Usage guidelines

Expression and

characterisation of recombinant

heparinase I expressed in *Escherichia coli*

Shelley James

A thesis submitted in partial fulfilment of the requirements
of the University of Sunderland for the degree of

Master of Philosophy

May 2022

Declaration

This dissertation and the experimental work herein is all my own, except where specifically indicated in the text. It has not been previously submitted for any qualification at any university.

Acknowledgments

Throughout the completion of this MPhil, I have received a great deal of support and assistance. I would like to express my deepest appreciation to Mr Albert Pattinson of Hart Biologicals and Innovate UK for providing funding for this research, and to Manchester Institute of Biotechnology for the use of both their lab facilities, specialist equipment and for kindly donating a selection of their propriety plasmids during the expression stage of the project.

I would like to offer my sincerest thanks to my academic supervisors Dr. Edward McKenzie from The University of Manchester, and Dr. Lewis Bingle and Prof. Noel Carter from The University of Sunderland. Their expert knowledge, guidance and feedback has been invaluable.

I give kind thanks to the Knowledge Transfer Partnership committee members Mr Albert Pattison, Dr Edward McKenzie, Prof. Alan Dickson, Dr Joanne Flannelly, Ms Jemma Elston, and Mr John Clayton.

Finally, special thanks to my family for supporting me during my studies, especially my fiancé Dave for his unwavering support and patience.

Abstract

Introduction: Heparin is a complex, sulphated polysaccharide which is essential for many biological activities, one of which is to prevent blood coagulation. Because of its immediate effect, heparin has long been the parenteral drug of choice when a rapid anticoagulant effect is needed, particularly in the treatment of thromboembolism and deep vein thrombosis. However, blood samples contaminated with heparin are known to affect the Prothrombin Time and the Activated Partial Thromboplastin Time in *in vitro* tests. It is therefore imperative to be able to halt heparin activity when desired. Heparinases produced by *Pedobacter heparinus* neutralise the effects of heparin and allow for the unhindered analysis of haemostasis. *P. heparinus* produces three variants of heparinase, namely I, II and III, of which heparinase I is the most effective at neutralising heparin. Heparinase I has many clinical applications, which has led to the demand for a pure, high yield product. However, purifying HepI has proven to be time consuming, laborious, and can result in a low yield product. HepI has been expressed by others with solubility enhancing fusion partners, it has not however been produced in a yield that is commercially viable. The leading commercial provider of HepI, IBEX Technologies, produce the enzyme in 250-500 µl aliquots with a specific activity of 90-110 IU/mg. Initially, we aimed to show that soluble recombinant HepI protein could be expressed in high yield, which was just as, if not more, biologically active against unfractionated heparin and low molecular weight heparin than the leading commercial provider. **Method and Results:** A library of seven expression vectors was created to maximise soluble expression of HepI in *Escherichia coli.* DNA concentration and purity of each expression vector was determined by UV spectrophotometry at 260 nm and 280 nm wavelengths. All the expression plasmids DNA sequences were verified by Sanger sequencing and the amino acid sequence matched to the native HepI sequence when using a pairwise alignment tool. The concentration of the expression plasmids varied between 34.8 ng/µl and 126.5 ng/µl and purity ranged from 1.79 to 1.85. The plasmids were used to transform five host *E. coli* strains under IPTG induction and auto-induction, as well as with two induction temperatures. The concentration of soluble HepI protein expressed ranged from 9.8 to 17.1 mg per 35 ml culture. Five vector / host combinations were chosen for small scale purification using metal IMAC beads. The vector / host combination which produced the highest yield of purified HepI protein was pNTHIS-SUMOCDOHepI + BL21 (DE3) at 161.9 mg, however activity was found to be around 16 times lower than reported in the literature. Insoluble expression and refolding trials were conducted and resulted in the expression of 280 mg/L of insoluble HepI, with a specific activity calculated between 1876.38 and 2514.98 U/L culture across three

experiments. **Conclusion:** Promising data were generated when expressing HepI in the insoluble form and results were comparable to the literature. Refolding trials produced between 1876.38 and 2514.98 U/L culture. Several avenues of optimising the process could be looked at such as new vector/host combinations, the use of a protamine anchor to increase Hep I affinity to heparin, utilising alternative expression hosts to aid in glycosylation and different methods of long-term storage such as lyophilising to reduce freeze-thaw denaturation.

Contents

List of Figures

List of Tables

Chapter 1. Introduction

1.1. Haemostasis and Anti-coagulants

Haemostasis is the process by which the body keeps thrombus formation and destruction at equilibrium. Under normal conditions, blood flows through the cardiovascular system in a liquid state; that is until injury to the vascular endothelium occurs (1). The vascular endothelium is a monolayer of endothelial cells which line the entire circulatory system, it acts as a protective layer and physical barrier between blood and tissues, provides a thromboresistant surface, and produces vascular regulatory receptors, molecules and hormones in response to stimuli (2). A plethora of pro- and anti-coagulant molecules have been identified as being produced (and / or released) from the vascular endothelium dependent on the environmental conditions (3). The dynamic role of the vascular endothelium means it plays an integral role in the maintenance of vascular tone, haemostasis, and thrombosis. Upon injury, blood is exposed to the sub-endothelium matrix and the highly pro-coagulant elements that lie within, such as tissue factor and collagen (4). Exposure to these elements has a dual effect, firstly in activating platelets, and secondly in activating the coagulation cascade. At the site of injury, platelets (thrombocytes) become attached via their surface proteins glycoprotein Ib-V-IX to von Willebrand factor (vWF) and Ia/IIa or IV to collagen (5). The binding to collagen activates platelets and causes them to release the contents of their α- and dense granules (6). Degranulation releases adenosine di-phosphate (ADP), thromboxane$_2$ (TXA$_2$), coagulation factor V (FV) thrombin and vWF, which amplifies the recruitment of more platelets and their subsequent activation. Activated platelets also bind to other platelets through fibrinogen bridges to produce a primary platelet 'plug' at the site of vascular injury. The second stage is the activation of the coagulation cascade which is a series of amplification reactions that ultimately converts fibrinogen to fibrin, stabilising the platelet mesh to form an insoluble thrombus (clot). The coagulation cascade is activated by the extrinsic pathway from tissue factor exposure and the intrinsic pathway from contact activation. When activated, the phospholipid membrane on the platelets surface change to allow for the binding of coagulation factors Va and VIIIa; VIIIa forms a complex with factors IXa, X and calcium ions known as a Xase on the platelet surface, which converts factor X to Xa. Xa then forms a complex with factor Va, FII (prothrombin) and calcium ions called a prothrombinase which converts prothrombin to thrombin (FII to FIIa) (7–9). Thrombin is initially

released in small amounts from the activated platelets, but this is amplified exponentially from the coagulation cascade, and consequently, more platelets and coagulation factors are activated, thus sealing the vessel wall, and preventing bleeding. Thrombus production is localised to the site of vascular injury due to the dilution effect from flowing blood, and a surrounding environment which favours anti-coagulant activity (10). Coagulation is balanced by negative feedback mechanisms to prevent over-coagulation, which would cause widespread thrombosis. As thrombin acts as a procoagulant, it also acts as a negative feedback by activating plasminogen into plasmin and stimulating the production of Antithrombin III (ATIII) and thrombomodulin (11). Whilst plasmin degrades the fibrin mesh, ATIII deactivates thrombin and factor X (12,13). Thrombomodulin activates protein C and this, along with co-factors including protein S, activated protein C degrades factor Va and factor VIIIa, thus slowing the rate of clotting (14). The coagulation cascade, including positive and negative feedback mechanisms is shown in **Figure 1**.
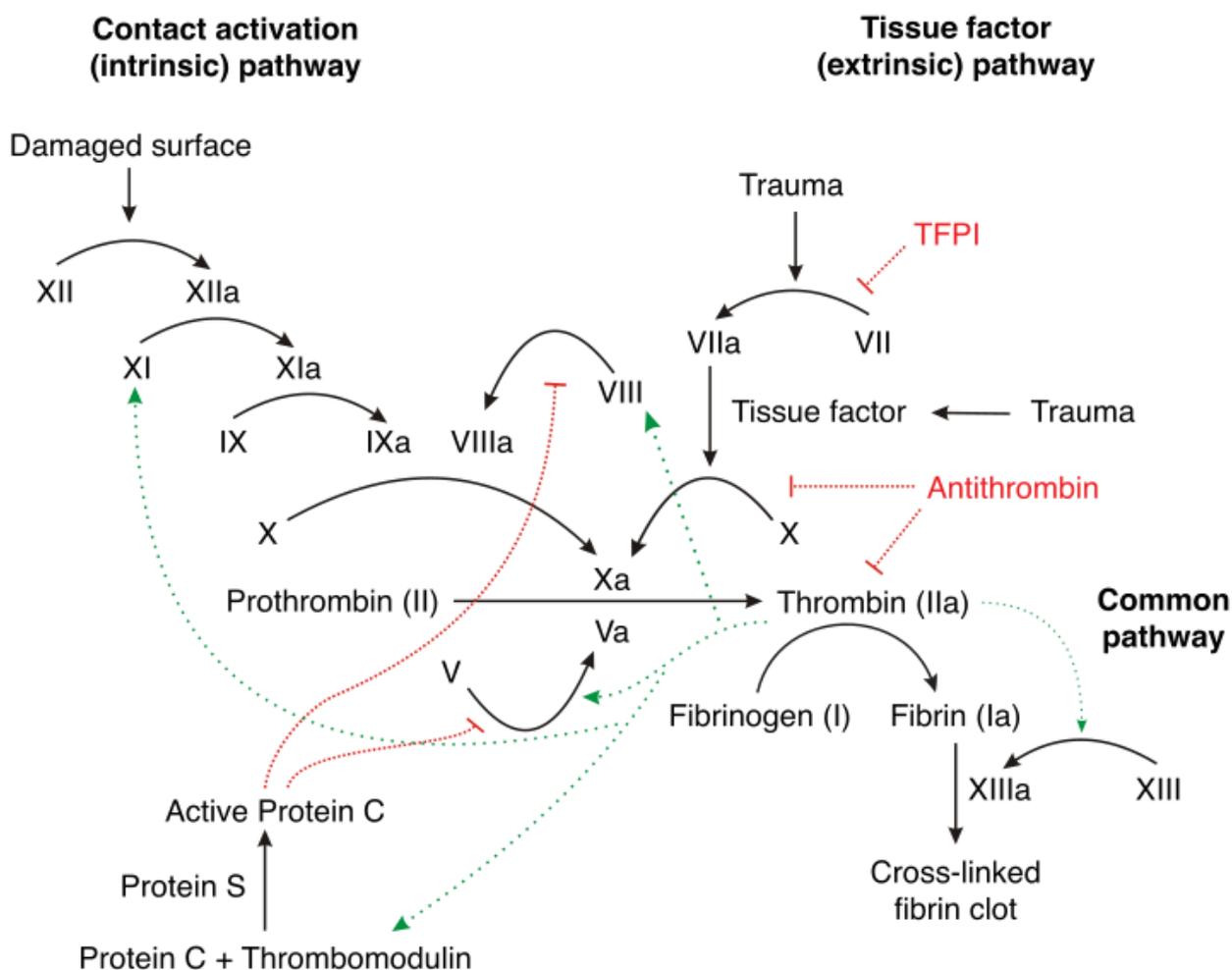
Figure 1: Overview of coagulation cascade. *Green arrows indicate pro-coagulant mechanisms which amplify coagulation and red bars indicate anti-coagulant mechanisms to prevent unregulated thrombus production. TFPI is Tissue Factor Pathway Inhibitor. Picture credited to https://en.wikipedia.org/wiki/File:Coagulation_full.svg*

Thrombosis is a pathological condition whereby an excessive pro-coagulation response is activated in the blood vessels and is usually caused by a loss of or dysfunction in the endothelium or blood components, genetic deficiencies, auto-immune disorders, or disturbed blood flow (15). Thrombosis can occur in any vessel within the circulatory system and can lead to congestion, ischemia or necrosis of the immediate vessel tissue or the thrombus can become dislodged (or parts thereof), known as an embolus, and travel to other parts of the cardiovascular system. A pulmonary embolism (PE) is caused when a thrombus travels to the lungs and blocks the pulmonary arteries and deep vein thrombosis (DVT) is the most common cause of PE (16,17). Pulmonary embolisms cause low blood pressure coupled with reduced blood oxygen levels and increased heart rate which can lead to both acute and chronic lung and heart damage (18,19). Acute PE is considered a medical emergency and contributed to an average of 12,640 deaths per annum in England, with a mortality rate of 21.7 per 100,000

between 2013-2015 (20). The treatment of PE or DVT is intravenous heparin or low molecular weight heparin (LMWH) followed by long term anticoagulant therapy (20,21) with Vitamin K antagonists such as warfarin, or with direct oral anticoagulants (DOAC's) such as Apixaban / Rivaroxaban (direct FX inhibitors) or Dabigatran / Argatroban (direct FII inhibitors) (22,23)

Heparin has been extensively studied since its discovery in 1916 and is one of the oldest medicines still in use today (24,25). Heparin is part of the glycosaminoglycan (GAG) family and is a complex, acidic, sulphated polysaccharide found in the extracellular matrix and on cell surfaces. Heparin is involved in many biological pathways and processes such as binding growth factors and cytokines, modulating signalling pathways and providing structural support and hydration to the extra cellular matrix (26–28) and thus plays an important role in proliferation, development, inflammation and disease (29–32). Heparin and heparin sulphate are composed of long repeating chains of disaccharides made up of α1-4 linked glucosamine and α/β1-4 linked uronic acid residues (L-iduronic or D-glucuronic acid) (**Figure 2**). The functional groups of these disaccharide units have a varying degree of modification making heparin heterogeneous (33–36). Such heterogeneity arises due to the variable presence of either acetyl or sulfo groups at the N – position of the glucosamine, sulfation at the 2-O – position on the uronic acid or 6-O – and 3-O – positions on the glucosamine, and epimerization at the C-5 of the uronic acid.
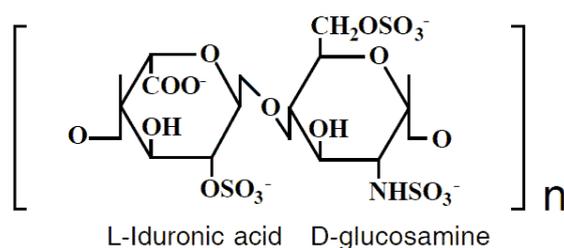
Figure 2: The major repeating disaccharide units in heparin. *The most frequent type of disaccharide unit in heparin is made up of alternating 1 to 4 linked, sulfated monosaccharide residues of L-iduronic acid and D-glucosamine. Picture credited to https://diapharma.com/heparin/.*

The molecular weight of heparin ranges from 3-30 kDa with a mean of 15 kDa (37) and is found in several tissues, with the highest concentrations found in the lungs, liver and in mast cells (38). About one-third of the

heparin chains produced by mast cells (unfractionated heparins) contain a unique pentasaccharide (**Figure 3**), which is a high-affinity binding site for ATIII. Binding to unfractionated heparin causes a conformational change in ATIII, which increases its ability to neutralise coagulation factors Xa and FIIa (thrombin) by up to 1000-fold (24). Consequently, this prevents clot formation via fibrin crosslinking and activation of platelets and other coagulation factors such as FV and FVIII by thrombin (40).
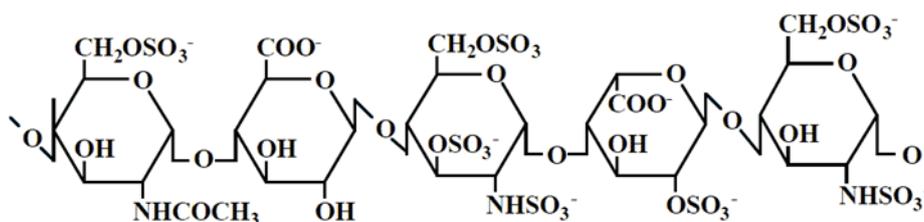
Figure 3: Heparin binding site for antithrombin III. *Picture credited to **https://diapharma.com/heparin/***

LMWH's are small fragments of unfractionated heparin which have been enzymatically cleaved to produce fragments of smaller molecular weight, typically fewer than 18 saccharides; these however do not have the same anticoagulant activity as unfractionated heparin as they do not increase the neutralisation effect of ATIII on thrombin (41)**.** The effect of ATIII, unfractionated heparin and LMWH on coagulation is shown below in **Figure 4**.
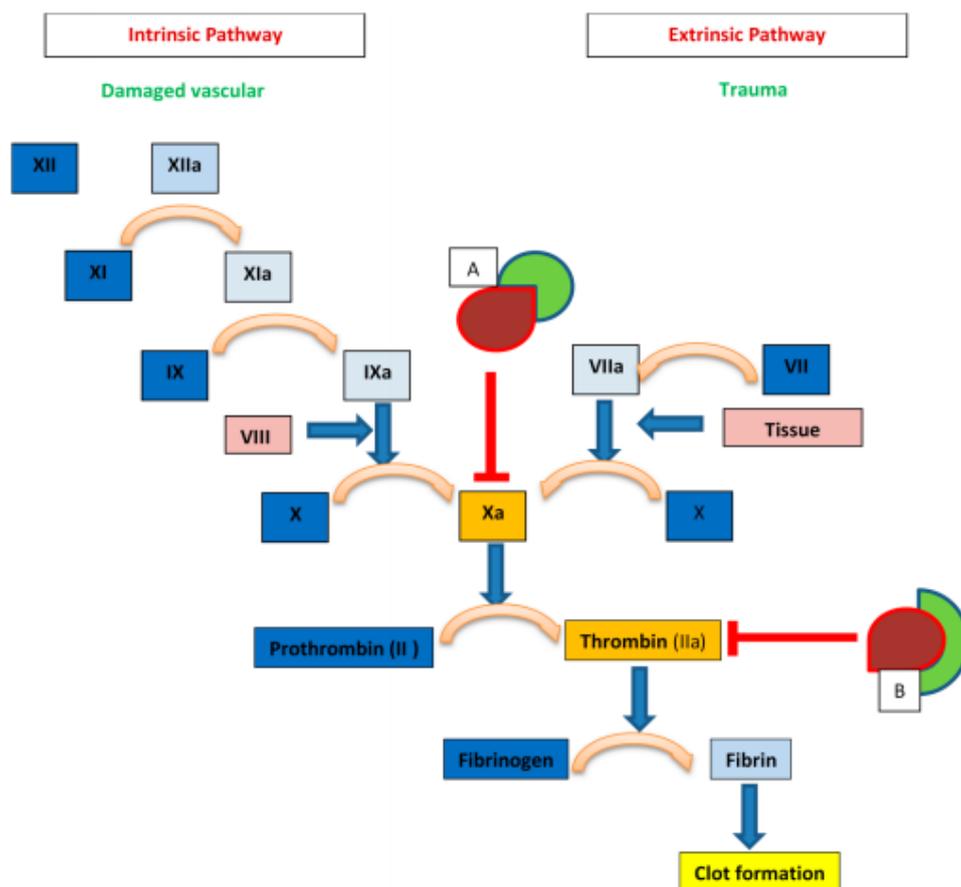
Figure 4: The effect of ATIII, unfractionated heparin, and LMWH on coagulation. *A: ATIII (red) when bound to unfractionated heparin / LMWH (green) has an increased capability to neutralise factor Xa. B: ATIII (red) when bound to unfractionated heparin (green) also has an increased affinity to neutralise thrombin (IIa) which prevents thrombin activating platelets, factor V and factor VIII. Picture credited to Oduah, Linhardt and Sharfstein (41).*

Because of its immediate effect, heparin has long been the parenteral drug of choice when a rapid anticoagulant effect is needed, especially in the treatment of PE and DVT (42). However, blood samples contaminated with heparin are known to elongate the Prothrombin Time (PT) and the Activated Partial Thromboplastin Time (APTT) in *in vitro* tests (43). As the PT and APTT tests are routinely used to determine the coagulation status of a patient, it is necessary to determine their baseline values without heparin interference. In addition to its anticoagulant effects, heparin is known to increase vessel wall permeability, suppresses vascular smooth muscle cell proliferation, and suppress osteoblast formation and activate osteoclasts, which indirectly promotes bone loss. It is therefore imperative to be able to halt heparin activity when desired.

Protamine has long been used to reverse the anticoagulant effect of heparin in patients. Protamine is a cationic polypeptide, rich in arginine, which electrostatically binds to unfractionated heparin to form an inactive

complex (44), thus reversing its anticoagulant effect (45). However, repeated small doses or singular high doses of protamine have been noted to cause severe adverse side effects including anaphylactic and anaphylactoid reactions (46), heparin rebound (47), systemic arterial hypotension / pulmonary artery hypertension (48), and an increased mortality risk after cardiac surgery (49). Studies have found cases of anaphylactic shock after slow intravenous administration of protamine sulfate to patients with diabetes mellitus which suggests a cross-allergy to the protamine present in protamine zinc insulin (50,51). When administered on its own, protamine inhibits platelet function, which can induce an anticoagulant effect in patients (52), it significantly effects intrinsically (APTT), and extrinsically (PT) activated clot times *in-vitro* (53,54). Additionally, protamine can only partially neutralise LMWH (55).

1.2. Heparinase I

Heparin lyases or heparinases are enzymes produced by various microbial species which catalyse the depolymerisation reactions of heparin (56). Based on substrate specificity, heparinases have been divided into three groups, heparinase I (EC4.2.2.7), heparinase II (EC 4.2.2.8) and heparinase III (EC 4.2.2.-). Heparinases are found in both Gram-negative and Gram-positive bacterial species including *Pedobacter heparinus,* (once classified as *Flavobacteria Heparinum*)*, Bacteroides spp., Acinetobacter calcoaceticus* and *Pseudomonas aeruginosa, Streptomyces, and Bacillus spp.* (57–61). Of the fungal strains, various species of *Aspergillus* are reported to produce heparinases (62). Heparinase production can act as a virulence factor for opportunistic pathogens such as *Bacteroides heparinolyticus* in abscess formation in the oral cavity (63), and *P. aeruginosa* in thermal injury (64). For other species, such as *P. heparinus and Aspergillus spp.* heparin can be used as a source of nitrogen, carbon and sulphur (62). Of the heparinases produced by bacteria, those by *P. heparinus* have been characterised and studied most extensively. *P. heparinus*, is a Gram-negative, non-pathogenic microorganism found in soil (65) which produces all three heparinase types (66). Heparinase I (42.8 kDa) neutralises the effect of unfractionated heparins, heparin sulphate and LMWH by cleaving at the highly sulphated polysaccharide domains containing 1-4 linkages between hexosamines and O sulphated uronic acid residues. Heparinase II (85.7 kDa) cleaves both high and low sulphated domains and heparinase III (73.2 kDa) specifically cleaves heparin sulphate (67–70) (**Figure 5A**). The linkages susceptible to heparinase I cleavage

reside directly within the ATIII binding site of heparin and LMWH (**Figure 6**), which makes it the most efficient of the heparinases in reducing anticoagulant activity (71). The cleavage of heparin by heparinases can be detected by UV spectrometry at 232 nm as it results in the formation of unsaturated uronic acid residues (72).

Structurally, heparinase I is composed of a β-jellyroll domain made of two concave β-sheets (eight antiparallel β-strands each) that have a glucanase fold and the thumb-like insertion domain extending from one side of the β-jellyroll (73) (**Figure 5B**). Heparinase I has a long positively charged 'canyon' where heparin binds (**Figure 5B**) and is subsequently depolymerized. A histidine residue at position 203 along with cystine[135] form the catalytic domain of heparinase I (74). Heparinase II composed of three domains: an N-terminal α-helical domain, a central two-layered β-sheet domain, and a C-terminal domain forming a two-layered β-sheet (75). Heparinase III is composed of an N-terminal α/α-barrel domain and a C-terminal antiparallel βsheet domain as its basic scaffold (76).

The enzymatic properties of recombinant heparinase I have been studied in detail (73,77–80) and show that the optimal conditions needed for maximum enzymatic activity are temperatures between 37-33 °C and at pH 7.0-9.0. Activity and stability of the enzyme is significantly affected at temperatures below 27 °C and above 33 °C and at pH below 7.0 and above 9.5. Metal ions such as calcium, magnesium, and copper increase activity, whereas manganese and lead negatively impact activity. When calcium chloride is used in purification stages acts as an activity stabiliser (81).
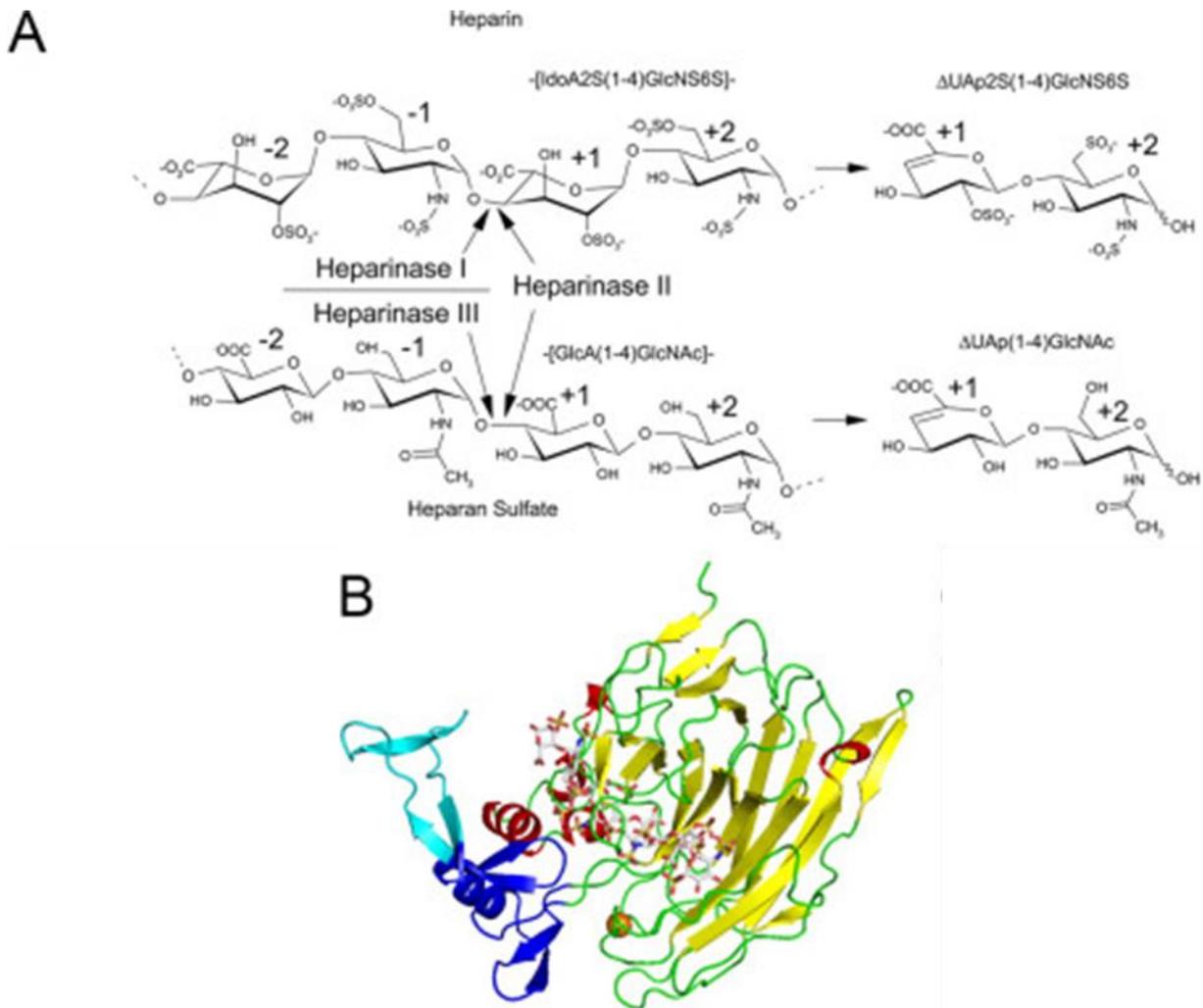
Figure 5: Cleavage of heparin by heparinases. *A) Degradation of heparin and heparin sulfate by heparinases I–III. The arrows indicate the cleavage site. B) Schematic representation of the heparinase I-heparin oligosaccharide complex. Heparin is shown in stick representation, and Ca²⁺ is shown as an orange sphere. Picture credited to Han et al* *(73).*
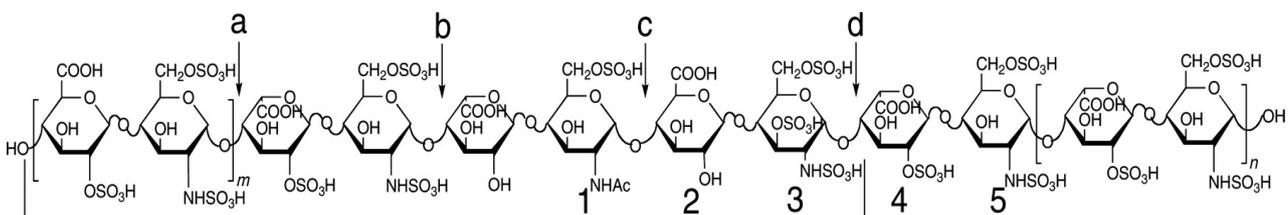
Figure 6: Heparin structure including ATIII binding site and heparinase I cleavage sites. *Linkages a, b and d are susceptible to heparinase I cleavage, whereas linkage c is not. Residues 1 to 5 denote the ATIII binding site. Linkage d resides within the ATIII binding site and when cleaved by heparinase I results in the reduced anticoagulant activity of heparin (Picture credited to Xiao et al* *(70).*

Heparinase I, unlike protamine, has been shown to have no haemodynamic effects when administered to patients in bolus form for the prophylaxis of DVT, PE, unstable angina pectoris and acute arterial occlusion (37,82). At dosages sufficient to reverse heparin, heparinase I shows no significant inhibition of platelets (83,84). When used in *in-vitro* coagulation tests, heparinase I can precisely detect trace amounts of heparin, without artefacts; even when blood is diluted, thus making it superior to protamine for *in-vitro* coagulation testing (54,85). Heparinase I is also used to enzymatically cleave heparin into LMWH for the prevention and treatment of deep vein thrombosis, pulmonary embolism, and myocardial infarction (66,86–88).

Heparinase I has many clinical applications, which has led to the demand for a pure, high yield product. However, purifying heparinase I from heparinase II and III as well as other co-expressed enzymes such as heparitinase, chondroitinase and hyaluronidase has proven time consuming, laborious, and can result in a low yield product (26). The first method describing the successful purification of heparinase I from *P. heparinus* is attributed to Yang *et al* (89), however, due to low yield, a recombinant expression system was developed in *E. coli* and optimised (26,90,91). However, a major disadvantage of expression in *E. coli* is that heparinase I can form into inclusion bodies within the host cell, which are particles of aggregated protein which leave the enzyme biologically inactive, this then requires a process of unfolding and refolding in order to get the protein into correct tertiary structure (92). Enhancement of soluble protein expression in *E. coli* can be achieved with the use of fusion tags (93). Such fusion tags include maltose binding protein (MBP), glutathione S-transferase (GST), N-utilisation substance protein A (NusA) and small ubiquitin-like modifier (SUMO) (94,95).

Because of the tendency to form inclusion bodies, researchers have previously tried to express heparinase I fused with both small polypeptides and chaperone proteins to encourage solubility (96). However, the focus of recent research is on the creation of LMWH for commercial purposes (97).

1.3. Overview and Aims

Heparinase I has many commercial and clinical implications which have increased demand for a pure, high yield, highly bioactive product. Because of difficulties in harvesting high yield, pure heparinase I from the native host, a recombinant expression system was soon adopted, however, heparinase I tends to aggregate into inclusion bodies within the host cell. This results in an inactive enzyme which needs to be extensively treated to unfold and refold the protein, a process which not only increases process time and cost but also reduces yield. Heparinase I has been expressed by others with solubility enhancing fusion partners and has been shown to be active against heparin in spectrophotometric assays, it has not however been produced in such a yield that is commercially viable. The current market leader in supplies of Heparinase I for Point-of-Care devices is IBEX technologies (Canada), whose enzyme is the only one approved by the FDA and EU (98). IBEX state that they produce a version of heparinase I using a proprietary expression system in *P. heparinus*, meaning that the recombinant heparinase I is glycosylated.

In this project we Initially aimed to show that soluble heparinase I could be expressed in high yield, which is just as, if not more, biologically active against UFH and LMWH than the current market standard preparation. Expressing soluble heparinase I using a variety of methods (solubility tags, expression conditions) was successful, however the protein lacked activity against heparin in *in vitro* tests, therefore an insoluble method of expression was adopted.

Chapter 2. Materials and methods

2.1. Protocols for protein production, purification, and characterisation

In this study a library of heparinase I constructs was created with different solubility enhancing tags. These constructs were expressed under different conditions as detailed in sections 2.1.2.2. Where insoluble protein was expressed, section 2.1.2.3 describes the three methods of inclusion body processing. Both insoluble and soluble protein was tested for activity as per 2.1.2.5. Bradford analysis of the protein and activity testing identified the optimal construct, strain, and growth condition combinations to produce the highest yield of active enzyme.

2.1.1. Materials

2.1.1.1. Media

All media used in these experimental procedures was made up to volume with de-ionised (DI) water (Fresenius Kabi) and autoclaved in a portable autoclave at 121 °C for 15 minutes before use (Dixons STXX28). Luria-Bertani (LB) medium is optimised for cell growth and yield in E. *coli* cultures and was used in the preparation of mini cultures, soluble protein expression and for agar plates. For LB media, 15g powder (Formedium w/o NaCl) was mixed with 10 g/L NaCl to produce a final composition of 10 g/L tryptone, 5 g/L yeast extract, pH 7.0. Auto-Inducing (AI) media composition is described in

. AI media has an LB base with the addition of trace elements to grow IPTG inducible strains of bacteria. Initial growth is without induction due to the presence of glucose in the media, however once depleted (usually near cell density saturation) lactose is the only available metabolite. Uptake of lactose activates the *lac* operon within the cell. Once ingested, lactose is converted to allolactose by β-galactosidase which releases the lac repressor from its binding site on the DNA (**Table 1**)

Table 1: Composition of AI Media.

Compound	Formula	Concentration
Ammonium Sulphate	$(NH_4)_2SO_4$	3.3 g/L
Potassium Dihydrogen Phosphate	KH_2PO_4	6.8 g/L
Di-sodium Hydrogen Phosphate	Na_2HPO_4	7.1 g/L
Magnesium Sulphate	$MgSO_4$	0.15 g/L
Glucose	$C_6H_{12}O_6$	0.5 g/L
Lactose	$C_{12}H_{22}O_{11}$	2 5 g /L
Trace Elements	---	0.03 g/L

Super Optimal broth with Catabolite repression Media: SOC media (NEB) was used to optimise transformation of competent E. *coli* cells, SOC media is a nutrient-rich microbial broth containing peptides, amino acids, water-soluble vitamins, and glucose in a low-salt formulation. The composition of SOC media is outlined in **Table 2.**

Table 2: Composition of SOC Media.

Compound	Chemical Formula	Concentration
Vegetable Peptone	---	2 % w/v
Yeast Extract	---	0.5 % w/v
Sodium Chloride	NaCl	0.58 g/L
Potassium Chloride	KCL	0.18 g/L
Magnesium Chloride	$MgCl_2$	0.95 g/L
Magnesium Sulphate	$MgSO_4$	1.2 g/L
Glucose	$C_6H_{12}O_6$	3.6 g/L

2.1.1.2. Selective LB Agar plates

Agar plates were made using 15 g/L LB powder, 15 g/L agar and 10 g/L NaCl made up to the relevant volume with DI water. The mixture was autoclaved in a portable autoclave at 121 °C for 15 minutes (Dixons STXX28) and allowed to cool to 50 °C before the appropriate antibiotic was added. Petri dishes were filled with approximately 30 ml of the agar/antibiotic mix and allowed to set under aseptic conditions.

2.1.1.3. Antibiotics

Kanamycin sulphate and ampicillin sodium salt (GBiosciences) were made into stock solutions of 100 mg/ml; working concentrations of ampicillin was 100 µg/ml and kanamycin was 25 µg/ml. The relevant antibiotic was dissolved in DI water and sterilised using a 0.22 µm filter (Whatman puradisk 25AS).

2.1.1.4. Bacterial Strains

E. coli strains used for general cloning and protein expression are described in **Table 3**

Table 3: E. coli strains used for cloning and recombinant protein expression.

Strain	Genotype	Supplier
DH5 alpha competent *E. coli*	*huA2 (argF-lacZ) U169 phoA glnV44 80 (lacZ) M15 gyrA96 recA1 relA1 endA1 thi-1 hsdR17*	New England Biolabs
E. coli BL21 (DE3)	fhuA2 [lon] ompT gal (λ DE3) [dcm] ΔhsdS λ DE3 = λ sBamHIo ΔEcoRI-B int::(lacI::PlacUV5::T7 gene1) i21 Δnin5 Kan^R (30 µg/ml)	New England Biolabs
E. coli BL21 (DE3) pLysS	miniF lysY (CamR) / fhuA2 lacZ::T7 gene1 [lon] ompT gal sulA11 R (mcr-73::miniTn10--TetS) 2 [dcm] R (zgb-210::Tn10--TetS) endA1 D(mcrC-mrr)114::IS10	New England Biolabs
E. coli Rosetta-gami (DE3)	fhuA2 lacZ::T7 gene1 [lon] ompT ahpC gal λatt::pNEB3-r1-cDsbC (SpecR, lacIq) ΔtrxB sulA11 R(mcr-73::miniTn10—TetS)2 [dcm] R(zgb-210::Tn10 –TetS) endA1 Δgor Δ(mcrC-mrr)114::IS10	New England Biolabs
E. coli JM109 (DE3)	endA1, recA1, gyrA96, thi, hsdR17 (rk–, mk+), relA1, supE44, λ–, Δ(lac-proAB), [F´, traD36, proAB, lacIqZΔM15], lDE3.	Promega / Manchester Institute of Biotechnology
E. coli JM109 (DE3) chaperone 4	endA1, recA1, gyrA96, thi, hsdR17 (rk–, mk+), relA1, supE44, λ–, Δ(lac-proAB), [F´, traD36, proAB, lacIqZΔM15], lDE3.	Promega / Manchester Institute of Biotechnology

NOTE: JM109 (DE3) chaperone 3 is a transformed JM109(DE3) strain (Promega) with chaperone construct pKJE7 added and JM109 (DE3) chaperone 4 is a transformed JM109 DE3 (Promega) strain with chaperone construct pG-Tf2 added (99).

2.1.1.5. Plasmid vectors

Modified pHIS (N terminal 6x Histidine), pHis trx (N terminal 6x Histidine + thioredoxin), pHIS NusA (N terminal 6x Histidine + N-utilization substance) plasmid vectors were kindly donated from Manchester Institute of Biotechnology (MIB). The pHis, pHis Trx and pHis NusA are derivatives of the pET-15b (Novagen) vector which was modified to allow easy cloning. They all have the same basic pHIS backbone (**Figure 7A)** with a simplified multiple cloning site (MCS) to only contain BamHI and EcoRI sites flanking the target protein. pHis Trx and pHis NusA (**Figure 7B** and **7C** respectively) have integrated solubility fusions which have been shown in the literature

to be beneficial in expressing soluble and correctly folded proteins (100). These expression vectors provided

the backbone to which the HepI gene was inserted to create a library of expression vectors.
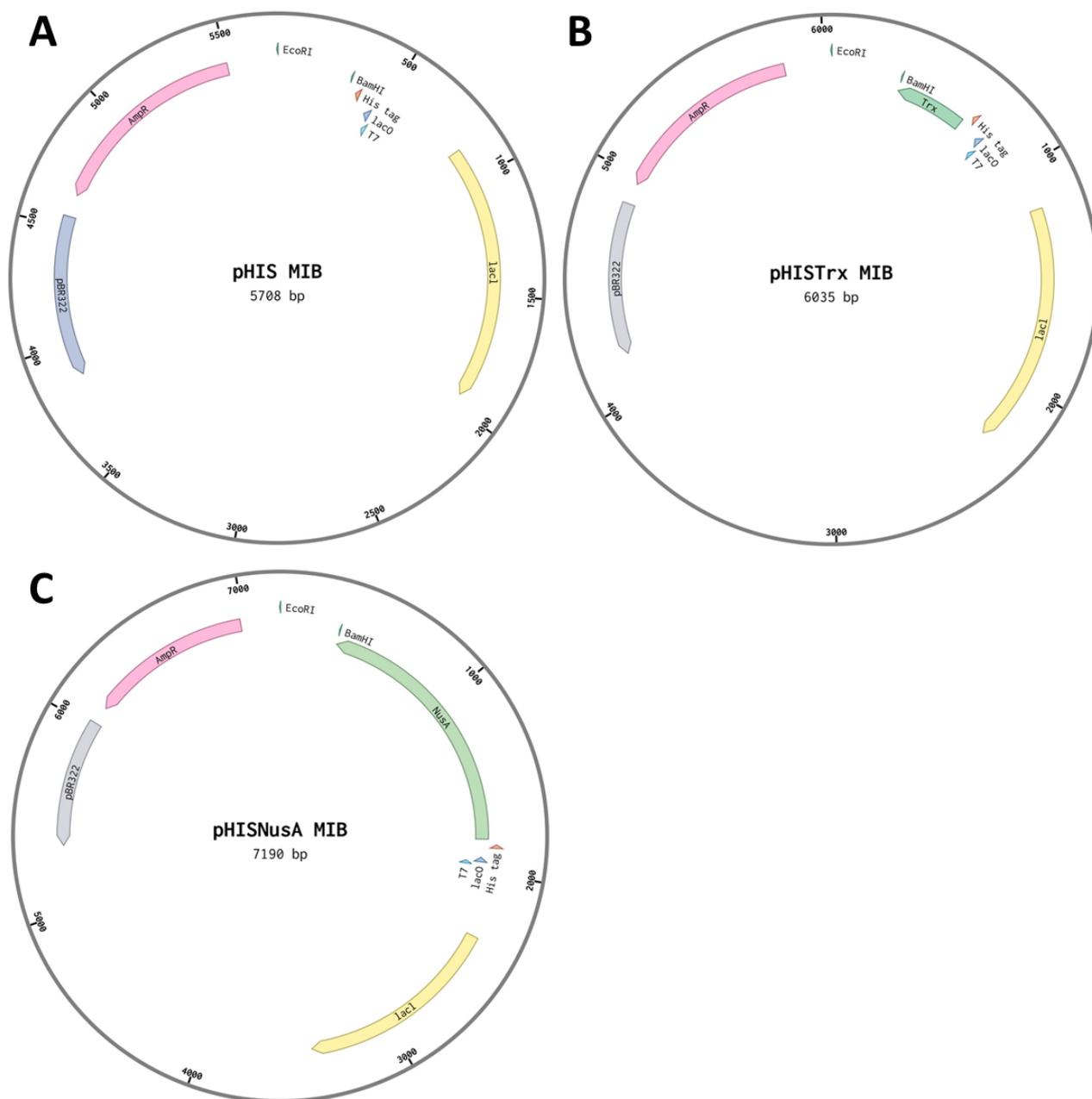
Figure 7: Map of pHIS MIB expression plasmids. *Key features include a lacI ORF for encoding lac repressor (yellow), an IPTG-inducible T7 promotor region (light blue arrow), lac operator (lacO) for lac repressor binding (dark blue), ampicillin resistance gene (pink), pBR122 (colEI) origin of replication (purple) and a His-tag coding sequence (peach arrow). All plasmids have a pET-15b (Novagen) backbone with a modified multiple cloning site (MCS) to only include BamHI and EcoRI restriction sites (green arrow). A. pHIS MIB is a basic expression plasmid to express target protein with N-terminal His-tag. B. pHISTrx MIB expression plasmid to express target protein with an N-terminal His-tag and thioredoxin fusion (green). C pHISNusA MIB expression plasmid to express target protein with an N-terminal His-tag and N utilising protein A fusion (green).*

Champion SUMO pET expression vector was purchased from Thermo Fisher, a plasmid map with key features

is shown in **Figure 8B**. The Heparinase I gene (minus leader sequence) was supplied by Integrated DNA

Technologies (IDT) and was previously sub-cloned in to a pET28a expression vector. The plasmid map for the

pET28a vector (**Figure 8A**).
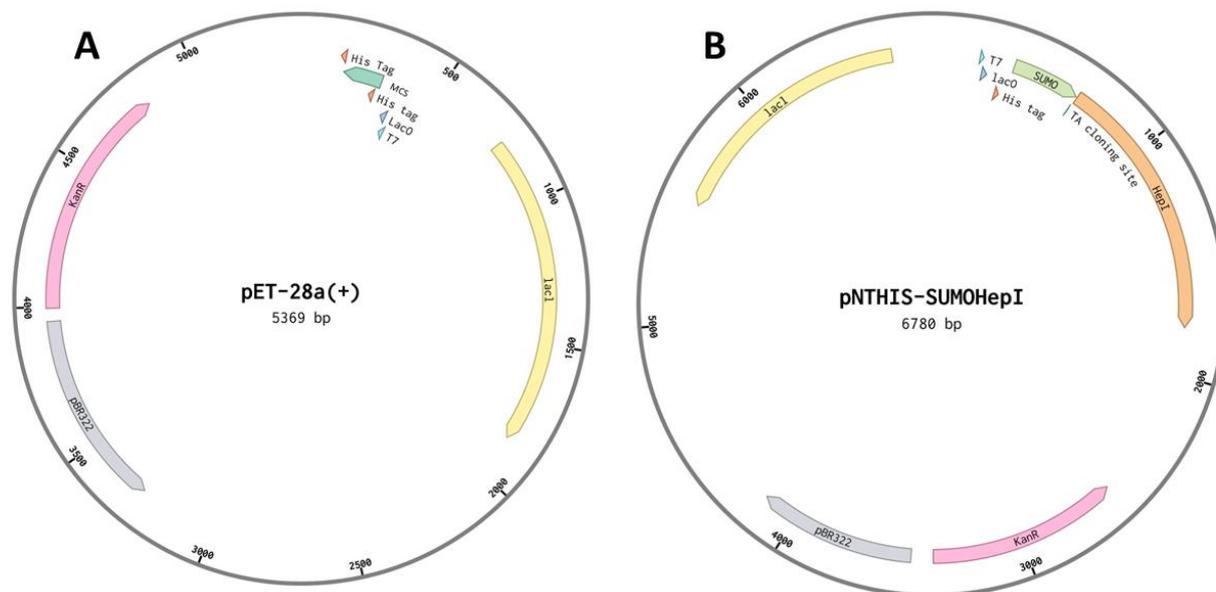
Figure 8: Map of pET28a and pET SUMO expression vectors. *Key features include a lacl ORF for encoding lac repressor (yellow), an IPTG-inducible T7 promotor region (light blue arrow), lac operator (lacO) for lac repressor binding (dark blue), kanamycin resistance gene (pink), pBR122 (colEI) origin of replication (purple) and a His-tag coding sequence (peach arrow). A. pET28a allows for the fusion of either an N or C terminal His-tag to the target protein and an MCS (green). B. pET SUMO allows for the fusion of SUMO protein to the target protein to increase soluble expression (light green), TA cloning of taq-amplified PCR insert (dark green) and an N terminal His-tag.*

2.1.2. Methods

2.1.2.1. Molecular Biology techniques

The native heparinase I gene sequence (detailed in Chapter 3), includes a 21 amino acid signal peptide which

is highly hydrophobic, this signal peptide transports the protein to the periplasm in the native host and is

cleaved prior to secretion and not essential for activity (90). Interestingly, the leader sequence is missing from

Heparinase I when expressed in other native hosts such as *Bacteroides thetaiotaomicron* (101). Due to this

hydrophobicity, the leader sequence was omitted from the gene sequence of the recombinant heparinase I

constructs for soluble expression and read Glu, Glu, Lys, lys, Ser…. The fusion constructs which were designed,

synthesised, and cloned in this project (**Figure 9)** and show the position of the His-tag, HepI protein and any

fusion proteins to aid in soluble expression.

5
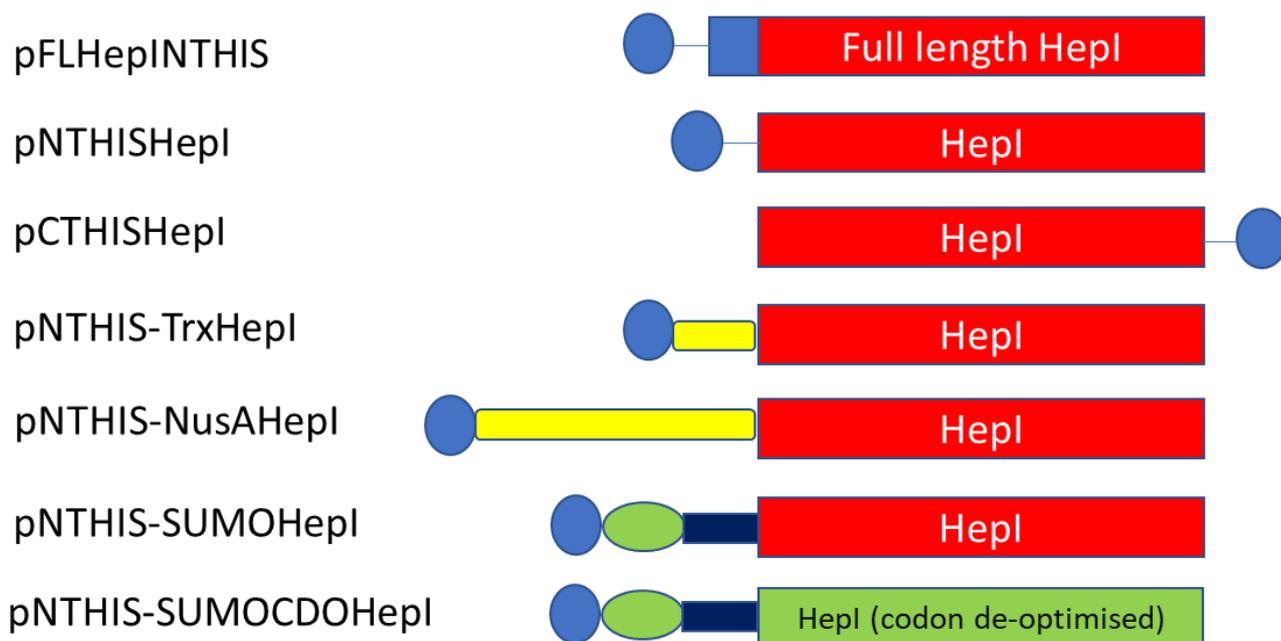
Figure 9: Plasmid constructs used to enhance soluble expression of recombinant heparinase I in *E. coli* (not to scale). *pFLHepINTHIS included an N-terminus 6X His-tag (blue circle) and the 29aa hydrophobic leader sequence (blue square), which was omitted from all other constructs for soluble expression. pNTHISHepI included an N-terminus 6X His-tag (blue circle). pCTHISHepI included a C-terminus 6X His-tag (blue circle). pNTHIS-TrxHepI included N-terminus 6X His-tag (blue circle) and thioredoxin (Trx) (yellow rectangle) solubility enhancing tag. pNTHIS-NusAHepI included N-terminus 6X His-tag (blue circle) and N-utilisation protein A (NusA) (long yellow rectangle) solubility enhancing tag. pNTHIS-SUMOHepI included N-terminus 6X His-tag (blue circle) small ubiquitin like modifier (SUMO) solubility enhancing tag (green oval) with a G4S linker and enterokinase cleavage site (dark blue rectangle). pNTHIS-SUMOCDOHepI was created whereby the codon sequence of the gene was de-optimised for expression in E. coli to act like a translation brake and included N-terminus 6X His-tag (blue circle) small ubiquitin like modifier (SUMO) solubility enhancing tag (green oval) with a G4S linker and enterokinase cleavage site (dark blue rectangle).*

2.1.2.1.1. Polymerase Chain Reaction (PCR)

The native heparinase I gene sequence (minus leader sequence) was synthesised and supplied in a pUC vector (Integrated DNA Technologies), this was used as a DNA template for all vector constructs, apart from the pFLHepINTHIS construct. In addition, the native full-length heparinase I gene sequence was synthesised and supplied in separate pUC vector (Integrated DNA Technologies), this was engineered to be cloned into a pHISMIB vector (MIB modified plasmid). PCR forward and reverse primers were created for cloning into the different expression vectors. All oligonucleotides used in this study are listed in **Table 4.**

Table 4: Cloning primers.

Primer name[1]	Sequence (5'-3')[2]
CTHISHepI F	CACCCATGGAA**CAGCAAAAAAAATCCGGTAACATC**
CTHISHepI R	CACCTCGAG**CTATCTGGCAGTTTCGCTGTAC**
NTHIS-TrxHepI F	CACAGATCT**CAGCAAAAAAAATCCGGTAACATC**
NTHIS-TrxHepI R	CACGAATTC**TTATTACTATCTGGCAGTTTCGCTGTAC**
NTHIS-NusAHepI F	CACAGATCT**CAGCAAAAAAAATCCGGTAACATC**
NTHIS-NusAHepI R	CACGAATTC**TTATTACTATCTGGCAGTTTCGCTGTAC**
NTHIS-SUMOHepI F	GGCGGCGGGGGCTCTGACGATGACGATAAA**CAGCAAAAAAAATCCGGTAACATC**
NTHIS-SUMOHepI R	**TTATTACTATCTGGCAGTTTCGCTGTAC**
NTHIS-SUMOCDOHepI F	GGCGGCGGGGGCTCTGACGATGACGATAAA**CAGCAGAAAAAAAGCGGGAACATTC**
NTHIS-SUMOCDOHepI R	**TTATTACCGGGCAGTTTCGGAGTAACC**

[1]Primer names are given to match with the construct design, N/C denotes the terminal position of the tags, if a solubility enhancing tag is used then it is stated after the dash. F indicates that the primer anneals at the 5' terminus of the template DNA and R indicates that the primer anneals at the 3' terminus of the template DNA.

[2]The region where the primer anneals to the template DNA is indicated in boldface. Blue bases indicate the G4S linker and green bases indicate the enterokinase cleavage site.

Standard PCR reaction mix: PCR reaction mix was made up as per NEB Q5 High Fidelity polymerase instructions (dNTP master mix, 100 pM forward and reverse primer, 55 ng DNA template, PCR Buffer 5X, Q5 Polymerase) and DI water to a final volume of 50 µl. Negative control reactions were made up by substituting water for either the forward or reverse primer.

SUMO PCR reaction mix: PCR reaction mix was made up as per Invitrogen Platinum *Taq*DNA Polymerase instructions (1.5 mM Mg, dNTP master mix, 200 pM forward and reverse primer, 55 ng DNA, PCR Buffer 10X, Platinum Polymerase), and DI water to a final volume of 50 µl. Negative control reactions were made up by substituting either the forward or reverse primer for DI water.

An Eppendorf MasterCycler Personal (5332) PCR machine was used for DNA amplification and set to the programme outlined in **Figure 10.**
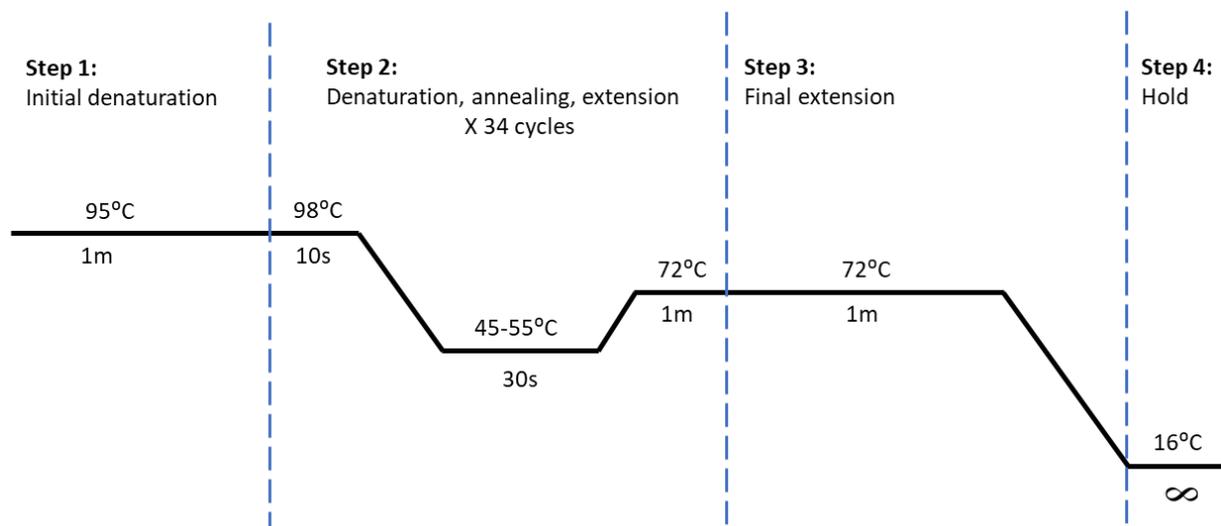
Figure 10: PCR programme: *Step 1 Initial denaturation at 95 °C for one minute, followed by Step 2, 34 cycles of denaturation at 98 °C for 10 seconds, annealing at 45-55 °C (depending on T$_m$ of oligos) for 30 seconds and extension at 72 °C for one minute. Step 3 final extension at 72 °C for one minute Step 4 Hold reactions at 16 °C at the end of the programme indefinitely.*

PCR Purification: PCR reactions had loading buffer added and were purified using agarose gel electrophoresis. The DNA band was excised from the gel using a sterile blade and DNA was recovered from the gel slice using a Qiaquick gel extraction kit (Qiagen) according to the manufacturer's instructions. Briefly, the gel was dissolved in a chaotropic buffer (5.5 M guanidine thiocyanate (GuSCN), 20 mM Tris HCl pH 6.6), and the mixture was then added to a column containing a silica membrane which is specifically designed to bind DNA fragments, impurities were washed away, and the purified DNA was eluted in 40 µl of a low salt buffer (10 mM Tris-HCl pH 8.5) and stored at -20 °C.

2.1.2.1.2. Agarose Gel Electrophoresis

TAE buffer (40 mM Tris, 20 mM Glacial acetic acid, 11 mM EDTA) had 1 % agarose added. The agarose and TAE mixture was microwaved in 10 second bursts with gentle swirling in between until the agarose was fully dissolved (usually within two minutes). The mixture cooled to 50 °C and nucleic acid dye (GelGreen®, Biotium) was added in a ratio of stain to buffer of 1:10,000 and thoroughly (but gently to avoid bubble formation) mixed. The mixture was poured into a mould, a gel comb was added and was then left to set on the benchtop before the comb was removed. DNA samples were prepared by mixing in a ratio of 10:1 DNA with loading buffer. Samples and DNA markers (NEB/Cambridge Biosciences) were loaded on to the gel (5 µl DNA marker and 10 µl sample per single well, and up to 40 µl sample per double well). DNA samples were electrophoresed at 80v

for one hour in 1X TAE buffer. Analytical gels were visualised and photographed in a C280 Azure imagine system and preparative gels visualised at 470 nm using a transilluminator (BluPAD, Bio-Helix). The DNA band was excised from the gel using a sterile blade and DNA was recovered from the gel slice.

2.1.2.1.3. Restriction Digestion

Restriction enzyme compatibility and reaction times were checked using the Thermo Fisher guide to FastDigest restriction enzymes (**https://bit.ly/2I8kyD6**). Restriction enzyme combinations used in this project are outlined in **Table 5**.

Table 5: Restriction enzymes.

Construct	Restriction enzyme combination
pHIS MIB	BamHI + EcoR1
pHISTrx MIB	BamHI + EcoR1
pHISNusA MIB	BamHI + EcoR1
pET28a (N-term HIS)	BamHI + NdeI
pET28a (C-Term HIS)	NcoI + XhoI

Double digests were carried out in 50 µl total volume with FastDigest Green Buffer (10X), FastDigest enzyme and either 1 µg plasmid DNA in no more than 2 µl or 0.2 µg PCR product in no more than 10 µl as per Thermo Fisher FastDigest protocol (**https://bit.ly/38QOmj9**). The reaction volume was made up to 50 µl with DI water and heated to 37 °C for five minutes typically (but this was enzyme dependent) and the reaction was terminated by heat inactivation for 5 minutes at 65 °C typically (but this was enzyme dependant). Reactions involving plasmid vectors had 1 U/µg of thermosensitive alkaline phosphatase (FastAP, ThermoFisher) added to dephosphorylate the newly created sticky ends. As per the manufacturer's instructions, the vectors were incubated with FastAP for 15 minutes at 37 °C, after which the FastAP was heat inactivated at 75 °C for five minutes.

2.1.2.1.4. Ligation

Phosphodiester bonds between DNA fragments to be joined were formed using a Rapid DNA ligation kit (ThermoFisher) containing T4 DNA ligase. Reactions were carried out in 20 µl volumes with insert DNA at 3:1 molar excess over vector which was calculated using the formula in **Figure 11**. Reactions were incubated at

9

room temperature for 15 minutes. The reaction mixture consisted of doubly digested, linearised, dephosphorylated vector, doubly digested DNA insert, T4 ligase master mix and DI water.

$$M_i = I{:}V \text{ molar excess} * (L_i/L_v)*M_v$$

Figure 11: An equation for determining the volumes required in a ligation reaction. *M_I Mass of insert in ng, I:V insert to vector ratio , typically 3 , L_I Length of insert in kb, L_v Length of vector in kb, M_v Mass of vector in kb.*

2.1.2.1.5. Plasmid DNA Extraction and Analysis

Plasmid DNA was extracted from overnight mini cultures. After overnight growth, mini cultures were centrifuged at 3400 g for 20 minutes to pellet cells (Centurion Scientific ProResearch K242R). The supernatant was decanted, and the tube inverted on to a tissue to remove any excess media. Plasmid DNA was extracted using a Quicklyse Miniprep Kit (Qiagen) as per manufacturer's instructions. Briefly the bacterial cells were lysed in complete lysis solution, and the mixture centrifuged at 17,000 g for one minute at room temperature (Eppendorf 5415D), the supernatant was added to a Quicklyse spin column containing a silica membrane which bound the plasmid DNA, impurities were washed away using the isopropanol-based buffer, QLW, and the purified plasmid DNA was eluted in low salt, QLE buffer (10 mM Tris-Cl and 0.1 mM EDTA, pH 8.5) and stored at -20 °C. DNA concentration and purity were determined by UV spectrophotometry at 260 nm and 280 nm wavelengths (Nanodrop Lite, ThermoFisher).

2.1.2.1.6. Plasmid sequencing

Plasmid DNA was doubly digested and purified using agarose gel electrophoresis. Plasmid samples which showed bands indicative in size to the vector and gene insert were selected for sequence verification. Sanger sequencing was performed by GATC Biotech and samples were prepared for sequencing by adding 4-500 ng DNA to a micro-centrifuge tube along with vector appropriate primer (forward and reverse at 20 μM) and made up to 10 μl with PCR water.

2.1.2.1.7. Bioinformatics

Bioinformatics tools were used in this project; Basic Local Alignment Tool (BLAST) was used to search for similarities between biological (nucleotide or protein) sequences (102). Benchling's molecular biology suite

was used to create vector maps (103). EMBOSS Needle was used or pairwise sequence alignment of amino acid sequences (standard parameters used) (104). EMBOSS sixpack was used to display DNA sequence with 6-frame translation and ORF's (standard parameters used) (105). Additionally, Expasy's translation tool was used for the translation of DNA sequence to protein sequence (standard parameters used) (106). Expasy's Swiss-Model tool was used to predict the 3D structure of the recombinant HepI protein and to show the position of Ser[39], Cys[297] and calcium binding site (**http://www.sbg.bio.ic.ac.uk/phyre2/html/page.cgi?id=index**).

2.1.2.2. Protein expression and analysis

Protein was expressed in a variety of *E. coli* hosts under different expression conditions, initially to maximise soluble expression and later to maximise insoluble expression of protein. Expression cells were transformed fresh each time with one of the heparinase I vectors and start up cultures were made.

2.1.2.2.1. Transformations

Aliquots (50 µl) of competent cells were thawed on ice, to which 50 ng of plasmid DNA was added under aseptic conditions. The DNA and cells were incubated on ice for 30 minutes, heat shocked for 30 seconds at 42 °C and placed back on ice for five minutes to allow for uptake of the plasmid DNA. The cells had 950 µl room temperature SOC outgrowth media (NEB) added and recovered to develop the relevant antibiotic resistance genes by incubating at 37 °C for one hour 200 rpm (Stuart S1500 orbital incubator). Selective LB agar plates were pre-warmed to 37 °C and 100 µl of the transformed cells was spread on to them. The plates were incubated overnight at 30 °C. Overnight colonies were used for mini-culture preparation immediately or stored for no more than one week at 4 °C.

2.1.2.2.2. Mini Cultures

A single colony was picked using a sterile toothpick from a LB agar plate and added to a mini bioreactor centrifuge tube (Corning) containing 5 ml of selective LB at 37 °C under aseptic conditions. The mixture was incubated overnight at 37 °C 200 rpm (Stuart S1500 orbital incubator). Mini cultures were used in two ways; 1, for plasmid extraction and DNA sequencing and 2, to inoculate large volumes of LB media for protein expression.

2.1.2.2.3. SDS-PAGE and western blot analysis of protein

For SDS-PAGE analysis of proteins, running buffer was made up by diluting 10X SDS-PAGE stock (250 mM Tris-base, 1.9 M glycine, 1 % (w/v) SDS) to working concentration with DI water. Samples were collected for SDS-PAGE and western blot analysis and labelled pre induced (PI), total protein (T), soluble protein after lysis (S) and insoluble protein after lysis (P). Prior to induction of the expression cultures, 1 ml was separated from the main culture and then left to grow under the same conditions as the induced cultures. No IPTG was added to the PI cultures. The PI samples were clarified at full speed, for one minute (Eppendorf 5415D micro-centrifuge) and resuspended in 200 µl lysis buffer and then diluted 1:20 with Laemmli buffer (BioRAD). After induction and lysis of the expression cultures, 50 µl of the cell lysate was removed and clarified, 5 µl of the supernatant was added to 95 µl of Laemmli buffer for the soluble fraction. The cell pellet was resuspended in 1 ml Laemmli buffer for the insoluble fraction. All protein samples were heated to 95 °C for five-seven minutes and 10 µl of each sample was loaded on to a 4-12 % pre-cast polyacrylamide gel housed inside of a tetra cell (Mini-PROTEAN system, BioRAD). A full range rainbow molecular weight marker (Amersham ECL) was also added to identify the molecular weight of the protein bands. The gel was covered in running buffer (25 mM Tris-base, 190 mM glycine, 0.1 % (w/v) SDS) and the protein was electrophoresed for one hour at 115 V. Gels were stained for one hour (Instantblue, Expedeon) and protein bands visualised and captured (C280 Azure imagine system).

For western blot analysis the proteins from the polyacrylamide gel were transferred to a polyvinylidene fluoride (PVDF) membrane using a semi-dry western transfer system (Cleaver Scientific SD10). The PVDF membrane was soaked in 100 % methanol to activate prior to transfer, and filter papers soaked in transfer buffer (192 mM glycine, 25 mM tris base, 0.037 % SDS, 20 % methanol, pH 8.3). The PVDF membrane and polyacrylamide gel were sandwiched between wet (but not soaking) filter papers and proteins were transferred for one hour at 45 V. A small amount of ponceau red (Sigma) was placed on to the membrane as a rapid but reversable way to determine successful transfer of proteins. After transfer, the membrane was placed in blocking buffer (150 mM NaCl, 20 mM tris base, 50 mM KCL, 0.2 % tween, 5 % milk powder, pH 7.6) for one hour at room temperature. Primary antibody, anti-His monoclonal raised in mouse (Proteintech) was added to fresh, cold, blocking solution (1:10,000) and the membrane incubated overnight at 4 °C. The next day the membrane was placed in fresh blocking solution containing secondary antibody, anti-mouse raised in

rabbit, conjugated to horseradish peroxidase (HRP) (Proteintech, 1:10,000) and left to incubate at room temperature for one hour. The membrane was washed three times in TBST buffer (150 mM NaCl, 20 mM tris base, 50 mM KCL, 0.2% tween, pH 7.6), and then coated in peroxidase and HRP substrate for two minutes at room temperature (Radience). Chemiluminescence visualisation and capture of proteins of interest was carried out in a C280 Azure imagine system.

2.1.2.2.4. Soluble protein expression

The OD_{600} of mini cultures was measured (Amersham Ultrospec 10 cell density meter), diluted to OD_{600} 0.2 in selective LB and incubated at 37 °C 200 rpm until OD_{600} reached 0.6. Cultures were chilled on ice to reach 18 °C and were induced with 0.2 mM Isopropyl β-D-1-thiogalactopyranoside (IPTG) overnight at 18 °C 200 rpm. Overnight cultures were clarified at 4 °C, 3400 g for 20 minutes (Centurion Scientific ProResearch K242R). The cell pellet was resuspended in five times reduced culture volume of ice-cold lysis buffer (300 mM NaCl, 25 mM Tris-HCl, 1 % NP40, 2 mM Ca^{2+}, 0.1 mM DTT, added fresh just before use, and protease inhibitor cocktail 1:500, pH 7.9) and sonicated for four minutes (Branson 250 sonifier cell disruptor with microtip, 30 % duty cycle, 20 % output). The protease inhibitor cocktail (PIC, Sigma) is specifically formulated to for isolating HIS-tagged proteins from E. *coli* cells; the components and their properties are outlined in **Table 6**. The cell lysate was clarified at 4 °C, 3400 g for 20 minutes. A sample of the supernatant was taken for SDS-PAGE, western blot, and activity analysis. The soluble fraction from the cell lysate was snap frozen in liquid nitrogen and stored at -80 °C.

Table 6: Components of Protease Inhibitor Cocktail.

Component name	Properties	Reversible / irreversible reaction	Reference
AEBSF 4-(2-Aminoethyl) benzenesulfonyl fluoride hydrochloride	Serine protease inhibitor	Irreversible	(107)
Bestatin (ubenimex	Aminopeptidase and enkephalinase inhibitor	Reversible	(108)
E-64 (epoxysuccinyl based inhibitor)	Epoxide which inhibits cystine peptidases and acid proteases	Irreversible	(109)
Pepstatin	Aspartyl protease inhibitor	Reversible	(110)
Phosphoramidon	Thermolysin and metallo-endoproteinase inhibitor	Reversible	(111)

2.1.2.2.5. Insoluble protein expression

Mini cultures were diluted to OD_{600} 0.2 in selective AI media and grown overnight at 30 °C 200 rpm. Overnight cultures were handled as described in section 2.1.2.2.2 up until the second clarification stage. At this point the supernatant was discarded and the insoluble inclusion bodies were washed with TBS (50 mM Tris, 150 mM NaCl, 5 mM $CaCl_2$ pH 7.2). Samples were taken for SDS-PAGE and western blot analysis. The inclusion bodies were snap frozen in liquid nitrogen and stored at -80 °C

2.1.2.3. Inclusion body processing

To recover a functionally active protein inclusion bodies were subjected to three refolding protocols; 1. Using a commercial refolding kit (Pierce Protein, Thermo Fisher), 2. A 'fast' refolding protocol, and 3. A 'slow' refolding protocol. A brief overview of each refolding process is outlined in **Figure 12**. Prior to refolding all inclusion bodies were resuspended in the same volume as lysis buffer of wash buffer (500 mM NaCl, 50 mM Tris-HCL, 2 % NP40. pH 7.2), clarified at 3400 g and the supernatant removed two times.
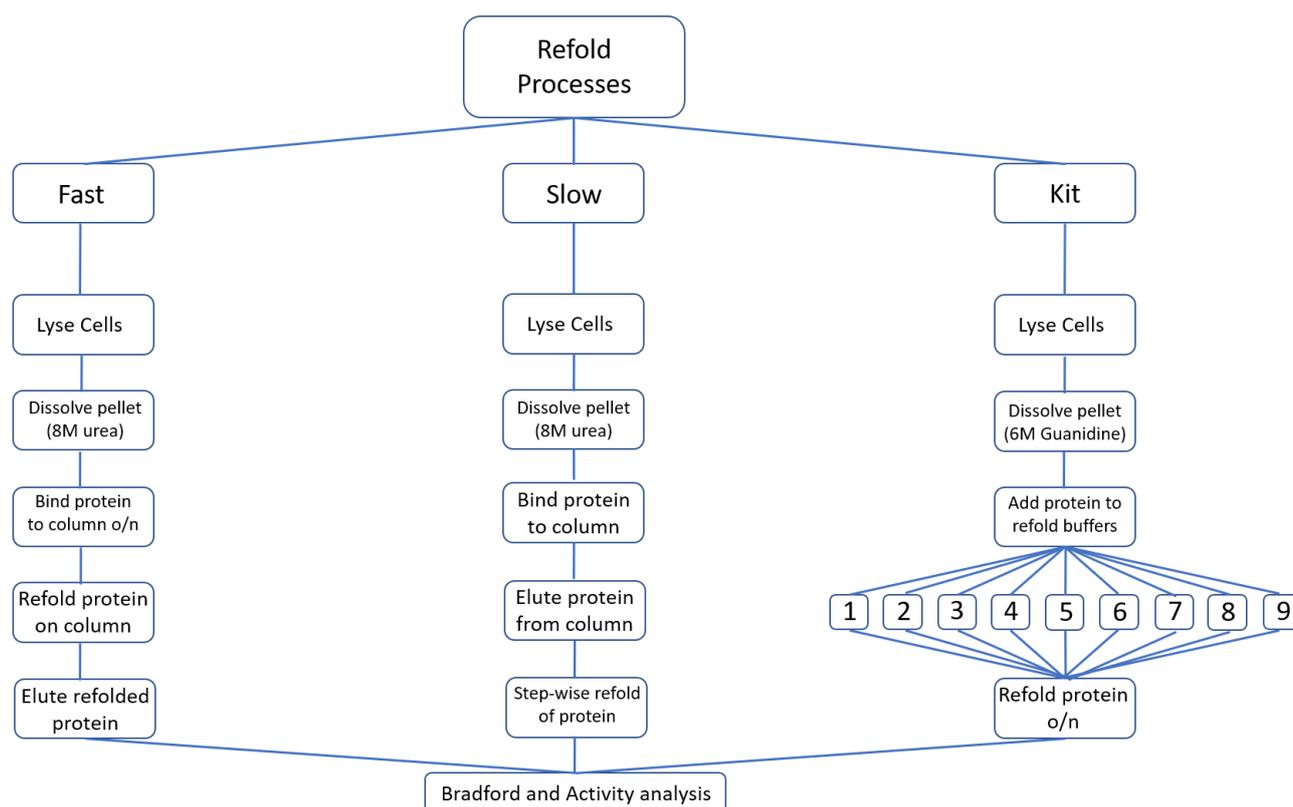
Figure 12: Outline of refolding processes. *Inclusion bodies were subjected to three different refolding protocols' Fast, Slow and using a commercial kit. The Fast process involved solubilising the inclusion bodies in 8 M urea, binding to either nickel coated beads or to a nickel FPLC column, washing with a refolding buffer followed elution of the refolded protein. The Slow process involved solubilising the inclusion bodies in 8 M urea, binding to either nickel coated beads or to a nickel FPLC column, elution of the protein under 8 M urea, followed by a step wise refold of the protein by dialysis. The commercial kit process involved dissolving the inclusion bodies in 6 M Guanidine then and then adding the protein to one of nine refolding buffers. Proteins were refolded overnight and dialysed to remove denaturants/reducing/oxidising reagents. All refolded protein was then subjected to Bradford (for protein concentration) and kinetic (activity) analysis.*

2.1.2.3.1. Fast refolding protocol

Inclusion bodies were maintained at 4 °C throughout the protocol and all buffers were pH 7.9. Inclusion bodies were solubilised in urea buffer (8 M urea, 500 mM NaCl, 20 mM Tris-HCL, 5 mM imidazole). Commercial nickel coated magnetic SepFast beads (Biotoolomics) were equilibrated with urea buffer and dried. Solubilised protein was added to the beads (200 µl beads per 1 ml solubilised protein). Samples were bound to the beads overnight at 4 °C, the supernatant was removed, and the beads washed twice (500 mM NaCl, 20 mM Tris-HCL, 60 mM Imidazole). The supernatant was removed, and the beads were washed with refolding buffer (500 mM NaCl, 20 mM Tris-HCL, 5 mM Imidazole) once and wash buffer twice. The protein was eluted (500 mM NaCl, 20 mM Tris-HCL, 1M Imidazole) and passed through a SuperSpin de-salting column (Biotoolomics) to remove the imidazole. Activity was tested and samples were snap frozen in liquid nitrogen and stored at -80 °C.

2.1.2.3.2. Slow refolding protocol

Inclusion bodies were maintained at 4 °C throughout the protocol and all buffers were pH 7.2. Inclusion bodies were solubilised in five times less urea buffer (8 M urea, 500 mM NaCl, 20 mM Tris-HCL, 5 mM Imidazole) than starting culture volume. The solubilised protein was added to either equilibrated commercial nickel coated magnetic SepFast beads (Biotoolomics) or an AKTA Pure FPLC with a pre-packed Ni SepFast 10 ml column (Biotoolomics) for refolding and purification. The protein was bound to the beads overnight at 4 °C in urea buffer (8 M urea, 500 mM NaCl, 20 mM Tris-HCL, 5 mM Imidazole) and eluted (8 M Urea, 500 mM NaCl, 20 mM Tris-HCL, 1 M Imidazole) and added to Pur-A-Lyzer dialysis cartridges (Merck, MWCO 3.5 kDa). FPLC programme is outlined in **Table 7**, eluted protein was collected over three fractions (4 ml each) and was added to pre-equilibrated (urea buffer) spectral POR3 dialysis tubing (Spectrum Labs, MWCO 3 kDa).

In both the bead and FPLC protocol the protein was dialysed in 100 times dialysis buffer (150 mM NaCl, 50 mM tris, 5 mM $CaCl_2$) + 4 M Urea for four hours, then + 2 M urea overnight, then + 1 M urea for four hours, and finally, overnight in dialysis buffer only.

After dialysis, the protein was tested for activity, snap frozen in liquid nitrogen and stored at -80 °C.

Table 7: AKTA Pure FPLC programme for slow refold of HepI inclusion bodies.

Buffers:	Composition					
A1	8 M urea, 150 mM NaCl, 50 mM tris, 5 mM $CaCl_2$					
A2	8 M urea, 150 mM NaCl, 50 mM tris, 10 mM imidazole, 5 mM $CaCl_2$					
A6	H_2O					
A7	20 % ethanol					
B1	8 M urea, 150 mM NaCl, 50 mM tris, 1M Imidazole, 5 mM $CaCl_2$					
Column:	Pre-packed Ni SepFast 10 ml					
Step	Procedure	Line	B1 (%)	Imidazole (mM)	Volume (ml)	Outlet
1	Column Wash	A1	0	0	20	Waste
2	Equilibration	A1	0.5	5	5	Waste
3	Sample application	Sample pump	0	0	10 x 20 cycles	Outlet 1 to sample vessel
4	Column wash	A2	0	10	100	Fraction collector
5	Elution	B1	100	1000	16	Fraction collector (4 ml)
6	Column wash	A6	0	0	100	Waste
7	Column cleaning and storage	A7	0	0	100	Waste

2.1.2.3.3. Kit refolding protocol

The kit comes with nine refolding buffers with seven buffer additives at three concentration levels to provide a matrix strategy for refolding recombinant proteins. The base refolding buffers provide different denaturing conditions to supress protein aggregation. The refolding kit is based on a fractional factorial matrix, factors one and two are defined by the buffers and factors three and four are defined by the user depending on the required refolding environment.

The final concentrations of components in each reaction are shown in **Table 8**, taking into consideration that the protein had been dissolved in 6 M guanidine as the kit suggests. The kit's recommended protocol for HIS tagged protein was followed. In brief, inclusion bodies were solubilised in guanidine buffer (6 M Guanidine HCL, 150 mM NaCl, 50 mM Tris-HCL, 5 mM $CaCl_2$. pH 7.2) for two hours at room temperature. The dissolved (2 mg in 25 µl) protein was then added to nine different refolding buffers (1:10) and refolded overnight at 4 °C. After refolding all samples were added to pre-equilibrated dialysis tubing and dialysed in 100 times dialysis buffer (150 mM NaCl, 50 mM tris, 5 mM $CaCl_2$. pH 7.2) for 24 hours. After dialysis protein samples were tested for activity.

Table 8: Composition of Pierce Protein Refolding Kit buffers (final volume of 500 μl). *Green text denotes a high concentration and red text denotes a low concentration as defined by the manufacturer.*

Refolding Buffer	Final concentration (mM)									
	Tris	NaCl	KCL	L-arginine	Guanidine	GSH	GSSG	NP40	DTT	[Protein]
1	49.50	18.90	0.8	-	-	-	-	-	5	2 mg/ml
2	49.50	18.90	0.8	396	-	1.95	0.2	0.05	-	2 mg/ml
3	49.50	18.90	0.8	792	-	1.95	0.4	0.2	-	2 mg/ml
4	49.50	18.90	0.8	-	795	1.95	0.2	0.2	-	2 mg/ml
5	49.50	18.90	0.8	396	795	1.95	0.4	-	-	2 mg/ml
6	49.50	18.90	0.8	792	795	-	-	0.05	5	2 mg/ml
7	49.50	18.90	0.8	-	1290	1.95	0.4	0.05		2 mg/ml
8	49.50	18.90	0.8	396	1290	-	-	0.2	5	2 mg/ml
9	49.50	18.90	0.8	792	1290	1.95	0.2	-	-	2 mg/ml

2.1.2.4. Protein purification

Small-scale purification trials were conducted using nickel coated magnetic SepFast Beads (Biotoolomics) and large-scale purification trials conducted on an AKTA Pure FPLC (GE healthcare). Purification trials were carried out on the soluble fraction from bacterial cell lysate, small-scale purification was carried out using soluble protein samples representing 5 ml of bacterial culture and large-scale purification was carried out using the soluble lysate from 1 L bacterial culture.

2.1.2.4.1. Small-scale purification

Nickel coated beads were washed twice in 10 times volume of equilibration buffer (50 mM Tris 500 mM NaCl, pH 7.4) and the supernatant removed between washes. Soluble protein samples (1 ml per 50 μL beads) were incubated for 1.5 hours at room temperature, and the supernatant removed. The beads were washed in 10 times volume wash buffer 1 (equilibration buffer + 50 mM imidazole) and the supernatant removed. The beads were washed in wash buffer 2 (equilibration buffer + 100 mM imidazole) and the supernatant removed. The bound protein was eluted using elution buffer (equilibration buffer + 500 mM imidazole) at 0.25 volumes

(compared to original sample). Unbound supernatant, wash supernatants, and eluted protein samples were taken for SDS-PAGE and western blot analysis.

2.1.2.4.2. Large-Scale purification

The soluble fraction from 1 L bacterial culture was loaded on to an AKTA Pure FPLC with a HisTrap crude #11-0004-58 1 mL column via the sample pump. The FPLC programme is outlined in **Table 9.** Samples were collected from the flow through (step 3), after column washes (steps 4-8), and after elution (step 9) for SDS-PAGE analysis. Fractions identifying HepI protein were pooled, concentrated up to 1 mg/mL and 1 mL sample was passed through a gel filtration column for further purification.

Table 9: AKTA Pure FPLC programme for Immobilized Metal Affinity Chromatography (IMAC) of soluble HepI.

Buffer	Buffer composition					
A1	50 mM Tris pH 7.9, 300 mM NaCl, 1 mM DTT					
A2	A1 + 0.1 % NP40					
A6	H_2O					
A7	20 % Ethanol					
B1	A1 + 1 M Imidazole					
Flow rate	2-3 ml/min					
Step	**Procedure**	**Line**	**B1 (%)**	**Imidazole (mM)**	**Volume (ml)**	**Outlet**
1	Column Wash	A6	0	0	5	Waste
2	Equilibration	A1	0.5	5	5	Waste
3	Sample application	A2	0.5	5	40	Outlet 1
4	Column wash	A2	1	10	5	Outlet 2
5	Column wash	A2	2	20	5	Outlet 3
6	Column wash	A2	3	30	5	Outlet 4
7	Column wash	A2	4	40	5	Outlet 5
8	Column wash	A2	5	50	5	Outlet 6
9	Step elution	A2	25	250	15	Fraction collector (1 ml)
10	Column wash	A2	100	1000	5	Outlet 7
11	Column cleaning	A7	0	0	5	Waste
12	Column cleaning	A7	0	0	5	Waste

2.1.2.5. Kinetic testing of protein

Heparinase I standard enzyme (IBEX) was diluted to 0.05, 0.15, 0.3, 0.6 and 1 U/ml in dilution buffer (150 mM NaCl, 46.8 mM Tris HCL, 3.2 mM Tris Base, 5 mM $CaCl_2$, pH 7.2); 50 µl of each calibrant was added to a cuvette containing 150 µl heparin substrate (150 U/ml in dilution buffer, SIGMA) the substrate and enzyme were mixed gently and the change in absorbance over 180 seconds measured at 232 nm in a spectrophotometer (Jenway 7295 UV/VIS). From the OD measurement at time 0 s and 180 s the ΔOD/Min was calculated and inputted to

a calibration graph. Samples were then tested in the same way as each calibrant sample and the ΔOD/Min

calculation from the line equation from the calibration graph. This was then used to calculate the specific

activity of each protein sample (U/ml or U/mg).

Chapter 3. Cloning and expression trials

3.1. Introduction

Expressing heparinase I from the native host (without any modifications) has not proven to be commercially viable due to high process costs and low yield of product (90,112). Because of this, a recombinant expression system in *E. coli* was quickly developed (26). This bacterium is the most convenient choice for recombinant protein expression (113), however, when expressed in *E. coli,* heparinase I tends to form into inclusion bodies within the cell (114). To increase the soluble expression of heparinase I in *E. coli* cells, several expression constructs were designed to fuse the HepI gene with different affinity tags and stable fusion partners which have successfully been expressed in *E. coli* previously and have shown to enhance correct conformation refolding of fusion proteins during expression. The addition of a 6X Histidine 'tag' (His-tag) to the N or C terminus of a recombinant protein aids in downstream purification of the protein using immobilised metal affinity chromatography (IMAC) (115). Because His-tags are only around 2.5 kDa they are unlikely to interfere with protein folding or function (116), however this can depend on where the active site lies within the protein (117,118). In this project His-tags were placed at both the N and C termini of the HepI protein to see if activity was affected (see **Figure 9**). Co-expressing proteins with fusion partners to enhance solubility is well documented in the literature and has proven to be a useful approach in expressing recombinant heparinase I (96,97,100,119).

3.2. Vector library construction

Soluble fusion partners such as Trx, NusA or SUMO are well documented as having the ability to increase soluble expression of recombinant proteins in *E. coli* (120)*,* (121) which is why they were chosen in this study. Thioredoxin (trx) is a heat stable protein of around 11.7 kDa in length found in nearly all known organisms and is known for its antioxidant role (122). In *E. coli* it is encoded by the *trxA* gene and plays a key role in cell redox homeostasis and other metabolic processes (123). Trx was the first fusion partner to successfully increase the solubility of a target protein in *E. coli* (124)*.* During expression Trx acts as a molecular chaperone and promotes functional folding support which can increase overall yield of soluble recombinant protein expression by over three-fold. (125). N-utilising substance A (NusA) is the largest of the fusion proteins at 54.8 kDa, it is a major

transcription factor in Bacteria and Archaea and is an essential protein involved in transcriptional elongation, termination and anti-termination as well as protecting against cold shock and stress-induced mutagenesis (100, 101). When used as a fusion host, NusA has shown to increase the soluble expression of proteins by up to 97 %, including those which were previously only expressed in insoluble form (128). One of the challenges with using NusA as a fusion host is its size in relation to the target protein; having a fusion protein which is larger than the target protein could impede activity as the active site could be blocked via steric hinderance. Small ubiquitin like modifiers (SUMO) are a family of eukaryotic proteins, encoded by single genes, which act as post translational modifiers to other proteins within the cell (129). The reversable attachment of SUMO to other protein, or SUMOylation, regulates activity, cellular location, stability and enhances protein to protein interactions (130). One of the major advantages of using SUMO as a fusion partner is that cleavage from the target protein leaves no residual amino acids at the N terminus, resulting in a native like protein (131). Additionally, the SUMO cleavage proteases are unique from other cleavage proteases in that they do not recognise a specific target sequence but rather a tertiary structure unique to SUMO (94).

3.2.1. pFLHepINTHIS

The full-length native gene sequence (GenBank) for heparinase I (**Figure 13)** was synthesised and supplied cloned in to a small, high copy number pUCIDT vector (Integrated DNA Technologies). The gene was extracted from the pUCIDT vector and subcloned into pHIS MIB vector (as described in Section 2.1.2.1.3) to create the vector construct pFLHepINTHIS. Following transformation into *E. coli* strain DH5α, plasmids were extracted from cells, digested, and visualised on an agarose gel for confirmation of vector and DNA insert. Ten digested plasmid samples were visualised on an agarose gel (**Figure 14)** with samples 4 and 9 indicating vector and gene insert sized bands of 6.5 kbp and 1.1 kbp respectively which were chosen for Sanger DNA sequencing.

```
     M  K  K  Q  I  L  Y  L  I  V  L  Q  Q  L  F  L  C  S  A  Y
   1 ATGAAAAAGCAAATTTTATATTTAATTGTCCTTCAACAATTGTTCTTATGCTCGGCCTAC 60

     A  Q  Q  K  K  S  G  N  I  P  Y  R  V  N  V  Q  A  D  S  A
  61 GCACAGCAAAAAAAATCCGGTAACATCCCTTACCGGGTAAATGTGCAGGCCGACAGTGCT 120

     K  Q  K  A  I  I  D  N  K  W  V  A  V  G  I  N  K  P  Y  A
 121 AAGCAGAAGGCGATTATTGACAACAAATGGGTGGCAGTAGGCATCAATAAACCTTATGCA 180

     L  Q  Y  D  D  K  L  R  F  N  G  K  P  S  Y  R  F  E  L  K
 181 TTACAATATGACGATAAACTGCGCTTTAATGGAAAACCATCCTATCGCTTTGAGCTTAAA 240

     A  E  D  N  S  L  E  G  Y  A  A  G  E  T  K  G  R  T  E  L
 241 GCCGAAGACAATTCGCTTGAAGGTTATGCTGCAGGAGAAACAAAGGGCCGTACAGAATTG 300

     S  Y  S  Y  A  T  T  N  D  F  K  K  F  P  P  S  V  Y  Q  N
 301 TCGTACAGCTATGCAACCACCAATGATTTTAAGAAATTTCCCCCAAGCGTATACCAAAAT 360

     A  Q  K  L  K  T  V  Y  H  Y  G  K  G  I  C  E  Q  G  S  S
 361 GCGCAAAAGCTAAAAACCGTTTATCATTACGGCAAAGGGATTTGTGAACAGGGGAGCTCC 420

     R  S  Y  T  F  S  V  Y  I  P  S  S  F  P  D  N  A  T  T  I
 421 CGCAGCTATACCTTTTCAGTGTACATACCCTCCTCCTTCCCCGACAATGCGACTACTATT 480

     F  A  Q  W  H  G  A  P  S  R  T  L  V  A  T  P  E  G  E  I
 481 TTTGCCCAATGGCATGGTGCACCCAGCAGAACGCTTGTAGCTACACCAGAGGGAGAAATT 540

     K  T  L  S  I  E  E  F  L  A  L  Y  D  R  M  I  F  K  K  N
 541 AAAACACTGAGCATAGAAGAGTTTTTGGCCTTATACGACCGCATGATCTTCAAAAAAAAT 600

     I  A  H  D  K  V  E  K  K  D  K  D  G  K  I  T  Y  V  A  G
 601 ATCGCCCATGATAAAGTTGAAAAAAAAGATAAGGACGGAAAAATTACTTATGTAGCCGGA 660

     K  P  N  G  W  K  V  E  Q  G  G  Y  P  T  L  A  F  G  F  S
 661 AAGCCAAATGGCTGGAAGGTAGAACAAGGTGGTTATCCCACGCTGGCCTTTGGTTTTTCT 720

     K  G  Y  F  Y  I  K  A  N  S  D  R  Q  W  L  T  D  K  A  D
 721 AAAGGGTATTTTTACATCAAGGCAAACTCCGACCGGCAGTGGCTTACCGACAAAGCCGAC 780

     R  N  N  A  N  P  E  N  S  E  V  M  K  P  Y  S  S  E  Y  K
 781 CGTAACAATGCCAATCCCGAGAATAGTGAAGTAATGAAGCCCTATTCCTCGGAATACAAA 840

     T  S  T  I  A  Y  K  M  P  F  A  Q  F  P  K  D  C  W  I  T
 841 ACTTCAACCATTGCCTATAAAATGCCCTTTGCCCAGTTCCCTAAAGATTGCTGGATTACT 900

     F  D  V  A  I  D  W  T  K  Y  G  K  E  A  N  T  I  L  K  P
 901 TTTGATGTCGCCATAGACTGGACGAAATATGGAAAAGAGGCCAATACAATTTTGAAACCC 960

     G  K  L  D  V  M  M  T  Y  T  K  N  K  K  P  Q  K  A  H  I
 961 GGTAAGCTGGATGTGATGATGACTTATACCAAGAATAAGAAACCACAAAAAGCGCATATC 1020

     V  N  Q  Q  E  I  L  I  G  R  N  D  D  D  G  Y  Y  F  K  F
1021 GTAAACCAGCAGGAAATCCTGATCGGACGTAACGATGACGATGGCTATTACTTCAAATTT 1080

     G  I  Y  R  V  G  N  S  T  V  P  V  T  Y  N  L  S  G  Y  S
1081 GGAATTTACAGGGTCGGTAACAGCACGGTCCCGGTTACTTATAACCTGAGCGGGTACAGC 1140

     E  T  A  R  *  *
1141 GAAACTGCCAGATAGT 1156
```

Figure 13: Full length gene and amino sequence for heparinase I. *Gene sequence (GenBank) encoding native heparinase I protein including the 21aa leader sequence (red), single letter amino acid code (**bold**) and triplet amino acid code (non-bold).*

pFLHepINTHIS Samples
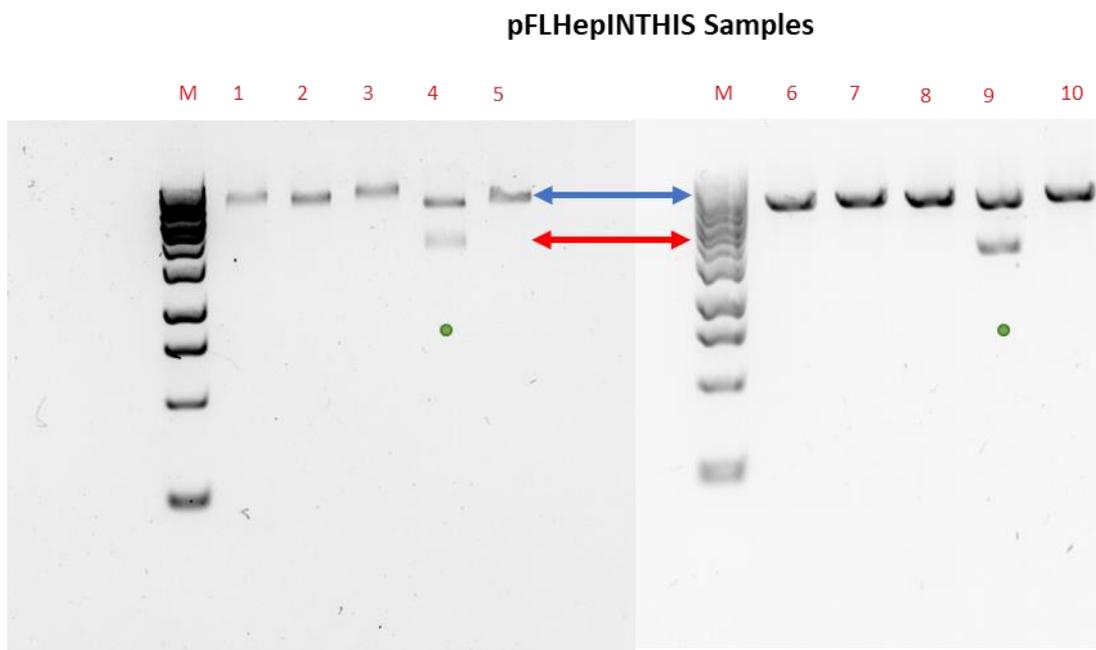
Figure 14: pFLHepINTHIS agarose gel. *Ten pFLHepINTHIS plasmid samples after digestion with BamHI and EcoRI endonuclease restriction enzymes. Samples 4 and 9 (green dots) show bands indicative in size of the digested vector (6.5 kbp, blue arrows) and DNA insert (1.1 kbp, red arrows). A DNA marker (NEB) was added to lane 1 (M) ranging between 20 bp and 6 kbp.*

3.2.2. pNTHISHepI

Integrated DNA Technologies supplied a lyophilised pUCIDT cloning vector, which contained the native heparinase I open reading frame (minus leader sequence) flanked with BamHI and NdeI endonuclease recognition sequences. Using these restriction enzymes, the heparinase I gene was subcloned in to the pET28a expression vector (Merck Millipore) to produce the N-terminal His-tagged heparinase I expression construct pNTHISHepI. Following transformation of *E. coli* DH5α cells, plasmids were extracted, digested, and visualised on an agarose gel for confirmation of vector and DNA insert. Agarose gel of six digested plasmid samples (**Figure 15**) indicate samples 1, 3, 4 and 5 as having pET28a plasmid and gene insert sized bands of 6.4kbp and 1.1 kbp respectively. Samples 2 and 6 indicate plasmids pET28a and pUCIDT only. Plasmid sample 3 was sent for forward and reverse Sanger DNA sequencing (GATC, Germany) and the sequence data were added to a BLAST (**https://blast.ncbi.nlm.nih.gov/Blast.cgi**) search for comparison against a bacterial DNA sequence database. DNA sequencing results showed a 100 % alignment to the native heparinase I protein from

Flavobacterium heparinum stored on the GenBank database (Accession number EU541216). The pNTHISHepI

verified gene sequence was used as a DNA template for all the following expression vector constructs.
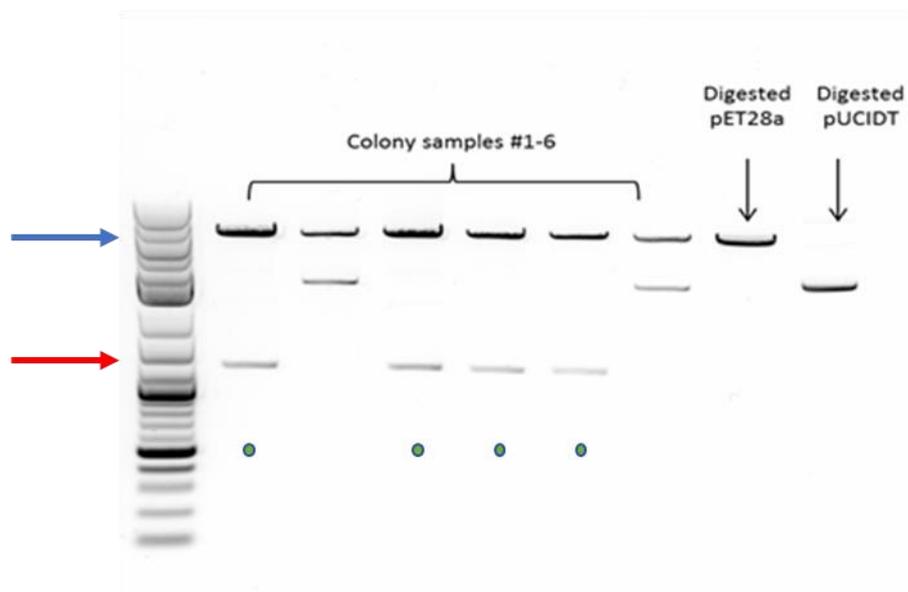
Figure 15: pNTHISHepI plasmid agarose gel. *Six pNTHISHepI plasmid samples. pET28a plasmid and pUCIDT plasmid after digestion with BamHI and NdeI endonuclease restriction enzymes. Samples 1, 3, 4 and 5 (green dots) show bands indicative of the pET28a vector (5.3 kbp, blue arrow) and DNA insert (1.1 kbp, red arrow). A DNA marker (2-log DNA ladder, NEB) was added to lane 1 (M) ranging between 0.1 and 10.0 kbp).*

3.2.3. pCTHISHepI

Heparinase I DNA was amplified for cloning into a pET28a expression vector. PCR primers were designed to

add NcoI (forward) and XhoI (reverse) restriction sites flanking the heparinase I DNA to position the His-tag at

the C terminus. After PCR amplification, samples, and negative controls, were electrophoresed on an agarose

gel and visualised. A large band indicative of the heparinase I DNA size in lane 2 and no DNA bands in lanes 3

and 4 (**Figure 16**).
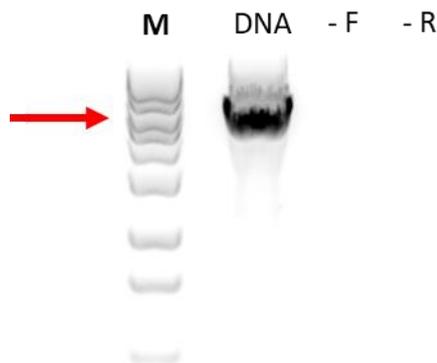
Figure 16: Agarose gel of PCR amplified heparinase I DNA. *Agarose gel of PCR amplified heparinase I DNA (lane 2), a negative control sample which omitted the forward primer from the reaction mix (lane 3) and a negative control sample which omitted the reverse primer from the reaction mix (lane 4). A DNA marker (NEB) for band size indication was added to lane 1 (M) ranging between 50 and 1200 bp. Red arrow indicates expected DNA band size at around 1 kbp.*

The amplified DNA and pET28a vector were digested with NcoI and XhoI and ligated together to create pCTHISHepI expression plasmid. Following transformation, six transformant colonies were selected, then plasmids were extracted from cells, digested, and visualised on an agarose gel for confirmation of vector and DNA insert. Agarose gel indicated samples 1, 3 and 4 with vector and gene insert sized bands of 6.3 kbp and 1.1 kbp respectively (**Figure 17**), plasmid sample 3 was chosen for DNA sequence testing.
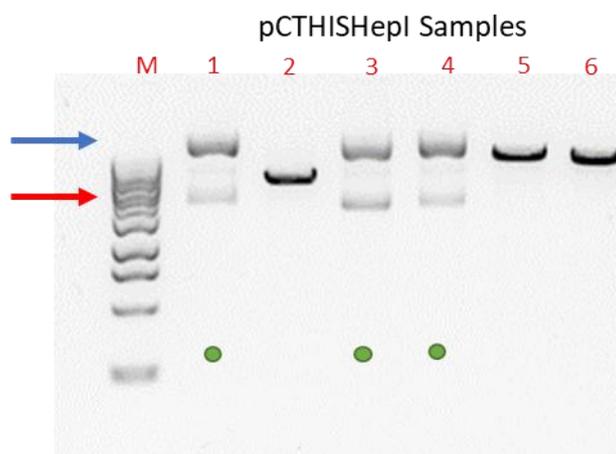
Figure 17: Agarose gel of digested pCTHISHepI plasmid samples. *Six plasmid samples after digestion with NcoI and XhoI endonuclease restriction enzymes. Samples 1, 3 and 4 (green dots) show bands indicative in size of the digested vector (5.3 kbp, blue arrow) and DNA insert (1.1 kbp, red arrow) and sample 3 was selected for Sanger sequencing. A DNA marker (NEB) was added to lane 1 (M) ranging between 20 bp and 6 kbp.*

3.2.4. pNTHIS-TrxHepI

Heparinase I DNA was amplified for cloning into a pHISTrx MIB expression vector. PCR primers were designed to add BamHI (forward) and EcoRI (reverse) restriction sites flanking the heparinase I DNA to position the DNA downstream of the His-tag and Trx chaperone protein coding sequences. After PCR amplification, samples, and negative controls, were electrophoresed on an agarose gel and visualised. A large band indicative of the heparinase I DNA size in lane 2 (**Figure 18**) and no DNA bands are visible in lanes 3 and 4.
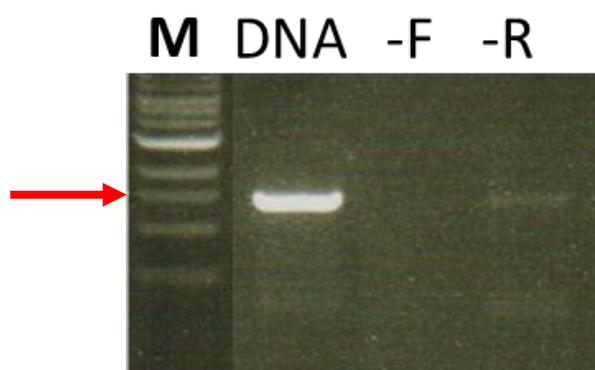
Figure 18. Agarose gel of PCR amplified heparinase I DNA. *Heparinase I DNA (lane 2), a negative control sample which omitted the forward primer from the reaction mix (lane 3) and a negative control sample which omitted the reverse primer from the reaction mix (lane 4). A DNA marker (NEB) was added to lane 1 (M) ranging between 20 bp and 6 kbp. Red arrow indicates expected DNA band of around 1 kbp.*

The amplified DNA and pHISTrx MIB vector were digested with BamHI and EcoRI and ligated together as per to create the expression plasmid pNTHIS-TrxHepI. Following transformation, plasmids were extracted from cells, digested, and visualised on an agarose gel for confirmation of vector and DNA insert. Agarose gel indicated four digested plasmid samples with vector + gene insert sized bands (**Figure 19)**. Samples 2, 3 and 4 were sent for Sanger sequencing.
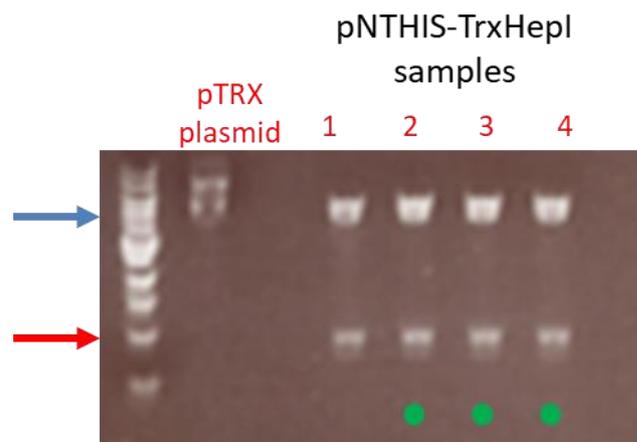
Figure 19: Agarose gel of digested pNTHIS-TrxHepI plasmids. *Four pNTHIS-TrxHepI plasmid samples after digestion with BamHI and EcoRI endonuclease restriction enzymes. All samples show bands indicative of the digested vector (6 kbp, blue arrow) and DNA insert (1.1 kbp, red arrow). Samples 2, 3 and 4 (green dots) were selected for Sanger sequencing. A DNA marker was added to lane 1 (M) ranging between 50 bp and 10 kbp.*

3.2.5. pNTHIS-NusAHepI

Heparinase I DNA was amplified for cloning into a pHISNusA MIB expression vector. PCR primers were designed to add BamHI (forward) and EcoRI (reverse) restriction sites flanking the heparinase I DNA to position the DNA downstream of the His-tag and NusA chaperone protein coding sequences. After PCR amplification, samples, and negative controls, were electrophoresed on an agarose gel and visualised. The amplified DNA and pHISNusA (MIB) vector were digested with BamHI and EcoRI, ligated together to create the expression plasmid pNTHIS-NusAHepI. Following transformation, four transformant colonies were selected, plasmid DNA was extracted, digested, and visualised on an agarose gel for confirmation of vector and DNA insert (data not shown). The plasmid sample indicative of vector and DNA insert was sent for Sanger sequencing.

3.2.6. pNTHIS-SUMOHepI

Heparinase I DNA was amplified for cloning into a pET SUMO expression vector. PCR primers were designed to add a G_4S linker followed by an enterokinase restriction site at the N terminus of the heparinase I ORF and a double STOP at the C terminus. The champion pET SUMO expression system (Invitrogen) uses TA cloning for the insertion of a target gene into the pET SUMO plasmid. This is achieved by taking advantage of the *Taq* polymerase non-template dependant activity of adding a single adenosine (A) residue to the 3' end of a PCR product (132). The pET SUMO plasmid is supplied pre-linearized with a single 3' thymidine (T) residue which

allows for direct cloning of a PCR product into the plasmid and negates the need for restriction digest (**Figure 20**).
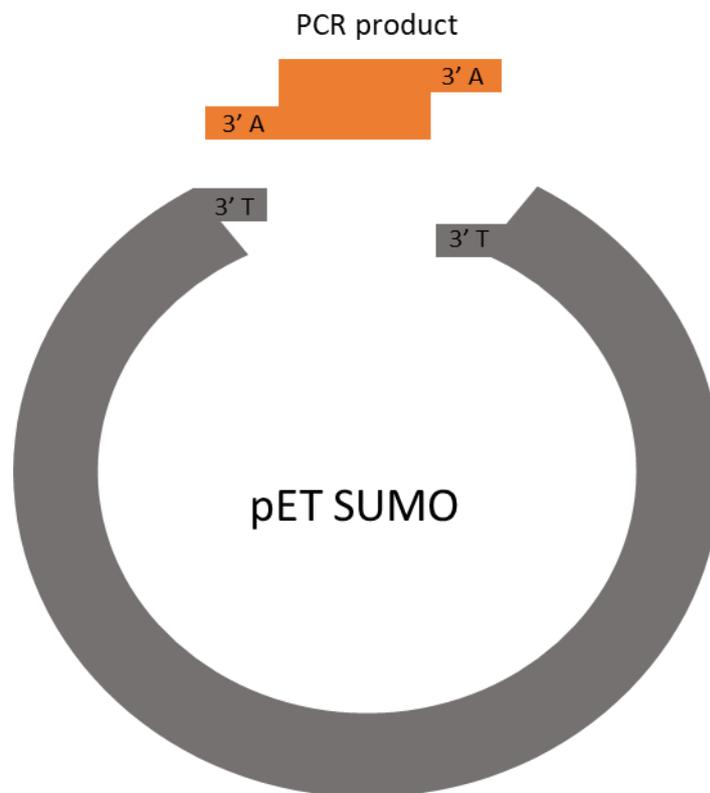
Figure 20. Champion pET SUMO plasmid and TA cloning (not to scale). *The pET SUMO plasmid has a 3' thymidine (T) residue which ligates to the 3' adenosine (A) residue on the PCR product.*

After PCR amplification, the DNA sample, and negative controls, were electrophoresed on an agarose gel, visualised, and purified. A large band in lane 2 is indicative in size of HepI and no DNA in the negative control lanes was found (lanes 3-6) (**Figure 21**). The amplified DNA and pET SUMO plasmid were ligated to create pNTHIS-SUMOHepI expression plasmid. Following transformation of competent cells and overnight growth, plasmids were extracted from cells, digested (BamHI and HindIII) and visualised on an agarose gel for confirmation of vector and DNA insert. Agarose gel shows four digested plasmid samples with all indicating vector and gene insert sized bands (**Figure 22**). Samples 2, 3 and 4 were sent for Sanger sequencing.
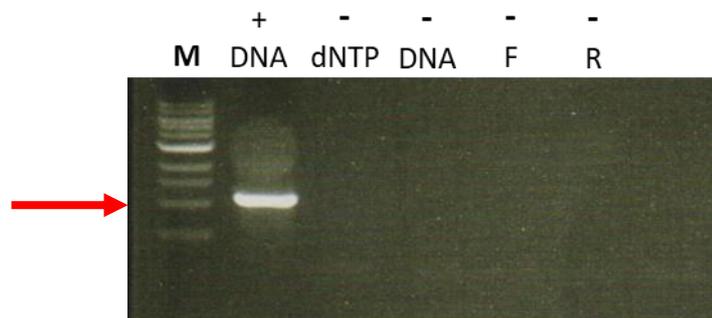
Figure 21. HepI DNA agarose gel. *PCR amplified HepI DNA (lane 2), a negative control sample which omitted the dNTP mix from the reaction mix (lane 3), a negative control sample which omitted the HepI DNA (lane 4), a negative control which omitted the forward primer (lane 5) and a negative control which omitted the reverse primer from the reaction mix (lane 6). A DNA (NEB) marker was added to lane 1 (M) 20 bp and 6 kbp. Red arrow indicates expected DNA band of 1.1 kbp.*
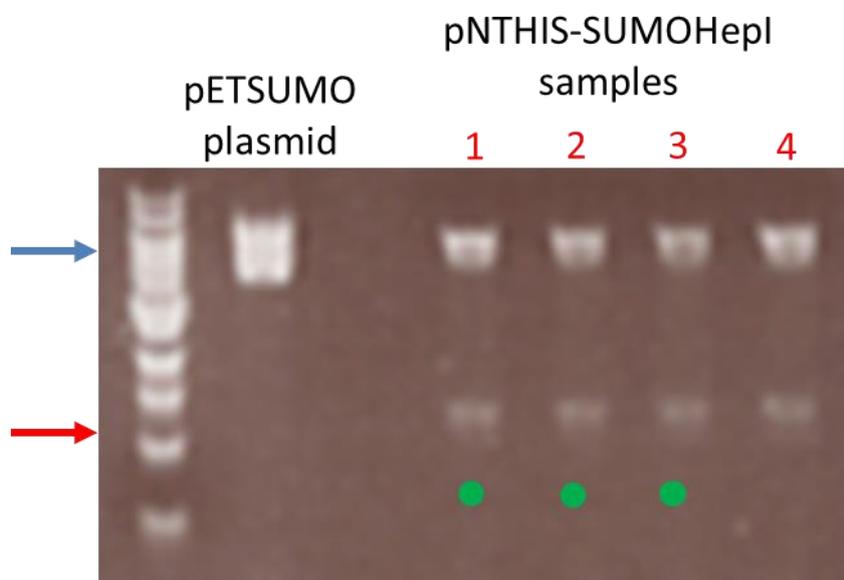
Figure 22: pNTHIS-SUMOHepI agarose gel. *Four pNTHIS-SUMOHepI plasmid samples after digestion with EcoRI and HindIII endonuclease restriction enzymes. All samples show bands indicative of the digested vector (5.6 kbp) and DNA insert (1.1kbp). Samples 1, 2, and 3 (green dots) were selected for Sanger sequencing. A DNA marker was added to lane 1 (M) between 50 bp and 10 kbp.*

3.2.7. pNTHIS-SUMOCDOHepI

When expressing recombinant proteins in *E. coli* it is common to codon optimise the gene sequence to those that are preferred by the bacterium to increase the success rate of expression, and while this does not change the overall amino acid sequence, it does change the DNA sequence and can increase the rate at which the protein is expressed (133). A consequence of this can be inclusion body formation as the protein is being expressed too fast (134). To act as a translational brake and slow down translation / expression of the HepI protein the codon sequence of the native HepI gene sequence was deliberately codon 'de-optimised' (CDO)

for *E. coli.* *T*wo DNA sequences for HepI were compared (**Figure 23**). The top row shows the native gene sequence (labelled HepI) and the bottom row shows the new DNA sequence which has been CDO for *E. coli* expression. The amino acid sequence is highlighted in red and shows that both sequences translate to the same amino acids. Mismatches in the nucleotides are highlighted by dots. The synthetic CDO HepI sequence was synthesised and supplied cloned into the pUCIDT vector (Integrated DNA Technologies). The DNA was amplified for cloning into a pET SUMO expression vector. PCR primers were designed to add a G_4S linker followed by an enterokinase restriction site at the N terminus of the heparinase I ORF and a double STOP at the C terminus.

```
            Q  Q  K  K  S  G  N  I  P  Y  R  V  N  V  Q  A  D
HepI     1  CAGCAAAAAAAATCCGGTAACATCCCTTACCGGGTAAATGTGCAGGCCGA        50
            |||||.||||||..|||.|||||.||.||.||.|||||.||.||.||.||
CDOI     1  CAGCAGAAAAAAAGCGGGAACATTCCCTATCGAGTAAACGTACAAGCAGA        50


            S  A  K  Q  K  A  I  I  D  N  K  W  V  A  V  G  I
HepI    51  CAGTGCTAAGCAGAAGGCGATTATTGACAACAAATGGGTGGCAGTAGGCA       100
            .|||||||||||||||||||||.||.|||||||||||.|||||.||.||.|||||
CDOI    51  TAGTGCTAAGCAGAAGGCTATCATTGACAACAAGTGGGTTGCCGTGGGCA       100


             N  K  P  Y  A  L  Q  Y  D  D  K  L  R  F  N  G
HepI   101  TCAATAAACCTTATGCATTACAATATGACGATAAACTGCGCTTTAATGGA       150
            |.|||||||||.||.||.||.|||||.|||||.||.||.||||||||||.
CDOI   101  TAAATAAACCCTACGCCTTGCAATACGACGACAAGCTCCGCTTTAATGGC       150


            K  P  S  Y  R  F  E  L  K  A  E  D  N  S  L  E  G
HepI   151  AAACCATCCTATCGCTTTGAGCTTAAAGCCGAAGACAATTCGCTTGAAGG       200
            ||.||.||.||.||.|||||.||.||.||.||||||.|||||.|||||||||
CDOI   151  AAGCCCTCATACCGGTTTGAACTCAAGGCGGAAGATAATTCTCTTGAAGG       200


             Y  A  A  G  E  T  K  G  R  T  E  L  S  Y  S  Y  A
HepI   201  TTATGCTGCAGGAGAAACAAAGGGCCGTACAGAATTGTCGTACAGCTATG       250
            .||||||.||.||.|||||.||.||.||||||.|||||.||.||||||||
CDOI   201  GTATGCCGCTGGGGAAACGAAAGGTCGGACAGAGCTGTCCTATAGCTATG       250


               T  T  N  D  F  K  K  F  P  P  S  V  Y  Q  N  A
HepI   251  CAACCACCAATGATTTTAAGAAATTTCCCCCAAGCGTATACCAAAATGCG       300
            |||||.||||||.||.|||||.|||||||||.||.|||||.||.||.||.||.
CDOI   251  CAACTACCAACGACTTTAAAAAAATTTCCTCCCAGCGTTTATCAGAACGCC       300


            Q  K  L  K  T  V  Y  H  Y  G  K  G  I  C  E  Q  G
HepI   301  CAAAAGCTAAAAACCGTTTATCATTACGGCAAAGGGATTTGTGAACAGGG       350
            ||||||.|.||||||.|||||.||||||||.||.|||||||||||.|||||
CDOI   301  CAAAAGTTGAAAACTGTTTACCATTACGGTAAGGGGATTTGTGAGCAGGG       350


             S  S  R  S  Y  T  F  S  V  Y  I  P  S  S  F  P  D
HepI   351  GAGCTCCCGCAGCTATACCTTTTCAGTGTACATACCCTCCTCCTTCCCCG       400
            .||...|||||...||.|||||.||.||.||||||||.|||||.||.||.|
CDOI   351  AAGTAGCCGCTCTTACACCTTCTCTGTATACATACCGTCCTCATTTCCAG       400


              N  A  T  T  I  F  A  Q  W  H  G  A  P  S  R  T
HepI   401  ACAATGCGACTACTATTTTTGCCCAATGGCATGGTGCACCCAGCAGAACG       450
            |.||.||.||.||.|||.||.|||||.||||||||||||.||.|||..||||||.
CDOI   401  ATAACGCAACAACCATATTTGCGCAATGGCATGGAGCCCCCTCCAGAACT       450


               L  V  A  T  P  E  G  E  I  K  T  L  S  I  E  E  F
HepI   451  CTTGTAGCTACACCAGAGGGAGAAATTAAAACACTGAGCATAGAAGAGTT       500
            ||.||.|||||.||||||||.||.||.||||||||.|||||||||.|||||
CDOI   451  CTCGTGGCTACTCCAGAGGGCGAGATCAAAACACTCAGCATAGAGGAGTT       500


               L  A  L  Y  D  R  M  I  F  K  K  N  I  A  H  D  K
HepI   501  TTTGGCCTTATACGACCGCATGATCTTCAAAAAAAATATCGCCCATGATA       550
            .||||||.||.|||||||||.|||||.|||||.||||||.||.|||||||||
CDOI   501  CTTGGCTTTGTACGACCGAATGATATTCAAGAAGAATATTGCACATGATA       550


                V  E  K  K  D  K  D  G  K  I  T  Y  V  A  G  K
HepI   551  AAGTTGAAAAAAAAGATAAGGACGGAAAAATTACTTATGTAGCCGGAAAG       600
            |.||.|||||||||||.|||||||.|||||||.||.||.||.||||||||||
CDOI   551  AGGTAGAAAAAAAAGACAAGGACGGCAAAATCACATACGTCGCCGGAAAG       600


            P  N  G  W  K  V  E  Q  G  G  Y  P  T  L  A  F  G
HepI   601  CCAAATGGCTGGAAGGTAGAACAAGGTGGTTATCCCACGCTGGCCTTTGG       650
            ||.||||||.|||||.||||||.||.||.||.||.||||||.||.||.||.||..|||||.|||||
```

```
CDOI        601 CCGAATGGTTGGAAAGTGGAGCAGGGTGGGTACCCGACATTGGCGTTTGG        650

                 F  S  K  G  Y  F  Y  I  K  A  N  S  D  R  Q  W  L
HepI        651 TTTTTCTAAAGGGTATTTTTACATCAAGGCAAACTCCGACCGGCAGTGGC        700
                ||||...|||||.||.||.|||||||.||.||.||.|||||||||||||
CDOI        651 TTTTAGCAAAGGTTACTTCTACATCAAAGCTAATTCAGACCGGCAGTGGC        700

                 T  D  K  A  D  R  N  N  A  N  P  E  N  S  E  V
HepI        701 TTACCGACAAAGCCGACCGTAACAATGCCAATCCCGAGAATAGTGAAGTA        750
                ||||.||||||||.|||||.||.||.|||||||||||.|||...||.|||
CDOI        701 TTACAGACAAAGCTGACCGCAATAACGCCAATCCCGAAAATTCCGAGGTA        750

                 M  K  P  Y  S  S  E  Y  K  T  S  T  I  A  Y  K  M
HepI        751 ATGAAGCCCTATTCCTCGGAATACAAAACTTCAACCATTGCCTATAAAAT        800
                ||||||||||||||||...||.||.||.|||...||.||.||.||.|||||
CDOI        751 ATGAAGCCCTATTCCAGTGAGTATAAGACTAGTACAATAGCTTACAAAAT        800

                 P  F  A  Q  F  P  K  D  C  W  I  T  F  D  V  A  I
HepI        801 GCCCTTTGCCCAGTTCCCTAAAGATTGCTGGATTACTTTTGATGTCGCCA        850
                |||.||.||.||.||.||.||.||.||.||.|||||||||.|||||.||.|
CDOI        801 GCCATTCGCGCAATTTCCAAAGGACTGTTGGATTACATTTGACGTGGCTA        850

                 D  W  T  K  Y  G  K  E  A  N  T  I  L  K  P  G
HepI        851 TAGACTGGACGAAATATGGAAAAGAGGCCAATACAATTTTGAAACCCGGT        900
                |||||.|||||||||.||||||.||||||||||||||||||.||...|.||||||.||.
CDOI        851 TAGATTGGACGAAGTATGGGAAAGAGGCCAATACGATCCTCAAACCAGGC        900

                 K  L  D  V  M  M  T  Y  T  K  N  K  K  P  Q  K  A
HepI        901 AAGCTGGATGTGATGATGACTTATACCAAGAATAAGAAACCACAAAAAGC        950
                ||.||.|||||.||||||||.||||||.||.|||||.||.||.||.|||||
CDOI        901 AAACTTGATGTCATGATGACGTATACTAAAAATAAAAAGCCTCAGAAAGC        950

                 H  I  V  N  Q  Q  E  I  L  I  G  R  N  D  D  D  G
HepI        951 GCATATCGTAAACCAGCAGGAAATCCTGATCGGACGTAACGATGACGATG        1000
                .|||||.||.|||||.|||||||||||.||.||.||.|||||.|||||.|
CDOI        951 CCATATAGTGAACCAACAGGAAATCCTCATTGGTCGGAACGACGACGACG        1000

                 Y  Y  F  K  F  G  I  Y  R  V  G  N  S  T  V  P
HepI       1001 GCTATTACTTCAAATTTGGAATTTACAGGGTCGGTAACAGCACGGTCCCG        1050
                |.||||||||||||||||||.||.||..|||||.||.||.|||||.|||||.
CDOI       1001 GATATTACTTCAAATTTGGGATCTATCGGGTAGGGAATAGCACTGTCCCC        1050

                 V  T  Y  N  L  S  G  Y  S  E  T  A  R  *  *
HepI       1051 GTTACTTATAACCTGAGCGGGTACAGCGAAACTGCCAGATAGTAA        1100
                ||.||.|||||||||.|||||.|||..||||||||||.|.||.|||
CDOI       1051 GTGACATATAACCTTAGCGGTTACTCCGAAACTGCCCGGTAATAA        1095
```

Figure 23: Codon de optimisation of HepI sequence. *The native HepI sequence (labelled HepI) compared to the deliberate codon de-optimised HepI sequence (labelled CDOI) Mis matches are highlighted by black dots and matches by lines. The translated amino acid sequence is shown in red.*

After PCR amplification, the DNA sample, and negative controls, were electrophoresed on an agarose gel, visualised, and purified. Agarose gel shows a large band in lane two indicative in size of HepI and no DNA in the negative control lanes (3 and 4) (**Figure 24**). The amplified DNA and pET SUMO plasmid were ligated to create pNTHIS-SUMOCDOHepI expression plasmid. Following transformation of competent cells and overnight growth, plasmids were extracted from cells, digested (BamHI and HindIII) and visualised on an agarose gel for

33

confirmation of vector and DNA insert. Agarose gel of four digested plasmid samples (**Figure 25**) indicates all

sample had vector and gene insert sized bands. Samples 1, 3 and 4 were sent for Sanger sequencing.
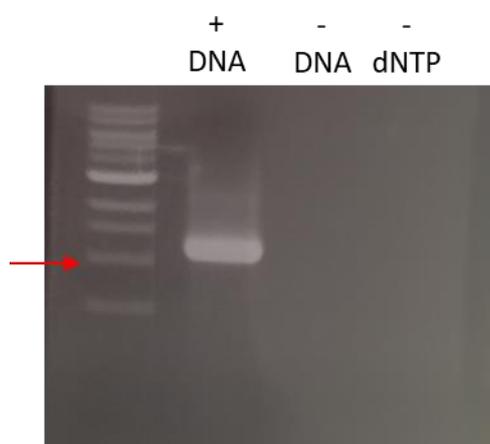
Figure 24: PCR amplified, codon de-optimised HepI DNA agarose gel. *Positive DNA amplified sample (lane 2), a negative control sample which omitted the DNA from the reaction mix (lane 3), and a negative control sample which omitted the dNTP mix from the reaction mix (lane 4). A DNA (NEB) marker was added to lane 1 (M) between 20 bp and 6 kbp. Red arrow indicates expected DNA size of 1.1 kbp.*
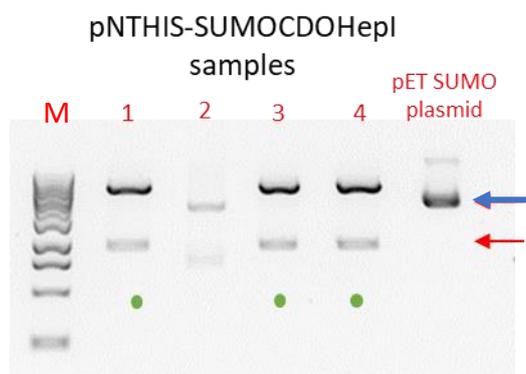
Figure 25: pNTHIS-SUMOCDO agarose gel. *Four pNTHIS-SUMOCDO HepI plasmid samples and a pET SUMO plasmid after digestion with EcoRI and HindIII endonuclease restriction enzymes. All pNTHIS-SUMOCDO samples show bands indicative of the digested vector (5.6 kbp) and DNA insert (1.1kbp). Samples 1, 3, and 4 (green dots) were selected for Sanger sequencing. Blue arrow indicates expected plasmid size of ~5 kbp and red arrow indicates expected DNA insert size of ~1 kbp. A DNA marker (NEB) was added (M) ranging between 20 bp and 6 kbp.*

3.2.8. Summary

A library of expression vectors was created to maximise soluble expression of HepI in *E. coli,* which included a

vector which contained a full-length version of the HepI DNA to determine if the 21 amino acid leader sequence

influenced solubility and / or activity. An approach to soluble HepI expression which has not been attempted

before was used whereby the HepI gene sequence was codon de-optimised for *E. coli* expression in the attempt

34

to act as a translation brake and encourage correct folding of the protein (135). Maps of the expression vectors constructed for this study were created (Benchling), and show the key features of the vectors, including size, antibiotic resistance genes, plasmid replicon, position of the target protein insert, and which endonuclease restriction sites were used for ligation (see **Figure 26**). DNA concentration and purity of each expression vector was determined by UV spectrophotometry at 260 nm and 280 nm wavelengths (Nanodrop Lite, ThermoFisher)**.** All the expression plasmids DNA sequence were verified by Sanger sequencing (GACT Germany) and the amino acid sequence matched to the native HepI sequence when using a pairwise alignment tool (EMBOSS Needle, **www.ebi.ac.uk**). The concentration of the expression plasmids varied between 34.8 ng/µl for the pNTHIS-TrxHepI and 126.5 ng/µl for the pNTHIS-NusAHepI (

Table *10*), however, these were exceptionally low and high concentrations with the other five vectors ranging between 43.4 and 66 ng/µl (

Table *10*). The purity of the vectors ranged from 1.79-1.85 with 1.8 being generally accepted as pure for DNA (136). A 260/280 ratio of higher than 1.8 would indicate RNA contamination and lower than 1.8 would indicate protein contamination.

A pFLHepINTHIS
6542 bp

B pNTHISHepI
6431 bp

C pCTHISHepI
6338 bp

D pNTHIS-TrxHepI
6815 bp
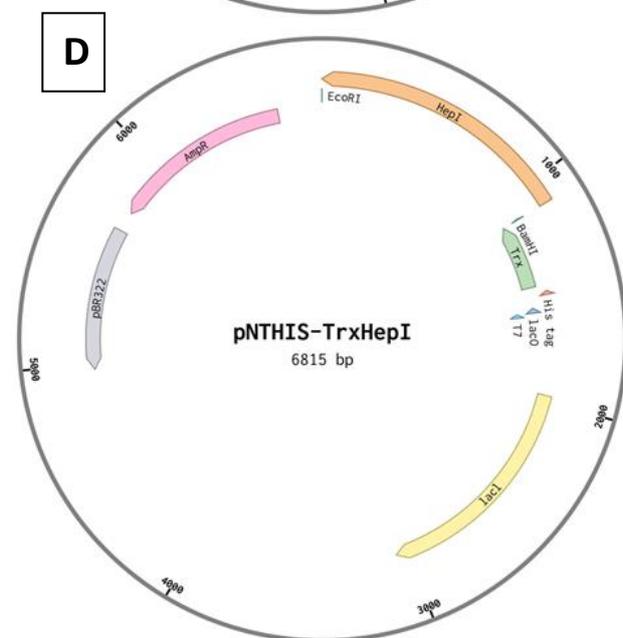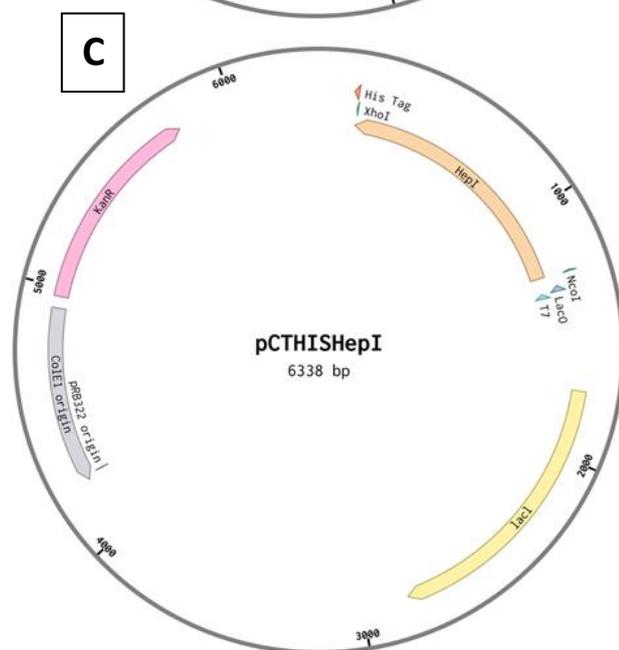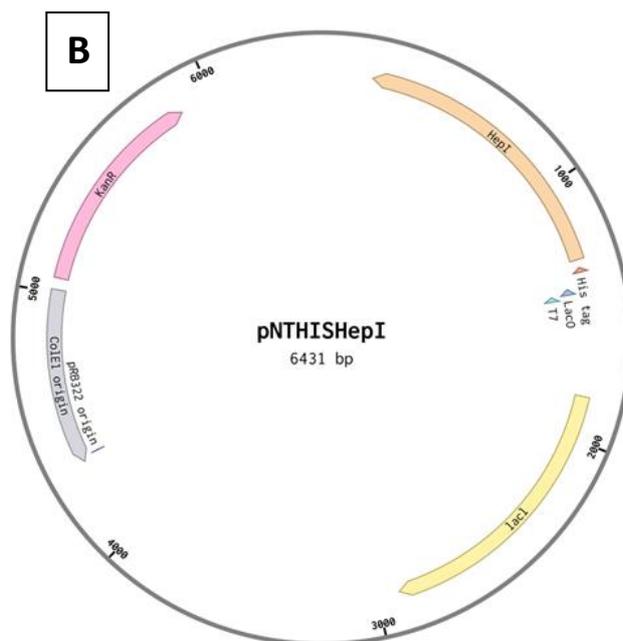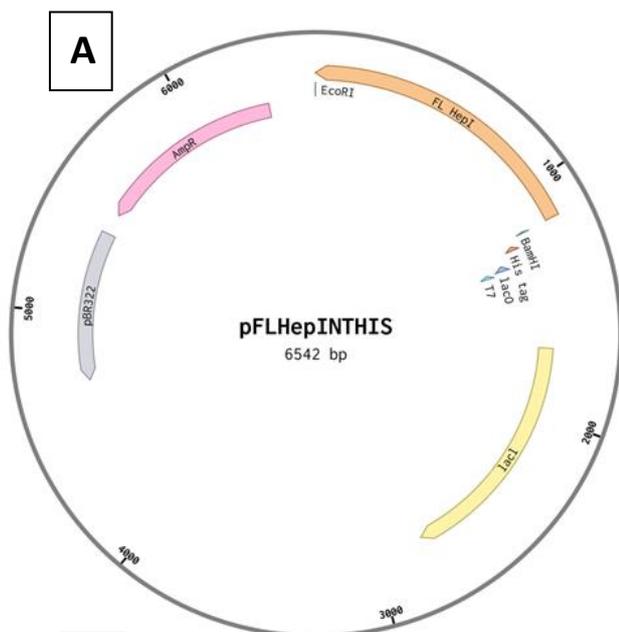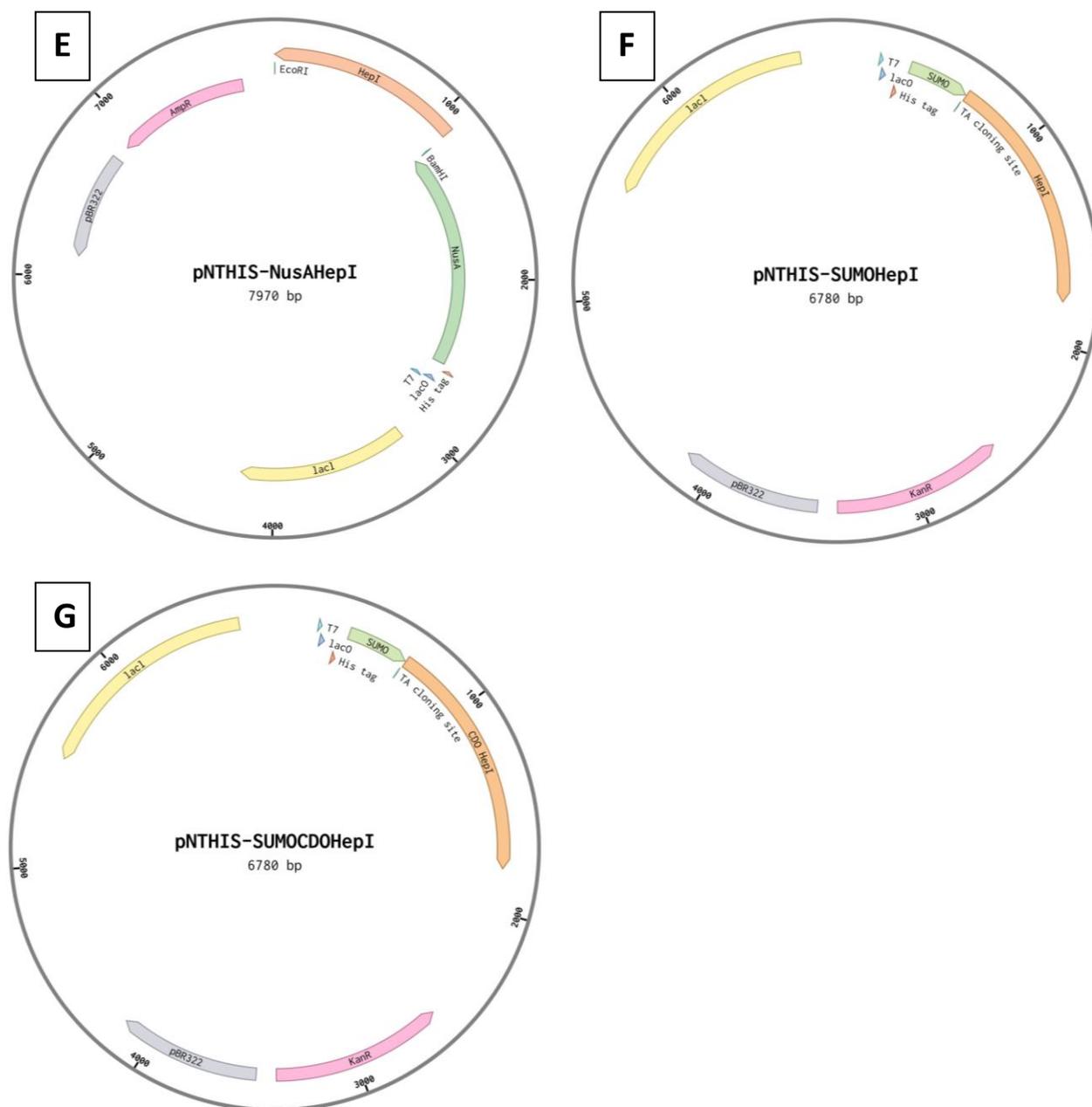
Figure 26: Expression vector maps. *Key features include a lacI ORF for encoding lac repressor (yellow), an IPTG-inducible T7 promotor region (light blue arrow), lac operator (lacO) for lac repressor binding (dark blue), antibiotic resistance gene (pink), pBR122 (colEI) origin of replication (purple) and a His-tag coding sequence (peach arrow) and endonuclease restriction sites (green arrow), and the position of the inserted HepI sequence (orange arrow). A. pFLHepINTHIS B. pNTHISHepI C. pCTHISHepI D. pNTHIS-TrxHepI E. pNTHIS-NusAHepI F. pNTHIS-SUMOHepI G. pNTHIS-SUMOCDOHepI.*

Table 10: Concentration and purity of seven vector constructs created for recombinant HepI expression.

Construct	Concentration (ng/µl)	Purity (A260/A280)
pFLHepINTHIS	51.2	1.83
pNTHISHepI	51.5	1.8
pCTHISHepI	54.4	1.79
pNTHIS-TrxHepI	34.8	1.86
pNTHIS-NusAHepI	126.5	1.85
pNTHIS-SUMOHepI	43.4	1.82
pNTHIS-SUMOCDOHepI	66.3	1.81

3.3. Expression trials

All expression vectors were cloned into the different host cells. *E. coli* BL21(DE3) is considered 'standard' for protein expression and was used in this study as the baseline host for protein expression. The BL21 (DE3) strain of *E. coli* is a common choice for an expression host because 1) It is a T7 phage lysogen and so carries the T7 polymerase needed for T7 promoter and 2) it is a protease deficient mutant. The other four host cells were expected to enhance expression of soluble protein. Two induction methods were used, IPTG induced expression and AI. For IPTG induction, all transformed hosts were induced using 0.2 mM IPTG and cultures grown overnight at 18 °C to allow for a slower, more controlled level of expression (137). For AI, the transformed hosts were added to selective AI media and grown overnight at 30 °C. A point to note is that during the expression trials, it was discovered that the strain of Rosetta-gami DE3 has an endogenous Kan^R gene which resulted in the inability to select for pNTHISHepI, pCTHISHepI, pNTHIS-SUMOHepI and pNTHIS-SUMOCDOHepI transformants.

The expression trials were expected to produce a total of 26 possible screens (6 plasmids expressed in 4 backgrounds and two AmpR plasmids also tested in the Rosetta background). Of the 26 screens, 10 vector/host combinations produced measurable protein. Due to time limitations, any vector/host combination which failed to produce colonies, or detectable protein were discarded. SDS-PAGE analysis of the insoluble and soluble protein (**Figure 27**) and soluble protein concentration as determined by Bradford assay (**Table 11**) shows that the vector / host combinations producing the highest percentage of soluble protein at 90 % were pNTHIS-SUMOCDOHepI plus BL21 (DE3) and pNTHIS-SUMOHepI plus JM109 (DE3) chaperone 4, however pNTHIS-SUMOCDOHepI plus BL21 (DE3) produced 18.9 mg total soluble protein compared to 13.0 mg soluble protein

from the combination pNTHIS-SUMOHepI plus JM109 (DE3) chaperone 4. When used to transform BL21 (DE3) the vector pNTHIS-NusAHepI produced the lowest amount of protein at 7.8 mg, and pNTHIS-TrxHepI + BL21 (DE3) produced 12.4 mg of soluble protein, however this only constituted 40 % of the total protein expressed. Nine of the vector / host combinations produced over 50 % soluble protein with concentration ranging between 1.9-2.9 g/l which is considered high yield for a non-fermentation process (138). The five vector / host combinations producing the highest concentration of soluble protein were selected for repeat expression and purification trials.

E. pNTHIS-SUMOHepI

F. pNTHIS-SUMOCDOHepI
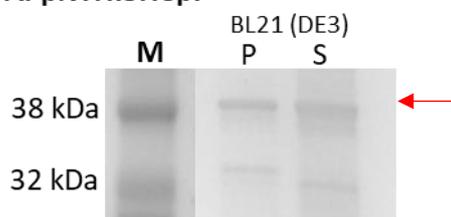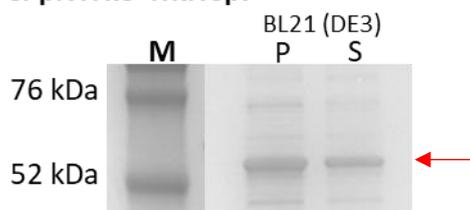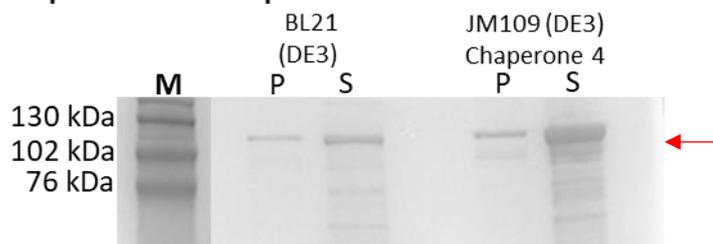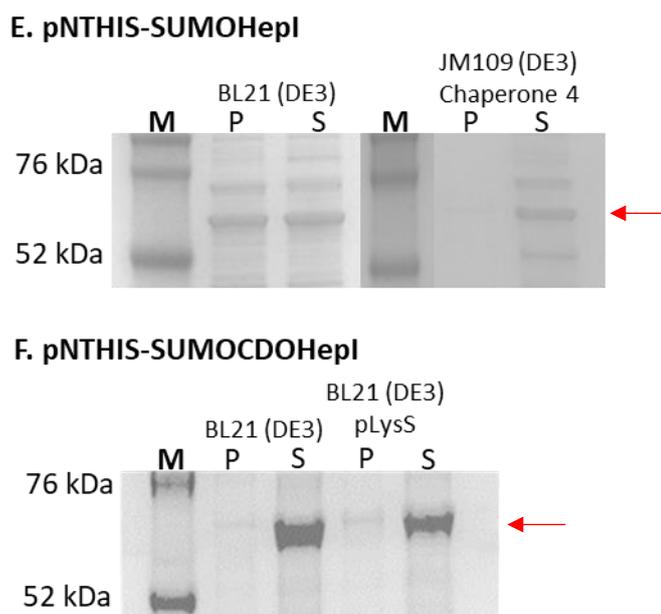
Figure 27: SDS-PAGE analysis of protein extracted from E. coli cultures during expression trials. *Expression plasmids (A-F) were used to transform host E. coli cells, after induction cultures were lysed and the soluble and insoluble fractions separated and analysed using SDS-PAGE. M. Molecular weight marker, P. Insoluble pellet fraction from protein lysate, and S. soluble fraction from protein lysate. Red arrow indicates expected MW of the HepI protein (+ fusion protein where indicated).*

Table 11: Analysis of HepI expression using a variety of plasmid and host combinations.

Vector and host combination	Insoluble fraction v soluble fractions (% weighting)	Concentration (g/L)	Total soluble protein expressed (mg)
pNTHISHepI + BL21 (DE3)	50/50	1.9	12.4
pCTHISHepI + BL21 (DE3)	30/70	2.3	15.0
pCTHISHepI + JM109 (DE3) chaperone 4	40/60	2.0	13.0
pNTHIS-TrxHepI + BL21 (DE3)	60/40	1.9	12.4
pNTHIS-NusAHepI + BL21 (DE3)	40/60	1.2	7.9
pNTHIS-NusAHepI + JM109 (DE3) chaperone 4	20/80	2.6	17.0
pNTHIS-SUMOHepI + BL21 (DE3)	50/50	2.1	13.7
pNTHIS-SUMOHepI + JM109 (DE3) chaperone 4	10/90	2.0	13.0
pNTHIS-SUMOCDOHepI + BL21 (DE3)	10/90	2.9	18.9
pNTHIS-SUMOCDOHepI + BL21 (DE3) pLysS	20/80	2.5	16.3

Five vector / host combinations from the expression trials which expressed soluble HepI in high yield (in the soluble fraction) were selected for repeat expression and purification trials. The vector / host combinations selected were pCTHISHepI + BL21 (DE3), pNTHIS-TrxHepI + BL21 (DE3), pNTHIS-NusAHepI + JM109 (DE3) chaperone 4, pNTHIS-SUMOHepI + BL21 (DE3), and pNTHIS-SUMOCDOHepI + BL21 (DE3). Two further expression trials were carried out and the expressed protein analysed by SDS-PAGE, western blot, and Bradford

40

assay for protein concentration (**Figure 28**). The expression vector and host combination which consistently expressed the highest total amount of soluble protein was pNTHIS-SUMOCDOHepI + BL21 (DE3) at 18.9, 15.6 and 16.9 mg across the three trials with an average of 17.1 mg (± 1.6). The expression host which consistently expressed the least amount of soluble protein was pNTHIS-TrxHepI + BL21 (DE3) at 12.4, 7.8 and 9.1 mg soluble protein with an average of 9.8 mg (± 2.3) across the three trials. The expression vector and host combination with the least degree of variation in amount of soluble protein expressed was pNTHIS-SUMOHepI + BL21 (DE3) with 13.7, 13.0 and 13.7 mg with an average of 13.4 mg (± 0.4).
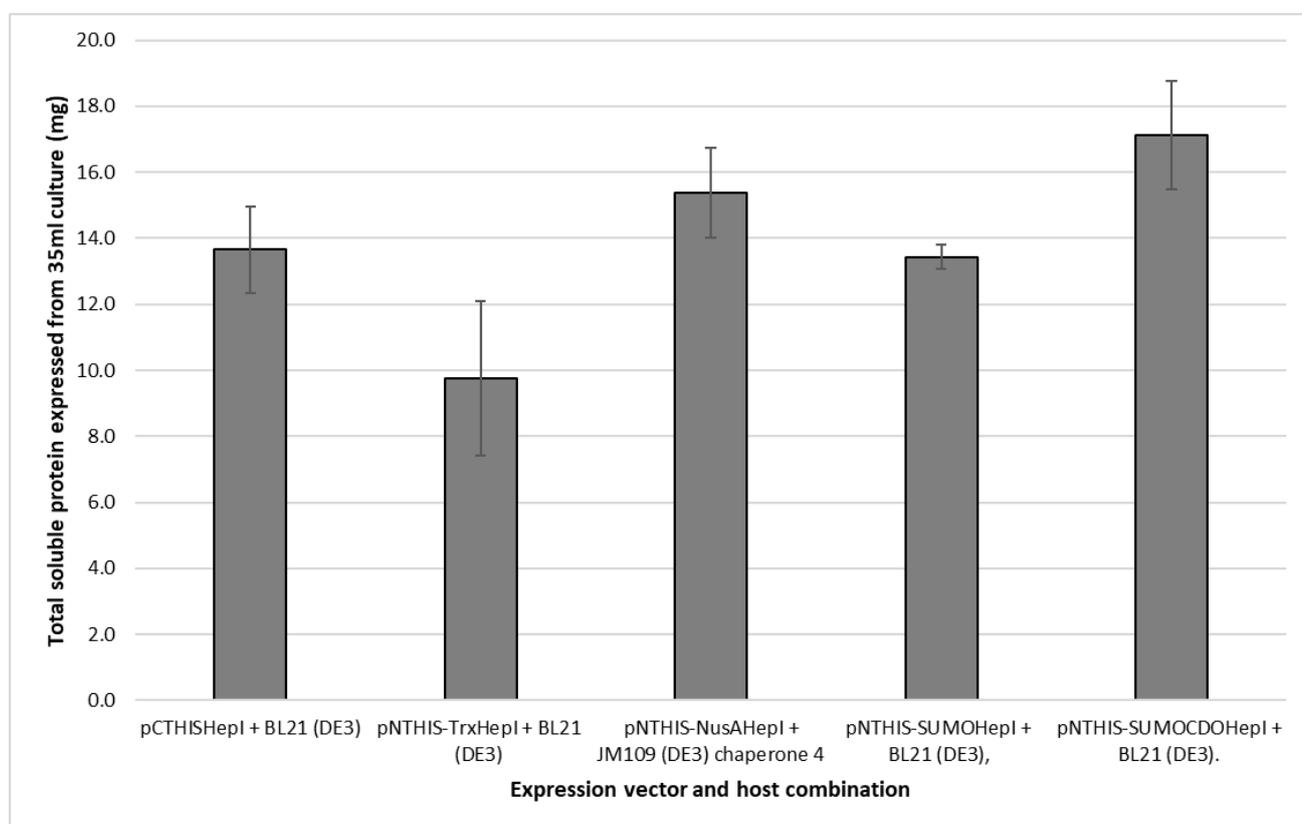
Figure 28: Soluble expression of HepI from different vector / host combination across three trials. *E. coli cultures were transformed with different expression vectors and induced with 0.2 mM IPTG and grown overnight at 18 °C. Cells were lysed and the total amount of soluble protein from 35 ml cultures was calculated (mean of three separate trials ±SD).*

As stated previously, the heparinase I gene contains a 21 amino acid signal peptide which is cleaved prior to secretion in the native host (139). The signal peptide directs transport of the protein via the Sec translocon to the periplasm for post translational modification (glycosylation) in the natural host (140). Because the peptide is removed prior to secretion and is highly hydrophobic, it is thought that it plays no role in protein function, expression or folding (141). In this study an expression trial was carried out to test the effect of leaving the

signal peptide intact, to determine if this a, reduces expression entirely or b, increases insoluble expression of the protein. SDS-PAGE analysis shows that overexpression of HepI was only detected for protein expressed in the host JM109 (DE3) chaperone 4 (

Figure 29). The concentration of protein in this fraction was 0.8 g/L determined by Bradford assay at 295 nm. Hosts BL32 (DE3), BL21 (DE3) pLysS, and JM109 (DE3) chaperone 3 transformed with pFLHepINTHIS show no over expression of protein where HepI would be expected (42.8 kDa).
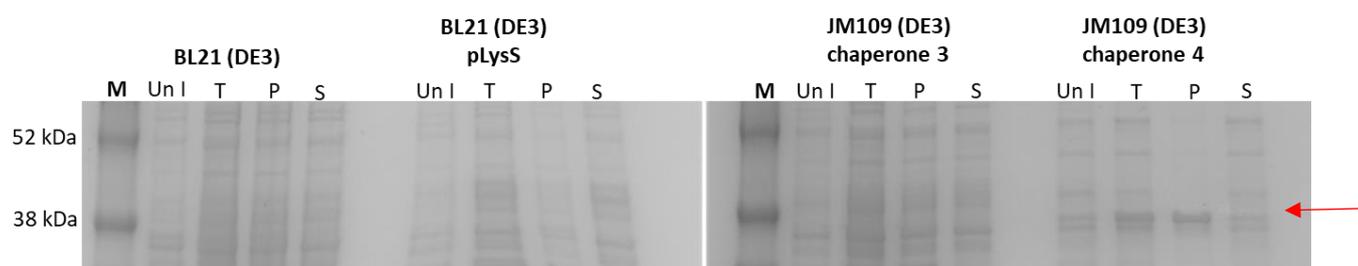
Figure 29: SDS-PAGE of protein extracted from E. coli cultures. *Host cells were transformed with pFLHepINTHIS containing the full length HepI gene and induced with 0.2mM IPTG overnight at 18 °C. Cells were clarified and samples separated to assess protein expression. M. Molecular weight marker, uninduced control (Un I), Total extract (T), Pellet fraction from lysate (P) and soluble fraction from lysate (S). Red arrow indicates expected size of HepI protein of 42 kDa.*

An autoinduction trial was carried out, plasmid constructs pNTHISHepI, pCTHISHepI, pNTHIS-TrxHepI, pNTHIS-SUMOHepI and pNTHIS-SUMOCDOHepI were transformed in to BL21(DE3) expression cells and grown overnight in AI media at 30 °C. The pNTHIS-NusAHepI was not included in the AI trial due to lack of available sample. After auto-induction, cultures were clarified, and the extracted protein analysed using SDS-PAGE. The AI trial resulted in an over expression of protein indicative of HepI, however almost 100 % was found in the insoluble fraction in four of the five samples (

Figure 30). Protein extracted from in BL21 (DE3) transformed with pNTHIS-SUMOHepI produced up to 80 % less protein than the other samples, however, 60 % was in the soluble fraction which overall, was the highest yield of soluble protein.
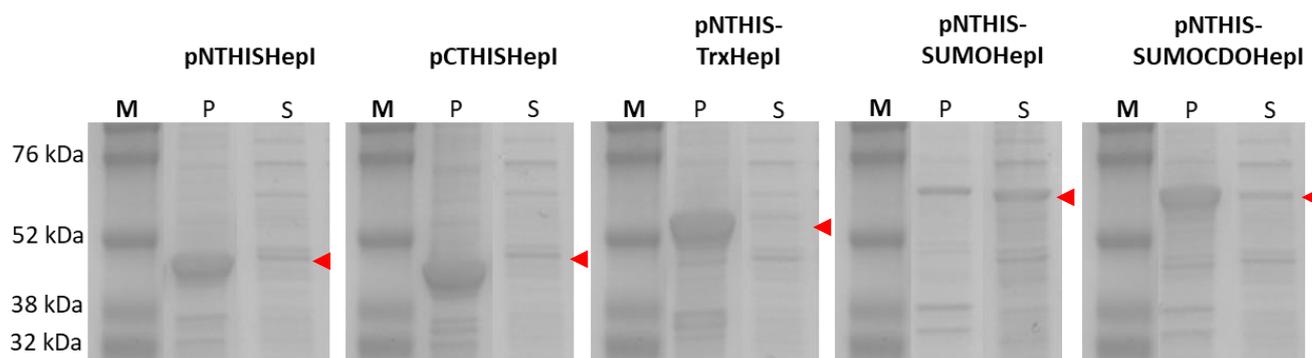
Figure 30: SDS-PAGE of protein extracted from auto-induced *E. coli* cultures. *E. coli BL21 (DE3) were transformed with expression plasmids (top label) and auto-induced overnight at 30 °C. Cells were clarified and samples separated in to insoluble (P) and soluble (S) fractions to assess protein expression. A protein marker was added to identify protein band size (M). Arrowhead indicates expected MW of the HepI protein (+ fusion protein where indicated).*

3.4. Summary

To maximise soluble expression of recombinant HepI protein a total of seven expression vectors were created with various combinations of fusion partners and tag positions. These were used to transform five host *E. coli* strains. For expression cultures, two methods of induction were used, IPTG and AI, as well as two induction temperatures. When induced at 37 °C recombinant HepI is expressed almost exclusively in the insoluble fraction (S. James previous work, data not shown), and it was thought that by reducing the induction temperature with the combination of a low concentration of IPTG would increase soluble expression of HepI protein. Reducing the IPTG concentration to 0.2 mM and expression temperature to 18 °C had a positive impact on soluble protein expression with over 50 % soluble expression being achieved with different vector and host combinations. The expression vector and host combination which produced the highest yield of soluble protein was pNTHIS-SUMOCDOHepI + BL21 (DE3) with an average of 17.1 mg per 35 mL culture or 480 mg/L culture.

Auto-induction was trialled in this project study as researchers have found that proteins can be expressed in the soluble fraction when they have previously formed into inclusion bodies during IPTG induced expression (142,143). AI media contains well-balanced mixture of carbon sources, glucose and lactose which promotes fast cell growth prior to induction, and as glucose levels are depleted lactose is consumed and recombinant protein expression is induced (142,144). Auto-induction is a method of producing high yields of recombinant protein without having to monitor cell density, unlike IPTG induction, which makes it less labour intensive. The results of the AI trial showed that although there was an over expression of HepI protein, it was found almost

exclusively in the insoluble fraction. A repeat of the AI trial could be conducted at 18 °C to see if a reduction in temperature increases soluble expression of HepI as indicated in the literature (113). The expression trial using the full-length version of HepI resulted in either no expression at all or inclusion body formation.

Chapter 4. Purification

4.1. Introduction

Purification trials were conducted using Immobilised Metal Affinity Chromatography (IMAC) at both small-scale and large-scale. IMAC is a powerful method of purifying target proteins based on the affinity between transitional metal ions such as Ni^{2+}, Zn^{2+} and Co^2 and exposed amino acid surface chains containing histidine. Expressing a target protein with a His-tag (several adjacent histidine residues) attached, increases the affinity to the immobilised metal ions multi-fold. As an alternative to column-based methods of purification (i.e., FPLC) magnetic IMAC beads allow for high throughput, flexible, and cost-effective purification of small quantities of His-tagged proteins with the additional benefit of easy separation of the solid and liquid phases during processing (145). Small-scale pilot studies for purification allow for the optimisation of buffer composition, incubation times, pH and temperatures whilst using a minimal amount of product. Once an optimal purification protocol is established it can be transferred to a more automated programme such as FPLC for large scale purification of up to kilogrammes of protein (116).

4.2. Protein purification

Crude HepI protein expressed in the soluble fraction, in high yield, from the five vector / host combinations was purified using metal IMAC beads on three separate occasions under native conditions. Samples taken during the process were analysed by SDS-PAGE (**Figure 31)** and western blot (**Figure 32**). Bradford analysis showed that NTHISCDOHepI + BL21 (DE3) produced the highest total amount of purified protein at 4.0 mg (± 0.3) across three experiments which was between 24 % and 170 % higher than the other samples (**Table 12**). Sample pNTHIS-NusAHepI + JM109 (DE3) chaperone 4 produced the lowest amount of purified soluble protein at 1.5 mg (± 0.7 SD) across the three experiments.
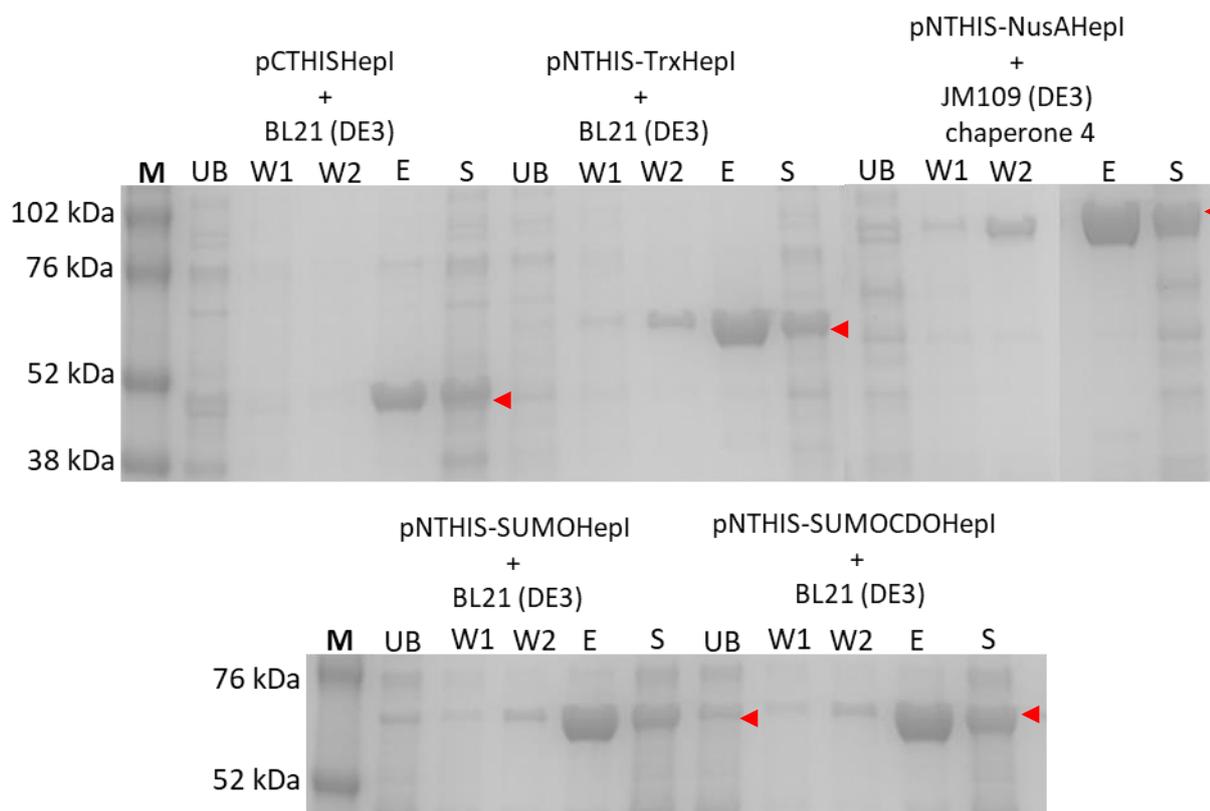
Figure 31: SDS-PAGE of soluble protein samples before, during and after metal IMAC purification. Soluble samples of HepI protein were purified using magnetic, nickel coated SepFast beads. Soluble protein was incubated with 50 µl beads for 1.5 hours at room temperature. The beads (and bound protein) were washed twice, and the protein eluted with 500 mM imidazole buffer. Samples were taken of the unbound protein (UB) after each wash (W1, W2) and after elution from the beads (E). A sample of the crude soluble lysate was also included (S) and a molecular weight marker (M). Expected size of fusion proteins are 45 kDa for pCTHISHepI + BL21 (DE3), 54.5kDa for pNTHIS-TrxHepI + BL21 (DE3), 97.6 kDa for pNTHIS-NusAHepI + JM109 (DE3) chaperone 4, and 57.8 kDa for pNTHIS-SUMOHepI + BL21 (DE3) and pNTHIS- SUMOCDOHepI + BL21 (DE3). Arrowhead indicates expected MW of the HepI protein (+ fusion protein where indicated). Please note that this figure has been manipulated to remove 'empty' lanes from the gel image.
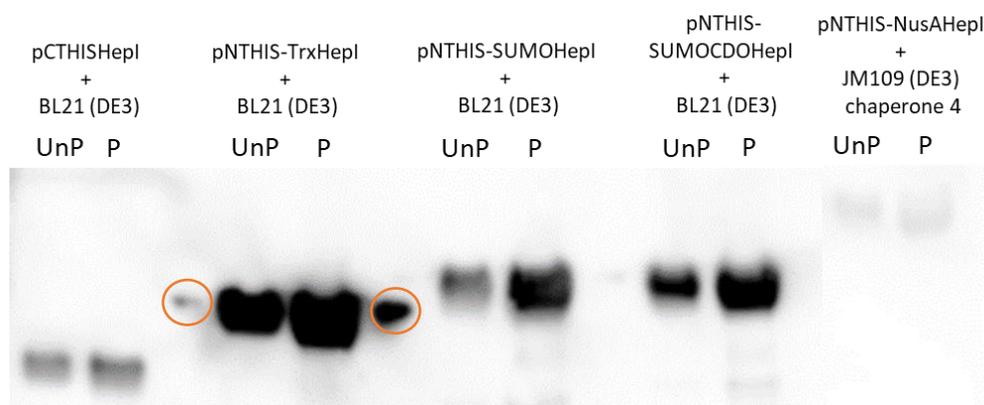
Figure 32: Western blot of HepI protein samples before and after purification. *Hep I protein samples were expressed from different vector / host combinations (top label) and purified by magnetic IMAC. The un-purified (UnP) and purified (P) samples were subjected to SDS-PAGE; the protein was then transferred to a PVDF membrane and incubated with 1° Ab. anti-His raised in mouse and then 2° Ab anti-mouse raised in rabbit with conjugated horse radish peroxidase (HRP). Protein bands were visualised by chemiluminescence. Please note that this figure has been manipulated to remove 'empty' lanes from the gel image. Circulated bands in lanes 3 and 6 are overspill from lanes 2 and 5.*

Table 12: Bradford analysis of HepI protein from 35 ml culture before and after purification.

Sample	Total protein pre-purification (mg ±SD)	Total protein post-purification (mg)
pCTHISHepI + BL21 (DE3)	13.7 (± 1.6)	2.4 (± 1.3)
pNTHIS-TrxHepI + BL21 (DE3)	9.8 (± 1.9)	1.8 (± 0.9)
pNTHIS-NusAHepI + JM109 (DE3) chaperone 4	15.4 (± 1.1)	1.5 (± 0.7)
pNTHIS-SUMOHepI + BL21 (DE3)	13.4 (± 0.3)	3.2 (± 0.1)
pNTHIS-SUMOCDOHepI + BL21 (DE3)	17.1 (± 1.3)	4.0 (± 0.3)

The vector / host combination pNTHIS-SUMOCDOHepI + BL21 (DE3) showed highest level of soluble protein expression and the highest total amount of purified protein after purification, so was chosen for 1 L expression trials and purification using an AKTA FPLC (GE healthcare). The 1 L culture produced 200 ml of crude soluble protein, of which 40 ml was added to the FPLC column. During each stage of the expression and purification trial, samples were taken for SDS-PAGE analysis (**Figure 33**). Some SUMOCDOHepI protein was lost in the wash stages due to the increased imidazole concentrations (**Figure 33B**), but there are large bands of overexpressed SUMOCDOHepI protein in the elution fractions. Peak fractions indicating SUMOCDOHepI protein were pooled, giving a total of 16 ml purified protein at an average concentration of 2.0 g/l (± 0.1SD). Data from three separate 1 L expression and AKTA purifications of SUMOCDOHepI protein (**Table** *13***)** shows on average 161.9 mg (± 9.3 SD) of purified SUMOCDOHepI protein was produced.
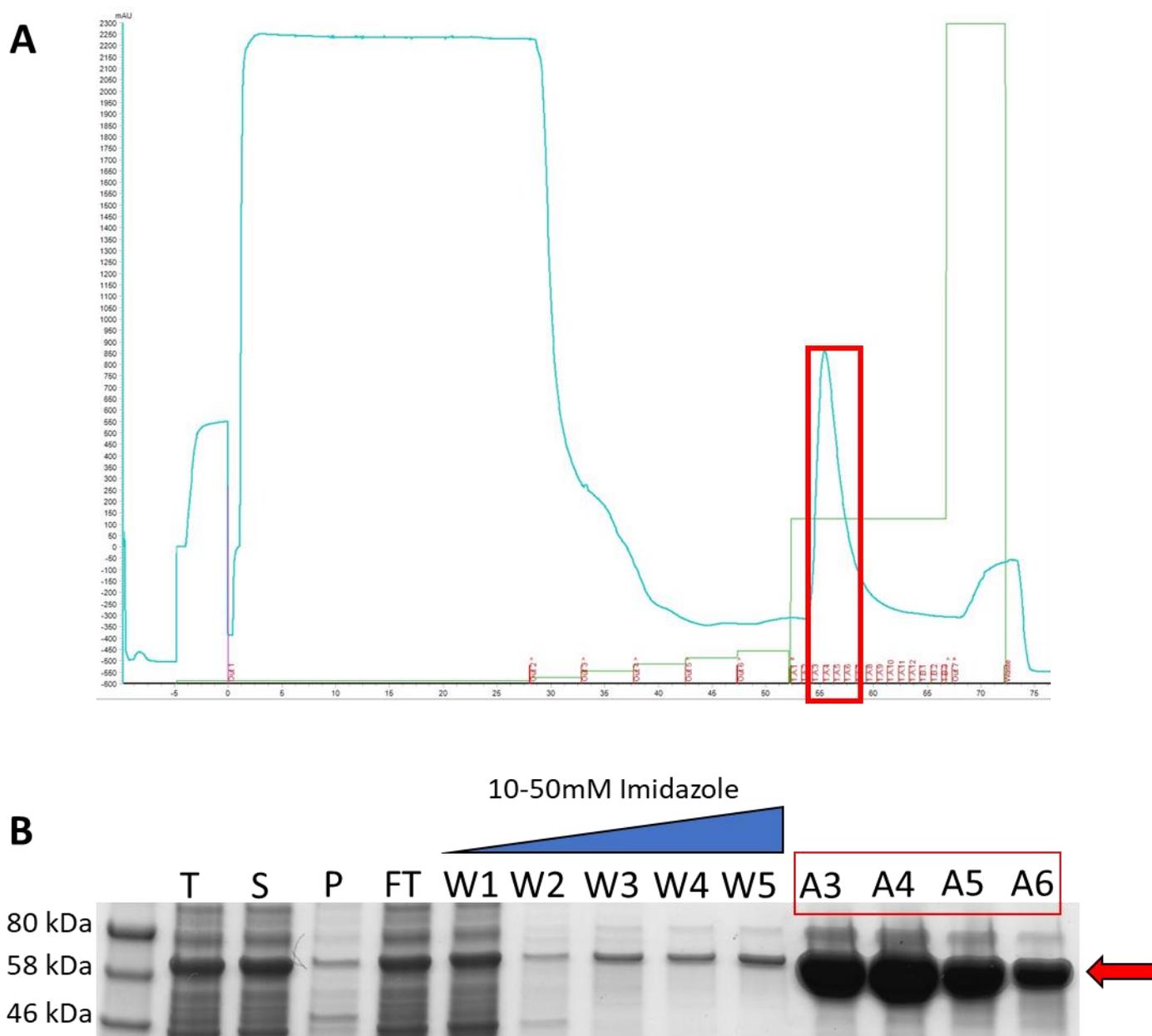
Figure 33: FPLC purification and SDS PAGE analysis of HepI protein expressed from 1 L *E. coli* culture. *E. coli BL21 (DE3) were transformed with pNTHIS-SUMOCDOHepI, induced, and cultured overnight. Cells were lysed and the soluble fraction purified using an AKTA Pure FPLC. The column was washed with increasing concentrations of imidazole (10-50 mM) and the protein eluted with 250 mM imidazole. A. Chromatograph of FPLC run. Blue line represents absorbance at 280 nm (protein concentration) and green line represents conductivity (salt concentration). The peak containing HepI protein is highlighted by a red box. B. SDS-PAGE analysis of protein during expression and purification. Total (T) soluble (S) insoluble pellet (P) from expression and the flow-through (FT), column washes (W1-5), and the protein peak fractions (A3-6) from purification. Red arrow indicates expected band size of SUMOCDOHepI protein.*

Table 13: Soluble SUMOCDOHepI protein from 1 L culture before and after purification

Sample	Soluble protein pre-purification (mg)	Soluble protein post (AKTA) purification (mg)
Culture 1	480.0	170.4

Culture 2	440.0	163.2
Culture 3	460.0	152.0
Average	460.0	161.9
SD	20.0	9.3

4.3. Summary

All expression constructs in this project had a 6x His-tag at either the N or C terminus to aid in downstream purification using IMAC. IMAC is advantageous as a purification technique at if offers a high selectivity and high capacity with milder buffer conditions compared to some other chromatography techniques (116). It can be used alone or as part of a multi-step approach to protein purification depending on the desired outcome or use of the product.

Five of the best performing vector / host combinations from the expression trials were chosen for small scale purification using metal IMAC beads. A culture volume of 35 ml was used which produced around 6.5 ml crude soluble lysate, of which 1 ml samples were purified. The vector / host combination which produced the highest yield of purified HepI protein was pNTHIS-SUMOCDOHepI + BL21 (DE3) at 4.0 mg (± 0.3). It was decided to use this vector / host combination for 1 L expression trials followed by FPLC purification. Over the three experiments an average of 161.9 mg (± 9.3) of purified HepI protein was produced from 1 L bacterial culture. The yield of soluble protein produced in this study exceeds that which is reported in the literature. When using SUMO as a fusion partner yields of between 1.4-1.8 mg/ml soluble protein are typical (69, 117), which is over 50% less than was produced in this project. When using other fusion partners such as GST, IF2 and MBP, yields range from 1.5, 1.4, and 1.3 respectively (74, 87, 118, 119) which is comparable to the yield observed in this study. This work, supported by the literature, indicates that SUMO is an optimal fusion partner to HepI and therefore, it could be hypothesised, that soluble HepI expression does not favour large fusion partners.

Chapter 5. Activity testing

5.1. Introduction

HepI cleaves, and therefore neutralises, the anticoagulant effect of unfractionated heparin, heparin sulphate, and low molecular weight heparin, with high specificity, to form oligo-, di- and mono-saccharides both *in vivo* and *in vitro* (149). One international unit (IU) of HepI is defined as the amount of enzyme that will liberate 1.0 µ mol unsaturated oligosaccharides from porcine mucosal heparin per minute at 37 °C and pH 7.0. There are several assays described in the literature to measure Hep I activity (150–153). The most widely used method is the 232 nm kinetic assay which measures the increase in ultra-violet (UV) absorbance, directly proportional to the production of uronic acid residues in the reaction (154). A limitation to this assay is that buffer components such as reducing or oxidising agents, detergents and salts interfere with absorbance measurements and mask any HepI activity being detected.

5.2. Kinetic Assay

The absorbance of the pure buffers used in this procedure were tested at 232 nm, these absorbances were compared to samples of different concentrations of a commercial HepI protein diluted in TBS. By determining the base absorbance of each buffer, or individual component, it was possible to establish at what stage in the experimental procedure, for example after lysis, or after purification, activity of the recombinant HepI protein could be accurately determined. The commercial HepI preparation in TBS measures an OD on average of 0.255 (± 0.004) at 0.05 U/ml, increasing to 0.322 (± 0.005) at 1 U/ml (**Figure 34**). TBS alone had an OD of 0.257 (± 0.004). The lysis buffer had an absorbance of 2.5 (maximum for the equipment) and so each component was tested separately, 1 mM DTT had an OD of 0.316 (± 0.004), 0.1 mM PMSF had an OD of 0.347 (± 0.003), the PIC diluted 1:500 in H_2O had an OD of 1.875 (± 0.046) and 1 % NP40 dissolved in H_2O absorbed over 2.5. The buffers A1, A2 and B2 from the FPLC protocol were tested, buffer A1 had an OD of 0.317 (±0.003), B1 which composed of A1 + 0.1 % NP40 had an OD of 2.19 (± 0.012), buffer B2 which composed A1 + 1 M imidazole had an OD of over 2.5, therefore samples of imidazole dissolved in H_2O between 1-150 mM were tested.
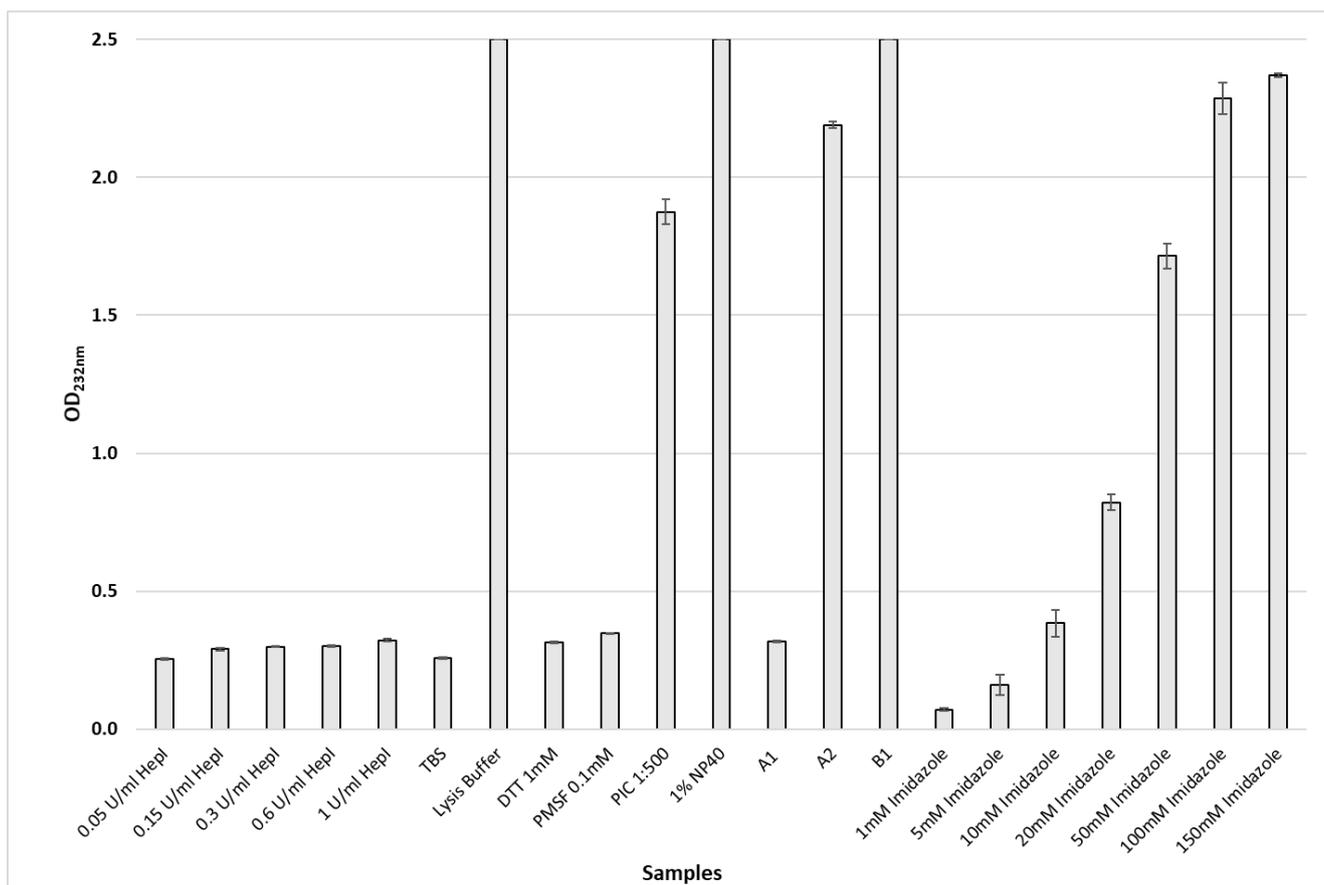
Figure 34: HepI activity interference testing. *The OD of several buffers / buffer components was measured at 232 nm; samples include increasing concentrations of commercial HepI, 0.05-1 U/ml in TBS. TBS (150 mM NaCl, 46.8 mM Tris HCL, 3.2 mM Tris Base, 5 mM CaCl$_2$, pH 7.2); Lysis buffer (300 mM NaCl, 25 mM Tris-HCl, 1 % NP40, 2 mM Ca^{2+}, 0.1 mM DTT, and protease inhibitor cocktail 1:500, pH7.9) A1 (300 mM NaCl, 50 mM Tris, 1 mM DTT, pH 7.9) A2 (A1 + 0.1 % NP40), B1 (A1+ 1M Imidazole). Each sample was measured in triplicate and the mean OD reported ±SD.*

To test if any of the buffers hindered HepI activity, 0.5 U/ml commercial HepI was added to each buffer, these solutions were then used as samples in the 232 nm kinetic assay against heparin, the change in absorbance measured over 180 seconds (ΔOD/Min) was used to calculate the amount of HepI activity in each sample. No activity could be determined from samples containing lysis buffer, 1 % NP40, A2 or B1 buffer. Samples containing 1 mM DTT, or A1 buffer overestimated the HepI activity by 32 % and 36 % respectively. Samples containing 0.1 mM PMSF underestimated the HepI activity by 12 % and PIC by 8 %. The most accurate determination of HepI activity with a 4 % over estimation was the sample containing TBS (**Table 14**).

Table 14: HepI activity interference testing. *0.5 U/ml of commercial HepI was added to various buffers, these were then used as samples in the kinetic 232 nm kinetic assay, the ΔOD/min was calculated and using the line equation from a calibration graph, the concentration of HepI present in the sample was determined from activity.*

Sample	ΔOD/Min	[Heparinase I] U/ml
Lysis Buffer	0.000	0.00
DTT 1 mM	0.341	0.66
PMSF 1 mM	0.230	0.44
PIC 1:500	0.243	0.46
A1	0.349	0.68
A2	0.000	0.00
B1	0.000	0.00
TBS	0.270	0.52

To test the activity of the recombinant HepI protein produced in this study a calibration of commercial HepI was carried out, where different concentrations of commercial HepI were tested against heparin in the assay. Studies suggest up to 1 U/ml HepI and between 25 and 2 mg/ml heparin should be used in the assay (90,112). Commercial HepI at concentrations of 0, 0.05, 0.15, 0.3, 0.6 and 1 U/ml were tested against 1 and 2 mg/ml heparin (equivalent of 150 and 300 IU). Testing showed no difference in the increase in absorbance between the higher and lower heparin concentrations, therefore it was decided to use 1 mg/ml for calibration testing. The activity of commercial HepI against 1 mg/ml heparin over 180 seconds was tested (**Figure 35**) data shows the average of five replicates.
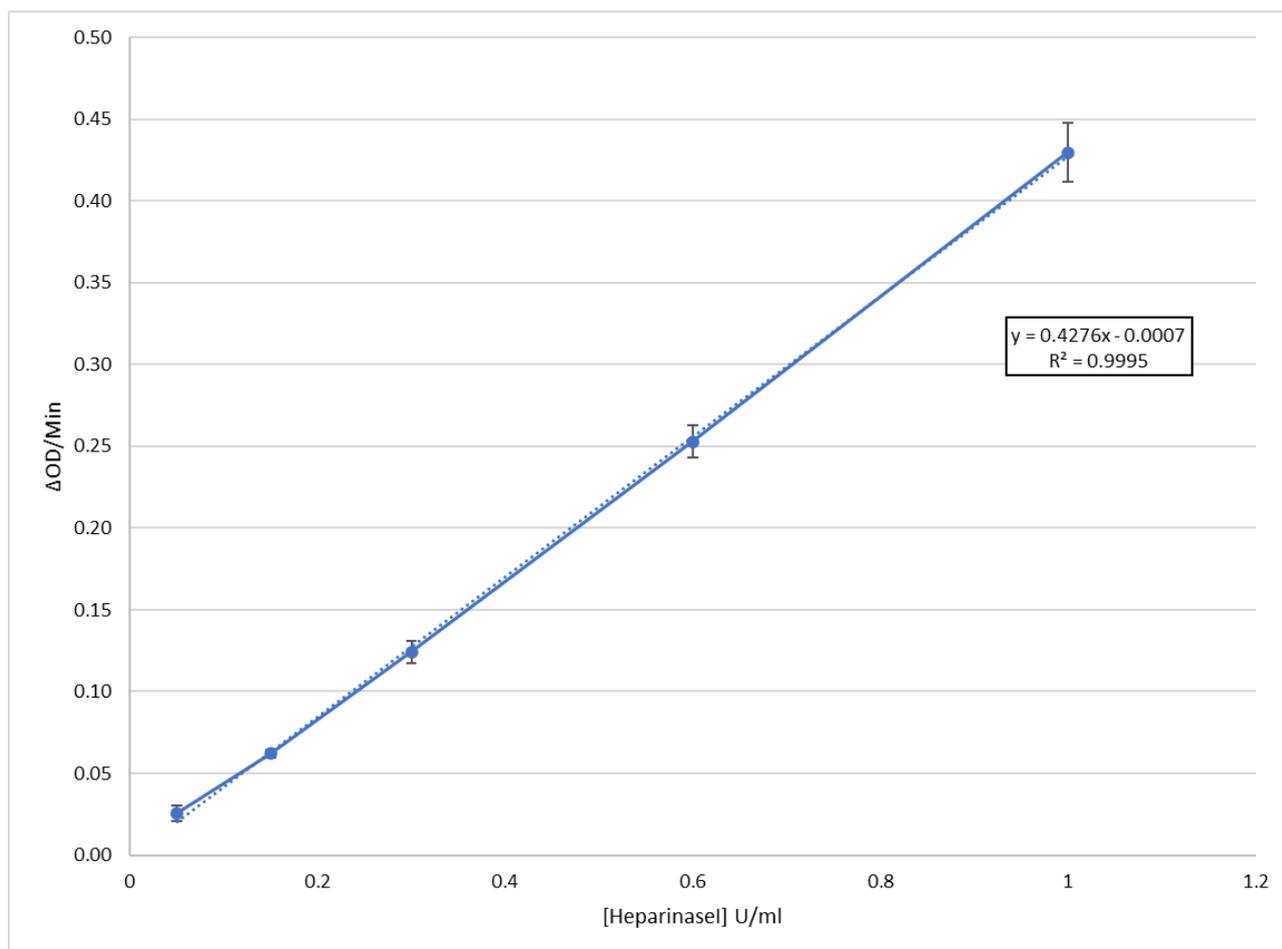
Figure 35: Activity of commercial Heparinase I. *Varying concentrations of Heparinase I (0.05-1 U/ml) was tested for activity against 1 mg/ml /150 U of heparin. OD was measured every 10 seconds for three minutes. Data shown is the mean ΔOD/min of five replicates.*

Protein samples from the IMAC and FPLC purification trials were tested for activity. Samples had to be passed through de-salting columns prior to testing to remove inhibitory substances such as imidazole and detergent.

From 35 ml culture samples, and after IMAC purification, 6.5 ml soluble protein and 1.3 ml purified protein per sample was produced (**Table 15**). Sample pCTHISHepI + BL21 (DE3) had no detectable enzyme activity, pNTHIS-TrxHepI + BL21 (DE3) had the highest activity at an average of 0.28 U/mg protein, which equates to 0.4 U in 35 ml culture or 11.30 U/L culture. From three separate 1L cultures, 200 ml soluble protein (each) and 80 ml purified protein (each) was produced. The specific activity of pNTHIS-SUMOCDOHepI + BL21 (DE3) HepI was 0.07 U/mg which equates to 11.4 U/L culture (**Table 16**).

Table 15: Activity testing of recombinant HepI samples after IMAC purification. *The activity of soluble HepI protein samples were tested using the 232 nm kinetic assay after being purified using IMAC magnetic beads. Samples were passed through a de-salting column to remove small molecule contaminants prior to testing. Activity testing was carried out in triplicate with the mean results reported ±SD.*

IMAC Purification Sample	Total Protein post purification (mg)	[Protein] post concentration (g/L)	Activity (U/ml)	Specific Activity (U/mg)	Total Units (35 ml Culture)	U/L Culture
pCTHISHepI + BL21 (DE3)	2.4 (± 1.3)	1.5 (± 0.7)	0	0.00	0.00	0.00
pNTHIS-TrxHepI + BL21 (DE3)	1.8 (± 0.9)	1.1 (± 0.5)	0.24 (±0.05)	0.28 (± 0.19)	0.40 (± 0.1)	11.30 (± 2.8)
pNTHIS-NusAHepI + JM109 (DE3) chaperone 4	1.5 (± 0.7)	0.9 (± 0.4)	0.07 (±0.09)	0.14 (± 0.23)	0.11 (± 0.18)	3.25 (± 5.23)
pNTHIS-SUMOHepI + BL21 (DE3)	3.2 (± 0.1)	2.0 (± 0.1)	0.18 (±0.14)	0.09 (± 0.09)	0.29 (± 0.28)	8.20 (± 8.13)
pNTHIS-SUMOCDOHepI + BL21 (DE3)	4.0 (± 0.3)	2.4 (± 0.2)	0.14 (±0.02)	0.06 (± 0.01)	0.23 (± 0.06)	6.65 (± 1.63)
pCTHISHepI + BL21 (DE3)	2.4 (± 1.3)	1.5 (± 0.7)	0	0.00	0.00	0.00

Table 16: Activity testing of recombinant HepI samples after FPLC purification. *The activity of soluble HepI protein samples were tested using the 232 nm kinetic assay after being purified using an AKTA Pure FPLC instrument. Samples were passed through a de-salting column to remove small molecule contaminants prior to testing. Activity testing was carried out in triplicate with the mean results reported ±SD.*

FPLC Purification Sample	Total protein post purification (mg)	[Protein] post purification (g/L)	Activity (U/ml)	Specific Activity (U/mg)	U/L Culture
pNTHIS-SUMOCDOHepI + BL21 (DE3)	161.9 (± 9.3)	2.0 (± 0.1)	0.14	0.07	11.4 (±3.1)

5.3. Summary

It is claimed in the literature that around 1.8 U/mg crude soluble protein and around 2.25 U/mg of purified HepI is achievable when expressed with a SUMO tag (146), however, in this study we were unable to replicate these findings despite following the same protocol for activity testing. Despite consistently expressing over 150 mg/L of soluble recombinant HepI protein, the activity was found to be around 16 times less as previously described which is a major concern. The 232 nm kinetic assay which is favoured for testing the activity of HepI in the literature has a disadvantage as it is hindered by many of the components used throughout the expression and purification process, this means that protein samples can only be tested after purification and removal of any contaminants such as imidazole, detergents, and PIC. It is possible to test HepI activity using a colourimetric assay, namely the Azure A / Toluidine Blue assay (114) but this is far less common than the 232 nm kinetic assay. The advantage to the Azure A / Toluidine Blue assay is that it is not hindered by imidazole or components of lysis buffer. However, this assay would need to be optimised and extensively tested before being used in this research.

The HepI samples which produced comparable results for activity at 11.30 (± 2.8) and 11.4 (± 3.1) U/L were N-terminal-His6-Trx-HepI and N-terminal-His6-SUMO-HepI (codon de-optimised) respectively. However, when compared to the commercial HepI formulation of 90-110 U/mg this is significantly lower and is therefore considered not to be commercially viable.

Chapter 6. Refolding

6.1. Introduction

As the soluble protein had little to no activity, it was assumed that it was misfolded during expression, and therefore biologically inactive. Because the protein would need to be denatured and then refolded under controlled conditions it was decided to maximise the yield of starting protein, i.e., express the HepI protein as inclusion bodies. Although denaturation and refolding increase the number of protocol steps, inclusion bodies can be advantageous for recombinant protein expression as they contain densely packed target proteins that are protected from proteolysis and degradation and are generally homogenous which reduces downstream purification steps (155). They also account for a higher level of expression of target protein compared to those in soluble form (156). As the autoinduction expression trial resulted in a high percentage of insoluble protein, it was decided to carry out an insoluble expression and refolding trial using HepI protein from the vector / host pCTHISHepI + BL21 (DE3) which produced the highest yield of insoluble protein during the AI trials (from now referred to as CTHISHepI) (

Figure 30). Three refolding methods were adopted and after either dialysis or de-salting, to remove inhibitory contaminants, the HepI samples were tested for activity using the 232 nm kinetic assay as previously described.

6.2. Results and Summary

The activity of recombinant *P. heparinus* HepI protein, produced from refolded inclusion bodies, in the literature varies from 114 U/mg or 1194 U/L culture by Ernst *et al* in 1996 (26), to 70 U/mg or 10,500 U/L by Shpigel *et al* in 1999 (91) which are significantly different. Both studies used the same expression system and activity assay but utilised different affinity tags which Shpigel *et al.* claimed to be the reason for the increase in expression in their study. Ernst *et al* expressed recombinant HepI with an N-terminal His-tag, whereas Shpigel *et al* constructed two vectors to enable the cloning and expression of HepI fused to the N- or C-terminus of cellulose-binding domain. They concluded that the C-terminal fusion of HepI to CBD was somewhat superior to the N-terminal fusion construct. The Ernst *et al* study produced approximately 67 mg of crude refolded inclusion bodies per 1 L culture and after further processing steps produced 10.23 mg of pure HepI protein per 1 L culture. The Shpigel *et al* study produced around 500 mg / L culture of crude inclusion bodies and after

refolding and purification produced 150 mg /L of pure HepI protein. It should be noted that the HepI protein produced in the afore-mentioned studies were by fermentation, which can increase recombinant protein expression by up to 20-fold compared to "shake flask culture" with IPTG induction in LB medium (157).

These studies have been used in comparison to this project as they used the International Unit of measurement for HepI activity, the HepI gene from the native host, expressed in *E. coli*. Other studies have been published which have investigated HepI expression and activity, however these are not comparable to this project. For example, Lohse and Linhardt (158) discuss expression of HepI from the native host, Kim *et al* (159) discuss a novel type of HepI purified from *Bacteroides stercoris* HJ-15, Nakamura *et al* (63) discuss the purification and properties of *Bacteroides heparinolyticus* heparinase. While a study by Yu and Wu (77) discuss recombinant expression of HepI in *E. coli* they do not use the recognised measurement of HepI activity but the 'relative activity', which is not clear how this compares to the IU of HepI.

As it was originally intended to express soluble HepI protein in this study, it was not possible, due to time constraints, to change the expression vectors or fusion tags or to optimise an insoluble expression protocol. Therefore, it was decided to use the insoluble material obtained during the AI expression trial (taken from - 80 °C storage) and repeat this method of expression if / when more protein was needed. Three methods of refolding were adopted and any method which produced over 1000 U/L culture in the initial trials was repeated and optimised.

The first trial to be conducted was the fast-refolding method, CTHISHepI protein at 1.4 mg/ml was de-natured in urea buffer and bound to nickel coated magnetic beads overnight, the bound protein was washed, refolded, and eluted. The specific activity of the refolded CTHISHepI protein was 3.21 U/mg (± 0.70) which equated to 68.85 (± 3.36) U/L of culture (**Table *17***).

Table 17: Fast refold trial. *Insoluble CTHISHepI protein was expressed using AI at a concentration of 1.4 mg/ml or 280 mg/L. The protein was de-natured and refolded as per the fast refold protocol. After refolding, the activity of the CTHISHepI was assessed using the 232 nm kinetic assay. Prior to activity testing the samples were passed through a de-salting column to remove any small molecule contaminants.*

Sample	Activity (U/ml)	Specific Activity (U/mg)	U/L culture
CTHISHepI	0.34 (± 0.02)	3.21 (± 0.70)	68.85 (± 3.36)

A second trial was conducted using a commercially available kit (Pierce Protein) which allowed for the testing of nine different refolding conditions; Insoluble CTHISHepI protein at 2 mg/ml (as per kit instructions) was denatured in guanidine buffer and refolded overnight. Refold conditions T3, T7 and T8 showed active HepI protein (**Table 18**), with T7 having the highest level of activity at 0.1 U/mg or 14.02 U/L culture. Samples T2 and T6 which effectively had 0 HepI activity were spiked with 0.5 U/ml commercial HepI protein and the activity measured. For sample T2 + 0.5 U/ml HepI activity was found to be 0.38 U/ml and for sample T6 + 0.5 U/ml HepI activity was found to be 0.47 U/ml which shows that the refolding buffer does not have an inhibitory effect on HepI activity. Sample T7 was snap frozen and stored at -80 °C, after 24 hours the sample was thawed and re-tested for activity, which was depleted to 0. The sample was spiked with 0.5 U/ml commercial HepI and the activity was found to be 0.48 U/ml which shows that the refolding buffer did not inhibit HepI activity, but the freeze-thawing process did. It is a common problem that freeze-thawing can deplete enzyme activity by up to 10 % per cycle (160) as freezing can induce several stresses on the protein causing it to denature such as freezing rate and thawing rate can induce cold denaturation (161), solute concentration changes due to the crystallisation of water will expose the protein to the ice-water interface (162), and crystallisation of buffer solutes cause precipitation and consequent pH changes (163). Much research has been reported emphasizing the use of cryoprotectants such as sugars or glycerol to increase the thermodynamic stability of proteins (164). However, the use of the protectants can sometimes be problematic in affecting the quality and purity of products. In this case, adding glycerol to the buffer solutions both during purification and before freezing did not improve activity of the enzyme after a freeze-thaw cycle which may be due to an increase in viscosity (data not shown). The commercial provider of HepI do not claim that glycerol is added to their formulation which suggests that HepI when expressed correctly is very stable during freeze-thaw cycles even without the addition of a cryoprotectant. An alternative method to freezing is to lyophilise (freeze-dry) protein samples for long-

term storage. This method is attractive as sample shelf-life is often increased, freeze-drying samples also reduces sample weight and volume, which can cut down on storage and shipping costs. Samples do not need to be transported on dry ice, as they are more stable at room temperature and reconstitution (usually with DI water) avoids denaturation caused by conventional defrosting methods (165). Freeze-drying the HepI samples was not practicable due to time limitations in this study, however, it is a promising avenue for further research.

Table 18: Kit refolding trial. *Insoluble CTHISHepI was expressed using AI at a concentration of 1.4 mg/ml or 280 mg/L. The protein was de-natured and refolded under nine different conditions (T1-T9) as per the Pierce Protein refolding kit protocol. After refolding, the activity of the CTHISHepI was assessed using the 232 nm kinetic assay. Samples showing 0 activity (T2 and T6) were spiked with 0.5 U/ml commercial HepI and the activity re-tested. Sample T7 was re-tested for activity after 24 hours, the sample was then spiked with 0.5 U/ml commercial HepI and the activity re-tested. Prior to activity testing the samples were passed through a de-salting column to remove any small molecule contaminants.*

Sample	Activity (U/ml)	Specific Activity (U/mg)	U/L culture
T1	0.00	0.00	0.00
T2	0.00	0.00	0.50
T3	0.01	0.08	10.84
T4	0.00	0.00	0.00
T5	0.00	0.00	0.00
T6	0.00	0.00	0.00
T7	0.01	0.10	14.02
T8	0.01	0.06	8.45
T9	0.00	0.00	0.00
T2 + 0.5U/ml HepI	0.38		
T6 + 0.5U/ml HepI	0.47		
T7 + 24hours	0.00		
T7 + 24hours + 0.5U/ml HepI	0.48		

The third refolding trial to be conducted was the slow refolding method. For small scale refolding the insoluble CTHISHepI protein was de-natured in urea buffer and bound to magnetic beads overnight. The bound (denatured) protein was eluted and slowly dialysed against decreasing concentrations of urea buffer to encourage slow refolding of the protein and to remove both urea and imidazole from the sample. CTHISHepI protein after refolding had activity of 14.57 U/mg or 1958.51 U/L culture (**Table 19**). The refolding protocol was repeated a second time and CTHISHepI protein showed a specific activity of 115.04 U/mg or 1876.38 U/L culture.

Table 19. Slow refold protocol using Ni coated magnetic beads. *Insoluble CTHISHepI protein was expressed using AI at a concentration of 1.4 mg/ml or 280 mg/L. The protein was de-natured in 8 M urea and bound to Ni coated magnetic beads. The de-natured protein was eluted using imidazole and then added to dialysis cartridges. The protein was dialysed in 100 volumes of buffer at 4 °C; 4 M Urea for four hours; 2 M urea overnight; 1 M urea for four hours; 0 urea overnight. After dialysis, the protein sample was tested for activity using the 232 nm kinetic assay. Results are reported as activity (U/ml), specific activity (U/mg) and total units per litre culture (U/L culture).*

Sample	Activity (U/ml)	Specific Activity (U/mg)	U/L culture
CTHISHepI round 1	2.04	14.57	1958.51
CTHISHepI round 2	1.95	15.04	1876.38

After activity was detected in two separate slow refolding trials using the magnetic beads, a refolding trial was conducted including a purification step using an AKTA FPLC machine. Activity results of a sample of CTHISHepI were determined after the inclusion bodies were de-natured in urea buffer, added to the FPLC column, eluted, and dialysed in refolding buffer containing decreasing amounts of urea. The specific activity of the CTHISHepI protein was 17.96 U/mg or 2514.98 U/L culture (**Table 20**).

Table 20: Slow refold trial using AKTA pure FPLC. *Insoluble CTHISHepI protein was expressed using AI at a concentration of 1.4 mg/ml or 280 mg/L. The protein was de-natured in 8 M urea and added to the FPLC column over 20 cycles. The de-natured protein was eluted using imidazole and then added to dialysis tubing. The protein was dialysed in 100 volumes of buffer at 4 °C; 4 M Urea for four hours; 2 M urea overnight; 1 M urea for four hours; 0 urea overnight. After dialysis, the protein sample was tested for activity using the 232 nm kinetic assay. Results are reported as activity (U/ml), specific activity (U/mg) and total units per litre culture (U/L culture).*

Sample	Activity (U/ml)	Specific Activity (U/mg)	U/L culture
CTHISHepI	2.51	17.96	2514.98

The results of the slow refolding trials show that from 280 mg/L of crude inclusion bodies, a yield of approximately 140 mg of purified CTHISHepI protein was achieved, which is a 50 % recovery rate and comparable to the yield reported by Shpigel *et al*.

Chapter 7. Discussion

Due to the clinical applications of HepI, demand for a pure product is high. Initially, however, separating HepI from heparinase II and III as well as other co-induced enzymes proved time consuming and laborious (26). These limitations led to the conclusion that a recombinant version of HepI, when expressed in a non-endogenous host would provide a raw material of high yield without contamination (90,91). A major disadvantage of expression in *E. coli* is that heparinase I aggregate into inclusion bodies within the host cell which leaves the enzyme biologically inactive (166). HepI has been expressed in recombinant form with solubility enhancing fusion partners and has shown to be active against heparin in spectrophotometric assays, it has not however been produced in such a yield that is commercially viable (146).

Initially, in this project we aimed to show that soluble HepI protein could be expressed in high yield, which was just as, if not more biologically active against UFH and LMWH than the current market standard preparation.

A library of expression vectors was created to maximise soluble expression of HepI in *E. coli*. A novel approach to soluble HepI expression was used whereby the HepI gene sequence was codon de-optimised for *E. coli* expression in the attempt to act as a translation brake and encourage correct folding of the protein.

To maximise soluble expression of recombinant HepI protein, seven expression vectors were created with various combinations of fusion partners and tag positions. The full-length version of HepI was also included in the vector library. All expression constructs in this study had a 6x His-tag at either the N or C terminus to aid in downstream purification using immobilised metal affinity chromatography (IMAC). The expression vectors were used to transform five host *E. coli* strains which have previously been shown to increase soluble expression of recombinant proteins. For expression cultures, two methods of induction were used, isopropyl β-d-1-thiogalactopyranoside (IPTG) and Auto-Induction (AI), as well as two induction temperatures. When induced at 37 °C recombinant HepI was expressed almost exclusively in the insoluble fraction, and it was thought that reducing the induction temperature in combination with a low concentration of IPTG would increase soluble expression of HepI protein (167). Reducing the IPTG concentration to 0.2 mM and expression temperature to 18 °C had a positive impact on soluble protein expression with over 50 % soluble expression being achieved with different vector and host combinations. The expression vector and host combination

which produced the highest yield of soluble protein on average was pNTHIS-SUMOCDOHepI in BL21 (DE3) at

2.9 g/L (**Table 11**). This vector/host combination created a SUMOHEPI fusion protein which had an N-terminal

His6. The HepI gene was deliberately de-optimised for *E. coli* expression to act as a translation brake.

The results of the AI trial showed that although there was an over expression of HepI protein, it was found

almost exclusively in the insoluble fraction. The expression trial using the full-length version of HepI resulted

in either no expression at all or inclusion body formation. Given the role of the leader sequence in the native

host, i.e., to facilitate movement of the protein to the periplasm for post translational modification and

secretion, it has been suggested that this could have been the case in this experiment. Unfortunately, it was

thought that secretion of the HepI protein would not happen in *E. coli* due to perceived differences in

transport/secretion pathways amongst bacterial species. Therefore, the absence of protein in the cellular

components was thought to have been because of another route cause. Upon reflection, it would have been

valuable to have tested the supernatant for the presence of HepI protein.

Five of the best performing vector / host combinations from the expression trials were chosen for small scale

purification using IMAC beads. A culture volume of 35 ml was used which produced around 6.5 ml crude soluble

lysate, of which 1 ml samples were purified. The vector / host combination which produced the highest yield

of purified HepI protein was again pNTHIS-SUMOCDOHepI in BL21 (DE3) at 4.0 mg (± 0.3). It was decided to

use this vector / host combination for 1 L expression trials followed by fast protein liquid chromatography

(FPLC) purification. Over the three experiments an average of 161.9 mg (± 9.3) of purified NTHISSUMOCDO

HepI protein was produced from 1 L bacterial culture.

It is claimed in the literature that around 1.8 U/mg crude soluble protein and around 2.25 U/mg of purified

HepI is achievable when expressed with a SUMO tag (146), however, this research was unable to replicate

these findings despite following the same protocol for activity testing. Despite consistently expressing over

150 mg/L of soluble recombinant HepI protein, the activity was found to be around 16 times less than that

which is claimed in the literature, which was a major concern. Because of these findings, it was decided to

follow an insoluble expression method along with refolding trials. As it was originally intended to express

soluble HepI protein, it was not possible due to time constraints to change the expression vectors or fusion

tags or to optimise an insoluble expression protocol. Therefore, it was decided to use the insoluble material obtained during the AI expression trial and solubilise it by different refolding methods.

The activity of recombinant *P. heparinus* HepI protein, produced from refolded inclusion bodies ranges from 114 U/mg or 1194 U/L culture by Ernst *et al* in 1996 (26), and 70 U/mg or 10,500 U/L by Shpigel *et al* in 1999 (91) which are significantly different. Both studies used the same expression system and activity assay but utilised different affinity tags which was claimed to be the reason for the increase in expression in the 1999 study. The Ernst *et al* study produced around 67 mg of crude refolded inclusion bodies per 1 L culture and after further processing steps produced 10.23 mg of pure HepI protein per 1 L culture. The 1999 study produced around 500 mg / L culture of crude inclusion bodies and after refolding and purification produced 150 mg /L of pure HepI protein. It should be noted that the HepI protein produced in the afore mentioned studies were by fermentation, which can increase recombinant protein expression by up to 20-fold compared to conventional "shake-flask culture" IPTG induction in LB medium (157).

Three methods of refolding were adopted in this study and any method which produced over 1000 U/L culture in the initial trials was repeated and optimised. The first trial to be conducted was the fast-refolding method as described in 2.1.2.3.1. In brief, insoluble HepI protein at a concentration of 1.4 mg/ml was denatured in urea buffer, bound to nickel coated magnetic beads overnight, washed, refolded, and eluted. The specific activity of the refolded CTHISHepI protein was 3.21 U/mg (± 0.70) which equated to 68.85 (± 3.36) U/L of culture.

A second trial was conducted using a commercial kit which allowed for the testing of 9 different refolding conditions. Insoluble CTHISHepI protein at 2 mg/ml was denatured in guanidine buffer and refolded overnight. Sample T7 had the highest level of activity at 0.1 U/mg or 14.02 U/L culture. Samples T2 and T6 which effectively had zero activity were spiked with 0.5 U/ml commercial HepI protein and activity was detected meaning that the refolding buffers do not have an inhibitory effect on HepI activity. After being snap frozen and re-thawed showed zero activity. The sample was spiked with 0.5 U/ml commercial HepI and the activity was found to be 0.48 U/ml which shows that the refolding buffer did not inhibit HepI activity, but the freeze-thawing process did. Freeze thawing of proteins can have a negative impact on stability and biological activity due to ice crystal formation, cold-denaturation, protein-solute interactions and protein-ice interactions (168).

Current methods of cryoprotection include lyophilisation or the use of osmolytes such as glycerol in the storage buffer. However, some disadvantage to using glycerol can be seen post-thawing and include sample viscosity, protein-protein interactions, and activity assay interference (169,170).

The third refolding trial to be conducted was the slow refolding method. For small scale refolding Insoluble CTHISHepI was de-natured in urea buffer and bound to magnetic beads overnight. The bound (denatured) protein was eluted and slowly dialysed against decreasing concentrations of urea buffer to encourage slow refolding of the protein and to remove both urea and imidazole from the sample. CTHISHepI after refolding had activity of 14.57 U/mg or 1958.51 U/L culture which was repeated a second time and produced comparable results.

After activity was detected in two separate slow refolding trials using the magnetic beads, a refolding trial was conducted using an AKTA FPLC machine. The specific activity of the CTHISHepI protein was 17.96 U/mg or 2514.98 U/L culture.

The results of the slow refolding trials show that from 280 mg/L of crude inclusion bodies, a yield of approximately 140 mg of purified recombinant HepI protein was achieved, which is a 50 % recovery rate and comparable to that seen in the literature (91).

7.1. Conclusion and future work

In this study we did show that a high yield of active recombinant HepI can be produced when it is expressed in the insoluble form and subjected to a slow refolding process. The results of the slow refolding trial are positive and would benefit from further optimisation to increase overall protein expression and recovery rate. If this avenue of testing was to be approached then it would make sense to create different expression vectors, as the ones created initially in this project contained fusion partners to increase soluble expression of the HepI protein and as such became superfluous. One possible fusion partner would be to use a part of the protamine sequence as an N-terminal or C-terminal fusion to act as a basic heparin anchor i.e., to increase the affinity of HepI for its heparin substrate and so increase the speed /efficiency of catalysis; this would be a novel approach to increasing HepI affinity to heparin as no literature is available for comparison. Full length protamine would likely have too high of an affinity to heparin and not allow the HepI to be processive i.e., cut-up heparin

fragments, release them and move onto the next heparin fragment. Protamine and heparin form an ionic bond which results in the formation of an inactive salt, which is metabolised by the liver/kidneys so any HepI-protamine complex would be removed from the blood before having any significant neutralisation effect (171,172). Creating a small protamine anchor may gave a balance of extra heparin affinity whilst allowing the enzyme to process other heparin fragments.

It would be interesting to explore the impact of glycosylation on the HepI protein, as the commercial provider (IBEX) now claim that their protein is glycosylated as it is in the native host (173,174). Historically, protein glycosylation was believed to be restricted to prokaryotes, but it is now well established that bacteria and even archaea also generate N- and O-linked glycosylated proteins (174–176), however this is usually related to pathogenicity (177). As expression in *E. coli* is typically limited to the production of non-glycosylated proteins it would be beneficial to utilise eukaryotic expression systems including plant and yeast cells, for example, *Nicotiana benthamiana* and *Pichia pastoris*, insect expression systems such as *baculovirus*, and various mammalian cells including Chinese Hamster Ovary or HeLa. Insect expression system allows for expression of high-quality glycosylated proteins that cannot be expressed using *E. coli* as they are a higher (eukaryotic) system and are therefore able to carry out more complex post translational modifications such as glycosylation (178). Mammalian cells are more complex than bacterial cells and have the most sophisticated protein production capabilities including mechanisms to assist protein folding into the proper conformation like chaperones and modifications (glycosylation, methylation, phosphorylation etc). Limitations to eukaryotic expression are however to be considered, such as high production cost, low yield, longer culture times, and demand for meticulous handling techniques (179). Several groups have reported an *E. coli*-based glycosylation platform which could alleviate the limitations found with eukaryotic expression (180–184).

Heparinases produced by *P. heparinum* are translocated to the periplasmic space and then post translationally modified to remove the N-terminal signal peptide. HepI is further modified by the addition of carbohydrate molecule, however there is little information on what benefit this may serve as the bacterium is non-pathogenic. Enzymatic digestion, NMR, and mass spectrometry has been used to establish the presence of an *O*-linked branched hepta-saccharide with a molecular mass of 1,161 Da at Ser[39] (after signal peptide removal) (**Figure *36***) (75).
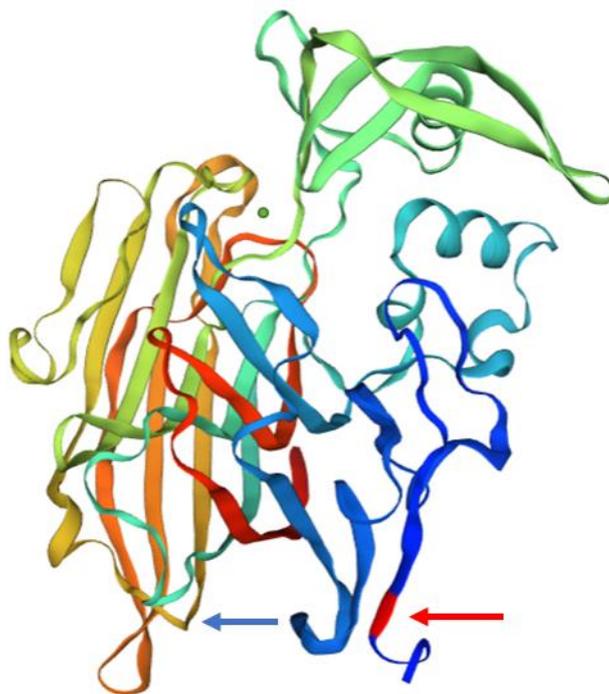
Figure 36: 3D structure of HepI protein and location of O-linked carbohydrate molecule. *The N-terminus of the protein is coloured in blue, through to the C-terminus coloured in red. The position of the branched O-linked hepta-saccharide at Ser^{39} is indicated by a red arrow / red block and Cys^{297} is indicated by a blue arrow. A calcium binding site is depicted by a green dot. The 3D structure was created using Swiss-Model prediction tool available at **https://swissmodel.expasy.org/**.*

Sasisekharan *et al* carried out a number of structural and functional studies of HepI from *F. heparinus* using chemical modifications and site-directed mutagenesis. They identified a number of residues which they proposed to play an essential role in heparin binding and its subsequent catalysis. They concluded that residues 195-221 constituted the primary heparin-binding site (185), and cysteine[135], histidine[203] and two putative calcium binding sites are essential for enzymatic activity (80,139,186).

A second cystine residue was identified at position 297 (**Figure 36**) and is part of the hydrophobic core and thought to play no role in activity (139) and could be the reason that the protein does not fold correctly when expressed as a recombinant version. If the two cysteine residues were to come in contact during folding / refolding they could form an internal di-sulfide bond or bind to another Cys^{297} in a neighbouring HepI protein. One suggestion would be to change Cys^{297} for a non-reactive alanine or serine which has shown to have a positive effect on improving total recovered protein percentage (187,188) These site-directed mutagenetic studies have shown activity of HepI can be increased by 30.06-122.05 % (74,187,188) compared to the native protein which is a promising avenue for further study.

A non-reducing SDS-PAGE analysis would show if dimerization (and trimerization) is occurring during the refolding process which could leave the enzyme inactive (189). Adding DTT to the refolding process could help to prevent the cysteines from creating a disulphide bond until the protein is in the correct conformational shape (190).

Structural analysis of HepI has been thus far based on bioinformatic modelling as demonstrated by Han *et al* (**Figure 5**), and in this project Figure *36*. There is no evidence of a crystal structure of HepI from *P. Heparinus.* There is however a crystal structure of HepI produced by *Bacteroides thetaiotaomicron* and overlapping the model of *P. Heparinus* HepI and the known structure of B**.** *thetaiotaomicron* HepI (**Figure 37***)* Zhang *et al* show that the two proteins share over 62.5 % homology (71).
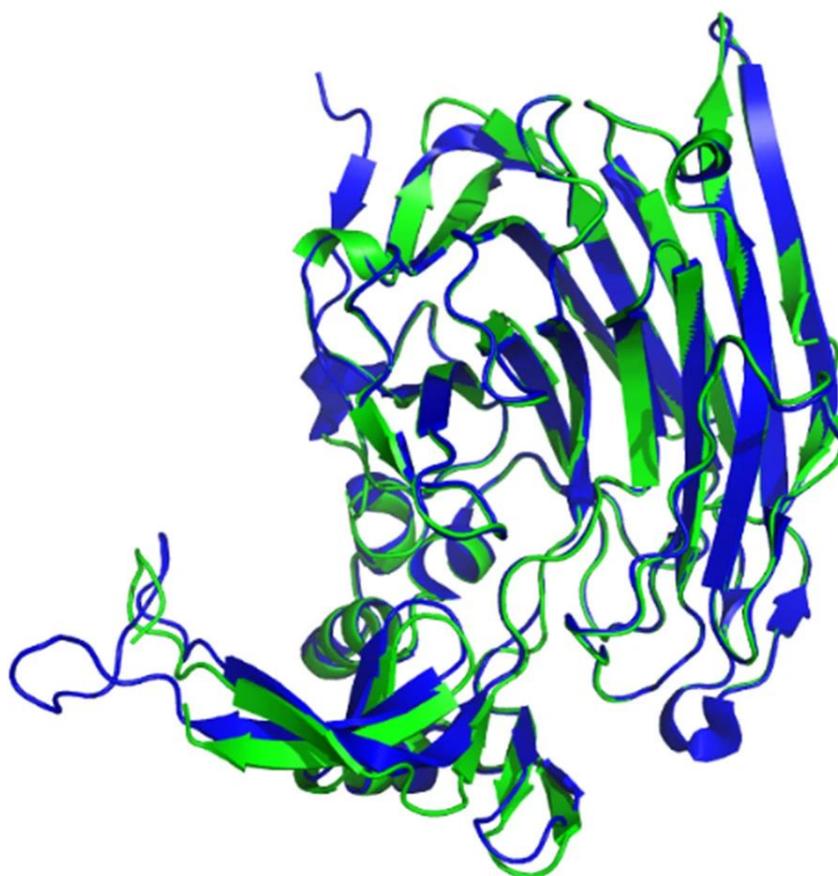
Figure 37: **The structure of HepI.** *Structure comparison of P. heparinus HepI (green) is superimposed with the template from Bacteroides thetaiotaomicron (blue). Picture credited to Zhang et al **(71)**.*

The 232 nm kinetic assay which is the preferred method for testing HepI activity was a disadvantage in this study as it is hindered many of the buffer components used throughout the expression and purification process. This meant that protein samples could only be tested for activity after purification. It is possible to

test HepI activity using a colourimetric assay, namely the Azure A / Toluidine Blue assay (114) but this is far less common than the 232 nm kinetic assay. The advantage to the Azure A / Toluidine Blue assay is that it is not hindered by imidazole or components of lysis buffer. It would be useful to optimise this assay for any future work.

In conclusion, ultimately, we failed to produce a highly active, soluble expressed, recombinant version of Heparinase I protein. However, promising results were generated when expressing HepI in the insoluble form and carrying out refolding steps. There are also many avenues available for further novel work to be carried out.

References

1.	Adams RLC, Bird RJ. Review article: Coagulation cascade and therapeutics update: Relevance to nephrology. Part 1: Overview of coagulation, thrombophilias and history of anticoagulants. *Nephrology*. 2009; 14 (5): 462–70.
2.	Rasche H. Haemostasis and thrombosis: an overview. *Eur Heart J Suppl*. 2001; 3: Q3–7.
3.	Wu KK, Thiagarajan P. Role of endothelium in thrombosis and hemostasis. *Annu Rev Med*. 1996; 47: 315–31.
4.	May A, Neumann F-J, Preissner K. The Relevance of Blood Cell-Vessel Wall Adhesive Interactions for Vascular Thrombotic Disease. *Thromb Haemost*. 1999; 82(08): 962–70.
5.	Preissner KT. Hemostatic Protease Receptors and Endothelial Cell Function: Insights from Gene Targeting in Mice. Semin *Thromb Hemost*. 2000; 26(05): 451–62.
6.	Yun S-H, Sim E-H, Goh R-Y, *et al*. Platelet Activation: The Mechanisms and Potential Biomarkers. *BioMed Res Int*. 2016: 9060143.
7.	Dahlbäck B, Villoutreix BO. Regulation of Blood Coagulation by the Protein C Anticoagulant Pathway: Novel Insights Into Structure-Function Relationships and Molecular Recognition. *Arterioscler Thromb Vasc Biol*. 2005; 25 (7): 1311–20.
8.	Palta S, Saroa R, Palta A. Overview of the coagulation system. *Indian J Anaesth.* 2014; 58 (5): 515–23.
9.	Madhusudhan T, Kerlin BA, Isermann B. The emerging role of coagulation proteases in kidney disease. *Nat Rev Nephrol*. 2016;12 (2): 94–109.
10.	Ezihe-Ejiofor JA, Hutchinson N. Anticlotting mechanisms 1: physiology and pathology. *Contin Educ Anaesth Crit Care Pain*. 2013; 13 (3): 87–92.
11.	Ustinov NB, Zav'yalova EG, Kopylov AM. Effect of thrombin inhibitors on positive feedback in the coagulation cascade. *Biochem Mosc.* 2016; 81(3): 242–8.
12.	Chapin JC, Hajjar KA. Fibrinolysis and the control of blood coagulation. *Blood Rev.* 2015; 29(1): 17–24.
13.	Odegård OR, Abildgaard U. Antithrombin III: critical review of assay methods. Significance of variations in health and disease. *Haemostasis.* 1978;7(2–3):127–34.
14.	Li Y-H, Kuo C-H, Shi G-Y, *et al*. The role of thrombomodulin lectin-like domain in inflammation. *J Biomed Sci*. 2012; 19(1): 34.
15.	Bick RL, Kaplan H. Syndromes Of Thrombosis And Hypercoagulability. *Medical Clinics of North America*. 1998; 82 (3): 409-458.
16.	Ramzi DW, Leeper KV. DVT and pulmonary embolism: Part I. Diagnosis. *Am Fam Physician.* 2004; 69(12): 2829-36.
17.	Tapson VF. Acute pulmonary embolism. *N Engl J Med*. 2008; 358(10): 1037-52.
18.	Stein PD, Matta F, Musani MH, Diaczok B. Silent pulmonary embolism in patients with deep venous thrombosis: a systematic review. *Am J Med*. 2010; 123(5): 426-31.
19.	Jiménez D. Simplification of the Pulmonary Embolism Severity Index for Prognostication in Patients With Acute Symptomatic Pulmonary Embolism. *Arch Intern Med.* 2010; 170 (15): 1383.
20.	Roberts LN, Whyte MB, Arya R. Pulmonary embolism mortality trends in the European region- too good to be true? *Lancet Respir Med.* 2020; 8(1): e2.
21.	Byrne JP, Geerts W, Mason SA, *et al*. Effectiveness of low-molecular-weight heparin versus unfractionated heparin to prevent pulmonary embolism following major trauma: A propensity-matched analysis. *J Trauma Acute Care Surg.* 2017; 82 (2): 252–62.
22.	Chen A, Stecker E, Warden BA. Direct Oral Anticoagulant Use: A Practical Guide to Common Clinical Challenges. *Journal of the American Heart Association.* 2020; 9(13).
23.	Vinogradova Y, Coupland C, Hill T, Hippisley-Cox J. Risks and benefits of direct oral anticoagulants versus warfarin in a real-world setting: cohort study in primary care. *BMJ*. 2018; 362: k2505.

24. Hemker HC. A century of heparin: past, present, and future. *J Thromb Haemost*. 2016;14(12):2329–38.

25. Lim GB. Discovery and purification of heparin. Nat Rev Cardiol. 2017. doi: 10.1038/nrcardio.2017.171. Epub ahead of print.

26. Ernst S, Venkataraman G, Winkler S, Godavarti R, *et al*. Expression in Escherichia coli, purification, and characterization of heparinase I from Flavobacterium heparinum. *Biochem J.* 1996; 315 (Pt 2): 589–97.

27. Bernfield M, Götte M, Park PW, *et al*. Functions of Cell Surface Heparan Sulfate Proteoglycans. A*nnu Rev Biochem.* 1999; 68 (1): 729–77.

28. Esko JD, Selleck SB. Order Out of Chaos: Assembly of Ligand Binding Sites in Heparan Sulfate. *Annu Rev Biochem*. 2002; 71 (1): 435–71.

29. Mailer RK, Kuta P, Renné T. An Update on Safe Anticoagulation. *Hämostaseologie*. 2022; 42(1): 65–72.

30. Ling L, Camilleri ET, Helledie T, *et al.* Effect of heparin on the biological properties and molecular signature of human mesenchymal stem cells. *Gene.* 2016; 576(1 Pt 2): 292–303.

31. Xu X, Dai Y. Heparin: an intervenor in cell communication. *J Cell Mol Med.* 2010; 14(1–2):175–80.

32. Gasimli L, Linhardt RJ, Dordick JS. Proteoglycans in stem cells. *Biotechnol Appl Biochem*. 2012; 59(2): 65–76.

33. Sasisekharan R. Heparin and heparan sulfate: biosynthesis, structure, and function. *Curr Opin Chem Biol.* 2000; 4 (6): 626–31.

34. Rabenstein DL. Heparin and heparan sulfate: structure and function. *Nat Prod Rep.* 2002; 19 (3): 312–31.

35. Choi S, Clements DJ, Pophristic V, *et al*. The Design and Evaluation of Heparin-Binding Foldamers. *Angew Chem Int Ed*. 2005; 44 (41): 6685–9.

36. Shriver Z, Capila I, Venkataraman G, Sasisekharan R. Heparin and Heparan Sulfate: Analyzing Structure and Microheterogeneity. Handb Exp Pharmacol. 2012; (207): 159–76.

37. Hirsh J, Anand SS, Halperin JL, Fuster V. Guide to Anticoagulant Therapy: Heparin: A Statement for Healthcare Professionals From the American Heart Association. Circulation. 2001; 103(24): 2994–3018.

38. Riley TV. Heparinase production by anaerobic bacteria. J Clin Pathol. 1987; 40(4): 384–6.

39. Maaroufi R. Mechanism of thrombin inhibition by antithrombin and heparin cofactor II in the presence of heparin. *Biomaterials.* 1997; 18(3): 203–11.

40. Hirsh J, Anand SS., Halperin JL., Fuster V. Mechanism of Action and Pharmacology of Unfractionated Heparin. *Arterioscler Thromb Vasc Biol*. 2001; 21 (7): 1094–6.

41. Oduah EI, Linhardt RJ, Sharfstein ST. Heparin: Past, Present, and Future. *Pharmaceuticals*. 2016; 9 (3): 38.

42. Smythe MA, Priziola J, Dobesh PP, *et al*. Guidance for the practical management of the heparin anticoagulants in the treatment of venous thromboembolism. *J Thromb Thrombolysis.* 2016; 41: 165.

43. Chen CC, You JY, Ho Ch. The aPTT assay as a monitor of heparin anticoagulation efficacy in clinical settings. *Adv Ther*. 2003; 20 (5): 231-6

44. Stafford-Smith M, Lefrak EA, Qazi AG. Efficacy and Safety of Heparinase I Versus Protamine in Patients Undergoing Coronary Artery Bypass Grafting with and without Cardiopulmonary Bypass: *Surv Anesthesiol.* 2006; 50 (2): 61–2.

45. Wolzt M, Weltermann A, Nieszpaur-Los M, *et al*. Studies on the neutralizing effects of protamine on unfractionated and low molecular weight heparin (Fragmin) at the site of activation of the coagulation system in man. *Thromb Haemost.* 1995; 73 (3): 439-43.

46. Porsche R, Brenner ZR. Allergy to protamine sulfate. *Heart Lung J Cardiopulm Acute Care.* 1999; 28 (6): 418–28.

47. Teoh KHT, Young E, Blackall MH, *et al.* Can extra protamine eliminate heparin rebound following cardiopulmonary bypass surgery? *J Thorac Cardiovasc Surg.* 2004; 128 (2): 211-9.

48. Jerath A, Srinivas C, Vegas A, *et al*. The successful management of severe protamine-induced pulmonary hypertension using inhaled prostacyclin. *Anesth Analg.* 2010; 110 (2): 365-9

49. Welsby IJ, Newman MF, Phillips-Bute B, *et al*. Hemodynamic Changes after Protamine Administration Association with Mortality after Coronary Artery Bypass Surgery. *Anesthesiology*. 2005; 102 (2): 308–14.

50. Chu Y-Q, Cai L-J, Jiang D-C, *et al*. Allergic shock and death associated with protamine administration in a diabetic patient. *Clin Ther.* 2010; 32(10): 1729–32.

51. Haastrup MB, Henriksen JE, Mortz CG, *et al.* Insulin allergy can be successfully managed by a systematic approach. *Clin Transl Allergy*. 2018; 8(1): 35.

52. Singla A, Sullivan MJ, Lee G, *et al*. Protamine-induced immune thrombocytopenia: Protamine-Induced Immune Thrombocytopenia. *Transfusion*. 2013; 53 (10): 2158-63.

53. Griffin MJ, Rinder HM, Smith BR, *et al*. The effects of heparin, protamine, and heparin/protamine reversal on platelet function under conditions of arterial shear stress. *Anesth Analg*. 2001; 93 (1): 20-7.

54. Mittermayr M, Margreiter J, Velik-Salchner C, *et al*. Effects of protamine and heparin can be detected and easily differentiated by modified thrombelastography (Rotem®): an in vitro study. *BJA Br J Anaesth.* 2005; 95 (3): 310–6.

55. van Veen JJ, Maclean RM, Hampton KK, *et al*. Protamine reversal of low molecular weight heparin: clinically effective? *Blood Coagul Fibrinolysis*. 2011; 22 (7): 565-70.

56. Tripathi CKM, Banga J, Mishra V. Microbial heparin/heparan sulphate lyases: potential and applications. *Appl Microbiol Biotechnol*. 2012; 94(2): 307–21.

57. Hong S-W, Shin H-Y, Kim YS, *et al*. Purification, and characterization of novel salt-active acharan sulfate lyase from *Bacteroides stercoris* HJ-15*. Eur J Biochem*. 2003; 270 (15): 3168–73.

58. Hong S-W, Shin H-Y, Kim YS, Kim D-H. Purification and characterization of novel salt-active acharan sulfate lyase from Bacteroides stercoris HJ-15. Eur J Biochem. 2003;270(15):3168–73.

59. Singh V, Haque S, Kumari V, *et al*. Isolation, Purification, and Characterization of Heparinase from *Streptomyces variabilis* MTCC 12266. *Sci Rep*. 2019; 9 (1): 6482.

60. Banga J, Tripathi CK. Rapid purification and characterization of a novel heparin degrading enzyme from Acinetobacter calcoaceticus. N Biotechnol. 2009 Oct 1;26(1-2):99-104.

61. Yoshida E, Sakai K, Tokuyama S, Miyazono H, Maruyama H, Morikawa K, Yoshida K, Tahara Y. Purification and characterization of heparinase that degrades both heparin and heparan sulfate from Bacillus circulans. Biosci Biotechnol Biochem. 2002 May;66(5):1181-4.

62. Banga J, Tripathi CKM, Bihari V. Growth and enzyme production kinetics of a heparinase-producing fungal isolate. *Med Chem Res*. 2008; 17 (2): 85–93.

63. Nakamura T, Shibata Y, Fujimura S. Purification and properties of *Bacteroides heparinolyticus* heparinase (heparin lyase, EC 4.2.2.7). *J Clin Microbiol.* 1988; 26 (5): 1070–1.

64. Dzvova N, Colmer-Hamood JA, Griswold JA, *et al*. Heparinase Is Essential for *Pseudomonas aeruginosa* Virulence during Thermal Injury and Infection. Bäumler AJ, editor. *Infect Immun*. 2017; 86(1): e00755-17.

65. Han C, Spring S, Lapidus A, *et al*. Complete genome sequence of Pedobacter heparinus type strain (HIM 762-3T). *Stand Genomic Sci*. 2009; 1 (1): 54–62.

66. Shriver Z, Sundaram M, Venkataraman G, *et al*. Cleavage of the antithrombin III binding site in heparin by heparinases and its implication in the generation of low molecular weight heparin. *Proc Natl Acad Sci.* 2000; 97 (19): 10365–70.

67. Bohlmann L, Chang C-W, Beacham I, von Itzstein M. Exploring Bacterial Heparinase II Activities with Defined Substrates. *ChemBioChem*. 2015; 16(8): 1205–11.

68. Bejoy J, Song L, Wang Z, Sang Q-X, Zhou Y, Li Y. Neuroprotective Activities of Heparin, Heparinase III, and Hyaluronic Acid on the Aβ42-Treated Forebrain Spheroids Derived from Human Stem Cells. ACS Biomater Sci Eng. 2018; 4(8): 2922–33.

69. Ernst S, Langer R, Cooney CL *et al*. Enzymatic Degradation of Glycosaminoglycans. *Critical Reviews in Biochemistry and Molecular Biology*. 1995; 30 (5): 387-444.

70. Xiao Z, Zhao W, Yang B, *et al*. Heparinase 1 selectivity for the 3,6-di-O-sulfo-2-deoxy-2-sulfamido-α-D-glucopyranose (1,4) 2-O-sulfo-α-L-idopyranosyluronic acid (GlcNS3S6S-IdoA2S) linkages. Glycobiology. 2011; 21(1): 13–22.

71. Zhang C, Yang B-C, Liu W-T, *et al*. Structure-based engineering of heparinase I with improved specific activity for degrading heparin. *BMC Biotechnol*. 2019; 19 (1): 59.

72. Limtiaco JFK, Beni S, Jones CJ, *et al*. NMR methods to monitor the enzymatic depolymerization of heparin. *Anal Bioanal Chem*. 2011; 399 (2): 593–603.

73. Han Y-H, Garron M-L, Kim H-Y, *et al*. Structural Snapshots of Heparin Depolymerization by Heparin Lyase I *. J Biol Chem. 2009 Dec 4;284(49):34019–27.

74. Godavarti R, Cooney CL, Langer R, Sasisekharan R. Heparinase I from Flavobacterium heparinum. Identification of a Critical Histidine Residue Essential for Catalysis As Probed by Chemical Modification and Site-Directed Mutagenesis. *Biochemistry*. 1996; 35(21): 6846–52.

75. Shaya D, Tocilj A, Li Y, *et al*. Crystal Structure of Heparinase II from Pedobacter heparinus and Its Complex with a Disaccharide Product *. *J Biol Chem.* 2006; 281(22): 15525–35.

76. Hashimoto W, Maruyama Y, Nakamichi Y, *et al*. Crystal Structure of Pedobacter heparinus Heparin Lyase Hep III with the Active Site in a Deep Cleft. *Biochemistry*. 2014; 53(4): 777–86.

77. Yu P, Wu Y. Expression of the heparinase gene from Flavobacterium heparinum in *Escherichia coli* and its enzymatic properties. *Carbohydr Polym*. 2012; 90(1): 348–52.

78. Zhang Q, Cao H-Y, Wei L, *et al*. Discovery of exolytic heparinases and their catalytic mechanism and potential application. *Nat Commun*. 2021; 12:1263.

79. Kuang Y, Xing X-H, Chen Y, *et al*. Production of heparin oligosaccharides by fusion protein of MBP–heparinase I and the enzyme thermostability. *J Mol Catal B Enzym.* 2006;1–4(43):90–5.

80. Liu D, Shriver Z, Godavarti R, Venkataraman G, Sasisekharan R. The calcium-binding sites of heparinase I from Flavobacterium heparinum are essential for enzymatic activity. J Biol Chem. 1999 Feb 12;274(7):4089–95.

81. Ma X, Wang Z, Li S, *et al. Ef*fect of CaCl2 as activity stabilizer on purification of heparinase I from *Flavobacterium heparinum. Journal of Chromatography B*. 2006; 843 (2): 209-215.

82. Heres EK, Horrow JC, Gravlee GP, *et al*. A Dose-Determining Trial of Heparinase-I (NeutralaseTM) for Heparin Neutralization in Coronary Artery Surgery. *Anesth Analg*. 2001; 93 (6): 1446–52.

83. Ammar T, Fisher CF. The Effects of Heparinase I and Protamine on Platelet Reactivity. *Anesthesiology.* 1997; 86 (6): 1382–6.

84. Kozek-Langenecker SA, Mohammad SF, Masaki T, *et al.* The Effects of Heparin, Protamine, and Heparinase I on Platelets in vitro Using Whole Blood Flow Cytometry. *Anesth Analg*. 2000; 90 (4): 808-12

85. Despotis GJ, Summerfield AL, Joist JH, *et al. In Vitro* Reversal of Heparin Effect with Heparinase: Evaluation with Whole Blood Prothrombin Time and Activated Partial Thromboplastin Time in Cardiac Surgical Patients. *Anesth Analg*. 1994; 79 (4): 670-4.

86. Maddineni J, Jeske WP, Iqbal O, *et al*. Differential Digestion of Different LMWHs by Heparinase-I and Heparinase-II: Drug Developmental Implications. *Blood*. 2004; 104 (11): 4084–4084.

87. Sommers CD, Ye H, Kolinski RE*, et al*. Characterization of currently marketed heparin products: analysis of molecular weight and heparinase-I digest patterns. *Anal Bioanal Chem*. 2011; 401 (8): 2445–54.

88. Xiao Z, Tappen BR, Ly M, *et al*. Heparin Mapping Using Heparin Lyases and the Generation of a Novel Low Molecular Weight Heparin. *J Med Chem*. 2011; 54 (2): 603–610

89. Yang VC, Linhardt RJ, Bernstein H, *et al*. Purification, and characterization of heparinase from *Flavobacterium heparinum. J Biol Chem.* 1985; 260 (3): 1849–57.

90. Sasisekharan R, Bulmer M, Moremen KW, *et al*. Cloning and expression of heparinase I gene from *Flavobacterium heparinum. Proc Natl Acad Sci*. 1993; 90 (8): 3660–4.

91. Shpigel E, Goldlust A, Efroni G, *et al*. Immobilization of recombinant heparinase I fused to cellulose-binding domain. *Biotechnol Bioeng*. 1999; 65 (1): 7.

92. Singh SM, Panda AK. Solubilization and refolding of bacterial inclusion body proteins. *J Biosci Bioeng*. 2005; 99 (4): 303–10.

93. Esposito D, Chatterjee DK. Enhancement of soluble protein expression through the use of fusion tags. *Curr Opin Biotechnol*. 2006; 17 (4): 353–8.

94. Butt TR, Edavettal SC, Hall JP, *et al*. SUMO fusion technology for difficult-to-express proteins. *Protein Expr Purif.* 2005; 43 (1): 1–9.

95. Marblestone JG, Edavettal SC, Lim Y, *et al.* Comparison of SUMO fusion technology with traditional gene fusion systems: Enhanced expression and solubility with SUMO. *Protein Sci.* 2006; 15 (1): 182-9.

96. Huang J, Cao L, Guo W, *et al*. Enhanced soluble expression of recombinant *Flavobacterium heparinum* heparinase I in *Escherichia coli* by fusing it with various soluble partners. *Protein Expr Purif*. 2012; 83 (2): 169–76. 80

97. Xu S, Qiu M, Zhang X, *et al*. Expression and characterization of an enhanced recombinant heparinase I with chitin binding domain. *Int J Biol Macromol*. 2017; 105: 1250–8.

98. IBEX Pharmaceuticals. Available from: https://www.ibex.ca/heparinase/

99. Chaperone plasmid set. Available from: https://www.takarabio.com/products/protein-research/expression-vectors-and-systems/protein-folding-kits/chaperone-plasmid-set.

100. Costa S, Almeida A, Castro A, *et al*. Fusion tags for protein solubility, purification, and immunogenicity in *Escherichia coli*: the novel Fh8 system. *Front Microbiol*. 2014; 5: 63

101. Luo Y, Huang X, McKeehan WL. High yield, purity, and activity of soluble recombinant *Bacteroides thetaiotaomicron* GST-heparinase I from *Escherichia coli**. *Arch Biochem Biophys.* 2007; 460 (1): 17–24.

102. BLAST: Basic Local Alignment Search Tool. Available from: https://blast.ncbi.nlm.nih.gov/Blast.cgi

103. Cloud-Based Informatics Platform for Life Sciences R&D. Benchling. Available from: https://www.benchling.com/

104. EMBOSS Needle Pairwise Sequence Alignment EMBL-EBI. Available from: https://www.ebi.ac.uk/Tools/psa/emboss_needle/

105. EMBOSS: sixpack. Available from: https://www.bioinformatics.nl/cgi-bin/emboss/sixpack

106. ExPASy - Translate tool. Available from: https://web.expasy.org/translate/

107. Ulvatne H. Proteases in *Escherichia coli* and *Staphylococcus aureus* confer reduced susceptibility to lactoferricin B. *J Antimicrob Chemother*. 2002; 50 (4): 461–7.

108. Poloz Y, Catalano A, O'Day DH. Bestatin Inhibits Cell Growth, Cell Division, and Spore Cell Differentiation in *Dictyostelium discoideum. Eukaryot Cell.* 2012; 11 (4): 545–57.

109. Matsumoto K, Mizoue K, Kitamura K, *et al*. Structural basis of inhibition of cysteine proteases by E-64 and its derivatives. *Pept Sci.* 1999; 51 (1): 99–107.

110. Marciniszyn Jr J, Hartsuck JA, Tang J. Pepstatin inhibition mechanism. *Adv Exp Med Biol*. 1977; 95: 199-210.

111. Matsumura Y, Hisaki K, Takaoka M, *et al.* Phosphoramidon, a metalloproteinase inhibitor, suppresses the hypertensive effect of big endothelin 1. *European Journal of Pharmacology*. 1990; 185 (1): 103-106.

112. Bernstein H, Yang VC, Cooney CL, *et al.* Immobilized heparin lyase system for blood deheparinization. *Methods in Enzymology*. 1988; 137: 515-529

113. Rosano GL, Ceccarelli EA. Recombinant protein expression in *Escherichia coli*: advances and challenges. *Front Microbiol*. 2014; 5: 172

114. Chen Y, Xing X-H, Lou K. Construction of recombinant *Escherichia coli* for over-production of soluble heparinase I by fusion to maltose-binding protein. *Biochem Eng J*. 2005; 23 (2): 155–9.

115. Aslantas Y, Surmeli NB. Effects of N-Terminal and C-Terminal Polyhistidine Tag on the Stability and Function of the Thermophilic P450 CYP119. *Bioinorg Chem Appl*. 2019; 2019: 1–8. 81

116. Gräslund S, Nordlund P, Weigelt J, *et al*. Protein production and purification. *Nat Methods*. 2008; 5 (2): 135–46.

117. Sabaty M, Grosse S, Adryanczyk G, *et al*. Detrimental effect of the 6 His C-terminal tag on YedY enzymatic activity and influence of the TAT signal sequence on YedY synthesis. *BMC Biochem.* 2013; 14 (1): 28.

118. Panek A, Pietrow O, Filipkowski P, *et al*. Effects of the polyhistidine tag on kinetics and other properties of trehalose synthase from *Deinococcus geothermalis. Acta Biochim Pol*. 2013; 60 (2).

119. Chen Y, Xing X-H, Ye F, *et al*. Production of MBP–HepA fusion protein in recombinant *Escherichia coli* by optimization of culture medium. *Biochem Eng J*. 2007; 34 (2): 114–21.

120. Terpe K. Overview of tag protein fusions: from molecular and biochemical fundamentals to commercial systems. *Appl Microbiol Biotechnol.* 2003; 60 (5): 523–33.

121. Peroutka Iii RJ, Orcutt SJ, Strickler JE, *et al.* SUMO fusion technology for enhanced protein expression and purification in prokaryotes and eukaryotes. *Methods Mol Biol Clifton NJ*. 2011; 705: 15–30.

122. Collet J-F, Messens J. Structure, Function, and Mechanism of Thioredoxin Proteins. *Antioxid Redox Signal*. 2010; 13 (8): 1205–16.

123. Arts IS, Vertommen D, Baldin F, Laloux G, Collet JF. Comprehensively Characterizing the Thioredoxin Interactome In Vivo Highlights the Central Role Played by This Ubiquitous Oxidoreductase in Redox Control. *Mol Cell Proteomics*. 2016; 15(6): 2125–40.

124. LaVallie ER, DiBlasio EA, Kovacic S, *et al.* A Thioredoxin Gene Fusion Expression System That Circumvents Inclusion Body Formation in the *E. coli* Cytoplasm. *Nat Biotechnol*. 1993; 11 (2): 187–93.

125. Kern R, Malki A, Holmgren A, Richarme G. Chaperone properties of *Escherichia coli* thioredoxin and thioredoxin reductase. *Biochem J*. 2003; 371 (Pt 3): 965–72.

126. Li K, Jiang T, Yu B, *et al. Escherichia coli* transcription termination factor NusA: heat-induced oligomerization and chaperone activity. *Sci Rep*. 2013; 3 (1): 2347.

127. Schmidt MC, Chamberlin MJ. nusA Protein of *Escherichia coli* is an efficient transcription termination factor for certain terminator sites. *J Mol Biol*. 1987; 195 (4): 809–18.

128. Harrison RG. Expression of soluble heterologous proteins via fusion with NusA protein. *Innovations*. 2013; 1: 4-7

129. Kurepa J, Walker JM, Smalle J, *et al*. The Small Ubiquitin-like Modifier (SUMO) Protein Modification System in Arabidopsis: ACCUMULATION OF SUMO1 AND -2 CONJUGATES IS INCREASED BY STRESS. *J Biol Chem.* 2003; 278 (9): 6862–72.

130. Dohmen JR. SUMO protein modification. *Biochim Biophys Acta BBA - Mol Cell Res*. 2004; 1695 (1–3):113–31.

131. Malakhov MP, Mattern MR, Malakhova OA, *et al*. SUMO fusions and SUMO-specific protease for efficient expression and purification of proteins. *J Struct Funct Genomics.* 2004; 5 (1–2): 75–86.

132. Zhou MY, Gomez-Sanchez CE. Universal TA cloning. *Curr Issues Mol Biol*. 2000; 2 (1): 1–7. 82

133. Angov E, Hillier CJ, Kincaid RL, *et al*. Heterologous Protein Expression Is Enhanced by Harmonizing the Codon Usage Frequencies of the Target Gene with those of the Expression Host. *PLoS ONE.* 2008; 14: 3-5.

134. Menzella HG. Comparison of two codon optimization strategies to enhance recombinant protein production in *Escherichia coli. Microb Cell Factories.* 2011; 10 (15).

135. Konczal J, Bower J, Gray CH. Re-introducing non-optimal synonymous codons into codon-optimized constructs enhances soluble recovery of recombinant proteins from *Escherichia coli. PLoS ONE.* 2019; 14: 4.

136. Desjardins P, Conklin D. NanoDrop Microvolume Quantitation of Nucleic Acids. *J Vis Exp JoVE.* 2010; 4: (45).

137. Vasina JA, Baneyx F. Expression of Aggregation-Prone Recombinant Proteins at Low Temperatures: A Comparative Study of the *Escherichia coli cspA* and tac Promoter Systems. *Protein Expr Purif.* 1997; 9 (2): 211-8.

138. Sivashanmugam A, Murray V, Cui C, *et al.* Practical protocols for production of very high yields of recombinant proteins using Escherichia coli. *Protein Sci Publ Protein Soc.* 2009; 18 (5): 936–48.

139. Sasisekharan R, Leckband D, Godavarti R, *et al.* Heparinase I from Flavobacterium heparinum: the role of the cysteine residue in catalysis as probed by chemical modification and site-directed mutagenesis. *Biochemistry.* 1995; 34 (44): 14441–8.

140. Godavarti R, Sasisekharan R A. A Comparative Analysis of the Primary Sequences and Characteristics of Heparinases I, II, and III from *Flavobacterium heparinum. Biochem Biophys Res Commun.* 1996; 229 (3): 770-7.

141. Kalinina AN, Borschevskaya LN, Gordeeva TL, *et al.* Construction of mutant heparinase I with significantly increased specific activity. *bioRxiv.* 2020; 1-3

142. EL-Baky NA, Linjawi MH, Redwan EM. Auto-induction expression of human consensus interferon-alpha in *Escherichia coli. BMC Biotechnol.* 2015; 15 (1): 14

143. Jia B, Jeon CO. High-throughput recombinant protein expression in Escherichia coli: current status and future perspectives. Open Biol. 2016; 6 (8): 160-196.

144. Li Z, Kessler W, van den Heuvel J, Rinas U. Simple defined autoinduction medium for high-level recombinant protein production using T7-based *Escherichia coli* expression systems. *Appl Microbiol Biotechnol.* 2011; 91(4): 1203–13.

145. Block H, Maertens B, Spriestersbach A, *et al.* Chapter 27 Immobilized-Metal Affinity Chromatography (IMAC). A Review. In: *Methods in Enzymology.* Academic Press. 2009; 463: 439-473.

146. Yang B-C, Zhang C, Wang C, *et al.* Soluble expression, and purification of heparinase I in Escherichia coli using a hexahistidine-tagged small ubiquitin-like modifier as a fusion partner. *Biotechnol Biotechnol Equip.* 2017; 31 (5): 1040–5.

147. Yin Chen, Xin-Hui Xing, Fengchun Ye, *et al.* Production of MBP–HepA fusion protein in recombinant *Escherichia coli* by optimization of culture medium. *Biochemical Engineering Journal.* 2007; 34 (2): 114-121.

148. Luo Y, Huang X, McKeehan WL. High yield, purity, and activity of soluble recombinant Bacteroides thetaiotaomicron GST-heparinase I from *Escherichia coli*. Arch Biochem Biophys.* 2007; 460 (1): 17–24.

149. Bhushan I, Alabbas A, Sistla JC. Heparin depolymerization by immobilized heparinase: A review. *Int J Biol Macromol.* 2017; 99: 721–30. 83

150. Zimmermann JJ, Langer R, Cooney CL. Specific plate assay for bacterial heparinase. *Appl Environ Microbiol.* 1990; 56 (11): 3593–4.

151. Rozenberg GI, Espada J, de Cidre LL*, et al.* Heparan sulfate, heparin, and heparinase activity detection on polyacrylamide gel electrophoresis using the fluorochrome tris (2,2'-bipyridine) ruthenium (II). *Electrophoresis.* 2001; 22 (1): 3–11.

152. Ahn SC, Kim BY, Oh WK, *et al.* Colorimetric heparinase assay for alternative anti-metastatic activity. *Life Sci.* 2006; 79 (17): 1661–5.

153. Schiemann S, Lühn S, Alban S. Development of both colorimetric and fluorescence heparinase activity assays using fondaparinux as substrate. *Anal Biochem.* 2012; 427 (1): 82–90.

154. Linhardt RJ, Loganathan D, Al-Hakim A, *et al.* Oligosaccharide mapping of low-molecular-weight heparins: structure and activity differences. *J Med Chem.* 1990; 33 (6): 1639–45.

155. Clark EDB. Refolding of recombinant proteins. *Curr Opin Biotechnol.* 1998; 9 (2): 157-63.

156. Ventura S, Villaverde A. Protein quality in bacterial inclusion bodies. *Trends Biotechnol*. 2006; 24 (4): 179–85.

157. Kim M, Elvin C, Brownlee A, *et al*. High yield expression of recombinant pro-resilin: Lactose-induced fermentation in *E. coli* and facile purification. *Protein Expr Purif.* 2007; 52 (1): 230-6.

158. Lohse DL, Linhardt RJ. Purification and characterization of heparin lyases from *Flavobacterium heparinum. J Biol Chem*. 1992 Dec 5; 267(34): 24347-55.

159. Kim B-T, Kim W-S, Shik Kim Y, Linhardt RJ, Kim D-H. Purification and Characterization of a Novel Heparinase from *Bacteroides stercoris HJ-151. J Biochem (Tokyo).* 2000; 128(2): 323–8.

160. Miura M, Jones TG, Hill PW. Freeze-thaw and dry-wet events reduce microbial extracellular enzyme activity, but not organic matter turnover in an agricultural grassland soil. *Applied Soil Ecology.* 2019; 44: 196-199,

161. Grabski AC. Chapter 18: Advances in Preparation of Biological Extracts for Protein Purification. In: *Methods in Enzymology*. Burgess RR, Deutscher MP, editors. Academic Press; 2009 p. 285–303.

162. D'Imprima E, Floris D, Joppe M, *et al*. Protein denaturation at the air-water interface and how to prevent it. *eLife.* 2019; 1 (8): 427-47.

163. Arsiccio A, Giorsello P, Marenco L, *et al*. Considerations on Protein Stability During Freezing and Its Impact on the Freeze-Drying Cycle: A Design Space *Approach. J Pharm Sci*. 2020; 109 (1): 464–75.

164. Fayter AER, Hasan M, Congdon TR, *et al*. Ice recrystallisation inhibiting polymers prevent irreversible protein aggregation during solvent-free cryopreservation as additives and as covalent polymer-protein conjugates. *Eur Polym J*. 2020; 140: 1100-36.

165. Molnar A, Lakat T, Hosszu A, *et al*. Lyophilization and homogenization of biological samples improves reproducibility and reduces standard deviation in molecular biology techniques. *Amino Acids*. 2021; 53 (6): 917–28.

166. Singh SM, Panda AK. Solubilization and refolding of bacterial inclusion body proteins. *J Biosci Bioeng*. 2005; 99(4): 303–10.

167. Kaur J, Kumar A, Kaur J. Strategies for optimization of heterologous protein expression in *E. coli:* Roadblocks and reinforcements. *Int J Biol Macromol*. 2018 Jan 1; 106: 803–22.

168. Garber Cohen IP, Castello PR, González Flecha FL. Ice-induced partial unfolding, and aggregation of an integral membrane protein. *Biochim Biophys Acta*. 2010; 1798(11): 2040–7.

169. Bhatnagar BS, Bogner RH, Pikal MJ. Protein stability during freezing: separation of stresses and mechanisms of protein stabilization. *Pharm Dev Technol*. 2007; 12(5): 505–23.

170. Mitchell DE, Fayter AER, Deller RC, *et al*. Ice-recrystallization inhibiting polymers protect proteins against freeze-stress and enable glycerol-free cryostorage. *Mater Horiz*. 2019; 6(2): 364–8.

171. Applefield D, Krishnan S. Protamine. In: StatPearls. *Treasure Island* (FL): StatPearls Publishing; 2021

172. Sokolowska E, Kalaska B, Miklosz J, *et al*. The toxicology of heparin reversal with protamine: past, present, and future. *Expert Opin Drug Metab Toxicol.* 2016; 12 (8): 897–909.

173. Iwashkiw JA, Vozza NF, Kinsella RL, *et al*. Pour some sugar on it: the expanding world of bacterial protein O-linked glycosylation. *Mol Microbiol.* 2013; 89 (1): 14–28.

174. Nothaft H, Szymanski CM. Protein glycosylation in bacteria: sweeter than ever. *Nat Rev Microbiol.* 2010; 8 (11): 765–78.

175. Nothaft H, Szymanski CM. Bacterial Protein N-Glycosylation: New Perspectives and Applications. *J Biol Chem*. 2013; 288 (10): 6912–20.

176. Szymanski CM, Wren BW. Protein glycosylation in bacterial mucosal pathogens. *Nat Rev Microbiol.* 2005; 3 (3): 225–37.

177. Feldman MF, Wacker M, Hernandez M, *et al*. Engineering N-linked protein glycosylation with diverse O antigen lipopolysaccharide structures in *Escherichia coli. Proc Natl Acad Sci.* 2005; 102 (8): 3016–21.

178. Liu F, Wu X, Li L, *et al*. Use of baculovirus expression system for generation of virus-like particles: Successes and challenges. *Protein Expr Purif*. 2013; 90 (2): 104–16.

179. Mueller P, Gauttam R, Raab N, *et al*. High-level in vivo mucin-type glycosylation in *Escherichia coli. Microb Cell Factories*. 2018; 17 (1): 168.

180. Prabhu SK, Yang Q, Tong X, *et al.* Exploring a combined *Escherichia coli-* based glycosylation and *in vitro* transglycosylation approach for expression of glycosylated interferon alpha. *Bioorg Med Chem*. 2021; 33: 116037.

181. Breyer CA, de Oliveira MA, Pessoa A. Expression of Glycosylated Proteins in Bacterial System and Purification by Affinity Chromatography. In: Picanço-Castro V., Swiech K. (eds) Recombinant Glycoprotein Production. *Methods in Molecular Biology*. 2018; 1674. Humana Press, New York, NY.

182. Prabhu SK, Yang Q, Tong X, *et al*. Exploring a combined *Escherichia coli*-based glycosylation and in vitro transglycosylation approach for expression of glycosylated interferon alpha. *Bioorg Med Chem*. 2021; 33: 116037. 85

183. Yates LE, Natarajan A, Li M, *et al*. Glyco-recoded *Escherichia coli*: Recombineering-based genome editing of native polysaccharide biosynthesis gene clusters. *Metab Eng*. 2019; 53: 59–68.

184. Keys TG, Aebi M. Engineering protein glycosylation in prokaryotes. *Curr Opin Syst Biol*. 2017; 5:23–31.

185. Sasisekharan R, Venkataraman G, Godavarti R, *et al*. Heparinase I from *Flavobacterium heparinum. J Biol Chem.* 1996; 271(6): 3124–31.

186. Godavarti R, Sasisekharan R. Heparinase I from *Flavobacterium heparinum*: ROLE OF POSITIVE CHARGE IN ENZYMATIC ACTIVITY *. *J Biol Chem*. 1998; 273(1): 248–55.

187. Chen S, Huang Z, Wu J, *et al*. Combination of site-directed mutagenesis and calcium ion addition for enhanced production of thermostable MBP-fused heparinase I in recombinant *Escherichia coli*. *Appl Microbiol Biotechnol*. 2013; 97(7): 2907–16.

188. Yu P, Jia T, Chen Y, *et al*. Improving the activity of heparinase I by the directed evolution, its enzymatic properties and optimal conditions for heparin degrading by recombinant cells. *Biochem Eng J*. 2016; 114: 237–43.

189. Rotoli SM, Caradonna SJ. Combining Non-reducing SDS-PAGE Analysis and Chemical Crosslinking to Detect Multimeric Complexes Stabilized by Disulfide Linkages in Mammalian Cells in Culture. J Vis Exp. 2019; (147): 59483.

190. Karyolaimos A, Ampah-Korsah H, Hillenaar T, *et al*. Enhancing Recombinant Protein Yields in the E. coli Periplasm by Combining Signal Peptide and Production Rate Screening. Front Microbiol. 2019; 10: 1511.

Appendix

Native Heparinase I gene sequence

ATG AAA AAG CAA ATT TTA TAT TTA ATT GTC CTT CAA CAA TTG TTC TTA TGC TCG GCC TAC GCA CAG CAA AAA

AAA TCC GGT AAC ATC CCT TAC CGG GTA AAT GTG CAG GCC GAC AGT GCT AAG CAG AAG GCG ATT ATT GAC

AAC AAA TGG GTG GCA GTA GGC ATC AAT AAA CCT TAT GCA TTA CAA TAT GAC GAT AAA CTG CGC TTT AAT

GGA AAA CCA TCC TAT CGC TTT GAG CTT AAA GCC GAA GAC AAT TCG CTT GAA GGT TAT GCT GCA GGA GAA

ACA AAG GGC CGT ACA GAA TTG TCG TAC AGC TAT GCA ACC ACC AAT GAT TTT AAG AAA TTT CCC CCA AGC

GTA TAC CAA AAT GCG CAA AAG CTA AAA ACC GTT TAT CAT TAC GGC AAA GGG ATT TGT GAA CAG GGG AGC

TCC CGC AGC TAT ACC TTT TCA GTG TAC ATA CCC TCC TCC TTC CCC GAC AAT GCG ACT ACT ATT TTT GCC CAA

TGG CAT GGT GCA CCC AGC AGA ACG CTT GTA GCT ACA CCA GAG GGA GAA ATT AAA ACA CTG AGC ATA GAA

GAG TTT TTG GCC TTA TAC GAC CGC ATG ATC TTC AAA AAA AAT ATC GCC CAT GAT AAA GTT GAA AAA AAA

GAT AAG GAC GGA AAA ATT ACT TAT GTA GCC GGA AAG CCA AAT GGC TGG AAG GTA GAA CAA GGT GGT TAT

CCC ACG CTG GCC TTT GGT TTT TCT AAA GGG TAT TTT TAC ATC AAG GCA AAC TCC GAC CGG CAG TGG CTT ACC

GAC AAA GCC GAC CGT AAC AAT GCC AAT CCC GAG AAT AGT GAA GTA ATG AAG CCC TAT TCC TCG GAA TAC

AAA ACT TCA ACC ATT GCC TAT AAA ATG CCC TTT GCC CAG TTC CCT AAA GAT TGC TGG ATT ACT TTT GAT GTC

GCC ATA GAC TGG ACG AAA TAT GGA AAA GAG GCC AAT ACA ATT TTG AAA CCC GGT AAG CTG GAT GTG ATG

ATG ACT TAT ACC AAG AAT AAG AAA CCA CAA AAA GCG CAT ATC GTA AAC CAG CAG GAA ATC CTG ATC GGA

CGT AAC GAT GAC GAT GGC TAT TAC TTC AAA TTT GGA ATT TAC AGG GTC GGT AAC AGC ACG GTC CCG GTT

ACT TAT AAC CTG AGC GGG TAC AGC GAA ACT GCC AGA TAG TAA GGA TCC

This translates to the following amino acid sequence:

MKKQILYLIVLQQLFLCSAYAQQKKSGNIPYRVNVQADSAKQKAIIDNKWVAVGINKPYALQYDDKLRFNGKPSYRFELKAED

NSLEGYAAGETKGRTELSYSYATTNDFKKFPPSVYQNAQKLKTVYHYGKGICEQGSSRSYTFSVYIPSSFPDNATTIFAQWHGA

PSRTLVATPEGEIKTLSIEEFLALYDRMIFKKNIAHDKVEKKDKDGKITYVAGKPNGWKVEQGGYPTLAFGFSKGYFYIKANSDR

QWLTDKADRNNANPENSEVMKPYSSEYKTSTIAYKMPFAQFPKDCWITFDVAIDWTKYGKEANTILKPGKLDV

MMTYTKNKKPQKAHIVNQQEILIGRNDDDGYYFKFGIYRVGNSTVPVTYNLSGYSETAR

Chapter 8. Abbreviations

ºC	Degrees Celsius
AI	Auto-Inducing Media
DI	De-ionised
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
dNTP	Deoxyribonucleotide triphosphate
DTT	Dithiothreitol
E. *coli*	*Escherichia coli*
EDTA	Ethylenediaminetetraacetic acid
F. heparinum	*Flavobacterium heparinum*
FPLC	Fast Protein Liquid Chromatograph
g/L	Gram per litre
HepI	Heparinase I
h	Hours
IDT	Integrated DNA technologies
L	Litre
LB	Luria-Bertani
M	Molar
µg	Microgram
mg/ml	Microgram per millilitre
MIB	Manchester Institute of Biotechnology
min	Minutes
ml	Millilitre
mM	Millimolar
µM	Micromolar

NEB	New England Biosciences
NMR	Nuclear magnetic resonance
Ni	Nickel
NusA	N-utilization substance
ORF	Open reading frame
PCR	Polymerase Chain Reaction
P. Heparanus	*Pedobacter heparinus*
PIC	Protease Inhibitor Cocktail
PMSF	Phenylmethylsulfonyl Fluoride
ROTEM	Rotational Thromboelastometry
s	Seconds
SDS	Sodium dodecyl sulfate
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
TAE	Tris, acetic acid and EDTA
TBS	Tris Buffered Saline
U	Unit
U/L	Units per Litre
U/mg	Units per milligram