Jimmy Renteria

Doug Brutlag

Biochem 118Q

6/4/01

Cut, Print:

Our Emerging Understanding of Alternative Splicing

1. One gene—30,000 (possible) proteins

Advances in gene sequencing technology and informatics have heralded the dawn of a new era in biology: the era of genomics. Now it is possible to sequence whole genomes of organisms, compare them, annotate them, and probe them for polymorphisms. With the dawn of this age of new knowledge, however, have come new mysteries. Chief among these mysteries, for many organisms, is the question, "Where did all the genes go?" The human genome, in particular, is now predicted to contain roughly 35,000 genes—tens of thousands less than initially expected (Human Genome Consortium 2001). How is it possible that organisms as complex as human beings could be produced from only 35,000 genes?

This mystery has been popping up over and over again as we sequence the genomes of more and more organisms. Take *Drosophila*, for instance; its genome is shorter than that of *C. elegans* (Graveley 2001). How can this be, given that *Drosophila* evolved much later than *C. elegans*, and is, by all accounts, a much more complex organism? The answer to this and the mystery of humanity's

disappearing genes lies in the demise of the 'one gene/one protein' belief so long held in biology. Alternative splicing, the pre-translation editing of RNA transcripts to include or exclude particular introns, gives one gene the theoretical potential to code for hundreds, thousands, even tens of thousands of proteins. In higher eukaryotes, particularly, this device is used to regulate and produce multiple proteins from single genes.

2. What is alternative splicing?

Alternative splicing hinges on the presence of introns in an organism's genome, a requirement met by all eukaryotes. The presence of these introns necessitates the editing of mRNA transcripts before they can be translated, and it is this step of editing that makes alternative splicing possible. There are five known kinds of alternative splicing (see figure 1). A gene may contain multiple 5' or 3' splice sites, allowing for variation in exon length. A gene may also contain a full exon that is either included in or excluded from the final, edited transcript. These self-contained exons are called cassette exons, and there may be multiple cassette exons such that the splicing machinery must choose one before continuing on to the next exon. Finally, introns can be retained in the final, translated mRNA (Graveley 2001).
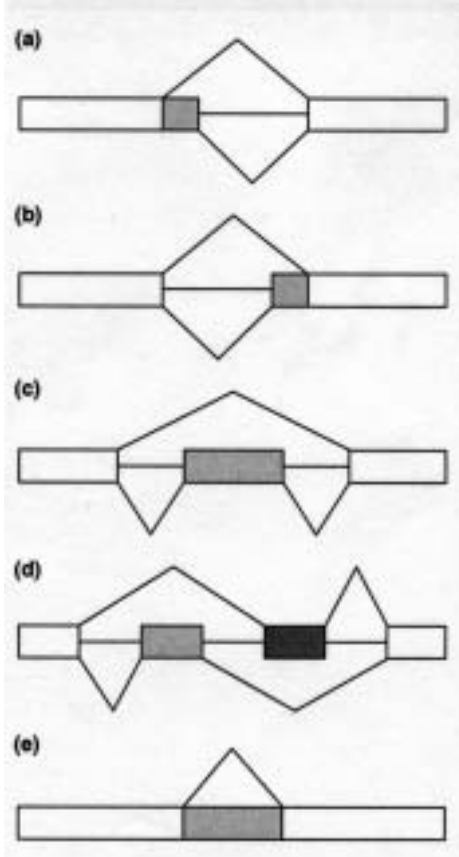
Figure 1
Taken from Graveley

Open boxes represent exons, while shaded boxes represent alternative splice exons or alternative splice sites. All lines above and below the boxes represent possible splice pathways.

(a) The first exon contains two alternate 5' splice sites
(b) The second exon contains two alternate 3' splice sites
(c) The middle exon can either be included or skipped
(d) One of two cassette exons will be expressed
(e) An intron may be retained for translation

A given gene may employ one, all, or any mixture of these devices to increase proteomic diversity. A startling example of this diversity can be found in the DSCAM gene of Drosophila. The protein encoded by this gene is known to function in the guiding of axons in nerve formation, and analization of cDNAs for the protein revealed differences indicative of alternative splicing. When geneticists looked at the actual DSCAM gene, they were amazed to find that exons 4, 6, 9, 17 were encoded by dozens of cassette exons. There are 12 cassettes that code for exon 4, 48 for 6, 33 for 9, and 2 for 17. Purely mathematically, the DSCAM gene is capable of producing 38,016 different proteins (Black 2000).

Though only 49 of the Drosophila DSCAM exon combinations have been observed thus far, this example illustrates the sheer capacity for proteomic dynamism from a single alternatively spliced gene. Current estimates are suspect at best, but so far at least 35% of

human genes are believed to be alternatively spliced(Black 2000). The mystery of the disappearing genes, then, is no mystery at all. Because of alternative splicing, genomic size can in no way be interpreted as indicative of proteomic or organismal complexity. The size of the human genome can even be seen as a testament to the importance of alternative splicing in it. Understanding the mechanisms by which alternative splicing occurs and is controlled are necessary steps in making sense of the genome, and linking genomics and proteomics.

The process of alternative splicing itself is somewhat similar to simple RNA splicing. It occurs co-transcriptionally, and begins with the binding of U1, a small nuclear ribonucleoprotein particle (snRNP) to the 5' end of the splice site, and U2 auxiliary factor (U2AF) to the 3' splice site. Splicing factor 1 also binds to the branch point, prepairing the pre-mRNA to the lariat-like structure characteristic of splicing (Smith and Valcarcel 2000).

So far, alternative splicing mirrors the typical splicing mechanism, so how does it take into account the competition of various, mutually exclusive splice sites? It is believed that a special class of serine-argenine rich proteins (SR) step in to affect selection of a given splice site or cassette over others. These proteins bind to specific RNA sequences called exon sequence enhancers (ESE). ESEs are nucleotide sequences found in exons, and are unique to each exon. Their individuality allows for great specificity in the splicing process, as only certain SR proteins will bind to certain ESEs. Binding of SR proteins to an ESE is believed to enhance recognition of the

upstream 3' splice site, making this mechanism crucial for selection of cassette exons (Lopez 1998).

3. Applications of alternative splicing

Though our understandings of the mechanisms of alternative splicing are rough, we already have sufficient information to begin targeting these pathways for regulation. ESEs, particularly, provide us with the opportunity to influence splicing. By binding to these short regulatory sequences, we could stimulate the selection of one splice variant, simultaneously inhibiting the production of another. But to what end would we apply this technology?

It is believed that 15% of genetic diseases are a result of mutations that affect splice sites (Philips and Cooper 2000). These diseases range from rare for inherited diseases (ALS) to common for such afflictions (spinal muscular atrophy occurs in 1 in 10,000 births) (Philips and Cooper 2000). Even the more common of these afflictions are relatively rare, however, and so are not appealing targets for which to develop a whole new class of drugs. Cancers, however, are another matter. American men have a 1 in 2 lifetime risk of developing cancer, women a 1 in 3 risk(American Cancer Society 2001). The market for chemotheraputic agents for cancer, then, is large and lucrative. Many cancers are also characterized by a shift in the ratios of alternative splice variants, making them an ideal target for developing drugs that regulate splicing(Mercatante and Kole 2000).

Apoptotic regulator and tumor suppressor proteins, in particular, exhibit abnormal splicing in cancers. The Bcl family of apoptotic regulatory proteins is well known to be involved in many types of lymphomas and invasive carcinomas. Overexpression of the Bcl-2α splice variant is believed to contribute to the resistance of some cancers to chemotherapy. Bcl-x is another anti-apoptotic signal that undergoes splicing. The Bcl-xL variant is strongly anti-apoptotic, whereas the Bcl-xS variant actually sensitizes a cell to undergo apoptosis. Bcl-x, then, is an attractive gene for regulation, since lowering the ratio of Bcl-xL to Bcl-xS would not only lower the cell's resistance to chemotherapy, it would actually sensatize the cell to undergo apoptosis more readily (Mercatante and Kole 2000).

Anti-sense oligonucleotides are being used in *in vitro* proofs of concept to demonstrate the effect of regulating these alternative splicing pathways. These oligonucleotides are a new class of drugs designed to hybridize with specific nucleotide sequences. This hybridization prevents the binding of other chemicals (like snRNPs and spliceosome proteins) to specific sequences in the genome. Anti-sense oligos targeted at a splice site in the Bcl-x gene has been shown to shift production towards the Bcl-xS pro-apoptotic variant in human cancer cell lines. The treated cells exhibited greater amounts of apoptosis, and were more sensitive to chemotherapeutic agents (Mercatante and Kole 2000). Similar proofs of concept have shown that splice regulations could potentially treat b–Thalassemia and cystic fibrosis (Sierakowska, Sambade et al. 1996; Mercatante and Kole 2000). Though early proofs of concept are by no means guarantees

that these drugs will ever reach market, they are encouraging and enlightening.

Our rudimentary ability to manipulate alternative splicing is by no means an indication that we have grasped its intricacies. More research in this area is needed, and is being conducted. Tantalizing hints exist that cell-specific splicing patterns may be at the root of many developmental processes (Lopez 1998). Probing these and other alternative splicing pathways will give us a clearer understanding of the proteomic significance and regulatory mechanisms of alternative splicing. Such and understanding, in turn, would lead us to better understanding of the human genome, and possibly provide us with new means of treating disease.

Works Cited

American Cancer Society (2001). Who Is at Risk of Developing Cancer? Online. http://www3.cancer.org/cancerinfo/sitecenter .asp ?ct=1&ctid=8&scp=8.3.1.40029&scs=4&scss= 3&scdoc=4004 1&pnt= 2&language=english . **2001**.

Black, D. L. (2000). "Protein diversity from alternative splicing: a challenge for bioinformatics and post-genome biology." Cell **103**(3): 367-70.

Graveley, B. R. (2001). "Alternative splicing: increasing diversity in the proteomic world." Trends in Genetics **17**(2): 100-7.

Human Genome Consortium, The (2001). "Initial Sequencing and analysis of the human genome." Nature **409**(15 Feburay 2001).

Lopez, A. J. (1998). "Alternative splicing of pre-mRNA: developmental consequences and mechanisms of regulation." Annu Rev Genet **32**: 279-305.

Mercatante, D. and R. Kole (2000). "Modification of alternative splicing pathways as a potential approach to chemotherapy." Pharmacol Ther **85**(3): 237-43.

Philips, A. V. and T. A. Cooper (2000). "RNA processing and human disease." Cell Mol Life Sci **57**(2): 235-49.

Sierakowska, H., M. J. Sambade, et al. (1996). "Repair of thalassemic human beta-globin mRNA in mammalian cells by antisense oligonucleotides." Proc Natl Acad Sci U S A **93**(23): 12840-4.

Smith, C. W. and J. Valcarcel (2000). "Alternative pre-mRNA splicing: the logic of combinatorial control." Trends Biochem Sci **25**(8):

381-8.