Expression Studies of an Antibacterial Protein in the Shore Crab *Carcinus maenas*

Virginia Brockton BSc. (Hons.)
Gatty Marine Laboratory
University of St. Andrews

Submitted for the Degree of Doctor of Philosophy
September 2004

Declarations

I. I, Virginia Brockton, hereby certify that this thesis, which is approximately 63,000 words in length, has been written by me, that it is the record of work carried out by me and that it has not been submitted in any previous application for a higher degree.

Candidates Signature: .....Date: 10/3/05

II. I was admitted as a research student in October, 2000 and as a candidate for the degree of Doctor of Philosophy in 2001; the higher study for which this is a record was carried out in the University of St. Andrews between 2000 and 2004.

Candidates Signature: Date: 10/3/05

III. I hereby certify that the candidate has fulfilled the conditions of the Resolution and Regulations appropriate for the degree of Doctor of Philosophy in the University of St. Andrews and that the candidate is qualified to submit this thesis in application for that degree.

Supervisors Signature:. Date: 10 Μαρτιος

In submitting this thesis to the University of St. Andrews, I understand that I am giving permission for it to be made available for use in accordance with the regulations of the University Library for the time being in force, subject to any copyright vested in the work not being affected thereby. I also understand that the title and abstract will be published, and that a copy of the work may be made and supplied to any bona fide library or research worker.

Candidates Signature: Date: 10/3/05

Acknowledgements

I will always be grateful to Dr. Valerie J. Smith for giving me the opportunity to conduct this research over the past four years and for her unerring support and enthusiasm throughout. She has given me the opportunity to enter a whole new world, learn new skills and has supported me though many challenges. I could not, however, have made the most of these opportunities without the encouragement and unrestricted assistance of several former members of the Comparative Immunology Group, including Dr June R. S. Chisholm, Dr. John A. Hammond, Dr. Ralph Bickerdike and Dr. Jorge Fernandez. I would particularly like to thank Dr. Chris Hauton for always being available to answer a "quick question" and his patience in explaining statistical details. Without his support and help, the last few months would have been very difficult. I would also like to acknowledge all the help given to me by Dr. G. D. Kemp with protein separation techniques and his support as a mentor.

Most of all I would like to express my deepest gratitude to Dr. Nigel T. Brockton who has continuously provided moral support, encouragement, constructive criticism, discussion, and endless patience. This time we have gained everything and lost nothing.

Finally, without the infinite encouragement, love and the choices made available to me by my parents and grandparents, I would not have been in the position to start this work never mind finish it.

This project was funded by the Natural Environment Research Council, grant number NER/S/A/2000/03633.

Abstract

Aquaculture of Crustacea and the revenues generated, can be devastated by disease. Antimicrobial proteins (AMPs) are central to the immunological defence mechanism of these organisms. Understanding the diversity, expression and regulation of AMPs may improve culture strategies, reduce the burden of disease and potentially provide biomarkers of "health".

"Carcinin" is a constitutively expressed AMP, from the shore crab, *Carcinus maenas*; its full transcript and sequence variants were determined. Four putative isoforms of this defensin-like moiety were identified, all of which exhibited a conserved signal sequence. The C-terminus has a unique cysteine array containing up to 6 possible disulphide bonds. This pattern is conserved in AMP transcripts of commercially important crustaceans.

The Carcinin protein was recombinantly expressed, using the "full" and putative "active" transcript sequences. This study is the first to express a crustacean AMP using a bacterial fusion system; the yield was 2-3 $\mu g\ ml^{-1}$ of culture. However, the "active" product was prone to autoproteolytic degradation. This was not observed for the full size protein owing to the presence of the signal sequence.

The present study was also the first to quantify the expression of a crustacean AMP transcript using real-time PCR. Expression levels were investigated at four temperatures and in response to the injection of bacterial antigen. Expression of the Carcinin transcript was significantly up-regulated by 2-4 fold at 5 °C and 20 °C, compared to 10 °C control samples. However, the injection of bacterial antigen did not elicit any significant change in expression levels 0.5-48 h post injection.

Therefore, this thesis concludes that the Carcinin protein probably has a high turnover *in vivo* and the "full" coding sequence could be easily expressed and used in antibody production. In addition, it is concluded that Carcinin is inappropriate as a biomarker of immunological status and may have alternative biological functions.

Contents

List of Figures

List of Tables

List of Abbreviations

ABI – Applied Biosystems International

ACAR – "active" Carcinin sequence

AMP – antimicrobial protein (s)

ANOVA – analysis of variance

AP – alkaline phosphotase

APS – ammonium persulphate

ARE – AU rich element

BCIP - 5 bromo-4-chloro-3-indolyl phosphate toluidinum salt in dimethylformamide

BMLSS – British Marine Life Survey Society

bp – base pair

BSA – bovine serum albumin

CCD - charge-coupled device

CD - circular dichroism

cDNA – complimentary deoxyribonucleic acid

cfu – colony forming units

cm - centimetres

CSαβ - cysteine stabilised α β fold

Ct – threshold value

dATP – adenosine triphosphates

DEPC – diethylpyrocarbonate

DIG - digoxygenin

DMSO - dimethylsulphoxide

DNA - deoxyribonucleic acid

dNTP – dinucleotide triphosphates

DTT - 1, 4 dithiothreitol

dUTP – uridine triphosphates

E – efficiency value

EBI – European Bioinformatics Institute

EDTA – ethylenediaminetetra-acetic acid

EMBLE – European Molecular Biology Laboratory

EST (s) – expressed sequence tags

FAO – Fisheries and Agriculture Organisation

FCAR – "full" Carcinin sequence

fmol – femto mole

FPLC – fast protein liquid chromatography

FRET – frequency resonance transfer

FWD - forward

g - gram

gDNA – genomic deoxyribonucleic acid

GLP – good laboratory practice

GRAS – Generally Regarded As Safe

GS4B – glutathione Sepharose 4B

GST - glutathione S-transferase

h - hour

His6 – ployhistidine

HLS –haemocyte lysate supernatant

HNP – human neutrophil peptide

HPLC – high pressure liquid chromatography

HTP – high through put

IEC – ion exchange chromatography

IMAC- immobilised metal ion affinity chromatography

IMPACT - Intein-Mediated Purification with an Affinity Chitin-binding Tag

IPTG - isopropyl β-D-thiogalactoside

IUPAC – International Union of Pure and Applied Chemistry

kDa – kilo Dalton

l (m.μ) – litre (milli, micro)

LB – Luria-Bertani

LBA - Luria-Bertani with ampicillin

LBAG - Luria-Bertani with ampicillin and glucose

LPS – lipopolysaccharide

m – meter

mM – milli molar

μM – micro molar

M – molar

MAC – marine anticoagulant

MAFF – Ministry of Agriculture Fisheries and Food

MCR – multiple cloning region

MGD1 – *Mytilus galloprovincialis* defensin 1

μ - micro

μmol – micro mole

min - minute

M-MLV – moloney murine leukaemia virus

MOPS - 1 X 3-[N-Morpholino] propanesulphonic acid

mRNA – messenger RNA

n– nano

nmol – nano mole

NBT - nitroblue tetraziolium salt in dimethylformamide

NCBI - National Center for Biotechnology Information

NCIMB - National Collections of Industrial Food and Marine Bacteria

NMR – nuclear magnetic resonance

NTC – no template control

OD – optical density

ORF – open reading frame

PBS – phosphate buffered saline

PCR- polymerase chain reaction

Pen-3 – penaeidin 3

PIR - Protein Information Resource

pmol – pico mole

PMSF - phenyl methylsulphonyl fluoride

PO - phenoloxidase

proPO– pro phenoloxidase

PTM – post-translational modification

R – correlation coefficient

RACE – rapid amplification of cDNA ends

RDA – radial diffusion assay

REV - reverse

RNA - ribonucleic acid

RPHPLC – reverse phase high pressure liquid chromatography

rpm – revolutions per minute

rRNA - ribosomal RNA

RT –room temperature

SD – sequence detection

SDS – sodium dodecyl sulphate

SDS PAGE - sodium dodecyl sulphate polyacrylamide gel electrophoresis

SNP – single nucleotide polymorphism

St.Dev – standard deviation

T – time

TBE – Tris borate

TEMED - N,N,N,',N',tetramethylethylenediamine

TFA – trifluoroacetic acid

THC (s) – total haemocyte count (s)

TSS - transfer and storage solution

US – United States

UTR – untranslated region

UV – ultra violet

V - volts

v/v – volume for volume

w/v – weight for volume

WAP – whey acidic protein

X-gal - 5-bromo-4-chloro-3-indolyl-β-D-glactopyranoside

CHAPTER 1

General Introduction

1.1 Crustacea and Aquaculture

Invertebrates comprise around ninety five percent of all species and the majority of these are arthropods (Barnes, 1986). The Crustacea represent more than 42,000 arthropod species, most of which are essentially marine in habitat. Of these species, the majority belong to the class of Malacostraca. Within this class, the decapods are the largest order and include crabs, shrimps, crayfish and lobsters (Barnes, 1986; Barnes *et al.*, 1993).

Several decapods are familiar as luxury food products and in the United Kingdom (UK), the value of the wild caught fishery in 1998 was estimated at £160.7 million (MAFF, 1998), with crabs in particular, representing a revenue of around £46.8 million per annum (FAO, 1998). As demand has continued to rise for these high value species over the last 15 years (Subasinghe *et al.*, 1998; FAO, 2002), their aquaculture has also increased considerably, although not quite as rapidly as the culture of other aquatic groups (Figure 1) (FAO, 2002).

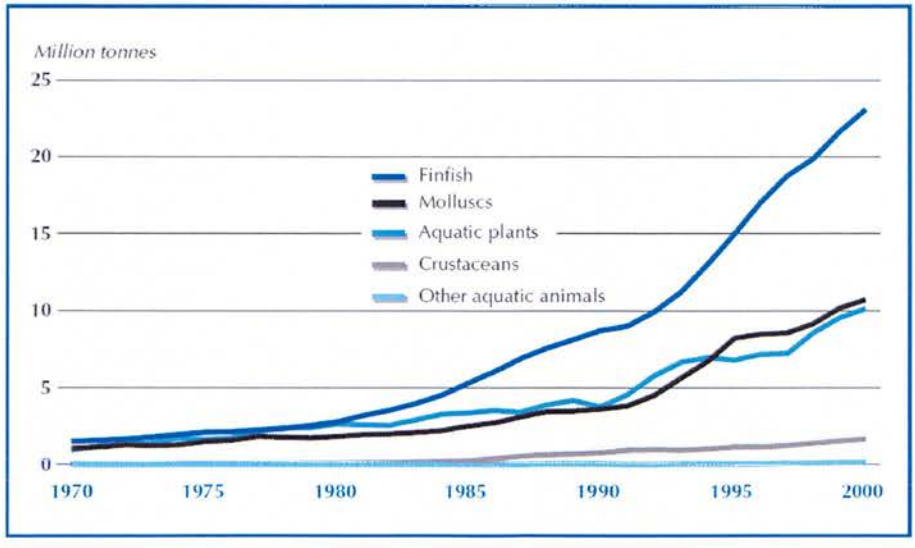


Figure 1: Trend of world aquaculture production by major species (image reproduced from FAO, 2002)

Crustacean aquaculture produces over 1.7 million tonnes worldwide (FAO, 2002) at an estimated value of ~$ 9.4 billion dollars (US) (Figure 2).

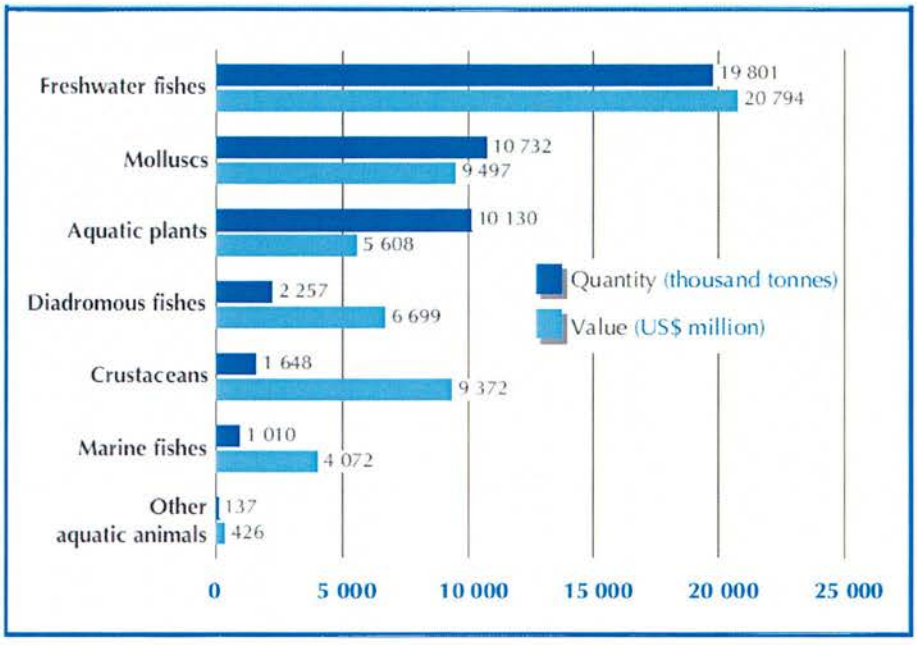


Figure 2: Aquaculture production of major species in groups (image reproduced from FAO, 2002)

Within the decapods, marine shrimp are the most important product (~1.1 million tonnes) and are farmed mainly in the Far East and South America (The Times, 1989; FAO, 2002). However, crabs are also considered a high value product, with a world production of ~140,000 tonnes in 2000, valued at ~$ 0.6 billion (FAO, 2002).

It is widely recognised that the main threat to successful aquaculture of these species is disease (Bachère *et al.*, 1995; Mialhe *et al.*, 1995; Roch, 1999; Aguirre and Ascencio, 2000; Bachère, 2000; Guzman *et al.*, 2000). As crustacean aquaculture is expected to increase in the coming years (FAO, 2002), research into disease prevention, diagnosis (Johnson, 1983; Mialhe *et al.*, 1995) as well as the development of treatments for known conditions has intensified (Rodriguez and Le Moullac, 2000). A greater understanding of their immune systems is required to combat and prevent the devastating effects on farmed stocks (Bachère *et al.*, 1995; Bachère, 1998; Roch, 1999;

Bachère *et al.*, 2004) and a considerable proportion of aquaculture research is focussed in this area. There is also a requirement for "markers" to assess the "health" of crustacean species; possible candidates include the antimicrobial proteins/peptides (Rodriguez and Le Moullac, 2000; Smith *et al.*, 2001; Bachère *et al.*, 2004). Understanding how these are affected by external factors is crucial to their validity. Also, the development of novel antibiotics based on these molecules, either by production of synthetic analogues, engineering transgenic disease-resistant animals (Mialhe *et al.*, 1995; Hancock and Lehrer, 1998) or enhancement of environmental and physiological conditions, may contribute to future disease prevention.

1.2 *Carcinus maenas* as a Model Organism

It is often not practical to investigate disease and immune capabilities in commercially or ecologically important species (Bachère *et al.*, 2004), due to restrictive culture conditions or the risk of introduction to the environment of non-indigenous species and diseases; therefore a suitable model organism is often used (Smith, 1991). In the present study, *Carcinus maenas,* the European Shore crab or Green crab was selected as the model organism (Figure 3).

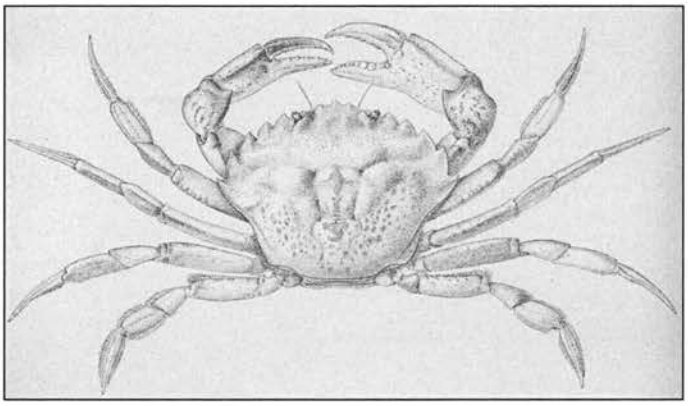


Figure 3: Drawing of the dorsal view of the European Shore Crab: *Carcinus maenas* (drawing by J.H Emerton reprinted from the National Oceanographic and Atmospheric Administration library at www.photolib.noaa.gov/historic/nmfs/figb0537.htm)

This species belongs to the infraorder of the Brachyura (Barnes, 1986; Barnes *et al.*, 1993) (Appendix 1) and although it is not a commercially important species in the UK, it has been extensively used for research since the 1940's (Williamson, 1940). This species is accessible, easily identified, sexed, measured and tolerates artificial conditions well (Crothers, 1966), making it an ideal study organism. It is known to inhabit a wide range of niches from the upper intertidal to the deep inshore waters of the coastal marine environment (Yonge, 1990; Mathieson and Berry, 1997). *C. maenas* also has a wide geographical distribution and can be found around the continental

shelves of Europe (Crothers, 1966; Bulnheim and Bahns, 1996; Mathieson and Berry, 1997) and the USA, where it is also affecting the balance of local biodiversity.

Adults usually have a carapace width of between 50-100 mm (Figure 4), a maximum weight of around 150 g (Crothers, 1966; Mathieson and Berry, 1997; British Marine Life Study Society, 2000), and can be found in several colour forms (Aagaard, 1996; Reid *et al.*, 1997; Styrishave *et al.*, 1999). This species grows by ecdysis (moulting) (Berrill, 1982), which continues throughout the life cycle (Abello *et al.*, 1997), forming a new soft exoskeleton, which later hardens.

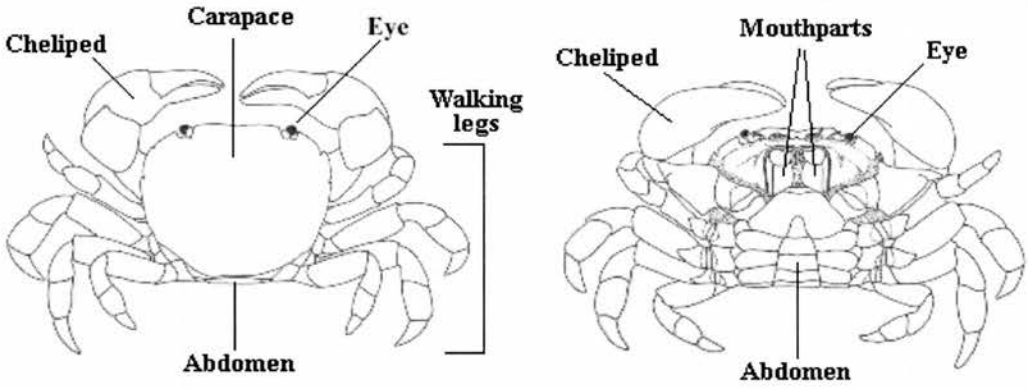


Figure 4: Dorsal and ventral body plan of *Carcinus maenas* (BMLSS, 2000)

C. maenas is particularly tolerant of changes in its environment and can often be found living in harsh conditions. This species thrives in a wide range of salinities (12-35 parts per thousand) (Mathieson and Berry, 1997), depths (0-50 m) and temperatures (3 – 25 °C) (Crothers, 1966). It is an epibenthic scavenger (Williamson, 1940) and its main diet across similar latitudes, normally consisting of bivalves, such as *Mytilus edulis* and *Mya arenaria* (Elner, 1981). In common with other crustaceans, *C. maenas* has an open vascular system and the "blood" or haemolymph consists of haemocytes and plasma (Crothers, 1966; Bachau, 1981). Its ability to inhabit a diverse range of habitats and tolerate a variety of challenging conditions makes it of considerable interest as a model

organism in immunological studies of aquaculture. An understanding of the factors affecting the immunological capability of a robust organism can provide insight into an "idealised" defence mechanism and provide information on the gross affecters of immunity in more sensitive organisms. Many aspects of *C. maenas* biology have been well documented, including behavioural, biochemical and cellular responses to environmental or physiological factors (Table 1). However, several factors may work in concert to bring about changes in the biology of these animals (Morris and Airriess, 1998). Based on evidence of the modulation of biochemical responses in this organism (Table 1), it is appropriate to explore whether immunological responses are similarly affected by external and internal factors.

Factor	Effect	Reference
Temperature	• Increased heart rate with increased temperature	(Aagaard, 1996; Styrishave *et al.*, 1999)
	• Increased locomotion with increased temperature	(Styrishave *et al.*, 1999)
	• Increased plasma membrane fluidity and lower cholesterol to phospholipid ratio, associated with decreased temperature	(Cuculescu *et al.*, 1995)
	• Affects development through feeding rate	(Breteler, 1975a, b; Dawirs, 1985; Dawirs and Dietrich, 1986)
	• Affects haemolymph proteins	(Terwilliger and Dumler, 2001)
Salinity	• Different colour crabs have different adaptability to salinity	(Reid *et al.*, 1997)
	• Affects larval development	(Anger *et al.*, 1998)
	• Affects biochemical makeup	(Torres *et al.*, 2002)
Tide	• Increased heart rate at high tide	(Styrishave *et al.*, 1999)
	• Circatidal moulting	(Abello *et al.*, 1997)
	• Regulates larval release	(Zeng and Naylor, 1997)
Contaminants	• Affects heart rate and osmoregulatory ability	(Bamber and Depledge, 1997a, b; Lundebye and Depledge, 1998)
	• Disrupts circulation and respiration	(Depledge, 1984)
Anoxia	• Reduces metabolism	(Hill *et al.*, 1991)
Infection / parasitism	• Reduces overall immunocompetence	(Hauton *et al.*, 1997) (Vivares and Cuq, 1981) (Henke, 1985)
	• Affects metabolism	
	• Affects biochemical makeup	
Age	• Heart rate higher in juvenile crab than adults	(Aagaard, 1996)
	• Influences reproductive success	(Dacoz, 1994)
Diet	• Affects biochemical makeup	(Busselen, 1970; Houlihan *et al.*, 1990; Harms *et al.*, 1994; Terwilliger and Dumler, 2001)
	• Affects larval growth	(Breteler, 1975a; Dawirs, 1984)
Illumination	• Affects larval release	(Zeng and Naylor, 1997)
Season	• Affects biochemical makeup	(Heath and Barnes, 1970)
Moult	• Affects biochemical makeup	(Sewell, 1955; Busselen, 1970; Heath and Barnes, 1970; El Haj and Houlihan, 1987)
Exposure	• Affects biochemical makeup	(Johnson and Uglow, 1985)

Table 1:Examples of environmental and physiological factors affecting biology in *Carcinus maenas*

1.2.1 Host Defences in C. maenas

C. maenas is an exceptionally successful coloniser of a variety of ecological niches despite exposure to temperature variation and a continuous assault by diverse bacterial communities. In the marine environment, bacterial numbers may vary from 10^3 to over 10^6 ml^{-1} in oceanic waters, with numbers reaching over 10^7 ml^{-1} in rich organic coastal waters (Li *et al.*, 1993; Li and Dickie, 1996). Bacterial numbers are also known to proliferate proportionally with increased temperature and decreased depth (Li and Dickie, 1996). Therefore, this species of crab must exhibit a particularly robust and "adaptable" immune defence mechanism (Millar and Ratcliffe, 1994) that provides constant protection and remains uncompromised by the changing environment. Overall, their immune defence mechanism is highly effective despite the fact that *C. maenas* does not posses a true adaptive response mechanism mediated through clonally derived T-lymphocyte sub-sets, as found in vertebrates. Some authors have argued, that crustaceans may have a quasi-adaptive system, independent of lymphocytes, but this idea has not yet received general acceptance (Teague and Friou, 1964; McKay and Jenkin, 1969; Arala-Chaves and Sequeira, 2000). Many comprehensive reviews of crustacean immunology have been published; these reviews describe both cellular and biochemical aspects of crustacean immune function (Sloan *et al.*, 1975; Bachau, 1981; McCumber and Clem, 1983; Cooper, 1985; Ratcliffe, 1985; Ratcliffe *et al.*, 1985; Ratcliffe, 1989; Smith, 1991; Smith and Chisholm, 1992; Söderhäll, 1992; Söderhäll and Cerenius, 1992; Söderhäll *et al.*, 1994; Battistella *et al.*, 1996; Söderhäll and Thornqvist, 1997; Smith *et al.*, 2001; Lee and Söderhäll, 2002).

In the Crustacea, as for most organisms, the initial defences against invading pathogens are the physiochemical barriers, for example, the shell or cuticle (Ratcliffe, 1985;

Lyndyard *et al.*, 2000). Internally, host defences reside mainly in the activity of the haemocytes (Smith and Ratcliffe, 1978; Ratcliffe and Rowley, 1979; Smith and Ratcliffe, 1980b, a, 1981; White and Ratcliffe, 1982; Smith and Ratcliffe, 1983; Smith and Söderhäll, 1983; Söderhäll and Smith, 1983; Smith *et al.*, 1984; White *et al.*, 1985; Smith and Söderhäll, 1986; Söderhäll *et al.*, 1986). These are involved in phagocytosis, agglutination, encapsulation, cell lysis and the secretion of bioactive compounds. Although several studies have been published which contradict each other on the number of cell types in *C. maenas* (Rabin, 1970; Johnston *et al.*, 1973; Williams and Lutz, 1975), the general consensus is that there are three types, the hyaline, semi-granular and granular cells (Rabin, 1970; Johnston *et al.*, 1973; Ghiretti-Magaldi *et al.*, 1977; Smith and Ratcliffe, 1978; Ratcliffe and Rowley, 1979; Bachau, 1981; Smith and Ratcliffe, 1981; Söderhäll and Smith, 1983; Johnson, 1987; Johansson *et al.*, 2000).

In *C. maenas* the hyaline cells are phagocytic (Smith and Ratcliffe, 1978, 1980b, a; Söderhäll and Smith, 1986; Söderhäll *et al.*, 1986) and this activity has been used as a measure of immunological capability (Smith and Ratcliffe, 1980b, a; Hauton *et al.*, 1997). These cells also exhibit respiratory burst characterised by the production of antioxidant enzymes (Bell and Smith, 1993, 1995). It has been shown that phagocytosis can be enhanced by β-glucans (Smith and Ratcliffe, 1983; Smith *et al.*, 1984) and other opsonins, released by degranulated granular cells (Smith and Söderhäll, 1983; Söderhäll *et al.*, 1986).

The granular cells are repositories of prophenoloxidase (proPO), which is activated by β 1-3 glucans or bacterial polysaccharides (Söderhäll *et al.*, 1990). Several reviews have been written on this subject (Söderhäll, 1982; Smith *et al.*, 1984; Smith and Söderhäll, 1986; Söderhäll and Smith, 1986; Söderhäll, 1992). Granulocytes are known to release

a variety of bioactive compounds, either by regulated exocytosis (as reviewed by Johansson and Söderhäll, 1985) or as a result of complete degranulation of the cell (as reviewed by Söderhäll and Smith, 1986). The bioactive compounds previously reported in decapods include antimicrobial proteins (Söderhäll and Smith, 1986; Chisholm and Smith, 1992; Schnapp *et al.*, 1996; Relf *et al.*, 1999), opsonins (Thornqvist *et al.*, 1994), agglutinins (Sritunyalucksana *et al.*, 1999), lytic enzymes and prophenoloxidase factors (Söderhäll, 1982; Smith and Söderhäll, 1983; Söderhäll and Smith, 1983, 1986; Aspan *et al.*, 1990; Söderhäll *et al.*, 1994; Smith, 1996). These biochemical components of host defence in crustaceans have also been thoroughly investigated, particularly encouraged in the wake of the increasing accessibility of molecular techniques (Söderhäll and Smith, 1986; Söderhäll *et al.*, 1990; Chisholm and Smith, 1992; Smith and Chisholm, 1992; Söderhäll, 1992; Chisholm, 1993; Söderhäll *et al.*, 1994; Smith and Chisholm, 2001; Bachère *et al.*, 2004).

The proteins that provide antimicrobial defences are fundamental to the host defence system and are of particular interest to the present study as possible biomarkers. These proteins and their transcripts are constitutively expressed in both adult and larval shrimp (Relf *et al.*, 1999; Destoumieux *et al.*, 2000; Munoz *et al.*, 2002; Munoz *et al.*, 2003). However, little information exists on the transcription and expression of antimicrobial protein (AMP) proteins in *C. maenas*. It is logical to speculate, that constitutive expression of these proteins and their transcripts must be energetically sustainable and may yet be shown to be modulated by a variety of environmental and physiological factors and these issues are further investigated in the present study.

Evidence already exists of the modulation of immunological function in crustaceans by environmental and physiological factors as reviewed by Le Moullac and Haffner (2000).

Of particular relevance, to the present study, are the reported effects of temperature and bacterial infection. Temperature appears to have a proportional relationship with immunological parameters such as total haemocyte count (THC), phenoloxidase activity and respiratory burst, (Dean and Vernberg, 1966; Truscott and White, 1990; Cheng and Chen, 1998; Vargas-Albores *et al.*, 1998; Cheng and Chen, 2000; Gomez-Jimenez *et al.*, 2000; Cheng *et al.*, 2002; Cheng *et al.*, 2003). Low temperatures reduce clotting times (Dean and Vernberg, 1966; Hamann, 1975), haemolymph protein titre (Peters and Long, 1973), and the morphology of the haemocytes (Ravindranath, 1975). A recent study of the relationship between viral pathogenicity and temperature concluded that higher temperatures reduced immune capability the thus increases the susceptibility of the animals to disease (Jiravanichpaisal *et al.*, 2004). To date, only a few studies have suggested a relationship between antibacterial activity and temperature in crustaceans (Smith and Chisholm, 1992; Chisholm and Smith, 1994), although direct relationships have been reported in other invertebrates such as the bivalves (Mitta *et al.*, 2000a; Hernroth, 2003). This relationship is to be further investigated in *C. maenas* in the present study.

It is generally accepted that the entry of bacteria into the body cavity of a crustacean elicits both cellular and biochemical responses, following the pioneering work of Evans *et al.* (1968) and Stewart and Zwicker (1971). The effects of sub-lethal doses can be further exacerbated by environmental parameters (Stewart and Zwicker, 1972; Stewart and Arie, 1973). Bacteraemia may also sometimes prove fatal to the organism (Cornick and Stewart, 1968). Cellular responses in *C. maenas* are characterised by rapid immobilisation and the elimination of the bacteria by phagocytosis and sequestration into clumps (Smith and Ratcliffe, 1980a, 1981; White and Ratcliffe, 1982; Johnson, 1987). There is also a concomitant reduction in the number of haemocytes, within 30

min, which returns to normal after 24-48 h (Smith and Ratcliffe, 1980b; Martin *et al.*, 1993; Hauton *et al.*, 1997). These cellular responses do not appear to differentiate between non-pathogenic strains (Smith and Ratcliffe, 1980b), although differentiation has been reported between species (Smith and Ratcliffe, 1978; Johnson, 1981). This may be related to the presence of commensal bacteria as described in the blue crab, *Callinectes sapidus,* by Colwell *et al.* (1975). The activation of the cellular response mechanism leads to the activation of humoral responses via the rapid degranulation of the cells and the release of bioactive compounds, including antimicrobial proteins (Smith and Ratcliffe, 1981; Smith and Söderhäll, 1983; Söderhäll and Smith, 1986; Söderhäll, 1992; Kanost, 1999; Sritunyalucksana *et al.*, 1999; Destoumieux *et al.*, 2000). Some authors have claimed that bactericidal components can be "induced" (Evans *et al.*, 1968; Evans *et al.*, 1969; Weinheimer *et al.*, 1969; Stewart and Zwicker, 1971; Stewart and Zwicker, 1972; Mori and Stewart, 1978; Adams, 1991). However, the use of the term "induction", may be misleading as it has been used to describe both the release and activation of bactericidal components (Destoumieux *et al.*, 2000) as seen in the shrimp, *L.vannamei*, and *de novo* protein synthesis as seen in insects in the fat body (Boman, 1991, 1995; Engstrom, 1999; Boman, 2003). In the present study, "induction" refers to increases in relative transcript abundance of an antimicrobial gene and not *de novo* synthesis of proteins or activation of inactive precursor proteins.

Antimicrobial molecules are thought to be one of the central factors in the immunological repertoire of crustacean biology (Bachère *et al.*, 2004) but the factors regulating the synthesis of these molecules are relatively unknown. In terms of this approaches' application to aquaculture, Le Moullac and Haffner (2000) have already suggested that studies of gene expression may be essential to elucidate the effects of external parameters on immune responses in Crustacea. Molecular approaches, such as

gene expression analysis, have been increasingly used to examine all aspects of *C. maenas* biology (Whiteley *et al.*, 1992; Towle *et al.*, 1995; Towle, 1997; Towle *et al.*, 1997; Lundebye and Depledge, 1998; Kotlyar *et al.*, 2000; Rewitz *et al.*, 2003), including biochemical and physiological aspects of immunity (Thornqvist *et al.*, 1994).

The present study aims to investigate the expression of a transcript for a single AMP in *C. maenas,* under selected environmental and physiological conditions, to determine if antimicrobial gene expression is modulated by bacteria and/or temperature changes. The responses for *C. maenas* will be compared to those reported for the shrimp (Munoz *et al.*, 2002). Some authors have suggested that single parameters are unreliable indicators of these effects and a suite of immunological tools would provide a better understanding of such responses (Hebel *et al.*, 1997). With this in mind, total haemocyte counts, a traditional indicator of immunological response activation, were used to corroborate gene expression observations. Thus, it may be possible to identify, or at least eliminate regulatory factors affecting expression of these molecules in crustaceans. Additionally, the present study will facilitate the collection of preliminary data on the suitability of this antimicrobial gene as a biomarker of immune health.

1.3 Antimicrobial Proteins

The first antimicrobial proteins (AMPs) were identified in plants in the early 1940's (Stuart and Harris, 1942). Since then, many (>1700) isolated proteins and nucleotide transcripts, with putative functions, have been described in vertebrates, invertebrates and plants (see reviews by Zasloff, 1992; Hancock and Scott, 2000; Brahmachary *et al.*, 2004); more than half of these have been described for arthropods (Dimarcq *et al.*, 1998). Eukaryotic AMPs are small (<100 residues) (Boman, 1995), ribosomally translated proteins, transcribed from single genes (Hancock and Chapple, 1999). Due to their small size, they can be produced by unspecialised cells (Lehrer and Ganz, 1999) and diffuse rapidly to the point of infection (i.e. direct effectors) (Boman, 1991). They are less susceptible to protease activity (Borregaard, 2001) and their considerable diversity, both within (Cuthbertson *et al.*, 2002; Nicolas *et al.*, 2004) and between species (Maxwell *et al.*, 2003), is thought to reflect adaptations to unique microbial environments (Zasloff, 2002). Most of our understanding of these molecules emanates from studies in insects (Michaut *et al.*, 1996a; Cho *et al.*, 1997; Hetru *et al.*, 1998; Kim *et al.*, 1998; Levashina *et al.*, 1998; Engstrom, 1999; Konno *et al.*, 2001; Luna *et al.*, 2002), although vertebrate studies are rapidly accumulating (Nizet *et al.*, 2001; Cunliffe, 2003; Devine, 2003; Philpott, 2003; Tollin *et al.*, 2003).

These AMPs generally have a net positive charge (cationic) and can function without specificity or memory (Boman, 1995). One mode of action of these molecules is a permeabilization of the bacterial membrane leading to leakage of its contents (White *et al.*, 1995; Andreu and Rivas, 1998; Dimarcq *et al.*, 1998; Hancock and Chapple, 1999; Tossi *et al.*, 2000; Zasloff, 2002; Ganz, 2003). One of the major ways in which this is brought about is based on charge interactions between the proteins and bacterial

membrane and is known as the Shai-Matsuzuki-Huang model (Shai, 1997; Epand and Vogel, 1999; Shai, 1999; Zasloff, 2002).

Research into antimicrobial proteins has also expanded due to ever-increasing problems associated with microbial resistance to traditional antibiotics in human medicine (Amabile-Cuevas *et al.*, 1995). This problem requires novel compounds to be developed (Elsbach, 1990; Hancock and Lehrer, 1998; Hancock and Scott, 2000; Levy, 2000) and several products based on natural antimicrobial proteins are already in clinical trials (Zasloff, 2002; Lucentini, 2003). Numerous reviews have been published, examining all aspects of AMPs in vertebrates and invertebrates (Zasloff, 1992; Andreu and Rivas, 1998; Barra *et al.*, 1998; Hetru *et al.*, 1998; Epand and Vogel, 1999; Giacometti *et al.*, 1999; Hancock and Chapple, 1999; Hancock and Diamond, 2000; Hancock and Scott, 2000; Levy, 2000; Tossi *et al.*, 2000; Zasloff, 2002; Boman, 2003; Brogden *et al.*, 2003; Koczulla and Bals, 2003; Powers and Hancock, 2003; Nicolas *et al.*, 2004). Several comprehensive antimicrobial protein databases also exist and catalogue functional, structural and evolutionary information, as well as providing links to relevant literature and bioinformatic tools (e.g. http://www.bbcm.univ.trieste.it/~tossi/amsdb.html, http://research.i2r.a-star.edu.sg/Templar/DB/ANTIMIC/), http://aps.unmc.edu/AP/main.php & http://oma.terkko.helsinki.fi:8080/~SAPD/ (Brahmachary *et al.*, 2004; Galperin, 2004; Wang and Wang, 2004). Due to the extensive information available on these molecules, the present overview focuses mainly on antimicrobial peptides identified in invertebrates and particularly from arthropods.

Studies on antimicrobial defences in invertebrates originate from the 1970's with Faye and Boman's work on the silk moth (*Hyalophora cecropia*) (Faye *et al.*, 1975),

although earlier, Evans *et al.* (1968) and Acton *et al.* (1969) reported "bacteriocidal" activity in the spiny lobster (*Panulirus argus*) American lobster (*Homarus americanus*) respectively. Invertebrate AMPs have since been extensively reviewed, with descriptions of signalling pathways (Levashina *et al.*, 1998; Aderem and Ulevitch, 2000; Naitza and Ligoxygakis, 2004), possible therapeutic uses (Lehrer and Ganz, 1999; Vizioli and Salzet, 2002; Wilson, 2002) and their links to vertebrate immunity (Dimarcq *et al.*, 1998; Salzet, 2001; Werling and Jungi, 2003). Fundamental differences between known AMPs, at structural and functional levels, have led some reviewers to focus on the AMPs found in particular species or taxonomic groups (Bulet *et al.*, 1999; Mitta *et al.*, 2000b; Smith and Chisholm, 2001; Bachère *et al.*, 2004; Naitza and Ligoxygakis, 2004).

Several AMPs have been purified from marine arthropods in recent years. In addition, several expressed sequence tags (EST), possibly coding for homologous proteins in other species, have been assigned putative antimicrobial functions based on sequence identity to known AMPs (Gross *et al.*, 2001; Bartlett *et al.*, 2002; Supungul *et al.*, 2002; Chen *et al.*, 2004; Stoss *et al.*, 2004; Vargas-Albores *et al.*, 2004). The ESTs from marine arthropods can be found both in the traditional EST databases at the National Center for Biotechnology Information (NCBI) (http://www.ncbi.nlm.nih.gov/), but also on the marine genomics database (http://www.marinegenomics.org/). Some of the first arthropod AMPs to be purified include the Tachyplesins (Nakamura *et al.*, 1988; Miyata *et al.*, 1989; Kawano *et al.*, 1990; Iwanaga, 2002) and the Tachycitins (Kawabata *et al.*, 1996) from the horseshoe crab (chelicerate), *Tachypleus tridentatus* (reviewed by Iwanaga *et al.*, 1994; Iwanaga, 2002). Of the crustaceans, the first of two proteins to be purified was a 6.5 kDa protein (Schnapp *et al.*, 1996) from *C. maenas* and the second from the same species, was an 11.5 kDa protein (Relf *et al.*, 1999) called Carcinin. An

AMP from the blue crab *Callinectes sapidus* called callinectin, was isolated in 1999 (Khoo *et al.*) and three penaeidins (Pen-1, Pen-2 and Pen-3) from the shrimp *Litopenaeus vannamei* were isolated in 1997 (Destoumieux *et al.*, 1997).

Many attempts have been made to classify AMPs, although the parameters defining each class are constantly changing to accommodate new proteins. Despite initial attempts to classify them on taxonomical basis, chemical and structural criteria (Boman, 1995) are now more commonly used (Andreu and Rivas, 1998; Dimarcq *et al.*, 1998). Some recent studies have suggested that classification should be based solely on activity, with two main groups emerging: proteins with activity against bacteria but not eukaryotic cells or those proteins active against both (Papo and Shai, 2004). However, the growing availability of high resolution tertiary structures (Brahmachary *et al.*, 2004; Wang and Wang, 2004) has ensured continued popularity of a mainly structural classification system, particularly as the inseparable relationship between structure and activity is well appreciated (Bulet *et al.*, 1999; Epand and Vogel, 1999; Powers and Hancock, 2003). Thus, a generalised classification based on four main structural patterns is recognised in both vertebrates (Hancock and Lehrer, 1998; Hancock and Diamond, 2000; Powers and Hancock, 2003) and invertebrates (Bulet *et al.*, 1999; Tossi *et al.*, 2000). These four classes are the β-sheet, α-helical, a combination of the β-sheets and α-helices and the extended/linear/loop proteins (Figure 5). The β-sheet class (c) includes all cysteine-rich proteins whose structures are folded and stabilised by several cysteine – cysteine (disulphide) bonds (White *et al.*, 1995), whereas α-helical proteins (a) are devoid of cysteines and readily form amphipathic structures (Bulet *et al.*, 1999; Tossi *et al.*, 2000). Structures which exhibit a combination of β-sheets and an α-helices (b) are known as cysteine stabilised αβ (CSαβ) fold structures and were first described in insects (Cornet *et al.*, 1995). The extended or loop proteins (d) are those

which do not appear to have any identifiable structural features and generally have an over-representation of a single protein residue (e.g. glycine or proline). Although partial sequence fragments of AMP have been isolated from *C. maenas* (Schnapp *et al.*, 1996), which may belong to this group (d), no tertiary structures have yet been elucidated. Until recently, the only known marine invertebrate AMPs structures were either β-sheet (Kawano *et al.*, 1990) or CSαβ folds (Yang *et al.*, 2000).

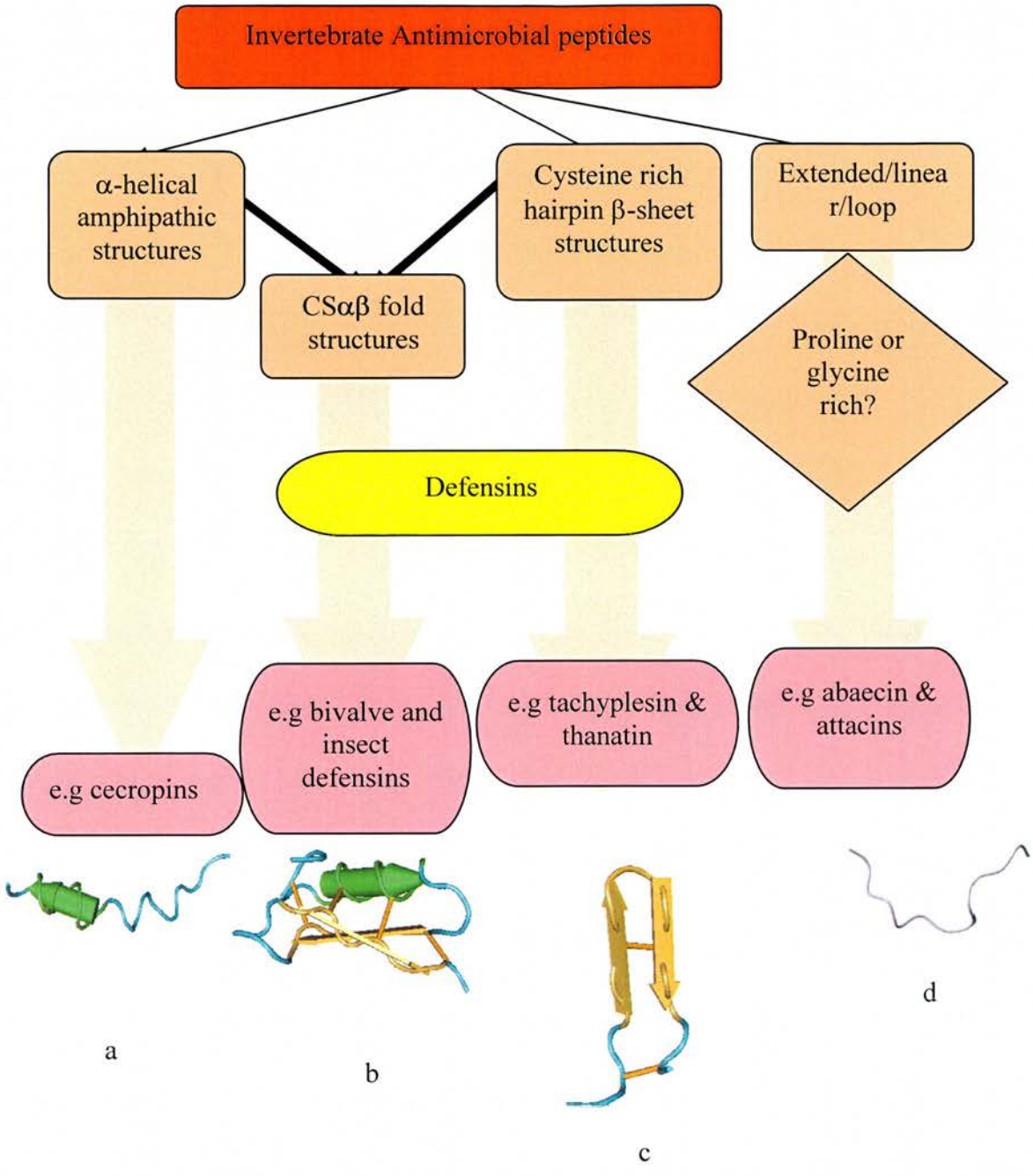


Figure 5: General classification of invertebrate antimicrobial proteins based on structure and residue data with examples created with Cn3D3 v. 4.1 of cecropin (a), MGD1 (b), tachyplesin (c) and a generalised extended structure (d) (Chen *et al.*, 2003; Powers and Hancock, 2003).

However, the publication of the unique structure of a recombinant antimicrobial peptide (Penaeidin 3) from the shrimp *L.vannamei* (Yang *et al.*, 2003), further supported earlier suggestions that it may belong to a novel group (Destoumieux *et al.*, 1997). This

structure combines an un-restricted proline domain with three disulphide bridges involved in the stabilising of an amphipathic α-helix and an extended segment but no β folds (Figure 6).

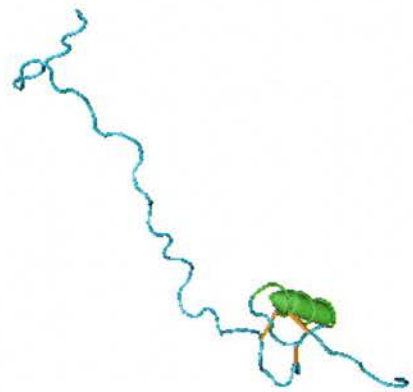


Figure 6: The tertiary structure of penaeidin 3 from the shrimp *L.vannamei* (Yang *et al.*, 2003).

Previously identified cysteine-stabilised structures, in insects, have been called "defensins" (Ganz and Lehrer, 1994; Hetru *et al.*, 1998) based on similarities of structure and efficacy (White *et al.*, 1995; Dimarcq *et al.*, 1998; Zhang and Kato, 2003) with mammalian defensins (see review by Raj and Dentino, 2002). The term "defensin" is now generally accepted to describe cysteine-rich antimicrobial proteins exhibiting either a CSαβ fold or a β-sheet arrangement. Based on this definition, the antimicrobial proteins isolated from diverse invertebrate species such as the mussel, *Mytilus galloprovincialis* (e.g. MGD1) (Charlet *et al.*, 1996; Hubert *et al.*, 1996; Mitta *et al.*, 2000a), the scorpion, *Androctonus australis* (Cociancich *et al.*, 1993; Ehret-Sabatier *et al.*, 1996), and the horse shoe crab, *Tachypleus tridentatus,* have also been called defensins. Although the penaeidins have been shown to have some disulphide bonds in their tertiary structure, they are not considered to be defensins as they are not overall cysteine-rich and do not have β-sheet folds. However, the 11.5 kDa antimicrobial protein isolated from *C. maenas* (Relf *et al.*, 1999) has been shown to have a high

cysteine representation throughout its sequence suggesting that it may form several disulphide bonds and be related to the defensin molecules; this assertion is unconfirmed. As part of the present study, further evidence for the relationship of Carcinin to defensin molecules will be explored.

Generally, defensins are synthesised as 90-100 residue preproproteins with a ~19 residue signal sequence at the amino (N)-terminal, an anionic propiece of around 45 residues (this may be absent in some proteins) and a mature protein of around 30 residues in length at the carboxy (C)-terminal end (Ganz, 2003). The transcription and subsequent translation of defensin genes results in the production of the inactive proteins which are stored in granular cells (Raj and Dentino, 2002; Boman, 2003; Ganz, 2003) where they are processed to active molecules upon stimulation. This has been observed for the penaeidins isolated from the shrimp (Destoumieux *et al.*, 2000), despite not being considered a defensin. Defensins from vertebrates and invertebrates have been suggested to have common ancestry dating back about 545 million years (Charlet *et al.*, 1996) and have been described as "ancestral effectors of innate immunity" (Lehrer *et al.*, 1991; Ganz and Lehrer, 1994; Yang *et al.*, 1999). This has prompted several reviews summarising their importance, classification, structure, mode of action and their potential in therapeutics (Ganz *et al.*, 1985; Ganz and Lehrer, 1994; White *et al.*, 1995; Dimarcq *et al.*, 1998; Yang *et al.*, 1999; Raj and Dentino, 2002; Dhaliwal *et al.*, 2003; Ganz, 2003; Philpott, 2003).

In the present study the full sequence of an AMP from *C. maenas* will be determined to help identify aspects of its gene structure, possible cleavage sites and signal sequences; these factors will help further determine its classification within the AMP.

1.3.1 Expression of Crustacean Antimicrobial Proteins

Antimicrobial proteins appear to be fundamental to the immunological repertoire of Crustacea. Of the crustacean AMPs, the penaeidins are some of the most thoroughly researched and the majority of transcriptional and translational studies have been conducted in the penaeids. The penaeidin proteins and transcripts have been shown to be produced in the granular haemocytes and demonstrate chitin binding activity (Destoumieux *et al.*, 2000). Penaeidin expression in different tissues and larval stages has also only recently been documented (Munoz *et al.*, 2002; Munoz *et al.*, 2003) as well as expression and distribution in response to microbial challenge. However, the mechanisms of recognition and subsequent activation of this immune response and the mode of action of crustacean AMPs remain unknown (Bachère *et al.*, 2004). Destoumieux *et al.* (2000) have suggested that microbial challenge temporarily down-regulates penaeidin transcript expression 3 h after stimulation. However, a subsequent study reported that microbial challenge does not affect the expression of penaeidin mRNA transcripts (Munoz *et al.*, 2002). Both these results were obtained using northern blotting; this is a rather insensitive technique, which can, at best, only differentiate ten-fold differences in transcription levels. The present study in *C. maenas* investigates the effect of external stimuli on transcript expression levels of Carcinin using quantitative real-time polymerase chain reaction (PCR), a technique sensitive enough to detect two-fold change changes in expression.

Since the purification of the penaeidin proteins, several penaeidin EST sequences, isolated from *Litopenaeus vannamei* and *Litopenaeus setiferus* (Gross *et al.*, 2001), have been submitted to GenBank. These sequences have since been divided into four main classes (Pen 1-4) (Cuthbertson *et al.*, 2002; Cuthbertson *et al.*, 2004). Overall,

antibacterial sequences were reported to represent a considerable proportion (~17-21 %) of the ESTs isolated. Further analysis of the EST libraries from both species indicated that, in addition to the penaeidin ESTs, a considerable proportion (4-9 %) of the transcripts (crustins) had some identity with Carcinin (Gross *et al.*, 2001; Bartlett *et al.*, 2002; Vargas-Albores *et al.*, 2004). Further EST sequences, with identity to both the crustin and Carcinin transcripts, have since been submitted to GenBank from a variety of other crustacean species including the shrimps, *Penaeus monodon* (BIO18072-74) (Supungul *et al.*, 2002) and *Marsupenaeus japonicus* (AB121740-44), the lobsters *Homarus americanus* (CN853187) and *Homarus gammarus* (AJ786653) as well as the crayfish *Panulirus leniusculus* (AF522504) and *P. argus* (AY340636). Therefore, it would appear that Carcinin-like transcripts may be common factors in the immune responses of most economically important Crustacea and expression studies in *C. maenas* may be directly relevant to all these species.

1.3.2 Antimicrobial Proteins from C. maenas

The ability of *C. maenas* to remove bacteria from circulation was first described in 1978 by Smith and Ratcliffe (1978). It was attributed, in the main, to cellular responses (Smith and Ratcliffe, 1980b, a, 1981; White *et al.*, 1985). Since then, antibacterial activity has been identified as a result of the biochemical components of the haemocytes and not only cellular responses; more specifically, these components are found only in the granular cells (Chisholm and Smith, 1992; Chisholm, 1993). This antibacterial activity was observed to be effective against both Gram-negative and Gram-positive marine bacteria, to be freeze and heat stable (~100 °C) and active at low concentrations (2 $\mu g\ ml^{-1}$) (Chisholm and Smith, 1992). Chisholm and Smith (1992) were the first to suggest that the factor or factors responsible for the antibacterial activity in *C. maenas*

may be related to the defensin molecules previously described in insects (Boman, 1991). In a later study, the vigour of the antibacterial activity in *C. maenas* was observed to vary throughout the year; it was unclear if this was a direct result of changes in temperature, total haemocytes count (THC) or other seasonal parameters (Chisholm and Smith, 1995). Qualitative screening methods were developed for the detection of antimicrobial activity in blood cell extracts and tissues *in vitro* (Chisholm and Smith, 1992). These were used to identify antibacterial activity in a number of species although the strength of activity was variable (Chisholm, 1993; Chisholm and Smith, 1995).

Subsequent investigation of antimicrobial activity in *C. maenas* led to the identification of several antimicrobial proteins with molecular masses of approximately >70 kDa, ~45 kDa, <25 kDa, ~11 kDa, ~14 kDa and ~6.5 kDa; these were active against both Gram-positive and Gram-negative bacteria (Schnapp *et al.*, 1996). Only two of these proteins have since been isolated, characterised and partially sequenced from *C. maenas* (Schnapp, 1996; Relf *et al.*, 1999).

The first of these was the ~6.5 kDa (Swiss Prot Accession No/ P82964) protein, which was purified to homogeneity and N-terminally sequenced (Schnapp *et al.*, 1996). This was a proline rich, 30-residue fragment, exhibiting considerable sequence identity to the bactenecin 7 precursor (P19661) sequence from bovine neutrophils. This partial sequence was later observed to have some identity to proline rich AMPs (penaeidins) isolated from the shrimp (Destoumieux *et al.*, 1997) although the full sequence was never obtained. The 6.5 kDa protein was not consistently evident in the haemolymph of *Carcinus maenas,* throughout all seasons, (Relf *et al.*, 1999); the significance of this

was unknown. Due to its inconsistent appearance in antibacterial fractions, research on antimicrobial proteins in *C. maenas* subsequently focussed on the larger proteins.

In 1999, an active "11.5 kDa" antibacterial protein was purified from pooled extracts of granular haemocytes. It was purified using both ion exchange chromatography (IEC) and reverse phase high performance liquid chromatography (RP-HPLC) (Relf *et al.*, 1999). The isolated protein was subsequently characterised (Relf *et al.*, 1999) (NCBI : AJ237947, trEMBL : Q9Y099) and named "Carcinin" (Appendix 2). This purified protein was reported to have a molecular mass of 11.534 kDa and the characterisation studies indicated that it was cationic, hydrophobic, cysteine-rich and only active against Gram-positive marine bacteria (including both *Aerococcus viridans,* the lobster pathogen, and *Planococcus citreus*). No start or stop codons were identified in the sequence thereby suggesting that the purified product was only a fragment of the full protein. It was speculated that Carcinin was probably produced as a larger precursor form, as had previously been described for most other antimicrobial proteins (Zasloff, 1992; Ganz and Lehrer, 1994; Hancock and Lehrer, 1998; Epand and Vogel, 1999; Hancock and Diamond, 2000; Hancock and Scott, 2000; Zasloff, 2002). However, this remains to be proven. At the time of its purification, Carcinin did not exhibit considerable sequence identity with any known proteins or translated transcripts except a weak identity (~30 %) with human anti-leukocyte proteinase sequence (Relf *et al.*, 1999), based on the identification of a whey acidic protein (WAP) domain. This domain, first described by Ranganathan *et al.* (1999), has since become recognised as central to the function of molecules with serine protease activity. It is found in many defensins and is thought to be central the antibacterial activity to these molecules (Ganz and Lehrer, 1994; Ganz, 2003; Chen *et al.*, 2004).

Although preliminary information regarding the protein sequence and the efficacy of Carcinin has been reported (Relf *et al.*, 1999), nothing is known of the gene, its tertiary structure, expression, posttranscriptional and post-translational modifications, mode of action or even the factors modulating increased production, storage and release of the active protein. By combining a bioinformatic approach with the development of transcript and protein tools, some of these issues will be investigated in the present study. By examining transcriptional responses of Carcinin to selected stimuli, at both transcriptional and translational levels, its application as a "biomarker" of immunological status will be assessed.

1.4 Specific Aims

- To determine the full transcript sequence of Carcinin in *C. maenas*
- Analyse this sequence using web based bioinformatic tools to further classify the protein, identify its gene structure, functional domains, and predict structural features.
- Identify variants of the Carcinin transcript and describe their distribution within a local population of *C. maenas*.
- Recombinantly express both the full coding sequence and the putative active fragment of the Carcinin protein, for use as antigen in antibody production to facilitate investigation into its protein expression.
- Design and optimise sensitive transcript probes to quantify changes in Carcinin transcript expression with temperature and bacterial antigen, normalised by endogenous control genes.
- Assess Carcinins' suitability as a biomarker of health in *C. maenas*.

CHAPTER 2

Carcinin Sequence Analysis

2.1 Introduction

The sequence information currently available on the Carcinin protein, has been derived from purified protein fragments and inferences from partial transcript sequencing. Physiochemical and efficacy data for the native protein have also been reported (Relf *et al.*, 1999). However, little is known of the full sequence, its variability, expression, post-translational modifications (PTM), structure or classification.

To begin investigating these issues, the full transcript and gene sequences were required. The size of the full transcript can be determined using hybridisation, colorimetric labelling, electrophoresis and northern blotting. By exploiting the known transcript sequence data and the properties of ribonucleic acid (RNA) polymerase enzyme, the full transcript sequence can be elucidated using the rapid amplification of the 5' and 3' complementary deoxyribonucleic acid (cDNA) ends (RACE) technique. This information can then be used to investigate the sequence variability by designing gene specific primers for amplification and sequencing of transcripts from different individuals. Transcript specific probes can also be developed to accommodate the variability and investigate expression (Chapters 4 and 5). Antimicrobial peptides, due to their sequence diversity and their widespread taxonomic distribution, have been difficult to categorise (Zasloff, 2002). However, web-based software tools can be used for bioinformatic analysis of the transcript sequence to suggest relationships between Carcinin and other known proteins as well as putative modifications and structural features. Through this type of analysis, the classification of this protein as a defensin molecule is investigated further. *In silico* analysis can also provide additional information about transcript control and processing, suggest putative functional and structural properties of the protein of interest and lead to a focussed experimental

approach. Bioinformatic approaches have been used to establish structural (Bulet *et al.*, 1999) and evolutionary links (Dimarcq *et al.*, 1998) between sequences, both within and between species (Gross *et al.*, 2001; Bartlett *et al.*, 2002; Cuthbertson *et al.*, 2002; Supungul *et al.*, 2002).

2.2 Specific Aims

The main aims of this chapter were:

- To obtain the full cDNA and gDNA sequence data for Carcinin.
- Assess sequence variability of the transcript between individual *C. maenas* and their relationship possible homologues in other species.
- Analyse sequence data using bioinformatic tools to hypothesise on structural characteristics and classification as a possible defensin molecule.

2.3 Methods and Materials

2.3.1 Sample Collection

Shore crabs (*Carcinus maenas*) were collected from St. Andrews Bay, Scotland in baited "creels" (lobster pots) deployed overnight in 10-20 m of water. The crabs were handled carefully on collection and placed immediately into buckets of seawater to minimise handling stress. Only crabs, which appeared superficially in good health, irrespective of shell colouration and size, were recruited.

The crabs were placed in large (2 m x 1.5 m x 1.5 m) flow-through, aerated, seawater tanks at (15 °C, ± 2 °C; ~1-2 °C higher than sea temperatures) at densities of ~10-20 individuals per cubic metre. Short sections (~15 cm long and 10 cm diameter) of plastic tubing provided shelter for the animals in the tanks; the tanks were also covered to minimise stress to the animals. The seawater circulation system in the aquarium ensured the water was replaced twice a day from St. Andrews Bay. The crabs were acclimatised at these temperatures for a minimum of 2 weeks and fed twice a week on chopped mussels and mackerel. The tanks were cleaned out once a week and any injured or dead individuals were removed. Increases in protein and RNA levels have been reported for up to 3 h after feeding (Houlihan *et al.*, 1990); therefore, the feeding and sampling regime was regulated to avoid sampling during spikes in protein and (RNA) levels. This was achieved by feeding the animals at the same time on the same days of each week and only sampling 2 days post feeding, when protein and RNA levels have been reported to return to levels similar to those observed for continuous feeding.

Crabs selected for sampling were placed in a bucket containing tank seawater. The water temperature was measured and maintained as close to the aquarium water

temperature as possible using an ice jacket or hot water bath replaced at regular intervals to avoid increased thermal stress. The buckets were gently aerated using a small air filter and covered to minimise handling stress.

All chemicals used in the present study were purchased from Sigma (Sigma-Aldrich Company Ltd., Dorset, UK) unless otherwise stated. Before collecting haemolymph, each animal was blotted dry; the right cheliped (Figure 4) was extended and swabbed with 70 % ethanol (BDH, VWR International Ltd., Dorset, UK). A 23 g needle (BD Biosciences, Oxfordshire, UK), on a 1 ml syringe, was inserted longitudinally and 0.5 ml of haemolymph was withdrawn; 0.3 ml of this was added to 1 ml of Trizol ® LS (Invitrogen Ltd., Paisley, UK) reagent in a 2 ml centrifuge tube and shaken vigorously. A second aliquot (0.1 ml) of the extracted haemolymph was diluted (1:20) with marine anticoagulant (MAC; 1020 mOsm kg^{-1}; 100 mM glucose; 30 mM trisodium citrate; 26 mM citric acid; 10 mM EDTA; 0.45 mM NaCl; pH 7.0: Söderhäll and Smith, (1983)). This diluted sample was used to establish total haemocyte counts (THC) using an improved Neubauer haemocytometer (Freshney, 2000). The samples in Trizol® LS were either processed immediately or frozen at –80 °C and stored for up to 2 months until RNA extractions could be completed.

2.3.2 RNA Isolation

All protocols, which involved RNA handling, were undertaken using baked glassware (170 °C for 6 h), gloves, nuclease-free water and molecular grade chemicals. All plastics used, were sterilised to reduce nucleases by autoclaving twice at 121 °C for 15 min. The nuclease-free water was prepared using double distilled water in baked bottles, to which was added diethylpyrocarbonate (DEPC) (0.01 % v/v) and left

overnight in a fume hood. The bottles were then autoclaved at 121 °C for 15 minutes and cooled before use in experiments.

The protocol for the isolation of total RNA using Trizol® LS (Invitrogen Ltd.), was adapted for use with crab haemolymph samples. The original isolation protocol is based on the method developed by Chomczynski and Sacchi, (1987), comprising a chaotropic lysis step followed by the addition of chloroform and centrifugation which separates the solution into and aqueous and organic phases. The RNA remains exclusively in the aqueous phase and can be subsequently precipitated with isopropyl alcohol (isopropanol).

Briefly, for the isolation of total RNA from *C. maenas*, fresh or defrosted samples (2.3.1) were shaken vigorously for 15 s and incubated for 3 min at room temperature (RT). The samples were then centrifuged at 12,000 x g for 15 min at 4 °C. The samples separated into three distinctly coloured phases. The lower, pink, phenol-chloroform phase contains the haemolymph proteins. The intermediate, white layer, contains the DNA and the lower clear layer, contains dissolved RNA. The clear aqueous layer was aspirated with a Pasteur pipette and transferred to cooled 1.5 ml labelled centrifuge tubes. Isopropanol (0.5 ml per 0.75 ml of Trizol® LS (Invitrogen)) was added and the tubes incubated for 10 min at RT. These samples were then centrifuged at 12,000 x g for 10 min at 4 °C resulting in a small white pellet (total RNA) at the bottom of the tube. The supernatant was carefully aspirated and discarded; 75 % (v/v) ethanol solution (1 ml per 0.75 ml Trizol® LS (Invitrogen)) was added to the tube, and the pellet resuspended. The sample was then centrifuged at 7500 x g for 5 min at 4 °C and the supernatant aspirated and discarded.

The pellet was air dried (~15 min) and re-dissolved in warm (~50 °C) nuclease-free water (20-50 µl). The concentration of the samples were quantified by absorbance at 260 nm (A_{260}) and quality assessed by denaturing agarose gel electrophoresis (Farrell, 1996) (2.3.2.1) and then stored at –80 °C.

2.3.2.1 QUANTITATIVE AND QUALITATIVE ASSESSMENT OF RNA

Quartz cuvettes, stored in 1:1 hydrochloric acid/methanol (VWR International Ltd.), were washed with nuclease-free water and their exteriors dried. Total RNA samples were defrosted (on ice) and a 4 µl diluted (1:200) with nuclease-free water. The diluted sample (800 µl) was added to the cuvette. The RNA concentration was measured by absorbance on an Ultrospec 3300 spectrophotometer (Amersham Biosciences) using the following equation.

$$\text{Total RNA } (\mu g\ ml^{-1}) = A_{260} \times 44.19 \times \text{dilution}$$

The relationship between total RNA concentrations and THC obtained in 5.3.1 below, was assessed by linear regression using Microsoft Excel ® software. A denaturing agarose gel electrophoresis was used to assess the quality of the RNA in the samples. An electrophoresis gel kit (Bio-Rad Laboratories Ltd., Hertfordshire, UK), including combs and end walls, was thoroughly rinsed with double distilled water and left to soak in 3 % (v/v) H_2O_2 (Phillip Harris Scientific, Leicestershire, UK) for 20 min before rinsing with autoclaved double distilled water. A 1 % (w/v) agarose (BDH) gel (50 ml) was prepared with 1 X 3-[N-Morpholino] propanesulphonic acid (MOPS), (41.8 % (w/v) MOPS, 16.6 % (v/v) 3 M sodium acetate, 20 % (v/v) 0.5 M EDTA, pH 7) and heated until the agarose was fully dissolved. Ethidium bromide was added (final conc. 1 µg ml^{-1}) when the gel had cooled to approximately 50 °C and mixed gently before

being cast. Ethidium bromide enables visualisation of the DNA in the agarose gel by interchelating between the stacked base pairs of the nucleic acid and fluorescing on exposure to ultraviolet (UV) radiation (Sambrook and Russell, 2001). Once the gel had set, the walls were removed and 1 X TBE buffer added to cover the gel. Finally, the comb was removed thereby forming the wells into which the samples were loaded.

Each sample to be analysed was prepared 1:2 with denaturing buffer (50 % (v/v) formamide, 16 % (v/v) (37 %) formaldehyde, 12.5 % 10 X MOPS buffer) in separate 0.5 ml microfuge tubes, and heated for 10 min at 65 °C, in a Touchdown or Sprint thermocycler (Hybaid Ltd., Middlesex, UK), to denature the RNA. Loading buffer (50 % (v/v) glycerol, 1mM EDTA, 0.4 % (w/v) bromophenol blue) was then added 1:1 to the denatured RNA sample, mixed and pipetted into the wells of the gel. The gel was electrophoresed at 70 V for 20 min or until the bromophenol dye front had migrated across the gel. The ethidium bromide stained RNA was visualised by ultraviolet (UV) light at 245 nm using a trans-illuminator.

2.3.3 DIG Oligonucleotide Labelling

The oligonucleotide sequence (C1; 5'-ATT GTC TTT GCA CCA GTA CTT ACA-3'), designed by Dr. J Chisholm (pers. comm.), was used as a probe (Appendix 2). This probe was labelled by 3' tailing with Digoxygenin-11-dUTP/dATP (DIG) (Roche Diagnostics GmbH, Mannheim, Germany), following the manufacturers methods, and used for analysis of Carcinin transcript size by northern blotting.

The kit supplied, contained 11 vials (Appendix 3; Table 32), with additional nuclease-free water and 20 mM EDTA (pH 8). The reagents were added to 0.5 ml microfuge tubes, chilled on ice, in the recommended order (Appendix 3; Table 3) to both the test

oligo and the control oligo (supplied) and briefly centrifuged in table top centrifuge at 500 x g for 30 s at RT.

After incubation at 37 °C for 15 min, 2 µl of the glycogen /EDTA mixture (1 µl glycogen & 0.2 M EDTA) was added to stop the reaction. The labelled probe was recovered by precipitation, as described in the manufacturers protocol using LiCl/ethanol; this was incubated at –70 °C for 30 min and centrifuged twice (<13,000 g for 15 min at 4 °C). The pellet was washed with 100 µl cold (-20 °C) (70 % (v/v)) ethanol before discarding the supernatant and air-drying. The pellet was redissolved in a minimal volume of nuclease-free water, to maintain a high DNA concentration. The labelling efficiency was checked by comparison of dot intensities of the test sample to the experimentally labelled control oligo (Appendix 3; Table 32; Vial 6) and the pre-labelled control oligo (Appendix 3; Table 32; Vial 7) by dot blotting (2.3.3.1)

2.3.3.1 QUANTIFICATION OF LABELLED PROBE

The reaction buffers were prepared as described in the manufacturers protocol (Appendix 3; Table 34) and a positively charged nylon membrane (Roche Biochemicals GmbH, Mannheim, Germany) was cut to accommodate the number of samples to be tested.

The labelled probe was diluted in nuclease-free water to 2.5 pmol µl-1 (assuming a yield of 100 pmol) and a dilution series prepared (50, 10, 2, 0.4, 0.08 and 0 fmol µl-1) for the experimentally labelled probe (C1), the control labelled oligo (Appendix 3; Table 32; Vial 6) and the pre-labelled control oligo (Appendix 3; Table 32; Vial 7). The dilutions were spotted onto the nylon membrane as shown (Figure 7) and left to dry in air before wrapping in cellophane and cross-linking for 3 min under ultraviolet (UV) light (245 nm).

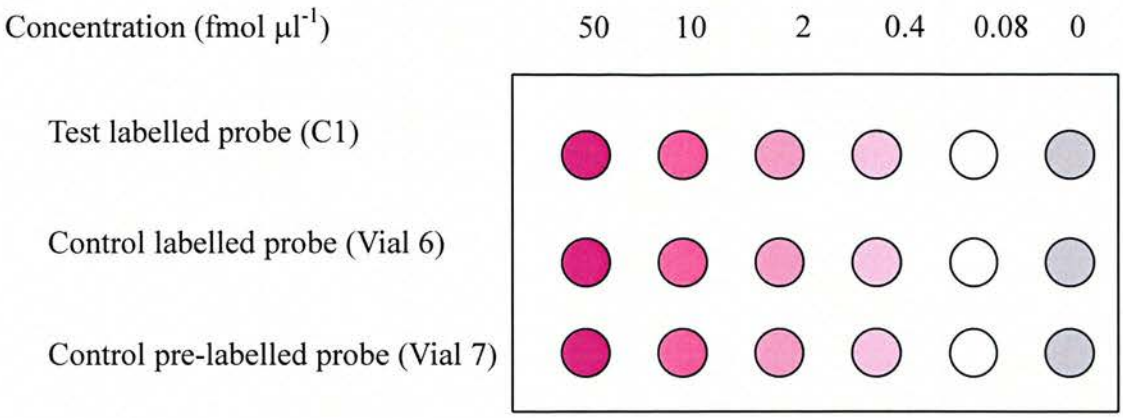


Figure 7: Graphical representation of the quantification of the C1 experimentally labelled probe. A spotted dilution series of C1 was compared to the experimentally labelled control probe and a pre-labelled probe of known concentration.

The spotted membrane was washed in washing buffer (Appendix 3; Table 34) and blocked with blocking solution for over 1 h at RT with gentle rocking. The membrane was then hybridised in blocking solution containing Anti DIG alkaline phosphotase (AP) (150 mU ml^{-1}) for 30 min at RT with gentle rocking and then washed twice in washing buffer (15 min per wash). The wash buffer was discarded and replaced by detection buffer for 2 min before the NBT/BCIP (2 % v/v) in detection buffer was added for development of the membrane in a hybridisation bag (in the dark) (Appendix 3; Table 34). Once the spots appeared on the membrane the test labelled probe was quantified against the control probe concentrations. The quantified probe was stored at –20 °C until required for use in northern blotting.

2.3.4 Northern Blotting

The electrophoresis gel kit was cleaned as described in section 2.3.3. A denaturing formaldehyde gel was prepared in a fume hood by boiling 74 ml of a 1 % (w/v) agarose/water suspension in a baked flask until the agarose had dissolved. The gel was then cooled to 60 °C before the addition of 10 ml of 10 X MOPS and 16 ml of 37 %

formaldehyde. A gel was cast to approximately 0.5-0.75 cm thick and left to set. The gaskets were removed and the gel was covered with autoclaved MOPS running buffer (10 % (v/v) 1 X MOPS, 3 % (v/v) (37 %) formaldehyde), before the comb was removed. The gel was then pre-run for 30 min at 70 V. The RNA samples (5-10 μg) were prepared and loaded, as described in section 2.3.2.1 and electrophoresed in the fume hood at 30-70 V for 2-3 h. RNA molecular weight markers III, (DIG labelled) 0.3-1.kb (Roche Diagnostics GmbH, Mannheim, Germany) were also similarly prepared and loaded into the first well for size comparison.

The gel was removed from the electrophoresis kit, washed with nuclease-free water and immersed in 20 X SSC (17.5 % (w/v) sodium chloride, 8.82 % (w/v) tri-sodium citrate, pH 7, filtered and autoclaved at 121 °C for 15 min) for 1h. The gel was removed from the solution and inverted. The inverted gel was placed onto a larger piece of chromatographic paper (3M, pre-soaked in 20 X SSC buffer) and this was placed onto a cling film covered support, standing in 20 X SSC buffer. The overhanging edges of the larger filter paper were dipped into the buffer. A positively charged nitrocellulose membrane (Roche Diagnostics GmbH, Mannheim, Germany) (cut to size of gel) was placed on the top of the gel, as shown below (Figure 8).

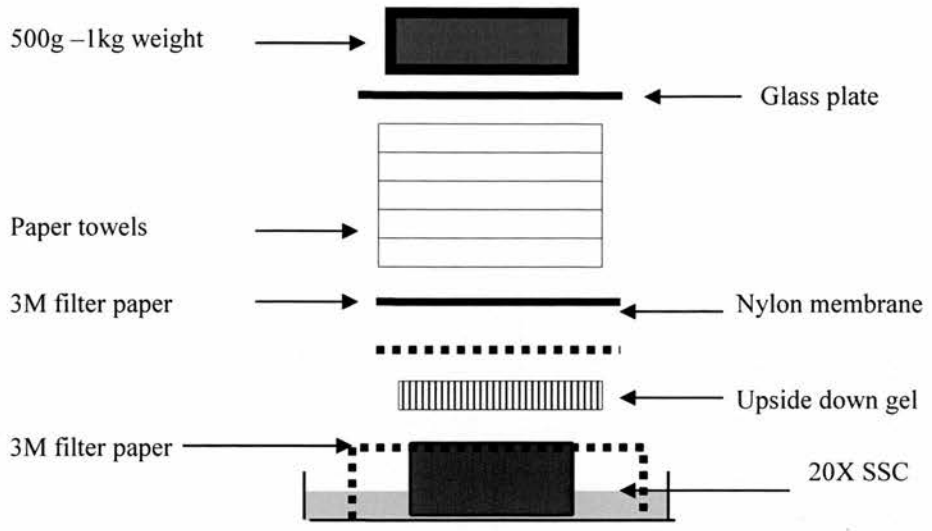


Figure 8: Northern blotting cassette construction

On top of the nitrocellulose membrane were placed two sheets 3 M filter paper or chromatographic paper. On top of this, several sheets of paper towels and a large weight (500-1000 g) were added. The gel was blotted for a max of ~16 h (transfer time normally about 10 min mm^{-1} of gel thickness) at RT, with several changes of paper towels. The cassette was then disassembled; the membrane was marked for orientation and was then wrapped in cling film and cross linked with UV light (245 nm) for 3 min. The transfer of RNA was confirmed by immersing the blotted gel in ethidium bromide solution (0.5 μg ml^{-1}) for 5 min and then visualising it under UV light.

The membrane was then pre-hybridised in high SDS buffer (7 % (w/v) SDS, 10 % (v/v) 20 X SSC, 20 % (v/v) 10 X Block, 5 % (v/v) sodium phosphate, 5 % (v/v) (10 %) N-laurylsarcosine dissolved in 250 ml DEPC water, autoclaved at 121 °C for 15 min; 250 ml formamide was added and the mixture stored at 4 °C) for 2 h at 43 °C (containing poly (A^+) at 0.1 mg ml^{-1}). The pre-hybridisation solution was replaced with hybridisation solution consisting of fresh high SDS buffer, containing DIG-tailed

oligonucleotide (0.4 pmol ml^{-1}), which was incubated overnight at 43 °C (Tm = 52.2 °C) in a sealed hybridisation bag.

The membrane was then washed twice (5 min) in 2 X wash (1 % v/v 20 X SSC, 0.5 % SDS (20 %)) at 43 °C and this process was repeat with 0.5 X wash (2.5 % v/v 20 X SSC, 0.5 % SDS (20 %)). The membrane was then equilibrated with basic wash buffer (99 % v/v maleic acid buffer, 0.3 % TWEEN 20) for one minute and then blocked with 1 X Blocking solution for 1 h. To fresh blocking solution, Anti-DIG-AP Fab fragments (Roche Diagnostics GmbH, Mannheim, Germany) (Appendix 3; Table 34) (1/5000) were added and the membrane was bathed in this solution for a further 30 min followed by two washes in basic wash buffer, each for 15 min. Final equilibration was achieved by placing the membrane in detection buffer (Appendix 3; Table 34) for 2 min. The membrane was then developed in detection buffer containing NBT/BCIP solution (Appendix 3; Table 34) (1:50), in a sealed hybridisation bag wrapped in foil and kept in the dark until bands developed (~3-4 h).

2.3.5 First Strand Synthesis

First strand synthesis, of all RNA samples collected, was achieved using the using recombinant M-MLV Reverse Transcriptase (M1302) from *Escherichia coli* at 200 units μl^{-1}. The protocol, provided with the enzyme, was used as indicated, with additional dNTP's (Promega, Southampton, UK) and oligo dT (Promega) as the primer. Oligo dT primer anneals to the 3' end poly (A^+) tail which is only present on mRNA transcripts and not on other nucleotide strands (e.g. gDNA or tRNA).

Briefly, the dNTP's (10 mM), oligo dT (1-5 μm) and total RNA (1-5 μg) were added in a thin walled polymerase chain reaction (PCR) tube (Axygen, California, USA) and

denatured at 70 °C for 10 min in a water bath. The samples were incubated at room temperature for a further 10 min to ensure elongation of the anchored oligo dT. Two microlitres of 10 X reverse transcriptase buffer (500 mM Tris-HCl, pH 8.3 with 500 mM KCl, 30 mM $MgCl_2$ and 50 mM 1, 4 dithiothreitol (DTT); supplied), 200 units M-MLV enzyme, and nuclease-free water to 20 µl were added to this mixture and further incubated at 37 °C for 50 min. The reaction was inactivated by heating to 85 °C for 10 min before cooling on ice.

To confirm the success of the reverse transcription reaction, the cDNA samples were amplified using polymerase chain reaction (PCR) with the selected primers. The resultant products were analysed by gel electrophoresis. Initially, gene specific primers (Figure 9; Carc1 primers) were designed on the known sequence (AJ237947) (Appendix 2) and these were used to make 5' and 3' RACE templates (forward and reverse primers) in section 2.3.6. After the full transcript sequence became available (AJ427538) (Appendix 7), new primers were designed (Figure 9) to include the full coding region (Carc2 primers) and these were used for all subsequent amplifications. The first primer pair (Carc1 primers) produced a 201 bp amplicon and the second pair (Carc2 primers), a 522 bp amplicon. All primers were synthesised by MWG Biotech (Ebersberg, Germany).

Carc1FWD	5'-GGT GCA AAG ACA ATC TAC TAG GA-3'
Car1REV	5'-GCT TCA GGC AGG TGT CAT-3'
Carc2FWD	5'-AGA CCA GAA CTG CAC CCT GT-3'
Carc2REV	5'-CAG TAC AGT AGG TAA CCA TGC GTC-3'

Figure 9: The two sets of gene specific primers used to amplify Carcinin transcript.

The cDNA (2.3.5) was amplified using these primers with Advantage® *Taq* polymerase mix (BD Biosciences., Hampshire UK). The Advantage® *Taq* polymerase mix comprises three main components: two polymerases (the first being KlenTaq-1 DNA polymerase primarily for amplification and the other with 3' to 5' "proof reading" activity) and a TaqStart™ Antibody providing "hot start" PCR conditions for the reaction (Kellogg *et al.*, 1994).

The cDNA template from each crab sampled was diluted 1 in 5 with nuclease-free water and 0.5 µl of this was used in the reaction mix in a 0.5 ml PCR tube (Axygen). With the tubes kept on ice, 0.5 µl of dNTP's (100mM) (Promega), 1 µl of both forward and reverse primers (10 pmol), 2.5 µl of the 10 X buffer (400 mM Tris-KOH pH 9 at RT, 150 mM KOAc, 35 mM $Mg(OAc)_2$), 37.5 µg ml^{-1} bovine serum albumin (BSA) (supplied) and 0.5 µl (0.55 µg) Advantage® *Taq* were added and the reaction mixture was made up to 25 µl with nuclease-free water. The tubes were briefly centrifuged to concentrate the sample. A simple thermocycling profile was followed, of an initial denaturing stage at 94 °C for 5 min, then 30 cycles of 94 °C for 30 s, an annealing temperature of 56 °C using the Carc1 primers or 57 °C using the Carc2 primers, both for 30 s and then 2 min at 72 °C, with a final extension time of 7 min at 72 °C.

To check the PCR reaction, 1 µl of each cDNA sample was mixed with 1 µl of 6 X DNA loading buffer (100 mM EDTA (VWR International, Dorset, UK); 25 mM Tris-HCl; 25 % glycerol; 0.05 % bromophenol blue (BDH); pH 7.0) (Sambrook and Russell, 2001) and 3 µl of nuclease-free water and run on a 2 % agarose gel. The agarose (BDH) was dissolved in 0.5 X TBE (45 mM Tris-Base, 1 mM EDTA, 44 mM Boric acid; pH 8.0) (Sambrook and Russell, 2001) and heated to dissolve the particles. Ethidium bromide (1µg ml^{-1}) was added to the cooled gel. The gel was electrophoresed in ~50-100 ml of 0.5 X TBE buffer running buffer. The gel was run at 70 V for approximately 30 min until the dye front had migrated most of the length of the gel. Molecular markers (Bioladder 100; Hybaid Ltd, Middlesex, UK and EZ Load 100 bp marker; BioRad Labs Ltd. Herts, UK) were loaded alongside the samples to determine the size of the PCR products. The resulting gels were viewed and photographed on a UV transilluminator and all cDNA and PCR samples were stored at –20 °C.

2.3.6 Product Isolation, Cloning, Plasmid Extraction and Sequencing

All "candidate" PCR products were excised with a sterile scalpel and the DNA purified from the agarose gel using the QIAQuick® gel extraction kit (Qiagen Ltd, West Sussex, UK) with slight modification of the proprietary protocol (Appendix 4).

The pCR® 2.1-TOPO® vector (Invitrogen, California, USA) (Appendix 5) was used to clone the PCR products. The manufacturers protocol was followed with only slight modification to the component volumes stated. One microlitre of the purified PCR product (~20 ng) was added to 0.25 µl of the supplied salt solution (1.2 M NaCl; 60 mM $MgCl_2$) and 0.25 µl of pCR® 2.1 vector (supplied). This mixture was gently mixed and incubated at RT for 5 min and then immediately placed on ice. The TOP10F' competent cells (12.5 µl) (supplied) were defrosted on ice, added to the cloning reaction

and incubated on ice for a further 30 min. The mixture was then heat shocked at 42 °C for 30 s (exactly) using a heated water bath (to allow entry of the vector through the bacterial membrane) and then the mixture was returned immediately to the ice. The warmed (RT) SOC medium (supplied) (65.5 µl) was then added to the mixture and the sample incubated at 37 °C for 1 h in a shaking (~200 rpm) water bath.

Luria-Bertani (LB) agar (ICN Biomedicals Inc, Ohio, USA) plates were prepared containing 50 µg ml^{-1} ampicillin. These were spread with 40 µl of isopropyl thiogalactoside (IPTG) (Promega, Southampton, UK) and 40 µl X-gal (5-bromo-4-chloro-3-indolyl-β-D-glactopyranoside) (Promega), using sterile glass spreaders. The plates were then dried and warmed under aseptic conditions to 37 °C in an air incubator.

Following the incubation at 37 °C, the cloning reaction was spread onto the preheated agar plates and incubated at 37 °C for a further 12 h or overnight. The plates were then screened for recombinant colonies using the blue/white visual screening; blue coloured colonies denotes unsuccessful transformations due to the expression, by the pUC 18 plasmid, of β-galactoside, which cleaves X-GAL to produce a blue product. With successful transformations (white colonies), the inserted DNA disrupts expression of the β-galactoside, so no coloured product is expressed.

White colonies were selected and screened by colony PCR using the M13 primers (supplied). The M13 primers anneal to either side of the multiple cloning region of the pCR® 2.1 vector (Appendix 5) and can be used to amplify the internal DNA sequence. The appropriate control reactions were also undertaken, including single primer reactions, blue colonies (no insert, so smaller product) and the amplification of control DNA (supplied) to check that no primer dimers were formed and that the PCR conditions were correct. The PCR reaction was set up as described in section 2.3.6

using the Advantage® cDNA polymerase mix (BD Biosciences). LB/ampicillin agar plates were prepared and labelled with a numbered grid corresponding to the colony and PCR reaction. Selected colonies were touched with a sterilised (autoclaved at 121 °C for 15 min) cocktail stick which was then touched onto a numbered grid position and then the stick was dipped into the PCR reaction mixture. The plate was incubated overnight as previously described (at 37 °C for 12 –16 h). The PCR cycling conditions involved an initial denaturation step at 95 °C for 2 min and then 30 cycles of denaturation at 95 °C for 30 s, annealing at 55 °C for 30 s and extension at 72 °C for 1 min. The additional elongation step for the addition of an adenosine (A) overhang at the 3' end was not required. The PCR products were electrophoresed and visualised with UV light, as previously described in 2.3.5 on a 2 % agarose gel.

Colonies that contained plasmids with an insert of the expected size (~700 bp) were grown and subjected to plasmid extraction. Plasmid extraction was performed using the Wizard™ plus SV minipreps DNA purification system (Promega UK, Southampton, UK) (Appendix 6). Sequencing of the extracted plasmids was performed by The Sequencing Service (University of Dundee, Dundee, Scotland) with both forward and reverse M13 primers (provided in the TOPO TA cloning kit). The results were analysed using Chromas (C. McCarthy, Queensland, Australia: version 1.45, 32 bit) and DNAMAN (version 5.2.0, Lynnon BioSoft, Quebec Canada) software packages.

2.3.7 Rapid Amplification of cDNA Ends (RACE)

To determine the 5' and 3' ends of the Carcinin transcript, RACE was performed by Dr. J. Hammond, using the SMART™ RACE cDNA amplification kit (BD Biosciences, Oxfordshire, UK). This technique was adapted (as described by Hammond, J.A.

(2002)) from the protocol provided with the kit and used with all the appropriate primer controls.

Briefly, reverse transcription was used to create 5' and 3' RACE templates from isolated total RNA (2.3.2). These templates were hybrid sequences comprising of Carcinus cDNA with a RACE specific sequence at either the 5' or 3' end. These were then used in primary RACE PCR, with both RACE adaptor specific and Carcinin specific primers using the Advantage® cDNA polymerase mix (BD Biosciences, Oxfordshire, UK) as described above (2.3.7). The resulting primary PCR product provided the template for a second round of PCR using gene specific and adaptor specific nested primers to obtain the unknown 5' or 3' ends. Products were visualised by gel electrophoresis (2.3.5) and those of interest were isolated as described in 2.3.6 above. The isolated DNA was cloned using the TOPO TA (Invitrogen) Cloning Kit following the manufacturer's protocols. Plasmids were isolated, sequenced and analysed as described in section 2.3.6.

2.3.8 Genomic DNA Sequencing

Genomic DNA (gDNA) was isolated from two individual animals using the DNAeasy Kit (Qiagen, Sussex, UK) as per the proprietary protocol. The Carcinin gDNA was amplified using the Carc2 primers (Figure 9); this product was purified, cloned and sequenced in both directions, as previously described (2.3.6).

2.3.9 Isoforms

Once the full transcript sequence became available (Appendix 7), the Carc2 primers (Carc2FWD 5'-AGA CCA GAA CTG CAC CCT GT-3' and Carc2REV 5'-CAG TAC AGT AGG TAA CCA TGC GTC-3') were used to amplify the full transcript sequence

(522 bp). Ten male individual crabs (acclimated to 16 °C ± 1 °C) were selected at random, mRNA was extracted as described 2.3.2. All crabs were similar in colour and approximate size (60-70 mm) and were collected at the same time of the year. First strand synthesis was performed as described in section 2.3.5. Using the reaction mixes (2.3.5), the following cycling conditions were used: 94 °C for 5 min and then 30 cycles of 94 °C for 30 s, 57 °C for 30 s and 2 min at 72 °C and then a final extension step of 7 min at 72 °C. Cloning, screening, plasmid isolation and sequencing was performed on single, insert positive (~700 bp) colonies from each individual, as described in 2.3.6.

The sequence data generated in the present study were compared to the data presented by Relf *et al.* 1999 (AJ237947) and a consensus sequence was suggested by multiple alignment, taking into account protein, cDNA, and gDNA sequences.

2.3.10 Sequence Analysis

The strategy employed in the present study, was to compare the transcript variants and the inferred protein sequences (2.3.9) with conserved sequence patterns and post translation modifications (PTM) of known antimicrobial proteins. Next, their sequence identity to other proteins and translated cDNA sequences were investigated. Finally, any conserved motifs, domains and structural features were identified. Although this type of analysis had been conducted with the initial sequence data published by Relf *et al.* (1999), re-analysis was justified in light of the results obtained in the present study. This was aided by the growing database of bioinformatic tools (Galperin, 2004).

The gDNA sequence was aligned with the consensus mRNA transcript sequence, using the Spidey software (www.ncbi.nlm.nih.gov/IEB/Research/Ostell/Spidey/) to identify intron-exon boundaries.

Using the ProtParam software, available from the EXPASY website (www.us.expasy.org/tools/protparam.html), the consensus protein sequence was analysed for composition, theoretical pI, probable mass, and predicted stability. The DNAMAN package was used to predict a hydrophobicity profile and charge data for the protein sequence. Possible signal sequences and cleavage sites were investigated using the SignalP v1.1 program (www.cbs.dtu.dk/services/SignalP/) (Neilsen *et al.*, 1997).

The inferred protein sequence was investigated for functional motifs and domains using the InterProScan package available from the EBI site (www.ebi.ac.uk/InterProScan/). This package integrates the PROSITE, PRINTS, Pfam, ProDom, SMART and TIGRFAMs databases (Zdobnov and Apweiler, 2001). These databases identify conserved motifs and domains, integrating biological context, sequence alignment and structural features (Westhead *et al.*, 2002).

Several types of post-translational modifications have been reported for AMPs including amidation, glycosylation and disulphide bond formation. However, activity is not always dependent on these modifications (Andreu and Rivas, 1998) and may even elicit the opposite effect. Basic post-translational modifications for Carcinin were investigated using the predictive software available from Predict Protein site (www.embl-heidelberg.de/predictprotein/predictprotein.html) and CBS prediction servers (www.cbs.dtu.dk/services/). CYSPRED analysis (Fariselli *et al.*, 1999) was used to predict the formation of disulphide bonding for the cysteine residues present in the sequence. The YinOYang 2.1 server was used to predict O-β-GlcNAc sites and the NetPhos 2 server was used to predict possible phosphorylation sites (Blom *et al.*, 1999).

Sequence similarity searches were carried out using the inferred consensus protein sequence (2.3.9) against the databases (SWISS-PROT, EMBL, TrEMBL and PIR)

maintained by the European Bioinformatics Institute (EBI) (www.ebi.ac.uk/). Using inferred protein sequences for similarity searching is considered more powerful than using the cDNA sequence due to the degenerate nature of codon usage (Westhead *et al.*, 2002). The BLAST (TBLASTN and BLASTP) programs used for this analysis were accessed through the Japanese GenomeNet (www.genome.ad.jp/) site. PSI-BLAST, which identifies more distant relatives of Carcinin using an iterative process (Westhead *et al.*, 2002), was accessed through the National Centre for Biotechnology Information (NCBI) site (www.ncbi.nlm.nih.gov/BLAST/Blast.cgi). The BLOSSUM62 (Henikoff and Henikoff, 1992) scoring matrix was used to identify related sequences using pairwise alignment with an E-value threshold of less than 0.001 for BLAST and 0.005 for PSI-BLAST. These values suggest the likelihood of homology between sequences not occurring by chance (i.e. E values closest to 0).

Secondary structure prediction was performed using the DNAMAN package, psi-pred (www.globin.bio.warwick.ac.uk/psipred/) and GOR 4 programs, also available from the EXPASY site.

It has been accepted that proteins may share the same structure if their sequences can be aligned to show at least 25 % identity (as reviewed by Westhead *et al.*, 2002). The protein sequences of known WAP domain containing proteins were aligned with Carcinin using the ClustalW pairwise alignment tool available from Pole Bioinformatic Lyonaise site (www.npsa-pbil.ibcp.fr/cgi-bin/npsa_automat.pl?page=npsa_clustalw.html). Suitable templates were also investigated for comparative 3D modelling of the Carcinin protein sequence using the CPH Models 2.0 program (Lund *et al.*, 2002) (www.cbs.dtu.dk/services/CPHmodels/).

2.4 Results

2.4.1 RNA isolation

Absorbance ($A_{260/280}$) ratios of 1.8 - 2.0 are widely accepted (Farrell, 1996) to indicate the purity of the RNA sample. The values obtained in the present study had a mean ratio of 1.53. Although this is lower than the commonly accepted value, this is a consistent result for good quality RNA (as determined by visualisation of electrophoresed gels) extracted from the species. Analysis by gel electrophoresis confirmed the quality of these samples, with clear 28S and 18S bands and mRNA smearing, as see in Figure 10 below. In the present study, it was commonly observed that the 18S rRNA band was usually brighter than the 28S which is in contrast to the pattern observed in vertebrate species (Farrell, 1996).

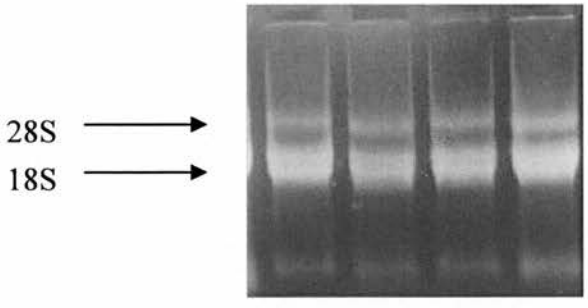


Figure 10: Denaturing agarose electrophoresis gel of *Carcinus maenas* RNA showing 28S, 18S rRNA .

No significant correlation (R = 0.446, p = 0.197) existed between RNA concentration and THC (Table 2) for the samples collected. This may be due to the small sample size and high inter animal variability in THC as previously observed by others (Chisholm, 1993).

Sample	RNA Concentration ($\mu g\ ml^{-1}$)	Temp (°C)	THC ($cm l^{-1}$)
55	1060	15	5.6×10^7
57	2080	15	6.11×10^7
58	1380	15	6.4×10^7
59	4400	15	6.8×10^7
60	600	17	7.2×10^7
80	200	17	8.6×10^5
81	640	17	1.3×10^7
82	1700	17	3.1×10^7
84	880	17	3.1×10^7
108	2160	17	2.4×10^7

Table 2: Total RNA concentrations and THC obtained from 0.3 ml of crab haemolymph

2.4.2 Northern Blotting

Colorimetric analysis was used to determine the size of the full transcript. Spot intensities were compared between the known concentrations of the control labelled oligo and the experimentally labelled oligo (C1) allowed approximation of labelling efficiency of the C1 probe. Labelling efficiency using this method for the C1 probe was between 40-100%.

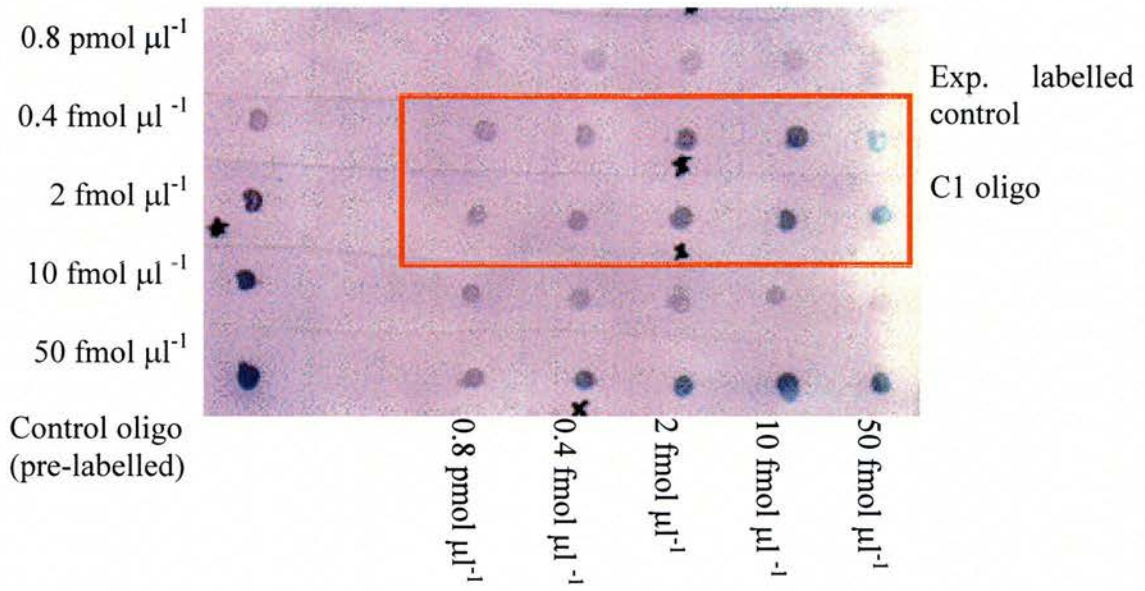


Figure 11: Efficiency of DIG labelling reaction analysed by comparing the spot intensities of the pre-labelled control oligo (supplied) with the experimentally labelled control oligo provided and the C1 oligo. Similar intensities were marked (*). C1 sample was quantified at ~2.5 pmol μl^{-1} of labelled probe (~100 % efficient).

The probe concentration was optimised for the production of good hybridisation plots and the final concentration used was 0.4 pmol μl^{-1} (Figure 12).

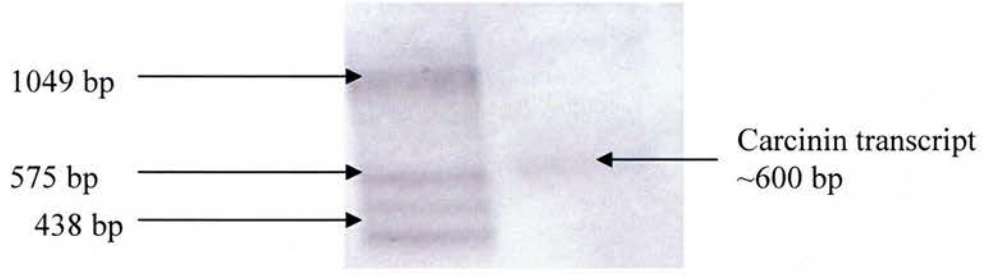


Figure 12: Northern blot of Carcinin mRNA transcript visualised using DIG oligo nucleotide labelled probes. Transcript size was compared to RNA molecular weight markers III, (DIG labelled) 0.3-1.5 kb (Roche Diagnostics GmbH).

The Carcinin transcript appeared to be approximately 600 bp in size. Although Carcinin was thought to be reasonably highly expressed, it appeared only faintly on

northern blots. Northern blotting proved to be a very laborious technique which required substantial optimisation to obtain reliable results.

2.4.3 First Strand Synthesis

Amplification of the first strand synthesis products, using Carc1 and Carc2 primer sets were visualised as bands at ~200 bp and ~500 bp, respectively (Figure 13 and Figure 14).

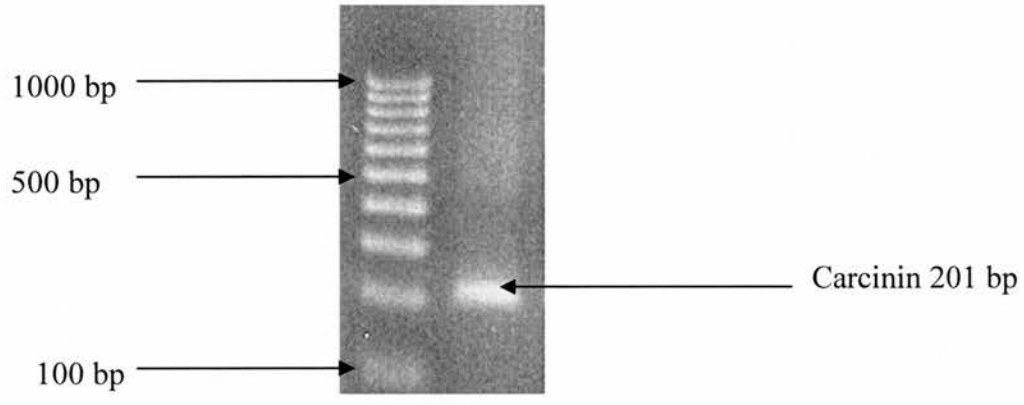


Figure 13: 2 % agarose gel of amplicons resulting from PCR using the Carc1 FWD and REV primers resulting in a 201 bp band using EZ load 100 bp marker (BioRad).

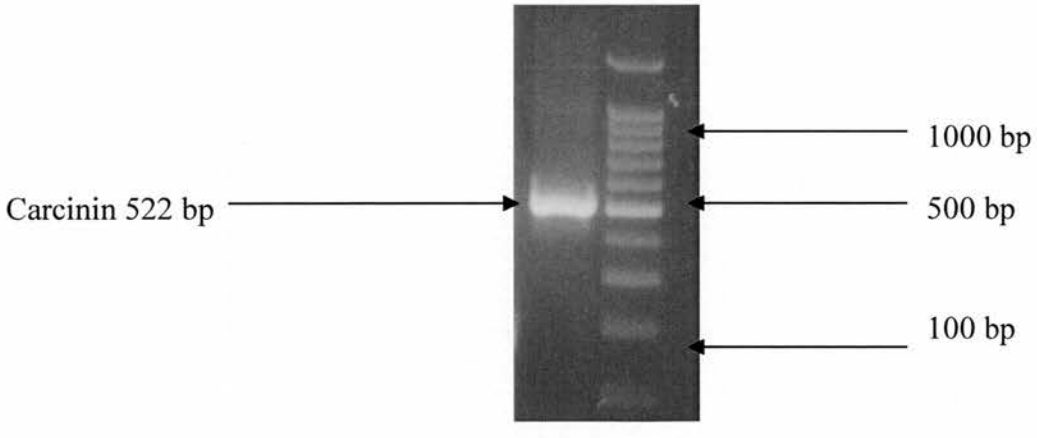


Figure 14: 2 % agarose gel of amplicon resulting from PCR using the Carc2 FWD and REV primers resulting in a 522 bp band using Bioladder 100 marker (Promega).

2.4.4 *Genomic DNA*

The amplified gDNA product was visualised as a ~920 bp band on a 2 % agarose gel.

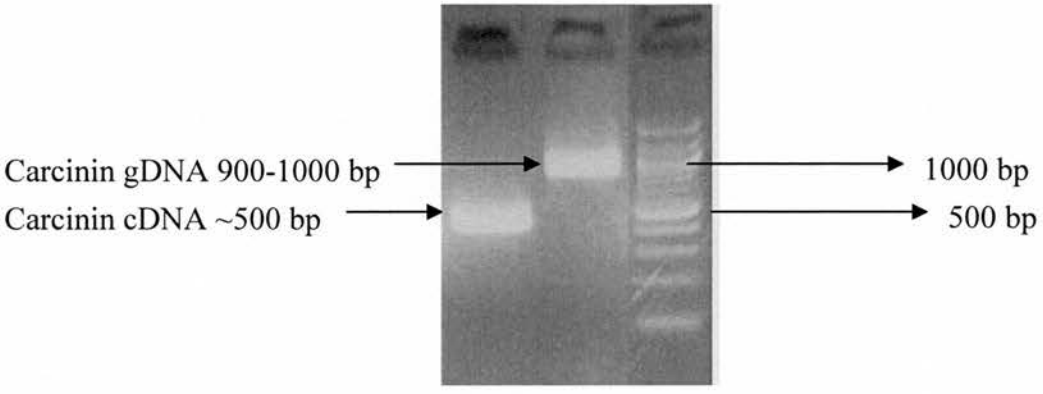


Figure 15: 2 % Agarose gel indicating the transcript (cDNA) amplicon and the gDNA amplicon using the same Carc2 primers

Alignment of the genomic and transcript sequences revealed the Carcinin transcript to comprise of 4 exons and 3 introns (Appendix 8). The genomic sequence was not complete for the 5' end of the transcript and consequently, the transcription initiation site could not be defined. Eukaryotic introns are highly conserved generally starting at the 5' end with GT and ending with AG (Farrell, 1996). Carcinin gDNA follows this pattern although the second intron does appear to end in a CA; this is thought to be a nucleotide assignment error.

2.4.5 *RACE*

The consensus sequence obtained using RACE (AJ427538) (Appendix 7) extended the previously submitted sequence (AJ237947) (Appendix 2) by 21 amino acids at the N-terminus and by 1 residue at the C-terminus. The 5' end of the nucleotide sequence was extended by 141 bp and the 3' end by 237 bp including the poly A^+ tail of 45 bp. The additional N-terminus sequence of the protein contained a valine rich low complexity

sequence with two methionine residues. The C-terminal end was extended by a single tyrosine residue before the TAG stop codon.

2.4.6 Isoforms

It was noticed that the assignment of some of the residues which were determined through protein sequencing (Relf *et al.*, 1999), were later changed by Relf when her cDNA sequences were used to supplement the gaps in the protein data (Appendix 9). Alignment of the full transcript sequence (AJ427538), obtained in the present study, with the sequence determined by Relf *et al.* 1999 (AJ237947), highlighted further differences between the sequences at four residues (Appendix 10). Using the transcript (Appendix 10) and gDNA (Appendix 8) sequence data, determined in the present study, these discrepancies were further analysed. The transcript data from all studies was consolidated to investigate suspected residue assignment errors and confirm the residues leading to isoforms. A general consensus protein sequence (Appendix 11) was determined which was used in subsequent bioinformatic analyses.

In total seven residues have either been corrected or identified as leading to alternative isoforms and these are summarised in Table 3 and their locations are indicated in Figure 16. All residue changes described below are described based on the most recent complete sequence, AJ427538 (Figure 16). The signal sequence (further described in 2.4.7) is 100 % conserved between the four isoforms, and this conservation of the signal sequence has also been shown within the putative penaeid sequences (Cuthbertson *et al.*, 2002).

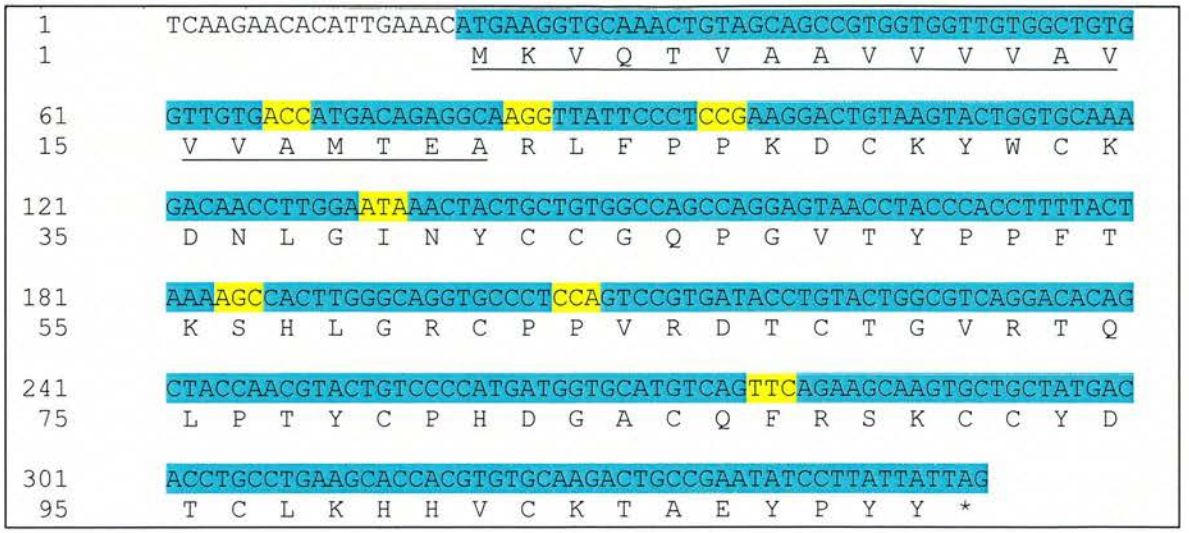


Figure 16: Translated AJ427538 sequence showing the full coding sequence (blue) and the variable codons (yellow) with the putative signal sequence underlined.

Both, translated gDNA and all subsequent cDNA data confirm that, the alanine at position 17 (A_{17}) should in fact be a threonine (T_{17}) (Table 3). It is thought that the assignment of this residue as an alanine in sequence AJ427538 had arisen due to a nucleotide assignment error.

Protein sequence data (Relf *et al.*, 1999), gDNA and cDNA data all concur that the arginine at position 22 (R_{22}) on the AJ427538 sequence should be a glycine (G_{22}) although either of these would maintain the cleavage site (2.4.7).

However, a marked discrepancy occurs between the protein sequence and all subsequent cDNA and gDNA sequences at position 26 (P_{26}); on the protein sequence, this is reported to be an asparagine (N) (Appendix 9), whereas it is a proline on all the other sequences. These residues are not thought to be commonly substituted for each other since proline is a hydrophobic residue and asparagine is a polar residue. This is highlighted further, by a PAM250 substitution matrix score of -1. Substitution matrices are used to show scores describing the likelihood on amino acid substitution between pairs of residues; high numbers indicate likely evolutionary substitutions and low

numbers indicate low probability substitutions (Westhead *et al.*, 2002). Other matrices, such as the BLOSUM series, are commonly used to describe greater evolutionary distances. The PAM250 matrix was used here to give an initial indication of the relationships between the candidate substitutions identified. The score obtained above suggests that the residue assignment reported by Relf *et al.* 1999 may be incorrect and should in fact be a proline. It was concluded that this was not a residue that leads to further isoforms (Table 3).

The isoleucine at position 39 (I_{39}) (Figure 16) was reported as a leucine (L_{39}) in the original protein data (Appendix 12 and Appendix 9). Subsequent cDNA sequences suggest either a valine or an isoleucine at this position; an isoleucine at this position is confirmed by the gDNA data. These three residues are commonly substituted for each other (Westhead *et al.*, 2002), with PAM250 scores all above 2. These residues have similar physicochemical properties (all hydrophobic) and similar sizes (~117-131 g mol^{-1}). This suggests that the different isoforms have arisen due to evolutionary substitutions, with little effect on conformation and therefore function.

The serine at position 56 (S_{56}) (Figure 16) may be a residue assignment error, since all other cDNA sequences suggest that this residue should be a lysine (K) or an asparagine (N). However, the gDNA data supports a serine residue at this position. The PAM250 substitution scores do not strongly support these residue substitutions. Therefore this position must be considered as only tentatively leading to additional isoforms of the protein (Table 3).

The proline at position 63 (P_{63}) (Figure 16) was also reported equal frequency as an alanine in the sequenced cDNA transcripts in the present study. Both alanine and proline are hydrophobic and are often substituted for each other as they have a PAM250

score of 1. No protein sequence data was available to support this, however, the gDNA sequence suggests a proline at this position.

Finally, a difference only observed between the gDNA sequences suggests that there may be a further substitution at F_{87} for a leucine. However, all the cDNA data suggests this residue to be a phenylalanine (F) and this is supported by the protein sequence data obtained by Relf *et al.* (1999). The PAM250 score for substitution of these two residues is quite high at 2 suggesting possible evolutionary substitution. However, the discrepancy between the gDNA sequences may also be a result of an erroneous nucleotide assignment. It is not clear whether or not this residue leads to an isoform (Appendix 10) but further sequencing could confirm this.

Position	Protein data	AJ237947	AJ427538	gDNA	cDNA	Consensus	Isoform?
A_{17}	-	-	A	T	T	T	NO
R_{22}	G	G	R	G	G	G	NO
P_{26}	-	N	P	P	P	P	NO
I_{39}	L	L	I	I	V/I	V/I/L	YES
S_{56}	-	K	S	S	N?/K	N?/K/S	YES
P_{63}	-	A	P	P	P/A	P/A	YES
F_{87}	F	F	F	F/L	F	F/L	YES/NO?

Table 3: Summary of evidence for residue substitutions, including data from protein fragments, cDNA and gDNA sequences.

The possible substitutions at these positions could theoretically give rise to up to >36 different isoforms of the same sequence. However, to date, only four putative isoform sequences (Accession numbers; AJ821886-AJ821889) have been identified from cDNA and two putative sequences from gDNA sequencing (Appendix 10). Only one of the isoforms (Appendix 10; AJ237947) is known to exhibit antibacterial activity (Relf *et al.*, 1999).

M K V Q T V A A V V V V A V V V T M T E A G L F P P K D C

K Y W C K D N L G I/V/L N Y C C G Q P G V T Y P P F T K S/N/K

H L G R C P P/A V R D T C T G V R T Q L P T Y C P H D G A C

Q F/L R S K C C Y D T C L K H H V C K T A E Y P Y Y *

Figure 17: Summary of residue substitution possibilities on the full Carcinin coding sequence.

Reliable cDNA sequences were obtained from seven of the ten individual crabs sequenced. Isoform 1 (Appendix 10) was the most common transcript and was

identified in 3 different crabs. Isoforms 2, 3 and 4 were identified once each in three different crabs. The gDNA sequences, which may code for two further isoforms were obtained from 2 different crabs (Appendix 10).

2.4.7 Sequence Analysis

The consensus inferred protein sequence (Appendix 11) comprises a coding region with two ATG translation start sites (ORF Finder at www.ncbi.nlm.nih.gov/gorf/orfig.cgi) although the first methionine (Appendix 7; M_1) was predicted to represent the start codon (DNAMAN). This codon is also thought more likely as the translational initiation site, as it is found immediately preceding an exon-intron boundary and that the "usual" essential purine (A) is situated three nucleotides upstream (Farrell, 1996).

The ProtParam tool suggested that the inferred putative full sequence has a molecular mass (without any PTM) of ~12.260 kDa, a theoretical pI of 8.73 and to be dominated by cationic residues. Although cysteine residues are highly represented in the sequence (10.9 %), as previously reported for the purified protein (Relf *et al.*, 1999), the full inferred sequence is dominated by valine residues (~11.8 %) particularly at the N-terminus. The instability indcx (Guruprasad *et al.*, 1990) was calculated at 35.02 which suggests this sequence to be stable with a relatively short half-life (10-30 h) (Baschmair *et al.*, 1986). If the alternative methionine residue (M_{18}) was the translational start site then the stability index drops and the protein sequence is classified as unstable so this further supports the first methionine as the translational initiator.

Sequence analysis using SignalP V1.1 software (http://us.expasy.org/tools) (Neilsen *et al.*, 1997) suggested the first 21 residues of the inferred protein sequence are a possible signal sequence with a clear cleavage site identified between residues 21 and 22 at

TEA-GL (Figure 18) immediately preceding the maximal Y score (a derivative of the combination of the signal (S) score and the cleavage (C) scores).

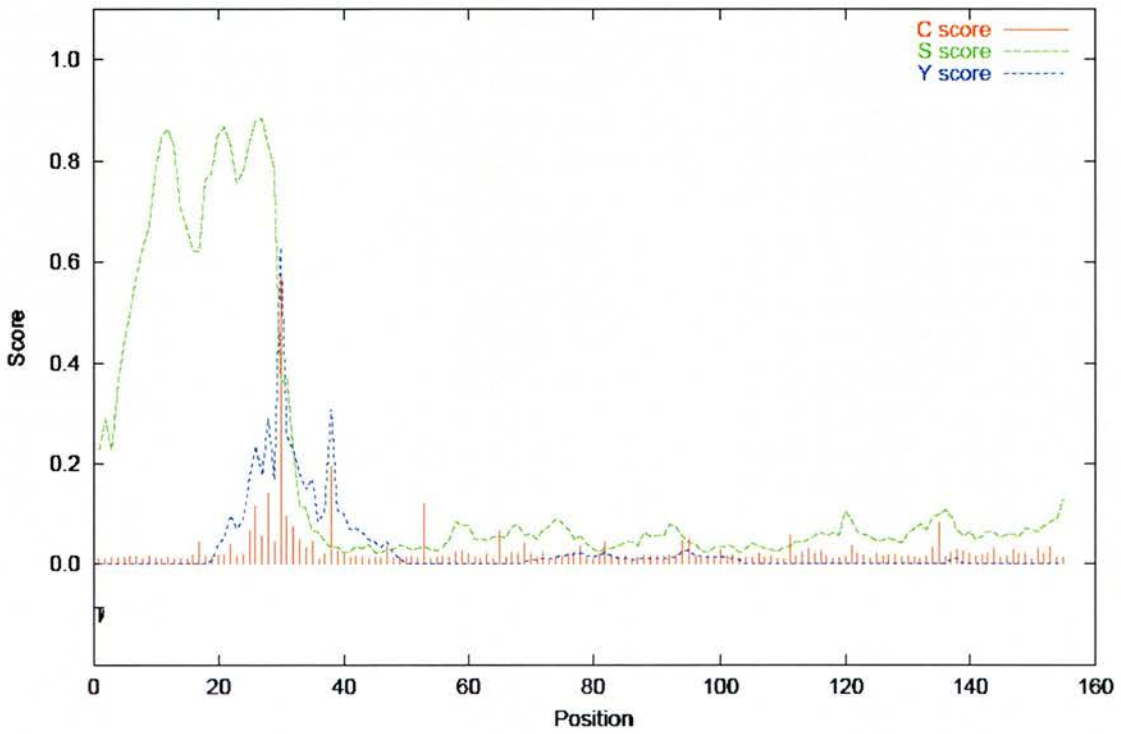


Figure 18: SignalP postscript graphic indicating signal peptide with cleavage site between residues 21 and 22.

This putative signal sequence is valine rich but exhibits no similarity to signal sequences found in other crustacean sequences (Bartlett *et al.*, 2002; Supungul *et al.*, 2002). This area of low complexity (Figure 19; yellow) was also identified by the DAS transmembrane prediction server (http://www.sbc.su.se/) and SEG software (Wootton and Federhen, 1996) as a possible transmembrane sequence. This function has been suggested to be characteristic of defensin molecule N-terminal signal sequences allowing insertion into the endoplasmic reticulum (Ganz and Lehrer, 1994). However, this is not thought to be its function in Carcinin due the expression differences between insect and crustacean AMP (discussed further in Chapters 5 and 6).

1 MKVQTxxxxxxxxxxxxMTEAGLFPPKDCKYWCKDNLGINYCCGQPGVTY

51 PPFTKKHLGRCPAVRDTCTGVRTQLPTYCPHDGACQFRSKCCYDTCLKH

100 HVCKTAEYPYY

Figure 19: Identification of area of low complexity (yellow) in the valine rich region of the N-terminal of Carcinin.

The hydrophobicity profile generated by the DNAMAN package predicts the putative signal peptide to be strongly hydrophobic and the rest of the sequence to be more hydrophilic (Figure 20). At the normal pH of *Carcinus maenas* haemolymph (pH = 7.5-7.8) the overall charge would be around 5 assuming up to six disulphide bonds (DNAMAN).

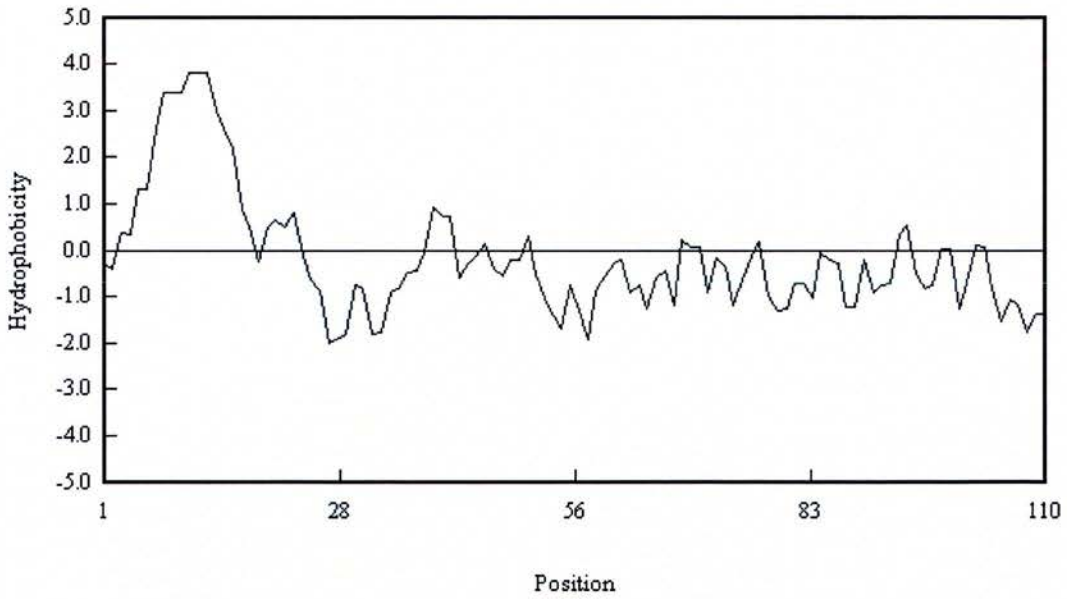


Figure 20: Hydrophobicity profile for consensus Carcinin sequence

If this signal peptide is cleaved, as commonly observed for other AMPS (Ganz and Lehrer, 1994; Ganz, 2003), by an endopeptidase (e.g. proteinase K) resulting in either an inactive proprotein or a "mature" active sequence (89 residues), then this sequence would have a molecular mass of 10.162 kDa without any PTMs. The "mature" protein

is predicted (by ProtParam) to have a similar pI (~8.74) and although maintains its cationic charge (+6), it is thought to be unstable (instability index ~ 40.03). This peptide would also have a glycine as the first residue which is a common feature of several antimicrobial peptides (Tossi *et al.*, 2000) and particularly arthropod AMPs (Dimarcq *et al.*, 1998).

Alpha (α) defensin proteins are synthesised as tripartite peptides (90-100 residues) comprising of a signal sequence at the N-terminal (~19 residues), an anionic propiece (~45 residues) and a C-terminal mature sequence (~30 residues) (Ganz, 2003). The β- and insect defensins (38-50 residues) follow a similar structure although the pro domain can be poorly conserved (Ganz and Lehrer, 1994; Dimarcq *et al.*, 1998). Although similar in number of residues, the Carcinin sequence does not appear to follow the overall α-defensin pattern and does not have an obvious acidic, anionic "pro" sequence. It does however appear to be more similar to the β- or insect defensin pattern, as well as the emerging crustacean defensin cysteine patterns (Kawabata *et al.*, 1996; Bartlett *et al.*, 2002; Supungul *et al.*, 2002). The prepropiece and the mature sections of defensins have also been suggested to be encoded by separate exons (Ganz and Lehrer, 1994) and for their overall charges to balance; this does not appear to be the case for Carcinin. It is possible to suggest that the C-terminal sequence predicted as the "mature" protein is indeed the active protein as it correlates to the amino acid sequence obtained for the purified native protein (Relf *et al.*, 1999) which exhibited antibacterial activity.

InterProScan confirmed a whey acidic protein (WAP) (first described by Ranganathan *et al.*, 1999), core region (residues 54-106, depending on the database used) based on consensus identification of a 4-disulphide core by PROSITE (E value = $8E^{-5}$), PRINTS and SMART databases (Figure 21).

C-x-(C)-[DN]-x(2)-C-x(5)-C-C

Figure 21: PROSITE consensus for the WAP-type 'four-disulfide core' domain signature

Four disulphide core proteins (Hennighausen and Sippel, 1982) usually contain 8 characteristically spaced cysteine residues which are involved in disulphide bonding but the over all sequence similarity between these proteins in low. It is also suggested that these proteins may share similar structural features as a result of their disulphide bonding.

ProDom clusters Carcinin with ESTs from *Litopenaeus vannamei* (TrEMBL entries; Q8WRN7-9) and *Litopenaeus setiferus* (TrEMBL entries; Q8WRP0-5) into a family of related proteins (PD523494) based on the alignment of the residues in the "mature" sequence (Figure 22).

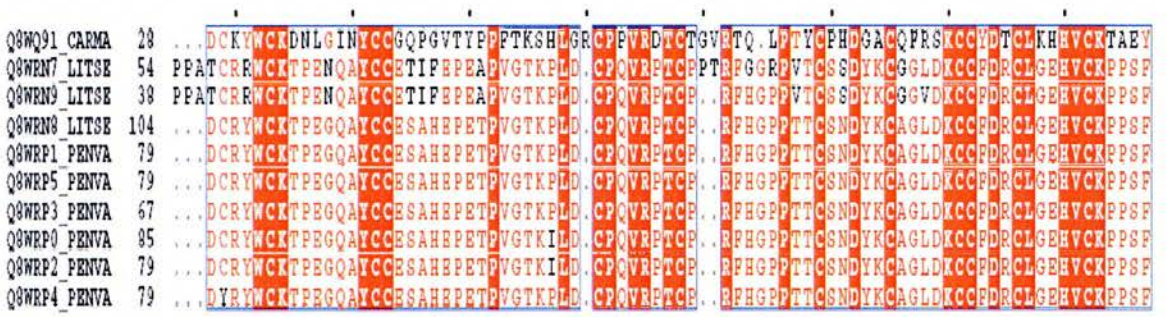


Figure 22: ProDom family alignment of crustacean antimicrobial inferred protein sequences. Residues that are 100 % conserved between the sequences are highlighted in red. Residues which are not highly conserved are indicated in black. The righthandside column indicates the TrEMBL codes for each sequence with a species distinction notation (CARMA – *Carcinus maenas*, LITSE – *Litopenaeus setiferus*, PENVA – *Penaeus vannamei*). The numbers following this refer to the number of the number of the first residue in the predicted sequence. The asterices indicate a separation along the sequence of 10 residues.

The conservation of twelve cysteines among other residues, between these species, may represent a new cysteine array among defensin proteins which is peculiar to the crustaceans (Figure 22). This pattern is conserved in transcripts with sequence identity to Carcinin from *Panulirus leniusculus* (AF522504), *Panulirus argus* (AY340636), *Marsupenaeus japonicus* (AB121740-44), *Homarus americanus* (CN853187), *Callinectes sapidus* (CV022228 & CV006490) and *P.monodon* (BIO18072-74) and most recently in *H. gammarus* (AJ 786653). All twelve cysteines are conserved in these sequences and suggest that these putative proteins may have similar 3D structures (>25 % sequence identity) (Westhead *et al.*, 2002).

The formation of disulphide bonds is one of the most likely PTMs considering the conservation of these residues in AMPS. CYSPRED analysis (Fariselli *et al.*, 1999) of the cysteines (Appendix 7; C_{29}, C_{33}, C_{42}, C_{43}, C_{61}, C_{68}, C_{79}, C_{85}, C_{91}, C_{92}, C_{96} and C_{102}) in the Carcinin sequence predicts that each is capable of forming a disulphide bond. This software uses conservation, charge, entropy and hydrophobicity parameters to achieve a consensus prediction of disulphide bonding. This result suggests, that in addition to the "disulphide core" domain involving four cysteines, a further 4 disulphide bonds may be possible but which are not in a WAP pattern. If all these cysteines are involved in disulphide bonding, Carcinin would be the first 6 disulphide bonded AMP protein. The most number of disulphide bonds to date has been found in the Tachycitin protein (73 residues) from the horseshoe crab *Tachypleus tridentatus* which has with 5 bonds (Kawabata *et al.*, 1996).

Amidation of antibacterial peptides is one of the most frequently reported post-translational modifications and peptides from terrestrial arthropods are described as "always amidated" (Tossi *et al.*, 2000). This modification has also been described in

the shrimp for the penaeidins (Destoumieux *et al.*, 1997) but not in the crustins (Bartlett *et al.*, 2002). Even so, Carcinin does not appear to have the G [RK] [RK] consensus sequence at the C-terminus which is required for this modification.

PTM	Max possible sites	Delta Mass value
Disulphide Bond Formation	**6**	**6 x (-2) = -12**
Phosphorylation	**2**	**2 x 80 = 160**
Glycosylation O-GlcNAc –β	**3**	**3 x 177 = 531**
Amidation	**1**	**1 x (-1) = -1**

Table 4: All predicted PTM and their associated delta mass values

Several sites were predicted as possible glycosylation (O-β-GlcNAc) sites on three threonines in the mature sequence (T_{49}, T_{54} and T_{69}). This modification has been suggested to be essential to the activity of some proteins (Andreu and Rivas, 1998), by influencing the mode of action. Three possible phosphorylation sites were also predicted at positions 31 for tyrosine (Y_{31}), and 54 (T_{54}) and 67 (T_{67}) for threonine, but no serine modifications were identified. T_{54} was been predicted to be both phosphorylated and glycosylated but only one of these modifications is possible at any one time and phosphorylation is thought to be less common than glycosylation (Andreu and Rivas, 1998).

Translational modification of tyrosine residues have been suggested to be essential to maintaining the activity of some AMP in some invertebrate peptides (Taylor, 2002).

Sequence	Mass
Experimentally obtained mass	**11.534 kDa**
Predicted mass of transcript from M_1	**12.226 kDa**
Predicted mass of putative active sequence	**10.162 kDa**

Table 5: Molecular masses predicted and experimentally obtained for the Carcinin protein

If all possible PTM were found in the active protein this would add 678 Da to the mass (Table 4). The experimentally obtained mass reported by Relf, J. M. (1999) of 11.534 kDa is markedly smaller than the mass predicted for the translated full transcript (12.226 kDa) by ProtParam and larger than the predicted mass of the mature protein (10.162 kDa) (Table 5). The mass added (678 Da) to the active protein by the possible predicted PTMs (Table 4) does not correlate with the experimentally obtained molecular mass of the native protein. The difference in mass could be attributed to further unidentified post-translational modifications or perhaps inclusion of some of the signal sequence.

PSI-BLAST and BLASTP analysis did not return any significant alignments (>40 % sequence identity) with other proteins from the databases searched (CDS translated, PDB, SWISSPROT, PIR & PRF released as of Jan 2004). The closest BLASTP matches ($E < 0.01$, ~33 % identity) were based on domain similarity with vertebrate whey acidic proteins and anti-leukoproteinase proteins as previously reported (Relf *et al.*, 1999). This domain was based on the identification of the 4 – disulphide core motif.

Sequence alignment using TBLASTN identified TrEMBL entries for ESTs from *Pacifastacus leniusculus*, *Litopenaeus vannamei*, *Litopenaeus setiferus* and *Penaeus monodon*. Sequence identity was only observed for these sequences from residue 25 onwards (P_{25}) of the Carcinin sequence. The sequence identities were between 32 % and 46 % depending on the length sequence with the highest identity with the *Pacifastacus leniusculus* EST sequences. Recently, Bartlett *et al.* (2002) suggested that the sequence identity at the inferred amino acid level was >40 % when comparing to the original sequence submitted by Relf *et al.* (1999). Since the full N-terminal sequence has been deduced for Carcinin the identity to the *Litopenaeus vannamei* ESTs has

dropped to 23-26%. This suggests that these sequences in the different species may have high sequence identity in the C-terminal fragment which conveys activity, but not in the signal sequence portion as described in other antimicrobial proteins (Zanetti *et al.*, 1995). Similarly, for *Litopenaeus setiferus,* the identity was initially suggested to be 40-43% comparing the C-terminal sequence but when comparing the full coding sequence this drop to 20-30%. As the Lv2 and Lv3 sequences do not exhibit the glycine repeat region identified in the other "crustin" sequences published the identity between sequences of Carcinin and *Litopenaeus setiferus* may imply a closer evolutionary link than between Carcinus and *Litopenaeus vannamei.*

Using the PredictProtein tool found at the European Bioinformatics Institute (EBI) web site (http://www.ebi.ac.uk/~rost/predictprotein/) theoretical secondary structure of Carcinin was investigated.

DNAMAN, prediction of secondary structure for the Carcinin sequence is that of a mainly random coiled structure with two possible β-sheets but no helices (Figure 23). This would suggest that this protein can be classified as a β-sheet or possibly a loop protein according to recent classification trends (Powers and Hancock, 2003). The prediction of the absence of a helix can further distinguish this type of crustacean defensin from the insect defensins (Ganz and Lehrer, 1994) and would suggest Carcinin to have greater structural similarity to the horse shoe crab defensins (Figure 5).

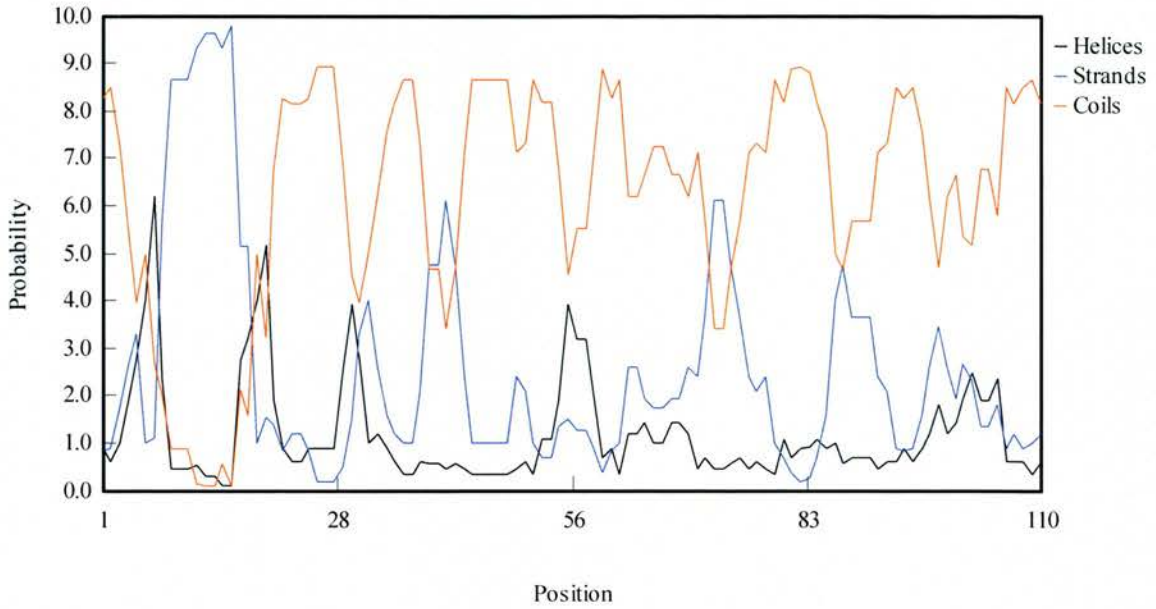


Figure 23: Secondary structure profile prediction for Carcinin by DNAMAN

More specifically, the PHDsec (Profile fed neural network systems from HeiDelberg) based on the methods of Rost and Sander (Rost and Sander, 1993, 1994) the valine rich sequence (11.8%) has a mainly looped (also known as coil or random coil) structure (79.1%) and some extended sheets (20.9%) and does not conform to either mainly an α helix or a β-sheet structure but into a "mixed" class. Again no helices were predicted in this sequence using the PHD software. The GOR4 software (Garnier *et al.*, 1996) analysis also suggests similar percentages, (Figure 24).

```
MKVQTVAAVVVVAVVVAMTEAGLFPPKDCKYWCKDNLGINYCCGQPGVTYPPFTKKHLG
Cccceeeeeeeeeeeehhhhccccccccceecccccccccceecccccccccceeeccccc
RCPAVRDTCTGVRTQLPTYCPHDGACQFRSKCCYDTCLKHHVCKTAEYPYY
ccceeeccccEeecccceecccccccceeeccceeeeccccccccccccee

Sequence length :   110
GOR4 :
   Alpha helix      Hh) :     4 is   3.64%
   3₁₀ helix        (Gg) :     0 is   0.00%
   Pi helix         (Ii) :     0 is   0.00%
   Beta bridge      (Bb) :     0 is   0.00%
   Extended strand  (Ee) :    36 is  32.73%
   Beta turn        (Tt) :     0 is   0.00%
   Bend region      (Ss) :     0 is   0.00%
   Random coil      (Cc) :    70 is  63.64%
   Ambiguous states (?)  :     0 is   0.00%
   Other states          :     0 is   0.00%
```

Figure 24: GOR4 results of secondary structure prediction

Predicted solvent accessibility of the sequence using the PHDacc software (Rost and Sander, 1994) suggests over ~70% of the residues have more than 16% of their surface area exposed. The sequence was also not predicted to be globular using the GLOBE: prediction of protein globularity software.

The Carcinin sequence and other crustacean AMP ESTs, are related to known structures based on the identification of a WAP signature. InterProScan analysis suggested that Carcinin belongs to the Elafin-like Superfamily of structures which include both 2REL and 1FLE. In the Structural Classification of Proteins (SCOP) these proteins are classified into the class of small proteins (dominated by disulphide bonds) and the fold of knottins (β hairpins with 2 adjacent disulphides).

Figure 25: 2Rel solution structure

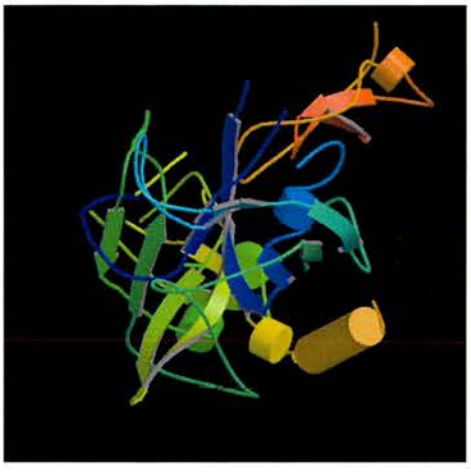


Figure 26: 1FLE (Elafin) structure

The sequences for these associated structures (1FLE and 2REL) did not achieve the 25 % identity threshold when aligned with Carcinin for use in comparative modelling although the pattern of cysteine conservation was extremely good. Known crustacean structures which have this domain include several metallothionine (e.g. 1DMC – crab and 1QJK – sea urchin), antifreeze proteins and protease inhibitor proteins (e.g. 2REL – human serine protease inhibitor) and overall 44 sequences in the Arthropoda have been identified. Pairwise alignment of Carcinin with the sequences of these known structures demonstrated that although cysteine conservation between the sequences was again quite high but none produced sequence identity over 25 %. Therefore, no suitable templates were identified for 3D modelling of the Carcinin sequence.

2.5 Discussion

In the present study the absorbance ratios which indicate the quality of the total RNA, were observed to be consistently lower (1.5 -2) than idealised values (1.8-2) (Farrell, 1996), therefore, the RNA samples were visualised by agarose gel electrophoresis. This confirmed that the samples were of good quality and showed little degradation. It was also observed that the concentrations of total RNA recovered, between samples, were highly variable and no significant correlation was observed between the total RNA concentrations and the THC. However, this was thought to be due to natural inter animal variability in THC, which has been previously described in this animal (Smith *et al.*, 2001) (a significant correlation between these factors was later confirmed in Chapter 4). Thus, in the present study, a known concentration (1μg) of total RNA was used to standardise the all first strand synthesis reactions.

As suggested by the northern blotting, the size of the full transcript sequence (AJ427538) was three times longer than that originally published (Relf *et al.*, 1999) (AJ237947). The identification of the polyadenylation (AAUAAA) (Farrell, 1996) signal at the 3' end (@ 472-477 bp; Appendix 7), the ~40 bp poly A^+ tail and the AUG translational start sites (starting @ 79 bp and 130 bp; Appendix 7) confirmed this as the full transcript. The mRNA does not appear to have an AU-rich element (ARE) sequence (AUUUA) which suggests the transcript is degraded slowly (Farrell, 1996) and this is supported by the predicted instability index (2.4.7). This issue is discussed further in Chapter 6 in light of results obtained here (Chapter 2) and in Chapter 3.

Despite the extension of the 5' UTR of the mRNA transcript and elucidation of the genomic sequence, the beginning of the first exon was not identified. It is likely that the missing upstream sequence contains the consensus arthropod transcriptional start site

(TCAGT) (Cherbas and Cherbas, 1993), as well as other regulatory elements (e.g., NF-κB binding site) (Ganz and Lehrer, 1994; Dimarcq *et al.*, 1998), although these sequences are not as well known in crustacea, as they are in other arthropods (Cho *et al.*, 1997). Future work on this gene could elucidate this upstream gDNA sequence (~1000 bp) using inverse PCR (Ochman *et al.*, 1988). This technique involves restriction digestion of the gDNA and subsequent ligation to circularise the fragments into plasmids. Primers, designed to the ends of the known gDNA sequence, can be used to amplify the sequences away from the known sequence, as a result of the circularisation (McPherson and Møller, 2000). The arthropod initiator and promoter sequences (by preparing deletion constructs) could then be identified and allow investigation of transcription control in marine arthropods. It has been suggested (Ganz, 2003), that separate exons encode the signal and mature sections of some defensin AMPs, however this is not thought to occur in the Carcinin sequence due to the position of the cleavage site (2.4.7).

The variability of the transcript sequence expressed by individual crabs was investigated by designing gene specific primers (Carc2 Fwd and Rev primers) used in amplification reactions. Every cDNA sample (from each crab) that was amplified with these Carcinin specific primers resulted in a positive PCR band of the correct size. It was therefore concluded that this transcript was constitutively expressed in the local population of this species. Further confirmation of this was established in Chapters 4 and 5 of the present study.

A compromise between sequence fidelity and efficient cloning led to the selection of Advantage ® *Taq* (BD Biosciences), over a non-proofreading *Taq* for the amplification of transcript sequences. Proofreading *Taq* is not recommended for use where the

resultant amplicons are used in cloning reactions (McPherson and Møller, 2000). This is because proof reading enzymes generally do not favour the addition of a non-template directed base (usually adenosine) that is fundamental to ligation into TOPO TA 2.1 cloning vector. In the present study, although the cloning efficiency may have been reduced, sufficient numbers of recombinant clones were produced to allow investigation of transcript variability with confidence in the sequence fidelity. Using the transcript sequences obtained in the present study in conjunction with gDNA sequence data, as well as both cDNA and protein data obtained by Relf (2000), allowed consolidation and investigation of sequence variability. Comparison of these sequences distinguished PCR or nucleotide assignment errors from the residues leading to isoform variation in this species. Of the seven residues initially identified as variable, three were thought to have arisen due to PCR or nucleotide assignment errors. The remaining four residues which appeared to vary led to 4 different isoforms (Appendix 10). Although only four distinct isoforms of Carcinin have been identified in the present study, several more may exist. To investigate further the variability in the Carcinin sequence, the frequency of expression of each isoform and to determine whether or not all transcripts are produced by all individuals a larger sequencing project could be undertaken in the future. The sequence diversity already observed in the present study may have arisen either by chance mutations or directed by evolution. Either way, it provides *C. maenas* with a repertoire of putative proteins which may all exhibit immunological defence functions. This type of diversity is increasingly being reported for transcripts with putative immunological capabilities in the crustacea (Cuthbertson *et al.*, 2002; and reviewed by Warr *et al.*, 2003; Cuthbertson *et al.*, 2004) and may reflect the animals ability to recognise to a large number of different pathogens even though these organisms are thought to have only a small number of genes.

A clear discrepancy has been highlighted in the present study, between the predicted masses for the deduced Carcinin protein sequences (putative inactive and mature sequences) and the mass which was determined experimentally (Relf *et al.*, 1999). The PTMs that were predicted by *in silico* analysis, could not account for the differences in molecular mass observed between the purified protein and the predicted masses of the full coding and putative mature sequence in Carcinin. To resolve this issue further, the native protein should be re-purified and analysed by mass spectrometry. Further analysis of the PTMs and three-dimensional structure could also be investigated using the either a purified, or a stabilised recombinantly expressed protein (see results from Chapter 3). Investigation of the number and position of possible disulphide bonds, as well as predicted glycosylation sites, could be undertaken using the methods described for use in other marine crustaceans (Muramoto and Kamiya, 1990) or a combination of cleavage, reverse phase high performance liquid chromatography (RP-HPLC) and electron ionisation mass spectrometry (EI-MS). In addition, the disulphide array could be determined using a combination of nuclear magnetic resonance (NMR) analysis and HPLC (Michaut *et al.*, 1996a). In addition, although amidation of the Carcinin protein does not seem likely, further processing of the C- terminus may make a glycine available for amidation.

Most AMPS have 2-4 disulphide bonds, however, an antibacterial protein from the Japanese horseshoe crab (*Tachypleus tridentatus),* has five, which is currently the highest number of disulphide bonds reported for AMPs (Andreu and Rivas, 1998). If, as predicted, all twelve of the cysteines are indeed bonded to each other then this would be the first six disulphide bonded antibacterial protein. In other AMPS, cysteines generally appear to be involved in disulphide bonding (Andreu and Rivas, 1998; Dimarcq *et al.*, 1998) and it is therefore thought likely that this may occur in Carcinin as

well. The results of the present study predict that the secondary structure of the Carcinin moiety consists mainly of β-sheets and probably no helical structures and it is thought that the results of the present study support the theory that Carcinin could be classified as belonging to the defensins. However, it is also hypothesised that Carcinin and related transcripts found in other species may form a new group of structures based on the conservation of the unique cysteine array.

Although sequence identity between the C-terminus sequences of Carcinin and transcripts in other crustaceans has been identified (Gross *et al.*, 2001; Bartlett *et al.*, 2002), the signal sequence are highly variable between suborders of animals (Pleoecyemata and the Denbrobranchiata). Between the crustins (Bartlett *et al.*, 2002) described from *Litopenaeus vannamei*, *Litopenaeus setiferus* and Carcinin from *C. maenas* the conservation appears to be in the C-terminus portion. However, within suborders (Denbrobranchiata) the signal sequences are more highly conserved (Supungul *et al.*, 2002). These differences may be due to autonomous module exon shuffling, as described for other arthropods (Froy and Gurevitz, 2003) and may be specific to the protein encoded or infer a functional property which have yet to be investigated.

Carcinin was previously identified to have a WAP domain (Relf *et al.*, 1999) and it was inferred that the mode of action of Carcinin may be similar to other protease inhibitors. Protease inhibition in proteins with WAP domains occurs by insertion of the inhibitory loop into the active site pocket and interference with the catalytic residues of the protease. This inhibitory loop feature, which confers protease activity to other WAP proteins, does not appear to be conserved in the Carcinin sequence based on predicted

secondary structure. This suggests that Carcinin does not function in a similar manner to these other protease inhibitor proteins.

Comparative modelling, using known resolved 3D structures, was unsuccessful as no known structures exhibited >25 % sequence identity (Westhead *et al.*, 2002). It could though be suggested that putative proteins identified in *Pacifastacus leniusculus*, *Penaeus monodon, Litopenaeus setiferus* and *Litopenaeus vannamei* among others, will all probably have a similar structure and they have >25% sequence identity and have all the cysteines conserved. Future investigations of the protein should include elucidation of the tertiary structure which will also help to identify the active portions (Yang *et al.*, 2000; Yang *et al.*, 2003).

In conclusion, although northern blotting was useful in determining the full size of the Carcinin transcript, it was not thought to be a suitable method for use in future expression analysis (Chapter 4). The elucidation of the full coding sequence and putative isoforms will enable the informed and accurate design of expression analysis probes (Chapter 4). Constitutive expression of this transcript has been confirmed and information on isoform expression in different individuals has been investigated. Using the consensus protein sequence, the analysis has suggested possible posttranslational processing, including the location of a cleavage site, the likelihood of disulphide bond formation and an indication of the stability of the resultant mature proteins. This information will be useful in supporting the decisions and results obtained in subsequent recombinant expression analysis (Chapter 3). Finally, the analysis lends further support to the premise that Carcinin belongs to the "defensin" family of proteins and that it may represent a new family of structures in this group. This chapter provides supporting information for the subsequent expression studies of Carcinin (Chapters 3-5).

2.6 Conclusions

- Carcinin transcript is >642 bp and codes for a 110 residue protein which is valine rich.
- The transcript is produced from splicing together four exons from a ~1kb gDNA sequence
- A cleavage site exists after a putative valine rich signal sequence resulting in a peptide with a glycine as the first residue.
- Several isoforms of the protein exist between individuals.
- High sequence identity exists with putative antibacterial ESTs in at least four other species of Crustacea which may form a single group of the first 6 disulphide bonded antibacterial peptides.
- No sequences of known structures were >25 % identical to the Carcinin sequence to allow comparative modelling.

CHAPTER 3

Recombinant Carcinin Expression

3.1 Introduction

There are several methods by which proteins can be obtained in sufficient quantities to allow the investigation of their functional and structural characteristics as well as the development of immunochemical "tools" such as antibodies. These methods include, purifying the native protein from source, synthetic production and recombinant expression. The choice of method depends on the subsequent use of the protein and the availability of information on the structure and function of the native protein.

For initial investigations into activity, characterisation and interactions with other proteins, the purification of the native molecule is often the first choice, although this can be resource limited (Bradley, 1990). Protein yields can be low, as substrate is lost at each purification stage, accompanied by increased possibility of proteolysis (Linn, 1990). However, purification does usually ensure that efficacy and functionality are maintained and the structure of the native protein can be determined (Lodish *et al.*, 1995) using a variety of techniques (e.g. nuclear magnetic resonance (NMR), Edman degradation and circular dichroism (CD) (Deutscher, 1990).

Synthetic production of full-length proteins is often too costly for initial studies and most small-scale applications, although it can be can be time efficient and affordable for small peptides or active fragments (Hara and Yamakawa, 1996). This approach is often used for proteins that are difficult to express *in vitro* or where large homogenous quantities are required (Raj *et al.*, 2000a; Cudic *et al.*, 2002). Manipulation of synthetic sequences is essential in structure based efficacy optimisation studies (Johansson *et al.*, 1998; Raj *et al.*, 2000b; Yang *et al.*, 2000; Yang *et al.*, 2002; Romestand *et al.*, 2003). Synthetic peptides/fragments are often used in antibody production (Lodish *et al.*, 1995)

where only a small portion of the protein may be necessary, although information about the antigenicity of the native protein may be vital to ensure their efficacy.

Recombinant expression is often used in preliminary studies to obtain product quickly and cheaply (Bradley, 1990; Zhang *et al.*, 1998). The advantages of recombinant expression in terms of simplicity and convenience have been widely recognised and used for the production of a variety of proteins (Bradley, 1990; Barrell *et al.*, 2004). Recombinant expression also gives the researcher control over several aspects of the production, including the sequence design, the purification method and the scale of the production. Conversely, using non-native cells as hosts may have limitations due to the importance of the post-translational modifications and possible toxicity issues. A summary of the main considerations in selecting a protein production method is outlined in Table 6.

Method	Most Appropriate Uses	Advantages	Disadvantages
Native Purification	**Functional efficacy** **Structural characterisation** **Antibody production**	**Efficacy maintained** **Native protein**	**Laborious** **Loss of material** **Resource limited**
Recombinant Production	**Functional efficacy** **Antibody production**	**Relatively inexpensive** **Control in design and production**	**Purity compromised** **Post Translational Modifications (PTM) may be compromised**
Synthetic Production	**Antibody production**	**Quick, purity** **Homogenous sample**	**Expensive** **Peptides only** **Efficacy may be compromised**

Table 6: Comparison of protein production methods

3.1.1 Recombinant Technology

Recombinant technology exploits the properties of naturally occurring enzymes to modify DNA sequences by restriction and ligation into vectors. This allows maximal expression of a protein of interest by optimal arrangement of genetic data. The use of recombinant protein expression has increased greatly in recent years, as have the advancements in the techniques routinely employed (Amersham Pharmacia Biotech, 2001). It is used mainly for the study of structural and functional properties of proteins (Choi *et al.*, 2002) and is becoming an essential part of drug discovery projects (Grandi, 2003). The method is also compatible with the systematic processing of biomolecules and optimisation of production methodologies (Dian *et al.*, 2002a). Genetic information used to produce these proteins can be naturally occurring or engineered chimeric sequences for production on a large scale with a specific function or efficacy (Dathe *et al.*, 2001; Muhle and Tam, 2001). The production of recombinant molecules relies greatly on the accuracy of the sequence and efficacy data gathered for the native protein; it allows verification that the expressed protein is representative and of comparable efficacy to the native protein.

Recombinant technology can be used with cell-free expression systems (Martemyanov *et al.*, 1997) as well as with a variety of host cell lines, including selected mammalian, insect, yeast, bacteria, fungi and plant cells (Table 7). The choice of system used, affects both amplification and purification methods and consideration must be given to the degree of purity, biological integrity and the possible toxicity to the host.

To complement these hosts, several vector options exist for each, allowing a degree of flexibility and control over the system (Amersham Pharmacia Biotech, 2001). There are also vectors specifically designed for initial studies, which can be moved from one host

type to another with relative ease, for example, the Gateway ™ Cloning Technology (Invitrogen Ltd., Paisley, UK).

Host	Advantages	Disadvantages
Bacteria	**Expression controllable** **Many vectors** **Historical data** **Easy to grow** **Secretion possible** **Large scale**	**No post translational modification** **Possible differences in immunogenicity and activity** **High endotoxin content in gram negatives** **Some strains have low expression ability** **Pathogenic**
Mammalian Cells	**Biological activity maintained** **Many vectors** **Large scale possible**	**Difficult and expensive** **Slow growth** **Genetic instability** **Low productivity**
Yeasts	**Low endotoxins** **Generally Regarded As Safe (GRAS)** **Fermentation relatively inexpensive** **Glycosylation and disulphide bond formation possible** **Large scale possible**	**Expression harder to control** **Glycosylation not identical to mammalian**
Insect	**Glycosylation and disulphide bond formation possible** **GRAS**	**Glycosylation not well characterised** **Not always fully functional** **Differences in function and antigenicity**
Fungi	**Established fermentation** **Inexpensive** **Some GRAS** **Mass secretion possible**	**High levels of expression not yet possible** **Genetics not well characterised** **No vectors available**
Plants		**Low transformation efficacy**

Table 7: Comparison of host choices for use in recombinant expression of proteins (Amersham Pharmacia Biotech, 2001)

3.1.1.1 RECOMBINANT BACTERIAL EXPRESSION

Bacterial expression systems, usually using *Escherichia coli (E.coli)* as a host, are often chosen for their ease of manipulation, propagation, storage and accessibility to the

researcher. *E.coli* systems have been well studied and extensively documented; they are often the initial system utilised in any expression study. A variety of vectors are available with selectable markers, controllable promoters and multiple cloning regions (MCR) engineered with several restriction sites. The choice of vector-host system selected for bacterial expression is influenced by the size of the protein, the amount to be produced and whether maintaining activity is essential for further experimentation. Although some expression systems can produce soluble active proteins, many are produced in an insoluble form and have to be purified from inclusion bodies and then refolded into their active form (Sambrook and Russell, 2001). In general, the more that is known about the protein of interest in terms of structural, chemical and biological function the easier it is to select the appropriate expression system which allows the production of maximum yield with speed and ease.

3.1.1.1.1 Fusion Protein Bacterial Expression Systems

A fusion protein expression system allows the isolation and purification steps to be concluded by exploiting the well-characterised properties of the protein tag, by affinity chromatography. Fusion genes are engineered by joining together two or more open reading frames (ORFs), generating a hybrid, where the protein of interest is attached to a carrier "tag" (Table 8) which has little effect on the conformational and functional properties of the protein of interest. The tag can then be easily removed and the protein of interest purified to homogeneity using standard chromatographic methods. For the production of small polypeptides, cytosolic proteins (<100 residues) and novel proteins, expression as a fusion protein is preferred as the carrier can often stabilise the protein of interest against degradation and prevent the formation of inclusion bodies; in some cases it can also direct the protein of interest to a specific cellular location (Sambrook and Russell, 2001).

Advantages	Disadvantages
Fusion Proteins	
Targeting information can be incorporated **Provides marker for expression** **Affinity purification** **Easy detection** **Refolding possible** **Better for <100aa proteins** **Ideal for secreted proteins**	**Tag can interfere with folding and activity** **Cleavage site not always specific**
Non-Fusion Proteins	
No cleavage necessary	**Purification and detection not as simple** **Solubility problems possible**

Table 8: Summary of the advantages and disadvantages of fusion and non fusion vectors (Amersham Pharmacia Biotech, 2001)

More than 20 fusion protein systems exist using both natural ligand-binding proteins such as glutathione-S-transferase (GST) and *Staphylococcus aureus* protein A (Zhang *et al.*, 1998), as well as novel artificial tags such as the polyhistidine (His6) and the FLAG® system from Sigma-Aldrich (Sigma-Aldrich Company Ltd. Dorset, UK). Other widely used systems include, the IMAC (immobilised metal ion affinity chromatography) system, the biotin/streptavidin systems from Sigma-Genosys (Sigma-Genosys Ltd., Suffolk, UK) and Promega (Promega, Southampton, UK) respectively, the IMPACT® (Intein-Mediated Purification with an Affinity Chitin-binding Tag) system from New England Biolabs (New England Biolabs (UK) Ltd., Hertfordshire, UK) and Novagen's T7 tag ® purification system (EMD Biosciences Inc., Wisconsin, USA) (Constans, 2002). Unfortunately, the removal of the affinity tag can lead to loss of target activity which can sometimes be recovered by refolding, although additional amino acids may remain attached to the target.

3.1.1.1.1.1 Glutathione S-transferase (GST) Fusion System

The GST fusion proteins are some of the most studied and commonly used fusion tag systems and are purified by affinity chromatography using glutathione coupled to a Sepharose 4B matrix (Figure 27) by epoxy activation (Amersham Pharmacia Biotech, 2001). The GST fusion system has often been chosen over the His system as it is very well documented, although in recent years the use of His tags has eclipsed that of GST with several of the major fusion vector suppliers promoting its superior ease of use. However, there have been reports of problems associated with the use of His tags with respect to solubility and possible instability (Constans, 2002).

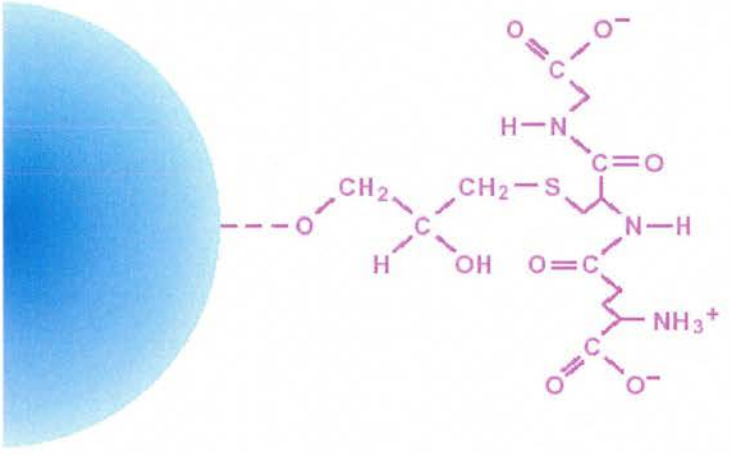


Figure 27: Glutathione coupled to Sepharose 4B matrix (Amersham Pharmacia Biotech, 2001)

The GST gene fusion system was chosen in the present study for its reported ease of manipulation, speed of purification and affordability. It has also been purported to be easily scaled up to accommodate high through put (HTP) methodologies (Dian *et al.*, 2002b). Amersham Biosciences produce several fusion vectors (pGEX) as well as a broad range of products for the subsequent amplification, purification and detection of GST fusion proteins in *E.coli*. These vectors provide chemically inducible, high level expression with mild elution conditions to release the fusion protein from affinity media

in order to reduce effects on antigenicity and functionality of the target protein (Amersham Pharmacia Biotech, 2001). They have been reported to favour the accumulation of fusions as soluble proteins although this is not always the case (Skosyrev *et al.*, 2003). This system relies on the correct folding of the GST moiety and its large size can hinder the solubility of certain proteins leading to the formation of inclusion bodies (Frangioni and Neel, 1993; Constans, 2002; Dian *et al.*, 2002a; Mercado-Pimentel *et al.*, 2002). Due to these factors, the tag is removed by enzymatic cleavage of a site between the tag and the protein of interest. A variety of these vectors are available with different protease recognition sites (Thrombin, Factor Xa and PreScission ™ protease) and restriction enzyme sites.

Expression in pGEX vectors is controlled by a hybrid *trp-lac* promoter (*tac*) (Sambrook and Russell, 2001) (Figure 28). The vectors are designed incorporating a *lac*I^q gene whose product represses the *tac* promoter by binding to its operator region which can be activated by the addition of IPTG, a lactose analogue.

For routine maintenance and propagation, the host bacteria recommended with this vector for cloning and full-length expression are *E.coli* BL21 (DE3) and *E.coli* JM105. Protease deficient strains are recommended for expression to avoid proteolysis of the product by the host. The inserted DNA sequence must have an open reading frame (ORF) less than 2 kb in length and have compatible ends with the linearised vector ends using two different restriction enzymes to facilitate directional cloning.

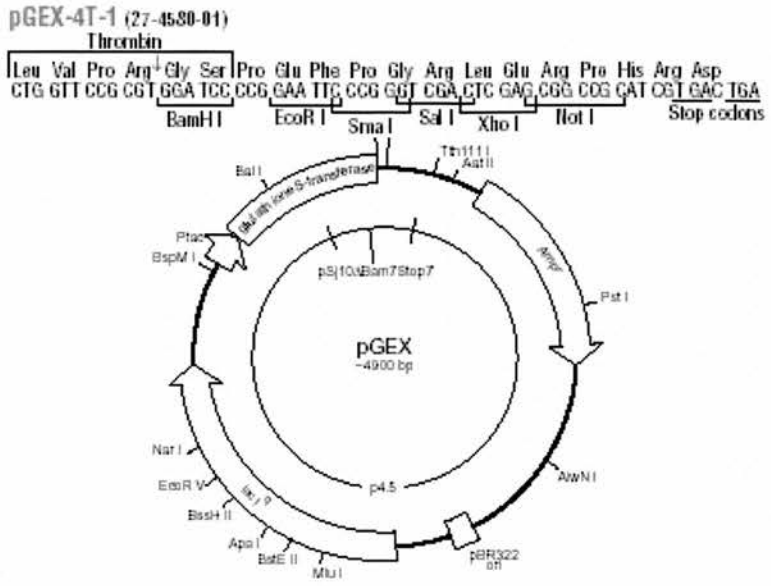


Figure 28: pGEX-4T-1 vector showing detail of multiple cloning site (Amersham Pharmacia Biotech, 2001)

Finally, these vectors have already been successfully used in the production of immunogens for subsequent study of protein expression (Mercado-Pimentel *et al.*, 2002) and were therefore considered to be suitable for use in the present study

3.1.1.2 RECOMBINANT EXPRESSION OF ANTIMICROBIAL PEPTIDES

Precedents do exist for the recombinant expression of antimicrobial peptides (AMP), although the majority of the work has been conducted on mammalian and insect derived molecules.

Recombinant antimicrobial peptides have been produced using several different hosts including: cell-free (Martemyanov *et al.*, 1997), bacterial (Piers *et al.*, 1993; Hara and Yamakawa, 1996; Zhang *et al.*, 1998; Choi *et al.*, 2002; Skosyrev *et al.*, 2003; Barrell *et al.*, 2004), baculovirus (Andersons *et al.*, 1991; Hellers *et al.*, 1991; Kim *et al.*, 1998), transgenic tobacco (Florack *et al.*, 1995; Choi *et al.*, 2003) and mammalian Chinese hamster lung cells lines (Pore and Pal, 2000). Many of these proteins were produced as

insoluble inclusion bodies, although this has been described as an advantageous result in terms of isolation (Zhang *et al.*, 1998). Of the bacterially produced proteins most were produced as fusion proteins using a range of different tags including GST, GFP, His6 and Protein A (Piers *et al.*, 1993; Zhang *et al.*, 1998; Skosyrev *et al.*, 2003; Barrell *et al.*, 2004). The GST fusion system has been used to express sarcotoxin IA, from the flesh fly *Sarcophaga peregrina* (Skosyrev *et al.*, 2003) as well as the human defensin, HNP1 (human neutrophil peptide 1) as well as cecropins/melittin hybrid proteins (Piers *et al.*, 1993). The expression of these proteins have been successful, partly, because some targets are derived from monoexonic transcripts and others do not require post-translational modification (e.g. disulphide bond formation) to maintain the activity of the resultant protein (Hara and Yamakawa, 1996; Kim *et al.*, 1998; Skosyrev *et al.*, 2003; Barrell *et al.*, 2004). However, some disulphide bonded defensin peptides from insects and mammals have been recombinantly expressed (Lepage *et al.*, 1991; Michaut *et al.*, 1996b; Porter *et al.*, 1997), and have been reviewed by Andreu *et al.*(1998).

To date the only reported recombinant expression of full length AMPs (Penaeidin 2 and 3a) from a crustacean (*Litopenaeus vannamei)* was achieved using the yeast *Saccharomyces cerevisiae*. This peptide falls into both proline rich and cyclical AMP categories (as reviewed by Hetru, *et al* (1998)). The recombinant peptides were expressed and purified to homogeneity by high performance liquid chromatography (HPLC) and were shown to exhibit comparable efficacy and activity spectrum to the native peptides. Some post-translational processing differences were observed between the native and the recombinant peptides but were thought to contribute to the stability rather than activity of these molecules. It was also suggested that the cysteine-rich C-terminal domain was central to the peptide's antimicrobial activity against mainly Gram-positive bacteria and filamentous fungi (Destoumieux *et al.*, 1999). This

recombinant protein was used to produce polyclonal antibodies for use in immunohistochemical studies (Destoumieux *et al.*, 1999; Destoumieux *et al.*, 2000) and the recombinant protein was also subsequently used to determine its tertiary structure (Yang *et al.*, 2003).

The primary aim of the present study was to produce sufficient recombinant Carcinin to facilitate the production of antibodies to this protein. Therefore, the post-translational modifications afforded by the more expensive and complex yeast expression system were not considered necessary and a bacterial expression system was chosen instead. In light of successful precedent studies (Piers *et al.*, 1993), a fusion bacterial expression system was chosen as earlier *in silico* analysis (Chapter 2) suggested that Carcinin may not be particularly stable.

Several antimicrobial peptides, including the defensins (Ganz and Lehrer, 1994; Ganz, 2003), have been shown to be translated as a prepropeptides which are later cleaved by an endopeptidase to release an active fragment from the signal sequence (Valore *et al.*, 1996; Kim *et al.*, 1998; Yiallouros *et al.*, 2002; Satchell *et al.*, 2003). The *in silico* analysis of Carcinin (Chapter 2) suggested that this protein is probably a defensin which is produced and activated in a similar fashion (2.4.7). Therefore, in the present study, both the pro- form and the putative mature or "active" sequence were cloned and expressed as fusion proteins. The present study is the first to describe the expression of a crustacean antimicrobial protein using a bacterial expression system.

3.2 Specific Aims

The aims of this chapter were to over express the antibacterial protein Carcinin, from the shore crab *C. maenas,* using a recombinant bacterial GST fusion system, for the subsequent production polyclonal antibodies in rabbits.

By expressing two different sequences of this protein, the full coding sequence and the "mature" sequence, initial investigation into the stabilising effect of the putative "signal" sequence on the "active" protein would be possible.

3.3 Methods and Materials

A GST fusion vector (pGEX) from Amersham Biosciences (Amersham Pharmacia Biotech Inc., New Jersey, USA) was purchased for recombinant expression of Carcinin.

This study was broken down into two main parts; the first part describes the construction of the expression vector (construct) and the second, the expression and purification of the fusion protein. The following figure outlines the main steps undertaken in the initial development of the construct (Figure 29).

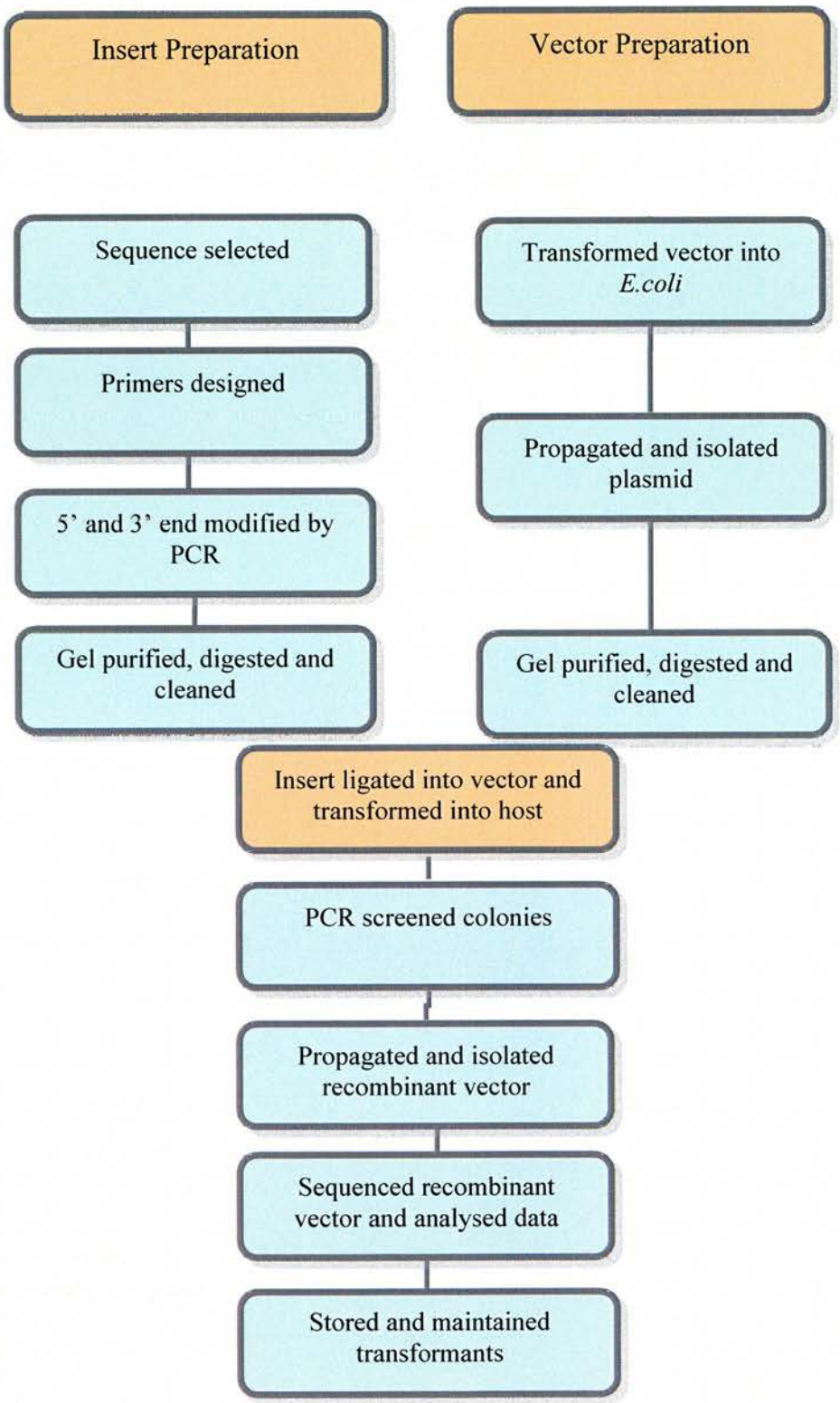


Figure 29: Summary of the steps involved in the recombinant construct production

3.3.1 Host Propagation and Storage

The recommended protease deficient (*Omp*T and *Lon*) bacterial strain, used for high level expression of fusion proteins was *E.coli* BL21 (Amersham Pharmacia Biosciences, Buckinghamshire, UK) (Appendix 13). *E.coli* BL21 is not recommended for routine maintenance and storage of vectors as it does not transform well and an alternative strain, *E.coli* JM105 (Amersham Pharmacia Biosciences, Buckinghamshire, UK) was used for this purpose (Appendix 13). It had been previously shown that Carcinin was not toxic to *E.coli* (Relf *et al.*, 1999) and was therefore a suitable expression host.

All chemicals used in the following methods were purchased from Sigma (Sigma-Aldrich Company Ltd., Dorset, UK) unless otherwise stated. Lyophilised *E.coli* BL21 and JM105 were resuspended in 1 ml of sterile Luria-Bertani (LB) broth (Fluka-Sigma-Aldrich, Dorset, UK) and grown overnight at 37 °C before plating out onto Luria-Bertani (LB) agar (ICN Pharmaceuticals, Hampshire, UK). The overnight culture plates were then stored until required at 4 °C.

A single colony of each strain was grown overnight at 37 °C in 10 ml of LB broth, streaked onto pre-warmed LB agar plates which were incubated at 37 °C for a further 16 h. Glycerol stocks were also prepared for both *E.coli* BL21 and JM105, by mixing 1 ml of overnight culture with 1 ml sterile, autoclaved, 80 % glycerol and frozen in 5 ml Bijou bottles at –80 °C. Frozen cultures were revived by streaking onto warmed LB agar plates, when required, using aseptic technique.

3.3.1.1 COMPETENT CELL PREPARATION

Competent cells of *E.coli* BL21 and *E.coli* JM105 (Amersham Biosciences) (Appendix 13) were prepared, using aseptic technique, from single colonies, grown overnight in 15 ml LB broth at 37 °C, in a shaking (~50 rpm) water bath (Sanyo Gallenkamp, Loughborough, UK) using 50 ml Falcon ® tubes (Fisher Scientific, Leicestershire, UK). Each bacterial strain was then sub cultured by transferring a 100 µl aliquot to 15 ml of fresh pre-warmed medium (LB) and grown for a further 3-6 h with shaking (~50 rpm). When the OD (A_{600}) reached ~0.4-0.5, as measured on an Ultrospec 3300 spectrophotometer (Amersham Pharmacia Biosciences, Buckinghamshire, UK), the cells were sedimented by centrifugation (2500 x g, 15min @ 5 °C). The supernatant was discarded and the cell pellets were then gently resuspended in 1.5 ml ice cold "transfer and storage solution" (TSS) (Appendix 14) and kept on ice for a maximum of 3 h in 1 ml aliquots in 50 ml Falcon ® tubes to be used in transformation of expression vectors and constructs.

3.3.2 Expression Vector – pGEX4T-1

The pGEX-4T-1 vector (Amersham Pharmacia Biosciences, Buckinghamshire, UK) (Figure 28), was selected for its open reading frame (ORF), choice of restriction enzyme sites and cleavage enzyme recognition sequence (thrombin) (Appendix 15). Carcinin was analysed for this recognition sequence using the restriction digest facility of the DNAMAN program version 5.2.0 (Lynnon BioSoft, Quebec Canada).

3.3.2.1 VECTOR TRANSFORMATION

Transformation of vectors into *E.coli* were achieved by chemical methods based on the procedures described by Chung *et al.*(1989). To each 1 ml aliquot of resuspended

competent cells (3.3.1.1), 20 µl of uncut pGEX–4T-1 (1 ng) vector were added, mixed gently and stored on ice for a further 45 min. For the negative control, 20 µl of sterile distilled water replaced the vector. The mixtures were then incubated exactly for 2 min at 42 °C in a circulating water bath (Grant) without mixing after which they were immediately returned to the ice. In separate 15 ml centrifuge tubes (Falcon ®), 900µl of LBG broth (Appendix 14), pre-warmed to 37 °C was inoculated with 100µl of the resultant transformations and incubated at 37 °C in a rocking hybridisation oven (Thermo-Hybaid Ltd, Middlesex, UK) for 1-2 h, until further growth could be observed, as judged by increased turbidity. Next, aliquots of 100 µl and 10 µl (diluted in 30 µl of LBAG medium) of the transformed cultures, were spread onto two separate pre-warmed LBAG (Appendix 14) plates, in duplicate, alongside 100 µl of untransformed competent cells as a negative control. Similarly, 100 µl of transformed cells (with un manipulated plasmid) were also plated onto LBG plates as a positive control. All plates were incubated overnight at 37 °C and single transformed colonies (growth on LBAG agar plates) were grown in LBAG broth and glycerol stocks were prepared as described in section 3.3.1.

Transformed bacterial strains containing the unmodified pGEX-4T-1 were used to propagate the expression vector and act as a positive control in subsequent expression experiments for comparison of recombinant protein expression efficiency.

3.3.2.2 VECTOR PROPAGATION AND ISOLATION

Transformed recombinant vectors were isolated using the Wizard® SV Minipreps DNA Purification System (Promega), as detailed by the manufacturers (Appendix 6). This protocol is based on the methods of Birnboim and Doly (1979) which relies on the selective alkaline denaturation of the gDNA and this leaves the circular plasmid DNA

intact. Isolated pGEX-4T-1 plasmids were quantified (A_{260}) using an Ultrospec 3300 spectrophotometer (Amersham Pharmacia Biotech) and agarose gel electrophoresis as described below. Initially, isolated plasmid samples were also linearised by digestion with BamH1 enzyme (Promega) as described below in section 3.3.3.3 and the resultant products were analysed by agarose gel electrophoresis alongside supercoiled plasmids.

The plasmid size and purity, as well as concentration, was determined by running a 1 µl sample, mixed with 1 µl 6X DNA loading buffer (100 mM EDTA (VWR International, Dorset, UK); 25 mM Tris-HCl; 25 % glycerol; 0.05 % bromophenol blue (BDH); pH 7.0) (Sambrook and Russell, 2001) and 4 µl of water. These samples were loaded into a 0.7 % agarose gel (BDH) / 0.5M TBE (45 mM Tris-borate; 1 mM EDTA; pH 8.0) (Sambrook and Russell, 2001), containing 1 µg ml^{-1} ethidium bromide with 0.5X TBE as a running buffer. The gel was run at 70V for approximately 30 min until the dye front had migrated across most of the length of the gel. The resultant gel was viewed using a UV transilluminator and the sizes of the plasmids were compared to molecular markers of known size (Lambda DNA *Hin*d III digestion and 1Kb DNA marker, both from Promega, Southampton, UK).

3.3.3 Insert Sequence Manipulation

The full cDNA (AJ427538) and genomic sequences for Carcinin were previously determined as described in Chapter 2. Two distinct coding sequences were selected for subsequent expression as fusion proteins. The first sequence comprised of the full coding region from the first methionine (M_1) codon to the first stop (TAG) codon (Appendix 7) and the second sequence from the glycine (G_{22}) codon, after the putative TEA↓GL cleavage site (Chapter 2), to the first stop (TAG) codon (Appendix 7).

3.3.3.1 PRIMER DESIGN AND INSERT MODIFICATION

To insert the selected sequences into the pGEX-4T-1 vector the 5' and 3' ends of the inserts were modified by polymerase chain reaction (PCR) to make them compatible with the linearised vector restriction site ends. The restriction sites selected for directional cloning were BamH1 (G/GATCC) and Xho1 (C/TCGAG) (Figure 28). Using the restriction analysis facility of the DNAMAN software the Carcinin inserts were analysed for these particular restriction sites.

Primers were designed to the 5' and 3'ends of both selected sequences described above (3.3.3) to add a BamH1 site to the 5'end and a Xho1 at the 3' end to ensure correctly orientated directional cloning (Figure 30).

All primers were synthesised by MWG (MWG Biotech A.G., Ebersberg, Germany) unless otherwise stated. The primers were designed to include a 4 bp GC clamp (GCGC), then the restriction site (BamH1 or Xho1) and then the Carcinin specific sequence at the 3' end. The GC clamp ensured efficient restriction digestion of the resultant PCR products as both BamH1 and Xho1 require 3 bp on either side of the restriction site for effective recognition (New England Biolabs, 2004).

Both resultant FCAR ("full" sequences with added restriction sites and GC clamp) and ACAR ("active" with added restriction sites and GC clamp) PCR products were smaller than the maximum insert size recommended for pGEX vectors (2kb) (Amersham Pharmacia Biotech, 2001) each with in-frame sequence and a stop codon at the 3' end.

FCARFWD 5' – **GCGCG↓GATCCATGAAGGTGCAAA** – 3'

ACARFWD 5' – **GCGCG↓GATCCAGGTTATTCCCTC**– 3'

FCARREV and ACARREV 5' – **GCGCC↓TCGAGTTACACGGGTCTG** – 3'

GGATCC – BamH1 restriction enzyme site

CTCGAG – Xho1 restriction enzyme site

GC – GC Clamp

XXXX- Carcinin specific Sequence

Figure 30: Primers designed for FCAR and ACAR fragments incorporating BamH1 and Xho1 sites at 5' and 3' ends respectively

Messenger ribonucleic acid (mRNA), isolated from a single healthy intermoult male crab and reverse transcribed to cDNA by first strand synthesis as described in Chapter 2. The cDNA was diluted 1 in 20 with nuclease-free water. PCR reactions of 25 µl were prepared in thin walled PCR tubes (Axygen, California, USA) using Advantage ™ cDNA polymerase mix (BD Biosciences, Oxfordshire, UK) and. The PCR reaction was set up a follows using the primers detailed in Figure 30: 0.25 µl of Advantage™ polymerase (5 units μl^{-1}), 2.50 µl 10x Buffer (500 mM KCl; 15 mM $MgCl_2$; 10 0mM Tris HCl), 1.00 µl FWD Primer (10 mM), 1.00 µl REV Primer (10 mM), 0.50 µl dNTPs (100mM) (Promega), 1.00 µl cDNA, and RNAse free H_2O to 25 µl. The cycling conditions used were as follows using a thermal cycler (Thermo-Hybaid Ltd, Middlesex, UK): 94 °C for 7 min, 30 cycles of 94 °C for 30 s, 55 °C for 30 s, 72 °C for 2 min and one cycle at 72 °C for 2 min.

3.3.3.2 PCR ANALYSIS

The resultant PCR products (FCAR and ACAR) were run out on a 2 % agarose gel as described for the 0.7 % gel in 3.3.2.2. The gel was run using a BioRad Gel Kit at 70V until the dye front had migrated considerably across the gel. The gel was viewed using a UV transilluminator and the size of the PCR products compared to molecular markers of known size and mass (100 bp DNA marker; New England Biolabs UK (Ltd.), Hertfordshire, UK).

The appropriate size band was then cut out with a clean scalpel and placed in a 1.5 ml centrifuge tube. Using the QIAQuick ® Gel Extraction Kit (Qiagen Ltd, West Sussex, UK) (Appendix 4) the resultant PCR products were isolated from the agarose slice and eluted with 30 µl autoclaved, Elga-MilliQ water (Veolia Water Systems, High Wycombe, UK). The elutants were quantified by absorbance $A_{260/280}$ and running a 2 % agarose gel as described above.

3.3.3.3 RESTRICTION DIGEST

The restriction sites on pGEX-4T-1 were at positions 954 (BamH1) and 930 (Xho1), (Appendix 16). In brief; vector DNA (~1000 ng) or the modified insert DNA (~200 ng) (3.3.3.2), 2 µl of 10 x Buffer D (6 mM HCl, 6 mM $MgCl_2$, 150 mM NaCl, 1 mM DTT, pH 7.9: supplied), 2 µl of bovine serum albumin (BSA) [1 mg ml^{-1}], 1.5 µl of BamH1 and Xho1 [both 10 units μl^{-1}] and distilled RNAse free water to 20 µl, were incubated at 37 °C for 1 h in thin walled PCR tubes. Although restriction digest of PCR products without product purification has been shown to be effective, it has not been recommended for use with subsequent cloning procedures (Turbett and Sellner, 1996). The enzymatic cleavage was halted by incubation at 65 °C for 10 min.

Reactions were then cleaned up using the QIAquick ® PCR Purification Protocol (Qiagen Ltd, West Sussex, UK) to remove excess nucleotides (Appendix 4). The elutant was quantified by gel electrophoresis as described above (3.3.3.2).

Initially, for the restriction (double) digest, the time required to complete the cleavage was optimised by taking small (2 µl) samples from the reaction at 2, 3, 6 and 16 h and running them on a 2 % agarose gel as described in section 3.3.3.2.

3.3.4 pGEX-4T-1F and pGEX4T-1A Recombinant Vector Construction

Ligation of the linearised vector and digested insert, using the T4 DNA Ligase enzyme from New England Biolabs (New England Biolabs (UK) Ltd., Hertfordshire, UK), was performed immediately after the digest reactions were cleaned up and quantified to avoid spontaneous re-annealing of the plasmid to itself (digests were stored on ice during quantification procedures).

The concentration of insert used in the ligation reaction was calculated using the following equation based on the amount of vector used.

$$\frac{\text{ng of vector} \times \text{kb of insert}}{\text{kb of vector}} \times \frac{5}{1} = \text{ng of insert}$$

e.g.

$$\frac{100 \times 0.372}{4.9} \times \frac{5}{1} = 37.95 \text{ ng insert}$$

Figure 31: Vector to insert calculation using base pair values from section Appendix 17

The reactions were performed with 100-500 ng of vector and 37.95-189.79 ng of insert (Figure 31), depending on the yield of the product after QIAQuick ® PCR Purification (Appendix 4). Initially, the source plasmid was used to create the initial constructs in

order to avoid mutations occurring in the sequence *in vivo*. The ligation reaction (20 µl) was prepared as per the manufacturers instructions with the supplied reagents (vector and insert DNA [~1 µM 5' termini each], 2 µl of 10 x buffer [50 mM Tris-HCl, 10 mM $MgCl_2$, 10 mM DTT, 1 mM ATP, 25 µg ml-1 BSA, pH 7.5 @ 25 °C], 1 µl of T4 DNA Ligase [400 units μl^{-1}] for 10 min at RT & reaction deactivated by incubating reaction at 65 °C for 10 min). The construct created using the FCAR insert was called pGEX4T-1F and the plasmid with the ACAR insert was called the pGEX4T-1A construct.

3.3.4.1 LIGATION CONFIRMATION

To check the success of the ligation reactions described above, 1µl of each reaction was used in PCR with pGEX sequencing primers (Figure 32: P4TF and P4TR; primer data sourced from Amersham Bioscience's pGEX 5' and pGEX3' sequencing primers). The 5' primer binds to the pGEX 4T-1 vector (Accession number: U13853) between bases 869-891 and the 3' primer between bases 1041-1019 (Appendix 18).

P4TF	5' – GGG CTG GCA AGC CAC GTT TGG TG – 3'
P4TR	5' – CCG GGA GCT GCA TGT GTC AGA GG – 3'

Figure 32: Sequencing primers for the pGEX4T-1 vector (Amersham Pharmacia Biotech, 2001)

The PCR reaction was prepared as follows; 1 µl DNA, 1 µl (10 pmol) of each P4TF and P4TR primers, 0.25 µl (4 units) *Taq* DNA polymerase (Amersham Biosciences UK Ltd.), 0.5 µl (100 mM) dNTPs (Promega), 10 x buffer (supplied with enzyme) and water to 25 µl. The cycling conditions were, 94 °C for 2 min, 30 cycles of 94 °C for 30 s, 55 °C for 30 s and 72 °C for 30 s, and finally one cycle at 72 °C for 2 min. The samples were then loaded onto a 2 % gel as described in section 3.3.3.1 at 70V for 20 min. Candidate bands were gel extracted as per section Appendix 4.

3.3.5 Expression Host Transformation and Verification

Competent cells of *E.coli* BL21 and JM105 were prepared as described in 3.3.1.1. The successful ligation reactions from 3.3.4 (pGEX-4T-1F and pGEX-4T-1A), which were shown to contain the insert, were transformed into both strains of *E.coli* as described in 3.3.2.1. Agar plates and glycerol stocks were prepared for each transformed strain as previously described in section 3.3.2.1.

Cultures of the transformed *E.coli* BL21 and JM105 strains containing the modified vectors (pGEX-4T-1F and pGEX-4T-1A) were prepared and grown to log phase (3.3.2.2). The recombinant plasmids were isolated using the Promega Wizard® Plus SV Miniprep DNA Purification System (Appendix 6). The extracted plasmids were sequenced with both forward and reverse primers by The Sequencing Service (University of Dundee, Tayside, Scotland) using 15 µl (~300 ng) of extracted plasmid and 1 µl (3.2 pmol) of either primer. Sequence data results were analysed using the freeware Chromas (C. McCarthy, Queensland, Australia: version 1.45, 32 bit) software to check sequencing calls and DNAMAN program for alignment.

3.3.6 Small Scale Expression

The expression of bacterial fusion proteins can be divided into three main stages, each of which can require optimisation to improve yield. This includes, inducing a log phase culture of transformed recombinant *E.coli* followed by resuspension and lysis to release the fusion protein from the cells. The last step is the isolation and purification of the protein of interest from the other proteins present (Figure 33).

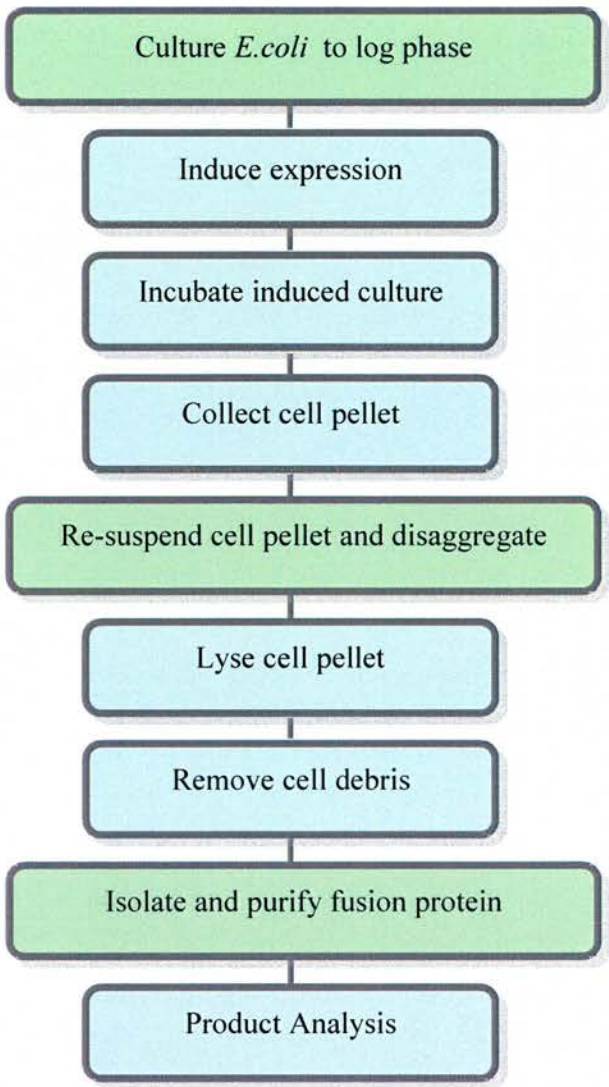


Figure 33: Main recombinant fusion protein production stages and optimisation points

The methods used to express, isolate and purify the Carcinin fusion proteins followed the vector manufacturer's protocols, as described in the GST Fusion System Handbook (Amersham Pharmacia Biotech, 2002), which is adapted (with permission) from Current Protocols in Molecular Biology, Vol. 2, Supplement 10, Unit 16.7. Copyright © by Current Protocols. All constructs were evaluated in optimisation studies with both *E.coli* BL21 and *E.coli* JM105 bacterial strains.

Construct Bacterial Strain	pGEX-4T-1 control vector	pGEX-4T1F construct 1	pGEX-4T-1A construct 2
E.coli JM105	pGEX-4T-1	pGEX-4T-1F	pGEX-4T-1A
E.coli BL21	pGEX-4T-1	pGEX-4T-1F	pGEX-4T-1A
Expression Product	GST	GST-FCAR	GST-ACAR

Table 9: Nomenclature of expression constructs and their products

The unaltered pGEX-4T-1 vector (Table 9 and Figure 34) was used as a positive control to express the GST tag protein (~26 kDa, see Appendix 19) for comparison of yield and homogeneity to the expressed recombinant fusion proteins. The experimental fusion products were the GST-FCAR and the GST-ACAR proteins produced by the coding sequences selected in section 3.3.3 attached to the GST tag by a thrombin cleavage recognition sequence (see Figure 34).

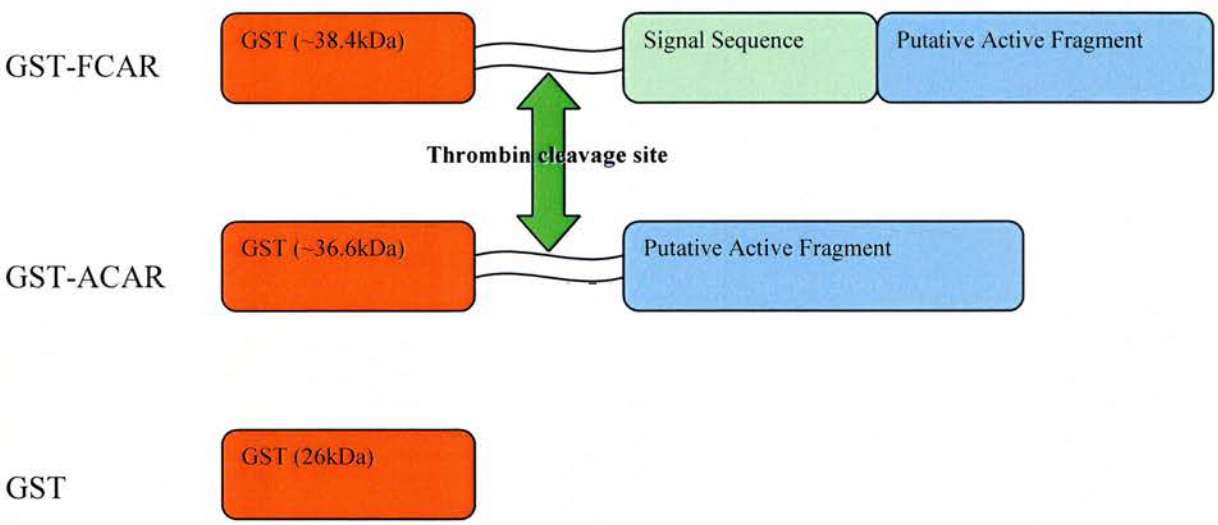


Figure 34: Expressed products (Appendix 19) from constructs used (pGEX-4T-1F, pGEX-4T-1A and pGEX-4T-1)

Expression and purification of a particular fusion protein can be achieved using a variety of different conditions, although the yield and efficiency of the process may vary. Small scale (2-50 ml) cultures were used for initial expression, isolation and

cleavage of the eluted or bound fusion protein and larger scale cultures (100 ml –2 l) were used to optimise the purification stages (see 3.3.7). Expression efficiency and yield were assessed by the collection of 10 µl samples that were analysed by sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS PAGE) (samples 1-7; see Table 10) (see 3.3.9 for protocol).

The enriched medium, 2 x YTA (Appendix 14) was used for culturing the transformed recombinant bacterial strains. Small cultures (2 ml) were prepared and inoculated with either *E.coli* BL21 or *E.coli* JM105, previously transformed with either pGEX-4T-1F, pGEX-4T-1A or pGEX-4T-1 vectors (prepared in 3.3.5 above). Inoculated cultures were grown to log phase in a shaking, water bath (37 °C @ 50 rpm) for ~3.5 h in separate 50 ml sterile Falcon® tubes. To calibrate the time required to reach log phase under these conditions, cultures were sampled periodically and the OD (A_{600}) measured to monitor growth as described in section 3.3.1.1.

Once the cultures had reached an OD of (A_{600}) ~ 0.6 (sample 1) they were then induced with 100 mM (2 µl) IPTG (Promega, Hampshire, UK) to a final concentration of 0.1 mM and incubated for a further 1.5 h with shaking (50 rpm) at 37 °C (sample 2). The induced cultures were then centrifuged to pellet the cells at 14,000 x g for 30 s at RT and the supernatant sampled and then discarded (sample 3). The cell pellets were either immediately processed or frozen at –80 °C for processing the following day.

Sample Number	Sample Type
sample 1	Pre-induction
sample 2	Post-induction with IPTG
sample 3	Culture supernatant
sample 4	Resuspended cells
sample 5	Sonicate / Lysate
sample 6	Lysate cell debris pellet
sample 7	Lysate cell supernatant

Table 10: Samples collected during expression optimisation steps for analysis by SDS PAGE

Frozen pellets were defrosted on ice for up to 2 h and then processed as described below for freshly prepared pellets. The cell pellets were resuspended with 300 µl of 1 x PBS (sample 4) and then sonicated in an ice water bath using a US70 Ultrasonic Homogeniser (Philip Harris Scientific, Staffordshire, UK). The sonication regime was performed using a cup horn (SH3213 - high gain horn, Philip Harris Scientific, Staffordshire, UK) for 30 s at 75 % power and with 5 s pulses.

The lysate (sample 5) was centrifuged to pellet the cell debris (sample 6) at 14,000 x g for 5 min at 4 °C and the supernatant sampled as before (sample 7).

The post induction samples collected, were analysed by SDS-PAGE (see 3.3.9) to confirm expression of the fusion protein, before proceeding to the isolation and purification stages (see 3.3.6.2).

3.3.6.1 GROWTH AND EXPRESSION OPTIMISATION

The culture medium (2 x YTA) was modified to 2 x YTAG (see Appendix 14) to include 2 % glucose to investigate the effect of reducing basal level expression on yield.

Optimisation Parameters	Range used
Culture Volume	2 ml, 15 ml or 50 ml
Temperature	20 °C, 24 °C or 37 °C
Incubation Time	1-19h
Shaking	~50, ~100 or ~250 rpm
A_{600}	0.5-0.9
Optimisation Parameters	Range used
Culture Volume	2 ml, 15 ml or 50 ml

Table 11 Culture optimisation parameters

The interdependent growth parameters were manipulated (i.e. temperature, aeration and cell density) to achieve the optimum growth conditions for the culture (Table 11). To further increase solubility of the fusion protein and reduce the possible formation of inclusion bodies (Schein and Noteborn, 1988), {as suggested Amersham fusion book and www.emble_heidleberg.de}, IPTG concentration, and cell density (Dian *et al.*, 2002b) were varied to improve yields (see Table 12). The aeration of the culture was also increased by increasing the rpm of the shaker and increasing the flask size to allow greater surface area to volume ratio.

Optimisation Parameters	Range tried
Final IPTG Concentration	0.05 mM, 0.075 mM, 0.1 mM, 0.2 mM, 1 mM & 2 mM
Temperature	~20 °C
Induction Time	1, 2, 3, 8, & 12.5h
Shaking	~250 rpm
Optimisation Parameters	Range tried

Table 12: Induction optimisation parameters

Two additional buffers (as well as PBS) were used to resuspend and lyse the cells (using 50 µl per 1 ml of culture volume prepared) of STE solution (10 mM Tris, pH 8.0, 150 mM NaCl & 1 mM EDTA) (Frangioni and Neel, 1993; Amersham Pharmacia Biotech, 2002; Mercado-Pimentel *et al.*, 2002)

Several additives (Table 13), added to the resuspended cells, were evaluated to aid solubility, prevent aggregation and aid binding of the fusion proteins to the GS4B matrix (Sambrook and Russell, 2001; Dian *et al.*, 2002b).

Additives
1 mM (v/v) phenyl methylsulphonyl fluoride (PMSF)
10 mg ml-1 (w/v) lysozyme in 25 mM (w/v) Tris-HCl pH 8.0
5 μg ml^{-1} of (w/v) RNAse and DNAse
1.5 % (w/v) N-lauroylsarcosine (ICN)
5 mM (v/v) 1, 4 dithiothreitol (DTT)
2 % (v/v) Triton X-100

Table 13: Additives used with resuspension buffers to aid solubility of fusion proteins.

Lysozyme (Sigma – chicken egg white) was added (10 mg ml^{-1} in 25 mM Tris-HCl, pH 8 to a final concentration ~1 mg ml^{-1} (w/v)) to the resuspended cells and incubated on ice for 30 min before sonication to increase cell disruption and aid the release of intracellular expressed proteins (Amersham Pharmacia Biotech, 2002).

Viscosity was observed to increase during lysis and in an attempt to disaggregate proteins, Triton at 1 %, RNAse (w/v 5 μg ml^{-1}) and DNAse (w/v 5 μg ml^{-1}) were added to the resuspended cells or lysate (Sambrook and Russell, 2001).

The use of detergents (i.e. Triton X-100 N-laurylsarcosine) at 1:1 or 1:2 (molar ratio) has been shown to improve solubility of GST fusion proteins maintaining glutathione matrix binding ability and inhibiting certain bacterial proteases (Frangioni and Neel, 1993). As suggested in the protocol, N-laurylsarcosine (1.5 %) and DTT (5 mM v/v) were added to cells resuspended in STE buffer before sonication (30 s @ 75 % & 5 s pulses). Triton X-100 (2 %) was added to the lysate supernatant before purification by affinity chromatography (see below in section 3.3.6.2). Triton X-100 sequesters N-

lauryolsarcosine (Sambrook and Russell, 2001) thus improving solubilisation but has little effect on the binding efficacy of the GS4B matrix used in subsequent purification steps. The resuspended cell pellet was then sonicated in an ice bath and the sonication regime optimised to increase lysis but avoiding denaturation of the protein using the regimes described in Table 14.

Probe and Power	Time and Bursts
SH213 high gain cup horn at 60-100 % into an ice water bath	10, 20 or 30 s at 5-9 %
"	1 or 3 min at 5-9 %
MS72 titanium tapered MicroTip at 10-30 % directly into resuspended sample placed in an ice-water bath	10, 20 or 30 s at 5- 9 %
"	1 or 3 min at 5-9 %

Table 14: Optimisation of sonication regimes used with different probes

The resultant lysate was then centrifuged at ~12,000 x g for 5 min at RT and the cell debris sampled and then stored frozen at –20 °C. The centrifugation speed was later increased to 31,000 x g for 10 min at 4 °C.

The supernatant was then filtered through a 0.45 μm nitrocellulose filter (Whatman), as recommended by Sambrook and Russell (2001) to prevent subsequent clogging of the resin.

A sample of the supernatant was also collected (Sample 8) and the remainder kept on ice until it could be added to glutathione Sepharose 4B affinity matrix for isolation of the fusion protein (see 3.3.6.2 below).

3.3.6.2 ISOLATION

The expressed fusion proteins were purified by affinity chromatography, from the lysate supernatant prepared in 3.3.6 above, using glutathione Sepharose 4B (Figure 27). As

some of the fusion protein may remain in the lysate pellet as inclusion bodies, recovery of these was also investigated (Appendix 20).

Depending on the desired amount of fusion protein to be purified, the column packing material for purification was varied and only small quantities were needed to optimise the process.

Protein Yield / Component	50,000 µg	10,000 µg	1000 µg	50 µg	25 µg	5 µg
Culture Volume	**20 l**	**4 l**	**400 ml**	**20 ml**	**10 ml**	**2 ml**
Volume of Sonicate	**1000 ml**	**200 ml**	**20 ml**	**1 ml**	**500 µl**	**300 µl**
Matrix Bed Volume	**10 ml**	**2 ml**	**200 ml**	**10 µl**	**10 µl**	**10 µl**
1X PBS	**100 ml**	**20 ml**	**2 ml**	**100 µl**	**100 µl**	**100 µl**
Elution Buffer	**10 ml**	**2 ml**	**200 µl**	**10 µl**	**10 µl**	**10 µl**

Table 15: Reagent volume requirements for different protein yields (Amersham Pharmacia Biotech, 2001)

Glutathione Sepharose 4B beads (GS-4B beads) were supplied as a dry powder which was hydrated before use, by adding 14 ml of distilled water per 70 mg of powder (200 ml g^{-1}) and incubating overnight (~16 h) at 4 °C to give ~1 ml of hydrated column material.

The hydrated GS-4B beads were then washed three times in 10 bed volumes of distilled water to remove storage chemicals by repeated centrifugation (500 x g for 5 min at RT) and re-suspension. The washed beads were then equilibrated with 10 volumes of cold (4 °C) 1 x PBS, by centrifugation and re-suspension as detailed above. Finally, the beads were resuspended in 1 M NaCl (BDH) to form a 50 % slurry and stored for up to one month (or for longer by adding 20 % ethanol (BDH)) at 4 °C, in a 14 ml centrifuge tube. The binding capacity of the resin is 5-10 mg of glutathione S-transferase per ml of resin. The beads were equilibrated before use by repeated washing with 1 x PBS (4 °C) and centrifugation and preparation as a 50 % slurry.

To the supernatant collected in section 3.3.6, 10 µl of the 50 % GS-4B slurry was added and mixed gently at RT for 5 min. The slurry mixture was then centrifuged at 500 x g for 5 min at RT. The supernatant was collected and sampled (Sample 8) for subsequent analysis of binding efficiency. The GST-4B beads were then washed three times with 100 µl ice-cold 1x PBS (10 µl samples collected at each after each centrifuge stage Sample 9-11) by repeated centrifugation (500 x g, 10 min at RT) and resuspension.

The bound protein (GST, GST-ACAR or GST-FACR) was then eluted from the pelleted matrix by adding 10 µl (1 ml elution buffer per 1 ml bed volume) of freshly prepared 10 mM glutathione elution buffer (Table 17) mixing at RT for 5 min and then centrifuging at 500 x g at RT for 5 min. The reduced glutathione in the elution buffer behaves as a competitor for the GST binding sites displacing the fusion protein which can then be purified from the supernatant (Dian *et al.*, 2002b). This process was repeated 3 times, each time collecting the supernatant (samples 12-14) and adding a further aliquot of elution buffer.

The eluted proteins were analysed by both SDS PAGE and spectrophotometry. The GS-4B beads were regenerated when supplies of fresh beads were exhausted by washing 3 times with 0.2 M NaOH and equilibrated with 1x PBS followed by storage at 4 °C with 20 % ethanol.

Sample Number	Sample Type
sample 8	Unbound lysate
sample 9	1st PBS wash
sample 10	2nd PBS wash
sample 11	3rd PBS wash
sample 12	1st elution 5min at RT
sample 13	2nd elution 10min at RT
sample 14	3rd elution 24hr at RT

Table 16: Samples collected for analysis by SDS-PAGE during isolation steps

Elution Buffer	Components
Glutathione elution buffer 1a	10 mM reduce glutathione in 50 mM Tris-HCl (pH 8.0)
Glutathione elution buffer 1b	20 mM reduce glutathione in 50 mM Tris-HCl (pH 8.0)
Glutathione elution buffer 2	75 mM HEPES @ pH 7.4, 150 mM NaCl, 10 mM reduced glutathione, and 5 mM DTT

Table 17: Elution buffers tried in fusion protein elution optimisation (Appendix 14 for method)

3.3.6.2.1 Isolation Optimisation

It has been suggested (Mercado-Pimentel *et al.*, 2002), that increasing the bed volume of the GS4B beads can considerably increase the amount of bound fusion protein. Two to ten fold greater bed volumes (than recommended in Table 15) were used in later expression isolations to investigate this, using the complementary volumes of washes and elution buffers.

Recommended incubation (supernatant with the GS4B beads) times for large scale cultures are about 30 min (Sambrook and Russell, 2001; Mercado-Pimentel *et al.*, 2002). After mixing, the number of washes of 1 x PBS (after binding of the fusion protein to the beads) was increased from 3 to 10, of 10 bed volumes (Mercado-Pimentel *et al.*, 2002), to aid the removal of unwanted host proteins from the beads.

Increasing the yield of fusion protein was also investigated by changing the composition and concentration of the elution buffer. The reduced glutathione in glutathione elution buffer 1a (Table 17) was increased from 10 mM to 20 mM, with the mixing and centrifugation conditions remaining the same. For the 20 mM glutathione elution buffer (1b) the concentration of Tris-HCl was increased to 100 mM from 50 mM to stabilise the pH of the solution. Fusion proteins released using the STE resuspension buffer were eluted from the beads using glutathione elution buffer 2 (Table 17).

The volume of the elution buffers were maintained in line with the bead volumes to avoid over dilution of the eluted protein. The number of elutions and the amount of time required for each elution were also varied (5 min –16 h) to achieve the optimum elution profile for this system. After each centrifugation step, the supernatant was sampled for analysis by SDS-PAGE and the remainder pooled and kept on ice in a 50 ml Falcon® tube or frozen at –20 °C.

Alternative methods of bead regeneration were investigated to improve the binding capabilities of the regenerated beads. The beads were regenerated following the manufacturers protocols (Sigma) by washing in 5 bed volumes of cleansing buffer 1 (0.1 M borate buffer, pH 8 to 9 containing 0.5 M NaCl) and then washing with 5 bed volumes of deionised water. The beads were then washed with cleansing buffer 2 (0.1 M acetate buffer, pH 4 containing 0.5 M NaCl) and then equilibrated with 1 x PBS.

When precipitated or denatured substances were visible or suspected to be affecting subsequent binding ability the beads, they were cleaned and regenerated by washing in two bed volumes of 6 M guanidine hydrochloride and then immediately washing with five volumes of 1 x PBS (pH 7.4) (Amersham Pharmacia Biotech, 2001).

All regenerated beads were then washed in 1x PBS and then stored with 20 % ethanol at neutral pH at 4 °C.

3.3.6.2.2 Solubilisation of Fusion Proteins from Inclusion Bodies

Some expressed GST fusion proteins are known to remain in the cell pellet as inclusion bodies using the pGEX range of vectors (Piers *et al.*, 1993; Zhang *et al.*, 1998). Solubilisation of inclusion bodies from the lysate pellet was investigated using the method described by Sambrook. and Russell, (2001) using urea, based on an adapted procedure from Schoner *et al.* (1992).

In brief, the cell debris lysate pellet collected in 3.3.6 was resuspended in 1 ml of distilled H_2O and 100 µl aliquots of these were transferred to separate microcentrifuge tubes. To these, different concentrations (0.5 M, 1 M, 2 M and 4 M) of urea (Sigma – Aldrich), dissolved in 0.05M Tris-HCl, were added and the pellet resuspended. These resuspensions were then centrifuged at 14,000 x g for 15 min at 4 °C and samples of the supernatant analysed by SDS PAGE to evaluate the concentration of urea which led to the best recovery of the inclusion bodies.

3.3.6.3 PURIFICATION

The expressed fusion protein was cleaved from the GST tag (Figure 34) using thrombin protease (Amersham Biosciences, Buckinghamshire, UK) initially at 25 units.

After elution of the intact fusion protein (3.3.6.2.1 above), 25 units of thrombin protease per ml of bed volume used (Amersham Pharmacia Biotech, 2002), was added to the sample and incubated at RT for 6 h, 10 h and 20 h and then sampled (Sample 15a, see Table 18) for analysis. Digestion of the fusion protein was also tried at 4 °C, but this was not found to be as effective, in agreement with other studies (Dian *et al.*, 2002a).

These digested "off-bead" samples were then dialysed with 1 x PBS using benzoylated dialysis tubing (D2272, Sigma), to remove the glutathione and equilibrate the sample in preparation of further chromatographic purification steps. The tubing was sealed at both ends with 4 cm dialysis clamps, and dialysed against 2000 volumes of cold (4 °C), stirred buffer (1 x PBS) for ~18-24 h at 4 °C with one change of buffer after 4 h. The dialysed sample was analysed SDS PAGE (Sample 16, Table 18) and the remainder stored at –20 °C.

Sample Number	Sample Type
Sample 15 a and b	**Digested sample (a) / supernatant (b)**
Sample 16	**Dialysed sample**

Table 18: Samples collected during cleavage steps

3.3.6.3.1 Purification Optimisation

The amount of thrombin added to cleave the fusion protein was increased from 25 to 80 units, as initial results indicated incomplete cleavage.

The number of elutions and the incubation time of the elution buffer with the beads was varied to investigate the effect on the yield of fusion protein. Up to five elutions were performed using the repeated incubation and centrifugation protocol previously described above. The incubation time was varied from 5 min to 24 h, with earlier elutions taking place at shorter time intervals and the final elutions conducted over longer periods of time.

To reduce on the number of steps in the purification stage and to avoid loss of the protein at each stage, cleavage of the fusion protein whilst still bound ("on –bead") to the GS4B matrix was investigated using thrombin protease (Amersham Pharmacia Biotech, 2002). GS-4B bound fusion proteins which were thoroughly washed with 1 x

PBS (as described in section 3.3.6.2.1) were pelleted by centrifugation at 500 x g for 5 min at RT. Thrombin (Amersham Biosciences) (80 units), diluted in one bed volume of 1 x PBS, was added to the pelleted GS4B beads and incubated at RT for 16 h with rocking. The slurry was then centrifuged (500 x g, 5 min, RT), resuspended 8 times with 1 volume 1 x PBS and supernatants were collected and sampled (Sample 15b, Table 18).

The beads were then regenerated as previously described (3.3.6.2). The digested sample obtained, was then dialysed as described in 3.3.6.3 above for the "off-bead" samples using a benzmaidine binding buffer (0.05 M Tris–HCl, 0.5 M NaCl, pH 7.4) (Amersham Pharmacia Biotech, 2002) and then either frozen at –20 °C immediately or kept briefly on ice until further purification steps could be performed (see below in section 3.3.8).

Due to some precipitation noted during thrombin digestion stages, the pH of the digestion buffer (50 mM Tris-HCl) was optimised to different pHs of 7.7, 8.0, 8.2 and 8.4, even though no differences were noted on digestion due to pH of the buffer by Dian and co workers (2002a).

3.3.7 Large Scale Expression and Isolation

Following optimization of the small scale cultures (see section 3.3.6.1) the expression system was scaled up to increase the yield of the fusion protein including the optimized growth, expression and isolation parameters.

Two liters of sterile 2 x YTAG medium (see Appendix 14) were prepared in a 5 l conical flask and stored until required at 4 °C.

A fifteen millilitre starter culture of 2 x YTAG was aseptically inoculated, as described in section 3.3.6 and grown at 37 °C for 5 h. The starter culture was aseptically introduced to the 2 l main culture medium and incubated at 20 °C for 19 h or until the optical density (OD) reached an absorbance (A_{600}) of ~0.5-1.0.

The culture was then induced with 2 ml of 100 mM IPTG (final concentration 0.1 M) and incubated at 20 °C for a further 3.5 h. The culture was then centrifuged at 7,700 x g for 10 min at 4 °C to pellet the cells.

The supernatant was discarded and the pellets were resuspended using a 5 ml pipette with 50 µl of 1 x PBS per 1 ml of culture (95 ml). The samples were pooled and sonicated at 60 % power for 30 s at 5 s pulses using the MS72 probe directly into the sample (no lysozyme was added). Triton X-100 (20 % in 1 x PBS) was added to a final concentration of 1 % (5 ml) and mixed gently in 50 ml Falcon® tubes for 30 min on ice after sonication. The lysate was then transferred to Beckman polycarbonate centrifuge tubes and centrifuged at 31,000 x g for 10 min at 4 °C. The supernatant was then filtered through a 0.45 µm nitrocellulose filter (Whatman, Kent, UK) by vacuum filtering into a chilled container into 50 ml Falcon® tubes.

A 50 % suspension of equilibrated (1 x PBS) GS-4B (2 ml) of was added to the elutant and mixed by gentle end-over-end mixing at room temperature (24 °C) for 30 min. The affinity matrix was then centrifuged (500 x g for 5 min @ RT) and the supernatant stored at –20 °C. The beads were then washed 10 times in 10 bed volumes of 1 x PBS (10 ml).

Fusion proteins were digested as described for the small scale expression trials (3.3.6.3) for the "off bead" samples, after elution with 20 mM glutathione elution buffer (see 3.3.6.3) (pH 8.0), and for the "on-bead" samples as described in 3.3.6.3.1.

3.3.8 Final Purification

After dialysis with the appropriate buffer, a number of chromatographic methods were employed to remove GST, serine proteases (e.g. thrombin) and any remaining host proteins. The dialysed (1 x PBS) "off-bead" samples (3.3.6.3) should contain Carcinin, GST, thrombin and possible residual host proteins. The dialysed (benzamidine binding buffer) "on-bead" samples (3.3.6.3.1) should contain Carcinin, thrombin and any remaining host proteins, but no GST.

3.3.8.1 THROMBIN REMOVAL

Both the "on-bead" and "off bead" samples (3.3.8), contained unused thrombin protease which was removed by affinity chromatography using benzamidine Sepharose 6B (1 ml binds 13 mg of trypsin) after dialysis against a benzamidine buffer (50 mM Tris-HCl, 0.5 M NaCl, pH 7.4) as described for PBS in 3.3.6.3. The matrix is supplied as a 5 ml slurry in a 0.9 % NaCl solution with 0.01 % thimerosal which was packed to into a 1 x 8 cm chromatographic column. The column was equilibrated with benzamidine binding buffer and the dialysed samples were loaded onto the column at 1 ml min^{-1}. The flow through was monitored at A_{280} (set to 1 AU FSD, Econo System; BioRad, Hertfordshire, UK) and collected as 1 ml fractions. These fractions were analysed by SDS-PAGE as described in section 3.3.9.

3.3.8.2 GST REMOVAL

The resultant "off-bead" sample (3.3.8.1) was then added to a 50 % preparation of glutathione Sepharose 4B as described in section 3.3.6.2 and incubated at RT for 30 min with gentle end over end mixing. The suspension was then centrifuged (500 x g, 5 min, RT), the supernatant collected and the beads resuspended ten times with 1 x PBS (1 bed

volume per wash). The supernatants from the washes were pooled and either kept on ice or frozen at –20 °C until required.

3.3.8.3 FAST PROTEIN LIQUID CHROMATOGRAPHY (FPLC)

GST removal from "off bead" samples was also attempted using FPLC. Fractions obtained above (3.3.8.1 above) containing ~10 kDa bands, were pooled and loaded onto a HR10/10 desalting column using a Super Loop (10 or 50 ml) and then run through an HR5/5 MonoS FPLC column (both from Pharmacia) at 1 ml min^{-1} with filtered 50 mM Tris-HCl (pH 8) (buffer A). The column was then eluted with filtered 50 mM Tris-HCl containing 1 M NaCl pH 8 (buffer B) with a stepwise gradient of 0-5 min of 100 % buffer A and 0 % buffer B, followed by 6-35 min of 50 % buffer A and 50 % buffer B and finally 36-40 min of 0 % buffer A and 100 % buffer B.

The same samples were also loaded onto an anion exchanger, HR5/5 MonoQ FPLC column, with the same buffers and elution gradient.

3.3.8.4 REVERSE PHASE HIGH PRESSURE LIQUID CHROMATOGRAPHY (RP-HPLC)

RP-HPLC was also investigated as an alternative method to purify the cleaved Carcinin moiety from the rest of the proteins in digested "off-bead" samples obtained in section 3.3.6.3.1 as per the conditions described by Relf *et al.* (1999) for the native protein. In brief, bound fusion proteins were digested with thrombin (80 units) dissolved in 50mM sodium phosphate buffer (pH 7) which was incubated for 24 h at 20 °C with gentle mixing. After centrifugation (500 x g, 5 min at RT) the supernatant was acidified with filtered 0.1 % TFA (BDH) in distilled water. One ml samples were loaded on to a 5 g Sep-Pak C18 cartridge (Waters, Hertfordshire, UK) at 1ml min^{-1} with 0.1 %TFA (buffer A). The column was eluted using a stepwise gravity gradient (c. 1 ml min^{-1}) with a

regime of 20 min of 100 % buffer A and 0 % buffer B, altered to 40 % buffer A and 60 % buffer B after 60 min and finally 0 % buffer A and 100 % of buffer B after 80 min. The elutant was collected in 1 ml fractions which were lyophilised and resuspended in 100 µl of deionised water and 10 µl samples of these were analysed by SDS-PAGE as detailed in section 3.3.9.1

Samples obtained after FPLC (section 3.3.8.3), containing protein bands around ~10 kDa, as observed by SDS-PAGE analysis, were acidified with filtered 0.1 % TFA (BDH, Dorset, UK) in distilled water. These samples were loaded as described above using a C18 column and the same buffers with a regime of 60 min of 100 % buffer A and 0 % buffer B after which the proportions were altered to 40 % buffer A and 60 % buffer B after 20 min and finally to 0 % buffer A and 100 % of buffer B after 90 min. The fractions were collected and analysed as described above.

3.3.9 Sample Analysis

3.3.9.1 SDS-PAGE

At each defined stage a 10 µl sample was collected to a 0.5 ml centrifuge tube (Samples 1-7; see Tables 10) and stored at –20 °C for subsequent analysis. Protein profiles of these samples were analysed by based on the methods described by Laemmli (1970).

A two part discontinuous gel method was utilised to analyse protein samples comprising of two standard stacking gels (Table 19 and Table 20) and two 12 % resolving gels as detailed below and used with a Mini-PROTEAN® II Electrophoresis Cell (BioRad, California, USA).

	Volume	Components
lower gel 4x buffer	2.5 ml	1.5 M Tris-HCl, pH 8.8, 0.4 % SDS
water	4.4 ml	Deionised
acrylamide solution 40 %	3.0 ml	
APS 10 %	50.0 µl	10 % Amonium persulphate in water (fresh)
TEMED	5.0 µl	N,N,N,',N',tetramethylethylenediamine

Table 19: Resolving gel (12 %) formulation

	Volume	Components
upper gel 4x buffer	2.5 ml	0.5 M Tris-HCl, pH 6.8, 0.4 % SDS
water	6.6 ml	Deionised
acrylamide solution 40 %	0.8 ml	
APS 10 %	100.0 µl	10 % Amonium persulphate in water (fresh)
TEMED	10.0 µl	N,N,N,',N',tetramethylethylenediamine

Table 20: Stacking gel formulation

The resolving gel (Table 19) was prepared first using the casting stand, the APS and TEMED which catalyse polymerisation of the acrylamide, were added last and the gel was gently poured between the glass plates. The gel was then overlayed with 40 % butanol solution and allowed to set for 30 min. The overlay was then removed and washed with distilled water. The stacking gel was then prepared as above (Table 20) and the stacking comb inserted immediately into the top of the gel and left to polymerise for a further 30 min. The gel cassettes were then fixed to the sandwich clamp support (upper buffer chamber) and placed into the lower buffer chamber with gel running buffer (25 mM Tris base, 192 mM glycine, 0.1 % SDS, pH 8.3). The comb was removed and the wells flushed with running buffer.

An equal volume of protein sample was added to the sample buffer (250 mM Tris, 2 % (w/v) SDS, 10 % (w/v) glycerol, 0.01 % (w/v) bromophenol blue (BDH), 2 % (v/v) β-mercaptoethanol (pH 6.8). The mixed samples were denatured by heated to 100 °C for

3 min and then the samples were directly loaded into the wells. The gels were run at a constant voltage (200 V) and 120 mA as recommended in the manual until the dye front was 1 cm from the end of the resolving gel. After electrophoresis the gels were stained with Coomassie (blue) as detailed below.

3.3.9.2 COOMASSIE STAINING

Gels were immersed for 7 min in staining solution (0.25 % (w/v) Coomassie brilliant blue R250 (BDH), 45 % (v/v) methanol (BDH) and 10 % (v/v) glacial ascetic acid (BDH) and then destained in a destaining solution (10 % (w/v) methanol and 5 % (w/v) acetic acid) for ~24 h.

3.3.9.3 RADIAL DIFFUSION ASSAY (RDA)

Antibacterial activity of the cleaved expressed ACAR protein was tested using a modified radial diffusion assay of Lehrer a *et al.* (1999) and using *Planococcus citreus,* as the test bacterium (Relf *et al.*, 1999). The 15 ml sterile marine broth was inoculated aseptically with a single colony of *P. citreus* and grown at 24 °C for ~16 h with shaking. The bacterium were centrifuged at 300 x g for 15 min at 4 °C to pellet the cells and these were then washed in 3.2 % marine saline and resuspended to ~10^7 bacteria ml-1 (calibrated by diluted spread plate counts).

An 1 % agarose underlay was prepared in 1/10 strength 2216E (Difco) and seeded with ~10^7 bacteria and poured evenly into square (144 cm^2) Petri dishes and allowed to set. Wells (approx. 2 mm diameter) were punched into the agarose with the fine end of a Pasteur pipette about 1 cm apart across the dish. Into each well either 3 µl of sample were pipetted along with single positive (3 µl of Penicillin G @ 50 µg ml^{-1}) and negative (3 µl sterile water) controls. The plate was then incubated at 4 °C for 3 h and then overlaid with 14 ml of sterile 1 % agarose containing double strength 2216E broth

before further incubation at 12-24 h at 20 °C. After incubation, clear zones around the wells were considered to indicate antibacterial activity.

3.3.9.4 PROTEIN MASS FINGERPRINTING

Protein samples analysed by SDS-PAGE gel which produced strong bands at around ~11 kDa after Coomassie staining were cut out of the gel with a clean scalpel and placed into a 0.5 ml microfuge tube. These samples were analysed by Dr. Catherine Botting (St. Andrews University, St. Andrews, UK) by mass spectrometric analysis using a MALDI-TOF MS (Micromass TofSpec-2E) after tryptic digest with an (Investigator ProGest Protein Digestion Station) using methods adapted from (Shevchenko *et al.*, 1996). Briefly the gel slice was washed, rehydrated and destained using a series of washing and incubation steps with acetonitrile, NH_4HCO_3, DTT and iodoacetamide. The DTT and iodoacetamide treatments ensure reduction of the disulphide bonds resulting in linear peptide fragments. The proteins were then digested with 12.5 ng μl^{-1} trypsin at 4 °C. After 45 min the remaining supernatant was removed and replaced with NH_4HCO_3 and incubated overnight at 37 °C. The proteins were then extracted with a series of incubations/washes with NH_4HCO_3, acetonitrile and 5 % formic acid. These extracts were vacuum dried (35-45 °C), redissolved in formic acid and submitted for mass spectrometry (see www.st-andrews.ac.uk/~bmsmspf/ for full details).

Molecular weights of the linear fragments obtained were then compared to theoretical expected fragments of digested Carcinin and sequences in NCBI database using Mascot software (www.matrixscience.com). This produces a Mowse probability score describing the similarity between the molecular weights of the linear peptides observed compared to theoretically trypsin-digested proteins from the database.

3.4 Results

3.4.1 Competent Cell Preparation and Transformation Analysis

Bacterial strains (*E.coli* BL21 and JM105) were cultured successfully in LB broth and stored as short and long term stocks (on agarose plates at 4 °C and in glycerols stocks at –80 °C). Strains were routinely re-cultured from the glycerol stocks to new LB agar plates to maintain healthy bacterial cell lines for use in the experiments.

During competent cell preparation, overnight cultures (15 ml) were grown to an OD (A_{600}) of 0.4-0.5, which, for *E.coli* BL21 was achieved after 3.5-4 h (Figure 35). Similar times were required for the JM105 strain.

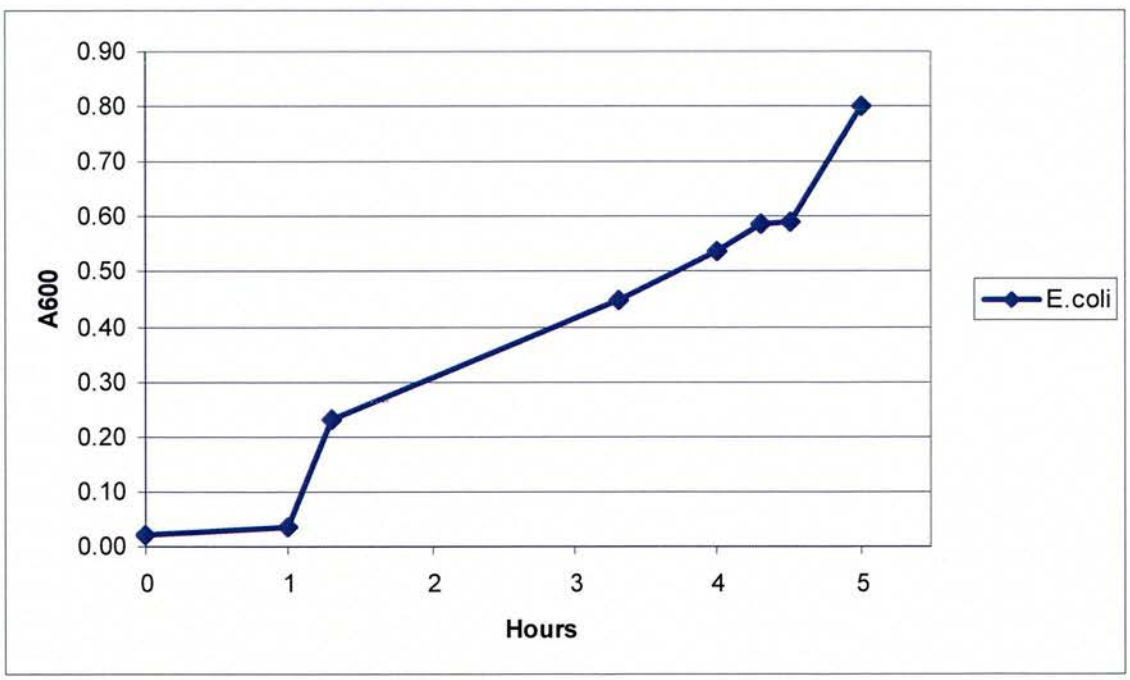


Figure 35: Exponential growth of 15 ml culture of *E.coli* BL21 at 37 °C in LB broth (~50 rpm) as measured by OD at A_{600}.

Competent cell preparation and transformation success (of the constructs into *E.coli* BL21 and *E.coli* JM105 strains) was confirmed by growth on the LBAG plates (as the

vector confers resistance to the bacteria through the β - lactamase Amp^r gene: gene region 1377-2235 bp) and also on LB plates as a positive control for bacterial viability. Negative transformation controls, were confirmed by the absence of growth on LBAG plates (i.e. no transformation), but growth was observed on LB plates (positive bacterial viability). Positive transformation controls, using the unaltered pGEX4T-1 vector, confirmed successful competent cell preparation and vector transformation by growth on both LBAG and LBG plates. No marked difference was noted in the transformation efficiency (number of colonies grown on LBAG plates) between the two bacterial strains, although this observation was not confirmed empirically.

Transformation success, purity and size of the vectors was confirmed by gel electrophoresis (Figure 36 and Figure 41).

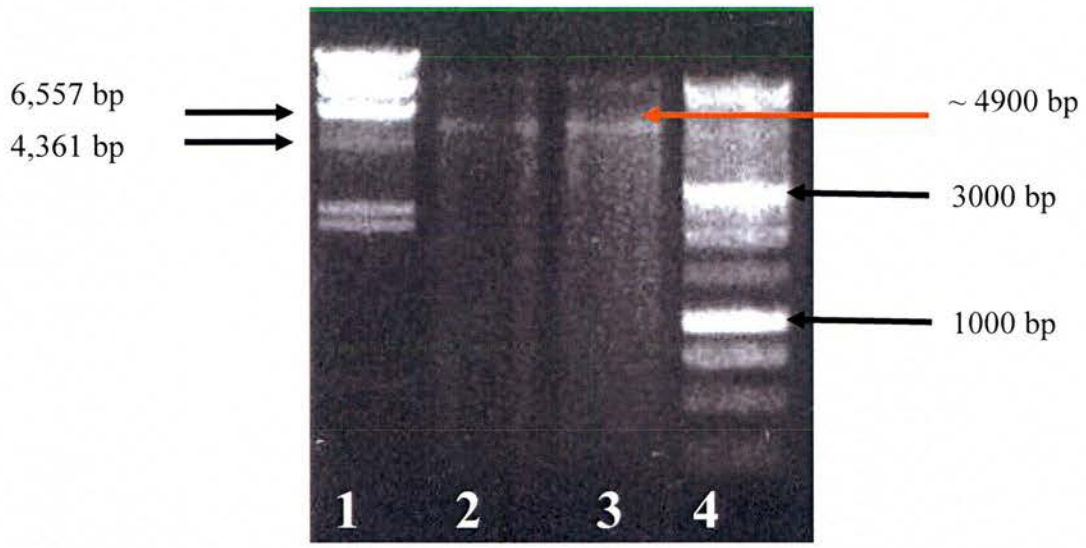


Figure 36: Recovery of pGEX4T-1 plasmid after Wizard® Plus SV Miniprep DNA Purification System (Lane 3: Promega) with HindIII digested and 1 kb DNA markers (Promega: Lanes 1 and 4 respectively) on a 0.7 % agarose gel.

The concentrations of the recovered pGEX4T-1 plasmids, after propagation and purification, were quantified using their absorbance (A_{260}) readings (Figure 37).

Quantification of DNA is based on the premise that a sample with an absorbance at 260 nm of 1 is equal to a concentration of 50 μg ml^{-1}.

$$\text{DNA Concentration } (\mu g\ ml^{-1}) = A_{260} * 50$$

Figure 37: Equation used to calculate DNA concentration (μg ml^{-1})

The plasmid yield using the Wizard ® SV System (Promega) was initially around 20 μg μl^{-1} but increased to around 1000μg μl^{-1} later in the study. The pGEX4T-1 plasmid exhibited a difference in gel migration due to the physical state of the plasmid (i.e. whether it was supercoiled or linearised by restriction digest). The supercoiled plasmid migrated slower and appeared to be therefore larger than the linearised vector (Figure 38).

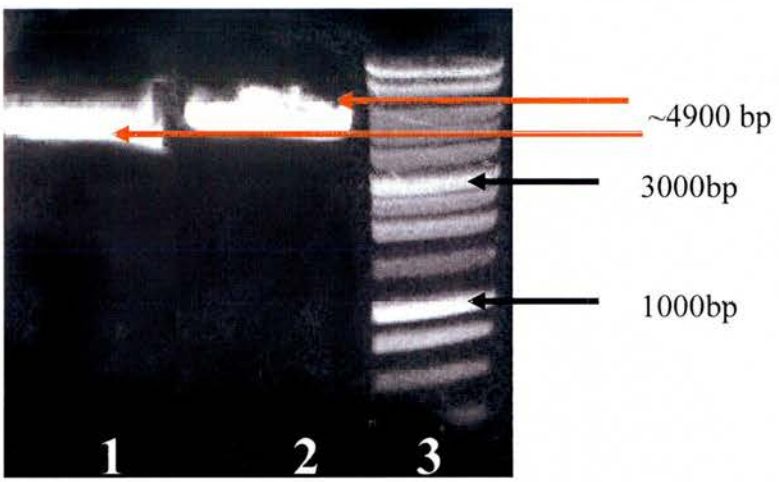


Figure 38: Linearised (Lane 1) and supercoiled (Lane 2) pGEX4T-1 plasmid with a 1 kb DNA marker (Lane 3: Promega) in a 0.7 % agarose gel.

3.4.2 Insert Analysis

Sequence analysis, using DNAMAN software, of the Carcinin inserts revealed no recognition sites for either thrombin, BamH1 or Xho1, in either insert sequence, except for those intentionally introduced at the 5' and 3' ends (Appendix 21). In addition, no

fortuitous Shine-Dalgarno sequences (Ivanov *et al.*, 1992), formed by tandemly repeated rare codons, were noted in these sequences.

Predicted sequences of PCR products are detailed in section Appendix 17. The expected band sizes of these products are detailed in Table 21 below.

Fragment Description	Base pairs (bp)
"full" fragment	**330**
"active " fragment	**267**
PCR modified "full" coding sequence (FCAR)	**372**
PCR modified putative "active" coding sequence (ACAR)	**309**
digested FCAR	**362**
digested ACAR	**299**

Table 21: Expected base pair sizes of fragments and PCR products of modified Carcinin sequence inserts

Successful PCR modification reaction was confirmed by gel analysis of the products (Figure 39 and Figure 40).

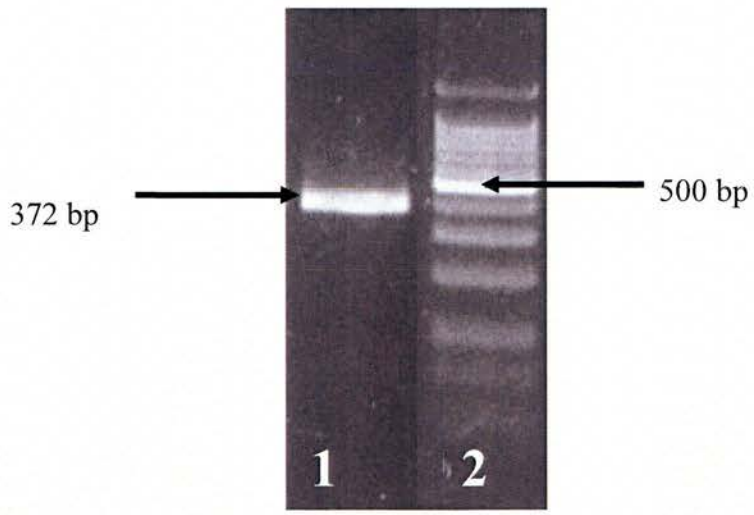


Figure 39: Image of gel purified, modified FCAR insert (372 bp) using FCAR primers (3.3.3.1) on a 2 % agarose gel compared to a 100 bp DNA marker (Lane 2: Promega).

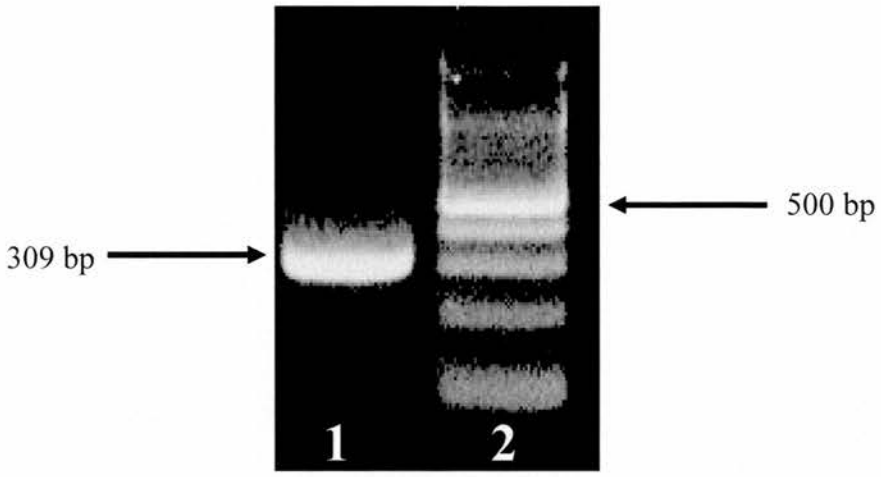


Figure 40: Image of gel purified, modified ACAR insert (309 bp) using ACAR primers (3.3.3.1) on a 2 % agarose gel compared to a 100 bp DNA marker (Lane 2: Promega).

Confirmation of the modification of the inserts was obtained by subsequent sequence analysis on the resultant constructs (see section 3.4.3 below).

Spectrophotometric analysis (A_{280}), of the recovered PCR products, was noted to be highly variable and unreliable as a form of measurement and recovered product was quantified by band intensity comparison to known masses, using the 100 bp DNA marker from New England Biolabs (UK) Ltd. (Figure 43).

3.4.3 Construct Analysis

Transformed bacterial colonies (*E.coli* BL21 and JM105) were screened by PCR amplification (using the "cocktail stick" method described in Chapter 2; 2.3.6) with pGEX4T-1 sequencing primers (Figure 32) to confirm the successful ligation and transformation of the expression vectors. Confirmation of the presence of the correct insert in the extracted plasmids, was by observation of appropriate band size (Table 22) on agarose gel electrophoresis (see Figure 42 and Figure 43).

Fragment Description	Base pairs (bp)
pGEX4T-1 unaltered plasmid	4969
FCAR ligated plasmid	5309
ACAR ligated plasmid	5244
PCR product of FCAR ligated plasmid (pGEC4T-1F)	505
PCR product of ACAR ligated plasmid (pGEX4T-1A)	442

Table 22: Base pair sizes of ligated plasmids and PCR products of each insert using pGEX4T-1 sequencing primers.

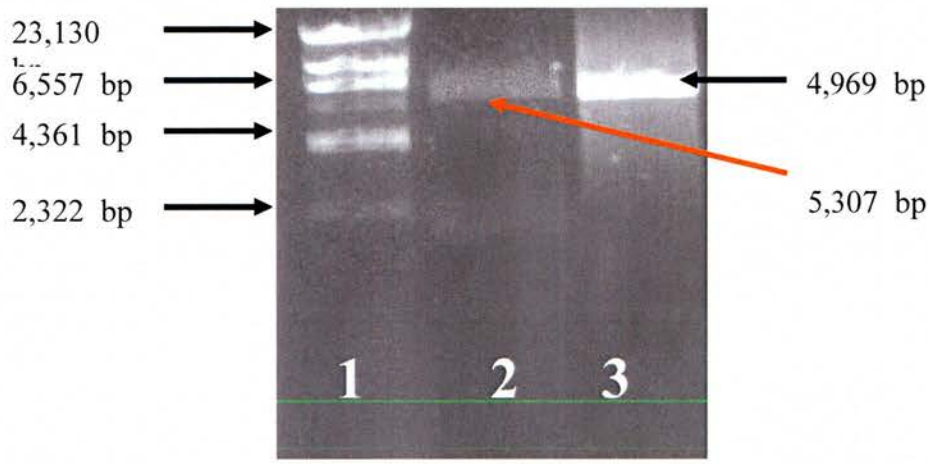


Figure 41: Ligated supercoiled pGEX4T-1F vector recovery after miniprep (Lane 2) with supercoiled original pGEX4T-1 vector (Lane 3) and Hind III digested DNA marker (Lane 1: Promega)

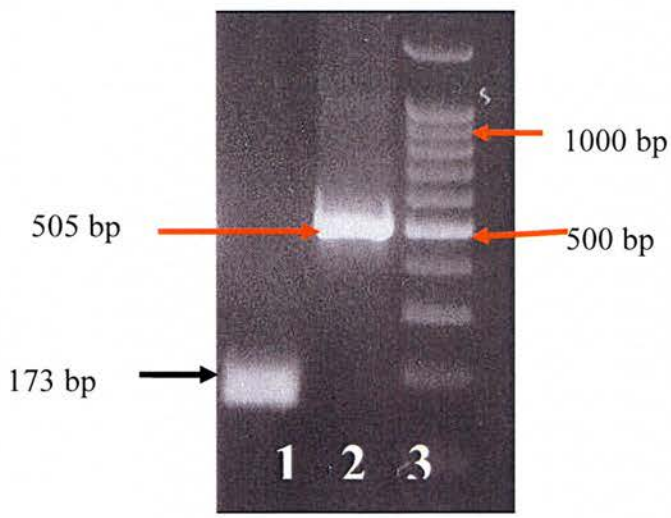


Figure 42: Colony screen of pGEX4T-1F transformed bacteria by PCR showing positive FCAR inserts (Lane 2 @ 505 bp) and a negative clone (Lane 1 @ 173 bp) with a 100 bp DNA marker (Lane 3: Promega).

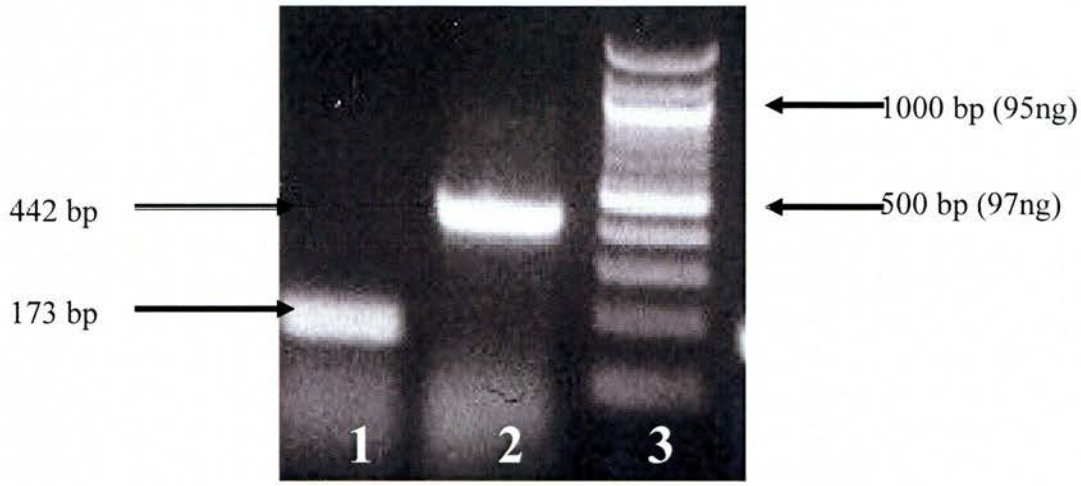


Figure 43: Image of excised and gel purified amplified ACAR insert (Lane 2: 442 bp) and a negative insert reaction (Lane 1: 173 bp) on a 2 % agarose gel compared to a quantified 100 bp DNA marker (Lane 3: New England Biolabs).

Successful PCR modification resulted in the restriction sites added to either end of the insert as shown in Figure 44.

FCAR between sequencing primers Vector –BamH1 site - ATG – 327 bp –TAGACATCGCAGACCCGTGTAA - Xho1 site –Vector
ACAR between sequencing primers Vector –BamH1site - AGG – 264 bp –TAGACATCGCAGACCCGTGTAA - Xho1 site- Vector

Figure 44: Overview of expected sequences obtained by PCR of ligated plasmids

Sequencing result files of the pGEX4T-1F and pGEX4T-1A constructs were analysed using DNAMAN software and were confirmed to have the expected sequences (Figures 44 to 46)

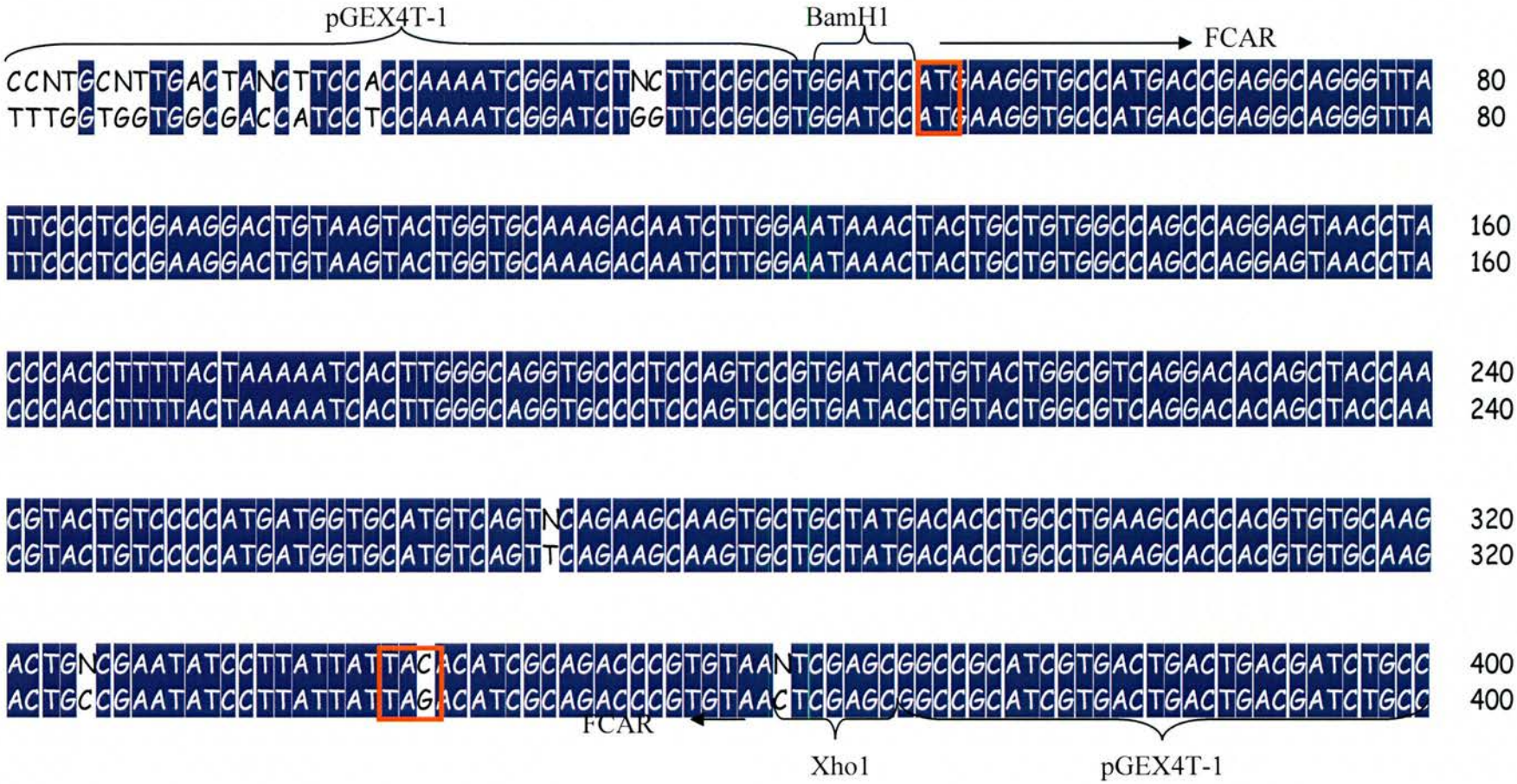


Figure 45: Raw sequence call data confirming in frame ligation of FCAR into pGEX4T-1F using both Fwd and Rev sequencing primers with red boxes indicating first codon and stop codon.

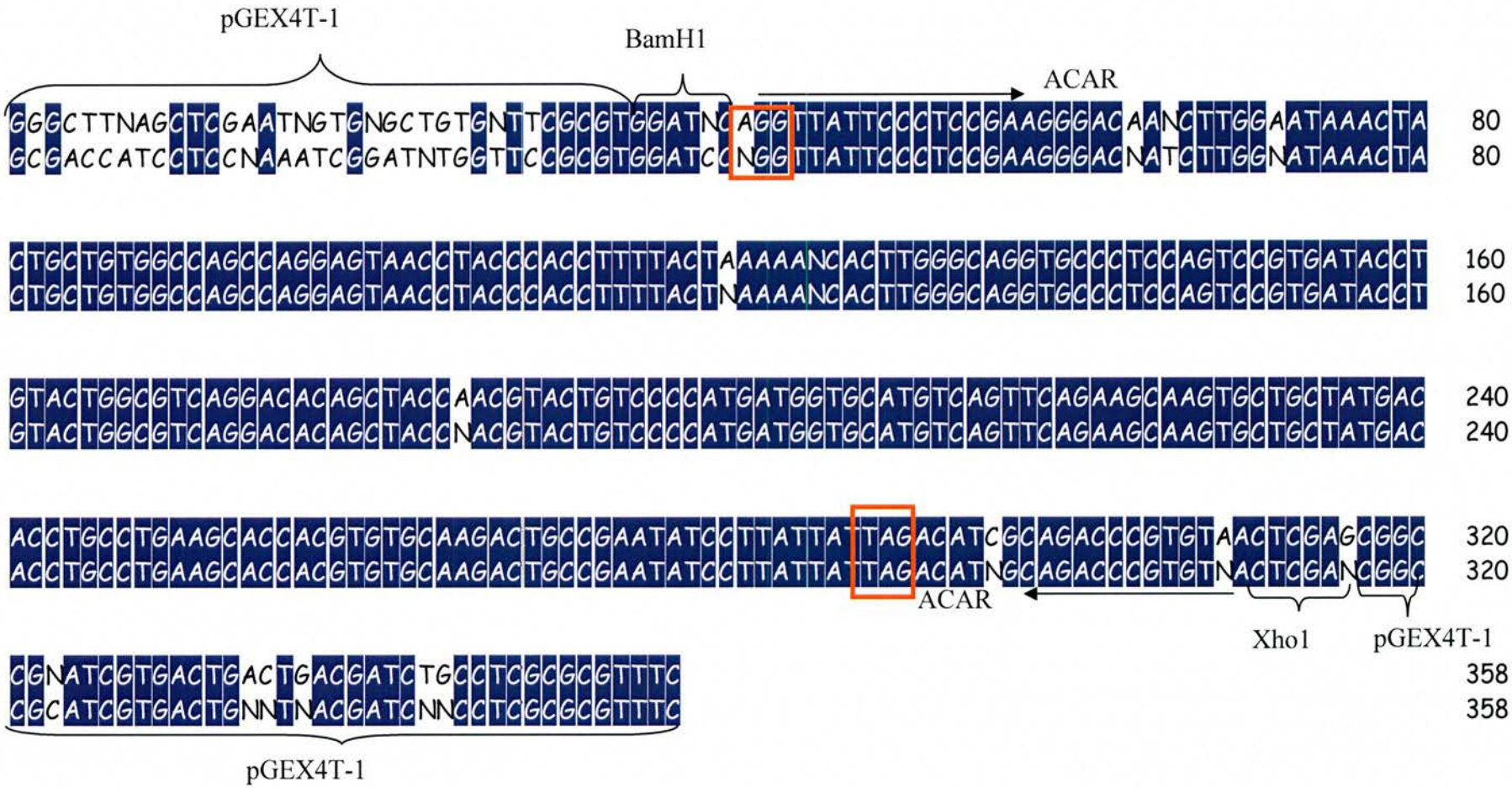


Figure 46: Raw sequencing call data confirming in frame ligation of ACAR into pGEX4T-1A using both Fwd and Rev sequencing primers with red boxes indicating the start and stop codons.

These sequences above confirmed the inclusion of the restriction sites on either insert at the 5' and 3' ends by successful restriction digest and subsequent ligation of insert maintaining the reading frame for the Carcinin sequences.

The nucleotide calls from the FCAR and ACAR raw sequencing data obtained (Figure 45 and Figure 46), were analysed using the Chromas and DNAMAN software packages. Erroneous calls were corrected before the two sequences were translated to show expected amino acid sequences when expressed and cleaved from the GST fusion tag with thrombin (Figure 47 and Figure 48). Both of these sequences coded for isoform 4 of Carcinin (AJ821889) (Appendix 10).

```
1          GGATCCATGAAGGTGCAAACTGTAGCAGCCGTGGTGGTTGTGGCTGTGGTTGTGACCATG
1           G  S  M  K  V  Q  T  V  A  A  V  V  V  V  A  V  V  V  T  M

61         ACCGAGGCAGGGTTATTCCCTCCGAAGGACTGTAAGTACTGGTGCAAAGACAATCTTGGA
21          T  E  A  G  L  F  P  P  K  D  C  K  Y  W  C  K  D  N  L  G

121        ATAAACTACTGCTGTGGCCAGCCAGGAGTAACCTACCCACCTTTTACTAAAAATCACTTG
41          I  N  Y  C  C  G  Q  P  G  V  T  Y  P  P  F  T  K  N  H  L

181        GGCAGGTGCCCTCCAGTCCGTGATACCTGTACTGGCGTCAGGACACAGCTACCAACGTAC
61          G  R  C  P  P  V  R  D  T  C  T  G  V  R  T  Q  L  P  T  Y

241        TGTCCCCATGATGGTGCATGTCAGTTCAGAAGCAAGTGCTGCTATGACACCTGCCTGAAG
81          C  P  H  D  G  A  C  Q  F  R  S  K  C  C  Y  D  T  C  L  K

301        CACCACGTGTGCAAGACTGCCGAATATCCTTATTATTAG
101         H  H  V  C  K  T  A  E  Y  P  Y  Y  *
```

Figure 47: Translation of corrected FCAR insert (AJ821889) sequence calls including residues remaining after thrombin cleavage at the 5' end (highlighted in yellow). Pink highlighted codon indicates difference between FCAR and ACAR insert sequences of first residue after putative cleavage site and star (*) indicates in frame stop codon.

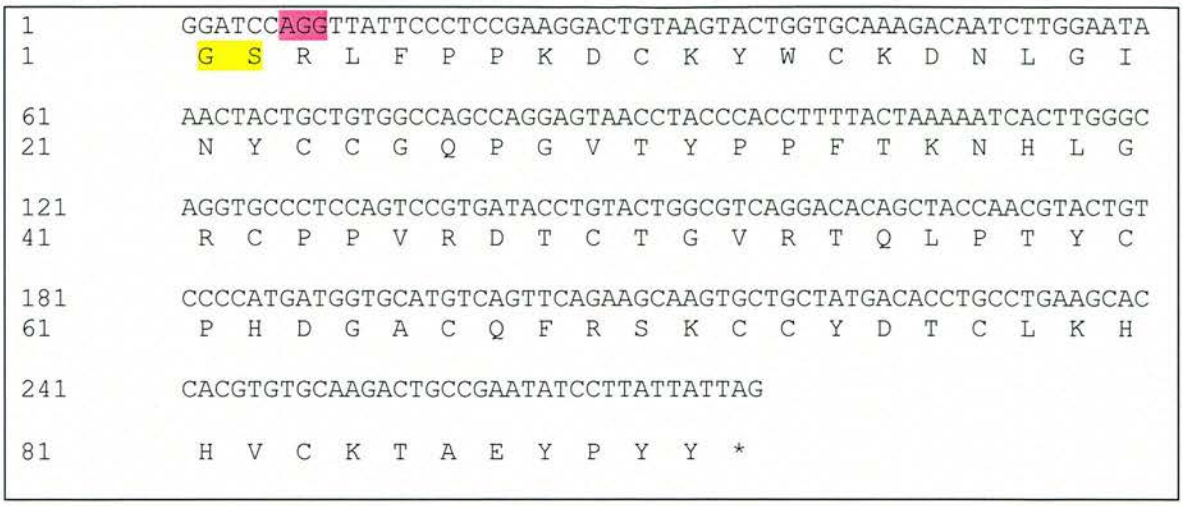


Figure 48: Translation of corrected ACAR insert (AJ821889), sequence calls including the residues remaining at the 5' end after thrombin cleavage (highlighted in yellow). Pink highlighted codon indicates difference between FCAR and ACAR insert sequences of first residue after putative cleavage site and star (*) indicates in frame stop codon.

Due to the vector design, two additional amino acid residues remained attached (highlighted in yellow) to both Carcinin sequences at the 5' end after cleavage with thrombin.

Differences between the sequences at G_{24} (Figure 47) and at R_3 (Figure 48) were also noted. The correct amino acid at this site should have been a glycine (G) (the first residue after the putative signal cleavage site) in both sequences and not an arginine (R) as shown in Figure 48. The error in the Figure 48 sequence was introduced during the design of the ACAR forward primer (see 3.3.3.1) as its design was based on an incorrect nucleotide assignment in the template cDNA sequence (AJ427538), before consensus sequence data was available (see Chapter 2).

3.4.4 Protein Expression

Initial expression cultures (~2 ml) followed the general growth trend presented in Figure 35 although, with smaller culture volumes, the optimum OD (OD_{600} = 0.5) was reached earlier than indicated on the figure.

Post induction samples (Table 10) were analysed by SDS PAGE and the resultant band sizes were compared to known protein size markers (BioRad Broad range marker) as illustrated below in Figure 49. Some baseline expression of GST was observed in the pre-induction sample (Lane 4), but not in the pre-induction of GST-FCAR (Lane 2). Initial expression of the fusion protein was immediately successful (Lane 1) but with relatively poor yield (Figure 49).

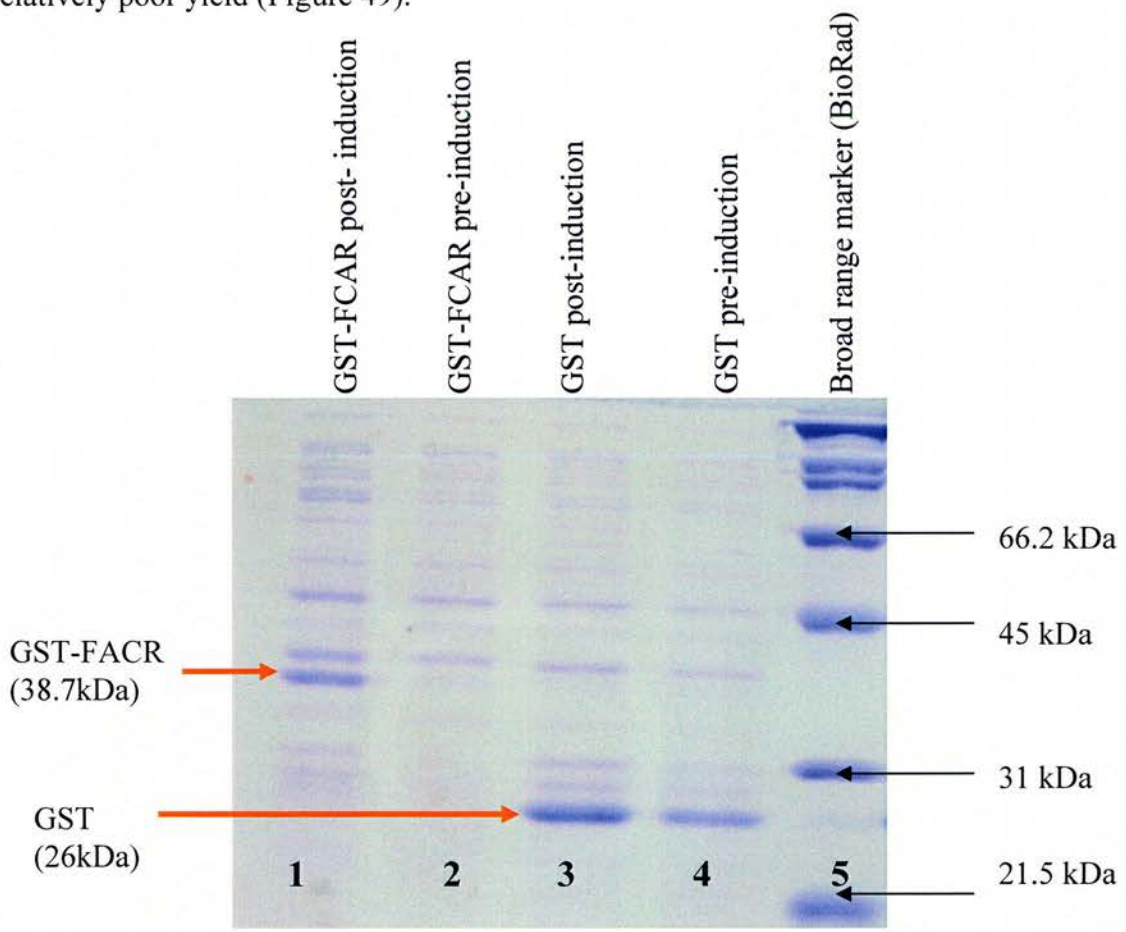


Figure 49: SDS PAGE of post- and pre-induction samples for both GST-FACR (38.7 kDa Lanes 1 and 2) and GST control (26 kDa in Lanes 3 and 4) expressed proteins @ 37 °C grown in 2 x YTA.

Initial expression experiments were taken through to the isolation stages and soluble fusion proteins were eluted successfully from the GS4B beads using the basic protocol outlined in section 3.3.6, although yields seemed quite low (Figure 50). To improve the yield of soluble fusion proteins, several expression parameters were optimised.

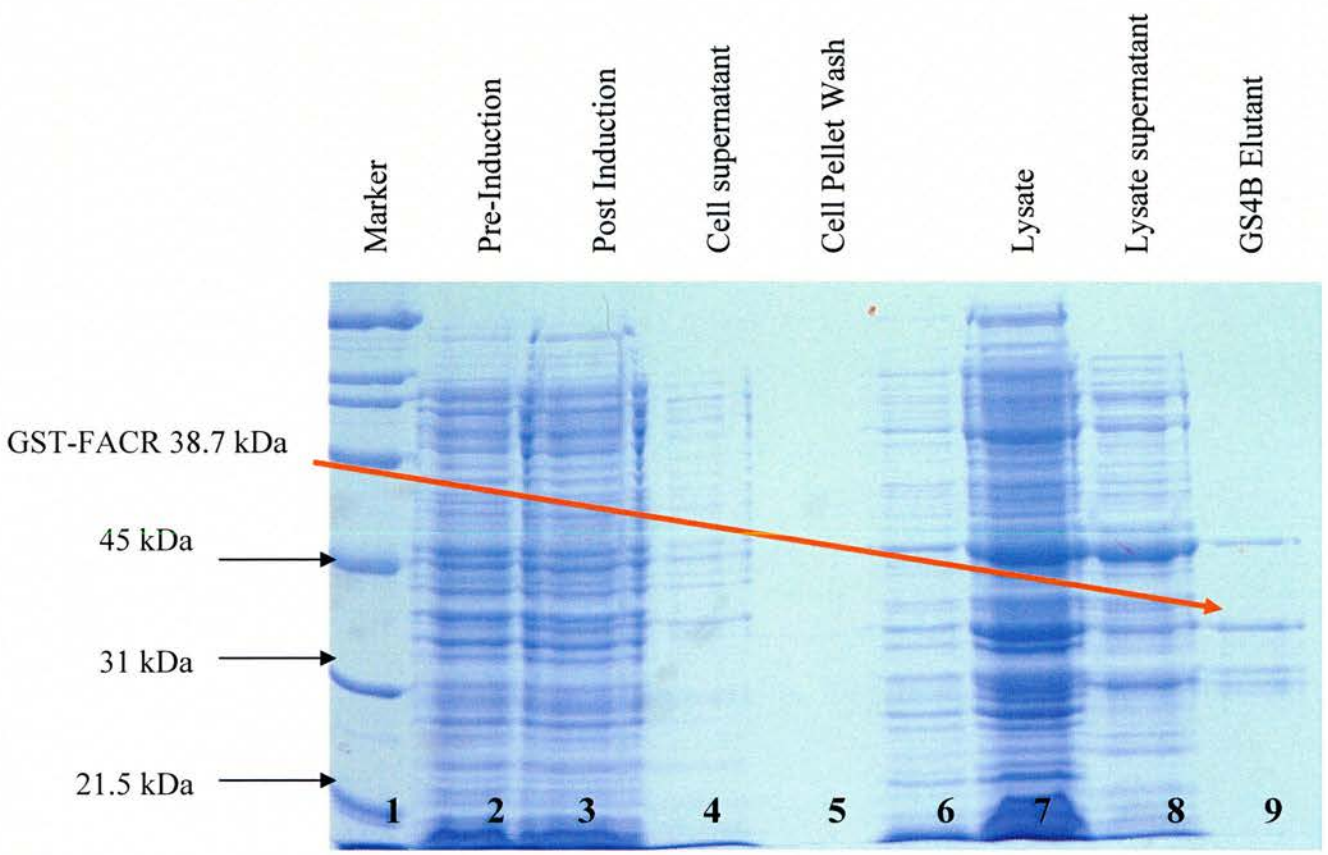


Figure 50: Coomassie stained SDS PAGE gel showing expression of GST-FACR at 37 °C in 2 x YTA from pre induction to affinity elution highlighting poor purification of fusion protein despite increased recovery of fusion protein after lysis condition optimisation.

3.4.5 Growth, Induction and Expression

A proportion of the expressed fusion protein (GST-FCAR and GST-ACAR) was thought to be insoluble (Lanes 7 & 8 of Figure 50) and this was later confirmed (see Figure 56).

The culture medium was changed to include 2 % glucose, which reduced the baseline expression, but did not adversely affect the expression of the fusion proteins (see Figure 49 and Figure 51).

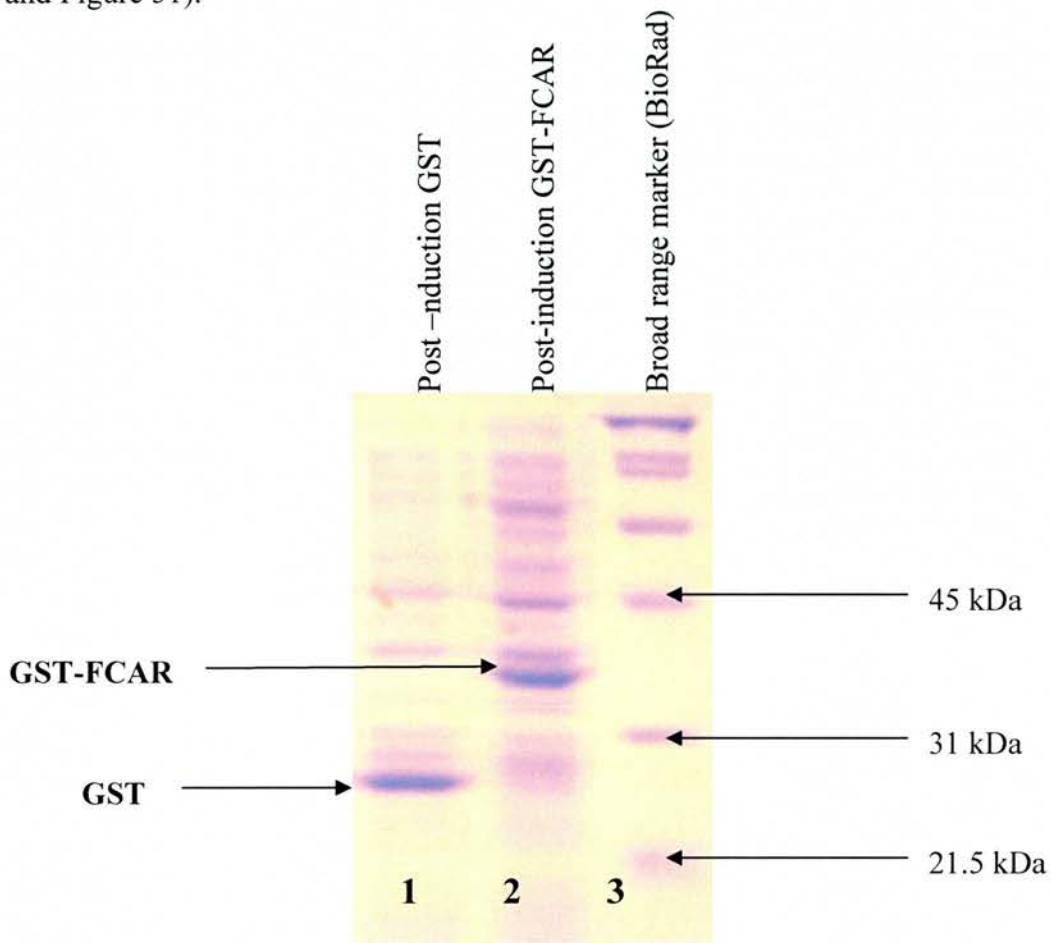


Figure 51: Post induction samples showing soluble expressed GST (Lane 1) and GST-FCAR (Lane 2) grown in 2 x YTAG medium at 20 °C.

Lower culture temperatures have been shown (Schein and Noteborn, 1988) to reduce the formation of inclusion bodies (Sambrook and Russell, 2001) and increase the yield of the soluble portion of the fusion protein. This was also was observed in the present study (i.e. improved yield at 20 °C compared to 37 °C, see Figure 49, Figure 50 and Figure 51).

Although not experimentally proven in the present study, increased aeration, (by increasing culture flask size and rpm) is known to reduce the formation of inclusion

bodies (Amersham Pharmacia Biotech, 2002). The aeration of the cultures grown at 20 °C was maximised by increasing the shaking of the culture to 250 rpm. Increasing the cell density of the expression cultures (OD_{600} @ ~0.9) increased the viscosity of the lysed cells and co-purification of host proteins after elution from the GS4B beads. During this isolation stage, it was observed that the translucency of the bead matrix decreased after the addition of the lysate supernatant and a "skin" formed on the top of the bead pellet even after washing and centrifugation. Lower cell density cultures were used instead (OD_{600} 0.5-0.6) throughout the remainder of the experiment. The time required to reach OD_{600} ~0.5-0.6 was dependent on the volume of the culture with smaller volumes requiring shorter time periods and larger cultures requiring longer.

Although expression of fusion proteins using an IPTG inducible promoter can be achieved with concentration ranges from 0.01 -5 mM IPTG, the ideal concentration must be determined by trial and error for each system (Sambrook and Russell, 2001). In this experiment, increasing the concentration of IPTG to 2 mM, or reducing it to 0.05 mM did not markedly affect the protein yield, as observed by SDS PAGE analysis. Therefore the recommended concentration of 1 mM (Amersham Pharmacia Biotech, 2002), was used throughout.

Different induction times were experimentally tested (1, 2, 3, 8 and 12.5 h using 50-200 ml cultures) with the two *E.coli* strains and two different lysis regimes both expressing the GST-ACAR fusion. Induction for 1 or 2 h, using *E.coli* BL21, produced similar yields (for example of 1 h induction expression see Figure 51) using the recommended resuspension and lysis conditions. Protein yield considerably increased after 3 h and 8 h induction times (see Table 23 and Figure 52: Lanes 1 and 3) and 3-4 h periods were

used in subsequent expression experiments due to concerns over proteolytic degradation associated with longer inductions.

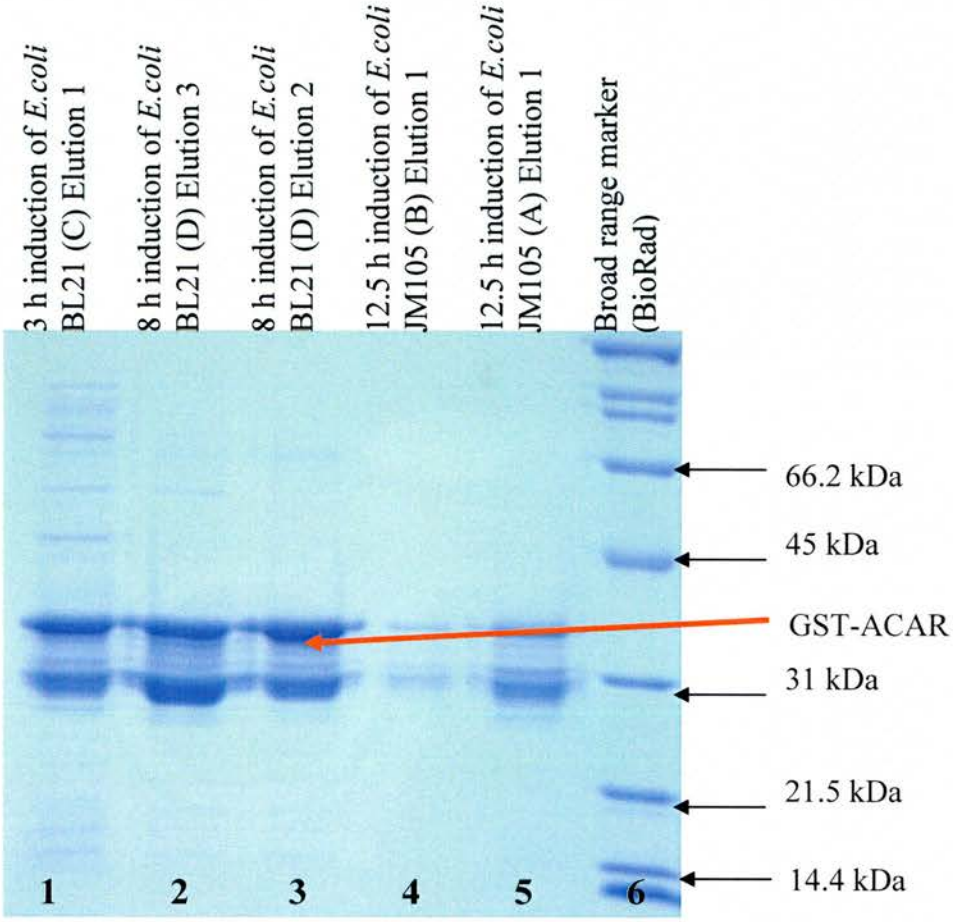


Figure 52: Elution of GST-ACAR (Lanes 1-5) from GS4B matrix after different inductions times and using different strains with increased bead volume and increased number of 1 x PBS washes. The gel also shows evidence to support proteolytic degradation of the GST-ACAR and accumulation of a band < 31 kDa.

The host proteins, visible in Lane 1 of Figure 52 (compared to Lane 6 of Figure 53), were eliminated in later purifications by increasing the number of washes from 3 to 10 of 10 volumes of 1 x PBS before elution stages were carried out.

The longer induction time (8 h), used with the *E.coli* BL21 strain, did result in approximately double the recovered protein concentrations compared to the 3 h induction (Table 23). Also, a greater yield of overall protein was obtained in the initial

elution samples from the 8 h compared to the 3 h induction. However, the longer induction was thought to have resulted in an increase in the degradation of the fusion protein. This was demonstrated by the presence of a strong < 31 kDa band in Lanes 1-3 of Figure 52.

Sample	*E.coli* Strain	Induced (h)	Lysozyme @ 10 mg ml^{-1} +/- sonication	1st Elution + 10 min ($\mu g\ ml^{-1}$)	2nd Elution + 16 h ($\mu g\ ml^{-1}$)	3rd Elution + 20 h ($\mu g\ ml^{-1}$)
A	JM105	12.5	Lys + Son	541.44	376.4	156
B	JM105	12.5	Lys only	167.4	95.4	9
C	BL21	3	Lys only	>1000	831	309
D	BL21	8	Lys only	>1000	1505	458

Table 23: Table of induction parameters results affecting total protein yields in elutants.

Each eluted sample (from the GS4B beads) was quantified using the Bradford method and the Ultrospec 3300 spectrophotometer and the results presented in Table 23. This particular experiment was based on 50 ml cultures of each bacterium. Based on the data shown for samples C and D in Table 23, the yield of total protein per ml of culture medium in this experiment was calculated to be between 2.14 and 2.96 μg. In addition, a large proportion of this was attributed to the additional band at < 31 kDa (Figure 52). Although unconfirmed, this band was thought to be a degraded product of the ACAR-GST.

E.coli JM105 is not a protease deficient strain and not generally recommended for expression of fusion proteins. Comparing the yields obtained in Table 23, even taking

into account the increased proteolysis due to the longer induction for JM105, the difference in expression efficacy between these strains is clear.

Lysate samples in this experiment did not appear to clarify after sonication as suggested in the manual (i.e. max 10 s) (Amersham Pharmacia Biotech, 2002), and it was thought that this method of cell disruption may have been insufficient and the addition of lysozyme (@ ~1 mg ml^{-1}) was investigated as longer sonications lead to frothing of the sample. The results obtained in the present study (Figure 52 and Table 23) highlight the relative differences observed using lysozyme with (A) or without (B) sonication (similar pattern observed for *E.coli* BL21). Sonication with lysozyme increased the yield of protein compared with lysozyme or sonication alone. This combination of lysis techniques was discarded though as there appeared to be increased proteolysis (Figure 52) of the fusion protein. After cleavage, lysozyme bands were also visible in the isolated sample and were not easily removed in subsequent purification steps (see Figure 53). Sonication without the addition of lysozyme was used for lysis steps in subsequent expression experiments.

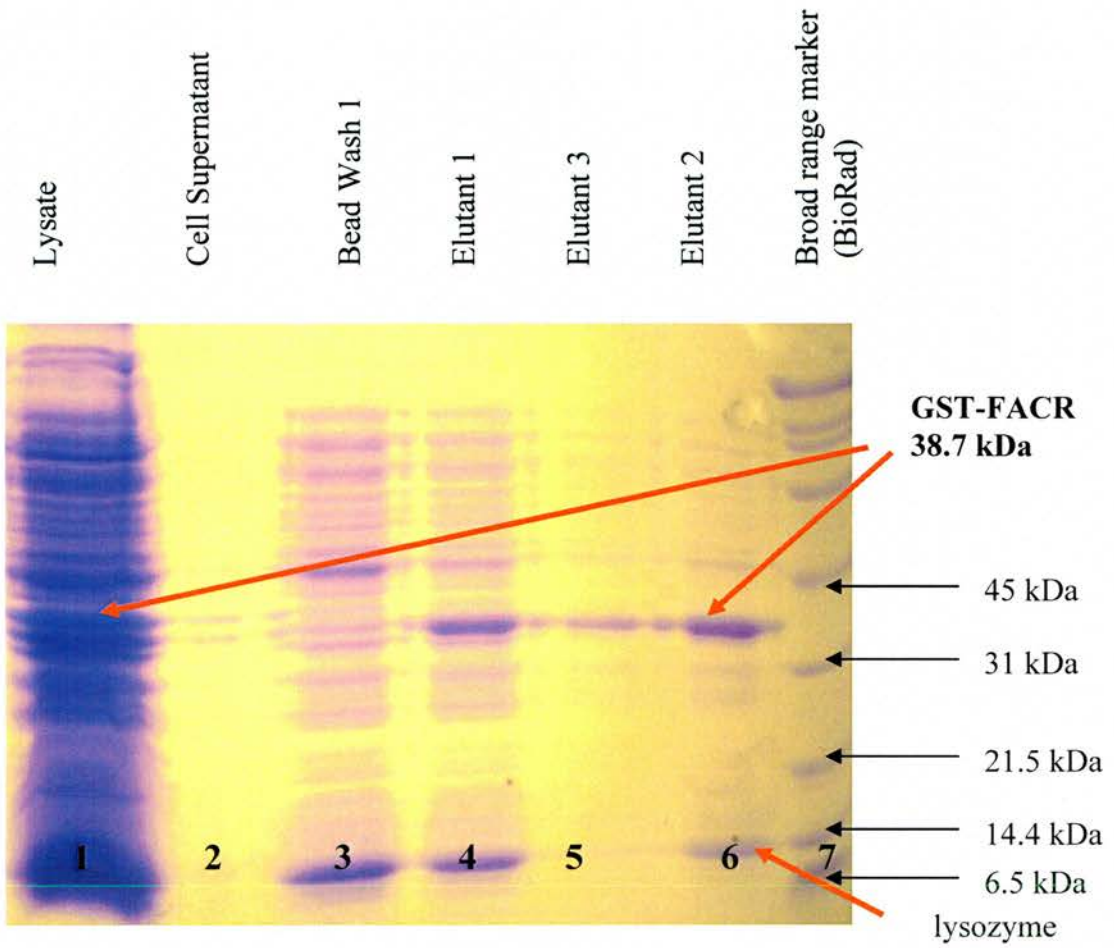


Figure 53: Expression of GST-FCAR fusion protein from control vector showing lysozyme (Lanes 3, 4 and 6) and the presence of extra host bands in elutant due to incomplete washing of GS4B beads after binding of lysate supernatant.

Different sonication regimes were tested to obtain minimum protein denaturation. Protein denaturation of the GST moiety has been suggested to reduce binding efficiency of fusion proteins (Mercado-Pimentel *et al.*, 2002) due to excessive sonication. Frothing of the samples (suggesting denaturation) occurred with the longer sonication regimes (3 m) as well as sonication using the tapered tip directly into the sample (for smaller samples) as observed by others (Frangioni and Neel, 1993). The regime which was eventually used for small scale samples was using the Cup horn (SH213) into a water bath for 30 s at 75 % and 5 s pulses and for the larger scale samples the tapered

probe (MS72) was used directly into the sample which was sonicated for 30 s at 60 % with 5 s pulses.

The other major protein bands which can be seen in Figure 52 (~26-29 kDa), may be present as a result of proteolysis of the fusion protein, as observed by others using a GST fusion tag expression system (Piers *et al.*, 1993; Zhang *et al.*, 1998), resulting in several bands of smaller proteins below the fusion protein (~38 kDa). To combat this, a protease inhibitor (PMSF @ 1 mM), was added to the lysed cells but had no great effect on the elution profiles analysed. PMSF is rapidly inactivated in aqueous solutions (Sambrook and Russell, 2001) (20 mM has a half life of only 35 min according to James (1978)) and may not have maintained its protease inhibitory activity in the later isolation steps leading to degradation of the fusion protein.

RNAse and DNAse (freshly prepared at 5 $\mu g\ ml^{-1}$ each) added to the lysed cells seemed to reduce the viscosity but the results were inconsistent.

The use of STE buffer with detergents (e.g. N-lauryolsarcosine with DTT) and sequestering agents (e.g. Triton X-100) appeared to increase the yield of fusion protein as described by others (Frangioni and Neel, 1993; Mercado-Pimentel *et al.*, 2002), but also increased the yield of the host proteins purified along with them (as illustrated by Figure 53 and Figure 54) and this lysis regime was not used with subsequent experiments.

The use of Triton X-100 (1 % v/v) in the resuspension buffer (1 x PBS) to help disaggregate the proteins, did not appear markedly change the yield of the fusion protein as judged by band intensity on SDS PAGE.

Increasing the centrifugation speed after lysis to 31,000 x g compacted the cell debris into a more solid pellet and reduced carryover to the isolation stages as observed in Figure 52 compared to Lane 4 of Figure 54.

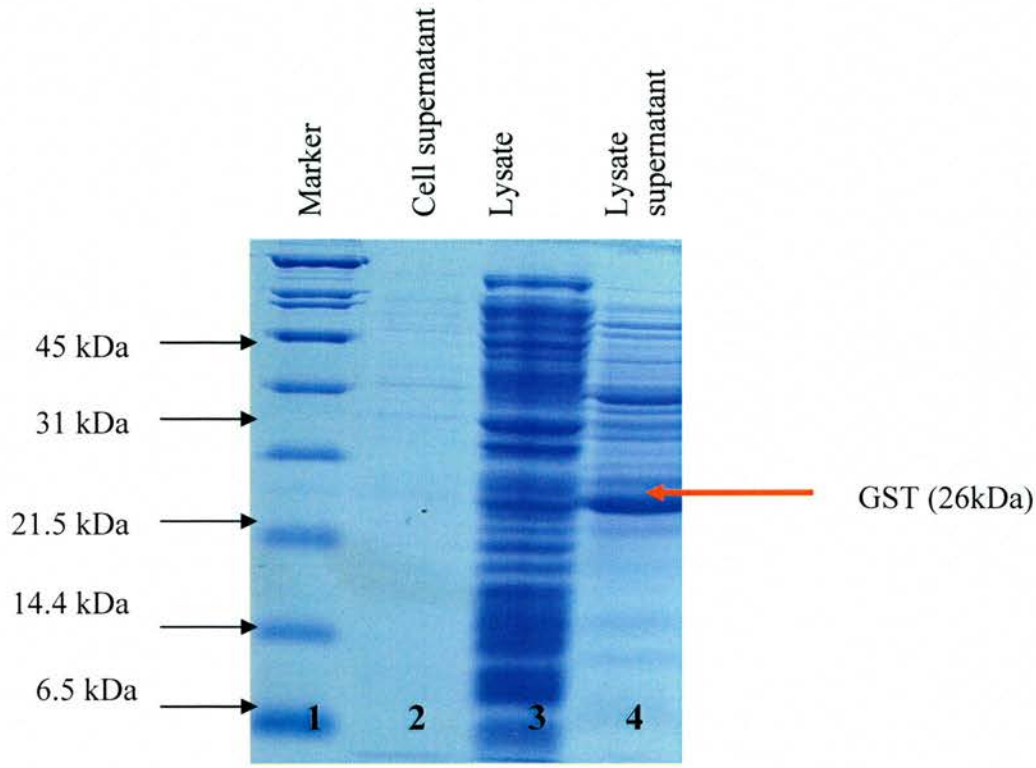


Figure 54: Several large molecular weight host proteins remaining in the lysate supernatant (Lane 4) of GST control expression using the STE lysis buffer, which were reduced by increasing the centrifugation speed to 31,000 x g using 1 x PBS lysis buffer instead.

Filtering the lysate supernatant trough a 45 μm filter improved the clarity of the sample by removing particles which were observed to clog the GS4B beads.

3.4.6 Isolation

The addition of Triton X-100 to the lysate supernatant has previously been shown to improve solubilisation of the fusion protein (Amersham Pharmacia Biotech, 2001) and this was used with large scale purifications although no dramatic differences were noted in the yield in this experiment.

Yield was improved though, using larger bed volumes (GS4B) than recommended in the protocol (Amersham Pharmacia Biotech, 2002) as observed by Mercado-Pimentel *et al.* (Mercado-Pimentel *et al.*, 2002). For small scale cultures, up to ten times the

recommended bed volume was used in isolation stages after which this was limited by the dilution effect on the protein by increasing the elution buffer volumes. With larger culture volumes the maximum amount of resin used was limited by its availability and 2 ml (2 x as recommended by the manual) of resin (bed volume) was the maximum ever used for large scale cultures (~2 l).

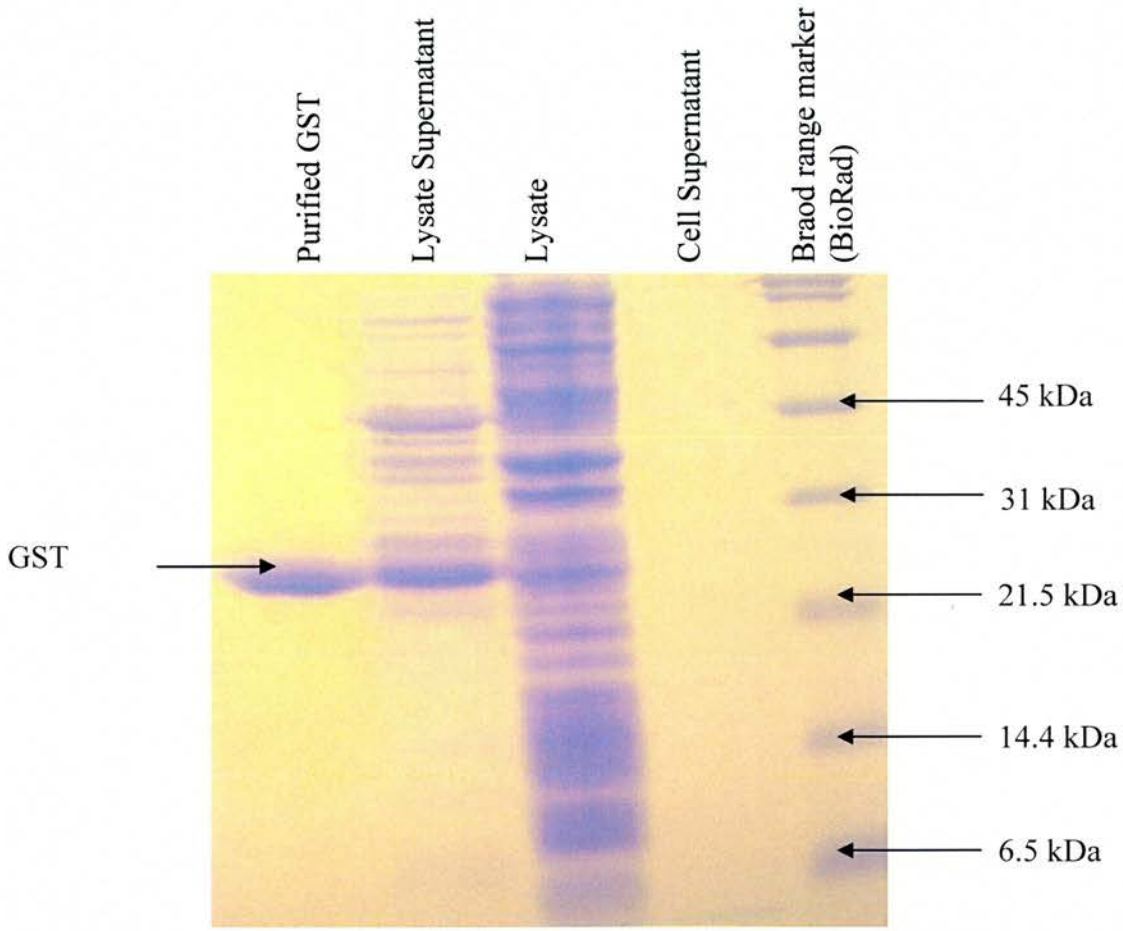


Figure 55: Coomassie stained SDS PAGE gel showing successful expression, lysis without lysozyme and affinity purification of GST protein (using GS4B beads) from pGEX4T-1 control vector with a BioRad Broad Range marker

The binding of cell lysate supernatant with the GS4B beads for 30 min improved the eluted yield of the fusion protein compared to 5 min incubations (see Figure 55 compared to (Figure 51) small and large scale cultures. As incubations were conducted

at room temperature and proteolysis was thought to be a problem affecting the yield, longer (>30 min) incubation times were not tried.

Increasing the number of washes of the beads after binding considerably reduced the number of host proteins found in the elutants (compared Lane 1 Figure 52 to Lane 6 Figure 53).

Varying the incubation times of the elution buffer with the fusion bound GS4B beads showed continued elution of the fusion protein up to 5 elutions later, with incubation times varying from 5 min to 24 h. As longer incubation with the elution buffer could also increase proteolysis of the fusion protein; elution times were kept to 30 min and the number of elutions kept to a maximum of 5.

Regeneration of the GS4B beads appeared to be most successful when using the 0.2 M NaOH protein denaturing wash. This was judged by the reduction in binding efficiency of the regenerated beads leading to loss of GST-FCAR in the bead supernatant when using the other protocols.

The solubilisation of fusion proteins from the lysate pellet using at 2 M and 4 M urea resulted in the highest yields of soluble fusion protein as analysed by SDS PAGE. Subsequent binding to the GS4B beads was not effective possibly due to denaturation of the GST.

Initial purifications showed incomplete cleavage of the fusion proteins due to the addition of only 25 units of thrombin. Successful cleavage of the eluted fusion protein ("off-beads" sample) was achieved with 80 units of thrombin protease per ml of bed volume (GS4B) as shown in Figure 56 (Amersham Pharmacia Biotech, 2002).

Thrombin (diluted in 2-3 ml 1 x PBS) added to the lysate pellet also confirmed considerable insolubility of the fusion protein.

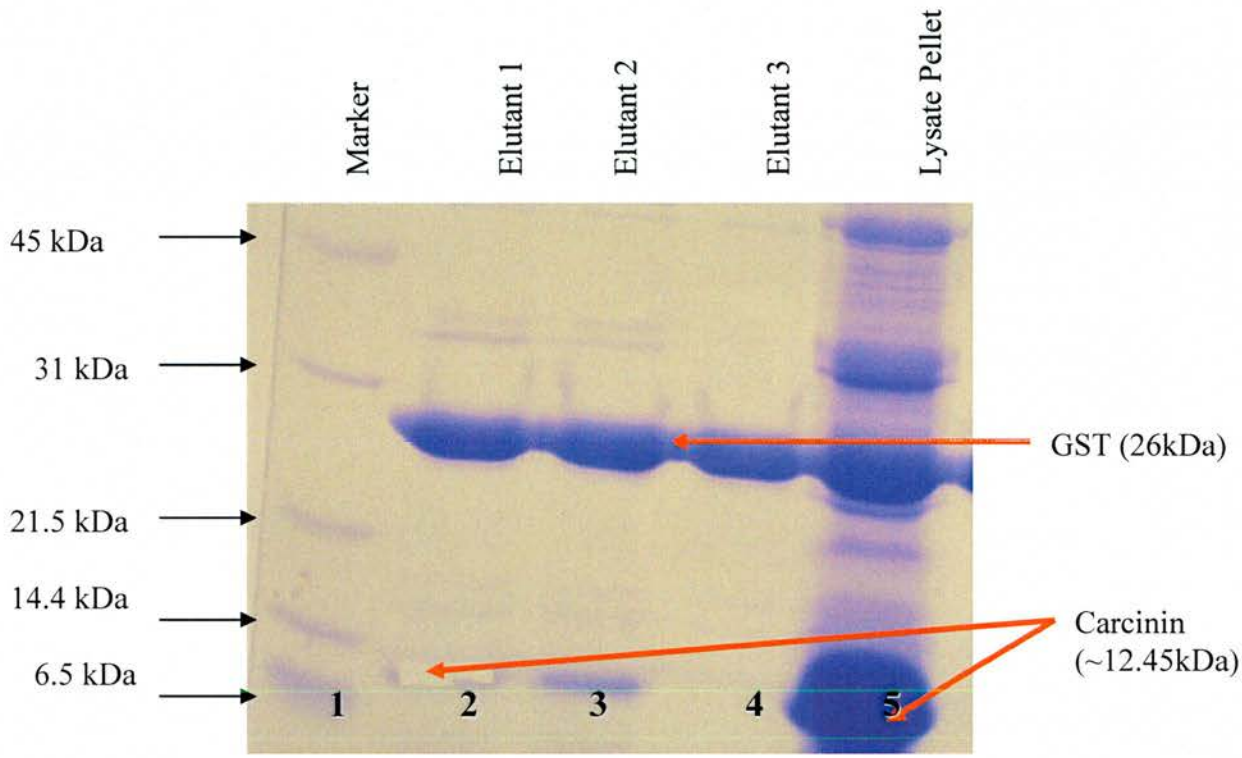


Figure 56: Confirmed cleaved ACAR Carcinin (by protein mass fingerprinting) from the fusion protein elutants. The lysate pellet sample clearly shows a substantial amount of insoluble or bound Carcinin lost in the lysate pellet. Digest times were varied, 6 h (Lane 2), 10 h (Lane 3) and 20 h (Lane 4) using 80 units thrombin.

Extended digestion times led to the degradation of the Carcinin moiety suggesting instability of the recombinant protein after cleavage.

Cleavage was also conducted on the bound fusion protein ("on-bead"), this resulted in extremely poor yields (see Figure 58) with some low molecular weight proteins (<~12 kDa) present below the putative Carcinin band. Reduced cleavage efficiency of the bound protein was not due to the presence of detergents as suggested by Dian *et al.* (2002b) as these additives were only tested with "off-bead" samples. The elution ("off-

bead") method was therefore used solely for the final large scale expression experiments.

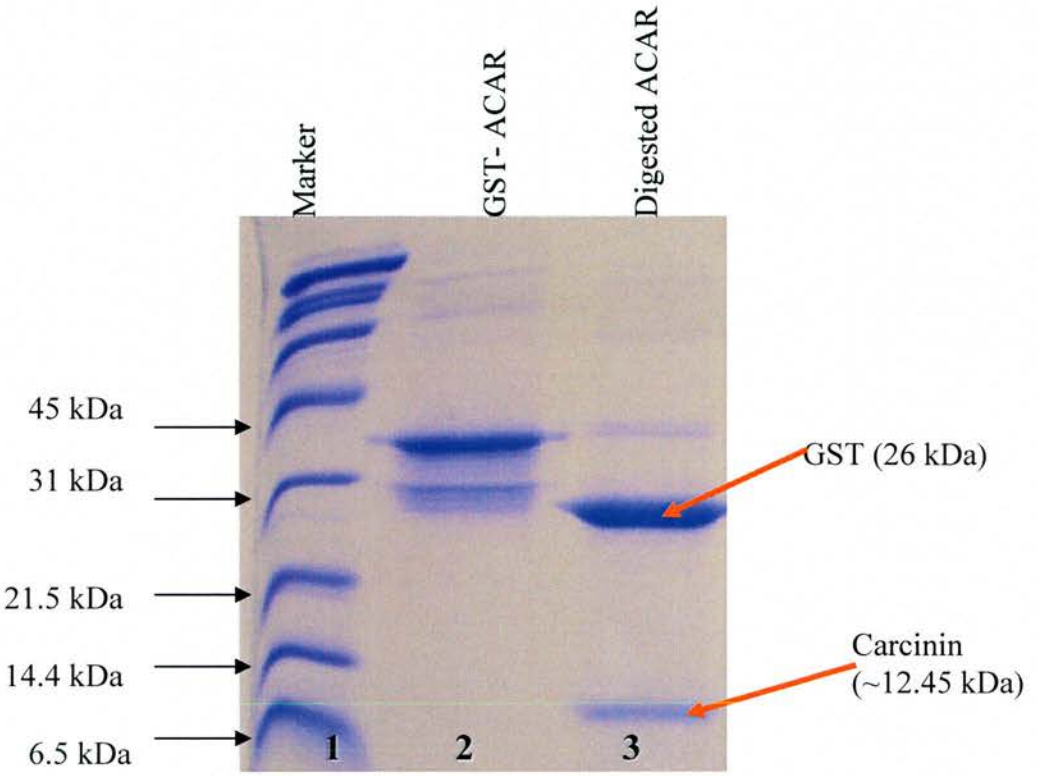


Figure 57: Cleavage of GST – ACAR showing eluted fusion product (Lane 2) and digested sample (Lane 3)

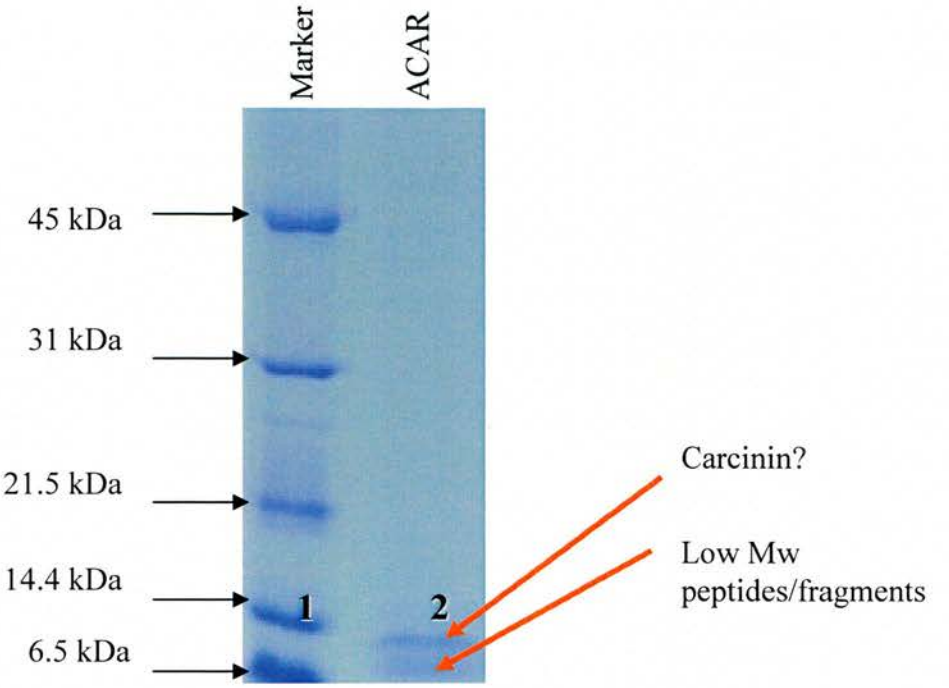


Figure 58: Released Carcinin from bound fusion protein cleavage (resulting from a 2 l culture) showing no residual thrombin but some very low Mw proteins.

3.4.7 Purification

Initially both the GST-FCAR and the GST-ACAR fusion proteins were expressed. The purification stages were less successful when using the GST-FCAR than the GST-ACAR and in the later purifications only the GST-ACAR fusion was used.

Dialysis of the digests resulted in some loss of protein from the "off-bead" samples as analysed by SDS PAGE, although this was not quantified. Dialysis of the "on-bead" sample resulted in complete loss of the Carcinin band even with large scale cultures (Figure 58). Only the elution method ("off-bead") was used for all subsequent purifications.

Further purification was required to remove thrombin, GST and any remaining host proteins from the "off-bead" samples and this was achieved by a combination of chromatographic methods.

3.4.7.1 AFFINITY CHROMATOGRAPHY

Removal of cleaved GST from "off-bead" samples after dialysis was not completely successful with considerable amounts of GST remaining in the supernatant (Figure 59). It was thought that the regenerated beads used, were incompletely regenerated, leading to inefficient binding and removal of GST. This theory was investigated using freshly prepared, unused GS4B beads with the same sample and similar results. Elution of GST fusion proteins from GS4B beads using reduced glutathione buffers may denature the GST moiety and lead to inefficient binding to the column matrix on the second occasion.

Flowthrough samples were then dialysed against the benzamidine buffer and applied to benzamidine Sepharose columns to remove excess thrombin and resultant fractions analysed by SDS PAGE.

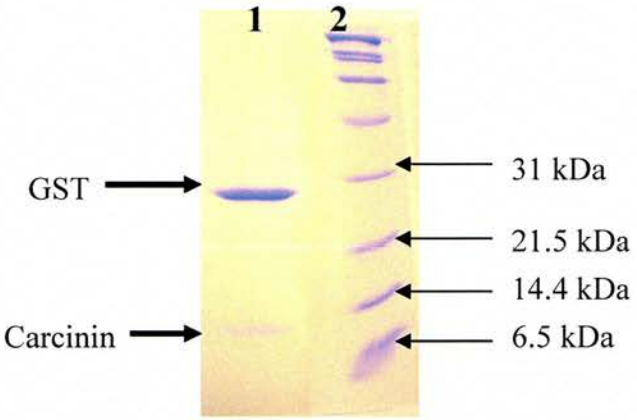


Figure 59: Removal of thrombin but unsuccessful removal of GST after glutathione and benzamidine and glutathione affinity chromatography

3.4.7.2 FPLC

Separation of the GST from Carcinin using HR5/5 MonoS (strong cation exchanger) or MonoQ (strong anion exchanger) chromatography columns was also unsuccessful and sample was lost at each stage (Figure 61) of the process.

The MonoS column was expected to retain cationic proteins (i.e. Carcinin) with neutral (i.e. GST) and any anionic proteins (i.e. GST-ACAR) collected in the flowthrough. The fractions were analysed by 12 % SDS PAGE and stained both with Coomassie and silver nitrate to increase detection resolution. Only a few fractions stained positive for proteins at the expected sizes for both GST and Carcinin, but no separation between them was achieved and the proteins were eluted in the same fractions at ~15-20 % B. Figure 61 illustrates the fractions obtained from a sample previously identified to contain both cleaved and uncleaved fusion protein and analysed by SDS PAGE.

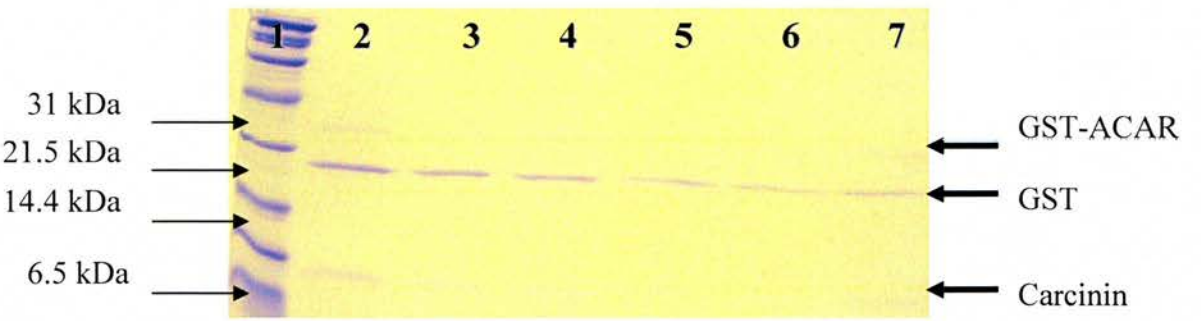


Figure 60: Fractions after MonoS FPLC of samples obtained using minimum thrombin enzyme to cleave fusion protein showing intact fusion protein, GST and cleaved Carcinin but no effective separation between all three.

The band around ~37 kDa is visible in some of the lanes (e.g. Lanes 2-4) and was thought to be uncleaved GST-ACAR. This was due to insufficient amounts of thrombin (~25 units instead of 80 units) having been added to the eluted fusion protein. It can be

noted that the majority of the GST is eluted with the Carcinin and the GST-ACAR in Lanes 2-4.

The MonoQ column was expected to retain negatively charged proteins with the cationic Carcinin appearing in the flowthrough and the GST and any remaining GST-ACAR eluting in later fractions. This profile was observed when fractions were analysed by SDS PAGE (Figure 61).

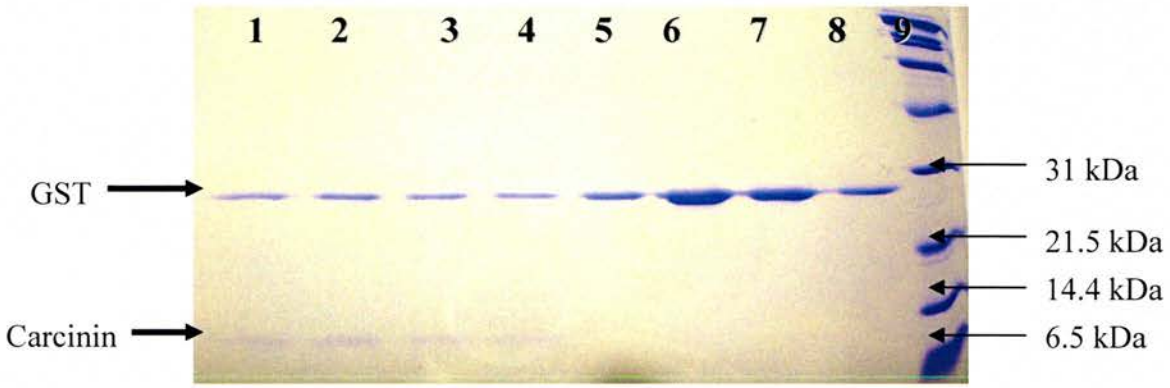


Figure 61: Fractions collected after affinity chromatography using MonoQ anion exchanger showing unsuccessful separation of GST from Carcinin and loss of sample.

The bulk of the GST (Lanes 5-8) did not separate from the Carcinin (Lanes 1-4). Bands of GST size can still be seen in Lanes 1-4 eluted at ~10-30 % B. Overall, the MonoQ anion exchanger produced fractions with better separation between the two proteins than the MonoS. This result may indicate some proteolysis of the samples or effect of denaturation.

3.4.7.3 RP-HPLC

Initial Coomassie stained SDS PAGE gels of the HPLC fractions, did not indicate the presence of any bands.

Analysis by SDS PAGE confirmed the presence of a very faint band around ~10-12 kDa and a more obvious one at ~26 kDa. Separation between these bands was not achieved using this method as the proteins appear to elute in the same fractions

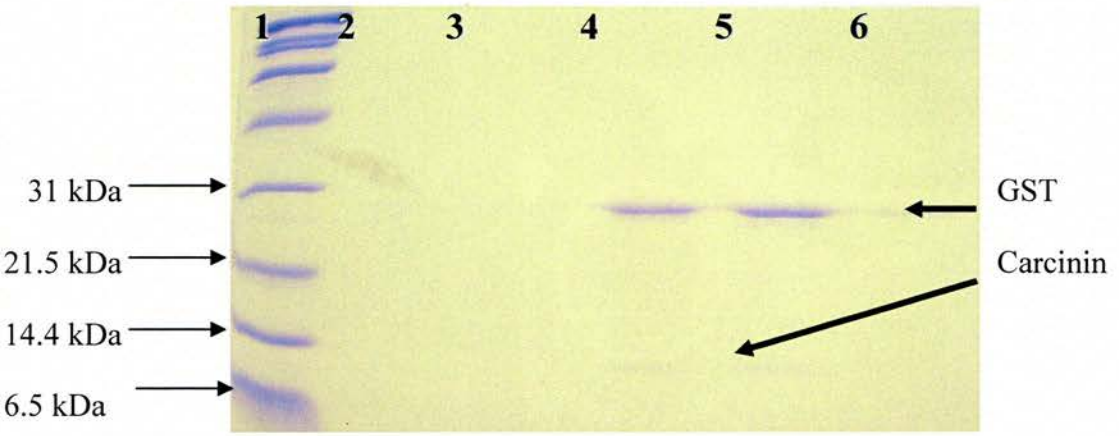


Figure 62: Lyophilised, resuspended HPLC fractions (L to R) 50, 51,53, 55 and 57 showing possible bands for GST (~26 kDa) and Carcinin A (~12 kDa) in Lanes 4 and 5 (i.e. fractions 53 and 55).

The elution profile obtained a small peak is visible for the first fractions and the remainder is eluted at 100. These proteins were eluted at two concentrations of buffer B ~35-37 % B and at 53-55 % B unlike the RP-HPLC results for the native Carcinin which eluted at 40-45 % B (Relf *et al.*, 1999).

3.4.8 Product Analysis

3.4.8.1 RDA

Proteins of~12 kDa eluted from the RP-HPLC above (section 3.4.7.3) did not appear to exhibit any antibacterial activity.

3.4.8.2 PROTEIN MASS FINGERPRINTING

Protein mass fingerprinting results confirmed the presence of Carcinin (CAB51030) in the band excised from Lane 2 of Figure 56. The Mascot analysis indicated that the highest Mowse scores were 114/140 and 138/140 with 8 and 9 peptides matched

respectively to entries gi 18157188 and gi 5531235 on NCBInr 20021205. The best match was achieved with 86 % sequence coverage.

3.5 Discussion

The main aim of the experiment was to use recombinant bacterial expression to obtain sufficient quantities of the Carcinin protein for use as antigen in polyclonal antibody production. Recombinant fusion proteins of Carcinin, an antibacterial protein from *C. maenas,* were easily produced but with relatively low yields. Considerable effort was invested to increase the yields and this was partly successful. However the stability of some of the expressed protein sequences hindered the production of sufficient protein to allow its use in the production of polyclonal antibodies.

Purification and characterisation of native Carcinin, as described by Relf *et al.* (1999), was shown to be resource limited. In addition, isolating the quantities required for antibody production was anticipated to be difficult to achieve by purification. Initial *in silico* sequence analysis (Chapter 2), suggested that the putative "active" protein may be unstable which can further diminish the yield of protein obtained by purification (Sambrook and Russell, 2001). Synthetic production was considered to be too expensive for the full sequence and no data was available on the antigenicity of the amino acid sequence to choose an appropriate fragment of this for production. Although yeast expression had been used to successfully express an active AMP from *L. vannamei* (Destoumieux *et al.*, 1999), maintaining the activity of the protein by folding or post translation modifications was not considered to be essential for the purposes of the present study. These issues contributed to the selection of a bacterial expression system for the expression of Carcinin.

Several antimicrobial peptides have recently been successfully over expressed using bacterial fusion systems (Piers *et al.*, 1993; Skosyrev *et al.*, 2003). Some of the resultant proteins have retained their activity, although this may be due to the relatively

simple structural characteristics of the proteins investigated. The selection of a fusion tag system in the present study was guided by the simplicity of the purification methods and the possible stabilising properties of the tag on the "active" sequence.

The pGEX vector family (Amersham Biosciences) in particular, was selected for the present study as these vectors are well documented and have been previously used to express other cationic antibacterial proteins (Piers *et al.*, 1993; Zhang *et al.*, 1998; Skosyrev *et al.*, 2003). The pGEX4T-1 vector was selected as the cleavage enzyme (thrombin) was relatively inexpensive and the purification methods associated to the GST system were readily available. Removal of the GST tag from the Carcinin protein using thrombin, resulted in the addition of two residues (glycine (G_1) and serine (S_2)) at the N-terminus of the expressed Carcinin sequence. Due to their small size and N-terminal position, this modification to the expressed Carcinin sequence was not considered to be of great consequence for the purposes of the present study.

The two Carcinin sequences, which were expressed (ACAR and FCAR) in the present study, were selected based on *in silico* data obtained in Chapter 2, which identified a putative signal sequence at the N-terminus. By expressing both proteins, one with and one without the signal sequence, the effect of the putative signal sequence on yield and stability of the recombinant protein, could be investigated further. The primers used to amplify the selected insert sequences for the expression vectors were designed on the AJ427538 sequence (Appendix 7). Due to a nucleotide assignment error at the 5' end of that sequence, the first residue on the expressed ACAR sequence was an arginine (R_3) and not a glycine (G_3). The primers designed for PCR of the FCAR insert sequence were not affected by this error. Changes in the N-terminal sequence of an antibacterial protein can considerably affect activity (Raj *et al.*, 2000a; Cuthbertson *et*

al., 2002; Skosyrev *et al.*, 2003). However, in the present study, the activity of the expressed ACAR was not essential for polyclonal antibody production and the primers were not redesigned once the error had been realised. Using either sequence for antibody production would result in polyclonals specific for both proteins, although the antibody titre may be higher using recombinant FCAR.

To ensure sequence fidelity of the inserts, a proof-reading *Taq* enzyme was used in their amplification. Proof-reading *Taq* has been shown to degrade the ends of primers and amplicons during PCR (Promega, 1996). In the present study, this could have resulted in the degradation of the GC clamp and the restriction sites. However, this did not appear to adversely affect the outcome of the present experiment, as demonstrated by successful digestion and ligation of the inserts into the expression vector. Nonetheless, the overall yield of the constructs may have been compromised.

Alternative restriction sites, present in the MCR of the pGEX4T-1 vector, were also assessed for suitability before BamH1 and Xho1 were finally selected for directional cloning of Carcinin. The BamH1 site was closest to the thrombin cleavage site and added the fewest additional residues to the resultant protein. EcoR1, Sal1, Sma1 and Not1 were either too close to the BamH1 site or required too many base pairs on either side of the recognition site to be useful in the present study. Assessment of buffer compatibility between the enzymes indicated BamH1 and Xho1 enzymes were the most suitable pairing (New England Biolabs, 2004) for use in double digests. Although no isochizomers of BamH1 existed in the envisaged construct, a single site (PaeR7 1) was found in the vector for the Xho1 enzyme, which could lead to additional bands being generated upon digestion. However in the present study none were noted, indicating that this site was not affected.

The two GST-Carcinin constructs pGEX-4T-1F and pGEX4T-1A) were relatively easy to produce, transform and propagate following the manufacturers instructions. Expression of soluble fusion protein was immediately successful without the "prerequisite" addition of detergents (Mercado-Pimentel *et al.*, 2002), but yields were thought to be relatively poor (compared to GST controls) and considerable effort was invested to optimise the conditions to achieve yields (~ 0.5 mg of fusion protein or ~600 μg of non-fusion protein) suitable for use in polyclonal antibody production.

The 2 x YTAG medium (i.e. 2 x YTA medium plus 2 % glucose), recommended (Amersham Pharmacia Biotech, 2001) for use with toxic inserts was used in the present study with *E.coli* BL21 and *E.coli* JM105 strains as hosts. Although the native protein was not toxic to one strain of *E.coli* (strain CGSC 5163, Yale University, New Haven, Connecticut, USA) (Relf, 2000), toxicity against *E.coli* BL21 and JM105 strains was not experimentally demonstrated. As these strains remained viable after transformation and induction, any toxic effects were thought to be negligible. In addition, RDA results did not indicate any antibacterial activity of the expressed protein, although this could have resulted equally from selection of an impotent insert sequence and or the N-terminal modification of the expressed proteins described.

Lower growth temperatures, increased aeration, and lower cell density cultures also improved recovery of the fusion protein from the cell lysate as shown for other peptides (Schein and Noteborn, 1988). High cell density cultures can lead to the formation of a "skin" on the GS4B beads, which reduces glutathione 4B binding efficiency. Whilst changing the concentration of the IPTG (0.05-2 mM) and using shorter induction times did not improve yields, longer induction times (~8 h) did. It is possible that even longer induction times (~12 h), could further improve yields although on the other hand it may

make little difference if non resistant bacteria flourish as the ampicillin is used up. Both high density cultures and longer induction times may also allow the accumulation of cell debris and host proteins. Concerns over proteolysis, associated with the longer induction times, led to a conservative 3.5 h induction time being used throughout.

The use of alternative resuspension / lysis buffers (STE) improved yields but was outweighed by the increased co-purification of host proteins. The use of some additives (e.g. RNAses and DNAses) in the lysis steps appeared to decrease the viscosity of the lysate and prevent the formation of insoluble aggregations of the fusion protein, but others, (e.g. Triton X-100 (1 %) to the 1 x PBS lysis buffer) had no marked effect. PMSF has a short half-life in aqueous solutions, so perhaps this could have been added at later stages of the purification process. In addition, alternative protease inhibitors could also have helped to curb this proteolysis.

Two molar and four molar urea were most effective at solubilisation of insoluble inclusion bodies in the lysate pellet. Urea may also have denatured the GST moiety preventing subsequent purification (Frangioni and Neel, 1993). Amersham (Amersham Pharmacia Biotech, 2002) assert that 2-3 M urea should not affect binding to the glutathione matrix, but this was not the case in the present study. Refolding methods to correct denaturation (Sambrook and Russell, 2001) were not attempted because of concerns about spontaneous proteolytic degradation of the fusion product.

Lysozyme was omitted from the lysis even though it improved the yield of protein considerably, because it masked the cleaved Carcinin protein bands in SDS PAGE gels and was difficult to remove from the purified samples. As the recommended (Amersham Pharmacia Biotech, 2002) 10 s sonication time did not appear to be sufficient several sonication regimes were tried. Using the fine probe directly into the

sample also appeared to increase the denaturation of the protein as has been noted by Frangioni, J.V. and Neel, B.G. (1993) so instead the water bath method was used. It was quite difficult to assess the effectiveness of the various regimes as several factors seemed to have influenced the resultant yield as viewed on SDS PAGE gels. A minimal sonication regime was employed to avoid excessive protein denaturation but maintain sufficient yield.

Increasing the centrifugation speed of the lysate to 31,000 x g from the recommended 12,000 x g, ensured minimal cell debris was transferred to the GS4B beads. In addition, filtering the lysate supernatant after the centrifugation step prevented a "skin" forming on the outer layer of the GS4B beads. Higher centrifugation speeds (70,000-300,000 x g) have been previously shown by Dian *et al.*(2002b) to ensure removal of cell debris, membrane components and aggregates using GST fusion proteins expressed in BL21. It is possible to speculate that using these higher ultracentrifugation speeds may render redundant the filtration step and help reduce the number of washes required post binding.

As shown by others (Mercado-Pimentel *et al.*, 2002) working with the GST affinity purification system, increasing the bed volumes of the beads over those recommended by the published protocols (Frangioni and Neel, 1993; Amersham Pharmacia Biotech, 2002) greatly increases the yield of the eluted proteins. This effect was observed even though the GST-Carcinin fusions were much smaller than those purified in the aforementioned study. In the present study it was shown that increasing both the number of elutions and the time the elution buffer is allowed to mix with the beads also increased the yield of the released protein, again confirming previous reports (Mercado-Pimentel *et al.*, 2002).

It was also observed in the present study, that although the number of steps were reduced by cleaving the fusion protein on the beads, the yield was considerably reduced compared to the "off-bead" method. Therefore, all subsequent isolations of the fusion protein were carried out using the "off-bead" method before further purification.

The effect of the putative "signal" sequence on the stability of the protein was demonstrated by comparing the expression profiles of the GST-FCAR and GST-ACAR proteins. Spontaneous proteolytic degradation was clearly observed (Figure 52) for the putative "active" fusion protein (GST-ACAR) but not the "full" protein (GST-FCAR; see Figure 49-33). Similar reports of instability of GST-antibacterial fusion proteins have previously been documented by Hancock's group (1993; 1998) and more recently by other workers (Yiallouros *et al.*, 2002; Skosyrev *et al.*, 2003). Piers, *et al.*(1993) and Zhang, *et al.* (1998) have also shown that the insertion of a pre-pro defensin sequence between the fusion tag and the cationic peptide confers complete protection against proteolytic degradation. It was concluded from the present study that the difference in stability between the GST-ACAR and the GST-FCAR proteins was due to the absence of the putative "signal" sequence in the GST-ACAR protein. At present, it is thought that crustacean granular cells do not actively transport the AMPs out of the cell (requiring a signal sequence to direct transmembrane transportation) but release peptides by degranulation and lysis (Munoz *et al.*, 2002) so the function of the signal sequence is thought to be in stability rather than translocation of the protein. This sequence may also have functions associated with the creation of disulphide bonds and this could be investigated in future studies.

The cleavage of the ACAR from the GST resulted in a marked difference between the yields of the two proteins, with considerably less ACAR than GST observed on the SDS

PAGE gels. This is further evidence of the suggested instability of the putative "active" Carcinin protein (ACAR). A strong GST band could still be observed in elution samples which were analysed 10-20 h after cleavage but no ACAR band was visible. A similar result was obtained in a separate study where sarcotoxin IA was expressed as a soluble GST fusion protein (Skosyrev *et al.*, 2003) but only a small proportion as full size fusion proteins. Cleavage of this fusion resulted in GST as the only stable product with complete degradation of the target, despite addition of protease inhibitor cocktails. In the present study it was not empirically demonstrated that the cleaved FCAR was more stable than the ACAR, but the accumulation of GST in SDS PAGE gels was not observed for the cleavage of the FCAR-GST protein.

Separation of the cleaved Carcinin proteins (FCAR and ACAR) from the GST tag proved to be quite labour intensive, requiring several cultures to be prepared, induced and purified. Initially, both constructs were tested alongside one another, but as the purification process became more complex, only the GST-ACAR fusion was investigated with a view to returning the GST-FCAR at a later date. Unfortunately, the lack of time and funds precluded this.

After cleavage of the fusion protein, the removal of thrombin was achieved by applying the sample to a benzamidine matrix, leaving any soluble Carcinin and GST in the flowthrough. Re-applying the sample to glutathione 4B affinity matrix to remove GST was not particularly effective and a considerable amount of GST was found to remain in the supernatant. It is possible that when the fusion protein was eluted the first time from the beads, some of the GST moiety became denatured and unable to bind a second time. Separation of GST from Carcinin was attempted using FPLC and HPLC methods. The RP-HPLC method used by Relf, *et al.*(1999) did not produce this separation and any

ACAR present may have been degraded as small peptides (<10 kDa) were identified by SDS PAGE. This could have been due to a combination the instability of the recombinant ACAR protein and the harsh purification conditions involved in RP-HPLC.

FPLC was then used to try and separate the proteins on the basis of charge. The ACAR protein should have been strongly cationic and the GST virtually neutral (if not slightly negative). Both a strong cation exchanger (MonoS) and a strong anion exchanger (MonoQ) were used to try to separate these protein, but neither was successful. The pI's of the two proteins may be too similar under the conditions used as a result of denaturation and degradation. It is possible that increased concentration of a protease inhibitor cocktail, during isolation and purification steps, could help to redress this and improve separation. Isoelectric focussing could then be used to establish the actual pI of the two proteins if they can be separated and by altering the pH and ionic concentration of the FPLC better separation can be achieved.

Neither the cleaved recombinant ACAR nor the full fusion construct (GST-ACAR) demonstrated antibacterial activity using RDA. This further supports the hypothesis that post translational modifications and folding are central to the antibacterial activity of Carcinin. However, this activity was not essential to the purposes of the current study.

It should be noted that only one of the isoforms identified in Chapter 2 (isoform 4) was experimentally expressed in this Chapter. It is possible that not all the isoforms exhibit antibacterial activity and they may also have other functions in their native form.

Defensins have been previously shown to be difficult to express in bacterial systems due to the necessity of the correct folding of disulphide bonds to retain activity (Piers *et al.*, 1993). Although Carcinin possess several cysteine residues (12), the formation of

disulphide bridges has not been confirmed experimentally. Evidence for the presence of a disulphide core is based on predictive software results alone.

In view of the results obtained in this chapter alternative expression vectors and systems could prove to be more effective (e.g. yeast expression systems). Contrary to this, some workers (Skosyrev *et al.*, 2003) have observed that although the expression of small peptides in heterlogous systems has inherent problems associated to it, for some peptides, bacterial expression yields are far greater than when using alternative systems. This study (Skosyrev *et al.*, 2003) also highlighted the varying successes obtained by expressing a cationic peptide using different fusion tags. The use of alternative fusion tags (e.g. His), as summarised in section 3.1.1.1.1, may prove more successful than the GST tag used in the present study.

It can be concluded from the present study that recombinant bacterial expression of a Carcinin fusion protein using the pGEX-4T-1 vector and BL21 expression host is easily achievable. However, expression of the "active" fusion protein (ACAR-GST), produced yields at the lower end of the scale (2.140 - 2.963 μg per ml of culture), compared to similar studies which have reported yields of between 2~10 μg per ml of culture (Hara and Yamakawa, 1996; Michaut *et al.*, 1996b; Mercado-Pimentel *et al.*, 2002). The low yield in the present experiment is thought to be due to a combination of factors, including the insolubility of a considerable proportion of the expressed fusion protein and the degradation of the putative "active" Carcinin sequence. These stability issues may preclude "active" protein sequences from being suitable candidates for use in polyclonal antibody production unless they can be stabilised by the production of chimeric hybrids. Although it was not empirically quantified in the present study, the recovered yields are thought to be greater for the expression of the FCAR fusion and

this construct should be used in the future to express a Carcinin fusion protein. The GST-FCAR fusion product does not appear to be prone to the spontaneous proteolytic degradation seen in the GST-ACAR and it can be speculated that this is due to the protective influences of the pre-protein or leader sequence. This stability may further support the hypothesis that this sequence could be the precursor form of the Carcinin protein, thought to be stored in the granulocytes. It is speculated that when the cell degranulates, these precursor proteins are rapidly activated by the cleavage of the signal peptide to release the active protein (possibly the ACAR sequence), which is then degraded. It is thought that the unrestricted accumulation of an immunologically active substance in the haemolymph may have detrimental effects to the crab. Therefore, this would support the need for rapid degradation of these active proteins. Evidence of this can be found in the shrimp *L.vannamei*, where the antibacterial activity in the plasma of penaeidins released from the granulocytes, abated 6 h post stimulation (Destoumieux *et al.*, 2000) and it is speculated that a similar situation is present in *C. maenas*.

The removal of the cleaved GST moiety was not achieved in this experiment and it is therefore concluded that only intact fusion proteins should be used for polyclonal antibody production. These antibodies could be used to investigate Carcinin protein expression, storage and processing loci as well as relate it to transcript studies (Chapter 4). These conclusions clearly highlight the need, for investigating a number of different expression systems and recombinant proteins before the most appropriate is chosen and that substantial optimisation may be required of each.

In future studies the quantification of expressed GST-FCAR fusion protein needs to be established and alternative methods of separation from the GST protein need to be investigated. The role of the signal sequence in controlling the disulphide bond

formation (i.e. acting as a prepro-protein and not just a pre-protein) should also be investigated, using alternative expression hosts and vectors, which allow secondary folding to occur. Alternative fusion partners could also be assessed as well as attaching them at either the C-terminal and N-terminal ends of Carcinin with a view to improving expressed yields. Additional prepro-protein sequences could be inserted between the GST tag and the Carcinin protein which could further stabilise the ACAR protein against proteolytic degradation.

Considering the optimisation required, the instability of the recombinant peptides and the absence of limitations imposed by an abundance of tryptophan or arginine residues (Zhang *et al.*, 1998), synthetic production, for the purposes of antibody production, may be preferable to recombinant production of Carcinin.

CHAPTER 4

Real-Time PCR Optimisation

4.1 Introduction

Northern blotting was used to quantify the relative expression of antimicrobial transcripts in the shrimp, *Litopenaeus vannamei* (Farfante and Kensley, 1997) but the results indicated that the technique lacked sensitivity and reproducibility (Destoumieux *et al.*, 2000; Munoz *et al.*, 2002); similar conclusions were reached in the present study in *Carcinus maenas* (Chapter 2; 2.3.2). Consequently, northern blotting was discarded as a method for quantifying transcript expression where sample size is limited. In the present, study gene expression in *C. maenas* was assessed using the latest technology in sequence detection (SD) analysis, namely quantitative real-time polymerase chain reaction (real-time PCR) using TaqMan® (Roche Molecular Systems Inc., USA) technology. This method has been used successfully to demonstrate changes in expression of antimicrobial transcripts in arthropods (Nakajima *et al.*, 2003), although none of these have been conducted in the Crustacea.

4.1.1 Central Principals

TaqMan® technology uses a modified PCR, in conjunction with fluorescent detection, to quantify reverse transcribed mRNAs (cDNA) sequences of interest. Whereas traditional endpoint PCR has a dynamic range of two orders of magnitude, real time PCR has a dynamic range greater than 8 (Schmittgen and Zakrajsek, 2000; Rasmussen, 2001). TaqMan® technology is based on the design of gene-specific primers and a labelled probe amplifying a relatively short amplicon (50-150 bp). The oligonucleotide probe anneals between the forward and reverse primers; this probe is labelled at the 5' end with a fluorescent reporter dye (e.g. 6 FAM™ or VIC™) and at the 3' end with a quencher dye (TAMRA™)(Figure 63). When the reporter and quencher dyes are in close proximity to each other, fluorescent resonant energy transfer (FRET) occurs from

the reporter to the quencher, thereby reducing the emission of the fluorescent signal from the reporter dye at 518 nm but increasing the emission of the quencher dye at 582 nm (Figure 63 and Figure 64).

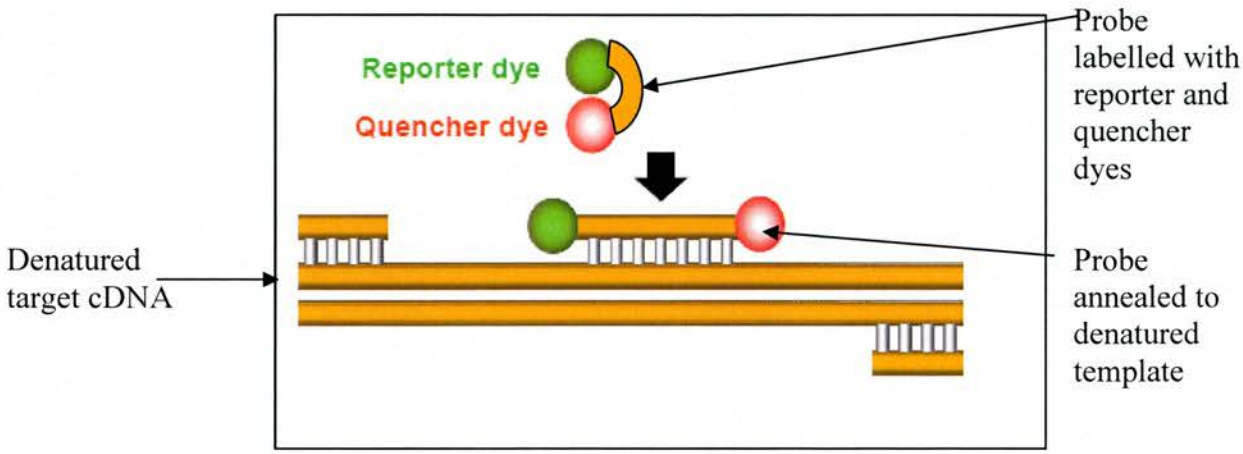


Figure 63:Annealing of a TaqMan® probe to denatured target cDNA (image reproduced from Applied Biosystems, 2004). FRET occurring as reporter and quencher dye in close proximity to each other which continues after annealing to template.

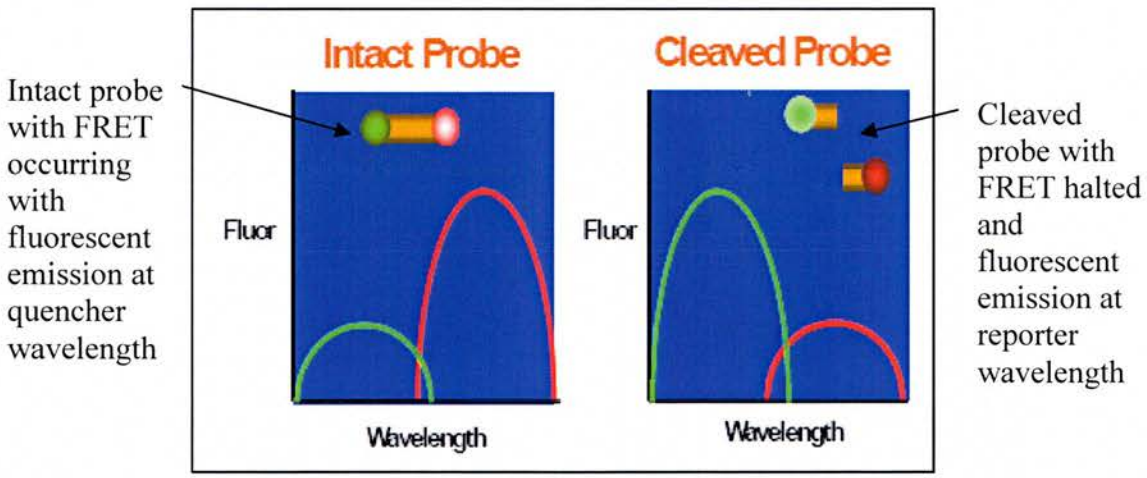


Figure 64: Reduction in emission at reporter signal wavelength (518 nm) as FRET occurs and then is halted due to 5' exo-nuclease activity of the AmpliTaq ® Gold resulting in increased emission at the reporter dye wavelength (518 nm) (image reproduced from Applied Biosystems, 2004).

The PCR master mix used in real-time PCR contains a DNA polymerase with 5' exo-nuclease activity, which removes nucleotides downstream of the polymerised amplicon (Figure 65). As the polymerase enzyme progresses along the cDNA template the reporter dye on the probe is cleaved by the 5' exo-nuclease activity of the AmpliTaq ® Gold and FRET is halted (Figure 65). This results in the emission of energy as fluorescence (Figure 64) which is detected by a charge-coupled device (CCD) camera.

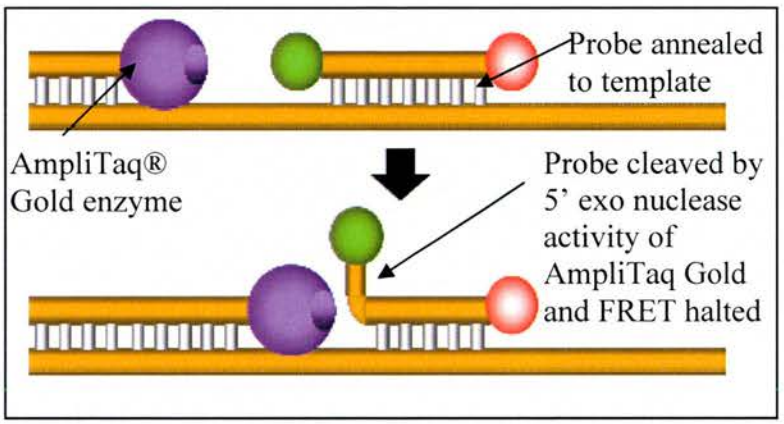


Figure 65: Cleavage of the reporter dye by the 5' exo-nuclease activity of AmpliTaq ® Gold enzyme (Applied Biosystems, 2004).

The non-PCR related fluctuations in fluorescent signal are normalised by the inclusion of a passive internal control dye ROX™. Dividing the signal from the reporter dye (e.g. 6 FAM™ or VIC™) by that of the passive reference dye (ROX™) gives a ratio known as Rn or R^+ (normalised reporter) for each reaction well. The Rn value obtained for the no template controls (NTC) is known as the R^- and the difference between R_n and R^- is known as the ΔR_n, and indicates the strength of the signal generated by the given set of PCR conditions (Applied Biosystems, 2002).

As the amplicons accumulate (doubling for each amplification round assuming 100 % efficiency of the reaction), so does the fluorescent signal in a directly proportional

manner. As the amplification round nears completion the probe is completely removed from the template (Figure 66).

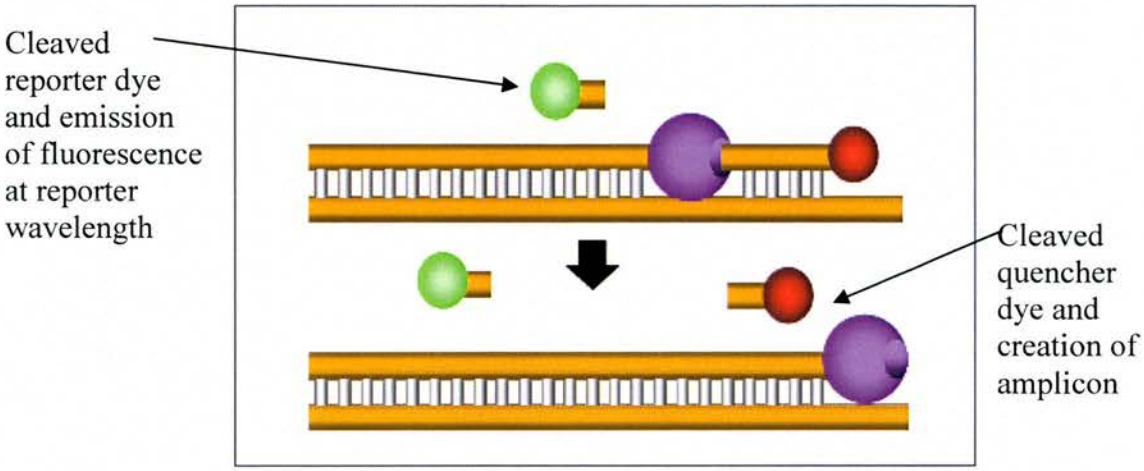


Figure 66: Removal of the probe from the template and creation of amplicon (Applied Biosystems)

The combination of FRET and the 5' exo-nuclease activity enables the SD instrument to collect data in "real-time" and this is displayed on an amplification plot as the fluorescent signal accumulates (Figure 67). At the point where there is an increase in ΔR_n, above the background levels, the amplification reaction enters the exponential phase and this point is known as the cycle threshold (Ct) value (Figure 67).

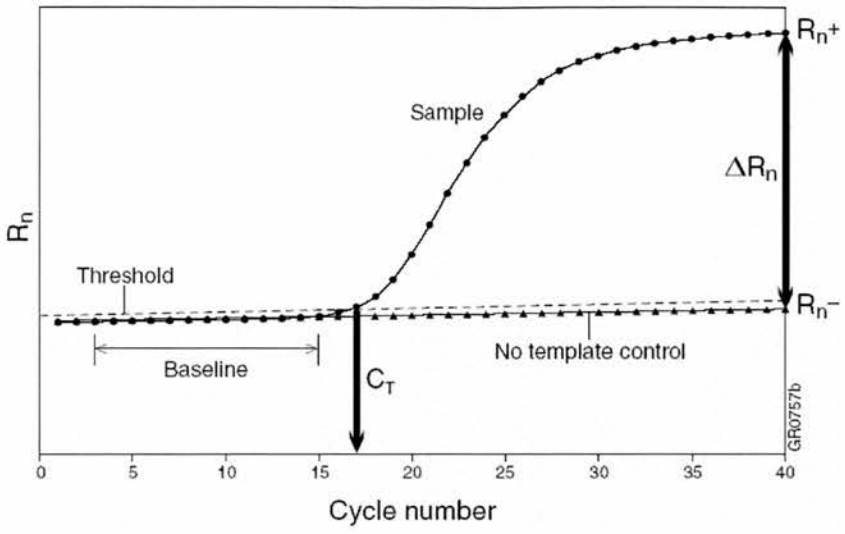


Figure 67: Representation of an amplification plot showing cycle number versus R_n displayed by an ABI PRISM 7000 SDS instrument (image reproduced from Applied Biosystems, 2002).

In relative gene expression, to control for different amounts of template in each reaction, the gene of interest is normalised by the co-amplification of an endogenous control gene. This gene should be expressed at constant levels in the tissue analysed and its expression level should remain unaffected by the experimental conditions (Bustin, 2000; Applied Biosystems, 2001; Pfaffl, 2004). In the present experiment the target and endogenous control genes were analysed in separate reaction tubes and the concentrations of the primers and probes were optimised for each gene amplified, as recommended (Applied Biosystems, 2001, 2002). The validity of relative quantification relies heavily on appropriate endogenous control selection and the normalisation and quantification calculations performed (Pfaffl *et al.*, 2002). Therefore, these issues were addressed in some detail in the present study.

4.1.2 Relative Quantification Data Analysis

Several analysis methods are available to quantify relative changes in expression data. These include the standard curve method (Applied Biosystems, 2001), the comparative

Ct method (ΔΔCt) (Pfaffl, 2001) and several software programs (e.g. Q-Gene) including REST- XL © (relative expression software tool) (Pfaffl *et al.*, 2002) which uses a mathematical model based on PCR efficiencies and mean Ct deviation between control and target groups. All three methods use the mean Ct value of the replicates from each sample in the subsequent calculations.

The standard curve method uses the construction of a standard curve by amplification of a serially diluted calibrator (control) cDNA sample, prepared (in triplicate) for both the target gene and the endogenous control gene. The mean transcript abundance of each experimental sample is then used to determine the standard curve equation for each gene. The mean abundance of the target gene is then divided by the mean of the endogenous control gene resulting in a normalised target gene value. The normalised target gene value is then divided by the normalised calibrator sample to give relative expression changes. The calibrator sample thus becomes the 1 X sample and the experimental samples are expressed as an n-fold difference to the calibrator. By dividing by the normalised calibrator quantity, the resulting value become unit less and only relative input quantities or dilutions need to be known (Applied Biosystems, 2002). This method can be used when the PCR amplification efficiencies of the target and the endogenous control are not equal.

The comparative method (Pfaffl, 2001) uses a mathematical equation ($2^{-\Delta\Delta Ct}$) for relative quantification and can only be applied if the PCR efficiencies of the target and endogenous control are similar in value. The ΔΔCt value is the difference between the normalised target ΔCt_{target} (Ct of target – Ct of endogenous control) and the normalised calibrator samples $\Delta Ct_{calibrator}$ (Ct of target calibrator – Ct of control calibrator endogenous). This ΔΔCt value is then used in the equation ($2^{-\Delta\Delta Ct}$) to obtain n-fold

changes in expression values. The similarity of their efficiencies is assessed by plotting the log of the cDNA input amount (or dilution) against the change in Ct (ΔCt) and the resultant regression line has a gradient of <0.1 (Applied Biosystems, 2001).

The REST-XL © software calculates PCR efficiency values (from standard curve data) for use in randomisation tests to establish the statistical significance of observed differences in expression between the target and control genes. The program calculates a PCR efficiency value from the mean (of sample replicates) Ct values, for target and control gene, using serial dilution plots (Pfaffl, 2001) as per the standard curve preparation. This value is used to calculate the statistical significance of differences between the mean Ct values of the genes, in a single data set, using a randomisation test (Pair Wise Fixed Reallocation Randomisation Test ©) (Pfaffl *et al.*, 2002) with 2000 iterations. The randomisation test determines whether the observed differences between the control and experimental groups are likely to have arisen by chance or by random reallocation of the values observed to each set, without assuming any distribution properties of the data. The p-value of the test is based on the proportion of the effects observed by the random reallocation process compared to those observed (Pfaffl *et al.*, 2002). In the present study, the standard curve method was used to evaluate initial optimisation experiments and to establish the amplification efficiencies of the target and endogenous control genes.

This chapter is concerned with the selection of endogenous controls, the optimisation of the real-time PCR reactions, the assessment of amplification efficiencies and the validation of the analysis methods to be used for the subsequent experimental samples.

4.2 Specific Aims

- To obtain transcript sequence information on at least two endogenous control sequences.

- To evaluate endogenous control genes for use in relative transcript expression analysis of Carcinin using real-time PCR .

- To optimise the reaction components and amplification reactions for the transcript tools developed.

- Establish relative expression levels of each gene and their amplification efficiencies.

4.3 Methods and Materials

4.3.1 Template Preparation

RNA was isolated and quantified from haemolymph samples as described previously in Chapter 2 (2.3.1 and 2.3.2). RNA was treated with RNAse-free DNAse (Promega, Madison, USA) according to the manufacturers protocol, to remove residual contaminating DNA. Briefly, RNA samples (1-2 μg) were digested with RNAse-free DNAse (DNAse 1) at 1unit per μg of RNA in 10X reaction buffer (400 mM Tris-HCl (pH 8), 100 mM $MgSO_4$, 10 mM $CaCl_2$; supplied) at 37 °C for 30 min. The DNAse enzyme was inactivated by the addition of 1 μl of RQ1 DNAse Stop Solution (20 mM EGTA; pH 8: supplied) and incubated at 65 °C for 10 min. The DNAse treated RNA (1-2 μg) was then reverse transcribed to cDNA (Chapter 2; 2.3.5) using random hexamers (Promega). RNA samples were then quantitatively and qualitatively assessed as described in section 2.3.2.1 of Chapter 2. cDNA samples were diluted with nuclease free water to ~2 ng μl^{-1} before gene expression was assessed. Samples were stored at -20 °C until required.

4.3.2 Endogenous Controls

Genes coding for β-actin and 18S rRNA transcript sequences were both evaluated as endogenous controls in the present study. These genes were selected for their popularity as endogenous controls in quantitative PCR studies in many species, including *C. maenas* (Towle *et al.*, 1995; Towle *et al.*, 1997), as well as the conservation of sequence information between related species (Spears *et al.*, 1992). These sequences in *C. maenas* are not published and were elucidated as follows.

Published 18S rRNA sequences from members of the Pleocyemata, including *Hepatus epheliticus* (M91053), *Callinectes sapidus* (M34360), *Homarus americanus* (AF235971) and *Nephrops norvegicus* (Y14812) (Appendix 1), were aligned using the DNAMAN software package (v 5.2, Lynnon BioSoft, Canada). Primers were then designed to conserved regions of these aligned sequences (Appendix 22). These primers were then synthesised by MWG Biotech (Ebersberg, Germany), diluted with 1 ml nuclease free water and dispensed into 20µl aliquots before storage at –20 °C.

PleoFWD 5' – TAA GGC GAA ACC GCG AAT GG -3'
PleoREV 5' – CTG GGA ATT CCT CGT TCA TGG – 3'

Table 24: Forward (PleoFWD) and reverse (PleoREV) 18S primers (T=thymine, A=adenine, G=guanine, C= cytosine) from the IUPAC code (www.dna.affrc.go.jp/misc/MPsrch/InfoIUPAC.html)

Total RNA was isolated, as described in 4.3.1 and reverse transcribed to cDNA using random hexamer primers; this template was then amplified using the "Pleo" primers detailed above (Table 24). These same primers were also used to amplify 18S genomic DNA sequences isolated in Chapter 2 (2.3.8). Proof reading Advantage® *Taq* polymerase mix (BD Biosciences, Oxford, UK) was used in the PCR, as described in Chapter 2 (2.3.5), using the following cycling conditions: 95 °C for 5min, then 30 cycles of 95 °C for 1 min, 55 °C for 30 s and 72 °C for 2 min, followed by 1 step of 5 min at 72 °C. The amplification products were visualised by electrophoresis on a 1.5 % agarose gel, as described in Chapter 2 (2.3.5). The resulting bands were purified, cloned and sequenced as described in Chapter 2 (2.3.6).

β-actin degenerate primer sequences were obtained from a study on the sodium/proton antiporter in *C. maenas*; these had been previously designed to conserved regions after alignment of published sequences from related arthropods (Towle, 1997).

ActinFWD 5' – GTC GGY GAY GAR GCW CAR A – 3'
ActinREV 5' – GGR CAR CGG AAW CGY TCA TT – 3'

Table 25: β-Actin degenerate primers (IUPAC code)

The "Actin" primers were used to amplify the β-actin transcript using previously generated cDNA as template (4.3.1). Advantage® *Taq* polymerase mix (BD Biosciences) was used and the cycling conditions were as follows: 95 °C for 5 min then 30 cycles of 95 °C for 1 min, 50 °C for 30 s and 72 °C for 2 min, followed by 1 step for 5 min at 72 °C. The amplification products were visualised on a 2 % agarose gel as in Chapter 2 (2.3.5). The resulting bands were purified, cloned and sequenced as described in Chapter 2 (2.3.6).

4.3.3 Transcript Tool Design

Once sufficient sequence and isoform data were obtained for the endogenous controls (4.3.2) and Carcinin (Appendix 10) respectively, primers and probes were designed using Primer Express ® software (version 2.0, Applera Corporation Inc. California, USA) according to the manufacturers instructions (Applied Biosystems, 2002). This software is integral to the Applied Biosystems PRISM 7000 Sequence Detection System (ABI PRISM 700 SDS) (Applied Biosystems Ltd, California, USA). Probes and primers were synthesised by Applied Biosystems (Applied Biosystems UK, Cheshire, UK) or MWG Biotech (MWG Biotech, Ebersberg, Germany). The freeze-dried

TaqMan® probes and primers were resuspended in 1ml of nuclease free water and dispensed into 0.5 ml microfuge tubes in 20 µl aliquots for storage at –20 °C until use.

The reporter dyes attached at the 5' ends of the probe were, 6-FAM ™ for the Carcinin and 18S rRNA probes and VIC ™ for the β-actin probe. The 3' ends of all probes were labelled with TAMRA ™ quencher dye (Applied Biosystems, 2002). Where gDNA data was available (Carcinin), either the probe or one of the primers was designed to bridge the identified exon-exon boundaries (Appendix 23). This ensured that only the mRNA and not any remaining gDNA contamination were amplified by the real-time PCR reaction.

4.3.4 Real-Time PCR Preparation

In the present study, real-time PCR was performed with an ABI PRISM 7000 SDS, using TaqMan® chemistries. The probes and primers (4.3.3) were used with TaqMan® Universal PCR Mastermix (Applied Biosystems) under universal cycling conditions. Double strength (2X) TaqMan® Universal Master Mix (Applied Biosystems), containing AmpliTaq Gold DNA Polymerase, dNTPs, ROX™ (passive reference dye) and optimised buffer components, was used for amplification reactions. Universal cycling conditions were used for all samples as follows: 1 step of 2 min at 50 °C then, 1 step of 95 °C for 10 min (to activate the AmpliTaq Gold), then 40 cycles of 95 °C for 15 s denaturation and 1 min at 60 °C for annealing and extension (Applied Biosystems, 2002). All reaction mixtures (50 µl) were prepared using DNAse / RNAse free plastics and nuclease-free water. MicroAmp ® 96 Well Reaction Plates (Applied Biosystems), with ABI Prism™ Optical Adhesive covers, were used only once each.

4.3.5 Primer and Probe Optimisation

In the present study, minimum primer and probe concentrations required for optimal amplification were established. These optimisation reactions also determined the appropriate concentration of template to be used in subsequent experiments as well as the minimum reaction volumes required. The amplification efficiencies of each gene were determined and used to evaluate the primer and probe sets. Optimisation reactions, as recommended by ABI, were undertaken for most of the primers and probes. However, the β-actin primers were used at maximum concentration (900 nM) in the initial amplifications to obtain preliminary results. Subsequently, the β-actin probe concentration was optimised as described, below, for the 18S rRNA probe.

Optimisation of the Carcinin and 18S rRNA primers was performed using varying concentrations of the two primers to determine the minimum primer concentration combination that resulted in the maximum ΔR_n (the change in R_n). The primer concentration range, recommended for testing by the manufacturer (Applied Biosystems, 2002) was 50-900 nM (Table 26).

	Forward Primer (nM)		
Reverse Primer (nM)	**50**	**300**	**900**
50	50/50	50/300	50/900
300	300/50	300/300	300/900
900	900/50	900/300	900/900

Table 26: Primer optimisation concentrations used for Carcinin and 18S genes.

Different primer concentrations were tested in quadruplicate for each gene alongside no template control (NTC) reactions. To minimise pipetting errors, initial reaction

mixtures were prepared for each gene in a single tube and included 2 X Universal Master Mix, 200 nM probe with 2-10 ng of a single control cDNA sample. This mixture was dispensed into 0.5 ml microfuge tubes. Forward and reverse primers (50-900 nM) were added to the microfuge tubes and made up to a total volume of 150 µl, with water, and vortexed. The NTCs were prepared with the highest primer concentrations (900nM forward and 900 nM reverse) for each gene. Fifty microlitre replicates of each reaction were dispensed onto a MicroAmp ® 96 well reaction plate which was sealed with an ABI Prism™ optical adhesive cover and placed into the block of the instrument.

Using the previously determined optimal primer concentrations (4.4.3) (or the maximum concentrations used for the β-actin primers), the probe concentration for each gene was optimised. The concentrations tested were 25 nM, 50 nM, 75 nM, 100 nM, 125 nM, 150 nM, 175 nM, 200 nM and 225 nM. The minimum concentration of probe which produced the minimum Ct value, was selected. Reactions mixtures were prepared as described for the primer optimisation reactions above, with the probe added last. The same sample volumes and universal cycling conditions were used throughout.

4.3.6 Amplification Efficiency

Standard curves for Carcinin and β-actin were produced using a pool of cDNA control (calibrator) samples as template. Ten cDNA samples from different individuals were selected to represent the range of temperatures most commonly experienced by the animals (10 – 20 °C). Two µl from each individual were pooled and prepared with nuclease-free water as a ten fold dilution series: undiluted, 1:10, 1:100, 1:1000 and 1:10,000. The reactions were prepared using primer and probe concentrations determined in 4.3.5. The amplification efficiencies and the validation of the

comparative ΔΔCt method were investigated by plotting the log of the dilutions vs. ΔCt (difference between the sample Ct averages for Carcinin and endogenous control). PCR efficiencies were also calculated using the REST-XL © software based on the mean Ct values obtained from the standard curves for each gene.

Standard curves for 18S rRNA and Carcinin were generated from a single cDNA sample from a control animal. This calibrator sample was prepared as a ten fold dilution series: undiluted, 1:10, 1:100, 1:1000 and 1:10,000. The primer and probe concentrations determined in 4.3.5 were used to prepare reaction mixtures for the standard curve reactions. Triplicate reactions of 50 µl were prepared as described in 4.3.4. The reactions were prepared and the amplification efficiencies assessed as described above.

4.4 Results

4.4.1 Endogenous Control Sequences

The 18S rRNA product was obtained and visualised as a band at approximately 1500 bp (Figure 68). The β-actin degenerate primers amplified a sequence approximately 750 bp (Figure 69).

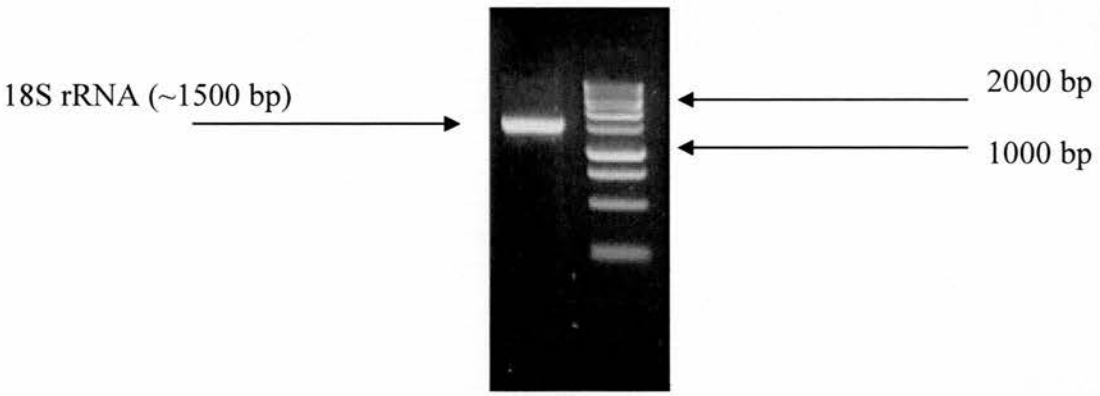


Figure 68: 1.5 % agarose gel of partial 18S rRNA transcript (~1500 bp) with a 1kb marker (Promega)

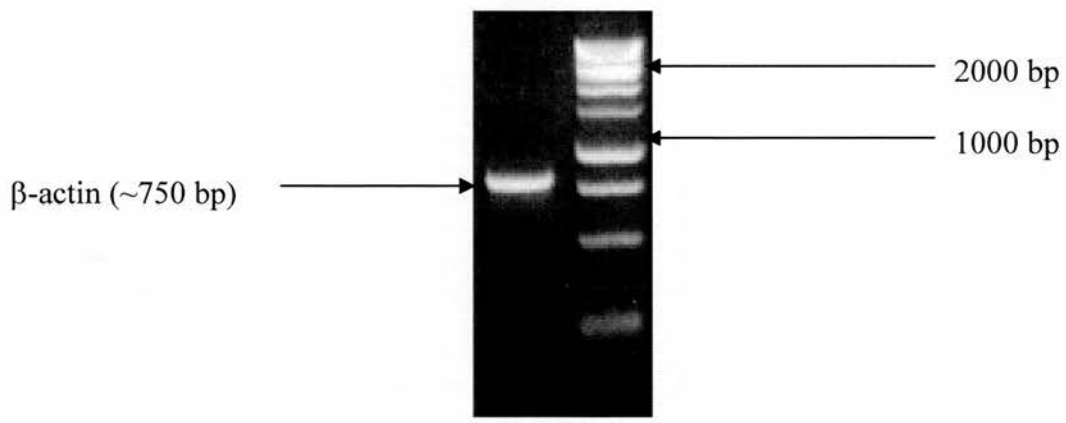


Figure 69: 2 % agarose gel of partial beta actin transcript (~750 bp) with a Promega 1kb marker.

Partial transcript sequences were obtained for both 18S rRNA and β-actin (Appendix 22 and Appendix 24). The 18S rRNA sequence shared 93.54 % identity, at the nucleotide level, with the Pleocyemata sequences (Appendix 22). Alignment of the partial β-actin sequence, obtained in the present study, with the β-actin nucleotide sequence obtained

by Towle D.W. (pers comm.) (Mount Desert Marine Laboratory, Maine, U.S.A.) exhibited an 83.25 % identity at the nucleotide level and 91.01 % sequence identity at the protein level (Appendix 25). The discrepancies between these two sequences are probably due to PCR fidelity or nucleotide assignment issues; these could be resolved in the future with further sequencing. Alignment of the amino acid sequence with published sequences of other related Crustacea revealed 95.92 % identity with β-actins from *Homarus gammarus* (CAE46725), *Gecarcinus lateralis* (AAL40077); these were distinct from the α or γ actin sequences.

4.4.2 Transcript Tools

The partial sequences obtained for both 18S rRNA and β-actin (Appendix 22 and Appendix 24) did not include any exon-exon boundaries when aligned to the gDNA sequences obtained. Consequently, neither the primers nor probe, for either gene, could be designed to bridge an exon-exon boundary. Therefore, all samples were DNAse treated to avoid amplification of contaminating gDNA. In addition, as only partial sequences were obtained for these genes, the choices of possible primer and probe annealing sites were limited. The primer and probe sets produced single amplicons of 112 bp for β-actin, 121 bp for 18S rRNA and 129 for Carcinin (Figure 70), by PCR performed under universal cycling conditions (4.3.4).

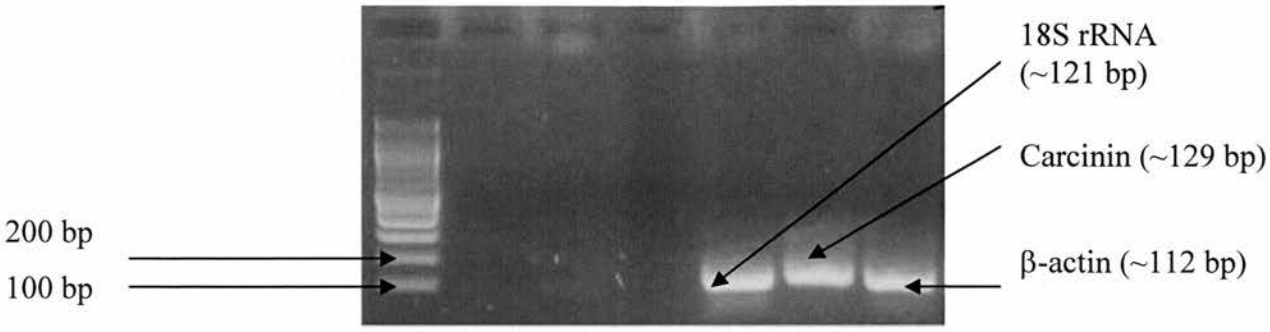


Figure 70: 2 % agarose gel electrophoresis of 18S rRNA and Carcinin amplicons using TaqMan® primers alongside a 100 bp marker (Promega).

Only one candidate site, on the Carcinin sequence satisfied the Primer Express® rules of primer and probe design; this was between nucleotides T^{264} and C^{291} (Appendix 23). This meant that the 5' end of the probe anneals in a region which includes a single nucleotide polymorphism (SNP) at nucleotide C^{265}. Since the binding of the probe at the 5' end is weaker than the 3' end (McPherson and Møller, 2000) and in light of the restrictions imposed by the Primer Express® software design rules, this location was accepted as the most appropriate candidate site for probe design. The probe was designed to bridge an exon-exon boundary ensuring only reverse transcribed transcript sequences would be amplified. The forward and reverse primers were also positioned so that transcripts arising from other previously identified SNPs in the sequence (Appendix 7) would not be excluded from amplification, in the current experiment. The resulting primers and probes designed for use in the current study were as follows:

TaqMan® PROBES	
Carcinin	5' 6-FAM-TCC AGT CCG TGA TAC CTG TAC TGG CGT C– TAMRA 3'
18S rRNA	5' 6-FAM – CGG AAT TAA CCA GAC AAA TCG CTC CAC C - TAMRA 3'
Actin	5' VIC - CCA TGT ACG TGG CCA TCC AGG CC – TAMRA 3'
PRIMERS	
Carcinin Fwd	5' – CCA GGA GTA ACC TAC CCA CCT TT –3'
Carcinin Rev	5' – TCT GAA CTG ACA TGC ACC ATC AT –3'
Actin Fwd	5' - CGT GAG AAG ATG ACC CAG ATC A – 3'
Actin Rev	5' – GCA CGA TGC CGG TGG TA –3'
18S rRNA Fwd	5' – GAC ACC GGA AGG ATT GAC AGA –3'
18S rRNA Rev	5' – GTA GGC CAG AGT CTC GTT CGT T –3'

Table 27: Real-time PCR primers and probes for Carcinin, β -actin and 18S rRNA

4.4.3 Optimisation

Gene	Forward primer nM	Reverse primer nM	Probe nM
Carcinin	**300**	**300**	**125**
18S rRNA	**300**	**900**	**150**
β-actin	**900**	**900**	**175**

Table 28: Primer and probe concentrations (nM) used in real-time PCR reactions

The lowest concentrations of primers eliciting the highest ΔR_n were selected and their concentrations are shown in Table 28. To quickly obtain preliminary information of the

relative expression levels and to conserve expensive reagents (e.g. master mix) the β-actin primers were not optimised prior to bacterial challenge experiments (Chapter 5, 5.4.2). As the probe was relatively expensive compared to the primers, these were optimised and the lowest Ct value obtained for the lowest primer concentration was 175 nM (Table 28).

4.4.4 Amplification Efficiency

Amplification plots and standard curves (Figures 71 to 73 and 75 to 77) were produced for both a Carcinin/β-actin and a Carcinin/18S rRNA template dilution series. The standard curve amplification plots were used to determine the appropriate dilution of template for the amplification of experimental samples.

The relative amplification efficiencies of the genes were assessed using both the standard curve plots (Figure 73 and Figure 77) and the ΔCt validation plots (Figure 74 and Figure 78) produced for each gene pair. These plots determined the most appropriate subsequent analysis.

4.4.4.1 AMPLIFICATION USING β-ACTIN ENDOGENOUS CONTROL

4.4.4.1.1 Standard Curve Method

The relative amplification plots of Carcinin and β-actin illustrate that Carcinin is highly expressed compared to β-actin (Figure 71 and Figure 72). The Ct values obtained for Carcinin were lower than β-actin by an average ΔCt of 6 (Figure 73). This suggests that Carcinin is expressed around 64 times the level of β-actin in the sample.

Appropriate threshold and baseline values were set for the Carcinin/β-actin amplification plot to obtain the Ct values. The mean Ct value for Carcinin, at the 10^{-3}

dilution, was 23.29 (Figure 71). This dilution was used for analysis of samples in the temperature experiment. The mean Ct value for β-actin, at the 10^{-3} dilution, was 29.9 (Figure 72).

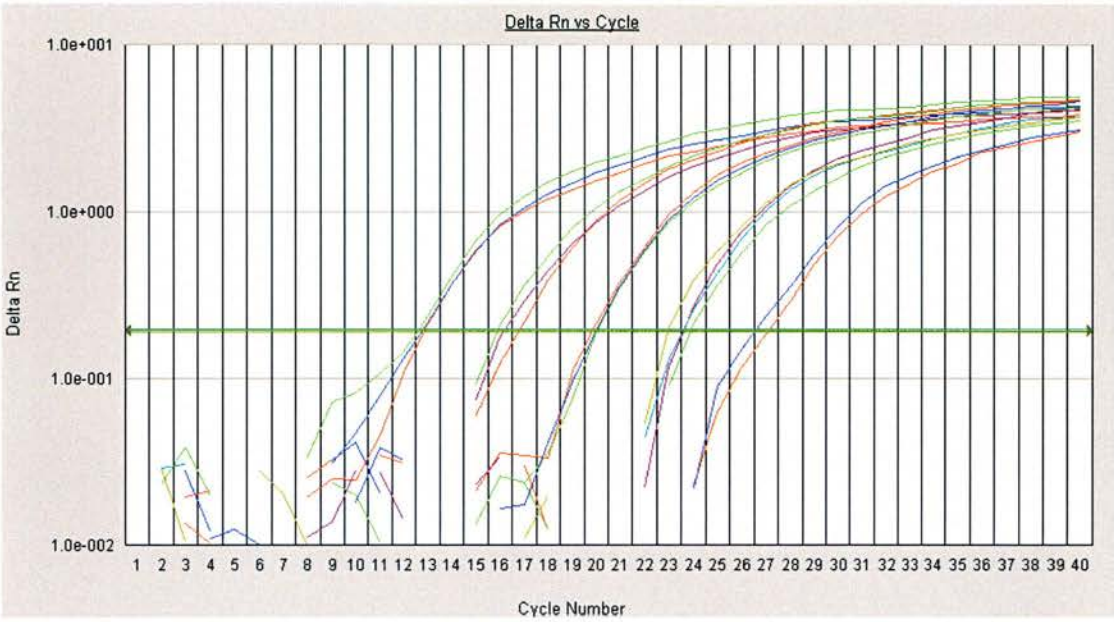


Figure 71: Amplification plot for the Carcinin standard curve, using 1 – 10^{-4} dilution of pooled cDNA template.

Figure 72: Amplification plot for the β-actin standard curve, using 1 – 10^{-4} dilution of pooled cDNA template.

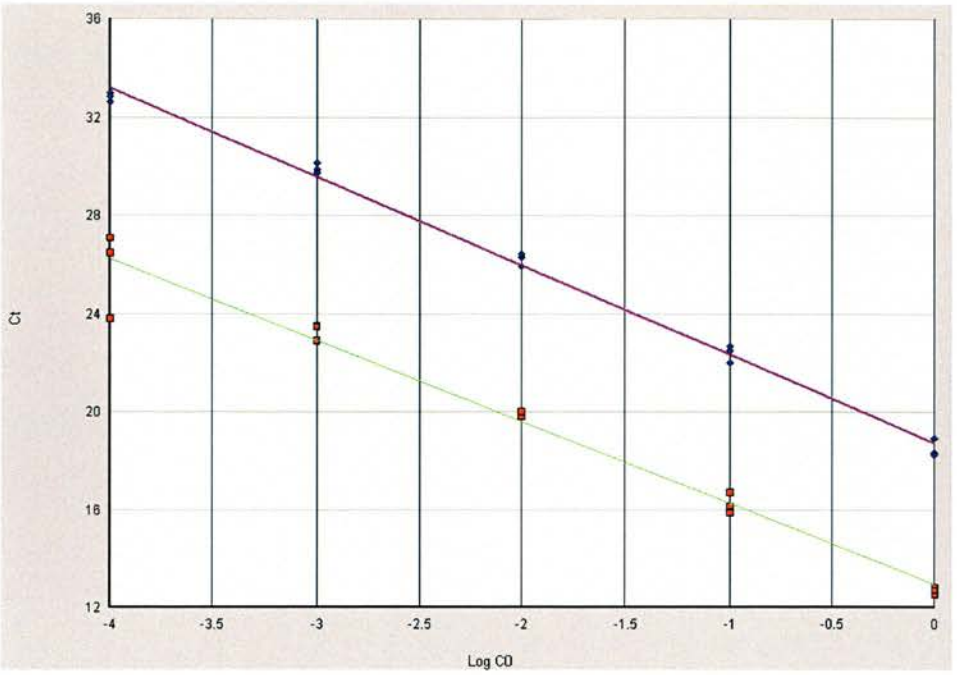


Figure 73: Standard curve plot of the log of the dilution plotted against the Ct values obtained with trendlines for Carcinin (red points green line) and β-actin (blue points purple line).

The standard curve plot for Carcinin (green) (Figure 73) produced a slope value of -3.3266 and R^2 of 0.9756. The standard curve plot for β-actin (purple) produced a slope value of -3.6166 and R^2 of 0.9952.

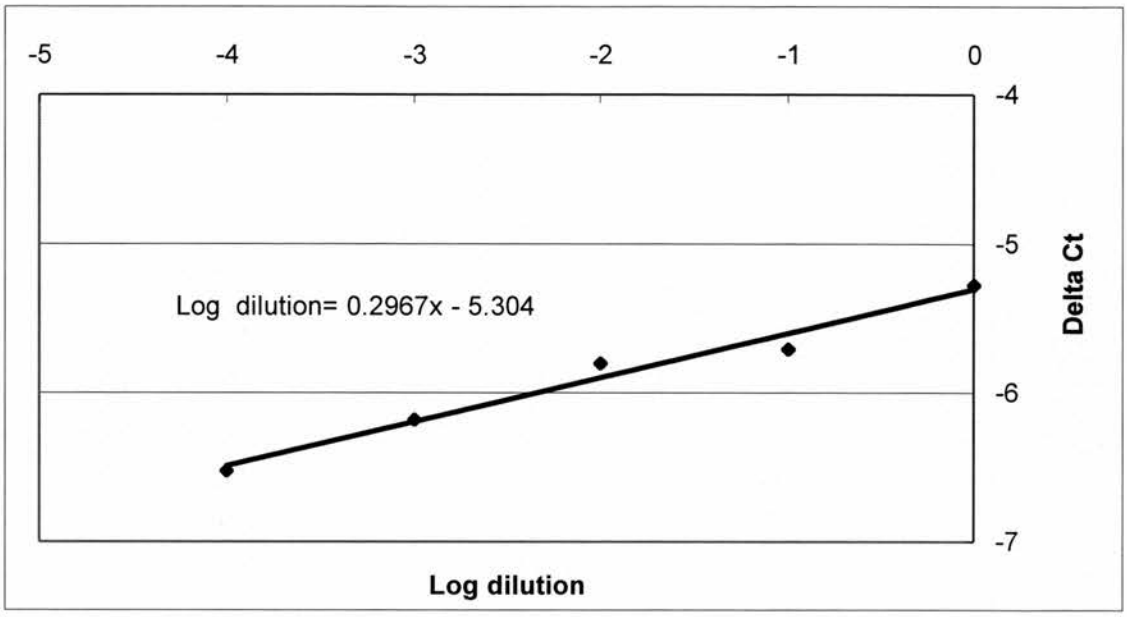


Figure 74:Plot of the log of the dilution versus ΔCt for validation of the comparative Ct analysis method for Carcinin with β-actin.

Investigation of relative amplification efficiencies showed that the Carcinin/β -actin plot has a slope value of 0.2967 (Figure 74). Therefore, as this slope value was greater than 0.1 (4.1.2), the comparative (ΔΔCt) method could not be used to analyse expression of these genes. Consequently, the relative expression analysis could only be investigated using the equation for the standard curve or the REST-XL © software.

4.4.4.1.2 REST-XL© Analysis

The amplification efficiency of these genes was calculated by REST-XL© using the equation described by Rasmussen (2001) of; Efficiency (E) = $10^{(-1/\text{slope})}$. This uses the slope value from the standard curve equation. The amplification efficiency of Carcinin was 2.00 (Table 29) (i.e. 100 % amplification efficiency) for this primer and probe set. The amplification efficiency obtained for β-actin was 1.89 (Table 29) (i.e. 94 % amplification efficiency). Using these amplification efficiencies, the REST-XL©

software program was used to assess the relative amplification of experimental temperature samples with Carcinin and β-actin probe sets.

cDNA dilution	Mean Reference Gene Ct β-actin	Mean Target Gene Ct Carcinin
(Undiluted)	17.96	12.68
0.1	21.95	16.24
0.01	25.74	19.94
0.001	29.47	23.29
0.0001	32.32	25.79
Slope	-3.62	-3.33
Efficiency	1.89	2.00
Correlation	-1.00	-1.00

Table 29: Amplification efficiencies calculated by the REST XL software package for Carcinin and β-actin.

4.4.4.2 AMPLIFICATION USING 18S RRNA ENDOGENOUS CONTROL

4.4.4.2.1 Standard curve method

The Carcinin amplification became detectable in later cycles than 18S rRNA amplification by an average ΔCt of 7 (Figure 77). This indicates that 18S rRNA is approximately 128 times more highly expressed than Carcinin (Figure 75 and Figure 76). The mean Ct value for Carcinin was 27.18, at a 10^{-3} template dilution (Figure 75). This template dilution was used for all samples analysed in the bacterial challenge experiments. A mean Ct value for 18S rRNA was 19.88 at a 10^{-3} template dilution (Figure 76).

Figure 75: Amplification plot for the Carcinin standard curve, using 1 – 10^{-4} dilution of a single cDNA template.

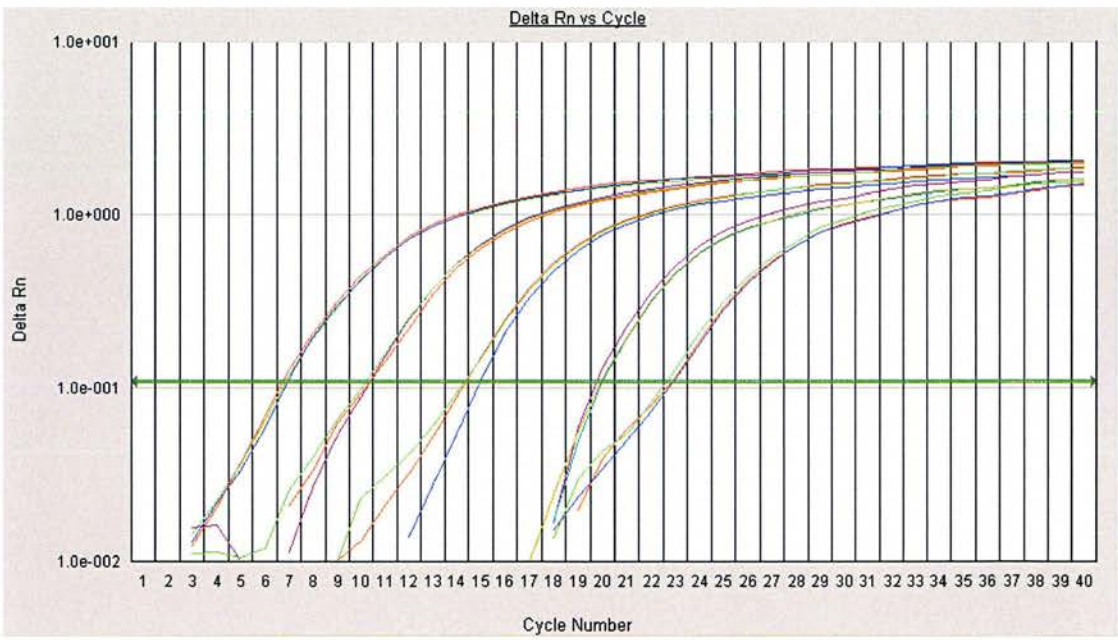


Figure 76: Amplification plot for the 18S rRNA standard curve, using 1 – 10^{-4} dilution of a single cDNA template.

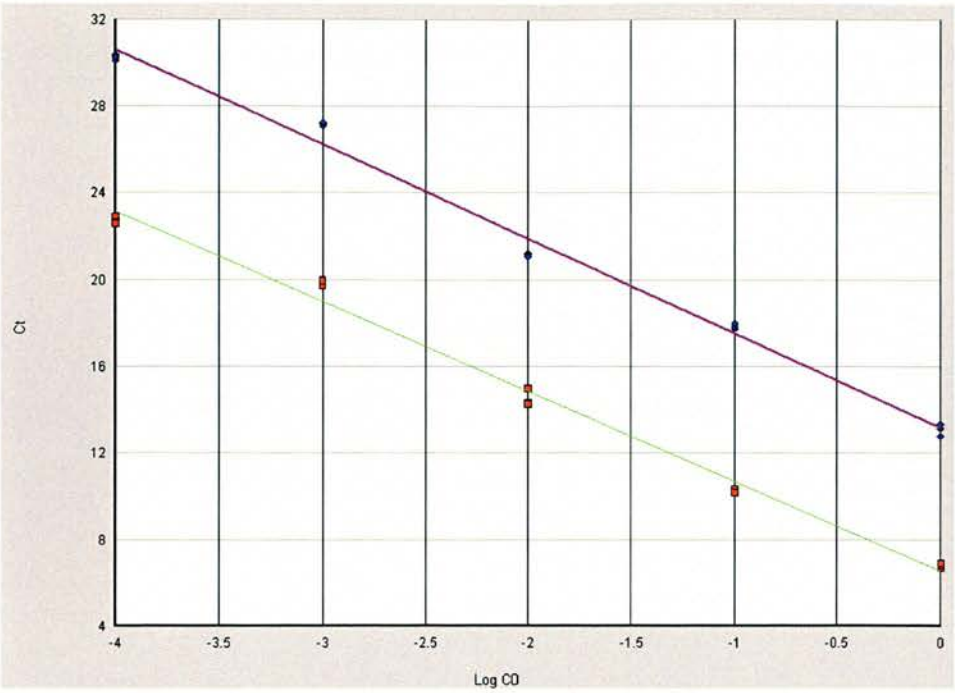


Figure 77: Standard curve plot of the log of the dilution plotted against the Ct values obtained with trendlines for Carcinin (purple) and 18S rRNA (green).

Standard curve plot (Figure 77) for Carcinin (purple) had a slope of – 4.3418 and an R^2 value of 0.9904. For 18S rRNA (green), the slope value was -4.1550 and the R^2 value was 0.9918.

4.4.4.2.2 REST-XL© Analysis

Investigation of relative amplification efficiencies showed that the Carcinin/18S rRNA plot has a slope value of 0.2017 (Figure 78). Therefore, the comparative ΔΔCt method could not be used to analyse the expression of these genes. Relative expression could only be assessed, for the Carcinin/18S rRNA experimental samples using the standard curve method or the REST-XL © software.

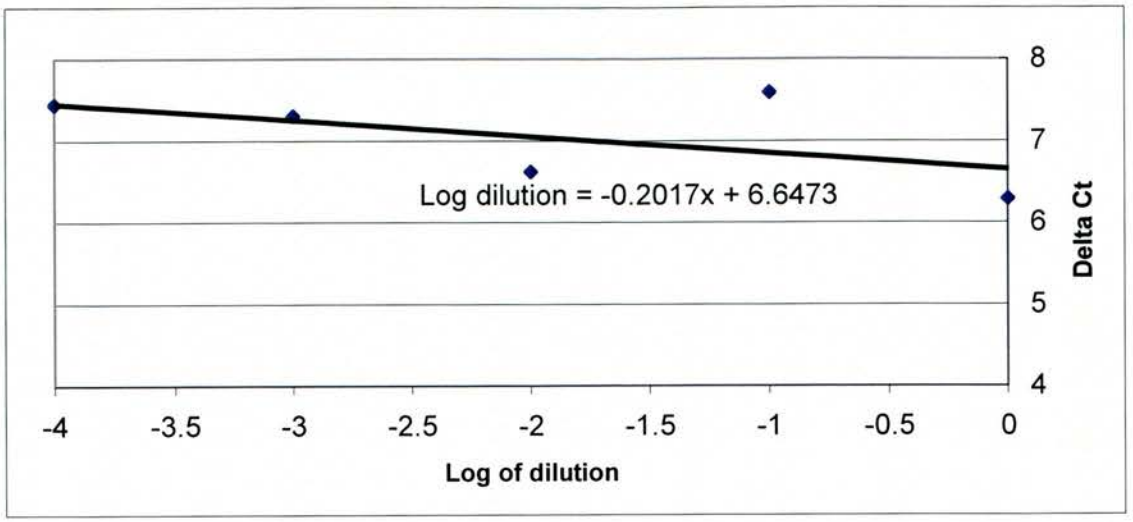


Figure 78: Plot of the log of the dilution versus ΔCt for validation of the comparative Ct analysis method for Carcinin with 18S rRNA.

PCR efficiencies of 1.74 and 1.70 were calculated for 18S rRNA and Carcinin, respectively (Table 30). This indicates that the amplification efficiencies were 87 % and 85 %, respectively.

cDNA dilution	Mean Reference Gene Ct 18S rRNA	Mean Target Gene Ct Carcinin
Undiluted	6.80	13.10
0.1	10.27	17.86
0.01	14.52	21.14
0.001	19.88	27.18
0.0001	22.77	30.22
Slope	-4.16	-4.36
Efficiency	1.74	1.70
Correlation	-1.00	-1.00

Table 30: Amplification efficiencies calculated by the REST-XL© software package for Carcinin and 18S rRNA.

4.5 Discussion

Partial transcript sequences were obtained for both β-actin and 18S rRNA from *C. maenas*. These sequences were used to design TaqMan® primers and probes to quantify relative expression. However, full transcript sequences could be obtained using RACE (Chapter 2) in conjunction with further cloning and sequencing.

The Primer Express ® program used to design the TaqMan® primer and probe set for Carcinin, returned a limited number of options. The candidate probe which was selected, annealed in a region which contained a known SNP (Chapter 2). Although this was not ideal, the probe was still used because the location of the primers ensured that all the other isoform transcripts identified in Chapter 2 would be amplified. It is possibility that transcripts with the alternative nucleotide at this site may not have been detected. However, the 5' position of the SNP reduces this likelihood.

Real-time PCR with shorter TaqMan® MGB probes can be used to quantify and distinguish transcripts which differ only by a single nucleotide (in the middle of the sequence). This could be useful for quantifying the relative expression of the variants of Carcinin transcripts in the future. The alleles are distinguished based on the difference in melting temperatures of the two probes (labelled with different dyes) and the mismatched probe is displaced by the correct sequence probe (Applied Biosystems, 2001) and then detected.

Although absolute quantification can be performed, and presents expression differences in terms of copy numbers, it is a more involved technique and relative quantification is often used to investigate trends in gene expression for single laboratory studies (Pfaffl *et al.*, 2002). Optimisation of the primer and probe concentrations is recommended by

ABI (Applied Biosystems, 2002). However, in the present study it was concluded that the costs associated with this process outweigh the advantages for the small sample sizes (<1000). The optimisation process in the current study, was useful to establish the minimum reaction volume (25 µl) and the optimum template dilution (10^{-3}) to be used for experimental samples. Both the 18S rRNA and the β-actin primer and probe sets could have been further optimised to improve the amplification efficiencies above 1.70 and 1.89, respectively but the resources required to do this were not available. Outliers observed on the standard curves (Figure 73) may explain the lower R^2 value and may be a result of pipetting inaccuracy between replicates. Optimisation of these outliers, by re constituting the standard curve reactions and running the amplification reaction again, was not possible.

It is widely accepted (as reviewed by Bustin, 2000), that 18S rRNA is often too highly expressed compared to many target genes and is therefore unsuitable as an endogenous control. However, it has also been recognised that this makes it suitable for use with highly expressed genes (Frost and Nilsen, 2003). The expression of Carcinin, relative to both candidate endogenous control genes, was unknown. However, it was expected that 18S rRNA expression would be prohibitively high. Consequently, β-actin was initially investigated for use as the endogenous control (Chapter 5, 5.4.1). Subsequent results, obtained in the bacterial challenge experiments (Chapter 5, 5.4.2 and 5.4.3), indicated that both control genes were expressed at comparable levels to Carcinin and could be used as endogenous controls without differential template dilutions.

The standard curve for Carcinin / 18S rRNA was prepared using a single cDNA template. It was subsequently realised that using a pool of cDNA templates would have been more appropriate but the cost of repeating this standard curve was prohibitive.

The advantages of using a pool of representative templates for the optimisation reactions and the standard curves were highlighted by the comparison between the amplification efficiencies obtained for Carcinin in the current study. Using pooled templates, an amplification efficiency of 2.0 was obtained but this dropped to 1.7 using a single template. This disparity may indicate that, although the probe and primer set for Carcinin are "good", the quality of the template used for the Carcinin / 18S rRNA standard curve was poor, leading to lower amplification efficiencies. Alternatively, the amplification efficiency might have been inhibited by unknown contaminants in the individual sample (McPherson and Møller, 2000). It could also be suggested that the amplification efficiency calculated for 18S rRNA is influenced by the quality of the template. Further optimisation would be required to clarify this issue, but this was beyond the scope of the present study. With hindsight, pooled templates would have been more appropriate for use in all optimisation experiments.

In summary, this chapter described the identification of two endogenous control sequences and the design of transcript tools used to quantify the transcript expression of each gene including the target gene, Carcinin. In addition, it was suggested that for sample size less than 1000 the benefit of optimisation of primers and probes did not outweigh the cost. It was also established that the expression levels of Carcinin were relatively high and that the transcript tools which were designed would be appropriate for use in subsequent investigations.

4.6 Conclusions

- Two endogenous control sequences from *C. maenas* were identified (18S rRNA and β actin) and efficient real-time PCR primers and probes were designed, optimised and tested.

- For small scale experiments of limited resources, optimisation was deemed unnecessary.

- Carcinin is expressed at extremely high levels in all individuals sampled in the local population.

- Due to differences in amplification efficiencies between the Carcinin / β-actin and Carcinin / 18S rRNA, REST-XL© was the appropriate method for analysis of the data.

CHAPTER 5

In Vivo Carcinin Gene Expression Analysis

5.1 Introduction

Bacterial challenge in the marine environment is predominantly attributable to Gram-negative bacteria, however, Carcinin appears to exhibit efficacy only towards Gram-positive marine bacteria (Relf *et al.*, 1999). The continuous high level production of a dedicated immunological protein, exhibiting specificity towards a group of organisms which are seldomly encountered by the animal, may seem to be an extravagant use of energy which may be required for other processes (moulting, reproduction etc). Indeed, studies in the mosquito, *Anopheles gambiae* have already shown that immunological activation can represent a considerable cost to reproductive fitness (Ahmed *et al.*, 2002). Thus, the expression of antimicrobial protein (AMP) transcripts in *C. maenas* may be expected to be regulated if this is their primary function. Conversely, the need to continuously mount a defence against many different types of pathogens, under different environmental conditions, should favour the unregulated production of proteins with a broad range of efficacy. Therefore, from an energetics point of view, it is possible to hypothesise that Carcinin is more likely to be closely regulated, compared to other broad range AMPs found in *C. maenas* (Schnapp *et al.*, 1996) and that expression levels may be modulated in response to external and/or internal cues.

To investigate this hypothesis in *C. maenas*, transcript expression of Carcinin was investigated in response to bacterial antigen challenge and temperature. These parameters, were selected because they have already been shown have influence on crustacean physiology (Breteler, 1975b; Dawirs and Dietrich, 1986; Moreira and Nelson, 1990; Cuculescu *et al.*, 1995; Aagaard, 1996; Whiteley *et al.*, 1997; Styrishave *et al.*, 1999; Terwilliger and Dumler, 2001; Coman *et al.*, 2002; Podrabsky and Somero, 2004) and immunology (Ravindranath, 1975; Truscott and White, 1990; Chisholm and

Smith, 1994; Vargas-Albores *et al.*, 1998; Le Moullac and Haffner, 2000; Mitta *et al.*, 2000a; Perazzolo *et al.*, 2002; Hernroth, 2003) of marine invertebrates. It has long been established, that transcript expression does not necessarily reflect expression of the functional protein. However, gene expression studies provide a relatively quick method of assessing the effects of selected parameters on gene regulation where protein quantification tools (e.g. antibodies) are unavailable.

Modulation of gene expression by temperature has been previously shown in other eurythermal animals (Goldspink, 1995; Kausel *et al.*, 1999; Mitta *et al.*, 2000a; Cadet *et al.*, 2002; Podrabsky and Somero, 2004), although no such studies have yet been published for the Crustacea. However, Chisholm and Smith (1994), have suggested that seasonal water temperature may influence the vigour of antibacterial activity in *C. maenas*. Therefore, the present study aims to investigate the relative expression levels of Carcinin at four temperatures within the physiologically tolerated temperature range of the local crab population. This would determine if there are considerable differences in expression between temperatures and whether or not temperature could be suggested to be a modulator of Carcinin expression.

Contrary to statements published by Raj and Dentino (2002), the relationship between defensin AMP transcript expression and microbial stimulation does not appear to be consistent throughout the Arthropoda,. In insects, AMP expression has been shown to be up-regulated in the fat body (Bulet *et al.*, 1999; Engstrom, 1999; Daibo *et al.*, 2001; De Gregorio *et al.*, 2002; Naitza and Ligoxygakis, 2004) and the granulocytes (Nakajima *et al.*, 2003), by microbial stimulation. On the other hand, haemolymph studies in chelicerates (e.g. *Androctonus australis*) have shown that gene expression is not regulated by microbial stimulation (Cociancich *et al.*, 1993; Ehret-Sabatier *et al.*,

1996). Several gene expression studies have been conducted in Crustacea (Whiteley *et al.*, 1992; El Haj and Harrison, 1994; Towle *et al.*, 1995; Towle *et al.*, 1997; Whiteley and El Haj, 1997; Lee *et al.*, 1998; Chang *et al.*, 1999; Kotlyar *et al.*, 2000; Frost and Nilsen, 2003; Rewitz *et al.*, 2003), although most have focussed on genes associated with the moult, growth or homeostasis. Immunologically focussed studies, are accumulating and have been conducted in penaeid shrimps (Destoumieux *et al.*, 1999; Destoumieux *et al.*, 2000; Munoz *et al.*, 2002; Rojtinnakorn *et al.*, 2002; Roux *et al.*, 2002; Munoz *et al.*, 2003; Rattanachai *et al.*, 2004), crayfish (Cerenius *et al.*, 2003; Söderhäll *et al.*, 2003) and very recently in the lobster (*Homarus gammarus*) (Hauton *et al.*, 2005). The present study is the first to report an immunologically based gene expression investigation in a brachyuran (true) crab. Of the studies in other species, for example using the penaeid shrimps (Destoumieux *et al.*, 1999; Destoumieux *et al.*, 2000; Munoz *et al.*, 2002; Munoz *et al.*, 2003; Vargas-Albores *et al.*, 2004), studies have focussed on the responses of AMP transcript expression to microbial stimulation, using traditional quantification methods. These studies have reported conflicting results on the modulatory effect of microbial stimulation on AMP transcript expression. In light of the contrasting reports on the responses elicited by microbial challenge in this phylum, relative Carcinin expression levels, resulting from microbial challenge, were investigated in *C. maenas* in the present study using a highly sensitive technique (real-time quantitative PCR). Although this technique has recently started to be used in Crustacea (Dhar *et al.*, 2001; Ouaissi *et al.*, 2002; Donald *et al.*, 2003; Hauton *et al.*, 2005), the present study is thought to be the first investigation to use this technique to assess antimicrobial gene expression in a marine invertebrate.

Investigating the responses of *C. maenas* to a range of parameters may help to predict the "immunological capability" of these organisms under different conditions.

Sequences which may be homologous to Carcinin have been published for some commercially important Crustacea such as: *Litopenaeus vannamei* (AF430071-76), *Litopenaeus setiferus* (AF430078), *Panulirus leniusculus* (AF522504), *Panulirus argus* (AY340636), *Marsupenaeus japonicus* (AB121740-44), *Homarus americanus* (CN853187), *Callinectes sapidus* (CV022228 & CV006490) and *P.monodon* (BIO18072-74) and most recently in *H. gammarus* (AJ786653). Therefore, the investigation of factors affecting the expression profile of these immunologically important genes may have considerable implications in aquacultural husbandry.

5.2 Specific Aims

- Confirm whether the Carcinin transcript is expressed "constitutively" and at assess these levels compared to endogenous controls.

- Quantify relative expression at different temperatures within the thermal tolerance range.

- Quantify relative expression changes resulting from the injection of bacterial antigen between 0 to 84 h after injection at both 5 °C (extreme exposure temperature) and 15 °C (frequent exposure temperature).

5.3 Methods and Materials

5.3.1 Temperature Experiment

Twenty female crabs were collected between August and September 2001 (Appendix 26), from St. Andrews Bay and maintained as described in Chapter 2 (2.3.1). Five crabs, selected for each treatment, were acclimatised (~2 °C min^{-1}) to a constant temperature (5 °C, 10 °C 15 °C or 20 °C) and maintained at these temperatures for two weeks, with feeding, as previously described (Chapter 2; 2.3.1). The 5 °C ± 2.0 °C and 10 °C ± 2.0 °C experiments were performed in specifically designed "cold rooms" and the 20 °C ± 2.1 °C temperature was maintained using an electrical water heater (Grants Instruments Ltd., Cambridge, UK). The 15 °C ± 2.2 °C samples were maintained at ambient aquarium water temperature, which was monitored daily using an electronic thermometer (KM812, Comark Ltd., Hertfordshire, UK).

5.3.2 Bacterial Challenge Experiments

One hundred and forty eight male crabs (*C. maenas*) were collected from St. Andrews Bay and maintained as described in Chapter 2 (2.3.1) and 5.3.1 above. Carcinin expression was analysed at pre-determined time intervals after the injection of either heat-killed bacteria (bacterial antigen) or a saline placebo control and all samples were processed in parallel. Crabs were not fed after experimental manipulation. Four male crabs were used in each treatment, at each time interval and were all sampled only once. The bacterial challenge experiment was conducted in two parts. First, a short-term experiment (n= 20) was performed with samples collected at 0 h, 0.5 h and 3 h intervals post injection (@ 10-15 °C). Next, two longer term experiments were performed (n= 128), with samples collected at 0 h, 6 h, 12 h, 24 h, 48 h and 84 h post injection at 5 °C

± 2.0 °C and also at 15 °C ± 2.2 °C. The 0 h group comprised of crabs, which were handled, but experienced no other experimental manipulation. In both of the longer experiments (0-84 h at 5 °C and 15 °C), an additional "wounding" control group was also added to the protocol and all sampled were processed in parallel.

The bacterium selected for injection was *Planococcus citreus* (NCIMB 1493), known to be killed by the antibacterial activity of Carcinin (Relf *et al.*, 1999). Previously, a monoculture of *P.citreus* was grown overnight (~12 h) from stock plates to logarithmic phase in 10 ml of marine broth (Difco 2216, Detroit, USA) at RT (~23 °C) on an orbital shaker. The bacteria were then harvested and washed with sterile 3.2 % (w/v) NaCl as described by Schnapp *et al.* (1996). After resuspension in 5 ml of sterile 3.2 % (w/v) NaCl, the bacteria were heat killed (bacterial antigen) by incubating at 60-80 °C for 1 h. Heat killing was verified by the absence of colony formation after spread-plating 40µl of the bacteria onto marine agar (Difco) plates and incubating at RT for 10 h. The bacterial antigen was then resuspended in sterile crab saline solution (580 mM NaCl, 13 mM KCl, 13mM $CaCl_2.2H_2O$, 50 mM Tris Base, 1 M HCl pH 7.4) (Smith and Ratcliffe, 1978) to a previously calibrated (with live bacteria) final concentration equivalent to ~1 x 10^8 colony forming units per ml (cfu ml^{-1} of live bacteria). This concentration was selected based on previous studies of immunological responses in *C. maenas* using sub lethal doses (10^5 - 10^9 cfu ml^{-1}) of bacteria (Smith and Ratcliffe, 1980b, a, 1981; White and Ratcliffe, 1982; White *et al.*, 1985; Hauton *et al.*, 1997).

Before injection, each crab was dried with a paper towel and the left cheliped swabbed with cotton wool soaked in 70 % ethanol (BDH). Experimental crabs were injected with 0.5 ml (~10^7 cfu) of *P.citreus* (bacterial antigen), in crab saline solution using a 23 g needle into the unsclerotised membrane between the carpus and the merus of the left

cheliped. Placebo controls were injected with an equivalent volume of sterile crab saline solution. The wounding controls were stabbed for 10 s with a 23 g needle. After treatment, all crabs were returned to the experimental tanks until sampled at the appropriate time interval except for the 0 h group which was immediately processed (5.3.3)

5.3.3 Template Preparation

The haemolymph extraction, cell counting and total RNA isolation were performed as described in Chapter 2 (2.3.1 and 2.3.2) and all animals were only bled once. As described in Chapter 2 (2.3.1 and 2.3.2), the concentration of each total RNA sample was measured and the quality was checked. The correlation between total RNA and total haemocyte counts (THC) was investigated and the differences in THC recorded at each temperature were tested using a one-way analysis of variance (ANOVA). As described in Chapter 4 (4.3.1), all RNA samples were subsequently DNAse treated and reverse transcribed with random hexamers and the resultant cDNA template stored at – 20 °C.

5.3.4 Transcript Amplification and Analysis

Amplification reactions were prepared in triplicate using the optimised primer and probe concentrations determined in Chapter 4 (4.4.3). In the temperature experiment (5.3.1), β-actin was used as the endogenous control. In the initial (0-3 h) bacterial antigen experiment (5.3.2), both β-actin and 18S rRNA were used as endogenous controls for evaluation purposes (Chapter 4; 4.3.6); in the longer term bacterial antigen experiment (0-84 h) only 18S rRNA was used.

The optimum template dilutions, established in Chapter 4 (4.4.3), were incorporated into the 25 µl optimised reaction volumes (as opposed to 50 µl optimised in Chapter 4; 4.3.4). The same reaction mixtures, instruments and universal cycling conditions were used for all samples as previously described (Chapter 4; 4.3.4).

5.3.5 Data Analysis

Using the mean Ct values for each experimental and control sample the relative expression of Carcinin was assessed using the REST-XL© software. The amplification efficiencies established for each gene in Chapter 4 (4.4.4) from the standard curves, were used in the REST-XL© calculations. The number of iterations used for each calculation for the Pair Wise Fixed Randomisation Test © was set at 2000.

The average sea temperature in St. Andrews Bay (Appendix 26) for the sampling year was 10 °C. Therefore, in the temperature experiment the 10 °C samples were used as the "control" samples in the REST-XL© analysis.

In the bacterial challenge experiments, the samples from the un-manipulated crabs represented the "control" samples within the treatments, whereas analysis between treatments was performed using the saline or wounding samples as "controls" in the REST-XL© analysis.

5.4 Results

Total RNA was successfully isolated from all haemolymph samples. As described in Chapter 2, the quality of each RNA sample was assessed by gel electrophoresis (Figure 79) and the quantity determined by absorbance (Farrell, 1996). All samples were assessed to be of sufficient quality for subsequent analysis (Figure 79); this was determined by the presence of clear 28S and 18S rRNA bands and slight smearing throughout lane, indicative of the presence of mRNA (Farrell, 1996) as seen in Chapter 2. Again, a greater intensity was observed for the 18S band compared to the 28S bands in all samples from *C. maenas*, similar to the results in Chapter 2. Overall, RNA recovery varied from 70 μg ml^{-1} to 3720 μg ml^{-1} in DEPC water (Appendix 27 and Appendix 28), for the same isolation procedure from the same amount of starting material (per 0.3 ml haemolymph, see Chapter 2; 2.3.1 for isolation and purification method).

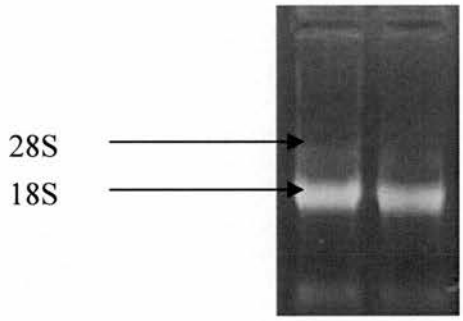


Figure 79: Typical total RNA isolated from *C. maenas* haemolymph showing 18S and 28S bands (all samples were similar).

The THC for the control animals sampled in the current experiment were found to vary between 3.60 x 10^6 to 1.14 x 10^8 per ml of haemolymph (Appendix 27 and Appendix 28)

5.4.1 Effect of Temperature on Carcinin expression and THC

Carcinin, normalised to the β-actin endogenous control gene, was significantly up-regulated by 3.56 fold ($p = 0.010$) in the 5 °C samples, compared to the 10 °C samples (Table 31). No change in expression was observed between the 15 °C samples and the 10 °C samples. Carcinin was significantly up-regulated at 20 °C, compared to the 10 °C control group by 2.917 ($p = 0.005$).

Temperature (°C)	Not Normalised β-Actin	p	Not Normalised Carcinin	p	Normalised Carcinin	p
5	↓ by 1.527	**0.366**	↑ by 2.334	**0.142**	↑ by 3.563	**0.010**
15	↓ by 4.538	**0.035**	↓ by 13.108	0.553	↑ by 1.461	0.494
20	↓ by 13.108	**0.003**	↓ by 4.494	0.035	↑ **by 2.917**	0.005

Table 31: Fold changes in expression of Carcinin and β-actin at different temperatures indicating non-normalised and normalised values. Significant changes are shown in bold with the corresponding p value.

THCs varied between 2-6 x 10^7 cells per ml of haemolymph for the temperature experiment samples (Figure 80). The THC exhibited only a very weak correlation ($R = 0.217$) with total RNA concentration, which was not significant ($p > 0.100$) (Appendix 27) ANOVA analysis showed that the 10 °C THC were significantly ($p < 0.010$) higher than the THC made for the other three temperature samples (Figure 80).

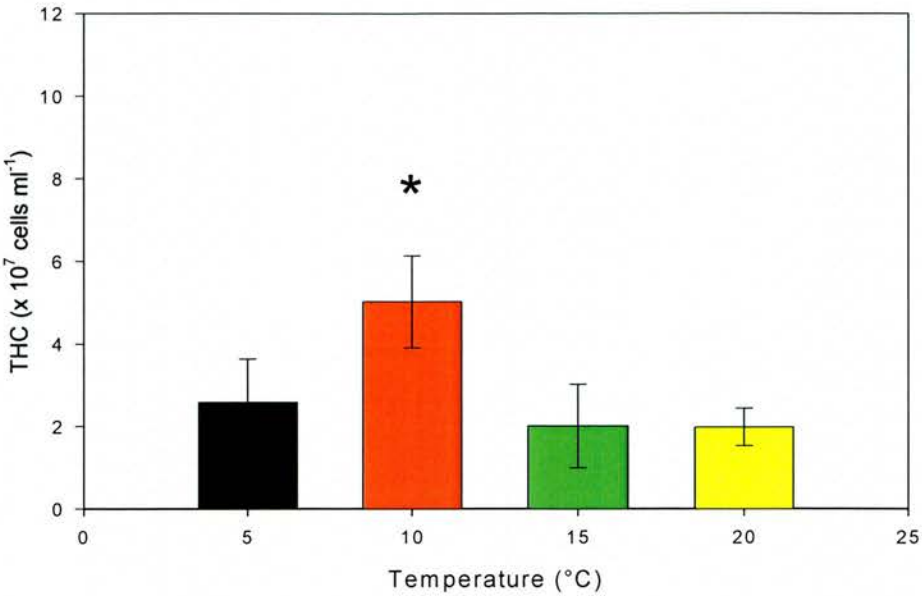


Figure 80: THC at different temperatures, asterisk (*) indicates a significant difference ($p = 0.01$) observed at 10 °C compared to the other three temperatures (+/- St.Dev., 5 observations in each temperature).

5.4.2 Change in Carcinin Expression and THC, 30 min and 3 h after Bacterial Antigen Injection

No significant changes in Carcinin expression (using either 18S rRNA or β-actin as endogenous control) were observed because of bacterial antigen (*P.citreus*) or placebo injection, at either 30 min or 3 h after injection.

There was also no significant difference in THC between the saline placebos and the controls (un-manipulated, T=0), 30 min or 3 h after injection (Figure 81). In contrast, crabs injected with bacterial antigen demonstrated a significant drop in cell counts compared to both the un-manipulated and placebo controls ($p < 0.001$ and $p < 0.010$ respectively) (Figure 81) after 30 min and 3 h. The THC in this part of the experiment, followed the general trend described by others (Smith and Ratcliffe, 1980b, 1983; Hauton *et al.*, 1997; Lorenzon *et al.*, 1999) in response to the injection of saline placebo

or bacterial antigen (Appendix 28). The small drop in the saline treated samples is probably attributable to natural variability and dilution effects. The observed THCs were not significantly different from the controls ($p > 0.01$). It must also be noted however that, at the point of inoculation, there will be a small localised immune response due to injury; this response would involve phagocytosis (Smith and Ratcliffe, 1980b), melanisation and the degranulation of cells which may contribute to the observed reduction in cell numbers (reviewed among others, by Bachau, 1981).

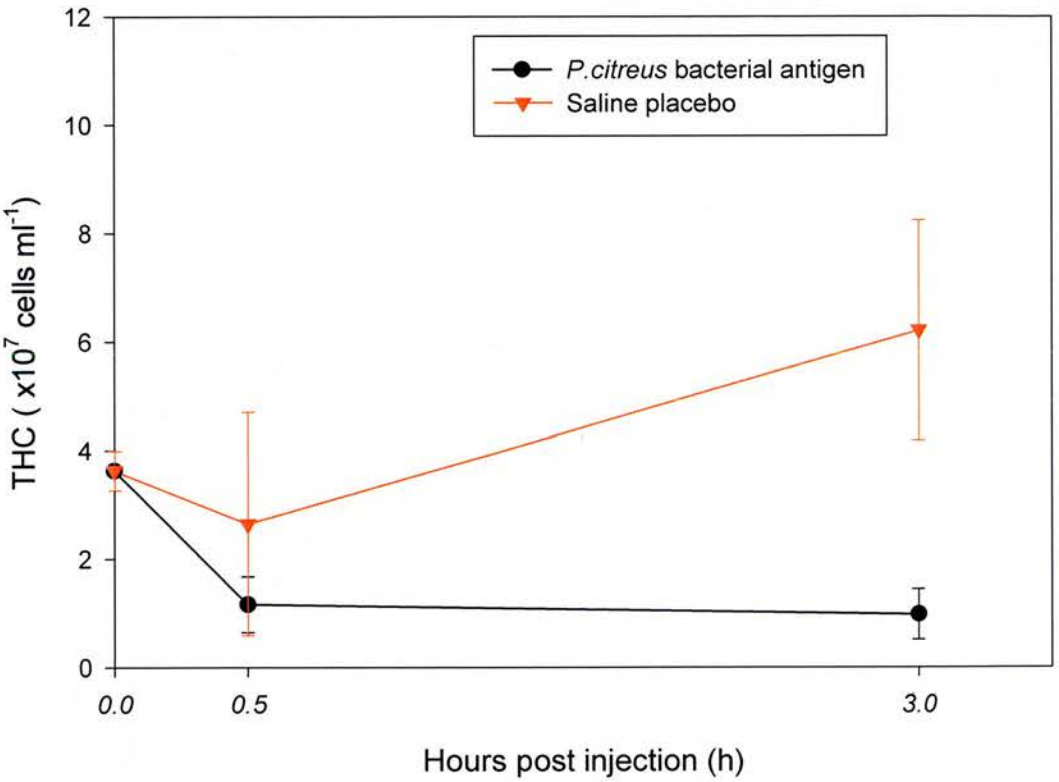


Figure 81: THC at 0 min, 30 min and 3 h after injection with *P. citreus* bacterial antigen (black line and circles) or a saline placebo (red line and triangles) showing the mean +/- St.Dev. of 4 observations.

5.4.3 Change in Carcinin Expression and THC, 6-84 h after Bacterial Antigen Injection

No significant changes in Carcinin transcript expression were observed within each treatment (i.e. within either bacterial antigen, saline placebo or wounding treatment), at

5 °C, at any time point after injection (T = 6, 12, 24, 48, 84 h) when compared to un-manipulated controls (T = 0). In addition, no significant changes in expression were observed at 5 °C, when each time point was compared between treatments (e.g. 6 h bacterial antigen compared to wounding at 6 h) (Figure 82). At 15 °C, comparison of the bacterial antigen injected samples to un-manipulated controls (T = 0) also failed to show significant changes in Carcinin expression after 6, 12, 24 and 48 h. However, at 84 h, a significant ($p = 0.023$) down-regulation of Carcinin, by 1.360 fold, was observed. However, this down-regulation, was not observed at 84 h for the wounding or placebo injection treatments when compared to their un-manipulated controls (T= 0 h). In addition, no significant changes in expression were observed between treatments, at any time point after manipulation (Figure 84). Finally, no significant difference in expression levels were observed between similarly manipulated samples or control samples between 5 °C and 15 °C, at any time interval.

Similar to the THC results for the temperature experiment (Appendix 27), a weak correlation was observed for the bacterial antigen challenged experiments with an R-value of 0.307 (Figure 82), however, this correlation was significant at $p = 0.01$.

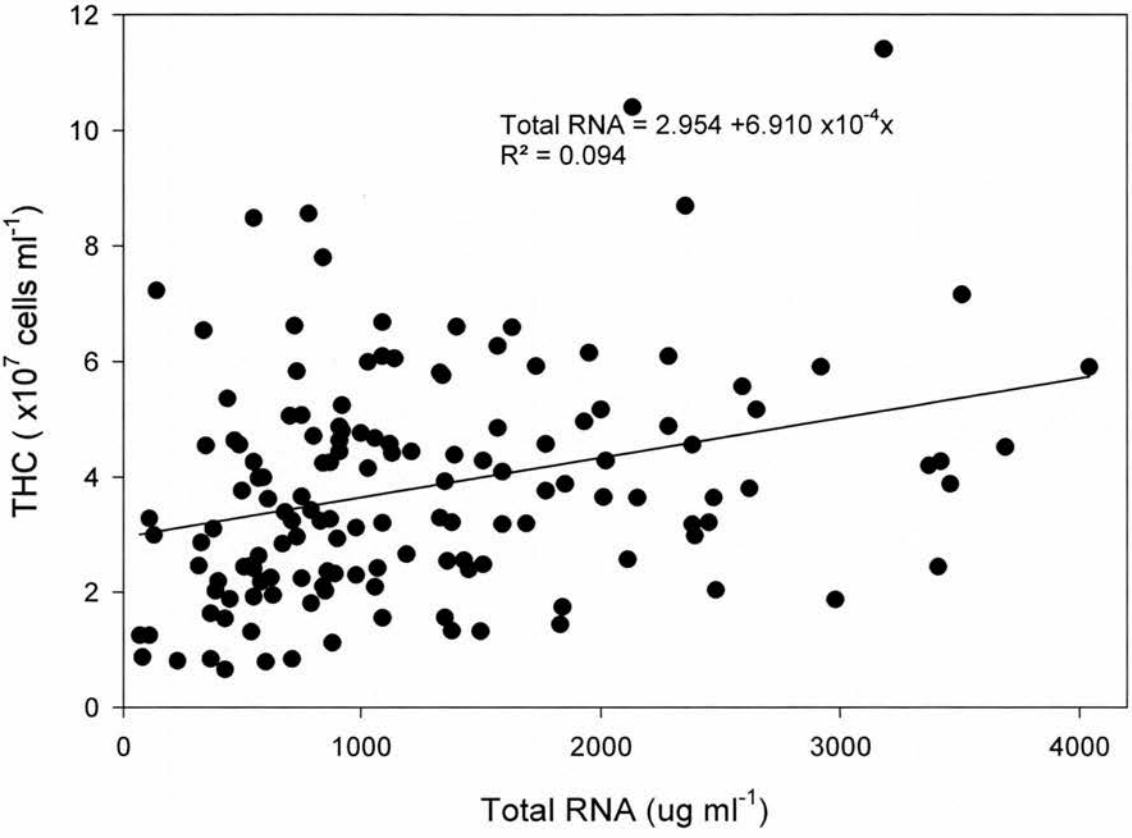


Figure 82: Correlation plot of cell counts to total RNA concentrations (μg ml^{-1}) for the 6-84 h heat-killed bacterial challenge experiments ($p > 0.01$ for 126 degrees of freedom).

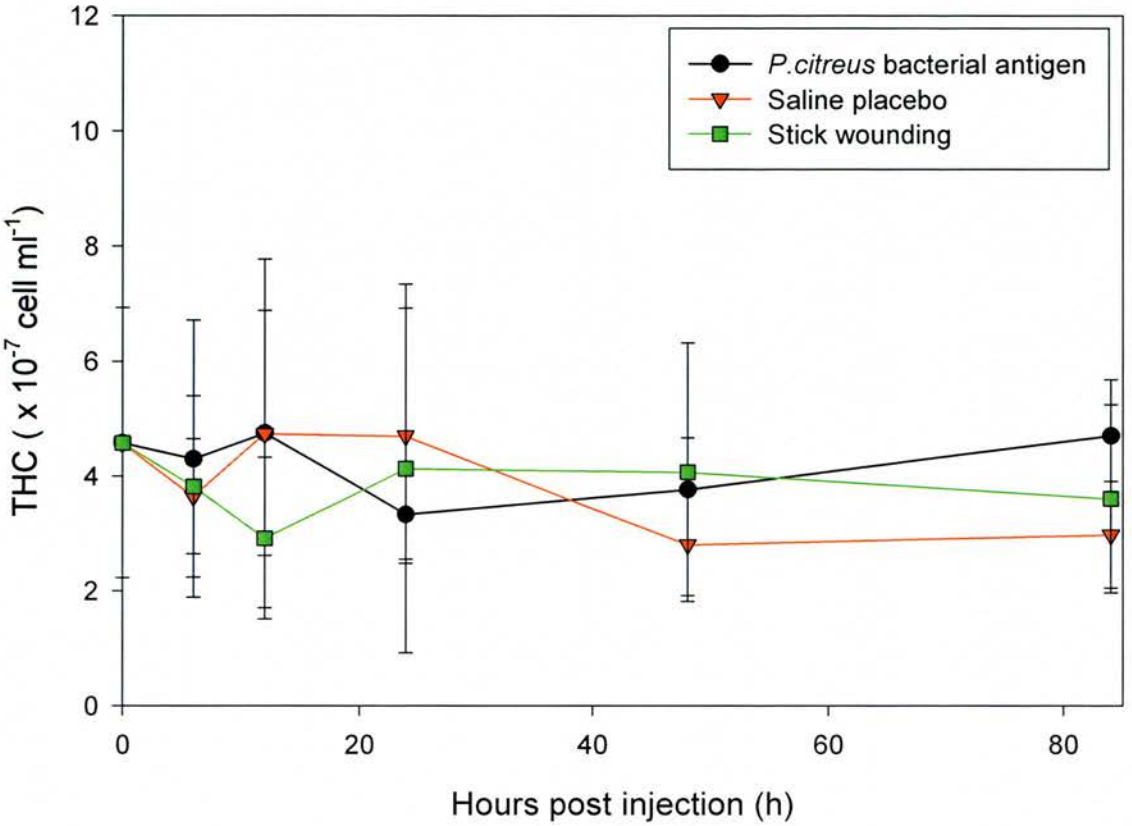


Figure 83: Mean THC at 5 °C, at time intervals (h) after injection of *P. citreus* bacterial antigen (black lines and circles), saline placebo (red lines and triangles) or stick wounding (green lines and squares) showing the mean +/- St.Dev. of 4 observations.

At both 5 °C and 15 °C, no significant differences in cell counts were observed either within or between treatments. The THC recorded, demonstrated a high level of inter-animal variability (Appendix 28) that is reflected in the size of the error bars at both temperatures (Figure 83 and Figure 84).

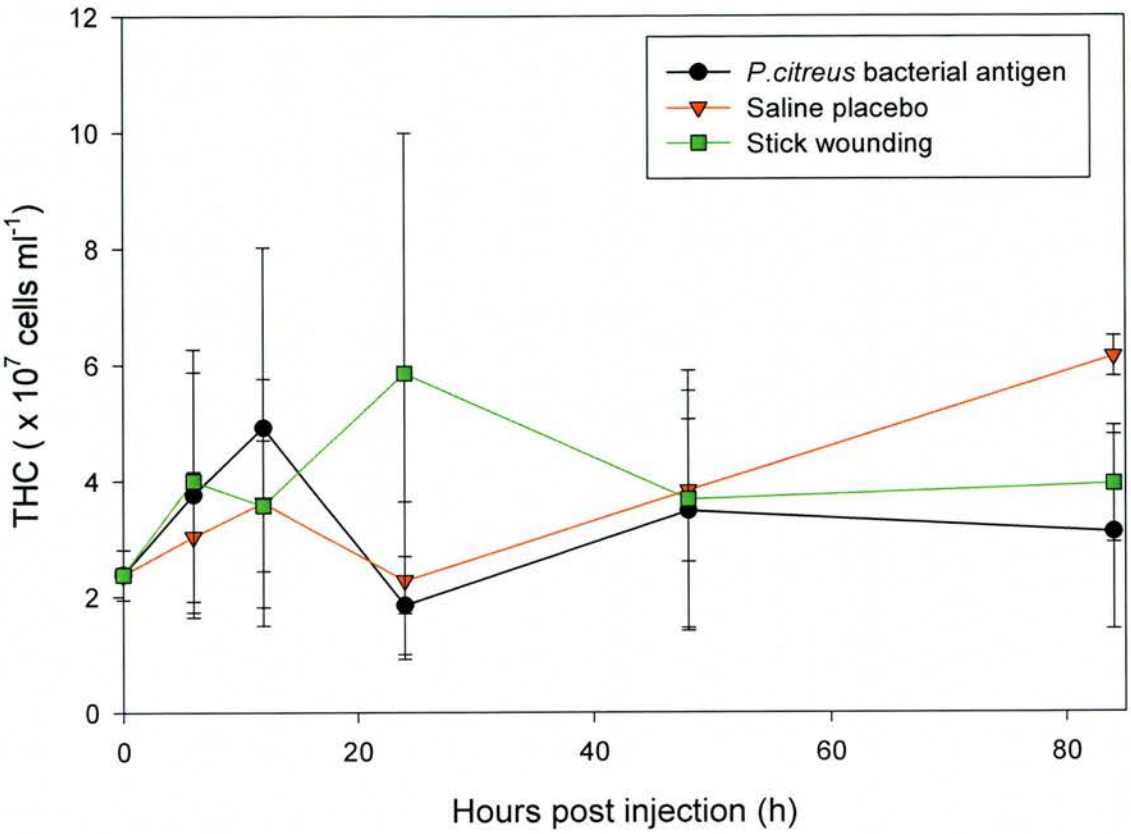


Figure 84: Mean (n=4) THC at 15 °C, at time intervals (h) after injection of *P. citreus* bacterial antigen (black lines and circles), saline placebo (red lines and triangles) or stick wounding (green lines and squares) showing the mean +/- St.Dev of 5 observations.

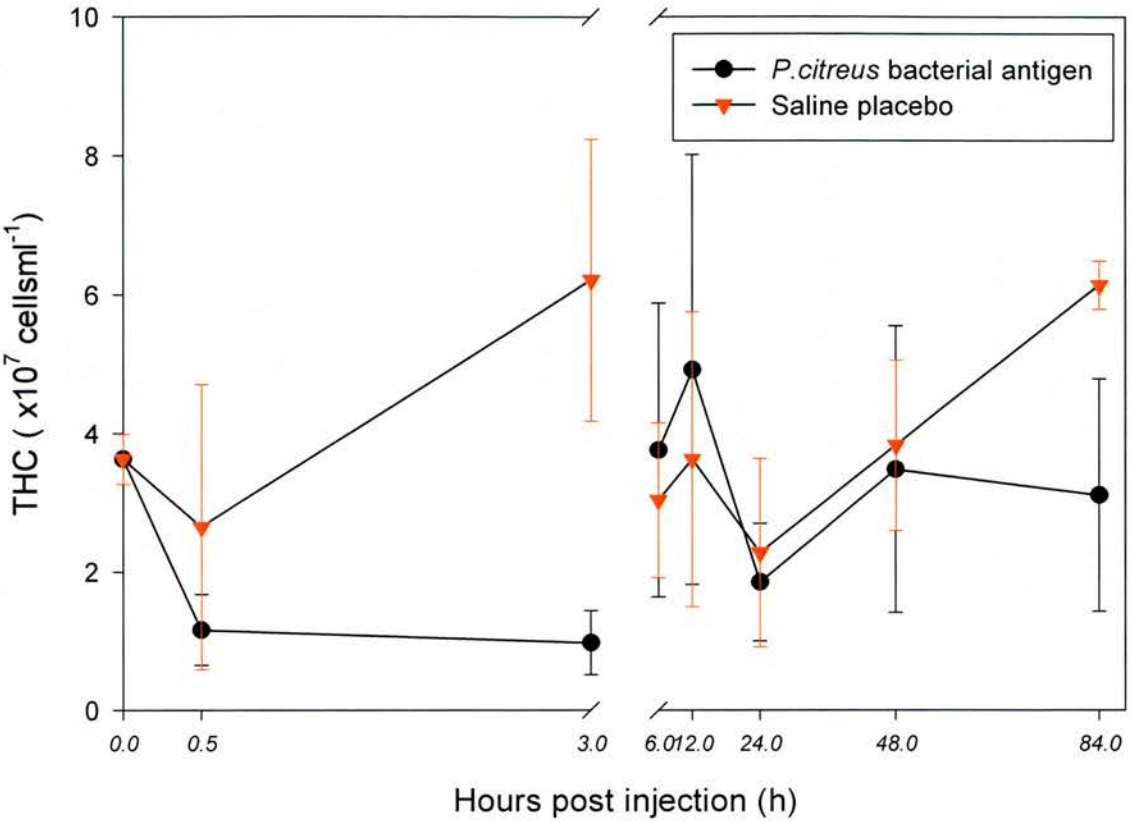


Figure 85: Mean THC of samples injected with *P.citreus* (black lines and squares) or saline placebo (red lines and triangles) at 15 °C from the two separate experiments (i.e. 0-3 h and 6-84 h).

When the results for both short term and long term experiments for the bacterial antigen injected crabs at 15 °C, were placed on the same figure (Figure 85), the THC could be suggested to return to normal levels between 3 and 6 hours, which is much earlier than reported in earlier studies (Smith and Ratcliffe, 1980b).

5.5 Discussion

Good quality RNA was extracted from all samples, although concentrations isolated from the same volume of haemolymph, were highly variable (70-3720 μg ml^{-1}), as were recorded THC. Although the correlation between these two parameters was relatively weak, it was significant, confirming that the variability in the total RNA concentrations was indeed a reflection of the THC and not the isolation method employed. However, the correlation could be improved further in future studies, by using a more precise RNA quantification method such as RiboGreen ® RNA Quantitation Reagent (Molecular Probes, Leiden, Holland). RiboGreen® is a fluorescent stain which binds only to double stranded nucleic acids and not to proteins or free nucleotides which can influence absorbance readings at 260 nm.

Endogenous control genes were used in the present study to normalise for fluctuations in mRNA resulting from differences in starting material and purification efficiency of total RNA. Some authors have suggested that mRNA levels should be correlated to cell counts (Munoz *et al.*, 2002; Pfaffl, 2004). Due to the weak correlation observed here between THC and total RNA, this was not considered an appropriate normalisation method for the present study. In Chapter 4, both 18S rRNA and β-actin were established as valid endogenous controls as they were constitutively expressed at comparable levels to Carcinin.

Initially, β-actin was used as endogenous control in the temperature experiments. The temperatures selected for investigation in the first part of this chapter, represent the seasonal range observed locally (Appendix 26). Although sea temperatures rarely reach 20 °C in Scotland, it is possible that *C. maenas* will experience this temperature if

stranded in a rock pool on a receding tide. The initial results of the temperature experiment showed that Carcinin is up-regulated by 3.565 and 2.917-fold, at 5 °C and 20 °C, respectively, compared to the 10 °C control samples. These results appear contrary to the most recent study which reported decreased antimicrobial vigour at these temperatures (Chisholm and Smith, 1994). However, Chisholm and Smith (1994) had used a Gram-negative marine bacterium (*Pyschrobacter immobilis*) to quantify antimicrobial vigour and not the Gram-positive *P.citreus*. Support for the up-regulation of Carcinin transcript observed in the present study, is provided by an earlier unpublished study by Karen Bell (Smith and Chisholm, 1992) in which the strongest antimicrobial activity towards *P.citreus* (Gram-positive) was exhibited by crabs acclimated at 5 °C, compared to crabs acclimated to 13 °C or 19 °C. Therefore, a possible hypothesis is that the increased transcription observed at 5 °C, in the present study, may lead to an increased abundance of anti Gram-positive AMPs, such as Carcinin. Therefore, the reduced vigour towards Gram-negative bacteria at this temperature, observed by Smith and Chisholm (1994), might be expected if haemocyte resources are focussed on the production of anti Gram-positive antibacterial compounds (e.g. Carcinin) and might help to explain the contrasting results.

At 15 °C, it was noted that the non-normalised level of β-actin was significantly down-regulated whereas the Carcinin was not. Although this did not affect the normalised Carcinin expression, it may suggest that β-actin is regulated by this temperature. At 20 °C, non-normalised β-actin and Carcinin were both significantly different from the 10 °C controls, although it cannot be determined whether this was due a lower template concentration added to the reaction mixture or actual regulation of the endogenous control. It is possible that β-actin may not be an appropriate endogenous control under

certain temperature conditions. Previous studies with Crustacea have reported changes in the expression of β-actin (Whiteley and El Haj, 1997) with temperature and growth (Whiteley *et al.*, 1992; Harrison and El Haj, 1994). However, this gene has been used successfully in other expression studies (Towle *et al.*, 1995; Towle *et al.*, 1997; Lee *et al.*, 1998; Kotlyar *et al.*, 2000; Rewitz *et al.*, 2003) although none of these have used real-time PCR methods. In light of the possible temperature dependent variability of β-actin, the up-regulation of Carcinin at 20 °C must be treated with some caution. Further confirmation of this regulation could be investigated by looking at expression of β-actin normalised to another endogenous control gene, but this was not possible within the scope of the present study.

Using northern blotting, Destoumieux *et al.* (2000) originally suggested, that the expression of penaeidins (in *P.vannamei*), normalised to total RNA, were down-regulated 3 h post injection (bacterial antigen), to 25 % of the original level. These levels were then thought to return to normal after 12 h and may even be up-regulated 48 h post injection, although the significance of this was unconfirmed. These changes in expression of the penaeidin transcripts were subsequently attributed to changes in circulating haemocytes numbers and not to changes in expression (Munoz *et al.*, 2002). In the present study, both β-actin and 18S rRNA were used individually as endogenous controls in the preliminary bacterial antigen study. Carcinin expression levels were not affected at 15 °C, 30 or 3 h min after the injection of either saline or bacterial antigen. This result concurs with the results published by Munoz (2002) for *P.vannamei*. Due to the costs associated with separately amplifying three genes and considering the possible regulation of β-actin at 15 °C and 20 °C, only 18S rRNA was selected for use in the subsequent longer term challenge experiment.

At 5 °C, in the present study, no changes in Carcinin transcript expression were identified between 6 and 84 h after the injection of bacterial antigen, saline placebo or by wounding. A similar result was observed for the 15 °C samples except a significant down-regulation of Carcinin transcript was observed at 84 h. This result was not due to a wounding response or a response to the injection of saline as no changes in expression were observed in these controls. Although the down-regulation could be attributed to the injection of *P.citreus,* it must be noted that the preceding samples (6-48 h) did not show any evidence of this trend. Therefore, the sudden down-regulation of this transcript at 84 h may not be part of a specific immunological response to *P.citreus.* One explanation may be, that the high level of transcription of this gene may decrease after sufficient Carcinin has been translated into protein and is stored in the granulocytes of the cells, but this hypothesis would have to be proven. The fact that this regulation was not observed in the 84 h samples at 5 °C may be a result of slower metabolism at lower temperatures as observed in other crustaceans (Breteler, 1975b; Moreira and Nelson, 1990). Alternatively, the expression of Carcinin may be controlled by physiological parameters beyond the scope of the present study, such as growth or tissue repair. Finally, this result could also be due to a Type 2 statistical error which suggests a significant difference where there is none (Fowler and Cohen, 1996). Considering these contrasting explanations it would be prudent to first repeat the experiment to confirm these results and then perhaps to expand the experiment in sample size and time frame, particularly at 5 °C. Additional measurements of physiological parameters (e.g. cell proliferation) may also help to identify other possible regulators of this transcription. As observed in the temperature experiment, no significant difference in expression in the bacterial antigen study was observed between samples at 5 °C and 15 °C.

A recent study, in *P.vannamei,* has reported differential regulation of two isoforms of a transcript with some identity to Carcinin. This study reported down-regulation in one of the isoforms, 6 h after the injection of bacterial antigen although the expression levels of the other isoform remained unaffected (Vargas-Albores *et al.*, 2004). The method used to quantify expression changes (intensity comparison between electrophoresed PCR amplicons) in this study suggested that the results may also be reflective of haemocyte numbers and not of changes in transcript expression (Munoz *et al.*, 2002).

Several workers (Bustin, 2002; Vandesompele *et al.*, 2002; Pfaffl, 2004; Radonic *et al.*, 2004) have recommended the simultaneous use of several endogenous control genes to normalise target gene expression but this is heavily dependent on available resources. The selection of appropriate endogenous control genes depends upon the tissue and experimental conditions being studied. For some studies, 18S rRNA may been be appropriate (Schmittgen and Zakrajsek, 2000; Goidin *et al.*, 2001; Dorak, 2003). However, for others, its expression is reported to be variable (Raaijmakers *et al.*, 2002) or be too highly expressed (Frost and Nilsen, 2003). Although a variety of endogenous control genes are available for use in species in which these sequences are known, the paucity of sequence information available for marine organisms, especially the Crustacea, limits the number of candidate genes. Highly conserved sequences across close phylogenetically related species, such as 18S rRNA (Spears *et al.*, 1992) and β-actin therefore make ideal candidates, although it is recognised that better ones may exist. Current opinion on the use of endogenous controls in real-time PCR varies depending on the experimental approach (Bustin, 2002). Although elegant studies have been published supporting individual genes (Schmittgen and Zakrajsek, 2000; Goidin *et al.*, 2001; Donald *et al.*, 2003; Frost and Nilsen, 2003; Radonic *et al.*, 2004), no single gene appears to be appropriate across different studies (Bustin, 2000) or even tissues

within a single species (Towle *et al.*, 1997). However, general opinion appears to favour the use of multiple genes (as reviewed by Pfaffl, 2004) over the use of a single endogenous control. Furthermore, validation methods are being developed to identify the best control genes for a given tissue and experiment (Vandesompele *et al.*, 2002). In the present study, the use of a single endogenous control in most of the experiments, was based purely based on financial constraints of the project.

Cellular responses in these species are fundamental to immunological defence (Rabin, 1970; Smith and Ratcliffe, 1978; Smith and Söderhäll, 1986; Battistella *et al.*, 1996) and thus the cell numbers can be indicative of immunological capability (Hauton, 1995). However, the results of the temperature experiment suggest that THC were significantly greater at 10 °C than at 5 °C, 15 °C or 20 °C. This was probably due to natural sampling inter-animal variability and has no effect on relative gene expression quantification.

The overall variability observed in the present study in THC (10^6 to 10^8 cells per ml) was within the expected range (Smith and Ratcliffe, 1978, 1980b; Truscott and White, 1990; Chisholm and Smith, 1994). This has been suggested to be a typical feature of this species (Smith *et al.*, 2001) reflecting the diversity of both biochemical and immunological status within a population. In the present study, THCs supported the expected cellular responses elicited by the injection of bacteria and saline as described in earlier studies (Smith, 1978; Smith and Ratcliffe, 1980b, 1983; Hauton *et al.*, 1997; Lorenzon *et al.*, 1999). Initially, the recorded THC appeared to follow the patterns described in earlier studies (Smith and Ratcliffe, 1980b, 1983; Hauton *et al.*, 1997; Lorenzon *et al.*, 1999) in the first 3 h post challenge compared to saline placebo injection. However, significant differences due to treatment effects were not observed

in the longer term challenge study (6-84 h). The present study suggests that the recovery of THC, to "normal" levels, occurs between 3 and 6 h, much earlier than the previously reported 24 - 48 h (Smith and Ratcliffe, 1980b, a; Smith *et al.*, 1984; Martin *et al.*, 1993; Hauton *et al.*, 1997; Lorenzon *et al.*, 1999). These results may be influenced by the high inter animal variability in the counts recorded, thereby diluting the legitimate trend towards a later recovery in THC.

The crabs used for each experiment were only selected if they appeared to be outwardly "healthy" and were not missing any appendages. These crude criteria provide no indication of the immunological history of the animals or their ability to mount a comprehensive immunological response. The effect of variable immunological histories within a population could be ameliorated if crabs were reared in captivity under gnotobiotic conditions, although this is very difficult to do for crustaceans (Chisholm, 1993). Several studies have also shown that the haemolymph of closely related crustacean species is not always sterile (Colwell *et al.*, 1975; Sizemore *et al.*, 1975; Tubiash *et al.*, 1975; Brandin and Pistole, 1985; Ueda *et al.*, 1993; Rivera *et al.*, 1999; Tsvetnenko *et al.*, 2001) and this may be the case for *C. maenas*. This may help to explain why Carcinin appears to be highly expressed and that no variation in expression is observed due to the injection of bacterial antigen. Immunological history may also exacerbate inter animal variability and therefore have a considerable impact on the interpretation of results obtained by highly sensitive techniques such as real-time PCR. Although the experimental samples used in real-time PCR could be further standardised by laboratory breeding of experimental animals and positive selection according to a particular biomarker this would not reflect, the situation in the "wild" as reported in the present study.

In conclusion, Carcinin expression is up-regulated within 2 weeks of acclimatisation to 5 °C and perhaps also to 20 °C. Carcinin expression levels are completely unaffected by the presence of bacterial antigen between 0 and 84 h after injection. Under these experimental conditions 18S rRNA is a suitable endogenous control gene and β-actin expression is suspected to be regulated by temperature. In light of the results of the present study, it is hypothesised that Carcinin transcript expression is modulated by acclimatisation to extreme exposure temperatures but is most probably related to non immunological processes associated with these temperatures.

5.6 Conclusions

- Carcinin is confirmed as constitutively expressed in the local population of *C. maenas* in both sexes and all sizes used.

- Carcinin is expressed at extremely high levels in all individuals sampled in the local population.

- 18S rRNA is a good endogenous control gene for haemolymph under these experimental conditions (heat-killed bacterial challenge and 5-15 °C).

- β-actin may be regulated at selected temperatures.

- Inter animal variability of Carcinin expression is quite high and reflects the diversity of immune capabilities in a single population.

- Small changes in expression may be obscured by inter animal variability

- Transcription of Carcinin is up-regulated by 2-4 fold at 5 °C and possibly also at 20 °C.

- Transcription levels are not significantly affected by heat-killed bacterial challenge between 30 min and 48 hours after challenge.

CHAPTER 6

General Discussion

6.1 General Discussion

The overall aims of this thesis were to extend the current knowledge of a small antibacterial protein (Carcinin) from the shore crab *C. maenas,* at both the transcript and protein level, using a combination of *in silico*, *in vitro* and *in vivo* approaches. Firstly, both protein and transcript "tools" were to be developed to quantify changes in the expression of this gene. Secondly, these tools would be used to assess the modulatory effects of environmental and /or physiological parameters on the expression of Carcinin *in vivo*. The inferences derived from this thesis might suggest changes to culture practices and potentially facilitate the natural enhancement of immunological capability, in related commercially important species which are thought to express homologous proteins. Additionally, results from this thesis would substantiate or disprove the use of this gene as a biomarker of crustacean "health".

To investigate the expression of a particular transcript such as Carcinin, a comprehensive picture of the gene was required. Using traditional and novel molecular techniques, the full sequence of the Carcinin transcript, its size, variants and gDNA sequence were determined and subjected to *in silico* analysis using bioinformatic tools (Chapter 2). *In silico* analysis identified a putative signal sequence (21 amino acids), at the N-terminus, followed by a definite cleavage site (TEA-GL). This type of arrangement clearly agrees with the primary structure and suggested processing of several antimicrobial proteins (Dimarcq *et al.*, 1998; Amiche *et al.*, 1999; Destoumieux *et al.*, 2000; Hancock and Diamond, 2000; Ganz, 2003). Despite this, Carcinin does not appear to have the distinct acidic propiece reported for several AMPs (Amiche *et al.*, 1999), including some defensins (Ganz and Lehrer, 1994; Ganz, 2003). However, the Carcinin transcript is composed of a multi-exon coding structure similar to that

observed in other defensins (Hancock and Diamond, 2000; Mitta *et al.*, 2000b). In terms of gene organisation, crustacean AMPs appear to be most similar to the defensins, identified in the scorpion and spider, which are thought to have arisen by exon shuffling (Froy and Gurevitz, 2003). However, contrary to observations of some classical defensins (Ganz, 2003), Carcinin does not appear to exhibit the pattern of separating the putative signal and mature sequences onto separate exons. This may be because in the Crustacea, these proteins are not thought to be translocated (as occurs in insects and vertebrates) but are thought to be released by spontaneous degranulation of the haemocytes (Destoumieux *et al.*, 2000; Munoz *et al.*, 2002). Therefore, the function of the Carcinin N-terminal sequence may be to inhibit the activity of the mature peptide until cleavage by endopeptidases, as described for some AMPs (Valore *et al.*, 1996; Yiallouros *et al.*, 2002; Satchell *et al.*, 2003; Shirafuji *et al.*, 2003). The results of the recombinant expression of this protein (Chapter 3) support this hypothesis.

The classification of AMPs as "defensins" has often been based solely on the cysteine framework (Ganz, 2003), particularly since little amino acid sequence conservation is often evident (Ganz and Lehrer, 1994; Ehret-Sabatier *et al.*, 1996; Andreu and Rivas, 1998; Barra *et al.*, 1998; Dimarcq *et al.*, 1998; Hetru *et al.*, 1998; Bulet *et al.*, 1999). Although Carcinin would appear to exhibit most similarity, in its basic structure and efficacy, with insect and β-defensin AMPs (Dimarcq *et al.*, 1998), it does appear to have a unique cysteine array (Chapter 2). The conservation of this cysteine framework across related crustacean species leads to the hypothesis that these molecules might belong to one or two new groups distinct from other known defensin structures (Powers and Hancock, 2003). The presence and location of the predicted disulphide bonds could be identified in future studies using similar methods to those described by Muramoto and Kamiya (1990) among others (Michaut *et al.*, 1996a). Alternatively, they could be

confirmed by determining the full structure as used in other marine invertebrates, for example the shrimp (*L. vannamei*) and the Mediterranean mussel (*M. galloprovincialis*) (Yang *et al.*, 2000; Yang *et al.*, 2003). The 12-cysteine framework in Carcinin and related crustins, appears to follow a new general consensus pattern (see below) which includes the WAP signature domains (yellow) as defined by Ranganathan *et al.* (1999):

C-x(3)-C-x(8-12)-C-C-x(16-17)–C–x(6)-C-x(9-10)–C–x(5)-C-x(5)-C-C-x(3)–C- x(5)-C

As most of these putative proteins exhibit >25 % sequence identity, it is probable that they share a similar tertiary structure (Westhead *et al.*, 2002). Moreover, unlike some AMPS which exhibit sequence conservation in the N-terminus (Zanetti *et al.*, 1995; Amiche *et al.*, 1999), the C-terminus of these transcripts is more conserved and the N-terminus is highly variable. Based on the evidence presented in this thesis (Chapter 2), it is suggested that Carcinin belongs to the defensins and probably has β-sheets stabilised by disulphide bonds in its tertiary structure.

Relf *et al.* (1999) previously described Carcinin as cysteine-rich, based on partial sequence data but the full translated coding sequence is actually valine-rich with a high representation of cysteines at the C-terminus. However, the valines all reside in the N-terminal sequence so the putative mature fragment is still cysteine-rich. *In silico* analysis (Chapter 2) revealed the putative full coding sequence of the protein to have a molecular mass closer to 12.260 kDa and the putative active fragment (after cleavage at TEA-GL) to be closer to 10.162 kDa. Neither of these values agree with the 11.534 kDa previously determined for the purified protein (Relf *et al.*, 1999). These discrepancies may suggest considerable post-translational modification of the processed active fragment, although the predictive *in silico* analysis, conducted in the present study, could not fully account for the observed difference of 1.372 kDa. Molecular

masses predicted from the possible alternative exon arrangements of this sequence also fail to correspond with the molecular weight of the native protein. The predicted molecular weights of the putative AMPs expressed in other crustaceans such as the lobster, *H. gammarus,* appear to suggest a larger molecule of ~12.2 kDa (Hauton, C. pers.comm) although even larger proteins (~16 kDa) have been predicted from related transcripts in the shrimp, *L. vannamei,* (Gross *et al.*, 2001). These observations, combined with the observed autoproteolysis of the putative active recombinant protein (Chapter 3), indicate that the molecular weight determined by Relf (1999) may have been of a degraded fragment of the full Carcinin coding sequence or that her sample may have been contaminated. To confirm this hypothesis, the determination of the molecular mass of the native protein and for the recombinantly expressed stabilised fragments, should be repeated.

Among the Carcinin transcripts cloned and sequenced, 4 possible isoforms were identified (Chapter 2) (Accession numbers; AJ821886-AJ821889). Each of these four identified sequences, differ at one or several of 4 residues. However, the same variant was expressed in several animals indicating that these variants have not arisen by somatic mutation within individuals and that their frequency may be population related, possibly in response to local bacterial challenge. Certainly, the expression of several variants of the same gene has been confirmed for other crustacean antimicrobial proteins (Destoumieux *et al.*, 1997; Gross *et al.*, 2001; Bartlett *et al.*, 2002; Cuthbertson *et al.*, 2002; Supungul *et al.*, 2002; Vargas-Albores *et al.*, 2004). Cuthbertson *et al.* (2002) have demonstrated that in *L. vannamei,* all identified penaeidin variants were expressed within each individual and this was suggested to be a result of alternate transcriptional mechanisms and not simple allelic polymorphism. The present observations are supported by the divergent nature of antimicrobial transcripts reported

in other crustacean species (Gross *et al.*, 2001; Cuthbertson *et al.*, 2002; Cuthbertson *et al.*, 2004) and vertebrates (Maxwell *et al.*, 2003; Nicolas *et al.*, 2004). To investigate this divergence in the future, a much larger scale sequencing project could be conducted in the future.

To develop a protein tool, namely polyclonal or monoclonal antibodies for use in immunohistochemical analysis, large amounts of purified Carcinin were required as antigen. The choice of recombinant bacterial expression was supported by earlier studies that used this method to successfully express AMPs (Piers *et al.*, 1993; Zhang *et al.*, 1998; Mercado-Pimentel *et al.*, 2002; Skosyrev *et al.*, 2003; Barrell *et al.*, 2004). *In silico* data obtained in Chapter 2, were used to construct a bacterial fusion vector with both the putative mature sequence and the putative precursor sequence of Carcinin (Chapter 3). The choice of a fusion system exploitated the stabilising effect of the GST tag on the Carcinin sequences, as the predicted half-lives of these proteins were relatively short (Chapter 2). Recombinant Carcinin-GST proteins were easily expressed and purified to homogeneity (Chapter 3). The yield of the fusion protein was optimised to ~2.5 μg per ml of culture medium. Further increases in the yield were attempted by the modification of the media, temperature and induction parameters but increases were not observed.

The use of a bacterial expression system in the present study, which does not perform PTM, produced proteins which exhibited no antibacterial activity. This result supports the hypothesis that Carcinin, with up to 6 possible disulphide bonds, probably requires the formation of these to ensure activity (Elsbach, 1990; Yang *et al.*, 2002) as observed for other AMPs (White *et al.*, 1995; Johansson *et al.*, 1998). Future studies, using yeast

and insect expression systems, which do perform PTM, including disulphide bond formation, would confirm this hypothesis.

It would have been possible to use the intact fusion protein expressed in the present study as antigen to obtain polyclonal antibodies to Carcinin. However, this was not initially thought to be ideal as the size and the antigenicity of GST would reduce the titre of the Carcinin specific antibodies and would require lengthy enrichment procedures. Separation of the GST from the recombinant Carcinin proteins (FACR and ACAR) was attempted, using a number of methods (i.e. IEC, HPLC and FPLC) but was not ultimately successful (Chapter 3). This was thought to be due to the denaturation of the GST moiety after cleavage, thereby altering its physico-chemical properties and prohibiting complete separation of the proteins. In addition, the cleavage of the putative active protein (ACAR) resulted in an unequal yield of the two resultant proteins. The considerably lower proportion of "active" Carcinin protein (compared to GST) suggested that spontaneous autoproteolytic degradation of the "active" Carcinin protein had occurred. Similar observations have been made by Skosyrev *et al.* (2003) and Piers *et al.* (1993). Although examples of precursor protein degradation do exist (Yiallouros *et al.*, 2002), the degradation of the cleaved FACR protein in the present study was not observed. This further supports the hypothesis of a protective role for the signal sequence, as suggested for other AMPs by Zhang *et al.* (1998).

The bacterial expression system used in the present study has been successful in expressing proteins which could be used to obtain antibodies, however, the use in the future of an insect or yeast expression system would also facilitate additional characterisation and activity studies on the expressed protein. A yeast system was successfully used by Destoumieux (1999) for the expression of a penaeidin from *L.*

vannamei for this type of investigation. An alternative approach might have been to use synthetic fragments (Hancock and Lehrer, 1998; Cudic *et al.*, 2002) to obtain higher titres of Carcinin specific antibodies, unfortunately the costs associated with this were prohibitive within the scope of a PhD study.

In summary, although recombinant expression of a Carcinin-GST fusion protein is feasible, obtaining purified recombinant Carcinin for use as an antigen is hampered by its propensity for spontaneous proteolytic degradation. These results suggest that the putative signal sequence (Chapter 2) confers increased stability on the active protein thereby inhibiting this degradation. The role of the signal peptide, the rate of autoproteolytic degradation and the significance of post-translational modifications could be further investigated using alternative expression systems and sequence manipulation. Although the development of a Carcinin specific protein tool (i.e. polyclonal antibodies) was not ultimately achieved in the present study, this preliminary expression study has identified one method for successfully producing these proteins and has identified characteristics of the protein which will be crucial in determining the path of future expression studies of this protein.

The design of a transcript tool (Chapter 4), to quantify the relative transcript expression of Carcinin, was successful. The design of these tools relies heavily on the quality and availability of sequence data (Chapter 2). The accurate interpretation of the real-time PCR data obtained, depends on good experimental design, the quality of nucleic acid samples collected and the consistent handling of templates to minimise experimental errors (Poisson's error law) (Pfaffl, 2004). At each stage of sample isolation and processing in the present study, every precaution was taken to minimise the introduction of experimental errors (Chapter 4). Furthermore, the high expression levels observed

may also have reduced the significance of this type of error as suggested by Pfaffl (2004). In spite of the meticulous manner in which the study was designed and conducted, improvements could be made in the future to further reduce inter and intra assay variability.

In expression studies on animals from the wild, it must be noted that the data can be affected by prior immunological experiences of the individuals. Although positive selection of crabs, which meet alternative selected immunological biomarker criteria, could be used to standardise the sample group, the animals would no longer represent a wild population. The best solution would be to considerably increase the sample sizes, to reduce the impact of inter-animal variability. It could be argued that some of the observed inter-animal variability (Chapters 4 and 5) may, in part, reflect the inconsistent quality of the templates amplified. In light of this, a vital consideration is the accurate quantification of the RNA and the pivotal role of the co-transcription of the endogenous control. Also, the inclusion of a standard sample with each reverse transcription reaction is recommended to normalise RT efficiency between runs. The reverse transcription step is considered to be one of the main sources of variability (Pfaffl, 2004), and, in the future, the use of one-step real-time PCR could minimise the influence of variable efficiencies between reverse transcription reactions. Several authors (Bustin, 2002; Pfaffl, 2004) have suggested that the best normalisation practise is to use total cellular RNA due to the problems recognised surrounding the selection of good endogenous controls. In the present study this approach was not considered appropriate as the RNA was quantified using the A_{260} absorbance method. In future studies, using RiboGreen® to quantify total RNA would lead to more precise concentrations and total RNA could then be used to normalise the expression data.

Primers and TaqMan® probes were designed and optimised to quantify the relative transcript expression of Carcinin in the haemolymph of *C. maenas* at different temperatures and with microbial challenge (Chapter 4). As a result of the present study, it was concluded that the optimisation of primers and probes, used in this technique, was not cost effective for small sample sizes (< 1000). However, amplification using the optimised primer and probe sets, resulted in excellent efficiency for the Carcinin set (E = 2) and acceptable amplification efficiencies for both β-actin (E = 1.89) and 18S rRNA (E = 1.74). For a larger study using the same genes, further optimisation, particularly of the 18S rRNA set, could improve efficiencies and the validation of the ΔΔCt method for analysis of the results. However, it must be noted, that the difference between the amplification efficiencies (E) of Carcinin using a single transcript (E= 1.7) and that obtained using the pool of transcripts (E = 2) highlighted the importance of the template quality. Therefore, the E value of the 18S rRNA could be assumed to be higher than that calculated by the REST-XL© software. β-actin is expressed at quite low levels in the haemolymph compared to Carcinin and 18S rRNA. Although Carcinin was expressed at higher levels than β-actin, it was expressed at lower levels 18S rRNA. The difference between the expression levels however was not sufficient so as to invalidate either endogenous control. Importantly, this showed that Carcinin was indeed extremely highly expressed in the haemolymph in all the samples.

To summarise, real-time PCR is a highly sensitive technique where small inconsistencies can exacerbate error and every care must be taken to validate results and to standardise assays wherever possible (Pfaffl, 2004). Sampling a natural population of animals where the immunological history, exact age, nutritional status and environmental exposure history are unknown, the standardisation of samples is more difficult.

The importance of good laboratory practice (GLP) becomes increasingly important when using such a sensitive technique and accurate pipetting is fundamental to reproducibility. In Chapters 4 and 5, the constitutive high level expression of this gene was confirmed in all samples, irrespective of water temperature, sex, colour or size of *C. maenas*. The high expression levels may be related to the suspected rapid turnover of the protein *in vivo,* requiring constant production of the transcript. This would appear to concur with results obtained for *L. vannamei* by Gross *et al.* (2001) where the high representation (17-20 %) of AMP ESTs reflected high expression levels observed *in vivo* by Destoumieux (1997; 2000). The expression level inter-animal variability observed in the present study is thought to be indicative of a genuine diversity of immunological capability within the population sampled (Chapter 4).

In Chapter 5, it was established that Carcinin transcription was up-regulated (2-3 fold) at 5 °C and 20 °C (Chapter 5). Increased temperatures have been reported to stimulate the expression of AMPs in another marine invertebrate (*M. galloprovincialis*) (Mitta *et al.*, 2000a; Hernroth, 2003). Temperature "shock" was not considered as a causal effector in the present study due to the controlled environmental conditions used. However, the 20 °C result must be interpreted cautiously due to the considerable regulation of both non normalised genes calculated by the REST-XL© software. Whilst this may have been due to increased template it may mask the regulation of the endogenous control under those conditions. Therefore, it was concluded that at 20 °C, β-actin may be an unsuitable endogenous control for real-time PCR and the inference that Carcinin is regulated at this temperature must be treated with caution (Chapter 5). This experiment should be repeated to establish if the up-regulation of Carcinin observed at 20 °C is reproducible. The up-regulation of Carcinin transcription observed at 5 °C is corroborated by an unpublished preliminary study by Karen Bell (c. 1989),

cited by Smith and Chisholm (1992), which showed enhanced killing of *P. citreus* by haemocyte extracts from crabs acclimated at 5 °C compared to those at higher temperatures. When viewed in conjunction with the reduced killing ability reported towards Gram-negative bacteria (Chisholm and Smith, 1994), it may be hypothesised that certain temperatures may favour the production of particular AMPs in *C. maenas*. Although the direct induction of Carcinin transcription by low temperatures seems most likely, it is possible that expression will also be modulated indirectly by temperature, due to, for example the onset of the moult and breeding season, cell proliferation and changes due to the tidal cycle. In light of mounting reports on the multifunctionality of these AMPs (Natori, 1990; Bals and Wilson, 2003; Farnaud and Evans, 2003; Kamysz *et al.*, 2003) and the association of Carcinin-like transcripts with cell proliferation (Stoss *et al.*, 2004), future investigations into changes in Carcinin expression may focus on the influences of these other factors.

In the present study, bacterial challenge did not result in any considerable change in expression levels for either 5 °C or 15 °C acclimatised crabs sampled between 0.5 and 48 h post injection with bacterial antigen. These results agree with similar studies in *L.vannamei* where bacterial antigen did not affect penaeidin transcript expression (Munoz *et al.*, 2002). This lack of gene expression inducibility by microbial challenge is contrary to the results obtained for other arthropods such as Drosophila and the soft tick, *Ornithodoros moubata* (Levashina *et al.*, 1998; Engstrom, 1999; Irving *et al.*, 2001; Nakajima *et al.*, 2003). However, this pattern appears to be consistent across the Crustacea. A recent paper on the Carcinin-like transcripts found in *L. vannamei*, suggested that some down-regulation may occur as a result of bacterial antigen injection (Vargas-Albores *et al.*, 2004). In light of the results obtained in the present study, and those reported for the penaeidins in *L. vannamei* by Munoz *et al.* (2002), the

conclusions of Vargas-Albores *et al.* (2004) are not supported and probably arise from differences in THC. In the present study in *C. maenas,* the relationship of THC with temperature or bacterial challenge, concur with earlier observations (Smith and Ratcliffe, 1980b, a; Chisholm and Smith, 1994; Hauton *et al.*, 1997), confirming the activation of the defence response. Despite this, these results support the theory that, in *C. maenas,* AMPs are released when the haemocytes degranulate, as shown in the shrimp (Destoumieux *et al.*, 2000) and that the transcription of these proteins is not induced by microbial challenge. However, since transcription of Carcinin is increased at 5 °C, it is concluded that low temperatures are a modulating factor in the expression of these proteins in this animal.

A recent study (Stoss *et al.*, 2004) has suggested that cell proliferation may be an alternative stimulus for the expression of these types of transcripts; this may be considered to be a wound response. A study in the spiny lobster (*Panulirus argus*), has identified a transcript similar to other crustin / Carcinin sequences with 34 % identity (Stoss *et al.*, 2004). Expression of this transcript was localised to the epithelial cells of the olfactory organ and expression was elevated after cell proliferation was induced by ablation. Defensins in vertebrates have often been localised to epithelial tissues (Zhao *et al.*, 1996; Harder *et al.*, 1997; Nizet *et al.*, 2001; Harder and Schroder, 2002; Zasloff, 2002; Philpott, 2003; Tollin *et al.*, 2003) and this may also represent a further repository of these proteins in the Crustacea. An investigation of the expression in epithelial tissues of *C. maenas* might help to develop this hypothesis further. In more closely related species, such as *Liocarcinus depurator*, there is evidence that proliferating cells are present in the haemolymph (Hammond and Smith, 2002), as previously reported in shrimp (Sequeira *et al.*, 1996), and that their numbers can be increased by stimulation with LPS (lipopolysaccharide). Whilst it may be possible that Carcinin transcript

expression might be similarly induced, no such regulation was observed in the present study in response to *P.citreus* despite cell numbers following the expected patterns. This experiment could be repeated using LPS in place of heat-killed bacteria to further investigate this theory. It must be noted however that the proportion of proliferating cells may be low and 18S rRNA has been shown to be regulated in proliferating cells (Thellin *et al.*, 1999; Raaijmakers *et al.*, 2002). The induction of Carcinin expression by cell proliferation could be investigated in a similar way to the study of Stoss *et al.* (2004), but it would first be essential to establish whether Carcinin is expressed in these epithelial tissues. If haematopoietic tissues identified in crustaceans (Ghiretti-Magaldi *et al.*, 1977; Hose *et al.*, 1992; Johansson *et al.*, 2000; Söderhäll *et al.*, 2003), could be stimulated to increase cell production, by wounding or damage, then it may be interesting to investigate the relationship between the expression of the Carcinin transcript and cell proliferation in the haemolymph.

Alternative functions, such as oxygen transport (Destoumieux-Garzon *et al.*, 2001) and the cleavage of other proteins (Nagai *et al.*, 2001) have been identified for several molecules with antibacterial properties in crustaceans. The multi-functional properties of these proteins are being increasingly recognised for many molecules (Natori, 1990; Kawabata *et al.*, 1996; Destoumieux-Garzon *et al.*, 2001; Nagai *et al.*, 2001; Adachi *et al.*, 2003; Bals and Wilson, 2003; Farnaud and Evans, 2003; Koczulla *et al.*, 2003) and good reviews have been published by Raj and Dentino (2002) and Kamysz *et al* (2003). The high percentage representation of this transcript combined with the results of other studies (Stoss *et al.*, 2004) may point to a more fundamental role for the translated proteins beyond innate immunity (e.g. moulting and repair); this has yet to be demonstrated but could prove to be a very interesting area for future study.

Carcinin-related transcripts are being increasingly identified in commercially important species (Gross *et al.*, 2001; Supungul *et al.*, 2002) and all appear to be expressed at high levels. As major effectors of the immune defence mechanism, their possible modulation by temperature is of great importance to the intensive culture of these species. First, it must be established whether these transcripts indeed result in a protein with antibacterial properties and if their transcript expression can also be modulated by temperature. Some larval stages of commercially important species are reared at high temperatures (e.g. lobster) (Sheeks, 1989) and it would be interesting to establish if a correlation exists between mortalities and high Gram-positive bacterial loads. This may suggest that the increased temperature may lead to selective expression of anti Gram-negative AMPs, resulting in a vulnerability to Gram-positive infection. Developing an understanding of the impact of environmental modulation of immune parameters may improve survival by allowing more precise control over conditions in key developmental stages. As a follow up to the present study in *C. maenas,* it would be useful to determine how quickly a change in temperature affects the expression of these transcripts. The results of the present study suggest that this effect takes place within 2 weeks, although it could be earlier. Repeating the 20 °C experiments would clarify the observations made in the present study at this temperature. In addition, it would be interesting to investigate the possible differential expression of some of the variants which have been identified.

Research into innate immunity has recently enjoyed a resurgence of interest within the scientific community. This has been led in part by suggestions of its evolutionary links to the adaptive immune system (Burnet, 1988; Bayne, 1990; Cooper *et al.*, 1992; Cooper, 1996; Medzhitov and Janeway, 1998; Salzet, 2001; Werling and Jungi, 2003) and the recognition of its common role as the first line of defence for both vertebrates

and invertebrates (Lyndyard *et al.*, 2000). Innate immunity is thought to be fundamental to both the early recognition of invading pathogens (Aderem and Ulevitch, 2000; Medzhitov, 2001; Akira and Hemmi, 2003; Werling and Jungi, 2003; Akira and Takeda, 2004) and the mobilisation of an effective immunological response (Brown, 2001). In addition, the increasing concern with global environmental toxicological issues (Smith, 1991; Snell *et al.*, 2003) and the rise of multi-drug resistant bacteria (Amabile-Cuevas *et al.*, 1995; Cudic *et al.*, 2002; Wilson, 2002; Blanchard, 2003; Cloete, 2003) have both promoted research into innate immunity to identify toxicological "biomarkers" and sources of novel therapeutic compounds (Hancock and Lehrer, 1998; Levy, 2000; Marshall and Arenas, 2003; Ulevitch, 2004). The fact that invertebrates rely solely on an innate defence mechanism has meant that a large proportion of the research has started to focus on these effector molecules.

In the Crustacea, the loss of revenue, resulting from disease, in commercially cultured species, has been sufficient to promote considerable research into understanding their immune systems and the factors modulating its efficacy. It has been recognised that changes in environmental and physiological parameters can have considerable effects on biological function and this includes immunological capability. To complement this, the value of "immunological tools" which enable the rapid assessment of the health "status" (Bachère, 2000) and level of immune capability, have also been recognised (Mialhe *et al.*, 1995; Rodriguez and Le Moullac, 2000). As antimicrobial proteins are considered central to the immunological defence mechanism of these organisms, research into validating their use as possible biomarkers of health has expanded rapidly. Part of this validation process has focussed on identifying factors which initiate and control their expression. This type of research has become increasingly possible due to

the growing accessibility of molecular techniques and bioinformatics tools, although within invertebrate research, their full exploitation is still in its infancy.

Carcinin transcripts do not appear to meet all the criteria required by a biomarker, as described by Smith *et al.* (Smith *et al.*, 2001). Nevertheless, it has already been demonstrated that this protein "relates directly to immune function" due to its antibacterial activity (Relf *et al.*, 1999). In addition, ESTs with identity to the Carcinin transcript have been found in many other crustaceans (Gross *et al.*, 2001; Bartlett *et al.*, 2002; Supungul *et al.*, 2002), these transcripts could be described as "broadly applicable to other related species". The results of the present thesis have shown that this transcript is expressed constitutively in all *C. maenas* individuals sampled, and is easily quantifiable using TaqMan® probes. However, this thesis has also shown, that transcript expression levels are not modulated by the presence of bacterial antigen, although they are modulated by temperature. This suggests, that this transcript may not be a good indicator of immunological capability as its expression does not respond to adverse immunological condition in the haemolymph, even though THCs suggested the immune response mechanism had been activated. Therefore, it is concluded that the Carcinin transcript cannot be used as a biomarker in *C. maenas,* and this may also be applicable to species expressing similar transcripts.

This present study concludes that Carcinin belongs to the "defensin" family of AMPs, and has a unique cysteine array. This array is conserved across ESTs found in other crustaceans and may constitute a new class of defensin structures. The Carcinin protein proabably exist as several isoforms and in its inactivated form, it is thought to be larger than 11.5 kDa. A recombinant fusion protein of the "inactive" protein has been successfully expressed using a bacterial expression system, at levels which could be

used to obtain antibodies for future immunohistochemical analysis. The putative "active" sequence of Carcinin is inherently unstable and rapidly degrades by autoproteolysis; this may explain the very high expression levels observed at the transcript level. The Carcinin transcript is expressed at very high levels in the haemolymph of *C. maenas*. Expression of this transcript is not modulated by bacterial antigen injection but is modulated by temperature. Commercially important animals which express homologous proteins, which are moved to lower temperatures during times of immunological stress, may benefit from increases in transcription of this AMP if it is translated to protein. However, the Carcinin transcript is not considered to be appropriate for use as an immunological biomarker of crustacean "health" as it does not respond to bacterial antigen challenge. Although this protein does have an immunological function it is thought to have additional functions which are not directly immunological and this may provide the most interesting area for future study.

Appendices

Appendix 1 Taxonomy

Arthropoda; Mandibulata; Pancrustacea; Crustacea; Malacostraca; Eumalacostraca; Eucarida; Decapoda; Pleocyemata

Pleocyemata

- **Astacidea** (True lobsters and crayfishes)
  - **Nephropoidea**
    - **Nephropidae** (clawed lobsters)
      - **Homarus**
        - ***Homarus americanus*** (American lobster)
      - **Nephrops**
        - ***Nephrops norvegicus*** (Norway lobster)
- **Brachyura** (Short-tailed crabs)
  - **Eubrachyura**
    - **Heterotremata/Thoracotremata group**
      - **Heterotremata**
        - **Leucosioidea**
          - **Calappidae**(box crabs)
            - **Hepatus**
              - *Hepatus epheliticus* (Calico box crab)
        - **Portunoidea**
          - **Portunidae** (swimming crabs)
            - **Callinectes**
              - *Callinectes sapidus* (Blue crab)
            - **Carcinus**
            - *Carcinus maenas* (Green crab)

Appendix 2 AJ237947

XXX – Purified protein (Relf *et al.*, 1999)

xxx - no cDNA sequence elucidated

XXX – variable codons later described

XXX – C1 oligonucleotide binding

XXX- Carc1primers (Chapter 2)

252 bp and 88 residues

```
0          xxxxxxxxxx
1          G  L  F  P

1          AATAAGGATTGTAAGTACTGGTGCAAAGACAATCTAGGACTAAACTACTGCTGTGGCCAG
5          N  K  D  C  K  Y  W  C  K  D  N  L  G  L  N  Y  C  C  G  Q

61         CCAGGAGTAACCTACCCACCTTTTACTAAAAAGCACTTGGGCAGGTGCCCTGCAGTCCGT
25         P  G  V  T  Y  P  P  F  T  K  K  H  L  G  R  C  P  A  V  R

121        GATACCTGTACTGGCGTCAGGACACAGCTACCAACGTACTGTCCCCATGATGGTGCATGT
45         D  T  C  T  G  V  R  T  Q  L  P  T  Y  C  P  H  D  G  A  C

181        CAGTTCAGAAGCAAGTGCTGCTATGACACCTGCCTGAAGCACCACGTGTGCAAGACCGCC
65         Q  F  R  S  K  C  C  Y  D  T  C  L  K  H  H  V  C  K  T  A

241        GAATACCCTTAT
85         E  Y  P  Y
```

Appendix 3 DIG labelling Kit Components

Supplied Vial	Solution	Components
Vial 1	**Reaction Buffer**	**1 M potassium cacodylate** **125 mM Tris –HCl** **1.25 mg.ml^{-1} bovine serum albumin** **pH 6.6 (25 °C)**
Vial 2	**$CoCl_2$ Solution**	**25 mM cobalt chloride**
Vial 3	**DIG-dUTP**	**1 mM Digoxigenin-11-dUTP**
Vial 4	**dATP**	**10 mM dATP solution in Tris buffer pH 7.5**
Vial 5	**Terminal Transferase**	**50 units µl^{-1} terminal transferase in 20 mM potassium cacodylate** **1 mM EDTA** **200 mM KCl** **0.2 mg ml^{-1} bovine serum albumin,** **50 % (v/v) glycerol; pH 6.5 (25 °C)**
Vial 6	**Unlab. Oligo Control**	**30-mer, 5'-TTG GGT AAC GCC AGG GTT TTC CCA GTC ACG OH-3'**
Vial 7	**Tailed Oligo Control**	
Vial 8	**Control DNA**	
Vial 9	**Glycogen Solution**	
Vial 10	**DNA Dilution Buffer**	
Vial 11	**Poly (A^+)**	

Table 32: Supplied contents of DIG Oligonucleotide tailing kit (Roche Diagnostics GmbH)

Reagents	Test Volume	Control Volume	Final Conc.
Vial 1	4 µl	4 µl	1x
Vial 2	4 µl	4 µl	5 mM
Vial 3	1 µl	1 µl	0.05 mM
C1 Oligo probe or Unlabeled probe (Vial 6)	100 pmol	5 µl (100 pmol)	5 pmol μl^{-1} 5 pmol μl^{-1}
Vial 4	1 µl	1 µl	0.5 mM
Vial 5	1 µl	1 µl	2.5 units μl^{-1}
H_2O	To 20 µl	4 µl	
TOTAL VOL	20 µl	20 µl	

Table 33: Reaction mixtures for DIG tailing of C1 oligonucleotide probe (Roche Diagnostics GmbH)

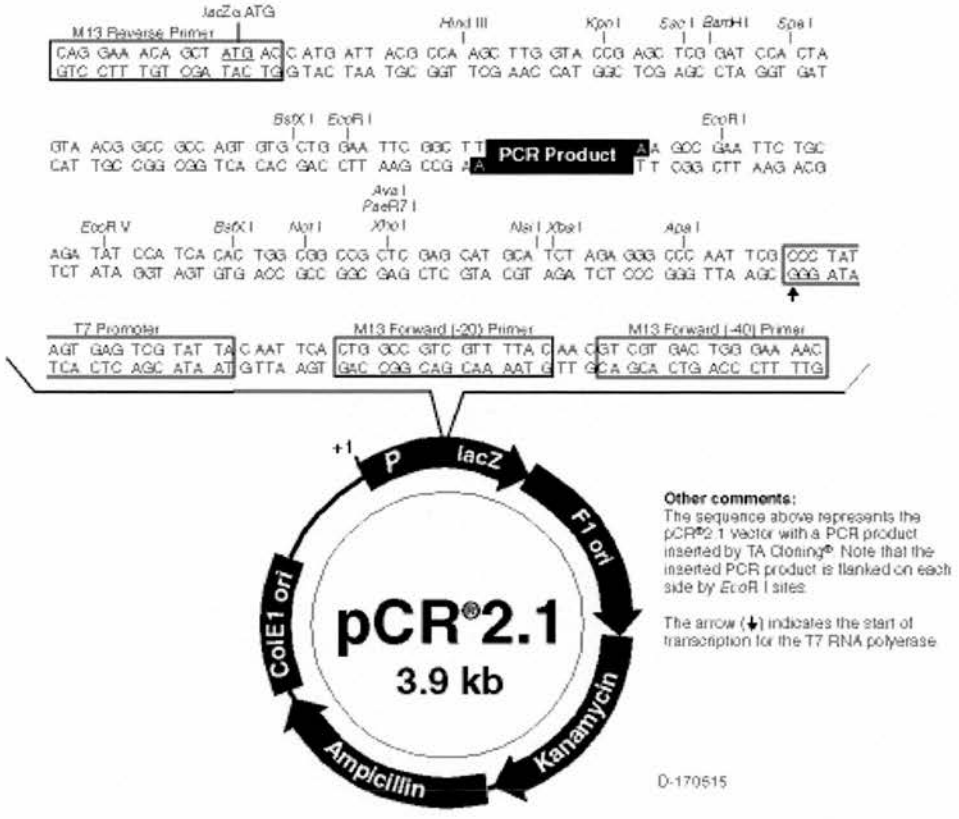
Solutions	Vial/Bottle Number	Description
Maleic Acid Buffer (1 l) (Dilute with H_2O)		**100 mM Maleic acid (11.6 g) 150 mM NaCl (8.8 g) pH 7.5 (10 M NaOH) @ 20 °C**
Washing Buffer (500 ml) (dilute with H_2O)		**Maleic acid buffer (498.5 ml) pH 7.5 @ 20 °C 0.3 % TWEEN 20 (1.5 ml) DO NOT AUTOCLAVE**
Blocking Solution (15 ml)		**1 % Blocking reagent (10 ml) Maleic acid buffer (5 ml) not filtered stable for two weeks @ 4 °C**
Detection Buffer (800 ml)		**100 mM Tris –HCl (24.2 g) 100 mM NaCl (11.6 g) pH 9.5 (HCl) @ 20 °**
DNA Dilution Buffer	**Vial 10 (tailing kit)**	**10 mM Tris-HCl pH 8 @ 20 °C 50µg ml^{-1} DNA from Herring sperm**
NBT & PCIB solution		**75 mg ml^{-1} nitroblue tetraziolium salt in dimethylformamide 50 mg ml^{-1}5 bromo-4-chloro-3-indolyl phosphate toluidinum salt in dimethylformamide**
Anti DIG solution		**1x blocking solution (5 ml) Anti-DIG-AP Fab fragments 1 µl**

Table 34: Reagents required for quantification of probe and northern hybridisation

Appendix 4 Qiagen QIAQuick® Gel Extraction / PCR Purification

Excised DNA obtained by PCR or DNA resulting from enzymatic cleavage, were purified using the "microcentrifuge method" following the manufacturers protocols. In brief, in the case of PCR bands, the excised band was weighed and 3 times the weight, in volume (100mg~100 µl) of binding and solubilisation buffer (QG), was added to the slice in an 0.5 ml centrifuge tube. This mixture was then incubated at 50 °C for ten minutes with intermittent mixing. In the case of enzymatically cleaved DNA, three times the volume of the reaction at are added in QG buffer although this mixture was not incubated as previously described. To the dissolved gel mixture (or cleavage mixture) 1 volume of isopropanol was added and the sample again mixed. The sample was then transferred to a filter tube inside a 2ml collection tube (provided in the kit) which was then sealed and centrifuged at 11,500 x g for 1min at RT. The flowthrough was discarded and a further 0.5 ml of QG buffer was added to the filter tube. After a second centrifugation step as described above 0.75 ml of Buffer PE was added, the solution allowed to stand for 3 min and then centrifuged again as before. The flowthrough was discarded and the tube centrifuged again using the same conditions. The filter tube was then placed into a clean 1.5 ml centrifuge tube and 30 µl of deionised water added and the column allowed to stand for a further five minutes before centrifugation at 13,000 x g for 1 min.

Appendix 5 TOPO TA 2.1 Cloning Vector



Appendix 6 Promega Wizard® Plus SV Miniprep DNA Purification System

A single colony containing the plasmid of interest was isolated and grown in 5-10 ml LBAG at 37 °C for 16 h in a shaking water bath to exponential growth phase (A_{600} ~2-4. The culture was centrifuged (10,000 x g, 5 min @ 5 °C) and supernatant discarded. The cell pellet was resuspended in 250 µl of Resuspension Solution and the suspension transferred to 1.5 ml centrifuge tube. Two hundred and fifty microlitres of Cell Lysis Solution were added to each tube and mixed gently by inversion (four times, ~5 min). Alkaline protease solution (10 µl) was added to each carefully opened tube, mixed gently by inversion and incubated at room temperature for 5 min. Each tube was opened carefully, to avoid nicking the DNA, 350 µl of Neutralisation Solution was added and mixed by inversion (four times). The tubes were then centrifuged at 14,000 x g in a cooled (5 °C) tabletop centrifuge for 10 min and the resultant supernatant decanted into a spin column in a 1.5 ml centrifuge tube. The spin column was sealed and centrifuged at 14,000 x g at RT for 1 min. The flow through was discarded and 750 µl of Wash Solution was added to the top of the spin column and centrifuged at 14,000 x g for 1 min at RT and repeated once using 250 µl. The spin column was transferred to a clean centrifuge tube and 80-100 µl of autoclaved MilliQ water was added to the top of each tube and centrifuged at 14,000 x g for 1 min at RT. The elutant was then quantified by spectrophotometry and purity assessed by gel electrophoresis as described in previous chapters.

Appendix 7 AJ427538

XXX - Coding Sequence

XXX - 5' UTR

XXX - 3'UTR

XXX – Putative precursor sequence cloned into fusion vector FCAR (333bp)

XXX – Putative active sequence cloned into fusion vector ACAR (270bp)

XXX – Variable codons

XXX – Protein sequence purified (Relf *et al.*, 1999)

XXXX – Carc1 and Carc2 primers FWD and REV

ACAGAGGCAAGGTTA (TEA ↓ GL) - putative signal cleavage site

Appendix 8 Carcinin gDNA Sequence

(Exons in bold and underlined)

```
1       AGACCAATCA GCTTCTTCAA GAACACATTG AAACATGGTA AGTTTATATT
51      ATGCTTCCTA AAAAAAAAAT AATAAATAAA AATAAAACAT GGTTGGATGA
101     AACTGCATTA ATACCACTTT TTCTGTTTCT TCAAACTTTT ATAATCAAAT
151     ATTTATTGTA AAGAGCTAAT TTCAATAATT CATGAACAGA AGGTGCAAAC
201     TGTAGCAGCC GTGGTGGTTG TGGCTGTGGT TGTGACCATG ACCGAGGCAG
251     GGTTATTCCC TCCGAAGGAC TGTAAGTACT GGTGCAAAGA CAATCTTGGA
301     ATAAACTACT GCTGTGGCCA GCCAGGAGTA ACCTACCCAC CTTTTACTAA
351     AAGTAAGTAT CCCTCATCTC GTTACCTTCA TCTTACTTTG CAATATGGCA
401     GTTTTAAAAT GCAACCTGTC ATTATTGATT CTTCTGTTAA TGCATTAATC
451     TTCTTAAAAT GTTGTATACG TGTTTATTGA TGAATAGATC AACCATAATT
501     TCTATGGTTC TCATATATTC ACACAATATT GTCATAAATT TCAGATCACT
551     TGGGCAGGTG CCCTCCAGTC CGTGATACCT GTACTGGCGT CAGGACACAG
601     CTACCAACGG TGAGTGAGAG AGAGAGAGAG TGAAGAAATT ATTTTTTCAT
651     GTCTTTATAA GAAAACAATT ACTTAAATAA ATTTCTTTCT TTTACAGTAC
701     TGTCCCCATG ATGGTGCATG TCAGCTCAGA AGCAAGTGCT GCTATGACAC
751     CTGCCTGAAG CACCACGTGT GCAAGACTGC CGAATATCCT TATTATTAGA
801     CATCGCAGAC CCGTGTAAGA AATCTTCCAC CTAGTACATA GATCAGATCT
851     GAAATAAGAA ACTCTGTAAT CTACGGAAAT TCTACAAACA CTATGACGCA
901     TGGTTACCTA CTGTACTGAA GGGCGAATTC CAGCACACTG GCGGCCGTTA
951     ACTAGTGGAT CACGAGCTCG
```

Appendix 9 Carcinin Protein Residue Assignment

X protein sequence

X changed from protein sequence to agree with cDNA (Relf *et al.* 1999)

X inferred from cDNA sequencing

```
0         xxxxxxxxxx
1         G  L  F  P

1       AATAAGGATTGTAAGTACTGGTGCAAAGACAATCTAGGACTAAACTACTGCTGTGGCCAG
5         N  K  D  C  K  Y  W  C  K  D  N  L  G  L  N  Y  C  C  G  Q

61      CCAGGAGTAACCTACCCACCTTTTACTAAAAAGCACTTGGGCAGGTGCCCTGCAGTCCGT
25        P  G  V  T  Y  P  P  F  T  K  K  H  L  G  R  C  P  A  V  R

121     GATACCTGTACTGGCGTCAGGACACAGCTACCAACGTACTGTCCCCATGATGGTGCATGT
45        D  T  C  T  G  V  R  T  Q  L  P  T  Y  C  P  H  D  G  A  C

181     CAGTTCAGAAGCAAGTGCTGCTATGACACCTGCCTGAAGCACCACGTGTGCAAGACCGCC
65        Q  F  R  S  K  C  C  Y  D  T  C  L  K  H  H  V  C  K  T  A

241     GAATACCCTTAT
85        E  Y  P  Y
```

Appendix 10 Carcinin Isoforms

AJ237947

MKVQTVAAVVVVAVVVAMTEAGLFPPKDCKYWCKDNLGLNYCCGQPGVTYPPFTKKHLGRC
PAVRDTCTGVRTQLPTYCPHDGACQFRSKCCYDTCLKHHVCKTAEYPYY

AJ427538

MKVQTVAAVVVVAVVVAMTEAGLFPPKDCKYWCKDNLGINYCCGQPGVTYPPFTKSHLGRCP
PVRDTCTGVRTQLPTYCPHDGACQFRSKCCYDTCLKHHVCKTAEYPYY

Isoform 1

MKVQTVAAVVVVAVVVAMTEAGLFPPKDCKYWCKDNLGVNYCCGQPGVTYPPFTKKHLGRC
PAVRDTCTGVRTQLPTYCPHDGACQFRSKCCYDTCLKHHVCKTAEYPYY

Isoform 2

MKVQTVAAVVVVAVVVAMTEAGLFPPKDCKYWCKDNLGINYCCGQPGVTYPPFTKKHLGRCP
AVRDTCTGVRTQLPTYCPHDGACQFRSKCCYDTCLKHHVCKTAEYPYY

Isoform 3

MKVQTVAAVVVVAVVVAMTEAGLFPPKDCKYWCKDNLGINYCCGQPGVTYPPFTKKHLGRCP
PVRDTCTGVRTQLPTYCPHDGACQFRSKCCYDTCLKHHVCKTAEYPYY

Isoform 4

MKVQTVAAVVVVAVVVAMTEAGLFPPKDCKYWCKDNLGINYCCGQPGVTYPPFTKNHLGRCP
PVRDTCTGVRTQLPTYCPHDGACQFRSKCCYDTCLKHHVCKTAEYPYY

Translated gDNA 1

MKVQTVAAVVVVAVVVAMTEAGLFPPKDCKYWCKDNLGINYCCGQPGVTYPPFTKSHLGRCP
AVRDTCTGVRTQLPTYCPHDGACQFRSKCCYDTCLKHHVCKTAEYPYY

Translated gDNA 2

MKVQTVAAVVVVAVVVAMTEAGLFPPKDCKYWCKDNLGINYCCGQPGVTYP

PFTKSHLGRCPPVRDTCTGVRTQLPTYCPHDGACQLRSKCCYDTCLKHHVCKT

AEYPYY

Appendix 11 Consensus Carcinin Protein Sequence

1 MKVQTVAAVV VVAVVVAMTE AGLFPPKDCK YWCKDNLGIN

41 YCCGQPGVTY PPFTKKHLGR CPAVRDTCTG VRTQLPTYCP

81 HDGACQFRSK CCYDTCLKHH VCKTAEYPYY

XX Hydrophobic Signal sequence

XX Cleavage site (TEA –GLF)

Appendix 12 Sequenced Purified Protein Fragments of Carcinin

(Relf *et al.*, 1999)

GLFPXKDCKYwCd

GLFPXKDXKYWCKDNXXInYCC

TQLPTYQPHDGAQQFR

TAEYPY

Appendix 13 Bacterial Genotypes

E.coli JM105	*thi*, *rps*L(Str^r), *end*A, *sbc*B15, Δ(*lac-pro*AB), [F', *tra*D36,*pro*AB+, *lac*I^{q}ZΔM15], *hsd*R4(r_K-, m_K+)
E.coli BL21	F$^-$, *omp*T, *hsd*S (r_B^-,m_B^-), *gal*, *dcm*

Appendix 14 Solutions

LB Broth

LB broth	20g

Dissolve ingredients in 900ml distilled water, pH to 7.2 and adjust volume to 1litre. Sterilise by autoclaving @ 121°C for 20 min.

LB Agar

LB agar	30g

Dissolve ingredients in 900ml distilled water and adjust volume to 1litre. Sterilise by autoclaving @ 121°C for 20 min and cool to 55°C before pouring into Petri dishes and storing at 4°C when cooled.

Transformation and Storage Solution (TSS) Buffer (100ml)

Tryptone	1.0g
Yeast Extract	0.5g
Sodium Chloride	0.5g
Polyethylene glycol (Mw 8000)	2.39g
Dimethylsulphoxide (DMSO)	5.0ml
Magnesium Chloride ($MgCl_2$)	5.0ml
MilliQ® water	70ml

Mix components one by one until dissolved in baked glassware and adjust pH to 6.5 with HCl or NaOH. Adjust volume to 100ml with MilliQ® water. Filter sterilise through a 0.2μm Whatman filter into baked glassware. Store @ 4°C for up to 6 months.

LBG broth and agar

LB broth and agar as described above plus 10ml of 2μm filtered (sterile) 2M Glucose (BDH) (final conc. 20mM) added after autoclaving and cooled to 55°C

LBAG broth and agar

LBG broth and agar as described above plus 100μg ml-1 ampicillin added to post autoclaved cooled mixture.

0.5X TBE Buffer

45 mM Tris-borate; 1 mM EDTA; pH 8.0

6X DNA loading buffer

100 mM EDTA ; 25 mM Tris-HCl; 25 % glycerol; 0.05 % bromophenol blue ; pH 7.0

2 x YTA and 2 x YTAG media

Tryptone	16g
Yeast Extract	10g
Sodium Chloride	5g
ampicillin	100μg ml-1

Dissolve tryptone, yeast extract and sodium chloride in 900ml distilled water and adjust pH to 7.0 with sodium hydroxide. Adjust volume to 1l, sterilise at 121°C for 15min and cool to 50°C before adding 1ml of stock ampicillin 100 mg ml^{-1} to a final concentration of 10μg ml^{-1}.
For 2 x YTAG, prepare medium as above and after autoclaving and cooling to 50°C add 100ml of sterile 20 % glucose solution (2 % final concentration).

Glutathione Elution Buffer 1 a (10mM) and b(20mM)

10mM-20mM reduced glutathione
50mM Tris-HCl

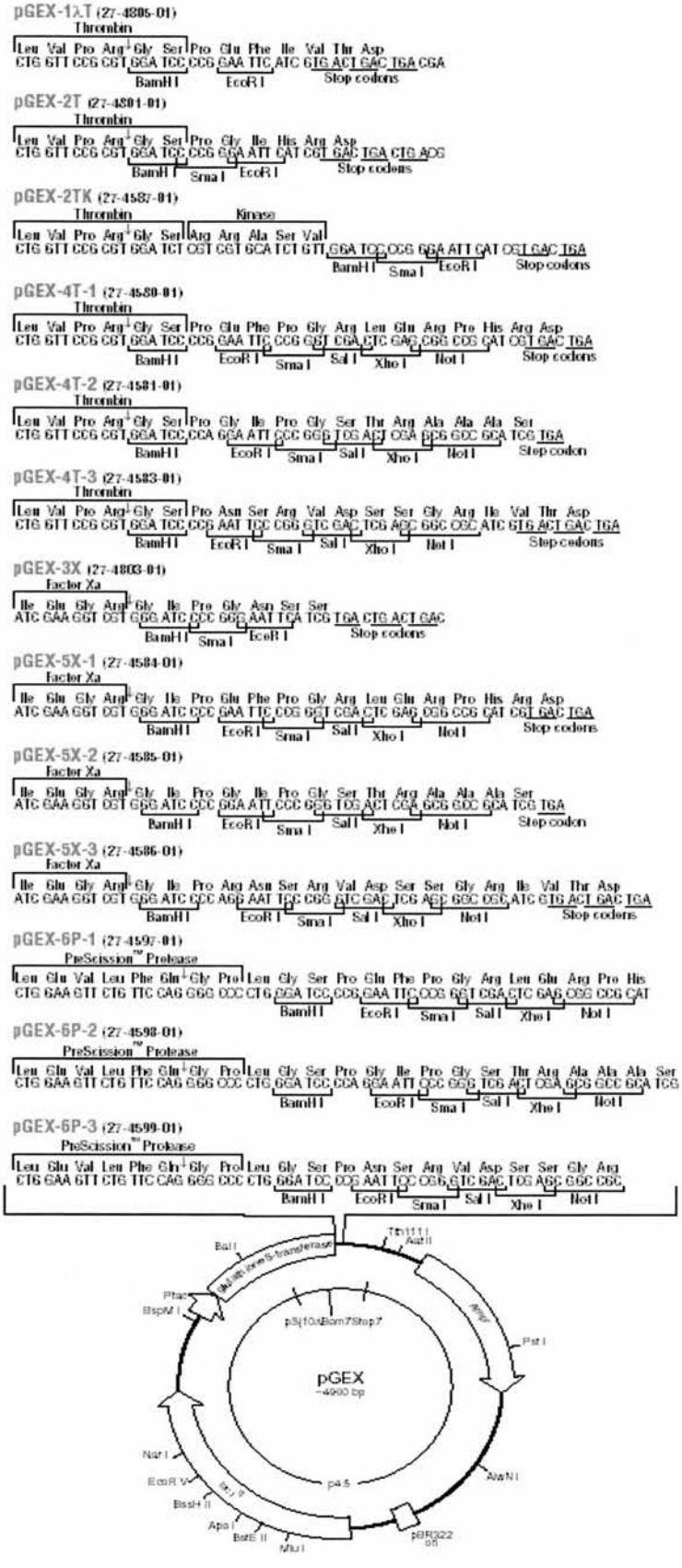
Prepare a 50mM Tris-HCl solution with distilled MilliQ water and add 10-20mM reduced glutathione and aliquot into 1ml aliquots in 1.5ml centrifuge tubes and store at -20°C until required.

Glutathione Elution Buffer 2

75mM HEPES (pH 7.4)
150mM NaCl
10mM reduced glutathione
5mM DTT
2 % N-octyl glucoside

Prepare a 150mM NaCl solution with distilled MilliQ water and add and aliquot into 1ml aliquots in 1.5ml centrifuge tubes and store at -20°C until required.

Appendix 15 pGEX vectors Multiple Cloning Region



Appendix 16 pGEX-4T-1 Restriction sites

Restriction sites for pGEX-4T-1

The following restriction site table was compiled using DNASIS software for sequence analysis. The enzymes chosen are those which we believe to have been commercially available in June 1992.

The vector pGEX-4T-1 was not tested with each enzyme and therefore the accuracy of the table cannot be guaranteed. Please contact your local Pharmacia subsidiary if a discrepancy is identified.

Please note that enzymes with non-palindromic recognition sequences are listed twice. You must combine both entries to obtain the total number of sites for these enzymes.

Locations represent the 5' end of the recognition sequence for each restriction enzyme.

Enzyme (# sites): Locations

Aat II (1): 1241
Acc I (1): 949
*Acc*65 I (0):
Act I (20): 924, 963, 1320, 1649, 1816, 2460, 2604, 3100, 3121, 3128, 3227, 3380, 3467, 3696, 4167, 4223, 4324, 4334, 4651, 4856
Act I (39): 225, 549, 959, 1065, 1104, 1114, 1156, 1273, 1419, 1528, 1605, 1770, 2007, 2098, 2469, 2714, 2835, 2854, 2981, 3009, 3171, 3188, 3214, 3268, 3419, 3435, 3444, 3486, 3558, 3622, 3642, 3645, 3873, 3957, 4113, 4251, 4390, 4493, 4727
Aff II (0):
Aff III (3): 496, 3052, 3668
Age I (0):
Aha II (5): 34, 1241, 1623, 3624, 4307
Alu I (25): 54, 178, 482, 631, 809, 1033, 1052, 1810, 1873, 1973, 2494, 2751, 2797, 2887, 3113, 3448, 3728, 4152, 4267, 4360, 4424, 4519, 4628, 4739, 4919
Alw I (7): 914, 930, 1517, 1838, 2302, 2400, 2486
Alw I (4): 926, 1530, 2310, 2407
*Alw*N I (1): 2638
Apa I (1): 3875
*Apa*L I (4): 18, 1492, 2738, 3648
Apo I (3): 939, 3233, 3943
Asc I (0):
Ase I (4): 189, 1988, 4352, 4411
Asu I (13): 290, 752, 1184, 1800, 2022, 2039, 2118, 3530, 3875, 3876, 4220, 4244, 4648
Asu II (1): 654
Ava I (2): 944, 954
Ava II (4): 752, 1800, 2022, 4220
Ava III (0):
Avr II (0):
Bal I (1): 463
*Bam*H I (1): 930
Ban I (6): 2211, 3588, 4307, 4437, 4717, 4729
Ban II (1): 3875
*Bbr*P I (0):
Bbs I (3): 1174, 3807, 4146
Bbs I (0):
Bbv I (11): 42, 1031, 1128, 1920, 2109, 2711, 3130, 3148, 4358, 4607, 4680
Bbv I (7): 1731, 2425, 2631, 2634, 3717, 4088, 4217
Bcg I (1): 3971
Bcg I (2): 299, 1637
Bcl I (2): 692, 3682
Bfa I (4): 270, 1971, 2306, 2559
Bfr I (0):
Bgl I (2): 2040, 4683
Bgl II (0):
Bpm I (3): 2074, 3974, 4745
Bpm I (1): 3521
*Bpu*1102 I (0):
Bsa I (1): 2092
Bsa I (0):
*Bsa*A I (1): 1144
*Bsa*B I (0):
*Bsa*H I (5): 34, 1241, 1623, 3624, 4307
*Bsa*J I (7): 934, 944, 2892, 4303, 4442, 4579, 4869
*Bse*A I (0):
Bsg I (2): 69, 523
Bsg I (3): 11, 3534, 3734
*Bsi*E I (6): 960, 1643, 1792, 2715, 3139, 4657
*Bst*W I (0):
*Bsf*Y I (11): 264, 861, 1037, 2568, 2847, 3013, 3031, 3244, 4090, 4334, 4785
Bsl I (11): 264, 861, 1037, 2568, 2847, 3013, 3031, 3244, 4090, 4334, 4785
Bsm I (0):
Bsm I (0):
*Bsm*A I (5): 2093, 3360, 3765, 3891, 4278
*Bsm*A I (2): 1036, 1323
*Bsp*1286 I (6): 18, 1492, 1577, 2738, 3648, 3875
*Bsp*D I (0):
*Bsp*E I (0):
*Bsp*H I (3): 1219, 1324, 2332
*Bsp*M I (0):
*Bsp*M I (1): 63
*Bsp*M II (0):
*Bsp*W I (26): 102, 866, 1920, 2040, 2428, 3000, 3114, 3179, 3345, 3435, 3531, 3633, 3720, 3857, 3864, 4065, 4072, 4074, 4216, 4310, 4394, 4438, 4653, 4683, 4692, 4718
Bsr I (14): 1514, 1953, 1996, 2114, 2520, 3482, 3584, 3821, 3973, 4283, 4382, 4543, 4569, 4791
Bsr I (6): 1138, 1688, 2636, 2649, 3323, 3522
*Bst*F I (2): 2079, 3354
*Bss*H II (1): 4079
*Bst*1107 I (0):
*Bst*B I (1): 654
*Bst*E II (1): 3849
*Bst*N I (11): 764, 2892, 2905, 3026, 3392, 3707, 4247, 4304, 4443, 4580, 4707
*Bst*U I (27): 360, 925, 992, 994, 1097, 1272, 1604, 2097, 2427, 3008, 3267, 3381, 3418, 3468, 3539, 3557, 3573, 3662, 3669, 3817, 3863, 3898, 4080, 4104, 4335, 4337, 4941
*Bst*X I (3): 3463, 3592, 3715
*Bst*Y I (8): 914, 930, 1517, 1534, 2302, 2314, 2400, 2411
*Bsu*36 I (1): 4760
*Cfr*10 I (2): 2079, 3354
Cla I (0):
Dde I (8): 624, 1237, 1663, 2203, 2369, 2778, 4240, 4761
Dpn I (23): 660, 693, 915, 931, 983, 1482, 1518, 1535, 1793, 1839, 1857, 2198, 2303, 2315, 2393, 2401, 2412, 2487, 3579, 3683, 4056, 4658, 4754
Dpn II (23): 660, 693, 915, 931, 983, 1482, 1518, 1535, 1793, 1839, 1857, 2198, 2303, 2315, 2393, 2401, 2412, 2487, 3579, 3683, 4056, 4658, 4754
Dra I (5): 683, 792, 1584, 2276, 2295
Dra II (2): 289, 1183
Dra III (0):
Drd I (2): 1081, 2944
Dsa I (1): 4869
Dsa V (21): 764, 935, 944, 945, 1037, 1072, 1627, 1978, 2674, 2892, 2905, 3026, 3287, 3392, 3707, 4096, 4247, 4304, 4443, 4580, 4707
Eae I (5): 463, 960, 1771, 4342, 4545
Eag I (1): 960
*Eam*1105 I (1): 2159
Ear I (0):
Ear I (5): 338, 1366, 3170, 3287, 4639
*Ech*136 II (0):
*Eco*47 III (0):
*Eco*57 I (1): 1477
*Eco*57 I (1): 2525
*Eco*N I (1): 264
*Eco*O109 I (2): 289, 1183
*Eco*R I (1): 939
*Eco*R II (11): 764, 2892, 2905, 3026, 3392, 3707, 4247, 4304, 4443,

27-4580-01

Appendix 17 Predicted Sequences of FCARR, FCARD ACARR and ACARD

PCR product incorporating full coding sequence including restriction sites and GC clamp (FCARR) (372bp)

GCGCG↓GATCCATG – 327 bp -

TAGACATCGCAGACCCGTGTAACTCGA↓GGCGC

Digested PCR product incorporating full coding sequence including restriction sites and GC clamp (FCARR) (372bp)

FCARRD (Digested PCR Product - 362bp)

↓GATCCATG – 327 bp - TAGACATCGCAGACCCGTGTAACTCGA↓

ACARR (PCR Product - 309bp)

GCGCG↓GATCCAGG – 264 bp -

TAGACATCGCAGACCCGTGTAACTCGA↓GGCGC

ACARRD (Digested PCR Product - 299bp)

↓GATCCAGG – 264bp – TAGACATCGCAGACCCGTGTAACTCGA↓

Appendix 18 pGEX4T-1 sequence

```
LOCUS  pGEX-4T-1  4969 bp
BASE COUNT 1225 A   1202 C   1292 G   1250 T
ORIGIN
    1 ACGTTATCGA CTGCACGGTG CACCAATGCT TCTGGCGTCA GGCAGCCATC GGAAGCTGTG
   61 GTATGGCTGT GCAGGTCGTA AATCACTGCA TAATTCGTGT CGCTCAAGGC GCACTCCCGT
  121 TCTGGATAAT GTTTTTTGCG CCGACATCAT AACGGTTCTG GCAAATATTC TGAAATGAGC
  181 TGTTGACAAT TAATCATCGG CTCGTATAAT GTGTGGAATT GTGAGCGGAT AACAATTTCA
  241 CACAGGAAAC AGTATTCATG TCCCCTATAC TAGGTTATTG GAAAATTAAG GGCCTTGTGC
  301 AACCCACTCG ACTTCTTTTG GAATATCTTG AAGAAAAATA TGAAGAGCAT TTGTATGAGC
  361 GCGATGAAGG TGATAAATGG CGAAACAAAA AGTTTGAATT GGGTTTGGAG TTTCCCAATC
  421 TTCCTTATTA TATTGATGGT GATGTTAAAT TAACACAGTC TATGGCCATC ATACGTTATA
  481 TAGCTGACAA GCACAACATG TTGGGTGGTT GTCCAAAAGA GCGTGCAGAG ATTTCAATGC
  541 TTGAAGGAGC GGTTTTGGAT ATTAGATACG GTGTTTCGAG AATTGCATAT AGTAAAGACT
  601 TTGAAACTCT CAAAGTTGAT TTTCTTAGCA AGCTACCTGA AATGCTGAAA ATGTTCGAAG
  661 ATCGTTTATG TCATAAAACA TATTTAAATG GTGATCATGT AACCCATCCT GACTTCATGT
  721 TGTATGACGC TCTTGATGTT GTTTTATACA TGGACCCAAT GTGCCTGGAT GCGTTCCCAA
  781 AATTAGTTTG TTTTAAAAAA CGTATTGAAG CTATCCCACA AATTGATAAG TACTTGAAAT
  841 CCAGCAAGTA TATAGCATGG CCTTTGCAGG GCTGGCAAGC CACGTTTGGT GGTGGCGACC
  901 ATCCTCCAAA ATCGGATCTG GTTCCGCGTG GATCCCCGGA ATTCCCGGGT CGACTCGAGC
  961 GGCCGCATCG TGACTGACTG ACGATCTGCC TCGCGCGTTT CGGTGATGAC GGTGAAAACC
1021 TCTGACACAT GCAGCTCCCG GAGACGGTCA CAGCTTGTCT GTAAGCGGAT GCCGGGAGCA
1081 GACAAGCCCG TCAGGGCGCG TCAGCGGGTG TTGGCGGGTG TCGGGGCGCA GCCATGACCC
1141 AGTCACGTAG CGATAGCGGA GTGTATAATT CTTGAAGACG AAAGGGCCTC GTGATACGCC
1201 TATTTTTATA GGTTAATGTC ATGATAATAA TGGTTTCTTA GACGTCAGGT GGCACTTTTC
1261 GGGGAAATGT GCGCGGAACC CCTATTTGTT TATTTTTCTA AATACATTCA AATATGTATC
1321 CGCTCATGAG ACAATAACCC TGATAAATGC TTCAATAATA TTGAAAAAGG AAGAGTATGA
1381 GTATTCAACA TTTCCGTGTC GCCCTTATTC CCTTTTTTGC GGCATTTTGC CTTCCTGTTT
1441 TTGCTCACCC AGAAACGCTG GTGAAAGTAA AAGATGCTGA AGATCAGTTG GGTGCACGAG
1501 TGGGTTACAT CGAACTGGAT CTCAACAGCG GTAAGATCCT TGAGAGTTTT CGCCCCGAAG
1561 AACGTTTTCC AATGATGAGC ACTTTTAAAG TTCTGCTATG TGGCGCGGTA TTATCCCGTG
1621 TTGACGCCGG GCAAGAGCAA CTCGGTCGCC GCATACACTA TTCTCAGAAT GACTTGGTTG
1681 AGTACTCACC AGTCACAGAA AAGCATCTTA CGGATGGCAT GACAGTAAGA GAATTATGCA
1741 GTGCTGCCAT AACCATGAGT GATAACACTG CGGCCAACTT ACTTCTGACA ACGATCGGAG
1801 GACCGAAGGA GCTAACCGCT TTTTTGCACA ACATGGGGGA TCATGTAACT CGCCTTGATC
1861 GTTGGGAACC GGAGCTGAAT GAAGCCATAC CAAACGACGA GCGTGACACC ACGATGCCTG
1921 CAGCAATGGC AACAACGTTG CGCAAACTAT TAACTGGCGA ACTACTTACT CTAGCTTCCC
1981 GGCAACAATT AATAGACTGG ATGGAGGCGG ATAAAGTTGC AGGACCACTT CTGCGCTCGG
2041 CCCTTCCGGC TGGCTGGTTT ATTGCTGATA AATCTGGAGC CGGTGAGCGT GGGTCTCGCG
2101 GTATCATTGC AGCACTGGGG CCAGATGGTA AGCCCTCCCG TATCGTAGTT ATCTACACGA
2161 CGGGGAGTCA GGCAACTATG GATGAACGAA ATAGACAGAT CGCTGAGATA GGTGCCTCAC
2221 TGATTAAGCA TTGGTAACTG TCAGACCAAG TTTACTCATA TATACTTTAG ATTGATTTAA
2281 AACTTCATTT TTAATTTAAA AGGATCTAGG TGAAGATCCT TTTTGATAAT CTCATGACCA
2341 AAATCCCTTA ACGTGAGTTT TCGTTCCACT GAGCGTCAGA CCCCGTAGAA AAGATCAAAG
2401 GATCTTCTTG AGATCCTTTT TTTCTGCGCG TAATCTGCTG CTTGCAAACA AAAAAACCAC
2461 CGCTACCAGC GGTGGTTTGT TTGCCGGATC AAGAGCTACC AACTCTTTTT CCGAAGGTAA
2521 CTGGCTTCAG CAGAGCGCAG ATACCAAATA CTGTCCTTCT AGTGTAGCCG TAGTTAGGCC
2581 ACCACTTCAA GAACTCTGTA GCACCGCCTA CATACCTCGC TCTGCTAATC CTGTTACCAG
2641 TGGCTGCTGC CAGTGGCGAT AAGTCGTGTC TTACCGGGTT GGACTCAAGA CGATAGTTAC
2701 CGGATAAGGC GCAGCGGTCG GGCTGAACGG GGGGTTCGTG CACACAGCCC AGCTTGGAGC
2761 GAACGACCTA CACCGAACTG AGATACCTAC AGCGTGAGCT ATGAGAAAGC GCCACGCTTC
2821 CCGAAGGGAG AAAGGCGGAC AGGTATCCGG TAAGCGGCAG GGTCGGAACA GGAGAGCGCA
2881 CGAGGGAGCT TCCAGGGGGA AACGCCTGGT ATCTTTATAG TCCTGTCGGG TTTCGCCACC
2941 TCTGACTTGA GCGTCGATTT TTGTGATGCT CGTCAGGGGG GCGGAGCCTA TGGAAAAACG
3001 CCAGCAACGC GGCCTTTTTA CGGTTCCTGG CCTTTTGCTG GCCTTTTGCT CACATGTTCT
3061 TTCCTGCGTT ATCCCCTGAT TCTGTGGATA ACCGTATTAC CGCCTTTGAG TGAGCTGATA
3121 CCGCTCGCCG CAGCCGAACG ACCGAGCGCA GCGAGTCAGT GAGCGAGGAA GCGGAAGAGC
3181 GCCTGATGCG GTATTTTCTC CTTACGCATC TGTGCGGTAT TTCACACCGC ATAAATTCCG
3241 ACACCATCGA ATGGTGCAAA ACCTTTCGCG GTATGGCATG ATAGCGCCCG GAAGAGAGTC
3301 AATTCAGGGT GGTGAATGTG AAACCAGTAA CGTTATACGA TGTCGCAGAG TATGCCGGTG
3361 TCTCTTATCA GACCGTTTCC CGCGTGGTGA ACCAGGCCAG CCACGTTTCT GCGAAAACGC
```

```
3421 GGGAAAAAGT GGAAGCGGCG ATGGCGGAGC TGAATTACAT TCCCAACCGC GTGGCACAAC
3481 AACTGGCGGG CAAACAGTCG TTGCTGATTG GCGTTGCCAC CTCCAGTCTG GCCCTGCACG
3541 CGCCGTCGCA AATTGTCGCG GCGATTAAAT CTCGCGCCGA TCAACTGGGT GCCAGCGTGG
3601 TGGTGTCGAT GGTAGAACGA AGCGGCGTCG AAGCCTGTAA AGCGGCGGTG CACAATCTTC
3661 TCGCGCAACG CGTCAGTGGG CTGATCATTA ACTATCCGCT GGATGACCAG GATGCCATTG
3721 CTGTGGAAGC TGCCTGCACT AATGTTCCGG CGTTATTTCT TGATGTCTCT GACCAGACAC
3781 CCATCAACAG TATTATTTTC TCCCATGAAG ACGGTACGCG ACTGGGCGTG GAGCATCTGG
3841 TCGCATTGGG TCACCAGCAA ATCGCGCTGT TAGCGGGCCC ATTAAGTTCT GTCTCGGCGC
3901 GTCTGCGTCT GGCTGGCTGG CATAAATATC TCACTCGCAA TCAAATTCAG CCGATAGCGG
3961 AACGGGAAGG CGACTGGAGT GCCATGTCCG GTTTTCAACA AACCATGCAA ATGCTGAATG
4021 AGGGCATCGT TCCCACTGCG ATGCTGGTTG CCAACGATCA GATGGCGCTG GGCGCAATGC
4081 GCGCCATTAC CGAGTCCGGG CTGCGCGTTG GTGCGGATAT CTCGGTAGTG GGATACGACG
4141 ATACCGAAGA CAGCTCATGT TATATCCCGC CGTTAACCAC CATCAAACAG GATTTTCGCC
4201 TGCTGGGGCA AACCAGCGTG GACCGCTTGC TGCAACTCTC TCAGGGCCAG GCGGTGAAGG
4261 GCAATCAGCT GTTGCCCGTC TCACTGGTGA AAAGAAAAAC CACCCTGGCG CCCAATACGC
4321 AAACCGCCTC TCCCCGCGCG TTGGCCGATT CATTAATGCA GCTGGCACGA CAGGTTTCCC
4381 GACTGGAAAG CGGGCAGTGA GCGCAACGCA ATTAATGTGA GTTAGCTCAC TCATTAGGCA
4441 CCCCAGGCTT TACACTTTAT GCTTCCGGCT CGTATGTTGT GTGGAATTGT GAGCGGATAA
4501 CAATTTCACA CAGGAAACAG CTATGACCAT GATTACGGAT TCACTGGCCG TCGTTTTACA
4561 ACGTCGTGAC TGGGAAAACC CTGGCGTTAC CCAACTTAAT CGCCTTGCAG CACATCCCCC
4621 TTTCGCCAGC TGGCGTAATA GCGAAGAGGC CCGCACCGAT CGCCCTTCCC AACAGTTGCG
4681 CAGCCTGAAT GGCGAATGGC GCTTTGCCTG GTTTCCGGCA CCAGAAGCGG TGCCGGAAAG
4741 CTGGCTGGAG TGCGATCTTC CTGAGGCCGA TACTGTCGTC GTCCCCTCAA ACTGGCAGAT
4801 GCACGGTTAC GATGCGCCCA TCTACACCAA CGTAACCTAT CCCATTACGG TCAATCCGCC
4861 GTTTGTTCCC ACGGAGAATC CGACGGGTTG TTACTCGCTC ACATTTAATG TTGATGAAAG
4921 CTGGCTACAG GAAGGCCAGA CGCGAATTAT TTTTGATGGC GTTGGAATT
```

Appendix 19 Predicted Expressed Protein Sequences

Expressed GST Control (26.31 kDa)

MSPILGYWKI KGLVQPTRLL LEYLEEKYEE HLYERDEGDK WRNKKFELGL EFPNLPYYID GDVKLTQSMA IIRYIADKHN MLGGCPKERA EISMLEGAVL DIRYGVSRIA YSKDFETLKV DFLSKLPEML KMFEDRLCHK TYLNGDHVTH PDFMLYDALD VVLYMDPMCL DAFPKLVCFK KRIEAIPQID KYLKSSKYIA WPLQGWQATF GGGDHPPKSD LVPRGS

Expressed GST-FCAR (~38.76 kDa)

(GST) - GSMKVQTVAA VVVVAVVVTM TEAGLFPPKD CKYWCKDNLG INYCCGQPGV TYPPFTKNHL GRCPPVRDDT CTGVRTQLPT YCPHDGACQF RSKCCYDTCL KHHVCKTAEY PYY

Cleaved GST-FCAR Sequence (~12.45 kDa)

GSMKVQTVAA VVVVAVVVTM TEAGLFPPKD CKYWCKDNLG INYCCGQPGV TYPPFTKNHL GRCPPVRDDT CTGVRTQLPT YCPHDGACQF RSKCCYDTCL KHHVCKTAEY PYY

Expressed GST-ACAR (~36.73 kDa)

(GST) - GSRLFP PKDCKYWCKD NLGINYCCGQ PGVTYPPFTK NHLGRCPPVR DTCTGVRTQL PTYCPHDGAC QFRSKCCYDT CLKHHVCKT AEYPYY

Cleaved GST-ACAR Sequence (~10.42 kDa)

GSRLFP PKDCKYWCKD NLGINYCCGQ PGVTYPPFTK NHLGRCPPVR DTCTGVRTQL PTYCPHDGAC QFRSKCCYDT CLKHHVCKT AEYPYY

Appendix 20 Recovery of Inclusion Bodies

Recovery of inclusion bodies from the cell pellet was performed using the adapted procedure of Schoner *et al* (1992) as described by Sambrook and Russell (2001) in Molecular Cloning Volume 3.

The cell lysate pellet was resuspended in distilled water (1ml per gram of lysate pellet) and 100 µl of this was aliquoted into microfuge tubes which were then cantirguged for 15 min at 4 °C at 12,000 x g. The supernatants were discarded and the pellets resuspended in 0.1M Tris-HCl (pH 8.5), which contained different concentrations of urea (0.5 M, 1 M, 2 M and 4 M). These were then centrifuged at 12,000 x g for 15 min at 4 °C with the resultant supernatants collected and pellet resuspended in distilled water. Samples (10 µl) of both the supernatant and the resuspended pellets were analysed by SDS-PAGE as described in section 3.3.9.

Appendix 21 Restriction Sites

```
Restriction analysis on FCAR insert pcr product

Screened with 117 enzymes, 16 sites found

ApaBI  1   GCANNNNN/TGC
      283

BalI  2   TGG/CCA
      59      141

BamHI  1   G/GATCC
      5

BglI  1   GCCNNNN/NGGC
      183

BspMI  2   ACCTGCNNNN/
      302     177

DraIII  1   CACNNN/GTG
      310

Eco72I  1   CAC/GTG
      310

MscI  2   TGG/CCA
      59      141

PmaCI  1   CAC/GTG
      310

ScaI  1   AGT/ACT
      101

SciI  1   CTC/GAG
      365

XhoI  1   C/TCGAG
      363

Eco57I  1
CTGAAGNNNNNNNNNNNNNNN/
      320
```

```
                                                                  BalI
          BamHI                                                   MscI
1         GCGCGGATCCATGAAGGTGCAAACTGTAGCAGCCGTGGTGGTTGTGGCTGTGGTTGTGGC
          CGCGCCTAGGTACTTCCACGTTTGACATCGTCGGCACCACCAACACCGACACCAACACCG

                                                  ScaI
61        CATGACAGAGGCAAGGTTATTCCCTCCGAAGGACTGTAAGTACTGGTGCAAAGACAACCT
          GTACTGTCTCCGTTCCAATAAGGGAGGCTTCCTGACATTCATGACCACGTTTCTGTTGGA

                          BalI
                          MscI                                    BspMI
121       TGGAATAAACTACTGCTGTGGCCAGCCAGGAGTAACCTACCCACCTTTTACTAAAAGCCA
          ACCTTATTTGATGACGACACCGGTCGGTCCTCATTGGATGGGTGGAAAATGATTTTCGGT

            BglI
181       CTTGGGCAGGTGCCCTCCAGTCCGTGATACCTGTACTGGCGTCAGGACACAGCTACCAAC
          GAACCCGTCCACGGGAGGTCAGGCACTATGGACATGACCGCAGTCCTGTGTCGATGGTTG

                                                      ApaBI
241       GTACTGTCCCCATGATGGTGCATGTCAGTTCAGAAGCAAGTGCTGCTATGACACCTGCCT
          CATGACAGGGGTACTACCACGTACAGTCAAGTCTTCGTTCACGACGATACTGTGGACGGA

                    DraIII
                    PmaCI
           BspMI    Eco72I     Eco57I
301       GAAGCACCACGTGTGCAAGACTGCCGAATATCCTTATTATTAGACATCGCAGACCCGTGT
          CTTCGTGGTGCACACGTTCTGACGGCTTATAGGAATAATAATCTGTAGCGTCTGGGCACA

             SciI
           XhoI
361       AACTCGAGGCGC
          TTGAGCTCCGCG
```

Appendix 22 Alignment of 18S rRNA nucleotide sequences

```
N.norvegicus    CCGATTTAAGGCGAAACCGCGAATGGCTCATTAAATCAGCTATGTTTCATTGGATCTGTA
H.americanus    CCGATTTAAGGCGAAACCGCGAATGGCTCATTAAATCAGCTATGTTTCATTGGATCTGTA
H.epheliticus   CCGACTTAAGGCGAAACCGCGAATGGCTCATTAAATCAGCTATGATTCATTGGATCTGTA
C.sapidus       CCGAATNAAGGCGAAACCGCGAATGGCTNNNTAAATCAGCTATGATTCATTNNATCTGTA
C.maenas        GCCCTTTAAGGCGAAATCGCGAATGGCTCATTAAATCAGCTATGATTCATTGGATCTGTA
                 *   * ********* ***********   ************* ******  *******

N.norvegicus    TAGTCTCTGACCGCAAGG-AAGAGCGCTTTTATTAGTTCAAAACTGGTCGGGCCTCGGTC
H.americanus    TGGTCTCTGACCGCAAGGGAAGAGCGCTTTTATTAGTTCAAAACTGGTCGGGCCCCGGTC
H.epheliticus   ACGTCTCTGACCGCAAGGGAAGAGCGCTTTTATTAGTTCAAAACCGGTCGGGCCTCGGTC
C.sapidus       ACGTCTCTGACCGCAAGGGAAGAGNGCTTTTATTAGTTCAAAACCGGTCGGGCCTCGGTC
C.maenas        ACGTCTCTGACCGCAAGGGAAGAGCGCTTTTATTAGTTCAAAACCGGTCGGGCCTCGGTC
                  **************** ***** ****************** ********* *****

N.norvegicus    CGTAACCCACCTGTGGTGAATCTGAATAACTTCCGGCTGAGCGCACGGTCTCCGCACCGG
H.americanus    CGTAACCCACCTGTGGTGAATCTGAATAACTTCGGGCTGAGCGCACGGTCTCCGCACCGG
H.epheliticus   CGTCACCCCACCGTGTTGAATCTGAATAACTTGTCGCTGAGCGCACGGTCTCCGCACCGG
C.sapidus       CGNNNCCCCACTGTGTTGAATCTGAATAACTTTTTGCTGAGCGCACGGTCTCNGCNC— G
C.maenas        CGTCACCCCAACGTGTTGAATCTGAATAACTTGTCGCTGAGCGCACGGTCTCTGCACCGG
                **   ***    *** ****************   ***************** ** *  *

N.norvegicus    CGCCGCTTCTTTCAAGTGTCTGCCTTATCAGCTTTCGATTGTAGGTTATGCGCCTACAAT
H.americanus    CGCCGCTTCTTTCAAGTGTCTGCCTTATCAGCTTTCGATTGTAGGTTATGCGCCTACAAT
H.epheliticus   CGCCGCCTCTTTCAAGTGTCTGCCTTATCAGCTTTCGATTGTAGGTTATACGCCTACAAT
C.sapidus       CGCNGCCTCTTTCAAGTGTCTGCCTTATCAGCTTTCGATTGTAGGTTATACGCCTACNAT
C.maenas        CGCCGCCTCTTTCAAGTGTCTGCCTTATCAGCTTTCGATTGTAGGCTATGCGCCTACAAT
                *** ** ************************************ *** ******* **
```

```
N.norvegicus    GGCTATAACGGGTAACGGGGAATCAGGGTTCGATTCCGGAGAGGGAGCCTGAGAAACGGC
H.americanus    GGCTATAACGGGTAACGGGGAATCAGGGTTCGATTCCGGAGAGGGAGCCTGAGAAACGGC
H.epheliticus   GGCTATAACGGGTAACGGGGAATCAGGGTTCGATTCCGGAGAGGGAGCCTGAGAAACGGC
C.sapidus       GGCTNTNACGGGTNACGGGGAATGAGGGTTCGATTCCGGAGAGNGAGCCTGAGAAACGGC
C.maenas        GGCTATAACGGGTAACGGGGAATCAGGGTTCGATTCCGGAGAGGGAGCCTGGGAAACGGC
                **** * ****** ********* ******************* ******* ********

N.norvegicus    TACCACATCTAAGGAAGGCAGCAGGCACGCAAATTACCCACTCCCGGCACGGGGAGGTAG
H.americanus    TACCACATCTAAGGAAGGCAGCAGGCACGCAAATTACCCACTCCCGGCACGGGGAGGTAG
H.epheliticus   TACCACATCTAAGGAAGGCAGCAGGCACGCAAATTACCCACTCCCGGCACGGGGAGGTAG
C.sapidus       TACCACNTCTNAGGNNGGCAGCAGGCACGCNNATTACCCACTCC-GGCNCGGGGAGGTAG
C.maenas        TACCACATCTAAGGAAGGCAGCAGGCACGCAAATTACCCACTCCCGGCACGGGGAGGTAG
                ****** *** ***  **************  ************ *** ***********

N.norvegicus    TGACGAAAAATAACGATGTGAGTCTCATCNGAGGCCTCGCAATCGGAATGAGTACACTTT
H.americanus    TGACGAAAAATAACGATGTGAGTCTCATCCGAGGCCTCGCAATCGGAATGAGTACACTTT
H.epheliticus   TGACGAAAAATAACGATGCGAGACTCATCCGAGGCCTCGCAATCGGAATGAGTACACTTT
C.sapidus       TGACNAAAAATAACGATGCGAGACTCATCCGNGGCCTCGNNATCGGAATGAGTNCACTTT
C.maenas        TGACGAAAAATAACGATGCGAGACTCATCCGAGGCCTCGCAATCGGAATGAGTACACTTT
                **** ************* *** ****** * *******  ************ ******

N.norvegicus    AAATCCTTTAACGAGTATCCATTGGAGGGCAAGTCTGGTGCCAGCAGCCGCGGTAATTCC
H.americanus    AAATCCTTTAACGAGTATCCATTGGAGGGCAAGTCTGGTGCCAGCAGCCGCGGTAATTCC
H.epheliticus   AAATCCTTTAACGAGGATCTATTGGAGGGCAAGTCTGGTGCCAGCAGCCGCGGTAATTCC
C.sapidus       AAATCCTTTNACGAGGATCTATTGGAGGGCNAGTCTGGTGCCNGCAGCCNCGGTNATTCC
C.maenas        AAATCCTTTAACGAGGATCCATTGGAGGGCAAGTCTGGTGCCAGCAGCCGCGGTAATTCC
                ********* ***** *** ********** *********** ****** **** *****
```

```
N.norvegicus    AGCTCCAATAGCGTATATTAAAGTTGTTGCGGTTAA-AAAGCTCGTAGTTGGATCTCAGT
H.americanus    AGCTCCAATAGCGTATATTAAAGTTGTTGCGGTTAA-AAAGCTCGTAGTTGGATCTCAGT
H.epheliticus   AGCTCCAATAGCGTATATTAAAGTTGTTGCGGTTAA-AAAGCTCGTAGTTGGATTTCAGT
C.sapidus       AGCTGCAATANNGTATATTAAAGTTGTTGCGGTTANNAAAGCTCGTAGTTNNATTTCAGT
C.maenas        AGCTCCAATAGCGTATATTAAAGTTGTTGCGGTTAA-AAAGCTCGTAGTTGGATTTCAGT
                **** *****  ************************  *************  ** *****

N.norvegicus    TCCGGACTGACGGTGCACCGCCCGGTGTTTACTGTCACGCTCCGAACAGCCGC-----CC
H.americanus    TCCGGACTGACGGTGCACCGCCCGGTGTTTACTGTCACGCTCCGAACAGCCGC-----CC
H.epheliticus   TCTGGACTGACGGTTCACCGCCCGGTGCATACTGTCACGCTCCGAACAGCCACAACAGCC
C.sapidus       TCTGGACTGACGGT-TNCCGCNNGGTGCACACTGTCACNCTCCGAACAGCCACAAC—AC
C.maenas        TCTGGACTGACGGGTCACCGCCCGGTGCATACTGTCACGCTCCGAACAGCCACAACAGCC
                ** **********    ****  ****   ******** ************ *      *

N.norvegicus    CGCCGGCTCGCACGGGATGCTCTTTGTCGAGTGTCCCGAGTGGCCGG-AG-GTTTACTTT
H.americanus    CGCCGGCTCGCACGGGATGCTCTTTGTCGAGTGTCCCGAGTGGCCGG-AGAGTTTACTTT
H.epheliticus   CGCTGGCTCGCACGGGGTGCTCTTCATCGAGTGTCCCGCGTGGCCGGCAGAGTTTACTTT
C.sapidus       CGCTGGCCNN--NGGGGTGCTCTTCNCCNGGTGTCCNNNNNNNNNNN-NNNNNNNNNNN
C.maenas        CGCTGGCTCGCACGGGGTGCTCTTCATCGAGTGTCCCGCGTGGCCGGCAGAGTTTACTTT
                *** ***      *** *******   *  ******

N.norvegicus    GAAAAAATTAGAGTGCTCAGAGCAGGCTATTTGAATGGCCCGAATGGTGATGCA-TGGAA
H.americanus    GAAAAAATTAGAGTGCTCAGAGCAGGCTATTTGAATGGCCCGAATGGTGATGCA-TGGAA
H.epheliticus   GAAAAAATTAGAGTGCTCAAAGCAGGCTACACTGACGGCCTGAATGCCTATGCA-TGGAA
C.sapidus       NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNCCTGAATGCCTATGCANTGGAA
C.maenas        GAAAAAATTAGAGTGCTCAAAGCAGGCTACACTGACGGCCTGAATGCCTATGCA-TGGAA
                                                      ** *****   ***** *****
```

```
N.norvegicus     TAATGGAATAGGACCTCGGTTCTATTTTG-TTGGTTTTCGGAACCAGAGGTAATGACTAA
H.americanus     TAATGGAATAGGACCTCGGTTCTATTTTG-TTGGTTTTCGGAACCAGAGGTAATGACTAA
H.epheliticus    TAATGGAATAGGACCTCGGTTCTATTTTG-TCGGTTTTCTGAACCCGAGGTAATGACTAA
C.sapidus        TAATGGAATAGGACCTCGGNNCTNTTTTG-TCGGTTTTCTGAACCCGAGGTAATGACTAA
C.maenas         TAATGGAATAG-ACCTCGGTTCTATTTTGGTCGGTTTTCTGAACCCGAGGTAATGACTAA
                 ********** *******  ** ***** * ******* ***** **************

N.norvegicus     TCGGAACAGGCGGGGGCATTCGTATTGCGACGCTAGAGGTGAAATTCTTGGACCGTCGCA
H.americanus     TCGGAACAGGCGGGGGCATTCGTATTGCGACGCTAGAGGTGAAATTGTTGGACCGTCGCA
H.epheliticus    TAGGAACAGGCGGGGGCATTCGTATTGCGACGCTAGAGGTGAAATTCTTGGACCGTCGCA
C.sapidus        TAGGAACNGGCGGGGGCNTTCGTATTGCGACGCTAGAGGTGAAATTCTTGGACCGTCGCN
C.maenas         TAGGAACAGGCGGGG-CATTCGTATTGCGACGCTAGAGGTGAAATTCTTGGACCGTCGCA
                 * ***** ******* * ***************************** ************

N.norvegicus     AGACGAACTACTGCGAAAGCATTTGCCAAGGATGTTTTCATTAATCAAGAACGAAAGTTA
H.americanus     AGACGAACTACTGCGATAGCATTTGCCAAGGATGTTTTCATTAATCAAGAACGAAAGTTA
H.epheliticus    AGACGAACTACTGCGAAAGCATTTGCCAAGGATGTTTTCATTAATCAAGAACGAAAGTTA
C.sapidus        AGACGAACTNCTGCGAAAGCATTTGCCNAGGATGTTT-CNTTNATCNAGAANGAAAGTTA
C.maenas         AGACGAACTACTGCGAAAGCATTTGCCAAGGATGTTTTCATTAATCAAGAACGAAAGTTA
                 ********* ****** ********** ********* * ** *** **** ********

N.norvegicus     GAGGTTCGAAGGCGATCAGATACCGCCCTAGTTCTAACCATAAACGATGCCAACTAGCGA
H.americanus     GAGGTTCGAAGGCGATCAGATACCGCCCTAGTTCTAACCATAAACGATGCCAACTAGCGA
H.epheliticus    GAGGTTCGAAGGCGATCAGATACCGCCCTAGTTCTAACCATAAACGATGCTGACCAGCGA
C.sapidus        GAGGTTCGAAGGCGATCAGATACCGCNNNNNNNNNAACCNTAAACGATGCTGACCAGCGA
C.maenas         GAGGTTCGAAGGCGATCAGATACCGCCCTAGTTCTAACCATAAACGATGCTGACCAGCGA
                 *************************          **** **********  ** *****
```

```
N.norvegicus     TCCGCCGGCGTTATTCCCATGACCCGGCGGGCAGCTTCCGGGAAACCAAAGTCTTTGGGT
H.americanus     TCCGCCGGCGTTATTCCCATGACCCGGCGGGCAGCTTCCGGGAAACCAAAGTCTTTGGGT
H.epheliticus    TCCGCCGGCGTTATTCCCATGACCCGGCGGGCAGCTTCCGGGAAACCAAAGTCTTTGGGT
C.sapidus        TCCGCCGGNNTTATTNNCATGACCCGGCCNCCAGCTTCCGGGAAACCAAAGTCTTTGGGT
C.maenas         TCCGCCGGTGTTACTCCCATGACCCGGCGGGCAGCTTCCGGGAAACCAAAGTCTTTGGGT
                 ********  *** *  ***********   ****************************

N.norvegicus     TCCGGGGGAAGTATGGTTGCAAAGCTGAAACTTAAAGGAATTGACGGAAGGGCACCACCA
H.americanus     TCCGGGGGAAGTATGGTTGCAAAGCTGAAACTTAAAGGAATTGACGGAAGGGCACCACCA
H.epheliticus    TCCGGGGGAAGTATGGTTGCAAAGCTGAAACTTAAAGGAATTGACGGAAGGGCACCACCA
C.sapidus        TCCGGGGGAAGTATGGTTGCAAAGCTGAAACTCAAAGGAATTGACGGNNNNNNNNNNNNN
C.maenas         TCCGGGGGAAGTATGGTTGCAAAGCTGAAACTTAAAGGGATTGACGGAAGGGCACCACCA
                 ********************************* ***** ********

N.norvegicus     GGAGTGGAGCCTGCGGCTTAATTTGACTCAACACGGGAAACCTCACCAGGCCCAGACACC
H.americanus     GGAGTGGAGCCTGCGGCTTAATTTGACTCAACACGGGAAACCTCACCAGGCCCAGACACC
H.epheliticus    GGAGTGGAGCCTGCGGCTTAATTTGACTCAACACGGGGAACCTCACCAGGCCCAGACACC
C.sapidus        NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNACACGGGGAACCTCACCAGGCCCAGACACC
C.maenas         GGAGTGGAGCCTGCGGCTTAATTTGACTCAACACGGGGAACCTCACCAGGCCCAGACACC
                                              ******* **********************

N.norvegicus     GGAAGGATTGACAGATTGAGAGCTCTTTCTCGATTCGGTGGGTGGTGGTGCATGGCC-GT
H.americanus     GGAAGGATTGACAGATTGAGAGCTCTTTCTCGATTCGGTGGGTGGTGGTGCATGGCC-GT
H.epheliticus    GGAAGGATTGACAGATTGAGAGCTCTTTCTCGATTCGGTGGGTGGTGGTGCATGGCC-GT
C.sapidus        GGAAGGANNGACAGATTGAGAGCTCNNTCTC-ATTNGGTGGGTGGTNGTGCATNNCGNGT
C.maenas         GGAAGGATTGACAGATTGAGAGCTCTTTCTTGATTCGGTGGGTGGTGGTGCATGGCC-GT
                 *******  ****************  ***  *** **********  ******  *  **
```

```
N.norvegicus     TCTTAGTTGGTGGAGCGATTTGTCTGGTTAATTCCGATAACGAACGAGACTCTGGCCTAC
H.americanus     TCTTAGTTGGTGGAGCGATTTGTCTGGTTAATTCCGATAACGAACGAGACTCTGGCCTAC
H.epheliticus    TCTTAGTTGGTGGAGCGATTTGTCTGGTTANNNNNNNNNNNNNNNNNNNNNTGGCCTAC
C.sapidus        TCTTAGTTGGTGGAGCGNNNNNNNNNNNNNNNNNCGATNACGAACGAGANNNNNNNNNNN
C.maenas         TCTTAGTTGGTGGAGCGATTTGTCTGGTTAATTCCGATAACGAACGAGACTCTGGCCTAC
                 ****************

N.norvegicus     TAAGTAGTCGACGGATCTCCAGAAAATG-GTGTCCAGTTCGCAACTTCTTCTTAGAGGGA
H.americanus     TAAGTAGTCGACGGATCTCCAGAAAATG-GTGTCCAGTTCGCAACTTCTTCTTAGAGGGA
H.epheliticus    TAACTAGTCGACGGATCTCCAGCAATTG-GTGTCCAGTTCGCAGCTTCTTCTTAGAGGGA
C.sapidus        NNNNNNNNNNNNNNNNNNNNNNNNNNNNNGTGTCCAGTTCGCAGCTTCTTCTTAGAGGGA
C.maenas         TAACTAGTCGACGGATCTCCAGCAATTG-GTGTCCAGTTCGCAGCTTCTTCTTAGAGGGA
                                             ************** ****************

N.norvegicus     TAAGCGGCAATTCTAGCCGCACGAGATTGAGCAATAACAGGTCTGTGATGCCCTTAGATG
H.americanus     TAAGCGGCAATTCTAGCCGCACGAGATTGAGCAATAACAGGTCTGTGATGCCCTTAGATG
H.epheliticus    TAAGCGGCAATTCTAGCCGCACGAGATTGAGCAATAACAGGTCTGTGATGCCCTTAGATG
C.sapidus        TAA-CGGCAATTCTAGCCGCACGAGATTGAGCAATAACAAGTCTGTGATGCCCTTAGATG
C.maenas         TAAGCGGCAATTCTAGCCGCACGAGATTGAGCAATAACAGGTCTGTGATGCCCTTAGATG
                 *** ********************************** ********************

N.norvegicus     TTCTGGGCCGCACGCGCGCTACACTGAAGGGATCAACGAGTTTTCCCCCTCCGAGAGGAG
H.americanus     TTCTGGGCCGCACGCGCGCTACACTGAAGGGATCAACGAGTTTTCCCCCTCCGAGAGGAG
H.epheliticus    TTCTGGGCCGCACGCGCGCTACACTGAAGGGATCAACGTGTCCTCCCCCTCCGAGAGGAG
C.sapidus        TTCTGGGC-GCACGCGCGCTACACTGAAGGGATCAACGTGTCCTCCCNCTCCGAGAGGAG
C.maenas         TTCTGGGCCGCACGCGCGCTACACTGAAGGGATCAACGTGTCCTCCCCCTCCGAGAGGAG
                 ******** ****************************** **  **** ************
```

```
N.norvegicus     CGGGTAACCCGTTCAAAGCCTTTCTTGATAGGGATTGGGGCTTGCAATTGTTTCCCATGA
H.americanus     CGGGTAACCCGTTCAAAGCCTTTCTTGATAGGGATTGGGGCTTGCAATTGTTTCCCATGA
H.epheliticus    CGGGTAACCCGTTGAAATCCTTTCATGATAGGGATTGGGGTTTGCAATTGTCTCCCATGA
C.sapidus        CGGGNNNCCCGTTGAAATCCNTTCATGATAGGGATTGGGGTTTGCAATTGTCTCCCATGA
C.maenas         CGGGTAACCCGTTGAAATCCTTTCATGATAGGGATTGGGGTTTGCAATTATCTCCCATGA
                 ****   ****** *** ** *** ************** ******** * ********

N.norvegicus     ACGAGGAATTCCCAGTAAGT
H.americanus     ACGAGGAATTCCCAGTAAGT
H.epheliticus    ACGAGGAATTCCCAGTAAGC
C.sapidus        ANNNGGAATTCCCAGTAAGC
C.maenas         ACGAGGAATTCCCAGAAGGG
                 *   *********** * *
```

Appendix 23 Carcinin Real-Time PCR Primer and Probe location

XXX - UTR

XXX – Carcinin REV QRTPCR primer

XXX – Carcinin FWD QRTPCR

XXX – Carcinin probe

XXX – QRTPCR amplicon (129 bp)

XX – exon / exon boundary

X – variable residue leading to isoforms

↓ - cleavage site

```
1       ACGCGGGGAGACCAGAACTGCACCCTGTGGTGGACACTTCTGTTTTGACCAACAGCTTCT

61      TCAAGAACACATTGAAACATGAAGGTGCAAACTGTAGCAGCCGTGGTGGTTGTGGCTGTG
 1                        M  K  V  Q  T  V  A  A  V  V  V  V  A  V
                                 ↓
121     GTTGTGACCATGACAGAGGCAAGGTTATTCCCTCCGAAGGACTGTAAGTACTGGTGCAAA
 15      V  V  A  M  T  E  A  R  L  F  P  P  K  D  C  K  Y  W  C  K

181     GACAACCTTGGAATAAACTACTGCTGTGGCCAGCCAGGAGTAACCTACCCACCTTTTACT
 35      D  N  L  G  I  N  Y  C  C  G  Q  P  G  V  T  Y  P  P  F  T

241     AAAAGCCACTTGGGCAGGTGCCCTCCAGTCCGTGATACCTGTACTGGCGTCAGGACACAG
 55      K  S  H  L  G  R  C  P  P  V  R  D  T  C  T  G  V  R  T  Q

301     CTACCAACGTACTGTCCCCATGATGGTGCATGTCAGTTCAGAAGCAAGTGCTGCTATGAC
 75      L  P  T  Y  C  P  H  D  G  A  C  Q  F  R  S  K  C  C  Y  D

361     ACCTGCCTGAAGCACCACGTGTGCAAGACTGCCGAATATCCTTATTATTAGACATCGCAG
 95      T  C  L  K  H  H  V  C  K  T  A  E  Y  P  Y  Y  *

421     ACCCGTGTAAGAAATCTTACACCTAGTACATAGATCAGATCTGAAATAAGAAACTCCGTA

481     ATCTACGGAAATTCTACAAACACTATGACGCATGGTTACCTACTGTACTGTATACTGTAT

541     GCAATTATAGGCAACAACAAAATTAATGATTAATAAACATTGTTGTGTTTAATGAGCAAA

542     AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
```

Appendix 24 C. maenas β-actin Sequence

```
1     GTCGGTGATG AGGCTCAGAG CAAGCGAGGT ATCCTGACTC TCAAGTACCC CATCGAGCAC
61    GGCATCATTA CCAACTGGGA CGACATGGAG AAGATCTGGC ATCACTCCTT CTACAACGAG
121   CTCCGCATTG CCCCTGAGGA GTCTCCCGTC CTTCTCACTG AGGCTCCCCT CAACCCCAAG
181   GCTAACCGTG AGAAGATGAC CCAGATCATG TTTGAAACCT TCAACACCCC CGCCATGTAC
241   GTGGCCATCC AGGCCGTGCT GTCCCTGTAC GCCTCTGGCC GTACCACCGG CATCGTGCTC
301   GATACTGGAG ATGGTGTTAC CCACACCGTC CCCATCTATG AAGGTTACTG CCTTCCCCAC
361   GCCATCCTGC GTCTCGATCT GGCTGGCCGT GACCTGACTG CCTACCTCAC CAAGATCATG
421   ACTGAGCGTG GCTACTCCTT CACCACCACA GCTGAGCGAG AAATCGTTCG CGATATCAAG
481   GAGAAACTTT GCTATGTCGC TCTTGATTTC GAGAGTGAGA TGAGTGTAGC CGCTGCTTCT
541   TCTTCTCTAG ATAAATCTTA CGAGCTTCCC GATGGTCAGG TTACCACCAT CGGCAATGAG
601   CGATTCCGCT GCCC
```

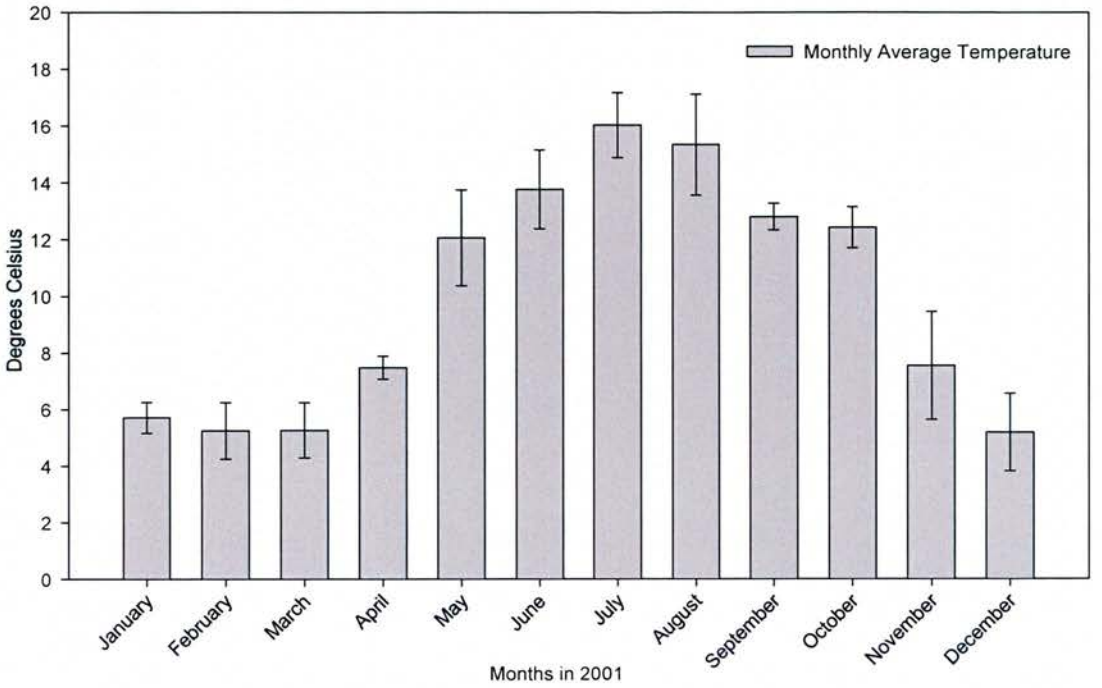
Appendix 25 Alignment of β-actin protein sequences

H.gammarus (CAE46725), C.lateralis(AAL40077), C.maenas(pending)

```
H.gammarus      MCDDEVAALVVDNGSGMCKAGFAGDDAPRAVFPSIVGRPRHQGVMVGMGQKDSYVGDEAQ
C.lateralis     MCDEDAESLVVDNGSGMVKAGFAGDDAPRAVFPSIVGRPRHQGVMVGMGQKDAYVGEQAQ
C.maenas        ----------------------------------------------ALVGDEAQ
                                                               . **..**

H.gammarus      SKRGILTLKYPIEHGIVTNWDDMEKIWHHTFYNELRVAPEEHPVLLTEAPLNPKANREKM
C.lateralis     SKRGILTLKYPIEHGIITNWDDMEKIWHHSFYNELRVAPEESPVLLTEAPLNPKANREKM
C.maenas        SKRGILTLKYPIEHGIITNWDDMEKIWHHSFYNELRIAPEESPVLLTEAPLNPKANREKM
                ***************.***********.******.**** ******************

H.gammarus      TQIMFETFNTPAMYVAIQAVLSLYASGRTTGIVLDSGDGVSHTVPIYEGYALPHAILRLD
C.lateralis     TQIMFEVFNTPAMYVAIQAVLSLYASGRTTGIVLDTGDGVTHTVPIYEGYCLPHAILRLD
C.maenas        TQIMFETFNTPAMYVAIQAVLSLYASGRTTGIVLDTGDGVTHTVPIYEGYCLPHAILRLD
                ****** ****************************.****.*********.*********

H.gammarus      LAGRDLTDYLMKILTERGYTFTTTAEREIVRDIKEKLCYVALDFEQEMTTAASSSSLEKS
C.lateralis     LAGRDLTAYLTKIMTERGYSFTTTAEREIVRDIKEKLCYVALDFESEMNVAAAPSSLEKS
C.maenas        LAGRDLTAYLTKIMTERGYSFTTTAEREIVRDIKEKLCYVALDFESEMSVAAASSSLDKS
                ******* ** **.*****.************************ **  **. ***.**
```

```
H.gammarus      YELPDGQVITIGNERFRCPEALFQPSFLGMESCGIHETTYNSIMKCDVDIRKDLYANTVL
C.lateralis     YELPDGRFITIGNERFRCPESLFQPSFLGMESVGIHETVYNSIMRCDIDIRKDLFANNVL
C.maenas        YELPDGQVTTIGNERFRC----------------------------------------
                ******.  *********

H.gammarus      SGGTTMYPGIADRMQKEITALAPSTMKIKIIAPPERKYSVWIGGSILASLSTFQQMWISK
C.lateralis     SGRTTMYPGIADRMQKEITALAPPTIKIKIIAPPERKYSVWIGGSILASLSTFQTMWITK
C.maenas        ------------------------------------------------------------

H.gammarus      QEYDESGPSIVHRKCF
C.lateralis     EEYDESGPGIVHRKCF
C.maenas        ----------------
```

Appendix 26 Sea Temperatures



Appendix 27 Temperature Sample Details

Temperature °C	Total Cells	RNA µg ml^{-1}
5	3.30E+07	2540
5	1.60E+07	1420
5	2.80E+07	2760
5	3.80E+07	3720
5	1.40E+07	1640
10	4.40E+07	1460
10	3.70E+07	1140
10	5.10E+07	400
10	6.70E+07	1440
10	5.20E+07	1620
15	1.54E+07	640
15	1.07E+07	1700
15	2.08E+07	420
15	1.44E+07	880
15	1.40E+07	200
20	1.19E+07	1660
20	1.20E+07	3210
20	1.60E+07	500
20	1.44E+07	500
20	1.70E+07	820

Table 35: Samples collected from crabs at different temperatures

Appendix 28 Heat-Killed Bacterial Challenge

	Bacterial challenge		Saline placebo	
Time (h)	Total Cells	RNA µg ml^{-1}	Total Cells	RNA µg ml^{-1}
0	5.01E+07	810	5.01E+07	810
0	7.50E+07	1240	7.50E+07	1240
0	7.39E+07	2260	7.39E+07	2260
0	9.11E+07	1250	9.11E+07	1250
0.5	5.80E+06	620	1.96E+07	1670
0.5	7.50E+06	70	7.70E+06	300
0.5	1.84E+07	90	6.12E+07	1280
0.5	1.47E+07	140	1.74E+07	1600
3	3.60E+06	520	2.71E+07	180
3	9.50E+06	320	7.30E+07	620
3	1.67E+07	900	7.68E+07	2270
3	9.15E+06	1200	7.16E+07	3450

Table 36: 30 min and 3 h cell counts and RNA concentration

		Bacterial challenge		Saline placebo		Stick wounding	
Time (h)	°C	Total Cells	RNA µg ml-1	Total Cells	RNA µg ml-1	Total Cells	RNA µg ml-1
0	5	2.39E+07	1450	2.39E+07	1450	2.39E+07	1450
0	5	4.71E+07	800	4.71E+07	800	4.71E+07	800
0	5	7.80E+07	840	7.80E+07	840	7.80E+07	840
0	5	3.42E+07	790	3.42E+07	790	3.42E+07	790
6	5	8.05E+06	230	4.28E+07	2020	3.21E+07	1380
6	5	5.91E+07	2920	4.44E+07	1210	6.15E+07	1950
6	5	5.92E+07	1730	3.66E+07	750	3.28E+07	110
6	5	4.57E+07	1770	2.24E+07	750	2.66E+07	1190
12	5	2.02E+07	850	1.32E+07	1500	1.12E+07	880
12	5	4.87E+07	910	4.67E+07	1060	2.84E+07	670
12	5	7.23E+07	140	4.28E+07	1510	4.54E+07	350
12	5	4.88E+07	2280	8.69E+07	2350	3.19E+07	1690
24	5	3.88E+07	3460	4.80E+07	1920	1.31E+07	540
24	5	3.12E+07	980	6.62E+07	720	8.48E+07	550
24	5	3.99E+07	590	5.83E+07	730	4.57E+07	1120
24	5	2.32E+07	890	1.56E+07	1350	2.18E+07	580
48	5	2.46E+07	540	2.44E+07	3410	1.95E+07	630
48	5	3.92E+07	1350	2.36E+07	860	7.16E+07	3410
48	5	4.44E+07	910	4.15E+07	1030	2.99E+07	130
48	5	4.27E+07	3420	2.30E+07	980	4.20E+07	3370
84	5	3.62E+07	610	1.74E+07	1840	2.93E+07	900
84	5	5.17E+07	2650	3.21E+07	2450	3.29E+07	1330
84	5	4.25E+07	750	2.98E+07	710	5.99E+07	1030
84	5	5.81E+07	1330	3.98E+07	570	2.25E+07	620

Table 37: 5 °C samples of 6-84 h heat-killed bacterial challenge experiment

		Bacterial challenge		Saline placebo		Stick wounding	
Time (h)	°C	Total Cells	RNA µg ml-1	Total Cells	RNA µg ml-1	Total Cells	RNA µg ml-1
0	15	2.44E+07	510	2.44E+07	510	2.44E+07	510
0	15	2.42E+07	1070	2.42E+07	1070	2.42E+07	1070
0	15	1.81E+07	790	1.81E+07	790	1.81E+07	790
0	15	2.86E+07	330	2.86E+07	330	2.86E+07	330
6	15	8.70E+06	80	2.02E+07	390	6.54E+07	340
6	15	5.91E+07	4040	4.56E+07	2380	4.96E+07	1930
6	15	4.52E+07	1610	3.18E+07	1590	1.25E+07	70
6	15	3.76E+07	1770	2.41E+07	550	3.27E+07	870
12	15	1.63E+07	370	7.90E+06	600	1.88E+07	450
12	15	6.27E+07	1570	5.17E+07	2000	3.76E+07	500
12	15	3.23E+07	830	5.36E+07	440	5.07E+07	750
12	15	8.56E+07	780	3.20E+07	1090	3.57E+07	600
24	15	2.63E+07	570	4.26E+07	550	1.54E+07	430
24	15	1.92E+07	550	2.09E+07	1060	4.56E+07	490
24	15	2.19E+07	400	1.44E+07	1830	1.14E+08	3180
24	15	6.55E+06	430	1.33E+07	1380	5.84E+07	2420
48	15	1.25E+07	110	2.48E+07	1510	4.24E+07	840
48	15	6.05E+07	1140	3.18E+07	1970	2.04E+07	2040
48	15	2.54E+07	1360	5.24E+07	920	1.87E+07	2980
48	15	4.09E+07	2470	4.41E+07	1130	6.59E+07	1630
84	15	4.85E+07	1570	5.76E+07	1340	2.57E+07	2110
84	15	3.10E+07	380	6.09E+07	2280	4.63E+07	470
84	15	8.40E+06	370	6.60E+07	1400	4.76E+07	1000
84	15	3.64E+07	2050	6.09E+07	1090	3.80E+07	2620

Table 38: 15 °C samples of heat-killed bacterial challenge experiment

References

Aagaard A. (1996) *In situ* variation in heart rate of the shore crab *Carcinus maenas* in relation to environmental factors and physiological condition. *Marine Biology* **125**, 765-772.

Abello P., Warman C. G. and Naylor E. (1997) Circatidal moulting rhythm in the shore crab *Carcinus maenas*. *Journal of the Marine Biological Association of the United Kingdom* **77**, 277-279.

Acton R. T., Weinheimer P. and Evans E. E. (1969) A bactericidal system in the lobster *Hommarus americanus*. *Journal of Invertebrate Pathology* **13**, 463-464.

Adachi K., Hirata T., Nishioka T. and Sakaguchi M. (2003) Hemocyte components in crustaceans convert hemocyanin into a phenoloxidase-like enzyme. *Comparative Biochemistry and Physiology B-Biochemistry & Molecular Biology* **134**, 135-141.

Adams A. (1991) Response of penaeid shrimp to exposure to Vibrio species. *Fish and Shellfish Immunology* **1**, 59-70.

Aderem A. and Ulevitch R. J. (2000) Toll-like receptors in the induction of the innate immune response. *Nature* **406**, 782-787.

Aguirre G. G. and Ascencio V. F. (2000) Infectious disease in shrimp species with aquaculture potential. *Recent Research Developments in Microbiology* **4**, 333-348.

Ahmed A. M., Baggott S. L., Maingon R. and Hurd H. (2002) The costs of mounting an immune response are reflected in the reproductive fitness of the mosquito *Anopheles gambiae*. *Oikos* **97**, 371-377.

Akira S. and Hemmi H. (2003) Recognition of pathogen-associated molecular patterns by TLR family. *Immunology Letters* **85**, 85-95.

Akira S. and Takeda K. (2004) Toll-Like Receptor Signalling. *Nature Reviews Immunology* **4**.

Amabile-Cuevas C. F., Cardenas-Garcia M. and Ludgar M. (1995) Antibiotic Resistance. *American Scientific* **83**, 320-329.

Amersham Pharmacia Biotech. (2001) *The Recombinant Protein Handbook*, Uppsala.

Amersham Pharmacia Biotech. (2002) *GST Gene Fusion System- Handbook*, Buckinghamshire.

Amiche M., Seon A. A., Pierre T. N. and Nicolas P. (1999) The dermaseptin precursors: a protein family with a common preproregion and a variable C-terminal antimicrobial domain. *Febs Letters* **456**, 352-356.

Andersons D., Engstrom A., Josephson S., Hansson L. and Steiner H. (1991) Biologically active and amidated cecropin produced in a baculovirus expression system from a fusion construct containing the antibody-binding part of protein A. *Biochemical Journal* **280 (Pt 1)**, 219-224.

Andreu D. and Rivas L. (1998) Animal Antimicrobial Peptides: An Overview. *Biopolymers* **47**, 415-433.

Anger K., Spivak E. and Luppi T. (1998) Effects of reduced salinities on development and bioenergetics of early larval shore crab, *Carcinus maenas*. *Journal of Experimental Marine Biology and Ecology* **220**, 287-304.

Applied Biosystems. (2001) Relative Quantitation of Gene Expression. Applied Biosystems Ltd.

Applied Biosystems. (2002) TaqMan Universal PCR Mastermix. Applied Biosystems.

Applied Biosystems. (2004) Applied Biosystems Website. Applied Biosystems.

Arala-Chaves M. and Sequeira T. (2000) Is there any kind of adaptive immunity in invertebrates? *Aquaculture* **191**, 247-258.

Aspan A., Sturtevant J., Smith V. J. and Söderhäll K. (1990) Purification and Characterization of a Prophenoloxidase Activating Enzyme From Crayfish Blood-Cells. *Insect Biochemistry* **20**, 709.

Bachau A. G. (1981) Crustaceans. In *Invertebrate Blood Cells 2* (Edited by Rowley N. A. R. a. A. F.), Vol. 12, p. 385-420. Academic Press, London.

Bachère E. (1998) Shrimp Immunity and disease control: An Integrated Approach. In *Advances in Shrimp Biotechnology* (Edited by Flegel T. W.). National Centre for Genetic Engineering and Biotechnology, Bangkok.

Bachère E. (2000) Shrimp and Immunity. *Aquaculture* **191**, 3-11.

Bachère E., Gueguen Y., Gonzalez M., de Lorgeril J., Garnier J. and Romestand B. (2004) Insights into the anti-microbial defense of marine invertebrates: the penaeid shrimps and the oyster *Crassostrea gigas*. *Immunological Reviews* **198**, 149-168.

Bachère E., Mialhe E., Noel D., Boulo V., Morvan A. and Rodriguez J. (1995) Knowledge and Research Prospects in Marine Mollusk and Crustacean Immunology. *Aquaculture* **132**, 17-32.

Bals R. and Wilson J. M. (2003) Cathelicidins--a family of multifunctional antimicrobial peptides. *Cellular and Molecular Life Sciences* **60**, 711-720.

Bamber S. D. and Depledge M. H. (1997a) Evaluation of changes in the adaptive physiology of shore crabs (*Carcinus maenas*) as an indicator of pollution in estuarine environments. *Marine Biology* **129**, 667-672.

Bamber S. D. and Depledge M. H. (1997b) Responses of shore crabs to physiological challenges following exposure to selected environmental contaminants. *Aquatic Toxicology* **40**, 79-92.

Barnes R. D. (1986) *Invertebrate Zoology*. Saunders College Publishing, Pennsylvania.

Barnes R. S. K., Calow P. and Olive P. J. W. (1993) *The Invertebrates - a new synthesis*. Blackwell Science Ltd., Oxford.

Barra D., Simmaco M. and Boman H. G. (1998) Gene-encoded peptide antibiotics and innate immunity. Do 'animalcules' have defence budgets? *Febs Letters* **430**, 130-134.

Barrell P. J., Liew O. W. and Conner A. J. (2004) Expressing an antibacterial protein in bacteria for raising antibodies. *Protein Expression and Purification* **33**, 153-159.

Bartlett T. C., Cuthbertson B. J., Shepard E. F., Chapman R. W., Gross P. S. and Warr G. W. (2002) Crustins, homologues of an 11.5-kDa antibacterial peptide, from two species of penaeid shrimp, *Litopenaeus vannamei* and *Litopenaeus setiferus*. *Marine Biotechnology* **4**, 278-293.

Baschmair A., Finley D. and Varshavsky A. (1986) In vivo half-life of a protein is a function of its amino-terminal residue. *Science* **234**, 179-186.

Battistella S., Bonivento P. and Amirante G. A. (1996) Hemocytes and immunological reactions in crustaceans. *Italian Journal of Zoology* **63**, 337-343.

Bayne C. J. (1990) Phagocytosis and Non-Self Recognition in Invertebrates. *BioScience* **40**, 723-731.

Bell K. L. and Smith V. J. (1993) *In vitro* Superoxide Production By Hyaline Cells of the Shore Crab *Carcinus maenas* (L). *Developmental and Comparative Immunology* **17**, 211-219.

Bell K. L. and Smith V. J. (1995) Occurrence and Distribution of Antioxidant Enzymes in the Hemolymph of the Shore Crab *Carcinus maenas*. *Marine Biology* **123**, 829-836.

Berrill M. (1982) The Life Cycle of the Green Crab *Carcinus maenas* at the Northern End of its Range. *Journal of Crustacean Biology* **2**, 31-39.

Birnboim H. C. and Doly J. (1979) A rapid alkaline extraction procedure for screening recombinant plasmid DNA. *Nucleic Acids Research* **7**, 1513-1523.

Blanchard J. S. (2003) Resisting bacterial drug resistance. *Chemistry and Biology* **10**, 104-106.

Blom N., Gammeltoft S. and Brunak S. (1999) Sequence and structure prediction of eukaryotic protein phosphorylation sites. *Journal of Molecular Biology* **294**, 1351-1362.

Boman H. G. (1991) Antibacterial Peptides - Key Components Needed in Immunity. *Cell* **65**, 205-207.

Boman H. G. (1995) Peptide Antibiotics and Their Role in Innate Immunity. *Annual Review of Immunology* **13**, 61-92.

Boman H. G. (2003) Antibacterial peptides: basic facts and emerging concepts. *Journal of Internal Medicine* **254**, 197-215.

Borregaard N. (2001) Encyclopedia of Life Sciences, Vol. 2001. Nature Publishing Group.

Bradley M. K. (1990) Overexpression of Proteins in Eukaryotes. In *Guide to Protein Purification* (Edited by Deutscher M. P.), Vol. 182, p. 112-132. Academic Press, London.

Brahmachary M., Krishnan S. P., Koh J. L., Khan A. M., Seah S. H., Tan T. W., Brusic V. and Bajic V. B. (2004) ANTIMIC: a database of antimicrobial sequences. *Nucleic Acids Research* **32 Database issue**, D586-589.

Brandin E. R. and Pistole T. G. (1985) Presence of Microorganisms in Hemolymph of the Horseshoe-Crab *Limulus polyphemus*. *Applied and Environmental Microbiology* **49**, 718-720.

Breteler W. C. M. K. (1975a) Food Consumption, Growth and Energy Metabolism of Juvenile Shore crabs, *Carcinus maenas*. *Netherlands Journal of Sea Research* **9**, 255-272.

Breteler W. C. M. K. (1975b) Oxygen Consumption and Respiratory Levels of Juvenile Shore Crabs, *Carcinus maenas*, In Relation to Weight and Temperature. *Netherlands Journal of Sea Research* **9**, 243-254.

British Marine Life Study Society. (2000) Shore Crab.

Brogden K. A., Ackermann M., McCray P. B. and Tack B. F. (2003) Antimicrobial peptides in animals and their role in host defences. *International Journal of Antimicrobial Agents* **22**, 465-478.

Brown P. (2001) Cinderella goes to the ball. *Nature* **410**, 1018-1020.

Bulet P., Hetru C., Dimarcq J. L. and Hoffmann D. (1999) Antimicrobial peptides in insects; structure and function. *Developmental and Comparative Immunology* **23**, 329-344.

Bulnheim H. P. and Bahns S. (1996) Genetic variation and divergence in the genus Carcinus (Crustacea, Decapoda). *Internationale Revue Der Gesamten Hydrobiologie* **81**, 611-619.

Burnet F. M. (1988) Evolution of the Immune Process in Vertebrates. *Nature* **218**, 426-430.

Busselen P. (1970) Effects of Moulting Cycle and Nutritional Conditions on Haemolymph proteins in *Carcinus maenas*. *Comparative Biochemistry and Physiology* **37**, 73-83.

Bustin S. A. (2000) Absolute quantification of mRNA using real-time reverse transcription polymerase chain reaction assays. *Journal of Molecular Endocrinology* **25**, 169-193.

Bustin S. A. (2002) Quantification of mRNA using real-time reverse transcription PCR (RT-PCR): trends and problems. *Journal of Molecular Endocrinology* **29**, 23-39.

Cadet P., Zhu W., Mantione K. J., Baggerman G. and Stefano G. B. (2002) Cold stress alters *Mytilus edulis* pedal ganglia expression of mu opiate receptor transcripts determined by real-time RT-PCR and morphine levels. *Molecular Brain Research* **99**, 26-33.

Cerenius L., Bangyeekhun E., Keyser P., Söderhäll I. and Söderhäll K. (2003) Host prophenoloxidase expression in freshwater crayfish is linked to increased resistance to the crayfish plague fungus, *Aphanomyces astaci*. *Cellular Microbiology* **5**, 353-357.

Chang E. S., Chang S. A., Keller R., Reddy P. S., Snyder M. J. and Spees J. L. (1999) Quantification of stress in lobsters: Crustacean hyperglycemic hormone, stress proteins, and gene expression. *American Zoologist* **39**, 487-495.

Charlet M., Chernysh S., Philippe H., Hetru C., Hoffmann J. A. and Bulet P. (1996) Innate immunity - Isolation of several cysteine-rich antimicrobial peptides from the blood of a mollusc, *Mytilus edulis*. *Journal of Biological Chemistry* **271**, 21808-21813.

Chen J., Anderson J. B., DeWeese-Scott C., Fedorova N. D., Geer L. Y., He S., Hurwitz D. I., Jackson J. D., Jacobs A. R., Lanczycki C. J., Liebert C. A., Liu C., Madej T., Marchler-Bauer A., Marchler G. H., Mazumder R., Nikolskaya A. N., Rao B. S., Panchenko A. R., Shoemaker B. A., Simonyan V., Song J. S., Thiessen P. A., Vasudevan S., Wang Y., Yamashita R. A., Yin J. J. and Bryant S. H. (2003) MMDB: Entrez's 3D-structure database. *Nucleic Acids Research* **31**, 474-477.

Chen J. Y., Pan C. Y. and Kuo C. M. (2004) cDNA sequence encoding an 11.5-kDa antibacterial peptide of the shrimp *Penaeus monodon*. *Fish and Shellfish Immunology* **16**, 659-664.

Cheng W. and Chen J. C. (1998) Enterococcus-like infections in *Macrobrachium rosenbergii* are exacerbated by high pH and temperature but reduced by low salinity. *Diseases of Aquatic Organisms* **34**, 103-108.

Cheng W. and Chen J. C. (2000) Effects of pH, temperature and salinity on immune parameters of the freshwater prawn *Macrobrachium rosenbergii*. *Fish and Shellfish Immunology* **10**, 387-391.

Cheng W., Juang F.-M., Li J.-T., Lin M.-C., Liu C.-H. and Chen J.-C. (2003) The immune response of the giant freshwater prawn *Macrobrachium rosenbergii* and its susceptibility to *Lactococcus garvieae* in relation to the moult stage. *Aquaculture* **218**, 33-45.

Cheng W., Liu C. H., Hsu J. P. and Chen J. C. (2002) Effect of hypoxia on the immune response of giant freshwater prawn *Machrobrachium rosenbergii* and its susceptibility to pathogen Enterococcus. *Fish and Shellfish Immunology* **13**, 351-365.

Cherbas L. and Cherbas P. (1993) The arthropod initiator: the capsite consensus plays an important role in transcription. *Insect Biochemistry and Molecular Biology* **23**, 81-90.

Chisholm J. R. S. (1993) Antibacterial activity in the blood cells on *Carcinus maenas* (L.) and other crustaceans. In *Department of Biology*. University of St. Andrews, St. Andrews.

Chisholm J. R. S. and Smith V. J. (1992) Antibacterial Activity in the Haemocytes of the Shore Crab, *Carcinus maenas*. *Journal of the Marine Biological Association of the United Kingdom* **72**, 529-542.

Chisholm J. R. S. and Smith V. J. (1995) Comparison of antibacterial activity in the haemocytes of different crustacean species. *Comparative Biochemistry and Physiology* **110a**, 39-45.

Chisholm J. R. S. and Smith V. J. S. (1994) Variation of Antibacterial Activity in the Haemocytes of the Shore Crab, *Carcinus maenas*, with temperature. *Journal of the Marine Biological Association of the United Kingdom* **74**, 979-982.

Cho W. L., Fu T. F., Chiou J. Y. and Chen C. C. (1997) Molecular characterization of a defensin gene from the mosquito, *Aedes aegypti*. *Insect Biochemistry and Molecular Biology* **27**, 351-358.

Choi K. C., Kim H. R., Park Y. S., Park S. M. and Kim J. H. (2002) Design and screening of in vivo expressed antimicrobial peptide library. *Biotechnology Letters* **24**, 251-256.

Choi S. M., Lee O. S., Kwon S. Y., Kwak S. S., Yu D. Y. and Lee H. S. (2003) High expression of a human lactoferrin in transgenic tobacco cell cultures. *Biotechnology Letters* **25**, 213-218.

Chomczynski P. and Sacchi N. (1987) Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. *Analytical Biochemistry* **162**, 156-159.

Chung C. T., Niemela S. L. and Miller R. H. (1989) One-step preparation of competent *Escherichia coli*: transformation and storage of bacterial cells in the same solution. *Proceedings of the National Academy of Sciences of the U S A* **86**, 2172-2175.

Cloete T. E. (2003) Resistance mechanisms of bacteria to antimicrobial compounds. *International Biodeterioration & Biodegradation* **51**, 277-282.

Cociancich S., Goyffon, M., Bontems, F., Bulet, P., Bouet, F., Menez, A. and Hoffmann, J. (1993) Purification and Characterization of a Scorpion Defensin, a 4kDa Antibacterial Peptide Presenting Structural Similarities with Insect Defensins and Scorpion Toxins. *Biochemical and Biophysical Research Communications* **194**, 17-22.

Colwell R. R., Tubiash H. S. and Wicks T. C. (1975) Comparative study of the bacterial flora of the hemolymph of *Callinecties sapidus*. *Marine Fisheries Review* **37**, 29-33.

Coman G. J., Crocos P. J., Preston N. P. and Fielder D. (2002) The effects of temperature on the growth, survival and biomass of different families of juvenile *Penaeus japonicus* Bate. *Aquaculture* **214**, 185-199.

Constans A. (2002) Protein Purification II: Affinity Tags. In *The Scientist*, Vol. 16.

Cooper E. L. (1985) Overview of humoral factors in invertebrates. *Developmental and Comparative Immunology* **9**, 577-583.

Cooper E. L. (1996) Diversity of Immune Systems. *Italian Journal of Zoology* **63**, 295-301.

Cooper E. L., Rinkevich B., Uhlenbruck G. and Valenbois P. (1992) Invertebrate Immunity: Another Viewpoint. *Scandinavian Journal of Immunology* **35**, 247-266.

Cornet B., Bonmatin J. M., Hetru C., Hoffmann J. A., Ptak M. and Vovelle F. (1995) Refined three-dimensional solution structure of insect defensin A. *Structure* **3**, 435-448.

Cornick J. W. and Stewart E. J. (1968) Interactions of the Pathogen *Gaffkya homari* with Natural Defense Mechanisms of *Homarus americanus*. *Journal of the Fisheries Research Board of Canada* **25**, 695-709.

Crothers J. H. (1966) The Biology of the Shore Crab *Carcinus maenas* (L.). *Field Studies* **2**, 407-434.

Cuculescu M., Hyde D. and Bowler K. (1995) Temperature-Acclimation of Marine Crabs - Changes in Plasma Membrane Fluidity and Lipid Composition. *Journal of Thermal Biology* **20**, 207-222.

Cudic M., Condie B. A., Weiner D. J., Lysenko E. S., Xiang Z. Q., Insug O., Bulet P. and Otvos J. L. (2002) Development of novel antibacterial peptides that kill resistant isolates. *Peptides* **23**, 2071-2083.

Cunliffe R. N. (2003) alpha-Defensins in the gastrointestinal tract. *Molecular Immunology* **40**, 463-467.

Cuthbertson B. J., Bullesbach E. E., Fievet J., Bachère E. and Gross P. S. (2004) A new class (penaeidin class 4) of antimicrobial peptides from the Atlantic white shrimp (*Litopenaeus setiferus*) exhibits target specificity and an independent proline-rich-domain function. *Biochemical Journal* **381**, 79-86.

Cuthbertson B. J., Shepard E. F., Chapman R. W. and Gross P. S. (2002) Diversity of the penaeidin antimicrobial peptides in two shrimp species. *Immunogenetics* **54**, 442-445.

Dacoz C. D. (1994) Seasonal Reproductive Activities of Different Size Classes of a Population of Green Shore Crabs, *Carcinus maenas* (Linnaeus, 1758) in the Southern Area of the North-Sea. *Cahiers De Biologie Marine* **35**, 1-13.

Daibo S., Kimura M. T. and Goto S. G. (2001) Upregulation of genes belonging to the drosomycin family in diapausing adults of *Drosophila triauraria*. *Gene* **278**, 177-184.

Dathe M., Nikolenko H., Meyer J., Beyermann M. and Bienert M. (2001) Optimization of the antimicrobial activity of magainin peptides by modification of charge. *Febs Letters* **501**, 146-150.

Dawirs R. R. (1984) Influence of Starvation On Larval Development of *Carcinus maenas* L (Decapoda, Portunidae). *Journal of Experimental Marine Biology and Ecology* **80**, 47-66.

Dawirs R. R. (1985) Temperature and Larval Development of *Carcinus maenas* (Decapoda) in the Laboratory - Predictions of Larval Dynamics in the Sea. *Marine Ecology Progress Series* **24**, 297-302.

Dawirs R. R. and Dietrich A. (1986) Temperature and Laboratory Feeding Rates in *Carcinus maenas* L (Decapoda, Portunidae) Larvae From Hatching Through Metamorphosis. *Journal of Experimental Marine Biology and Ecology* **99**, 133-147.

De Gregorio E., Spellman P. T., Tzou P., Rubin G. M. and Lemaitre B. (2002) The Toll and Imd pathways are the major regulators of the immune response in Drosophila. *EMBO Journal* **21**, 2568-2579.

Dean J. M. and Vernberg F. J. (1966) Hypothermia and the Blood of Crabs. *Comparative Biochemistry and Physiology B-Biochemistry & Molecular Biology* **17**, 19-22.

Depledge M. H. (1984) Disruption of Circulatory and Respiratory Activity in Shore Crabs (*Carcinus maenas* (L)) Exposed to Heavy-Metal Pollution. *Comparative Biochemistry and Physiology C-Pharmacology Toxicology & Endocrinology* **78**, 445-459.

Destoumieux D., Bulet P., Loew D., Van Dorsselaer A., Rodriguez J. and Bachère E. (1997) Penaeidins, a new family of antimicrobial peptides isolated from the shrimp *Penaeus vannamei* (Decapoda). *Journal of Biological Chemistry* **272**, 28398-28406.

Destoumieux D., Bulet P., Strub J. M., Van Dorsselaer A. and Bachère E. (1999) Recombinant expression and range of activity of penaeidins, antimicrobial peptides from penaeid shrimp. *European Journal of Biochemistry* **266**, 335-346.

Destoumieux D., Munoz M., Cosseau C., Rodriguez J., Bulet P., Comps M. and Bachère E. (2000) Penaeidins, antimicrobial peptides with chitin-binding activity, are produced and stored in shrimp granulocytes and released after microbial challenge. *Journal of Cell Science* **113**, 461-469.

Destoumieux-Garzon D., Saulnier D., Garnier J., Jouffrey C., Bulet P. and Bachère E. (2001) Crustacean immunity. Antifungal peptides are generated from the C terminus of shrimp hemocyanin in response to microbial challenge. *Journal of Biological Chemistry* **276**, 47070-47077.

Deutscher M. P. (1990) *Guide to Protein Purification*. Academic Press, London.

Devine D. A. (2003) Antimicrobial peptides in defence of the oral and respiratory tracts. *Molecular Immunology* **40**, 431-443.

Dhaliwal W., Bajaj-Elliott M. and Kelly P. (2003) Intestinal defensin gene expression in human populations. *Molecular Immunology* **40**, 469-475.

Dhar A. K., Roux M. M. and Klimpel K. R. (2001) Detection and quantification of infectious hypodermal and hematopoietic necrosis virus and white spot virus in shrimp using real-time quantitative PCR and SYBR Green chemistry. *Journal of Clinical Microbiology* **39**, 2835-2845.

Dian C., Barath P., Knaust R., McSweeney S., Moss E., Hristova M., Ambros V. and Birse D. (2002a) Overcoming protein instability problems during fusion protein cleavage. In *Life Science News*, Vol. 10. Amersham Bioscience.

Dian C., Eshaghi S., Urbig T., McSweeney S., Heijbel A., Salbert G. and Birse D. (2002b) Strategies for the purification and on-column cleavage of glutathione-S-transferase fusion target proteins. *Journal of Chromatography B-Analytical Technologies in the Biomedical and Life Sciences* **769**, 133-144.

Dimarcq J. L., Bulet P., Hetru C. and Hoffmann J. (1998) Cysteine-rich antimicrobial peptides in invertebrates. *Biopolymers* **47**, 465-477.

Donald K. M., Day A. J., Smerdon G. R., Cross L. J. and Hawkins A. J. (2003) Quantification of gene transcription and enzyme activity for functionally important proteolytic enzymes during early development in the Pacific oyster *Crassostrea gigas*. *Comparative Biochemistry and Physiology B-Biochemistry & Molecular Biology* **136**, 383-392.

Dorak M. T. (2003) Real-Time PCR.

Ehret-Sabatier L., Loew D., Goyffon M., Fehlbaum P., Hoffmann J. A., van Dorsselaer A. and Bulet P. (1996) Characterization of novel cysteine-rich antimicrobial peptides from scorpion blood. *Journal Biological Chemistry* **271**, 29537-29544.

El Haj A. J. and Harrison P. (1994) Regulation of Actin Expression During Muscle Growth Over the Molt Cycle in Lobsters. *Journal of Cellular Biochemistry*, 496-496.

El Haj A. J. and Houlihan D. F. (1987) *In vitro* and *in vivo* protein synthesis raters in a crustacean muscle during the moult cycle. *Journal of Experimental Biology* **127**, 413-426.

Elner R. W. (1981) Diet of Green Crab *Carcinus maenas* (L.) from Port Herbert, Southwetern Nova Scotia. *Journal of Shellfish Research* **1**, 89-94.

Elsbach P. (1990) Antibiotics from within: antibacterials from human and animal sources. *Tibtech* **8**, 26-30.

Engstrom Y. (1999) Induction and regulation of antimicrobial peptides in Drosophila. *Developmental and Comparative Immunology* **23**, 345-358.

Epand R. M. and Vogel H. J. (1999) Diversity of antimicrobial peptides and their mechanisms of action. *Biochimica Et Biophysica Acta-Biomembranes* **1462**, 11-28.

Evans E. E., Cushing J. E., Sawyer S., Weinheimer P. F., Acton R. T. and McNeely J. L. (1969) Induced bactericidal response in the California spiny lobster *Panulirus interruptus*. *Proceedings of the Society of Experimental and Biological Medicine* **132**, 111-114.

Evans E. E., Painter B., Evans M. L., Weinheimer P. and Acton R. T. (1968) An induced bactericidin in the spiny lobster, *Panulirus argus*. *Proceedings of the Society of Experimental and Biological Medicine* **128**, 394-398.

FAO. (1998) FAOSTAT Fisheries Data, Vol. 1998. FAO.

FAO. (2002) The State of the World Fisheries and Aquaculture. FAO, Sofia.

Farfante P. and Kensley B. (1997) *Penaeoid and Sergestoid Shrimps and Prawns of the World: Keys and Diagnoses for the Families and Genera*. Muséum national d'Histoire naturelle, Paris, France.

Fariselli P., Riccobelli P. and Casadio R. (1999) Role of evolutionary information in predicting the disulfide-bonding state of cysteine in proteins. *Proteins* **36**, 340-346.

Farnaud S. and Evans R. W. (2003) Lactoferrin-a multifunctional protein with antimicrobial properties. *Molecular Immunology* **40**, 395-405.

Farrell R. E. (1996) *RNA Methodologies*. Academic Press, London.

Faye I., Pye A., Rasmuson T., Boman H. G. and Boman I. A. (1975) Insect immunity. 11. Simultaneous induction of antibacterial activity and selection synthesis of some hemolymph proteins in diapausing pupae of *Hyalophora cecropia* and *Samia cynthia*. *Infection and Immunity* **12**, 1426-1438.

Florack D., Allefs S., Bollen R., Bosch D., Visser B. and Stiekema W. (1995) Expression of giant silkmoth cecropin B genes in tobacco. *Transgenic Research* **4**, 132-141.

Fowler J. and Cohen L. (1996) *Practical Statistics for Field Biology*. John Wiley & Sons Ltd., Chichester.

Frangioni J. V. and Neel B. G. (1993) Solubilization and purification of enzymatically active glutathione S-transferase (pGEX) fusion proteins. *Analytical Biochemistry* **210**, 179-187.

Freshney R. I. (2000) *Culture of Animal Cells - A Manual of Basic Techniques*. A John Wiley Sons Inc., Publication.

Frost P. and Nilsen F. (2003) Validation of reference genes for transcription profiling in the salmon louse, *Lepeophtheirus salmonis*, by quantitative real-time PCR. *Veterinary Parasitology* **118**, 169-174.

Froy O. and Gurevitz M. (2003) Arthropod and mollusk defensins--evolution by exon-shuffling. *Trends in Genetics* **19**, 684-687.

Galperin M. Y. (2004) The Molecular Biology Database Collection: 2004 update. *Nucleic Acids Research* **32**.

Ganz T. (2003) Defensins: Antimicrobial Peptides of Innate Immunity. *Nature Reviews Immunology* **3**, 710-720.

Ganz T. and Lehrer R. I. (1994) Defensins. *Current Opinion in Immunology* **6**, 584-589.

Ganz T., Selsted M. E., Szklarek D. and al. e. (1985) Defensins: natural peptide antibiotics of human neutrophils. *Journal of Clinical Investigation* **76**, 1427-1435.

Garnier J., Gibrat J. F. and Robson B. (1996) GOR method for predicting protein secondary structure from amino acid sequence. *Methods in Enzymology* **266**, 540-553.

Ghiretti-Magaldi A., Milanesi C. and Tognon G. (1977) Hemopoiesis in Crustacea Decapoda: Origin and Evolution of Haemocytes and Cyanocytes of *Carcinus maenas*. *Cell Differentiation*.

Giacometti A., Cirioni O., Barchiesi F., Del Prete M. S. and Scalise G. (1999) Antimicrobial activity of polycationic peptides. *Peptides* **20**, 1265-1273.

Goidin D., Mamessier A., Staquet M. J., Schmitt D. and Berthier-Vergnes O. (2001) Ribosomal 18S RNA prevails over glyceraldehyde-3-phosphate dehydrogenase and beta-actin genes as internal standard for quantitative comparison of mRNA levels in invasive and noninvasive human melanoma cell subpopulations. *Analytical Biochemistry* **295**, 17-21.

Goldspink G. (1995) Adaptation of Fish to Different Environmental-Temperature By Qualitative and Quantitative Changes in Gene-Expression. *Journal of Thermal Biology* **20**, 167-174.

Gomez-Jimenez S., Uglow R. F. and Gollas-Galvan T. (2000) The effects of cooling and emersion on total haemocyte count and phenoloxidase activity of the spiny lobster *Panulirus interruptus*. *Fish and Shellfish Immunology* **10**, 631-635.

Grandi G. (2003) Rational antibacterial vaccine design through genomic technologies. *International Journal for Parasitology* **33**, 615-620.

Gross P. S., Bartlett T. C., Browdy C. L., Chapman R. W. and Warr G. W. (2001) Immune gene discovery by expressed sequence tag analysis of hemocytes and hepatopancreas in the Pacific White Shrimp, *Litopenaeus vannamei*, and the Atlantic White Shrimp, *L. setiferus*. *Developmental and Comparative Immunology* **25**, 565-577.

Guruprasad K., Reddy B. V. and Pandit M. W. (1990) Correlation between stability of a protein and its dipeptide composition: a novel approach for predicting *in vivo* stability of a protein from its primary sequence. *Protein Engineering* **4**, 15-161.

Guzman A., Ascencio V. F. and Ascencio G. (2000) Infectious disease in shrimp species with aquaculture potential. *Recent Research Devlelopments in Microbiology* **4**, 333-348.

Hamann A. (1975) Stress-Induced Changes in Cell Titre of Crayfish Hemolymph. *Zeitschraft fur Naturforschung* **30**, 850.

Hammond J. A. (2002) Ancestral Genes of the Complement System in the Ascidian, *Ciona intestinalis*. In *Department of Biology*. St. Andrews University, St. Andrews.

Hammond J. A. and Smith V. J. (2002) Lipopolysaccharide induces DNA-synthesis in a sub-population of haemocytes from the swimming crab, *Liocarcinus depurator*. *Developmental and Comparative Immunology* **26**, 227-236.

Hancock R. E. W. and Chapple D. S. (1999) Peptide antibiotics. *Antimicrobial Agents and Chemotherapy* **43**, 1317-1323.

Hancock R. E. W. and Diamond G. (2000) The role of cationic antimicrobial peptides in innate host defences. *Trends in Microbiology* **8**, 402-410.

Hancock R. E. W. and Lehrer R. (1998) Cationic peptides:a new source of antibiotics. *Tibtech* **16**, 82-88.

Hancock R. E. W. and Scott M. G. (2000) The role of antimicrobial peptides in animal defenses. *Proceedings of the National Academy of Sciences of the U S A* **97**, 8856-8861.

Hara S. and Yamakawa M. (1996) Production in *Escherichia coli* of moricin, a novel type antibacterial peptide from the silkworm *Bombyx mori*. *Biochemical and Biophysical Research Communications* **220**, 664-669.

Harder J., Bartels J., Christophers E. and Schroder J. M. (1997) A peptide antibiotic from human skin. *Nature* **387**, 861.

Harder J. and Schroder J. M. (2002) RNase 7, a novel innate immune defense antimicrobial protein of healthy human skin. *Journal of Biological Chemistry* **277**, 46779-46784.

Harms J., Meyerharms B., Dawirs R. R. and Anger K. (1994) Growth and Physiology of *Carcinus maenas* (Decapoda, Portunidae) Larvae in the Field and in Laboratory Experiments. *Marine Ecology Progress Series* **108**, 107-118.

Harrison P. and El Haj A. J. (1994) Actin mRNA levels and myofibrillar growth in leg muscles of the European lobster (*Hommarus gammarus*) in response to passive stretch. *Molecular Marine Biology and Biotechnology* **3**, 35-41.

Hauton C. (1995) An Investigation of Potential Immunological and Metabolic Indices of Environmental Water Quality in the Shore Crab, *Carcinus maenas*. In *Department of Oceanography*, p. 271. Southampton University, Southampton.

Hauton C., Hammond J. A. and Smith J. V. (2005) Real-time PCR quantification of the *in vitro* effects of crustacean immunostimulants on gene expression in lobster (*Homarus gammarus*) granular haemocytes. *Developmental and Comparative Immunology* **29**, 33-42.

Hauton C., Williams J. A. and Hawkins L. E. (1997) The effects of a live in vivo pathogenic infection on aspects of the immunocompetence of the common shore crab, *Carcinus maenas* (L). *Journal of Experimental Marine Biology and Ecology* **211**, 115-128.

Heath J. R. and Barnes H. (1970) Some Changes in Biochemical Composition with Season and During The Moulting Cycle of the Common Shore Crab, *Carcinus maenas* (L.). *Journal of Experimental Marine Biology and Ecology* **5**, 199-233.

Hebel D. K., Jones M. B. and Depledge M. H. (1997) Responses of Crustaceans to Contaminant Exposure: a Holistic Approach. *Estuarine, Coastal and Shelf Science* **44**, 177-184.

Hellers M., Gunne H. and Steiner H. (1991) Expression of post-translational processing of preprocecropin A using a baculovirus vector. *European Journal of Biochemistry* **199**, 435-439.

Henikoff S. and Henikoff J. G. (1992) Amino acid substitution matrices from protein blocks. *Proceedings of the National Academy of Sciences of the U S A* **89**, 10915-10919.

Henke G. A. (1985) Variations of Hemolymph-Proteins of the Common Shore Crab *Carcinus maenas* after a Bacterial-Infection. *Marine Biology* **87**, 181-183.

Hennighausen L. G. and Sippel A. E. (1982) Mouse whey acidic protein is a novel member of the family of "four-disulphide core" proteins. *Nucleic Acids Research* **10**, 2677-2684.

Hernroth B. (2003) The influence of temperature and dose on antibacterial peptide response against lipopolysaccharide in the blue mussel, *Mytilus edulis*. *Fish and Shellfish Immunology* **14**, 25-37.

Hetru C., Hoffmann D. and Bulet P. (1998) Antimicrobial peptides from insects. In *Molecular Mechanisms of Immune Resposnses in Insects* (Edited by Brey P. T. and Hultmark D.), p. 40-66. Chapman and Hall, London.

Hill A. D., Taylor A. C. and Strang R. H. C. (1991) Physiological and Metabolic Responses of the Shore Crab *Carcinus maenas* (L) During Environmental Anoxia and Subsequent Recovery. *Journal of Experimental Marine Biology and Ecology* **150**, 31-50.

Hose J. E., Martin G. G., Tiu S. and McKrell N. (1992) Patterns of Hemocyte Production and Release Throughout the Molt Cycle in the Penaeid Shrimp *Sicyonia ingentis*. *Biological Bulletin* **183**, 185-199.

Houlihan D. F., Waring C. P., Mathers E. and Gray C. (1990) Protein-Synthesis and Oxygen-Consumption of the Shore Crab *Carcinus maenas* After a Meal. *Physiological Zoology* **63**, 735-756.

Hubert F., Noel T. and Roch P. (1996) A member of the arthropod defensin family from edible Mediterranean mussels (*Mytilus galloprovincialis). European Journal of Biochemistry* **240**, 815-815.

Irving P., Troxler L., Heuer T. S., Belvin M., Kopczynski C., Reichhart J. M., Hoffmann J. A. and Hetru C. (2001) A genome-wide analysis of immune responses in Drosophila. *Proceedings of the National Academy of Sciences of the U S A* **98**, 15119-15124.

Ivanov I., Alexandrova R., Dragulev B., Saraffova A. and AbouHaidar M. G. (1992) Effect of tandemly repeated AGG triplets on the translation of CAT-mRNA in E. coli. *Febs Letters* **307**, 173-176.

Iwanaga S. (2002) The molecular basis of innate immunity in the horseshoe crab. *Current Opinion in Immunology* **14**, 87-95.

Iwanaga S., Muta T., Shigenaga T., Miura Y., Seki N., Saito T. and Kawabata S. (1994) Role of hemocyte-derived granular components in invertebrate defense. *Annals of the New York Academy of Sciences* **712**, 102-116.

James G. T. (1978) Inactivation of the protease inhibitor phenylmethylsulfonyl fluoride in buffers. *Analytical Biochemistry* **86**, 574-579.

Jiravanichpaisal P., Söderhäll K. and Söderhäll I. (2004) Effect of water temperature on the immune response and infectivity pattern of white spot syndrome virus (WSSV) in freshwater crayfish. *Fish and Shellfish Immunology* **17**, 265-275.

Johansson J., Gudmundsson G. H., Rottenberg M. E., Berndt K. D. and Agerberth B. (1998) Conformation-dependent antibacterial activity of the naturally occurring human peptide LL-37. *Journal of Biological Chemistry* **273**, 3718-3724.

Johansson M. W., Keyser P., Sritunyalucksana K. and Söderhäll K. (2000) Crustacean Haemocytes and Haematopoiesis. *Aquaculture* **191**, 45-52.

Johansson M. W. and Söderhäll K. (1985) Exocytosis of the prophenoloxidase activating system from crayfish haemocytes. *Journal of Comparative Physiology* **156**, 175-181.

Johnson I. and Uglow R. F. (1985) Some Effects of Aerial Exposure On the Respiratory Physiology and Blood Chemistry of *Carcinus maenas* (L) and *Liocarcinus puber* (L). *Journal of Experimental Marine Biology and Ecology* **94**, 151-165.

Johnson P. T. (1981) Histopathology of *Aerococcus viridans* var. *homari* Infection (Gaffkemia) in the Lobster, *Homarus americanus*, and a Comparison with Histological Reactions to a Gram-Negative Species, *Pseudomonas perolens*. *Journal of Invertebrate Pathology* **38**, 127-148.

Johnson P. T. (1983) Diagnosis of crustacean diseases. *Rapports et procès-verbaux des réunions Commission internationale pour l'exploration scientifique de la Mer Méditerranée* **182**, 54-57.

Johnson P. T. (1987) A Review of Fixed Phagocytic and Pinocytic Cells of Decapod Crustaceans, with Remarks on Hemocytes. *Developmental and Comparative Immunology* **11**, 679-704.

Johnston M. A., Elder H. Y. and Davies P. S. (1973) Cytology of Carcinus Haemocytes and their Function in Carbohydrate Metabolism. *Comparative Biochemistry and Physiology* **46**, 569-581.

Kamysz W., Okroj M. and Lukasiak J. (2003) Novel properties of antimicrobial peptides. *Acta Biochim Pol* **50**, 461-469.

Kanost M. R. (1999) Serine proteinase inhibitors in arthropod immunity. *Developmental and Comparative Immunology* **23**, 291-301.

Kausel G., Vera M. I., San Martin R., Figueroa J., Molina A., Muller M., Martial J. and Krauskopf M. (1999) Transcription factor Pit-1 expression is modulated upon seasonal acclimatization of eurythermal ectotherms: Identification of two Pit-1 genes in the carp. *Journal of Cellular Biochemistry* **75**, 598-609.

Kawabata S., Nagayama R., Hirata M., Shigenaga T., Agarwala K. L., Saito T., Cho J., Nakajima H., Takagi T. and Iwanaga S. (1996) Tachycitin, a small granular component in horseshoe crab hemocytes, is an antimicrobial protein with chitin-binding activity. *Journal of Biochemistry (Tokyo)* **120**, 1253-1260.

Kawano K., Yoneya T., Miyata T., Yoshikawa K., Tokunaga F., Terada Y. and Iwanaga S. (1990) Antimicrobial peptide, tachyplesin I, isolated from hemocytes of the horseshoe crab (*Tachypleus tridentatus*). NMR determination of the beta-sheet structure. *Journal of Biological Chemistry* **265**, 15365-15367.

Kellogg D. E., Rybalkin I., Chen S., Mukhamedova N., Vlasik T., Siebert P. D. and Chenchik A. (1994) TaqStart Antibody: "hot start" PCR facilitated by a neutralizing monoclonal antibody directed against Taq DNA polymerase. *Biotechniques* **16**, 1134-1137.

Khoo L., Riobinette D. W. and Noga E. J. (1999) Callinectin, an antibacterial peptide from the blue crab, *Callinectes sapidus,* hemocytes. *Marine Biotechnology* **1**, 44-51.

Kim S. H., Park B. S., Yun E. Y., Je Y. H., Woo S. D., Kang S. W., Kim K. Y. and Kang S. K. (1998) Cloning and expression of a novel gene encoding a new antibacterial peptide from silkworm, *Bombyx mori. Biochemical and Biophysical Research Communications* **246**, 388-392.

Koczulla A. R. and Bals R. (2003) Antimicrobial peptides: current status and therapeutic potential. *Drugs* **63**, 389-406.

Koczulla R., von Degenfeld G., Kupatt C., Krotz F., Zahler S., Gloe T., Issbrucker K., Unterberger P., Zaiou M., Lebherz C., Karl A., Raake P., Pfosser A., Boekstegers P., Welsch U., Hiemstra P. S., Vogelmeier C., Gallo R. L., Clauss M. and Bals R. (2003) An angiogenic role for the human peptide antibiotic LL-37/hCAP-18. *Journal of Clinical Investigation* **111**, 1665-1672.

Konno K., Hisada M., Fontana R., Lorenzi C. C., Naoki H., Itagaki Y., Miwa A., Kawai N., Nakata Y., Yasuhara T., Ruggiero Neto J., de Azevedo W. F., Jr., Palma M. S. and Nakajima T. (2001) Anoplin, a novel antimicrobial peptide from the venom of the solitary wasp *Anoplius samariensis. Biochimica Et Biophysica Acta-General Subjects* **1550**, 70-80.

Kotlyar S., Weihrauch D., Paulsen R. S. and Towle D. W. (2000) Expression of arginine kinase enzymatic activity and mRNA in gills of the euryhaline crabs *Carcinus maenas* and *Callinectes sapidus. Journal of Experimental Biology* **203**, 2395-2404.

Laemmli U. K. (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227**, 680-685.

Le Moullac G. and Haffner P. (2000) Environmental factors affecting immune responses in Crustacea. *Aquaculture* **191**, 121-131.

Lee K. J., Watson R. D. and Roer R. D. (1998) Molt-inhibiting hormone mRNA levels and ecdysteroid titer during a molt cycle of the blue crab, *Callinectes sapidus*. *Biochemical and Biophysical Research Communications* **249**, 624-627.

Lee S. Y. and Söderhäll K. (2002) Early events in crustacean innate immunity. *Fish and Shellfish Immunology* **12**, 421-437.

Lehrer R. I. and Ganz T. (1999) Antimicrobial peptides in mammalian and insect host defence. *Current Opinion in Immunology* **11**, 23-27.

Lehrer R. I., Rosenman M., Harwig S. S., Jackson R. and Eisenhauer P. (1991) Ultrasensitive assays for endogenous antimicrobial polypeptides. *Journal of Immunological Methods* **137**, 167-173.

Lepage P., Bitsch F., Roecklin D., Keppi E., Dimarcq J. L., Reichhart J. M., Hoffmann J. A., Roitsch C. and Vandorsselaer A. (1991) Determination of Disulfide Bridges in Natural and Recombinant Insect Defensin-A. *European Journal of Biochemistry* **196**, 735-742.

Levashina E. A., Ohresser S., Lemaitre B. and Imler J.-L. (1998) Two Distinct Pathways can Control Expression of the Gene Encoding the Drosophila Antimicrobial Peptide Metchnikowin. *Journal of Molecular Biology* **278**, 515-527.

Levy O. (2000) Antimicrobial proteins and peptides of blood: templates for novel antimicrobial agents. *Blood* **96**, 2664-2672.

Li W. K. W. and Dickie P. M. (1996) Distribution of bacteria in marine waters. Canadian Fisheries.

Li W. K. W., Dickie P. M., Harrison W. G. and Irwin B. D. (1993) Biomass and production of bacteria and phytoplankton during the spring bloom in the western North Atlantic Ocean. *Deep Sea Research Part II: Topical Studies in Oceanography* **40**, 307-327.

Linn S. (1990) Strategies and Considerations for Protein Purification. In *Guide to Protein Purification* (Edited by Deutscher M. P.), Vol 182, p. 9-15. Academic Press, London.

Lodish H., Baltimore D., Berk A., Zipursky S. L., Matsudaira P. and Darnell J. (1995) *Molecular Cell Biology*. Scientific American Books, New York.

Lorenzon S., de Guarrini S., Smith V. J. and Ferrero E. A. (1999) Effects of LPS injection on circulating haemocytes in crustaceans *in vivo*. *Fish and Shellfish Immunology* **9**, 31-50.

Lucentini J. (2003) Antibiotics Arms Race Heats Up. In *The Scientist*.

Luna C., Wang X. L., Huang Y. M., Zhang J. A. and Zheng L. B. (2002) Characterization of four Toll related genes during development and immune responses in *Anopheles gambiae*. *Insect Biochemistry and Molecular Biology* **32**, 1171-1179.

Lund O., Nielsen M., Lundegaard C. and Worning P. (2002) X3M a Coputer Program to Extract 3D Models.

Lundebye A. K. and Depledge M. H. (1998) Molecular and physiological responses in shore crabs *Carcinus maenas* following exposure to copper. *Marine Environmental Research* **46**, 567-572.

Lyndyard P. M., Whelan A. and Fanger M. W. (2000) *Instant Notes in Immunology*. BIOS Scientific Publishing Ltd.

MAFF. (1998) Sea Fisheries Statistics, p. 29-34. Ministry of Agiculture , Food and Fisheries.

Marshall S. H. and Arenas G. (2003) Antimicrobial peptides: A natural alternative to chemical antibiotics and a potential for applied biotechnology. *Electronic Journal of Biotechnology* **6**.

Martemyanov K. A., Spirin A. S. and Gudkov A. T. (1997) Direct expression of PCR products in a cell-free transcription/translation system: synthesis of antibacterial peptide cecropin. *Febs Letters* **414**, 268-270.

Martin G. G., Poole D., Poole C., Hose J. E., Arias M., Reynolds L., McKrell N. and Whang A. (1993) Clearance of Bacteria Injected into the Hemolymph of the Penaeid Shrimp, *Sicyonia ingentis*. *Journal of Invertebrate Pathology* **62**, 308-315.

Mathieson S. and Berry A. J. (1997) Spatial, temporal and tidal variation in crab populations in the Forth Estuary, Scotland. *Journal of the Marine Biological Association of the United Kingdom* **77**, 167-183.

Maxwell A. I., Morrison G. M. and Dorin J. R. (2003) Rapid sequence divergence in mammalian beta-defensins by adaptive evolution. *Molecular Immunology* **40**, 413-421.

McCumber L. J. and Clem L. W. (1983) Recognition of Non-Self in Crustaceans. *American Zoologist* **23**, 173-183.

McKay D. and Jenkin C. R. (1969) Immuity in the Invertebrates: II Adaptive Immunity in the Crayfish (*Parachaeraps bicarinatus*). *Immunology* **17**, 127-137.

McPherson M. J. and Møller S. G. (2000) *PCR*. Bios Scientific Publishers Ltd., New York.

Medzhitov R. (2001) Toll-like receptors and innate immunity. *Nature Reviews in Immunology* **1**, 135-145.

Medzhitov R. and Janeway C. A. (1998) An ancient system of host defense. *Current Opinion in Immunology* **10**, 12-15.

Mercado-Pimentel M. E., Jordan N. C. and Aisemberg G. O. (2002) Affinity purification of GST fusion proteins for immunohistochemical studies of gene expression. *Protein Expression and Purification* **26**, 260-265.

Mialhe E., Bachère E., Boulo V., Cadoret J. P., Rousseau C., Cedeno V., Saraiva E., Carrera L., Calderon J. and Colwell R. R. (1995) Future of biotechnology-based control of disease in marine invertebrates. *Molecular Marine Biology and Biotechnology* **4**, 275-283.

Michaut L., Fehlbaum P., Moniatte M., Van Dorsselaer A., Reichhart J. M. and Bulet P. (1996a) Determination of the disulfide array of the first inducible antifungal peptide from insects: drosomycin from *Drosophila melanogaster*. *Febs Letters* **395**, 6-10.

Michaut L., Fehlbaum P., Moniatte M., VanDorsselaer A., Reichhart J. M. and Bulet P. (1996b) Determination of the disulfide array of the first inducible antifungal peptide from insects: Drosomycin from Drosophila melanogaster. *Febs Letters* **395**, 6-10.

Millar D. A. and Ratcliffe N. A. (1994) Invertebrates. In *Immunology: A Comparative Approach* (Edited by Turner R. J.), p. 29-68. John Wiley & Sons Ltd.

Mitta G., Hubert F., Dyrynda E. A., Boudry P. and Roch P. (2000a) Mytilin B and MGD2, two antimicrobial peptides of marine mussels: gene structure and expression analysis. *Developmental and Comparative Immunology* **24**, 381-393.

Mitta G., Vandenbulcke F. and Roch P. (2000b) Original involvement of antimicrobial peptides in mussel innate immunity. *Febs Letters* **486**, 185-190.

Miyata T., Tokunaga F., Yoneya T., Yoshikawa K., Iwanaga S., Niwa M., Takao T. and Shimonishi Y. (1989) Antimicrobial peptides, isolated from horseshoe crab hemocytes, tachyplesin II, and polyphemusins I and II: chemical structures and biological activity. *Journal of Biochemistry (Tokyo)* **106**, 663-668.

Moreira G. S. and Nelson S. G. (1990) The Effects of Temperature On the Respiratory Metabolism of *Calcinus laevimanus* (Randall, 1839) (Crustacea, Anomura), in Different Salinities. *Journal of Thermal Biology* **15**, 37-39.

Mori K. and Stewart J. E. (1978) Natural and Induced Bactericidal Activities of the Hepatopancreas of the Amercican Lobster, *Homarus americanus*. *Journal of Invertebrate Pathology* **32**, 171-176.

Morris S. and Airriess C. N. (1998) Integration of physiological responses of crustaceans to environmental challenge. *South African Journal of Zoology* **33**, 87-106.

Muhle S. A. and Tam J. P. (2001) Design of gram-negative selective antimicrobial peptides. *Biochemistry* **40**, 5777-5785.

Munoz M., Vandenbulcke F., Gueguen Y. and Bachère E. (2003) Expression of penaeidin antimicrobial peptides in early larval stages of the shrimp *Penaeus vannamei*. *Developmental and Comparative Immunology* **27**, 283-289.

Munoz M., Vandenbulcke F., Saulnier D. and Bachère E. (2002) Expression and distribution of penaeidin antimicrobial peptides are regulated by haemocyte reactions in microbial challenged shrimp. *European Journal of Biochemistry* **269**, 2678-2689.

Muramoto K. and Kamiya H. (1990) The position of the disulphide bonds and the glycosylation site in a lectin of the acorn barnacle *Megabalanus rosa*. *Biochimica et Biophysica Acta* **1039**, 52-60.

Nagai T., Osaki T. and Kawabata S. (2001) Functional Conversion of Hemocyanin to Phenoloxidase by Horseshoe Crab Antimicrobial Peptides. *Journal of Biological Chemistry* **276**, 27166-27170.

Naitza S. and Ligoxygakis P. (2004) Antimicrobial defences in Drosophila: the story so far. *Molecular Immunology* **40**, 887-896.

Nakajima Y., Saido-Sakanaka H., Taylor D. and Yamakawa M. (2003) Up-regulated humoral immune response in the soft tick, *Ornithodoros moubata* (Acari: Argasidae). *Parasitology Research* **91**, 476-481.

Nakamura T., Furunaka H., Miyata T., Tokunaga F., Muta T., Iwanaga S., Niwa M., Takao T. and Shimonishi Y. (1988) Tachyplesin, a class of antimicrobial peptide from the hemocytes of the horseshoe crab (*Tachypleus tridentatus*). Isolation and chemical structure. *Journal of Biological Chemistry* **263**, 16709-11613.

Natori S. (1990) Dual functions of insect immunity proteins in defence and development. *Research in Immunology* **141**, 938-939.

Neilsen H., Engelbrecht J., Brunak S. and von Henije G. (1997) Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. *Protein Engineering* **10**, 1-6.

New England Biolabs. (2004) *New England Biolabs Product Catalogue*. New England Biolabs Inc., Hertfordshire.

Nicolas P., Vanhoye D. and Amiche M. (2004) Molecular strategies in biological evolution of antimicrobial peptides. *Peptides* **24**, 1669-1680.

Nizet V., Ohtake T., Lauth X., Trowbridge J., Rudisill J., Dorschner R. A., Pestonjamasp V., Piraino J., Huttner K. and Gallo R. L. (2001) Innate antimicrobial peptide protects the skin from invasive bacterial infection. *Nature* **414**, 454-457.

Ochman H., Gerber A. S. and Hartl D. L. (1988) Genetic applications of an inverse polymerase chain reaction. *Genetics* **120**, 621-623.

Ouaissi A., Ouaissi M. and Sereno D. (2002) Glutathione S-transferases and related proteins from pathogenic human parasites behave as immunomodulatory factors. *Immunology Letters* **81**, 159-164.

Papo N. and Shai Y. (2004) Can we predict biological activity of antimicrobial peptides from their interactions with model phospholipid membranes? *Peptides* **24**, 1693-1703.

Perazzolo L. M., Gargioni R., Ogliari P. and Barracco M. A. A. (2002) Evaluation of some hemato-immunological parameters in the shrimp *Farfantepenaeus paulensis* submitted to environmental and physiological stress. *Aquaculture* **214**, 19-33.

Peters R. I. and Long G. E. (1973) Some Alterations in Hemolymph Proteins of *Hemigrapsus nudus* Associated with Temperature Acclimation. *Comparative Biochemistry and Physiology* **46**, 207-213.

Pfaffl M. W. (2001) A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Research* **29**.

Pfaffl M. W. (2004) Chapter 3: Quantification strategies in real-time PCR. In *A-Z of quantitative PCR* (Edited by Bustin S. A.). International University Line (IUL), La Jolla, CA, U.S.A.

Pfaffl M. W., Horgan G. W. and Dempfle L. (2002) Relative expression software tool (REST©) for group-wise comparison and statistical analysis of relative expression results in real-time PCR. *Nucleic Acids Research* **30**.

Philpott M. P. (2003) Defensins and acne. *Molecular Immunology* **40**, 457-462.

Piers K. L., Brown M. H. and Hancock R. E. (1993) Recombinant DNA procedures for producing small antimicrobial cationic peptides in bacteria. *Gene* **134**, 7-13.

Podrabsky J. E. and Somero G. N. (2004) Changes in gene expression associated with acclimation to constant temperatures and fluctuating daily temperatures in an annual killifish *Austrofundulus limnaeus*. *Journal of Experimental Biology* **207**, 2237-2254.

Pore N. and Pal S. (2000) Expression of the antibacterial peptide, cecropin, in cultured mammalian cells. *Biotechnology Letters* **22**, 151-155.

Porter E. M., vanDam E., Valore E. V. and Ganz T. (1997) Broad-spectrum antimicrobial activity of human intestinal defensin 5. *Infection and Immunity* **65**, 2396-2401.

Powers J.-P. S. and Hancock R. E. W. (2003) The relationship between peptide structure and antibacterial activity. *Peptides* **24**, 1681-1691.

Promega. (1996) *The Source for Discovery -Protocols and Applications Guide*, Southampton.

Raaijmakers M. H., van Emst L., de Witte T., Mensink E. and Raymakers R. A. (2002) Quantitative assessment of gene expression in highly purified hematopoietic cells using real-time reverse transcriptase polymerase chain reaction. *Experimental Hematology* **30**, 481-487.

Rabin H. (1970) Hemocytes, Hemolymph, and Defence Reactions in Crustaceans. *Journal of Reticuloendothelial Scoiety* **7**, 195-207.

Radonic A., Thulke S., Mackay I. M., Landt O., Siegert W. and Nitsche A. (2004) Guideline to reference gene selection for quantitative real-time PCR. *Biochemical and Biophysical Research Communications* **313**, 856-862.

Raj P. A., Antonyraj K. J. and Karunakaran T. (2000a) Large-scale synthesis and functional elements for the antimicrobial activity of defensins. *Biochemistry Journal* **347 Pt 3**, 633-641.

Raj P. A. and Dentino A. R. (2002) Current status of defensins and their role in innate and adaptive immunity. *FEMS Microbiology Letters* **206**, 9-18.

Raj P. A., Karunakaran T. and Sukumaran D. K. (2000b) Synthesis, microbicidal activity, and solution structure of the dodecapeptide from bovine neutrophils. *Biopolymers* **53**, 281-292.

Ranganathan S., Simpson K. J., Shaw D. C. and Nicholas K. R. (1999) The whey acidic protein family: A new signature motif and three-dimensional structure by comparative modeling. *Journal of Molecular Graphics & Modelling* **17**, 106.

Rasmussen R. (2001) *Quantification on the LightCycler*. Springer Press, Heidleberg.

Ratcliffe N. A. (1985) Invertebrate Immunity - A Primer for the Non-Specialist. *Immunology Letters* **10**, 253-270.

Ratcliffe N. A. (1989) The Biological Significance of Immunity. *Developmental and Comparative Immunology* **13**, 273-283.

Ratcliffe N. A. and Rowley A. F. (1979) Role of Hemocytes in Defence Against Biological Agents. In *Insect Hemocytes* (Edited by Gupta A. P.). Cambridge University Press, Cambridge.

Ratcliffe N. A., Rowley A. F., Fitzgerald S. W. and Rhodes C. P. (1985) Invertebrate Immunity - Basic Concepts and Recent Advances. *International Review of Cytology-a Survey of Cell Biology* **97**, 183-350.

Rattanachai A., Hirono I., Ohira T., Takahashi Y. and Aoki T. (2004) Molecular cloning and expression analysis of alpha 2-macroglobulin in the kuruma shrimp, *Marsupenaeus japonicus*. *Fish and Shellfish Immunology* **16**, 599-611.

Ravindranath M. H. (1975) Effects of Temperature on the Morphology of Hemocytes and Coagulation Process in the Mole-crab *Emerita (=Hippa) asiatica*. *Biological Bulletin* **148**, 286-302.

Reid D. G., Abello P., Kaiser M. J. and Warman C. G. (1997) Carapace colour, inter-moult duration and the behavioral and physiological ecology of the shore crab *Carcinus maenas*. *Estuarine, Coastal and Shelf Science* **44**, 203-211.

Relf J. M. (2000) Antibacterial Proteins in *Carcinus maenas* (Crustacea: Decapoda). In *Department of Biology*, p. 190. St. Andrews University, St. Andrews.

Relf J. M., Chisholm J. R. S., Kemp G. D. and Smith V. J. (1999) Purification and characterisation of a cysteine-rich 11.5kDa antibacterial protein from the granular haemocytes of the shore crab, *Carcinus maenas*. *European Journal of Biochemistry* **264**, 350-357.

Rewitz K., Styrishave B. and Andersen O. (2003) CYP330A1 and CYP4C39 enzymes in the shore crab *Carcinus maenas*: sequence and expression regulation by ecdysteroids and xenobiotics. *Biochemical and Biophysical Research Communications* **310**, 252-260.

Rivera A., Santiago K., Torres J., Sastre M. P. and Rivera F. F. (1999) Bacteria associated with hemolymph in the crab *Callinectes bocourti* in Puerto Rico. *Bulletin of Marine Science* **64**, 543-548.

Roch P. (1999) Defense mechanisms and disease prevention in farmed marine invertebrates. *Aquaculture* **172**, 125-145.

Rodriguez J. and Le Moullac G. (2000) State of the art of immunological tools and health control of penaeid shrimp. *Aquaculture* **191**, 109-119.

Rojtinnakorn J., Hirono I., Itami T., Takahashi Y. and Aoki T. (2002) Gene expression in haemocytes of kuruma prawn, *Penaeus japonicus*, in response to infection with WSSV by EST approach. *Fish and Shellfish Immunology* **13**, 69-83.

Romestand B., Molina B., Richard V., Roch P. and Granier C. (2003) Key role of the loop connecting the two beta strands of mussel defensin in its antimicrobial activity. *European Journal of Biochemistry* **270**, 2805-2813.

Rost B. and Sander C. (1993) Prediction of protein secondary structure at better than 70% accuracy. *Journal of Molecular Biology* **232**, 584-599.

Rost B. and Sander C. (1994) Combining evolutionary information and neural networks to predict protein secondary structure. *Proteins* **19**, 55-72.

Roux M. M., Pain A., Klimpel K. R. and Dhar A. K. (2002) The lipopolysaccharide and beta-1,3-glucan binding protein gene is upregulated in white spot virus-infected shrimp (*Penaeus stylirostris*). *J Virol* **76**, 7140-7149.

Salzet M. (2001) Vertebrate innate immunity resembles a mosaic of invertebrate immune responses. *Trends in Immunology* **22**, 285-288.

Sambrook J. and Russell D. (2001) *Molecular Cloning Volume 3 -A Laboratory Manual*. Cold Spring Harbour Laboratory Press.

Satchell D. P., Sheynis T., Shirafuji Y., Kolusheva S., Ouellette A. J. and Jelinek R. (2003) Interactions of Mouse Paneth Cell alpha -Defensins and alpha -Defensin Precursors with Membranes. Prosegment Inhibition of Peptide Association with Biometric Membranes. *Journal Biological Chemistry* **278**, 13838-13846.

Schein C. H. and Noteborn M. H. M. (1988) Formation of Soluble Recombinant Proteins in *Escherichia coli* Is Favored by Lower Growth Temperature. *Biotechnology* **6**, 291-294.

Schmittgen T. D. and Zakrajsek B. A. (2000) Effect of experimental treatment on housekeeping gene expression: validation by real-time, quantitative RT-PCR. *Journal of Biochemical and Biophysical Methods* **46**, 69-81.

Schnapp D. (1996) Antiviral Defence and Antibacterial Proteins in the Shore Crab, *Carcinus maenas*. In *Gatty Marine Laboratory*, p. 286. University of St. Andrews, St. Andrews.

Schnapp D., Kemp G. D. and Smith V. J. (1996) Purification and characterisation of a proline-rich antibacterial peptide, with sequence similarity to bactenecin-7, from the haemocytes of the shore crab, *Carcinus maenas*. *European Journal of Biochemistry* **240**, 532-539.

Schoner R. G., Ellis L. F. and Schoner B. E. (1992) Isolation and purification of protein granules from *Escherichia coli* cells overproducing bovine growth hormone. 1985. *Biotechnology* **24**, 349-352.

Sequeira T., Tavares D. and Arala-Chaves M. (1996) Evidence for circulating hemocyte proliferation in the shrimp *Penaeus japonicus*. *Developmental and Comparative Immunology* **20**, 97-104.

Sewell M. T. (1955) Lipo Protein Cells in the Blood of *Carcinus maenas* and their cycle of Activity Correlated with the Molt. *Quarterly Journal of Microscopical Science* **96**, 73-83.

Shai Y. (1997) Mode of action of antibacterial peptides. In *Molecular Mechanisms of Immune Responses in Insects* (Edited by Brey P. T. and Hultmark D.). Chapman and Hall, London.

Shai Y. (1999) Mechanism of the binding, insertion and destabilization of phospholipid bilayer membranes by alpha-helical antimicrobial and cell non-selective membrane-lytic peptides. *Biochimica et Biophysica Acta (BBA) - Proteins & Proteomics* **1462**, 55-70.

Sheeks R. B. (1989) Taiwan's aquaculture - at the crossroads. *Infofish International* **6**, 38-43.

Shevchenko A., Wilm M., Vorm O. and Mann M. (1996) Mass spectrometric sequencing of proteins silver-stained polyacrylamide gels. *Analytical Chemistry* **68**, 850-858.

Shirafuji Y., Tanabe H., Satchell D. P., Henschen-Edman A., Wilson C. L. and Ouellette A. J. (2003) Structural Determinants of Procryptdin Recognition and Cleavage by Matrix Metalloproteinase-7. *Journal of Biological Chemistry* **278**, 7910-7919.

Sizemore R. K., Colwell R. R., Tubiash H. S. and Lovelace T. E. (1975) Bacterial flora of the hemolymph of the blue crab, *Callinectes sapidus*: numerical taxonomy. *Applied Microbiology* **29**, 393-399.

Skosyrev V. S., Kulesskiy E. A., Yakhnin A. V., Temirov Y. V. and Vinokurov L. M. (2003) Expression of the recombinant antibacterial peptide sarcotoxin IA in *Escherichia coli* cells. *Protein Expression and Purification* **28**, 350-356.

Sloan B., Yocum C. and Clem L. W. (1975) Recognition of self from non-self in crustaceans. *Nature* **258**, 521-523.

Smith V. J. (1978) Cellular Defence Reactions of *Carcinus maenas* (Crustacea). In *Department of Zoology*, p. 242. University College of Swansea, Swansea.

Smith V. J. (1991) Invertebrate Immunology: Phylogenetic, Ecotoxicological and Biomedical Implications. *Comparative Haematology International* **1**, 61-76.

Smith V. J. (1996) The Prophenolosidase Activating System: A Common Defence Pathway for Deutrostomes and Protostomes? In *Advances in Comparative and Environmental Physiology* (Edited by Heidelberg S.-V. B.).

Smith V. J. and Chisholm J. R. S. (1992) Non-cellular immunity in crustaceans. *Fish and Shellfish Immunology* **2**, 1-31.

Smith V. J. and Chisholm J. R. S. (2001) Antimicrobial Proteins in Crustaceans. *Advances in Experimental Medical Biology* **484**, 95-112.

Smith V. J., Chisholm J. R. S. and Brown J. H. (2001) Developmental Ecology of Immunity in Crustaceans. *Bios*.

Smith V. J. and Ratcliffe N. A. (1978) Host Defence Reactions of the Shore Crab, *Carcinus maenas* (L.) *in vitro*. *Journal of the Marine Biological Association of the United Kingdom* **58**, 367-379.

Smith V. J. and Ratcliffe N. A. (1980a) Cellular Defense Reactions of the shore crab *Carcinus maenas.* In vivo haemocytic and histopathological responses to injected bacteria. *Journal of Invertebrate Pathology* **35**, 65-74.

Smith V. J. and Ratcliffe N. A. (1980b) Host Defence Reactions of the Shore Crab *Carcinus maenas* (L.): Clearance and Distrinution of injected Test Particles. *Journal of the Marine Biological Association of the United Kingdom* **60**, 89-102.

Smith V. J. and Ratcliffe N. A. (1981) Pathological Changes in the Nephrocytes of the Shore Crab *Carcinus maenas*, following Injection of Bacteria. *Journal of Invertebrate Pathology* **38**, 113-121.

Smith V. J. and Ratcliffe N. A. (1983) b-1,3 Glucan Activation of Crustacean hemocytes in vitro and *in vivo*. *Biological Bulletin* **164**, 299-314.

Smith V. J. and Söderhäll K. (1983) Induction of De-Granulation and Lysis of Hemocytes in the Fresh-Water Crayfish, *Astacus astacus* By Components of the Prophenoloxidase Activating System *in vitro*. *Cell and Tissue Research* **233**, 295-303.

Smith V. J. and Söderhäll K. (1986) Cellular Immune Mechanisms in the Crustacea. In *Symposium of the Zoological Society of London* (Edited by London T. Z. S. o.), Vol. 56, London.

Smith V. J., Söderhäll K. and Hamilton M. (1984) b-1,3,Glucan Induced Cellular Defence Reactions in the Shore Crab, *Carcinus maenas*. *Comparative Biochemistry and Physiology* **77a**, 635-639.

Snell T. W., Brogdon S. E. and Morgan M. B. (2003) Gene expression profiling in ecotoxicology. *Ecotoxicology* **12**, 475-843.

Söderhäll I., Bangyeekhun E., Mayo S. and Söderhäll K. (2003) Hemocyte production and maturation in an invertebrate animal; proliferation and gene expression in hematopoietic stem cells of *Pacifastacus leniusculus*. *Developmental and Comparative Immunology* **27**, 661-672.

Söderhäll K. (1982) Prophenoloxidase activating system and melanisation - a recognition mechanism in arthropods - a review. *Developmental and Comparative Immunology* **6**, 601-611.

Söderhäll K. (1992) Biochemical and Molecular Aspects of Cellular Communication in Arthropods. *Bollettino Di Zoologia* **59**, 141-151.

Söderhäll K., Aspan A. and Diuvic B. (1990) The proPO-System and Associated Proteins: Role in Cellular Communication in Arthropods. *Research in Immunology* **141**, 896-907.

Söderhäll K. and Cerenius L. (1992) Crustacean Immunity. *Annual Review of Fish Diseases*, 3-23.

Söderhäll K., Cerenius L. and Johansson M. W. (1994) The Prophenoloxidase Activating System and Its Role in Invertebrate Defense. In *Primordial Immunity: Foundations For the Vertebrate Immune System*, Vol. 712, p. 155-161.

Söderhäll K. and Smith V. J. (1983) Separation of the Hemocyte Populations of *Carcinus maenas* and Other Marine Decapods, and Prophenoloxidase Distribution. *Developmental and Comparative Immunology* **7**, 229-239.

Söderhäll K. and Smith V. J. (1986) The Prophenoloxidase Activating System: The Biochemistry of its Activation and Role in Arthropod Cellular Immunity, with Special Reference to Crustaceans. In *Immunity in Invertebrates* (Edited by Brehe'lin M.), p. 208-223. Springer-Verlag Berlin Heidelberg, Berlin.

Söderhäll K., Smith V. J. and Johansson M. W. (1986) Exocytosis and Uptake of Bacteria By Isolated Hemocyte Populations of 2 Crustaceans - Evidence For Cellular Cooperation in the Defense Reactions of Arthropods. *Cell and Tissue Research* **245**, 43-49.

Söderhäll K. and Thornqvist P. O. (1997) Crustacean Immunity - A Short Review. *Development of Specifications for Biotechnology and Pharmaceutical Products* **90**, 45-51.

Spears T., Abele L. G. and Kim W. (1992) The Monophyly of Brachyuran Crabs: A Phylogenetic Study Based on 18S rRNA. *Systematic Biology* **41**, 446-461.

Sritunyalucksana K., Sithisarn P., Withayachumnarnkul B. and Flegel T. W. (1999) Activation of prophenoloxidase, agglutinin and antibacterial activity in haemolymph of the black tiger prawn, *Penaeus monodon*, by immunostimulants. *Fish and Shellfish Immunology* **9**, 21-30.

Stewart E. J. and Arie B. (1973) Paradoxical effects of Salinity Reductions on Lobsters (*Homarus americanus*) infected with *Gaffkya homari*. *Comparative Biochemistry and Physiology* **45A**, 717-730.

Stewart E. J. and Zwicker B. M. (1971) Induction of Internal Defence Mechanisms in the Lobster: *Homarus americanus*. In *Contemporary Topics in Immunobiology* (Edited by Cooper E. L.), Vol. 4, p. 233-239.

Stewart J. E. and Zwicker B. M. (1972) Natural and induced bactericidal activities in the hemolymph of the lobster, *Homarus americanus*: products of hemocyte-plasma interaction. *Canadian Journal of Microbiology* **18**, 1499-1513.

Stoss T. D., Nickell M. D., Hardin D., Derby C. D. and McClintock T. S. (2004) Inducible transcript expressed by reactive epithelial cells at sites of olfactory sensory neuron proliferation. *Journal of Neurobiology* **58**, 355-368.

Stuart L. S. and Harris T. H. (1942) Bactericidal and Fungicidal Properties of a Crystalline Protein Isolated From Unbleached Wheat Flour. *Cereal Chemistry* **19**, 288-300.

Styrishave B., Aagaard A. and Andersen O. (1999) In situ studies on physiology and behavior in two colour forms of the shore crab *Carcinus maenas* in relation to season. *Marine Ecology Progress Series* **189**, 221-231.

Subasinghe R. P., Bartley D. M., McGladdery S. and Barg U. (1998) Sustainable Shrimp Culture Development: Biotechnological Issues and Challenges. In *Advances in Shrimp Biotechnology* (Edited by Flegel T. W.). National Centre for Engineering and Biotechnology, Bangkok.

Supungul P., Klinbunga S., Pichyangkura R., Jitrapakdee S., Hirono I., Aoki T. and Tassanakajon A. (2002) Identification of Immune-Related Genes in Hemocytes of Black Tiger Shrimp (*Penaeus monodon*). *Marine Biotechnology* **4**, 487-494.

Taylor S. W. (2002) Chemoenzymatic synthesis of peptidyl 3,4-dihydroxyphenylalanine for structure-activity relationships in marine invertebrate polypeptides. *Analytical Biochemistry* **302**, 70-74.

Teague P. O. and Friou G. J. (1964) Lack of Immunological Responses by an Invertebrate. *Comparative Biochemistry and Physiology* **12**, 471-478.

Terwilliger N. B. and Dumler K. (2001) Ontogeny of decapod crustacean hemocyanin: Effects of temperature and nutrition. *Journal of Experimental Biology* **204**, 1013-1020.

The Times. (1989) *Atlas and Encyclopedia of the Sea*. Times Books Ltd., London.

Thellin O., Zorzi W., Lakaye B., De Borman B., Coumans B., Hennen G., Grisar T., Igout A. and Heinen E. (1999) Housekeeping genes as internal standards: use and limits. *Journal of Biotechnology* **75**, 291-295.

Thornqvist P. O., Johansson M. W. and Söderhäll K. (1994) Opsonic Activity of Cell-Adhesion Proteins and Beta-1,3-Glucan Binding-Proteins From 2 Crustaceans. *Developmental and Comparative Immunology* **18**, 3-12.

Tollin M., Bergman P., Svenberg T., Jornvall H., Gudmundsson G. H. and Agerberth B. (2003) Antimicrobial peptides in the first line defence of human colon mucosa. *Peptides* **24**, 523-530.

Torres G., Gimenez L. and Anger K. (2002) Effects of reduced salinity on the biochemical composition (lipid, protein) of zoea 1 decapod crustacean larvae. *Journal of Experimental Marine Biology and Ecology* **277**, 43-60.

Tossi A., Sandri L. and Giangaspero A. (2000) Amphipathic, A-Helical Antimicrobial Peptides. *Biopolymers* **55**, 4-30.

Towle D. W. (1997) Molecular approaches to understanding salinity adaptation of estuarine animals. *American Zoologist* **37**, 575-584.

Towle D. W., Jordan M. K. and Shearer D. W. (1995) Estimation of Sodium/Proton Antiporter mRNA in Gills of the Green Shore Crab *Carcinus maenas* by Quantitative RT-PCR. *Bulletin of Mount Desert Island Biological Lab* **34**, 64-65.

Towle D. W., Rushton M. E., Heidysch D., Magnani J. J., Rose M. J., Amstutz A., Jordan M. K., Shearer D. W. and Wu W. S. (1997) Sodium/proton antiporter in the euryhaline crab *Carcinus maenas*: Molecular cloning, expression and tissue distribution. *Journal of Experimental Biology* **200**, 1003-1014.

Truscott R. and White K. N. (1990) The Influence of Metal and Temperature Stress On the Immune System of Crabs. *Functional Ecology* **4**, 455-461.

Tsvetnenko E., Fotedar S. and Evans L. (2001) Antibacterial activity in the haemolymph of western rock lobster, *Panulirus cygnus*. *Marine and Freshwater Research* **52**, 1407-1412.

Tubiash H. S., Sizemore R. K. and Colwell R. R. (1975) Bacterial flora of the hemolymph of the blue crab, *Callinectes sapidus*: most probable numbers. *Appl Microbiol* **29**, 388-392.

Turbett G. R. and Sellner L. N. (1996) Digestion of PCR and RT-PCR Products With Restriction Endonucleases Without Prior Purification or Precipitation. In *Promega Notes Magazine*, Vol. 60, p. 23.

Ueda R., Sugita H. and Deguchi Y. (1993) Microflora in the Hemolymph of Japanese Coastal Crustacea. *Nippon Suisan Gakkaishi* **59**, 1637-1637.

Ulevitch R. J. (2004) Theraputics Targeting The Innate Immune System. *Nature Reviews Immunology* **4**, 512-520.

Valore E. V., Martin E., Harwig S. S. L. and Ganz T. (1996) Intramolecular Inhibition of Human Defensin HNP-1 by Its Propiece. *Journal of Clinical Investigation* **97**, 1624-1629.

Vandesompele J., De Preter K., Pattyn F., Poppe B., Van Roy N., De Paepe A. and Speleman F. (2002) Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biologyl* **3**.

Vargas-Albores F., Hinojosa-Baltazar P., Portillo-Clark G. and Magallon-Barajas F. (1998) Influence of temperature and salinity on the yellowleg shrimp, *Penaeus californiensis* Holmes, prophenoloxidase system. *Aquaculture Research* **29**, 549-553.

Vargas-Albores F., Yepiz-Plascencia G., Jimenez-Vega F. and Avila-Villa A. (2004) Structural and functional differences of *Litopenaeus vannamei* crustins. *Comparative Biochemistry and Physiology Part B: Biochemistry and Molecular Biology* **138**, 415-422.

Vivares C. P. and Cuq J. L. (1981) Physiological and Metabolic Variations in *Carcinus mediterraneus* (Crustacea - Decapoda) Parasitized By *Thelohania maenadis* (Microsporida) - an Ecophysiopathological Approach. *Journal of Invertebrate Pathology* **37**, 38-46.

Vizioli J. and Salzet M. (2002) Antimicrobial peptides from animals: focus on invertebrates. *Trends in Pharmacological Science* **23**, 494-496.

Wang Z. and Wang G. (2004) APD: the Antimicrobial Peptide Database. *Nucleic Acids Research* **32 Database issue**, D590-592.

Warr G. W., Chapman R. W. and Smith L. C. (2003) Evolutionary immunobiology: new approaches, new paradigms. *Developmental and Comparative Immunology* **27**, 257-262.

Weinheimer P. F., Acton R. T., Sawyer S. and Evans E. E. (1969) Specificity of the induced bactericidin of the West Indian spiny lobster, *Panulirus argus*. *Journal of Bacteriology* **98**, 947-948.

Werling D. and Jungi T. W. (2003) TOLL-like receptors linking innate and adaptive immune response. *Veterinary Immunology and Immunopathology* **91**, 1-12.

Westhead D. R., Parish J. H. and Twyman R. M. (2002) *Bioinformatics*. BIOS Scientific Publishing Ltd., Leeds.

White K. N. and Ratcliffe N. A. (1982) The segregation and elimination of radio- and fluorescent-labelled marine bacteria from the haemolymph of the shore crab, *Carcinus maenas*. *Journal of the Marine Biological Association of the United Kingdom* **62**, 819-833.

White K. N., Ratcliffe N. A. and Rossa M. (1985) The Antibacterial Activity of Hemocyte Clumps in the Gills of the Shore Crab, *Carcinus maenas*. *Journal of the Marine Biological Association of the United Kingdom* **65**, 857-870.

White S. H., Wimley W. C. and Selsted M. E. (1995) Structure, function, and membrane integration of defensins. *Current Biology* **5**, 521-527.

Whiteley N. M. and El Haj A. J. (1997) Regulation of muscle gene expression over the moult in crustacea. *Comparative Biochemistry and Physiology B-Biochemistry & Molecular Biology* **117**, 323-331.

Whiteley N. M., Taylor E. W. and El Haj A. J. (1992) Actin Gene-Expression During Muscle Growth in *Carcinus maenas*. *Journal of Experimental Biology* **167**, 277-284.

Whiteley N. M., Taylor E. W. and El Haj A. J. (1997) Seasonal and latitudinal adaptation to temperature in crustaceans. *Journal of Thermal Biology* **22**, 419-427.

Williams A. J. and Lutz P. L. (1975) Blood Cell Types in *Carcinus maenas* and their Physiological Role. *Journal of the Marine Biological Association of the United Kingdom* **55**, 671-674.

Williamson H. C. (1940) *The Crab Fishery*. St. Andrews Natural History Department, Dundee.

Wilson J. F. (2002) Renewing the Fight Against Bacteria. In *The Scientist*, Vol. 16, p. 22-23.

Wootton J. C. and Federhen S. (1996) Analysis of compositionally biased regions in sequence databases. *Methods in Enzymology* **266**, 554-571.

Yang D., Chertov O., Bykovskaia S. N., Chen Q., Buffo M. J., Shogan J., Anderson M., Schroder J. M., Wang J. M., Howard O. M. and Oppenheim J. J. (1999) Beta-defensins: linking innate and adaptive immunity through dendritic and T cell CCR6. *Science* **286**, 525-528.

Yang S. T., Yub Shin S. Y., Kim Y. C., Kim Y., Hahm K. S. and Kim J. I. (2002) Conformation-dependent antibiotic activity of tritrpticin, a cathelicidin-derived antimicrobial peptide. *Biochemical and Biophysical Research Communications* **296**, 1044-1050.

Yang Y., Poncet J., Garnier J., Zatylny C., Bachère E. and Aumelas A. (2003) Solution structure of the recombinant penaeidin-3, a shrimp antimicrobial peptide. *Journal of Biological Chemistry* **278**, 36859-36867.

Yang Y. S., Mitta G., Chavanieu A., Calas B., Sanchez J. F., Roch P. and Aumelas A. (2000) Solution structure and activity of the synthetic four-disulfide bond Mediterranean mussel defensin (MGD-1). *Biochemistry* **39**, 14436-14447.

Yiallouros I., Kappelhoff R., Schilling O., Wegmann F., Helms M. W., Auge A., Brachtendorf G., Berkhoff E. G., Beerman B., Hinz H.-J., Konig S., Peter-Katalinic J. and Stocker W. (2002) Acivation Mechanism of Pro-astacin: Role of the Pro-peptide, Tryptic and Autoproteolytic Cleavage and Importance of Precise Amino-terminal Processing. *Journal of Molecular Biology* **324**, 237-246.

Yonge C. M. (1990) *The Sea Shore*. Bloomsbury Books, London.

Zanetti M., Gennaro R. and Romeo D. (1995) Cathelicidins: a novel protein family with a common proregion and a variable C-terminal antimicrobial domain. *Febs Letters* **374**, 1-5.

Zasloff M. (1992) Antibiotic peptides as mediators of innate immunity. *Current Opinion in Immunology* **4**, 3-7.

Zasloff M. (2002) Antimicrobial peptides of multicellular organisms. *Nature* **415**, 389-395.

Zdobnov E. M. and Apweiler R. (2001) InterProScan - an integration platform for the signature-recognition methods in InterPro. *Bioinformatics* **19**, 847-848.

Zeng C. S. and Naylor E. (1997) Rhythms of larval release in the shore crab *Carcinus maenas* (Decapoda: Brachyura). *Journal of the Marine Biological Association of the United Kingdom* **77**, 451-461.

Zhang H. and Kato Y. (2003) Common structural properties specifically found in the CSalphabeta-type antimicrobial peptides in nematodes and mollusks: evidence for the same evolutionary origin? *Developmental and Comparative Immunology* **27**, 499-503.

Zhang L., Falla T., Wu M., Fidai S., Burian J., Kay W. and Hancock R. E. (1998) Determinants of recombinant production of antimicrobial cationic peptides and creation of peptide variants in bacteria. *Biochemical Biophysical Research Communications* **247**, 674-680.

Zhao C., Wang I. and Lehrer R. I. (1996) Widespread expression of beta-defensin hBD-1 in human secretory glands and epithelial cells. *Febs Letters* **396**, 319-322.