DESIGNING NOVEL LEAD MOLECULES FOR HUMAN COAGULATION FACTOR XIII B INVOLVED IN CARDIOVASCULAR DISEASES

A thesis submitted in partial fulfillment of the

Requirements for the degree of

Bachelor of Technology

In

Biomedical engineering

By

Abhishek Nayak (111BM0011)

Department biotechnology and medical engineering

National Institute of Technology, Rourkela

Rourkela, Odisha – 769008

NATIONAL INSTITUTE OF TECHNOLOGY

ROURKELA 769008, INDIA

Certificate of Approval

This is to certify that the thesis entitled "**DESIGNING NOVEL LEAD MOLECULES FOR HUMAN COAGULATION FACTOR XIII B INVOLVED IN CARDIOVASCULAR DISEASES**." submitted to the National Institute of Technology, Rourkela by **ABHISHEK NAYAK**, **Roll No. 111BM0011** for the award of the Degree of Bachelor of Technology in Biomedical Engineering is a record of bona fide research work carried out by then under my supervision and guidance. The results presented in this thesis has not been, to the best of my knowledge, submitted to any other University or Institute for the award of any degree or diploma. The thesis, in my opinion, has reached the standards fulfilling the requirement for the award of the degree of Bachelor of technology in accordance with regulations of the Institute.

Prof. (Dr.) MUKESH KUMAR GUPTA

Associate Professor

Department of Biotechnology and medical engineering

National Institute of Technology, Rourkela

Date:

Acknowledgement

I have taken efforts in this project. However, it would not have been possible without the kind support and help of many individuals and the department. I would like to extend my sincere thanks to all of them.

I am highly indebted to Department of Biotechnology and medical engineering for providing necessary information and guidance regarding the project and also for their support in completing the project.

I would like to express my gratitude and special thanks to my project guide **Prof. (Dr.) Mukesh Kumar Gupta**, Associate Professor, Department of Biotechnology and medical engineering, NIT Rourkela, for all the cooperation and time. I would also like to thank Mr. Praveen Guttula for giving me all the time I needed for understanding the software required for the project.

Last but not the least I would like to thank my parents and National Institute of Technology Rourkela for giving me this wonderful opportunity.

Abhishek Nayak (111BM0011)

CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABBREVIATIONS

Name	Full form
CVD	Cardiovascular diseases
2D	Two Dimensional
3D	Three Dimensional
MI	Myocardial Infarctions
CVA	Cerebrovascular accident
XIIIF	Coagulation factor XIII
VS	Virtual Screening
CADD	Computer Aided Drug Designing

Abstract

The coagulation of blood is very important characteristics of the blood since it protects us from excess bleeding by forming clots and binding it. This is done by protein called fibrin stabilizing protein (XIIIB) and plays an important role in clot stabilization by crosslinking fibrin chains. However it has been found that these fibrin stabilizing factor also known as blood coagulation factor XIII is responsible for many cardiovascular diseases. But very little is known about the role of FXIII in vascular diseases. Although it showed increased levels of FXIII A-subunit antigen in patients with obliterative atherosclerosis of the lower limbs and in patients with diabetic angiopathy, small patient numbers make the interpretation of these data difficult. However, increased plasma concentration of cross-linked fibrin polymers in acute myocardial infarction has been described, assuming the presence of increased plasma FXIII plasma activity in patients with coronary artery disease. The research work aims to develop a novel lead molecule for coagulation factor XIIIB involved in cardiovascular diseases. The proposed work defines the protein structure, identifies the proposed ligands and develops the molecule by docking.

Keywords: Blood, clot, Stabilizing, Cardiovascular, Coagulation factor, ligands, docking.

Chapter – 01

INTRODUCTION AND OBJECTIVE

1. Introduction and Objective

Cardiovascular systems consists of blood, heart and veins. Cardiovascular sickness (CVDs) are the class of infections that include veins and heart. The term cardiovascular sickness deals with any ailment that are linked with cardiovascular network such as hypertension, atherosclerosis, myocardial dead tissue, aneurysm and stroke. CVD is considered to be lead cause of death of 14 million people throughout the world (Mather 2006, Murray and Lopez 1996, WHO 2002). In 1980-2003 the National Heart, Lung and Blood Institute (NHLBI) and Framingham Heart Study (FHS) assessed around 81,100,000 Indian adults to have one or more sorts of CVD. In India report of CVD death, rate of death is 34.3 percent, which is the very reason of death every year. (Nelson *et al*., 2006).

Hypertension happens when blood goes through the vessels at a higher velocity than normal. Atherosclerosis is a condition of collection of oily materials including cholesterol, underneath the internal linings of blood vessels. Myocardial infarctions (MI), also called as heart attack, prompts degradation of a part of heart muscle due to a non-exposure to oxygen. Cerebrovascular accident (CVA) in like manner called as stroke, routinely occurs as somewhat cranial arteriole blasts or is hindered by an embolus (Thomas *et al*., 2000). Cardiac death is more found in smokers, an undesirable eating routine is a vital peril component for unending cardiovascular diseases (Maton *et al*., 1993). Fast food and refreshments containing colorants and added substances which are dangerous for the heart. Intake of excessive alcohol can induce hypertension, heart blocking, alcoholic cardiomyopathy and strokes. Drugs like heroin, cocaine and similar kinds cause choking in the heart valve and deteriorates the circulatory system. Family history also plays an essential factor in cardiovascular disease risk. The ordinary side effects of cardiovascular illness are related mainly with chest pain with shortness of breath. Individuals with coronary artery infection generally have steady movement of their symptoms over the long run. The seriousness or severity of symptoms increases as a vessel contracts over the long run. The indications of a stroke include: sudden weakening of the face, leg and arm

frequently on one side of the body, sudden complexity, inconvenience while talking or comprehension, eye, sudden trouble in strolling, wooziness, loss of balance or coordination, serious migraine with no known reason.

The treatment of cardiovascular disease is based upon clinical history and physical examination of the patient. Supporting tests incorporate electrocardiography and estimation of the level of serum creatine kinase catalysts which are discharged into the blood by the injured tissue. Next technique is the heart imaging procedure which is called angiography (infusion of color into the vessels followed by X-rays). Lowering the pulse by decreasing the volume of blood pumped by the heart can be done by the beta-blockers such as propranolol (Inderal), metoprolol (Lopressor, Toprol), nadolol (Corgard) and atenol (Tenormin). These medications might likewise lessen the danger of a resulting heart attacks in patients who have endured once.

Coagulation factor XIII B higher levels were seen during cardiovascular illness conditions (Bottenus *et al*., 1990). Over articulation of the protein Coagulation factor XIII B plasma prompts sicknesses like atherosclerosis, myocardial infarction (Lorand *et al*., 2001).

Myocardial infarctions (MI) usually known as a heart attack, results from the interference of blood supply to a piece of the heart, resulting in the heart cells to die (Francis *et al*., 1980) . Due to impaired transglutaminase activity or a disproportionate increase of metalloproteinases may reduce the fibrous cap's strength of an atheroma thereby increasing its vulnerability and favoring plaque rupture, a main cause of myocardial infarction (Devine *et al*., 1993*)*. Dimethylbiguanide (metformin) interferes with factor XIII activation and with fibrin polymerization and atherosclerosis is caused (Shebusk *et al*., 1990). At the point when the illness is treated with metformin more than 3 months, factor XIII antigen and movement levels were diminished and it controls the illness (Ritchie *et al*., 2000).

Computer aided drug design (CADD) utilizes computational science to find, upgrade, or study biologically active particles and molecules. The important part is the anticipation of the binding strength of a given molecule and its target. Consequently, ligand based virtual screening convention of CADD system was induced in the present study to propose potential inhibitors of Coagulation factor XIII B. Along these lines the present study was completed with the following aim and objectives.

1.1 Aim and Objectives of the Work

The aim of the work can be determined as follows:

- Designing novel lead molecules for human Coagulation factor XIII B involved in cardiovascular diseases.

The objectives are as follows:

1. Proteomic analysis of human Coagulation factor XIII B.
2. Delineating Phylogenetic relationship of 30 selected human proteins causing cardiovascular disease.
3. Metabolic pathway analysis.
4. Ligand binding site prediction.
5. Lead identification and optimization.
    - Ligand based virtual screening.
    - Docking and scoring using VLife software suite.

Chapter - 02

LITERATURE REVIEW

2. Literature Review

The capacity of the body to control the stream of blood after vascular damage is very important to proceed with survival. The procedure of blood coagulating and after that the resulting disintegration of the coagulation, taking after repair of the damaged tissue, is termed hemostasis (Chung *et al*., 1974). Hemostasis made out of 4 noteworthy events that happen in a set request after the loss of vascular stability. The first step of the procedure is vascular constriction. This constrains the stream of blood to the area of damage. Next, platelets get enacted by thrombin and gather at the site of damage, forming an interim, free platelet plug (Nagy *et al*., 1980). Stimulating of platelet clumping is mainly done by a protein known as fibrinogen. Clumping of platelets occurs by tying to collagen that is exposed followed by rupture of endothelial covering of vessels (Muszbek *et al*., 1987). Nucleotide, ADP and the eicosanoid, TXA2, serotonin, phospholipids, lipoproteins, and different types of proteins imperative for coagulation cascade only after activation discharged by the platelets. Platelets, beside induced secretion can also change their shape in order to facilitate the formation of plug (Ando *et al*., 1987). Stability of the initial platelet plug is ensured by formation of a fibrin mesh (clot formation) and it surrounds the plug. At the end, the coagulation or the clot must be dissolved for blood stream to resume after tissue repair. The disintegration of the coagulation happens through the activity of plasmin (Hornyak *et al*., 1989).

Intrinsic and extrinsic pathways lead to the development of fibrin coagulation (Muller *et al*., 1984). The two pathways merge on a typical pathway that prompts coagulation development even if their initiation mechanisms are different (Sakata *et al*., 1980). Both are complex pathways and include various distinctive proteins termed as clotting factor. Under typical physiological conditions fibrin clot formation is the most relevant event that occurs during hemostasis (Mosher *et al*., 1980). Most clinically significant is the initiation of the natural pathway by contact of the artery wall with lipoprotein particles, VLDLs and chylomicrons (Mimuro *et al*., 1986).

A much more stimulating goal is to develop an inhibitor for the chain in treatment of cardiovascular diseases, such as atherosclerosis, myocardial infarction (Morgan *et al.,* 2007). A judicious and efficient way of inhibitor identification is high throughput virtual screening. Virtual Screening (VS) methods are fundamental part of the drug discovery system (Bajorath *et al.,* 2002). Discovery of novel inhibitors (Umamaheswari *et al.,* 2011) based on ligand based virtual screening method has been a great discovery. This methods precisely detects the small inhibitor property enabled molecules towards a target specific drug.

Blood coagulation factor XIII holds an important part in blood clotting and stabilization by crosslinking the fibrin chains (Ichinose *et al.*, 1994). Blood coagulation XIII is basically a transglutaminase consisting two A and two B subunits rotating as a tetramer (A_2B_2) in the plasma (Yee *et al.*, 1994). At around 21 μg/ml concentration, all A-subunit molecules are in complex with the B-subunit in the plasma while the B-subunit is in both free and complex forms. The overall concentration of B is about 21 μg/ml and nearly 50% of it freely circulates in plasma (10 μg/ml). Almost 90% of FXIII is attached to fibrinogen by a linking site for the B-subunit (Jujikawa *et al.*, 1986) during circulation. Fibrinogen acts as a transporter for the zymogen plasma through this process. Activation peptide consisting of 37 amino acid is released after activation of FXIII to FXIIIA′. Dissociation of A'_2B_2 tetramer is facilitated in the presence of calcium. Both the phases (dissociation of the subunits and release of activation peptide) are important for the activation of the FXIII zymogen (Board *et al.*, 1980) completely. In presence of fibrinogen the tetramer dissociation is increased and the activation peptide is released from the A-subunit which is enhanced by the proteolytic products of fibrinogen (polymerized des-A, des-A, B fibrinogen) (Dykes *et al.*, 1987). When FXIII-catalyzed cross-linked fibrin appears, this effect on the activation peptide is diminished. Activated factor XIII increases the mechanical toughness of the fibrin by catalyzing the covalent cross linking of the γ and α chains, and crosslinking the antiplasmin which increases its resistance to fibrinolysis. Some examples of proteins which form the substrates for factor XIII are factor FVII, fibronectin, α_2-antiplasmin, thrombospondin, von Willebrand factor, collagen and vitronectin (Wilson *et al.*, 1998).

Factor XIII being a protransglutaminase, is activated by thrombin which catalyzes the formation of e-(cglutamyl) - lysine bonds in fibrin polymerization. Fibrin crosslinking reduces the clot mechanically, chemically as well as resilient to fibrinolysis [1, 2] (Kamath *et al.*, 2003). FXIII is a heterotetramer consisting of two A and two B subunits that are associated non-covalently as A_2B_2 during the time of circulation [1, 2]. The A subunit has active site of

the enzyme (Kaptoge *et al.*, 2003). Activation of FXIII is done by cleaving of the peptide bond among Arg37 and Gly38 of the A-subunit followed by B-subunit dissociation from the A-subunit in presence of Ca2+ [3, 4]. Both the phases are essential in order to access the substrate to the catalytic triad of FXIII which are enhanced by the presence of the main substrate polymerized fibrin (Spencer *et al.*, 2000). The B subunit is involved in equilibrium of the A-subunit and has no enzymatic activity in the human plasma aqueous atmosphere [5, 6]. It consists of 641 amino acids divided into 10 cyclic repeats of about 60 amino acids each, which are called GPI structures or Sushi domains [7, 8]. Each of the Sushi domain is preset by a single exon. The function of these domains are thought to serve mainly as protein-binding module but is not yet known properly (H.P. Kohler, 1 February 1999) (Sinning *et al.*, 2006). The A and B subunits of factor XIII interact as a complex in plasma, the genes that encode them reside on chromosomes 6p and 1q respectively, and are genetically unlinked. A Val34Leu polymorphism of the factor XIIIA gene located near the thrombin activation site has been associated with an altered rate of factor XIII activation, abnormal fibrin clot structure, and a decreased risk of arterial and venous thrombotic events. Komanasin recently identified a His95Arg polymorphism of the factor XIIIB gene that is related with an increased dissociation rate of the factor XIII A_2B_2 tetramer followed by activation by thrombin, but the role of His95Arg in arterial thrombosis has yet to be known.

The dynamic site of the catalyst is contained in A-subunit and is incorporated by hepatocytes, monocytes, and megakaryocytes (Tzoulaki *et al*., 2007). The A-subunit flowing in plasma is taken from both liver and bone marrow is verified by protein phenotype analysis after bone marrow and liver transplantation. The B-subunit acts as a transporter for the reactant A-subunit in plasma, is arranged by the liver, and is emitted as a monomer that ties free A in plasma (Cremer *et al*., 2007). The A-subunit is isolated into 4 space (Robert A. S. Ariëns, August 1, 2002), assigned the catalyst center, barrel 1 and barrel 2. It contains an initiation peptide of 37 amino acids that forces the entrance of the substrate to the dynamic site cysteine. This structure is balanced out by a few hydrogen bonds and bridges of salt between the initiation peptide and the catalyst center of one subunit and the reactant center and barrel of the second subunit (Losner *et al*., 2003). While the A-subunit contains 6 potential asparagine-connected glycosylation destinations, none of these have carb connections, as decided by recoloring with occasional corrosive Schiff base reagent14 or by mass spectrometry3 interestingly, sugar adds to pretty nearly 8.5% of the aggregate sub-atomic weight of factor XIII B-subunit. The B-subunit is a protein made out of 10 repeated Sushi or glycoprotein-1 domains. Each Sushi area contains 2 disulfide spans that manage its tertiary structure, adding

up to a sum of 40 cysteine deposits and 20 disulfide, connecting the B-subunit protein (Phear *et al*., 2001). The fundamental function of B-subunit is the adjustment and transport of the hydrophobic A-subunit in the human plasma environment (Levy *et al*., 2001). A substantially more difficult objective is to create inhibitor for the Coagulation element XIII B chain in treatment of cardiovascular sickness, for example, atherosclerosis, myocardial dead tissue (Morgan *et al*., 2007). An effective method for inhibitor distinguishing proof is high throughput virtual screening. Virtual screening (VS) methods play an essential role in the fields of drug research and discovery (Bajorath *et al*., 2002). One of the advancements in this field is exemplified with showed works performed in the field of ligand based virtual screening in disclosure of novel inhibitors against different illness (Umamaheswari *et al*., 2011). This procedure is time saving and in the meantime profoundly solid to recommend little atoms with sufficient resistance towards a predetermined medication target.

The literature study revealed that there was no experimentally determined structure for human Coagulation factor XIII B. Therefore a reliable three dimensional model for the human Coagulation factor XIII B can be modelled using bioinformatics tools. As there is no specific inhibitor for human Coagulation factor XIII B and published inhibitor metformin obtained through literature search were used for ligand based high throughput virtual screening.

Chapter - 03

METHODOLOGY

3. Methodology

In this chapter, the software used and the process has been discussed.

3.1 Retrieval of human coagulation factor XIII B sequence

UniProt (Universal protein asset) is a store of protein information made by consolidating Swiss-Prot, TrEMBL and PIR. The UniProt involves the European Bioinformatics Institute (EBI), the Swiss Institute of Bioinformatics (SIB) and the protein information resource (PIR). It is the most essential database among all protein grouping databases (Leinonen *et al*., 2009). The UniProt Knowledgebase (UniProtKB) is the focal access area for broad curated protein data, comprising of capacity, arrangement and cross-references. Genome sequencing activities are the primary hotspot for UniProt. These proteins arrangement data accessible at UniProt hold the keys to get learning on some key elements on the medication targets which are basic for computer aided drug designing. UniProtKB Protein information base comprises of two segments SwissProt and TrEMBL.

The protein sequence of human coagulation factor XIII B was retrieved from the UniProt accessed at *http://www.uniprot.org*

3.2 Proteomic analysis of human coagulation factor XIII B

Proteins are basically the polymers of L-alpha amino acids. Proteins fold into one or more particular spatial adaptations, determined by various non-covalent interactions, for example ionic interactions, van der Waals forces, hydrogen bonding and hydrophobic packing to perform their biological function (Durbin *et al*., 2006). Proteomic examination gives a premise to *in silico* portrayal of the protein towards physio-chemical properties, functional sites, secondary structural elements and structure function relationship.

The ExPASy or (expert protein analysis system) is a proteomics server of the Swiss Institute of Bioinformatics (SIB). It investigates protein arrangements and structures (Gasteiger *et al*., 2003). The server capacities in a joint effort with the European Bioinformatics Institute. ExPASy additionally delivers the protein grouping knowledgebase, UniProtKB/Swiss-Prot and its computer aided supplement, UniProtKB/TrEMBL.

Primary sequence analysis

The primary structure of a protein consists of a linear chain of amino acids and peptide bonds connecting each residue. Number of online tools for performing primary structure analysis are available in ExPASy like ProtParam, ProtScale, SAPS (Statistical Analysis of protein Sequence) etc.

ProtParam (Gasteiger *et al*., 2005) is a tool which allows computation of various physio chemical parameters for a given protein stored in Swiss-Prot or TrEMBL or for a user entered sequence. The computed parameters include the molecular weight, theoretical pI, amino acid composition, atomic composition, extinction coefficient, estimated half-life, instability index, aliphatic index and grand average of hydropathicity (GRAVY).

The protein sequence of human coagulation factor XIII B was submitted to ProtParam *at http://www.expasy.ch/tools/ProtParam.html* and results were analyzed.

Prediction of functional elements

Protein consists of functional parts such as motifs, patterns, domains and BLOCKS. Patterns are the conserved sequence of amino acids having characteristic trait of family. A motif is a nucleotide or amino acid sequence pattern that is vast and has, a biological significance where domain is referred as part of protein sequence and structure that can evolve the specific function. BLOCKS (Henikoff and Henikoff, 1996) are unspaced multiple alignment of segments of related protein sequences that relates to the most conserved regions of proteins. The database, for example, PROSITE, BLOCKS and Pfam were utilized to anticipate useful components from essential arrangement of the medication target.

Prediction of Pattern

Uncharacterized proteins interpreted from genomic or cDNA arrangements are done by a database known as PROSITE (Bairach *et al*., 2002). With appropriate computational methods any biologically important patterns and sites can be identified along with its family. At times, the arrangement of an unknown protein is too remotely identified with any protein of known structure for recognizing its similarity by general grouping arrangement, yet it can be recognized by the event in its succession of a specific bunch of deposit sorts which is differently known as a signature, motif, pattern or fingerprint. These motifs arise as a result of specific necessities on the structure of particular regions of a protein which may be essential for their enzymatic movement. These necessities force tight imperatives on the development of those constrained (in size) however essential portions of a protein arrangement. The patterns are recognized by utilizing PROSITE database.

The protein sequence of human coagulation factor XIII B was submitted to PROSITE database (*http://www.expasy.ch/prosite/*) and the results were analyzed.

Prediction of BLOCK

The BLOCKS database is a collection of blocks representing known protein families that can be used to compare a protein of DNA sequence with documented families of protein. The blocks for the BLOCKS database are made usually by looking for most highly conserved regions in group of proteins documented in InterPro (Attwood *et al*., 2000).

The protein sequence of human coagulation factor XIII B was submitted to BLOCKS database (*http://blocks.fhcrc.org/blocks/blocks_search.html*) and the results were analyzed.

Prediction of Domain

The Pfam (Eddy et al., 1997) database is a huge accumulation of protein families, each represented by multiple sequence alignments (MSAs) and hidden markov models (HMMs). Generally, proteins are made out of one or more functional groups called domains. Diverse mixes of domains offer ascent to the various scope of proteins found in nature. The capacity of a protein can be known by recognizable proof of their areas that happen inside of the protein.

There are two parts of Pfam: Pfam-A and Pfam-B. Pfam-A entries are high caliber, physically curated families. In spite of the fact that these Pfam-A entries cover a substantial extent of the arrangements in the basic grouping database, to give a more thorough scope of known proteins to create a supplement naturally utilizing the PRODOM database. These naturally produced entries are called Pfam-B. In spite of the fact that of lower quality, Pfam-B families can be helpful for distinguishing practically moderated districts when no Pfam-A sections are found. Pfam likewise creates more elevated amount groupings of related families, known as groups. A clan is an accumulation of Pfam-A sections which are connected by likeness of grouping, structure or profile-HMM.

The protein sequence of human coagulation factor XIII B was submitted to Pfam database (*http://pfam.sanger.ac.uk/search*) and results were analyzed.

3.3 Phylogenetic Analysis

Phylogenetic analysis is the study of the evolutionary relationships among a group of related (nucleic acid or protein) sequences by producing a tree representation of the relationships known as phylogenetic tree. To construct a phylogenetic tree a multiple sequence alignment file of a group of 30 human cardiovascular disease causing sequences is needed. The various programs used for phylogenetic tree construction are PHYLIP, MEGA, PAML, TREE-PUZZLE, DAMBE, Hennig86, RNA, NONA etc.

Multiple sequence alignment

Arranging three or more biological sequences, usually protein, RNA or DNA can be done by multiple sequence alignment (MSA). Ordinarily, the input set of query sequences are assumed to have shared a direct lineage. After doing the MSA, sequence homology can be done and phylogenetic analysis can be conducted to use the sequences shared on the basis of evolutionary origins. In order to assess sequence conservation of protein domains, tertiary and secondary structures and even individual amino acids or nucleotides multiple sequence alignment (MSA) is often used.

Phylogenetic tree construction

A phylogenetic tree a fanning outline or "tree" demonstrating the induced transformative connections among different natural species or different substances based upon equalities and contrasts in their physical and/or hereditary attributes. In an established phylogenetic tree, every hub with relatives speaks to the surmised latest regular predecessor of the relatives and such detail is not present in an unrooted tree. The methods used for phylogenetic tree construction are:

1. Maximum parsimony method -for a set of closely relates sequences.

2. Maximum likelihood method -for a set of distantly related sequences.

3. Distance matrix method -if the distance of relatedness relation is unknown or the sequences are moderately related.

3.4 Pathway Analysis

KEGG (http://www.genome.jp/kegg/) is a database of natural frameworks that incorporates genomic, synthetic and systemic useful data. KEGG gives a reference learning base to connecting genomes to life through the procedure of pathway mapping, which is to guide, for instance, a genomic or transcriptomic substance of qualities to KEGG reference pathways to derive systemic practices of the cell or the living being. Furthermore, KEGG gives a reference learning base to connecting genomes to nature, for example, for the examination of medication target connections, through the procedure of brite mapping. KEGG brite is a cosmology database speaking to useful progressive systems of different natural articles, including atoms, cells life forms, illnesses and medications, and connections among them. KEGG pathway is currently supplemented with another worldwide guide of metabolic pathways, which is basically a joined guide of around 120 current pathway maps. Also, littler pathway modules are characterized and put away in KEGG module that likewise contains other useful units and edifices. The KEGG asset is being extended to suit the requirements for reasonable applications.

KEGG database was used to analyze role of human coagulation factor XIII B in metabolic pathway.

3.5 Homology modelling

Homology modelling otherwise called comparative modelling is a class of technique for developing a model of a protein in atomic resolution from its amino acid succession (the "query sequence" or "target"). All homology modelling strategies depend on the distinguishing proof of one or more known protein structures (known as "templates" or "parent structures") liable to look like the structure of the query arrangement. The basic assumption of homology modelling is the unknown structure and the homologous template protein of known structure have nearly identical backbone structure in the aligned regions. Homology modelling may also be used in CADD when a good template structure is available for the target sequence. Homology modelling is used only when the target sequence have >30% identity with the template sequence. On production of an alignment that maps residues in the query sequence to residues in the template sequence a model can be generated using different software making the alignment as input file.

3.6 3-D structure prediction of coagulation factor XIII B protein

The knowledge of the three-dimensional structure is a requirement for experimental approach. Structural information often greatly enhances our understanding of how proteins function and how they interact with each other. X-ray crystallography, NMR spectroscopy, and Electron Microscopy are the ways to obtain detailed structural information. But, all these techniques involve elaborate technical procedures and many proteins fail to crystallize at all and/or cannot be obtained or dissolved in large enough quantities for NMR measurements. The size of the protein is also a limiting factor for NMR. Therefore when structural information regarding a protein is not available from these experimental procedures rely on computational modeling to identify the tertiary structure of the protein. The computational modeling for structure prediction includes Homology modeling, threading and ab initio methods. Homology modeling is widely used procedure of aligning target and template to generate a model.

3.7 Input data for Homology Modelling

- Target sequence.
- Template sequence.
- Alignment of target and template sequence.
- Structure alignment of multiple templates, if available.
- Experimental information.

Modeller is open source software used for protein modelling. SGI Fuel Workstation with 2.6 GHz processor, 8GB RAM, 750 GB Hard drive and an Nvidia GeForceGT 540M graphics card running in Windows 7 operating system was set as experimental environment for the present study. Bioinformatics software such as Modeller9v8 (Eswar *et al*., 2008; Sali and Blundell, 1993), and online resources were employed systematically to propose the outcome of the study.

3.8 Modelling of the unknown protein

Modeller 9.14

Modeller is a computer program that models three-dimensional structures of proteins and their assemblies by satisfaction of spatial restraints (Sali and Blundell, 1993). The distance and dihedral angle restraints on the target sequence are calculated from its alignment with template 3D structure. The form of these restraints was obtained from a statistical analysis of relationship between many pairs of homologous structures. This analysis relied on a database of 105 family alignments that include 416 proteins with known 3D structure (John and Sali, 2003). A 3D model is obtained by optimization of a molecular probability density function (pdf). The molecular pdf for comparative modeling is optimized with the variable target function procedure in Cartesian space that employs methods of conjugate gradients and molecular dynamics with simulated annealing.

Modeller can also perform multiple comparisons of protein sequences and/or structures, clustering of proteins, and searching of sequence databases. Modeller is most frequently used for homology or comparative protein structure modeling. The user provides an alignment of a sequence to be modeled with known related structures and modeller will automatically

calculate a model with all non-hydrogen atoms. Modeller9v8 is the latest version and is currently maintained by BENWEBB, California University, and Sanfransisco, U.S.A.

Template identification: A template is the one which has >30% identity with the target or query sequence to build a reliable model. More the identity more will be the accuracy of the model. Template can be identified using BLASTP search.

BLAST: Basic local alignment search tool (BLAST) is a sequence similarity search tool. BLAST (Altschul *et al*., 1990) is available from http://www.ncbi.nlm.nih.gov/blast.

BLAST at NCBI has been recently re-engineered to improve usability and performance (Johnson *et al*., 2008), it compares all combinations of protein or nucleotide queries with nucleotide or protein database. There are different flavors of BLAST like Blastp, Blastn, Blastx, tBlastn, and tBlastx. Of which Blastp is uses query protein sequence and searches for similarity in the protein database. Blastp was used to find the template sequence.

The human coagulation protein was queried against protein databank (PDB) using BLASTP to identify the template structure.

Model generation

Modeller 9.14v was used for generating a 3 dimensional model coagulation factor XIII B chain protein. Modeller builds the model based on homology modeling, it requires three files those are Alignment file, PDB or Atomic file and Script file. Solution structure of the c-terminal SCR-16/20 fragment of complement factor H of human was taken as template sequence (having PDB ID 2QFH).

Alignment file contain sequence alignment of target and template, this file should have an extension '*.ali'. The resulted PIR file generated by ClustalX was modified as below and saved as 'qsec1.ali'.

```
P1;P05160
sequence:P05160:1:A:661:A::::
MRLKNLTFIIILIISGELYAEEKPCGFPHVENGRIAQYYYTFKSFYFPMS
IDKKLSFFCLAGYTTESGRQEEQTTCTTEGWSPEPRCFKK--CTKPDLSN
GY-ISDVKLLYKIQENMRYGCASGYKTTGGKDEEVVQCLSDGWSSQPTCR
KEHETCLAPELYNGNYSTTQKTFKVK--DKVQYECATGYYTAGGKKTEEV
```

```
ECLTYGWSLTPKC--TKLKCSSLRLIENGYFHPVK-QTYEEGDVVQFFCH
ENYYLSGSDLIQCYNFGWYPESPVCEGRRNRCPPPPLPINSKIQTHS-TT
YRHGEIVHIECELNFEIHGSAEIRCEDGKWTEPPKCIEGQEKVACEEPPF
IENGAANLH----SKIY-YNGDKVTYACKSGYLLHGSN---EITCNRGKW
TLPPECVENNENCKHPPVVMNGAVADGILASYATGSSVEYRCNEYYLLRG
SKISRCEQGKWSSPPVCLEPCTVNVDYMNRNNIEMKWKYEGKVLH--GDL
IDFVCKQGYDLSPLTPLSELSVQCNRGEVKYPLCTRKESKGMCTSPPLIK
HGVIISSTVDTYENGSSVEYRCFDHHFLEGSREAYCLDGMWTTPPLCLEP
CTLSFTEMEKNNLLLKWDFDNRPHILHGEYIEFICRGDTYPAELYITGSI
LRMQCDRGQLKYPRCIPRQSTLSYQEPLRT
*
>P1;2QFHa
StructureN:2QFHa:935:A:1226:A::::
---------------------------PEISHGVVAHMSDSYQ-------
YGEEVTYKCFEGFGIDG---PAIAKCLGEKWSHPPSCIKTDCLSLPSFEN
AIPMGEKKDVYKAGEQVTYTCATYYKMDGASN---VTCINSRWTGRPTCR
--DTSCVNPPTVQNAYIVSRQMSKYPSGERVRYQCRSPYEMFGD---EEV
MCLNGNWTEPPQCKDSTGKCGPPPPIDNGDITSFPLSVYAPASSVEYQCQ
NLYQLEGNKRITCRNGQWS-EPPKCL--HPCVISRE--IMENYNIALRWT
AKQKLYSRTGESVEFVCKRGYRLSSRSHTLRTTCWDGKLEYP--------
--------------------------------------------------
--------------------------------------------------
--------------------------------------------------
-----------------------------------------
*
>P1;2QFHb
StructureN:2QFHb:929:A:1231:A::::
--------------------------------------------------
--------------------------------------------------
--------------------------------------------------
--------------------------------------------------
----------LPCKSPPEISHGVVAHMSDSYQYGEEVTYKCFEGFGIDGP
AIAKCLGEKWSHPPSCI----KTDCLSLPSFENAIPMGEKKDVYKAGEQV
TYTCATYYKMDGASNVTCINSRWTGRPTCRD----TSCVNPPTVQNAYIV
SRQMSKYPSGERVRYQCRSPYEMFGDEEVMCLNGNWTEPPQCKDSTGKCG
PPPPIDNGDITSFPLSVYAPASSVEYQCQNLYQLEGNKRITCRNGQWSEP
PKCLHPCVISREIMENYNIALRWTAKQKLYSRTGESVEFVCKRGYRLSSR
S--HTLRTTCWDGKLEYPTCAKR
*
```

```
>P1;2QFHc
StructureN:2QFHc:924:A:1231:A::::
-------------------------------------------------
-------------------------------------------------
-------------------------------------------------
-------------------------------------------------
-------------------------------------------------
-------------------------------------------------
----------------------------EAEFGLPCKSPPEISHGVVAH
MSDSYQYGEEVTYKCFEGFGIDGPAIAKCLGEKWSHPPSCIKTD--CLSL
PSFENAIPMGEKKDVYKAGEQVTYTCATYYKMDGASNVTCINSRWTGRPT
CRDTSCVNPPTVQ--NAYIVSRQMSKYPSGERVRYQCRSPYEM-----FG
DEEVMCLNGNWTEPPQC--KDSTGKCGPPPPIDNGDITSFPLSVYAPASS
VEYQCQNLYQLEGNKRITCRNGQWSEPPKCLHPCVISREIMENYNIALRW
TAKQKLYSRTGESVEFVCKRG---YRLSSRSHTLRTTCWDGKLEYPTCAK
R*
```

Script file is a program file written in python script. The alignment file, known protein structure name and unknown sequence name was set as they were kept in alignment file. Other requisite parameters were set properly and the file was saved as "qsec1.py" script file was prepared to generate 20 models for coagulation factor XIII B chain with dope score.

Running Modeller: The script file (qsec1.py) was executed using Modeller command prompt. The model with lowest score was chosen for later validation and docking studies.

Validation of modelled protein structure

The three dimensional structure obtained using Modeller can be validated using validation servers like WHATIF and PROCHECK.

The PROCHECK analysis provides an idea of stereo-chemical quality of all protein chains in a given PDB structure. They highlight regions of proteins which appear to have unusual geometry and provide an overall assessment as the structure as a whole.

The stereo chemical validation of model structure of proteins is an important part of molecular modeling. Firstly, the selection of high quality structures for inclusion in loop dictionaries is important for the simple reason that these coordinate sets will be used to build future models. Secondly, the structure evaluation of comparative modeling output must be used to identify possible problematic regions. It is good to use verification tool for refinement which will improve quality of new structures and probably make structure validation easier.

The RCSB Validation Server allows to check the format of coordinate and structure factor files and to create a variety of validation reports about the structure. The aim of PROCHECK is to assess how normal, or conversely how unusual the geometry of the residues in a given protein structure is, as compared with stereo-chemical parameters derived from well-refined high resolution structures. This also validates protein structure by calculating covalent geometry, Planarity, Dihedral angles, Chirality, Non-bonded interactions, Main-chain hydrogen bonds, Disulphide bonds, Stereochemical parameters and Parameter comparisons etc.

PROCHEK was accessed from http://www.rcsb.org/pdb/home/home.do. Deposit and validate option was selected, X-ray method was selected and then by selecting the options PDB and validate, Ramachandran plot was analyzed.

3.9 Lead identification

Ligand based high-throughput virtual screening:

The smile formats for the existed ligand were submitted to Ligand Info database at http://ligand.info/ to produce the library of chemical compounds. The smile format for these three ligands were generated using Molinspiration cheminformatics (JME Molecular Editor).

The Ligand-Info framework is in view of the supposition that little particles with comparative structure have comparable practical (binding) properties. The created framework empowers a quick and sensitive record based quest for comparable mixes in substantial databases. A Java-based device for grouping and checking of little molecules has been made. The instrument can intuitively gather sets of molecules and make file profiles on the client side and consequently download comparable atoms from a database of 250000 mixes. The published inhibitor (metaformin) of human protein was searched through pubmed, pubchem

and literature search and submitted to Ligand. Info Meta- Database to prepare an in house library of many compounds from more than one million entries (Umamaheswari *et al*., 2010a; Umamaheswari *et al*., 2010b; Grotthuss *et al*., 2003). The framework is accessible at http://Ligand.Info. The application requires the Java Runtime Environment 1.4, which can be automatically installed during the first use on desktop systems. A standalone version of the program is available from the authors upon request.

Virtual screening (VS) is a computational procedure utilized as a part of medication disclosure research. It includes the fast in silico appraisal of expansive libraries of substance structures to distinguish those structures that bind the drug target mostly, commonly a protein receptor or chemical (Rollinger, Stuppner and Langer, 2008). The purpose of virtual screening to come up with hits of novel chemical structure that bind to the macromolecular target of interest. Hence, accomplishment of a virtual screen is characterized as far as discovering new platforms as opposed to numerous hits. Interpretations of virtual screening accuracy should therefore be considered with caution (Sun, 2009). Low hit rates of scaffolds are clearly preferable over high hit rates of already known scaffolds. There are two broad categories of screening techniques: ligand-based and structure-based (McInnes, 2007).

3.10 Computational docking

Docking is a system which predicts the favored introduction of one particle to a second when bound to one another to frame a steady complex (Lengauer et al., 1996). In the field of drug designing configuration, first particle is normally protein while the second one is small organic molecule, potential drug target is utilized to anticipate the quality of affiliation or binding affinity between two molecules using for example scoring functions. Hence, docking plays an important role in the rational drug design (Kitchen *et al.,* 2004). A number of powerful software programs, e.g. VLifeDock, Schrodinger software (Glide), Auto Dock, HEX, GOLD, FlexX, DOCK, Surflex, LigandFit, have been developed over the past several decades to carry out the docking.

VLifeDock:

VLifeDock provides its users with a choice of methods for molecular docking. Grid based docking, GA docking and VLife's own GRIP docking are the three methods that offer unique 'Accuracy - Speed' options to the user from rapid screening to complete precision docking. VLifeDock also has batch docking facility to enable prioritizing the ligands based on

their binding scores easily. The choice of systematic and stochastic methods with an array of scoring functions makes VLifeDock a truly versatile application. VLifeDock is available for both Linux and Windows® operating systems. The docking score and binding orientations of the proposed leads, and known coagulation factor XIII B chain ligand were compared to validate the docking predictions.

Chapter – 04

RESULTS AND DISCUSSION

4. Results and discussion

4.1 Retrieval of FXIII B human sequence

The protein sequence for human Coagulation factor FXIII B was aquired from UniProt**.** The Fasta sequence was displayed in table 1.

Table1. Sequence of human FXIII B retrieved

>sp|P05160|F13B_HUMAN Coagulation factor XIII B chain OS=Homo sapiens GN=F13B PE=1 SV=3

MRLKNLTFIIILIISGELYAEEKPCGFPHVENGRIAQYYYTFKSFYFPMSIDKKLSFFCL

AGYTTESGRQEEQTTCTTEGWSPEPRCFKKCTKPDLSNGYISDVKLLYKIQENMRYGCAS

GYKTTGGKDEEVVQCLSDGWSSQPTCRKEHETCLAPELYNGNYSTTQKTFKVKDKVQYEC

ATGYYTAGGKKTEEVECLTYGWSLTPKCTKLKCSSLRLIENGYFHPVKQTYEEGDVVQFF

CHENYYLSGSDLIQCYNFGWYPESPVCEGRRNRCPPPPLPINSKIQTHSTTYRHGEIVHI

ECELNFEIHGSAEIRCEDGKWTEPPKCIEGQEKVACEEPPFIENGAANLHSKIYYNGDKV

TYACKSGYLLHGSNEITCNRGKWTLPPECVENNENCKHPPVVMNGAVADGILASYATGSS

VEYRCNEYYLLRGSKISRCEQGKWSSPPVCLEPCTVNVDYMNRNNIEMKWKYEGKVLHGD

LIDFVCKQGYDLSPLTPLSELSVQCNRGEVKYPLCTRKESKGMCTSPPLIKHGVIISSTV

DTYENGSSVEYRCFDHHFLEGSREAYCLDGMWTTPPLCLEPCTLSFTEMEKNNLLLKWDF

DNRPHILHGEYIEFICRGDTYPAELYITGSILRMQCDRGQLKYPRCIPRQSTLSYQEPLRT*

4.2 Proteomic analysis of human FXIII B

Primary sequence analysis:

FXIII B contains 661amino acids with a molecular weight of 75510.6 Daltons. The Isoelectric point value is 6.01 the total number of positively and negatively charged residues is 83 and 73 respectively. Aliphatic index is 65.60 and grand average of hydropathicity (GRAVY) was -0.551. The physio-chemical parameters, amino acid and atomic compositions of FXIII B.

Table 2. Predicted physio-chemical parameters of human FXIII B through ProtParam

S.No	Physico-chemical Property	Values
1.	Number of Amino Acids	661
2.	Molecular weight	75510.6
3.	Total number of positively charged residues(Arg+Lys)	83
4.	Total number of negatively charged residues(Asp+Glu)	73
5.	Theoretical pI	6.01
6.	Formula	$C_{3343}H_{5111}N_{885}O_{1013}S_{50}$
7.	Total number of atoms	10402
8.	Instability index	46.13
9.	Aliphatic index	65.60
10.	Grand average of hydropathicity (GRAVY)	-0.551

Aliphatic index is the relative volume occupied by aliphatic side chains like alanine, valine, isoleucine and leucine, which is regarded as a positive factor for the increase of Thermo stability of globular proteins. GRAVY (Grand average of hydropathicity index) shows the solubility of the proteins: positive GRAVY (hydrophobic), negative GRAVY (hydrophilic). As the GRAVY value is -0.551, it indicates that the molecule is hydrophilic.

Table 3. Amino acid composition of human FXIIIB predicted through ProtParam

Amino acid	Composition
Ala (A) 18	**2.7%**
Arg (R) 28	**4.2%**
Asn (N) 31	**4.7%**
Asp (D) 22	**3.3%**
Cys (C) 40	**6.1%**
Gln (Q) 20	**3.0%**
Gly (G) 49	**9.2%**
Gly (G) 49	**7.4%**
His (H) 17	**2.6%**
Ile (I) 34	**5.1%**
Leu (L) 51	**7.7%**
Lys (K) 45	**6.8%**
Met (M) 10	**1.5%**
Phe (F) 21	**3.2%**
Pro (P) 41	**6.2%**
Ser (S) 46	**7.0%**
Thr (T) 45	**6.8%**

Trp (W) 10	**1.5%**
Tyr (Y) 43	**6.5%**
Val (V) 29	**4.4%**
Pyl (O) 0	**0.0%**
Sec (U) 0	**0.0%**

Prediction of functional elements of human FXIII B:

The human FXIII B sequence was scanned against PROSITE patterns and profiles, which showed the patterns along with their ID's and sites. BLOCKS database was used to find the conserved regions and Pfam for the domains of the protein.

Pattern prediction:

The use of protein sequence patterns or motifs to determine the function of proteins is an essential tool of sequence analysis. Significant patterns are not found in FXIII B.

BLOCKS prediction:

The possible hits (motif) reported through BLOCKS database for FXIII B are 2. The first hit with ID IPB000436, with protein name Sushi domain/SCR domain/CCP module, 3of 3 blocks were observed with 1.3e-06 Value indicating that it is most significant match. ID of the second hit is IPB005533, with protein name AMOP domain 1 of 16 blocks were observed with 0.0005 E-value indicating that is less significant.

Table 4. Prediction of BLOCKS regions

PREDICTION	No. OF HITS	FAMILY	BLOCKS	COMBINED E-VALUE
FXIII B	2	Sushi domain/SCR domain/CCP module	3 of 3	1.3e-06
		AMOP domain	1 of 16	0.0005

Pfam:

Human, coagulation factor XIII B chain has three significant matches of pfam-A and no match with pfam-B. The pfam-A matches was shown in table 5.

Table 5: pfam-A matches for FXIII B

Source	Name	Start	End
Pfam	**Sushi**	25	87
	Sushi	91	146
	Sushi	153	208
	Sushi	213	267
	Sushi	274	327
	Sushi	336	389
	Sushi	524	578

4.3 Phylogenetic analysis of protein

Multiple sequence alignment:

The Multiple sequence alignment is performed for FXIII B and other 29 cardiovascular causing proteins using ClustalW. The 30 cardiovascular disease causing proteins and their corresponding UniProt are shown in the table 6.

Table 6. List of 30 cardiovascular disease causing proteins with their UniProt IDS.

S.No	Protein names	Uniprot IDS
1	Complement component c8 alpha chain	P07357
2	Cathepsin D	P07339
3	Macrophage colony stimulating factor 1 receptor	P07333
4	Growth factor receptor bound protein-2	P62993

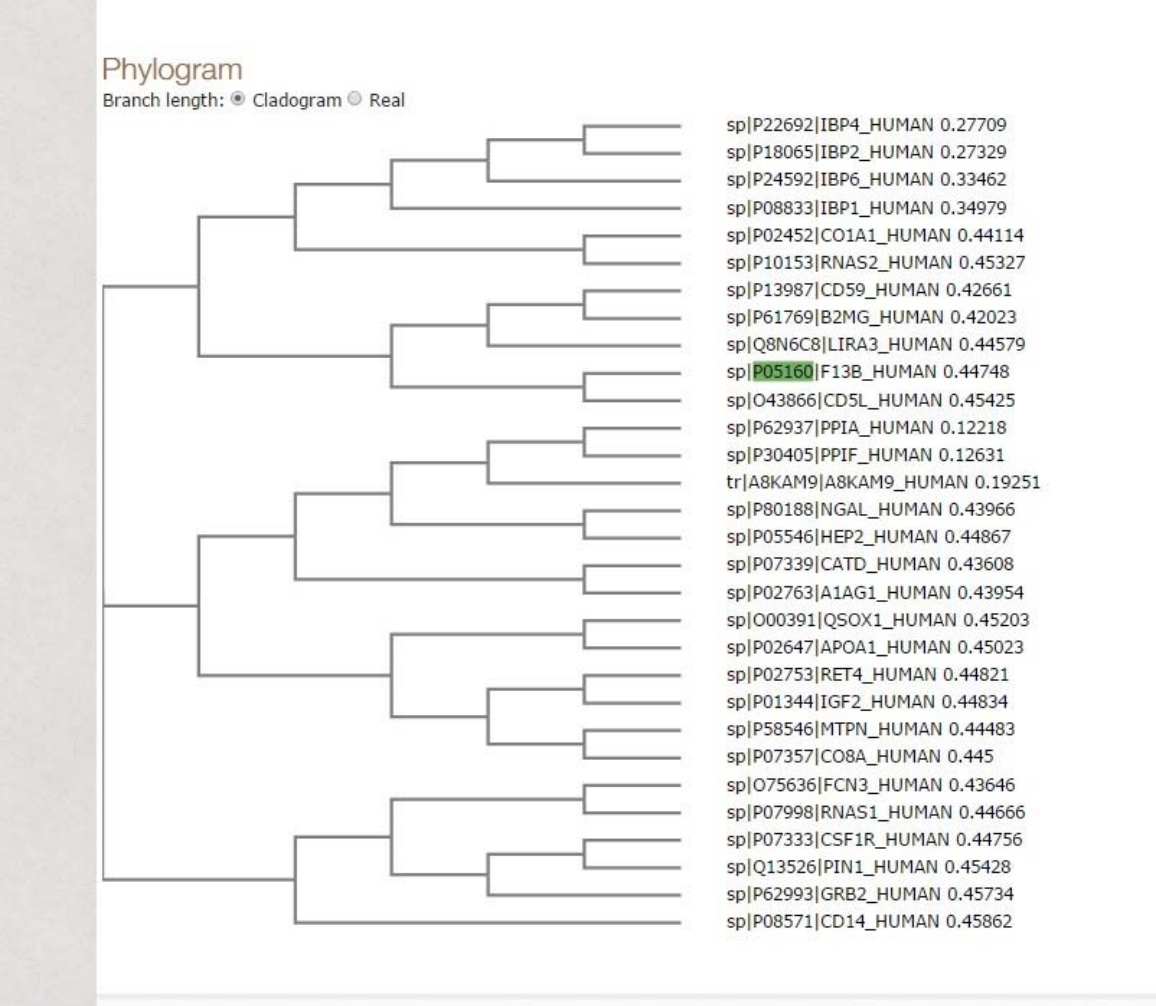
5	Ficolin-3	O75636
6	Sulphydryl oxidase-1	O00391
7	Heparin cofactor-2	P05546
8	Apolipoprotein A-1	PO2647
9	Insulin like growth factor binding protein -1	P08833
10	Alpha -1 acid glycoprotein 1	P02763
11	Peptidyl-prolyl cis trans isomerase F mitochondrial	P30405
12	Coagulation factor XIII B chain	P05160
13	Collagen alpha -1(1) chain	P02452
14	Neutrophil gelatinase-associated lipocalin(NGAL)	P80188
15	Peptidyl-prolyl cis-trans isomerase A(PPIA)	P62937
16	CD5 antigen-like(CD5L)	O43866
17	Retinol-binding protein 4(RET 4)	P02753
18	Insulin-like growth factor-binding protein 2(IBP2)	P18065
19	Leukocyte immunoglobulin-like receptor subfamily A member	Q8N6C8
20	Non-secretory ribonuclease(RNAS2)	P10153
21	Insulin-like growth factor-binding protein 6(IBP6)	P24592
22	Insulin-like growth factor II(IGF2)	P01344
23	Peptidyl-prolyl cis-trans isomerase NIMA-interacting 1(PIN 1)	Q13526
24	Myotrophin(MTPN)	P58546
25	Ribonuclease pancreatic(RNAS 1)	P07998

26	Monocyte differentiation antigen CD14(CD14)	P08571
27	cDNA FLJ78607(A8KAM9)	A8KAM9
28	CD59 glycoprotein(CD59)	P13987
29	Insulin-like growth factor-binding protein 4(IBP4)	P22692
30	Beta-2-microglobulin(B2M)	P61769

Multiple sequence alignment was done by using ClustalW for phylogenetic tree construction.

Phylogenetic tree construction:

The phylogenetic tree was checked in ClustalW website accessed at www.ebi.ac.uk after completing the multiple sequence alignment.



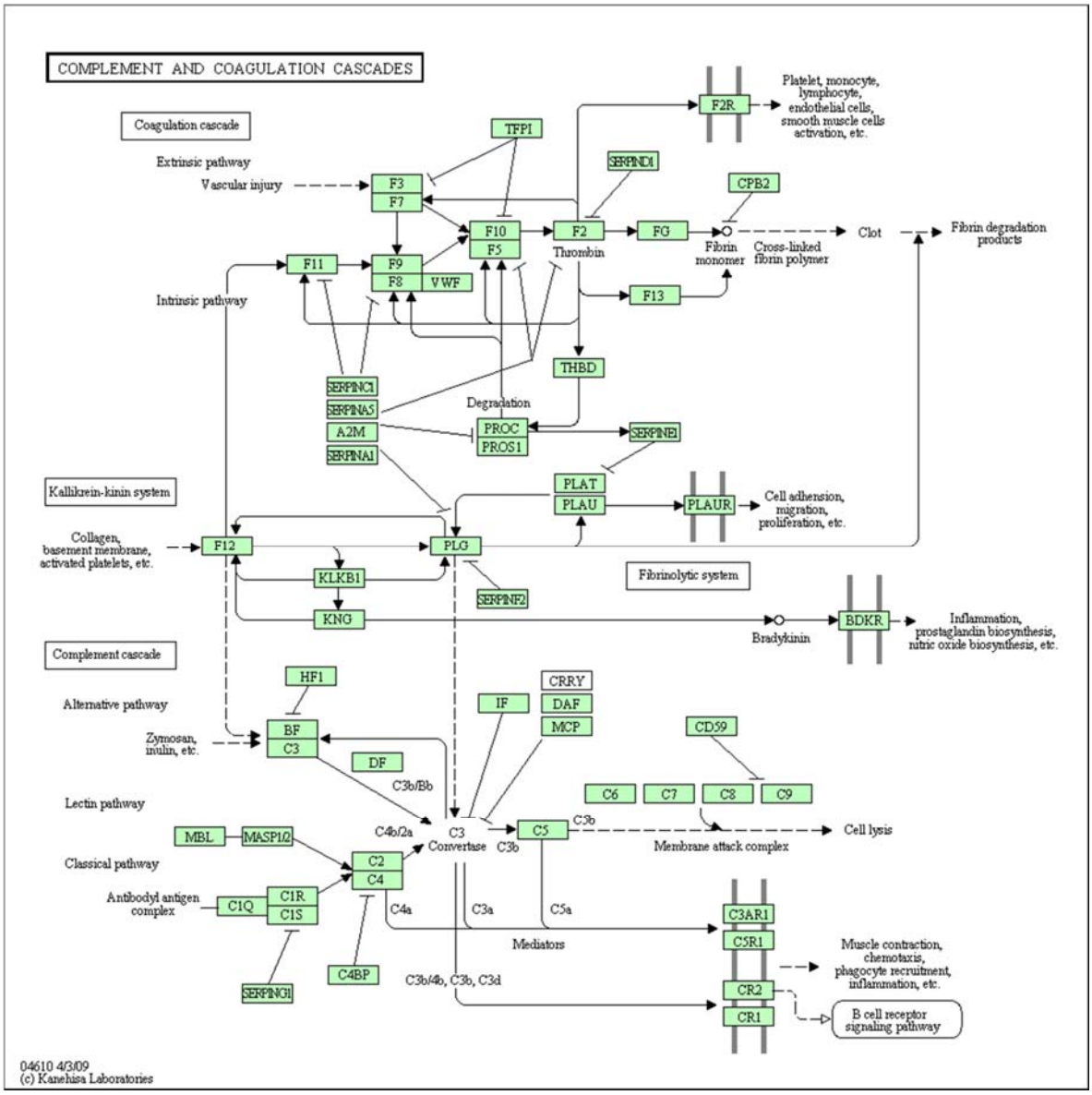
.

Figure. 1 Rooted tree for FXIII B and 29 cardiovascular disease causing proteins

Figure. 1 shows that FXIII B (p05160) and CD5 antigen-like (O43866) are much closely related. All the above 30 proteins are the cardiovascular causing proteins. So this tree also reveals that the human coagulation protein belongs to the family of the cardiovascular disease causing proteins.

4.4 Pathway analysis

The complement and coagulation cascades pathway in which the FXIII protein is involved was retrieved from KEGG Database. The complement and coagulation cascades pathway and the position of the FXIII are shown in the figure 2.



.

Figure 2. Complement and coagulation cascade Pathway

4.5 Three dimensional structure prediction of human FXIII B

The human coagulation factor XIII B protein was queried against PDB using BLASTP. Chain Solution structure of the c-terminal SCR-16/20 fragment of complement factor H of human was taken as template sequence (having PDB ID 2QFH). The structure of template was downloaded from PDB. The alignment file was generated using ClustalW. Alignment file, PDB file of template with were initialized in the modeller script file. The program saved as "tseq1.py" (script file) was run through Modeller to generate model for human coagulation factor XIII B protein.

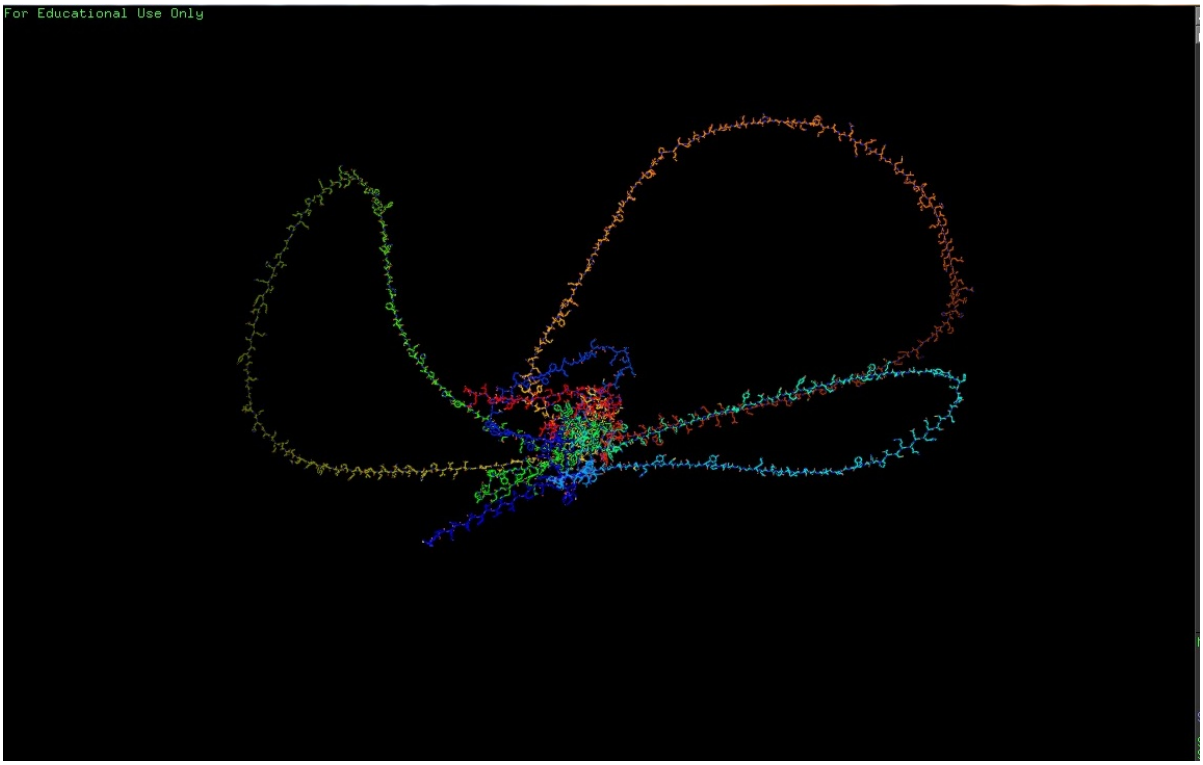


Figure 3. 3D structure of the XIIIB protein

4.6 Lead identification

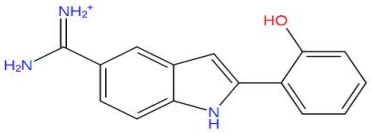
Ligand based high-throughput virtual screening:

The active site residues are ILE-9 which was searched through literature for structure based virtual screening from the 1151 compounds of in house ligand library. LigPrep was used to generate multiple conformations (toutomers, sterioisomers) 2561 compounds of in house library to dock into binding site generated for virtual screening.

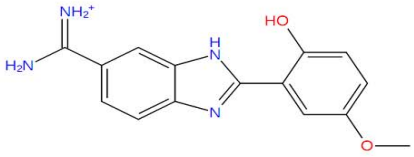
4.7 Docking Studies

The conformations after filtering for ADME (absorption, distribution, metabolism, excretion) properties and reactive functional group was subjected to docking procedure those uses softer to high precision penalty to steric classes, were followed subsequently. 6 lead molecules were selected based on lower G score through virtual screening. Docking is done using VLife software. All the 6 lead molecules were subjected to docking with the receptor protein human coagulation factor XIII B. Structures of 6 lead molecules generated through virtual screening were shown in figure below.

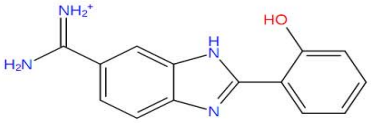
Analysis of docking complex based on docking scores revealed that the six leads showed good binding affinities towards human coagulation factor XIII B. Among the six leads two leads have better docking score than that of metaformin and it also has good capability of interacting with the ligand binding site. Hence, in the present study two leads were proposed as potential inhibitors against human coagulation factor XIII B. The highest binding affinity and the least docking score of -5.321 Kcal/mol was observed as a best lead than that of metaformin with the docking score of -5.239 Kcal/mol. The structures of six proposed leads with their docking scores were shown in figure 4.



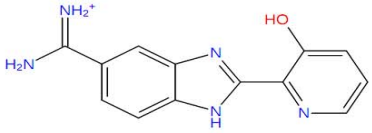
Title: lead1
Score: -5.32



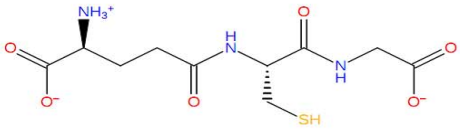
Title: lead2
Score: -5.27



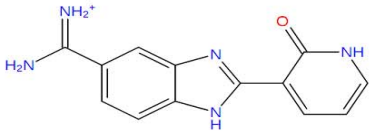
Title: lead3
Score: -4.64



Title: lead4
Score: -4.17



Title: lead5
Score: -2.63



Title: lead6
Score: -3.81

Figure 4. Structures of six proposed lead

Arg-34, Glu-31, Thr-7 residues were involved in formation of hydrogen bond network. Out of the 6 lead molecules, the lead '1', having the least docking score of -5.21 was taken. The docking complex of human coagulation factor XIII B with lead '1' molecule and the hydrogen bonds formed by lead '1' molecule with human coagulation factor XIII B were given in figure5 respectively. The amino acid residues involved in binding with the lead 1 were

similar to the residues predicted at the binding site which evidences the correct prediction in sequence analysis.

.

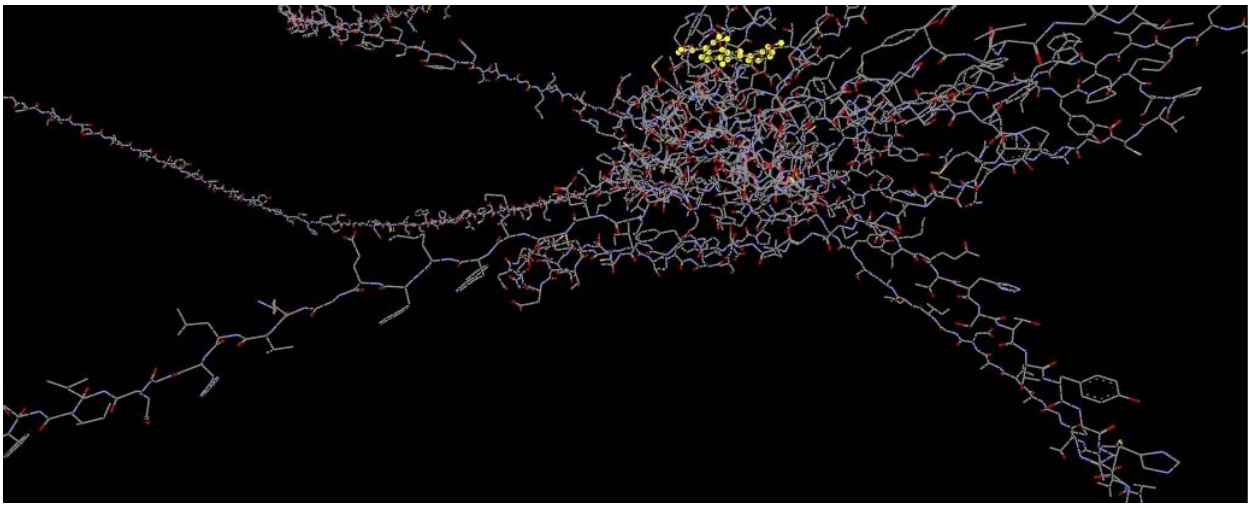


Figure 5. Docking complex of human coagulation factor XIII B with lead ‘1’ molecule.

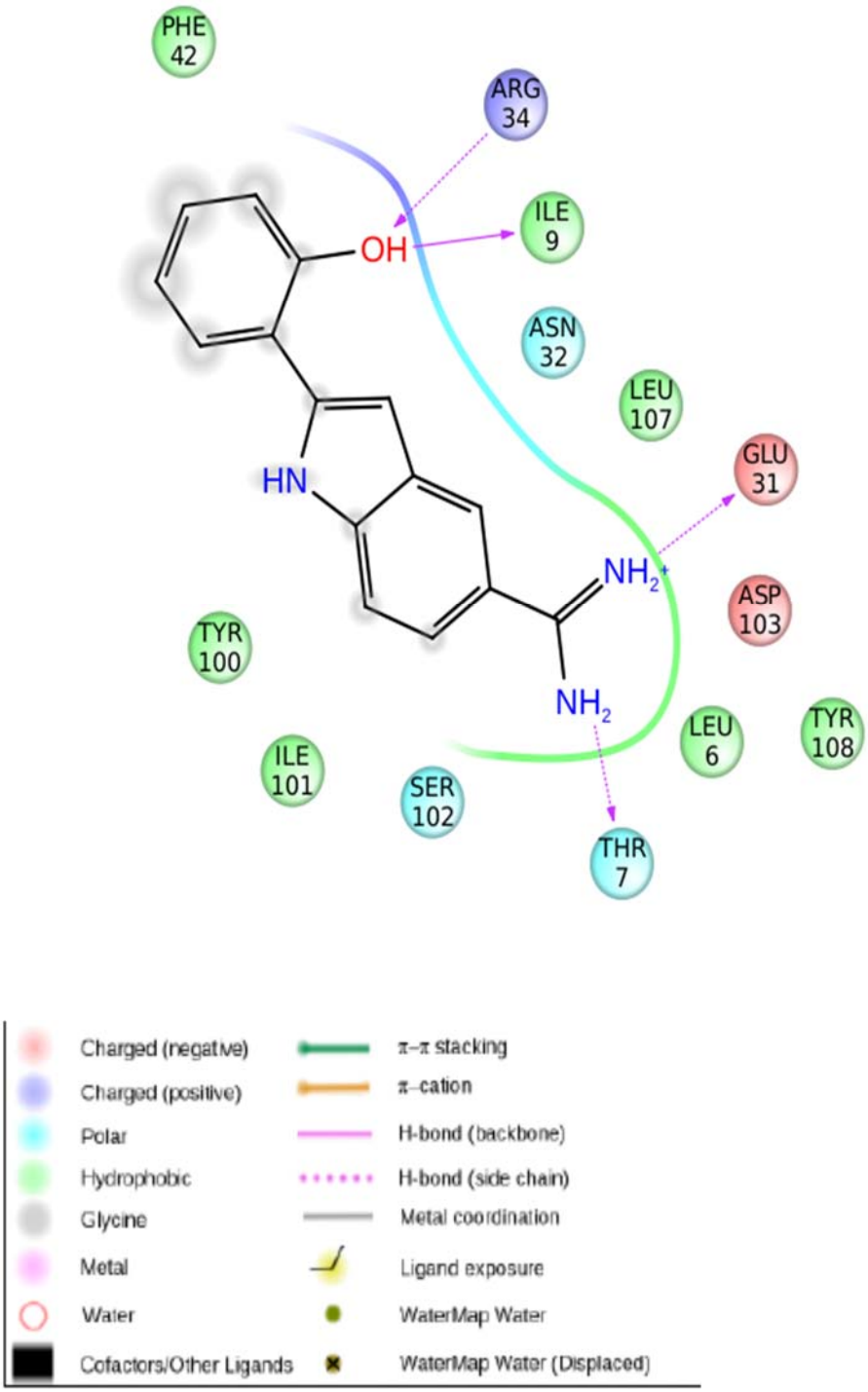


Figure 6. Docking interactions of lead '1' molecule with the human coagulation factor XIII B.

4.8 Prediction of ADME properties

The physically significant descriptors and pharmaceutically relevant properties of inhibitors, were molecular weight, H-bond donors, H-bond acceptors, log P (octanol/water), log PMDCK, log Kp (skin permeability), human oral absorption and their position according to Lipinski's rule of 5 were analysed. Lipinski's rules of 5 is a rule of thumb to evaluate drug likeness, or determine if a chemical compound with a certain pharmacological or biological activity has properties that would make it a likely orally active drug in humans. The rule describes molecular properties important for drug's pharmacokinetics in the human body including its ADME. However the rule does not predict if a compound is pharmacologically active (Lipinski 2001). Based on Lipinski's rule of 5 the two leads which were proposed have showed allowed values for the properties analysed and exhibited drug-like characteristics, and also has good pharmacological properties. Among the proposed leads, the lead 1 molecule with the lowest docking score effectively binds with coagulation factor XIII B and blocks the functional activity. Human coagulation factor XIII B up regulation was observed in atherosclerosis, myocardial infarction etc. The lead 1 molecule can regulate the elevated levels of coagulation factor XIII B efficiently during the cardiovascular disease condition.

Table 7. Principle descriptors of six leads using QikProp3.2

Lead no.	MW	rotor	dipole	SASA	FOSA	FISA	PISA	WPSA	PSA	volume	donorHB	acceptHB	IP (eV)	EA (eV)	glob
1	251.2	4	6.3	486.9	0.00	162.4	337.1	0.00	115.8	828.0	5.00	2.2	8.1	0.4	0.85
2	282.3	5	6.3	499.6	92.7	180.9	258.0	0.00	79.8	891.6	5.00	4.5	8.4	0.7	0.84
3	252.2	4	4.5	531.8	0.00	188.2	308.8	0.00	101.2	891.2	5.00	3.7	8.4	0.4	0.85
4	253.2	4	9.4	497.0	0.00	199.8	289.9	0.00	95.4	804.8	5.00	4.7	8.4	0.7	0.85
5	307.3	11	5.0	489.7	175.4	296.1	0.00	62.9	104.4	917.7	4.5	8.0	9.7	0.3	0.85
6	253.2	3	10.4	534.3	0.00	220.7	266.1	0.00	189.5	800.3	4.0	4.5	8.4	0.8	0.85

The abbreviations used to specify different principal descriptors of ten probable leads in the table and their range in 95% of the available drugs are given. The Range for properties of 95% drug is given based on QikProp3.2.

(Range 95% of Drugs)

MW = Molecular Weight	(130.0 / 725.0)
Rotor=No. of Rotatable Bonds	(0.0 / 15.0)
Dipole=Dipole Moment	(1.0 / 12.5)
SASA =Total solvent accessible surface area	(300.0 /1000.0)
FOSA =Hydrophobic solvent accessible surface area	(0.0 / 750.0)
FISA =Hydrophilic solvent accessible surface area	(7.0 / 330.0)
PISA =Carbon Pi solvent accessible surface area	(0.0 / 450.0)
WPSA=Weakly Polar solvent accessible surface area	(0.0 / 175.0)
PSA=vdW Polar surface area	(7.0 / 200.0)
Volume=Molecular Volume (A^3)	(500.0 /2000.0)
Donor= Donor - Hydrogen Bonds	(0.0 / 6.0)
AccptHB= Acceptor - Hydrogen Bonds	(2.0 / 20.0)
IP (eV) =Ionization Potential	(7.9 / 10.5)
EA (eV) =Electron Affinity	(-0.9 / 1.7)
Glob=Globularity	(0.75 / 0.95)

Table 8. ADME properties of proposed two leads using QikProp 3.2

Lead no.	log o/w	logs	Clogs	logBB	logKp	logKhsa	PMDCK	PCaco	%human oral absorption	Rule OfFive	Rule Of Three
1	1.8	-2.8	-3.7	-1.1	-5.0	-0.2	127.5	285.3	81.5	0	0
2	1.3	-3.0	-3.7	-1.4	-5.5	-0.3	82.3	190.3	75.4	0	0
3	1.1	-3.0	-2.7	-1.4	-5.8	-0.3	69.4	162.5	73.0	0	0
4	0.6	-3.7	-3.3	-1.5	-7.9	-0.4	52.7	126.2	68.3	0	0
5	-2.3	-2.5	-3.1	-2.2	-6.4	-1.4	0.2	0.104	0.00	1	2
6	0.7	-2.8	-3.3	-1.6	-5.0	-0.3	32.2	79.8	65.2	0	0

(Range 95% of Drugs)

LogP o/w = log P for octanol/water	(-2.01 / 6.5)
LogS= log S for aqueous solubility	(-6.53 / 0.5)
CIlogS= log S - conformation independent	(-6.55 / 0.5)
LogBB= log BB for brain/blood	(-3.03/ 1.2)
Log KP = log KP for skin permeability	(KP in cm/hr)
Log Khsa= log K hsa Serum Protein Binding	(-1.51 / 1.51)
Lipinski Rule of 5 Violations	(maximum is 4)
Jorgensen Rule of 3 Violations	(maximum is 3)
% Human Oral Absorption in GI (+-20%)	(<25% is poor)
Apparent Caco-2 Permeability (nm/sec)	(<25 poor, >500 great)
Apparent MDCK Permeability (nm/sec)	(<25 poor, >500 great)

Chapter - 05

SUMMARY AND CONCLUSION

5. Summary and Conclusion

This chapter concludes the technical sum-up of the thesis work on *Designing novel lead molecules for human Coagulation factor XIII B involved in cardiovascular diseases.*

5.1 Summary

Blood coagulation factor XIII (likewise called fibrin-stabilizing component) assumes an imperative part in cluster adjustment by cross connecting fibrin chains. Factor XIII (FXIII) was found more than 70 years back and studies showed the insolubility of fibrin clusters in the vicinity of calcium. Blood coagulation FXIII is a transglutaminase made out of two A- and two B-subunits circling in plasma as a tetramer (A_2B_2). The B-subunit serves as a transporter for the catalytic A-subunit in plasma, is incorporated by the liver, and is emitted as a monomer that ties free A in plasma. The A-subunit is partitioned into 4 spaces, assigned the _-sandwich, the catalytic center, barrel 1, and barrel 2. It contains an initiation peptide of 37 amino acids that confines the entrance of the substrate to the dynamic site cysteine.

The B-subunit is a measured protein made out of 10 repeated Sushi or glycoprotein-1 domains 16, 17. Each Sushi areas contains 2 disulfide bridges that support its tertiary structure, adding up to a sum of 40 cysteine deposits and 20 disulfide connects in the developed B-subunit protein. The primary function of the B-subunit is the stabilization and transport of the hydrophobic A-subunit in the aqueous environment of human plasma. Elevated factor XIII levels causes an illness called atherosclerosis, though dimethylbiguanide (metformin) interferes with component XIII initiation and with fibrin polymerization. At the point when the illness is treated with metformin more than a 12 week period, component XIIIB antigen and reactivity levels were diminished and it controls the infection.

5.2 Conclusion

Coronary heart illness and stroke rate is higher in south and eastern Indian populaces. The cardiovascular system is mainly damaged due to one of the membrane bound protein, where a wide range of evidence indicates that Coagulation factor XIII B is important in the complement and coagulation cascade of vascular complications. Human Coagulation factor XIII B is supplement framework and its effect was seen in atherosclerosis, myocardial infarction and so on. Keeping in mind the end goal to find strong inhibitors subsequently an *in silico* work has been completed to distinguish powerful inhibitors for the protein. In the present study proteomic analysis, phylogenetic analysis, docking was performed to Coagulation factor XIII B. In Proteomic analysis examples, BLOCKS, catalytic domains were predicted. Coagulation factor XIII B with the other 29 cardiovascular disease causing genes, delineated close relationship among them justifying role of Coagulation factor XIII B as a cardiovascular protein. Docking was performed and results were analyzed where 6 potential leads were identified. Among them lead 1 molecule with least docking score -5.321 was identified as best lead. It is a new immunosuppressive drug; it has fewer side effects. *In silico* pharmacokinetics study for the proposed two leads are within the range of parameters involved in both pharmacological and ADME properties. The lead 1 molecule can inhibit the elevated levels of Coagulation factor XIII B efficiently during the cardiovascular disease condition. Hence, lead 1 act as an effective drug against Coagulation factor XIII B if synthesized and validated in animal models.

REFERENCES

[1] Kolher H.P, Grant P.J, February 1, 1999; "The role of factor XIIIVal34Leu in cardiovascular disease"; from http://dx.doi.org/10.1093/qjmed/92.2.67

[2] Robert A. S. Ariëns, August 1, 2002; "role of factor XIII in fibrin clot formation and effects of genetic polymorphism"; from http://qjmed.oxfordjournals.org/content/qjmed/92/2/67.full.pdf

[3] Kolher H.P, Grant P.J, February 1, 1999; "The role of factor XIIIVal34Leu in cardiovascular disease"; from http://dx.doi.org/10.1093/qjmed/92.2.67

[4] Robert A. S. Ariëns, August 1, 2002 "role of factor XIII in fibrin clot formation and effects of genetic polymorphism"; from http://qjmed.oxfordjournals.org/content/qjmed/92/2/67.full.pdf

[5] Okada K, *et al*. Clinical Usefulness of Additional Treatment with Ezetimibe in Patients with Coronary Artery Disease on Statin Therapy. Circ J. 2011; 75: 2496-2504.

[6] Wang SB, *et al*. Redox Regulation of Mitochondrial ATP Synthase: Implications for CardiacResynchronization Therapy. Circ Res. 2011; 109: 750-757.

[7] Kloczko J, Wojtukiewicz M, Bielawiec M, Zarzycka B, Kinalska I. Plasma Factor XIII and some other haemostasis parameters in patients with diabetic angiopathy. Acta Haemat 1986; 76:81–5.

[8] Francis CW, Connaghan DG, Scott WL, Marder VJ. Increased plasma concentration of cross-linked fibrin polymers in acute myocardial infarction. Circulation 1987; 6:1170–7.