Chapter 3

The complex and interactive pathway from (trans)genes to proteins

DAVID QUIST,[1] KAARE M. NIELSEN[1, 2] AND TERJE TRAAVIK[1, 3]
[1]THE NORWEGIAN INSTITUTE OF GENE ECOLOGY (GENØK), TROMSØ, NORWAY
[2]DEPARTMENT OF PHARMACY, UNIVERSITY OF TROMSØ, NORWAY
[3]DEPARTMENT OF MICROBIOLOGY AND VIROLOGY, UNIVERSITY OF TROMSØ, NORWAY

Multicellular organisms, such as animals and plants, consist of hundreds of different cell types. Each cell type (e.g. the various types found in liver, heart and lung organs) contributes a specific function to the organism. Yet, all of these cells contain the same number of identical genes (i.e. 20,000–50,000). The high cell diversity is achieved not through gene content, but through the tightly controlled regulation of expression of a subset of the genes in each cell type. This chapter introduces the different steps on the pathway from the gene to functional protein(s), and shows that a single gene can give rise to a high number of more or less related yet functionally distinct proteins. Knowledge of the broad range of factors governing gene expression in various cellular and environmental conditions is necessary to understand how genetic engineering may introduce novel risk aspects of genetically modified organisms (GMOs). From a basic science standpoint, genetic engineering has been an important development in science to uncover the inherent complexity of factors regulating gene expression. Hence, the limited understanding of how these factors relate within a biosafety context is an important source of uncertainty in the risk assessment of GMOs.

The pathway from genes to proteins in higher (eukaryotic) organisms (outlined in Chapter 2) involve a complex series of pathways divided into the following steps:

1. Regulation of gene transcription
- 1.1. Promoter recognition
- 1.2. RNA transcript modifications
- 1.3. Stability of RNA transcript
- 1.4. mRNA transport to the cytoplasm

2. Regulation of mRNA translation

3. Regulation of protein activity and stability
- 3.1. Protein folding, cleavage and chemical modification
- 3.2. Higher order protein interactions
- 3.3. Regulated protein degradation

1. Regulation of gene transcription

1.1. Promoter recognition

The first step from gene to protein involves the transcription of a gene's DNA code into messenger RNA (mRNA). The start site (switch) of mRNA production is called

the promoter.[1] The various genes in a cell have different promoters ensuring gene expression is «on» or «off» in response to specific developmental and environmental conditions. A variety of proteins, known as *transcription factors* (TFs), bind to DNA in a sequence-specific manner, that initiate and regulate transcription.[2] The transcription factors bind to DNA either in the promoter or further upstream of the gene.[3] A given promoter is composed of a variety of partly overlapping binding areas for different TFs. The occurrence of relevant TFs in a given cell type will determine whether, and to what extent, a particular gene is transcribed and proteins are produced. Figure 3.1 presents an outline of a generic eukaryotic promoter.

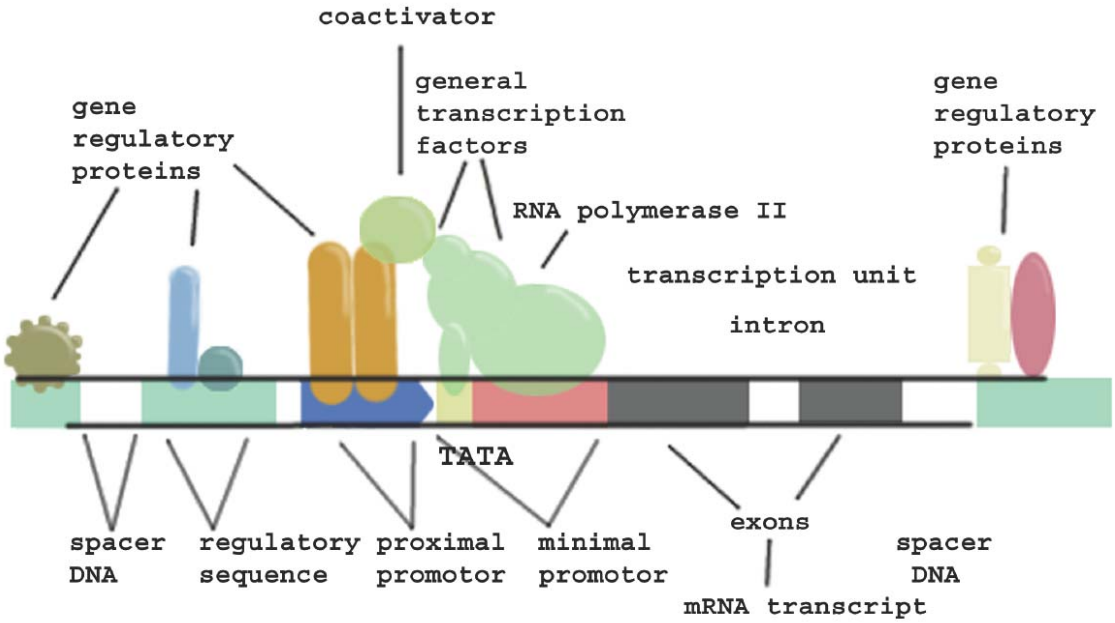

Figure 3.1. Outline of a generic eukaryotic promoter.

In addition to the promoter, three other types of DNA elements bind transcription factors and regulate cell type specific gene expression:

– *Enhancers* are DNA sequences that serve as specific binding sites for transcription factors to up-regulate the rate of transcription initiation. Enhancer regions are usually relatively short (30–500 base pairs), and have several binding sites for TFs.

– *Silencers* are DNA sequences that serve as specific binding sites for transcription factors that upon binding will down-regulate transcription initiation.

[1]A promoter is defined as a segment of DNA to which the RNA polymerase II enzyme attaches. The promoter binds general and specific transcription factors (proteins) that guide the polymerase to the initiation site and regulate the rate of transcription. The minimal promoter is the DNA sequence at which the general transcription factors and RNA polymerase II assemble.

[2]Each TF recognizes and binds to a specific 5–20 bp long DNA region. One important member of this class of proteins is TFIID. This protein binds to a short AT-rich sequence, the 'TATA box', found approximately 30 nucleotides upstream of the transcription initiation site in many eukaryotic genes. The primary function of TFIID is to direct RNA polymerase II to the initiation site.

[3]The fact that 5% of an organism's genes encode TFs underscores the importance of this protein family in biology.

– *Insulators* are DNA sequence elements that prevent inappropriate interactions between adjacent chromatin domains.

These DNA motifs, called functional elements, may all be located upstream or downstream of the gene, or within an *intron* (non-coding DNA sequence). While promoters have defined sequence orientations, enhancers and silencers can be turned around, and still exert their biological functions. The combination of these regulatory elements and their locations relative to the promoter are different for each gene. In a GMO context, it is important to realize that all the mentioned regulatory elements may influence the transcription of more than one gene, including non-target native genes. Furthermore, the transgenesis process inserts new promoters and/or other functional elements at unpredictable sites in an established genome. Finally, small parts of the inserted transgenic construct (e.g. plasmid backbone sequences) may contain functional regulatory elements.

1.2 RNA transcript modifications

Once appropriate TFs are bound to the promoter, enzymes called RNA polymerases[4] will produce single-stranded mRNA (transcripts). This RNA transcript undergoes a series of modifications in the cell nucleus before it is translocated to the cytoplasm for subsequent translation into a protein strand. Both ends of the primary mRNA transcript are modified.[5] Moreover, non-protein coding RNA regions (introns) are removed from the RNA strand, leaving only the regions that contain information left to be transcribed (exons). This intron removal is called *splicing*. The DNA signals, which direct the splicing, flank the intron.[6] Most genes in higher eukaryotes contain one or more introns, which are generally longer than the exons. Hence, the major part of the primary RNA transcript is removed in order to generate a functional mRNA ready to be translated in the cytoplasm. This can occur in a number of combinations (Figure 3.2), leading to the production of different mRNAs, and hence protein products, from the same initial DNA sequence[7]. The combination of introns, which are removed by splicing, varies between cell types, thus allowing a single gene to produce transcripts coding different protein sizes. This form of post-transcriptional regulation is called *differential splicing* or *alternative splicing.*

Because alternative splicing allows individual genes to produce multiple protein types with variable post-transcriptional RNA modifications, stability and function, the 'one gene, one protein' rule of the Central Dogma is erroneous. The genetic composition of an organism cannot therefore be used to predict *a priori* the actual protein composition (*proteome*) of a cell at a given life stage or under different sets of ecological or biological conditions. Alternative splicing is the most important process

[4]In eukaryotes there are three separate types of RNA polymerases (enzymes) which are responsible for the production of different kinds of RNA. Messenger RNA (mRNA) is synthesized by RNA polymerase II, while RNA polymerases I and III synthesize structural RNAs.

[5]The earliest processing step in the formation of mRNA is the enzymatic addition of a cap, which occurs almost simultaneously with the initiation of transcription. The site in the genomic DNA at which transcription starts is commonly known as the cap site. Close to the cap site in the DNA are recognition sites for DNA binding transcription factors which cause RNA polymerase II to initiate transcription. While transcription initiation may be reasonably well understood, the termination process has been less well defined. Transcription proceeds beyond the eventual 3' end of the mature mRNA, and the resultant primary transcript is then cleaved internally to generate the mRNA precursor. Cleavage takes place 10–20 nucleotides downstream of a specific AU-rich sequence, AAUAAA, which is highly conserved in all eukaryotic mRNAs. An enzyme called poly (A) polymerase then synthesizes the poly (A) tail at the 3' terminus of the mRNA.

[6]Almost all introns have a GU dinucleotide pair at their 5' boundary, and an AG dinucleotide pair at their 3' boundary. These dinucleotides form part of a larger consensus DNA sequence that overlaps the intron-exon boundaries. Pre-mRNA splicing operates towards at least 95% of the primary transcript pool.

[7]This is only one of several ways in which the same 'gene' or region of DNA can lead to the production of many different protein products in the same organism.

that generates a large number of mRNA and protein types from the surprisingly low number of genes. Unlike variable promoter activity, alternative splicing changes the structure of transcripts and their encoded proteins, thereby also affecting the protein binding properties, intracellular localization, enzymatic activity, stability, and post-translational modifications. The magnitude of the effects of alternative splicing ranges from a complete loss of protein function, acquisition of a new function, to very subtle modifications in function. Evidence is now accumulating that alternative splicing coordinates physiologically meaningful changes in protein expression and is a key mechanism to generate the complex proteome of multicellular organisms (Stamm et al., 2005). In the most extreme case of alternative splicing described to date, the Down's syndrome cell adhesion molecule (Dscam) gene alone could potentially encode more than 38,000 different protein isoforms (Zipursky et al., 2006).

Additionally, less understood processes act on the RNA transcript prior to translation. These are collectively called *RNA editing* and result in sequence modifications of the original RNA molecule. Alterations can include substitutions, insertions or removal of nucleotides and bases. RNA editing can be regulated in a developmental stage or tissue-specific manner.

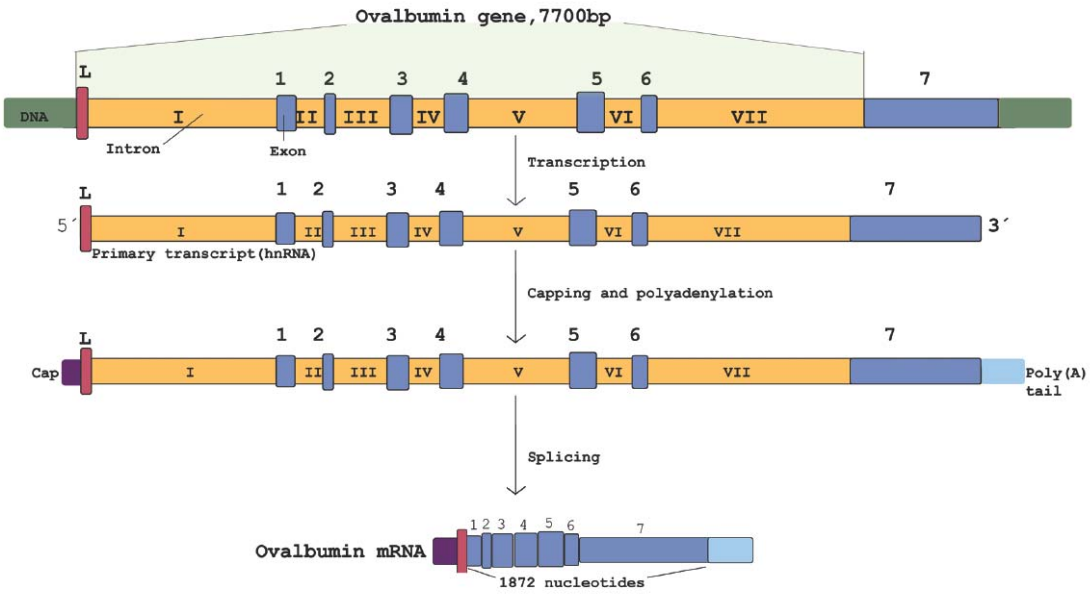

Figure 3.2. Processing of the primary RNA transcript for the gene encoding the protein ovalbumin. The intron regions (yellow) are spliced out, leaving only the exon regions (blue), which code for the protein in the final transcript.

1.3 Stability of mRNA transcript

The initial mRNA transcript needs to survive enzymatic degradation during modification, transport and translation. The limited lifetime of RNA transcripts allows a cell to change its pattern of protein synthesis continuously to changing physiological needs. Several types of molecules affect the stability of the RNA transcript. A particularly interesting group of regulatory molecules are small (19–28 nucleotides long), *non-protein encoding RNA molecules*. Such molecules are derived from cleavage of double-stranded RNAs (dsRNAs). Small RNAs can induce *gene silencing* through specific base pairing (binding) with the targeted mRNA transcript, thereby preventing protein expression. Small RNA-mediated gene silencing has been

observed in a number of eukaryotes for almost two decades, but the fundamental role of small RNA molecules in regulating gene expression has been unravelled only recently (Mattick 2003).

Degradation rates of mRNAs are important determinants of transcript availability for protein synthesis. The degradation rates of mRNAs differ in different cell types: In the gut bacterium *E. coli*, typical mRNAs' half-lives are *c.*15 minutes. In mammalian cells, unstable mRNA has about the same half-life as the *E. coli* mRNA, while stable mRNA such as the transcript of the β-globin gene has a half-life that exceeds one day. The control of mRNA degradation is an important component of the regulation of gene expression since the concentration of mRNA is determined both by the rates of transcription and rates of decay.[8]

1.4. mRNA transport to the cytoplasm

Once RNA processing is complete, mature mRNAs are exported to the cell's cytoplasm, where they serve as the blueprints for protein synthesis by ribosomes. Specific mRNAs may be directed to and anchored at specific subcellular locations, where they may be temporarily withheld from the translation apparatus and have their 3' ends trimmed or extended. From there the modified RNA may associate with other mRNAs encoding proteins of related function, and be scrutinized by protein complexes that serve as 'the quality-control police'. Hence, mRNAs in multiple cell types are subject to a diverse array of regulatory activities affecting essentially every aspect of their short lives and contribution to protein synthesis.

Throughout their existence, mRNAs are escorted by a complement of proteins and small non-protein coding RNAs (e.g. miRNAs), some of which remain stably bound while others are subject to dynamic association. Together with mRNA, these constitute *the messenger ribonucleoprotein particle* (mRNP). Individual mRNP components can be thought of as adaptors mediating the mRNAs' activity. Some adaptors make positive interactions and thereby serve as activators of a particular process, whereas others disrupt the positive interactions and act as repressors. By containing binding sites for diverse adaptors, individual mRNAs can respond to a myriad of regulatory inputs, allowing their expression to be selectively fine-tuned in response to changing conditions. The result is an elaborate web of regulatory networks of equal, if not greater, complexity to those controlling initial mRNA synthesis.

Box 3.1 Examples showing how the complex characteristics of transcription affect the understanding of the biology of GMOs

a. *Promoters*. A lack of in-depth understanding of promoter regulation and activities has led to the frequent insertion of strong promoters from pathogenic microorganisms and viruses into genetically modified (GM) plants. For instance, the use of the 35S CaMV plant virus promoter leads to a continual expression of the transgenes in the

[8]Two general pathways of mRNA decay have been described in eukaryotes. Both pathways share the exonucleolytic removal of the poly(A) tail (deadenylation) as the first step. In one pathway, deadenylation is followed by the hydrolysis of the cap and processive degradation of the mRNA by a 5' exonuclease. In the second pathway, the mRNA is degraded by a complex of 3' exonucleases before the remaining cap structure is hydrolyzed.

GM plant; the promoter can be active in a range of other organisms (Myhre et al. 2006).

b. *Enhancers*. The introduction of viral DNA sequences containing an enhancer into a GM plant can lead to unexpected results such as a change in the transcription of other unrelated genes. Recent studies provide evidence that the CMV enhancer may activate other unrelated promoters. Introduced genetic material may thus produce unexpected changes in expression of various genes localized far away from the transgene insert site (D'Aiuto et al. 2006).

c. *Transcript length variability*. Inefficient termination of transcription in a GM soybean variety led to the presence of various unexpected transcripts, and potentially also proteins (Rang et al. 2005).

---

2. Regulation of mRNA translation

After the processing of the mRNA, the translation machinery localized in the *ribosomes* converts the RNA information into the specified protein. The proteins are produced by ribosomes reading the *codon triplets* of the mRNA strand. The codon triplet is a sequence of three bases in the RNA that gives instructions to ribosomes to produce a specific individual amino acid to be assembled into a linear amino acid strand that makes up the protein. The individual amino acids are transported to the ribosome by *transfer RNAs*, small RNAs that are specialized in providing each of the 20 naturally occurring amino acids. The genetic codes of these triplets are universal for all organisms (Figure 3.3) although species-specific preferences on codon usage exist when there is more than one codon specifying a given amino acid (redundancy).

	U		C		A		G		
U	UUU UUC	Phenyla-lanine	UCU UCC UCA UCG	Serine	UAU UAC	Tyrosine	UGU UGC	Cysteine	U C A G
	UUA UUG	Leucine			UAA UAG	Stop codon Stop codon	UGA	Stop codon	
							UGG	Tryptophan	
C	CUU CUC CUA CUG	Leucine	CCU CCC CCA CCG	Proline	CAU CAC	Histidine	CGU CGC CGA CGG	Arginine	U C A G
					CAA CAG	Glutamine			
A	AUU AUC AUA	Iso-leucine	ACU ACC ACA ACG	Threonine	AAU AAC	Asparagine	AGU AGC	Serine	U C A G
	AUG	Methionine; start codon			AAA AAG	Lysine	AGA AGG	Arginine	
G	GUU GUC GUA GUG	Valine	CGU GCC GCA GCG	Alanine	GAU GAC	Aspartic acid	GGU GGC GGA GGG	Glycine	U C A G
					GAA GAG	Glutamic acid			

Figure 3.3. The codons specifying the amino acid compositions of proteins. Genetic information in genes becomes encoded into mRNA in three-letter units known as codons, comprised of the bases uracil (U), cytosine (C), adenine (A), and guanine (G).

The regulation of gene expression at the level of translation is an important, but still not completely understood process. Several recent studies using comparative proteomic profiling of cells have documented a lack of correlation between the mRNA level and composition and the corresponding protein levels of numerous genes. This indicates that post-transcriptional control is more important in the regulation of the protein content of a cell than often assumed. Regulation at this level allows for an immediate and rapid response to changes in environmental, physiological or pathological conditions (for example, heat shock, oxygen deprivation, pollution with *endocrine disrupters*, nutrient deprivation). In eukaryotes, translation is divided into three distinct phases *initiation, elongation* and *termination*. Although all three phases are subject to regulatory mechanisms, under most circumstances the rate limiting step is initiation.

- *Initiation.* A single mRNA transcript can have several translation initiation codons, thus several lengths of a protein can potentially be translated from a single mRNA transcript. A small, yet growing number of mammalian mRNAs have been shown to initiate translation from other sites than the standard AUG nucleotide start codon (Figure 3.3). These start codons may be downstream in frame or out of frame AUG or CUG codons. Translation initiation on such mRNAs results in the synthesis of proteins with different sizes (i.e. harbouring different amino terminal domains), potentially conferring distinct protein functions.[9]

- *Elongation.* The straightforward codon-by-codon translation of an mRNA is looked upon as the standard way in which proteins are synthesized. An increasing number of unusual elongation events are, however, being discovered. One of them is *frame shifting*, a process occurring when a ribosome pauses in the middle of an mRNA, moves back one nucleotide or, less frequently, forward one nucleotide, and then continues translation. The result is that the codons that are read after the pause are not contiguous with the preceding set of codons: they lie in a different triplet codon reading frame.

Spontaneous frame shifts occur randomly, are commonly non-functional and perhaps deleterious, because the protein synthesized after the frame shift has the incorrect amino acid sequence. However, not all frame shifts are not spontaneous. Some mRNAs utilize *programmed frame shifting* to induce the ribosome to change to a specific point within the transcript. Programmed frame shifting occurs in all types of organisms, from bacteria through to humans, as well as during expression of a number of viral genomes.

[9]The biological significance of the non-AUG alternative initiation is demonstrated by the different subcellular localizations and/or distinct biological functions of the protein isoforms translated from a single mRNA. Use of non-AUG codons appears to be governed by several features, including the sequence context and the secondary mRNA structure surrounding the codon (Touriol et al., 2003).

– *Termination*. Termination signals of protein synthesis are encoded in the gene and are also present in the mRNA transcript in the form of three different base triplets referred to as termination, stop or nonsense codons[10]. Inefficient translation termination can lead to variation in the size of translated protein products, which in turn may result in new proteins with unexpected biological functions.

Box 3.2 Examples showing how the complex characteristics of translation can affect the biological understanding of GMOs

a. Given the unique combinations of factors regulating protein production from each mRNA transcript, it is clear that changing the cellular environment of a given mRNA transcript (as done in GMOs) will affect its stability and translational properties.
b. Alternative translation start codons that are normally not recognized in one organism may become active when the gene is modified and inserted into another organism. The result is that translation of certain gene products might be turned on, off, or up- or down-regulated abnormally within the GM recipient cell.

---

3. Regulation of protein activity and stability

3.1 Protein folding, cleavage and chemical modification

After the protein is produced by the cell, the protein undergoes a series of modifications to its structure to ensure that it functions properly and that it is directed to the correct region of the eukaryotic cell. Such *post-translational modifications* are essential processes in the regulation of eukaryotic protein functions. The types of modifications that occur can have dramatic effects on the bioactivity, specificity and stability of the modified protein. Four types of post-translational processing are common:

– *Protein folding*. The protein emerging from the ribosome machinery may require the assistance of specialized proteins called *chaperones* to become folded into its functional 3-dimensional structure.

– *Proteolytic cleavage*. Some proteins are cleaved by enzymes called *proteases* that may remove segments from one or both ends of the polypeptide chain, resulting in a shortened active form of the protein. Alternatively, proteases may cut the polypeptide into a number of different segments, all or some of which are biologically active.

[10]When a stop codon has been translocated into the ribosomal A-site by the action of elongation factors, it is decoded at the small ribosomal subunit. The chemical reaction that is triggered by a stop signal leads to cleavage of the ester bond between the peptidyl and tRNA moieties of the peptidyl-tRNA complex. This occurs within the large ribosomal subunit at the peptidyl transferase centre (PTC) of the ribosome. How a stop signal can be transduced from the small to the large ribosomal subunit and trigger hydrolysis of peptidyl-tRNA remains unknown, and alternative hypotheses are still being discussed in the literature (Mitkevitch et al., 2006).

– *Intein splicing.* Inteins are intervening sequences in some proteins, similar to introns in mRNAs. They have to be removed, and the *exteins* (similar to exons) ligated in order for the protein to become functional.

– *Chemical modification.* Individual amino acids in the polypeptide chain may be modified by attachment of new chemical groups. The modifications may influence the folding of the proteins and their interactions with other proteins.

Chemical modifications are often introduced on the surface of the proteins at different amino acid sites. Modifications at single or multiple sites occur in different ways, by inserting additional side chains. Some examples include glycosylation, phosphorylation, acetylation, methylation, ubiquitination, sumoylation, and citrullination. Multi-site modification on a protein constitutes a complex regulatory programme that resembles a dynamic 'molecular barcode'. The chemical modification patterns hence encode 'loss-of-function' and 'gain-of-function' processes that affect bioactivity and protein stability. Recruitment of these modifying groups on proteins is often modulated by chemical modifications occurring at neighbouring and distant sites on the affected molecule. Multi-site modifications thus coordinate intra- and inter-molecular signalling for the qualitative and quantitative control of protein function.[11]

One of the most common and least understood post-translational chemical modifications of proteins is *glycosylation.* Proteins may be glycosylated with a bewilderingly array of complex N- and O-linked sugar molecules.[12] Glycosylation of proteins is highly regulated and changes during differentiation, development, under different physiological and cell culture conditions, and in disease. When a given transgene is expressed in different organisms, the glycosylation patterns may be very different, and this may add or retract biochemical and biological activities from the proteins. This may be the case even for the same gene expressed in different crop plants (see example in Box 3.3).

3.2 Higher order protein interactions

Many proteins have multiple functions that may be exerted by discrete parts of the proteins called *active domains* or *active sites.* Most proteins are conceived as globular beads on a string, where the 'beads' represent domains that range in length from 50 to 250 amino acid residues. Each domain may perform a specific biochemical function. Some protein activities, however, are performed at the interface between two or more domains situated on two different protein molecules. The structure is called a *homodimer* if the two molecules are the same, otherwise it is a *heterodimer*. There may be multiple, different proteins in the active complex. The self-association of proteins to form dimers and higher-order oligomers (the formation of protein chains consisting of many shorter proteins linked together) is a common phenomenon. Dimerization and oligomerization requirements for protein function allow regulation to occur by interfering in the assembly process. Whether and to what extent transgenic proteins engage in such interactions are unknown.

[11]Post-translational modifications are often modulating and coordinating the activities of transcription factors (discussed earlier). Chemical modifications can rapidly and reversibly regulate virtually all transcription factors, including subcellular localization, stability, interactions with co-factors and transcriptional activities, and thus have important regulatory function on protein production as well, illustrating the circular and multi-dimensional regulation of gene expression.

[12]Originating from the regulated activity of enzymes within the endoplasmatic reticulum and Golgi apparatus of eukaryotic cells.

3.3. Regulated protein degradation

Protein degradation (*proteolysis*) is a means to remove obsolete or damaged proteins. Proteolysis is mediated by specific enzymes called proteases, which vary from small proteins such as extracellular trypsin and the intracellular caspases to large, ATP-dependent, multifunctional proteases called *proteasomes*. Protein degradation is mediated by conjugation of the protein to the signal molecule *ubiquitin* that is regulated by specific degradation signals (*degrons*) in short-lived proteins. Regulated ubiquitin-dependent degradation processes are thought to play a major role in controlling the levels and menus of intracellular proteins, a function previously thought to be mediated almost exclusively at the transcription or translation stages.

Box 3.3 Examples showing how the complex regulation of protein activity and stability affects the biological understanding of GMOs

The effects of variable extent and type of post-translational processing are important when considering the biological properties of GMOs.

a. Different host organisms of a particular gene may process the resulting protein in non-similar ways; that can affect protein activity, stability and composition. For example, recombinant human insulin, for the treatment of diabetes, is produced in GM bacteria and yeasts. However, because the insulin protein does not fold to the active conformation when produced in a microorganism, an extra enzyme must be added to re-fold the protein before it can be administered to humans.

b. The changed glycosylation patterns that can occur in the recombinant proteins produced by GMOs are of critical importance. Glycosylation profoundly affects the protein's biological activity, function, clearance from circulation, and antigenicity. The cells of non-human species, particularly plants, do not glycosylate their proteins in the same way as human cells do. Different plants may even glycosylate the same protein in different ways. In many cases, the differences are profound. Furthermore, there may be important differences in the processing and degradation of glycosylated proteins between mammalian and plant cells. Thus, expressing recombinant proteins in novel cell contexts may substantially alter the biological properties of the proteins produced by the transgene (Prescott et al. 2005).

---

4. Genome-scale factors affecting gene expression

4.1 Genome structure

Chromatin (Fig. 3.4) is one of the hallmarks of eukaryotic life. DNA in eukaryotes is tightly associated with a group of proteins called *histones*. Two molecules each of the four different core histones (H2A, H2B, H3, and H4) form a histone octamer, around which 146 bp of DNA are wrapped to form a nucleosomal core particle. A linker histone (H1, H5 and a number of histone-like proteins) binds to the free ('linker') DNA between two nucleosomal core particles, and this finally makes up the *nucleosome*. Given the length of the haploid human genome (3.3×10^9 base pairs), every diploid cell nucleus contains roughly 5×10^7 nucleosomes. Any molecular process entailing genomic DNA or the nucleus by default provokes or depends on

chromatin structural dynamics on various space and timescales. Chromatin dynamics are a result of changes in the properties of the chromatin constituents themselves or in the nuclear environment (Benecke, 2006). The transgenic process itself, with integration of foreign DNA and insertional mutagenesis as a key element, may change chromatin dynamics, and hence influence the expression of the endogenous genes profoundly (Recillas-Targa, 2006).

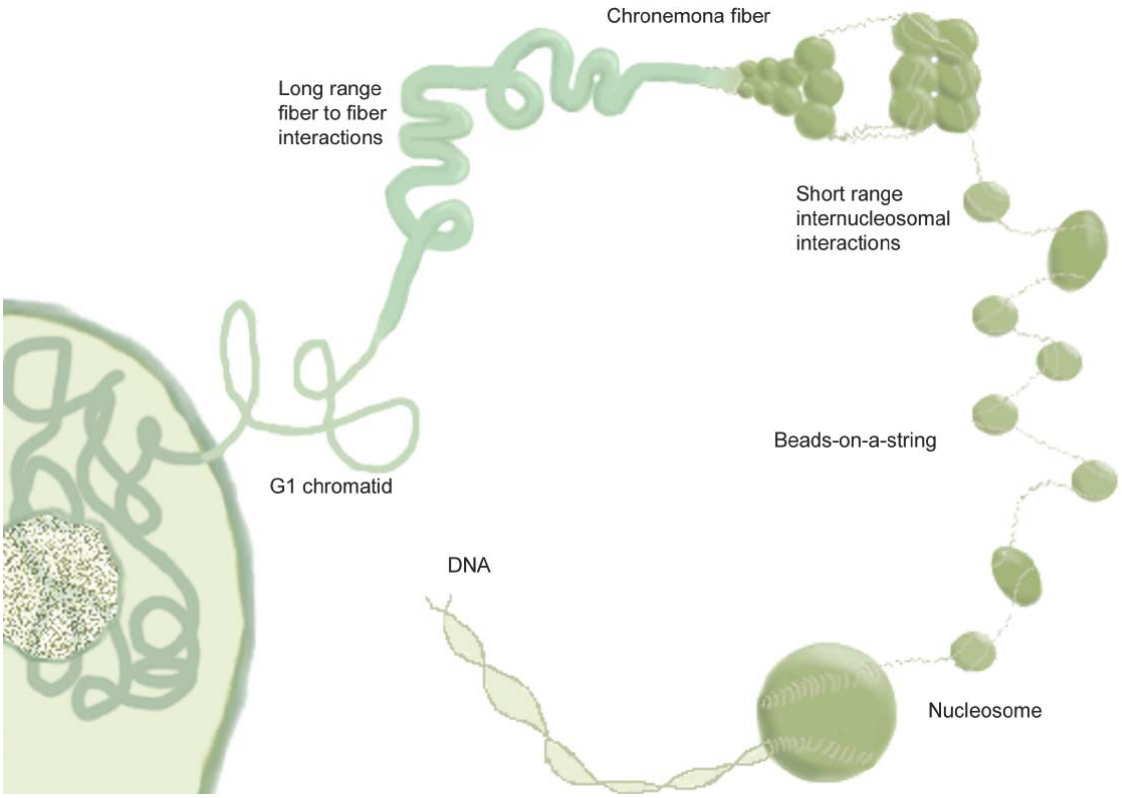

Figure 3.4. A schematic presentation of chromatin components and topology.

Upon transcriptional activation, the DNA strand with the activated gene can loop so that the gene will be present in nuclear locations enriched in the enzyme RNA polymerase and the larger transcriptional machinery, known as *transcription factories*. Gene promoters have been observed to be in close physical contact with enhancer elements upon transcription. Thus, intricate networks of DNA strands, their regulatory elements and the transcription factories will form during gene transcription. Thus, positional effects of genes in a genome can be conceived of as important for gene regulation and function.

Chromatin is also subject to a diverse array of chemical modifications that can regulate access and transcriptional activity of the underlying DNA.[13] Introduction of methyl-groups (methylation) is very common at specific DNA locations in most organisms. Methylation causes two major effects: 1) to displace transcription factors that normally bind to the DNA, and 2) to attract methyl binding proteins that are

[13]The complete DNA strand in a chromosome, called the chromatin fiber, is composed of multiple specialized domains, each of which contains a distinct subset of proteins such as nucleosomes, linker histone variants and non-histone proteins.

functionally associated with gene silencing. DNA methylation is discussed in more detail in Chapter 5.

5. Synthesis and conclusions

What can we surmise from the information on complex regulatory networks governing gene expression presented here? The dynamic changes taking place within a eukaryotic genome, and the dynamic interplay between the genome and its outside world, is slowly coming to light (Leitch, 2007). Clearly, contemporary science is evolving a picture of the genome and its regulation that is much different from the reductionist paradigm (DNA-RNA-protein) that has guided the biological understanding of DNA function over the last half century (including the foundational basis of genetic engineering). It is important to recognize that much of what is presented in this chapter is a conceptually and mechanistically framed understanding of genes and genome developed independently of the environmental or biological *context*. By placing the genes (DNA and downstream regulatory processes) in context, new layers of information coming from both within and outside the cell, influence these fundamental processes of gene activity. As a result, the prevailing paradigm of a mechanistically and deterministically defined gene regulation that forms the conceptual basis of genetic engineering is now widely understood to be invalid. Such a static view of transgenes as inert to their genomic and regulatory context needs to be revised not only in theory, but also in practical terms. This is particularly germane to developing scientifically sound GMO products and policies. Currently, significant levels of uncertainty and gaps in knowledge in the behaviour of transgene expression in the GMO itself require greater investments into biosafety research in order to assure their safe use.

We have yet to develop models, concepts and metaphors that can inform us about how this molecular orchestrating comes about. If the organism is not caused by its molecular parts, but these parts themselves are orchestrated in concert by the organism and ecosystem as a whole, there are a lot of concepts and thinking habits that have to be reconstructed. The field of ‘systems biology’ attempts to take a more holistic approach to understanding organismal gene expression and development. A revised view of the factors governing gene expression and, hence, organismal properties will also impact the fundamental rationale of genetic engineering; namely the Central Dogma inspired idea that the organism can be precisely controlled by the engineer (e.g. adding one gene for the addition of one single trait, without further genomic effects).

The Central Dogma represents the guiding idea underlying genetic engineering. This idea was conceptualized some 35 years ago when understanding of gene expression and function was in its infancy. The Central Dogma does not deal with the complex interactions leading to protein production as we observe them today. *It seems now more relevant to think of genes as the tools of the organism, rather than as the cause of the organism.*

We observe from genetic engineering that the introduction of a new gene into a new host or into a new location in the genome of the same hosts can:

- Significantly alter the phenotype of the host organism beyond what is expected from the inserted/moved trait. This can, for instance, occur by up-regulation or down-regulation of production and chemical composition of unrelated gene products.

- Result in one or more proteins different from the protein produced in the original organism (from where the gene was found).

These changes can occur in GMOs without the genetic engineers being able to *a priori* predict the outcome. Thus, the multiple levels of environmental and cellular interactions guiding gene expression and protein functionality are not represented in the narrow interpretation of the Central Dogma. The relevant questions guiding further development, and investigation, of the safety and monitoring of GM crops in the environment require an adoption of a more holistic concept of (trans)genes in their new contexts. A critical analysis of the concepts, methods and paradigmatic models of thinking that have predominated the field of transgene biology is required. The emerging new holistic methods and models in transgene biology will not replace the reductionistic approaches, but will complement them with the multidimensional interactions between genomes and their environments. With this broader understanding, we can start asking important biosafety questions that include context and changing conditions.

How are the GMO, transgene expression and recombinant protein compositions affected when an organism is put into new organismal and environmental contexts? How do these multi-scale changes interact? These basic questions require a methodological approach that considers all levels of biological organization and ecological interactions.

What this chapter aims to illustrate is that the simplistic, unidirectional and deterministic cause-and-effect understandings gene expression, which forms the basis for genetic engineering, has become scientifically invalid. The connection between genotype and phenotype is not solely determined by DNA, but is dependent on a multilayered informational network of, for example, proteins, RNA, genomes, and environmental stimuli, all of which are context dependent, and their outcome cannot yet be predicted *a priori*. With this in mind, it can be seen that *a single gene delivers only part of the identity and function of a protein*. In this connection it has to be remembered that the development of the first generation of GM plants was based on the knowledge of the 1970s and 1980s.

Lastly, the random insertion of foreign genes into an organism during transgenesis does not comprehend the importance or effect of the organizational placement and interacting factors upon the inserted gene, nor its long-term implications for the host cell, environment, or interacting species. Genes, including transgenes, are not autonomous units and should not be treated as such in a scientific or even regulatory sense. For example, the complex pathways to protein synthesis discussed mean that a number of different recombinant proteins may be produced from the same transgene, leading to changes in allergenic potential, target bioactivity, or influence on host biochemical composition, function and survival. Currently, there is little data and research that shed light on these important processes as they occur in GMOs. The

research field of gene ecology seeks to address these fundamental knowledge gaps to improve the biological understanding of, and hence, the safety of GMOs.

Some biosafety issues related to transgenic organisms are further discussed and exemplified in Chapters 8–14.

Resources

Karp GE. Cell and Molecular Biology. Concepts and Experiments. John Wiley & Sons Inc., 2005 (ISBN 0-471-65665-8, Wiley International Edition)

Campbell NA and Reece JB. Biology., Seventh Edition. Pearson-Benjamin Cummings, 2005 (ISBN 0-321-26984-5, International Edition)

References

Benecke A. (2006) Chromatin code, local non-equilibrium dynamics, and the emergence of transcription regulatory programs. Eur. Phys. J. E 19: 353-366

Kapranov, P, Willingham A, T, Gingeras, T. R. 2007. Genome-wide transcription and the implication for genomic organization. Nature Rev. Genet. 8, 413-423.

Leitch, AR (2007) The dynamic interplay between genomic DNA and the outside world. Heredity, 98: 61-62

Mattick, J. S. . (2003). Challenging the dogma: the hidden layer of non-protein-coding RNAs in complex organisms. Bioessays 2: 930-939, 2003.

Mitkevitch VA et al. (2006) Termination of translation in eukaryotes is mediated by the quaternary eRF1*eRF3*GTP*Mg2+ complex. The biological roles of eRF3 and prokaryotic RF3 are profoundly distinct. Nucleic Acids Research 34: 3947-3954

Recillas-Targa F. (2006) Multiple strategies for gene transfer, expression, knockdown, and chromatin influence in mammalian cell lines and transgenic animals. Mol. Biotechnol. 34: 337-354

Stamm S et al. (2005) Function of alternative splicing. Gene 344: 1-20

Touriol C et al. (2003) Generation of protein isoform diversity by alternative initiation of translation at non-AUG codons. Biology of the Cell 95: 169-178

Zipursky, SL et al. (2006) Got diversity? Wiring the fly brain with Dscam. Trends Biochem Sci 31: 581-588