing agent (Sørensen *et al.*, 2005). The culturing conditions were optimized in order to slow down expression and thus enable correct folding of the HA1 protein that would increase its solubility. Such factors as culturing temperature and concentration of IPTG used for induction were tested. However, the HA1 protein was expressed exclusively in an insoluble form.

In order to produce the HA1 protein in soluble form two separate approaches were undertaken. One was based on alteration of the expression system in order to obtain a soluble protein. The second approach was focused on solubilisation of the initially expressed aggregates of HA1 protein.

As previously reported, the solubility of heterologous proteins expressed in *E. coli* is often increased by targeting to the periplasmic space as well as by fusion protein utilization (Pryor *et al.*, 1997; Dyson *et al.*, 2004). Both of the above strategies were used to promote solubility of the HA1 protein. The Sec translocation system was initially utilized for HA1 targeting into the periplasmic space. Although successfully expressed with a signal sequence the HA1 protein was insoluble. This may be attributed to the mechanism of transloaction by the Sec system which is a posttranslational system (Fekkes *et al.*, 1999). Thus, nascently produced polypeptide is translocated in this case after being fully synthesized. The aggregation that possibly occurred before the termination of the translation process might have prevented subsequent translocation.

Fusion proteins are often used for heterologous protein expression in *E. coli* offering several advantages like increased expression rates or simplified and efficient purification and also, most importantly for this study, increased solubility of the product (Pryor *et al.*, 1997; Dyson *et al.*, 2004). The maltose binding protein (MBP) was found to be the most effective fusion partner for promoting solubility (Dyson *et al.*, 2004). Thus the *ha1* gene was successfully cloned in frame with the *malE* gene encoding MBP in pMALp2E vector. The expression of MBP fused HA1 was unsuccessful as the overexpression of the MBP-HA1 was not detected by SDS-PAGE analysis neither in the soluble nor in the insoluble fractions of the cells. The lack of HA1 protein may be explained by the protease activity, which is present in the XL10-Gold strain of *E. coli* (Stratagene). However, the HA1 protein was successfully expressed in this strain before. The addition of MBP as a fusion partner seems to hamper the production of the protein. It is possible that the large size of the transcript which is 2235 bp corresponding to 745 amino acids (including the signal sequence) impairs or even makes the production of such a polypeptide impossible. Such results of HA1 expression in *E.coli* are in concert with previous reports where the entire influenza virus hemagglutinin protein was expressed with β-galactosidase as a fusion partner in a soluble form but in a very low quantity (Davis *et al.*, 1981).

Production of soluble mammalian proteins in *E. coli* by solubilisation and refolding of polypeptides expressed in the form of inclusion bodies is another widely applied practice (Vallejo *et al.*, 2004). Such a strategy was applied in parallel with previously discussed expression alterations as the results of this part of the study would be beneficial in two ways with regards to the ultimate goal of the project. Firstly, optimized

solubilization conditions would enable efficient HA1 purification and, as a result, promote refolding of the soluble active product (Fekkes *et al.*, 1999). Secondly, effective denaturation followed by successful refolding may be utilized for ligand release to regenerate the ligand free form of the protein. Concentrated solutions of urea as well as guanidine hydrochloride, respectively, were used in this study as they are reported to be the most potent chaotropic agents for solubilisation of protein aggregates (Vallejo *et al.*, 2004). The solubilisation of the HA1 protein using urea at various temperatures of incubation was unsuccessful. It is possible that urea is not a strong enough chaotropic agent to solubilise this particular protein. The unsuccessful solubilisation may also be attributed to the reported impact of the pH on the solubilisation process using urea, as urea at various pH values influences the solubilisation of a given protein differently (Estapé *et al.*, 1996).

The application of guanidine hydrochloride as a denaturant was initially unsuccessful with the concentration of the chaotropic agent of 6 M, which is the most widely used concentration reported (Vallejo *et al.*, 2004). As the unsuccessful solubilisation of HA1 protein by 6 M GnHCl may be attributed to the denaturant being still too weak to break the protein aggregates down, a higher concentration of GnHCl was applied. The solubilisation of HA1 protein by 8 M GnHCl was successful and soluble polypeptide was obtained.

In order to purify the solubilised HA1 protein, which would promote efficient refolding of the polypeptide, Immobilised Metal Affinity Chromatography (IMAC) was used. However, the solubilised HA1 protein was found not to bind to the preequilibrated resin of Ni-NTA agarose (Qiagen). The lack of binding may be attributed to the high concentration of GnHCl (8 M) during the procedure. Such a high concentration may be incompatible with the resin, as the manufacturer suggests using a 6 M concentration of GnHCl at most; nevertheless, a lower concentration of GnHCl was found to be ineffective for HA1 protein solubilisation. Application of other resins (Westburg, Invitrogen) might be more successful. However, at this point in the project it was concluded that HA1 protein is not readily amenable to production for the purpose of the project and it was decided to use another sialic acid binding protein.

The Sia-binding protein, SiaP, of *H. influenzae* was modified, produced, characterized and applied for sialic acid detection. The *siaP* gene was initially cloned into a pET-21b derived expression vector. Efficient expression of the gene was driven by the strong T7

promoter. The C-terminal 6His-tag was utilized to facilitate further efficient purification and immobilisation of the SiaP protein. Additionally C-terminal Strep2 and 6Lys tags were introduced which were subsequently utilized for immobilisation on the biacore chips.

As an expression host *E. coli* KRX (Promega) cells were chosen. The T7 polymerase encoded in the genome of KRX cells is under control of the rhamnose promoter/operator. This results in not only high but also regulated protein expression (Guzman *et al.*, 1995). Furthermore the ompT$^-$ and ompP$^-$ mutations eliminate proteolysis of the overexpressed proteins, whereas a lack of the most common nuclease that copurifies with plasmid DNA, achieved by endA$^-$ mutation, is an advantage during cloning steps (Promega).

The expression of the SiaP protein was optimized with regards to expression time, concentration of inducing agent and culture cell density at the induction time. The decreased concentration of soluble SiaP protein in cultures induced with a higher than optimal level of rhamnose and/or in cultures after prolonged (overnight) expression time was observed. This can be attributed to the deposition of SiaP in inclusion bodies which is a common phenomenon of overexpressed proteins in *E. coli* (Vallejo and Rinas, 2004). This reasoning is supported by the findings of the increased concentration of the SiaP protein present in the insoluble fractions of the respective cultures.

The optimal conditions were applied for expression of the soluble SiaP protein that was later purified using affinity chromatography. During the initial manual IMAC purification using NiNTA agarose columns it was found that the 6His-tagged SiaP 201 protein has low affinity for the resin and was not efficiently captured by the resin. This may be attributed to the low accessibility of the 6His-tag to the environment as it is directly at the C-terminus of the 35.5 kDa globular protein and the steric contraction may block it from interaction with nickel ions of the resin. In order to increase the accessibility of the 6His-tag the SiaP protein was engineered by introduction of the (Gly Gly Gly Ser)$_2$ linker/spacer between the globular protein and the tag. The introduction of a flexible peptide linker composed of Gly and Ser was previously reported to improve immunoreactivity of a fusion peptide by increasing the distance between domains of the polypeptide (Hu *et al.*, 2004). Extending the tag from the globular part of the SiaP protein by the utilization of the linker did indeed improve the purification efficiency by both nickel and StrepTactin affinity chromatography. The StrepTactin is a

streptavidin derivative protein with a high affinity for the biotin mimicking Strep2-tag (Skerra *et al.*, 2000). Purification over the StrepTactin coated resin also verifed the functionality of the C-terminal Strep2-tag.

Subsequently the molecular weight of the purified SiaP protein was accurately determined using MS. The results of these studies were in concert with the *in silico* predicted values. Interestingly, the SiaP protein analysed, which was isolated from a total cellular fraction, corresponded in size to the mature protein after cleavage of the initial 23-amino acid leader sequence directing the polypeptide to the periplasm. No larger polypeptides, containing leader sequence, were detected and thus further isolation of the SiaP was based on the total cell extracts, as there was no requirement for more complicated and time consuming periplasmic protein isolation. The highly precise MS method facilitated the distinction between molecules with similar molecular masses; hence, it was found that two - sialic acid-free and sialic acid-bound species of SiaP protein were present in the analysed material. The slight deviation of the analysed molecules from the predicted molecular masses may be attributed to the adducts of both water and ammonium formate. The presence of two forms of sialic acid binding SiaP protein is a result of the complex medium (LB) containing sialic acid utilized for expression. The impact of sialic acid present in the media on the SiaP protein was reported previously (Severi *et al.*, 2005). It was found that SiaP protein expressed by cells grown in LB medium was saturated with Sia.

SiaP with sialic acid bound is undesired for further application and in order to eliminate it the production process for the SiaP protein was modified. Two separate approaches were undertaken to obtain a homogenous, sialic acid free population of SiaP. The first was based on the fact that *E. coli* cells metabolise sialic acid and utilize it as a sole source of carbon (Vimr *et al.*, 1985). Post-induction expression time was extended in order to allow cells to metabolize sialic acid present in the expression media (LB) which would otherwise be bound by SiaP protein (Severi *et al.*, 2005). The methodology involving extension of the culturing time was found to be effective for sialic acid free production of the SiaP protein. In the second-developed method for sialic acid-free production of the SiaP protein the ligand was washed off from the protein by denaturation using chemical denaturants. The SiaP protein after immobilisation on the NiNTA agarose was denatured with urea or GnHCl. As the native three dimensional structure of the ligand binding pocket of the protein is required for sialic acid binding

the denaturation of the protein leads to sialic acid release. The efficiency of this method was verified by MS. It was found that application of this protocol leads to complete removal of the sialic acid from sialic acid-saturated SiaP sample. The activity of the sialic acid free molecule was also positively verified by MS. As the SiaP protein was found to refold spontaneously in lysis buffer, there was no requirement for application of special refolding methods or conditions, which may be time consuming and not always succesfull (Middleberg, 2002; Vallejo and Rinas, 2004; Katoh and Kaoh, 2000). This suggested the possibility of regeneration of sialic acid-free protein could be reused in analytical devices after diagnostic tests. Both of the developed methods for sialic acid-free SiaP protein production were found to be successful. The clear advantage of the first method is its simplicity, as no additional preparation steps are required. However, a limiting factor may be the changeable composition of the complex medium with regards to sialic acid content if provided by different manufacturers or when different lots of the media are utilised. Elevated levels of sialic acid in the medium may lead to decreased efficiency of this method. The second method of sialic acid free SiaP production based on the utilisation of chaotropic agents requires an additional step of protein denaturation. This method, however, offers reduced production time as the post-expression time is 19 hours shorter in comparison to the alternativ method (5 and 24 hours, respectively). Furthermore, the second protocol is much less prone to be affected by changeable free sialic acid content in the expression medium. In large scale production, the cost effectiveness of relatively cheap urea addition in the second method would most probably exceed the cost effectiveness of much longer expression time in the first method of the sialic acid free SiaP generation. It is also possible to utilise minimal medium for the protein expression and thus eliminate the sialic acid contamination (Severi *et al.*, 2005); however, the time and cost effectiveness of such protocols would be limiting factors. Thus, the denaturation method is suggested if high amounts of the SiaP protein are required.

Subsequently, the specificity of the SiaP protein for terminal sialic acid of glycoconjugates was tested using ELLA. The sialylation of the glycoproteins used in the tests was confirmed by positive results of the tests with commercial lectins (WGA and SNA). It was found that the His-tagged SiaP protein can be readily detected using HRP-conjugated anti-His antibodies when immobilised on the surface of the plate used in the tests. Hence, the negative results of the protein tested against sialylated

glycoprotein can be attributed to the lack of SiaP-glycan interaction and not to the detection problems. It was reported previously that the SiaP protein during the ligand accommodation drastically changes its conformation, burying the ligand deep into the binding pocket (Johnston *et al.*, 2007). Therefore, it is possible that steric constrains limit the binding of terminal sialic acid of the complex glycans by the SiaP, as the ligand is not accessible for the internally situated binding site. In order to test the SiaP specificity for less complex sialic acid conjugates the biotinylated sialic acid was used. The polyacrylamide linker connecting the two components is a much less complex structure than a glycan chain and thus should interfere less with the SiaP at the binding site. The negative results of ELLA experiments supported by similar findings using surface plasmon resonance, indicates lack of interaction between the SiaP protein and sialic acid conjugates. It is possible that the structure and mechanism of ligand accommodation of the SiaP protein limit the scope of the ligand to free entities and not conjugated ones. The elevated signal for WGA and SNA lectins observed in the same experiment may indicate that the biotinylated sialic acid was successfully immobilised using neutravidin coated surface. The increased signal was also found in the negative control for the anti-biotin antibodies. This also suggests successful immobilisation of the biotinylated sialic acid on the surface. However, this finding makes it unclear whether SNA and WGA lectins bound specifically to the sialic acid or underwent an unspecific interaction. The SNA and WGA were previously, successfully applied for detection of Sia in glycoconjugates using Biacore system (Kelly *et al.*, 2007). In order to investigate the specificity of SNA and WGA for biotinylated sialic acid a surface plasmon resonance analysis with immobilised target compound is suggested. Nonetheless, the biotinylated sialic acid was not captured by the SiaP protein and the interaction with other sialic acid binders is in question.

As it was found that the SiaP protein in the above described tests expressed exclusive specificity to free sialic acid, further experiments were focused on application of the protein for free sialic acid detection. The MS platform was found to efficiently detect various concentrations of sialic acid. The detection was based on the fact that sialic acid saturated SiaP has a higher molecular mass than the sialic acid free SiaP. As it is possible to distinguish between these two species using MS, it was found that addition of sialic acid to the analysed SiaP protein caused relative increase of the concentration of sialic acid saturated species in comparison to the sialic acid-free species in the

sample. By monitoring the relative increase of one species and the concomitant decrease of the second, different responses of SiaP are observed for different concentrations of sialic acid. Although detection of sialic acid was found to be possible using SiaP on the MS platform, the efforts to quantify the ligand were unsuccessful as the correlation between sialic acid concentration and the response was not linear. The creation of a standard curve for sialic acid quantification was impossible due to the lack of linear response. This may be attributed to either intrinsic features of the SiaP protein and its mechanism of binding or to the equipment imperfections. As the resolution of the MS is limited the precise differentiation between two molecules with very similar m/z values may occure difficult. Future advances in MS technology would certainly improve the resolution essential for this type of assay and therefore further tests with more advanced MS devices are suggested.

Further application of the SiaP protein for free sialic acid detection and quantification involved utilization of the surface plasmon resonance (SRP) analysis. The SRP platform has been previously applied for various assay development such as telomerase activity assay (Maesawa *et al.*, 2003), DNA-gyrase assay (Maxwell *et al.*, 2006), immunoassay for nut protein presence (Bremer *et al.*, 2009) and the assay to test coliphages contamination as a fecal pollution in water (Garcia-Aljaro *et al.*, 2008). Both indirect, with immobilised biotinylated sialic acid, and direct assays for free sialic acid detection were tested and different formats were employed. For the indirect assay development neutravidin was immobilised on the CM5 biacore chip and was used to capture biotinylated sialic acid. Subsequently, the SiaP protein was injected in order to capture the immobilised ligand. The pre-incubation of the SiaP protein with the sample containing free sialic acid would decrease the amount of SiaP protein able to bind to the surface of the chip. This was designed to lead to the development of the competitive assay for free sialic acid. Furthermore, the relatively large molecular mass of the SiaP would increase the precision of the assay as any changes in the amount of such a large compound are clearly detectable using the Biacore system. The increased signal baseline after biotinylated sialic acid injection indicates successful immobilisation of the conjugate on the neutravidin coated surface of the chip. However, subsequent testing with the SiaP protein injection yielded no stable increase of the signal. This indicates that the SiaP protein was not binding to the surface. The baseline drift of 35 RU observed during the SiaP injection can be attributed to the unspecifically bound

biotinylated compound leaching off the surface. These findings are in concert with the previously discussed results of ELLA experiments where no interaction between SiaP and sialylated conjugates was observed.

In the subsequently-tested formats for direct sialic acid detection, the SiaP protein was immobilised on the CM5 chip and used to capture the analyte. For the immobilisation of the SiaP protein on the CM5 chip three different approaches were utilized: immobilisation via Strep2-tag and Strep2-tag-specific neutravidin, immobilisation via Strep2-tag and Strep2-tag-specific antibodies and immobilisation via 6Lys-tag directly on the dextran surface of the CM5 chip. The immobilisation of ligands on the Biacore chip surfaces using streptavidin-biotin interaction as well as via monoclonal antibodies has been previously reported (Kazemier *et al.*, 1996; Jensen *et al.*, 1997). This type of immobilisation offers milder conditions and reduce the probability of negative effects of the immobilisation process on the ligand (L\rofas S., *et al*., 1995). The immobilizations of the ligand via 6xHis or 10xHis tag and iminodiacetic acid or nitrilotriacetic acid were also reported; however, the moderate affinity of the chelate–Ni^{2+}–histidine ternary interaction means that there is sometimes considerable decay in the level of immobilised ligand (Nieba *et al.*, 1997; Cooper M., 2002).

The immobilisation via neutravidin or antibodies was selected, as these would also offer the possibility of the regeneration of the detection device after the diagnostic test. In these formats the SiaP protein, degraded over time or irreversibly saturated, would be replaced with a fresh batch of the protein, hence regenerating the surface. Regeneration of the Strep2-tag specific StrepTactin, which as neutravidin, is a streptavidin derivative can be achieved by the cycle of desthiobiotin and sodium hydroxide washes which was reported previously (Schmidt *et al.*, 2007). The immobilisation of SiaP protein via anti-Strep2-tag antibodies (StrepMAB-Immo) is irreversible in physiological conditions. However, the regeneration of the surface may be achieved by utilization of primary goat/rabbit anti-mouse antibodies as a support for mouse StrepMAB-Immo. As the primary/secondary antibody binding is pH dependent it is possible to recharge the surface with StrepMAB-Immo antibodies and subsequently with SiaP.

Initial tests with a neutravidin coated CM5 chip suggested the possibility of the Strep2-tagged SiaP immobilisation. The C-terminal tag was used to promote the orientated immobilisation of the SiaP protein as the C-terminus of the protein is located directly

opposite the binding site of the globular protein. Such orientation would be preferred for further analyte capture on the surface of the chip. However, unstable baseline in the form of significant drift after the SiaP protein injection on the neutravidin coated surface was observed. This can be attributed to the insufficient affinity of neutravidin for the Step2-tag of the SiaP protein in the conditions applied and thus under the flow of the HBS buffer (10 µl/min) the captured SiaP protein was leaching off the surface of the chip. The subsequent injections of sialic acid at various concentrations did not render a stable increase in the signal which indicated that for further development of this type of assay an efficient immobilisation of SiaP protein is crucial.

In order to increase the affinity of the CM5 biacore chip surface for the SiaP protein primary goat anti-mouse and secondary mouse anti-Strep2-tag antibodies were utilized. The initial immobilisation of Strep2-tagged SiaP protein on the StrepMAB-Immo coated surface resulted in an increase in the baseline. It was observed that the baseline was more stable than the previously tested neutravidin surface; however, a drift of the baseline was still observed. This indicates that the antibody surface has higher affinity for Strep2-tagged SiaP than neutravidin, which is in concert with previous reports (Schmidt *et al.*, 2007), where it was found that monoclonal antidobies are a preferential surface for Biacore application when targeting Strep2-tagged proteins. Subsequently, injection of 5 mM sialic acid resulted in a stable increase in the signal which was the first indication that it is possible to detect sialic acid using this methodology. The possibility of chip regeneration was also tested by NaOH utilization which is commonly used for analyte removal (Biacore; Karlsson *et al.*, 2004). The regeneration scouting process is commonly used in Biacore application in order to regenerate the active surface. The concentration of sodium hydroxide (or other chemicals) needs to be optimised. In this study the regeneration of the chip may be attributed to the removal of only sialic acid from the surface, as after the regeneration with NaOH the baseline returned to the level close to the one before sialic acid injection. This indicates that only sialic acid was removed and the SiaP protein stayed on the surface of the chip. The possibility of sialic acid removal by SiaP protein denaturation was previously discussed and is in concert with this result. The removal of sialic acid was facilitated by the flow of the NaOH containing HBS buffer that denatures/relaxes the SiaP protein. The subsequent repeated injection of an equal sample of sialic acid rendered a similar effect, which may indicate that the SiaP protein spontaneously refolds in HBS buffer and

regains its activity after sialic acid removal by NaOH. In order to test different concentrations of sialic acid a 1 mM sample was analysed; however, the result was negative as no stable increase in the baseline was observed. The subsequent tests were continued with 5 mM sialic acid to verify the reproducibility of the results. As the subsequent injection resulted in gradually decreasing values of RU it was concluded that, using this format, it would be difficult to obtain reproducible results. Furthermore, after a sequence of analyte and NaOH injection an overall drift of the baseline was observed. This may be attributed to the surface degradation although it is not clear if the SiaP protein elution off the chip, the elution of one of the antibodies or maybe the elution of all of the protein off the chip was responsible for this effect.

The utilization of indirect immobilisation of the SiaP protein on the CM5 biacore chip using neutravidin/antibodies would offer the potential for advantageous regeneration; however, this approach was found to be unsuccessful due to the problems with protein stability on the surface. In order to immobilise the SiaP protein on the CM5 biacore chip surface and maximise the stability of the binding, the SiaP protein was directly immobilised on the dextran surface. 6Lys-tag SiaP protein was covalently bound to the dextran layer of the chip by primary amine chemistry. The poly lysine tag was selected as the lysine residue contains a primary amine group in its side chain. Primary amine groups were utilized in the immobilisation chemistry (EDC/NHS coupling). Therefore, a string of six lysine residues at the C-terminus of the SiaP protein would increase the possibility of the desired, orientated immobilisation. Otherwise the protein would be immobilised using any of the available primary amines on the surface. The utilization of a poly-lysine tag was previously shown to be an effective method of orientated immobilisation in comparison to the untagged proteins. The studies with modified HIV proteins demonstrated that an increase of the charge and amine density in the tag enhances the coupling yield, the most efficient tag being a six lysine one (Allard *et al.*, 2001). Furthermore, it was found to be an effective method for increasing the protein concentration on the surface. It previous studies under modified HIV proteins (Ladaviere *et al.*, 1998) it was found that the maximum loading capacity of the polymer, on which the protein was immobilized, depended on whether the protein bore the lysine tag or not. It was demonstrated that 6xLys tag increased the effectiveness of immobilisation, which would also be advantageous in this project.

The EDC/NHS chemistry was selected as a method of choice as it is the most

widespread method of direct immobilisation on CM5 chips (Biacore); however, immobilisation using dimethyl pimelimidate dihydrochloride was found to be more efficient (Catimel B., *et al*, 1997).

The application of this format generated the most stable baseline for sialic acid testing among all of the utilized formats. The activity of the SiaP protein was confirmed and the surface was tested with various concentrations of sialic acid. Injections of sialic acid in the range of concentrations from 1 mM to 8 mM generated an almost linear response. This proves that it may be possible to quantify free sialic acid using this method; however, the protocol has to be modified to optimise detection of possibly lower concentrations and also extend the operational range of the assay. The subsequent studies aimed at identification of the upper detection limit of the system generated unusual results for 10 mM sialic acid. The unexpectedly high signal generated by the 10 mM sialic acid sample may indicate that at some concentration of the ligand the transfer of flowing molecules over the chip becomes easier and that the ligand becomes more accessible for the immobilised protein. The subsequent injections of higher concentrations of sialic acid resulted in lower than 10 mM signal. This can indicate the saturation of the surface or the gradual degradation of the SiaP protein. However, the ultimate injection of 8 mM sialic acid generated an identical result to the previous test with the same concentration, which indicates that the response of the surface did not change.

The specificity of the method was tested with 10 mM galactose. It was found that the 10 mM galactose injection generates a similar response to the HBS buffer, which is the first indication of the specificity of the used methodology. This result is in concert with previous reports (Severi *et al.*, 2005; Muller *et al.*, 2006), where the SiaP protein was found to be highly specific for Neu5Ac, with K_d value of 120 nM in comparison to Neu5Gc (K_d=280nM) and α(2→3)sialyl lactose (K_d=18μM). Unspecific interactions of other monosaccharides are highly improbable as no such findings have been previously reported. However, further, in-depth, investigation of this matter using various sugars as well as other substances is needed.

The specificity of the SiaP protein and thus the potential specificity of the SiaP-based assay for sialic acid, in conjunction with the short analysis time, create the possibility for this method to outclass the existing methods (Warren, 1959; Svennerholm, 1956, 1957; Sugahara, 1980; Teshima *et al.*, 1988; Marzouk *et al.*, 2007). The SiaP-based

assay, because of the intrinsic features of the SiaP protein, would be more specific for Neu5Ac than widely applied colorimetric methods for sialic acid quantification which suffer from interference from other members of the sialic acid family (Romero *et al.*, 1997). With regards to relatively expensive enzymatic methods (Sugahara, 1980; Teshima *et al.*, 1988), the assay with immobilised SiaP protein which is capable of regeneration, and thus multiple use, would be competitive, especially where the cost is an issue, such as would be the case for multiple sample analysis in medical diagnostics. The clear advantage of the SiaP-based method over the HPLC method (Morimoto *et al.*, 2001) for sialic acid quantification is the time required. The SiaP based assay requires only five minutes and it may be possible to reduce this time.

The lower detection limit of the sialic acid concentration (1 mM) of the SiaP protein using Biacore technology is unexpectedly high. As the affinity of the SiaP protein for Neu5Ac is high (K_d = 28 nM (Johnston *et al.*, 2007)) the detection limit was expected to also be in the nano or micro molar range. The relatively high lower detection limit may be attributed to the SiaP protein behaviour being altered once immobilised in comparison to the molecule in the solution. This would be consistent with the literature reports of the altered activity and other features of proteins after immobilization. Metapyrocatechase was reported to be only 30% as active following immobilization as a free enzyme; however its stability was greatly improved (Iwaki et al., 1982). The altered activity following immobilization was reported for rabbit muscle pyruvate kinase, where the pH optimum for catalytic activity shifted into a more alkaline direction (Simon *et al.*, 1985). Another explanation of the unexpected detection range may lie in the concentration of the SiaP protein on the surface of the chip. In the experiments with 6Lys-tagged SiaP immobilized directly on the CM5 chip it was found that for detection of sialic acid to occur a certain amount of SiaP protein is required on the chip (12336 RU was found to be sufficient). Insufficient amount of the protein results in no response for sialic acid. These findings suggest the need for investigation of the optimal concentration of the SiaP protein on the CM5 surface for precise and reliable sialic acid detection and quantification.

The research presented in this thesis was aimed at development of a sialic acid detection and quantification method based on sialic acid-specific proteins. To an extent this was successful with the SiaP protein. The SiaP based method for free sialic acid analysis has the potential to significantly influence medical diagnostics where the free sialic acid

level in samples serves as an important indicator of physiological states and diseases (Salvolini *et al.*, 1998; Diamantopoulou *et al.*, 1999; Paszkowska *et al.*, 2000; Wongkham *et al.*, 2003; Marzouk *et al.*, 2007). For the pharma industry the implications of the SiaP based method may be even greater as the rapid analysis may enable real time monitoring of production processes, which is of great importance for biopharmaceutical production in cell culture. As the sialylation of glycoproteins significantly improves their activity, immunogenetic characteristic and blood retention time, it is highly desirable that this process occurs during the cell culture based production process (Egrie and Browne, 2001; Fernandes and Gregoriadis, 2001). The decrease in free sialic acid level in the medium due to its incorporation in the glycan structures and the subsequent sialic acid level rise as a result of product degradation after reaching a certain stage of culturing are important indications of the point where harvesting should occur. For this the real time, specific and reliable analysis of the free sialic acid in the production medium is crucial. Additionally, as presented and discussed above, SiaP is a free sialic acid binder. For media monitoring this is a serious advantage, as no interference from sialylated structures is possible since the protein does not bind sialylated glycoconjugates.

Furthermore, the SiaP protein application for sialic acid detection may be further extended for the purpose of sialylation analysis of various glycoconjugates such as EPO (Egrie and Browne, 2001) and asparaginase (Fernandes and Gregoriadis, 2001). The significance of sialic acid presence in glycans for human physiology and pathology, which were described above (see section 1.1.3), implies a requirement for its in-depth analysis. Paradoxically, the free sialic acid-specific SiaP protein has the potential to be applied for sialylated structure analyses by utilization of upstream processing of the analyte. The sialic acids may be liberated from the glycoprotein by neuraminidase treatment and the resulting free sialic acid subsequently captured by SiaP, enabling its detection and quantification (Figure 4.1 A).

Here it is important to emphasise that the linkage between terminal sialic acid and the penultimate monosaccharide of human glycans predominantly occurs in two different conformations, namely α(2→3) and α(2→6) (Lehmann *et al.*, 2006). The application of neuraminidases capable of specific cleavage of one of the linkages would liberate only one class of sialic acid (Figure 4.1 B and C). This would enable precise measurement of the sialic acid that is linked in a known conformation.

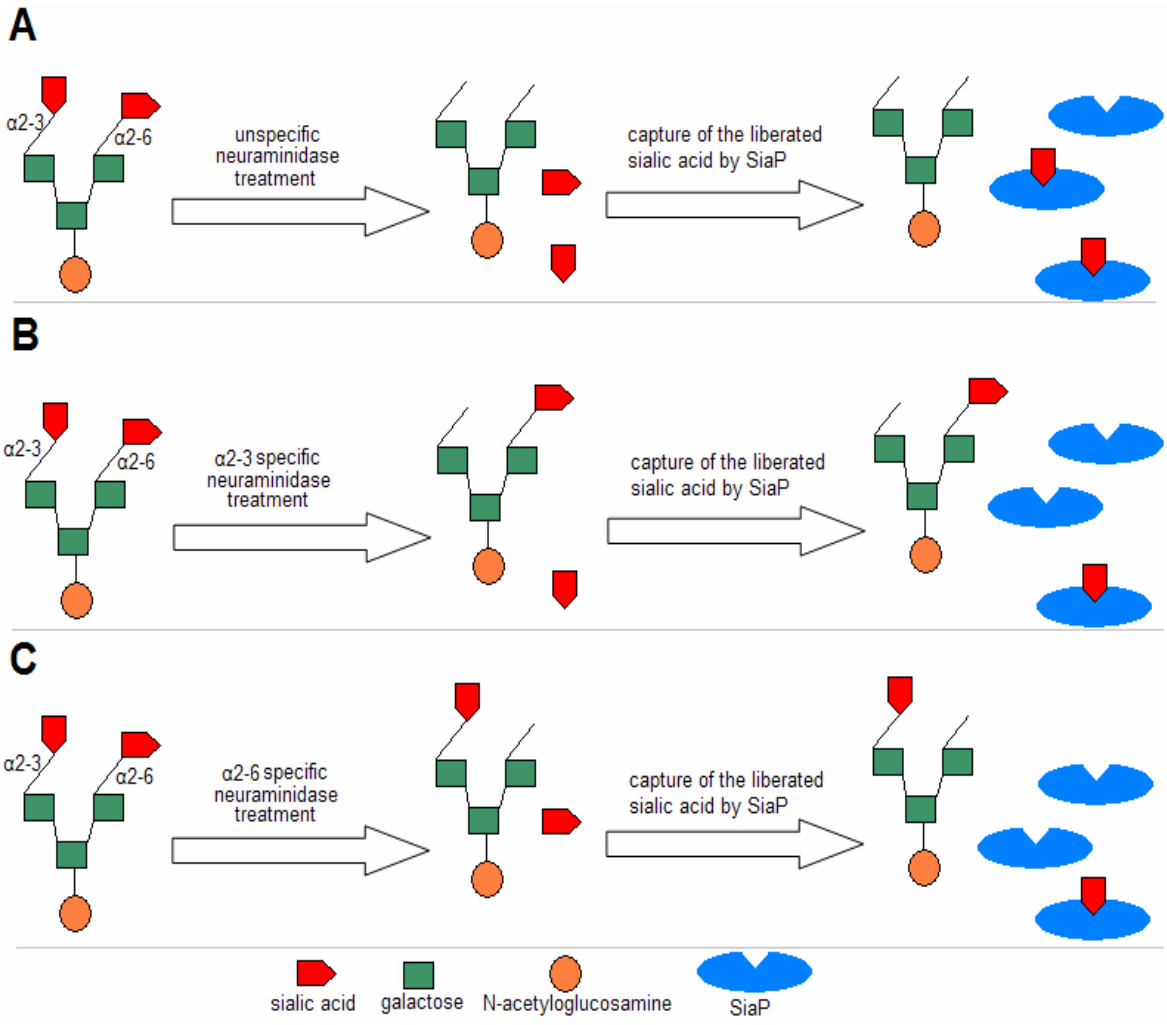


Figure 4.1. The schematic representation of the neuraminidase treatment of sialylated glycans. Unspecific neuraminidase treatment (A) liberates all terminal sialic acids from the glycans, treatment with α2-3 specific neuraminidase (B) liberates α2-3 linked sialic acids and treatment with α2-6 specific neuraminidase (C) liberates α2-6 linked sialic acids. The liberated sialic acid is subsequently captured by SiaP and measured. As the SiaP is specific for free sialic acid, there is no requirement for separation of the neuraminidase-treated products (glycans and sialic acid) as these glycans will not interact with SiaP.

Both free sialic acid present in the medium and the liberated sialic acid from sialylated glycans would be subjected to analysis by the SiaP protein immobilised on biacore CM5 chip. The possible methodology would involve sampling and filtering of the production culture in order to remove cells. The material could be then directly injected onto the SiaP covered chip in order to measure the free sialic acid content in the expression medium (Figure 4.1). In order to analyse the sialylation, the product would be purified from the medium, sialic acid liberated from the product by neuraminidase treatment and

the measurement of liberated sialic acid content would indicate the sialylation level of the product.

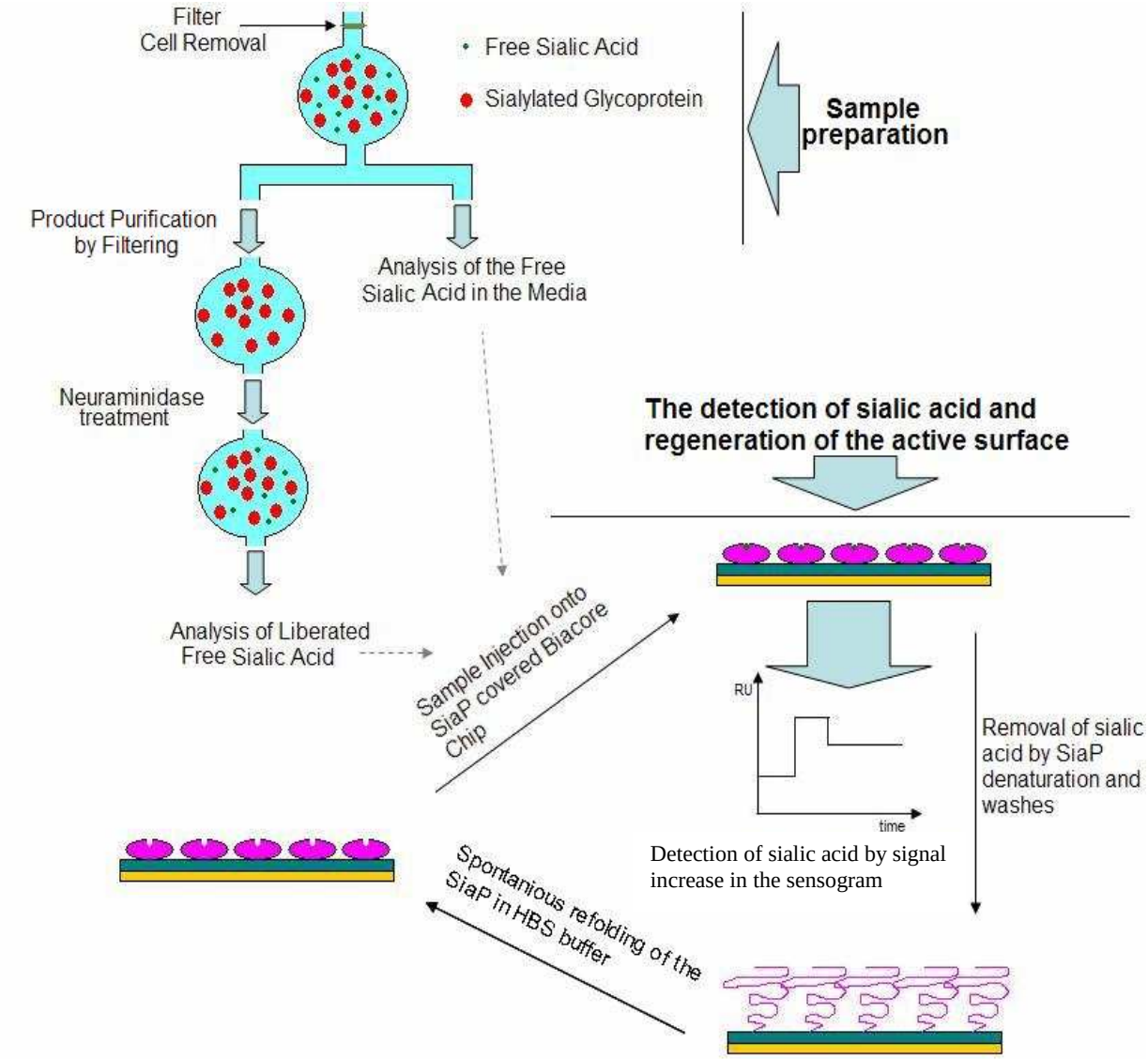


Figure 4.2. Sample preparation and detection of free sialic acid using the SiaP protein immobilised on a biacore CM5 chip. Preparation of sample for measuring the free sialic acid in the medium and preparation of the sample for measurement of the glycoprotein sialylation is presented. Both samples can be subsequently analysed by injection onto the SiaP coated CM5 chip. The cycle of sialic acid capture by SiaP and removal of sialic acid from the surface is presented.

The sialic acid concentration may be calculated by comparison of the signal generated by the sample on the CM5 chip (in RU) with the standard curve of chip response to various sialic acid levels. As presented above, it is possible to remove the sialic acid bound to the chip surface, and thus regenerate the surface. This suggests that the chip may be reusable, which is advantageous with regards to the cost efficiency of this method.

The interaction confirmed in this study, between sialic acid and SiaP, as well as development of the methodology for monitoring of this interaction, creates the possibility for implementation of SiaP based assays exploiting other analytical platforms. One of the envisaged applications of the SiaP protein may be a fluorimetric assay for sialic acid quantification. This assay could be based on the conformational change that SiaP undergoes after sialic acid accommodation. Such a conformational change influences the intrinsic fluorescence of the SiaP protein, namely the intensity of emission at 310 nm (Muller *et al.*, 2006). The free sialic acid-containing sample would be applied on the well of a 96-well plate with pre-immobilised SiaP protein (Figure 4.3). The sialic acid would be captured by the protein and the sample subsequently removed. The increase of the emission at 310 nm in comparison to the blank sample would be used to quantify the sialic acid concentration in the sample.

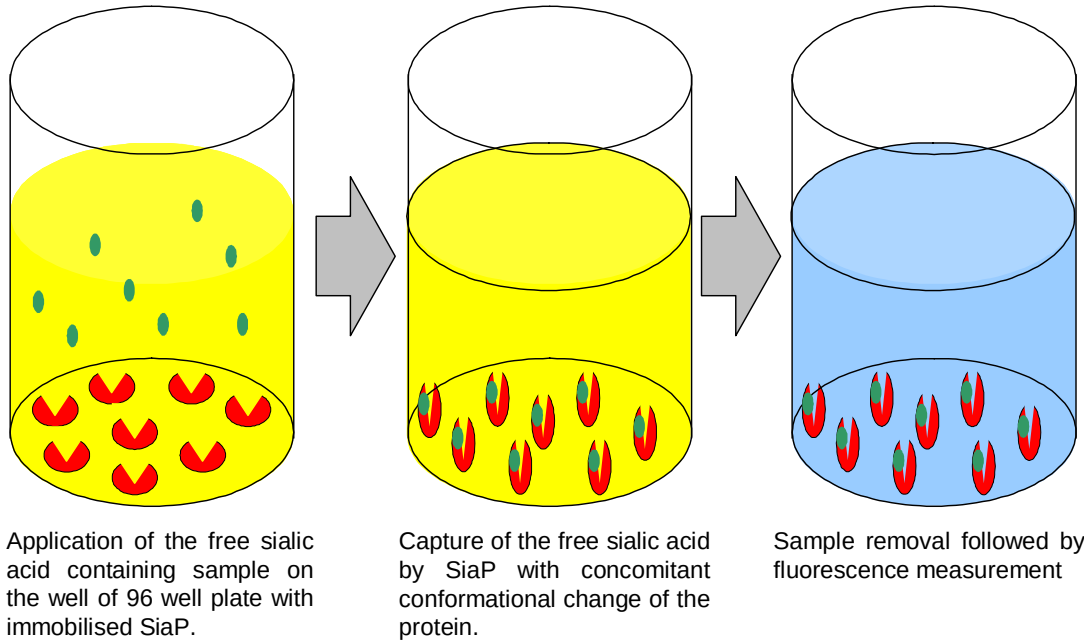


Figure 4.3. Fluorescence based quantification of sialic acid using SiaP.

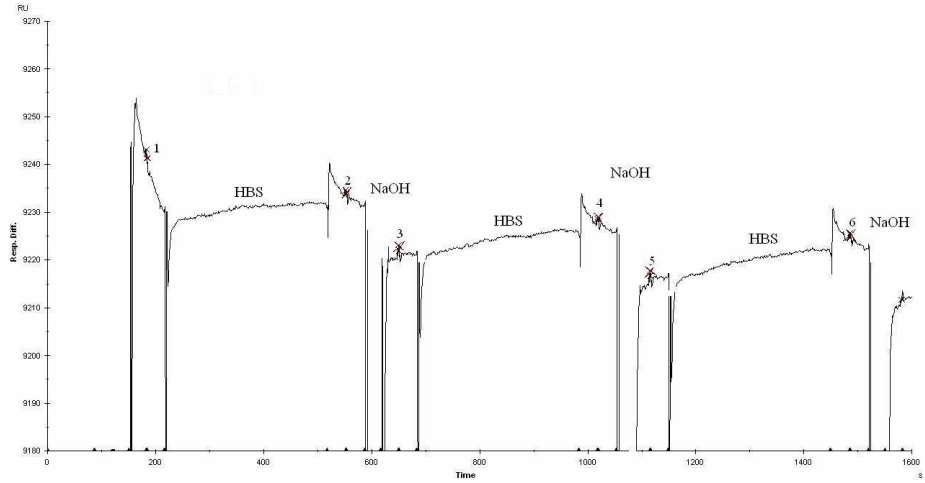
Another approach for sialic acid quantification using SiaP protein would be exploitation of the intrinsic features of the ligand. As the sialic acid is highly negatively charged, the accommodation of such a ligand changes the net charge of the capturing protein and the net charge of the surface covered with this protein. Accurate detection of such changes would enable precise quantification of the sialic acid captured by the protein. The immobilisation of the SiaP protein on monolithic columns and subsequent capture of charged sialic acid would change the electrical characteristics of the column. The

increase of the charge of the column by bound sialic acid would positively impact the passage of electrons through the column from one electrode fixed on one side of the column to the second electrode fixed on the second side of the column. The ability of the monolithic column to conduct electrons can be measured using a method called capacitively coupled conductivity detection (C^4D) (Connolly *et al.*, 2007). The C^4D is a particular type of conductivity-based detector where the electrodes are not in direct contact with the measured solution. Two tubular electrodes placed over the capillary. Each of the electrodes forms a capacitor with the buffer solution inside the capillary through which the signal can be coupled into and out of the capillary.

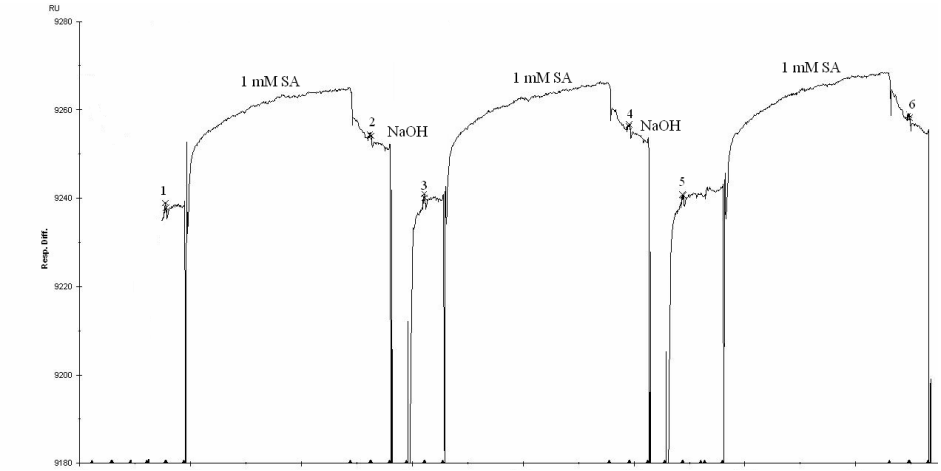
Thus the injection of a free sialic containing sample onto the monolithic column with immobilised SiaP protein would result in ligand capture, change of the column conductivity and a rise in the signal. Initial studies aimed at SiaP immobilisation on monolithic column have been successful and the signal changes after sialic acid application were also detected (O'Shea, Ph.D Thesis, DCU 2009).

In conclusion, as the aim of this study was to develop a method for sialic acid detection and quantification based on sialic acid-specific proteins, the results of this study show the possibility for sialic acid detection and quantification based on the SiaP protein immobilised on a Biacore chip. The SiaP protein was successfully produced in a prokaryotic expression system and its purification was optimised. The secondary goal of Sia-free SiaP production was also achieved. The Sia-free SiaP protein was analysed using MS, ELLA and surface plasmon resonance. The SPR platform was selected for SiaP protein application for Sia detection and quantification. The methodology requires further optimisation with regards to specificity, detection range and protein concentration on the active surface. There also is a vast potential for application of the SiaP protein on other platforms and preliminary results are promising. The successful evolution of this assay may contribute to development of similar assays based on transport proteins for monosaccharides and other molecules, enabling their detection and quantification.

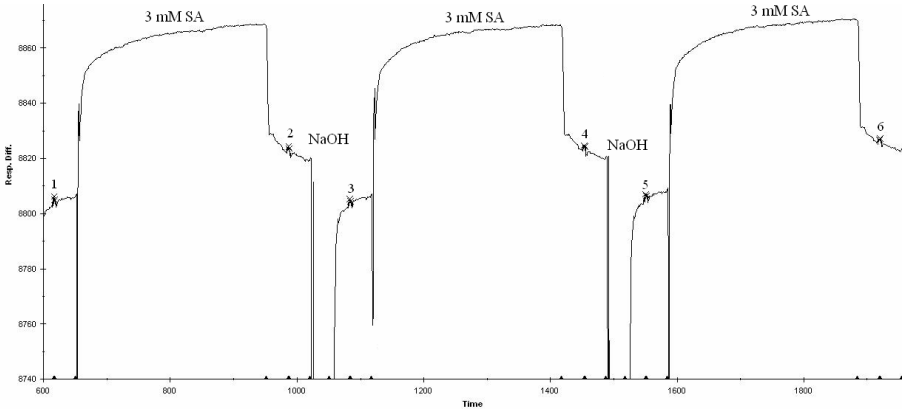
Appendix



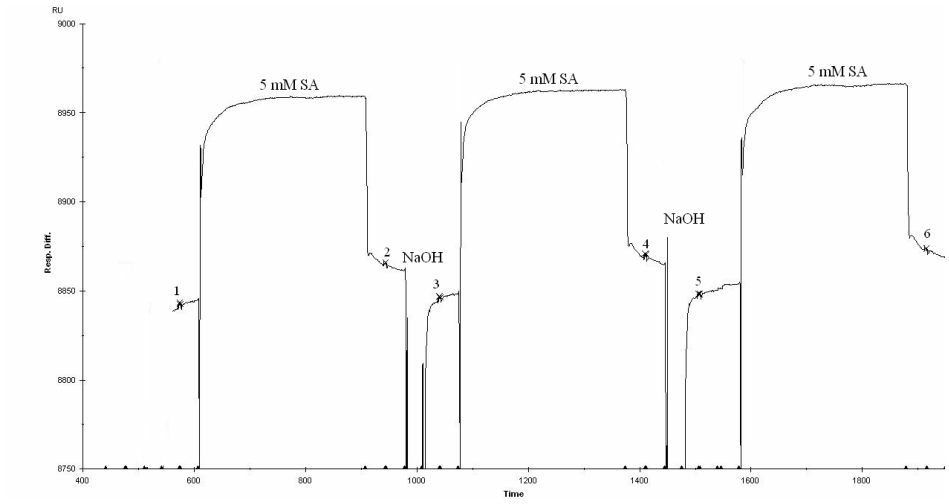
Sensogram representing the injection of 50 µl of HBS buffer (0 Sia) onto the active surface and NaOH [5 µl of 5 mM] regeneration of the surface. The sample was passed over both the SiaP surface and the reference surface (blocked dextran) and the sensogram represents data after reference subtraction. The flowrate = 10 µl/min.



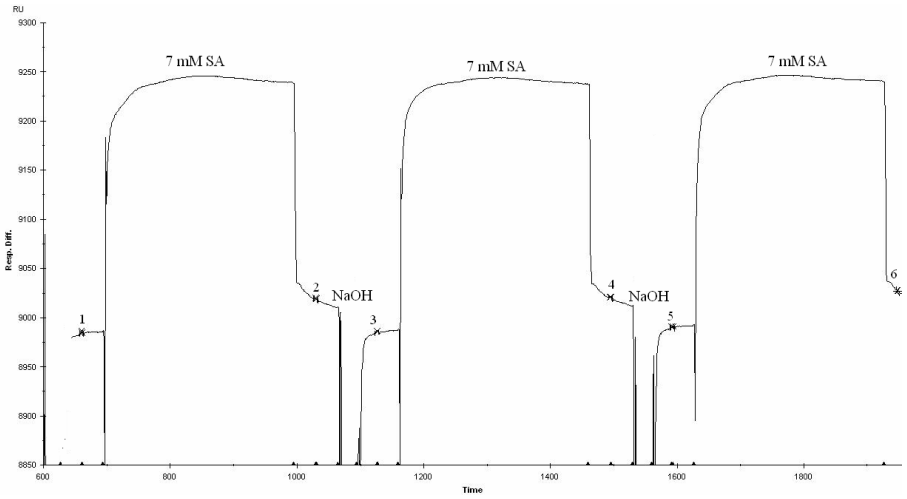
Sensogram representing the injection of 50 µl of 1 mM sialic acid onto the active surface and NaOH [5 µl of 5 mM] regeneration of the surface. The sample was passed over both the SiaP surface and the reference surface (blocked dextran) and the sensogram represents data after reference subtraction. The flowrate = 10 µl/min.



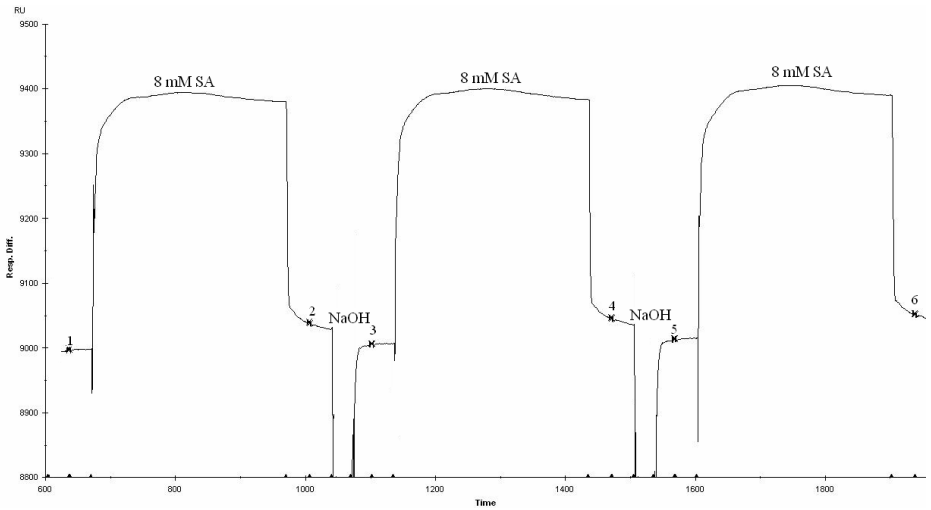
Sensogram representing the injection of 50 µl of 3 mM sialic acid onto the active surface and NaOH [5 µl of 5 mM] regeneration of the surface. The sample was passed over both the SiaP surface and the reference surface (blocked dextran) and the sensogram represents data after reference subtraction. The flowrate = 10 µl/min.



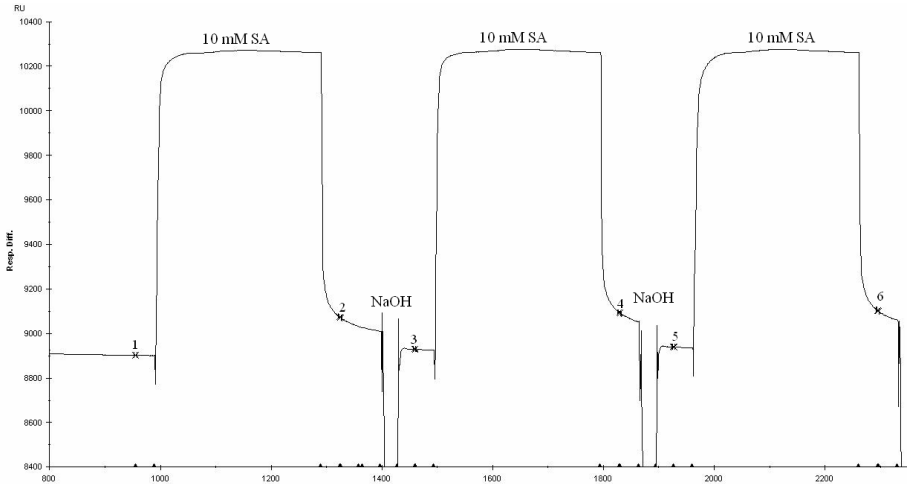
Sensogram representing the injection of 50 µl of 5 mM sialic acid onto the active surface and NaOH [5 µl of 5 mM] regeneration of the surface. The sample was passed over both the SiaP surface and the reference surface (blocked dextran) and the sensogram represents data after reference subtraction. The flowrate = 10 µl/min.



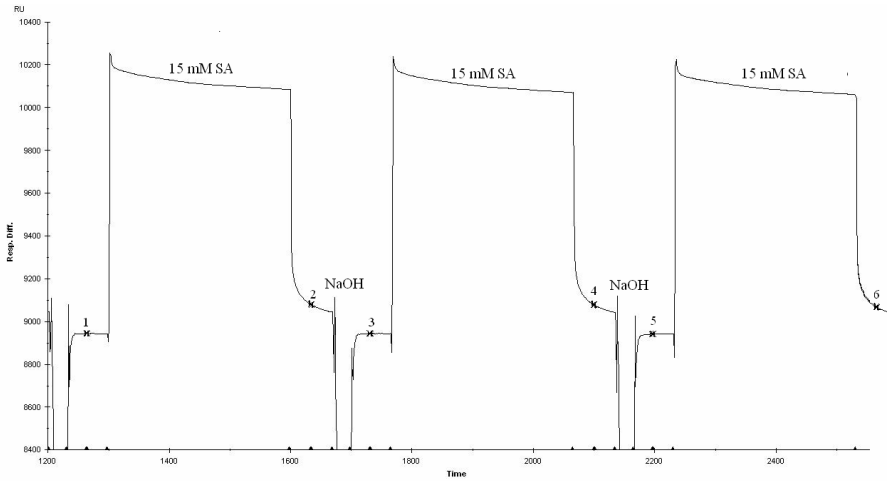
Sensogram representing the injection of 50 µl of 7 mM sialic acid onto the active surface and NaOH [5 µl of 5 mM] regeneration of the surface. The sample was passed over both the SiaP surface and the reference surface (blocked dextran) and the sensogram represents data after reference subtraction. The flowrate = 10 µl/min.



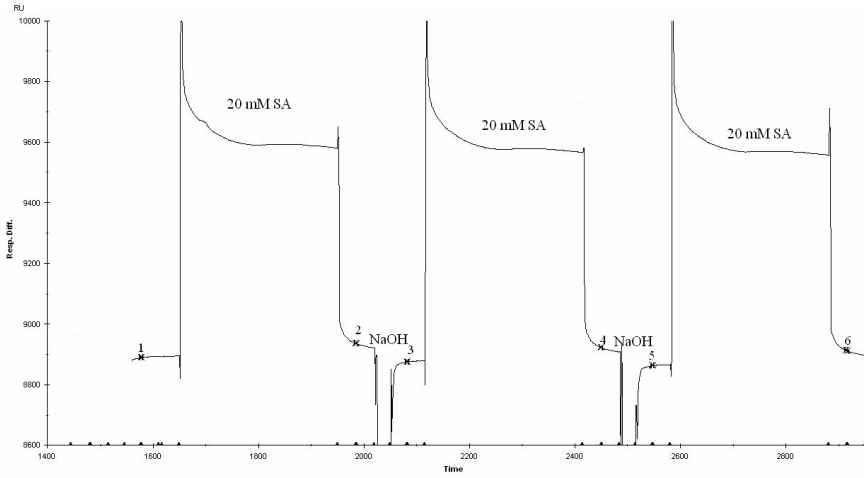
Sensogram representing the injection of 50 µl of 8 mM sialic acid onto the active surface and NaOH [5 µl of 5 mM] regeneration of the surface. The sample was passed over both the SiaP surface and the reference surface (blocked dextran) and the sensogram represents data after reference subtraction. The flowrate = 10 µl/min.



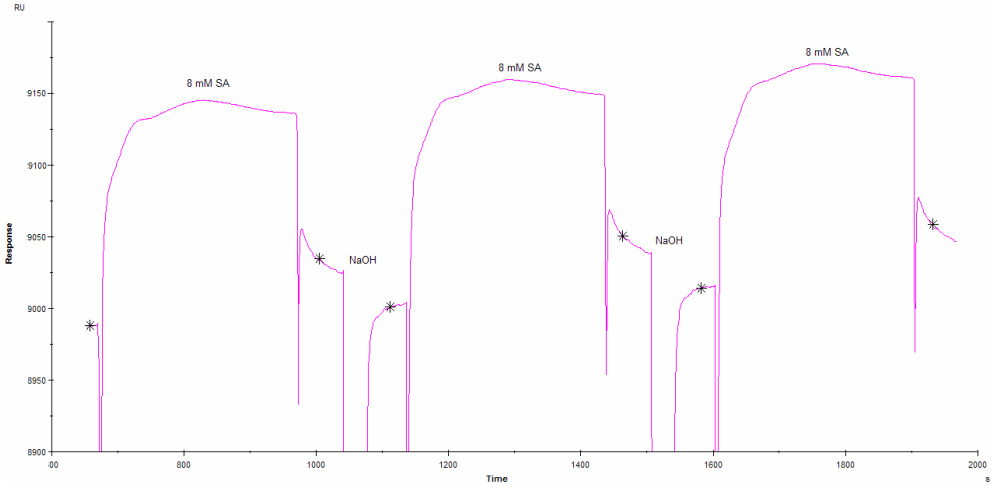
Sensogram representing the injection of 50 µl of 10 mM sialic acid onto the active surface and NaOH [5 µl of 5 mM] regeneration of the surface. The sample was passed over both the SiaP surface and the reference surface (blocked dextran) and the sensogram represents data after reference subtraction. The flowrate = 10 µl/min.



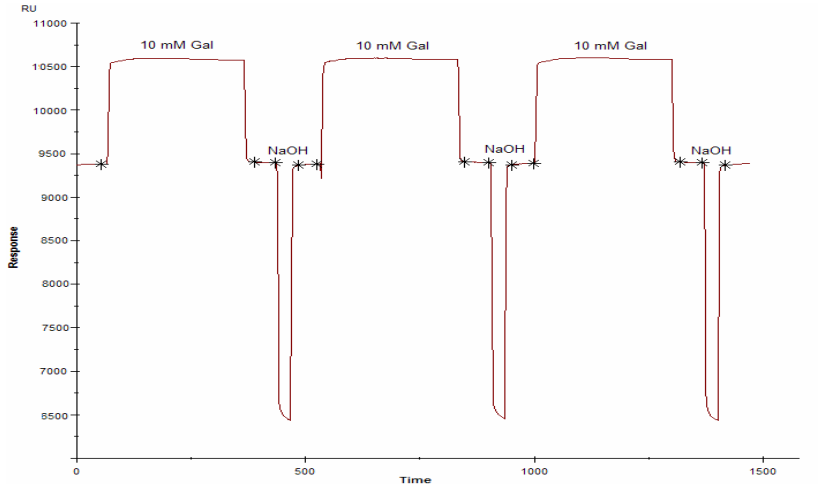
Sensogram representing the injection of 50 µl of 15 mM sialic acid onto the active surface and NaOH [5 µl of 5 mM] regeneration of the surface. The sample was passed over both the SiaP surface and the reference surface (blocked dextran) and the sensogram represents data after reference subtraction. The flowrate = 10 µl/min.



Sensogram representing the injection of 50 µl of 20 mM sialic acid onto the active surface and NaOH [5 µl of 5 mM] regeneration of the surface. The sample was passed over both the SiaP surface and the reference surface (blocked dextran) and the sensogram represents data after reference subtraction. The flowrate = 10 µl/min.



Sensogram representing the injection of 50 µl of 8 mM sialic acid onto the active surface and NaOH [5 µl of 5 mM] regeneration of the surface. The sample was passed over both the SiaP surface and the reference surface (blocked dextran) and the sensogram represents data after reference subtraction. The flowrate = 10 µl/min.



Sensogram representing the injection of 50 µl of 10 mM galactose onto the active surface and NaOH [5 µl of 5 mM] regeneration of the surface. The sample was passed over both the SiaP surface and the reference surface (blocked dextran) and the sensogram represents data after reference subtraction. The flowrate = 10 µl/min.