

Alternative splicing as regulator of protein-protein interactions

A. Colantoni, F. Ferrè✉, M. Helmer-Citterich

Department of Biology, University of Tor Vergata, Rome, Italy

Motivations

In recent years, the role of alternative splicing in protein function regulation has been widely investigated. Examples have been reported in which alternative splicing modulates the interaction between two proteins by removing interacting regions. However, despite the large amount of data available about alternative splicing isoforms and protein-protein interactions, only a few systematic studies have been carried to assess how much widespread and general is this form of regulation. Some of these works lack specificity, while others are based on a relatively small number of cases; in addition, they show conflicting results. In this work, we have created a non-redundant residue-level dataset of protein-protein interfaces derived from the three dimensional structure of human protein complexes in order to determine whether the removal of protein-protein interfaces via alternative splicing in the human proteome could have evolved as a form of interaction regulation.

Methods

To create a non-redundant dataset of human protein-protein interfaces, we first obtained from the Protein Data Bank (PDB) all the Biological Units containing at least two human protein chains that could be assigned to Uniprot entries according to SIFTS. From this dataset we extracted all the protein-protein interfaces (defined as sufficiently large sets residues involved in Van der Waals contacts) and filtered out those that were not present in the Asymmetric Unit. Using this procedure, we obtained a set of about 12100 protein-protein interfaces, from which, we derived a non-redundant dataset of nearly 2300 interfaces, using a multi-level clustering procedure based on sequence identity between chains, evaluated using the BLASTClust software (<ftp://ftp.ncbi.nih.gov/blast/documents/blastclust.html>), and on a simple measure of the similarity between two interfaces. For each interface, we mapped the interacting residues to each Ensembl splicing

isoform (if present) of the gene encoding it using residue genomic coordinates retrieved using Biomart, excluding those isoforms which undergo nonsense mediated decay. To evaluate whether splicing has a significant effect on the availability of the interface in the mature protein product, we compared the rate of splicing-mediated interface deletion with that of random controls using a chi-square test. Controls were created picking groups of residues that shared the same dispersion in the sequence with the corresponding interface. An interface, or a control, was considered to be removed from a splicing isoform if the corresponding protein product did not contain at least a certain fraction of its residues: we repeated the procedure using different values of this cutoff. This general procedure was differently adapted for hetero- and homo-dimeric interfaces, taking into account their different nature that might also imply a different evolution of their regulation.

Results

Initial results showed that, if we consider all the isoforms, there isn't any statistically significant difference between the frequency of removal of real interfaces and random controls, both for heterodimeric and homodimeric interactions. However, discarding all the isoforms that determine a large change in the sequence of the interacting protein (which are not likely to regulate specific functions of that protein), we detected a significant tendency to avoid removal of small fractions of the homodimeric interfaces, suggesting that interaction interfaces are protected from being partially spliced and are more often completely included or removed in the mature transcript. This effect was also seen for compact heterodimeric interfaces. This effect is mainly ascribed to interfaces that are small and whose residues are relatively less scattered on the sequence, suggesting that alternative splicing is less likely to have evolved as regulative agent for large interfaces having more dispersed residues.