Genomics & Proteomics

Proteins are Often Spiced by a Slice of Alternative Splice

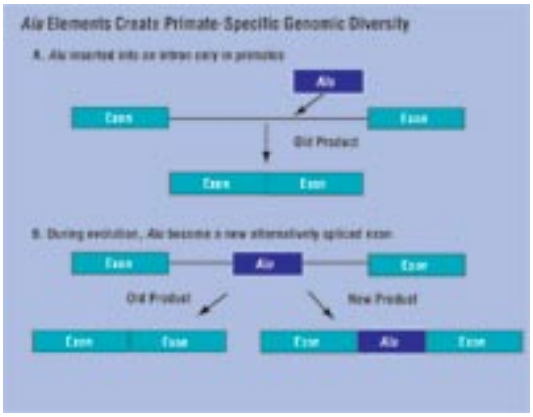
Once thought to number 150,000 genes, the human genome is closer to having 30,000 protein-coding genes. Alternative splicing is a significant factor in creating proteomic and functional complexity from this genomic minimalism.

By Vivien Marx

Who, except a small number of genomic clochards, would like to dig through genomic garbage? asked Wojciech Makalowski, PhD, from the Institute of Molecular Evolutionary Genetics at Pennsylvania State University, State College, Pa., in an article in the journal Science [W. Makalowski, vol. 300, pp. 1246-1247 (23 May 2003)]. The group of rather proud clochards, which includes Makalowski, is growing as scientists increasingly realize that within the code of the derided 'junk DNA' sequences lies insight into protein complexity as well as an enhanced understanding of certain genetic diseases.

This work is blurring the boundaries between exons and introns. It is showing researchers how dynamic the genome is and how the repetitive elements of so-called junk DNA are actually, as Makalowski puts it, "a genomic treasure" and "a source of 'ready-to-use-motifs' " increasing an organism's evolutionary flexibility.

In RNA splicing, the responsible cellular machinery, the spliceosome, recognizes a given intron's beginning and end, its 5' and 3' sites; exons are hooked together and introns are cut out. But the editing is not always as straightforward as that. Instead of genetic information flowing straight from DNA to one transcripted RNA to one protein product, as was long believed, alternative splicing patterns can lead to different mRNAs and thus varying proteins in the transcription of one gene. As Barmak Modrek, PhD, and Christopher Lee, PhD, of the departments of chemistry and biochemistry at UCLA and others have pointed out, 35 to 59% of human genes show evidence of alternative splicing, with a majority of the alternative mRNA leading to changed protein product and are thus often of functional importance. Fyodor



click the image to enlarge

Gil Ast, PhD, and his colleagues at Tel Aviv University were able to show the birth of an alternatively spliced exon by looking at human exonized *Alu* elements. They charted the molecular steps and the regulation of how an intronic *Alu* element, that are sequences of 300 nucleotides, turns into an exon. The genome is able to adapt to the insertion of the element. Splice variants are added to the transcriptome without disturbing the function of the original protein. (Source: Gil Ast, PhD, Tel Aviv University)

Proteins are Often Spiced by a Slice of Alternative Splice

Kondrashov and Eugene Koonin, PhD, of the National Center for Biotechnology Information at the National Institutes of Health say that alternative splicing is "thought to be a major source of functional diversity in animal proteins." [Kondrashov *et al., Trends in Genetics,* vol. 19, pp. 115-119 (March 2003)].

In complex genes, alternative splicing can generate dozens or even hundreds of different mRNA isoforms from a single transcript. As the scientists indicate, 80% of all alternative splicing in humans is caused by length difference alternative splicing (LDAS), in which an exon is either used or not used at all, leading to a long or a short isoform. "We asked the question, which one is the ancestral state," says Kondrashov who is now finishing his PhD in evolutionary biology at the University of California at Davis, "because we are wondering how alternative splicing might have evolved." Down the line, the hope is that knowing the origin contributes to understanding function.

Drawing on data from the Swiss-Prot database, the researchers compared the long human isoform deoxyhypusine synthase, known to contain an alternatively spliced chunk, to prokaryotes, which probably never had introns and yeast, which are thought to have lost introns. They also aligned the short isoforms of chicken and rat casein kinase with yeast and bacteria orthologs.

The comparison revealed that with human deoxyhypusine synthase, the long isoform corresponds to the ancestral state and with casein kinases, it is the short isoform that is the ancestral form.

click the image to enlarge

Alternative splicing can involve substituting one protein sequence segment for another one. It can also involve inserting/deleting a part of the protein sequence, called length difference alternative splicing (LDAS). About 80% of all cases of alternative splicing in humans are LDAS and it may be an important contributor to functional diversity of proteins. Part of understanding LDAS is figuring out which isoform corresponds to the ancestral state from which it evolved. (A) In

Regulating the change

Honing in on how alternative splicing is regulated, Gil Ast, PhD, of the department of Human Genetics and Molecular Medicine at Tel Aviv University in cooperation with Compugen Ltd., both of Tel Aviv, Israel, has been able to trace how intronic elements morph into new exons. (*Molecular Cell,* in press)

Ast focuses on *Alu,* elements around 300 nucleotides long, with two parts separated by a short A-rich linker and with a long poly-adenylate tail. As Ast says, *Alu* was inserted into the genome through retroposition. "Somehow during evolution it was duplicated from one monomer into two, so what we see currently are two monomeric forms of *Alu* and this is unique only to primates," he says. Five percent of human alternatively spliced exons are *Alu*-derived, he says. In the human genome, there are approximately 1.4 million copies of it, so it makes up nearly 10% of the entire sequence. "About half of them are in our introns; they pretty much love to get into our introns," he says. Half are in sense and half in antisense orientation, he says, the latter of which is probably the best orientation for the exonization process.

All *Alu* elements originated from one parental sequence, each of which was inserted into different locations in the genome," Ast says. *Alu* elements are connected to a number of genetic disorders.

Proteins are Often Spiced by a Slice of Alternative Splice

human deoxyhypusine synthase (DHYS_HUMAN), the long isoform corresponds to the ancestral state, it aligns with yeast and prokaryotic orthologs. (B) In the case of casein kinases, the short isofom from rat and chicken aligns with yeast and prokaryote orthologs. (Source: Fyodor Kondrashov and Eugene Koonin, PhD)

Two positions along the inverted *Alu* sequences usually act as 3' splice sites, proximal and distal AG dinucleotides, located four nucleotides from each other. There are two main *Alu* families: J, in which the distal AG is mainly chosen for alternative splicing and S, in which the proximal AG is selected. The scientists examined *Alu* sequences in order to find out, as Ast says, "the molecular steps that are required to take an intronic sequence and turn it into a new exon."

What they found was "quite amazing," he says. It was a seven-nucleotide motif, which is essential to control the level of alternative splicing. The scientists performed experimental verification to substantiate their bioinformatics analysis.

"The big picture is that the last seven nucleotides of the 3' splice site of the *Alu* element is a regulatory sequence that provides a weak 3' splice site, . . . essential to maintain alternative splicing." The splicing machinery may recognize it and the next time it will skip it, he says. So in the case of *Alu* it is the actual splice site that controls the level of alternative splicing. While current dogma asserts that the 3' and 5' splice sites are not regulatory sequences, with *Alu* the 3' splice site is indeed a controller.

Ast and his colleagues recently looked at 250,000 *Alu* elements in introns as well as those that have become exons, looking at the distribution of nucleotides. They found there are changes at only two positions. They alone distinguish intronic *Alu* elements from exonic *Alu* elements and they are at the 5' splice site. "So the only difference between those *Alu*s sitting in our introns . . . to those that became new exons are only two positions," Ast says. They also found a total of 8,000 *Alu* s in introns that most probably are one step away from becoming exons.

Don't call it junk

Alu -based changes are not necessarily just bad news, in that they lead to disease. Many insertions are nonfunctional or they produce proteins not currently essential for survival. But, Ast says, under

Glimpsing the Future ...

"Approximately a year ago, the mouse genome came out. It has turned out that mouse and human contain almost the same number of genes. Ninety-nine percent of the mouse genes have a human homolog . . . basically, we and mouse and every other mammalian organism produce the same proteins. The excitement about the exonization of *Alu* is the ability to explain what is unique in our genome." Gil Ast, PhD, Department of Human Genetics and Molecular Medicine, Tel Aviv University

"We have reported that technologies enabling the usage of alternative splicing information are emerging, but still far from being mature. More ways of implementing alternative splicing knowledge, such as in the fields of rational structure-based drug design, chemical screening, pharmacogenomics, pathways and toxicity analyses, will probably emerge in the future. This new window to the transcriptome and the proteome adds an additional dimension of accuracy and reduces the costly levels of uncertainty embedded in the [drug discovery] process." Erez Levanon, PhD, Rotem Sorek, PhD, Compugen Ltd., from Targets, vol. 2, no. 3, p. 114 (June 2003)

altering environmental conditions, they might offer an edge in some way. Junk DNA is a horrible term, he says. "This is most probably the reservoir of options we have for a rainy day," Ast says. It is a dynamic region, the playground of the genome, which can give organisms the opportunity to survive when the conditions change, he says.

"The excitement about the exonization of *Alu* is the ability to explain what is unique in our genome," Ast says. The mouse genome contains 2.5 billion nucleotides, the human genome around 3 billion. "The quarter of a billion nucleotides, [or] the difference between human and mouse, is mostly [due to] retrotransposable elements like *Alu* ," he says.

Understanding and exploiting these mechanisms is of interest to a number of biotechnology companies. ExonHit Therapeutics SA, Paris, seeks to apply alternative splicing knowledge for their "isoform- specific drug discovery." Compugen is combining computation and molecular biology to study alternative splicing in order to add therapeutically and diagnostically interesting proteins to their pipeline. One person's junk is, at least potentially, another person's gold.

Resources for Investigating Alternative Splicing
Emerging methods to study alternative splicing involve bioinformatics analysis as well as experimental techniques. Some experimental techniques include genomic tiling arrays and exon arrays to identify co-regulated exons, expression arrays with multiple probes help to identify exons that are differentially included. Researchers at the University of California in Santa Cruz developed splicing specific microarrays suited for studying yeast and studying splicing of every intron in parallel. (www.biology.ucsc.edu/faculty/ares.html)

Comparative genomic methods to detect alternative splicing include a search for discriminative features of alternative splicing developed at Kyushu University, Japan, and a specific alignment program developed at University of California Los Angeles.

Some databases of splicing information:
ASD, The Alternative Splicing Database Project:
Established by researchers at the European Bioinformatics Institute with the aim to help understand the mechanism of alternative splicing, the project currently comprises three databases: AltSplice, AltExtron and AEdb. AltSplice is a computer-generated data set of human transcript-confirmed splice patterns, alternative splice events, and the associated annotations. AltExtron is a computer-generated dataset of human transcript-confirmed constitutive and alternative exons and introns. AEdb is the manually curated database equivalent drawing its information from the literature. EBI plans, over the longer term, to solicit Web submission of data to AEdb from laboratory scientists.
(www.ebi.ac.uk/asd/altsplice/)

ProSplicer: a database of alternative splicing information set up by bioinformatics researchers and life scientists at National Central University, Taiwan, derived by using the alignment of proteins, mRNA sequences and ESTs against the human genomic sequences.
(http://bioinfo.csie.ncu.edu.tw/ProSplicer/)

SpliceNest: a Web-based graphical tool created by researchers at the Max-Planck-Institute for Molecular Genetics, Germany, for exploring gene structure, including alternative splicing, based on mapping EST consensus sequences (contigs) from GeneNest to the complete human genome. SpliceNest is integrated with GeneNest, a visualization of gene indices of human, mouse, *Arabidopsis* , zebrafish, *Drosophila* , and the SYSTERS protein sequence cluster set into one framework.
(http://splicenest.molgen.mpg.de/)
(http://genenest.molgen.mpg.de/)

AsMamDB: a database that aims to facilitate the systematic study of alternatively spliced genes of mammals. Version 1.0 of AsMamDB contains 1,563 alternatively spliced genes of human, mouse, and rat. The main information provided by AsMamDB includes gene's alternative splicing patterns, their structures, their chromosomal locations, their products and the tissues in which they are expressed.
(http://166.111.30.65/ASMAMDB.html)

ASAP, Alternative Splicing Annotation Project: an online database set up by bioinformaticists at UCLA for researchers to access and mine alternative splicing information. ASAP is based on genome-wide analyses of alternative splicing in the human genome (30,793 alternative splice relationships found) from detailed alignment of expressed sequences onto the genomic sequence. There are plans to expand the database to cover other genomes as well. ASAP provides information about gene exon-intron structure, alternative splicing, tissue specificity of alternative splice forms, and protein isoform sequences resulting from alternative splicing. The database can also help biologists design probe sequences for distinguishing specific mRNA isoforms and quantitate them.
(www.bioinformatics.ucla.edu/ASAP)

ASDB, Alternative Splicing Database: Established at Lawrence Berkeley National Laboratory to assemble information about alternatively spliced genes, their products and expression patterns. The current ASDB format is considered a prototype. It contains two sections: ASDB (proteins), which contains amino acid sequences, and ASDB(nucleotides) with genomic sequences.
(http://hazelton.lbl.gov/~teplitski/alt/)

The Intronerator: a collection of tools compiled by researchers at the University of California at Santa Cruz for exploring RNA splicing and gene structure in *Caenorhabditis elegans.* It includes a display of cDNA alignments with the genomic sequence, a catalog of alternatively spliced genes and a database of introns. For example, Tracks Display permits the viewing of splicing diagrams for any gene in the Sanger *C. elegans* database alongside cDNA and EST alignments. Worm Align lets users paste in a cDNA sequence and check genomic alignment. There is an Alt-Splicing Catalog for which the cDNA and EST evidence indicates alternative splicing. And with Intron Database, the ends of introns and neighboring exons can be scanned for patterns.
(www.cse.ucsc.edu/~kent/intronerator/)
(http://www.cse.ucsc.edu/~kent/intronerator/)

Proteins are Often Spiced by a Slice of Alternative Splice
