

# Regulation of gene expression in Health and Disease

By Alexandre M. Carmo and Alexandra Moreira, Institute for Molecular and Cell Biology, Porto, Portugal



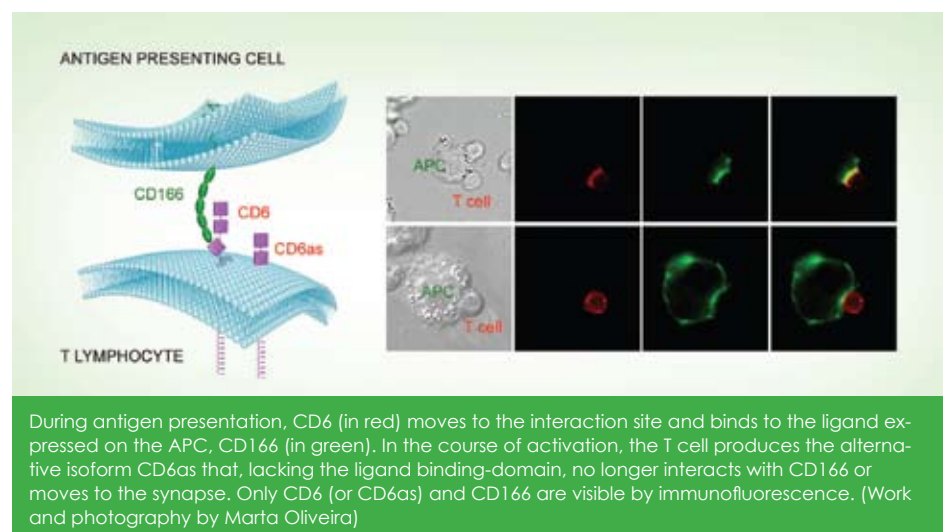
**THE FULL SEQUENCING** of the Human genome opened new avenues for research in biology and medicine, but also brought with it a few surprises. Instead of making our prospects of scientific development simpler, it started a whole new range of discussions and controversies. All of the sudden, we were “downscaled” to the level of most other mammals. We knew that our genes were very similar to those of apes (98% similarity) and even rats and mice (85%), but we now realized that we have virtually the same number of genes as most other mammals. More disconcerting perhaps, we only have a mere 50% higher number of genes than worms. That is, for each 3 genes we have, worms have 2, despite the fact that our genome is 30 times bigger than theirs. And we are, in fact, more complex organisms.

Molecular biologists were, however, not too astounded. Having been studying for decades the mechanisms of alternative processing of messenger RNA (mRNA), i.e., the different modes by which a transcript of a gene can be modified to generate diverse protein products from a single transcription unit, and started their own global projects of ESTs sequencing,

it seemed only natural that a huge diversity of structural proteins, enzymes and cellular receptors could arise simply due to the complexity of the mechanisms of gene expression regulation. Complexity could now be explained by the variability in the number of proteins coded by a single transcription unit, and not solely by the crude number of genes within a genome. And in fact, the first genome-wide screens of alternative isoforms of mRNAs indicated that complex systems of mammalian physiology, such as the nervous and the immune systems, were particularly rich in generating alternative mRNA products.

While it is apparently true that over two thirds of human genes have alternative forms, most of them arising from alternative splicing of exonic sequences, the actual number of different protein isoforms resulting from translation of alternative spliced messengers, and having a clear distinct biological function, is disappointingly low. In immunology, the well-studied cases are just a handful, such as the various isoforms of the broadly expressed leucocyte common antigen (CD45), whose different extracellular isoforms correlate with the developmental or activation stage of lympho-

» The full sequencing of the Human genome opened new avenues for research in biology and medicine, but also brought with it a few surprises «





cytes. Another well-characterized example is that of the cell adhesion molecule CD44, a T lymphocyte receptor that is expressed in at least 6 different protein isoforms, despite having more than 20 mRNA varied transcripts. The differential expression of CD44 isoforms correlates with the malignancy of certain tumours, and yet no clear mechanistic characterization has been advanced on how the several isoforms work at the molecular level.

The converse is also true, sometimes we are presented with situations where we can clearly assign a differential molecular function for different isoforms, but their physiological significance or biological outcome are still to be determined. A remarkable illustration can be perhaps given by the T cell surface receptor CD6 that we study in our laboratory. When antigen presenting cells (APC) interact with T lymphocytes to display foreign antigens to the T cell receptor, a number of other adhesion and co-receptor molecules stabilise the APC:T cell interface through a dynamic structure called the immunological synapse. CD6 targets to the immunological synapse due to its interaction with the receptor CD166 expressed at the surface of the presenting cell. During this

process, however, signalling to the T cell nucleus induces T cells to stop producing full-length CD6, and replace it with a CD6 isoform that lacks only the part of the molecule that binds to CD166. The resulting CD6 variant also gets expressed at the surface of T cells, but no longer moves to the synapse. Why this mechanism happens is still to be determined. One would think that if the molecule is no longer required, then the cheapest way for the cell would be to stop producing it altogether.

So if only a few known examples pose such complex questions with no clear answers in the horizon, how can we then tackle the thousands of other possible examples of differential gene function for which we are accumulating data only at the mRNA level? Clearly, high-throughput experimentation will be required to, first confirm the protein correlation to the RNA forms, and then, to establish relationships between changes in gene expression profiles and modifications with cellular or organism behaviour. But undoubtedly, one will have to present strong and credible cases for the different gene/protein products so that the whole of the scientific and medical communities can support this gigantic endeavour.

One of the fields with the biggest potential to benefit from the in-depth studies in alternative mRNA processing is unquestionably the area of human genetics and genetic disorders, both acquired as well as inherited. It is evident that the possibility of introduction of errors in the pre-established splicing machinery can result in the production of gross faults in the biology and biochemistry of the cell. For example non-coding nucleotide expansions originate a number of disorders, including several cancers and neurological diseases by messing with the sequence sites where splicing factors are known to process the RNA molecule. Metachromatic leukodystrophy, spinal muscular atrophy, Menkes disease, adenosine deaminase deficiency and Sandhoff disease all result from the introduction of mistakes in intron/exon boundaries, which are then incorrectly recognized by the splicing machinery. Microarray platforms are now being developed to account for the detection of changes in the splicing patterns of several genes correlated with known diseases. Together with state of the art computational programs, the correct mRNA products of a gene expected to be produced in different organs can be mapped ➤



*Drosophila melanogaster* is a useful biological experimental model, where the expression of genes can be easily regulated. The fly on the bottom had only one 3' non-coding polyadenylation site of one gene mutated, yet it developed abnormally. (Work and photography by Pedro Pinto)



and, detecting changes in splicing patterns, one can predict the development of diseases allowing for the establishment of protocols and strategies to an efficient therapy.

It is quite straightforward to appreciate that the production of alternative spliced-isoforms of a given gene may be translated into different protein products, each with the potential to have a different biological function. However, there are other forms of diverse mRNA processing whose outcome is not always that predictable. For instance, alternative polyadenylation, that is, the production of several mRNA molecules from the same gene, whose difference is only to have a poly adenosine tail inserted in different 3' ends of the mature RNA, is also common to about half of human genes. And, whereas in some cases the protein end products are different (for example, immunoglobulin class switch is primarily determined by a choice of different polyadenylation

(pA) sites that, in conjunction with alternative splicing, produces either soluble or membrane-bound immunoglobulins), in other cases the resulting mRNA transcripts differ only at the 3' untranslated region, having identical coding sequences. In short, many genes produce different mRNA molecules, which will then manufacture exactly the same protein. Why is this the case?

This is probably one of the most challenging areas for the proximal future. Undeniably, polyadenylation is a central feature of the whole gene expression apparatus, and mutations in pA sites are responsible for a number of well known disorders, such as thalassemias and hereditary thrombophilia. The extra sequences of the untranslated region present in some mRNA isoforms and not in others, and may have a role in determining which cell types are able to use that specific RNA for translation. Alternatively, these sequences may

be important for the correct transport and localization of the mRNA molecule within the cell. For example, in neurons some proteins should only be expressed at neurites, far away from the nucleus, so non-coding mRNA sequences may be required to bind to transporters that will carry the RNA molecule to the proper site, where, in the correct environment, translation may take place. However, most of the research tackling alternative polyadenylation does so in a mechanistic point of view, using mostly cell lines and in vitro systems. These do not advance much in the determination of a physiological function of a macromolecule whose function may alter completely the behaviour of an entire organ or organism.

Using the fruit fly *Drosophila melanogaster* as an experimental model, we are now addressing the function of alternative polyadenylation-based isoforms of genes involved in the cell cycle and development. We managed to express mRNA molecules lacking one or the other polyadenylation sites, and remarkably, in flies where we remove one of the pA sites of some important genes we can induce major defects in the development of the organism. Surprisingly enough, the levels of protein are usually close to normal in the whole of the fly, suggesting that only in small populations or in defined tissues, the mRNA defective of the ablated pA site can not be recognized and further process to translation. We are just starting to evaluate the precise mechanism. However, all the difficulties facing us ahead are nothing compared to the joy we will get in the next discovery.

It is estimated that only a tiny fraction of the total mRNA produced by a cell is employed in the translation into protein products. The rest could be regulatory products, debris or just plain noise. The truth is, only now are we beginning to understand the wonderful complexity of RNA biology. ■

» The truth is, only now are we beginning to understand the wonderful complexity of RNA biology «

