IMPACT OF POINT MUTATION ON THE STRUCTURE AND FUNCTION OF GBSS-1 ENZYME DETERMINING THE QUALITY OF VARIOUS RICE CULTIVARS

BY:

PRIYASHREE VERMA - 091511

RAHUL ACHARYA - 091514

PROJECT SUPERVISOR – DR. PRADEEP NAIK

May – 2013

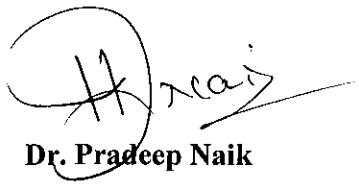
Submitted in partial fulfillment of the Degree of
Bachelor of Technology
In
BIOINFORMATICS

DEPARTMENT OF BIO-TECHNOLOGY AND BIO- INFORMATICS
JAYPEE UNIVERSITY OF INFORMATION TECHNOLOGY,
WAKNAGHAT

Table of Contents

CERTIFICATE

This is to certify that the work titled " **Impact of point mutation on the structure and function of GBSS-1 enzyme determining the quality of various rice cultivars**" submitted by ***Priyashree Verma*** and ***Rahul Acharya*** in partial fulfillment for the award of degree of B.Tech. of Jaypee University of Information Technology, Waknaghat has been carried out under my supervision. This work has not been submitted partially or wholly to any other University or Institute for the award of this or any other degree or diploma.



Dr. Pradeep Naik

Associate Professor

Department of Biotechnology and Bioinformatics

Jaypee University of Information Technology

Waknaghat, Solan (H.P)

Date:

ACKNOWLEDGEMENT

It was a great pleasure and honor working under the faculty and staff of the Department of Biotechnology and Bioinformatics at Jaypee University of Information Technology during our study as B. Tech students. We extend our heartfelt gratitude to **Dr. Pradeep Naik**, our Project guide, for his constant support and guidance, and his painstaking efforts to develop our scientific temper and helping us convert this project from an idea to reality.

We would also like to extend our warmest gratitude to **Ms. Seneha Santoshi** and **Ms. Charu Suri**, pursuing their ph.D at JUIT, Waknaghat, who took out time from their busy academic schedules in order to help us, for clearing our doubts and supporting us at every point. Their constant support gave us the confidence to keep moving ahead.

The list of the people who have helped us in the accomplishment of this project is long and we express our sincerest appreciation for all of them. We are indebted to all those who believed in us, provided reviews, pointed out our mistakes and gave us the clarity of thought that we most needed. Last but not the least, we owe it a great deal to our teachers whose efforts since the beginning of our tenure at JUIT, made us capable enough to undertake this final year project.

The blessings of our parents and God, the almighty are the driving forces of our very lives.

Sincere apologies to all those, whose names we may have failed to mention. Their efforts shall live on with us, inspiring each moment of our lives.

Priyashree Verma

(091511)

Rahul Acharya

(091514)

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

GBSS-1 : Granule Bound Starch Synthase-1,

GROMACS : GROningen MAchine for Chemical Simulations

MD simulation : Molecular Dynamics Simulation

ABSTRACT

The basic aim of our project was to study the overall impact of point mutations, if any, upon the structure and function of GBBS-1 enzyme, the enzyme that is responsible for the synthesis of amylose in the cells. We used a number of computational methods to carry out the study, which include Homology modelling, Molecular dynamics simulation, pairwise sequence alignment and Ligand docking. Our results were analysed with the help of statistical data obtained from these computational methods which we converted to graphs for better understanding and comparative analysis of all the 19 cultivars provided to us. The project helped us analyse differences in the protein of different cultivars at the molecular level and helped us understand the impact of point mutations in the same.

CHAPTER 1

INTRODUCTION

Rice is one of the most widely consumed cereals in the world and serves as the staple food of around 3 billion people in the world who consume it every day. One fifth of the world's population—more than a billion people—depend on rice cultivation for their livelihoods. Asia, where about 90% of world's rice is grown, has more than 200 million rice farms. Rice-based farming is the main economic activity for hundreds of millions of rural poor in this region.In most of the developing world, rice is equated with food security and closely connected to political security. Changes in rice availability, and hence price, have caused social unrest in several countries.The importance and need of rice as a food crop can further be understood by studying the nutritional value this crop provides which primarily includes carbohydrates (starch), proteins and several other vitamins and minerals that are essential for healthy living.

Rice has different varieties, usually depending upon the region of cultivation. There are Indian varieties, Pakistani varieties, japanese, french, canadian, thai and so on and so forth. Each variety has a distinct flavour, sometimes a distinct colour, taste and size, shape and quality of the grain. But, overall, there are only four varieties of rice in the world and all the other varieties, which we shall henceforth be referring to as cultivars share most or atleast some of their traits which puts them in one of the four categories, which are:

1. Indica
2. Japonica
3. Aromatic
4. Glutinous

Yet, when we talk about the quality of rice, or more precisely, quality of the cooked grain of rice, we find two kinds:

1. Waxy or Sticky
2. Non-waxy or non-sticky

The waxy nature of rice is primarily found associated with the variety japonica, named so because of it being the native variety of japan. The non-waxy type, the one which exhibits separate grains when

cooked, is the characteristic trait of the variety Indica, named so, as the name suggests, because of it being the native variety of India.

The waxy nature of rice or more precisely, the sticky nature of the grain of the rice when cooked is attributed to the fact that it imbibes a lot of water while being cooked, swells and "something inside it" leads to it being sticky. On the other hand, as Indians, as we take pride and associate ourselves with the "Basmati rice" that is so prevalent in India, it is implicit that there is another quality that imbibes less water, swells less and there is "something inside it" that leads to the different grains being separate from each other and therefore, no stickiness. This curiousity of what is that "something inside" that differentiates the two and makes them so distinct led us to start digging from the layman realm to the scientific nadir and led us to finding out more than one facts about this so common a grass.

An important question to be addressed is that what makes one kind of rice so sticky and the other so non-sticky in nature. The answer lies in the presence of one of the two compounds, Amylose or Amylopectin, both being forms of starch, a primary nutrient found in rice. Amylose, an unbranched a-glucan, is responsible for non-waxy nature of Rice Cultivars while Amylopectin, a branched chain a-glucan, is responsible for the waxy nature of rice cultivars.

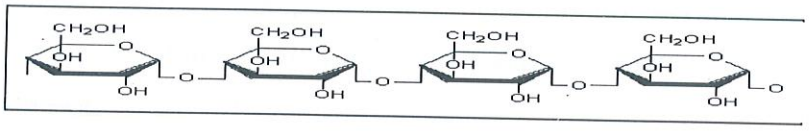


Fig 1.1, showing the molecular structure of Amylose, an unbranched a-glucan. Note the 1-4 linkage in the structure.

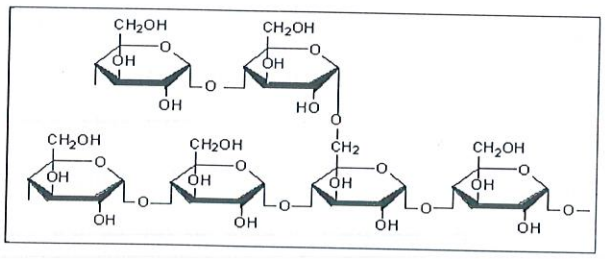


Fig 1.2, showing the molecular structure of Amylopectin, a branched a-glucan. Note the 1-6 linkage in the structure.

The greater presence of amylose in the rice granules leads to the non-waxy nature of the rice granule and thus when cooked, as in the case of basmati, leads the granule to imbibe less water. This kind of starch is more resistant to digestion and is least soluble in water.On the other hand, amylopectin, a branched chain a-glucan, imbibes more water and gives the rice (as in case of the japonica variety) a sticky nature. Amylopectin is more soluble in water than amylose and can be

digested easily.But what leads to the presence or absence of a particular kind of starch in the rice granule or what is the factor that increases the quantity of one kind of starch in the rice giving it a particular nature. The answer to this lies deeper inside the cell if we look at the process of formation of starch inside the cell, which introduces us to GBSS-1, the primary enzyme that is involved in the formation of Amylose and thus becomes the factor which must be considered when we talk about the molecular aspects of the quality of rice.

GBSS-1, Granule bound starch synthase-1, is the enzyme found in the cell that is responsible for synthesis of amylose which, as already stated, gives rice its non-waxy nature. GBSS-1 acts on the long, straight, unbranched glucose polymer, Starch and forms amylose molecules from it. The lack of GBSS-1, thus, in a molecule, will lead to the starch being converted to amylopectin and thus render the grain sticky when cooked. GBSS-1, is completely granule bound and is always found within the granule.

GBSS-1, thus becomes extremely important parameter to be considered when we talk about the quality of rice. In this study, we hypothesize that the differences in the qualities of rice could have some relation with the GBSS-1 enzyme. We propose that a mutation in the sequence could affect the structure of the enzyme GBSS-1. The comparison of sequences of different types of rice is the first step towards the study. The results of this study will provide enough information which could contribute to improve the quality of rice.

In this study we have used the sequences for GBSS-1 enzyme for a certain number of rice cultivars, in this case 19, and with the help of computational techniques like, sequence analysis, homology modelling etc, we have found out the differences in the sequences of the 19 cultivars and thus the structure of the enzyme. This differences were further analysed for their affect in the binding site of the enzyme i:e the site where the substrate (in this case starch) binds. As we analyze the difference in the binding site, we would be able to find out which point mutation (change in amino acids at a single point) would be most fundamental in changing the structure of the binding site and it's properties and thus affecting the enzyme and it's action.

To analyze the effect of mutations at the molecular level, various computational techniques were employed. Homology modelling was used to model the structure of GBSS-1 from the 19 rice cultivars. various minimization and refinement techniques like molecular dynamics simulations are employed to refine the structures. Further we have studied the interaction of the protein with starch

using molecular docking calculations to find out the affect of mutations on the binding of starch. We have investigated the mechanistic details of binding of starch with the proteins and tried to find out hot spot amino acid involved in the interaction of the enzyme GBSS-1 and starch and the implication of change of amino acids , if any.

CHAPTER 2

METHODOLOGY

KEGG Database

In our study, to understand the importance of GBSS-1 protein in the starch synthesis pathway in *oryza sativa,* we have used KEGG Database. **KEGG (Kyoto Encyclopedia of Genes and Genomes)** is a bioinformatics resource for understanding higher-order functional meanings and utilities of the cell or the organism from its genome information. It is an integrated resource consisting of three types of databases for genomic, chemical and network information, and associated software, which are all developed by the **Kanehisa Laboratory** (now part of the Bioinformatics Center) in the Institute for Chemical Research, Kyoto University .KEGG is a database of biological systems, consisting of genetic building blocks of genes and proteins (KEGG GENES), chemical building blocks of both endogenous and exogenous substances (KEGG LIGAND), molecular wiring diagrams of interaction and reaction networks (**KEGG PATHWAY**), and hierarchies and relationships of various biological objects (KEGG BRITE). KEGG provides a reference knowledge base for linking genomes to biological systems and also to environments by the processes. There are approximately 100 reference metabolic pathway maps in KEGG, each showing a biochemical pathway such as glycolysis or peptidoglycan biosynthesis. KEGG also contains organism-specific pathway maps for each sequenced organism.

Modelling the Template

NCBI

Reference protein sequence for GBSS1 enzyme in *oryza sativa* subsp. indica was extracted out using NCBI. (gi|25992195|).Since the structure of GBSS1 enzyme in *oryza sativa* is not available, the sequence was compared with the entire PDB database ,using **Psi BLAST**, to find out the most suitable template for modelling the structure of GBSS1 protein. **Psi BLAST** is an algorithm that is used for Protein BLAST. It stands for **Position Specific Iterated BLAST**. It derives a position-specific scoring

matrix (PSSM) or profile from the multiple sequence alignment of sequences detected above a given score threshold using protein–protein BLAST. This PSSM is used to further search the database for new matches, and is updated for subsequent iterations with these newly detected sequences.Thus, PSI-BLAST provides a means of detecting distant relationships between proteins.

The template structure was found to have missing residues which were added using multiple template based homology modelling technique in **Accelerys' Discovery Studio Client 3.5.** the structure was further minimized using Charm M forcefield. The overall quality of the structure, stereochemical values and non-bonded interactions were tested using PROCHECK [4], ERRAT [5] and VERIFY3D [6].

The minimized template was further used to model the structure of the refrence protein. The structure obtained was energy minimized using OPLS 2005 force field with Polak-Ribiere Conjugate Gradient (PRCG) algorithm. The minimization was stopped either after 5,000 steps or after the energy gradient converged below 0.001 kcal/mol. All atom molecular dynamics (MD) simulation of protein structure in explicit water was carried out using the GROMACS 4.5.4 software.

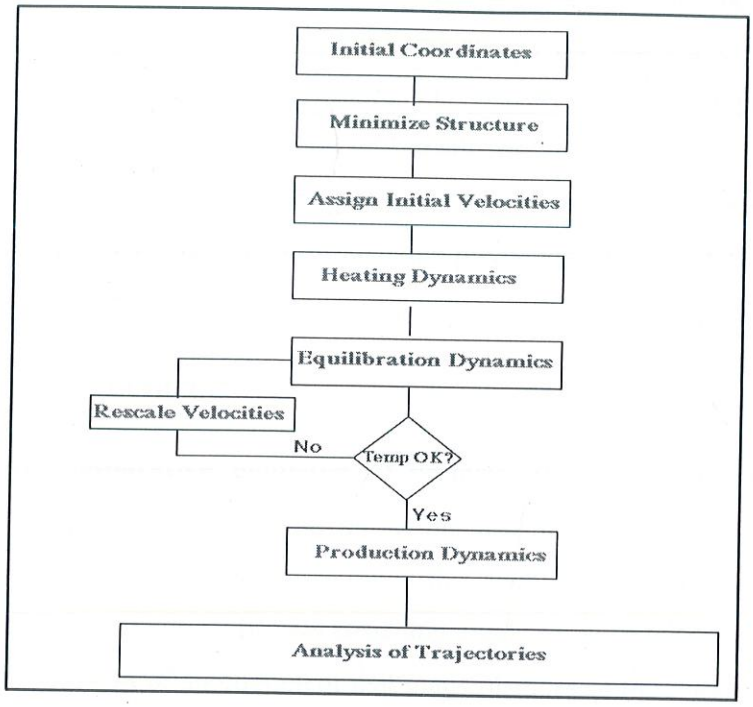
Molecular dynamics (MD) is a computer simulation of physical movements of atoms and molecules. The atoms and molecules are allowed to interact for a period of time, giving a view of the motion of the atoms. In the most common version, the trajectories of atoms and molecules are determined by numerically solving the Newton's equations of motion for a system of interacting particles, where forces between the particles and potential energy are defined by molecular mechanics force fields.

Because molecular systems consist of a vast number of particles, it is impossible to find the properties of such complex systems analytically; MD simulation circumvents this problem by using numerical methods. However, long MD simulations are mathematically ill-conditioned, generating cumulative errors in numerical integration that can be minimized with proper selection of algorithms and parameters, but not eliminated entirely.

For conducting MD simulation of GBSS1, we used the **GROMACS package**. **GROMACS (GROningen MAchine for Chemical Simulations)** was primarily designed for simulations of proteins, lipids and nucleic acids. GROMACS is extremely fast compared to other software packages due to its optimization. GROMACS is operated via the command-line, and can use files for input and output. Once a configuration file for the simulation of several molecules (possibly including solvent) has been created, the actual simulation run (which can be time consuming) produces

a trajectory file, describing the movements of the atoms over time. This trajectory file can then be analyzed or visualized with a number of supplied tools.

The various steps of Molecular Dynamics Simulation are:



We have used theGROMOS96 force field for a time scale of 10 ns. Three-dimensional periodic boundary conditions were imposed, enclosing the molecule in a dodecahedron solvated with the SPC216 water model provided in the GROMACS package and energy minimized using 1000 steps of steepest descent. The system was neutralized with 18 Na+ counter ion and was locally minimized using 100 steps of steepest descent. The electrostatic term was described using the Particle Mesh Ewald algorithm [2]. The LINCS [3] algorithm was used to constrain all bond lengths and cut-off distances for the calculation of the coulombic and van der Waals interactions at 1.0 nm. The system was equilibrated by100ps of MD runs with position restraints on the protein to allow the relaxation of the solvent molecules at 300 K and normal pressure. The system was coupled to the external bath by the Berendsen thermostat with a coupling time of 0.1 ps with default setting. The final MD calculations were performed for 1.0ns under the same conditions, except the position restraints were removed. The overall quality of the model, stereo-chemical values and non-bonded interactions were tested using PROCHECK, ERRAT and VERIFY3D.

Sequence assembly and analysis

As already stated, our study consisted of 19 cultivars of rice and thus, the starting material for our study consisted of 38 sequences of primers, 19 reverse primers and 19 forward primers. We used **EMBOSS** online tool in order to assemble those sequences and translate them into protein sequences for modelling. The results of Emboss gave us a total of 3 results per sequence shifting one frame per result (Frame 1 starts from position 1 of the sequence, Frame 2 starts from the second position). We chose the longest frame because it is considered the most accurate prediction of translation.Sequence alignment of each sequence with the original GBSS-1 model sequence was done with the help of **CLUSTALW**.

Modelling of Protein structures:

Also known as **comparative modeling** of proteins, it refers to constructing an atomic-resolution model of the "*target*" protein from its amino acid sequence and an experimental three-dimensional structure of a related homologous protein (the "*template*"). Homology modeling relies on the identification of one or more known protein structures likely to resemble the structure of the query sequence, and on the production of an alignment that maps residues in the query sequence to residues in the template sequence. It has been shown that protein structures are more conserved than protein sequences amongst homologues, but sequences falling below a 25-30% sequence identity can have very different structure. Homology modeling is a multistep process and it can be summarized in following steps:

1. Template recognition and initial alignment
2. Alignment correction
3. Backbone generation
4. Loop modeling
5. Side-chain modeling
6. Model optimization
7. Model validation

We used **MOE (Molecular Operating Environment)** for aligning and modelling the all the 19 using the already prepared template structure of GBSS-1. All the structures obtained was energy minimized using OPLS 2005 force field with Polak-Ribiere Conjugate Gradient (PRCG) algorithm. The minimization was stopped either after 5,000 steps or after the energy gradient converged below 0.001 kcal/mol.

Validation of protein structures.

Protein structures might contain errors independently if they have been created from experimental information or from modeling. Therefore it is necessary to have some possibility to check for errors in a given model independent on how this model was created. To verify the model generated, we used SAVES server, in which we used three programmes PROCHECK, ERRAT, VERIFY_3D.

The use of these programs gives us the results in the form of a Ramachandran Plot (PROCHECK) and graphs (ERRAT, VERIFY 3D).

Ramachandran plot (also known as a **Ramachandran diagram** or a **[φ,ψ] plot**), originally developed in 1963 by G. N. Ramachandran, C. Ramakrishnan, and V. Sasisekharan, is a way to visualize backbone dihedral angles ψ against φ of amino acid residues in protein structure.

The Ramachandran plot[7] is the 2d plot of the φ-ψ torsion angles of the protein backbone. It provides a simple view of the conformation of a protein. The φ-ψ angles cluster into distinct regions in the Ramachandran plot where each region corresponds to a particular secondary structure. There are four basic types of Ramachandran plots, depending on the stereo-chemistry of the amino acid: generic (which refers to the 18 non-glycine non-proline amino acids), glycine, proline, and pre-proline (which refers to residues preceding a proline. The generic and proline Ramachandran plots are now well understood but the glycine and pre-proline Ramachandran plots are not.

ERRAT_is a program for verifying protein structures determined by crystallography. Error values are plotted as a function of the position of a sliding 9-residue window. The error function is based on the

statistics of non-bonded atom-atom interactions in the reported structure (compared to a database of reliable high-resolution structures). The ERRAT score i.e. the overall quality factor, is considered to be good if above 80%. The program works by analyzing the statistics of non-bonded interactions between different atom types. A **single** output plot is produced that gives the value of the error function *vs.* position of a 9-residue sliding window. By comparisons with statistics from highly refined structures, the error values have been calibrated to give **confidence** limits. This is extremely useful in making decisions about reliability.

Verify3D: The Verify3D (*Eisenberg et al.*) method assess protein structures using three-dimensional profiles. This program analyzes the compatibility of an atomic model (3D) with its own amino acid sequence (1D). Each residue is assigned a structural class based on its location and environment (alpha, beta, loop, polar, apolar etc). Then a database generated from good structures is used to obtain a score for each of the 20 amino acids in this structural class. The vertical axis in the plot represents the average 3D-1D profile score for each residues in a 21-residue sliding window. The scores ranges from -1 (bad score) to +1 (good score).

Active site Prediction

The active site of the proteins was found out through the use of Schrodinger Maestro's Sitemap tool with default parameters which was instructed to give top 5 regions as the areas where the active site of the protein can be present per structure. This software generates information on the binding site's characteristics using novel search and analytical facilities: a **SiteMap** calculation begins with an initial search step that identifies or characterizes—through the use of grid points— one or more regions on the protein surface that may be suitable for binding ligands to the receptor. To select a site the site_score should be 1 and the site should also include all the important interacting residues.

Ligand preparation: Molecular structures of starch (compounds 1–6) was built using molecular builder of Maestro (version 8.5). The structure was energy minimized in vaccum using Impact (version 5.6, Schro¨dinger, LLC). Each structure was assigned an appropriate bond order using\ Ligprep (version 2.4, Schro¨dinger LLC).Ligprep is a tool in Schrodinger suite used to generate the 3D structures from 2D representation, searching for tautomers, steric isomers, ionizations states using Epik followed by geometry minimization of ligands based upon OPLS 2005 force field. Out of these orientations, the most stable ones for both the ligands are taken.

Grid Generation: The shape and properties of the receptor are represented on a grid by several different sets of fields that provide progressively more accurate scoring of the ligand pose. The binding site is defined in terms of two concentric cubes: the bounding box, which must contain the mass center of any acceptable ligand pose, and the enclosing box, which must contain all the atoms of a ligand pose for successful docking into the binding site.

On the basis of active site decided according to the analysis of SiteMap results, important residues were selected for grid generation. The grid was generated on the basis of analysed active site and residues within the radius of 5 A°. In this work the bounding box of size 10 °A ×10 °A ×10 °A was defined in tubulin and centered on the residue information from the sitemap analysis. The residue numbers selected for grid generation are:

1, 281,283,284,286,287,288,289,290,298,299,300,421,422,423,424,489,490,491,492,493,494,495,497

The larger enclosing box with an edge length of 10 °A was also defined.

DOCKING

Docking is a procedure to predict the preferable binding orientation between the two molecules to form the stable complex.

All docking calculations were performed using the "Standard Precision (SP) and refined using "Extra Precision" (XP) mode of Glide docking [25,26] (version 4.5, Schrodinger Inc.) with the 2005 implementation of the OPLS-AA force field. XP provides the most accurately docked poses. As we already knew the most probable ligand we took only the most stable conformation of the most probable ligand and docked it into the receptor using XP. The detailed algorithm of Glide docking has been described previously [25,26]. Briefly, Glide approximates a systematic search of positions, orientations, and conformations of the ligand in the receptor binding site using a series of hierarchical filters. The shape and properties of the receptor are represented on a grid by several different sets of fields that provide progressively more accurate scoring of the ligand pose. The binding site is defined in terms of two concentric cubes: the bounding box, which must contain the mass center of any acceptable ligand pose, and the enclosing box, which must contain all the atoms of a ligand pose for successful docking into the binding site. Glide also performed conformational searches for each input structure during docking process. A set of initial ligand conformations is generated through exhaustive

search of the torsional minima and the conformers are clustered in a combinatorial fashion. Each cluster, characterised by a common conformation of the core and an exhaustive set of side-chain conformations, is docked as a single object in the first stage. The search begins with a rough positioning and scoring phase that significantly narrows the search space and reduces the number of poses to be further considered to a few hundred. These selected poses are energy minimized on pre-computed OPLS-AA van der Waals and electrostatic grids for the receptor. In the final stage, the 5–10 lowest-energy poses obtained in this fashion are subjected to a Monte Carlo sampling in which nearby torsional minima are examined, and the orientation of peripheral groups of the ligand is refined. The minimized poses are then rescored using the GlideScore function.

CHAPTER 3

RESULTS

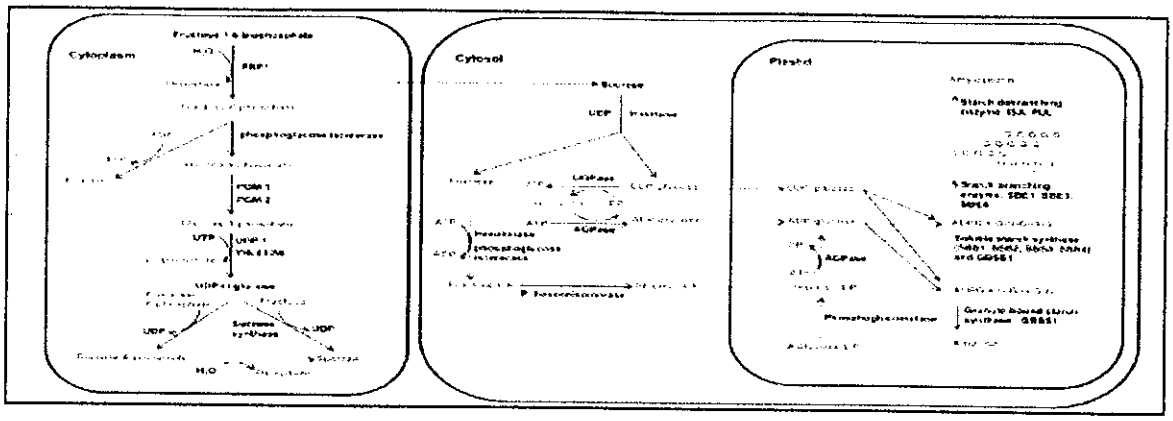


Fig 3.1, showing the amylose and amylopectin synthesis pathway.

The figure given above shows that GBSS-1 is an important enzyme in the production of either forms of starch. The site of its reaction in the pathway can be seen from the figure obtained from KEGG which shows that GBSS-1 forms amylose from a long, straight chain of glucose molecules.

Modelling the template

The template we required for modelling the 19 structures from their sequences was extracted by us from PDB (GBSS-1 (Reference sequence, gi|25992195|) and template sequence as obtained from PDB (ID:3VUE) Before carrying out the modelling, it is necessary to find out if the structure is as accurate according to the sequence provided to us. Because in certain cases gaps might be present in the original structure while the sequence might be complete. This problem was faced by us when we modelled the structure and we found a gap that too distinctly in the region of the active site which is present from residue number 182 to 609. Notably, the gap was present from residue number 173 to 192. The figure given below shows the structure as obtained by us from MOE containing the gap that has been shown and the residues have been marked.

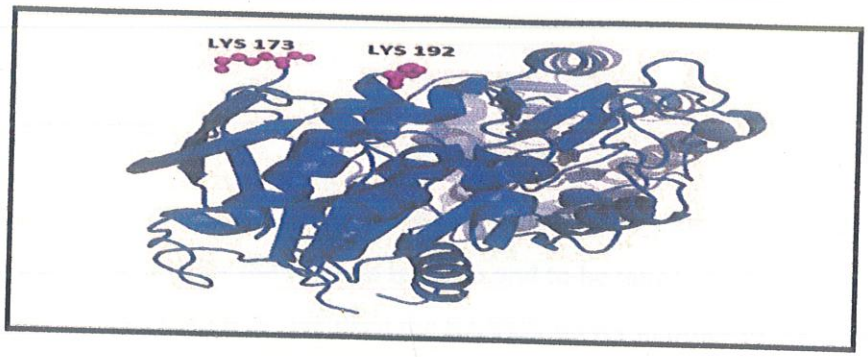


Fig 3.2: giving the pictorial representation of the gaps found in the structure of reference protein GBSS-1. As it is clearly visible the gap is present between residue numbers 173 to 192.

Crystal Structure Of Rice Granule Bound Starch Synthase I Catalytic Domain (PDB id: 3VUE; Res: 2.70) was found with 86% query coverage and 99% identity with GBSS1. It's a single chained protein with length 536 and having catalytic domain, UNP residues 83-609. As a good homologue was available for GBSS1 enzyme, homology modelling was done for building mode

The gap of residues caused a substantial challenge to the modelling of the protein as the gap was present in the region corresponding to the catalytic domain of GBSS-1. As the gap was too long (18 residues), it was not feasible to use a gap filling software as software would not be able to provide accurate results for such a long gap of residues. In order to avoid inaccuracy in the resulting structure, the modelling of GBSS-1 was done using multiple templates. The two additional templates along with the 3VUE, were used to model the complete structure.

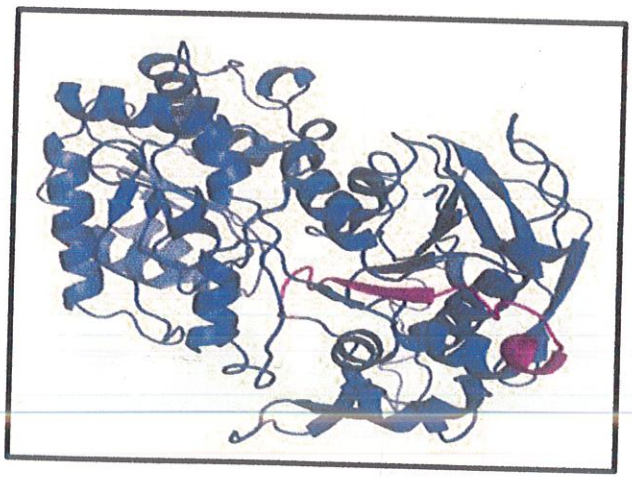
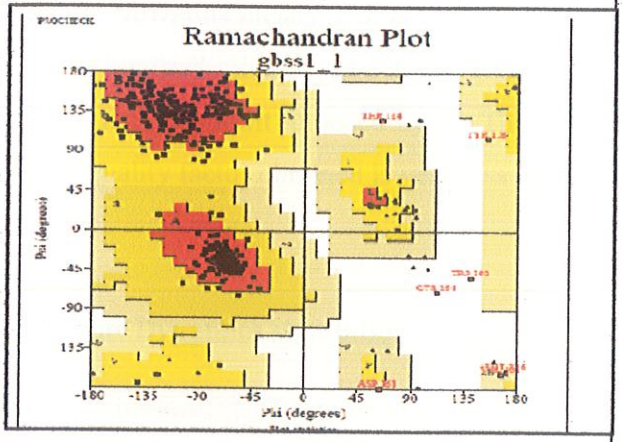


Fig 3.3: showing the modelled structure of GBSS-1, using multiple templates

Protein Validation

Once the protein is created, it was important for us to validate the protein and find out if the protein that we have modelled is feasible for the work to be done and to be taken as a template. Hence, as already discussed under METHODOLOGY, we used the SAVES server in order to carry out protein validation. The results are given as under.

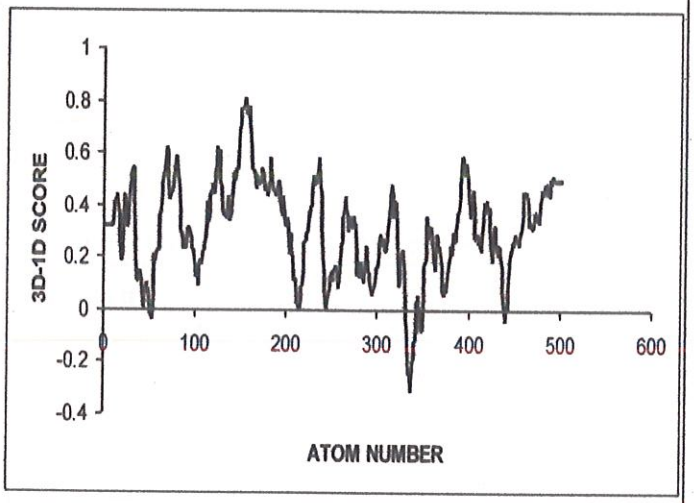
RAMACHANDRAN PLOT



Residues in most favored regions	90.1%
Residues in additional allowed regions	8.3%
Residues in disallowed regions	0.7%

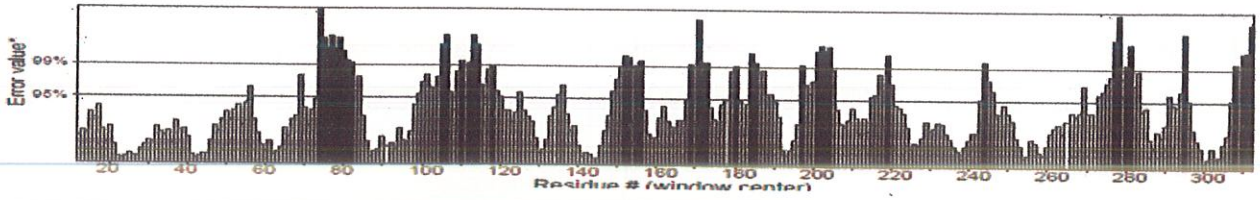
(A)

VERIFY_3D



(B)

ERRAT



(C)

Fig: (A) showing the Ramachandran plot for GBSS-1 as generated by SAVES (PROCHECK), **(B)** showing the graph generated by VERIFY_3D showing 88.17% of the residues had an averaged 3D-1D score > 0.2, **(C)** showing the graph generated by ERRAT showing an Overall quality factor 67.606

The Ramachandran Plot generated by Procheck gave us 90.1% of the residues in the most favoured regions (Regions in red are the most favoured region). Even though this is a good percentage of amino acids, we need the other analysis for further information.

The ERRAT score i.e. the overall quality factor, is considered to be good if above 80%. The program works by analyzing the statistics of non-bonded interactions between different atom types. A single output plot is produced that gives the value of the error function *vs.* position of a 9-residue sliding window. Now, this means, that according to our plot, only the amino acids indicated by the black bars pass the required criteria of having a good quality i:e having the structure where the confidence limit exceeds the 99% limit imposed by us. This comes out to be around 67.606 quality factor. Clearly the overall quality factor obtained is very less and hence indicating the structure to be less reliable.

Molecular Dynamics simulation results

After we performed Molecular dynamics upon the structures with the help of Gromacs, we found a substantial change in the quality of the structures. The results are as follows:

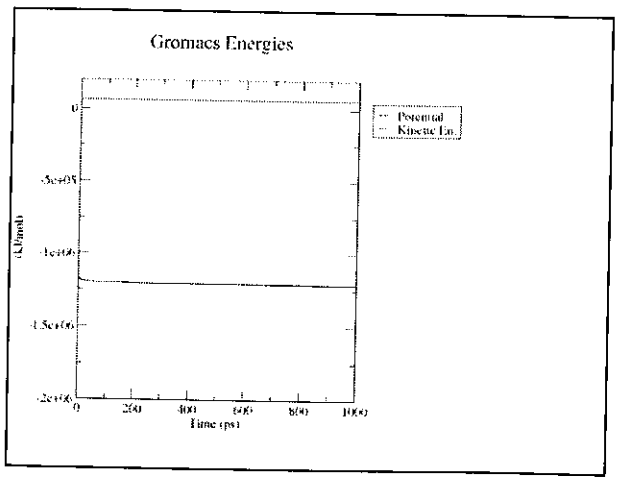


Fig 3.7: potential energy and kinetic energy graph against time

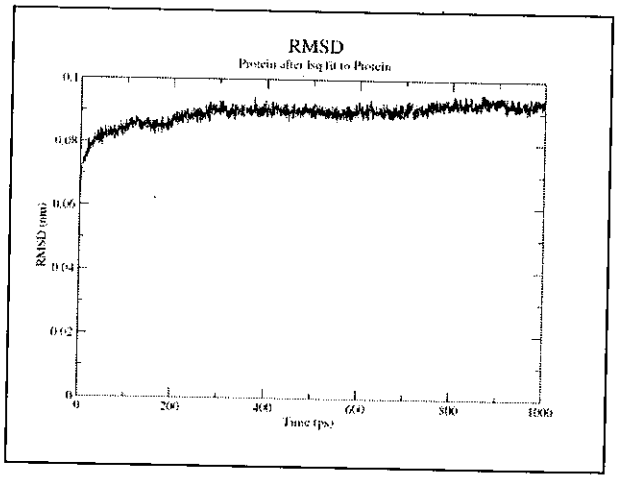


Fig 3.8: showing RMSD plot obtained

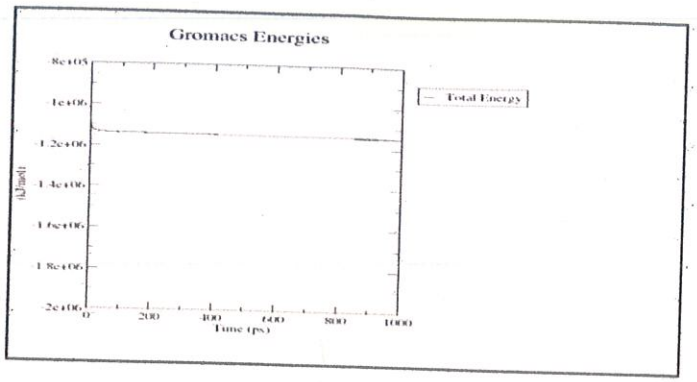


Fig 3.9: showing total energy plot against time

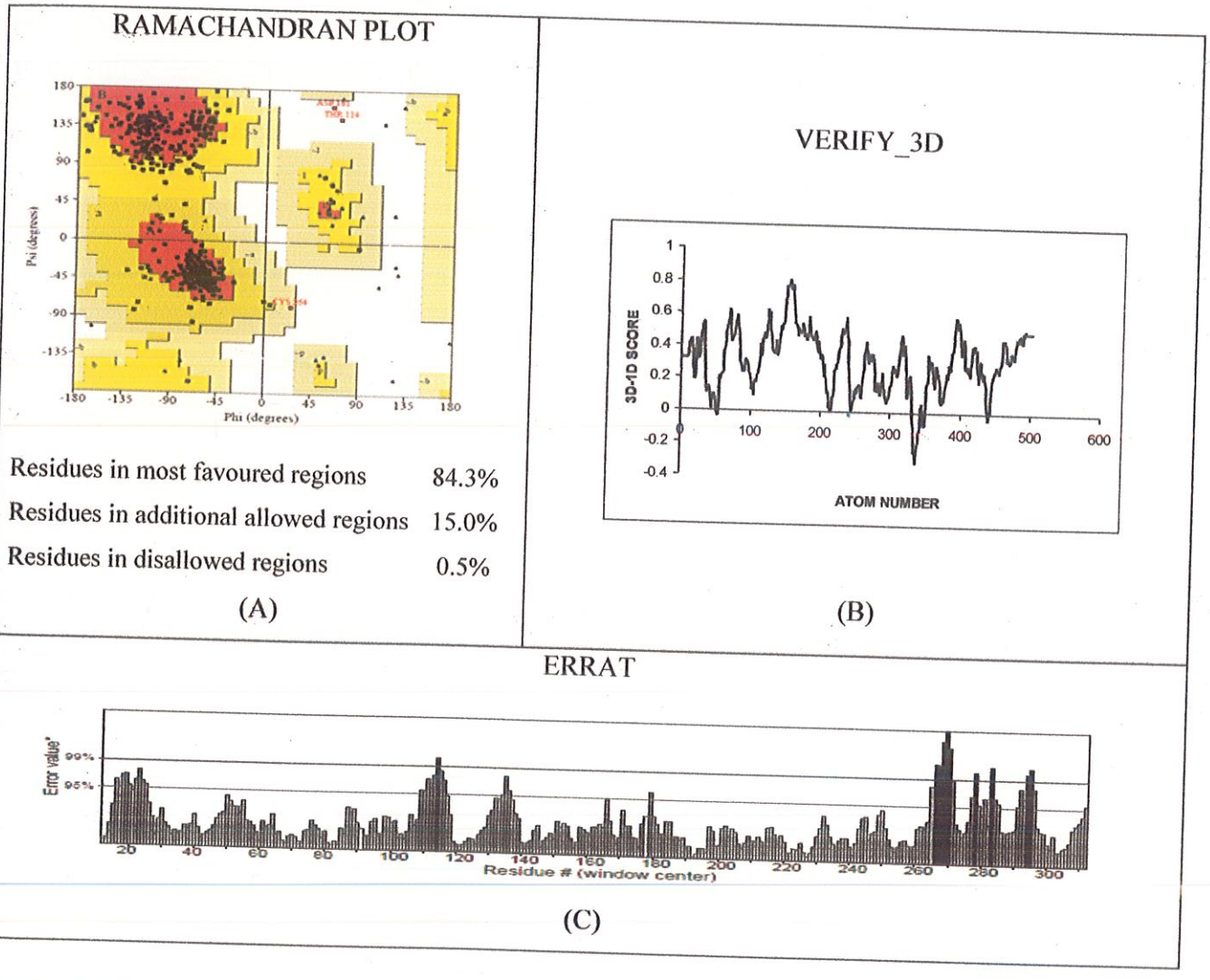


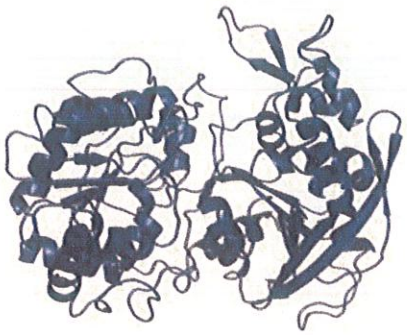
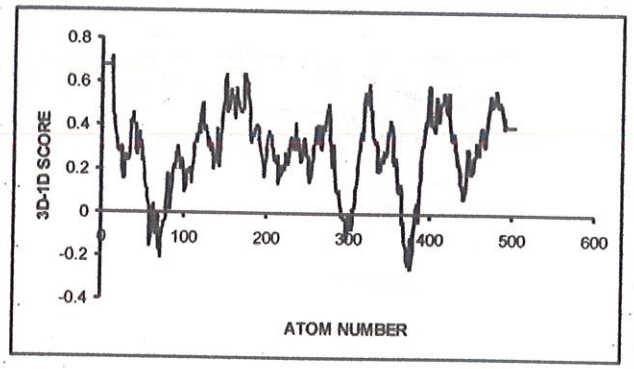
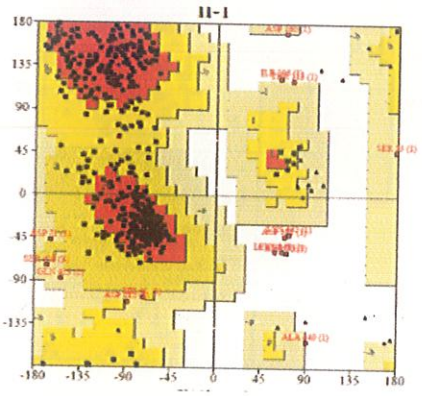
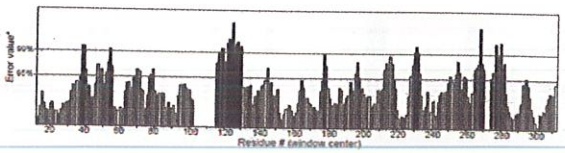
FIG:(A): Depicting Ramachandran Plot for the template protein. **(B)** showing the graph generated by

VERIFY_3D showing 95.07% of the residues had an averaged 3D-1D score > 0.2 . **(C)** graph generated by VERIFY_3D showing 95.07% of the residues had an averaged 3D-1D score > 0.2

A clear contrast was seen in the results obtained after molecular dynamics. The ERRAT score improved significantly to 88.726 and a huge improvement was also seen in the Verify 3D scores. This improvement in the scores makes the structure more reliable.

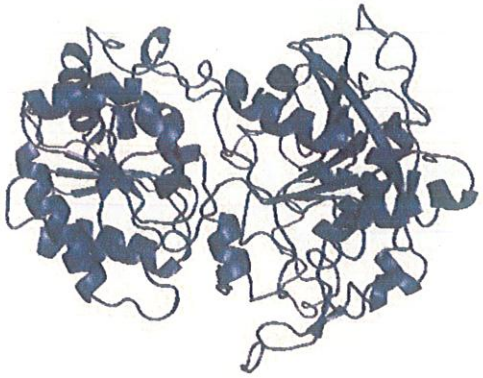
Modelling of GBSS-1 for other cultivars

Once our template was procured, modelled, minimized and validated, our next job was to model the sequences of the rest of the 19 cultivars. We followed a similar procedure for the modelling, carrying out modelling in MOE and then minimizing the structures and obtaining the ERRAT, VERIFY 3D and PROCHECK scores. The results in the form of tables for each protein are as follows:

II-1	
Structure of II-1(GBSS1 protein) 	**Structure of II-1(blue) superimposed with reference(yellow) is the having RMSD 0.902 Å** 
75.64% of atoms have 3D-ID score>0.2 	**Ramachandran plot** 
Errat score 77.149 	Residues in most favored regions = 79.8% Residues in generously allowed regions= 1.3% Residues in disallowed regions =2.0%

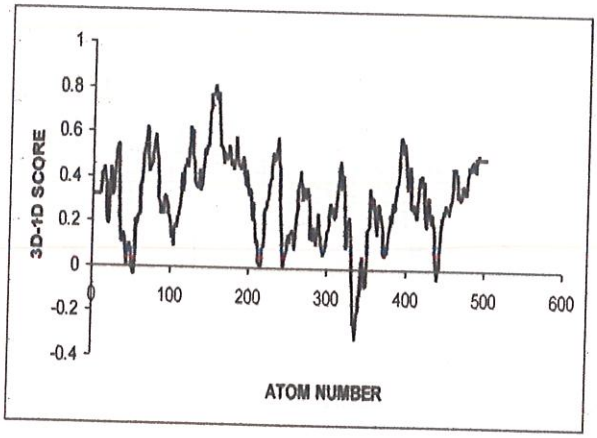
II-5

Structure of II-5(GBSS1 protein)

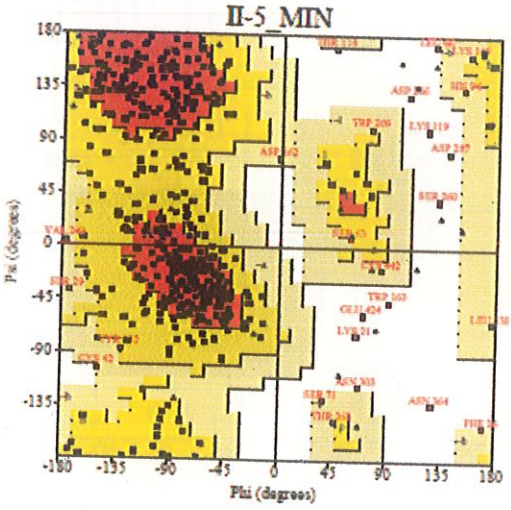


Structure of II-5(blue) superimposed with reference(yellow) is the having RMSD 1.076 Å

72.87% of atoms have 3D-ID score>0.2



Ramachandran plot

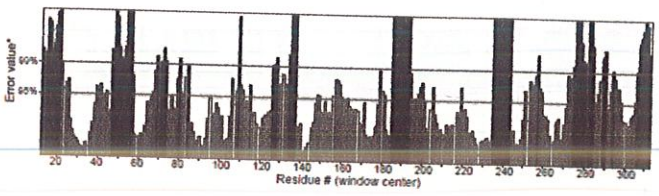


Residues in most favored regions 61.7%

Residues in generously allowed regions 3.1%

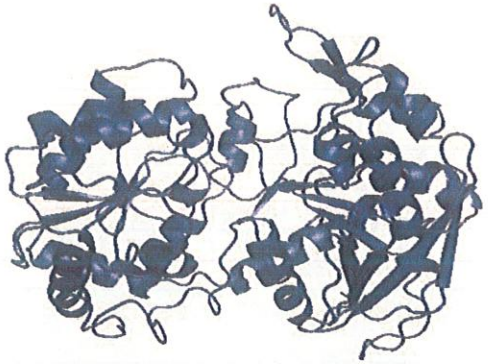
Residues in disallowed regions 2.4%

Errat score 59.467

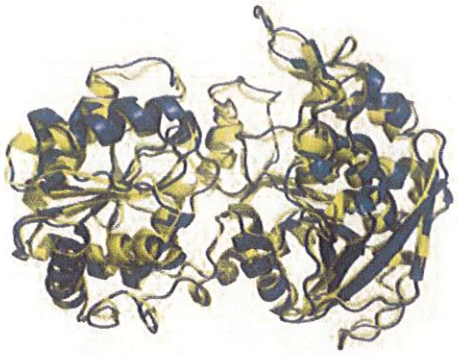


II-6

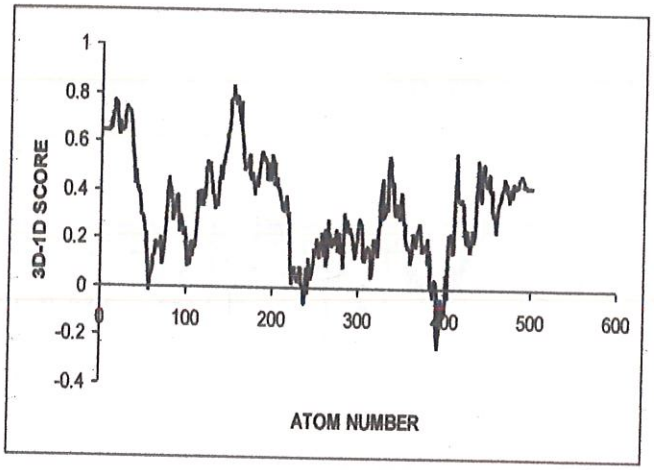
Structure of II-6(GBSS1 protein)



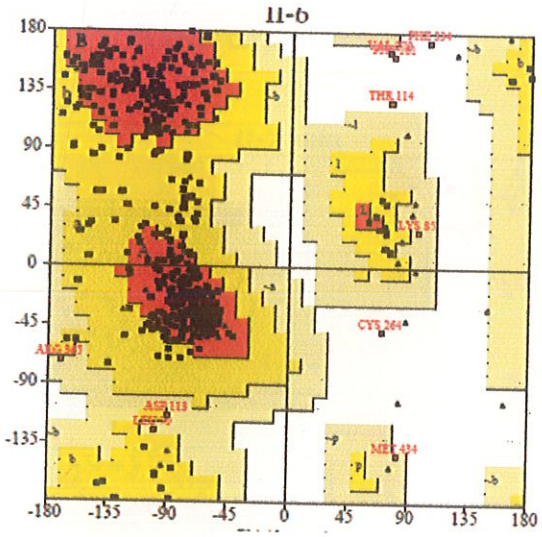
Structure of II-6(blue) superimposed with reference(yellow) is the having RMSD 0.926Å



69.17% of atoms have 3D-ID score>0.2



Ramachandran plot

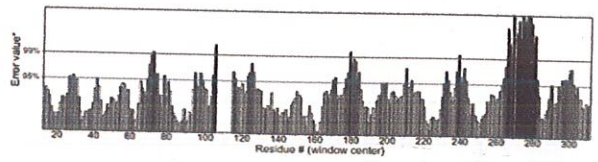


Residues in most favored regions 77.3%

Residues in additional allowed regions 20.4%

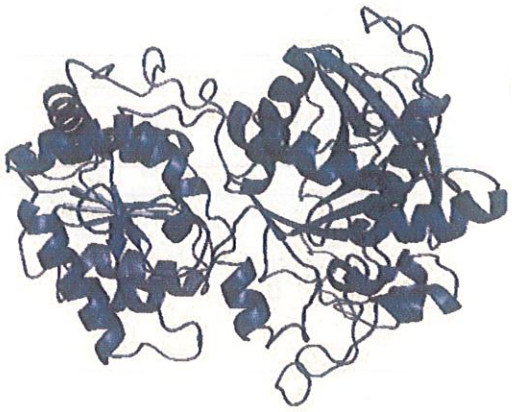
Residues in disallowed regions 1.1%

Errat score 77.5



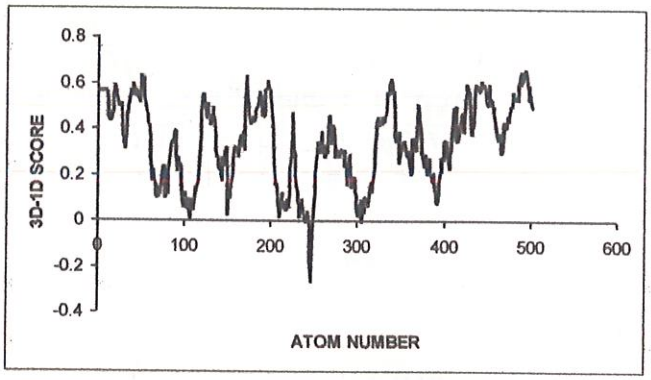
II-7

Structure of II-7(GBSS1 protein)

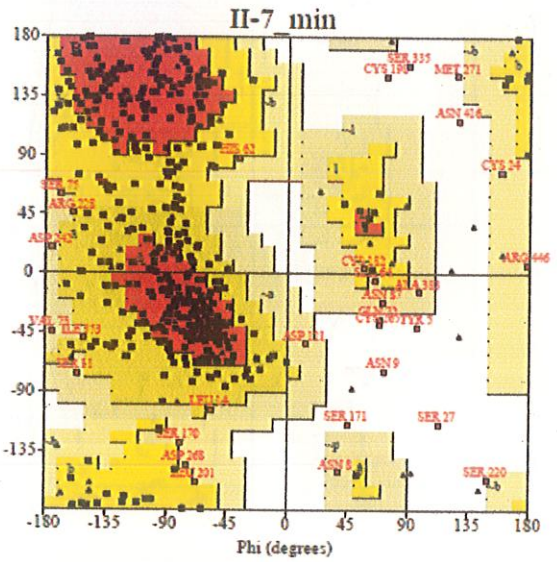


Structure of II-7(blue) superimposed with reference(yellow) is the having RMSD Å

79.18% of atoms have 3D-ID score>0.2



Ramachandran plot

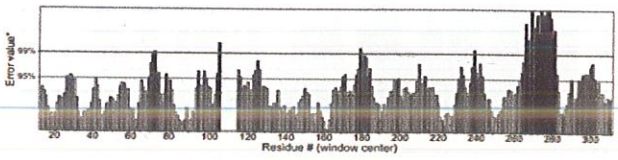


Residues in most favored regions = 59.8%

Residues in additional allowed regions = 33.6%

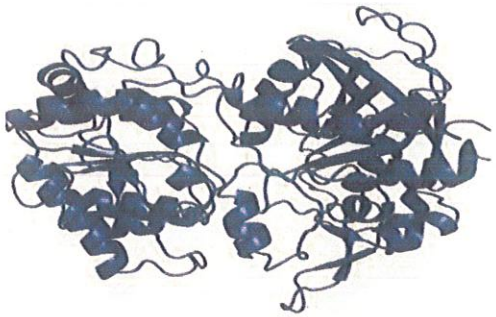
Residues in disallowed regions= 2.4%

Errat score 76.832

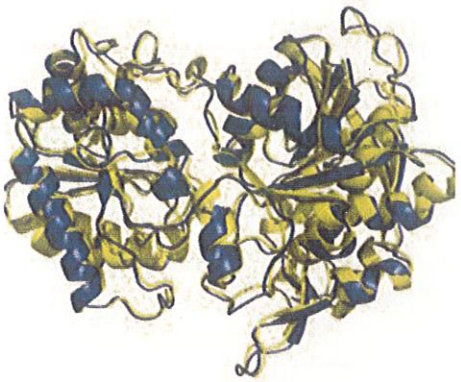


II-8

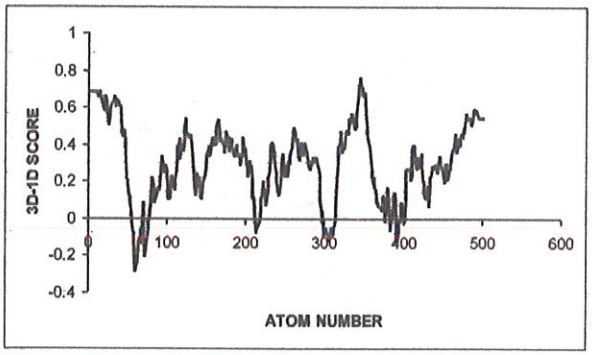
Structure of II-8(GBSS1 protein)



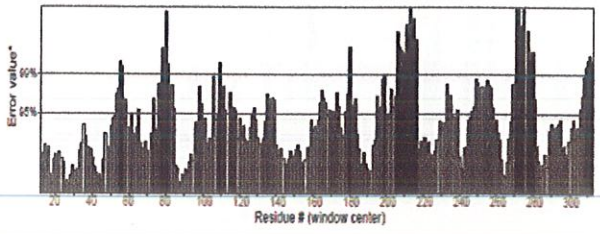
Structure of II-8(blue) superimposed with reference(yellow) is the having RMSD 0.911Å



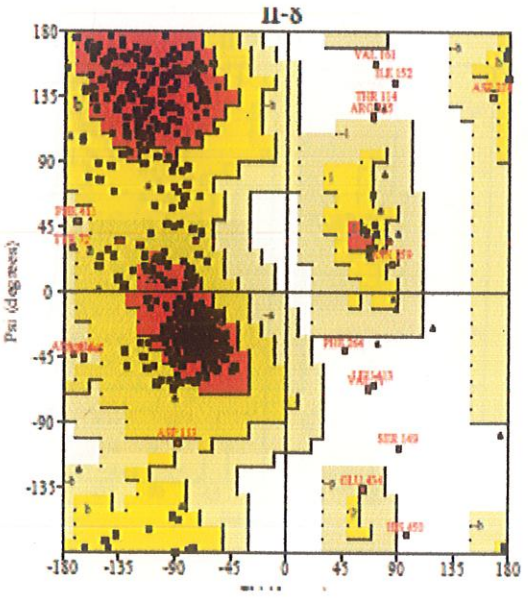
70.16% of atoms have 3D-ID score>0.2



Errat score 65.849



Ramachandran plot



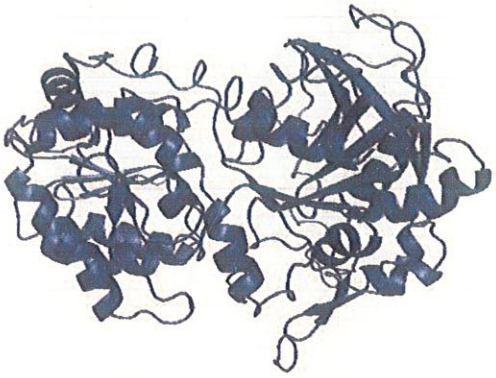
Residues in most favored regions =76.3%

Residues in additional allowed regions= 20.0%

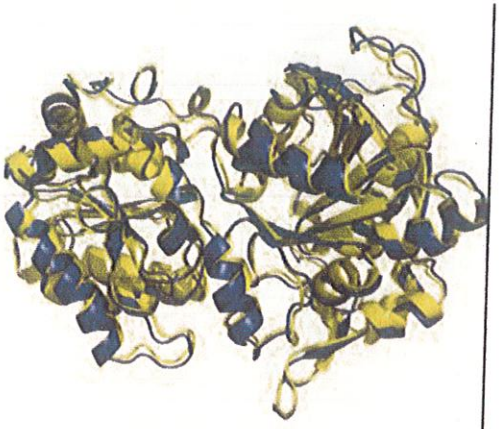
Residues in disallowed regions =2.0%

II-9

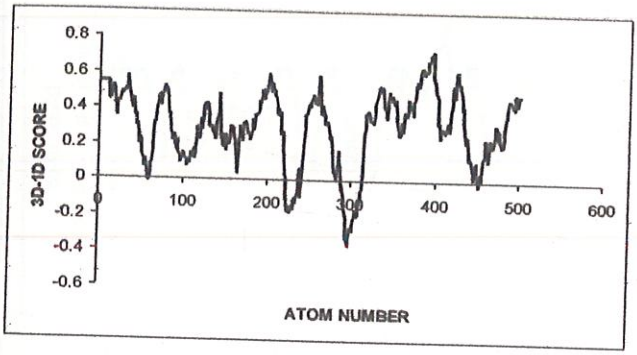
Structure of II-9(GBSS1 protein)



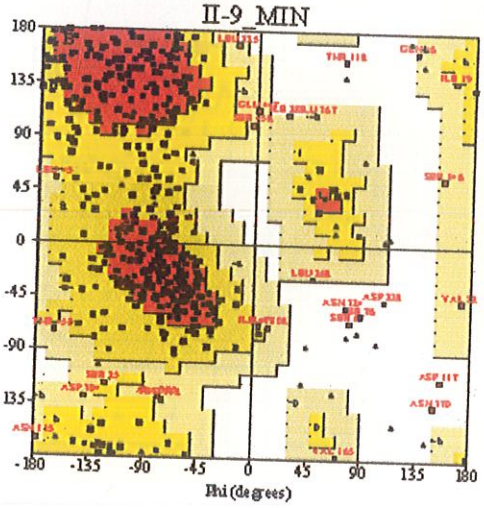
Structure of II-9(blue) superimposed with reference(yellow) is the having RMSD 0.968Å



75.39% of atoms have 3D-ID score>0.2



Ramachandran plot

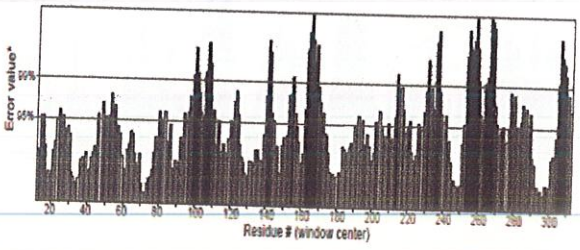


Residues in most favored regions = 59.5%

Residues in additional allowed regions = 34.8%

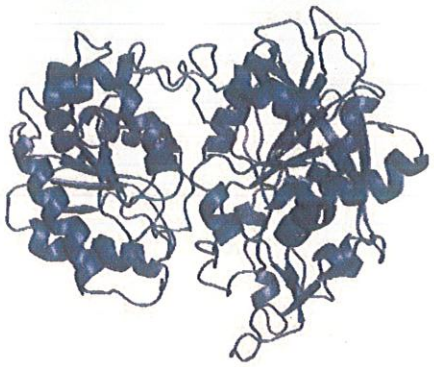
Residues in disallowed regions= 1.5%

Errat score 63.02

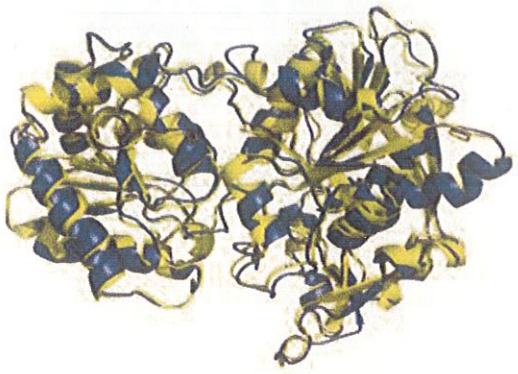


III-1

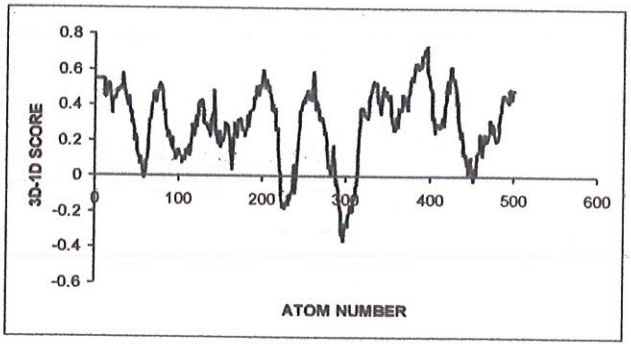
Structure of III-1(GBSS1 protein)



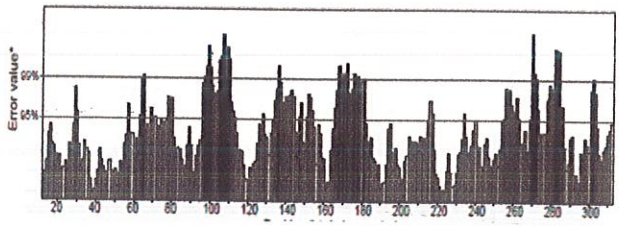
Structure of III-1(blue) superimposed with reference(yellow) is the having RMSD 1.018 Å



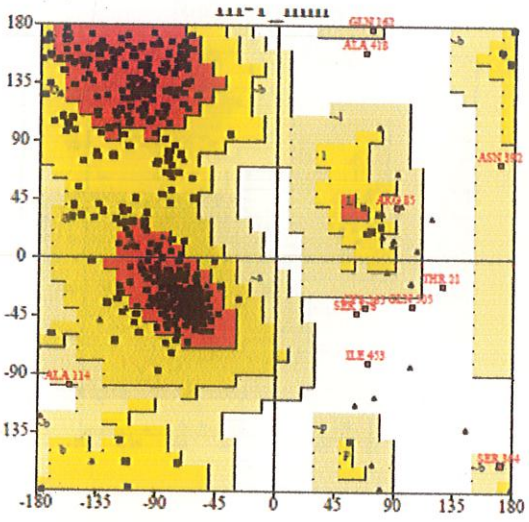
73.63% of atoms have 3D-ID score>0.2



Errat score 68.93



Ramachandran plot



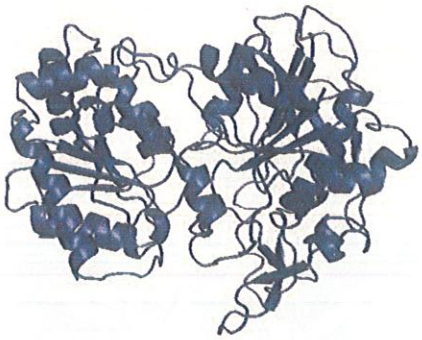
Residues in most favored regions= 77.3%

Residues in additional allowed regions= 20.2%

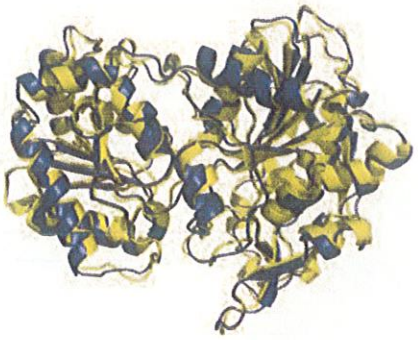
Residues in disallowed regions= 1.4%

III-2

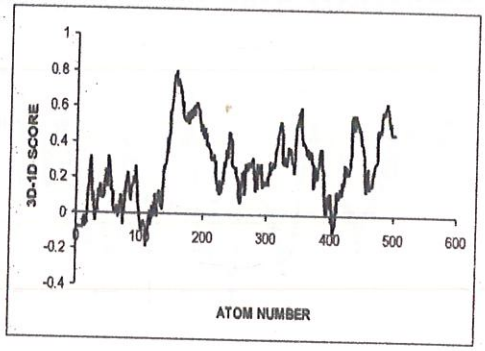
Structure of III-2(GBSS1 protein)



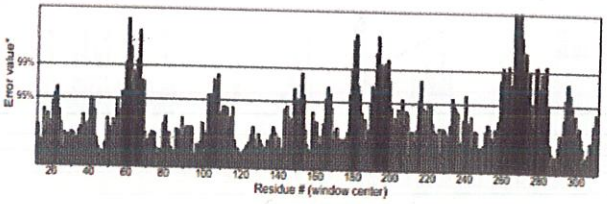
Structure of III-1 superimposed with Template. RMSD 1.036Å



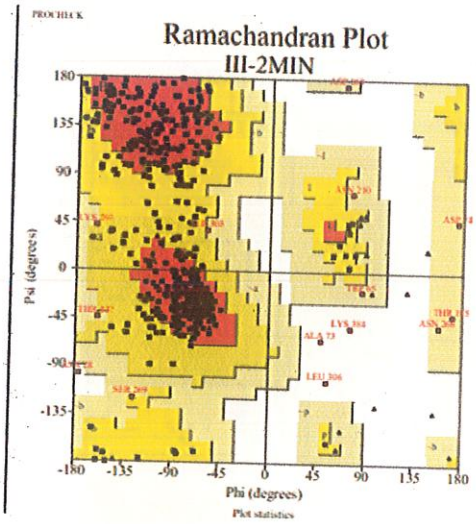
60.83% of atoms have 3D-ID score>0.2



Errat score 72.94



Ramachandran plot



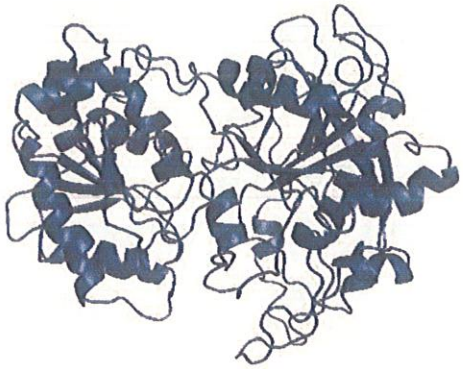
Residues in most favored regions = 73.1%

Residues in additional allowed regions = 23.8%

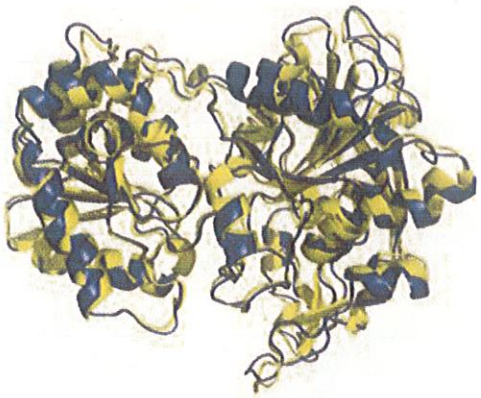
Residues in disallowed regions = 0.9%

III-7

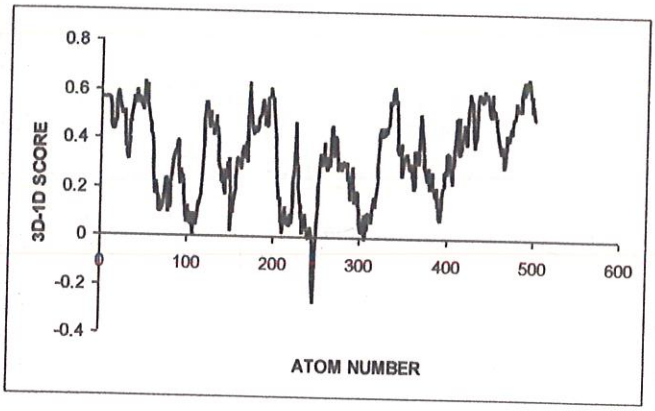
Structure of III-7(GBSS1 protein)



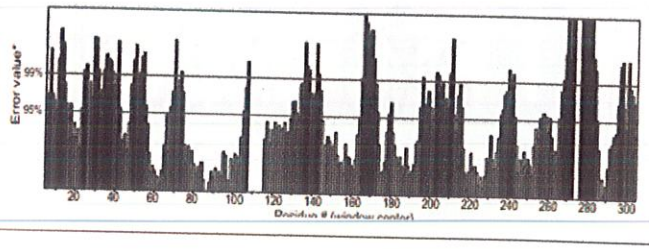
Structure of III-1 superimposed with Template. RMSD 1.108Å



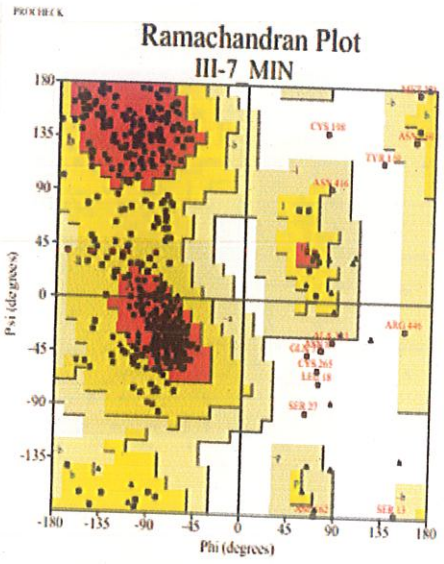
76.61% of atoms have 3D-ID score>0.2



Errat score 57.75



Ramachandran plot



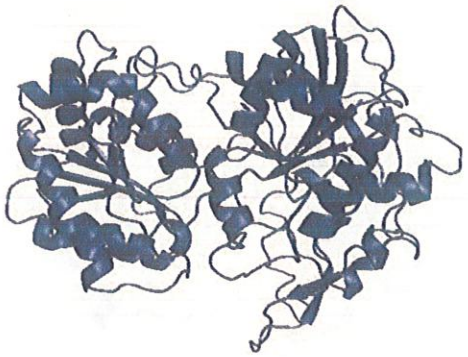
Residues in most favored regions =73.3%

Residues in additional allowed regions= 23.6%

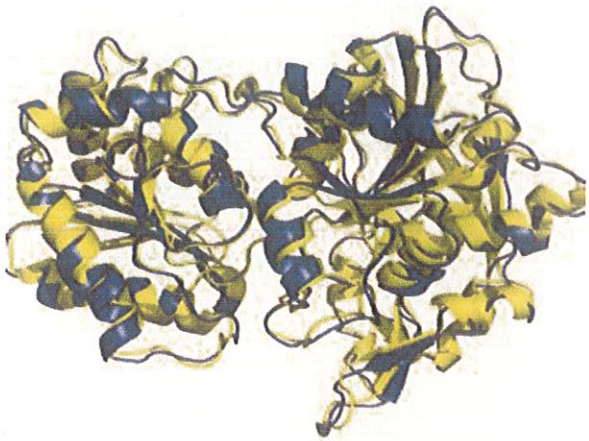
Residues in disallowed regions = 2.0%

III-8

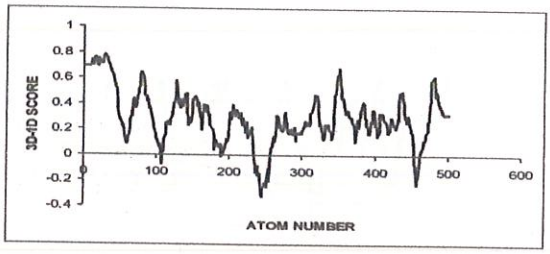
Structure of III-8(GBSS1 protein)



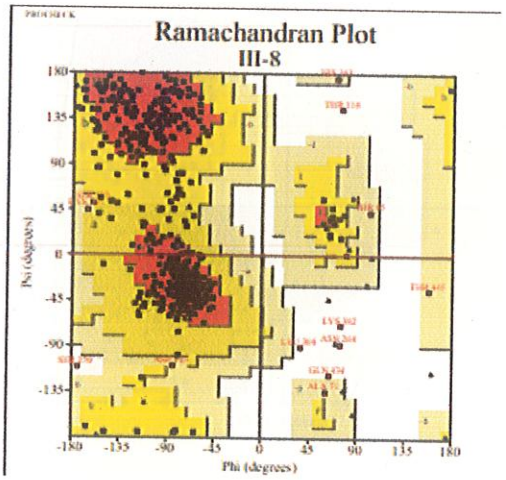
Structure of III-1 superimposed with Template. RMSD 1.015 Å



70.55% of atoms have 3D-ID score>0.2



Ramachandran plot

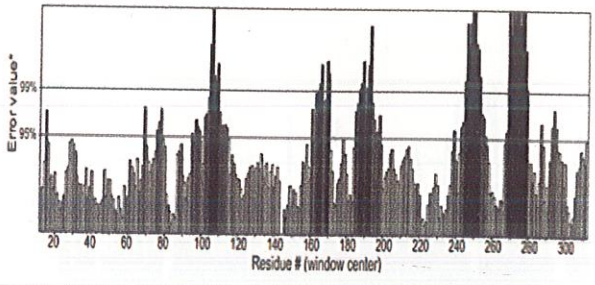


Residues in most favored regions =78.2%

Residues in additional allowed regions =18.9%

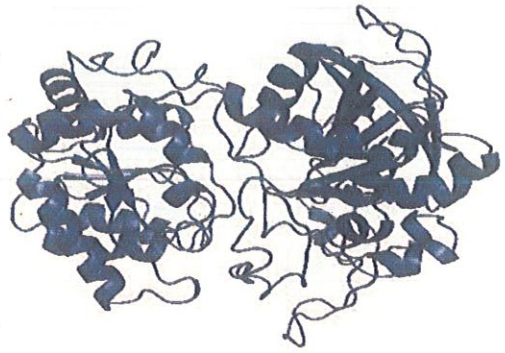
Residues in disallowed regions =1.4%

Errat score 72.874

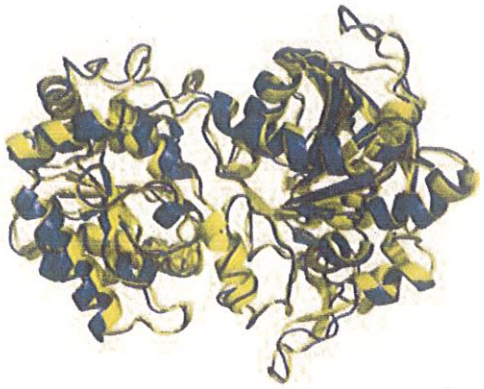


III-9

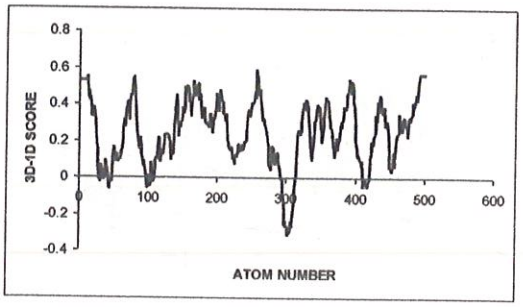
Structure of III-9(GBSS1 protein)



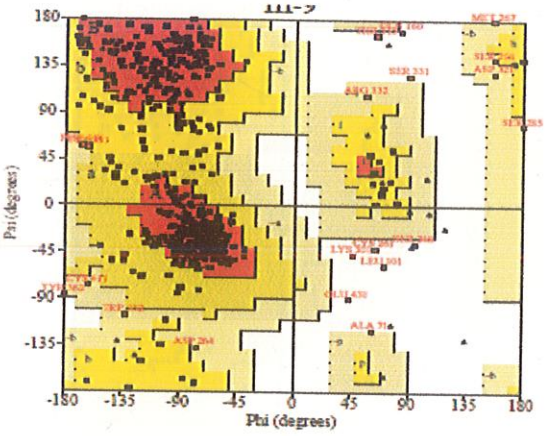
Structure of III-1 superimposed with Template. RMSD 0.904Å



61.63% of atoms have 3D-ID score>0.2



Ramachandran plot

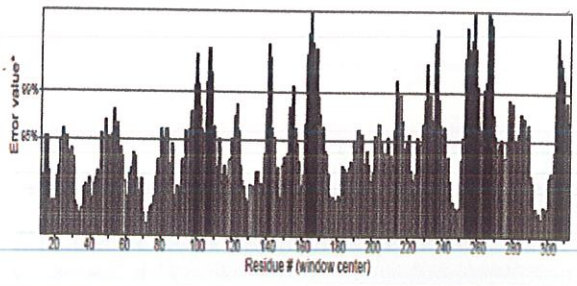


Residues in most favored regions =74.9%

Residues in additional allowed regions =20.5%

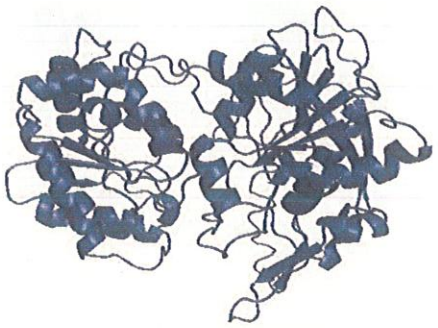
Residues in disallowed regions = 1.8%

Errat score 62.34



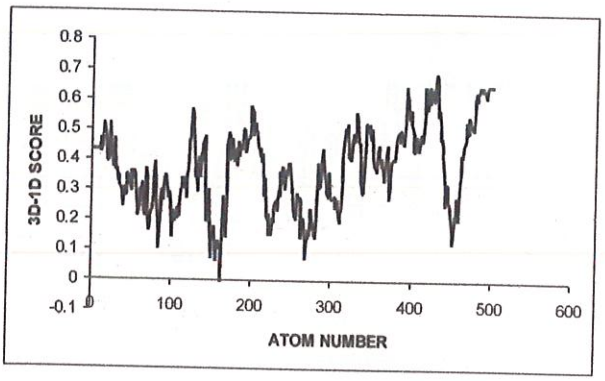
V-1

Structure of V-1(GBSS1 protein)

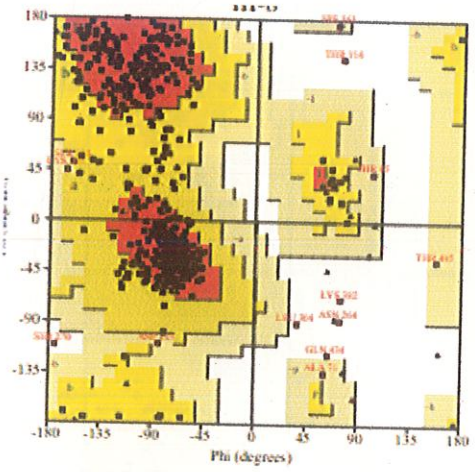


Structure of III-1 superimposed with Template. RMSD 0.723Å

89.72% of atoms have 3D-ID score>0.2



Ramachandran plot

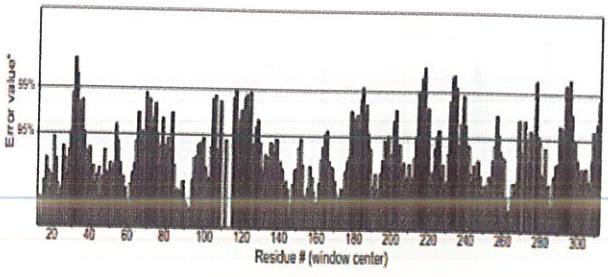


Residues in most favored regions =78.2%

Residues in additional allowed regions = 18.9%

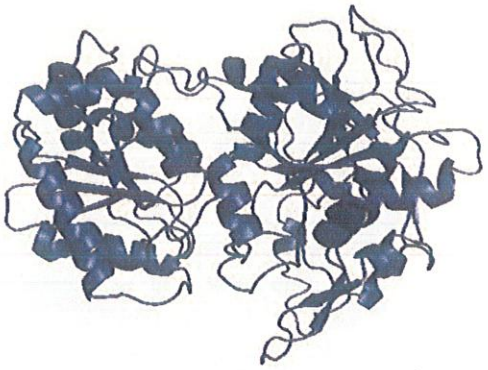
Residues in disallowed regions =1.4%

Errat score 72.802



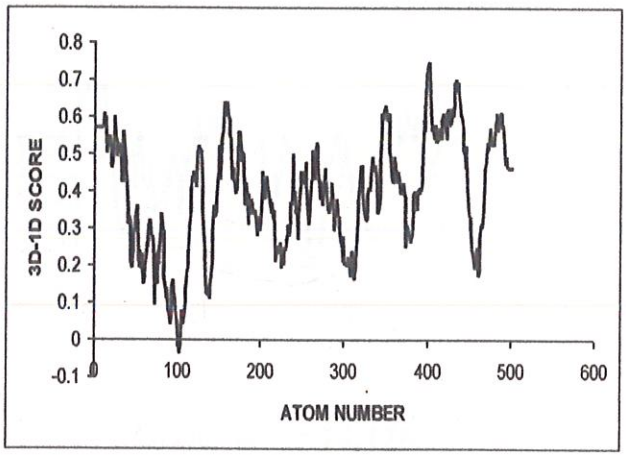
V-2

Structure of V-2(GBSS1 protein)

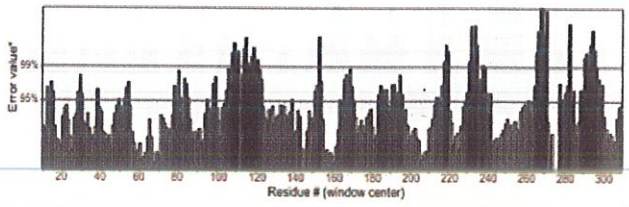


Structure of III-1 superimposed with Template. RMSD 0.880Å

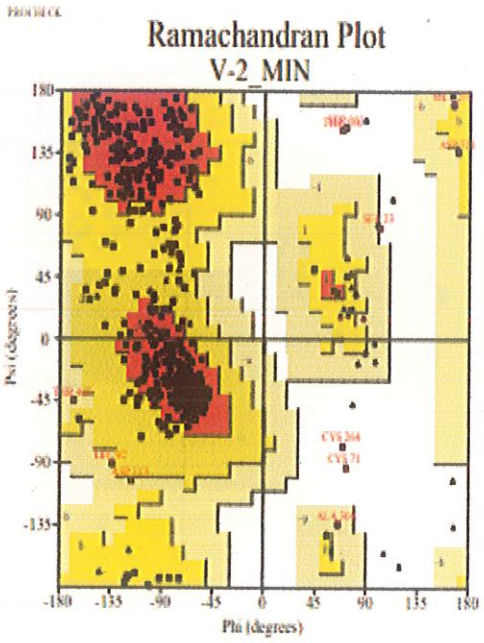
59.76% of atoms have 3D-ID score>0.2



Errat score 67.74



Ramachandran plot



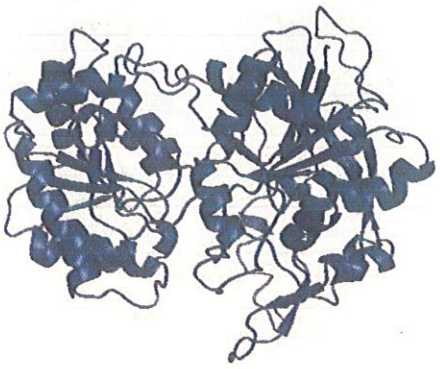
Residues in most favored regions =80.0%

Residues in additional allowed regions =17.4%

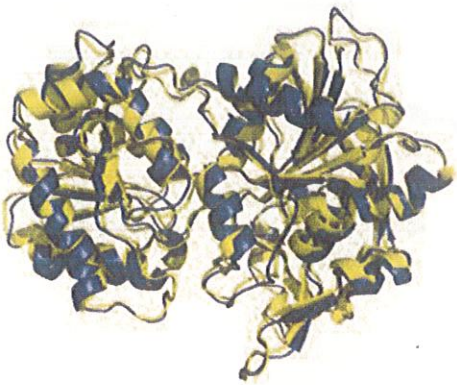
Residues in disallowed regions =1.1%

V-4

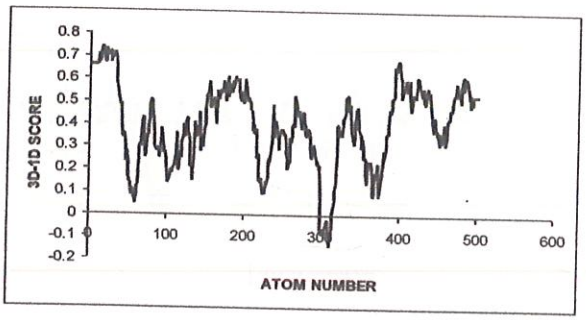
Structure of V-4(GBSS1 protein)



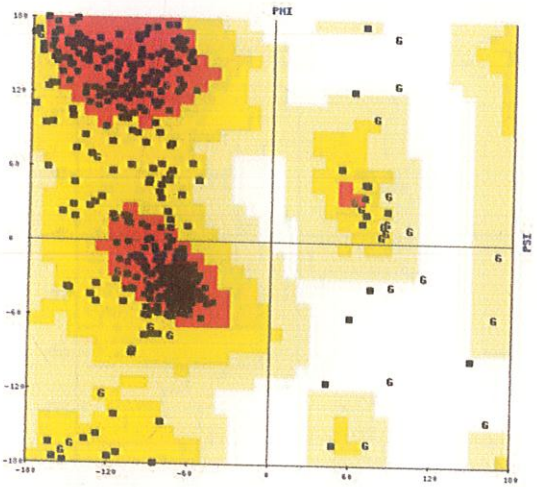
Structure of III-1 superimposed with Template. RMSD 3.504Å



86.39% of atoms have 3D-ID score>0.2



Ramachandran plot

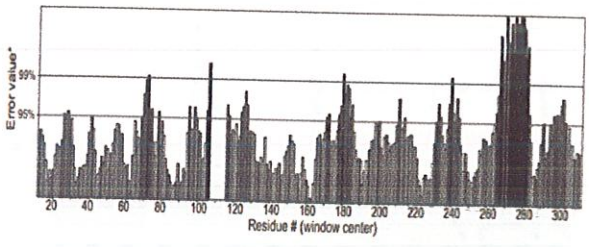


Residues in most favored regions =78.9%

Residues in additional allowed regions =18.8%

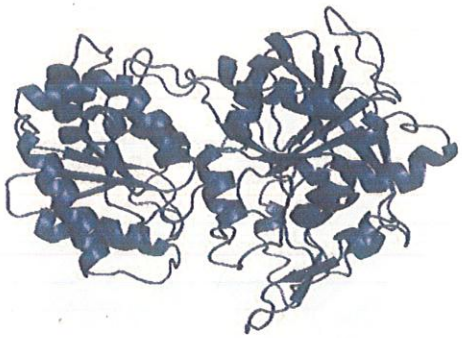
Residues in disallowed regions =1.1%

Errat score 74.120

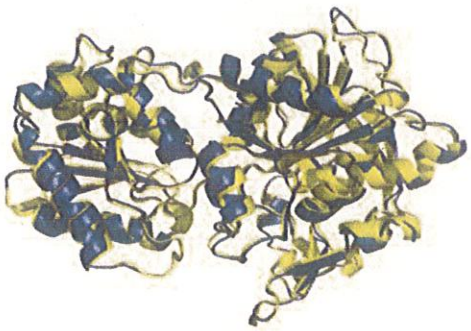


V-5

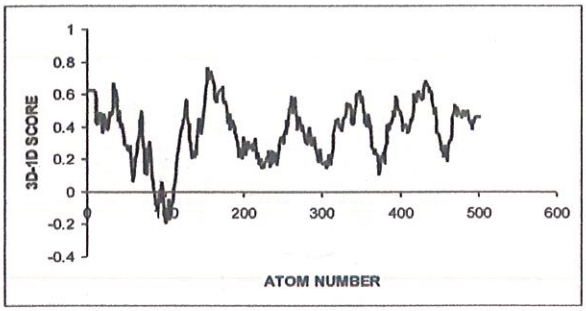
Structure of V-5(GBSS1 protein)



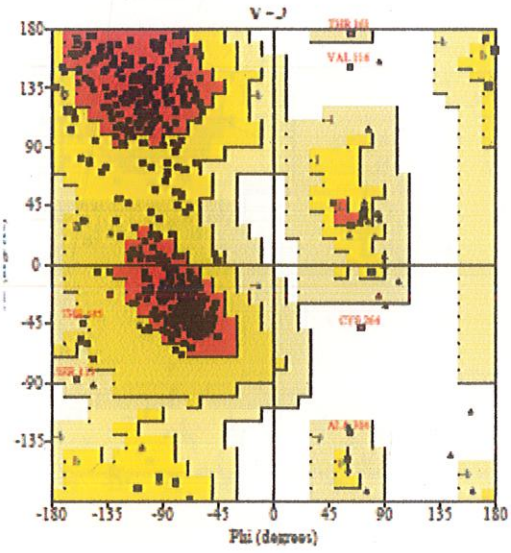
Structure of III-1 superimposed with Template. RMSD 0.680Å



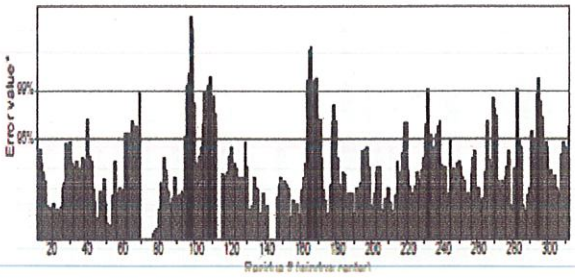
84.98% of atoms have 3D-ID score>0.2



Ramachandran plot



Errat score 79.418



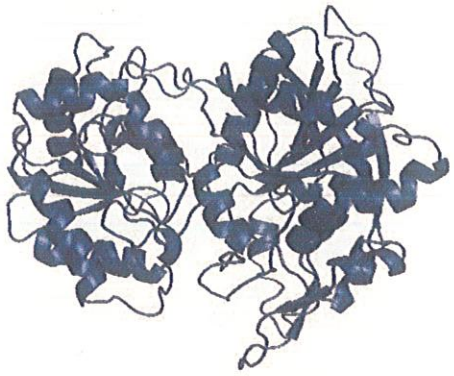
Residues in most favored regions= 82.6%

Residues in additional allowed regions= 16.1%

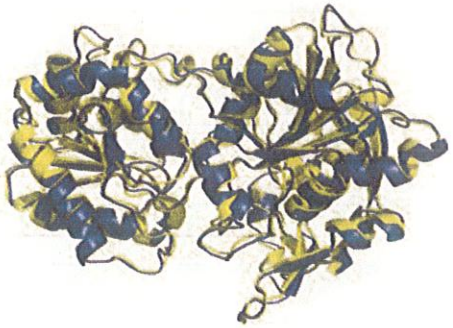
Residues in disallowed regions= 0.5%

V-6

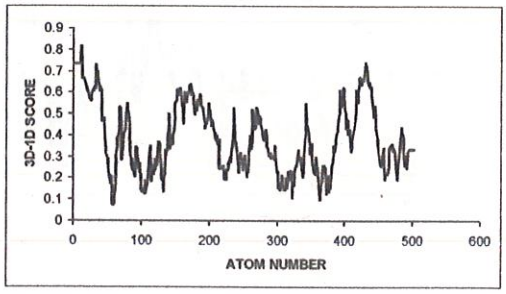
Structure of V-6(GBSS1 protein)



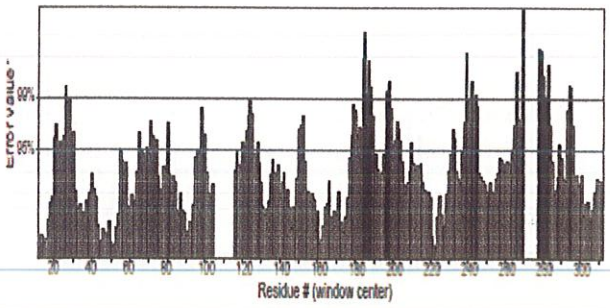
Structure of III-1 superimposed with Template. RMSD 0.909Å



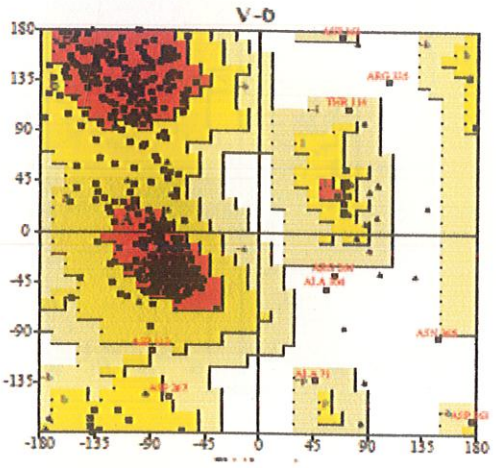
89.13% of atoms have 3D-ID score>0.2



Errat score 74.214



Ramachandran plot



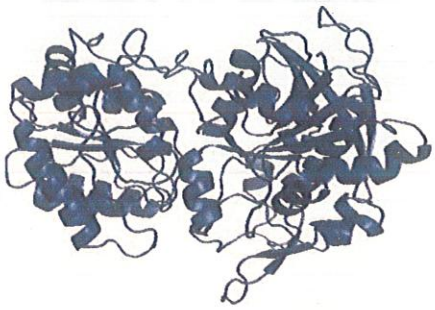
Residues in most favored regions =79.8%

Residues in additional allowed regions =17.9%

Residues in disallowed regions =0.9%

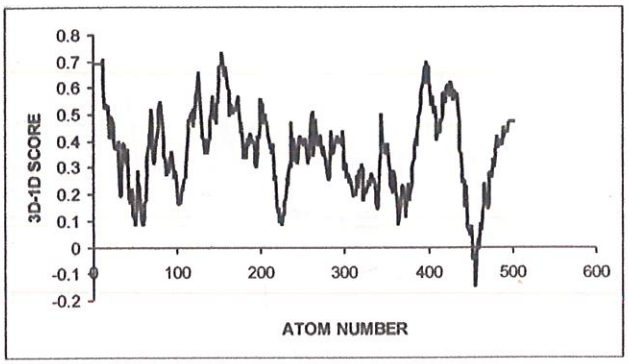
V-7

Structure of V-7(GBSS1 protein)

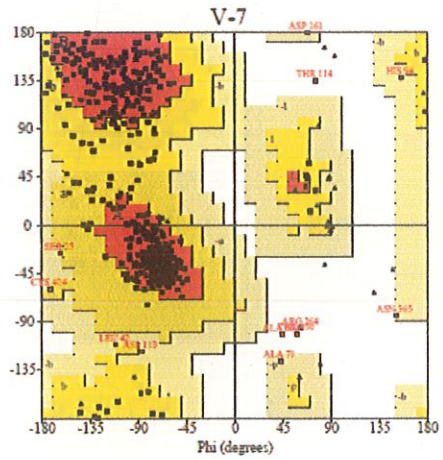


Structure of III-1 superimposed with Template. RMSD 0.820Å

84.19% of atoms have 3D-ID score>0.2



Ramachandran plot

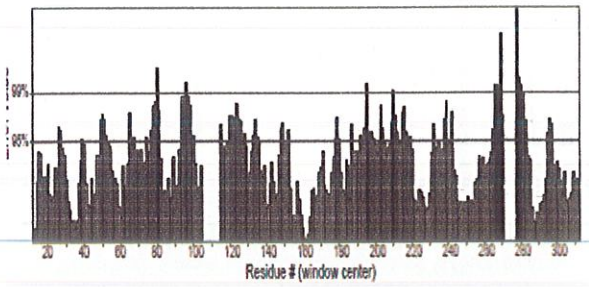


Residues in most favored regions =80.08%

Residues in additional allowed regions =16.5%

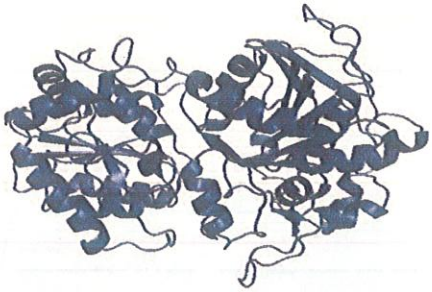
Residues in disallowed regions =0.9%

Errat score 70.084

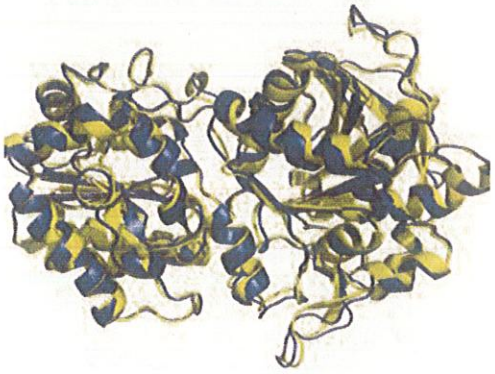


V-8

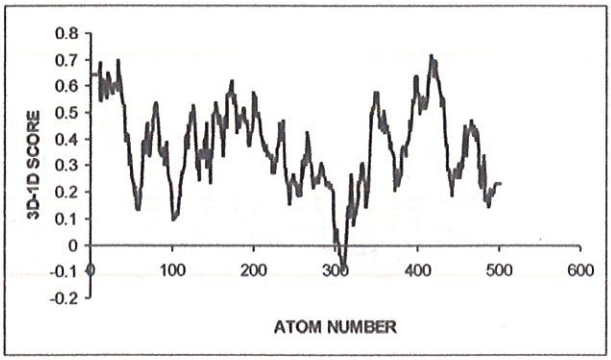
Structure of V-8(GBSS1 protein)



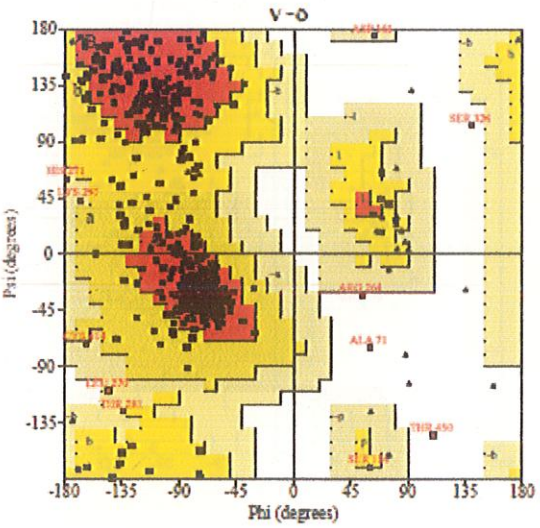
Structure of III-1 superimposed with Template. RMSD 0.831Å



87.55% of atoms have 3D-ID score>0.2



Ramachandran plot

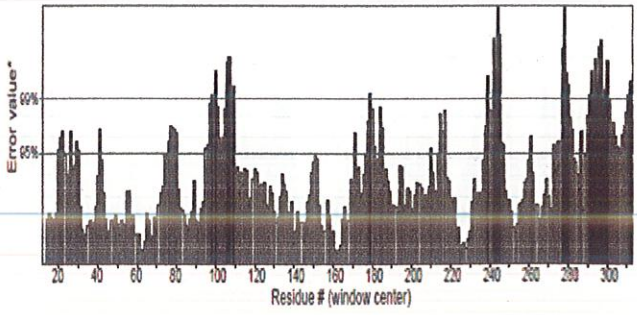


Residues in most favored regions =78.2%

Residues in additional allowed regions =19.3%

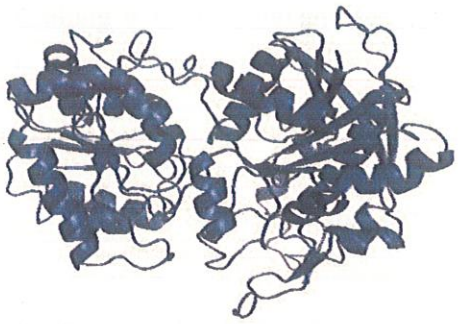
Residues in disallowed regions =0.9%

Errat score 68.219

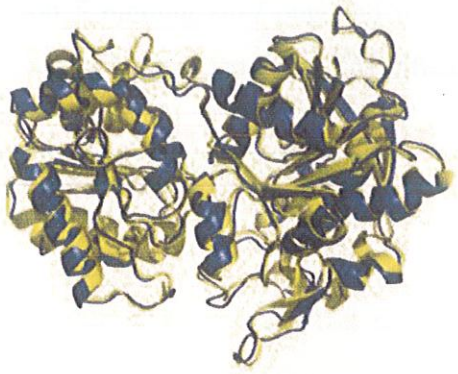


V-9

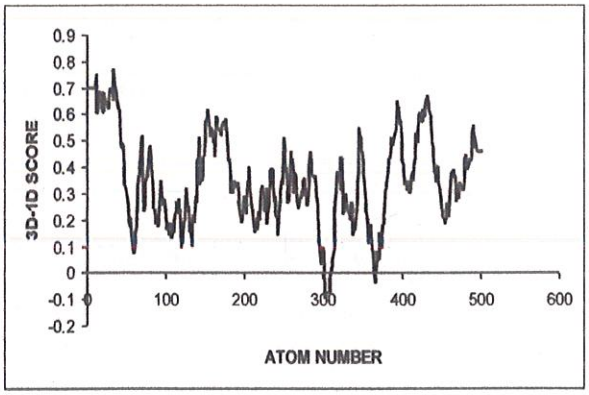
Structure of V-9(GBSS1 protein)



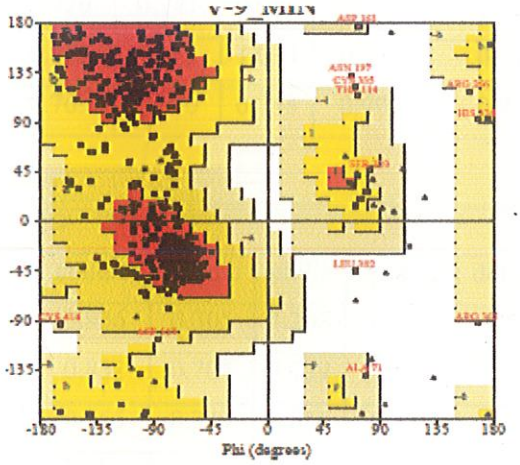
Structure of III-1 superimposed with Template. RMSD 0.921Å



82.21% of atoms have 3D-ID score>0.2



Ramachandran plot

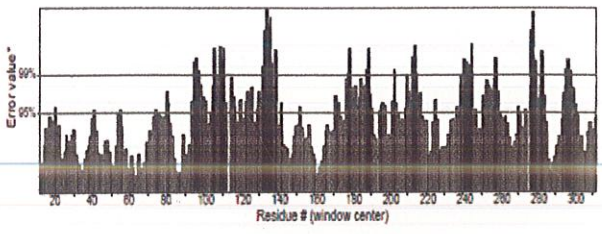


Residues in most favored regions =79.5%

Residues in additional allowed regions =17.7%

Residues in disallowed regions = 0.9%

Errat score 64.63



Binding Site Prediction

To find out about the interaction between Starch Molecule and GBSS-1 Enzymes that have been modelled, we need to find out the regions in the GBSS-1 enzyme where the substrate would bind to form the products. We used Sitemap tool to determine the top 5 locations in the enzyme for binding of the starch molecule. The table given below lists the top 5 binding sites obtained for all the 19 proteins along with the relevant data and scores. On the basis of these scores we identified the best possible site for each protein in which it was possible to carry out the docking with the starch molecule.

Following are the 19 tables depicting SITEMAP scores

II-1										
	SiteScore	size	volume	exposure	enclosure	contact	phobic	philic	balance	don/acc
site_1	1.077	158	454.475	0.564	0.759	0.921	0.972	0.817	1.189	0.746
site_4	0.811	53	176.302	0.728	0.606	0.744	0.818	0.640	1.278	0.882
site_3	0.771	43	108.731	0.616	0.715	0.920	0.241	0.939	0.257	1.100
site_2	0.752	51	140.63	0.675	0.641	0.882	0.353	1.034	0.341	0.572
site_5	0.678	35	81.291	0.615	0.591	0.769	0.680	0.753	0.903	0.959

II-5										
	SiteScore	size	volume	exposure	enclosure	contact	phobic	philic	balance	don/acc
site_1	1.035	101.000	294.637	0.526	0.750	0.936	0.305	1.203	0.254	0.533
site_3	0.980	85.000	259.651	0.685	0.672	0.831	0.704	0.729	0.966	0.546
site_2	0.946	89.000	330.309	0.720	0.635	0.765	0.339	0.851	0.399	0.811
site_4	0.788	41.000	117.649	0.707	0.628	0.684	0.806	0.508	1.589	1.354
site_5	0.682	33.000	79.919	0.764	0.550	0.658	0.526	0.533	0.987	0.745

II-6										
Title	SiteScore	size	volume	exposure	enclosure	contact	phobic	philic	balance	don/acc
site_1	1.112367	232	661.647	0.549	0.752	0.924	1.326	0.621	2.134	0.514
site_2	1.030259	72	146.804	0.437	0.785	1.081	1.535	0.586	2.619	2.019
site_5	1.009027	46	78.547	0.494	0.824	1.055	3.822	0.199	19.199	13.579
site_3	0.895448	35	55.223	0.435	0.981	1.223	2.095	0.975	2.149	0.584
site_4	0.784977	38	123.48	0.728	0.635	0.774	1.326	0.461	2.876	0.288

II-7										
Title	SiteScore	size	volume	exposure	Enclosure	contact	phobic	Philic	balance	don/acc
site_1	1.132	218	456.876	0.377	0.830	1.064	1.384	0.782	1.769	0.473
site_3	0.821	44	117.306	0.653	0.686	0.896	1.514	0.619	2.445	0.707
site_2	0.758	49	89.866	0.65	0.601	0.790	0.293	0.786	0.373	0.717
site_5	0.678	28	83.692	0.768	0.561	0.704	1.089	0.422	2.580	0.551
site_4	0.669	35	74.431	0.646	0.635	0.872	0.734	0.948	0.774	0.748

II-8									
Title	SiteScore	size	volume	Exposure	enclosure	contact	phobic	philic	balance
site_1	1.047	210	575.211	0.519	0.754	0.932	0.715	0.952	0.751
site_3	1.012	75	109.417	0.506	0.713	0.887	1.972	0.498	3.952
site_2	0.751	44	76.489	0.551	0.695	0.882	0.129	1.363	0.094
site_5	0.676	38	69.286	0.722	0.561	0.695	0.256	0.755	0.339
site_4	0.665	32	77.175	0.594	0.673	0.912	0	1.198	0

II-9										
Title	SiteScore	size	volume	exposure	Enclosure	contact	phobic	philic	balance	don/acc
site_1	1.091	123	371.126	0.5	0.787	1.035	1.237	0.840	1.472	0.335
site_3	1.021	39	71.687	0.338	0.986	1.277	3.351	0.480	6.973	1.424
site_2	0.907	65	140.63	0.522	0.772	1.007	0.532	1.050	0.506	2.103
site_5	0.893	64	239.071	0.597	0.725	0.861	0.481	0.892	0.539	1.402
site_4	0.744	32	117.992	0.644	0.733	0.858	0.550	0.805	0.683	5.511

III-1										
Title	SiteScore	size	volume	exposure	enclosure	contact	phobic	philic	balance	don/acc
site_1	1.050	53	73.745	0.293	0.979	1.286	2.815	0.689	4.084	0.596
site_3	1.017	315	990.584	0.546	0.724	0.922	0.373	1.141	0.326	0.572
site_2	0.982	80	152.635	0.509	0.775	1.066	0.776	0.958	0.810	3.967
site_5	0.883	67	127.939	0.521	0.722	0.944	0.259	1.256	0.206	1.106
site_4	0.804	55	148.519	0.764	0.592	0.767	0.704	0.679	1.036	0.873

III-2										
Title	SiteScore	size	volume	exposure	Enclosure	contact	phobic	philic	balance	don/acc
site_1	1.051	224	752.542	0.521	0.769	1.000	0.699	0.980	0.713	0.820
site_3	0.812	53	110.789	0.490	0.715	1.098	0.176	1.418	0.124	0.310
site_2	0.761	47	88.151	0.646	0.581	0.828	0.847	0.650	1.303	0.667
site_5	0.633	25	95.011	0.816	0.582	0.720	0.578	0.611	0.945	0.576
site_4	0.591	18	56.252	0.684	0.676	0.849	0.797	0.859	0.927	1.205

III-7										
Title	SiteScore	size	volume	exposure	enclosure	contact	phobic	philic	balance	don/acc
site_1	1.179	200	365.295	0.307	0.892	1.145	1.955	0.751	2.601	0.945
site_3	1.081	91	159.152	0.409	0.809	1.119	1.515	0.793	1.909	1.405
site_2	0.864	63	122.451	0.622	0.665	0.800	1.072	0.814	1.315	0.298
site_5	0.781	61	149.891	0.733	0.608	0.761	0.438	0.990	0.442	1.116
site_4	0.638	40	68.943	0.689	0.522	0.744	0.333	0.872	0.382	0.740

III-8										
Title	SiteScore	size	volume	exposure	Enclosure	contact	phobic	philic	balance	don/acc
site_1	1.027	260	905.863	0.580	0.734	0.893	0.484	0.984	0.492	0.531
site_3	0.744	51	198.94	0.741	0.629	0.802	0.128	1.001	0.127	0.777
site_2	0.687	42	80.605	0.631	0.594	0.853	0.216	0.925	0.233	0.780
site_5	0.685	47	54.537	0.801	0.538	0.597	0.126	0.888	0.142	0.463
site_4	0.677	39	67.914	0.7	0.540	0.613	0.600	0.709	0.845	0.335

III-9										
Title	SiteScore	size	volume	exposure	enclosure	contact	phobic	philic	balance	don/acc
site_1	1.125	43	57.624	0.348	0.995	1.321	6.284	0.106	58.861	1.177
site_3	1.060	48	63.455	0.294	0.966	1.311	5.044	0.469	10.742	0.390
site_2	1.056	304	777.581	0.440	0.783	1.019	0.596	1.038	0.577	0.455
site_5	0.944	67	201.684	0.603	0.814	1.029	0.810	1.100	0.736	0.912
site_4	0.814	46	108.731	0.544	0.773	1.018	0.135	1.260	0.107	0.785

V-1										
Title	SiteScore	size	volume	exposure	enclosure	contact	phobic	philic	balance	don/acc
site_1	1.045	216	603.68	0.503	0.766	0.956	0.498	1.015	0.491	0.463
site_3	0.836	48	113.19	0.538	0.789	1.069	0.974	1.011	0.963	1.210
site_2	0.802	42	101.185	0.628	0.683	0.896	0.751	0.645	1.164	1.348
site_5	0.692	34	108.388	0.682	0.669	0.822	0.685	0.910	0.753	1.091
site_4	0.681	34	66.542	0.569	0.678	0.925	0.173	1.481	0.117	1.159

V-2										
Title	SiteScore	size	volume	exposure	Enclosure	contact	phobic	philic	balance	don/acc
site_1	1.027	291	882.882	0.548	0.737	0.931	0.623	0.995	0.626	0.518
site_3	0.917	85	241.129	0.700	0.606	0.740	0.331	0.821	0.402	1.115
site_2	0.771	30	86.093	0.610	0.739	1.055	0.945	0.624	1.514	1.549
site_5	0.768	51	105.301	0.643	0.613	0.765	0.318	0.825	0.385	0.840
site_4	0.652	39	117.649	0.775	0.590	0.772	0.113	1.087	0.104	0.763

V-4										
Title	SiteScore	size	volume	exposure	enclosure	contact	phobic	philic	balance	don/acc
site_1	1.082	278	798.847	0.5	0.760	0.894	1.065	0.795	1.338	0.474
site_3	0.842	70	162.582	0.699	0.599	0.766	0.377	0.856	0.441	0.466
site_2	0.836	77	181.447	0.745	0.570	0.733	0.367	0.943	0.390	0.515
site_5	0.728	49	112.504	0.701	0.601	0.823	0.561	0.934	0.600	0.743
site_4	0.694	43	82.32	0.711	0.561	0.664	0.307	0.808	0.379	1.134

V-5										
Title	SiteScore	size	volume	exposure	enclosure	contact	phobic	philic	balance	don/acc
site_1	1.095	250	631.806	0.352	0.841	1.043	0.896	1.07	0.835	0.379
site_2	0.961	113	257.936	0.690	0.595	0.764	0.525	0.850	0.618	1.070
site_3	0.841	55	117.649	0.697	0.611	0.722	0.910	0.558	1.629	0.387
site_4	0.809	61	141.659	0.743	0.579	0.747	0.574	0.752	0.763	0.625
site_5	0.751	47	122.451	0.598	0.671	0.805	0.037	1.135	0.0327	1.020
V-6										
Title	SiteScore	size	volume	exposure	enclosure	contact	phobic	philic	balance	don/acc
site_1	1.053	226	824.915	0.573	0.763	0.942	0.624	0.954	0.654	0.60
site_2	0.860	62	142.002	0.685	0.681	0.909	0.345	0.864	0.400	1.077
site_3	0.769	39	81.291	0.602	0.631	0.770	0.968	0.553	1.743	1.225
site_5	0.639	27	84.035	0.721	0.619	0.853	0.611	0.779	0.784	0.555
site_4	0.622	29	111.818	0.794	0.597	0.697	0.267	0.858	0.311	1.363

V-7										
Title	SiteScore	size	volume	exposure	enclosure	contact	phobic	philic	balance	don/acc
site_1	1.050	279	727.503	0.552	0.751	0.924	0.732	0.926	0.790	0.549
site_2	0.970	74	126.224	0.482	0.705	0.935	1.154	0.659	1.750	0.71
site_3	0.947	40	69.286	0.506	0.806	1.028	3.813	0.279	13.644	1.934
site_4	0.857	50	119.707	0.519	0.771	1.010	1.406	0.881	1.594	0.968
site_5	0.634	27	86.436	0.742	0.630	0.747	0.208	0.838	0.248	1.046

V-8										
Title	SiteScore	size	volume	exposure	enclosure	contact	phobic	philic	balance	don/acc
site_1	1.065	189	883.911	0.604	0.747	0.859	0.732	0.840	0.871	0.353
site_2	0.754	48	124.852	0.636	0.667	0.836	0.503	1.065	0.472	0.485
site_3	0.722	42	77.861	0.618	0.593	0.759	0.719	0.748	0.961	1.122
site_4	0.694	40	91.238	0.672	0.644	0.805	0.276	1.141	0.242	0.797
site_5	0.630	35	65.17	0.726	0.536	0.642	0.337	0.811	0.415	1.329

V-9										
Title	SiteScore	size	volume	exposure	Enclosure	contact	phobic	philic	balance	don/acc
site_1	1.035	276	732.648	0.580	0.708	0.850	0.603	0.855	0.706	0.678
site_2	0.953	73	185.22	0.665	0.652	0.804	1.072	0.547	1.959	1.094
site_4	0.902	40	83.006	0.473	0.955	1.153	1.260	1.013	1.242	0.658
site_3	0.847	68	174.587	0.726	0.621	0.719	0.525	0.861	0.609	0.628
site_5	0.683	35	77.175	0.726	0.559	0.668	0.655	0.624	1.049	1.192

Since there are mutations in the binding site we could easily see a change in the physicochemical properties of the binding pocket in the different structures.

Molecular docking interaction

Ligand Docking is the process in which the modelled protein (GBSS-1) is made to interact with the substrate molecule (Starch). The entire process is simulated computationally and the energies with respect to the pairwise interaction of the ligand and the protein are calculated. This interaction and the energy graphs plotted after that with the help of the values obtained help us to analyse the important amino acids in the interaction and the difference in the interaction because of a particular mutation (change in an amino acid).

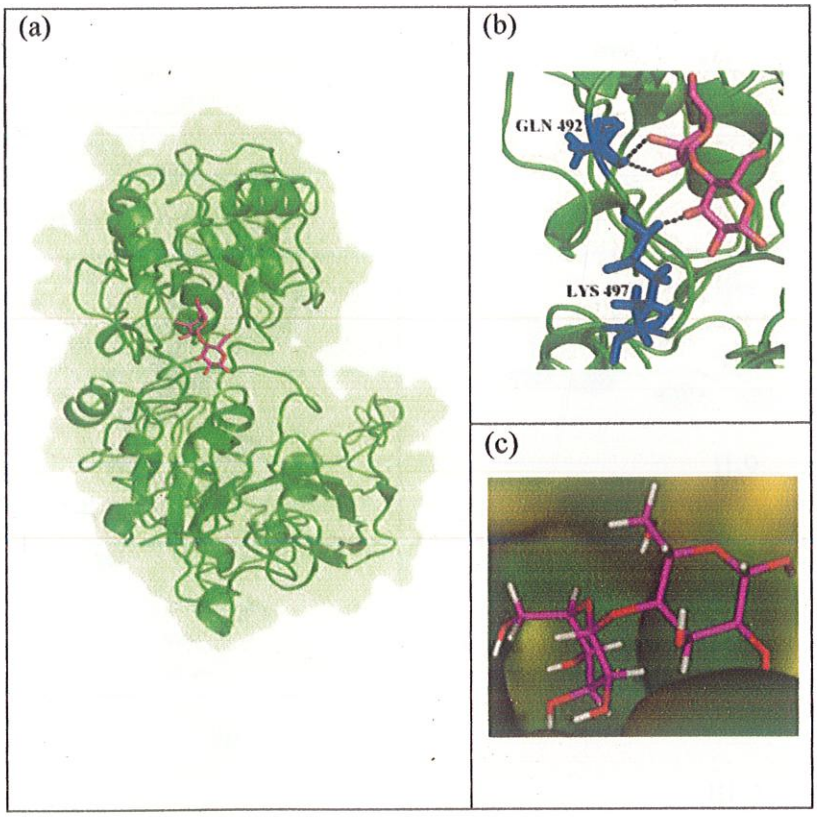
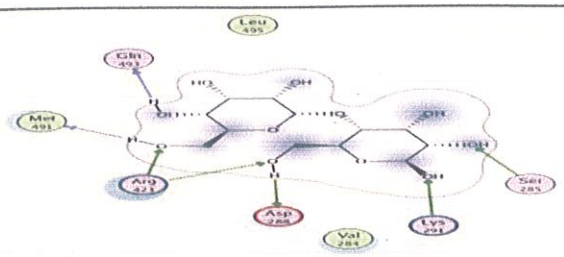
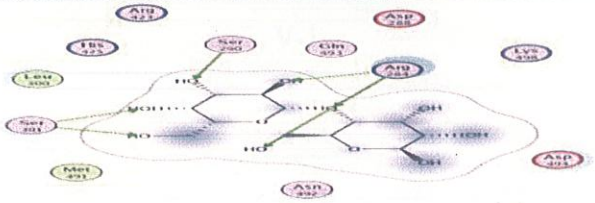


Fig 3.13: (A) showing docked starch molecule with GBSS-1, cultivar II-1; (B) starch molecule docked with II-1 with the nearest residues marked; (C) Zoomed view of docking of starch with II-1 (space filled representation)

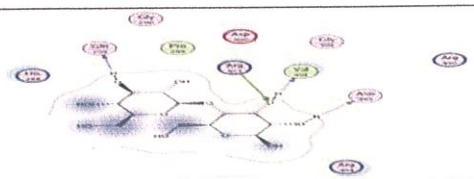
The docking results reveal that the mutations have affected the binding of starch to GBSS1 protein.we were motivated to analyze the details of the docking interactions.ligplot was used to analyze the hydrogen bonding pattern of starch with GBSS1 . The results are shown in the figures given below:



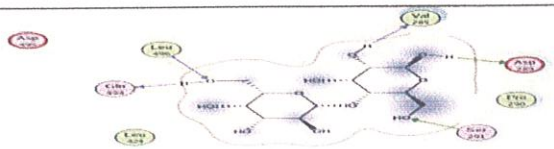
II-1



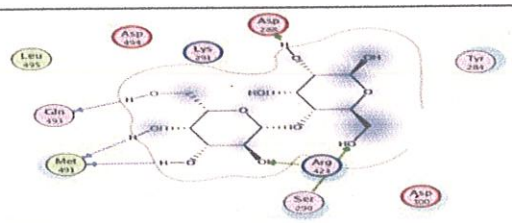
II-5



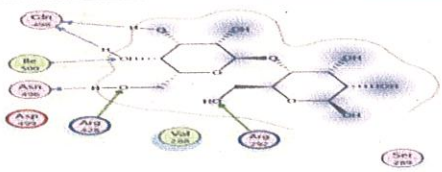
II-6



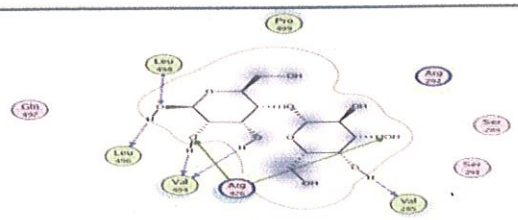
II-7



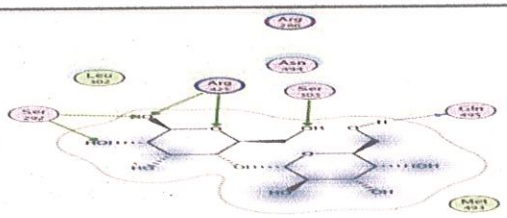
II-8



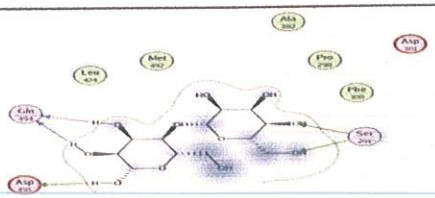
II-9



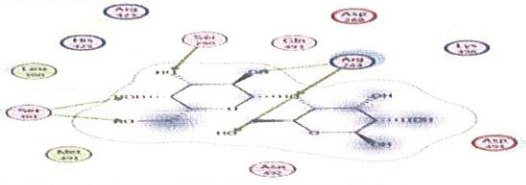
III-1



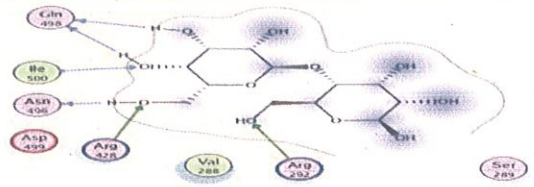
III-2



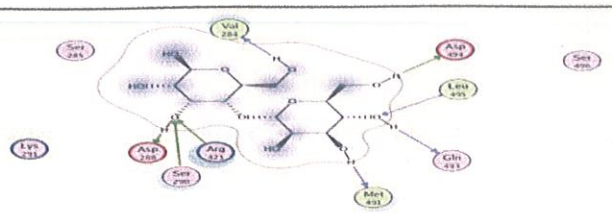
III-7



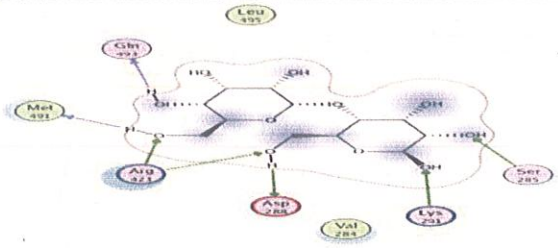
III-8



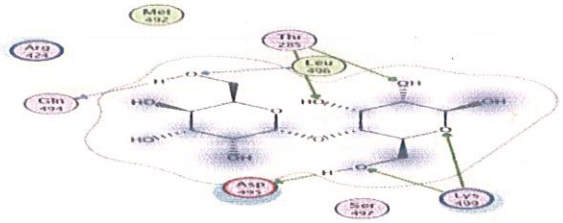
III-9



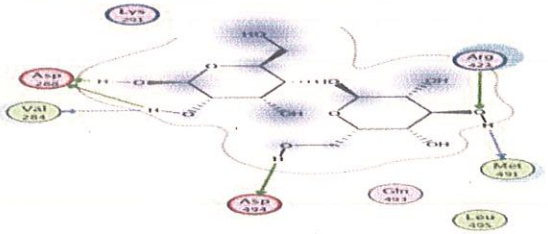
V-1



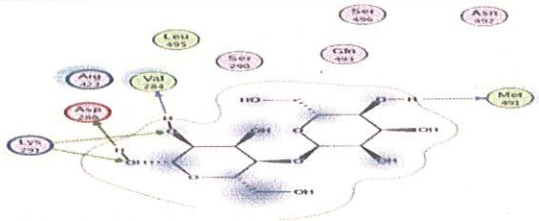
V-2



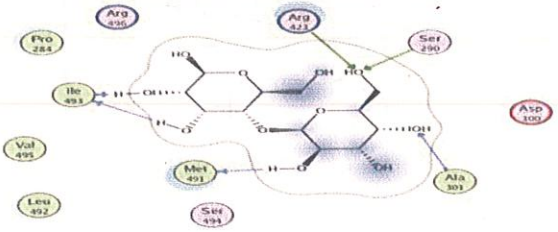
V-4



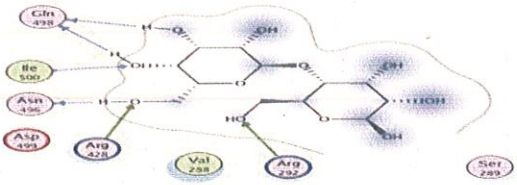
V-5



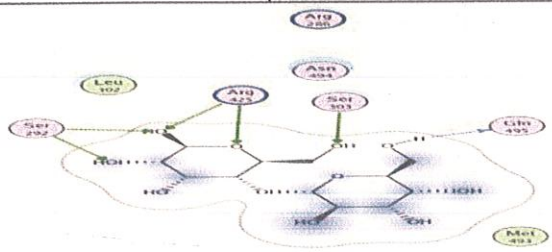
V-7



V-8



V-9



V-6

To analyze the interaction of the starch molecule with the GBSS-1 enzyme of various cultivars, we also needed to find out the amino acids which were found to be the most substantially contributing in the docking process. Hence, we analyzed the interaction by plotting graphs between the residue number and its corresponding docking energy. The graphs are presented below:

The table given below shows the docking scores obtained for all the cultivar sequences:

TITLE	DOCKING SCORE	GLIDE GSCORE
III-7_MIN	-6.655331	-6.655331
III-2_MIN	-7.056223	-7.056223
III-1_MIN	-8.551607	-8.551607
II-9_MIN	-6.456373	-6.456373
II-7_MIN	-7.294466	-7.294466
II-6_MIN	-7.588458	-7.588458
II-1_MIN	-6.620772	-6.620772
II-8_MIN	-6.464576	-6.464576
III-8_MIN	-7.179732	-7.179732
II-9_MIN	-6.456373	-6.456373
V-5_MIN	-6.70477	-6.70477
V-6_MIN	-7.054808	-7.054808
V-7_MIN	-8.192043	-8.192043
V-8_MIN	-7.788419	-7.788419
V-1_MIN	-9.12373	-9.12373
V-2_MIN	-7.864315	-7.864315
V-4_MIN	-3.508568	-3.508568

CHAPTER 4

DISCUSSION

The basic aim of our study was to understand how point mutations in the sequence of Rice cultivars. And how it affects the interaction of the enzyme with the ligand molecule, if the former is present in the active site of the enzyme. The point mutations in various regions of the sequence were found to affect the total activity of the enzyme and were found to affect its interaction with the starch molecule as revealed by the docking studies. The change in the amino acids at one or more points reflected upon the interaction of the entire molecule. In cultivar V-6, for instance, the presence of phenylalanine as residue number 44 gave us the docking energy for that point as -19.49 kcal/mol, which incidentally was also the lowest for the cultivar, at the same time, the presence of Asparagine, at the same place, residue number 44 in cultivar V-7, resulted in the docking energy of the point to be -29.39 kcal/mol which, again was the lowest for the molecule. Even in cultivars II-6 and III-1, the difference was noted at point 8, where the presence of alanine in II-6 and glycine in III-1 resulted in the docking energies contribution to be -21.12 kcal/mol and -16.58 kcal/mol respectively. Thus, it is evident the starch molecule shows variation in the docking score with different structures. Further investigating the details of the interaction, using Ligplot analysis we found out that the amino acids within the binding site interact differently with the starch molecule in different structures. We have also evaluated the energy contribution of the amino acids within the binding site and the results indicate the mutations have affected the overall energy contribution of each amino acid within the binding site of the different GBSS-1 structures. So we could conclude that point mutations are effective in determining the quality of rice in different cultivars.

CONCLUSION

GBSS-1 enzyme is an important factor in studying the difference between various rice cultivars. While experimental methods have been used for a long time to study association of phenotype with the genetic make-up of an organism, computational methods like sequence alignment, homology modelling, docking etc provide a faster, more reliable and comprehensive view of the factors that affect the phenotype at the molecular level. This assumes greater importance with the scientific need for understanding the working of every system at the molecular level and the need for clear visualization of the factors like interaction and energy contribution. We found that point mutations have a significant impact upon the phenotype, structure and function of GBSS-1 enzyme and its interaction with it's substrate- starch. The docking scores and docking energies obtained by us clearly tell us that change in the amino acids at certain points in the sequence of the enzyme have a tremendous impact upon its structure and function. This study has been helpful for us to understand clearly the extent of effect certain amino acids have at different points in the sequence and the effect of different amino acids at the same point in different cultivars which has helped us to understand the basic differences which cause rice to have sticky and non-sticky mutants.

REFERENCES

1. Berendsen HJC, van der Spoel D, van Drunen R (1995) GROMACS: A message passing parallel molecular dynamics implementation. Computer Physics Communications 91: 43-56.
2. Daren T, York D, Pedersen L (1993) Particle mesh Ewald: An N-log(N) method for Ewald sums in large systems. Journal of Chemical Physics 98: 10089-10092.
3. Hess B, Bekker H,Berendsen HJC, Fraaije JGEM(1997) LINCS: A linear constraint solver for molecular simulations. J. Comput. Chem 18: 1463-1472.
4. Laskowski RA, MacArthurMW, Moss DS, Thornton JM(1993) PROCHECK: a program to check the stereochemical quality of protein structures. J. App. Crystal26: 283-291.
5. Colovos C, Yeates TO (1993) Verification of protein structures: Patterns of non-bonded atomic interactions. Protein Sci 2: 1511-1519.
6. Eisenberg D, Luthy R, Bowie, JU(1997) VERIFY3D: assessment 694 of protein models with three-dimensional profiles. Methods Enzymol 277: 396–404.
7. Ramachandran GN,Ramakrishnan C, SasisekharanV(1963) Stereochemistry of polypeptide chain configurations. J. Mol. Biol 7: 95–99.
8. Friesner RA, Banks JL, Murphy RB, Halgren TA, Klicic JJ, Mainz DT, Repasky MP, Knoll EH, Shelley M, Perry JK, ShawDE, Francis P, Shenkin PS (2004) Glide: a new approach for rapid, accurate docking and scoring. 1. method and assessment of docking accuracy. J. Med. Chem 47: 1739–1749.
9. Halgren TA, Murphy RB, Friesner RA, Beard HS, FryeLL, Pollard WT, Banks JL(2004) Glide: a new approach for rapid, accurate docking and scoring. 2. Enrichment factors in database screening. J. Med. Chem 47: 1750-1759

BIODATA

PRIYASHREE VERMA

I am currently pursuing B TECH in bioinformatics. My current research interests encompasses of protein modeling, docking and molecular aspects of protein analysis. I intend to pursue career in docking techniques, computational and molecular dynamics simulations for protein.

RAHUL ACHARYA

I am currently pursuing B. Tech in Bioinformatics. My goal is to pursue career in the domain of proteomics. I have hands on experience in visualisation and understanding the energetics of protein and docking techniques.