

Exons are non-randomly associated to transmembrane regions in single- and multi-spanning proteins

Albert Krewinkel^{1,*} and William FitzHugh² and Thomas Martinetz¹ and Steffen Möller^{1,*}

¹Institute of Neuro- and Bioinformatics, University of Lübeck, Ratzeburger Allee 160, 23538 Lübeck, Germany ²Celera Genomics, 45 West Gude Drive, Rockville, MD, USA

*to whom correspondence should be addressed: {krewink,moeller}@inb.uni-luebeck.de

Abstract: For a considerable number of transmembrane or peripheral membrane proteins the existence of a soluble variant has been reported. Solubilisation may be induced post-translationally by proteolytic shedding or pre-translationally by alternative splicing [HHPB05, CSW⁺04]. In the latter case, the soluble variants lack the exons that code for membrane-spanning regions. The solubilisation is an intrinsic process for the regulation of intercellular signalling, particularly in inflammatory processes and the soluble variants are frequently discussed as disease markers. Beside the understanding of disease pathogenicity, investigation of solubilisation processes is valuable since the data yields a good basis for the investigation of alternative splicing. Yet no method has been developed to predict the existence of a soluble variant from sequence or structure.

This work presents a classifier for the prediction of soluble splice variants by linking the exon-intron structure of a gene with the protein's feature to span the membrane. From this data it can be expected that at least 3% of single-spanning proteins have a soluble splice-variant. Conversely, a specific single-spanning protein may be assigned the likelihood to have a soluble splice-variant with respect to its exon structure. This work indicates that the organisation of the nucleotide sequence is indicative for the existence of soluble splice variants. Beyond this specific interpretation, this work may stimulate further research into the statistical association of protein and gene structure and may stimulate the research on soluble splice-variants to model the control of alternative splicing in general.

To develop a statistical measure for the probability to find the exons aligned with membrane-spanning regions (MSRs) in proteins, the evaluation was compared with the same on 50000 randomly positioned MSRs on the same gene structure. The p-value of exon-MSR alignments was found to be remarkable low the in single- and undecaspanning proteins.

References

- [CSW⁺04] M. S. Cline, R. Shigeta, R. L. Wheeler, M. A. Siani-Rose, D. Kulp, and A. E. Loraine. The effects of alternative splicing on transmembrane proteins in the mouse genome. *Pac. Symp. Biocomput.*, 9:17–28, 2004.
- [HHPB05] M. Hiller, K. Huse, M. Platzer, and R. Backofen. Creation and disruption of protein features by alternative splicing – a novel mechanism to modulate function. *Genome Biol.*, 6(7):R58, 2005.