

# Improving mammalian cell factories: The selection of signal peptide has a major impact on recombinant protein synthesis and secretion in mammalian cells

Beate Stern<sup>1,2</sup>, Lene C. Olsen<sup>1</sup>, Christiane Tröbe<sup>2</sup>, Hanne Ravneberg<sup>2</sup> and Ian F. Pryme<sup>1,2,\*</sup>

<sup>1</sup>Department of Biomedicine, University of Bergen, Jonas Lies vei 91, N-5009 Bergen, Norway.

<sup>2</sup>UniTargetingResearch AS, Thormøhlensgate 55, N-5008 Bergen, Norway

## ABSTRACT

Although mammalian cell factories are now widely used in production of biomedically important proteins they are still relatively ineffective compared with those from prokaryotic organisms. In an attempt to improve the CHO cell line we have used a biotechnological approach to investigate whether or not the choice of signal peptide may have a major impact on the synthesis/secretion of a recombinant protein (*Gaussia* luciferase) in transfected cells. In experiments where the native luciferase signal peptide was replaced with that derived from a series of sources the amount of luciferase produced varied by a factor of more than 100. The most effective signal peptide being that derived from *Gaussia* luciferase. Surprisingly two signal peptides of mammalian origin, human interleukin-2 and albumin, proved to be far inferior to the signal peptide derived from the marine organism. A reduction in level of mRNA available for translation proved not to be responsible for the major decrease observed in luciferase production. When tested in a hepatic cell line (HepG2) the marine luciferase signal peptide again proved to be superior to that from albumin. The *Gaussia* luciferase signal peptide was also much more effective than that derived from *Vargula* luciferase in production of *Vargula* luciferase. This

demonstrates that a native signal peptide is not necessarily always the most effective. The signal peptide from *Metridia*, seven oikosin signal peptides derived from *Oikopleura* and two prokaryotic signal peptides all proved to be far inferior to the *Gaussia* luciferase signal peptide with respect to recombinant luciferase production in CHO cells. Overall the results described here indicate the vital importance of selection of the signal peptide when aiming to produce maximal amounts of recombinant protein in a mammalian expression system.

**KEYWORDS** : CHO cells, *Gaussia princeps*, *Oikopleura dioica*, *Vargula hilgendorffii*, *Metridia longa*, signal sequence, oikosins, luciferase

## INTRODUCTION

During the last two decades the number of recombinant proteins for therapeutic and research purposes on the market has increased substantially, and a further dramatic growth in the demand for protein and peptide drugs and specific antibodies is expected within the next 10-20 years [1, 2].

Prokaryotic cells were originally the most commonly used host for recombinant protein production, since they are easy to work with and divide rapidly, making it possible to produce large amounts of protein at low costs. However, when eukaryotic proteins are produced in prokaryotes, they might lack the proper folding and the post-translational modifications (PTMs) necessary for

---

\*Corresponding author  
ian.pryme@biomed.uib.no

their function [3]. Another problem with production of proteins in prokaryotes, or even simple eukaryotes such as yeast, is the possibility of allergic reactions in humans if the proteins are to be utilised for medicinal purposes. Therapeutic use of proteins purified from bacteria can also lead to a build-up of endotoxins and other possibly harmful substances [4].

To obtain fully functional and therapeutically safe proteins, production in mammalian cells is therefore often imperative. Mammalian cells are much more advanced than prokaryotes; they have the appropriate organelles containing enzymes to perform PTMs (e.g. glycosylation and acetylation) and chaperones to ensure proper folding. Production in mammalian cells results in recombinant proteins with biophysical, biochemical, and therefore also the functional properties that are comparable to those of their natural counterparts. For these reasons, 60-70% of all recombinant, therapeutic proteins are currently being produced in mammalian cells [5], with Chinese Hamster Ovary (CHO) cells being one of the most commonly used cell types [2].

Mammalian cells are much more demanding and expensive to cultivate than bacteria. While a bacterial cell can divide in 20 minutes, it may take 24 hours or more for a mammalian cell to complete mitosis. Mammalian cells are also much more sensitive to contamination and to growth conditions, requiring expensive growth medium and incubators. To ease the use and minimize the expense of mammalian cell cultures, it would be desirable if each cell in a population could produce as much of the protein of interest as possible, and for it to be efficiently secreted into the medium for convenient collection and purification.

In the past few years a lot of work has been put into increasing the production of protein from mammalian cells, both with regards to optimising growth conditions as well as to using genetic engineering to modify the vectors and specific features of the host cells themselves [2]. However, more work is still needed, particularly when it comes to optimising vectors especially with respect to post-transcriptional phenomena.

An overview of the various stages of protein production in mammalian cells, with a focus on secreted proteins, is presented. Relevant regulatory

elements involved at each level being highlighted. A brief section concerns the reporter protein used, a naturally secreted luciferase from the copepod *Gaussia princeps*. The final section reviews the work performed to study the effects of signal peptides from various sources on production of recombinant proteins in transfected CHO cells.

## OVERVIEW

### Improvement of mammalian cell factories through vector optimisation

Modifying the vector (both the backbone and the insert containing the gene of interest) can greatly influence the total amount of protein produced from each cell. Most attention has so far been directed to the promoter and its associated elements in order to increase transcription levels. The promoters used are generally derived from the genes of animal viruses or highly expressed mammalian genes, with the viral CMV and SV40 promoters being the most commonly used.

An example of another technological advancement is the addition of chromatin modulator elements, such as insulator or matrix attachment region (MAR) elements, in the vector. MAR elements are AT-rich DNA segments which act as attachment points of chromatin loops to the nuclear matrix. Each loop has independent transcription activity, and the MAR element is thought to insulate the gene(s) in the loop from the influence of the surrounding chromatin. Including such elements in a vector usually results in increased transcription of the gene of interest, probably because the insulation effect will reduce the effect of the transgene's positioning in the genome of the transfected host, and because of MAR's ability to inflict topological changes on the chromatin [6-9].

The sequences in the transcript generated from the transgene could also be important for the level of protein produced from a transfected cell. With increasing knowledge about regulatory elements within the 5' and the 3' untranslated regions (UTR) of the transcripts with regards to mRNA, targeting, stability and translatability [10-13], it is clear that these sequences need to be considered when generating a DNA construct for protein expression [14].

### **Regulation of transcription and translation**

A cell requires different concentrations of its various proteins for proper function. Some proteins are constantly required, such as actin, whereas others are needed only in short bursts, e.g. regulatory proteins such as c-myc. To finely tune the levels of proteins a cell has various ways of regulating the quantity of the different transcripts and the proteins they encode in most of the steps undertaken until the release of the polypeptide from the polysome. Proteins for secretion are transported through the endoplasmic reticulum (ER) and Golgi apparatus where they may be subject to PTMs before release from the cell. To obtain a maximal output of a protein from a cell, as is desirable for an efficient mammalian cell factory, ideally each of the following processes should be optimised: transcriptional control, RNA processing control (capping, splicing, cleavage, polyadenylation), mRNA export from the nucleus, transport and targeting of mRNA, control of mRNA stability and translational control. In the following sections the current knowledge of these steps that influence protein production is outlined.

### **Transcription and processing of the transcript**

A widely used form of gene regulation in eukaryotes is regulation of the initiation of transcription. This is initially dependent on the opening of the chromatin structure and the creation of a favorable conformation of the DNA by chromatin remodeling factors so the transcription factors (TFs) and the RNA polymerase II (RNAP II) can assemble at the promoter of the gene to be transcribed [15].

The RNAP II promoters are complex, being involved in regulating the initiation and efficiency of the transcription process. In addition to the core promoter, a gene will have regulatory elements upstream or downstream from the promoter which can influence the establishment and maintenance of transcription, such as activators, enhancers, repressors and silencers [16-18]. A strong, constitutive promoter will lead to high levels of the primary transcript (pre-mRNA), and a larger population of processed transcripts (mRNA) could ensure higher levels of available templates for translation.

Transcription of eukaryotic genes begins with the recognition and binding of core promoter elements.

Many genes have a TATA box (consensus TATAWAWR, where W is A or T, and R is a purine) [19] and/or an initiator element (Inr, consensus YYA<sub>+1</sub>NWYY, where Y is a pyrimidine, W is A or T, and N any nucleotide) [16]. The TATA box is generally located between 28 and 33 base pairs upstream of the transcription initiation site [20], and is bound by the TATA-binding protein (TBP) [19]. The binding of the TBP promotes the binding of TFs and RNAP II, forming the pre-initiation complex (PIC) [21]. The Inr and the TATA box function in synergy when they are both present in a promoter [22].

Transcription initiates with formation of the first phosphodiester bond of the transcript. The switch to elongation is a rate-limiting step, and certain promoter properties [15], as well as interactions between RNAP II and the initially transcribed sequences and/or the template [20, 23, 24], can ease the transition by providing a conformation favourable for progression. Elongation is assisted by the addition of elongation factors to the transcription complex [15], and pauses in the elongation process contribute to the regulation of the rate of transcription [25].

The processes from transcription to maturation and export of mRNA out of the nucleus are very much coupled. The proteins involved in the different transcript processing reactions often influence other steps in the generation of a mature mRNP particle. The most important factor linking these processes is the C-terminal domain of the RNAP II, which provides a platform for recruiting RNA-processing factors during transcription.

After about 17-22 nucleotides have been transcribed, a methylated guanosine cap is added to the 5' end of the nascent transcript. The cap is thought to protect the RNA from degradation by 5'-3'-exonucleases [26], and it is required for efficient splicing and 3' formation of the mRNA [27]. In the cytosol the cap is essential for the translation of most mRNAs.

Another co-transcriptional event is the start of splicing, involving the excision of intervening introns by a complex aggregate of ribonuclear proteins, known as the spliceosome. The exonic segments are then ligated together, yielding a

translatable mRNA molecule. Splicing is also interlinked with, and affects other transcriptional events. Spliceosomal U-snRNPs can for example enhance elongation, and splice sites close to the promoter will enhance the loading of the polymerase onto the gene [27-29]. Although the presence of excisable introns is not essential for the recombinant expression of most human genes, experiments have shown that in general, mRNAs generated by splicing will produce a higher level of the protein encoded than cDNA-derived mRNAs. One study showed that the presence of an intron enhanced gene expression by 6.3 fold on average, with a range of 1.5 to over 20 fold increase [30]. This could be because splicing causes the deposition of the exon junction complex (EJC) upstream of the splice site, which influences the efficiency of the 3' end formation of the mRNA, mRNA stability, and even the efficiency of translation [31]. There have also been reports of transcriptional enhancers and nucleosome-positioning elements present in introns, which can affect the efficiency of the initiation of transcription [32].

The final steps before the transcript can be released, are cleavage of the transcript, termination of transcription, and addition of a poly(A)-tail to the transcript. The cleavage site lies ~10-25 bases downstream of a highly conserved poly(A)-signal (consensus AAU AAA), and ~15-30 bases upstream of a GU/U-rich element [33]. Other elements that either enhance or repress the 3' end processing also exist in many genes [33, 34]. If the conformation around the poly(A)-signal is unstructured, this will facilitate the binding of the CPSF/CstF factors controlling the cleavage and termination and increase the 3' end processing [33, 35]. The recognition of the poly(A)-site is inhibited by any proximal splice site [26]. Kim *et al.* [36] found that the expression levels of a reporter protein improved 3-4 fold by changing the sequence downstream of the poly(A)-site.

Yeung *et al.* [37] characterised two types of transcription termination – poly(A)-assisted and poly(A)-driven termination. In poly(A)-assisted termination, e.g. in the SV40 and mouse  $\beta^{\text{maj}}$ -globin genes, RNAP II has shown no detectable decrease in processivity until encountering special termination

elements. These can in some instances be several kilobases downstream of the poly(A)-site. However, in addition to these termination elements, the upstream poly(A)-signal is also required. In contrast, for the poly(A) driven termination, e.g. in the chicken  $\beta^{\text{H}}$ - gene, the poly(A)-signal not only assists termination; it also revokes the processivity of the polymerase directly downstream of the poly(A)-site, almost regardless of which downstream sequence is used. To achieve release from the RNAP II complex, the transcript needs a poly(A)-tail [38].

If the cleavage and polyadenylation process is disrupted by *cis*- or *trans*-acting mutations, there will be an accumulation of transcripts in the nucleus. This proves that polyadenylation is a required step before an mRNA can be exported out of the nucleus [39].

### Export of mRNA from the nucleus

All mature mRNA molecules are transported from the nucleus to the cytosol for translation through the nuclear pore complex (NPC) in the nuclear envelope in a highly regulated process [40]. One of the proteins in the EJC, Aly, will recruit the Tap-Nxt heterodimer which targets the spliced mRNA to the NPC [30]. If the RNA molecule is incompletely spliced, it will be retained in the nucleus by the Mlp1p protein of the NPC [41]. mRNAs that have not undergone splicing (mRNAs transcribed from genes lacking introns and mRNAs derived from cDNA) seem to be recognised by distinct export factors [30]. However, the deposition of the EJC seems to enhance the export of at least some mRNAs [42].

Throughout transcription and mRNA maturation, proteins are deposited on the pre-mRNA, and together they make up the mRNP particle. The heteronuclear ribonuclear proteins (hnRNPs) in this particle change during these processes, and constantly undergo quality control. The deposition of certain hnRNPs will determine the export pathway of the mRNP, and this could eventually influence the translation of the mRNA [30]. Other hnRNPs, i.e. those making up the EJC, are involved in the cytoplasmic targeting of the mRNA [43]. The hnRNPs therefore do not only control many events in the nucleus, but some also control the fate of an mRNA in the cytoplasm [44, 45].

### Targeting of mRNA to polysome populations

Once the mRNA has reached the cytosol, it needs to be targeted to the appropriate polysome population for translation. Depending on the final destination and use of the protein to be produced, the mRNA is targeted to one of three general classes of polysomes, namely free polysomes, cytoskeletal-bound polysomes (CBP) and membrane-bound polysomes (MBP), which are associated with the ER [46-50]. Membrane proteins and proteins for secretion are synthesised on the latter class of polysomes. It is believed that it is more energy efficient to transport an mRNA molecule that can be translated many times at a specific location instead of transporting all the proteins produced elsewhere in the cell to the desired destination.

The localisation of the mRNP is an active process, and cytoskeletal motor proteins such as myosins, dyneins and kinesins are used to transport a large fraction of the mRNPs in the cytoplasm [51]. The mRNA does not interact directly with the cytoskeleton, but through proteins present in the RNP particle. After the mRNP has passed through the NPC and reached the cytosol, most of the shuttling hnRNP proteins are replaced with cytosolic mRNA-binding proteins and ribosomes. However, a subset of the hnRNPs will stay attached to the mRNA, and may help target the mRNA to its destination in the cell as well as regulate the translation and the stability of the mRNA. The first step in producing a transport particle containing the mRNA is therefore believed to occur in the nucleus [52-54]. Many mRNAs are believed to be in a translationally repressed state during transport [51]. When the mRNP arrives at its destination, auxiliary proteins will help anchor the mRNA on the ribosome such that translation can begin.

The types of *trans*-acting factors that are recruited to the RNP particles and participate in the localisation of the mRNA in the cytosol depends on *cis*-acting elements in the 3' UTR of the mRNA molecule. The *cis*-acting elements are unique in the different mRNAs, and the primary, and the secondary sequence (in particular RNA stem loop structures) and tertiary structures all appear to be important for the binding of the *trans*-acting factors [51, 55]. The *trans*-acting factors help fold

the mRNA into a proper configuration so auxiliary proteins can bind to the complex [53].

Partridge *et al.* [56] observed that there is competition between the localisation elements in the 3' UTR and the signal peptide. When the coding region of the signal peptide of human albumin, a naturally secreted protein, was fused to the coding region of  $\beta$ -globin, the mRNA was directed to MBP. However, when the 3' UTR of c-myc, an intracellular protein, whose transcript is known to be targeted to CBP [57], was coupled to the transcript, the extent of the redirection was decreased, and more of that species of mRNA was directed to CBP. This observation showed the importance of the 3' UTR with respect to targeting of the transcript with the nascent signal peptide to the ER.

### Stability of the mRNA molecule

The concentration of a particular mRNA species depends on both its rate of synthesis and its rate of degradation [58]. If the rate of degradation is low for an mRNA molecule, it allows for build-up of that species in the cell, and for each molecule to go through many rounds of translation. Therefore, the stability of an mRNA molecule can also influence the amount of protein a cell produces. Some proteins need to be produced in short bursts, and the main mechanism the cell uses for such proteins is for its corresponding transcripts to be unstable. Such mRNAs will often contain multiple and sometimes overlapping copies of the consensus sequence AUUUA in their 3' UTR, which function as instability elements. The presence of these AU-rich elements (ARE) causes the poly(A)-tail to degrade more rapidly, probably through mediation of the exosome, and they can influence translational efficiency [11, 59, 60]. In addition to the AREs, elements having a destabilising effect on the mRNA have been found in the 5' UTR, the coding sequence, and 3' UTR [58]. Elements stabilising an mRNA molecule also exist, and have been found in the 3' UTRs of transcripts encoding proteins required constantly and in high quantities, such as  $\beta$ -globin, the transferrin receptor, histones, and cytokines [61].

In some instances, general mRNA degradation with stability only at certain locations has been shown to cause local accumulation of an mRNA.

These mRNAs contain *cis*-acting degradation elements in their 3' UTR, but at their proper location there are binding factors that protect the mRNAs from being degraded [53, 54].

## Translation

Translational regulation is an important step in controlling protein production. The initiation step in particular is a key determinant for the efficiency of mRNA translation. Most of the control elements for initiation are found within the UTRs of the transcript, but features all along the mRNA can affect translational efficiency [11].

A vast majority of mRNAs are translated by a cap-dependent translation mechanism. The 5'<sup>m</sup>7G(5')ppp(5')N cap of the mRNA is bound by eukaryotic initiation factor 4F (eIF4F), which consists of subunits 4E, 4A, and 4G [62].

The 43S pre-initiation complex forms next, consisting of the 40S ribosomal subunit, eIF2-GTP, and Met-tRNA. The 43S complex scans the 5' UTR for the start codon, and upon binding to the start AUG (sAUG), the 60S ribosomal subunit can join the complex, and translation can begin [63].

The sequence around the sAUG codon has been found to be important to help the scanning ribosomal subunit pause at the start site so proper codon-anticodon recognition can take place. The optimal sequence varies between species, with the consensus sequence for higher vertebrates being GCCRCCaugG. The purine (usually A) at position -3 and the G at +4 are the most crucial bases [64, 65], and mutations in these positions can cause disease [66]. A stable stem-loop structure near the sAUG will block the access, prohibiting translation from starting at all.

The scanning mechanism can be impeded, but not completely precluded, by complex secondary structure in a long 5' UTR sequence [67]. The 40S migratory complex will then become trapped on the 5' side of a secondary structure [65]. Most likely because of this, human 5' UTRs are fairly short on average, usually between 100 and 200 nucleotides [11, 68]. Secondary structure near the cap blocks the access of eIF4F, and is therefore the most inhibitory feature [62, 67]. If there are hairpin structures close to the 5' end, the access of

the eIF4F to the cap will be blocked, resulting in lower levels of translation.

The presence of AUGs upstream (uAUGs) of the start site can also affect the rate of translation in a negative way. Meijer and Thomas [12] estimated that 20-48% of human mRNAs contain at least one uAUG. These are often found in mRNAs which encode regulatory proteins [69]. If an upstream open reading frame (uORF) is present, it will be translated, and translation will have to be reinitiated at the sAUG. This reinitiation mode is very inefficient, but it improves as the distance increases between the stop codon and the re-start site [67]. A cell can therefore use this mechanism to direct how many protein molecules will be produced from each mRNA. Regulatory proteins, such as growth factors, transcription factors, and proto-oncogenes often have structured 5' UTRs and uAUGs to prevent their over-expression, which would be harmful to the cell [70].

It has been claimed that the 3' UTR participates in the regulation of translation as well [13, 63], but this is a matter of some controversy. The poly(A)-binding protein (PABP), which is bound to the poly(A)-tail of the mRNA, interacts with the N-terminal part of eIF4G, and this interaction increases the affinity of eIF4E for the cap structure by 40-fold [71]. These protein interactions cause the mRNA to adopt a pseudocircular complex, with proximation of head to tail, enabling the ribosome to restart translation more rapidly, thereby contributing to increased translational efficiency [13, 63].

Others, however, argue that the fact that eIF4G binds both the PABP and eIF4E merely proves that mRNA is capable of circularizing in this fashion, not that it actually happens during the initiation of translation [72]. There are claims that 3' UTR binding proteins play a role in regulating translation, but some believe that the effect on protein synthesis is merely the effect of 3' UTRs on localisation and/or stability of the mRNA [72].

The rate of elongation also influences the total protein production in a cell. If a codon is rarely used in the genome of an organism, it can cause a pause in translation because of the corresponding low concentration of the tRNA required for that particular codon. Several studies have shown that

the levels of recombinant protein synthesised may increase many fold if the codon usage of the gene correlates to the codon bias in the organism in which the protein is expressed [73, 74]. However, the influence of the codon change on mRNA stability and on the efficiency of mRNA export into the cytoplasm has not been properly investigated [74, 75]. The tendency for increased protein synthesis is clear enough though for codon optimisation to be one of the most common tools used by molecular biologists for enhanced recombinant gene expression [76].

Whatever the cause of increased translational efficiency, it may in itself prevent breakdown of mRNA since the protection provided by being bound to the ribosome decreases the exposure of the molecule to endo-RNases [76].

As mentioned above proteins destined for secretion are translated on MBP. In a mammalian cell factory it is thus important to take steps to ensure that the mRNA coding for the recombinant protein in question is efficiently targeted to this polysome population and then translated as effectively as possible such that optimal amounts of product can be secreted from the cell.

### Secreted proteins

A protein to be secreted from a eukaryotic cell first enters the ER before transport into the Golgi complex. A signal peptide at the N-terminal end of the nascent polypeptide directs the molecule into the ER lumen. This is cleaved off by signal peptidase once it has served its purpose of targeting the protein to, and importing it into, the ER.

These signal peptides are generally between 15 and 30 amino acid residues in length, but can have more than 50 such residues [77,78]. They display substantial variation in the amino acid sequence, but they do share some common features; most of them consist of an N-terminal polar (n-) region, a hydrophobic core (h-) region, and a C-terminal polar cleavage (c-) region. The basic n-region is 2-5 amino acids long, and has a net positive charge. The h-region normally contains between 6 and 15 amino acid residues, and has been found to be the most essential part of the signal peptide for targeting to, and insertion into, the ER membrane [78].

Through increasing the basicity of the n-region and the hydrophobicity of the h-region in the IL-2 signal peptide, Zhang *et al.* [79] observed up to a 3.5 fold increase in secretion levels of a reporter protein. The c-region is imperative for the signal peptide to be cleaved off after the nascent polypeptide has been imported into the ER. For cleavage to occur, the -1 position needs to contain an amino acid with a short side chain, and charged amino acids should not be present in the -3 position. The c-region also often contains helix-breaking Pro and Gly residues [78].

When the nascent signal peptide emerges from the ribosome, it is recognised by the signal recognition particle (SRP). It has been proposed that the h-region forms an alpha helix, which functions as a recognition element for the SRP together with the ribosome [79]. The affinity of the signal peptide for the SRP will determine whether or not the nascent polypeptide chain will enter the ER. Since the different signal peptides show such sequence diversity, it is expected that they also show different affinities to the SRP, and that this dictates the efficiency with which a nascent polypeptide enters the secretory pathway [79, 80].

Translation then temporarily halts while the ribosome-signal peptide-SRP complex is directed to the ER membrane where it is recognised by the SRP receptor. This interaction leads to ribosome binding to the translocon and the insertion of the nascent polypeptide into the translocon pore. The translocon is a protein complex which provides a gateway into the ER, and which catalyses the transfer of growing peptide chains across the ER membrane. The SRP is then released, and there is a co-translational translocation of the nascent polypeptide through the internal pore of the translocon into the ER lumen [77, 81, 82]. On the luminal side of the ER membrane, the cleavage site in the c-region of the signal peptide is recognised by signal peptidases, and the signal peptide is cleaved off [83].

As the nascent polypeptide enters the ER, it is bound by molecular chaperones which prevent aggregation of the protein and assist in its folding. Any initial glycosylation also occurs in the ER. The folded, partially glycosylated protein then transits to the transitional ER (tER), where it is

packaged into vesicular tubular complexes (VTC), and translocated to the Golgi apparatus by dynein. At the Golgi, the VTC fuses with the *cis*-Golgi cisterna, or, according to an alternative theory, with other VTCs to form the *cis* cisterna. The cargo from the VTC (i.e. the protein) is transported through the Golgi stacks, where it is subjected to further PTMs such as glycosylation and proteolytic processing, to the *trans*-Golgi network (TGN). The TGN sorts and packages proteins into secretory granules based on their biochemical properties. Depending on coat proteins and regulatory domains, the secretory granules are thought to take different transport routes to the plasma membrane, to which they fuse and release their contents to the exterior of the cell through exocytosis. Proteins destined for regulated secretion are packaged into immature secretory granules and stored in the cytosol until the cell receives a signal triggering their transport to the plasma membrane [82, 84, 85].

### Experimental background

Through the understanding of the biological processes involved in transcription, transport, targeting, translation, and secretion of a protein, expression vectors can be tailored to increase the levels of protein produced by and secreted from a cell. Recent focus has been to alter the vector construct sequences being transcribed and translated in order to improve protein synthesis and secretion in mammalian cells. The standard expression vector pTRE2hyg, which was developed by Gossen and Bujard [86, 87], has been used as part of a model system in this quest for higher expression levels. It has an ampicillin resistance gene ( $Amp^r$ ) for selection in bacteria, and a hygromycin resistance gene ( $Hyg^r$ ) for selection in mammalian cells. Any recombinant gene inserted into the MCS will be under the influence of the strong CMV promoter located upstream of the insert. Through a transactivator sequence known as the Tet-regulatory element the expression of the inserted gene can be regulated (induced) by the addition of tetracycline. Downstream of any insert is the polyadenylation signal from  $\beta$ -globin. All the vector constructs contained identical promoter and polyadenylation signals.

The model system consists of the pTRE2hyg vector with the 5' UTR, coding sequence and 3' UTR from *Gaussia* luciferase (GLuc) cDNA (gi:12621053), along with the native luciferase signal peptide coding region or signal peptide coding regions from various natural sources, inserted into the MCS using a seamless cloning technique. In the text the constructs are termed p followed by a four letter code where the first, third and fourth letters refer to the 5' UTR, the coding region and the 3' UTR from *Gaussia* luciferase, respectively. The second letter represents the nature of the signal peptide inserted in the construct (G = *Gaussia* luciferase, A = human albumin, T = trypsinogen-2, cT = chymotrypsinogen, Mt = *Metridia longa* luciferase, V = *Vargula hilgendorfi*). When used \*refers to codon optimisation for mammalian gene expression [88]. Both *Metridia* and *Vargula*, like *Gaussia*, synthesise and secrete a luciferase.

*Gaussia* is a calanoid copepod of the *Metridiidae* family, and its habitat is from 200 to more than 400 m below sea level in the mesopelagic zone of the ocean [89]. This depth is characterised by low light levels, decreasing temperature, slower currents, lower food availability, and increasing salinity and pressure. When disturbed, *Gaussia* produces a brilliant blue bioluminescent display, probably as a defense mechanism against predators following some form of stimulus [90]. This bioluminescence is caused by the bulk secretion of the luciferase, most likely together with its substrate coelenterazine, from glands containing secretory vesicles. The light is produced from a reaction that oxidizes coelenterazine into coelenteramide catalysed by GLuc. The reaction displays flash kinetics, with the photon emission rapidly decaying with time. Even though GLuc and its substrate are believed to coexist inside the cell, the bioluminescence reaction is thought to be prevented from occurring in this environment due to the lack of cofactors such as  $O_2$ , which is present in sea water [91].

With its 185 amino acids (19.9 kDa), the monomeric GLuc is the smallest luciferase isolated until now. The reaction it catalyses is  $Na^+$  dependent, but unlike the reaction of the widely used firefly luciferase, it is ATP independent. Compared to the firefly luciferase, which remains inside the cell and catalyses a reaction causing an

intracellular light emission, GLuc differs as it is naturally secreted. A signal peptide of 17 amino acids in length directs its mRNA to the ER, causing it to enter the secretory pathway. *Gaussia* luciferase is heat stable, is resistant to acidic and alkaline conditions (pH optimum 7.7), and has a broad emission spectrum extending to 600 nm with a light peak of 480 nm. The only known luciferase with sequence similarity to GLuc is from *Metridia*, another copepod in the *Metridiidae* family [92].

When expressed in mammalian cells, recombinant *Gaussia* luciferase (rGLuc) encoded by humanised cDNA was found to generate far higher signal intensity (more than 100 fold) than other codon-humanised luciferases, such as the firefly and *Renilla* luciferases [88]. Importantly, rGLuc does not display any toxic effect in mammalian cells [88], and because it is not native to the cells, the background signal is negligible. These properties make rGLuc useful as a reporter protein in biological studies. In this work it was used as a tool to measure gene expression.

## RESULTS AND DISCUSSION

Various vector constructs were transfected into CHO cells in order to investigate the effect of signal peptides on rGLuc synthesis/secretion. *Gaussia* luciferase activity was measured in cell extracts and medium samples taken from stably transfected cell populations.

The observations, accumulated from a series of experiments, are presented in Table 1. The results show that the different signal peptides tested varied extensively in their relative abilities to generate rGLuc.

Data from experiments where mammalian signal peptides replaced the native *Gaussia* sequence are presented in Table 1A. It was surprising that the signal peptides derived from human albumin and human interleukin-2 (Il-2) were relatively inefficient compared to the native *Gaussia* luciferase sequence because both these proteins are known to be secreted with great efficiency from mammalian cells. In fact interleukin-derived signal sequences are in common use both in commercial protein production and in gene therapy research [93-99]. Recently, Zhang *et al.* [79] observed that the

secretion of both alkaline phosphatase and endostatin could be improved by up to 3.5 fold by increasing the basicity and the hydrophobicity of the Il-2 signal peptide. These experiments demonstrated that the performance of the Il-2 signal peptide can be improved substantially, suggesting that the native sequence does not operate optimally.

The fact that the human albumin signal peptide was apparently very ineffective was difficult to explain in the light of its important role in the hepatic cell where substantial amounts of albumin are synthesised and secreted into the blood circulation from the liver on a daily basis. It was considered possible that since the human albumin signal peptide did not function well in CHO cells then it would perhaps function more efficiently in a hepatic cell line. In an experiment where pGG\*G\*G and pGAG\*G were transfected into HepG2 cells and both medium samples and cell extracts were assayed for *Gaussia* luciferase activity, it was observed that the results were very similar to those obtained using CHO cells [111]. It was thus evident that the albumin signal peptide did not operate more effectively in a liver cell line than in CHO cells. Furthermore, the *Gaussia* luciferase signal peptide, i.e. a non-mammalian sequence, was able to operate very successfully in a cell line of human origin.

Signal peptides derived from trypsinogen 2 and chymotrypsinogen were chosen for comparison purposes because these proteins are naturally “bulk” secreted like *Gaussia* luciferase. In this context a bulk secreted protein is defined as a protein which, after being synthesised, is stored in secretory vacuoles until a physiological stimulus triggers exocytosis resulting in a release of vacuolar content. Thus both trypsinogen and chymotrypsinogen are produced and “bulk” secreted in large amounts by exocrine cells of the pancreas following hormonal stimulation. CHO cells transfected with pGT2G\*G and pGcTG\*G showed a lower level of synthesis and secretion of luciferase than cells transfected with pGG\*G\*G (Table 1A). None of the four mammalian signal peptides tested i.e. interleukin-2, albumin, trypsinogen 2 and chymotrypsinogen, proved to be as effective as that derived from *Gaussia* luciferase with respect to producing high levels of synthesis and secretion of rGLuc in CHO cells.

That the luciferase signal peptide derived from *Gaussia* should be far superior to four mammalian signal peptides in CHO cells with respect to rGluc production was an unexpected finding. In order to investigate whether effects caused by signal peptides take place at the stage of transcription or translation, the levels of *Gaussia* luciferase mRNA found in CHO cells stably transfected with constructs containing signal peptide coding regions derived from *Gaussia* luciferase, Il-2 or albumin were estimated. Interestingly, replacement of the signal peptide coding region from *Gaussia* with that either from Il-2 or albumin did not have a large impact on the amount of mRNA transcript since the corresponding levels were 89% and 77% respectively [111]. These values were in great contrast to rGLuc activity in cell extracts which were only 59% and 5% respectively, of those observed using the *Gaussia* signal peptide (Table 1A). These results would suggest that the dramatic impact on the quantity of active recombinant protein was not merely due to a shortage of mRNA but was rather coupled to a post-transcriptional event. Why there should be a substantial quantity of mRNA transcript available in the transfected cells but that it appears to be poorly utilised is not yet understood. It would seem, however, that more important than quantity is the quality of the mRNA and the potential of the transfected cell to utilise it in an efficient manner.

In experiments to test whether or not the signal peptide derived from *Gaussia* luciferase was as efficient as other signal peptides of marine origin, comparison was made using the luciferase signal peptides derived from *Metridia* (a copepod) and *Vargula* (an ostracod). Both copepods live in sea water at a temperature of 4°C, *Gaussia* living at depths of 200-400 m, while *Metridia* lives at 50-100m. *Vargula* is found in the shallow coastal waters of Japan. There is more than 60% sequence similarity between the *Gaussia* and the *Metridia* luciferases. All three species store the luciferase and its substrate in vesicles. In response to an appropriate stimulus the contents of the vesicles are released into the seawater and an intense flash of light is produced, frightening potential predators.

Cells transfected with either pGMtG\*G or pGVG\*G gave levels of rGluc which were about 50 and 80% respectively, of those found using the

native *Gaussia* luciferase peptide (Table 1B). Since *Metridia* and *Gaussia* share some similarities, it was unexpected that the signal peptide derived from *Metridia* luciferase proved to be only about 50% as effective as that of *Gaussia* luciferase.

If high-level secretion of a certain protein under difficult environmental conditions is mediated by a “powerful” signal peptide, it was considered that signal peptides derived from the marine organism *Oikopleura dioica* (a urochordate) may provide such a source. This organism secretes a very complex mucous extracellular structure, the house, the function of which is to filter seawater for food particles. The house is replaced every 3-4 hours, and is produced by a fixed number of cells that form a monolayer covering the trunk of the organism, termed the oikoplastic epithelium. The proteins that compose the house are called oikosins, of which there are seven species [100,101]. Since the cells of the oikoplastic epithelium have an extremely high secretory activity the signal peptides derived from the seven oikosin proteins were therefore chosen for study because they presented the possibility of comparing the relative effectiveness of seven closely related signal peptides on production of *Gaussia* luciferase. It can be seen from Table 1B that there were great individual differences between the signal peptides with respect to production of rGLuc. Only oikosin 1 signal peptide resulted in the generation of a substantial amount of luciferase while the other six demonstrated extremely limited levels of production.

In comparison to oikosin 1, use of oikosin 2-7 signal peptides only gave rise to extremely low levels of luciferase activity in medium samples (0,01-0,5% of values, achieved using the *Gaussia* luciferase signal peptide - Table 1B). Northern blotting, however, showed the presence of high quantities of rGLuc mRNA (Tröbe *et al.*, manuscript in preparation) irrespective of which oikosin signal peptide was employed. These observations were thus similar to those obtained using Il-2 or albumin signal sequences and again demonstrated that there is not necessarily a direct correlation between mRNA level and amount of recombinant protein produced.

Two prokaryotic signal peptides were tested in order to be able to compare their effectiveness

with marine/mammalian counterparts. The signal peptide of killer toxin K28 (K28-SP), a preprotoxin, from the yeast virus M28, has been found to allow the efficient secretion of GFP from different yeast strains [102], making it an interesting candidate for testing in mammalian cells. The signal peptide of Slmj 1 from *Methanococcus jannaschii*, an extreme thermophile methanogenic Archaeobacterium, living at deep sea levels under high-pressure conditions, was also considered as an interesting candidate for study. Slmj 1 is a surface layer protein that is part of the cellular envelope [103]. The capacity of secretion under such extreme conditions could be explained by the action of a powerful signal peptide. It can be seen from Table 1C, however, that neither of the two signal peptides tested were particularly effective with respect to production of rGLuc in transfected CHO cells.

From the results presented in Table 1A-C it is evident that there are extremely large variations in the levels of luciferase activity found in medium samples of cell populations stably transfected with vector constructs containing a series of signal peptide coding sequences. It was unexpected to find that some of the signal peptides investigated, chosen because of their expected potential to promote high-level synthesis/ secretion, produced strikingly low levels of rGLuc compared to the *Gaussia* luciferase signal peptide. The performance of a signal peptide is therefore not necessarily predictable from the features of the corresponding protein, or the role this protein plays in the organism it is derived from.

Secretory signal peptides from prokaryotic and eukaryotic organisms share a common overall architecture [104], and it has been demonstrated that several signal peptides from one organism could function in another [105-108]. However, not all signal peptides function efficiently in foreign organisms [109]. A number of murine signal peptides, functional in a mouse cell line, are, however, not efficient in yeast [110]. These findings indicate that the exploitation of signal peptides in organisms different from their natural origin can be problematic. The weak performances of both Mj-SP and K28-SP described here in CHO cells might be attributed to such a host-dependent effect. It is not clear, though, if such an effect

may also explain the low degree of efficiency seen with the seven oikosin signal peptides derived from *Oikopleura*, which is, compared to archaeobacteria and yeast, in closer phylogenetic relation to mammals. It is evident, however, that the signal peptide derived from *Gaussia* (also in closer relation to mammals) functions extremely well in mammalian cells [111,112].

Since none of the signal peptides tested to direct rGLuc production were able to match the performance of the native *Gaussia* signal peptide, the ability of this peptide to function in the synthesis of another reporter protein was monitored, namely *Vargula* luciferase. When the *Vargula* luciferase signal peptide coding region was replaced by that derived from *Gaussia* luciferase in a vector which otherwise contained *Vargula* sequences, luciferase activity determination in transfected CHO cell extracts and the culture medium showed that the *Gaussia* signal peptide was far more effective than the native *Vargula* luciferase signal peptide. Furthermore, the *Gaussia* signal peptide proved to be superior in the production of *Vargula* luciferase when compared to signal peptides derived from albumin, interleukin-2 and follistatin (Ravneberg *et al.*, submitted). Taken together the results suggest that the *Gaussia* luciferase signal peptide has extremely interesting properties.

Due to the fact that the signal peptide derived from *Gaussia* luciferase proved to be more effective than four mammalian signal peptides, nine of marine origin and two prokaryotic signal peptides with respect to recombinant luciferase synthesis/ secretion in CHO cells, it was considered important to examine whether or not this signal peptide was effective in production of a mammalian protein. Human endostatin was chosen for these studies where a comparison was made between the *Gaussia* signal peptide and that from albumin. The results showed that the construct containing the *Gaussia* luciferase signal peptide gave rise to a several fold higher yield of endostatin than the albumin signal peptide [111]. It is thus evident that the signal peptide of marine origin can also be successfully used for production of mammalian proteins.

Since it had been observed that the human albumin signal peptide did not function well in

**Table 1.** A comparison of the effectiveness of signal peptides from different sources with that derived from *Gaussia* luciferase on recombinant luciferase production.

## A : Mammalian signal peptides

Origin of signal peptide	Amino acid sequence	% relative activity
<i>Gaussia</i> luciferase	MGVKVLFALICIAVAEA	100
Human albumin	MKWVTFISLLFLFSSAYS	5
Human chymotrypsinogen	MAFLWLLSCWALLGTTFG	59
Human interleukin-2	MQLLSICIALALV	37
Human trypsinogen-2	MNLLLILTFVAAAVA	71

## B : Marine organisms

Origin of signal peptide	Amino acid sequence	% relative activity
<i>Gaussia</i> luciferase	MGVKVLFALICIAVAEA	100
<i>Metridia</i> luciferase	MDIKVVFTLVFSALVQA	47
Oikosin 1*	MLLLSALLGLAHGYS	45
Oikosin 2A	MKLLASVLTIAAADYACC	0,5
Oikosin 3	MKISAGLLGVALGQNEGSAEA	0,01
Oikosin 4A	MKLFAALSASFASVEA	0,01
Oikosin 5A	MKLLCSVLLGTVFG	0,04
Oikosin 6A	MKISPLLVVTAVVG	0,01
Oikosin 7A	MKIAATFAALASATEWQG	0,01
<i>Vargula</i> luciferase	MKIILSVILAYCVTDNC	26

\* Oikosins 1-7A derived from *Oikopleura dioica*

## C : Prokaryotic

Origin of signal peptide	Amino acid sequence	% relative activity
<i>Gaussia</i> luciferase	MGVKVLFALICIAVAEA	100
<i>Methanococcus jannaschii</i> Slmj1	MAMSL <b>KK</b> IGAIAVGGAMVATALASGVAA	0,01
M28 virus killer toxin K28	MESVSSLFNIFSTIMVNYKSLV <b>L</b> ALLSVSNLKYARG	10

Constructs were made based on the pTRE2hyg expression vector system. The basic construct contained the 5'UTR, the signal peptide coding region, the coding region and the 3'UTR all derived from *Gaussia* luciferase. Constructs were prepared where the signal peptide coding region of *Gaussia* luciferase was exchanged with the coding regions of signal peptides derived from A : mammalian, B : marine or C : prokaryotic sources. Luciferase activity was measured in medium samples obtained from stably transfected cells. The total activity in the medium was related to that found using the signal peptide coding region of *Gaussia* luciferase and expressed as percentage relative activity. The first four amino acids in the sequence (reading from left to right) and the final three are regarded as constituting the n- and c-regions, respectively. Letters in bold indicate hydrophilic amino acids in the h-region.

CHO cells it was considered possible that it may function effectively in a cell line of liver origin. *Gaussia* luciferase signal peptide and albumin signal peptide constructs were transfected into HepG2 cells and rGLuc activity was measured in cell extracts and medium samples. Almost identical results were obtained to those using CHO cells [111]. These observations were interesting in two ways, firstly in that the albumin signal peptide did not appear to function more efficiently in a liver cell line than in CHO cells, and secondly, that the luciferase signal peptide, derived from a marine copepod, was able to operate effectively not only in a human cell line but also in one originating from the hamster (i.e. CHO cells). It would appear, therefore, that when wanting to achieve high levels of recombinant protein then the choice of signal peptide is extremely important, more so than the actual cell line to be used.

As discussed in the section above on “Secreted and Membrane Proteins” it has been suggested that the h-region of the signal peptide forms an alpha helix producing some type of recognition element for the SRP. The h-region may therefore be important in some form of regulation at the level of attachment to the SRP. The high level of synthesis/secretion caused by using the *Gaussia* luciferase signal peptide than other signal peptides could thus be due to it having a higher affinity for the SRP.

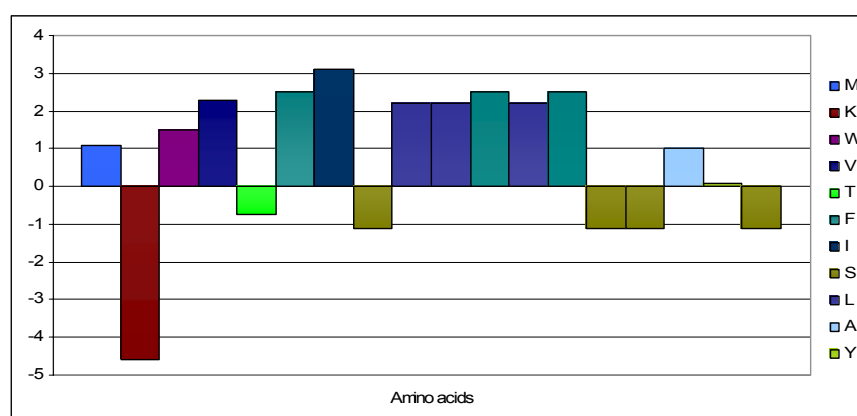
Figure 1 shows the hydropathic plots of two signal peptides, namely human albumin and *Gaussia* luciferase. It can be seen that the h-regions differ since that of *Gaussia* luciferase does not have any hydrophilic amino acids while that of albumin has a total of four (one threonine and three serine residues). Similar plots of the other signal peptides tested (amino acid sequences are depicted in Tables 1A-C) showed that the *Gaussia* luciferase signal peptide was the only one that did not have hydrophilic amino acid residues in the h-region. Whether or not this is important with respect to efficiency of function of the signal peptide is not yet known.

Recent results have demonstrated that by substituting hydrophilic amino acids in the h-region of the *Metridia* luciferase signal peptide with more hydrophobic residues in a vector construct containing the *Metridia* signal peptide coding

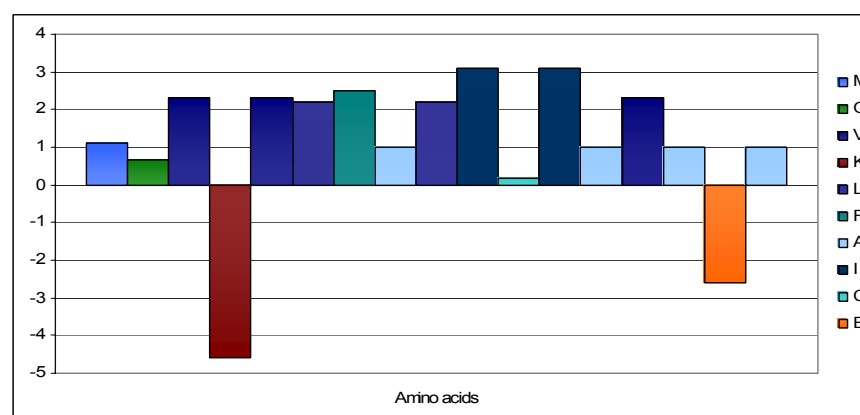
sequence fused to the coding sequence of rGLuc (pGMtG\*G), with the change to a cysteine being the most important, the levels of rGLuc synthesis and secretion increased up to 1.9 fold. The improvement of secretion when changing the hydrophobicity of the signal peptide of Il-2 has also been observed by Zhang *et al.* [79]. The h-region of albumin does indeed include hydrophilic residues, but does not contain a cysteine residue. These differences in properties might explain why the levels of rGLuc synthesis and secretion are lower when using the albumin signal peptide instead of the native *Gaussia* luciferase signal peptide.

The observations reviewed here demonstrate that there are great differences with respect to the functionality of individual signal peptides in production of a recombinant protein. The results clearly suggest that the choice of signal peptide is imperative when designing a “secretion cassette” with the object of optimising synthesis/secretion of a recombinant protein. Although the amount of recombinant protein produced was greatly affected by the nature of the signal peptide utilised, this appeared to have no major influence on the level of mRNA available for translation, suggesting that the signal peptide plays some form of regulatory role at the post-transcriptional level. We believe that this is the first report indicating a control mechanism involving the signal peptide.

The finding that a signal peptide originating from a marine organism was far more effective in a mammalian expression system than commonly used mammalian signal peptides such as human Il-2 and albumin was quite surprising. It is also interesting to note that there was considerable variation between the individual signal peptides tested with regard to “performance”. For instance, four of the oikosin signal peptides produced only 0,01% of the level of *Gaussia* luciferase that was achieved using the native signal sequence. Taken together the results described here demonstrate that the actual choice of signal peptide can have a considerable impact on the amount of protein synthesised/secreted from mammalian cells. Future work will be designed to examine the comparative effectiveness of other signal peptides from alternative sources and also



Human albumin signal peptide : MKWVTFISLLFLFSSAYS

*Gaussia* luciferase signal peptide : MGVKVLFALICIAVAEA

**Figure 1.** Hydropathy plots of the amino acids in the signal peptides of human albumin and *Gaussia* luciferase. Plots were made using hydropathy values for amino acids taken from Eisenberg *et al.* [113].

to understand their manner of function in post-transcriptional control of protein synthesis. It has become evident that care needs to be exercised when selecting a signal peptide for use in a vector when one is aiming for bulk production/secretion of a recombinant protein in a mammalian cell system since this can have a major influence on the outcome.

#### ACKNOWLEDGEMENTS

Professor John E. Hesketh (University of Newcastle, UK) and Professor Albert Tauler (University of Barcelona, Spain) are thanked for constructive discussions during the course of this work. Sarsia Innovation is acknowledged for financial support. Dr. Thomas Pruß is thanked for carefully reading the manuscript.

#### REFERENCES

1. Chu, L., and Robinson, D. K. 2001, *Curr. Opin. Biotechnol.*, 12, 180.
2. Andersen, D. C., and Krummen, L. 2002, *Curr. Opin. Biotechnol.*, 13, 117.
3. Kwaks, T. H., and Otte, A. P. 2006, *Trends Biotechnol.*, 24, 137.
4. Fischer, R., and Emans, N. 2000, *Transgenic Res.*, 9, 279.
5. Wurm, F. M. 2004, *Nat. Biotechnol.*, 22, 1393.
6. Banan, M., Rojas, I. C., Lee, W. H., King, H. L., Harriss, J. V., Kobayashi, R., Webb, C. F., and Gottlieb, P. D. 1997, *J. Biol. Chem.*, 272, 18440.

7. Van Leeuwen, W., Mlynarova, L., Nap, J. P., van der Plas, L. H., and van der Krol, A. R. 2001, *Plant Mol. Biol.*, 47, 543.
8. Phi-Van, L., and Stratling, W. H. 1996, *Biochemistry*, 35, 10735.
9. Bode, J., Goetze, S., Heng, H., Krawetz, S. A., and Benham, C. 2003, *Chromosome Res.*, 11, 435.
10. Ringner, M., and Krogh, M. 2005, *PLoS Comput. Biol.*, 1, e72.
11. Mignone, F., Gissi, C., Liuni, S., and Pesole, G. 2002, *Genome Biol*, 3, REVIEWS 0004.1-0004.10.
12. Meijer, H. A., and Thomas, A. A. 2002, *Biochem. J.*, 367, 1.
13. Mazumder, B., Seshadri, V., and Fox, P. L. 2003, *Trends Biochem. Sci.*, 28, 91.
14. van der Velden, A W., Voorma, H. O., and Thomas, A. A. 2001, *Biotechniques*, 31, 572, 574, 576-80.
15. Darzacq, X., Singer, R. H., and Shav-Tal, Y. 2005, *Curr. Opin. Cell Biol.*, 17, 332.
16. Verrijzer, C. P., Chen, J. L., Yokomori, K., and Tjian, R. 1995, *Cell* 81, 1115.
17. Fiering, S., Whitelaw, E., and Martin, D. I. 2000, *Bioessays*, 22, 381.
18. Lee, T. I., and Young, R. A. 2000, *Annu. Rev. Genet.*, 34, 77.
19. Faiger, H., Ivanchenko, M., Cohen, I., and Haran, T. E. 2006, *Nucleic Acids Res.* 34, 104.
20. Carninci, P., Sandelin, A., Lenhard, B., Katayama, S., Shimokawa, K., Ponjavic, J., Semple, C. A., Taylor, M. S., Engstrom, P. G., Frith, M. C., Forrest, A. R., Alkema, W. B., Tan, S. L., Plessy, C., Kodzius, R., Ravasi, T., Kasukawa, T., Fukuda, S., Kitazume, Y., Kawaji, H., Kai, C., Nakamura, M., Konno, H., Nakano, K., Mottagui-Tabar, S., Arner, P., Chesi, A., Gustincich, S., Persichetti, F., Suzuki, H., Grimmond, S. M., Wells, C. A., Orlando, V., Wahlestedt, C., Liu, E. T., Harbers, M., Kawai, J., Bajic, V. B., Hume, D. A., Kanamori-Katayama, M., and Hayashizaki, Y. 2006, *Nat. Genet.*, 38, 626.
21. Muller, F. and Tora, L. 2004, *EMBO J.* 23, 2.
22. Gershenzon, N. I. and Ioshikhes, I. P. 2005, *Bioinformatics*, 21, 1295.
23. Pal, M., McKean, D., and Luse, D. S. 2001, *Mol. Cell Biol.*, 21, 5815.
24. Keene, R. G., and Luse, D. S. 1999, *J. Biol. Chem.*, 274, 11526.
25. Sims, R. J. I., Belotserkovskaya, R., and Reinberg, D. 2004, *Genes Dev.*, 18, 2437.
26. Proudfoot, N. J., Furger, A., and Dye, M. J. 2002, *Cell*, 108, 501.
27. Orphanides, G., and Reinberg, D. 2002, *Cell*, 108, 439.
28. Bentley, D. L. 2005, *Curr. Opin. Cell Biol.*, 17, 251.
29. Zorio, D. A., and Bentley, D. L. 2004, *Exp. Cell Res.*, 296, 91.
30. Cullen, B. R. 2003, *J. Cell Sci.*, 116, 587.
31. Luo, M. J., and Reed, R. 1999, *Proc. Natl. Acad. Sci.*, 96, 14937.
32. Le Hir, H., Nott, A., and Moore, M. J. 2003, *Trends Biochem. Sci.*, 28, 215.
33. Arhin, G. K., Boots, M., Bagga, P. S., Milcarek, C., and Wilusz, J. 2002, *Nucleic Acids Res.*, 30, 1842.
34. Gu, H., Das Gupta, J., and Schoenberg, D. R. 1999, *Proc. Natl. Acad. Sci.*, 96, 8943.
35. Gilmartin, G. M. 2005, *Genes Dev.* 19, 2517.
36. Kim, D., Kim, J. D., Baek, K., Yoon, Y., and Yoon, J. 2003, *Biotechnol. Prog.* 19, 1620.
37. Yeung, G., Choi, L. M., Chao, L. C., Park, N. J., Liu, D., Jamil, A., and Martinson, H. G. 1998, *Mol. Cell Biol.*, 18, 276.
38. Saguez, C., Olesen, J. R., and Jensen, T. H. 2005, *Curr. Opin. Cell Biol.*, 17, 287.
39. Lei, E. P., and Silver, P. A. 2002, *Genes Dev.*, 16, 2761.
40. Rout, M. P., and Aitchison, J. D. 2001, *J. Biol. Chem.*, 276, 16593.
41. Sommer, P., and Nehrbass, U. 2005, *Curr. Opin. Cell Biol.*, 17, 294.
42. Tange, T. O., Nott, A., and Moore, M.J. 2004, *Curr. Opin. Cell Biol.*, 16, 279.
43. Reed, R. 2003, *Curr. Opin. Cell Biol.*, 15, 326.
44. Shyu, A. B., and Wilkinson, M.F. 2000, *Cell*, 102, 135.
45. Kuersten, S., and Goodwin, E. B. 2005, *Biol. Cell*, 97, 469.
46. Hovland, R., Hesketh, J. E., and Pryme, I. F. 1996, *Int. J. Biochem. Cell Biol.*, 28, 1089.

47. Hovland, R., Pryme, I. F., Campbell, G., and Hesketh, J.E. 1995, *Biochem. J.*, 310, 193.
48. Pryme, I. F., Partridge, K., Johannessen, A. J., Jodar, D., Tauler, A., and Hesketh J. E. 1996, *Gen. Eng. & Biotechnol.*, 16, 137.
49. Vedeler, A., Pryme, I. F., and Hesketh, J. E. 1991, *Mol. Cell. Biochem.*, 100, 183.
50. Hesketh, J. E., and Pryme I. F. 1991, *Biochem. J.*, 277, 1.
51. Jansen, R. P. 1999, *FASEB J.*, 13, 455.
52. Jansen, R. P. 2001, *Nat. Rev. Mol. Cell. Biol.*, 2, 247.
53. Kloc, M., Zearfoss, N. R., and Etkin, L. D. 2002, *Cell*, 108, 533.
54. Palacios, I. M., and St Johnston, D. 2001, *Annu. Rev. Cell Dev. Biol.*, 17, 569.
55. Bassell, G. J., Oleynikov, Y., and Singer, R. H. 1999, *FASEB J.*, 13, 447.
56. Partridge, K. A., Johannessen, A., Tauler, A., Pryme, I. F., and Hesketh, J. E. 1999, *Cytotechnology*, 30, 37.
57. Chabanon, H., Mickleburgh, I., Burtle, B., Pedder, C., and Hesketh, J. E. 2005, *Biochem. J.*, 392, 475.
58. Day, D. A., and Tuite, M. F. 1998, *J. Endocrinol.*, 157, 361.
59. Wilusz, C. J., and Wilusz, J. 2004, *Trends Genet.*, 20, 491.
60. Mukherjee, D., Gao, M., O'Connor, J. P., Rajmakers, R., Pruijn, G., Lutz, C. S., and Wilusz, J. 2002, *EMBO J.*, 21, 165.
61. Jiang, Y., Xu, X. S., and Russell, J. E. 2006, *Mol. Cell Biol.*, 26, 2419.
62. Roberts, S. J., Chung, K. N., Nachmanoff, K., and Elwood, P. C. 1997, *Biochem. J.*, 326, 439.
63. Preiss, T. W., and Hentze, M. 2003, *Bioessays* 25, 1201.
64. Kozak, M. 1987, *Nucleic Acids Res.* 15, 8125.
65. Kozak, M. 1999, *Gene*, 234, 187.
66. Kozak, M. 2002, *Mamm. Genome.*, 13, 401.
67. Kozak, M. 2005, *Gene*, 361, 13.
68. Suzuki, Y., Ishihara, D., Sasaki, M., Nakagawa, H., Hata, H., Tsunoda, T., Watanabe, M., Komatsu, T., Ota, T., Isogai, T., Suyama, A., and Sugano, S. 2000, *Genomics*, 64, 286.
69. Davuluri, R. V., Suzuki, Y., Sugano, S., and Zhang, M. Q. 2000, *Genome Res.*, 10, 1807.
70. Gray, N. K., and Wickens, M. 1998, *Annu. Rev. Cell Dev. Biol.*, 14, 399.
71. Gallie, D. R., Ling, J., Niepel, M., Morley, S. J., and Pain, V. M. 2000, *Nucleic Acids Res.*, 28, 2943.
72. Kozak, M. 2004, *Gene*, 343, 41.
73. Hu, S., Li, L., Qiao, J., Guo, Y., Cheng, L., and Liu, J. 2006, *Protein Expr. Purif.*, 47, 1640.
74. Narum, D. L., Kumar, S., Rogers, W. O., Fuhrmann, S. R., Liang, H., Oakley, M., Taye, A., Sim, B. K., and Hoffman, S. L. 2001, *Infect. Immun.*, 69, 7250.
75. Kurland, C. G. 1991, *FEBS Lett.*, 285, 165.
76. Gustafsson, C., Govindarajan, S., and Minshull, J. 2004, *Trends Biotechnol.*, 22, 346.
77. Johnson, A. E., and van Waes, M. A. 1999, *Annu. Rev. Cell Dev. Biol.*, 15, 799.
78. Martoglio, B., and Dobberstein, B. 1998, *Trends Cell Biol.*, 8, 410.
79. Zhang, L., Leng, Q., and Mixson, A. J. 2005, *J. Gene Med.*, 7, 354.
80. Zheng, N., and Gierasch, L. M. 1996, *Cell*, 86, 849.
81. Swanton, E., and Bulleid, N. J. 2003, *Mol. Membr. Biol.*, 20, 99.
82. van Vliet, C., Thomas, E. C., Merino-Trigo, A., Teasdale, R. D., and Gleeson, P. A. 2003, *Prog. Biophys. Mol. Biol.*, 83, 1.
83. Walter, P., and Johnson, A. E. 1994, *Annu. Rev. Cell Biol.*, 10, 87.
84. Murshid, A., and Presley, J. F. 2004, *Cell. Mol. Life Sci.*, 61, 133.
85. Ponnambalam, S., and Baldwin, S. A. 2003, *Mol. Membr. Biol.*, 20, 129.
86. Gossen, M., and Bujard, H. 1992, *Proc. Natl. Acad. Sci.*, 89, 5547.
87. Gossen, M., Freundlieb, S., Bender, G., Muller, G., Hillen, W., and Bujard, H. 1995, *Science*, 268, 1766.
88. Tannous, B. A., Kim, D. E., Fernandez, J. L., Weissleder, R., and Breakefield, X. O. 2005, *Mol. Ther.*, 11, 435.
89. Bowlby, M. R., and Case, J. F. 1991, *Marine Biology*, 110, 329.
90. Wiles, S., Ferguson, K., Stefanidou, M., Young, D. B., and Robertson, B. D. 2005, *Appl. Environ. Microbiol.*, 71, 3427.

91. Bowlby, M. R., and Case, J. F. 1991, *Biological Bulletin*, 180, 440.
92. Markova, S. V., Golz, S., Frank, L. A., Kalthof, B., and Vysotski, E. S. 2004, *J. Biol. Chem.*, 279, 3212.
93. Bamford, R. N., DeFilippis, A. P., Azimi, N., Kurys, G., and Waldmann, T. A. 1998, *J. Immunol.*, 160, 4418.
94. Komada, F., Nishiguchi, K., Tanigawara, Y., Iwakawa, S., and Okumura, K. 1999, *Biol. Pharm. Bull.*, 22, 846.
95. La Flamme, A. C., Buckner, F. S., Swindle, J. Ajioka, J., and van Voorhis, W. C. 1995, *Mol. Biochem. Parasitol.*, 75, 25.
96. Liu, J. O'Kane, D. J., and Escher, A. 1997, *Gene* 203, 141.
97. Sasada, R., Onda, H., and Igarashi, K. 1987, *Cell. Struct. Funct.* 12, 205.
98. Sasada, R., Marumoto, R., and Igarashi, K. 1988, *Cell. Struct. Funct.*, 13, 129.
99. Suzuki, K., Nakazato, H., Matsui, H., Hasumi, M., Shibata, Y., Ito, K., Fukabori, Y., Kurokawa, K., and Yamanaka, H. 2001, *J. Leukoc. Biol.*, 69, 531.
100. Spada, F., Steen, H., Troedsson, C., Kallesoe, T., Spriet, E., Mann, M., and Thompson, E. M. 2001, *J. Biol. Chem.*, 276, 20624.
101. Thompson, E. M., Kallesoe, T., and Spada, F. 2001, *Dev. Biol.* 238, 260.
102. Eiden-Plach, A., Zagorc, T., Heintel, T., Carius, Y., Breinig, F., and Schmitt, M. J. 2004, *Appl. Environ. Microbiol.*, 70, 961.
103. Akca, E., Claus, H., Schultz, N., Karbach, G., Schlott, B., Debaerdemaeker, T., Declercq, J. P., and Konig, H. 2002, *Extremophiles*, 6, 351.
104. von Heijne, G. 2002, In *Protein targeting, transport and translocation*, R. E. Dalbey, and G. von Heijne (Ed.), Academic Press, 35.
105. Muller, M., Ibrahimi, I., Chang, C. N., Walter, P., and Blobel, G. 1982, *J. Biol. Chem.*, 257, 11860.
106. Sleep, D., Belfield, G. P., and Goodey, A. R. 1990, *Biotechnology*, 8, 42.
107. Talmadge, K., Stahl, S., and Gilbert, W. 1980, *Proc. Natl. Acad. Sci.*, 77, 3369.
108. Yost, C. S., Hedgpeth, J., and Lingappa, V. R. 1983, *Cell*, 34, 759.
109. Kojima, T., and Kitamura, T. 1999, *Nat. Biotechnol.*, 17, 487.
110. Galliciotti, G., Schneider, H., Wyder, L., Vitaliti, A., Wittmer, M., Ajmo, M., and Klemenz, R. 2001, *J. Membr. Biol.*, 183, 175.
111. Knappskog, S., Ravneberg H., Gjerdrum, C., Stern, B., and Pryme, I. F. 2007, *J. Biotechnol.*, 128, 705.
112. Stern, B., Gjerdrum, C., Knappskog, S., Minsaas, L., Nylund, S., Olsen, L. C., Olsen, L., Ravneberg, H., Tröbe, C., Vollsund, E., Tauler, A., Hesketh, J. E., and Pryme, I. F. 2007, *Cell Technology for Cell Products*, R. Smith (Ed.). Springer, Dordrecht, 261.
113. Eisenberg, D., Weiss, R. M., Terwilliger, T. C., and Wilcox, W. 1982, *Faraday Symp Chem Soc.*, 17, 109.