

Alternative splicing in human transcriptome: Functional and structural influence on proteins

Kei Yura ^{a,1}, Masafumi Shionyu ^{b,1,2}, Kei Hagino ^c, Atsushi Hijikata ^c, Yoshinori Hirashima ^c, Taku Nakahara ^c, Tatsuya Eguchi ^c, Kazuki Shinoda ^c, Akihiro Yamaguchi ^d, Ken-ichi Takahashi ^d, Takeshi Itoh ^{e,f}, Tadashi Imanishi ^f, Takashi Gojobori ^{f,g}, Mitiko Go ^{d,h,*}

^a *Quantum Bioinformatics Team, Center for Computational Science and Engineering, Japan Atomic Energy Agency, 8-1 Umemidai, Kizu-cho, Souraku-gun, Kyoto 619-0215, Japan*

^b *Department of Science of Biological Supramolecular Systems, Graduate School of Integrated Science, Yokohama City University, 1-7-29, Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan*

^c *Division of Biological Science, Graduate School of Science, Nagoya University, Furo-cho, Chikusa-ku, Nagoya 464-8602, Japan*

^d *Department of Bio-Science, Faculty of Bio-Science, Nagahama Institute of Bio-Science and Technology, 1266, Tamura-cho, Nagahama, Shiga 526-0829, Japan*

^e *Bioinformatics Laboratory, Genome Research Department, National Institute of Agrobiological Sciences, Kannondai 2-1-2, Tsukuba-shi, Ibaraki 305-8602, Japan*

^f *Biological Information Research Center, National Institute of Advanced Industrial Science and Technology, TIME24 Bldg. 10F, Aomi 2-45, Koto-ku, Tokyo 135-0064, Japan*

^g *Center for Information Biology and DDBJ, National Institute of Genetics, Mishima 411-8540, Japan*

^h *Ochanomizu University, 2-1-1 Otsuka, Bunkyo-ku, Tokyo 112-8610, Japan*

Received 20 February 2006; accepted 3 May 2006

Available online 2 June 2006

Abstract

Alternative splicing is a molecular mechanism that produces multiple proteins from a single gene, and is thought to produce variety in proteins translated from a limited number of genes. Here we analyzed how alternative splicing produced variety in protein structure and function, by using human full-length cDNAs on the assumption that all of the alternatively spliced mRNAs were translated to proteins. We found that the length of alternatively spliced amino acid sequences, in most cases, fell into a size shorter than that of average protein domain. We evaluated comprehensively the presumptive three-dimensional structures of the alternatively spliced products to assess the impact of alternative splicing on gene function. We found that more than half of the products encoded proteins which were involved in signal transduction, transcription and translation, and more than half of alternatively spliced regions comprised interaction sites between proteins and their binding partners, including substrates, DNA/RNA, and other proteins. Intriguingly, 67% of the alternatively spliced isoforms showed significant alterations to regions of the protein structural core, which likely resulted in large conformational change. Based on those findings, we speculate that there are a large number of cases that alternative splicing modulates protein networks through significant alteration in protein conformation.

© 2006 Published by Elsevier B.V.

Keywords: Exon–intron structure; Full-length cDNA; Gene function; Protein stability; Signaling pathway; Three-dimensional structure

Abbreviations: AS, alternative splicing; EST, expressed sequence tag; FLcDNA, full-length cDNA; 3D, three dimensional; PDB, Protein Data Bank; GPCR, G-protein coupled receptor; NMD, nonsense-mediated mRNA decay.

* Corresponding author. Department of Bio-Science, Faculty of Bio-Science, Nagahama Institute of Bio-Science and Technology, 1266 Tamura-cho, Nagahama, Shiga 526-0829, Japan. Tel.: +81 749 64 8153; fax: +81 749 64 8126.

E-mail address: m_goh@nagahama-i-bio.ac.jp (M. Go).

¹ These authors contributed equally to this work.

² Present address: Department of Bio-Science, Faculty of Bio-Science, Nagahama Institute of Bio-Science and Technology, 1266, Tamura-cho, Nagahama, Shiga 526-0829, Japan.

1. Introduction

A large number of introns exist in the genomes of higher eukaryotes. The human genome contains about 53,000 confirmed introns (Lander et al., 2001). Introns can perform various important functions such as exon shuffling and alternative splicing (AS) (Fedorova and Fedorov, 2003). Exon shuffling has been discussed as a mechanism to combine exons and build a gene by a new combination of exons, thereby contributes to the expansion of gene diversity during biological evolution (Gilbert, 1978; Sato et al., 1999). Evidence for exon shuffling was reported on many genes. Exon shuffling apparently occurs by recombination in introns on a genomic DNA (Long et al., 2003). On the other hand, AS changes the exon usage on mature mRNA by altering splicing patterns of introns. AS is a mechanism to increase the number of the structure of transcripts by various combinations of exons in the splicing process.

The number of protein coding genes in the human genome is estimated to total around 20,000–25,000 based on genome sequencing data (International Human Genome Sequencing Consortium, 2004), far fewer than the predicted maximum number of protein coding genes from expressed sequence tag (EST) and CpG island analyses, which yield figures between 80,000 and 120,000 (Antequera and Bird, 1993; Liang et al., 2000). The difference between two kinds of estimations of gene numbers is explained by AS that produces multiple proteins from a single gene (Maniatis and Tasic, 2002). Based on microarray analyses, the proportion of human genes that undergo AS is estimated to be at least 74% (Johnson et al., 2003). Thus, AS shortens the numerical gap between the estimations. However, the question arises whether AS contributes to diversifying functions of mRNAs and/or proteins. Many studies on modification of protein functions by AS have been reported (Stamm et al., 2005), and on the limited cases, the details of the three-dimensional (3D) structural modifications causing functional differences of AS products have been studied (reviewed in Stetefeld and Ruegg, 2005).

Since the number of AS products is expected to be so large, experimental analysis of all AS products is impractical in a short period. Computational analysis is expected to play a valuable role in identification of the possible effects of AS (Black, 2000) on protein structure and function. Assuming that AS products are translated in proteins, the following studies were carried out so far. Kriventseva et al. (2003) have analyzed the effects of AS on protein structure using the AS products registered in Swiss-Prot (Boeckmann et al., 2003). Wen et al. (2004) searched public databases for human alternative splice variants to compile products of “very short alternative splicing (VSAS),” products resulted from AS of segments less than 50 nucleotides long, and proposed that VSAS may alter the 3D structure of proteins. Homma et al. (2004) studied protein 3D structures of AS products in human brain and concluded that most of the alternative splice junctions coincided with domain junctions and that small proportion of cases were found that the junction resided within protein domain. These studies were informative, yet they were carried out using a limited number of AS products.

A valuable resource for studying AS has been produced by H-Invitational (H-Inv (<http://www.h-invitational.jp>)), an interna-

tional collaboration aimed at systematically and functionally validating human genes using a unique set of high quality full-length cDNA (FLcDNA) clones. These FLcDNAs are derived from tissues in which the RNA has undergone the maturation process. Of the 41,118 FLcDNAs mapped to the human genome, 8553 FLcDNAs were turned out to be presumptive AS isoforms. They were mapped onto 3181 loci, with each locus producing 2.7 isoforms on average (Imanishi et al., 2004). H-Inv provides an opportunity to analyze a large number of AS isoforms deduced from human FLcDNAs obtained through transcriptome analysis.

Here, based on exhaustive integrative characterization of cDNA in H-Inv data, we computationally investigated the length distribution of altered region in amino acid sequence by AS, and analyzed the effects of AS on protein 3D structures and functions comprehensively. The number of cases we studied here based on H-Inv data exceeded that of the previous studies of AS on protein 3D structures. Through protein 3D structures, we report how AS modulates gene functions by increasing the number of products from a gene.

2. Materials and methods

2.1. Identification of AS in transcripts and proteins

We selected FLcDNAs that were annotated to be products of AS from H-Inv (Imanishi et al., 2004). We have independently identified exon junctions of these FLcDNAs by mapping them onto the human genome sequence using sim4 program (Florea et al., 1998). Exon structure of each FLcDNA derived from the same locus was compared against that of the representative FLcDNA defined by H-Inv. We defined the differences between the representative FLcDNA and the other FLcDNAs as insertions,

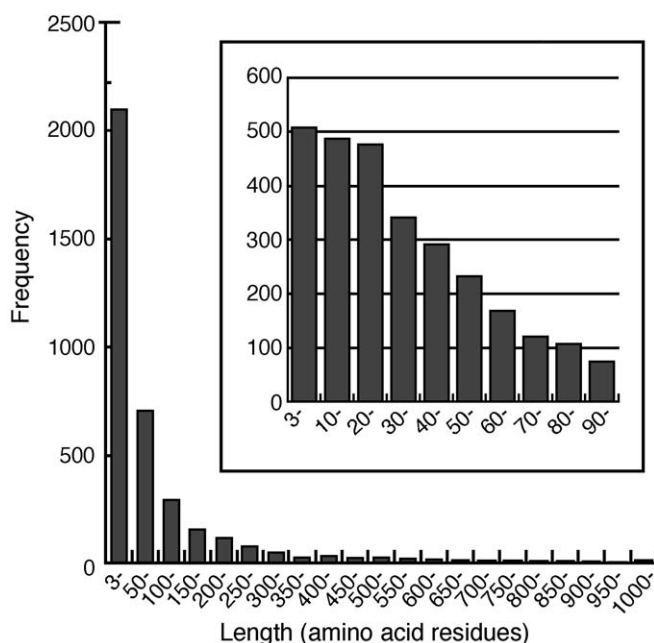


Fig. 1. Distribution of polypeptide length affected by AS in proteins. The inlet is an enlargement of the length distribution for spliced segments shorter than 100 amino acid residues. Cases in which the length was shorter than three residues were not scored, because such small differences could be caused by ambiguity in sequence alignments.

Download English Version:

<https://daneshyari.com/en/article/2820051>

Download Persian Version:

<https://daneshyari.com/article/2820051>

[Daneshyari.com](https://daneshyari.com)