Studies of Recombinant Protein Expression: Targeting Signals, 3'untranslated Regions and Trans-acting Factors

Siavash Khazaipoul

Institute for Cell and Molecular Biosciences
Newcastle University

Thesis submitted for the degree of Doctor of Philosophy

June 2010

DECLARATION

I hereby declare that I alone have submitted this thesis and except where stated the work presented herein is my own.

June 2010

Acknowledgments

I would firstly like to thank John for giving me the opportunity to work on this project, for his help and invaluable guidance throughout and in the revising of this thesis. I would also like to thank Beate, Ian and everybody at UniTargetingResearch As, Bergen for their kind support and making my time in Bergen an unforgettable experience every time. I would especially like to thank Ian and Beate for revising my thesis and for providing valuable feedback.

I would like to thank past and present members of John Hesketh and Dianne Ford's labs for their help, support especially Brian and Mark for their invaluable assistance. Also thanks goes to the members of Jeremy Brown and Zosia Chrzanowska-Lightowlers's labs for their kind support and assistance.

My gratitude and thanks goes to my parents who have always encouraged and supported me and a big thank you to my brother Rahman and his family for accommodating me during my several visits to Bergen.

And last but not least, a big thank you to Eileen for her support.

This project was found by BBSRC and UniTargetingResearch AS, Bergen.

Abstract

At present mammalian cell factories are being employed for recombinant protein production. However, the yields of proteins produced from such systems are often poor. This thesis describes experiments to study the effects of altering targeting signals (signal peptide) and 3' untranslated regions (3'UTR) in an expression vector on protein expression. A variety of gene constructs containing *Gaussia princeps* luciferase as reporter were created using a seamless cloning method. In these constructs a variety of signal peptides, some with altered hydrophobicity, were combined with the *Gaussia* luciferase coding region and either the native *Gaussia* luciferase or human albumin 3'UTR. These were then transfected into CHO AA8 Tet-Off cells to measure how modification of the signal peptide/3'UTR affects protein expression. The results indicate that the Albumin 3'UTR, in conjunction with an appropriate signal peptide, boosts protein production by approximately 3 fold compared to the native *Gaussia* luciferase 3'UTR. Deletion analysis of the Albumin 3'UTR showed that deletion of regions Δ1-50, Δ1-100, Δ1-150, Δ101-150 significantly reduces protein production compared with deletion of regions Δ51-100, Δ51-150 and Δ1-50&101-150. Interestingly, mRNA abundance levels were significantly decreased for constructs containing deletions in regions 1-50, 1-150 and 1-50&101-150. UV Cross linking and electrophoretic mobility gel shift competition assays showed strong competition by RNA transcripts from the deletion construct Δ1-50, which was then used as bait for isolating bound protein/s from a CHO cell extract. Three proteins, including CUG-BP1 an RNA-binding protein involved in mRNA stability and translation were identified by mass spectrophotometry analysis. Knock down of CUG-BP1 expression using siRNA, led to impairment of complex formation between CHO cell protein extract and Albumin 3'UTR RNA transcripts, and in addition it led to an increase in the

reporter activity and mRNA expression level in cells expressing the reporter gene with the full length Albumin 3'UTR and deletion variant 51-100. It is hypothesised that the differences in mRNA expression levels and secreted luciferase activity were due to CUG-BP1 binding to the Albumin 3'UTR. Further work is needed to explore the effects of CUG-BP1 on mRNA translation and stability.

List of Figures

List of Tables

Glossary of terms

A: Adenosine

APS: Ammonium Persulphate

ATP: Adenosine 5'- Triphosphate

bp: Base Pair

BSA: Bovine Serum Albumin

˚C: Degrees Celsius

CDS: Coding Sequence

CHO: Chinese Hamster Ovary

DEPC: Diethyl Pyrocarbonate

DMEM: Dulbecco's Modified Eagles Medium

DMSO: Dimethyl Sulphoxide

dNTP: Deoxynucleotide Triphosphate

DTT: Dithiothreitol

ER: Endoplasmic Reticulum

EOR: ER Overload Response

FCS: Fetal Calf Serum

G: *Gaussia princeps*

GDP: Guanine Diphosphate

GTP: Guanine 5'-Triphosphate

Hr: Hour(s)

kDa: Kilodalton

M: Mole(s)

MBq: Megabecquerel

mg: Milligram(s)

min: Minute(s)

ml: Millilitre(s)

mRNA: Messenger Ribonucleic Acid

MW: Molecular Weight

PAGE: Polyacrylamide Gel Electrophoresis

PBS: Phosphate Buffered Saline

PCR: Polymerase Chain Reaction

Poly (A): Poly Adenosine

PVDF: Polyvinylidene Difluoride

RBPs: RNA Binding Proteins

rpm: Revolution per Minute

RT-PCR: Reverse Transcriptase PCR

S: Second

SA-PMPs: Streptavidin MagneSphere Paramagnetic Particles

SP: Signal Peptide

SR: Signal Recognition Particle Receptor

SRP: Signal Recognition Particle

TAE: Tris-Acetae EDTA

TBE: Tris-Borate EDTA

TEMED: N,N,N',N'-tetramethylethylene-diamine

UPR: Unfolded Protein Response

UTR: Untranslated Region

UV: Ultra Violet

V: Volt(s)

V/V: Volume per Unit Volume

W/V: Weight per Unit Volume

1. Introduction

1.1 Recombinant protein production and biotechnology

Recombinant protein production has increasingly become a thriving industry in the past three decades. Since their introduction, recombinant proteins have made a huge impact on modern medicine by providing treatment for varying conditions ranging from cancer to infertility (Victibix, Follistim/Gonal-F). For any given recombinant protein to be biologically active, it requires the correct folding and where possible appropriate post-translational modifications. If the recombinant proteins are to be used as therapeutics or for medical diagnostic purposes, it is crucial that they are produced to the highest degree of similarity to their naturally occurring equivalent.

Table1.1 A selection of commercialised recombinant proteins produced in Chinese Hamster Ovary cell lines (CHO) [1] .

Product	Therapeutic use	Year of approval
Vectibix	Metastatic colorectal cancer	2006
Myozyme	Pompe disease	2006
Orencia	Rheumatoid arthritis	2005
Luveris	Infertility	2004
Avastin	Lung and colorectal cancer	2004
Advate	Hemophilia A	2003
Xolair	Asthema	2003

Although much has changed since the early days of recombinant protein production, the general approach for the large scale production of recombinant proteins has been based on the

development of genetically engineered expression systems that contain the genetic code for the protein of interest. Mammalian expression systems offer advantages and are superior to other systems (bacteria, yeast) in that they produce proteins with post-translational modifications (glycosylation, phosphorylation, etc) which are essential for many proteins to be fully biologically active [2]**.** However a major obstacle for mammalian expression system has been the relatively low production of protein from these cells as well as being costly and complex to maintain [3]. To overcome this obstacle various methods have been utilised, such as expression vectors which contain a heterologous promoter in conjunction with polyadenylation signals [4]**.** The production efficiency of recombinant proteins in mammalian systems can be increased if the chosen cell line is compatible with the promoter [5]**.** In addition to this method, gene amplification to increase the gene copy number has also been widely used**.** An example of the latter is the dihydrofolate reductase (dhfr) gene that enables the amplification of the transfected DNA in host cells [6]. This method is one of the most stringent methods and frequently used processes available to select for cells that express the gene of interest at high level. However, a disadvantage to this system is the long selection process involving a steady increase of methotrexate (an inhibitor specific to the dfhr gene) that is needed to amplify the gene of interest that results in the generation of cell lines with high productivity [7].

Generally, to promote the expression of recombinant proteins a strong viral or cellular promoter/enhancer is used [8]. An example includes the cytomegalovirus (CMV) immediate early (IE) promoter-enhancer which is renowned for its strong activity in a variety of cell lines. The baculovirus vectors have also been utilised in mammalian and insect systems for large scale expression of recombinant proteins. The Baculovirus expression system offers

many advantages, such as insertion of up to 38 kb of DNA into the host genome which in turn allows the delivery of multiple genes from a single vector. Furthermore, this system allows the production of non-toxic products and can also carry out post-translational modifications to the proteins [9-13].

By optimising the coding region (codon optimisation) it is also possible to improve the level of gene expression in mammalian cell factories. For example, a mammalian gene may not be expressed at a high level if a rare codon is used. Therefore, by replacing the rare codon for one that is more abundant, high level protein expression can be achieved [14].

The transcription of recombinant genes is also heavily influenced by the site of integration [15]. This phenomenon is known as the "position effect". It is known that the integration into inactive heterochromatin leads to little or no levels of transgene expression, whereas the integration into active euchromatin allows transgene expression to occur more frequently. A drawback to the latter is that this method may not be sufficient to ensure long term expression of the recombinant gene. Also, it has been found that transgene expression in mammalian cells is rapidly inactivated or silenced in many cases. This could be due to the influence of nearby condensed chromatin, as gene silencing correlates with an increase in CpG methylation in the promoter region of the transgene, histone hypoacetylation and methylation of lysine 9 of histone 3 [16, 17].

Alternatively, prokaryotic systems can be employed to drive expression of transgenes. Examples of their use include that of the production of growth hormones and insulin, which are both produced for medicinal purposes on an industrial scale from *E.coli.* However, there are some limitations to this system. The lack of post-translational machinery to perform vital protein modification is a major caveat in this process. The lack of appropriate chaperones in

bacteria may also lead to the misfolding of proteins, which could potentially lead to the production of proteins that are very different in terms of biological activity compared to the native protein [3]. The use of known prokaryotic systems is thereby restricted to the preparation of proteins that are not naturally glycosylated or the natively glycosylated proteins which are pharmacologically active without being glycosylated. These include interleukins, interferons and tumor necrosis factor [18].

A new approach is currently being employed by a small biotechnology company (UniTargetingResearch AS, Bergen, http://www.unitargeting.com/) to improve protein production by focusing mainly on post-transcriptional mechanisms that influence mRNA/protein trafficking and expression. Rather than focusing on traditional approaches for increasing the level of recombinant proteins, such as optimisation of growth medium, increasing the transcriptional activity of the transgene by employing a strong promoter/enhancer or even amplification of gene copy number, this new technology tackles the obstacle of low yield of protein production from Chinese Hamster Ovary cells (CHO) by altering the signal peptide and the 3'untranslated region in the transgene. This novel approach is heavily dependent on two key elements within the expression vector (the signal peptide and the 3' untranslated region (3'UTR)): because of their crucial importance in mRNA targeting to the endoplasmic reticulum (ER) and regulation of expression through mRNA stability and translation [19].

In this chapter I will describe the biogenesis of secretory proteins focusing mainly on the signal peptide, signal recognition particle and the translocation of protein across the mammalian ER. I also review the current knowledge on mRNA translation, localisation and stability with respect to untranslated regions within the RNA transcript.

1.2 Synthesis of secreted proteins - an overview

Since a substantial fraction of synthesised proteins must be transported across or integrated into the membrane this raises the question of how the permeability barrier of the membrane is maintained if macromolecules cross the barrier, or how proteins could cross a barrier that usually prevents such movement? [20]. The process of protein transport from the cytosol into the ER is an extremely crucial point in the biogenesis of many proteins, comprising that of secretory proteins and plasma membrane proteins [21]. It has been known for decades that secretory proteins are synthesised with a hydrophobic N-terminal extension (signal peptide) that then guides proteins into the secretion pathway. Protein targeting can occur either co-translationally or post-translationally (which will not be discussed here).

In a co-translationally targeting manner, protein synthesis is tightly coupled to the translocation process. This requires a close partnership between the ribosome and components of the translocation machinery [22]. The advantage of this mode of targeting is the prevention of partial folding or the mis-folding of secretory proteins in the cytosol prior to their targeting. Therefore, the structures that could hamper the translocation of the protein cannot be formed [23]. The co-translational targeting to the ER is initiated as soon as the newly synthesised polypeptide emerges from the ribosome. Once the signal peptide and growing polypeptide emerge from the exit subunit of ribosome, it is recognised by a universally conserved protein complex known as the signal recognition particle (SRP).

Next, the complex of SRP and the newly synthesised polypeptide is targeted to the ER membrane via the affinities of the signal recognition particle for its receptor on the ER membrane. Once the complex binds to the ER, a series of GTPase hydrolysis reactions result in the release of the signal recognition particle from its receptor. This leads to the transfer of

the signal peptide to a protein conducting channel known as the translocon. This alleviates the elongation arrest leading to the resumption of translation of the polypeptide and the transfer of the synthesised chain into the ER membrane [22].

1.2.1 The signal peptide

A major breakthrough in cell biology was the discovery that secreted proteins contain N-terminal signal peptide that mediates their translocation across the ER during synthesis [24]**.** It has been discovered that the signal peptides are often interchangeable, able to tolerate a wide range of mutations, and can direct the secretion in evolutionary distant organisms [25-27]. It has been estimated that ~20% of random sequences can function as signal peptides and promote the secretion of yeast invertase [28]. Although signal peptides vary in amino acid content and length, they all share similar structural features (Figure 1.1) and consist of three regions: the N-domain (1-5 residues in length which contains only charged amino acid residues), a hydrophobic core (7-9 residues in length, also known as the H-domain) and finally a short C-domain containing the signal peptidase recognition site [29, 30]. Other than that the signal peptides have no known sequence homology.

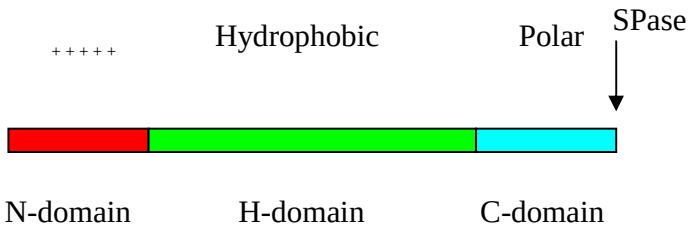


Figure 1.1 Schematic representation of signal peptide and its three distinct domains. Signal peptides share common structural features: a central hydrophobic core region (green) flanked by a hydrophilic N-region (red) with net positive charge and a C-terminal region (blue). The arrow shows the cleavage site by the signal peptidase (SPase).

The N-region of the signal peptide is positively charged and it has been reported that mutations that cause reduction of the net positive charge or the distance between the basic amino acid and the H-domain lead to an accumulation of proteins in the cytoplasm, accompanied with the reduction of post-translational export across the ER that can profoundly affect the signal peptide cleavage [31, 32]. The hydrophobic core region (H-domain), however, is by far the longest and most dominant of all three regions. Mutational analysis of signal peptides by various groups has shown that the H-domain is the most important for targeting of secretory/membrane bound proteins [29]. It has been reported that the hydrophobic core region along with the N-region play a crucial role in the translocation of proteins by anchoring the signal peptide firmly onto the membrane of ER [33].

Furthermore, it has also been reported that deletion of the hydrophobic core or insertion of positively charged amino-acids in this region severely disrupts protein translocation and protein secretion [34]. The C-domain, which immediately follows the hydrophobic H-domain, contains neutral but polar residues such as asparagine and serine, in addition to small non-polar and small uncharged amino acids at position -3 and -1 respectively (with respect to the cleavage site). The small non-polar and uncharged amino acids crucially determine the site of cleavage of the signal peptide by the signal peptidase [35]. It has been claimed that alanine is the most commonly found amino acid residue at these two positions [36, 37]. In general terms, in spite of having a tripartite structure and physical similarities, the signal peptides differ greatly in their functional properties. For instance, it has been claimed that signal peptides may differ in their entrance into the translocon [38] or their reliance on the translocation factors that are accessory to the process of translocation [39]. It has also been

reported that signal peptides can differ greatly in their overall efficiency in mediating the translocation process [20] or their various degree of sensitivity to translocation inhibitors [40].

1.2.2 Modification of the signal peptide and its consequences

Altering signal peptide hydrophobicity or even basicity is one approach with potential to augment protein secretion levels. However, there are some drawbacks as well as positive effects of these alterations. Since the hydrophobic core and the basic N-terminus of the signal peptide play a critical role in the translocation of proteins, by anchoring the signal peptide onto the ER membrane, it has been suggested that increasing the net positive charge of the basic region, or increasing the hydrophobicity of the H-domain, might augment the secretion level of the translocated protein [41]. Furthermore it has been reported that modification to the basic and hydrophobic domains may potentially alter the cleavage site, presumably by changing the configuration of the signal peptide inside the translocon or its release into the lipid bilayer [41, 42]. As a result, the cleavage of the signal peptide could potentially occur at cryptic cleavage sites. This could potentially decrease the biological activity of a recombinant therapeutic protein. The signal peptides interact with the SRP via their ability to form α-helical structures [43, 44].

It has been postulated that the helical potential of the H-domain may act as a unique factor for recognition by SRP and other elements of the export pathway of secretory proteins. Reducing the net positive charge of the N-terminus can be compensated by increasing the hydrophobicity of the H-domain [31, 45]. It has also been claimed that the positive charge of the N-terminus domain of the signal peptide may play a crucial role in associating the signal

peptide with the ER membrane as a result of electrostatic interaction with the negatively charged phospholipids [45].

Work by Zhang *et al* (2005) showed that alteration to both the basic N-terminus and hydrophobic H-domain increases the secretion level of protein more than that of the modification to each domain separately. However, further alterations (reducing the positive charge and increasing the hydrophobicity) to both domains did not lead to a further increase in the level of protein secretion. Firstly, this pointed to a hydrophobicity limit for the signal peptide. Secondly, it seems that the hydrophobic core of a signal peptide and the basic domain may act as a unit and share some overlapping functions. When the total hydrophobicity of the unit is ideal, it increases the secretion level of protein perhaps by having a stronger affinity for binding to SRP. On the other hand, when the total hydrophobicity is less than ideal, the affinity of the signal peptide for SRP may not be sufficient enough for proper export of the nascent chain [41].

1.2.3 Cleavage of the signal peptide and its fate

The signal peptide is cleaved by the signal peptidase as after its insertion into the translocation channel or the luminal side of the membrane. The cleavage depends on many aspects of the signal peptide, but the most important factor is the amino acids at positions -3 and -1 at the N-terminus of the cleavage site. Cleavage proceeds when charged amino acids are absent from -3 position and an amino acid with a short side chain is present at position -1. Other signal peptide features such as the length of the H-region and properties of its basic region may also influence the cleavage of the signal peptide [46].

It has been proposed that signal peptides may regulate the timing of the cleavage as a means of controlling events such as glycosylation and proper folding as well as the protein exit from the ER [47]. For instance, the slow cleavage of the native HIV-gp 120 signal peptide results in the extended retention of the secretory protein in the ER. Consequently, this prolonged retention prevents a premature exposure of the HIV-gp 120 to the immune system or to the adjacent cells. The unusually higher number of positively charged amino acids in the N-terminus of the HIV-gp 120 has been proposed to cause this delay [47, 48].

Little, however, is known about the fate of the signal peptide after cleavage, but it has been suggested that the cleaved fragments of the signal peptides may have functional role/s either in the lumen of the ER or the cytosol [46]. For instance, it has been demonstrated that the signal peptide of bovine preprolactin is first partly processed by the signal peptidase then further processed by an uncharacterised signal peptide peptidase [49]. This particular signal peptide is initially processed into a 20 amino acid long fragment, which contains the entire N-region and half of its hydrophobic core region. The fragment is later released into the cytosol where it binds to calmodulin in a Ca^{2+} dependent manner. Since this small acidic protein (calmodulin) regulates many cellular processes in the presence of Ca^{2+} signalling pathways [50]**,** this indicates that the signal peptide fragments may have a regulatory function. In addition, it has been suggested by others that fragments derived from signal peptide cleavage may play a critical role in the immunity surveillance of the cells. For example, the signal peptide fragments could be presented to the natural killer cells or the cytotoxic T cells by the major histocompatibility complex (MHC) class 1molecules [51].

1.2.4 The signal recognition particle (SRP)

The translocation of secretory proteins across the membrane of the ER is facilitated by a universally conserved protein complex named the signal recognition particle (SRP) [52, 53]. This 11S ribonucleoprotein binds to the newly synthesised proteins destined for secretion or insertion into the membrane as soon as they emerge from the large ribosomal polypeptide exit tunnel (Figure 1.2) [54]. The SRP then binds to their N-terminal signal peptide and forms a SRP-ribosome nascent chain complex (RNC). The association of the signal peptide with the SRP at this point causes elongation arrest or slow-down of translation of the newly synthesised protein. Subsequently, the SRP-RNC is directed to a protein conducting channel (translocon) in the ER by association of SRP with its receptor (SR), which is only found on rough ER [55, 56].

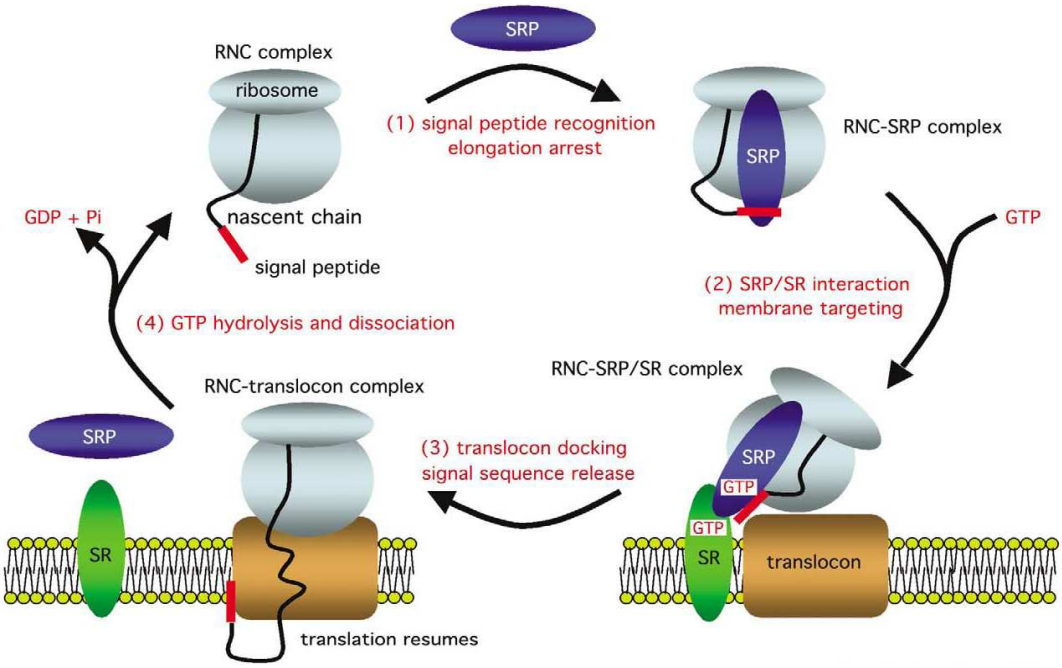


Figure 1.2 Translocation of secretory/membrane bound proteins across the mammalian ER. The SRP recognises the emerging signal peptide, causing elongation arrest (1). The SRP-RNC complex is then transferred to membrane in a GTP dependent manner and binds to its receptor, SR (2). After the docking of SRP on its receptor, the signal peptide is then released from SRP (3). Following GTP hydrolysis, the SRP disassociates from its receptor (4) [52]. Illustration taken from [52].

Interestingly, both the signal recognition particle and its receptor contain GTPase domains and alternate between active and inactive states that also reciprocally activate each other [57]**.** The GTP hydrolysis of the SRP and its receptor leads to dissociation of SRP from its receptor and the release of the RNC to the translocon. SRP is then recycled and reinitiated into another cycle of protein translocation [55, 56].

The SRP protein-RNA complex in eukaryotes can be divided into two domains: the S domain that contains the binding site for the signal peptide [58] and provides an interface for the interaction with the SR via coordinated GTP binding [59], and the Alu domain that is involved in elongation arrest [60-62]. The S domain comprises of the proteins SRP19, SRP54, SRP68, SRP72 as well as nucleotides 100-250 of the 7S RNA [63]. Crucially, the

most conserved part of the SRP complex, namely helix 8 and SRP54 are part of the S domain [64]. In eukaryotes, the assembly of the SRP complex takes place in the nucleus as well as in the cytoplasm. Initially, the SRP RNA molecule is transcribed in the nucleus while the protein subunits of SRP are synthesised in the cytoplasm. These are then imported into the nucleus with the exception of SRP54. The imported SRP proteins assemble onto the SRP RNA in the nucleus and then this pre-SRP is exported into the cytoplasm where SRP54 binds to the complex [65, 66].

The Alu domain of mammalian SRP however, includes the heterodimer SRP9/SRP14, which are bound to the extreme 5' and 3' ends of the 7S RNA [62]. It has been reported that the Alu domain is involved in peptide synthesis elongation arrest by interacting with the binding site of the elongation factor eEF2 that plays an essential role in protein synthesis [67, 68]. This leads to the antagonisation of translation as result of the interference with eEF2 binding to the ribosome [69]. The signal peptide recognition of the newly synthesise protein in a GTP-dependent interaction with SR is regulated by SRP54. The SRP54 subunit consists of 3 domains, the amino-terminal domain (N-domain), a central GTPase domain (G-domain) and a C-terminal M-domain [70]. It has been reported that the N-domain associates with the G-domain and may have a regulatory functional role [71]. Functional studies and cross-linking experiments have implicated the M-domain of SRP54 as primary site for the binding of the signal peptide to SRP [58] as well as the interaction of SRP54 to helix 8 of SRP RNA [70].

1.2.5 Targeting pathways

A variety of approaches has been employed to investigate the many aspects of protein translocation across the ER membrane. There are multiple pathways by which the signal peptides can direct proteins to the membrane. The translocation can be dependent on the SRP and the SRP receptor (SR), or can be fully independent of SRP [72]. If the targeting is independent of SRP, as can be found in bacteria and yeast, the translocation and targeting involves the SecB protein in bacteria and Sec62-Sec63 complex in yeast [73](see Figure 1.3). The SRP-dependent targeting pathway (as can be found in mammals) however is regulated by the 54kDa subunit of the SRP and two subunits of the SRP receptor (α,β). Interestingly, at the membrane the two different targeting pathways use the same targeting pore which is composed of the SecA and SecY-E-G complex in bacteria and Sec61 and the BiP protein in eukaryotic cells [74-76]. The key factor that determines the selection of SRP-dependent or SRP-independent targeting pathways is the signal peptide [46]. More specifically, the hydrophobicity of the signal peptide is the main parameter for selecting either of the pathways. It has been shown that signal peptides that direct proteins to a SRP-dependent pathway, contain a more hydrophobic H-domain than those signal peptides which promote a SRP-independent pathway [77, 78]. Interestingly, a targeting pathway independent of SRP and the Sec complex have also been identified in the plasma membrane of bacteria and the thylakoid membrane of plant chloroplasts [79, 80]. The key feature of this system is the existence of a twin-arginine motif upstream of the H-domain of the signal peptide. It has been demonstrated by several groups that most proteins targeted by the twin-arginine translocation system, known as Tat-system, are those which bind to redox-cofactors and fold or even oligomerise before targeting takes place [46].

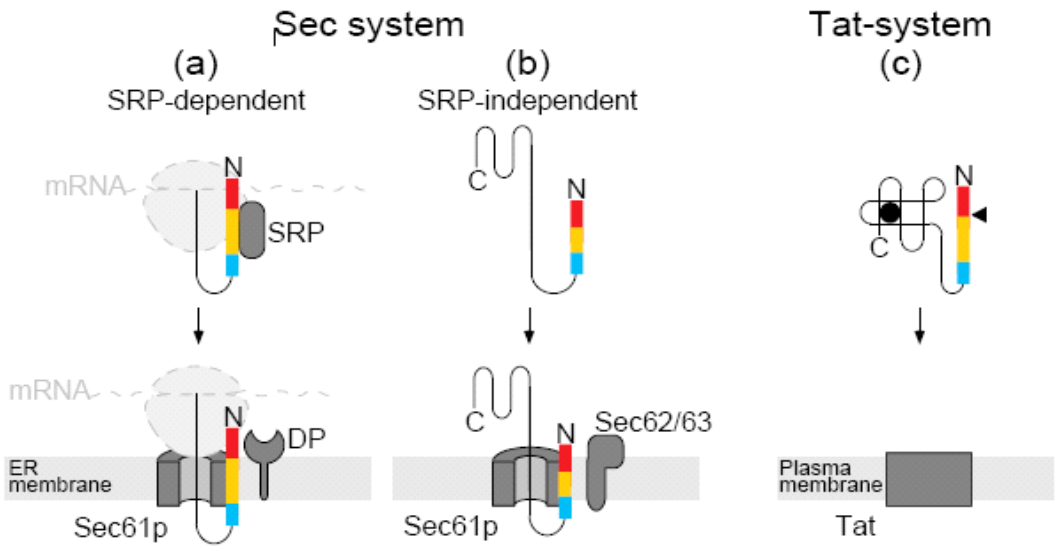


Figure 1.3 Recognition of different targeting pathways by the signal peptide. The hydrophobic region of the signal peptide (yellow) discriminates between an SRP-dependent (a) and SRP-independent pathway (b). In eukaryotic systems both pathways use the Sec61 complex. Twin-arginine translocation (Tat) so far has only been identified in plasmids and prokaryotes. Folded proteins with their bound co-factors (filled circle) use the twin-arginine translocation system (Tat-system), whereas unfolded proteins utilise the Sec61-dependent system. Illustration taken from [46].

The fact that there are at least three different known targeting pathways indicates the need for different rates of translocation. For instance, in rapidly growing organisms such as yeast, the post-translationally translocation pathway is perhaps more favorable. This is because it correlates with a higher rate of protein secretion than co-translational translocation. The higher rate of protein translocation is achieved because it is greater than the rate of protein synthesis. Therefore, it has been postulated that cells can make optimal use of the perhaps limited number of available Sec complexes. Hence in bacteria and yeast, the SRP-dependent pathway may only be used for proteins that quickly lose their translocation capability [81, 82].

1.2.6 Translocation channel (translocon)

Secretory/membrane bound proteins are translocated across the ER via an aqueous channel termed the translocon [72]. This multiprotein pore with a diameter of 9 to 15 Å when inactive and 40 to 60 Å when operational is a complex of proteins that is directly or indirectly involved in the translocation of proteins [83, 84]. The main component of this pore, the heterotrimeric Sec61, surrounds the aqueous pore that spans the lipid bilayer of the ER. Sec61 consists of Sec61α, Sec61β and Sec61γ and its associated proteins [85].

While the nascent chain remains in the translocon, it is kept inaccessible to either the luminal proteins or the cytoplasmic proteolytic proteins. The luminal side of the translocon is sealed by BiP until the nascent chain grows to a threshold of ~70 amino acids long [86]. The growing polypeptide opens the luminal end and co-translational translocation proceeds through the translocon (Figure 1.4). The BiP protein is a member of Hsp70 family of luminal chaperones which govern the gating of the aqueous pore [87].

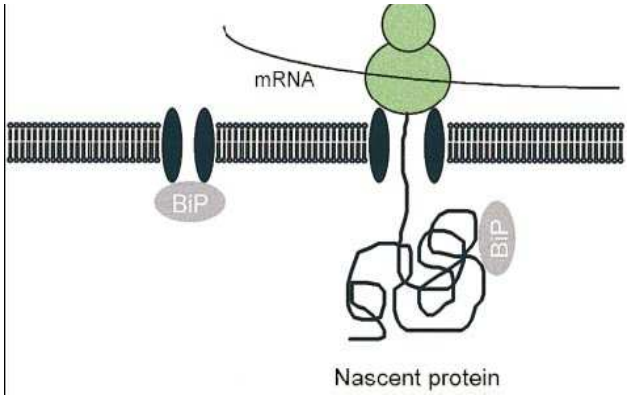


Figure1.4 Translocation of proteins across the ER. The inactive protein pore (left) is sealed with BiP whereas the active pore (right) assists the translocation of the unfolded nascent chain. When a ribosomal subunit binds to the cytoplasmic end of the translocon, the translocon widens (40 to 60 A°). The nascent chain is released into the pore where it starts folding perhaps assisted by BiP. Illustration taken from [87].

1.2.7 Secretory protein trafficking: from ER to Golgi

The secretory membrane system that modulates the delivery of lipids, carbohydrates and newly synthesised proteins to the cell surface is crucial for cell growth and homeostasis. This system comprises of distinct organelles including the ER, Golgi complex and the plasma membrane [88]. The ER, a continuous membrane structure found throughout the cytoplasm of eukaryotes, has evolutionary acquired a wide range of functions ranging from the integration of proteins into membrane, the translocation of secretory proteins across the membrane, the synthesis of phospholipids, modification, folding and degradation of proteins, the storage of calcium ions and their release into the cytosol [89, 90]. This complex structure, depending on intracellular localisation and cell type, can have tubular morphology or a sheet like structure [91]. Large sections of the ER (known as rough ER), are associated with ribosomes involved in the synthesis of proteins that are targeted to the lumen of the ER or to the membrane. In contrast, other parts (known as smooth ER) regulate ca^{2+} concentration, metabolise steroids

and carbohydrates, synthesise lipids, and are devoid from ribosomes [91]. Proteins that are targeted for secretion out of cells or are integrated into the plasma membrane must be correctly folded, assembled into oligomeric structures and glycosylated within the ER before being trafficked into vesicles for the Golgi complex. While in Golgi complex prior to the transport of proteins to the cell surface, the proteins are further modified and processed by glycosyltransferases or are returned to the ER [92]. The Golgi complex provides a mechanism for cells to distinguish various cargoes intended for different compartments, ranging from endosomes and lysosomes to the basolateral and apical plasma membrane. Most crucially the Golgi complex can act as a filter to separate proteins that should remain in the ER from those destined to be exported out of the cells [92, 93].

1.2.8 Protein retrieval in the ER

Newly synthesised proteins interact with luminal chaperones (examples include the protein disulfide isomerase, calnexin, calreticulin and BiP) upon their co-or post-translational insertion into the ER membrane. The role of these chaperones is to facilitate folding reactions necessary for protein oligomerisation and maturation [88]. These chaperones exclusively reside in the ER and must be separated and retained from their substrates upon cargo transport from the ER to Golgi complexes [94]. The retention of folding enzymes and ER-resident chaperones is brought about by the presence of a specific C-terminal sequence (KDEL tetrapeptide) that interacts with soluble receptors or the integral membrane [95]. The KDEL tetrapeptide is part of protein sequence that is recognised by a membrane-bound receptor for the KDEL tetrapeptide (Erd2) that binds to the ER-resident chaperones in the Golgi complex and therefore facilitate their retrieval to in the ER [96]. Incorrectly folded proteins are also

retained in the ER. It has been postulated that the retention of incorrectly folded proteins occur as a result of extensive interactions with the ER chaperones. This leads to the immobilisation and subsequent degradation of these proteins [97]. A receptor-mediated model of transport suggests that correctly folded proteins are recognised by the transport receptors which in turn target these proteins for secretion into transport vesicles. Next, proteins transported into Golgi and continue their transport. However, incorrectly folded protein or mis-folded proteins cannot be targeted by transport receptors and therefore aggregate within the ER [97].

1.2.9 Modulation of secretory pathway: unfolded protein response

The ER not only directly regulates the process of protein maturation, but also has a major role in transferring signals to the rest of the cell that may subsequently influence the rate of protein secretion [98]. It has been shown that within the ER there is a sensitive surveillance mechanism to ensure firstly that mis-folded proteins are prevented from exiting the secretory pathway and then secondly the mis-folded proteins are directed towards a degradation pathway (in the ER or cytosol) [99]. However, when homeostasis within the ER is disturbed (for instance, by accumulation or influx of mis-folded polypeptides that exceeds the folding capacity of the ER, or glucose or nutrient reduction and perturbations in ca^{2+} homeostasis) this leads to the activation of a signaling network known as unfolded protein response (UPR) [100]. The activation of UFR provides a check point for cells by which cell can survive and adapt or commit to a program of cell death under conditions of chronic stress [101]. In mammals, UPR is comprised of two parts: the transcriptional activation or up regulation of genes that increase the ER folding capacity (such as that of folding enzymes and ER

chaperones) and general translational attenuation of genes which overload the ER [102]. The activation of UPR occurs through an orchestrated action through transcription, involving ER-membrane bound UPR sensors, IRE1 (inositol requiring-1), ATF6 (activating transcription factor 6) and translation, involving the protein kinase-like ER kinase (PERK) [102].

In unstressed cells, BiP interacts with the luminal domain of ATF6, IRE1 and PERK. Upon aggregation of unfolded proteins in the ER lumen, IRE1 disassociate from BiP and dimerises to induce its RNase and kinase activity. This leads to the splicing of X-box DNA-binding protein (XBP1) mRNA, therefore creating a strong transcriptional activator [99, 103]. Genes whose induction requires the IRE1/XBP1 pathway includes the ER-associated degradation encoding gene. Similarly, ATF6 disassociates from BiP and is transported to the Golgi complex, where it is cleaved by S1 and S2 proteases to produce a cytosolic fragment that is transported to the nucleus in order to further up regulate transcription of UPR-responsive genes [99, 104]. Finally, disassociation of PERK from BiP results in its dimerisation and autophosphorylation. The dimerisation and autophosphorylation of PERK leads to the phosphorylation of eukaryotic translation initiation factor 2α (eIF2α) leading to general attenuation of translation initiation [99].

1.2.10 Modulation of secretory pathway: ER overload response

ER overload response (EOR) however, is a stress response to deal quickly with over-accumulation of synthesised protein in the ER membrane that leads to an increase in ca^{2+} permeability [98]. The EOR, just like the UPR requires a series of transcription factors and protein kinases that results in the adaptation in gene expression. However, the EOR uses separate mediators in the targeting and signaling to different genes to cause reaction [98]. The

EOR response causes activation of the κB kinase (IKK) which in turn leads to the degradation of κB inhibitor (I- κB). This induces the activation of transcription factor NF- κB (nuclear factor- κB) that promotes the expression of several genes encoding products involved in inflammatory responses or survival/death decisions [105].

1.3 Eukaryotic mRNA and untranslated regions (UTRs)

It has been estimated that in higher eukaryotic organisms a small fraction of the genetic material codes for proteins [106]. This suggests that most of the genomic DNA is implicated in the regulation of gene expression. This regulation can be exerted at the level of transcription or at the post-transcriptional level, in which the translation efficiency, the subcellular localisation of the transcript and the stability of the mRNA are regulated [107]. Transcriptional factors, enhancers, RNA polymerases, silencers and promoters regulate transcription of DNA to produce pre-mRNA molecules, which in turn undergo a series of stepwise processes to become mature mRNAs [106].

Pre-mRNA transcripts in eukaryotic organisms possess a common structural feature that include a special modified base at the 5' end known as the cap structure, 5' and 3'untranslated regions (UTRs), introns and exons, and a stretch of adenine residues at the 3'end known as poly (A) region. *Cis*-acting regulatory elements (signals) can occur within the coding region as well as the untranslated regions. Processes that lead to mRNA maturation include the removal of introns, addition of a cap structure to the 5'end of the first exon as well as the inclusion of 100-250 adenine residues at the 3'end. It is believed that most of the mRNA regulatory elements are found within the 5' and 3'UTRs. These regulatory elements mostly

function as binding sites for binding proteins and generate ribonucleoparticles (mRNPs) [108]. Whereas the primary role of the 5'UTR has been claimed to be mainly the regulation of mRNA translation, on the other hand the 3'UTR modulates several aspects of mRNA metabolism, such as subcellular localisation, translational efficiency and mRNA stability (Figure 1.5) [109, 110]. Interestingly, unlike DNA- mediated information (enhancers, promoters etc) which is essentially contained in the primary sequence, the regulatory motifs at the RNA level can be contained in the secondary as well as primary structure [107].

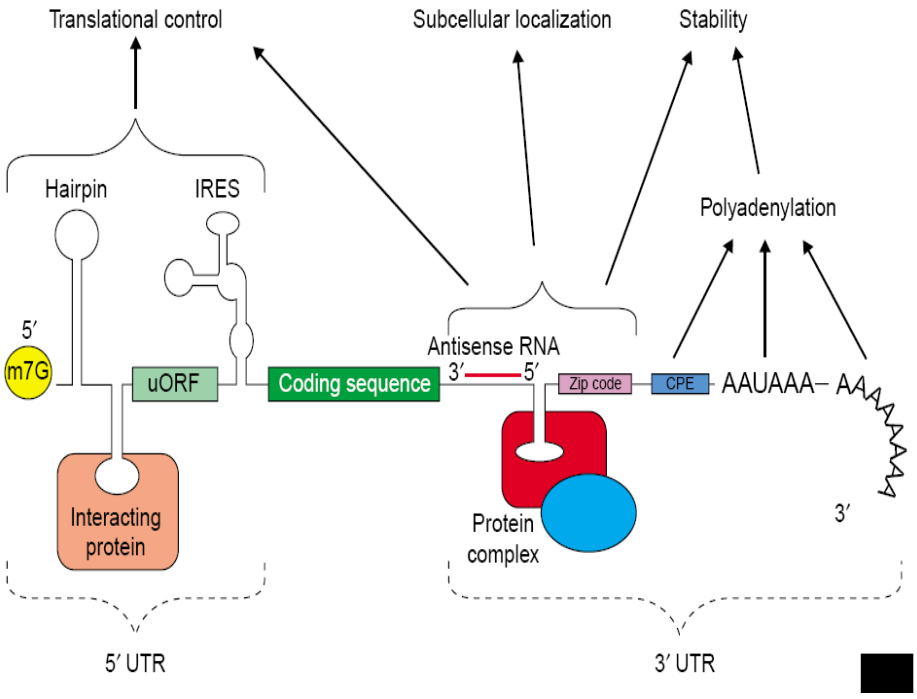


Figure 1.5 A eukaryotic mRNA and its general tripartite structure. Some possible elements that regulate post-transcriptional regulation are shown. 5' and 3'-untranslated regions mediate translation, subcellular localisation and mRNA stability through regulatory elements, such as cap-structure, hairpin and secondary structure, upstream open reading frames, internal ribosome entry sites (IRES), RNA protein interactions, the formation of multi-protein complexes, poly (A) tail and finally cytoplasmic polyadenylation elements. Illustration taken from[106].

1.3.1 Structural features of UTRs

Comparative analysis of partial and completed genome sequences has shown that 5'UTRs and 3'UTRs contain some conserved aspects. In general, 3'UTRs are longer than 5'UTRs. The average length of 3'UTRs varies, ranging from 50-800nts, whereas the average length of 5'UTRs is relatively constant over diverse taxonomic classes (100-200nts) [106]**.** Intriguingly, even the length of untranslated regions can differ within a species ranging from a dozen to a few thousand nucleotides [107]. It has even been claimed that a single nucleotide *in vitro* can act as a 5'UTR and initiate translation [111]. It has been reported that the intron content of gene regions corresponding to 5'UTRs is greater than that of 3'UTRs, implying that exons at 3'UTRs are perhaps much longer [112]. Furthermore, it has been found that the G+C content of UTRs may also vary. It is estimated that in warm-blooded vertebrates, G+C content for the 5'UTR is 60% whereas the 3'UTR is 45% [112].

1.3.2 UTRs and mRNA subcellular localisation

mRNA subcellular localisation is an important aspect of cell development that is mostly achieved by active transport of transcripts to various subcellular locations or by producing transcripts that differ in stability [113]. The localisation of mRNA within the cytoplasm plays a key role in dictating the polarity of the cell in both oocytes and somatic cells in addition to its key role in cell fate determination and pattern formation during embryogenic development in species such as *Drosophila*. The latter is a result of site-specific synthesis of proteins and the formation of a morphogen gradient, typically of a transcription factor which determines the overall body pattern after fertilisation [114-117].

Localisation of proteins through the targeting of mRNA rather than localising the protein itself has several advantages [118]: firstly by localising mRNA cells prevent protein translation elsewhere. This is highly important for cytoplasmic determinants, which are capable of altering the pattern of the embryo if translated in the wrong regions. For instance, mislocalisation of *oskar* and *nanos* in *Drosophila* causes the formation of a second abdomen in the region of the head and thorax. Secondly, it would be extremely efficient for the cells to localise a RNA molecule and hence perform several rounds of translation than targeting the synthesised protein multiple times. Finally by localising the mRNAs, cells can set up a protein gradient which can maintain or establish cell diversification and polarity. In many instances the key to translation of the mRNA transcript is the localisation of the transcript to the correct destination, which is modulated by *cis*-acting elements within the 3'UTR. For instance, translation of *oskar* mRNA in *Drosophila* is inhibited when it is mislocalised in the posterior pole of the embryo as a result of the interaction of *trans*-acting factors with three regions within the 3'UTR of *oskar* mRNA [119, 120]. Whichever way a transcript is localised, it must contain the targeting signal or the *cis*-acting localisation element. These elements are recognised by the *trans*-acting proteins that couple them to the localisation machinery [121-123]. The length of *cis*-acting elements may vary ranging from a few nucleotides to over 1 kb. It has been claimed that *cis*-acting elements are often located within the 3'UTR either as a combination of different elements or several copies of the same element [124]. It has also been reported that the *cis*-acting elements can also be situated in the 5'UTR as well as the coding region, in spite of the obvious necessity for the coding of amino acids [125, 126].

1.3.3 UTRs and control of translation efficiency

The structural features of the 5'UTR can have a major role in the regulation of mRNA translation. For example, mRNAs encoding proteins involved in developmental processes (transcription factors, growth factors or proto-oncogenes) often have 5'UTRs longer than average and they contribute to the fine regulation of their translation [106]. These 5'UTRs contain upstream open reading frames (uORFs), upstream initiation codons and stable secondary structures that can repress translation efficiency [127]. Interestingly, the inhibitory effect of the secondary structures can be overcome with an increase in the level of eIF4A, an RNA helicase that facilitates ribosome binding and its passage along the 5'UTR towards the initiation codon [128]. In addition, the UTRs regulate the efficiency of translation through interactions between *trans*-acting binding proteins and the sequence elements that are targets for these proteins. One of the best studied examples is the iron response element (IRE) in the 5'UTR region of mRNA coding for proteins involved in iron metabolism [129]. These elements impede translation by preventing the 40S subunit from carry out its scanning function[107, 130]**.**

The actual process of mRNA translation requires events that take place at both ends of the mRNA transcript. Broadly speaking it begins by the assembly of the ribosome subunit (40S) on the mRNA transcript [131]. The cap dependent translation in eukaryotic organisms initiates with the recognition of the m^7G cap structure at the 5'-end of mRNA transcript with the initiation factor eIF4F (4F). This heterotrimeric complex consists of 3 subunits: the cap binding protein eIF4E, eIF4A and eIF4G, where eIF4G associates not only with eIF4E and eIF4A but also with eIF3, a multi-subunit complex that associates with proteins interacting with the mRNA's 5'end and with the 40S ribosomal subunit [132]. Association of the 40S

subunit with multiple initiation factors (eIF1, 3 and 5) and also the $tRNA^{met}$-eIF2 complex forms a larger complex knows as the 43S pre-initiation complex that scans the 5'end of the mRNA transcript for the initiation codons. Furthermore, the association of eIF4A with another general translation initiation factor known as eIF4B in an ATP dependent manner leads to the unwinding of secondary structures in the 5'UTR region [133, 134].

Next, eIF5 triggers the hydrolysis of the GTP bound by the ternary complex. This results in the release of initiation factors and joining of the large ribosomal subunit (60S) to form the 80S ribosome that is capable of initiating elongation [135]. When the 43S pre-initiation complex encounters the AUG start codon, it initiates translation. However, it has been claimed that a great proportion of the 5'UTRs contain upstream AUGs ranging from 15% to 50%, the percentage dependent on the type of organism. This implies that a scanning model of the ribosome for the first AUG start codon selection is ignored in mRNA transcripts with multiple upstream AUG start codons [106].

Whenever the scanning subunit faces a poor initiation start codon which lacks the optimal context for translation initiation (Kozak sequence), it may bypass this furthest upstream codon and initiate translation at a more distal AUG start codon. This mechanism (leaky scanning) allows for a variety of proteins to be obtained from the same mRNA transcript [136, 137]. Furthermore, it has been discovered that there is a direct correlation between the length of the 5'UTR and the presence of multiple AUG codons. The longer the length of the 5'UTR, the greater the number of upstream AUGs found. The shorter the 5'UTR, the less number of AUGs present. This suggests that upstream AUGs may be involved in keeping the basal translation level of an mRNA low [138, 139].

Furthermore, the initiation of translation can also occur in a cap-independent manner in which the 40S ribosomal subunit is recruited to the close proximity of the AUG start codon without being influenced by the cap structure**.** In this mechanism the mRNA transcript of RNA and DNA viruses contains an element known as an internal ribosome entry site (IRES) that allows ribosomes to initiate effectively on highly structured 5'UTR regions. The IRES-dependent translation provides cells with a crucial mechanism for protein expression to continue where the cap-dependent translation is prohibited, for instance by structural elements (found within the 5'UTR) that prevent scanning by ribosomes [140]. Although the mechanism of cap-independent translation initiation was first discovered in the expression of encephalomyocarditis virus [141] and polio virus [142] since then it has been proposed that about 3-5% of cellular genes such as mRNAs of proteins implicated in regulating gene expression during cell growth, differentiation, development, cell cycle progression, stress response and apoptosis may utilise the IRES-dependent translation [143, 144]. It has also been reported that IRES predominantly becomes activated when in a situation such as DNA damage and cellular stress under which cap-dependent translation is greatly decreased [145].

Translational control can also be exerted by the 3'UTR. The 3'UTR can influence the translation of mRNA in several ways. It contains short sequences (motifs) that can interact with the specific RNA binding proteins which in turn influence translation efficiency [146, 147]. It is not surprising that the 3'UTRs are considered as important regions for regulation of translation by offering a variety of regulatory mechanisms given its proximity to the termination stop codon and the poly (A) tail region. Given the position of the 3'UTR and from a mechanistic point of view, it is very unlikely that this region is ever scanned by the scanning subunit. Therefore any interaction between the 3'UTR and the RNA binding

proteins could potentially have long lasting influence over translation and enabling the modulation at any given time [148].

It has been claimed that translation of mRNA transcripts can be influenced by the cytoplasmic changes in the poly (A) tail length, where an increase in length would normally correlate with greater efficiency of translation [149-152]. The poly (A) tail is thought to stimulate translation in coordination with the m^7G cap by circularising the mRNA transcript. This would lead to the association of the poly (A) tail binding protein (PABP) with the translation initiation factor eIF4G, which in turn binds to the cap binding protein eIF4E. This association between PABP and eIF4G is an essential step in translation regulation since evidence from several groups has shown that interfering with this interaction (for instance, by over expression of mutant eIF4G or the truncation of eIF4G) leads to a decrease in poly (A) mediated translation [153, 154]. The poly (A) tail and its length are modulated by elements within the 3'UTR. Polyadenylation can take place in the cytoplasm as well as in the nucleus. In order for cytoplasmic polyadenylation to take place, two elements within the 3'UTR are essential: the hexanucleotide polyadenylation signal (AAUAAA) and the nearby cytoplasmic polyadenylation element (CPE) where the latter is usually located within 20-30 nucleotides of the polyadenylation signal**.** It has been reported that CPE can only mediate polyadenylation efficiently if it is located within 100 nucleotides from the polyadenylation signal [155, 156].

1.3.4 UTRs and control of mRNA stability

A crucial step in the post-transcriptional regulation of gene expression is the control of mRNA stability and turnover. A variety of mechanisms have been suggested to define mRNA stability and turnover, such as the shortening of the poly (A) tail at the 3'-end or

equally the removal of the 5'cap structure [157]. It has been reported that in the 3'UTR of short-lived mRNA transcripts, such as those coding for lymphokines and cytokines, there are sequence elements rich in AU that have been identified. These AU-rich elements (ARE) that promote the shortening of the poly (A) tail, which results in degradation of the target mRNA [158, 159]. The ARE elements are normally found in the 3'UTRs of mRNA transcripts. The presence of these elements promotes instability of the mRNA transcript and functions as binding sites for RNA binding proteins [160, 161].

The ARE elements are grouped into 3 classes based on their functional and structural properties. Classes I and II share a common structural feature in which both classes contain multiple copies of the pentanucleotide AUUUA. This pentanucleotide sequence is absent in class III [162, 163]. Cytoplasmic shortening and deadenylation of the poly (A) is controlled by the class I AREs. Class I AREs are mainly found in mRNA transcripts encoding nuclear transcription factors (such as c-Myc and c-Fos) and some cytokines such as interleukin 4 and 6 [106, 164]. Unlike class I synchronous deadenylation of mRNA, class II AREs deadenylate the poly (A) tail asynchronously at a different time and rate. This produces mRNAs without poly (A) tails. The main characteristic of class II is the presence of the pentanucleotide AUUUA with a minimum of 3 in tandem. Furthermore, an AU-rich region is found upstream of these nucleotides [162]. mRNA transcripts encoding class III AREs such as c-Jun lack the AUUUA pentamer, instead contain a U-rich segment [165].

1.4 mRNA-binding proteins

Post-transcriptional control of mRNA stability, translation, localisation, polyadenylation and splicing is commonly achieved by interaction between the non coding region of an mRNA

and the RNA-binding proteins (RBPs) [166]. Although all RBPs can bind to the RNA, the interaction is achieved with different levels of affinity and RNA-sequence specificities (RNA-binding domains) [167]. RNA-binding proteins in general, interact with the sequences within the 3'UTR of target mRNA transcripts, however there is evidence of interaction between the RNA binding proteins and the 5'UTR or even the coding region [168]. It has been reported that in the *C.elegans* genome, there are approximately 500 genes (2%) that code for RBPs with one or more known RNA-binding domains such as the K Homology domain, RNA Recognition Motif (RRM), zinc finger domain**.** While some domains only predict RNA-binding function, others may imply a further role in the molecular function of RBPs such as the DEAD/DEAH box for RNA helicase activity [169]. The RNA Recognition Motif (also known as RNA Binding Domain) is very often found within a single protein as multiple repeats as seen in the poly (A) binding protein (PABP) [170]. This enables recognition of more complex and larger RNA targets and enhancing the specificity of binding [171].

1.4.1 The diverse function of mRNA-binding proteins

1.4.1.1 mRNA stabilisers: HU proteins

The HU family of RNA-binding proteins, is widely becoming recognised as a regulator of post-transcriptional gene expression. It has been reported that the members of this family of RNA-binding proteins (HU R, HU B, HU C and HU D) interact with the target mRNAs containing the AU and U-rich sequences through their three RNA-recognition motifs. This interaction leads to modification of gene expression as a result of altering the stability of the target mRNA transcripts [172]. It has also been shown that the HU family of proteins can

increase cell division by stabilising mRNA transcripts that encode cell cycle control and proliferation genes such as c-fos, cyclin A and B1 [172, 173].

1.4.1.2 mRNA destablisers: KSRP (K homology splicing regulatory protein)

There are many RNA-binding proteins that promote mRNA degradation such as KSRP. This RNA-binding protein is a member of the far upstream binding protein (FBP) family which contains four RNA binding K homology (KH) motifs. It has also been shown that KSRP promotes degradation by associating with the ARE-containing mRNAs via an exosome mediated pathway [174].

1.4.1.3 mRNA localisers: She2 and She3

ASH1 mRNA (*S.cerevisiae*) localisation to the bud of daughter cells during cell division is an example where mRNA binding proteins facilitate the localisation process. This localisation is achieved by association with two more proteins, She2 and She3 [175]. It has been reported that She2 associates as a dimer with the localisation elements in the 3'UTR and coding region of ASH1 mRNA [176].

1.4.1.4 mRNA exporters: TAP/NXF1:p15 heterodimer

mRNAs are exported from the nucleus to the cytoplasm as soon as the pre-mRNA processing (transcription, splicing and 3'end formation) is finalized [177]. Therefore cells require a mechanism to ensure the export of fully processed pre-mRNA out of the nucleus. mRNA

binding proteins (heterodimer TAP/NXF1: p15) are an example of where their association with the constitutive transport element (CTE) facilitate mRNA export [167].

1.4.1.5 mRNA translational repressors : TIAR

It has been claimed that in the absence of environmental stress, mRNA binding protein TIAR acts a translational repressor for ARE-containing mRNA targets such as the tumor necrosis factor alpha (TNF-α). TIAR achieves this by associating with the 40S ribosomal subunit and initiation factors eIF1 and 3. This leads to formation of the preinitiation complex (containing the small ribosomal subunit 40S but lacking the initiation factors eIF2 and 5) that is translationally inactive [178, 179]. It has also been reported that under environmental stress, TIAR can act as a transient repressor for general mRNA pools [179].

1.5 Gaussia princeps and Luciferase reporter system

Gaussia princeps is a marine organism with many secretory glands capable of producing Luciferase in response to environmental signals. This copepod is 10mm in length and lives in the Pacific Ocean. The secretory glands of this organism are made from one single cell; however its cytoplasm is packed with many secretory vesicles. The secretory vesicles of this organism are released through a pore to the plasma membrane in response to nerve stimuli. It has been reported that the secretory vesicles of *Gaussia princeps* contain both the Luciferase and its substrate, Coelentrazine. Luciferase and its substrate react only when co-factors Ca^{2+} and O_2 are present [180].

It has been postulated that this organism uses bioluminescence as a defence mechanism when approached by predators or to attract and draw the attention of a mate. Amongst other Luciferase reporter systems, the *Gaussia* Luciferase is the smallest identified Luciferase consisting of 185 amino acids and has a molecular weight of 19.8 kDa [181]. The naturally secreted *Gaussia* Luciferase can be readily expressed in mammalian cell lines and emits light at a peak of 480nm [182]. Although this system is exceedingly costly to run and requires a specialised luminometer to measure Luciferase activity in a short space of time, there are many advantages to the *Gaussia* Luciferase compared to that in other commercially available reporter systems. For example, the *Gaussia* Luciferase is a secretory protein with a signal peptide, therefore upon synthesis it is secreted out of the cell. This is enormously beneficial and time saving since the lysis of the cell is not required. Furthermore, it has been reported that the bioluminescence produced by the *Gaussia* Luciferase is 1000-fold stronger in signal intensity than other reporters such as *Renilla* and firefly Luciferases. In addition, it allows one to perform sensitive assays due to the stability of the secreted protein [181].

1.6 The aims of this project

The aims of this project were to study the extent to which modifications to the signal peptide/3'UTR sequences affect the production of recombinant reporter protein and the underlying mechanisms. A series of novel constructs containing an appropriate Luciferase reporter and various signal peptide/3'UTR were created and expressed in an appropriate mammalian cell line (CHO AA8 Tet-Off). To study whether increasing the hydrophobicity of signal peptide would influence activity of the reporter protein, the Chymotrypsinogen signal peptide hydrophobicity was increased and the reporter activity measured. In addition, preliminary results from the JEH laboratory suggested that the Albumin 3'UTR in combination with an appropriate signal peptide promotes Luciferase activity more readily than the native *Gaussia* 3'UTR; it was not known whether a particular region/motif within the Albumin 3'UTR is responsible for the higher reporter activity and Luciferase secretion. To examine this, a series of constructs with the mutated Albumin 3'UTR were created and subsequently the secreted Luciferase activity and mRNA expression was measured. Finally, the project aimed at studying the effects of protein(s) binding to the Albumin 3'UTR and the extent to which the reporter activity and specific protein binding to the Albumin 3'UTR can be correlated.

2 Materials and methods

2.1 Chemicals

Name	Supplier
Acrylamide/bis-Acrylamide (40% stock solution) 19:1	Sigma
Agar	Sigma
Anti Rabit IgG (whole molecule) Peroxidase Antibody	Sigma
Ammonium Persulphate 98%	Sigma
ATP, [α-32P]- 800Ci/mmol 10mCi/ml , 250 µCi	PerkinElmer
Biotin-16-UTP (250nmol)	Roche
Biotaq^{TM} DNA Polymerase	Bioline
Blue/Orange Loading Dye,6x	Promega
Bovine Serum Albumin (10mg/ml)	New England BioLab
Bradford Reagent	Sigma-Aldrich
Bromophenol Blue	Sigma
Chloroform	Sigma
Coeloentrazine	UniTargetingResearch AS
ColorBurst^{TM} Electrophoresis Marker	Sigma
Complete Mini,EDTA-free Protease Inhibitor	Roche
CUG-BP1 (Rabbit polyclonal) Antibody	Abcam
Diethyl Pyrocarbonate (DEPC)	Sigma
Dimethylsulphoxide (DMSO)	Sigma-Aldrich
Dithiothreitol (DTT)	Invitrogen
dNTP mix	Bioline

DMEM Medium	Invitrogen
α- MEM Medium (optimum medium)	Invitrogen
Ethanol	Sigma
Ethidium Bromide (50 µg/µl)	Sigma
Fetal Calf serum	Sigma
Ficoll	Sigma
Glycerol 99.5%	Sigma-Aldrich
Glycine	Signam-Aldrich
ECL solution	GE Health care
Isopropanol 99%	Sigma
Lipofectamin ^{TM} 2000	Invitrogen
Magnesium Chloride Solution 1.M	Sigma
N,N,N',N'-tetramethylethylene-diamine (TEMED)	Sigma
Nonidet P-40 (NP-40)	Sigma
Optiphase Hisafe 2 (scintillation liquid)	PerkinElmer
Phosphate buffered Saline	Invitrogen
Phenol:Chloroform:Isoamyl Alcohol (25:24:1)	Sigma
(Saturated with 10mM Tris, Ph 8.0,1mM EDTA)	
Powder Milk	Marvel
RNase A (7000 u/ml)	Qiagen
RNase T1 (1000 u/µl)	Ambion
RNasin Ribonuclease Inhibitor (40u/ul)	Promega
Sodium Chloride	Fluka

Sodium Chloride Solution 5M	Sigma
Sodium Citrate	Sigma
Sodium Dodecyl Sulphate (SDS)	Sigma
Stealth RNAi	Invitrogen
Stealth RNAi (Negative Universal Control)	Invitrogen
TAE: Tris-Acetae EDTA	Promega
TBE: Tris-Borate EDTA	Sigma
Trizol	Invitrogen
Trypan Blue	Sigma
Trypton	Fluka
Trypsin 10x solution(0.5% Trypsin and 0.2% EDTA)	Sigma
Chloride	Sigma
Tween 20	Sigma
UltraPure™ Agarose	Invitrogen
Yeast Extract	Fluka
Yeast tRNA (10mg/ml)	Ambion

2.2 Antibiotics

Name	**Supplier**
Ampicillin Sodium Salt	Sigma-Aldrich
Gentamycin (50 mg/ml)	Invitrogen
Hygromycin B (50 mg/ml)	Invitrogen
Penicillin-Streptomycin mix	Invitrogen

(containing 10,000 U of Penicillin G Sodium and 10,000 µg of Streptomycin Sulphate in 0.85% Saline /ml)

2.3 Commercial kits

Name	Supplier
Colloidal Blue Staining Kit	Invitrogen
Expand High Fidelity PCR System	Roche
Light Cycler 480 SYBR Green I Master	Roche
MEGAshortscript™ T7 Kit	Ambion
MinElute Gel Extraction kit	Qiagen
Miniprep kit	Eppendorf
QIAquick PCR Purification Kit	Qiagen
Primer for cDNA synthesis p (dT) 15	Roche
Rapid DNA Ligation Kit	Roche
Transcriptor Reverse Transcriptase, 250 U	Roche

2.4 Buffers

Binding buffer (EMSA)

½ tablet EDTA-free complete protease inhibitor

10 µl 1M DTT

4.44 ml DEPC-treated water

0.5 ml of 10x 40 mM lysis buffer

10x 40 mM NaCl lysis buffer (EMSA)

1.3 M NaCl

50 mM $MgCl_2$

0.3 M Tris-HCl, pH 7.6

130 mM lysis buffer 1 (S-100 CHO cell extraction)

130 mM NaCl

5 mM $MgCl_2$

30 mM Tris-HCl, pH 7.6

2 mM DTT

1 minitab EDTA-free protease inhibitor

130 mM lysis buffer 2 (S-100 CHO cell extraction)

130 mM NaCl

5 mM $MgCl_2$

30 mM Tris-HCl, pH 7.6

0.5% [v/v] NP-40

40 mM lysis buffer

40 mM NaCl

5 mM $MgCl_2$

30 mM Tris-HCl, pH 7.6

1x lysis buffer (CHO S 100 cell extraction)

0.5 ml of 10x 40 mM NaCl lysis buffer

4.5 ml of DEPC-treated water

10 µl of 1M DTT

½ minitab EDTA-free protease inhibitor

25 µl of NP-40 (0.5%)

Protein sample buffer (5x)

2.25 ml of 1M Tris (pH 6.8)

5 ml of Glycerol

0.5 g of Sodium Dodecyl Sulphate (SDS)

5 mg of Bromophenol Blue

2.5 ml of 1M Dithiothreitol

5% β-Mercaptoethanol

Protein electrophoresis running buffer (5x) 1L

60.6 g of Tris Base

144.1 g of Glycine

5 g of Sodium Dodecyl Sulphate (SDS)

Transfer buffer

80% (v/v) 1X protein electrophoresis running buffer

20% (v/v) methanol

Membrane wash solution

0.05% (v/v) Tween-20 in 1x PBS

Blocking solution

5% (w/v) milk powder in membrane wash solution

Lysis buffer (*Gaussia* Luciferase measurement)

10 mM TrisHCl, pH 7.4

10 mM NaCl

1.5 mM $MgCl_2$

0.5% NP-40

Renilla dilution buffer

0.5M NaCl

1M K_2HPO_4 } 0.1M Potassium Phosphate, pH 7.6

1M KH_2PO_4 }

1mM EDTA, pH 8.0

0.02% BSA (w/v)

Adjust the pH to 7.6

Mili-Q H_2O to final volume of 100ml

Storage: at 4˚C after autoclaving

Renilla dilution solution

Required amount of Renilla buffer + 1 µl of GAR2 stabiliser (*Gaussia* Luciferase assay reagent) /50 µl of Renilla buffer

Coelentrazine (CTZ) substrate buffer

50 mM Tris-HCl, pH 7.6

10 mM EDTA, pH 8.0

0.5 M NaCl

200 µg of coelentrazine

Storage: at 4˚C after autoclaving

2x dissociation buffer

90 mM Tris-HCl, pH 6.8

20% [w/v] Glycerol

2% [w/v] SDS

2% [w/v] β-mercaptoethanol

0.02% [w/v] Bromophenol Blue

0.5x SSC Buffer (protein identification)

4.38 mg/ml of NaCl

2.205 mg/ml of Sodium Citrate, pH 7.0

DEPC-treated water

dH_2O to final volume of 250 ml

0.05% of DEPC [v/v]

Sterilised by autoclaving

2.5 Growth medium

Bacteria growth medium

LB (Lysogeny Broth) Agar

Preparation of 500 ml

10 g of Agar (2%)

5 g of Tryptone

5 g of NaCl

2.5 g of Yeast Extract

dH_2O to final volume of 500 ml

Sterilised by autoclaving

LB (Lysogeny Broth) Medium

Preparation of 500 ml

5 g of Tryptone

5 g of NaCl

2.5 g of Yeast Extract

dH_2O to final volume of 500 ml

Sterilised by autoclaving

Cell culture growth medium

500 ml of DMEM Medium

50 ml of FCS (10%)

5 ml of Penicillin/Streptomycin mix (1%)

300 µl of Gentamycin

The growth medium used for transfected cells contains 400 µg/ml Hygromycin B.

Medium for freezing cells

Preparation of 50 ml

45 ml of FCS

5 ml of Dimethylsulphoxide (DMSO)

2.6 Enzymes

Enzyme	Supplier
Transcriptor Reverse Transcriptase, 250 U	Roche
Protector RNase Inhibitor 2000 U	Roche
*Acc*I (20000 U/ml)	New England BioLabs
*BamH*I (20000 U/ml)	New England BioLabs
*Pflm*I (8,000 U/ml)	New England BioLabs

2.7 Electrophoresis and agarose gels

Preparation of 1% gel

0.250 g of Agarose

25 ml of TAE

1 µl of Ethidium Bromide

Acrylamide gel:

Urea Stock

18.9 g of Urea

9 ml of 5x TBE

Adjust to final volume of 45 ml of DEPC treated- water

Acrylamide stock

8.4 g of Urea

10 ml of 40% Acrylamide/bis (19:1)

Adjust to final volume of 20 ml with DEPC- treated-water

To make gel:

7.5 ml of Urea Stock

2.5 ml of Acrylamide stock

10 µl of TEMED

80 µl of APS (10%) solution [w/v]

SDS PAGE GEL

Separating gel

	10%	12.5%
40% Acrylamide /bis - Acrylamide Stock (37.5:1)	1.75ml	2.19 ml
2.5x Separating Gel Buffer [1.875 M Tris Base pH8.9, 0.25% SDS]	2.8ml	2.19 ml
Distilled Water	2.4 ml	1.94 ml
TEMED	6 µl	6 µl
APS (10%) solution [w/v]	65 µl	65 µl

Stacking gel

40% Acrylamide /bis-Acrylamide Stock (37.5:1)	0.25 ml
5x Stacking Buffer	0.4 ml
[0.3 M Tris, pH6.7, 0.5% SDS]	
Distilled Water	1.32 ml
TEMED	2.5 µl
APS (10%) solution [w/v]	18 µl

2.8 Bacterial Strain

One Shot ® TOP10 Competent Cells (Invitrogen)

Genotype: F- *mcr*A Δ(*mrr-hsd*RMS-*mcr*BC) φ80*lac*ZΔM15 Δ*lac*X74 *rec*A1 *ara*D139 Δ(*araleu*) 7697 *gal*U *gal*K *rps*L (StrR) *end*A1 *nup*G

2.9 Cell line

CHO AA8 Tet-Off (Clontech)

2.10 Primers and oligonucleotides

Table 2.1 primers used for general purpose PCRs (Chapter 3 and 4).

oligo	Sequences (all in 5’ to 3’-direction)
Gau.for 5’	TCCTAATAGAATACTGCATAACTG
Gau.rev 5'	GGATTGCATAAATTATATTTATAGGAATTAC
G3Da.for 5’	CCAAGATGAAGAAGTTCATC CCAGG
CDs.for 5’	GCGACCTTTGCCAGCAAGATCC
pTRE2_rev 5’	CCATTCTAAACAACACCCTG
Glob.rev 5’	GTATTACCGCCTTTGAGTGAGC
Chy.rev 5’	CGTTGTTCTCGGTGGGCTTGGAATAAGCCGAGGTAA
Chy.for 5’	GTATCTTCTGGCAGGGAAAATGAAGTGGGTAACCTT
pTRE2.for 5’	CGCCTGGAGACGCCATC
pTRE2a.for 5’	GGCTCCTCTTCTGCTGGGCC
pTRE2a.rev 5’	GGCGCAGAAGAGGAGCC
pTRE2b.for 5’	CCTCCTGGGTGCCACCTTCG
pTRE2b.rev 5’	CGAAGGTGGCACCCAGGAGG
Alb3.rev 5’	AGATTCTTTCCATTTTTTATTAATTGAAGC
G3Db.rev 5’	TTAGTCACCACCGGCCCCCTTG
BCD.rev 5’	ACCGCCTTTGAGTGAGCTGATACC

Table 2.2 Primers used for the Albumin 3'UTR deletion analysis work (Chapter 3 and 5).

oligo	Sequences (all in 5' to 3'- direction)
Alb151.100For 5'	TTCTTTTTCGTTGGTGTAAAGCCTCTGTGCTTCAAT
Alb.151.50.For 5'	CCATGAGAATAAGAGAAAGAAACTCTGTGCTTCAAT
Alb50.151.Rev 5'	CCATTTTTTATTAATTGAAGCACAGAGTTTCTTTCT
Alb50.100.Rev 5'	ATGTTTTTTAGACAGGGTGTTGTTTCTTTCTCTTAT
Alb51.luc.For 5'	CAAGGGGGCCGGTGGTGACTAAATGAAGATCAAAAG
Alb100.151.Rev 5'	CCATTTTTTATTAATTGAAGCACAGAGGCTTTACAC

Table 2.3 Primers used for performing the Real-Time PCR experiment (Chapter 4, 5 and 6).

Diagnostic oligo	Sequences (all in 5' to 3'- direction)
RTluc.for 5′	GTTCTGACCTGCTCAAGAAGTGG
G3Db.rev 5′	TTAGTCACCACCGGCCCCCTTG
CHOGAPDH .For 5'	ATGGTGAAGGTCGGCGTGAACG
CHOGAPDH.rev 5'	GGTCATTGATGGCAACAACTTCCTCTTTGCC

Table 2.4 Primers used for performing seamless cloning procedure (Chapter 3).

oligo	Sequences (all in 5' to 3'- direction)
5'UTR.for 5'	GGCGTGTACGGTGGGAGGCC
5'UTR.rev 5'	TTTCCCTGCCAGAAGATAC
CDC.for 5'	AAGCCCACCGAGAACAACGAAGACTTCAACATCG
CDC.rev 5'	TTAGTCACCACCGGCCCCCTTGATCTTGTCC
OliAlbSP.for 5'	ATGAAGTGGGTAACCTTTAT TTCCCTTCT

OliAlbSP.rev 5’	GGAATAAGCCGAGCTAAAGAGAAAAAGAAGAGAAAA AGAAGGGAAATAAAGG
AlbSP.for 5’	GTATCTTCTGGCAGGGAAAATGAAGTGGGTAACCTTTTTCC
AlbSP.rev 5’	GGAATAAGCCGAGCTAAAGAGCGTTGTTCTCGGTGGGCTT
Gausig.for 5’	GGAGTCAAAGTTCTGTTTGCC
Alb3’UTR.for 5’	CATCTACATTTAAAAGCATCTCAGC

Table 2.5 Primers used for the protein binding work. Regions in blue refer to T7 minimum promoter sequence (Chapter 6).

oligo	**Sequences (all in 5’ to 3’- direction)**
Full3UTR.for 5’	TAATACGACTCACTATAGGCATCTACATTTAAAAGC
Full3UTR.rev 5’	AGATTCTTTCCATTTTTTATTAATTGAAGC
Del(1-50).for 5’	TAATACGACTCACTATAGGATGAAGATCAAAAGC
Del(1-100).for 5’	TAATACGACTCACTATAGGCAACACCCTGTCTAAAAAAC
Del(1-150).for 5’	TAATACGACTCACTATAGGCTCTGTGCTTCAATTAATA
Del(1-50&101-150).for 5’	TAATACGACTCACTATAGGATGAAGATCAAAAGCTTATTC
GloFor141_T7 5’	TAATACGACTCACTATAGGGCAGGCTGCTGGTTGTCTAC
GloRev368 5	CCTGAAGTTCTCAGGATCC

2.11 Plasmids

Names	Source
pTRE2hygGG*G*G	UniTargetingResearch AS, Bergen
pTRE2hygGXG*G	UniTargetingResearch AS, Bergen
pTRE2hygGX1G*G	UniTargetingResearch AS, Bergen
pTRE2hygGX2G*G	UniTargetingResearch AS, Bergen
pTRE2hygGG*G*Alb 3Δβ	UniTargetingResearch AS, Bergen
pTRE2hygGXG*A5 3Δβ	UniTargetingResearch AS, Bergen
pTRE2hygGX1G*A5 3Δβ	UniTargetingResearch AS, Bergen
pTRE2hygGX2G*A5 3Δβ	UniTargetingResearch AS, Bergen
pTRE2hygGG*G*Alb 3Δβ (Δ1-50)	JEH Laboratory
pTRE2hygGG*G*Alb 3Δβ (Δ1-100)	JEH Laboratory
pTRE2hygGG*G*Alb 3Δβ (Δ1-150)	JEH Laboratory
pTRE2hygGG*G*Alb 3Δβ (Δ51-100)	JEH Laboratory
pTRE2hygGG*G*Alb 3Δβ (Δ101-150)	JEH Laboratory
pTRE2hyg	Clontech

2.12 Storage of materials

Bacteria strain (Top10cells)	-80˚C
Buffers and solutions	Room temperature

Cell lines	Liquid Nitrogen
Enzymes	-20˚C
Enzyme buffers	-20˚C
Growth media	+4˚C
Luciferase medium and cell extract samples	-80˚C
Oligonucleotides	-20˚C
Plasmids	-80˚C

2.13 DNA techniques

2.13.1 Ligation

The Rapid DNA ligation kit (Roche) was used in this study to ligate the DNA fragments to be inserted with double digested vector. To maximise the chance of ligation, various ratios of insert to vector (3:1, 5:1 and 6:1) were used. The calculation to obtain the ratio was as follows:

$$\text{Ratio of insert} = \frac{\text{ng of vector x kb of insert x 3}}{\text{kb of vector}}$$

Where ng is nanogram and kb is kilobase. To set up a ligation reaction the following steps were carried out:

1. 1x DNA dilution buffer was added to digested vector and insert (according to the ratio) to a final volume of 10 µl.
2. 10 µl of T4 DNA ligation buffer was mixed immediately before use and added to the reaction.
3. After thorough mixing, 1 µl of T4 DNA ligase was added to the reaction vial and mixed.
4. The reaction mix was incubated for 5 min at RT.

The ligation reaction products were used directly for transformation of competent cells (One Shot ® TOP10 *E.coli*).

2.13.2 Electrophoresis of DNA

Agarose gel electrophoresis was used to separate digestion products and to purify PCR products. To make the gel and achieve the desired percentage, the appropriate amount of agarose powder was mixed with 1x TAE electrophoresis buffer and heated in a microwave oven until it was completely melted. The melted agarose was allowed to cool down for 5-10 min. At this point, 1 µl of ethidium bromide (10 µg/µl) was added to the melted agarose to facilitate visualisation. The solution was poured into an appropriate gel casting tray containing comb/s and it was allowed to cool down for a further 15-20 min and solidify at RT. After the gel was solidified, the comb/s was removed and 1x TAE electrophoresis buffer was poured in to fill up the gel tank. Then 4 µl of samples and 1 µl loading buffer were loaded into wells of the gel. The electrophoresis was performed with an appropriate voltage (60 V) for 30-35 min. Finally, the gel was photographed under UV-light.

2.13.3 Restriction digestion

Restriction sites for *BamHI* and *Acc*I were located in the vector sequences by using the web-based program NEB cutter 2.0 (http://tool.neb.com/NEBcutter2/index/php). The vector and seamless cloning product were digested with the same restriction enzymes.

General reaction mix for a typical digestion is shown below

10x Buffer (according to the enzyme)	= 5 µl
BSA (100x)	= 0.5 µl
Restriction digestion enzyme (s)	= 1 µl (of each)
DNA (miniprep product)	=10 µl
dH_2O (final volume of 50µl)	= 32.5 µl

The reaction mix was incubated at 37ºC for 105 to 120 min.

2.13.4 Gel extraction

After separation of DNA fragments by gel electrophoresis, extraction of DNA fragment was carried out by using a MinElute Gel Extraction kit (Qiagen) to obtain the desired fragments of restriction enzyme digestion products. The digestion products were separated on 0.8% agarose gels. The visualisation of DNA took place under UV-light and gel extraction was carried out as according to the manufacturer's instruction.

- The desired band was cut out of the gel and the fragment weighed.
- 3 volumes of binding buffer were added for every volume of excised gel fragment.
- The gel slice was placed in the binding buffer and incubated at 50ºC for 5-10 min.

- One volume of isopropanol was added and 800 µl of mixture added to a spin column in a 2 ml collection tube and centrifuged for 1 min at 10000 g, the filtrate was discarded and 750 µl of wash buffer was added to the spin column.
- The spin column was centrifuged for 1 min at 10000 g, the filtrate was discarded and the spin column centrifuged again for an additional 1 min.
- The spin column was placed in a fresh collection tube, 30 µl of elution buffer added to the centre of the column and the column centrifuged for 1 min at 10000 g.
- Finally the spin column was discarded and the purified product was collected from the bottom of the collection tube.

2.13.5 Polymerase Chain Reaction (PCR)

PCR was performed to amplify the DNA fragments in vectors to either verify the presence of specific sequences or as an integral part of the seamless cloning procedure. The Eppendorf Mastercycler Gradient PCR machine was used throughout.

General reaction mix for a typical PCR reaction (Expand High Fidelity PCR Kit)

10x buffer	=5µl
dNTP mix (2.5 mM)	=2µl
Forward primer (10 pM)	=5µl
Reverse primer (10 pM)	=5µl
DNA template	=1µl
DNA polymerase	=1µl
dH_2O	=31µl

General thermocycling program for a Typical PCR

Step	TM	Time (m/s)		cycle
1	95 ºC	2m		1
2	95 ºC	30s		30
2	58 ºC	30s	30	30
2	72 ºC	1m		30
3	72 ºC	5m		1

Finally hold at 4ºC

The optimal annealing temperature was obtained by performing gradient PCR.

2.13.6 PCR purification

Throughout this study PCR products were purified using QIAquick PCR purification kit and this was used according to the manufacturer's instructions. To purify PCR products, 5 volumes of buffer BP was added to one volume of PCR sample. A QIAquick spin column was placed into a 2ml collection tube and the sample mix transferred to the spin column. The column mix was centrifuged at 13000 rpm for 30-60 s. Next, the flow-through was discarded and the spin column placed back into the collection tube. To wash, 750 µl of buffer PE was added to the spin column and the column centrifuged at 13000 rpm for 30-60 s. The flow-through was discarded, the spin column placed back into the collection tube and centrifuged for additional 1 min. The spin column was placed into a microcentrifuge tube and 50µl of elution buffer EB transferred to the centre of spin column. After 1 minute of incubation, the microcentrifuge tube was centrifuged at 13000 rpm for 30-60 s. The purified PCR product was collected from the bottom of the tube.

2.13.7 Quantification of DNA

DNA was diluted in dH_2O water (1:50) and the absorption was measured spectrophotometrically using a Biophotometer™ (Eppendorf®). The calculation of concentration was based on the knowledge that an OD=1.0 at 260nm corresponds to 50 µg/ml DNA (double stranded) or 37 µg/ml DNA (single stranded). Concentrations of DNA were expressed as µg/ml and a good quality sample was judged to be characterised by a 260/280 ratio of 1.7-1.9.

2.13.8 DNA sequencing

Sequencing of all the constructs was carried out by MWG bio-tech (Germany). Prior to the sequencing, ~1 µg of DNA samples were placed in microcentrifuge tubes and dried for 30-45 min in a desiccator. Finally, 10 µl of forward and 10 µl of reverse primers per reaction were sent to MWG along with the samples.

2.14 Bacterial techniques

2.14.1 Transformation of *E. coli* cells (One Shot® TOP 10)

TOP10 competent cells were thawed on ice prior to transformation. For each reaction, 20-30 µl of competent cells were used. 0.5 µl of plasmid (irrespective of concentration) was used for a single transformation experiment. To each aliquot of TOP10 cells, 0.5 µl of chosen plasmid was added. As a negative control, one aliquot of cells was left without plasmid to check that the selection procedure is accurate and that there is no background resistance. The

cells were tapped gently to mix and immediately kept on ice for a minimum of 30 min. The mixture was heat shocked for 30 s at 42ºC and then immediately placed back on ice. 80 µl of pre-warmed SOC medium was added to the reaction mix. Tube(s) were taped together and placed on their side in a shaker at 37ºC for 60 min. 50 µl of reaction mix was then spread on pre-made LB agar containing Ampicillin with a final concentration of 100 µg/ml. This was carried out with a continuous flame nearby to prevent contamination. Using a glass spreader, cells were spread evenly across the agar plates to cover the whole surface with a glass spreader. The spreader was flamed for 30 s prior to use, then placed on agar to cool it down. Plates were incubated overnight at 37ºC.

2.14.2 Miniprep, small scale plasmid purification

Eppendorf miniprep kits were used for small scale plasmid preparations following the manufacturer's instructions. A single colony from the overnight plates was selected and placed in a tube containing 10 ml of LB medium and of 10 µl of Ampicillin with a final concentration of 100 µg/ml. The tube was incubated overnight at 37ºC with shaking at 200 rpm. 1.5 ml of overnight culture was centrifuged at 13000 rpm for 5 min. The supernatant fluid was discarded. 0.2 ml of buffer P1 containing RNase was then added and the pellet and the buffer mixed a few times with a pipette. In order to lyse the bacteria 0.250 ml of buffer P2 was added. Tubes were mixed gently by inversion until the colour changed to blue. 0.35 ml of buffer P3 was added next, mixed by inversion and centrifuged for 10 min at 13000 rpm. The supernatant fluid was transferred to the columns (ideally 850 µl is taken up and poured into the column). The supernatant fluid was then centrifuged for 1min at 13000 rpm. Next, 0.5 ml of buffer PB was added and centrifuged for 1 min at 13000 rpm. The liquid at the

bottom of columns was discarded and 0.75 ml of buffer PE was added next. The columns were centrifuged for 1 min and the liquid discarded. This step was repeated. The columns were then placed in 1.5 ml microcentrifuge tubes and 50 µl of elution buffer was added. The elution buffer was added to the centre of column and left for 1 min and then centrifuged for 1 min at 13000 rpm. Finally, the purified plasmid was collected from the bottom of microcentrifuge tube. The column was discarded.

2.15 Cell culture techniques

2.15.1 General cultivation of cells

CHO AA8 Tet-Off cells (Clontech) were grown at 37ºC in an atmosphere containing 5% CO_2. Cells were cultivated in T-25 or 75 flasks in a monolayer and fed with DMEM medium (Dulbecco's Modified Eagle Medium + GlutaMax) supplemented with 10% [v/v] FCS, 1% [v/v] penicillin-streptomycin mix and 300 µl of Gentamycin. Cells were grown until they reached ~90% confluency then subcultured (this normally takes 2-3 days for non-transfected CHO cells). They were washed twice with 1x PBS and then treated with 1 ml of 1x trypsin for T-25 flask and 2-3 ml of trypsin for T-75 flask in order to detach the cells. Cells were incubated with trypsin for 2-4 min and then observed under the microscope. When cells were fully detached, fresh growth medium was added to a final volume of 10 ml and the resuspended cells were taken up and poured into universal tubes and centrifuged for 5 min at 1500 rpm. The cell pellet was resuspended in 10ml of fresh DMEM. Next, depending on the required dilution, 2 or 3 ml of resuspended cells were dispensed into new flasks. 8 or 12 ml of fresh medium was added depending on the flask size.

2.15.2 Stable transfection of CHO cells

Stable transfection was performed over a 4 day period. Day one of the transfection procedure began by plating out the cells in 6-well plates. $6x10^5$ cells were seeded into each well so as to reach ~90 confluency by the next day. Three wells of each plate were used for transfecting the CHO cells with one construct. On day two of transfection, using a 96 well (microtiter) plate a complex of lipofectamine^{TM} 2000 and DNA was made. To form the complex, 100 µl of optiMEM medium was added to two wells of the microtiter plate in two rows. Then the required amount of DNA (up to 4 µg) was added to the wells of the first row of the plate. Next 10 µl of lipofectamine^{TM} 2000 was added to the wells of second row. Then the contents of the wells in the first row were transferred to the well of the second row to form the DNA-lipofectamine complex. To achieve this, the contents of the wells in the first row were pipetted out and transferred to the corresponding wells of second row then pipetted up once more and finally pipetted back gently into the well of the second row. This procedure was repeated only once. Once the DNA-lipofectamine complex was formed, the cells were incubated for 30 min at room temperature. Next, 100 µl of optiMEM was added to give a final volume of volume 0.320 ml.

The medium from the 6-well plates was removed and CHO cells were washed with1 ml of 1x PBS. Then the 320 µl volume of DNA-lipofectamine complex was transferred gently to each well of the 6-well plate and a further 0.5 ml of optiMEM was added to each well. The cells were then incubated with the DNA-lipofectamine complex for 5-6 hours. After this period, 2 ml of DMEM growth medium was added to each well.

On day 3, the transfected cells were washed and trypsinised with 1x PBS and 1x trypsin respectively. Cells were transferred to T-25 flasks. On day 4 the transfected cells were

washed and trypsinised again, re-seeded into a T-75 flask and cells were then grown in a new growth medium containing 400 µg/ml Hygromycin B to maintain selection pressure. Cells were selected against Hygromycin, since the vector contained the gene that confers resistance against hygromycin. The effective concentration of hygromycin had been previously determined by the industrial partner (UniTargetingResearch AS, Bergen). After this step cells were cultivated for ~ 4 weeks in medium containing hygromycin. Finally, the transfected cells were harvested for the measurement of Luciferase activity and RNA extraction.

2.15.3 Transient co-transfection

In transient co-transfection experiments equal volumes (2 µg of each) of a vector containing firefly Luciferase sequence and an appropriate plasmid expressing *Gaussia* Luciferase were used. The procedure was the same as described above in 2.15.2 for stable transfection, except that transfected cells were kept in 6 well plates post-transfection and no selection media was applied.

2.15.4 siRNA transfection

CHO cells (500,000) were seeded in a 6 well plate 24 hours before transfection in order to reach 30-50% confluency. On the day of transfection, the growth medium was replaced with fresh medium containing only 5% Fetal Calf Serum. 80 pmol of siRNA (4 µl) was added to 246 µl of OptiMEM medium for each well and incubated for 10 min at room temperature. Equally, for each well 5 µl of Lipofectamine^{TM} 2000 and 245 µl of OptiMEM were mixed and incubated for 10 min at room temperature. Next, the two solutions were mixed and incubated for further 25 min and then added gently to the cells in the 6 well plate. The medium was

replenished the next day and cells were collected on day 1-4 after siRNA transfection to maximise the chance for observing the siRNA knock down. The same procedure was applied for the transfection of non-specific siRNA as a control.

2.15.5 Sample collection and cell harvesting (stable transfection)

2.15.5.1 Medium samples for Luciferase activity measurement

From each well, 0.9 ml of medium was collected and divided into 2 aliquot of microcentrifuge tubes. Samples were then centrifuged at 10000 g for 10 min at 4ºC (this was carried out to remove dead cells or debris) and 0.4 ml of the supernatant fluid was transferred to 2 new microcentrifuge tubes. Samples were stored at -80˚C until to be assayed.

2.15.5.2 Cell samples for Luciferase activity measurement

To harvest cells for the preparation of cell extracts, after collecting the culture medium cells were washed once with 2.5 ml of cold 1x PBS. The wash buffer was removed completely and 250µl of lysis buffer added to the cells. Cells were incubated for 5 min at room temperature and then 250 µl of suspension transferred to a microcentrifuge tube on ice. To collect the remaining cell material, 250 µl of 1xPBS was added to each well of the 6 well plate and wells were rinsed carefully by pipetting up and down the 1xPBS. The solution was transferred to the original microcentrifuge tube on ice (now with a final volume of 500 µl) and centrifuged at 10000 g for 10 min at 4˚C to pellet the cell debris. Finally, 450 µl of cell extract was transferred into a new microcentrifuge tube. The samples were stored at -80˚C.

2.15.6 Cryogenic storage

To store cell stocks cells were grown to ~90% confluency or higher, trypsinised and pelleted by centrifugation at 1500 g for 5 min. The pellet was resuspended in 3 ml of freezing medium (90% [v/v] FCS +10% [v/v] DMSO) and 1ml of suspension transferred to Cryo-vials. Tubes were kept in the freezer (-20ºC) for two hours, then placed at -80ºC. Next day tubes were transferred to a liquid nitrogen storage tank.

2.15.7 Thawing of cryogenic storage cells

Frozen cells were thawed by placing the Cryo-vials in a water bath (37˚C), and then transferred to a universal tube and 5 ml of complete growth medium was added. The cell resuspension was centrifuged at 1500 rpm for 5 min to remove the DMSO from the cells. The cell pellet was resuspended in 5ml of complete growth medium and transferred to a T-25 flask. Once the cells reached ~90% confluency, the cells were split and re-seeded in T-75 flask. The cells were split every second day thereafter.

2.15.8 Luciferase assay

Samples from culture medium and cell extracts were measured for *Gaussia* Luciferase activity. First an optimum dilution with minimal inhibitory effect from other components of the samples was found by performing a series of dilutions from 1:10 through to 1:21870 in *Renilla* dilution buffer in a 96- well plate. The optimum dilution was important as the plate reader (plate Chameleon^{TM} multilabel counter, Hidex) cannot measure accurately a luminescence signal which is too weak or too intense. Once an appropriate dilution was

found, culture medium and cell extract samples were diluted in *Renilla* dilution buffer accordingly. Next, 25 µl of each sample dilution was dispensed to 2 wells of a white microtiter plate (Nunc, Roskilde, Denmark). The white microtiter plate was used to avoid signal interference between wells. The plate was covered with aluminium foil and kept at room temperature for 10-15 min prior to measurement. The plate reader was programmed to dispense 150 µl of coelentrazine substrate buffer to each well. The results of measurement were given in RLU in which 1 RLU is equal to 1 photon hitting the detector. The RLU was then converted to percentage of *Gaussia* Luciferase activity observed from the standard construct (cell population) GG*G*G.

2.15.9 Cell count

The confluent CHO AA8 Tet-Off cells were washed once with 1x PBS and trypsinised with 2 ml of 1x trypsin. To remove the trypsin, 8 ml of growth medium was added and cells were centrifuged at 1500 rpm for 5 min. The cells were then resuspended in 2 ml of growth medium. To count the number of live cells, Trypan Blue exclusion method was used. Equal volume of Trypan Blue solution (Sigma) and cell suspension were mixed together, and the number of cells excluding Trypan Blue were counted using a haemocytometer. The subsequent cell concentration and total number of cells per ml was determined by using the following calculation:

Cells per ml: the average count per square x the dilution factor x 10^4

2.16 RNA techniques

Throughout this study while working with RNA, great care was taken to keep the samples RNase free. For this reason all the solutions were made with DEPC treated water.

2.16.1 RNA extraction

Cells were grown in 6-well plates to ~90% confluency and total RNA extracted by using Trizol (Invitrogen). After removal of culture medium, cells were washed with 1x PBS and 1ml of Trizol added and incubated for 5 min at room temperature. Next, 0.2 ml of chloroform was added, the tubes were mixed vigorously for 15 s and then incubated for 2-3 min at room temperature. The samples were then centrifuged at 12000 g for 15 min at 4ºC. This leads to a formation of a red phenol-chloroform phase at the bottom of the tube and a colourless aqueous phase on top. The top layer containing the RNA was carefully removed to a new microcentrifuge tube and 0.5 ml of isopropanol added to it. After mixing by inversion samples were incubated for 10 min and centrifuged at 12000 g for 15 min at 4ºC to pellet the RNA at the bottom of the tube. The supernatant fluid was then carefully removed and the pellet washed with 1ml of 75% ethanol. Tubes were vortexed and centrifuged at 7500 g for 5 min at 4ºC. The supernatant fluid was removed and the pellet was then air-dried. Finally, the air-dried pellet was resuspended in 50-100 µl of DEPC-treated water and incubated at 55-60ºC until the pellet was completely dissolved.

2.16.2 Measuring RNA concentration by spectrophotometry

RNA samples were diluted in DEPC water (1:50) and absorbance was measured spectrophotometrically using a Biophotometer™ (Eppendorf®). The calculation of concentration was based on knowledge that an OD=1.0 at 260nm corresponds to 40 µg/ml RNA. Concentrations of the measured RNA were therefore expressed as µg/ml and good quality of sample was attributed by a 260/280 ratio of 1.7-1.9.

2.16.3 RT-PCR

Throughout this study, reverse transcription was carried out using a Transcriptor Reverse Transcriptase kit (Roche). To set up a 20 µl reaction, the following components were pipetted into a small thin-walled Rnase and Dnase free reaction tube on ice according to the manufacturer's instruction:

Reaction mix for a typical RT-PCR

Template RNA	1µg of total RNA
Oligo $(dT)_{15}$ primer	1µl
Water, PCR grade	up to 13µl
Transcriptor RT buffer	4µl
RNase inhibitor	0.5µl
dNTP mix (1mM each)	2µl
Transcriptor Reverse Transcriptase	0.5µl

The reaction was mixed by vortexing. Tubes were briefly centrifuged and the samples incubated for 30 min at 55˚C. The reaction tubes were stored at -15-25˚C.

2.16.4 Real-Time PCR

2.16.4.1 Background

Real-Time PCR is a technique that uses a fluorescent dye (e.g. Light Cycler 480 SYBR Green 1 Master, Roche) to measure in real time the amount of double stranded DNA products that are produced over 45 cycles. Real-time PCR uses cDNA as starting material with a typical amplicon being between 100–200 base pairs. For this study primers were designed to produce 100 base-pair amplicons for both the target gene (*Gaussia* Luciferase) and reference gene (Glyceraldehyde 3 phosphate dehydrogenase also known as GAPDH). The Real-Time PCR measurements in this study were carried out by using a LightCycler® 480 (Roche).

Since the fluorescent dye has the capability of binding to any double helix structure, it will fluoresce at 483-533nm upon binding to DNA at the major groove. This light can be calculated while the actual experiment is running. Furthermore, the SYBR Green 1 Master has the capability to produce small amount of fluorescence in its unbound state and this crucially allows the measurement of the point at which the log-linear part of amplification curve crosses the background noise or crossover points [183]. The crossover point or Cp value is proportional to the initial cDNA concentration of samples [184].

2.16.4.2 Standard curves

PCR was used to amplify a double stranded 100bp region from all the *Gaussia* Luciferase and GAPDH samples. The amplified products were purified by using QIAquick PCR purification kit (Qiagen) and then diluted in a series of 1:10 serial dilutions (15 times) to create a standard curve for *Gaussia* Luciferase and GAPDH samples. In addition, 1:10 dilution of those samples to be quantified were made and analysed alongside the standard curve in order to be sure that the crossover points of the samples fell within the curve. Furthermore, a calibrator sample could be found from this data if a standard sample showed a crossover point (Cp value) similar to those obtained for the selection of unknown samples. The calibrator was used by the LightCycler® 480 to correct for PCR efficiency.

2.16.4.3 Sample measurement

Once it was established that 1:10 dilution for all the unknown samples was appropriate, the samples were diluted accordingly. The primers that were utilised for the Real-Time PCRs were produced by Eurofins (MWG, Germany) by using MWG-HYPER gel purification system to obtain primers with the highest purity. To carry out a Real-Time experiment, a master mix was made which consisted of 10 µl of SYBR Green 1Master mix (Roche), 1 µl of forward and reverse primers and 3 µl of PCR grade water for each reaction. The mix was loaded into a LightCycler® 480 96-well plate (Roche). Next, 5 µl of diluted samples to be quantified was added to each well. As negative control 5 µl of water was added to one well along with 15 µl of reaction mix. The plates were sealed by using a LightCycler® 480 sealing foil and then centrifuged for 2 min at 1500 g before carrying out the measurements.

2.17 Protein binding techniques

2.17.1 S-100 cell extraction of CHO cells

CHO cells were grown in T-175 flasks until cells reached ~90% confluency. Then 5ml of 1x trypsin was added and cells incubated for 5-10 min at 37° C. 5 ml of growth medium was added and cells pelleted by centrifugation at 210 g for 5 min and resuspended in 20 ml of PBS. Cells were centrifuged at 210 g for 5 min and the resulting pellet was washed and re-suspended with 2ml of 130 mM NaCl lysis buffer 1 (see 2.4). 1ml of cell re-suspension was transferred to a 1.5ml microcentrifuge tube, centrifuged (pulse spin at 4500 g) and the supernatant fluid removed. The remaining half of re-suspended cells was added and another round of pulse spin centrifugation (4500 g) performed. The pellet was resuspended in 400 µl of 130 mM lysis buffer 2 (see 2.4). Cells were lysed by passing them through a 21-gauge needle several times, collected and centrifuged at 5000 g at 4°C for 5 min. The supernatant fluid was diluted with 3 volumes of 40 mM lysis buffer (see 2.4). The diluted fluid was divided equally into two ultra centrifuge tubes and ultracentrifugation was carried out at 53000 g at 4°C for 60 min. The supernatant fluid was collected and stored in -80°C.

2.17.2 *In-vitro* transcription to synthesise non-labelled and radiolabelled RNA probe

To perform *in-vitro* transcription and obtain template of the Albumin 3'UTR RNA transcript, initially a typical PCR was carried out where the minimal T7-polymerase promoter sequence was incorporated at the 5' position of the forward primer (see Table 2.5)**.** Purified PCR products (QIAquick PCR purification kit, Qiagen) were then used as template to synthesise non-labelled and radiolabelled RNA transcripts of interest. Following purification of PCR

products, 0.8 µg of DNA was used in *in-vitro* transcription using MEGAshortscript™ Kit (Ambion). Reaction mixtures were prepared in an RNase free microcentrifuge tube at room temperature in the order shown below according to the manufacturer's instruction:

Reaction mix for a typical *in-vitro* transcription for production of non-labelled RNA

Water (Nuclease free)	to 25 µl of final volume
10x T7reaction buffer	2.5 µl
CTP solution (75mM)	2 µ
GTP solution (75mM)	2 µl
UTP solution (75mM)	2 µl
ATP solution (75mM)	2.5 µl
Template DNA	0.8 µg
T7 enzyme mix	2 µl

Reaction mix for a typical *in-vitro* transcription for production of radiolabelled RNA

Water (Nuclease free)	to 25 µl of final volume
10x T7reaction buffer	2.5 µl
CTP solution (75mM)	2 µl
GTP solution (75mM)	2 µl
UTP solution (75mM)	2 µl
[α-^{32}P] ATP (PerkinElmer) 800 ci/mmol	2.5 µl
Template DNA	0.8 µg
T7 enzyme mix	2 µl

The reaction was mixed by gently flicking the tube and then it was centrifuged briefly to collect the reaction mixture at the bottom of the tube. The reaction was incubated for 3 hr. To remove the template, 1 µl of turbo DNase was added and mixed well. The reaction was incubated at 37°C for 15 min at room temperature. Finally 70 µl of DEPC-water was added to the reaction mixture.

2.17.3 Phenol-Chloroform extraction

Phenol-Chloroform extraction was performed following the in-vitro transcription procedure to remove possible free nucleotides and enzymes present in the reaction mix. Initially 70 µl of Phenol: Chloroform: Isoamyl Alcohol 25:24:1 (Sigma) was added to the reaction and mixed by inverting the tube. The reaction was centrifuged at 13000 rpm for 10 min at 4°C. The aqueous phase was removed and mixed with 70 µl of chloroform. The mixture was centrifuged at 13000 rpm for 10 min at 4°C. The aqueous phase was added to 500 µl of chilled ethanol containing 15 µl ammonium acetate (5M) to precipitate the transcript. The mixture was stored overnight at -80°C to precipitate the RNA and the following day it was centrifuged at 15000 g for 60 min at 4°C. The supernatant fluid was removed carefully and RNA precipitated at the bottom of the tube as pellet. The resulted pellet was air dried and 20 µl of nuclease–free water added to dissolve the precipitated RNA transcript. Finally the concentration of the transcript was measured by spectrophotometry.

2.17.4 Measuring concentration and activity of radiolabelled RNA transcript

The incorporation of ^{32}P-ATP into the RNA transcript was determined by scintillation counting. A 1/10 dilution of RNA transcript to be measured was spotted (0.1 µl) in duplicate onto Whatman filter paper and then was placed into scintillation vials with approximately 5-10 ml of Optiphase Hisafe™ scintillation cocktail (Perkin-Elmer®). In addition a blank sample without any RNA was also set up. The ^{32}P radioactivity of RNA transcripts was measured in counts per min (cpm), in a TRI-CARB™ 2007 TR (Perkin Elmer® liquid scintillation counter. The output by the scintillation counter machine was averaged and used to calculate the total activity and concentration in fmols/µl of RNA transcript.

Activity of sample in MBq

A = (readout 1+ readout 2)/2 = cpm average of two duplicates (see 2.17.4)

B = A x 200(dilution factor) = total cpm of radiolabelled RNA transcript

Test by JEH laboratory showed that 77360300cpm = 0.74 MBq

=> (B / 77360300) x 0.74 = activity of whole sample (MBq)

Theoretical maximum yield = 2.22 x 10^6

=> Isotope stock used = 10 mCi/ml or 10 µCi/µl

Assuming 100% integration of isotope in RNA transcript→ 2.5 µl (amount of radio active ATP used per reaction) x 10 = 25 µCi

=> 2.22 x 10^6 x 25 = 55500000 dpm (disintegrations per min)

Since ^{32}P has a half life of 14 days, a decay factor (Y) must be included in the calculation based on time from reference date.

Decay factor (Y) = activity at present time/ activity at reference date

Calculating percentage of isotope incorporation in RNA transcript => C

C = [total sample activity/ (total dpm x decay factor)] x 100

C = [B / (55500000 x Y)] x 100

Calculating maximum possible yield of RNA Transcript (µg) => D

Amount of ATP used per *in-vitro* transcription reaction = nmoles

Proportion of A bases = number of A bases in RNA transcript / total number of all four bases in RNA transcript

Molecular weight of each nucleotide = 330 (Dalton)

=> D = C/100 x ((total amount of ATP (nmoles)/ proportion of A bases in RNA transcript) x 330

Calculating total yield (ng) => E

E = D/20 x 1000 where 20 = the actual volume of transcript

Calculating concentration of fmol/µl of transcript => F

First calculating the concentration in pmol = E/(186/330) x 1000000 where 186 = total number of nucleotides in the transcript

Then concentration in pmole/100 x 1000 => concentration of fmol/µl of transcript

2.17.5 Assessing the integrity of radiolabelled RNA transcript

Prior to protein binding assays, integrity of RNA transcript was checked by subjecting a sample of radiolabelled RNA transcript to electrophoresis in a denaturing urea polyacrylamide gel and then visualised by autoradiography.

General reaction mix for a denaturing urea polyacrylamide gel

7.5 ml of urea stock (18.9 g of Urea +9 ml of 5x TBE to final volume of 45 ml with DEPC-treated water)

2.5 ml urea acrylamide stock (8.4 g of urea + 10 ml of 40% acrylamide/bisacrylamide (29:1) to the final volume of 20 ml with DEPC-treated water))

10 µl of TEMED

80 µl APS (10%) solution [w/v]

After solidification of the gel (15-20 min), samples of RNA transcript were loaded in 2x gel loading buffer and the gel was run for 90-120 min at 120 V. Next, the gel was wrapped and exposed to a Kodak Biomax XAR X-ray film for up to an hour. The film was developed in a Konia®SRX-101™ film processor.

2.17.6 Electrophoretic mobility shift assay (EMSA) and competition assay

EMSA assays were carried out in which binding of radiolabelled RNA to cell protein extracts was assessed by electrophoresis through a non-denaturing polyacrylamide gel. The required amount of radioactively labelled transcript (12 fmol per reaction) and non-labelled transcript if applicable (as competitor) were heated at 70°C then allowed to cool down to ~40°C. This allows the RNA transcripts to be denatured and theoretically refold into their native form. Prior to assembling the binding reaction, the native polyacrylamide gel was pre-run in 0.5x TBE buffer at 120 V.

The reaction mix was incubated for 15 min at room temperature. 40 U of RNase T1 (1/25 dilution of RNase T1 in binding buffer) was added to each reaction and samples incubated for

further 5 min at room temperature. 2 µl of 20% Ficoll was added before loading samples on to the gel. As a marker, bromophenol blue was loaded in one lane. Electrophoresis was carried out in 0.5x TBE buffer for 2 hr until the bromophenol blue dye had run completely off the bottom of the gel. Finally, the gel was loaded between 2 sheets of GelAir Cellophane Support and placed in a gel air dryer (115 V, 60 Hz,Bio-Rad) until it was completely dried out. The gel was exposed to Kodak Biomax XAR film and developed in Konia®SRX-101™ film processor.

General reaction mix for a native polyacrylamide gel

1.25 ml of 40% acrylamide/bisacrylamide stock (79:1)

[3.75 ml of 40% (29:1) acrylamide/bisacrylamide + 6.25 ml of 40% acrylamide]

1 ml of 5x TBE buffer

7.65 ml of DEPC- treated water

10 µl of TEMED

90 µl of ammonium persulphate

General reaction mix for a typical EMSA

Protein extracts	2-2.5 µg
Labelled RNA transcript	12 fmol
Competitor RNA Transcript	?µl excess fold of competitor (only applicable for competition assay)
40 mM lysis buffer (see 2.4)	to a final volume of 8 µl

2.17.7 UV Cross-linking assay

The required amount of labelled RNA probe (50 fmol per reaction) and non-labelled RNA transcript (required excess fold based on experiment) were heated at 70°C and allowed to cool down to ~40°C. This allows RNA transcripts to be denatured and theoretically refold into their native form. The denatured and refolded RNA transcripts (non-labelled and labelled) were added to a mixture of protein extracts and 40 mM lysis buffer to make the reaction mix. The reaction mix was incubated at room temperature for 15 min. 40 U of RNase T1 (1/25 dilution of RNase T1 in binding buffer) was added and the reaction mix incubated for a further 5 min at room temperature. The reaction mix was then cross-linked for 12 min in SpectroLinker XL 1000 V Cross-Linker at 4,500 µW/cm^2 on ice. Samples were incubated with 10 µg of RNase A (1 µl) at 37°C for 1 hr. 3 µlof 5X Laemmli protein sample buffer was added to each reaction and boiled for 5 min. Samples were loaded onto a 10% SDS-PAGE (29:1) gel and run initially at 120 V for 30 min and then 150 V for 2.5 hr in 1x protein electrophoresis running buffer (see 2.4). 5 µl of colorBurst^{TM} Electrophoresis Marker (8000-22000 kDa) was loaded into one lane as marker. The gel was removed and placed in between 2 sheets of GelAir Cellophane Support and inserted in a gel air dryer (115 V, 60 Hz,Bio-Rad) until it was completely dried out. The gel was exposed to Kodak Biomax XAR film and developed in the Konica SRX-101A developer.

<u>General reaction mix for a typical UV Cross-linking</u>

Protein extracts	2.5-3 µg
Competitor (non-labelled transcript)	1-2µl (depending on the excess fold of competitors)
50 fmol of radiolabelled RNA transcript	2-4 µl

40 mM lysis buffer (see 2.4) up to final volume of 13 µl

General reaction mix for 10% SDS-PAGE gel

Separating gel (75-80% of whole gel)

40% Acrylamide/*Bis* Acrylamide Stock (29:1)	1.75 ml
Separating buffer (1.875M Tris Base pH 8.9, 0.25%SDS)	2.8 ml
Distilled water	2.4 ml
TEMED	6 µl
APS (10%) solution [w/v]	65 µl

Stacking gel(20-25% of whole gel)

40% Acrylamide/*Bis* Acrylamide Stock (29:1)	0.25 ml
5x stacking buffer stock (0.3M Tris Base pH 6.7, 0.5%SDS)	0.4 ml
Distilled water	1.32 ml
TEMED	2.5 µl
APS (10%) solution [w/v]	18 µl

The separating gel mixture was immediately poured into the gel cassettes using a 5 ml pipette and filled up to ~80% of gel cassettes. 2 ml of distilled water was added and the gel was allowed to polymerise for 20-30 min at room temperature.

Once the separating gel was set, the top layer of distilled water was poured out and the gel was rinsed with 2 ml of fresh distilled water. The inner side of the cassettes was dried with clean paper towel and the stacking mixture was immediately poured into the cassettes onto the top of the separating gel.

2.17.8 Isolation of RNA-binding proteins

To synthesise the biotinylated probe, the same procedure in 2.17.2 was followed with the exception of adding 2 µl of Biotin-UTP-16. The synthesised biotinylated transcripts were subjected to phenol/chloroform extraction (see 2.17.3) and quantified by spectrophotometry. 20 µg of biotinylated transcript were heated at 70°C for 5 min, then at 40°C for 20 min and then allowed to cool down at room temperature. The MagneSphere Streptavidin-Coated Paramagnetic Particles (SA-PMP) (0.6 ml with a final concentration of 1 mg/ml) were incubated with 100 µl of 0.5x SSC buffer, 10 µg of BSA and 10 µg of yeast tRNA for 60 min at room temperature with shaking.

The SA-PMPs (beads) were washed twice with 300 µl of 0.5x SSC buffer. Next, the SA-PMPs were incubated with 20 µg of biotinylated transcript in 300 µl of 0.5x SSC buffer for 10 min at room temperature. For a negative control no biotinylated transcript was added. To remove the unbound transcript, beads were washed with 0.3 ml of 0.5x SSC buffer. Then the beads were incubated with 1 mg of CHO cell extract in 500 µl of 40 mM lysis buffer (see 2.4) with an additional 25 µg of yeast tRNA, 10 µg of BSA and 800 U/ml of RNasin at 4°C for 60 min with shaking. In order to separate beads from non-bound material, beads were pelleted magnetically by placing them onto MagnaRack^{TM} (Invitrogen). The supernatant fluid was removed and SA-PMPs washed 5 times with 1ml of 40 mM lysis buffer. Beads were then re-suspended in 25 µl of 40 mM lysis buffer. 10 µl of re-suspension was mixed with 2x dissociation buffer (see 2.4) in order to release bound protein/s. Protein/s were denatured for 5 min at 95°C and loaded into a 10% SDS-PAGE gel. The gel was visualised by Novex Colloidal Blue (Invitrogen).

General reaction mix for a typical *in-vitro* transcription of biotinylated RNA probe

Water (Nuclease free)	to 25 µl final volume
10x T7 reaction buffer	2.5 µl
ATP solution (75mM)	2 µl
CTP solution (75mM)	2 µl
GTP solution (75mM)	2 µl
UTP solution (75mM)	2 µl
Biotin-UTP-16	2 µl
Template DNA	0.8 µg
T7 enzyme mix	2 µl

2.17.9 Staining the 10% SDS-PAGE gel

Colloidal blue staining solution was prepared according to the manufacturer's instructions (Invitrogen), as follows: 55 ml of de-ionised water was added to 20 ml of ethanol, 5 ml of stainer B and 20 ml of stainer A. The gel was placed in a clean container with staining solution and shaken for 3-12 hr. The staining solution was decanted and replaced with 200ml of de-ionised water and shaken for a minimum of 7 hr to de-stain the gel.

2.18 Protein techniques

2.18.1Bradford assay

The protein content of cell extract samples was measured by Bradford assay in which Bovine Serum Albumin (BSA, New England BioLab) was used as a standard over a range from 0.02,

0.04, 0.06, 0.08 and 0.1 mg/ml. Samples were diluted in a range of 1:20 to 1:120 and an appropriate dilution selected by colour comparison to the chosen dilution of the standard curve. 50 µl of samples and standards were transferred into a 96-well microtiter plate. 200 µl of Bradford reagent (Sigma) was added to each well and the plate incubated at room temperature for 10 min before the absorbance was measured by Multiskan Ascent microplate reader at 595nm.

2.18.2 Western blotting and analysis

Western Blotting was carried out to identify proteins following SDS-PAGE. After electrophoresis, the proteins were transferred to a PVDF membrane (Roche) by a semi dry transfer method. Initially, the membrane was placed in a clean container containing 100% methanol for 30 seconds. The membrane was then placed in a clean container containing distilled water at room temperature for 2 min while being shaken. Next the membrane was transferred to a clean container containing transfer buffer at room temperature for 10-15 min. Three layers of Whatman chromatography paper (grade 3MM) were soaked in transfer buffer and placed on the transfer apparatus (atta AE6675 semi-dry transfer blotter). The air bubbles were removed by using a clean glass tube before the membrane was laid on top of them. Next, the separating part of the gel was placed on top of the membrane before placing three more layers of soaked Whatman paper in transfer buffer. Transfer was carried out at 15 V for 50 min. The PVDF was then removed and placed overnight in blocking solution at 4°C with shaking to prevent non-specific binding of the antibodies to the membrane.

The membrane was blocked with 5% (w/v) of dried milk powder (Marvel) in PBS containing 0.05% (v/v) Tween 20 (Sigma). The CUG-BP1 primary antibody was then diluted (1/1000)

in blocking solution. Then the membrane was incubated with the primary antibody for 1 hr at room temperature with shaking. The membrane was washed 4 times for 10 min and then incubated with the secondary antibody (anti Rabit IgG (whole molecule) peroxide antibody) diluted 1:5000 in blocking solution for 1 hr with shaking. The membrane was then washed 5 times for 10 min each to remove any non-specific binding. An Amersham™ ECL+ western blotting detection system (GE Healthcare) was applied to visualise the secondary antibody. The ECL solutions were let to warm up for 5 min at room temperature prior to mixing. 2ml of solution A and 50 µl of solution B were mixed gently in a universal tube wrapped with foil. The membrane was placed onto a plastic film gently blot dried with Whatman paper. The mixed solution was poured onto the membrane and was spread evenly by placing a second plastic film on top of the membrane. The membrane incubated at room temperature for 5 min. The plastic films were removed and the membrane was gently blot dried with Whatman paper. A fresh dry plastic film was placed on top of the membrane. The membrane was exposed to Kodak Biomax XAR film (Sigma) in a Hypercassette (GE Healthcare) in the dark for 5 to 15 min (depending on the intensity of the signal). The films were developed by Konica SRX-101A developer.

2.18.3 Protein identification by mass spectrometry analysis (courtesy of NEPAF)

Bands were excised from the gel using a razorblade in a clean air cabinet. Proteins in the bands were digested in the gel using a GenomicSolutions ProGest robot and the following standardised procedure was followed. The gel pieces were cut into smaller pieces using a pointed forceps and transferred into a pierced PCR plate. Gel pieces were incubated for 10

min at room temperature with 25 mM NH_4HCO_3 (ammonium bicarbonate) and for 10 min with 100% MeCN (acetonitrile). This washing cycle was repeated once. Then, gel pieces were treated with DTT (10 mM, 1 hr, 40 °C), the reaction was allowed to cool down for 20 min and cysteines were alkylated using ICH_2CONH_2 (10 mM, 30 min, 20 °C). Gel pieces were washed and dehydrated using MeCN and incubated with 1.5µg of trypsin (Promega, 3 hr, 37°C). Proteolytic peptides were extracted using water and MeCN and the extracts were concentrated under vacuum to 10µl and finally 3 µl of this mixture was used for LCMS analysis (Liquid Chromatography Mass Spectrometry). To perform the LCMS analysis an aliquot of the digests was transferred into an autosampler vial and analysed using an Ultimate 3000 Nano-HPLC system (Dionex, UK) coupled online to an LTQ XL orbitrap mass spectrometer. The following parameters were used for the analysis:

Column:	Dionex Pepmap C18 column, 3um ID 25cm
Mobile Phase A	0.05% Formic Acid in LCMS H_2O
Mobile PhaseB	0.05% Formic Acid in 80% Acetonitrile
Flow rate	300ul/min
Gradient:	4% to 50% B in 102 min
Segment runtime:	120 min
MS Resolution:	30,000
FTMS ACG Targets:	$5x\ 10^5$

Top 3 FT MS/MS at 35% normalized collision energy

MS/MS Resolution:	30,000, 1 microscan
FT MS/MS AGC Target:	$3x\ 10^5$

Dynamic Exclusion:	Repeat Count=2,
Exclusion duration :	180 seconds
MS/MS Threshold:	10,000
Max IT FTMS:	50.000 ms
Full MS mass range:	400- 1600 m/z
MS/MS mass range:	Full

To perform data analysis thermo raw binary data files were converted into MSMS peak lists (Mascot Generic Format, *.mgf) using Thermo ProteomeExplorer 1.0. Next Data were searched using an in house installed version of X!Tandem (x! tandem TORNADO (2009.04.01.1). All processing of gel slices and protein identification was carried out by NEPAF, Newcastle University.

3 Creation of gene constructs to investigate effects of signal peptides/3'UTR on reporter activity

3.1 Introduction

To examine elements that influence post-translational regulation such as signal peptide and 3'UTR, the *Gaussia princeps* Luciferase was used as reporter system in a series of novel gene constructs that were created by utilising a seamless cloning method. The names and codes that have been used to identify constructs in this study are based on the different elements that make up any given construct. The expression vector $_{p}$TRE2hyg (Clontech) was used throughout to make all the gene constructs (Figure 3.1) and all the constructs names begin with the name of the expression vector $_{p}$TRE2hyg. This is followed by the 5' untranslated region (5'UTR), the signal peptide coding region, the *Gaussia* Luciferase coding region and the 3' untranslated region (3'UTR).

In all constructs the 5'UTR region is from the *Gaussia princeps* Luciferase gene. Immediately after the 5'UTR region is the signal peptide coding region. In this thesis the Human Chymotrypsinogen signal peptide, the Human Albumin signal peptide and finally the native *Gaussia* Luciferase signal peptides were utilised as the signal peptide of a given gene construct. The coding region of the signal peptide is then followed by the *Gaussia princeps* Luciferase coding region/sequence in all of the constructs. The asterisk "*" on the letter G indicates that this region has been codon optimised with respect to mammalian codon usage without changing the amino acid composition. This is termed the humanisation of the coding region. The region following the *Gaussia* Luciferase coding region is the 3'UTR. In this thesis the Albumin and *Gaussia* Luciferase 3'UTRs were utilised to make the constructs.

Finally, sequences at the end of the 5'and 3'-untranslated regions were upstream and downstream sequences respectively derived from the pTRE2hyg vector. The codes used throughout were:

G: *Gaussia*

X: Chymotrypsinogen signal peptide

X1: Chymotrypsinogen signal peptide with 1 mutation

X2: Chymotrypsinogen signal peptide with 2 mutations

A: Albumin signal peptide

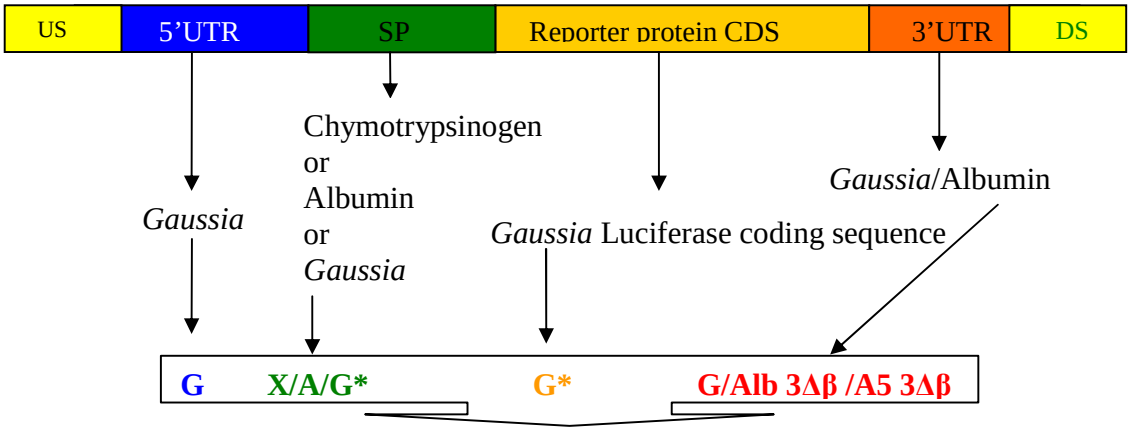
A5: Albumin 3'UTR (with 5 point mutations)

Alb: Albumin 3'UTR (wild type):

3Δβ: without β-globin vector-derived sequences

US: Upstream sequences from vector

DS: Downstream sequences from vector



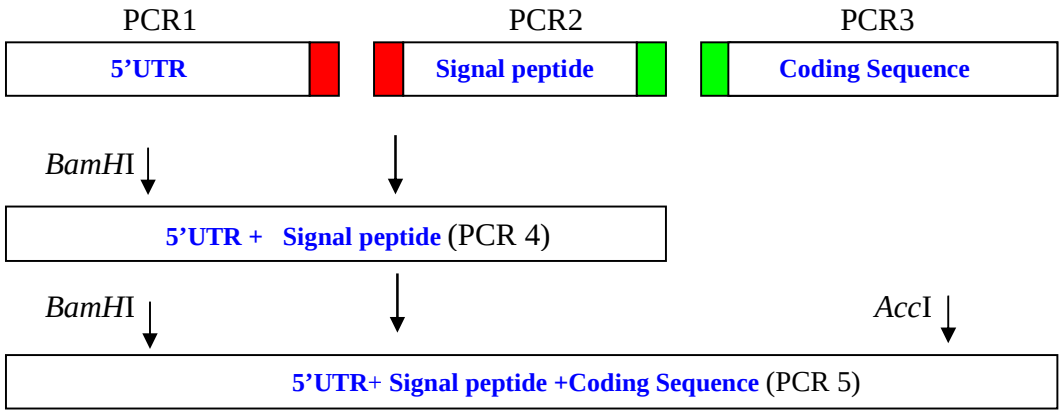
$_P$TRE2hyg G— X/A/G*— G *— G/Alb 3Δβ /A5 3Δβ

Figure 3.1 Schematic representation of the terminology used in this study to identify constructs. In all the constructs, the 5'UTR and the coding region were obtained from the *Gaussia* princeps Luciferase. The signal peptides of Chymotrypsinogen, Human Albumin and *Gaussia* Luciferase were used in this study. Finally, either the endogenous *Gaussia* Luciferase or Albumin 3'UTR were included in the constructs.

3.2 Seamless cloning

Seamless cloning is a PCR based cloning method in which a series of PCRs (5 separate PCRs in the present study) are performed to amplify a region in order to insert it into a vector without using linker sequences or incorporating new restriction sites [185]. The strategy by which this method was applied in the present work is shown schematically in Figure 3.2. To begin this procedure 3 separate PCRs were performed to amplify the *Gaussia* 5'UTR, a signal peptide coding region of interest and finally the *Gaussia* Luciferase coding region (sequence) separately. To amplify the actual signal peptide without using a template, the forward and reverse primers are designed in such a way that the primers share a region of homology between them (see Table 2.4, OliAlbSP.for 5', OliAlbSP.rev 5'). This region of homology is used by the polymerase to amplify the signal peptide without using template.

Once these 3 regions were amplified, then the signal peptide coding region was linked to the *Gaussia* coding sequence and the *Gaussia* 5'UTR region by carrying out further PCRs with specific forward and reverse primers for the signal peptide (see Table 2.4, AlbSP.for 5',AlbSP.rev5') which contain regions of homology with the 3'-end of the 5'UTR (Figure 3.2, highlighted in red) and with the 5'-end of *Gaussia* coding region (Figure 3.2, highlighted in green). These overlapping regions theoretically allow the polymerase to use these regions as a template so that PCR leads to a product in which the coding region of the signal peptide is attached to the *Gaussia* 5'UTR (PCR number 4 in Figure 3.2). The forward primer for the *Gaussia* 5'UTR was designed specifically in such a way that the amplicon contains *BamHI* unique restriction site up stream of the 5'UTR. A further PCR reaction (PCR 5) was carried out in which subsequently the *Gaussia* Luciferase coding region was linked to the *Gaussia* 5'UTR and signal peptide (the product of PCR 4). The reverse primer for the *Gaussia* coding region was designed in such a way that it contained the *Acc*I unique restriction site. Since any given gene construct in this thesis and the product of PCR 5 contained the *BamH*I and *Acc*I unique restriction site at the same positions, their double digestion and subsequent ligation led to complete replacement of the gene construct signal peptide (Figure 3.2).



Double digestions of both the final product of the PCR5 and any given gene construct with the *BamH*I and *Acc*I

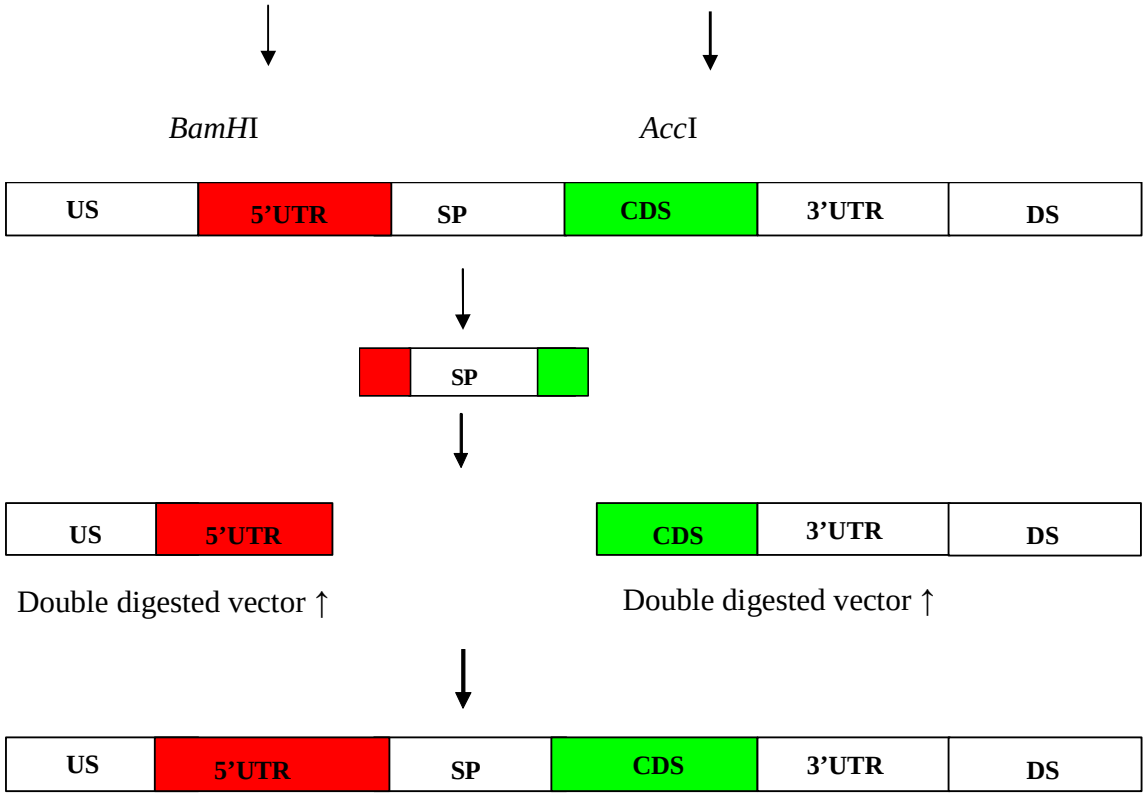


Figure 3.2 Schematic representation of the seamless cloning procedure. The vector and product of PCR5 are digested with the restriction enzymes *BamH*I and *Acc*I that can then be ligated together. As a consequence the signal peptide of the construct will be replaced.

3.3 Results

3.3.1 Construction of pTRE2hyg GXG*Alb 3Δβ

To correct the Albumin 3'UTR for the 5 point mutations in construct GXG*A5 3Δβ, it was logical to replace the A5 3'UTR of the previously made construct GXG*A5 3Δβ with the Albumin 3'UTR but without the 5 point mutations in construct GG*G*Alb 3Δβ. The strategy used to make the constructs was to find two unique restriction sites at either end of the 3'UTR region and replace the A5 3'UTR (containing the 5 point mutations) with the Alb 3'UTR (corrected Albumin 3'UTR for the 5 point mutations). Constructs GXG*A5 3Δβ and GG*G*Alb 3Δβ were treated with restriction enzymes *Acc*I and *PflM*I (Figure 3.3). The *Acc*I restriction site is in the middle of the *Gaussia* Luciferase coding region and the *PflM*I restriction site is in the downstream sequences. Therefore, by digesting the plasmid, the 3'UTR from both constructs was removed. Next, ligation was performed between the vector (lane 2) and insert (lane7). The product of the ligation was used directly for transformation.

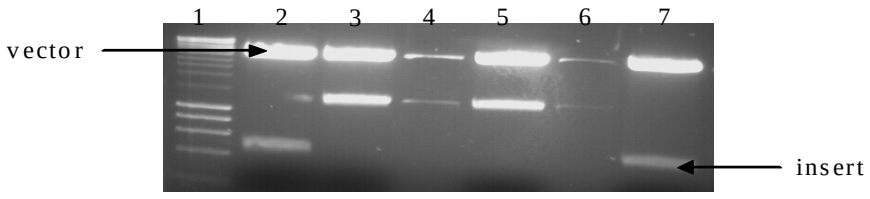


Figure 3.3 Double digestion of construct GXG*A5 3Δβ (lane 2) and construct GG*G*Alb 3Δβ (lane 7) using restriction enzymes *Acc*I and *Pflm*I. Lanes 3 and 5 represent the double digestion of construct GX1G*G and construct GX2G*G (containing the additional vector-derived sequences). Lane 4 and 6 represent empty wells. Lane 1 represents the DNA Hyper Ladder I.

From each transformation plate 5 colonies were selected for transformation of TOP10 competent cells and colony PCR was performed using several single colonies as template (Figure 3.4). Samples together with forward and reverse primers (pTRE2hyg.for and BCD.rev respectively, 10 µl each with a concentration of 10 pmol) were sent to MWG for sequencing. The results of pair-wise alignment confirmed the successful construction of GXG*Alb 3Δβ.

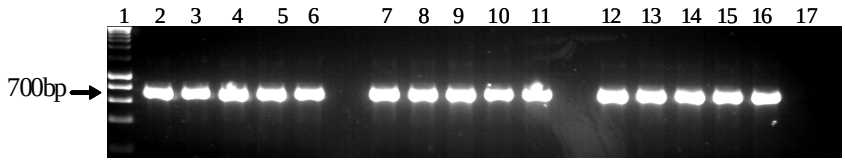


Figure 3.4 Agarose gel electrophoresis showing the screening single colonies for the presence of construct GXG*Alb 3Δβ by colony PCR. Primers CDC.for and Alb.3 rev were used. The expected product of these PCRs was ~700 bp long. Lane 1 is DNA HyperLadder I and Lanes 2-6, 7-11 and 12-16 show the expected band. Lane 17 represents a negative control.

3.3.2 Construction of pTRE2hyg GAG*Alb 3Δβ

The strategy to make this construct was to use the seamless cloning procedure as discussed previously in this chapter. This required removing the Chymotrypsinogen signal peptide coding region from the construct GXG*Alb3Δβ and replacing it with the Albumin signal peptide. The sequence of the Human Albumin Signal peptide was obtained from the gene bank database (Accession Number NM_000477). Initially, oligonucleotides were designed as primers for PCR so that the forward and reverse primers (OliAlbSP.for and OliAlbSP.rev) share 14 base pairs in homology. The aim of this initial PCR was to obtain the complete

sequence of the Albumin signal peptide as the template for carrying out the seamless cloning. To initiate making of construct GAG*Alb3Δβ and performing the seamless cloning procedure (see 3.2), further primers were designed for the Albumin signal peptide so that the forward primers (AlbSP.for) shared sequence homology with the *Gaussia* 5'UTR and the reverse primers (AlbSP.rev) shared sequence homology with the *Gaussia* Luciferase coding region. To achieve this, 5 separate PCRs were performed. First, PCR1 was performed to amplify a region upstream of *Gaussia* 5'UTR to the end of this region using appropriate primers (5'UTR.for and 5'UTR.rev) in order to accommodate the restriction site *BamH*I (Figure 3.5 lane 2).Figure Next, PCR2 was performed using the Albumin signal peptide forward and reverse primers (OliAlbSP.for and OliAlbSP.rev) (Figure 3.5 lane 3). Then PCR3 was carried out using *Gaussia* Luciferase coding region's forward and reverse primers (CDS.for and CDS.rev) to amplify the *Gaussia* Luciferase coding region (Figure 3.5 lane4).

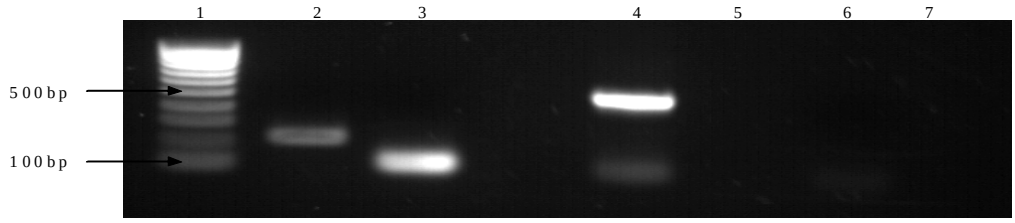


Figure 3.5 Agarose gel electrophoresis showing products of 3 PCRs carried out to amplify the *Gaussia* Luciferase 5'UTR (PCR1), the Albumin signal peptide (PCR2) and the *Gaussia* Luciferase coding region (PCR3, see Figure 3.2 for details). Lane 1 represents DNA HyperLadder IV. Lane 2 shows the product of PCR1. The expected product for this PCR was 210bp long. Lane 3 shows the product of PCR 2 (the signal peptide). The expected product for this PCR was 54 bp long. The *Gaussia* Luciferase coding region was obtained by performing PCR3 (lane5). The expected product for this PCR was 507bp long. Lanes 5, 6 and 7 represent the negative controls for PCRs1-3 respectively.

The next step was to amplify in tandem the *Gaussia* 5'UTR and Albumin signal peptide regions by performing a separate PCR (PCR4) using the 5'UTR.for and AlbSP.rev primers.

As shown in Figure 3.6 (left panel), this PCR produced a product of approximately 300 base pairs. To perform the PCR 4, product of PCR1 and 2 were used as template after performing the PCR purification with Qiagen PCR purification Kit. Once this region was amplified it was linked with the *Gaussia* Luciferase coding sequence**.** For this purpose, 5'UTR.for and CDS.rev primers were used (Figure 3.6, right panel).

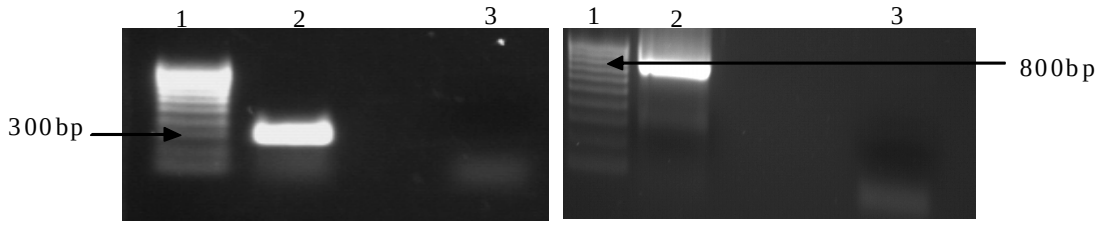


Figure 3.6 Agarose gel electrophoresis showing the product of PCR 4 and 5. PCR4 (left panel) was performed to amplify a region upstream of the *Gaussia* 5'UTR to the end of the Albumin signal peptide (see Figure 3.2 for details). The expected product of this PCR was 296bp long (left panel, lane 2). PCR5 (right panel) was performed to amplify in tandem the product of PCR 3 (see Figure 3.5) and 4 (right panel, lane 2). The expected product for this PCR was 771bp long. Lane 1 in both gels represents DNA HyperLadder IV. The lane 2 in both gels represents the expected products and the lane 3 represents (in both gels) the negative control for the two PCRs.

The next step was to digest both construct GXG*Alb3Δβ (Figure 3.7) and the amplified region of PCR5 (Figure 3.8) with two restriction enzymes (*BamH*I and *Acc*I). The double digestion of the product of PCR 5 produced 2 bands (Figure 3.8). The lower band of ~400bp was ligated with the linearised vector. The transformation, colony PCR and sequencing to confirm the identity of construct GAG*Alb 3Δβ was carried out in similar way to that for GXG*Alb 3Δβ.

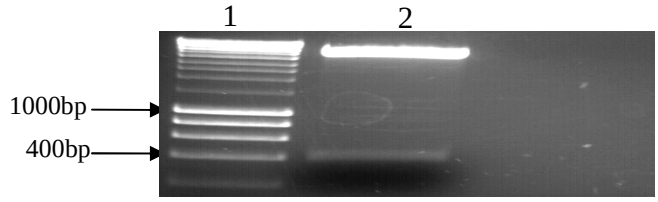


Figure 3.7 Agarose gel electrophoresis showing the result of the double digestion of construct GXG*Alb3Δβ. Digestion using *BamH*I and *Acc*I produced two fragments that included a ~400bp insert. Lane 1 represents the DNA ladder I and lane 2 shows the digested plasmid with its two fragments.

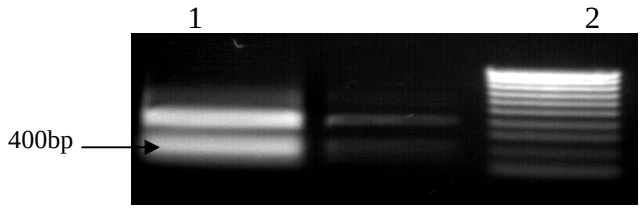


Figure 3.8 Agarose gel electrophoresis showing the double digestion of the product of PCR5 (see figure 3.6 for details). The cleavage site for *BamH*I is located upstream of *Gaussia* 5'UTR and the cleavage site for *Acc*I is in the middle of *Gaussia* Luciferase coding region. Double digestion of the final amplified region produced two fragments. Lane 1 represents the double digestion products of PCR5. Lane 2 shows DNA HyperLadder IV.

3.3.3 Construction of pTRE2hyg GAG*G 3Δβ

The strategy to make constructs GAG*G 3Δβ was to double digest the construct GG*G*G 3Δβ and replace its signal peptide with the Albumin signal peptide. This required the double digestion of previously made construct GAG*Alb3Δβ in order to obtain the Albumin signal peptide for the ligation process. Constructs were double digested with restriction enzymes *BamH*I and *Acc*I. The product of digestion (insert) was used for ligation with the double digested GG*G*G 3Δβ vector (Figure 3.9).

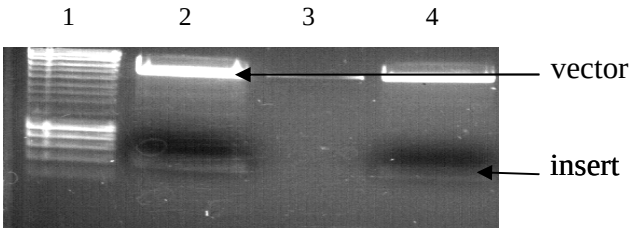


Figure 3.9 Agarose gel electrophoresis showing the double digestion of constructs GG*G*G 3Δβ (lane 2) and GAG*Alb 3Δβ (lane 4). Double digestion of both constructs using *BamH*I and *Acc*I linearised the vectors and produced for both a ~400bp band that represented the insert. Lane 1 is DNA ladder I and lanes 2, 4 show the digested vectors. Lane 3 is an empty well.

The transformation, colony PCR and sequencing to confirm the identity of construct GAG*G* 3Δβ was carried out in similar way to that for GXG*Alb 3Δβ.

3.4 Albumin 3'UTR deletion constructs

3.4.1 Construction of pTRE2hyg GG*G*Alb 3Δβ (Δ51-150)

Two novel deletion constructs were made using the construct GG*G*Alb 3Δβ with deletion of 100 nucleotides within their 3'UTR (Figure 3.10). The strategy was to use primers that amplify specific regions of the Albumin 3'UTR region but exclude the amplification of region/s base 51-150 for construct GG*G*Alb 3Δβ (Δ51-150) or regions 1-50 &101-150 for construct GG*G*Alb 3Δβ (Δ1-50&101-150). Crucially, the amplified region/s must cover the restriction sites for *Acc*I and *Pflm*I.

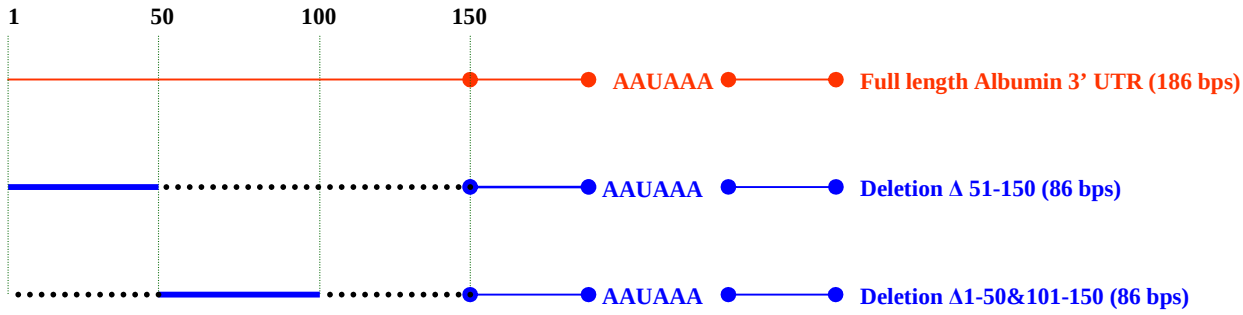


Figure 3.10 Schematic representation of the full length Albumin 3'UTR (red) and its deletion constructs (blue). The broken lines represent the deleted regions. The hexamer AAUAAA represent the polyadenylation signal.

To make the construct GG*G*Alb 3Δβ (Δ51-150), hereafter referred to as Δ51-150, initially 2 PCRs were carried out in order to amplify a product that covered the restriction sites for *Acc*I and *BamH*I (Figure 3.11). The reverse primer for the PCR1 (Alb50.151.Rev) was designed so that its 3'- tail would bind to base 50-42 and its 5'-portion would bind to bases 177-151 in the Albumin 3'UTR. The forward primer for this PCR (G3Da.for) was designed so that it would bind to a region in the Luciferase coding sequence in order to accommodate the *Acc*I restriction recognition site. This PCR produced a product that was approximately 400bp long (Figure 3.12 lane 2, left panel). For the second PCR, the forward primer (Alb.151.50.For) was designed so that its 5'-portion would bind to bases 151-164 and its tail shared sequence homology with bases 29-50.

The reverse primer (BCD. rev) for this PCR was designed so that it would bind to a region in downstream sequences to accommodate the restriction site *Pflm*I. This PCR produced a 530bp product (Figure 3.12 lane 3, left). Finally to anneal the product of PCR 1 and PCR 2, the third PCR was carried out with the forward primer of PCR 1 and the reverse primer of PCR 2 following PCR purification of both products (Qiagen PCR purification Kit)**.** This PCR produced a ~900bp product (Figure 3.13 left panel). Next the product of PCR 3 was purified

with the Qiagen PCR Purification Kit and double digested with *Acc*I and *Pflm*I restriction enzymes (Figure 3.13**,** right panel lane 2).

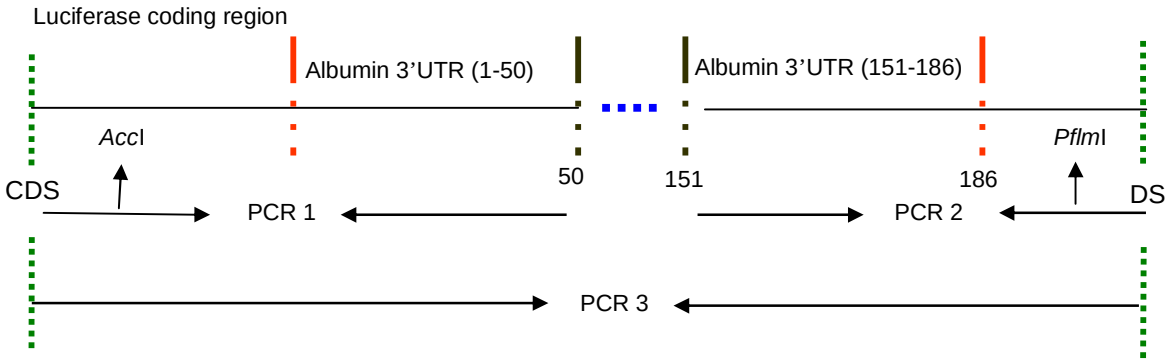


Figure 3.11 Schematic representation of the strategy used to make the deletion construct Δ 51-150. The broken line in blue represents the deleted region within the Albumin 3'UTR. The vertical red lines represent the beginning and end of the Albumin 3'UTR. The double vertical green lines represent the Luciferase coding sequence/region and downstream sequences (CDS and DS) respectively. The two arrows indicate the restriction site for *Acc*I and *Pflm*I.

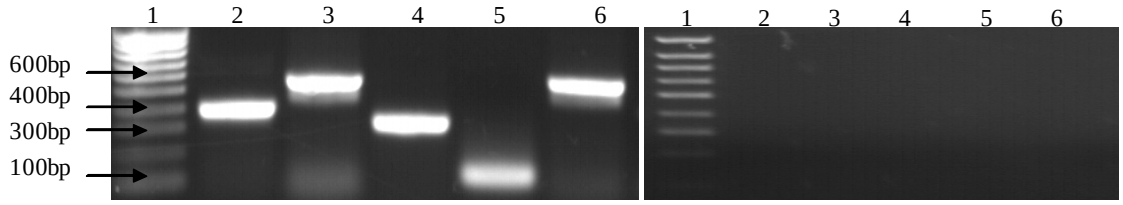


Figure 3.12 Agarose gel electrophoresis showing the results of 5 PCRs (left panel) and their negative controls (right panel) to create the Albumin deletion constructs. To make deletion construct Δ51-150, 2 separate PCRs were performed (lane 2 and 3, left panel). These products were then subjected to PCR purification and linked together to obtain the final product (PCR3) using the forward primer used for PCR 1 and the reverse primer used for PCR 2. Lane 1 (in both gels) represents DNA HyperLadder IV. Lanes 2&3 in the right panel show the negative controls for PCR1&2. Lane 4, 5 and 6 from left panel represent the products of 3 PCRs that were used for the construction of Δ1-50&101-150. Lane 4, 5 and 6 (right panel) represent the respective negative controls for the 3 PCRs.

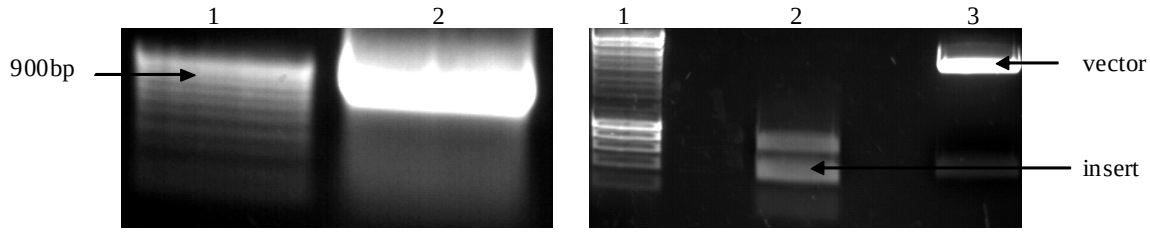


Figure 3.13 Agarose gel electrophoresis showing the final product of PCR3 (see Figure 3.11 for details) for deletion construct Δ51-150 (left panel). Lane 1 (left panel) represents the DNA HyperLadder1V. Lane 2 (left panel) represents the final product. The gel on right shows the double digestion (lane 2) of PCR3 and construct GG*G*Alb 3Δβ (lane 3) to obtain the insert and the linearised vector for the ligation. Lane 1 (right panel) represents the DNA HyperLadder I.

The transformation, colony PCR and sequencing to confirm the identity of construct GG*G*Alb 3Δβ (Δ51-150) was carried out in similar way to that for GXG*Alb 3Δβ.

3.4.2 Construction of pTRE2hyg GG*G*Alb 3Δβ (Δ1-50&101-150)

To make the deletion construct GG*G*Alb 3Δβ (Δ1-50&101-150), hereafter referred to as Δ1-50&101-150, 5 PCRs were performed (Figure 3.14).

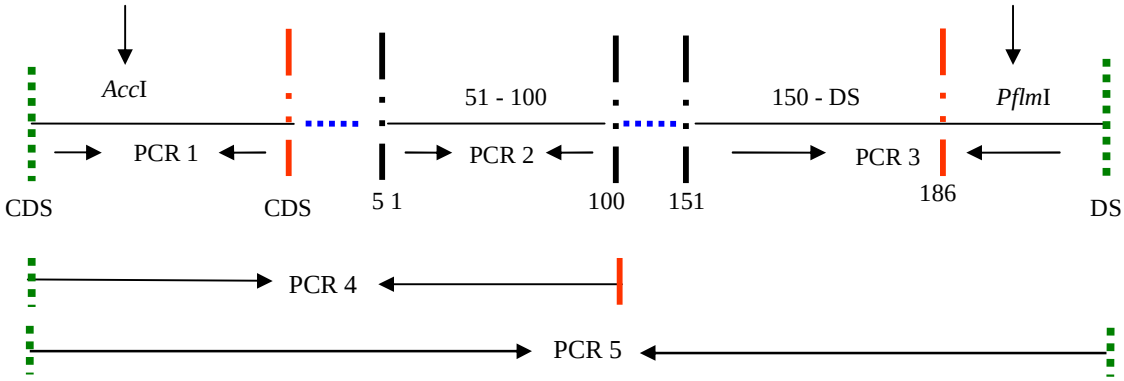


Figure 3.14 Schematic representation of the strategy for making the double deleted construct Δ1-50&101-150. 5 PCRs were performed to amplify multiple regions. Regions in broken blue represent the deleted regions. Regions in vertical red lines represent the beginning and end of the Albumin 3'UTR. Two vertical green lines at either ends represent the Luciferase coding sequence/region and downstream sequences (CDS and DS). The two vertical arrows indicate the restriction sites for *Acc*I and *Pflm*I.

PCR1 was performed with primers (G3Da.for and G3Db.rev) that amplify a region within the *Gaussia* Luciferase coding region toward its very end. This was carried out to accommodate the *Acc*I restriction site. The amplified product of this PCR was approximately 300bp long (figure 3.12 lane 4). Next to perform PCR 2, the forward primer (Alb51.for) for this PCR was designed in such a way so that its 3'-end would share sequence homology with a region within the *Gaussia* coding region to the very end of this region (nucleotides 486-507) and its 5'-portion would bind to nucleotides 51-64 of the Albumin 3'UTR. The reverse primer (Alb100.151.Rev) for this PCR was designed so that its 3'-end would bind to positions 101-92 and its 5'-portion shared sequence homology with bases 177-153 of the Albumin 3'UTR.

The amplified product of this PCR was ~ 100bp long (Figure 3.12 lane 5, left panel). PCR 3 was performed to amplify base 151 to a region within the downstream sequences in order to accommodate the *Pflm*I restriction site. The forward primer (Alb151.100For) for this PCR was designed so that its 5'-portion would bind to bases 151-164 of the Albumin 3'UTR and its 3'-end shared sequence homology with bases 79-100 of the Albumin 3'UTR. The expected product of this PCR was ~ 550bp long (Figure 3.12 lane6, left panel). The next task was to carry out PCR 4 and link the products of PCR 1 and 2 following PCR purification of both products with the Qiagen PCR purification kit. To perform this PCR, the forward primer of PCR1 and reverse primer of PCR 2 were used. The expected product of this PCR was ~ 400bp long (Figure 3.15, left panel). Finally, PCR 5 was performed with the forward primer of PCR 1 and the reverse primer of the PCR 3 (Figure 3.15, right panel). This PCR was carried out to link products of PCR 1-3. The amplified region was ~ 950bp long. The product of PCR 5 was purified with Qiagen PCR purification kit and then treated with restriction enzymes *Acc*I and *Pflm*I (Figure 3.16) along with the construct GG*G*Alb 3Δβ, and then the product ligated into the vector.

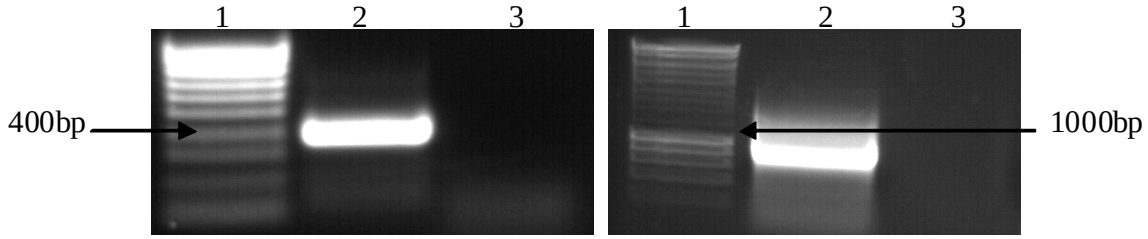


Figure 3.15 Agarose gel electrophoresis showing the product of PCR 4 (lane 2, left panel, see Figure 3.14 for details) and the product of PCR 5 (lane 2, right panel, see Figure 3.14 for details). The final product PCR 5 was purified with PCR Purification Kit (Qiagen) and then treated with the restriction enzymes *Acc*I and *Pflm*I to obtain the insert. Lane 1 of both gels represents the DNA ladder (HyperLadder IV in left panel, HyperLadder I in right panel). Lane 3 in either picture represents the negative control.

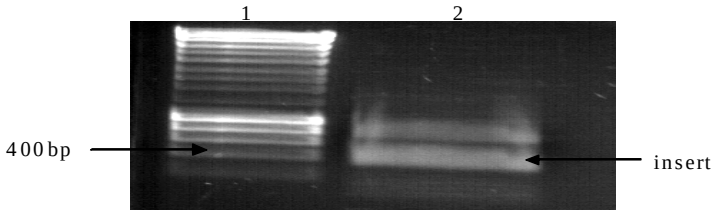


Figure 3.16 Agarose gel electrophoresis showing double digestion of the product of PCR 5 (see Figure 3.14 for details) with the restriction enzymes *Acc*I and *Pflm*I. Lane 1 represents the DNA HyperLadder I. Lane 2 represents the digested product of PCR 5.

The transformation, colony PCR and sequencing to confirm the identity of construct GG*G*Alb 3Δβ (Δ1-50&101-150) was carried out in similar way to that for GXG*Alb 3Δβ.

3.5 Summary

To study the elements that influence post-transcriptional regulations, such as signal peptide and 3'UTR, a series of gene constructs were created. The gene constructs were made by using a seamless cloning method that allows two or more DNA fragments precisely to be linked together, therefore avoiding the inclusion of unwanted sequences [186]. To identify the presence of the sequence of interest, colony PCR was performed by using individual colony that were obtained after ligation of vector and inserts. Finally, pair wise alignment was carried out to verify the creation of constructs.

4 Manipulation of signal peptide/ 3'UTR and its effect on protein expression and mRNA abundance

4.1 Introduction

Reporter systems have turned into priceless tools to study gene expression. They have rapidly become an integral part of scientific work in various fields, from molecular biology to biomedical and pharmaceutical research. Luciferases, as some of the most common and widely used reporter systems are being frequently used to examine the regulation of gene expression. Recently *Gaussia* Luciferase has been used to assess effects of modification of expression vectors on recombinant protein synthesis/secretion so as to develop vectors and mammalian cell expression technology [187]. In this study *Gaussia princeps* Luciferase was chosen as reporter system to assess elements such as signal peptide and 3'UTR that influence post-translation regulatory mechanisms and subsequently affect the expression/secretion of a reporter protein.

Novel gene constructs were made by utilising seamless cloning procedure which contained the *Gaussia* Luciferase 5'UTR, a signal peptide coding region, *Gaussia* coding region and various 3'UTR of interest (see chapter 3). These novel gene constructs were then stably or transiently transfected into CHO cells. Following transfection, culture medium and cell extract samples were obtained to assess activity of the reporter protein. To study the effect of signal peptide hydrophobicity on Luciferase activity and mRNA abundance level, cells were transfected with constructs that contained the Chymotrypsinogen or Albumin signal peptide or the native *Gaussia* Luciferase signal peptide. Chymotrypsinogen is a precursor of the digestive enzyme, Chymotrypsin that is involved in the breaking down proteins into amino acids. This inactive enzyme is produced by pancreatic acinar cells. Human serum albumin,

in contrast, is the most abundant protein in blood plasma and is produced in the liver. The serum albumin is involved in maintaining the osmotic pressure and the transport of fatty acids, thyroid hormones, unconjugated bilirubin and drugs.

The selection of the Chymotrypsinogen and the Albumin signal peptides was based mainly on their pattern of expression. Chymotrypsinogen is produced in response to stimuli, which mimics the bulk secretion of the *Gaussia princeps* Luciferase in nature. This is in direct contrast to the constitutive production of albumin directly into blood from hepatic cells [187].

To determine Luciferase activity by cells transfected with the gene constructs of interest, detectable light energy (Relative Light Unit) was measured by a Luminometer using both culture medium and cell extracts. In many experiments but not all, the construct containing the native *Gaussia* signal peptide and *Gaussia* 3'UTR was chosen as standard reference construct (GG*G*G) and Luciferase activity in these cells was set as 100%. The Luciferase activity in culture medium and cell extract for other constructs was expressed as a percentage of Luciferase activity for the standard reference construct GG*G*G in culture medium. Furthermore, total RNA extraction was performed to assess whether manipulation of 3'UTR correlates with protein expression. To achieve this Real-Time PCR was performed to quantify mRNA abundance relative to the level of housekeeping gene, glyceroldehyde-3-phosphate dehydrogenase (GAPDH).

4.2 Results

4.2.1 Increasing signal peptide hydrophobicity and its effects

Prior to this study a series of constructs had been created (JEH laboratory) in which mutations (see table 4.1) were made in the H-domain of the Chymotrypsinogen signal peptide ($_{P}$TRE2hyg GXG*A5 3Δβ, GX1G*A5 3Δβ, GX2G*A5 3Δβ, GXG*G, GX1G*G and GX2G*G). First, a mutation was made so that the amino acid serine was replaced with the highly hydrophobic amino acid phenylalanine as a single mutation (X1). To increase further the hydrophobicity of the Chymotrypsinogen signal peptide, the first threonine residue in the H-domain of Chymotrypsinogen was also substituted with the amino acid alanine as well as the substitution of serine to phenylalanine as double mutation (X2). However during cloning, 5 point mutations were randomly introduced within the Albumin 3'UTR (A5) and mutations were detected by sequencing the constructs.

Table 4.1 Amino acid sequences of the wild type human Chymotrypsinogen signal peptide (X) and its two mutant variants that were investigated in this study. To make X1 Chymotrypsinogen signal peptide, the residue serine was substituted with phenylalanine. To make X2 Chymotrypsinogen signal peptide, the first threonine residue was replaced with alanine as well as the substitution of serine to phenylalanine. The substitute residues are underlined and highlighted.

Signal peptide	Amino acid sequence
X= wild type	M A F L W L L S C W A L L G T T F G MetAlaPheLeuTrpLeuLeuSerCysTrpAlaLeuLeuGlyThrThrPheGly
X1= with 1 mutation	M A F L W L L **F** C W A L L G T T F G MetAlaPheLeuTrpLeuLeu**Phe**CysTrpAlaLeuLeuGlyThrThrPheGly
X2= with 2 mutations	M A F L W L L **F** C W A L L G **A** T F G MetAlaPheLeuTrpLeuLeu**Phe**CysTrpAlaLeuLeuGly**Ala**ThrPheGly

To examine the effect of signal peptide hydrophobicity on Luciferase activity and mRNA abundance level, CHO AA8 Tet off cells were stably transfected with constructs $_{p}$TRE2hyg GXG*A5 3Δβ, GX1G*A5 3Δβ and GX2G*A5 3Δβ. These constructs contained the Human Albumin 3'UTR into which 5 point mutations had been randomly introduced during cloning. Furthermore these constructs did not contain the Rabbit β-globin vector-derived sequences. The β-globin vector-derived sequences labelled as β-globin poly (A) by Clontech, were made of 2 exons and one intron from the β-globin gene as well as the 3'UTR from the same gene. After transfecting the cell line with these constructs, cells were grown for approximately 4 weeks as a mixed population with growth medium containing Hygromycin to maintain selection pressure on the transfected cells. Transfection was carried out on two batches of CHO cells 4 days apart. Culture medium was harvested and cell extracts were prepared. Luciferase activity was measured (see chapter 2) in all samples. For every transfected cell line, 3 parallel cell populations were obtained separately and therefore the results presented here are the average of 6 parallel measurements from a single construct unless otherwise stated. These procedures were followed for all the transfections throughout this study unless otherwise stated.

It would appear from Figure 4.1 that increasing hydrophobicity of the signal peptide (Chymotrypsinogen) had no significant effect on the expression of *Gaussia* Luciferase. One plausible explanation for the observed finding here is that perhaps the native Chymotrypsinogen signal peptide was already sufficiently hydrophobic that converting the amino acid serine to amino acid phenylalanine or amino acid threonine to amino acid alanine had no significant effect on overall hydrophobicity and thus did not influence the reporter activity.

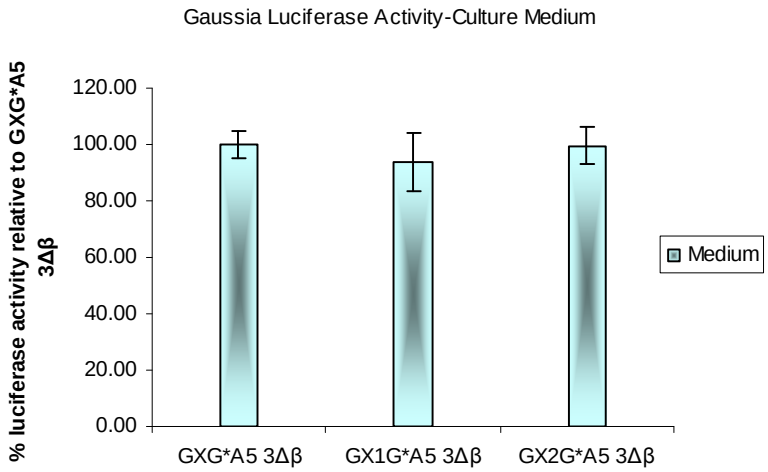


Figure 4.1 *Gaussia* Luciferase activity in the culture medium samples from cells transfected with constructs containing the Albumin 3'UTR (with 5 point mutations) and Chymotrypsinogen signal peptide (with or without mutations). Data are shown as percentage of Luciferase activity for construct GXG*A5 3Δβ. The results showed that increasing signal peptide hydrophobicity had no detectable effect on the activity of the Luciferase reporter. The error bars represent the standard error of the mean.

The data obtained from the *Gaussia* Luciferase activity in the cell extract (Figure 4.2) did not show any statistically significant difference between constructs GXG*A5 3Δβ, GX1G*A5 3Δβ and GX2G*A5 3Δβ.

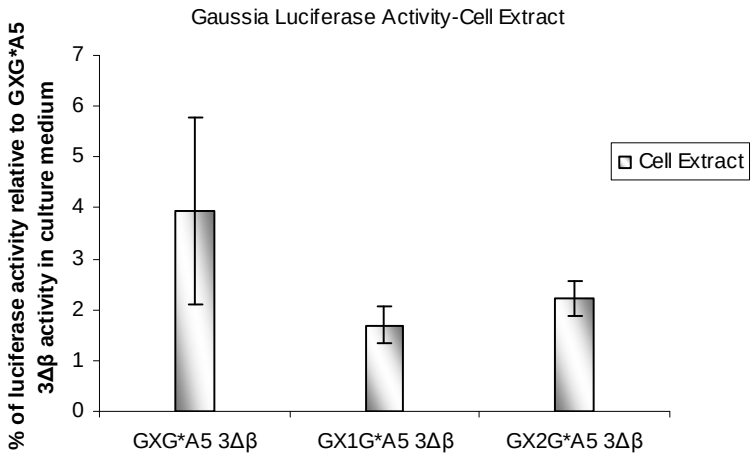


Figure 4.2 *Gaussia* Luciferase activity in the cell extract samples from cells transfected with constructs containing the Albumin 3'UTR and Chymotrypsinogen signal peptide. Data are shown as percentage of Luciferase activity for construct GXG*A5 3Δβ in the culture medium. The error bars represent the standard error of the mean.

In addition CHO AA8 Tet off cells were also stably transfected with constructs GXG*G, GX1G*G and GX2G*G (Figure 4.3). The signal peptide in these constructs was the Chymotrypsinogen signal peptide without mutation in construct GXG*G, with one mutation (substitution of serine with phenylalanine) in construct GX1G*G and two mutations (substitution serine with phenylalanine and first Threonine residue with alanine) in construct GX2G*G. However these construct contained the β-globin vector-derived sequences with *Gaussia* 3'UTR.

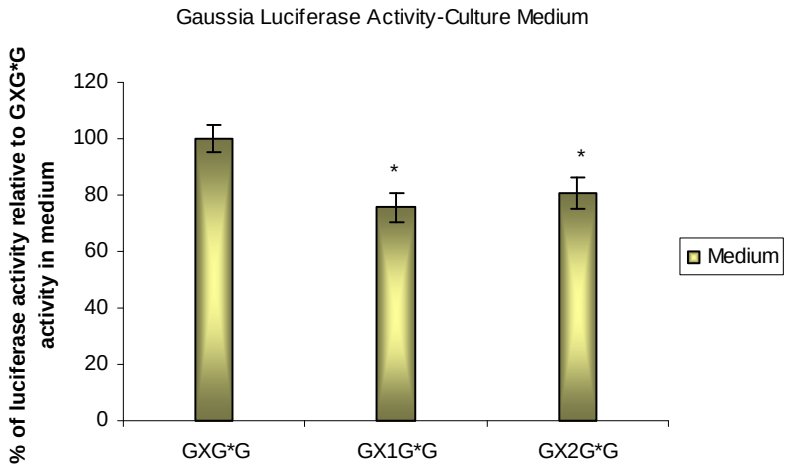


Figure 4.3 *Gaussia* Luciferase activity in the culture medium samples from cells transfected with constructs containing the *Gaussia* 3'UTR and Chymotrypsinogen signal peptide. Data are shown as percentage of Luciferase activity for construct GXG*G. The data presented here suggests that increasing the hydrophobicity of signal peptide significantly reduced the activity of the Luciferase reporter in the culture medium. Statistical significance relative to the construct GXG*G was calculated by using the Mann-Whitney U test where *=p≤0.05 where n=6. The error bars represent the standard error of the mean.

The result in Figure 4.3 suggests that increasing total hydrophobicity of the signal peptide in constructs containing the *Gaussia* 3'UTR significantly reduced the Luciferase activity in the culture medium. This was in contrast to constructs containing the Albumin 3'UTR where increasing the hydrophobicity of signal peptide had no detectable effect on the activity of the reporter protein. In contrast, no statistical difference was observed for the activity of Luciferase reporter in cell extract samples (Figure 4.4).

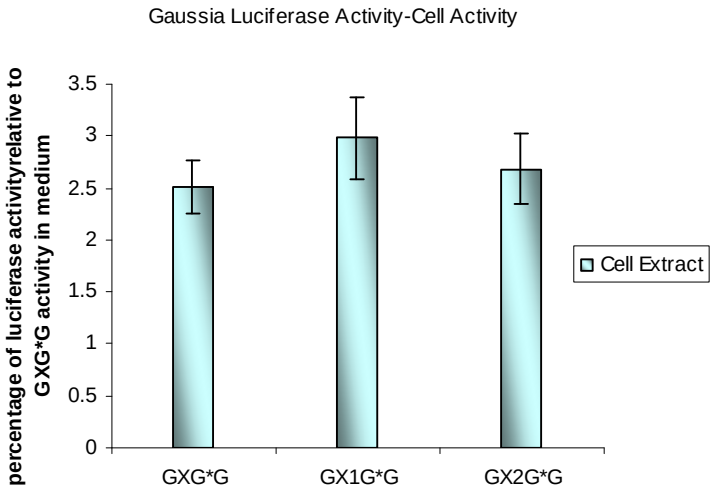


Figure 4.4 *Gaussia* Luciferase activity in the cell extract samples from cells transfected with constructs containing the *Gaussia* 3'UTR and Chymotrypsinogen signal peptide. Data are shown as percentage of Luciferase activity for construct GXG*G in the culture medium. The error bars represent the standard error of the mean.

Overall the data obtained from the cell lines expressing constructs with the Albumin and *Gaussia* 3'UTRs showed that Luciferase activity was significantly higher in constructs containing the Albumin 3'UTR. It must be noted that the Albumin 3'UTR containing constructs did not contain vector-derived sequences. This is in contrast to constructs containing the *Gaussia* 3'UTR that also contained vector-derived sequences (Figure 4.5). Therefore, higher Luciferase activity could be attributed to the deletion of vector-derived sequences, as well as the choice of 3'UTR.

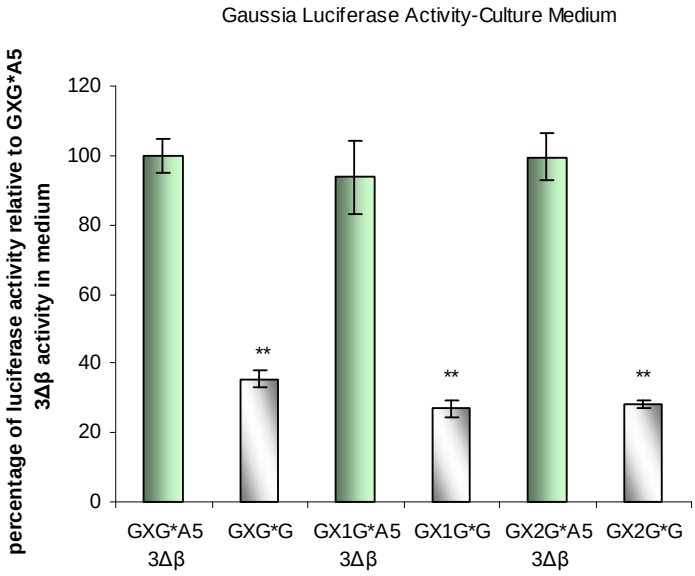


Figure 4.5 Luciferase activity in culture medium samples from cells transfected with constructs containing the Albumin and *Gaussia* 3'UTRs. Statistical significance relative to the construct GXG*A53Δβ was calculated by using the Mann-Whitney U test where **=p≤0.01 and where n=6. The error bars represent the standard error of the mean.

Interestingly, the results of Luciferase activity in cell extract for both groups of construct (Figure 4.6) mimicked the Luciferase activity in the culture medium.

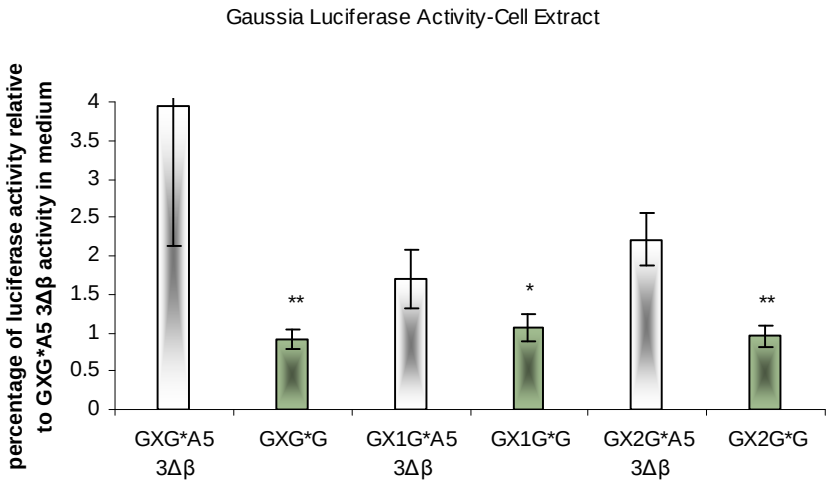


Figure 4.6 Luciferase activity in cell extract samples from cells transfected with constructs containing the Albumin and *Gaussia* 3'UTRs. Luciferase activity is shown as percentage of the Luciferase activity relative to GXG*A5 3Δβ activity in culture medium. Statistical significance relative to the construct GXG*A5 3Δβ was calculated by using the Mann-Whitney U test where **=p≤0.01 and where n=6. The error bars represent the standard error of the mean.

However, it should be noted that downstream vector-derived sequences had been removed from constructs containing the Albumin 3'UTR but not from those with the *Gaussia* Luciferase 3'UTR and therefore significantly higher Luciferase activity could be attributed either to removal of vector-derived sequences or the difference in 3'UTR. The significantly higher Luciferase activity observed for constructs containing the Albumin 3'UTR compared to *Gaussia* 3'UTR could not have been due to utilisation of the Chymotrypsinogen signal peptide since both groups of constructs had Chymotrypsinogen as the signal peptide. It is however possible that β-globin vector-derived sequences could alter the secondary structure of the mRNA and as a result de-stabilise the mRNA transcripts, therefore potentially leading to lower Luciferase activity.

4.2.2 Removal of vector-derived sequences and its effect on the secretion of Luciferase

Overall it would appear from Figures 4.5 that constructs containing the Albumin 3'UTR showed significantly higher Luciferase activity than constructs containing *Gaussia* 3'UTR. The data also suggested that increasing the signal peptide hydrophobicity had little or no significant effect on the reporter activity in constructs containing the Albumin 3'UTR. The higher Luciferase activity in constructs containing the Albumin 3'UTR could have been due to the choice of 3'UTR or the removal of β-globin vector-derived sequences since the two groups of constructs differed only in absence or presence of vector-derived sequences and the choice of 3'UTR.

To investigate the positive/negative effect of β-globin vector-derived sequences on the Luciferase secretion and reporter activity, CHO AA8 Tet-Off cells were stably transfected with construct GG*G*G (containing the vector-derived sequences) and construct GG*G*G 3Δβ (without vector-derived sequences). The results showed that reporter activity and Luciferase secretion improved significantly by removing the β-globin vector-derived sequences in GG*G*G 3Δβ (Figure 4.7). This was in agreement with the findings by UniTargetingResearch AS, Bergen [188].

The significantly higher Luciferase activity in cells expressing the construct GG*G*G 3Δβ compared to construct GG*G*G could be directly attributed to the deletion of vector-derived sequences since both constructs utilised the endogenous *Gaussia* Luciferase signal peptide and 3'UTR and were identical except for loss of vector-derived sequences.

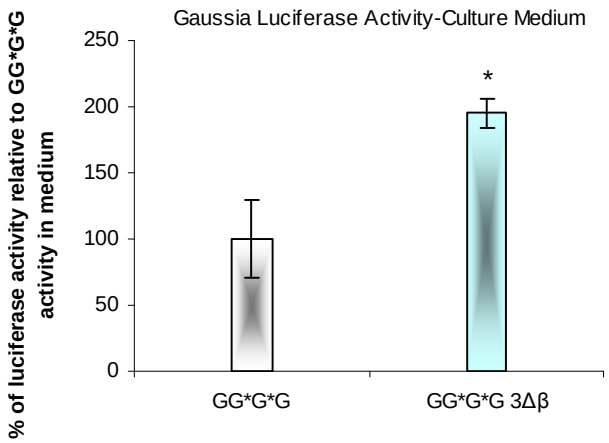


Figure 4.7 *Gaussia* Luciferase activity in the culture medium samples shown as percentage of construct GG*G*G Luciferase activity. The results showed that the removal of vector-derived sequences significantly improved the Luciferase activity in construct GG*G*G 3Δβ. Statistical significance relative to the construct GG*G*G was calculated by using the Mann-Whitney U test where *=p≤0.05 and where n=3. The error bars represent the standard error of the mean.

4.2.3 Effect of Albumin signal peptide on Luciferase activity

It would appear from the results in this chapter that when a hydrophobic signal peptide (Chymotrypsinogen signal peptide) was used in combination with the Albumin 3'UTR, significantly higher Luciferase secretion and reporter activity was observed. However, little was known about the Albumin signal peptide and its effect on Luciferase activity. Work by others [187] had shown that using the Human Albumin signal peptide, in conjunction with the native *Gaussia* 3'UTR, had negative effect on reporter activity and Luciferase secretion.

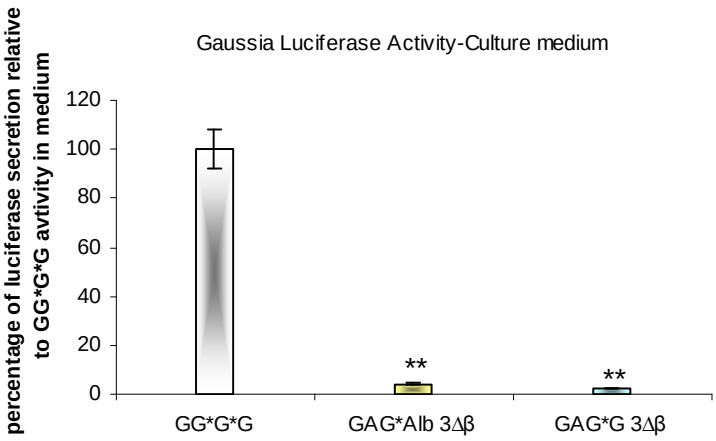


Figure 4.8 Luciferase activity in the culture medium samples shown as percentage of the standard reference construct GG*G*G Luciferase activity. The results show that replacing the hydrophobic *Gaussia* signal peptide with the Albumin signal peptide reduces the reporter activity. Statistical significance relative to the construct GG*G*G was calculated by using the Mann-Whitney U test where *=p≤0.05, **=p≤0.01 and where n=6. The error bars represent the standard error of the mean.

The significantly lower Luciferase activity in constructs GAG*Alb 3Δβ and GAG*G 3Δβ could have been due to the lower Albumin signal peptide hydrophobicity compared to the native *Gaussia* Luciferase signal peptide, as demonstrated in Figure 4.9, and deletion of vector-derived sequences. Furthermore, it is conceivable to consider the possibility of mis-targeting of the mRNA transcripts by an insufficiently hydrophobic Albumin signal peptide. Since the reporter activity of the Albumin signal peptide has been estimated by UniTargetingResearch AS, Bergen, at 2% of the native *Gaussia* signal peptide, it remains to be seen whether replacing less hydrophobic amino acids within the Albumin 3'UTR with more hydrophobic amino acids such as alanine would positively influence reporter activity.

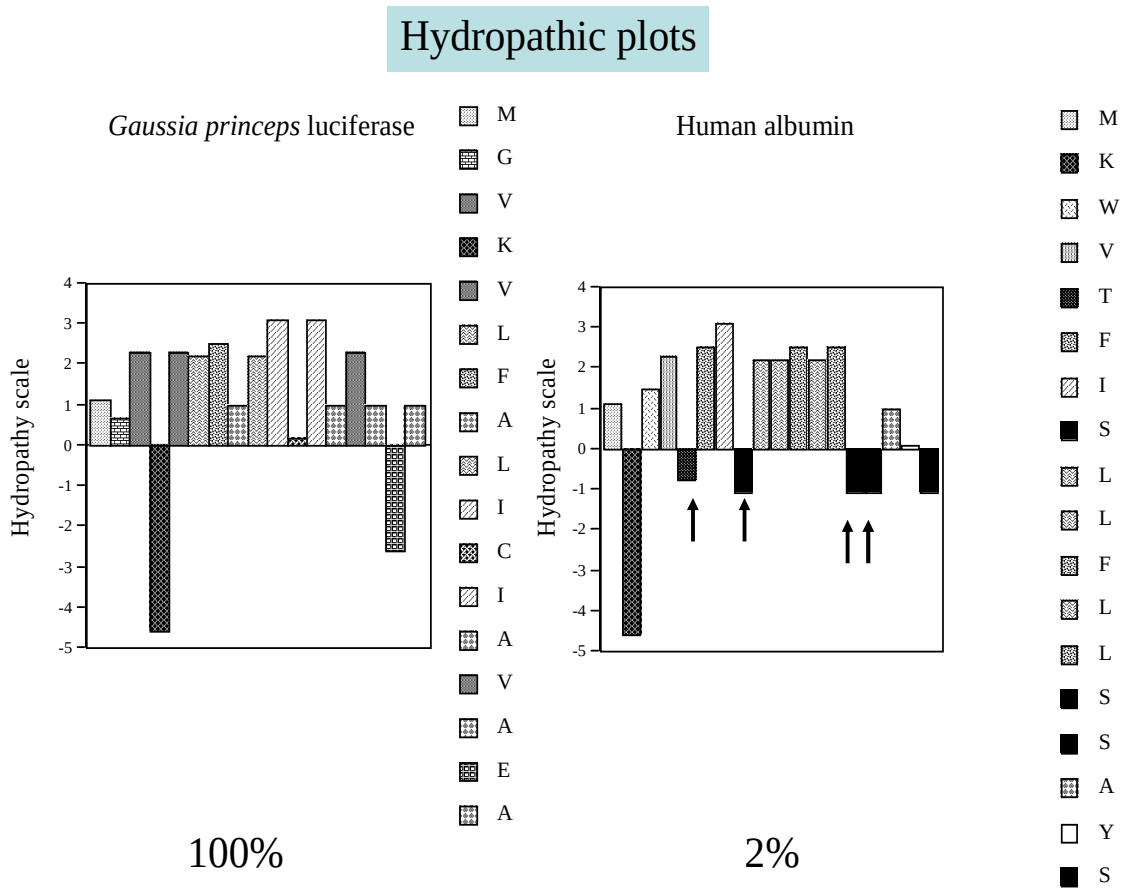


Figure 4.9 Human Albumin and the native *Gaussia* Luciferases hydrophobicity plot based on hydrophathy scale for amino acid residues adopted from Eisenberg *et al* 1982 [189] (courtesy of UniTargetingResearch AS, Bergen). Human Albumin signal peptide activity is shown as percentage of the native *Gaussia* Luciferase signal peptide. The arrows in the Albumin signal peptide indicate possible targets for making amino acid substitutions.

4.2.4 Effect of signal peptide hydrophobicity/3'UTR on the mRNA expression level

To investigate the effect of signal peptide hydrophobicity/3'UTR on mRNA expression level, RNA was prepared (see chapter 2) from cells expressing all constructs that were previously used in transfection experiments. For every transfected cell line, 3 parallel cell populations were obtained separately. Then RT-PCR was carried out to obtain the cDNA. Finally Real-Time PCR was performed on the cDNA to obtain a normalised relative ratio of *Gaussia* Luciferase over the housekeeping gene GAPDH. Next the mean of normalised relative ratio for each set of samples was introduced as percentage of the mean of construct GG*G*G samples.

The data obtained from the mRNA quantification (Figure 4.10) showed a similar pattern to the results of *Gaussia* Luciferase activity in the culture medium in that mRNA abundance level was significantly higher in constructs *containing* the Albumin 3'UTR (without vector-derived sequences) than construct with the *Gaussia* 3'UTR (with vector-derived sequences). The data also suggested that deletion of β-globin vector-derived sequences in construct GG*G*G 3Δβ also significantly improved the expression level of mRNA. Interestingly, as previously seen with the Luciferase activity, increasing the hydrophobicity of Chymotrypsinogen signal peptide had little or no detectable effect on the mRNA expression level in both the constructs containing the Albumin 3'UTR/ *Gaussia* 3'UTR with the Chymotrypsinogen signal peptide. Finally no significant difference was observed in the level of mRNA expression between construct GG*G*G and the constructs with the Chymotrypsinogen signal peptide (with or without mutations) and *Gaussia* 3'UTR.

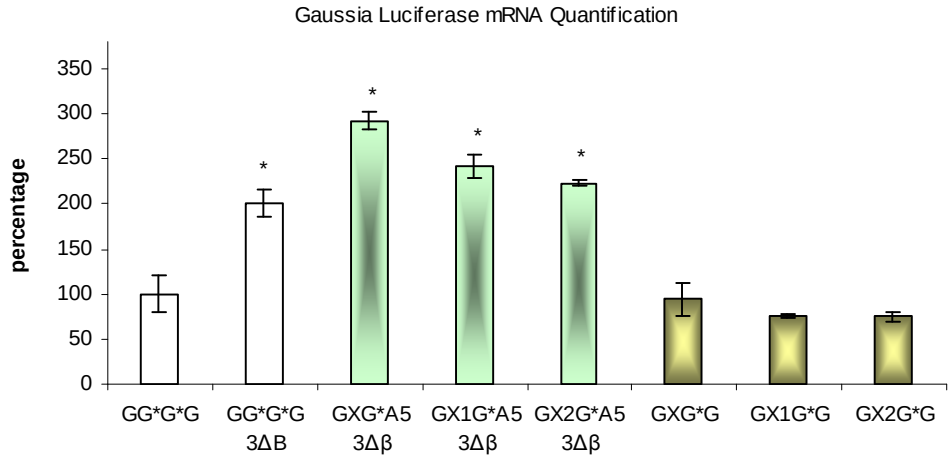


Figure 4.10 Relative quantification of mRNA abundance compared to standard reference construct GG*G*G. The quantification of mRNA is relative to the house keeping gene GAPDH. The data showed that mRNA expression was significantly higher in constructs containing the Albumin 3'UTR. The results also showed that deletion of β-globin vector-derived sequences significantly improved the expression of mRNA in constructs GG*G*G 3Δβ. Statistical significance relative to the construct GG*G*G was calculated by using the Mann-Whitney U test where *=p≤0.05 and where n=3. The error bars represent the standard error of the mean.

4.2.5 Effect of Albumin signal peptide on mRNA expression

Interestingly, there were no significant differences in mRNA expression levels between cells expressing construct GG*G*G and constructs GAG*G 3Δβ or GAG*Alb 3Δβ (Figure 4.11). The fact that mRNA expression levels for constructs GAG*G 3Δβ and GAG*Alb 3Δβ did not differ significantly from construct GG*G*G, perhaps highlighted once again the importance of signal peptide hydrophobicity in its targeting role. The data presented in Figures 4.8 and 4.11 indicated huge discrepancies between the mRNA expression level and protein synthesis. Furthermore, normalisation of *Gaussia* Luciferase activity to the mRNA abundance (see Figure 4.12) for cells expressing constructs GAG*G 3Δβ and GAG*Alb 3Δβ relative to construct GG*G*G clearly showed that Luciferase activity did not correlate with the mRNA abundance. A plausible explanation for this could have been the effects of the Albumin signal

peptide. It is postulated that the Albumin signal peptide causes mis-targeting or inefficient targeting of the mRNA transcript during translocation across the ER (see introduction). This could be due to the low hydrophobicity level and therefore potentially poor activity of the Albumin signal peptide. It would be interesting to know whether replacing the less hydrophobic amino acids within the H-domain of Albumin signal peptide (3 serine residues and 1 threonine) with more hydrophobic amino acids such as phenylalanine improves its targeting function.

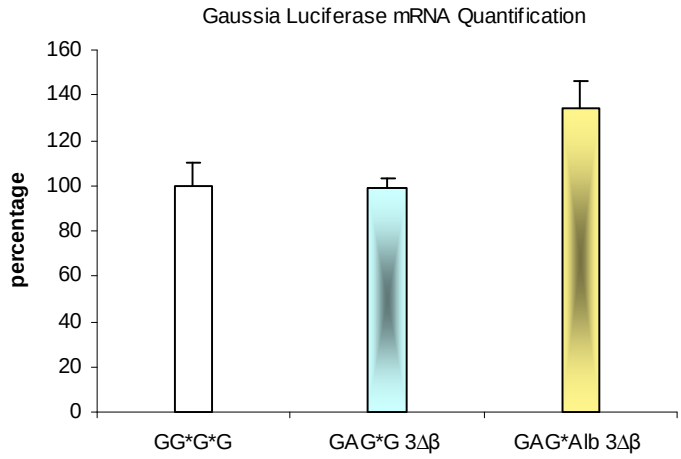


Figure 4.11 Relative quantification of mRNA abundance compared to standard reference construct GG*G*G. The quantification of mRNA is relative to the house keeping gene GAPDH. The data show that no significant difference was observed in mRNA expression between standard reference construct GG*G*G and constructs GAG*G 3Δβ and GAG*Alb 3Δβ. The error bars represent the standard error of the mean.

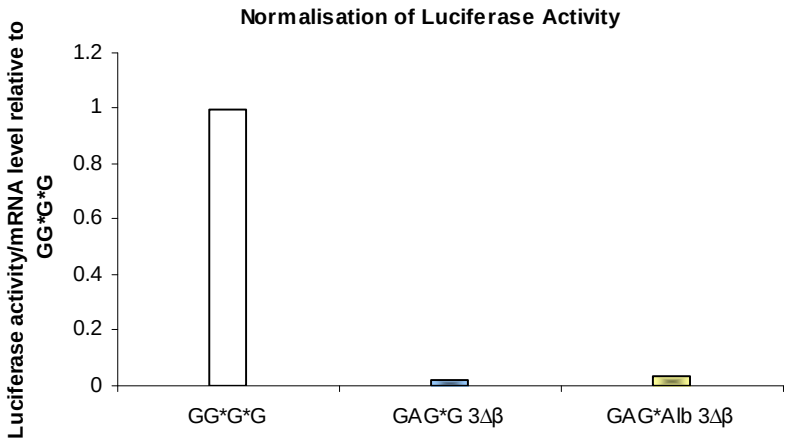


Figure 4.12 Normalisation of Gaussia Luciferase activity over mRNA abundance. The figure shows that the luciferase activity for cells expressing constructs GAG*G 3Δβ and GAG*Alb 3Δβ did not correlate with the mRNA abundance.

4.3 Transient co-transfection

To exclude the site of integration of the transgene as a contributing factor for the observed differences in secreted Luciferase activity in cells expressing the various constructs, transient *co*-transfection of the constructs of interest and firefly Luciferase was performed. Samples of medium and cell extracts were collected and a ratio of *Gaussia* Luciferase activity to firefly Luciferase was established. CHO AA8 Tet-Off cells were co-transiently transfected with construct GG*G*G and constructs containing the *Gaussia* 3'UTR but without vector-derived sequences (GG*G*G 3Δβ, GAGG* 3Δβ) and constructs containing the corrected Albumin 3'UTR for 5 point mutations (GAG*Alb 3Δβ, GXG*Alb 3Δβ).

As it was previously observed in this study, the data presented in Figure 4.13 showed that deletion of β-globin vector-derived sequence improved the secreted Luciferase activity in construct GG*G*G 3Δβ. The data also indicated that highest ratio of *Gaussia* Luciferase activity to firefly Luciferase was observed in construct containing the Albumin 3'UTR and Chymotrypsinogen signal peptide. Furthermore, as it was seen previously in this chapter, replacing the hydrophobic signal peptides *Gaussia* or Chymotrypsinogen with the Albumin signal peptide reduced the ratio of the secreted Luciferase activity over firefly Luciferase. The data obtained from the cell extract samples also showed a similar pattern to those of culture medium samples.

Since the pattern of Luciferase activity observed in various constructs in the transient *co*-transfection was essentially the same as found with the stable transfections, the results would indicate that the positional effect of the plasmid in the genome could not have been a determining factor for the pattern of secreted Luciferase activity for the stable transfections.

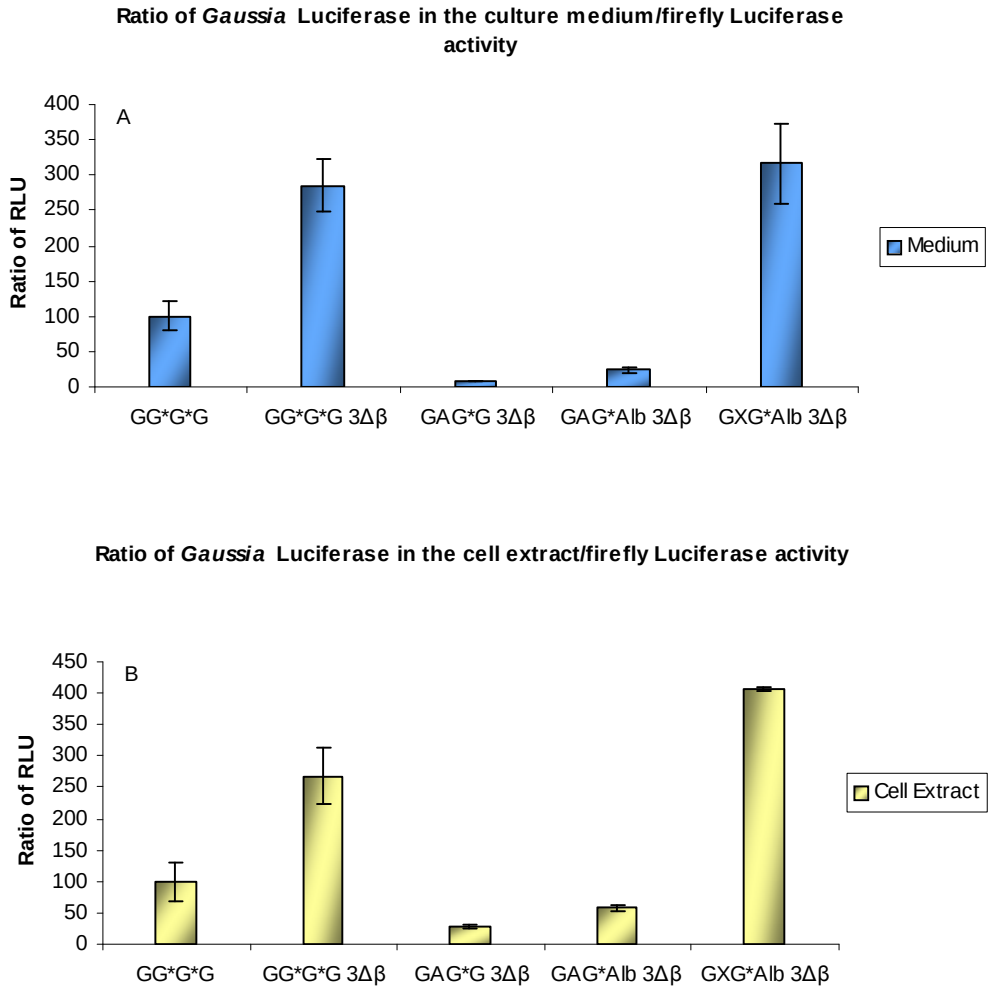


Figure 4.13 Transient co-transfection. Ratio of Relative Light Unit (RLU) values of *Gaussia* Luciferase activity in the culture medium samples to the firefly Luciferase activity in the cell extract (A) and the *Gaussia* Luciferase activity in the cell extract against the firefly Luciferase activity in the cell extract (B). The data presented here suggest that the pattern of *Gaussia* Luciferase activity in transient co-transfection is similar to that observed for stable transfections. The error bars represent the standard error of the mean.

4.4 Summary

It was shown in this chapter that increasing the hydrophobicity of the Chymotrypsinogen signal peptide had little or no detectable influence on reporter protein expression. One plausible explanation for this could be the hydrophobicity limits of the signal peptide in general, that beyond which is not effective as shown by others [41]**.** Overall the data presented in this chapter suggested that the secreted *Gaussia* Luciferase activity and mRNA expression were significantly higher in construct containing the Albumin 3'UTR. The results also strongly suggested that the deletion of vector-derived sequences significantly improved the Luciferase activity and mRNA abundance. This could have been due to alteration or conformation changes in the secondary structure of the mRNA transcript, therefore leading to higher stability of transcript. The data also showed that the highest level of reporter activity and mRNA abundance is achieved where the Chymotrypsinogen signal peptide (with or without mutations) is used in combination with the Albumin 3'UTR. The results in this chapter also suggested that replacing a hydrophobic signal peptide (native *Gaussia* or Chymotrypsinogen) with a less hydrophobic signal peptide (Albumin signal peptide), significantly reduces the protein expression. This not only once more highlighted the importance of hydrophobic signal peptide but interestingly its combination with the 3'UTR. The data presented in this chapter also suggested that differences in secreted Luciferase activity and mRNA expression were not due to differences in the site of integration.

5 Albumin 3'UTR: a deletion analysis

5.1 Introduction

It has widely become recogonised that the 3'UTR is a crucial post-transcriptional regulatory element and regulates mRNA transcripts[190]. 3'UTR sequences regulates gene activity through the interaction of RNA binding proteins and small non-coding RNAs such as miRNAs [190, 191]. These interactions could potentially influence mRNA stability or alter translation and localisation. In the past recent years mRNA stability has emerged as a key mechanism for the rapid regulation of gene expression [192].

It was seen in the previous chapter that constructs containing the Albumin 3'UTR (without vector-derived sequences) consistently promoted high levels of mRNA and reporter activity compared to constructs containing the *Gaussia* Luciferase 3'UTR (with vector-derived sequences) when it was used with an appropriate signal peptide (either the Chymotrypsinogen or native *Gaussia* Luciferase signal peptides). The data presented in chapter 4 showed that the greater level of *Gaussia* Luciferase activity in constructs containing the Albumin 3'UTR compared to control cell populations (construct GG*G*G) was not attributable to differences in the site of integration of plasmids within the genome. However, the question remains - is a particular region of the Albumin 3'UTR responsible for promoting the high level of secreted Luciferase that was observed?

In order to understand how the Albumin 3'UTR or a particular region/s promote Luciferase secretion and reporter activity, a series of deletions were made within the Albumin 3'UTR. The regions deleted were 50,100 and 150 nucleotides long (Figure 5.1). No deletions were made in the last 36 nucleotides of the Albumin 3'UTR.

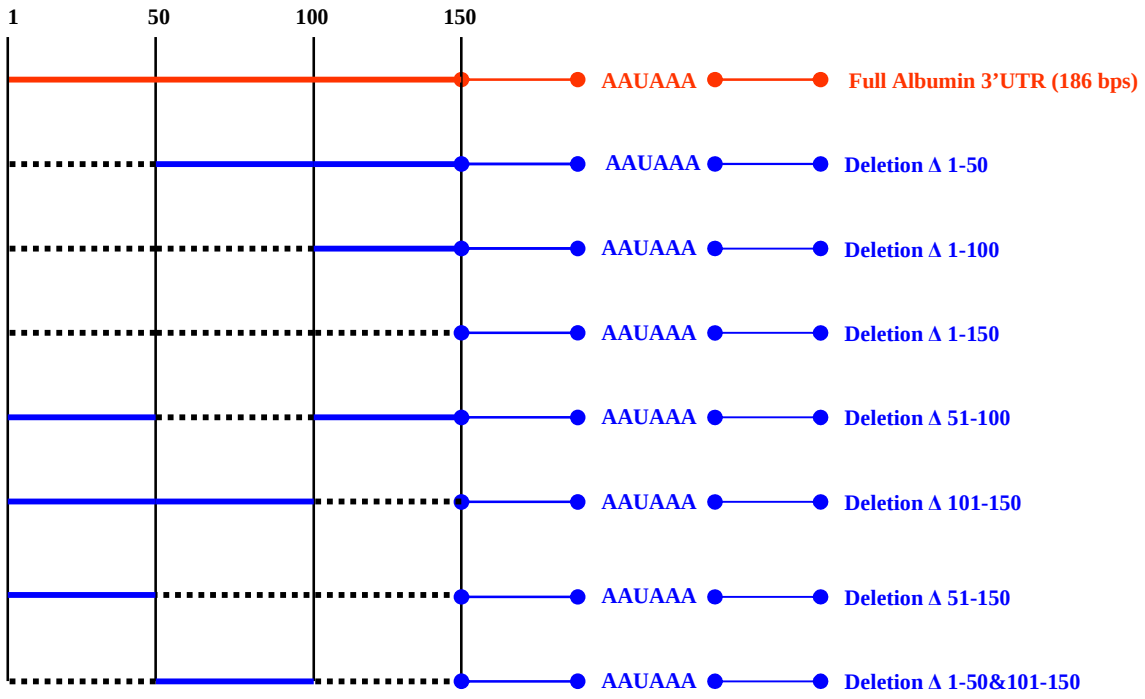


Figure 5.1 Schematic representation of the full Albumin 3'UTR construct (in red) and its mutant variants (in blue). All the deletions were made within the first 150 bps. Regions in broken lines represent the deleted nucleotides. The hexamer AAUAAA represent the polyadenylation signal. Only Deletion construct 51-150 and deletion construct 1-50 & 101-150 were made in this study (see chapter 3). Other deletion constructs were obtained from JEH laboratory.

5.2 Results

5.2.1 Albumin 3'UTR deletion analysis: Luciferase activity

CHO AA8 Tet-Off cells were stably transfected with constructs containing the full length Albumin 3'UTR and its deletion mutants and grown for approximately 4 weeks as a mixed population using growth medium containing Hygromycin to maintain selection pressure on the transfected cells. Transfection was carried out on two batches of CHO cells 4 days apart.

Culture medium was harvested and cell extract samples prepared, and Luciferase activity was measured. For every transfected cell line, 3 parallel cell populations were obtained separately and therefore the results were the average of 6 parallel measurements from a single construct unless stated otherwise.

It would appear from Figure 5.2 that *Gaussia* Luciferase activity is significantly reduced in all the deletion constructs with the exception of Δ51-100 and Δ51-150. The reduction of Luciferase activity in cells expressing the deletion construct 1-50 &101-150 was found to be close to significance (p value = 0.055). The data presented in Figure 5.2 suggest that by deleting the region 51-100, Luciferase production is maintained and deleting other regions leads to either a not significant reduction (Δ51-150), significant reduction (or close to significant reduction) of the reporter activity and Luciferase secretion.

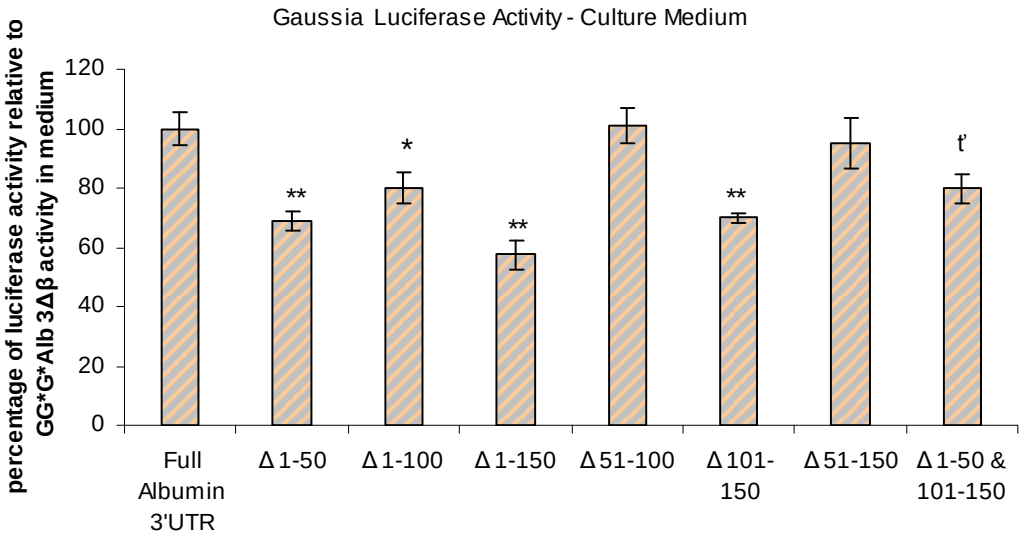


Figure 5.2 *Gaussia* Luciferase activity in the culture medium samples from cells transfected with construct containing the full length Albumin 3'UTR and its deletion variants shown as percentage of Luciferase activity for construct GG*G*Alb 3Δβ. The data presented here show that high levels of Luciferase activity were maintained in deletion constructs 51-100 and 51-150. Statistical significance relative to the construct GG*G*Alb 3Δβ was calculated by using the Mann-Whitney U test where *=p≤0.05, **=p≤0.01, t'=0.055 and where n=6. The error bars represent the standard error of the mean.

The data presented in Figure 5.2 showed that deletion of nucleotides 1-150 leads to highest reduction in Luciferase activity. This suggests that elements that regulate mRNA translation, are located within the first 150 nucleotides of the Albumin 3'UTR. The data also showed that the greatest reduction in secreted Luciferase activity amongst the constructs with a single region deletion was observed with deletion construct Δ1-50 and Δ 101-150, indicating that these regions are required to maintain reporter activity.

Interestingly, the data presented in Figure 5.3 indicated that Luciferase activity in cell extract samples showed a similar pattern to that observed with the culture medium samples. Activity in extracts from cells expressing the deletion constructs Δ1-50 (p value = 0.016), Δ1-150, Δ101-150 and Δ1-50 & 101-150 was significantly reduced. The results presented in Figure

5.3 also showed that the greatest reduction of Luciferase activity was observed in cells expressing the deletion construct Δ1-150. This too was in agreement with the activity pattern to that observed for the culture medium samples for these cells. The data also showed that the reduction of luciferase activity in cells expressing the deletion constructs Δ1-100, Δ51-100 and Δ51-150 relative to the cells expressing the full length Albumin 3'UTR was found not to be significant.

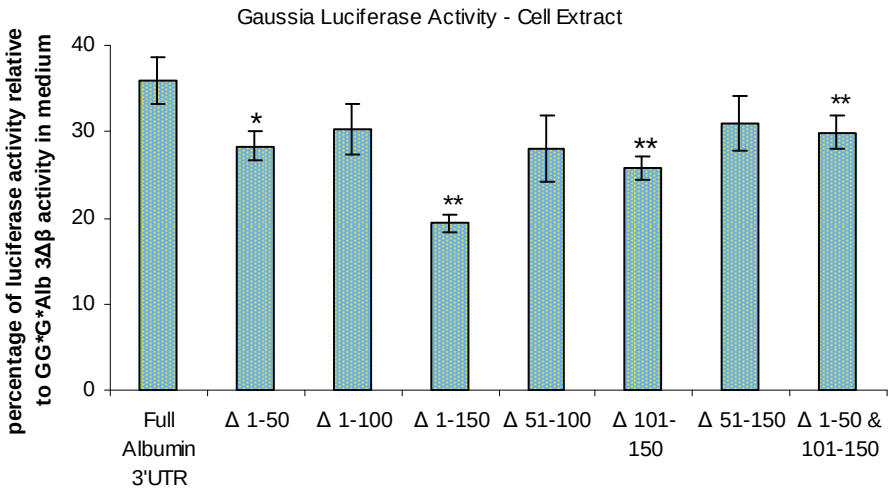


Figure 5.3 *Gaussia* Luciferase activity in cell extract samples prepared from cells transfected with constructs containing the full Albumin 3'UTR and its mutant variants shown as percentage of Luciferase activity for construct GG*G*Alb 3Δβ in the culture medium. Statistical significance relative to the construct GG*G*Alb 3Δβ was calculated by using the Mann-Whitney U test where *=p≤0.05, **=p≤0.01 and where n=6. The error bars represent the standard error of the mean.

Overall, the observations from the Luciferase activity in the culture medium and cell extract indicate that deletion of nucleotides 1-50 or 101-150 has a significantly negative effect on the reporter activity. This could mean that deleting these regions negatively influences the stability or indeed the translation of transcripts. On the other hand it appears that deleting the

region that contains nucleotide 51-100 has little or no negative effect on the stability or translation of transcript.

5.2.2 Albumin 3'UTR deletion analysis: mRNA quantification

To study the effect of Albumin 3'UTR deletions at the mRNA expression level, CHO AA8 Tet-Off cells were stably transfected as described previously with construct containing the full length Albumin 3'UTR and its deletion variants , cells harvested and mRNA extracted. Relative quantification of the mRNA expression for all the Albumin 3'UTR deletion constructs was compared to the *Gaussia* Luciferase mRNA expression of construct GG*G*Alb 3Δβ.

In agreement with the data presented in Figure 5.3, the mRNA expression level was significantly lower in deletion constructs Δ1-50, Δ1-150 and Δ1-50 & 101-150. The data presented in Figure 5.4 showed that reduction of mRNA abundance in cells expressing the deletion constructs Δ1-100, Δ51-100, Δ101-150 and Δ51-150 relative to cells expressing the full length Albumin 3'UTR was found not to be significant.

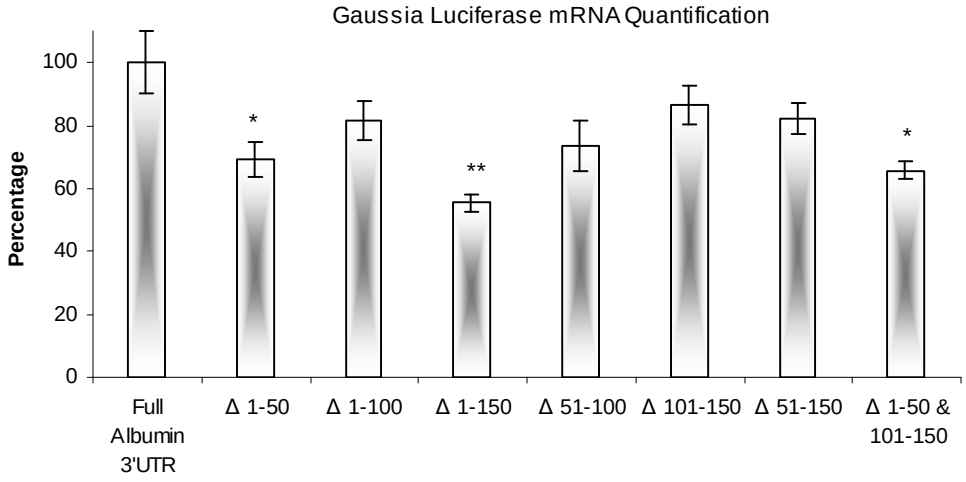


Figure 5.4 Relative quantification of mRNA abundance compared to construct GG*G*Alb 3Δβ. The quantification of mRNA is relative to the house keeping gene GAPDH. The data presented here showed that the level of mRNA expression was significantly lower in deletion constructs Δ1-50, Δ 1-150 and Δ1-50 & 101-150. This mirrored the *Gaussia* Luciferase activity level for these deletion variants. Statistical significance relative to the construct GG*G*Alb 3Δβ was calculated by using the Mann-Whitney U test where *=p≤0.05, **=p≤0.01 and where n=6 for Alb, Del 1-50, Del 1-00, Del 1-150 Del 51-150 and n=5 for Del 51-100, Del 101-150 and Del 1-50&101-150. The error bars represent the standard error of the mean.

The normalisation of *Gaussia* Luciferase activity to the mRNA level (see Figure 5.5) also showed that Luciferase activity for cells expressing constructs Δ1-50, Δ 1-100 and Δ1-50 correlated with the mRNA level, whereas for cells expressing constructs Δ 51-100, Δ101-150, Δ51-150 and Δ1-50&101-150 did not.

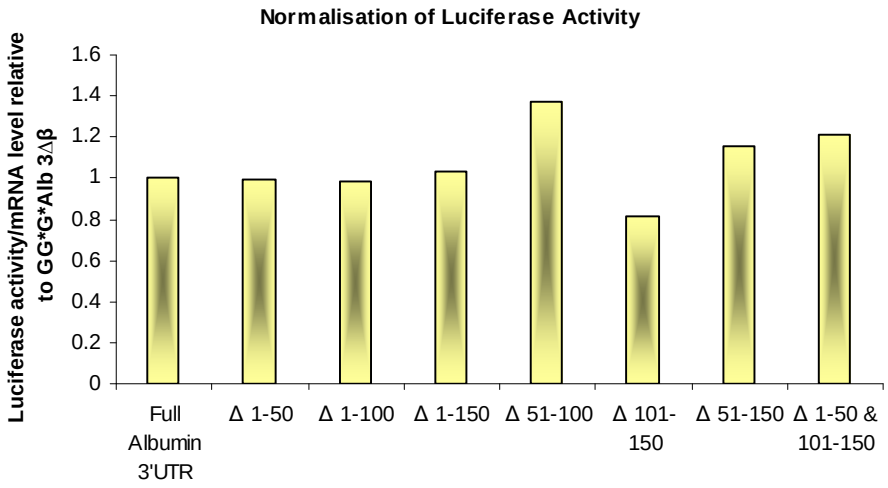


Figure 5.5 The normalisation of *Gaussia* Luciferase activity to mRNA abundance. The figure shows that Luciferase activity for deletion constructs Δ1-50, Δ1-100 and Δ1-150 correlated with the mRNA abundance for those constructs. The figure also shows that reporter activity in the deletion constructs Δ51-100, Δ51-150 and Δ1-50&101-150 was greater than the mRNA abundance whereas this was lower in the deletion construct Δ101-150.

5.3 Summary

Overall it can be observed from the data presented in this chapter that the *Gaussia* Luciferase activity in the culture medium is only maintained where the region containing nucleotide 51-100 is deleted. The data also suggested that deletion of nucleotides 1-50 and 1-100, 1-150 and 101-150 leads to a statistically significant reduction in *Gaussia* Luciferase activity. The reduction of reporter activity in cells expressing the deletion construct Δ1-50 & 101-150 was found not to be statistically significant (p value = 0.055). The data also showed that deletion of nucleotides 1-50, 1-150 and 1-50&101-150 leads to significant reduction in mRNA level.

The greatest reduction in mRNA expression and Luciferase activity was observed by deletion of nucleotides 1-150. One plausible explanation for the significantly lower mRNA expression level and the reporter activity in the culture medium and cell extract for cells expressing the deletion construct Δ1-150 could be due to destabilisation of the mRNA transcripts, which is a result of the deletion of most of the Albumin 3'UTR. It has long been known that for any given RNA molecule, functionality depends on primary sequence as well as the secondary folding[193] (see 8.2.3).

On the other hand, it is speculated that as a result of alteration to the secondary structure of the RNA transcript (see 8.2.2), inhibitory protein/s could bind to a deletion variant such as Δ1-50 more readily than other Albumin 3'UTR deletion variants. This could be plausibly a valid explanation for the significantly lower *Gaussia* Luciferase activity and mRNA abundance in cells expressing the deletion construct Δ1-50.

6. The Albumin 3'UTR and protein binding

6.1 Introduction

It is becoming increasingly apparent that RNA binding proteins play an important role in the regulation of mRNA stability, translation and localisation by associating with the regulatory sequences that are present in mRNA transcripts such as AU-rich sequences, stem loop structures and the poly (A) tail [194]. These *trans*-acting factors influence the regulation of gene expression by modulating events such as nuclear degradation of pre-mRNA, alternative splicing, RNA editing, RNA stability and ultimately the translation of mRNA transcripts. This influence is achieved by association of the *trans*-acting factors with specific sequences/motifs which are present typically in the 5'UTR or the 3'UTR [107].

In chapter 4 and 5 results revealed that using constructs containing the Albumin 3'UTR (without vector-derived sequences) led to a higher level of Luciferase mRNA and secreted protein compared to constructs containing the *Gaussia* Luciferase 3'UTR (with vector-derived sequences). It was postulated that differences in mRNA and protein levels were due to specific binding of protein/s to possible regulatory element/s within the Albumin 3'UTR. In this chapter experiments are described where techniques such as Electrophoresis Mobility Gel Shift Assay (EMSA) and UV Cross-linking were applied to investigate evidence of protein/s binding to the Albumin 3'UTR. Subsequently mass spectrometry analysis was carried out to identify protein/s which bound to the Albumin 3'UTR. This result was then confirmed by utilisation of several experimental techniques such as siRNA knock down and western blotting. In addition, this chapter discusses evidence of siRNA knock down on the

candidate protein/s and its effects on the promotion of Luciferase activity and mRNA abundance is investigated.

6.2 Results

6.2.1 Optimisation of EMSA reaction

To characterise the interaction of proteins in CHO cell extracts with the Albumin 3'UTR, EMSA were performed. The principle for EMSA is that protein/RNA complexes migrate more slowly than the free RNA fragments through a non-denaturing polyacrylamide gel. To allow the protein-RNA interaction, S-100 CHO cell extracts were incubated with the Albumin 3'UTR radiolabelled RNA transcript and the protein and RNA complexes were subjected to electrophoresis in a non denaturing polyacrylamide gel. To optimise the EMSA protocol, different relative ratios of radiolabelled RNA transcript (fmols) to S-100 CHO cell extract (µg); 12 fmols of RNA transcript was incubated with 1, 1.5 and 2µg of protein extracts (see Materials and Methods). As shown in Figure 6.1, complex formation was seen as a distinct shift in migration as the amount of protein extract increased. Complex formation was achieved with 1.5 and 2 µg of protein. In this study hereafter 12 fmols of RNA transcript and 2 µg of S-100 CHO cell extracts were used for an EMSA reaction to ensure complex formation.

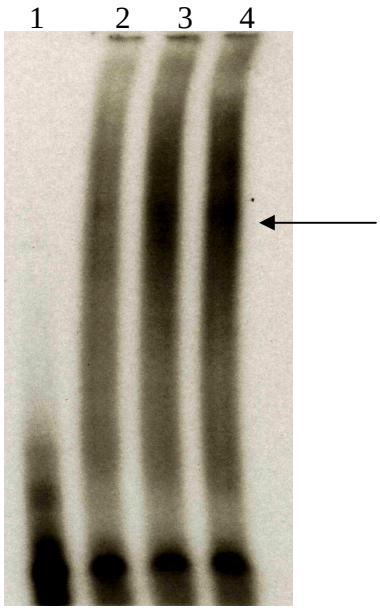


Figure 6.1 Optimisation of EMSA reaction to investigate protein binding to the full Albumin 3'UTR. Complex formation is seen as distinct shift in molecular weight after increasing the amount of S-100 CHO cell extract. Lane 1 shows the radio-labeled RNA transcript alone (control reaction). Lanes 2-4 represent RNA-protein binding complex formation with 1, 1.5 and 2 µg of protein, respectively. The arrow indicates a distinct shift in migration of complex formation in comparison to control reaction.

6.2.2 Determining the specificity of protein binding to the Albumin 3'UTR-EMSA

To determine the specificity of the RNA-protein interaction and complex formation in EMSA, competition reactions were carried out using non-labelled full Albumin 3'UTR and β-globin 3'UTR as specific and non-specific competitors respectively. Both specific and non specific competitors were synthesised with the MEGAshortscript^{TM} Kit (Ambion) by performing *in-vitro* transcription (see 2.17.2) and subsequently quantified by spectrophotometry. Both specific and non-specific competitors were added to the EMSA reaction in 20, 40 and 80 molar excess. As show in Figure 6.2, less complex formed as the amount of specific Albumin 3'UTR competitor was increased.

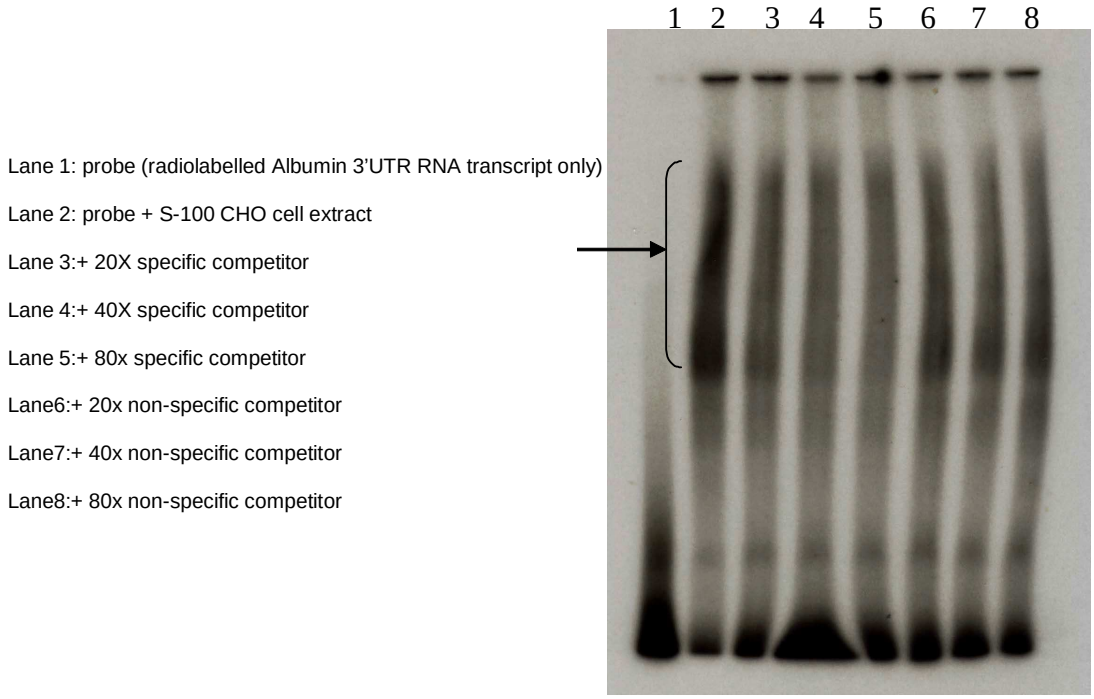


Figure 6.2 Specificity of protein binding to the Albumin 3'UTR. The intensity of protein binding and complex formation decreases as the molar ratio of specific competitor (non-labelled full length Albumin 3'UTR RNA transcript) increases (lane 3-5) whereas increasing the non-specific competitor (non-labelled β-globin RNA transcript) had no effect on the complex formation (lane 6-8). The arrow indicates complex formation decreases as the molar ratio of specific competitor increases.

This indicates that there was greater competition between the radiolabelled Albumin 3'UTR and its self-competitor for binding to the protein/s as the molar ratio of self competitor (non-labelled full length Albumin 3'UTR) was increased. As shown in Figure 6.2, competition for binding to the protein/s reaches its highest when an 80 molar excess of the specific competitor was applied. In contrast, increasing amounts of a non-specific competitor had less effect on complex formation. The data presented in Figure 6.2 strongly indicate that binding of CHO cell proteins to the radiolabelled Albumin 3'UTR RNA transcript is specific.

6.2.3 Competition analysis with the Albumin 3'UTR deletion regions: EMSA

Having established CHO cell protein binding to the Albumin 3'UTR and its specificity, a competition experiment was performed with unlabelled truncated transcripts of the Albumin 3'UTR in which specific regions were deleted. The aim was to identify region/s within the Albumin 3'UTR that were responsible for protein binding. The Albumin 3'UTR of deletion constructs (see chapter 5) were used as competitors in this experiment. In addition the full length Albumin 3'UTR RNA transcript was used as control. Based on the previous results (Figure 6.2) a 20 fold molar excess ratio of all competitors was used.

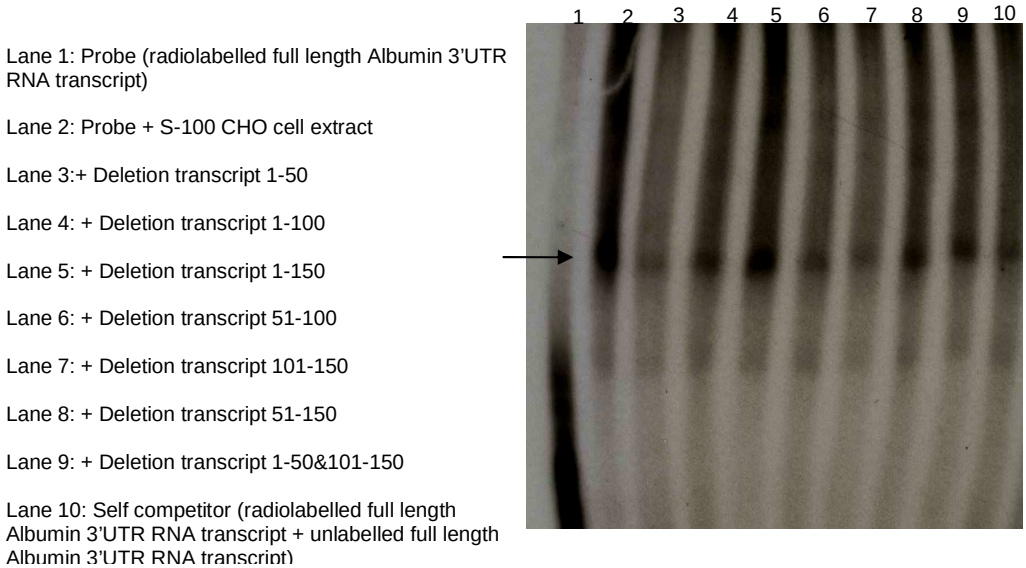


Figure 6.3 Competition assay with the Albumin 3'UTR deletion constructs as competitors. The Albumin 3'UTR deletion constructs were used in an EMSA experiment to identity deletion construct/s which compete for binding to protein extract with the radiolabelled full length Albumin 3'UTR RNA transcript. The strongest competition appeared to be with the deletion regions 1-50 and 101-150 (lanes 3 and 7). The arrow indicates the protein-RNA complex formation.

As shown in Figure 6.3 little or no competition was observed by deletion region 1-150. This indicates that all the possible binding motifs for the candidate protein/s were located within the first 150 nucleotides. It would also appear from Figure 6.3 that the strongest competition was provided by the deletion regions 1-50 and 101-150. This indicates firstly that protein binding to the Albumin 3'UTR occurs primarily outside the first 50 nucleotides. Then secondly, any candidate protein/s potentially bind to nucleotides 51-100 within the Albumin 3'UTR. The data also showed a lesser degree of competition with deletion regions 1-100, 51-100, 51-150 and 1-50 & 101-150. This could suggest that candidate protein/s may bind to other region/s within the Albumin 3'UTR.

To quantify the intensity of complex formation in Figure 6.3, densitometry analysis of autoradiograph image of an EMSA competition assay was carried out. Autoradiography image of EMSA competition experiment was analysed by software (UV MW band). The intensity of distinct bands was used to assess competition for binding to candidate protein/s between the Albumin 3'UTR and the deleted regions. The intensity of the bands was expressed as a percentage of the band intensity for the control (radiolabelled full length Albumin 3'UTR RNA transcript).

The procedure was performed independently for complex observed in the EMSA competition assay (Figure 6.3). The data obtained from the quantification of the complex formation in the EMSA assay was in agreement with the visual inspection of the autoradiography image. As shown in Figure 6.3, the strongest competition for binding to protein/s was seen with the deletion regions 1-50 and 101-150 competitors. Interestingly, densitometry analysis showed yet again that there is a varying degree of competition for the binding to candidate protein/s amongst other competitors (deletion regions 1-100, 51-100, 51-150 and 1-50&101-150)

indicating that possibly there are multiple binding site/s within the Albumin 3'UTR for candidate protein/s.

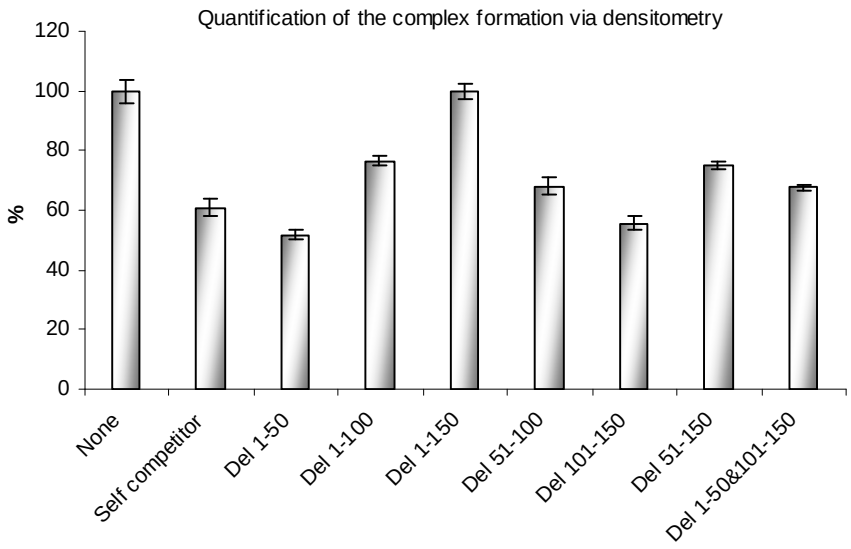


Figure 6.4 Quantification of the complex formation in the EMSA competition assay via densitometry. The intensity of bands was quantified and expressed as a percentage of control (radiolabelled full length Albumin 3'UTR RNA transcript) band intensity. The error bars represent the standard error of the mean.

In agreement with the visual inspection of the EMSA competition assay, the data presented in Figure 6.4 also showed that little or no competition was observed with the deletion region 1-150. This potentially indicates that the major binding sites for the protein/s are placed within the first 150 nucleotides of the Albumin 3'UTR. The quantification of complex formation was in overall agreement with the visual inspection of the data.

6.2.4 Protein binding to the Albumin 3'UTR: UV Cross-linking

6.2.4.1 Optimisation of UV Cross-linking reaction

To further investigate the interaction between protein and the Albumin 3'UTR transcripts, UV Cross-linking experiments were performed. In a typical UV Cross-linking experiment, an RNA transcript is cross-linked to proteins and then separated as distinct bands and visualised on a denaturing SDS-PAGE gel. The experiments are carried out by cross linking the RNA transcript and proteins (see 2.17.7) prior to RNaseA treatment and then separating materials on SDS-PAGE gel. The UV Cross-linking analysis separated independent protein components present in the cross-linking of RNA transcript and proteins and by using a molecular marker, the approximate molecular weight of the cross linked proteins and nucleic acid was determined.

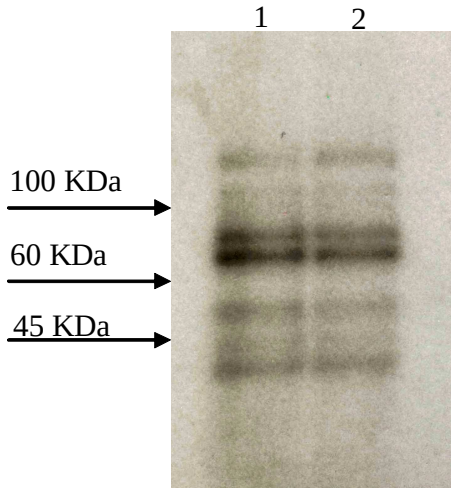


Figure 6.5 UV Cross-linking experiment to reveal protein binding to the Albumin 3'UTR. The RNA-protein interaction is shown as distinct bands (lane 1 and 2) which are separated based on molecular weight. Lanes 1 and 2 show the cross-linking of radiolabelled full length Albumin 3'UTR RNA transcript with 2.5 µg and 2 µg of S-100 CHO cell extract, respectively.

As shown in Figure 6.5, UV Cross-linking followed by SDS-PAGE analysis showed presence of several distinct bands (5) ranging from ~45 kDa to 100 kDa. It appeared from the data that when 2.5 µg of S-100 CHO cell extract was used with the radiolabelled full length Albumin 3'UTR RNA transcript, stronger bands were observed. In this study hereafter 2.5 µg of S-100 CHO cell extract was used for UV Cross-linking experiment.

6.2.4.2 Specificity of protein binding to the Albumin 3'UTR: UV Cross-linking

To confirm the results of EMSA for specificity of protein binding to the Albumin 3'UTR, UV Cross-linking experiments were performed where non-labelled full Albumin 3'UTR and β-globin 3'UTR were used as specific and non-specific competitors at 20, 40 and 80 fold molar excess.

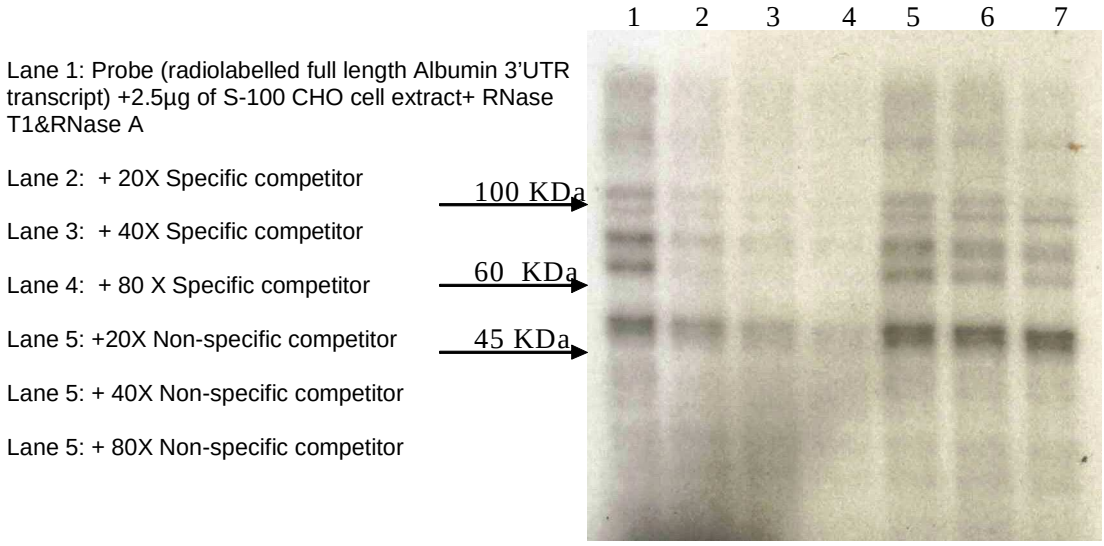


Figure 6.6 Examining the specificity of protein binding to the Albumin 3'UTR. The intensity of distinct bands representing the interaction between the CHO cell extract and the Albumin 3'UTR decreases as the molar ratio of specific competitor (non-labelled full length Albumin 3'UTR RNA transcript) increases (lanes 2-4) whereas increasing the molar ratio of non-specific competitor (non-labelled β-globin RNA transcript) had little or no significant effect on the intensity of complex formation and bands (lanes 5-7). The arrow indicates approximate molecular weight of complexes between the protein and labelled RNA transcript.

As shown in Figure 6.6, 5 prominent bands were seen cross-linked to the radiolabelled RNA transcript. The data presented in Figure 6.6 also showed that protein binding to the Albumin 3'UTR was specific as it had been seen previously in this chapter with EMSA. The data also showed that by increasing the molar ratio of specific competitor (non-labelled full length Albumin 3'UTR RNA transcript) from 20 to 80 fold excess, the intensity of bands with approximate molecular weights ranging from 45 to 100 kDa was reduced and at 80 molar excess all but disappeared. This result mirrored the result of EMSA for specificity of protein binding (Figure 6.2), confirming that interaction between the proteins and RNA transcript appeared specific.

6.2.4.3 Competition analysis with the Albumin 3'UTR deletion regions: UV Cross-linking

In order to further investigate protein binding to the various regions of the Albumin 3'UTR, UV Cross-linking competition experiments were carried out using unlabelled full length Albumin 3'UTR RNA transcripts with deletion transcripts as competitors. In addition an unlabelled full length Albumin 3'UTR was used as a positive control.

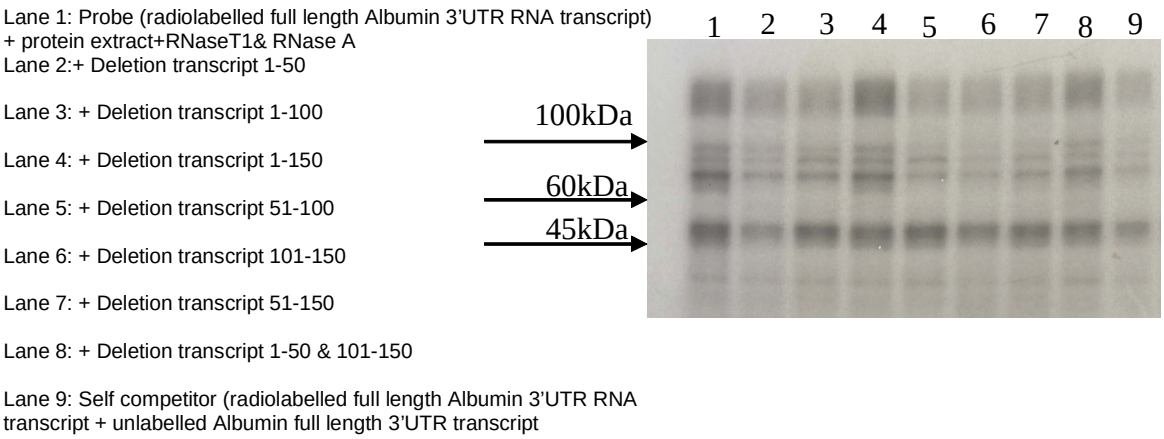


Figure 6.7 UV Cross-linking competition assay with the Albumin 3'UTR deletion constructs as specific competitors. The Albumin 3'UTR deletion constructs were used in a UV Cross-linking experiment to identity deletion construct/s which competes for binding to protein extract with the radiolabelled full length Albumin 3'UTR RNA transcript (probe).

As shown in Figure 6.7, the lowest competition amongst all the competitors was observed with the deletion region 1-150 as it was seen with the EMSA competition assay (see Figure 6.3). This suggests that binding motif/s for candidate protein with a molecular weight of ~50, are probably located within the first 150 nucleotides of the Albumin 3'UTR. The data presented in Figure 6.7 also showed that the strongest competition for binding to the protein with a molecular weight of ~50 kDa was provided by the deletion construct 1-50. This was in agreement with the previous finding in this chapter (see 6.2.3) although the EMSA competition experiment gave no information on the precise proteins involved in the interaction with the RNA transcripts. This yet again could suggest that the region which contains nucleotides 1-50 is not required for binding to any candidate protein/s.

The data presented in Figure 6.7 also showed that deletion region 101-150 provided the second strongest competition for binding to protein with a molecular weight of ~50 kDa. This could imply that binding motif/s for this candidate protein is probably located in region which contains nucleotides 51-100. This was based on observation that firstly, the deletion region 1-

150 showed little or no competition for binding to candidate protein/s. Secondly, the deletion region 1-50 showed the strongest competition. Figure 6.7 also showed that deletion regions 101-150 and 51-150 provided the strongest competition for binding to protein with the molecular weight of ~70 kDa. This could also indicate the existence of more than one possible binding site (motif) for the candidate protein/s within the Albumin 3'UTR.

The results of the EMSA and UV Cross-linking competition experiments suggested a varying degree of competition for binding to protein/s by various competitors with respect to the positive control (unlabelled full Albumin 3'UTR). This would suggest a preference of protein/s to bind to certain region/s of the Albumin 3'UTR. To gather a more precise measure of competition, densitometry analysis of autoradiograph image of two independent UV Cross-linking assays was carried out. The intensity of signal for distinct bands over two experiments was used to assess competition for binding to protein/s amongst the Albumin 3'UTR RNA transcripts. Autoradiography image of EMSA competition experiments were analysed by software (UV MW band) to quantify the intensity of bands. The intensity of the bands was expressed as a percentage of the appropriate band for the control (radiolabelled full length Albumin 3'UTR RNA transcript). To perform densitometry analysis, signals for the bands with molecular weight of ~50 kDa in Figure 6.7 were quantified and then averaged and expressed as a percentage of the control's band intensity (radiolabelled full length Albumin 3'UTR RNA transcript).

Interestingly, densitometry analysis data for UV Cross-linking experiments (Figure 6.8) showed a similar pattern to the densitometry analysis of the EMSA competition assay. As shown in Figure 6.8 the strongest competition for binding to protein/s was observed with the deletion region 1-50 unlabelled RNA transcripts and the competition was statistically

significant. The data also showed that the second strongest competition was observed with the deletion region 101-150. This competition of deletion region 101-150 was also found to be significant. The data also showed that competition of deletion regions 51-100 and 1-50& 101-150 for binding to protein/s was also significant. Interestingly, deletion regions 1-100, 1-150 and 51-100 showed no competition for binding to protein/s.

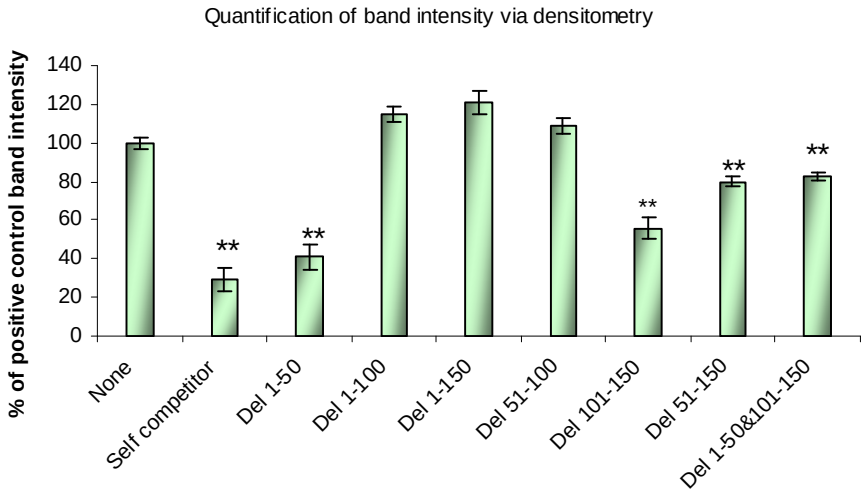


Figure 6.8 Quantification of band intensity by densitometry analysis. The intensity of bands obtained from two independent UV Cross-linking competition experiments for binding to protein/s with a molecular weight of ~50 kDa was quantified and expressed as percentage of the control (radiolabelled full length Albumin 3'UTR RNA transcript) band intensity. Statistical significance relative to the control was calculated by using the Mann-Whitney U test where **=p≤0.01 and where n=6. The error bars represent the standard error of the mean.

In addition, densitometry analysis was performed for bands immediately above the 60 kDa molecular weight observed in Figure 6.7. To perform densitometry analysis, signals obtained from a single experiment were quantified and then averaged and expressed as a percentage of the control's band intensity (the full length Albumin 3'UTR radiolabelled transcript). In agreement with previous findings in this chapter, unlabelled 1-150 transcript showed no

competition (Figure 6.9). This could suggest that the majority of the binding sites for candidate proteins are located within the first 150 nucleotides of the Albumin 3'UTR. In a contrast to a previous finding in this chapter, the strongest competition was observed with the unlabelled 101-150 RNA transcript not the unlabelled 1-50 RNA transcript. Unlabelled 51-100 and 51-150 RNA transcripts also showed strong competition indicating that various regions of the Albumin 3'UTR show varying degrees of competition for binding to perhaps a variety of protein/s. Interestingly, unlabelled 1-100 and 1-50&101-150 showed the lowest competition for binding to protein/s with molecular weight of ~70 kDa.

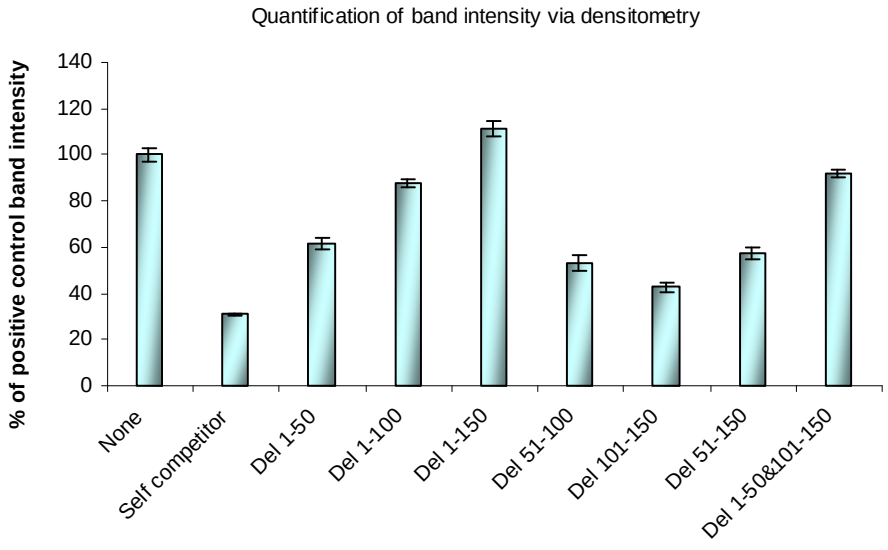


Figure 6.9 Quantification of band intensity by densitometry analysis. The intensity of bands for a protein/s with molecular weight of ~70 kDa was obtained from a typical UV Cross-linking competition experiment. The intensity of signals was then quantified and expressed as a percentage of the control's band intensity (radiolabelled full length Albumin 3'UTR RNA transcript). The error bars represent the standard error of the mean.

6.2.5 Streptavidin MagneSphere Paramagnetic Particles (SA-PMPs) and protein separation

RNA affinity techniques followed by mass spectrometry have been widely used to isolate and identify RNA-binding proteins. Streptavidin MagneSphere Paramagnetic Particles (SA-PMPs) were used in this study to isolate bound proteins utilising the strong binding affinity between biotin and Streptavidin which has long been known to be a very strong and stable interaction [195].

EMSA, UV Cross-linking competition assays showed that unlabelled 1-50 deletion transcripts gave the strongest competition for binding to protein/s when they were used as a specific competitor against radiolabelled full length Albumin 3'UTR RNA transcript. Therefore, it was decided to select deletion construct 1-50 for the "bait" with which to isolate and identify protein/s that can bind to the Albumin 3'UTR RNA. The unlabelled 1-50 and globin 3'UTR (as a negative control) transcripts were labelled with biotin (see 2.17.8) and then immobilised on Streptavidin MagneSphere Paramagnetic Particles (SA-PMPs). The biotinylated control transcript was used in order to distinguish the specificity of protein binding to the Albumin 3'UTR RNA transcript.

The magnetic beads were then incubated with S-100 CHO cell extract, washed to remove unbound particles and after separation by 10% SDS-PAGE, bound protein/s were then visualised by staining the gel with Colloidal Coomassie. Staining of the SDS-PAGE gel revealed three distinct bands each of a molecular weight of under ~60 kDa. They were only present in the unlabelled 1-50 biotinylated transcript (Figure 6.10). The focus of protein identification in this study was only on specific protein/s that bind to the Albumin 3'UTR, therefore only the region with the triplet bands in unlabelled 1-50 transcript and its immediate

corresponding region in control transcript were subjected to mass spectrometry analysis. The actual process of protein identification and excision of bands was performed by a private company (NEPAF, Newcastle University).

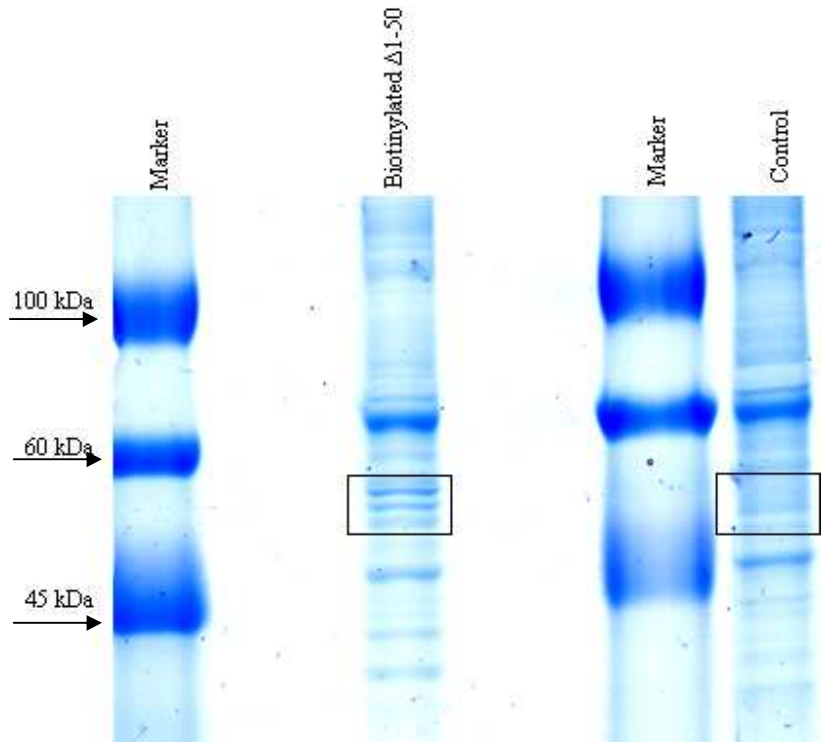


Figure 6.10 Isolation of protein/s attached to biotinylated 1-50 RNA transcript. Attached protein/s to the biotinylated 1-50 and control (biotinylated globin 3'UTR) transcript were subjected to 10% SDS-PAGE gel separation. The regions selected in small rectangular boxes were subjected to mass spectrometry analysis.

Following mass spectrometry analysis, three proteins were identified that were only present in the sample containing the deletion region 1-50 biotinylated RNA transcript. The identified proteins were absent in sample containing the negative control. The identified proteins were as follows:

1) Far upstream element (FUSE) binding protein 3
2) Regulator of differentiation 1 (Rod 1)
3) CUG-BP1

1) **Far upstream element (FUSE) binding protein 3:** This protein is a member of an ancient family of single stranded DNA binding proteins which are required for proper regulation of *c-myc* proto-oncogene and may activate gene expression. It has been shown that this family of proteins bind to a variety of RNAs and therefore is likely to be multifunctional [196].

Protein id	Mass	log (e)	pI	rI	log (I)
ENSRNOP00000021871	61.9	-17.1	8.4	3	6.10

protein sequence:

Matched peptides are highlighted and underlined:

1 MAELVQGQSAPVGLKAEDFVDALHRVRQIAAKIDSIPHLNNSTPLVDPSVYGYGVQKRPL 60

61 DDGVGNQLGALVHQRAVITEEFKVPDK**MVGFIIGR**GGEQISRIQAESGCKIQIASESSGI 120

121 PERPCVLTGTPESIEQAKRLLGQIVDRCRNGPGFHNDIDGNSTIQELLIPASKVGLVIGK 180

181 GGETIKQLQERTGVKMVMIQDGPLPTGADKPLRITGDPFKVQQAR**EMVLEIIR**EKDQADF 240

241 RGVRSDFTSR**AGGGSIEVSVPR**FVVGIVIGRNGEMIKKIQNDAGVRIQFKPDDGISPERA 300

301 AQVMGPPDRCQHAARIINELILTAQEREILGGLTGTRGRGRGRGDWSVGTPGGIQEITYT 360

361 VPADKCGLVIGKGGENIKSINQQSGAHVELQRNPPPNTDPNLRIFTIRGAPQQIEVARHL 420

421 IDEKVGGASLGAPTAFGQSPFSQPPAAPHQNTFPPRAFPNIAAKVNGNPHSTPVSGPPAF 480

481 LTQGWGSTYQAWQQPTQQVPSQQSQPQNSQPDYSKAWEDYYKKQGHTTSAAPQASSPPDY 540

541 TMAWAEYYRQQAAFYGQTLGQAQAHSQEQ 569

Figure 6.11 Mass Spectrometry analysis result for far upstream element (FUSE) protein 3. The data shows the identified protein's specification and sequence. Log (e) represents the expectation of finding protein stochastically, pI represents the isoelectric value for the intact gene product, rI represents the number of found peptides after digestion with trypsin and log (I) represents the sum of raw spectrum intensities. Mass Spectrometry analysis carried out by comparing the sequence of peptide fragments with the following 5 protein sequence databases : cavPOR 3 ENSEMBLE 53, cRAP with Sigma-Aldrich Universal Protein Standard Sequences, monDom 5 ENSEMBLE 50.5e, RGSC 3,4 ENSEMBLE 52, NCBIM37 ENSEMBLE 52.

2) Regulator of differentiation 1 (Rod 1): an RNA binding protein which suppresses differentiation and binds to RNA preferentially to both poly (U) and Poly (G) sequences[197].

NCBI id	Mass	log (e)	pI	rI	log (I)
NP_005147.3	55.7	-16.0	9.2	12	7.20

Partial protein sequence:

Matched peptides are highlighted and underlined:

1 ANGNDSKKFKGDRPPCSPSRVLHLRKIPCDVTEAEVISLGLPFGK**VTNLLMLK**GKSQAFL 60

61 EMASEEAAVTMVNYYTPVTPHLRSQPVYIQYSNHRELKTDNLPNQARAQAALQAVSAIQS 120

121 GNLTLHGAPSNEVTILPGQSPVLRIIIENLFYPVTLEVLHQIFSKFGTVLKIITFTKNNQ 180

181 FQALLQYADPVNAHYAKMALDGQNIYNACCTLRIDFSKLTSLNVKYNNDKSRDFTRLDLP 240

241 SGDGQPSLEPPMAAAFGAPGIISSPYAGAAGFAPAIGFPQATGLSVPAVPGALGPLALTS 300

301 SAITGRMAIPGASGIPGNSVLLVTNLNPDLITPHGLFILFGVYGDVHRVKIMFNKKENAL 360

361 VQMADANQAQLAMNHLSGQRLYGKVLRATLSKHQTVQLPR**EGQEDQGLTKDFSNSPLHR**F 420

421 KKPGSKNFQNIFPPSATLHLSNIPPSVTMDDLKNLFTEAGCSVKAFKFFQKDRKMALIQL 480

481 GSVEEAIQALIELHNHDLGENHHLRVSFSKSTI 513

Figure 6.12 Mass Spectrometry analysis result for regulator of differentiation 1 (Rod 1). The data shows the identified protein's specification and sequence. Log (e) represents the expectation of finding protein stochastically, pI represents the isoelectric value for the intact gene product, rI represents the number of found peptides after digestion with trypsin and log(I) represents the sum of raw spectrum intensities. Mass Spectrometry analysis carried out by comparing the sequence of peptide fragments with the following 5 protein sequence databases : cavPOR 3 ENSEMBLE 53, cRAP with Sigma-Aldrich Universal Protein Standard Sequences, monDom 5 ENSEMBLE 50.5e, RGSC 3,4 ENSEMBLE 52, NCBIM37 ENSEMBLE 52.

3) CUG-BP1: CUG-BP1, is the founder member of CELF/Bruno-like family of RNA binding proteins that has been reported to be involved in myotonic dystrophy type 1 (DM1) [198]. The known members of CELF/Bruno-like family of RNA binding proteins are extremely comparable in their structural organization. They contain 3 RNA recognition motifs (RRM), one in the C-terminal region and two in the N-terminal region with high degrees of conservation. Interestingly for all the members, the sequence content of the linker region between the RRM 2 and 3 has been less conserved [199]. It has also been reported that the primary transcripts of several members of the CELF family are subject to alternative splicing with 4 known protein isoforms for human CUG-BP1 and 7 protein isoforms for CUG-BP2 [200, 201]. Following a computational analysis in the 3'UTRs of many short-lived transcripts, it was found that these transcripts share an 11-mer nucleotide region (UGUUUGUUUGU) termed the GU-rich element (GRE). It was also found that CUG-BP1 binds specifically to the GRE [202].

Due to the similarity of this 11-mer nucleotide sequence with a region within the Albumin 3'UTR (UGUUUUCUUUU), which was removed in deletion constructs 51-100 and 51-150 (nucleotides 75-85) and also the considerable evidence for the binding of CUB-BP1 to CUG and UG dinucleotide repeats that function as a regulator of translation and stability of mRNA transcripts, it became of interest to further study the CUG-BP1 protein and its possible binding to the Albumin 3'UTR.

NCBI id	Mass	log (e)	pI	rI	log (I)
NP_941989.1	55.1	-8.10	8.6	4	6.47

Partial protein sequence:

Matched peptides are highlighted and underlined:

1 MAAFKLDFLPEMMVEHCSLNSSPVSKKMNGTLDHPDQPDLDAIK**MFVGQVPR**TWSEKDLR 60

61 ELFEQYGAVYEINVLRDRSQNPPQSKGCCFVTFYTRKAALEAQNALHNMKVLPGMHHPIQ 120

121 MKPADSEKNNAVEDRKLFIGMISKKCTENDIR**VMFSSFGQIEECR**ILRGPDGLSRGCAFV 180

181 TFTTRAMAQTAIKAMHQAQTMEGCSSPMVVKFADTQKDKEQKRMAQQLQQQMQQISAASV 240

241 WGNLAGLNTLGPQYLALYLQLLQQTASSGNLNTLSSLHPMGGLNAMQLQNLAALAAAASA 300

301 AQNTPSGTNALTTSSSPLSVLTSSGSSPSSSSSNSVNPIASLGALQTLAGATAGLNVSSL 360

420 AGMAALNGGLGSSGLSNGTGSTMEALTQAYSGIQQYAAAALPTLYNQNLLTQQSIGAAGS 480

481 QKEGPEGANLFIYHLPQEFGDQDLLQMFMPFGNVVSAKVFIDKQTNLSKCFGFVSYDNPV 540

541 SAQAAIQSMNGFQIGMKRLKVQLKRSKNDSKPY 513

Figure 6.13 Mass Spectrometry analysis result for CUG-BP1. The data shows the identified protein's specification and sequence. Log (e) represents the expectation of finding protein stochastically, pI represents the isoelectric value for the intact gene product, rI represents the number of found peptides after digestion with trypsin and log (I) represents the sum of raw spectrum intensities. Mass Spectrometry analysis carried out by comparing the sequence of peptide fragments with the following 5 protein sequence databases : cavPOR 3 ENSEMBLE 53, cRAP with Sigma-Aldrich Universal Protein Standard Sequences, monDom 5 ENSEMBLE 50.5e, RGSC 3,4 ENSEMBLE 52, NCBIM37 ENSEMBLE 52.

6.2.6 Confirmation of the Albumin 3'UTR and CUG-BP1 interaction

In attempts to confirm the results of mass spectrometry analysis and the presence of CUG-BP1 in EMSA RNA-protein complexes, 2 experiments were carried out; a supershift and EMSA assay with extracts after knock down of CUG-BP1 by siRNA. The supershift experiment was carried out where Rabbit Polyclonal CUG-BP1 antibody (Abcam) was used in a typical EMSA reaction. Theoretically the binding of specific antibody to the complex would further increase the molecular weight of the complex so resulting in a greater shift. The supershift assay was performed as described in chapter 2 but with 0.5,1 and 2 µg of

CUG-BP1 antibody incubated with CHO cell extract and RNA transcript overnight (or for 48 hours) at 4˚C.

It would appear from Figure 6.14 that the overnight incubation of CUG-BP1 antibody with the radiolabelled RNA transcript + CHO cell extract led to a slight but crucially not significant shift (lanes 3-5) in mobility of the complex as the volume of CUG-BP1 increased. It would also appear from Figure 6.14 that longer incubation of the complex with specific antibody (lanes 6-8) did not create any further shift in formed complex. In addition, further repeats and alteration of experiment's conditions did not induce a significant shift (data not shown).

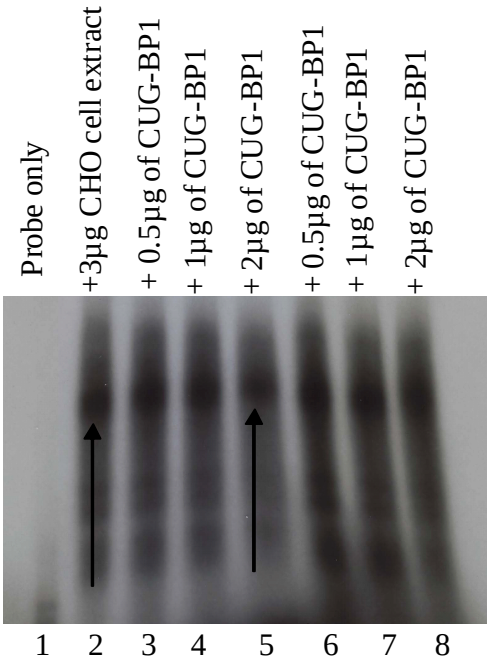


Figure 6.14 Supershift assay with CUG-BP1. Autoradiography image showing the supershift assay result. A standard EMSA reaction was performed by using radiolabelled Δ1-50 RNA transcript and S-100 CHO cell extracts incubated overnight with CUG-BP1 antibody (lane 3-5) or 48 hours (lane 6-8). Lane 1 represents the RNA transcript only and lane 2 represents the interaction formed between the RNA transcript and S-100 CHO cell extract.

The result of supershift assay did not appear to be conclusive and therefore an alternative approach was used in which CUG-BP1 was knocked down by siRNA and extracts were used in an EMSA reaction to assess whether complex formation with the Albumin 3'UTR was impaired. Since the genome of the Chinese Hamster has not yet been fully sequenced and the protein sequence of CUG-BP1 in the Chinese Hamster was not known, it was decided to design the siRNA based on CUG-BP1 mRNA sequence in *Rattus norvegicus* (Accession number:NM_001025421) by using Blok-iT^{TM}RNAi Designer software (Invitrogen) and then perform a blast search for the sequence alignment with *Mus Musculus* 2 isoforms of CUG-BP1(Accession number: NM_198683.1 & NM_017368.2), *Homo sapian* 3 isoforms of CUG-BP1 (Accession number: NM_006560, & NM_198700 & NM_001025596 and finally 2 variants of CUG-BP1 in *Canis familiaris* (Accession number:XM_533186 & NM_855451) to find sequence homology and the extent of conservation. Following identification of the most conserved sequence of the CUG-BP1 amongst all the species for known CUG-BP1 sequences, a specific siRNA (NM_001025421_stealth_1297 (RNA)-AAACCUUGGCAGACACGACAUUCCC), (NM_001025421_stealth_1297 (RNA)-GGGAAUGUCGUGUCUGCCAAGGUUU) was designed by the Blok-iT^{TM}RNAi Designer software (Invitrogen) and this was used in siRNA transfection with CHO cells.

To determine the extent of CUG-BP1 knock down, CHO cells were treated with specific CUG-BP1 siRNA and a negative universal control siRNA (Invitrogen). As shown in Figure 6.14, siRNA transfection of CHO cells successfully knocked down CUG-BP1 in the cell extracts collected 1-4 days after siRNA transfection. It appeared from the data that the siRNA transfection had knocked down the ~55 KDa CUG-BP1 in cell populations from day 1-4.

Recognition of CUG-BP1 by the specific antibody was seen as distinct bands which were only present in cell populations treated with the negative universal control siRNA.

The data also showed that treatment of cell populations with the negative universal control siRNA had no obvious negative effect on the CUG-BP1. It also appeared from Figure 6.15 that the CUG-BP1 antibody identifies 2 distinct bands which were only observed in cell extracts treated with the negative universal control siRNA. A plausible explanation for observing multiple bands is the possibility of the existence of different splice variants that share the same epitopes and could be from the same family of protein and therefore could be identified by the CUG-BP1 antibody. Furthermore, it is tempting to consider that the protein could have multiple modified forms *in vivo* as a result of events such as glycosylation, phosphorylation, methylation or acetylation. This could have also led to the observation of multiple bands seen in Figure 6.15.

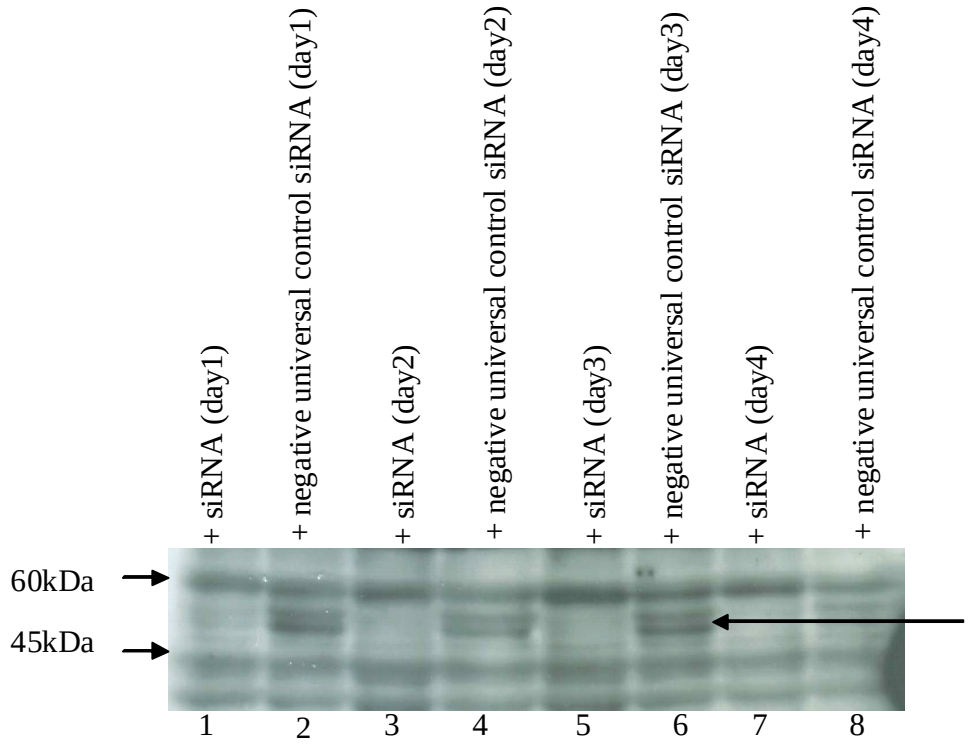


Figure 6.15 Western blot of CHO cells transfected with siRNA probed with CUG-BP1 antibody. CHO cells were transfected with siRNA for CUG-BP1 and S-100 cell extraction was performed on day 1-4 after transfection. siRNA negative universal control was applied to cells day 1-4 as a negative control. The arrows indicate bands which were only present in cell extracts treated with the negative universal control siRNA.

Following successful siRNA transfection and CUG-BP1 knock down, it was decided to perform the siRNA transfection on stably transfected CHO cells with construct GG*G*Alb 3Δβ. The aim of this experiment was firstly to visualise siRNA knock down on the stably transfected cell line. Then secondly, to investigate whether is possible to knock down CUG-BP1 and therefore impair interactions that form between the radiolabelled RNA transcript and the CHO cell extract.

To achieve this, stably transfected cells (GG*G*Alb 3Δβ) were transfected with siRNA and negative universal control siRNA. Transfection was followed by S-100 CHO cell extraction

from a day 3 cell population and western blotting was performed**.** The selection of day 3 samples was based on visual observation of band intensity and the manufacturer's recommendation.

As shown in Figure 6.16 the specific siRNA transfection successfully knocked down the CUG-BP1 whereas treatment of stably transfected cell line with the negative universal control siRNA had no obvious negative effect on the ~55 KDa CUG-BP1 recognition by CUG-BP1 antibody.

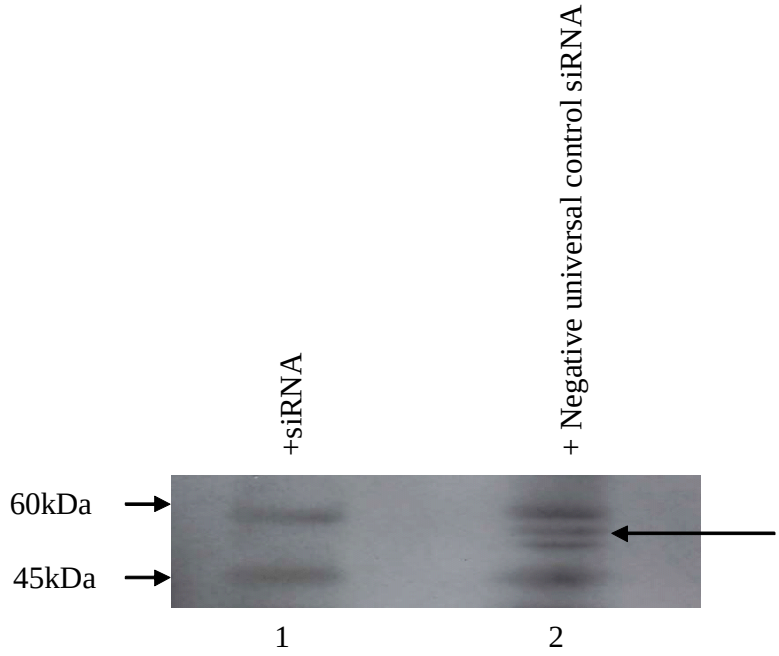


Figure 6.16 Knock-down of CUG-BP1 in cells stably transfected with GG*G*Alb3delB. S-100 cell extracts were prepared from cells transfected with specific siRNA and negative universal control siRNA. siRNA treated stably transfected cells line were probed with CUG-BP1 antibody. The arrow pointing at bands which were only present in cell extracts treated with the negative universal control siRNA but crucially not specific siRNA.

Interestingly, treatment of the stably transfected cells with specific CUG-BP1 siRNA led to a loss in complex formation between the radiolabelled full length Albumin 3'UTR RNA transcript and cell extracts (Figure 6.17), firstly confirming that CUG-BP1 knock down

affects protein binding to the Albumin 3'UTR and secondly suggesting that it is directly involved in complex formation of protein-Albumin 3'UTR.

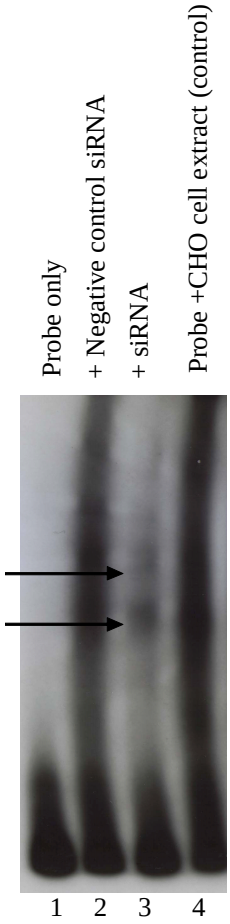


Figure 6.17 EMSA with CHO cell extract from stably transfected cell line (GG*G*Alb 3Δβ) treated with siRNA specific to CUG-BP1 and negative universal control siRNA (negative control). The image showed that siRNA impairs the complex formation between RNA transcript and CHO cell extract. The arrows show the regions where complex formation was affected. Lane 1 represents the RNA transcript (probe). Lane 2 shows the complex formation between the RNA transcript and CHO cell extract obtained from cell line treated with negative universal control siRNA. Lane 3 represents the complex formation between the RNA transcript and CHO cell extract treated with specific CUG-BP1 siRNA. Lane 4 represents the complex formation between the RNA transcript and CHO cell extract obtained from cell line without siRNA treatment (control).

6.2.7 siRNA knock down of CUG-BP1 and its effect on Luciferase activity

Following successful CUG-BP1 knock down, it was decided to investigate the negative/positive effects of CUG-BP1 down regulation on Luciferase production and mRNA abundance levels in cells transfected with GG*G*Alb 3Δβ and GG*G*Alb 3Δβ (Δ 51-100). As it was seen previously in this chapter, UV Cross-linking and EMSA competition assays as well as siRNA knock down, revealed evidence of CUG-BP1 binding to the Albumin 3'UTR. It was therefore hypothesised that CUG-BP1 knock down would influence Luciferase protein secretion and mRNA abundance in the stably transfected cells with construct GG*G*Alb 3Δβ but not construct GG*G*Alb 3Δβ (Δ51-100). CHO cell lines were transfected with the siRNA and negative universal control siRNA. Culture medium sample collection and total RNA extraction were performed on day 3 after siRNA transfection.

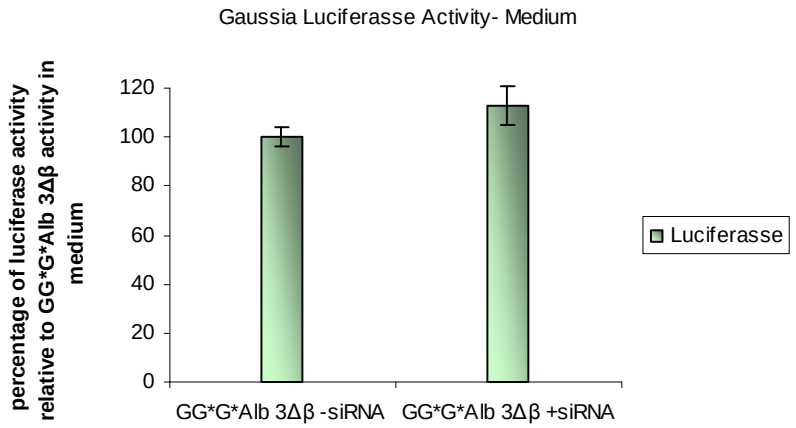


Figure 6.18 Luciferase activity in the culture medium samples shown as percentage of construct GG*G*Alb 3Δβ (treated with negative universal control siRNA) Luciferase activity. The data showed that measured secreted Luciferase did not increase significantly after treatment with siRNA. The error bars represent the standard error of the mean.

As shown in Figure 6.18 differences in Luciferase activity in cell populations treated with specific siRNA and negative universal control siRNA were not statistically significant. Interestingly, mRNA abundance level increased significantly after treatment with specific siRNA (Figure 6.19). This was compatible with findings by other groups that CUG-BP1 destabilises the mRNA targets [202], suggesting that CUG-BP1 knock down significantly influenced mRNA abundance level. The data presented here suggest that although siRNA treatment and down regulation of CUG-BP1 stabilises the mRNA targets and result in a significantly higher abundance of mRNA compared to the negative control siRNA, the higher mRNA level, however, did not correlate with increased protein synthesis as it can be seen in Figure 6.20 following the normalisation of *Gaussia* Luciferase activity to the mRNA level. The imbalance between the higher level of mRNA than protein in Figure 6.19 and 6.18 respectively, could be due to induction of the ER stress responses; the UPR and EOR (see

1.2.9 and 1.2.10). It is speculated for instance that specific siRNA treatment of stably transfected cells with construct GG*G*Alb 3Δβ leads to the activation of UPR as a result of aggregation of unfolded proteins in the ER. Therefore this could lead to PERK-mediated phosphorylation of eIF2α and subsequent repression of translation.

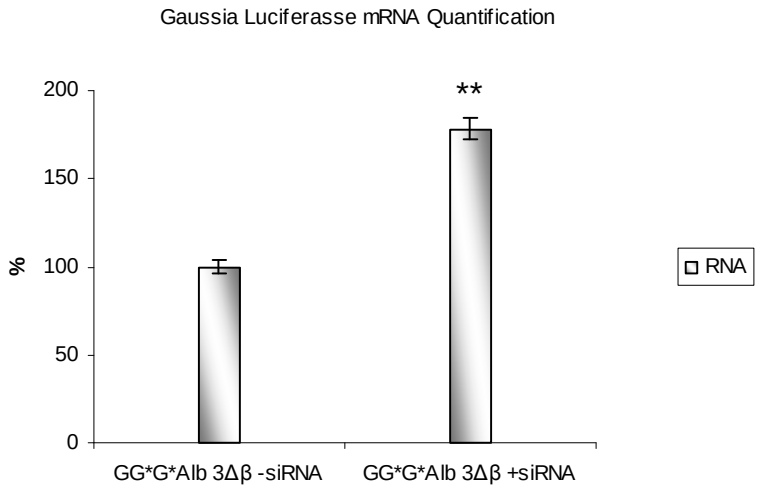


Figure 6.19 Relative quantification of mRNA abundance in stably transfected CHO cells with construct GG*G*Alb 3Δβ after treatment with specific siRNA and negative universal control siRNA. The quantification of mRNA is relative to the house keeping gene GAPDH. The data presented here showed that the level of mRNA expression was significantly higher after treatment with specific siRNA. Statistical significance relative to the construct GG*G*Alb 3Δβ was calculated by using the Mann-Whitney U test where **=p≤0.01 and where n=6. The error bars represent the standard error of the mean.

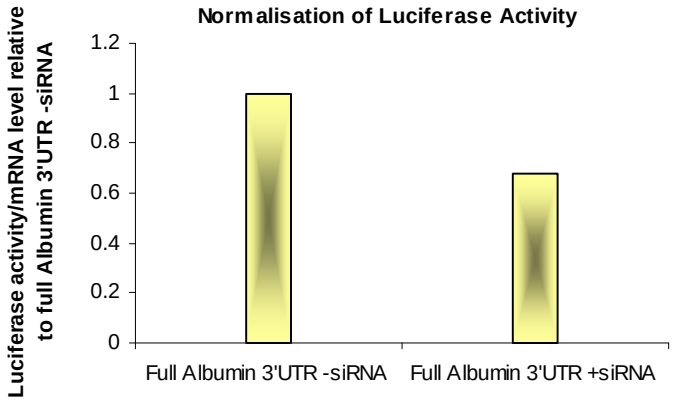


Figure 6.20 The normalisation of *Gaussia* Luciferase activity to mRNA abundance. The figure shows that reporter activity did not correlate with the mRNA abundance following the siRNA treatment.

In contrast to the construct with the full Albumin 3'UTR (GG*G*Alb 3Δβ) where siRNA treatment did not cause a significant increase in Luciferase activity, siRNA knock down and down regulation of CUG-BP1 in a stably transfected cell population with GG*G*Alb (Δ 51-100) significantly increased the activity of reporter protein (Figure 6.21) and mRNA abundance (Figure 6.22). Normalisation of *Gaussia* Luciferase activity to the mRNA level for cells expressing construct GG*G*Alb (Δ 51-100) clearly showed that reporter activity correlated with the mRNA abundance following treatment with the siRNA (see Figure 6.23).

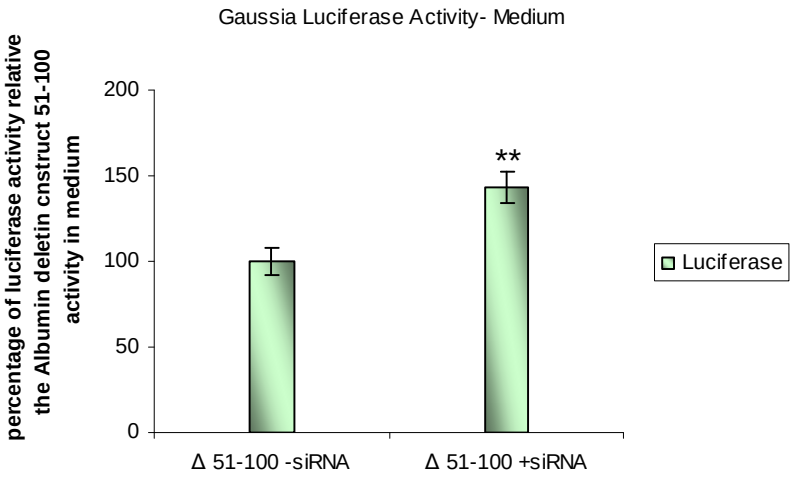


Figure 6.21 Luciferase activity in the culture medium samples shown as a percentage of construct GG*G*Alb (Δ51-100) 3Δβ (treated with negative universal control siRNA) Luciferase activity. The data showed that measured secreted Luciferase significantly increased after treatment of the cell line with specific siRNA. Statistical significance relative to the construct GG*G*Alb (Δ51-100) 3Δβ was calculated by using the Mann-Whitney U test where **=p≤0.01 and where n=6. The error bars represent the standard error of the mean.

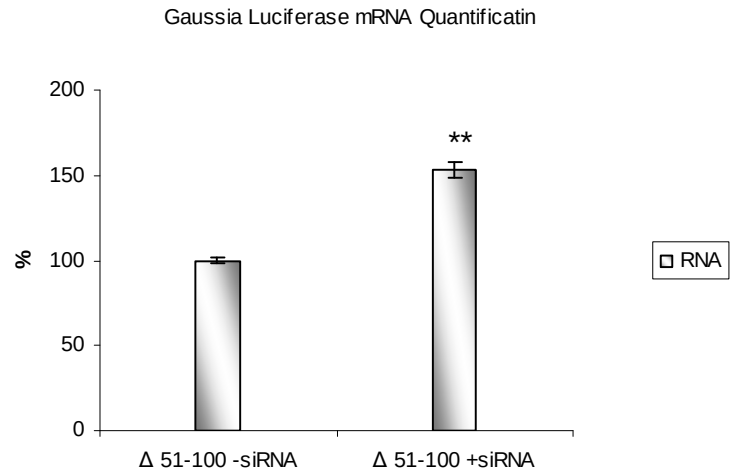


Figure 6.22 Relative quantification of mRNA abundance in stably transfected CHO cells with GG*G Alb (Δ51-100) 3Δβ after treatment with specific siRNA and negative universal control siRNA. The quantification of mRNA is relative to the house keeping gene GAPDH. The data presented here showed that the level of mRNA abundance was significantly higher after treatment with specific siRNA. Statistical significance relative to the construct GG*G*Alb (Δ51-100) 3Δβ was calculated by using the Mann-Whitney U test where **=p≤0.01 and where n=6. The error bars represent the standard error of the mean.

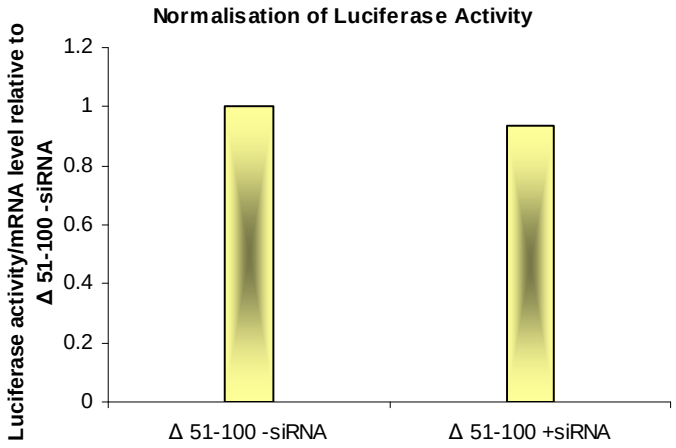


Figure 6.23 The normalisation of *Gaussia* Luciferase activity to mRNA abundance. The figure shows that Gaussia Luciferase activity correlated with the mRNA abundance following treatment with the siRNA. This was in contrast to cells expressing the full length Albumin 3'UTR.

6.3 Summary

Overall, in this chapter evidence of protein/s binding to the Albumin 3'UTR was presented. Experiments such as EMSA and UV Cross-linking were performed to identify the region/s to which protein/s would bind. Here it was shown that strongest competition for binding to protein/s was seen with unlabelled 1-50 deletion variant when it was used as a specific competitor against radiolabelled full length Albumin 3'UTR RNA transcript. Based on findings obtained from EMSA and UV Cross-linking competition experiments, to isolate protein/s that can bind to the Albumin 3'UTR, it was decided to select deletion construct 1-50 as "bait". Mass spectrometry analysis was used to identify bound protein/s to the Albumin 3'UTR. Three proteins were identified by mass spectrometry analysis. The identified proteins were as follows:

1) Far upstream element (FUSE) binding protein 3
2) Regulator of differentiation 1 (Rod 1)
3) CUG-BP1

To study CUG-BP1 further and confirm the role of CUG-BP1 in CHO cell protein-Albumin 3'UTR complex formation, siRNA was used to knock down CUG-BP1 expression. To design siRNA for Chinese Hamster CUG-BP1, Blok-iT^{TM}RNAi Designer software (Invitrogen) was used and specific siRNA for *Rattus norvegicus* CUG-BP1 was designed. Here it was shown that CUG-BP1 in CHO cells could be knocked down following siRNA transfection and subsequently the effects of CUG-BP1 knock down were investigated. Here it was shown that siRNA knock down of CUG-BP1 leads to severe loss of complex formation between CHO cell extracts and the radiolabelled full length Albumin 3'UTR RNA transcript, revealing

firstly that protein binding to the Albumin 3'UTR is affected by CUG-BP1 siRNA knock down and secondly it is directly involved in complex formation of protein-Albumin 3'UTR.

7. Discussion

Prior to this study, a series of gene constructs containing the *Gaussia princeps* Luciferase as reporter were created by utilising a seamless cloning method where the Chymotrypsinogen signal peptide was used in combination with either the native *Gaussia* Luciferase 3'UTR or the Albumin 3'UTR. In the present work these constructs were used, and in some cases further modified to remove mutations introduced during the earlier cloning, as a basis for further changes in the expression vector by which the signal peptide, 3'UTR or both were altered. This led to constructs in which the Albumin 3'UTR was corrected for 5 point mutations and in which the Chymotrypsinogen signal peptide was replaced with the Human Albumin signal peptide. The aims were to express these constructs in CHO cells and to investigate the extent to which modifications to the signal peptide/3'UTR sequences affect the production of recombinant reporter protein and the underlying mechanisms.

Using the generated constructs it was shown that increasing the hydrophobicity of the Chymotrypsinogen signal peptide had little or no detectable influence on protein expression when it was used in conjunction with the Albumin 3'UTR (without vector-derived sequences). Since hydrophobicity is recognised as an important factor regarding the operational efficiencies of signal peptides [20], one plausible explanation for this observation is that the Chymotrypsinogen signal peptide is already sufficiently hydrophobic and therefore increasing its total hydrophobicity even further had little or no detectable influence on the level of reporter produced.

The data presented in chapter 4 also showed that the highest level of reporter activity and mRNA abundance is achieved when the native or mutated Chymotrypsinogen signal peptide is used in combination with the Albumin 3'UTR. On the other hand, replacing the

Chymotrypsinogen signal peptides with the Albumin signal peptide significantly reduced the level of expression of the reporter protein. There could be multiple reasons for the lower *Gaussia* Luciferase activity in constructs containing the Albumin signal peptide. The most likely reason could be the lower level of Albumin signal peptide hydrophobicity compared to the native *Gaussia* or Chymotrypsinogen signal peptide (see chapter 4). This could lead to inefficient recognition by SRP or the mis-targeting of mRNA transcripts.

If time had allowed, it would have been interesting to investigate whether increasing the hydrophobicity of the Albumin signal peptide improves protein expression. There are four amino acids within the H-domain of the Albumin signal peptide that can be replaced with more hydrophobic amino acids. For instance, less hydrophobic amino acids threonine (1 residue) and serine (3 residues) could be replaced sequentially with more hydrophobic amino acids such as alanine and phenylalanine) using site-directed mutagenesis. This could produce several constructs with different levels of hydrophobicity for the Albumin signal peptide. Subsequently, the reporter activity using constructs containing the Albumin signal peptide varied with respect to signal peptide hydrophobicity, could be tested and any increase/decrease in protein production could be correlated with altered signal peptide hydrophobicity.

It is tempting to speculate that by gradually increasing the hydrophobicity of the Albumin signal peptide, protein production should be positively influenced. However, there will possibly be a point where increasing the hydrophobicity would have little or no further detectable effect on protein production, as was observed in this study for constructs containing the mutated Chymotrypsinogen signal peptide and the *Gaussia* Luciferase/Albumin 3'UTR.

The results shown in chapter 4 indicated that the deletion of vector-derived sequences significantly improved *Gaussia* Luciferase activity and mRNA abundance in construct GG*G*G 3Δβ compared to that of GG*G*G. This could have been due to the alteration of, or conformation changes in the secondary structure of the mRNA transcript following the deletion of vector-derived sequences (the partial rabbit β- globin coding region comprising of 2 exons, one intron and the full rabbit β-globin 3'UTR), therefore leading to a higher stability of transcript. In addition, it is speculated that the theoretical existence of 2 different 3'UTRs (the rabbit β-globin 3'UTR and either the *Gaussia* or Albumin 3'UTR) at the same time in a given transcript might cause an unknown interaction between the *Gaussia* Luciferase 5'UTR and the β-globin 3'UTR, as well as the *Gaussia* Luciferase 5'UTR and the *Gaussia* Luciferase/Albumin 3'UTR. This could lead to an interaction that reduces the translatability of the mRNA transcript generated by a construct containing the vector-derived sequences.

The data presented in this study also suggested that differences in secreted Luciferase activity and mRNA expression between constructs containing various combinations of signal peptide/3'UTR were not due to differences in the site of integration. The pattern of Luciferase activity obtained from cells transiently transfected with the various constructs, correlated with the pattern of secreted *Gaussia* Luciferase activity in stably transfected cells. It was observed from the data obtained from the transiently transfected constructs that the highest Luciferase activity was seen in constructs containing the Chymotrypsinogen signal peptide and Albumin 3'UTR (without vector-derived sequences). This was in agreement with the data obtained from stably transfected cells. It was also observed that deletion of vector-derived sequences (GG*G*G 3Δβ) improved protein secretion. This too was in agreement with the data obtained from the stably transfected cells. Therefore, the differences observed

in various constructs are unlikely to have been due to differences in the site of transgene integration.

In the present work it was shown that a construct containing the Albumin 3'UTR (without vector-derived sequences) in combination with a sufficiently hydrophobic signal peptide (Chymotrypsinogen signal peptide), promotes higher levels of Luciferase activity compared to the native *Gaussia* Luciferase 3'UTR (containing vector-derived sequences).

To understand the underlying mechanism by which the Albumin 3'UTR influences protein production in the context of this study and, in particular, to find out whether a specific region or motif within the Albumin 3'UTR is responsible for higher protein production and mRNA expression, a series of deletions were made within the first 150 nucleotides of this 3'UTR. The data obtained from cell lines stably transfected with the full length Albumin 3'UTR and its deletion variants showed that protein production is significantly reduced by the deletion of regions containing nucleotides 1-50, 1-100, 1-150, 101-150 (see table 7.1). However, the high-level presence of Luciferase in culture medium samples was maintained following the deletion of nucleotide 51-100 and 51-150.

The reduction of Luciferase activity in cells expressing the deletion construct 1-50 & 101-150 was also found to be close to significance (p value = 0.055). Interestingly, the data presented in chapter 5 also showed that Luciferase activity in cell extracts from cells expressing the deletion constructs Δ1-50, Δ1-150, Δ101-150 and Δ1-50 & 101-150 showed a similar pattern to those observed with the culture medium samples in that the Luciferase activity was significantly reduced. The data presented in chapter 5 also showed that the deletion of nucleotides 1-50, 1-150 and 1-50 &101-150 significantly reduces the abundance of Luciferase mRNA.

To test the hypothesis that the differences observed in secreted *Gaussia* Luciferase activity and mRNA expression for the Albumin 3'UTR deletion variants were correlated with the ability of the variants to bind trans-acting factor/s, EMSA and UV Cross-linking were employed to assess binding to proteins in CHO cell extracts. As shown in chapter 6, there was specific binding of proteins to the Albumin 3'UTR.

The data obtained from the EMSA competition assay and subsequent quantification data showed that the strongest competition was observed by the deletion variant 1-50. This indicates that protein binding to the Albumin 3'UTR occurs primarily outside the first 50 nucleotides. The data obtained in chapter 5 also consistently showed that deletion of nucleotides 1-50 cause a significant reduction in Luciferase activity and mRNA abundance (see Table 7.1). It appeared from the data presented in chapter 6 that little or no competition was observed by the deletion variant 1-150. This indicated that all possible binding motifs/sites for the candidate protein/s are located within the first 150 nucleotides. In agreement with the binding data, deletion of nucleotides 1-150 consistently showed the greatest reduction in Luciferase activity, both in the culture medium and cell extracts, and in mRNA level. It should be noted that the observed differences in Luciferase activity and mRNA expression, are unlikely to have been as a result of the variation in length of the poly (A) region due to presence of a polyadenylation signal at nucleotide 166 to 171 of the Albumin 3'UTR, the position of which is similar for all deletion variants.

The deletion variant 101-150 showed the second strongest competition with the radiolabelled full length Albumin 3'UTR RNA transcript and significantly reduced Luciferase activity. This suggested that potential regulatory motif/s for candidate protein/s are located within nucleotides 51-100 of the Albumin 3'UTR. The binding data also showed a lesser degree of

competition with the other competitors (the deletion variant 1-100, 51-100, 51-150 and 1-50 & 101-150). This suggested that candidate protein/s may bind to other region/s within the Albumin 3'UTR with varying degrees of affinity.

In agreement with the data obtained from EMSA competition assay, the results of the UV Cross-linking experiments also showed that the strongest competition for binding to a protein with the molecular weight of ~50 kDa was provided by the deletion variant 1-50. This suggested yet again that the region containing nucleotides 1-50 is not required for binding to any candidate protein/s. The deletion region 101-150 provided the second strongest competition for binding to the protein/s with a molecular weight of ~50kDa. This suggested that binding motif/s for the candidate protein is probably located in the region that contains nucleotides 51-100.

Therefore, based on the findings from EMSA and UV Cross-linking competition assays (chapter 6), the deletion variant Δ1-50 was selected as bait to isolate and subsequently identify protein/s that bind to the Albumin 3'UTR RNA transcript. Three proteins were identified by mass spectrophotometry analysis, and these were identified as follows:

1) Far upstream element (FUSE) binding protein 3
2) Regulator of differentiation 1 (Rod 1)
3) CUG-BP1

In this study CUG-BP1 was studied further, due to the existence of considerable previous evidence for the binding of CUG-BP1 to CUG and UG dinucleotide repeats and also due to the similarity of a region (nucleotides 75-85) within the Albumin 3'UTR (UGUUUUCUUUU) with a reported binding site for CUG-BP1 (UGUUUGUUUGU) . The CUG-BP1 is a member of the CELF/Bruno-like family of evolutionary conserved RNA

binding proteins that play essential roles in post-transcriptional gene regulation. These are amongst the most highly conserved RNA binding proteins, all containing three RNA recognition motifs [203]. CUG-BP1 was the first member of this family to be identified [198] and can act as a regulator for alternative splicing [203-206] , translation and stability [199, 207, 208] and the deadenylation of target mRNA transcripts [209, 210]. CUG-BP1, was first identified as a protein which binds to CUG repeats in the onset of DM1 [208, 211], but contrary to its name it is has been shown that CUG-BP1 more readily binds to UG motifs than CUG repeats in a yeast three-hybrid system [212, 213]. The binding of CUG-BP1 to the CUG repeats could explain the aberrant pattern of splicing which is the hallmark of DM1 [206]. Interestingly, it was shown recently that a conserved 11-mer nucleotide sequence (UGUUUGUUUGU) termed GU-rich element (GRE), regulates mRNA decay by binding to CUG-BP1 [202].

To confirm the role of CUG-BP1 in forming CHO cell protein-Albumin 3'UTR complexes, CHO cells were treated with siRNA specific to CUG-BP1 and a negative universal control siRNA prior to S-100 CHO cell extraction. Treatment of the stably transfected CHO cells with CUG-BP1 siRNA led to the severe loss of complex formation between the RNA transcript and cell extracts revealing that the knock down of CUG-BP1 expression affects protein binding to the Albumin 3'UTR, suggesting that CUG-BP1 is directly involved in complex formation.

It was also shown in this study that siRNA treatment of stably transfected cells with the construct GG*G*Alb 3Δβ did not lead to a significant increase in *Gaussia* Luciferase activity. The opposite was seen following siRNA treatment of stably transfected cells with the construct GG*G*Alb (Δ51-100) 3Δβ, where a significant increase in *Gaussia* Luciferase

activity was seen. In contrast to the reporter activity, the mRNA expression level was significantly higher after siRNA treatment of the stably transfected cells with construct GG*G*Alb 3Δβ. In addition, siRNA treatment of stably transfected cells with construct GG*G*Alb (Δ51-100) 3Δβ also led to a significant increase in the mRNA abundance level. The mechanism/s which causes the discrepancy between the level of mRNA target and the effectiveness of its translation is not yet understood. The differences in protein production and mRNA expression observed after siRNA knock down of CUG-BP1 is speculated to be a result of CUG-BP1 binding to its different possible binding sites or motifs within the Albumin 3'UTR.

Table 7.1 Comparative effects of Albumin 3'UTR deletions on reporter activity, mRNA expression and protein binding where ↓* indicates a significant reduction, Ŗ indicates no significant reduction, ↓ indicates 2 significant reductions, ↑** indicates increase of 2 significance , ND indicates not determined, ↔ indicates maintained activity, - indicates no competition, + + + + indicates the strongest competitor, + + + indicates the second strongest competitor, + + indicates the third strongest competitor, + indicates the fourth strongest competitor.**

Transcript	Medium	Cell Extract	mRNA	EMSA	UV Cross-linking	+ siRNA (medium)	+ siRNA (mRNA)
Δ 1-50	↓**	Ŗ	↓*	++ ++	+ + + +	ND	ND
Δ 1-100	↓*	Ŗ	Ŗ	+	-	ND	ND
Δ 1-150	↓**	↓**	↓**	-	-	ND	ND
Δ 51-100	↔	Ŗ	Ŗ	+ +	-	↑**	↑**
Δ 101-150	↓**	↓**	Ŗ	+ + +	+ + +	ND	ND
Δ 51-150	Ŗ	Ŗ	Ŗ	+	+ +	ND	ND
Δ 1-50 & 10150	↓close to*	↓**	↓*	+ +	+ +	ND	ND

Based on the data obtained from EMSA and UV Cross-linking competition assays carried out in this study, it is highly tempting to speculate that CUG-BP1 binds more readily to the 11-mer nucleotide sequence (UGUUUGUUUGU) termed GU-rich elements within the Albumin 3'UTR (nucleotides 75-85) with a greater affinity than to other available binding sites. When this binding site is deleted (deletion variant 51-100 or construct GG*G*Alb3Δβ (Δ51-100)), CUG-BP1 may bind with a lower affinity to other possible binding site/s or motifs within the Albumin 3'UTR. This, therefore, could be a possible explanation for the observed differences in mRNA abundance and protein secretion in Albumin 3'UTR deletion variants.

It is not evident from this study whether the mechanisms by which the Albumin 3'UTR and subsequent protein binding involve mRNA stability or translation of transcripts. To assess the stability of the transcripts (with the full length Albumin 3'UTR and the deletion variants), actinomycin could be used to inhibit transcription and measure half-life of various transcripts. Such experiments would allow a correlation to be made between the stability of mRNA transcript and various deletions within the Albumin 3'UTR. To address the effects of Albumin 3'UTR and protein binding on translation of the transcript, the association of various mutated mRNA transcripts of the Albumin 3'UTR with ribosomes could be examined by separating them on sucrose gradients (polysome profiling) and then analysing presence of Luciferase transcripts throughout the gradient [214].

To summarise, in this thesis evidence for specific protein binding to the Albumin 3'UTR has been presented, and its probable binding site has been identified (Figure 7.1). Furthermore, inclusion of the Albumin 3'UTR in a reporter vector (without vector-derived sequences) was shown to lead to increased Luciferase activity compared to when the native *Gaussia* Luciferase 3'UTR was utilised in a vector containing the vector-derived sequences. The

actual relationship between protein binding to the Albumin 3'UTR and the level of protein expression need to be explored in depth.

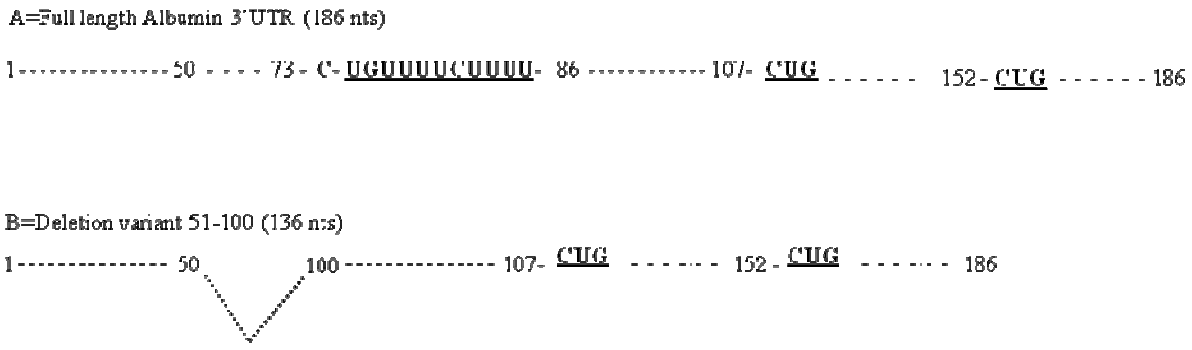


Figure 7.1 Schematic representation of probable CUG-BP1 binding sites within the full length Albumin 3'UTR (A) and the deletion variant 51-100 (B). The proposed binding sites are underlined.

To examine this, various deletions can be made within the Albumin 3'UTR. For instance, the CUG repeats and the 11-mer nucleotides UGUUUUCUUUU can be deleted and then the interaction between the CUG-BP1 and the Albumin 3'UTR can be examined by performing UV Cross-linking and EMSA competition assays.

8. Appendices

8.1 DNA sequences

8.1.1 Gaussia Luciferase 5'UTR:

GGTACTCAAAGTATCTTCTGGCAGGGAAA

8.1.2 Gaussia Luciferase signal peptide:

ATGGGAGTCAAAGTTCTGTTTGCCCTGATCTGCATCGCTGTGGCCGAGGCC

8.1.3 Gaussia Luciferase coding region:

AAGCCCACCGAGAACAACGAAGACTTCAACATCGTGGCCGTGGCCAGCAACTTC
GCGACCACGGATCTCGATGCTGACCGCGGGAAGTTGCCCGGCAAGAAGCTGCCG
CTGGAGGTGCTCAAAGAGATGGAAGCCAATGCCCGGAAAGCTGGCTGCACCAGG
GGCTGTCTGATCTGCCTGTCCCACATCAAGTGCACGCCCAAGATGAAGAAGTTCA
TCCCAGGACGCTGCCACACCTACGAAGGCGACAAAGAGTCCGCACAGGGCGGCA
TAGGCGAGGCGATCGTCGACATTCCTGAGATTCCTGGGTTCAAGGACTTGGAGCC
CATGGAGCAGTTCATCGCACAGGTCGATCTGTGTGTGGACTGCACAACTGGCTGC
CTCAAAGGGCTTGCCAACGTGCAGTGTTCTGACCTGCTCAAGAAGTGGCTGCCGC
AACGCTGTGCGACCTTTGCCAGCAAGATCCAGGGCCAGGTGGACAAGATCAAGG
GGGCCGGTGGTGACTAA

8.1.4 Gaussia Luciferase 3'UTR:

TCCTAATAGAATACTGCATAACTGGATGATGATATACTAGCTTATTGCTCATAAAA
TGGCCATTTTTTGTAACAAATCGAGTCTATGTAATTCAAAATACCTAATTAATTGT
TAATACATATGTAATTCCTATAAATATAATTTATGCAATCC

8.1.5 Human Albumin signal peptide:

ATGAAGTGGGTAACCTTTATTTCCCTTCTTTTTCTCTTTAGCTCGGCTTATTCC

8.1.6 Human Albumin 3'UTR:

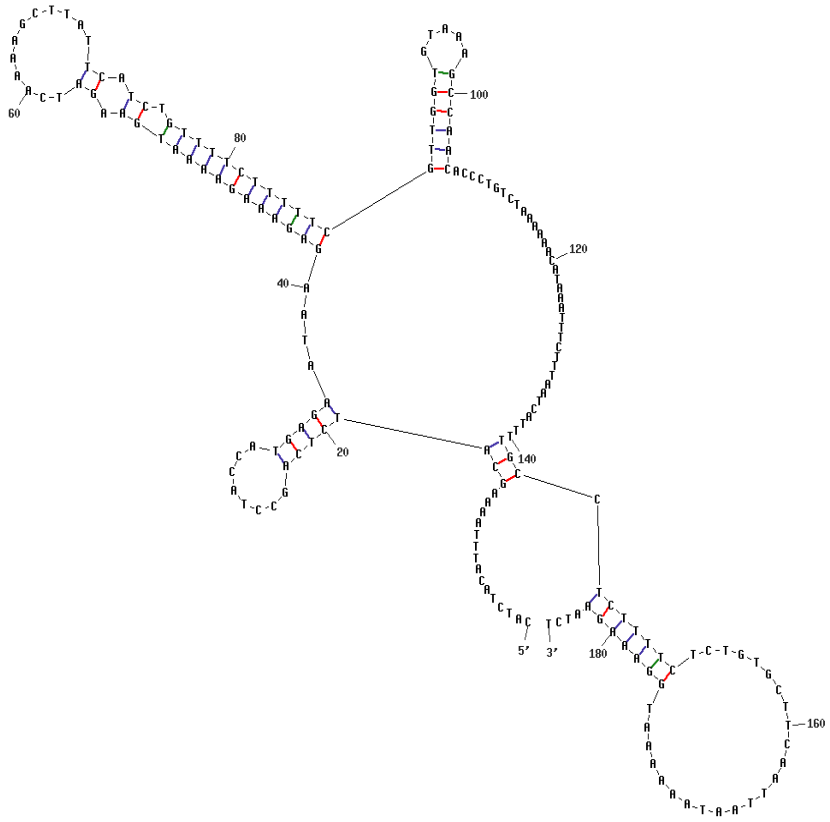
CATCTACATTTAAAAGCATCTCAGCCTACCATGAGAATAAGAGAAAGAAAATGA
AGATCAAAAGCTTATTCATCTGTTTTCTTTTTCGTTGGTGTAAAGCCAACACCCTG
TCTAAAAAACATAAATTTCTTTAATCATTTTGCCTCTTTTCTCTGTGCTTCAATTAA
TAAAAAATGGAAAGAATCT

8.1.7 Human Chymotrypsinogen signal peptide:

ATGGCTTTCCTCTGGCTCCTCTCCTGCTGGGCCCTCCTGGGTACCACCTTCGGC

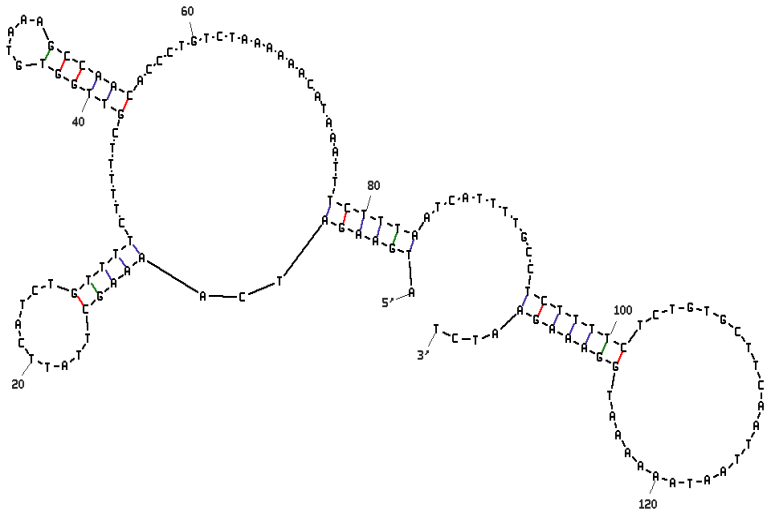
8.2 mFold mRNA secondary structure prediction:

8.2.1 Albumin 3'UTR (wild type):



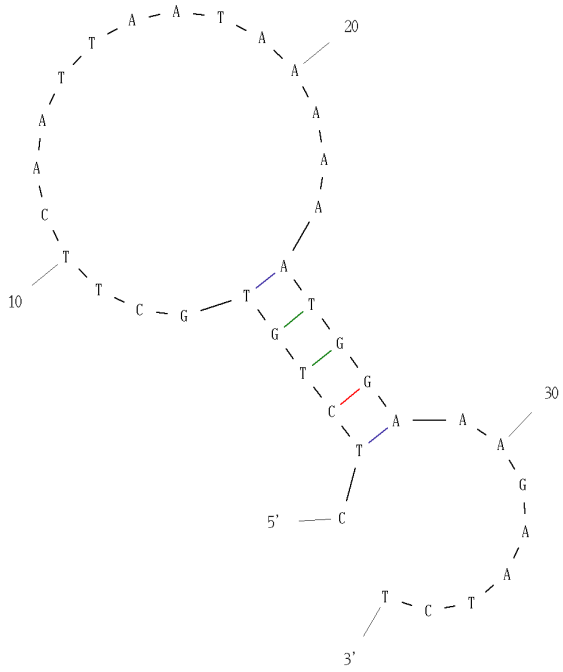
dG = -28.59 [initially -28.50] unknown

8.2.2 Δ 1-50:



dG = -14.11 [initially -16.60] unknown

8.2.3 Δ 1-150:



dG = -0.3 [initially -0.30] unknown

8.3 Sequencing results:

Construct GXG*Alb 3Δβ

```
Query  311   CTCGATGCTGACCGCGGGAAGTTGCCCGGCAAGAAGCTGCCGCTGGAGGTGCTCAAAGAG  370
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  59    CTCGATGCTGACCGCGGGAAGTTGCCCGGCAAGAAGCTGCCGCTGGAGGTGCTCAAAGAG  118

Query  371   ATGGAAGCCAATGCCCGGAAAGCTGGCTGCACCAGGGGCTGTCTGATCTGCCTGTCCCAC  430
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  119   ATGGAAGCCAATGCCCGGAAAGCTGGCTGCACCAGGGGCTGTCTGATCTGCCTGTCCCAC  178

Query  431   ATCAAGTGCACGCCCAAGATGAAGAAGTTCATCCCAGGACGCTGCCACACCTACGAAGGC  490
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  179   ATCAAGTGCACGCCCAAGATGAAGAAGTTCATCCCAGGACGCTGCCACACCTACGAAGGC  238

Query  491   GACAAAGAGTCCGCACAGGGCGGCATAGGCGAGGCGATCGTCGACATTCCTGAGATTCCT  550
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  239   GACAAAGAGTCCGCACAGGGCGGCATAGGCGAGGCGATCGTCGACATTCCTGAGATTCCT  298

Query  551   GGGTTCAAGGACTTGGAGCCCATGGAGCAGTTCATCGCACAGGTCGATCTGTGTGTGGAC  610
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  299   GGGTTCAAGGACTTGGAGCCCATGGAGCAGTTCATCGCACAGGTCGATCTGTGTGTGGAC  358

Query  611   TGCACAACTGGCTGCCTCAAAGGGCTTGCCAACGTGCAGTGTTCTGACCTGCTCAAGAAG  670
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  359   TGCACAACTGGCTGCCTCAAAGGGCTTGCCAACGTGCAGTGTTCTGACCTGCTCAAGAAG  418

Query  671   TGGCTGCCGCAACGCTGTGCGACCTTTGCCAGCAAGATCCAGGGCCAGGTGGACAAGATC  730
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  419   TGGCTGCCGCAACGCTGTGCGACCTTTGCCAGCAAGATCCAGGGCCAGGTGGACAAGATC  478

Query  731   AAGGGGGCCGGTGGTGACTAACATCTACATTTAAAAGCATCTCAGCCTACCATGAGAATA  790
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  479   AAGGGGGCCGGTGGTGACTAACATCTACATTTAAAAGCATCTCAGCCTACCATGAGAATA  538

Query  791   AGAGAAAGAAAATGAAGATCAAAAGCTTATTCATCTGTTTTCTTTTTCGTTGGTGTAAAG  850
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  539   AGAGAAAGAAAATGAAGATCAAAAGCTTATTCATCTGTTTTCTTTTTCGTTGGTGTAAAG  598

Query  851   CCAACACCCTGTCTAAAAAACATAAATTTCTTTAATCATTTTGCCTCTTTTCTCTGTGCT  910
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  599   CCAACACCCTGTCTAAAAAACATAAATTTCTTTAATCATTTTGCCTCTTTTCTCTGTGCT  658

Query  911   TCAATTAATAAAAAATGGAAAGAATCTAATAGTGTGTGGGAATTTTTTGTGTCTCTCACT  970
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  659   TCAATTAATAAAAAATGGAAAGAATCTAATAGTGTGTGGGAATTTTTTGTGTCTCTCACT  718

Query  971   CGGAAGGACATATGGGAGGGCAATCATTTAAAACATCAGAATGAGTATTTGGTTTAGAGT  1030
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  719   CGGAAGGACATATGGGAGGGCAATCATTTAAAACATCAGAATGAGTATTTGGTTTAGAGT  778

Query  1031  TTGGCAACATATGCCCATATGCTGGCTGCCATGAACAAAGGTTGGCTATAAAGAGGTCAT  1090
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  779   TTGGCAACATATGCCCATATGCTGGCTGCCATGAACAAAGGTTGGCTATAAAGAGGTCAT  838

Query  1091  CAGTATATGAAACAGCCCCCTGCTGTCCATTCCTTATTCCATAGAAAAGCCTTGACTTGA  1150
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  839   CAGTATATGAAACAGCCCCCTGCTGTCCATTCCTTATTCCATAGAAAAGCCTTGACTTGA  898
```

```
Query  1151  GGTTAGAttttttttatattttgttttgtgttatttttttctttAACATCCCTAAAATTT  1210
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  899   GGTTAGATTTTTTTTATATTTTGTTTTGTGTTATTTTTTTCTTTAACATCCCTAAAATTT  958

Query  1211  TCCTTACATG  1220
             ||||||||||
Sbjct  959   TCCTTACATG  968
```

Construct GAG*G 3Δβ

```
Query  61   CCTCCGCGGCCCCGAATTCGAGCTCGGTACCCGGGGATCCCCCGGGCTGCAGGAATTCGG  120
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  94   CCTCCGCGGCCCCGAATTCGAGCTCGGTACCCGGGGATCCCCCGGGCTGCAGGAATTCGG  153

Query  121  CACGAGGGTACTCAAAGTATCTTCTGGCAGGGAAAATGAAGTGGGTAACCTTTATTTCCC  180
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  154  CACGAGGGTACTCAAAGTATCTTCTGGCAGGGAAAATGAAGTGGGTAACCTTTATTTCCC  213

Query  181  TTCTTTTTCTCTTTAGCTCGGCTTATTCCAAGCCCACCGAGAACAACGAAGACTTCAACA  240
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  214  TTCTTTTTCTCTTTAGCTCGGCTTATTCCAAGCCCACCGAGAACAACGAAGACTTCAACA  273

Query  241  TCGTGGCCGTGGCCAGCAACTTCGCGACCACGGATCTCGATGCTGACCGCGGGAAGTTGC  300
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  274  TCGTGGCCGTGGCCAGCAACTTCGCGACCACGGATCTCGATGCTGACCGCGGGAAGTTGC  333

Query  301  CCGGCAAGAAGCTGCCGCTGGAGGTGCTCAAAGAGATGGAAGCCAATGCCCGGAAAGCTG  360
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  334  CCGGCAAGAAGCTGCCGCTGGAGGTGCTCAAAGAGATGGAAGCCAATGCCCGGAAAGCTG  393

Query  361  GCTGCACCAGGGGCTGTCTGATCTGCCTGTCCCACATCAAGTGCACGCCCAAGATGAAGA  420
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  394  GCTGCACCAGGGGCTGTCTGATCTGCCTGTCCCACATCAAGTGCACGCCCAAGATGAAGA  453

Query  421  AGTTCATCCCAGGACGCTGCCACACCTACGAAGGCGACAAAGAGTCCGCACAGGGCGGCA  480
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  454  AGTTCATCCCAGGACGCTGCCACACCTACGAAGGCGACAAAGAGTCCGCACAGGGCGGCA  513

Query  481  TAGGCGAGGCGATCGTCGACATTCCTGAGATTCCTGGGTTCAAGGACTTGGAGCCCATGG  540
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  514  TAGGCGAGGCGATCGTCGACATTCCTGAGATTCCTGGGTTCAAGGACTTGGAGCCCATGG  573

Query  541  AGCAGTTCATCGCACAGGTCGATCTGTGTGTGGACTGCACAACTGGCTGCCTCAAAGGGC  600
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  574  AGCAGTTCATCGCACAGGTCGATCTGTGTGTGGACTGCACAACTGGCTGCCTCAAAGGGC  633

Query  601  TTGCCAACGTGCAGTGTTCTGACCTGCTCAAGAAGTGGCTGCCGCAACGCTGTGCGACCT  660
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  634  TTGCCAACGTGCAGTGTTCTGACCTGCTCAAGAAGTGGCTGCCGCAACGCTGTGCGACCT  693

Query  661  TTGCCAGCAAGATCCAGGGCCAGGTGGACAAGATCAAGGGGGCCGGTGGTGACTAATCCT  720
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  694  TTGCCAGCAAGATCCAGGGCCAGGTGGACAAGATCAAGGGGGCCGGTGGTGACTAATCCT  753

Query  721  AATAGAATACTGCATAACTGGATGATGATATACTAGCTTATTGCTCATAAAATGGCCATT  780
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Sbjct  754  AATAGAATACTGCATAACTGGATGATGATATACTAGCTTATTGCTCATAAAATGGCCATT  813

Query  781  TTTTGTAACAAATCGAGTCTATGTAATTCAAAATACCTAATTAATTGTTAATACATATGT  840
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  814  TTTTGTAACAAATCGAGTCTATGTAATTCAAAATACCTAATTAATTGTTAATACATATGT  873

Query  841  AATTCCTATAAATATAATTTATGCAATCC  869
            |||||||||||||||||||||||||||||
Sbjct  874  AATTCCTATAAATATAATTTATGCAATCC  902
```

Construct GAG*Alb 3Δβ

```
Query  61   CCTCCGCGGCCCCGAATTCGAGCTCGGTACCCGGGGATCCCCCGGGCTGCAGGAATTCGG  120
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  77   CCTCCGCGGCCCCGAATTCGAGCTCGGTACCCGGGGATCCCCCGGGCTGCAGGAATTCGG  136

Query  121  CACGAGGGTACTCAAAGTATCTTCTGGCAGGGAAAATGAAGTGGGTAACCTTTATTTCCC  180
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  137  CACGAGGGTACTCAAAGTATCTTCTGGCAGGGAAAATGAAGTGGGTAACCTTTATTTCCC  196

Query  181  TTCTTTTTCTCTTTAGCTCGGCTTATTCCAAGCCCACCGAGAACAACGAAGACTTCAACA  240
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  197  TTCTTTTTCTCTTTAGCTCGGCTTATTCCAAGCCCACCGAGAACAACGAAGACTTCAACA  256

Query  241  TCGTGGCCGTGGCCAGCAACTTCGCGACCACGGATCTCGATGCTGACCGCGGGAAGTTGC  300
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  257  TCGTGGCCGTGGCCAGCAACTTCGCGACCACGGATCTCGATGCTGACCGCGGGAAGTTGC  316

Query  301  CCGGCAAGAAGCTGCCGCTGGAGGTGCTCAAAGAGATGGAAGCCAATGCCCGGAAAGCTG  360
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  317  CCGGCAAGAAGCTGCCGCTGGAGGTGCTCAAAGAGATGGAAGCCAATGCCCGGAAAGCTG  376

Query  361  GCTGCACCAGGGGCTGTCTGATCTGCCTGTCCCACATCAAGTGCACGCCCAAGATGAAGA  420
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  377  GCTGCACCAGGGGCTGTCTGATCTGCCTGTCCCACATCAAGTGCACGCCCAAGATGAAGA  436

Query  421  AGTTCATCCCAGGACGCTGCCACACCTACGAAGGCGACAAAGAGTCCGCACAGGGCGGCA  480
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  437  AGTTCATCCCAGGACGCTGCCACACCTACGAAGGCGACAAAGAGTCCGCACAGGGCGGCA  496

Query  481  TAGGCGAGGCGATCGTCGACATTCCTGAGATTCCTGGGTTCAAGGACTTGGAGCCCATGG  540
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  497  TAGGCGAGGCGATCGTCGACATTCCTGAGATTCCTGGGTTCAAGGACTTGGAGCCCATGG  556

Query  541  AGCAGTTCATCGCACAGGTCGATCTGTGTGTGGACTGCACAACTGGCTGCCTCAAAGGGC  600
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  557  AGCAGTTCATCGCACAGGTCGATCTGTGTGTGGACTGCACAACTGGCTGCCTCAAAGGGC  616

Query  601  TTGCCAACGTGCAGTGTTCTGACCTGCTCAAGAAGTGGCTGCCGCAACGCTGTGCGACCT  660
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  617  TTGCCAACGTGCAGTGTTCTGACCTGCTCAAGAAGTGGCTGCCGCAACGCTGTGCGACCT  676

Query  661  TTGCCAGCAAGATCCAGGGCCAGGTGGACAAGATCAAGGGGGCCGGTGGTGACTAACATC  720
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  677  TTGCCAGCAAGATCCAGGGCCAGGTGGACAAGATCAAGGGGGCCGGTGGTGACTAACATC  736

Query  721  TACATTTAAAAGCATCTCAGCCTACCATGAGAATAAGAGAAAGAAAATGAAGATCAAAAG  780
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  737  TACATTTAAAAGCATCTCAGCCTACCATGAGAATAAGAGAAAGAAAATGAAGATCAAAAG  796
```

```
Query  781  CTTATTCATCTGTTTTCTTTTTCGTTGGTGTAAAGCCAACAC  822
            |||||||||||||||||||||||||||||||||||||||||||
Sbjct  797  CTTATTCATCTGTTTTCTTTTTCGTTGGTGTAAAGCCAACAC  838
```

Albumin 3'UTR deletion construct 51-150

```
Query  87   GGGGATCCCCCGGGCTGCAGGAATTCGGCACGAGGGTACTCAAAGTATCTTCTGGCAGGG  146
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  68   GGGGATCCCCCGGGCTGCAGGAATTCGGCACGAGGGTACTCAAAGTATCTTCTGGCAGGG  127

Query  147  AAAATGGGAGTCAAAGTTCTGTTTGCCCTGATCTGCATCGCTGTGGCCGAGGCCAAGCCC  206
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  128  AAAATGGGAGTCAAAGTTCTGTTTGCCCTGATCTGCATCGCTGTGGCCGAGGCCAAGCCC  187

Query  207  ACCGAGAACAACGAAGACTTCAACATCGTGGCCGTGGCCAGCAACTTCGCGACCACGGAT  266
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  188  ACCGAGAACAACGAAGACTTCAACATCGTGGCCGTGGCCAGCAACTTCGCGACCACGGAT  247

Query  267  CTCGATGCTGACCGCGGGAAGTTGCCCGGCAAGAAGCTGCCGCTGGAGGTGCTCAAAGAG  326
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  248  CTCGATGCTGACCGCGGGAAGTTGCCCGGCAAGAAGCTGCCGCTGGAGGTGCTCAAAGAG  307

Query  327  ATGGAAGCCAATGCCCGGAAAGCTGGCTGCACCAGGGGCTGTCTGATCTGCCTGTCCCAC  386
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  308  ATGGAAGCCAATGCCCGGAAAGCTGGCTGCACCAGGGGCTGTCTGATCTGCCTGTCCCAC  367

Query  387  ATCAAGTGCACGCCCAAGATGAAGAAGTTCATCCCAGGACGCTGCCACACCTACGAAGGC  446
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  368  ATCAAGTGCACGCCCAAGATGAAGAAGTTCATCCCAGGACGCTGCCACACCTACGAAGGC  427

Query  447  GACAAAGAGTCCGCACAGGGCGGCATAGGCGAGGCGATCGTCGACATTCCTGAGATTCCT  506
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  428  GACAAAGAGTCCGCACAGGGCGGCATAGGCGAGGCGATCGTCGACATTCCTGAGATTCCT  487

Query  507  GGGTTCAAGGACTTGGAGCCCATGGAGCAGTTCATCGCACAGGTCGATCTGTGTGTGGAC  566
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  488  GGGTTCAAGGACTTGGAGCCCATGGAGCAGTTCATCGCACAGGTCGATCTGTGTGTGGAC  547

Query  567  TGCACAACTGGCTGCCTCAAAGGGCTTGCCAACGTGCAGTGTTCTGACCTGCTCAAGAAG  626
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  548  TGCACAACTGGCTGCCTCAAAGGGCTTGCCAACGTGCAGTGTTCTGACCTGCTCAAGAAG  607

Query  627  TGGCTGCCGCAACGCTGTGCGACCTTTGCCAGCAAGATCCAGGGCCAGGTGGACAAGATC  686
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  608  TGGCTGCCGCAACGCTGTGCGACCTTTGCCAGCAAGATCCAGGGCCAGGTGGACAAGATC  667

Query  687  AAGGGGGCCGGTGGTGACTAACATCTACATTTAAAAGCATCTCAGCCTACCATGAGAATA  746
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  668  AAGGGGGCCGGTGGTGACTAACATCTACATTTAAAAGCATCTCAGCCTACCATGAGAATA  727

Query  747  AGAGAAAGAAACTCTGTGCTTCAATTAATAAAAAATGGAAAGAATCT  793
            |||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  728  AGAGAAAGAAACTCTGTGCTTCAATTAATAAAAAATGGAAAGAATCT  774
```

Albumin 3'UTR deletion construct 1-50&101-150

```
Query  89   GCTGCAGGAATTCGGCACGAGGGTACTCAAAGTATCTTCTGGCAGGGAAAATGGGAGTCA  148
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  82   GCTGCAGGAATTCGGCACGAGGGTACTCAAAGTATCTTCTGGCAGGGAAAATGGGAGTCA  141

Query  149  AAGTTCTGTTTGCCCTGATCTGCATCGCTGTGGCCGAGGCCAAGCCCACCGAGAACAACG  208
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  142  AAGTTCTGTTTGCCCTGATCTGCATCGCTGTGGCCGAGGCCAAGCCCACCGAGAACAACG  201

Query  209  AAGACTTCAACATCGTGGCCGTGGCCAGCAACTTCGCGACCACGGATCTCGATGCTGACC  268
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  202  AAGACTTCAACATCGTGGCCGTGGCCAGCAACTTCGCGACCACGGATCTCGATGCTGACC  261

Query  269  GCGGGAAGTTGCCCGGCAAGAAGCTGCCGCTGGAGGTGCTCAAAGAGATGGAAGCCAATG  328
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  262  GCGGGAAGTTGCCCGGCAAGAAGCTGCCGCTGGAGGTGCTCAAAGAGATGGAAGCCAATG  321

Query  329  CCCGGAAAGCTGGCTGCACCAGGGGCTGTCTGATCTGCCTGTCCCACATCAAGTGCACGC  388
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  322  CCCGGAAAGCTGGCTGCACCAGGGGCTGTCTGATCTGCCTGTCCCACATCAAGTGCACGC  381

Query  389  CCAAGATGAAGAAGTTCATCCCAGGACGCTGCCACACCTACGAAGGCGACAAAGAGTCCG  448
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  382  CCAAGATGAAGAAGTTCATCCCAGGACGCTGCCACACCTACGAAGGCGACAAAGAGTCCG  441

Query  449  CACAGGGCGGCATAGGCGAGGCGATCGTCGACATTCCTGAGATTCCTGGGTTCAAGGACT  508
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  442  CACAGGGCGGCATAGGCGAGGCGATCGTCGACATTCCTGAGATTCCTGGGTTCAAGGACT  501

Query  509  TGGAGCCCATGGAGCAGTTCATCGCACAGGTCGATCTGTGTGTGGACTGCACAACTGGCT  568
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  502  TGGAGCCCATGGAGCAGTTCATCGCACAGGTCGATCTGTGTGTGGACTGCACAACTGGCT  561

Query  569  GCCTCAAAGGGCTTGCCAACGTGCAGTGTTCTGACCTGCTCAAGAAGTGGCTGCCGCAAC  628
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  562  GCCTCAAAGGGCTTGCCAACGTGCAGTGTTCTGACCTGCTCAAGAAGTGGCTGCCGCAAC  621

Query  629  GCTGTGCGACCTTTGCCAGCAAGATCCAGGGCCAGGTGGACAAGATCAAGGGGGCCGGTG  688
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  622  GCTGTGCGACCTTTGCCAGCAAGATCCAGGGCCAGGTGGACAAGATCAAGGGGGCCGGTG  681

Query  689  GTGACTAAATGAAGATCAAAAGCTTATTCATCTGTTTTCTTTTTCGTTGGTGTAAAGCCT  748
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  682  GTGACTAAATGAAGATCAAAAGCTTATTCATCTGTTTTCTTTTTCGTTGGTGTAAAGCCT  741

Query  749  CTGTGCTTCAATTAATAAAAAATGGAAAGAATCT  782
            ||||||||||||||||||||||||||||||||||
Sbjct  742  CTGTGCTTCAATTAATAAAAAATGGAAAGAATCT  77
```

8.4 Programme Settings for Real-Time PCR

Detection Format	**Block Type**	**Reaction Volume**
SYBR Green	96 (LightCycler®480)	20µl

Programs

Programs Name	**Cycle**	**Analysis Mode**
Pre-Incubation	1	None
Amplification	60	Quantification
Melting Curve	1	Melting Curves
Cooling	1	None

Temperature Targets

Target (°C)	**Acquisition Mode**	**Hold (hh:mm:ss)**	**Ramp rate**	**Acquisition (per °C)**
Pre-Incubation:				
95	None	00:05:00	4.4	_
Amplification:				
95	None	00:00:10	4.4	_
63	None	00:00:15	4.4	_
72	Single	00:00:05	2.2	_
Melting Curve				
95	None	00:00:05	4.4	_
65	None	00:10:00	2.2	_
97	Continues	_	_	5
Cooling				
4	None	00:00:10	1.5	-

9. References

1. Jayapal KP. WK, Hu WS: **Recombinant Protein Therapeutics from CHO Cells-20 years and Counting**. *CHO Consortium*.
2. Katakura Y. SP, Miura T, Ohashi H, Teruya K, Shirahata S: **Productivity enhancement of recombinat protein in CHO cells via specific promotor activation by oncogenes**. *Cytotechnology* 1999, **31**:103-109.
3. Schmidt FR: **Recombinant expression systems in the pharmaceutical industry**. *Appl Microbiol Biotechnol* 2004, **65**(4):363-372.
4. Kronman C, Velan B, Gozes Y, Leitner M, Flashner Y, Lazar A, Marcus D, Sery T, Papier Y, Grosfeld H: **Production and secretion of high levels of recombinant human acetylcholinesterase in cultured cell lines: microheterogeneity of the catalytic subunit**. *Gene* 1992, **121**(2):295-304.
5. Wenger RH, Moreau H, Nielsen PJ: **A comparison of different promoter, enhancer, and cell type combinations in transient transfections**. *Anal Biochem* 1994, **221**(2):416-418.
6. Kemball-Cook G, Garner I, Imanaka Y, Nishimura T, O'Brien DP, Tuddenham EG, McVey JH: **High-level production of human blood coagulation factors VII and XI using a new mammalian expression vector**. *Gene* 1994, **139**(2):275-279.
7. Gandor C, Leist C, Fiechter A, Asselbergs FA: **Amplification and expression of recombinant genes in serum-independent Chinese hamster ovary cells**. *FEBS Lett* 1995, **377**(3):290-294.
8. Gopalkrishnan RV, Christiansen KA, Goldstein NI, DePinho RA, Fisher PB: **Use of the human EF-1alpha promoter for expression can significantly increase success in establishing stable cell lines with consistent expression: a study using the tetracycline-inducible system in human cancer cells**. *Nucleic Acids Res* 1999, **27**(24):4775-4782.
9. Condreay JP, Kost TA: **Baculovirus expression vectors for insect and mammalian cells**. *Curr Drug Targets* 2007, **8**(10):1126-1131.
10. Kost TA, Condreay JP, Jarvis DL: **Baculovirus as versatile vectors for protein expression in insect and mammalian cells**. *Nat biotechnol* 2005, **23**(5):567-575.
11. Spenger A, Ernst W, Condreay JP, Kost TA, Grabherr R: **Influence of promoter choice and trichostatin A treatment on expression of baculovirus delivered genes in mammalian cells**. *Protein Expr Purif* 2004, **38**(1):17-23.
12. Merrihew RV, Kost TA, Condreay JP: **Baculovirus-mediated gene delivery into mammalian cells**. *Methods Mol Biol* 2004, **246**:355-365.
13. Kost TA, Condreay JP: **Recombinant baculoviruses as mammalian cell gene-delivery vectors**. *Trends Biotechnol* 2002, **20**(4):173-180.
14. Makrides SC: **Components of vectors for gene transfer and expression in mammalian cells**. *Protein Expr Purif* 1999, **17**(2):183-202.
15. Wurm FM: **Production of recombinant protein therapeutics in cultivated mammalian cells**. *Nat biotechnol* 2004, **22**(11):1393-1398.

16. Mutskov V, Felsenfeld G: **Silencing of transgene transcription precedes methylation of promoter DNA and histone H3 lysine 9**. *EMBO J* 2004, **23**(1):138-149.
17. Richards EJ, Elgin SCR: **Epigenetic codes for heterochromatin formation and silencing: rounding up the usual suspects**. *Cell* 2002, **108**(4):489-500.
18. Upreti RK, Kumar M, Shankar V: **Bacterial glycoproteins: functions, biosynthesis and applications**. *Proteomics* 2003, **3**(4):363-379.
19. Partridge K A. JA, Tauler A., Pryme I F., Hesketh J E: **Competition between the signal sequence and a 3'UTR localisation signal during redirection of beta-globin mRNA to the endoplasmic reticulum:implications for biotechnology**. *Cytotechnology* 1999, **30**:37-47.
20. Shaffer KL, Sharma A, Snapp EL, Hegde RS: **Regulation of protein compartmentalization expands the diversity of protein function**. *Dev Cell* 2005, **9**(4):545-554.
21. Kalies KU, Hartmann E: **Protein translocation into the endoplasmic reticulum (ER)--two similar routes with different modes**. *European Journal of Biochemistry* 1998, **254**(1):1-5.
22. Pool MR: **Getting to the membrane: how is co-translational protein targeting to the endoplasmic reticulum regulated?** *Biochem Soc Trans* 2003, **31**(Pt 6):1232-1237.
23. De Gier JW, Valent QA, Von Heijne G, Luirink J: **The E. coli SRP: preferences of a targeting factor**. *FEBS Lett* 1997, **408**(1):1-4.
24. Blobel G, Dobberstein B: **Transfer of proteins across membranes. I. Presence of proteolytically processed and unprocessed nascent immunoglobulin light chains on membrane-bound ribosomes of murine myeloma**. *J Cell Biol* 1975, **67**(3):835-851.
25. Gierasch LM: **Signal sequences**. *Biochemistry* 1989, **28**(3):923-930.
26. Talmadge K, Stahl S, Gilbert W: **Eukaryotic signal sequence transports insulin antigen in Escherichia coli**. *Proc Natl Acad Sci U S A* 1980, **77**(6):3369-3373.
27. Muller M, Ibrahimi I, Chang CN, Walter P, Blobel G: **A bacterial secretory protein requires signal recognition particle for translocation across mammalian endoplasmic reticulum**. *J Biol Chem* 1982, **257**(20):11860-11863.
28. Kaiser CA, Preuss D, Grisafi P, Botstein D: **Many random sequences functionally replace the secretion signal sequence of yeast invertase**. *Science* 1987, **235**(4786):312-317.
29. von Heijne G: **Signal sequences. The limits of variation**. *Journal of Molecular Biology* 1985, **184**(1):99-105.
30. Rane NS, Chakrabarti O, Feigenbaum L, Hegde RS: **Signal sequence insufficiency contributes to neurodegeneration caused by transmembrane prion protein**. *J Cell Biol*, **188**(4):515-526.
31. Chen M, Nagarajan V: **Effect of alteration of charged residues at the N termini of signal peptides on protein export in Bacillus subtilis**. *J Bacteriol* 1994, **176**(18):5796-5801.
32. Peterson JH, Woolhead CA, Bernstein HD: **Basic amino acids in a distinct subset of signal peptides promote interaction with the signal recognition particle**. *J Biol Chem* 2003, **278**(46):46155-46162.

33. Sagiya Y, Yamagata H, Udaka S: **Direct high-level secretion into the culture medium of tuna growth hormone in biologically active form by Bacillus brevis**. *Appl Microbiol Biotechnol* 1994, **42**(2-3):358-363.
34. Hikita C, Mizushima S: **The requirement of a positive charge at the amino terminus can be compensated for by a longer central hydrophobic stretch in the functioning of signal peptides**. *J Biol Chem* 1992, **267**(17):12375-12379.
35. von Heijne G: **Life and death of a signal peptide.[comment]**. *Nature* 1998, **396**(6707):111.
36. Lety MA, Frehel C, Beretti JL, Berche P, Charbit A: **Modification of the signal sequence cleavage site of listeriolysin O does not affect protein secretion but impairs the virulence of Listeria monocytogenes**. *Microbiology* 2003, **149**(Pt 5):1249-1255.
37. Paetzel M, Dalbey RE, Strynadka NC: **The structure and mechanism of bacterial type I signal peptidases. A novel antibiotic target**. *Pharmacol Ther* 2000, **87**(1):27-49.
38. Kim SJ, Mitra D, Salerno JR, Hegde RS: **Signal sequences control gating of the protein translocation channel in a substrate-specific manner**. *Dev Cell* 2002, **2**(2):207-217.
39. Fons RD, Bogert BA, Hegde RS: **Substrate-specific function of the translocon-associated protein complex during translocation across the ER membrane**. *J Cell Biol* 2003, **160**(4):529-539.
40. Garrison JL, Kunkel EJ, Hegde RS, Taunton J: **A substrate-specific inhibitor of protein translocation into the endoplasmic reticulum**. *Nature* 2005, **436**(7048):285-289.
41. Zhang L, Leng Q, Mixson AJ: **Alteration in the IL-2 signal peptide affects secretion of proteins in vitro and in vivo**. *Journal of Gene Medicine* 2005, **7**(3):354-365.
42. Nilsson I, Whitley P, von Heijne G: **The COOH-terminal ends of internal signal and signal-anchor sequences are positioned differently in the ER translocase**. *J Cell Biol* 1994, **126**(5):1127-1132.
43. Morioka-Fujimoto K, Marumoto R, Fukuda T: **Modified enterotoxin signal sequences increase secretion level of the recombinant human epidermal growth factor in Escherichia coli**. *J Biol Chem* 1991, **266**(3):1728-1732.
44. Izard JW, Doughty MB, Kendall DA: **Physical and conformational properties of synthetic idealized signal sequences parallel their biological function**. *Biochemistry* 1995, **34**(31):9904-9912.
45. Izard JW, Rusch SL, Kendall DA: **The amino-terminal charge and core region hydrophobicity interdependently contribute to the function of signal sequences**. *J Biol Chem* 1996, **271**(35):21579-21582.
46. Martoglio B, Dobberstein B: **Signal sequences: more than just greasy peptides**. *Trends Cell Biol* 1998, **8**(10):410-415.
47. Li Y, Luo L, Thomas DY, Kang CY: **The HIV-1 Env protein signal sequence retards its cleavage and down-regulates the glycoprotein folding**. *Virology* 2000, **272**(2):417-428.
48. Li Y, Bergeron JJ, Luo L, Ou WJ, Thomas DY, Kang CY: **Effects of inefficient cleavage of the signal sequence of HIV-1 gp 120 on its association with calnexin,**

folding, and intracellular transport. *Proc Natl Acad Sci U S A* 1996, **93**(18):9606-9611.
49. Weihofen A, Binns K, Lemberg MK, Ashman K, Martoglio B: **Identification of signal peptide peptidase, a presenilin-type aspartic protease.[see comment]**. *Science* 2002, **296**(5576):2215-2218.
50. Martoglio B, Graf R, Dobberstein B: **Signal peptide fragments of preprolactin and HIV-1 p-gp160 interact with calmodulin**. *EMBO J* 1997, **16**(22):6636-6645.
51. Braud V, Jones EY, McMichael A: **The human major histocompatibility complex class Ib molecule HLA-E binds signal sequence-derived peptides with primary anchor residues at positions 2 and 9**. *Eur J Immunol* 1997, **27**(5):1164-1169.
52. Egea PF, Stroud RM, Walter P: **Targeting proteins to membranes: structure of the signal recognition particle**. *Curr Opin Struct Biol* 2005, **15**(2):213-220.
53. Halic M, Becker T, Pool MR, Spahn CMT, Grassucci RA, Frank J, Beckmann R: **Structure of the signal recognition particle interacting with the elongation-arrested ribosome**. *Nature* 2004, **427**(6977):808-814.
54. Walter P, Ibrahimi I, Blobel G: **Translocation of proteins across the endoplasmic reticulum. I. Signal recognition protein (SRP) binds to in-vitro-assembled polysomes synthesizing secretory protein**. *J Cell Biol* 1981, **91**(2 Pt 1):545-550.
55. Koch HG, Moser M, Muller M: **Signal recognition particle-dependent protein targeting, universal to all kingdoms of life**. *Rev Physiol Biochem Pharmacol* 2003, **146**:55-94.
56. Keenan RJ, Freymann DM, Stroud RM, Walter P: **The signal recognition particle**. *Annu Rev Biochem* 2001, **70**:755-775.
57. Powers T, Walter P: **Reciprocal stimulation of GTP hydrolysis by two directly interacting GTPases**. *Science* 1995, **269**(5229):1422-1424.
58. Zopf D, Bernstein HD, Johnson AE, Walter P: **The methionine-rich domain of the 54 kd protein subunit of the signal recognition particle contains an RNA binding site and can be crosslinked to a signal sequence**. *EMBO J* 1990, **9**(13):4511-4517.
59. Rapiejko PJ, Gilmore R: **Empty site forms of the SRP54 and SR alpha GTPases mediate targeting of ribosome-nascent chain complexes to the endoplasmic reticulum**. *Cell* 1997, **89**(5):703-713.
60. Walter P, Blobel G: **Translocation of proteins across the endoplasmic reticulum III. Signal recognition protein (SRP) causes signal sequence-dependent and site-specific arrest of chain elongation that is released by microsomal membranes**. *J Cell Biol* 1981, **91**(2 Pt 1):557-561.
61. Mason N, Ciufo LF, Brown JD: **Elongation arrest is a physiologically important function of signal recognition particle**. *EMBO J* 2000, **19**(15):4164-4174.
62. Siegel V, Walter P: **Elongation arrest is not a prerequisite for secretory protein translocation across the microsomal membrane**. *J Cell Biol* 1985, **100**(6):1913-1921.
63. Gundelfinger ED, Krause E, Melli M, Dobberstein B: **The organization of the 7SL RNA in the signal recognition particle**. *Nucleic Acids Res* 1983, **11**(21):7363-7374.
64. Yin J, Huang Q, Pakhomova ON, Hinck AP, Zwieb C: **The conserved adenosine in helix 6 of Archaeoglobus fulgidus signal recognition particle RNA initiates SRP assembly**. *Archaea* 2004, **1**(4):269-275.

65. Grosshans H, Deinert K, Hurt E, Simos G: **Biogenesis of the signal recognition particle (SRP) involves import of SRP proteins into the nucleolus, assembly with the SRP-RNA, and Xpo1p-mediated export**. *J Cell Biol* 2001, **153**(4):745-762.
66. Ciufo LF, Brown JD: **Nuclear export of yeast signal recognition particle lacking Srp54p by the Xpo1p/Crm1p NES-dependent pathway**. *Curr Biol* 2000, **10**(20):1256-1264.
67. Spahn CM, Gomez-Lorenzo MG, Grassucci RA, Jorgensen R, Andersen GR, Beckmann R, Penczek PA, Ballesta JP, Frank J: **Domain movements of elongation factor eEF2 and the eukaryotic 80S ribosome facilitate tRNA translocation**. *EMBO J* 2004, **23**(5):1008-1019.
68. Gomez-Lorenzo MG, Spahn CM, Agrawal RK, Grassucci RA, Penczek P, Chakraburtty K, Ballesta JP, Lavandera JL, Garcia-Bustos JF, Frank J: **Three-dimensional cryo-electron microscopy localization of EF2 in the Saccharomyces cerevisiae 80S ribosome at 17.5 A resolution**. *EMBO J* 2000, **19**(11):2710-2718.
69. Lakkaraju AKK, Mary C, Scherrer A, Johnson AE, Strub K: **SRP keeps polypeptides translocation-competent by slowing translation to match limiting ER-targeting sites**. *Cell* 2008, **133**(3):440-451.
70. Keenan RJ, Freymann DM, Walter P, Stroud RM: **Crystal structure of the signal sequence binding subunit of the signal recognition particle**. *Cell* 1998, **94**(2):181-191.
71. Freymann DM, Keenan RJ, Stroud RM, Walter P: **Functional changes in the structure of the SRP GTPase on binding GDP and Mg2+GDP**. *Nat Struct Biol* 1999, **6**(8):793-801.
72. Rapoport TA: **Protein translocation across the eukaryotic endoplasmic reticulum and bacterial plasma membranes**. *Nature* 2007, **450**(7170):663-669.
73. Walter P, Johnson AE: **Signal sequence recognition and protein targeting to the endoplasmic reticulum membrane**. *Annu Rev Cell Biol* 1994, **10**:87-119.
74. Or E, Rapoport T: **Cross-linked SecA dimers are not functional in protein translocation**. *FEBS Lett* 2007, **581**(14):2616-2620.
75. Or E, Boyd D, Gon S, Beckwith J, Rapoport T: **The bacterial ATPase SecA functions as a monomer in protein translocation**. *J Biol Chem* 2005, **280**(10):9097-9105.
76. Oliver J, Jungnickel B, Gorlich D, Rapoport T, High S: **The Sec61 complex is essential for the insertion of proteins into the membrane of the endoplasmic reticulum**. *FEBS Lett* 1995, **362**(2):126-130.
77. Zheng N, Gierasch LM: **Signal sequences: the same yet different**. *Cell* 1996, **86**(6):849-852.
78. Ng DT, Brown JD, Walter P: **Signal sequences specify the targeting route to the endoplasmic reticulum membrane**. *J Cell Biol* 1996, **134**(2):269-278.
79. Bogsch E, Brink S, Robinson C, Bogsch E, Brink S, Robinson C: **Pathway specificity for a delta pH-dependent precursor thylakoid lumen protein is governed by a 'Sec-avoidance' motif in the transfer peptide and a 'Sec-incompatible' mature protein**. *EMBO J* 1997, **16**(13):3851-3859.
80. Sambasivarao D, Turner RJ, Simala-Grant JL, Shaw G, Hu J, Weiner JH: **Multiple roles for the twin arginine leader sequence of dimethyl sulfoxide reductase of Escherichia coli**. *J Biol Chem* 2000, **275**(29):22526-22531.

81. Huber D, Boyd D, Xia Y, Olma MH, Gerstein M, Beckwith J: **Use of thioredoxin as a reporter to identify a subset of Escherichia coli signal sequences that promote signal recognition particle-dependent translocation**. *J Bacteriol* 2005, **187**(9):2983-2991.
82. Hegde RS, Bernstein HD: **The surprising complexity of signal sequences**. *Trends Biochem Sci* 2006, **31**(10):563-571.
83. Haigh NG, Johnson AE: **A new role for BiP: closing the aqueous translocon pore during protein integration into the ER membrane**. *J Cell Biol* 2002, **156**(2):261-270.
84. Hamman BD, Chen JC, Johnson EE, Johnson AE: **The aqueous pore through the translocon has a diameter of 40-60 A during cotranslational protein translocation at the ER membrane**. *Cell* 1997, **89**(4):535-544.
85. Beckmann R, Bubeck D, Grassucci R, Penczek P, Verschoor A, Blobel G, Frank J: **Alignment of conduits for the nascent polypeptide chain in the ribosome-Sec61 complex**. *Science* 1997, **278**(5346):2123-2126.
86. Crowley KS, Liao S, Worrell VE, Reinhart GD, Johnson AE: **Secretory proteins move through the endoplasmic reticulum membrane via an aqueous, gated pore**. *Cell* 1994, **78**(3):461-471.
87. Vitale A. DJ: **The Endoplasmic Reticulum-Gateway of the Secretory Pathway**. *The Plant Cells* 1999, **11**:615-628.
88. Lippincott-Schwartz J, Roberts TH, Hirschberg K: **Secretory protein trafficking and organelle dynamics in living cells**. *Annual Review of Cell & Developmental Biology* 2000, **16**:557-589.
89. McMaster CR: **Lipid metabolism and vesicle trafficking: more than just greasing the transport machinery**. *Biochem Cell Biol* 2001, **79**(6):681-692.
90. Ma Y, Hendershot LM: **The unfolding tale of the unfolded protein response**. *Cell* 2001, **107**(7):827-830.
91. Murshid A, Presley JF: **ER-to-Golgi transport and cytoskeletal interactions in animal cells**. *Cell Mol Life Sci* 2004, **61**(2):133-145.
92. Pfeffer SR: **Unsolved mysteries in membrane traffic**. *Annu Rev Biochem* 2007, **76**:629-645.
93. Pelham HR, Rothman JE: **The debate about transport in the Golgi--two sides of the same coin?** *Cell* 2000, **102**(6):713-719.
94. Lippincott-Schwartz J: **Dynamics of secretory membrane trafficking**. *Ann N Y Acad Sci* 2004, **1038**:115-124.
95. Gomord V, Wee E, Faye L: **Protein retention and localization in the endoplasmic reticulum and the golgi apparatus**. *Biochimie* 1999, **81**(6):607-618.
96. Lee S, Park B, Ahn K: **Determinant for endoplasmic reticulum retention in the luminal domain of the human cytomegalovirus US3 glycoprotein**. *J Virol* 2003, **77**(3):2147-2156.
97. Hendershot LM: **Giving protein traffic the green light**. *Nat Cell Biol* 2000, **2**(6):E105-106.
98. Cudna RE, Dickson AJ: **Endoplasmic reticulum signaling as a determinant of recombinant protein expression**. *Biotechnol Bioeng* 2003, **81**(1):56-65.
99. Malhotra JD, Kaufman RJ: **The endoplasmic reticulum and the unfolded protein response**. *Semin Cell Dev Biol* 2007, **18**(6):716-731.

100. Schroder M, Kaufman RJ: **The mammalian unfolded protein response**. *Annu Rev Biochem* 2005, **74**:739-789.
101. Hamanaka RB, Bennett BS, Cullinan SB, Diehl JA: **PERK and GCN2 contribute to eIF2alpha phosphorylation and cell cycle arrest after activation of the unfolded protein response pathway**. *Mol Biol Cell* 2005, **16**(12):5493-5501.
102. Baryshev M, Sargsyan E, Wallin G, Lejnieks A, Furudate S, Hishinuma A, Mkrtchian S: **Unfolded protein response is involved in the pathology of human congenital hypothyroid goiter and rat non-goitrous congenital hypothyroidism**. *J Mol Endocrinol* 2004, **32**(3):903-920.
103. Kaufman RJ: **Orchestrating the unfolded protein response in health and disease**. *J Clin Invest* 2002, **110**(10):1389-1398.
104. Wu J, Kaufman RJ: **From acute ER stress to physiological roles of the Unfolded Protein Response**. *Cell Death Differ* 2006, **13**(3):374-384.
105. May MJ, Ghosh S: **Signal transduction through NF-kappa B**. *Immunol Today* 1998, **19**(2):80-88.
106. Mignone F, Gissi C, Liuni S, Pesole G: **Untranslated regions of mRNAs**. *Genome Biology* 2002, **3**(3):REVIEWS0004.
107. Pesole G, Mignone F, Gissi C, Grillo G, Licciulli F, Liuni S: **Structural and functional features of eukaryotic mRNA untranslated regions**. *Gene* 2001, **276**(1-2):73-81.
108. Andreassi C, Riccio A: **To localize or not to localize: mRNA fate is in 3'UTR ends**. *Trends Cell Biol* 2009, **19**(9):465-474.
109. Moore MJ: **From birth to death: the complex lives of eukaryotic mRNAs**. *Science* 2005, **309**(5740):1514-1518.
110. Pickering BM, Willis AE: **The implications of structured 5' untranslated regions on translation and disease**. *Semin Cell Dev Biol* 2005, **16**(1):39-47.
111. Hughes MJ, Andrews DW: **A single nucleotide is a sufficient 5' untranslated region for translation in an eukaryotic in vitro system**. *FEBS Lett* 1997, **414**(1):19-22.
112. Pesole G, Grillo G, Larizza A, Liuni S: **The untranslated regions of eukaryotic mRNAs: structure, function, evolution and bioinformatic tools for their analysis**. *Briefings in Bioinformatics* 2000, **1**(3):236-249.
113. Hamilton RS, Davis I: **RNA localization signals: deciphering the message with bioinformatics**. *Semin Cell Dev Biol* 2007, **18**(2):178-185.
114. Crofts AJ, Washida H, Okita TW, Ogawa M, Kumamaru T, Satoh H: **Targeting of proteins to endoplasmic reticulum-derived compartments in plants. The importance of RNA localization**. *Plant Physiology* 2004, **136**(3):3414-3419.
115. Jansen RP: **mRNA localization: message on the move**. *Nat Rev Mol Cell Biol* 2001, **2**(4):247-256.
116. Du TG, Schmid M, Jansen RP: **Why cells move messages: the biological functions of mRNA localization**. *Semin Cell Dev Biol* 2007, **18**(2):171-177.
117. Cohen RS: **The role of membranes and membrane trafficking in RNA localization**. *Biol Cell* 2005, **97**(1):5-18.
118. St Johnston D: **Moving messages: the intracellular localization of mRNAs**. *Nat Rev Mol Cell Biol* 2005, **6**(5):363-375.
119. Bashirullah A, Cooperstock RL, Lipshitz HD: **RNA localization in development**. *Annu Rev Biochem* 1998, **67**:335-394.

120. Mhlanga MM, Bratu DP, Genovesio A, Rybarska A, Chenouard N, Nehrbass U, Olivo-Marin JC: **In vivo colocalisation of oskar mRNA and trans-acting proteins revealed by quantitative imaging of the Drosophila oocyte**. *PLoS ONE* 2009, **4**(7):e6241.
121. St Johnston D: **The intracellular localization of messenger RNAs**. *Cell* 1995, **81**(2):161-170.
122. Bermano G, Shepherd RK, Zehner ZE, Hesketh JE: **Perinuclear mRNA localisation by vimentin 3'-untranslated region requires a 100 nucleotide sequence and intermediate filaments**. *FEBS Lett* 2001, **497**(2-3):77-81.
123. Hesketh JE: **mRNA targeting: signals in the 3'-untranslated sequences for sorting of some mRNAs**. *Biochem Soc Trans* 1996, **24**(2):521-527.
124. Jambhekar A, Derisi JL: **Cis-acting determinants of asymmetric, cytoplasmic RNA transport**. *Rna* 2007, **13**(5):625-642.
125. Brennan CM, Steitz JA: **HuR and mRNA stability**. *Cell Mol Life Sci* 2001, **58**(2):266-277.
126. Jacobson A, Peltz SW: **Interrelationships of the pathways of mRNA decay and translation in eukaryotic cells**. *Annu Rev Biochem* 1996, **65**:693-739.
127. McClelland S, Shrivastava R, Medh JD: **Regulation of Translational Efficiency by Disparate 5' UTRs of PPARgamma Splice Variants**. *PPAR Res* 2009, **2009**:193413.
128. Jackson RJ, Hellen CU, Pestova TV: **The mechanism of eukaryotic translation initiation and principles of its regulation**. *Nat Rev Mol Cell Biol*, **11**(2):113-127.
129. Volz K: **The functional duality of iron regulatory protein 1**. *Curr Opin Struct Biol* 2008, **18**(1):106-111.
130. Zhang DL, Hughes RM, Ollivierre-Wilson H, Ghosh MC, Rouault TA: **A ferroportin transcript that lacks an iron-responsive element enables duodenal and erythroid precursor cells to evade translational repression**. *Cell Metab* 2009, **9**(5):461-473.
131. Kuersten S, Goodwin EB: **The power of the 3' UTR: translational control and development**. *Nat Rev Genet* 2003, **4**(8):626-637.
132. Groppo R, Richter JD: **Translational control from head to tail**. *Curr Opin Cell Biol* 2009, **21**(3):444-451.
133. Wilkie GS, Dickson KS, Gray NK: **Regulation of mRNA translation by 5'- and 3'-UTR-binding factors**. *Trends Biochem Sci* 2003, **28**(4):182-188.
134. Pierrat OA, Mikitova V, Bush MS, Browning KS, Doonan JH: **Control of protein translation by phosphorylation of the mRNA 5'-cap-binding complex**. *Biochem Soc Trans* 2007, **35**(Pt 6):1634-1637.
135. Algire MA, Lorsch JR: **Where to begin? The mechanism of translation initiation codon selection in eukaryotes**. *Curr Opin Chem Biol* 2006, **10**(5):480-486.
136. Xiong W, Hsieh CC, Kurtz AJ, Rabek JP, Papaconstantinou J: **Regulation of CCAAT/enhancer-binding protein-beta isoform synthesis by alternative translational initiation at multiple AUG start sites**. *Nucleic Acids Res* 2001, **29**(14):3087-3098.
137. Racine T, Barry C, Roy K, Dawe SJ, Shmulevitz M, Duncan R: **Leaky scanning and scanning-independent ribosome migration on the tricistronic S1 mRNA of avian reovirus**. *J Biol Chem* 2007, **282**(35):25613-25622.

138. Rogozin IB, Kochetov AV, Kondrashov FA, Koonin EV, Milanesi L: **Presence of ATG triplets in 5' untranslated regions of eukaryotic cDNAs correlates with a 'weak' context of the start codon**. *Bioinformatics* 2001, **17**(10):890-900.
139. Wegrzyn JL, Drudge TM, Valafar F, Hook V: **Bioinformatic analyses of mammalian 5'-UTR sequence properties of mRNAs predicts alternative translation initiation sites**. *BMC Bioinformatics* 2008, **9**:232.
140. Filbin ME, Kieft JS: **Toward a structural understanding of IRES RNA function**. *Curr Opin Struct Biol* 2009, **19**(3):267-276.
141. Jang SK, Krausslich HG, Nicklin MJ, Duke GM, Palmenberg AC, Wimmer E: **A segment of the 5' nontranslated region of encephalomyocarditis virus RNA directs internal entry of ribosomes during in vitro translation**. *J Virol* 1988, **62**(8):2636-2643.
142. Pelletier J, Sonenberg N: **Internal initiation of translation of eukaryotic mRNA directed by a sequence derived from poliovirus RNA**. *Nature* 1988, **334**(6180):320-325.
143. Baird SD, Turcotte M, Korneluk RG, Holcik M: **Searching for IRES**. *Rna* 2006, **12**(10):1755-1785.
144. Bonnal S, Boutonnet C, Prado-Lourenco L, Vagner S: **IRESdb: the Internal Ribosome Entry Site database**. *Nucleic Acids Res* 2003, **31**(1):427-428.
145. Komar AA, Hatzoglou M: **Internal ribosome entry sites in cellular mRNAs: mystery of their existence**. *J Biol Chem* 2005, **280**(25):23425-23428.
146. Winter J, Roepcke S, Krause S, Muller E-C, Otto A, Vingron M, Schweiger S: **Comparative 3'UTR analysis allows identification of regulatory clusters that drive Eph/ephrin expression in cancer cell lines**. *PLoS ONE* 2008, **3**(7):e2780.
147. Mazumder B, Seshadri V, Fox PL: **Translational control by the 3'-UTR: the ends specify the means**. *Trends Biochem Sci* 2003, **28**(2):91-98.
148. Poyry TA, Kaminski A, Jackson RJ: **What determines whether mammalian ribosomes resume scanning after translation of a short upstream open reading frame?** *Genes Dev* 2004, **18**(1):62-75.
149. Jiang C, Schuman EM: **Regulation and function of local protein synthesis in neuronal dendrites**. *Trends Biochem Sci* 2002, **27**(10):506-513.
150. Osborne HB: **An insight into the post-transcriptional control of gene expression in cell function**. *Biol Cell* 2003, **95**(3-4):125-127.
151. Kean KM: **The role of mRNA 5'-noncoding and 3'-end sequences on 40S ribosomal subunit recruitment, and how RNA viruses successfully compete with cellular mRNAs to ensure their own protein synthesis**. *Biol Cell* 2003, **95**(3-4):129-139.
152. Bradrick SS, Walters RW, Gromeier M: **The hepatitis C virus 3'-untranslated region or a poly(A) tract promote efficient translation subsequent to the initiation phase**. *Nucleic Acids Res* 2006, **34**(4):1293-1303.
153. Wakiyama M, Imataka H, Sonenberg N: **Interaction of eIF4G with poly(A)-binding protein stimulates translation and is critical for Xenopus oocyte maturation**. *Curr Biol* 2000, **10**(18):1147-1150.
154. Thoma C, Fraterman S, Gentzel M, Wilm M, Hentze MW: **Translation initiation by the c-myc mRNA internal ribosome entry sequence and the poly(A) tail**. *Rna* 2008, **14**(8):1579-1589.

155. Mendez R, Richter JD: **Translational control by CPEB: a means to the end**. *Nat Rev Mol Cell Biol* 2001, **2**(7):521-529.
156. Radford HE, Meijer HA, de Moor CH: **Translational control by cytoplasmic polyadenylation in Xenopus oocytes**. *Biochim Biophys Acta* 2008, **1779**(4):217-229.
157. Schmid M, Jensen TH: **The exosome: a multipurpose RNA-decay machine**. *Trends Biochem Sci* 2008, **33**(10):501-510.
158. Chen CY, Shyu AB: **AU-rich elements: characterization and importance in mRNA degradation**. *Trends Biochem Sci* 1995, **20**(11):465-470.
159. Bolognani F, Perrone-Bizzozero NI: **RNA-protein interactions and control of mRNA stability in neurons**. *J Neurosci Res* 2008, **86**(3):481-489.
160. Bakheet T, Frevel M, Williams BR, Greer W, Khabar KS: **ARED: human AU-rich element-containing mRNA database reveals an unexpectedly diverse functional repertoire of encoded proteins**. *Nucleic Acids Res* 2001, **29**(1):246-254.
161. Bakheet T, Williams BR, Khabar KS: **ARED 3.0: the large and diverse AU-rich transcriptome**. *Nucleic Acids Res* 2006, **34**(Database issue):D111-114.
162. Barreau C, Paillard L, Osborne HB: **AU-rich elements and associated factors: are there unifying principles?** *Nucleic Acids Res* 2005, **33**(22):7138-7150.
163. Peng SS, Chen CY, Shyu AB: **Functional characterization of a non-AUUUA AU-rich element from the c-jun proto-oncogene mRNA: evidence for a novel class of AU-rich elements**. *Mol Cell Biol* 1996, **16**(4):1490-1499.
164. Gingerich TJ, Feige J-J, LaMarre J: **AU-rich elements and the control of gene expression through regulated mRNA stability**. *Anim Health Res Rev* 2004, **5**(1):49-63.
165. Paillard L, Legagneux V, Maniey D, Osborne HB: **c-Jun ARE targets mRNA deadenylation by an EDEN-BP (embryo deadenylation element-binding protein)-dependent pathway**. *J Biol Chem* 2002, **277**(5):3232-3235.
166. Ghosh M, Aguila HL, Michaud J, Ai Y, Wu M-T, Hemmes A, Ristimaki A, Guo C, Furneaux H, Hla T: **Essential role of the RNA-binding protein HuR in progenitor cell survival in mice**. *J Clin Invest* 2009, **119**(12):3530-3543.
167. Glisovic T, Bachorik JL, Yong J, Dreyfuss G: **RNA-binding proteins and post-transcriptional gene regulation**. *FEBS Lett* 2008, **582**(14):1977-1986.
168. Abdelmohsen K, Kuwano Y, Kim HH, Gorospe M: **Posttranscriptional gene regulation by RNA-binding proteins during oxidative stress: implications for cellular senescence**. *Biol Chem* 2008, **389**(3):243-255.
169. Lee MH, Schedl T: **RNA-binding proteins**. *Wormbook* 2006:1-13.
170. Lunde BM, Moore C, Varani G: **RNA-binding proteins: modular design for efficient function**. *Nat Rev Mol Cell Biol* 2007, **8**(6):479-490.
171. Maris C, Dominguez C, Allain FHT: **The RNA recognition motif, a plastic RNA-binding platform to regulate post-transcriptional gene expression**. *FEBS J* 2005, **272**(9):2118-2131.
172. Lopez de Silanes I, Zhan M, Lal A, Yang X, Gorospe M: **Identification of a target RNA motif for RNA-binding protein HuR**. *Proc Natl Acad Sci U S A* 2004, **101**(9):2987-2992.
173. Lu JY, Schneider RJ: **Tissue distribution of AU-rich mRNA-binding proteins involved in regulation of mRNA decay**. *J Biol Chem* 2004, **279**(13):12974-12979.

174. Gherzi R, Lee KY, Briata P, Wegmuller D, Moroni C, Karin M, Chen CY: **A KH domain RNA binding protein, KSRP, promotes ARE-directed mRNA turnover by recruiting the degradation machinery**. *Mol Cell* 2004, **14**(5):571-583.
175. Landers SM, Gallas MR, Little J, Long RM: **She3p possesses a novel activity required for ASH1 mRNA localization in Saccharomyces cerevisiae**. *Eukaryot Cell* 2009, **8**(7):1072-1083.
176. Gonsalvez GB, Urbinati CR, Long RM: **RNA localization in yeast: moving towards a mechanism**. *Biol Cell* 2005, **97**(1):75-86.
177. Carmody SR, Wente SR: **mRNA nuclear export at a glance**. *J Cell Sci* 2009, **122**(Pt 12):1933-1937.
178. Mazan-Mamczarz K, Lal A, Martindale JL, Kawai T, Gorospe M: **Translational repression by RNA-binding protein TIAR**. *Mol Cell Biol* 2006, **26**(7):2716-2727.
179. Anderson P, Kedersha N: **Visibly stressed: the role of eIF2, TIA-1, and stress granules in protein translation**. *Cell Stress Chaperones* 2002, **7**(2):213-221.
180. Bowlby MR, and J.F. Case: **Flash kinetics and spatial patterns of bioluminescence in the copepodGaussia princeps** *Marine Biology* 1991, **110**:329-336.
181. Tannous BA, Kim DE, Fernandez JL, Weissleder R, Breakefield XO: **Codon-optimized gaussia luciferase cDNA for mammalian gene expression in culture and in vivo**. *Molecular Therapy* 2005, **11**(3):435-443.
182. Verhaegen M, Christopoulos TK: **Bacterial expression of in vivo-biotinylated aequorin for direct application to bioluminometric hybridization assays**. *Anal Biochem* 2002, **306**(2):314-322.
183. Wittwer CT, Ririe KM, Andrew RV, David DA, Gundry RA, Balis UJ: **The LightCycler: a microvolume multisample fluorimeter with rapid temperature control**. *Biotechniques* 1997, **22**(1):176-181.
184. Wittwer CT, Herrmann MG, Moss AA, Rasmussen RP: **Continuous fluorescence monitoring of rapid cycle DNA amplification**. *Biotechniques*, **22**(1):130-131.
185. Chen GJ, Qiu N, Karrer C, Caspers P, Page MG: **Restriction site-free insertion of PCR products directionally into vectors**. *Biotechniques* 2000, **28**(3):498-500.
186. Lu Q: **Seamless cloning and gene fusion**. *Trends Biotechnol* 2005, **23**(4):199-207.
187. Knappskog S, Ravneberg H, Gjerdrum C, Trosse C, Stern B, Pryme IF: **The level of synthesis and secretion of Gaussia princeps luciferase in transfected CHO cells is heavily dependent on the choice of signal peptide**. *Journal of Biotechnology* 2007, **128**(4):705-715.
188. Olsen LC: **The influence of native and vector-derived sequences in transcripts on mRNA targeting and levels of synthesis and secretion of Gaussia Luciferase.** *Master Thesis* 2006.
189. Eisenberg D, Weiss, R. M.,, Teruiliger TC, Wilcox, W: **Hydrophobic Moments and Protein Structure**. *Faraday Symp Chem Soc* 1982, **17**:109-120.
190. Mangone M, Macmenamin P, Zegar C, Piano F, Gunsalus KC: **UTRome.org: a platform for 3'UTR biology in C. elegans**. *Nucleic Acids Res* 2008, **36**(Database issue):D57-62.
191. Misquitta CM, Iyer VR, Werstiuk ES, Grover AK: **The role of 3'-untranslated region (3'-UTR) mediated mRNA stability in cardiovascular pathophysiology**. *Mol Cell Biochem* 2001, **224**(1-2):53-67.

192. Maret D, Boffa MB, Brien DF, Nesheim ME, Koschinsky ML: **Role of mRNA transcript stability in modulation of expression of the gene encoding thrombin activable fibrinolysis inhibitor**. *J Thromb Haemost* 2004, **2**(11):1969-1979.
193. Noe G, De Gaudenzi JG, Frasch AC: **Functionally related transcripts have common RNA motifs for specific RNA-binding proteins in trypanosomes**. *BMC Mol Biol* 2008, **9**:107.
194. Pullmann R, Jr., Kim HH, Abdelmohsen K, Lal A, Martindale JL, Yang X, Gorospe M: **Analysis of turnover and translation regulatory RNA-binding protein expression through binding to cognate mRNAs**. *Mol Cell Biol* 2007, **27**(18):6265-6278.
195. Holmberg A, Blomstergren A, Nord O, Lukacs M, Lundeberg J, Uhlen M: **The biotin-streptavidin interaction can be reversibly broken using water at elevated temperatures**. *Electrophoresis* 2005, **26**(3):501-510.
196. Weber A, Kristiansen I, Johannsen M, Oelrich B, Scholmann K, Gunia S, May M, Meyer HA, Behnke S, Moch H *et al*: **The FUSE binding proteins FBP1 and FBP3 are potential c-myc regulators in renal, but not in prostate and bladder cancer**. *BMC Cancer* 2008, **8**:369.
197. Yamamoto H, Tsukahara K, Kanaoka Y, Jinno S, Okayama H: **Isolation of a mammalian homologue of a fission yeast differentiation regulator**. *Mol Cell Biol* 1999, **19**(5):3829-3841.
198. Mori D, Sasagawa N, Kino Y, Ishiura S: **Quantitative analysis of CUG-BP1 binding to RNA repeats**. *Journal of Biochemistry* 2008, **143**(3):377-383.
199. Barreau C, Paillard L, Mereau A, Osborne HB: **Mammalian CELF/Bruno-like RNA-binding proteins: molecular characteristics and biological functions**. *Biochimie* 2006, **88**(5):515-525.
200. Good PJ, Chen Q, Warner SJ, Herring DC: **A family of human RNA-binding proteins related to the Drosophila Bruno translational regulator**. *J Biol Chem* 2000, **275**(37):28583-28592.
201. Li D, Bachinski LL, Roberts R: **Genomic organization and isoform-specific tissue expression of human NAPOR (CUGBP2) as a candidate gene for familial arrhythmogenic right ventricular dysplasia**. *Genomics* 2001, **74**(3):396-401.
202. Vlasova IA, Tahoe NM, Fan D, Larsson O, Rattenbacher B, Sternjohn JR, Vasdewani J, Karypis G, Reilly CS, Bitterman PB *et al*: **Conserved GU-rich elements mediate mRNA decay by binding to CUG-binding protein 1.[see comment]**. *Mol Cell* 2008, **29**(2):263-270.
203. Ladd AN, Charlet N, Cooper TA: **The CELF family of RNA binding proteins is implicated in cell-specific and developmentally regulated alternative splicing**. *Mol Cell Biol* 2001, **21**(4):1285-1296.
204. Singh G, Charlet BN, Han J, Cooper TA: **ETR-3 and CELF4 protein domains required for RNA binding and splicing activity in vivo**. *Nucleic Acids Res* 2004, **32**(3):1232-1241.
205. Charlet BN, Savkur RS, Singh G, Philips AV, Grice EA, Cooper TA: **Loss of the muscle-specific chloride channel in type 1 myotonic dystrophy due to misregulated alternative splicing**. *Mol Cell* 2002, **10**(1):45-53.

206. Philips AV, Timchenko LT, Cooper TA: **Disruption of splicing regulated by a CUG-binding protein in myotonic dystrophy.[see comment]**. *Science* 1998, **280**(5364):737-741.
207. Timchenko NA, Iakova P, Cai ZJ, Smith JR, Timchenko LT: **Molecular basis for impaired muscle differentiation in myotonic dystrophy**. *Mol Cell Biol* 2001, **21**(20):6927-6938.
208. Timchenko NA, Welm AL, Lu X, Timchenko LT: **CUG repeat binding protein (CUGBP1) interacts with the 5' region of C/EBPbeta mRNA and regulates translation of C/EBPbeta isoforms**. *Nucleic Acids Res* 1999, **27**(22):4517-4525.
209. Paillard L, Legagneux V, Beverley Osborne H: **A functional deadenylation assay identifies human CUG-BP as a deadenylation factor**. *Biol Cell* 2003, **95**(2):107-113.
210. Moraes KC, Wilusz CJ, Wilusz J: **CUG-BP binds to RNA substrates and recruits PARN deadenylase**. *Rna* 2006, **12**(6):1084-1091.
211. Timchenko LT, Miller JW, Timchenko NA, DeVore DR, Datar KV, Lin L, Roberts R, Caskey CT, Swanson MS: **Identification of a (CUG)n triplet repeat RNA-binding protein and its expression in myotonic dystrophy**. *Nucleic Acids Res* 1996, **24**(22):4407-4414.
212. Takahashi N, Sasagawa N, Suzuki K, Ishiura S: **The CUG-binding protein binds specifically to UG dinucleotide repeats in a yeast three-hybrid system**. *Biochem Biophys Res Commun* 2000, **277**(2):518-523.
213. Kino Y, Mori D, Oma Y, Takeshita Y, Sasagawa N, Ishiura S: **Muscleblind protein, MBNL1/EXP, binds specifically to CHHG repeats**. *Hum Mol Genet* 2004, **13**(5):495-507.
214. Johannes G, Carter MS, Eisen MB, Brown PO, Sarnow P: **Identification of eukaryotic mRNAs that are translated at reduced cap binding complex eIF4F concentrations using a cDNA microarray**. *Proc Natl Acad Sci U S A* 1999, **96**(23):13118-13123.