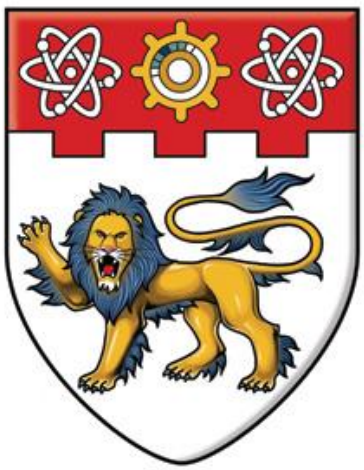

NANYANG TECHNOLOGICAL UNIVERSITY

DEVELOPMENT OF RATIONAL BIOPROCESS DESIGN STRATEGIES FOR A CLINICALLY RELEVANT PROTEIN CANDIDATE

ANINDYA BASU
SCHOOL OF CHEMICAL AND BIOMEDICAL ENGINEERING
2012

DEVELOPMENT OF RATIONAL BIOPROCESS DESIGN STRATEGIES FOR A CLINICALLY RELEVANT PROTEIN CANDIDATE

ANINDYA BASU

SCHOOL OF CHEMICAL AND BIOMEDICAL ENGINEERING

A thesis submitted to the Nanyang Technological University in partial fulfilment of the requirement for the degree of Doctor of Philosophy

Acknowledgement

I would like to express my heartfelt gratitude to my supervisor Dr. Susanna Leong Su Jan for allowing me to be a part of her team. I know no words to thank her for her expert guidance and unrestricted support that constantly kept me motivated during the toughest times of my project. Without her valuable suggestions, this thesis would not have been possible.

My special thanks and appreciation goes to Prof. William Chen for providing the plasmid, which formed the starting point of the current thesis.

I am highly indebted to Dr. Chen Yu, the senior member of my lab. Without his selfless helping hand and patient considerations to all my queries, this work could not have been possible.

I would also like to acknowledge the kind support from all my other lab-mates, especially Dr. A. K. R. Gobinath, Mr. Li Xiang, Mr. Derrick T.K. Sing and Dr. Foo Jee Loon for their unremitted help whenever I needed.

I would like to thank the School of Chemical and Biomedical Engineering, Nanyang Technological University, Singapore for providing me the scholarship without which this work would not have been possible.

Last but not the least I am grateful to each and every member of my family for their unconditional support and encouragement to keep me upbeat and enthusiastic to do a good work.

Contents

List of Abbreviations

6His	6-Histidine
APS	Ammonium persulphate
BSA	Bovine serum albumin
CQA	Critical to Quality Attributes
CV	Column Volume
DoE	Design of Experiments
DTT	Dithiothreitol
ELISA	Enzyme Linked Immuno-Sorbent Assay
FPLC	Fast Protein Liquid Chromatography
GSH	Glutathione
GST	Glutathione S Transferase
HBV	Hepatitis B Virus
HBx	Hepatitis B Virus X
HCC	Hepatocellular carcinoma
His	Histidine
HPLC	High performance liquid chromatography
IBs	Inclusion Bodies
IMAC	Immobilised Metal Affinity Chromatography
IPTG	Isopropyl-1-thio-β-D-galactoside
MBP	Maltose Binding Protein
nRTIs	nucleos(t)idic reverse transcriptase inhibitors
OD	Optical density
PBS	Phosphate buffered saline
pI	Isoelectric pH
QbD	Quality by Design
RP HPLC	Reverse phase high performance liquid chromatography
SDS	Sodium dodecyl sulphate
SDS PAGE	Sodium dodecyl sulphate poly-acryl amide gel electrophoresis
SEC	Size Exclusion Chromatography
SEC HPLC	Size Exclusion High Performance Liquid Chromatography
SLS	Static Light Scattering
SVC	Second Virial Coefficient
TEMED	Tetramethylethylenediamine
TFA	Tri-Fluoroacetic Acid

Abstract

The Hepatitis B Virus X (HBx) protein has been associated with the initiation and development of hepatocellular carcinoma (HCC), a killer disease affecting millions of lives worldwide. Being a multifunctional viral regulator, the HBx protein has been found to modulate all major host cellular metabolic pathways, causing scientists to hypothesize that it can be a potential drug target for the disease. However, the HBx protein is expressed at very low levels within the infected host cells. Increase in HBx yield can be achieved by recombinant production in bacteria host systems but this often results in the insoluble expression of the protein. The lack of pure bioactive HBx continues to hinder research progress to study the protein's structure-function and hence the development of new anti-HBx drug candidates. Moreover, the absence of native HBx also prevents quantitative bioactivity determination of the protein, making bioprocess design and scale-up studies difficult to perform. To overcome this roadblock, my research project aims to develop a scalable bioprocess for HBx production at amounts that are sufficient for subsequent structural characterisation and drug designing studies.

HBx expression and bioprocess development studies commenced with the use of a glutathione S transferase (GST) tagged HBx construct. The use of this construct, however, was challenged by inefficient Factor Xa cleavage and poor economics. Thereafter, a new 6His-HBx protein construct was designed which only comprised a 6-histidine (His)tag to ease recovery of the protein using immobilised metal affinity chromatography (IMAC). As the 6His-HBx protein was expressed as insoluble inclusion bodies (IBs), rational strategies employing second virial coefficient (SVC) measurements in combination with a Statistical design of experiments (DoE) platform were developed to facilitate rapid determination of optimal physicochemical conditions necessary to retain HBx solubility and stability. The SVC studies clearly indicated the importance of a net reducing environment combined with L-arginine (an aggregation inhibitor) for improved solubility of HBx, a highly hydrophobic protein with 9 cysteine residues. The SVC results guided the rational design of a HBx refolding buffer to maintain HBx in a stable soluble state, leading to the development of a dilution refolding based bioprocess for HBx. Further improvement of the process was impeded by the absence of an analytical platform to evaluate HBx refolding yields. To overcome this roadblock, a novel ELISA platform for HBx was subsequently developed to

quantitatively determine HBx refolding yields. The ELISA platform was designed based on the well-characterised interaction between HBx and the tumour suppressor protein, p53. GST-p53 molecules were immobilised on a glutathione (GSH) functionalised maleimide surface to avail the p53 ligand for binding with bioactive 6His-HBx. The amount of bioactive HBx bound to the p53 molecules was determined by generating a 6His-GST calibration curve, where the 6His-GSTproteins were also immobilised on the maleimide plate surface. Armed with a protein stabilising buffer and a quantitative analytical method, the stage was set for the design and development of an intensified process for large scale production of HBx. An optimum IMAC refolding-based bioprocess was then developed for the HBx protein using the Statistical DoE methodology, which readily achieved the production of bioactive HBx protein at 0.6 mg/ml and >95% purity (at an overall refolding yield of 54%). The DoE methodology is advantaged by its capability to provide detailed understanding of how process parameters like column incubation times, protein load and refolding buffer exchange gradients influenced the critical to quality attributes (CQA) of the target product. This process design approach also allows the attainment of data needed to generate the process design space, crucial for subsequent regulatory filing of the product.

In conclusion, using HBx as a model protein candidate, this thesis presents rational strategies that can be used for determining optimal refolding conditions for hydrophobic and cysteine-rich protein candidates. Furthermore, the HBx ELISA platform presented here is the first of its kind, where both the soluble and refolding yields of the protein could be individually determined. The successful development of an efficient bioprocess platform for HBx production reported in this thesis is expected to be readily adopted as a routine platform/flowsheet for the production of other novel protein candidates of biopharmaceutical relevance.

CHAPTER 1: Introduction

1. Introduction

HCC is a killer disease affecting over 350 million lives every year [3-4]. Though highly prevalent in Asia and sub-Saharan Africa [5-7], incidences of HCC are also on a steep rise in developed countries like Japan and United States [6, 8]. In fact, HCC is considered to be the fifth most common malignancy in the world and the third most common cause of cancer-related death worldwide [9]. Through intense research in this field, scientists have identified that Hepatitis B Virus (HBV) infection as a major causative factor for the initiation and development of the HCC within humans. Currently the seven anti-viral medications approved in the USA for treatment of this condition include (i) two type-1 interferons, and (ii) five nucleos(t)idic reverse transcriptase inhibitors (nRTIs), none of which has been found to completely eradicate the infection [10]. Furthermore, long term usage of these medications, particularly the nRTIs may lead to development of virus resistance and cross-tolerance, thereby rendering these strategies ineffective. Therefore, new drug therapeutic strategies are urgently needed to reduce patient suffering and mortality rates associated with HCC [11]. Over the years, different studies have indicated that the HBx protein expressed by the HBV is strongly associated with HCC development. Various studies have suggested the involvement of HBx in HBV transcription, replication and interaction with host cellular machineries to support its proliferation [12-14], leading to strong interests in developing HBx-targeted therapeutic strategies for HCC treatment. However, the real bottleneck in identifying a structure-based lead molecule against HBx is the lack of knowledge of the HBx structure [12], thereby making its structural characterization a difficult task even with *in silico* modelling and subsequent drug designing studies [15]. Furthermore, the low inherent expression of the protein within the infected host cells makes structural characterization of natively isolated proteins, a highly challenging endeavour. Biosynthesis of the HBx protein in high-yielding bacteria platforms thus seems to be an obvious solution. However, previous attempts to recombinantly produce the protein in different expression systems such as *E. coli*, have consistently resulted in the production of insoluble and inactive HBx [16-19] necessitating the development of a refolding based HBx bioprocess "[20]".

Apart from the protein insolubility problem, development of a bacterial expression platform for the HBx protein is not straightforward because of the following problems:

i. Due to the absence of a native protein standard, it would be difficult to verify structural characteristics of the recombinantly-derived protein.

ii. There are no analytical platforms reported to quantify HBx bioactivity, thus hindering bioprocess design and scale-up studies for the HBx protein.

The unmet demand for HBx merits detailed research into the development of a bioprocess capable of large scale production of the active protein, which forms the basis of this thesis. The outcome of this study is aimed to pave the way for the development of better targeted antibodies against HBx or the design of anti-HBx drug candidates. Production of the HBx protein in sufficient quantities will also allow easy structural characterization of the molecule for better drug designing opportunities.

Refolding the HBx protein is expected to be challenging because of its high hydrophobicity and the number of cysteine residues (i.e. between eight to ten in different variants) [21]. For HBx variants containing odd number cysteines, the tendency of the free cysteine moieties to interact covalently and form non-native conformations such as dimers or other misfolded conformations due to cysteine mispairing increases bioprocessing complexity. Therefore, the successful design of the HBx bioprocess demands a highly systematic and rational approach involving a thorough understanding of the protein behaviour under different physicochemical environments leading to rational bioprocess design, which will be researched in this thesis. The resulting bioprocess will not only meet rising demands for the HBx protein, but also other complex proteins that are rapidly being discovered in this post-genomic era [22-23].

1.1. Aims and objectives

The overall objective of this Ph.D. research project is to design a robust bioprocess that can be easily scaled up for the large scale production of the active HBx protein. To achieve this goal, a set of studies with the following specific aims were conducted:

- Studying the effect of a GST fusion partner on HBx expression, solubility behaviour and bioprocessing considerations.
- Refolding of HBx and physiochemical optimisation of refolding conditions for high bioactive protein yield guided by second virial coefficient measurements.
- Development of an ELISA bio-analytical platform for evaluating HBx soluble and refolding yields

- Development of a scalable and intensified chromatography refolding bioprocess aligned with Quality by Design (QbD) principles for improved HBx productivity.

The motivations of this research are stated below:

- The outcome of this research will form the basis for developing a generic bioprocess platform for the rapid, efficient, cost effective and scalable production of other HBx variants for structural and clinical studies of the protein.
- Studies from this project will also help in understanding how molecular complexity of a protein such as the presence of odd number of cysteine residues can influence protein refolding, thereby laying the foundations for a general approach for refolding processes of similar protein molecules.
- The adoption of rational and systematic approaches such as Statistical design of experiments (DoE) methodologies and QbD concepts to every stage of the process development studies will narrow existing knowledge gap related to designing scalable bioprocesses for novel protein based drug candidates which will be of significant value to the pharmaceutical industry.

1.2. Thesis Organization

The following chapter (i.e. Chapter 2) reviews the role of HBx in HCC development and discusses protein folding theories based on which current protein refolding strategies have been designed. Chapter 2 also reviews previous studies reported on the production of HBx protein, and discusses bioprocess limitations identified in those studies. Chapter 3 reports the studies of the effects of a GST tag on the solubility, yield and bioprocessing ease of the HBx protein. Chapter 4 reports the design of a new HBx protein construct (i.e. 6His-HBx) and the rational development of a dilution refolding process for this construct guided by static light scattering experiments using a Statistical DoE platform. Chapter 5 reports the development of an ELISA based analytical platform for quantitative soluble and refolding yield determination of 6His-HBx. Chapter 6 presents the development of a chromatography based refolding bioprocess for 6His-HBx following principles of QbD concepts. Future work and conclusions are presented in Chapter 7. A brief overview of the thesis flow rationale (Chapters 2-6) is presented in Figure 1-1.

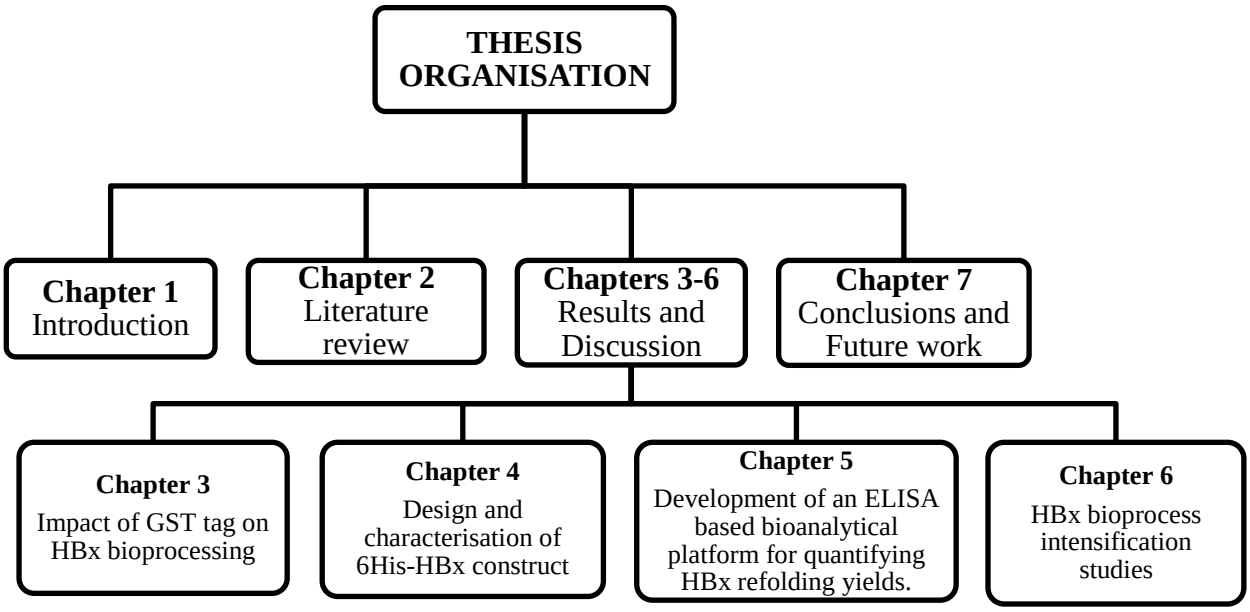

Figure 1-1: Thesis organisation.

1.3. List of publications

1. A. Basu, W.N. Chen, and S.S.J. Leong, (2011) *A rational design for hepatitis B virus X protein refolding and bioprocess development guided by second virial coefficient studies.* Applied Microbiology and Biotechnology, **90**(1): 181-191.

2. A. Basu, L. Xiang, S.S.J. Leong, (2011) *Refolding of proteins from protein inclusion bodies: Rational design and recipes.* Applied Microbiology and Biotechnology. **92**(2): 241-251.

3. A. Basu, S.S.J. Leong, (2011) *Development of an ELISA based analytical method for HBx refolding yield quantification.* Analytical Biochemistry. **418**(1): 155-157.

4. A. Basu, S.S.J. Leong, (2011-2) *High productivity chromatography refolding process for Hepatitis B Virus X (HBx) protein guided by Statistical Design of Experiment studies.* Journal of Chromatography A. **1223**(0): p. 64-71.

CHAPTER 2: Literature Review

Abstract

This chapter reviews the importance of the HBx protein in relation to HCC and presents some basic protein refolding concepts that are essential to rationally design a bio-manufacturing platform for the HBx protein. From a pharmacological perspective, the ability of the HBx protein to modulate all major host cellular controlling systems (including cell cycle, gene expression and defence systems), indicates its potential as a therapeutic target against HCC. However, since the protein is consistently produced as insoluble inclusion bodies in the bacterial host systems, bulk manufacturing of the protein would necessitate a refolding based bioprocessing strategy. To facilitate the development of a scalable HBx bioprocess, a brief overview of the different theories related to protein folding mechanisms are discussed in this chapter.

The protein folding phenomenon has been explained using different models to quantitatively and qualitatively describe the events leading to the successful folding of a protein from its denatured-reduced form. From a protein refolding perspective, the common point proposed in all the theories is the presence of refolding intermediates which can interact with each other leading to off pathway aggregation reactions and subsequent protein precipitation. Different technologies have emerged to facilitate protein refolding reactions for bulk manufacturing of target proteins. The most conventional method is dilution refolding, although simple to perform, the huge dilution fold required to keep protein concentrations low incurs huge capital investments and associated operating costs to scale-up the process. Other refolding technologies have subsequently been developed to overcome the limitations of dilution refolding which include chromatography and high pressure refolding methods. In parallel to the development of new technologies, significant efforts have also been directed towards understanding the impact of different additives on protein refolding.

Lack of a universal *in situ* monitoring platform however, makes the protein refolding process largely trial and error based. Light scattering tools have shown to be effective in the quick screening of the optimum refolding conditions for proteins, and their potential in monitoring and controlling protein refolding reactions are also discussed in this chapter. Such tools can be instrumental in developing rational protein refolding strategies for novel protein candidate like HBx, lacking commercially available native standards. This will be particularly important since earlier reports related to HBx have attempted to produce truncated or mutated

versions of the protein and also lack quantitative data which is needed for establishing scalable bioprocesses. The success and the failures associated with these earlier attempts to produce HBx are also reviewed in this chapter.

2.1. Role of HBx in HCC

Chronic HBV infection has been well associated with initiation and development of HCC. Within the small3 kb genome of HBV, HBx has attracted particular attention amongst scientists chiefly because of its capacity to affect myriads of metabolic pathways within the infected cells. However, its mechanism of action is yet to be firmly established due to the differential and sometimes contradictory roles played by the protein through its interaction with other protein binding partners[24]. Based on current understanding, it is believed that HBx can activate both viral and cellular transcription through direct binding to transcription factors, co-activators, and components of the basal transcription machinery (summarised in Figure2-1). It also has the capability to cause a mitochondria-mediated increase in cytosolic calcium levels, triggering transcription indirectly through stimulation of signalling pathways like nuclear factor kappa-B (NFκB), Janus kinase (JAK) /signal transducers and activators of a transcription (STAT), and mitogen-activated protein kinase (MAPK). In addition, HBx interacts with proteasome subunits to regulate protein degradation and stability. Importantly, HBx can interact with various cellular proteins like damaged DNA binding protein 1 (DDB1) and p53, thereby providing access to regulation of important cellular events like apoptosis, cell cycle, transformation and viral replication [25-28]. Based on these findings, it seems that the HBx protein not only facilitates HBV replication but can also lead to cellular transformations within the infected hosts.

Until now, the main focus of HBx research seems to be directed towards unearthing its interaction networks and understanding the cluster of genes or proteins regulated by it. Studies like these can help in understanding the probable pathways by which the HBx establishes its control within the host cells and thus allows identification of potential drug targets for the treatment of HCC. To develop suitable drug candidates against HCC, different strategies can be adopted: the major pathway in which the HBx is involved can be manipulated so that the carcinoma is not induced. An alternative and a much more direct strategy would be to design suitable ligands that can specifically bind to HBx within the cellular environment, thereby inactivating it. In either approach, particularly for the latter case, a reliable and cost effective source of the active protein is required, without which research on HBx and HCC may be impeded. Unfortunately, the structural and biochemical characteristics of the HBx protein are largely unknown due to the difficulties associated with producing the protein in soluble form [1, 25]. My Ph.D. research therefore aims to address

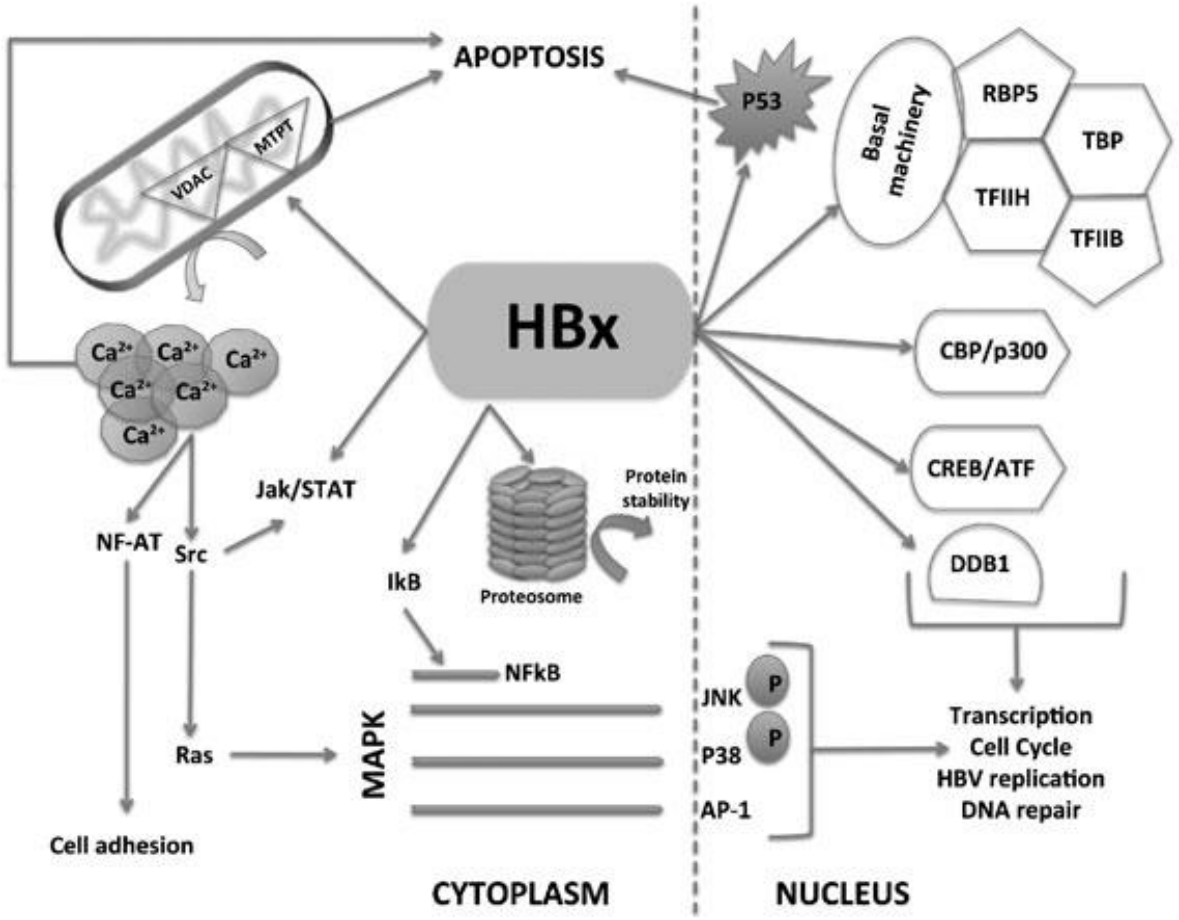

Figure 2-1.Host cellular pathways affected by the HBx protein[1].ATF:activating transcription factor**;** CBP:CREB-binding protein; CREB: cAMP response element binding protein 1; IκB: nuclear factor kappa-B inhibitor; JNK: Jun N-terminal kinase;NFAT: nuclear factors of activated T cells; Pyk2: proline-rich tyrosine kinase; RBP5: retinol-binding protein 5; Src: avian sarcoma viral oncogene; TFIIB: transcription factor IIB; TFIIH: transcription factor IIH

this problem by development of a scalable bioprocess for producing high yields of a recombinant variant of the active HBx protein *in vitro* for further research and development of drugs against HCC.

2.2. In vitro expression of HBx

Just like any other heterologous protein synthesis, the basic strategy for *in vitro* production of the HBx protein would involve incorporation of a HBx-gene carrying vector into a suitable host organism followed by its expression. This so called recombinant protein production can be carried out using various cell culture strategies as well as higher organisms, the latter mainly comprising of transgenic plants and animals. Bioprocessing advantages and disadvantages with respect to a routine HBx protein production using the above-mentioned strategies are briefly discussed in Sections 2.2.1 and 2.2.2.

2.2.1. Protein production in higher organisms

Amongst higher organisms plants have gained wide popularity for large scale recombinant protein production chiefly because of their safety and cost-effectiveness. Protein production in plants would require water, minerals and sunlight. However, protein production in plants suffers from several disadvantages like difficulty in genetic manipulation and frequent contamination with other secondary metabolites due to less well-characterized plant genomes. Moreover, problems associated with low expression yields and protein stability within the harvested crops [29] imply that more investigation are required before such systems can be used as a reliable protein production platform.

Like plants, transgenic animals (cows, goats, pigs, etc.) have been used for production of recombinant proteins in milk, egg white, blood, urine, seminal plasma and silk worm cocoons. Though high titres of the target proteins have been achieved through the use of transgenic animals, the low production throughput (i.e. about 3.5 months for mice to 3 years for cows) disadvantages their use for protein production. Moreover, production in transgenic animals is associated with several ethical or regulatory challenges and has thus practically stalled after the closure of the PPL therapeutics of Scotland which, along with the Roslin Institute, cloned Dolly, the sheep [30].

Thus considering the complications associated with large organisms it would be easier to consider cell culture based protein production platforms for designing a scalable HBx bioprocess.

2.2.2. Cell culture based protein production

In this strategy, the target genes are over-expressed within cellular hosts which can be either eukaryotic or prokaryotic. The cells are first transfected with the desired gene which is then used to produce the protein (depending on the expression system) within the cytoplasm, periplasm or are secreted by the cell. The most popular production platforms using the eukaryotic systems include: yeasts, insect cells and mammalian cells. The chief advantage of using eukaryotic platforms is their capability to produce post-translationally modified proteins. Post-translational modifications are not important as far as the HBx protein production is concerned but being a cysteine rich small protein, expression in eukaryotic hosts might provide significant benefits. However, as discussed later in section 2.6 attempts

to produce the HBx protein in yeast and insect cells led to some protein instability and insolubility problems and hence were not considered in this research. Protein production in mammalian cultures on the other hand takes a significantly longer time for growth and multiplication compared to bacterial expression systems thereby presenting considerable challenges in maintaining sterility. This constraint calls for high operating costs because of the need to maintain a carefully controlled environment for cell growth in an expensive medium [30].

Compared to mammalian expression systems, production of proteins in bacterial hosts is highly economical not only because of cheap growth media, but also because of the higher productivity. Bacterial systems however, are disadvantaged due to protein folding problems chiefly when expressing cysteine-rich proteins because of the reducing environment presented by the bacterial cytoplasm. To overcome these problems, different bacteria strains have been engineered to facilitate heterologous protein production within bacterial hosts, which can be easily adopted for HBx production. *E. coli* (Gram negative) is widely used as a host system for large scale protein production as it is a very well-studied system. Other host systems that are well-studied such as *Staphylococcus aureus* (Gram positive) are less favoured because they express considerable amounts of proteases which impede heterologous protein stability [30]. For cost and productivity reasons, *E. coli* has been used for HBx protein production in this study.

In vitro production of the recombinant HBx protein in the *E.coli* would involve the over-expression of the protein using suitably engineered plasmids. The recombinantly produced protein can be extracted from the bacteria cells, purified and subsequently, if required, refolded to produce the desired product in a pure and biologically active form. Producing the protein of interest in the correctly folded and active state often poses a great challenge, particularly as seen in the case of the HBx protein (discussed in Section 2.6), where its over-expression has been found to form IBs within the bacteria cells. Insoluble protein expression necessitates an *in vitro* refolding step which often gives sub-optimal yields. A biomolecule-centred process which is tailored around specific HBx molecular characteristics is thus considered important to enhance downstream product yield. Therefore, the scope of my research demands a proper understanding of the mechanism of protein folding. A review of

some of the most accepted mechanisms of protein folding and protocols for the protein refolding based on these mechanisms is presented in the following sections.

2.3. Protein folding mechanisms: Forces involved

The idea of folding proteins from an unfolded state emerged in the 1950s when a ribonuclease protein was successfully refolded from denaturing conditions [31]. This famous experiment by Anfinsen indicated that the amino acid sequence of a polypeptide chain must contain all the information required for its proper folding. However, the idea led to another great realization, which is popularly known as the Levinthal's paradox [32]. It is an accepted fact today that Levinthal rightly said that the protein folding process cannot be a completely random trial-and-error process. There must be shortcut pathways along which the polypeptide chain passes through to acquire its native state within the stipulated time. It is believed that forces like electrostatic interactions, hydrogen bonding, hydrophobic interactions and disulfide bridging play a key role in stabilizing the protein molecule in its native state. Based on the understanding of the roles of different forces in protein folding, scientists have hypothesized various models which can be used to understand how a given protein acquire the native stable secondary, tertiary and quaternary structures in a given environment [2, 33].

2.3.1. Protein folding models

A crucial aim for all these models is to understand how a protein can find the 'right' pathway and avoid others. All these models are based on the evidence obtained through the metastable or partially folded states of proteins from various experimental methods [2]. Some of the important models are discussed below.

2.3.1.1. The thermodynamic hypothesis

The thermodynamic approach for understanding the protein folding mechanism probably came to the forefront right after Anfinsen's work. According to this hypothesis, folding of polypeptides is driven by the need to attain a thermodynamically stable state, where the folded conformation, under physiological conditions, is more stable than the unfolded state. This theory does not take into consideration that the native state acquired may be the lowest energy state of the protein molecule, which is accessible within the physiologically appropriate time scale. Anfinsen's hypothesis thus may be accepted as the basis for protein folding processes through self assembly in general. However, the theory says very little about

the actual mechanism of protein folding [34]. This issue was addressed to a certain extent in the Molten Globule Hypothesis.

2.3.1.2. Molten Globule hypothesis

The Molten Globule hypothesis presents a very methodical approach to understand the mechanism of protein folding. The protein chain is thought to first collapse under the influence of hydrophobic interactions between the hydrophobic amino acid residues. The elimination of the ordered water molecules from the hydrophobic sites can be considered to be the driving force for this process. A non-polar core is thus formed within the protein molecule. The stability of the polar peptide moieties within the core is achieved by the formation of suitable H-bonds between the carbonyl and amide groups of the peptide backbone in a particular pattern. These interactions lead to the formation of regular secondary structures or non-regular conformations which are compatible with the polarity of the globule. The resulting conformation of the protein is called the *molten globule state* [35-36].

The molten globule is considered to have a conformation close to the native structure, the difference being that it is inactive and may not be as compact as the native structure. According to Creighton, various combinations of disulfide bonds [37] are formed, broken and reformed within this collapsed state until the correct linkage is formed leading to correct folding. Within the molten globule all the sulphide groups may not be easily accessible. The Molten Globule hypothesis was highly successful in explaining the protein folding process. However, with time and based on new experimental outcomes, scientists have proposed other models to describe the protein folding mechanism as well.

2.3.1.3. Other similar models

Like the Molten Globule theory, the protein folding models that subsequently followed have their roots from the Anfinsen's protein folding concept itself, where the thermodynamic forces provide a major guidance towards attainment of the correctly folded state of the protein. Amongst these models, the Nucleation Growth Model deserves particular attention. It was proposed that the residues adjacent in the sequence can itself form a nucleus for initiating the folding process and ultimately leading to native structure formation. The Framework Model, on the other hand, suggested that at first the basic framework for native confirmation is provided by local elements of secondary structure initiating the formation of native tertiary

structures through diffusion-collision mechanisms. Some scientists on the other hand are sceptical to accept the presence of only one pathway for the protein folding. According to them, there are multiple pathways by which a protein can fold just like how the jigsaw puzzle can be solved in a large variety of ways, which is popularly known as the Jigsaw Model. [34]. All the models discussed above give some ideas regarding the plausible routes a protein molecule can take to arrive at its destined native structure. Figure 2-2 is an attempt to summarize the theories discussed so far [2]. These models laid the foundation for the energy landscape models, which were subsequently developed.

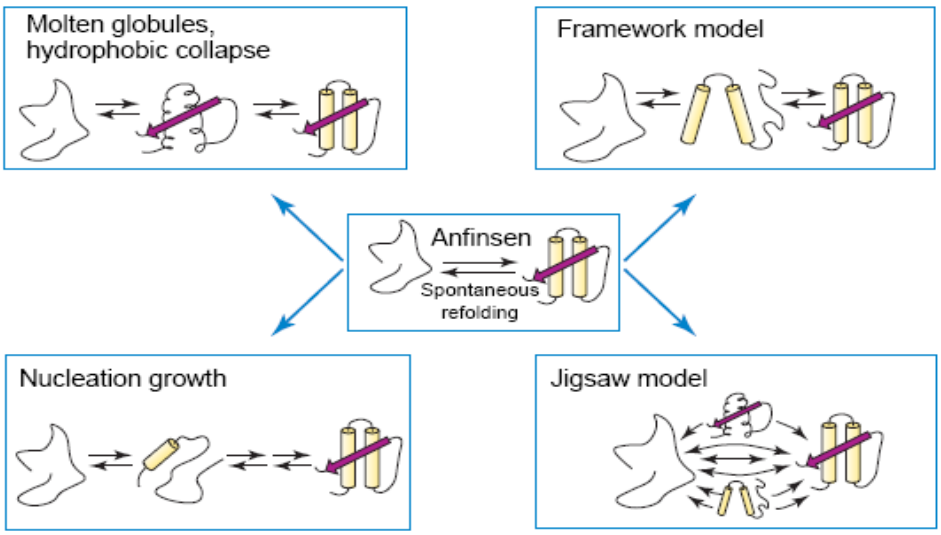

Figure 2-2. Some of the early models of protein folding that originated after the Anfinsen's spontaneous refolding model [2]. Secondary structures like alpha helix and beta sheets are indicated by the cylinders and arrows, respectively.

2.3.1.4. Energy Landscape theory

With the advent of new experimental techniques for protein structure analysis and the development of powerful computers, major changes in our earlier perceptions of folding have developed during the past decade. The Energy Landscape theory is a consequence of these developments.

In contrast to the classical view of protein folding which relies on discreet intermediate refolding species, the Energy Landscape theory describes folding using an ensemble of partially folded structures. According to this theory, the refolding intermediates progressively organise themselves to attain the final native folded state and the energy profile of the these refolding intermediates can be described through an energy landscape diagram as shown in Figure 2-3 [38]. Denatured protein molecules, having the highest energy remain at the top of the landscape, which ultimately guides them towards the final (low energy) conformation by a 'funnel'. On initiation of folding infinite pathways can take the denatured protein to its native conformation, a few of which might involve transient intermediates or local energy

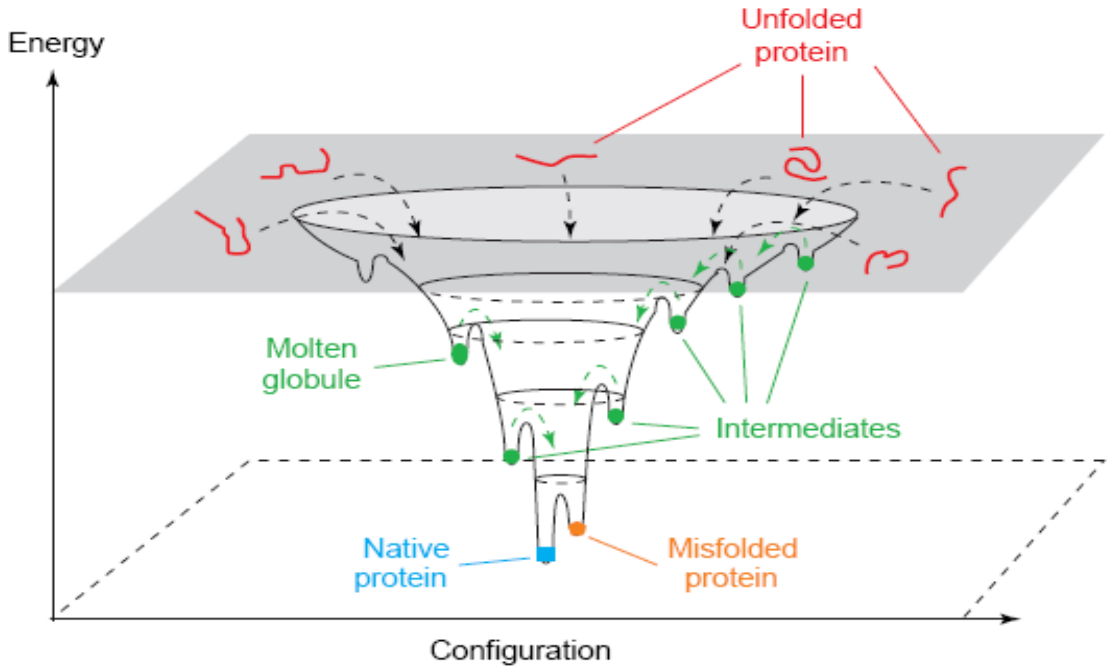

Figure 2-3. Schematic diagram of a folding energy landscape [2].

minima. However, some of the routes can lead to formation of misfolded states leading to significant kinetic traps. It is therefore considered that lesser the number of intermediates, the smoother the surface of the funnel. Thus the shape of the landscape directs all conformations towards the native state (the bottom of the funnel) and avoids the need to sample all possible conformations [2].

The energy landscape theory has several advantages over the previously proposed theories: First, this model proposes that the process of folding will be difficult to change by mutation as long as its stability is maintained based on the funnelling landscape. Mild mutation will thus probably lead to only slight shifts in the folding routes. Second, the funnelling

phenomenon limits the space the protein needs to search for the proper folding path. Thus to fold, the system has to pass through ensembles of structures, called 'transition state ensembles' which consist of high energy intermediate structures. The transition state ensembles derive their stability from interactions present in the native structure. The protein would finally proceed to the bottom of the funnel through this 'transition state ensemble' to attain the stable native state [33].

2.3.1.5. Protein folding inside the cells: role of chaperones

The discussion on the protein folding process presented so far has been restricted to *in vitro* systems. However, understanding of the folding process is incomplete without the knowledge of the challenges confronted by proteins inside the cellular environment, where the folding and biosynthesis of proteins are intimately coupled to each other. The newly emerging polypeptide chains, particularly in the case of large protein molecules, can start to fold before synthesis is complete [39]. Hence, the lack of complete sequence information can lead to incorrect tertiary structure formation leading to the exposure of hydrophobic surfaces resulting in aggregation. To make situations worse, the over-crowded cellular environment highly favours aggregation of partially folded protein molecules [40]. All these factors indicate that proteins are highly prone to aggregation during the protein folding operations within the cell [41].

To address this aggregation problem, living cells use an ingenious machinery called chaperones [42]. These proteins can sequester newly synthesised peptides until it has folded to a state that that is much less prone to aggregation thereby smoothening the folding energy landscape [43-44]. Chaperones are also reported to play important roles in many cellular events like signal transduction, protein targeting and degradation. Different types of chaperones have been reported in the prokaryotic and eukaryotic cells; for instance, in the case of *E. coli*, the GroEL-GroES system forms an important component of the chaperone machinery. Within the eukaryotic cells on the other hand, the hsp70 and hsp60 chaperones have been reported to co-operate *in vivo*, forming a chaperone relay.

The importance of the chaperonin systems is widely accepted today and has influenced many scientists to use different molecular species that have the capability to behave like artificial chaperones for protein refolding operations.

2.4. Protein refolding methods

Refolding is generally referred to as the change in protein conformation from the unfolded to folded state. In this research, the choice of refolding method will be highly dependent on the HBx structure. Hence, understanding the protein folding mechanism will be very important for proper design of *in vitro* refolding systems. Aggregates, popularly known as IBs, are a major hurdle for the biopharmaceutical industry in the efficient manufacture of commercially important proteins. However, increasing demands of different protein-based biopharmaceuticals has urged the industry to focus their attention towards development of sustainable bioprocesses capable of producing active proteins from IBs. Unfortunately, refolding of IBs is not a straightforward process: The following review on protein refolding methods describes the way in which contemporary knowledge of protein folding and aggregation mechanisms have influenced scientists for devising protocols for *in vitro* protein refolding.

Undoubtedly, the first step in refolding proteins from IBs would involve solubilisation of protein aggregates in a suitable denaturant to unfold the polypeptide and render a monomolecular dispersion of polypeptides, prior to refolding. The solubilisation efficiency and structure of the proteins in the denatured state as well as in the subsequent refolding process generally depends on the choice of the solubilising agents. Due to their capability to interfere with solute-solvent interactions, high concentrations of common chaotropes like urea or guanidine hydrochloride (GdnHCl), has been reported to reduce the hydrophillicity of an aqueous medium. In the presence of chaotropes, the environment becomes sufficiently hydrophobic so as to stabilize the hydrophobic amino acid residues within the protein molecules and thus unfolds the protein. Consequently, urea or GdnHCl in high concentrations have become an obvious choice as solubilising agents for IBs. Detergents are also used occasionally as denaturants. However, the main drawback of using only detergents for solubilisation is that unlike urea or GdnHCl, modulating the binding of detergents with proteins for more efficient refolding is rather difficult [45] and detergent carry-over, a common phenomenon, is highly undesirable in secondary processing steps.

Solubilising-denaturing proteins containing disulphide bonds necessitates the use of reducing agents like Dithiothreitol (DTT) or β Mercapto-ethanol to resist the formation of unwanted

disulphide bonds during solubilisation [46]. Even in the unfolded state and in the presence of high concentrations of denaturants, non-native disulphide bonds are often easily formed, leading to the formation of non-native structures. These unstable structures have a strong tendency for aggregation or misfolding upon removal of denaturants during the subsequent refolding steps. The solubilised molecules can then be subjected to refolding by reducing the concentration of denaturants. The success of refolding lies in the success to minimise other competing side-reactions which lead to misfolding and aggregation which occur in parallel to the refolding process.

The rates of refolding and other side-reactions generally depend on the procedure used to reduce the denaturant concentration and the solvent conditions surrounding the protein as the protein is subjected to different physicochemical environment. A review of the most widely used refolding protocols is discussed below.

2.4.1. Dilution refolding

In dilution refolding, native structure formation of the protein is induced by diluting the denatured protein solution into a refolding buffer. Dispersion of the protein in large volumes will reduce protein concentration and hence the second or higher ordered aggregation pathways. A refolding reaction, which is of first order reaction is described using Equation 2-1.

$$\frac{dU}{dt} = -\left(k_1 U + k_2 N U^n\right) \qquad \text{Equation 2-1}$$

where k_1 is the net rate constant of refolding, k_2 is the net rate constant of aggregation, U is the concentration of unfolded protein, t is refolding time, N is aggregation number and n is the reaction order of aggregation. Thus at higher protein concentrations, the protein aggregation reaction would dominate over the refolding reactions, while at lower protein concentrations, the refolding reaction would dominate as shown in Figure 2-4 [47-48].

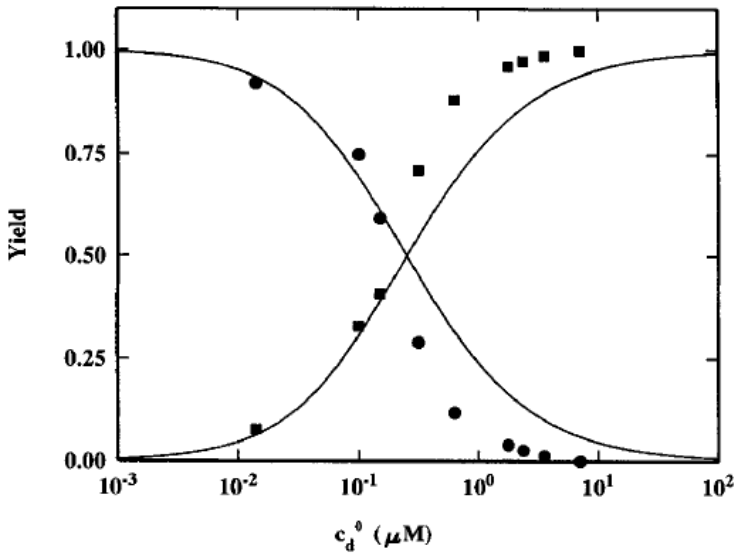

Figure 2-4. Effect of concentration of denatured protein (c_d) on the extent of *in vitro* refolding (◕) and aggregation (■) [48]

To minimize unwanted irreversible intramolecular interactions, low denaturant concentrations are often introduced in the refolding buffer to minimise the propensity of aggregation during refolding. For most proteins, rapid dilution is favoured over a slow dilution process to keep protein concentration low at the onset of refolding [49-52].

Pulsed dilution or fed-batch dilution refolding is also popular, where the addition of denatured protein is interspersed with the aim to maintain low concentrations of the denatured protein species at any given point of time. For example, denatured protein is added to the refolding mixture only when 80% of the refolding yield is reached. Factors like the increase of denaturant concentration with each cycle and the amount of protein added per pulse need to be carefully considered with the use of such protocols [53-54].

Though it is simple to scale-up a dilution refolding process using a stirred-tank reactor, the main disadvantage of this strategy lies in its exorbitant scale-up costs. Process scale up would require the utilization of huge volumes of buffers and stirring tanks that can increase the capital and operating costs of the overall process. The need for subsequent treatment of such huge volumes of buffers can further multiply the production costs to a significant extent [55].

2.4.2. Dialysis Methods

Instead of using an approach that drastically reduces the denaturant concentration which may guide protein molecules into a wrong folding pathway, many researchers like to use a much more gradual approach for the removal of denaturants such as dialysis, to obtain the native proteins at good yields [54, 56]. However, the dialysis refolding process is time consuming. Additionally, the protein being subjected to a relatively slow change in denaturant concentration can also provide ample opportunities for protein refolding intermediates to interact with each other. Refolding intermediates which still has a loosely packed hydrophobic core also often get adsorbed onto the dialysis membrane, resulting in protein loss. In a comparative study of the dilution and dialysis methods for the refolding of α-fetoprotein, a highly disulphide-bonded biopharmaceutical candidate, the former process was found to give higher refolding yields than the latter [57]. However, each protein having a unique sequence may exhibit significant extents of amphiphilicity and may respond differently to different refolding methods [55, 58-60].

2.4.3. Chromatographic techniques

In recent years, column chromatography has gained an increased interest among researchers for protein refolding. The ease of automation and scaling up using chromatography techniques has increased the number of refolding studies based on chromatography tools in the past decade. The overall objective of on-column refolding is to adsorb solubilised denatured proteins on a matrix to facilitate spatial isolation of the molecules. The restricted movement of the protein molecules within the immobilized matrix allows them to remain dispersed, avoiding inter-molecular interactions, thus preventing aggregate formation. In this refolding method, the chromatography column is first equilibrated with denaturant to allow protein adsorption in the denatured form, followed by a solvent exchange which favours refolding of the protein molecules on-column. After refolding, the refolding buffer is then displaced with an appropriate elution buffer to elute the refolded proteins. Different chromatographic matrices have been successfully used for on-column refolding including size-exclusion chromatography, ion-exchange chromatography, affinity chromatography and hydrophobic interaction chromatography [61-66]. Due to the spatial isolation of the immobilised protein molecules within the packed beads, chromatography refolding has the advantage of facilitating refolding at much higher protein concentrations compared with other dilution refolding techniques without incurring extensive aggregation. Simultaneous

purification and refolding can also be readily achieved on-column upon adequate process optimisation [67].

Successful refolding using this strategy demands optimisation of several operational parameters like proper selection of the chromatographic matrix, column dimensions and solvent flow rates. Particularly, the rate of change of the denaturing environment within the column into a refolding one can significantly influence the topography of the energy landscape of the refolding protein [60].

2.4.4. High hydrostatic pressure (HHP) technology

Another way to promote native refolding reactions is to create a refolding environment that can destabilise off-pathway aggregation. HHP technology was developed with this strategy in mind, where hydrostatic pressure is applied to 'ensure' that only proteins refolded to the native form and hence having the lowest specific volumes, are stabilised [68-69]. Application of pressure can also help to solubilise the IBs directly and refold the target protein spontaneously. However, modulation of hydrogen bonds and disulphide bridges needed for solubilising the IBs and obtaining the correctly refolded product can only be achieved by an optimised buffer system [70].

Based on the existing refolding techniques (summarised in Table 2-1) it is clear that the design of existing refolding methods is focused mainly on controlling refolding concentrations to obtain high refolding yields, but optimisation of the physicochemical environment still needs to be independently performed for each of these methods. In recent years, high throughput-screening of refolding buffers for protein refolding based on protein solubility has been reported [71]. Some of these screening kits have been commercialized; for example, Refold Master^{TM} developed by Novexin Ltd (Cambridge, UK) comprises a matrix of different buffer compositions which allows screening of protein foldability in 96 well-plate formats. These screening kits however, do not allow *in situ* monitoring of the refolding reaction, which can provide important insights into the influence of different refolding additives on refolding yields. The major hurdle for such monitoring is that different proteins require different detection strategies and a generic screening platform is not practically possible. It would, therefore, be considerably more advantageous if new screening methods could be developed to simultaneously provide quantitative information on aggregation

Table 2-1. A brief overview of existing refolding technologies.

Refolding Strategy	Rationale	Advantages	Disadvantages
Dialysis	1) Minimise denaturant carry over from the solubilised IB solutions. 2) Equilibrate the proteins within a buffer capable of refolding the protein.	Helpful when the denatured-reduced protein concentration is low.	1) Protein loss to the dialysis membranes. 2) Slow buffer exchange kinetics can induce aggregation of the refolding intermediates.
Dilution	Reduced interaction amongst the refolding intermediates by lowering protein concentration.	Better refolding yields compared to dialysis in most cases.	1) Success largely depends on the choice of the refolding buffer composition. 2) Large consumptions of expensive buffers can negatively impact the overall process economics.
Chromatography	1) Spatial isolation of refolding intermediates bound on chromatographic matrices, eliminates inter-molecular interactions. 2) Refolded proteins can be eluted from the column at relatively high concentrations.	Useful for process intensification.	1) May inhibit protein refolding due to loss of flexibility of the bound polypetide chain. 2) Intra-molecular protein interactions can lead to unproductive refolding of the bound proteins.
High Hydrostatic Pressure Technology	1) High pressure helps to solubilise the IBs. 2) The pressure allows preferential stabilisation of the natively folded protein molecules.	One-step conversion of IBs into refolded products.	Hydrogen and disulphide bonds within the IBs cannot be broken by applying pressure. Successful solubilisation and refolding require an optimised buffer system.

kinetics under different refolding environments, thus providing a better correlation between physicochemical parameters and refolding yield for a given protein "[72]". The following two sections present a brief review on how second virial coefficient (SVC) measurements can be used not only as a valuable tool for measuring such data but also to develop rational process designs and recipes for protein refolding.

2.5. SVC measurement: a rational approach to designing protein refolding recipes

Based on our discussion in Section 3.2, the random movement of protein molecules during refolding is influenced by the inter-particle and hydrodynamic interactions amongst them. These interactions can be quantitatively assessed using a thermodynamic parameter known as SVC. Importance of SVC measurement in understanding protein behaviour came to the forefront during the early 1990s, when George and Wilson found that successful crystallization takes place only when the SVC of the solutions lie within a specific value [73]. The ability to quantitatively predict environments conducive for protein crystallization indicates that a convenient SVC measurement platform can ~~also~~ be instrumental for studying protein-protein interaction within a refolding system by providing thermodynamic information on protein phase behaviour. Before discussing the applications of SVC in the rational design of refolding recipes, concepts of SVC is briefly reviewed to set the context of the discussion "[72]".

2.5.1. Second virial coefficient: theory and applications

SVC is a thermodynamic parameter which originates from the virial expansion of solution osmotic pressure, given in Equation 2-2,

$$\frac{\pi}{RT} = \rho + B_2\rho^2 + C\rho^3 + . \qquad \text{Equation 2-2}$$

where π is the solution osmotic pressure, R is the gas constant, T is the temperature, ρ is the solute density and B_2, C, etc. are the osmotic virial coefficients. B_2 in the above equation is known as the second virial coefficient and is a measure of two-body interactions within a solution. Throughout the rest of the thesis, the term *"B_2"* and "SVC" will be used synonymously. According to statistical mechanics, SVC can be expressed as an integration of the potential of mean force, $w_2(r)$, between two molecules:

$$B_2 = -\frac{2\pi N_A}{M_w^2}\int_0^\infty \left(e^{-w_2(r)/k_B T} - 1\right) r^2 dr \qquad \text{Equation 2-3}$$

In Equation 2-3, N_A stands for the Avogadro's constant, M_w is the molecular weight, k is the Boltzmann constant and r is the centre-to-centre distance between two molecules [74]. Therefore, the SVC takes into account the same interactions (like electrostatic, Van der Waals, excluded volume, hydration forces, and hydrophobic interactions) which are known to regulate protein phase behaviour [75]. Hence this parameter should correlate well with protein solubility which is also evident from Equation 2-4, developed by Guo et.al [76]

$$B_2 = \frac{-\Delta\mu_p}{RT}\frac{1}{2M_wS} - \frac{\ln S}{2M_wS} \qquad \text{Equation 2-4}$$

where S is the protein solubility and $\Delta\mu_p$ is defined as the difference of the standard chemical potentials of protein between two states (solution and crystal). Equation 2-4 was further refined by Ruppert et al. [77] to obtain a direct relationship between the solubility and experimentally determined SVC.

Based on these findings, over the past decade, experimentally determined SVC values have been used by scientists to shed light on methods to manipulate protein thermodynamic properties and phase behaviour [78-79] and even obtain partial structural information [80]. SVC values have thus been used as an unbiased estimator for protein-specific behavioral patterns to guide determination of optimal crystallization conditions for proteins [73, 81-82]. In view of protein refolding applications, SVC values can accurately correlate the relationship between refolding additives and protein interactions [83-86], and has been exploited to predict (i) the propensity for aggregation during protein refolding [51] and (ii) potential refolding yields of IB proteins [87]. Therefore, SVC measurements of refolding mixtures carried out at statistically determined design points can rapidly provide quantitative correlations between the choice and concentration of additives on the interactive behaviour of proteins [75, 88]. Hence, in the case of proteins which do not have standard bioactive product(s) nor any analytical platforms for the quantitative bioactivity assay such as HBx, SVC can be a good analytical tool for determining protein stability and/or solubility.

Amongst diverse methods for SVC measurement, both static light scattering (SLS) and static interaction chromatography (SIC) have obtained wide popularity within the scientific community. SLS experiments are based on the measurement of the light intensities scattered by the randomly moving protein molecules within a solution (discussed in Chapter 4). The SIC methodology on the other hand, involves measurement of the retention time between

protein molecules flowing over the same molecules, covalently immobilised over a surface (chromatographic bead or a chip); the retention time in turn is linked to B_2 through the potential of mean force [81]. SLS measurements being a non-invasive technique, is very simple to conduct particularly with the modern commercially available equipment but is often disadvantaged by the lack of throughput. SIC in this regard can provide higher throughput but the need to immobilise the protein on a surface might be problematic on many occasions. In this research project, SLS was used to monitor the SVC values of the HBx protein solutions (discussed in Chapter 4) "[72]".

2.6. Earlier attempts to produce recombinant HBx in microbial systems

A definitive proof for the existence of the HBx protein was established only in the late 1980s from two independent studies [89-90]. Since then, several attempts have been directed towards soluble production of the protein for its structural and biochemical characterisation. However, the protein is consistently expressed in the insoluble form under all expression conditions for all types of protein production platforms studied including insect cells [91-93]. Even attempts to produce the full length protein within yeast cells lead to some stability problems [94]. Since *E. coli* is the most popular protein expression platform due to cost and speed factors, various strategies have been adopted by scientists to improve the soluble expression of HBx in bacteria, which include but are not limited to: (i) co-expression of chaperones to aid in protein folding [95-97]; (ii) producing truncated version of the whole length protein; (iii) directing protein expression in the periplasmic space, to provide an oxidising environment for cysteine oxidation to form disulphide bonds; or (iv) using solubility enhancing fusion tags [98-102]. Table 2-2 summarises some recent attempts to express HBx in the soluble form, where the most successful methodology involving the use of maltose binding protein (MBP) fusion protein strategy by Liu D. *et al.*, could only provide sub-optimal yields of the HBx protein (10 mg HBx from 25 L of culture).

The exact cause for insoluble HBx production is yet to be fully established, but amino acid sequence analysis of the protein using the ProtParam program, available from www.ExPasy.org, reveals some interesting results. With approximately 40% of hydrophobic amino acid residues, the HBx sequence possesses a positive grand average of hydropathicity (GRAVY) index, indicating its high hydrophobicity. Moreover, the HBx protein sequence often possesses an odd number of cysteine residues, which further adds to its instability

during recombinant production, as indicated by the instability index of 60.39, classifying the protein as unstable. Based on these results, it is not surprising that until now, protein scientists have had little success in expressing the HBx protein in the soluble form at reasonable yields. The absence of quantitative yields reported in earlier refolding studies of HBx also further impedes the development of a sustainable bioprocess for large-scale production of HBx. For example, Marczinovits and co-workers reported successful refolding of a GST-fused truncated HBx protein using dilution refolding method [17]. However, a complete analysis of the final yields of the bioactive protein was not reported in their study. Other groups reported the use of sequential dialysis in PIPES buffer systems to successfully refold HBx and its mutated forms [103-104] but again, the lack of quantitative yields impedes implementation of bioprocess design for scale-up. It is clear that the lack of quantitative HBx refolding analyses coupled with ill-defined bioprocesses will continue to hinder the supply of bioactive HBx in sufficient amounts for advanced structural and drug designing studies.

Based on the discussion presented above on the earlier works, it appears that successful bioprocesses can be established to improve HBx solubility during expression by using solubility-enhancing fusion tags. Considering that the widely known MBP fusion tag only led to suboptimal HBx yields, other tags like NusA, Trx or GST can be studied to obtain better results. Tags like NusA or Trx, however, are disadvantaged by the lack of any biospecific ligand affinity, leading to purification complexities [105]. In comparison, although the GST tag may or may not solve the solubility problem [106-107], the bioactive target protein can be readily purified using a glutathione affinity chromatography step. Furthermore, since a truncated version of the HBx protein has been successfully produced with a GST-tag [17], we deemed the use of a GST-tag for full length expression of HBx worth studying.

Table 2-2. A brief summary of the attempts to express HBx in the soluble form [N.A. implies not applicable / not mentioned in the study].

Parameter	Jameel S.[108]	De-Medina T.[109]	Marczinovits I.[17]	Rui E.[104]	Hwang G.Y.[18]	Zhang S.M.[19]	Liu D.[110]
E. coli strain	BL21 (DE3)	BL21 (DE3) pLysS	DH5α	BL21	DH5α	BL21 (DE3)	JM109
Construct	pET-8C vector	1. pET-8c vector 2.pGEX1 vector	Truncated variant of HBx cloned into pGEX-3X vector	psW202 vector	pGEX-5X-1 vector	pGEX-4T-1 vector	pMAL-c2x
Fusion tag	No	1. No 2. GST	GST	1. 6-Histidine 2. GST	GST	GST	MBP
Expression medium	N.A	2xYT	Luria broth	LB	LB	LB	LB
Expression temperature	37°C	37°C	37°C	37°C	30°C	28°C	37°C
Point of induction	N.A.	OD 0.5	A_{600} 0.6-0.7	OD 0.7	N.A.	A_{550} 0.6	OD_{600} 0.6-0.7
Induction method	0.4 mM IPTG	1 mM IPTG	0.5 mM IPTG	1 mM IPTG	1 mM IPTG	0.1 mM IPTG	0.03 mM IPTG
Post induction time	3 h	2 h	2.5 h	0.5 - 4 h	5 h	3 h	2.5 h

Parameter	Jameel S.[108]	De-Medina T.[109]	Marczinovits I.[17]	Rui E.[104]	Hwang G.Y.[18]	Zhang S.M.[19]	Liu D.[110]
Lysis buffer	50 mM Tris (pH8.5), 10 mM EDTA.	50 mM HEPES (pH8.0), 100 mM NaCl, 0.5 mM EDTA, 5 μg/ml PMSF, 10 μg/ml aprotinin, 20 μg/ml benzamindine, and 2 μg/ml pepstatin.	10 mM Tris (pH 8.0), 150 mM NaCl, 1 mM EDTA, incubation with 100 g/ml lysozyme on ice for 15 min, 1% N-lauryl sarcosine.	Addition of 1 mg/ml lysozyme.	10 mM Tris (pH8.0), 150 mM NaCl, 1 mM EDTA), addition of 100 μg/ml lysozyme, incubated on ice for 15 min, 5 mM DTT, sarkosyl.	PBS, pH 7.3, 0.1% Triton X-100.	20 ml of 1.0 M Tris-HCl, pH 7.4, 11.7 g of NaCl, 2 ml of 0.5 M EDTA, 154 mg of DTT.
Homogenization method	One freeze thaw cycle, sonication.	Sonication	Sonication	Incubation at room temperature for 20 and 3 freeze thaw cycles.	Sonication	Sonication	Sonication
Buffer to dissolve insoluble fraction	50 mM Tris (pH 8.5), 10 mM EDTA, 6 M urea and 5 mM 2-mercaptoethanol.	7 M urea	8M urea (pH 6.5)	6 M guanidinium-HCl, 100 mM NaH_2PO_4, 10 mM Tris-HCl (pH 8.0).	N.A.	N.A.	N.A.

Parameter	Jameel S.[108]	De-Medina T.[109]	Marczinovits I.[17]	Rui E.[104]	Hwang G.Y.[18]	Zhang S.M.[19]	Liu D.[110]
Soluble HBx	No	1. No 2. Partly soluble.	Partly soluble.	1. No. 2. Partly soluble.	N.A.	Yes	Yes. >90% of MBP-HBx expressed.
Approach to increase soluble expression	No	No	1. Lower incubation temperature for induction. 2. Mild disruption of the cells to avoid denaturation and aggregation of protein. 3. Use of detergent and thiol agent for solubilisation. 4. Reduce the number of hydrophobic amino acids by truncation of the HBx protein.	1. Lower expression temperature. 2. Co-expression of chaperones. 3. Directing HBx protein into the periplasm (an oxidising environment which might help in the correct formation of disulphide bonds). 4. Shorter mutants of GST-HBx.	No	No	1. Lower incubation temperature for induction. 2. Mild disruption of the cells to avoid denaturation and aggregation of proteins. 3. Use of detergent and thiol-group-containing reagent for solubilisation.

Parameter	Jameel S.[108]	De-Medina T.[109]	Marczinovits I.[17]	Rui E.[104]	Hwang G.Y.[18]	Zhang S.M.[19]	Liu D.[110]
Successful in increasing soluble expression	N.A.	N.A.	No	Truncated proteins with fusion tags were partly soluble, but the full length HBx protein was exclusively found in the insoluble fraction.	N.A.	N.A.	No
Hypothesis on why the method failed to increase soluble HBx expression	N.A.	N.A.	Due to structural characteristics of HBx (52% of the amino acids is hydrophobic and 4 disulphide bonds).	1. Chaperones used were not involved in the folding pathway of the HBx protein. 2. HBx might also form IBs in periplasmatic space.	N.A.	N.A.	N.A.

In this research I aim to study both molecular and process-related factors which could affect refoldability and solubility of HBx, which have not been quantitatively studied in the earlier works. The sequence of the HBx model protein in this thesis contains 9 cysteine residues and has close similarities with the protein from the HBV-A3 variant, recently reported to be prevalent in places like Cameroon, Gambia and Gabon in Africa (Isolate Cameroon/CMR983/1994) [111]. To the best of my knowledge, it is the first time refolding studies of this HBx variant are studied.

CHAPTER 3: The Effect of a GST Fusion Partner on HBx Expression and Bioprocessing in an *E. coli* Microbial Platform

Abstract

Earlier efforts to produce the HBx protein by optimising culture conditions using various protein expression systems have consistently resulted in the expression of HBx in the insoluble form. The hydrophobicity of HBx and the presence of odd cysteine residues are likely causes for the protein's propensity to form aggregates in solution, which pose a roadblock to the development of a large scale HBx production platform for structure-characterization studies. This chapter investigates the impact of a solubility enhancing fusion partner i.e. GST, on the soluble expression and bioprocessing strategies for the HBx protein. Extensive fermentation optimisation strategies failed to improve the expression of the GST-HBx in the soluble form in *E. coli*. Two downstream processing strategies for the GST-HBx protein were subsequently designed and employed with the aim to recover purified HBx in the soluble form. Firstly, GST-HBx was purified in the denatured-reduced state, followed by protein solubilisation and Factor Xa cleavage under non-denaturing conditions to recover HBx. Secondly sarkosyl was introduced into the cell lysate to facilitate selective solubilisation of GST-HBx, followed by purification and Factor Xa cleavage to recover HBx. The results from these studies revealed that designing an economically viable bioprocess for HBx production would be extremely challenging using the GST-HBx construct. In particular, solubility of the fusion protein or the HBx protein alone was difficult to achieve under various physicochemical environment due to the pI range of the fusion protein components. Additionally, the low efficiency of the Factor Xa cleavage reaction makes subsequent recovery of bioactive HBx difficult and costly. These results point to the need for a new HBx construct design to facilitate more efficient and economical bioprocessing of this protein, which was subsequently investigated in Chapter 4.

3.1. Introduction

The insolubility of recombinantly produced HBx in *E. coli* is a major impeding factor for large scale biomanufacturing of this protein. The results from earlier HBx production studies presented in Chapter 2 indicate that the way forward for developing a scalable HBx bioprocess might be through the use of solubility enhancing fusion tags or optimised protein refolding strategies for the HBx protein. Both these strategies were studied in this thesis wherein studies were first directed towards enhancing soluble protein production using a GST-HBx construct, which forms the basis of the studies reported in this chapter.

Co-expression of HBx with fusion partners like GST or MBP has been reported in earlier HBx production studies [17, 104, 110]. Although the co-expression of HBx with MBP resulted in the soluble expression of HBx, the low cleavage efficiency coupled with high step yield losses resulted in a sub-optimal HBx yields. Despite its well-known solubility-enhancing property, the use of MBP is also disadvantaged by the increased metabolic burden to produce such a large fusion partner (42 kDa) [112], thus resulting in lower peptide yields due to the high fusion protein to peptide mass ratio. Due to these problems, the GST protein was thought to be a more suitable fusion partner for HBx production, since it has been reported to improve recombinantly expressed protein solubility and facilitate its purification as well [113]. Previously, researchers have tried to produce the GST-HBx protein using mutated or truncated HBx variants but insoluble expression was consistently observed [17, 104].

Studies reported in this chapter were first aimed at improving the soluble expression of GST-HBx in *E. coli* by varying culture conditions and media compositions. Two approaches were subsequently studied to purify and solubilise the IB-derived HBx protein: Strategy I) pre-refolding purification of GST-HBx under denatured-reduced conditions followed by dilution refolding and Strategy II) protein solubilisation using suitable detergents under non-denaturing / non-reducing conditions, followed by purification. The purified fusion protein, stabilised under non-denaturing conditions, was then subjected to Factor Xa cleavage for removal of the GST-tag in both strategies.

3.2. Methods and materials

3.2.1. HBx plasmid transformation

A pGEX-5X-1 plasmid containing the gene sequence of HBx-6His in frame with that of GST was kindly provided by Prof. William Chen Wei Ning (School of Chemical and Biomedical Engineering, NTU, Singapore). The plasmid was dissolved in Tris-EDTA (TE) buffer and transformed into *E. coli* BL21DE3 (Stratagene) by the heat shock method. The bacteria cells were made competent by $CaCl_2$-$MgCl_2$ treatment using standard protocols [114]. The transformed cells were allowed to grow overnight on an ampicillin selective LB agar medium. A single colony from the plate was picked and used to inoculate a liquid LB medium (5 ml) containing 0.1mg/ml of ampicillin and the culture was grown overnight. The cultures were then harvested for subsequent DNA sequencing and protein expression studies.

3.2.2. DNA sequencing

Plasmids from 5 ml of the transformed *E. coli* overnight culture were extracted using a plasmid extraction kit (Qiagen) for DNA sequencing of the HBx insert to determine the precise amino acid sequence of the protein. Sequencing was performed by the National Neuroscience Institute (NNI), Singapore, using pGEX-5' and pGEX-3' primers (GE Healthcare). The resulting DNA sequence data was translated into the corresponding amino acid sequence using the online software 'Translate' [http//www.expasy.org/tools/#translate].

3.2.3. GST-HBx protein expression

Overnight culture containing the transformed *E. coli* (BL21 DE3) cells was used to inoculate a fresh sterile 400 ml LB medium contained in a 2 L culture flask, to obtain a 1% (v/v) bacterial suspension. The bacteria cells were then allowed to grow in a shaking incubator at 37°C and 200 rpm shaking condition until the late log phase was attained (i.e. OD_{600} = 0.7). Isopropyl-1-thio-β-D-galactoside (IPTG) (1st Base) was then added to a final concentration of 0.25 mM to induce recombinant protein expression. The cells were harvested at OD_{600} = 2.0. The 2 L culture was divided into 500 ml centrifuge tubes and centrifuged at 5200 g for 30 min. The cell pellets were stored in a -20°C freezer (for a maximum duration of 1 week) and thawed before cell lysis.

For protein expression using the M9 medium, the cell cultures (grown at 37°C in LB medium) were centrifuged after reaching the late log phase (OD_{600} = 0.7), as specified earlier.

The cell pellets were then re-suspended in fresh M9 minimal medium containing 1% (w/v) casamino acids and 1% (w/v) glucose. The M9 minimal medium consisted of 12.8 g Na_2HPO_4, 3 g KH_2PO_4, 0.5 g NaCl, 1.0 g NH_4Cl, 2 ml of 1 M $MgSO_4$, 0.1 ml of 1 M $CaCl_2$ in 1 L sterile deionised water. For low temperature protein expression (i.e. at 30°C and 18°C), the cultures were grown at 30°C or 18°C for 30 min, and induced with 0.1 mM IPTG after reaching the late log phase (i.e. OD_{600} = 0.7).

3.2.4. Cell disruption to recover GST-HBx

The cell pellets were washed with phosphate buffered saline (PBS, pH 7.2) and then re-suspended in 200 ml of 50 mM Tris-HCl buffer (pH 8.2). The suspended cells were disrupted by subjecting the cells to (i) French Press (Thermo Electron Corporation) at16000 psi pressure in 3 consecutive passes, or (ii) sonication (Branson sonifier 450, 400 watts) for 3 min (pulses: 5s on and 7s off) at 30% amplitude (maximum frequency = 20000 Hz). The soluble and insoluble protein fractions were separated by centrifugation, before further analysis of the proteins.

3.2.5. GST-HBx purification, refolding and cleavage

Despite extensive optimisation of culture conditions, the GST-HBx was inherently expressed as IBs in *E. coli.* A refolding based downstream processing strategy was thus developed to recover GST-HBx in a soluble form. Since the Factor Xa protease activity is dependent on the native structural conformation of the fusion protein, studies to solubilise GST-HBx under non-denaturing conditions were performed. Two strategies were employed (i.e. Strategies I and II) as described in Sections 3.2.5.1 and 3.2.5.2, to solubilise GST-HBx, followed by FactorXa cleavage.

N.B. In the following sections of this chapter, solubilisation of GST-HBx under non-denaturing conditions is referred to as a refolding step.

3.2.5.1. GST-HBx processing using Strategy I (pre-refolding purification of GST-HBx under denatured-reduced conditions followed by dilution refolding)

The GST-HBx protein expressed in LB medium (at 37°C, Section 3.2.3) was used for further downstream processing studies. The IBs recovered from the disrupted cell suspensions were centrifuged at 6500 g for 30 min. Pre-refolding purification of the GST-HBx was performed

using IMAC with a 1 ml Ni^{2+} Sepharose HisTrap FF column (GE Healthcare). All chromatography experiments were conducted using an AKTA Explorer Fast Protein Liquid Chromatography (FPLC) System (GE Healthcare). HBx IBs were solubilised in the IMAC binding buffer (Buffer A) comprising 8 M urea, 0.5 M NaCl, 20 mM Imidazole, 10 mM DTT, in 20 mM sodium phosphate buffer (pH 7.4), and filtered through a 0.45 μm filter. After equilibrating the IMAC column with 10 column volumes (CVs) of Buffer A at 1.0 ml/min, approximately 40 mg of the protein sample was loaded into the column. The bound protein was then further washed with 10 CVs of Buffer A at 1.0 ml/min to allow complete removal of all unbound proteins. GST-HBx was then eluted with 60% Buffer A and 40% elution buffer (Buffer B) (8 M urea, 0.5 M NaCl, 300 mM Imidazole, 10 mM DTT in 20 mM sodium phosphate buffer, pH 7.4) in a single elution step over 10 CVs at 1.0 ml/min. 1.5 ml fractions of the eluate collected using the Frac-950 (AKTA, GE Healthcare) fraction collector were subjected to refolding studies. Urea used in all the work reported in this thesis was pre-treated with Amberlite (GE Healthcare) to capture cyanates which can carbamylate proteins [115]. Amberlite pre-treatment was achieved by stirring 9 M urea solution containing 1% (w/v) Amberlite at room temperature for 1 h. The urea solution was subsequently diluted to the desired concentration.

Purified GST-HBx was buffer exchanged into 8 M urea and 50 mM Tris (pH 7.5±0.5) using a PD 10 desalting column (GE Healthcare) to remove Imidazole and DTT which can interfere with HBx refolding. The buffer-exchanged protein solution was rapidly diluted into the refolding buffer at different folds to obtain refolding protein concentrations ranging from 25 to 100 μg/ml. Four refolding buffer compositions were examined for protein solubilisation optimization studies: (i) Buffer 1: 2 M urea, 1 mM reduced glutathione (GSH) and 0.1 mM oxidised glutathione (GSSG), 50 mM Tris, pH 8.0±0.5; (ii) Buffer 2: 2 M urea, 1 mM GSH and 0.1 mM GSSG, 0.3 M L-arginine, 50 mM Tris, pH 8.0±0.5; (iii) Buffer 3: 2 M urea, 0.3 M L-arginine, 50 mM Tris, pH 8-9; (iv) Buffer 4: 2 M urea, 0.3 M L-arginine in 50 mM acetate-citrate buffer, pH 6. In all the refolding studies, the proteins were incubated at 4°C for 1 to 3 h.

After incubation, the refolded GST-HBx samples were subjected to enzymatic cleavage by Factor Xa (Part No. P8010L, New England Biolabs). The cleavage reaction was conducted at room temperature with an optimised enzyme to substrate ratio of 1:2.5 (by mass) for 48 h.

The cleavage buffer was prepared by adding NaCl and $CaCl_2$ to the refolding buffer to a final concentration of 150 mM and 1 mM, respectively. The cleavage reaction was arrested by adding the supernatant of the cleavage sample into 2 x Laemmli buffer and storing at -20°C. HBx recovery from the cleavage reaction was determined by sodium dodecyl sulphate poly-acrylamide gel electrophoresis (SDS PAGE) analysis. The gels were stained either by the Coomassie stain or the Silver staining kit (Catalogue No. 161-0449, Biorad).

3.2.5.2. GST-HBx bioprocessing using Strategy II (protein solubilisation using sarkosyl under non-denaturing / non-reducing conditions, and purification using Glutahione-Sepharose resins)

In Strategy II, the GST-HBx IBs were solubilised with three types of detergents (i.e. sarkosyl, SDS and Triton X-100) during the cell lysis step and the proteins were then bound to Glutahione-Sepharose 4B (GS) resins (GE Healthcare), followed by the Factor Xa cleavage reaction.

The cell pellets (1 g of cells obtained from LB culture as described in Section 3.2.3) were re-suspended in 40 ml of STE buffer (150 mM NaCl, 1 mM EDTA, 10 mM Tris-HCl, pH 8.0) and divided into 15 ml aliquots in 50 ml Falcon tubes. DTT and sarkosyl were then added to the mixture to a final concentration of 5 mM and 0.2% (w/v) respectively, before the sample was subjected to sonication for 5 min at 25% amplitude (pulses: 5s on, 7s off). The cell lysates were transferred into 1.5 ml eppendorf tubes and centrifuged at 16000 g at 4°C for 25 min. The supernatant which contained the clarified cell lysate was recovered and the insoluble proteins were re-solubilised in 8 M urea for subsequent SDS PAGE analysis.

Excess sarkosyl in the 1 ml clarified cell lysate, was sequestered by adding Triton X-100 to a final concentration of 2%. The resultant mixture was added to 100 – 300 µl of 50% (v/v) slurry of GS beads suspended in PBS buffer. The mixture was then incubated with gentle agitation at room temperature for 2 h and the GS beads were then separated by centrifugation at 500 g for 5 min. After washing the bound proteins with Factor Xa cleavage buffer (50 mM Tris-HCl, 150 mM NaCl, 1 mM $CaCl_2$, pH 7.4), the mixture was treated by Factor Xa at an enzyme : substrate ratio of 1:5 and incubated up to 16 h. After cleavage, the GS beads were removed from the protein mixture by centrifugation and the bound proteins were eluted with 3 ~~bed~~ column volumes of elution buffer (50 mM Tris-HCl, 20 mM GSH, pH 8.0) [113].

3.2.6. Analytical methods

Recombinant protein expression was qualitatively and quantitatively analysed by SDS PAGE. 12% Bis-Tris SDS PAGE resolving gel was prepared by mixing 1.88 ml of deionized water, 1.5 ml of 40% acrylamide, 1.52 ml of 1.25 M Bis-Tris (pH6.7), 50 µl of 10% (w/v) SDS (1st BASE, Singapore), 50 µl of 10% (w/v) ammonium persulphate (APS) and 2 µl of tetramethylethylenediamine (TEMED) (Biorad). This mixture was allowed to polymerize for 30 min. For preparing the 5% stacking gel, 1.1 ml of deionized water, 0.25 ml of 40% acrylamide, 0.61 ml of 1.25 M Bis-Tris (pH6.7), 20 µl of 10% (w/v) SDS and 10% (w/v) APS were mixed and poured on top of the solidified resolving gel. The gels were run at a fixed voltage of 200 V for 50 min, using MOPS running buffer containing 1% SDS. The gels were stained with Coomassie Blue Stain for 5-10 min and then incubated overnight in water for de-staining. The destained gels were scanned by the GS800 Calibrated Densitometric scanner (BioRad) and analyzed by the GelPro (Version 3.0) densitometry software.

Bradford assay was used together with the GelPro3 software to estimate the concentration of the target protein in each sample. 1 ml Bradford reagent (Biorad) was added to 20 µl protein sample and incubated at room temperature for 15 min in the dark. Sample absorbance was then measured at optical density at 595 nm (OD_{595}) using a spectrophotometer (Beckmann Coulter, DU 800). A standard curve of OD_{595} against bovine serum albumin (BSA) concentrations was plotted. The total amount of proteins in each sample was calculated by linear regression of the standard curve [116].

The HBx fusion protein was also analysed by reverse phase high performance liquid chromatography (RP HPLC) to determine protein hydrophobicity. A C5 Jupiter RP column (5 mm particle size, 300 A˚ pore size, 150 x 4.6 mm, Phenomenex) was used on a Shimadzu LC-20 AT HPLC system (Shimadzu Corporation). Tri-Fluoroacetic Acid (TFA) counter-ion (0.05% v/v) was added to all HPLC buffers. Solvent delivery and formation of acetonitrile gradients were achieved using a Shimadzu LC-20 AT solvent delivery module. The RP column was first subjected to an isocratic 20% v/v acetonitrile–water gradient for 20 min, followed by a 20-80% v/v acetonitrile–water gradient over the next 40 min. The column was then subjected to another isocratic 80% v/v acetonitrile-water gradient for 20 min before a final 20 min wash with 100% acetonitrile. The solvent flow-rate was kept constant at 0.5

ml/min. Absorbance was measured at 214 nm using a UV detector (Shimadzu SPD-20A). Samples were passed through a 0.45 µm filter prior to injection into the column. Data acquisition was performed using the Shimadzu LC solution software [LC Solutions, Shimadzu Pte. Ltd].

Aggregation kinetics during refolding was monitored by measuring sample turbidity via absorbance measurement at OD_{450} using a spectrophotometer. The protein sample was diluted in the refolding buffer and incubated at 4°C for different time points up to 10 to 12 h.

Refolding of the GST fusion protein was monitored using a Chloro-Di-Nitro Benzene (CDNB) assay kit (Merck). Bioactivity determination of GST was performed using the protocol specified by the supplier.

The pI of the fusion protein and its fragments were calculated using online Proteomics software from the website www.ExPASY.com.

All chemicals used for the experiments, unless specified, were obtained from Sigma-Aldrich.

3.3. Results and discussions

3.3.1 Characterization of the pGEX plasmid harbouring GST-HBx

DNA sequencing of the pGEX-5X-1 plasmid harbouring the GST-HBx sequence confirmed that the HBx protein possessed 154 amino acid residues. The GST fusion protein was located at the N-terminus of the HBx protein while 6 histidine residues were present at the C terminus. About 40% of the HBx sequence comprised hydrophobic amino acids while the presence of a total of 9 cysteine residues in the HBx protein sequence suggested that the HBx protein studied is a variant from the HBV subtype A3.

3.3.2. Optimising culture conditions for higher soluble GST-HBx yield

Optimisation studies to increase the soluble expression of GST-HBx were performed under various culturing conditions. The transformed BL21DE3 strains containing the pGEX-5X-1 plasmid harbouring GST-HBx were grown in LB medium, and the effects of varying IPTG concentrations and post-induction times were studied. Protein solubility or insolubility was determined by SDS PAGE analysis, where it is clear that the HBx protein was expressed in

the insoluble form in all cases (Figure 3-1). Varying IPTG concentrations from 0.25 mM to 1.0 mM did not significantly affect GST-HBx expression yield. However, upon increasing the post-induction time, protein expression increased alongside the expression of contaminant proteins for the insoluble protein fractions (Figure 3-1B).

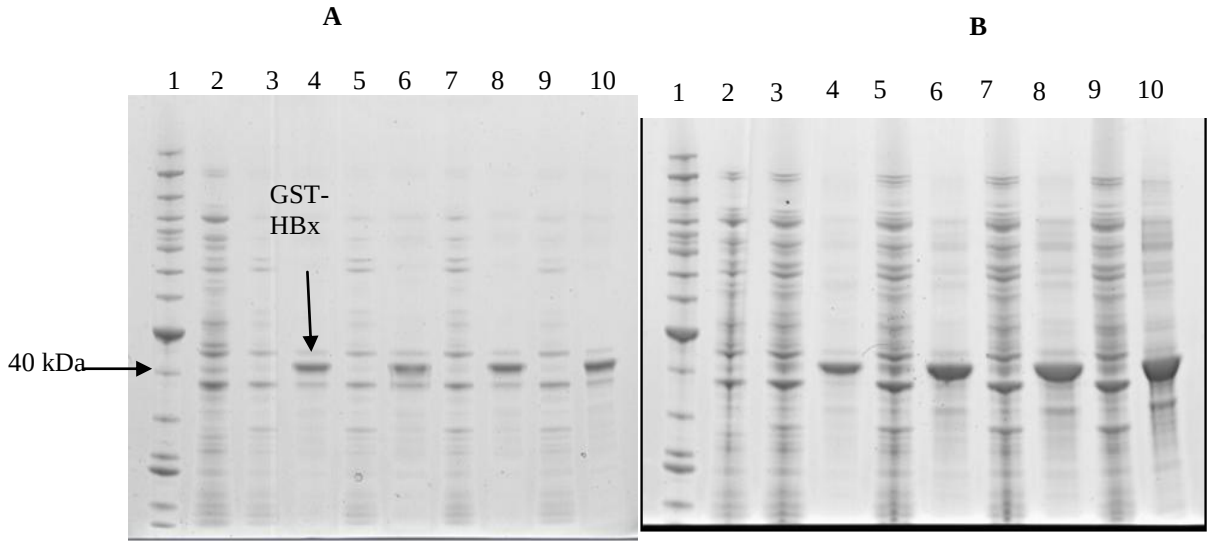

Figure 3-1. SDS PAGE analysis showing the expression of GST-HBx in LB medium with varying (A) IPTG concentrations and (B) post-induction times. Harvested cells were centrifuged, and the supernatant (soluble) and pellet (insoluble) fractions were loaded on SDS gels. Lane 1: protein marker; lane 2: un-induced culture. (A) Lanes 4, 6, 8, 10: insoluble protein fractions after induction with 0.25, 0.5, 0.75 and 1.0 mM IPTG, respectively for 1 h; lanes 3, 5, 7 and 9: soluble protein fractions after induction with 0.25, 0.5, 0.75 and 1.0 mM IPTG respectively for 1 h. (B) Cultures induced by 0.25 mM IPTG. Lanes 4, 6, 8, 10: insoluble protein fractions after 1, 2, 3 and 4 h, respectively; lanes 3, 5, 7 and 9: soluble protein fractions after 1, 2, 3 and 4 h, respectively.

Based on these results, recombinant protein expression was carried out in LB and induced with 0.25 mM IPTG at OD_{600}= 0.7 for all subsequent experiments. The IPTG-induced cells were harvested at OD_{600} = 2.0 (~1.5 to 2 h post-induction). The final yield obtained was 0.1±0.03 g GST-HBx per litre of culture.

3.3.2.1. Effect of temperature on soluble GST-HBx expression

To investigate the effects of temperature on soluble expression of GST-HBx, induction of protein expression at lower temperature and IPTG concentration (i.e. 0.1 mM) was studied in both LB and M9 minimal medium. Reduction of culture temperature has been reported to improve protein solubility [117] while the coupled use of lowered temperature and M9 medium has reportedly improved soluble protein expression by lowering overall cellular metabolic rate [106, 118]. The SDS PAGE gels of whole cell lysates cultured at 30°C and 18°C are shown in Figure 3-2. It is clear that the GST-HBx was expressed as IBs in all the

four conditions (i.e. A, B, C and D). The GST-HBx protein yield decreased when the culturing temperature was lowered to 18°C compared to 30°C. Even the reduced IPTG concentration used for inducing the protein expression, did not help in improving GST-HBx soluble yields.

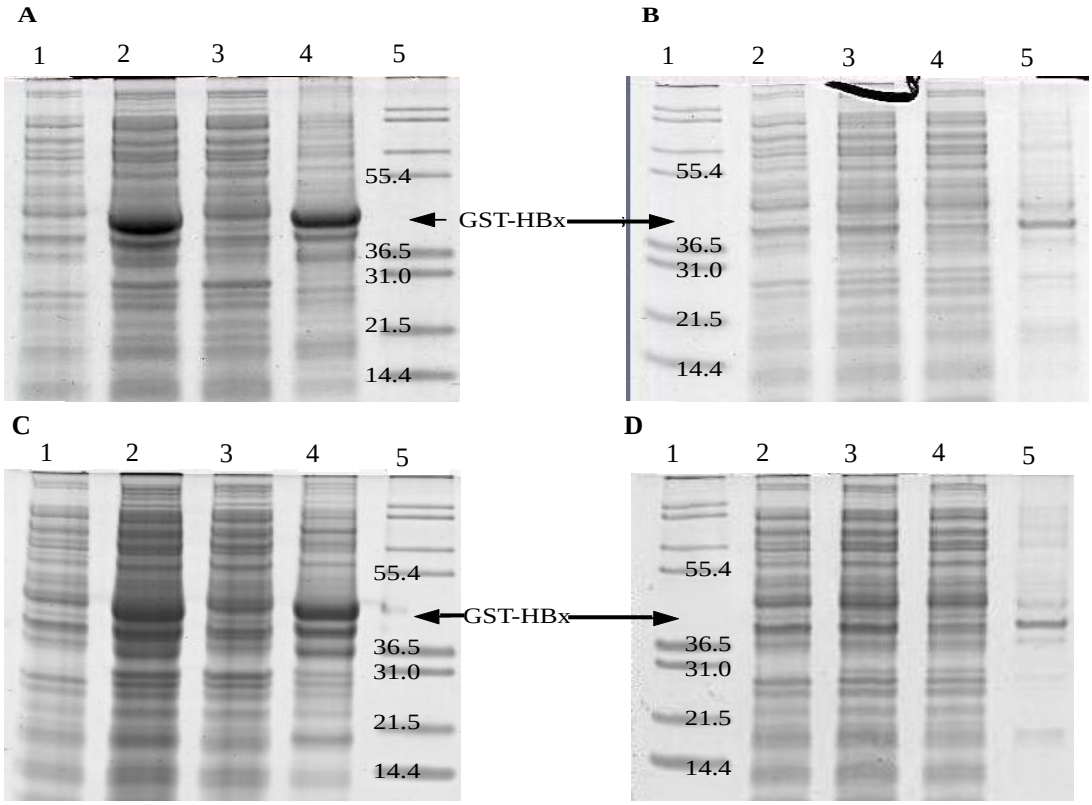

Figure 3-2. SDS-PAGE analyses of GST-HBx expression at different cell culture conditions, induced with 0.1 mM IPTG using (A) LB medium at 30°C (Lanes 1, 2, 3, 4 and 5: non-induced whole cell lysate, induced whole cell lysate, soluble cell lysate fraction, insoluble cell lysate fraction and protein marker, respectively); (B) LB medium at 18°C (Lanes 1, 2, 3, 4 and 5: protein marker, non-induced whole cell lysate, induced whole cell lysate, soluble cell lysate fraction and insoluble cell lysate fraction, respectively); (C) M9 medium at 30°C (Lanes 1, 2, 3, 4 and 5: non-induced whole cell lysate, induced whole cell lysate, soluble cell lysate fraction, insoluble cell lysate and protein marker, respectively); (D) M9 medium at 18°C (Lanes 1, 2, 3, 4 and 5: protein marker, non-induced whole cell lysate, induced whole cell lysate, soluble cell lysate and insoluble cell lysate, respectively).

3.3.2.1.1. Effect of carbon source on soluble expression of GST-HBx

The impact of carbon source on the expression of GST-HBx in M9 medium was tested at 30°C using two different nutrients: glucose and glycerol. Glucose is a common carbon source [114] and has been effective in inducing soluble expression of GST-tagged proteins using M9 medium [106]. Glycerol, on the other hand, is often used as an alternative carbon source since it produces less acetate when metabolised as compared to glucose [119]. Acetate accumulation might affect the pH within the cells and hence affect soluble protein production. From Figure 3-3, it is clear that most of the GST-HBx protein was expressed in

the insoluble fraction for both carbon sources, indicating that varying carbon sources had no significant effect on the solubility of GST-HBx.

Since the choice of medium (i.e. LB vs. M9) did not improve soluble expression of GST-HBx, no further attempts were made to express the protein in any other culture medium and LB was used throughout the work reported in this thesis.

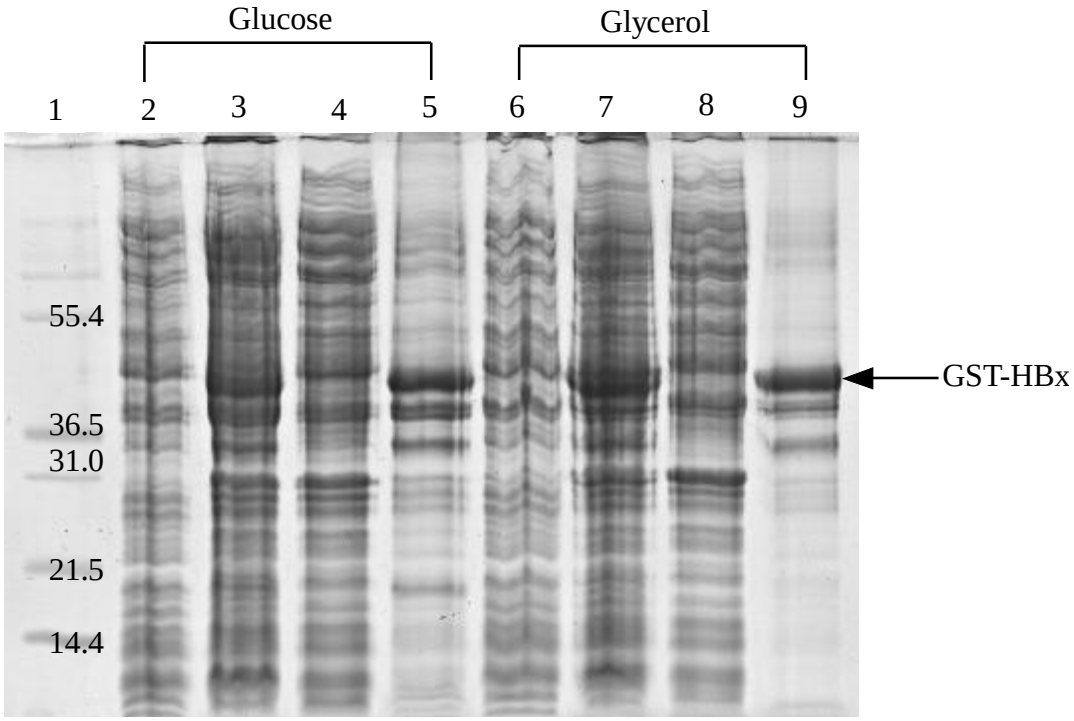

Figure 3-3. SDS-PAGE analysis of the expression of GST-HBx using glucose and glycerol as carbon source at 30°C for 5 h. Lane 1: protein marker, lanes 2 and 6: non-induced whole cell lysate, lanes 3 and 7: whole cell lysate, lanes 4 and 8: soluble cell lysate, lanes 5 and 9: insoluble cell lysate.

3.3.3. Downstream processing of GST-HBx using Strategy I

Following the completion of culture optimization studies, a conceptual downstream bioprocess flowsheet was designed for bioprocessing of GST-HBx to recover the protein in a soluble and bioactive form. A bioprocess flowsheet that involves IB solubilisation followed by a refolding step was designed (Figure 3-4). The GST tag could then be removed from the fusion protein by cleavage using Factor Xa protease. Sections 3.3.3.1 to 3.3.3.2.3 discuss the results obtained from each processing step of this strategy.

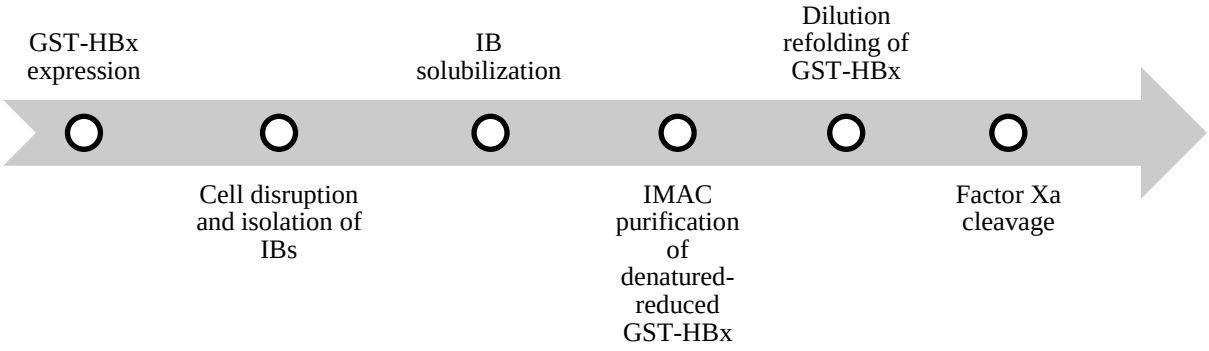

Figure 3-4. Steps involved in bioprocessing of GST-HBx using Strategy I.

3.3.3.1. Solubilisation and purification of GST-HBx IBs

The recovered IBs were first solubilised in the IMAC binding buffer (Buffer A) to facilitate subsequent IMAC purification by FPLC. Urea was chosen over guanidine HCl (GuHCl) as the preferred chaotrope for IB solubilisation in this study because it is less corrosive, cheaper and can efficiently solubilise proteins [47]. Urea often degrades at high solution concentrations and temperatures to produce negatively charged cyanates that can lead to carbamylation of proteins. To eliminate this problem, all urea solutions were treated with Amberlite which is an ion exchange resin that can mop up the negatively charged cyanates from the urea solutions [115]. DTT (10 mM) was also added in the IB solubilisation buffer to reduce disulphide bonds, rendering the target protein completely unfolded and free from intra- or intermolecular disulphide bonds.

Following the solubilisation step, a pre-refolding purification of the target protein was performed. Since the HBx protein is highly hydrophobic, the protein is expected to possess a strong tendency for aggregation with its own refolding intermediates and also other contaminating proteins. Additionally, the free cysteine residue of the HBx protein variant used in this study will increase likelihood for intermolecular covalent interactions. Therefore, purification of the GST-HBx protein prior to refolding is critical to minimise contaminant-induced aggregation of the HBx proteins at the onset of the bioprocess. Using the IMAC purification protocol described in Section 3.2.5.1, protein recovery of ≥50% at concentrations of 4-5 mg/ml and sample purity of >90% was obtained from each IMAC purification cycle.

3.3.3.2. Solubilisation studies of GST-HBx under non-denaturing conditions by dilution refolding, followed by Factor Xa cleavage

In designing the GST-HBx refolding process, it was important to consider the physicochemical environment requirements of the cleavage reaction. The supplier's recommended cleavage conditions for Factor Xa are as follows:

a) The substrate (i.e. GST-HBx protein) should be soluble in the cleavage buffer to allow hydrolysis activity of Factor Xa on the substrate.

b) The optimum pH range for Factor Xa activity is 6-8 (i.e. cleavage reaction needs to be performed at this pH range).

Since the theoretical isoelectric pH (pI) of GST-HBx is 6.98, GST-HBx is likely to lose its charged state and thus aggregate when solubilised at the pH 6-8 range and the refolding intermediates would thus be susceptible to pI induced aggregation within the optimal pH range (i.e. 6-8). To maintain solubility of GST-HBx, it was therefore considered necessary to conduct the cleavage reaction at pH values away from pH 7±1 but still within the pH range that supports Factor Xa activity. As such, refolding had to be performed at pH 8 or 6. Besides pH, variation in the refolding buffer composition (i.e. redox and refolding additives) was also performed to increase HBx solubility and Factor Xa cleavage efficiency. The effects of varying the refolding buffer composition on HBx solubility are presented in Figure 3-5 and discussed in the following sections.

3.3.3.2.1. GST-HBx solubility studies at pH 8-9

Denatured-reduced HBx was diluted into 3 refolding buffer systems to a final GST-HBx concentration of 0.05 to 0.1mg/ml. Buffer 1 consisted of 2 M urea, 50 mM Tris, 1 mM GSH and 0.1 mM GSSG at pH 8. The addition of GSSG and GSH was considered important to facilitate oxido-shuffling of the disulphide bonds which is favoured at a pH range of 8-9, where the GSH molecules can easily form thiolate anions. Generally, a GSH and GSSG ratio of 10:1 is widely used in renaturation buffers, with the GSH concentration in the range of 0.1–1 mM [120-121]. In most cases, a slightly net reducing environment helps in oxido-disulphide shuffling, where mispaired disulphide bonds easily re-shuffle to form the correct

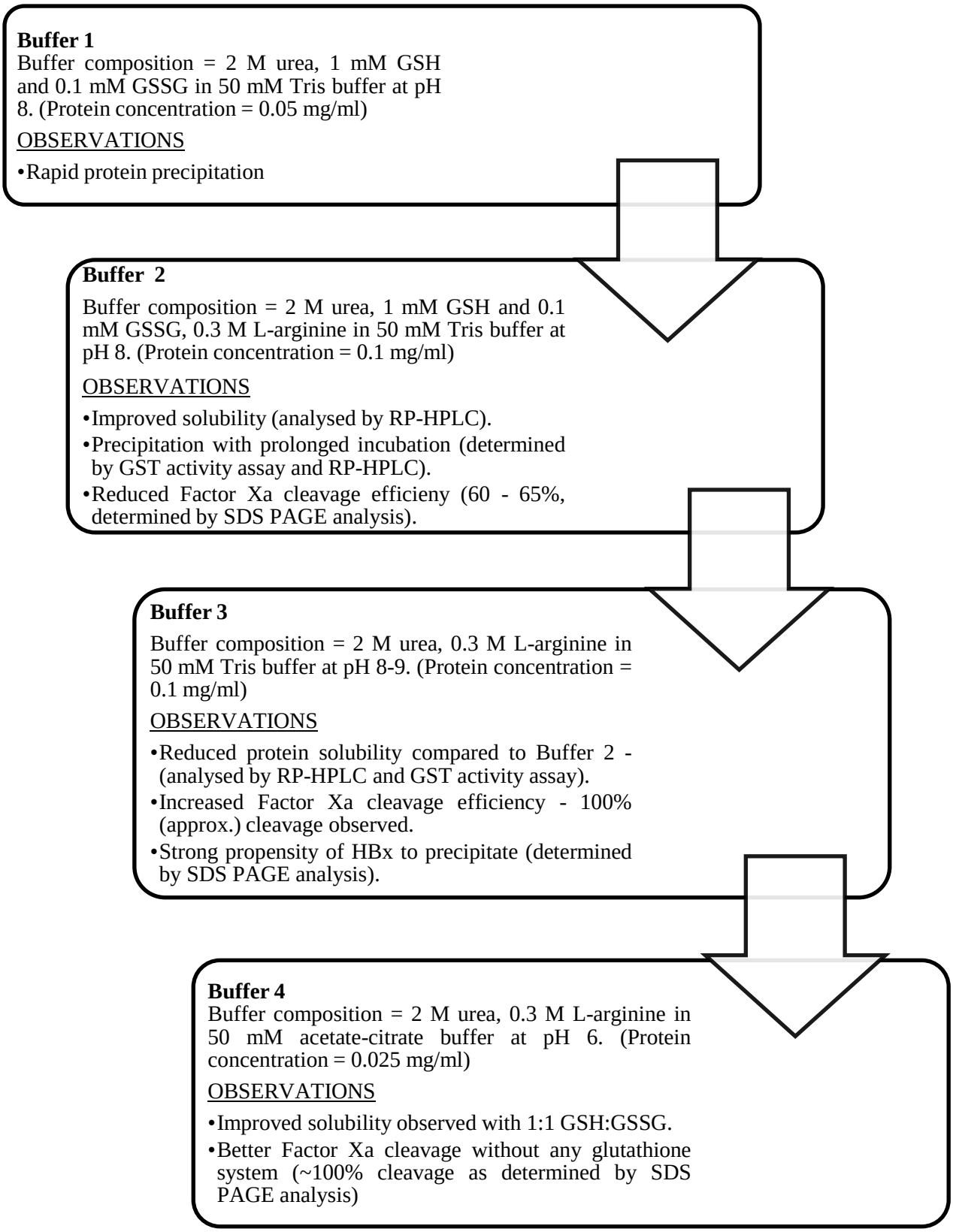

Figure 3-5. The effect of varying refolding buffer composition on HBx protein solubility.

native disulphide bond. A non-denaturing urea concentration (i.e. 2 M) was used to maintain a mild denaturing environment within the refolding buffer to minimise hydrophobicity-induced aggregation. Buffer 1 was unable to support GST-HBx solubility, as indicated by the visible turbidity of the protein solution when denatured protein was diluted into the refolding buffer. Since this rapid increase in turbidity suggested that the protein had aggregated, Buffer 1 was not considered further in subsequent experiments. To suppress aggregation, a widely used protein aggregation inhibitor, L-arginine, was added into the refolding buffer [122] and this buffer composition was named Buffer 2. Buffer 2 increased protein solubility, where no visible aggregates were observed after 10–12 h of incubation at 4^{o}C following dilution of the denatured protein into the buffer (although aggregates started appearing at 24 h of incubation). To prevent compromising the bioactivity of Factor Xa protease (which has one disulphide bond) by the net reducing environment imposed by excess GSH, the glutathione redox couple was removed from Buffer 2, which formed Buffer 3 (2 M urea, 0.3 M L-arginine in 50 mM Tris buffer, pH 8; i.e. Buffer 2 without GSH and GSSG). Buffer 3, however, was found to be poorer in protein solubilisation ability compared to Buffer 2 (i.e. solution turbidity was observed within 10-12 h of incubation) indicating the importance of disulphide shuffling in maintaining solubility and stability of the GST-HBx protein.

Apart from visual observations, GST-HBx solubility in Buffers 2 and 3 was also determined using RP HPLC. The reduction in elution peak area for GST-HBx of the RP HPLC chromatograms of the refolding protein samples (Figure 3-6) indicated reduced quantities of the target protein in solution. The fact that a higher elution peak area was observed for GST-HBx in Buffer 2 compared to Buffer 3 suggests that GST-HBx has a higher solubility in Buffer 2 than Buffer 3. The consistent retention time of the GST-HBx peak indicated that there was no change in the hydrophobicity of the protein following incubation in the refolding buffer.

The protein samples incubated in Buffers 2 and 3 were also subjected to GST activity assays. The assay results indicated reduction in GST activity with higher incubation times for both Buffers 2 and 3 (Figure 3-7), which corresponds with decreased sample solubility with increasing incubation time, as observed by the turbidity studies.

3.3.3.2.2. Factor Xa cleavage efficiency in Buffers 2 and 3

Buffers 2 and 3 were used for subsequent Factor Xa cleavage studies. The cleavage reaction was performed in the refolding buffer after addition of NaCl and $CaCl_2$ to a final concentration of 150 mM and 1 mM, respectively. Due to the correlation between Factor Xa protease activity and binding to Na^+ and Ca^{2+} ions, the addition of NaCl and $CaCl_2$ was important to improve the efficiency of the cleavage reaction [123-124]. The cleavage reaction was first performed in Buffer 2. The enzyme : substrate ratio was varied from 1:5 to 1:50 by mass, but complete cleavage of GST-HBx was not observed even after 72 h of incubation. The maximum cleavage achieved at the highest enzyme : substrate ratio was approximately 60 – 65% after 72 h of incubation.

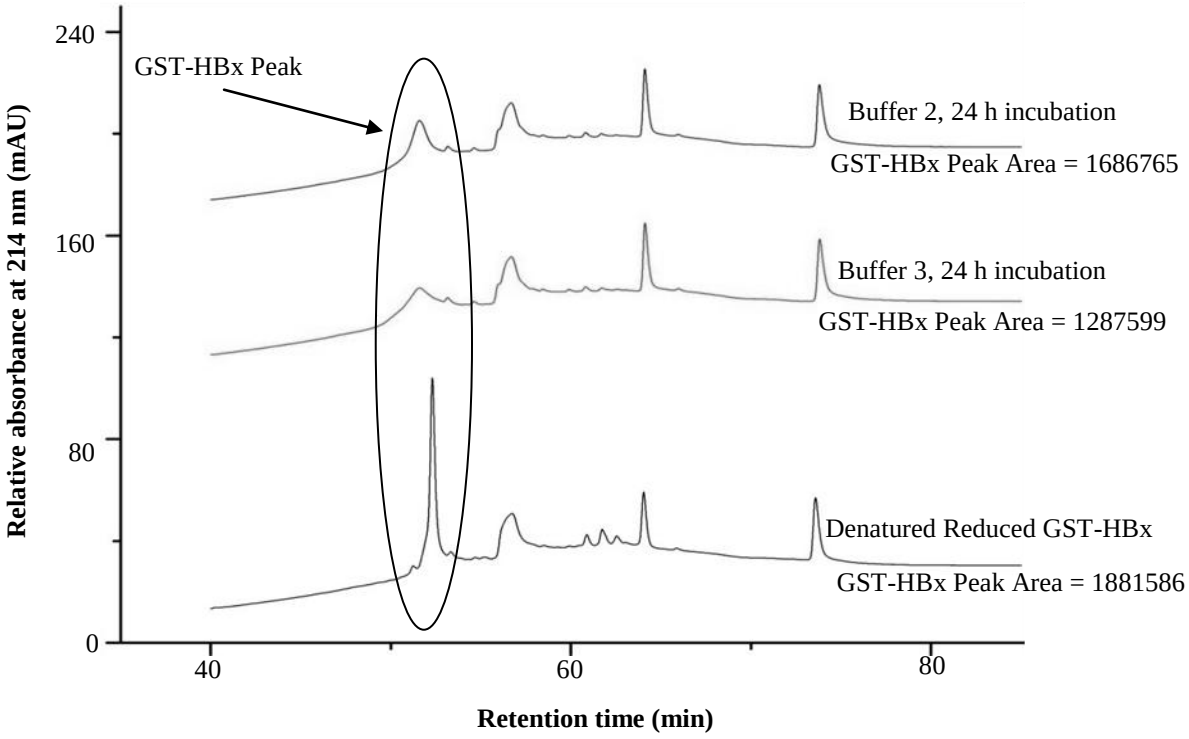

Figure 3-6. RP HPLC chromatogram of the refolded fusion protein after incubation in different refolding buffers. Quantity of protein loaded in all cases was 1 µg.

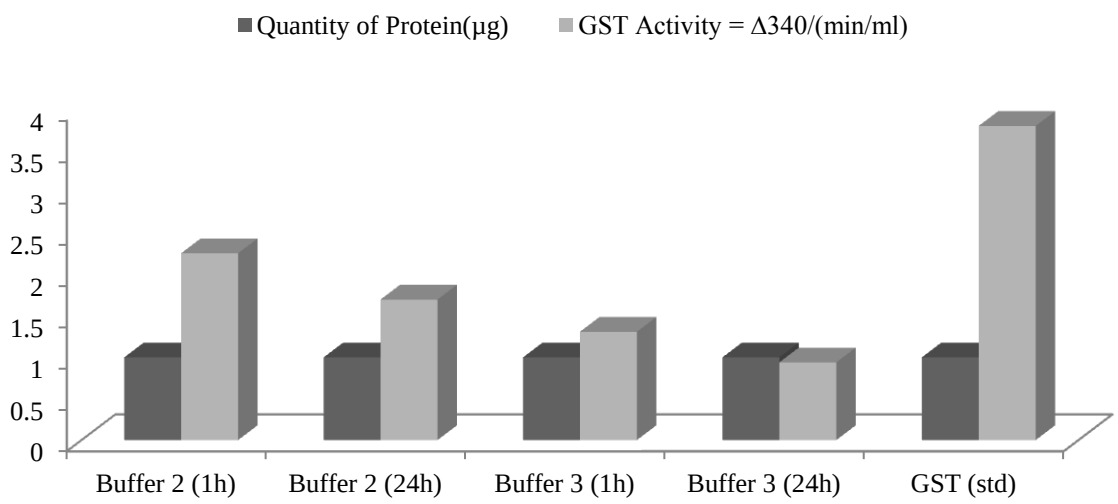

Figure 3-7. GST activity of the GST-HBx protein after incubation in Buffer 2 and 3; the respective incubation periods are indicated in the x-axis. The last set of bars corresponds to the amount and activity of a GST protein standard.

Cleavage of GST-HBx was next studied in Buffer 3. Since the post-refolding sample solubility was lower in Buffer 3 compared to Buffer 2 (discussed in Section 3.3.3.2.1), it was necessary to identify the optimum time following incubation in the refolding buffer after which the cleavage reaction should be started. Since the cleavage reaction cannot occur in aggregating molecules, the optimum time point for the initiation of the reaction would be the point when the aggregation induced turbidity of the solution is minimum. The turbidity kinetics of GST-HBx when refolded in Buffer 3 was subsequently studied and the results are presented in Figure 3-8. In order to identify the impact of pH on the fusion protein solubility, the pH of Buffer 3 was also varied from 8-9.

The results of the turbidity experiments indicated that irrespective of the refolding buffer pH, increase in turbidity of the protein solution could be expected after 2 to 3 h of protein dilution into refolding buffer. In fact the solutions started becoming visibly turbid at OD450 0.02-0.03, following which fluctuation in the solution turbidity was observed. This may be due to the precipitation of the formed aggregates. Based on these results, the time point of 2 h post-refolding was chosen for the initiation of the cleavage reaction, where complete cleavage of the substrate was obtained at an enzyme : substrate ratio of 1:2.5 (by mass) at room temperature after 48 h incubation (Figure 3-9). To maintain activity of the Factor Xa protease, the final pH of the refolding and cleavage buffer was kept at 8.

Since the molecular mass of the GST protein is approximately 1.5 times that of HBx, the expected (nominal) quantities of HBx recovered post-cleavage was calculated based on the amount of GST present in the SDS-PAGE gel, quantified by the Gel Pro software. The quantity of HBx recovered, as quantified by the Gel Pro software was determined (Table 3-1), where approximately close to 50% of HBx could be recovered post-cleavage in Buffer 3.

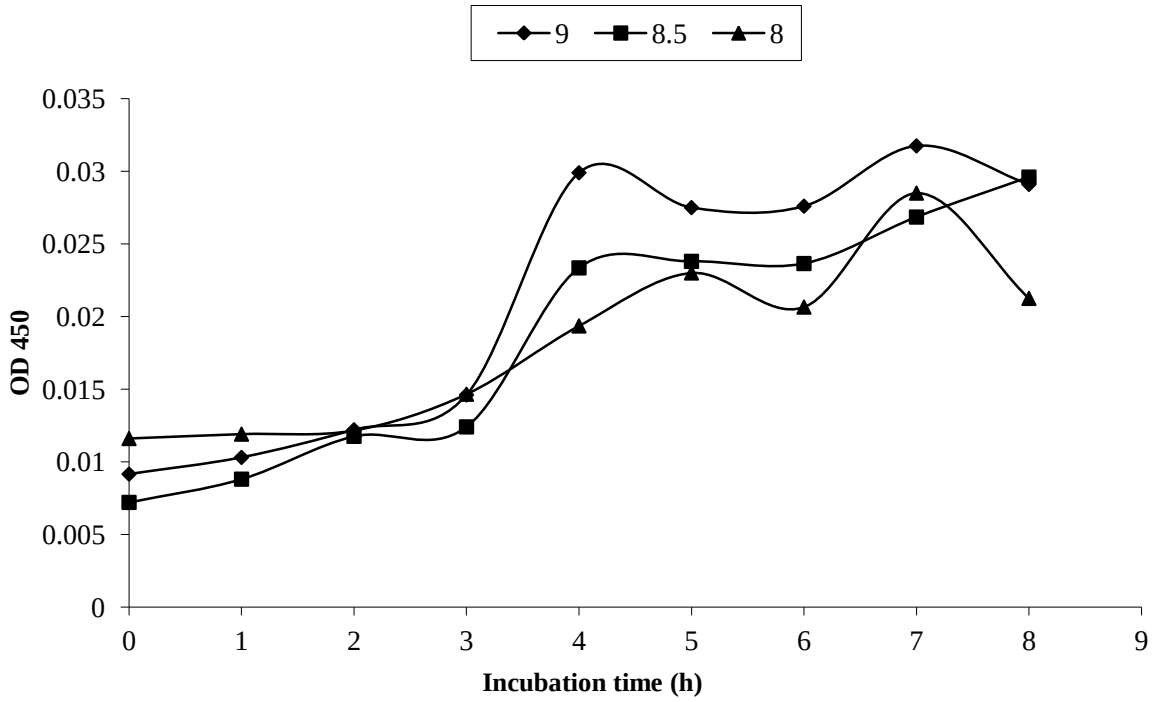

Figure 3-8. Turbidity kinetics of the refolding sample diluted to 0.1 mg/ml in Buffer 3 at different pH values.

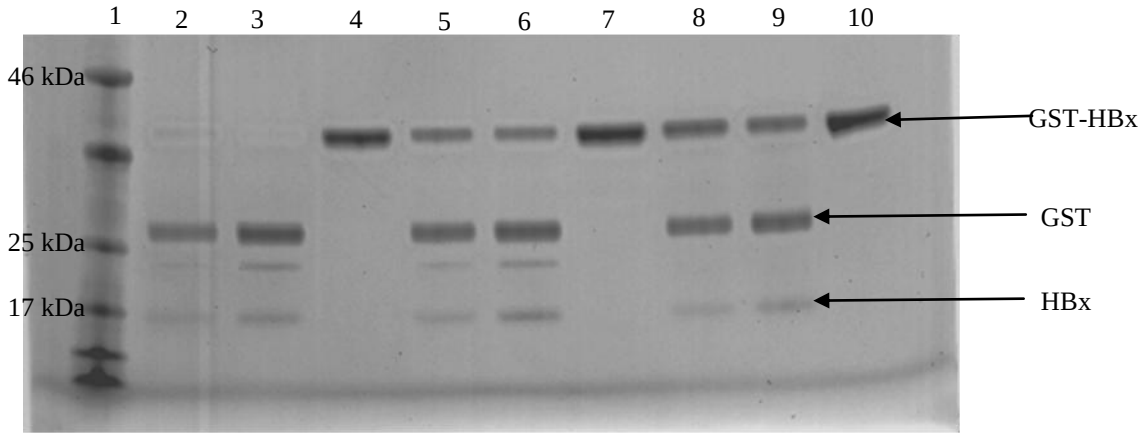

Figure 3-9. SDS-PAGE analysis of the GST-HBx cleavage efficiency at a Factor Xa : substrate ratio of 1: 2.5 (by mass) with varying cleavage incubation time at pH = 8. Lane 1: protein marker; lanes 2 and 3: 48 h, lanes 5-6: 24 h and 8-9: 12 h, lanes 4, 7 and 10: corresponding controls without any treatment with Factor Xa.

Table 3-1.HBx protein recovery after cleavage.

Amount of GST detected (ng)	Nominal amount of HBx (ng)	Amount of HBx detected (ng)	% Recovery of HBx
263	172	80	47
264	172	97	56
310	202	97	48
353	231	113	49

Considering that the refolding pH was within ± 0.5 pH units to the HBx pI (i.e. 8.6), it is therefore highly possible that the loss of HBx could be attributed to pI-induced precipitation [121]. It was hypothesized that lowering the refolding buffer pH could help to reduce the occurrence of HBx aggregation. Therefore, a refolding pH of 6 was subsequently studied (i.e. Buffer 4), where the HBx protein would, in theory, be less prone to pI-induced precipitation.

3.3.3.2.3. GST-HBx solubilisation and cleavage studies at pH 6

Amongst the various buffer systems that could be used to maintain pH at 6, the acetate-citrate buffer system was chosen for cost reasons. Two buffer systems were selected for dilution refolding: (i) Buffer 4 (2 M urea and 0.3 M L-arginine in 50 mM acetate-citrate buffer), and (ii) Buffer 4A (0.1 M urea and 0.3 M L-arginine in 50 mM acetate-citrate buffer), both at pH 6. The final protein concentration within the refolding buffer was kept at 0.025 mg/ml. Turbidity was observed after 12 h and 6 h of protein incubation in Buffers 4 and 4A, respectively, which indicates the importance of keeping urea concentration modestly high to suppress hydrophobic-induced aggregation during refolding.

The Factor Xa cleavage efficiency of the fusion protein in the presence of Buffer 4 and Buffer 4A is shown in Figure 3-10. From the SDS PAGE results, complete cleavage could be observed upon incubation of GST-HBx with Factor Xa for 48 h in Buffer 4 containing 0.15 M NaCl and 1 mM $CaCl_2$. Complete cleavage could not be observed in other buffer systems, which contained

varying ratios of GSH-GSSG at a total molar concentration of 1.1 mM. Based on these results, Buffer 4 was found to be the most optimum refolding and cleavage buffer system for HBx bioprocessing. However, subsequent scale-up efforts to perform the cleavage reaction at higher volumes and protein concentrations resulted in reduced yield coupled with protein precipitation.

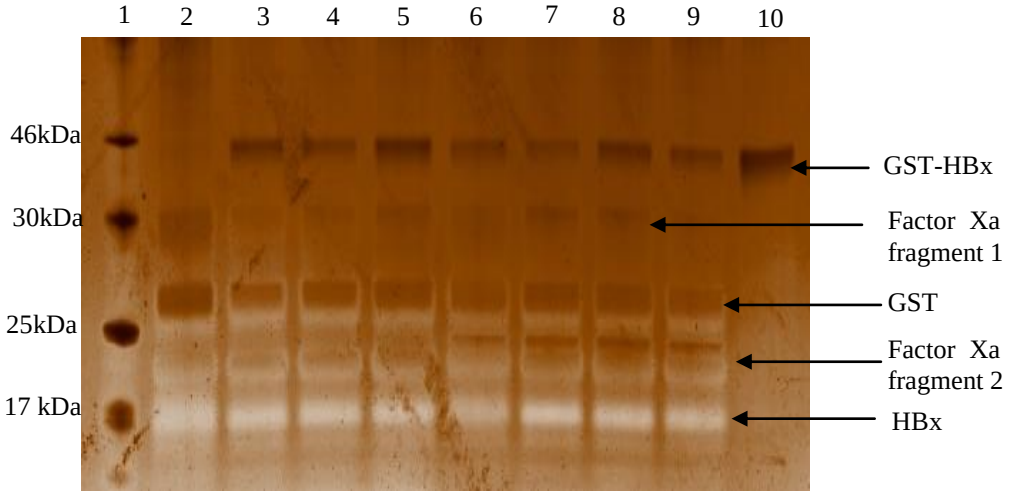

Figure 3-10. SDS PAGE analysis of the GST-HBx cleavage reaction performed in Buffers 4 and 4A stained using a silver staining kit. NaCl and $CaCl_2$ were added to a final concentration of 150 mM and 1 mM respectively, to all cleavage reaction mixtures before Factor Xa addition. Lane 1: protein marker, lane 2: cleavage in Buffer 4; lanes 3-5: cleavage in Buffer 4 at various GSH and GSSG ratios (10:1, 1:1 and 1:10, respectively); lane 6: cleavage in Buffer 4A; lanes 7-9: cleavage in Buffer 4A at various concentrations of GSH and GSSG (10:1, 1:1 and 1:10, respectively); lane 10: no Factor Xa treatment.

3.3.4. Downstream processing of GST-HBx using Strategy II

Considering that Strategy I failed to recover HBx in the soluble form, a separate strategy based on HBx solubilisation in the cell lysis step was studied. The conceptual bioprocess design for detergent-based solubilisation of GST-HBx (i.e. Strategy II) is illustrated in Figure 3-11. The aim of this strategy is to recover HBx in the soluble and preferably bioactive form by solubilising the GST-HBx IBs under non-denaturing conditions. Sarkosyl, which has been reported to selectively solubilise GST-tagged proteins was used to stabilise GST-HBx under non-denaturing conditions in this study [107]. The solubilised fusion protein could then be purified in a single step by binding the fusion protein to GS beads, followed by Factor Xa cleavage.

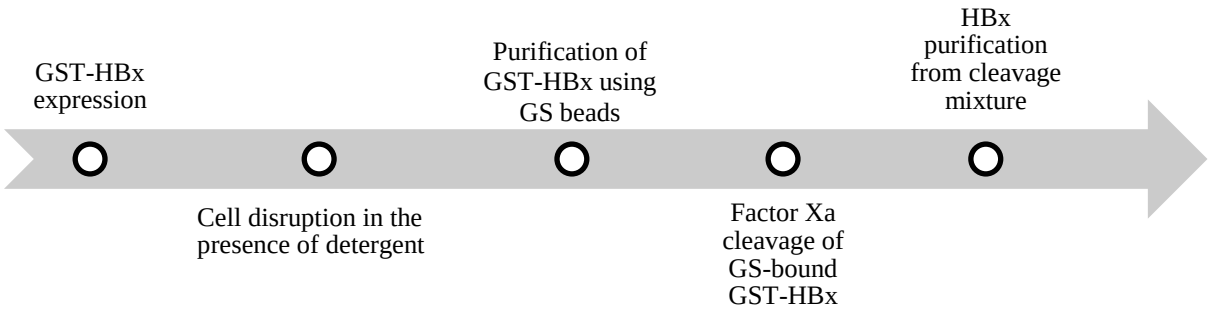

Figure 3-11. Conceptual bioprocess flowsheet for Strategy II.

3.3.4.1. Effect of detergents on GST-HBx protein solubility during cell lysis

The first objective in this set of studies was to establish whether the detergents studied could solubilise the GST-HBx protein during the cell lysis step. Sarkosyl and other detergents including SDS and Triton X-100 were added to the pre-sonicated cell suspensions. For initial studies, sarkosyl concentration in STE lysis buffer was set to 1.5% (w/v) and the SDS concentration was maintained to be the same to that of sarkosyl. Triton X-100 concentration was set to 1.05% (w/v) based on the concentration reported in other studies for protein solubilisation purposes [107]. The SDS PAGE profile of the respective soluble and insoluble protein fractions clearly indicates that sarkosyl was more selective than the other detergents in solubilising GST-HBx (Figure 3-12). The solubility of the GST-HBx protein also seems to be significantly

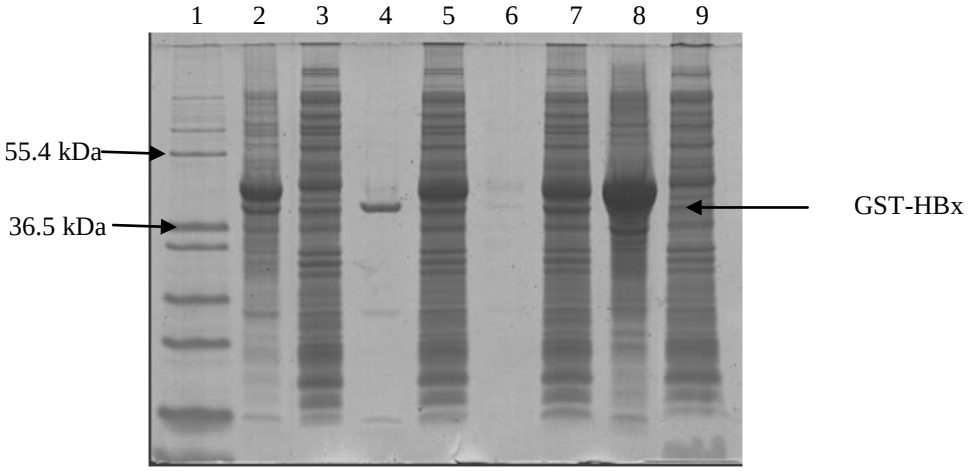

Figure 3-12. Effect of different detergents used in solubilising GST-HBx. Lane 1: protein marker; lanes 2, 4, 6, 8: insoluble proteins from cell lysates which do not contain any detergent, 1.5% (w/v) sarkosyl, 1.5% (w/v) SDS and 1.05% (w/v) Triton X-100, respectively; lanes 3, 5, 7, 9: cell lysate supernatant with no detergent, 1.5% (w/v) sarkosyl, 1.5% (w/v) SDS and 1.05% (w/v) Triton X-100, respectively.

improved as indicated by a higher fraction of the protein recovered in the soluble fraction compared to the insoluble fraction (i.e. Figure 3-12, lanes 4 and 5, respectively). The absence of protein bands in lane 6 reflects the denaturing effect of SDS which can non-specifically solubilise all proteins within the insoluble cell lysates fraction. Triton-X 100 did not have any effect on improving the GST-HBx solubility. These results confirmed the capability of sarkosyl to selectively solubilise the GST-HBx protein in the cell lysates. Recovery of the HBx protein in the correctly refolded or bioactive conformation is highly desirable to eliminate the need for subsequent refolding step, which can lower yields. Further optimisation studies on sarkosyl concentration showed that a final concentration of 0.2% (w/v) yielded complete solubilisation of the GST-HBx IBs. A lower concentration of the detergent would not only benefit the overall process economics, but at concentrations lower than its critical micelle concentration (CMC) (i.e. 0.428% (w/v) [125]), sarkosyl would have less interference on the binding of GST-HBx on the GS beads used in the subsequent protein purification step.

3.3.4.2. Binding and cleavage of GS-immobilised GST-HBx

Purification of GST-HBx from other contaminant proteins was performed using GS beads. To minimise the interference of sarkosyl on the binding of GST-HBx to the GS beads, Triton X-100, a non-ionic detergent, was employed to sequester sarkosyl carried over from the lysis step. The clarified cell lysates were incubated with GS beads at different concentrations of Triton X-100 and the bound GST-HBx was eluted at a high concentration of glutathione (20 mM). Figure 3-13A shows that majority of the GST-HBx bound to the GS beads, and could be recovered after elution from the beads. For binding of the GST tag to the glutathione moiety on the beads, the GST must be in a conformationally active form. Therefore, it can be concluded that the GST-tag of the fusion protein was in the refolded conformation to allow binding to the glutathione moiety on the beads. No significant improvement in the binding capacity of GST-HBx could be observed beyond 2% (w/v) of Triton X-100. Subsequent optimization studies also showed that increasing the volume of GS beads beyond 200 µl did not further increase the amount of GST-HBx bound. This result suggests that the misfolded GST moieties present in the sample failed to bind to the GS beads. The GS beads having bound GST-HBx were washed and re-suspended in the cleavage buffer before subsequent addition of the Factor Xa enzyme to the desired enzyme / substrate ratio.

Upon cleavage of the fusion protein, the HBx protein is expected to remain soluble in the cleavage buffer and hence can be easily separated from the GS beads by centrifugation. The bottleneck for the cleavage reaction is the high enzyme to substrate ratio required which is prohibitively expensive. The SDS PAGE gel characterising the cleavage reaction at 1:5 enzyme : substrate ratio is presented in Figure 3-13B. It is clear that only partial cleavage of the fusion protein occurred despite a prolonged cleavage incubation time of 16 h. It is possible that the HBx protein have either precipitated or remained strongly bound to the GS beads. Recovery of HBx from this partially cleaved protein mixture would require further separation operations and probably refolding of the HBx protein, which can lead to yield losses. Improved Factor Xa cleavage efficiency was not observed when purification was performed on Ni-NTA beads instead of GS beads (data not shown).

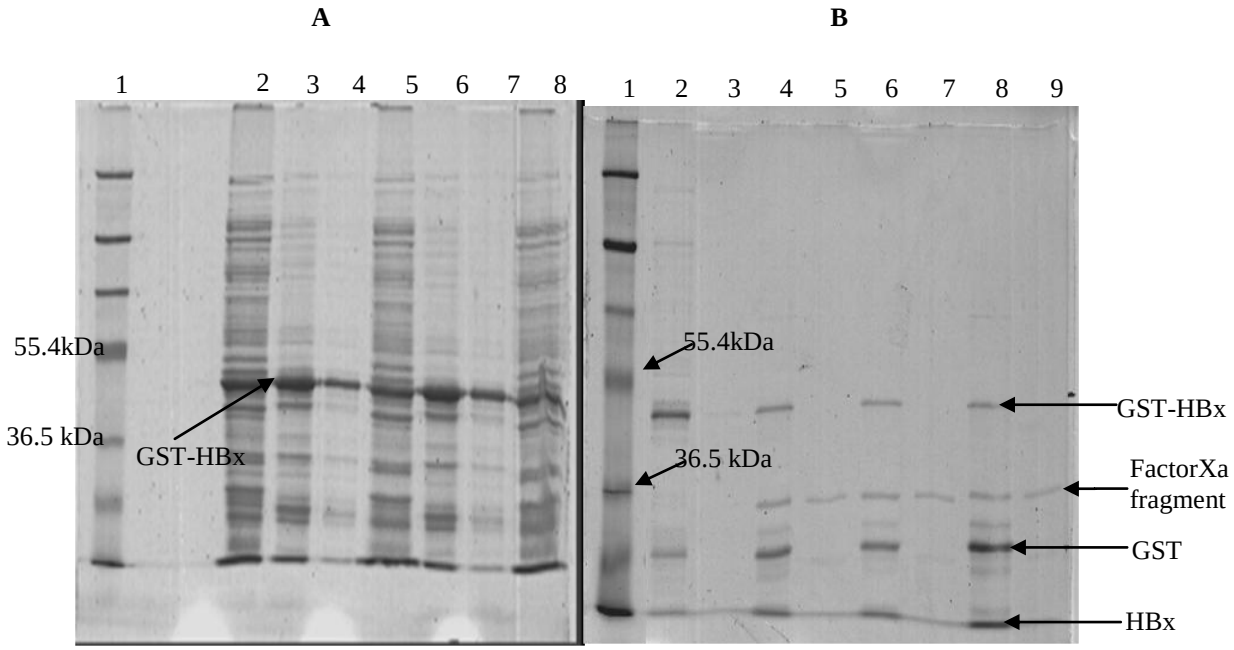

Figure 3-13. (A) Analysis of GST-HBx binding to GS beads with 2% (w/v) and 4% (w/v) of Triton X-100 in binding buffer containing 0.2% (w/v) sarkosyl. Lane 1: protein marker, lane 2: cell lysate prior to addition of Triton X-100, lanes 3 and 6: GS beads with proteins bound in the presence of 4% (w/v) and 2% (w/v) Triton X-100 present in the binding buffer, respectively, lanes 4 and 7: eluted protein fractions obtained from GS beads corresponding to lanes 3 and 6, respectively, lanes 5 and 8: flowthrough obtained after binding with 4% (w/v) and 2% (w/v) Triton X-100 present in the binding buffer, respectively. (B) Determination of cleavage efficiency at enzyme to substrate ratio = 1:5. Lane 1: protein marker, lanes 2 and 3: GS beads and the supernatant after 16 h incubation with no Factor Xa added, respectively, lanes 4, 6, 8: GS beads after 4, 8 and 16 h incubation, respectively, lanes 5, 7, 9: supernatant of GS beads cleavage mixture after 4, 8 and 16 h incubation, respectively.

3.4. Conclusions

This chapter shows the importance of a close interlink between product design and its bioprocessing considerations. Inspite of extensive optimisation of culture conditions, the GST-HBx could only be produced as IBs, necessitating a refolding based bioprocess. The successful cleavage of the GST-HBx protein was observed both at pH 6 and 8 (using Strategy I), but the high enzyme to substrate ratio required makes scale-up efforts difficult and costly using this strategy. Strategy II successfully solubilised the GST-HBx protein, which facilitates the separation of misfolded GST-HBx on the GS beads. However, recovery of the 'Factor Xa'-cleaved HBx still requires subsequent purification and concentration steps, thus making the bioprocess significantly complex.

It is apparent that the Factor Xa cleavage step, being disadvantaged by the narrow (pH) window of operation and optimal activity conditions, is the main bottleneck for bioprocessing the GST-HBx, thus necessitating an improved construct design for easy protein bioprocessing. Considering the pI of the different components of the GST-HBx construct (Figure 3-14), it is apparent that one of its components (pI of GST, GST-HBx and HBx are 5.9, 7 and 8.6,

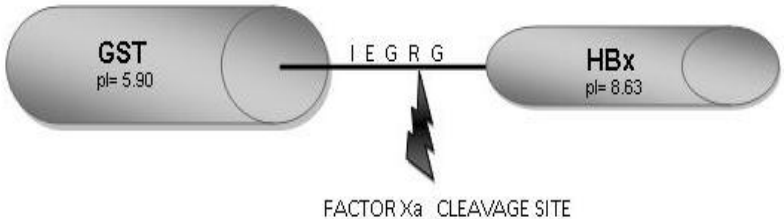

Figure 3-14. pI values of the fusion protein components.

respectively) would have no net charge at the pH range of 6-8, which is needed for optimum Factor Xa protease activity, leading to protein precipitation. Furthermore, the need for Factor Xa to bind Na^{+} and Ca^{2+} ions for maximum activity demands an environment of relatively high ionic strength, further aggravating the propensity of pI-induced precipitation of the protein moieties. Thus, despite the successful stabilisation of the GST-HBx in solution using sarkosyl employed by Strategy II, a complete cleavage of the fusion protein could not be achieved.

The results presented in this chapter are to some extent in concert with the earlier findings of the HBx production strategy with MBP-fusion partner [110]. As discussed previously, being a highly solubility enhancing fusion partner MBP helped in the soluble expression of the HBx gene, but the inefficient fusion protein cleavage coupled with expensive post-cleavage purification steps led only to sub-optimal HBx yields. It is thus reasonable to say ~~clear~~ from this study that product design for novel protein candidates demands careful consideration of its bioprocessing requirements. A highly complex downstream processing flowsheet such as that necessitated by the GST-HBx construct, which was kindly provided by a collaborator, is undesirable because it is expensive, laborious and difficult to scale-up or validate. These challenges merit further studies to design a new HBx construct without a fusion partner, which forms the aim of Chapter 4. The absence of a fusion partner will ease out the bioprocessing challenges confronted with protein cleavage operations.

CHAPTER 4: Study of Protein Intermolecular Interactions to Determine Rational Methods for HBx Refolding

Abstract

The chief objective of this chapter is to develop a scalable bioprocess for the bulk production of the HBx protein construct devoid of any solubility enhancing fusion tag. The HBx protein was expressed in the insoluble form in spite of extensive culture optimization studies involving modifications of expression conditions as well as the *E. coli* strains (i.e. through chaperone co-expression systems). A refolding based bioprocess was thus imperative for designing a scalable HBx bioprocess. To guide refolding studies of HBx *in vitro*, it was considered important to establish a quantitative method to correlate HBx solubility with the protein's physicochemical environment. Specifically, SVC measurements were employed to quantify the influence of refolding-relevant physicochemical environment on HBx solubility behaviour *in vitro*. Denatured-reduced HBx was diluted into buffers having different compositions, and SVC measurements were performed on each sample. The negative SVC values exhibited by HBx in a highly denaturing environment such as 6 M urea (without adding the GSH:GSSG redox couple) indicated an attractive HBx protein-protein interaction results mainly due to intramolecular and/or intermolecular cysteine mispairing. The increase in positive SVC values with increasing GSH:GSSG ratio indicated that a net reducing environment improves HBx solubility by enhancing the oxido-shuffling of disulphides. This observed redox-dependent variation in SVC is not unexpected, considering that the model HBx protein is stabilised by 4 disulphide bonds and a free cysteine residue. Thus, using an SVC optimized refolding buffer, the dilution refolded HBx remained stable in solution over a prolonged period of time (at least 60 h). In line with the SVC results, size exclusion chromatography HPLC (SEC-HPLC) and reducing and non-reducing SDS-PAGE analyses of the HBx refolding samples revealed that the protein was not subjected to disulphide-induced inter-molecular aggregation. Circular dichroism (CD) spectroscopy showed that the refolded protein adopted a predominantly random coil secondary structure. The ability of the refolded HBx protein to bind with its substrate, the p53 protein, confirmed the biological activity of the refolded protein. The optimised refolding based bioprocess presented in this chapter can be used for bulk manufacturing of the HBx protein for further structural and drug designing studies. This chapter also presents the applicability of SVC measurements for rapid determination and optimisation of refolding conditions for novel protein candidates like HBx.

4.1. Introduction

In Chapter 3 of this thesis, problems associated with bioprocessing a GST-tagged HBx protein were discussed, where the use of a GST tag as a fusion protein failed to improve soluble expression of HBx. The ensuing complexities linked to poor enzymatic cleavage efficiency of Factor Xa, an expensive commercial enzyme, coupled with additional post-cleavage protein recovery steps that increased step yield losses indicated that bioprocessing the GST-HBx construct is likely to be futile. Efforts were thus directed towards designing a process for a new HBx construct without any solubility-enhancing fusion partner. The HBx protein used in the new construct is from HBV subtype A3. HBx expression from this new construct resulted in formation of HBx IBs despite efforts to optimise culture conditions to promote soluble HBx expression. From a bioprocessing perspective, IB formation can be beneficial in many respects: apart from high product yields, resistance towards proteolytic cleavage, minimised product toxicity towards the host cell [126], and relative ease of purification of the target protein from other contaminants are strong advantages [127] for bulk manufacturing of the protein "[72]". However, designing a refolding process for a novel protein like HBx, which lacks detailed intermolecular interaction studies, is particularly challenging because the physicochemical forces influencing the protein refolding reactions are unknown. Such information is critical in determining the optimum conditions necessary for refolding a highly hydrophobic protein consisting of an odd cysteine residue system, where the refolding reaction would be complicated by the hydrophobic and covalent interactions leading to off-pathway protein aggregation reactions. The primary objective of this chapter is thus to characterise HBx stability (in solution) at the molecular level under different environmental conditions.

As discussed in Chapter 2, the use of a SLS analytical platform allows SVC measurements of the HBx protein solutions, which provide vital thermodynamic information on protein solubility [75]. Since a positive SVC value of the HBx solution would mean that the protein molecules repel each other because of their affinity towards the solvent molecules and a negative value would mean the opposite, SVC measurements can be used to evaluate the impact of diverse environmental conditions on HBx through a statistical DoE methodology. Statistical tools used in the DoE methodology can yield valuable information on the impact of experimental parameters which include: (i) the solution environment consisting of various refolding-assisting

components and (ii) their corresponding interactions on the dependent variable (or the response variable, SVC) and hence are widely used for bioprocess optimization studies [128]. Such experimental strategies can thus be used to determine the type of interactions between the HBx molecules within different environmental conditions. Subsequently, the results obtained from the DoE studies can be verified by independent protein solubility or aggregation studies, and used to determine the optimum refolding conditions for the HBx protein, thus laying the foundations for a rationally designed dilution refolding based process for HBx production.

4.2. Materials and methods

All chemicals used for the experiments performed were obtained from Sigma Aldrich unless stated otherwise.

4.2.1. HBx plasmid transformation

The sequence of HBx which originates from the HBV subtype A3 (Isolate Cameroon) is given as follows (cysteine residue underlined):

MAARLYCQLDSSRDVLCLRPVGAESRGRPFSGPLGTLSSPSPSAVSSDHGAHLSLRGLPV
CAFSSAGPCALRFTSARCMATTVNAHQILPKVLHKRTLGLPAMSTTDLEAYFKDCLFKD
WEELGEEIRLKVFVLGGCRHKLVCAPSSCNFFTSA

The codon-optimised HBx gene was cloned between the NcoI and XhoI restriction sites of a pET28b+ vector by Geneart. The plasmid harbouring the HBx gene, which is referred to as pET28b-X throughout the thesis, would yield HBx proteins with 6-Histidine (6-His) tags at the C-terminus. The pET28b-X plasmid was transformed into *E. coli* BL21DE3RIL (Stratagene) by heat shock. The transformed cells were grown overnight on a kanamycin (50 µg/ml) selective LB agar medium and inoculated into LB medium (50 ml), also containing 50 µg/ml kanamycin. The cells were then grown overnight (12 h) in a shaking incubator at 37^{o}C and 200 rpm.

4.2.2. HBx cell culture optimisation studies

To investigate the effects of temperature and co-expression of chaperones on the soluble expression of HBx, the pET28b-X plasmid was also incorporated into two different *E. coli* strains, (i) *E. coli* BL21 DE3 (Stratagene) harbouring the pG-KJE8 chaperones plasmid (Takara

Bio Inc.), and (ii) ArcticExpress™ (DE3) RIL (Stratagene), respectively. The pG-KJE8 chaperone plasmid system employed can co-express a system of five different chaperones (i.e. dnaK, dnaJ, grpE, groES and groEL). The ArcticExpress™ (DE3) RIL strain is specifically engineered to induce soluble expression of recombinant proteins at low temperatures [129]. The effects of medium composition on soluble expression of HBx was also investigated, where the expression profile of HBx cultured in LB and M9 were compared at high (30°C) and low (18°C) temperatures.

4.2.3. Recombinant HBx protein expression and IB recovery

Overnight culture (obtained as described in Section 4.2.1) was inoculated into 400 ml LB medium containing 50 µg/ml kanamycin to obtain a 1% bacterial suspension. The bacterial cells were grown at 37^{o}C and 200 rpm until late log phase (OD_{600} = 0.7±0.1), followed by 0.5 mM IPTG induction to induce HBx expression. The cells were harvested 2 h post-induction and the culture medium containing the cells was centrifuged at 5200 g for 30 min. The cell pellets obtained were washed with PBS and then re-suspended in 80 ml of 50 mM Tris-HCl buffer (pH 8). 10 ml of this bacterial suspension was disrupted using a sonicator (Branson sonifier 450, 400 watts) at 30% amplitude strength (Maximum frequency = 20000 Hz) for 3 min with 5 s pulse on and 7 s pulse off. The HBx IBs were recovered by centrifugation at 6500 g for 30 min at 4^{o}C.

4.2.4. HBx protein identification using Western Blot

Western blot of the insoluble protein fraction after cell lysis was performed using monoclonal antibodies (Millipore) generated against the HBx antigen and anti-mouse IgG-AP (alkaline phosphatase). Visualization of the bands was performed with Chemiluminescent Alkaline Phosphatase (AP) substrate (Immobilon Western, Millipore) detection system. Luminescence from the stained bands was captured by exposing the blots to a medical film (Konica Minolta) and scanned by an image scanner (Image Scanner II, Amersham Biosciences).

4.2.5. HBx protein preparation for SVC measurements

For SVC measurements, the HBx IBs were IMAC purified using a 5 ml Ni^{2+}Sepharose HisTrap FF column (GE Healthcare) using an AKTA Explorer FPLC System (GE Healthcare). The IBs

were solubilised under denaturing-reducing conditions in IMAC Binding Buffer (8 M urea, 1 M NaCl, 10 mM DTT, 20 mM Imidazole, 50 mM Tris, pH 7.4) and then bound to the Ni^{2+}Sepharose FF column which was also equilibrated in the same buffer, at 1.0 ml/min mobile phase flow rate. Column washing was performed with 2 CV of IMAC Binding Buffer, and bound proteins were eluted with 100% IMAC Elution Buffer (8 M urea, 1 M NaCl, 10 mM DTT, 1 M Imidazole, 50 mM Tris, pH 7.4) at 5.0 ml/min flow rate. NaCl, Imidazole and DTT were removed from the protein solutions using a PD10 desalting column (GE Healthcare) in 8 M urea and 50 mM Tris, pH = 7.4. The HBx solutions were subsequently diluted in refolding buffer containing varying concentrations of urea, arginine, and reduced to oxidised glutathione (GSH:GSSG) ratios, to the desired HBx concentrations (0.01 – 0.5 mg/ml) for SLS experiments. All protein solutions were centrifuged at 12000 g for 25 min and filtered through 0.2 µm filters prior to SLS measurements. All buffers used for SLS experiments were also filtered using 0.2 µm filters.

4.2.6. Determination of SVC using SLS

The SVC value for each HBx protein sample was measured by performing SLS experiments using a Zetasizer Nano (ZS) instrument (Malvern Instruments Ltd.). The Zetasizer uses a non-invasive back scatter (NIBS) technology for performing SLS measurements, where a 633 nm laser is focused on the HBx samples and the Rayleigh ratio measured using a detector at an angle of 173°. All measurements were performed at room temperature. SVC values of protein solutions were determined using a Debye plot according to Equation 4-1 [130].

$$\frac{Kc}{R(\theta)} = \frac{1}{M_w} + 2B_2c \qquad (4\text{-}1)$$

R(θ) is the excess Rayleigh ratio (cm^{-1}) of the HBx protein solution under investigation, *K* (mol cm^2/g) is the light-scattering optical constant, *c* is the protein concentration (g/ml). *K* depends on the solutions' scattering properties and is given by Equation 4-2.

$$K = \frac{4\pi^2[n_s(dn/dc)]^2}{N_A\lambda^4} \qquad (4\text{-}2)$$

N_A is the Avogadro's number (mol^{-1}), λ is the wavelength of the incident light (cm), n is the refractive index of the protein solution and n_s is the refractive index of the solvent. n_s for each solvent used was determined using a pocket refractometer (Atago). HPLC grade toluene was used as the scattering standard for all light scattering experiments. HBx concentrations were obtained by measuring sample absorbance at 280 nm using extinction coefficient ε_{280} = 0.5 L/g.cm [calculated using the ProtParam tool in the Expasy server (http://expasy.org/tools/-protparam.html)].

4.2.7. Studying the effect of physicochemical environment on HBx solubility by DoE methodology

A DoE-based study was designed to study and determine a combination of the experimental variables (i.e., the concentrations of urea, L-arginine and the redox potential) that can yield a positive value of SVC for the HBx solution. Two main reasons have been hypothesized for the insoluble behaviour of the expressed recombinant HBx proteins. Firstly, the amino acid composition of the HBx protein renders the protein strongly hydrophobic, thereby leading to hydrophobicity induced aggregation. Secondly, the significant number of cysteine residues (i.e. eight to ten depending on the variant of HBx, nine in the construct reported in this study) in the HBx primary structure can lead to mispairing or incomplete oxidation of cysteines. An optimum physicochemical environment that can address these instability problems is critical in designing the HBx refolding process. The concentrations of urea, L-arginine and GSH:GSSG ratio were varied in the refolding buffer. Urea was chosen as a variable that can control protein hydrophobicity in solution, arginine is an anti-aggregating agent capable of preventing non-specific protein aggregation reactions, while GSH:GSSG can control the redox environment to induce oxido-shuffling of the cysteine residues.

SVC values for the HBx protein solutions following IMAC purification were measured under statistically designed experimental conditions. Using the Minitab 15 software (Minitab Inc.), a two level full factorial statistical DoE approach (consisting of one centre point) was adopted for this study. The three parameters used in the DoE study were: (i) urea concentration, (ii) L-arginine concentration, and (iii) GSH:GSSG molar ratio (Table 4-1). All SVC measurements

were performed thrice to determine the influence of each parameter and parameter interactions on HBx solubility. The turbidity of HBx samples was determined by absorbance measurement at 600 nm using a spectrophotometer (Beckman Coulter, DU 800).

4.2.8. Dilution refolding of HBx

The recovered HBx IBs were first denatured-reduced and solubilised in 8 M urea and 10 mM DTT (pH 7.4) and buffer-exchanged into 8 M urea using a PD10 column (GE Healthcare). The DTT-free denatured HBx protein was rapidly diluted into the optimised refolding buffer (2 M urea, 0.1 mM GSH, 0.01 mM GSSG and 0.25 M L-arginine in 50 mM Tris buffer, pH 7.4) to a final HBx concentration of 0.1 mg/ml under stirring conditions. After stirring for 2 min, the solution was incubated in the refolding buffer at 4°C for at least 60 h.

4.2.9. SEC-HPLC analysis

An analytical SEC column (BioSep-SEC-S3000, 300 x 7.8mm) (Phenomenex) was used to analyse the HBx samples at different time points. SEC analysis was performed using a HPLC system (LC-20 AT, Shimadzu) at room temperature. A constant mobile phase flow rate of 1.0 ml/min was used throughout the analysis. The mobile phase composition was 2 M urea, 0.25 M L-arginine in 50 mM Tris buffer (pH 7.0). HPLC peak integration was performed using the LC solution software (Shimadzu).

4.2.9.1. Protein monomericity determination by SEC-FPLC

To determine the protein composition in chromatogram peaks obtained from SEC-HPLC analyses, refolded HBx samples were fractionated using a preparative SEC column (Superdex 200 10/300 GL, GE Healthcare) operated on an AKTA Explorer FPLC system (GE Healthcare). The flow rate of the mobile phase (2 M urea, 0.25 M L-arginine in 50 mM Tris buffer, pH 7 – 7.4) was kept at 0.5 ml/min. Protein fractions from the SEC column were concentrated using Vivaspin concentrators (Sartorious) before SDS PAGE analysis.

4.2.10. p53 interaction test for HBx bioactivity determination

After 60 h of incubation in the optimised refolding buffer, refolded HBx bioactivity was determined by measuring the interaction of the refolded protein with the human tumour

suppressor protein p53. Ni-NTA resin (30 µl) (BioRad) was added to three 100 µl aliquots of refolded HBx (buffer exchanged into 50 mM Tris-HCl) solution. The first and second refolded HBx aliquots contained 5.7 µg GST and 15 µg GST-p53 (Millipore) respectively, which yielded equimolar GST concentrations in both aliquots. The third refolded HBx aliquot did not contain any GST and served as a negative control. After 2 h incubation under mild stirring conditions, the Ni-NTA resins were recovered by centrifugation, and washed with 50 mM Tris-HCl buffer. Bound proteins were subsequently eluted by incubating the resins with equal volumes of 2x Laemmli buffer and heating at 100°C for 10 min. The eluate was subjected to SDS-PAGE analysis.

4.2.11. RP-HPLC analysis

Conformational changes of the HBx protein following refolding was analysed by RP HPLC. The RP column was equilibrated in 10% v/v acetonitrile–water gradient for 10 min, and a 10-80% v/v acetonitrile–water gradient was employed over the next 30 min, followed by column stripping with 80 % v/v acetonitrile-water gradient for 10 min. Absorbance was measured at 214 nm using a UV detector (Shimadzu SPD-20A). All samples were passed through a 0.2 µm filter prior to injection into the column.

4.2.12. Circular dichroism (CD) spectroscopy

Far- and near-UV CD spectra of the refolded and denatured HBx protein were measured using a spectropolarimeter (Chirascan, Applied Photophysics) at room temperature. Quartz cuvettes with 1 and 10 mm path-lengths (Hellma) were used for far- and near-UV CD measurements, respectively. All CD spectra were corrected by subtracting the buffer baseline, and then averaged over 5 scans for all measurements. Denatured HBx was solubilised in 8 M urea and 1 mM DTT (pH 7.4). 0.5 mg/ml and 1 mg/ml HBx protein were used for far- and near-UV spectra, respectively. The far-UV CD spectrum of refolded HBx was deconvoluted using the CDNN software package provided with the CD equipment. The predicted data was verified using the Hierarchical Neural Network (HNN) secondary structure prediction tool from the Expasy server (http://expasy.org/tools/#proteome).

4.2.13. Other analytical protocols

Aggregation kinetics for the HBx protein molecules, when incubated within different refolding buffers, was tested by measuring the OD600 at different time points. Protein refolding concentrations were maintained at 0.1 mg/ml for all the samples.

All protein samples were centrifuged and the supernatant filtered using a 0.45 µm filter to remove insoluble proteins prior to SDS-PAGE and Bioanalyzer analyses.

All gel electrophoreses were performed with 12% Bis-Tris gels and 50 mM MOPS running buffer, pH = 7.0 (containing 1% Sodium Dodecyl Sulphate) at 200 V constant voltage for 40 min. For analysis of samples under denatured-reduced conditions, equal volumes of the sample and 2x Laemmli buffer were mixed and heated at 100°C before loading into the gels. For analysis of samples under denatured but non-reduced conditions, the sample was mixed with Nupage 4x LDS sample loading buffer (Invitrogen) which did not contain reducing agent. Protein band intensity was determined using a calibrated densitometer (GS-800 Densitometer, BioRad), while HBx purity was estimated using the Gel Pro3 software to correlate protein band intensity with protein amount.

HBx soluble yield and purity at different processing steps were determined using an Agilent 2100 Bioanalyzer® in combination with the Protein 80 Plus LabChip® kit. The Agilent 2100 Bioanalyzer® detection is based on laser-induced fluorescence of an intercalating dye, which interacts with protein-SDS complexes. Denatured-reduced HBx (from 20 to 100 µg/ml, measured using Bradford assay with BSA a standard) was used to generate a calibration curve. The samples and chips were prepared according to the protocol provided with the Protein 80 Plus LabChip® kit. All samples used for the analysis were centrifuged and filtered using 0.2 µm filter, prior to analysis.

4.3. Results and discussions

4.3.1. Optimising HBx gene expression conditions

Recombinant HBx was expressed predominantly in the insoluble form in the *E. coli* BL21(DE3)RIL strain (Figure 4-1A, lane 5). The identity of the full length HBx protein was confirmed by Western Blot (Figure 4-1B). Extensive optimisation studies were conducted to understand the effects of varying culture conditions on HBx protein solubility during expression (Figure 4-2A). The use of lower temperatures (i.e. 30°C and 18°C) for protein expression was hypothesised to improve correct folding because the formation of properly folded proteins, having lower entropy than their denatured counterparts, would be favoured at lower temperatures [117]. The effect of media composition on HBx expression was also investigated, where earlier studies have shown that the osmotic stress provided by the chemically defined minimal media often promotes soluble expression of recombinant proteins [106, 131]. At 30°C, HBx was still predominantly expressed in the insoluble fraction, while at 18°C, the expression yield of HBx was extremely poor (Figure 4-2B). These results showed that soluble HBx expression could not be improved by modifying culture media and temperature, which were in agreement with previously reported protein expression studies, albeit for different HBx constructs [17, 104]. Attempts to improve protein solubility were continued by expressing the HBx protein at lower

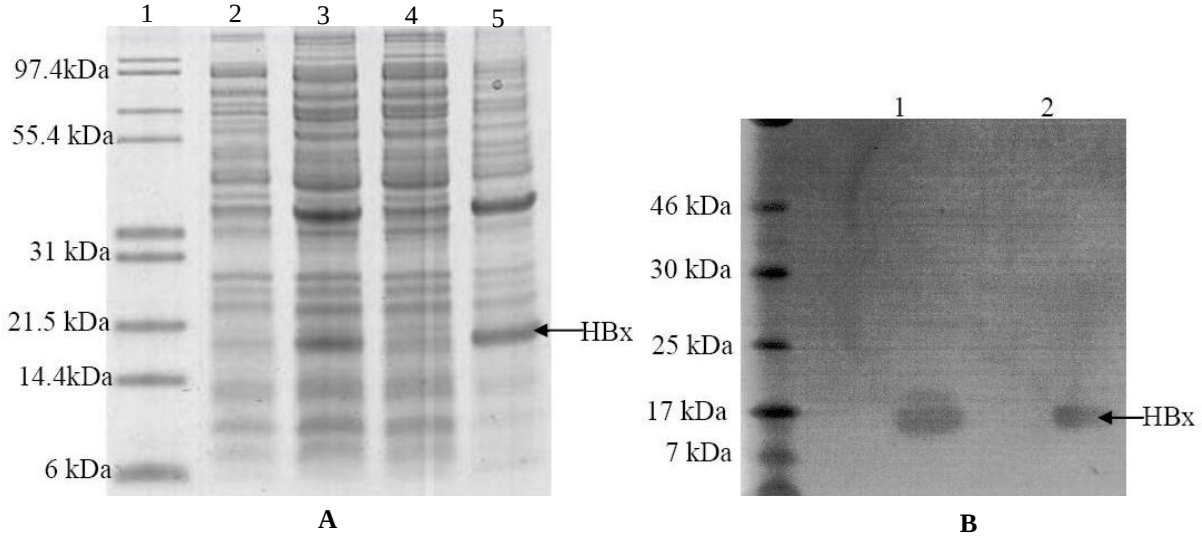

Figure 4-1. (A) SDS PAGE analysis of the protein expression profile of *E. coli* (BL21 DE3 RIL), harbouring the plasmid containing HBx-6His. Lane 1: protein marker; lanes 2 and 3: un-induced and induced whole cell proteins, respectively; lanes 4 and 5: soluble and insoluble protein fractions from IPTG-induced cells, respectively. (B) Identification of the HBx protein within the insoluble protein fraction using Western Blotting. Lanes 1and 2 represent blot corresponding to 1.5 µg and 1 µg total insoluble protein fraction load.

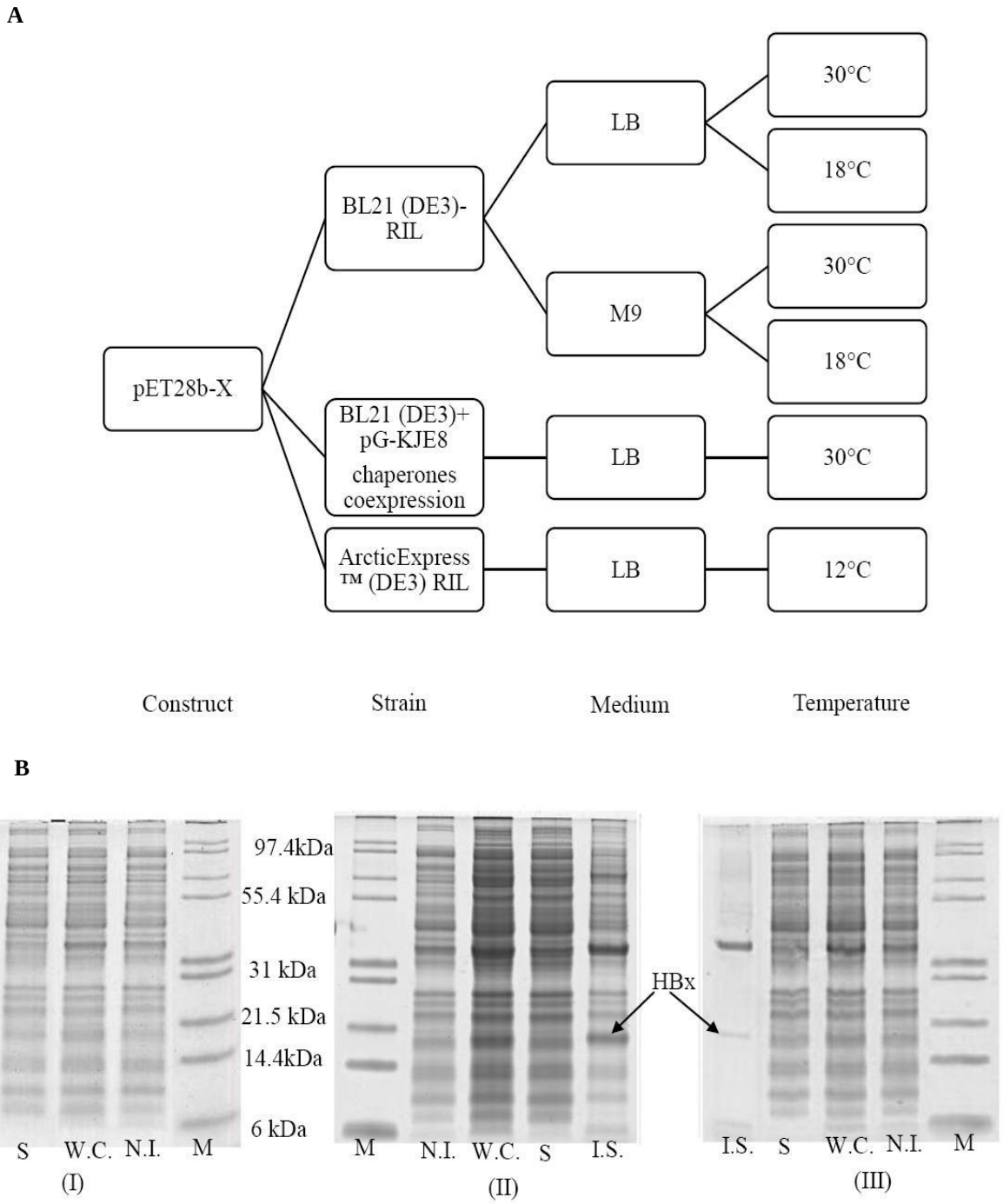

Figure 4-2. (A) Investigation of culture parameters to promote soluble HBx expression. (B) SDS-PAGE analyses of the expression of HBx-6His at different cell culture conditions. (I) LB medium, 18°C. (II) M9 medium, 30°C. (III) M9 medium, 18°C. (M) - Protein Marker, (N.I) - Non-induced cell lysate, (W.C) – Induced whole cell lysate, (S) - Soluble fraction, (I.S) - Insoluble fraction.

temperatures (12°C) using the (i) Arctic Express DE3 strain, and (ii) chaperone co-expression system, where the HBx protein was expressed in an *E. coli* strain harbouring the pG-KJE8 plasmid which produces the following five chaperones: Dna-K, Dna-J, Grp-E, Gro-EL and Gro-ES. As shown in Figure 4-3, HBx expression was significantly repressed at 12°C and no improvement in expression yield could be observed even when IPTG induction was extended to 24 h, suggesting that low-temperature stress was unfavourable for HBx expression. The use of the chaperone co-expressing strain also did not improve the soluble expression of the HBx protein. These results clearly demonstrate that the HBx protein was recalcitrantly insoluble in *E. coli* and hence a refolding-based bioprocess was the most reasonable way forward for large scale HBx production, which forms the basis for the mode of HBx production reported in this chapter and subsequent chapters.

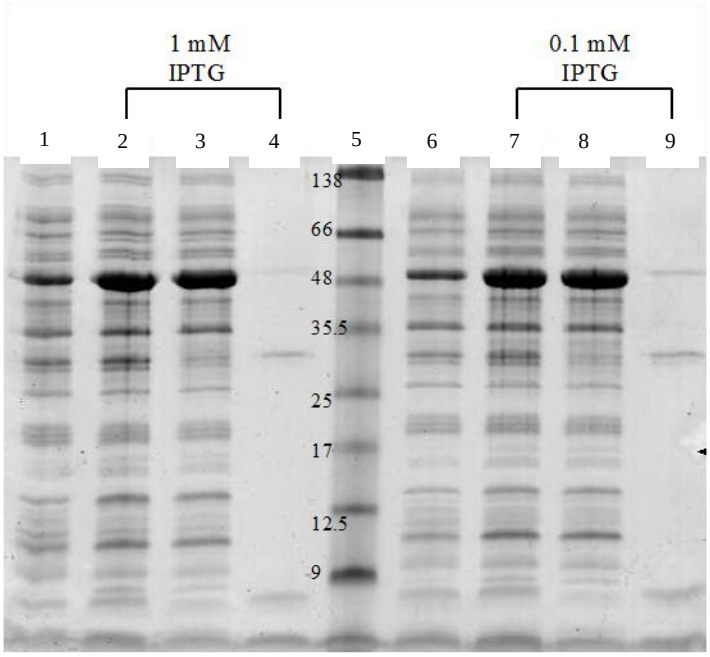

Figure 4-3. SDS PAGE analysis of the expression of HBx-6His using ArcticExpress™ strain. Lanes on the left half of the gel represent cells induced by 1 mM IPTG, while those on the right represent cells induced by 0.1 mM IPTG. Lanes 1 and 6: non-induced cell lysates; lanes 2 and 7: induced whole cell lysates; lanes 3 and 8: soluble cell lysates fraction; lanes 4 and 9: insoluble cell lysates fraction; lane 5: protein marker. The molecular mass units of the protein markers are in kDa.

A HBx IB expression yield of 49±6 mg HBx per liter of cell culture was obtained under optimised cell culture conditions (i.e. 37°C, 0.5 mM IPTG, LB medium, 2 h post induction time). This expression yield was approximately 40-fold higher than the recently reported soluble HBx expression yield in *E. coli* [110]. Besides an expression yield advantage, insoluble expression of

HBx also yielded a relatively pure target protein (i.e. 70% purity), which can be directly used for dilution refolding, thus significantly simplifying bioprocessing of HBx, compared to a fusion protein derived HBx.

4.3.2. HBx solubility studies by SVC measurements

SVC measurements (as obtained from the SLS experiments) were employed to determine the effects of different physicochemical environment on HBx solubility using the statistical DoE methodology. Following IB solubilisation, the denatured-reduced HBx protein was purified by Ni^{2+} metal affinity chromatography purification to remove cellular contaminants (detailed in Section 4.2.5). The purified denatured-reduced HBx was diluted into buffers containing varying concentrations of urea, L-arginine and GSH:GSSG ratio, for SVC measurements. The SVC results are presented in Table 4-1. The relative magnitude and statistical significance of each varying parameter, as well as parameter interaction effects obtained from the analysis of data are shown using a Pareto chart in Figure 4-4. It is clear that GSH:GSSG ratio, as a single parameter, had the most significant impact on the SVC values of the HBx solutions, while L-arginine had the least impact. It is also clear that parameter interactions played a significant role in varying the HBx SVC values, which confirmed a complex dependence of SVC on the range of refolding additives studied.

In accordance to the order of importance (Figure 4-4), the interaction effects of GSH:GSSG ratio versus urea and L-arginine, were next studied. Figures 4-5A and 4-5B show the intricacies of interaction amongst parameters, where the changes in mean SVC values with urea (Figure 4-5A) and L-arginine (Figure 4-5B) concentrations at a fixed GSH:GSSG ratio were plotted. Each point in Figure 4-5A was normalised against L-arginine concentration at any given combination of GSH:GSSG ratio and urea concentration (i.e. the plot was obtained from the mean of all the SVCs recorded at data points that contain 0 and 0.5 M L-arginine in combination with the concerned levels of urea and GSH:GSSG ratio). Therefore, Figure 4-5A represents the average change in the SVC due to a change in the urea concentration at a given GSH:GSSG ratio, irrespective of L-arginine concentration. Similarly, in the case of Figure 4-5B, the data points were normalised against urea concentration at any given combination of GSH:GSSG ratio and L-

Table 4-1. DoE study to determine the effects of urea, arginine and GSH:GSSG redox on SVC of HBx samples in 50 mM Tris (pH 7.4)

Factor Settings[a]			SVC
Urea (M)	GSH:GSSG ratio[b]	Arginine (M)	(ml mol/g^2)
6	0	0	-0.04382
6	10	0.5	0.0704
6	10	0	-0.01353
2	10	0	0.000633
2	0	0	-0.0226
4	5	0.25	0.0222
2	10	0.5	0.030614
6	0	0.5	-0.351
2	0	0.5	0.003679
6	0	0	-0.0363
6	10	0.5	0.080824
6	10	0	-0.01255
2	10	0	0.000628
2	0	0	-0.02219
4	5	0.25	0.0228
2	10	0.5	0.038047
6	0	0.5	-0.17964
2	0	0.5	0.004479
6	0	0	-0.03996
6	10	0.5	0.0699
6	10	0	-0.01275
2	10	0	0.000622
2	0	0	-0.02427
4	5	0.25	0.0238
2	10	0.5	0.03194
6	0	0.5	-0.3459
2	0	0.5	0.00397

arginine concentration. From Figure 4-5A, it is clear that in the presence of 10:1 GSH:GSSG ratio, increasing urea concentration from 2 to 6 M had a relatively small effect on the SVC values, while in the absence of the redox couple, increasing urea concentrations alone

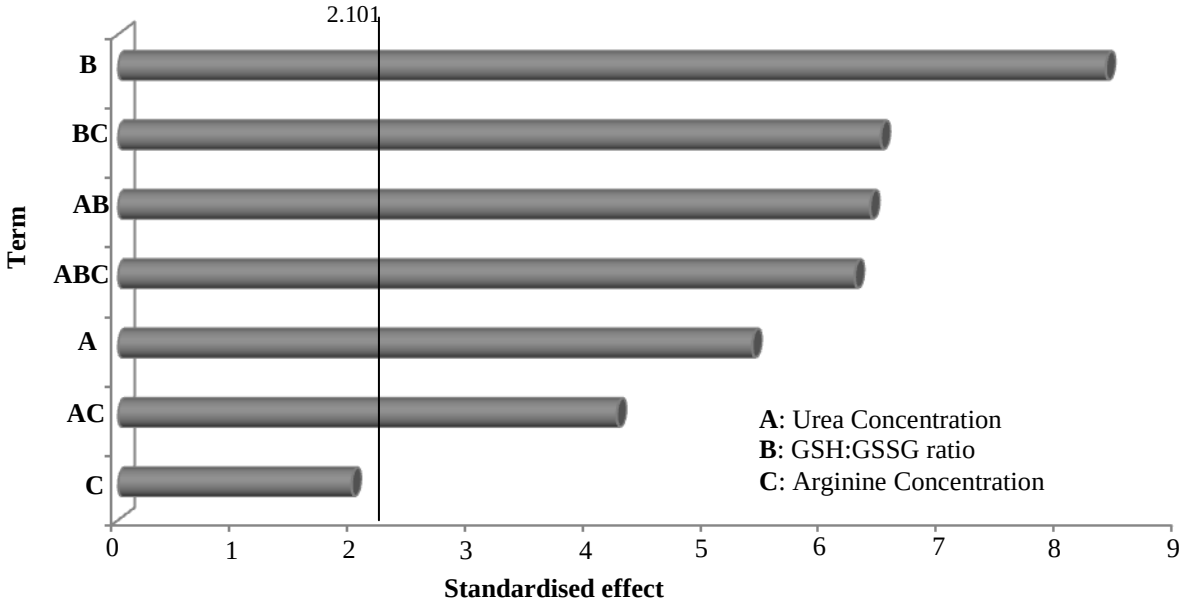

Figure 4-4. Pareto Chart for the relative effects on SVC due to varying experimental parameters. Any term with a standardized effect (a measure of how far an observation lies from its mean, in units of standard deviation) ≥ 2.101 is considered to be significant at 95% confidence level.

increased the attractive interaction between the HBx proteins and hence will not promote HBx solubility. Figure 4-5B shows that in the absence of the glutathione redox couple, the addition of L-arginine from 0 to 0.5 M also decreased SVC values. In the presence of 10:1 GSH:GSSG ratio, however, increasing L-arginine concentration increased HBx solubility, as indicated by a linearly increasing mean SVC value. Urea, at non-denaturing concentrations (i.e, 2 M), is widely known to reduce protein-protein interaction by preferentially interacting with the hydrophobic amino acid residues [132]. However, it appears that a mild reducing environment consisting of non-denaturing concentrations of urea was insufficient to completely eliminate possibilities of protein aggregation. L-arginine, on the other hand, incurred a stronger influence on HBx solubility (Figure 4-5B), as indicated by a steeper increase in SVC values with increasing concentrations compared to urea. L-arginine has been reported to reduce aggregation rates of the protein transition states and improve the equilibrium solubility of the protein molecules in solution by reducing the chemical potential of the refolding species [133-135]. The positive synergistic effect of L-arginine and the glutathione redox couple to yield a net increase in SVC values

suggests that besides covalent aggregation, other types of aggregates also contribute to HBx protein destabilization and hence insolubility.

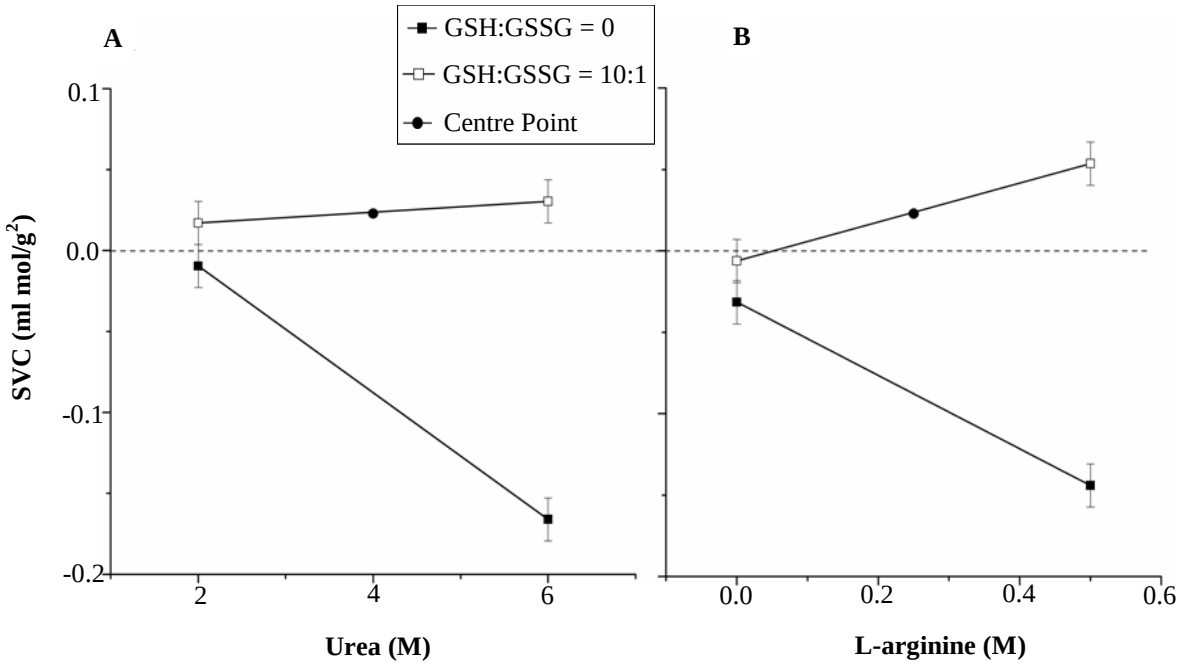

Figure 4-5. (A) Interaction plot showing the dependence of SVC on urea and GSH:GSSG ratio. (B) Interaction plot showing the dependence of SVC on L-arginine and GSH:GSSG ratio. The centre-point (●) represents the SVC for HBx in a solution containing 4 M Urea, 5:1 GSH:GSSG ratio, 0.25 M L-arginine.

Considering that both L-arginine and urea were ineffective in improving HBx solubility in the absence of the glutathione redox couple, it is likely that protein aggregates due to disulphide mispairing were present, which cannot be disrupted by L-arginine or urea. Despite the extensive use of L-arginine, urea and redox couples in protein refolding, the mechanism by which these additives interact with each other to aid refolding remains largely unreported, and not well-understood. The SVC results obtained in this study allow quantification of physicochemical environment effects on HBx solubility, thus opening the way for the rational design of a refolding-based bioprocess for HBx production.

4.3.3. HBx solubility and refolding studies

The SVC results obtained were used to guide the design of an optimised refolding buffer for *in vitro* HBx refolding. It is clear from Figures 4-5A and 4-5B that HBx solubility was enhanced by

a simultaneous increase in net reducing environment, as well as L-arginine and urea concentration. Although the use of 10:1 GSH:GSSG in the presence of 6 M urea gave the highest positive SVC value (Figure 4-5A), the highly denaturing environment incurred by 6 M urea will be unfavourable for HBx refolding. 2 M urea is instead more suitable for refolding and has been reported to be effective in solubilising hydrophobic-driven aggregates in earlier refolding studies [115, 120, 136-137], and was hence employed in the HBx refolding buffer for this study. The location of the centre points in both Figures 4-5A and 4-5B suggest that decreasing the net reducing environment from 10:1 to 5:1 while increasing urea or L-arginine concentrations to 4 M and 0.25 M, respectively, retained positive SVC values. This result indicates that a net reducing environment was dominant in governing SVC values, and remained unaffected by changing concentrations of other additives, within the concentration range tested. Based on these SVC results, it was hypothesised that HBx solubility would increase in the following order: Condition 1 (2 M urea) < Condition 2 (2 M urea +10:1 GSH:GSSG) < Condition 3 (2 M urea + 10:1 GSH:GSSG + 0.25 M L-arginine) < Condition 4 (2 M urea + 10:1 GSH:GSSG + 0.5 M L-arginine). To test this hypothesis, HBx solubility under Conditions 1 to 4 was determined by absorbance measurement at 600 nm (A_{600}). The turbidity kinetics data obtained confirmed the SVC-inferred HBx solubility behaviour, where relatively rapid precipitation of the HBx protein was observed for Conditions 1 and 2 (i.e. visible aggregates were observed at A_{600}> 0.04) compared to Conditions 3 and 4, where no turbidity was observed even after an incubation time of 6 h (Figure 4-6). Condition 3 was subsequently chosen as the optimised refolding condition which was subsequently employed for *in vitro* dilution HBx refolding throughout this thesis.

HBx was thus refolded by diluting the denatured-reduced HBx IBs into the optimised refolding buffer (2 M urea, 0.1 mM GSH, 0.01 mM GSSG and 0.25 M L-arginine, pH 7.5) to attain a final HBx refolding concentration of 0.1 mg/ml. The progress of HBx refolding over time was also monitored using SEC-HPLC (Figure 4-7). Figure 4-7A shows the SEC-HPLC chromatogram profiles of HBx refolded over time. Three protein peaks were consistently observed.

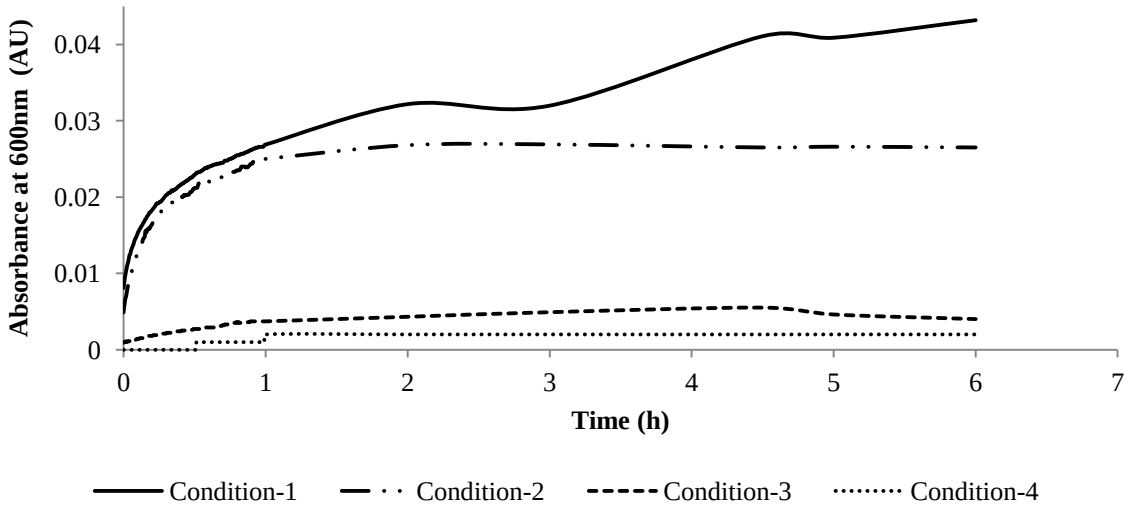

Figure 4-6. Turbidity kinetics for HBx protein fractions under varying physicochemical environment. Condition 1 = 2 M urea; Condition 2 = 2 M urea + 10:1 GSH:GSSG; Condition 3 = 2 M urea + 10:1 GSH:GSSG + 0.25 M L-arginine; Condition 4 = 2 M urea + 10:1 GSH:GSSG + 0.5 M L-arginine. All fractions contained 0.1 mg/ml HBx and were buffered at pH 7.5 with 50 mM Tris.

From the calibration curve plotted using standard proteins of known molecular masses, monomeric refolded HBx was expected to be present in Peak 2. To confirm the protein composition of each SEC-HPLC peak, the 60 h refolded HBx sample was introduced into a preparative SEC column, which was operated on an FPLC system, and protein fractions from SEC-FPLC were collected for SDS-PAGE analysis. A comparable chromatogram profile (Figure 4-7B) which contained 3 distinct peaks was obtained after SEC-FPLC fractionation. SDS-PAGE analysis of the SEC-FPLC eluate showed that all the HBx protein fractions were found in peak 2 (Figure 4-8A, lane 3) and neither peak 1 nor peak 3 contained HBx (Figure 4-8A lanes 2 and 4, respectively). Peaks 1 and 3 might thus contain other contaminants present in the diluted HBx IB solutions. To study the occurrence of disulphide-induced inter-molecular protein aggregation, the reduced and non-reduced refolding samples were analysed at different time points by SDS-PAGE analysis. A similar SDS PAGE profile was observed for both the reduced and non-reduced protein samples (Figure 4-8B), which indicates the absence of any disulphide-induced protein aggregation after HBx was refolded under Condition 3 for 60 h.

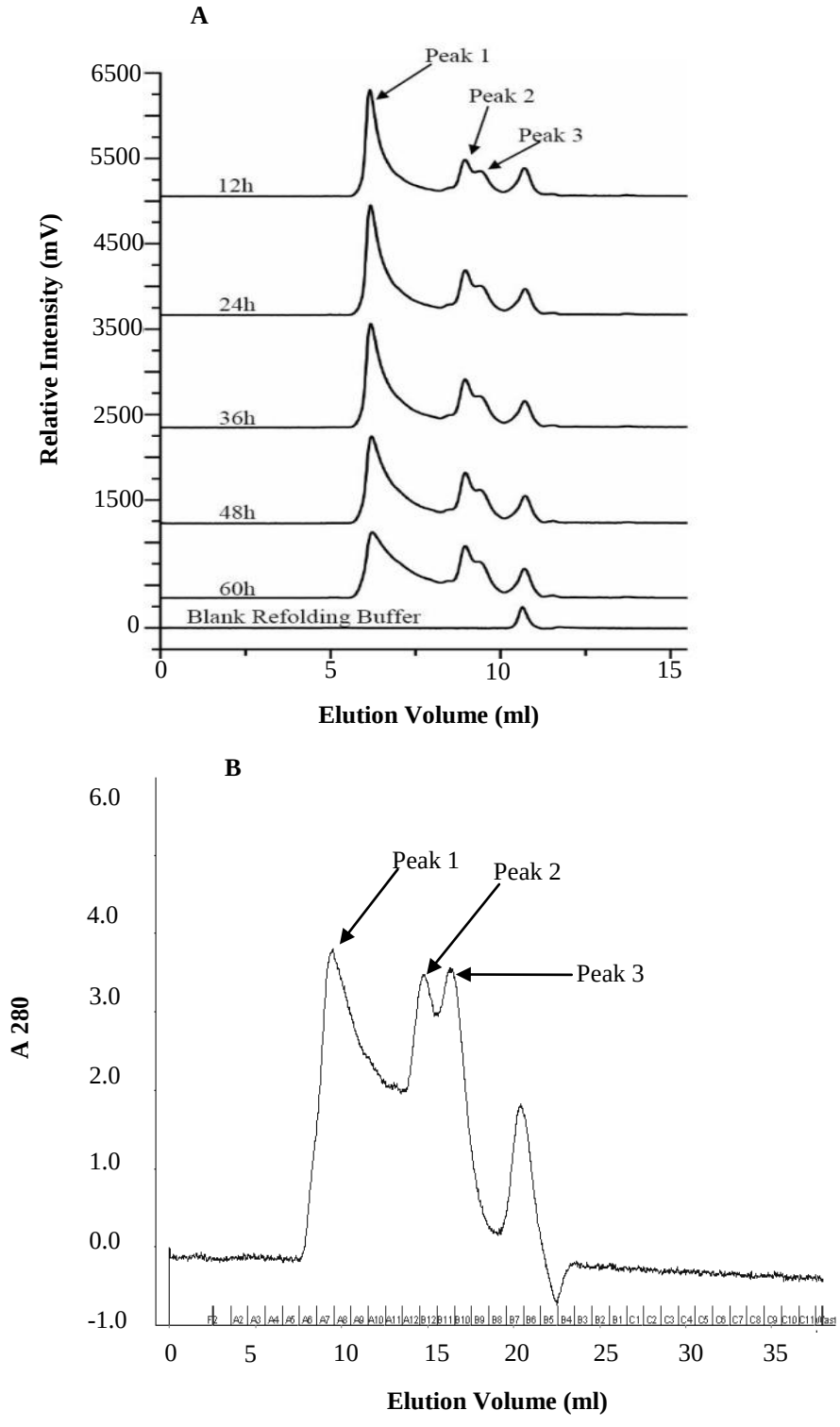

Figure 4-7. (A) SEC-HPLC chromatograms of refolded HBx at different time-points. (B) SEC-FPLC chromatogram of the 60 h refolded HBx protein sample.

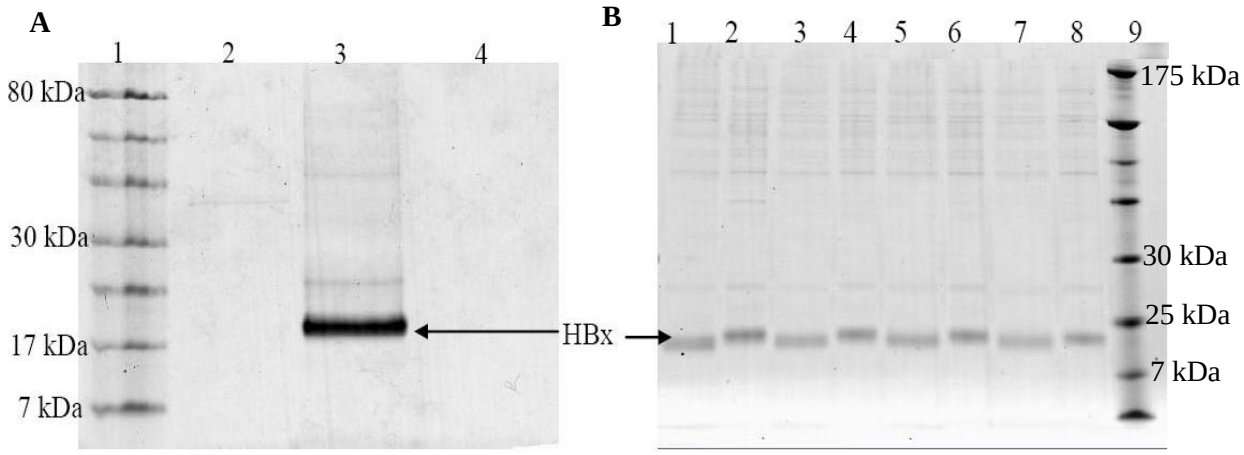

Figure 4-8. (A) SDS PAGE analysis of refolded HBx samples which were fractionated by SEC-FPLC (Figure 4-7B). Lane 1: protein marker; lane 2: sample corresponding to peak1; lane 3: sample corresponding to peak 2; lane 4: sample corresponding to peak 3. (B) SDS PAGE profile of the refolding samples analysed under reducing and non-reducing conditions. Equal volumes of the samples were loaded in each well. Lanes 1, 3, 5 and 7 represent the refolded samples incubated in the refolding buffer for 60, 48, 24 and 12 h respectively, which were analysed under non-reducing conditions. Lanes 2, 4, 6 and 8 represent the refolded samples incubated in the refolding buffer for 60, 48, 24 and 12 h respectively, which were analysed under reducing conditions. Lane 9: protein marker.

Figure 4-9A illustrates the overall bioprocess flowsheet that was developed for the production of recombinant HBx. Figure 4-9B shows the protein profile from Bioanalyzer analysis for determination of HBx soluble yield after refolding and buffer exchange steps. Table 4-2 shows the soluble HBx yields and respective protein purity obtained after each process step. HBx yields in this study were quantified by Bioanalyzer analysis. Due to the absence of any native commercial HBx standards for calibration purposes, solubilised denatured-reduced HBx protein were used as in-house standards for yield calculations during Bioanalyzer analysis.

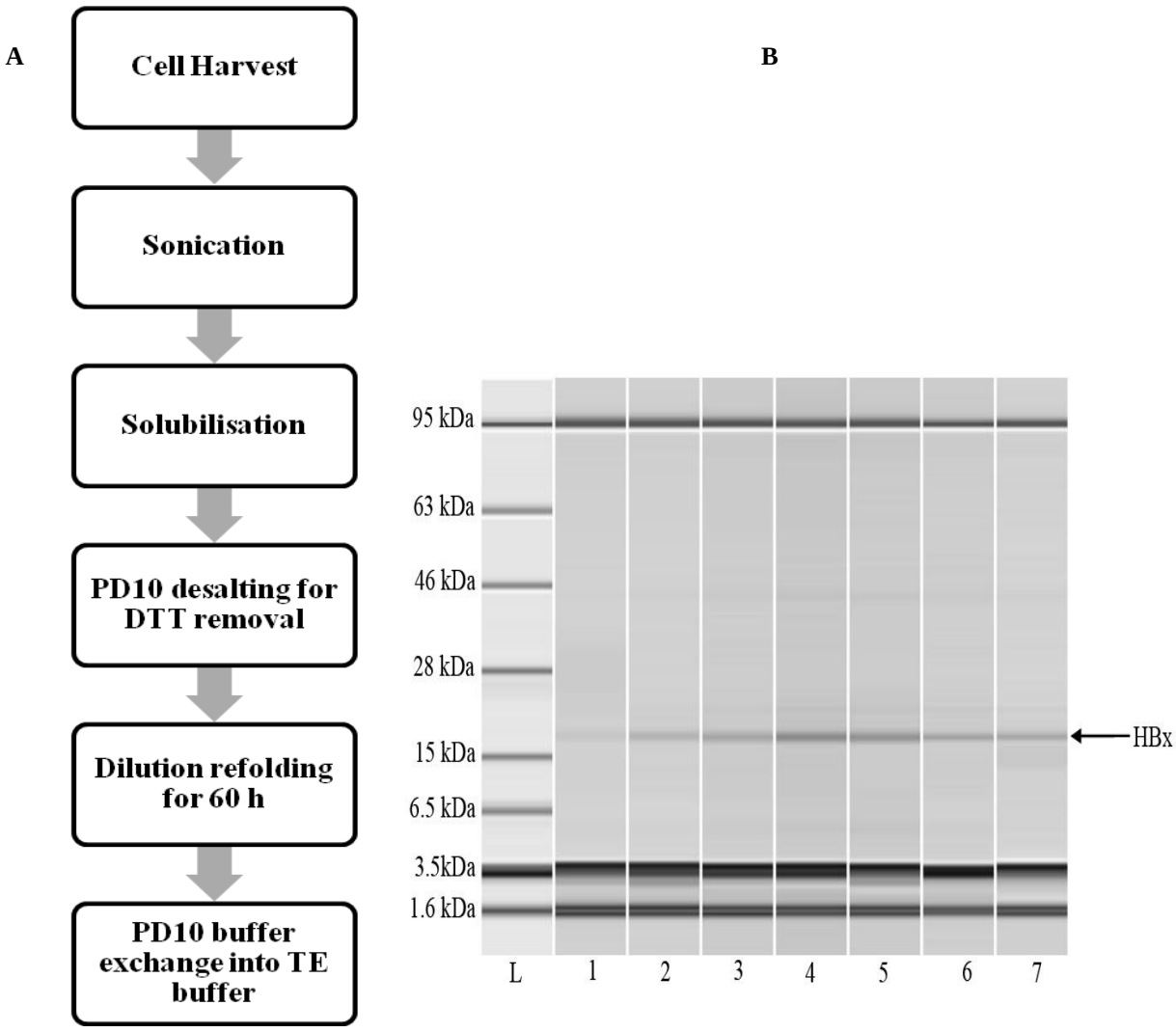

Figure 4-9. (A) Process flowsheet for HBx production *via* a dilution refolding based bioprocess. (B) Bioanalyzer analysis to determine HBx soluble yields. L: protein marker; lanes 1-5: 20, 40, 60, 80 and 100 µg/ml of solubilised denatured-reduced HBx protein, respectively; lane 6: refolded HBx sample after 60 h incubation in the refolding buffer (centrifuged and filtered before analysis); lane 7: refolded HBx in 50 mM Tris and 1 mM EDTA buffer (centrifuged and filtered before analysis).

4.3.4. Refolded HBx bioactivity determination

Refolded HBx bioactivity was determined by testing the ability of the refolded HBx to interact with the p53 tumour suppressor protein [14, 138-139]. It is believed that for survival within the host cells, the Hepatitis B virus needs to evade the host cellular apoptotic mechanism [12]. Native HBx has been widely shown to interact *in vitro* and *in vivo* with the tumour suppressor p53, which leads to inactivation of this tumour suppressor protein [14, 138-139]. The interaction

Table 4-1. HBx yield and purity obtained at each bioprocess step.

Process step	Recovered HBx (mg)	Step recovery (%)	HBx purity (%)
Harvest (from 1L culture)	49.3	NA	22.5
Sonication and solubilisation	23.2	46.9	71.8
PD10 desalting	17.1	73.9	NA*
Soluble yield after dilution refolding	16.6	96.8	86.8
Buffer exchange	16.6	100	87.1

between His-tagged HBx and p53 has also been reported in earlier studies [103, 140]. In this study, the 6-Histidine tag at the C-terminus of the HBx protein construct was utilised for binding the refolded HBx protein to charged nickel resins (Ni^{2+}-resins) for p53 interaction studies. GST and GST-p53 were then separately added to Ni^{2+}-resin bound HBx protein. After a 2 h incubation period, the resins were thoroughly washed and the bound proteins were eluted by heating in a strong denaturant and analysed by SDS-PAGE. Refolded HBx showed selective interaction with GST-p53 (Figure 4-10, lane 7) and not with GST, where majority of the GST protein was found in the supernatant fraction (Figure 4-10, lane 6). The selective interaction of HBx to p53 indicates that the refolded HBx was bioactive.

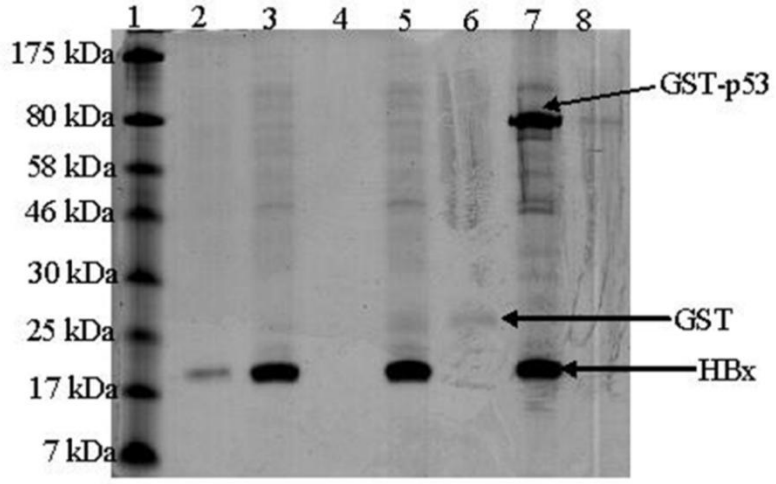

Figure 4-10. SDS-PAGE analysis to study the interaction of refolded HBx with p53 protein. Lane 1: protein marker; 2: refolded HBx in solution before binding to Ni^{2+} resins; 3: refolded HBx incubated with Ni^{2+} resins (re-suspended insoluble fraction after centrifugation); 4: supernatant of sample in lane 3 after centrifugation (7-fold diluted compared to lane 3); 5: refolded HBx incubated with Ni^{2+} resins and GST; 6: supernatant of sample in lane 5 after centrifugation (7-fold diluted compared to lane 5); 7: refolded HBx incubated with Ni^{2+} resins and GST-p53; 8: supernatant of sample in lane 7 after centrifugation (7-fold diluted compared to lane 7).

4.3.5. Structural characterisation of soluble HBx post-refolding

4.3.5.1. RP-HPLC analysis

Comparison of conformational changes between the refolded and denatured-reduced HBx protein were analysed using RP HPLC. RP HPLC is a powerful method for analysing conformations of a given protein, including the analysis of disulphide-induced protein conformational changes [141-144]. Figure 4-11 shows that denatured-reduced and DTT-treated refolded HBx obtained post-refolding eluted at the same retention time, while refolded HBx eluted at a slightly earlier retention time (Figure 4-11). The RP-HPLC chromatogram indicates that both HBx proteins had quite a comparable hydrophobicity profile.

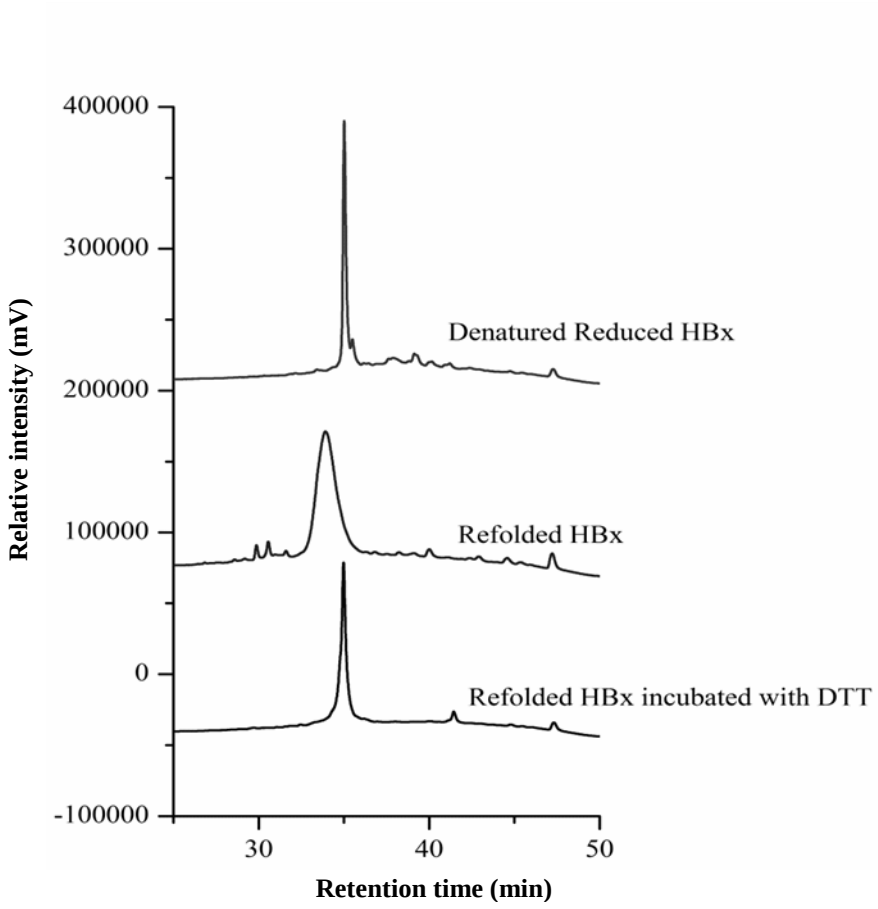

Figure 4-11. RP-HPLC chromatogram for denatured-reduced, refolded and DTT-reduced HBx after refolding

4.3.5.2. Secondary and tertiary HBx structure characterisation using CD spectroscopy

Refolded HBx was analysed by far- and near-UV CD spectroscopy to obtain the secondary and tertiary structure fingerprints of the protein, respectively. The CD spectra in the near-UV region

(250–300 nm) originates from aromatic amino acid residues and disulphide bonds, thus providing a unique fingerprint of the protein tertiary structure [145]. Figure 4-12A represents the tertiary structure fingerprint of the HBx protein measured by near-UV CD spectroscopy. The far–UV CD spectrum of the refolded HBx protein (Figure 4-12B) shows good conformity with the spectrum reported by a recent study [110]. Deconvolution of the far-UV CD spectrum of

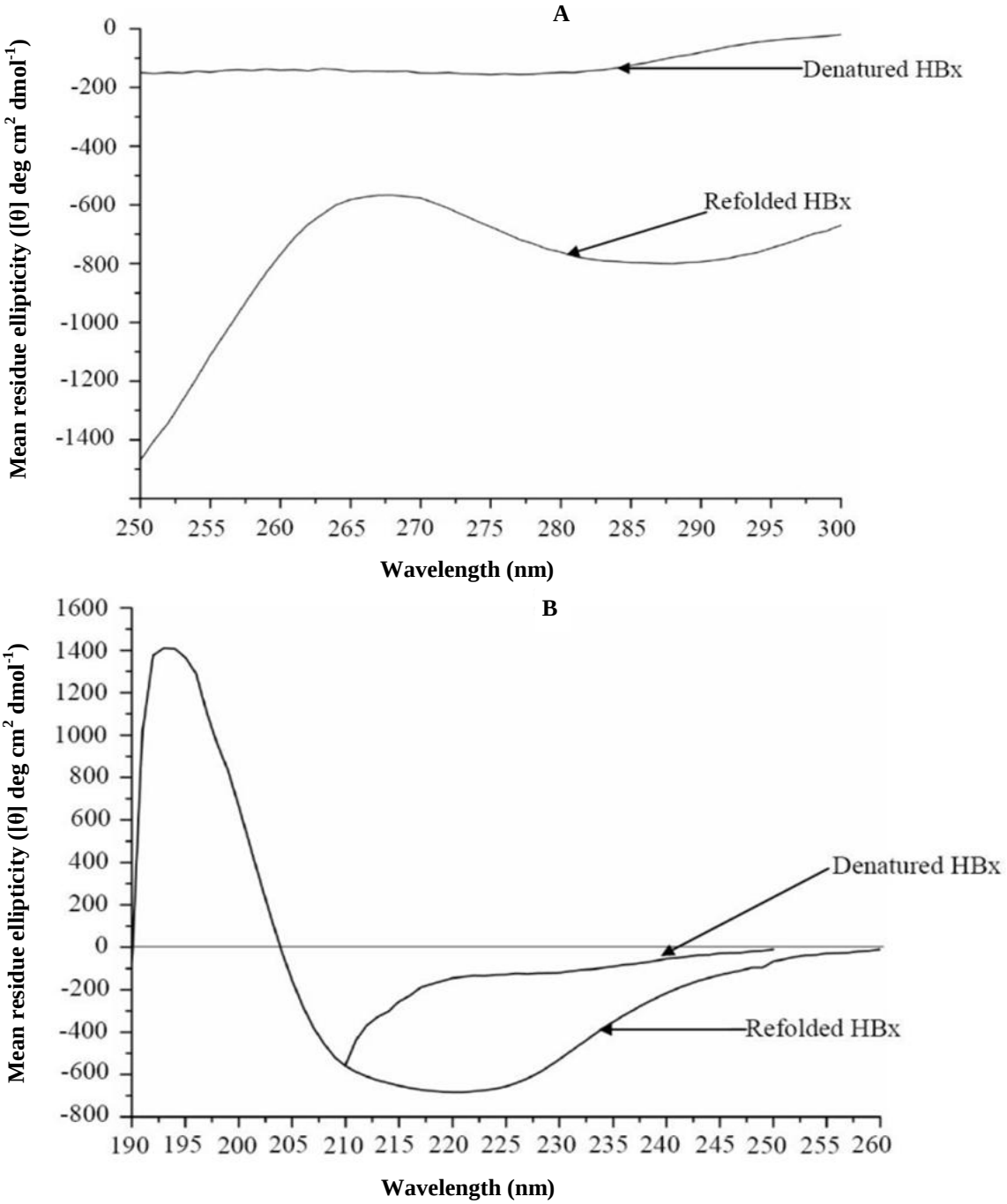

Figure 4-12. (A) Near-UV and (B) far-UV CD spectra of denatured and refolded HBx protein.

refolded HBx using the CDNN software predicted that approximately 42% of the refolded HBx secondary structure is random coil while 23% of the remaining refolded protein is alpha helical. This spectral data analysis is in close conformity with the secondary structure of the HBx predicted using the hierarchical neural network (HNN) tool (for structure prediction) which also predicts the native HBx protein to possess a random coil structure (54%) with traces of alpha helix (25%).

4.4. Conclusions

In this chapter, the combination of employing SVC measurements and a DoE platform proved to be crucial for a better understanding of the HBx inter-molecular interactions, which laid the foundations for the development of a rationally designed refolding process. Although protein concentration and physicochemical environment of the refolding buffer deserve equal considerations while designing an optimal protein refolding strategy, little progress has been made in tool development to rapidly and rationally optimise the latter. Although huge amounts of information on the impact of different additives on the refolding reactions are available from earlier studies, from which different refolding buffer systems can be designed, the lack of quantitative data on aggregation kinetics to complement solubility data which is often used to infer protein foldability in existing refolding screening methods increases the difficulty in design and scale-up of refolding processes. Accurate and systematic measurements of thermodynamic parameters like SVC at statistically determined design points can breach this gap, thus expediting process development strategies for novel protein candidates like HBx, as demonstrated by the results of this chapter.

An overall soluble HBx yield of 34% at a protein purity of 87% was obtained from the refolding-based bioprocess developed in this chapter. The bioactivity of the refolded HBx protein was demonstrated by its selective interaction with the tumour suppressor protein, p53 “[88]”. Due to the lack of native standards however, only HBx soluble yield could be quantified in this chapter while quantification of refolding or bioactive yield was not possible. Since quantitative refolding yield is a pre-requisite for subsequent design and bioprocess scale-up studies, the following chapter (i.e. Chapter 5) reports efforts to develop a quantitative analytical platform for determining HBx refolding (or bioactive) yield. The selective interaction of HBx with the p53

protein, as described in Section 4.3.4, is exploited in developing the ELISA platform, which will be presented in detail in Chapter 5.

CHAPTER 5: Development of an ELISA Based Analytical Platform for Determination of HBx Refolding Yields

Abstract

This chapter reports the development of a novel ELISA platform to quantitate HBx protein refolding yields, which is critical for rational design and scale up of the HBx bioprocess. HBx refolding yields were measured by determining the amount of HBx bound to immobilized GST-p53 on a GSH-functionalised maleimide surface. This ELISA platform allows refolding yields to be distinguished from soluble yields, where the latter was determined by measuring total HBx protein bound to a maleimide surface under reducing conditions. The accuracy of the developed ELISA platform was further validated using standard statistical techniques and data trend analysis. The ELISA platform is amenable to scale-up, and will facilitate bioprocess development for large scale production of HBx for structural and clinical studies, thereby expediting the development of HBx-based therapy for liver cancer.

5.1. Introduction

In Chapter 4, a dilution refolding based HBx bioprocess was developed using an SVC optimized refolding buffer. However, further process improvement and scale-up efforts are hindered due to the lack of suitable bioanalytical platforms to quantitatively assess HBx refolding yields. Although *in vitro* activity of recombinantly produced HBx protein has been reported in some of the previous studies [110, 146], no quantitative method for HBx refolding yield determination has been reported to date. Confronted by this roadblock, this chapter aims to develop an ELISA platform for quantitative refolding yield determination of HBx (possessing a 6-His tag). The design rationale of the ELISA analytical platform is based on the well-characterized HBx cellular activities, which are reported to be mediated through various protein-protein interactions [147]. Amongst the different cellular tumor suppressor factors modulated by HBx, its interaction with the p53 protein is one of the most well-established [14, 103, 138-140, 147]. The ELISA platform was therefore designed to correlate HBx-p53 interaction with the HBx protein's biological activity and hence refolding yield. The basic strategy for the ELISA system would include immobilisation of GST-p53 on a suitably functionalised micro-plate surface such that it can bind with bioactive HBx (possessing a 6-His tag). Thereafter, using the 6His-tag as a common detection platform, the amount of immobilised bioactive HBx can be detected from a calibration curve generated using commercially available standard 6His-tagged proteins (such as 6His-GST) which can also be immobilised on a similar surface.

Each step of the developed ELISA platform can be analysed through well established statistical methods to identify the source of variation in the observed data such that it can be concluded whether the measurement methodology or the sample variation leads to variability in the observed signals. In this study, it would be reasonable to assume that the designed measurement system (i.e. the HBx ELISA platform) is valid if the data obtained from any given sample produces a consistent result (i.e., reproducible results are obtained using the strategy for any given sample), indicating its precision. Additionally, the accuracy of the measurement system can be gauged through the conformity of the ELISA derived HBx yield data to expected refolding trends from the results of earlier studies.

5.2. Materials and methods

5.2.1. Construction of a calibration curve to quantify bound HBx

A calibration curve to quantify the amount of bound HBx on the ELISA plate was constructed using commercially available 6His-GST protein (Abcam). 50 µl of 6His-GST protein solutions at varying known concentrations (0 to 20 µg/ml) were covalently bound to maleimide-functionalized 96-well microplates (Sulfhydryl Bind™, Corning). As per the supplier's instructions, the 6His-GST protein solutions were prepared in Tris-EDTA (TE) buffer (50 mM Tris and 1 mM EDTA, pH 6.5) containing 10^{-3} mM DTT, before being added to the plates, where the samples were incubated for 1 h under moderate shaking conditions at room temperature. The wells containing the proteins were then thoroughly washed with TE buffer and incubated with a 5% non-fat milk solution for 1 h at 37°C for blocking. The excess milk was washed off and the wells were incubated with 50 µl of horse radish peroxidase (HRP)-conjugated 6His primary antibody (Abcam, diluted 5000 fold in 5% non-fat milk solution) for 1.5 hour at 37°C. The amounts of 6His-GST protein bound to the plate surfaces were then quantified by adding equal volumes of TMB substrate developing solution (Millipore) and measuring sample absorbance at 650 nm in a microplate reader following 15 min incubation.

5.2.2. Determination of HBx refolding yield

HBx refolding yield was determined based on the interaction of correctly refolded HBx with p53. 50 µl of 0.1 M GSH solution (in TE buffer) was incubated on the maleimide-functionalised plate surface for 1 h at room temperature under moderate shaking conditions. After washing away the unbound GSH molecules, the wells were blocked for at least 30 min by adding 100 µl of 0.2% non-fat milk as per the plate supplier's protocol. The milk was decanted and 50 µl of 10 µg/ml GST-p53 (Millipore) solution, in 50 mM Tris buffer, was added to the wells for immobilisation to the GSH-functionalised surface. After 2 h of incubation at room temperature, the unbound GST-p53 proteins were washed away. The wells were then blocked for 1 h by incubation with 5% milk at 37°C and after washing away the excess milk, the 'GST-p53'-immobilized surfaces were incubated with IMAC purified 6His-HBx protein (refolded as described in Section 4.2.8, Chapter 4) for 1.5 h at room temperature.

To prepare the refolded HBx samples for loading into the ELISA wells, the refolded HBx samples which had been incubated at different time intervals (from 1 to 5 days) in the refolding buffer (2 M urea, 0.25 M L-arginine, 0.1 mM GSH and 0.01 mM GSSG in 50 mM Tris, pH 7.4), were desalted into 50 mM Tris buffer (pH 7.4) and diluted 1000 times. A fixed HBx refolding concentration of 0.1 mg/ml was used throughout this study. The unbound proteins were washed away and the immobilized HBx protein surfaces were incubated with the HRP-conjugated 6-His primary antibody (diluted 5000 times in 5% milk) for 1.5 h at 37°C. After washing away the unbound proteins from the plate surface, the immobilized antibodies were detected by adding 100 µl of the TMB substrate. The substrate was incubated for 15 min at room temperature under stirring conditions before measuring sample absorbance at 650 nm in a micro-plate reader. Using this strategy, HBx refolding yields were determined from Equation 5-1 below.

$$\text{Refolding Yield}(\%) = \frac{[\text{Mass of refolded HBx bound to GST-p53}] \times 100\%}{[\text{Mass of denatured reduced HBx}]} \qquad \text{(Equation 5-1)}$$

To determine the total denatured-reduced HBx concentration (Equation 5-1), solubilised HBx protein (in 8 M urea and 10 mM DTT) was first diluted in 8 M urea to the desired refolding concentration of 0.1 mg/ml, and then diluted 1000-fold in TE buffer (pH 7.4) for loading onto the ELISA plate. Denatured-reduced HBx can covalently bind to the maleimide surface under reducing conditions by means of interaction of the reduced cysteine present in the HBx protein and the maleimide surface. The amount of denatured-reduced HBx which bound on the plate was quantified using the 6His-tag antibody detection strategy as described in Section 5.2.1. The binding specificity of refolded HBx to GST-p53 was tested by measuring the ability of the refolded HBx to bind to BSA, which was covalently immobilized on the maleimide surface. As described in Section 5.2.1, BSA was dissolved in the TE buffer and surface-immobilised under mild reducing conditions at 1 mg/ml.

5.2.3. Determination of HBx soluble yields

The ELISA platform design was also aimed to distinguish HBx refolding yields from HBx soluble yields. HBx soluble yields were determined using Equation 5-2.

$$\text{Soluble Yield}(\%) = \frac{[\text{Mass of HBx bound on the maleimide surface}] \times 100\%}{[\text{Mass of total denatured reduced HBx}]} \qquad \text{(Equation 5-2)}$$

To determine the mass of HBx bound on the maleimide surface for determination of HBx soluble yields, the refolded HBx samples which were diluted 1000-fold in TE buffer was incubated in 10^{-3} mM DTT to enable covalent binding between the free cysteine residues of HBx and the maleimide surface. The amount of HBx bound was determined by using the 6-His tag antibody detection strategy (as described in Sections 5.2.1 and 5.2.2).

5.2.4. Analytical methods

To determine whether the standard proteins (i.e. 6His-GST and GST-p53) have bound to the ELISA plates surface, 6His-GST and the GST-p53 proteins were added to the maleimide and the GSH-functionalized maleimide surfaces respectively, as per the methods described in Sections 5.2.1 and 5.2.2. Samples from the wells were then injected into a RP HPLC system for analysis as per the method described in Section 4.2.11, Chapter 4. The molecular weights of the standard proteins (6His-GST and GST-p53) were further confirmed by SDS PAGE.

Soluble yield data trend as obtained from the ELISA strategy was further validated independently using a Bioanalyzer (protocol detailed in Section 4.2.13, Chapter 4).

5.2.5. Measurement System Analysis

The two most important criteria for an ideal measurement system are (i) precision and (ii) accuracy [148]. This section explains the methods by which "precision" and "accuracy" of the developed ELISA platform was assessed.

5.2.5.1. Precision

The total variation within the observed response from any measurement system, can be considered as the sum of variation due to samples and the measurement system (Equation 5-3)

$$S^2_{\text{Total}} = S^2_{\text{sample}} + S^2_{\text{measurement}} \qquad (5\text{-}3)$$

where S^2 represents the data variation under consideration. Therefore, from Equation 5-3, it can be implied that a measurement system is "precise" if the $S^2_{\text{measurement}}$ forms only a miniscule

fraction (<10%) of the total observed variation (S^2_{Total}). Only in such a situation can changes in the response from the measurement system "accurately" reflect the changes within the tested samples (i.e., $S^2_{Total} \approx S^2_{sample}$).

For generation of the calibration curve, signals (i.e. absorbance measurements at 650 nm) obtained from the immobilized 6His-GST using the ELISA platform were directly used for measurement system analysis using the Gage Repeatability and Reproducibility program in the Minitab 15 software. The test consisted of five samples (containing 6His-GST samples within the concentration range of 1.25 – 15 µg/ml) with three repeats each. For determining the precision of the soluble and refolding yields data, the yield values obtained using Equations 5-1 and 5-2 respectively, were used for the analysis, where a total of seven different samples were considered for each case along with three corresponding repeats. Data variation for all the test cases was analysed using the Analysis of Variance (ANOVA) method.

5.2.5.2. Accuracy determination

The measurement system "accuracy" can be qualitatively determined through the conformity of the observed response from the ELISA strategy to expected trends. For example, the accuracy of the calibration curve constructed using the 6His-GST protein, can be assessed from the correlation coefficient (R^2) of the protein concentration versus absorbance at 650 nm (A650) plot.

5.3. Results and discussion

5.3.1. Immobilisation of proteins on the maleimide surface of the ELISA plates

Maleimide-functionalised surfaces were chosen for use in this study because of their interaction specificity for sulphydryl groups within the pH range of 6.5 to 7.5, which allows binding of cysteine-containing proteins like GST and HBx. Maleimides (or maleic acid imides) (Figure 5-1, I) are obtained when maleic anhydride is reacted with ammonia or an amine derivative. Thereafter, as shown in Figure 5-1, one of the double bonded carbon atoms in molecule I (the maleimide) can undergo nucleophillic attacks by sulphydryl groups to form stable thioether bonds (Figure 5-1, II). Based on this strategy, maleimide-functionalised surfaces are widely used to specifically immobilize proteins or peptides containing reduced cysteine residues [149]. In this

study, the use of a maleimide surface will facilitate covalent binding of the 6His-GST and HBx molecules which both contain cysteine residues, under mild reducing conditions. Hence based on this strategy, the developed ELISA methodology can be schematically represented by Figure 5-2 (A and C), showing how the 6His-GST and HBx can be covalently bound to the maleimide surface under mild reducing conditions.

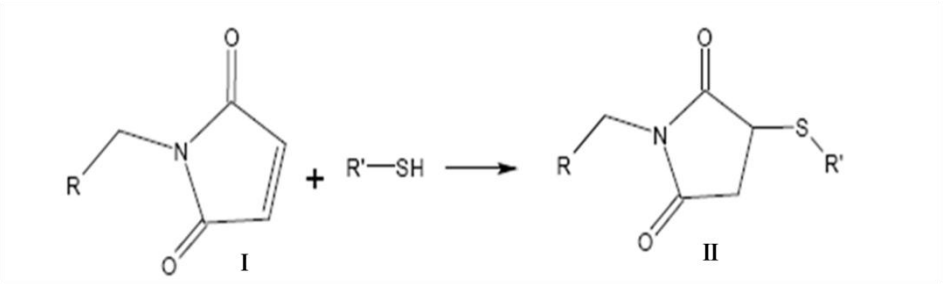

Figure 5-1 Binding mechanism of sulphydryl groups on maleimide surfaces.

To determine HBx bioactivity and hence refolding yield, GSH was first immobilized on the maleimide-functionalized microplate *via* covalent binding of the free cysteine residue within GSH and the maleimide surface (Figure 5-2B). Incubation of the GST-p53 protein on this surface would allow binding of the GST protein to GSH, and thus avail the p53 molecule to bind with biologically active HBx proteins present in the refolding samples.

To determine the amounts of 6His-GST and GST-p53 binding to the ELISA microplate surface, the proteins were incubated in the maleimide and the GSH-functionalised maleimide surfaces of the microplates respectively, as per the methods described in Sections 5.2.1 and 5.2.2. After incubating the samples on the respective surfaces, the supernatant of the sample recovered from the ELISA plates (which would contain unbound proteins, if any) was injected into a RP HPLC column for analysis. Peak integration of the RP HPLC chromatograms of the protein samples before and after binding on the plates indicated that all the 6His-GST and GST-p53 proteins that were loaded into the ELISA plates bound to the plate surfaces (Figure 5-3A and B). To further confirm the identity of the eluted proteins, the molecular weights of the corresponding protein samples were verified using SDS PAGE analysis (Figure 5-3C).

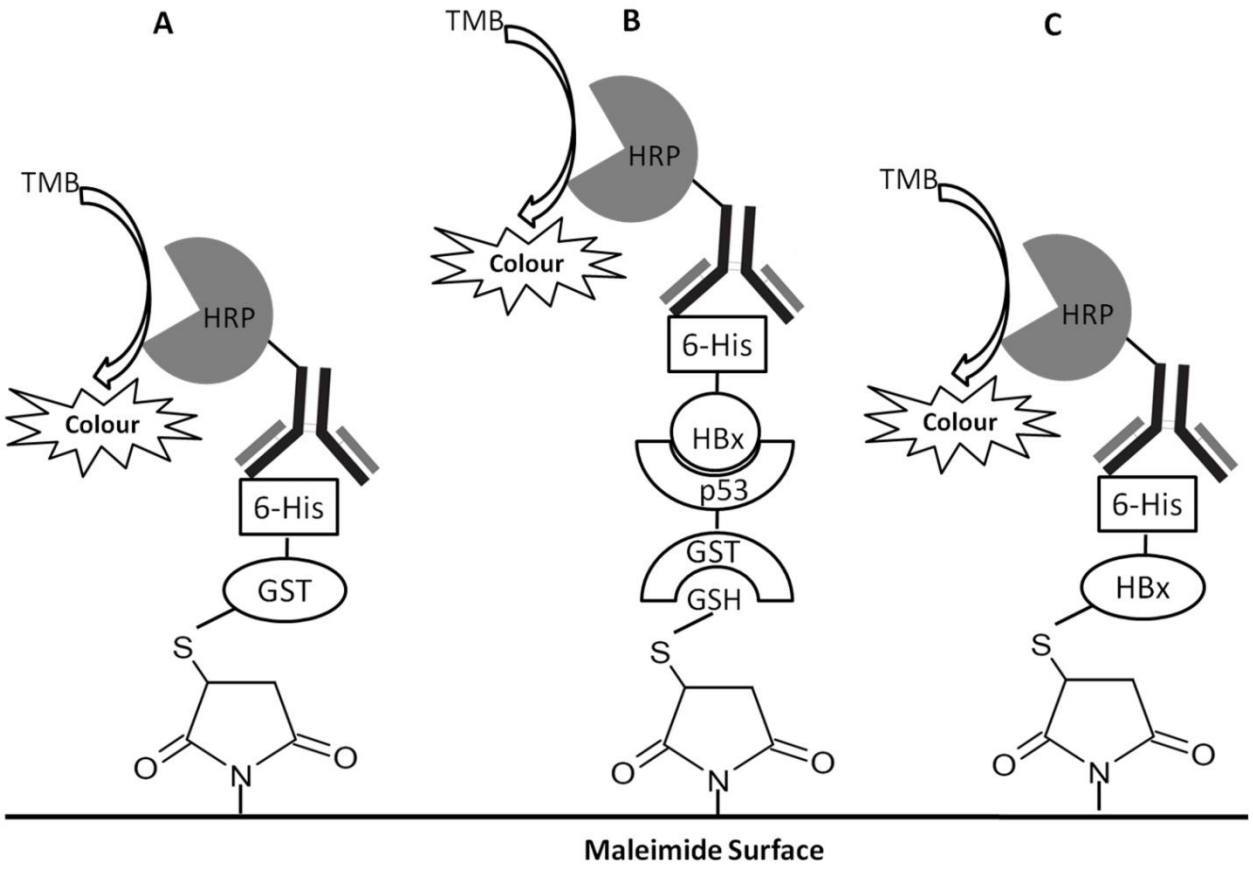

Figure-5-2 Schematic of the developed ELISA platform to quantify HBx refolding and soluble yields. The 6-His tags were used as a common platform for detection of bound proteins. (A) Strategy for the construction of calibration curve: 6His-GST protein at different known concentrations were covalently attached to the maleimide surface and incubated with the HRP-conjugated 6His primary antibody. Upon addition of the TMB substrate, the intensity of the colour development was measured by recording absorbance at 650 nm, where increasing absorbance correlates with increasing 6His primary antibody concentration and 6His proteins bound on the plate. (B) Strategy to determine HBx refolding yields: GST-p53 proteins were first bound to a GSH-functionalized maleimide surface, to which the refolded HBx sample was added. The amount of HBx protein bound to GST-p53 was determined by incubation with HRP-conjugated 6His primary antibody. (C) Soluble HBx yield was determined by covalently binding the HBx proteins to the maleimide surface (under mild reducing conditions), followed by incubation with the HRP-conjugated 6His primary antibody for subsequent detection and quantitation of bound HBx. **N.B.** The schematic diagram may not reflect the true binding stoichiometry between the different proteins.

5.3.2. Generation of calibration curve for the ELISA system using 6His-GST protein

A major hurdle for the development of a quantitative analytical platform for a protein like HBx is the absence of native standards needed for generating the calibration curves to quantify the amount of HBx bound on the ELISA plates. This roadblock can be overcome by using the 6His tag as a common platform to detect the amount of HBx protein bound on a given surface, i.e., using a commercially available 6His-tagged protein like the 6His-GST that can specifically bind on the microplate surface under similar conditions. The results in Figure 5-3 indicate negligible

protein loss due to non-binding. Therefore, the absorbance units (at 650 nm) corresponding to known concentrations of 6His-GST can be used to correlate the amount of bound protein (nM), which is employed to generate the calibration curve to quantify the amounts of immobilized HBx

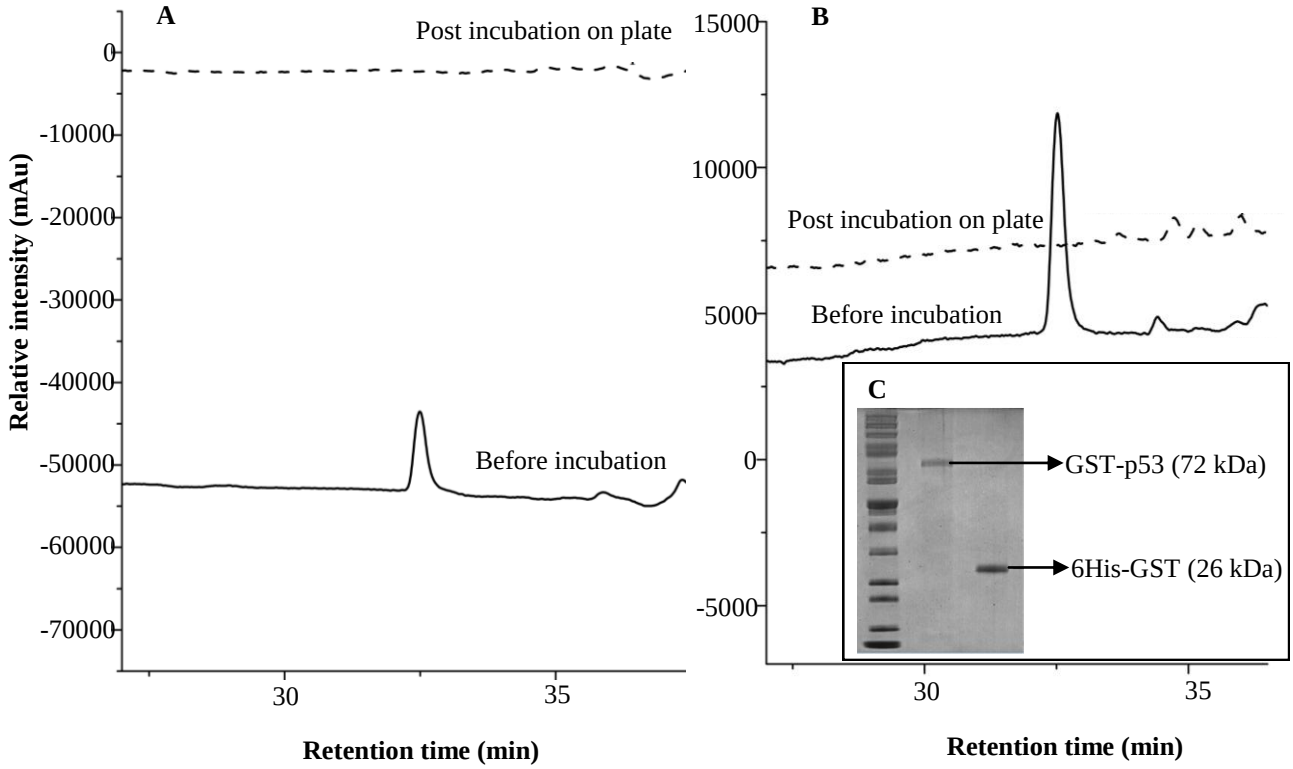

Figure 5-3. RP-HPLC chromatograms of (A) pre-incubation and post-incubation of 6His-GST protein samples on the maleimide-functionalized surface; (B) pre-incubation and post-incubation GST-p53 samples on the GSH-functionalized maleimide surface. Both samples in (A) and (B) were diluted to same protein concentrations; (C) (Insert) SDS PAGE analysis of 6His-GST and GST-p53 pre-incubation samples used in RP-HPLC analysis in (A) and (B).

on the ELISA plate. Figure 5-4A shows that majority of the variation (>99%) in the observed ELISA response for calibration curve is due to variation within samples indicating the precision of the system (i.e. $S^2_{Total} \approx S^2_{Sample}$). The obtained calibration curve shown in Figure 5-4B also possessed a high correlation coefficient ($R^2 = 0.963$), which indicates a strong correlation of the observed signal with the amount of immobilized protein. Based on these results, it can be concluded that the concentration of the immobilized 6His-tagged proteins can be measured accurately and precisely using the calibration curve presented in Figure 5-4B.

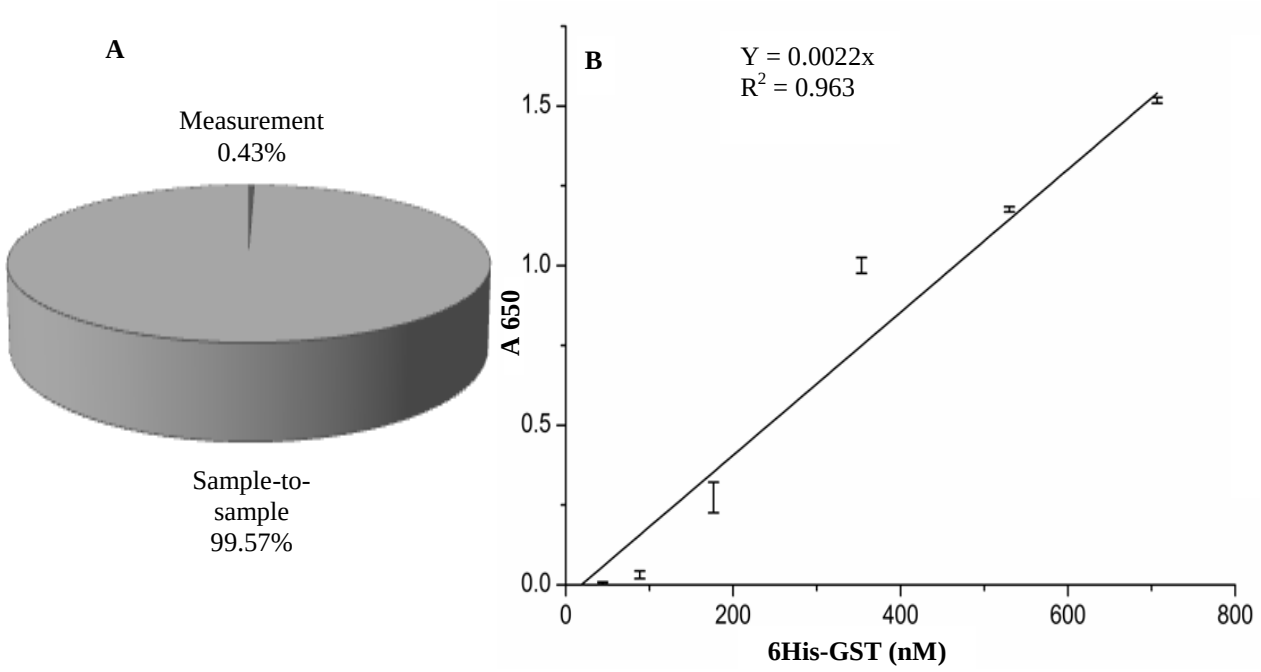

Figure 5-4. (A) The percentage contributions to the total variation by the measurement system and the samples for the calibration curve generated; (B) Calibration curve obtained by covalent immobilization of 6His-GST on the maleimide-functionalized surface of the ELISA plate.

5.3.3. Determination of HBx soluble and refolding yields

Upon establishment of a calibration curve to quantify the amount of HBx bound to GST-p53, the ELISA platform was subsequently used to determine the soluble and refolding yields of the HBx protein. Denatured-reduced HBx samples were incubated over different incubation periods in the refolding buffer, and introduced to the ELISA platform for soluble and refolding yield determination. From Figures 5-5A and B, it is clear that the measurement system contributes to a very small proportion of the total variation (<4%), which indicates the relative accuracy of the results obtained. The trend of increasing soluble yield with refolding incubation time (Figure 5-5C), agrees well with the HBx solubility behaviour with increasing incubation time reported in Section 4.3.3, Chapter 4 and is further validated using an independent method of analysis, as shown in Figure 5-6. The decrease in refolding yield after 3 days is likely to be due to unproductive interaction between misfolded proteins and the refolded protein and/or oxidation of the GSH [88], which is expected considering that the SVC results presented in Section 4.3.2, Chapter 4 indicated that the stability of the HBx protein largely depended on the successful

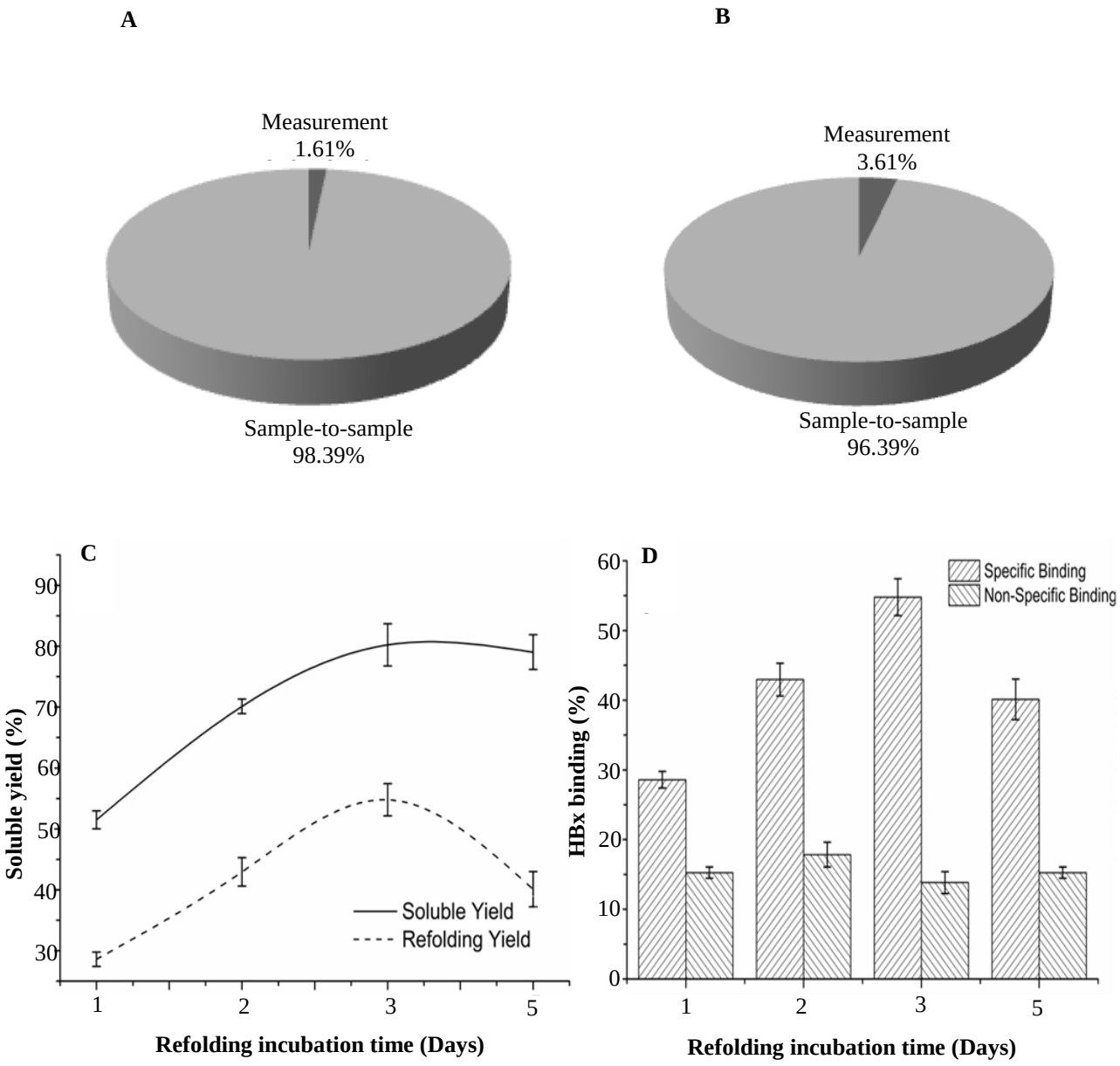

Figure 5-5 (A) Percentage contributions to the total variation by the measurement system and the samples obtained for the HBx soluble yield determination step; (B) percentage contributions to the total variation by the measurement system and the samples obtained for the HBx refolding yield determination step; (C) HBx soluble and refolding yields determined using the ELISA platform; (D) comparison of the specificity of HBx binding to GST-p53 and BSA. All data represented within the figure shows the mean ± standard deviation of the observed values.

maintenance of a net reducing environment. Figure 5-5D compares the amount of HBx protein bound to GST-p53 with non-specific binding of the HBx protein to BSA. Non-specific binding of HBx to BSA (surface-immobilized at 1mg/ml) for each point, was found to be significantly smaller ($p<0.01$, data tested with Student's T-test) compared to HBx binding to GST-p53

(surface-immobilized at 10 µg/ml). Based on these results, it is clear that the developed measurement system produced precise data and the data trends are in good agreement with the HBx solubility behavior (as a function of time), thus indicating its accuracy.

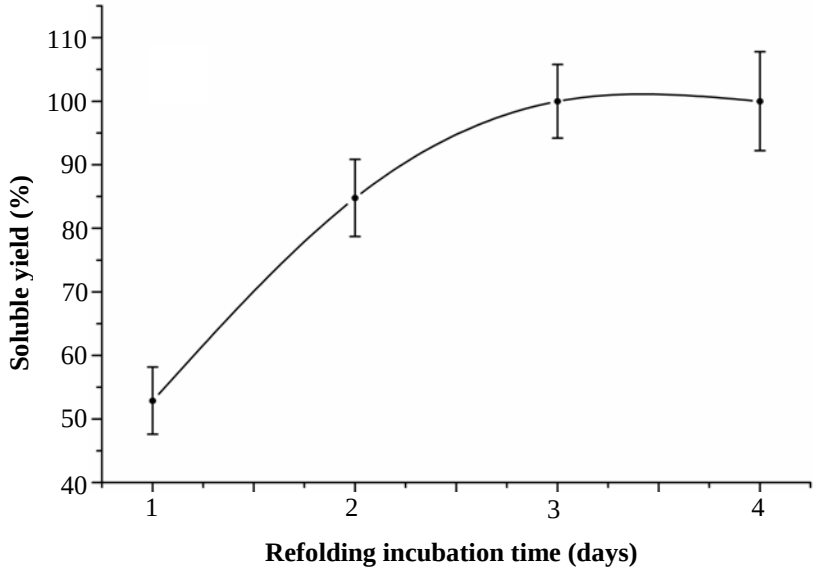

Figure 5-6. Soluble HBx yield determined by Bioanalyzer analysis.

5.4. Conclusion

This study reports the successful development of a novel ELISA analytical platform to determine HBx refolding and soluble yields based on the specific interaction of HBx with the tumour suppressor protein, p53 [103, 140]. Since there are no high purity HBx standards that are commercially available to date, this ELISA platform will be instrumental in facilitating future design, optimisation and scale-up studies of HBx bioprocesses. The maleimide attachment chemistry [149] was used to create a GSH surface that was needed to immobilize the p53 protein molecules so that it remains available for binding with the HBx. The success of the strategy is well-reflected not only by the reproducibility of the values at each data point (coefficient of variation < 6% for all the measured soluble and refolded HBx concentration measurements), but also the data trends which agreed well with the previous findings on HBx solubility studies based on SVC measurements, as reported in Chapter 4. To the best of my knowledge, this is the first report on the development of an ELISA based analytical platform for evaluation of recombinant HBx bioactivity, where the distinction between the soluble and the refolding yields of the HBx

protein could be established. Such an analytical platform will be critical not only for evaluating the quality of recombinantly produced HBx, but also allows quantitative refolding yield determination to facilitate HBx bioprocess design, development and scale-up studies “[20]”, as reported in Chapter 6.

CHAPTER 6: High Productivity Chromatography Refolding Process for HBx Guided by Statistical DoE Studies

Abstract

This chapter reports how the rationally designed refolding buffer system and HBx bio-analytical platform established in chapters 4 and 5 respectively, can synergistically guide the development of a 'chromatography refolding'-based bioprocess for HBx using immobilised metal affinity chromatography (IMAC). In line with the principles of QbD, a Statistical DoE methodology was used to design an optimum IMAC refolding step which successfully delivered bioactive HBx at a productivity of 0.22 mg/ml/h and refolding yield of 54% (at 10 mg/ml refolding concentration). The use of a DoE-directed approach to design and optimise the IMAC refolding process facilitated the rational development of an intensified IMAC-based bioprocessing platform which enabled the production of bioactive HBx at ~ 350 mg/L at >95% purity in a highly streamlined process flowsheet. Compared with an intensified dilution refolding process, the optimised IMAC refolding process achieved HBx refolding yield and productivity which was higher by 250% and 400%, respectively. The DoE-based screening methodology adopted for the study also provided important insights into the effects of different bioprocess parameters on the HBx productivity and quality. Such a bioprocess design approach can play a pivotal role in the development of intensified processes for other novel proteins, and hence expedite resolution of scale-up and validation challenges faced by the biopharmaceutical industry today.

6.1. Introduction

Having established the physicochemical environment essential for HBx refolding (in Chapter 4) and an ELISA platform to quantify HBx refolding yield (in Chapter 5), the next objective of this project was to establish a bioprocess that is robust enough to ensure efficient and scalable supply of the HBx protein for subsequent structural characterisation and drug designing studies. In this chapter, the C-terminus 6His-tag of the HBx protein was exploited to design a scalable bioprocess based on IMAC refolding. The aim of developing an IMAC refolding strategy was to allow simultaneous purification and refolding of 6His-tagged proteins, thus minimising the number of unit operations involved to improve productivity and process costs [66, 150]. Instead of adopting a trial and error approach for optimising the IMAC refolding process, a Statistical DoE methodology was used to extract maximum information about the impact of different refolding process parameters on the quality of the HBx product. IMAC on-column protein refolding processes reported in the literature so far [151-152] lack this systematic DoE approach for obtaining optimised processes, which is crucial for identifying the interaction effects of the different process parameters on the critical to quality attributes (CQA) of the product. Besides saving time and resources, DoE approaches for designing and optimising the IMAC refolding process align with the requirements by the drug regulatory authorities to implement Quality by Design (QbD) principles in biopharmaceutical manufacturing [153-154]. In the DoE-based optimisation study of IMAC refolding, important process parameters affecting the product CQA were first determined after thorough consideration of each step of the probable process design. Selected parameters which might significantly affect the on-column refolding process were then chosen as the independent variables within the DoE strategy.

6.2. Methods and materials

All chemicals were obtained from Sigma-Aldrich (Singapore), unless stated otherwise.

6.2.1. HBx protein expression and solubilisation

Expression and recovery of the HBx IBs were performed using the same methods as those described in Section 4.2.3, Chapter 4 [88]. The prepared HBx IBs were solubilised in a denaturing-reducing buffer (8 M urea, 1 M NaCl, 10 mM DTT, 20 mM Imidazole, 50 mM Tris, pH 7.4) at 0.8 to 1.2 mg/ml protein concentration for 3 h. The denaturing-reducing buffer used

for solubilising the HBx protein was also used for pre-equilibrating the IMAC columns and would be referred to as IMAC binding buffer in the following sections.

6.2.2. IMAC refolding of the HBx protein

Refolding studies for the HBx protein were performed on IMAC columns using the AKTA Explorer FPLC system (GE Healthcare). All IMAC refolding experiments were carried out using 1 ml Ni^{2+}Sepharose HisTrap HP columns (GE Healthcare) at 1.0 ml/min flow-rate. 5-10 mg of solubilised denatured-reduced HBx IBs were loaded on the IMAC columns which were pre-equilibrated with IMAC binding buffer. After protein loading, the column mobile phase was changed from IMAC binding buffer to refolding buffer (2 M urea, 0.25 M L-arginine, 0.1 mM GSH and 0.01 mM GSSG in 50 mM Tris, pH 7.4) over 0-5 column volumes (CVs), followed by equilibration with 15 CVs of refolding buffer. The IMAC columns containing the bound proteins were then incubated at 4°C for different time periods (1 to 3 days) and then eluted in a single step with the refolding elution buffer (2 M urea, 0.25 M L-arginine, 0.1 mM GSH and 0.01 mM GSSG, 1 M Imidazole in 50 mM Tris, pH 7.4) for refolding yield determination. The column was then washed with 10 CVs of IMAC binding buffer and a stripping step was performed using the stripping buffer (8 M urea, 1 M NaCl, 10 mM DTT, 1 M Imidazole, 50 mM Tris, pH 7.4) to remove strongly bound proteins.

6.2.3. Dilution refolding of HBx IBs

The prepared HBx IBs were first purified using a 1 ml Ni^{2+}Sepharose HisTrap HP column on an AKTA Explorer FPLC system at a constant mobile phase flow rate of 1.0 ml/min. 10 mg of the solubilised denatured-reduced HBx IBs were loaded onto the column pre-equilibrated in the IMAC binding buffer. The bound proteins were then eluted with the stripping buffer. Dilution refolding was performed through a buffer exchange step using a Hi-Trap Desalting column (GE Healthcare) equilibrated in the refolding buffer (2 M urea, 0.25 M L-arginine, 0.1 mM GSH and 0.01 mM GSSG in 50 mM Tris, pH 7.4). The purified HBx IBs were loaded into the desalting column and the eluate which was 1.33-fold diluted was collected, and incubated overnight at 4°C, followed by determination of the HBx refolding yield (Section 6.2.4.2).

6.2.4. Analytical methods

6.2.4.1. Determination of HBx soluble yields

HBx soluble yield and purity were measured by a chip-based electrophoresis method performed on an Agilent 2100 Bioanalyzer® in combination with the Protein 230 Plus LabChip® kit as discussed in Chapter 4 (Section 4.2.13). Due to the absence of any useable commercially available native HBx standards, for each chip based analysis, the solubilised HBx IBs (concentrations determined by Bradford assay) were diluted to different concentrations and used to generate a calibration curve for determining HBx concentrations after dilution and IMAC refolding. The samples and chips were prepared according to the protocol provided with the Protein 230 Plus LabChip® kit. All samples used for the analysis were centrifuged and filtered using a 0.2 μm filter prior to analysis. HBx soluble yields were determined by Equation 6-1:

$$\text{Soluble Yield} = \frac{\text{Mass of soluble HBx as determined by Bioanalyser}}{\text{Mass of denatured-reduced HBx IB loaded on column}} \times 100\ \% \quad \text{(Eq 6-1)}$$

6.2.4.2. Determination of HBx refolding yields

HBx protein refolding yields were determined by the ELISA kit developed in-house for exclusively determining the refolding yields of 6His-tagged HBx proteins (described in Chapter 5) [20]. Refolding yields were calculated based on Equation 6-2.

$$\text{Refolding Yield} = \frac{\text{Mass of bioactive HBx as determined by ELISA}}{\text{Mass of denatured-reduced HBx IB loaded on column}} \times 100\ \% \quad \text{(Eq 6-2)}$$

6.2.5. DoE-directed design of an optimised IMAC refolding process

The design of the IMAC refolding process was guided by the Statistical DoE methodology. The impact of different process parameters on the product CQA (i.e. HBx protein yields and concentration) were determined by an Ishikawa diagram, which helps to identify all the steps and or sub-steps affecting the CQA of the process [155-156]. Optimisation of the IMAC refolding process was performed by employing a two level full factorial central composite experimental design with three operating parameters as the independent variables: (i) on-column incubation

period (number of days), (ii) protein load on the column, and (iii) the rate of change of denaturing buffer to refolding buffer (number of CVs used to perform the buffer exchange). The levels of the factors considered and the corresponding coded symbols used for the DoE analyses are summarised in Table 6-1. The experimental plan consisted of 25 trials including one centre point as detailed in Table 6-2. The response variables (i.e. HBx protein yield and concentration) were fitted into a polynomial equation of the form shown in Equation 6-3:

$$Y_i = \beta_o + \sum \beta_i x_i + \sum \beta_{ij} x_i x_j + \sum \beta_{ijk} x_i x_j x_k \qquad \text{(Eq 6-3)}$$

where Y_i is the predicted response variable and $x_i x_j$ are the independent variables. β_o is the offset term, β_i is the *i*th linear coefficient for the main effects, β_{ij} is the coefficient for the two-way interactions while β_{ijk} is the interaction coefficient for the three-way interactions. The polynomial coefficients were calculated and analysed using the 'Minitab 15' (Minitab Inc.) statistical software package. Statistical analyses of the models were performed by the analysis of variance (ANOVA).

Table 6-1. Independent (process) variables and the corresponding coded values used for the DoE stud**y.**

Factors	**Coded Symbol**	**Actual Values of the Coded Levels**		
		-1	0	1
Incubation period (days)	X_1	1	2	3
Protein load (mg)	X_2	5	7.5	10
Buffer exchange gradient (CV)	X_3	0	2.5	5

6.3. Results and discussion

6.3.1. DoE scheme and data analysis method for HBx bioprocess development

The DoE approach adopted to guide the development of the HBx IMAC refolding process is summarised in Figure 6-1. The overall aims for employing the DoE approach are to i) systematically determine how different variables or process parameters of the IMAC refolding

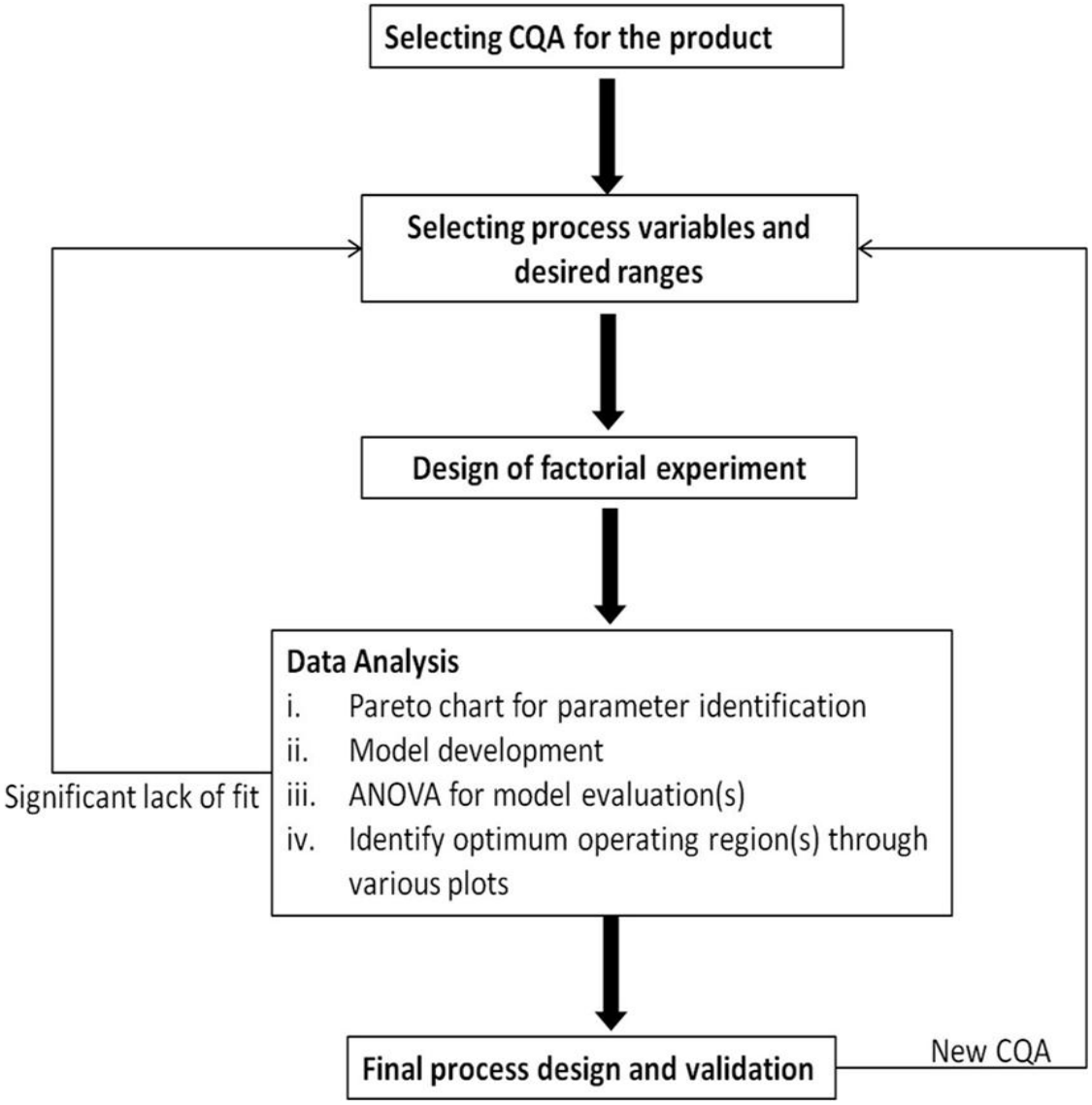

Figure 6-1 DoE scheme employed for rational HBx bioprocess development.

process affect the process CQA (i.e. HBx refolding yield and concentration) and ii) obtain the design space which can provide the optimum CQAs for the target product. An Ishikawa diagram was constructed to identify different process parameters that can directly impact the process CQA for the DoE analysis [154-155]. Based on the number of parameters selected, a set of experiments were then designed at statistically determined points to obtain maximum information with minimum number of experiments. The experimental design, in general, largely depends on the number of parameters to be tested and the feasibility of the experimental set-up required or the desired resolution [128]. In this study, a two-level full factorial central composite

design was used along with one centre point to obtain full resolution data for the HBx refolding process. A standardised effects plot was then constructed to identify which of the parameters, or their combined interactions, significantly affected the response variables. The terms that significantly influenced the response variables (or the CQAs) were then chosen to develop semi-empirical models which would provide the predicted responses of the CQA parameters within the tested range of the process variables. Statistical quality of the models was assessed by ANOVA studies which obtained: a) the significance of the parameters used within the models, and b) the source of residual errors obtained from the models. In general, the residual error possesses two important components: lack of fit (error due to improper model fitting) and pure error (experimental error). The exclusion of any critical process parameters from this study would result in a significant lack of fit error ($P < 0.05$), necessitating a reconsideration of the process variables tested. Graphical plots from the developed models were then used to determine optimum process design which maintains the CQA parameters within the desired range. The model-predicted outcomes were subsequently validated with experimental results.

6.3.2. Selection of IMAC refolding parameters for DoE studies

As mentioned in Section 6.1, the chief objective of this chapter is to develop a bioprocess that would allow the production of HBx protein at concentrations and purity which will enable their direct use in structural and drug designing studies. To this end, the two most important CQA requirements that the process must achieve are (i) high HBx refolding yield, and (ii) high HBx protein concentration. To accelerate data acquisition for the optimisation study of IMAC refolding parameters, HBx soluble yields were measured instead of HBx refolding yield for preliminary screening of refolding conditions to indicate correct refolding. This decision was made based on the previous results (discussed in Section 5.3.3, Chapter 5), which showed that HBx soluble yield positively correlated with HBx refolding yield. Using the DoE approach, the process optimisation strategy was designed after thorough consideration of the different parameters that can influence the IMAC column refolding process, as shown in the Ishikawa diagram (Figure 6-2).

Since all column refolding experiments were performed on pre-packed 1 ml Ni^{2+}Sepharose columns (with fixed dimensions) at a consistent flow-rate, parameters such as column

equilibration, bed height and flow rate were fixed in this study. Protein load was chosen as the first process variable to be studied, where protein binding behaviour on-column has been reported to significantly affect refolding yield [151-152, 157]. It was hypothesised that a higher protein load may reduce protein intermolecular distances leading to unwanted protein aggregation during on-column refolding. The highest protein concentration used in previously reported IMAC refolding studies was approximately 5mg protein per ml of resin [152]. Since the

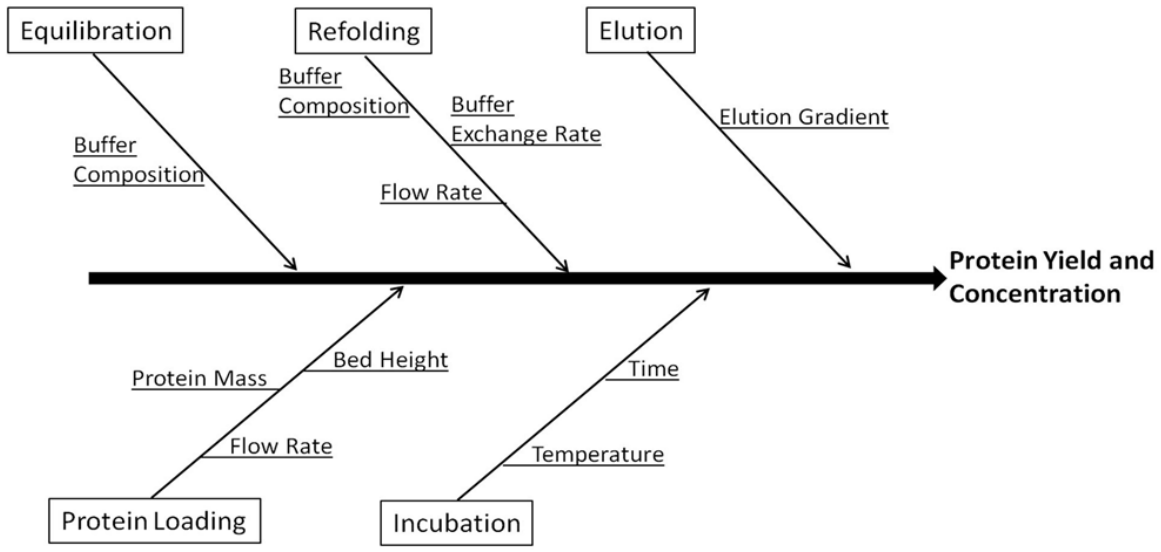

Figure 6-2. Ishikawa diagram showing the various factors that can affect the performance of a column refolding process.

aim of this study is to develop an intensified refolding platform to increase HBx productivity, a high protein load is desired. The effects of protein concentration in the range of 5 – 10 mg/ml on HBx refolding yield and concentration was therefore chosen to be studied.

The success of any refolding process relies on the optimal refolding buffer design. The SVC-optimised dilution refolding buffer, established in Chapter 4, was used for all the on-column refolding experiments. The choice of buffer was based on the consideration that upon immobilisation of the denatured-reduced HBx protein molecules on the Ni^{2+} resin, the majority of the protein molecule would remain suspended in the mobile phase of the column. Hence successful protein refolding would depend on the flexibility of the HBx polypeptide to refold

within the physicochemical environment provided [150, 158]. Furthermore, it was also hypothesised that refolding may be influenced by the rate of change of mobile phase from denaturing to refolding, which would largely depend on rate of mobile phase buffer exchange. Buffer exchange gradients were hence chosen as another process variable for the DoE study. With the aim to minimise refolding time for increased productivity, buffer exchange gradients from 0 - 5 CV were employed for the IMAC refolding studies. As reported in Section 5.3.3, Chapter 5, dilution refolding achieved a maximum refolding yield of ~55% at a refolding concentration of 0.1 mg/ml, after 3 days of protein incubation at 4ºC [20], which reflects slow refolding kinetics for the HBx protein in the dilution refolding step. Hence, in this work, the incubation period (number of days) was chosen as the third process variable for the DoE study. The maximum incubation time under refolding conditions was kept at 3 days; while the lower level of the factor was kept at 1 day. To maximise protein concentration during elution, step elutions rather than gradient elutions were employed for all the IMAC refolding experiments. The process variables (independent variables) thus used for the DoE along with the levels and coded symbols are presented in Table 6-1.

6.3.3. Development and optimisation of an IMAC refolding bioprocess for HBx production

By employing the DoE format, refolding of HBx on the IMAC column was performed by varying the three process variables chosen in Section 6.3.2 over a given range. The impact of different combinations of the process variables on HBx soluble yield and the final protein concentration are summarised in Table 6-2. Preliminary ANOVA studies of the experimental data did not show any significant contribution to data variability due to curvature or non-linearity ($P > 0.05$) of the response variables within the limits of the tested independent variables. Since the centre point (X_1, X_2, $X_3 = 0$, i.e. incubation period = 2 days, protein load = 7.5 mg and buffer exchange gradient = 2.5 CV, respectively) in the DoE design is representative of the curvatures within the experimental range, this point was excluded during subsequent data analyses. Figure 6-3 is a pareto chart which compares the importance of the different single and interaction parameters studied on HBx soluble yield and protein concentration during IMAC as obtained from Table 6-2. Figure 6-3 shows that the protein incubation period has a major impact on both the soluble yield and concentration of the HBx protein. The parameters having a significance

level higher than 95% (as indicated by the vertical line in Figure 6-3) were then chosen for model development to predict HBx soluble yield and protein concentration, as shown in Equations 6-4 and 6-5 below:

$$\text{Soluble Yield (\%)} = 40.04 - 13.96X_1 + 6.13X_2 + 5.21X_3 - 8.79X_1X_3 - 9.04X_2X_3 + 2.79X_1X_2X_3 \quad \text{(Eq 6-4)}$$

[$\mathbf{R^2}$ = 98.85% $\mathbf{R^2(pred)}$ = 97.70% $\mathbf{R^2(adj)}$ = 98.44%]

$$\text{Concentration (mg/ml)} = 0.3 - 0.092X_1 + 0.135X_2 + 0.019X_3 - 0.054X_1X_2 - 0.066X_1X_3 - 0.038X_2X_3 \quad \text{(Eq 6-5)}$$

[$\mathbf{R^2}$ = 99.31%; $\mathbf{R^2(pred)}$ = 98.62%; $\mathbf{R^2(adj)}$ = 99.06%]

The quality of the developed models are indicated by the R^2, R^2(pred) and R^2(adj) values, where R^2 indicates the net variability within the data that is explained by the respective model. Therefore, the closer the value is to 100%, the better the model. The predicted R^2 [R^2(pred)] indicates how well the model will describe any future data, while the Adjusted R^2 [R^2(adj)] is a modified form of R^2 that takes into consideration the number of terms used within the model. Hence the models developed through this DoE strategy explain the obtained experimental results

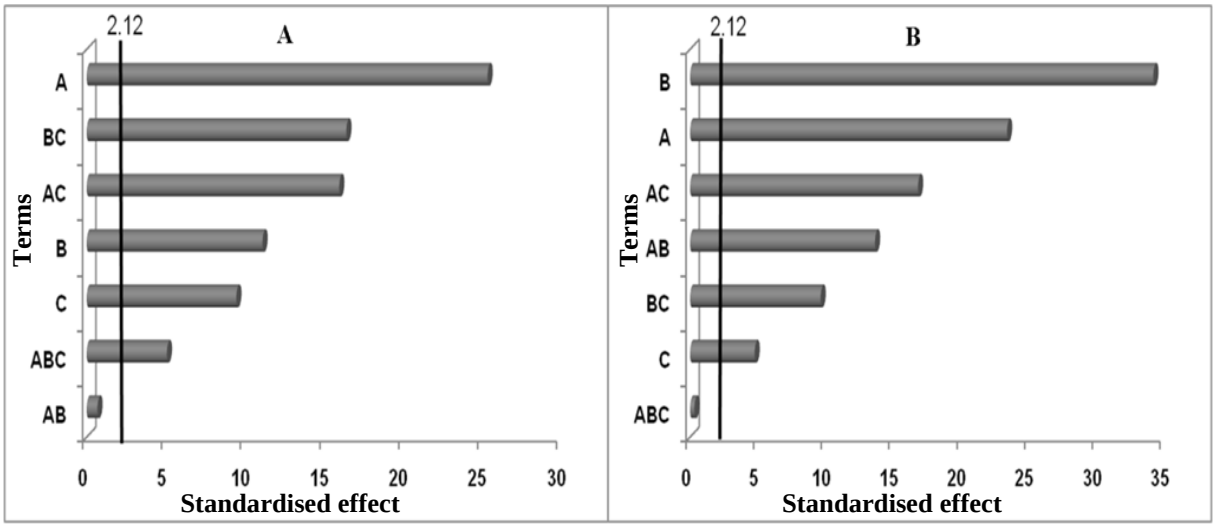

Figure 6-3. Pareto plot showing the influence of the parameters (in order of their importance) on (A) HBx soluble yield and (B) concentration of the eluted HBx protein (under refolding conditions). A: Incubation period (days); B: Protein loaded on to the column (mg); C: Rate of buffer exchange (CVs). Bars crossing the 2.12 line are significant at 95% confidence level.

Table 6-2. DoE experimental design for HBx IMAC refolding and corresponding results.

Coded Levels*			Soluble Yield (%)	Concentration (mg/ml)
X_1	X_2	X_3		
-1	+1	-1	55	0.5
+1	-1	+1	20	0.1
-1	-1	+1	70	0.3
+1	+1	-1	40	0.37
-1	-1	-1	20	0.07
+1	+1	+1	21	0.2
+1	-1	-1	20	0.15
+1	-1	-1	18	0.12
+1	+1	+1	25	0.2
-1	+1	-1	60	0.55
-1	+1	+1	65	0.65
+1	+1	+1	20	0.2
-1	+1	-1	60	0.54
-1	-1	-1	25	0.08
-1	+1	+1	61	0.61
+1	+1	-1	40	0.4
+1	-1	+1	24	0.12
-1	-1	+1	75	0.32
-1	-1	-1	20	0.09
-1	-1	+1	75	0.35
0	0	0	43	0.315
+1	-1	+1	25	0.13
-1	+1	+1	62	0.62
+1	-1	-1	15	0.13
+1	+1	-1	45	0.35

*Details of symbol explained in Table 6-1.

well (shown in Table 6-2), and could therefore be used to predict the optimum IMAC refolding strategy. Furthermore, ANOVA results for our developed models indicated that the error due to lack of fit was insignificant ($P > 0.05$) (Table 6-3), which indicated that the variable process parameters chosen for this investigation were sufficient to explain the observed data variability observed in Table 6-2.

Table 6-3A. ANOVA table for Equation 6-4.

Source	DF	Seq SS	Adj SS	Adj MS	F	P
Main Effects	3	6227.5	6227.46	2075.82	295.61	0
2-Way Interactions	2	3817.1	3817.08	1908.54	271.79	0
3-Way Interactions	1	187	187.04	187.04	26.64	0
Residual Error	17	119.4	119.38	7.02		
Lack of Fit	1	3.4	3.38	3.38	0.47	0.505
Pure Error	16	116	116	7.25		
Total	23	10351				

Table 6-3B. ANOVA table for Equation 6-5.

Source	DF	Seq SS	Adj SS	Adj MS	F	P
Main Effects	3	0.64665	0.64665	0.21555	613.7	0
2-Way Interactions	3	0.20918	0.20918	0.06973	198.52	0
3-Way Interactions	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.
Residual Error	17	0.00597	0.00597	0.00035		
Lack of Fit	1	3.7E-05	3.7E-05	3.7E-05	0.1	0.755
Pure Error	16	0.00593	0.00593	0.00037		
Total	23	0.8618				

DF: Degrees of Freedom; **Seq SS**: Sequential sums of squares; **Adj SS**: Adjusted sum of squares; **Adj MS**: Adjusted mean squares

6.3.4. Designing the optimum IMAC refolding strategy and comparison with dilution refolding

Having developed reliable predictive models represented by Equations 6-4 and 6-5 for the HBx IMAC refolding process, the next aim was to determine the optimum refolding parameters for HBx refolding on-column. From Figure 6-3A, it is clear that the most important parameter affecting the soluble yield of IMAC-refolded HBx was the incubation period, where extended incubation time was detrimental due to the probable oxidation of the GSH molecules in the refolding buffer catalysed by the presence of Ni^{2+} ions [159]. This observation is in agreement with the findings discussed in Section 4.3.2, Chapter 4, where a net reducing environment was crucial for the improved stability of the HBx protein. As the physicochemical environment surrounding HBx in the column becomes more oxidizing over time, the efficiency of disulphide

re-shuffling may be reduced. The importance of incubation period on protein concentration is also demonstrated in Figure 6-3B, where it is the second most important parameter affecting the concentration of IMAC-refolded HBx. The results in Table 6-2 also show that incubation time negatively impacts the two CQA parameters. Therefore, the optimum on-column incubation time for the HBx protein was taken to be 1day.

The investigation was next focused towards understanding how different protein loads and buffer exchange gradients affect HBx soluble yield and concentration at a fixed on-column incubation period of 1 day in the refolding buffer. The results based on the prediction models (Equations 6-4 and 6-5) are shown in the form of contour plots in Figure 6-4. Although the best soluble yields would be expected to be attained at lower protein loads (5 mg) and by using 5 CV buffer exchange gradients (Figure 6-4A), the resulting protein concentration obtained would however, be relatively low (0.3 to 0.4 mg/ml), as shown in Figure 6-4B.

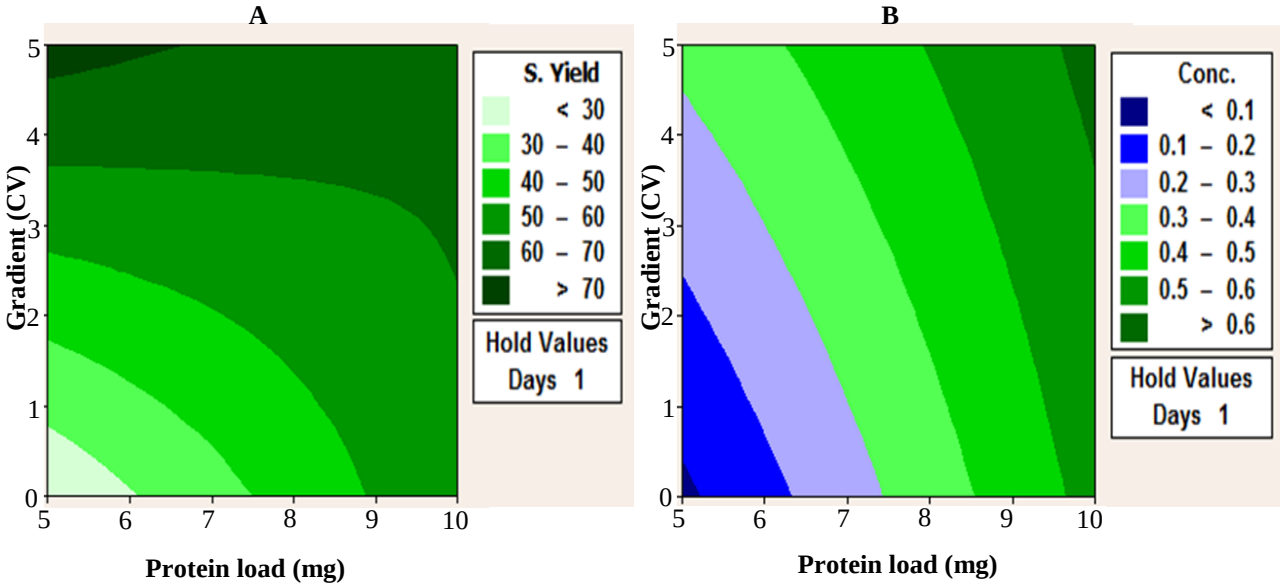

Figure 6-4 (A) Contour plot of soluble yield (S. Yield) versus buffer exchange gradient and protein load at 1 day of incubation. (B) Contour plot of product concentration (Conc.) versus buffer exchange gradient and protein load at 1 day of incubation.

It is also clear from Figure 6-4 that at a fixed protein load, soluble yields can be improved by increasing the length of buffer exchange gradients. The refolding kinetics of different proteins have been found to be significantly affected by the slope of the free energy landscape which in turn is directly influenced by the refolding physicochemical environment [2, 160-162]. In this

study, experimental results clearly demonstrate the importance of optimising the topography of the free energy landscape in on-column refolding studies, where the protein remains largely flexible to refold as only the 6His tag portion of the protein remains bound to the charged Ni^{2+} resins. Therefore, the rate of change in the physicochemical environment will have a similarly critical effect on governing the refolding behaviour such as efficiency and kinetics of the bound HBx proteins, as it would have on free HBx proteins in solution.

Considering that a soluble yield of 60-70% at a concentration of >0.6 mg/ml would provide an acceptable product amount that enables their direct use for structural and other characterisation studies, it was concluded that the following parameters, i.e. 10 mg protein load with 5 CV refolding buffer exchange gradient and 1 day protein incubation, would provide the optimal parameter combination for the IMAC refolding process of the HBx protein. Subsequently, using these conditions, a HBx soluble yield of 63% and eluted HBx concentration of 0.64 mg/ml, was obtained which confirmed the validity of the predicted models (overall process scheme shown in Figure 6-5, Scheme I). HBx soluble yield and concentration were analysed by the Bioanalyzer (Figure 6-6). The bioactivity of the refolded HBx fraction was determined by ELISA, which

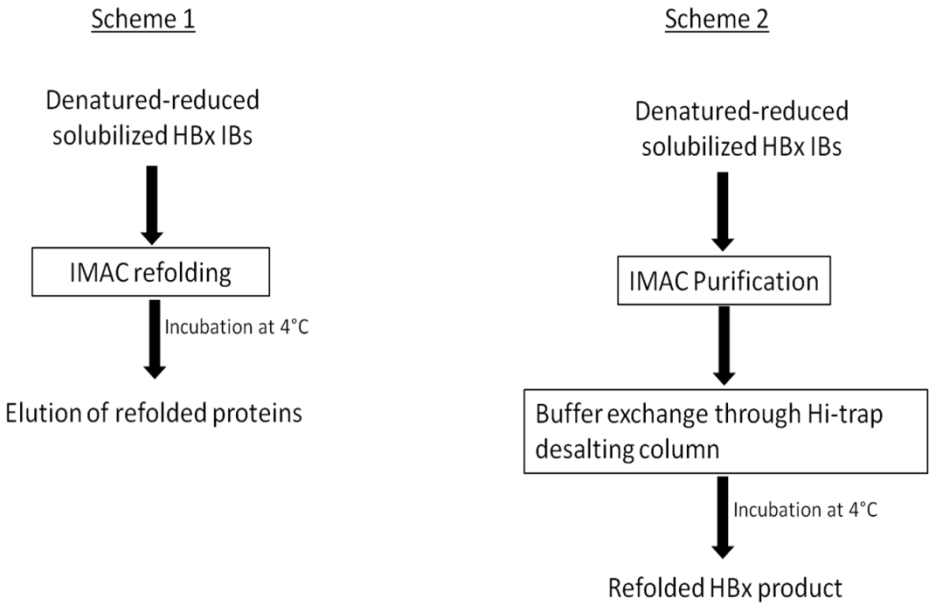

Figure 6-5. HBx refolding process flowsheets based on (1) IMAC refolding and (2) dilution refolding.

gave a HBx refolding yield of 54 %. Although an iterative refolding strategy can be employed to refold the recovered HBx proteins in the stripping step to further improve process yield [152], the eluted protein concentration is likely to be lower than the expected range and hence was not considered.

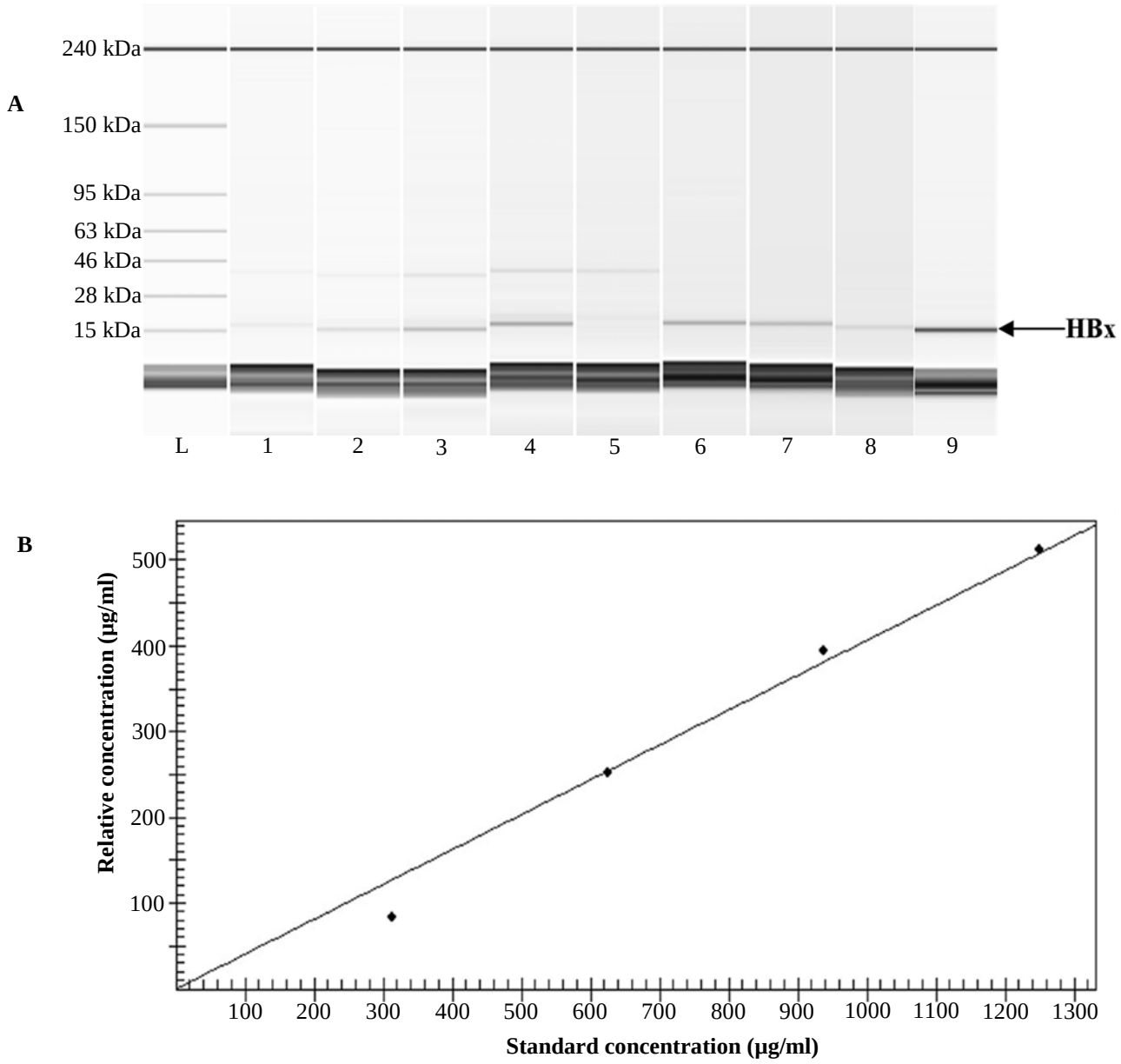

Figure 6-6. (A) Electrophoresis results obtained from Bioanalyzer analysis of the optimised IMAC and dilution refolding processes. L: Protein marker; lanes 1-4: denatured reduced HBx IBs diluted to different concentrations (consisting of 25, 50, 75 and 100% respectively of the solubilised HBx IB solution); lane 5: flowthrough fractions; lanes 6-7: eluted fractions after IMAC refolding; 8: column stripping fractions; 9: dilution-refolded HBx. (B) Calibration curve obtained from denatured-reduced HBx samples in lanes 1-4: $Y = 0.41*X$; $R^2 = 0.9797$

Therefore compared to dilution refolding (at 0.1 mg/ml), refolding the HBx protein using the optimised IMAC refolding process achieved similar refolding yields in 1 day (at 10 mg/ml protein concentration). Spatial isolation of the HBx molecules, achieved through protein immobilisation on an affinity chromatography based platform, thus seems to be vital for improving refolding productivity with the conventional dilution refolding method. To verify this hypothesis, HBx was further refolded and purified using two process schemes i.e. a dilution refolding based process and an IMAC-based process in parallel (Figure 6-5), starting with a consistent initial protein amount of 10 mg. The IMAC refolding process (Figure 6-5, Scheme I) achieved a bioactive HBx amount and productivity that were 2.6 (260%) and 4.4-fold higher than the dilution refolding based process (Figure 6-5, Scheme II), respectively (Table 6-4). These results are not unexpected, considering the fact that the protein concentration range at which the dilution refolding is performed (i.e. ≥5 mg/ml) would almost certainly readily induce intermolecular protein interaction, thereby accelerating non-native off-pathway protein aggregation reactions. HBx protein production was thus found to be significantly improved through the optimised IMAC refolding process. The outcome of this study shows the effectiveness of employing a DoE methodology to generate process parameter information to successfully design, develop and optimise HBx bioprocess, in a time- and resource-saving manner. Such a methodology can be suitably adopted in future for designing the optimum protein refolding process for any other novel proteins like HBx.

Table 6-4. Comparison of the HBx product characteristics obtained from process schemes 1 and 2 at a fixed denatured-reduced protein load of 10 mg.

	Process Scheme	
	1	2
HBx purity (%)	>95	>95
HBx soluble yield (%)	63 ± 3	40 ± 4
HBx refolding yield (%)	54 ± 12	21 ± 3
HBx protein concentration(mg/ml)	>0.6	~ 2.0
Overall bioactive HBx (mg)	5.4 ± 1.2	2.1 ± 0.3
Refolding process productivity (mg/ml/h)	0.21 ± 0.05	0.05 ± 0.01

6.4. Conclusions

Significant efforts are being directed to establish proper design spaces for the manufacturing of biopharmaceuticals which forms a crucial part in establishing QbD approaches, a mandate for current pharmaceutical manufacturing practices to meet stringent regulatory requirements. The development and optimisation of a bioprocess platform based on a DoE design methodology reported in this study is beneficial in establishing the optimised bioprocess right at the development stage, which saves resources whilst maximising information output. As shown in this work, the DoE approach provides critical information on the important parameters affecting the chromatography refolding behaviour for the target HBx protein, while simultaneously providing the design space needed for the subsequent regulatory approval procedures in the near future. The systematic methodology adopted for the current study is the first study of its kind for refolding the HBx protein, leading to the first intensified chromatography-based refolding platform for HBx production, to the best of my knowledge. The models developed through this study can be used to facilitate scale-up of the HBx IMAC refolding process which will then open the way for an efficient and cost-effective supply of the HBx protein to the scientific community striving to develop new drug candidates against HCC “[163]”.

CHAPTER 7: Conclusions and Future Work

7.1. Conclusions

The HBx protein is an example of many candidate proteins that encounter bioprocessing challenges typically faced in many academic and industrial laboratories. The lack of easily available native standards and a reliable bioanalytical platform (to assess the bioactivity of the recombinantly produced protein) are particularly challenging for proteins like HBx that are inherently insoluble and require a refolding based bioprocessing strategy. This thesis presents a methodical and rational approach to develop streamlined bioprocessing strategies for the HBx protein. It is clear from this study that product design and bioprocessing considerations are closely interlinked, where the process shapes the product and vice versa. As discussed in Chapter 3, the presence of a GST fusion partner for a highly hydrophobic protein like HBx failed to improve protein solubility even after extensive fermentation optimisation attempts. The physicochemical environment which was necessary for optimum Factor Xa cleavage activity like the use of Na^{+} and Ca^{2+} ions within a pH range of 6-8 appeared to compromise HBx stability in solution. Faced by these 'product design'-associated obstacles, a new HBx construct was designed to comprise only a C-terminus 6His tag to ease purification and detection. Using this construct, a dilution refolding based HBx bioprocess was successfully developed in Chapter 4. The combination of SVC and DoE proved crucial for a better understanding of HBx solubility and protein-protein interactions at the molecular level, leading to the development of an optimised refolding buffer for the HBx protein. This outcome is significant considering that although huge amounts of information on the impact of different types of additives on the refolding reactions are available, little progress has been made in tool development for rapid optimisation of protein refolding buffers. Although the HBx protein was found to be stable in the soluble form during dilution refolding, the absence of native standards or any bioactivity analytical platform prevented accurate determination of bioactive protein yield, which presented another roadblock for subsequent scale-up efforts.

This roadblock is addressed in Chapter 5, where a quantitative ELISA platform for determination of HBx refolding (or bioactive) yield was successfully developed. The well-studied selective interaction of the HBx protein with the p53 molecule was exploited to determine the amount of bioactive 6His-HBx bound to the ELISA wells which were immobilised with GST-p53. Protein quantification was based on a calibration curve generated using 6His-GST immobilized on the

ELISA wells coupled with colorimetric detection of anti-6His antibodies. This novel ELISA platform allowed the determination of HBx refolding and soluble yields based on the 6His tag detection strategy. In the absence of commercially available high purity HBx standards, the developed ELISA platform will be instrumental in facilitating future design, optimisation and scale-up studies of HBx bioprocesses. To the best of my knowledge, this is the first study on the development of an ELISA based analytical platform for evaluation of recombinant HBx bioactivity, where the distinction between the soluble and the refolding yields of the HBx protein could be established.

Armed with a rationally designed protein refolding buffer system and bio-analytical platform, the stage was set to steer ahead with process intensification studies for the HBx protein using an IMAC based on-column refolding strategy, which is reported in Chapter 6. The use of the statistical DoE approach helped to rapidly establish how critical process parameters like buffer exchange gradients, protein load and on-column incubation period under refolding conditions together affect the chromatography refolding behaviour of the HBx protein. Based on the DoE results, the use of longer buffer–exchange gradients (5 CV) in combination with smaller on-column incubation period (1 day) at a higher protein load (10 mg) was able to yield HBx at concentrations >0.6 mg/ml at a purity of >95%. The final refolding yield (54±12%) obtained from the optimised process was similar to that obtained using the dilution refolding based process albeit at an incubation period of 3 days and a refolding concentration of 0.1 mg/ml, thus indicating improved process productivity using the IMAC based process. The QbD aligned bioprocess design strategy adopted in Chapter 6 also provides the foundations for generating the design space needed for subsequent regulatory approval of other target products and hence can be readily used for other novel protein candidates. To the best of my knowledge, the systematic methodology adopted for HBx refolding and bioprocess design and development (summarised in Figure 7-1) is the first study of its kind for the HBx protein, leading to the first intensified chromatography-based refolding platform for HBx production.

The overall findings of this thesis will not only open the way for establishing an efficient and cost-effective supply of the HBx protein to the scientific community but can also solve some of

the big challenges faced by the modern bio-pharmaceutical industry, which is striving to deliver novel products and meeting increasing demands from the drug regulatory authorities.

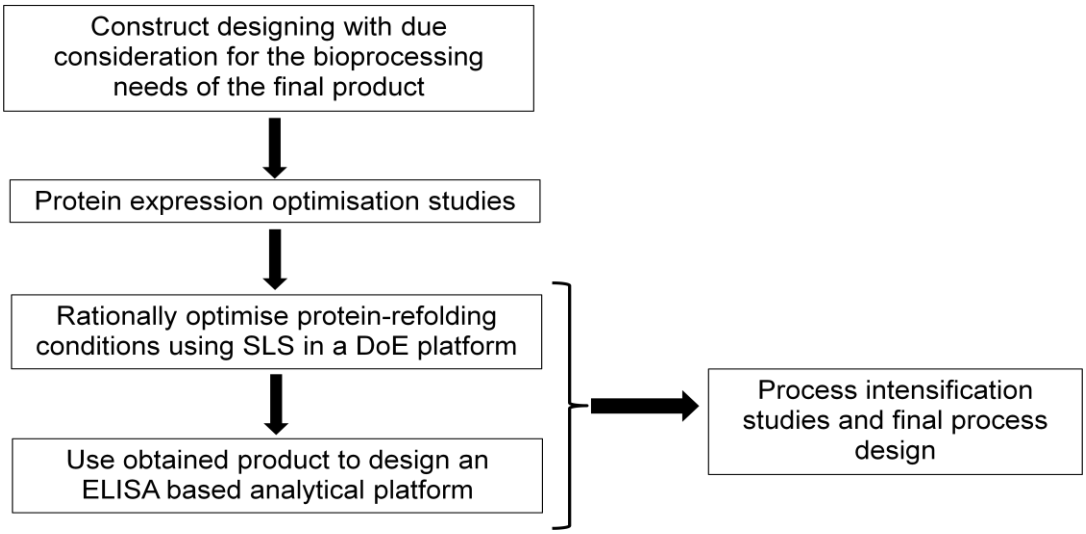

Figure 7-1. Bioprocess design flowsheet for the HBx protein.

7.2. Future Work

With a scalable bioprocess now established for the HBx protein, the subsequent research goals for the project are as follows: (i) HBx stability studies for formulation development of the protein, and (ii) HBx protein structural characterisation.

7.2.1. Formulation and stability studies

Research efforts would be directed towards formulation studies for the HBx protein to ensure delivery of the product in its bioactive form to the end-user. The protein can be delivered in a solubilised form or in a lyophilised form depending on the end-user requirement and final processing costs. The following two sections discuss some of the tentative strategies that would be taken for formulation development and stability testing for the HBx protein in solution and lyophilised forms, respectively.

7.2.1.1. Formulation of HBx solution and stability studies

Producing the HBx protein in the stable solution form would facilitate *in vitro* bioactivity studies of the protein for improved understanding of the clinical functions/applications of the product. However, obtaining an optimum formulation in the liquid form would be challenging particularly

due to stability issues. During storage, protein molecules remain prone to chemical degradation and/or aggregation reactions which is dependent on the physicochemical environment surrounding the protein including temperature, pH and stabilising agents within the formulation. Thus, detailed stability profiles for the designed protein formulations must be obtained to determine the optimum storage conditions and corresponding shelf-lives of the protein solution.

To determine the optimum formulation for the HBx protein in solution form, the refolded HBx protein (obtained through the optimised process presented in this thesis) would be buffer exchanged into different protein stabilising buffers. The use of protein refolding agents and other protein stabilizers such as sugars and glycerol will be studied. Chaotropes under non-denaturing concentrations, however, are known to promote aggregation reactions because of its capability to destabilise the protein by allowing the protein molecules search for other conformations, and hence will not be used. As discussed in Chapter 4, SLS tools can guide the rational design of the formulation buffer, where buffers showing positive SVC values would be only chosen for further protein stability studies.

Following successful formulation development, the next challenge would be to determine the optimum storage temperature of the protein. This is particularly important for determining the experimental strategies to be adopted for subsequent stability studies. Considering the tendency of the HBx protein to aggregate, particular attention would be devoted towards understanding its aggregation kinetics. As such, dynamic light scattering (DLS) methods can be used for *in situ* monitoring of the aggregation reaction process at different temperature. Light scattering tools are particularly advantageous in this regard because of its amenability to modelling studies for methodology iteration. Thermodynamic parameters like the association rate constants under different environments can be obtained through DLS data [164], which in turn can be used to calculate the free energy change for HBx aggregation reactions. Under the same conditions SVC values (obtained through SLS experiments) can be used to gauge the potential of mean force (PMF) for such type of reactions. All these information can be used as input for subsequent Molecular Dynamics simulation studies. These studies can provide detailed mechanistic insight into the molecular interactions leading to HBx aggregation reactions. Such knowledge will not only open up ways to devise the best formulations for the protein but also help in obtaining

mathematical models for the HBx protein folding pathway, which in turn can be used for further improvement and better control of the HBx bioprocessing platform.

The main aim of the HBx stability studies would be to determine aggregation rates for the HBx protein when subjected to different stress conditions associated with temperature and protein concentration, for example. Generally, isothermal and non-isothermal approaches are taken for determining protein aggregation reaction rates. However, considering that non-isothermal methods allow faster determination of temperature dependence of the reaction rate constants while also using lesser amounts of protein samples, the latter method could be more readily employed. Furthermore, non-isothermal methods are advantaged by their compatibility to different mathematical and / or statistical treatments for obtaining the best estimates of the product shelf-life [165]. However, the rate of change of temperature is likely to influence the aggregation rates during the non-isothermal treatment of the HBx solutions. To overcome this problem, light scattering equipment provided with linear, reciprocal, logarithmic and exponential heating algorithms will be used for monitoring the aggregation reactions to obtain the reaction rates at different temperatures for final prediction purposes [165]. Formulations possessing the lowest aggregation rates would be used for subsequent stability studies as per the ICH Q1A-(R2) guidelines to determine the final shelf-life of the product. The bioactivity of the HBx protein within these formulations would be determined using the ELISA methodology developed in Chapter 5, for determining the most feasible HBx formulation [166].

7.2.1.2. HBx lyophilisation and stability studies

Just like the behaviour of many proteins, lyophilised HBx is expected to be more stable than HBx in solution. Apart from the chemical stability, lyophilised protein molecules would be particularly resistant to conformational changes associated with shear stress or surface partitioning frequently encountered in protein solutions. Therefore, alongside the development of a liquefied HBx formulation, efforts would be directed towards developing a lyophilised HBx product as well.

The lyophilisation process involves the removal of water molecules from a frozen protein solution by vacuum sublimation at temperatures below the triple point temperature of water.

Although apparently simple to perform, achieving an optimised formulation for a lyophilisation process is not straightforward chiefly because freezing often induces denaturation of proteins leading to irreversible aggregate formation. The freezing rate of the protein solution thus needs to be carefully optimised, particularly because slow cooling might lead to crystallisation of buffer constituents, which can instigate crystallisation-induced protein aggregation reactions. In other situations, precipitation of buffer constituents can compromise the pH of the protein solution, thus triggering a pH induced aggregation reaction. Suitable stabilising agents would thus be studied for designing the optimum formulation for the lyophilised product. In general, sugars like terhalose and sucrose or other polymers and volatile salts are widely used as stabilising agents for formulating lyophilised proteins. The use of these compounds can be advantageous in two respects: first, these additives prevent over-drying (induced by lyophilisation) of the protein molecules by forming hydrogen bonds between the protein-adhered water molecules and the peptide chain, thereby preventing hydrophobic residues aggregating with each other; second, these compounds form glasses under low temperatures thus preserving the liquid properties even under freezing conditions and preventing undesirable crystallisation and / or chemical side reactions [143, 167].

Stability of the reconstituted protein samples would then be studied with the optimised lyophilisation process. The process which provides maximum HBx activity, following its reconstitution, would be chosen as the final lyophilisation process. Following successful optimisation of the HBx lyophilisation process, the dried HBx product would be again subjected to stability testing in accordance to the ICH Q1A-(R2) guidelines to obtain the shelf-life of the final product [166].

7.2.2. HBx structural studies

Structural elucidation of the HBx protein is vital for subsequent drug designing efforts for HCC treatment. Having laid the foundations for the production of sufficient amounts of the HBx protein at reasonably high concentrations, efforts would be directed towards high resolution structural elucidation through crystallisation and nuclear magnetic resonance (NMR) methodologies. The approach for the structural elucidation studies is briefly described in the following two sections.

7.2.2.1. HBx structural characterisation through crystallisation studies

The primary requirement for any protein crystallisation studies are high purity and high protein concentrations. The high purity HBx obtained from the process described in Chapter 6 can be directly used for crystallisation trials. Following achievement of sufficient quantities of the protein, different buffer compositions would be screened using commercially available high throughput crystallisation screening kits to determine optimum crystallising conditions for HBx. Alternatively, aligned with the DoE strategy adopted in Chapter 4, the effect of different additives can be assessed using SLS measurements to obtain semi-empirical models that can be used to design SVC-optimised HBx crystallisation conditions.

Besides HBx crystallisation studies, efforts would also be directed towards studying crystallisation of the HBx-p53 complex. This study will have to be preceded by attempts to co-express the HBx protein along with the p53 ligand to obtain the HBx-p53 complex in the bacteria host system. A suitable bioprocess will then need to be developed to obtain high concentrations of the complex in purified form for subsequent crystallisation studies. From a drug designing perspective, successful crystallization of this complex can provide vital information about the binding sites of the HBx protein, which in turn can be used to design tailor made anti-HBx lead compounds. Such studies would also allow better characterisation of the HBx-p53 interaction, which can be used for designing other bioanalytical platforms for the HBx protein as well.

7.2.2.1. HBx structural characterisation through NMR spectroscopy

Alongside HBx crystallisation attempts, efforts will be directed towards structural characterisation of the protein using NMR spectroscopy. NMR spectroscopy is advantaged by its capability to provide structural information in the protein solution form. This method requires the use of HBx molecules doped with N^{15} and C^{13} isotopes, which in turn can be obtained by culturing the *E. coli* in chemically defined medium enriched with such isotopes. Considering substantial HBx bioprocessing information is provided within this thesis, revalidating the HBx bioprocess with the labelled isotopes is not expected to pose a great challenge.

The outcomes from these future works will help in subsequent drug designing studies which will open up new therapeutic strategies for treatment of HCC to hopefully eradicate this disease.

References

1. Martin-Vilchez, S., et al., *The molecular and pathophysiological implications of hepatitis B X antigen in chronic hepatitis B virus infection.* Reviews in Medical Virology, 2011. **21**(5): p. 315-329.
2. Radford, S.E., *Protein folding: progress made and promises ahead.* Trends in Biochemical Sciences, 2000. **25**(12): p. 611-618.
3. Lavanchy, D., *Hepatitis B virus epidemiology, disease burden, treatment, and current and emerging prevention and control measures.* Journal of Viral Hepatitis, 2004. **11**(2): p. 97-107.
4. Parkin, D.M., P. Pisani, and J. Ferlay, *Estimates of the worldwide incidence of 25 major cancers in 1990.* International Journal of Cancer, 1999. **80**(6): p. 827-841.
5. Bosch, F.X., et al., *Epidemiology of hepatocellular carcinoma.* Clinics in Liver Disease, 2005. **9**(2): p. 191-211.
6. Kew, M.C., *Epidemiology of hepatocellular carcinoma.* Toxicology, 2002: p. 181-182.
7. Bosch, F.X., J. Ribes, and J. Borràs, *Epidemiology of primary liver cancer.* Seminars in Liver Disease, 1999. **19**(3): p. 271-285.
8. Bergsland, E.K. and A.P. Venook, *Hepatocellular carcinoma.* Current Opinion in Oncology, 2000. **12**(4): p. 357-361.
9. Metodieva, S.N., *Molecular pathogenesis of hepatocellular carcinoma.* Balkan Journal of Medical Genetics, 2007. **10**(2): p. 15-22.
10. Lok, A.S., *Navigating the Maze of Hepatitis B Treatments.* Gastroenterology, 2007. **132**(4): p. 1586-1594.
11. Nassal, M., *New insights into HBV replication: New opportunities for improved therapies.* Future Virology, 2009. **4**(1): p. 55-70.
12. Benhenda, S., et al., *Chapter 4 Hepatitis B Virus X Protein. Molecular Functions and Its Role in Virus Life Cycle and Pathogenesis,* in *Advances in Cancer Research*. 2009. p. 75-109.
13. David H. Nguyen, L.L., and Jianming Hu, *Hepatitis B Virus–Cell Interactions and Pathogenesis.* Journal of Cellular Physiology. , 2008. **216**: p. 289-294.
14. Tang, H., et al., *Molecular functions and biological roles of hepatitis B virus x protein.* Cancer Science, 2006. **97**(10): p. 977-983.
15. Mandal, S., M. Moudgil, and S.K. Mandal, *Rational drug design.* European Journal of Pharmacology, 2009. **625**(1-3): p. 90-100.
16. Gupta, A., et al., *Assignment of disulphide bonds in the X protein (HBx) of hepatitis B virus.* Biochemical and Biophysical Research Communications, 1995. **212**(3): p. 919-924.
17. Marczinovits, I., et al., *An alternative purification protocol for producing hepatitis B virus X antigen on a preparative scale in Escherichia coli.* Journal of Biotechnology, 1997. **56**(2): p. 81-88.
18. Hwang, G.Y., et al., *Detection of the Hepatitis B Virus X Protein (HBx) Antigen and Anti-HBx Antibodies in Cases of Human Hepatocellular Carcinoma.* Journal of Clinical Microbiology, 2003. **41**(12): p. 5598-5603.

19. Zhang, S.M., et al., *HBx protein of hepatitis B virus (HBV) can form complex with mitochondrial HSP60 and HSP70.* Archives of Virology, 2005. **150**(8): p. 1579-1590.
20. Basu, A. and S.S.J. Leong, *Development of an enzyme-linked immunosorbent assay analytical platform for refolding yield determination of recombinant hepatitis B virus X (HBx) protein.* Analytical Biochemistry, 2011. **418**(1): p. 155-157.
21. AparnaGupta , T.K.M., N.Jayasuryan and Virander S.Chauhan *Assignment of Disulphide Bonds in the X Protein (HBx) of Hepatitis B Virus.* Biochemical and Biophysical Research Communications, 1995. **212**(3): p. 919-924.
22. Eisenberg, D., et al., *Protein function the post-genomic era.* Nature, 2000. **405**(6788): p. 823-826.
23. Zhou, C., et al., *High potency and broad-spectrum antimicrobial peptides synthesized via ring-opening polymerization of α-Aminoacid-N-carboxyanhydrides.* Biomacromolecules, 2010. **11**(1): p. 60-67.
24. Ma, J., et al., *The role of hepatitis B virus X protein is related to its differential intracellular localization.* Acta Biochimica et Biophysica Sinica, 2011. **43**(8): p. 583-588.
25. Murakami, S., *Hepatitis B Virus X Protein: Structure, Function and Biology.* Intervirology, 1999. **42**: p. 81–99.
26. Hong Tang, et al., *Molecular functions and biological roles of hepatitis B virus x protein.* Cancer Research, 2006. **97**(10): p. 977-983.
27. Bouchard, J.M. and J.R. Schneider, *The Enigmatic X Gene of Hepatitis B Virus.* Journal of Virology, 2004 **78**: p. 12725-12734.
28. Murakami, S., *Hepatitis B virus X protein: a multifunctional viral regulator.* Journal of Gastroenterology, 2001. **36**(10): p. 651-660.
29. Fischer, R., et al., *Plant-based production of biopharmaceuticals.* Current Opinion in Plant Biology, 2004. **7**(2): p. 152-158.
30. Demain, A.L. and P. Vaishnav, *Production of recombinant proteins by microbes and higher organisms.* Biotechnology Advances, 2009. **27**(3): p. 297-306.
31. Sela, M., F.H. White, and C.B. Anfinsen, *REDUCTIVE CLEAVAGE OF DISULFIDE BRIDGES IN RIBONUCLEASE.* Science, 1957. **125**(3250): p. 691-692.
32. Lehninger, A.L., D.L. Nelson, and M.M. Cox. *Lehninger principles of biochemistry student cd-rom*. 2005; 4th:[1 CD-ROM].
33. Sayam SenGupta, T.S., *Folding, Self-assembly and Conformational Switches of Proteins*. Protein Folding-Misfolding Some Current Concepts in Protein Chemistry., ed. T.S. J.P. Zbilut. 2007: Nova Schience Publishers, Inc. 1-33.
34. Pain, R., *Protein folding for pleasure and for profit.* Trends in Biochemical Sciences, 1987. **12**: p. 309-312.
35. Dolgikh, D.A., et al., *[alpha]-lactalbumin: compact state with fluctuating tertiary structure?* FEBS Letters, 1981. **136**(2): p. 311-315.
36. Mikio Ohgushi and A. Wada, *Molten-globule state' : a compact form of globular proteins with mobile side-chains.* FEBS Letters, 1983. **164**(1): p. 21-24.
37. Creighton, T.E., *Experimental studies of protein folding and unfolding.* Progress in biophysics and molecular biology, 1978. **33**(3): p. 231-297.
38. Dill, K.A. and H.S. Chan, *From Levinthal to pathways to funnels.* Nat Struct Mol Biol, 1997. **4**(1): p. 10-19.

39. Agashe, V.R. and F.U. Hartl, *Roles of molecular chaperones in cytoplasmic protein folding.* Seminars in Cell & Developmental Biology, 2000. **11**(1): p. 15-25.
40. Minton, A.P., *Implications of macromolecular crowding for protein assembly.* Current Opinion in Structural Biology, 2000. **10**(1): p. 34-39.
41. Bert van den Berg, R.J.E.a. and C. M.Dobson, *The EMBO Journal.* 1999. **18**(24): p. 6927-6933.
42. Ellis, J., *Proteins as molecular chaperones.* Nature, 1987. **328**(6129): p. 378-379.
43. Hue Sun Chan, K.A.D., *A simple model of chaperonin-mediated protein folding.* Proteins: Structure, Function, and Genetics, 1996. **24**(3): p. 345-351.
44. Todd, M.J., G.H. Lorimer, and D. Thirumalai, *Chaperonin-facilitated protein folding: optimization of rate and yield by an iterative annealing mechanism.* Proceedings of the National Academy of Sciences of the United States of America, 1996. **93**(9): p. 4030-4035.
45. Kouhei Tsumoto, et al., *Practical considerations in refolding proteins from inclusion bodies.* Protein Expression and Purification, 2003 **28**: p. 1-8.
46. Bhavesh, N.S., et al., *NMR identification of local structural preferences in HIV-1 protease tethered heterodimer in 6 M guanidine hydrochloride.* FEBS Letters, 2001. **509**(2): p. 218-224.
47. Jungbauer, A. and W. Kaar, *Current status of technical protein refolding.* Journal of Biotechnology, 2007. **128**(3): p. 587-596.
48. Kiefhaber, T., et al., *Protein aggregation in vitro and in vivo: A quantitative model of the kinetic competition between folding and aggregation.* Bio/Technology, 1991. **9**(9): p. 825-829.
49. Vallejo, L.F. and U. Rinas, *Optimized Procedure for Renaturation of Recombinant Human Bone Morphogenetic Protein-2 at High Protein Concentration.* Biotechnology and Bioengineering, 2004. **85**(6): p. 601-609.
50. Fischer, B., et al., *A novel sequential procedure to enhance the renaturation of recombinant protein from Escherichia coli inclusion bodies.* Protein Engineering, 1992. **5**(6): p. 593-596.
51. Ho, J.G.S., et al., *The likelihood of aggregation during protein renaturation can be assessed using the second virial coefficient.* Protein Science, 2003. **12**(4): p. 708-716.
52. Terashima, M., K. Suzuki, and S. Katoh, *Effective refolding of fully reduced lysozyme with a flow-type reactor.* Process Biochemistry, 1996. **31**(4): p. 341-345.
53. De Bernardez Clark, E., E. Schwarz, and R. Rudolph, *Inhibition of aggregation side reactions during in vitro protein folding.* Methods in Enzymology, 1999. **309**: p. 217-236.
54. Maeda, Y., et al., *Effective renaturation of reduced lysozyme by gentle removal of urea.* Protein Engineering, 1995. **8**(2): p. 201-205.
55. Sahdev, S., S.K. Khattar, and K.S. Saini, *Production of active eukaryotic proteins through bacterial expression systems: A review of the existing biotechnology strategies.* Molecular and Cellular Biochemistry, 2008. **307**(1-2): p. 249-264.
56. Varnerin, J.P., et al., *Production of leptin in Escherichia coli: A comparison of methods.* Protein Expression and Purification, 1998. **14**(3): p. 335-342.
57. Leong, S.S.J. and A.P.J. Middelberg, *Dilution Versus Dialysis: A Quantitative Study of the Oxidative Refolding of Recombinant Human Alpha-fetoprotein.* Food and Bioproducts Processing, 2006. **84**(1): p. 9-17.

58. West, S.M., J.B. Chaudhuri, and J.A. Howell, *Improved protein refolding using hollow-fibre membrane dialysis.* Biotechnology and Bioengineering, 1998. **57**(5): p. 590-599.
59. Yoshii, H., et al., *Refolding of denatured/reduced lysozyme at high concentration with diafiltration.* Bioscience, Biotechnology and Biochemistry, 2000. **64**(6): p. 1159-1165.
60. Gu, Z., et al., *Chromatographic Methods for the Isolation of, and Refolding of Proteins from, Escherichia coli Inclusion Bodies.* Protein Expression and Purification, 2002. **25**(1): p. 174-179.
61. Batas, B., H.R. Jones, and J.B. Chaudhuri, *Studies of the hydrodynamic volume changes that occur during refolding of lysozyme using size-exclusion chromatography.* Journal of Chromatography A, 1997. **766**(1-2): p. 109-119.
62. Li, M., G. Zhang, and Z. Su, *Dual gradient ion-exchange chromatography improved refolding yield of lysozyme.* Journal of Chromatography A, 2002. **959**(1-2): p. 113-120.
63. Li, M. and Z.-G. Su, *Refolding of superoxide dismutase by ion-exchange chromatography.* Biotechnology Letters, 2002. **24**(11): p. 919-923.
64. Anne B.Mason, Q.-Y.H., Peter J. Halbrooks, Stephen J. Everse, Dmitry R. Gumerov, Igor A. Kaltashov, Valerie C. Smith, Jeff Hewitt, Ross T. A. MacGillivray, , *Differential effect of a His tag at the N- and C-termini: Functional studies with recombinant human serum transferrin.* Biochemistry 2002. **41**(30): p. 9448-9454.
65. Shi, Y., et al., *One-step on-column affinity refolding purification and functional analysis of recombinant human VDAC1.* Biochemical and Biophysical Research Communications, 2003. **303**(2): p. 475-482.
66. Li, M., Z.-G. Su, and J.-C. Janson, *In vitro protein refolding by chromatographic procedures.* Protein Expression and Purification, 2004. **33**(1): p. 1-10.
67. Geng, X., et al., *Refolding and purification of interferon-gamma in industry by hydrophobic interaction chromatography.* Journal of Biotechnology, 2004. **113**(1-3): p. 137-149.
68. Malavasi, N.V., et al., *Protein refolding at high pressure: Optimization using eGFP as a model.* Process Biochemistry, 2011. **46**(2): p. 512-518.
69. Seefeldt, M.B., et al., *Application of high hydrostatic pressure to dissociate aggregates and refold proteins.* Current Pharmaceutical Biotechnology, 2009. **10**(4): p. 447-455.
70. Qoronfleh, M.W., L.K. Hesterberg, and M.B. Seefeldt, *Confronting high-throughput protein refolding using high pressure and solution screens.* Protein Expression and Purification, 2007. **55**(2): p. 209-224.
71. Vincentelli, R., et al., *High-throughput automated refolding screening of inclusion bodies.* Protein Science, 2004. **13**(10): p. 2782-2792.
72. Basu, A., X. Li, and S. Leong, *Refolding of proteins from inclusion bodies: rational design and recipes.* Applied Microbiology and Biotechnology, 2011. **92**(2): p. 241-251.
73. George, A. and W.W. Wilson, *Predicting protein crystallization from a dilute solution property.* Acta Crystallogr. D, 1994. **50**: p. 361-365.
74. Pan, X. and C.E. Glatz, *Solvent effects on the second virial coefficient of subtilisin and solubility.* Crystal Growth and Design, 2003. **3**(2): p. 203-207.
75. Valente, J.J., et al., *Colloidal behavior of proteins: Effects of the second virial coefficient on solubility, crystallization and aggregation of proteins in aqueous solution.* Current Pharmaceutical Biotechnology, 2005. **6**(6): p. 427-436.
76. Guo, B., et al., *Correlation of second virial coefficients and solubilities useful in protein crystal growth.* Journal of Crystal Growth, 1999. **196**(2-4): p. 424-433.

77. Ruppert, S., S.I. Sandler, and A.M. Lenhoff, *Correlation between the osmotic second virial coefficient and the solubility of proteins.* Biotechnology Progress, 2001. **17**(1): p. 182-187.
78. Curtis, R.A. and L. Lue, *A molecular approach to bioseparations: Protein-protein and protein-salt interactions.* Chemical Engineering Science, 2006. **61**(3): p. 907-923.
79. Rosenbaum, D.F. and C.F. Zukoski, *Protein interactions and crystallization.* Journal of Crystal Growth, 1996. **169**(4): p. 752-758.
80. Neal, B.L., D. Asthagiri, and A.M. Lenhoff, *Molecular origins of osmotic second virial coefficients of proteins.* Biophysical Journal, 1998. **75**(5): p. 2469-2477.
81. Tessier, P.M. and A.M. Lenhoff, *Measurements of protein self-association as a guide to crystallization.* Current Opinion in Biotechnology, 2003. **14**(5): p. 512-516.
82. Velev, O.D., E.W. Kaler, and A.M. Lenhoff, *Protein interactions in solution characterized by light and neutron scattering: Comparison of lysozyme and chymotrypsinogen.* Biophysical Journal, 1998. **75**(6): p. 2682-2697.
83. Kulkarni, A.M., et al., *Effects of polyethylene glycol on protein interactions.* Journal of Chemical Physics, 2000. **113**(21): p. 9863-9873.
84. Kulkarni, A. and C. Zukoski, *Depletion interactions and protein crystallization.* Journal of Crystal Growth, 2001. **232**(1-4): p. 156-164.
85. Curtis, R.A., et al., *Protein–protein interactions in concentrated electrolyte solutions.* Biotechnology and Bioengineering, 2002. **79**(4): p. 367-380.
86. Liu, W., et al., *Effect of alcohols on aqueous lysozyme-lysozyme interactions from static light-scattering measurements.* Biophysical Chemistry, 2004. **107**(3): p. 289-298.
87. Ho, J.G.S. and A.P.J. Middelberg, *Estimating the potential refolding yield of recombinant proteins expressed as inclusion bodies.* Biotechnology and Bioengineering, 2004. **87**(5): p. 584-592.
88. Basu, A., W.N. Chen, and S.S.J. Leong, *A rational design for hepatitis B virus X protein refolding and bioprocess development guided by second virial coefficient studies.* Applied Microbiology and Biotechnology, 2011. **90**(1): p. 181-191.
89. Pugh, J.C., et al., *Expression of the X gene of hepatitis B virus.* Journal of Medical Virology, 1986. **20**(3): p. 229-246.
90. Kay, A., et al., *The HBV HBX gene expressed in E. coli is recognised by sera from hepatitis patients.* EMBO Journal, 1985. **4**(5): p. 1287-1292.
91. Song, C.Z., Z.L. Bai, and Q.W. Wang, *Aggregate formation of hepatitis B virus X protein affects cell cycle and apoptosis.* World Journal of Gastroenterology, 2003. **9**(7): p. 1521-1524.
92. Henkler, F., et al., *Intracellular localization of the hepatitis B virus HBx protein.* Journal of General Virology, 2001. **82**(4): p. 871-882.
93. Urban, S., et al., *Isolation and molecular characterization of hepatitis B virus X-protein from a baculovirus expression system.* Hepatology, 1997. **26**(4): p. 1045-1053.
94. Melegari, M., P.P. Scaglioni, and J.R. Wands, *Cloning and characterization of a novel hepatitis B virus x binding protein that inhibits viral replication.* JOURNAL OF VIROLOGY, 1998. **72**(3): p. 1737-1743.
95. Nishihara, K., et al., *Chaperone coexpression plasmids: Differential and synergistic roles of DnaK-DnaJ-GrpE and GroEL-GroES in assisting folding of an allergen of Japanese cedar pollen, Cryj2, in Escherichia coli.* Applied and Environmental Microbiology, 1998. **64**(5): p. 1694-1699.

96. Kirschner, A., J. Altenbuchner, and U.T. Bornscheuer, *Cloning, expression, and characterization of a Baeyer-Villiger monooxygenase from Pseudomonas fluorescens DSM 50106 in E. coli.* Applied Microbiology and Biotechnology, 2007. **73**(5): p. 1065-1072.
97. de Marco, A., et al., *Chaperone-based procedure to increase yields of soluble recombinant proteins produced in E. coli.* BMC Biotechnology, 2007. **7**.
98. Bryksa, B.C., et al., *A C-terminal glycine suppresses production of pleurocidin as a fusion peptide in Escherichia coli.* Protein Expression and Purification, 2006. **45**(1): p. 88-98.
99. Pazgier, M. and J. Lubkowski, *Expression and purification of recombinant human α-defensins in Escherichia coli.* Protein Expression and Purification, 2006. **49**(1): p. 1-8.
100. Xu, X., et al., *Expression and purification of a recombinant antibacterial peptide, cecropin, from Escherichia coli.* Protein Expression and Purification, 2007. **53**(2): p. 293-301.
101. Huang, L., et al., *Production of bioactive human beta-defensin 5 and 6 in Escherichia coli by soluble fusion expression.* Protein Expression and Purification, 2008. **61**(2): p. 168-174.
102. Huang, L., S.S.J. Leong, and R. Jiang, *Soluble fusion expression and characterization of bioactive human beta-defensin 26 and 27.* Applied Microbiology and Biotechnology, 2009. **84**(2): p. 301-308.
103. de Moura, P.R., et al., *The cysteine residues of the hepatitis B virus onco-protein HBx are not required for its interaction with RNA or with human p53.* Virus Research, 2005. **108**(1-2): p. 121-131.
104. Rui, E., P.R. de Moura, and J. Kobarg, *Expression of deletion mutants of the hepatitis B virus protein HBx in E. coli and characterization of their RNA binding activities*. Virus Research, 2001. **74**(1-2): p. 59-73.
105. Esposito, D. and D.K. Chatterjee, *Enhancement of soluble protein expression through the use of fusion tags.* Current Opinion in Biotechnology, 2006. **17**(4): p. 353-358.
106. Khattar, S.K., et al., *Enhanced soluble production of biologically active recombinant human p38 mitogen-activated-protein kinase (MAPK) in Escherichia coli.* Protein and Peptide Letters, 2007. **14**(8): p. 756-760.
107. Frangioni, J.V. and B.G. Neel, *Solubilization and purification of enzymatically active glutathione S-transferase (pGEX) fusion proteins.* Analytical Biochemistry, 1993. **210**(1): p. 179-187.
108. Jameel, S., et al., *Hepatitis B virus X protein produced in Escherichia coli is biologically functional.* Journal of Virology, 1990. **64**(8): p. 3963-3966.
109. De-Medina, T., et al., *The X protein of hepatitis B virus has a ribo/deoxy ATPase activity.* Virology, 1994. **202**(1): p. 401-407.
110. Liu, D., et al., *High-level expression and large-scale preparation of soluble HBx antigen from Escherichia coli.* Biotechnology and Applied Biochemistry, 2009. **54**(3): p. 141-147.
111. Mbayed, V.A., et al., *Molecular characterization of hepatitis B virus genotype A from Argentina and Brazil.* Archives of Virology, 2009. **154**(3): p. 525-529.
112. Hoffmann, F. and U. Rinas, *Stress induced by recombinant protein production in Escherichia coli.* Advances in biochemical engineering/biotechnology, 2004. **89**: p. 73-92.

113. Harper, S. and D.W. Speicher, *Expression and purification of GST fusion proteins.* Current Protocols in Protein Science, 2008(SUPPL. 52): p. 6.6.1-6.6.26.
114. Sambrook, J. and D.W. Russell, *Molecular cloning : a laboratory manual*. 3rd ed. 2001, Cold Spring Harbor, N.Y.: Cold Spring Harbor Laboratory Press.
115. Leong, S.S.J. and A.P.J. Middelberg, *A simplified bioprocess for human alpha-fetoprotein production from inclusion bodies.* Biotechnology and Bioengineering, 2007. **97**(1): p. 99-117.
116. Bradford, M.M., *A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein dye binding.* Analytical Biochemistry, 1976. **72**(1-2): p. 248-254.
117. Privalov, P.L., *Thermodynamics of protein folding.* Journal of Chemical Thermodynamics, 1997. **29**(4): p. 447-474.
118. Francis, D.M. and R. Page, *Strategies to optimize protein expression in E. coli.* Current Protocols in Protein Science, 2010(SUPPL. 61): p. 5.24.1-5.24.29.
119. Flickinger, M.C. and S.W. Drew, *Encyclopedia of bioprocess technology : fermentation, biocatalysis, and bioseperation*. 1999, New York: John Wiley. 5 v. (2756 p.).
120. Middelberg, A.P.J., *Preparative protein refolding.* Trends in Biotechnology, 2002. **20**(10): p. 437-443.
121. Fischer, B., I. Sumner, and P. Goodenough, *Isolation, renaturation, and formation of disulfide bonds of eukaryotic proteins expressed in Escherichia coli as inclusion bodies.* Biotechnology and Bioengineering, 1993. **41**(1): p. 3-13.
122. Tsumoto, K., et al., *Role of arginine in protein refolding, solubilization, and purification.* Biotechnology Progress, 2004. **20**(5): p. 1301-1308.
123. Jenny, R.J., K.G. Mann, and R.L. Lundblad, *A critical review of the methods for cleavage of fusion proteins with thrombin and factor Xa.* Protein Expression and Purification, 2003. **31**(1): p. 1-11.
124. Underwood, M.C., et al., *Thermodynamic linkage between the S1 site, the Na+ site, and the Ca2+ site in the protease domain of human coagulation factor Xa. Studies on catalytic efficiency and inhibitor binding.* Journal of Biological Chemistry, 2000. **275**(47): p. 36876-36884.
125. Aldrich, S. *N-Lauroylsarcosine sodium salt*. 2010 [cited 2010 16 March 2010]; N-Lauroylsarcosine sodium salt properties]. Available from: http://www.sigmaaldrich.com/catalog/ProductDetail.do?lang=en&N4=L9150|SIGMA&N5=SEARCH_CONCAT_PNO|BRAND_KEY&F=SPEC.
126. Greenshields, A.L., et al., *Strategies for recombinant expression of small, highly disulphide-bonded, cationic antimicrobial peptides.* Protein and Peptide Letters, 2008. **15**(9): p. 985-994.
127. Fahnert, B., H. Lilie, and P. Neubauer, *Inclusion bodies: formation and utilisation.* Advances in biochemical engineering/biotechnology, 2004. **89**: p. 93-142.
128. Mandenius, C.F. and A. Brundin, *Bioprocess optimization using design-of-experiments methodology.* Biotechnology Progress, 2008. **24**(6): p. 1191-1203.
129. Ferrer, M., et al., *Chaperonins govern growth of Escherichia coli at low temperatures [2]*. Nature Biotechnology, 2003. **21**(11): p. 1266-1267.
130. Stockmayer, W.H., *Light scattering in multi-component systems.* The Journal of Chemical Physics, 1950. **18**(1): p. 58-61.

131. Blackwell, J.R. and R. Horgan, *A novel strategy for production of a highly expressed recombinant protein in an active form.* FEBS Letters, 1991. **295**(1-3): p. 10-12.
132. Tsumoto, K., et al., *Practical considerations in refolding proteins from inclusion bodies.* Protein Expression and Purification, 2003. **28**(1): p. 1-8.
133. Tischer, A., et al., *L-arginine hydrochloride increases the solubility of folded and unfolded recombinant plasminogen activator rPA.* Protein Science, 2010. **19**(9): p. 1783-1795.
134. Reddy K, R.C., et al., *L-Arginine increases the solubility of unfolded species of hen egg white lysozyme.* Protein Science, 2005. **14**(4): p. 929-935.
135. Hevehan, D.L. and E. De Bernardez Clark, *Oxidative renaturation of lysozyme at high concentrations.* Biotechnology and Bioengineering, 1997. **54**(3): p. 221-230.
136. Leong, S.S.J. and A.P.J. Middelberg, *The refolding of different α-fetoprotein variants.* Protein Science, 2006. **15**(9): p. 2040-2050.
137. Huang, Z. and S.S.J. Leong, *Molecular-assisted refolding: Study of two different ionic forms of recombinant human fibroblast growth factors.* Journal of Biotechnology, 2009. **142**(2): p. 157-163.
138. Feitelson, M.A., et al., *Hepatitis B x antigen and p53 are associated in vitro and in liver tissues from patients with primary hepatocellular carcinoma.* Oncogene, 1993. **8**(5): p. 1109-1117.
139. Truant, R., et al., *Direct interaction of the hepatitis B virus HBx protein with p53 leads to inhibition by HBx of p53 response element-directed transactivation.* JOURNAL OF VIROLOGY, 1995. **69**(3): p. 1851-1859.
140. Lin, Y., et al., *The hepatitis B virus X protein is a co-activator of activated transcription that modulates the transcription machinery and distal binding activators.* Journal of Biological Chemistry, 1998. **273**(42): p. 27097-27103.
141. Van Den Berg, B., et al., *The oxidative refolding of hen lysozyme and its catalysis by protein disulfide isomerase.* EMBO Journal, 1999. **18**(17): p. 4794-4803.
142. Daly, N.L., R.J. Clark, and D.J. Craik, *Disulfide folding pathways of cystine knot proteins: Tying the knot within the circular backbone of the cyclotides.* Journal of Biological Chemistry, 2003. **278**(8): p. 6314-6322.
143. Wang, W., *Protein aggregation and its inhibition in biopharmaceutics.* International Journal of Pharmaceutics, 2005. **289**(1-2): p. 1-30.
144. Arolas, J.L., et al., *Folding of small disulfide-rich proteins: clarifying the puzzle.* Trends in Biochemical Sciences, 2006. **31**(5): p. 292-301.
145. Kelly, S.M., T.J. Jess, and N.C. Price, *How to study proteins by circular dichroism.* Biochimica et Biophysica Acta (BBA) - Proteins & Proteomics, 2005. **1751**(2): p. 119-139.
146. Rui, E., et al., *Expression and spectroscopic analysis of a mutant hepatitis B virus onco-protein HBx without cysteine residues.* Journal of Virological Methods, 2005. **126**(1-2): p. 65-74.
147. Kew, M.C., *Hepatitis B virus x protein in the pathogenesis of hepatitis B virus-induced hepatocellular carcinoma.* Journal of Gastroenterology and Hepatology, 2011. **26**(SUPPL. 1): p. 144-152.
148. Vogel, A.I. and J. Mendham, *Vogel's textbook of quantitative chemical analysis*. 6th ed. 2008, India: Pearson Education Limited. 834 p.

149. Hermanson, G.T., *Bioconjugate techniques*. 2nd ed. 2008, Amsterdam, Boston, London ; Burlington, MA: Elsevier Academic Press. xxx, 1202 p.
150. Chen, Y. and S.S.J. Leong, *Adsorptive refolding of a highly disulfide-bonded inclusion body protein using anion-exchange chromatography.* Journal of Chromatography A, 2009. **1216**(24): p. 4877-4886.
151. Hutchinson, M.H. and H.A. Chase, *Refolding strategies for ketosteroid isomerase following insoluble expression in Escherichia coli.* Biotechnology and Bioengineering, 2006. **94**(6): p. 1089-1098.
152. Hutchinson, M.H. and H.A. Chase, *Adsorptive refolding of histidine-tagged glutathione S-transferase using metal affinity chromatography.* Journal of Chromatography A, 2006. **1128**(1-2): p. 125-132.
153. Rathore, A.S. and H. Winkle, *Quality by design for biopharmaceuticals.* Nat Biotech, 2009. **27**(1): p. 26-34.
154. Harms, J., et al., *Defining Process Design Space for Biotech Products: Case Study of Pichia pastoris Fermentation.* Biotechnology Progress, 2008. **24**(3): p. 655-662.
155. Jiang, C., et al., *Defining process design space for a hydrophobic interaction chromatography (HIC) purification step: Application of quality by design (QbD) principles.* Biotechnology and Bioengineering, 2010. **107**(6): p. 985-997.
156. Shukla, A.A., et al., *Process characterization for metal-affinity chromatography of an Fc fusion protein: a design-of-experiments approach.* Biotechnology and Applied Biochemistry, 2001. **34**(2): p. 71-80.
157. Hutchinson, M.H., et al., *Production of enzymatically active ketosteroid isomerase following insoluble expression in Escherichia coli.* Biotechnology and Bioengineering, 2006. **95**(4): p. 724-733.
158. Langenhof, M., et al., *Controlled oxidative protein refolding using an ion-exchange column.* Journal of Chromatography A, 2005. **1069**(2): p. 195-201.
159. Geng, X. and C. Wang, *Protein folding liquid chromatography and its recent developments.* Journal of Chromatography B, 2007. **849**(1-2): p. 69-80.
160. Dobson, C.M., *Protein folding and misfolding.* Nature, 2003. **426**(6968): p. 884-890.
161. Dobson, C.M., *Principles of protein folding, misfolding and aggregation.* Seminars in Cell & Developmental Biology, 2004. **15**(1): p. 3-16.
162. Oliveberg, M., et al., *The changing nature of the protein folding transition state: implications for the shape of the free-energy profile for folding.* Journal of Molecular Biology, 1998. **277**(4): p. 933-943.
163. Basu, A. and S.S.J. Leong, *High productivity chromatography refolding process for Hepatitis B Virus X (HBx) protein guided by statistical design of experiment studies.* Journal of Chromatography A, (In Press).
164. Kuehner, D.E., et al., *Interactions of lysozyme in concentrated electrolyte solutions from dynamic light-scattering measurements.* Biophysical Journal, 1997. **73**(6): p. 3211-3224.
165. Oliva, A., M. Llabrés, and J.B. Fariña, *Data analysis in stability studies of biopharmaceutical drugs with isothermal and non-isothermal assays.* TrAC - Trends in Analytical Chemistry, 2011. **30**(5): p. 717-730.
166. International Conference on Harmonization, I., *Harmonized Tripartite Guideline for the Stability Testing of New Drug Substances and Products Q1A-(R2)*. 2002, IFPMA: Geneva, Switzerland.

167. Frokjaer, S. and D.E. Otzen, *Protein drug stability: A formulation challenge.* Nature Reviews Drug Discovery, 2005. **4**(4): p. 298-306.