Mapping the Interactomics of the BBSome Complex; The Driving Motor of the Cell

By:

Madeeha Jehan Azhar

(Reg # 26/FBAS/MSBI/F09)

Supervisor:

Dr. Asif Mir

Assistant Professor

Department of Environmental Sciences

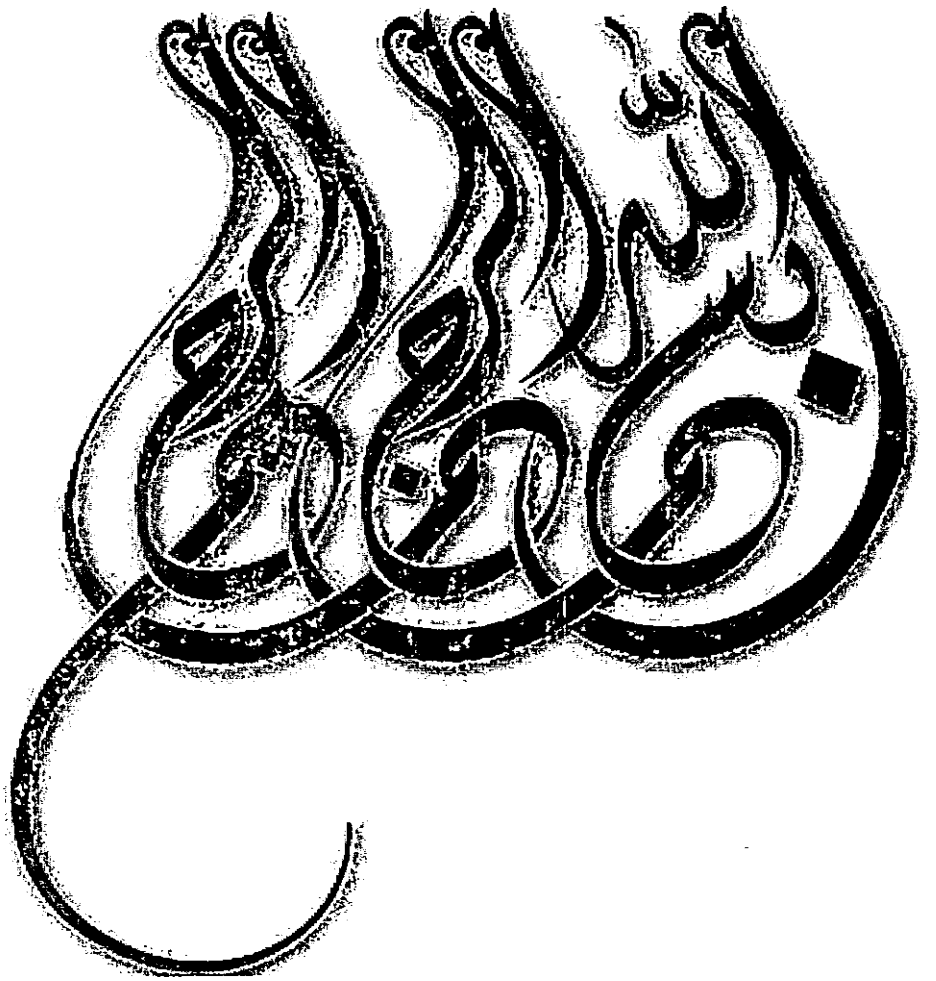
Faculty of Basic & Applied Sciences

International Islamic University, Islamabad

2012

1- Driving motor vehicles; recreation

2- Automobile driving : recreation



Department of Environmental Sciences

International Islamic University, Islamabad

Faculty of Basic and Applied Sciences

Dated: _____

Final Approval

It is certified that we have read the project report submitted by Ms. Madeeha Jehan Azhar Reg No. 26/FBAS/MSBI/F09. It is our judgment that this project is of sufficient standard to warrant its acceptance by the International Islamic University, Islamabad for Masters Degree in Bioinformatics (MSBI).

Committee

External Examiner:

Dr. M. Ismail

Dy. Director

Institute of Biomedical & Genetic Eng.

KRL, Islamabad

Internal Examiner:

Dr. Asma Gul

Assistant Professor

Chairperson, DES

IIUI, Islamabad

Supervisor:

Dr. Asif Mir

Assistant Professor (DES)

IIUI, Islamabad

A dissertation submitted to the

Department of Environmental Sciences,

International Islamic University, Islamabad

As a partial fulfillment of the requirements

For the award of the degree of

Masters in Bioinformatics

Declaration

I hereby declare that this research neither as a whole nor as a part thereof has been copied out from any source. It is further declared that I have carried out this research entirely on the basis of my personal efforts made under the sincere guidance of my teachers and supervisor. No portion of the work presented in this report has been submitted in support of any application for any other degree or qualification of this or any other university or institute of learning. If any plagiarism is found in this report, my supervisor/examiners will not be responsible.

Madeeha Jehan Azhar

26/FBAS/MSBI/F09

Acknowledgement

All praise to Almighty Allah, who gave me the understanding, courage and patience to complete this research.

I express my gratitude to my kind supervisor **Dr. Asif Mir,** who provided me opportunity to learn and enhance my knowledge. As my supervisor, he had been ready to help and guide me throughout the research work.

I would also like to thank my teachers in the department especially **Ms. Hina,** for the moral support of my efforts.

And last of all I would like to acknowledge the support of my family members and friends. I would like to admit that I owe all my achievements to my truly, sincere and most loving parents, whose prayers are a source of determination for me throughout my educational career.

In the end I want to present my unbending thanks to all those hands that prayed for my betterment and serenity.

Madeeha Jehan Azhar

(26/FBAS/MSBI/F09)

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

BLAST	Basic Local Alignment Search Tool
FASTA	Fast All
HM	Homology Modeling
PDB	Protein Data Bank
SPP	Scratch Protein Predictor
BBSome	Bardet-Biedl Syndrome
MSA	Multiple Sequence Alignment
SPF	Spherical Polar Fourier
FFT	Fast Fourier Transform
PIR	Protein Information Resource
NN	Neural Network
HMM	Hidden Markov Model
GRAMM	Global Range Molecular Matching

ABSTRACT

Bardet-Biedl Syndrome is a ciliopathic disorder (a genetic disorder of the cellular cilia or the cilia anchoring structures, the basal bodies) that is characterized by a number of various disorders most common of these being obesity, mild hypertonia, mental retardation, rod-cone dystrophy, postaxial polydactyly, male hypogonadism, complex female genitourinary malformations, and renal abnormalities. This syndrome is a pleiotropic disorder with variable expressivity and a wide range of clinical variability. Trafficking flaws to the ciliary membrane are involved in this syndrome.

Beside the identification of 14 novel genes being held responsible for this syndrome, the biochemical mechanisms that lead to Bardet-Biedl syndrome are still unclear.

During the course of this research, protein models involved in the formation of this crucial complex have been designed. Different protein models were generated through various standard molecular modeling bioinformatics tools. Each obtained model was then evaluated through the established in silico methods to obtain the most accurate protein model. In the second phase of study interaction pattern of these proteins was figured out to explore the possible role of each protein in this complex and their interactions with other possible entities. Antigenic sites and interaction pockets have been identified on proteins of the BBSome complex. Major causes for high interaction profiles of members of this complex were mapped and mutation analysis was also done.

Chapter 1

Introduction

1. INTRODUCTION

The biochemical basis of life stands on a variety of different compounds and molecules. Of these proteins can be surely said to as the building blocks of the cell and hence of life too. Proteins are the biochemical compounds; consist of multiple polypeptide units and a proper 3-Dimensional structure for appropriate functioning. The structure of a protein can be well understood at different hierarchical levels i.e. Primary, secondary, tertiary and quaternary, each corresponds to its specific structural elements. The polypeptide units of a protein are held together by a variety of different forces including hydrogen bonding, Van Der Waals forces, ionic interactions and various other non covalent interactions.

The advent of recombinant DNA technology has led to many notable advances in mapping the amino acid sequence of a protein however when it comes to work out the secondary and tertiary structures of a protein, things get quite complicated, intricate and time consuming as well.

The major ways of determining the 3-Dimensional structure of a protein include NMR spectroscopy, X-Ray crystallography and to some extent interferometry is also used for structure determination. Keeping in view the pros and cons of these predictive methods, much stress has been laid on devising computational methods and algorithms to find out the secondary structure of a protein. Protein structure prediction is one of the most highlighted objectives of modern biology, bioinformatics, biotechnology and some vital areas of theoretical chemistry like drug designing. Its applications in the above mentioned disciplines are of great significance. 3D protein structures are of immense concern for the drug designing in many different types of biological and medical experiments, such as

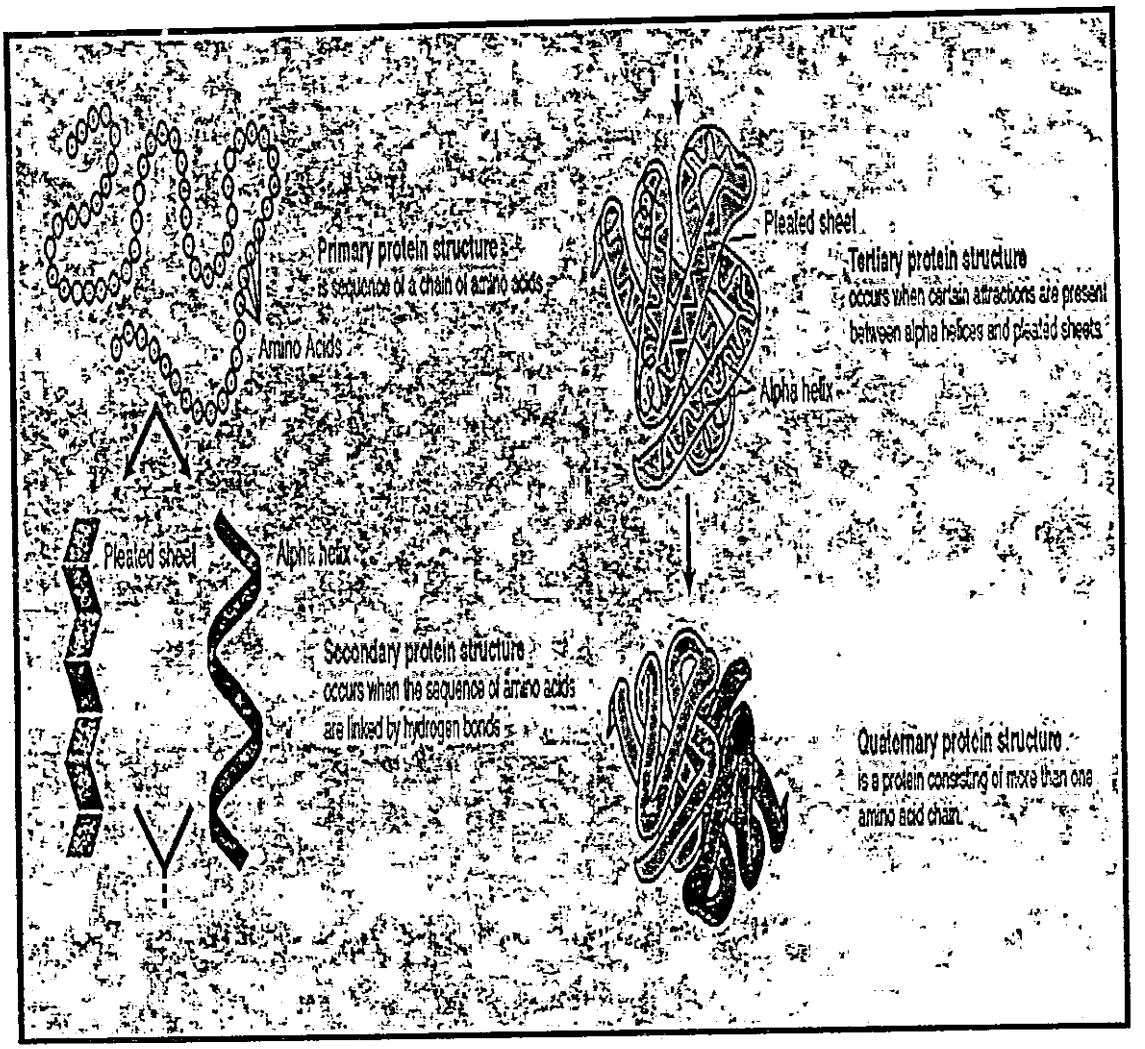


Figure 1.1. Four levels of protein structure assembly (www.umass.edu/molvis/workshop/prot1234.htm)

site-directed mutagenesis or discovery of structures of specific inhibitors (Schwede *et al.* 2003). Among all the theoretical approaches known today, comparative modeling is the only technique that can generate a 3D model of a protein (target) from its linear amino acid sequence of the protein in a quite reliable manner (Tramontano *et al.* 2001).

1.1 Protein-Protein Interactions:

The function of a protein is most commonly defined by the interactions it makes with other proteins and ligands. Biochemical, biological and biophysical studies carried on proteins have revealed the important role of interactions between proteins. PPIs are operative at almost every level of cellular function, in various biological processes and pathways, signal transductions, muscle contractions, genetic expression, to name a few. They are critical for normal cell functioning, structural and genetic organization and apoptosis. Due to their central role in majority of cellular functions, they have been the subject of intensive research for many years. The prediction of viable protein-protein interactions and their role in elucidation of the cellular processes is still a big challenge for biologists today. Computational techniques for the detection and classification of functional pockets on proteins have increasingly become an area of interest for scientists today.

This is mainly due to the various newly solved structures that have inadequately characterized biochemical functions or molecular interactions related. Faced with a speedily mounting number of known protein structures, it has become more significant to have investigative and analytical tools that identify functional sites.The factors that affect the arrangement of protein-protein complexes are usually explored in four different ways of protein-protein complexes. These are homodimeric proteins, enzyme inhibitor

complexes, heterodimeric proteins, and antibody-protein complexes (Jones and Thornton 1996).

1.2 Methods for 3D Structure Prediction of a Protein:

Knowledge of the 3-dimensional organization of a protein is a prerequisite for the rational design of site-directed mutations in a protein and can be of great significance for the design of drugs. At present three approaches are mostly followed for the predicting 3-Dimensional structure of proteins. These are:

1.2.1 Threading:

In this method, the amino acid sequence of a protein whose structure is unknown is examined and tested for its ability or compatibility to fit into a known 3-D structure. If a reasonable extent of compatibility is found with a known structural core, the query protein is predicted to fold into a parallel 3-D configuration. Sequences with identity less than or equal to 30% are mostly subject to this method of structure determination. These methods are undergoing a significant degree of progress at the present time.

1.2.2 Ab initio structure prediction:

This approach takes into account the energy values of the protein in question. Basically it involves modeling all the energetic involved in the process of protein folding. The 2nd step involves selection of the structure with the lowest free energy. However for larger proteins, this technique may become computationally more expensive.

1.2.3 Homology Modeling (HM):

Homology modeling (comparative modeling) is an efficient technique for predicting tertiary structure of a protein, provided there exist homologous proteins whose 3-

Dimensional structural conformations are known. This approach is widely becoming the method of preference for obtaining 3-D coordinates of proteins. The method works for proteins with 25% or greater similarity with sequences of known 3-D structure (Blundell *et al.* 1987). Basically it involves a number of different steps, and each step relies strongly on the results/ outcomes of the previous step. Therefore, errors may be unintentionally introduced and propagated. Homology Modeling is a relatively easy technique, as compared to other two. It does not require expensive experimental facilities. It is entirely a computational process, much easier to implement than the experimental procedures.

Steps involved in HM:

Major steps involved in the process are as under:

1.2.3.1 Template Selection:

The first and the most critical steps of the process is selection of a suitable template. Template identification can be done by running a search query of the leading sequence and structure databases available. The notable such databases are: Protein Data Bank (PDB), Structural Classification of Proteins (SCOP) (Lo Conte *et al.* 2002), Distance-matrix Alignment Comparative Protein Structure Modeling 837 (DALI) (Holm and Sander 1999) and Class, Architecture, Topology and Homology (CATH) (Orengo *et al.* 2002)

Template selection for the query protein can also be categorized into three steps on the basis of sequence alignment approach used. Simplest one being the serial pairwise sequence alignments and it can be aided by sequence search protocols and algorithms like FASTA and BLAST (Altschul *et al.* 1990; Pearson 1990). The process of template

selection greatly depends on the percentage identity between the sequences of interest. 25-30% and greater are required for most of the sequence alignment programs. Similarity less than this threshold can end up with the use of advance techniques like threading.

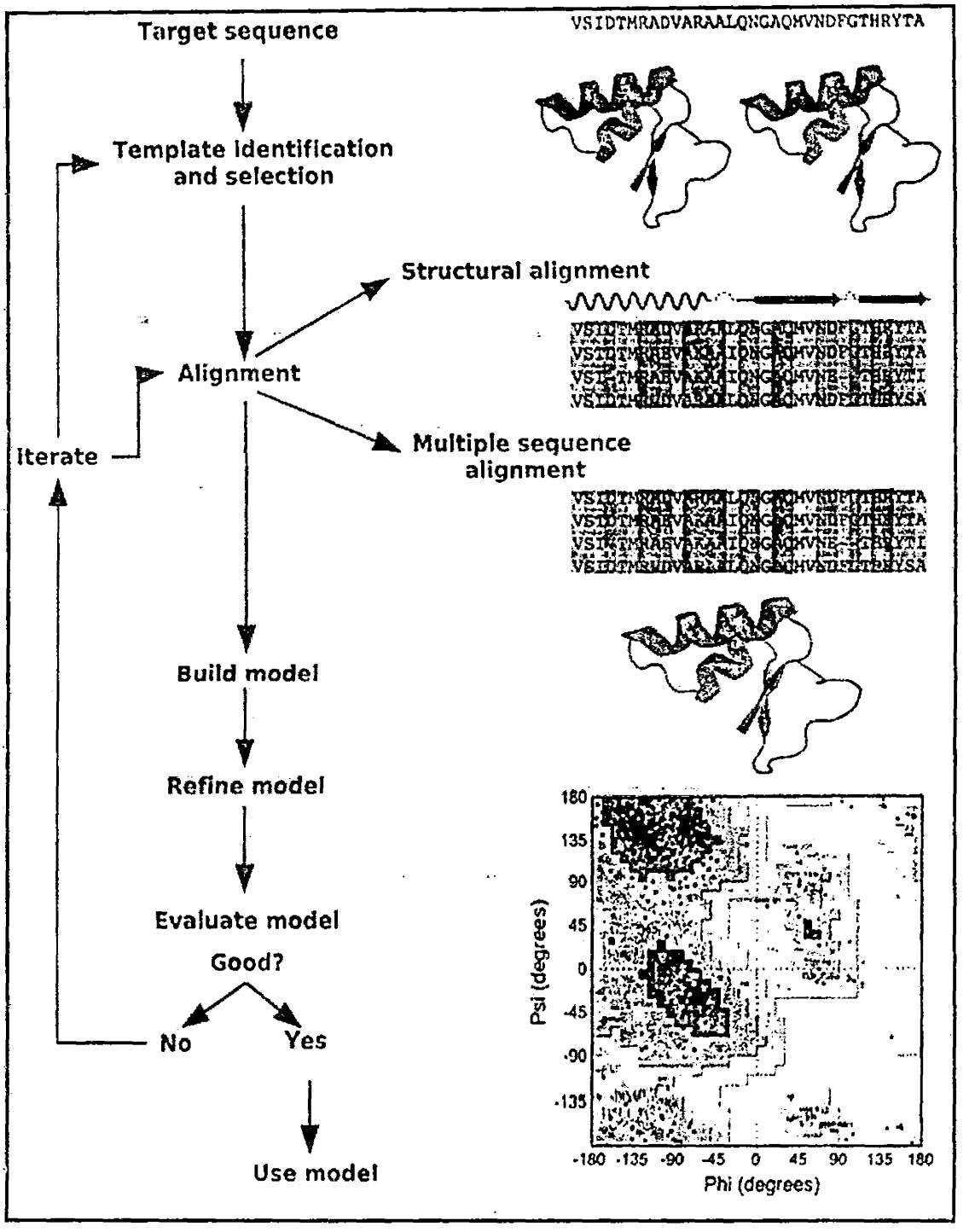


Figure 1.2. Comparative modeling workflow (Bishop *et al.* 2008)

1.2.3.2 Target–Template Alignment:

The next step after the template selection process is to align the target and template sequences. Many techniques are present nowadays for aligning sequences, and sometimes the most puzzling task is deciding which methods to apply to generate an optimal alignment between both sequences. This step is subject to slip-ups and can lead to a faulty model. The (true) alignment signifies the process of evolution giving rise to the diverse sequences starting from the common ancestor sequence and then shifting through mutations (insertions, deletions, and substitutions).

1.2.3.3 Model Building:

a) Modeling by assembly of rigid bodies

This approach uses the information of ample number of small rigid bodies that were previously obtained through alignment of proteins (Blundell *et al.* 1987; Browne *et al.* 1969).

b) Modeling by segment matching or coordinate reconstruction:

As the name indicates, modeling by segment matching involves the identification of guiding positions from the template structure. These positions mostly comprise of the Cα which are conserved in the alignment of template and target proteins. This technique can be employed to model both the side and main chain of a query protein

c) Modeling by satisfaction of spatial restraints:

Modeling by spatial restraints involves many restraints of the template selected and these are mostly angles and distances between aligned residues of template and target sequence. Modeling is done by minimizing the violations associated

with these restraints. The very same technique is implied in MODELLER to obtain models of proteins and it came out to be a promising one of all the comparative modeling techniques.

d) Loop modeling:

Loops often participate in defining the functional specificity of a given protein scaffolds, as they play a major role in deciding the binding sites on a protein. Loop modeling emerges out to be an efficient comparative modeling technique as its usefulness is marked by its role in protein-protein interactions, identification of protein binding sites and in docking studies (Marti-Renom *et al.* 2000).

1.2.3.4 Model Evaluation:

The evaluation phase is a mandatory phase in the course of a comparative modeling project. It is in fact a review checkpoint for the project. Most importantly, the evaluation phase involves for the possible errors in the fold of the newly designed protein structure. Accurate and precise 3D modeling requires both a correct fold assignment and a roughly correct target–template alignment (Sanchez and Sali 1998). As this project includes the modeling of various proteins of the BBSome complex, so the evaluation phase involves checking for the possible errors in the newly designed models, their energy values as well as the favored and non-favored regions within the structure.

The evaluation phase involves reviewing all the save points in the course of a project. The major save points in the model evaluation are:

- Sequence alignment
- Z-Score evaluation

- Ramachandran plot evaluation

Tools used for testing the 3-D model of protein structures are PROSA (Sippl1993) and VERIFY3D (Luthy *et al.* 1992). These programs evaluate the contribution and behavior of each residue in a model with respect to the expected one as found in the high-resolution X-ray structures.

1.3 Bardet-Biedl Syndrome:

Bardet-Biedl Syndrome is a genetically and clinically heterogeneous disorder caused by mutations in almost 14 different proteins that have been identified so far. It is a human genetic disorder which is mostly characterized by obesity, polyadyctaly, mild hypertonia, rod-cone dystrophy, complex female genitourinary malformations, male hypogonadism, nephropathy and retinal degeneration. Limb abnormalities present in the case of Bardet-Biedl Syndrome are majorly those of fingers. Bardet-Biedl Syndrome is also categorized as a developmental and ciliopathic disorder (a genetic disorder of the cellular cilia or the cilia anchoring structures, the basal bodies). The prevalence of this disease is quite low in general population (1/100,000 live births) as compared to various other genetic anomalies. However the reported pervasiveness is quite higher in some isolated regions of the world like Kuwait and Newfoundland. This syndrome is considerable genetic reason of chronic and end-stage renal collapse in children (Tobin and Beales 2007).

The diagnosis of this disorder in most of the cases reported so far is initiated as the vision begins to degrade. Beside all this, the biochemical basis of the syndrome is still unclear. The symptoms of this disorder and their associated percent prevalence can be summarized in table 1.1.

1.3.1 Genes Involved:

The biochemical mechanisms that lead to Bardet-Biedl syndrome are still unclear however fourteen genes responsible for BBS have been identified so far. The products of these genes are called BBS proteins, and multiple evidences have indicated that they are located in basal body and cilia of the cell. Many studies have suggested an important role for BBS proteins in ciliary dysfunction as well as their association with intraflagellar transport (IFT) (Ansley *et al.* 2003). The most significant function of these proteins regarding IFT is that they are an active participant of the process of stabilization of IFT proteins. This process mainly comprises of stabilizing the association between two motor proteins of the IFT mechanism, OSMotic avoidance abnormal 3 (OSM-3) and kinesin (Ou *et al.* 2005). The other vital biological system affected by these BBS proteins is the melanosome transport system.

Recent studies have shown that the proteins involved in BBS are assembled into a complex organization called "BBSome". This complex is known to have been responsible for the transport of intracellular vesicles to the base of cilia.

The active partakers of BBSome complex are:

BBS1, BBS2, ARL6/BBS3, BBS4, BBS5, MKKS/BBS6, BBS7, TTC8/BBS8, B1/BBS9, BBS10, TRIM32/BBS11, BBS12, MKS1/BBS13, and CEP290/BBS14

The contribution of each of these 14 genes to the total reported cases of Bardet-Biedl Syndrome is represented in the form of a pie-chart by Katsanis 2004 (figure 1.3.).

Beside these 14 subunits involved in this syndrome, there are some other contributing factors that add to the complexity of the syndrome. Of these important are the ARF like GTPase, E3 Ubiquitin Ligase and some type II chapheronins. The molecular basis of the

syndrome remains indefinable and quite elusive. While the underlying pathology of all the symptoms of the Bardet-Biedl syndrome remains confusing and enigmatic, animal models designed for investigating the root causes have pointed towards the role of ciliary dysfunction involved. A tabular representation of the inner details of the genes involved in the formation of BBSome complex is given in table 1.2.

Table 1.1. Prevalence summary of Bardet-Biedl Syndrome (Beales *et al.* 1999)

Features	Percent Prevalence
Primary Features	
Rod-cone dystrophy	93%
Post-axial polydactyly	69%
Obesity	72%
Hypogonadism	98%
Renal malfunctions	24%
Secondary Features	
Learning Difficulties	62%
Developmental delay	50%
Speech abnormalities	54%
Diabetes Mellitus	6%
Hearing Loss	21%

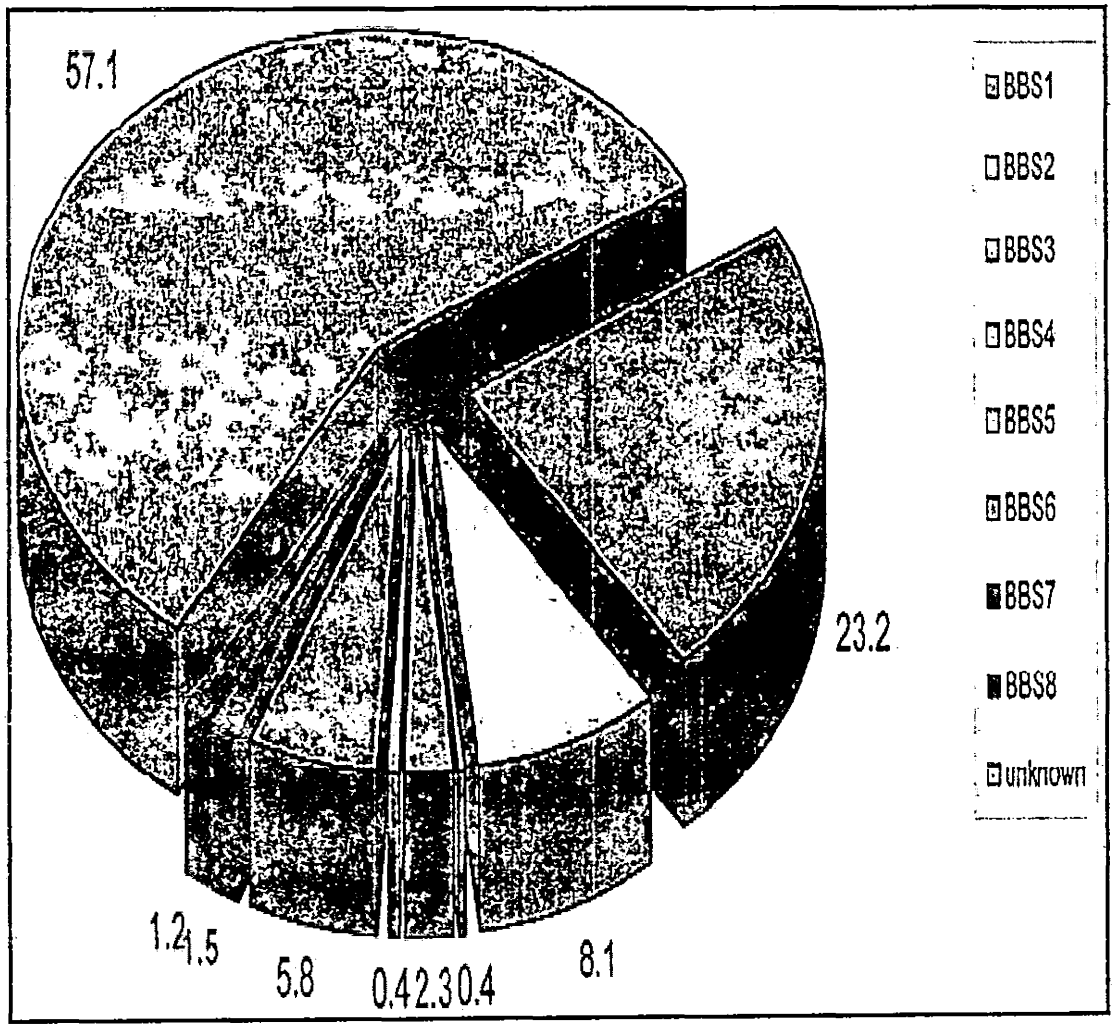


Figure 1.3. Contribution of related BBS gene to the total reported cases

(Katsanis 2004)

1.3.2 BBSome Complex:

BBSome complex is a protein assemblage that is known to have its function in membrane trafficking to and inside the primary cilium. This complex basically encompasses seven proteins of all BBSome proteins known so far. These seven are BBS1, BBS2, BBS4, BBS5, BBS7, BBS8 and BBS9 (Table 1.2). The presence of a novel protein BBIP10 also known as BBS10 is also reported recently. The apparent molecular mass of the complex came out to be 438 kD, and its sedimentation coefficient is 14S. The BBSome subunits are highly conserved and are largely distributed among those organisms which possess cilia, thus signifying that the functionality of the BBSome Complex has also been conserved. Non ciliated organisms like fungi, plants and amoebas lack this complex assembly of proteins.

Experiments have revealed that the BBSome complex is mainly localized to the primary cilium of the cell. Of the 7 subunits of the complex, three are localized primarily to the cilium whereas while the remaining members of the complex have shown their part in the intraflagellar transport system of the cell. The BBSome complex collaborates with the GTPase Rab8 to promote the ciliogenesis of the cell (Nachury *et al.* 2007). As it is evident from the structural details of the cilia, cilia lack ribosomes and thus they lack the sufficient machinery to transport all the necessary proteins required for their assembly and construction. These vital proteins need to be imported first for the cilia to perform properly (Blacque and Leroux 2006). So the BBS proteins basically work as adaptors to help loading of these cargo proteins at the proximal cytoplasmic end to be carried to the distal tip (Blacque *et al.* 2004). The participants of BBSome complex are also thought to have a vital role in shuttling the proteins back to cytoplasm for recycling. Mutations in

these genes result in loss of function of the proteins and hence destroy the whole mechanism of transport and disruption of the IFT machinery (Tobin and Beales 2007).

The interaction pattern of this strange complex of proteins is still unclear that how out of 14 known proteins, only a specific set of protein clumps together? The types of interactions between the components of the complex are elusive enough to uncover the inner details and causative mechanisms of the Bardet-Biedl Syndrome. Moreover their functionality in the proper assembly of primary cilia and intraflagellar transport also needs to be elaborated further. The details of the enzymatic activity of the member proteins of BBSome complex is also an enigma so far.

Table 1.2. Overview of the Structure and Function of Known Genes of the BBSome C

Sr.No.	Gene	MIM No.	Chromosome	No. of Exons	Function	
1.	BBS1	209901	11q13	17 exons	Interaction with RABIN8, ciliary membrane growth	Fetal ti adipose
2.	BBS2	606151	16q21	17 exons	Encodes a protein of unknown function	Brain, k thyroid
3.	BBS4	600374	15q22.3-q23	16 exons	Associated with insulin resistance, functions as an adaptor of the p150 (glued) subunit of the dynein transport machine.	Cellular with mo
4.	BBS5	603650	2q31	12 exons	Necessary for the generation of both cilia and flagella	Ciliated
5.	BBS7	607590	4q27	19 exons	Involved in eye, limb, cardiac and reproductive system development	Almost tissues
6.	BBS8	608132	14q32.1	14 exons	Involved in the formation of cilia,	Testis, pancrea
7.	BBS10	610148	12q21.2	2 exons	May affect the folding or stability of other ciliary or basal body proteins, Assists the folding of proteins upon ATP hydrolysis	Adipog

Chapter 2

Material & Methods

2. MATERIALS & METHODS

2.1 Similarity Search:

Similarity search has become a stronghold of Bioinformatics domain. Similarity search has been effective in determining the function of those genes whose sequence has been resolved in silico. In this basic technique, the query sequence is compared against every sequence in the database and those appearing similar to it are identified (Figure 2.1).

2.1.1 BLAST:

Basic Local Alignment Search Tool is the most employed algorithm for similarity search of a sequence against a huge number of database sequences. It is an algorithm for comparing the primary information present in a sequence whether it is that of a protein or of DNA. The algorithm follows a heuristic based strategy to find out the alignments. HSPs or high scoring segment pairs present in an alignment is the basic idea behind BLAST. The algorithm basically searches for the regions of high scoring alignments between both the query and the database sequence (Altschul *et al.* 1990).

Various alterations in the basic algorithm have led to the formation of multiple BLAST programs like Psi blast, blastp, WU blast, blastx, tblastn etc.

2.1.2 FASTA:

FASTA is an algorithm for aligning the DNA sequences and those of proteins. FASTA algorithm basically looks for similar regions of comparatively shorter

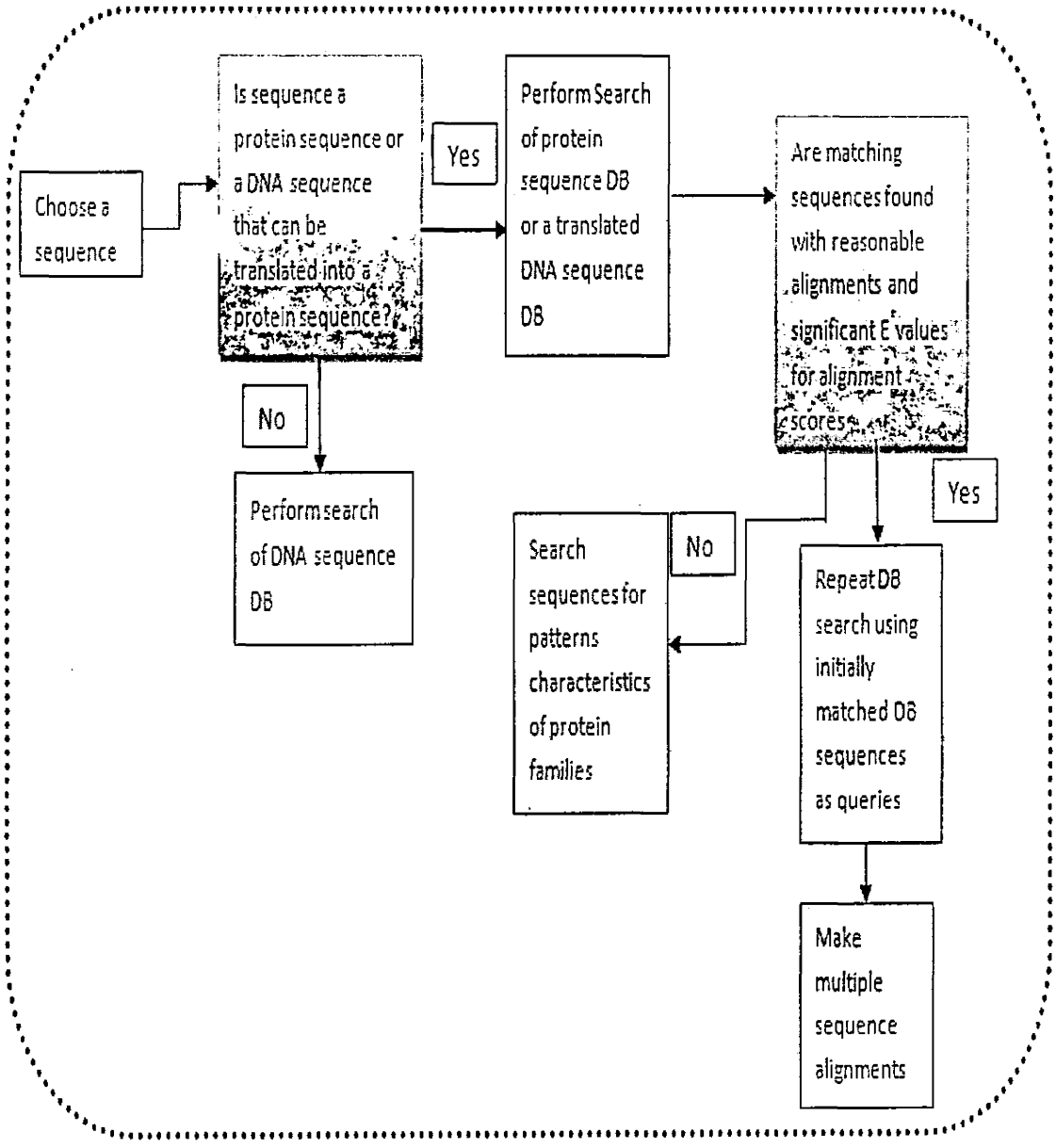


Figure 2.1. Database searching for similar sequences

length in the two sequences and after that it tries to extend those regions of similarity on both sides. FASTA provides a speedy way to locate short stretches of similar sequence between a new sequence and any sequence in a database. The word or k-tuple method is being used by this program. It is also a heuristic based method and usually provides a reliable alignment. Due to the capability of the algorithm to find matching sequences in a sequence database with high speed, FASTA is useful for regular database searches (Pearson and Lipman 1988).

2.2 Post Translational Modifications:

Proteins after being translated undergo a variety of different chemical alterations before being transported to their destination. These changes in the protein structure are broadly termed as post translational modifications. These modifications are greatly responsible for rendering the newly formed protein as a functional one. Some of the crucial post translational modifications are glycosylation, alkylation, acetylation, phosphorylation, sulfation and C-terminus amidation etc.

2.2.1 SIGNALP:

SIGNALP is a server to identify the location and presence of potential cleavage sites in a protein structure. It produces both classification and cleavage site assignment. Novel amino acid composition units and sequence position units have been incorporated in the neural network layer of this server to improve the performance. The input to the SIGNALP server is sequence in FASTA format. The server provides a number of different parameters to formulate the query in order to obtain the required output (Bendtsen *et al.* 2004)

2.3 Topology Prediction:

Topology is usually defined as those features of a structure which remain unchanged even if the structure is distorted. In the world of proteins, topology defines an important focal point between the amino-acid sequence and the entirely folded three-dimensional structure.

2.3.1 TMPRED:

The TMpred program is basically used for predicting membrane-spanning sections and their orientation. The algorithm used for this purpose is based on the statistical analysis of TMbase. The input to the server is the amino acid sequence of the protein. The prediction is made using a amalgamation of numerous weight-matrices for scoring (Hofmann and Stoffel 1993).

PROTEIN STRUCTURE:

Proteins are the macromolecules consisting of a single or multiple polypeptides that fold into specific conformations to achieve the biochemical functionality associated with them. The proteins are well explained and understood with reference to their structure. The structure of a protein is mainly categorized as primary, secondary, tertiary and quaternary structures.

2.4 Primary Structure:

The primary structure of a protein is simply linear sequence of the amino acid constituting that protein. Each protein has its specific amino acid sequence that is in most cases defined by the nucleotide sequence of the encoding gene.

2.4.1 PROTPARAM:

ProtParam is a tool for analyzing the primary structure of a protein. The tool is provided by Expasy Proteomics Server. It computes various physio-chemical properties of a protein including its molecular weight, the amino acid make up of the protein in question, atomic composition and estimated half life of the protein. The input to ProtParam can be the amino acid sequence of the query protein or it can be specified by the Swiss-Prot/TrEMBL accession number (Gasteiger *et al.* 2005).

2.5 Secondary Structure:

When local ordering is introduced into the primary structure of a protein, mainly through hydrogen bonding, the structure is then referred to as the secondary structure of a protein. The most frequent secondary structure elements found in proteins are the alpha helices and the beta sheets (sometime also called β-pleated sheet). The secondary structure of a protein provides useful insight into its functional behavior. Thus predicting the structure of protein holds an important position in the various techniques of computational biology.

2.5.1 PREDICT PROTEIN:

Predict Protein is an automatic source/service for analyzing protein sequence and for the prediction of the protein's structure as well as its function. It integrates the feature prediction for secondary structure elements, transmembrane helices, disordered regions, domain boundaries, sub cellular localization, disulfide bonds and protein-protein and protein-DNA binding sites. The input to Predict protein is the amino acid sequence of the protein and the output may contain a variety of different features of the protein including MSA, sequence motifs, low complexity

regions, domains present in the protein. It also gives a complete insight into the protein's secondary structure elements, transmembrane helices, intra residue contacts, beta barrels and cysteine predictions etc. the evaluation of the prediction accuracy can be obtained upon sending a request to the server (Rost *et al.* 2004).

2.5.2 SCRATCH PROTEIN PREDICTOR:

Scratch Protein Predictor is an online service for predicting the secondary structure elements of a protein, provided the primary sequence. The SPP suite includes multiple options for predicting the structural details of a protein, the most important of them being residue contacts, extent of antigenicity, disulfide bonds, and domain predictions. The performance constraints of SPP can be best judged by the fact that it is more effective in those cases when we have targets without high levels of homology to any of the known structures (Cheng *et al.* 2005).

2.6 Tertiary Structure:

The ultimate folding of the protein chains and the secondary structure elements (alpha helices and beta pleated sheets) result in tertiary conformation of a protein. Disulfide bonds, hydrogen bonding, salt bridges and the non-polar hydrophobic interactions contribute to the stability of the tertiary structure. The most important characteristic of tertiary conformation of a protein structure is that all the polar residues/groups are on the surface while the non-polar ones are buried inside.

2.6.1 MODELLER:

Designing tertiary conformation of protein through MODELLER is a multi step process which involves identification of suitable templates through hits returned by Psi-BLAST. After the selection of template, alignment is performed between

the target and template and after the alignment MODELLER is then used to design tertiary structures of each query protein. Out of many models designed by the MODELLER, best one is selected for further use on the basis of lowest value for the Modeller Objective Function (Sali and Blundell 1993).

2.6.2 SWISS MODEL:

Swiss Model is a fully automated protein homology modeling server that facilitates the user by providing multiple options varying from the most basic tasks of sequence alignment to the highly complex tasks of designing a complete model of a protein. The server also assists the user in building protein models at different levels of intricacy (Kiefer *et al.* 2009; Schwede *et al.* 2003).

2.6.3 ESYPRED3D:

ESyPred3D is also an automated homology modeling program. ESyPred3D gets benefits of increased alignment performance which is based on neural networks. It normalizes the alignment score by combining the alignment results of various programs. After that MODELLER is used to design the final model of the protein in question. Once it gets a model designed, the model is then assessed using PROCHECK package (Lambert *et al.* 2002).

2.6.4 MODBASE:

MODBASE is a relational database that contains annotated comparative protein structure models for all the proteins available. The models are derived by ModPipe, which is a programmed modeling pipeline relying on the packages PSI-BLAST and MODELLER. The database also includes fold assignments and

alignments and the newly built models are based on them. MODBASE contains theoretically calculated models, which may contain significant errors, not experimentally determined structures. The database also contains necessary information about ligand binding sites in a protein. It also provides an overview of protein-protein interactions (Pieper *et al.* 2006).

2.7 Evaluation:

The evaluation phase is a mandatory phase in almost all projects. It is in fact a review checkpoint for the project. As this project includes the modeling of various proteins of the BBSome complex, so the evaluation phase involves checking for the possible errors in the newly designed models, their energy values as well as the favored and non-favored regions within the structure.

2.7.1 PROSA:

The major problem in structural biology is to pick out the errors and slip-ups in the models of protein especially those designed through homology modeling.

ProSA is a web based protein evaluation tool that helps to evaluate the protein by calculating the z-score of the protein, plotting the scores of the residues and their energies as well. It is mostly used in the refinement and validation of experimental protein models (Wiederstein and Sippl 2007).

Plot of residue scores:

ProSA provides the user with a plot depicting the energies of the protein structure with detailed information about each and every residue. High negative values in a plot correspond to favored regions of the input structure and vice versa.

2.7.2 RAMACHANDRAN PLOT 2.0:

The Ramachandran plot shows the phi-psi torsion angles for all residues in a protein structure. It depicts the likely conformations of the φ and ψ angles for a protein ensemble. The server which I have used to figure the Ramachandran plots of the newly modeled protein structures is Ramachandran plot 2.0. It is a web based server that has an assortment of improved options for displaying the torsion angles in various regions. It is a program for visualizing and assessing the Ramachandran plot of a protein structure. It provides a complete overview of the composition of the protein. The input to the server is the PDB file and the output contains detail of residues in the favored regions, allowed regions and outlier regions (Gopalakrishnan *et al.* 2007).

2.8 Hydrophobicity Plots:

Hydrophobicity Plots are mostly used in analyzing the membrane spanning regions which are highly hydrophobic and potential antigenic sites on a protein that are likely to be exposed. The algorithm used in this case is Hopp & Woods hydrophobicity scale which was developed to pick out the antigenic sites of proteins through the polarity and charge of residues present in a protein (Hoop and Woods 1981).

2.9 Protein Interactions:

Protein interaction studies enjoy an imperative position among protein structural and conformational studies. Protein interactions are considered to be a hub of information for the investigation of participation of proteins in various biological pathways. Moreover, protein-protein interactions are also important in mapping the drug interactions, disease studies and various other biochemical processes.

2.9.1 STRING:

STRING is a database of protein interactions which include both the known as well as the predicted interactions. The interactions mapped out by this database rely mostly on the experimental data available, co-expression of proteins and genomic context. The input to the database can either be the amino acid sequence of the protein in question or just the name of the gene/ protein. It then prompts the user about the source of the protein (organism details). The final output is in the form of a connected graph of the query protein to various other known proteins and ligands. The final output can be restricted according to user's choice by checking the display variables (Jensen *et al.* 2009).

2.9.2 GRAMM-X:

The Global Range Molecular Matching (GRAMM) methodology is an experimental approach to smoothing the energy function (intermolecular) by varying the range of the atom-atom potentials. To predict the structure of a complex, it needs only the atomic coordinates of the two molecules. No information about the binding sites is needed. The software then performs an exhaustive 6 dimensional search for relative positions and rotations of molecules. An important aspect of GRAMM software is its ability to smooth the protein surface representation to relate possible conformational changes upon binding within the rigid body docking technique (Tovchigrechko and Vasker 2006).

2.9.3 PYMOL:

PyMOL is protein visualization software as well as a molecular modeling program to facilitate the understanding of the compositional details of a protein

(DeLano 2002). It can also be used to visualize the interactions between docked complexes of proteins. Structure comparison can also be done by using this biological toolkit. With the help of PyMOL, surfaces and electrostatic potentials, polar interactions between residues of one molecule or between multiple molecules can also be found out. The pseudo code used for finding hydrogen bonds between the docked complexes is given below:

» zoom center, 50

» select ChA, chain A

» select ChB, chain B

» dist name, ChA, ChB, mode=2

2.9.4 Hex:

Hex is an interactive molecular visualization tool that can be used to display as well as calculate feasible modes of docking between two molecules. SPF (Spherical Polar Fourier) correlations and FFTs (Fast Fourier Transform) are employed by this software package to calculate the docking correlations.

2.10 Objectives and Aims of Study:

The current study is designed to determine the 3D structure of proteins involved in BBSome complex as none of them are already available. Moreover the interactions of the 9 BBSome genes and proteins will be investigated and the behavior of interacting proteins with respect to secondary and tertiary structure distortion will also be an important consideration of this project.

Chapter 3

Results

3. RESULTS

3.1 Similarity Search:

3.1.1 BLAST:

BLAST is the most commonly employed algorithm for similarity search. It follows a heuristic based approach to find out alignments for unknown proteins against a huge number of database sequences. BLAST results for known BBS proteins are as under:

Table 3.1. BLAST results for known proteins involved in BBS

Gene Name	Acc. No.	Protein Name	Max Score	E-value
BBS1	2BTZ_A	Chain A, Crystal Structures Of Human Pyruvate Dehydrogenase Kinase 2 Containing Physiological And Synthetic Ligands	30.8	1.6
	2ZKJ_A	Chain A, Crystal Structure Of Human Pdk4-Adp Complex	30.8	1.9
BBS2	1C8B_A	Chain A, Crystal Structure Of A Novel Germination Protease From Spores Of Bacillus Megaterium: Structural Rearrangements And Zymogen Activation	33.1	0.45
	1V7V_A	Chain A, Crystal Structure Of Vibrio Proteolyticus Chitobiose Phosphorylase	30.4	3.1
BBS4	2FO7_A	Chain A, Crystal Structure Of An 8 Repeat Consensus Tpr Superhelix (Trigonal Crystal Form)	80.1	2e-15
	3KD7_A	Chain A, Designed Tpr Module (Ctpr390) In Complex With Its Peptide- Ligand (Hsp90 Peptide)	70.1	2e-12

BBS5	2WGQ_A	Chain A, Zinc Substituted E Coli Copper Amine Oxidase, A Model For The Precursor For 2,4,5-Trihydroxyphenylalaninequinone Formation	30	1.7
	1OAC_A	Chain A, Crystal Structure Of A Quinoenzyme: Copper Amine Oxidase Of Escherichia Coli At 2 Angstroems Resolution	30	1.7
BBS7	3L3P_A	Chain A, Crystal Structure Of The C-Terminal Domain Of Shigella Type Iii Effector Ipah9.8, With A Novel Domain Swap	30.8	2.2
	2UV8_A	Chain A, Crystal Structure Of Yeast Fatty Acid Synthase With Stalled Acyl Carrier Protein At 3.1 Angstrom Resolution	30	4.0
BBS8	2FO7_A	Chain A, Crystal Structure Of An 8 Repeat Consensus Tpr Superhelix (Trigonal Crystal Form)	57.8	1e-08
	1W3B_A	Chain A, The Superhelical Tpr Domain Of O-Linked Glcnac Transferase Reveals Structural Similarities To Importin Alpha	56.6	2e-08
BBS10	3IYG_E	Chain E, Ca Model Of Bovine TricCCT DERIVED FROM A 4.0 ANGSTROM Cryo-Em Map	60.1	3e-09
	1Q3R_A	Chain A, Crystal Structure Of The Chaperonin From Thermococcus Strain Ks-1 (Nucleotide-Free Form Of Single Mutant	59.7	5e-09

3.2 Post Translational Modifications:

3.3.1 SignalP:

SignalP is a web based service to pick out the post translational modifications of a protein. It gives a prediction signal peptide cleavage sites in a query protein. Currently SignalP works in form of two basic modules, Neural Network and Hidden Markov Model. The output of the NN module is in the form of C-Score, S-Score and Y-Score while the output of other module is in the form of posterior probabilities for the cleavage site and signal peptides at each position.

SignalP outputs for various proteins of the BBSome complex are given in table 3.2 while NN plots for these proteins can be found in figure 3.1 to figure 3.7.

Table 3.2. SignalP outputs for known proteins of BBS

Gene Name	Measure	Position	Value	Cutoff	Signal Peptide?
BBS1	Max. C	22	0.062	0.32	No
	Max. Y	45	0.027	0.33	No
	Max. S	1	0.091	0.87	No
	mean S	1-44	0.024	0.48	No
	D	1-44	0.025	0.43	No
BBS2	Max. C	23	0.096	0.32	No
	Max. Y	23	0.143	0.33	No
	Max. S	12	0.710	0.87	No
	mean S	1-22	0.288	0.48	No
	D	1-22	0.216	0.43	No
BBS4	Max. C	19	0.199	0.32	No
	Max. Y	3	0.035	0.33	No
	Max. S	1	0.174	0.87	No
	mean S	1-2	0.159	0.48	No
	D	1-2	0.097	0.43	No
BBS5	Max. C	20	0.065	0.32	No
	Max. Y	20	0.033	0.33	No
	Max. S	4	0.099	0.87	No
	mean S	1-19	0.047	0.48	No

	D	1-19	0.040	0.43	No
BBS7	Max. C	18	0.248	0.32	No
	Max. Y	18	0.073	0.33	No
	Max. S	3	0.168	0.87	No
	mean S	1-17	0.066	0.48	No
	D	1-17	0.070	0.43	No
BBS8	Max. C	24	0.069	0.32	No
	Max. Y	24	0.120	0.33	No
	Max. S	14	0.403	0.87	No
	mean S	1-23	0.254	0.48	No
	D	1-23	0.187	0.43	No
BBS10	Max. C	24	0.096	0.32	No
	Max. Y	24	0.167	0.33	No
	Max. S	12	0.842	0.87	No
	mean S	1-23	0.410	0.48	No
	D	1-23	0.289	0.43	No

Visual plots for all these known proteins of the BBSome complex are given below:

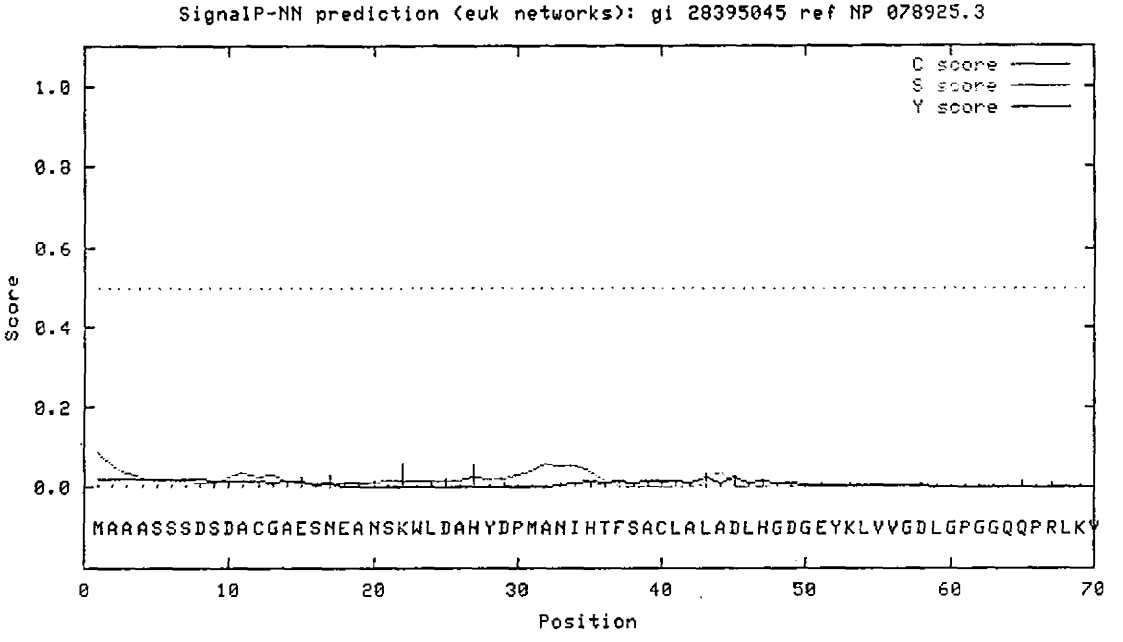


Figure 3.1. SignalP-NN result for BBS1

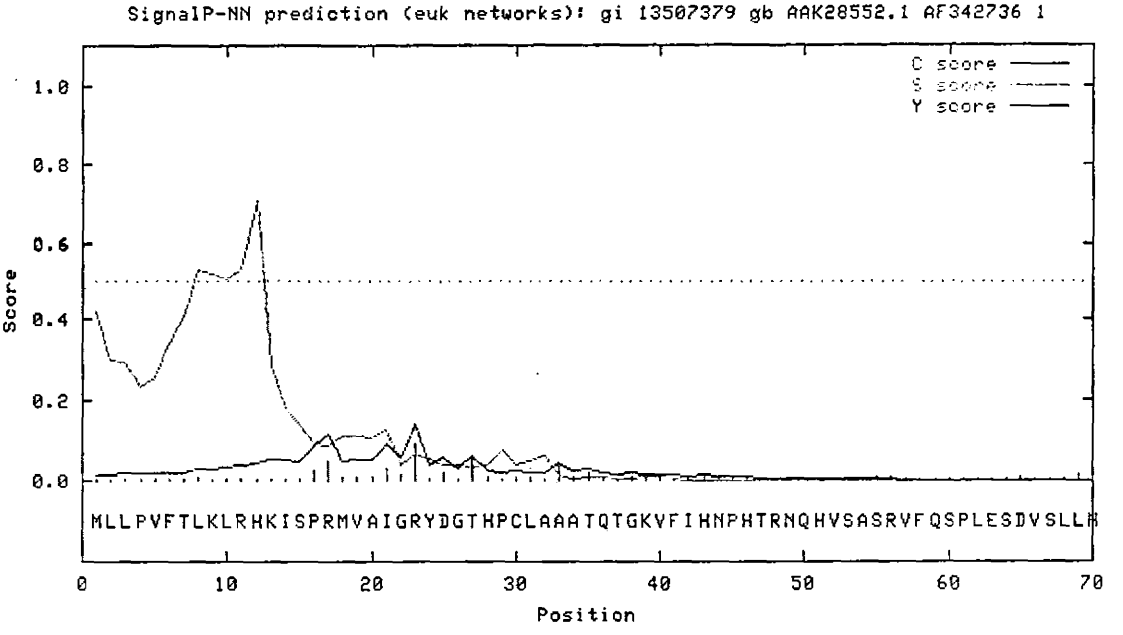


Figure 3.2. SignalP-NN result for BBS2

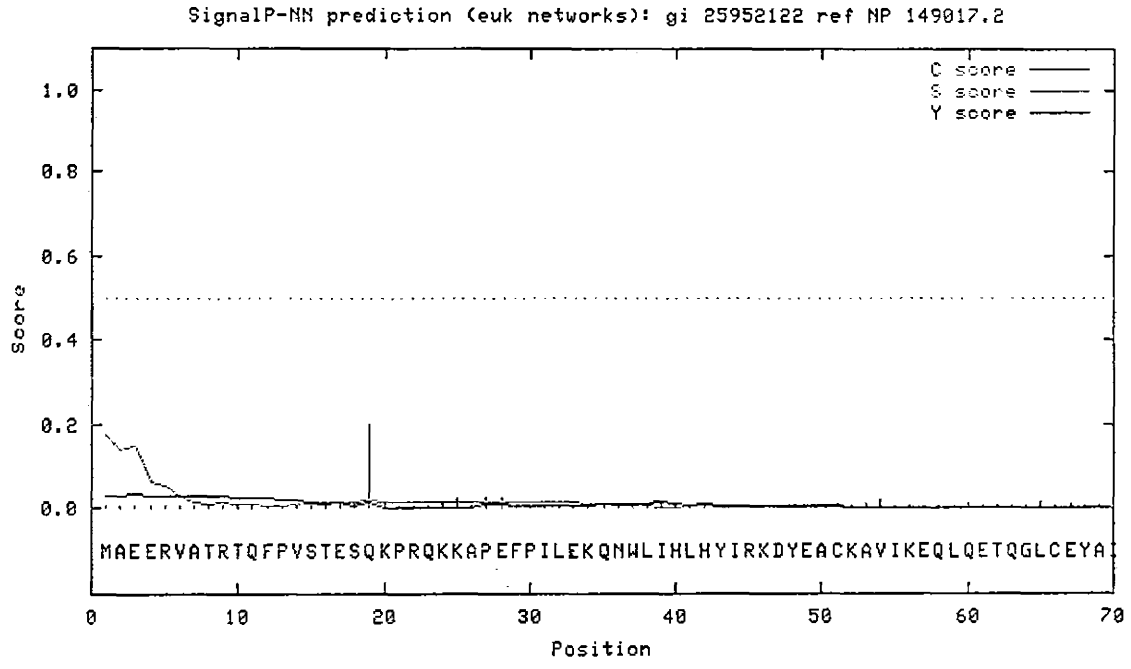


Figure 3.3. SignalP-NN result for BBS4

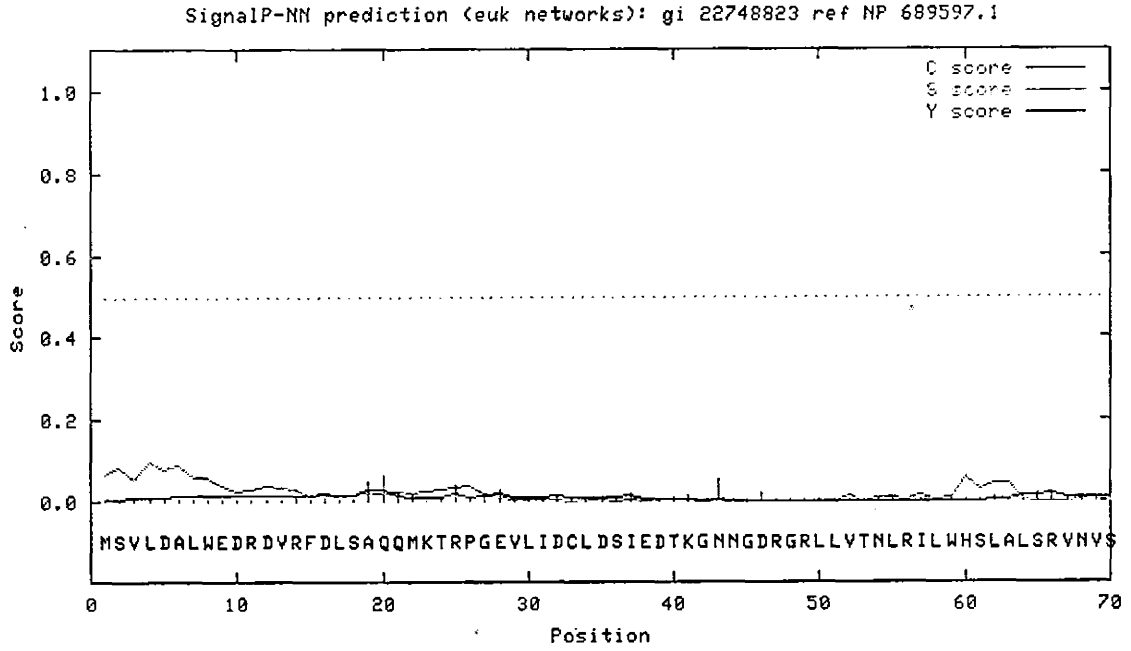


Figure 3.4. SignalP-NN result for BBS5

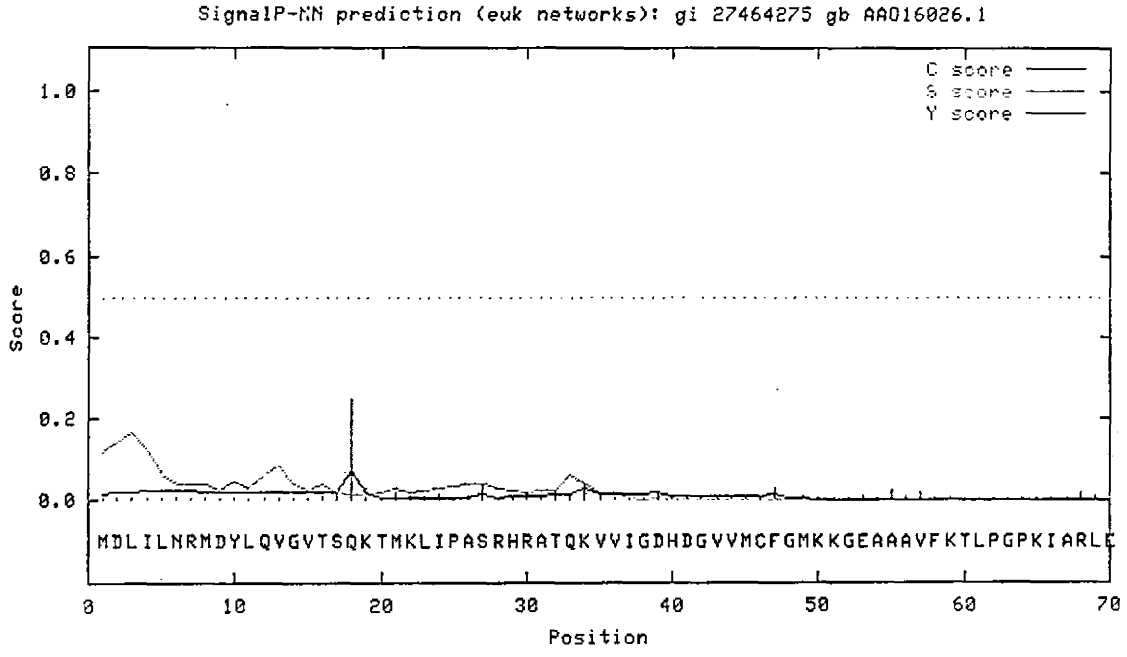


Figure 3.5. SignalP-NN result for BBS7

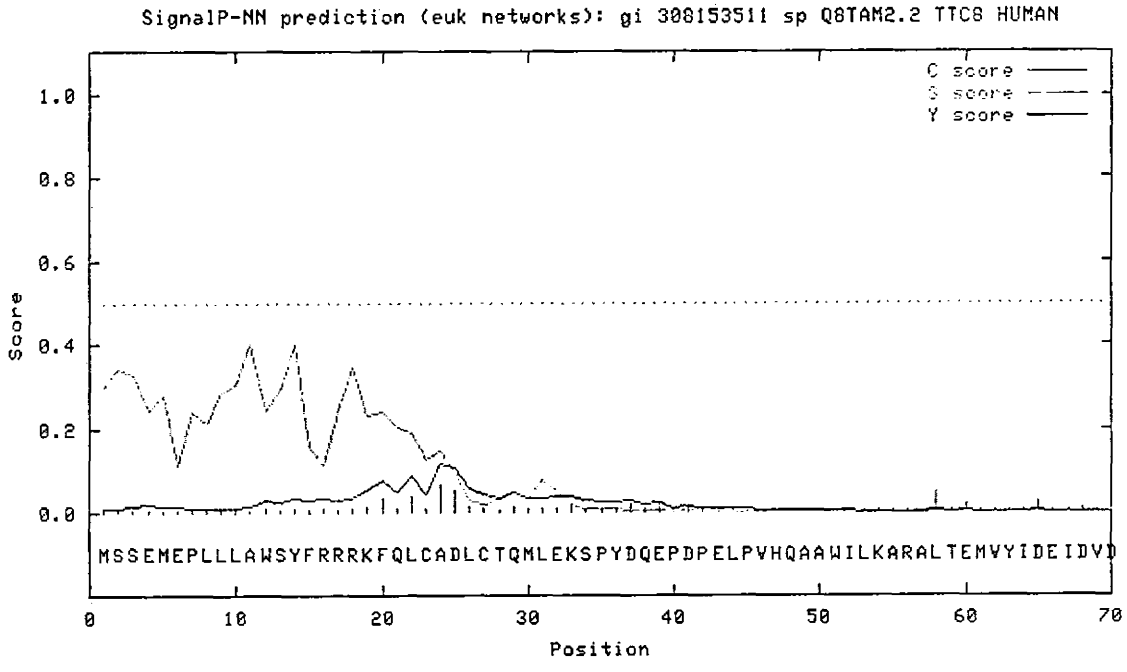


Figure 3.6. SignalP-NN result for BBS8

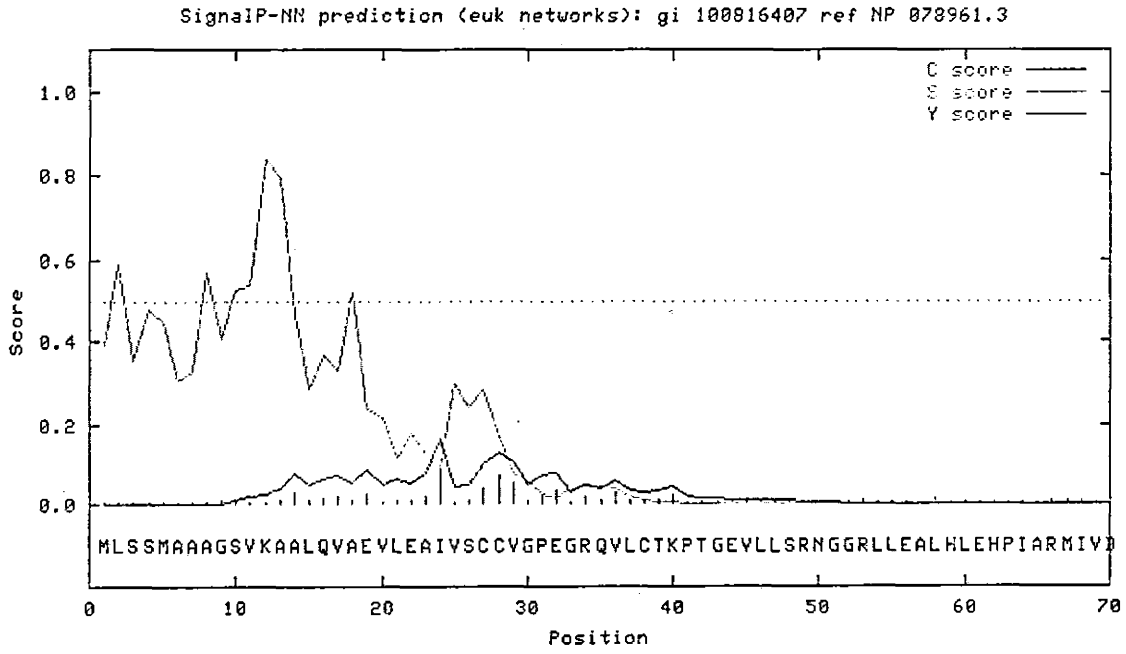


Figure 3.7. SignalP-NN result for BBS10

3.3 Topology Prediction:

Topology prediction in this project is done through TMpred.

3.3.1 TMPRED:

TMpred results of various genes of the BBSome complex are as under:

BBS1:

Table 3.3. Transmembrane models of BBS1

STRONGLY Preferred Model:	**From**	**To**	**Length**	**Score**	**Orientation**
N-terminus outside	507	531	(25)	1085	o-i
Total Score				1085	
Alternative Model	509	530	(22)	550	o-i
Total Score				550	

BBS2:

Table 3.4. Transmembrane models of BBS2

STRONGLY Preferred Model:	**From**	**To**	**Length**	**Score**	**Orientation**
N-terminus inside	68	89	(22)	705	i-o
	123	147	(25)	843	o-i
Total Score				1548	
Alternative Model	123	147	(25)	843	o-i
Total Score				843	

BBS4:

Probably no transmembrane protein - no possible model found.

BBS5:

Table 3.5. Transmembrane models of BBS5

STRONGLY Preferred Model:	From	To	Length	Score	Orientation
N-terminus outside	57	77	(21)	694	o-i
Total Score				694	
Alternative Model	50	73	(24)	538	i-o
Total Score				538	

BBS7:

Table 3.6. Transmembrane models of BBS7

STRONGLY Preferred Model:	From	To	Length	Score	Orientation
N-terminus inside	113	132	(20)	517	i-o
Total Score				517	
Alternative Model	No strong transmembrane helices predicted				

BBS8:

Table 3.7. Transmembrane models of BBS8

STRONGLY Preferred Model:	**From**	**To**	**Length**	**Score**	**Orientation**
N-terminus inside	378	402	(25)	782	i-o
Total Score				782	
Alternative Model	No strong transmembrane helices predicted				

BBS10:

Table 3.8. Transmembrane models of BBS10

STRONGLY Preferred Model:	**From**	**To**	**Length**	**Score**	**Orientation**
N-terminus outside	384	403	(20)	1039	o-i
	589	611	(23)	552	i-o
	626	647	(22)	915	o-i
Total Score				2506	
Alternative Model	382	401	(20)	1368	i-o
	589	615	(27)	1103	o-i
Total Score				2471	

3.4 Primary Structure:

3.4.1 ProtParam:

ProtParam computes various physio-chemical properties of a protein including its molecular weight, the amino acid make up of the protein in question, atomic composition and estimated half life of the protein.

Table 3.9. ProtParam results for the known proteins of BBSome Complex

Name Of Gene	Amino Acids	Molecular Weight	Theoretical PI	Formula	Instability Index
BBS1	593	65083.4	8.02	$C_{2911}H_{4688}N_{794}O_{846}S_{23}$	46.18
BBS2	721	79870.6	5.74	$C_{3520}H_{5580}N_{988}O_{1079}S_{27}$	30.07
BBS4	519	588281.9	6.90	$C_{2625}H_{4125}N_{689}O_{769}S_{20}$	47.07
BBS5	341	38755.0	5.39	$C_{1728}H_{2721}N_{469}O_{523}S_{10}$	49.61
BBS7	715	80337.7	5.70	$C_{3583}H_{5682}N_{952}O_{1090}S_{25}$	37.23
BBS8	541	61534.2	6.33	$C_{2750}H_{4272}N_{752}O_{804}S_{25}$	39.03
BBS10	723	80837.5	7.95	$C_{3598}H_{5740}N_{966}O_{1060}S_{42}$	38.98

3.5 Secondary Structure:

3.5.1 PREDICT PROTEIN:

The Predict Protein Server helps analyzing the in depth study of the secondary structure of a protein on the basis of primary structure (raw sequence) provided.

Table 3.10. Results obtained for Known BBS proteins through Predict Protein

GENE NAME	SECONDARY STRUCTURE (PREDICTED)	
BBS1	**SEQUENCE LENGTH**	593
	Secondary structure	Helix=17.37%, Strand=36.76%, Loop=45.87%
	window size	70
	Structure content cutoff	12%
	Residues exposed	47.72
	Confidence of disulfide bonding state prediction	7, 8, 9 (High)
	NORS region predicted	None
	Low Complexity Regions	Present
	SEQUENCE LENGTH	721
	Secondary structure	Helix=23.99%, Strand=35.64%, Loop=40.36%
	window size	70

BBS2	**Structure content cutoff**	12%
	Residues exposed	48.96
	Confidence of disulfide bonding state prediction	8, 9 (High)
	NORS region predicted	None
	Low Complexity Regions	Present
BBS4	**SEQUENCE LENGTH**	519
	Secondary structure	Helix=79.38%, Strand=0%, Loop=20.62%
	window size	70
	Structure content cutoff	12%
	Residues exposed	50.87
	Confidence of disulfide bonding state prediction	8, 9 (High)
	NORS region predicted	None
	Low Complexity Regions	Present
BBS5	**SEQUENCE LENGTH**	341
	Secondary structure	Helix=21.99%, Strand=37.54%, Loop=40.47%
	window size	70
	Structure content cutoff	12%
	Residues exposed	51.03

	Confidence of disulfide bonding state prediction	8 (High)
	NORS region predicted	None
	Low Complexity Regions	Absent
BBS7	**SEQUENCE LENGTH**	715
	Secondary structure	Helix=22.8%, Strand=40.56%, Loop=36.64%
	window size	70
	Structure content cutoff	12%
	Residues exposed	47.13
	Confidence of disulfide bonding state prediction	8, 9 (High)
	NORS region predicted	None
	Low Complexity Regions	Present
BBS8	**SEQUENCE LENGTH**	541
	Secondary structure	Helix=68.0%, Strand=0.0%, Loop=32.0%
	window size	70
	Structure content cutoff	12%
	Residues exposed	52.68
	Predicted disulfide bonding state	No disulfide bonds

	Confidence of disulfide bonding state prediction	8, 9 (High)
	NORS region predicted	59-134
	Low Complexity Regions	Present
BBS10	**SEQUENCE LENGTH**	723
	Secondary structure	Helix=49.8%, Strand=11.2%, Loop=39.0%
	window size	70
	Structure content cutoff	12%
	Residues exposed	49.65
	Predicted disulfide bonding state	No disulfide bonds
	NORS region predicted	None
	Confidence of disulfide bonding state prediction	9 (High)
	Low Complexity Regions	Present

3.5.2 Scratch Protein Predictor:

Scratch Protein Predictor gives the information for secondary structure, single mutation stability, disordered regions, domains, relative solvent accessibility, disulfide bridges and details of residue contacts for the protein. The following table shows the secondary structure predictions for proteins in the BBSome complex.

Table 3.11. Secondary structure elements for proteins of BBSome Complex by SPP

Name Of Gene	Predicted Domains	Total Cys	Predicted Disulfide Bonds	Cys1	Cys2	Probability of Antigenicity
BBS1	Domain1: 1 - 103 Domain2: 104 - 449 Domain3: 450 - 593	8	3	12 267 377	40 285 520	0.449276
BBS2	Domain1: 1 - 344 Domain2: 345 - 672 Domain3: 673 - 721	11	4	662 30 142 442	698 77 169 534	0.756270
BBS4	Domain1: 1 - 461 Domain2: 462 - 519	Nil	Nil	Nil	Nil	0.735580
BBS5	Domain1: 1 - 131 Domain2: 132 - 341	3	1	33	75	0.638001
BBS7	Domain1: 1 - 354 Domain2: 355 - 451 Domain3: 452 - 715	13	5`	527 432 243 133 135	540 447 296 152 167	0.764474

BBS8	Domain 1: 1 - 93 Domain 2: 94 - 541	8	3	23 396 357	27 445 395	0.631723
BBS10	Domain1: 1 - 482 Domain2: 483 - 577 Domain3: 578 - 723	22	8	390 578 360 110 201 621 166 27	400 598 371 123 204 694 179 28	0.755688

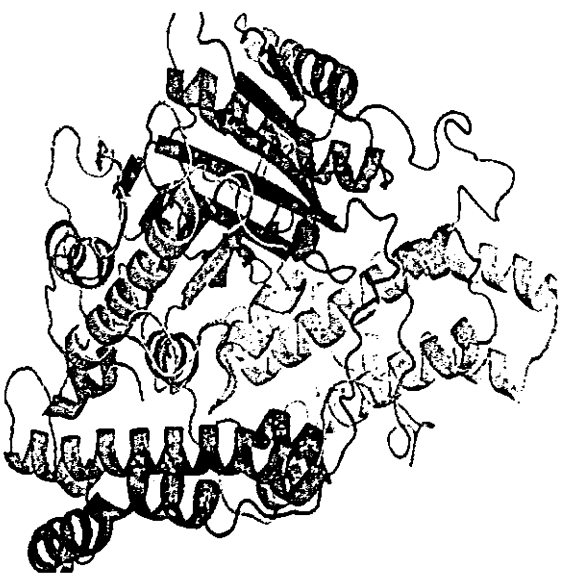
3.6 Tertiary Structure:

3.6.1 MODELLER:

MODELLER is used to design models for known BBS genes. The details of the models designed against each protein of BBSome complex are shown in table 3.12. the visual details of these models are also shown in figure 3.

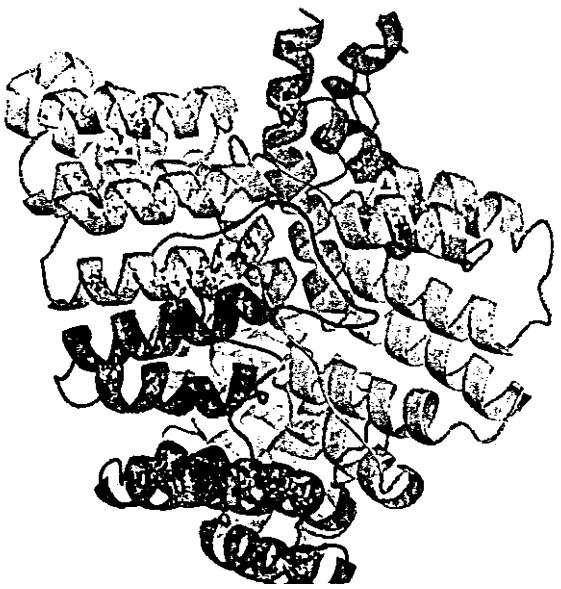
Table 3.12. Model details for known BBS proteins

Name of gene	Model residue range	No of residues modeled	MODELLER Objective Function	Based on template
BBS1	1-593	593	4998.3853	2ZKJ_A
BBS2	1-721	721	5388.9409	1C8B_A
BBS4	1-519	519	15220.1328	3KD7_A
BBS5	1-341	341	2509.8127	2WGQ_A
BBS7	1-715	715	6055.8643	3L3P_A
BBS8	1-541	541	3119.1487	1W3B_A
BBS10	1-723	723	4598.3145	1Q3R_A



a)

b)



c)

d)

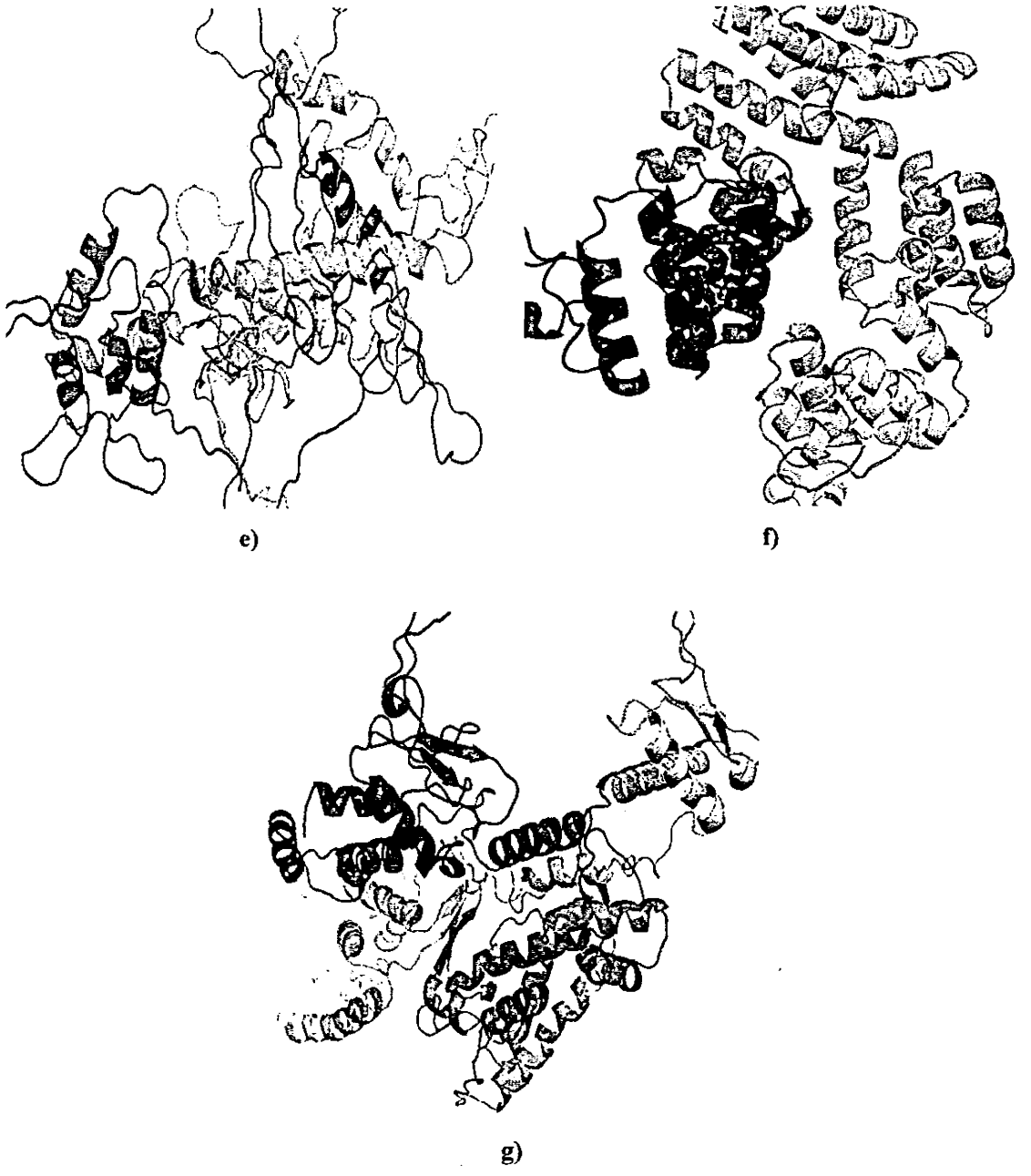


Fig. 3.8. Cartoon Representation of proteins of the BBSome complex designed by Modeller, Viewed in PyMol (a) BBS1 (b) BBS2 (c) BBS4 (d) BBS5 (e) BBS7 (f) BBS8 (g) BBS10

SWISS MODEL:

Swiss Model workspace is an easy-to-use web based modelling workbench. It facilitates the user to evaluate and build protein models. The Swiss Model results for the proteins of BBSome complex are given in table 3.13.

Table 3.13. Swiss Model results for known genes of BBSome Complex

NAME OF GENE	RESIDUE RANGE	BASED ON TEMPLATE	E-VALUE	QMEAN Z-SCORE:
BBS1	218-399	1tbgA	7.90e-5	-5.474
BBS2	72-107	3iiyA	6.60e-5	-1.69
BBS4	101-234	2fo7A	7.30e-16	-1.801
BBS7	205-271	1erjC	7.20e-5	-1.103
BBS8	426 to 523	2fo7A	2.90e-12	-1.56
BBS10	232 to 401	1gmIC	1.60e-26	0

3.6.2 ESYPRED3D:

Esypred3D is an online service for protein tertiary structure prediction. It normalizes the alignment score by combining the alignment results of various programs.The details of models designed by Esypred3D are given in table 3.14.

Table 3.14. Esypred3D output for known genes of BBSome Complex

NAME OF GENE	MODELLED REGION	SEQUENCE ALIGNMENT		Z-SCORE	TARGET PROTEIN LENGTH
		QUERY	MODEL		
BBS1	85-541	1-355	85-541	-1.52	593
BBS2	10-712	1-320	10-712	25.5	721
BBS4	251-617	1-102	251-617	14.03	519
BBS5	1-341	1-720	1-341	-0.57	341
BBS7	14-712	1-263	14-712	13.89	715
BBS8	1-418	1-388	1-418	-6.58	541
BBS10	1-549	1-518	1-549	-6.73	723

3.6.3 MODBASE:

Modbase is an online service for efficient calculation of tertiary conformations of a protein. Moreover it also allows the user to assess the models of a protein. The models designed by Modbase may cover only a segment or the entire target sequence.

The details of models calculated through this service are presented in table 3.15.

Table 3.15. Modbase output for genes of BBSome Complex

NAME OF GENE	MODELLED REGION	TEMPLATE	E-VALUE	TARGET PROTEIN LENGTH
BBS1	198-368	3emhA	0.0069	593
BBS2	5-321	3odtA	0	721
	35-312	1flgA	0.0019	721
BBS4	23-514	2xpiA	4.8e-11	519
	175-223	1elwA	0.027	519
BBS5	148-250	2cayA	0	341
	159-219	2c5aA	0.81	341
BBS7	573-636	1gvnA	0.58	715
BBS8	264-489	2xpiA	0.0087	541
BBS10	3-708	1q3qA	0	723

3.7 Model Evaluation:

The evaluation phase is a mandatory phase in almost all projects. It is in fact a review checkpoint for the project. As this project includes the modeling of various proteins of the BBSome complex, so the evaluation phase involves checking for the possible errors in the newly designed models, their energy values as well as the favored and non-favored regions within the structure.

The evaluation phase involves reviewing all the save points in the course of a project. The major save points in the model evaluation are:

- Z-Score evaluation
- Ramachandran plot evaluation

These steps are performed through various available servers.

3.7.1 Z-Score Evaluation:

Swiss-Model:

Table 3.16. Z-Score details of BBS proteins through Swiss-Model

Model	Z-Score	
	Template	Model
BBS1_2ZKJ	0.169	-8.14
BBS2_1C8B	-3.827	-7.055
BBS4_3KD7	-0.664	-5.019
BBS5_2WGQ	2.032	-8.996
BBS7_3L3P	-1.071	-7.427
BBS8_1W3B	-0.331	-4.554
BBS10_1Q3R	-0.323	-6.01

ProSA:

Table 3.17. Z-Score details of BBS proteins through ProSA

Model	Z-Score	
	Template	Model
BBS1_2ZKJ	-8.52	-2.05
BBS2_1C8B	-3.74	-0.4
BBS4_3KD7	-5.96	-0.47
BBS5_2WGQ	-11.2	0.66
BBS7_3L3P	-6.87	0.86
BBS8_1W3B	-8.8	-6.94
BBS10_1Q3R	-9.15	-4.88

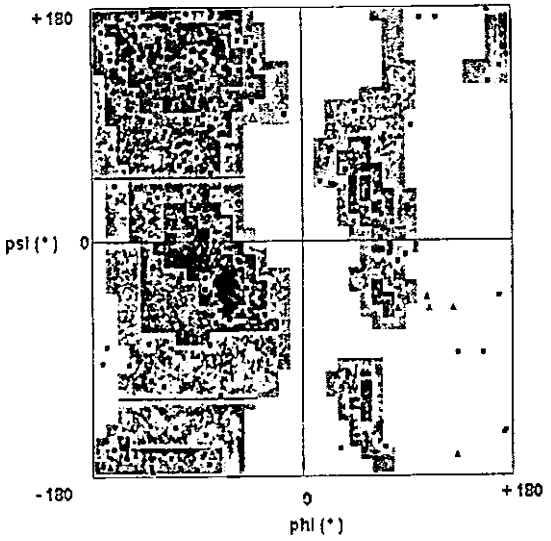
3.7.2 RAMACHANDRAN PLOT 2.0:

The server which I have used to figure the Ramachandran plots of the newly modeled protein structures is Ramachandran plot 2.0. It is a web based server that has an assortment of improved options for displaying the torsion angles in various regions. The residue details provided by Ramachandran plots for the proteins of BBSome complex are given in table 3.18.

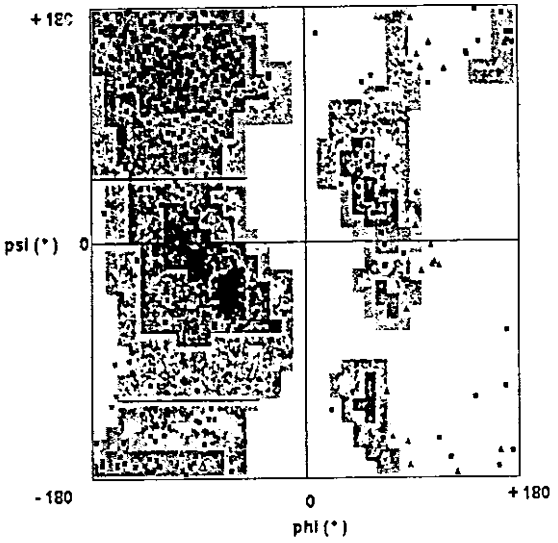
Table 3.18. Residue details of BBS proteins as obtained through Ramachandran Plot 2.0

Model	Fully Allowed Regions	Additionally Allowed Regions	Generously Allowed Regions	Outside Region
BBS1_2ZKJ	81.05%	12.01%	4.23%	2.71%
BBS2_1C8B	74.27%	16.27%	5.01%	4.45%
BBS4_3KD7	79.11%	12.19%	5.03%	3.68%
BBS5_2WGQ	77.88%	15.63%	4.13%	2. 36%
BBS7_3L3P	69.57%	18.93%	8.13%	3.37%
BBS8_1W3B	84.60%	8.91%	4.08%	2.41%
BBS10_1Q3R	80.44%	12.48%	4.30%	2.77%

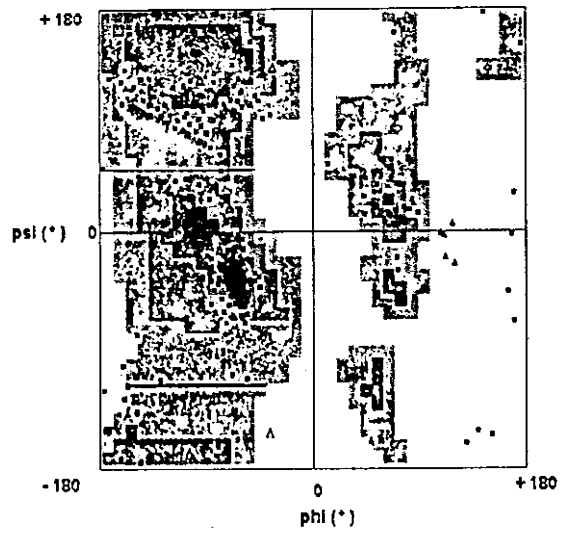
The Ramachandran plots obtained are as under:



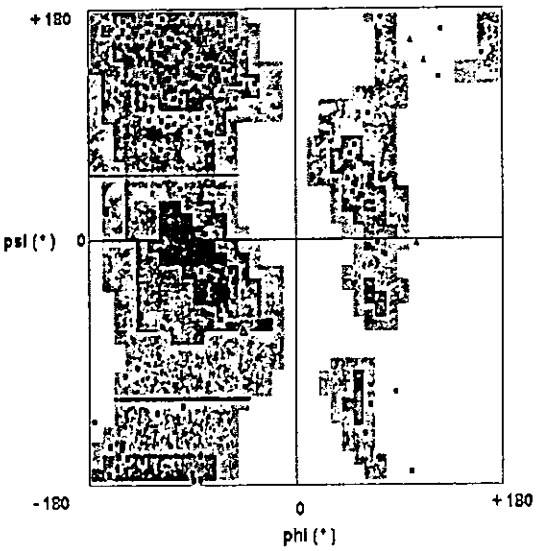
a)



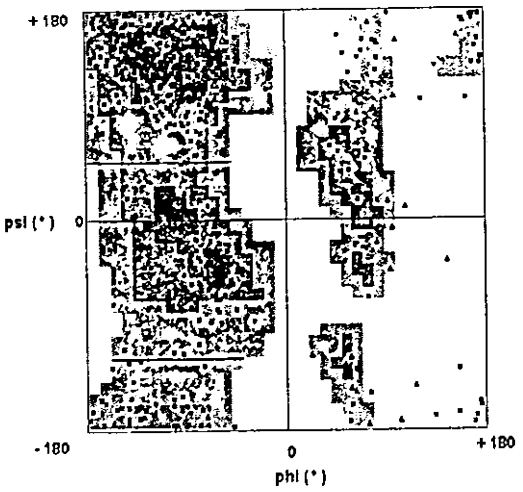
b)



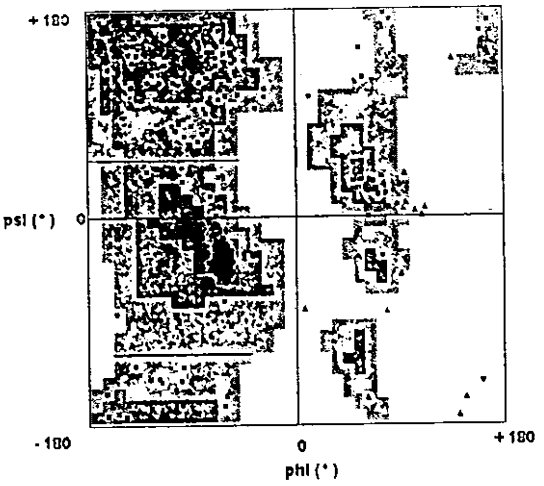
c)



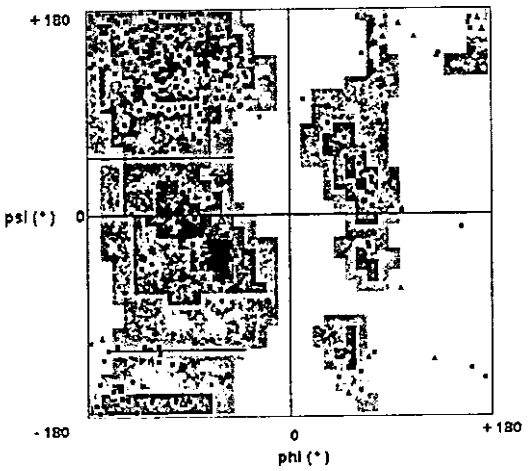
d)



e)



f)

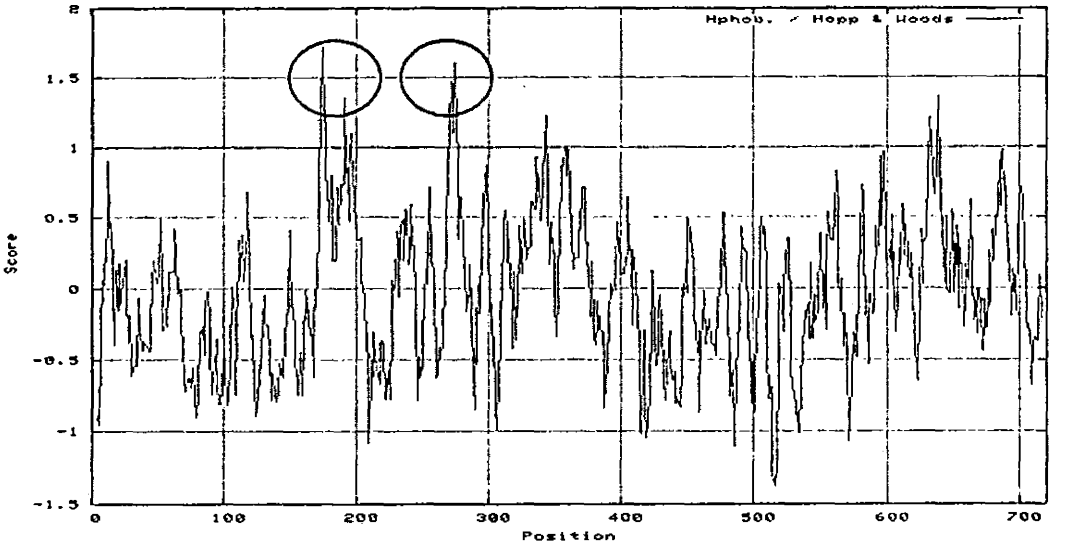


g)

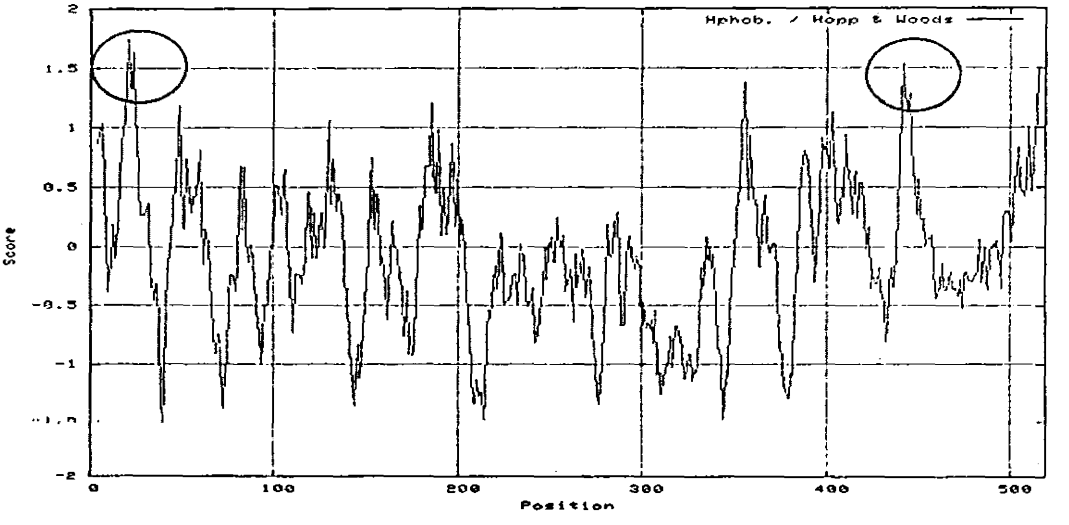
Fig. 3.9. Ramachandran plots of (a) BBS1 (b) BBS2 (c) BBS4 (d) BBS5 (e) BBS7 (f) BBS8 (g) BBS10

3.8 Hydrophobicity Plots:

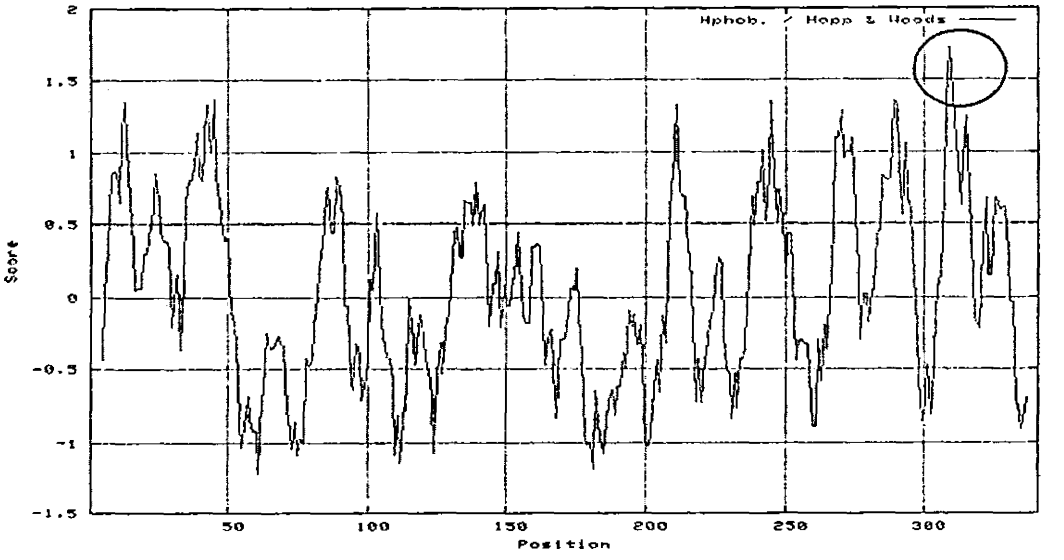
Hopp & Woods Hydrophobicity scale was used to get an insight into the antigenic sites and surface details of the proteins of the BBSome complex. Preditop package was used for antigenic prediction. BBS2, BBS4 and BBS5 showed some higher peak values in residue plots corresponding to potential antigenic sites of this complex.



(a)



(b)



(c)

Fig. 3.10. Hydrophobicity plots of the proteins of BBSome complex using Preditop package showing antigenic sites in (a) BBS2 (b) BBS4 (c) BBS5

3.9 Protein Interactions:

Protein Interactions involved in Bardet-Biedl Syndrome are obtained through STRING database.

3.8.1 STRING:

Interactions obtained through STRING are based on experimental data, text mining, co-expression of proteins and genomic context. The interactions of the proteins involved in BBSome complex obtained through STRING are given in figure 3.11 to figure 3.17.

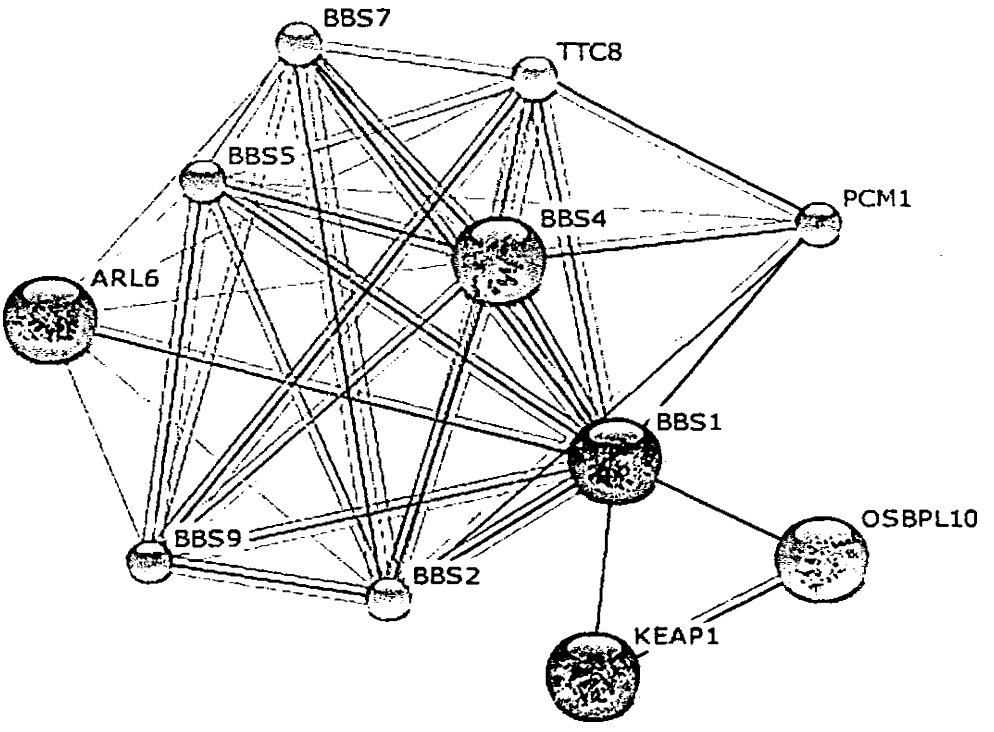


Figure 3.11. Interactions obtained through STRING for BBS1

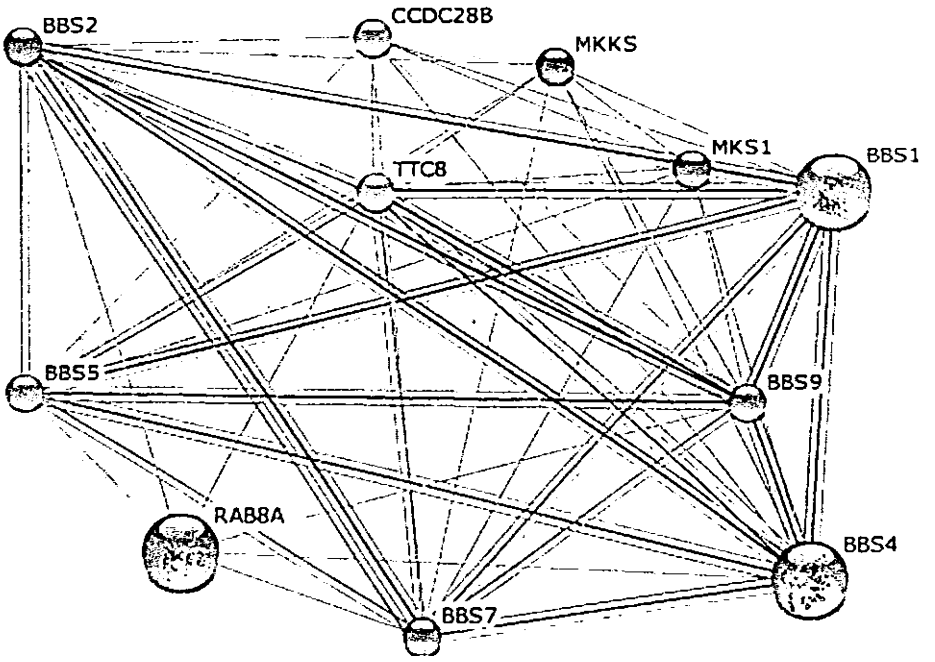


Figure 3.12. Interactions obtained through STRING for BBS2

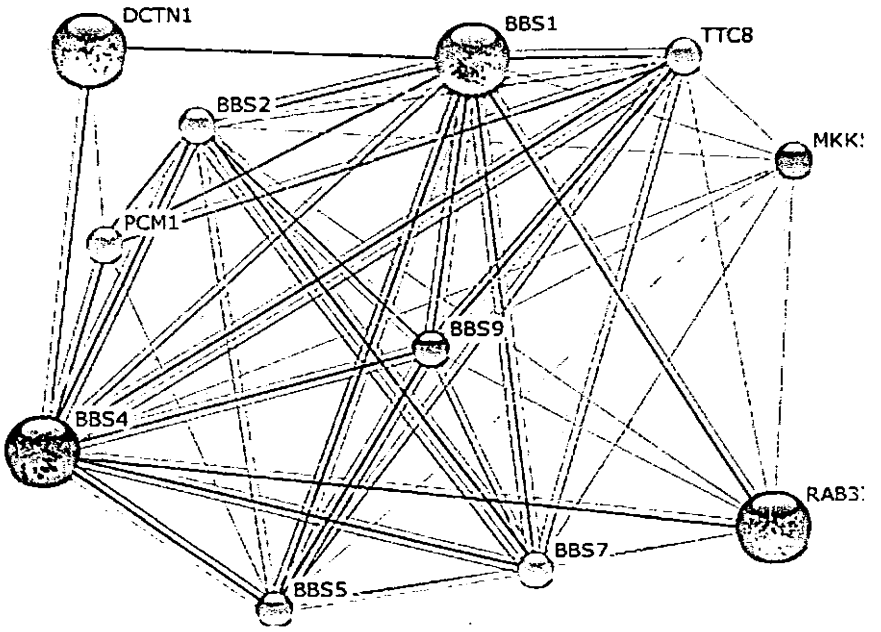


Figure 3.13. Interactions obtained through STRING for BBS4

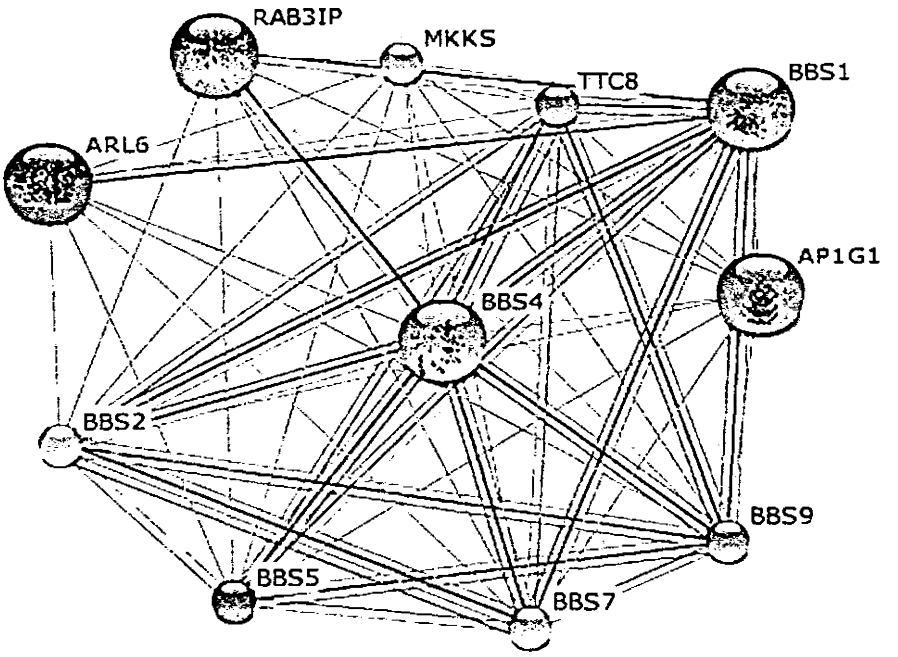


Figure 3.14. Interactions obtained through STRING for BBS5

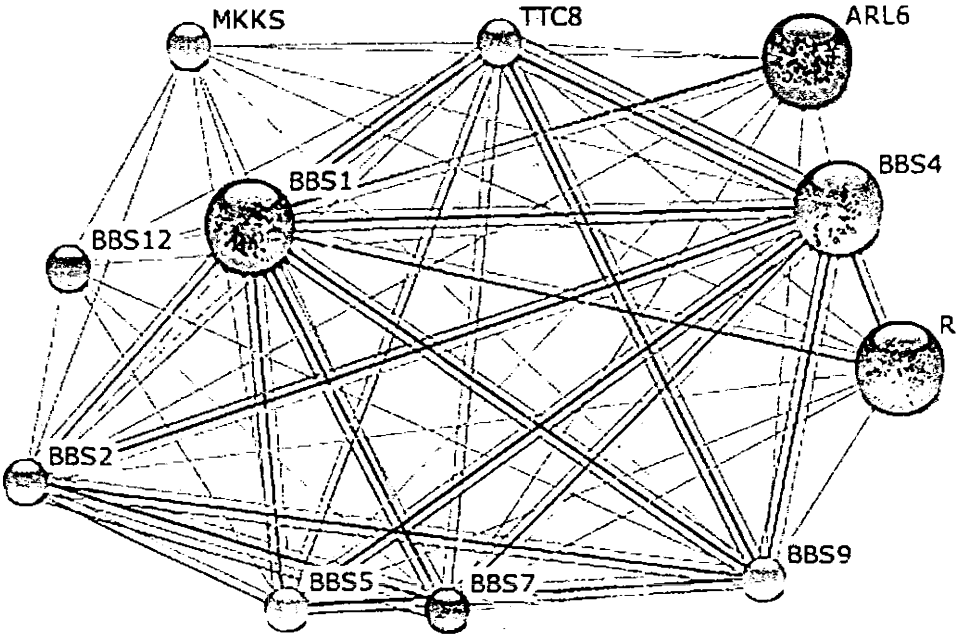


Figure 3.15. Interactions obtained through STRING for BBS7

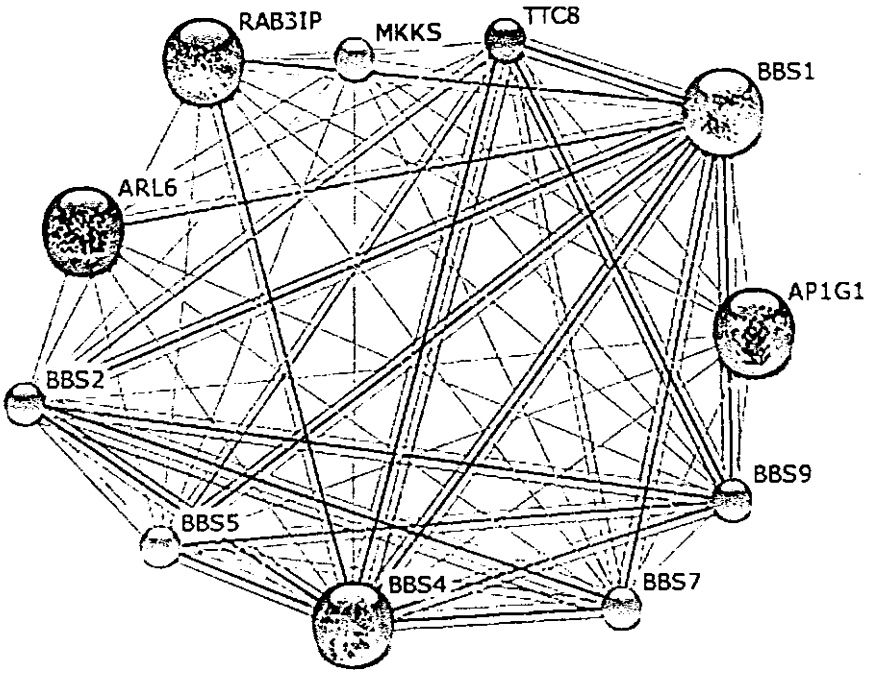


Figure 3.16. Interactions obtained through STRING for BBS8

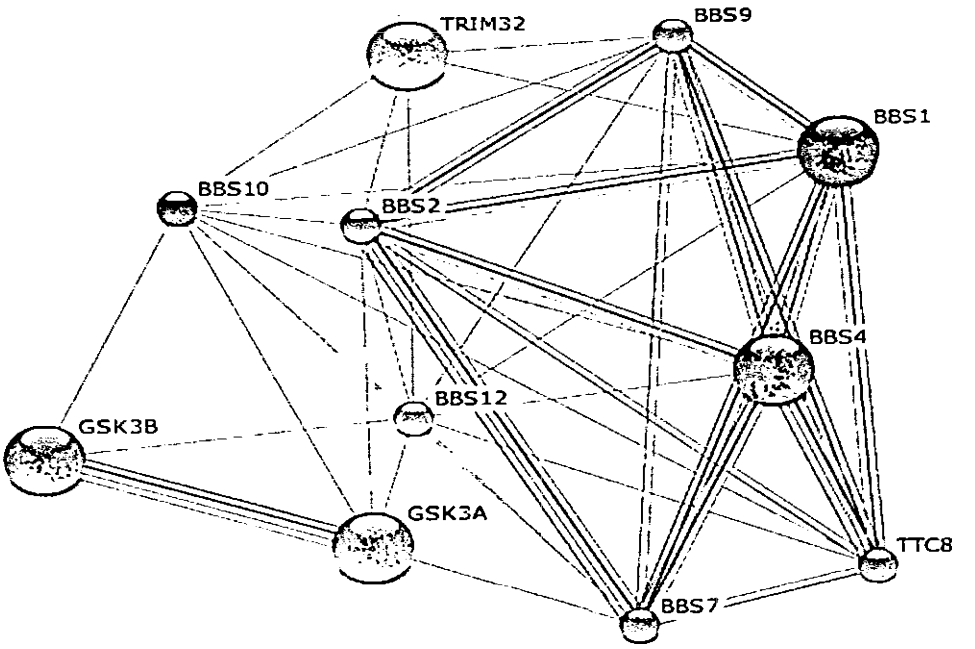


Figure 3.17. Interactions obtained through STRING for BBS10

3.8.2 PyMol:

Selection Algebra:

On the basis of above found interactions of the BBSome complex, several interaction possibilities have been mapped out for docking studies. These possible interactions are further elucidated by the extent of interaction found between various proteins. These are:

BBS1-BBS2:

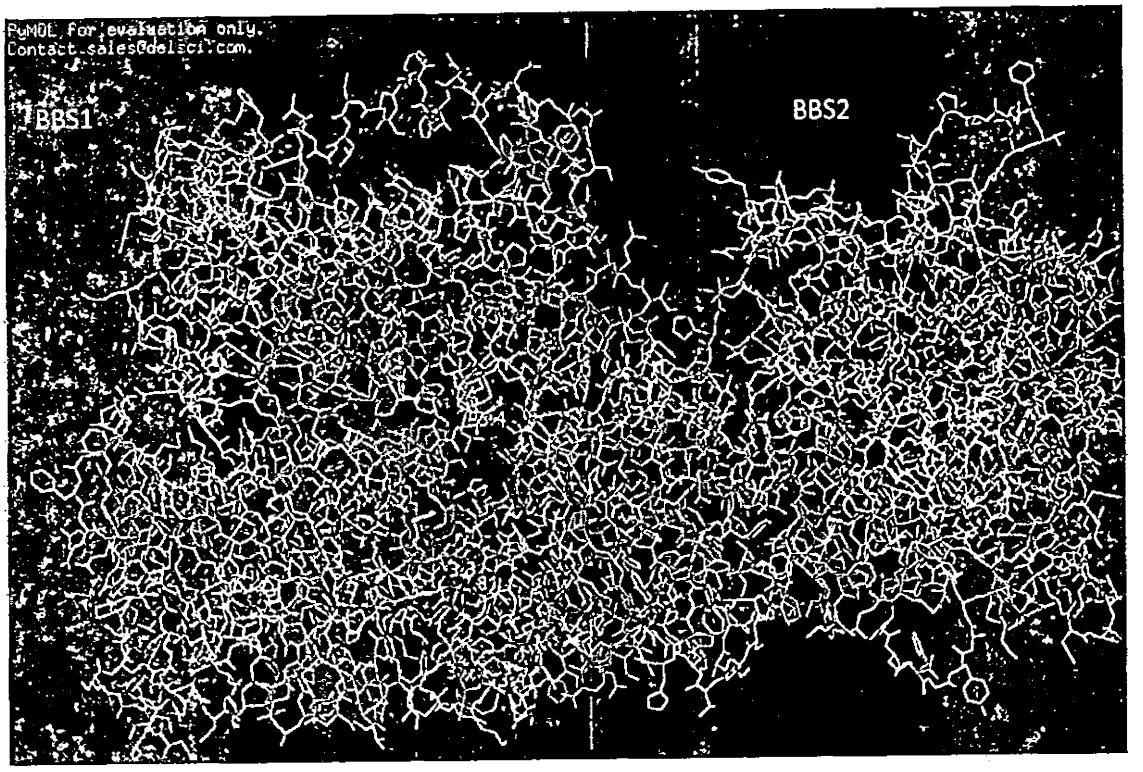


Figure 3.18. Result of BBS1-BBS2 docking, Showing Hydrogen Bonds

The details of hydrogen bonds of the BBS1-BBS2 complex are as under:

Table 3.19. Hydrogen Bond positions for BBS1-BBS2 (docked)

BBS1	Gln	Arg	Lys	Arg	Arg	Ser	Tyr	Tyr	Tyr
→	141	100	241	483	483	529	528	528	528
BBS2	Ile	Pro	Ser	Ser	Val	Arg	Arg	Ser	Ser
→	443	444	440	721	716	703	703	704	704
Bond Distance	3.2	3.0	3.4	3.3	2.8	3.4	3.3	3.4	3.3

BBS1-BBS4:

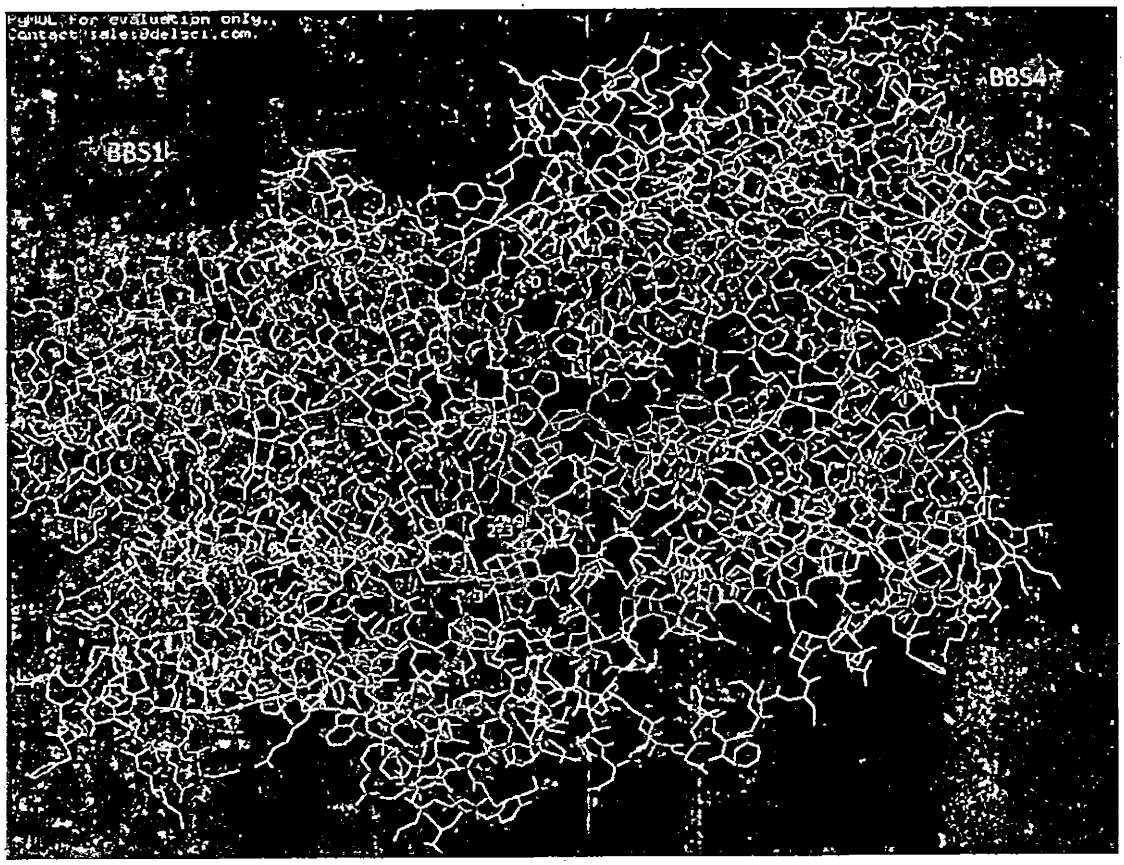


Figure 3.19. Result of BBS1-BBS4 docking, Showing Hydrogen Bonds

Following table shows the details of hydrogen bonds present in the BBS1-BBS4 complex:

Table 3.20. Hydrogen Bond positions for BBS1-BBS4 (docked)

BBS1	Ser	Ala	Ser	Ser	Gln	Asn	Pro	Asp	Lys
→	21	11	7	243	141	524	370	371	317
BBS4	Asn	Gln	Thr	Leu	Leu	Thr	Gln	Gln	Glu
→	297	247	248	215	215	452	448	448	352
Bond Distance	3.5	3.1	3.3	2.9	2.3	3.0	2.4	2.7	3.5

BBS1-BBS5:

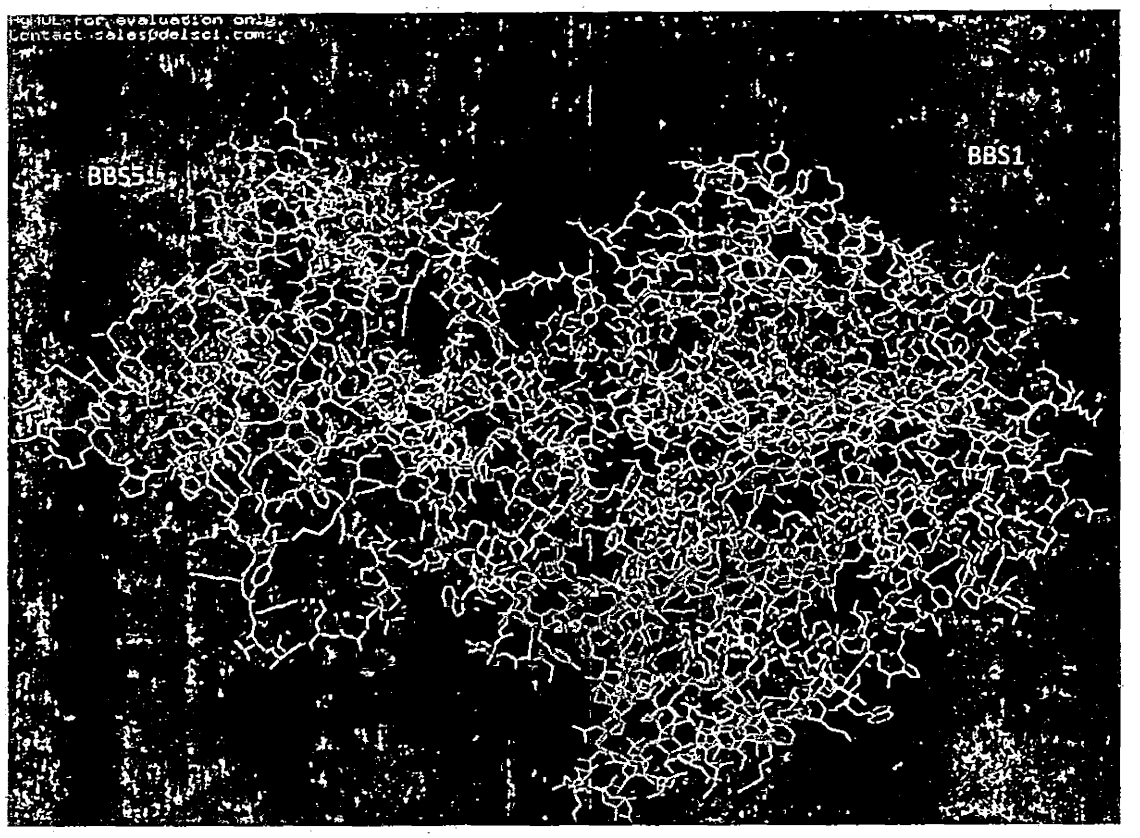


Figure 3.20. Result of BBS1-BBS5 docking, Showing Hydrogen Bonds

The details of hydrogen bonds found in the BBS1-BBS5 complex are as under:

Table 3.21. Hydrogen Bond positions for BBS1-BBS5 (docked)

BBS1	Arg	Arg	His	Thr
→	262	268	281	498
BBS5	Gln	Gln	Tyr	His
→	148	157	161	190
Bond Distance	3.2	3.1	2.9	3.4

BBS1-BBS7:

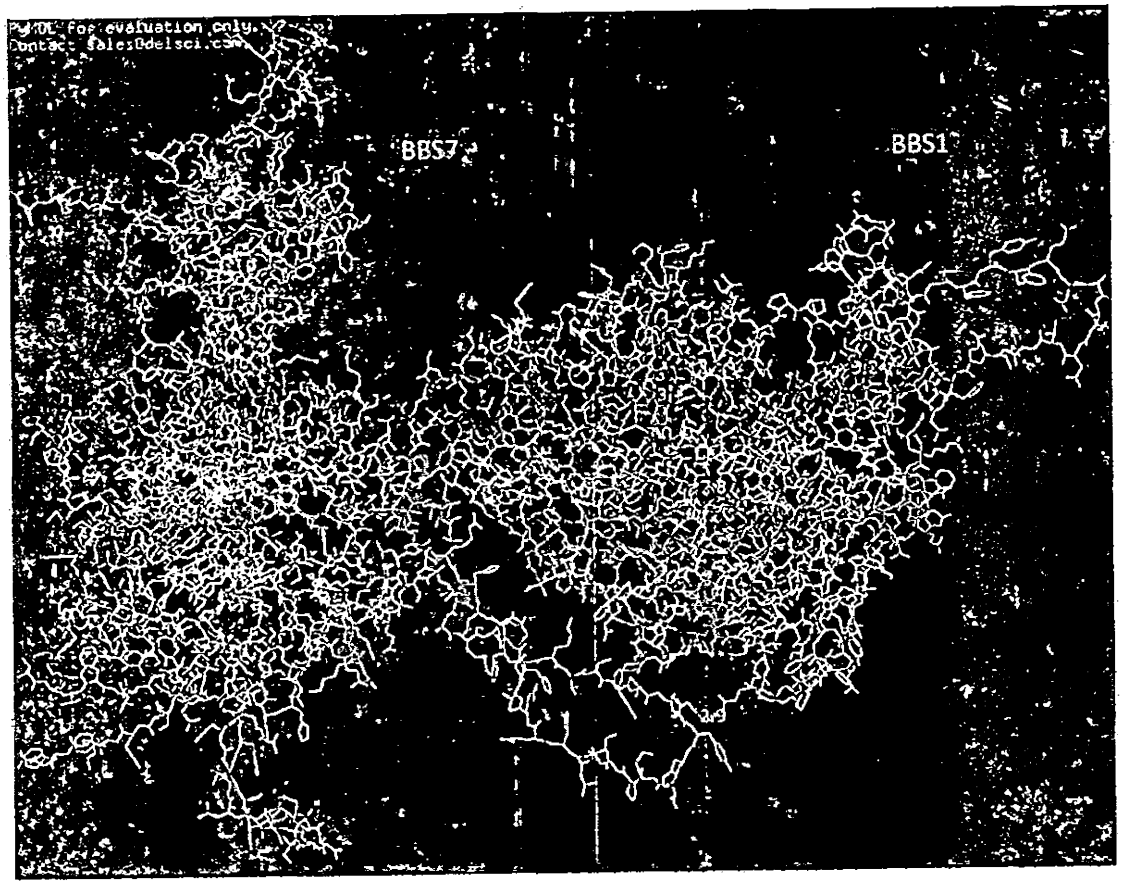


Figure 3.21. Result of BBS1-BBS7 docking, Showing Hydrogen Bonds

The hydrogen bonds found in this complex are as under:

Table 3.22. Hydrogen Bond positions for BBS1-BBS7 (docked)

BBS1	GLU	ASN	GLY	GLU	ASP
→	51	33	63	164	258
BBS7	Ser	Arg	Phe	Arg	Cys
→	567	559	526	484	527
Bond Distance	2.8	3.2	3.0	2.9	3.4

BBS1-BBS8:

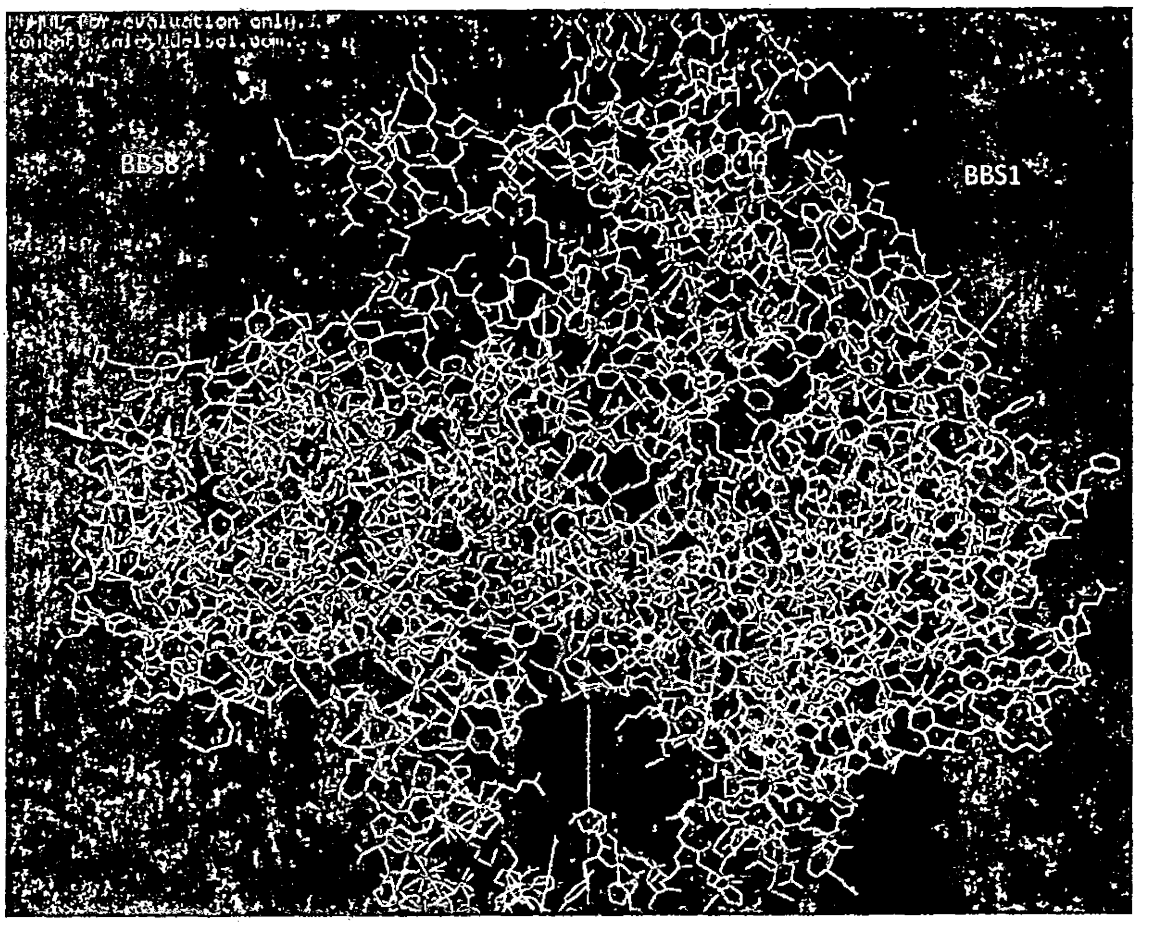


Figure 3.22. Result of BBS1-BBS8 docking, Showing Hydrogen Bonds

A look at the BBS1-BBS8 complex reveals the following hydrogen bonds:

Table 3.23. Hydrogen Bond positions for BBS1-BBS8 (docked)

BBS1	HIS	ASP	LEU
→	316	371	522
BBS8	Gln	Thr	Gln
→	380	407	401
Bond Distance	2.3	3.1	2.2

BBS2-BBS4:

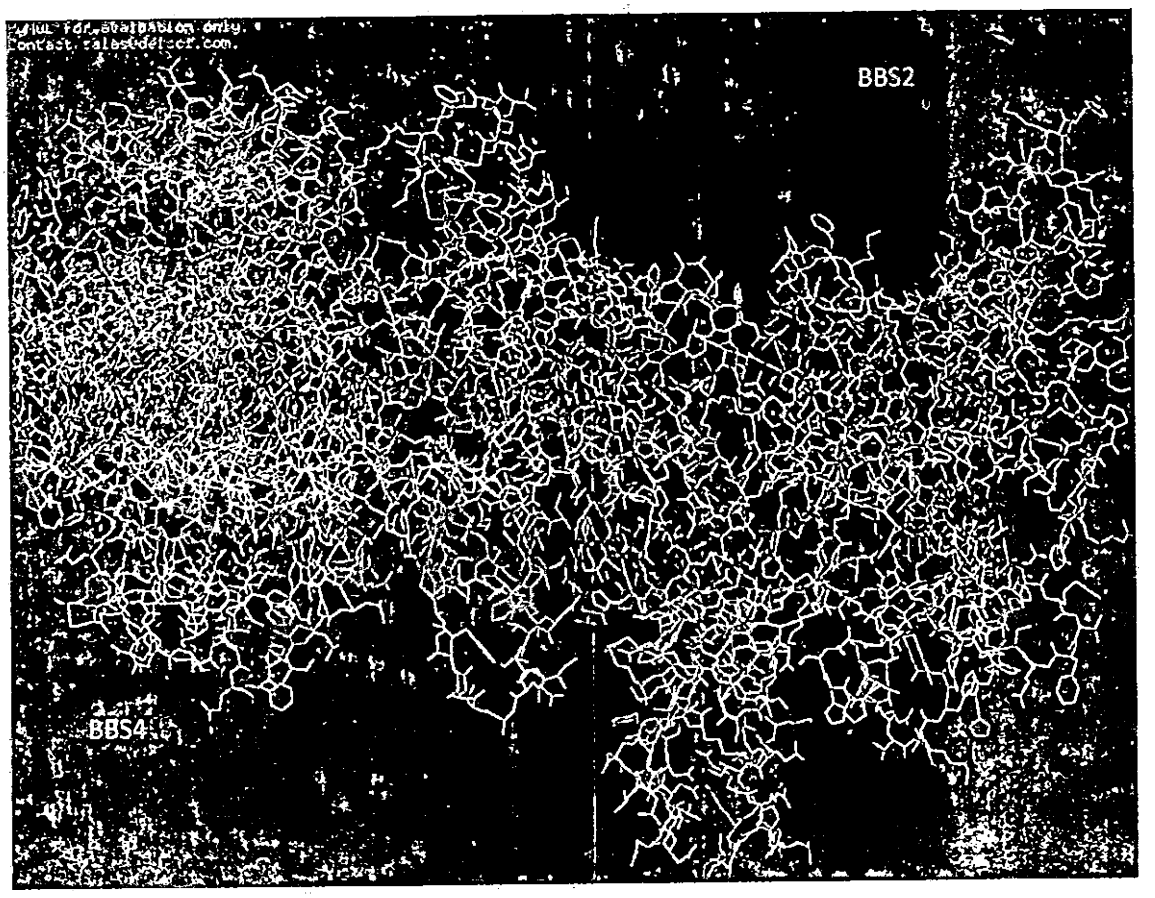


Figure 3.23. Result of BBS2-BBS4 docking, Showing Hydrogen Bonds

The hydrogen bonds found in this complex are as under:

Table 3.24. Hydrogen Bond positions for BBS2-BBS4 (docked)

BBS2	Asp	Ala	Arg	Thr
→	454	719	703	718
BBS4	Ser	Lys	Asn	Thr
→	110	116	227	230
Bond Distance	2.8	2.8	3.2, 3.3	2.8, 2.7

BBS2-BBS7:

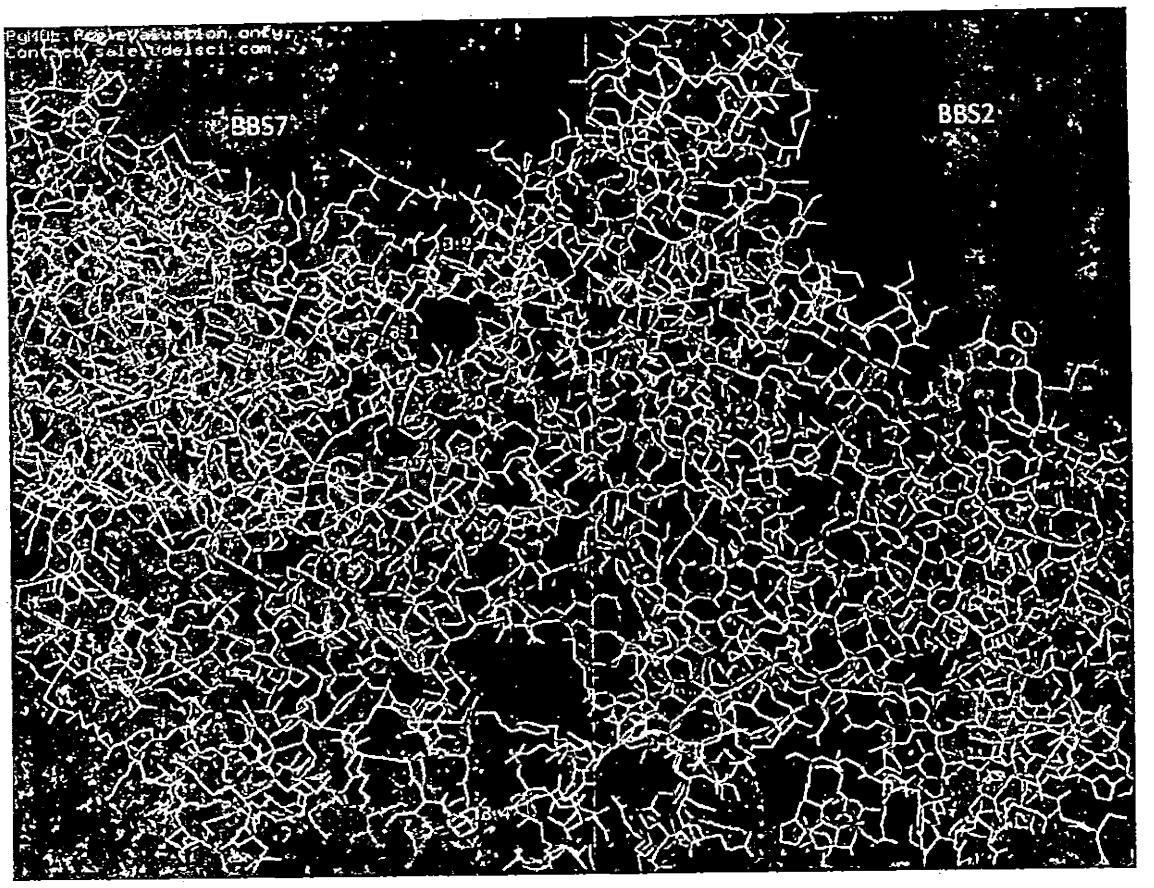


Figure 3.24. Result of BBS2-BBS7 docking, Showing Hydrogen Bonds

The details of hydrogen bonds found in the BBS2-BBS7 complex are as under:

Table 3.25. Hydrogen Bond positions for BBS2-BBS7 (docked)

BBS2	Ala	Ser	Ile	Ser	Leu	Ala	Ser
→	419	440	443	488	647	719	720
BBS7	Tyr	Leu	Thr	Arg	Gly	Arg	Arg
→	615	317	319	346	63	261	261
Bond Distance	2.7	3.2	2.0	3.2	3.4	3.1	3.4

BBS4-BBS5:

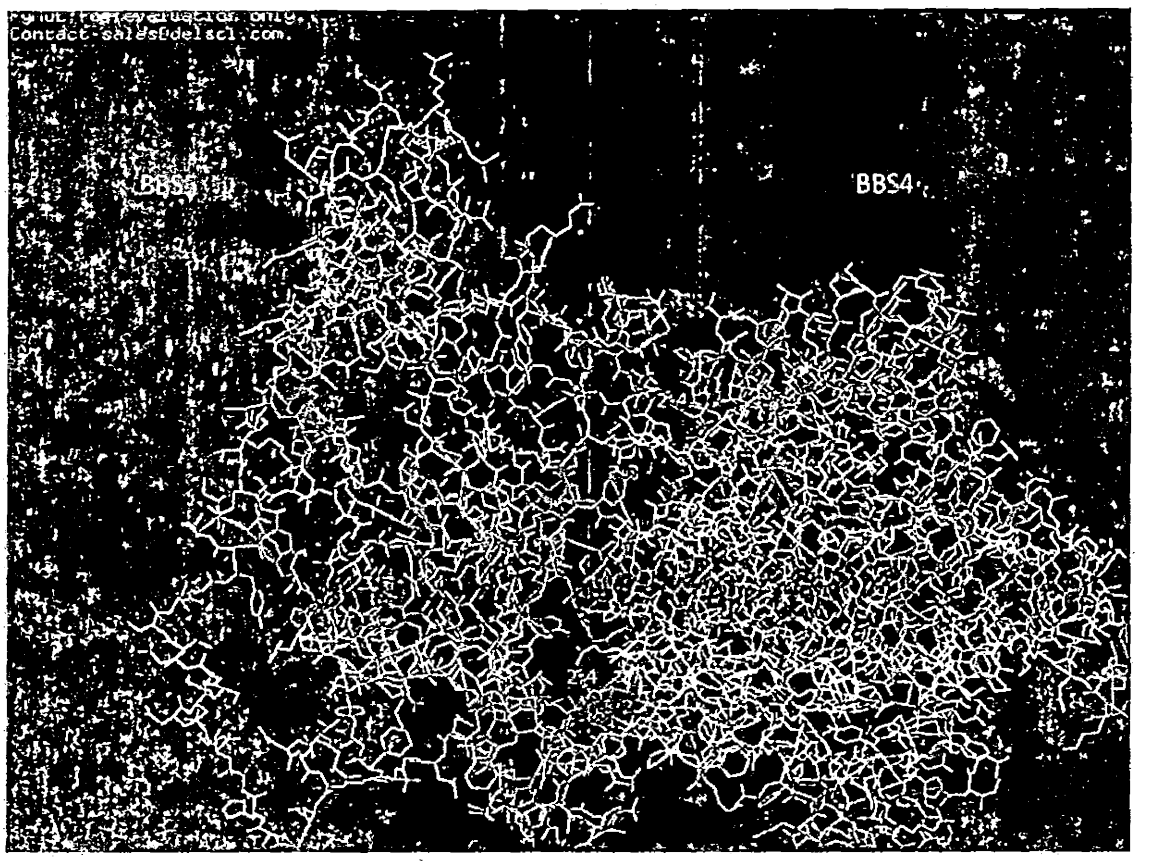


Figure 3.25. Result of BBS4-BBS5 docking, Showing Hydrogen Bonds

The BBS4-BBS5 complex showed following hydrogen bond interactions when viewed through Pymol:

Table 3.26. Hydrogen Bond positions for BBS4-BBS5 (docked)

BBS4	Asn	Glu	Arg	Glu	Thr
→	332	352	359	363	449
BBS5	Glu	Tyr	Arg	Leu	Glu
→	277	137	153	154	319
Bond Distance	3.5	3.3	2.4	3.2	2.4

BBS4-BBS7:

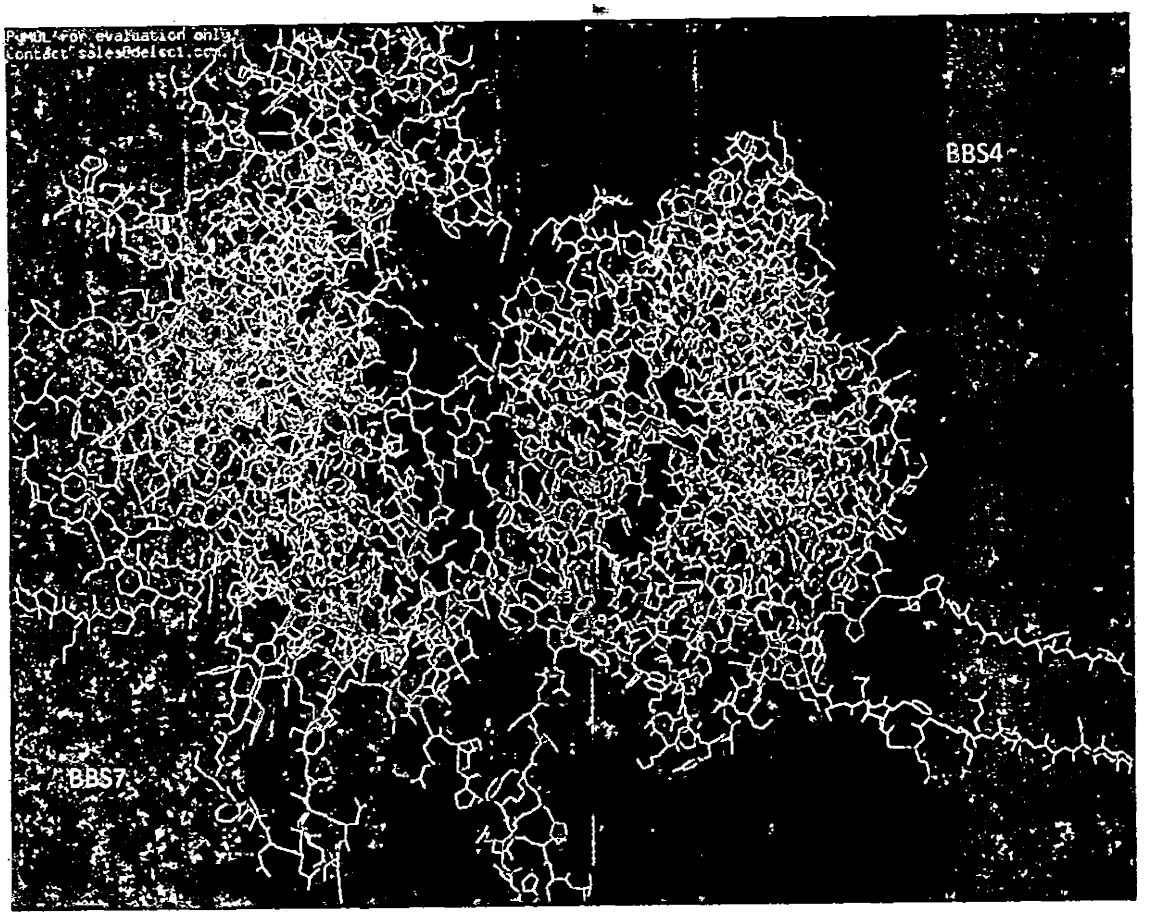


Figure 3.26. Result of BBS4-BBS7 docking, Showing Hydrogen Bonds

A look at the BBS4-BBS7 complex reveals the following hydrogen bonds:

Table 3.27. Hydrogen Bond positions for BBS4-BBS7 (docked)

BBS4	Leu	Tyr	Thr	Asn	Asn	Thr	Thr
→	215	218	248	275	297	449	452
BBS7	Val	Ser	Ser	Val	Glu	Lys	Lys
→	520	522	522	525	530	577	586
Bond Distance	2.3	1.9	3.5	2.9	2.0	2.8	3.3

BBS4-BBS8:

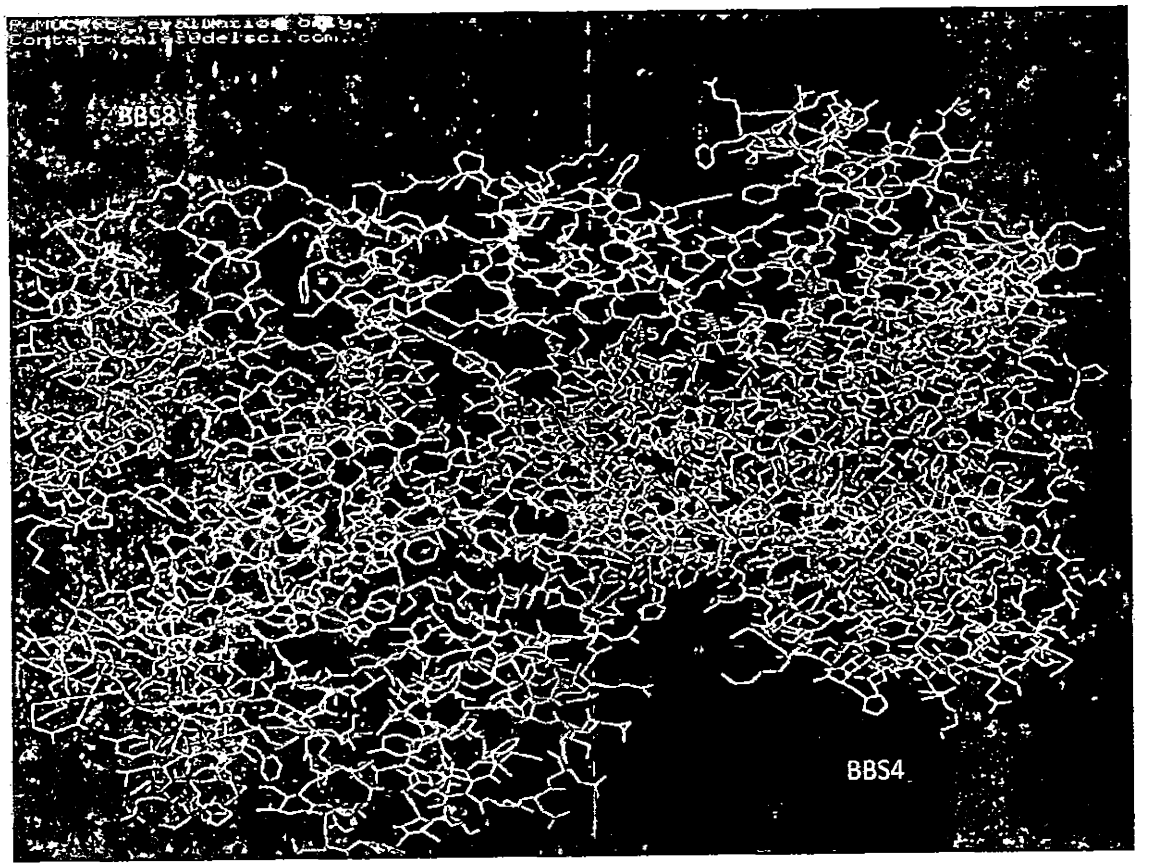


Figure 3.27. Result of BBS4-BBS8 docking, Showing Hydrogen Bonds

The docked complex of BBS4 and BBS8 confirmed the following hydrogen interactions:

Table 3.28. Hydrogen Bond positions for BBS4-BBS8 (docked)

BBS4 →	Thr 62	Gln 63	Asn 227
BBS8 →	Ile 64	Ile 67	Phe 222
Bond Distance	2.5	3.5	2.5

Chapter 4

Discussion

4. DISCUSSION

Protein structure recognition and modeling is one of the most intriguing and challenging task as far as structural and computational biology is concerned. Despite our knowledge of the biochemical properties and behavior of the proteins, the interaction prototypes of proteins are still quite hard to pin down. Comparative modeling techniques are being employed quite often to address the mystery of protein interactions

Here I have investigated the properties of functional module i.e., BBSome complex, designing the models for the eight proteins in question, one being a molecular chaperone and the other one is the Rab8 GTPase. The BBS proteins pose a different situation to the advancement of the research by showing very low sequence similarity to those of known structures as shown in BLAST results of these proteins (Table 3.1). The complexity of the process is increased to a greater extent by the fact that these proteins interacting together to form the BBSome complex show very less similarity to the other members of the complex too, thus making the process a more elusive one. Protein-Protein interactions are fundamental to the biological processes in a cell. Many important cellular functions are implemented by protein complexes.

During the study, post translational modifications in the case of BBS proteins are viewed through SignalP. None of the proteins appeared to be as a signal peptide as it is evident from the data obtained for these proteins (Table 3.2 & Figure 3.1 to 3.7). Thus, on the basis of our findings it is proposed that they do not possess the mechanism to reach to their final destination and need to be transported to the base of cilia through some signaling pathway. These findings about the BBSome proteins appear to provide an answer to the question of involvement of BBSome proteins in transport of signaling

receptors and that whether the BBSome is directly involved in vesicular trafficking to the cilium (Jin and Nachury 2009). The results for known proteins of the BBSome complex are then cross checked through SecretomeP server to check for the possibility of Non-classical secretion.

The primary structure details obtained through ProtParam have revealed that BBS1, BBS4 and BBS5 showed potentially higher values (>45) for instability indexes (Table 3.9). These members of the complex display somewhat unstable behavior, while the rest (BBS2, BBS7, BBS8, BBS10) are reasonably stable thus rendering the whole complex as a stable one.

For analyzing the potential antigenic sites of these proteins, ProtScale has been used out of various options available. As we know that antigenic components of proteins are mostly the surface features so Hopp & Woods hydrophobicity scale was employed to get an insight into the antigenic and surface details of the proteins of the BBSome complex. High peak values in case of BBS2, BBS4 and BBS5 correspond to greater antigenic activity than the remaining members of the complex. The regions of high antigenicity in the case of BBS2 are 170-180 and 270-280 residues. In case of BBS4 the antigenic site lies near 20-25 residues while in case of BBS5, the region is 310-315. Antigenicity and catalytic activity of proteins are correlated (Yan and Harding 1997). Moreover, the antigenic sites may also support a possible role as a surface-exposed receptor binding site (Langedijk *et al.* 1997). The antigenic regions identified in these proteins corresponded well to the active residue sites as identified by PyMOL in interaction studies of these proteins (Table. 3.19 to Table. 3.28). Thus, confirming the possibility of enzymatic and antigenic activity in this complex of proteins.

The secondary structure analysis of the BBS proteins has revealed that these proteins possess quite fewer secondary components as is evident from Table 3.10 and Table 3.11. This may be due to the interactions between proteins. Because, when proteins interact together to form complexes, their interacting sequences may be essentially disordered; they lack the crucial secondary and tertiary structure components to ensure firm, stable and potentially viable interactions.

All the members of the BBSome complex except BBS5 appear to possess low complexity regions (Table. 3.10). Low complexity regions are known as regions of proteins with biased amino acid composition. The proteins having low complexity regions tend to have a greater participation in interactions than those lacking these regions (Coletta *et al.* 2010). Moreover, the low complexity regions of these proteins appear to be central rather than terminal, which indicate role of these proteins in transcription and transcription regulation processes.

The models of BBS proteins selected in this project have been designed mainly through Modeller while various other modeling servers have also been used to generate models for these proteins. However, Modeller utilizes all the ways to minimize the energy of the newly designed models, as well as it provides models for the full length sequence of the protein in question. A fair number of models (45) against each protein were designed through Modeller. When analyzed through ProSA, Swiss Model and Ramachandran plot 2.0, the models designed through Modeller gave best results as compared to all others in almost all of the cases. Esypred3D has also produced fairly comparable models for BBS8, BBS10 and quite compromising models in case of BBS1 and BBS5.

When the modeling process was over, proteins of the BBSome complex were docked through GRAMMX (Tovchigrechko and Vakser 2006). The pattern for docking was originally based on the interactions present in the BBSome complex as provided by STRING database. The selection algebra for these docked complexes was basically based on the parameters used by STRING database. The docked complexes were than analyzed for possible ionic interactions, most importantly being hydrogen bonding through PyMOL. Fair number of interactions was observed in all the docked complexes. Almost 50% of the residues that took part in these interactions were aromatic amino acids, other participants being polar ones (Table 3.19 to Table 3.28). When the residues participating actively in these interactions were analyzed, they came out to be fairly related to the antigenic sites of these member proteins. In case of BBS1, actively participating residue regions that appear to interact most promisingly in all the docked complexes are 522-529, 483-498, 370-371 and 241-268. These regions of BBS1 may pose for potential interacting pockets. BBS2 when checked for similar interaction patterns revealed two interactive sites over the entire sequence of the protein i.e., 419-454 and 703-721. A look at BBS4 docking and interaction prototypes showed residues at 215-248 and 448-452 to be the active participants of interactions. Interaction pattern found in BBS7 revealed residues at 520-586 position as potential interactive site. BBS5 and BBS8 came out to be less interactive than other members of the complex while BBS10 which is a molecular chaperone did not participate in any direct interaction with the proteins of the BBSome complex. When these interaction sites in all the member proteins were analyzed, their amino acid composition demonstrated a combined behavior in terms of polarity of these sites. Amino acids at these positions are mostly polar, being neutral as well as charged.

Thus the proteins of the BBSome complex displayed more polar or hydrophilic character than hydrophobic one, as it is evident from ProtScale results and is also verified through docking studies. This character of proteins is important in determining their role in protein-protein interactions as well as in the structure determination of the protein.

Interestingly, member proteins of the BBSome complex are reported to interact with the Rab8 small GTPase factor for the proper cilium assembly and for the extension of ciliary membranes (Nachury *et al.* 2007). Rab8 is reported to uphold the docking and fusion of vesicles near the ciliary membrane (Leroux 2007). Investigation of the interaction pattern of this GTPase with the BBSome proteins revealed some interesting patterns in the complex. Firstly the docking of BBSome Proteins was done on the basis of STRING data as described in the PyMOL section. After that, docking of these proteins was carried out with the Rab8 GTPase. The docked complexes were then docked again according to the pattern obtained by STRING. The next step in the process was to analyze the simple dockings through Hex on the basis of their energy values. The energy values were then compared to those obtained after docking with Rab8 GTPase and a striking difference was observed. The binding energies of the Rab8 complexes appeared to be less than those of the simply docked complexes, thus giving a hint about the positive interaction impact of Rab8 GTPase. On the basis of these findings, it is proposed that the BBS proteins first interact with Rab8 GTPase and then interact with each other to form a stable complex.

FUTURE WORK:

Protein–protein interactions (PPI) participate in many biological progressions such as gene expression control, enzyme inhibition, signal transduction, antibody–antigen recognition or even the assemblage of multi-domain proteins. Interactomics not only

attempts to characterize the interaction between proteins, but between all molecules in the cell. While genomes are stable, interactomes may vary between tissues and developmental stages. That is why their study involves much detailed information of the genome. Much of the signal transduction processes get affected by binding interactions of proteins. Thus mapping protein-protein interactions can play an important role in identifying potential drug targets. As for the proteins of the BBSome complex, much remains still unexplored. So study of BBS proteins with respect to the fact that whether they are involved in one way transport to cilia or otherwise is an important and viable research area.

BIBILIOGRAPHY:

Altschul S. F., Gish W., Miller W., Myers E. W., and Lipman D. J., 1990. Basic local alignment search tool. J. Mol. Biol. 215: 403–410.

Ansley S.J., et al. 2003. Basal body dysfunction is a likely cause of pleiotropic Bardet-Biedl syndrome. Nature. 425:628–633.

Beales P.L., Elcioglu N., Woolf A.S., Parker D., and Flinter F.A. 1999. New criteria for improved diagnosis of Bardet-Biedl syndrome: results of a population survey. J. Med. Genet. 36:437–446.

Bendtsen J.D., Nielsen H., Brunak. S., 2004. Improved prediction of signal peptides SignalP 3.0 J Mol Biol. 340(4):783-795

Bishop A.O.T., Beer T.A.P., Joubert F., 2008. Protein homology modeling and its use in South Africa. S. Afr. j. sci. 104: 2-6

Blacque O.E., and Leroux M.R., 2006. Bardet-Biedl syndrome: an emerging pathomechanism of intracellular transport. Cell. Mol. Life Sci. 63: 2145–2161

Blacque O.E., Reardon M.J., Li C., McCarthy J., Mahjoub M.R., Ansley S.J., Badano J.L., Mah A.K., Beales P.L., Davidson W.S., Johnsen R.C., Audeh M., Plasterk R.H.A., Baillie D.L., Katsanis N., Quarmby L.M., Wicks S.R., Leroux M.R., 2004. Loss of C. elegans BBS-7 and BBS-8 protein function results in cilia defects and compromised intraflagellar transport. Genes Dev. 18(13):1630–1642

References

BIBLIOGRAPHY:

Blundell T.L., Sibanda B.L., Sternberg M.J.E., Thornton J.M., 1987. Knowledge-based prediction of protein structures and the design of novel molecules. Nature. 323:347–352

Browne W. J., North A. C. T., Phillips D. C., et al. 1969. A possible three-dimensional structure of bovine lactalbumin based on that of hen's egg-white lysosyme. J. Mol. Biol.42: 65–86.

Cheng J., Randall A.Z., Sweredoski M.J., Baldi P., 2005. SCRATCH: a protein structure and structural feature prediction server. Nucleic Acids Research. 33: W72-W76

Coletta A., Pinney J.W., Solis D.Y.W., Marsh J., Pettifer S.R., Attwood T.K., 2010. Low-complexity regions within protein sequences have position-dependent roles. BMC Systems Biology. 4:43

DeLano W.L., The PyMOL Molecular Graphics System, 2002.

European Research Consortium for Informatics and Mathematics. Ab Initio Methods for Protein Structure Prediction: A New Technique based on Ramachandran Plots. http://www.ercim.org/publication/Ercim_News/enw43/bernasconi.html [accessed 30.07.2011]

Gasteiger E., Hoogland C., Gattiker A., Duvaud S., Wilkins M.R., Appel R.D., Bairoch A., 2005. Protein Identification and Analysis Tools on the ExPASy Server John M. Walker (ed): The Proteomics Protocols Handbook, Humana Press. pp. 571-607

Gopalakrishnan K., Sowmiya G., Sheik S.S., Sekar K., 2007. Ramachandran plot on the web (2.0) Protein Pept Lett. 14(7):669-71.

Hofmann K., and Stoffel W., 1993. TMbase - A database of membrane spanning proteins segments Biol. Chem. Hoppe-Seyler. 374:166

Holm L., and Sander C., 1999. Protein folds and families: sequence and structure alignments. Nucleic Acids Res. 27: 244–247.

Hooft R. W., Vriend G., Sander C., and Abola E. E., 1996. Errors in protein structures. Nature. 381:272.

http://www.umass.edu/molvis/workshop/prot1234.htm [accessed 30.07.2011]

Jensen L.J., Kuhn M., Stark M., Chaffron S., Creevey C., Muller J., Doerks T., Julien P., Roth A., Simonovic M., Bork P., Mering C.V., 2009. STRING 8-A global view on proteins and their functional interactions in 630 organisms. Nucleic Acids Res. 37:D412-6.

Jin H., and Nachury M.V., 2009. The BBSome. Current Biology. 19(12):R472-3.

Jones, S., and Thornton, J.M., (1996) Principles of protein-protein interactions. Proc. Natl. Acad. Sci. USA. Vol. 93, pp. 13-20.

Katsanis N., 2004. The oligogenic properties of Bardet-Biedl syndrome. Hum. Mol. Genet. 13:R65–R71.

Kiefer F., Arnold K., Künzli M., Bordoli L., Schwede T., 2009. The SWISS-MODEL Repository and associated resources. Nucleic Acids Research. 37: D387-D392.

Kneller D.G., Cohen F. E., and Langridge R., 1990. Improvements in Protein Secondary Structure Prediction by an Enhanced Neural Network. J. Mol. Biol. 214:171-182

Lambert C., Leonard N., Bolle X.D., Depiereux E., 2002. ESyPred3D: Prediction of Protein 3D Structures. Bioinformatics. 18:1250-1256

Langedijk J.P.M., Daus F.J., Oirschot J.T.V., 1997. Sequence and Structure Alignment of Paramyxoviridae Attachment Proteins and Discovery of Enzymatic Activity for a Morbillivirus Hemagglutinin. Journal of Virology. 71:6155–6167

Laskowski R. A., Rullmann J. A., MacArthur M. W., Kaptein R., and Thornton J. M., 1996. AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. J. Biomol. NMR. 8:477–486.

Laskowski R.A., MacArthur M.W., and Thornton J.M., 1998. Validation of protein models derived from experiment. Curr. Opin. Struct. Biol. 8:631-639

Leroux M.R., 2007. Taking Vesicular Transport to the Cilium. Cell. 129:1041-1042.

Lo Conte L., Brenner S. E., Hubbard T. J., Chothia C., and Murzin A. G., 2002. SCOP database: refinements accommodate structural genomics. Nucleic Acids Res. 30:264–267.

Luthy R., Bowie J. U., and Eisenberg D., 1992. Assessment of protein models with three-dimensional profiles. Nature 356:83–85.

Marti-Renom M. A., Stuart A. C., Fisar A., Sanchez R., Melo F., and Sali A., 2000. Comparative protein structure modeling of genes and genomes. Annu. Rev. Biophys. Biomol. Struct. 29:291–325

Nachury M.V., Loktev A.V., Zhang Q., Westlake C.J., Peränen J., Merdes A., Slusarski D.C., Scheller R.H., Bazan J.F., Sheffield V.C., Jackson P.K., 2007. A core complex of BBS proteins cooperates with the GTPase Rab8 to promote ciliary membrane biogenesis. Cell. 129:1201–1213.

Orengo C.A., Bray J.E., Buchan D.W., 2002. The CATH protein family database: a resource for structural and functional annotation of genomes. Proteomics. 2:11–21.

Ou G., Blacque O.E., Snow J.J., Leroux M.R., and Scholey J.M., 2005. Functional coordination of intraflagellar transport motors. Nature. 436:583–587.

Pearson W.R., and Lipman D.J., 1988. Improved tools for biological sequence comparison. Proc. Natl. Acad. Sci. 85: 2444–2448.

Pearson W. R., 1990. Rapid and sensitive sequence comparison with FASTP and FASTA. Methods Enzymol. 183:63–98.

Pieper U., Webb B.M., Barkan D.T., Duhovny D.S., Schlessinger A., Braberg H., Yang Z., Meng E.C., Pettersen E.F., Huang C.C., Datta R.S., Sampathkumar P., Madhusudhan M.S., Sjolander K., Ferrin T.E., Burley S.K., Sali A., 2006. MODBASE: a database of annotated comparative protein structure models and associated resources. Nucleic Acids Research. 34: D291–D295

Rost B., Yachdav G., Liu J., 2004. The Predict Protein server. Nucleic Acids Research. 32:W321-W326.

Sali A., and Blundell T.L., 1993. Comparative protein modelling by satisfaction of spatial restraints. J. Mol. Biol. 234:779-815.

Sanchez R., and Sali A., 1998. Large-scale protein structure modeling of the Saccharomyces cerevisiae genome. Proc. Natl. Acad. Sci. USA. 95:13597–13602.

Schwede T., Kopp J., Guex N., Peitsch M.C., 2003. SWISS-MODEL: An automated protein homology-modeling server. Nucleic Acids Res. 31:3381-3385.

Sippl M.J., 1993. Recognition of errors in three-dimensional structures of proteins. Proteins. 17:355-362.

Tobin J.L., and Beales P.L. 2007. Bardet-Biedl syndrome: beyond the cilium. Pediatr. Nephrol. 22:926–936.

Tovchigrechko A., Vakser I.A., 2006. GRAMM-X public web server for protein-protein docking. Nucleic Acids Res. 34:W310-4.

Tramontano A., Leplae R., and Morea V., 2001. Analysis and assessment of comparative modeling predictions in CASP4. Proteins. 45:22–38.

Wiederstein M., and Sippl M.J., 2007. ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. Nucleic Acids Research. 35:W407–W410

Yan H., and Harding J.J., 1997. Glycation-induced inactivation and loss of antigenicity of catalase and superoxide dismutase. Biochem. J. 328:599-605