Modeling the interaction of Trehalose-6-Phosphate synthase and UDP-glucose bycomputational methods

J. Sarvanan[1], Sourav Dutta[2], Asit B. Mandal[3]*

[1]ICAR-Central Island Agricultural Research Institute, Port Blair, Andaman and Nicobar Islands, India, Pin-744101.
[2]Techno India University, Biotechnology Dept, EM-4/1, EM Block, Sector V, Salt Lake, Kolkata, West Bengal, India, Pin-700091.
[3]ICAR-Central Island Agricultural Research Institute, Port Blair, Andaman and Nicobar Islands, India, Pin-744101.
[3]Biotechnology Unit, ICAR-Central Research Institute for Jute and Allied Fibres (CRIJAF), Barrackpore, Kolkata -700120, WB, India
(Former Project Director, Directorate of Seed Research, Mau 275101, UP, India).
jsaravanan@gmail.com[1], vic_clueless@rediffmail.com[2], amandal2@rediffmail.com[3]

Abstract

Background: Trehalose, an integral component of the cell-wall glycolipids of *Mycobacterium*is an unusual non-reducing disaccharide, which plays a variety of biological roles from food storage to cellular protection under diverse environmental stresses (e.g. high pressure, heat-shock, reactive oxygen species, free radicals etc.). Primary enzymatic route to trehalose first involves the transfer of glucose from an UDP-glucose donor to glucose-6-phosphate to form alpha, alpha-1 and 1 trehalose-6-phosphate. It serves as a stress protectant in a range of organisms especially in resurrection of plants and many microorganisms. In addition, the principal role of trehalose is as an important metabolite due to its unique properties like high hydrophobicity, chemical stability, non-hygroscopic glass formation and no internal hydrogen bond formation is worthy to mention.

Methods: MODELLER, a command-line bioinformatics tools and various other bioinformatics programmes and servers like RAMPAGE, PRO-CHECK, Dallite, Errat, Verify 3D were used for analysis and validation of final structures created through homology modeling.

Results: The present work was done for virtual screening of the enzyme Trehalose-6-Phosphate Synthase and studying the interaction with ligand UDP. The crystal structure of TPS-UDP glucose was used for studying the protein-ligand interaction. Further homology modeling of the protein was done in detail by deploying basic operation of Argus Lab; Chimer graphics interface to display and manipulate modeled protein.

Application: We report here virtual screening of the enzyme Trehalose-6-Phosphate Synthase to unzip protein-ligand interaction through homology search or comparative modeling to pinpoint potential of the modeled/created protein and to analyze and validate the final structures using various informatics tools.

Keywords:Trehalose-6-Phosphate, UDP-Glucose, Homology modeling, MODELLER, PROCHECK

1. Introduction

Trehalose (α-D-glucopyranosyl-(1→1)-α-D-glucopyranoside) is a disaccharide comprising of two molecules of glucose. In spite of the fact that its biosynthesis is similar to that of sucrose, its evolutionary origin is probably more ancient because it is omnipresent in all kingdoms, be it plants, animals, microbes and arachnids. Lack of reducing ends in trehalose makes it highly resistant to heat and pH. Furthermore, trehalose imparts a strong stabilizing effect on biomolecules and thereby forms a glass-like structure after dehydration. Owing to such unique characteristics, trehalose is deemed to be a useful stabilizer in foods and an additive in cosmetics and pharmaceuticals. The biosynthesis and degradation of trehalose and trehalose 6-phosphate is similar to that of sucrose [Figure1].

TPP- Trehalose-p phosphatase, TPS- Trehalose-p synthase, UDP Glucose-Uridinediphosphate glucose, GDP- Guanosinediphosphoglucose, Glc-6-P- Glucose-6-phosphate, BLAST-Basic local alignment search tool, PIR- Protein information resources, pdf- Probability density function, PRO-Proline, TRP-Tryptophan, ILE-Isoleucine, HIS-Histidine, ARG-Arginine, LYS-Lysine, GLY-Glycine, LEU-Leucine, TYR-Tyrosine, ASP-aspartic acid, ASN-asparagine, GLU-glutamine, amino acid code A-Alanine, B-Aspartic acid, C-Cysteine, D-Aspartic, E-Glutamic, F-Phenvlalanine, G-Glycine, H-Histidine, I-Isoleucine, K-Lysine, L-Leucine, M-Methionine, N-Asparagine, O-Pyrrolysine, P-Proline, Q-Glutamine, R-Arginine, S-Serine, T-Threonine, S-Selenocysteine, V-Valine, W-Tryptophan, Y-Tyrosine, Z-Glutamine, X-Any.

The building blocks of trehalose are UDP-glucose and glucose 6-phosphate synthase, which are linked by the enzyme trehalose-6-phosphate phosphatase (TPP') 4. Although unspecific phosphatases are also able to dephosphorylate t6p, in *E.coli* the two enzymatic activities involved in trehalose biosynthesis are encoded by Ots A (TPS activity). This is in contrast to the situation in the yeast *Sacchromycescerevisiae*, where trehalose synthase complex is involved in the formation of trehalose.Trehalose is undoubtedly one of the key players in the protection of living organisms against extremity ofenvironmental stress, perhaps most well characterized under conditions of water stress and is found in abundance in case of 'Anhydrobiotic' organisms i.e., organisms that can live without water (e.g.,archea, eubacteria) and also in 'Resurrection' plants i.e., plants which employ trehalose as osmoprotectants to help in the regulation of drought stress response by inducing osmotolerance, which is achieved by closing of the stomata to avoid complete dehydration and hence ensure survival under extreme desiccation, even when 99% of their water is reported to be removed. Under these conditions cell viability is maintained through the synthesis of trehalose, which can reach levels up to 10-20% of the dry weight of these organisms. Stabilization reflects two major factors: the substitution for solvent water in H-bonding to proteins and the prevention of lipid-phase transitions at biological membranes. Even in so-called 'normal'organisms, trehalose synthesis is elevated under stress conditions, including dehydration, stationary phase culture, salt- stress, heat-shock, high pressure, extreme cold, exposure to oxygen radicals and anoxia.

Owing to its physicochemical properties, trehalose finds its use in varied biotechnological applications, the likes of which include preservation of flowers, foods, antibodies, enzymes, cells, tissues and organs. In order to engineer and acquire better crop yield under aberrant conditions such as water shortages (which have been demonstrated in tobacco, rice, potato and *Arabidopsis*) either the Trehalose biosynthesis pathway can be modified or the heterologous expression of bacterial or yeast trehalose biosynthesis genes can be achieved, thereby exploiting the osmoprotectant nature of trehalose.Adverse effects of enhanced and prolonged trehalose production usually result in altered phenotypes and negative effects on their physiology like aberrant leaf morphologies and decreased plant biomass yield justifying its referral as the 'double-faced' molecule [1, 2, 3, 4, 5, 6, 7, 8].

Modern NMR techniques are being extensively used to determine the structures of every larger moleculeincluding carbohydrates, nucleic acids and small to average-sized proteins. Computational biology isconceptualizing biology in terms of molecules and applying informatics techniques to understand andorganize the information associated with these molecules on a large scale. In recent years the proteinstructure prediction was made easier by the development of the computational biology. A mission for findingan alternative was accomplished, when homology modeling came into light. Molecular modeling is acollective term that refers to theoretical methods and computational techniques to model or mimic the behavior of molecules. The techniques are used in the fields of computational biology, computational chemistry and material science for studying molecular systems ranging from small chemical systems to large biological molecules and material assemblies. One of the major aspects of structural biology is to predict the 3-dimensional (3D) structure from the given sequence. X-ray crystallography provides an imaging technique capitalizing the diffraction of X-rays patterns through closely spaced lattice of atoms in a crystal is listed and then analyzed to reveal the nature of that lattice to unzip the 3D structure very effectively. Trehalose, a sugar produced by a wide variety of organisms, has long been known for its role in protecting certain organisms from diverse stresses especially from desiccation. Recent work in yeast indicates that trehalose also promotes survival under conditions of extreme heat, by enabling proteins to retain their native conformation at highly elevated temperatures and suppressing the aggregation of denatured proteins. The latter property, however, seems to impair the recovery of cells from heat shock if those fail to degrade trehalose after the stress is over. These multiple effects of trehalose on protein stability and folding suggest a host of promising applications. Three different pathways described very elaborately till date for the biosynthesis of trehalose, of whichfirstly described and most widely distributed pathway involves the transfer of glucose from UDP-glucose (or GDP-glucose in some cases) to glucose 6-phosphate to form trehalose-6-phosphate and UDP. The second pathway reported in a few unusual bacteria involves the intermolecular rearrangement of maltose (4-O-α-D-Glucopyranosyl-D-glucose) to convert the 1, 4-linkage to the 1, 1-bond of trehalose. Basically, trehalose synthase catalyzes this reaction and provides free trehalose as the initial product. A third pathway containing batteries of different enzymes is also evident. In this pathway the glucose at the reducing end of a glycogen chain converts the α l, 4-linkage to and α, α 1,1-bond, owing to synergistic effects of the enzyme pool. A second enzyme subsequently releases the trehalose disaccharide from the reducing end of the glycogen molecule. The most common route to trehalose is through the formation of trehalose-6-phosphate from UDP-Glc and Glc-6-P, catalyzed by the retaining"glycosyltransferaseOtsA"[9, 10].

2.Materials and Methods

The development of techniques in molecular biology allows rapid identification, isolation and sequencing of genes and proteins. A major goal of structural biology is to obtain the 3D structure of the proteins either by experimental methods such as X-ray crystallography, NMR and electron diffraction methods for deducing the 3D structure of the protein. Computational methods can then be applied to generate a reasonable model of proteins structure by homology modeling. This approach is known as comparative modeling or knowledge based modeling. One of the major hypotheses in homology modeling is that for a set of proteins that are found to be homologous, their 3D structure found to be conserved to a greater extent than their primary structure. This observation has been used to generate models of protein from homologues even with very low sequence similarity. But the conservation of folds among homologous proteins is the major factor that guides the entire method of comparative modeling.

2.1.Homology modeling usually consists of the following steps

1. Search for target sequence.
2. Search for template structure.
3. Confirming the search results.
4. Target template alignment.
5. Model building.
6. Model evaluation.

2.1.1. Search for the target sequence

The SWISSPROT database was searched against trehalose-6-phosphate synthase. From the result page the protein sequence not having a PDB entry was selected as the target.

2.1.2. Search for template sequence

Homology modeling requires at least one sequence of known 3D structure with significant similarity with the target sequence. The resulting target sequence was used as a query against NCBI Blast P by selecting PDB database in BlastP option. From result page the sequence having lowest e-value and high score value was selected as the template sequence is 1uqt.

2.1.3. Confirming the search result

Homology modeling requires sequence of known 3D structure and a target sequence having more than 35% similarity. In order to confirm the selection, a pair wise sequence alignment of template and target was performed and the sequence identity was found to be 38.395 and 40.319% (ungapped), respectively.

2.1.4.Target template alignment

To develop a model, all homology- modeling programmebasically depends upon a list of inputs that establishes structural resemblances between the target and the template residues and is defined by the alignment of the target and template sequence. The alignment is found to be relatively elementary to obtain when the target template sequences identity is more than 35%. When the target template sequence identity lies lower than 35%, the alignment generally retains gaps and needs manual mediation to minimize the number of misaligned residues. In these low sequences identity, accuracy plays the most important factor architecting the quality of the resulting model.

2.1.5. Model building

Once an initial target template isdone, a variety of methods can be used to involve constructing a 3D model for the target protein. Automated comparative model building by adopting the Program MODELLER 6v2 has been employed for modeling the target protein.

A.Homology Modeling with Program MODELLER

MODELLER is a computer program that develops 3D structure of proteins by fulfilling spatial restraints. The user provides an alignment of a sequence for modeling with known analogous structures. Modeler automatically calculates a model with all non-hydrogen atoms. In general, the inputs used to operate the program areconstraints

on the spatial structure of the amino acid sequence and ligands to be modeled. The output displays a 3D structure that satisfies these as well as possible constraints, which can in principle be deemed to be derived from a number of ancillary sources, which include the likes of related protein structures (combinatorial modeling), NMR experiments (NMR refinements), image reconstruction in electron microscopy,rules of secondary structure packing (combinatorial modeling), site-directed mutagenesis,cross- linking experiments, fluorescence spectroscopy, residue- residue and atom-atom potentials of mean force etc. The constraints operate angles, dihedral angles, distances, pairs of dihedral angles, and some other spatial features defined by pseudo atoms. Presently, MODELLER automatically identifies the constraints from known homologous structures and their alignment with the target sequence. It is to be mentioned that MODELLER also performscomparative assessment of protein sequences and/or structures, searching of sequence databases andclustering of proteins. The program is generally used with a scripting language and the use of graphics is excluded. It is written in standard FORTRAN 90 and is compatible to run in a UNIX computer [11].

B.Finding structures and sequences related to the target sequences

Before any modeling can be initiated, the sequences and segment with known 3D structures related to the sequence are modeled initially. Comparative protein modeling requires at least one sequence of known 3D structure with significant similarity with the target sequence. In order to determine if the modeling of the target protein can be carried out, it should be compared with a database of sequence of known protein derived from Brookhaven protein databank involving programs such as FASTA and BLAST. The archetypal structures are likely to have less than 40% sequence identity to each other and the length difference is at least 30% of the shorter chain or 30 amino acids residues, whichever is smaller. Template matching or threading methods can be used for more difficult modeling problems.

C.Preparing input Files

Modeller requires three input files:
1. Atom File (*.atm) Protein Data Bank atom files with coordinates for development of the template structures.
2. Alignment File (*.ali) the alignment file, the template structures aligned with the target sequence. And
3. Top File (* top) the MODELLER command or script file.

Atoms Files

Each atom file is named (*.atm) where * is a short protein code usually representing an alphanumeric code, preferably the PDB code e.g., 1uqt.atm. The code is used as that protein's identifier throughout the modeling process.

Alignment file

One of the formats for the alignment file is related to the PIR (Protein Information Resource) database format, which is generated by aligning the target sequence with the template sequence for pair wise alignment using ClustalX. This is the most preferred format for comparative modeling and isalso the most preferred and suitable format for executing MODELLER. This alignment file prepared is shown below:

.Ali File:

>P1; luqt

StructureX: luqt: 1: B: 456: B: TREHALOSE-6-PHOSPHATE SYNTHASE: ESCHERICHIACOLI: 2.00:0.22

SRLVWSNR------------
IAPPAGGEAVGILGALKAAGGLWFGWSGETGNEDQPLKKVKKGNITWASFNLSEQDLDEYYNQFSNAVLWPAFHYRLDL

VQFQRPAWDGYLRVNALLADKLLPLLQDDDIIWIHDYHLLPFAHELRKRGVNNRIGFFLHIPEPTPEIFNALPTYDTLLEQLCDYDLLGFQ

TENDRLAFLDCLSNLTRVTTRSAKSFTTAWGKAFRTEVYPIGIEPKJEIAKQAAGPEPPKLAQLKAELKNVQNIFSVERLDYSKGLPERFAL
AYEALLEKYPQHHGKIRYTQIAPTSRGDVQAYQDIRHQLENEAGRINGKYGQEGW

TPLYYENQHFDRKELMKIFRYSDVGLVTPLRDGMNEVAKEYVAAQDPANPGVLVLSQFAGAANELTSALIVNPYDRDEVAAALDRAL

TMSLAERISRHAEMLDVIVKNDINHWQECFISDLKQIVPR

>P1; T6PS

Sequence: T6PS1457: TREHALOSE-6-PHOSPHATESYNTHASE: DEINOCOCCUSGEOTHERMALIS: 2.00: 0.22

MGLIWSNREPYAPKRGEAGQLAWVPSIGGLTAALDPALQRSGGTWIAWGEERPEVGEVDLPQGAPRYRLKRLRLSEAEVRDFYYGFA
NRALWPMSHYFIGRATYQTSTWRTYVNVNRRF

AQAAVESFRTGDUWVHDYQLALVPRLIREALPGARtGFFWHIPWPSTEVFRTLPWDRELLEGLLGADLIGMHTDEYVAH
FLSACRRVLGAET-EGN TVRWQGRESRWARPIGIEVETY

EALAANPEVEEAADRtRRTLQTQILLGVDRLDYTKGIPERLEAFDAFLDRYPEARRRVTLLQIAVPSRERVESYRQERAQVEGLVGRINGK
HTRDGWSPVQYIYRGVPREELVAHYRAAD

VMLVTPLRDGLNLVAKEFVASSR---
DGVLILSRFAGAADELPEALQVNVDGLAEALLEALRMPLDEKKARLQRLRERLRQSDLHAWAEGFLRELAGT---

Script file
Since MODELLER is a command-line solo tool, with no graphical user interface, it requires a script file containing MODELLER commands. This is an ordinary Python 2.3 script. The script file contains commands for MODELLER, in the TOP language.
A script files model- default top to produce one model of sequence T6PS from the known structure of 1uqt and from the alignment between the two sequences is shown as follows:
.top File and its Description

Homology modeling by the MODELLER TOP routine "model".

INCLUDE   #Include the predefined TOP routines
SET OUTPUT- CONTROL = 1 1 1 1 1   #uncomment to produce a large log
File
SET ALNFILE = 'alignment.'   #alignment file name
SET KNOWNS = 'luqt'   #codes of the templates
SET SEQUENCE = 'T6PS'   #codes of the target
SET ATOM_ FILES_ DIRECTORY = '. /:.. / Atom_ files' #directories for input atom files
SET STARTING_ MODEL = 1   #index of the first model
SET ENDING_ MODEL = 10   #index of the last model
SET PDB _ EXT = '. Pdb'   #determines how many models to calculate
SET ALIGNMENT_ FORMAT = 'PIR'
CALL ROUTINE = 'model'   #do homology modeling

The first line of the top file includes MODELLER variable and routine definition. The following file lines set parameters values for the subsequent model routine. ALNFILE name the file that contains the target-template alignment in the PIR format. KNOWS define the known template structure in ALNFILE (alignment). SEQUENCE defines the name of the target sequence in ALNFILE. STARTING -MODEL and ENDING- MODEL defines the Number of models that are calculated. The first line of the file calls the model routine that actually calculates the model. The most important output files are model- single [12].

D. Method for comparative protein structure modeling by MODELLER
MODELLER implements an automated approach to undertake comparative assessment of protein structure modeling by satisfying spatial restraints.In a nutshell, the core modeling procedure initiates with an alignment of the sequence to be modeled (target) with related known 3D structures (templates). This alignment usually provides the inputs to the program. The output is a 3D model for the target sequence containing all main chain and side chain non-hydrogen atoms. At a given alignment, the model is obtained without any user intervention. Many distance and dihedral angle constraints on the target sequence are calculated from its alignment with template 3D structures.The form of these constraints was obtained by deploying a statistical analysis to unzip relationships between many pairs

of homologous structures. This analysis reliesheavily on a database of 105 family alignments that included 416 proteins with known 3D structures [13, 14].

By scanning the database, tables quantifying various correlations were obtained, such as the correlations between two equivalent C_α-C_α distances, or between equivalent main chain dihedral angles from two related proteins. The relationships were expressed as conditional probability density functions (pdf's) and can be used directly as spatial restraints, e.g. probabilities for different values of the main chain dihedral angles are calculated from the type of a residue considered, from main chain conformation of an comparable residue, and from sequence similarity between the two proteins. Another notable example is the pdf for a certain C_α-$C_\alpha$distance given equivalent distances in two related protein structures. An important feature of the method is that the spatial restraints are obtained empirically, from a database of protein structure alignments. Next, the spatial restraints and CHARMM energy terms enforcing proper stereochemistryare fused into an objective function. Finally, the model is achieved by optimizing the objective function in Cartesian space. The optimization is carried out by the use of the variable target function methodemploying methods of conjugate gradients and molecular dynamics with simulated annealing. Varying the initial structure can give rise to several slightly different models. The variation among those models could be used to estimate the quantum of errors in the corresponding regions of the fold. There are additional specialized modeling protocols, such as the modeling of loops [15, 16, 17, 18, 19].

E. Criteria for analysis of correctness

Program likes PROCHECK and Verify 3D is used to analyze the final structure created through homology modeling. Dihedral angles, peptide bonds, side chain conformation, hydrogen bonding, and hydrophobic/ hydrophilic residues are analyzed with the help of this program. All models are built by homology are prone to errors induced during many stages of model building. This model verification forms an important step in homology modeling methods. A model is considered to be wrong if at least part of its structural features is misplaced relatively to rest of the model. Error of that type can very easily slip into a model when erroneous sequence alignments are used during the model building procedure. Such models can nevertheless have proper stereochemistry thus care should be taken to this aspect during the building procedure. Models should be checked for the normality of torsion angles, bond angles and bond lengths. Also the bandedness of each amino acids and the inside /outside distribution of amino acids in case of trans membrane proteins should be checked, which helps in identifying misfolded regions in a trans membrane protein.

2.1.6. Model evaluation
A. Servers used for structure validation

Improved novel methods are available for differentiating between correctly and incorrectly determined regions of protein structures, which is purely based on characteristic atomic interaction. Different types of atoms are distributed non-randomly with respect to each other in proteins. Errors in model building lead to more randomized distributions of the different atom types, which can be distinguished from correct distributions by statistical methods. Most of the servers do structure validation on the basis of statistical methods. Some of the servers used are RAMPAGE SERVER, VERIFY 3D, PROCHECK, DALILITE and ERRAT.

RAMPAGE

Rampage is a program for visualizing and assessing the Ramachandran plot of a protein structure. It works on the basis of a manually curated set of high quality protein structures and a number of filters, reference phi/psi plots are derived for gly, pro, pre-pro and general residue type and subdivided into 'favoured', 'allowed' and 'outlier' regions will be listed, and a picture of the Ramachandran plotis displayed. The output, high-resolution multi-color Adobe PDF or PostScript file contains the general plot with critical data showing the percentage of residues that occurred in different regions.

PROCHECK

PROCHECK aims to assess how normal, or conversely how unusual, the geometry of the residues in a given protein structure, is compared with stereo chemical parameters derived from well-refined, high-resolution structures. The input to PROCHECK is a single file containing the coordinates of the protein structure. The outputs comprise a number of plots, together with a detailed residue-by-residue listing.

VERIFY 3D

Verify 3D provides a visual analysis of the quality of the putative crystal structure for a protein. It distinguishes between a mistraced or wrongly folded model, and that is basically correct. An effective test of the accuracy of the 3D protein model in comparisons of the model to its own amino acid sequence, using a 3D profile, computed from the atomic co-ordinates of the structure. Verify 3D works best with protein with at least 100 residues.

DALILITE

Dalilite server uses the Daliliteprogram, which performs structure comparison, wherein it optimizes a weighed sum of similarities of intermolecular distances. Sequence identity and root-mean-square deviation of C-alpha atoms after rigid-body superimposition are principal component to be outlined. It performs suboptimal alignments that do not superimpose the optimal alignment or each other. Suboptimal alignments observed by the program are reported only if the Z-score is above 2. In the C- alpha traces, the chains of the first (template) and second (target derived from modeler) structure are named 'Q'and 'S', respectively. The best match to each chain in the second structure is generally highlighted. Z-scores below 2 are found not to be significant.

ERRAT

ERRAT is a program for verifying protein structures determined by crystallography. Error values will be plotted as a function of the position of a sliding 9- residue window. According to the analysis by ERRAT, the final model is significantly improved relative to the initial model. This program examines a PDB file, and generates a score based on the quality of the local structure surrounding each residue, based on the typical ranges of dihedral angles and side chain contacts observed in real proteins, generally speaking, the method is sensitive to smaller errors than 3-D profile analysisbut is more forgiving than Procheck.

B. Superimposing the Model structure with the crystal structure

A correct sequence alignment is crucial for homology modeling. It can be achieved only when a structure based sequence alignment is available. Superimposition plays an important role in comparing two structure of the same protein, for example, one structure in complex with a ligand, and other without a ligand. Often such complex formation results in conformational transition, when the same part of the structure (domain or loop) moves to accommodate the ligand.

C. Docking method

The discovery of new drug has evolved from a random process of screening natural products that include component from computational and structural chemistry. The availability of high- resolution data on enzymes involved in critical metabolic pathways has triggered the development of techniques utilizing such data in the quest for novel compound for therapeutic relevance. Various algorithms have been developed to tackle the docking problem. A well-known example of such an algorithm is the DOCK program of Kuntz and co-worker.The DOCK algorithm - The fundamental version of this algorithm was developed by Kuntz *et al.,* at the University of California, USAand despite its long age, is still considered to be a successful means for solving the rigid docking problem and is used as a reference point to other algorithms. The algorithm is time consuming and thus suffers from difficulties in handling large surfaces. Therefore it is primarily used to interpret ligand-protein docking while trying to focus on 'interesting' sites on the surface of the molecules [20].

3. Results and Discussion

3.1. Analysis of the crystal structure

The crystallographic structure of 1uqt reveals the structure as two identical Chains A bound with Ligand UDP glucose and Chain B. The chain has 2 sheets, 6 beta hairpins, 1 psi loop, 3 beta bulges, 16 strands, 24 helices, 30 helix-helix interactions, 27 beta turns and 4 gamma turns[Figure 2].

3.2. Active site of TPS showing UDP binding cavity

The UDP binding site for chain A is described by TRP 85, HIS 54, GLN 185, ILE 225, ARG 262, LYS 267, PRO 297, HIS 338, PHE 339, ARG 341, LEU 344, ASP 361, GLY 362, MET 363, ASN 364, LEU 365 and GLU 369, HOH Z 52, HOH Z 52, HOH Z 69, HOH Z 70, HOH Z 70, HOH Z 72, HOH Z 73. The catalytic functions of this protein transferase are confirmed to amino acid residues ASP 361- HIS 154[Figure 3].

3.3. Output of Analysis of Verify 3D:[Figure 4a-4j]

3.4. Validation Results of Models/PROCHECK Results: [Figure 5][Figure 6], [Figure 7],[Figure 8a-8b]

3.5. DALILITE Results: [Figure 9],[Figure 10], [Figure 11]

3.6.ERRAT Results:[Figure 12]

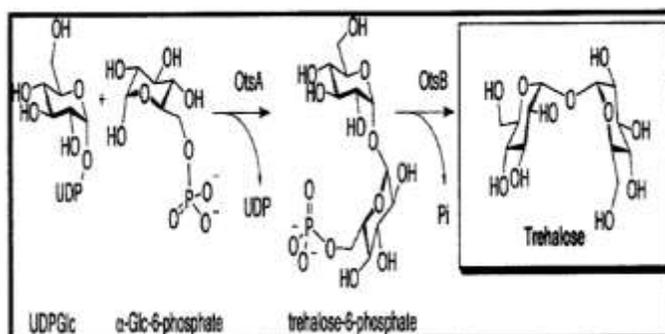**3.7. Patch Dock Results:** [Figure 13]

Figure 1. Biosynthesis of trehalose

Figure 1: Building blocks of trehalose are UDP-glucose and glucose 6-phosphate synthase

Figure 2. Trehalose-6-Phosphate Synthase (TPS) bound with the substrate ligand UDP-glucose

Figure 2: A ribbon diagram of Trehalose-6-Phosphate Synthase (TPS) bound with the substrate ligand UDP-glucose showing the arrangement of alpha helix and beta sheets.

Figure 3. UDP binding site of the chain A.
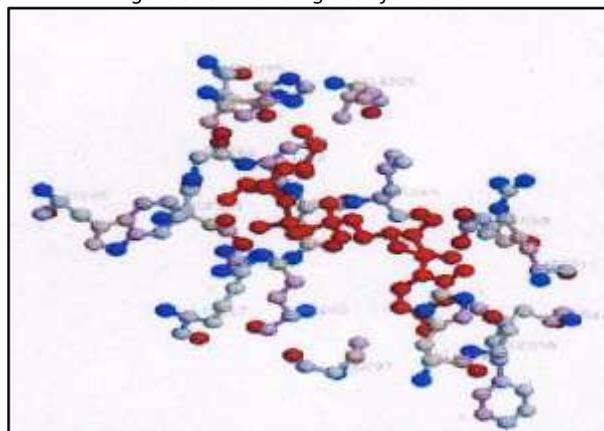
Figure 3: Active site of TPS showing UDP binding cavity

Figure 4a. Output of the Verify 3D for T6PS.B99990001.pd
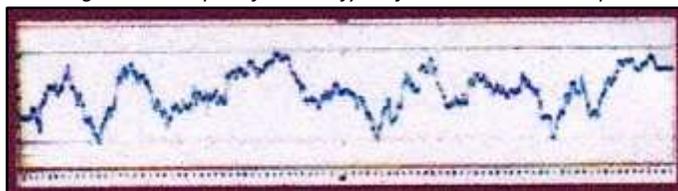
Figure 4b. Output of the Verify 3D for T6PS.B99990002.pd

Figure 4c. Output of the Verify 3D for T6PS.B99990003.pd

Figure 4d. Output of the Verify 3D for T6PS.B99990004.pd
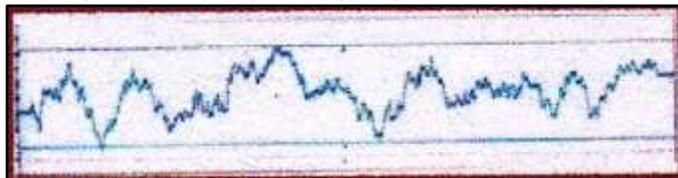
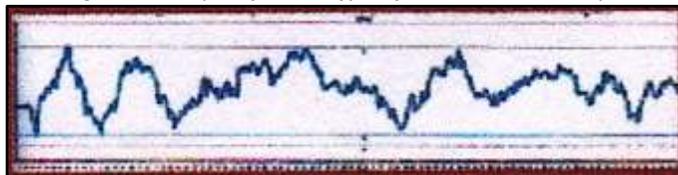*Figure 4e. Output of the Verify 3D for T6PS.B99990005.pd*

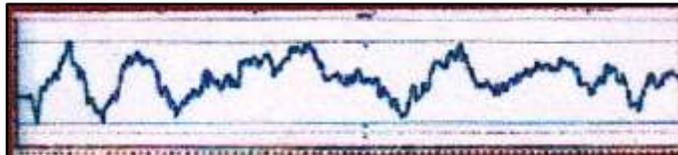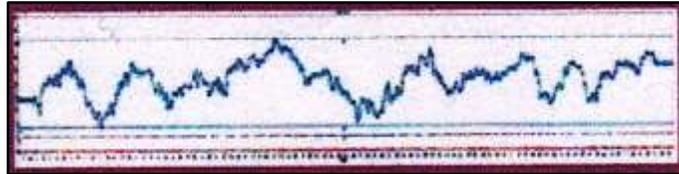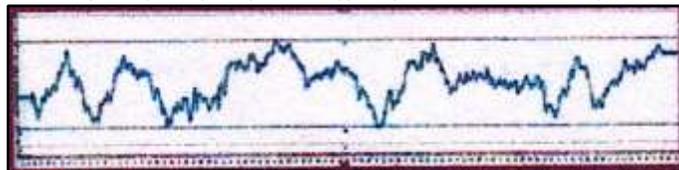*Figure 4f. Output of the Verify 3D for T6PS.B99990006.pd*

Figure 4g. Output of the Verify 3D for T6PS.B99990007.pd

Figure 4h. Output of the Verify 3D for T6PS.B99990008.pd

Figure 4i. Output of the Verify 3D for T6PS.B99990009.pd

Figure 4j. Output of the Verify 3D for T6PS.B99990010.pd

Figure 4a-4j: In these graphs, the X-axis shows the residue index present in the model protein and the Y-axis shows the pseudo energy. The upper red line shows the maximum frequency and the lower green line shows the minimum frequency of the model. The best model should be retrieved on the basis that the graph of the model should not cross the second bottom line. Since most of the models shown above are found to be good, selecting the best one becomes cumbersome. Hence in order to select the best among the better results various server results are validated.

Figure 5. Validation Results of Models

Protein ID (.pdb)	Favored Region		Allowed Region		Generously Allowed Regions		Disallowed regions		No of Residues (Excl. Gly ,Pro)	No. of Glycine	No. of Proline
	Residues	%	Residues	%	Residues	%	Residues	%			
T6PS.B99990001	360	90.9	30	7.6	4	1.0	2	0.5	2	35	24
T6PS.B99990002	361	91.2	27	6.8	7	1.8	1	1.0	2	35	24
T6PS.B99990003	356	89.9	33	8.3	3	0.8	4	1.0	2	35	24
T6PS.B99990004	364	91.9	27	6.8	3	0.8	2	0.5	2	35	24
T6PS.B99990005	363	91.7	25	6.3	6	1.5	2	0.5	2	35	24
T6PS.B99990006	361	91.2	29	7.3	5	1.3	1	0.3	2	35	24
T6PS.B99990007	360	90.9	28	7.1	6	1.5	2	0.5	2	35	24
T6PS.B99990008	363	91.7	26	6.6	4	1.0	3	0.8	2	35	24
T6PS.B99990009	358	90.4	32	8.1	4	1.0	2	0.5	2	35	24
T6PS.B99990010	363	91.7	26	6.6	4	1.0	3	0.8	2	35	24

Figure 5: The model T6PS.B99990004 shows the highest % and number of residues in the favored region

Figure 6. Validation Results of Models

Protein ID(.pdb)	Number of residues in favored region (~98.0% expected)		Number of residues in allowed region (~2.0% expected)		Number of residues in outlier region	
	Residues	%	Residues	%	Residues	%
T6PS.B99990001	432	94.9	16	3.5	7	1.5
T6PS.B99990002	432	94.9	13	2.9	10	2.2
T6PS.B99990003	430	94.5	19	4.2	6	1.3
T6PS.B99990004	436	95.8	13	2.9	6	1.3
T6PS.B99990005	430	94.5	18	4.0	7	1.5
T6PS.B99990006	429	94.3	16	3.5	10	2.2
T6PS.B99990007	431	94.7	14	3.1	10	2.2
T6PS.B99990008	429	94.3	18	4.0	8	1.8
T6PS.B99990009	433	95.2	12	2.6	10	2.2
T6PS.B99990010	431	94.7	15	3.3	9	2.0

Figure 6: The model T6PS.B99990004 displays the highest number of residues in the favored region (~98.0% expected)
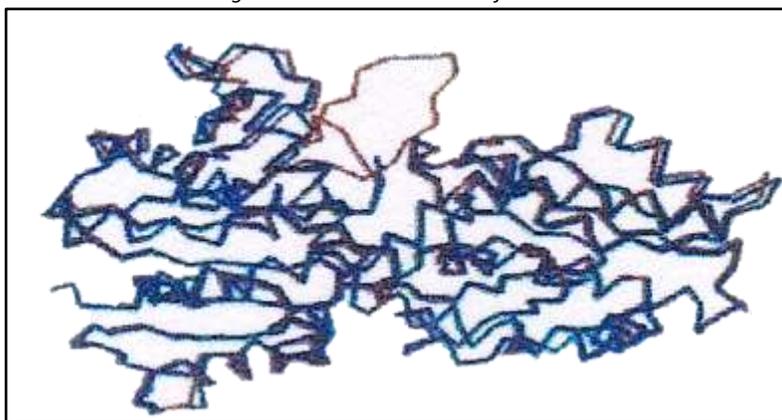
Figure 7. Validation Results of Models

Figure 7: Superimposed structure of 1uqt and the model T6PS.B99990004 depicting any deviation in the structural topology
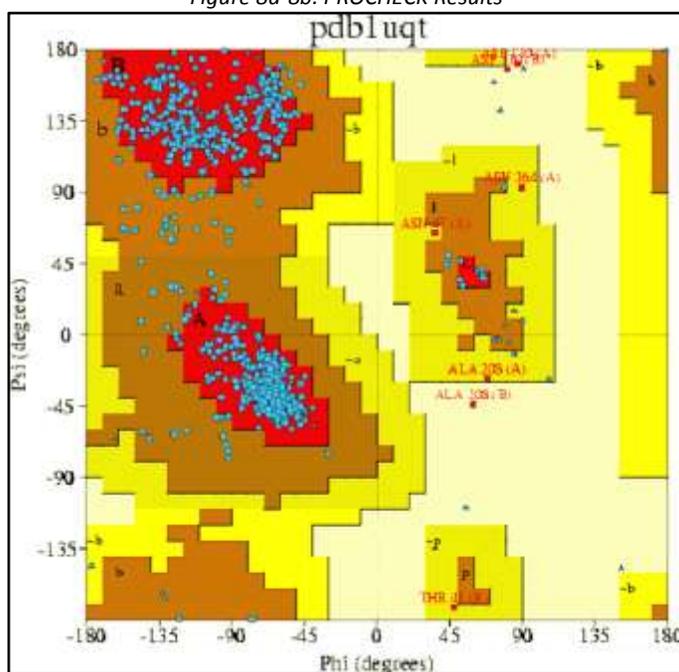
Figure 8a-8b. PROCHECK Results

Figure 8a: Ramachandran Plot of 1uqt showing that most of the residues lie in the allowed region

Figure 8b: Statistics of the Plot showing that ~90% residues lie in the favoured regions

Figure 9. DALILITE Results

Query Structure Submitted	PDB Template Used(1uqt)	Number of residues Used for alignment	Sequence ID (%)	No. of secondary Structure element	Z-Score	RMSD [A]
T6PS.B99990001	Chain B	443	38	31	61.6	0.9
T6PS.B99990002	Chain B	443	38	31	61.9	0.7
T6PS.B99990003	Chain B	443	38	31	62.2	0.6
T6PS.B99990004	Chain B	443	38	31	62.2	0.6
T6PS.B99990005	Chain B	443	38	31	61.7	0.7
T6PS.B99990006	Chain B	443	38	31	61.7	0.8
T6PS.B99990007	Chain B	443	38	31	61.9	0.7
T6PS.B99990008	Chain B	443	38	31	61.6	0.9
T6PS.B99990009	Chain B	443	38	31	62.0	0.7
T6PS.B99990010	Chain B	443	38	31	62.1	0.6

Figure 9: Model T6PS.B99990003and Model T6PS.B99990004 depicted as the best match results
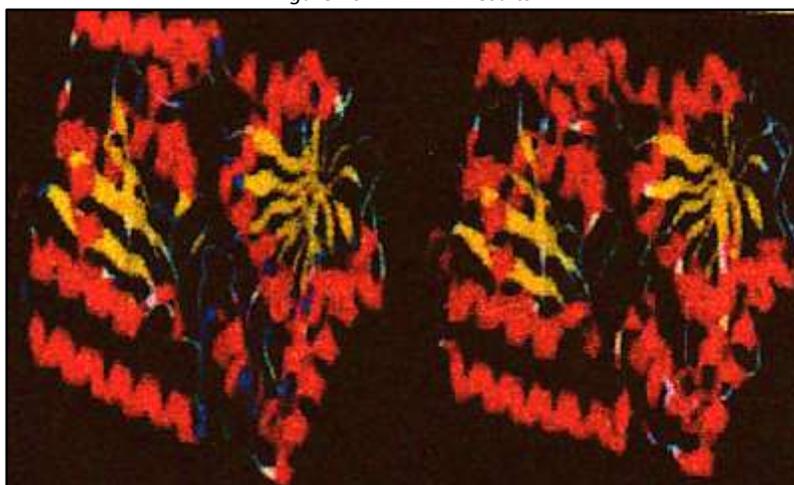
Figure 10. DALILITE Results

Figure 10: Ribbon structure of the Model T6PS.B99990004 and 1uqt
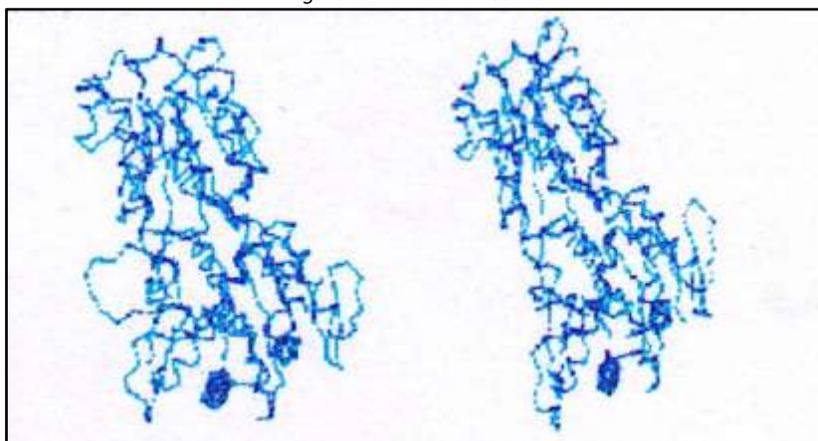
Figure 11. DALILITE Results

Figure 11: Backbone structure of the Model T6PS.B99990004 and 1uqt

Figure 12. Using various structure validation server results the best model among 10 models generated by MODELLER 6V2 was selected on the basis of various criteria specified by server results.

Protein ID	1UQT	T6PS 1	T6PS 2	T6PS 3	T6PS4	T6PS5	T6PS6	T6PS7	T6PS8	T6PS9	T6PS10
Quality factor	99.5	73.71	69.48	69.57	73.71	68.82	66.66	76.73	63.25	70.37	67.03

Figure 12: According to the results mentioned above, T6PS.B99990004, showing good results with high score in most of the server results was selected and used for docking ligand using Patch Dock.
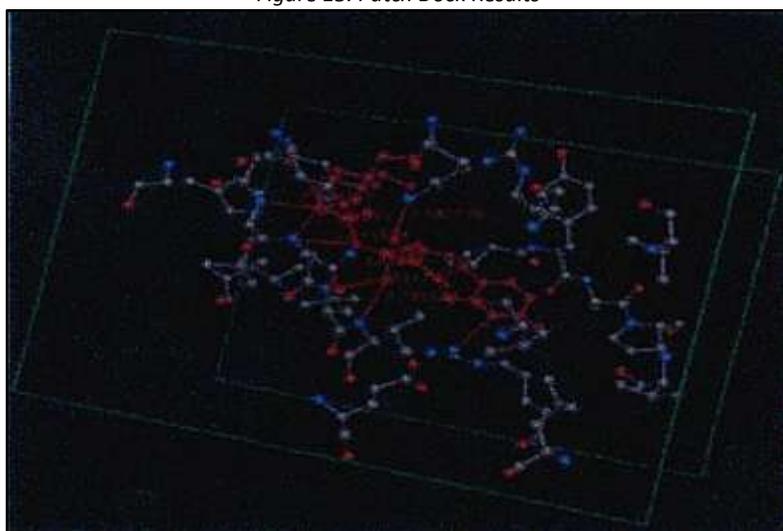
Figure 13. Patch Dock Results

Figure 13: UDP Glucose within the binding pocket of Modeled Protein T6PS.B99990004. Manual Grid Docking generated in Argus lab showing the UDP-glucose docked within the pocket (color by atom) generated by Patch Dock server. Grid points size calculated (28.41,28.251,20.1920), which generated a total Grid points of 273312 showed the van der wall energy for 105 different pose, of which the best pose having lowest energy value was shown. At best pose having lowest energy (-7.41) showed the best dock result, which elapsed at 7 seconds.

4. Conclusion

In absence of experimentally determined data, homology modeling proves to be promising in elucidating the structure by homology or comparative modeling, not undermining the associated shortcomings, which are apparent while modeling of loop regions.The Program MODELLER is found to be a very powerful protein-modeling environment, which features numerous algorithms for model constructionas it was used by many for homology modeling and found to be the best one. Large-scale comparative protein modeling of *E.coli* was made using MODELLER Programby an automated Modeling server Swiss-Model [21, 22, 23, 24, 25].

In the present investigation one such effort was taken to model the structure of a protein sequence retrieved from SWISSPROT data. This protein structure is framed by a homology modeling program Modeler 6v2 using 1uqt as template structure. As the input specified in the .top file resulted in 10 structures (1-10) of which single structure having best score value is selected after validating the structure using various validation servers.The selected model was superimposed onto the crystallographic structure to confirm the structural topology, the resulting structure showed some loops emerging out that signifies their evolutionary divergence. Loops often contribute to binding sites and determine the functional specificity of a given protein framework.However identifying the binding pocket of the modeled protein is essential for docking the ligand into its active site, for which the model structure was submitted to patch dock, a docking server. Out of 120 results, the best one was selected based on the high score. Binding Pocket was generated by Argus lab - a Molecule Building and Drug Docking Software. The binding pocket of the modeled protein is defined by 20 amino acids viz., PRO 14, TRP93, ILE 163, HIS 193, ILE232, ARG 269, LYS 274, VAL304, GLY 345, VAL 346, ARG348, LEU 351, TYR 355, ASP 368, GLY 369, LEU370, ASN 371, LEU372, VAL 373, and GLU 376.

5. Acknowledgement

6. References

1. E.Cabib, L.F.Leloir.The biosynthesis of Trehalose phosphate.*The Journal of Biological Chemistry.* 1958; 231, 259-275.
2. Ines Delorge, Michal Janiak, SebastienCarpentier, Patrick Van Dijck. Fine tuning of trehalose biosynthesis and hydrolysis as novel tools for the generation of abiotic stress tolerant plants.*Frontiers in plant science.*2014; 5.
3. Juan Carlos Argüelles. Physiological roles of trehalose in bacteria and yeasts: a comparative analysis. *Archives of Microbiology.*2000*;* 174(4), 217-224.
4. Olivier Fernandez, Linda Béthencourt, Anthony Quero, Rajbir S. Sangwan, Christophe Clément. Trehalose and plant stress responses: friend or foe?.*Trends in plant science.*2010; 15(7), 409-417.
5. Robert P. Gibson, Chris A. Tarling, Shirley Roberts, Stephen G. Withers, Gideon J. Davies. The donor subsite of trehalose-6-phosphate synthase binary complexes with UDP-glucose and UDP-2-deoxy-2-fluoro-glucose at 2 Å resolution. *The Journal of Biological Chemistry.*2004; 279(3), 1950-1955.
6. S.Ilhan, F. Ozdemir, M. Bor. Contribution of trehalose biosynthetic pathway to drought stress tolerance of CapparisovataDesf.*Plant Biology*. 2015; 17(2), 402-407.
7. Walter Bell, Weining Sun, Stefan Hohmann, StefaanWera, AnkeReinders, Claudio De Virgilio, Andres Wiemken, Johan M. Thevelein. Composition and functional analysis of the Saccharomyces cerevisiae trehalose synthase complex.*TheJournal of Biological Chemistry.*1998; 273(50), 33311-33319.
8. ZeliaSilva, Susana Alarico, Milton S. da Costa. Trehalose biosynthesis in Thermusthermophilus RQ-1: biochemical properties of the trehalose-6-phosphate synthase and trehalose-6-phosphate phosphatase.*Extremophiles.* 2005; 9(1), 29-36.
9. Alan D.Elbein, Y. T. Pan, Irena Pastuszak, David Carroll. New insights on trehalose: a multifunctional molecule.*Glycobiology.* 2003; 13(4), 17R-27R.
10. Gabriel Iturriaga, Ramón Suárez, Barbara Nova-Franco. Trehalose metabolism: from osmoprotection to signaling.*International journal of molecular sciences*.2009; 10(9), 3793-3810.
11. What is MODELLER?http://salilab.org/modeller/8v0/manual/node7.html. Date accessed: May 2015.
12. Running MODELLER. http://structbio.vanderbilt.edu/comp/soft/modeller/manual/node14.html. Date accessed: May 2015.
13. Andrej Šali, Tom L. Blundell. Comparative protein modelling by satisfaction of spatial restraints.*Journal of molecular biology.*1993; 234(3), 779-815.
14. Andrej ŠAli, John P. Overington. Derivation of rules for comparative protein modeling from a database of protein structure alignments.*Protein Science.* 1994; 3(9), 1582-1596.
15. A.D.MacKerell, D.Bashford, M. L. D. R. Bellott, R. L. Dunbrack, J.D. Evanseck, M.J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F.T.K. Lau, C. Mattos, S. Michnick, T. Ngo, D.T. Nguyen, B. Prodhom, W.E. Reiher, B. Roux, M. Schlenkrich, J.C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiorkiewicz-Kuczera, D. Yin, M. Karplus. All-atom empirical potential for molecular modeling and dynamics studies of proteins.*The Journal of Physical Chemistry B*. 1998; 102(18), 3586-3616.
16. Benjamin Webb, Andrej Sali. Comparative protein structure modeling using Modeller.*Current protocols in bioinformatics*.2014; 5-6.
17. W.Braun, N. Gō. Calculation of protein conformations by proton-proton distance constraints: A new efficient algorithm.*Journal of molecular biology.*1985; 186(3), 611-626.
18. Basic modeling tutorial – Andrej Sali Lab. http://salilab.org/modeller/tutorial/basic.html. Date accessed: May 2015.
19. Method for comparative structure modeling by MODELLER. http://www.msg.ucsf.edu/local/programs/modeller/node11.html. Date accessed: May 2015.
20. Irwin D. Kuntz, Jeffrey M. Blaney, Stuart J. Oatley, Robert Langridge, Thomas E. Ferrin. A geometric approach to macromolecule-ligand interactions.*Journal of molecular biology.*1982; 161(2), 269-288.
21. A.Šali, R. Sánchez, A. Y. Badretdinov, A. Fiser, F. Melo, J. P. Overington, E. Feyfant, M. A. Martı-Renom. MODELLER, a protein structure modeling program, release 5.*URL http: 00guitar. rockefeller. edu0* (1999).

22. Manuel C. Peitsch. Large Scale Protein Modeling and Model Repository.*American Association for Artificial Intelligence*.1997; 234-236.
23. Nicolas Guex, Manuel C. Peitsch. SWISS-MODEL and the Swiss-Pdb Viewer: an environment for comparative protein modeling.*Electrophoresis.*1997; 18(15), 2714-2723.
24. Sarah A.Osmani, SørenBak, Birger Lindberg Møller. Substrate specificity of plant UDP-dependent glycosyltransferases predicted from crystal structures and homology modeling.*Phytochemistry.* 2009; 70(3), 325-347.
25. R. O. B. E. R. T. O.SAnchez, Andrej Sali. Large-scale protein structure modeling of the Saccharomyces cerevisiae genome.*Proceedings of the National Academy of Sciences.*1998; 95, 23, 13597-13602.