



UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Faculty of Natural and Agricultural Sciences

Investigating the potentially expanded target repertoire of murinized Internalin of *Listeria monocytogenes*

DANIEL SENZILE NDIMA

A thesis submitted in partial fulfilment of the requirements for the degree *Magister Scientiae* (MSc) in Biochemistry in the Faculty of Natural and Agricultural Sciences, Department of Biochemistry, University of Pretoria

29 February 2016

Supervisor: Prof. Wolf-Dieter Schubert

Co-Supervisor: Dr. Precious Gugulethu Motshwene

Submission declaration

I, Daniel Sensile Ndimma declare that the dissertation, which I hereby submit for the degree (Magister Scientiae) at the Department of Biochemistry, at the University of Pretoria, contains my own work and has not previously been submitted by me for a degree at this or any other tertiary institution.

Daniel Sensile Ndimma

MSc. candidate

December 2015

Plagiarism declaration

Full name _____ Student number _____

Title of the work _____

Declaration

1. I understand what plagiarism entails and am aware of University's policy in this regard.
2. I declare that this dissertation is my own, original work. Where someone's work was used (whether from the printed source, the internet or any other source) due acknowledgement was given and reference was made according to department requirements.
3. I did not make use of another student's previous work and submit as my own.
4. I did not allow and will not allow anyone to copy my work with the intention of presenting it as his or her own work.

Signature _____

Date _____

Acknowledgements

- Firstly, I would like to thank God Almighty for the Grace I have been receiving throughout my studies and the strength for completing this degree.
- My sincere gratitude goes to my supervisor Prof. Wolf-Dieter Schubert for accepting me in his research laboratory and supported me throughout my research, Dr. Shabir Ahmad Mir and Dr. Precious Gugulethu Motswene for research support and advice.
- I would like to thank the following organisations who provided me with financial support, Allan Gray Orbis Foundation, The Mandela Rhodes Foundation, National Research Foundation and University of Pretoria. Their funding was highly appreciated and without their financial support, completing this dissertation would have been impossible.
- I would like to thank University of Pretoria for allowing me to do research in their facilities and Department of Biochemistry for accepting and allowing me to be part of the department.
- My research group was exceptionally helpful, advised and encouraged me during difficult moments when there were difficulties during the research. Special thanks go to Thuso Mapotsane, Mthawelanga Ndengane, Clive Mketsu, Valentine Anye, Lungelo Mandyoli, Clifford Ntui and my honours student Erna Freyer.
- Lastly but not least, I thank my family for the courageous support and prayers. My mother Tseleng Ndima who has sacrificed a lot for my studies up to this far, my two beautiful sisters (Nonki and Mpho) and handsome little brother (Teboho) who kept motivating me. My nephew Neo and niece Mpho.

I dedicate this work to my two cousins and other family members who passed away. It was stressful working in the laboratory after I lost them and this work is also dedicated to them.

Summary

The ability of intracellular pathogens to invade and spread from non-phagocytic cell to another is an imperative mechanism broadly investigated in cellular biology. *Listeria monocytogenes* (*Lm*) –one example of intracellular pathogens, invades specifically human epithelial cells using its surface proteins Internalin A (InlA) and InlB, respectively. InlA alone is sufficient to internalise the pathogen into the host cells by interacting with human E-cadherin –specifically the N-terminal domain 1 (hEC1). The InlA variant (InlA^m) that was previously made to increase the binding affinity to hEC1 was successfully engineered in this study. This variant was found to interact with N-terminal domain 1 of murine E-cadherin (mEC1) by isothermal titration calorimetry (ITC). Previously, the InlA^m was reported to allow *Lm* invasion into M villous cells that express murine N-cadherin –possibly via the N-terminal domain 1 (mNC1). In this study, InlA^m did not have affinity for mNC1 or N-terminal domain 1 of human N-cadherin (hNC1) when analysed by ITC –possibly due to amino acid sequences variation from both mEC1 and hEC1. However, by structurally engineering the complexes (InlA^m/mNC1 and InlA^m/hNC1) and studying their interaction interfaces, it was revealed that mNC1 and hNC1 can be recognised by InlA^m just like hEC1. This was supported by the distances between interacting amino acid residues in InlA^m/hEC1 crystal structure complex, which were also conserved in the engineered complexes. These observations related to the fact that the N-terminal domains of E- and N-cadherin are structurally conserved, therefore that could have attributed to similarities observed in the engineered complexes. Therefore, future studies would aim at using alternative methods that could support or disprove one of the two findings, that is whether InlA^m and any of the N-terminal domains of N-cadherin interact or not.

Keywords: *Listeria monocytogenes*, murinized InlA, murine N-cadherin, human N-cadherin, protein-protein interactions

Contents

Submission declaration	i
Plagiarism declaration.....	ii
Acknowledgements.....	iii
Summary.....	v
List of Figures.....	ix
List of Tables	xi
Abbreviations.....	xii
Chapter 1.0: Introduction	1
1.1 Introducing <i>Listeria monocytogenes</i>	1
1.2 Pathogenic route of <i>Lm</i>	1
1.3 Intra and intercellular spread of <i>Lm</i>	2
1.4 Virulence factors important during intercellular spread of <i>Lm</i>	4
1.4.1 Listeriolysin O.....	4
1.4.2 Host cellular actin capture by ActA.....	4
1.4.3 Host cell-cell junctions are weakened by InlC protein.....	5
1.5 InlA interacts with human E-cadherin	6
1.6 Human E-cadherin functions and interacting partners	7
1.7. Downstream effects of InlA and E-cadherin complex	8
1.8 Downstream effects triggered by murinized InlA.....	9
1.9 N-cadherin: The newly proposed receptor for InlA^m	10
1.10 Rationale of the study	12
1.11 Aims and Objectives of the study.....	13
1.11.1 Aims	13
1.11.2 Objectives	13
Chapter 2.0: Methods and materials.....	14
2.1 Chemicals and reagents	14
2.2 Primers	14
2.3 Plasmid DNA used in the study.....	16
2.3.1 Normal plasmid DNA	16
2.3.2 Recombinant plasmids	16
2.4 Bacterial cells.....	17
2.5 Preparation of media and solutions.....	18

2.5.1 LB broth and agar media.....	18
2.5.2 Buffers and solutions.....	18
2.5.3 Making bacterial cultures.....	18
2.6 Molecular biology experiments.....	18
2.6.1 Polymerase chain reaction (PCR).....	18
2.6.2 Agarose gel electrophoresis.....	20
2.6.3 Site Directed Mutagenesis.....	20
2.6.4 Preparation of chemically competent <i>E. coli</i> cells.....	22
2.6.5 Cloning and transformation of competent cells.....	22
2.6.6 Plasmid DNA isolation.....	24
2.6.7 Sanger sequencing experiment.....	24
2.6.8 Transformation of competent cells for protein expression.....	25
2.6.9 Preparation of glycerol stocks.....	25
2.7 Protein production and purification.....	26
2.7.1 Preparation of pre-culture.....	26
2.7.2 Preparation of main-culture.....	26
2.7.3 Harvesting, lysis of cells and protein collection.....	26
2.7.4 Protein purification.....	27
2.8 Interaction studies by Isothermal Titration Calorimetry (ITC).....	29
2.8.1 Isothermal titration calorimetry experiment.....	29
2.8.2 ITC Data analysis.....	30
2.9 Homology modelling.....	30
2.10 Engineering of complexes.....	31
Chapter 3.0: Results.....	32
Part 1 A: Cloning and DNA analysis of mNC1 and hNC1 gene fragments.....	32
3.1. PCR and colony PCR products.....	32
Part 1 B: Protein production and purification.....	33
3.2 Protein production in <i>Escherichia coli</i> cells.....	33
3.2.1 Production and purification of mNC1 and hNC1 proteins.....	33
3.2.3 Production and purification of mEC1 and hEC1 proteins.....	38
3.2.4 Anion exchange chromatography of mEC1 and hEC1 proteins.....	40
3.2.5 Generation of InlA ^m variant by mutagenesis.....	43
3.2.6 Production and purification of InlA variants.....	44

3.3 Conclusion.....	52
Part 2: Biophysical studies by isothermal titration calorimetry	53
3.4 Introduction	53
3.5 ITC control experiments.....	53
3.6 Interactions of InlA variants and E-cadherin domains (mEC1 and hEC1)	56
3.7 Interaction of InlA ^m and N-cadherin domains (mNC1 and hNC1 proteins)	60
3.8 Conclusion.....	61
Part 3: Homology modelling and structural analysis of complexes	62
3.9 Introduction	62
3.10 Homology modelling of hNC1	62
3.11 Comparisons of InlA ^{wt} /hEC1 and InlA ^m /mEC1 complexes.....	63
3.12 Re-engineering InlA ^m /mNC1 complex.....	66
3.13 Re-engineering InlA ^m /hNC1 complex.....	67
3.14 Conclusion	68
Chapter 4.0: Discussion	69
4.1 Protein production and purification of E- and N-cadherins	69
4.2 Site directed mutagenesis on InlA functional gene	70
4.3 Production and purification of InlA variants	70
4.4 Biophysical characterisation of proteins by isothermal titration calorimetry	71
4.5 Engineered complexes and the insights from interaction interfaces	73
4.5.1 Modelling and overall structural analyses of the complexes	73
4.5.2 Analysing the complexes' interaction interfaces.....	74
4.6 Conclusion and outlook	76
References.....	77
APPENDICES	84

List of Figures

Figure 1.1: The pathogenic route of *Listeria monocytogenes* infection following consumption of contaminated food

Figure 1.2: The infection cycle and spread of *Listeria monocytogenes* from cell to cell

Figure 1.3: A diagram showing species specificity of both InlA and InlB

Figure 1.4: The interacting partners of E-cadherin N-terminal domain

Figure 1.5: The crystal structure of murine N-cadherin domain 1 and 2

Figure 3.1: The 1% agarose gel used to analyse PCR products of mNC1 and hNC1 gene fragments

Figure 3.2: Production and GS affinity chromatography analyses of mNC1 protein by SDS-PAGE

Figure 3.3: Production and GS affinity chromatography analyses of hNC1 protein by SDS-PAGE

Figure 3.4: Anion exchange chromatography and SDS-PAGE analysis of mNC1 protein

Figure 3.5: Anion exchange chromatography and SDS-PAGE analysis of hNC1 protein

Figure 3.6: Production and GS affinity chromatography analyses of mEC1 protein by SDS-PAGE

Figure 3.7: Production and GS affinity chromatography analyses of hEC1 protein by SDS-PAGE

Figure 3.8: Anion exchange chromatography and SDS-PAGE analysis of mEC1 protein

Figure 3.9: Anion exchange chromatography and SDS-PAGE analysis of hEC1 protein

Figure 3.10: The Sanger sequencing results of GST-InlA^m construct

Figure 3.11: Production, GS affinity chromatography and SDS-PAGE analysis of InlA^{wt} protein

Figure 3.12: Anion exchange chromatography and SDS-PAGE analysis of InlA^{wt} protein

Figure 3.13: Production, GS affinity chromatography and SDS-PAGE analysis of InlA^{S192N} protein

Figure 3.14: Anion exchange chromatography and SDS-PAGE analysis of InlA^{S192N} protein

Figure 3.15: Production, GS affinity chromatography and SDS-PAGE analysis of InlA^{Y369S} protein

Figure 3.16: Anion exchange chromatography and SDS-PAGE analysis of InlA^{Y369S} protein

Figure 3.17: Production, GS affinity chromatography and SDS-PAGE analysis of InlA^m protein

Figure 3.18: Anion exchange chromatography and SDS-PAGE analysis of InlA^m protein

Figure 3.19: Examples of non-binding and binding ITC experiments

Figure 3.20: The calorimetric titration isotherm of the interaction between InlA^m and mEC1 proteins

Figure 3.21: The calorimetric titration isotherm of the interaction between InlA^{wt} and hEC1 proteins

Figure 3.22: The heat of dilution signals formed during the calorimetric titration experiment of InlA^m and N-cadherin domains (mNC1 and hNC1)

Figure 3.23: The model and crystal structures of hNC1 and mNC1 proteins

Figure 3.24: Comparison of InlA^m/mEC1 and InlA^{wt}/hEC1 superimposed complexes

Figure 3.25: Engineered complexes of InlA^m/mNC1 and InlA^{wt}/hEC1

Figure 3.26: Engineered complexes of InlA^m/hNC1 and InlA^{wt}/hEC1

List of Tables

Table 2.1: The list of primers for PCR and SDM experiments

Table 2.2: The list of plasmids used in the study with additional information

Table 2.3: The list of recombinant plasmids used in the study

Table 2.4: The list of bacteria strains used in this study

Table 2.5: The components for PCR experiments

Table 2.6: The PCR parameters

Table 2.7: The SDM reaction mixture

Table 2.8: The PCR parameters for all SDM experiments

Table 2.9: The double digestion components of all genes and plasmids

Table 2.10: The ligation reaction for N-cadherin domains and pGEX-6P-2

Table 2.11: The reagents needed for a Sanger sequencing experiment

Table 2.12: The precipitation solution for PCR product

Abbreviations

AEC	Anion exchange chromatography
APS	Ammonium persulphate
BLAST	Basic local alignment search tool
BV	Bed volume
CaCl ₂	Calcium chloride
CDCs	Cholesterol-dependent cytolysins
CV	Column volume
ddH ₂ O	double distilled water
DMSO	Dimethyl sulfoxide
dNTP	deoxyribose nucleotide triphosphates
DTT	Dithiothreitol
<i>E. coli</i>	<i>Escherichia coli</i>
EC	Extracellular
E-cadherin	Epithelial cadherin
EDTA	Ethylene diamine tetra acetic acid
F- actin	Filamentous (F-) actin
GC	guanine cytosine
GS	Glutathione Sepharose
GST	Glutathion S-transferase
HCl	Hydrochloric acid
hEC1	Human E-cadherin N-terminal domain 1
HEPES	Hydroxyethyl piperazineethanesulfonic acid
HIV/AIDS	Human Immunodeficiency Virus/Acquired Immune Deficiency Syndrome
hNC1	Human N-cadherin N-terminal domain 1
Ig-like	Immunoglobulin-like
InlA	Internalin A
InlA ^m	Muritized InlA

InIA ^{wt}	Wild-type InIA
InIB	Internalin B
InIC	Internalin C
IPTG	Isopropyl-D-1-thiogalactopyranoside
ITC	Isothermal Titration Calorimetry
LB	Lysogeny broth
LLO	Listeriolysin
<i>Lm</i>	<i>Listeria monocytogenes</i>
LRRs	Leucine-rich repeats
MCS	Multiple cloning site
mEC1	Murine E-cadherin N-terminal domain 1
MgCl ₂	Magnesium chloride
mNC1	Murine N-cadherin N-terminal domain 1
MWCO	Molecular weight cut off
N-cadherin	Neural cadherin
NFκ-β	Nuclear factor kappa-light-chain-enhancer of activated B cells
N-WASP	Neuronal-Wiskott-Aldrich Syndrome Proteins
PBS	Phosphate-buffered saline
PC-PLC	Phospholipase C
PCR	Polymerase chain reaction
PDB	Protein data bank
Phyre2	Protein Homology/analogy Recognition Engine 2
PI-PLC	Phosphatidylinosito-specific phospholipase C
SDM	Site directed mutagenesis
SDS	Sodium dodecyl sulphate
SDS-PAGE	Sodium dodecyl sulphate-polyacrylamide gel electrophoresis
SEC	Size exclusion chromatography
SH3	Src Homology 3
TAE	Tris-acetate EDTA

Chapter 1.0: Introduction

1.1 Introducing *Listeria monocytogenes*

Listeria monocytogenes (*Lm*) is a rod-shaped, spore forming and Gram positive bacterium (Hof, 2003). It is a foodborne pathogen and contains a low GC content genome (Jeyaletchumi *et al.*, 2010). Other members of the genus *Listeria* are *L. ivanovii*, *L. innocua*, *L. seeligeri*, *L. welshimeri*, *L. grayi*, *L. marthii* and *L. rocourtiae*. Only *Lm* and *L. ivanovii* are pathogenic: *Lm* infects humans, Guinea pigs and gerbils (Bonazzi *et al.*, 2009a; Cossart, 2011), *L. ivanovii* infects ruminants such as cattle and sheep to mention a few (Low and Donachie, 1997; Ramaswamy *et al.*, 2007).

Lm infects humans through consumption of contaminated fresh meats, milk and raw vegetables *inter alia* (Wing and Gregory, 2002). The pathogen survives inside human gut because its optimal growth temperature corresponds with that of mammals (Barbau-piednoir *et al.*, 2013). The disease caused by *Lm* is known as listeriosis and is characterised by symptoms such as sepsis and meningitis amongst others (Wing and Gregory, 2002). Listeriosis is most severe in individuals with compromised immune systems (Ramaswamy *et al.*, 2007).

Humans that are most prone to *Lm* infection include pregnant women, newly born babies, elderly, people living with HIV and acquired AIDS and patients with chronic diseases (Wing and Greory, 2002; Jeyaletchumi *et al.*, 2010; Allerberger and Wagner, 2010; de Noordhout *et al.*, 2014). Listeriosis may be fatal with 30% mortality rate in diagnosed patients (Ramaswamy *et al.*, 2007; Camejo *et al.*, 2011). Listeriosis results in more hospital cases than other food borne infections (McLauchlin *et al.*, 2004; Oevermann *et al.*, 2010).

1.2 Pathogenic route of *Lm*

The main route of infection for *Lm* is shown in **Figure 1.1, steps 1-7**. Consumption of contaminated food (**step 1**) is the main source of infection (Parida *et al.*, 1998). *Lm* is absorbed from intestinal lumen (**step 2**), crosses tightly regulated epithelial cells layer on its way into mesenteric lymph nodes and bloodstream (**step 3**). It then reaches the liver (**step 4a**) or spleen (**step 4b**) where it replicates inside while shielding itself from host immune responses. The bacterium re-enters bloodstream (**step 5**) and crosses blood brain barrier (**step**

6) where it causes meningitis or placental barrier (**step 7**) where it causes spontaneous abortion in pregnant women (Cossart and Toledo-Arana, 2008; Camejo *et al.*, 2011; Travier *et al.*, 2013). *Lm* may also targets the heart, gall bladder and bone marrow (Xayarath and Freitag, 2013).

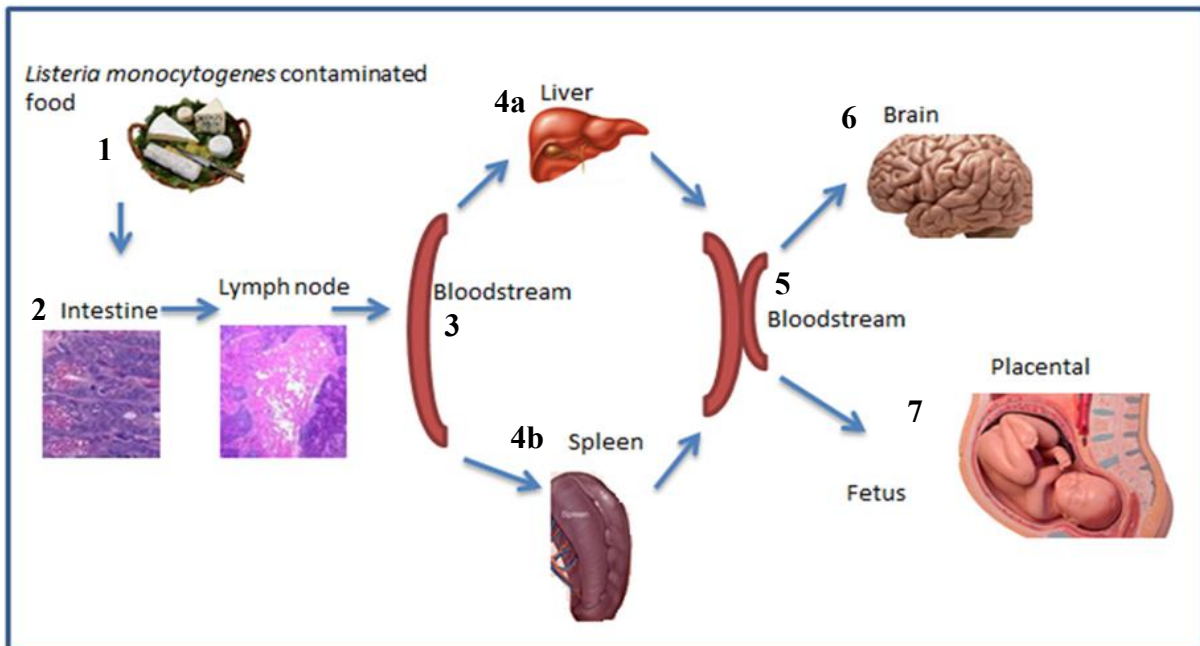


Figure 1.1: The pathogenic route of *Listeria monocytogenes* infection following consumption of contaminated food. The diagram shows that the pathogen crosses three tightly regulated cell layers, the intestinal, blood brain and placental barriers to reach parts of the body where symptoms of listeriosis may be common (the diagram is adapted from Cossart and Toledo-Arana, 2008 with some modifications).

1.3 Intra and intercellular spread of *Lm*

The infection cycle of *Lm* is shown in **Figure 1.2, steps 1-7**. The pathogen uses two surface proteins; Internalin A (InlA) and B (InlB) (**step 1**) to be internalized into host epithelial cells (Gregory *et al.*, 1996). The two proteins ensure this entry by interacting with mammalian cells using human E-cadherin and tyrosine kinase Met (c-Met) as receptors, respectively (Shen *et al.*, 2000; Bierne and Cossart, 2002). The binding of these two proteins to host cells receptors lead to formation of phagocytic vacuole that engulfs and transports *Lm* to the cytoplasm (Cossart and Sansonetti, 2004). This vacuole is then digested by a secreted toxin

listeriolysin (LLO) and phospholipase C (PC-PLC) causing the pathogen to escape into the cytosol (**step 2**) (Ireton, 2013). The pathogen then proliferates inside host cells (**step 3**) and use ActA protein to propel itself (**step 4**) so that it can protrude (**step 5**) into neighbouring cells (Camejo *et al.*, 2011). The pathogen inside double vacuole membrane formed during protrusion gets transported into cytoplasm (**step 6**) of neighbouring cell and LLO, PC-PLC and phosphatidylinositol-specific phospholipase C (PI-PLC) are secreted to digest double membrane vacuole (**step 7**). This leads to pathogen re-establishing its intercellular infection cycle again (Cossart, 2011; Ireton, 2013).

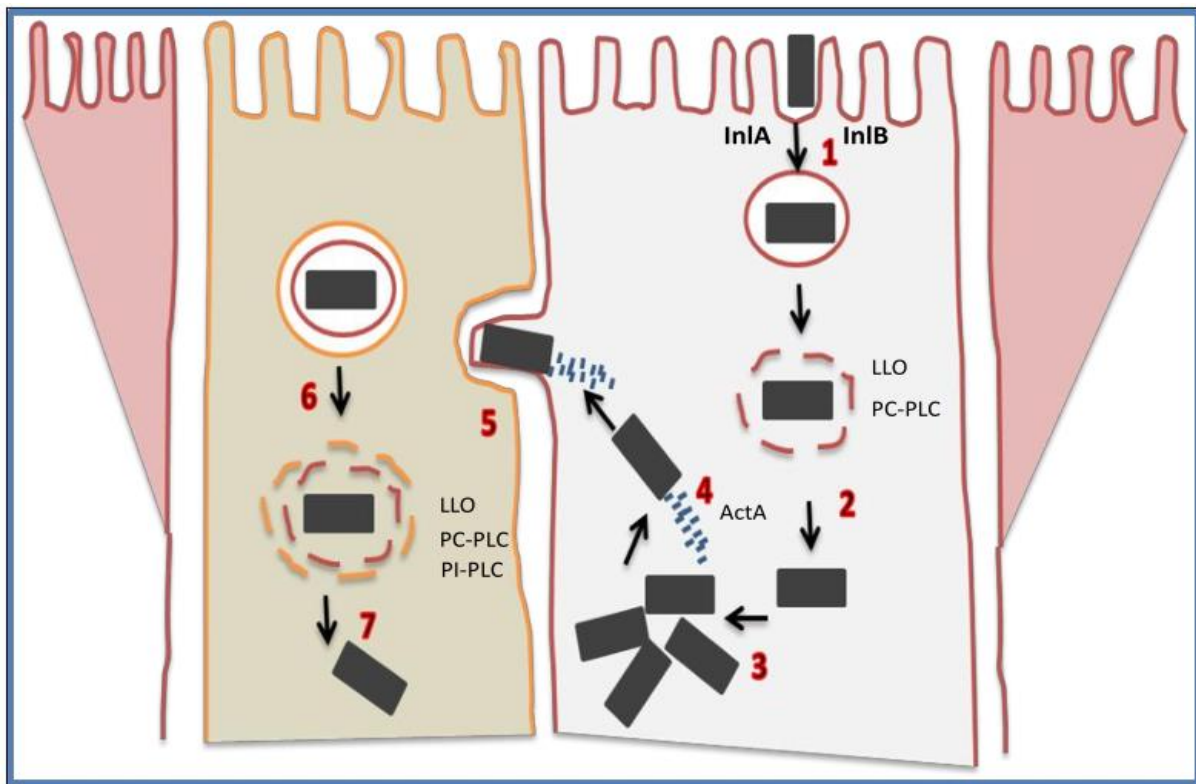


Figure 1.2: The infection cycle and spread of *Listeria monocytogenes* from cell to cell. The seven steps of infection are shown together with virulence factors that are involved. **Step 1** is adhesion and entry of the bacterium, **(2)** is digestion of vacuole with the aid of LLO and PC-PLC, **(3)** is evasion of host defences and intracellular multiplication, **(4)** is intracellular motility, **(5)** is protrusion into adjacent cell, **(6)** is engulfed bacterium by double membrane vacuole and **(7)** is lysis of double membrane vacuole using LLO, PI-PLC and PC-PLC (The diagram was adopted from Ireton’s review of 2013 with some modifications).

1.4 Virulence factors important during intercellular spread of *Lm*

1.4.1 Listeriolysin O

Listeriolysin O (LLO) belongs to a family of cholesterol-dependent cytolysins (CDCs) (Hamon *et al.*, 2012). The CDCs family members share DNA sequence similarity up to 70% and have analogous secondary structures. Most CDCs producing bacteria belong to Gram-positive species (Melton-Witt *et al.*, 2012). The bacterial species include *Clostridium perfringens* which produces a toxin called perfringolysin O or PFO, *Bacillus anthracis* (Anthrolysin O or ALO), *Streptococcus pyogenes* (Streptolysin O or SLO) and *Streptococcus pneumoniae* (Pneumolysin O or PLY) *inter alia* (Schnupf and Portnoy, 2007; Hamon *et al.*, 2012; Melton-Witt *et al.*, 2012). All CDCs can assemble into very large complexes that create pores when encountering a mammalian cell (Melton-Witt *et al.*, 2012).

Lm is internalised into the cytoplasm by membranous vacuole (Mostowy and Cossart, 2012) and released after secretion of LLO toxin (Dussurget *et al.*, 2004; Bischofberger *et al.*, 2012). The toxin is aided by PC-PLC and PI-PLC to disrupt vacuole by creating pores on the membrane enabling the pathogen to escape into the cytosol (Alberti-Segui *et al.*, 2007; Hamon *et al.*, 2012). Usually, the first vacuole (primary vacuole) is disrupted by partnership of LLO and PC-PLC while disruption of the secondary vacuole requires these virulence factors as well as PI-PLC (Alberti-Segui *et al.*, 2007). Both LLO aiding virulence factors directly break down membranes via hydrolysis of phospholipids (Lam *et al.*, 2011).

LLO has been implicated in stimulation of host signalling pathways that lead to host cellular apoptosis by causing histone modification (Melton-Witt *et al.*, 2012). LLO also stimulates nuclear factor kappa-light-chain-enhancer of activated B cells (NF κ - β) and mitogen-activated protein kinase (MAPK) pathways which serve to reprogram host cells during gene transcription (Stavru *et al.*, 2011). In this way, *Lm* is able to manipulate host cells for its survival and avoid innate immune responses (Lam *et al.*, 2011).

1.4.2 Host cellular actin capture by ActA

Actin polymerisation is the process that occurs when cells assemble or nucleate actin in order to respond to intracellular signals that command various cellular processes and responses (Rohatgi *et al.*, 1999). However, actin polymerization can be exploited by intracellular pathogens for their distinct purposes (Jasnin *et al.*, 2013).

An ActA, a surface protein of *Lm*, polymerises actin of host cells. It does so by interacting and recruiting host actin filaments (Filamentous or F-actin) in the host cytosol (Le Monnier *et al.*, 2007). The F-actin creates “comet tails” behind the pathogen (Ireton, 2013) that push it to random directions (Kocks *et al.*, 1992). Therefore, a motile pathogen would create protrusions upon an encounter with host plasma membrane and subsequently get delivered into a neighbouring cell as seen in **Figure 1.2**, steps 4 and 5 (Ireton, 2013).

A recent study has reported ActA-ActA complexes to be responsible for *Lm* persistence of mice intestines and faecal shedding (Travier *et al.*, 2013). It has also been reported that ActA enables *Lm* to escape autophagy (Travier *et al.*, 2013; Travier and Lecuit, 2014). This is a process whereby intracellularly invading pathogens are engulfed and destroyed by immune system (Rajabian *et al.*, 2009).

1.4.3 Host cell-cell junctions are weakened by InlC protein

Inside human epithelial cells, Neuronal-Wiskott-Aldrich Syndrome Proteins (N-WASP) interacts with adaptor protein Tuba (Hamon *et al.*, 2006). This interaction regulates the morphology and tight cellular junctions between cells (Leung *et al.*, 2013). Tuba contains six Src Homology 3 (SH3) domains. The four SH3 domains bind to dynamin I and the sixth SH3 domain (SH3-6) at C-terminal binds to N-WASP (Salazar *et al.*, 2003). The induction of RNA interference (RNAi) experiment targeting Tuba or N-WASP affected the cell morphology and subsequently reduced cortical tension that strengthen cellular plasma membrane, substantiating the importance of Tuba/N-WASP complex formation in cells (Rajabian *et al.*, 2009a).

The Tuba/N-WASP complex is separated by InlC protein secreted by *Lm* following recruitment of actin by ActA (Leung *et al.*, 2013). InlC displaces N-WASP from Tuba SH3-6 domain, doing so positioning itself to the binding interface (Rajabian *et al.*, 2009a). These results in dramatic change in host cells morphology, reducing cortical tension between host cells and resulting in wobbly cells that have no plasma membrane contractibility (Gouin *et al.*, 2010; Leung *et al.*, 2013). Therefore InlC promotes bacterial protrusion and spreading into adjacent cells (Leung *et al.*, 2013).

Even though intracellular growth, autophagy avoidance and intercellular spreading are critical infection strategies, they alone do not account for overall pathogenesis of *Lm*.

Therefore the pathogen has primarily virulence factors that enable its internalisation into host cells. Therefore, the invasion of non-phagocytes by *Lm* clearly points out key upstream virulence factors that promote its internalisation into these cells. Thus InlA is one of the key virulence factors that enable *Lm* to enter host cells (Gaillard *et al.*, 1991).

1.5 InlA interacts with human E-cadherin

InlA is a surface protein that consists of an N-terminal signalling peptide –also known as a cap domain, a central leucine-rich repeats (LRRs) domain with tandem repeats of 22 amino acids (Cabanès *et al.*, 2002) and variable C-terminus of immunoglobulin-like (Ig-like) domain (Gouin *et al.*, 2010). The LRRs domain serves as adhesion and plays critical role in signalling and ligand-receptor interactions. Together with the Ig-like domain, the LRRs domain is critical for *Lm* internalisation into non-phagocytic cells (Cabanès *et al.*, 2002; Genheden and Eriksson, 2013).

The crystal structure of InlA interacting with human epithelial (E)-cadherin extracellular domain 1 (hEC1) has been solved (Schubert *et al.*, 2002). The structure shows hEC1 fitting in the concave interface of InlA's LRRs domain. Key amino acid residues involved are InlA serine on position 192 interacting with hEC1 proline on position 16 (Lecuit *et al.*, 1999; Schubert *et al.*, 2002) (**Figure 1.3, A**). The interaction between these two proteins looking at involved amino acids is critical for studying bacterial invasion *in vivo* and the resulting listeriosis (Wollert *et al.*, 2007).

The molecular interaction between InlA and hEC1 is different from its internalizing partner InlB. InlA binds covalently to hEC1 (Hamon *et al.*, 2006; Genheden and Eriksson, 2013) while InlB binds non-covalently to c-Met, globular C1q and proteoglycans receptors (Hamon *et al.*, 2006; Oevermann *et al.*, 2010). InlA and InlB allow *Lm* to target various body organs where their receptors are present. InlA interaction with hEC1 enables *Lm* to cross the intestinal barrier and blood brain barrier (Bonazzi *et al.*, 2009a). However, both InlA and InlB assist the pathogen to cross the placental barrier (Bakardjiev *et al.*, 2004). Moreover, InlA and InlB exhibit species specificity. They both target human and gerbil epithelial cells through binding their respective receptors. However InlB targets mouse epithelial cells in contrast to InlA which targets Guinea pig and rabbit (**Figure 1.3, B**) (Bonazzi *et al.*, 2009b).

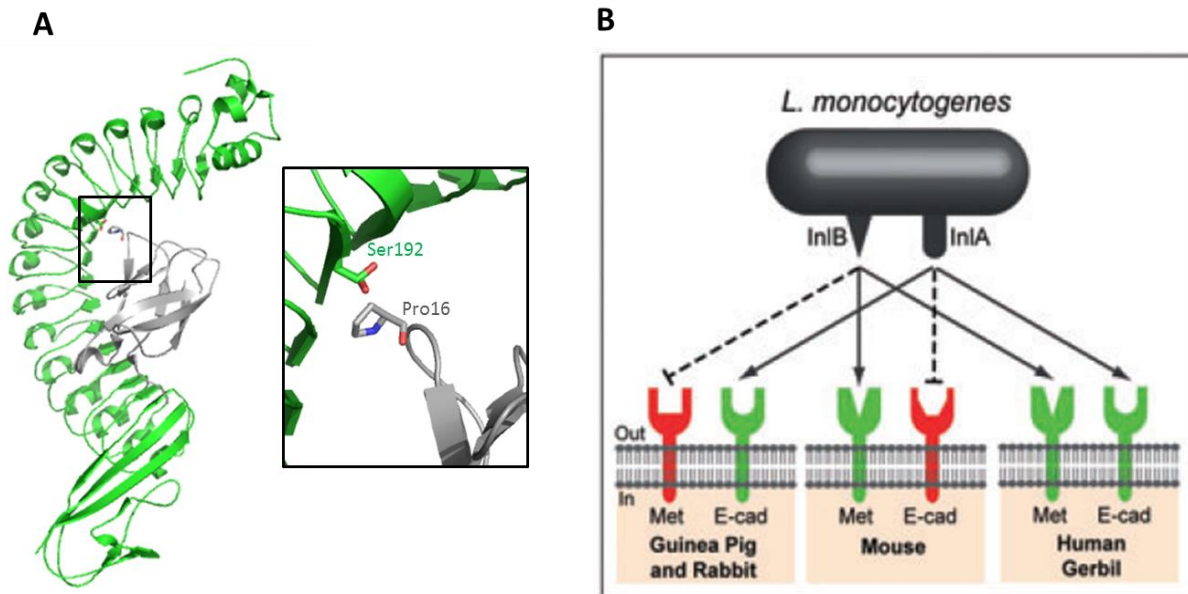


Figure 1.3: A diagram showing species specificity of both InIA and InIB. (A) The complex of InIA (green) and human E-cadherin (grey) is specific due to proline 16 on E-cadherin shown by cross section. The presence of this amino acid ensures hydrophobic interactions between InIA and E-cadherin. (B) Species targeted by *Lm* and receptors responsible for invasion into epithelial cells. Both InIA and InIB display species specific interaction based on receptors amino acid sequences differences (Bonazzi *et al.*, 2009b).

1.6 Human E-cadherin functions and interacting partners

Cadherin is a protein that regulates tight linkage between mammalian cells (Mengaud *et al.*, 1996). Cadherin consists of five extracellular domains or ectodomains (EC1 to EC5) linked together with a cytoplasmic tail domain that interacts with actin cytoskeleton, ensuring stable cellular junctions (Bonazzi *et al.*, 2008; Ciatto *et al.*, 2010). Normally, cadherins of neighbouring cells interact with each other via EC1 domains, ensuring tight cell to cell contacts (Bonazzi *et al.*, 2009b). This interaction occurs when EC1 domains swap their N-terminal β -strands (Ciatto *et al.*, 2010). Tryptophan residue on position 2 of one EC1 is inserted into tryptophan binding pocket of another (Lee *et al.*, 2014). This ensures strong arrangement of extracellular matrix with a traction force that maintains tight cell to cell linkage (Mertz *et al.*, 2012). InIA interacts with the N-terminal EC1 domain to enable *Lm* internalisation (Mengaud *et al.*, 1996).

There are other proteins that bind to EC1 domain (**Figure 1.4**) such as *Bacillus fragilis* toxin (BFT) that cleaves E-cadherin and subsequently weakening cell to cell contacts (Wu *et al.*, 2006; Remacle *et al.*, 2014). *Candida albicans* invasion protein Als3 also binds to EC1 to induce fungal endocytosis and subsequent diseases in host cells (Phan *et al.*, 2007). Moreover, hemagglutinin (HA) complex produced together with botulinum neurotoxin complex from *Clostridium botulinum* has been reported to bind to both EC1 and EC2 domains, enabling the toxin to enter epithelial cells and subsequently entering blood stream to get into Axon Terminal leading to muscle paralysis (Lee *et al.*, 2014).

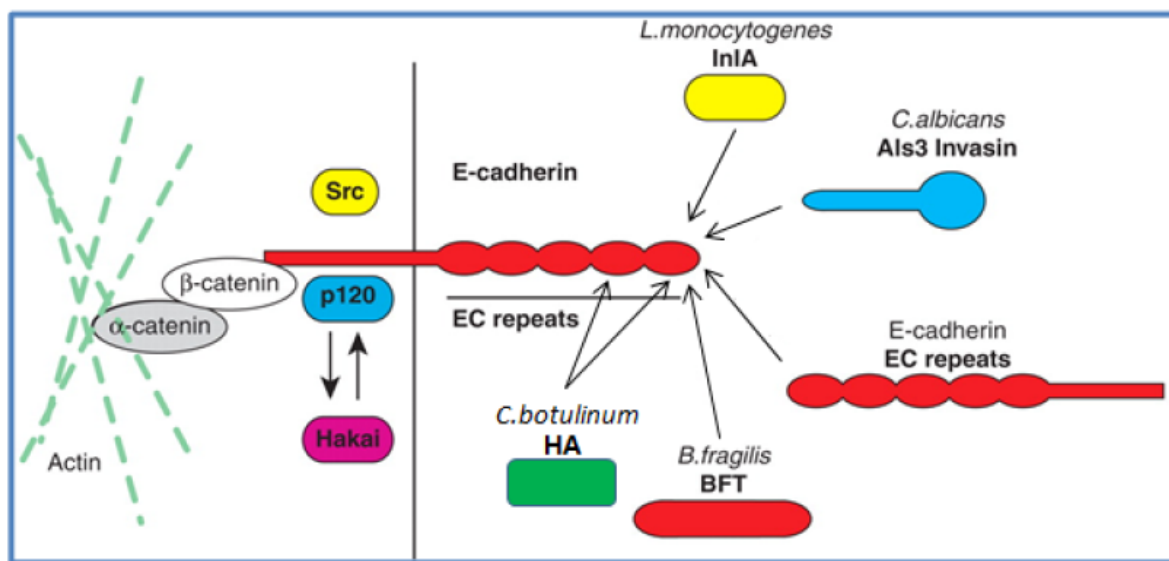


Figure 1.4: The interacting partners of E-cadherin N-terminal domains. There are four known binding partners of EC1 which include InlA from *Lm*, Als3 from *Candida albicans*, BFT from *Bacillus fragilis* and opposite EC1 domain from neighbouring cells. All these proteins have a weaker binding affinity to EC1 except for EC1-EC1 interaction which regulates tighter association of adjacent cells. The diagram was adapted from Bonazzi *et al.*, 2009a with additional information.

1.7. Downstream effects of InlA and E-cadherin complex

The interaction between InlA and EC1 domain of E-cadherin leads to rearrangement of host cellular cytoskeleton, which helps *Lm* to enter into non-phagocytic cells. This process is initiated when signalling pathways of β - and α - catenins are activated and recruited to the site of infection (**Figure 1.4**). Furthermore, InlA and E-cadherin interaction leads to phosphorylation of Src-mediated tyrosine (Bonazzi *et al.*, 2009a). This is followed by

ubiquitination of E-cadherin through Ubiquitin-ligase Hakai. These two post-translational modifications (PTMs) on E-cadherin are important for the *Lm* to manipulate host cells mechanisms for its internalisation strategy (Bonazzi *et al.*, 2008).

It has been reported that the wild type InlA (InlA^{wt}) interaction with E-cadherin molecule is a species specific (Lecuit *et al.*, 1999; Bonazzi *et al.*, 2009a). The interaction spans to a number of species but not murine E-cadherin (Lecuit *et al.*, 1999). Mice are widely used animal models for infectious diseases and host immune responses (Lecuit, 2005). However, there is impaired interaction specificity between InlA^{wt} and EC1 of murine E-cadherin (Lecuit *et al.*, 1999; Lecuit *et al.*, 2001) and this led researchers to engineer mice models that express humanized and genetically modified murine E-cadherin to study listeriosis (Lecuit *et al.*, 2004; Nikitas *et al.*, 2011).

In another approach, *Lm* strains were generated that can adapt to mice by changing two specific amino acids; serine at position 192 was replaced by asparagine (S192N) and tyrosine on position 369 replaced by serine (Y369S) on *inlA* gene (Wollert *et al.*, 2007). The virulence factor InlA protein was said to be murinized, hence denoted InlA^m. The *Lm* expressing InlA^m was able to use murine E-cadherin pathway for its internalisation and protein-protein interaction studies displayed similar binding affinity to that of the wild type InlA/hEC1 complex (Wollert *et al.*, 2007; Monk *et al.*, 2010). Despite side effects observed in mice as a result of this new strain, listeriosis in mice could be studied (Nikitas *et al.*, 2011).

1.8 Downstream effects triggered by murinized InlA

The studies involving murinized InlA observed that *Lm* expressing *inlA^m* gene is responsible for higher bacterial loads in orally infected mouse strain compared to *Lm* expressing *InlA^{wt}* gene. The *Lm* expressing InlA^m is able to increase binding affinity by 2-fold, hence tighter adhesion and subsequent invasion into murine epithelial cells (Wollert *et al.*, 2007; Monk *et al.*, 2010). In another study, the effects of *Lm* expressing InlA^m in mice strains (A/J OlaHsd, BALB/cj, C3HeB/FeJ and C57BL/6J) chosen to acquire listeriosis were investigated (Bergmann *et al.*, 2013). The researchers noted that this *Lm* mutant strain caused increased invasiveness and pronounced infections in all chosen mice strains compared to those infected with *Lm* expressing InlA^{wt}. From four mice strains used, three were found to be highly susceptible to infection by *Lm* expressing InlA^m. This was supported by the evidence of faster *Lm* spreading, huge microbial loads in internal tissues and high level of serum in a form of

interleukin 6 (IL-6), interferon gamma (IFN- γ), tumor necrosis factor alpha (TNF- α) and chemokine ligand 2 (CCL2).

The recent studies have reported that *Lm* expressing *InlA*^m does not only allow internalisation of the pathogen via murine E-cadherin internalisation pathways, but also employs murine N-cadherin internalisation pathway, thus targeting villous M cells which express accessible N-cadherin protein (Tsai *et al.*, 2013). In their study, the researchers argued the relevance of using this particular strain to study listeriosis in mice because of increased receptor repertoire as compared to wild type strain.

1.9 N-cadherin: The newly proposed receptor for *InlA*^m

Neural (N)-cadherin is a homophilic cell to cell adhesion molecule that belongs to the cadherin family of proteins (Tamura *et al.*, 1998). N-cadherin is localized at adheren junctions of chick beating avian hearts and also found in mouse and chicken nervous system cells (Linask, 1992). In addition, Navarro *et al.* (1998) reported that N-cadherin is found in mammalian nervous tissue cells, myocytes and vascular smooth muscle cells. The protein is also found concentrated on the apical and basal membranes in neuronal cells and fibroblasts (Amsellem *et al.*, 2014).

N-cadherin is an essential glycoprotein (Jin *et al.*, 2012), playing a major role together with E-cadherin during neural tube division from embryonic ectoderm –a process also known as neurulation. This glycoprotein also majors in processes such as neurite growth and synapse that occur during brain development (Tamura *et al.*, 1998). Additionally, N-cadherins perform similar functions like E-cadherins by interacting with each other to ensure tight junctions that regulate neighbouring cells (Jin *et al.*, 2012).

N-cadherin shares structural similarities with other cadherin family members, which are five ECs interfaces binding to calcium ions (Tamura *et al.*, 1998). Domain 1 (EC1) and 2 (EC2) are coordinated by three calcium ions while domain 3 (EC3) and domain 4 (EC4) have similar coordination. However, EC2 and EC3 are linked together by a “kinked” linker without calcium ions (Jin *et al.*, 2012). The structure of N-cadherin as published by Jin *et al.* (2012) shows the inter-dependent EC1 and EC2 existing as a two domains protein, with three calcium ions located between as indicated by the black arrows (**Figure 1.5**).

The N-cadherin EC1 shows a conformation described as a quasi- β helix, characterized by a conserved stretch of proline and glycine rich sequence section, forming a succession of β -turns and β -like hydrogen bonds (Tamura *et al.*, 1998). The EC1 domains from opposite cells are known to form what is called “X-dimer”, designated as such because the X shape is formed during the assembly of all dimers (Elledge *et al.*, 2010). The *trans* strand swapping interface can also form between EC1 domains, while a *cis* swapping interface is formed between EC1 of one protein and EC2 of an adjacent protein (Harrison *et al.*, 2011).

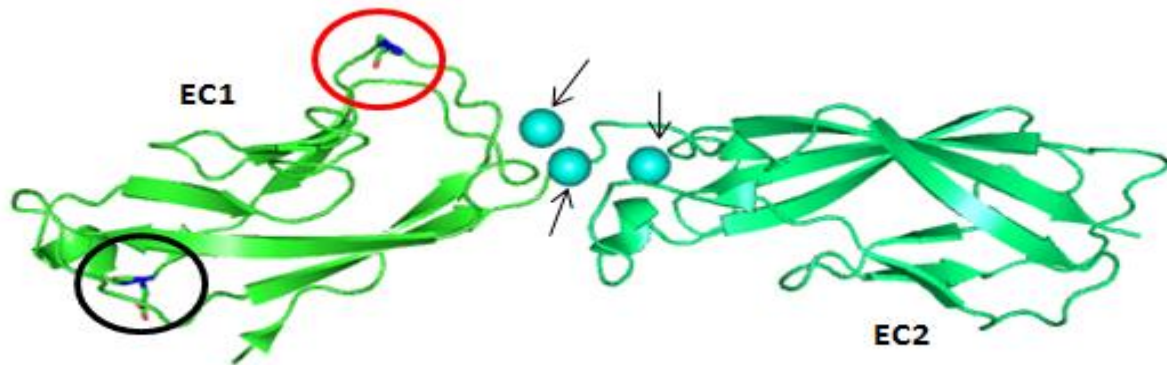


Figure 1.5: The crystal structure of murine N-cadherin domain 1 and 2. The crystal structure shows N-cadherin domains 1 and 2 (EC1 and EC2) with three calcium ions bound in between. Encircled in red is proline 16 residues and encircled in black is asparagine 27 and the two residues are responsible for its counterpart E-cadherin binding to InlA^{wt}. The structure was obtained from Protein Data Bank.

1.10 Rationale of the study

Lm expressing InlA^m has improved pathogenesis (Monk *et al.*, 2010). Moreover, the pathogen targets other unanticipated receptor N-cadherin (Tsai *et al.*, 2013) and also induces harmful effects which are not observed when using *Lm* expressing InlA^{wt} in mice strains (Disson and Lecuit, 2013). The effects include intestinal barrier damage in addition to intrinsic immune reactions, which were not observed before (Monk *et al.*, 2010; Tsai *et al.*, 2013).

The biophysical studies between InlA^m protein and murine or human N-cadherin extracellular domain 1 (mNC1 or hNC1) have not yet been done. The significance of this particular study will be to substantiate whether InlA^m interact with mNC1 or hNC1. This study in proposition would also further provide opportunities to scrutinize the relationship between two proteins belonging to cadherin family (E- and N-cadherins) upon interacting with InlA^m protein.

It is hypothesised that InlA^m will interact with both mNC1 and hNC1 because proline 16 found in hEC1 is also present in both domains. The second hypothesis disputes this because amino acid sequence variations have been detected between E- and N-cadherins.

1.11 Aims and Objectives of the study

1.11.1 Aims

The first aim of the study is to investigate the interaction of InlA^m with mNC1 and hNC1 proteins, by biophysically characterizing the complexes using isothermal titration calorimetry (ITC).

The second aim will be to study the interaction interfaces between InlA variants and both mNC1 and hNC1, respectively, in comparisons to hEC1. The amino acids that are critical for protein-protein interactions will be demonstrated in the interaction interfaces.

1.11.2 Objectives

- Cloning of mNC1 and hNC1 into pGEX-6P-2
- Site directed mutagenesis (SDM) on pGEX-6P-1/*inlA* construct to produce pGEX-6P-1/InlA^m construct
- Protein production and purification using Glutathione Sepharose (GS) affinity chromatography, anion exchange chromatography and size exclusion chromatography
- Biophysical characterization of complexes using isothermal titration calorimetry (ITC)
- Crystal screening of complexes by X-ray crystallography
- Comparative structural similarities studies between E- and N-cadherins interaction interfaces with InlA variants

Chapter 2.0: Methods and materials

2.1 Chemicals and reagents

The list of chemicals, reagents and suppliers used in this study is shown in **Appendix A**.

2.2 Primers

The site directed mutagenesis (SDM) and polymerase chain reaction (PCR) primers were designed manually. However, the SDM primers were analysed online (<http://www.rf-cloning.org/>). SDM primers were designed in a manner in which they satisfied the following rules;

- 1) Length: 25 to 45 nucleotides long
- 2) The melting temperature (T_m): $\geq 78^\circ\text{C}$
- 3) Both primers be self-complement, with the mutation nucleotides in the middle

For PCR and sequencing primers, the following rules were met;

- 1) Length: 15 to 30 nucleotides
- 2) $T_m = 4(G+C)+2(A+T)$, where G, C, A and T are relative number of nucleotides found in a primer
- 3) The T_m of forward and reverse primer were designed to differ only by 5°C

All primers were synthesised at Inqaba Biotechnical Industries (Pty) Ltd (South Africa). The lyophilised primers were dissolved with certain amounts of double distilled water to make stocks solutions with the final concentrations of $100\ \mu\text{M}$ (following supplier's instructions). From all primers stocks solutions, $10\ \mu\text{M}$ of the working solutions were made in order to be used during the subsequent experiments. All primers solutions were stored at -20°C until they were used.

Table 2.1: The list of primers for PCR and SDM experiments

Primer name	Length (bases)	Sequence
mNC1 forward primer	21*	5'-TGGGTCATCCCGCCAATCAAC-3'
mNC1 reverse primer	24*	5'-GTCAATAACATTGATGACAATGTC-3'
hNC1 forward primer	18*	5'-TGGGTCATCCCTCCAATC-3'
hNC1 reverse primer	24*	5'-GTCAATAACATTGATGACAATGTC-3'
InlA S192N SDM forward primer	37	5'- CTAGTCTACAGCAATTA AACT TTGGTAATCAAG TGAC-3'
InlA S192N SDM reverse primer	37	5'- GTCACTTGATTACCAAAG TTTA ATTGCTGTAGA CTAG-3'
InlA S192N/Y369S forward primer	46	5'- GCTTCAAAGATTATTTTT TCAGCA ATAACAAGGT AAGTGACGTAAGC-3'
InlA S192N/Y369S reverse primer	46	5'- GCTTACGTCACTTACCTTGT TATTGCT GAAAAAT AATCTTTGAAGC-3'
pGEX-5' primer	23	5'-GGGCTGGCAAGCCACGTTTGGTG-3'
pGEX-3' primer	23	5'-CCGGGAGCTGCATGTGTCAGAGG-3'
T7 primer	20	5'-TAATACGACTCACTATAGGG-3'
T7 reverse primer	17	5'-CTAGTTATTGCTCGGTG-3'

*Indicates the primers whose parameters were calculated with exclusion of restriction nucleotides sites and 5' end additional nucleotides. Highlighted in grey are codons encoding amino acids to be introduced on the gene. Reverse primer for both mNC1 and hNC1 is one due to identical nucleotide sequences.

2.3 Plasmid DNA used in the study

2.3.1 Normal plasmid DNA

Names, features and suppliers of plasmids used in this study are listed in Table 2.2. These plasmids were readily available in the laboratory at the inception of the study.

Table 2.2: The list of plasmids used in the study with additional information

Plasmid	Size (kb)	Selection	Cleavage site	Tag	Position	Supplier
pGEX-62-1	4.9	Amp ^R	3C protease	GST	N-terminal	GE Healthcare
pGEX-6P-2	4.9	Amp ^R	3C protease	GST	N-terminal	GE Healthcare

2.3.2 Recombinant plasmids

The recombinant plasmids used in this study are listed in Table 2.3. The information includes restriction sites on both sites of target genes. The InlA variants were made using wild type variant as a template. The restriction enzyme *Bgl*III was used instead of *Bam*HI because the sequence encoding the latter was found within mNC1 sequence. *Bgl*III produces complementary sticky ends to that of the *Bam*HI.

Table 2.3: The list of recombinant plasmids used in the study

Recombinant plasmid	Restriction sites	Source of recombinant plasmids
pGEX-6P-1-InlA ^{wt}	<i>Bam</i> HI and <i>Not</i> I	Previous work
pGEX-6P-1-InlA ^{S192N}	<i>Bam</i> HI and <i>Not</i> I	Current work
pGEX-6P-1-InlA ^{Y369S}	<i>Bam</i> HI and <i>Not</i> I	Current work
pGEX-6P-1-InlA ^{S192N/Y369S}	<i>Bam</i> HI and <i>Not</i> I	Current work
pGEX-6P-2-mNC1	<i>Bgl</i> II and <i>Xho</i> I	Current work
pGEX-6P-2-hNC1	<i>Bam</i> HI and <i>Xho</i> I	Current work
pGEX-6P-1-mEC1	<i>Bam</i> HI and <i>Not</i> I	Previous work
pGEX-6P-1-hEC1	<i>Bam</i> HI and <i>Not</i> I	Previous work

2.4 Bacterial cells

The list of competent *Escherichia coli* (*E. coli*) cells used in this study is shown in Table 2.4. Some bacterial cells were used for DNA propagation while some used to produce proteins. The constructs and *E. coli* strains transformed with are shown in **Appendix F**.

Table 2.4: The list of bacteria strains used in this study

Strains	Purpose	Suppliers
* <i>E. coli</i> Top 10 cells	DNA propagation	Stratagene
<i>E. coli</i> BL21 (DE3)	Protein production	Stratagene
<i>E. coli</i> BL21 CodonPlus	Protein production	Stratagene

**E. coli* Top 10 cells are normally used for protein production, but were used for one purpose in this study which was to propagate recombinant DNA following ligation and after SDM experiments.

2.5 Preparation of media and solutions

The lysogeny broth (LB) media, LB agar media and all buffers were autoclaved at 121°C for 20 minutes and then cooled down before storage or use. Other buffers which are heat-sensitive were only filtered using 0.2 µM syringe filters (Sartorius, Germany) and degassed by Vacuum Blotting Pump (LKB BROMMA, Sweden).

2.5.1 LB broth and agar media

LB broth was prepared by dissolving 10 g tryptone or peptone, 5 g sodium chloride and 5 g yeast extract in 500 mL distilled water. The solution was filled up to make 1 L broth, followed by autoclaving. The LB agar media was prepared with the similar composition as LB broth, but with additional 12 g of bacteriological agar. The agar medium was cooled down to 40°C before adding ampicillin. The sterile Petri dishes were filled with 30 mL of agar medium while swirling smoothly and letting to set for 1 hour before use or storage at 4°C.

2.5.2 Buffers and solutions

The list of buffers and solutions is shown in **Appendix B**.

2.5.3 Making bacterial cultures

All bacterial cultures were prepared in the baffled Erlenmeyer flasks (Duran, Germany) that allowed aeration. The ampicillin was added accordingly (1 mL of ampicillin per 1 L of bacterial culture) and bacterial cultures were incubated inside Multitron Shaker Incubator (INFORS, USA) with desired temperatures. This was done for all experiments requiring bacterial induction and protein production, plasmid DNA isolation, making of pre-cultures, making main cultures and glycerol stocks.

2.6 Molecular biology experiments

2.6.1 Polymerase chain reaction (PCR)

The constructs mNC1-2 and hNC1-2 domains were shipped from University of California, San Francisco (USA) (the sequences are shown in **Appendix E**). The N-terminal domains (mNC1 and hNC1); (highlighted in red, **Appendix E**) were amplified by PCR. The first

amino acid (Asp¹) from the sequences was excluded by primers. The removal of this residue was to ensure that tryptophan on second position (Trp²) is exposed at the N-terminus. All constructs were sequenced by Sanger method (section 2.6.7), followed by aligning them using online software ClustalW2 (<http://www.ebi.ac.uk/Tools/msa/clustalw2>). The PCR reagents kit components are listed in a Table 2.5 and parameters listed in Table 2.6. All PCR experiments were done using T100TM Thermal Cycler (Bio-Rad, USA).

Table 2.5: The components for PCR experiments

Reaction components for	Concentration	Volume (μL)
mNC1 and hNC1		
10 x Phusion HF buffer	1 x	5
10 mM dNTP mix	10 μM	1
10 μM forward primer	0.5 μM	2.5
10 μM reverse primer	0.5 μM	2.5
DNA template	-varied	1.4
ddH₂O	-	35.6
100 % DMSO	2 %	1
Phusion DNA polymerase	1U	1
Total		50

Table 2.6: The PCR parameters

PCR step	Purpose	Temperature	Time
Step 1	Initial denaturation	95°C	2 minutes
*Step 2	Denaturation	95°C	40 seconds
*Step 3	Annealing	58°C	40 seconds
*Step 4	Elongation	72°C	3 minutes
Step 5	Final elongation	72°C	6 minutes
Step 6	Hold	4°C	∞

*Steps 2 to 4 were repeated 35 times.

2.6.2 Agarose gel electrophoresis

The agarose gel electrophoresis was run to analyse DNA samples following amplifications by PCR, isolation of plasmid DNA and restriction digestion of samples. In this study, 1% (w/v) agarose gel was used and the samples were mixed with DNA loading buffer (**Appendix B**) containing gel red (Biotium, Inc. USA) (10 μ L DNA and 2 μ L loading buffer). The samples were then loaded alongside the GeneRuler™ 100 bp Plus DNA ladder. The agarose gel electrophoresis was run at 80 volts (v) for 50 minutes and the gel visualized and photographed using Molecular Imager® Gel Doc™ XR+ (Bio-Rad).

2.6.3 Site Directed Mutagenesis

Site directed mutagenesis (SDM) was done to mutate InIA^{wt} functional gene while cloned in pGEX-6P-1. The primers were designed (Table 2.1) to overlap the specific regions on the gene of interest. Conventional PCR experiment was performed to generate the first and second mutations. Serine on position 192 was replaced with asparagine, hence mutation was denoted (S192N). After this SDM experiment, the construct was sequenced to confirm the mutation. InIA^{S192N} construct was used as a template to do a second SDM experiment which was to replace tyrosine on position 369 with serine, representing (Y369S). The construct was now called InIA^{S192N/Y369S} or InIA^m. Following each SDM-PCR experiment, the samples were digested with *DpnI* enzyme for 1 hour at 37°C. The enzyme was heat inactivated for 5 minutes at 80°C. The enzyme normally digests only methylated parental DNA and leaves the

newly amplified DNA for downstream application. The SDM reaction components and parameters are listed in Table 2.7 and Table 2.8, respectively.

Table 2.7: The SDM reaction mixture

Reaction components	Concentration	Volume (μL)
10 x Phusion HF buffer	1 x	5
10 mM dNTP mix	10 μM	1
10 μM forward primer	0.5 μM	2.5
10 μM reverse primer	0.5 μM	2.5
DNA template (InIA)	100 ng/ μL	0.5-1
ddH₂O	-	35.5
MgCl₂	1 mM	1
Phusion DNA polymerase	1U	0.5
Total		50

Table 2.8: The PCR parameters for all SDM experiments

Steps	Purpose	Temperature	Time
Step 1	Initial denaturation	95°C	2 minutes
*Step 2	Denaturation	95°C	40 seconds
*Step 3	Annealing	60°C	40 seconds
*Step 4	Elongation	74°C	6 minutes
Step 5	Final elongation	74°C	8 minutes
Step 6	Hold	4°C	∞

*Steps 2 to 4 were repeated 25 times.

2.6.4 Preparation of chemically competent *E. coli* cells

The *E. coli* strains (stored as glycerol stocks) were plated on LB agar plates without any antibiotic and incubated overnight at 37°C. A single colony was isolated and inoculated into 15 mL LB broth without antibiotic. The culture was incubated inside the Multitron Shaker Incubator with the speed of 160 rpm overnight at 37°C. The culture was used to inoculate 200 mL fresh LB broth. The bacterial cells were grown until OD_{600nm} of 0.4, followed by cooling on ice while occasionally swirling. Then the bacterial cells were harvested by centrifugation in F9-6 X 1000 LEX rotor using SORVALL LYNX 6000 Centrifuge (Thermo Fischer Scientific), at 4°C with a speed of 5,213 x g for 15 minutes. The supernatant was discarded, pelleted cells resuspended in 20 mL buffer 1 (80 mM magnesium chloride and 100 mM calcium chloride). This cell suspension was allowed to stand on ice for 30 minutes and centrifuged again as above for 10 minutes. Then the supernatant was discarded and pelleted cells resuspended in 20 mL ice-cold buffer 2 (100 mM calcium chloride). This was followed by another centrifugation of cell suspension for 5 minutes using the same parameters as above. The supernatant was discarded and the pellet resuspended in 2 mL ice-cold solution 3 (85 mM calcium chloride and 15% (v/v) glycerol). About 50 µL aliquots were poured into 1.5 mL Eppendorf tubes, flash cooled inside liquid nitrogen and stored at -80°C.

2.6.5 Cloning and transformation of competent cells

2.6.5.1 Cloning

Following amplification of mNC1 and hNC1 gene fragments by PCR; the DNA samples were purified using GENEJet PCR Purification Kit (Thermo Fisher Scientific) following manufacturer's instructions. This was followed by restriction digestion of DNA samples and pGEX-6P-2 with corresponding restriction enzymes (Table 2.3). All samples were analysed on 1% agarose gel and DNA bands excised from the gel, followed by purification using GENEJet Gel Extraction Kit (Thermo Fisher Scientific). This was followed by ligation of digested DNA fragments into digested pGEX-6P-2 vector. The components of restriction digestion and ligation experiments are listed in Tables 2.9 and 2.10, respectively.

Table 2.9: The double digestion components of all genes and plasmids

Reagent	Concentration	Volume (µL)
10 x Fast digestion buffer	-varied	4
DNA samples (pure PCR product or plasmid)	-varied	10 to 20
Enzyme 1	1 U	1
Enzyme 2	1 U	1

Table 2.10: The ligation reaction for N-cadherin domains and pGEX-6P-2

Reagents	Concentration	Volume (µL)
5 x T4 DNA ligase buffer	1 x	4
T4 DNA ligase	0.2 U	1 µL
ddH₂O	-	-varied
*mNC1 or hNC1	300 ng	-varied
*pGEX-6P-2	100 ng	-varied

*The ligation reaction was done with 3:1 (insert: plasmid) weight basis ratio. Before transformation, the ligation mixture was incubated at 16°C for 16 hours.

2.6.5.2 Transformation of bacterial cells by heat-shock

Heat-shock is the technique used to propagate recombinant plasmid DNA into competent *E. coli* cells. The recombinant DNA (1.5 µL) was mixed with 50 µL of bacterial competent cells and left on ice for 30 minutes. The tubes were placed in a 42°C heating block (Techne) for 45 seconds and back on ice for 2 minutes. The pre-warmed LB broth was added to the cells, allowed to stand at 37°C for 5 minutes, followed by shaking on Thermomixer 5436 (Eppendorf) for 55 minutes. The cells were spread on LB agar plates containing ampicillin

and left to dry in the laminar flow cabinet. This was followed by incubation of agar plates at 37°C overnight.

Colonies that were able to grow on agar plates containing ampicillin were considered to be positive and one or more colonies were screened by colony PCR. In the colony PCR, same reagents in Table 2.7 were used except that a colony was diluted with 2 mL LB broth and served as DNA template. Following confirmation of insert, this colony was grown further to make glycerol stock and plasmid isolation culture. The agar plates with remaining positive colonies were stored at 4°C.

2.6.6 Plasmid DNA isolation

A single colony from agar plates was inoculated into 10 mL of LB broth containing 0.2 µg/mL of ampicillin. The bacterial culture was allowed to grow at 37°C overnight on a shaker incubator. Plasmid DNA was isolated using GenJET Plasmid Mini-prep Kit (Thermo Fisher Scientific) following the manufacturer's instructions. All plasmid DNA samples were eluted with double distilled water (ddH₂O) and quantified using a NanoDrop ND-1000 Spectrophotometer (Biotechnologie GmbH, Germany) at absorption of 260 nm (A_{260}). The plasmid DNA was analysed with aid of agarose gel electrophoresis.

2.6.7 Sanger sequencing experiment

The sequencing of all recombinant plasmid DNA was done at Forestry and Agricultural Biotechnology Institute (FABI) sequencing unit, University of Pretoria. To prepare samples for Sanger sequencing, Table 2.11 provides list of reagents used to perform PCR. The cycle-PCR mixture was made to 10 µL and after the experiment; about 10 µL of ddH₂O was added to make up 20 µL solution. The PCR product was then supplemented with 80 µL of precipitation solution (Table 2.12). This was followed by spinning the samples using Centrifuge 5417C with 30-place fixed angle rotor for 1.5-2.0 mL tubes (Eppendorf) for 30 minutes at 20817 x g. The supernatant was gently discarded and precipitated DNA washed twice with 70% ethanol. The samples were allowed to dry in the laminar flow cabinet for 20 minutes in order to eliminate excess ethanol and smell. The samples were then taken to FABI sequencing unit.

Table 2.11: The reagents needed for a Sanger sequencing experiment

Reagent	Concentration	Volume (μL)
5 x Sequencing buffer	0.5 x	1
Plasmid DNA	60-100 ng	-varied
*Big Dye	-	1.2
Primer (forward/reverse)	3.2 pmol	2

*The Big Dye contains *Taq* DNA polymerase, dNTPs and ddNTPs which are reagents required for both amplification and chain termination process in a Sanger sequencing reaction.

Table 2.12: The precipitation solution for PCR product

Reagent	Concentration	Volume (μL)
Sodium acetate (pH 4.5)	3 M	3
99.9 % ethanol	78%	62.5
ddH ₂ O	-	14.5

2.6.8 Transformation of competent cells for protein expression

Following the confirmation of recombinant plasmid DNA by sequencing, *E. coli* strains; BL21 (DE3) and BL21 (DE3) CodonPlus competent cells were transformed for protein production. The positive colonies were screened by colony PCR.

2.6.9 Preparation of glycerol stocks

In this study, the new glycerol stocks were prepared from single colonies (of GST-mNC1 and GST-hNC1 constructs) and from old glycerol stocks of GST-mEC1, GST-hEC1 and GST-InlA^{wt} constructs. To make a glycerol stock, 1 mL of bacterial culture was mixed with 1 mL of 80% glycerol (autoclaved) in a laminar flow cabinet. The solutions were mixed well by vortexing using Vortec-2-Gene (Scientific Industries, USA) and stored in -80°C freezer.

2.7 Protein production and purification

2.7.1 Preparation of pre-culture

All pre-cultures were made by inoculating glycerol stocks into Erlenmeyer flask with 50 mL fresh LB broth containing ampicillin. The cultures were incubated inside the shaker incubator at 37°C overnight with the speed of 160 rpm.

2.7.2 Preparation of main-culture

In order to make the culture for protein production, a pre-culture was inoculated into 1 L fresh LB broth containing ampicillin. The bacterial cells were allowed to grow by shaking the flasks at the speed of 170 rpm, at 37°C, until OD_{600nm} reading between 0.6 and 0.8. To measure this absorbance reading of bacterial culture, 1 mL of fresh LB broth was poured into a cuvette and used to blank the CO8000 Cell Density Meter (WPA biowave, USA). The same amount of growing bacterial culture was used to measure the OD_{600nm} reading. The bacterial cells were then induced for protein production or gene expression by adding 0.1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG). All bacterial cells were induced at 28°C overnight with incubator shaking speed of 170 rpm. Following induction, 1 mL of bacterial culture was taken and labelled as after induction sample.

2.7.3 Harvesting, lysis of cells and protein collection

The bacterial cells were harvested using F9-6 X 1000 LEX rotor within Thermo Scientific SORVALL LYNX 6000 Centrifuge while applying manufacturer's instructions; speed: 5,213 x g, time: 15 minutes, temperature: 4°C. Following the harvesting of cells, the supernatant was discarded as waste and the pellets were resuspended in 1 x PBS (**Appendix B**) and gently dissolved.

A sonicator (QSONICA SONICATORS, USA) was used to disrupt cells by applying 5 cycles of 30 seconds disruption step and 30 seconds break, while bacterial cells were placed on ice in order to prevent protein denaturation. The disrupted cells were transferred into the SS34 tubes (Thermo Fischer Scientific) which were also chilled on ice. The insoluble proteins and cell debris were separated from soluble proteins (target proteins) by centrifugation using Thermo Scientific SORVALL LYNX 6000 Centrifuge, with F21-8 x 50 Y rotor, at the speed of 37,500 x g for 1 hour at 4°C. The soluble fraction was obtained from the supernatant while

insoluble fraction from cell pellet. About 1 μ L of soluble fraction was taken and labelled, while the insoluble fraction pellet was resuspended in 8 M urea buffer.

2.7.4 Protein purification

The proteins were produced as N-terminus glutathione-S-transferase (GST) tag fusions. All proteins were first purified using Glutathione Sepharose (GS) affinity chromatography and anion exchange chromatography, followed by buffer exchange using size exclusion chromatography. Briefly, the GST-tagged 3C protease was obtained from stored sample that was previously prepared in the laboratory for general use.

2.7.4.1 GS Affinity chromatography

About 2 mL bed volume (BV) of Glutathione Sepharose (GS) beads (Sigma, USA) was added to 15 mL GST SpinTrapTM column (GE Healthcare, USA). The beads were washed with 10 x BV of 20% ethanol. This was followed by washing with 10 x BV of double distilled water and equilibration with 10 x BV of 1 x PBS. The supernatant from section 2.7.3 was mixed with GS beads in 50 mL Falcon tubes (Eppendorf, Germany) and incubated at 4°C overnight on the MX-T6-S roller mixer (Dragon Lab, China).

The supernatant and GS beads mixture was poured into GST SpinTrapTM column while allowing unbound proteins and impurities to flow through out. The flow through fraction was collected into clean 50 mL Falcon tubes and labelled accordingly. This was followed by washing of impurities and unbound proteins three times with ice cold 1 x PBS buffer, while collecting three wash fractions. The GS beads with bound proteins were supplemented in 6 mL ice cold 1 x PBS buffer and the 3C protease was added. This was followed by mixing the solution by inverting the column up and down and subsequently incubating at 4°C overnight on the roller mixer. To elute target proteins while leaving GST still bound to GS beads, the solution in the column was allowed to flow out making elution 1 sample. Extra 6 mL of ice cold 1 x PBS buffer was added two more times into the purification column and the elution fractions were collected again. The samples were quantified using a NanoDrop 2000 UV-Vis Spectrophotometer (Thermo Fischer Scientific), following manufacturer's instructions. All protein fractions were concentrated at 4°C using Amicon[®] Ultra concentrators (Merck Millipore, USA) and Eppendorf 5417 C Centrifuge at 20817 x g.

2.7.4.2 Anion Exchange Chromatography

Ion exchange chromatography is a method used to purify proteins based on their charge and there are two derivatives which are anion exchange chromatography (AEC) and cation exchange chromatography (CEC). In both methods, proteins bind to the column based on their ionic properties in specific buffers. The difference between these methods is that negatively charged proteins bind to positively charged column during AEC. In contrast, proteins with net positive charges bind to negatively charged column in CEC (Cummins *et al.*, 2011). In both cases, an increasing salt concentration results in the displacement and hence elution of the target proteins (Wei *et al.*, 1999).

All proteins were concentrated (varied in concentrations from 10 mg/mL to 20 mg/mL) and purified by anion exchange chromatography (AEC) in an 8 mL MonoQ HR 10/100 GL column (GE Healthcare) on an AKTA Protein Purification System (Amersham Biosciences, UK). The proteins were loaded onto column at flow rate of 0.8 mL/min. About 2 to 4 column volume (CV) of low salt buffer (**Appendix B**) was used to wash out unbound proteins. The elution was monitored with linear salt gradient of high salt buffer (**Appendix B**). Following this purification, elution peaks were visualised on 15% SDS-PAGE to determine purity of proteins.

Proteins that were not successfully purified by this method were added to GS beads for second GS affinity chromatography. These proteins were eluted using 1 x PBS buffer and concentrated. The mEC1, hEC1, mNC1 and hNC1 were pooled down or concentrated using 10 kDa molecular weight cut off (MWCO) Amicon[®] Ultra concentrators (Merk Millipore) at 4000 x g using Thermo Scientific Heraeus Megafuge SR Centrifuge (F15- 6X100 Carbon Fiber rotor), while occasionally pipetting up and down until 0.5-1 mL sample was left. The InlA variants were concentrated using a 30 kDa MWCO concentrators instead. All proteins were quantified using NanoDrop 2000 UV-Vis Spectrophotometer (Thermo Fischer Scientific).

2.7.4.3 Size Exclusion Chromatography

Size exclusion chromatography (SEC) is used to separate proteins based on their sizes. Large proteins elute first before the small proteins. In this study, SEC was used both for buffer exchange purpose and as final purification step. All proteins were concentrated (varied in

concentrations from 10 mg/mL to 20 mg/mL) and purified using SEC in HEPES buffer (**Appendix B**). A 24 mL Superdex 75 16/60 column (GE Healthcare) was washed with 4 CV of 20% ethanol followed by 4 CV of double distilled water. This was followed by equilibration of the column using HEPES buffer before loading proteins. The proteins were loaded into the column at the flow rate of 0.8 mL/min. The elution of proteins was monitored at absorbance of 280 nm. The fractions were collected, pooled down using appropriate spin concentrators, quantified similarly to above and stored at 4°C for subsequent experiments.

2.7.4.4 Resolving proteins on SDS-PAGE

All protein samples were mixed with 8 x SDS-PAGE loading buffer (**Appendix B**) at 12 µL protein sample: 2 µL loading buffer ratio. The protein samples were denatured at 100°C for 5 minutes on the Thermostat 5320 drybath heating block (Eppendorf). The samples were loaded alongside the protein markers (Thermo Fisher Scientific) onto the 15% SDS-PAGE. All protein samples were resolved at 40 mA for 35-40 minutes (40 mA per one gel). The SDS gels were stained in a staining solution (**Appendix B**) for 10 minutes. The visibility of protein bands on the gels was ensured by adding destaining solution (**Appendix B**) and shaking for 10 minutes. All SDS gels were imaged and photographed using Molecular Imager® Gel DocTM XR+ machine (Bio-Rad).

2.8 Interaction studies by Isothermal Titration Calorimetry (ITC)

2.8.1 Isothermal titration calorimetry experiment

Isothermal titration calorimetry (ITC) is used to measure the rate of reaction taking place during ligand-target macromolecule interaction (Pierce *et al.*, 1999; Freyer and Lewis, 2008). The types of reactions measured include exothermic or endothermic reactions (Pierce *et al.*, 1999). ITC is advantageous compared to other biophysical characterisation techniques because the binding parameters; number of binding sites (n), dissociation constant (K_d), thermodynamic binding parameters; change in enthalpy (ΔH) and change in entropy (ΔS) are measured in a single experiment (Freyer and Lewis, 2008).

ITC data collection was collected at the Medical Pathology Laboratory, University of Cape Town (Republic of South Africa). The MicroCal iTC₂₀₀ machine (Microcal, USA) with

Origin v7.0 software was used for protein-protein interactions analyses. The InlA variants were used as ligands while mEC1, hEC1, mNC1 and hNC1 were used as target samples. Prior and between individual experiments, the sample cell and titrating syringe were washed 3 times with double distilled water to avoid interference of contaminants from previous runs. While filling the sample cell with target protein solution, avoidance of bubbles was ensured by gently pipetting the sample. After filling the syringe with ligand protein solution, two purge-refill cycles were performed to avoid air bubbles in the syringe. The ligand in a syringe with an amount of 300 μL (250 μM) was titrated into 200 μL of the target protein (100 μM) inside the sample cell. The instrument settings for the experiments were as the following: injections = 18, initial delay = 60 seconds, spacing = 180 seconds, filter period = 5 seconds, first injection volume = 1 μL , other injections = 2 μL , measurement temperature = 25°C, reference power = 5 $\mu\text{cal s}^{-1}$ and stirring speed = 1000 rpm.

2.8.2 ITC Data analysis

The integration of the peaks was corrected manually by adjusting the baseline to eliminate the effect of bubbles or drifting of the baseline. The blank experiment values were used to remove dilution effects and the last heat points of the binding curve were made to be zero. The data was fitted and analysed using one site binding model using Origin v7.0 with aid of MicroCal ITC add-on software.

2.9 Homology modelling

Homology modelling describes a computational technique that is used to predict the protein structure using its amino acids sequence (Venselaar *et al.*, 2000). Normally, when experimental structures determination fail, homology modelling serves as a solution to predict the protein structures (Krieger *et al.*, 2003).

During homology modelling, a template recognition and alignment process is initiated by programs such as basic local alignment search tool (BLAST) that identifies template by comparing the query sequence to all the sequences of known structures in the protein data bank (PDB) (Bishop *et al.*, 2008). This process ensures that a query sequence (protein sequence to be modelled) is aligned to a template sequence (already existing protein structure) from one homologous protein with highest sequence similarity (Marks *et al.*, 2012). The backbone of target amino acid sequence is generated by transfer of atomic

coordinates (N, C α , C and O) of template residues (Krieger *et al.*, 2003). During modelling, gaps are generated as a result of deletions during alignment and are closed-up on the continuous backbone. Native orientation of side chains conformations is also critical during homology modelling. These side chains are fitted on the model backbone while positioning atoms, fixing bond lengths and torsion angles to obtain quality modelled structure (Krieger *et al.*, 2003; Raman *et al.*, 2010).

The crystal structures of mEC1 (PDB name: 2QVF), hEC1 (2O72) and mNC1 (3Q2W) proteins already existed in the PDB database, with exception of hNC1. Homology modelling was done to predict the structure of hNC1. The hNC1 amino acid sequence was submitted to online software named Protein Homology/analogy Recognition Engine 2 (Phyre2) (<http://www.sbg.bio.ic.ac.uk/phyre2/html/page.cgi?id=index>). The template used to build this model was that of mNC1, with confidence of 99.9% and coverage of 98%. The model was downloaded as PyMol graphic file (<https://www.pymol.org/>). Comparisons were made between hNC1 modelled and mNC1 crystal structures.

2.10 Engineering of complexes

The complexes of InA^{wt}/hEC1 and InA^m/mEC1 crystal structures were obtained from PDB and presented as PyMol graphics files. The complexes were superimposed to illustrate similarities in the interaction interfaces. The interaction interfaces at specific locations involving the mutations on InA^m were illustrated. These complexes were used to engineer formation of interaction interfaces when mNC1 and hNC1 were present. Therefore, the interaction interfaces at mutation positions on InA^m were also illustrated to study possible interactions between this protein and N-cadherin domains.

Chapter 3.0: Results

Part 1 A: Cloning and DNA analysis of mNC1 and hNC1 gene fragments

3.1. PCR and colony PCR products

The PCR amplified mNC1 and hNC1 gene fragments and pGEX-6P-2 vector were digested with restriction enzymes as indicated in chapter 2, section 2.3.2. Each digested gene fragment was mixed with pGEX-6P-2 in a ligation experiment and plasmid DNA isolation was confirmed by 1% agarose gel electrophoresis (**Appendix G**) and Sanger sequencing. Sanger sequencing results confirmed that both mNC1 and hNC1 gene fragments were cloned in frame with the GST tag of pGEX-6P-2 (see vector map in **Appendix D**). The colony PCR screening of mNC1 and hNC1 gene fragments confirmed ~300 bp products lengths, coding for 100 amino acids (**Figure 3.1, lanes 3, 4, 7 and 8**, respectively). The conventional PCR products served as size controls (**Figure 3.1, lanes 2 and 6**).

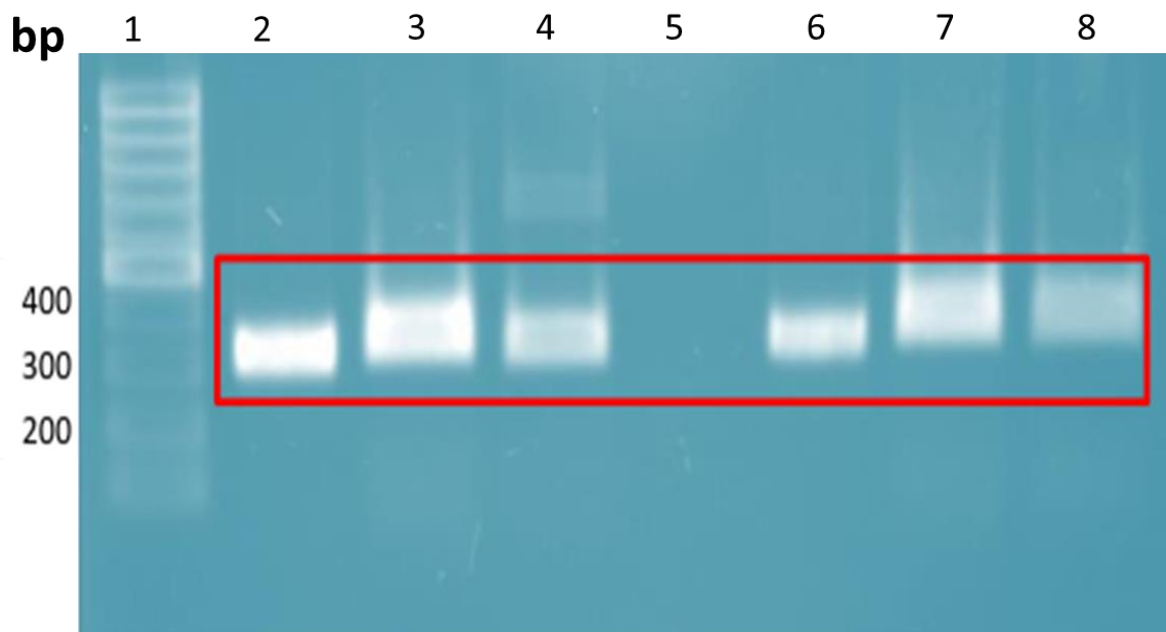


Figure 3.1: The 1% agarose gel used to analyse PCR products of mNC1 and hNC1 gene fragments. The red box marks the bands of PCR products from all experiments. Lane 1: GeneRulerTM 100 bp Plus DNA ladder with DNA band sizes indicated on the left in base pairs (bp); lane 2: mNC1 PCR product double digested before cloning into pGEX-6P-2; lanes 3-4: two colonies PCR products for mNC1; lane 6: hNC1 PCR product double digested before cloning into pGEX-6P-2; lane 7-8: two colonies PCR for hNC1. All digested PCR products and colony PCR products migrated at molecular weight of ~300 bp.

Part 1 B: Protein production and purification

3.2 Protein production in *Escherichia coli* cells

Escherichia coli (*E. coli*) cells are usually used for protein production because they are grown in relatively cheap media and can produce large quantities of proteins. Their limitations include the potential for proteins to aggregate in inclusion bodies, hindering protein immobilisation and purification (Rosano and Ceccarelli, 2014). In this case, solubilisation methods are often used in order to solubilise the protein. Most of the proteins produced in this study were highly soluble with some partially insoluble. The amount of proteins required at the end of purification experiments was ~20 mg per protein prior subsequent studies (**Appendix H1**). All proteins were analysed on SDS-PAGE after size exclusion chromatography and before isothermal titration calorimetry (ITC) (**Appendix H2**).

3.2.1 Production and purification of mNC1 and hNC1 proteins

The GST-mNC1 (fusion protein) was expected to have a molecular weight of ~38 kDa (~26 kDa of GST and ~12 kDa of mNC1) and such a protein band was observed in the SDS-PAGE analysis of a sample taken after induction by IPTG (**Figure 3.2, lane 2**). Furthermore, the fusion protein was found to be produced solubly in *E. coli* cells (**Figure 3.2, lane 4**). The bands corresponding to the fusion protein were also seen in the flow through (**Figure 3.2, lane 5**) and samples taken after wash steps (**Figure 3.2, lanes 6-7**) during purification, implying that some of the protein was not immobilised on the Glutathione Sepharose (GS) beads. After washing unbound proteins, 3C protease was added to cleave mNC1 protein from GST. Cleavage was confirmed by SDS-PAGE, where three bands corresponding to fusion protein, target protein and GST were observed (**Figure 3.2, lane 8**). The mNC1 protein was successfully eluted from the GS beads (**Figure 3.2, lanes 9-11**). However, the protein was contaminated with GST, fusion protein and other unidentified proteins (**Figure 3.2, lane 9**).

The SDS-PAGE analysis representing production and GS affinity chromatography of hNC1 protein was demonstrated in **Figure 3.3**. The results were similar to **Figure 3.2**. However, the smear covering the protein bands in lane 8 made it difficult to visualise protein bands corresponding to fusion, GST and hNC1 proteins. Moreover, the elution fractions (**Figure 3.3, lanes 9-11**) were observed containing hNC1 protein which was also contaminated with fusion protein, GST and other unidentified proteins.

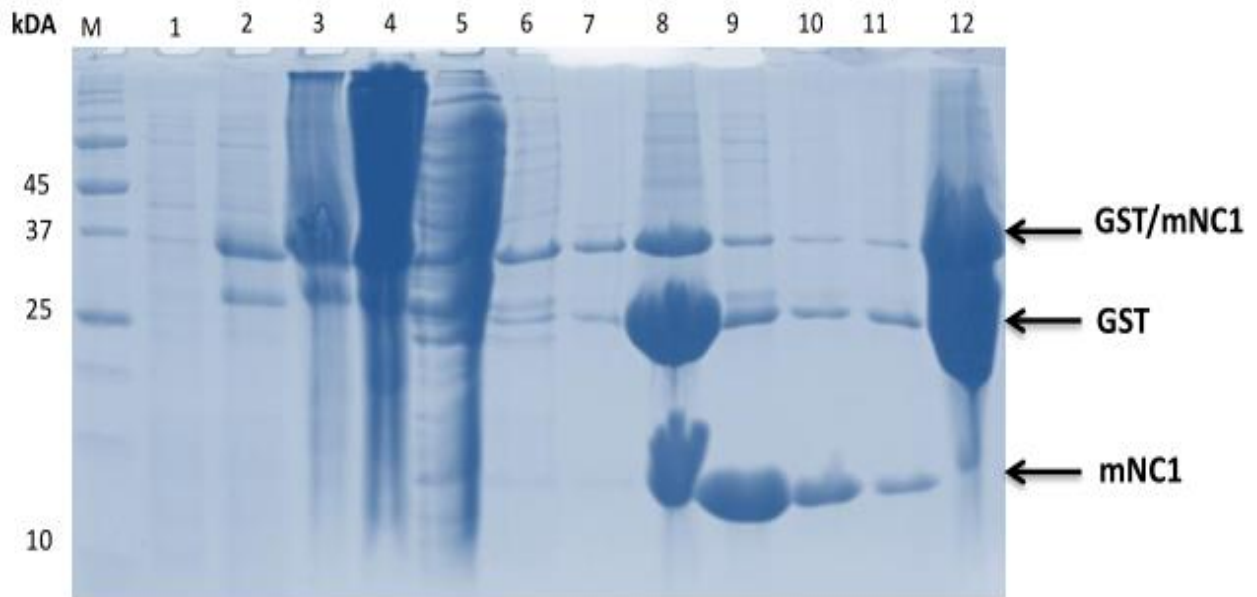


Figure 3.2: Production and GS affinity chromatography analyses of mNC1 protein by SDS-PAGE. M represents sizes of PageRuler Unstained protein marker (indicated on the left in kDa). Lanes 1-12 represent before induction, after induction, insoluble, soluble, flow through, wash 1, wash 2, beads after cleavage, elution 1, elution 2, elution 3 and beads after elution, respectively. The fusion protein, GST and mNC1 are indicated by the black arrows.

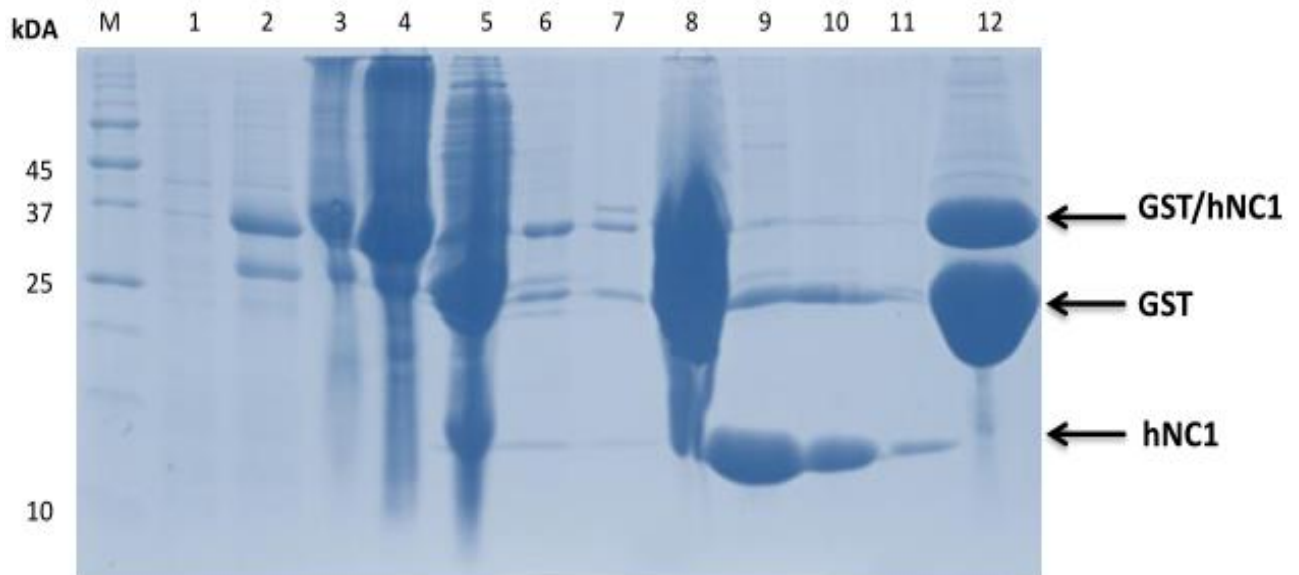


Figure 3.3: Production and GS affinity chromatography analyses of hNC1 protein by SDS-PAGE. The loading pattern is similar to that of previous image of mNC1 production and purification. The fusion protein, GST and hNC1 are also indicated by the black arrows, respectively.

3.2.2 Anion exchange chromatography of mNC1 and hNC1 proteins

The elution fractions (**figure 3.2, lanes 9-11**) were pooled and purified by anion exchange chromatography (AEC). The mNC1 (**figure 3.4, A**, blue arrow) eluted during washing step before the linear salt gradient (light green arrow) started. A small peak was observed in A2 fraction. This and all other fractions were analysed by SDS-PAGE and confirmed to be of pure mNC1 protein (**figure 3.4, B**). The red arrow indicated the mNC1 protein band on SDS-PAGE. The lane representing small peak of fraction A2 had two bands, looking like protein degradation occurred. The 4 mL fractions were pooled and quantified, yielding 26 mg protein.

The AEC of hNC1 protein was demonstrated in **figure 3.5, (A)**. Similar observations from **figure 3.4** were also made here. The fractions were collected and analysed on SDS-PAGE for purity, thus pure hNC1 protein (red arrow) was observed (**figure 3.5, B**). The 4 mL fractions were pooled and quantified to 29 mg, followed by storage at 4°C.

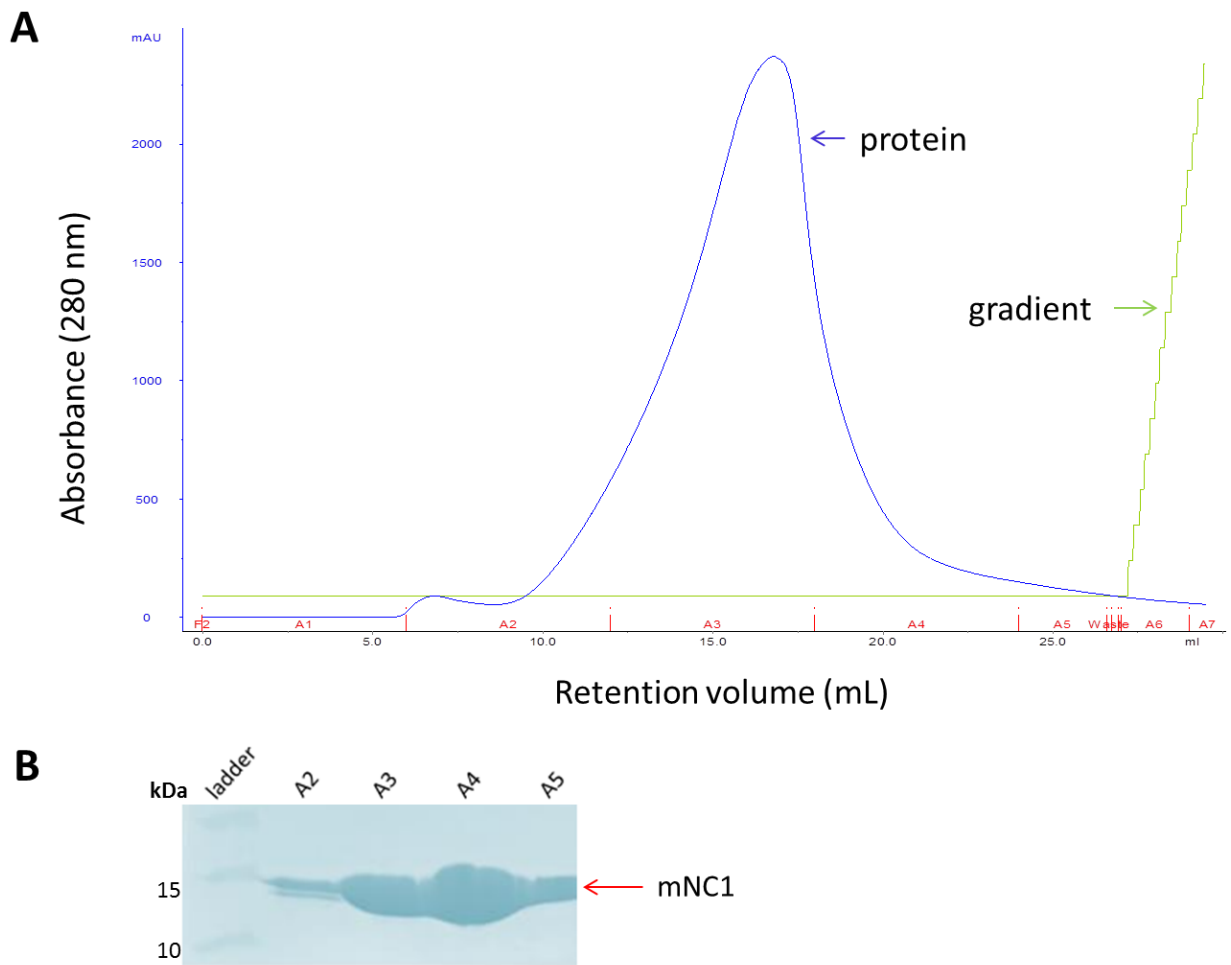


Figure 3.4: Anion exchange chromatography and SDS-PAGE analysis of mNC1 protein. (A) The purification of mNC1 by AEC. The protein eluted during washing step before salt gradient started. The protein absorbance and salt concentration gradient are shown as blue and light green arrows, respectively. (B) The fractions were analysed on SDS-PAGE for purity. All fractions analysed were of pure mNC1 protein as indicated by the red arrow even though possible degradation might have occurred in fraction A2.

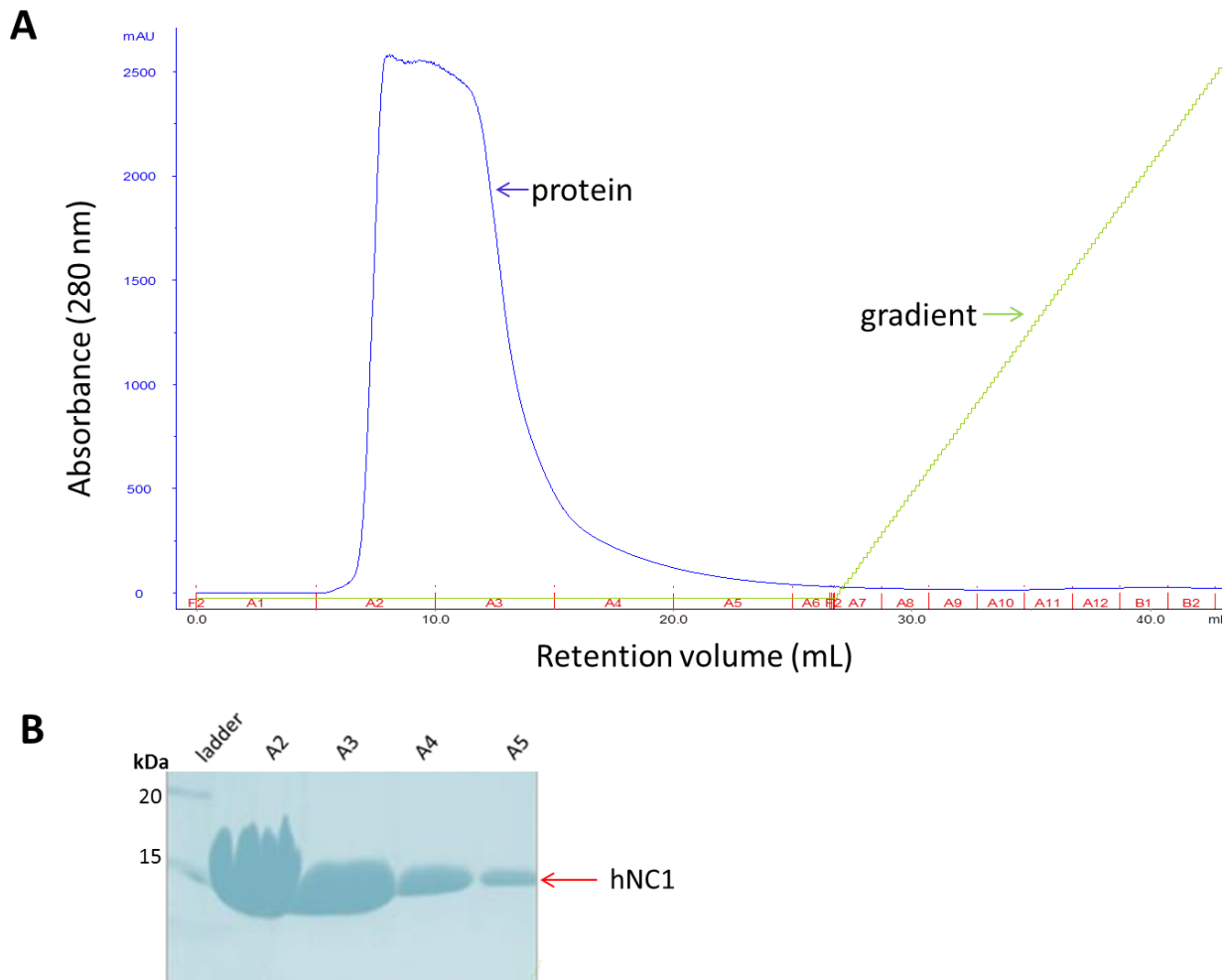


Figure 3.5: Anion exchange chromatography and SDS-PAGE analysis of hNC1 protein. (A) The protein (absorbance indicated by blue arrow) eluted during washing step before linear salt gradient (that is indicated by a light green arrow) started. (B) The fractions assessed on SDS-PAGE for purity were confirmed to be of pure hNC1 protein as indicated by the red arrow.

In these experiments, what was highly required was to purify target proteins and this was achieved by AEC regardless of proteins eluting in the void volumes. The buffer exchange experiments were performed using size exclusion chromatography (SEC). The essence of doing this was to ensure that all proteins were in HEPES buffer required for isothermal titration calorimetry (ITC) experiments. Therefore the fractions from SEC buffer exchange experiments were pooled and quantified, yielding 21 mg and 26 mg for mNC1 and hNC1 proteins, respectively.

3.2.3 Production and purification of mEC1 and hEC1 proteins

The gene fragments coding for mEC1 and hEC1 proteins were cloned into pGEX-6P-1 (see vector map in **Appendix C**). The constructs were confirmed by Sanger sequencing, followed by protein production similarly to section 3.2.1.

The GST-mEC1 (fusion protein) has theoretic molecular weight of ~38 kDa and corresponding band was observed in the lane of sample taken after induction by IPTG (**Figure 3.6, lane 2**). This band was not visualised in the lane representing soluble fraction due to a smear (**Figure 3.6, lane 4**). However, the corresponding fusion protein band was observed in flow through (**Figure 3.6, lane 5**) and washes lanes (**Figure 3.6, lane 6-7**) indicating that some of the protein was not bound to the GS beads. Following 3C protease cleavage, three bands were observed (**Figure 3.6, lanes 9 to 11**) representing from the top to bottom; fusion protein, GST and mEC1 protein, respectively. These bands were indicated by black arrows. Therefore, the elution fractions were pooled and quantified to 16 mg and stored at 4°C.

The production and purification analysis of hEC1 (**Figure 3.7**) was done similarly to mEC1 (**Figure 3.6**). The GST-hEC1 (fusion protein), GST and unidentified protein bands were observed in elution fraction together with hEC1 (**Figure 3.7, lane 8**). The elution fractions (including those not shown) were pooled and quantified to 13 mg of protein.

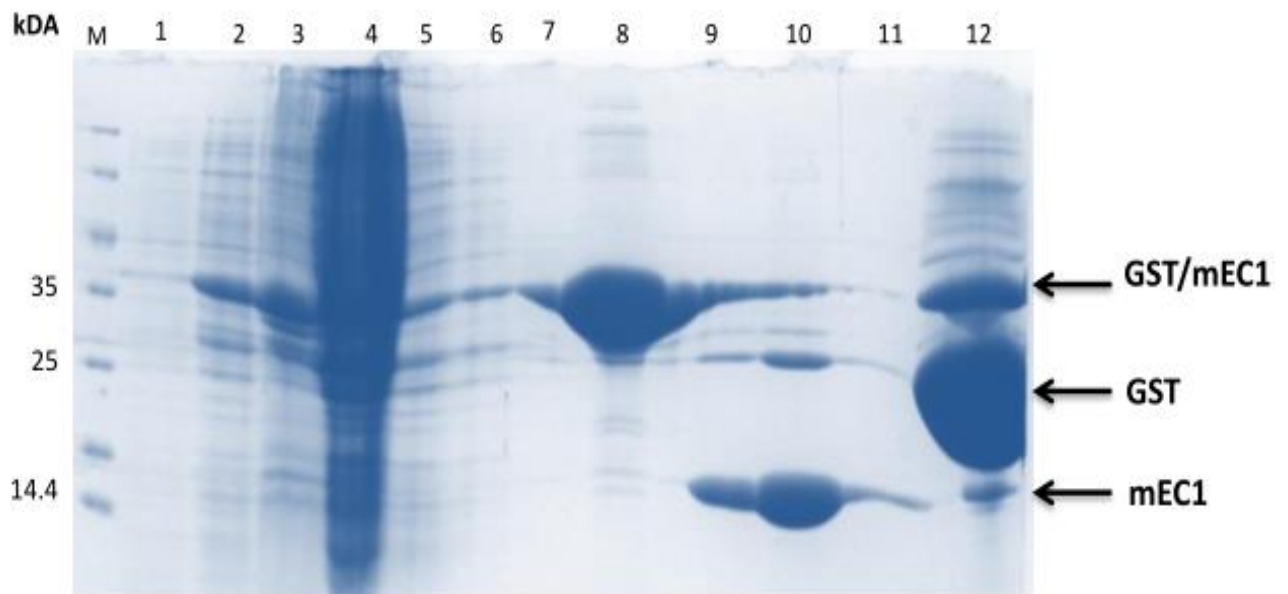


Figure 3.6: Production and GS affinity chromatography analyses of mEC1 protein by SDS-PAGE. M represents sizes of Pierce Unstained protein marker (indicated on the left in kDa). Lanes 1-12 represent before induction, after induction, insoluble, soluble, flow through, wash 1, wash 2, beads before cleavage, elution 2, elution 1, elution 3 and beads after elution samples, respectively.

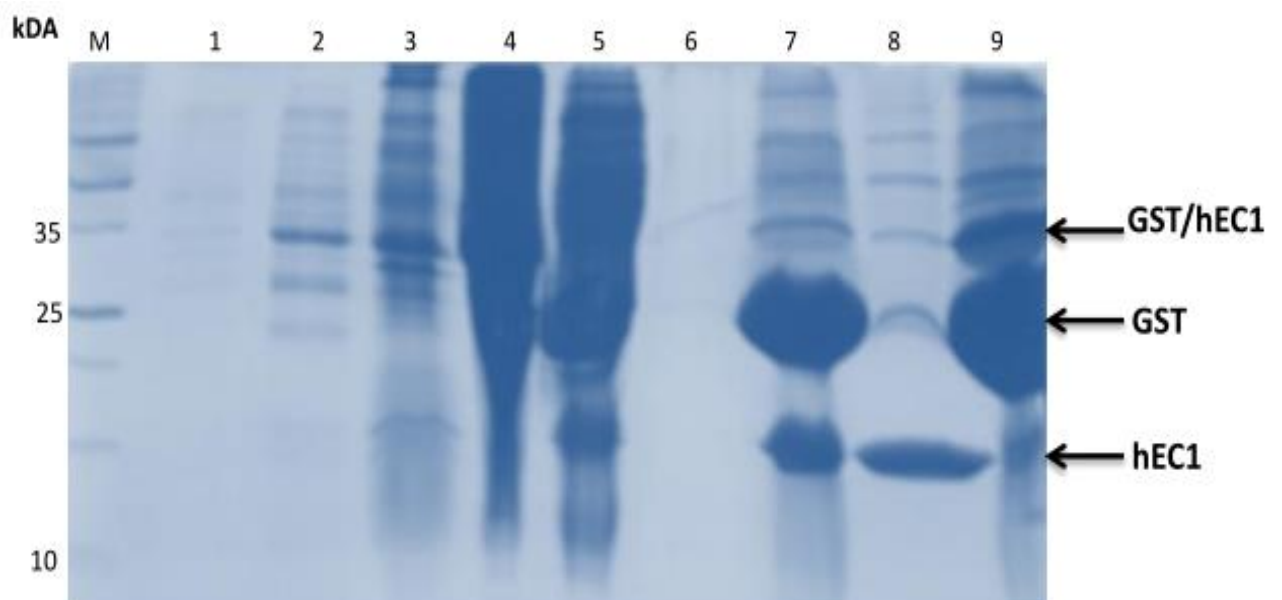


Figure 3.7: Production and GS affinity chromatography analyses of hEC1 protein by SDS-PAGE. The loading pattern is as follows; M represents Pierce Unstained protein marker, lanes 1-9 represent before induction, after induction, insoluble, soluble, flow though, wash 1, beads after cleavage, elution 1 and beads after elution, respectively.

3.2.4 Anion exchange chromatography of mEC1 and hEC1 proteins

The mEC1 was purified by AEC as shown in **Figure 3.8 (A)**. Therefore, the protein (blue arrow) eluted as a sharp peak at 38% of high salt buffer during linear salt gradient (green arrow). The peak fractions from A12 to B5 were assessed by SDS-PAGE and found to be of pure mEC1 protein (**Figure, 3.8, B**), also indicated by red arrow. The fractions were then pooled in order to do buffer exchange for ITC experiment. Buffer exchange experiment was done in HEPES buffer using SEC. The pooled samples were quantified to 9 mg and stored at 4°C.

During the purification of hEC1 (**Figure 3.9**), two small peaks (black arrow) in the void volume were observed and were thought to be that of two unidentified bands seen in **Figure 3.6**. However, the hEC1 protein (blue arrow) eluted at 40% of linear salt gradient (light green arrow) (**Figure 3.9, A**). The elution fractions (B9-B1) contained pure hEC1 (red arrow) when analysed on SDS-PAGE (**Figure 3.9, B**). The fractions were then pooled, followed by buffer exchange. After this, the fractions were pooled again, quantified to 11 mg and stored at 4°C.

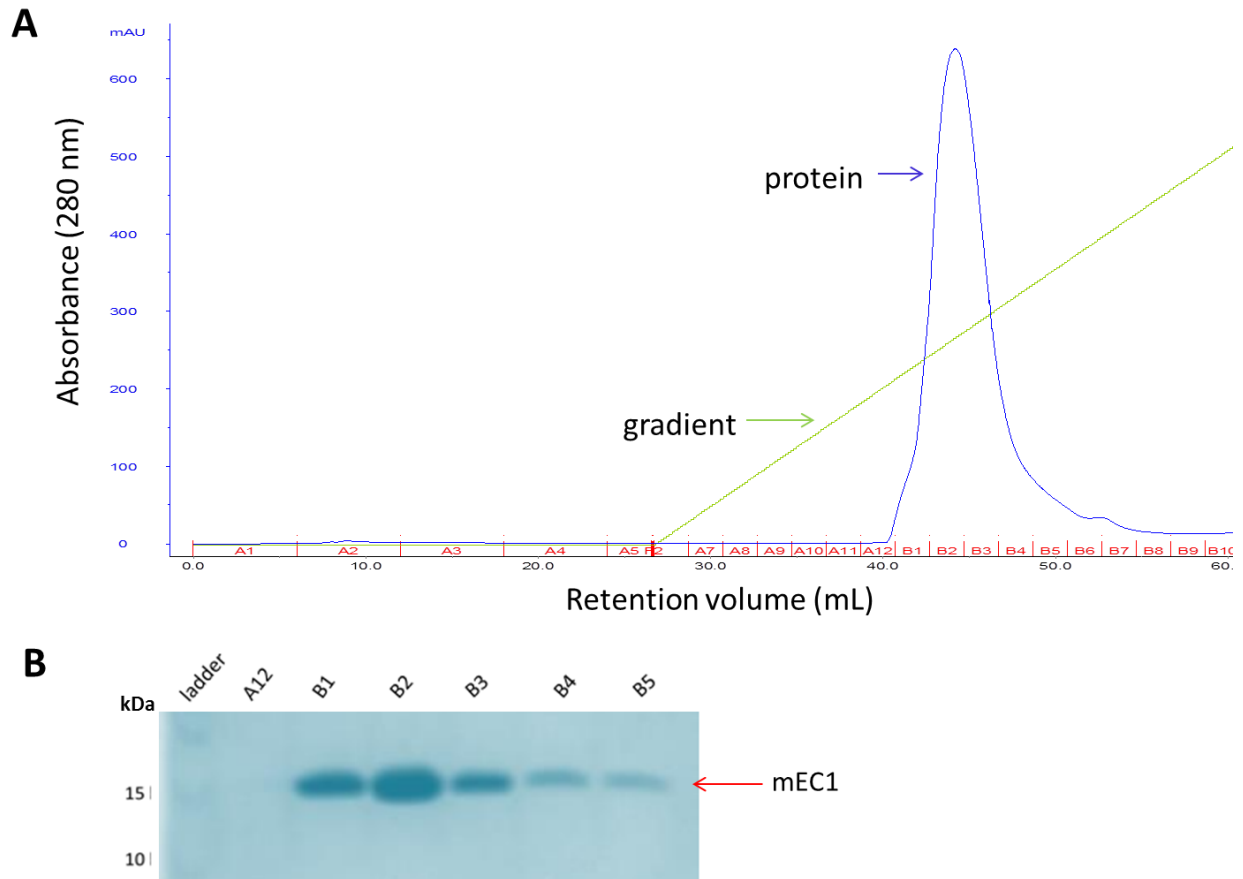


Figure 3.8: Anion exchange chromatography and SDS-PAGE analysis of mEC1 protein. (A) The results showing purification of mEC1 by AEC. The protein eluted as sharp peak at a salt concentration of 38% (of high salt buffer). The salt gradient and protein absorbance are indicated by light green and blue arrows, respectively. (B) The fractions were collected and assessed on SDS-PAGE for purity. The pure bands of mEC1 protein as indicated by a red arrow migrated at ~12 kDa.

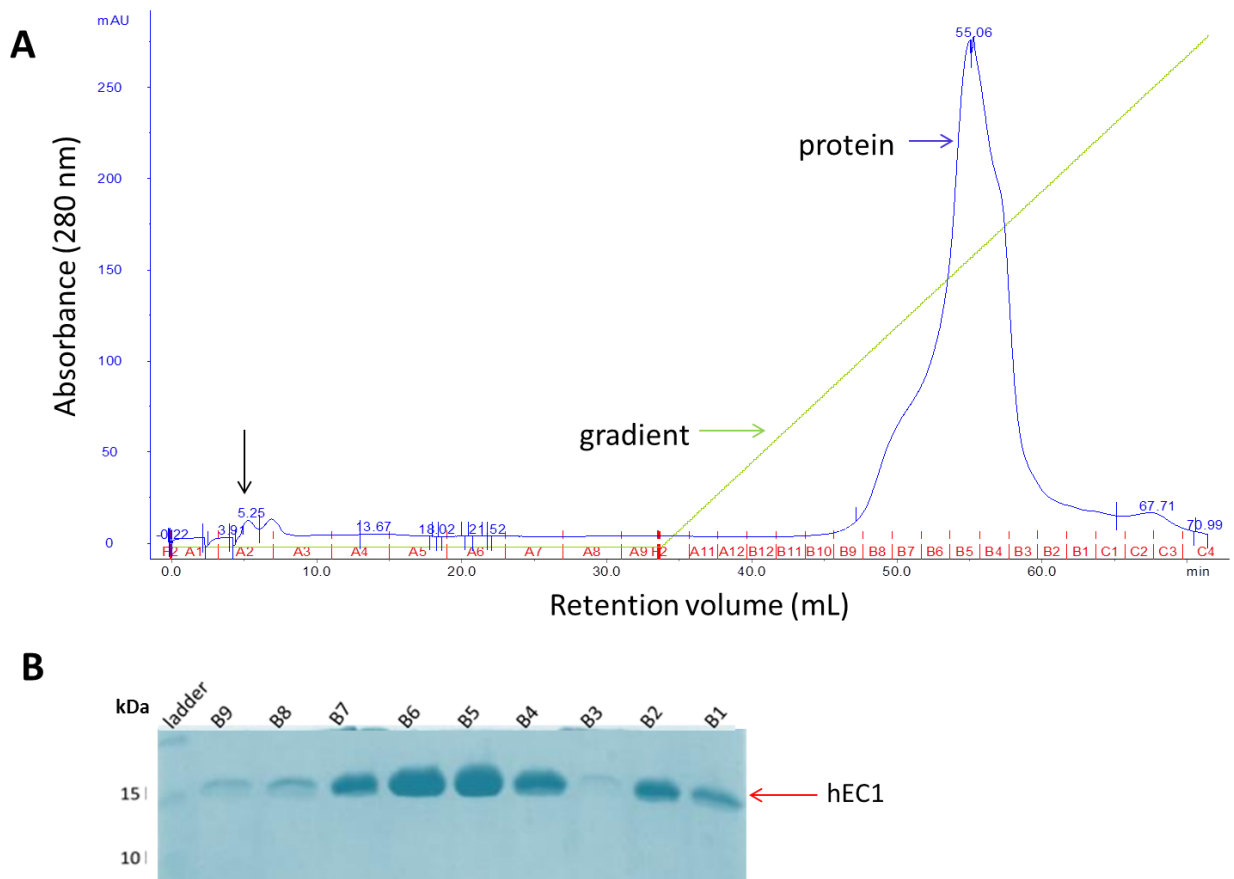


Figure 3.9: Anion exchange chromatography and SDS-PAGE analysis of hEC1 protein. (A) The purification of hEC1 by AEC shows that the protein eluted at 40% salt gradient. The salt gradient and protein absorbance were indicated by light green and blue arrows, respectively. (B) The SDS-PAGE analysis for purity assessment shows pure hEC1 protein bands (red arrow), migrated at 12 kDa.

At the end of these experiments, mEC1 and hEC1 proteins were stored after meeting purity requirements but the concentrations did not meet targeted value of ~20 mg or more. Therefore both proteins were produced and purified again in the same manner as above and the combine fractions were quantified to 29 mg and 22 mg for mEC1 an hEC1 proteins, respectively. The proteins were stable for long periods of time without any sign of degradation or proteolysis during storage at 4°C.

3.2.5 Generation of InA^m variant by mutagenesis

The InA^m protein was engineered to investigate whether it binds to either mNC1 or hNC1 protein during ITC. To achieve this, GST-InA^{wt} construct was sequenced using Sanger method. Following this, the InA^{wt} sequence was used to design primers for site directed mutagenesis similarly to previous study (Wollert *et al.*, 2007). The two point mutations were introduced by conventional PCR, thus the codon for serine 192 was replaced to asparagine resulting in GST-InA^{S192N} variant or construct. This construct was used again to replace tyrosine on position 369 with serine making InA^{S192N/Y369S} or InA^m variant.

The site directed mutagenesis results were confirmed by Sanger sequencing (**Figure 3.10**). The red arrows indicated the modified nucleotides and below were the corresponding amino acids. The forward primer was used to highlight the nucleotide sequence wherein mutation (S192N) was introduced in *inlA* gene (**Figure 3.10, A**). The reverse sequence of reverse primer was used to highlight nucleotide sequence wherein mutation (Y369S) was introduced (**Figure 3.10, B**).

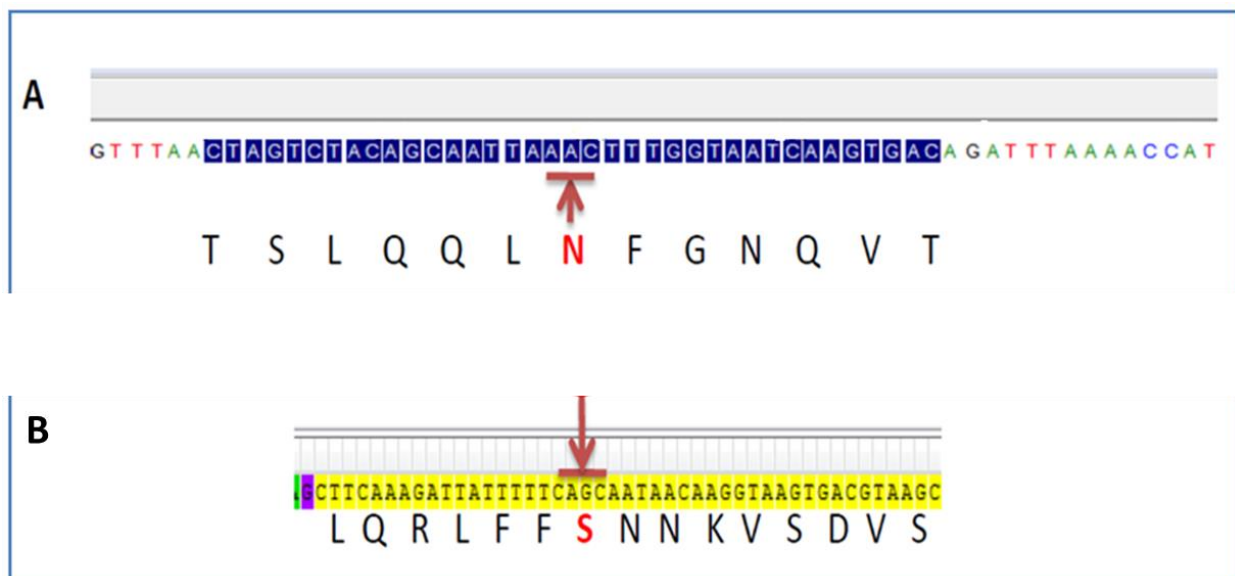


Figure 3.10: The Sanger sequencing results of GST-InlA^m construct. (A) The sequence obtained from Sanger sequencing using pGEX-5' primer to confirm S192N mutation. As indicated with translated amino acids, serine 192 was replaced by asparagine. (B) The sequence obtained when pGEX-3' primer was used to confirm Y369S mutation. The reverse sequence was reverse complimented and as indicated, tyrosine 369 was replaced by serine after translating nucleotide sequence to amino acids.

3.2.6 Production and purification of InlA variants

The InlA variants were previously produced as soluble proteins in *E. coli* cells (Schubert *et al.*, 2002; Wollert *et al.*, 2007). The productions and purifications of InlA variants were carried out similarly to those of N- and E- cadherin N-terminal domains.

3.2.6.1 Production and purification of InlA variants

The expected molecular weight of GST-InlA^{wt} (fusion protein) was ~75 kDa (~26 kDa of GST and ~49.5 kDa of InlA^{wt} protein) and the representative bands were indicated by black arrows, respectively (**Figure 3.11**). The fusion protein was successfully produced after induction of *E. coli* BL21 (DE3) cells by IPTG (**Figure 3.11, lane 2**). The fusion protein was soluble and successfully cleaved by 3C protease; separating InlA^{wt} and GST (**Figure 3.11, lanes 4 and 10**, respectively). The elution fractions (**Figure 3.11, lanes 11-13**) contained mostly InlA^{wt} protein with a molecular weight of ~49.5 kDa. However, the lanes representing these elution fractions also contained faint protein bands that corresponded to GST. The elution fractions were pooled and quantified to 16 mg and stored at 4°C until further purification of the protein was done.

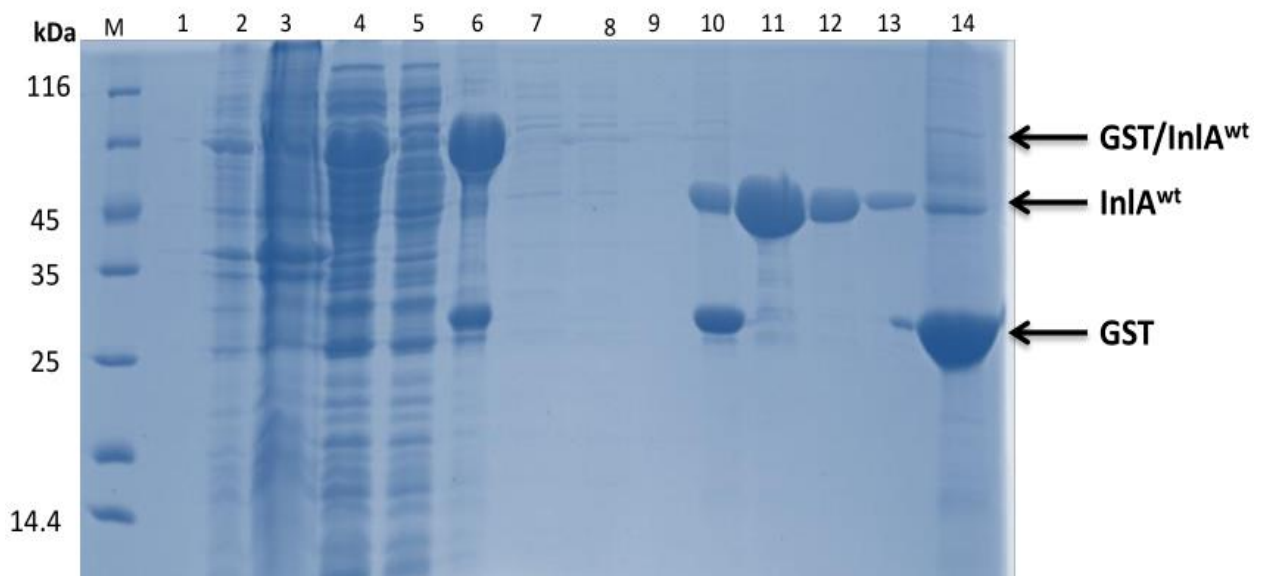


Figure 3.11: Production, GS affinity chromatography and SDS-PAGE analysis of InlA^{wt} protein. M represents sizes of Pierce Unstained protein marker (indicated on the left in kDa). Lanes 1-14 represent before induction, after induction, insoluble, soluble, flow through, beads before wash, wash 1, wash 2, wash 3, beads after cleavage, elution 1, elution 2, elution 3 and beads after elution samples, respectively. The fusion protein, InlA^{wt} and GST are indicated by the black arrows.

The pooled and quantified elution fractions were further purified by the AEC (**figure 3.12, A**). The protein absorbance and salt gradient were indicated by blue and light green arrows, respectively. The InlA^{wt} protein eluted at 50% of high salt buffer as a sharp peak during linear salt gradient. All 3 mL fractions were analysed on SDS-PAGE and confirmed to contain pure InlA^{wt} protein (**figure 3.12, B**). Therefore, all fractions were pooled and quantified to 13 mg. Since the amount of InlA^{wt} protein was lower than ~20 mg, the protein production and purification was repeated in order to reach this target concentration. The protein solutions were pooled and buffer exchange done. The fractions were pooled again and quantified to 28 mg.

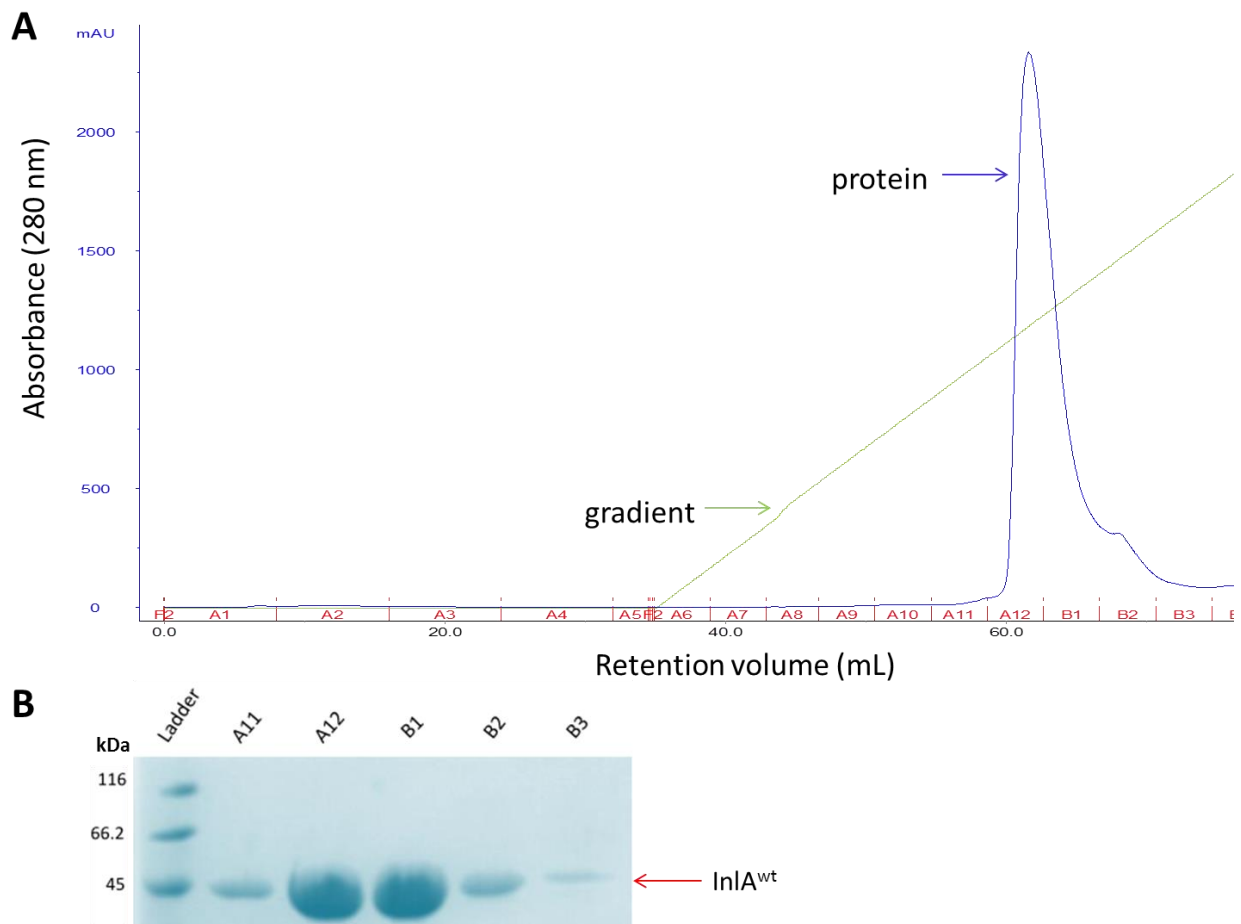


Figure 3.12: Anion exchange chromatography and SDS-PAGE analysis of InlA^{wt} protein. (A) The purification of InlA^{wt} protein by AEC. The protein eluted as sharp peak at 50% of salt gradient (from 1 M salt concentration). Indicated by blue and light green arrows were protein absorbance and salt gradient, respectively. (B) The fractions were collected and assessed on SDS-PAGE for purity. The pure InlA^{wt} protein bands as indicated by a red arrow migrated at ~49.5 kDa.

The InIA^{S192N} variant was produced and purified similarly to **figure 3.11**. However, the fusion protein was not clearly observed in lane of bacterial sample taken after induction by IPTG and lane of soluble fraction, respectively (**figure 3.13, lanes 2 and 4**). This is due to loading inadequate and excessive amounts of samples on the gel, respectively. The fusion protein was cleaved by 3C protease (**figure 3.13, lane 5**) and target protein bands were visualised (**figure 3.13, lanes 6-7**). The elution fractions were clearly not pure as faint GST and unknown proteins bands were observed. These elution fractions, including the third one not shown here, were pooled and quantified to 18 mg. The protein production and purification was repeated to increase the concentration. The amount obtained prior to AEC was 34 mg.

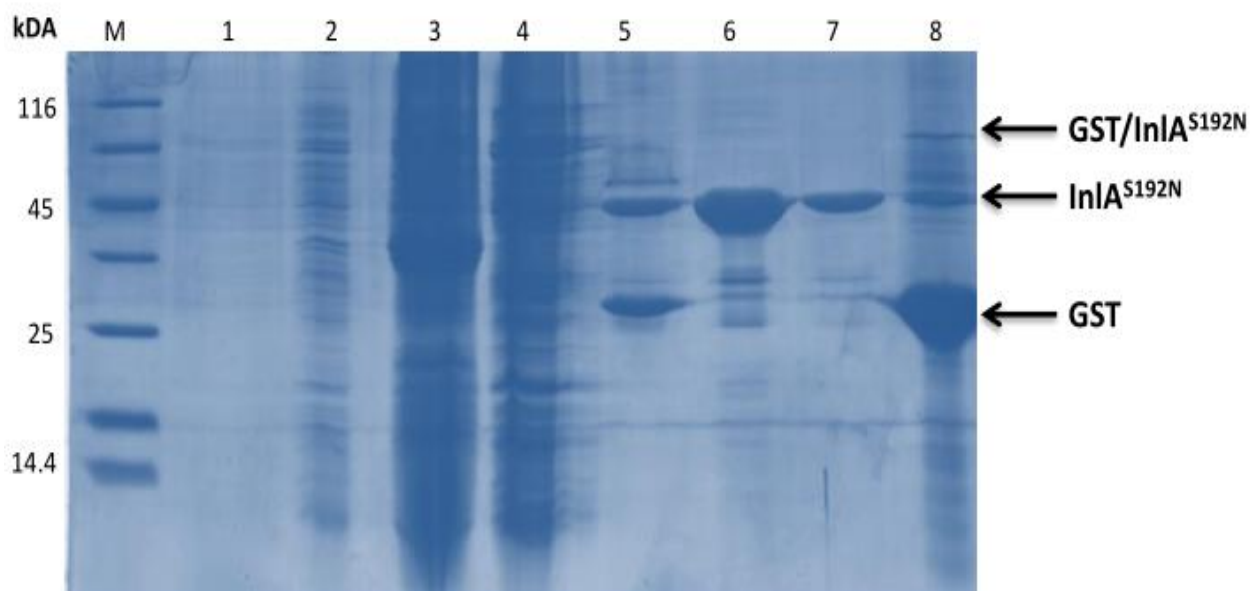


Figure 3.13: Production, GS affinity chromatography and SDS-PAGE analysis of InIA^{S192N} protein. M represents sizes of Pierce Unstained protein marker. Lanes 1-8 represent before induction, after induction, insoluble, soluble, beads after cleavage, elution 1, elution 2 and beads after elution samples, respectively. The fusion protein, InIA^{S192N} protein and GST were indicated by the black arrows.

The elution fractions (**figure 3.13, lanes 6-7**) were purified by the AEC (**figure 3.14, A**). During linear salt gradient, three peaks were observed. The target protein eluted as a sharp peak at 50% of high salt buffer and was confirmed by SDS-PAGE analysis (**figure 3.14, B**). The target protein, however, was contaminated by GST and unidentified proteins, thus these

contaminants were observed in fractions from second and third peaks (**figure 3.14, B, lanes B10 and B12**, respectively).

Since AEC did not purify the target protein successfully, the GS affinity chromatography was done by adding the protein solution from fractions B6 to B10 into clean GS beads. This was done to allow GST to bind to GS beads while target protein flows through. The target protein was collected as flow through. Therefore the protein solution was pooled again in order to do buffer exchange. The protein was pooled, quantified to 29 mg and stored at 4°C.

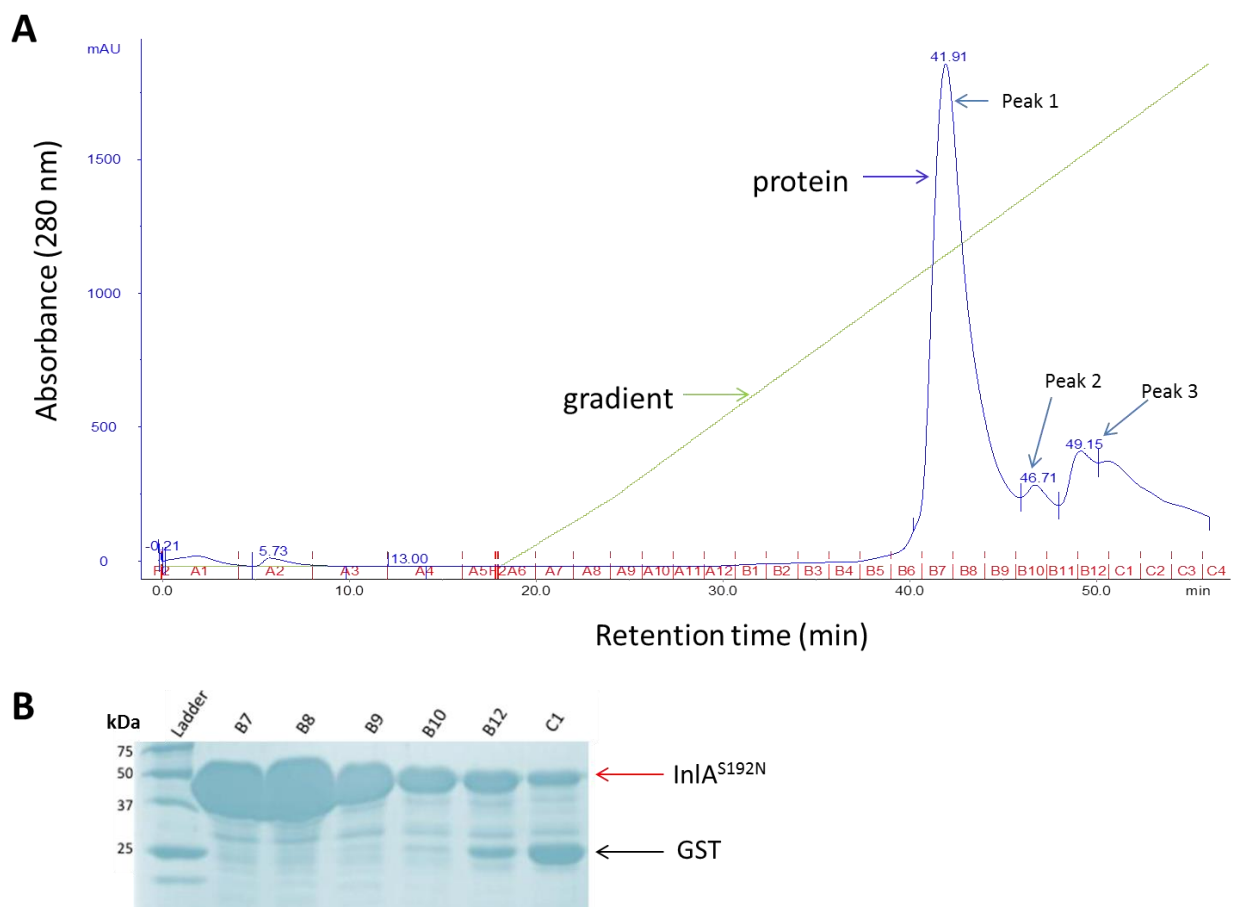


Figure 3.14: Anion exchange chromatography and SDS-PAGE analysis of InIA^{S192N} protein. (A) The protein eluted as sharp peak at a salt concentration of 50% (from 1 M salt concentration) and two other peaks followed. The protein absorbance and salt gradients were indicated by blue and light green arrows, respectively. (B) The fractions were collected and assessed on SDS-PAGE for purity. InIA^{S192N} protein bands as indicated by a red arrow migrated at ~49.5 kDa and were not pure. All fractions were contaminated with GST (black arrow) and unidentified protein bands.

It was realised that the first two InIA variants were obtained with high concentrations from repeating the experiments. As a result it was then decided to produce GST-InIA^{Y369S} (fusion protein) in BL21 (DE3) cells in increased volumes. The InIA^{Y369S} protein was produced and purified similarly to other InIA variants. The fusion protein was not visualised in lanes representing after IPTG induction sample and soluble fraction due to excessively dilution and overloading of samples on the gel, respectively (**figure 3.15, lanes 2 and 4**). However, the InIA^{Y369S} protein was successfully eluted from GS beads after 3C protease cleavage (**figure 3.15, lanes 8-10**), but was contaminated with fusion protein and GST. The elution fractions were pooled, quantified to 29 mg and stored at 4°C.

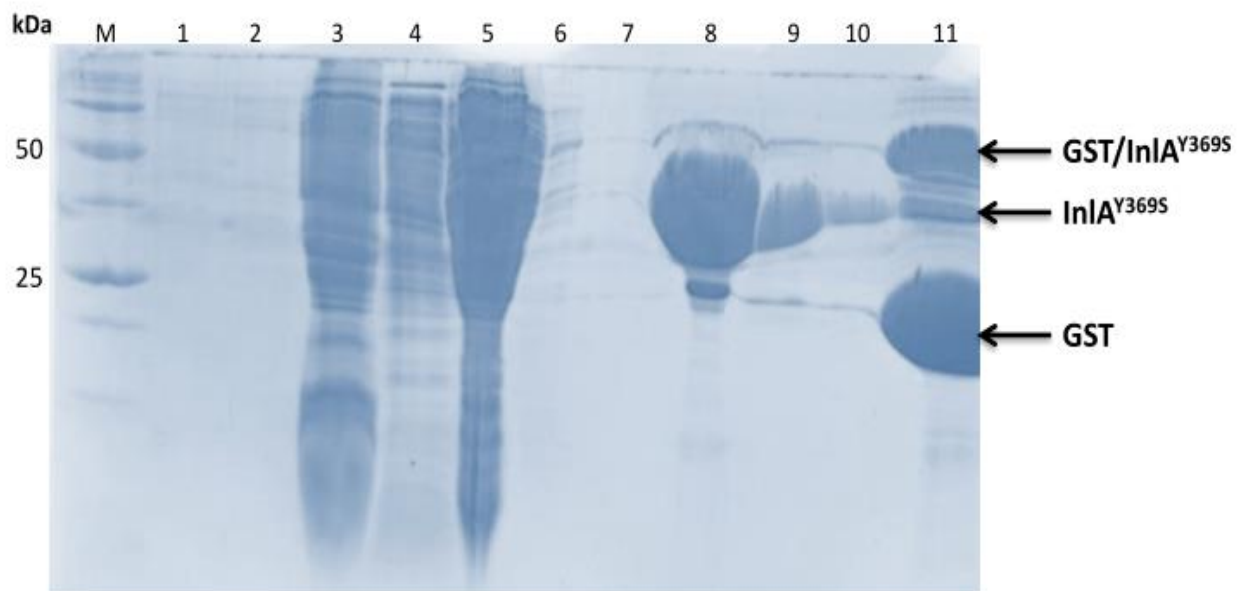


Figure 3.15: Production, GS affinity chromatography and SDS-PAGE analysis of InIA^{Y369S} protein. The 15% SDS-PAGE gel showing samples taken during protein production in *E. coli* BL21 cells and after purification by GS affinity chromatography. M represents sizes of PageRuler Unstained protein marker. Lanes 1-11 represent before induction, after induction, insoluble, soluble, flow through, wash1, wash 2, elution 1, elution 2, elution 3 and beads after elution samples, respectively. The fusion protein, InIA^{Y369S} and GST are indicated by the black arrows.

The AEC of InIA^{Y369S} protein was done to eliminate the contaminants (**figure 3.16, A**) and two peaks were observed during linear salt gradient. These peaks eluted at high salt buffer concentration of 50% and 65%, respectively. Therefore, the collected fractions were assessed for purity on SDS-PAGE. The protein bands from the first peak were mainly of target protein

(red arrow) even though faint GST bands were observed (**figure 3.16, B, B6-B11**). However, the second peak contained both target protein and GST (black arrow). Then fractions from B6 to B12 were chosen for another purification experiment.

The 3 mL fractions (B6-B12) were pooled and added to unused GS beads to bind excess GST to the GS beads. The flow through was collected and then pooled to do buffer exchange. After buffer exchange, the target protein was pooled, quantified to 23 mg and stored at 4°C.

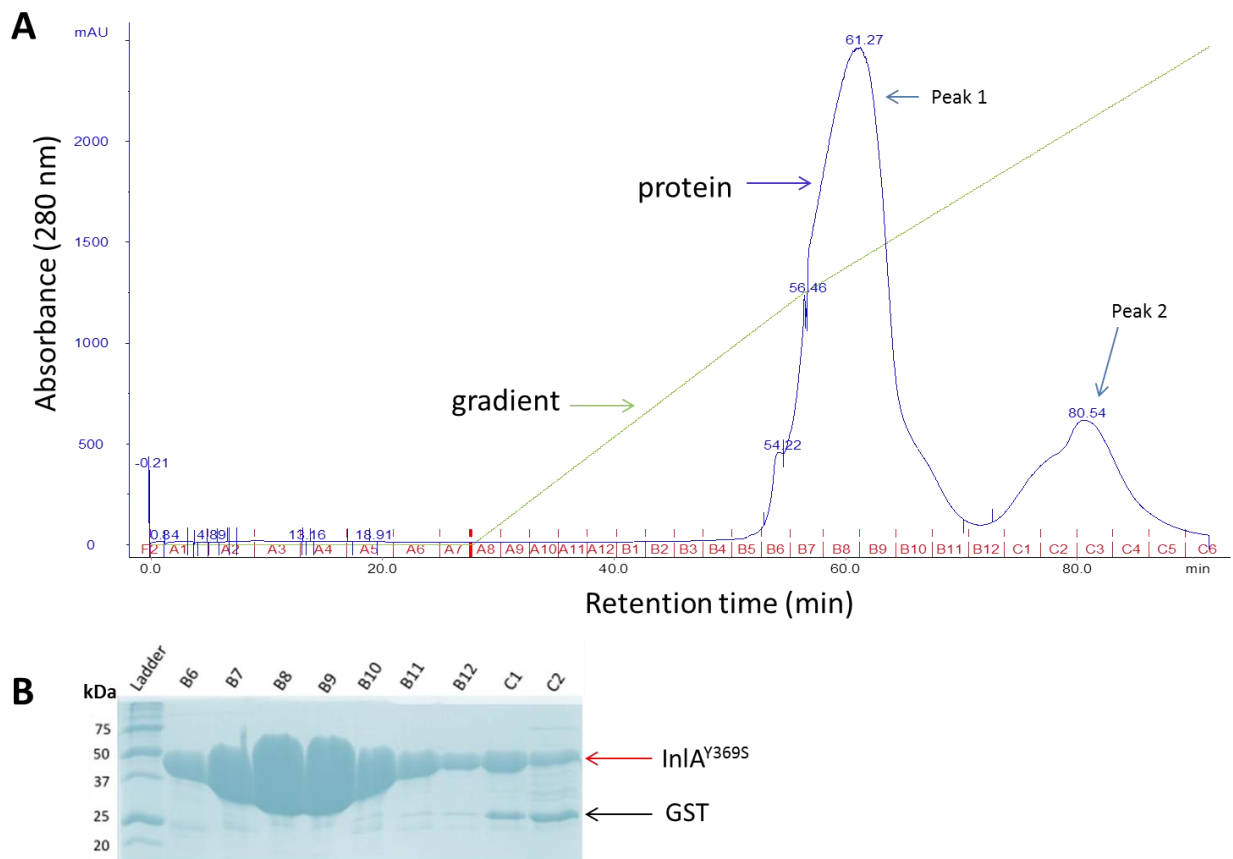


Figure 3.16: Anion exchange chromatography and SDS-PAGE analysis of InIA^{Y369S} protein. (A) The purification of InIA^{Y369S} protein using AEC. The protein eluted at salt gradient of 50% (from 1 M salt concentration) and one other peak followed. (B) The fractions were collected and analysed on SDS-PAGE for purity. InIA^{Y369S} protein bands as indicated by a red arrow migrated at ~49.5 kDa (red arrow). The second peak contained InIA^{Y369S} as well as GST (black arrow).

The GST-InIA^m fusion protein (~75 kDa), InIA^m (~49.5 kDa) and GST (~26 kDa) were indicated by black arrows, respectively on the SDS-PAGE (**figure 3.17**). The small band

corresponding to fusion protein was detected (**figure 3.17, lane 2**) and this observation suggested that bacterial cells were successfully induced by IPTG to produce soluble protein (**figure 3.17, lane 4**). However, the protein was partially insoluble (**figure 3.17, lane 3**). The InIA^m protein was cleaved from GST by 3C protease as three bands representing these two proteins plus fusion protein band were visualised (**figure 3.17, lane 7**). The InIA^m protein was successfully eluted (**figure 3.17, lane 8**), but the sample was not pure because unidentified faint bands including fusion protein and GST were present. This elution fraction and other two not analysed on this SDS-PAGE were pooled and quantified to 26 mg.

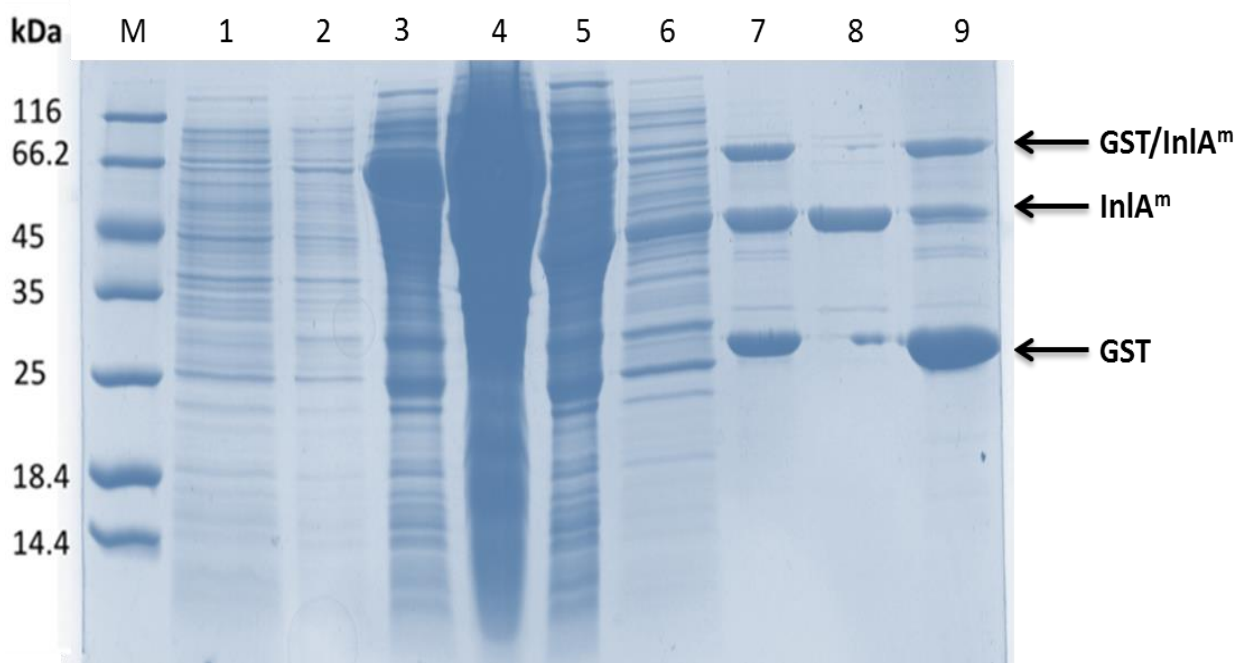


Figure 3.17: Production, GS affinity chromatography and SDS-PAGE analysis of InIA^m protein. M represents sizes of Pierce Unstained protein marker. Lanes 1-9 represent before induction, after induction, insoluble, soluble, flow through, wash 1, beads after cleavage, elution 1 and beads after elution samples, respectively. The fusion protein, InIA^m and GST are indicated by the black arrows.

The AEC was used to purify about 1.5 mL of InIA^m protein (**figure 3.18, A**). Three peaks were observed, the first one eluting at 50% of high salt buffer, starting with a shoulder before conforming to a sharp peak. The 3 mL fractions collected from these peaks were analysed on SDS-PAGE (**figure 3.18, B**). The InIA^m protein bands (red arrow) were observed at molecular weight of ~49.5 kDa from the first peak (**figure 3.18, B, lanes B8-B12**). The

fractions from peak 2 and 3 were observed containing fusion protein (blue arrow), target protein and GST (black arrow) in addition to other unidentified protein bands, respectively. Therefore the third purification experiment was required to eliminate impurities that were contaminating InIA^m protein.

Therefore, the fractions from B8 to B12 were pooled and purified by GS affinity chromatography. The buffer exchange was done using SEC. All fractions were pooled and quantified to 30 mg and stored at 4°C.

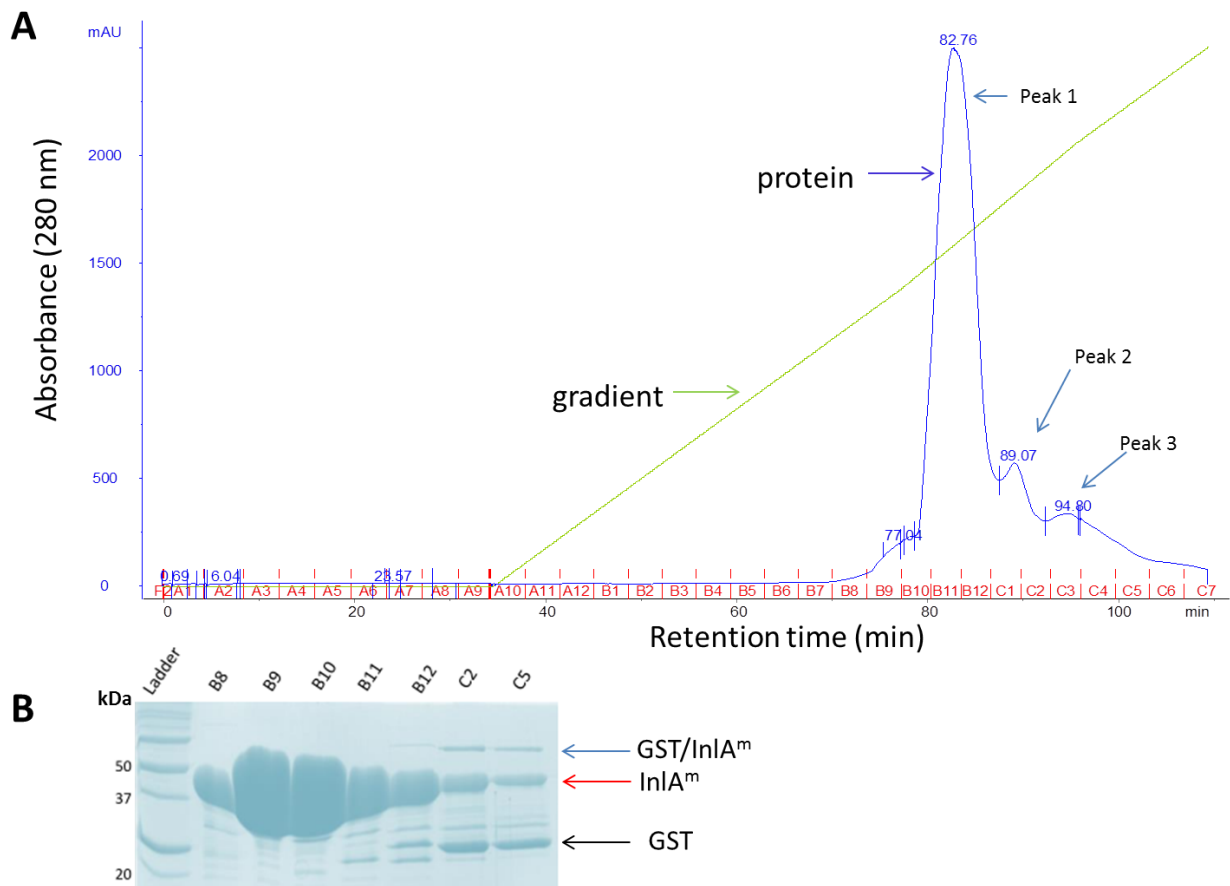


Figure 3.18: Anion exchange chromatography and SDS-PAGE gel analysis of InIA^m protein. (A) The purification of InIA^m by AEC. The protein eluted as sharp peak starting with a shoulder at 50% salt gradient and two other peaks followed. The protein absorbance and salt gradient were indicated by blue and light green arrows, respectively. (B) The fractions were assessed on SDS-PAGE for purity. Protein bands of InIA^m as indicated by a red arrow migrated at ~49.5 kDa. The two other peaks contained InIA^m protein as well as fusion protein and GST. However, all collected fractions were not pure as observed on SDS-PAGE.

3.3 Conclusion

At the end of part 1, we had stored all proteins with good purities and concentrations. The purity of all proteins after SEC experiment and concentrating was analysed on SDS-PAGE (**Appendix H**).

Part 2: Biophysical studies by isothermal titration calorimetry

3.4 Introduction

Isothermal titration calorimetry (ITC) was used in this study to determine the binding affinities or dissociation constants (K_d) between InlA variants and cadherin domains. In a typical ITC experiment, a solution of a ligand “biomolecule” is titrated into a solution of its binding partner or target molecule(s) inside the cell (Leavitt and Freire, 2001). In this study, the ligand was one of a range of InlA variants (InlA^{wt}, InlA^{S192N}, InlA^{Y369S} and InlA^m proteins) while the target protein solutions were of mEC1, hEC1, mNC1 and hNC1 proteins, respectively.

In an ITC experiment, each peak represents the endo- or exothermic reaction occurring as a result of interacting proteins (Leavitt and Freire, 2001). Heat signals return to the baseline when there is complete protein-protein interaction (Leavitt and Freire, 2001; Duff *et al.*, 2011). When the target protein is saturated, the heat signal diminishes until only the heat of dilution is observed –indicating that no further binding will occur. A binding curve is then generated by the software which integrates all heat points from each injection against the ratio of ligand and target protein (Doyle, 1997). A binding curve is then fitted to a desired binding model to estimate the binding parameters (Leavitt and Freire, 2001).

3.5 ITC control experiments

The first ITC experiments done were controls in order to ensure that the system parameters set were reliable. These control experiments were done to demonstrate how the binding data will look compared to non-binding data. Before experiment was run, the ligand was filled in the syringe and several purge cycles were performed to eliminate air bubbles in the solution. The air bubbles were also avoided in the sample cell by slowly pipetting the target solution into the sample cell. Then the experiment was run using parameters indicated in chapter 2, section 2.8.1.

The non-binding control was demonstrated (**Figure 3.19, A**), in which the InlA^{Y369S} protein was titrated into a sample cell containing HEPES buffer. What we expected was the heat of dilution signals that indicate no interaction occurring. Here what was initially observed was that the first three titrations resulted in heat signals pointing up and down and these were

thought to be as a result of possible air bubbles present in one of the samples. However, the rest of ligand titrations resulted with heat of dilution signals as expected.

The binding experiment of CaCl₂ solution as a ligand and EDTA solution as a target solution was demonstrated (**Figure 3.19, B**). The EDTA is a chelating agent that has been reported to form strong complexes with divalent cation metals such as Ca²⁺ in aqueous solutions. This complex formation has been studied previously by ITC, resulting in an exothermic reaction (Griko, 1999). In this study, the binding of Ca²⁺ ions to EDTA was characterised by the similar reaction (upper panel). The first titration heat signal was short because 1 μL of ligand solution was titrated into the sample cell while the rest of injections were of 2 μL and resulted in longer signals. The saturation of target solution was reached after nine injections as heat signals were diminishing. The binding curve (lower panel) was generated from the plot of isotherms and fitted to one site binding model in order to determine the K_d which was found to be 0.6 μM. All binding parameters were shown in the bottom right box and K_d determination calculation example shown below.

K_d calculation example

$$\begin{aligned}
 K_d &\equiv \frac{1}{K_a} \\
 &= \frac{1}{1.73 \times 10^6} = 5.78 \times 10^{-7} \text{ M} \times 1000 = 5.78 \times 10^{-4} \text{ mM} \times 1000 \\
 &K_d = 0.6 \mu\text{M}
 \end{aligned}$$

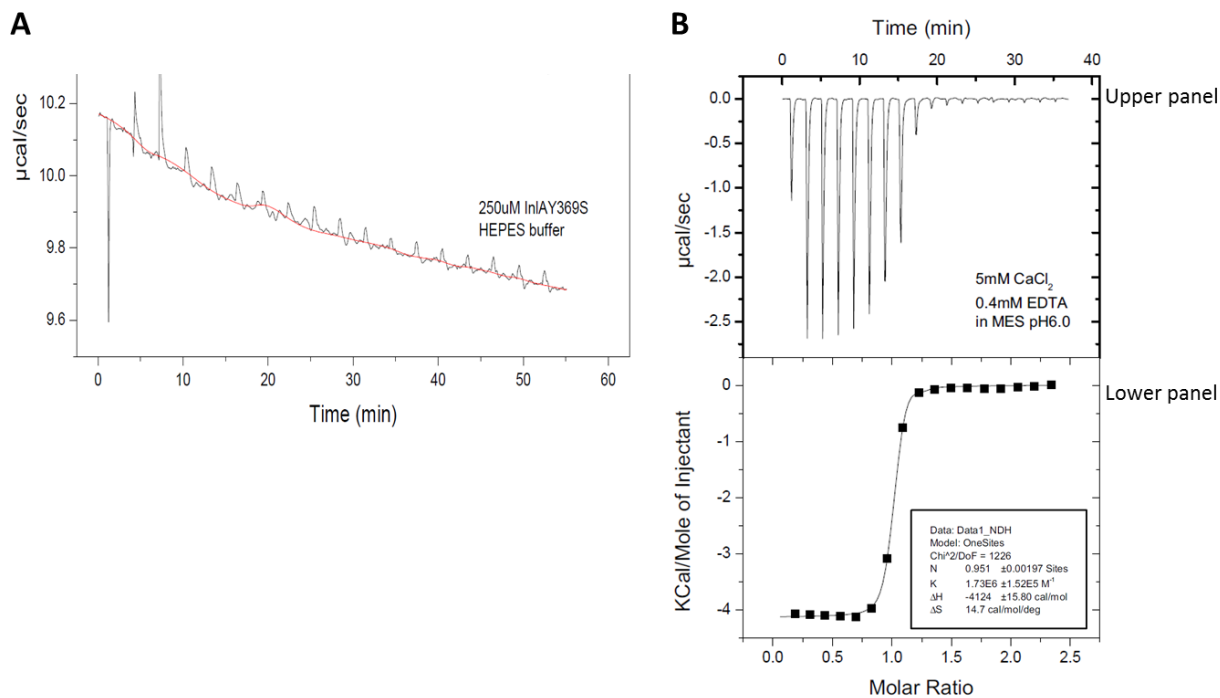


Figure 3.19: Examples of non-binding and binding ITC experiments. (A) Calometric titration of InLA^{Y369S} into sample cell containing HEPES buffer demonstrates the non-binding ITC experiment. The first three titrations resulted in heat-like signals, possibly as a result of air bubbles in one of the samples. (B) Calometric titration isotherms of the interaction between EDTA solution and CaCl₂ solution demonstrate the binding ITC experiment. The isotherms (upper panel) show that the interaction was occurring and this was measured on the axis of full scale deflection power in $\mu\text{cal/sec}$ vs Time. The lower panel shows the binding curve which was plotted on the plot of isotherms corresponding to each titration of ligand solution into target solution. The one site binding model plot resulted in the K_d of 0.6 μM . The other binding parameters are shown at bottom right of lower panel.

3.6 Interactions of InlA variants and E-cadherin domains (mEC1 and hEC1)

Since control experiments were demonstrated, it was now decided to carry out the interactions studies of InlA^{wt} and InlA^m proteins as ligand for hEC1 and mEC1 proteins, respectively. The ratio concentrations required for both ligand and target proteins were determined as shown in the calculation below, using InlA^m and mEC1 proteins as examples in a 10:1 ratio. The heat of dilution signals were detected during the experiment and repetition of this particular experiment resulted in similar results.

Since the K_d from two experiments was not determined as a result of heat of dilution signals, ratio concentrations were optimised by decreasing a concentration of the ligand. By doing this, the interaction between the proteins was characterised by heat signals resulting from an exothermic reaction. However, the improved ITC data was obtained when the ligand concentration was decreased to 250 μ M while that of the target protein remained as 100 μ M, making 2.5:1 ratio. Reproducible results were obtained when this ratio was used.

Ratio concentration calculation example

	InlA^m		mEC1
Protein concentration	→ 60 mg/mL		29 mg/mL
Protein MW in Da	→ 49500 Da		12000 Da
	$= 1.212 \times 10^{-3} \text{ M} \times 1000$		$= 2.417 \times 10^{-3} \text{ M} \times 1000$
	$= 1.212 \text{ mM} \times 1000$		$= 2.417 \text{ mM} \times 1000$
Final concentration	$= 1212 \mu\text{M}$		$= 2417 \mu\text{M}$

$$C_1 \times V_1 = C_2 \times V_2$$

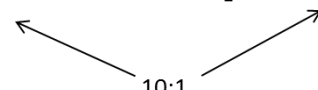
$$1212 \mu\text{M} \times V_1 = 1000 \mu\text{M} \times 220 \mu\text{L}$$

$$V_1 = 181.5 \mu\text{L} + 38.5 \mu\text{L of HEPES buffer}$$

$$C_1 \times V_1 = C_2 \times V_2$$

$$2417 \mu\text{M} \times V_1 = 100 \mu\text{M} \times 320 \mu\text{L}$$

$$V_1 = 13.2 \mu\text{L} + 306.7 \mu\text{L of HEPES buffer}$$



10:1

The ITC data obtained during the titration of InlA^m into sample cell containing mEC1 resulted in an exothermic reaction (**figure 3.20**). The isotherms indicating that the interaction between these proteins was occurring were placed in the upper panel. The plot of heat signals observed in the lower panel was generated by the Origin v7 software. The software was then used to correct the baseline and plot the binding curve (lower panel) with one site binding model. The best fit binding parameters as estimated by the software: n , K , ΔH and ΔS were shown inside top left box (lower panel). To obtain K_d for this binding, K value was converted in the same manner shown in the K_d calculation (section 3.5). The K_d determined for this reaction was 2 μM , a value that was lower but comparable to previous ITC findings of 10 μM (Wollert *et al.*, 2007).

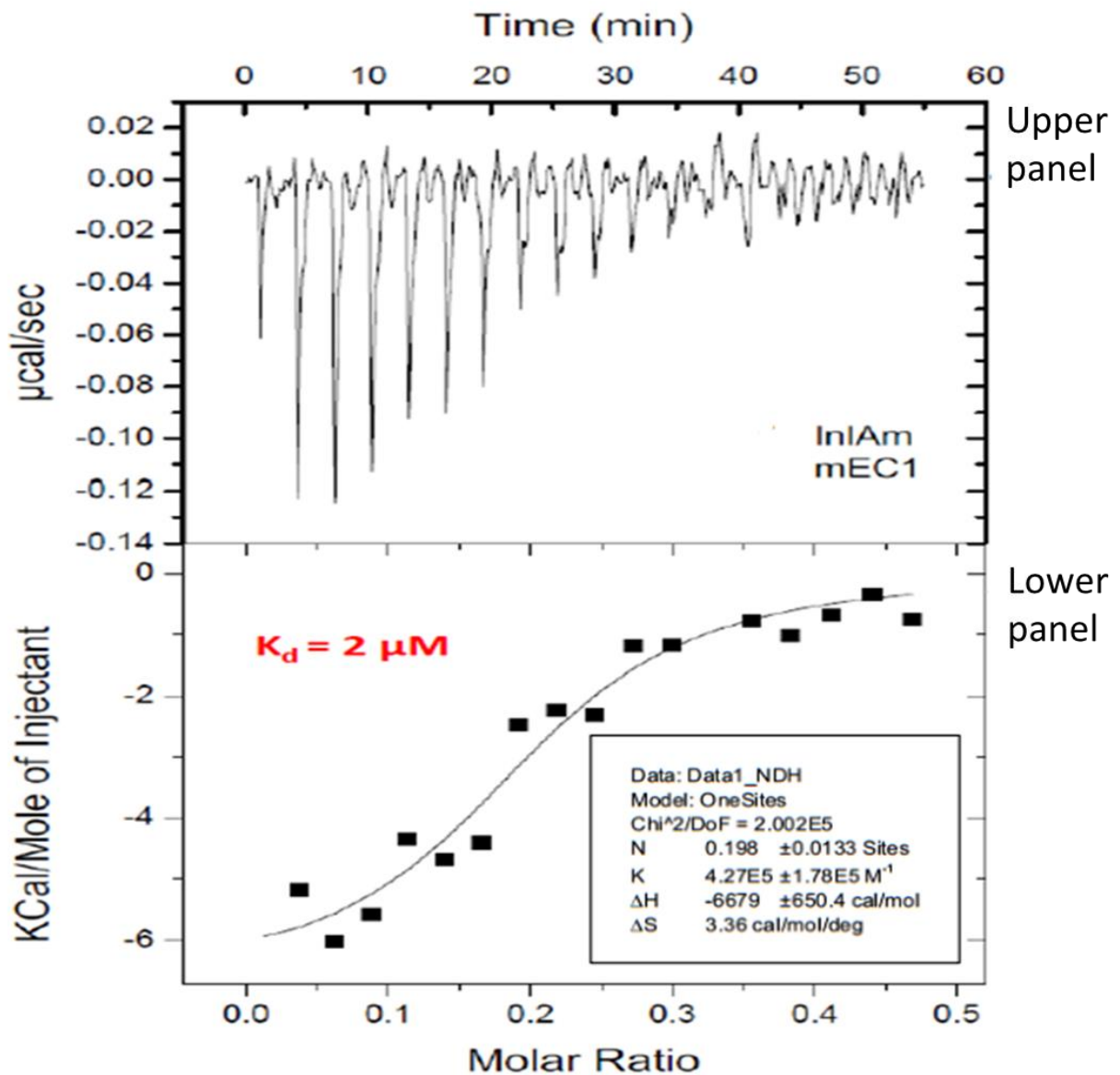


Figure 3.20: The calometric titration isotherm of the interaction between InIA^m and mEC1 proteins. The isotherms (upper panel) indicate that InIA^m and mEC1 proteins were interacting. The upper panel also indicates the full scale deflection power in $\mu\text{cal/sec}$ vs. Time. The isotherm data was used to plot the binding curve (lower panel) which was fitted to a one site binding model. The binding parameters are shown inside the bottom right box and the K_d was determined to be $2 \mu\text{M}$.

The interaction between InIA^{wt} and hEC1 proteins also resulted in an exothermic reaction (**figure 3.21**). The experiment and analyses were done similarly to **figure 3.20**. However, the K_d in this reaction was determined to be 4 μM , the binding affinity comparable to that analysed by analytical ultracentrifugation in previous study, of 8 μM (Schubert *et al.*, 2002).

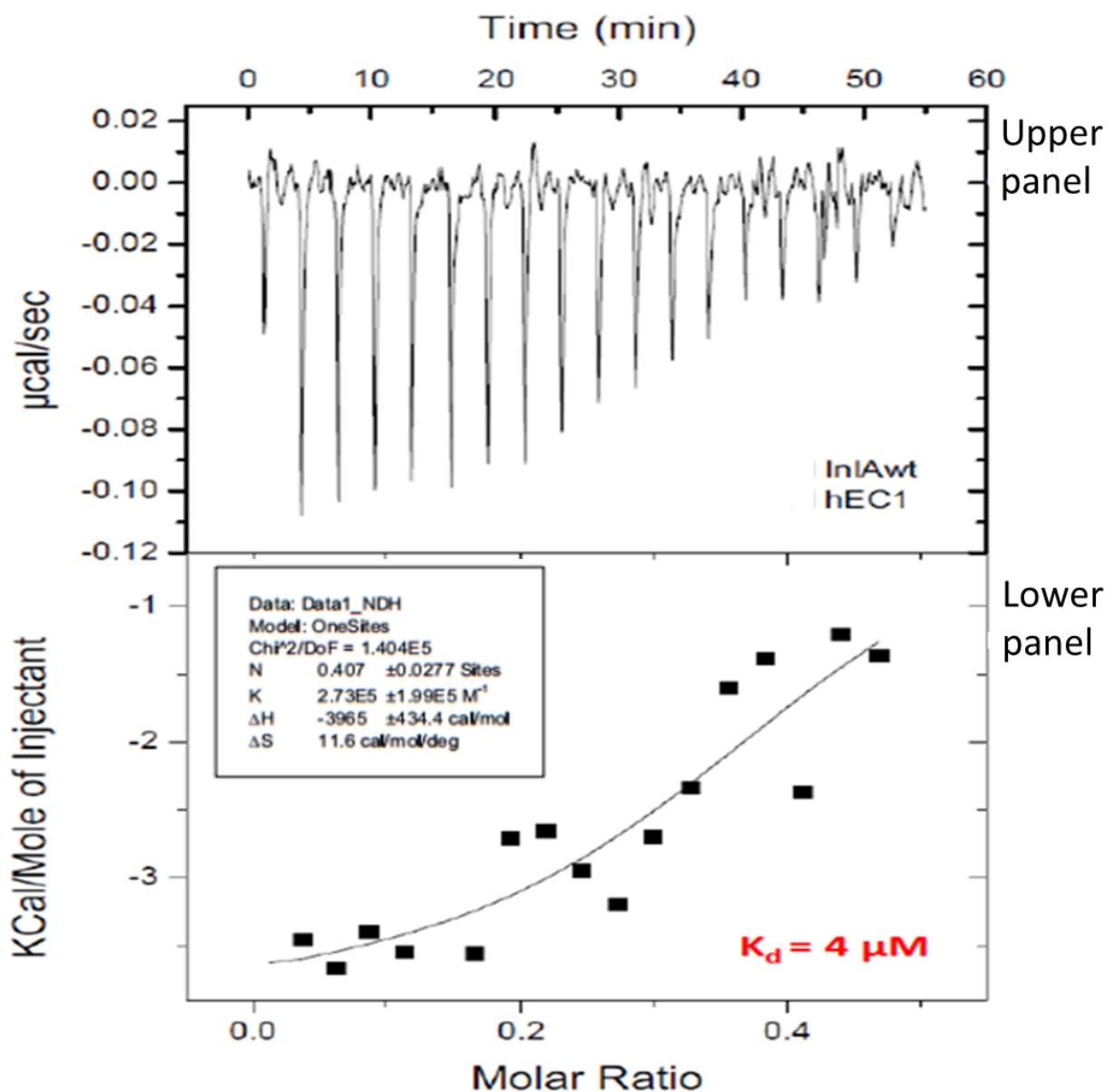


Figure 3.21: The calometric titration isotherm of the interaction between InIA^{wt} and hEC1 proteins. The isotherms (upper panel) indicate that InIA^{wt} and hEC1 proteins were interacting. The full scale deflection power was indicated as $\mu\text{cal/sec}$ vs Time (upper panel). The isotherms data was used to plot the binding curve (lower panel) fitted to a one site binding model. The K_d of 4 μM was determined and highlighted in red. The other binding parameters are shown inside the top left box.

3.7 Interaction of InlA^m and N-cadherin domains (mNC1 and hNC1 proteins)

The main focus of the current study was to investigate the interactions between InlA^m and N-terminal domains of N-cadherins (mNC1 and hNC1 proteins, respectively). Therefore, ITC experiments for already known protein-protein interactions (section 3.6) were done for two reasons: firstly, as positive controls to demonstrate the protein-protein interaction experiments, from which the binding affinities could be determined, secondly to compare these binding affinities to those of this section. The interactions under this section have not been reported before by any means of biophysical characterisation. What was only reported was the possibility that *Listeria monocytogenes* expressing *inlA^m* gene has additional pathogenesis as it targets M villous cells expressing murine N-cadherin (Tsai *et al.*, 2013). It was therefore proposed in this study that this interaction could be as a result of a direct interaction between InlA^m protein and the N-terminal domain of murine N-cadherin (mNC1 protein) which is conserved like mEC1 and hEC1 proteins (Mengaud *et al.*, 1996; Gallin, 1998).

It was then decided to investigate first the interaction between InlA^m and mNC1 proteins (**Figure 3.22, B**). The first ratio concentrations used (10:1) did not show any binding between the proteins, thus the experiment was optimised just like in section 3.6. However, the heat of dilution signals were observed when InlA^m protein was being titrated into sample cell containing mNC1 protein. After repetition of these experiments, the similar trend was observed and conclusions were made signifying that InlA^m does not recognise mNC1 protein as its binding partner.

The interaction between InlA^m and hNC1 proteins was done (**Figure 3.22, A**). By using both 10:1 and 2.5:1 ratios, observations indicated that InlA^m and hNC1 proteins were not interacting and what was observed was the heat of dilution signals observed in **Figure 3.22 (B)**. Therefore, the ITC data indicated that both mNC1 and hNC1 proteins have no affinity for InlA^m protein.

Since mNC1 and hNC1 proteins did not interact with InlA^m protein during ITC experiments, it was then decided not to use InlA single variants (InlA^{S192N} and InlA^{Y369S} proteins) to investigate their possibilities to interaction with these domains.

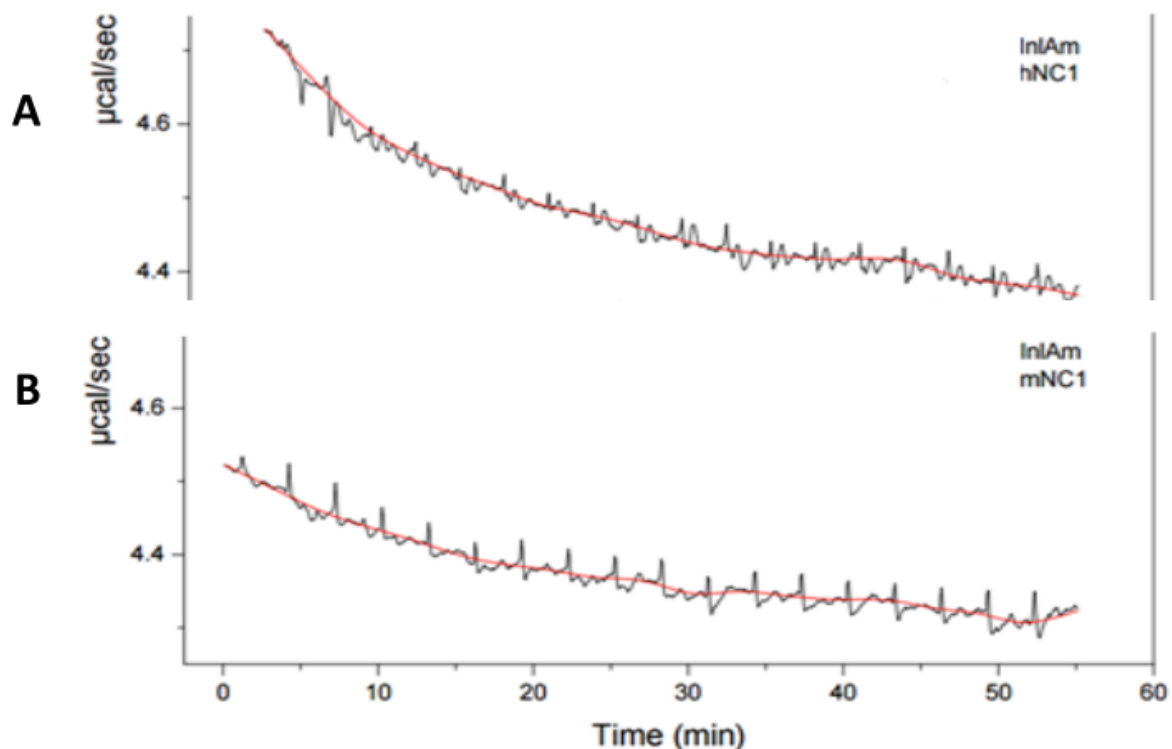


Figure 3.22: The heat of dilution signals formed during the calorimetric titration experiment of InlA^m and N-cadherin domains (mNC1 and hNC1). (A) The results show that hNC1 has no affinity for InlA^m protein. The first two injections seemed like the interaction was occurring but the rest of injections showed heat of dilution signals. (B) The ITC results showing that there was no interaction between InlA^m protein and mNC1 protein. The heats of dilution signals were detected during ligand titrations into target protein.

3.8 Conclusion

At the end of part 2, two key observations were made. The first observation was that already known protein-protein interactions binding affinities were successfully quantified just like in previous studies. The second observation was that InlA^m protein did not interact with both mNC1 and hNC1 proteins, respectively based on current findings. It was then hypothesised that these observations could be contributed possibly by unique structural elements of mNC1 and hNC1 proteins, respectively. Therefore it was decided to commence another study to analyse structural details of these domains in engineered complexes with InlA^{wt} and InlA^m.

Part 3: Homology modelling and structural analysis of complexes

3.9 Introduction

When experimental structures are not available, homology modelling of a particular protein provides three-dimensional (3D) structural information (Bishop *et al.*, 2008). The homology modelling predicts a structure for a protein of interest based on atomic coordinates of the template structure already existing in the protein data bank (PDB) (Krieger *et al.*, 2003; Bishop *et al.*, 2008). The accuracy of homology modelling relies heavily on amino acid sequence identity similarities of both the template and the target. Therefore the model structure with high sequence similarity to the template sequence provides reliable information (Krieger *et al.*, 2003). Phyre2 was used to model hNC1 structure for which experimental structure is not available. The SWISS-MODEL (www.swissmodel.org) was also used to compare the model structure of hNC1 to that obtained by Phyre2 and these models were similar.

3.10 Homology modelling of hNC1

The model structure of hNC1 was obtained using experimentally determined murine N-cadherin (PDB name: 3Q2W) as a template. The model report supported the multiple sequence alignment (MSA) results (**Appendix I**) by indicating that mNC1 and hNC1 proteins share 99% amino acid sequence similarity. The sequence coverage of mNC1 used to model hNC1 was 99%. Moreover, the report suggested that hNC1 was modelled to 99.9% confidence level, which indicated accuracy of the model. It was observed that both domains were highly conserved between humans (**Figure 3.23, A**) and mice (**Figure 3.23, B**). The proteins adopted similar fold and all structural elements were conserved despite a single amino acid variation between the two. This included the unique 3_{10} helices between β -strands b and c.

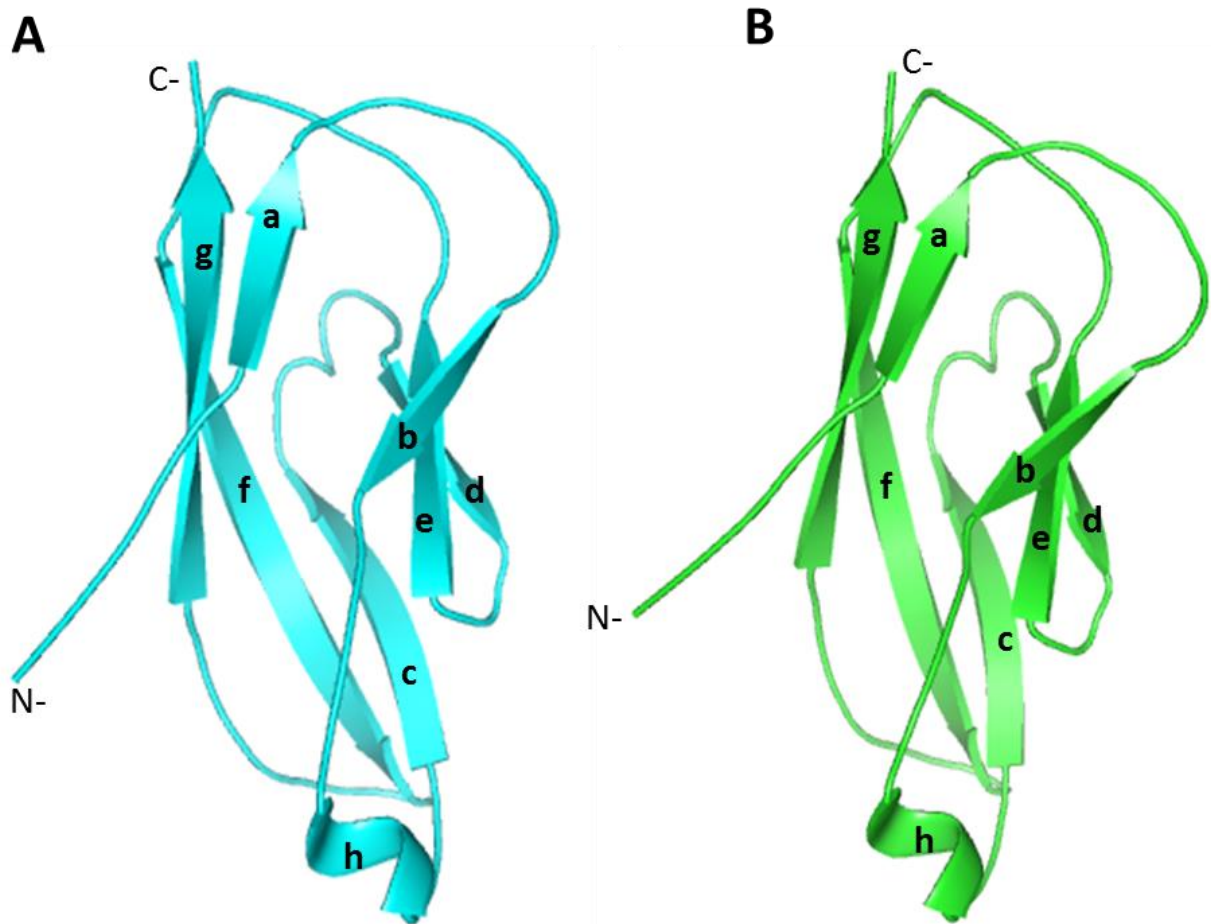


Figure 3.23: The model and crystal structures of hNC1 and mNC1 proteins. (A) The structural prediction of hNC1 (cyan) indicates that the protein has seven β -strands labelled a-g and 3_{10} helix. (B) The crystal structure of mNC1 (green) has similar structural elements to those of hNC1 model from which both 3_{10} helices are located between β -strand b and c.

3.11 Comparisons of InIA^{wt}/hEC1 and InIA^m/mEC1 complexes

The crystal structures of InIA^{wt}/hEC1 and InIA^m/mEC1 complexes were superimposed to illustrate the interaction interfaces, so that the information can be compared to engineered structural complexes involving mNC1 and hNC1. By superimposing these complexes (**Figure 3.24, A**), it was observed that the single mutations in InIA did not affect the overall structural orientation or geometry. The structural elements of mEC1 and hEC1 were perfectly aligned, similarly occupying and filling the concave central cavities of LRR domains of both InIA variants. These concave structures of LRR domains were made up by 15 strands of parallel β -sheets of both InIA variants (also numbered). The atomic changes and interaction

interfaces were analysed near the InlA mutations: S192N and Y369S, respectively. The specific information required with these interaction interfaces was to show whether these mutations provide direct hydrogen or water mediated bonds between InlA^m and both mEC1 and hEC1 interacting residues, respectively.

The interaction interface involving InlA (S192N) mutation and the residues on the loops of mEC1 and hEC1 was demonstrated in **Figure 3.24 (B)**. The direct hydrogen bond formed between Asn192 in InlA^m and main chain carbonyl of Phe17 in both mEC1 and hEC1. However, the distances between these residues slightly varied; Asn192 in InlA^m and Phe17 in hEC1 had a distance of 2.8 Å, while corresponding distance between Phe17 in mEC1 was 2.9 Å. This could attest to the improved binding affinity between InlA^{S192N}/hEC1 compared to InlA^{wt}/hEC1 by ITC (Wollert *et al.*, 2007). The second interaction interface involving InlA (Y369S) mutation and Asn27 in both mEC1 and hEC1 was shown in **Figure 3.24 (C)**. The water-mediated hydrogen bonds in this interface have been reported (Wollert *et al.*, 2007). Here, the bonds between these residues were not shown but rather measured as distances. The distance between Ser369 in InlA^m and Asn27 in mEC1 was 4.3 Å, while between the corresponding Asn27 in hEC1 was 4.5 Å.

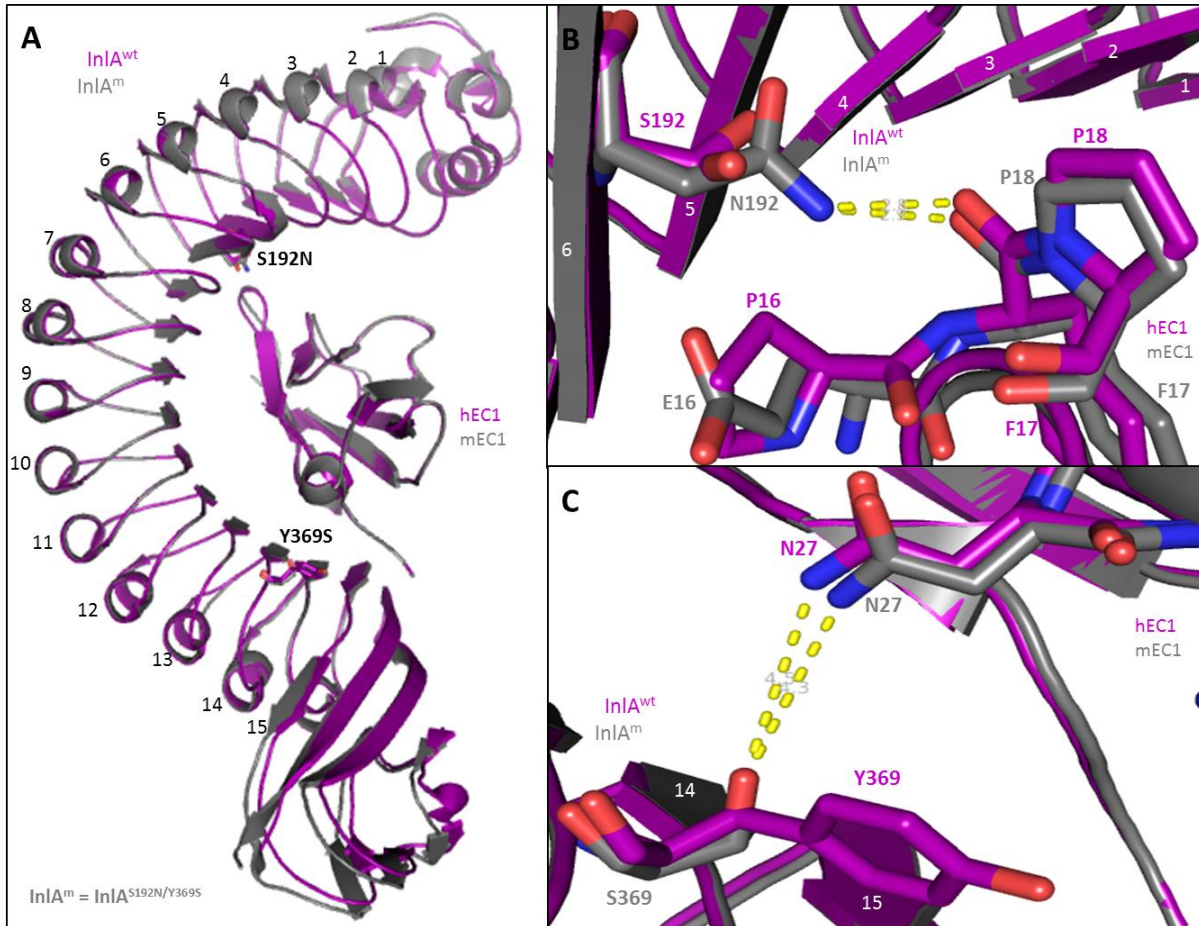


Figure 3.24: Comparison of InLAtm/mEC1 and InLA^{wt}/hEC1 superimposed complexes. (A) The superimposition of the complexes shows each complex with its colour. The LRRs are numbered and mutated residues in InLA are shown as sticks. (B) The interaction interface involving Asn192 in InLAtm shows that this residue result in a direct hydrogen bond between its side chain and carboxyl groups of Phe17 in mEC1 and hEC1, respectively. (C) The interaction interface involving Ser369 in InLAtm showing water-mediated interactions with Asn27 of mEC1 and hEC1, respectively.

3.12 Re-engineering InlA^m/mNC1 complex

The mNC1 and hEC1 were superimposed while the latter was in complex with InlA variants (**Figure 3.25, A**). The hEC1 and mNC1 were observed to share similar fold and all structural elements. However, these elements were not perfectly aligned and this was thought to be as a result of amino acid sequence variations between the domains (Mengaud *et al.*, 1996; Gallin, 1998). Nonetheless, both domains fitted in the central cavity within LRR domains of InlA variants. The conformational change to hEC1 β -strand a, that is normally induced by InlA^{wt} (Schubert *et al.*, 2002) was also observed in mNC1.

Figure 3.25 (B) demonstrated the interaction interface involving InlA (S192N) mutation site and residues from hEC1 and mNC1 proteins. The three amino acids (Pro16, Phe17 and Pro18) from hEC1 and mNC1 at this interface are similar but were not perfectly aligned. However, it was observed that Asn192 in InlA^m created direct hydrogen bonds to the main chain carbonyls of Phe17 in both hEC1 and mNC1. The distances measured between these interacting residues were equally 2.8 Å. This suggested that despite imperfect residues alignment between hEC1 and mNC1, the strength of the bonds at this interface was equivalent. It was then proposed that the water-mediated interactions between Ser192 in InlA^{wt} and Pro16 in hEC1 (Schubert *et al.*, 2002) would also be ubiquitous in the case of mNC1. This assumption was made based on the importance of Pro16 ensuring that InlA/hEC1 interactions were occurring (Lecuit *et al.*, 1999) and this residue was also present in mNC1.

The interaction interface involving the second mutation (Y369S) on InlA^m and these domains was shown in **Figure 3.25, (C)**. The residues 27 of mNC1 and hEC1 had their back-bones and side chains up to the methylene carbon perfectly aligned. However, the carbamoyl group of Asn27 in mEC1 and carboxyl group of Asp27 in mNC1 faced in opposite directions. Furthermore, the Asn27 twisted around its side chain amino group from its original position that was observed in **Figure 3.24 (C)**. Therefore, it was proposed that this arrangement was possibly influenced by the presence of mNC1. This is evidenced by the dramatic change in distance measured between Ser369 in InlA^m and Asn27 in hEC1. It was also clear that Asp27 in mNC1 would have formed a bond via hydroxyl group of Ser369 in InlA^m. These observations to a large extent contradicted with ITC results which recorded that InlA^m does not interact with mNC1.

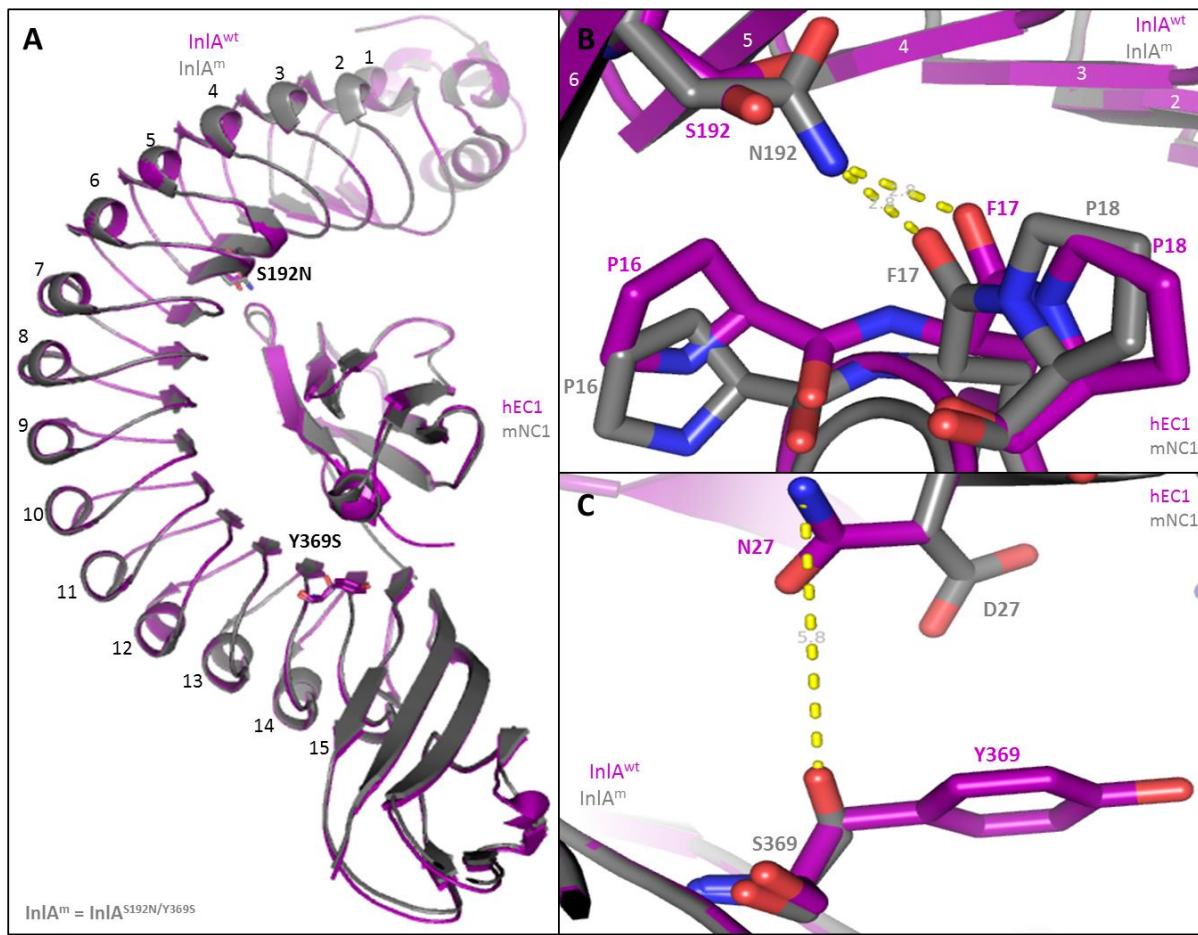


Figure 3.25: Engineered complexes of InA^m/mNC1 and InA^{wt}/hEC1. (A) The superimposed complexes are shown with their colour. (B) The interaction interface involving Asn192 in InA^m shows that this residue results in a direct hydrogen bond between its side chain and carboxyl groups of Phe17 in mNC1 and hEC1. (C) The interaction interface involving Ser369 in InA^m and Asn27 in hEC1. The water-mediated hydrogen bond distance is different from the previous complex.

3.13 Re-engineering InA^m/hNC1 complex

It was then decided to superimpose hNC1 model structure to the hEC1 structure while the latter was in complex with InA variants (**Figure 3.26, A**). The structural elements of both hEC1 and hNC1 were perfectly aligned with noticeable variations on their 3₁₀ helices. Overall, the superimposed image was similar to **Figure 3.25 (A)**. The interaction interface observations in **Figure 3.26 (B)** were similar to those made in **Figure 3.25 (B)**, of which the direct hydrogen bond distances between interacting residues were 2.8 Å, respectively. The interaction interface involving Ser369 in InA^m and corresponding residues in hEC1 and hNC1 (**Figure 3.26, C**) was exactly the same as that which was observed in **Figure 3.25 (C)**.

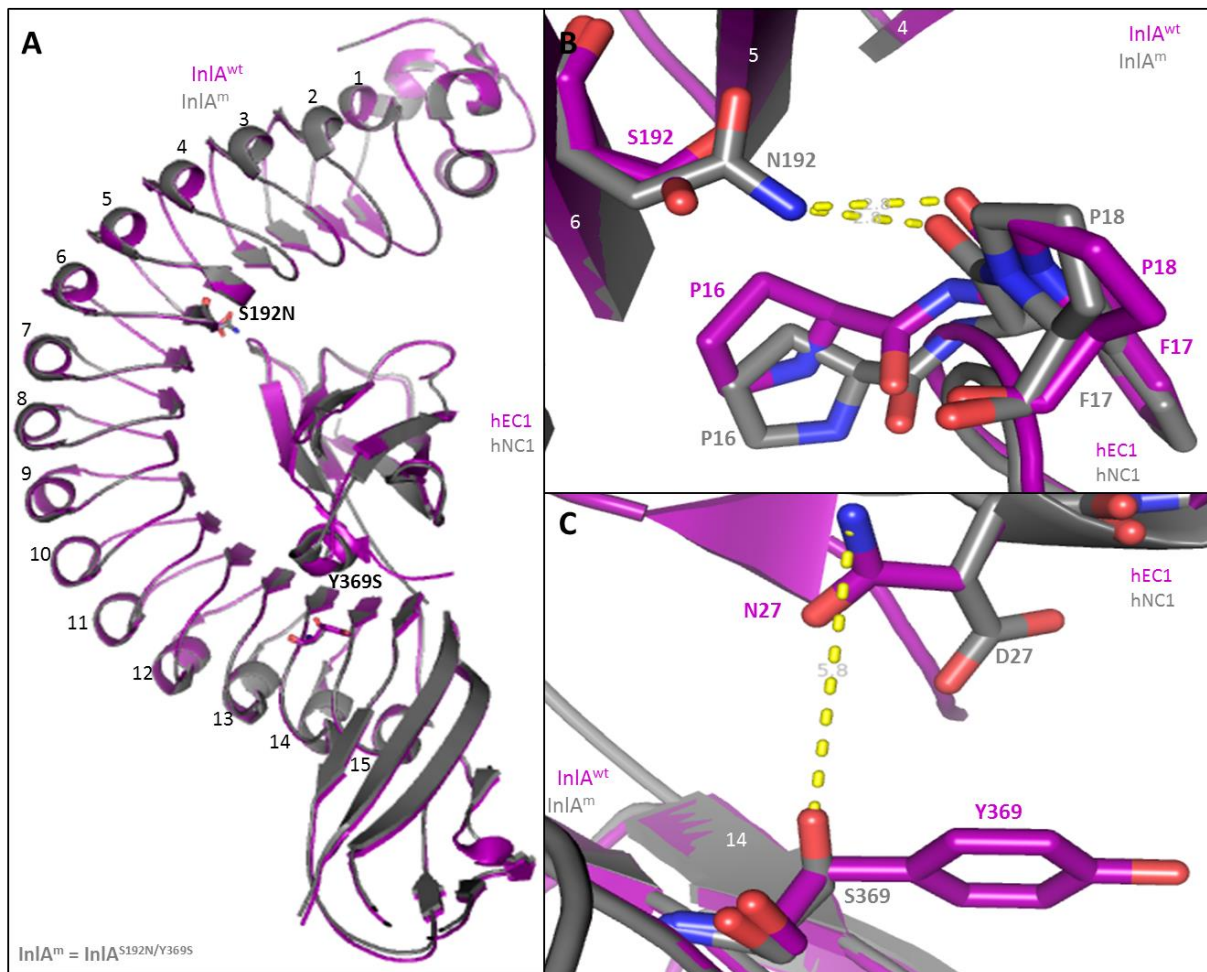


Figure 3.26: Engineered complexes of InlA^m/hNC1 and InlA^{wt}/hEC1. (A) The superposition of the complexes is shown with its colour. (B) The interaction interface involving Asn192 in InlA^m is shown. (C) The interaction interface involving Ser369 in InlA^m and Asn27 in hEC1 with water-mediated hydrogen bond.

3.14 Conclusion

It was observed that E- and N-cadherin N-terminal domains share high degree of structural similarities. Furthermore, N-cadherin domains fit perfectly into the concave central cavity of InlA variants in the superimposed complexes just like hEC1. However, amino acid sequence variations between E- and N-cadherin domains significantly contributed to conformational changes observed in hEC1 in the interaction interface involving Ser369 in InlA^m. All interaction interfaces strongly suggested that N-cadherin domains could be similarly recognised by InlA^m as receptors just like E-cadherin domains.

Chapter 4.0: Discussion

4.1 Protein production and purification of E- and N-cadherins

The purpose of cloning mNC1 and hNC1 in pGEX-6P-2 vector was to produce corresponding glutathione S-transferase (GST) tagged proteins like mEC1 and hEC1, which were already cloned in pGEX-6P-1 vector. The GST moiety is located at the N-terminus of the protein of interest, acting as a chaperone to ensure protein folding (Harper and Speicher, 2011) and consequently resulting in soluble protein (Einarson *et al.*, 2005). All proteins in this study were produced as highly soluble GST fusion proteins even though some exhibited partial insolubility. The GST fusion proteins are normally captured by immobilised Glutathione Sepharose (GS) beads (Einarson *et al.*, 2005; Harper and Speicher, 2011) and this allows unbound proteins to be washed out. The 3C protease cleaves between Gln and Gly amino acids of GST C-terminal sequence (Leu-Phe-Gln/Gly-Pro) to separate GST from target protein. This cleavage allows the target protein to be eluted easily while the GST remains on the GS beads (Harper and Speicher, 2011). However, it was observed that GST and fusion proteins eluted together with target proteins. It was proposed that this could have been as a result of GS beads leakage, possibly caused during mixing of 3C protein and GS beads before cleavage incubation.

During anion exchange chromatography (AEC), weakly bound proteins were expected to elute first at low salt concentrations while strongly bound proteins elute lastly at high salt concentrations (Wei *et al.*, 1999). Unexpectedly, mNC1 and hNC1 eluted in the void volume before linear salt gradient. It was then assumed that perhaps there were going to be other peaks representing these proteins together with impurities during salt gradient only if the experiment was not terminated prematurely. This is because GST was determined to have lower theoretical isoelectric point (pI), meaning that it required high salt concentration to elute. We therefore named this type of purification; the negative AEC, to imply the case whereby impurities bind to the column while target protein elutes in the void volume. The mEC1 and hEC1 eluted at equivalent salt gradient and these proteins have been previously purified and reported to be highly stable in high salt buffers (Schubert *et al.*, 2002; Wollert *et al.*, 2007).

4.2 Site directed mutagenesis on InlA functional gene

Site directed mutagenesis (SDM) is a strategy used to introduce mutations on the gene while cloned in the vector (Liu and Naismith, 2008; Li *et al.*, 2008). This strategy is easily achieved, but depend on the number of nucleotides that are needed to be mutated (Li *et al.*, 2008; Munteanu and Braun, 2012). The strategy to create InlA^m was successfully achieved because only three nucleotides, giving rise to two different amino acids at two specific positions on *inlA* gene were replaced –one experiment at the time. A conventional PCR contributed significantly towards achieving this aim, with primers allowing the amplification of entire vector while replacing desired nucleotides. This strategy was previously used to create the similar InlA variant. Moreover, *Listeria monocytogenes* was genetically mutated to express the gene that encodes for InlA^m protein, allowing passage of *Listeria monocytogenes* (*Lm*) via mEC1-dependent pathway (Wollert *et al.*, 2007) and later mNC1-dependent pathway (Tsai *et al.*, 2013).

The justification behind creating these mutations at these specific positions on amino acids sequence was due to side chain length of Ser192 and unspecific binding properties of Tyr369. In their explanation, Wollert et al (2007) stated that Ser192 in InlA^{wt} has short side chain which could not form a direct hydrogen bond with Pro16 in hEC1 protein. Thus replacing it by Asn could result in a direct hydrogen bonding to Phe17 in hEC1. On the other hand, the bulky Tyr369 in InlA^{wt} created unfavourable interaction with Asn27 in hEC1 and was replaced by hydrophilic amino acid serine, thus a water-mediated hydrogen bond formed between these residues. These two mutations were reported to increase binding affinity between these proteins by 6700-fold more than the wild type interaction and in addition InlA^m protein interacted with mEC1 protein.

4.3 Production and purification of InlA variants

InlA^{wt} was produced as soluble protein at 37°C by Mengaud et al., (1996). In this study, bacterial cells were successfully induced at 28°C for all InlA variants production. Therefore, the proteins were soluble and stable when kept in high salt buffers, similarly to storage conditions in previous studies (Schubert *et al.*, 2002; Wollert *et al.*, 2007). The proteins were purified by GS affinity chromatography, but elution fractions were contaminated with fusion proteins and GST. Since this was the similar trend observed during GS affinity

chromatography of E- and N-cadherin N-terminal domain proteins, further purification was required.

The AEC has been used previously to purify other *Lm* proteins such as ActA (Zalevsky *et al.*, 2001), InlB (Reinl *et al.*, 2009) and InlC (Polle *et al.*, 2014), respectively. During AEC of InlA variants, the peaks came out immediately after each other with the exception of InlA^{wt} purification. This could be explained by the fact that GST and InlA variants have almost similar theoretical pIs of 4.91 and 4.45, respectively. The fractions under these peaks were analysed on SDS-PAGE and confirmed as InlA variants, fusion proteins and GST. On the other hand, the introduction of unknown proteins could be as a result of inadequate cleaning of the MonoQ column after previous experiments.

To eliminate the GST and fusion proteins from target proteins, GS affinity chromatography is done to allow the proteins to bind to GS beads again (Herper and Speicher, 2011). After adding protein solutions to GS beads, target proteins were successfully eluted without traces of GST, GST-fusion and surprisingly additional proteins introduced during AEC were not observed. Alternatively, AEC purification could have been repeated, but the process was thought to be time consuming. The GS affinity chromatography on other hand is a quick and effective process, in which target proteins are simply obtained from the flow through.

4.4 Biophysical characterisation of proteins by isothermal titration calorimetry

Isothermal titration calorimetry (ITC) experiment requires that all proteins be pure, highly concentrated (Doyle, 1997) and in exactly the same buffer (Leavitt and Freire, 2001). Failure to match the buffers may result in heat of dilution effects (Duff *et al.*, 2011) and this could lead to inaccurate analyses regarding the interactions between the proteins. In this study, all proteins were in HEPES buffer containing CaCl₂ just like in previous studies (Schubert *et al.*, 2002; Wollert *et al.*, 2007). The calcium ions play coordination roles between EC domains interfaces of both E- and N-cadherins (Jin *et al.*, 2012). Furthermore, a buffer containing calcium ions has been proven to bring forth favourable conditions for protein-protein interactions (Schubert *et al.*, 2002).

The previously investigated interactions of InlA^{wt}/hEC1 (Schubert *et al.*, 2002) and InlA^m/mEC1 (Wollert *et al.*, 2007) complexes were repeated in this study. This was done to

ensure consistency of the ITC statistics, so that reliable conclusions may be drawn when analysing other complexes. Therefore, the previously analysed complexes were again proved to form by ITC, with slightly different, but yet comparable binding affinities to those reported. The current study recorded the dissociation constant for InLA^{wt}/hEC1 complex to be 4 μM , comparable to 8 μM (Schubert *et al.*, 2002). In another experiment where InLA^m was in complex with mEC1, dissociation constant recorded was 2 μM , a decreased value from 10 μM (Wollert *et al.*, 2007). We propose that these decreases in K_d could be influenced by technology advancement of ITC equipment and much improved analyses software that comes with it. However, this is unlikely explanation and other possibilities were pursued.

The differences in binding affinities could have been influenced by factors involving protein-protein ratios used. In previous studies, the ratio of 10:1 (e.g. 800 μM ligand: 80 μM target protein) was used. In this study, the same ratio was used but no protein-protein interactions were recorded. The optimisations that involved increasing ratio concentrations were made (1000 μM ligand: 100 μM target protein) and gave promising results. However, lowering the concentration of ligand to 250 μM provided better results. The ratio that was finally used was 2.5:1 (250 μM ligand: 100 μM target protein). The differences in K_d did not disqualify the current results but dictated that much improved binding affinities of the complexes were recorded for the first time.

Listeria monocytogenes mutant that expresses *inlA^m* gene was reported to be internalised into villous M cells that express accessible murine N-cadherin (mNcad) protein (Tsai *et al.*, 2013). According to our knowledge, the InLA^m protein has never been biophysically characterised to interact with mNC1 or hNC1 proteins. Therefore, the main focus of this study was to investigate whether these interactions occur and to compare their binding affinities to those studied previously (Wollert *et al.*, 2007). However, ITC results suggested that InLA^m does not recognise both mNC1 and hNC1 proteins as receptors. This was supported by heat of dilution effects similarly to when InLA^{Y369S} variant was titrated into HEPES buffer. Heat of dilution effects in principle may suggest possible buffer mismatch, but since all proteins were in similar buffer, the only assumption that could be made at this point was that these proteins have no binding affinities for each other.

To our knowledge, none of the two proposed receptors for InLA^m have been reported to interact with any particular protein by ITC. The information in literature about mNC1 and hNC1 was not useful in supplying possible judgement behind these findings. At this stage, we

hypothesised that dimerization might be the possible explanation. However, this suggestion was not substantial because size exclusion chromatography done during buffer exchange experiments did not show any possible dimerization of these proteins. A better explanation was required to make conclusions as to why these domains were not recognised by InLA^m like mEC1 and hEC1 even though they are highly conserved structurally (Mengaud *et al.*, 1996).

Despite the structural similarities between E-cadherin domains (Mengaud *et al.*, 1996), one amino acid variation on position 16 between mEC1 and hEC1 was enough to disallow the former to be recognised by InLA^{wt} (Schubert *et al.*, 2002). But mEC1 was similarly recognised by InLA^m just like hEC1 (Wollert *et al.*, 2007). The crystal structure of mNcad, especially the N-terminal domain (mNC1) is highly similar to that of mEC1 and hEC1. However, there is a degree of amino acid sequence variations between these cadherin domains. Thus, failure for InLA^m to recognise mNC1 and hNC1 could have been as a result of these variations since a single mutation in the amino acid sequence affect the functional properties of a particular protein (Berrondo, 2010). Therefore, imperative structural information about mNC1 and hNC1 in comparisons with either hEC1 or mEC1 was required.

4.5 Engineered complexes and the insights from interaction interfaces

4.5.1 Modelling and overall structural analyses of the complexes

The mNC1 and hNC1 share exactly the same structural fold, with all structural elements perfectly matching between the two. This was anticipated as there is only one amino acid variation between the two. However, this did not translate to similar functional properties of both proteins provided that a single amino acid variation may alter the function of any of them (Berrondo, 2010). The highly conserved structural fold between E- and N-cadherin domains (Mengaud *et al.*, 1996) could be as a result of common ancestral genome (Gallin, 1998). The E- and N-cadherin domains have 3₁₀ helices made up of three residues differed by a single amino acid between them. The side chains of these residues in 3₁₀ helices are normally aligned and interact via hydrogen bonds to form a right-handed helical structure (Schwaiger *et al.*, 2011). The similar structural elements detected between these domains were considered not enough to suggest whether the N-cadherin domains may or may not be recognised by InLA^{wt} or InLA^m similarly to E-cadherin domains.

The information obtained from superimposed crystal complexes of InlA^{wt}/hEC1 and InlA^m/hEC1 has been extensively explained previously (Schubert *et al.*, 2002; Wollert *et al.*, 2007). Concisely, the InlA^{wt} has been reported to induce the first β -strand's (β -strand a) conformational change in hEC1 (Schubert *et al.*, 2002) and this was observed in similar cases involving InlA^m protein (Wollert *et al.*, 2007). We wanted to highlight any possible conformational changes that can be induced by InlA^{wt} or InlA^m on mNC1 and hNC1 structures. The slight conformational changes involving the same β -strand in mNC1 and hNC1 were observed. This was minor because superimposed domains within the central cavities of LRR domains had slightly poorer alignment of structural elements –possibly due to amino acid sequences variations between the cadherin family members. However, the overall orientations of complexes whereby mNC1 or hNC1 were present were similar to the original complexes.

4.5.2 Analysing the complexes' interaction interfaces

The residues on position 16 in mEC1 and hEC1 from InlA^m/mEC1 and InlA^{wt}/hEC1 complexes adopted different conformations. Pro16 in hEC1 adopted a strained *cis* conformation, positioning its side chain near the LRR domain repeat 6 β -sheet. Glu16 in mEC1, by contrast, adopted a relaxed *trans* conformation by orientating its backbone in a manner that allowed the side chain to be aligned to the aromatic group of Pro16 in hEC1. Similar conformations have been observed in all superimposed complexes involving mNC1 and hNC1, but these domains were not perfectly aligned with hEC1. The interface involving Pro16 of hEC1 and Glu16 of mEC1 was shown as hydrophobic binding pocket in InlA^{wt}/InlA^m superimposed structures (Wollert *et al.*, 2007). We propose that structural alignment forced atomic coordinates of the mNC1 and hNC1 to adopt similar conformations as those of hEC1 domain. However, variations in amino acids sequences between N-cadherin domains and hEC1 could have affected secondary elements to align perfectly. Moreover, the complexes suggested that there are possible interactions between InlA^m and N-cadherin domains based on possible bonds that may form.

The water-mediated hydrogen bonds forming between Ser369 in InlA^m and residues in position 27 of mEC1 and hEC1 varied. This could be as a result of variations in neighbouring residues between these domains. Neighbouring amino acids tend to drive orientations of other residues, thus this could either affect or contribute to stronger binding between interacting proteins (Schubert *et al.*, 2002). Therefore, this correlated to the fact that InlA^m has strong

binding affinity for hEC1 than mEC1 (Wollert *et al.*, 2007). This is regardless of the two E-cadherin domains having equivalent distances of direct hydrogen bonds with InLA^m protein. The changes in water-mediated distances observed when mNC1 or hNC1 were superimposed to hEC1 could have been due to alignment constraints between different domains. The N-cadherin domains might have affected the orientation of hEC1 atoms at this interface, such that the former domains can be accommodated in the interface. Nonetheless, the N-cadherin domains were seemingly having close contacts with InLA^m just like hEC1.

There are other factors that were reported to contribute to the stronger binding of InLA^{wt} to hEC1. These included various individual residues found in both hEC1 and InLA^{wt} (Schubert *et al.*, 2002), and they were found to be common during mEC1 and InLA^m interactions (Wollert *et al.*, 2007). The individual residues were analysed on mNC1 and hNC1 that could have influenced their interaction with InLA^m. Concisely, the surface exposed aromatic amino acids of InLA^m were observed to be directly interacting with various residues in mNC1 and hNC1 in the engineered complexes.

Ions at the interaction interfaces have been reported to positively regulate complex formation between InLA^{wt} and hEC1 (Schubert *et al.*, 2002). For example, the Ca²⁺ ions coordinate InLA^{wt}/hEC1 complex formation via five water molecules and Glu326 in InLA^{wt} with average distances of 2.45 Å. Moreover, the Cl⁻ was also found to form a salt bridge to hEC1, thus bridging both InLA^{wt} and hEC1 together (Schubert *et al.*, 2002). Therefore the octahedral coordination by Ca²⁺ and bridging by Cl⁻ could have uniquely contributed to the increase in the binding affinity between InLA^m and hEC1 (Wollert *et al.*, 2007). These contributing factors were also observed and could possibly facilitate the interactions between InLA^m and N-cadherin domains. Moreover, these observations meant that InLA^m has additional factors that could enable it to recognise N-cadherin domains besides the mutation sites interfaces. Therefore, our findings suggested that both mNC1 and hNC1 could be receptors of InLA^m and possibly InLA^{wt} as they shared binding interface coordinates with hEC1.

4.6 Conclusion and outlook

This study investigated whether the InlA variant designed to increase its binding affinity to hEC1 had additional receptors apart from mEC1 (Wollert *et al.*, 2007). The InlA^m/mNC1 or InlA^m/hNC1 complexes formation could provide exciting avenues towards understanding *Listerial* infection via these pathways and subsequent listeriosis in cells expressing murine or human N-cadherin.

This work reported that InlA^m does not recognise mNC1 or hNC1 as its potential receptors by ITC. Therefore, the justifications were that amino acid variations between E- and N-cadherin were driving forces behind these observations. However, engineered structural complexes suggested contradicting results from ITC, indicating that InlA^m can possibly interact with both mNC1 and hNC1. Therefore, the two findings correlated to both hypotheses made in this study.

We therefore propose two studies which would render conclusive arguments adding to the current findings. Firstly, biophysical characterisation should be repeated by using other techniques such as Surface Plasmon Resonance spectroscopy. Secondly, the crystallisation experiments of the complexes may also contribute in provision of conclusive information about whether the InlA^m form complexes with mNC1 and hNC1, respectively.

References

- Alberti-Segui, C., Goeden, K.R., and Higgins, D.E. (2007) "Differential function of *Listeria monocytogenes* listeriolysin O and phospholipases C in vacuolar dissolution following cell-to-cell spread", *Cell Microbiol*, **9**: 179–195.
- Amsellem, V., Dryden, N.H., Martinelli, R., Gavins, F., Almagro, L.O., Birdsey, G.M., Haskard, D.O., Mason, J.C., Turowski, P., and Randi, A.M. (2014) "ICAM-2 regulates vascular permeability and N-cadherin localization through ezrin-radixin-moesin (ERM) proteins and Rac-1 signalling", *Cell Commun Signal*, **12**: 1–14.
- Bakardjiev, A.I., Stacy, B.A., Fisher, S.J., and Portnoy, D.A. (2004) "Listeriosis in the Pregnant Guinea Pig: A Model of Vertical Transmission", *Infect Immun*, **72**: 489–497.
- Barbau-Piednoir, E., Botteldoorn, N., Yde, M., Mahillon, J., and Roosens, N.H. (2013) "Development and validation of qualitative SYBR Green real-time PCR for detection and discrimination of *Listeria* spp. and *Listeria monocytogenes*", *Appl Microbiol Biotechnol*, **97**: 4021–4037.
- Bergmann, S., Beard, P.M., Pasche, B., Lienenklaus, S., Weiss, S., Gahan, C.G.M., Schughart, K., and Lengeling, A. (2013) "Influence of internalin A murinisation on host resistance to orally acquired listeriosis in mice", *BMC Microbiol*, **13**: 1–16.
- Berrondo, M. (2010) "Predicting the Structure and Function of Protein Mutants", Thesis Manuscript, 1-162.
- Bierne, H., and Cossart, P. (2002) "InlB, a surface protein of *Listeria monocytogenes* that behaves as an invasin and a growth factor", *J Cell Sci*, **115**: 3357–3367.
- Bischofberger, M., Iacovache, I., and Gisou Van Der Goot, F. (2012) "Pathogenic pore-forming proteins: Function and host response", *Cell Host Microbe*, **12**: 266–275.
- Bishop, A.O.T., de Beer, T.A.P., and Joubert, F. (2008) "Protein homology modelling and its use in South Africa", *S Afr J Sci*, **104**: 2–6.
- Bonazzi, M., Lecuit, M., and Cossart, P. (2009a) "*Listeria monocytogenes* internalin and E-cadherin: from bench to bedside", *Cold Spring Harb Perspect Biol*, **1**: 1–15.
- Bonazzi, M., Lecuit, M., and Cossart, P. (2009b) "Microreview *Listeria monocytogenes* internalin and E-cadherin: from structure to pathogenesis", *Cell Microbiol*, **11**: 693–702.
- Bonazzi, M., Veiga, E., Pizarro-Cerdá, J., and Cossart, P. (2008) "Successive post-translational modifications of E-cadherin are required for InlA-mediated internalization of *Listeria monocytogenes*", *Cell Microbiol*, **10**: 2208–2222.
- Cabanes, D., Dehoux, P., Dussurget, O., Frangeul, L., and Cossart, P. (2002) "Surface proteins and the pathogenic potential of *Listeria monocytogenes*", *Trends Microbiol*, **10**: 238–245.

- Camejo, A., Carvalho, F., Reis, O., Leitão, E., Sousa, S., and Cabanes, D. (2011) "The arsenal of virulence factors deployed by *Listeria monocytogenes* to promote its cell infection cycle", *Virulence* **2**: 379–394.
- Ciatto, C., Bahna, F., Zampieri, N., VanSteenhouse, H.C., Katsamba, P.S., Ahlsen, G., Harrison, O.J., Brasch, J., Jin, X., Posy, S., Vendome, J., Ranscht, B., Jessell, T.M., Honig, B., and Shapiro, L. (2010) "T-cadherin structures reveal a novel adhesive binding mechanism", *Nat Struct Mol Biol*, **17**: 339–347.
- Cossart, P. (2011) "Illuminating the landscape of host-pathogen interactions with the bacterium *Listeria monocytogenes*", *Proc Natl Acad Sci U S A*, **108**: 19484–91.
- Cossart, P., and Sansonetti, P.J. (2004) "Bacterial invasion: the paradigms of enteroinvasive pathogens", *Science*, **304**: 242–8.
- Cossart, P., and Toledo-Arana, A. (2008) "*Listeria monocytogenes*, a unique model in infection biology: an overview", *Microbes Infect*, **10**: 1041–1050.
- Cummins, P.M., Dowling, A., and O'Connor, B.F. (2011) "Ion-Exchange Chromatography: Basic Principles and Application to the Partial Purification of Soluble Mammalian Prolyl Oligopeptidase", Ion-Exchange Chapter, Final Draft, 1-15
- Disson, O., and Lecuit, M. (2013) "In vitro and in vivo models to study human listeriosis: mind the gap", *Microbes Infect*, **15**: 971–80.
- Doyle, L. (1997) "Characterization calorimetry of binding interactions by isothermal titration", *Anal Biotechnol*, **8**: 31–35.
- Duff, M.R., Grubbs, J., and Howell, E.E. (2011) "Isothermal Titration Calorimetry for Measuring Macromolecule-Ligand Affinity", *Journal of Visualized experiments*, **55**: 2–5.
- Dussurget, O., Pizarro-Cerda, J., and Cossart, P. (2004) "Molecular determinants of *Listeria monocytogenes* virulence", *Annu Rev Microbiol*, **58**: 587–610.
- Einarson, M.B, Pugacheva, E.N, Orlinick, J.R. (2005) "Preparation of GST Fusion Proteins", *Protein-protein Interact*, **2**: 1–6.
- Elledge, H.M., Kazmierczak, P., Clark, P., Joseph, J.S., Kolatkar, A., and Kuhn, P. (2010) "Structure of the N terminus of cadherin 23 reveals a new adhesion mechanism for a subset of cadherin superfamily members", *PNAS*, **107**: 10708–10712.
- Freyer, M.W., and Lewis, E.A. (2008) "Isothermal Titration Calorimetry: Experimental Design, Data Analysis, and Probing Macromolecule/Ligand Binding and Kinetic Interactions", *Methods Cell Biol*, **84**: 79–113.
- Gaillard, J.L., Berche, P., Frehel, C., Gouin, E., and Cossart, P. (1991) "Entry of *L. monocytogenes* into cells is mediated by internalin, a repeat protein reminiscent of surface antigens from gram-positive cocci", *Cell*, **65**: 1127–1141.

Gallin, W.J. (1998) "Evolution of the "Classical" Cadherin Family of Cell Adhesion Molecules in Vertebrates", *Mol. Biol. Evol.*, **15**:9, 99-107.

Genheden, S., and Eriksson, L.A. (2013) "Of mice and men: Dissecting the interaction between *Listeria monocytogenes* Internalin A and E-cadherin", *Comput Struct Biotechnol J*, **6**: 1–13.

Gouin, E., Adib-Conquy, M., Balestrino, D., Nahori, M.-A., Villiers, V., Colland, F., Dramsi, S., Dussurget, O., and Cossart, P. (2010) "The *Listeria monocytogenes* InlC protein interferes with innate immune responses by targeting the I κ B kinase subunit IKK α ", *Proc Natl Acad Sci U S A*, **107**: 17333–17338.

Gregory, S.H., Sagnimeni, A.J., and Wing, E.J. (1996) "Expression of the inlAB operon by *Listeria monocytogenes* is not required for entry into hepatic cells in vivo", *Infect Immun*, **64**: 3983–3986.

Griko, Y.V. (1999) "Energetics of Ca²⁺ . EDTA interactions : calorimetric study", *Biophys Chem*, **79**: 117–127.

Hamon, M., Bierne, H., and Cossart, P. (2006) "*Listeria monocytogenes*: a multifaceted model", *Nat Rev Microbiol*, **4**: 423–34.

Hamon, M.A., Ribet, D., Stavru, F., and Cossart, P. (2012) "Listeriolysin O: The Swiss army knife of *Listeria*", *Trends Microbiol*, **20**: 360–368.

Harper, S., and Speicher, D.W. (2011) "Purification of proteins fused to glutathione S-transferase", *Methods Mol Biol*, **681**: 1–15.

Harrison, O.J., Jin, X., Hong, S., Bahna, F., Brasch, J., Wu, Y., Vendome, J., Felsovalyi, K., Hampton, C.M., Troyanovsky, R.B., Ben-shaul, A., Frank, J., Troyanovsky, S.M., Shapiro, L., and Honig, B. (2011) "The extracellular architecture of adherens revealed by crystal structures of type I cadherins", *NIH Public Access*, **19**: 244–256.

Hof, H. (2003) "History and epidemiology of listeriosis", *FEMS Immunol Med Microbiol*, **35**: 199–202.

Ireton, K. (2013) "Molecular mechanisms of cell-cell spread of intracellular bacterial pathogens", *Open Biol*, **3**: 1-15.

Jasnin, M., Asano, S., Gouin, E., Hegerl, R., Plitzko, J.M., Villa, E., Cossart, P., and Baumeister, W. (2013) "Three-dimensional architecture of actin filaments in *Listeria monocytogenes* comet tails", *Proc Natl Acad Sci U S A*, **110**: 20521–20526.

Jeyaletchumi, P, Tunung, R, Margaret, S.P, Son, R, Farinazleen, M.G, and Cheah, Y. (2010) Detection of *Listeria monocytogenes* in foods. *Int Food Res J* **17**: 1–11.

Jin, X., Walker, M. a., Felsovalyi, K., Vendome, J., Bahna, F., Manneppalli, S., Cosmanescu, F., Ahlsen, G., Honig, B., and Shapiro, L. (2012) "Crystal structures of *Drosophila* N-cadherin ectodomain regions reveal a widely used class of Ca²⁺-free interdomain linkers", *Proc Natl Acad Sci*, **109**: 127–134.

- Kocks, C., Gouin, E., Tabouret, M., Berche, P., Ohayon, H., and Cossart, P. (1992) "*L. monocytogenes*-induced actin assembly requires the actA gene product, a surface protein", *Cell*, **68**: 521–531.
- Krieger, Elmar, Nabuurs, Sander. B and Vriend, G. (2003) "Homology modeling", *Struct Bioinforma*, 507–521.
- Lam, G.Y., Fattouh, R., Muise, A.M., Grinstein, S., Higgins, D.E., and Brumell, J.H. (2011) "Listeriolysin O suppresses phospholipase c-mediated activation of the microbicidal NADPH oxidase to promote *Listeria monocytogenes* infection", *Cell Host Microbe*, **10**: 627–634.
- Leavitt, S., and Freire, E. (2001) "Direct measurement of protein binding energetics by isothermal titration calorimetry", *Curr Opin Struct Biol*, **11**: 560–566.
- Lecuit, M. (2005) "Understanding how *Listeria monocytogenes* targets and crosses host barriers", *Clin Microbiol Infect*, **11**: 430–436.
- Lecuit, M., Dramsi, S., Gottardi, C., Fedor-Chaiken, M., Gumbiner, B., and Cossart, P. (1999) "A single amino acid in E-cadherin responsible for host specificity towards the human pathogen *Listeria monocytogenes*", *EMBO J*, **18**: 3956–3963.
- Lecuit, M., Nelson, D.M., Smith, S.D., Khun, H., Huerre, M., Vacher-Lavenu, M-C., Gordon, J.I., and Cossart, P. (2004) "Targeting and crossing of the human maternofetal barrier by *Listeria monocytogenes*: Role of internalin interaction with trophoblast E-cadherin", *Proc Natl Acad Sci U S A*, **101**: 6152–6157.
- Lecuit, M., Vandormael-Pournin, S., Lefort, J., Huerre, M., Gounon, P., Dupuy, C., Babinet, C., and Cossart, P. (2001) "A transgenic model for listeriosis: role of internalin in crossing the intestinal barrier", *Science*, **292**: 1722–1725.
- Lee, K., Zhong, X., Gu, S., Krueel, A.M., Dorner, M.B., Perry, K., Rummel, A., Dong, M., and Jin, R. (2014) "Molecular basis for disruption of E-cadherin adhesion by botulinum neurotoxin A complex", *Science*, **344**: 1405–1410.
- Leung, N., Gianfelice, A., Gray-Owen, S.D., and Ireton, K. (2013) "Impact of the *Listeria monocytogenes* protein InlC on infection in mice", *Infect Immun*, **81**: 1334–1340.
- Li, J., Li, C., Xiao, W., Yuan, D., Wan, G., and Ma, L. (2008) "Site-directed mutagenesis by combination of homologous recombination and Dpn I digestion of the plasmid template in *Escherichia coli*", *Anal Biochem*, **373**: 389–391.
- Linask, K.K. (1992) "N-Cadherin Localization in Early Heart Development and Polar Expression of Na⁺, K⁺-ATPase, and Integrin during Pericardial Coelom Formation and Epithelialization of the Differentiating Myocardium", *Dev Biol*, **224**: 213–224.
- Liu, H., and Naismith, J.H. (2008) "Multiple-site plasmid mutagenesis protocol", *BMC Biotechnol*, **8**: 1–10.
- Low, J.C and Donachie, W. (1997) "A review of *Listeria monocytogenes* and Listeriosis", *Vet J* **153**: 9–29.

- Marks, D.S., Hopf, T.A., and Sander, C. (2012) "Perspective protein structure prediction from sequence variation", *Nat Publ Gr*, **30**: 1072–1080.
- Melton-Witt, J.A, McKay, S.L., and Portnoy, D.A. (2012) "Development of a Single-Gene, Signature-Tag-Based Approach in Combination with Alanine Mutagenesis To Identify Listeriolysin O Residues Critical for the In Vivo Survival of *Listeria monocytogenes*", *Infect Immun*, **80**: 1980–1986.
- Mengaud, J., Ohayon, H., Gounon, P., Mege, R.M., and Cossart, P. (1996) "E-cadherin is the receptor for internalin, a surface protein required for entry of *L. monocytogenes* into epithelial cells", *Cell*, **84**: 923–932.
- Mertz, A.F., Che, Y., Banerjee, S., Goldstein, J.M., Rosowski, K.A., and Revilla, S.F. (2012) "Epithelial cell – matrix traction forces", *PNAS*, **110**: 842–847.
- Monk, I.R., Casey, P.G., Hill, C., and Gahan, C.G.M. (2010) "Directed evolution and targeted mutagenesis to murinize *Listeria monocytogenes* internalin A for enhanced infectivity in the murine oral infection model", *BMC Microbiol* **10**: 318, 1-13.
- Monnier, A. Le, Autret, N., Join-Lambert, O.F., Jaubert, F., Charbit, A., Berche, P., and Kayal, S. (2007) "ActA is required for crossing of the fetoplacental barrier by *Listeria monocytogenes*", *Infect Immun*, **75**: 950–957.
- Mostowy, S., and Cossart, P. (2012) "Virulence factors that modulate the cell biology of *Listeria* infection and the host response", *Advances in Immunology*, **113**: 19-32
- Munteanu, B., and Braun, M. (2012) "Improvement of PCR reaction conditions for site-directed mutagenesis of big plasmids", *Zhejiang University Press*, **13**: 244–247.
- Navarro, P., Ruco, L., and Dejana, E. (1998) "Differential Localization of VE- and N-Cadherins in Human Endothelial Cells: VE-Cadherin Competes with N-Cadherin for Junctional Localization", *J Cell Biol*, **140**: 1475–1484.
- Nikitas, G., Deschamps, C., Disson, O., Niaux, T., Cossart, P., and Lecuit, M. (2011) "Transcytosis of *Listeria monocytogenes* across the intestinal barrier upon specific targeting of goblet cell accessible E-cadherin", *J Exp Med*, **208**: 2263–2277.
- Oevermann, A., Zurbriggen, A., and Vandeveld, M. (2010) "Rhombencephalitis caused by *Listeria monocytogenes* in humans and ruminants: A zoonosis on the rise?", *Interdiscip Perspect Infect Dis*, 1-21.
- Parida, S.K., Domann, E., Ronde, M., Müller, S., Darji, A., Hain, T., Wehland, J., and Chakraborty, T. (1998) "Internalin B is essential for adhesion and mediates the invasion of *Listeria monocytogenes* into human endothelial cells", *Mol Microbiol*, **28**: 81–93.
- Phan, Q.T., Myers, C.L., Fu, Y., Sheppard, D.C., Yeaman, M.R., Welch, W.H., Ibrahim, A.S., Edward, J.E., and Filler, S.G. (2007) "Als3 Is a *Candida albicans* Invasin That Binds to Cadherins and Induces Endocytosis by Host Cells", *PLoS Biol*, **5**: 543–557.

- Pierce, M.M., Raman, C.S., and Nall, B.T. (1999) "Isothermal Titration Calorimetry of Protein – Protein Interactions", **221**: 213–221.
- Rajabian, T., Gavicherla, B., Heisig, M., Müller-Altrock, S., Goebel, W., Gray-Owen, S.D., and Ireton, K. (2009) "The bacterial virulence factor InlC perturbs apical cell junctions and promotes cell-to-cell spread of *Listeria*", *Nat Cell Biol*, **11**: 1212–1218.
- Raman, S., Lange, O.F., Rossi, P., Tyka, M., Wang, X., Liu, G., Ramelot, T., Eletsy, A., Szyperski, T., Kennedy, M., Prestegard, J., Montelione, G.T., and Baker, D. (2010) "NMR Structure Determination for Larger Proteins Using Backbone-ONLY Data", *NIH Public Access*, **327**: 1014–1018.
- Ramaswamy, V., Cresence, V.M., Rejitha, J.S., Lekshmi, M.U., Dharsana, K.S., Prasad, S.P., and Vijila, H.M. (2007) "*Listeria*-review of epidemiology and pathogenesis", *J Microbiol Immunol Infect*, **40**: 4–13.
- Reinl, T., Nimtz, M., Hundertmark, C., Johl, T., Kéri, G., Wehland, J., Daub, H., and Jansch, L. (2009) "Quantitative phosphokinome analysis of the Met pathway activated by the invasin internalin B from *Listeria monocytogenes*", *Mol Cell Proteomics*, **8**: 2778–2795.
- Remacle, A.G., Shiryayev, S.A., and Strongin, A.Y. (2014) "Distinct Interactions with Cellular E-Cadherin of the Two Virulent Metalloproteinases Encoded by a *Bacteroides fragilis* Pathogenicity Island", *PLoS One*, **9**: 1–7.
- Rohatgi, R., Ma, L., Miki, H., Lopez, M., Kirchhausen, T., Takenawa, T., and Kirschner, M.W. (1999) "The interaction between N-WASP and the Arp2/3 complex links Cdc42-dependent signals to actin assembly", *Cell*, **97**: 221–231.
- Rosano, G.L., and Ceccarelli, E.A. (2014) "Recombinant protein expression in *Escherichia coli*: advances and challenges", **5**:172, 1-17
- Salazar, M. A., Kwiatkowski, A.V., Pellegrini, L., Cestra, G., Butler, M.H., Rossman, K.L., Serna, D.M., Sondak, J., Gerther, F.B., and Camilli, P.D. (2003) "Tuba, a novel protein containing bin/amphiphysin/Rvs and Dbl homology domains, links dynamin to regulation of the actin cytoskeleton", *J Biol Chem*, **278**: 49031–49043.
- Schnupf, P., and Portnoy, D. A. (2007) "Listeriolysin O: a phagosome-specific lysin", *Microbes Infect*, **9**: 1176–1187.
- Schubert, W.D., Urbanke, C., Ziehm, T., Beier, V., Machner, M.P., Domann, E., Wehland, J., Chakraborty, T., and Heinz, D.W. (2002) "Structure of internalin, a major invasion protein of *Listeria monocytogenes*, in complex with its human receptor E-cadherin", *Cell*, **111**: 825–836.
- Schwaiger, Christine S, Bjelkmar, Par, Hess, Berk and Lindahl, E. (2011) " $^3_{10}$ -Helix Conformation Facilitates the Transition of a Voltage Sensor S4 Segment toward the Down State", *Biophys J*, **100**: 1446–1454.
- Shen, Y., Naujokas, M., Park, M., and Ireton, K. (2000) "InIB-dependent internalization of *Listeria* is mediated by the Met receptor tyrosine kinase", *Cell* **103**: 501–510.

- Stavru, F., Bouillaud, F., Sartori, A., Ricquier, D., and Cossart, P. (2011) "*Listeria monocytogenes* transiently alters mitochondrial dynamics during infection", *Proc Natl Acad Sci U S A*, **108**: 3612–3617.
- Tamura, K., Shan, W.S., Hendrickson, W.A., Colman, D.R., and Shapiro, L. (1998) "Structure-function analysis of cell adhesion by neural (N)-cadherin", *Neuron*, **20**: 1153–1163.
- Travier, L., Guadagnini, S., Gouin, E., Dufour, A., Chenal-Francisque, V., Cossart, P., Olivio-Marin, J-C., Ghigo, J-M., Disson, O., and Lecuit, M. (2013) "ActA Promotes *Listeria monocytogenes* Aggregation, Intestinal Colonization and Carriage", *PLoS Pathog*, **9**: 1-16
- Travier, L., and Lecuit, M. (2014) "*Listeria monocytogenes* ActA: A new function for a “classic” virulence factor", *Curr Opin Microbiol*, **17**: 53–60.
- Tsai, Y.H., Disson, O., Bierne, H., and Lecuit, M. (2013) "Murinization of Internalin Extends Its Receptor Repertoire, Altering *Listeria monocytogenes* Cell Tropism and Host Responses", *PLoS Pathog*, **9**: 1–16.
- Venselaar, H., Krieger, E., and Vriend, G. (2000) Homology Modeling. Chapter 25, 1–21.
- Wei, G., Liu, F., and Wang, C.R.C. (1999) "Shape Separation of Nanometer Gold Particles by Size-Exclusion Chromatography", **71**: 2085–2091.
- Wing, E.J., and Gregory, S.H. (2002) "*Listeria monocytogenes*: clinical and experimental update", *J Infect Dis*, **185**: 18–24.
- Wollert, T., Pasche, B., Rochon, M., Deppenmeier, S., Heuvel, J. van den, Gruber, A.D., Heinz, D.W., Lengeling, A., and Schubert, W-D. (2007) "Extending the Host Range of *Listeria monocytogenes* by Rational Protein Design", *Cell* **129**: 891–902.
- Wu, S., Shin, J., Zhang, G., Cohen, M., Franco, A., and Sears, C.L. (2006) "The *Bacteroides fragilis* Toxin Binds to a Specific Intestinal Epithelial Cell Receptor", *Infect Immun*, **74**: 5382–5390.
- Xayarath, B., and Freitag, N.E. (2013) "Optimizing the balance between host and environmental survival skills: lessons learned from *Listeria monocytogenes*", *Futur Microbiol*, **7**: 839–852.
- Zalevsky, J., Grigorova, I., and Mullins, R.D. (2001) "Activation of the Arp2/3 complex by the *Listeria* ActA protein", *J Biol Chem*, **276**: 3468–3475.

APPENDICES

Appendix A: The list of chemicals and suppliers

Reagents	Suppliers
40% 37.5:1 bis-acrylamide	Bio-Rad
β-mercaptoethanol	Sigma
Acetic acid	Merck
Agarose	Lonza
Ammonium persulphate (APS)	Merck
Ampicillin	Roche
Bacteriological agar	Merck
Bromophenol blue	Sigma
Calcium chloride	Merck
Coomasie brilliant blue R250	Sigma
GeneRuler™ 100 bp Plus DNA ladder	Thermo Fischer Scientific
Dithiothreitol (DTT)	Roche
Ethylene diamine tetra acetic acid (EDTA)	Merck
Ethanol	Sugar Illovo Limited
Gel red	Biotium
Glacial acetic acid	Merck
Glutathione Sepharose (GS) beads	Qiagen
Glycine	Merck

Glycerol	Merck
Hydrochloric acid (HCl)	Merck
Hydroxyethyl piperazineethanesulfonic acid (HEPES)	Sigma
Isopropyl-D-1-thiogalactopyranoside (IPTG)	Roche
Reduced Glutathione	Sigma
Magnesium chloride	Merck
PCR reagents	Kappa Biosystems
Potassium chloride	Merck
Potassium phosphate	Merck
Protein molecular weight makers	Bio-Rad
Restriction enzymes	Fermentas
Sodium acetate	Merck
Sodium chloride	Merck
Sodium dodecyl sulphate (SDS)	Promega
Sodium phosphate	Merck
Sucrose	Merck
T4 DNA ligase	Fermentas
N, N, N', N' - Tetra methylethylene-diamine (TEMED)	Promega
Tris (hydroxymethyl) aminoethan (Tris)	Sigma
Tryptone and peptone	Merck
Urea	Sigma

 Yeast extract

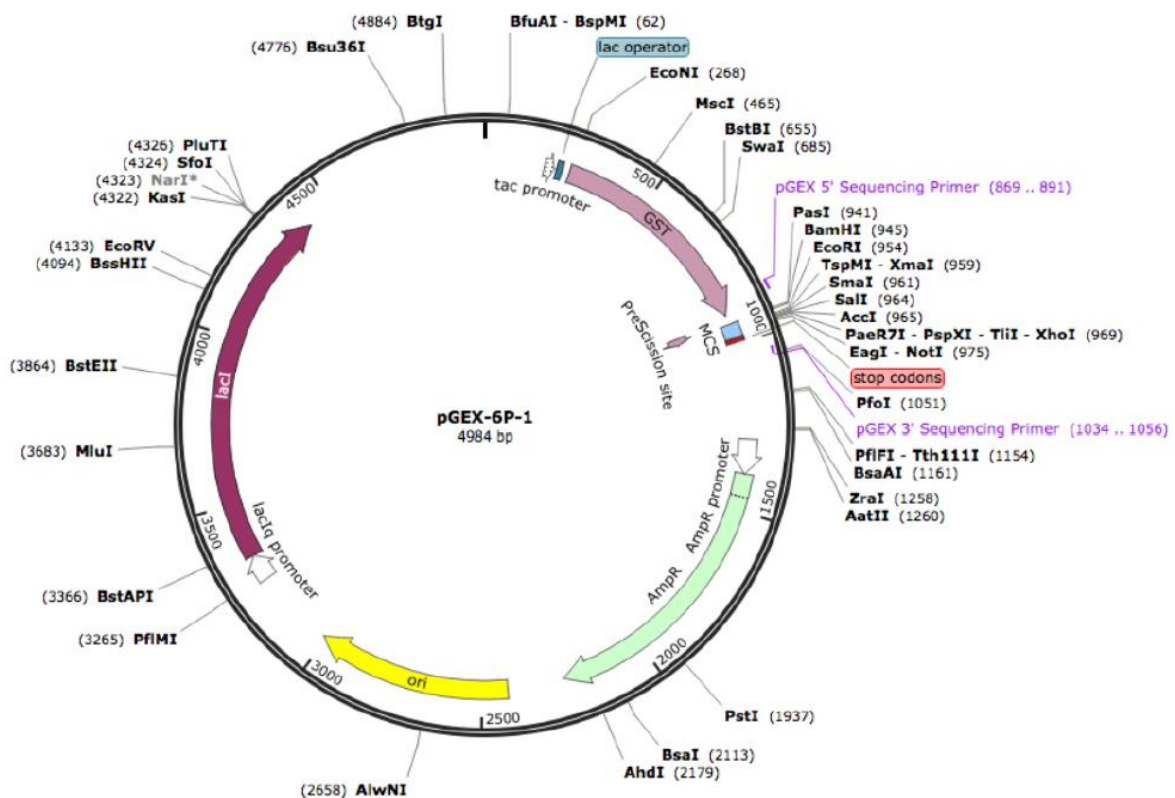
 Merck

Appendix B: The list of buffers and solutions with their compositions

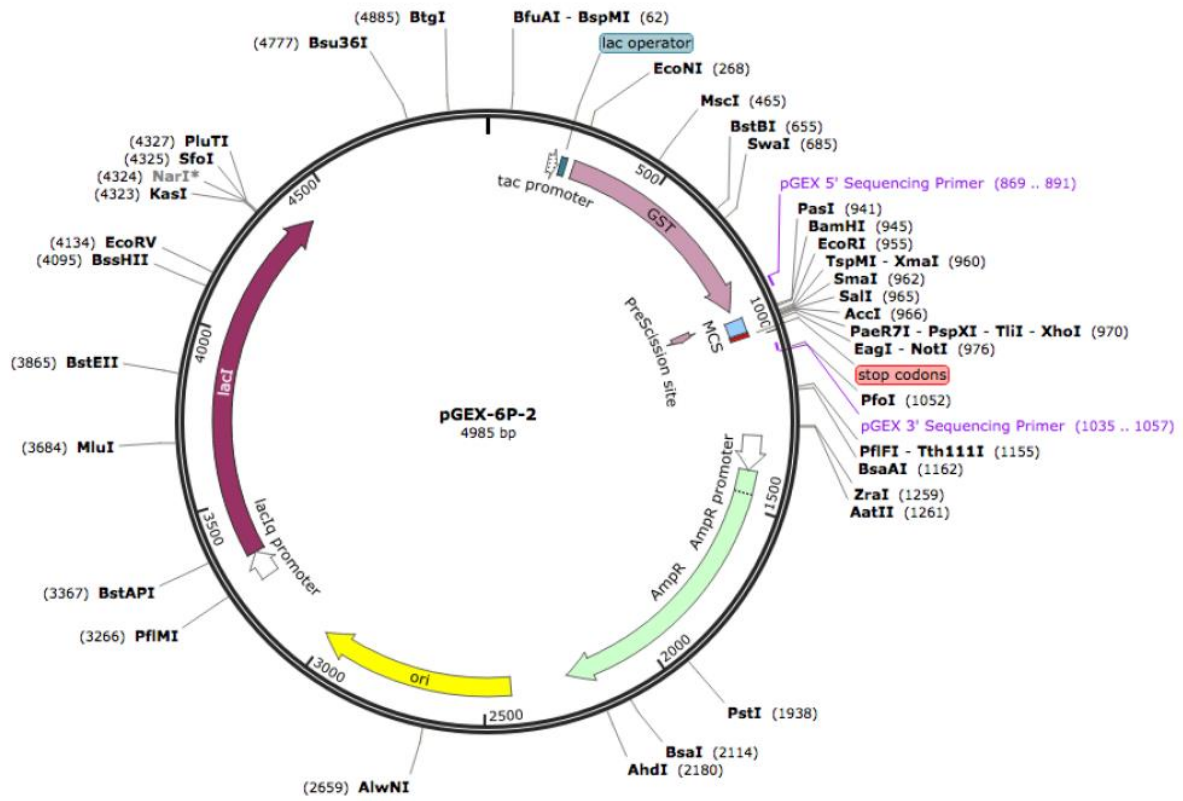
Buffers/Solutions	Composition
Phosphate-buffered saline (PBS) buffer (10 x)	80 g NaCl, 2 g KCl, 14.4 g Na ₂ HPO ₄ , 2.4 g KH ₂ PO ₄ pH 7
3C storage buffer	50 mM Tris-HCl pH 8, 150 mM NaCl, 10 mM EDTA, 20% (v/v) glycerol
Glutathione buffer	50 mM Tris/HCl pH 7.5, 25 mM L-reduced glutathione
HEPES buffer	50 mM HEPES pH 7.5, 20 mM CaCl ₂
Low salt buffer	25 mM Tris/HCl pH 8, 20 mM NaCl
High salt buffer	25 mM Tris/HCl pH 8, 1M NaCl
TAE buffer	40 mM Tris/HCl pH 7.5, 20 mM sodium acetate and 1mM EDTA adjusted to pH 8.5 with acetic acid
DNA loading buffer (10 x)	70% (w/v) sucrose, 0.25% (w/v) Bromophenol blue, 0.1 M EDTA, gel red
SDS-PAGE running buffer (10 x)	25 mM Tris-HCl pH 8, 192 mM glycine, 0.1% (w/v) SDS
SDS-PAGE loading buffer (8 x)	1.5 mM Tris/HCl pH 6.8, 10% (w/v) SDS, 5% (v/v) glycerol, 1% (v/v) β-mercaptoethanol and 0.25 % Bromophenol blue

SDS-PAGE stacking gel solution (15%)	30% bisagrylamide, 1.25 mL of 0.5 M Tris/HCl pH 6.8, 4.75 mL of dH ₂ O, 7.5 μL TEMED, 25 μL of 25% APS
SDS-PAGE resolving gel solution	30% bisagrylamide, 5 mL of 1.5 M Tris/HCl pH 8.8, 200 μL of 10% SDS, 7.2 mL of dH ₂ O, 20 μL TEMED, 50 μL of 25% APS
Staining solution	40 % (v/v) ethanol, 10% (v/v) acetic acid and 0.1 % (w/v) Coomassie brilliant blue R250
Destaining solution	40% (v/v) ethanol, 10% (v/v) acetic acid

Appendix C: The pGEX-6P-1 vector map



Appendix D: The pGEX-6P-2 vector map



Appendix E: The nucleotide and amino acids sequences of murine and human N-cadherin EC1-2 (wt) domains

Mouse N-cadherin EC 1-2 (wt)

gactgg
D W
gtcatccccgccaatcaacttgccagaaaaactccagaggacccttttctcaagagcttgtc
V I P P I N L P E N S R G P F P Q E L V
agaatcaggtctgatagagataaaaaacctttccctgagatacagcgtcactgggcagga
R I R S D R D K N L S L R Y S V T G P G
gctgaccagcctccaacgggcatcttcattatcaacccccatctcaggacagctgtcagtc
A D Q P P T G I F I I N P I S G Q L S V
acaaagcctctggatcgagagctgatagcccggtttcacttgagagcacatgacagtggac
T K P L D R E L I A R F H L R A H A V D
atcaatggcaatcaagtggagaacccccattgacattgtcatcaatggttattgacatgaat
I N G N Q V E N P I D I V I N V I D M N
gataacagacctgagtttctgcaccaggtttggaatgggtctggttcagagggatcaaaq
D N R P E F L H Q V W N G S V P E G S K
cctgggacgtatgtgatgacgggtcactgcccattgatgaggatgatecaaatgccctgaat
P G T Y V M T V T A I D A D D P N A L N
ggaatgctgvggtacaggtacctgaccaggcgccagcacacaccttccaccaacatggtt
G M L R Y R I L S Q A P S T P S P N M F
acaatcaacaatgagactggggacatcatcactgtggcagctgggtctggatcgagagaaa
T I N N E T G D I I T V A A G L D R E K
gtgcaacagtatacgttaataattcaagccacagacatggaaggcaatcccacttatggc
V Q Q Y T L I I Q A T D M E G N P T Y G
ctttcaaacacagccacagccgtcatcacgggtgacagatgtctaa
L S N T A T A V I T V T D V -

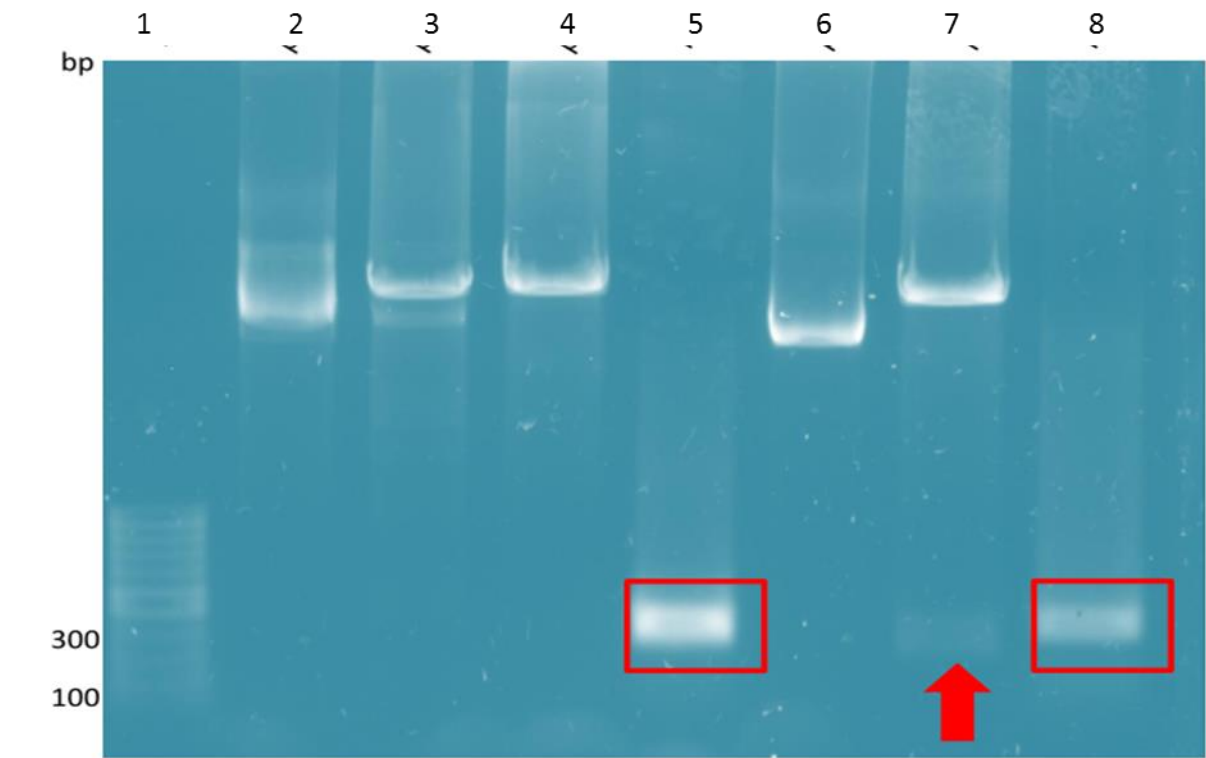
Human N cadherin EC 1-2 (wt)

gactggggtcatccctccaatcaacttgccagaaaaactccaggggaccttttctcaaa
D W V I P P I N L P E N S R G P F P Q
gagcttgcaggtcaggtctgatagagataaaaaacctttcactgaggtacagtgtaact
E L V R I R S D R D K N L S L R Y S V T
gggccaggagctgaccagcctccaactggtatcttcattatcaacccccatctcgggtcag
G P G A D Q P P T G I F I I N P I S G Q
ctgtcgggtgacaagcccctggatcgcgagcagatagcccggtttcatttgagggcacat
L S V T K P L D R E Q I A R F H L R A H
gcagtagatattaatggaaatcaagtggagaacccccattgacattgtcatcaatggttatt
A V D I N G N Q V E N P I D I V I N V I
gacatgaatgacaaacagacctgagttcttacaccaggtttggaatgggacagttcctgag
D M N D N R P E F L H Q V W N G T V P E
ggatcaaaq
G S K
cctggaacatatgtgatgacccgtaacagcaattgatgctgacgatcccaatgccctcaat
P G T Y V M T V T A I D A D D P N A L N
gggatggtgaggtacagaatcgtgtctcaggctccaagcacccttccaccaacatggtt
G M L R Y R I V S Q A P S T P S P N M F
acaatcaacaatgagactgggtgacatcatcacagtgccagctggacttgatcgagaaaaa
T I N N E T G D I I T V A A G L D R E K
gtgcaacagtatacgttaataattcaagctacagacatggaaggcaatcccacatatggc
V Q Q Y T L I I Q A T D M E G N P T Y G
ctttcaaacacagccacggccgtcatcacagtgacagatgtcaattag
L S N T A T A V I T V T D V N -

Appendix F: The constructs and bacterial strains

Construct	<i>E. coli</i> strain
pGEX-6P-1/mEC1	BL21 (DE3) CodonPlus-RP
pGEX-6P-1/hEC1	BL21 (DE3) CodonPlus-RP
pGEX-6P-2/mNC1	BL21 (DE3) CodonPlus-RP
pGEX-6P-2/hNC1	BL21 (DE3) CodonPlus-RP
pGEX-6P-1/InA ^{wt}	BL21 (DE3)
pGEX-6P-1/InA ^{S192N}	BL21 (DE3)
pGEX-6P-1/InA ^{Y369S}	BL21 (DE3)
pGEX-6P-1/InA ^m	BL21 (DE3)

Appendix G: Agarose gel electrophoresis of restriction digestion and ligations of mNC1 and hNC1 with pGEX-6P-2

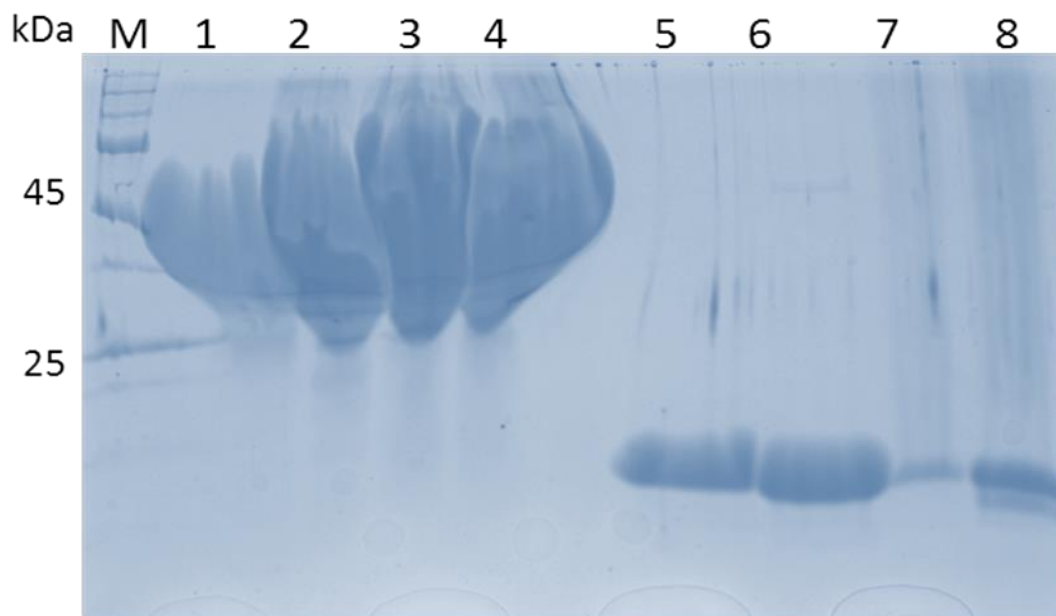


PCR products, plasmid DNA isolation and double digestion experiments of mNC1 and hNC1 constructs. The sizes of both genes are ~300 bp without endonuclease restriction enzymes cut sites and additional nucleic acids added to enhance cleavages. The first lane is GeneRuler™ 100 bp Plus DNA ladder with DNA bands sizes indicated on the left. The second lane is mNC1 construct plasmid DNA isolated. The third and fourth lanes are endonuclease restriction enzymes digested samples with *Bam*HI and *Xho*I and *Bgl*III and *Xho*I for mNC1 construct, respectively. The fifth lane is PCR product of mNC1 before digesting with endonuclease restriction enzymes. The sixth lane is hNC1 plasmid DNA isolated. The seventh lane represents double digestion of hNC1 construct by *Bam*HI and *Xho*I with red arrow indicating an insert from the plasmid DNA. The eighth lane is PCR product of hNC1 prior to digestion by endonuclease restriction enzymes.

Appendix H1: The table showing proteins concentrations prior to ITC

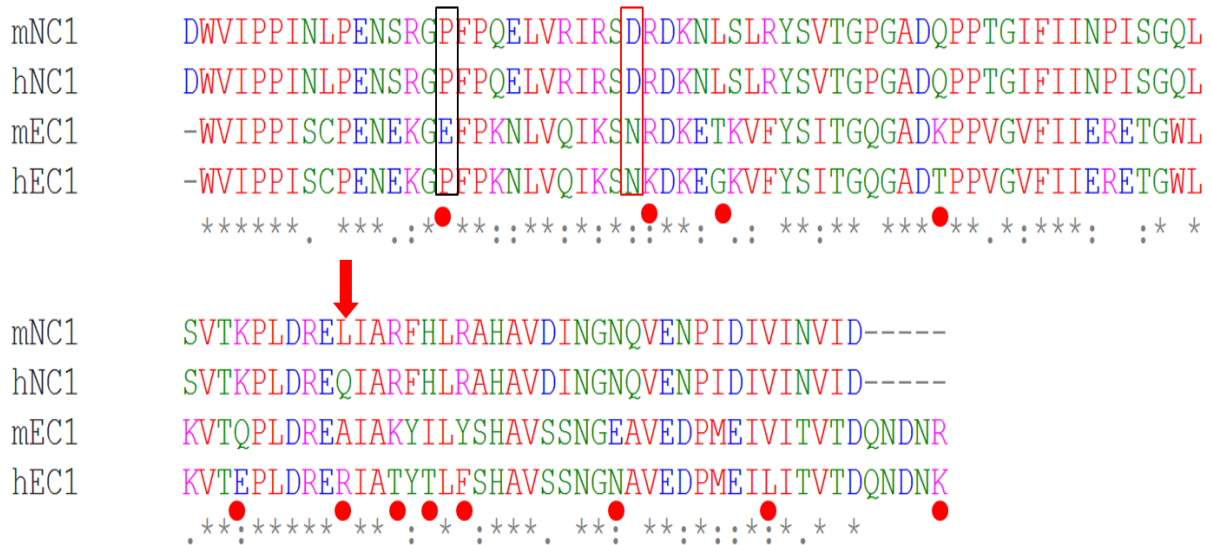
Protein	Concentration	Volume
InlA ^{wt}	28 mg/mL	0.8 mL
InlA ^{S192N}	36 mg/mL	0.8 mL
InlA ^{Y369S}	46 mg/mL	0.5 mL
InlA ^{S192N/Y369S} or InlA ^m	60 mg/mL	0.5 mL
mNC1	21 mg/mL	1 mL
hNC1	26 mg/mL	1 mL
mEC1	29 mg/mL	1 mL
hEC1	22 mg/mL	1 mL

Appendix H2: SDS-PAGE analysis of all samples prior to ITC



The proteins taken to ITC analysis following the SEC purification and concentrating. Lane M is unstained protein marker. Lane 1,2,3,4 are InlA^{wt}, InlA^{S192N}, InlA^{Y369S} and InlA^{S192N/Y369S} variants, respectively. Lane 5,6,7,8 are mNC1, hNC1, mEC1 and hEC1 domains, respectively.

Appendix I: Multiple sequence alignment of E- and N-cadherin N-terminal domains



The amino acid alignment of mEC1, hEC1, mNC1 and hNC1 domains. The black box indicates residues on position 16 crucial in a case of hEC1 binding to InlA and this amino acid is present in both N-cadherins. The red box indicates the asparagine residues on position 27 for both mEC1 and hEC1, and aspartic acids in both mNC1 and hNC1 domains. The red dots indicate the different amino acids in that position for mEC1 and hEC1 while the red arrow indicate the different amino acid in that position for mNC1 and hNC1 domains.