A Protocol for Mapping Alternate Splice Variants onto Protein Structures

Chitra L. Madhwacharyula[1] and Shoba Ranganathan [2,3]

Keywords: Alternative Splice Variants, Mapping, Protein Structure

1 Introduction

Alternate splice variants are generated in eukaryotic genomes by the combination of different genomic elements at the mRNA level. Alternative splicing generates many products from one gene and high frequency of alternative splicing has been reported in the human genome. It is also well established that such splice variants either cause or are indicative of many diseases. While it is recognized that protein structure determines its function, there is no literature evidence at all that splice variants have been mapped to structure, to determine the nature of the modifications caused and to correlate these modifications to diseased states.

This is a new project aimed at systematic organization of known alternate splice variants of protein with available structural information. Structural data may be either for that protein itself or for a close relative, permitting homology-based computational 3D structural modeling. Data collection, curation and the development of a queriable alternate splice variant database, with available experimental structural information is envisaged. This step will be followed by the development of 3D structural models for all alternate splice variants, using bio computing tools. Correlation between sequence and structural variations with diseased states will permit the development of a predictive algorithm for those proteins for which no pathological data is available.

2 Proposed Algorithm

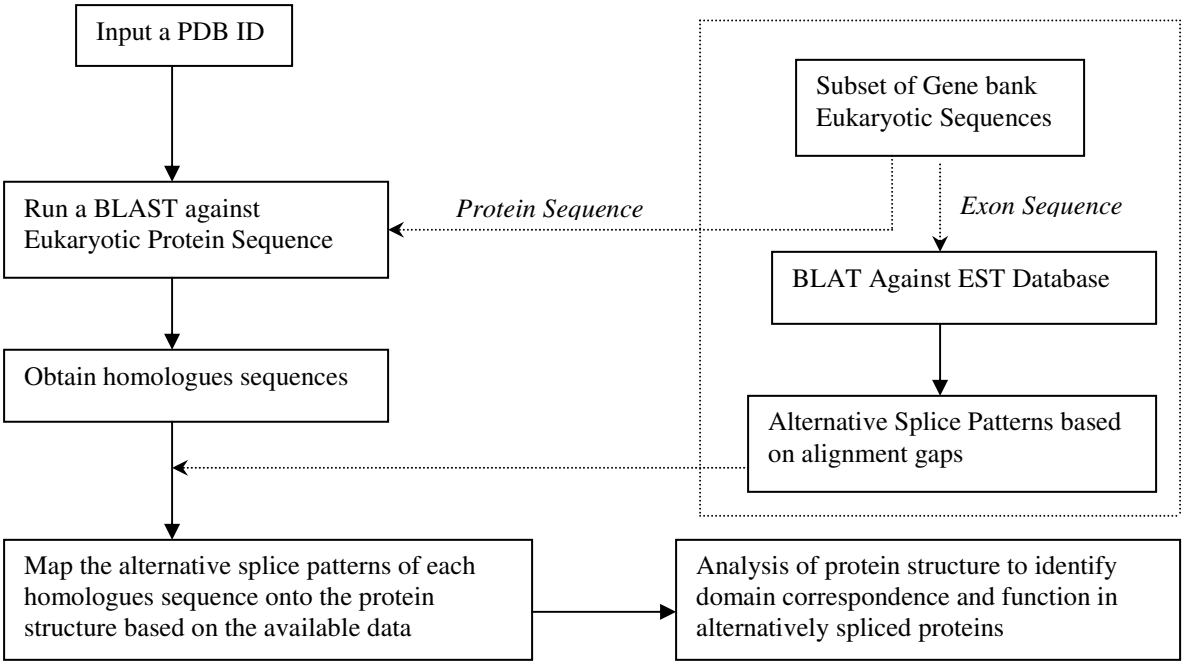

Figure 1: Steps for mapping alternate splice variants onto protein structures

[1] Department of Biological Sciences, National University of Singapore, Singapore, E-mail: chitra@bic.nus.edu.sg
[2] Department of Biochemistry, National University of Singapore, Singapore, E-mail: shoba@bic.nus.edu.sg
[3] Research Institute for Biotechnology, Macquarie University, Sydney, Australia E-mail: shoba@els.mq.edu.au

3 Scope of Work

The main objectives of the project are as follows:

1. Organize all available human proteins known to exhibit alternately spliced variants into a relational database.
2. Identify which of these proteins have structural information.
3. Map the splice variants onto the 3D structure using homology modeling methods.
4. Develop a predictive methodology that can help provide clues as to the functional implications of splice variants.

It is well known today that the human genome has relatively few genes (Lander et al., 2001) but is able generate more than one protein from most genes by generating splice variants (Modrek et al. 2001). It is also well documented that such alternate splice variants have altered functionality and are implicated in many diseases (Stoilov et al., 2002). With the growing interest in genome analysis using bioinformatics and bio computing tools, a large number of alternate splice databases have been developed such as ASAP (the Alternative Splicing Annotation Project: http://www.bioinformatics.ucla.edu/ASAP; Lee et al., 2003). These have sought to collect and organize the available data on alternatively spliced gene products. ASAP provides precise gene exon-intron structure, alternative splicing, tissue specificity of alternative splice forms and protein isoform sequences resulting from alternative splicing Although a few studies have mapped alternate splice variants onto protein structure (Kriventseva et al,2003), a genome-wide scale approach to this problem has not been attempted, to the best of our knowledge.

Thus the proposed project is both timely and novel and addresses the following issues:

- *Data mining* of the human genome
- *Integration of biological databases*: Sequence, Structure, Domain, Disease
- *Bio computing*: Generation of structural models for proteins and/or its splice variants using state-of-the-art structure modeling tools
- *Methodology development*: Utilizing the analysis results to develop a predictive tool so that given a protein and its variants, the functional modification can be computed *a priori*.

4 References

[1] Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J., *et al.* (2001). Initial sequencing and analysis of the human genome. *Nature* **409**, 860–921.

[2] Lee,C., Atanelov,L.,Modrek,B., and Yi,X.(2003).ASAP: the Alternative Splicing Annotation Project. *Nucleic Acids Research*, **31(1)**, 101-105.

[3] Kriventseva,E.V.,Koch,I.,Apweiler,R., et al.(2003). Increase of functional diversity by alternative splicing. *TRENDS in Genetics*, **19(3)**, 124-127.

[4] Modrek, B., Resch, A., Grasso, C., and Lee, C. (2001). Genome-wide detection of alternative splicing in expressed sequences of human genes. *Nucleic Acids Res.* **29**, 2850–2859.

[5] Stoilov, P., Meshorer, E., Gencheva,M., Glick, D., Soreq, H. and Stamm, S. (2002) Defects in Pre-mRNA Processing as Causes of and Predisposition to Diseases. *DNA Cell Biol.* **21**, 803-808.