

**CONSTRUCTION OF cDNA LIBRARIES, AND THE
SELECTION AND EXPRESSION OF PROTEINS AND
PEPTIDES INVOLVED IN HAEMOSTASIS**

By

KAREN DE BRUIN

Submitted in fulfilment of the requirements for the degree

Magister in Medical Science (M.Med.Sc)

In the Faculty of Health Sciences
Department of Haematology and Cell Biology
At the University of the Free State
Bloemfontein
South Africa

JUNE 2004

PROMOTER: Prof. S.M. Meiring
CO-PROMOTER: Prof. H. Deckmyn

ACKNOWLEDGEMENTS

Thanks and praise to our Lord in heaven for His unconditional love and grace!

I further wish to express my sincere gratitude and appreciation to the following people:

@ My supervisors, Prof. Muriel Meiring and Prof. Hans Deckmyn, for their guidance, encouragement and friendship.

@ Nienke-Nanje, Oubaas and the rest of my colleagues in the molecular laboratory for their interest and moral support.

@ All my colleagues and friends in the “Labo voor Thrombose Onderzoek” at the KULAK for their kindness, support and assistance. A special thanks to Hendrik and Karen for their guidance and tolerance.

@ The bilateral agreement between the University of the Free State (UFS) and the KU Leuven Campus Kortrijk (KULAK) in Belgium for making my studies abroad possible.

@ My family for their wonderful love, support and tolerance.

TABLE OF CONTENTS

	Page
ABBREVIATIONS	i
LIST OF FIGURES	iv
LIST OF TABLES	iv
CHAPTER 1	
INTRODUCTION	1
CHAPTER 2	
LITERATURE REVIEW	5
2.1 CDNA LIBRARIES	5
2.1.1 Introduction to cDNA libraries	5
2.1.2 cDNA library synthesis	6
2.1.3 Full-length cDNA libraries	12
2.2 PHAGE DISPLAY TECHNOLOGY	16
2.2.1 Introduction to phage display technology	16
2.2.2 The filamentous phage	18
2.2.2.1 <i>Structure</i>	18
2.2.2.2 <i>Life cycle</i>	19
2.2.2.3 <i>Phagemid cloning vectors</i>	20
2.2.3 Types of phage-display systems	23
2.2.4 Phage-displayed peptide libraries	25
2.2.4.1 <i>Random peptide libraries</i>	25
2.2.4.2 <i>Applications of phage-displayed peptide libraries</i>	26
2.2.5 Phage display of proteins	28
2.2.5.1 <i>cDNA libraries</i>	29
2.2.5.2 <i>Antibody libraries</i>	31
2.2.5.3 <i>Applications of phage-displayed proteins</i>	33
2.2.5.3.1 <i>Engineering proteins</i>	34
2.2.5.3.2 <i>Substrate-binding phage display</i>	35

2.2.6	Screening phage-displayed libraries	36
2.2.7	Recent innovations in phage display technology	40
2.2.7.1	<i>Selectively infective phage (SIP)</i>	40
2.2.7.2	<i>Landscape phage libraries</i>	41
2.2.7.3	<i>Ribosome display</i>	43
2.2.8	Applications of phage display in haemostasis	44
2.3	EXPRESSION OF CLONED GENES IN <i>ESCHERICHIA COLI</i>	51
2.3.1	Introduction to expression systems	51
2.3.2	Protein expression in <i>Escherichia coli</i>	53
2.3.2.1	<i>Expression vector components</i>	55
2.3.2.1.1	<i>Promoters</i>	58
2.3.2.1.2	<i>Transcriptional terminators</i>	59
2.3.2.2	<i>mRNA stability</i>	60
2.3.2.3	<i>Translation initiation</i>	61
2.3.3	Choice of cellular compartment for protein expression	62
2.3.3.1	<i>Cytoplasmic expression</i>	63
2.3.3.2	<i>Periplasmic expression</i>	64
2.3.3.3	<i>Extracellular secretion</i>	65
2.3.4	Host design considerations	67
2.3.4.1	<i>Fusion proteins</i>	67
2.3.4.2	<i>Molecular chaperones</i>	70
2.3.4.3	<i>Codon usage</i>	71
2.3.4.4	<i>Stress response</i>	73
2.3.5	Refolding recombinant proteins	74
2.3.5.1	<i>Inclusion body isolation, purification and solubilisation</i>	76
2.3.5.2	<i>Renaturation and refolding of the solubilised protein</i>	77
2.4	THROMBOSIS AND ANTITHROMBOTICS	79
2.4.1	The search for antithrombotic agents	79
2.4.2	Thrombin and thrombin inhibitors	80
2.4.3	Antihaemostatic compounds from haematophagous animals	81
2.4.4	Platelets, von Willebrand factor and ADAMTS-13	83

CHAPTER 3

MATERIALS AND METHODS

3.1 CONSTRUCTION OF CDNA LIBRARY	85
3.1.1 <i>Isolation of total RNA from Hippobosca rufipes</i>	85
3.1.2 <i>Generation of mRNA</i>	85
3.1.3 <i>cDNA synthesis</i>	86
3.1.4 <i>Construction of cDNA phage display library</i>	87
3.1.5 <i>Direct polymerase chain reaction (PCR)</i>	87
3.2 SELECTION OF THROMBIN-BINDING PHAGES	88
3.2.1 <i>Preparation of TG1 cultures</i>	88
3.2.2 <i>Biopanning against α-thrombin</i>	89
3.2.3 <i>Global ELISA</i>	90
3.2.4 <i>Growing and amplification of single colonies</i>	91
3.2.5 <i>Binding ELISA</i>	92
3.2.6 <i>Dilution ELISA</i>	92
3.2.7 <i>Competition ELISA</i>	92
3.2.8 <i>Thrombin Time (TT)</i>	93
3.3 CLONING OF THE CUB2 DOMAIN OF ADAMTS-13	93
3.3.1 <i>Amplification of the CUB2 domain</i>	93
3.3.2 <i>Preparation of the CUB2 insert</i>	95
3.3.3 <i>Preparation of the pET-26b(+) expression vector</i>	95
3.3.4 <i>Construction of recombinant plasmid</i>	96
3.3.5 <i>Direct polymerase chain reaction (PCR)</i>	97
3.3.6 <i>Plasmid preparation</i>	97
3.3.7 <i>Expression host transformation</i>	97
3.4 EXPRESSION OF RECOMBINANT PEPTIDE	98
3.4.1 <i>Expression of recombinant CUB2</i>	98
3.4.2 <i>Localisation of expressed recombinant CUB2</i>	100
3.4.3 <i>Optimisation of expression and purification</i>	102
3.4.4 <i>Large scale production of recombinant CUB2</i>	102
3.4.5 <i>Peptide purification and refolding from inclusion bodies</i>	103

CHAPTER 4	
RESULTS	
4.1 CONSTRUCTION OF CDNA LIBRARY	105
4.1.1 <i>cDNA synthesis from Hippobosca rufipes</i>	105
4.1.2 <i>Construction of cDNA phage display library</i>	105
4.2 SELECTION OF THROMBIN-BINDING PHAGES	106
4.2.1 <i>Biopanning against α-thrombin</i>	106
4.2.2 <i>Competition ELISA</i>	108
4.2.3 <i>Thrombin time (TT)</i>	109
4.3 CLONING OF THE CUB2 DOMAIN OF ADAMTS-13	110
4.3.1 <i>Amplification of CUB2</i>	110
4.3.2 <i>Transformation of the non-expression host</i>	112
4.3.3 <i>Transformation of the expression host</i>	113
4.4 EXPRESSION OF RECOMBINANT PEPTIDE	114
4.4.1 <i>Expression of recombinant CUB2</i>	114
4.4.2 <i>Localisation of expressed recombinant CUB2</i>	114
4.4.3 <i>Large scale expression and refolding of recombinant CUB2</i>	115
CHAPTER 5	
DISCUSSION	117
CHAPTER 6	
ABSTRACT	123
CHAPTER 7	
ABSTRAK	125
REFERENCES	127

ABBREVIATIONS

ADAMTS	a disintegrin-like and metalloprotease with thrombospondin type-1 motifs
ALV	avian leukaemia virus
bp	nucleotide base-pair
BSA	bovine serum albumin
CBB	Coomassie Brilliant Blue
cDNA	complementary DNA
CDR	complement-determining regions
CUB	complement subcomponents C1r/C1s, Uegf, Bmpl
DNA	deoxyribonucleic acid
dNTP	deoxynucleoside triphosphate
ds	double-stranded (DNA)
DTE	dithioerythritol
DTT	dithiothreitol
<i>E. coli</i>	<i>Escherichia coli</i>
EDTA	ethylenediamine tetra-acetic acid
ELISA	enzyme-linked immunoabsorbent assay
EPO	erythropoietin
EST	expressed-sequence-tag
Fab	variable sequence fragment of immunoglobulin
Fv	variable region fragments
g3p	gene for protein-3 / pIII
g6p	gene for protein-6 / pVI
g7p	gene for protein-7 / pVII
g8p	gene for protein-8 / pVIII
g9p	gene for protein-9 / pIX
GdmCl	guanidinium chloride
GTP	guanosine-5'-triphosphate
His	L-histidine
HRP	horse radish peroxidase
HSP	heat shock protein
IB	inclusion body

IPTG	isopropyl- β -D-1-thiogalactopyranoside
kb	kilobase
kDa	kilodalton
LB-agar	Luria-Bertani agar
Mo-MLV	Moloney strain of murine leukaemia virus
MQ water	Millipore milli-Q water
mRNA	messenger RNA
OD	optical density
oligo(dT)	oligo-thymidine
OPD	<i>ortho</i> -phenylenediamine
ORF	open reading frame
Ori	origin of replication
PBS	phosphate buffered saline
PCR	polymerase chain reaction
PEG	polyethylene glycol
pfu	plaque forming unit
PNPase	polynucleotide phosphorylase
RACE	rapid amplification of cDNA ends
RBS	ribosome-binding site
RF	replicative form
rCUB2	recombinant CUB2 domain
RNA	ribonucleic acid
RNase	ribonuclease
scFv	single-chain Fv
SD	Shine-Dalgarno
SDS	sodium dodecyl sulphate
SDS-PAGE	sodium dodecyl sulfate polyacrylamid gel electrophoresis
SEC	size-exclusion chromatography
SIP	selective infective phage
SM	skimmed milk
ss	single-stranded (DNA)
TBE	Tris-borate-EDTA
TBS	Tris-buffered saline
TBST	Tris-buffered saline with Tween

TCA	trichloroacetic acid
TCP	total cell protein
tPa	tissue plasmin activator
Tris	tris(hydroxymethyl)-amino-methane
tRNA	transfer ribonucleic acid
TT	thrombin time
TTI	tsetse thrombin inhibitor
UTR	untranslated region
V _H	variable region, heavy chain
V _L	variable region, light chain
vWF	Von Willebrand factor

List of Figures

	Page
Figure 2.1 Preparation and screening of a cDNA library	7
Figure 2.2 Full-length cDNA cloning methods	15
Figure 2.3 Schematic representation of the Ff phage	19
Figure 2.4 M13 phagemid display vector system	22
Figure 2.5 Types of phage display systems	24
Figure 2.6 Screening phage-displayed libraries	37
Figure 2.7 Control elements of the pET system	57
Figure 4.1 Restriction endonuclease digestion of cDNA	105
Figure 4.2 Direct PCR on randomly picked single TG1 colonies	106
Figure 4.3 Global ELISA of four biopanning rounds against α -thrombin	107
Figure 4.4 Dilution ELISA of the six strongest thrombin-binding phage colonies	108
Figure 4.5 Competition ELISA of colonies 46 and 48 performed in the presence of different hirudin concentrations	109
Figure 4.6 Prolongation of thrombin times	110
Figure 4.7 Schematic representation of insert and vector design	111
Figure 4.8 Restriction endonuclease digestion of vector and insert	112
Figure 4.9 Direct PCR on transformed TOP10 colonies	113
Figure 4.10 Direct PCR on randomly picked BL21 single colonies	113
Figure 4.11 Induction of protein expression by addition of IPTG	114
Figure 4.12 Expression of recombinant CUB2 in different cellular compartments	115
Figure 4.13 Refolding of recombinant CUB2	116

LIST OF TABLES

	Page
Table 2.1 Thrombin inhibitors derived from haematophagous animals	82

CHAPTER 1

INTRODUCTION

Cardiovascular disease remains the leading cause of mortality and morbidity in industrialised countries. The thrombotic complications of atherosclerosis, such as acute coronary events and ischemic stroke, can be fatal. Patients who survive such events have a far greater risk of future cardiovascular events. This huge medical need cries out for improved novel anticoagulants, antiplatelet agents/drugs, and profibrinolytic agents. These agents must successfully respond to the medical need by providing safe, effective, and easily administered treatments that have little, if any, drug and food interactions and that require minimal monitoring (Hirsh, 2003).

From early on, the discovery of new drugs involved a trail-and-error approach using naturally derived materials and substances. The first half of the twentieth century witnessed systematic pharmacological evaluations of both natural and synthetic compounds. With the exponential development of molecular biology on the one hand, and computer technology on the other, from 1980 onwards drug discovery has seen many changes. Cloning of genes has led to the development of methodologies for specific receptor-directed and enzyme-directed drug discoveries. Advances in recombinant DNA and transgenic technologies have enabled the production of endogenous biomolecules as new drugs (Kaul, 1998).

Over the last couple of decades haemostasis and thrombosis research has emerged as an interdisciplinary field that draws on biochemistry, physiology, structural biology, clinical chemistry, cell biology, molecular genetics, recombinant DNA technology and biochemical engineering to unravel the complex processes involved in maintaining the delicate haemostatic balance. The evolution of molecular biology techniques in particular has taken place at a

staggering pace and the introduction of powerful new methodologies has revolutionised haemostasis research.

The past century has witnessed a breathtaking array of discoveries in the biological sciences, in particular in the general area of molecular biology. The concepts of genetic engineering and recombinant DNA are often erroneously viewed as a relatively new field of scientific pursuit, but in fact the techniques that are in use today are the result of a series of landmark discoveries that were made over a period spanning more than 125 years.

Mullis introduced the polymerase chain reaction (PCR), a novel method of amplifying large amounts of a specific DNA fragment starting with very small amounts of source DNA, in the late 1980s (Mullis & Faloona, 1987). PCR has revolutionized modern biology and has widespread applications in the areas of forensics, diagnostics, and gene expression analysis. The 15-year Human Genome Project formally began in 1990 (Watson, 1990), and today the scientific world is reaping the benefits, drawing on information stored in huge databases, most of which are accessible by the public. Many genes of particular interest have already been discovered in recent years due to the rapid progress in genomic sequencing and cDNA library characterisation. The challenge facing this enormous amount of genomic information is to now elucidate the cellular localizations and biological functions of the predicted proteins and to identify those that can serve as targets for therapeutic intervention.

In the 1980s, combinatorial mutations and display of peptide or small protein libraries became important tools for production of molecules on filamentous bacteriophage mutated tips. Phage display technology is a powerful *in vitro* selection technique, in which a peptide or protein is genetically fused to a coat protein of a bacteriophage (Smith, 1985). The phage display approach is based on key developments in molecular biology. It combines (i) the ability to functionally express gene fragments in *E. coli* (Skerra & Plückthun, 1988), (ii)

molecular techniques such as PCR technology to create very large peptide or antibody gene repertoires (Orlandi *et al.*, 1989), and (iii) the ability to express protein fragments on the surface of bacteriophages (McCafferty *et al.*, 1990).

The expression of recombinant proteins is an important step toward elucidating many genes discovered through genomic sequencing projects and also for validating gene targets. Both prokaryotic and eukaryotic heterologous expression systems are employed for the production of recombinant proteins because of the convenience of manipulation of these systems, and their ability to achieve levels of production many times higher than the native source of the protein. Today, the field of heterologous expression is in a mature phase, with most of the progress being incremental rather than ground breaking.

The biotechnology revolution of the 1980s brought along the ability to extract DNA encoding anticoagulant proteins isolated from various haematophagous (blood-feeding) animals and insects, and to use this DNA to transfect cells that produce the target protein in large quantities (Hirsh & Weitz, 1999). Indeed, the screening of expression libraries from haematophagous animals has a revived role in the exploration of genomic data for the development of novel anticoagulants and haemostatic regulators (Urata *et al.*, 2003).

The evolutionary success of haematophagous animals depends on their ability to maintain blood in a fluid state during acquisition and storage in the gut canal during digestion. To this end, they secrete compounds that block the haemostasis of the host to prevent blood from clotting (Basanova *et al.*, 2002). Haematophagous animals and their mechanisms to interfere with the constituents of the coagulation cascade have been studied since the end of the nineteenth century (Dodt *et al.*, 1996). Anticoagulant activity in the salivary glands of haematophagous insects was demonstrated in 1914 using simple coagulation tests (Cornwall & Patton, 1914). Antihaemostatic compounds isolated from these animals include inhibitors of vascular-platelet haemostasis,

inhibitors of the activation of intrinsic mechanisms of blood coagulation and proteins of the prothrombinase complex, and regulators of fibrin formation, including inhibitors of thrombin and FXIIIa, fibrinolytic enzymes and activators of fibrinolysis (Arocha-Pinango *et al.*, 1999) In recent years much interest has arisen in these substances for their potential clinical use in treating thromboembolic diseases. Numerous antithrombotic proteins have been purified, their cDNAs cloned, and the protein expressed in heterologous systems for subsequent analyses (Ribeiro, 1995).

In this study, the use of molecular biology techniques in the quest to better understand and control the intricate processes of thrombosis, and search for novel antithrombotic compounds, was demonstrated. In the first section, a cDNA library was constructed from the haematophagous louse fly *Hippobosca rufipes*, and phage display technology was employed to select for possible antithrombotic agents specifically directed against thrombin. Different single phage colonies were picked, grown, and their ability to bind to and inhibit thrombin was tested. In the second section, a recombinant peptide was constructed by cloning the second CUB-domain of the human metalloprotease ADAMTS-13 into an *Escherichia coli* expression vector. The recombinant CUB2-domain was expressed, isolated and purified to serve as an important tool in the further analysis of the ADAMTS-13. Biotechnology tools such as molecular cloning, nucleic acid hybridisation, gel electrophoresis, restriction enzyme digestion, PCR, biotinylation, ELISA (enzyme-linked immunoabsorbent assay), biopanning, etc. were used.

Harnessing the many tools and techniques produced by the ongoing biotechnology explosion, allows the researcher to apply a wealth of new information to both hereditary and acquired haemostatic and thrombotic diseases, and to increase understanding of the biochemical processes involved in these disorders.

CHAPTER 2

LITERATURE REVIEW

The main focus of this dissertation is the implementation of molecular biology methods in the interdisciplinary field of haemostasis research. Employing the techniques of cDNA library construction, phage display and protein expression in thrombosis research, in particular, will be examined in detail. The literature review will thus commence by reviewing the aforementioned techniques, and end with a brief introduction to thrombosis and antithrombotics.

2.1 CDNA LIBRARIES

2.1.1 Introduction to cDNA libraries

In any attempt to study the gene function of an organism, the incorporation of a cDNA-based approach is unavoidable. Indeed, one of the fundamental tools of molecular biology is the enzymatic conversion of poly(A)⁺ messenger RNA (mRNA) to double stranded (ds) DNA, and the insertion of this DNA into both prokaryotic and eukaryotic vectors (Kimmel & Berger, 1987).

Complementary DNA (cDNA) is the reverse transcriptase product of mRNA and represents the coding sequence of all transcribed genes at the time of mRNA isolation (Kimmel & Berger, 1987). In order to isolate and study a particular eukaryotic gene, transcribed at a certain time or under certain conditions, a cDNA clone is usually isolated. The general method involves the construction of a cDNA library representing the mRNA population using poly(A)⁺ RNA, extracted from the appropriate tissue or cell type, followed by the identification of the cDNA clone of interest. The cDNA clone is selected from the library by screening with synthetic oligonucleotide probes, cDNA probes representing differentially expressed mRNAs, or antibody probes (Huynh *et al.*, 1988). Many techniques for the construction of ds cDNA from mRNA are available, some of which are so

advanced that it is possible to construct a cDNA library from a single cell (Sambrook & Russell, 2001).

The structural features of mRNA play an important, and often limiting, role in transcribing ds cDNA. A common feature of all eukaryotic mRNAs is the presence of a cap structure at the 5'-end (Furuichi & Miura, 1975) and a poly(A) stretch at the 3'-end (Darnell *et al.*, 1971). The cap is a 5'-terminal 7-methylated GTP, attached to the first nucleotide of the mRNA through two pyrophosphates, which is added early during transcription of RNA polymerase II genes in the nucleus (Furuichi & Miura, 1975). The cap is required for several steps of mRNA biogenesis which include protecting the mRNA against 5'-exonucleases, stimulating translation, stimulating precursor mRNA splicing, enhancing nucleocytoplasmic transport, and facilitating 3'-end processing (Sonnenberg, 1988).

2.1.2 cDNA library synthesis

Gubler and Hoffman developed the original method for generating cDNA libraries by combining classical first-strand synthesis with RNase H DNA polymerase I mediated second-strand synthesis. This allows the conversion of first-strand to second-strand ds cDNA by RNA-primed nick-translation without significant loss of sequence information (Gubler & Hoffman, 1983).

Various modifications followed over the years to improve the method's versatility (Rutledge *et al.*, 1988), but the procedure to synthesise cDNA has not been extensively changed, with the possible exception of the construction of uniform-abundance (normalised) cDNA libraries (Patanjali *et al.*, 1991). A flowchart illustrating processes involved in the preparing and screening of a cDNA library is given in figure 2.1.

RNA-dependent DNA polymerase reverse transcriptases use poly(A)⁺ RNA or mRNA as template to synthesise the first-strand of cDNA. Different forms of

reverse transcriptases are commercially available including avian reverse transcriptase, purified from particles of an avian leukaemia virus (ALV), and murine reverse transcriptase, derived from the Moloney strain of murine leukaemia virus (Mo-MLV) (Gerard, 1998).

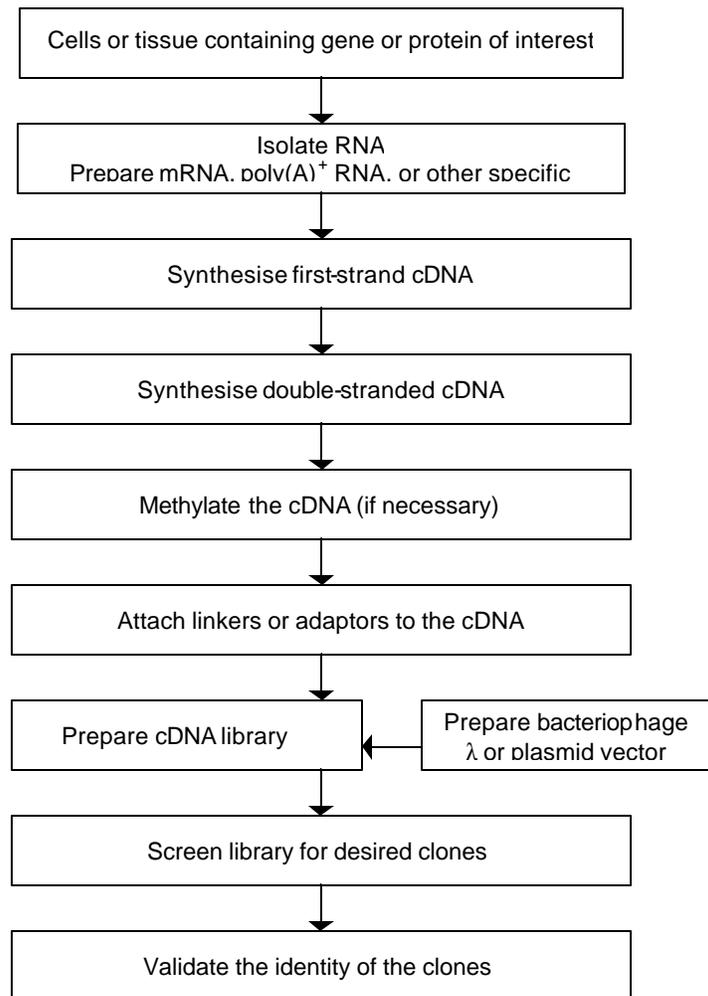


Figure 2.1. Preparation and screening of a cDNA library (Sambrook & Russell, 2001)

A major obstacle in cDNA synthesis is the strong secondary structure of mRNA, which cause the reverse transcriptase to stop the synthesis, and subsequently to be released from the hybrid mRNA/incomplete cDNA. Both denaturing the sample before the cDNA synthesis reaction, and increased temperature reaction have been employed to overcome problems associated with the secondary structure of mRNA. However, attempts to overcome the problem by heat

destabilisation, or treatment of mRNA with methylmercury hydroxide were not always successful, especially to obtain full length cDNA from very long transcripts (Carninci *et al.*, 1996).

In addition, all reverse transcriptases have high error rates of polymerisation, causing the enzyme to stumble. Therefore, even in the absence of RNA secondary structures and RNase H activity, the synthesis of full-length from long mRNA transcripts still remain a challenge (Gerard, 1998). In an attempt to overcome these limitations, a recent study demonstrated the advantages of coupling a 3'→5' exonuclease function during reverse transcription. The incorporation of a proofreading activity, when used in conjunction with denaturant buffers and RNase H-deficient reverse transcriptases, was shown to successfully generate full-length cDNAs (Hawkins *et al.*, 2003).

A wide range of technical and theoretical advances over the last couple of decades has enabled cDNA libraries to be constructed from small quantities of mRNA. A variety of reliable methods have been developed to identify cDNA clones corresponding to extremely rare species of mRNA (Sambrook & Russell, 2001). Methods exist to generate and amplify cDNA libraries, and amplify genes and single RNA transcripts from a single cell without cloning (Jena *et al.*, 1996). When isolating RNA or mRNA from a small number, or even individual cells, certain techniques and precautions can be taken to establish cDNA libraries of useful size. These include extracting RNA immediately from freshly harvested material and using a scaled-down version of standard RNA isolation protocols (Brady *et al.*, 1990), using total RNA rather than poly(A)⁺ mRNA as template for first-strand cDNA synthesis (Lambert & Williamson, 1993), using a single buffer for synthesis of first-strand and second-strand cDNA, and addition of linkers (Brady *et al.*, 1990), and amplifying either first-strand or double-strand cDNA by PCR (McCarrey & Williams, 1994).

The primers for first-strand cDNA synthesis can be specifically designed to hybridise to a particular target gene, or they can bind generally to all mRNAs. Different primers used for first-strand cDNA synthesis include oligo(dT) primers, primer-adaptors, plasmid-linked primers, and random hexamer primers (Kimmel & Berger, 1987). Oligo(dT) primers are 12-18 nucleotides in length, bind to the endogenous poly(A) tails at the 3'-terminus of eukaryotic cellular mRNAs, and are often used as a universal primer for conventional first-strand cDNA synthesis (Resuehr & Spiess, 2003). Primer-adaptors contain a homopolymeric oligo(dT) tail at the 3'-terminus to prime first-strand synthesis, and an additional restriction site at the 5'-terminus. This allows the DNA to be ligated to a vector before second-strand synthesis, or alternatively enables a second primer-adaptor to prime the synthesis of second-strand cDNA (Coleclough & Erlitz, 1985).

Okayama and Berg developed a method in which priming of first-strand synthesis is carried out by a oligo(dT) tail that is covalently attached to a plasmid (Okayama & Berg, 1983). Libraries constructed using the original lengthy and demanding protocol are generally moderate in complexity and have a high proportion of full-length cDNAs. The classical method has been simplified over the years, for example by the introduction of a short synthetic oligonucleotide as a second-strand adaptor (Boel *et al.*, 1991). These newer procedures, that also use asymmetrically tailed plasmid to prime synthesis of first-strand cDNA, are much simpler and yield libraries of higher complexity (Spickofsky & Margolskee, 1991).

Random hexanucleotides, which are capable of priming cDNA synthesis at numerous points along the length of the RNA template, generate fragmentary copies of the entire population of RNA molecules. A benefit of using random hexamer primers is that they are capable of avoiding possible secondary mRNA structures such as loops and stems (Hawkins *et al.*, 2003). These secondary structures of mRNA can often cause difficulties for the reverse transcriptase,

causing the enzyme to stall and end its synthesis well ahead of the 5'-end (Brooks *et al.*, 1995).

Rapid amplification of cDNA ends (RACE) has been developed as a cloning strategy to overcome many of the difficulties encountered in obtaining full-length cDNA clones of low-abundance mRNAs. In essence, cDNAs are generated by using PCR to amplify copies of the region between a single point in the transcript and the 3'- or 5'-end. The minimum information required for this amplification is a single short stretch of sequence within the mRNA to be cloned (Frohman *et al.*, 1988). If the use of random primers is not successful, alternative methods, such as 5'-RACE and 3'-RACE, may be used to generate cDNA clones containing 5'- and 3'- terminal regions of mRNAs (Schaefer, 1995).

Replacement synthesis of second-strand cDNA was introduced by Okayama and Berg and modified by Gubler and Hoffman. The primers for second-strand synthesis are created by RNase H, which introduces nicks into the RNA moiety of the cDNA/mRNA hybrids. *E. coli* DNA polymerase I then extends the 3'-hydroxyl termini of these RNA primers, using the first-strand cDNA as a template, and replacing the remaining segments of mRNA in the cDNA/mRNA hybrid with the newly synthesised second-strand cDNA (Okayama & Berg, 1982; Gubler & Hoffman, 1983). Residual nicks in the DNA/DNA hybrid are then repaired by *E. coli* ligase, and frequently T4 DNA polymerase or a thermostable polymerase such as *Pfu* is added to polish the frayed termini of the completed double-stranded DNAs (Gerard & D'Alessio, 1933).

An alternative method for the synthesis of second-strand cDNA involves combination of the switch mechanism at the 5'-end of RNA templates (SMART) with reverse transcription, followed by PCR. A comparative study, however found that conventional second-strand cDNA synthesis is the better method for amplification of limited amounts of RNA (Wang *et al.*, 2003).

After cDNA library synthesis, most cDNAs were originally cloned by adding complementary homopolymeric tails to the ds cDNA and to a plasmid vector. The vector and the cDNA were then joined by hydrogen bonding between the complementary homopolymers to form open circular hybrid molecules capable of transforming *E. coli* (Peacock *et al.*, 1981). Although this strategy was used with success for many years, homopolymeric tailing has now fallen into disuse. Today the cloning of cDNA into a vector is facilitated by the addition of various tails, linkers, or adaptor sequences to the ends of cDNAs. In addition to equipping the termini of cDNA for cloning, linkers can also serve as binding sites for PCR primers, for use in the construction of large cDNA libraries from very small amounts of cDNA, where a PCR step is added (McCarrey & Williams, 1994).

The development of efficient ways to synthesise cDNA and the growing availability of linkers, adaptors, methylases, and packaging strategies, has made it possible to use bacteriophage lambda (λ) as a cloning vector, benefiting from the high efficiency and reproducibility of *in vitro* packaging of bacteriophage λ DNA into infectious particles. The resulting libraries are often large enough to be screened directly without amplification. Alternatively, they may be amplified, stored indefinitely without loss of viability, and screened with either nucleic acid probes, antibodies, or other ligands, depending on the particular vector (Niwa *et al.*, 2000).

Intracellular expression technology is often coupled with filter screening, which presents several problems: (i) the number of clones that can be screened is limited by the number of plaques or colonies that can be fitted on each filter, which makes it difficult or impossible to completely survey very large libraries; (ii) large amounts of screening reagents are required; and (iii) hydrophobic peptides which tend to aggregate will form inclusion bodies and will not be available for ligate recognition (Santi *et al.*, 2000). However, with the advent of phage display technology, affinity selectable biological repertoires have become the preferred

system to identify proteins and ligands (Dunn, 1996). This will be discussed in detail in section 2.2.5, in the literature review of phage display technology.

The selection of cDNA repertoires displayed on the surfaces of phages allow the rapid isolation of interacting partners overcoming slower screening procedures, and have the advantage that the conditions for selection can be completely controlled. Interacting molecules with either modest or high affinities can be recovered, which has been reported to span a greater range than that which can be obtained with simple filter screening (Hufton *et al.*, 1999). Despite initial difficulties in displaying cDNA libraries on filamentous phages, such as the presence of stop codons in full-length cDNA, and problems with direct fusion to the C-terminus of coat proteins, methods have been developed to display oligo(dT)-primed cDNA libraries successfully (Crameri & Walter, 1999).

2.1.3 Full-length cDNA libraries

Since full-length cDNAs carry complete protein coding sequences and UTRs, they are indispensable for the identification of genes and for the determination of primary protein structure. In particular, full-length cDNAs represent a valuable resource for functional gene studies (Draper *et al.*, 2002).

Several efforts for large-scale sequencing of cDNA libraries are in progress, in which most data are generated by single-pass sequencing of randomly selected cDNA clones through expressed-sequence-tag (EST) projects (Adams *et al.*, 1991). The EST data collection is enormous, and ESTs are used in many genetic studies and gene mapping projects. ESTs are also used in genomic sequencing projects to identify splicing sites and overlapped transcription units. However, the usefulness of EST clones are limited, because many EST clones lack the complete sequences of mRNAs, they cannot be used to reveal the primary structures of entire genes and encoded proteins. An alternative to gene discovery strategies based on ESTs followed by cloning of individual full-length cDNAs, is to construct full-length-cDNA libraries and to sequence the libraries.

These full-length clones not only serve as a resource for functional analysis but also give valuable information such as transcriptional start sites (Suzuki *et al.*, 2002).

Traditionally, generating full-length cDNA libraries presented two major technical problems. The first is reduced efficiency, mainly due to the limited processivity of reverse transcriptase and the stalling induced by the secondary structure of mRNAs. The second limitation is the lack of an efficient technique for selecting only full-length cDNA. Usually, due to the reduced representation of full-length clones, several rounds of screening are needed to select the cDNAs carrying the complete sequence (Carninci *et al.*, 1996).

In addition, when a full-length clone is not obtained, a researcher must obtain the 5'-end of the message by further screening a random primed library or using methods such as primer extension (Thompson *et al.*, 1979) or RACE (Frohman *et al.*, 1988). Neither of these methods provides a full-length message that is useful for further research and these screening procedures are often inefficient, costly, and time consuming (Sugahara *et al.*, 2001). Several methods for producing libraries that are enriched for full-length clones have been reported. These methods, employing the cap structure to select full-length cDNAs, are oligo-capping (Maruyama & Sugano, 1994), Capfinder-SMART technology (Zhu *et al.*, 2001), CAPture (Edery *et al.*, 1995) and CAP trapper (Carninci *et al.*, 1996). Figure 2.2 illustrates oligo-capping, CAPture and biotinylated CAP trapper.

In oligo-capping, bacterial alkaline phosphatase (BAP) is used to remove phosphates from the 5'-ends of uncapped RNA molecules leaving a hydroxyl group in their place, while the cap structure on full length RNAs is unaffected. Tobacco acid pyrophosphatase (TAP) is then used to remove the cap, leaving a single phosphate in its place. An oligoribonucleotide is then ligated to the RNA molecules using RNA ligase. Because this enzyme requires a 3'-hydroxyl

acceptor and a 5'-phosphate donor for substrate ligation it will not add the oligonucleotide to the partially degraded RNAs that lack a 5'-phosphate. Once the oligoribonucleotide mRNA has been established, a primer homologous to it can then be used to create full-length enriched libraries, or 5'-end enriched libraries through PCR amplification (Maruyama and Sugano, 1994).

With the original oligo-capping method, it was, however, not possible to construct a high-quality full-length cDNA library using small amounts of mRNA as starting material. Consequently, an improved oligo-capping method was developed using total RNA instead of mRNA as starting material. The large reservoir of RNA seems to act as a carrier and protects the smaller amount of mRNA from degradation. The remaining mRNA can be purified from the total RNA, directly used as a template for first-strand cDNA synthesis, and amplified by PCR (Oh *et al.*, 2003).

The use of PCR amplification to obtain a reasonable number of clones may, however, lead to selective amplification of some populations of clones, resulting in a strongly biased library in which rare or long cDNAs can be lost. To eliminate such drawbacks, the method has been modified by omitting the use of PCR (Kato *et al.*, 1994).

The CAPture (cap retention) procedure uses an affinity selection scheme which allows mRNAs to be purified via the 5'-cap structure. After RNase A treatment of the first-strand cDNA/mRNA hybrid, only the full-length cDNAs are selected by the cap-binding protein, since RNase A does not remove the RNA near the cap structure if it is protected by a full-length first-strand cDNA. CAPture can be used to enrich for clones containing the authentic mRNA 5'-end, as well as to facilitate identification of sites of transcription initiation (Edery *et al.*, 1995).

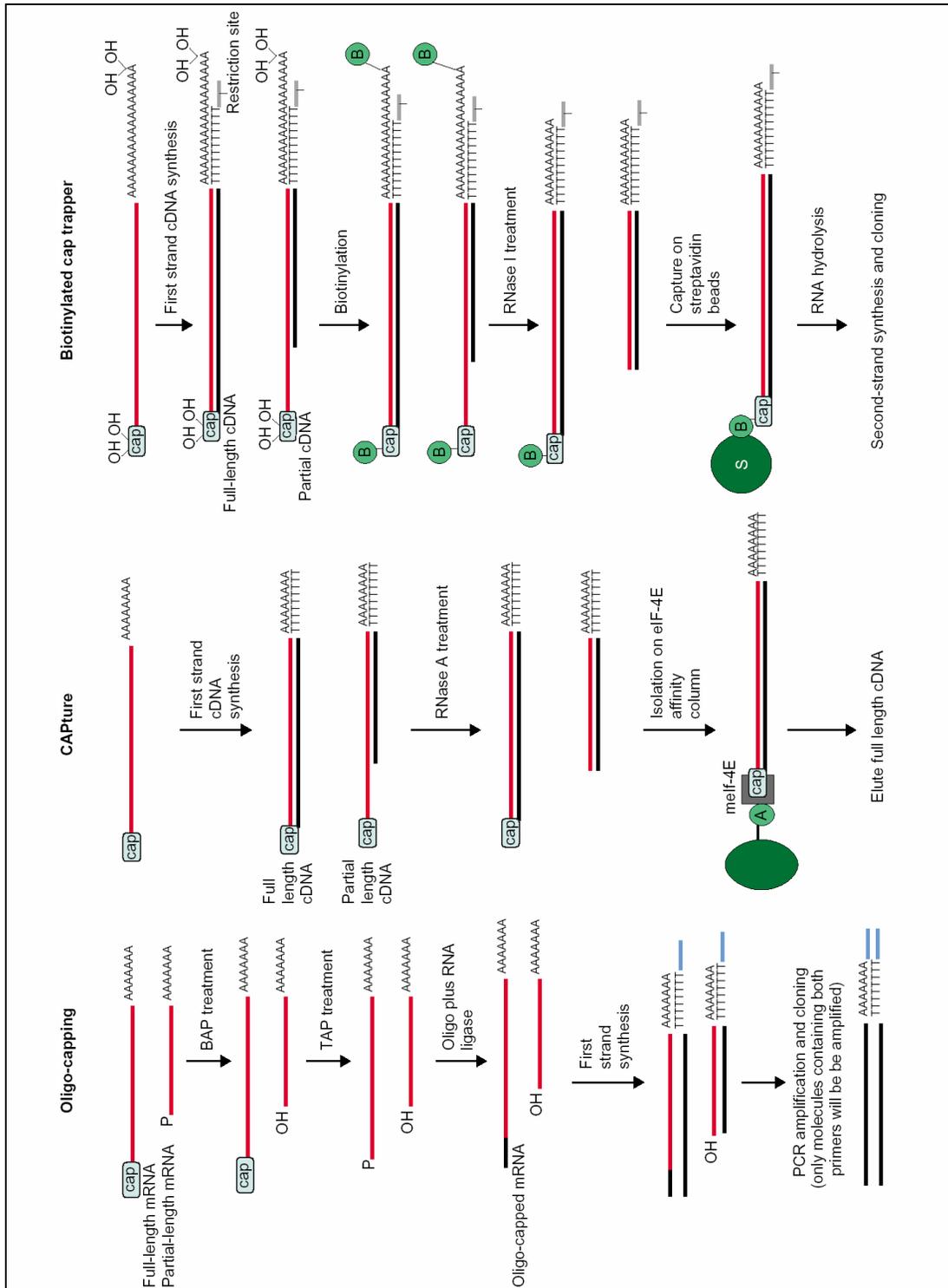


Figure 2.2. Full-length cDNA cloning methods (Bashiardes & Lovett, 2000)

The CAP trapper strategy is similar to CAPture in that it targets the retention of the 5'-termini of the mRNA/cDNA hybrid, and enables full-length cDNA to be enriched (Bashiardes & Lovett, 2000). In this method the cap structure is chemically labelled with a biotin group, and by using streptavidin coated magnetic beads, only the full-length first-strand cDNA/mRNA hybrids are selectively recovered after RNase I treatment. CAP trapper allows the preparation of high-content full-length cDNA libraries, even from relatively small quantities of tissues or early embryos, with no bias in representation since no PCR amplification step has been introduced (Carninci *et al.*, 1996).

A potential disadvantage of the original biotinylated CAP trapper protocol was the exposure of mRNA to chemical and enzymatic attacks during biotinylation of the cap structure, before first-strand cDNA synthesis and selection of full-length cDNA by biotinylated cap (Carninci *et al.*, 1996). Consequently, the protocol for cap structure biotinylation and capture of full-length cDNA was improved to allow long cDNAs to be cloned more efficiently. By performing biotinylation on the mRNA/cDNA hybrid produced by the first-strand cDNA synthesis reaction, the mRNA remains protected from chemical and enzymatic degradation, making it possible to select full-length cDNAs of longer average size (Carninci *et al.*, 1997). To improve complexity even more, most CAP trapper libraries use size-fractionation or normalisation/subtraction (Carninci *et al.*, 2000). On average, CAP trapper cDNA libraries show 2 or 3-fold higher gene discovery than both oligo-capping and Capfinder libraries (Sugahara *et al.*, 2001).

2.2 PHAGE DISPLAY TECHNOLOGY

2.2.1 Introduction to phage display technology

The ability to display peptides and proteins on the surface of the filamentous bacteriophage M13 has had a major impact on the fields of immunology, cell biology and protein engineering. George P. Smith first published the phage display concept in 1985 (Smith, 1985) as a powerful method for selecting and

engineering proteins and polypeptides with desired binding specificity. Phage display constructs of genetically tagged peptides, proteins or protein fragments allow researchers to convert pools of combinatorial nucleotides, mRNAs or fragmented genomes into populations of viruses that contain the nucleotides coding for the elements that are displayed on their viral surfaces (Benhar, 2001).

Since 1990, the scale and scope of phage display has rapidly evolved and has become a widely used technology in life sciences. Today, natural and synthetic peptides, proteins and protein domains, synthetic antibodies, and single-chain Fv (scFv) and Fab antibody libraries can be displayed. The success of phage display is due to two main reasons, namely the linkage between genotype and phenotype allowing screening of large libraries based on the power of affinity selection, and the possibility to construct large diverse synthetic or natural combinatorial libraries (Rosander *et al.*, 2002)

The enormous success of M13 phage display has prompted the development of numerous alternative display systems. These include systems that utilize other *E. coli* specific phages, such as lambda (λ) phage (Santini *et al.*, 1998) and T4 phage (Ren and Black, 1998), and also systems that use eukaryotic viruses (Possee, 1997). In addition, polypeptides have been displayed on the surfaces of bacteria and yeast (Georgiou *et al.*, 1997). Although these alternative systems have proven advantageous in special applications, M13 phage display remains the dominant technology (Sidhu, 2000).

Although originally mainly employed in discovering high-affinity ligands, phage display technology has in recent years been developed beyond its usage as a ligand-binding tool, to find application in various aspects of therapeutic and diagnostic areas. Phage display is playing an increasingly important role in the functional genomics area, in which ligands or antibody fragments are crucial in determining the functions of the hundreds and thousands of genome-derived

proteins and deciphering various therapeutically important pathways (Willats, 2002).

2.2.2 The filamentous phage

The filamentous phage (*Inovirus*) constitute a large family of bacterial viruses that infect a variety of Gram-negative bacteria. The best characterised are the very similar M13, fd and f1 phages, that infect *E. coli* via F-pili (Russel *et al.*, 1997). The relative simplicity of these viruses and the ease with which they can be genetically manipulated have made them fruitful models to study macromolecular structure and interactions (Sidhu, 2001).

2.2.2.1 Structure

The M13 phage particles are rods about 6-10 nm in diameter and 800-2000 nm in length. The viral particle consists of a single-stranded (ss), closed circular DNA core surrounded by a protein coat. Prior to virus assembly, the coat proteins are fixed in the bacterial membrane by transmembrane domains. During assembly, viral DNA is extruded through the membrane and concomitantly enveloped by coat proteins (Russel, 1991). The particle tube formed by several thousand copies of the major coat protein, protein-8 (g8p, pVIII) covers the length of the particle. Four minor coat proteins are present at about 5 copies per particle. The one end of the particle is capped by protein-7 (g7p, pVII) and protein-9 (g9p, pIX), while protein-3 (g3p, pIII) and protein-6 (g6p, pVI) cap the other end (Marvin, 1998). The basic structure of a filamentous phage is illustrated in figure 2.3.

All five coat proteins contribute to the structural stability of the phage particle, but g3p plays an additional important role in host cell recognition and infection. Consequently, g3p is the largest and most complex of the coat proteins and it contains three distinct domains (Armstrong *et al.*, 1981). The N-terminal domain initiates translocation of the viral DNA into *E. coli* during infection, while the second domain confers host cell recognition by binding to the F-pilus on the *E.*

coli surface (Jakes *et al.*, 1988). The C-terminal domain interacts with other phage coat proteins, and is thus responsible for the integration of g3p into the phage coat (Rakonjac *et al.*, 1999).

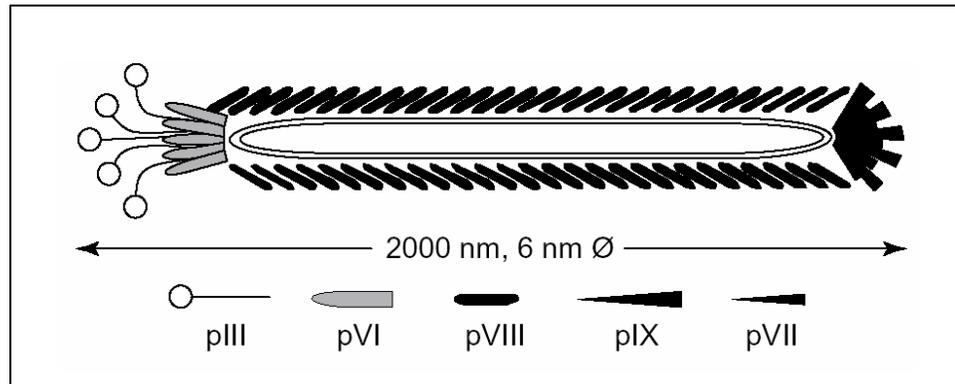


Figure 2.3. Schematic representation of the Ff phage (Konthur & Walter, 2002)

2.2.2.2 Life cycle

Filamentous phages do not produce a lytic infection in *E. coli*, but rather induce a lysogenic state in which the infected bacteria produce and secrete phage particles without undergoing lysis. Infection is initiated by the attachment of the phage g3p to the F pilus of a male *E. coli* (Caro & Schnös, 1972). When the circular phage ssDNA enters the bacterium, it is converted by the host DNA replication machinery into the double-stranded plasmid like replicative form (RF). The RF undergoes rolling circle replication to make ssDNA and also serves as a template for expression of phage proteins g3p and g8p. Phage progeny are assembled by packaging of ssDNA into protein coats and extruded through the bacterial membrane into the medium (Marvin *et al.*, 1994).

Recombinant antibodies, and folded proteins, are typically expressed as g3p (but occasionally also as g6p) fusion proteins and are displayed at the tip of the M13 phage (Hoogenboom *et al.*, 1991). When screening such a library the displayed proteins bind to the antigen or ligand. These bound phages can be detected with

an HRP-labelled antibody that recognises the g8p coat proteins. Since several thousand copies of g8p exist on the phage surface, it effectively amplifies the detection signal. On the other hand, peptides may be displayed as fusions to either g3p or g8p. If peptides were fused to g8p, bound phage can be detected using monoclonal antibodies that recognise an epitope localised in the N-terminal portion of g3p (Dente *et al.*, 1994).

2.2.2.3 Phagemid cloning vectors

The two key physical elements of phage display are firstly the libraries of nucleotide sequences encoding peptides or proteins, and secondly the phage vehicles on which these sequences are expressed. The simplest way to achieve the expression of a foreign protein is simply to create a fusion between the nucleotide sequence to be expressed and a coat protein gene within the viral genome (Willats, 2002).

Using this direct approach all the copies of the chosen coat protein become fusion proteins (Winter *et al.*, 1994). This can be advantageous in terms of the numbers of expressed foreign proteins, but if the functionality of the chosen coat protein is compromised by the fusion, phage viability may be affected, especially since no wild-type versions of the coat protein are retained. This can be avoided if hybrid phages are produced carrying some versions of a given coat protein which are wild-type and some which are fused to a foreign protein. In some hybrid phage systems, the gene fusion is an additional element of the phage genome which ensures that a wild-type copy of the coat protein gene is retained and phage particles express both wild-type and fusion proteins (Sidhu, 2001).

Alternatively, hybrid phages may be created using a phagemid-based system. A phagemid vector is a plasmid that carries the origins of replications for both the M13 phage and *E. coli*, a leader sequence, appropriate multiple cloning sites, and an antibiotic-resistance gene (Mead & Kemper, 1988). The phagemid further contains an additional copy of the one coat protein that will be fused to the

polypeptide that is to be displayed. Phagemids replicate in *E. coli* as a double-stranded plasmid, but co-infection with a helper phage results in the production of single-stranded phagemid DNA, which is packaged into phage particles. The helper phage provides all the proteins necessary for phage assembly, including wild-type copies of all the coat proteins (Bass *et al.*, 1990). The resulting phage thus contains both the wild-type coat protein from the helper phage and also the fusion coat protein from the phagemid. As a result, the heterologous protein is displayed on the phage surface, while the deleterious effects of the fusion are attenuated by the presence of helper-derived wild-type coat proteins (Sidhu, 2001). The M13 phagemid display vector system is illustrated in Figure 2.4.

Hybrid display systems have enabled the development of many phage display applications and platforms that were not possible with earlier phage-based systems. With the phagemid vector system, large proteins can be readily displayed as N-terminal fusions to g3p or g8p (Bass *et al.*, 1990), and g9p (Gao *et al.*, 1999). In addition, C-terminal display has been achieved with g6p, g3p, and g8p (Fu & Sidhu, 2000). The phagemid vector system enables coupling of affinity selection, based on the display repertoires of peptides or antibody fragments, to the recovery of the packaged gene encoding that peptide or antibody. Although this system imposes a few limitations such as gene deletion and plasmid instability, it has been successfully used to isolate antibody fragments against a wide range of proteins, cell-surface markers, viruses, and parasites. Phagemid vectors also allow either the conditional display of antibody on phage, or the secretion of the antibody in the periplasmic space of *E. coli* in a form that can be easily detected through the incorporation of peptides tags such as c-myc and polyhistidine (Sidhu, 2001).

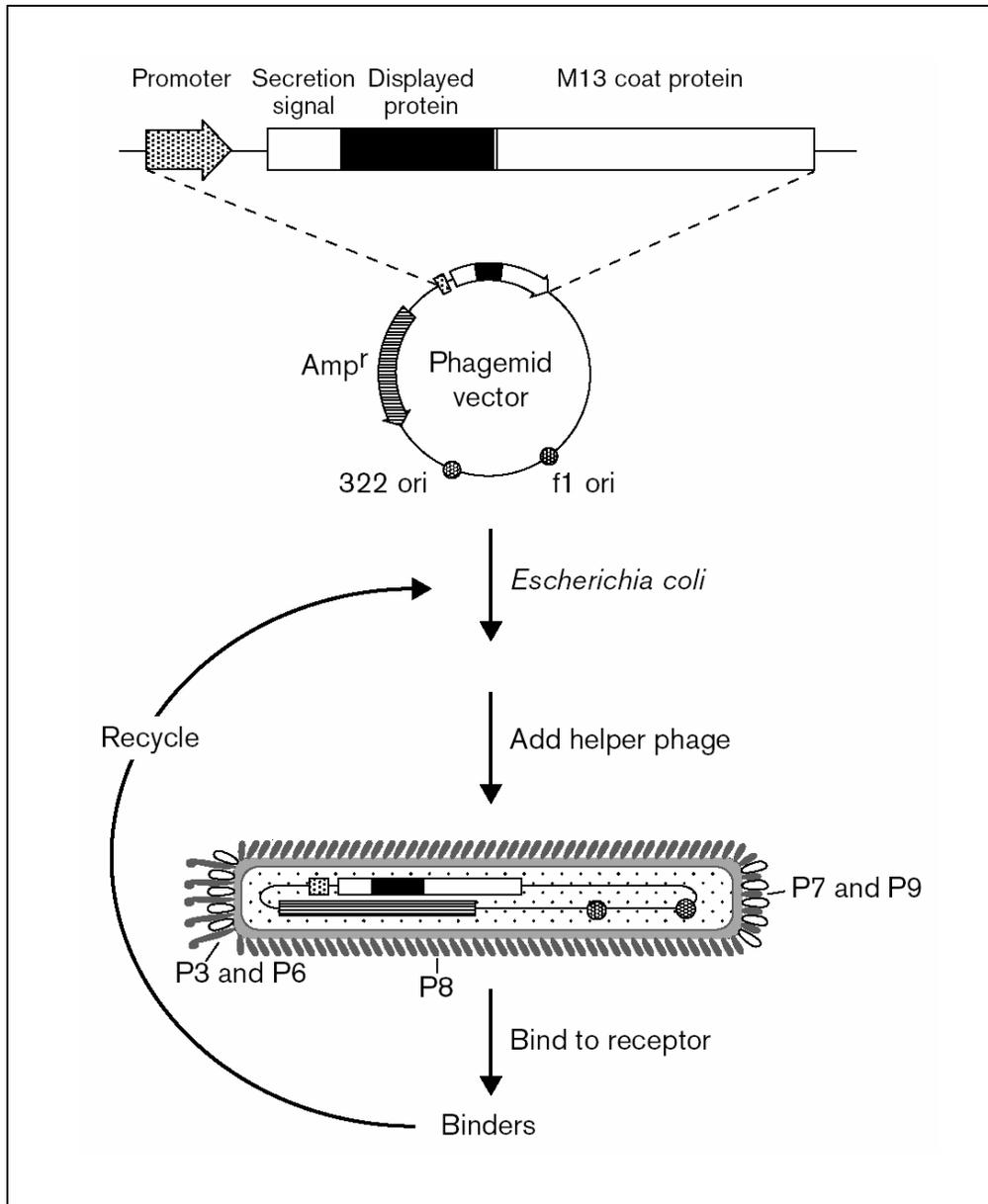


Figure 2.4. M13 phagemid display vector system (Sidhu, 2000)

A potential disadvantage of hybrid phage systems, however, is that the average number of displayed fusion proteins is reduced because of competition for incorporation into the phage particle between wild-type and fusion coat proteins (Winter *et al.*, 1994). Low valency can be used as a strategy to select for high avidity binders during biopanning selection. If coat protein functionality is not completely compromised by fusion to a foreign protein, then valency can be

increased in phagemid systems by the use of modified helper phage that lack the gene for the chosen coat protein (Rondot *et al.*, 2001). Moreover, the choice of coat protein fusion partners has recently been extended through the development of new mutant variants of coat proteins and even completely artificial coat proteins (Sidhu, 2001). The number of expressed proteins therefore depends on the coat protein chosen as a fusion partner, the display system used (phage or phagemid) and, if a phagemid system is used, the choice of helper phage. A refinement of some phage display systems is the insertion of an amber stop codon between the sequences encoding the coat protein and the displayed foreign protein. This allows a soluble, non-phage bound, version of the foreign protein to be produced if the phage are propagated in an appropriate non-suppressing strain of host bacteria (Winter *et al.*, 1994).

2.2.3 Types of phage display systems

Phage display systems can be classified according to the arrangement of the coat protein genes. Fusion to all five coat proteins has been exploited for phage display, resulting in monovalent or multivalent display of combinatorial libraries (Sidhu, 2001). Different types of phage display systems are illustrated in figure 2.5.

In a type 3 vector, there is a single phage chromosome bearing a single gene III which accepts foreign DNA inserts and encodes a single type of g3p molecule. The foreign peptide encoded by the insert is theoretically displayed on all five g3p molecules on the virion, though in practice normal proteolytic enzymes in the host bacterium often remove the foreign peptide from some or even most copies of g3p, especially if the foreign peptide is large. Similarly, type 8 and type 6 vectors display foreign peptides on every copy of g8p and g6p, respectively. Only short foreign peptides can be displayed on every copy of g8p, and even so, the peptide comprises a substantial fraction of the virion's mass and can dramatically change its physical and biological properties (Kishchenko *et al.*, 1994).

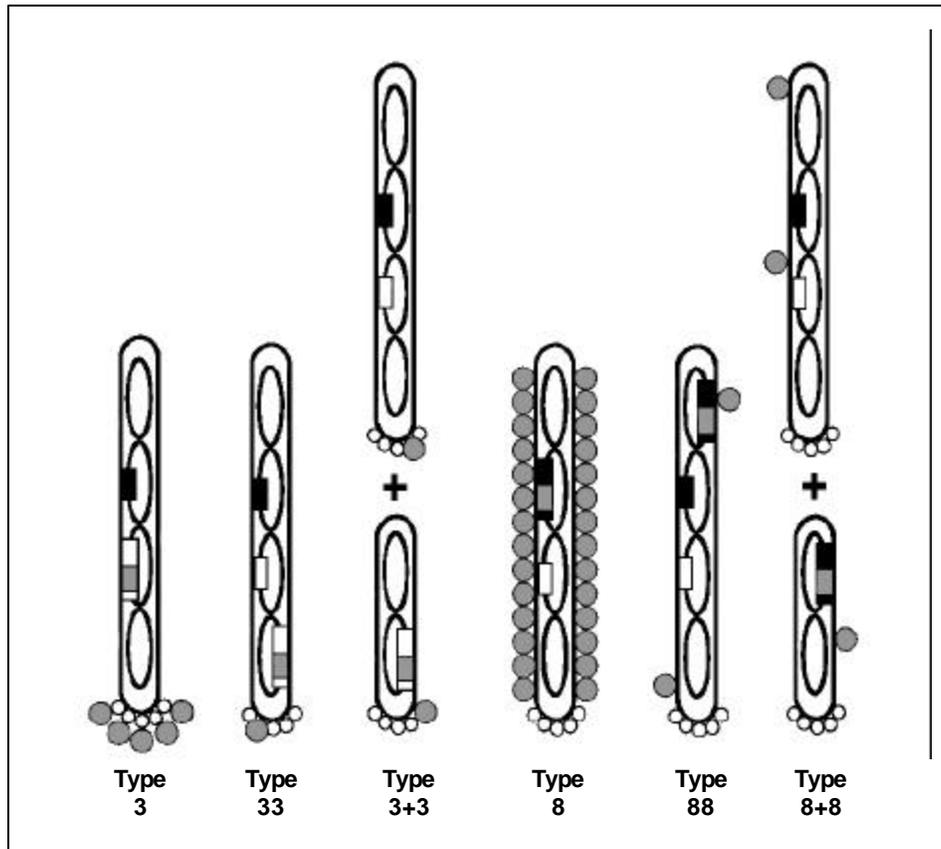


Figure 2.5. Types of phage display systems (Smith & Petrenko, 1997)

In a type 88 vector, the phage genome bears two genes VIII, encoding two different types of g8p molecule, one being ordinarily recombinant and the other one wild-type. The resulting virion is a mosaic, its coat comprised of both wild-type and recombinant g8p molecules (the former usually predominating). This allows hybrid g8p proteins with quite large foreign peptides to be displayed on the virion surface, even though the hybrid protein by itself cannot support phage assembly. Similarly, a type 33 vector bears two genes III, one of which is recombinant. A type 8+8 system differs from a type 88 system in that the two genes VIII are on separate genomes. The wild-type version is on a phage, usually the helper phage, while the recombinant version is on a phagemid (Smith & Petrenko, 1997).

The phagemid carries the origins of replication for *E. coli* and the filamentous phage (which is inactive until the cell is infected with the helper phage), the phage replication protein acts not only on the phage origin on the helper phage DNA, but also on the phage origin on the phagemid DNA (Fuh & Sidhu, 2000). Two types of progeny virions are thus secreted, namely particles carrying helper phage DNA, and particles carrying phagemid DNA. Both these virions, like the type 88 virions, are mosaics, whose coats are composed of a mixture of recombinant and wild-type g8p molecules. When a phagemid virion infects a cell, the cell acquires the antibiotic resistance carried by the phagemid. When a helper phage virion infects a cell, the cell goes on to produce progeny helper virions in the normal manner. The resulting progeny virions are not mosaic, since the helper phage carries only a single gene VIII. Type 3+3 and 6+6 systems are like type 8+8 systems, except that the phagemid carries an insert-bearing recombinant gene III or VI, respectively, rather than gene VIII. The recombinant g3p encoded by a type 3+3 phagemid is usually missing the N-terminal domain, since cells expressing this domain are resistant to superinfection by helper phage (Smith & Scott, 1993).

Most phage display vectors are designed to be introduced into *E. coli* cells as naked DNA by electroporation (Dower *et al.*, 1988), which is particularly well-suited to making very large libraries. Special display vectors that can be packaged *in vitro* into phage λ particles have also been reported (Hogrefe *et al.*, 1993).

2.2.4 Phage-displayed peptide libraries

2.2.4.1 Random peptide libraries

The construction of combinatorial peptide libraries is an important application of phage display technology. Peptide libraries displaying many millions of random peptide sequences on the surface of the filamentous bacteriophages fd and M13 are used to define ligand-binding sites that are difficult to identify by conventional techniques (Szardenings, 2003).

Synthetic oligonucleotides with a constant length but with unspecified codons, randomised through site-directed mutagenesis using degenerate oligodeoxynucleotides, are cloned as fusions to one of the M13 phage coat proteins, where they are expressed as peptide-capsid fusion proteins. Phage-borne peptides exhibit a wide mimicking potential to linear, conformational, and nonproteinaceous epitopes (Smith, 1991). These random peptide libraries can be tested for binding to target molecules of interest. The display of random peptides on filamentous bacteriophage as fusion to either g3p or g8p coat proteins (Sternberg & Hoess, 1995), has allowed the identifications of peptides that specifically bind to a variety of targets (Turk & Cantley, 2003). Moreover, display of small peptides on the surface of phage particles can increase their immunogenicity and consequently their potential as vaccine candidates (Azzazy & Highsmith, 2002).

2.2.4.2 Applications for phage-displayed peptide libraries

Phage display of random peptide or gene fragment libraries offer a unique approach to the elucidation of protein interaction networks and signalling pathways in the absence of a three dimensional structure (Stephen & Lane, 1992). Many enzymes that act on proteins, recognise their substrates on the basis of the sequence context surrounding the site of modification. Likewise, protein-protein interactions are often mediated by modular domains that interact with short linear stretches of protein sequence, often in the context of post-translational covalent modifications to the binding partner (Pawson *et al.*, 2002).

As extracellular protein-protein interactions typically involve large contact surfaces void of significant concavity, conventional small-molecule screening efforts have largely failed to identify antagonists for these interactions (Cochran, 2000). In contrast, phage-displayed peptide libraries can be used to isolate peptides that bind with high specificity and affinity to virtually any target, and have proven remarkably successful in generating both antagonists and agonists for numerous extracellular targets (Sidhu *et al.*, 2003). This strategy thus offers a

practical alternative to the laborious individual analysis of impossibly large numbers of peptides or mutant proteins (Cwirla, *et al.* 1990). These binding peptides can then be used as reagents to understand molecular recognition, as minimised mimics for receptors, or as lead molecules in drug design (Turk & Cantley, 2003).

Library members are screened with immobilised proteins by alternating rounds of affinity selection and viral replication, followed by sequence determination of selected phage particles. Displayed peptide sequences are then compared to the protein sequence of the binding partner, with the assumption that the immobilised protein will identify phage particles from the library carrying amino acid strings with which it normally makes molecular contact (Rodi & Makowski, 1999). In cases where the selected peptides bear obvious resemblance to the natural peptide ligand, but bind just as tightly to the screening target as the natural peptide, they are termed mimotopes (Smith & Scott, 1993).

Random peptide libraries have been used successfully in investigating different types of protein interactions. Peptide sequences identified by phage display have been shown to act as agonists and antagonists of receptors (Doorbar & Winter, 1994). Peptides that neutralise immunoglobulins may be employed as diagnostic reagents or used as therapeutic agents for controlling autoimmune diseases (Blank *et al.*, 1999). Random peptide libraries have been used for epitope mapping of monoclonal and polyclonal antibodies, even in cases in which the antigen is not available or even not yet known (Hill & Stockley, 1996), to elicit antibodies against the coat proteins of parasites and viruses (Azzazy & Highsmith, 2002), and to develop substrate phage to define substrate sites for different enzymes (Matthews & Wells, 1993). Library screenings of phage-displayed peptides have identified peptide ligands for carbohydrate moieties (Szardenings, 2003), protein kinase substrates (Dente *et al.*, 1997), cell surface receptors (Wrighton *et al.*, 1996), receptor ligands (Balass *et al.*, 1997), folded domains within larger proteins (Zwick *et al.*, 1998), and cultured cells and serum

samples (Cortese *et al.*, 1996). Peptide mimics of nonpeptide ligands (Devlin *et al.*, 1990), and peptides that bind small molecules have also been isolated using this methodology (Rozinov & Nolan, 1998).

Another exciting development has been the use of phage display to select organ-specific peptides *in vivo*. Following injection of a peptide library into mice, the organs of interest are harvested and washed, and the eluted phages used in subsequent rounds of injection and selection (Pasqualini & Ruoslathi, 1996). Peptides selected in this manner have been successfully used to specifically deliver drugs to tumour cells (Arap *et al.*, 1998).

2.2.5 Phage display of proteins

Despite the great potential of filamentous phage display, inherent limitations imposed by the phage assembly mechanism limit the scope of proteins that can be displayed with this phage class (Dunn, 1996). Display with lytic phage, such as bacteriophage λ , may complement the filamentous phage system, especially for the functional cloning of intracellular enzymes (O'Neil & Hoess, 1995). Protein libraries displayed on filamentous M13, however, still leads the way as prototype for protein display (Forrer *et al.*, 1999).

Certain functional homodimeric or homotrimeric proteins have been successfully displayed on filamentous phage, where association between the subunits, fused with the phage g3p product, occurs in the periplasm with at least one of the subunit fusions incorporated into a phage particle (Clackson & Wells, 1994). Heterodimers such as Fab antibody fragments can be displayed by coexpressing both subunits but with only one fused to the appropriate phage protein (Winter *et al.*, 1994). Hormones (Lowman *et al.*, 1991), protease inhibitors (Roberts *et al.*, 1992), and DNA binding proteins have successfully been displayed (Soumillion *et al.*, 1994).

2.2.5.1 cDNA libraries

Over the past decade, cDNA expression libraries have become a ubiquitous tool for the identification of genes encoding for ligands to proteins, antibodies and nucleic acids (Hufton *et al.*, 1999). With the realisation that phage display is a very powerful protein engineering tool, it was soon anticipated that there would be applications for cDNA library analysis (Crameri *et al.*, 1994). The main application of cDNA phage display libraries thus far has been the identification of natural binders to antibodies derived from patients suffering from allergy, autoimmune diseases or certain cancers. However, phage displayed cDNA libraries are also increasingly used for the identification of protein–protein interaction partners (Crameri & Kodzius, 2001).

While the M13 phage is successful for generating short peptide libraries or specialised libraries, it is not the ideal presentation vehicle for complex repertoires from natural sources, such as cDNA libraries (Smith & Petrenko, 1997). This may be due to some peculiar biological features of the phage, such as the requirement for the fusion products to be secreted prior to phage assembly. This may introduce a bias during phage production because of inefficient recombinant protein translocation, which in turn would lead to under-representation, or even the absence, of many polypeptides in the library (Malik *et al.*, 1996). Expression products of oligo(dT)-primed cDNA libraries have, however been successfully been fused to the C-terminus of coat protein g6p (Jespers *et al.*, 1995) or displayed by using a Jun-Fos leucine zipper system on g3p (Crameri & Walter, 1999).

All filamentous phage cDNA surface libraries are subject to the same constraints limiting the transport and secretion of proteins in such phage assembly. However, despite their limitations, it seems clear that they can nevertheless yield very useful results. Surface cDNA libraries in one of the existing λ display systems may complement whatever shortcomings exist with corresponding filamentous phage libraries (Allen *et al.*, 1995). Methods to display cDNA

libraries as C-terminal fusions utilizing display vectors based on phage λ (Santini *et al.*, 1998) and T4 and T7 phages have been developed (Sche *et al.*, 1999). Although all these systems show considerable promise, no single display format has proven universally applicable. This is probably because of the fact that only a subset of eukaryotic proteins can be efficiently expressed in *E. coli*, and that only a fraction of these can be efficiently displayed with any given system (Noren & Noren, 2001).

Efficient screening methods based on selective enrichment of clones expressing desired gene products, allow the isolation of all ligand-specific clones that are present in a library (Santi *et al.*, 2000). Manual identification of clones by restriction analysis and random codon sequencing is unlikely to be successful for the isolation of gene products derived from rare mRNA species. Combining cDNA phage display and the power of robotic-based high-throughput screening, however, allow the rapid handling of large numbers of individual clones. The strategy is based on a combination of phage display and high-density arrays, and represents a widely applicable method for rapid high-throughput identification of all individual cDNAs presented in selectively enriched libraries (Cramer & Kodzius, 2001).

2.2.5.2 Antibody libraries

The analysis of the potentially hundreds of thousands of different proteins constituting the proteome requires high numbers of versatile and highly specific tools such as antibodies for protein detection, in-depth expression profiling and functional studies (Kretzchmar & Von Rüden, 2002). Phage display technology is well suited for high-throughput generation of antibodies for research purposes such as massive target identification and validation programs (Holt *et al.*, 2000).

The first monoclonal antibodies were generated from mouse B-cell hybridomas almost 30 years ago (Köhler & Milstein, 1975), and in 1989 monoclonal antibodies were isolated from the first combinatorial antibody library expressed

by phage lambda (Huse *et al.*, 1989). Driven by the success of this initial report, incorporation of the filamentous phage display method into general practice was rapid, and over the past decade, phage-displayed antibody fragments have been the subject of intensive research (Rader & Barbas, 1997). Today, it is possible to mimic the key features of the *in vivo* antibody production and antigen-driven affinity maturation processes, by expressing antibody fragment gene repertoires on the surface of phages. High affinity antibodies can be made without prior immunisation, and their binding properties can be further manipulated *in vitro* (Hoogenboom *et al.*, 1998).

As a result, antibody phage display is considered a very popular alternative to hybridoma technology for the production of monoclonal antibodies (Corisdeo & Wang, 2004). Antibody libraries have become practical tools for drug discovery and several phage-derived therapeutic monoclonal antibodies have been developed for a broad range of indications and even more are being tested in advanced clinical trials. It is estimated that phage display has provided approximately 30% of all human antibodies currently in clinical development (Reichert, 2000).

Antibody fragments can be displayed on the surface of phage in different formats including Fab fragments (Cabilly, 1989), variable region fragments (Fv) (Skerra & Plückthun, 1988), single-chain Fv's (scFv) (McCafferty *et al.*, 1990), Fv's with an engineered intermolecular disulphide bond to stabilise the V_H - V_L pairs (dsFv) (Brinkmann *et al.*, 1995), and diabody fragments (Holliger *et al.*, 1993). Filamentous bacteriophage such as M13 are most often used, although T7 bacteriophage has been shown to also allow antibody display (Kretzschmar & Von Rüden, 2002).

In contrast to classic hybridoma approaches for monoclonal antibody generation, phage display can produce antibodies against weakly or nonimmunogenic, conserved proteins (Winter *et al.*, 1994). Three types of libraries are typically

displayed on phages, namely immune, single pot naïve and synthetic antibody libraries (Hoogenboom *et al.*, 1998). An immune library, constructed from a host immunised with a target antigen, reflects the diversity and maturation of the host immune response and favours selection of high-affinity, specific antibodies (Clackson *et al.*, 1991).

Single pot libraries, also known as naïve, nonimmune or universal human antibody libraries, contain an unbiased repertoire of variable regions from V_H and V_L from cDNA of many antibody genes (Marks *et al.*, 1991). Theoretically, a universal naïve antibody library can be used multiple times for diverse applications, however, it typically needs to be large (more than 10⁸ clones) to select high-affinity antibodies (Sheets *et al.*, 1998). Moreover, studies imply that universal libraries are a richer source of antibodies against cell surface markers as compared to immune libraries, since *in vivo* tolerance mechanisms could have deleted antibody clones reactive against cell surface antigen from the immune repertoire (Roovers *et al.*, 2001).

Although it is possible to obtain specific antibodies directly from naïve phage-displayed repertoires, another important application for phage display technology has been the humanization and affinity maturation of antibodies, lacking sufficient affinity for therapeutic applications (Sidhu, 2000). Affinity maturation essentially involves the introduction of diversity in the V-genes of the antibody, to create a secondary library, selection of higher affinity from the low affinity variants, and screening to discriminate between antibody variants with differences in affinity or kinetics of binding (Lowman & Wells, 1993).

An important advance has been the engineering of recombinant antibody forms including the development of high-quality libraries with completely synthetic complement-determining regions (CDR3 regions) (Knappik *et al.*, 2000). Synthetic antibodies are built by *in vitro* assembly of V-gene segments and D/J segments. V-genes may be assembled by introducing a predetermined level of

randomisation of CDR regions into germline V-gene fragments (Barbas *et al.*, 1992).

Large antibody libraries can be screened with speed and flexibility, and the recombinant antibodies can then be expressed in bacterial or mammalian culture system (Li & Aitken, 2004). The selection of antibodies by phage display relies on several factors: (i) the ability to isolate or synthesize antibody gene pools to construct large, highly diverse libraries; (ii) the possibility to express functional antibody fragments in the periplasmic space of *E. coli* (Skera & Plückthun, 1988); and (iii) the efficient coupling of expression and display of the antibody protein with the antibody's genetic information being packaged in the very same *E. coli* bacteriophage (McCafferty *et al.*, 1990).

2.2.5.3 Applications of phage-displayed proteins

Phage display intrinsically selects reasonably thermodynamically stable and folded members from a protein library, provided, of course, that such members exist in the library, because of the interdependence of functional display and affinity selection (Ruan *et al.*, 1998). A new selection concept for the isolation of proteins with increased stability links the protease resistance of folded species directly to infectivity. This method has the great advantage that it is broadly applicable to all polypeptides without the need for a binding ligand, yet at the price of not maintaining the selection pressure for function (Sieber *et al.*, 1998).

2.2.5.3.1 Engineering proteins

Once a population of protein ligands has been isolated, further layers of modification and selection can be applied in order to enhance or manipulate binding properties or affinities. Several groups have used phage display to engineer zinc-finger domains with designed DNA-binding specificities that can be used to control gene expression (Wolfe *et al.*, 1999). In another application, SH3 domains with improved or altered binding properties were selected, thus demonstrating that phage display can be used to engineer signalling protein

interaction domains (Hiipakka *et al.*, 1999). In addition, complex signalling pathways can be simplified by using phage display and rational design to generate receptor-selective variants of a pleiotropic factor to elucidate the specific role of each receptor (Li *et al.*, 2000).

Techniques for the identification and enrichment of proteins with enzymatic activity have been developed (Pedersen *et al.*, 1998). The main value of this strategy appears to lie in the identification of novel protein catalysts from repertoires on phages, such as, for example, the selection of a specific enzyme out of a cDNA library or a catalytic antibody out of an appropriate antibody library. In this case, the proximity of the substrate and catalyst on the phage particle has the potential advantage that poor catalysts are also selectable (Forrer *et al.*, 1999).

The strategy of directing a population of proteins towards specific properties by creating random sequence variation is known as directed evolution. In contrast to rational approaches for manipulating the properties of proteins, directed evolution has the advantage that proteins can be manipulated without the need for a prior knowledge of molecular structure, or of the details of molecular action. Through the use of directed evolution it has been possible to identify stronger binding ligands to receptors, and to produce novel enzyme inhibitors and DNA binding proteins (Lowman & Wells, 1993).

The products of convergent evolution experiments can be a fruitful source of variants upon which further diversity can be imposed. Using the sequences encoding isolated peptides as a starting point, a second combinatorial library may be generated that is varied around selected sequences. The starting point for directed evolution can also be a protein of which the function is already known and characterised. A number of strategies are employed to introduce limited variation, including error prone PCR, the amplification of phage populations in mutator strains of host bacteria and DNA and family shuffling (Willats, 2002).

2.2.5.3.2 Substrate binding phage display

Phage display is highly useful not only for the selection and evolution of enzymes, but also for the selection and evolution of polypeptide substrates (Forrer *et al.*, 1999). Substrate phage display, an adaptation of phage display, was first described by Matthews and Wells, and became a prototype approach for identifying novel substrates for proteases. A library of fusion proteins is constructed containing the N-terminal domain used to bind to an affinity support, followed by a randomized protease substrate sequence and the C-terminal domain of M13 g3p. Each fusion protein is displayed as a single copy on filamentous phagemid particles (substrate phage) (Matthews & Wells, 1993). Phages are then bound to an affinity support and treated with the protease of interest. Phages that display good protease substrates are released, whereas phage with substrates that resists proteolysis remain bound. This approach is valuable in identifying novel substrates or optimizing existing ones (Ohkubo *et al.*, 1999).

In a more general way, the product-containing phages may be bound, for example, by an antibody, similar to the catELISA approach, in which the occurrence of an enzymatic reaction is detected by an antibody specific for the product (Tawfik *et al.*, 1993). In a similar fashion, bacteriophage λ can be used for the display of random polypeptide fragments derived from a biotin-accepting protein domain. Phages displaying a functional domain, which can be biotinylated in the cytoplasm of *E. coli*, are identified after affinity selection with avidin, demonstrating that it is feasible to select for the occurrence of an enzymatic reaction by product binding (Stolz *et al.*, 1998).

2.2.6 Screening phage-displayed libraries

Once a phage display library has been constructed, the library must be screened in such a way that the original very high diversity of the library is reduced to a manageable number of clones which can then be analysed in detail. Most screening procedures are based on affinity selection and involve the following

fundamental steps: (i) the library is amplified and phage particles produced; (ii) phage particles are exposed to a target for which a binding protein is sought; (iii) non-binding phages are removed by washing; and (iv) binding phages are eluted, infected into host bacteria and thereby amplified (Willats, 2002). Figure 2.6 illustrates the general selection procedure for phage-displayed libraries.

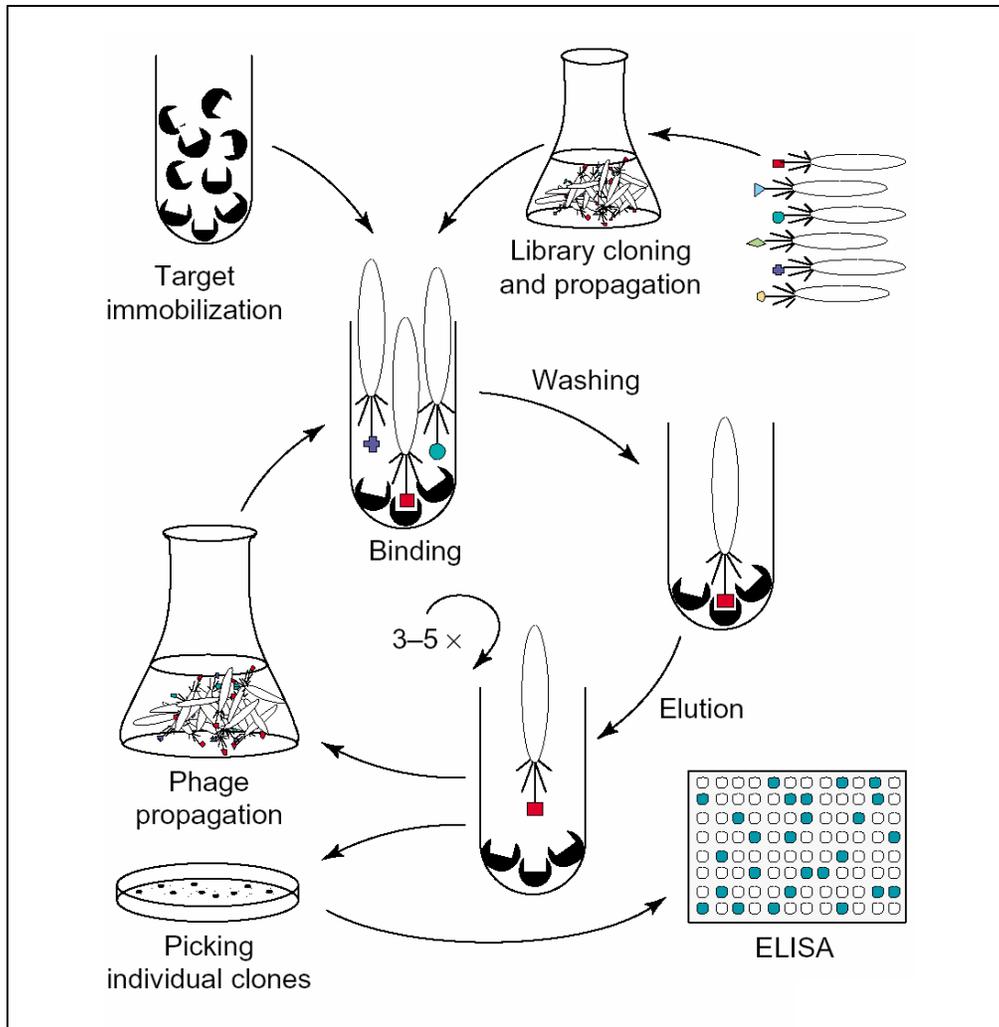


Figure 2.6. Screening phage-displayed libraries (Konthur & Walter, 2002)
 Phage display allows the *in vitro* selection of specific peptides by enrichment of binding phages during cycles of biopanning and propagation. This enables the rapid isolation of desired molecules from hugely diverse libraries.

Although libraries with very high diversities are available, some expressed sequences are incompatible with phage propagation, whilst others are highly susceptible to proteolysis during propagation. These factors impose constraints on effective diversity and it is therefore desirable to start with a library that is as diverse as possible (Rodi *et al.*, 2002). The possibility that some expressed sequences may be somewhat deleterious to phage propagation can be militated to some extent by including a growth step in each panning round that creates less competitive growth conditions, for example by growing on solid media rather than exclusively in liquid culture (McGregor, 1996).

Crucial to the success of a panning experiment is the complexity of the library. The more sequences contained within the library, the more likely that a sequence that binds to the target is present. Library size is effectively limited by the volume requirements and solubility of M13 phage, as well as the logistics of large-scale ligation and transformation of combinatorial libraries (Rodi *et al.*, 2002). A practical upper limit for the number of phages to work with in a panning experiment is in the order of 10^{11} (Smith & Scott, 1993).

Electroporation of a library of random DNA sequences in *E. coli* needs to be carried out under conditions where the cells are in excess, to prevent multiple sequences from being taken up by a single cell (Dower *et al.*, 1988). This effectively limits the complexity of the primary unamplified library to 10^9 – 10^{10} electroporated clones. The use of 10^{11} amplified phages in a panning experiment thus corresponds to 10–100 copies of each sequence present in the primary library. Library complexities in the order of 10^9 can be easily achieved if both the ligation and the electroporation steps are as efficient as possible (Gaskin *et al.*, 2001). For a typical cloning experiment, a single clone of the right identity is often sufficient. For cloning combinatorial peptide libraries, however, the equivalent of a billion clones are needed, necessitating careful optimization of ligation and transformation conditions (Noren & Noren, 2001).

One of the strengths of phage display is that screening protocols can readily be tailored to the particular requirements of many different target molecules. The selection of ligand-binding phage particles has most often been achieved by screening of the library using antibody bound to a solid phase. Two approaches are commonly used. Either the antibody is attached directly to the surface of microtitre plates (Heiskanen *et al.*, 1999) or polystyrene beads (Keller *et al.*, 1993), or the antibody is first biotinylated and then immobilised on streptavidin-coated plates (Parmley & Smith, 1988). More recently, library screening has been successfully achieved by binding antibody to a BIAcore sensor chip (Malmberg *et al.*, 1996). After immobilisation of the antibody, the library is reacted with bound immunoglobulin. The unbound phages are then removed and the bound phages eluted at low pH. The eluted phages are amplified in bacterial cells prior to subsequent rounds of biopanning (Parmley & Smith, 1988). Thus, by successive binding and elution, preparations of phages are selectively enriched and amplified by repeated rounds of biopanning. After three to four rounds of this process, the majority of clones bind to the selecting antibody (Scott & Smith, 1990).

However, these procedures tend to select mainly medium to low affinity-binding clones (Cwirla *et al.*, 1990). A modified screening procedure established polystyrene beads as the optimum surface for the immobilisation of antibody during the biopanning process. Together with a stepwise decrease in the pH of the elution buffer in the final round of biopanning, this method results in the elimination of non-binding clones and an increase in the efficiency in isolating high affinity binding clones (D'Mello & Howard, 2001).

More innovative screening methods have also been employed including panning against whole fixed or living cells, tissue sections or even within living animals (Watters *et al.*, 1997). Screens may also be designed such that specific complexes can be selected, for example, infectivity screening is based on phages bearing truncated, non-infective fusion coat proteins. Infectivity is

restored only if a complex is formed with a binding partner that has the capacity to restore infective functionality to the truncated coat protein (Willats, 2002).

The basic purpose of washing is to remove non-binding phages from the selection process so that binding phages are selectively enriched. Most phage display libraries of whatever sort are likely to contain clones with a spectrum of avidities for any particular target. Some may be strong binders with low specificities, others the reverse. If washing is too stringent then highly specific, but weak binders may be lost. If washing is not stringent enough then populations of selected clones may be dominated by strong binders with low specificity. In practice this balance is achieved by adjusting washing times, detergent concentrations and using regimes in which washing stringencies are progressively increased. A number of treatments can be used to elute bound phages from targets. Dramatically lowering or increasing the pH is often employed, or reducing agents may be used to disrupt disulphide-based links between supports and targets. A more subtle approach using enzymatic cleavage may be employed where there are concerns about the effects on phage integrity of harsh elution conditions (Rader & Barbas, 1997). Enzyme cleavage sites can be incorporated into the fusion protein, for example a trypsin cleavage site can be inserted between M13 g3p and the displayed fusion protein (Willats, 2002).

Following elution of bound phages, it is essential to then amplify the recovered phage population before the next round of biopanning, and indeed virtually all protocols include this step. However, this methodology may be worth careful examination since some reports indicate that directly using eluted phages without amplification may reduce background problems and help reduce the number of non-specific phages that are inevitably carried through the panning process (McGregor, 1996). The rationale is that during amplification, phages with inferior avidities for the target but better growth characteristics may be preferentially amplified. This has some important practical implications. The *in vivo*

amplification steps are the most time consuming part of phage display library screening, and if they could be avoided the time required for each screen would therefore be greatly reduced (Rader & Barbas, 1997). Moreover, without the *in vivo* steps it is much easier to envisage how the whole screening process could eventually be completely automated (Rhyner *et al.*, 2002).

2.2.7 Recent innovations in phage display technology

2.2.7.1 Selectively infective phage (SIP)

The development of selectively infective phage (SIP) technology has provided a novel method for the *in vivo* selection of interacting protein-ligand pairs (Krebber *et al.*, 1995). It consists of two components, namely a phage particle made noninfective by replacing its g3p N-terminal domains with ligand-binding proteins, and an adapter molecule in which the ligand is linked to the N-terminal domains of g3p which are missing from the phage particle (Gramatikoff *et al.*, 1994). Infectivity is restored when the displayed protein binds to the ligand and thereby attaches the missing N-terminal domains of g3p to the phage particle. Phage propagation is thus strictly dependent on the protein-ligand interaction (Krebber *et al.*, 1997).

Advantages of SIP technology include low background and elimination of the need for inefficient physical separation of the specific binders from the non-specific binders. SIP is a one-step procedure that appears to be very efficient and rapid in selecting for high affinity ligands (Azzazy & Highsmith, 2002). In addition, this strategy can possibly be extended to eukaryotic systems, where it might be used to design a vector for gene therapy, in which a virus is only infective in the presence of an exogenously added adapter protein (Krebber *et al.*, 1997).

SIP technology has already been modified in the development of SIP/polyphage system, where the two interacting partners are encoded on two different vectors which are co-packaged into phage particles. The system involves a combination

of the SIP technology, to select protein interactions, with the use of polyphages to shuffle and link the genetic information contained in two libraries. SIP/polyphage combines several advantageous features to make it a promising candidate system for high throughput interaction screening (Rudert *et al.*, 1998).

2.2.7.2 Landscape phage libraries

Landscape libraries are peptide libraries displayed on modified phage known as landscape phage. In these phage amino acids 2 to 4 on every wild-type g8p major capsid protein is replaced with random octamers (Petrenko *et al.*, 1996). The phage-peptide structures may have emergent properties that are mainly dependent on the introduced variable peptide and its immediate surroundings on the phage surface. Further, if the foreign peptide is displayed on every subunit of g8p, additional properties may arise owing to the global architecture of the entire phage surface landscape (Petrenko & Smith, 2000).

Landscape phages were shown to serve as substitutes for antibodies against various soluble and cell-displayed antigens and receptors (Petrenko & Smith, 2000). The foreign random peptides spliced into the phage scaffold can be likened to antigen-binding regions of antibodies. They are highly variable and, because they are forced to lie up against the virus body, they are usually constrained by interactions with neighbouring wild-type residues to form a defined organic landscape (Kishchenko *et al.*, 1994). A landscape phage thus captures much of the essential features of the antibody molecule (Petrenko & Vodyanoy, 2003).

The randomised amino acids that form the active site of a landscape phage comprise up to 25% by weight of the particle and up to 50% of its surface area, which is an extraordinarily high fraction compared to natural proteins, including antibodies. In many applications, the extreme multivalency of landscape phage, that is the thousands of binding sites per particle, may be a great advantage (Petrenko *et al.*, 1996). The phage structure is extraordinarily robust, being

resistant to heat (Holliger *et al.*, 1999), many organic solvents (Olofsson *et al.*, 2001), up to 6M urea, acid, alkali and other stresses. Moreover, purified phage can be stored indefinitely at moderate temperatures without losing infectivity and probe-binding activity making them ideal antibody substitutes in various diagnostic assays (Petrenko & Vodyanoy, 2003).

An extension of the landscape library, is the development of the alpha (α) landscape library, in which biological selection helps to generate a great variety of conformationally biased α -helical ligands (Petrenko *et al.*, 2002). In this system, the randomised amino acids lie within the α -helical portion of g8p rather than at the N-terminus and are thus conformationally homogenous (Bianchi *et al.*, 1995). Strong biological censoring during phage growth probably prevents the appearance of amino acids which disturb the α -helical architecture of the major coat protein. Phage-borne α -helical peptides have a very rigid structure and cannot be used with sure success for the selection of ligands for any receptors and antibodies. However, if selected they provide precise information about the structure of the ligand, including its conformation, and give a clue for the design of lead compounds for this receptor. Moreover, the alpha library can be used as partner in combination with other landscape libraries, for the generation of mosaic phage libraries which have a very high diversity of antigen-binding sites, limited only by the volume of infected bacteria (10^{12} clones being a realistic number), and thus substituting the use antibodies (Petrenko *et al.*, 2002).

2.2.7.3 Ribosome display

Ribosome display was developed to overcome two major limiting steps in construction a very large phage display library, namely transformation efficiency, and the use of mutator strains to induce diversity in the library members (Hanes & Plückthun, 1997). This *in vitro* technology aims for simultaneous selection and evolution of proteins from diverse libraries without any bacterial transformation. In ribosome display, DNA which encodes a protein library is first transcribed to mRNA that is then purified and used for *in vitro* translation. *In vitro* translation of

mRNA is designed to prevent dissociation of mRNA, ribosomes, and the translated peptide. Such mRNA-ribosome-peptide complexes are then used for affinity selection on an immobilized target where only the complexes that do not encode a binding polypeptide, and specifically recognises the target antigen, are removed by washing. The mRNA, that encodes a polypeptide cognate for the target is then dissociated from the ribosomal complexes and reverse transcribed into cDNA. The prepared cDNA is then amplified by PCR and used for the next cycle of enrichment and PCR or analysed by sequencing (Hanes & Plückthun, 1999).

Because no transformation is necessary, large libraries can be constructed and used for selection. Additionally, library diversification is suitably introduced either before starting or in between cycles of ribosome display via DNA shuffling (Stemmer, 1994) or error-prone PCR (Cadwell & Joyce, 1994). Ribosome display has been applied with success to both peptides (Mattheakis *et al.*, 1994) and folded proteins (Hanes & Plückthun, 1997).

2.2.8 Applications of phage display in haemostasis

Over the past two decades, phage display technology has continually evolved and expanded giving rise to many possibilities in biotechnology and biomedical research. A great many publications employing phage display in the fields of thrombosis and haemostasis appear yearly, confirming that this technology is a key research tool in these fields (Arza & Féléz, 1998). Phage libraries have been used successfully to identify ligands of receptors, to define enzyme substrates, to increase anticoagulant activities (Yang *et al.*, 2002), and to select high affinity proteinase inhibitors (Tanaka *et al.*, 1999).

Phage display has been used with success in the following areas: (i) targeted therapeutics, (ii) *in vivo* phage display, (iii) genetic characterisation of alloantibodies and autoantibody inhibitors, (iv) epitope mapping, (v) receptor agonists and antagonist interactions, and (vi) the modulation of functional

pathways (Mullaney & Pallavicini, 2001). A very brief overview and selected examples of these applications will be given.

Arguably, the most important application of phage display is the display of antibody chains and their selection. Antibody phage display has been applied extensively to develop antibody reagents for subpopulation discrimination and to probe the immune response (Dall'Aqua & Carter, 1998). The first haematological applications of phage display were in the generation of antibodies against ABO, Rh, and Kell red blood cell antigens for agglutination and immunohistochemistry assays (Marks *et al.*, 1993). These studies established the suitability of the technology in generating antibodies against red blood cells (Siegel & Silberstein, 1994), platelets (Lang *et al.*, 1996), and several clotting factors (Suggett, *et al.*, 2000).

A successful therapeutic antibody is nonimmunogenic and has the ability to recognise a cell-surface molecule specific for the disease state. It further has a direct toxic effect or localises delivery of therapy to the targeted cell. Phage-displayed antibodies using human immunoglobulin genes are ideally suited to rapidly generate candidate therapeutic antibodies (Sheets *et al.*, 1998). Short lead-generation times and the ability to further optimise these lead candidates, contributes to the possibility of generating high potency therapeutics through phage display. Moreover, the time-consuming process of chimerisation or humanisation required for traditional commercial therapeutic monoclonal antibodies is avoided when using phage antibodies (Osbourn *et al.*, 2003).

Phage-displayed scFv libraries for anti-cancer antibody therapy are typically selected against tumour cell-surface molecules and then conjugated with a drug or toxin. scFv-targeted immunoliposomes couple the ability of the scFv to preferentially target a tumour cell with an encapsulated gene or chemotherapeutic agent with cellular internalisation (Nielsen & Marks, 2000). Cell surface-binding scFv with inherent internalisation capabilities may localise

the targeting chemotherapy drug closer to its mechanism of action. While the emphasis of most scFv targeting approaches has been directed toward cancer therapy, antibody-based targeting is also pursued for anti-inflammatory diseases (Bendas *et al.*, 1998) and for site-selected delivery in the cardiovascular system (Spragg *et al.*, 1997).

Although *in vitro* assays are typically employed for peptide and antibody phage selection, *in vivo* phage display makes it possible to select and identify peptides that interact with tissues and cells in their native three-dimensional microenvironment. Peptides were identified that mediate selective localisation of phages to brain and kidney blood vessels with a selectivity higher than 13-fold for these organs. Based on one of the brain-binding phage sequences, a soluble peptide was synthesised which specifically inhibited the localisation of homologous phages into the brain. When coated onto glutaraldehyde-fixed red blood cells, the peptide induced selective localisation of intravenously injected cells into the brain (Pasqualini & Ruoslahti, 1996).

Characterisation of self-antigens is essential for understanding the immunopathological mechanisms in autoimmunity, which may in turn lead to the development of effective therapies and novel diagnostic assays for autoimmune disease. Phage-displayed antibodies are used to investigate the immune response of patients with autoimmune thrombocytopenic purpura and haemophilia. DNA sequence analysis of phage antibodies facilitates identification of target antigens and may increase understanding of the autoimmune disease process (Osbourn *et al.*, 2003). In addition, antibody phage libraries also lend insight into the genetic basis of the alloimmune response and complexity of the humeral response (Mullaney & Pallavicini, 2001).

Autoantibodies directed against platelet membrane proteins are found in the serum of patients with chronic immune thrombocytopenic purpura. Peptides selected from a phage-displayed library of random linear hexapeptides against

these autoantibodies were shown to share sequence homologies with known autoantigens, GPIIb/IIIa and GPIb (Bowditch *et al.*, 1996). Identification of specific peptides that react with polyclonal antiserum can suggest autoantibody targets (Arza & Féléz, 1998).

In response to haemophilia A treatment with factor VIII concentrates, patients often develop antibodies against the A2, A3, and C2 epitopes of factor VIII. Antibodies directed against the C2 domain of factor VIII (anti-C2 antibodies) were isolated from the phage-displayed immunoglobulin repertoire of a patient with acquired haemophilia. Clones from a large fragment library were selected with an overlapping minimal amino acid region corresponding to the C2 domain (Van den Brink *et al.*, 2000a). The next study also used phage-displayed haemophilia patient antibodies, but this time isolated and defined the primary structure of anti-A2 antibodies. In addition, the study also suggested a previously undescribed epitope for anti-factor VIII antibodies located in the A2 domain (Van den Brink *et al.*, 2000b).

The use of peptide decoys may be a promising new approach for the neutralisation of pathologic antibodies. A random phage-displayed peptide library of surrogate epitopes was constructed to disrupt the interaction between the C2 domain of factor VIII and anti-C2 antibodies. Peptide decoys that mimic factor VIII epitopes, and interfere with the recognition of factor VIII by anti-factor VIII antibodies, were selected and characterised (Villard *et al.*, 2003).

Peptide, gene fragments and cDNA epitope libraries are applied to study several protein-protein interactions, to map epitopes and determine antibody-combining sites in haematology. The localisation of these binding regions has the potential to improve targeted therapeutics and vaccines by selection or development of small-molecule drugs (Mullaney & Pallavicini, 2001). The identification of functional groups presented by their side chains can provide the information needed for the design of nonproteinaceous ligands, such as the rigid organic

molecules engineered for binding platelet glycoprotein IIb/IIIa (McDowell *et al.*, 1994).

Numerous groups have constructed epitope libraries by either biological (Cwirla *et al.*, 1990) or chemical means (Houghten *et al.*, 1991). These libraries have proven to be useful for the rapid identification of epitopes to monoclonal antibodies, and have been used with success to identify peptide sequences that bind to a variety of protein targets (Hill & Stockley, 1996).

Phage display is used in current pharmacological in the search for agonists or antagonists of natural ligand-receptor interactions. Drug screening is an area for developmental application of phage display because small molecule libraries can be used to displace phage peptides bound to receptor, which may lead to identification of improved clinical lead compounds (Sidhu, 2000).

Monoclonal antibodies may be employed not only for mapping antibody recognising determinants, but also for screening epitope libraries to identify peptides that inhibit a receptor-ligand interaction. This was illustrated in a study that used an antibody, that neutralises the vWF-GPIb interaction, to successfully screen and identify peptides from a large epitope library that inhibit this receptor-ligand interaction. Synthetic peptides derived from the displayed sequences not only inhibited antibody binding to vWF, but also inhibited ristocetin-induced vWF binding to the GPIb receptor (South *et al.*, 1995).

Apart from screening with antibodies, epitope libraries can be also selected against the purified or recombinant receptor protein, and used to identify proteins and peptides that modulate receptor activity. In addition to their usefulness in the study of receptor function, biologically active proteins and peptides may also be used therapeutically (Hartley, 2002). Small cyclic peptides that bound to, and activated, the erythropoietin receptor were selected against a recombinant receptor. *In vitro* and *in vivo* biological assays suggested that these peptides

mimicked the functional properties of erythropoietin, even though the amino acid sequence was unrelated. Determination and analysis of the 3D structure of the complex between one of the mimetic peptides and the erythropoietin receptor, revealed the mechanism by which the peptides could induce the erythropoietin response (Wrighton *et al.*, 1996).

A peptide antagonist of the thrombin receptor was isolated by selecting directly against platelets. A library displaying peptide sequences based on the tethered ligand sequence was constructed and selected for binding to platelets. Phage clones were selected that encoded peptides able to immunoprecipitate the thrombin receptor and also shared common features with the ligand. These peptides were then chemically synthesised and studied further (Doorbar & Winter, 1994).

The purified extracellular domains of the $\alpha_5\beta_1$ fibronectin receptor and the $\alpha_{IIb}\beta_3$ platelet fibrinogen receptor were used to select peptides related to the natural ligand fibronectin, with recognition RGD motifs (Koivunen *et al.*, 1999). A constrained peptide library was used to isolate ligands displaying high affinity for the $\alpha_{IIb}\beta_3$ integrin and preventing platelet aggregation. This was achieved by flanking a library of hexapeptides with cysteine residues to introduce a degree of conformational constraint into random peptides (O'Neil *et al.*, 1992). A random cyclic hexapeptide library was also used to characterise the peptide binding specificity of the $\alpha_5\beta_1$ integrin (Koivunen *et al.*, 1994). The selected peptides were shown to antagonise integrin-mediated cell adhesion (Koivunen *et al.*, 1999).

More recently, monoclonal antibodies specific to the $\alpha_{IIb}\beta_3$ integrin were selected from a synthetic human antibody library. The selected antibodies strongly inhibited the interaction between $\alpha_{IIb}\beta_3$ and fibrinogen, and also inhibited platelet aggregation *ex vivo*. These may be the first human monoclonal antibodies that

are specific to $\alpha_{IIb}\beta_3$ and can potently inhibit platelet aggregation (Chung *et al.*, 2004).

Monoclonal antibodies directed against platelet membrane receptors are used extensively in the analysis of receptor structure and function (Mullaney & Pallavicini, 2001). In a recent study, a library of phage-displayed scFv antibodies were selected against whole platelets, and different platelet-binding human monoclonal antibodies were isolated. Such function-modulating antibodies may be useful in the development of potential antiplatelet drugs (Hagay *et al.*, 2003)

Applications of peptide phage libraries in agonist, antagonist, and allosteric selection strategies promise to facilitate identification of candidate epitopes without a prior knowledge of the protein interaction. Random peptide libraries have proven to be powerful tools for the selection of peptides that mimic linear, conformational, and even nonproteinaceous epitopes (Lowman, 1997). Peptides that induce a signalling response can be identified by selecting a random peptide phage library against purified cells and screening binders for their ability to influence or induce signalling responses (Brown, 2000).

Phage peptides that modulate enzyme activity involved in coagulation were obtained by selecting phage-displayed peptide libraries against the tissue factor–factor VIIa complex. The selected phage peptide acted as a highly specific and potent noncompetitive inhibitor of factor VIIa activity. Extensive mutational studies, nuclear magnetic resonance and crystal structure analyses revealed that this peptide bound to a novel exosite region, distinct from the active site. These results are of substantial therapeutic relevance, since in past, it has been extremely difficult to selectively inhibit individual coagulation proteases (Dennis *et al.*, 2000).

Other phage-derived anticoagulant leads include sequences that disrupt the binding of von Willebrand factor and collagen (Depraetere *et al.*, 1998). In one

study, a phage-displayed cDNA library was screened for binding to collagen, and the encoded collagen-binding protein was recloned and overexpressed in *E. coli*. The recombinant protein exhibited binding to collagen in a specific, dose-dependent and saturable manner (Viaene *et al.*, 2001).

A cyclic thrombin inhibitory peptide, containing the same sequence as that displayed on the phage surface, was synthesised after peptides from a cyclic heptapeptide phage display library were selected on their binding characteristics to α -thrombin. The peptide displayed promising *in vitro* antithrombotic activity and was shown to inhibit thrombin by binding to its active site, with an inhibitory effect on platelet activation and adhesion under certain conditions (Meiring *et al.*, 2002).

Displaying functionally active proteins and enzymes on phages without the loss of activity allows the generation and investigation of new protein variants with increased affinity and specificity (Dunn, 1996). Novel protein-protein interactions were created by combining phage display and loop grafting. Amino acids in a surface loop of human tissue-type plasminogen activator (tPa) were replaced with a recognition region from an antibody specific to $\alpha_{IIb}\beta_3$ integrins. The resultant tPa variant bound $\alpha_{IIb}\beta_3$, retained full enzymatic activity, and was stimulated normally by the physiological cofactor fibrin (Smith *et al.*, 1995).

An important prerequisite to the functional display of molecules is their suitability for either N-terminal or C-terminal extensions, created by fusion with the bacteriophage coat protein of choice. Hirudin, a potent thrombin inhibitor, was shown to tolerate no extensions at the N-terminus (Wallace *et al.*, 1989). It could, however, be C-terminally extended without substantially impairing its antithrombotic function. The expressed active hirudin can be utilised as a functional module in the design of multifunctional effector molecules in thrombolytic therapy (Wirsching *et al.*, 1997).

2.3 EXPRESSION OF CLONED GENES IN *ESCHERICHIA COLI*

2.3.1 Introduction to protein expression systems

The biotechnology industry has rapidly expanded in recent years, and as a result the expression of a spectrum of recombinant proteins in different systems for a wide variety of purposes has been a major challenge. In some applications, a large array of proteins are needed in relatively small quantities for screening applications, whereas in other cases, quantities approaching the metric ton scale are needed for specific therapeutic applications (Andersen & Krummen, 2002). The development of efficient methods for obtaining the necessary amounts of purified material thus serves as a source of significant competitive advantage in both academic and industrial laboratories (Yokoyama, 2003).

For example, the current postgenomic era is placing much emphasis on the analyses of protein structure, function and regulation. Structure-function studies utilise physical techniques such as calorimetry, nuclear magnetic resonance and X-ray diffraction, and these methods are often limited by the availability of sufficient amounts of purified protein (Yokoyama, 2003). Moreover, many proteins of interest are present only in very small quantities in their natural sources. This necessitates the use of recombinant methods, coupled with heterologous expression in a foreign host, to provide acceptable quantities of protein for structure-function studies. These recombinant expression systems are furthermore invaluable for the production and characterisation of the ever-increasing number of open-reading frames (ORFs) identified in the ongoing genome sequencing programs (Kost, 1997).

The expression of cloned genes, however, remains one of the fields in molecular cloning where success is not guaranteed. Investigators often do everything perfect, but still see the experiments fail (Sambrook & Russell, 2001). One of the reasons for this is that still too little is known about the mechanisms of folding of proteins in different organisms to predict which host-vector system might be best for a given protein. Moreover, few universal methods are available to increase

the efficiency of folding or to prevent the aggregation, denaturation, and/or degradation of foreign proteins expressed in environments that are unnatural for them. In the absence of a golden set of guiding principles, the expression of every cDNA or gene presents a unique set of problems which must be solved empirically (Rai & Padh, 2001).

No single expression system serves as a panacea for the problem of producing recombinant proteins in the tens to hundreds of milligram amounts typically required for thorough biophysical studies. Over the last couple of decades, five major gene expression systems have been developed to a point where they are successfully reproducible from one laboratory to the other, and where the necessary compounds are commercially available. The host cells for these expression systems are the bacteria *Escherichia coli* and *Bacillus subtilis*, the yeast *Saccharomyces cerevisiae*, cultured insect cells, and cultured animal cells (Shatzman, 1995).

Expression of cloned genes in bacteria is by far the simplest and most inexpensive means to produce large amounts of the desired product, and is consequently widely used both in research for the production of proteins for structural and/or biochemical studies, and in industry for the production of pharmaceutical proteins. The short generation times of the bacteria and efficiency of nutrient conversion to biomass is unequalled by eukaryotic-based expression systems, with the possible exception of those systems utilising yeasts (Hodgson, 1993). However, while *E. coli* remains the organism of choice for the high-level production of recombinant proteins, genomic information is rapidly being generated on organisms whose genes are not expressed well in *E. coli*. It is thus imperative that alternative host-vector systems are investigated for use in non-enteric bacteria (Hauser & Zylstra, 2001).

2.3.2 Protein expression in *Escherichia coli*

The Gram-negative bacterium *Escherichia coli*, has been a laboratory workhorse for many years. Extensive knowledge and practical experience exist with respect to this organism's genetics, biochemistry, and physiology (Gold, 1990). Genetic manipulations are straightforward, cultures grow rapidly and at high density on inexpensive substrates, and many foreign proteins are well tolerated and may be expressed at high levels. This, together with the availability of an increasing large number of cloning vectors and mutant host strains, makes *E. coli* the system of first choice for the expression of heterologous proteins (Olins & Lee, 1993).

However, in spite of this extensive knowledge surrounding *E. coli*, not every gene can be expressed efficiently in this organism. This may be due to the following factors: (i) the unique and subtle structural features of the gene sequence; (ii) the stability and translational efficiency of mRNA; (iii) the ease of protein folding; (iv) degradation of the protein by the host cell proteases; (v) the major differences in codon usage between the foreign gene and native *E. coli*; and (vi) the potential toxicity of the protein to the host. Fortunately, through thorough research over the last couple of decades, some empirical rules that can guide the design of expression systems and limit the unpredictability of this operation in *E. coli* have emerged (Hockney, 1994).

The major drawbacks of *E. coli* as an expression system include the inability to perform many of the post-translational modifications found in eukaryotic proteins, the lack of a secretion mechanism for the efficient release of protein into the culture medium, and the limited ability to facilitate extensive disulphide bond formation (Makrides, 1996). On the other hand, many eukaryotic proteins retain their full biological activity in a nonglycosylated form and therefore can be produced in *E. coli* (Sarmientos *et al.*, 1989).

Plasmid transformation of *E. coli* was first demonstrated by Cohen in the early 1970's (Cohen *et al.*, 1972), when he applied the observation made by Mandel that the combination of *E. coli* and bacteriophage lambda (λ) in a solution of CaCl_2 at 0°C produced infection (actually transfection) (Mandel & Higa, 1970). During the 'Great Recombination DNA Debate' which followed in 1975, the idea that there might be problems in constructing strains of *E. coli* that would make proteins to order was hardly questioned. In fact, much of the discussion centred around the degrees of physical and biological containment that would be required to protect the world from bacteria that expressed a certain foreign gene (Hodgson, 1993).

A considerable number of excellent studies have addressed the process of plasmid transformation, aiming to improve the frequencies of transformation, and at the same time characterise the parameters involved (Gold, 1990). In the last decade, much progress has been made in the understanding of protein folding, protein translocation across biological membranes, and the role of molecular chaperones in these processes. This improved understanding has led to a capability to accumulate proteins in a soluble form, secrete proteins from the cell cytoplasm, accumulate proteins in the cytoplasmic membrane, and direct proteins to the outer membrane of the cell for surface display (Olson *et al.*, 1998). Developments in *E. coli* expression systems include not only vector and strain developments, and control of transcription and translation, but also metabolic engineering of the cell's central metabolism. Coexpression of protein subunits, foldases and chaperones, protein folding, location and purification schemes, and *in vitro* refolding strategies are now available to aid in the success of an efficient expression system for active heterologous proteins (Balbas, 2001). In the section to follow, the components of an expression vector will be discussed in more detail.

2.3.2.1 Expression vector components

The two major processes involved in the heterologous production of proteins are the introduction of foreign DNA into the host cells, and the transcription of the foreign DNA in the chosen expression system. The first step involves the following considerations: (i) identification and isolation of the DNA to be introduced; (ii) vector choice and construction of recombinant vector; and (iii) identification of the suitable expression system to receive the rDNA (Marino, 1989).

A well-designed prokaryotic expression vector contains a set of optimally configured genetic elements that affect both transcriptional and translational aspects of protein production. The promoter is typically positioned approximately 10 to 100 bp upstream of the ribosome-binding site (RBS) and is under the control of a regulatory (repressor) gene, which may be present on the vector itself or integrated in the host chromosome (Hawley & McClure, 1983). The RBS consists of the Shine-Dalgarno (SD) sequence followed by an AT rich translational space that has an optimal length of approximately 8 bases. The SD sequence interacts with the 3' end of the 16S rRNA during translation initiation. The transcription terminator serves to stabilise the mRNA and the vector. In addition, the inclusion of an antibiotic-resistance gene facilitates phenotypic selection of the vector, and the origin of replication (Ori) determines the vector copy number (Makrides, 1996).

Multi copy plasmids have been extensively used as vectors for recombinant protein expression. When the production strategy involves protein secretion, the translation rate should be fine-tuned in order to avoid protein accumulation in the cytoplasm (Simmons & Yansura, 1996) and the formation of inclusion bodies (Swartz, 2001), which can happen if very high copy number systems are used. Low copy number plasmids have a number of advantages over high copy number plasmids, such as tight control of gene expression, the ability to replicate

large pieces of DNA, and low metabolic burden on host strains (Carrier *et al.*, 1998).

Expression vectors commonly used for overexpression of foreign genes in *E. coli* can be categorised according to the type of promoter used. The systems commonly used are driven by IPTG-inducible, bacteriophage T7, or bacteriophage lambda p_L promoters (Sambrook & Russell, 2001).

Plasmids carrying isopropyl- β -D-thiogalactoside (IPTG)-inducible promoters, based on the *lac* operon, are capable of expressing proteins at levels that exceed 30% of total mass of bacterial protein. These plasmids are well suited for small-scale laboratory experiments, but the high cost of IPTG prevents their use for large-scale production of foreign genes (Goldstein & Doi, 1995). Promoters used in these vectors include the *lac* promoter (Calos, 1978) and two hybrid promoters containing elements of the *trp* and *lacUV5* promoters, namely the *trp-lac (tac)* promoter (De Boer *et al.*, 1983) and *trp-lac (trc)* promoter (Amann & Brosius, 1985).

Expression systems using the bacteriophage T7 promoter, first developed by Tabor and Richardson (1985) and Studier and Moffatt (1986), employ transcription signals derived from the bacteriophage T7 genome. In 1990 Studier and colleagues developed the popular pET series of expression vectors which allow regulated expression of foreign genes by bacteriophage T7 polymerase (Studier *et al.*, 1990), as illustrated in Figure 2.7.

In recent years, the pET expression system (commercialised by Novagen, Madison, WI) have gained increasing popularity and is one of the most widely used systems for the cloning and *in vivo* expression of recombinant proteins in *E. coli*. This is due to the high selectivity of the pET system's bacteriophage T7 RNA polymerase for its cognate promoter sequences, the high level of activity of the polymerase and the high translation efficiency mediated by the T7 gene 10

translation initiation signals. The protein coding sequence of interest is cloned downstream of the T7 promoter and gene 10 leader sequences, and then transformed into *E. coli* strains. In addition to the T7 promoter, the vectors carry the colicin E1 (colE1) replicon that confers antibiotic resistance to either ampicillin or kanamycin (Novagen, 2003).

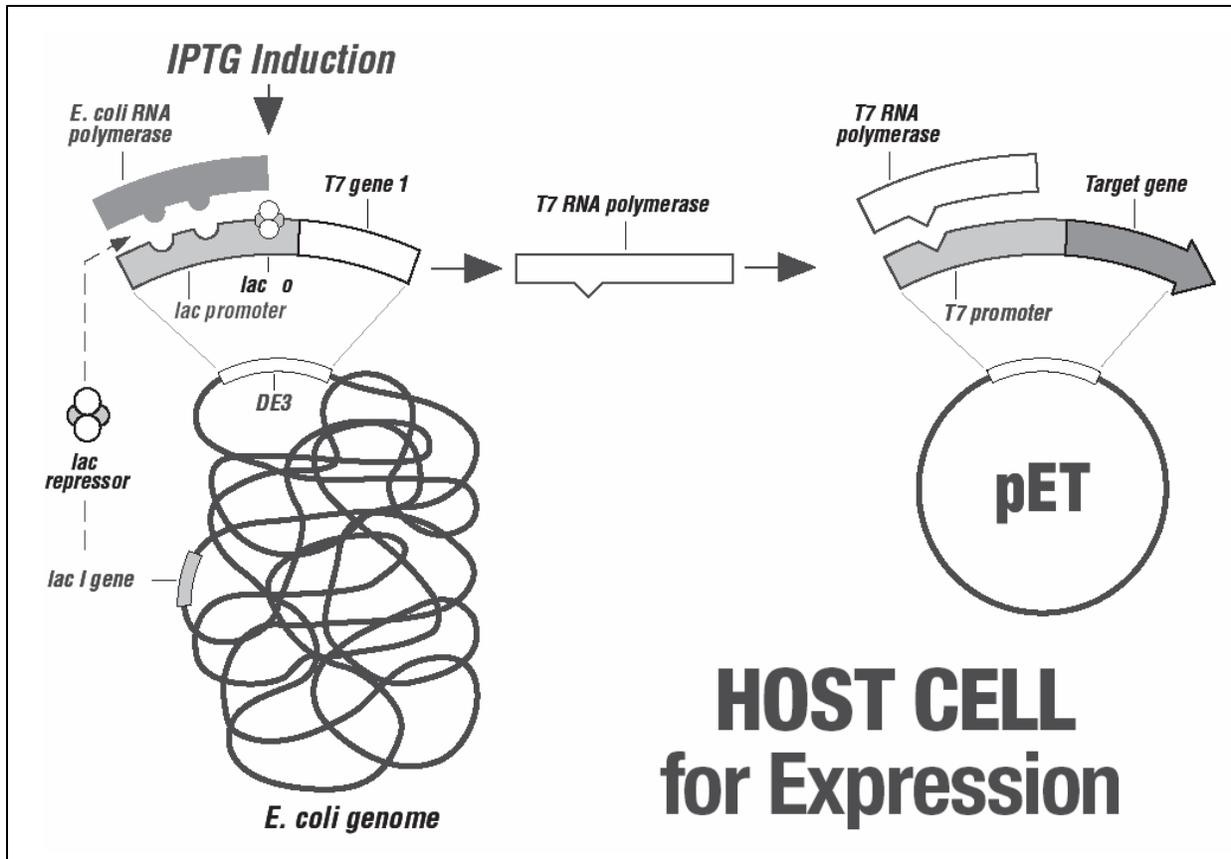


Figure 2.7. Control elements of the pET system (Novagen, 2003)

In the pET system, protein expression is achieved either by IPTG induction of a chromosomally integrated cassette in which the T7 RNA polymerase is expressed from the *lacUV5* promoter, or by infection with the polymerase-expressing bacteriophage lambda CE6 (Studier & Moffat, 1986). Due to the specificity of the T7 promoter, basal expression of cloned target genes is extremely low in strains lacking a source of T7 RNA polymerase. Upon induction the highly active polymerase essentially out-competes transcription by the host

RNA polymerase (Studier, 1991). This phenomenon, together with high-efficiency translation, achieves expression levels in which the target protein may constitute the majority of the cellular protein, achieved after only a few hours (Pan & Malcolm, 2000).

In vectors carrying the bacteriophage lambda (λ) ρ_L promoter, the promoter is regulated by a temperature-sensitive repressor, *clts857*, which represses ρ_L -driven transcription at low but not at elevated temperatures. *E. coli* strains harbouring the *clts857* gene must therefore be used as hosts with vectors carrying the λ ρ_L promoter. These vectors are particularly useful if the expressed gene product is toxic to *E. coli* (Bernard & Helinski, 1979).

2.3.2.1.1 Promoters

The ideal promoter is directed by the following criteria. Firstly, it should be strong, capable of protein production in excess of 10-30% of the total cellular protein. Secondly, it should be tightly regulated to exhibit a minimal level of basal transcription in order to minimise metabolic burdens and toxic effects on the host. Thirdly, it should be inducible to varying degrees by either a low-cost chemical inducers or by shifting the growth conditions (Hannig & Makrides, 1998).

However, many promoters are not ideally suited for the large-scale production of recombinant proteins, and share common problems such as leaky expression under non-inducing conditions, and the negative effect of induction on the physiology of the host (Weickert *et al.*, 1996). Leakiness often results in the overgrowth of plasmid-free cells and/or cells that carry a reduced capacity for high-level expression, an impediment for high-level protein production (Mertens *et al.*, 1995a).

There is variety of promoters available for gene expression in *E. coli*, including those derived from Gram-positive bacteria and bacteriophages. The most widely used promoters for large-scale protein production use either chemical inducers or

thermal induction. Chemically inducible promoters include, amongst others, those induced by IPTG addition (Goldstein & Doi, 1995); or by arabinose addition, the *araBAD* promoter (Guzman *et al.*, 1995). Then there are promoters induced by phosphate starvation, the *phoA* promoter (Kikuchi, 1981); or tryptophan starvation, the *trp* promoter (Russell & Bennett, 1982). Among the cold-responsive promoters is the λ p_L promoter (Bernard and Helinski, 1979); the tandem p_R , p_L promoters (Elvin *et al.*, 1990); and the cold shock promoter, *cspA* (Vasina & Baneyx, 1996).

Improvements made in promoter control in recent years include the incorporation of a thermosensitive *lac* repressor gene, *lacTs*, into expression plasmids to provide a tightly regulated, low-cost expression system (Adari *et al.*, 1995); the use of a reversible transcription termination system derived from phage λ to control leaky expression of T7 RNA polymerase (Mertens, *et al.* 1995a); expression plasmids with better stability provided by tandem transcription terminators that prevent read-through transcription into vector sequences (Mertens *et al.*, 1995b); the development of feed strategies for tight low-cost control of the *trp* promoter in large scale fermentations (Yoon *et al.*, 1996); and the development of thermally regulated, runaway replication plasmids that allow more efficient repression of the *trp* promoter (Kidwell *et al.*, 1996).

2.3.2.1.2 Transcriptional terminators

In prokaryotes, two different types of mechanisms effect transcription termination. Rho-dependent transcription termination depends on the hexameric protein rho, which causes the release of the nascent RNA transcript from the template. In contrast, rho-independent termination depends on signals encoded in the template, specifically, a region of dyad symmetry that encodes a hairpin or stem-loop structure in the nascent RNA, and a second AT-rich region (Wilson & Von Hippel, 1995).

Although often overlooked in the construction of expression plasmids, efficient transcriptional terminators are indispensable elements of expression vectors, because they serve several important functions. Transcription through a promoter may inhibit its function, a phenomenon known as promoter occlusion. The proper placement of a transcriptional terminator downstream of the coding sequence, to prevent continued transcription through another promoter, can prevent this interference (Adhya & Gottesman, 1982).

Similarly, a transcription terminator placed upstream of the promoter minimises background transcription (Nishihiara *et al.*, 1994). It is also known that transcription from strong promoters can destabilise plasmids, owing to overproduction of the ROP protein involved in the control of plasmid copy number, as a result of transcriptional read-through into the replicating region (Stueber & Bujard, 1982). In addition, transcription terminators can enhance mRNA stability and can substantially increase the level of recombinant protein production (Makrides, 1996).

2.3.2.2 mRNA stability

The process of mRNA degradation provides a major control point of gene expression in virtually all organisms. The concept of mRNA and its liability was established over 40 years ago (Jacob & Monod, 1961), and despite the many perplexing questions surrounding the mechanisms of mRNA decay, impressive progress have been made in understanding this important biological process (Coburn & Mackie, 1999).

A large multiprotein complex, now called the RNA degrasome, was discovered during the purification of *E. coli* RNase E (Carpousis *et al.*, 1994). The major components of the RNA degrasome include RNase E, polynucleotide phosphorylase (PNPase) and the DEAD-box RNA helicase, Rh1B (Miczak *et al.*, 1996). The degradation of mRNA is mediated by the combined action of the endonucleases (RNase E, RNase K and RNase II) and two 3'-5' exonucleases

(RNase II and PNPase) (Coburn & Mackie, 1999). mRNA degradation is not effected randomly by non-specific endonucleolytic cleavage, since there is no correlation between mRNA length and half-life (Chen & Belasco, 1990).

Considerable controversy exists over whether RNase E-dependent mRNA decay proceeds in the 5'-3' or in the opposite direction. In either case, stable secondary structures present the 5' untranslated region (UTR) of certain transcripts, as well as stem-loop structures in 3' rho-independent terminators, can both increase mRNA stability, though fine features modulate their efficiency (Baneyx, 1999). For example, specific sequences in the 5' UTR of certain mRNAs, such as the *ompA* transcript, have been shown to prolong the half-life of several labile heterologous mRNAs, as has the addition of a protective hairpin structure at the 5' terminus (Carrier & Keasling, 1999). These results can be compared with another study in which the C-terminal portion of RNase E was truncated to inactivate its RNase activity, thereby significantly decreasing total mRNA degradation (Lopez *et al.*, 1999).

However, none of the stabilising sequences identified to date functions as a universal stabiliser in heterologous mRNAs, but their integration into otherwise highly unstable transcripts may be effective (Swartz, 2001).

2.3.2.3 Translation initiation

Various factors affect the initiation of protein synthesis and the mere complexity of the process allows it to have a huge impact on expression efficiency. For example, control of initiation was used in a study to slow translation so that the secretion apparatus was not overwhelmed. Surprisingly, however, higher product secretion and accumulation were found to result from less effective ribosomal binding (Simmons & Yansura, 1996).

Research has shown that the wide range of efficiencies in the translation of different mRNAs is due to a consensus SD sequence, complementary to the 3'

end of 16S rRNA that enhances the formation of the translation initiation complex of the 30S ribosomal subunit with mRNAs. Proper spacing and sequence before the initiation codon also play a role, as does a downstream box (DB). The DB was shown to form a complex with anti-DB in the 16S rRNA to enhance translation initiation of DB-containing mRNA, in addition to the SD sequence (Etchegaray & Inouye, 1999).

However, an even more important factor may be possible unique secondary structures at the 5' end of each RNA species that block ribosome binding (Carrier & Keasling, 1999). These secondary structures can be disrupted by RNA helicases such as RhIB in *E. coli* (Linder & Daugeron, 2000). Rh1B is a member of the DEAD-box proteins, a family of putative ATP-dependent RNA binding helices that have a conserved motif including eight highly conserved motifs including the amino acids Asp(D)-Glu(E)-Ala(A)-Asp(D) (Schmid & Linder, 1992).

2.3.3 Choice of cellular compartment for protein expression

The decision to target the overexpressed protein to a specific cellular compartment, that is, to the cytoplasm, periplasm or the culture medium, rests on balancing the advantages and disadvantages of each compartment. Although for most applications it is desirable to achieve maximal production within the cytoplasm. Targeting the protein to extracellular compartments may offer an interesting alternative, especially when cytoplasmic expression results in toxicity or improper folding (Baneyx, 1999)

2.3.3.1 Cytoplasmic expression

Overproduction of heterologous proteins in the cytoplasm of *E. coli* is often accompanied by their misfolding and segregation into insoluble aggregates known as inclusion bodies. Several factors contribute to the inability of overexpressed proteins to fold into their authentic configuration, even in the presence of molecular chaperones. These include the reducing environment, the

lack of disulphide bonds and/or the absence of post-translational modifications (Wilkinson & Harrison, 1991).

In *E. coli*, two pathways contribute to the reduction of disulphide bonds, namely the thioredoxin system, which is composed of thioredoxin reductase and thioredoxin, and the glutaredoxin system, which consists of glutathione reductase, glutathione and three glutaredoxins (Prinz *et al.*, 1997). Strategies to generate a less-reducing cytoplasmic environment that facilitates disulphide bond formation include the use of *E. coli* strains deficient in thioredoxin reductase (*trxB*), which contributes to the sulfydryl-reducing potential (Bessette *et al.*, 1999).

A traditional approach to reduce protein aggregation and aid the formation of the native three-dimensional protein structure is through fermentation engineering, most commonly by reducing the cultivation temperature. The realisation that *in vivo* protein folding is assisted by molecular chaperones by foldases, which accelerate rate-limiting steps along the folding pathway, has provided powerful new tools to combat the problem of inclusion body formation (Richardson *et al.*, 1998). Although the production of several monomeric and multimeric proteins may be increased with overexpression of one or more chaperone or foldase with the protein of interest, the success of this strategy appears to be protein specific (Wall & Plückthun, 1995).

2.3.3.2 Periplasmic expression

Targeting proteins to the periplasm has both advantages and disadvantages. A major drawback of periplasmic expression is that the space is limited, and consequently yields of recombinant proteins generally never match those obtained upon cytosolic expression. However, in the case of proteins that bear multiple disulphide bonds of nonlinear connectivities in their natural conformations and that are resilient to renaturation of inclusion body material,

expression in the periplasmic space may offer the method of choice (Missiakas & Raina, 1997).

The periplasm affords ease and cost-effectiveness for target protein purification from a significantly smaller pool of bacterial proteins than compared with the cytoplasm. In addition, the oxidizing environment of the periplasm facilitates proper protein folding, and the *in vivo* cleaving of the signal peptide during translocation to the periplasm is more likely to yield the authentic N-terminus of the target protein (Makrides, 1996). Signal peptides of prokaryotic and eukaryotic origin have been utilised successfully for this purpose, but the presence of a signal peptide does not always ensure efficient protein translocation through the inner membrane because several structural features are involved in membrane transport (Andrews *et al.*, 1996).

Several strategies for improved translocation of proteins to the periplasm have been reported. These include the supply of components involved in protein transport and processing for example the overproduction of the signal peptidase I (Van Dijk *et al.*, 1991), the manipulation of the β -lactamase gene to reduce protein-expression levels in order to prevent the overloading of the translocation machinery (Mertens *et al.*, 1995b), and the coproduction of several proteins such as *prfF*, *prlA4* and *secE* genes that participate in membrane-transport processes (Makrides, 1996).

Another challenge is to minimise the protein degradation in the periplasm. It has been hypothesized that the misfolding and degradation of proteins results from their inefficient chaperoning to the translocase, either because they fold (or misfold) too rapidly in the cytoplasm, or because the necessary chaperones become limiting (Bergès *et al.*, 1996). A systematic search for periplasmic factors improving phage display (Bothmann & Plückthun, 1998) led to the identification of Skp/OmpH, a protein previously implicated in the folding of outer membrane proteins (Missiakas *et al.*, 1996). In contrast to specialized

periplasmic chaperones, Skp appears to be a broad substrate range chaperone, and its overexpression improves folding of a number of aggregation-prone single-chain antibody fragments (Bothmann & Plückthun, 1998).

Attempts to co-overproduce the molecular chaperones, SecB, DnaK-DnaJ and GroEL/ES, have met with variable success, and improved secretion depends heavily on the signal-sequence-mature protein combination. This suggests that the signal sequence influences secondary and tertiary structure formation in the mature region of secretory proteins, which in turn affects chaperone recognition. It may therefore be necessary to try several signal sequences and/or overproduce different chaperones to optimize the translocation of any given heterologous protein (Bergès *et al.*, 1996).

2.3.3.3 Extracellular secretion

A system that targets synthesised proteins for secretion to the culture medium presents significant advantages. Purification of the protein of interest is simplified, the culture medium provides a larger space for accumulation of the protein, and release of the protein will not result in cell death or lysis, as often occurs in high-level cytoplasmic production of recombinant proteins (Hsiung *et al.*, 1989). Unfortunately, *E. coli* normally secretes very few proteins and the manipulation of the various transport pathways to facilitate secretion of foreign proteins is a formidable task (Blight *et al.*, 1994).

Proteins that are targeted for secretion into the extracellular medium have to cross the two membranes of the cellular envelope, the cytoplasmic and outer membranes. Passage through these membranes is a highly discriminating process that allows the export of only a selected number of proteins. A set of complex biochemical pathways govern the translocation of proteins through the membranes has evolved for this purpose, making manipulated secretion of a recombinant protein a challenge that often requires extensive effort (Sandkvist & Bagdasarian, 1996).

In general, methods for protein secretion into the culture medium fall into two categories, namely the utilization of existing pathways for secreted proteins, and the use of signal sequences, fusion partners and permeabilising agents that effect protein secretion as a result of selective and limited permeability of the outer membrane (Stader & Silhavy, 1990). The first approach offers the advantage of specific secretion of the protein of interest and hence minimum contamination by nontarget proteins. Perhaps the best known example is the hemolysin gene, which has been used for construction of secreted hybrid proteins (Blight *et al.*, 1994).

The second approach relies on the induction of limited leakage of the outer membrane to cause protein secretion (Obukowicz *et al.*, 1988). One such a system uses the bacteriocin release protein (BRP) in the release of recombinant proteins. BRP is a small lipoprotein which is produced as a precursor with signal peptide and is secreted across the cytoplasmic membrane, where it is N-acetylated and inserted into the outer membrane. In the outer membrane, BRP can activate the detergent-resistant phospholipase A, resulting in the formation of permeable zones in the cell envelope, through which the target protein can pass and be released into the culture medium (Fu *et al.*, 2003). Controlled expression of BRP has been used for the release of several heterologous proteins from *E. coli* (Hsiung *et al.*, 1989).

In addition, indirect methods leading to an increase in the concentration of chromosomal heat shock-proteins have proven beneficial to the folding and secretion of certain overexpressed proteins. These methods include growing the cells at high temperatures (Goloubinoff *et al.*, 1989), mutations in negative regulators of the heat-shock response (Yura *et al.*, 1993), and co-overexpression of plasmid-encoded σ^{32} (Thomas & Baneyx, 1996).

An interesting alternative for secretion is the use of bacterial L-forms (mutants devoid of outer membrane and murein sacculus) in order to get products that

normally are targeted to the periplasm directly into the medium (Gumpert & Hoischen, 1998). In this system, polypeptides fused to a normal Sec-recognised amino-terminal signal sequence cross the cytoplasmic membrane via the Sec machinery and arrive in the extracellular space because of the absence of periplasm. Interestingly, correct folding of the proteins released into the medium takes place, probably because periplasmic enzymes such as DsdA are also released into the medium (Cornelis, 2000).

2.3.4 Host design considerations

The complete sequence of the *E. coli* chromosome and a variety of genetic tools now allow almost any modification of this host organism. For example, a method has been developed that allows the precise insertion of DNA into the chromosome without leaving a drug resistance or other marker (Bass *et al.*, 1996). An exciting application is the possibility of modifying the production cell for more efficient metabolism, for stabilising of the protein product, and for more efficient protein folding (Swartz, 2001).

2.3.4.1 Fusion proteins

Several sophisticated strategies have been developed to optimise the expression of heterologous genes that involve fusing the gene of interest downstream of a second gene to produce a fusion protein displaying the combined properties of the original gene products. Fusion can take place on either or both sides of the target gene depending on the specific application, but the majority of the described fusion protein systems place the protein of interest at the C-terminal of a well characterised fusion partner (LaVallie & McCoy, 1995).

Gene fusions are a long-accepted way of optimising expression in *E. coli*. Translation fusions were among the earliest methods used to obtain high-level expression in *E. coli*. These fusions were often of limited utility as the resulting gene products were not authentic. Fortunately, translational fusion technology has evolved to allow the production of high levels of fusion gene products which

can rapidly be converted to non-fusion authentic proteins through the use of specific proteases that remove the fusion partner (Olins, 1996). The many different applications for this generation of fusion partners include facilitated purification of the target gene, means to decrease proteolysis of the target protein, display of proteins on surfaces of bacterial cells and phages, construction of reporter molecules for the monitoring of gene expression and protein localisation, and lengthening of the circulation half-life of protein therapeutics (Nilsson *et al.*, 1997). However, as pioneered in the early 1980's, the most frequent application of gene fusions has been for the purpose of affinity purification of recombinant proteins. Fusion proteins can often be purified to near homogeneity from crude biological mixtures by a single, fusion-partner-specific, affinity chromatography step (Uhlén *et al.*, 1983).

The four main gene fusion expression systems used in *E. coli* are based on fusions to *Staphylococcus* protein A (Nilsson & Abrahmsén, 1990); *Schistosoma japonicum* glutathione-S-transferase (GST) which allows affinity purification by binding to glutathione (Smith & Johnson, 1988); *E. coli* maltose-binding protein which allows affinity purification by binding to amylose (DiGuan *et al.*, 1988); and *E. coli* thioredoxin which enhances solubility and allows affinity purification by binding to phenylarsine oxide (LaVallie *et al.*, 1993). These four systems have been the most successful in producing correctly folded and soluble heterologous proteins in the bacterial cytoplasm. The reasons for this success are probably due to the physical properties of these fusion partners allowing them to act as covalently linked chaperones and enhance the folding of the fused protein (LaVallie & McCoy, 1995).

If a fusion partner possess no property that can be exploited for purification or detection (such as GST which binds to glutathione), a peptide or polypeptide tag can be added. Probably the most commonly used are the polyhistidine (His₆) tag which allows affinity purification by binding to immobilized cobalt or nickel (Crowe *et al.*, 1994), and the FLAG (AspTyrLysAsp₄Lys) tag which is an epitope for a

monoclonal antibody and facilitates detection and affinity purification (Hopp *et al.*, 1988).

The introduction of a recognition sequence for a chemical agent or a protease between the fusion partner and the target protein allows for site-specific cleavage of the fusion protein to remove the affinity fusion partner (Nilsson *et al.*, 1997). Protease-recognition sequences have been engineered downstream of the fusion partner in many commercially available expression systems. Examples are thrombin, factor Xa, enterokinase, subtilisin, and viral proteases. However, cleavage is not always complete, and the proteases may cleave within the fused protein (Chong *et al.*, 1997).

Expressing proteins as fusions to ubiquitin offers the advantage of an often dramatic increase in yield, and the ability to produce any desired amino-terminal residue upon ubiquitin cleavage. The remarkable increase in protein yield is probably due to protection of the target protein from proteolysis, improved folding, and efficient mRNA translation (Butt *et al.*, 1989). Ubiquitin or the ubiquitin metabolic pathway is absent in prokaryotic organisms. To remove the ubiquitin moiety from fusion proteins, the ubiquitin-specific protease Ubp2 (ubiquitin-carboxy-terminal hydrolase) is coexpressed in *E. coli*, thus effecting the cotranslational cleavage of ubiquitin from the fusion protein (Baker, 1996).

One drawback of fusion protein expression is the covalent linkage of the two proteins, where the presence of the fusion partner may interfere with subsequent uses of the protein. The conditions for purification differ from system to system, and the environment tolerated by the target protein is an important factor for deciding which affinity fusion partner to choose. In addition, other factors including protein localization, costs of the affinity matrix and buffers, and the possibilities of removing the fusion partner by site-specific cleavage should also be considered (Nilsson *et al.*, 1997).

2.3.4.2 Molecular chaperones

It is well established that the efficient posttranslational folding of proteins, the assembly of polypeptides into oligomeric structures, and the localization of proteins are mediated by a universally conserved class of proteins termed molecular chaperones. Many chaperones are also stress or heat-shock proteins, whose rate of synthesis accelerates under various protein-damaging conditions (Clarke, 1996).

The major chaperone systems present in the cytoplasm of *E. coli* are the DnaK-DnaJ-GrpE and GroEL-GroES folding machines (Hartl, 1996). The co-overexpression of components of either set of the major chaperones can significantly improve the solubility and/or secretion of many structurally and functionally unrelated recombinant polypeptides (Wall & Plückthun, 1995). The demonstration that efficient production and assembly of prokaryotic ribulose biphosphate carboxylase in *E. coli* required both GroES and GroEL proteins, led to an increasing interest in the use of molecular chaperones for high-level expression (Goloubinoff *et al.*, 1989). Since then a number of studies reinvestigated this approach with respect to the folding of various other proteins. Interestingly, this chaperone-mediated improvement in folding ultimately led to a smaller increase in yield than engineering of the protein for better folding (Caspers *et al.*, 1994). Interestingly, cosecretion of ATP-independent chaperones (DnaJ) together with the use of low-molecular-size medium additives to the culture medium, can further dramatically increase the yield of native eukaryotic proteins with complex disulphide patterns in the periplasm of *E. coli* (Schäffner *et al.*, 2001).

Three bacterial chaperones are thought to participate in the folding of newly translated polypeptides in the cytosol, namely trigger factor (TF), DnaK/DnaJ and the chaperonin GroEL/ES. Studies have shown that TF appears to be the first player in the folding of nascent chains, recognizing relatively short hydrophobic stretches and protecting them from aggregation. DnaK/DnaJ can then bind to

longer chains and allow larger polypeptides to fold (Albanèse & Frydman, 2002). Disassociation of DnaK and DnaJ from the polypeptide is facilitated by protein GrpE, which triggers nucleotide exchange. TF and DnaK/DnaJ seem to have particularly overlapping functions, and mutants that are defective in the chaperone functions of both proteins exhibit a synthetic lethal phenotype. Finally, GroEL/ES functions post-translationally to assist folding of a subset of cytosolic proteins (Hartl & Hayer-Hartl, 2002).

While direct co-overexpression of components of the DnaK-DnaJ-GrpE and GroEL-GroES molecular chaperone machines clearly can improve the proper folding or secretion of a number of recombinant proteins in *E. coli*, chaperone overproduction remains ineffective in other cases (Georgiou & Valax, 1996). Possible explanations for this phenomenon include the following: (i) an incorrect choice of the overproduced chaperone(s); (ii) a need for additional cofactors which may not have been identified to date; or (iii) a requirement for the cooperative or network action of several chaperones systems (Thomas & Baneyx, 1997).

2.3.4.3 Codon usage

All species display a bias in the non-random usage of the 61 available amino acid codons. In each cell, the tRNA population closely reflects the codon bias of the mRNA population (Ikemura, 1981). Cloned heterologous genes often differ markedly in codon bias, and even in the crude relative amino acid content of their encoded proteins, from that of their host (Kane, 1995).

Within codon families, *E. coli* shuns certain codons such as AGG/AGA/CGA (arginine), AUA (isoleucine), CUA (leucine), and CCC (proline) that mammalian genes use more lavishly. Therefore, the high-level expression of a heterologous gene may place demands on the host protein synthesis apparatus that are not matched to its normal tRNA population (Kurland & Gallant, 1996). This may lead to translational errors as a result of ribosomal stalling at a position requiring

incorporation of amino acids coupled to minor tRNAs, or even at sites requiring major tRNAs, but which are depleted because of overutilisation of a particular amino acid. The mistranslational events related to rare tRNAs are observed as codon misreadings and as processing errors, which manifests themselves as amino acid substitutions or frameshift events (McNulty *et al.*, 2003).

Rare codons often are often grouped together in clusters. One such a subset of codons, namely AGG/AGA, AUG, CUA, CGA and CCC, was shown to reduce both the quantity and quality of the synthesised protein. It appears that these clusters cause most of the translational errors, although simply the presence of a large number of these codons may introduce translational errors as well (Kane, 1995). Furthermore, it appears that the presence of rare codons near the 5' end of a transcript affects translational efficiency, and the GC content of the 5' coding region of certain genes appears to influence expression (Pedersen-Lane *et al.*, 1997).

Two alternative strategies have been developed to minimise the effects of preferential codon usage in *E. coli*. The first relies on genetically altering rare codons in the target gene, without modifying the encoded protein product, in order to reflect the specific codon bias of the host (Pedersen-Lane *et al.*, 1997). In the second strategy, the intracellular tRNA pool is expanded by coexpressing genes that encode rare tRNAs such as the *argU* (*dnaY*) gene that encodes the minor tRNA^{Arg(AGG/AGA)} (Andrews *et al.*, 1996). However, results from several studies employing this approach have been inconsistent, and unambiguous guidelines have not been established to correlate codon usage and translation of a transcript (Makrides, 1996).

Expression trails of wild-type rare codon genes in various modified *E. coli* strains are the recommended first choice, while a fusion gene approach is a considerable robust alternative preventing problems encountered by rare codon or specific RNA secondary structure formation. The costly and tedious synthesis

of synthetic genes appears to be the least desirable alternative for production of heterologous proteins (Wu *et al.*, 2004).

2.3.4.4 Stress response

The production of foreign proteins in *E. coli* is a challenging, complex and dynamic process. The accumulation of misfolded proteins, often due to the overexpression of heterologous genes, results in a rapid stress response. A key feature of this response is increased protease activity (Harcum & Bentley, 1993). In addition, the strong competitive effect of product synthesis on the synthesis of the housekeeping host proteins leads to malfunction of the protein synthetic machinery, increased plasmid instability, (Gill *et al.*, 2000), and a change in the central carbon and energy metabolism resulting in growth inhibition and acetate formation (Jürgen *et al.*, 2000).

This response is generally referred to as the metabolic burden associated with the production of nonessential proteins (Glick, 1995). The phenotypic characteristics of this response are similar to the responses *E. coli* have developed to heat shock and amino acid limitation (Andersson *et al.*, 1996). Interestingly, the stress response of *E. coli* includes an increase in genetic variation (transposition events) which is presumed to provide the host with a tool for adaptation to environmental changes, a feature not desired when maintaining reproducibility among repeated fermentations (Taddei *et al.*, 1997).

In response to heterologous protein overproduction, a variety of regulatory pathways are effected, each of which contributes to the reduced cell growth rate. These include the SOS (Lee *et al.*, 2002), stationary phase, and heat-shock responses (Dong *et al.*, 1995); the bacteriophage λ life cycle (Glick, 1995); and transcription from a transposition-related gene (Taddei *et al.*, 1997). It is further suggested that *E. coli* cells communicate the burden associated with overexpressing genes through cell-cell communications or quorum sensing (DeLisa *et al.*, 2001). A tremendous overlap between the recombinant protein-

induced response and other characterized responses thus clearly exists (Gill *et al.*, 2000).

2.3.5 Refolding recombinant proteins

Overexpression of recombinant proteins in bacteria, and other host systems such as yeast or higher eukaryotes, often leads to the accumulation of the protein product in inactive insoluble deposits inside the cell. These inclusion body (IB) aggregates comprise of dense, amorphous protein deposits that can be found both in the cytoplasmic and periplasmic space of bacteria (Gribskov & Burgess, 1983).

Inclusion bodies are formed because the expression of recombinant proteins in *E. coli* entail the creation of an unnatural situation where folding intermediates are present at very high concentrations, which results in a greater tendency to aggregate. Many overexpressed proteins exist in their natural environments only in low amounts, and thus little evolutionary pressure may have existed to optimise their sequences for efficient folding. Moreover, recombinant proteins are often expressed at levels that are orders of magnitude higher than their normal expression levels (Wall & Plückthun, 1995). There is thus no direct correlation between the propensity of a certain protein to aggregate and its intrinsic properties such as molecular weight, hydrophobicity and folding pathways. Only in the case of disulphide bonded proteins can IB formation be anticipated if the protein is produced in the bacterial cytosol, as formation of disulphide bonds does usually not occur in this reducing cellular compartment (Lilie *et al.*, 1998).

Faced with the prospect of producing an insoluble and inactive protein, researchers usually attempt to improve solubility by manipulating the parameters known to influence inclusion body formation. These include the transcription rate of the gene of interest, the growth temperature, the composition and pH of the culture medium, and the cellular localization of the overexpressed protein

(Hockney, 1994). Alterations in the intracellular concentration of folding modulators can also have a significant impact on the folding of many recombinant gene products, and researchers will often coexpress the protein of interest with chaperones and foldases, or use solubilising fusion partners (Wall & Plückthun, 1995).

Importantly, the expression of heterologous proteins in inclusion bodies form has certain advantages. Large amounts of highly enriched proteins can be expressed as IBs, and trapped in insoluble aggregates, these proteins are for the most part protected from proteolytic degradation by host cell enzymes. If the protein of interest is toxic or lethal to the host cell, then IB expression may be the best available production method. The IBs can readily be separated from the soluble proteins of the host cells by differential centrifugation, filtration or size-exclusion chromatography, which also provide a useful concentration and purification step. Furthermore, expression of the target protein as IBs can be directly observed by phase contrast microscopy avoiding the need for initial identification by electrophoresis after cell disruption (Carrió & Villaverde, 2002). The challenge is to take advantage of the high-expression levels of IB proteins by being able to convert inactive and misfolded IB proteins into soluble bioactive products (Misawa & Kumagai, 1999).

The general strategy used to recover active protein involves the following steps: (i) inclusion body isolation and washing; (ii) solubilisation of the aggregated protein, which causes denaturation; and finally (iii) refolding of the solubilised protein. While the efficiency of the first two steps can be relatively high, folding yields may be limited by the production of inactive misfolded species as well as aggregates (Mukhopadhyay, 1997).

2.3.5.1 Inclusion body isolation, purification and solubilisation

Methods for preparing denatured protein involve isolation of the inclusion bodies with some removal of contaminants, followed by solubilisation using concentrated chemical denaturants (Tsumoto *et al.*, 2003).

Cells containing IBs are typically disrupted by high-pressure homogenization or a combination of mechanical and chemical methods. Occasionally a lytic enzyme, such as lysozyme, may be added before cell disruption to increase efficiency and reduce power requirements. The resulting suspension is then treated by either differential low-speed centrifugation or filtration to separate the dense IBs from the lighter cell-membrane components and soluble contaminants (Georgiou & Valax, 1999). Washing steps are performed with buffers containing EDTA, and either low concentrations of chaotropic agents such as urea or guanidinium chloride (GdmCl), or detergents such as Triton X-100 (Cardamone *et al.*, 1995), sodium deoxycholate or octylglucoside (Burgess, 1996).

This washing step is designed to remove contaminants such as membrane-associated proteins that are released upon cell breakage, and other proteins that have adsorbed onto the hydrophobic IBs during processing and could affect protein refolding yield. Alternatively, sucrose gradient centrifugation may be performed to purify IBs and separate them from other cellular components. After washing, IBs are solubilised using a variety of strong denaturants and/or detergents (Middelberg, 2002). The most commonly used solubilising agents are denaturants such as urea, GdmCl (Misawa & Kumagai, 1999), and thiocyanate salts. Using these denaturants, solubilisation is accomplished by the complete disruption of the protein structure (unfolding) or by the disruption of intermolecular interactions with partial unfolding of the protein (De Bernardez Clark, 2001).

Detergents commonly used to solubilise IBs are sodium dodecyl sulfate (SDS) (Patra *et al.*, 2000), *n*-cetyl trimethylammonium bromide (CTAB) (Cardamone *et*

al., 1995), sarkosyl (Burgess, 1996), and sodium n-laurosyl sarcosine (Kurucz *et al.*, 1995). Detergents offer the advantage that the solubilised protein may already display biological activity, thus avoiding the need for a refolding step. If this is the case, it is important to remove contaminating membrane-associated proteases in the IB washing step to avoid proteolytic degradation of the solubilised IB protein (Georgiou & Valax, 1999). One drawback of the use of detergents as solubilising agents is that they may interfere with downstream chromatographic steps, and therefore it essential to remove all solubilising detergents (Middelberg, 2002).

A key to the solubilisation process is the addition of a reducing agent to maintain cysteine residues in the reduced state and thus prevent non-native intra- and inter-disulphide bond formation in highly concentrated protein solutions at alkaline pH. Typically used reducing agents are dithiothreitol (DTT), dithioerythritol (DTE), and β -mercaptoethanol. These reducing agents are added in slight excess to ensure complete reduction of all cysteine residues (Lilie *et al.*, 1998). Temperatures above 30°C are typically used to facilitate the solubilisation process. A chelating agent such as EDTA can be included in the solubilisation buffer to scavenge metal ions, which could cause unwanted oxidation reactions. Solubilisation can also be accomplished by the addition of acids such as 70% formic acid (De Bernardez Clark, 1998).

2.3.5.2 Renaturation and refolding of the solubilised protein

After the inclusion bodies have been solubilised, renaturation and refolding is then accomplished by the removal of excess denaturants by either one of three methods. These are dilution, a buffer-exchange step (such as dialysis, diafiltration, gel-filtration chromatography), or immobilisation onto a solid support (Tsumoto *et al.*, 2003).

Dilution of the solubilised protein directly into an appropriate renaturation buffer is the most commonly used method in small-scale refolding studies because of its

simplicity. The main disadvantages of dilution refolding for commercial applications are the need for larger vessels and additional concentration steps after renaturation (Mukhopadhyay, 1997). The key to successful dilution refolding is to control the rate of the addition of denatured protein to renaturation buffer and to provide good mixing. An intermediate concentration of denaturants should be maintained to induce folding while the solubility and flexibility of the proteins molecules are maintained and aggregation prevented (Tsumoto *et al.*, 2003). Dilution refolding can also be accomplished in multiple steps, also known as pulse renaturation, in which aliquots of denatured reduced protein are added to renaturation buffer at successive time intervals (Lilie *et al.*, 1998), or semicontinuously via fed-batch addition of the denatured reduced protein to refolding buffer (Katoh & Katoh, 2000).

Buffer exchange to remove high denaturant concentrations can also be accomplished by diafiltration (Varnerin *et al.*, 1998) and dialysis (West *et al.*, 1998) using ultrafiltration membranes. Renaturation yields using these membrane-based methods may be significantly affected by protein binding to the membranes. Binding can be minimised by using highly hydrophilic materials, such as cellulose acetate, which are more compatible with unfolded protein molecules. With typical hydrophobic membrane materials, such as polyether sulfone, the majority of the denatured proteins bind to the membrane and significant losses of unfolded protein occur via transmission through the membrane (West *et al.*, 1998). These problems have led to the increased interest in size-exclusion chromatography (SEC) as an alternative buffer-exchange method to remove high denaturant concentrations and promote renaturation. SEC restricts diffusion of various protein forms in the refolding mixture, thereby facilitating the separation of correctly folded and aggregated species (Li *et al.*, 2004).

Buffer exchange to remove high denaturant concentrations is also achievable through reversible adsorption of the denatured proteins onto a solid support.

Intermolecular interactions leading aggregation are minimised when the refolding molecules are isolated through binding to the support. Freedom for structure formation during renaturation is facilitated by binding through fusion partners such as a His-tag (Rogl *et al.*, 1998) or the cellulose-binding domain (Berdichevsky *et al.*, 1999), which retain their binding capabilities under denaturing conditions required for loading the solubilised IB protein onto the column. *In situ* purification is achieved by washing the bound protein before elution (Middelberg, 2002).

2.4 THROMBOSIS AND ANTITHROMBOTICS

The experimental part of this study involves the isolation of a thrombin inhibitor from a phage displayed cDNA library derived from the haematophagous louse fly *Hippobosca rufipes*, and the expression of the second C-terminal CUB domain from ADAMTS-13. Therefore a brief background, highlighting the role of and importance these elements, will be given.

2.4.1 The search for antithrombotic agents

The development of a specific inhibitor for a single coagulation factor could reduce side effects and improve the therapeutic profile. The structure of the active site in all these proteases is very similar, which makes the development of a specific small molecule active site inhibitor challenging (Weitz & Hirsh, 2001). Although relatively nonselective with respect to small chromogenic substrates, these proteases are highly specific for their natural macromolecular substrates. In order to achieve this, exosites on these enzymes play an important role in substrate recognition and catalysis. Blocking such important interactions could result in the specific inhibition of a single protease in this pathway (Baugh *et al.*, 2000).

Progress in molecular biology techniques has stimulated interest in the structure and function of thrombin. It has improved the understanding of the central role of

thrombin in thrombogenesis, and clarified the molecular events of inhibitor binding. This development has resulted in the production of recombinant hirudins and hirudin analogous (Knapp *et al.*, 1992). It has also allowed the molecular design of synthetic antithrombins, and encouraged the development of these products for clinical use (Weitz & Hirsh, 2001).

2.4.2 Thrombin and thrombin inhibitors

Thrombin is a multifunctional trypsin-like serine protease that exhibits both pro- and anti-coagulant activity. The catalytic activity of thrombin is regulated physiologically by serpins such as antithrombin III, heparin cofactor II, protease nexin I, and by the general protease scavenger α_2 -macroglobulin (Tulinsky, 1996). Thrombin activates platelets, promotes its own generation by activating factors V, VIII, and XI, and converts fibrinogen to fibrin, where it remains enzymatically active and relatively protected from inactivation by fluid-phase inhibitors (Weitz *et al.*, 1990).

Thrombin plays a central role in blood coagulation and in the activation of various cell types including platelets, endothelial cells, fibroblasts, glomerular epithelial and mesangial cells, and smooth tissue cells (Fenton *et al.*, 1998). So crucial is the role played by this serine protease that the inappropriate activity of it ultimately leads to thrombosis, whether through venous thromboembolism, myocardial infarction or stroke. The number and type of intrinsic and extrinsic natural mechanisms of targeting thrombin that have evolved, validate thrombin as an important physiological target, and provide strategies to knock it out (Huntington & Baglin, 2003).

Inhibitors of thrombin can be classified as either direct or indirect. Indirect thrombin inhibitors (heparins and vitamin K antagonists) block the generation and action of thrombin, either by activating naturally occurring thrombin inhibitors, or by inactivating specific factors in the coagulation system that subsequently impact on thrombin generation or activity. In contrast, direct thrombin inhibitors

(hirudin, bivalirudin, argatroban and melagatran) achieve their anticoagulant effect by directly binding to the thrombin molecule, thereby preventing it from interacting with its substrates (Agnelli & Sonaglia, 1999).

Protease inhibitors directed mainly at thrombin and factor Xa are found widespread in the saliva of haematophagous animals. These protease inhibitors have been used as models to study physiological and pathophysiological pathways, and develop selective strong anticoagulant inhibitors (Stark & James, 1996). Several studies have been performed by mutating inhibitors (Tanaka *et al.*, 1999) or by designing new molecules based on the three dimensional structures of classical protease inhibitors (Knapp *et al.*, 1992).

2.4.3 Antithrombotic compound from haematophagous animals

Antithrombotic compounds found in haematophagous animals can be divided into five groups: (i) antithrombotic agents, (ii) inhibitors and activators of the prothrombinase complex by directly inhibiting factor Xa, (iii) substances that affect platelet function, (iv) substances that affect the fibrinolytic mechanism, and (v) a group of miscellaneous agents whose activities are difficult to group together (Arocha-Pinango *et al.*, 1999).

The most prominent natural anticoagulant is hirudin, a 65 amino acid single-chain polypeptide with high specificity toward thrombin. Hirudin was isolated from the salivary glands of the European medicinal leech *Hirudo medicinalis* for the first time in the 1950s (Markwardt, 1957). In the years following, the isolation and purification were refined and the complete amino acid structure was determined (Markwardt & Walsmann, 1967). Markwardt used hirudin as an antithrombotic drug in the 1970s, but its evaluation in clinical trials was delayed until large amounts could be produced using recombinant DNA technology (Stone & Hofsteenge, 1986). Throughout the years, hirudin has served as a standard for developing naturally occurring coagulation inhibitors into anticoagulant drugs. Today the recombinant forms of hirudin, Lepirudin (Refludan™) and Desirudin

(Revasc™), are FDA-approved anticoagulants which are administered to patients with heparin-induced thrombocytopenia or thrombosis (Hirsh, 2003).

Other naturally occurring thrombin inhibitors derived from haematophagous animals are summarised in table 1. In the strategies employed by haematophagous invertebrates to overcome host haemostatic systems, thrombin is one of the most important targets (Dodt *et al.*, 1996).

Table 2.1 Thrombin inhibitors derived from haematophagous animals

Inhibitor	Organism	Species	Reference
Hirudin	European medical leech	<i>Hirudo medicinalis</i>	Stone & Hofsteenge, 1986
Haemadin	Indian land-living leech	<i>Haemadipas sylvestris</i>	Strube <i>et al.</i> , 1993
Theromin	Rhynchobdellid leech	<i>Theromyzon tessulatum</i>	Salzet <i>et al.</i> , 2000
Dipetalogastin	Reduviid bug	<i>Dipetalogaster maximus</i>	Lange <i>et al.</i> , 1999
Dipetalogastin II	Reduviid bug	<i>Dipetalogaster maximus</i>	Van de Locht <i>et al.</i> , 1995
Rhodniin	Triatomine bug	<i>Rhodnius prolixus</i>	Friedrich <i>et al.</i> , 1993
Triabin	Triatomine bug	<i>Triamota pallidipennis</i>	Noeske-Jungblut <i>et al.</i> , 1995
Ornithodorin	Soft tick	<i>Ornithodoros moubata</i>	Van de Locht <i>et al.</i> , 1996
Savignin	Soft tick	<i>Ornithodoros savignyi</i>	Mans <i>et al.</i> , 2002
Americanin	Lone star tick	<i>Amblyomma americanum</i>	Zhu <i>et al.</i> , 1997
Thrombostasin	Horn fly	<i>Haematobia irritans</i>	Zhang <i>et al.</i> , 2002
Anophelin	Mosquito	<i>Anopheles albimanus</i>	Valenzuela <i>et al.</i> , 1999
Tsetse thrombin inhibitor (TTI)	Tsetse fly	<i>Glossina morsitans morsitans</i>	Cappello <i>et al.</i> , 1996

2.4.4 Platelets, von Willebrand factor and ADAMTS-13

Platelets carry the burden of initiating primary plug formation through accumulating and adhering to the site of vessel wall injury, and setting the complex cascade of blood coagulation into action. However, if something in the delicate maintenance of the sequence of events goes array, the potential for the formation of unwanted aggregates and thrombi, resulting in ischemic occlusion,

becomes a serious problem. Platelets are thus role players in both physiologic haemostasis and pathologic thrombosis (Wagner & Burger, 2003).

The glycoprotein von Willebrand factor (vWF) plays a critical role in capturing the circulating platelets to the site of injury, thereby initiating coagulation and arresting bleeding. This is mediated by vWF forming a bridge between collagen and other components in the damaged vessel wall, and the glycoprotein receptors (GpIb and GpIIb-IIIa) on the platelets (Ruggeri, 2000). Mature vWF is generated in the endothelial cells and released as “unusually large” UL-vWF or alternatively remain attached to the subendothelium as long string-like aggregates. The larger multimers have an increased thrombotic potential since they have more available binding sites, and are thus more potent in binding the platelet glycoproteins. The size of vWF multimers therefore has to be physiologically regulated to prevent unwanted thrombus generation (Tsai, 1996).

In 1996, the enzyme responsible for the proteolysis of vWF was identified as a metalloprotease which cleaves the peptide bond between amino acid residues Tyr842 and Met843 within the A2 domain of vWF (Furlan *et al.*, 1996). Two groups, independently and simultaneously, succeeded in 2001 in purifying plasma-derived vWF cleaving protease in quantities sufficient to obtain partial amino acid sequences (Fujikawa *et al.*, 2001; Gerritsen *et al.*, 2001). The protease was identified as a member of the ADAMTS (a disintegrin-like and metalloprotease with thrombospondin type-1 motifs) family, and designed the name ADAMTS-13. It is similar in structure to other ADAMTS proteases, but with unique features that might suggest distinct modes of ligand binding. ADAMTS-13 is the only known family member that contains CUB domains, it has a shorter pro-domain, and it lacks two of the three cysteine residues common to others in the ADAMTS family (Zheng *et al.*, 2001).

CUB (complement subcomponents Clr/Cls, Uegf, Bmpl) domains are found widespread in developmentally regulated proteins of higher eukaryotes, were

they are generally involved in protein-protein and protein-carbohydrate interactions (Bork & Beckmann, 1993). The physiological relevance of the C-terminal CUB domains of ADAMTS-13 is, however, still uncertain. Recently it was found that the CUB domains are not required for vWF cleavage proteinase activity measured under static conditions (Soejima *et al.*, 2003), and that certain mouse strains possess a variant form of murine ADAMTS-13 that lacks the CUB domains (Banno *et al.*, 2003), supporting the idea that these domains are dispensable *in vivo*. Another study hinted at a possible role played by the domains in the secretion of newly synthesised ADAMTS-13 (Pimanda *et al.*, 2003). On the other hand, a functional role for the CUB domains have been suggested by a report that peptides from the CUB domains inhibit vWF cleaving proteinase activity under flow, but not static, conditions (Bernardo *et al.*, 2003).

CHAPTER 3

MATERIALS AND METHODS

3.1 Construction of cDNA library

3.1.1 Isolation of total RNA from *Hippobosca rufipes*

Live *Hippobosca rufipes* specimens were collected from horses from a stable in South Africa, under the supervision of Prof. S.M. Meiring. Whole specimens were frozen and stored at -70°C . From the frozen *H. rufipes* material, 3 ml was thawed in a sterile 50 ml tube. To this 30 ml TRIZOL[®] reagent (Amersham Biosciences, Buckinghamshire, UK) was added, and the mixture was thoroughly homogenised using a Polyton[®] PT 1200 homogeniser (Kinematica AG, Luzern, Switzerland). The sample was then incubated for 5 min at room temperature. To allow the dissociation of nucleoprotein complexes, 6 ml chloroform was added, the tubes vigorously vortexed for 15 s, and incubated for 3 min at room temperature. The sample was next centrifuged for at 12000 g for 15 min at 4°C to separate the different phases of the sample. The aqueous phase was transferred to a new tube and 15 ml isopropyl alcohol was added to precipitate the RNA. The sample was incubated for 10 min at room temperature and centrifuged at 12000 g for 10 min at 4°C . The supernatant was removed and the pellet washed by adding 30 ml 75 % ethanol, vortexing, and centrifuging at 7500 g for 5 min at 4°C . After decanting the supernatant, the pellet was air-dried for 10 min. The RNA was finally dissolved in 50 μl RNase-free water.

3.1.2 Generation of mRNA

Poly(A)⁺ mRNA was isolated from total RNA with the Dynabeads[®] mRNA Purification kit (Dyna, Oslo, Norway) following the manufacturer's instructions. The system is designed for the rapid isolation of highly purified, intact mRNA from eukaryotic total RNA, with the use of superparamagnetic polymer spheres called Dynabeads. Efficient mRNA isolation relies upon base-pairing between the 25-nucleotide long chains of deoxy-thymidylate (dT) residues that are

covalently coupled to the surface of Dynabeads oligo(dT)₂₅, and the poly(A) at the 3'-end of the mRNA.

3.1.3 cDNA synthesis

cDNA was synthesised from the mRNA with the cDNA Synthesis Kit (Boehringer Mannheim Biochemica, Mannheim, Germany) according to the manufacturer's instructions. RNA was transcribed into cDNA essentially according to the Gubler-Hoffmann method (Gubler & Hoffman, 1983). First-strand synthesis started at the 3'-end of the poly(A)⁺ containing mRNA by using the oligo(dT) primer, and second-strand synthesis took place using the mRNA/DNA as substrate. Mild treatment with RNase H inserted nicks into the RNA, providing 3'OH-primers for DNA polymerase I present in the reaction. The 5'→3' exonuclease activity of DNA polymerase I removed the primer stretches in the direction of synthesis, which were then replaced with new nucleotides by the polymerase activity. T4 DNA polymerase removed any overhanging 3'-ends, to ensure that the cDNA was blunt-ended for further cloning steps.

The cDNA was precipitated by adding 100 µl phenol to 100 µl (~4 µg) cDNA, vortexing, and centrifuging for 1 min at 13000 *g*. The supernatant was transferred to a new tube, to which 10 µl (1/10 volume) 3 M NaAc and 200 µl (2× volume) 100 % ethanol was added. The sample was incubated at -20°C for 30 min, centrifuged for 10 min at 13000 *g*, and the supernatant discarded. The pellet was washed by adding 300 µl 70 % ethanol and centrifuging at 13000 *g* for 1 min. The supernatant was discarded and the pellet dried for 2 min in a Savant SpeedVac[®] (Thermo Electron Corporation).

To fragment the cDNA and generate compatible 5' sticky ends, the cDNA was digested with the *Sau3A* restriction enzyme (Boehringer Mannheim Biochemica). The following components were added in a microcentrifuge tube: 48 µl (4 µg) cDNA, 1 µl *Sau3A*, and 6 µl SuRE/cut[™] Buffer A (Boehringer Mannheim Biochemica), to a final volume of 60 µl. After incubation at 37°C for 15 min the

reaction was returned to ice. A 2 µl aliquot of the digestion reaction was analysed by gel electrophoresis, using 2 µl XIV molecular weight marker (Roche Molecular Biochemicals, Mannheim, Germany) as a molecular standard. Electrophoresis was carried out in TBE electrophoresis buffer on a 0.8 % agarose gel containing 0.5 µg/ml ethidium bromide.

3.1.4 Construction of cDNA phage display library

A cDNA phage display library was constructed using the EZnet™ phage display cDNA library construction kit PDL-5001 (Maxim Biotech, San Francisco, USA) following the manufacturer's instructions. The multiple cloning site of the pHage 3.2 phagemid vector contains a M13 gene III leader sequence, which allows the cDNA/gene III fusion products to be displayed on the M13 filamentous phage particle tip. The pHage 3.2 vector was digested with *Bgl*II to generate compatible sticky-ends. The vector was then dephosphorylated, purified, and precipitated. The fragmented cDNA inserts were ligated into the vector at a 2:1 ratio, in a 15 µl ligation reaction consisting of 1 µl T4 ligase (Promega, Madison, USA), 1.5 µl ligation buffer (Promega) and 1.5 µl ATP. The ligation reaction was used to transform *E. coli* TG1 cells (K12 $\lambda(lac-proAB)$ *supE thi hsdD5 / F' traD36 proA⁺B lac^f lacZ[?]M15*) by electroporation. The primary library was amplified in a TG1 culture in liquid 2×YT medium before it was superinfected with M13K07 helper phage (Pharmacia Biotech, New York, USA). The superinfection and packaging in M13 further amplified the cDNA library.

3.1.5 Direct colony polymerase chain reaction (PCR)

The presence of the cDNA library insert in the phagemid vector was established by direct colony PCR performed on recombinant TG1 colonies. To select for ampicillin-resistant transformants, aliquots of the transformation reaction were spread on sterile pre-warmed 2×YT agar plates containing 100 µg/ml ampicillin (Roche Molecular Biochemicals) and incubated overnight at 37°C. The following day, 20 single colonies were randomly picked and each was added to 20 µl MQ

water, vortexed vigorously, incubated for 5 min at 96°C, and then centrifuged for 1 min at 13000 *g*. To serve as template for the PCR reaction, 10 µl supernatant was added to the amplification reaction containing 0.2 µM of the forward primer (G3F: 5'-ATTACCTCGAAAGCAAGCTG-3'), 0.2 µM of the reverse primer (G3R: 5'-ACCCTCATAGTTAGCGTAACG-3'), 1.5 units of *Taq* DNA polymerase, 0.2 mM of each dNTP, 10× Mg-free buffer (1× final concentration), and 1.5 mM MgCl₂. The primers were custom ordered from Roche Molecular Biochemicals, and Promega supplied all the other reagents. Two negative control reactions were included, one from which *Taq* DNA polymerase was omitted, and the other containing MQ water instead of template.

The PCR amplification was carried out on a Gene Amp PCR System 2400 (Applied Biosystems, CA, USA) under the following conditions: 96°C denaturation for 2 min, followed by 25 cycles of 94° for 1 min, 57°C for 2 min and 72°C for 3 min. After cycling the reaction was held at 72°C for a further 10 min and then maintained indefinitely at 4°C. Gel electrophoretic analysis of the PCR product was performed on a 0.8 % agarose gel as described earlier.

The cDNA library was amplified by inoculating a sample of the same colonies used for PCR, into 3 ml 2×YT medium supplemented with ampicillin, and incubating at 37°C with shaking overnight. The next day, the overnight culture was used to prepare glycerol cell stock by adding 100 µl 75 % glycerol to 300 µl culture, mixing well, and storing at –70°C until further use.

3.2 Selection of thrombin-binding phages

3.2.1 Preparation of TG1 cultures

A primary streak plate of recombinant TG1 cells was created by spreading 5 µl TG1 glycerol cell stock on a sterile pre-warmed 2×YT agar plate containing 100 µg/ml ampicillin and incubating the plate overnight at 37°C. To prepare a

working stock plate, a single colony was picked from the overnight primary streak plate and inoculated onto a M9 Minimal agar plate (Sambrook & Russell, 2001), supplemented with 0.4 % glucose, and 0.01 % thiamine. After overnight incubation at 37°C, a single colony was picked from the working stock plate, inoculated into 5 ml 2×YT media containing 50 µg/ml ampicillin, and incubated at 37°C with shaking until the OD₆₀₀ was between 0.4 and 0.6. This log-phase liquid culture was then stored at 4°C for up to a week for and used as needed in the subsequent biopanning experiments.

3.2.2 Biopanning against α -thrombin

Human α -thrombin was a kind gift from Dr. Pöttsch, Kerckhoff Klinik (Bud Nauhem, Germany). The inside of Maxisorb™ immune-tube (Nalge Nunc International, Roskilde, Denmark) was coated with 1 mg α -thrombin by adding a solution of thrombin and PBS (1 mg thrombin/1 ml PBS) to the immune-tube, and rotating the tube for 1 h at room temperature, followed by overnight incubation at 4°C. The following day, the coating solution was discarded and the immune-tube blocked with 4 % skimmed milk (SM) in a PBS solution, rotated for 2 h at room temperature, and washed 3 times with PBS, 0.05 % Tween-20. Thereafter 2×10^{12} phages were added, the tube rotated for 4 h at room temperature. The non-binding phages removed and kept as the “input” phages. After washing the tube 4 times with PBS, 0.05% Tween-20, 1 ml pre-chilled log phage TG1 cells were added, and the tube was incubated at 37°C for 30 min. Following incubation, these thrombin-binding “output” phages were divided into 10 µl and 500 µl aliquots. Dilutions of the 10 µl aliquot of the “output” phages were made in 2×YT media up to 10^{-4} , and 100 µl of each dilution was plated out on 2×YT plates supplemented with 100 µg/ml ampicillin and 1.8 mg/ml glucose, and incubated overnight at 37°C.

The “output” phages were amplified by adding 500 µl “output” phages to 10 ml 2×YT media containing 100 µg/ml ampicillin and 1.8 mg/ml glucose. Immediately

thereafter, 5×10^9 pfu (300 μ l) M13K07 helper phages (Pharmacia Biotech) were added to the “output” phages, followed by incubation at 37°C for 1 h with gentle agitation. The culture was centrifuged for 10 min at 4000 g and the supernatant was discarded. The pellet was resuspended in 40 ml 2 \times YT media containing 100 μ l protease inhibitor cocktail, 50 μ g/ml ampicillin and 50 μ g/ml kanamycin, and incubated at 37°C overnight with shaking. The protease inhibitor cocktail was used to prevent degradation of the displayed peptides by proteases.

The following day, the cell cultures were centrifuged at 23000 g for 20 min at 4°C to remove the TG1 *E. coli* cells. The supernatant was transferred to a new tube, and the phages in the supernatant were precipitated with the addition of 20% polyethyleneglycol/sodium chloride (0.03 M PEG, 2.5 M NaCl) for 2 h. Following centrifugation for 20 min at 23000 g at 4°C, the phage pellet was dissolved in 1 ml PBS to which 20 μ l protease cocktail was added. The phage concentration was determined by measuring the OD₂₆₀ and calculating the phage concentration (phages/ml) as follows:

$$\text{phages/ml} = \text{OD}_{260} \times \text{dilution} \times \text{constant} \times 2$$

where the constant is $2.214 \times 10^{11} \times 2$, since the OD₂₆₀ of $2.214 \times 10^{11} \times 2$ phages is equal to 1.

Each time 2×10^{11} purified phages were used for the next round of selection. Four rounds of selection were done in total.

3.2.3 Global ELISA

A global enzyme-linked immunoabsorbent assay (ELISA) was performed in duplicate on the amplified phages from each round of panning. The starting phage concentration was 5×10^{10} which was then diluted 1:2 to a final concentration of 7.5×10^8 .

One half (48 wells) of a 96-well Nunc-Immuno™ Maxisorp™ surface plate (Nalge Nunc International) was coated overnight with 100 μ g/ml α -thrombin at 4°C in a

humidified container. The plate was blocked with a 4 % SM solution for 2 h at room temperature, and washed 3 times with PBS, 0.1 % Tween-20. From each panning round, 5×10^{10} phages were added in a final concentration of 2% SM to the first well of each column of both the coated and non-coated (control) half of the ELISA plate, and diluted 7 times 1:2 into the remaining wells of the respective columns. No phages were added to the last well of each column, which served as a negative control. After incubation at room temperature for 2 h, the plate was washed 6 times with PBS, 0.1 % Tween-20. A 1:5000 dilution of a polyclonal anti-M13 phage horseradish peroxidase (HRP)- conjugated antibody (Amersham Biosciences) was added, and the plate incubated at room temperature for 1 h. After 9 wash steps with PBS, 0.1 % Tween-20, 100 μ l *ortho*-phenylenediamine dihydrochloride (OPD) (Sigma) and 100 μ l peroxidase (H_2O_2) were added per well. After incubation room temperature for 10 min, the reaction was stopped with the addition of 30 μ l 3 M H_2SO_4 , and the absorbance was measured at 490 nm with an EL312e Microplate Bio-Kinetics reader (Bio-Tek Instruments, Vermont, USA).

3.2.4 Growing and amplification of single colonies

Single phage colonies from the selection round that contained the highest concentration of thrombin-binding phages, were selected and amplified.

From the “output” phage colonies grown after the fourth biopanning selection round on 2 \times YT/ampicillin/glucose plates, 48 single colonies were selected. The single colonies were picked and inoculated into 2 ml of a 1:50 diluted pre-culture (grown overnight in liquid 2 \times YT/ampicillin media), and incubated at 37°C with shaking overnight. The following day, the overnight cultures were split into two, and one half was centrifuged for 5 min at 13000 g. The supernatant containing the amplified phages were used in an ELISA to determine which individual colonies bind to α -thrombin (binding ELISA). Glycerol cell stock was prepared from the remaining half of the overnight cultures, by adding 100 μ l 75 % glycerol to 300 μ l culture, mixing well and storing the stock at $-70^\circ C$ for later use.

3.2.5 Binding ELISA of single colonies

To distinguish between actual α -thrombin binders and non-specific plastic binders, each colony was assayed in a α -thrombin-coated and a non-coated well. This binding assay was performed in duplicate. Again 48 wells of a 96-well Nunc-Immuno™ plate was coated overnight with 100 μ g/ml α -thrombin, blocked and washed as previously described. 100 μ l of the supernatant of each round 4 single colony generated by the previous step, was added to a coated and non-coated well of the ELISA plate in a final concentration of 2 % SM. The plates were incubated at room temperature for 2 h, detection was performed by the addition of anti-M13 antibody and visualised as described previously. The absorbance at 490 nm was measured with the most intense colouration indicating either the strongest binding phages, or alternatively a high concentration of weak α -thrombin binding phages.

3.2.6 Dilution ELISA

The single colonies that bound to α -thrombin with the highest affinity were subsequently amplified from frozen cell stock, and assayed for concentration dependent binding to α -thrombin in a dilution ELISA. The highest concentration of phages added was 5×10^{10} , and phages were diluted 1:2 as in the global ELISA. Each colony was diluted on coated and non-coated wells to again distinguish between α -thrombin binding phages and non-specific plastic binding phages. The assay was performed in duplicate, following the same protocol as the previously described global ELISA.

3.2.7 Competition ELISA

To investigate the potential of the phage colonies to inhibit thrombin, a competition assay was performed to determine whether r-hirudin, a known thrombin inhibitor, was able to prevent the strongest binding single colonies from binding to α -thrombin. If the phage colonies bound in the hirudin-binding area on thrombin, they would have a good change of inhibiting thrombin, and this would be illustrated in the competition assay.

The recombinant hirudin used was a generous gift from Dr. B. Rosenkranz (HBW 023; Hoeschst AG / Behringwerke AG, Germany). The ELISA plate was coated with 100 µg/ml α -thrombin as previously described, to which r-hirudin was added at the highest concentration of 500 µg/ml and serially diluted 1:2. After incubation at room temperature for 15 min, 5×10^{10} phages in 2% SM from the selected single colonies were added, and the plates incubated for 2 h at room temperature. Detection and visualisation was done as previously described. The assay was performed in duplicate.

3.2.8 Thrombin time (TT)

The effect the of the strongest binding phage colonies on the thrombin time (TT) was determined. Normal platelet poor plasma was used. This was prepared by centrifuging 10 ml citrated blood, of each of the 20 normal volunteers, at 2000 g for 10 min, and aspirating the plasma. The volunteers must not have used aspirin for 2 weeks prior to sampling. From each colony, 100 µl rescued phages were added to platelet poor plasma. Two negative control reactions were included, in one PBS was added to plasma, and in the other non-binding phages were added. The thrombin used in the test was Dade Thrombin Clotting Time Test[®] (Dade Behring, Marburg, Germany). The TT was determined by incubating 100 µl plasma with 100 µl phages. Next 100 µl Dade thrombin was added, and the clotting time was measured on the SStart[®]4 coagulation timer (Diagnostica Stago, Asnieres, France).

3.3 Cloning of the CUB2 domain of ADAMTS-13

3.3.1 Amplification of the CUB2 domain

Plasmid vector containing full length ADAMTS-13 (pNUT-ADAMTS13) was a kind gift from Dr. J.E. Sadler (Washington School of Medicine, St. Louis, USA), and served as template for the amplification of the second CUB-domain.

Primers were designed to allow amplification of the CUB2 insert and simultaneously add appropriate sites for restriction endonucleases at either ends of the target sequence. The CUB2Fp' forward primer contained a restriction site for *EcoR1* (5'-GCGGAATTCATGTGACATGCAGCTCTTTGG-3'), while the CUB2Rp reverse primer (5'-ATGCTCGAGGGTTCCTTCCTTTCCCTTCC-3') contained a *Xho1* restriction site. Primers were custom made by Invitrogen, Paisley, UK.

Different reagent concentrations and reaction times and temperatures were tested but only the optimum conditions will be given in this section. In the 50 μ l amplification reaction, 12 ng template was added to a PCR mixture (all reagents from Invitrogen) on ice containing 0.3 μ M of each primer, 2.5 units of Platinum *Pfx* DNA polymerase, 0.3 mM of each dNTP, *Pfx* amplification buffer and PCR_x enhancer solution (1 \times final concentration) and 1 mM MgSO₄. Two control reactions were included, one in which *Pfx* DNA polymerase was omitted, and the other containing MQ water instead of template. The reactions were subjected to the following program on a Peltier Thermal Cycler (BIOzymTC, Landgraaf, The Netherlands): 94°C denaturation for 2 min, followed by 30 cycles of 94°C for 15 s, 57°C for 30 s and 68°C for 1 min. After cycling the reaction was maintained at 25°C and then placed on ice.

Gel electrophoresis was carried out in TBE electrophoresis buffer on a 1 % agarose gel containing 0.5 μ g/ml ethidium bromide, and using 5 μ l DNA smart ladder (Eurogentec, Seraing, Belgium) as a molecular standard. The product was purified with the QIAquick[®] Gel Extraction kit (QIAGEN, Hilden, Germany) according to the manufacturer's suggestions. An aliquot of the purified product was then used for quality analysis on a 1 % agarose gel and the quantity of product was estimated by DNA smart ladder comparison, according to the manufacturer's suggestions.

3.3.2 Preparation of the CUB2 insert

The purified PCR product (CUB2 insert containing endonuclease restriction sites) was digested in a 30 μ l double-digestion reaction containing 22 μ l of the PCR product, 3 μ l 10 \times NEBuffer2, 1 mg/ μ l BSA, 1 μ l *EcoR*1 and 1 μ l *Xho*1 (all enzymes and buffers from New England BioLabs, Beverley, USA). After a 3 h 30min incubation at 37°C, the endonucleases were heat-inactivated for 20 min at 65°C.

Gel electrophoretic analysis of the insert was performed as described earlier on a 1% agarose gel to assess the size and quality of the insert. The insert was then purified from the gel with the QIAquick[®] gel extraction kit following the manufacturer's instructions and an aliquot was again visualised on a 1 % agarose gel to determine the relative quantity of the insert.

3.3.3 Preparation of the pET-26b(+) expression vector

One Shot[®] TOP10 (*F*⁻ *mcrA* Δ (*mrr-hsdRMS-mcrBC*) ϕ 80*lacZ* Δ M15 Δ *lacX74* *deoR* *recA1* *araD139* Δ (*ara-leu*)7697 *galU* *galK* *rpsL* (Str^R) *endA1* *nupG*) chemically competent cells from Invitrogen, were transformed with the expression vector pET-26b(+) from Novagen (Madison, USA), to facilitate the propagation of the plasmid.

Briefly, a vial of TOP10 cells was thawed on ice to which 2 μ l vector was added, and returned to ice for a further 30 min. The cells were heat-shocked at 42°C for 30 s and placed on ice, to which 250 μ l pre-warmed SOC medium (Invitrogen) was added. The vial was then incubated horizontally at 37°C for 1 h with shaking.

To select for kanamycin-resistant transformants, 150 μ l and 100 μ l aliquots of the transformation reaction were spread on pre-warmed sterile LB agar plates containing 100 μ g/ml kanamycin sulfate (Roche Molecular Biochemicals) and incubated at 37°C overnight. The following day, single colonies were selected

and inoculated into 3 ml LB medium containing 30 µg/ml kanamycin, and then incubated overnight at 37°C with shaking. Vector plasmid was isolated with the QIAprep® Spin Miniprep kit (QIAGEN) following the manufacturer's protocol for using a QIAvac® 24 vacuum manifold (QIAGEN). The concentration of the purified vector was determined spectrophotometrically on a Gene Quant II spectrophotometer (Amersham Biosciences).

The vector was prepared for cloning by digesting 3 µg vector with 1 µl each *EcoR1* and *Xho1* restriction enzymes in a double-digestion reaction containing 3 µl 10× NEBuffer2 and 1 mg/ml BSA. After being incubated at 37°C for 3 h 30 min, the enzymes were inactivated by heating the reaction to 65°C for 20 min. The cut vector was analysed for correct size by gel electrophoresis, purified from the gel, and the relative concentration was estimated by running an aliquot on a agarose gel, all as previously described.

3.3.4 Construction of recombinant plasmid

The cut insert was ligated into the prepared vector to create the recombinant pET-CUB2 plasmid. As suggested by the pET system manual, 64 ng vector and 0.2 pmol insert were used in the ligation reactions. The following components were assembled in a microcentrifuge tube: 6 µl prepared CUB2 insert, 4 µl prepared pET-26b(+) vector, 2 µl 10× ligase buffer Ligase (Roche Molecular Biochemicals), 1 µl DNA T4 Ligase (Roche Molecular Biochemicals) and MQ water to a total reaction volume of 20 µl. A control reaction, in which the insert was omitted, was set up to check for non-recombinant background. Both ligation reactions were incubated at 16°C for 16 h.

Aliquots of the ligation mixtures were cloned into One Shot® TOP10 chemically competent cells, following a heat shock transformation protocol as previously described. Transformation mixture aliquots of 100 µl and 150 µl were plated out

on freshly prepared sterile LB agar plates containing 50 μ l/ml kanamycin, and plates were incubated overnight at 37°C.

3.3.5 Direct polymerase chain reaction (PCR)

The kanamycin-resistant transformants were screened by direct PCR on single colonies to verify the presence of the recombinant pET-CUB2 plasmid. Random single colonies were picked and added to the following PCR mixture: 5 pmol insert-specific primer (2Fp'), 5 pmol vector-specific primer (T7 terminator primer), 1.5 units of Platinum *Pfx* DNA polymerase, 0.3 mM of each dNTP, *Pfx* amplification buffer and PCR_x enhancer solution (1 \times final concentration), 1 mM MgSO₄ and MQ water to a total reaction volume of 20 μ l. The reaction was then denatured for 10 min at 94°C, followed by 25 cycles of 94°C for 15 s, 57°C for 30 s and 68°C for 1 min. After cycling the reaction was maintained at 25°C and then placed on ice. Two negative control reactions were also included, one omitting template and the other omitting *Pfx* DNA polymerase. PCR products were visualised on a 1 % agarose gel as described earlier, and evaluated for plasmid presence and size.

3.3.6 Plasmid preparation

Single colonies containing the recombinant pET-CUB2 plasmid were selected and inoculated in liquid LB medium supplemented with 30 μ l/ml kanamycin for overnight growth at 37°C with shaking. A small volume of the overnight culture was used to prepare glycerol cell stock as previously described. pET-CUB2 plasmid was purified from the remainder of the overnight culture by following the QIAprep[®] spin miniprep kit protocol for using a vacuum manifold. The DNA concentration was measured spectrophotometrically.

3.3.7 Expression host transformation

An aliquot of purified pET-CUB2 construct was used to transform BL21 Star[™] (DE3) One Shot[®] (*F⁻ ompT hsdS_B (r_B⁻m_B⁻) gal dcm rne131 (DE3)*) chemically

competent cells (Invitrogen) to facilitate high-level expression of the target insert. A heat shock transformation protocol was followed as described earlier. Transformation mixture aliquots of 150 μ l and 100 μ l were plated out on LB agar plates containing 50 μ l/ml kanamycin and incubated overnight at 37°C to select for kanamycin-resistant transformants.

To obtain pure single clones, single colonies were picked the following day, again streaked out on LB agar plates containing 50 μ l/ml kanamycin and incubated at 37°C overnight. Glycerol cell stocks of the transformants were prepared as previously described.

The transformants were again screened by direct colony PCR on randomly selected single colonies to verify the presence and size of the recombinant pET-CUB2 plasmid in the BL21 cells. Direct colony PCR using the CUB2Fp' and T7-terminator primers were carried out as described earlier. PCR products were visualised by agarose gel electrophoresis as previously described.

3.4 Expression of recombinant peptide

3.4.1 Expression of recombinant CUB2

Expression of the recombinant CUB2 peptide (rCUB2) was induced by the addition of IPTG (isopropyl- β -D-thiogalactopyranoside) to a culture of BL21 cells in the exponential growing phase. A single colony was picked from a freshly streaked plate of BL21 cells (containing recombinant pET-CUB2), and inoculated into 3 ml LB media supplemented with 30 μ l/ml kanamycin and incubated with shaking overnight at 37°C. The following day, 1 ml of the overnight culture was inoculated in 100 ml fresh LB media supplemented with 30 μ l/ml kanamycin and incubated with shaking at 37°C until the OD₆₀₀ reached 0.5–1.0. The optical density of the culture was determined on a Ultrospec 1000 UV/visible Spectrophotometer (Amersham Biosciences) as follows. The culture was shaken

well to ensure a homogeneous suspension, two 100 μ l aliquots were removed and diluted into 500 μ l and 1000 μ l of the same medium used for growth, to ensure a OD₆₀₀ reading of between 0.1 and 0.8. The same medium was also used to zero the spectrophotometer, the OD₆₀₀ was noted and the average of the two readings was determined, taking the dilution factor into account.

Prior to induction, cultures were split into two 50 ml flasks, one culture served as an uninduced control and to the other, IPTG was added to a final concentration of 1 mM for induction of the *lac*-promoter. Both continued incubation at 37°C with shaking for 3 h. After incubation the flasks were placed on ice for 5 min. To obtain a medium sample and a total cell protein (TCP) sample, the cells were harvested by centrifugation at 3000 *g* for 20 min at 4°C and then returned to ice.

Of the supernatant, 1ml was transferred to a microcentrifuge tube and concentrated by trichloroacetic acid (TCA) precipitation. To the 1 ml medium fraction, 100 μ l (1/10 volume) of 100 % TCA (w/v) was added, the sample vortexed for 1 min and then placed on ice for 15 min. After centrifugation at 13000 *g* for 10 min, the supernatant was decanted and the pellet washed twice with 100 μ l acetone. This was performed by adding the acetone, mixing, and then centrifuging for 5 min (13000 *g*). The acetone was then discarded and the final pellet dried by spinning briefly in a Savant SpeedVac[®]. The pellet was resuspended in 100 μ l PBS and 100 μ l reducing SDS sample buffer was added.

The cell pellet was resuspended in 10 ml cold PBS, and sonicated with a Probe Sonicator (Instruments Scientifiques Analis, Namur, Belgium) for a total of 3 min (6 \times 30s, with 30s rest in between) at 18–24 amplitude. Thereafter, 100 μ l of the TCP sample was transferred to a microcentrifuge tube to which 100 μ l reducing SDS sample buffer was added. Both samples were vigorously vortexed, boiled for 5 min and stored at –20°C until SDS-PAGE analysis.

The medium and TCP samples of the induced culture, and aliquots of the uninduced cultures were analysed under reducing conditions on a SDS polyacrylamide gel consisting of a 15 % running gel and a 4 % stacking gel, in SDS-PAGE electrophoresis buffer. As molecular standard, 5 μ l Low Molecular Weight Marker (Amersham Biosciences) was added to 20 μ l non-reducing sample buffer and boiled for 5 min before it was loaded onto the gel. The proteins were detected with Coomassie Brilliant Blue™ (CBB) staining solution during 15 min, followed by destaining with CBB destaining solution. The gel was then dried according to the manufacture's instructions using the DryEase® Mini-Gel Drying System (Invitrogen).

3.4.2 Localisation of expressed recombinant CUB2

The localisation of the expressed rCUB2 in either the culture medium, or a cellular compartment of the *E. coli* host cell, was determined by analysing the different cellular fractions. Fresh recombinant BL21 cultures were grown, and protein expression was induced as described earlier. To facilitate localization of the expressed rCUB2, fractions of the TCP, medium, periplasm, soluble cytoplasm and insoluble cytoplasm were prepared. After preparation, all fraction samples were boiled for 5 min and stored at -20°C until SDS-PAGE analysis.

The TCP sample was collected prior to harvesting the cells: 1 ml of well-mixed culture was removed, centrifuged for 1 min at 13000 *g*, the supernatant discarded and the pellet air-dried. The pellet was resuspended in 100 μ l PBS, 100 μ l reducing SDS sample buffer added, and the sample was passed several times through a 27 gauge needle.

The remainder of the culture was harvested by centrifugation at 10000 *g* for 10 min at 4°C . For the medium sample, 1 ml of the supernatant was removed and concentrated by TCA precipitation, as already described.

The periplasmic cell fraction was prepared by dissolving the cell pellet in 30 ml 30 mM Tris-HCl, pH 8.0, 20% sucrose, adding 60 μ l 0.5M EDTA, pH 8.0, and stirring slowly at room temperature for 10 min. The cells were collected by centrifugation for 10 min at 10000 *g* at 4°C, the supernatant discarded, the pellet resuspended in 30 ml ice-cold 5 mM MgSO₄ and slowly stirred at 4°C for 10 min. The cells were centrifuged at 4°C for 10 min at 10000 *g*, after which a 1 ml aliquot was transferred from the supernatant and concentrated through TCA precipitation as described for the medium sample.

The cell pellet from the previous step was used for the preparation of the soluble and insoluble cytoplasmic fractions. The pellet was resuspended in 10 ml ice-cold 20 mM Tris-HCl, pH 7.5, and the cells lysed by adding 100 μ l/ml lysozyme (Roche Molecular Biochemicals) and incubating for 15 min at 30°C. The lysate was sonicated for a total of 3 min, after which a 1.5 ml aliquot was centrifuged for 10 min at 13000 *g*. To isolate the soluble cytoplasmic fraction, 100 μ l of the supernatant was transferred to a microcentrifuge tube to which 100 μ l reducing SDS sample buffer was added. The pellet was kept to prepare the insoluble cytoplasmic fraction.

Next, the insoluble pellet containing the inclusion bodies was resuspended in 750 μ l 20 mM Tris-HCl, pH 7.5, centrifuged at 10000 *g* for 5 min, the supernatant decanted and the wash step repeated. The final pellet was resuspended in 1.5 ml 1% SDS with heating and vigorous vortexing, after which a 100 μ l aliquot was transferred to a microcentrifuge tube as the insoluble cytoplasmic fraction sample. To the sample 100 μ l reducing SDS sample buffer was added.

Gel electrophoretic analysis of the different fraction samples were performed in duplicate on two SDS polyacrylamide gels consisting of a 15 % running gel and a 4 % stacking gel. A 5 μ l aliquot of Precision Plus Protein™ Standards Dual Color (Bio-Rad Laboratories, Hercules, USA) was loaded on each gel as a molecular marker. As a control for the hybridisation, λ domain labelled with His-Tag was

used. The His used as control was a kind gift from A. Schoolmeester (IRC, Kortrijk, Belgium). One gel was stained with CBB staining solution and dried as described earlier. The other gel was used for Western blot analysis and transferred to a nitrocellulose membrane (Schleicher & Schuell, Dassel, Germany). After blocking the membrane with a 5 % skimmed milk powder solution in TBS, the membrane was incubated with INDIA™ HisProbe-HRP (Pierce, Rockford, USA) in a dilution of 1:5000 in TBS, 0.05 % Tween-20 (TBST) during 90 min, followed by 4 wash steps with TBST. The membrane was developed using the ECL Western blotting detection reagents (Amersham Biosciences), and ECL Hyperfilm (Amersham Biosciences) according to the manufacturer's suggestions.

3.4.3 Optimisation of expression and purification

To optimise expression, the number of bacteria inoculated into the growth medium, the length of time cells were grown before induction, the density to which cells were grown after induction, as well as different IPTG concentrations were monitored and evaluated for the best quality and quantity of rCUB2 expression. The isolation and solubilising of inclusion bodies were optimised by comparing different methods of mechanical and chemical cell lysis and using wash buffers and solubilisation buffers of different chemical compositions. Renaturation and refolding of rCUB2 were evaluated by dialysing the soluble inclusion bodies against different buffers and for different lengths of time to find the best method to gradually remove the denaturing agents and detergents. Optimisation of expression and purification was firstly done on small scale and then applied to the large scale production of rCUB2.

3.4.4 Large scale production of recombinant CUB2

After optimisation of the expression and purification of the rCUB2 peptide from inclusion bodies, the large scale production of the recombinant peptide was addressed. A stab culture, taken from frozen glycerol cell stock, was inoculated into 2 ml LB medium supplemented with 30 µl/ml kanamycin. After incubation for

5 h with shaking at 37°C, 500 µl culture was transferred to 10 ml fresh LB medium containing 30 µl/ml kanamycin and incubated overnight at 37°C with shaking.

The following day, 5 ml of the overnight culture was inoculated in 500 ml LB medium supplemented with 30 µl/ml kanamycin, and incubated with shaking at 37°C until the culture reached an OD₆₀₀ of between 0.6 and 1.0. The culture was split into 2× 250 ml just prior to IPTG being added at a final concentration of 1 mM after which incubation continued for 3 h at 37°C with shaking. Again no IPTG was added to one 250 ml culture to serve as an uninduced control.

Prior to pelleting the cells, a TCP sample was removed as described earlier. The rest of the culture was harvested by centrifugation at 6500 g for 15 min at 4°C, and all supernatant discarded. The pellet was resuspended in ice-cold 25 ml inclusion body (IB) wash buffer (prepared in TBS) containing 200 mM Tris-HCl, pH 7.5, 10 mM EDTA, and 1 % (v/v) Triton X-100. To efficiently lyse the cells, lysozyme was added to a final concentration of 100 µl/ml, followed by 15 min incubation at 30°C, and sonication on ice for a total of 3 min at an amplitude of 24.

The inclusion bodies were collected by centrifuging the lysed cells for 15 min at 6000 g. The pellet was resuspended in 25 ml IB wash buffer, centrifuged at 6000 g for 15 min and the supernatant again discarded.

3.4.5 Peptide purification and refolding from inclusion bodies

The inclusion body pellet was washed in 25 ml 20 mM Tris-HCl, pH7.5, and centrifuged for 15 min at 6000 g. The supernatant was decanted and the pellet dissolved in 100 ml IB solubilisation buffer (prepared in TBS) containing 6 M GdmHCl, 1.5 M urea and a redox system consisting of 0.6 mM reduced

glutathione and 0.3 mM oxidized glutathione. To dissolve the inclusion bodies as best as possible the solution was mixed vigorously at intervals.

The solubilised peptides were then renatured and refolded by stepwise dialysis to remove the GdmHCl and urea. Dialysis started at 4°C against 10 litre TBS containing 8 M urea. After 48 h, the concentration of urea was gradually decreased over the next 48 h (each time using fresh buffer) with final dialysis against TBS for 12 h at 4°C.

To monitor the efficiency of renaturation and refolding of rCUB2, 1.5 ml samples were taken after dialysis and centrifuged for 5 min at 13000 *g*. A 100 µl aliquot of the supernatant was removed and the rest of the supernatant was discarded. The pellet was resuspended in 100 µl TBS, and 100 µl of this was removed as pellet sample. Both samples were prepared by adding 100 µl reducing sample buffer, and boiling for the samples for 5 min before being loaded on two SDS gels. Dual color protein standard was used as molecular marker, His served as the HisProbe control. The TCP sample taken from the culture before inclusion body purification was also loaded onto the gels. Samples were analysed in duplicate by SDS-PAGE as previously described. Proteins were detected by both CBB staining and Western blot also as described earlier.

CHAPTER 4

RESULTS

4.1 CONSTRUCTION OF cDNA PHAGE DISPLAY LIBRARY

4.1.1 cDNA synthesis from *Hippobosca rufipes* RNA

mRNA was prepared from total RNA isolated from whole *Hippobosca rufipes* specimens, and converted into double stranded cDNA. The cDNA was digested with *Sau3A* endonuclease to create compatible 5' extensions of 4 bases in length to facilitate ligation into the pHage 3.2 phagemid vector, which contains a restriction site for *Bgl*II. cDNA fragments of approximately 500 bp were created which was ideal since the vector can tolerate fragments ranging between 500 bp and 3000 kb. The digestion products were visualised by agarose gel electrophoresis, which is shown in figure 4.1.

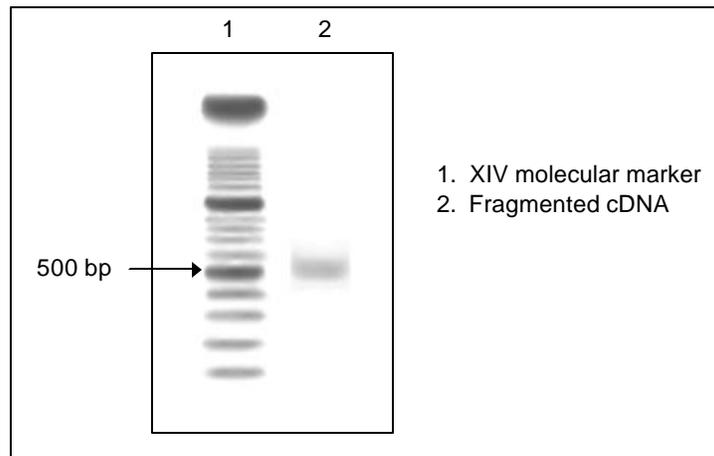


Figure 4.1. Restriction endonuclease digestion of cDNA. Digestion of cDNA with *Sau3A* generated a smear of cDNA fragments of approximately 500 bp in size.

4.1.2 Construction of cDNA phage display library

The constructed cDNA phage display library was transformed into *E. coli* TG1 cells. To determine whether TG1 transformants contained the cDNA library or empty phagemid vectors, direct colony PCR was performed on 20 randomly

picked single colonies. Primers were designed to hybridise to gene III of the M13 phage and only amplify a fragment of the gene (~ 250 bp). This ensured amplification of the pHage 3.2 insert, since the multiple cloning site of the phagemid vector was located just downstream of the gene III leader sequence on the vector. Interestingly, all the colonies tested contained a cDNA insert of approximately 500 bp suggesting that all these phages had the same cDNA fragment inserted. The cDNA insert was readily separated by gel electrophoresis from the smaller gene III fragment. This is portrayed in figure 4.2.

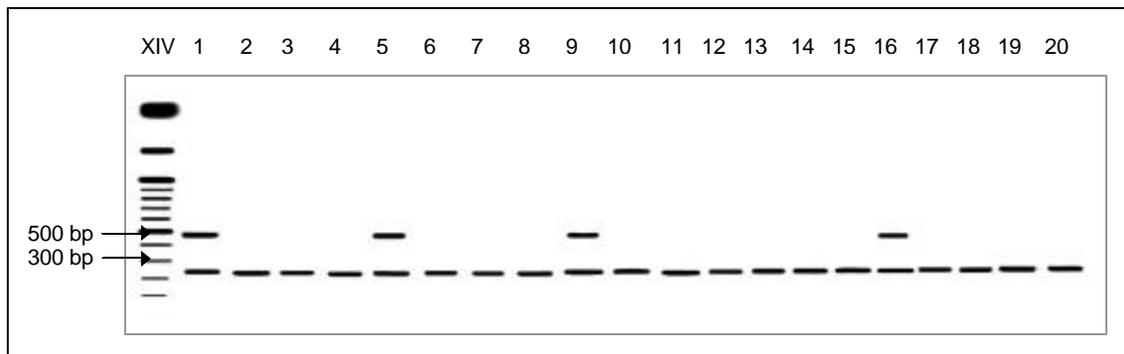


Figure 4.2. Direct PCR on randomly picked single TG1 colonies. Of the 20 colonies tested, 4 colonies contained a cDNA insert of approximately 500 bp in size.

The cDNA library was amplified and used in the biopanning experiments. This was to establish the presence of a thrombin-binding substance possibly encoded by these cDNA fragments.

4.2 SELECTION OF THROMBIN-BINDING PHAGES

4.2.1 Biopanning against thrombin

Biopanning was performed against immune-tubes coated with α -thrombin. The concentration of the amplified phages increased as from the second round of panning. The concentrations of the amplified phages per ml were 2×10^{12} for the first round, 2×10^{12} for the second round, 3×10^{12} for the third round, and 3.2×10^{12} for the fourth round.

The global ELISA indicated that the phages of round 4 contained the highest concentration of thrombin-binders, since the optical density of phages from this round was the highest. These results are illustrated in figure 4.3. Phages from round 4 were therefore amplified and used in further binding studies.

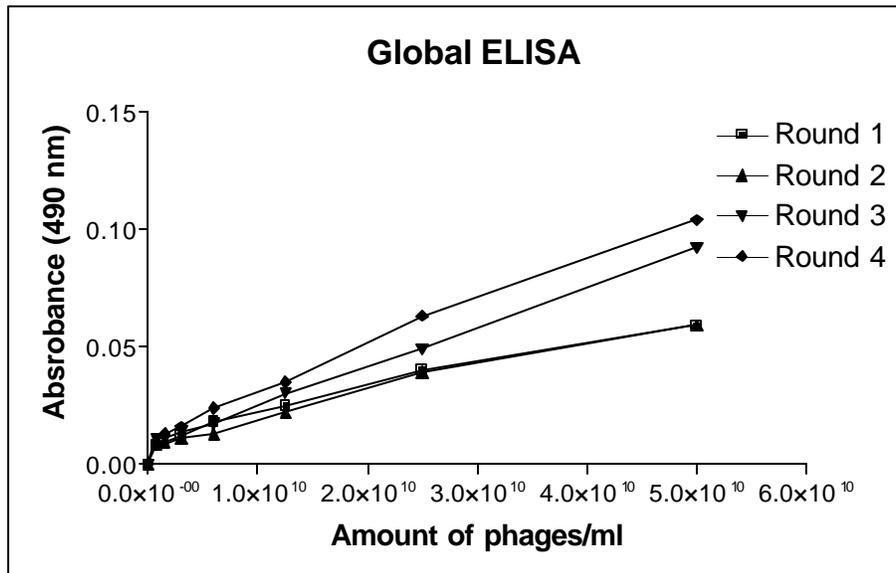


Figure 4.3. Global ELISA of four biopanning rounds against α -thrombin. The concentration of the amplified phages increased from the second panning round, indicating that phages were enriched for binding to thrombin.

With the binding ELISA performed with single colonies, 6 phage colonies exhibited strong binding to thrombin. These 6 colonies were amplified and a dilution ELISA assay was performed. Colony 46 bound to thrombin with the highest affinity, while the OD₄₉₀ increased with increasing phage concentrations. The assay was done in duplicate, but to simplify the graph, the average of the values obtained are plotted in figure 4.4.

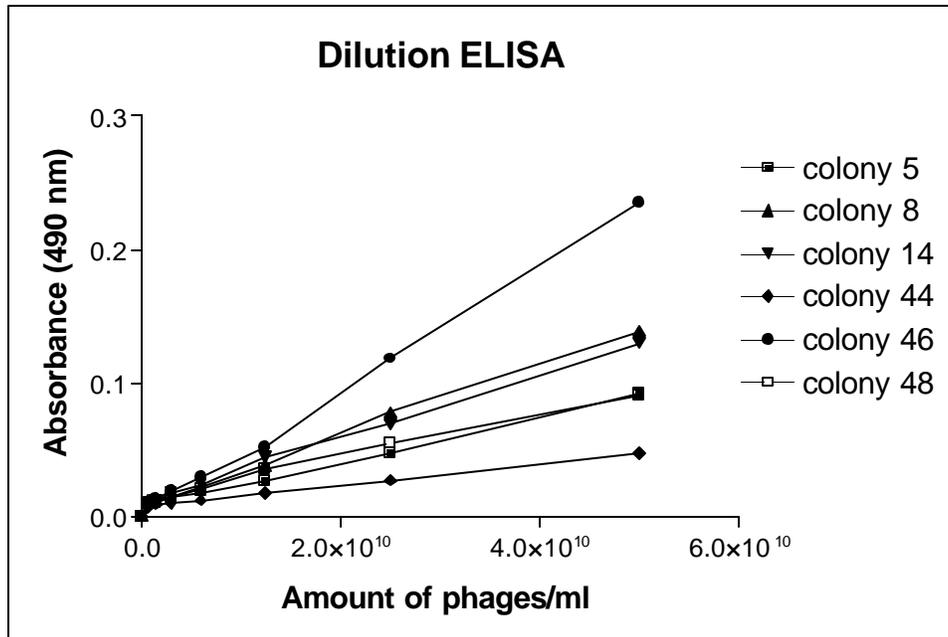


Figure 4.4. Dilution ELISA of the six strongest thrombin-binding phage colonies. The OD_{490} increased with increasing phage concentrations, indicating stronger binding to thrombin. Colony 46 bound to thrombin with the highest affinity. The OD_{490} of the uncoated control wells were subtracted from the values of the thrombin-coated wells. These values were plotted on the graph.

4.2.2 Competition ELISA

A competition ELISA was performed on the 6 selected phage colonies to determine whether the phages bound in the hirudin-binding area on thrombin, and whether hirudin was able to prevent the colonies from binding to thrombin. Only the binding of colony 46 was moderately inhibited by hirudin. Since the binding of the other colonies was not inhibited, only the results of colony 46 and 48 are plotted. Again the average values of the two experiments are shown in figure 4.5.

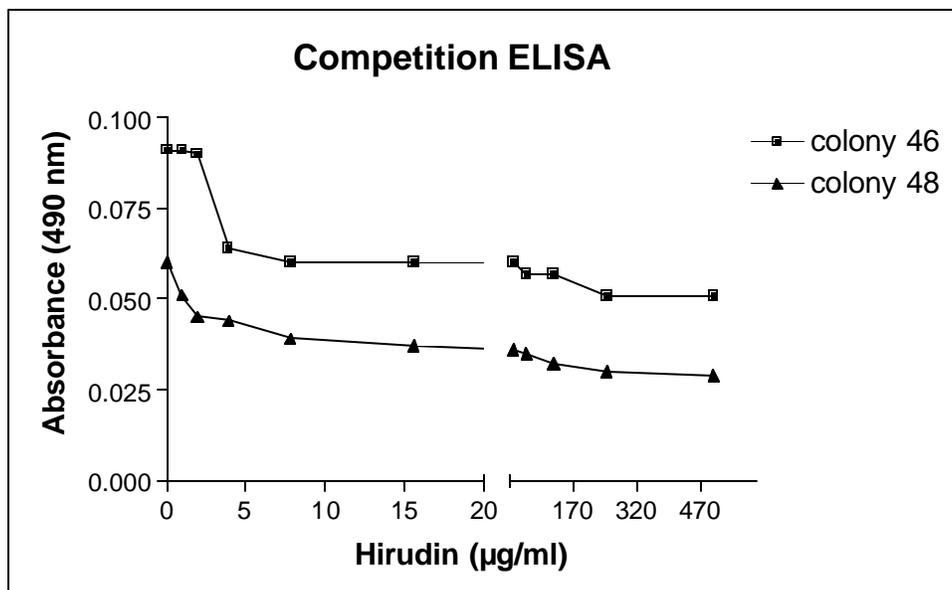


Figure 4.5. Competition ELISA of colonies 46 and 48 performed in the presence of different hirudin concentrations.

4.2.3 Thrombin time (TT)

Thrombin times (TT) were performed in the presence of different concentrations of the 4 colonies that bound strongest to thrombin. In the control reaction, PBS was added to the plasma. As another negative control, a non-binding phage colony was added to the plasma. Colonies lengthened the TT concentration dependently, with colony 46 having a greater effect than the other colonies. This is the same colony of which the binding was inhibited by hirudin and that also bound to thrombin with the highest affinity. Results are shown in figure 4.6.

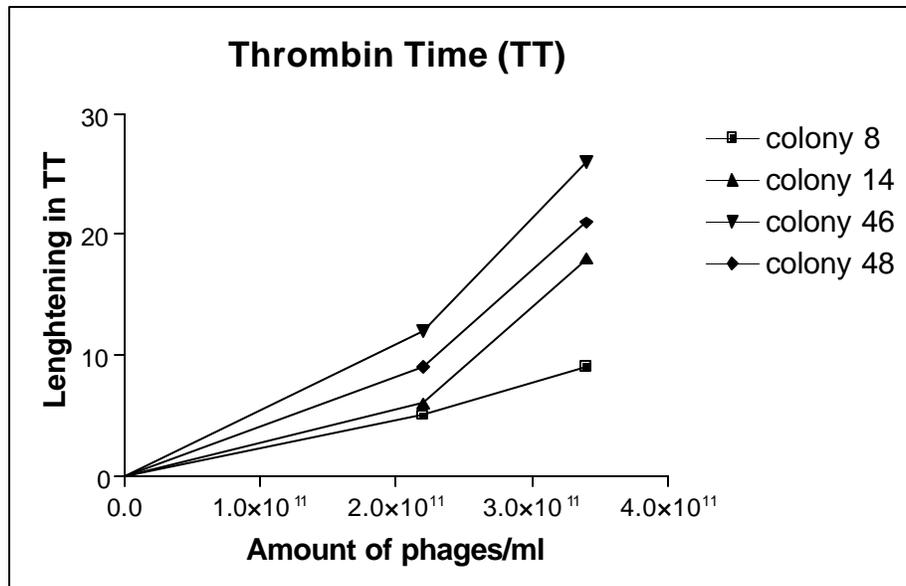


Figure 4.6. Prolongation of thrombin times. TTs were performed in the presence of increasing concentrations of the 4 strongest thrombin-binding phage colonies. Colony 46 lengthened the TT by 26 s. The lengthening in TT was calculated by subtracting the TT value of the control reaction from the TT value of the different colonies.

4.3 CLONING OF THE CUB2 DOMAIN OF ADAMTS-13

4.3.1 Amplification of CUB2

Drawing on published data, primers were designed to specifically amplify the CUB2 domain. At the same time, the primers added 5' and 3' restriction enzyme sites compatible with sites in the bacteriophage T7 promoter expression plasmid pET-26b(+) at either ends of the target sequence. This ensured the production of a PCR product consisting of the CUB2 domain with *EcoR1* and *Xho1* restriction sites to assist in further cloning steps. The vector and insert design are illustrated in figure 4.7.

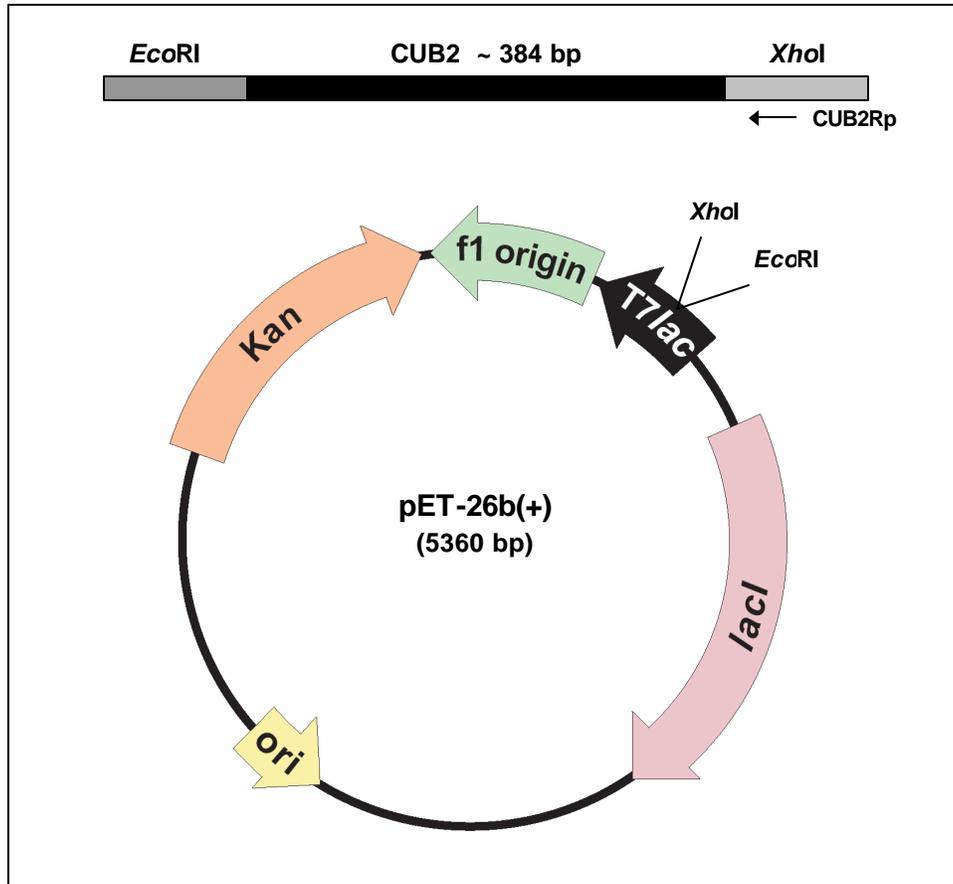


Figure 4.7. Schematic representation of insert and vector design.

Full-length recombinant plasmid DNA (pNUT-ADAMTS13) was used as a template from which the CUB2 domain was successfully amplified by PCR. In separate double-digestion reactions, 220 ng CUB2 and 3 μ g pET-26b(+) vector was digested with the restriction endonucleases *EcoR1* and *Xho1* to create the appropriate cloning sites. After digestion, the cut CUB2 insert and pET-26 vector were evaluated for size and purity, and the relative concentrations were estimated through agarose gel electrophoresis. The agarose gel is shown in figure 4.8.

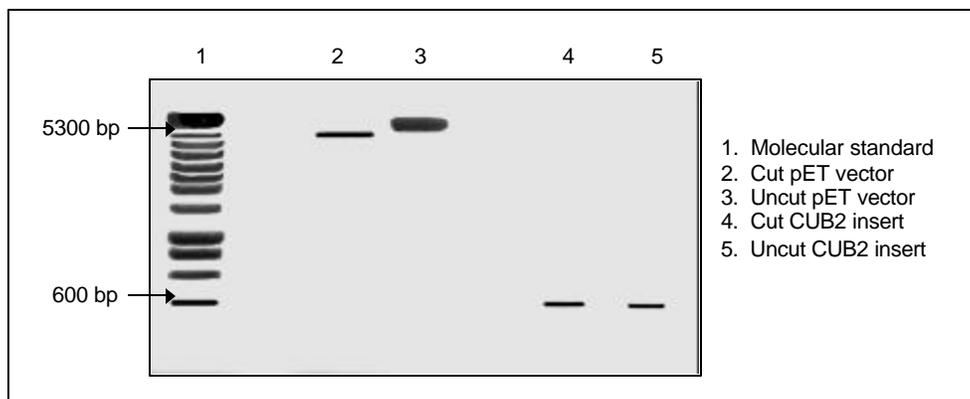


Figure 4.8. Restriction endonuclease digestion of vector and insert. Visualisation of digestion reaction on a agarose gel showing cut and uncut pET-26 vector, and cut and uncut CUB2 insert.

4.3.2 Transformation of the non-expression host

The recombinant pET-CUB2 plasmid was constructed by ligating 0.198 pmol prepared CUB2 insert to 64 ng prepared pET-26b(+) expression vector. A non-expression host (TOP10 *E. coli*) was transformed with the recombinant pET-CUB2 plasmid to facilitate the analysis of the construct. A control transformation was also done using self-ligated pET-26 plasmid. Kanamycin-resistant transformants were selected by plating aliquots of the transformation reaction on agar plates containing kanamycin.

Single colonies were observed on plates of the cells transformed with the recombinant plasmid, but no colonies were visible on the transformation control plates indicating a successful transformation and an absence of non-recombinant background. To verify the presence of the recombinant pET-CUB2 plasmid, the transformants were screened by direct colony PCR using an insert-specific primer (CUB2Fp') and a vector-specific primer (T7-terminator primer). The PCR products were evaluated on a 1% agarose gel by electrophoresis. All of the 16 randomly picked colonies tested positive for the presence and correct size (approximately 600 bp) of the pET-CUB2 plasmid. The agarose electrophoresis gel is shown in figure 4.9.

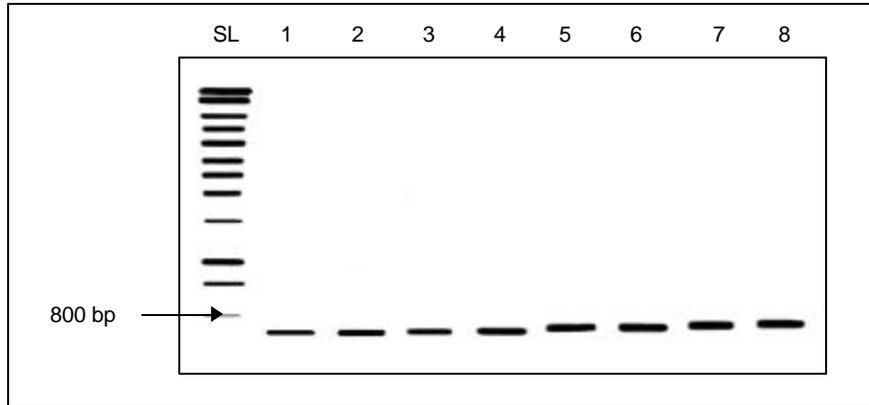


Figure 4.9. Direct PCR on transformed TOP10 colonies. All 16 single colonies tested positive for the presence, correct size (~600 bp) and orientation of pET-CUB2. SL = DNA Smart Ladder molecular standard (Eurogentec).

4.3.3 Transformation of the expression host

After positive clones had been identified and the recombinant plasmid verified, the plasmid was isolated and purified from the non-expression host. The expression host (BL21 *E. coli*) was transformed with the pET-CUB2 construct, and pure single clones were obtained. Single colonies were again screened by direct colony PCR and all of the 20 randomly picked transformants tested positive for insert presence and size. PCR products were visualised on a agarose gel, which can be seen in a figure 4.10.

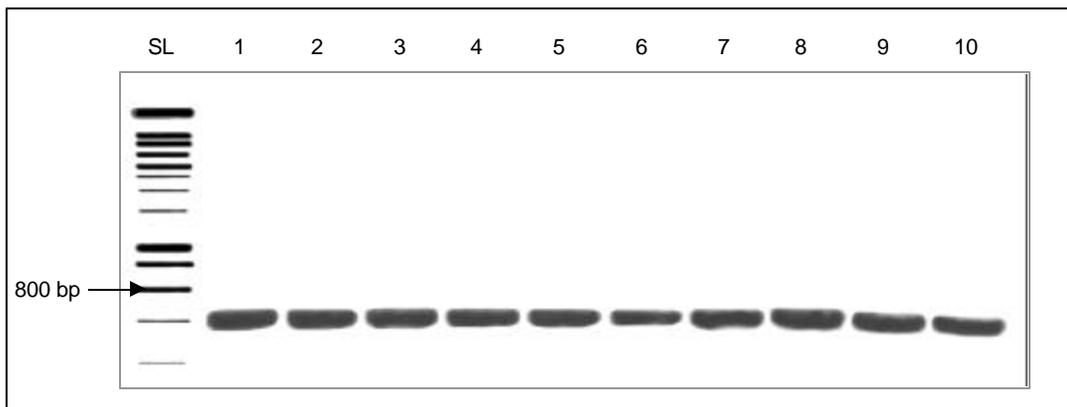


Figure 4.10. Direct PCR on randomly picked BL21 single colonies. All the colonies tested positive for the correct size (~600 bp) and orientation of the pET-CUB2 construct. SL = DNA Smart Ladder molecular standard.

4.4 EXPRESSION OF RECOMBINANT CUB2

4.4.1 Expression of the recombinant peptide

Expression of the recombinant CUB2 peptide was induced by the addition of IPTG (isopropyl- β -D-thiogalactopyranoside) to a final concentration of 1mM to a growing recombinant BL21 culture after it had reached the appropriate OD₆₀₀. Medium and total cell protein (TCP) samples were analysed for the expression of the recombinant peptide by SDS-PAGE analysis and staining with Coomassie blue (CBB). Increased protein expression was observed in the TCP sample, with possible leakage into the medium (Figure 4.11). Even though the protein-band compared with the size of rCUB2 (~16 kDa or 400 bp), the protein had not yet been definitively shown to be rCUB2 by the techniques employed thus far. More sensitive methods needed to be used and moreover, the protein seemingly expressed in the medium also had to be investigated further.

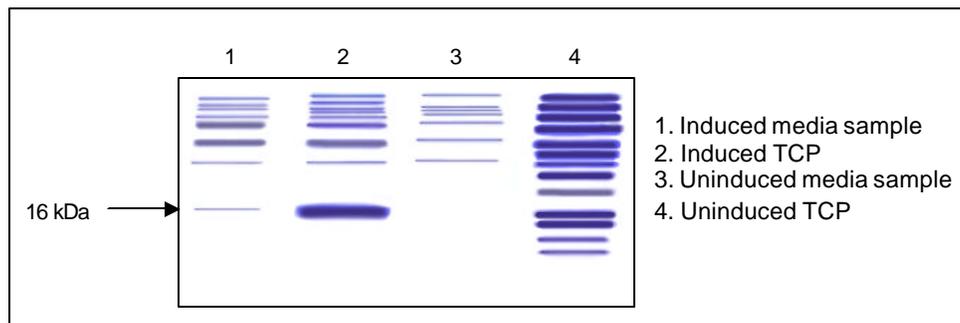


Figure 4.11. Induction of protein expression by addition of IPTG. A protein band of the correct size (~16 kDa) for rCUB2 was observed in the induced samples.

4.4.2 Localisation of expressed peptide

The following step was to determine in which cellular compartment of the host cell the peptide was being expressed, and identify the expressed protein as rCUB2. Small scale analysis of TCP and of four sequential cellular fractions namely medium, periplasm, soluble cytoplasm and insoluble cytoplasm was carried out. Each fraction was subjected to SDS-PAGE, and both CBB staining and Western blot analysis. The latter was done to verify the protein expressed

as rCUB2 by using a HRP-labelled HisProbe to hybridise with the His-Tag contained in the pET-CUB2 plasmid. The Western blot confirmed the higher protein levels in the TCP and insoluble cytoplasmic fractions, as seen with CBB staining, as expressed rCUB2. It further showed that the proteins seen in the media fraction were bacterial proteins and not rCUB2, and thus there was no leakage into the medium. Results of the SDS-PAGE analyses can be seen in figure 4.12.

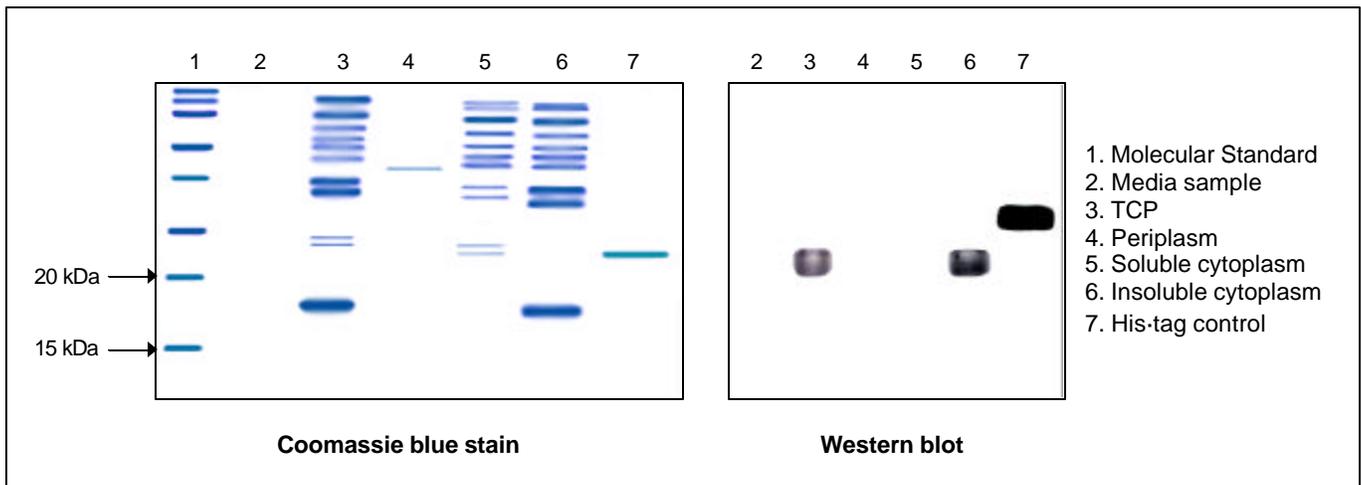


Figure 4.12. Expression of recombinant CUB2 in different cellular compartments. rCUB2 expression was detected in the TCP and insoluble cytoplasmic fractions. Western blot analysis showed that the 16 kDa fragment contains a His-Tag, and thus is rCUB2.

4.4.3 Large scale production and refolding of rCUB2

Once the expression, purification and refolding of rCUB2 was optimised, production of the peptide was done on large scale. The inclusion bodies were recovered by centrifugation and washed with inclusion body (IB) wash buffer containing Triton-X and EDTA and then solubilised in IB solubilisation buffer containing GdmHCl, urea and a glutathione redox couple. The refolding solution, however, remained cloudy with visible white aggregates even after numerous attempts to completely dissolve the inclusion bodies. Refolding was carried out by stepwise dialysis against TBS containing decreasing concentrations of urea over a period of 96h.

The efficiency of refolding and renaturing of rCUB2 was assessed by SDS-PAGE with CBB staining and western blot analysis. After 96h of dialysis rCUB2 was detected in both the supernatant and the pellet of a centrifuged sample. The bacterial proteins contaminating these fractions were successfully removed. This was established through the SDS-PAGE analyses seen in figure 4.13. The white aggregates observed previously were still visible in the solution.

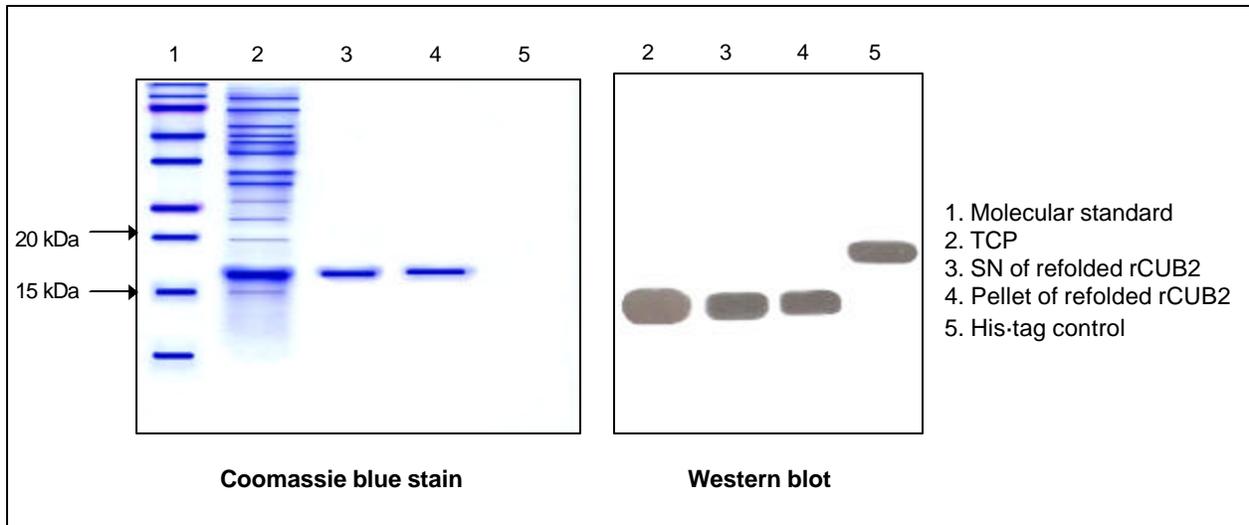


Figure 4.13. Refolding of recombinant CUB2. Inclusion bodies were purified and contaminating bacterial proteins were removed. rCUB2 was isolated from the inclusion bodies, solubilised in a IB solubilisation buffer, and refolded through stepwise dialysis.

CHAPTER 5

DISCUSSION

Advances made in understanding the mechanisms that regulate blood coagulation have brought about the identification of novel, and potentially more effective, targets for the pharmacological treatment of thrombosis. The main driving force behind this is the persistence of thrombosis as a leading cause of mortality and morbidity, despite current treatments, and the safety limitations of the antithrombotic agents currently available. This has stimulated the search for new alternatives based on selective inhibition of serine proteases (Agnelli & Sonaglia, 1999). The coagulation cascade consists of several highly homologous serine proteases, their cofactors and inhibitors (Mann, 1999). Commonly used anticoagulants, such as coumarin and heparin, lack specificity and have a rather narrow therapeutic window. This may lead to undesirable side effects such as bleeding (Hirsh, 2003).

A thrombin inhibitor was isolated from the haematophagous louse fly *Hippobosca rufipes*. This study was the first investigation into the antihaemostatic compounds of *H. rufipes*, which is a common parasitic insect found on large mammals throughout Southern Africa. Although *H. rufipes* are mainly found on horses and cattle, the insect is known to also feed on humans which suggests that it may also contain anti-human antithrombotic compounds.

An approach was followed that allowed proteins to be selected based on their binding capacities to haemostatic agents. Antithrombotic substances are found not only in the saliva of haematophagous insects, but also throughout the rest of the digestive system. For instance, the thrombin inhibitor (TTI), originally isolated from tsetse fly saliva, is also expressed in the gut of the insects following a bloodmeal (Cappello *et al.*, 1998). Therefore it was decided to isolate total RNA

from whole *H. rufipes* specimens, and not just from the salivary glands. Moreover, this provided a larger amount of starting material.

Antithrombotic agents have been isolated mainly using classical chromatography-based techniques, but the minute amounts of starting material has made the use of these techniques an extremely difficult. In an attempt to circumvent these problems, advantage was taken of a phage display method that allows the expression of cDNA libraries on phage surfaces. This phage display method has been used to identify two serine protease inhibitors from a cDNA library derived from the *Ancylostoma caninum* canine hookworm (Jespers *et al.*, 1995), and isolate a collagen-binding protein from a cDNA library of the human hookworm *Necator americanus* (Viaene *et al.*, 2001).

Phage display technology creates a physical linkage between a vast library of random cDNA fragments, and the cDNA encoding each sequence, allowing rapid identification of ligands for a variety of target molecules. More than 10^9 different sequences can be screened which gives phage display a major advantage over other selection methods. A further advantage over other methods for studying protein-protein interactions, is the fact that it allows proteins to fold correctly on the phage surface (Azzazy & Highsmith, 2002).

A cDNA library derived from *H. rufipes* was fused to the gene encoding the phage coat protein III of the filamentous M13 phage. Phages rescued from the cDNA expression library were selected for their binding ability to human α -thrombin. Thrombin was chosen since this protease plays such a central role in thrombosis and haemostasis. Other thrombin inhibitors have been isolated from the haematophagous tsetse fly (Cappello *et al.*, 1996) and the horn fly (Zhang *et al.*, 2002). In subsequent studies, the libraries might be panned against other coagulation factors such as factor Xa, fibrinogen, tissue factor, plasminogen, and collagen to select compounds directed against these coagulation agents.

Even though only 25 % of the single colonies tested contained an inserted cDNA library, it was still possible to select phages which bound to α -thrombin. Based on the number of base pairs, it seemed possible that these colonies contained the same cDNA library. Four selection rounds for thrombin-binding phages were done using a technique called biopanning. After the four rounds of enrichment (biopanning), single colonies were tested for binding to α -thrombin. The six strongest binding single colonies were identified and amplified for further testing. One of the six colonies bound strongly and in a dose-dependent manner to thrombin. The phage clones that bound strongest to thrombin were tested for binding to thrombin in competition with hirudin. The binding of one clone was inhibited by hirudin. The colonies were also evaluated for their influence on thrombin time (TT), and showed a dose-dependent lengthening. The colony which bound to thrombin with the highest affinity, further also showed the greatest lengthening of the TT, in a dose-dependent manner. However, these are still very preliminary results and assays should be more repeated with empty phages and made more specific to ensure reliable results.

Future steps in the investigation of this thrombin binding compound will include the sequencing of the cDNA library insert in these thrombin-inhibiting phages, and the cloning and expression of the inserts in an attempt to isolate and further characterise the antithrombotic agent. The isolated thrombin inhibitor could then be evaluated for development as a potential antithrombotic therapeutic.

This agent is clearly still at a very early stage in its development as an antithrombotic. Therefore it was decided not to continue with the cloning and expression of this antithrombotic agent in this particular study, but rather to express a fragment from a known protein to illustrate the expression of proteins and peptides in a prokaryotic expression system. Even though this was also the beginning of a much larger project, these experiments were ideal to illustrate the expression of a recombinant protein. Heterologous protein expression as part of thrombosis research, was demonstrated by creating a recombinant domain from

the recently identified metalloprotease ADAMTS-13. This enzyme differs from the rest of its family members in that it contains two C-terminal CUB domains of still unknown function.

Investigation into the C-terminal CUB domains of ADAMTS-13 was begun by generating recombinant CUB2 (rCUB2). The recombinant domain can be used in future studies to unravel the possible function of the domain as a contributor to the enzymatic activity of ADAMTS-13, and also be employed in functional studies as target binding (e.g. vWF and platelets) either in flow cytometry or ELISA assays. Moreover, it could serve as a valuable tool in monitoring *in vitro* proteolytic activity of ADAMTS-13.

rCUB2 was expressed on a large scale using a prokaryotic *E. coli* expression system, rather than in a eukaryotic system. This was done because of the simplicity, ease, speed and cost-effectiveness involved in using an *E. coli* expression system, and because of the small size (approximately 400 bp) of rCUB2. In addition, CUB2 contains only two cysteine residues and no rare codons that may present a problem in translation by *E. coli*.

CUB2 was amplified from ADAMTS-13 cDNA and cloned into a pET-26b(+) expression vector, in a position downstream of a T7 *lac* operon. The pET-CUB2 construct was then used to transform *E. coli* TOP10 cells. The rCUB2 peptide could, however, not be expressed in TOP10 cells, since these cells lack T7 polymerase. Recombinant plasmids were selected, purified and used to transform BL21 *E. coli* where rCUB2 expression could be induced through the addition of IPTG. Expressed recombinant peptides had an estimated molecular weight of 16 kDa, which is in agreement with the molecular weight of the CUB2 domain of ADAMTS-13 purified from human plasma (Zheng *et al.*, 2001)

Optimisation of expression, isolation and purification of the recombinant peptide was done once it was established that the recombinant peptide was expressed in

the insoluble cytoplasmic fraction. This fraction mainly consists of cell debris and cytoplasmic granules, and inclusion bodies composed of insoluble aggregates of the expressed protein. The inclusion bodies can quite readily be purified from most soluble and membrane-bound bacterial proteins, however, the product may be contaminated at some level with other proteins and nucleic acids. Therefore several different wash and solubilisation buffers were evaluated to create the best combination of buffers used for the isolation and sufficient solubilisation of inclusion bodies. Since rCUB2 contains only two cysteine residues, it was unlikely that the disulphide bridges would form in an incorrect manner, and therefore misfolding of the peptide was improbable. After the removal of contaminants, the inclusion bodies were solubilised using concentrated chemical denaturants (urea and guanidinium chloride), where after renaturation and refolding was accomplished by the removal of excess denaturants by buffer-exchange stepwise dialysis. Refolded rCUB2 was detected by Western blot analysis with a HisProbe HRP conjugate, since cloning into the expression vector used allowed the attachment of a His-tag to the peptide. In this study the renaturation and refolding of rCUB2 was not presented graphically (as a function of dialysis time or denaturant concentration) but this may be done in future work.

Since no anti-CUB2 antibodies are available, one manner in which to further positively identify the expressed peptide as CUB2, would be to sequence the purified fragment and compare the nucleotide sequence with the published sequence of ADAMTS-13 (Zheng *et al.*, 2001), which will be the next step in this project. The rCUB2 produced will then be used to raise polyclonal antibodies against the CUB2 domain of ADAMTS-13 in a rabbit model. The polyclonal antisera will be tested against rCUB2 and rADAMTS-13, and a good indication of the proper folding of rCUB2 would be if all the antibodies are able to recognise rCUB2 and rADAMTS-13. These anti-CUB2 antibodies can then be employed in the further study of the CUB domain, and used to develop a method to detect ADAMTS-13 titers in plasma.

In conclusion, these studies demonstrated the use of molecular biology methods, at different stages of research studies, into the isolation of novel antithrombotics, and investigation of the functioning of the complex process of thrombosis (of which much is still not clearly understood). This proved that the understanding and prevention of cardiovascular and thrombotic diseases, and the generation of new therapeutic protein products to treat these conditions, benefit tremendously from the continued development, and advances made in the use of molecular biology techniques. Without the biotechnological advances of the last couple of decades, the medical world would still be searching around in the dark, forever bumping into its own shortcomings and misconceptions.

CHAPTER 6

ABSTRACT

The need to find new manners in which to combat cardiovascular disease and associated thrombotic complications, remains a high priority in industrialised countries. Even in third-world countries the implications and associated risks of these diseases are being felt more and more. The advent of the biotechnology era and employment of recombinant DNA techniques has brought about exponential advances in understanding the complex mechanisms of haemostasis, and is employed to find new ways to combat pathological thrombotic complications.

The challenge is to harness the many tools and techniques produced by the ongoing biotechnology explosion, and apply them to elucidate questions still unanswered and explore areas still unknown. In this study it was illustrated that modern molecular biology techniques can be applied in many areas of thrombosis and haemostasis research.

The display of cDNA libraries on the surfaces of filamentous bacteriophages was used in the search for novel antithrombotic compounds from a haematophagous insect *Hippobosca rufipes*. Phages displaying the cDNA libraries were panned against human α -thrombin and selected according to their binding affinity and inhibition ability. To illustrate the use of a *Escherichia coli* expression system, a domain of an enzyme was cloned, expressed, and the recombinant peptide isolated and refolded. ADAMTS-13 was recently identified as an important role player in the realm of von Willebrand factor activity, including primary haemostasis and pathological disorders. The second carboxy-terminal CUB domain of ADAMTS-13 was amplified from full-length cDNA, cloned into an expression vector system, and expressed as insoluble inclusion bodies in the cytoplasm of *E. coli*, from where it was isolated and refolded.

In this study, molecular techniques were used in different phases of research into the specific activity and interactions of a particular component of the haemostatic system. This illustrated the marriage of biotechnology with fundamental medical research in an era of interdisciplinary sciences.

CHAPTER 7

ABSTRAK

Die voortdurende soeke na nuwe maniere om kardiovaskulêre siektes die hoof te bied geniet hoë prioriteit in eerste-wêreld lande, terwyl die impak van hierdie siekte toestande meer en meer in die derde-wêreld gevoel word. Die aanvang en onlangse ontploffing van die biotegnologiese era het die toepassing van rekombinante DNA beginsels en tegnieke vandag alledaags gemaak. Mediese biotegnologie het reuse vooruitgang gebring in die diepte waarin die ingewikkelde wisselwerking van die komponente betrokke in hemostase verstaan word, en word uiteindelik ingespan om ongewenste trombose te beveg.

Die uitdaging is daarin geleë om al die nuwe tegnologie en tegnieke wat uit die biotegnologiese revolusie voortspring, in te span om onbeantwoorde vrae en raaisels op te los. Hierdie studie illustreer hoe moderne tegnieke in molekulêre biologie aangewend kan word in verskillende gebiede van hemostase en trombose navorsing.

Die blootlegging van komplementêre DNA (cDNA)-biblioteke op die oppervlakte van filamentieuse bakteriofage was gebruik in die soeke om nuwe, unieke antistolmiddels uit die bloedsuiende perde-luisvlieg *Hippobosca rufipes* te isoleer. cDNA-blootleggende fage is getoets teen menslike α -trombin en verder geselekteer op grond van hul affiniteits bindings- en inhibisie vermoëns. Ten einde die werking van die *Escherichia coli* uitdrukkingssitem te illustreer, is 'n domein van 'n ensiem gekloneer en uitgedruk, en die rekombinante peptied is geïsoleer en hervou. ADAMTS-13 is onlangs aangewys as 'n belangrike rolspeler in die bepaling van von Willebrand-faktor aktiwiteit in beide primêre hemostase en siekte toestande. Die tweede CUB domein aan die karboksie-einde van ADAMTS-13 is vanaf volledige cDNA geamplifiseer, gekloneer in 'n uitdrukkingss-

vektorsisteem, en uitgedruk in onoplosbare liggaampies in die *E. coli* sitoplasma, waarvandaan dit geïsoleer en hervou is.

In hierdie studie is molekulêre tegnieke ingespan in verskillende navorsingsfases, in die poging om die spesifieke aktiwiteite en wisselwerkinge waarby 'n bepaalde komponent van die hemostase stelsel betrokke is, beter te verstaan. Dit dien as voorbeeld van effektiewe samewerking tussen biotegnologie en basiese mediese navorsing in hierdie opwindende tye van interdisiplinêre wetenskappe.

References

- Adams, M.D., Kelley, J.M., Gocayne, J.D., Dubnick, M., Polymeropoulos, M.H., Xiao, H., *et al.* (1991) Complementary DNA sequencing: expressed sequence tags and Human Genome Project. *Science* **252**, 1651-1656.
- Adari, H., Andrews, B., Ford, P.J., Hanning, G., Brosius, J. and Makrides, S.C. (1995) Expression of the human T-cell receptor V β 5.3 in *Escherichia coli* by thermal induction of the *trc* promoter: nucleotide sequence of the *lacI*ts gene. *DNA Cell Biol.* **14**, 945-950.
- Adhya, S. and Gottesman, M. (1982) Promoter occlusion: transcription through a promoter may inhibit its activity. *Cell* **29**, 939-944.
- Agnelli, G. and Sonaglia, F. (1999) Clinical status of direct thrombin inhibitors. *Crit. Rev. Oncol. Hematol.* **31**, 97-117.
- Albanèse, V. and Frydman, J. (2002) Where chaperones and nascent polypeptides meet. *Nat. Struct. Biol.* **9**, 716-718.
- Allen, J.B., Walberg, M.W., Edwards, M.C. and Elledge, S.J. (1995) Finding prospective partners in the library – the two-hybrid system and phage display find a match. *Trends Biochem. Sci.* **20**, 511-516.
- Amann, E. and Brosius, J. (1985) “ATG vectors” for regulated high-level expression of cloned genes in *Escherichia coli*. *Gene* **40**, 183-190.
- Andersen, D.C. and Krummen, L. (2002) Recombinant protein expression for therapeutic applications. *Curr. Opin. Biotechnol.* **13**, 117-123.
- Andersson, L., Yang, S., Neubauer, P. and Enfors, S-O. (1996) Impact of plasmid presence and induction on cellular responses in fed-batch cultures of *Escherichia coli*. *J. Biotechnol.* **46**, 255-263.
- Andrews, B. Adari, H., Hanning, G., Lahue, E., Gosselin, M., Martin, S., *et al.* (1996) A tightly regulated high level expression vector that utilizes a thermosensitive *lac* repressor: production of the human T cell receptors V β 5.3 in *Escherichia coli*. *Gene* **182**, 101-109.

Arap, W., Pasqualini, R. and Ruoslahti, E. (1998) Cancer treatment by targeted drug delivery to tumor vasculature in a mouse model. *Science* **279**, 377–380.

Armstrong, J., Perham, R.N. and Walker, J.E. (1981) Domain structure of the bacteriophage fd adsorption protein. *FEBS Lett.* **135**, 167-172.

Arocha-Pinango, C.L., Marchi, R., Carvajal, Z. and Guerrero, B. (1999) Invertebrate compounds acting on the hemostatic mechanism. *Blood Coagul. Fibrinolysis* **10**, 43-68.

Arza, B. and Féléz, J. (1998) The emerging impact of phage display technology in thrombosis and haemostasis. *Thromb. Haemost.* **80**, 354-365.

Azzazy, H.M.E. and Highsmith, W.E. (2002) Phage display technology: clinical applications and recent innovations. *Clin. Biochem.* **35**, 425-445.

Baker, R.T. (1996) Protein expression using ubiquitin fusion and cleavage. *Curr. Opin. Biotechnol.* **7**, 541-546.

Balass, M., Katchalski-Katzir, E. and Fuchs, S. (1997) The alpha-bungarotoxin binding site on the nicotinic acetylcholine receptor: analysis using a phage-epitope library. *Proc. Natl. Acad. Sci. USA* **94**, 6054-6058.

Balbas, P. (2001) Understanding the art of producing protein and nonprotein molecules in *Escherichia coli*. *Mol. Biotechnol.* **19**, 251-267.

Baneyx, F. (1999) Recombinant protein expression in *Escherichia coli*. *Curr. Opin. Biotechnol.* **10**, 411-421.

Banno, F., Kaminaka, K., Soejima, K., Kokame, K. and Miyata, T. (2003) Identification of strain-specific variants of mouse ADAMTS 13 gene encoding von Willebrand factor-cleaving protease. *J. Thromb. Haemost.* E-pub. Manuscript M314184200.

Barbas III, C.F., Bain, J.D., Hoekstra, D.M. and Lerner, R. (1992) Semisynthetic combinatorial antibody libraries: a chemical solution to the diversity problem. *Proc. Natl. Acad. Sci. USA* **89**, 4457-4461.

Basanova, A.V., Baskova, I.P. and Zavalova, L.L. (2002) Vascular-platelet and plasma hemostasis regulators from bloodsucking animals. *Biochemistry (Moscow)* **67**, 143-150.

Bashiardes, S. and Lovett, M. (2000) cDNA detection and analysis. *Curr. Opin. Chem. Biol.* **5**, 15-20.

Bass, S., Gu, Q. and Christen, A. (1996) Multicopy suppressors of prc mutant *Escherichia coli* include two HtrA (DegP) protease homologs (HhoAB), DskA, and a truncated R1pA. *J. Bacteriol.* **178**, 1154-1161.

Bass, S.H., Greene, R. and Wells, J.A. (1990) Hormone phage: an enrichment method for variant proteins with altered binding properties. *Proteins* **8**, 309-314.

Baugh, R.J., Dickinson, C.D., Ruf, W. and Krishnaswamy, S. (2000) Exosite interactions determine the affinity of factor X for the extrinsic Xase complex. *Biochemistry* **275**, 28826-28833.

Bendas, G., Krause, A., Schmidt, R., Vogel, J. and Rothe, U. (1998) Selectins as new targets for immunoliposome-mediated drug delivery. *Pharma. Acta. Helv.* **73**, 19-26.

Benhar, I. (2001) Biotechnological applications of phage and cell display. *Biotech. Adv.* **19**, 1-33.

Berdichevsky, Y., Lamed, R., Frenkel, D., Gphna, U., Bayer, E.A., Yaron, S., *et al.* (1999) Matrix-assisted folding of single-chain Fv cellulose binding domain fusion proteins. *Protein Express. Purif.* **17**, 249-259.

Bergès, H., Joseph-Liauzun, E. and Fayet, O. (1996) Combined effects of the signal sequence and the major chaperone proteins on the export of human cytokines in *Escherichia coli*. *Appl. Environ. Microbiol.* **62**, 55-60.

Bernard, H.U. and Helinski, D.R. (1979) Use of the lambda phage promoter p_{λ} to promote gene expression in hybrid plasmid cloning vehicles. *Methods Enzymol.* **68**, 482-492.

Bernardo, A., Nolasco, L., Ball, C., Wang, Y., Moake, J., López, J., *et al.* (2003) Peptides from the C-terminal regions of ADAMTS-13 specifically block cleavage of ultra-large von Willebrand factor multimers on the endothelial surface under flow. *J. Thromb. Haemost.* Supplement 1 July, OC405

Bessett, P.-H., Åslund, F., Beckwith, J. and Georgiou, G. (1999) Efficient folding of proteins with multiple disulphide bonds in *Escherichia coli* cytoplasm. *Proc. Natl. Acad. Sci. USA* **96**, 13703-13708.

Bianchi, E., Folgori, A., Wallace, A., Nicotra, M., Acali, S., Phalipon, A., *et al.* (1995) A conformationally homogeneous combinatorial peptide library. *J. Mol. Biol.* **247**, 154–160.

Blank, M., Shoenfeld, Y., Cabilly, S., Heldman, Y., Fridkin, M. and Katchalski-Katzir, E. (1999) Prevention of experimental antiphospholipid syndrome and endothelial cell activation by synthetic peptides. *Proc. Natl. Acad. Sci. USA* **96**, 5164-5168.

Blight, M.A., Chervaux, C. and Holland, I.B. (1994) Protein secretion pathway in *Escherichia coli*. *Curr. Opin. Biotechnol.* **5**, 468-474.

Bode, W., Mayr, I., Baumann, U., Huber, R., Stone, S.R. and Hofsteenge, J. (1989) The refined 1.9 Å crystal structure of human α -thrombin: interaction with D-Phe-Pro-Arg chloromethylketone and significance of the Tyr-Pro-Pro-Trp insertion segment. *EMBO J.* **8**, 3467-3475.

Boel, E., Hjorth, A.L., Moller, K.B. and Moller, P.H. (1991) A short synthetic adaptor as second-strand primer in the construction of cDNA libraries by the vector-primer method. *Biotechniques* **11**, 26-28.

Bork, P. and Beckmann, G. (1993) The CUB domain: a widespread module in developmentally regulated proteins. *J. Mol. Biol.* **231**, 539-545.

Bothmann, H. and Plückthun, A. (1998) Selection for a periplasmic factor improving phage display and functional periplasmic expression. *Nat. Biotechnol.* **16**, 376-380.

Bowditch, R.D., Tani, P., Fong, K.C. and McMillan, R. (1996) Characterization of autoantigenic epitopes on platelet glycoprotein IIb/IIIa using random peptide libraries. *Blood* **88**, 4579-4584.

Brady, G., Barbara, M. and Iscove, N. (1990) Representative *in vitro* cDNA amplification from individual hematopoietic cells and colonies. *Methods Mol. Cell. Biol.* **2**, 17-25.

Brinkmann, U., Chowdhury, P.S., Roscoe, D.M. and Pastan, I. (1995) Phage display of disulfide-stabilized Fv fragments. *J. Immunol. Methods* **182**, 41-50.

Brooks, E.M., Sheflin, L.G. and Spaulding, S.W. (1995) Secondary structure in the 3'-UTR of EGF and the choice of reverse transcriptase affect the detection of message diversity by RT-PCR. *Biotechniques* **19**, 806-815.

Brown, K.C. (2000) New approaches for cell-specific targeting: identification of cell-selective peptides from combinatorial libraries. *Curr. Opin. Chem. Biol.* **4**, 16-21.

Burgess, R.R. (1996) Purification of overproduced *Escherichia coli* RNA polymerase σ factors by solubilizing inclusion bodies and refolding from sarkosyl. *Methods Enzymol.* **273**, 145-149.

Butt, T.R., Jonnalagadda, S., Monia, B.P., Sternberg, E.J., Marsh, J.A., Stadel, J.M., *et al.* (1989) Ubiquitin fusion augments the yield of cloned gene products in *Escherichia coli*. *Proc. Natl. Acad. Sci. USA* **86**, 2540-2544.

Cabilly, S. (1989) Growth at sub-optimal temperatures allows the productions of functional, antigen-binding Fab fragments in *Escherichia coli*. *Gene* **85**, 553-557.

Cadwell, R.C. and Joyce, G.F. (1994) Mutagenic PCR. *PCR Methods Appl.* **3**, S136-S140.

Calos, M.P. (1978) DNA sequence for a low-level promoter of the *lac* repressor gene and an "up" promoter mutation. *Nature* **274**, 762-765.

Cappello, M., Bergum, P.W., Vlasuk, G.P., Furnidge, B.A., Pritchard, D.I. and Aksoy, S. (1996) Isolation and characterization of the tsetse thrombin inhibitor: a potent antithrombotic peptide from the saliva of *Glossina morsitans morsitans*. *Am. J. Trop. Med. Hyg.* **54**, 475-480.

Cappello, M., Li, S., Chen, X., Li, G.B., Harrison, L., *et al.* (1998) Tsetse thrombin inhibitor: bloodmeal-induced expression of an anticoagulant in salivary glands and gut tissue of *Glossina morsitans morsitans*. *Proc. Natl. Acad. Sci. USA* **95**, 14290-14295.

Cardamone, M., Puri, N.K. and Brandon, M.R. (1995) Comparing the refolding and reoxidation of recombinant porcine growth hormone from a urea denatured state and from *Escherichia coli* inclusion bodies. *Biochemistry* **34**, 5773-5794.

Carninci, P., Kvam, C., Kitamura, A., Ohsumi, T., Okazaki, Y., Itoh, M., *et al.* (1996) High-efficiency full-length cDNA cloning by biotinylated CAP trapper. *Genomics* **37**, 327-336.

Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., *et al.* (2000) Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. *Genome Res.* **10**, 1617-1630.

Carninci, P., Westover, A., Nishiyama, Y., Ohsumi, T., Itoh, M., Nagaoka, S., *et al.* (1997) High efficiency selection of full-length cDNA by improved biotinylated cap trapper. *DNA Res.* **4**, 61-66.

Caro, L.G. and Schnös, M. (1972) The attachment of the male specific bacteriophage f1 to sensitive strains of *E. coli*. *Proc. Natl. Acad. Sci. USA* **56**, 126-132.

Carpousis, A.J., Van Houwe, G., Ehretsmann, C. and Krisch, H.M. (1994) Copurification of *E. coli* RNase E and PNPase: evidence for a specific association between two enzymes important in RNA processing and degradation. *Cell* **76**, 889-900.

Carrier, T.A., Jones, K.L. and Keasling, J.D. (1998) mRNA stability and plasmid copy number effects on gene expression from an inducible promoter system. *Biotechnol. Bioeng.* **59**, 666-672.

Carrier, T.A. and Keasling, J.D. (1999) Library of synthetic 5' secondary structures to manipulate mRNA stability in *Escherichia coli*. *Biotechnol. Prog.* **15**, 58-64.

Carrió, M.M. and Villaverde, A. (2002) Construction and deconstruction of bacterial inclusion bodies. *J. Biotechnol.* **96**, 3-12.

Caspers, P., Stieger, M. and Burn, P. (1994) Overproduction of bacterial chaperones improves the solubility of recombinant protein tyrosine kinases in *Escherichia coli*. *Cell Mol. Biol.* **40**, 635-644.

Chen, C-Y. and Belasco, J.G. (1990) Degradation of *pufLMX* mRNA in *Rhodobacter capsulatus* is initiated by non-random endonucleolytic cleavage. *J. Bacteriol.* **172**, 4578-4586.

Chong, S., Mersha, F.B., Comb, D.G., Scott, M.E., Landry, D., Vence, L.M., *et al.* (1997) Single-column purification of free recombinant proteins using a self-cleavable affinity tag derived from a proteins splicing element. *Gene* **192**, 271-281.

Chung, J., Rader, C., Popkov, M., Hur, Y.M., Kim, H.K., Lee, Y.J., *et al.* (2004) Integrin alpha IIb beta 3-specific synthetic human monoclonal antibodies and HCDR3 peptides that potently inhibit platelet aggregation. *FASEB J.* **18**, 361-363.

Clackson, T. and Wells, J.A. (1994) *In vitro* selection from protein and peptide libraries. *Trends Biotechnol.* **12**, 173-184.

Clackson, T., Hoogenboom, H.R., Griffiths, A.D. and Winter, G. (1991) Making antibody fragments using phage display libraries. *Nature* **352**, 624-628.

Clarke, A.R. (1996) Molecular chaperones in protein folding and translocation. *Curr. Opin. Struct. Biol.* **6**, 43-50.

Coburn, G.A. and Mackie, G.A. (1999) Degradation of mRNA in *Escherichia coli*: an old problem with some new twists. *Prog. Nucleic Acid Res. Mol. Biol.* **62**, 55-108.

Cochran, A.G. (2000) Antagonists of protein-protein interactions. *Chem. Biol.* **7**, R85-R94.

Cohen, S.N., Chang, A.C. and Hsu, L. (1972) Nonchromosomal antibiotic resistance in bacteria: genetic transformation of *Escherichia coli* by R-factor DNA. *Proc. Natl. Acad. Sci. USA* **69**, 2110-2114.

Coleclough, C. and Erlitz, F.L. (1985) Use of primer-restriction-end adapters in a novel cDNA cloning strategy. *Gene* **34**, 305-314.

Corisdeo, S. and Wang, B. (2004) Functional expression and display of an antibody Fab fragment in *Escherichia coli*: study of vector designs and culture conditions. *Protein Express. Purif.* **34**, 270-279.

Cornelis, P. (2000) Expressing genes in different *Escherichia coli* compartments. *Curr. Opin. Biotechnol.* **11**, 450-454.

Cornwall, J.W. and Patton, W.S. (1914) Some observations on the salivary secretions of the commoner blood-sucking insects and ticks. *Indian J. Med. Res.* **2**, 569-593.

- Cortese, R., Monaci, P., Luzzago, A., Santini, C., Bartoli, F., Corteses, I., *et al.* (1996) Selection of biologically active peptides by phage display of random peptide libraries. *Curr. Opin. Biotechnol.* **7**, 616-621.
- Crameri, R. and Kodzius, R. (2001) The powerful combination of phage surface display of cDNA libraries and high throughput screening. *Comb. Chem. High Throughput Screen.* **4**, 145-155.
- Crameri, R. and Walter, G. (1999) Selective enrichment and high-throughput screening of phage surface-displayed cDNA libraries from complex allergenic systems. *Comb. Chem. High Throughput Screen.* **2**, 63-72.
- Crameri, R., Jaussi, R., Menz, G. and Blaser, K. (1994) Display of expression products of cDNA libraries on phage surfaces. *Eur. J. Biochem.* **226**, 53-58.
- Crowe, J., Döbeli, H., Gentz, R., Hochuli, E., Stüber, D. and Henco, K. (1994) 6x His-Ni-NTA chromatography as a superior technique in recombinant protein expression/purification. *Methods Mol. Biol.* **31**, 371-387.
- Cwirla, S.E., Peters, E.A., Barret, R.W. and Dower, W.J. (1990) Peptides on phage: a vast library of peptides for identifying ligands. *Proc. Natl. Acad. Sci. USA* **87**, 6378-6382.
- D'Mello, F. and Howard, C.R. (2001) An improved selection procedure for the screening of phage display peptide libraries. *J. Immunol. Methods* **247**, 191-203.
- Dall'Aqua, W. and Carter, P. (1998) Antibody engineering. *Curr. Opin. Struct. Biol.* **8**, 443-450.
- Darnell, J.E., Wall, R., and Tushinski, R.J. (1971) An adenylic acid-rich sequence in messenger RNA of HeLa cells and its possible relationship to reiterated sites in DNA. *Proc. Natl. Acad. Sci. USA* **68**, 1321-1325.
- De Bernardez Clark, E. (1998) Refolding of recombinant proteins. *Curr. Opin. Biotechnol.* **9**, 157-163.
- De Bernardez Clark, E. (2001) Protein refolding for industrial processes. *Curr. Opin. Biotechnol.* **12**, 202-207.

De Boer, H.A., Comstock, L.J. and Vasser, M. (1983) The *tac* promoter: a functional hybrid derived from the *trp* and *lac* promoters. *Proc. Natl. Acad. Sci. USA* **80**, 21-25.

DeLisa, M.P., Valdes, J.J. and Bentley, W.E. (2001) Quorum signaling via A1-2 communicates the "metabolic burden" associated with heterologous protein production in *Escherichia coli*. *Biotechnol. Bioeng.* **75**, 439-450.

Dennis, M.S., Eigenbrot, C., Skelton, N.J., Ultsch, M.H., Santell, L., Dwyer, M.A., *et al.* (2000) Peptide exosite inhibitors of factor VIIa as anticoagulants. *Nature* **404**, 465-470.

Dente, L., Cesareni, G., Micheli, G., Felici, F., Folgori, A., Luzzago, A., *et al.* (1994) Monoclonal antibodies that recognize filamentous phage: tools for phage display technology. *Gene* **148**, 7-13.

Dente, L., Vetriani, C., Zucconi, A., Pelicci, G., Lanfrancone, L., Pelicci, P.G., *et al.* (1997) Modified phage peptide libraries as a tool to study specificity of phosphorylation and recognition of tyrosine containing peptides. *J. Mol. Biol.* **269**, 694-703.

Depraetere, H., Viaene, A., Deroo, S., Vauterin, S. and Deckmyn, H. (1998) Identification of peptides, selected by phage display technology, that inhibit von Willebrand factor binding to collagen. *Blood* **92**, 4207-4211.

Devlin, J.J., Panganiban, L.C. and Devlin, P.E. (1990) Random peptide libraries: a source of specific protein binding molecules. *Science* **249**, 404-406.

DiGuan, C., Li, P., Riggs, P.D. and Inoye, H. (1988) Vectors that facilitate the expression and purification of foreign peptides in *Escherichia coli* by fusion to maltose-binding protein. *Gene* **67**, 21-30.

Dodt, J., Otte, M., Strube, K-H. and Friedrich, T. (1996) Thrombin inhibitors of bloodsucking animals. *Sem. Thromb. Hemost.* **22**, 203-208.

Dong, H., Nilsson, L. and Kurland, C.G. (1995) Gratuitous overexpression of genes in *Escherichia coli* leads to growth inhibition and ribosome destruction. *J. Bacteriol.* **177**, 1497-1504.

Doorbar, J. and Winter, G. (1994) Isolation of a peptide antagonist to the thrombin receptor using phage display. *J. Mol. Biol.* **244**, 361-369.

Dower, W.J. and Mattheakis, L.C. (2002) *In vitro* selection as a powerful tool for the applied evolution of proteins and peptides. *Curr. Opin. Chem. Biol.* **6**, 390-398.

Dower, W.J., Miller, J.F., and Ragsdale, C.W. (1988) High-efficiency transformation of *E. coli* by high-voltage electroporation. *Nucleic Acids Res.* **16**, 6127-6145.

Draper, M.P., August, P.R., Connolly, T., Packard, B. and Call, K.M. (2002) Efficient cloning of full-length cDNAs based on cDNA size fractionation. *Genomics* **79**, 603-607.

Dunn, I.S. (1996) Phage display of proteins. *Curr. Opin. Biotechnol.* **7**, 547-553.

Ederly, I., Chu, L.L., Sonenberg, N. and Pelletier, J. (1995) An efficient strategy to isolate full-length cDNAs based on an mRNA cap retention procedure (CAPture). *Mol. Cell. Biol.* **15**, 3363-3371.

Elvin, C.M., Thompson, P.R., Argall, M.E., Hendry, P., Stamford, N.P.J., Lilley, E., *et al.* (1990) Modified bacteriophage lambda promoter vectors for overproduction of proteins in *Escherichia coli*. *Gene* **87**, 123-126.

Etchegaray, J-P. and Inouye, M. (1999) Translational enhancements by an element downstream of the initiation codon in *Escherichia coli*. *J. Biol. Chem.* **274**, 10079-1085.

Fenton, J.W., Ofosu, F.A., Brezniak, D.V. and Hassouna, H.I. (1998) Thrombin and antithrombotics. *Semin. Thromb. Hemost.* **24**, 87-92.

Forrer, P., Jung, S. and Plückthun, A. (1999) Beyond binding: using phage display to select for structure, folding and enzymatic activity in proteins. *Curr. Opin. Struct. Biol.* **9**, 514-520.

Friedrich, T., Kröger, B., Bialojan, S., Lemaire, H.G., Höffken, H.W., *et al.* (1993) A Kazal-type inhibitor with thrombin specificity from *Rhodnius prolixus*. *J. Biol. Chem.* **268**, 16216-16222.

Frohman, M.A., Dush, M.K. and Martin, G.R. (1988) Rapid production of full-length cDNAs from rare transcripts: amplification using a single gene-specific oligonucleotide primer. *Proc. Natl. Acad. Sci. USA* **85**, 8998-9002.

- Fu, Z., Hamid, S.B.A., Razak, C.N.A., Basri, M., Salleh, A.B. and Rahman, R.N.Z.A. (2003) Secretory expression in *Escherichia coli* and single-step purification of a heat-stable alkaline protease. *Protein Express. Purif.* **28**, 63-68.
- Fuh, G. and Sidhu, S.S. (2000) Efficient phage display of polypeptides fused to the carboxy-terminus of the M13 gene-3 minor coat protein. *FEBS Lett.* **480**, 231-234.
- Fujikawa, K., Suzuki, H., McMullen, B. and Chung, D. (2001) Purification of human von Willebrand factor-cleaving protease and its identification as a new member of the metalloproteinase family. *Blood* **98**, 1662-1666.
- Furlan, M., Robles, R. and Lämmle, B. (1996) Partial purification and characterization of a protease from human plasma cleaving von Willebrand factor to fragments produced by *in vivo* proteolysis. *Blood* **87**, 4223-4234.
- Furuichi, Y. and Miura, K. (1975) A blocked structure at the 5' terminus of mRNA from cytoplasmic polyhedrosis virus. *Nature* **253**, 374-375.
- Gao, C., Mao, S., Lo, C.H., Wirsching, P., Lerner, R.A. and Janda, K.D. (1999) Making artificial antibodies: a format for phage display of combinatorial heterodimeric arrays. *Proc. Natl. Acad. Sci. USA* **96**, 6025-6030.
- Gaskin, D.J.H., Starck, K., Turner, N.A. and Vulfson, E.N. (2001) Phage display combinatorial libraries of short peptides: ligand selection for protein purification. *Enzyme Microbiol. Technol.* **28**, 766-772.
- Georgiou, G. and Valax, P. (1996) Expression of correctly folded proteins in *Escherichia coli*. *Curr. Opin. Biotechnol.* **7**, 190-197.
- Georgiou, G. and Valax, P. (1999) Isolating inclusion bodies from bacteria. *Methods Enzymol.* **309**, 48-58.
- Gerard, G.F. (1998) Reverse transcriptase: a historical perspective. *Focus* **20**, 65-67.

- Gerard, G.F. and D'Alessio, J.M. (1993) Reverse transcriptase (EC2.7.7.49): the use of cloned Moloney murine leukaemia virus reverse transcriptase to synthesize DNA from RNA. *Methods Mol. Biol.* **16**, 73-94.
- Gerritsen, H.E., Robles, R., Lämmle, B. and Furlan, M. (2001) Partial amino acid sequence of purified von Willebrand factor-cleaving protease. *Blood* **98**, 1654-1661.
- Gill, R.T., Valdes, J.J. and Bentley, W.E. (2000) A comparative study of global stress gene regulation in response to overexpression of recombinant proteins in *Escherichia coli*. *Metabolic Eng.* **2**, 178-189.
- Glick, B. (1995) Metabolic load and heterologous gene expression. *Biotechnol. Adv.* **13**, 247-261.
- Gold, L. (1990) Expression of heterologous proteins in *Escherichia coli*. *Methods Enzymol.* **185**, 11-14.
- Goldstein, M.A. and Doi, R.H. (1995) Prokaryotic promoters in biotechnology. *Biotechnol. Annu. Rev.* **1**, 105-128.
- Goloubinoff, P., Gatenby, A.A. and Lorimer, G.H. (1989) GroE heat-shock proteins promote assembly of foreign prokaryotic ribulose biphosphate carboxylase oligomers in *Escherichia coli*. *Nature* **337**, 44-47.
- Gramatikoff, K., Georgiev, O. and Schaffner, W. (1994) Direct interaction rescue, a novel filamentous phage technique to study protein-protein interactions. *Nucleic Acids Res.* **22**, 5761-5762.
- Gribskov, M. and Burgess, R.R. (1983) Overproduction and purification of the sigma subunit of *Escherichia coli* RNA polymerase. *Gene* **26**, 109-118.
- Griffiths, A.D. and Duncan, A.R. (1998) Strategies for selection of antibodies by phage display. *Curr. Opin. Biotechnol.* **9**, 102-108.
- Gubler, U. and Hoffman, B.J. (1983) A simple and very effective method for generating cDNA libraries. *Gene* **25**, 263-269.

Gumpert, J. and Hoischen, C. (1998) Use of cell-wall-less bacteria (L-forms) for efficient expression and secretion of heterologous gene products. *Curr. Opin. Biotechnol.* **9**, 506-509.

Guzman, L-M., Belin, D., Carson, M.J. and Beckwith, J. (1995) Tight regulation, modulation, and high-level expression by vectors containing the arabinose P_{BAD} promoter. *J. Bacteriol.* **177**, 4121-4130.

Hagay, Y., Lahav, J., Levanon, A. and Panet, A. (2003) Function-modulating human monoclonal antibodies against platelet-membrane receptors isolated from a phage-display library. *J. Thromb. Haemost.* **1**, 1829-1836.

Hanes, J. and Plückthun, A. (1997) *In vitro* selection, and evolution of functional proteins by using ribosome display. *Proc. Natl. Acad. Sci. USA* **94**, 4937-4942.

Hanes, J. and Plückthun, A. (1999) *In vitro* selection methods for screening of peptide, and protein libraries. *Curr. Top. Microbiol. Immunol.* **243**, 107-122.

Hannig, G. and Makrides, S.C. (1998) Strategies for optimising heterologous protein expression in *Escherichia coli*. *Trends Biotechnol.* **16**, 54-60.

Harcum, S.W. and Bentley, W.E. (1993) Response dynamics of 26-, 34-, 39-, 54-, and 80-kDa proteases in induced cultures of recombinant *Escherichia coli*. *Biotechnol. Bioeng.* **42**, 675-685.

Hartl, F.U. (1996) Molecular chaperones in cellular protein folding. *Nature* **381**, 571-579.

Hartl, F.U. and Hayer-Hartl, M. (2002) Molecular chaperones in the cytosol: from nascent chain to folded protein. *Science* **295**, 1852-1858.

Hartley, O. (2002) The use of phage display in the study of receptors and their ligands. *Recept. Signal. Transduct. Res.* **22**, 373-392.

Hauser, H. and Zylstra, G.J. (2001) Mammals, plants, bacteria, and soil: common problems and solutions in gene expression and analysis. *Curr. Opin. Biotechnol.* **12**, 437-438.

Hawkins, P.R., Jin, P. and Fu, G.K. (2003) Full-length cDNA synthesis for long-distance RT-PCR of large mRNA transcripts. *Biotechniques* **34**, 768-773.

Hawley, D.K. and McClure, W.R. (1983) Compilation and analysis of *Escherichia coli* promoter DNA sequences. *Nucleic Acid Res.* **11**, 2237-2255.

Heiskanen, T., Lundkvist, A., Soliymani, R., Koivunene, E., Vaheri, A. and Lankinene, H. (1999) Phage-displayed peptides mimicking the discontinuous neutralisation sites of Puumala Hanta virus envelope glycoproteins. *Virology* **262**, 321-332.

Hiipakka, M., Poikonen, K. and Saksela, K. (1999) SH3 domains with high affinity and engineered ligand specificities targeted to HIV-1 Nef. *J. Mol. Biol.* **293**, 1097-1106.

Hill, H.R. and Stockley, P.G. (1996) Phage presentation. *Mol. Microbiol.* **20**, 685-692.

Hirsh, J. (2003) Current anticoagulant therapy – unmet clinical needs. *Thromb. Res.* **109**, S1-S8.

Hirsh, J. and Weitz, J.I. (1999) Thrombosis and anticoagulation. *Semin. Hematol.* **36**, 118-132.

Hockney, R.C. (1994) Recent developments in heterologous protein production in *Escherichia coli*. *Trends Biotechnol.* **12**, 456-463.

Hodgson, J. (1993) Expression systems: a user's guide. *Biotechnology* **11**, 887-893.

Hogrefe, H.H., Amberg, J.R., Hay, B.N., Sorge, J.A. and Shopes, B. (1993) Cloning in a bacteriophage lambda vector for the display of binding proteins on filamentous phage. *Gene* **137**, 85-91.

Holliger, P., Prospero, T. and Winter, G. (1993) Diabodies: small bivalent and bispecific antibody fragments. *Proc. Natl. Acad. Sci. USA.* **90**, 6444-6448.

Holliger, P., Richmann, L. and Williams, R.L. (1999) Crystal structure of the two N-terminal domains of g3p from the filamentous phage fd at 1.9 Å: Evidence for conformational lability. *J. Mol. Biol.* **288**, 649-657.

Holt, L.J., Enever, C., De Wildt, R.M.T. and Tomlinson, I.M. (2000) The use of recombinant antibodies in proteomics. *Curr. Opin. Biotechnol.* **11**, 445-449.

Hoogenboom, H.R., De Bruïne A.P., Hufton, S.E., Hoet, R.M., Arends, J.W. and Roovers, R.C. (1998) Antibody phage display and its applications. *Immunotechnology* **4**, 1–20.

Hoogenboom, H.R., Griffiths, A.D., Johnson, K.S, Chiswell, D.J., Hudson, P. and Winter, G. (1991) Multi-subunit proteins on the surface of filamentous phage: methodologies for displaying antibody (Fab) heavy and light chains. *Nucleic Acid Res.* **19**, 4133-4137.

Hopp, T.P., Prickett, K.S., Price, V., Libby, R.T., March, C.J., Cerretti, P., *et al.* (1988) A short polypeptide marker sequence useful for recombinant protein identification and purification. *Biotechnology* **6**, 1205-1210.

Houghten, R.A., Pinella, C., Blondelle, S.E., Appel, J.R., Dooley, C.T. and Cuervo, J.H. (1991) Generation and use of synthetic peptide combinatorial libraries for basic research and drug discovery. *Nature* **354**, 84-86.

Hsiung, H.M., Cantrell, A., Luirink, J., Oudega, B., Veros, A.J. and Becker, G.W. (1989) Use of bacteriocin release protein in *E. coli* for excretion of human growth hormone into the culture medium. *Biotechnology* **7**, 267-271.

Hufton, S.E., Moerkerk, P.T., Meulemans, E.V., De Bruïne, A., Arends, J.W. and Hoogenboom, H.R. (1999) Phage display of cDNA repertoires: the pVI display system and its applications for the selection of immunogenic ligands. *J. Immunol. Methods* **231**, 39-51.

Hungtington, J.A. and Baglin, T.P. (2003) Targeting thrombin – rational drug design from natural mechanisms. *Trends Pharm. Sci.* **24**, 589-595.

Huse, W.D., Sastry, L., Iverson, S.A., Kang, A.S., Alting-Mees, M., Burton, D.R., *et al.* (1989) Generation of a large combinatorial library of the immunoglobulin repertoire in phage lambda. *Science* **226**, 1315-1317.

Huynh, T.V., Young, R.A. and Davis, R.W. (1988) Construction and screening cDNA libraries in λ gt10 and λ gt11. In *DNA cloning: a practical approach, Volume I*. Glover, D.M. (ed). IRL Press Limited, Oxford.

Ikemura, T. (1981) Correlation between the abundance of *Escherichia coli* transfer RNAs and the occurrence of the respective codons in its protein genes. *J. Mol. Biol.* **146**, 1-21.

Jacob, F. and Monod, J. (1961) Genetic regulatory mechanisms in the synthesis of proteins. *J. Mol. Biol.* **3**, 318-356.

Jakes, K.S., Davis, N.G. and Zinder, N.D. (1988) A hybrid toxin from bacteriophage f1 attachment protein and colicin E3 has altered cell receptor specificity. *J. Bacteriol.* **170**, 4231-4238.

Jena, P.K., Liu, A.H., Smith, D.S. and Wysocki, L.J. (1996) Amplification of genes, single transcripts and cDNA libraries from one cell and direct sequence analysis of amplified products derived from one molecule. *J. Immunol. Methods* **190**, 199-213.

Jespers, L., Messens, J.H., De Keyser, A., Eeckhout, D., Van den Brande, I., Gansemans, Y.G., *et al.* (1995) Surface expression and ligand-based selection of cDNAs fused to filamentous phage gene VI. *Biotechnology* **13**, 378-382.

Jürgen, B., Lin, H.Y., Riemschneider, S., Scharf, C., Neubauer, P., Schmid, R., *et al.* (2000) Monitoring of genes that respond to overproduction of an insoluble recombinant protein in *Escherichia coli* glucose-limited fed-batch fermentations. *Biotechnol. Bioeng.* **70**, 217-224.

Kane, J.F. (1995) Effects of rare codon clusters on high-level expression of heterologous proteins in *Escherichia coli*. *Curr. Opin. Biotechnol.* **6**, 494-500.

Kato, S., Sekine, S., Oh, S-W., Kim, N-S., Umezawa, Y., Abe, N., *et al.* (1994) Construction of a human full-length cDNA bank. *Gene* **150**, 243-250.

Katoh, S. and Katoh, Y. (2000) Continuous refolding of lysozyme with fed-batch addition of denatured protein solution. *Process. Biochem.* **35**, 1119-1124.

Kaul, P.N. (1998) Drug discovery: past, present and future. *Prog. Drug Res.* **50**, 90-105.

Keller, P.M., Arnold, B.A., Shaw, A.R., Tolman, L.R., Van Middlesworth, F., Bondy, S., *et al.* (1993) Identification of HIV vaccine candidate peptides by screening random phage epitope libraries. *Virology* **193**, 709-716.

Kidwell, J., Kolibachuk, D. and Dennis, D. (1996) Regulated expression of the *Alcaligenes eutrophus pha* biosynthesis genes in *Escherichia coli*. *Biotechnol. Bioeng.* **50**, 108-114.

- Kikuchi, Y., Yoda, K., Yamasaki, M. and Tamura, G. (1981) The nucleotide sequence of the promoter and the amino-terminal region of alkaline phosphatase structural gene (*phoA*) of *Escherichia coli*. *Nucleic Acids Res.* **9**, 5671-5678.
- Kimmel, A.R. and Berger, S.L. (1987) Preparation of cDNA and the generation of cDNA libraries. *Methods Enzymol.* **152**, 307-316.
- Kishchenko, G., Batliwala, H. and Makowski, L. (1994) Structure of a foreign peptide displayed on the surface of bacteriophage-M13. *J. Mol. Biol.* **241**, 208-213.
- Knapp, A., Degenhardt, T. and Dodt, J. (1992) Hirudisins. Hirudin-derived thrombin inhibitors with disintegrin activity. *J. Biol. Chem.* **267**, 24230-24234.
- Knappik, A., Ge, L., Honegger, A., Pack, P., Fischer, M., Wellnhofer, G., *et al.* (2000) Fully synthetic human combinatorial antibody libraries (HuCAL) based on modular consensus frameworks and CDRs randomized with trinucleotides. *J. Mol. Biol.* **296**, 57-86.
- Köhler, G. and Milstein, C. (1975) Continuous cultures of fused cells secreting antibody of predefined specificity. *Nature* **256**, 495-497.
- Koivunen, E., Arap, W., Rajotte, D., Lahdenranta, J. and Pasqualini, R. (1999) Identification of receptor ligands with phage display peptide libraries. *J. Nucl. Med.* **40**, 883-888.
- Koivunen, E., Wang, B. and Ruoslahti, E. (1994) Isolation of a highly specific ligand for the alpha 5 beta 1 integrin from a phage display library. *J. Cell Biol.* **124**, 373-380.
- Konthur, Z. and Cramer, R. (2003) High-throughput applications of phage display in proteomic analyses. *Targets* **2**, 261-270.
- Konthur, Z. and Walter, G. (2002) Automation of phage display for high-throughput antibody development. *Targets* **1**, 30-36.
- Kost, T.A. (1997) Expression systems: gene expression systems in the genomic era. *Curr. Opin. Biotechnol.* **8**, 539-541.
- Krebber, C., Spada, S., Desplancq, D. and Plückthun, A. (1995) Co-selection of cognate antibody-antigen pairs by selectively-infective phages. *FEBS Lett.* **377**, 227-231.

Krebber, C., Spada, S., Desplancq, D., Krebber, A., Ge, L. and Plückthun, A. (1997) Selectively-infective phage (SIP): a mechanistic dissection of a novel *in vivo* selection for protein-ligand interactions. *J. Mol. Biol.* **268**, 607-618.

Kretzchmar, T. and Von Rüden, T. (2002) Antibody discovery: phage display. *Curr. Opin. Biotechnol.* **13**, 598-602.

Kurland, C. and Gallant, J. (1996) Errors of heterologous protein expression. *Curr. Opin. Biotechnol.* **7**, 489-493.

Kurucz, I., Titus, J.A., Jost, C.A. and Segal, D.M. (1995) Correct disulphide pairing and efficient refolding of detergent-solubilized single-chain Fv proteins from bacterial inclusion bodies. *Mol. Immunol.* **12**, 1443-1452.

Lambert, K.N. and Williamson, V.M. (1993) cDNA library construction from small amounts of RNA using paramagnetic beads and PCR. *Nucleic Acids Res.* **21**, 775-776.

Lang, I.M., Barbas, C.F. and Schleef, R.R. (1996) Recombinant rabbit Fab with binding activity to type-1 plasminogen activator derived from a phage-display library against human α -granules. *Gene* **172**, 295-298.

Lange, U., Keilholz, W., Schaub, G.A., Landmann, H., Markwardt, F. and Nowak, G. (1999) Biochemical characterization of a thrombin inhibitor from the bloodsucking bug *Dipetalogaster maximus*. *Haemostasis* **29**, 204-211.

LaVallie, E.R. and McCoy, J.M. (1995) Gene fusion expression systems in *Escherichia coli*. *Curr Opin. Biotechnol.* **6**, 501-506.

LaVallie, E.R., DiBlasio, E.A., Kovacic, S., Grant, K.L., Schendel, P.F. and McCoy, J.M. (1993) A thioredoxin gene fusion expression system that circumvents inclusion body formation in the *E. coli* cytoplasm. *Biotechnology* **11**, 187-193.

Lee, J., Kim, H.C., Kim, S.W., Kim, S.W., Hong, S.I. and Park, Y.H. (2002) Interplay of SOS induction, recombinant gene expression, and multimerization of plasmid vectors in *Escherichia coli*. *Biotechnol. Bioeng.* **80**, 84-92.

Legendre, D. and Fastrez, J. (2002) Construction and exploitation in model experiments of functional selection of a landscape library expressed from a phagemid. *Gene* **290**, 203-215.

Li, B., Fuh, G., Meng, G., Xin, X., Gerritsen, M.E., Cunningham, B., *et al.* (2000) Receptor-selective variants of human VEGF: generation and characterization. *J. Biol. Chem.* **275**, 29823-29828.

Li, F. and Aitken, R. (2004) Cloning of porcine scFv antibodies by phage display and expression in *Escherichia coli*. *Vet. Immunol. Immunopathol.* **97**, 39-51.

Li, M., Su, ZG. and Janson, J-C. (2004) *In vitro* refolding by chromatographic procedures. *Protein Express. Purif.* **33**, 1-10.

Lilie, H., Schwarz, E. and Rudolph, R. (1998) Advances in refolding of proteins produced in *E. coli*. *Curr. Opin. Biotechnol.* **9**, 497-501.

Linder, P. and Daugeron, M-C. (2000) Are DEAD-box proteins becoming respectable helicases? *Nat. Struct. Biol.* **7**, 97-99.

Lopez, P.J., Marchand, I., Joyce, S.A. and Dreyfus, M. (1999) The C-terminal half of RNase E, which organizes the *Escherichia coli* degradosome, participates in mRNA degradation but not rRNA processing *in vivo*. *Mol. Microbiol.* **33**, 188-199.

Lowman, H.B. (1997) Bacteriophage display and discovery of peptide leads for drug development. *Annu. Rev. Biophys. Biomol. Struct.* **26**, 401-424.

Lowman, H.B. and Wells, J.A. (1993) Affinity maturation of human growth hormone by monovalent phage display. *J. Mol. Biol.* **234**, 564-578.

Lowman, H.B., Bass, S.H., Simpson, N. and Wells, J.A. (1991) Selecting high-affinity binding proteins by monovalent phage display. *Biochemistry* **30**, 10832-10838.

Makrides, S.C. (1996) Strategies for achieving high-level expression of genes in *Escherichia coli*. *Microbiol. Rev.* **60**, 512-538.

Malik, P., Terry, T.D., Gowda, L.R., Langara, A., Petukhov, S.A., Symmons, M.F., *et al.* (1996) Role of capsid structure and membrane protein processing in determining the size and copy

number of peptides displayed on the major coat protein of filamentous bacteriophage. *J. Mol. Biol.* **260**, 9-21.

Malmborg, A.C., Duenas, M., Ohlin, M., Soderlind, E. and Borrebaeck, C.A. (1996) Selection of binders from phage displayed antibody libraries using the BIAcore biosensor. *J. Immunol. Methods* **198**, 51-57.

Mandel, M. and Higa, A. (1970) Calcium-dependent bacteriophage DNA infection. *J. Mol. Biol.* **53**, 159-162.

Mann, K.G. (1999) Biochemistry and physiology of blood coagulation. *Thromb. Haemost.* **82**, 165-174.

Mans, B.J., Louw, A.I. and Neitz, A.W.H. (2002) Amino acid sequence and structure modelling of savignin, a thrombin inhibitor from the tick, *Ornithodoros savignyi*. *Insect Biochem. Mol. Biol.* **32**, 821-828.

Marino, M.H. (1989) Expression systems for heterologous protein production. *BioPharm* **2**, 18-33.

Marks, J.D., Hoogenboom, H.R., Bonnert, T.P., McCafferty, J., Giffiths, A.D. and Winter, G. (1991) By-passing immunization: human antibodies from V-gene libraries displayed on phage. *J. Mol. Biol.* **222**, 581-597.

Marks, J.D., Ouwehand, W.H., Bye, J.M., Finnern, R., Gorick, B.D., Voak, D., *et al.* (1993) Human antibody fragments specific for human blood group antigens from a phage display library. *Biotechnology* **11**, 1145-1149.

Markwardt, F. (1957) Die isolierung und chemische charakterisierung des hirudins. *Hoppe Seylers Z. Physiol. Chem.* **312**, 85-89.

Markwardt, F. and Walsmann, P. (1967) Reindarstellung und analyse des thrombin inhibitors hirudin. *Hoppe Seylers Z. Physiol. Chem.* **348**, 1381-1386.

Maruyama, K. and Sugano, S. (1994) Oligo-capping: a simple method to replace the cap structure of eukaryotic mRNAs with oligoribonucleotides. *Gene* **138**, 171-174.

Marvin, D.A. (1998) Filamentous phage structure, infection and assembly. *Curr. Opin. Struct. Biol.* **8**, 150-158.

Marvin, D.A., Hale, R.D., Nave, C. and Citterich, M.H. (1994) Molecular models and structural comparisons of native and mutant class I filamentous bacteriophages: Fd (fd, f1, M13), If1 and IKe. *J. Mol. Biol.* **235**, 260-286.

Mattheakis, L.C., Bhatt, R.R. and Dower, W.J. (1994) An *in vitro* polysome display system for identifying ligands from very large peptide libraries. *Proc. Natl. Acad. Sci. USA* **91**, 9022–9026.

Matthews, D.J. and Wells, J.A. (1993) Substrate phage: selection of protease substrates by monovalent phage display. *Science* **260**, 1113-1117.

McCafferty, J., Griffiths, A.D., Winter, G. and Chiswell, D.J. (1990) Phage antibodies: filamentous phage displaying antibody variable domains. *Nature* **348**, 552-554.

McCarrey, J.R. and Williams, S.A. (1994) Construction of cDNA libraries from limiting amounts of material. *Curr. Opin. Biotechnol.* **5**, 34-39.

McDowell, R.S., Blackburn, B.K., Gadek, T.R., McGee, L.R., Rawson, T., *et al.* (1994) From peptide to non-peptide: the de novo design of potent, non-peptidal inhibitors of platelet aggregation based on a benzodiazepinedione scaffold. *J. Am. Chem. Soc.* **116**, 5077–5083.

McGregor, D. (1996) Selection of proteins or peptides from libraries displayed on filamentous bacteriophage. *Mol. Biotechnol.* **6**, 155-162.

McNulty, D.E., Claffee, B.A., Huddleston, M.J. and Kane, J.F. (2003) Mistranslational errors associated with the rare arginine codon CGG in *Escherichia coli*. *Protein Express. Purif.* **27**, 365-374.

Mead, D.A. and Kemper, B. (1988) Chimeric single-stranded DNA phage-plasmid cloning vectors. *Biotechnology* **10**, 85-102.

Meiring, M.S., Litthauer, D., Hársfalvi, J., Van Wyk, V., Badenhorst, P.N. and Kotzé, H.F. (2002) *In vitro* effect of a thrombin inhibition peptide selected by phage display technology. *Thromb. Res.* **107**, 365-371.

Mertens, N., Remaut, E. and Fiers, W. (1995a) Tight transcriptional control mechanism ensures stable high-level expression from T7 promoter-based expression plasmids. *Biotechnology* **13**, 175-179.

Mertens, N., Remaut, E. and Fiers, W. (1995b) Versatile, multi-featured plasmids for high-level expression of heterologous genes in *Escherichia coli*: overproduction of human and murine cytokines. *Gene* **164**, 9-15.

Miczak, A., Kaberdin, V.R., Wei, C.L. and Lin-Chao, S. (1996) Proteins associated with RNase E in a multicomponent ribonucleolytic complex. *Proc. Natl. Acad. Sci. USA* **93**, 3865-3869.

Middelberg, A.P.J. (2002) Preparative protein folding. *Trends Biotechnol.* **20**, 437-443.

Misawa, S. and Kumagai, I. (1999) Refolding of therapeutic proteins produced in *Escherichia coli* as inclusion bodies. *Bioploymers* **51**, 297-307.

Missiakas, D. and Raina, S. (1997) Protein refolding in the bacterial periplasm. *J. Bacteriol.* **179**, 2465-2471.

Missiakas, D., Betton, J-M. and Raina, S. (1996) New components of protein folding in extracytoplasmic compartments of *Escherichia coli* SurA, FkpA and Skp/OmpH. *Mol. Microbiol.* **21**, 871-884.

Mukhopadhyay, A. (1997) Inclusion bodies and purification of proteins in biologically active forms. *Adv. Biochem. Eng. Biotechnol.* **56**, 61-109.

Mullaney, B.P. and Pallavicini, M.G. (2001) Protein-protein interactions in haematology and phage display. *Exp. Hematol.* **29**, 1136-1146.

Mullis, K.B. and Faloona, F.A. (1987) Specific synthesis of DNA *in vitro* via a polymerase-catalysed reaction. *Methods Enzymol.* **155**, 335-350.

Nielsen, U.B. and Marks, J.D. (2000) Internalizing antibodies and targeted cancer therapy: direct selection from phage display libraries. *Pharma. Sci. Technol. Today* **3**, 282-291.

Nilsson, B. and Abrahmsén, L. (1990) Fusions to staphylococcal protein A. *Methods Enzymol.* **185**, 144-161.

Nilsson, J., Ståhl, S., Lundeberg, J., Uhlén, M. and Nygren, P-A. (1997) Affinity fusion strategies for detection, purification, and immobilization of recombinant proteins. *Protein Express. Purif.* **11**, 1-16.

Nishihara, T., Iwabuchi, T. and Nohno, T. (1994) A T7 promoter vector with a transcriptional terminator for stringent expression of foreign genes. *Gene* **145**, 145-146.

Niwa, M., Maruyama, H., Fujimoto, T., Dohi, K. and Maruyama, I.N. (2000) Affinity selection of cDNA libraries by λ phage surface display. *Gene* **256**, 229-236.

Noeske-Jungblut, C., Haendler, B., Donner, P., Alagon, A., Possani, L. and Schleuning, W.D. (1995) Triabin, a highly potent exosite inhibitor of thrombin. *J. Biol. Chem.* **270**, 28629-28634.

Noren, K.A. and Noren, C.J. (2001) Construction of high-complexity combinatorial phage display peptide libraries. *Methods* **23**, 169-178.

Novagen (2003) *pET system manual* 10th ed. Madison, WI, USA.

O'Neil, K.T. and Hoess, R.H. (1995) Phage display: protein engineering by directed evolution. *Curr. Opin. Struct. Biol.* **5**, 443-449.

O'Neil, K.T., Hoess, R.H., Jackson, S.A., Ramachandran, N.S., Mousa, S.A. and DeGrado, W.F. (1992) Identification of novel peptide antagonists for GPIIb/IIIa from a conformationally constrained phage peptide library. *Proteins* **14**, 509-515.

Obukowicz, M.G., Turner, M.A., Wong, E.Y. and Tacon, W.C. (1988) Secretion and export of IGF-1 in *Escherichia coli* strain JM101. *Mol. Gen. Genet.* **215**, 19-25.

Oh, J.H., Kim, Y.S. and Kim, N.S. (2003) An improved method for constructing a full-length enriched cDNA library using small amounts of total RNA as a starting material. *Exp. Mol. Med.* **35**, 586-590.

Ohkubo, S., Miyadera, K., Sugimoto, Y., Matsuo, K., Wierzba, K. and Yamada, Y. (1999) Identification of substrate sequences for membrane type-1 matrix metalloproteinase using bacteriophage peptide display library. *Biochem. Biophys. Res. Commun.* **266**, 308-313.

Okayama, H. and Berg, P. (1982) High-efficiency cloning of full-length cDNA. *Mol. Cell. Biol.* **2**, 161-170.

Okayama, H. and Berg, P. (1983) A cDNA vector that permits expression of cDNA inserts in mammalian cells. *Mol. Cell. Biol.* **3**, 280-289.

Okayama, H., Kawaichi, M., Brownstein, M., Lee, F., Yokota, T. and Arai, K. (1987) High-efficiency cloning of full-length cDNA: construction and screening of cDNA expression libraries for mammalian cells. *Methods Enzymol.* **154**, 3-28.

Olins, P.O. (1996) Quantity versus authenticity of heterologously produced proteins: an inevitable compromise? *Curr. Opin. Biotechnol.* **7**, 487-488.

Olins, P.O. and Lee, S.C. (1993) Recent advances in heterologous gene expression in *Escherichia coli*. *Curr. Opin. Biotechnol.* **4**, 520-525.

Olofsson, L., Ankarloo, J., Andersson, P.O. and Nicholls, I.A. (2001) Filamentous bacteriophage stability in non-aqueous media. *Chem. Biol.* **8**, 661-671.

Olson, P., Zhang, Y., Olsen, D., Owens, A., Cohen, P., Nguyen, K., *et al.* (1998) High-level expression of eukaryotic polypeptides from bacterial chromosomes. *Protein Express. Purif.* **14**, 160-166.

Orlandi, R., Gussow, D.H., Jones, P.T. and Winter, G. (1989) Cloning immunoglobulin variable domains for expression by the polymerase chain reaction. *Proc. Natl. Acad. Sci. USA* **86**, 3833-3837.

Osbourn, J., Jermutus, L. and Duncan, A. (2003) Current methods for the generation of human antibodies for the treatment of autoimmune diseases. *Drug Discov. Today* **8**, 845-851.

Pan, S.H. and Malcolm, B.A. (2000) Reduced background expression and improved plasmid stability with pET vectors in BL21 (DE3). *Biotechniques* **29**, 1234-1238.

Parmley, S.F. and Smith, G.P. (1988) Antibody-selectable filamentous fd phage vectors: affinity purification of target genes. *Gene* **73**, 305-318.

Pasqualini, R. and Ruoslahti, E. (1996) Organ targeting *in vivo* using phage display peptide libraries. *Nature* **380**, 364–366.

Patanjali, S., Parimoo, S. and Weissman, S.M. (1991) Construction of a uniform-abundance (normalized) cDNA library. *Proc. Natl. Acad. Sci. USA* **88**, 1943-1947.

Patra, A.K., Mukhopadhyay, R., Muhhija, R., Krishnan, A., Garg, L.C. and Panda, A.K. (2000) Optimization of inclusion body solubilization and renaturation of recombinant human growth hormone from *Escherichia coli*. *Protein Express. Purif.* **18**, 182-192.

Pawson, T., Raina, M. and Nash, P. (2002) Interaction domains: from simple binding events to complex cellular behavior. *FEBS Lett.* **513**, 2-10.

Peacock, S.L., Mclver, C.M. and Monahan, J.J. (1981) Transformation of *E. coli* using homopolymer-linked plasmid chimeras. *Biochim. Biohyps. Acta* **655**, 243-250.

Pedersen, H., Holder, S., Sutherlin, D.P., Schwitter, U., King, D.S. and Schultz, P.G. (1998) A method for directed evolution and functional cloning of enzymes. *Proc. Natl. Acad. Sci. USA* **95**, 10523-10528.

Pedersen-Lane, J., Maley, G.F., Chu, E. and Maley, F. (1997) High-level expression of human thymidylate synthase. *Protein Express. Purif.* **10**, 256-262.

Petrenko, V.A. and Smith, G.P. (2000) Phages from landscape libraries as substitute antibodies. *Protein Eng.* **13**, 589–592.

Petrenko, V.A. and Vodyanoy, V.J. (2003) Phage display for detection of biological threat agents. *J. Microbiol. Methods* **53**, 253-262.

Petrenko, V.A., Smith, G.P., Gong, X. and Quinn, T. (1996) A library of organic landscapes on filamentous phage. *Protein Eng.* **9**, 797–801.

Petrenko, V.A., Smith, G.P., Mazooji, M.M. and Quinn, T. (2002) α -Helically constrained phage display libraries. *Protein Eng.* **15**, 943-950.

Pimanda, J.E., Maekawa, A., Wind, T., Paxton, J., Chesterman, C.N. and Hogg, P.J. (2003) Congenital thrombotic thrombocytopenic purpura in association with a mutation in the second CUB domain of ADAMTS 13. *Blood* **103**, 627-629.

Possee, R.D. (1997) Baculoviruses as expression vectors. *Curr. Opin. Biotechnol.* **8**, 569-572.

Poul, M.A., Becerril, B., Nielsen, U.B., Morisson, P. and Marks, J.D. (2000) Selection of tumor-specific internalizing human antibodies from phage libraries. *J. Mol. Biol.* **301**, 1149-1161.

Prinz, W.A., Åslund, F., Holmgren, A. and Beckwith, J. (1997) The role of the thioredoxin and glutaredoxin pathways in reducing protein disulfide bonds in the *Escherichia coli* cytoplasm. *J. Biol. Chem.* **272**, 15661-15667.

Rader, C. and Barbas III, C.F. (1997) Phage display of combinatorial antibody libraries. *Curr. Opin. Biotechnol.* **8**, 503-508.

Rai, M. and Padh, H. (2001) Expression systems for production of heterologous proteins. *Curr. Science* **80**, 1121-1128.

Rakonjac, J., Feng, J. and Model, P. (1999) Filamentous phage are released from the bacterial membrane by a two-step mechanism involving a short C-terminal fragment of pIII. *J. Mol. Biol.* **289**, 1253-1265.

Reichert, J.M. (2000) New biopharmaceuticals in the USA: trends in development and marketing approvals 1995-1999. *Trends Biotechnol.* **18**, 364-369.

Ren, Z and Black, L.W. (1998) Phage T4 SOC and HOC display of biologically active, full-length proteins on the viral capsid. *Gene* **215**, 439-444.

Resuehr, D. and Spiess, A-N. (2003) A real-time polymerase chain reaction-based evaluation of cDNA synthesis priming methods. *Anal. Biochem.* **322**, 287-291.

Rhyner, C., Kodzius, R. and Cramer, R. (2002) Direct selection of cDNAs from filamentous phage surface display libraries: potential and limitations. *Curr. Pharm. Biotechnol.* **3**, 13-21.

Ribeiro, J.M.C. (1995) Blood-feeding arthropods: live syringes or invertebrate pharmacologists? *Infect. Agents Disease* **4**, 143-152.

- Richardson, A., Landry, S.J. and Georgopoulos, C. (1998) The ins and outs of a molecular chaperone machine. *Trends Biochem. Sci.* **23**, 138-143.
- Roberts, B.L., Markland, W., Ley, A.C., Kent, R.B., White, D.W., Guterman, S.K., *et al.* (1992) Directed evolution of a protein: selection of potent neutrophil elastase inhibitors displayed on M13 fusion phage. *Proc. Natl. Acad. Sci. USA* **89**, 2429-2433.
- Rodi, D.J. and Makowski, L. (1999) Phage-display technology – finding a needle in a vast molecular haystack. *Curr. Opin. Biotech.* **10**, 87-93.
- Rodi, D.J., Soares, A.S. and Makowski, L. (2002) Quantitative assessment of peptide sequence diversity in M13 combinatorial peptide phage display libraries. *J. Mol. Biol.* **322**, 1039-1052.
- Rogl, H., Kosemund, K., Kuühlbrandt, W. and Collison, I. (1998) Refolding of *Escherichia coli* produced membrane protein inclusion bodies immobilised by nickel chelating chromatography. *FEBS Lett.* **432**, 21-26.
- Rondot, S., Koch, J., Breitling, F. and Dübel, S. (2001) A helper phage to improve single-chain antibody presentation in phage display. *Nature Biotechnol.* **19**, 75–78.
- Roovers, R.C., Van der Linden, E., Zijlema, H., De Bruïne, A., Arends, J.W. and Hoogenboom, H.R. (2001) Evidence for a bias toward intracellular antigens in the local humoral anti-tumor immune response of a colorectal cancer patient revealed by phage display. *Int. J. Cancer* **93**, 832-840.
- Rosander, A., Bjerketorp, J., Frykberg, L. and Jacobsson, K. (2002) Phage display as a novel screening method to identify extracellular proteins. *J. Microbiol. Methods* **51**, 43-55.
- Rozinov, M.N. and Nolan, G.P. (1998) Evolution of peptides that modulate the spectral qualities of bound, small-molecule fluorophores. *Chem. Biol.* **5**, 713–728.
- Ruan, B., Hoskins, J., Wang, L. and Bryan, P.N. (1998) Stabilizing the subtilisin BPN' pro-domain by phage display selection: how restrictive is the amino acid code for maximum protein stability? *Protein Sci.* **7**, 2345-2353.

Rudert, F., Woltering, C., Frisch, C., Rottenberger, C. and Ilag, L.L. (1998) A phage-based system to select multiple protein-protein interactions simultaneously from combinatorial libraries. *FEBS Lett.* **440**, 135-140.

Ruggeri, Z.M. (2000) Role of von Willebrand factor in platelet thrombus formation. *Ann. Med.* **32** S1, 2-9.

Russel, M. (1991) Filamentous phage assembly. *Mol. Microbiol.* **5**, 1607-1613.

Russel, M., Linderoth, N.A. and Šali, A. (1997) Filamentous phage assembly: variation on a protein export theme. *Gene* **192**, 23-32.

Russell, D.R. and Bennett, G.N. (1982) Cloning of small DNA fragments containing the *Escherichia coli* tryptophan operon promoter and operator. *Gene* **17**, 9-18.

Rutledge, R.G., Seligy, V.L., Cote, M.J., Dimock, K., Lewin, L.L. and Tenniswood, M.P. (1988) Rapid synthesis and cloning of complementary DNA from any RNA molecule into plasmid and phage M13 vectors. *Gene* **68**, 151-158.

Salzet, M., Chopin, V., Baert, J., Matias, I. and Malecha, J. (2000) Theromin, a novel leech thrombin inhibitor. *J. Biol. Chem.* **275**, 30774-30780.

Sambrook, J. and Russell, D.W. (2001) *Molecular cloning: a laboratory manual*, 3rd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor.

Sandkvist, M. and Bagdasarian, M. (1996) Secretion of recombinant proteins by Gram-negative bacteria. *Curr. Opin. Biotechnol.* **7**, 505-511.

Santi, E., Capone, S., Mennuni, C., Lahm, A., Tramontano, A., Luzzago, A., *et al.* (2000) Bacteriophage lambda display of complex cDNA libraries: a new approach to functional genomics. *J. Mol. Biol.* **296**, 497-508.

Santini, C., Brennan, D., Mennuni, C., Hoess, R.H., Nicosia, A., Cortese, R., *et al.* (1998) Efficient display of an HCV cDNA expression library as C-terminal fusion to the capsid protein D of bacteriophage lambda. *J. Mol. Biol.* **282**, 125-135.

Sarmientos, P., Duchesne, M., Denèfle, P., Boiziau, J., Fromage, N., Delporte, N., *et al.* (1989) Synthesis and purification of active human tissue plasminogen activator from *Escherichia coli*. *Biotechnology* **7**, 495-501.

Schaefer, B. (1995) Revolution in rapid amplification of cDNA ends: new strategies for polymerase chain reaction cloning of full-length cDNA ends. *Anal. Biochem.* **227**, 255-273.

Schäffner, J., Winter, J., Rudolph, R. and Schwarz, E. (2001) Cosecretion of chaperones and low-molecular-size medium additives increase the yield of recombinant disulfide-bridged proteins. *Appl. Environ. Microbiol.* **67**, 3994-4000.

Sche, P.P., McKenzie, K.M., White, J.D. and Austin, D.J. (1999) Display cloning: functional identification of natural product receptors using cDNA-phage display. *Chem. Biol.* **6**, 707-716.

Schmid, S.R. and Linder, P. (1992) DE-A-D protein family of putative RNA helicases. *Mol. Microbiol.* **6**, 283-292.

Scott, J.K. and Smith, G.P. (1990) Searching for peptide ligands with an epitope library. *Science* **249**, 386-390.

Shatzman, A.R. (1995) Expression systems. *Curr. Opin. Biotechnol.* **6**, 491-493.

Sheets, M.D., Amersdorfer, P., Finnern, R., Sargent, P., Lindquist, E., Schier, R., *et al.* (1998) Efficient construction of a large nonimmune phage antibody library: the production of high-affinity human single-chain antibodies to protein antigens. *Proc. Natl. Acad. Sci. USA* **95**, 6157-6162.

Sidhu, S.S. (2000) Phage display in pharmaceutical biotechnology. *Curr. Opin. Biotechnol.* **11**, 610-616.

Sidhu, S.S. (2001) Engineering M13 for phage display. *Biomol. Eng.* **18**, 57-63.

Sieber, V., Plückthun, A. and Schmid, F.X. (1998) Selecting proteins with improved stability by a phage-based method. *Nat. Biotechnol.* **16**, 955-960.

Siegel, D.L. and Silberstein, L.E. (1994) Expression and characterization of recombinant anti-Rh(D) antibodies on filamentous phage: a model system for isolating human red blood cell antibodies by repertoire cloning. *Blood* **83**, 2334-2344.

- Simmons, L.C. and Yansura, D.G. (1996) Translation level is a critical factor for the secretion of heterologous proteins in *Escherichia coli*. *Nat. Biotechnol.* **14**, 629-634.
- Skerra, A. and Plückthun, A. (1988) Assembly of a functional immunoglobulin Fv fragment in *Escherichia coli*. *Science* **240**, 1038-1041.
- Smith, D.B. and Johnson, K.S. (1988) Single-step purification of polypeptides expressed in *Escherichia coli* as fusions with glutathione S-transferase. *Gene* **67**, 31-40.
- Smith, G.P. (1985) Filamentous fusion phage: novel expression vectors that display cloned antigens on the virion surface. *Science* **228**, 1315-1317.
- Smith, G.P. (1991) Surface presentation of protein epitopes using bacteriophage expression systems. *Curr. Opin. Biotechnol.* **2**, 668-673.
- Smith, G.P. and Petrenko, V.A. (1997) Phage display. *Chem. Rev.* **97**, 391-410.
- Smith, G.P. and Scott, J.K. (1993) Libraries of peptides and proteins displayed on filamentous phage. *Methods Enzymology* **217**, 228-257.
- Smith, J.W., Tachias, K. and Madison, E.L. (1995) Protein loop grafting to construct a variant of tissue-type plasminogen activator that binds platelet integrin α IIb β 3. *J. Biol. Chem.* **270**, 30486-30490.
- Soejima, K., Matsumoto, M., Kokame, K., Yagi, H., Ishizashi, H., Maeda, H., *et al.* (2003) ADAMTS 13 cysteine-rich/spacer domains are functionally essential for von Willebrand factor cleavage. *Blood* **102**, 3232-3237.
- Sonenberg, N. (1988) Cap-binding proteins of eukaryotic messenger RNA: functions in initiation and control of translation. *Prog. Nucleic Acid Res. Mol. Biol.* **35**, 173-207.
- Soumillion, P., Jespers, L., Boucher, M., Marchand-Brynaert, J., Winter, G. and Fastrez, J. (1994) Selection of α -lactomase on filamentous bacteriophage by catalytic activity. *J. Mol. Biol.* **237**, 415-422.

- South, V., Searfoss, G.H., French, S., Cheadle, C., Murray, E., Howk, R., *et al.* (1995) Identification of novel peptide antagonists for von Willebrand factor binding to the platelet glycoprotein Ib receptor from a phage epitope library. *Thromb. Haemost.* **73**, 144-150.
- Spickofsky, N. and Margolskee, R.F. (1991) A highly efficient directional cDNA cloning method utilizing an asymmetrically tailed linker-primer plasmid. *Nucleic Acids Res.* **19**, 7105-7111.
- Spragg, D.D., Alford, D.R., Greferath, R., Larsen, C.E., Lee, K.D., Gurtner, G.C., *et al.* (1997) Immunotargeting of liposomes to activated vascular endothelial cells: a strategy for site-selective delivery in the cardiovascular system. *Proc. Natl. Acad. Sci. USA* **94**, 8795-8800.
- Stader, J.A. and Silhavy, T.J. (1990) Engineering *Escherichia coli* to secrete heterologous gene products. *Methods Enzymol.* **185**, 166-187.
- Stark, K.R. and James, A.A. (1996) Anticoagulants in vector arthropods. *Parasitology Today* **12**, 430-437.
- Stemmer, W.P. (1994) Rapid evolution of a protein *in vitro* by DNA shuffling. *Nature* **370**, 389-391.
- Stephen, C.W. and Lane, D.P. (1992) Mutant conformation of p53: precise epitope mapping using a filamentous phage epitope library. *J. Mol. Biol.* **225**, 577-583.
- Sternberg, N. and Hoess, R.H. (1995) Display of peptides and proteins on the surface of bacteriophage lambda. *Proc. Natl. Acad. Sci. USA* **92**, 1609-1613.
- Stolz, J., Ludwig, A. and Sauer, N. (1998) Bacteriophage lambda surface display of a bacterial biotin acceptor domain reveals the minimal peptide size required for biotinylation. *FEBS Lett.* **440**, 213-217.
- Stone, S.R. and Hofsteenge, J. (1986) Kinetics of inhibition of thrombin by hirudin. *Biochemistry* **25**, 4622-4628.
- Strube, K.H., Kröger, B., Bialojan, S., Otte, M. and Dodt, J. (1993) Isolation, sequence analysis, and cloning of haemadin: an anticoagulant peptide from the Indian leech. *J. Biol. Chem.* **268**, 8590-8595.

Studier, F.W. (1991) Use of bacteriophage T7 lysozyme to improve an inducible T7 expression system. *J. Mol. Biol.* **219**, 37-44.

Studier, F.W. and Moffat, B.A. (1986) Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. *J. Mol. Biol.* **189**, 113-130.

Studier, F.W., Rosenberg, A.H., Dunn, J.J. and Dubendorff, J.W. (1990) Use of T7 RNA polymerase to direct expression of cloned genes. *Methods Enzymol.* **185**, 60-89.

Stueber, D. and Bujard, H. (1982) Transcription from efficient promoters can interfere with plasmid replication and diminish expression of plasmid specified genes. *EMBO J.* **1**, 1399-1404.

Sugahara, Y., Carninci, P., Itoh, M., Shibata, K., Konno, H., Endo, T., *et al.* (2001) Comparative evaluation of 5'-end-sequence quality of clones in CAP trapper and other full-length-cDNA libraries. *Gene* **263**, 93-102.

Suggett, S., Kirchhofer, D., Hass, P., Lipari, T., Moran, P., Nagel, M., *et al.* (2000) Use of phage display for the generation of human antibodies that neutralize factor IXa function. *Blood Coagul. Fibrinolysis* **11**, 27-42.

Suzuki, Y., Yamashita, R., Nakai, K. and Sugano, S. (2002) DBTSS: Database of human transcriptional start sites and full-length cDNAs. *Nucleic Acids Res.* **30**, 328-331.

Swartz, J.R. (2001) Advances in *Escherichia coli* production of therapeutic proteins. *Curr. Opin. Biotechnol.* **12**, 195-201.

Szardenings, M. (2003) Phage display of random peptide libraries: applications, limits, and potential. *J. Recept. Transduct. Res.* **23**, 307-349.

Tabor, S. and Richardson, C.C. (1985) A bacteriophage T7 RNA polymerase/promoter system for controlled exclusive expression of specific genes. *Proc. Natl. Acad. Sci USA* **82**, 1074-1078.

Taddei, F., Vulic, M., Radman, M. and Matic, I. (1997) Genetic variability and adaptation to stress. *EXS* **83**, 271-290.

- Tanaka, A.S., Silva, M.M., Noguti, M.A.E., Sampaio, C.A.M., Fritz, H. and Auerswald, E.A. (1999) Functional phage display of leech-derived tryptase inhibitor (LDTI): construction of a library and selection of thrombin inhibitors. *FEBS Lett.* **458**, 11-16.
- Tawfik, D.S., Green, B.S., Charp, R., Sela, M. and Eshhar, Z. (1993) catEILSA: a facile general route to catalytic antibodies. *Proc. Natl. Acad. Sci. USA* **90**, 373-377.
- Thomas, J.G. and Baneyx, F. (1996) Protein misfolding and inclusion body formation in recombinant *Escherichia coli* cells overproducing heat-shock proteins. *J. Biol. Chem.* **271**, 11141-11147.
- Thomas, J.G. and Baneyx, F. (1997) Divergent effects of chaperone overexpression and ethanol supplementation on inclusion body formation in recombinant *Escherichia coli*. *Protein Express. Purif.* **11**, 289-296.
- Thompson, J.A., Radonovich, M.F. and Salzman, N.P. (1979) Characterization of the 5'-terminal structure of simian virus 40 early mRNA's. *J. Virology* **31**, 437-446.
- Tsai, H.M. (1996) Physiologic cleavage of von Willebrand factor by a plasma protease is dependent on its conformation and requires calcium ion. *Blood* **87**, 4235-4244.
- Tsumoto, K., Ejima, D., Kumagai, I. and Arakawa, T. (2003) Practical considerations in refolding proteins from inclusion bodies. *Protein Express. Purif.* **28**, 1-8.
- Tulinsky, A. (1996) Molecular interactions of thrombin. *Semin. Thromb. Hemost.* **22**, 117-124.
- Turk, B.E. and Cantley, L.C. (2003) Peptide libraries: at the crossroads of proteomics and bioinformatics. *Curr. Opin. Chem. Biol.* **7**, 84-90.
- Uhlén, M., Nilsson, B., Guss, B., Lindberg, M., Gatenbeck, S. and Philipson, L. (1983) Gene fusion vectors based on staphylococcal protein A. *Gene* **23**, 369-378.
- Urata, J., Shojo, H. and Kaneko, Y. (2003) Inhibition of haematophagous invertebrate compounds acting on the host blood coagulation and platelet aggregation pathways. *Biochimie* **85**, 493-500.

Valenzuela, J.G., Francischetti, I.M.B. and Ribeiro, J.M.C. (1999) Purification, cloning, and synthesis of a novel salivary anti-thrombin from the mosquito *Anopheles albimanus*. *Biochemistry* **38**, 11209-11215.

Van de Locht, A., Lamba, D., Bauer, M., Huber, R., Friedrich, T., Kröger, B., *et al.* (1995) Two heads are better than one: crystal structure of the insect derived double domain Kazal inhibitor rhodniin in complex with thrombin. *EMBO J.* **14**, 5149-5157.

Van de Locht, A., Stubbs, M.T., Bode, W., Friedrich, T., Bollschweiler, C., Hoffken, W., *et al.* (1996) The ornithodorin-thrombin crystal structure, a key to the TAP enigma? *EMBO J.* **15**, 6011-6017.

Van Den Brink, E.N., Turenhout, E.A., Bank, C.M.C., Fijnvandraat, K., Peters, M. and Voorberg, J. (2000b) Molecular analysis of human anti-factor VIII antibodies by V gene phage display identifies a new epitope in the acidic region following the A2 domain. *Blood* **96**, 540-545.

Van Den Brink, E.N., Turenhout, E.A., Davies, J., Bovenschen, N., Fijnvandraat, K., Ouwehand, W.H., *et al.* (2000a) Human antibodies with specificity for the C2 domain of factor VIII are derived from VH1 germline genes. *Blood* **95**, 558-863.

Van Dijk, J.M., De Jong, A., Smith, H., Bron, S. and Venema, G. (1991) Signal peptidase I overproduction results in increased efficiencies of export and maturation of hybrid secretory proteins in *Escherichia coli*. *Mol. Gen. Genet.* **227**, 40-48.

Varnerin, J.P., Smith, T., Rosenblum, C.I., Vongs, A., Murphy, B.A., Nunes, C., *et al.* (1998) Production of leptin in *Escherichia coli*: a comparison of methods. *Protein Express. Purif.* **14**, 335-342.

Vasina, J.A. and Baneyx, F. (1996) Recombinant protein expression at low temperatures under the transcriptional control of the major *Escherichia coli* cold shock promoter *cspA*. *Appl. Environ. Microbiol.* **62**, 1444-1447.

Viaene, A., Crab, A., Meiring, M.S., Pritchard, D.I., Jepers, L. and Deckmyn, H. (2001) Identification of a collagen-binding protein from *Necator americanus* by using a cDNA-expression phage display library. *J. Parasitology* **87**, 619-625.

- Villard, S., Lacroix-Desmazes, S., Kieber-Emmons, T., Piquer, D., Grailly, S., Benhida, A., *et al.* (2003) Peptide decoys selected by phage display block *in vitro* and *in vivo* activity of human anti-FVIII inhibitor. *Blood* **102**, 949-952.
- Wagner, D.D. and Burger, P.C. (2003) Platelets in inflammation and thrombosis. *Arterioscler. Thromb. Vasc. Biol.* **23**, 2131-2137.
- Wall, J.G. and Plückthun, A. (1995) Effects of overexpressing folding modulators on the *in vivo* folding of heterologous proteins in *Escherichia coli*. *Curr. Opin. Biotechnol.* **6**, 507-516.
- Wallace, A., Dennis, S., Hofsteenge, J. and Stone, S.R. (1989) Contribution of the N-terminal region of hirudin to its interaction with thrombin. *Biochemistry* **28**, 10079-10084.
- Wang, J., Hu, L., Hamilton, S.R., Coombes, K.R. and Zhang, W. (2003) RNA amplification strategies for cDNA microarray experiments. *Biotechniques* **34**, 394-400.
- Watson, J.D. (1990) The human genome project: past, present, and future. *Science* **248**, 44-49.
- Watters, J.M., Telleman, P. and Junghans, R.P. (1997) An optimized method for cell based phage display panning. *Immunotechnology* **3**, 21-29.
- Weickert, M.J., Doherty, D.H., Best, E.A. and Olins, P.O. (1996) Optimization of heterologous protein production in *Escherichia coli*. *Curr. Opin. Biotechnol.* **7**, 494-499.
- Weitz, J.I. and Hirsh, J. (2001) New anticoagulant drugs. *Chest* **119**, S95-S107.
- Weitz, J.I., Hudoba, M., Massel, D., Maraganore, J. and Hirsh, J. (1990) Clot-bound thrombin is protected from inhibition by heparin-antithrombin III but is susceptible to inactivation by antithrombin III-independent inhibitors. *J. Clin. Invest.* **86**, 385-391.
- West, S.M., Chaudhuri, J.B. and Howell, J.A. (1998) Improved protein refolding using hollow-fibre membrane dialysis. *Biotechnol. Bioeng.* **57**, 590-599.
- Wilkinson, D. and Harrison, R.G. (1991) Predicting the solubility of recombinant proteins in *Escherichia coli*. *Biotechnology* **9**, 443-448.
- Willats, W.G. (2002) Phage display: practicalities and prospects. *Plant Mol. Biol.* **50**, 837-854.

- Wilson, K.S. and Von Hppel, P.H. (1995) Transcription termination at intrinsic terminators: the role of the RNA hairpin. *Proc. Natl. Acad. Sci. USA* **92**, 8793-8797.
- Winter, G., Griffiths, A.D., Hawkins, R.E. and Hoogenboom, H.R. (1994) Making antibodies by phage display technology. *Annu. Rev. Immunol.* **12**, 433–455.
- Wirsching, F., Opitz, T., Dietrich, R. and Schwienhorst, A. (1997) Display of functional thrombin inhibitor hirudin on the surface of phage M13. *Gene* **204**, 177-184.
- Wolfe, S.A., Greisman, H.A., Ramm, E.I. and Pabo, C.O. (1999) Analysis of zinc fingers optimized *via* phage display: evaluating the utility of a recognition code. *J. Mol. Biol.* **285**, 1917-1934.
- Wrighton, N.C., Farrell, F.X., Chang, R., Kashyap, A.K., Barbone, F.P., Mulcahy, L.S., *et al.* (1996) Small peptides as potent mimetics of the protein hormone erythropoietin. *Science* **273**, 458–463.
- Wu, X., Jörnvall, H., Berndt, K.D. and Oppermann, U. (2004) Codon optimization reveals critical factors for high-level expression of two rare codon genes in *Escherichia coli*: RNA stability and secondary structure but not tRNA abundance. *Biochem. Biophys. Res. Commun.* **313**, 89-96.
- Yang, J., Lee, G.F., Riederer, M.A. and Kelley, R.F. (2002) Enhancing the anticoagulant potency of soluble tissue factor mutants by increasing their affinity to factor VIIa. *Thromb. Haemost.* **87**, 450–458.
- Yokoyama, S. (2003) Protein expression systems for structural genomics and proteomics. *Curr. Opin. Chem. Biol.* **7**, 39-43.
- Yoon, S.K., Kang, W.K. and Park, T.H. (1996) Regulation of *trp* promoter for production of bovine somatotropin in recombinant *Escherichia coli* fed-batch fermentation. *J. Ferment. Bioeng.* **81**, 153-157.
- Yura, T., Nagai, H. and Mori, H. (1993) Regulation of the heat-shock response in bacteria. *Annu. Rev. Microbiol.* **47**, 321-350.

Zhang, D., Cupp, M.S. and Cupp, E.W. (2002) Thrombostasin: purification, molecular cloning and expression of a novel anti-thrombin protein from horn fly saliva. *Insect Biochem. Mol. Biol.* **32**, 321-330.

Zheng, X., Chung, D., Takayama, T.K., Majerus, E.M., Sadler, J.E. and Fujikawa, K. (2001) Structure of von Willebrand factor-cleaving protease (ADAMTS 13), a metalloprotease involved in thrombotic thrombocytopenic purpura. *J. Biol. Chem.* **276**, 41059-41063.

Zhu, K., Bowman, A.S., Brigham, D.L., Essenberg, R.C., Dillwith, J.W. and Sauer, J.R. (1997) Isolation and characterization of Americanin, a specific inhibitor of thrombin, from the salivary glands of the lone star tick *Amblyomma americanum*. *Exp. Parasitology* **87**, 30-38.

Zhu, Y.Y., Machleder, E.M., Chenchik, A., Li, R. and Siebert, P.D. (2001) Reverse transcriptase template switching: a SMART approach for full-length cDNA library construction. *Biotechniques* **30**, 892–897.

Zwick, M.B., Shen, J. and Scott, J.K. (1998) Phage-displayed peptide libraries. *Curr. Opin. Biotechnol.* **9**, 427-436.