Chapter 9
Computational Analysis and Characterization of Marfan Syndrome Associated Human Proteins

K. Sivakumar
Sri Chandrasekharendra Saraswathi Viswa Maha Vidyalaya University, India

ABSTRACT

Novel computational procedures and methods have been used to analyze, characterize and to provide more detailed definition of some Marfan syndrome associated human Fibrillin 1 proteins retrieved from NCBI Entrez protein database. Primary structure analysis reveals that the Marfan syndrome associated proteins are rich in cysteine and glycine residues. Extinction Coefficients of Marfan syndrome associated proteins at 280nm is ranging from 1490 to 259165 M^{-1} cm^{-1}. Expasy's ProtParam classifies most of the Marfan syndrome associated human Fibrillin 1 proteins as unstable on the basis of Instability index (II>40) and few proteins (AAB25244.1, 1EMO_A, Q504W9) as stable (II<40) proteins in the room temperature. The aliphatic index infers that the Fibrillin 1 proteins may become unstable at high temperature. GRAVY index of all the proteins indicates that all these proteins may interact equally and easily with water. The number of basic and acidic amino acids in each Marfan syndrome associated human Fibrillin 1 proteins correlates well with the corresponding pI computed. Secondary structure analysis shows that human Fibrillin 1 proteins are found to be with mixed secondary structural content. The average molecular weight of Marfan syndrome associated proteins calculated is 134086 Da. Scanprosite server identified EGF-like domain, TGF-beta binding domain and extracellular sushi domain profiles in Marfan syndrome associated proteins.

INTRODUCTION

Marfan syndrome (OMIM, 2000) [OMIM Number 154700 - http://www.ncbi.nlm.nih.gov/sites/entrez?db=omim] is a systemic, rare connective tissue disorder that affects primarily the bones, joints, heart, blood vessels and the eyes. The main characteristic features of the Marfan syndrome are tall and thin stature with long arms, slender legs, elongated fingers and toes disproportionately long

DOI: 10.4018/978-1-60566-768-3.ch009

in relation to rest of the body and exceeds the body height. Marfan syndrome often leads to the eye problems where one or both lenses of the eye may dislocate, heart and blood vessel problems that may result in heart murmurs, irregular heartbeat, or in severe cases aortic aneurysm. Marfan syndrome affects numerous other structures and organs including the lungs, eyes, dural sac surrounding the spinal cord, and hard palate. Marfan syndrome affects males and females equally, and the mutation shows no geographical bias. According to National Marfan Foundation (http://www.marfan.org) approximately 1 in 5000 people have Marfan syndrome, including men, women and children of all races and ethnic groups. The parents with Marfan syndrome have high chance (50:50) of passing Marfan syndrome on to a child due to its autosomal dominant nature. Signs and symptoms of Marfan syndrome vary from one person to another, even within the same family. Some people have mild signs and symptoms, while others may have severe problems and discomfort, which occur, in many parts of the body. Genetic mutation studies reveal that the Marfan syndrome is due to the mutations in gene located on chromosome 15, locus 15q, 21.1 (Magenis, R.E., 1991) (http://www.ncbi.nlm.nih.gov/projects/mapview/). This gene is involved in the production and processing of Fibrillin 1 (OMIM, 2008) [OMIM Number 134797 - http://www.ncbi.nlm.nih.gov/sites/entrez?db=omim] protein. Mutations in this gene disrupt the production, processing, or assembly of Fibrillin 1 protein. Many researchers have reported clinical and genetics studies on Marfan syndrome. However, sequence analysis and physicochemical characterization of Marfan syndrome associated human Fibrillin 1 proteins have not been done so far. In this chapter we report the computational analysis and characterization studies on Marfan syndrome associated 14 human Fibrillin 1 proteins retrieved from NCBI Entrez protein database to provide more detailed description of Marfan syndrome associated proteins.

MATERIALS AND METHODS

Protein Sequence Databases

Protein sequence databases consist of protein sequences and information about proteins. These databases consist of protein sequences that have been translated from the nucleotide sequences and also sequenced by methods like N-terminal sequencing. Many protein sequence databases are available today and all of these databases allow free download of full content. Any researcher from all over the world can download these protein sequences to study the properties of encoded proteins and utilize them for healthcare, disease identification, drug discovery and development. Protein sequence databases contain the amino acid sequence of proteins; the constituent amino acids are represented by single letter amino acid code. Short descriptions of various protein databases used in this study are given in the following sections.

UniProtKB/Swiss-Prot

SwissProt (http://www.expasy.org/sprot) is a manually curated protein sequence database that provides a high level of annotation (e.g. the description of the function of a protein, its domain structure, etc), while maintaining a minimal level of redundancy and a high level of integration with other databases. Release 56.5 of (Nov-2008) contained 402,482 entries.

UniProtKB/TrEMBL

UniProtKB/TrEMBL (http://www.uniprot.org/database/knowledgebase.shtml) consists of computer-annotated entries, which are derived from the translation of coding sequences present in the EMBL/GenBank/DBJ Nucleotide Sequence Databases and also protein sequences extracted from the literature or submitted to UniProtKB. Release 39.5 (Nov-2008) contained 6796,837