

THERMOSTABILIZATION OF PANOMYCOCIN, A NOVEL EXO-BETA-1,3-
GLUCANASE ISOLATED FROM *PICHIA ANOMALA* NCYC 434, BY USING
EXCIPIENTS AND COMPUTATIONAL METHODS

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES
OF
MIDDLE EAST TECHNICAL UNIVERSITY

BY

MUHAMMED TILAHUN MUHAMMED

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR
THE DEGREE OF MASTER OF SCIENCE
IN
BIOMEDICAL ENGINEERING

DECEMBER 2015

Approval of the thesis:

THERMOSTABILIZATION OF PANOMYCOCIN, A NOVEL EXO-BETA-1,3-GLUCANASE ISOLATED FROM *PICHTIA ANOMALA* NCYC 434, BY USING EXCIPIENTS AND COMPUTATIONAL METHODS

submitted by **MUHAMMED TILAHUN MUHAMMED** in partial fulfillment of the requirements for the degree of **Master of Science in Department of Biomedical Engineering**, Middle East Technical University by,

Prof. Dr. Gülbin Dural Ünver

Director, Graduate School of **Natural and Applied Sciences**

Prof. Dr. Hakan I. Tarman

Head of Department, **Biomedical Engineering**

Prof. Dr. Fatih İzgü

Supervisor, **Biology Department, METU**

Assoc. Prof. Dr. Çağdaş D. Son

Co-Supervisor, **Biology Department, METU**

Examining Committee members:

Prof. Dr. Semra Kocabıyık

Biology Dept., METU

Prof. Dr. Fatih İzgü

Biology Dept., METU

Prof. Dr. Ayşegül Ç. Gözen

Biology Dept., METU

Prof. Dr. Cumhuri Çökmüş

Biology Dept., KFAU

Assoc. Prof. Dr. Çağdaş D. Son

Biology Dept., METU

Date: 30.12.2015

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Last name: Muhammed Tilahun, Muhammed

Signature:

ABSTRACT

THERMOSTABILIZATION OF PANOMYCOCIN, A NOVEL EXO-BETA-1,3-GLUCANASE ISOLATED FROM *PICHIA ANOMALA* NCYC 434, BY USING EXCIPIENTS AND COMPUTATIONAL METHODS

Muhammed, Muhammed Tilahun

M.S., Department of Biomedical Engineering

Supervisor: Prof. Dr. Fatih İzgü

Co-Supervisor: Assoc. Prof. Dr. Çağdaş D. Son

December 2015, 126 pages

As the risks for fungal infections increased, the prevalence of invasive fungal infections increased. Therefore, the demand for antifungal agents has risen. Moreover, the currently used antifungal agents have serious side effects and resistance development resulting from their mechanisms of action. Thus, novel antifungal agents with mechanisms that will not affect the host mammalian cells are in need. Yeast killer proteins which are naturally occurring toxins are good candidates for such types of agents. Panomycocin is an example for this type of killer proteins. Panomycocin is a killer toxin of *Pichia anomala* NCYC 434 (K5). It has exo- β -1,3-glucanase activity. It kills the sensitive cells by hydrolyzing β -1,3-glucans that are crucial in maintaining the integrity of fungal cell wall. However, its activity decreases above 37°C. In this work various types of excipients were used to increase the thermostability of Panomycocin so that it would be active at higher temperatures. Gradient concentrations of these excipients along with the protein were tested on *Saccharomyces cerevisiae* NCYC 1006 at increasing temperatures. If the excipients tested increase the thermostability, this will give us the opportunity to choose suitable excipients that can be used in the formulation of Panomycocin as a novel antifungal

drug. Since the effect of excipients on thermostability is limited, computational methods were also used to design a thermostable protein at much higher temperatures. Homology modeling of the protein was performed first. After the binding site of the protein was predicted, the best thermostabilizing positions in the model generated were detected utilizing various computer programs and servers. Although the excipients tested did not increase the thermostability of the protein, we found the best amino acid residues in the model whose substitutions can increase the thermostability of the protein to the desired level. In the allosteric part of the protein Leu52Arg, Phe223Arg and Gly254Arg were found to be the best thermostabilizing mutations with 6.26 K, 6.26 K and 8.27 K temperature increases respectively. In the binding site Glu186Arg was found to be the best thermostabilizer mutation with 9.58 K temperature increase. Using the results mutant thermostable protein can be obtained and this will enable the formulation of Panomycocin as a novel antifungal drug with high thermostability.

Keywords: Computational; excipient; homology modeling; Panomycocin; thermostability.

ÖZ

***PICHIA ANOMALA* NCYC 434'TEN İZOLE EDİLMİŞ YENİ BİR EKZO-BETA-1,3-GLUKANAZ OLAN PANOMİKOSİN'İN YARDIMCI MADDELER VE BİLGİSAYAR PROGRAMLAR KULLANILARAK TERMOSTABİLİTESİNİN ARTIRILMASI**

Muhammed, Muhammed Tilahun

Yüksek Lisans, Biyomedikal Mühendisliği Bölümü

Tez Yöneticisi: Prof. Dr. Fatih İzgü

Ortak Tez Yöneticisi: Doç. Dr. Çağdaş D. Son

Aralık 2015, 126 sayfa

Mantar enfeksiyonlarına neden olan risklerin artmasıyla, invaziv mantar hastalığı yaygınlaşmıştır. Bundan dolayı antifungal ilaçlarına talep artmıştır. Ayrıca şu anda kullanılmakta olan antifungal ilaçların etki mekanizmalarından dolayı ciddi yan etkileri bulunmaktadır ve onlara karşı direnç gelişmektedir. Dolayısıyla etki mekanizmaları memeli hücrelerini etkilemeyen yeni antifungal ilaçlara ihtiyaç duyulmaktadır. Doğal ortamda oluşan toksik, maya öldürücü proteinler bunun gibi ilaçlara iyi bir örnek teşkil etmektedir. Panomikosin de bu tür öldürücü proteinlere örnektir. Panomikosin *Pichia anomala* NCYC 434 maya suşunun (K5) öldürücü toksinidir. Ekzo- β -1,3-glukanaz aktivitesi bulunmaktadır. Mantar hücresinin duvar bütünlüğünün korunmasında önemli olan β -1,3-glukanı hidrolize ederek proteine karşı hassas olan hücreleri öldürür. Ancak 37°C'den sonra proteinin aktivitesi azalmaktadır. Bu çalışmada Panomikosin'nin termostabilitesinin çeşitli yardımcı maddeler kullanılarak artırılmasıyla daha yüksek sıcaklıklarda etkin olması amaçlanmıştır. Çeşitli konsantrasyonlarda hazırlanan yardımcı maddeler protein ile birlikte *Saccharomyces cerevisiae* NCYC 1006 üzerinde sıcaklığı artırarak denenmiştir. Yardımcı maddeler protein termostabilitesini arttırmasını sağlarsa Panomikosin'in

yeni bir antifungal ilaç olarak formülasyon geliştirilmesinde kullanılacak yardımcı maddeleri seçme fırsatı verecektir. Yardımcı maddelerin termostabilite üzerindeki etkileri sınırlı olduğundan daha yüksek sıcaklıklarda termostabil olan bir protein tasarlamak için bilgisayar programlar da kullanılmıştır. Önce proteinin homoloji modellenmesi yapılmıştır. Proteinin bağlanma noktaları bulunduğundan sonra çeşitli programlar ve serverler kullanılarak geliştirilmiş model üzerindeki en iyi termostabilize eden noktalar tespit edilmiştir. Test edilen yardımcı maddelerin proteinin termostabilitesini artırmamasına rağmen mutasyon ile termostabiliteyi istenilen dereceye artırabilen model üzerindeki en iyi amino asit noktaları tespit edilmiştir. Proteinin alosterik bölgesinde Leu52Arg, Phe223Arg ve Gly254Arg mutasyonları sırasıyla 6.26 K, 6.26 K ve 8.27 K sıcaklık artışı sağlayabilecek en iyi termostabilize eden noktalar oldukları bulunmuştur. Bağlanma noktasında Glu186Arg mutasyonu 9.58 K sıcaklık artışı sağlayabilecek en iyi termostabilize eden nokta olduğu bulunmuştur. Sonucu kullanarak mutant termostabil protein elde edilebilir ve bu da termostabilitesi yüksek olan yeni bir antifungal ilaç olan Panomikosin formülasyonunun yapılmasında yardımcı olacaktır.

Anahtar kelimeler: Bilgisayar programlar; homoloji modelleme; Panomikosin; termostabilite; yardımcı maddeler.

To My Mother

ACKNOWLEDGEMENTS

The author wishes to express his deepest gratitude to his supervisor Prof. Dr. Fatih IZGÜ for his guidance, advice and insight throughout the research. The author would also like to thank his cosupervisor Assoc. Prof. Dr. Çağdaş D. Son for his guidance and suggestions.

The author would like to thank those who provided the technical support in the Lab works. The author would also like to thank developers of the programs and servers used in this work for providing free service for researchers.

TABLE OF CONTENTS

ABSTRACT	v
ÖZ	vii
ACKNOWLEDGEMENTS	x
TABLE OF CONTENTS	xi
LIST OF TABLES	xv
LIST OF FIGURES	xvi
LIST OF SYMBOLS AND ABBREVIATIONS	xviii
CHAPTERS	
1. INTRODUCTION	1
1.1 Antifungal Agents	1
1.2 Resistance Development and Side Effects of Currently Used Antifungal Agents.....	4
1.3 Yeast Killer Proteins and Their Applications.....	5
1.3.1 Structure, Processing and Secretion of Yeast Killer Proteins	6
1.4 K5 Type Yeast Killer Protein, Panomycocin	7
1.5 Protein Thermostability	9
1.6 Excipients	12
1.7 Homology Modeling	15
1.7.1 Steps in Homology Modeling	15
1.7.2 Softwares for Homology Modeling	21
1.7.2.a MODELLER	21
1.7.2.b I-TASSER	21
1.7.2.c SWISS MODEL	22
1.7.2.d PrISM	22
1.7.2.e ORCHESTRAR	23
1.7.2.f MOE	23
1.7.2.g COMPOSER	23
1.7.2.h ROSETTA.....	24
1.7.3 Applications of Homology Modeling	24

1.8	Comparison of Models	27
1.8.1	DOPE	27
1.8.2	TM Score.....	27
1.8.3	RMSD	28
1.9	Molecular Dynamics and Energy Calculations	29
1.9.1	GROMOS.....	29
1.9.2	GROMACS	30
1.9.3	CHARMM	30
1.9.4	AMBER	30
1.9.5	OPLS	31
1.10	Protein Stability Predictors.....	31
1.10.1	CNA	32
1.10.2	CUPSAT	32
1.10.3	I-MUTANT	33
1.10.4	FoldX	33
1.10.5	AUTO-MUTE.....	34
1.10.6	MUpro.....	34
1.10.7	Eris	34
1.11	Aim of the Study	35
2.	MATERIALS AND METHODS	37
2.1	Materials	37
2.1.1	Strains	37
2.1.2	Culture Media	37
2.1.3	Chemicals.....	38
2.1.4	Buffers and Solutions.....	38
2.1.5	Programs and Servers	38
2.2	Laboratory Methods	38
2.2.1	Sterilizations	38
2.2.2	Maintainance of the Yeast Cultures.....	38
2.2.3	Production of Panomycocin.....	39
2.2.4	Preparation of Crude Panomycocin	39
2.2.5	Determination of Killer Activity of Panomycocin	40
2.2.6	Panomycocin Purification with Gel Filtration Chromatography.....	41
2.2.7	Determination of Protein Concentration.....	41
2.2.8.	SDS Polyacrylamide Gel Electrophoresis	41

2.2.9 Protein Detection in Gels by Coomassiae Brilliant Blue Staining	44
2.2.10 Thermostability Tests by Using Excipients	44
2.3 Computational Methods	45
2.3.1 Protein Similarity Search	45
2.3.2 Determination of Signal Peptide and KEX2 Cleavage Site.....	45
2.3.3 Homology Modeling.....	46
2.2.3.a Homology Modeling by MODELLER.....	46
2.3.3.a.1 Generating Models	46
2.3.3.a.2 Comparing the Generated Models	48
2.3.3.a.3 Loop Modeling	50
2.3.3.a.4 Optimization, Verification and Validation	50
2.2.3.b Homology Modeling by I-TASSER	51
2.3.4 Determination of Binding Site	51
2.3.5 Thermostabilization of Models	52
3. RESULTS	55
3.1 Laboratory Results	55
3.1.1 Production of Crude Panomycocin	55
3.1.2 Determination of Killer Activity of Panomycocin	55
3.1.3 Isolation of Panomycocin	56
3.1.4 Determination of Protein Concentration.....	57
3.1.5 SDS Polyacrylamide Gel Electrophoresis	57
3.1.6 Results of Thermostability Tests Using Excipients.....	59
3.2 Computational Results	62
3.2.1 Protein Similarity Search	62
3.2.2 Signal Peptide and KEX2 Cleavage Site	63
3.2.3 Homology Modeling.....	65
2.2.3.a Homology Modeling by MODELLER.....	65
3.2.3.a.1 Generating Models	65
3.2.3.a.2 Determination of the Best Model	70
3.2.3.a.3 Loop Modeling	72
3.2.3.a.4 Optimization, Verification and Validation	75
2.2.3.b Homology Modeling by I-TASSER	77
3.2.4 Binding Sites.....	78
3.2.5 Thermostabilization of the Models	80
4. DISCUSSION	87

5. CONCLUSION	99
REFERENCES.....	101
APPENDICES	
A. CHEMICALS AND THEIR SUPPLIERS.....	121
B. BUFFERS AND SOLUTIONS.....	123
C. PROGRAMS AND SERVERS.....	125

LIST OF TABLES

TABLES

Table 1.1 Recent applications of protein engineering or thermostabilization.....	11
Table 1.2 Pharmaceutical excipients commonly used in protein formulations	14
Table 1.3 Recent applications of homology modeling.....	26
Table 2. 1 Separating gel mixture	43
Table 2. 2 Stacking gel mixture.	43
Table 3.1 Results of the killing activity of the protein with excipients at various temperatures.	60
Table 3.2 TM scores for the models.....	71
Table 3.3 RMSD values of the models.	72
Table 3.4 DOPE score and energies of loop modeling at different positions.....	73
Table 3.5 The best ten stabilizing positions in the allosteric region of the loop model	80
Table 3.6 Combinations that gave a higher energy.....	81
Table 3.7 The best ten stabilizing positions in the I-TASSER model.	85
Table B.1 SDS PAGE Gel Components and Staining Solutions.	123
Table C.1 Programs and servers used.	125

LIST OF FIGURES

FIGURES

Figure 1.1 Mechanisms of currently used antifungal agents and their cellular targets.	3
Figure 1.2 Secretory Pathway of killer Toxin K28 in <i>S. cerevisiae</i> .	7
Figure 1.3 pH stability of Panomycocin.	8
Figure 1.4 Steps of homology modeling.	16
Figure 1.5 The two zones of sequence alignments.	17
Figure 1.6 Applications of homology modeling in drug discovery and its outline.	25
Figure 2.1 Ultrafiltration system for Panomycocin.	40
Figure 2.2 Query sequence and templates loaded into EasyModeller.	47
Figure 2.3 Models loaded to VMD.	49
Figure 3.1 Killer activity of Panomycocin.	56
Figure 3.2 Elution profile of Panomycocin on TSK G2000SW column.	57
Figure 3.3 Denatured SDS-PAGE.	58
Figure 3.4 Results of thermostability tests with excipients.	61
Figure 3.5 FASTA search results of matrices with high target identity.	62
Figure 3.6 FASTA search results of BLOSUM50 which has low target identity.	62
Figure 3.7 Signal peptide results of SignalP 4.1	64
Figure 3.8 KEX2 cleavage site estimation of SMART server.	65
Figure 3.9 PSI-BLAST result of the protein.	66
Figure 3.10 Alignment of the templates.	67
Figure 3.11 Alignment of the query with the templates.	68
Figure 3.12 Generated models and their DOPE profile.	69
Figure 3.13 Graph of DOPE score of the models.	70

Figure 3.14 Graph of energy of the models.	71
Figure 3.15 DOPE score profile of model 1 relative to the templates	73
Figure 3.16 DOPE profile of model 1 (turquoise) and the model after loop modeling was undertaken (red).	74
Figure 3.17 ERRAT generated result of the loop model where 95 % indicates rejection limit.	75
Figure 3.18 Verify 3D result generated for the loop model.....	75
Figure 3.19 3D structure of the best model after loop modeling.	76
Figure 3.20 The match between MODELLER's best model (blue) and I-TASSER model (gray).	77
Figure 3.21 A) The table of the area and the volume for different active sites of the model. B) The three dimensional structure of the best active site. C) Binding site analysis by CASTp server. Green color illustrates the binding site position.....	79
Figure 3.22 Global indices for the thermal unfolding of the loop model.	82
Figure 3.23 CNA result of the loop model.....	83
Figure 3.24 CNA result after Leu52Arg mutation was done.	83
Figure 3.25 MUpro result of Leu52Arg mutation.....	84
Figure 4.1 Surface of the model generated	97

LIST OF SYMBOLS AND ABBREVIATIONS

APS	Ammonium Persulfate
BLAST	Basic Local Alignment Search Tool
BSA	Bovine Serum Albumin
CASTp	Computed Atlas of Surface Topography of Proteins
CNA	Constraint Network Analysis
DOPE	Discrete Optimized Protein Energy
FPLC	Fast Protein Liquid Chromatography
GROMOS	GRoningen MOlecular Simulations
I-TASSER	Iterative Threading ASSEmbly Refinement
K	Kelvin
kDa	Kilo dalton
NCBI	National Center for Biotechnology Information
NCYC	National Collection of Yeast Cultures
NMR	Nuclear Magnetic Resonance
°C	Degree Celsius
PDB	Protein Data Bank
RMSD	Root Mean Square Deviation
SAVES	Structure Analysis and Verification Server
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
SMART	Simple Modular Architecture Research Tool
SPDBViewer	Swiss Protein Data Bank Viewer
TEMED	Tetramethylethylenediamine
T_m	The temperature at which half of the unfolding occurs
TM	Template Modeling
VMD	Visual Molecular Dynamics
YEPD	Yeast Extract Peptone Dextrose

CHAPTER 1

INTRODUCTION

Participation of fungi in the aetiology of infections has increased significantly [1,2]. Extensive use of wide spectrum antibiotics for a long time, acquired resistance development of pathogenic fungi towards commonly used drugs, immunosuppressive diseases like AIDS, cancer and organ transplant patients, the use of chemotherapeutic drugs, the use of corticoids for a long time and invasive medical interventions increased the risk of fungal infections [3-5]. Furthermore advances in medical technology has increased the survival of patients with harsh and life threatening diseases has led to a rapid increase in the immunosuppressed population [6]. There is a correlation between these changes and the high increase in the rate of invasive fungal infections. World Health Organization (WHO) reports also show the increase in the prevalence of these infections in the last decades. As a result the demand for antifungal agents to combat with these infections has risen. According to BCC Research report the world market for human antifungal agents was approximately 11.6 billion USD in 2012 and 11.8 billion USD in 2013. The global market is expected to grow approximately to 13.9 billion USD in 2018 with a compound annual growth rate of 3.2% from 2013 to 2018 [7].

1.1 Antifungal Agents

Developments in this area started with the discovery of griseofulvin, which was the earliest chemical substance declared to show inhibition of fungi in a selective manner. In 1955 a significant development was amphotericin B and in 1961 flucytosine was reported as an antifungal agent. The discovery of the azole groups was an important

advance in the area. In 1981 ketoconazole was introduced as an oral agent which is an important advance in the exploration of new antifungal agents for the treatment of systemic fungal infections. In the early 2000s echinocandins entered the market. These antifungal agents are natural products which are safe, have minimal drug interactions and good pharmacokinetics [8-10].

Currently used antifungal drugs can be generally classified according to their mechanisms of action (figure 1.1).

Antifungal agents that target cell membrane specifically act on ergosterol, the major sterol of the cell membrane of many pathogenic fungi and essential for membrane structural unity. They bring their actions by inhibiting 14- α -demethylase in the biosynthetic pathway of ergosterol which is important in the conversion of lanosterol to ergosterol; binding to ergosterol that results in the formation of porin channels and leakage of plasma components. Hence it causes transmembrane potential loss, problematic cellular function and ergosterol biosynthesis inhibition. Azoles, polyenes, allylamines and morpholines act in this way.

Those that target cell wall bring their action by binding and inhibiting β -1,3-D-glucan synthase enzyme complex responsible for β -1,3-D-glucan polysaccharides synthesis, the crucial structural components of the cell wall in various commonly seen pathogenic fungi. As a result of the actions of those agents, fungal cells can not maintain their integrity as a result of the loss of glucans which leads to osmotic stress and later to fungal cell lysis, especially in cells that are under rapid growth. Echinocandins are examples for this.

Pyrimidine analogues exert their action by interfering with the metabolism of pyrimidine which impairs DNA, RNA and protein synthesis. They are taken up into the fungal cell selectively, deamination into 5-fluorouracil (5-FU) with cytosine deaminase and then conversion into 5-fluorouridylic acid takes place. Finally this is phosphorylated by uracil phosphoribosyl transferase that inhibits thymidylate synthase. Incorporation of 5-FU into RNA causes RNA miscoding, early chain termination and disruption of protein synthesis. There is also conversion of 5-

Fluorouracil to 5-fluorodeoxyuridine monophosphate which is an inhibitor of thymidylate synthase. Thymidylate synthase takes part in DNA synthesis and nuclear division and thus its inhibition subsequently causes disruption of DNA synthesis. Mitotic inhibitors act by binding to tubulin and interfering with microtubule formation [11-15].

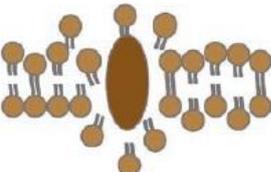
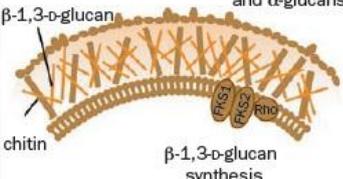
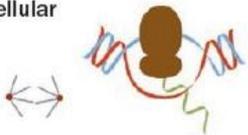
Mechanism	Drug class	Drugs
Cell membrane  <i>Ergosterol inhibitors/binders</i>	Azoles (14- α -demethylase inhibitors)	Imidazoles Ketoconazole, miconazole Triazoles Fluconazole, itraconazole, voriconazole, posaconazole, isavuconazole
	Polyenes (ergosterol binding)	Amphotericin B
	Allylamines (squalene monooxygenase)	Terbinafine
Cell wall  β -1,3-D-glucan Mannoproteins and α -glucans chitin β -1,3-D-glucan synthesis	Echinocandins (β -1,3-D-glucan synthesis inhibitors)	Anidulafungin, caspofungin, micafungin
Intracellular 	Pyrimidine analogues/ thymidylate synthase inhibitor	Flucytosine
	Mitotic inhibitor	Griseofulvin

Figure 1.1 Mechanisms of currently used antifungal agents and their cellular targets [13].

1.2 Resistance Development and Side Effects of Currently Used Antifungal Agents

Antifungal drug resistance is the absence of sensitivity of pathogenic fungi to an antifungal drug resulting in the failure of antifungal treatments. Primary, secondary and clinical antifungal resistances are types of resistance. Primary resistance is present naturally among specific fungi species without exposure to antifungal drug before. Secondary resistance is gained by initially susceptible fungal strains after exposure to the antifungal drug. Clinical resistance is the failure of antifungal therapy or reoccurring of an infection by a fully sensitive fungi to the antifungal drug used [14,15].

Antifungal resistance can develop specifically by increased efflux of the drug by up regulation of multidrug transporter genes leading to decreased drug concentrations; alteration of target site due to mutations that reduces the binding of the drug to the target site; up regulation of target enzyme resulting from gene amplification, high rate of transcription or low rate of the degradation of the gene product; development of bypass pathways in the ergosterol biosynthetic pathway due to the change of particular steps; reduced uptake of drug due to decreased permeability (polyenes and pyrimidine analogues); decreased conversion to toxic antimetabolites (pyrimidine analogues) [12,14,16].

Hepatotoxicity, GI (gastrointestinal) disturbances, nephrotoxicity, fever, headache, chills, phlebitis, anaemia, bone marrow suppression, pruritus, thrombophlebitis and Stevens-Johnson syndrome are the common side effects of currently used antifungal agents [13,17].

In addition to resistance and many side effects there is unfavourable interaction with other medications, their spectrum of activity is low, limited formulation, many of them are fungistatic. This is mainly because fungi cells are eukaryotic and therefore they have common biochemical pathways and subcellular structures with mammalian cells. Antifungal drugs which target cell wall components of fungi have some special superiorities over the action mechanisms of the other antifungal agents because they

act selectively on the fungi and the acquired resistance development against the other antifungal drugs is low. Their wide spectrum fungicidal activity with minimal effect on mammalian cells and their potential to act on resistant strains which are now appearing due to therapy using the current antifungal drugs with a low tendency to raise resistance makes them a promising, attractive novel group of antifungal agents [11,18-20].

1.3 Yeast Killer Proteins and Their Applications

Yeast killer proteins are naturally produced toxins which are secreted by killer yeast strains. They kill sensitive yeast cells and their related genera but they do not kill their own genera resulting from their autoimmunity mechanisms. They are low molecular mass proteins that are lethal to sensitive yeast and fungal strains without cell–cell contact. Studied yeast killer proteins are generally sensitive to heat, susceptible to proteases, and show their activity within narrow pH and temperature ranges, and display their action at acidic environments [21-24].

Based on their killing spectra and immunity cross reactions of the strains, yeast killer proteins are classified into eleven groups (K1-K11) [25,26]. In addition to killer yeasts there are non-killer strains that do not secrete a killer protein but immune to a particular killer protein.

There are various yeast killer toxins with different modes of action, structures and growth processes. They mainly exert their action by hydrolyzing or inhibiting the synthesis of β -1,3-glucans which are the major cell wall components and forming ion channels on plasma membrane leading to ion leakage. Furthermore they act by inhibiting the DNA synthesis and blocking the cell cycle in G1 phase [27].

Yeast killer proteins have been used as starter culture to combat with contaminating yeasts in the production of bread, beer and wine, biocontrol systems for food preservation and as secretion vectors. In addition to biotechnological applications, yeast killer proteins have been suggested as novel, attractive potential antifungal agents [22,23,27].

1.3.1 Structure, Processing and Secretion of Yeast Killer Proteins

Toxin secretion pathways of K1 and K28 which are secreted by *S. cerevisiae* are fully identified. Although these two killer proteins are different in amino acid compositions and mechanisms of action, their synthesis, processing and secretion displays significant similarities. K1 and K28 toxins consist of two distinct disulfide bonded unglycosylated subunits, termed α and β and these domains flank a segment called γ , which is not part of the mature toxin and assumed to be the immunity determinant. Killer toxins are initially translated as preprotoxin which undergoes post-translational modifications within the endoplasmic reticulum and the Golgi complex until it is finally secreted as mature α/β heterodimeric protein toxin [23].

After synthesis the preprotoxin enters the endoplasmic reticulum with the help of a highly hydrophobic signal peptide in the N-terminal region. The signal peptide is cut by a peptidase that perhaps cleaves to produce protoxin. In the endoplasmic reticulum, the γ domain is N-glycosylated and folds into a form suitable for translocation to the Golgi and for further processing. In the Golgi apparatus the products of genes KEX1 and KEX2 are also apparently involved. KEX1 and KEX2 genes were found to encode proteases that are important for the processing of the killer toxin and α factor precursor proteins. The combined action of these two proteases yields mature toxin from protoxin. Kex2p, the gene product of KEX2, is an endopeptidase and cleaves the pro-region, removes the intramolecular γ sequence. Kex1p, the product of KEX1, is a carboxypeptidase which cleaves the C terminal basic dipeptide exposed by Kex2p action. Then mature toxin is transferred to a secretory vesicle and secreted out of the cell. Secretion process is possible with the products of SEC genes [23,28-30]. The secretion pathway of K28 is depicted in figure 1.2.

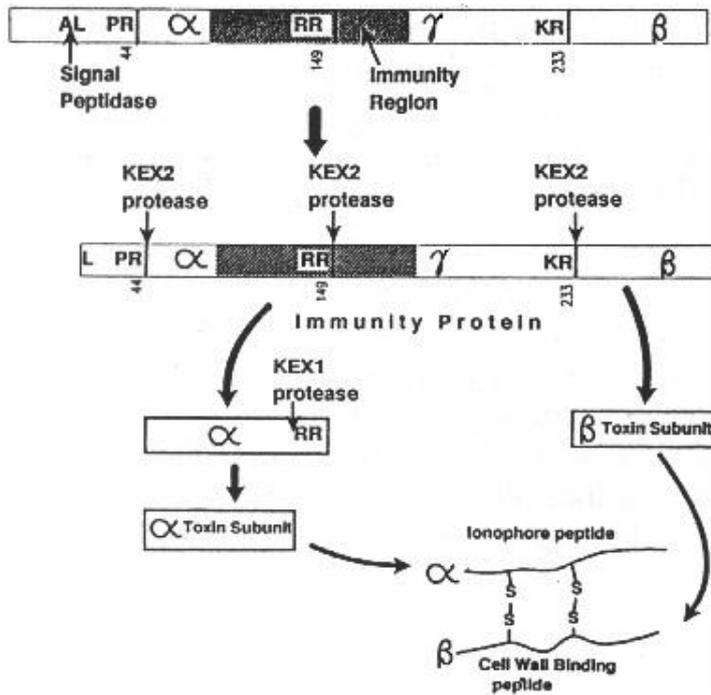


Figure 1.2 Secretory pathway of killer toxin K28 in *S. cerevisiae* [23].

1.4 K5 Type Yeast Killer Protein, Panomycocin

K5 type yeast killer protein is produced by *Pichia anomala* NCYC 434. *P. anomala* toxins have been found to have a broad killing spectrum with relatively high stability. Among *P. anomala*, the NCYC 434 strain has been extensively studied for various applications. The purification and characterization of the killer protein of *P. anomala* NCYC 434 was made and named as Panomycocin [27,31,32].

Panomycocin is a glycosylated monomeric protein. Its molecular mass is 49 kDa and it has a pI value of 3.7. Temperature and pH stability testing of the toxin shows that the optimum pH value is 4.5 and the optimum temperature for the toxin activity is 25 °C [33]. The pH stability is given in figure 1.3.

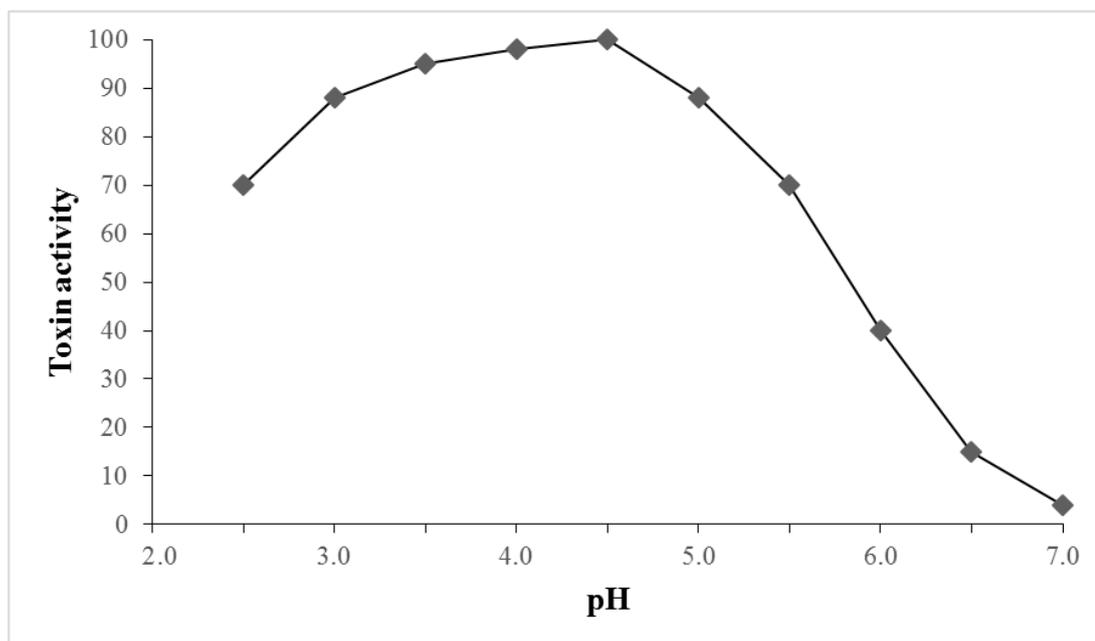


Figure 1.3 pH stability of Panomycocin [33].

The N-terminal of Panomycocin was sequenced and gave GDYWDYQNDKIR sequence. The internal amino acid sequencing also gave LNDFWQQGYHNL, IPIGYWAFQLLDNDPY and YGGSDYGDVVIGIELL sequences. These have 100% similarity with the mature secreted exo- β -1,3-glucanase (with accession number AJ222862) of *P. anomala* strain K which is a glycoprotein with molecular mass of 45.7 kDa and has a pI value of 4.7 [27,33].

Panomycocin exerts its lethal effect by hydrolyzing β -1,3- glucan residues of the cell wall of sensitive fungal cells and causes loss of cell wall rigidity which leads to cell death due to the osmotic pressure. Further studies on its mode of action shows that it exerts hydrolytic activity on the β -1,3- glucans in an exo like manner [32].

Study on the effects of metal ions on the activity of the toxin showed that its activity is inhibited by Hg^{+2} , but increases with other metal ions like Ba^{+2} , Ni^{+2} , Cr^{+2} , Zn^{+2} ; most of all by Pb^{+2} [32].

In vitro activity of Panomycocin was tested against nine dermatophyte strains and all of them were found to be susceptible to Panomycocin [34]. Moreover, Panomycocin was found to have the *in vitro* killing activity against human isolates of pathogenic *Candida spp.* that cause candidiasis and showed *in vitro* and *in vivo* activities against *Penicillium digitatum* and *P. italicum* isolates that cause green mold and blue mold diseases in citrus fruits [35,36].

Panomycocin is a potent antifungal protein that can be used in the therapy of fungal infections due to its novel mechanism of action, high selectivity and relative stability. However, its activity decreases above 37 °C. Thermal stability of Panomycocin should be increased so that formulations that are stable at high temperatures can be prepared.

1.5 Protein Thermostability

Thermostability is the resistance to irreversibility of chemical or physical changes of a substance due to elevation in temperature. Protein thermostability is, therefore, the preservation of the unique structure and chemical properties of polypeptide chains under extreme temperatures [37]. Proteins with high thermostability are needed in various industrial, bioanalytical and pharmaceutical applications [38].

The ability to produce proteins of desired sequence *in vitro* and *in vivo* has led to attempts for production of proteins with increased thermostability [39]. Thermostable proteins may be obtained by protein engineering or by looking for homologs in the thermophiles. When a protein is a eukaryotic one without homologs in the thermophilic organisms, protein engineering would be necessary [38]. The aim of engineering a thermostable protein is to retain its original activity and specificity, but increase its thermostability.

Thermostabilization may keep the protein stable at temperatures that would normally destabilize it and denature it or may keep the protein stable at moderate temperatures over prolonged time (increase in the half life).

Thermodynamic thermostability can be specified by its T_m . T_m is the temperature at which half of the protein molecules are in native state and the remaining half is in the denatured state. Molecules in the denatured state are influenced by proteolysis, adsorption, aggregation and precipitation more easily. The increase in the free energy difference between the native and denatured states may increase the thermodynamic stability [38].

Kinetic thermostability which can be specified by half life or rate constant is generally of greater importance in pharmaceutical applications. The increase in the free energy difference between the native and transition states may increase the kinetic thermostability [38].

Stabilization of the native state may decrease the free energy. This can be achieved by increasing the electrostatic and van der Waals interactions between atoms, by optimization of hydrogen bonding and spatial distribution of polar and hydrophobic residues. The free energy of the denatured state can be increased by decreasing the conformational entropy. This can be achieved by cross linking, post translational modifications, immobilization, adsorption and confinement in nanoparticles [38].

To increase thermostability of proteins various methods can be used. Site directed mutagenesis, segment deletion and ligation, directed evolution, novel thermostable scaffolds, chemical modification of proteins and addition of excipients and confinement in nanoparticles are the principal methods used to improve the thermostability of proteins [38]. Recent applications of protein engineering related with thermostability are given in table 1.1.

Table 1.1 Recent applications of protein engineering or thermostabilization [38].

Protein	Method	Result
Protease from <i>Aspergillus oryzae</i>	Encapsulation within biomimetically generated silicate nanospheres	Enhanced thermostability
Pyranose 2-oxidase from <i>Trametes multicolor</i>	Designed triple mutant	Increased half life at 60°C
Endo β-glucanase EgI499 from <i>Bacillus subtilis</i> JA18	Deletion of C-terminal region	Increases half life from 10 to 29 mins at 65°C
Lipases from <i>Yarrowia lipolytica</i>	Surface display on <i>Saccharomyces cerevisiae</i>	Increased thermal stability and activity
Alkaline protease from <i>Aureobasidium pullulans</i> HN2-3	Surface display on yeast <i>Yarrowia lipolytica</i>	Increased thermal stability and decreased pH stability
Xylanase XT6 from <i>Geobacillus stearothermophilus</i>	Directed evolution and site directed mutagenesis	52 fold increase in thermal stability, catalytic efficiency increase by %90
Tyrosine phenol lyase from <i>Symbiobacterium toebi</i>	Directed evolution	Higher thermostability and activity
Phytase from <i>Penicillium</i>	Random mutation	Increased thermal stability
Arabinose isomerase	Immobilization on amino propyl glass modified with glutaraldehyde	Kinetic thermal stability increases by 138 fold
Papain	Immobilization on cotton fabric with anhydrides	Improved thermal stability and stability to alkali
Cellobiase	Immobilization with chitosan-alginate	Increased thermal stability
Invertase	Inter- and intra-molecular cross linking with diisocyanate reagents	First order thermal denaturation rate constant reduced
Adenylate kinase	Structural entropy optimization	Higher thermostability
L-Asparaginase	<i>In vitro</i> directed evolution	Increase in half life
Human glucocorticoid receptor ligand binding domain (hGR-LBD)	Random mutagenesis and high throughput screening	Thermal stability increases upto 8°C
MDM2	Site directed mutagenesis	Increased thermal stability
HIV 4E10 epitope	Flexible backbone remodeling	Improved thermal stability
Human epidermal growth factor specific antibodies	Graft antigen binding segments from antibody to human EGF (HER-2) onto a designed thermostable three helix bundle	HER-2 binding ability retained after heating to 90°C, three times. Binding constant 76pm
Granulocyte colony stimulating factor	Fusion with gelatin like protein polymer	Increased thermal stability, slower plasma clearance rate
Heme peroxidase from <i>cinereus</i>	<i>In silico</i> Rosetta design and site directed mutation	2/8 of designed enzyme were more stable
Cytochrome P450	Rational mutagenesis	Increased thermal stability
Human erythropoietin	Polyethyleneglycosylation	Increased thermal stability, Prolonged half life

1.6 Excipients

Excipients are substances that are formulated together with the active substance of a pharmaceutical therapy. The quality of drugs depends on the active substance, excipients and the production processes. The classical concept of the excipient as any component other than the active substance has undergone a change from an 'inactive' carrier to an essential component of the formulation. Pharmaceutical regulations and standards require that all excipients in drugs and their decomposition products should be identified and shown to be safe [40].

Pharmaceutical excipients that are used to formulate and stabilize protein therapeutics are classified as amino acids, buffering agents, osmolytes, sugars, polymers and proteins, salts, surfactants, antioxidants and chelators, preservatives and specific ligands (table 1.2).

Buffering agents: Solution pH is important in determining the integrity of amino acid residues and in the maintenance of higher order structure of a protein. Buffering agents are used to maintain solution pH and therefore increased protein stability [41].

Stabilization of proteins with amino acids is by preferential hydration, direct binding, through its buffering capacity and antioxidant properties [42].

Osmolytes: Polyols which are representatives of this group have been found to decrease the surface tension of water and act through the solvophobic effect. And this effects increase protein stability and affects solubility. This also changes their effects on protein folding [43]. In addition to this mechanisms observed in carbohydrates and sugars such as preferential hydration are seen as there are common examples.

Sugars and carbohydrates stabilize proteins probably by preferential hydration at high concentrations, by a combination of specific interactions with proteins and formation of glassy matrices with high viscosity. They can also be used as bulking agents and carrier molecules [41].

Proteins and polymers like rHSA (recombinant human serum albumin) prevent protein aggregation by hindering unwanted adsorption onto vial surfaces, prevent self-association of protein therapeutics through preferential hydration or exclusion effects and to prevent the formation of micro level particles to enhance the solubility of poorly soluble peptide drugs [44].

Surfactants can be used in protein formulations to prevent protein aggregation especially resulting from agitation or shaking. This is primarily due to their ability to outcompete protein molecules for hydrophobic surfaces such as air-water interfaces and therefore prevent proteins from unfolding at these interfaces. Moreover, they can prevent protein molecules from adsorbing to hydrophobic surfaces that exist in the processing [45].

Chelators and antioxidants prevent oxidation by binding with metal ions or other oxidative agents [44].

Preservatives are often needed in protein liquid formulations to ensure sterility during its shelf life. However, preservatives often induce aggregation of protein in aqueous solution [46].

Specific ligands that are inherent ligand binding sites are effective methods in improving the conformational stability of protein drugs [41].

Table 1.2 Pharmaceutical excipients commonly used in protein formulations [41].

Category	Representative Examples	General Comments
Buffering agents	Citrate, phosphate, Tris, sulphate	Control solution pH Buffer ion specific interactions with protein
Amino acids	Lysine, arginine, glycine, proline,	Particular interactions with proteins Free radical scavengers Tonicifying agents Buffering agents
Osmolytes	Trehalose, sorbitol, glycerol, urea, sucrose	Natural compounds that improve protein stability
Sugars and carbohydrates	Mannitol, dextrose, lactose	Protein stabilizing agents Bulking agents Lactose serves as a vehicle for inhalation drugs
Proteins and polymers	HSA, gelatin, PVB, PEG	Protein adsorption inhibitors Bulking agents in lyophilization Serve as carriers in drug delivery vehicles
Salts	Sodium chloride, potassium chloride, sodium sulphate	Have protein stabilizing effects They have also destabilizing effects on proteins Tonicifying agents
Surfactants	Tween 20 and 80	Protein adsorption inhibitors Competitive inhibitors of surface denaturation in proteins Used in liposomes as drug delivery vehicles
Chelators and antioxidants	EDTA, DTPA, ethanol	Bind metal ions Free radical scavengers
Preservatives	Benzyl alcohol, phenol	Prevents microbial growth in multiple doses
Specific ligands	Metals, ligands, polyanions	Increases protein stability by preventing stress induced unfolding and by enhancing conformational stability

1.7 Homology Modeling

Homology modeling is aimed at predicting a structure of a protein from its sequence with a high accuracy which is comparable to the results obtained experimentally. 3D structure prediction of a protein can be performed by using experimental methods such as NMR spectroscopy and X-ray diffraction. It can also be done by using homology modeling. For NMR analysis protein molecules should be small and for X-ray diffraction these molecules should be crystallized. Homology modeling is the only alternative to predict the structure of a protein if experimental methods fail. Moreover, homology modeling is a time saving and easy method [47].

The basis for homology modeling are two major findings:

1. The structure of a protein is particularly determined by its amino acid sequence [48].
2. The structure of proteins is more conserved and changes much slower than the related sequence during evolution. As a result similar sequences fold into identical structures and even sequences with low relation take similar structures [49,50].

1.7.1 Steps in Homology Modeling

Homology modeling is a process that consists of multiple steps (figure 1.4) that can be summarized in the following steps (1) identification of templates and initial alignment; (2) sequence alignments and alignment correction; (3) building the model; (4) loop modeling; (5) side chain modeling; (6) optimization of the model, and (7) validation of the model [47,51].

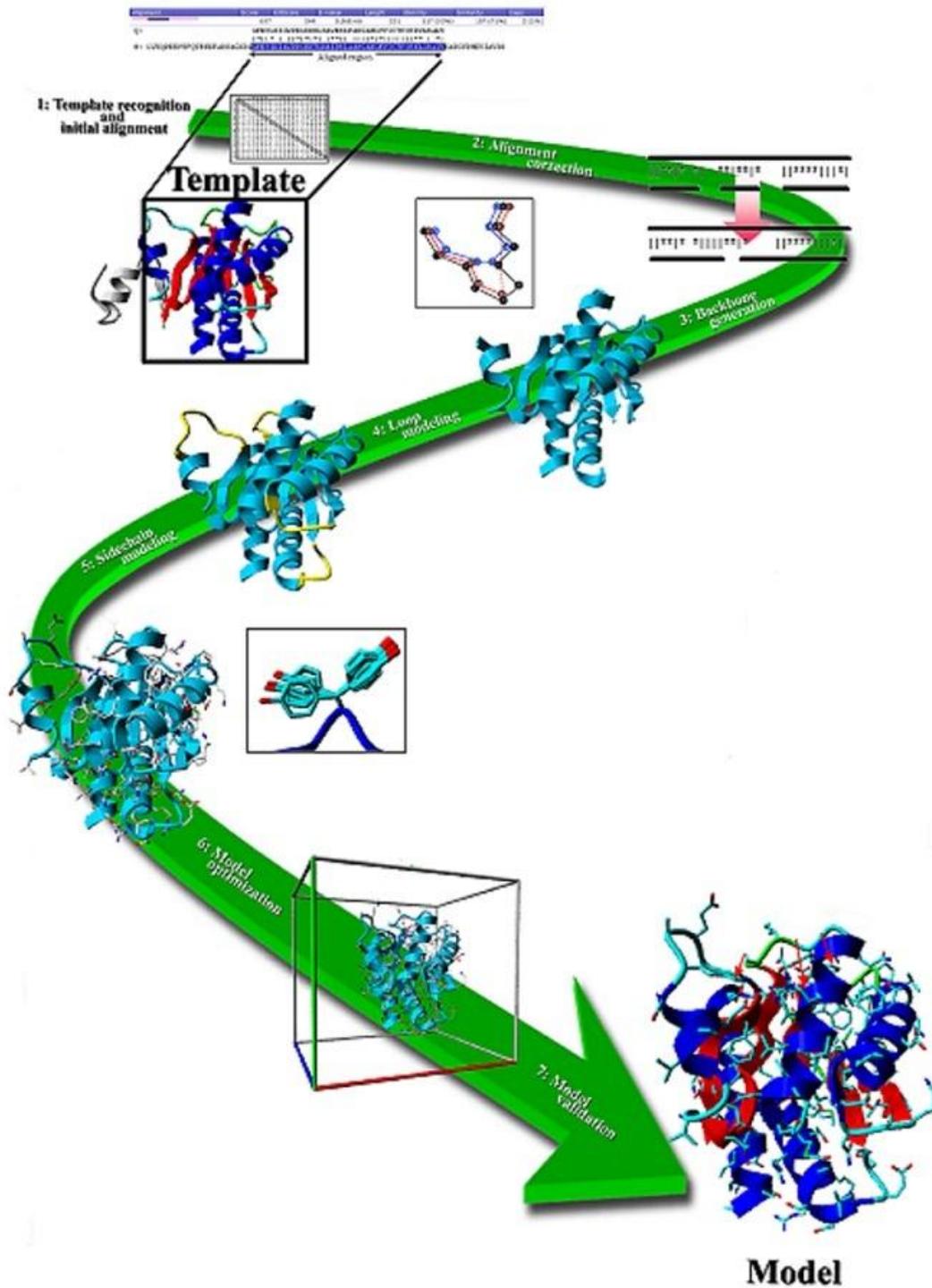


Figure 1.4 Steps of homology modeling [47].

Step 1: Identification of Templates and Initial Alignment

This is the first step in which the query sequence is compared with known structures stored in PDB (Protein Data Bank) by using programs or servers. The most popular server used is BLAST (Basic Local Alignment Search Tool). A search with BLAST against the database for suitable local alignments with the query sequence give a number of proteins that matches the sequence. From the list of the results only those which structures are known and expected to fold into the same structure should be selected as templates (figure 1.5). More sensitive alignment methods based on iteration such as PSI-BLAST, Hidden Markov Models (HMMs) or profile profile alignments have been developed. And these methods are more preferable for homology searches these days.

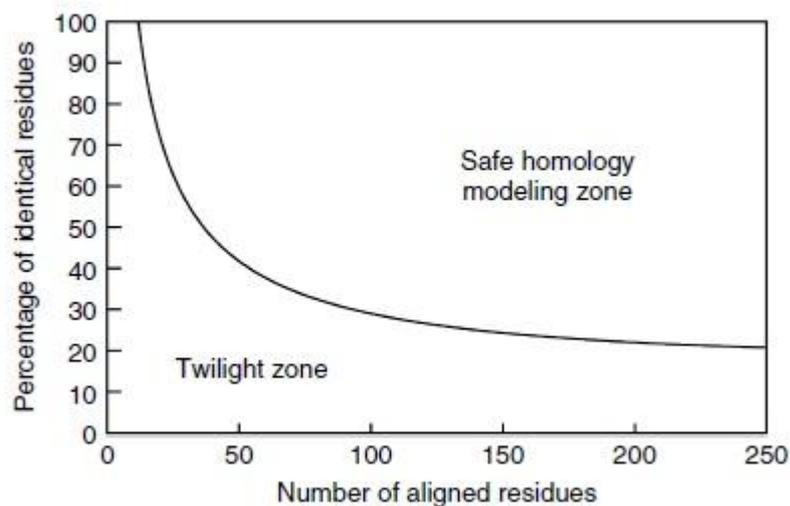


Figure 1.5 The two zones of sequence alignments. Two proteins are expected to fold in to the same structure if the sequence length and percentage identity fall into the region which is marked as safe [52].

Step 2: Sequence alignments and alignment correction

After the templates are found, there will be alignments and correction of them if necessary. When the percentage sequence identity is very low, alignment of the two sequences in a region may be difficult. In such conditions sequences from homologous proteins can be used to find a solution. For example suppose we want to align the sequence GAGAGAGA with RKRKRKRK. These are two completely different sequences that may cause the formation of a gap in this region. However, a sequence like ARARARAR that aligns to both of them can solve the problem. This is the basis for the concept of multiple sequence alignment [47].

Multiple alignments are heuristic that are known as progressive alignment [53]. Progressive alignments are easy to conduct and allow alignments of large sequences with low relation to be built. ClustalW program allows the alignment of distantly related sequences with high accuracy [54]. HMMs are a class of probabilistic models which have application mostly to linear sequence [55]. Profile HMM are highly efficient in detecting conserved regions in multiple sequences [56].

Step 3: Model building

Experimentally determined protein structures may have errors. To develop software that detects these errors, lots of studies have been performed. PDBREPORT database at www.cmbi.nl/gv/pdbreport is an example [57]. So, templates with the fewest errors should be chosen to build a good model.

Various methods are used to build a model for the query sequence. Rigid body assembly [58–60], segment matching [61], spatial restraint [52] and artificial evolution [62] methods are used to build the model. Rigid body assembly model building is based on the dissection of the protein structure into conserved core regions, loops and side chains. A subset of atomic positions obtained from template structures are used as leading positions in segment matching. Modeling by satisfaction of spatial restraints depends on the generation of many restraints on the structure of query sequence by using its alignment to similar protein structures as a guide.

Step 4: Loop modeling

Gaps or insertions called loops are present in sequences of homologous proteins. The structures of loops are not conserved during evolution. Even without deletions or insertions different loop conformations in query and template are often found. This is the consequence of surface loops that have the potential to participate in crystal contacts, exchange of side chains below the loop that pushes it aside or mutation of a residue to proline or from glycine to any other residue in a loop [47].

The functional specificity of a protein structure is often determined by the loops. The accuracy of loop modeling is an important factor which determines the value of homology models for further applications [63]. Since loops show higher structural variability than strands and helices, the prediction of their structure is more difficult than strands and helices.

Knowledge based and energy based approaches can be used for loop modeling. In the knowledge based approaches prediction of the loop structure measures the orientation and separation of the backbone segments, flanking the region to be modelled and eventually look for segments with the same length that span a region of similar structure in the PDB that will be copied.

Loop construction by random search mechanisms is the main alternative to this [64]. In the energy based approaches an energy function is used to measure the quality of a loop. Then this function is minimized using Monte Carlo [65] or molecular dynamics [66] techniques to arrive at the best loop conformation.

Step 5: Side chain modeling

Side chain modeling is usually done by putting side chains onto the backbone coordinates that are obtained from a parent structure and/or derived from *ab initio* modeling simulations. In practice side chain prediction works at high levels of sequence identity. Protein side chains are present in a limited number of structures

with low energy known as rotamers. Depending on defined energy functions and search strategies, rotamers are selected in accordance with the preferred protein sequence and the given backbone coordinates. The accuracy of prediction is usually high for residues in the hydrophobic core but low for water exposed residues on the surface [47,51].

Step 6: Model optimization

Optimization of the model usually begin with an energy minimization utilizing molecular mechanics force fields [67,68]. At each energy minimization a few big errors are removed but many small errors are introduced at the same time and start accumulating. Therefore, restraining the atom positions, applying only a few hundred steps of energy minimization and using more precise force fields like quantum force fields [69] and self parameterizing force fields [70] can be used to decrease the errors in model optimization. For further model optimization methods such as molecular dynamics and Monte Carlo can be used [71,72].

Step 7: Model validation

Depending on the percentage identity between the sequences of the query and the templates and the quality of the templates, the generated models have errors. Thus verification and validation of models is necessary. Errors are estimated by calculating the energy of the model based on force fields and by using normality indices [73]. To determine the normality indices the normality of bond lengths, bond and torsion angles can be checked [74,75]; core/surface distributions of polar and nonpolar residues can be investigated [76]; the radial distribution and direction of atomic contacts can be calculated [77,78].

Many programs that are available freely can be used for verification and validation of homology models. To mention WHAT_CHECK solves crystallographic problems [79]. Programs such as PROCHECK and WHATIF check for appropriate higher order structure and structural packing quality; and programs such as Verify 3D and PROSAIL check the compatibility of sequence to structure by assigning a score for each residue

compatible in its current environment. The Ramachandran plot is also powerful determinant of the quality of protein structure [80].

1.7.2 Softwares for Homology Modeling

There are many programs and servers that are used to build models from query sequences. MODELLER, I-TASSER, SWISS MODEL, PrISM, ORCHESTRAR, MOE, COMPOSER and ROSETTA are some examples.

1.7.2.a MODELLER

MODELLER is a popular and widely used homology modeling tool which is available freely, has powerful features and gives reliable results. MODELLER utilizes the query structures to build constraints on atomic distances and dihedral angles. Then these are combined with statistical distributions obtained from many homologous structure pairs in the PDB. But for most users it is some what difficult to begin with MODELLER as it is based on command line and requires basic knowledge about Python scripting to use it effectively. Hence a graphical user interface to MODELLER called EasyModeller has been introduced. EasyModeller provides modeling, assessment, visualization, and optimization of protein models in a simple and straightforward way for users. EasyModeller has the following features: 1) tab based logical modeling steps; 2) allows to load unlimited number of templates; 3) colorful alignment viewer with alignment editor; 4) MODELLER code editing; 5) incorporated DOPE profile viewer, Ramachandran plot viewer, loop modeling, model optimization and dynamics for a selected model [81,82,83].

1.7.2.b I-TASSER

I-TASSER (Iterative Threading ASSEmbly Refinement) is a server that provides an internet based service for protein structure predictions that allows academic users to make high quality structure predictions. Critical Assessment of Structure Prediction (CASP) experiments have been developed to obtain an objective assessment of the

performance of modeling tools. I-TASSER was found to be one of the best methods in the servers section of the CASP experiments. In the default mode the query sequence is submitted and the program obtains templates and does the alignments. It is also possible to specify the templates and alignments by selecting the option ‘specify template with alignment’ in the I-TASSER server. Depending on the sequence length, the I-TASSER modeling procedure takes a maximum of 48 hours. The output of the I-TASSER server includes up to five models, estimated accuracy of the predicted models including TM score and RMSD value for the first model, GIF images of the predicted models and top 10 proteins in PDB which have structural similarity with the predicted models [84,85,86].

1.7.2.c SWISS MODEL

SWISS MODEL is a protein structure prediction method that gives the model of a protein from its amino acid sequences. It is a web based service that is used widely. It provides users a friendly web interface that allows users to generate 3D models for their proteins of interest without installing or downloading large databases. A recently developed interactive web interface allows users to easily look for appropriate templates using sensitive Hidden Markov Models (HMM) searches against SMTL (SWISS MODEL Template Library), analyse the alternative templates and alignments and compare the resulting models using QMEAN potential. The out put model is colored by model quality estimates calculated by QMEAN to highlight regions of the model according to the quality of the modeling. If many models have been generated for a query sequence, these models can be interactively superposed and visualized [87,88].

1.7.2.d PrISM

PrISM (Protein Informatics System for Modeling) is a homology modeling method by using alignment to build a composite template by selecting each secondary structure from the most suitable template. Loop modeling is done by using *ab initio* methods and side chain dihedrals are obtained from the template or predicted structure based

on main chain torsion angles and a neural network algorithm. Advanced sequence template alignment techniques in PrISM are important in conditions where standard pairwise dynamic programming algorithms fail to make any reasonable global alignment [89].

1.7.2.e ORCHESTRAR

ORCHESTRAR consists of programs that are used to align homologs, generate conserved regions, find structurally variable regions and add side chains. In these programs the concept of using existing knowledge from multiple templates and the underlying basic knowledge about protein environment is utilized [90].

1.7.2.f MOE

MOE (Molecular Operating Environment) is the combination of segment matching methods [61] and the modeling of insertion/deletion regions approach [91]. The applications in homology modeling are powerful, intuitive and easy to use. MOE has these features: 1) 3D structure prediction from sequence; 2) fold detection / template selection with searchable structural databases; 3) advanced loop modeling; 4) advanced alignment methods; 5) powerful alignment visualizer and editor [92].

1.7.2.g COMPOSER

Composer is based on structural alignments information obtained from multiple templates to detect structurally conserved regions (SCRs) across all homologs. From this information it builds a partial model. The remaining gaps are built using a loop modeling algorithm. Its features include: 1) detection of SCRs; 2) loop modeling; 3) customizable database of high quality 3D protein structures extracted from the PDB; 4) similarity matrices for scoring sequence homology; 5) support for various file formats; 6) graphical editing of selection of homologs and sequence alignment; 7) displays multiple aligned structures along with sequence percent identity and statistical significance of the alignment simultaneously [93,94].

1.7.2.h ROSETTA

Rosetta a comprehensive freely available software suite for modeling molecular structures. It has been a strong performer in the CASP prediction exercises. Rosetta is a structure prediction tool that offers various efficient sampling algorithms to search for the backbone, side chain and gaps. It also includes tested scoring functions. Rosetta uses information from known structures to build a model from a query sequence [95].

1.7.3 Applications of Homology Modeling

Homology modeling has a wide range of applications and its importance is increasing as the number of structures determined increase. It has applications in: (1) structure based drug design process; (2) investigation of the effects of mutations; (3) identification of binding sites; (4) looking for ligands and designing of novel ligands; (5) modeling of substrate specificity; (6) protein–protein docking simulations; (7) molecular replacement in experimental structural refinements; (8) rationalizing of known experimental results and (9) planning of future computational experiments by using the generated models [48].

Homology modeling has many applications in drug discovery process. This makes the discovery process faster, easier, cheaper and more practical (figure 1.6 and table 1.3).

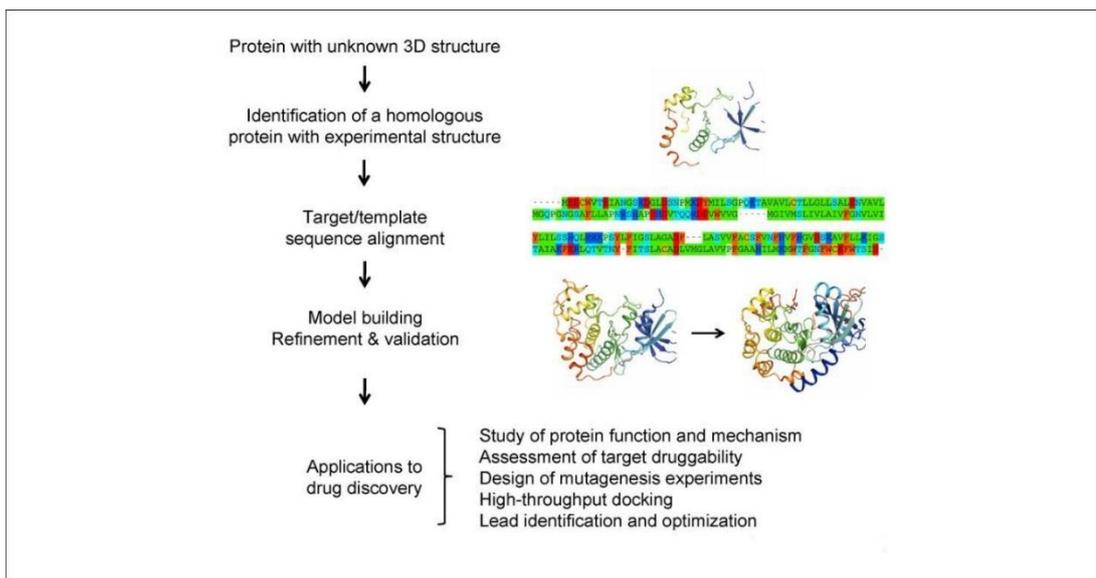


Figure 1.6 Applications of homology modeling in drug discovery and its outline [96].

Table 1.3 Recent applications of homology modeling.

Protein	Application	Program/Server
M Antigen [97]	Study of protein function	BLAST, SWISS MODEL, CLUSTALW, Swiss PDBViewer, PROCHECK
Nod like Receptors [98]	Understanding of the protein mechanism in the immune response	FFAS, Predictprotein, m-coffee, MUSCLE, SCWRL, ProSa, Pymol, ConSurf
ECE-2 [99]	Understanding of catalytic activity loss of ECE-2	MODELLER, PROCHECK, ProSa, Pymol, AutoDock
MCH-R1 [100]	Structure-based discovery of antiobesity drugs	ICM, MODELLER, AutoDock
1,2,4-Triazole Analogs [101]	Study of Structure-function relationships	BLAST, SSpro, PRIME, SITEMAP, PROCHECK, WHATIF, MAESTRO
Human Dopamine (D_{2L} and D3) receptors [102]	Exploration of structure and ligand based interactions	MODELLER, SYBYL, PROCHECK
Dopamine D2 receptor [103]	Structure identification for the design of new protease antagonists	CLUSTALX, BLAST, MODELLER, SYBYL
Leishmanial Farnesyl Pyrophosphate Synthases[104]	Useful in the identification of hits with novel scaffolds as antileishmanial agents	BLAST, MODELLER, SWISS-PROT, SYBYL, InsightII
hASIC1a ion channel [105]	Dsiplays the route to <i>in silico</i> search for developed ASIC1a channel blockers	MODELLER, DOT, SHAKE
Cytochrome P450 sterol 14α-demethylase [106]	Identification of new hit compounds with similar inhibitory activities	Amber, CPHmodels, FlexX, PROCHECK, SYBYL, PMF, DOCK
Lysophosphatidic acid LPA4 receptor [107]	Contributes for the recognition of LPA within the LPA4 receptor	SYBYL, GOLD, SCWRL
Cysteine Protease YopT from <i>Yersinia pestis</i> [108]	Important for plague regulation and further drug designing	Phyre2, ModRefiner, SOPMA, CASTP, PROCHECK, ERRAT, Verify 3D
Human angiotensin II type I (Ang II-AT1) receptor [109]	Guide for the design of novel compounds as Ang II AT1 receptor antagonists	BLAST, CLUSTALW, SYBYL, MODELLER, I-TASSER, PROCHECK, SurfexDock
Chitinase II from <i>Thermomyces lanuginosus</i> [110]	Offers insights into the structure and stability of the protein	BLAST, SYSTEMS, I-TASSER, MODELLER, PsiPred, COFACTOR, MetaPocket, COACH, SAVES, CDOCKER
Transketolase [111]	Guide for the design of the enzyme inhibitors with potential antimalarial activity	SOPMA, MODELLER, QMEAN, PROCHECK, CASTP, GlobPlot, AutoDock

1.8 Comparison of Models

During modeling process many models of a query sequence may be generated. At this time to identify the best model, the generated models are compared using various parameters. Moreover, the best models are compared with the templates to measure the quality of the modeling process. In addition to this such parameters can also be used to compare two different protein structures that are generated through modeling or practical methods. DOPE (Discrete Optimized Protein Energy) score, TM (Template Modeling) score and RMSD (Root Mean Square Deviation) value are example parameters that are used for comparison.

1.8.1 DOPE Score

DOPE is a statistical potential based on atomic distance calculated from a group of native protein structures that is related with the probability theory. So, it is dependent on the spherical shape of the structures. It is implemented in the MODELLER and used to measure the energy of the model which is generated through many iterations by MODELLER. DOPE is used to measure the quality of the structure as a whole. It can also generate a residue-by-residue DOPE profile for a model so that the region that is problematic in the model can be detected. DOPE is the best function that is used in the detection of the native state, comparison of models and determination of the best model [112].

1.8.2 TM Score

TM score is used to measure the similarity of topologies of protein structures. TM score has a value in the interval (0,1]. Here 1 is an indication of a perfect match between two structures. TM score below 0.17 shows two randomly selected proteins which are unrelated whereas a score above 0.5 shows two structures take the same fold.

$$\text{TM-score} = \max \left[\frac{1}{L_{\text{target}}} \sum_i^{L_{\text{aligned}}} \frac{1}{1 + \left(\frac{d_i}{d_0(L_{\text{target}})} \right)^2} \right]$$

Where, L_{target} : length of the query protein,
 L_{aligned} : length of the aligned region,
 d_i : distance between the i^{th} pair of residues and
 d_0 : distance scale that normalizes distances.

$$d_0(L_{\text{target}}) = 1.24 \sqrt[3]{L_{\text{target}} - 15} - 1.8$$

TM score gives information about global fold similarity and it is less sensitive to structural differences at local level. TM score does not depend on the length of the protein as a result of normalization that depends on the size. So, it solves the problem in classical measurements such as RMSD [113].

1.8.3 RMSD

RMSD is used to measure the differences between values of a model and the values of templates. RMSD represents the standard deviation of the differences between predicted values and actual values.

$$\text{RMSD} = \sqrt{\frac{\sum_{i=1}^{N \text{ atoms}} (r_i(t1) - r_i(t2))^2}{N \text{ atoms}}}$$

Where N_{atoms} : number of atoms being compared and
 $r_i(t)$: position of atom i at time t .

1.9 Molecular Dynamics and Energy Calculations

There are various methods and programs to calculate the energy and molecular dynamics of proteins. Molecular dynamics is used to assess the state of a protein in various environmental conditions. Energy calculations are used to compare models and to minimize the energy to make the model more stable. The principle of working and the force fields used are different for each one of them. Among them GROMOS (GRONingen MOlecular Simulations), GROMACS (GRONingen MACHine for Chemical Simulations), CHARMM (Chemistry at HARvard Macromolecular Mechanics), AMBER (Assisted Model Building with Energy Refinement) and OPLS (Optimized Potentials for Liquid Simulations) are commonly used in various programs.

1.9.1 GROMOS

GROMOS has the capability of simulation of biomolecules using the molecular dynamics and analysis of molecular configurations and energies obtained by computer simulation or model building based on experimental data. Since it has been developed, various versions of it were introduced. Among them GROMOS96 was comprehensive and was released with manual. In GROMOS96 the force field, the potential energy function that describes the interaction between the particles or atoms. Here aliphatic CH_n groups are represented as combined atoms with van der Waals interactions reparametrized based on a series of molecular dynamics simulations of model liquid alkanes. It also includes energy minimization and stochastic dynamics [114,115].

1.9.2 GROMACS

GROMACS provides molecular dynamics, energy minimization and stochastic dynamics. Free energy determinations and nonequilibrium dynamics are also incorporated in it. Even if it does not have its own force fields, it is compatible with the force fields of GROMOS96, AMBER and OPLS. It can treat polarizable shell models and flexible constraints. Force fields can be added by the user so that functions can be specified. Since the program is versatile, analyses can be customized easily. The program is also fast and free [116].

1.9.3 CHARMM

CHARMM provides molecular dynamics simulations, free energy estimates, molecular energy minimization and analysis techniques. It has its own force fields with various versions. It is a flexible and widely used program. It can be utilized with various energy functions and models. It has been ported to many platforms. It has many applications with special emphasis on the study of biological molecules [117].

1.9.4 AMBER

AMBER consists of a set of molecular mechanical force fields and a suite of molecular simulation programs. It is possible to implement AMBER force fields with AMBER programs. Even if the codes are distributed under license agreement, the force fields are free. The force field function includes the bond length due to covalent bonds, the angle or geometry of the covalent bonds, the torsion of the bonds and the nonbonded energy between all atom pairs. It has been implemented in several other computer packages. It has applications in proteins, nucleic acids, drugs, ligands and carbohydrates [118,119].

1.9.5 OPLS

OPLS consists of a set of force fields that can be used for computer simulations of proteins in their native state. These force fields are adopted from AMBER. It has versions of parameters like OPLS-UA and OPLS-AA. These parameters were obtained from Monte Carlo simulations and they were tested primarily in combination with them [120].

1.10 Protein Stability Predictors

Stability predictors are important to determine the stability of a protein in a given environmental conditions. Furthermore when a mutation is undertaken over a protein, the change in its stability can be estimated with these predictors. They predict stability changes due to single site mutations by using physical, statistical, empirical or machine learning approaches. Most of the methods require 3D structure of the protein. However, some can do the prediction only from the sequence of the protein. Most of them give the general trend of stability. Some others can give the exact value of the increase or decrease in stability.

There are various protein stability predictors. To mention CNA (Constraint Network Analysis), CUPSAT (Cologne University Stability Analysis Tool), I-MUTANT, FoldX, AUTO-MUTE and MUpro.

1.10.1 CNA

CNA enables to identify the rigid and flexible regions in a biomolecule, provides a refined modeling of thermal unfolding simulations, allows rigidity analyses and calculates global and local indices for quantifying biomacromolecular stability. CNA has applications in the investigation of protein thermostability, identification of weak spots in a protein and linking the protein flexibility with function [121,122,123].

CNA web server has been developed. The CNA web server has a user friendly interface. It accepts PDB IDs or PDB format of a protein as an input. If there is a deficiency of hydrogens and/or residues in the input protein, it should be added before submitting the input to the server. Analysis type and simulation parameters are chosen by the user. Results are available in the web server or can be fetched from the mail given to the server. The result includes global and local indices with plots along with mappings onto the 3D structure. Flexible regions predicted for a protein structure are also depicted on the structure. It can calculate the exact temperature changes due to a point mutation. In short CNA web server provides detail analysis results [123].

1.10.2 CUPSAT

CUPSAT is used to predict stability changes in a protein due to single site mutations. In CUPSAT mean force potentials are used to predict protein stability; amino acid-atom potentials are used; torsion angle potentials are obtained from the distribution of main torsion angles ϕ and ψ ; gaussian apodisation function has been used to accommodate torsion angle perturbation in protein mutants [124,125].

CUPSAT accepts PDB format of the protein as an input. The input PDB structure should be in the quality of proteins that are put into PDB. The output includes information about mutation site and detail information about changes in protein stability for the 19 substitutions of a specific amino acid residue. It also analyses the ability of the mutated amino acids to adapt the observed torsion angles [126].

1.10.3 I-MUTANT

I-Mutant2.0 is a support vector machine (SVM)-based tool that is used to predict stability changes due to single site mutations. It accepts PDB format or sequence of a protein as an input. Giving protein stability change predictions based only on the sequence makes it special. I-MUTANT2.0 predicts the sign of protein stability change upon mutation and estimates the related $\Delta\Delta G$ values. The output consists of the position under consideration and the changes undertaken, the predicted value or the sign of free energy change, the temperature and the pH at which the prediction has been carried out [127]. I-MUTANT3.0 is another version of I-MUTANT. I-Mutant3.0 has similar working principles with the other version. The prediction result of I-MUTANT can be classified as: neutral mutation ($-0.5 \leq \Delta\Delta G \leq 0.5$), large decrease (< -0.5) and large increase (> 0.5) [128].

1.10.4 FoldX

FoldX is a force field that is used in the investigation of the effects of mutations on the stability, folding and dynamics of proteins. It allows the computation of the stability of a protein, calculation of the positions of the protons, prediction of binding sites and analysis of the free energy of complex formation. FoldX gives good results whenever the two structures being compared are known. Otherwise the results may have a large error [129].

The FoldX web server accepts PDB structures as an input after the users join the web as guests or registered users. It is also possible to download the executive software. After the calculations are done according to the specified options, the results are sent to the specified mail in zip format [129].

1.10.5 AUTO-MUTE

AUTO-MUTE provides the combined power of computational mutagenesis using a four body and cutting-edge machine learning methodologies and tools. As a result it provides more accurate prediction for mutant using various protein functions. For each type of function prediction, a variety of classification and regression models have been developed and are available. It accepts PDB ID or PDB structure of a protein as an input. The parameters to be calculated and the conditions in which this calculations are undertaken are specified by the user. The output is in the form of increase or decrease stability [130].

1.10.6 MUpro

MUpro predicts thermostability changes due to single site mutations by using SVM and neural networks. Inputs can be in 3D format or just the sequence of the protein. Furthermore experimental results showed that the accuracy of predictions using sequence alone is similar with the accuracy of predictions using tertiary structure information. This is important when the tertiary structure of a protein is not determined. The output shows the general trend of stability. The result is in the form of +/- free energy value changes [131].

1.10.7 Eris

Eris is a server used to calculate protein stability changes resulting from mutations utilizing the Medusa modeling suite. Inputs can be submitted in PDB format. After calculations are done in accordance with the conditions specified by the user, the results are sent to the e-mail of registration of the user. Studies that compared results of Eris with the experimental data showed that there are significant correlations among of them. Eris also allows refinement of the protein structure when there is a problem with the quality of the 3D structure of the protein [132].

1.11 Aim of the Study

The aim of this study is to thermostabilize Panomycocin so that it can be formulated as a drug which will be stable at high temperatures (above 37 °C). Various excipients will be tested to thermostabilize the protein. For this purpose Panomycocin will be produced and isolated. The excipients will be mixed with the isolated protein and the mixture will be tested whether it is stable at high temperatures. Since excipients can increase thermostability only to some extent, computational methods are also going to be used to design a thermostable protein. Its 3D structure will be built by using homology modeling. Computational site directed mutations will be undertaken and the effect of the mutations will be assessed by using various programs and servers. By doing this the positions that may increase the thermostability of the protein will be detected.

CHAPTER 2

MATERIALS AND METHODS

2.1 Materials

2.1.1 Strains

Pichia anomala (NCYC 434) which is the source of Panomycocin and *Saccharomyces cerevisiae* (NCYC 1006) which is the sensitive strain were purchased from the National Collection of Yeast Cultures, Norwich, U.K.

2.1.2 Culture Media

P. anomala and *S. cerevisiae* were grown and maintained in YEPD medium consisting of 1 % Bacto-yeast extract, 1 % Bacto-peptone, 2 % dextrose and 2 % Bacto-agar.

During production of Panomycocin, *P. anomala* cells were grown in YEPD medium buffered to pH 4.5 with phosphate citrate buffer in the presence of 5 % glycerol as a protein stabilizer. YEPD medium together with 2 % Bacto-agar buffered to pH 4.5 was used to determine the killer activity.

2.1.3 Chemicals

The chemicals used and their suppliers are listed in Appendix A.

2.1.4 Buffers and Solutions

The buffers and solutions used are listed in Appendix B.

2.1.5 Programs and Servers

The programs and servers used and their websites are listed in Appendix C.

2.2 Laboratory Methods

2.2.1 Sterilizations

Liquid cycle sterilization at 121 °C for 15 minutes was used to sterilize the glasswares, the media, buffers and distilled water used in all steps. Buffers used for the chromatographic steps were filtered using 0.45µm cellulose acetate filters (Sartorius, AG, Germany) prior to sterilization.

2.2.2 Maintenance of the Yeast Cultures

Stock cultures of *P. anomala* (NCYC 434) and *S. cerevisiae* (NCYC 1006) were maintained in YEPD agar plates. These plates were stored at 4 °C. New stocks were propagated for the replacement of the stock in every 2 months [33].

2.2.3 Production of Panomycocin

Production, concentration and isolation of Panomycocin was done by the method described by İzgü and Altınbay [33]. *P. anomala* NCYC 434 cells were cultivated into 10 mL of YEPD medium. After the yeast cells were incubated at 25 °C for approximately 7 hours on a gyratory shaker (Innova 4330, New Brunswick Scientific, USA) at 120 rpm, one mL cell suspension was obtained from the above inoculation and this was inoculated into 100 mL of the same medium. This was incubated at 25 °C overnight at 120 rpm on a gyratory shaker (Innova 4330, New Brunswick Scientific, USA). Then 10 mL cell suspension was transferred to 1L of YEPD medium with 5 % glycerol whose pH was made 4.5 with acetic acid. This was further incubated at 18 °C for 48 hours at 120 rpm on a gyratory shaker (Innova 4330, New Brunswick, USA). After 24 hours of incubation, %20 glucose was added to 1L YEPD medium with glycerol. To obtain the cell free medium, centrifugation (KR 22i, Jouan, France) was done at 9500 rpm for nearly 10 minutes at 4 °C. The supernatant obtained was filtered using 0.45µm first and then 0.2µm cellulose membrane filters (Sartorius, AG, Germany) to sterilize the medium.

2.2.4 Preparation of Crude Panomycocin

Crude protein was prepared by ultrafiltration systems. The cell free culture medium consisting of Panomycocin was concentrated by using first 30 kDa and then 5 kDa molecular weight cut-off ultrafiltration systems (Vivaflow 200, Sartorius AG, Goettingen, Germany) having polyethersulfone membranes operating with peristaltic pump at nearly 2.5 bar pressure. The total 4 L medium was concentrated to approximately 50 mL.



Figure 2.1 Ultrafiltration system for Panomycocin a) feed, b) return, c) filtrate.

2.2.5 Determination of Killer Activity of Panomycocin

To determine the killer activity of Panomycocin YEPD agar plates with pH 4.5 were prepared. *S. cerevisiae* (NCYC 1006) cells in sterile water at 0.5 McFarland standard cell density (1 to 5 million cells/mL) were prepared. These cells were spread and cultivated into the plates with cotton buds. Fifty μL Panomycocin samples were spotted onto the YEPD plates and incubated at 25°C. These were checked for clear zone of growth inhibition after 24 hours of incubation to determine the killer activity.

2.2.6 Panomycocin Purification with Gel Filtration Chromatography

Concentrated crude Panomycocin was subjected to gel filtration chromatographies by using a fully automated FPLC system (Biocad 700E Perseptive Biosystems, USA) with a fraction collector (SF-2120 Super Fraction Collector, Advantec MFS, Japan). The flow was detected with UV absorbance at 280 nm at 20 °C.

The concentrated Panomycocin sample was subjected to gel filtration chromatography using a TSK G2000 SW (7.5 mmD/30 cmL TosoHaas, Japan) column. Prior to the injection of the sample, column was equilibrated with 0.1 M Na₂HPO₄-citric acid buffer at pH 4.5 containing 0.1 M Na₂SO₄ at a flow rate of 1 mL/min. 80 µL of the sample was injected into the column and the same buffer at a flow rate of 1 mL/min was used for elution. After the fractions that consist of Panomycocin was determined, these fractions were collected. The collected fractions were concentrated and buffer exchanged with pH 4.5 acetate buffer. Fifty µL of the purified protein was spotted on to YEPD (pH 4.5) agar plates seeded with 0.5 standard McFarland *S. cerevisiae* NCYC 1006 cells to determine the killer activity of the protein.

2.2.7 Determination of Protein Concentration

UV-Vis spectrophotometer (Nanodrop 2000, Thermo Scientific, USA) at 280 nm was used to determine the concentration of Panomycocin.

2.2.8 SDS Polyacrylamide Gel Electrophoresis

Gel electrophoresis was done according to Laemmli [134] protocol by using SE250/SE260 Mighty Small slab gel unit (Hoefer, USA). The purity of the protein was measured by using electrophoresis on a 12.5 % linear, 0.75 mm thick polyacrylamide gel in a discontinuous buffer system using a vertical slab gel electrophoresis unit SE 250 (Hoefer, USA).

12.5 % separating gel was prepared and poured into the electrophoresis unit. This was covered with n-butanol to avoid contact of the gel with air and then left for polymerization for overnight. When the polymerization of the separating gel was over, n-butanol was removed by washing first with water and then stacking buffer. Then the stacking gel was poured and left for polymerization. Separating and stacking gel components and their amounts are given in table 2.1 and 2.2.

Protein samples were heated for 5 minutes with boiling water in equal volume of sample buffer (0.125 M Tris-Cl, 20 % glycerol, 4 % SDS, 0.02 % bromophenol blue at pH 6.8) and 10% 2- β - mercaptoethanol was added to the sample buffer for SDS-PAGE. Finally the samples were loaded onto the gel after the polymerization of the stacking gel was over.

Electrophoresis was done at 15 mA/0.75 mm thick gel (Power supply PP4000, Biometra, Germany) at constant current for an hour.

Table 2.1 Separating gel mixture (12.5%).

Components	Amount
Acrylamide-bisacrylamide (30:0.8)	8.3 mL
4X Separating Gel Buffer (1.5M Tris-Cl, pH:8.8)	5.0 mL
10% SDS	0.2 mL
ddH ₂ O	6.4 mL
10% Amonium Persulfate*	100 µL
TEMED*	6.7 µL

*APS and TEMED were added after deaeration.

Table 2.2 Stacking gel mixture.

Components	Amount
Acrylamide-bisacrylamide (30.8% T, 2.7% C _{bis})	0.44 mL
4X Separating Gel Buffer (0.5M Tris-Cl, pH:6.8)	0.83 mL
10% SDS	33.0 µL
ddH ₂ O	2.03 mL
10% Amonium Persulfate*	16.7 µL
TEMED*	1.7 µL

*APS and TEMED were added after deaeration.

2.2.9 Protein Detection in Gels by Coomassie Brilliant Blue Staining

Protein bands on the gel were visualized by using coomassie brilliant blue standard staining method described by Wilson [135]. After the gel was removed, it was put into staining solution over night with a gentle shake on a rotary shaker. After staining was made, the gel was put into destaining solution I to remove the bulky stain over it. Then the gel was put into destaining solution II. Destaining solution II was refreshed until the gel back ground was clear and the blue protein bands were detected.

2.2.10 Thermostability Tests by Using Excipients

Fresh stock solutions of the excipients were prepared. From these stock solutions 50 μL sample was prepared in which 25 μL was protein and 25 μL was stock solution diluted with sterile buffer for the required concentration of the excipient. Gradient concentrations of the samples were put into 4 °C for 24 hours. YEPD medium along agar at pH 4.5 were poured into petri dishes. The YEPD petri dishes were waited until the media solidifies and then they were dried for an hour. During this time 0.5 McFarland standard (1 to 5 million cells/mL) *S. cerevisiae* (NCYC 1006) cells were prepared in a sterile water. The dried YEPD petri dishes were brought and the 0.5 McFarland *S. cerevisiae* cells were spread over the dishes using sterile cotton buds. The cells were spread evenly so that the killing activity of Panomycocin would be clear. Fifty μL samples were spotted on the petri dishes. Finally the petri dishes were put at increasing temperatures for incubation. The results were checked for a clear zone of killing after 48 hours. For each test a control plate was also prepared. In the control plates the sample spotted consist of 25 μL protein and 25 μL buffer.

These tests were done at 37.5°C, 38°C, 38.5°C and 39 °C. Excipients from various groups of excipients were tested at various concentrations. Sucrose, glucose, lactose, sorbitol, trehalose, lysine, glycine, proline and glutamic acid were tested at 0.005M, 0.02M, 0.05M, 0.1M, 0.2M, 0.5M, 1M and 1.25M concentrations. Mannitol and arginine were tested at 0.005M, 0.02M, 0.05M, 0.1M, 0.2M, 0.4M, 0.5M and 0.625M concentrations. Tween 80 and Pluronic F-68 were tested at %0.005, %0.01, %0.025,

%0.05, %0.1, %0.25, %0.5 and %1 concentrations. Histidine was tested at 0.005M, 0.002M, 0.05M, 0.1M, 0.15M, 0.2M and 0.25M concentrations. Glycerol was tested at %0.5, %1, %2, %2.5, %5, %10, %15 and %20 concentrations. Excipients were tested upto their maximum saturated solubility levels in water. Furthermore tests were also performed by mixing these excipients with each other at various concentrations.

2.3 Computational Methods

2.3.1 Protein Similarity Search

The homology of the internal and N-terminal amino acid sequences of Panomycocin was checked. FASTA program [136] was used to find protein similarity from the sequences. FASTA was opened with <http://www.ebi.ac.uk/Tools/sss/fasta/>. UniProtKB/Swiss-Prot was chosen as protein database. The amino acid sequences were put as an input protein. SSEARCH (local search) was selected as a program, submitted and the result was obtained. Various matrices were also used in ‘the additional options’ and the results were recorded.

2.3.2 Determination of Signal Peptide and KEX2 Cleavage Site

After FASTA search the signal peptide of the homologue protein (AJ222862) was determined by using various servers such as SignalP 4.1 [137], Signal-CF [138], PrediSi [139] and Signal-3L [140]. The species type was selected as eukaryotic, FASTA format of the protein was pasted as an input and ‘submit’ was clicked in these servers. The results appeared in a short period of time.

Cleavage site predictors such as PROSPER(PROtease Specificity Prediction serVER) [141] and SMART (Simple Modular Architecture Research Tool) [142] were used to predict the KEX2 cleavage site. According to the estimated cleavage sites, various models with different cleavage sites were generated. These models were compared with each other by their DOPE (Discrete Optimized Protein Energy) score and energy

results. In addition to this literatures related with the protein were used to decide the exact KEX2 cleavage site.

2.3.3 Homology Modeling

2.2.3.a Homology Modeling by MODELLER

2.3.3.a.1 Generating Models

After the KEX2 cleavage site was determined, the upstream sequence was cleaved and the modeling of downstream sequence was started. The FASTA format of the protein was obtained from UniProt [143] with its accession number (AJ222862). The sequence was analysed by BLAST (Basic Local Alignment Search Tool) [144]. NCBI (National Center for Biotechnology Information) [145] was opened. BLAST which is on the top right hand side under ‘popular resources’ menu was clicked. Among the list of programs, ‘protein blast’ was clicked. FASTA sequence, after upstream of the cleavage site was cleaved, was pasted into the query box. PSI-BLAST (Position Specific Iteration BLAST) algorithm was selected among the algorithms listed and ‘BLAST’ was clicked.

To download structures from PDB (Protein Data Bank) [146], www.rcsb.org/pdb was opened. PDB IDs of the templates were written in to the search box. In the window of the respective template, the arrow of ‘download files’ which is found in the right upper hand side was clicked. PDB file (text) was downloaded. Then the file was saved.

In order to access MODELLER [81] license agreement was filled and academic license key was obtained through e-mail. 32 bit version of the windows installer was downloaded and saved to desktop. MODELLER 9.11-32 bit was double clicked to start the installer. ‘User account control’ window warning about an unidentified program popped up and ‘allow’ was clicked and continued. The place to install MODELLER was specified and MODELLER license key was entered when

prompted. After the installation was over, Python 2.7 [147] was installed from Python website. At this point every thing was ready for modeling. Modeling steps are presented as follows.

In the first step query sequence and its templates were loaded. EasyModeller 4.0 [83] was opened. Under the tab 'load inputs' the FASTA format of the query sequence downstream of the KEX2 cleavage site was pasted into load query sequence. The sequence was checked whether there was a gap or not. The templates were loaded using 'add template' button under 'load template structures'.

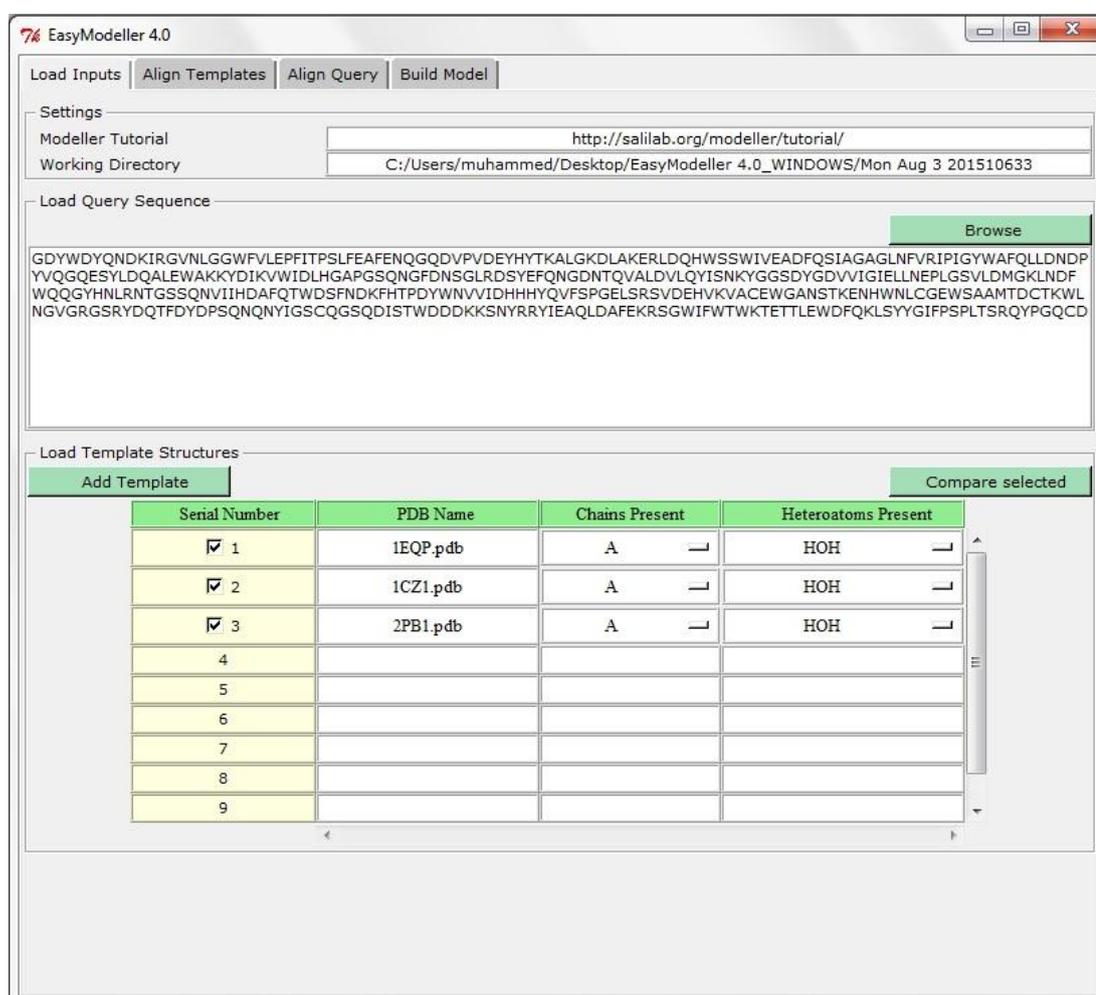


Figure 2.2 Query sequence and templates loaded into EasyModeller.

In the second step templates were aligned. The selected templates were aligned using 'align template' button under 'align templates' tab by clicking them. By doing this salign command of MODELLER was implemented. When 'align templates' was clicked, selected templates popped up and 'proceed' was clicked. In seconds information that shows the task is complete was popped up and 'ok' was clicked.

In the third step the query sequence was aligned with the templates. 'Align query' tab was clicked and then 'align query with templates' button was clicked to make the alignment.

In the last step models were generated. 'Build model' tab and 'generate models' button was clicked respectively to generate the models. In the box that popped up the number of models to be generated was increased to nine. The other options in the box were used in the default form. The effect of changing parameters on the model was also investigated later. This step is slower than the above steps and it takes a few minutes. 'DOPE profile' was clicked and all the models were selected to generate the DOPE profile plot under the DOPE profile viewer. After this 'view model' was clicked and each model was investigated and saved using UCSF Chimera [148].

2.3.3.a.2 Comparing the Generated Models

DOPE score, energy, TM (Template Modeling) score and RMSD (Root Mean Square Deviation) values were used to compare the generated models. MODELLER gives the DOPE score under the DOPE score viewer. On the other hand the rest of the parameters were calculated using other various methods.

The energy of the models was calculated with GROMOS96 (GRoningen MOlecular Simulation96) [114] in the SPDBViewer (Swiss PDB Viewer) [149]. The 'compute energy' under the 'tools' was clicked to calculate the energy. To find the minimized energy of the models each model was selected and 'energy minimization' under the 'tools' was clicked.

TM score calculator [113,150] was used to measure the TM score. Models and templates were given as inputs for comparison. Then 'run TM score' was clicked. And this was done for each model in relative to the three templates used.

VMD (Visual Molecular Dynamics) [151] was used to calculate the RMSD values of all models. At first VMD 1.9.2 was downloaded. VMD 1.9.2 was double clicked and VMD main was opened. In the 'VMD main file' the 'new molecule' was clicked. Molecule file browser appeared. Molecules were browsed and 'load' was clicked to upload the molecules as shown in figure 2.3. Then the mouse cursor was put over 'extensions and analysis' and then 'RMSD calculator' was clicked. RMSD tool appeared. The residue length and other specifications were changed according to the protein. Finally 'RMSD' in the RMSD tool was clicked. This was done for all models in relative to the templates.

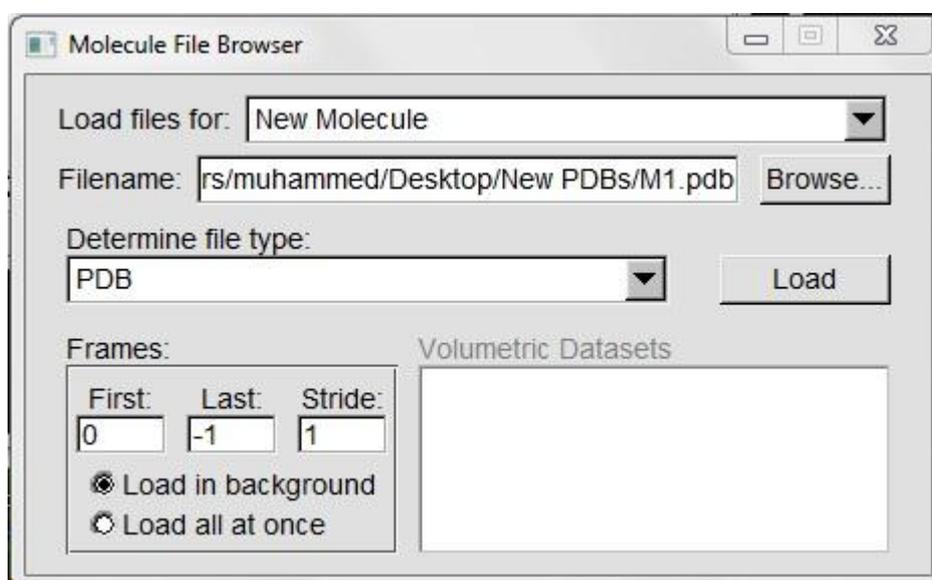


Figure 2.3 Models loaded to VMD.

2.3.3.a.3 Loop Modeling

MODELLER gives the opportunity of loop modeling. After the best model was determined by the above methods, loop modeling was undertaken over the best model (model 1). The DOPE profile of model 1 with respect to templates, structural analysis and the analysis of the alignment between the query and the templates were used to determine the regions that loop modeling might give a better model.

Loop modeling was undertaken at various positions. The effect of loop modeling at these positions was analysed by the DOPE score and energy of the resulting models. Furthermore TM score and RMSD values were calculated. Then the positions that gave a better result were considered in the following steps. The results obtained were compared with model 1. However, to compare the effect of loop modeling up to the last stage, thermostabilization was done for both models.

2.3.3.a.4 Optimization, Verification and Validation

Optimization of the best model (model 1) and its loop model was done by MODELLER. The models were selected and 'optimize' was clicked to get the optimized model.

The optimized models were evaluated with SAVES (Structure Analysis and Verification Server). PDB format of the models was uploaded in the SAVES server. 'Run all programs on this file' was clicked. Then the results obtained were recorded.

2.2.3.b Homology Modeling by I-TASSER

In order to compare and learn how accurate the model generated with MODELLER was, another model was generated by I-TASSER (Iterative Threading ASSEMBLY Refinement) server [84,85,86]. At first registration to I-TASSER server was done. Using the e-mail and password of registration, FASTA format of the protein downstream of the KEX2 cleavage site was submitted to the server. In two days the result of modeling was available. After the I-TASSER model was generated, its energy, TM score and RMSD values were calculated with the methods explained before. I-TASSER model was also evaluated with SAVES. Then the two models which were generated with MODELLER and I-TASSER were compared.

2.3.4 Determination of Binding Site

COACH [152], COFACTOR [153], MetaPocket [154], CASTp (Computed Atlas of Surface Topography of Proteins) [155] and DoGSiteScorer [156] were used to predict the binding site of the models. FASTA format of the protein downstream of the KEX2 cleavage site was submitted to COACH. COACH gave combined results with COFACTOR, FINDSITE and ConCavity too. PDB format of the models was uploaded and 'run COFACTOR' was clicked to do COFACTOR calculations. For MetaPocket PDB format was uploaded and the number of pockets to be generated was specified. For CASTp 'calculation request' was clicked and PDB format was submitted. The results of the above methods were obtained through e-mail. For DoGSiteScorer PDB format was uploaded and 'calculate and analyse pockets' was clicked. The result was presented soon and it was saved.

2.3.5 Thermostabilization of Models

Thermostabilization of the models was first done with GROMOS96 in the SPDBViewer. Energy of each position in the models was predicted and energy profiles of the alignment were investigated. All of the amino acids in the unstable regions were substituted with the rest nineteen amino acids. Those which gave lower energy were recorded. Among them the best ten were selected. 'Compute energy' under the 'tools' was clicked to calculate the energy. To find the minimized energy of the models the models were selected and 'energy minimization' under the 'tools' was clicked. Mutations were done with SPDBViewer. The respective models were opened in the SPDBViewer. To color the position that will be mutated differently 'wind' and then 'control panel' was clicked. In the control panel the box beside the amino acid was clicked and the suitable color was selected. 'Mutate' was clicked and went over the marked amino acid in the opened model. As the mouse cursor is over it, it was clicked. At this time all the twenty amino acids appeared and by clicking on one of them the amino acid that will substitute the position was selected. 'Mutate' was clicked again. Then a box that gives a general information about the mutation popped up. By clicking 'yes' the mutation of the position with the amino acid selected was accepted.

After the ten best positions were determined, the combinations among them were also tested. Those that gave a higher energy were removed. In addition to this the best three thermostabilizer positions in the binding site were determined.

The above calculations were in KJ/mol. These changes were changed in to temperature (K) by CNA (Constraint Network Analysis) [123] server. Various programs were used to make the models ready for analysis. The PDB format was opened with SPDBViewer. 'All' under 'select' menu was clicked to select the model. The energy was minimized using 'energy minimization' under the 'tools' menu and was saved. And then this was opened with UCSF Chimera. In the Chimera 'tools' menu was clicked and went over 'structure editing' and 'add H' was clicked to add hydrogen. This was also saved and opened with SPDBViewer again. Since the CNA server showed that there was deficiency of residues at isoleucine's of the models, deficient

residues were added. To do this 'select' menu was clicked and went over 'group kind' and then 'Ile (I)' was clicked. At last 'build' and then 'add residues' was clicked. As the model was ready, it was submitted to CNA server. Single network-single structure was marked as analysis structure and stability map was requested for the output. The other parameters were set as default. The results were obtained through e-mail.

Furthermore to increase the accuracy of the predictions made various servers that are used to estimate the effect of mutation on a protein were also used. I-Mutant2.0 [157] is one of the commonly used methods for this purpose. I-Mutant2.0 main page was opened and 'protein sequence' was marked as an input for single point mutation stability changes. Protein sequence was pasted, the positions and new residues were entered and free energy change calculations were requested. The results were obtained through e-mail. Next PDB format of the protein was given, the mutations made were specified and this was submitted to AUTO-MUTE [130]. The results were available soon. Eris [132] server was the other program used. First membership registration was done. PDB file of the protein was given as an input. The mutations were done over the sequence presented by the server. The other parameters were set as default and the task was submitted. The results were sent to the e-mail of registration. For MUpro [131] protein sequence which is recommended by the server was given as an input. The positions and substitutions were entered one by one. The results were available soon.

CHAPTER 3

RESULTS

3.1 Laboratory Results

3.1.1 Production of Crude Panomycocin

Panomycocin was produced and isolated according to the method described by Izgü and Altınbay [33]. Since the production of Panomycocin depends on the pH of the cultivation medium and incubation temperature, *P. anomala* cells were grown in YEPD pH 4.5 medium with the addition of 5% glycerol as protein stabilizer at 18 °C to maintain high production and thus the highest degree of killing activity [33].

3.1.2 Determination of Killer Activity of Panomycocin

To determine the killer toxin activity of Panomycocin YEPD agar plates with pH 4.5 were prepared. *S. cerevisiae* (NCYC 1006) cells with a 0.5 McFarland standard (1 to 5 million cells/mL) were prepared in a sterile water. These cells were spread and cultivated into the plates with cotton buds. Fifty µL Panomycocin samples were spotted onto petri dishes and incubated at 25°C. A clear growth inhibition zone was observed after 24 hours of incubation at 25 °C (figure 3.1).

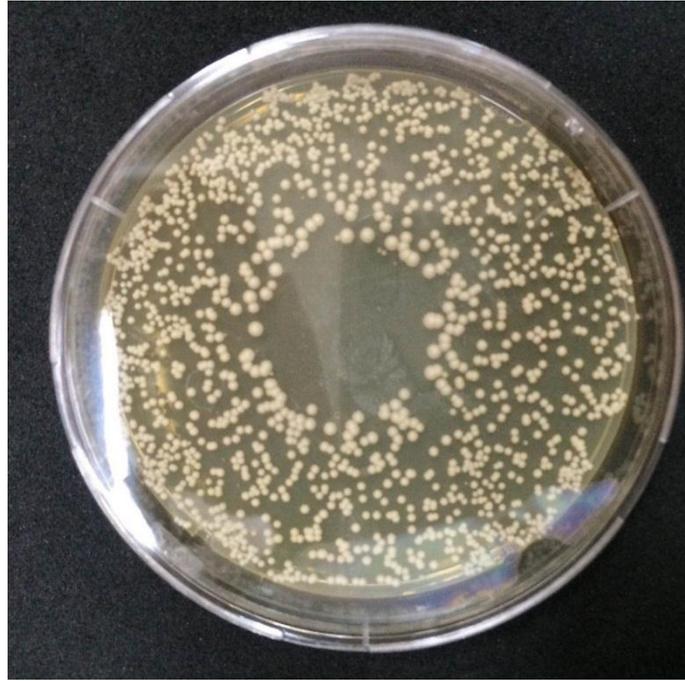


Figure 3.1 Killer activity of Panomycocin.

3.1.3 Isolation of Panomycocin

Centrifugation was used to remove *P. anomala* cells from the culture medium, cellulose membrane filters were used to sterilize the supernatant obtained from centrifugation and then concentration was made by ultrafiltration systems. Furthermore concentrated killer protein was purified by using a fully automated FPLC system (Biocad 700E Perseptive Biosystems, USA) with a gel filtration column TSK G 2000SW (Particle size 10 μm , Pore size 125 \AA , Sample MW 5 - 100 kDa) (figure 3.2). The fraction that consists of Panomycocin was eluted at approximately 11.5 mL as it is depicted in figure 3.2.

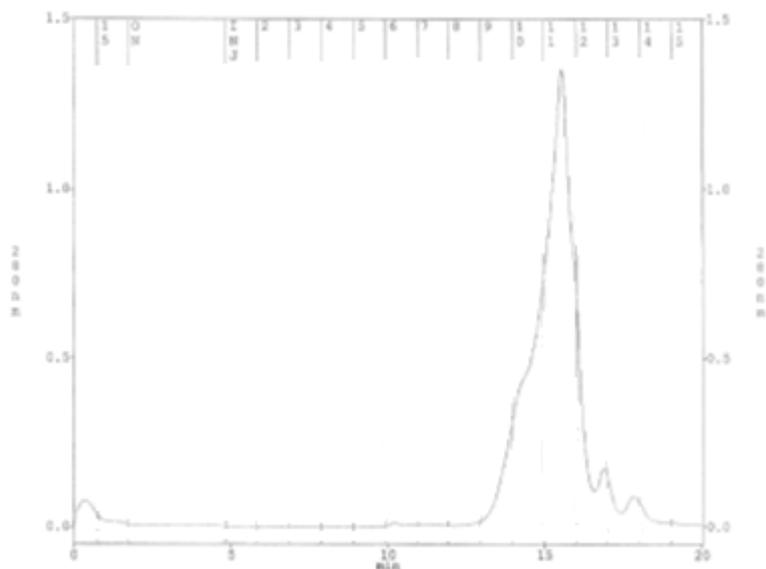


Figure 3.2 Elution profile of Panomycocin on TSK G2000SW column. Column size: 7.5 mmD/ 30 cmL; Sample: 80 μ L; Elution buffer: 100 mM Na_2HPO_4 pH 4.5 with 100 mM Na_2SO_4 ; flow rate 1 mL/min; Detection: 280 nm UV. The fraction containing Panomycocin was eluted at 11.5 mL.

After some runs the active fractions collected were concentrated by using 5 kDa molecular mass cut off ultrafilters (Vivaspin VS2021, Sartorius, AG, Germany) to a desired protein concentration.

3.1.4 Determination of Protein Concentration

Nanodrop 2000 UV-Vis spectrophotometer at a wavelength of 280 nm was used to determine the protein concentration after purification and it was measured as 8 μ g/mL.

3.1.5 SDS Polyacrylamide Gel Electrophoresis

For purity test nearly ten μ g isolated Panomycocin was electrophoresed on a 12.5% linear SDS-PAGE gel in a discontinuous buffer system under denaturing conditions. Single protein band was observed on the coomassie brilliant blue stained gel that

shows the absence of contamination in the sample. The protein band on lane 2 between d and e markers was found to be the band that belongs to Panomycocin (figure 3.3) since the molecular mass of Panomycocin is 49 kDa.

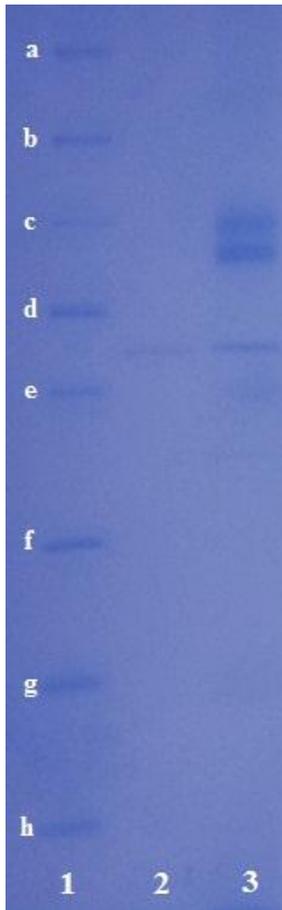


Figure 3.3 Denatured SDS-PAGE. Lane 2 is pure Panomycocin, lane 3 is crude Panomycocin and lane 1 is molecular mass markers (kDa) a) α 2-macroglobulin (170), b) β -galactosidase (116.353), c) fructose-6-phosphate kinase (85.204), d) glutamate dehydrogenase (55.562), e) aldolase (39.212), f) triose phosphate isomerase (26.626), g) trypsin-inhibitor (20.100), h) lysozyme (14.307).

3.1.6 Results of Thermostability Tests Using Excipients

Fifty μL samples which consist of 25 μL protein and 25 μL excipients with gradient of concentrations were prepared. These samples were spotted on dried YEPD petri dishes at pH 4.5 in which 0.5 McFarland standard (1 to 5 million cells/mL) *S. cerevisiae* (NCYC 1006) cells were spread over. These petri dishes were put at increasing temperatures from 37.5 °C to 39 °C for incubation. The results were checked for a clear zone of killing after 48 hours.

These tests were done using excipients from various groups of excipients that includes sucrose, glucose, lactose, mannitol, sorbitol, trehalose, tween 80, pluronic F-68, histidine, lysine, arginine, glycine, proline and glutamic acid. However, none of them increased the thermostability of Panomycocin (table 3.1). The results of glycine are presented in figure 3.4 to demonstrate.

Table 3.1 Results of the killing activity of the protein with excipients at various temperatures.

Excipients	Concentration of excipients	Amount of excipient solution (in μL)	Amount of Panomycocin (in μL)	Killing activity of the total mixture (50 μL) at various temperatures (in $^{\circ}\text{C}$)			
				37.5	38	38.5	39
Glucose, Glutamic acid, Glycine, Lactose, Lysine, Proline, Sorbitol, Sucrose, Trehalose	0.005M	25	25	-	-	-	-
	0.02M	25	25	-	-	-	-
	0.05M	25	25	-	-	-	-
	0.1M	25	25	-	-	-	-
	0.2M	25	25	-	-	-	-
	0.5M	25	25	-	-	-	-
	1M	25	25	-	-	-	-
	1.25M	25	25	-	-	-	-
Arginine, Mannitol	0.005M	25	25	-	-	-	-
	0.02M	25	25	-	-	-	-
	0.05M	25	25	-	-	-	-
	0.1M	25	25	-	-	-	-
	0.2M	25	25	-	-	-	-
	0.4M	25	25	-	-	-	-
	0.5M	25	25	-	-	-	-
	0.625M	25	25	-	-	-	-
Histidine	0.005M	25	25	-	-	-	-
	0.02M	25	25	-	-	-	-
	0.05M	25	25	-	-	-	-
	0.1M	25	25	-	-	-	-
	0.15M	25	25	-	-	-	-
	0.2M	25	25	-	-	-	-
	0.25M	25	25	-	-	-	-
Pluronic F-68, Tween 80	%0.005	25	25	-	-	-	-
	%0.01	25	25	-	-	-	-
	%0.025	25	25	-	-	-	-
	%0.05	25	25	-	-	-	-
	%0.1	25	25	-	-	-	-
	%0.25	25	25	-	-	-	-
	%0.5	25	25	-	-	-	-
	%1	25	25	-	-	-	-
Glycerol	%0.5	25	25	-	-	-	-
	%1	25	25	-	-	-	-
	%2	25	25	-	-	-	-
	%2.5	25	25	-	-	-	-
	%5	25	25	-	-	-	-
	%10	25	25	-	-	-	-
	%15	25	25	-	-	-	-
	%20	25	25	-	-	-	-

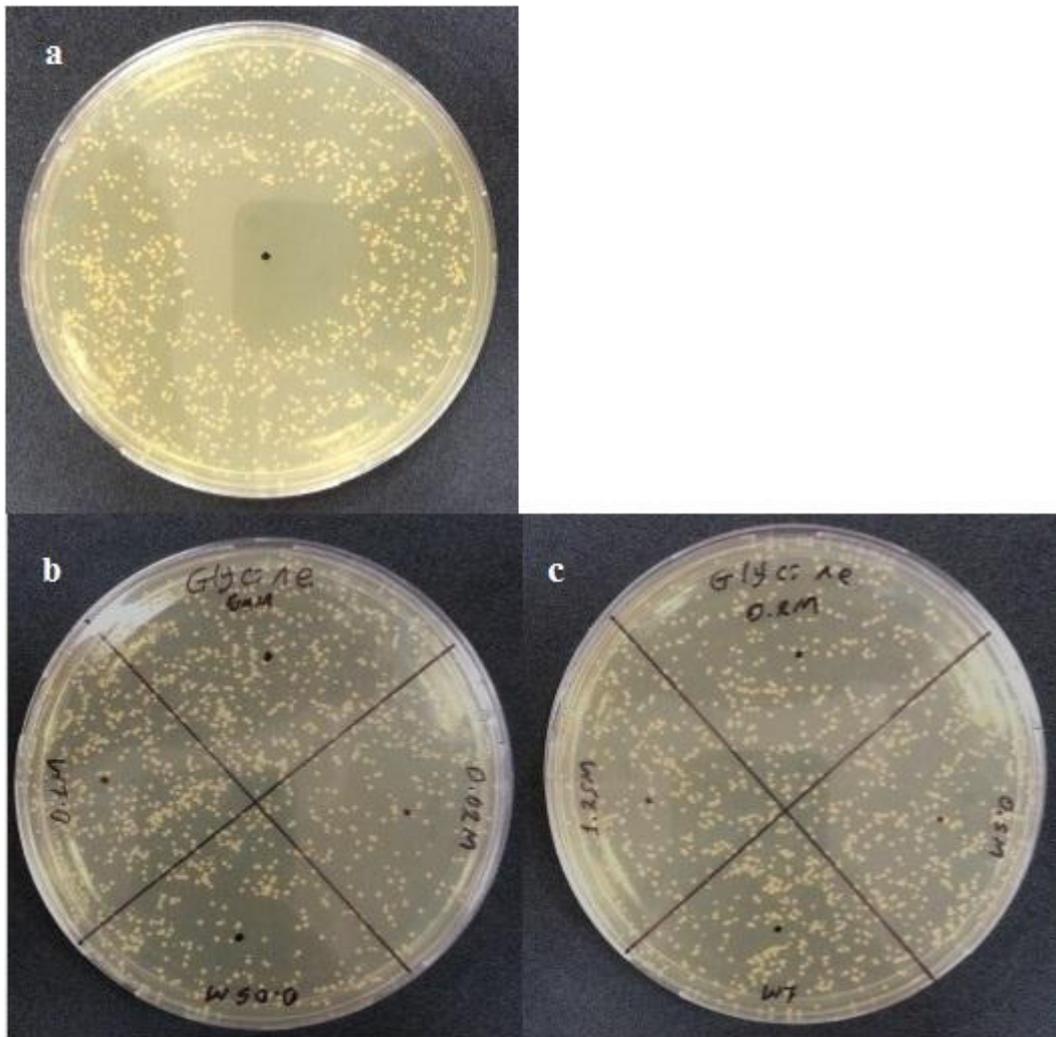


Figure 3.4 Results of thermostability tests with excipients a) killing activity of Panomycocin at 25°C, b) results of glycine at 0.005M, 0.02M, 0.05M and 0.1M concentrations, c) results of glycine at 0.2M, 0.5M, 1M and 1.25M concentrations.

3.2 Computational Results

3.2.1 Protein Similarity Search

Matrices with high target identity like MDM10 and VTML10 gave %100 identities and positives with some proteins as shown in figure 3.5.

Align.	DB:ID	Source	Length	Score	Identities %	Positives %	E()
<input checked="" type="checkbox"/> 1	SP:EXG_CANOL	Glucan 1,3-beta-glucosidase OS=Candida oleophila GN=EXG1 PE=3 SV=1 <i>Cross-references and related information in:</i> ▶ Small molecules ▶ Nucleotide sequences ▶ Enzymes ▶ Samples & ontologies ▶ Protein families ▶ Literature ▶ Protein sequences	425	129	100.0	100.0	2.3E-11
<input checked="" type="checkbox"/> 2	SP:EXG2_HANAN	Glucan 1,3-beta-glucosidase 2 OS=Hansenula anomala GN=EXG2 PE=3 SV=1	427	129	100.0	100.0	2.3E-11
<input checked="" type="checkbox"/> 3	SP:EXG1_CANAL	Glucan 1,3-beta-glucosidase OS=Candida albicans (strain SC5314 / ATCC MYA-2876) GN=XOG1 PE=1 SV=4	438	129	100.0	100.0	2.4E-11

Figure 3.5 FASTA search results of matrices with high target identity.

But matrices with low target identity gave lower percentage of identities and positives. For example, the default matrix BLOSUM50 gave %81 identities and %90.5 positives as the best homologue (figure 3.6).

Align.	DB:ID	Source	Length	Score	Identities %	Positives %	E()
<input checked="" type="checkbox"/> 1	SP:EXG2_HANAN	Glucan 1,3-beta-glucosidase 2 OS=Hansenula anomala GN=EXG2 PE=3 SV=1 <i>Cross-references and related information in:</i> ▶ Small molecules ▶ Nucleotide sequences ▶ Enzymes ▶ Samples & ontologies ▶ Protein families ▶ Protein sequences	427	129	81.0	90.5	2.6E-4
<input checked="" type="checkbox"/> 2	SP:EXG1_CANAL	Glucan 1,3-beta-glucosidase OS=Candida albicans (strain SC5314 / ATCC MYA-2876) GN=XOG1 PE=1 SV=4 <i>Cross-references and related information in:</i> ▶ Small molecules ▶ Nucleotide sequences ▶ Enzymes ▶ Samples & ontologies ▶ Protein families ▶ Literature ▶ Macromolecular structures ▶ Protein expression data ▶ Protein sequences	438	129	54.5	68.2	2.7E-4

Figure 3.6 FASTA search results of BLOSUM50 which has low target identity.

The FASTA program analysis that was performed showed that $\text{exo-}\beta\text{-1,3-glucanase}$ from *Hansenula (Pichia) anomala* is the best homologue of the sequences of the protein. The sequence was obtained from UniProt using its accession number (AJ222862) [143]. The sequence found is gene sequencing with 427 amino acids. The sequence obtained from UniProt is presented below.

```
MLISTFISSLLSIALANPIPSRGGTQFYKRGDYWDYQNDKIRGVNLGGWFVL
EPFITPSLFEAFENQGGQDVPVDEYHYTKALGKDLAKERLDQHWSSWIVEADF
QSIAGAGLNFVRIPIGYWAFQLLDNDPYVQGQESYLDQALEWAKKYDIKVV
IDLHGAPGSQNGFDNSGLRDSYEFQNGDNTQVALDVLQYISNKYGGSDYGD
VVIGIELLNEPLGSVLDMGKLNDFWQQGYHNLNRNTGSSQNVIHDAFQTWD
SFNDKFHTPDYWNVVIDHHHYQVFSPGELSRVDEHVKVACEWGANSTKE
NHWNLCEGWSAAMTDCTKWLNGVGRGSRYDQTFDYDPSQNQNYIGSCQG
SQDISTWDDDKKSNYRRYIEAQLDAFEKRSWIFWTWKTETTLEWDFQKLS
YYGIFPSPLTSRQYPGQCD
```

As a result modeling and thermostabilization works were done with this sequence.

3.2.2 Signal Peptide and KEX2 Cleavage Site

SignalP 4.1, Signal-CF, PrediSi and Signal-3L results showed that the signal peptide includes the first seventeen amino acids. SignalP 4.1 results are presented in figure 3.7.

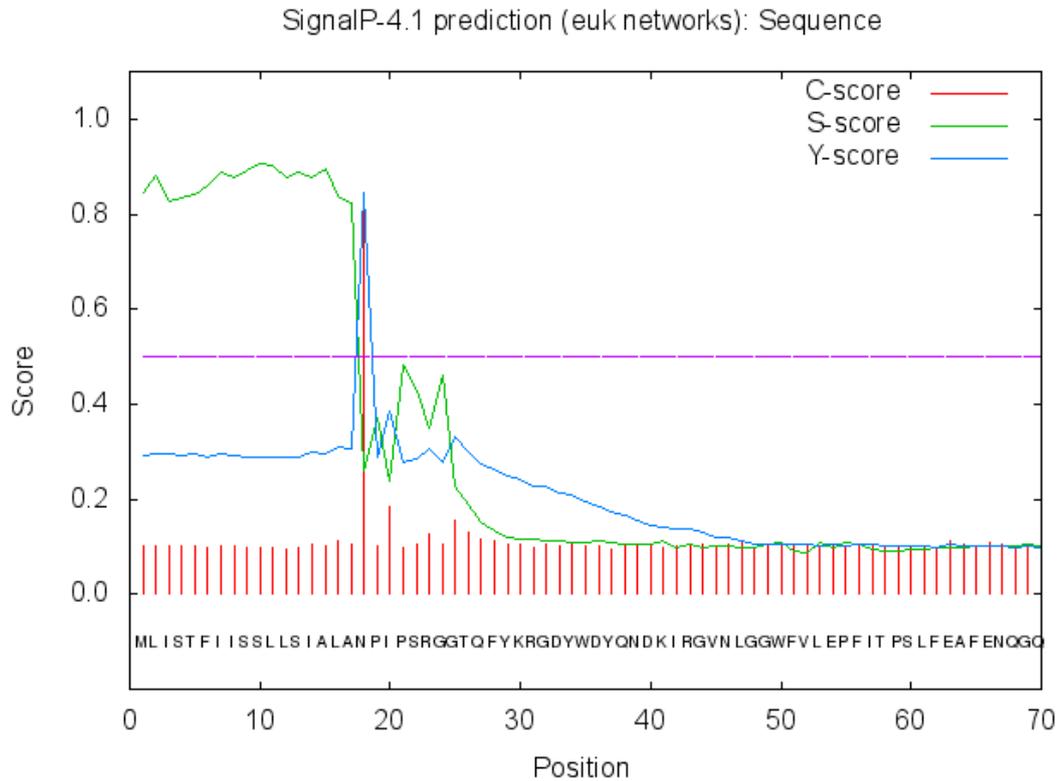


Figure 3.7 Signal peptide results of SignalP 4.1: C-score is high immediately after the cleavage site; S-score is high at all positions before the cleavage site; Y-score is the estimation of cleavage site that is optimized by considering regions where the C-score is high and the S-score changes from high to low value.

As it was estimated by SMART server, KEX2 cleavage may occur at 30-32 positions as shown in figure 3.8. In addition to this amino acid sequencing at the N-terminal showed that the secreted mature protein starts with glycine at 32. position. As a result the cleavage occurs after the 31. position on -Arg-Gly- bond.

Elm Name	Instances (Matched Sequence)	Positions	View in Jmol	Elm Description	Cell Compartment	Pattern	PHI-Blast Instance Mapping	Structural Filter Info	Probability
CLV_PCSK_KEX2_1	KRG	30-32 [A]	-	Yeast kexin 2 cleavage site (K-R-I-X or R-R-I-X).	extracellular, Golgi apparatus	[KR]R.	-	-	7.973e-03
CLV_PCSK_PC1ET2_1	KRG	30-32 [A]	-	NEC1/NEC2 cleavage site (K-R-I-X).	extracellular, Golgi apparatus, Golgi membrane	KR.	-	-	3.903e-03

Figure 3.8 KEX2 cleavage site estimation of SMART server.

3.2.3 Homology Modeling

3.2.3.a Homology Modeling by MODELLER

3.2.3.a.1 Generating Models

The BLAST result showed that 1EQP, 1CZ1 and 2PB1 are the best homologs with the protein with %98 coverage and %66 identity.

Description	Max score	Total score	Query cover	E value	Ident	Accession	Select for PSI blast	Used to build PSSM
<input type="checkbox"/> RecName: Full=Glucan 1,3-beta-glucosidase 2; AltName: Full=Exo-1,3-beta-glucanase 2; Flags: Precursor [Wickerha	821	821	100%	0.0	100%	Q93983.1	<input checked="" type="checkbox"/>	
<input type="checkbox"/> exo-1,3-beta-glucanase [Wickerhamomyces anomalus]	816	816	100%	0.0	99%	AGC67022.1	<input checked="" type="checkbox"/>	
<input type="checkbox"/> exo-1,3-beta-glucanase [Wickerhamomyces anomalus]	814	814	100%	0.0	99%	AFK33205.1	<input checked="" type="checkbox"/>	
<input type="checkbox"/> exo-beta-1,3-glucanase [Wickerhamomyces anomalus]	813	813	100%	0.0	99%	ABK40520.1	<input checked="" type="checkbox"/>	
<input type="checkbox"/> putative beta-1,3-glucanase [Wickerhamomyces anomalus]	813	813	100%	0.0	99%	AEZ66637.1	<input checked="" type="checkbox"/>	
<input type="checkbox"/> exo-1,3-beta-glucanase [Wickerhamomyces anomalus]	813	813	100%	0.0	99%	AFK33203.1	<input checked="" type="checkbox"/>	
<input type="checkbox"/> putative beta-1,3-glucanase [Wickerhamomyces anomalus]	811	811	100%	0.0	99%	AEZ66638.1	<input checked="" type="checkbox"/>	
<input type="checkbox"/> exo-1,3-beta-glucanase [Wickerhamomyces anomalus]	808	808	100%	0.0	98%	CDF77671.1	<input checked="" type="checkbox"/>	
<input type="checkbox"/> Glucan 1,3-beta-glucosidase 2 [Wickerhamomyces ciferrii]	630	630	98%	0.0	75%	XP_011273243.1	<input checked="" type="checkbox"/>	
<input type="checkbox"/> CYFA0S09a02652g1_1 [Cyberlindnera fabianii]	615	615	99%	0.0	74%	CDR42395.1	<input checked="" type="checkbox"/>	
<input type="checkbox"/> exo-beta-1,3-glucanase [Cyberlindnera saturnus]	572	572	98%	0.0	69%	ACP74152.2	<input checked="" type="checkbox"/>	
<input type="checkbox"/> XOG1 [Cyberlindnera iadinii]	565	565	98%	0.0	69%	CEP22307.1	<input checked="" type="checkbox"/>	
<input type="checkbox"/> glucan 1,3-beta-glucosidase [Candida albicans P76067]	538	538	98%	0.0	66%	KHC43133.1	<input checked="" type="checkbox"/>	
<input type="checkbox"/> hypothetical protein CaO19.10507 [Candida albicans SC5314]	538	538	98%	0.0	66%	XP_721216.1	<input checked="" type="checkbox"/>	
<input type="checkbox"/> hypothetical protein CaO19.2990 [Candida albicans SC5314]	538	538	98%	0.0	66%	XP_721488.1	<input checked="" type="checkbox"/>	
<input type="checkbox"/> glucan 1,3-beta-glucosidase [Candida albicans P34048]	537	537	98%	0.0	66%	KGU33678.1	<input checked="" type="checkbox"/>	
<input checked="" type="checkbox"/> Chain A, Exo-B-(1,3)-Glucanase From Candida Albicans	536	536	98%	0.0	66%	1EQP_A	<input checked="" type="checkbox"/>	
<input checked="" type="checkbox"/> Chain A, Exo-B-(1,3)-Glucanase From Candida Albicans At 1.85 A Resolution	535	535	98%	0.0	66%	1CZ1_A	<input checked="" type="checkbox"/>	
<input type="checkbox"/> exo-1,3-beta-glucanase [III] putative; glucan 1,3-beta-glucosidase [III] precursor; putative [Candida dubliniensis CD36]	535	535	98%	0.0	66%	XP_002416951.1	<input checked="" type="checkbox"/>	
<input type="checkbox"/> beta-glucanase [Candida albicans]	535	535	98%	0.0	65%	CAA21969.1	<input checked="" type="checkbox"/>	
<input checked="" type="checkbox"/> Chain A, Exo-b-(1,3)-glucanase From Candida Albicans In Complex With Unhydrolysed And Covalently Linked 2,4-di	535	535	98%	0.0	66%	2PB1_A	<input checked="" type="checkbox"/>	

Figure 3.9 PSI-BLAST search result of the protein. The marked proteins 3D structure has been determined.

There are proteins which have upto %100 identity and coverage with the protein but their 3D structure has not been determined yet. As a result those proteins whose structures were determined and put into PDB were selected as templates. The three proteins marked in figure 3.9, with PDB ID: 1EQP, 2PB1 and 1CZ1, were selected as templates. They have %66 identity and %98 coverage in relative to the protein.

There was not any gap in the protein sequence. So the next steps were pursued and the templates were aligned (figure 3.10). The templates were found to have almost the same sequence.

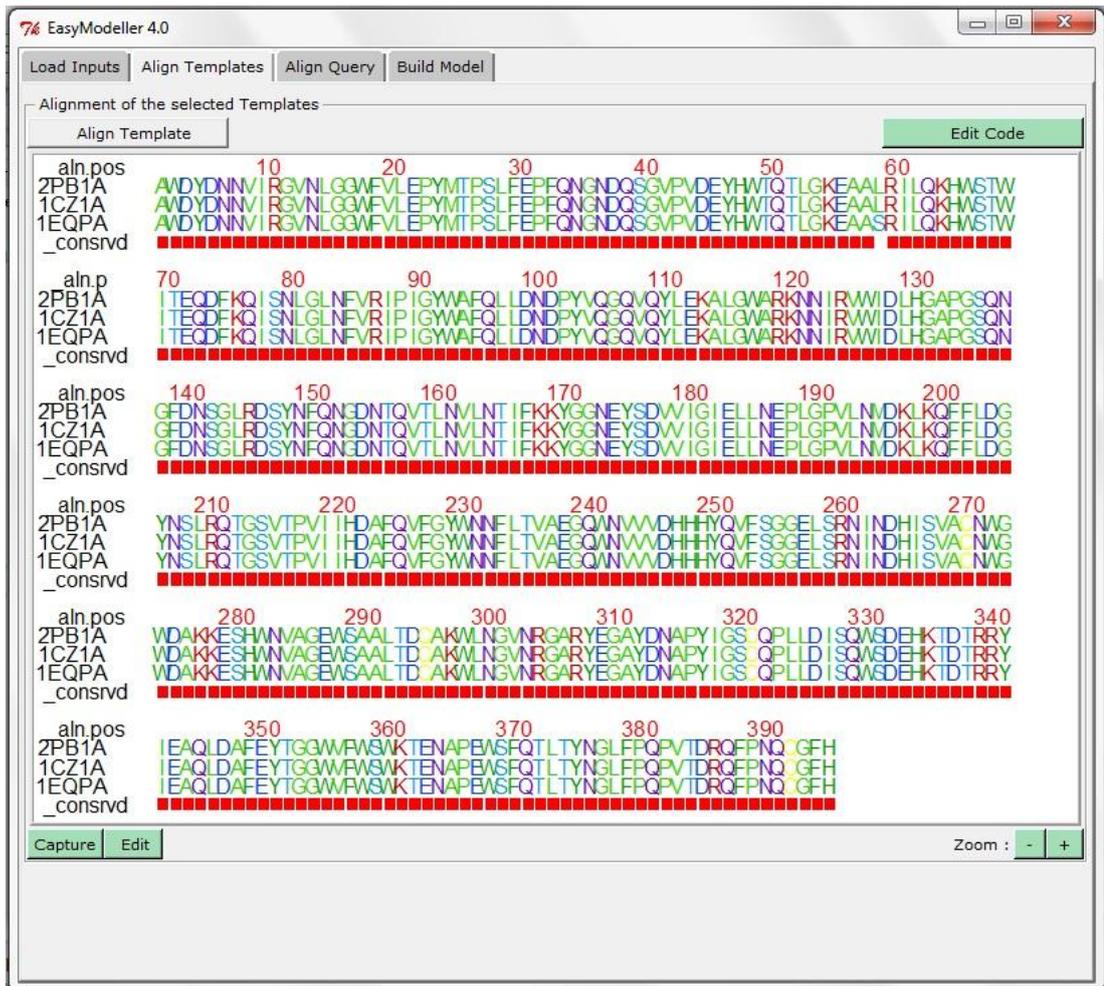


Figure 3.10 Alignment of the templates. The amino acids in the alignment were colored in accordance with the similarity and the conserved residues were marked by a red square below the alignment.

Then the query sequence was aligned with the templates. Since a wide gap was not observed in the predicted secondary structure regions of the alignment, the code of the alignment was not edited.

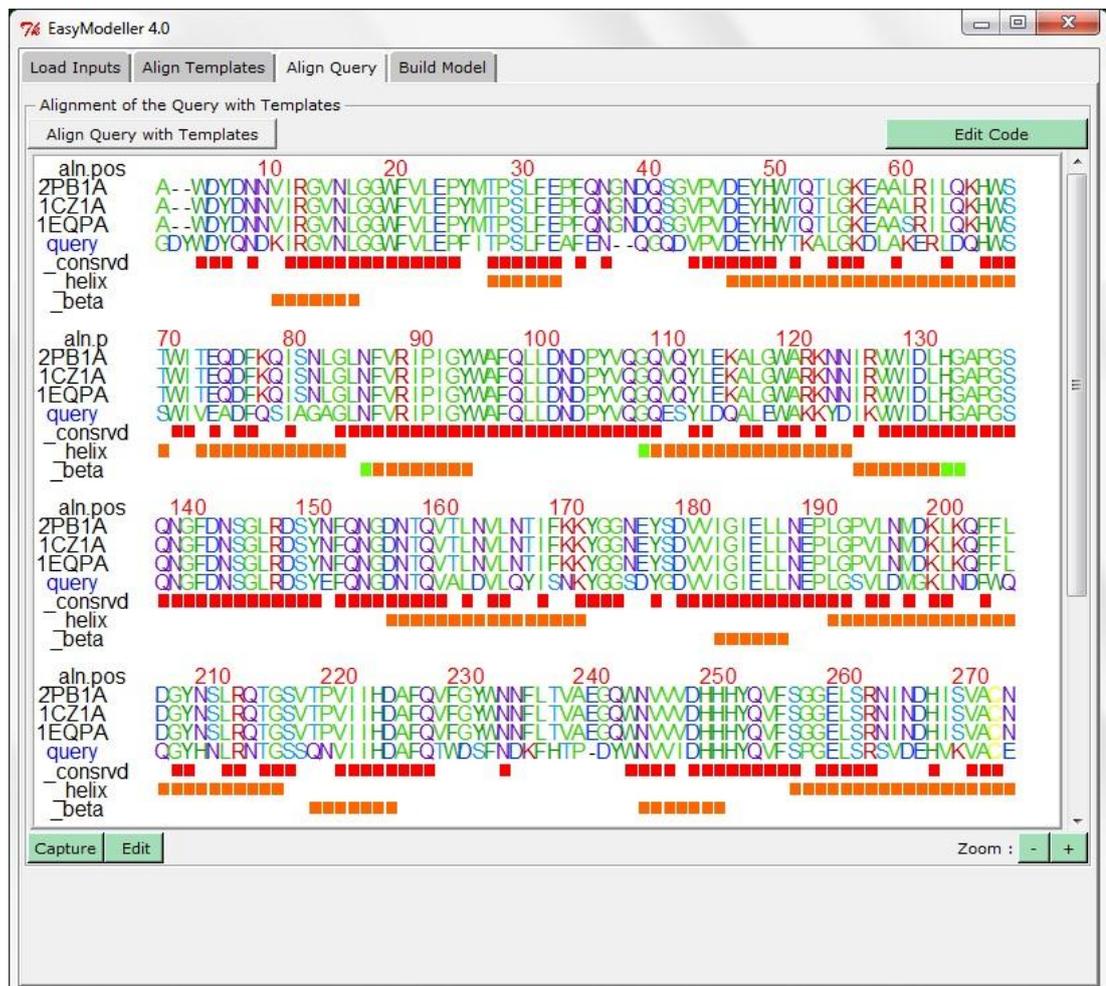


Figure 3.11 Alignment of the query with the templates. In the result window conserved residues in the alignment are shown in red squares. Alpha and beta helices are also predicted. The occurrence of these secondary structures is a function of color such that deeper red indicates high confidence and deeper green indicates low confidence.

At last the models were generated. DOPE profile was also plotted as shown in figure 3.12.

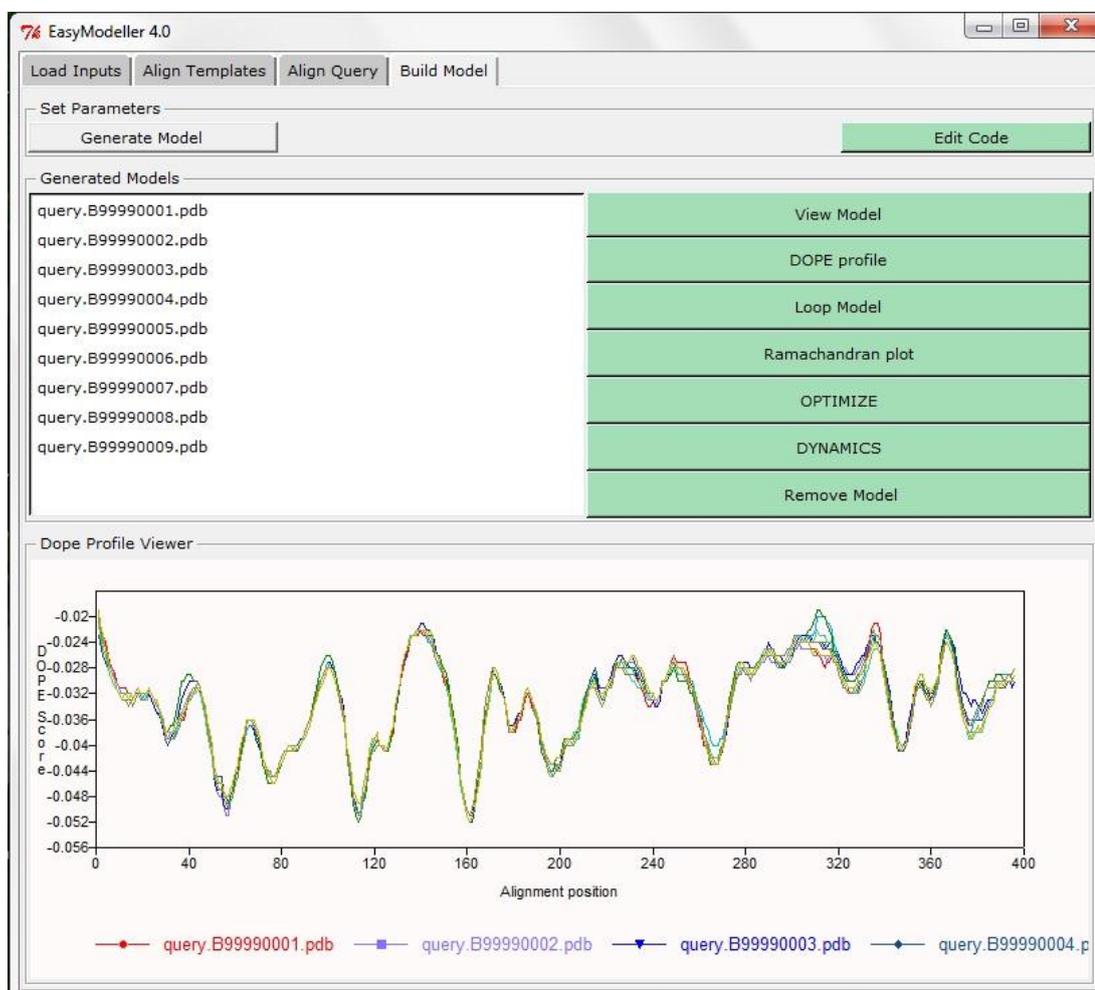


Figure 3.12 Generated models and their DOPE profile.

3.2.3.a.2 Determination of the Best Model

Generated models were compared with DOPE score, energy, TM score and RMSD values. MODELLER gives DOPE score under DOPE profile viewer when it is requested. The results were represented in a graph (figure 3.13).

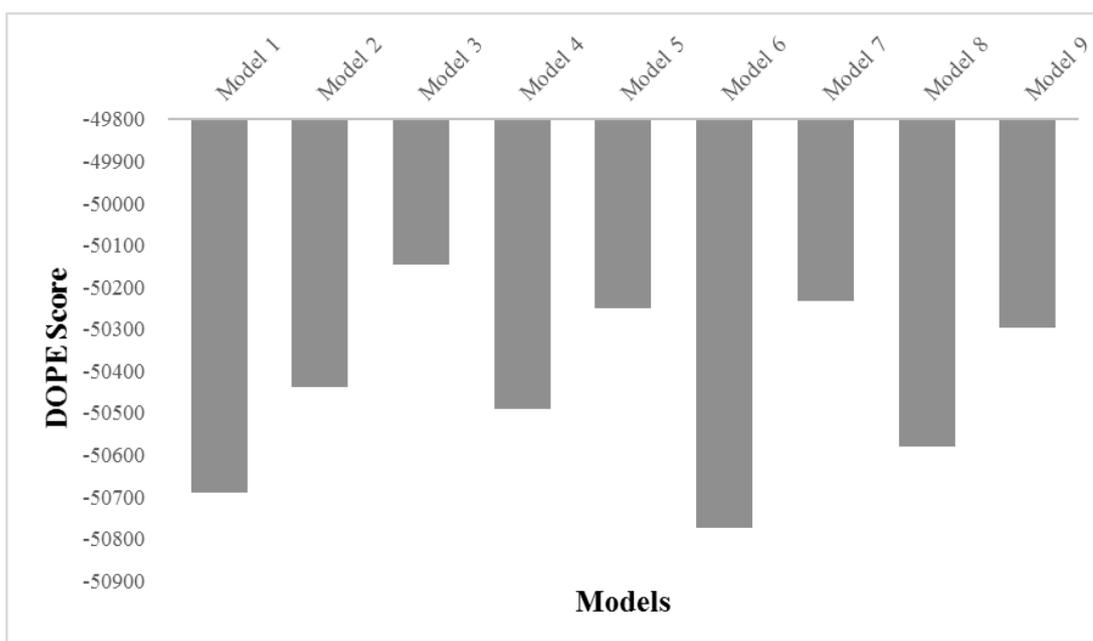


Figure 3.13 Graph of DOPE score of the models.

Energies of the models were calculated with GROMOS96 in Swiss PDBViewer. It was calculated separately for each model and the results were put in a graph (figure 3.14).

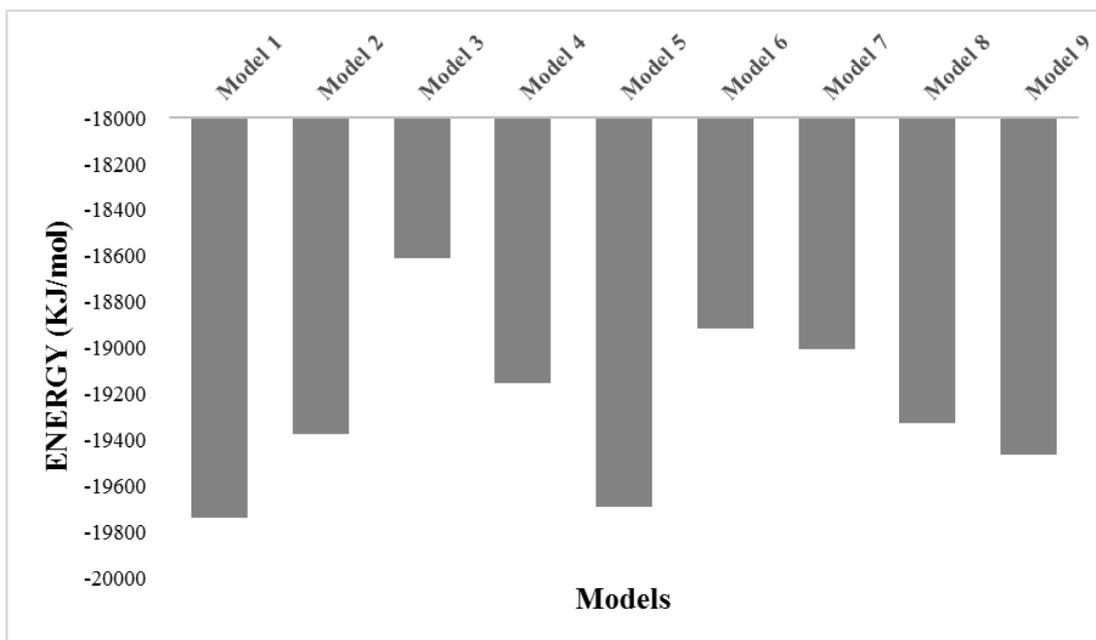


Figure 3.14 Graph of energy of the models.

TM scores were calculated for each model in relative to the three templates used. The results are shown in table 3.2.

Table 3.2 TM scores for the models.

Model	1	2	3	4	5	6	7	8	9
Mean TM	0.4099	0.4108	0.4113	0.4109	0.4118	0.4117	0.4121	0.4123	0.4099

RMSD values were calculated with VMD. Calculations were done for all models with respect to a template. This was done for all the templates. The average of the resulting RMSD values were calculated and shown in table 3.3.

Table 3.3 RMSD values of the models.

Model	1	2	3	4	5	6	7	8	9
Mean RMSD (Å)	5.55 40	6.32 75	5.03 80	5.48 37	5.68 13	7.20 66	5.92 65	6.20 07	6.44 63

As it was shown in the above graphs and tables, model 1 gave the best result in energy and second in DOPE score. In addition to this it gave a good result in measurements related with topology and position even if it was not the best. Since model 1 gave a good result in all measurements, it was chosen as the best model. So that the next stages were continued with it.

3.2.3.a.3 Loop Modeling

DOPE profile of model 1 with respect to templates as shown in figure 3.15, structural analysis and analysis of the alignment between the query and the templates were used to determine the regions that loop modeling might give a better model. As a result loop modeling was undertaken at 1-5, 228-242, 311-320, 330-340 and 358-362 positions.

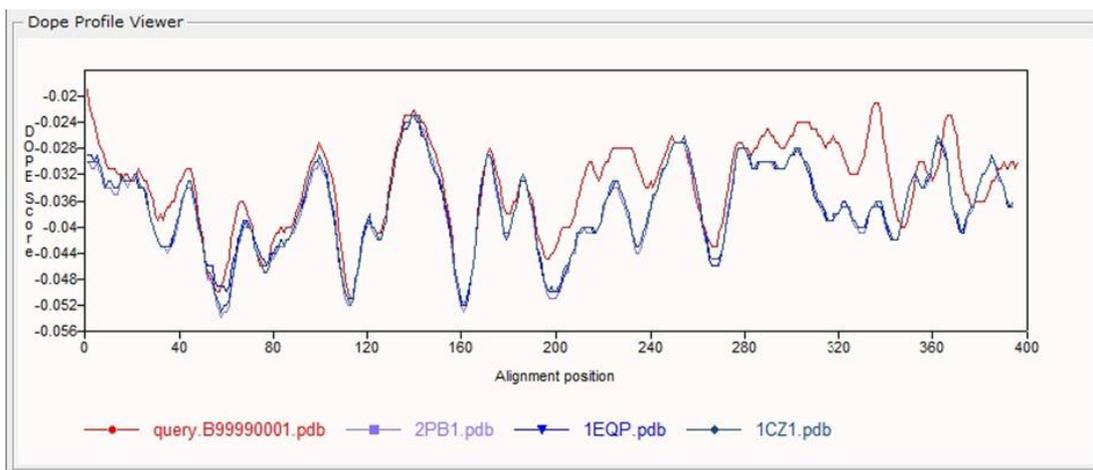


Figure 3.15 DOPE score profile of model 1 relative to the templates. Red is profile of model 1 and the others are templates.

DOPE score and energy of the models were calculated and the results were summarized in table 3.4.

Table 3.4 DOPE score and energies of loop modeling at different positions.

Loop Modeling Positions	DOPE Score	Energy (KJ/mol)
1-5	-50587.976562	-19859.064
228-242	-49398.257812	-17632.781
311-320	-50808.406250	-19918.740
330-340	-50226.585938	-19458.766
358-362	-50715.792969	-19540.637

Loop modeling at positions 311-320 gave a better result in both measurements. The first and the fifth regions in the table showed a decrease which is not significant in one of the parameters but an increase in the other. The rest two gave worse result than model 1 in both measurements. As a result only loop modeling at position 311-320 was undertaken over model 1. The resulting DOPE profile is shown in figure 3.16.

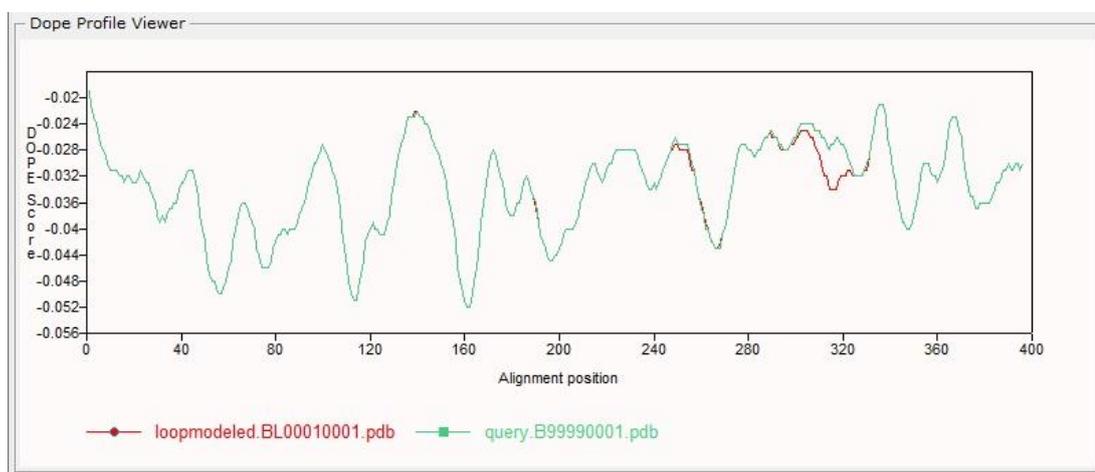


Figure 3.16 DOPE profile of model 1 (turquoise) and the model after loop modeling was undertaken (red).

The energy and DOPE score of the loop model was lower than model 1. TM score was found to be 0.4083 which is near but lower than the result of model 1 (0.4099) and RMSD 5.5950 Å which is near to but higher than the result of model 1 (5.5540 Å). Since the result was not good in topology and position measurements, thermostabilization was performed with both models.

3.2.3.a.4 Optimization, Verification and Validation

Optimization gave a model with a better characteristics. DOPE score and energy was lower. TM score and RMSD values were also better. The optimized model was verified and validated with SAVES. The results are presented below.

The overall quality factor estimated by ERRAT was 89.175. This value is acceptable since the results are out of the rejection limit (%95) in most areas.

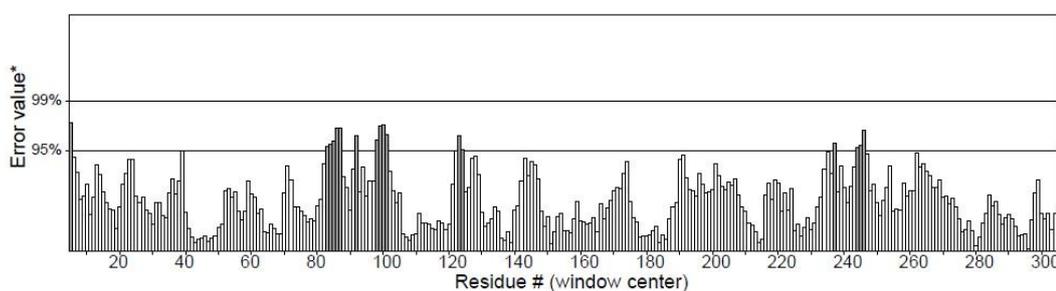


Figure 3.17 ERRAT generated result of the loop model where 95 % indicates rejection limit.

Verify 3D result showed that 96.97% of the residues had an average 3D to 1D score greater than or equal to 0.2. This indicates that the loop model generated is reliable.

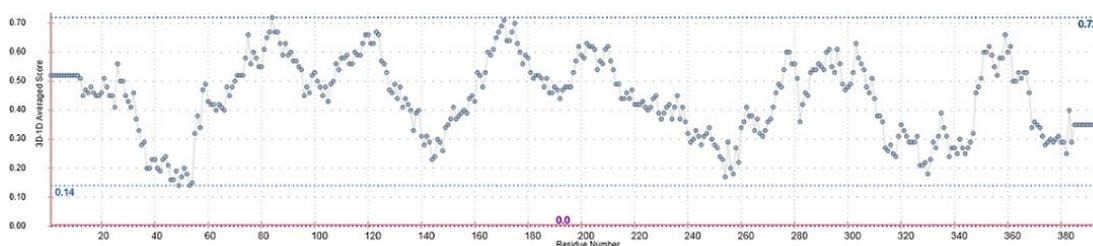


Figure 3.18 Verify 3D result generated for the loop model.

In addition to this most of the results of WHATCHECK were either green or yellow that show the reliability of the model. As a result the over all results showed that the loop model generated is reliable and the model is presented in figure 3.19.

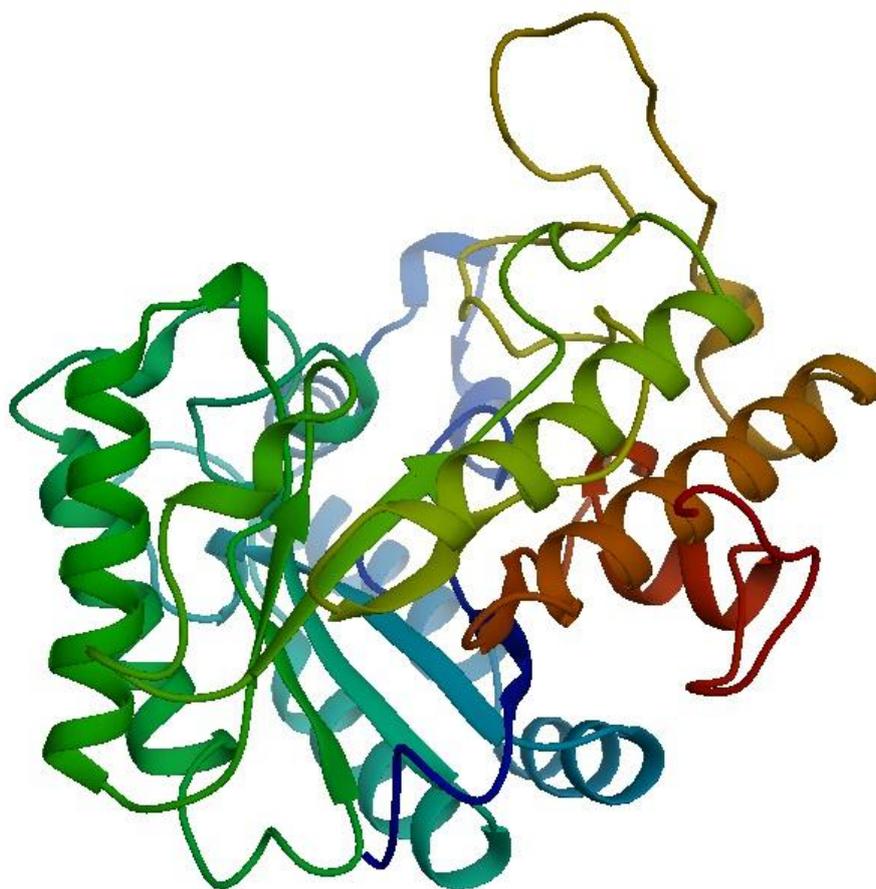


Figure 3.19 3D structure of the best model after loop modeling.

3.2.3.b Homology Modeling by I-TASSER

After the I-TASSER model was generated, its energy, TM score and RMSD values were calculated. The energy was -16787.990 KJ/mol and with energy minimization it was -22267.980 KJ/mol. The average TM score and RMSD value were 0.3480 and 78.0966 Å respectively. The energy was lower than the loop model which is good. But TM score and RMSD values of the loop model were better. The match between MODELLER and I-TASSER model is presented in figure 3.20. I-TASSER model was also evaluated with SAVES. The overall quality factor estimated by ERRAT was 90.979. Verify 3D showed that 96.72% of the residues had an average 3D to 1D score of greater than or equal 0.2. Most of the results of WHATCHECK were either green or yellow that show the reliability of the model. These results show that the generated model is reliable.



Figure 3.20 The match between MODELLER's best model (blue) and I-TASSER model (gray).

3.2.4 Binding Sites

From the results of the COACH, COFACTOR and MetaPocket binding site pocket may include Glu23, Phe25, His129, Asn140, Asn185, Glu186, Tyr248, Phe251, Glu285, Trp361 and Trp371. Binding site predictions by CASTp and DoGSiteScorer also gave a similar result but with a wider coverage. The best binding site of CASTp is presented in figure 3.21. All of the positions which are listed above are part of the best pocket of CASTp. In addition to this *exo*- β -1,3-glucanase has active sites at Glu217 and Glu316 positions as indicated by UniProt. These are Glu186 and Glu285 positions of the protein which was modeled after the first 31 sequences were cleaved by KEX2. These sequences are part of the binding site pocket which was calculated by the computational methods mentioned above.

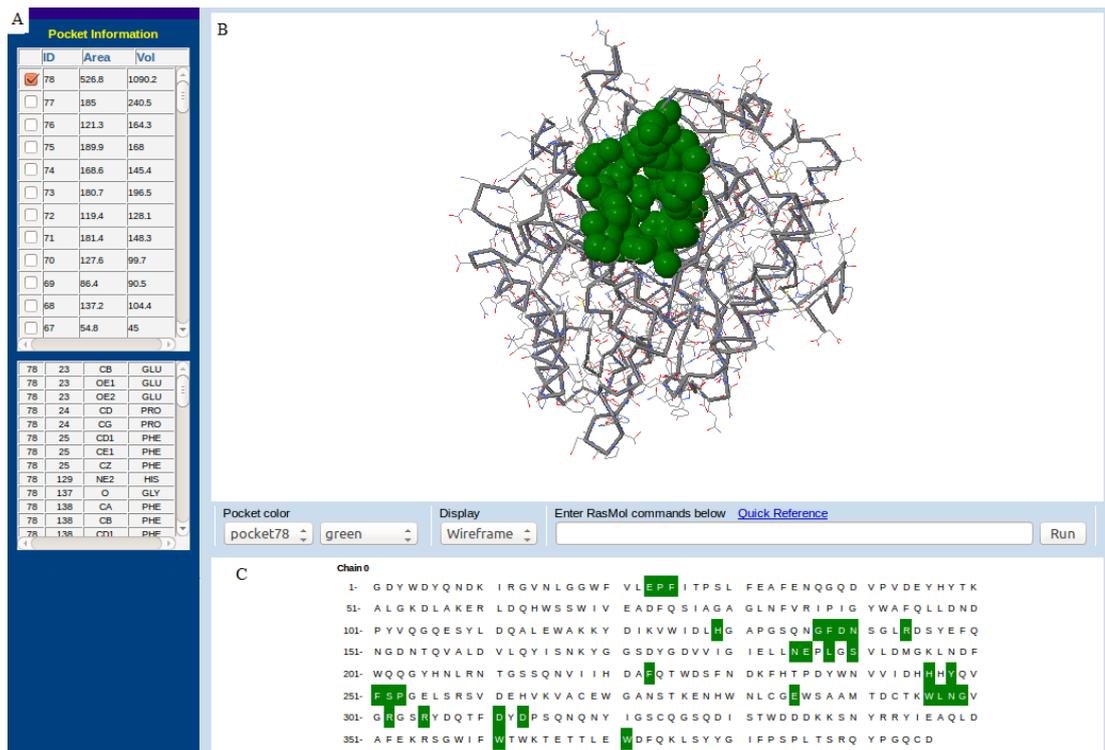


Figure 3.21 A) The table of the area and the volume for different binding sites of the model. B) The three dimensional structure of the best binding site. C) Binding site analysis by CASTp server. Green color illustrates the binding site position.

3.2.5 Thermostabilization of the Models

Thermostabilization was first done with GROMOS96 in SPDBViewer. The results are summarized in table 3.5.

Table 3.5 The best ten stabilizing positions in the allosteric region of the loop model.

Mutation	Leu52Arg	Tyr120Arg	Gly130Ala	Gly195Pro	Gly204Arg
Energy change (KJ/mol)	-580.623	-531.240	-584.131	-541.701	-627.356

Table 3.5 (continued).

Mutation	Phe223Arg	Ile243Arg	Gly254Arg	Cys293Arg	Gly299Arg
Energy change (KJ/mol)	-549.100	-469.840	-799.483	-510.106	-553.352

The total energy with site directed mutations was measured as -20993.869 KJ/mol. This means -1129.170 KJ/mol energy minimization was achieved with site directed mutation at 10 positions.

Then combinations were tested among these best stabilizing positions. Combinations that gave a higher energy are listed in table 3.6.

Table 3.6 Combinations that gave a higher energy.

Nonmutated position	52	120	204	223	254
Energy (KJ/mol)	-20743.613	-20431.449	-20803.203	-20791.016	-20698.871

Table 3.6 (continued).

Nonmutated positions	254,293	254,299	195,223	223,299	223,293
Energy (KJ/mol)	-20719.742	-20682.580	-20775.824	-20767.078	-20812.621

Table 3.6 (continued).

Nonmutated positions	130,204	52,130	52,243	53,130	130,223
Energy (KJ/mol)	-20759.754	-20700.242	-20814.043	-20211.770	-20746.279

As it is shown in table 3.6 positions 130, 243, 293 and 299 in combination with others gave a higher energy but that was lower than the values gained by changing only the positions that gave a higher energy solely. As a result 52, 120, 204, 223 and 254 positions can be mutated if the other methods give a similar result. By doing this the total energy was -21047.369 KJ/mol. This means -1182.670 KJ/mol energy minimization was achieved with site directed mutations at 5 positions. The result is better than what was gained by mutating the ten best positions.

In addition to this the best three stabilizing positions in the binding site were estimated. Phe25Arg, Glu186Arg and Trp371Arg were found to be the best three stabilizers in order with -322.826 KJ/mol, -285.887 KJ/mol and -244.387 KJ/mol changes respectively.

These changes were estimated in terms of K by CNA (figure 3.22).

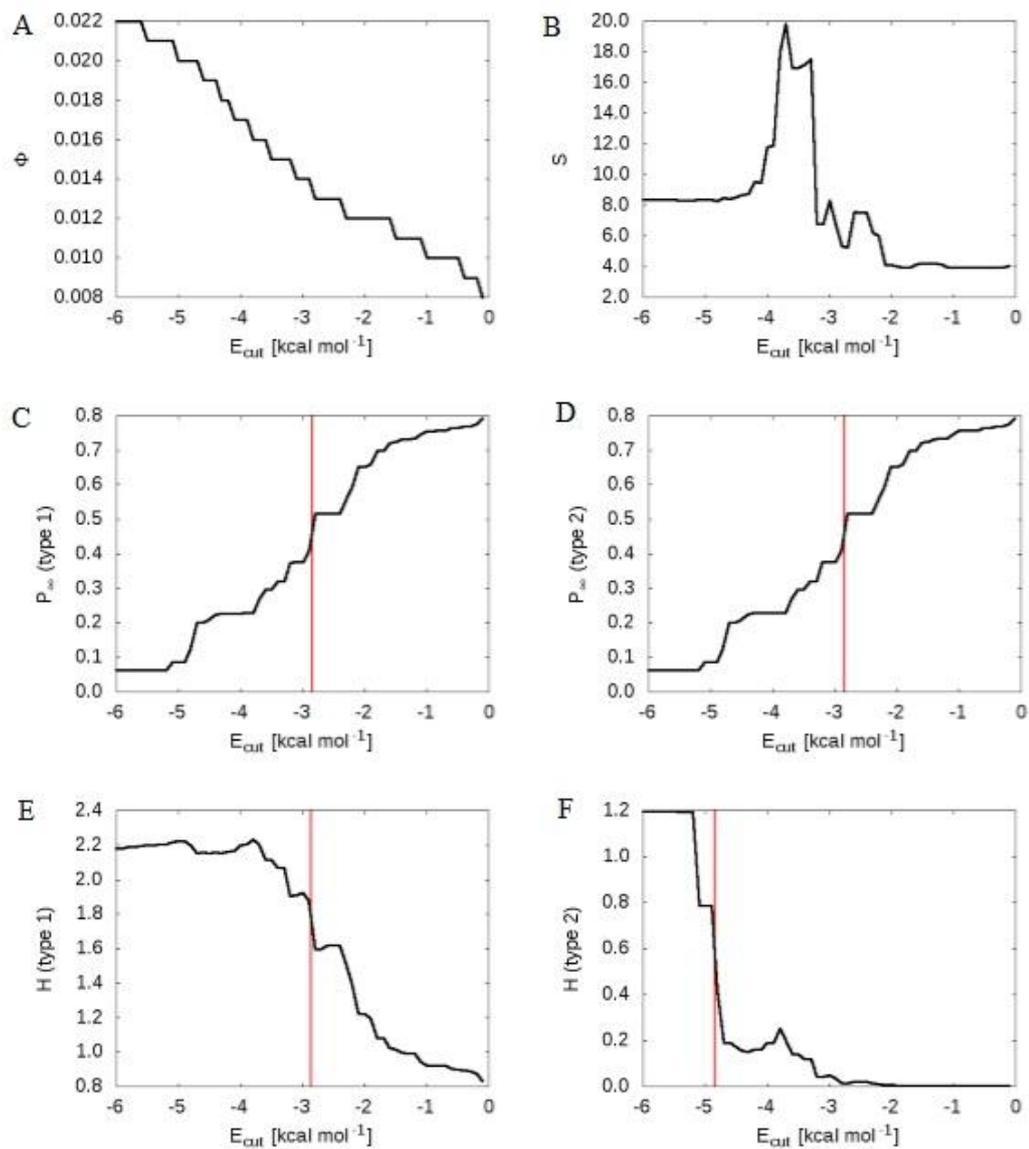


Figure 3.22 Global indices for the thermal unfolding of the loop model: (A) floppy mode density F ; (B) mean rigid cluster size S ; rigidity order parameter P_1 (C) type 1 and (D) type 2; cluster configuration entropy H (E) type 1 and (F) type 2. The red vertical lines (C–F) indicates the phase transitions.

The cluster configuration entropy type 2 (H type 2) which have been found to be related with melting point of the protein was -4.84 kcal/mol equivalent to 396.74 K without any mutation.

Index	$E_{cut}[\text{kcal mol}^{-1}]$	T[K]
Rigidity order parameter type 1 c)	-2.85	356.96
Rigidity order parameter type 2 d)	-2.85	356.96
Cluster configuration entropy type 1 e)	-2.86	357.13
Cluster configuration entropy type 2 f)	-4.84	396.74

Figure 3.23 CNA result of the loop model.

Values after those positions mutated and estimated were compared with the nonmutated result. As an example Leu52Arg mutation gave the following result.

Index	$E_{cut}[\text{kcal mol}^{-1}]$	T[K]
Rigidity order parameter type 1 c)	-2.86	357.13
Rigidity order parameter type 2 d)	-2.86	357.13
Cluster configuration entropy type 1 e)	not detected	not detected
Cluster configuration entropy type 2 f)	-5.15	403.00

Figure 3.24 CNA result after Leu52Arg mutation was done.

H type 2 was found to be -5.15 kcal/mol equivalent to 403.00 K. This means there is an increase of 6.26 K by Leu52Arg mutation. Similarly 0, -0.06 K, 6.26 K and 8.27 K changes were achieved with mutations Tyr120Arg, Gly204Arg, Phe223Arg and Gly254Arg respectively. From this we can say that mutations at 52, 223 and 254 positions gave consistent result with the above method.

I-Mutant2.0 gave a result that shows most of the mutations for the best ten stabilizers are stabilizing. Eris results showed that most of them are stabilizers. AUTO-MUTE also gave a similar result. According to MUpro results all the ten positions are stabilizers. Results were obtained for each position separately. To illustrate this MUpro result for Leu52Arg is presented in figure 3.25.

From the three best stabilizing positions in the active site, Glu186Arg was found to be stabilizer by most methods used. 9.58 K increase can be achieved with this mutation as CNA estimated it.

Structure stability prediction for mutation:

Mutation Request:

Name: Panomycocin

Sequence:

```
GDYWDYQNDKIRGVNLLGGWVLEPFITPSLFEAFENQGQDVPVDEYHYTKALGKDLAKER
LDQHSSWIVEADFQSIAGAGLNFVRIPIGYWAFQLLDNDPYVQGQESYLDQALEWAKKY
DIKVVIDLHGAPGSQNGFDNSGLRDSYEFQNGDNTQVALDVLQYISNKYGGSDYGDVVIG
IELLNEPLGSLVDMGKLNDFWQQGYHNLNRTGSSQNVIIHDAFQTWDSFNDKFHTPDYWN
VVIDHHYQVFSPELRSVDEHVKVACEWGANSTKENHWNLCGEWSAAMDCTKWLNGV
GRGSRDQTFDYDPSQNYIGSCQGSQDISTWDDKKSNYRRYIEAQLDAFEKRSGWIF
WTWKTETTLWDFQKLSYYGIFPSPLTSRQYPGQCD
```

Position: 52

Original Amino Acid: L

Substitute Amino Acid: R

Prediction Results:

1. Predicted both value and sign of energy change using SVM and sequence information only

deta delta G = (INCREASE stability)

Figure 3.25 MUpro result of Leu52Arg mutation.

These results were obtained for I-TASSER model too. The results are summarized below.

Table 3.7 The best ten stabilizing positions in the I-TASSER model.

Mutation	Gly38Arg	Gly53Arg	Gly81Arg	Gly105Arg	Gly137Arg
Energy change (KJ/mol)	-377.844	-365.002	-375.975	-355.661	-394.256

Table 3.7 (continued).

Mutation	Gly175Arg	Gly212Arg	Gly301Arg	Gly322Arg	Gly326Arg
Energy change (KJ/mol)	-364.934	-342.795	-344.219	-354.147	-354.383

The total energy with site directed mutations was measured as -25396.039 KJ/mol. This means -3128.059 KJ/mol energy minimization was achieved with site directed mutations at 10 positions.

In addition to this combinations were tested among these best stabilizing positions. All combinations among these positions gave higher than -25000 KJ/mol. This shows us that all of the best ten positions can be mutated if the other methods give a consistent result.

However, only Gly53Arg gave 29.46 K increase according to CNA results. The other positions did not give a consistent result.

CHAPTER 4

DISCUSSION

Increased use of wide spectrum antibiotics, secondary resistance development towards conventional drugs, immunosuppressive infections or diseases, the use of chemotherapeutic agents, long term use of corticoids, invasive medical interventions and high level of immunosuppressed population due to the survival of patients with severe illnesses resulting from advanced medical technology increased the risk of fungal infections [3-5]. These changes has increased the rate of invasive fungal infections substantially as WHO reports show. As a result the demand for antifungal agents has risen and the global market for these agents is expected to reach nearly to \$13.9 billion in 2018 [7].

Many of the currently used antifungal agents target the cell membrane or intracellular molecules and processes like the mitotic division of fungi and some other new agents act on the cell wall. Fungi cells share many of the biochemical pathways and subcellular structures with mammalian cells because they are eukaryotic cells. This means the antifungal agents affect the host cells too if their targets are these common pathways and structures. Therefore, many of the currently used antifungal agents have many side effects and there is high resistance development. Furthermore they interact with other therapeutics unfavourably, have a narrow spectrum of activity, have limited formulation and are fungistatic. Thus antifungal agents that act on fungal cell wall compounds are superior to other antifungal drugs since they are selective to the fungi, the cross-resistance with the other antifungal agents does not develop and have wide spectrum fungicidal activity. As a result antifungal agents that target the cell wall are a promising, attractive novel antifungals [11,12,16].

Yeast killer proteins are naturally occurring substances secreted by killer yeast strains that are lethal to sensitive yeast cells. They have various modes of action but the common mechanisms of action are hydrolyzing or inhibiting the synthesis of β -1,3-glucans which are the major cell wall components and forming ion channels on cell membrane leading to ion leakage. Their way of action and sources show that they are good candidates as novel and potent antifungal agents. K5 type yeast killer protein, Panomycocin, is an example [27].

Panomycocin is a K5 type yeast killer protein which is produced by *Pichia anomala* NCYC 434 strain. Panomycocin is a glycosylated monomer protein. Its molecular mass is 49 kDa and it has a pI value of 3.7. Panomycocin has a wide range pH (2.5-5.5) stability and has high affinity to β -1,3- glucans which is the component of fungal cell wall. As a result Panomycocin has a high potential to be used as a therapeutic antifungal agent in the medications of fungal infections due to its novel mechanism of action, selectivity and stability. However, its activity decreases above 37 °C. Thermostability of Panomycocin should be increased so that formulations that are stable at high temperature can be prepared. Various excipients and computational methods were used to thermostabilize the protein [27,33].

Excipients from various groups which have various mechanisms of interactions are used to increase thermostability of proteins. However, depending on the concentration of excipients and the working environment, excipients may have different effects on the thermostability. Therefore, gradient concentrations of the excipients with increasing temperatures were tested in this work. Buffering agents are used to keep the pH at the point in which the protein is highly active and stable. Since the optimum pH for Panomycocin is 4.5, the test medium petri pH was made 4.5 with citrate phosphate buffer.

Thermostabilization of proteins using amino acids is achieved through preferential hydration and direct binding. They are also buffering agents and have antioxidant properties that are important in stabilizing proteins [42]. Stabilizing effects of amino acids are based on the interactions of amino acids with proteins. These interactions are

highly influenced by the nature of amino acids and proteins, concentrations and the environment. Arginine is one of the amino acids which are used to stabilize proteins. It interacts with proteins in various ways. It may interact through hydrogen bonding, electrostatic interactions, direct interactions with the backbone structure, association with the hydrophobic patches of the protein and formation of cation- π interactions with aromatic amino acids of the protein [41]. However, these interactions might differ based on concentrations as it was observed for arginine-BSA (Bovine Serum Albumin) interactions [41]. Arginine's effect on thermostability was found to vary depending on the protein. For example, Arakawa *et al.* have seen arginine has no effects on the melting point of lysozyme and RNase but Thakkar *et al.* seen varying effects of it on monoclonal antibodies [42,158]. Glycine is another amino acid which is used in the thermostabilization of proteins. It has various interactions with various proteins depending on the concentration and type of protein. The interactions of glycine are mainly with charged side chains at low concentrations, with peptide backbone at intermediate concentrations and with competition for water. Platts *et al.* observed that glycine stabilizes BSA, myoglobin and lysozyme proteins at concentrations above 0.1M but destabilizes myoglobin at lower concentrations and it has varying effects on the others. This shows how the effect of glycine is protein specific and unpredictable [159]. In our work none of the amino acids used to increase the thermostability of Panomycocin gave a positive result. This might be the consequence of the complicated interactions and thus effects of amino acids on proteins.

The stabilizing effects of carbohydrates and sugars result from the interactions between these excipients and proteins. The interactions are mainly hydrogen bonding, dispersive interactions, preferential hydration and aromatic interactions such as hydrophobic and CH- π interactions. Special conformational interactions with proteins and formation of viscous matrices which are important in the improvement of thermostability are the effects of these interactions. A study by Zhong *et al.* showed that sucrose, glucose and lactose can improve the thermostability of whey protein-maltodextrin conjugates by increasing the viscosity of continuous phase and stability of conjugates [160]. However, they have also destabilizing effects depending on the condition and the nature of excipients and proteins. Some of them can degrade or may contain impurities that destabilize proteins. For example, sucrose may hydrolyze to

glucose and fructose. This results in the formation of a reducing sugar that may cause protein glycation at lysine residues [161]. The carbohydrates and sugars tested did not increase the thermostability of Panomycocin perhaps due to these complicated interaction mechanisms and effects on it.

Osmolytes act similar to carbohydrates and sugars since they have similar nature. However, polyols which are representatives for this group including glycerol decrease the surface tension of water and therefore their stabilizing effect is mainly solvophobic in nature [43]. Depending on temperature, pH, concentration and the nature of the protein the thermostabilizing effects of osmolytes vary. For example, although trehalose and sorbitol lead to preferential hydration of RNase A at low temperatures, they have been found to bind weakly to RNase A at high temperatures [162]. As a result they may not show their thermostabilizing effect at high temperatures like it was observed in our work.

Surfactants' major ways of interactions with proteins are electrostatic and dispersive interactions. These are essentially used to prevent aggregation due to agitation or shaking. In addition to this Tween 80 and Pluronic F-68 has been found to increase the thermostability of toxoid A and toxoid B. These excipients too act variably. Wang *et al.* have observed that Tween 80 has a dual effect on the stability of IL-2 (Interleucin-2). Tween 80 has been found to inhibit the aggregation of IL-2 due to shaking but adversely affected the stability of IL-2 in the solution storage form. These effects depend highly on temperature and formulation [163]. So, in a similar way these excipients did not thermostabilize the protein in our work.

All of the excipients tested did not thermostabilize Panomycocin over 37 °C. Thus computational methods were pursued. Computational methods have the additional advantage of stabilizing the protein through its synthesis process resulting in higher production and purification yields. Prior to thermostabilization, modeling of the protein was undertaken. MODELLER and I-TASSER (Iterative Threading ASSEmbly Refinement) server were used in homology modeling of the protein.

MODELLER is a popular and widely used homology modeling tool. MODELLER has many advantages over the other programs used in homology modeling. It is freely

available, has many powerful features and gives reliable results. It has freely available graphic user interface called EasyModeller. EasyModeller enables users to build a model easily even if they did not have experience in modeling because it is simple and straightforward. Moreover, assessment, visualization and optimization of the generated model is possible. EasyModeller has many features that makes it user friendly and reliable. These features are: 1) tab based logical modeling steps which are easy to implement; 2) provides the opportunity of loading unlimited number of templates; 3) enables users to generate many models; 4) colorful alignment viewer with alignment editor; 5) MODELLER code editing; 6) inbuilt DOPE profile viewer, Ramachandran plot viewer, loop modeling, model optimization and dynamics for a selected model [83].

I-TASSER server is an internet based service for protein structure predictions. It allows academic users to generate high quality 3D structure predictions. CASP (Critical Assessment of Structure Prediction) experiments have been developed to get an objective assessment of the state of modeling tools. The competition takes place every two years since 1994. I-TASSER server participated in the competition since 2006 and was found to be one of the best modeling tools in the servers section of the CASP experiments. In addition to this I-TASSER has many good features that makes it superior to other modeling servers. I-TASSER: 1) enables users to select and align templates; 2) can also predict the function of the protein; 3) can give upto five models; 4) gives a confidence score, TM score and RMSD value for the first model; 5) the output includes GIF images of the predicted models and top 10 proteins in PDB that have similar structures to the predicted models [84].

Using the internal and N-terminal sequences made, protein similarity search was performed with FASTA program. Exo- β -1,3-glucanase from *Pichia anomala* strain K with accession number AJ222862 was found to be the best homologue of the sequences as it was also indicated by Izgü *et al.* before [27]. The sequence of the homologue protein was obtained from UniProt [143] with its accession number. The sequence found is gene sequencing which was done by Jigakli *et al.* [164] and thus it contains the signal peptide and amino acids that are cleaved by KEX2 before the secretion of the protein. So, the sequences which are not part of the mature protein

should be determined and cleaved before beginning homology modeling. The signal peptide was determined by SignalP 4.1, Signal-CF, PrediSi and Signal-3L. It was found to include the first seventeen amino acids. Before secretion of the protein, a few amino acids are cleaved by KEX2. To determine the KEX2 cleavage site SMART (Simple Modular Architecture Research Tool) was used. SMART estimated the KEX2 cleavage site to be between 30-32 positions (-Lys-Arg-Gly-). Since N-terminal sequencing of the protein yielded GDYWDYQNDKIR [27], the secreted mature protein starts with Gly. So, the KEX2 cleavage takes place after the 31. position at the -Arg-Gly- bond. Then the sequence upstream of the KEX2 cleavage site was cleaved and modeling was performed with the downstream sequence.

First templates that would be used in the homology modeling of the protein were found. PSI-BLAST (position specific iteration BLAST) algorithm of BLAST (Basic Local Alignment Search Tool) in the NCBI (National Center for Biotechnology Information) was used to find the templates. The template search showed that there are homologue proteins to the protein with %100 identity and coverage but their 3D structure has not been determined yet. As a result those proteins whose structures were determined and put into PDB were selected as templates. These are α -1,3-glucanases from *Candida albicans* with PDB IDs 1EQP, 2PB1 and 1CZ1 whose structures were determined by X-ray diffraction. They have %66 identity and %98 coverage in relative to the protein. According to Rost *et al* for the protein being modelled which consists of 396 amino acids downstream of the KEX2 cleavage site, above %30 identity is in the safe homology modeling zone [52]. So, %66 identity is a high identity level and thus the query and templates are expected to fold in to the same structure.

MODELLER licence was obtained. Then MODELLER 9.11 was installed. As all the necessary inputs and programs for homology modeling by MODELLER were ready, modeling began. First the query sequence and the templates were loaded in EasyModeller 4.0, the graphic user interface for MODELLER. The query sequence was checked whether there is a gap in it or not. Since there was not any gap, the next steps were pursued. Second the templates were aligned and it was observed that their sequences are almost the same. Then the query sequence was aligned to the templates.

The predicted secondary structure was investigated but since there was not a wide gap no need of MODELLER code editing. At last nine models were generated together with DOPE profiles under the DOPE profile viewer.

Among the generated models, the best model was chosen by comparing them with their DOPE score, energy, TM (Template Modeling) score and RMSD (Root Mean Square Deviation) values. DOPE score was obtained from MODELLER; energy was calculated with GROMOS96 (GRONingen MOlecular Simulation96) in the SPDBViewer (Swiss PDB Viewer); TM score was calculated with TM score calculator; RMSD was calculated with VMD (Visual Molecular Dynamics). The results of these calculations showed that model 1 gave the best result in energy and second in DOPE score. In addition to this it gave a good result in measurements related with topology (TM score) and position (RMSD) even if it was not the best. In short the results showed model 1 is good in all the measurements used. So, it was chosen as the best model and thus the next stages were performed with it.

Since MODELLER gives the opportunity of loop modeling, loop modeling was performed over the best model (model 1). In order to estimate the regions that loop modeling might give a better model, analysis of the alignment between the query and the templates, observation of the 3D structure of the model and analysis of the DOPE profile of model 1 with respect to templates were used. Then loop modeling was undertaken at various positions and its effects were analyzed by comparing the DOPE score, energy, TM score and RMSD values of the resulting models to the best model. Only loop modeling at positions 311-320 gave a better result in DOPE score and energy. However, the TM score and RMSD values of the loop model were worse than the best model. Therefore, computational thermostabilization was performed with both models.

The best model and loop model were optimized by MODELLER to get models with better properties. The optimized models were verified and validated with SAVES (Structure Analysis and Verification Server). The overall quality factor estimated by ERRAT was 89.175. Colovos *et al* put that a model should be out of the rejection limit (%95) at least with 80 overall quality factor [165]. Therefore, the result obtained

is in the acceptable region. Verify 3D result showed that 96.97% of the residues had an average 3D to 1D score of greater than or equal to 0.2. According to Eisenberg *et al.* this value is a measurement of the compatibility of a protein model with its sequence. It can have values from -100% (bad value) to $+100\%$ (good value) [166]. So, our model is good according to this criteria. In addition to this most of the results of WHATCHECK were either green or yellow that show the reliability of the model. As a result the overall results showed that the model generated is reliable.

A model was also generated by I-TASSER and compared with the model which was generated by MODELLER. Although I-TASSER model was found to be better in energy, it was worse than the MODELLER model in TM score and RMSD values. I-TASSER model was also evaluated with SAVES. The overall quality factor estimated by ERRAT was 90.979. Verify 3D showed that 96.72% of the residues had an average 3D to 1D score of greater than or equal to 0.2. Most of the results of WHATCHECK were either green or yellow that show the reliability of the model. These results show that the generated model is reliable.

The roles of binding sites and allosteric regions in the activity of proteins are different. Binding site plays a crucial role in the binding of a protein to its substrate. Therefore, binding site was determined and its thermostabilization was made separately as it needs special treatment. COACH, COFACTOR, MetaPocket, CASTp (Computed Atlas of Surface Topography of Proteins) and DoGSiteScorer were used to predict the binding site of the models. From the predictions made by these methods binding site pocket may include Glu23, Phe25, His129, Asn140, Asn185, Glu186, Tyr248, Phe251, Glu285, Trp361 and Trp371. Moreover, UniProt indicates exo- β -1,3-glucanase has active sites at Glu217 and Glu316 positions that are Glu186 and Glu285 positions of the protein which was modeled after the first 31 sequences were cleaved by KEX2. These are part of the binding site pocket which was estimated by the computational methods.

Thermostabilization of the models was first done with GROMOS96 in the SPDBViewer. Energy of each residue in the models was predicted and energy profiles of the alignment were investigated. The energy calculated here is the sum of the

energies of angles, bonds, torsions, improper interactions, nonbondeds, electrostatics and constraints. All of the amino acids in the unstable regions were substituted with the rest nineteen amino acids using SPDBViewer and the energies were calculated. Those residues which gave lower energy than the nonmutated ones were recorded. Among them the best ten were chosen. Leu52Arg, Tyr120Arg, Gly130Ala, Gly195Pro, Gly204Arg, Phe223Arg, Ile243Arg, Gly254Arg, Cys293Arg, Glyc299Arg were found to be the best ten stabilizing mutations in the allosteric region of the protein. Then the combinations among the best ten stabilizing positions were tested. The interactions between the newly mutated amino acids can be destabilizing since the properties of the amino acids differ from each other. Those that are destabilizing gave a higher energy. Leu52Arg, Tyr120Arg, Gly204Arg, Phe223Arg and Gly254Arg were found to be stabilizing in the combinations too and thus these positions can be mutated to get a thermostable protein if the other methods give a similar result. The total energy decrease achieved by mutating only these five positions was better than that of the ten best positions. The energy calculations made by GROMOS96 in SPDBViewer are in KJ/mol. Therefore, calculations were made in terms of temperature (K) by CNA (Constraint Network Analysis) server. CNA results together with various servers that are used to estimate the effect of mutations on a protein such as I-MUTANT2.0, AUTO-MUTE, Eris and MUpro were used to increase the accuracy of the predictions made. CNA predictions showed that Leu52Arg, Phe223Arg and Gly254Arg mutations can stabilize the protein with 6.26K, 8.27K and 6.26K temperature increase respectively. Moreover, the other methods used also predicted these positions as thermostabilizing. So, substitution of 52, 223 or 254 positions with arginine is expected to increase the thermostability of Panomycocin to the desired level.

The thermostability problem may result from the deformation and thus deactivation of the binding site of the protein. In case Panomycocin has a similar problem, thermostabilizing positions in the binding site were also predicted and Phe25Arg, Glu186Arg and Trp371Arg were found to be the best three thermostabilizers. From the three best thermostabilizing positions in the binding site, Glu186Arg was found to be thermostabilizer by the other methods too. 9.58 K temperature increase can be achieved with this point mutation as estimated by CNA.

The aim of the thermostabilization work is to design a thermostable protein without changing its activity. The 3D structure and binding site of the protein did not change after the best stabilizing positions mutated. So, the mutant strain protein is expected to bind to the same substrate and show the same activity with the wild strain protein.

All of the above mentioned calculations were performed on the I-TASSER model one by one. All of the best ten positions were substitutions of glycine by arginine. In the combination tests, all the ten positions were found to be thermostabilizers in the combined form too. However, only a position gave a consistent result with the other methods used.

In our computational thermostabilization work, most of the thermostabilizing mutations were substitutions of a residue by arginine. Since several studies before showed that the change of hydrophobicity to hydrophilicity of residues in solvent exposed surface of proteins is a good strategy of stabilizing proteins, substituting a hydrophobic residue by arginine which is a positively charged hydrophilic amino acid is expected to increase the stability of proteins in general and thermostability in particular. Strub *et al.* increased the stability of acetylcholinesterase by substituting solvent exposed hydrophobic residues by arginine [167]. Even substitutions of hydrophilic residues on the surface by arginine may increase the thermostability of proteins. Sokalingam *et al.* increased the stability of green fluorescent protein (GFP) by substituting solvent exposed surface lysines by arginines. Although lysine and arginine are both positively charged basic amino acids, the guanidinium group of arginine permits interactions in three directions that enables it to form a higher number electrostatic interactions and its basic residue has higher pKa that may generate more stable ionic interactions [168]. Mortazavi *et al.* enhanced the thermostability of firefly luciferases by substituting solvent exposed hydrophobic residues by arginine [169]. Zhou *et al.* increased the thermostability of xylanase II from *Aspergillus usamii* E001 by replacing serines and threonines on the solvent exposed surface of the enzyme with arginines [170]. Moreover, bioinformatics analysis showed that one of the most striking features of thermostable proteins is the higher proportion of arginine in the exposed surfaces [169]. For example, Kumwenda *et al.* observed high frequency of

arginine (on the surfaces) and alanine (in well buried areas) in thermostable proteins of *Thermus thermophilus* HB27 [171]. The three thermostabilizing positions detected in the allosteric region of the protein are all substitutions of amino acids with arginine. As it is depicted in figure 4.1 all the three positions are on water exposed surface of the protein. Moreover, leucine and phenylalanine are nonpolar that increases the hydrophobicity of the residues being mutated and glycine is a flexible amino acid in which its substitution may contribute to the thermostabilization of proteins. So, the result obtained is in accordance with the computational and practical researches conducted before.

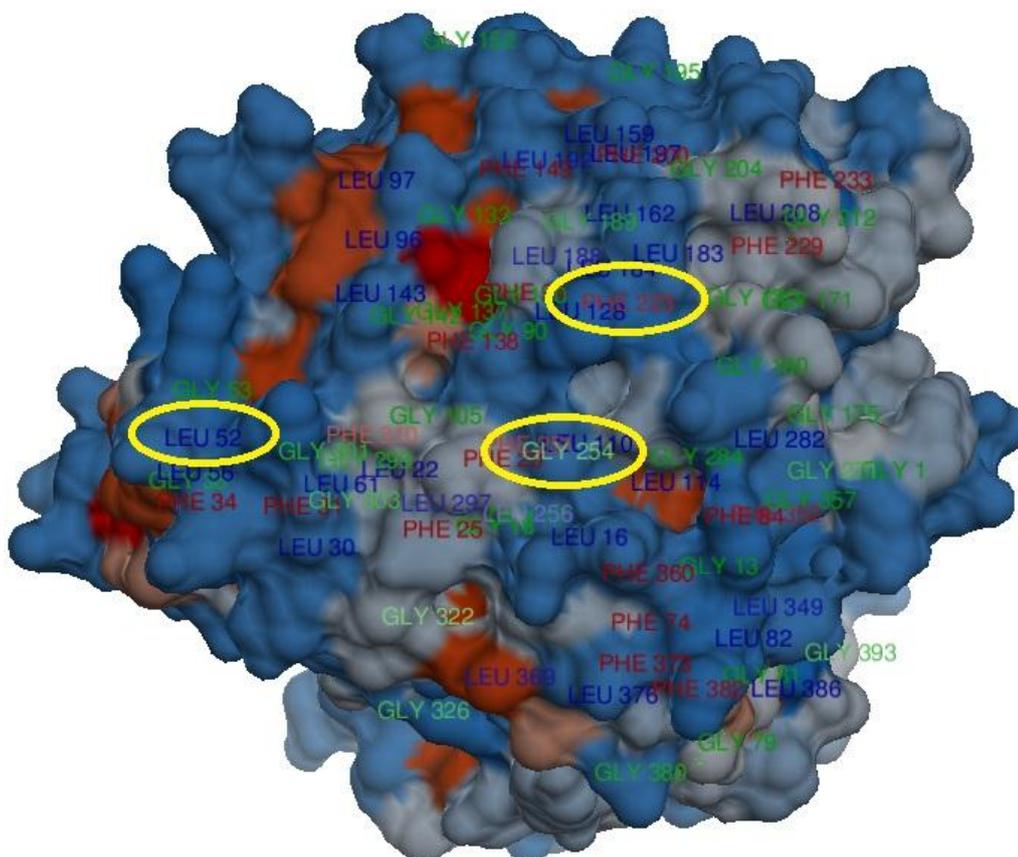


Figure 4.1 Surface of the model generated. The blue regions are expected to be hydrophobic.

CHAPTER 5

CONCLUSION

- 1) Various excipients were tested on *S. cerevisiae* NCYC 1006 at increasing temperatures with gradient of concentrations but none of the excipients increased the thermostability of Panomycocin.
- 2) The signal peptide of the protein was determined by SignalP 4.1, Signal-CF, PrediSi and Signal-3L and it was found to include the first 17 amino acids.
- 3) The KEX2 cleavage site was determined using SMART and the mature secreted protein sequencing made before and was found to be after the 31. amino acid at the -Arg-Gly- .
- 4) Binding site was predicted by COACH, COFACTOR, MetaPocket, CASTp and DoGSiteScorer. Binding site may include Glu23, Phe25, His129, Asn140, Asn185, Glu186, Tyr248, Phe251, Glu285, Trp361 and Trp371.
- 5) Thermostabilization of the models was done by GROMOS96, CNA, I-MUTANT2.0, AUTO-MUTE, Eris and MUpro. In the allosteric part of the protein Leu52Arg, Phe223Arg and Gly254Arg were found to be the best thermostabilizing mutations with 6.26 K, 6.26 K and 8.27 K temperature increases respectively. In the binding site Glu186Arg was found to be the best thermostabilizer mutation with 9.58 K temperature increase.

- 6) In this study single site mutations that can increase the thermostability of Panomycocin to the desired level were detected computationally. In future studies these thermostable mutant strains can be produced and isolated. This mutant strains can be used in the formulation of Panomycocin as a novel antifungal drug which is stable at high temperatures.

REFERENCES

- [1] Nucci M., Queiroz-Telles F., Alvarado-Matute T., Tiraboschi I.N., Cortes J., Zurita J., Guzman-Blanco M., Santolaya M.E., Thompson L., Sifuentes-Osornio J., (2013). Epidemiology of candidemia in Latin America: A laboratory-based survey. *PLoS One*, 8, e59373.
- [2] Arendrup M.C., Dzajic E., Jensen R.H., Johansen H.K., Kjældgaard P., Knudsen J.D., Kristensen L., Leitz C., Lemming L.E., Nielsen L. (2013). Epidemiological changes with potential implication for antifungal prescription recommendations for fungaemia: Data from a nationwide fungaemia surveillance programme. *Clin. Microbiol. Infect.*, 19, 343–353.
- [3] Venkatesan P., Perfect J. R., Myers, S. A. (2005). Evaluation and management of Fungal infections in Immunocompromised patients. *Dermatologic Therapy*, 18, 44–57.
- [4] Vicente M. F., Basilio A., Cabello A., Peláez F. (2003). Microbial natural products as a source of antifungals, *Clin. Microbiol. Infect.*, 9(1), 15-32.
- [5] Di Santo R. (2010). Natural products as antifungal agents against clinically relevant pathogens. *Nat. Prod. Rep.*, 27(7), 1084-1098.
- [6] Butts A., Krysan, D.J. (2012). Antifungal drug discovery: Something old and something new. *PLoS Pathog.*, 8, e1002870.

- [7] Corpano B., Bombourg N., Goncalves J., Cadi K. (2014). Summary of BCC Research report about antifungal agents. <http://www.reportlinker.com/p02212016-summary/view-report.html> last accessed on 12/10/2015.
- [8] Odds F.C. (2003). Antifungal agents: Their diversity and increasing sophistication. *Mycologist*, 2, 17, 51–55.
- [9] Heeres J., Meerpoel L., Lewi, P. (2010). Conazoles. *Molecules*, 15, 4129–4188.
- [10] Cappelletty D.; Eiselstein-McKittrick K. (2007). The echinocandins. *Pharmacotherapy*, 27, 369–388.
- [11] Carrillo-Muñoz A. J., Giusiano G., Ezkurra P. A., Quindós, G. (2006). Antifungal agents: Mode of action in yeast cells. *Rev. Esp. Quimioterap.*, 19(2), 130-139.
- [12] Kanafani Z. A., Perfect J. R. (2008). Antimicrobial resistance: resistance to antifungal agents: mechanisms and clinical impact. *Clin. Infect. Dis.*, 46(1), 120-128.
- [13] Lewis R. E. (2011). Current Concepts in Antifungal Pharmacology. *Mayo Clin. Proc.*, 86(8), 805-817.
- [14] Loeffler J., Stevens D. A. (2003). Antifungal Drug Resistance. *Clinical Infectious Diseases*, 36(1), 31–41.
- [15] Kontoyiannis D. P., Lewis R. E. (2002). Antifungal drug resistance of pathogenic fungi. *Lancet*, 359, 1135–1144.
- [16] Anderson J. B. (2005). Evolution of antifungal drug resistance: mechanisms and pathogen fitness. *Nature Reviews, Microbiology*, 3, 547-556.

- [17] Ashley E. S. D., Lewis R., Lewis J. S., Martin C., Andes D. (2006). Pharmacology of Systemic Antifungal Agents. *Clinical Infectious Disease*, 43, 28–39.
- [18] Jucker E. (2004). Antifungal Agents: Advances and Problems. Switzerland: Birkhauser Verlag.
- [19] Sangamwar A. T., Deshpande U. D., Pekamwar S. S. (2008). Antifungals: Need to search for a new molecular target. *Indian J. Pharm. Sci.*, 70, 423-430.
- [20] Canuto M. M., Rodero F. G. (2002). Antifungal drug resistance to azoles and polyenes. *Lancet Infect. Dis.*, 2, 550-563.
- [21] Meinhardt F., Klassen R. (2009). Yeast Killer Toxins: Fundamentals and Applications, Physiology and Genetics. In Anke, T., Weber, D. (Ed.). *The Mycota*, 15, (107-130). Berlin Heidelberg: Springer Verlag.
- [22] Marquina D., Santos A., Peinado J. M. (2002). Biology of killer yeasts. *Int. Microbiol.*, 5, 65–71.
- [23] Schmitt M. J., Breinig F. (2002). The viral killer system in yeast: from molecular biology to application. *FEMS Microbiology Reviews*, 26, 257-276.
- [24] Magliani W., Conti S., Travassos L. R., Polonelli L. (2008). From yeast killer toxins to antibiobodies and beyond. *FEMS Microbiol Lett*, 288, 1–8.
- [25] Young T.W., Yagiu M. (1978). A comparison of the killer character in different yeasts and its classification. *Antonie Van Leeuwenhoek J. Microbiol. Serol.* , 44, 59-77.
- [26] Wickner R. B. (1986). Double-stranded RNA replication in yeast: the killer system. *Annu. Rev. Biochem.*, 55, 373–395.

- [27] Izgü F., Altınbay D., Acun T. (2006). Killer toxin of *Pichia anomala* NCYC 432; purification, characterization and its exo- β -1,3-glucanase activity. *Enzyme Microb. Technol.*, 39, 669–676.
- [28] Goffeau A. *et al.* (1996). Life with 6000 genes. *Science*, 274, 546-567.
- [29] Sommer S.S., Wickner R.B. (1982). Yeast L dsRNA consists of at least three distinct RNAs; evidence that the non-Mendelian genes [HOK], [NEX] and [EXL] are on one of these dsRNAs. *Cell*, 31, 429- 441.
- [30] Bussey H. D., Saville D., Greene D. J., Tipper K. A., Bostian S. (1983). Secretion of yeast killer toxin: processing of the glycosylated precursor. *Mol. Cell. Biol.*, 3, 1362–1370.
- [31] Walker G. M. (2011). *Pichia anomala*: cell physiology and biotechnology relative to other yeasts. *Antonie Van Leeuwenhoek*, 99(1), 25-34.
- [32] Izgü F., Altınbay D., Sertkaya A. (2005). Enzymatic activity of the K5-type yeast killer toxin and its characterization. *Biosci. Biotechnol. Biochem.*, 69(11), 2200-2206.
- [33] Izgü F., Altınbay D. (2004). Isolation and characterization of the K5 type yeast killer protein and its homology with an exo- β -1, 3-Glucanase. *Biosci. Biotech. Biochem.*, 68(3), 685-693.
- [34] Izgü F., Altınbay D., Türeli E. (2007). In vitro activity of Panomycocin, a novel exo- β -1,3-glucanase isolated from *Pichia anomala* NCYC 434, against dermatophytes. *Mycoses*, 50, 31-34.
- [35] Izgü F., Altınbay D., Türeli E. (2007). In vitro susceptibilities of *Candida spp.* To Panomycocin, a novel exo- β -1,3-glucanase isolated from *Pichia anomala* NCYC 434. *Microbiol. Immunol.*, 51(9), 797-803.

- [36] Izgü A. D., Kepekci R. A., İzgü F. (2011). Inhibition of *Penicillium digitatum* and *Penicillium italicum* in vitro and in planta with Panomycocin, a novel exo- β -1,3-glucanase isolated from *Pichia anomala* NCYC 434. *Antonie van Leeuwenhoek*, 99, 85-91.
- [37] Zhou X-X., Wang Y-B., Pan Y-J., Li W-F. (2008). Differences in amino acids composition and coupling patterns between mesophilic and thermophilic proteins. *Amino Acids*, 34, 25–33.
- [38] Talluri S. (2011). Advances in engineering of proteins for thermal stability. *Int. J. Adv. Biotechnol. Res.*, 2(1), 190-200.
- [39] O’Fagain C. (2011). Engineering protein stability. *Methods Mol. Biol.*, 681, 103-136.
- [40] Pifferi G., Santoro P., Pedrani M. (1999). Quality and functionality of excipients. *Farmaco*, 54, 1-14.
- [41] Kamerzell T.J., Esfandiary R., Joshi S.B., Middaugh C.R., Volkin D.B. (2011). Protein–excipient interactions: Mechanisms and biophysical characterization applied to protein formulation development. *Adv. Drug Del. Rev.*, 63, 1118-1159.
- [42] Arakawa T., Tsumoto K., Kita Y., Chang B., Ejima D. (2007). Biotechnology applications of amino acids in protein purification and formulations. *Amino Acids*, 33, 587–605.
- [43] Kaushik J.K., Bhat R. (1998). Thermal stability of proteins in aqueous polyol solutions: Role of the surface tension of water in the stabilizing effect of polyols. *J. Phys. Chem. B.*, 102, 7058-7066.
- [44] Challener C.A. (2015). Excipient selection for protein stabilization. *Pharmaceutical Tech. Eu.*, 26-33.

- [45] Mollmann S.H., Elofsson U., Bukrinsky J.T., Frokjaer S. (2005). Displacement of adsorbed insulin by Tween 80 monitored using total internal reflection fluorescence and ellipsometry. *Pharm. Res.*, 22, 1931–1941.
- [46] Chi Y.E., Krishnan S., Randolph T.W., Carpenter J.F. (2003). Physical stability of proteins in aqueous solution: Mechanism and driving forces in nonnative protein aggregation. *Phar. Res.*, 20(9), 1325-1333.
- [47] Krieger E., Nabuurs S.B., Vriend G. (2003). Homology modeling. *Structural Bioinformatics*, 507-519.
- [48] Epstein C.J., Goldberger R.F., Anfinsen C.B. (1963). *Cold Spring Harb Symp. Quant. Biol.*, 28, 439.
- [49] Chothia C., Lesk A.M. (1986). The relation between the divergence of sequence and structure in proteins. *EMBO J.*, 5, 823–836.
- [50] Sander C., Schneider R. (1991). Database of homology-derived protein structures and the structural meaning of sequence alignment. *Proteins*, 9, 56–68.
- [51] Vyas V.K., Ukawala R.D., Ghate M., Chinthia C. (2012). Homology modeling a fast tool for drug discovery: Current perspectives. *Indian J. Pharm. Sci.*, 74 (1), 1-17.
- [52] Rost B. (1999). Twilight zone of protein sequence alignments. *Protein Eng.*, 12, 85–94.
- [53] Feng D.F., Doolittle R.F. (1987). Progressive sequence alignment as a prerequisite to correct phylogenetic trees. *J. Mol. Evol.*, 25, 351-360.

- [54] Thompson J.D., Higgins D.G., Gibson T.J. (1994). CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.*, 22, 4673-4680.
- [55] Karplus K., Barrett C., Hughey R. (1998). Hidden Markov models for detecting remote protein homologies. *Bioinformatics*, 14, 846-856.
- [56] Eddy S.R. (1998). Profile hidden Markov models. *Bioinformatics*, 14, 755-763.
- [57] Hoof R.W.W., Vriend G., Sander C., Abola E.E. (1996). Errors in protein structures. *Nature*, 381, 272.
- [58] Collura V., Higo J., Garnier J. (1993). Modeling of protein loops by simulated annealing. *Protein Sci.*, 2, 1502-1510.
- [59] Browne W.J., North A.C.T., Phillips D.C., Brew K., Vanaman T.C., Hill R.C. (1969). A possible three-dimensional structure of bovine α -lactalbumin based on that of hen's egg white lysozyme. *J. Mol. Biol.*, 42, 65-86.
- [60] Greer J. (1990). Comparative modeling methods: application to the family of the mammalian serine proteases. *Proteins*, 7, 317-334.
- [61] Levitt M. (1992). Accurate modeling of protein conformation by automatic segment matching. *J. Mol. Biol.*, 226, 507-533.
- [62] Xiang Z. (2006). Advances in homology protein structure modeling. *Curr. Prot. Pept. Sci.*, 7, 217-227.
- [63] Lee J., Lee D., Park H., Coutsias E.A., Seok C. (2010). Protein loop modeling by using fragment assembly and analytical loop closure. *Proteins*, 78, 3428-3436.

- [64] Moult J., James M.N. (1986). An algorithm for determining the conformation of polypeptide segments in proteins by systematic search. *Proteins, 1*, 146-163.
- [65] Simons K.T., Bonneau R., Ruczinski I., Baker D. (1999). Ab initio structure prediction of CASP III targets using ROSETTA. *Proteins, Suppl. 3*, 171–176.
- [66] Fiser A., Do R.K., Sali A. (2000). Modeling of loops in protein structures. *Protein Sci., 9*, 1753–1773.
- [67] Zhu J., Fan H., Periole X., Honig B., Mark A.E. (2008). Refining homology models by combining replica-exchange molecular dynamics and statistical potentials. *Proteins, 72*, 1171-1188.
- [68] Cornell W.D., Cieplak P., Bayly C.I., Gould I.R., Merz K.M., Ferguson D.M. (1995). A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J. Am. Chem. Soc., 117*, 5179-5197.
- [69] Liu H., Elstner M., Kaxiras E., Frauenheim T., Hermans J., Yang W. (2001). Quantum mechanics simulation of protein dynamics on long timescale. *Proteins, 44*, 484–489.
- [70] Krieger E., Koraimann G., Vriend G. (2002). Increasing the precision of comparative models with YASARA NOVA—a self-parameterizing force field. *Proteins, 47*, 393-402.
- [71] Das R., Qian B., Raman S., Vernon R., Thompson J., Bradley P. (2007). Structure prediction for CASP7 targets using extensive all-atom refinement with Rosetta. *Proteins, 69(S8)*, 118-128.
- [72] Han R., Leo-Macias A., Zerbino D., Bastolla U., Contreras-Moreira B., Ortiz A.R. (2008). An efficient conformational sampling method for homology modeling. *Proteins, 71*, 175-188.

- [73] Novotny J., Rashin A.A., Bruccoleri R.E. (1988). Criteria that discriminate between native proteins and incorrectly folded models. *Proteins*, 4, 19–30.
- [74] Morris A.L., MacArthur M.W., Hutchinson E.G., Thornton J.M. (1992). Stereochemical quality of protein structure coordinates. *Proteins*, 12, 345–364.
- [75] Czaplewski C., Rodziewicz-Motowidlo S., Liwo A., Ripoll D.R., Wawak R.J., Scheraga H.A. (2000). Molecular simulation study of cooperativity in hydrophobic association. *Protein Sci.*, 9, 1235–1245.
- [76] Baumann G., Frommel C., Sander C. (1989). Polarity as a criterion in protein design. *Protein Eng.*, 2, 329–334.
- [77] Sippl M.J. (1993). Recognition of errors in three dimensional structures of proteins. *Proteins*, 17, 355–362.
- [78] Vriend G., Sander C. (1993). Quality control of protein models: directional atomic contact analysis. *J. Appl. Crystallogr.*, 26, 47–60.
- [79] Huang E., Subbiah S., Levitt M. (1995). Recognizing native folds by the arrangement of hydrophobic and polar residues. *J. Mol. Biol.*, 252, 709-720.
- [80] Gerstein M., Levitt M. (1998). Comprehensive assessment of automatic structural alignment against a manual standard, the scop classification of proteins. *Protein Sci.*, 7, 445-456.
- [81] Sali A., Blundell T.L. (1993). Comparative protein modelling by satisfaction of spatial restraints. *Mol. Biol.*, 234, 779-815.
- [82] Sanchez R., Sali A. (1997). Evaluation of comparative protein structure modeling by MODELLER-3. *Proteins*, 1, 50-58.

- [83] Kuntal B.K., Aparoy P., Reddanna P. (2010). EasyModeller: A graphical interface to MODELLER. *BMC Research Notes*, 3, 226.
- [84] Yang J., Yan R., Roy A., Xu D., Poisson J., Zhang Y. (2015). The I-TASSER Suite: Protein structure and function prediction. *Nature Methods*, 12, 7-8.
- [85] Roy A., Kucukural A., Zhang Y. (2010). I-TASSER: a unified platform for automated protein structure and function prediction. *Nature Protocols*, 5, 725-738.
- [86] Zhang Y. (2008). I-TASSER server for protein 3D structure prediction. *BMC Bioinformatics*, 9, 40.
- [87] Arnold K., Bordoli L., Kopp J., Schwede T. (2006). The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. *Bioinformatics*, 22, 195–201.
- [88] Biasini M., Bienert S., Waterhouse A., Arnold K., Studer G., Schmidt T., Kiefer F., Cassarino T.G., Bertoni M., Bordoli L., Schwede T. (2014). SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. *Nucleic Acids Research*, 42, W252-258.
- [89] Yang A.S., Honig B. (1999). Sequence to structure alignment in comparative modeling using PrISM. *Proteins*, 37, 66-72.
- [90] Dolan M.A., Noah J.W., Hurt D. (2012). Comparison of common homology modeling algorithms: Application of user-defined alignments. *Homology Modeling: Methods and Protocols, Methods in Molecular Biology*, 857, 399-412.
- [91] Fechteler T., Dengler U., Schomburg D. (1995). Prediction of protein three-dimensional structures in insertion and deletion regions: A procedure for searching data bases of representative protein fragments using geometric scoring criteria. *J. Mol. Biol.*, 253, 114–131.

- [92] Almagro J.C., Beavers M.P., Hernandez-Guzman M.J., Shaulsky J., Butenhof K., Labute P., Thorsteinson N., Kelly K., Teplyakov A., Luo J., Sweet R., Gilliland G.L. (2011). Antibody modeling assessment. *Proteins: Struct. Func. Bioinf.*, 79, 3050–3066.
- [93] Sutcliffe M.J., Haneef I., Carney D., Blundell T.L. (1987). Knowledge-based modeling of homologous proteins, Part I: Three-dimensional frameworks derived from the simultaneous superposition of multiple structures. *Prot. Eng.*, 1, 377-384.
- [94] Sutcliffe M.J., Hayes F.R., Blundell T.L. (1987). Knowledge-based modeling of homologous proteins, Part II: Rules for the conformations of substituted sidechains. *Prot. Eng.*, 1, 385.
- [95] Misura K.M., Chivian D., Rohl C.A., Kim D.E., Baker D. (2006). Physically realistic homology models built with ROSETTA can be more accurate than their templates. *PNAS*, 103(14), 5361–5366.
- [96] Cavasotto C.N., Phatak S.S. (2009). Homology modeling in drug discovery: current trends and applications. *Drug Discov. Today*, 14, 676-683.
- [97] Guimaraes A.J., Hamilton A.J., de M. Guedes H.L., Nosanchuk J.D., Zancopé-Oliveira R.M. (2008). Biological function and molecular mapping of M antigen in yeast phase of *Histoplasma capsulatum*. *PLoS ONE*, 3(10), e3449.
- [98] Proell M., Riedl S.J., Fritz J.H., Rojas A.M., Schwarzenbacher R. (2008). The Nod-like receptor (NLR) family: a tale of similarities and differences. *PLoS ONE*, 3(4), e2119.
- [99] Gagnidze K., Sachchidanand, Rozenfeld R., Mezei M., Zhou M., Devi L.A. (2008). Homology modeling and site-directed mutagenesis to identify selective inhibitors of endothelin-converting enzyme-2. *J. Med. Chem.*, 51, 3378–3387.

- [100] Cavasotto C.N., Orry A.J.W., Murgolo N.J., Czarniecki M.F., Kocsi S.A., Hawes B.E., O'Neill K.A., Hine H., Burton M.S., Voigt J.H., Abagyan R.A., Bayne M.L., Monsma F.J., (2008). Discovery of novel chemotypes to a G-protein-coupled receptor through ligand-steered homology modeling and structure-based virtual screening. *J. Med. Chem.*, 51, 581–588.
- [101] Sangshetti J.N., Lokwani D.K., Sarkate A.P., Shinde D.B. (2011). Synthesis, antifungal activity and docking study of some new 1,2,4-triazole analogs. *Chem. Biol. Drug Des.*, 78, 800–809.
- [102] Wang Q., Mach R.H., Luedtke R.R., Reichert D.E. (2010). Subtype selectivity of dopamine receptor ligands: insights from structure and ligand-based methods. *J. Chem. Inf. Model*, 50, 1970-1985.
- [103] Wang Y., Su Z., Hsieh C., Chen C. (2009). Predictions of binding for dopamine D2 receptor antagonists by the SIE method. *J. Chem. Inf. Model*, 49, 2369-2375.
- [104] Mukherjee P., Desai P.V., Srivastava A., Tekwani B.L., Avery M.A. (2008). Probing the structures of leishmanial farnesyl pyrophosphate synthases: homology modeling and docking studies. *J. Chem. Inf. Model*, 48, 1026-1040.
- [105] Pietra F. (2009). Docking and MD simulations of the interaction of the tarantula peptide psalmotoxin-1 with ASIC1a channels using a homology model. *J. Chem. Inf. Model*, 49, 972-977.
- [106] Zhang Q., Li D., Wei P., Zhang J., Wan J., Ren Y. (2010). Structure-based rational screening of novel hit compounds with structural diversity for cytochrome P450 sterol 14 α -demethylase from *Penicillium digitatum*. *J. Chem. Inf. Model*, 50, 317-325.
- [107] Li G., Mosier P.D., Fang X., Zhang Y. (2009). Toward the three-dimensional structure and lysophosphatidic acid binding characteristics of the LPA4/ p2y9/GPR23 receptor: A homology modeling study. *J. Mol. Graphics Model.*, 28, 70-79.

- [108] Hasan Md. A., Alauddin S.M., Al Amin M., Nur S.M., Mannan A. (2014). *In silico* molecular characterization of cysteine protease YopT from *Yersinia pestis* by homology modeling and binding site identification. *Drug Target Insights*, 8, 1–9.
- [109] Vyas V.K., Ghate M., Patel K., Qureshi G., Shah S. (2015). Homology modeling, binding site identification and docking study of human angiotensin II type I (Ang II-AT1) receptor. *Biomedicine & Pharmacotherapy*, 74, 42–48.
- [110] Khan F.I., Govender A., Permaul K., Singh S., Bisetty K. (2015). Thermostable chitinase II from *Thermomyces lanuginosus* SSBP: Cloning, structure prediction and molecular dynamics simulations. *Journal of Theoretical Biology*, <http://dx.doi.org/10.1016/j.jtbi.2015.03.035>.
- [111] Hasan Md.A., Mazumder Md.H.H., Chowdhury A.S., Datta A., Khan Md.A. (2015). Molecular-docking study of malaria drug target enzyme transketolase in *Plasmodium falciparum* 3D7 portends the novel approach to its treatment. *Source Code for Biology and Medicine*, 10(7), DOI 10.1186/s13029-015-0037-3.
- [112] Shen M., Sali A. (2006). Statistical potential for assessment and prediction of protein structures. *Protein science*, 15, 2507-2524.
- [113] Zhang Y., Skolnick J. (2004). Scoring function for automated assessment of protein structure template quality. *Proteins*, 57, 702–710.
- [114] Van Gunsteren W.F., Billeter S.R., Eising A.A., Hunenberger P.H., Kruger P., Mark A.E., Scott W.R.P., Tironi I.G. (1996). *Biomolecular Simulation: The GROMOS96 Manual and User Guide*, 1-1042.
- [115] Schuler L.D., Daura X., van Gunsteren W.F. (2001). An improved GROMOS96 force field for aliphatic hydrocarbons in the condensed phase. *J. Comput. Chem.*, 22 (11), 1205–1218.

- [116] Van Der Spoel D., Lindahl E., Hess B., Groenhof G., Mark A.E., Berendsen H.J.C. (2005). GROMACS: Fast, flexible and free. *J. Comp. Chem*, 26(16), 1701-1718.
- [117] Brooks B.R., Bruccoleri R.E., Olafson B.D., States D.J., Swaminathan S., Karplus M. (1983). CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. *J. Comp. Chem.*, 4 (2), 187–217.
- [118] Pearlman D.A., Case D.A., Caldwell J.W., Ross W.S., Cheatham T.E. III, DeBolt S., Ferguson D., Seibel G., Kollman P. (1995). AMBER, a package of computer programs for applying molecular mechanics, normal mode analysis, molecular dynamics and free energy calculations to simulate the structural and energetic properties of molecules. *Comp. Phys. Commun.*, 91, 1–41.
- [119] Case D.A., Cheatham T., Darden T., Gohlke H., Luo R., Merz K.M. Jr., Onufriev A., Simmerling C., Wang B., Woods R. (2005). The Amber biomolecular simulation programs. *J. Computat. Chem.*, 26, 1668–1688.
- [120] Jorgensen W.L., Tirado-Rives J. (1988). The OPLS potential functions for proteins: Energy minimizations for crystals of cyclic peptides and crambin. *J. Am. Chem. Soc.*, 110(6), 1657-1666.
- [121] Pflieger C., Radestock S., Schmidt E., Gohlke H. (2013). Global and local indices for characterizing biomolecular flexibility and rigidity. *J. Comp. Chem.*, 34, 220–233.
- [122] Radestock S., Gohlke H. (2011). Protein rigidity and thermophilic adaptation. *Proteins*, 79, 1089–1108.
- [123] Krüger D. M., Rathi P. C., Pflieger C., Gohlke H. (2013). CNA web server: rigidity theory-based thermal unfolding simulations of proteins for linking structure, (thermo-)stability and function. *Nucleic Acids Res.*, 41, W340-W348.

- [124] Parthiban V., Gromiha M.M., Hoppe C., Schomburg D. (2007). Structural analysis and prediction of protein mutant stability using distance and torsion potentials: role of secondary structure and solvent accessibility. *Proteins*, 66(1), 41-52.
- [125] Parthiban V., Gromiha M. M., Abhinandan M., Schomburg D. (2007). Computational modeling of protein mutant stability: analysis and optimization of statistical potentials and structural features reveal insights into prediction model development. *BMC Structural Biology*, 7(54), 1-9.
- [126] Parthiban V., Gromiha M.M., Schomburg D. (2006). CUPSAT: prediction of protein stability upon point mutations. *Nucleic Acids Research*, 34, W239-W242.
- [127] Capriotti E., Fariselli P., Casadio R. (2005). I-Mutant2.0: predicting stability changes upon mutation from the protein sequence or structure. *Nucleic Acids Res.*, 33, W306-W310.
- [128] Capriotti E., Fariselli P., Rossi I., Casadio R. (2008). A three-state prediction of single point mutations on protein stability changes. *BMC Bioinformatics*, 9 (Suppl. 2), S6.
- [129] Schymkowitz J., Borg J., Stricher F., Nys R., Rousseau F., Serrano L. (2005). The FoldX web server: an online force field. *Nucleic Acids Research*, 33, W382-W388.
- [130] Masso M., Vaisman I.I. (2010). AUTO-MUTE: web-based tools for predicting stability changes in proteins due to single amino acid replacements. *Prot. Eng. Design Selec.*, 23(8), 683–687.
- [131] Cheng J., Randall A., Baldi P. (2006). Prediction of protein stability changes for single-site mutations using support vector machines. *Proteins*, 62, 1125-1132.

- [132] Yin S., Ding F., Dokholyan N.V. (2007). Eris: an automated estimator of protein stability. *Nature Methods*, 4, 466-467.
- [133] Kirsop B. (1987). Maintenance of yeast cultures. In D. R. Berry, I. Russell, G. G. Stewart (Eds.). *Yeast Biotechnology*, (3-32). Boston: Springer Netherlands.
- [134] Laemmli U.K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature*, 227, 680-685.
- [135] Wilson C.M. (1983). Staining of proteins on gels: comparisons of dyes and procedures. *Methods in Enzymol.*, 91, 236-247.
- [136] Lipman D.J., Pearson W.R. (1985). Rapid and sensitive protein similarity searches. *Science*, 227 (4693), 1435–1441.
- [137] Petersen T.N., Brunak S., von Heijne G., Nielsen H. (2011). SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nature Methods*, 8, 785-786.
- [138] Chou K.H., Shen H.B. (2007). Signal-CF: A subsite-coupled and window-fusing approach for predicting signal peptides. *Biochem. Biophys. Res. Comm.*, 357, 633-640.
- [139] Hiller K., Grote A., Scheer M., Münch R., Jahn D. (2004). PrediSi: prediction of signal peptides and their cleavage positions. *Nucleic Acids Res.*, 32, W375-W379.
- [140] Shen H.B., Chou K.C. (2007). Signal-3L: a 3-layer approach for predicting signal peptides. *Biochemical and Biophysical Res. Commun.*, 363, 297-303.
- [141] Song J., Tan H., Perry A.J., Akutsu T., Webb G.I., Whisstock J.C., Pike R.N. (2012). PROSPER: an integrated feature-based tool for predicting protease substrate cleavage sites. *PLoS ONE*, 7(11), e50300.

- [142] Letunic I., Doerks T., Bork P. (2015). SMART: recent updates, new developments and status in 2015. *Nucleic Acids Res.*, *43*, D257–D260.
- [143] Uniprot C. (2010). Ongoing and future developments at the Universal Protein Resource. *Nucleic Acids Res.*, *39*, D214–D219.
- [144] Gish W., Miller W., Muers E., Lipman D. (1990). *J. Mol. Biol.*, *215* (3), 403–410.
- [145] NCBI Resource Coordinators (2015). Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.*, *43*, D6–D17.
- [146] Berman H.M., Westbrook J., Feng Z., Gilliland G., Bhat T.N., Weissig H., Shindyalov I.N., Bourne P.E. (2000). The Protein Data Bank. *Nucleic Acids Res.*, *28*, 235–242.
- [147] van Rossum G., Drake F.L. (2001). Python Reference Manual. PythonLabs, Virginia, USA.
- [148] Pettersen E.F., Goddard T.D., Huang C.C., Couch G.S., Greenblatt D.M., Meng E.C., Ferrin T.E. (2004). UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.*, *25*(13), 1605–1612.
- [149] Guex N., Peitsch M.C. (1997). SWISS-MODEL and the Swiss-PdbViewer: An environment for comparative protein modeling. *Electrophoresis*, *18*, 2714–2723.
- [150] Xu J., Zhang Y. (2010). How significant is a protein structure similarity with TM score=0.5? *Bioinformatics*, *26*, 889–895.
- [151] Humphrey W., Dalke A., Schulten K. (1996). VMD -Visual Molecular Dynamics. *J. Molecular Graphics*, *14*, 33–38.

- [152] Yang J., Roy A., Zhang Y. (2013). Protein-ligand binding site recognition using complementary binding-specific substructure comparison and sequence profile alignment. *Bioinformatics*, 29, 2588-2595.
- [153] Roy A., Yang J., Zhang Y. (2012). COFACTOR: an accurate comparative algorithm for structure-based protein function annotation. *Nucleic Acids Research*, 40, W471-W477.
- [154] Huang B. (2009). MetaPocket: a meta approach to improve protein ligand binding site prediction. *OmicS*, 13(4), 325-330.
- [155] Dundas J., Ouyang Z., Tseng J., Binkowski A., Turpaz Y., Liang J. (2006). CASTp: computed atlas of surface topography of proteins with structural and topographical mapping of functionally annotated residues. *Nucleic Acid Research*, 34, W116-W118.
- [156] Volkamer D. K., Grombacher T., Rippmann F., Rarey M. (2012). Combining global and local measures for structure-based druggability predictions. *J. Chem. Inf. Model.*, 52, 360-372.
- [157] Capriotti E., Fariselli P., Casadio R. (2005). I-Mutant2.0: predicting stability changes upon mutation from the protein sequence or structure. *Nucleic Acids Res.*, 33, W306-W310.
- [158] Thakkar S.V., Joshi S.B., Jones M.E., Sathish H.A., Bishop S.M., Volkin D.B., Middaugh C.R. (2012). Excipients differentially influence the conformational stability and pretransition dynamics of two IgG1 monoclonal antibodies. *J. Pharm. Sci.*, 101, 3062–3077.
- [159] Platts L., Falconer R.F. (2015). Controlling protein stability: Mechanisms revealed using formulations of arginine, glycine and guanidinium HCl with three globular proteins. *Int. J. Pharmaceutics*, 486, 131-135.

- [160] Wang W., Zhong Q. (2014). Improved thermal stability of whey protein-maltodextrin conjugates at pH 5.0 by D-Glucose, sucrose, D-cellobiose, and lactose. *Food Hydrocolloids*, 41, 257-264.
- [161] Banks D.D., Hambly D.M., Scavezze J.L., Siska C.C., Stackhouse N.L., Gadgil H.S. (2009). The effect of sucrose hydrolysis on the stability of protein therapeutics during accelerated formulation studies. *J. Pharm. Sci.*, 98, 4501–4510.
- [162] Kaushik J.K., Bhat R. (2003). An analysis of the thermal stability of proteins in the presence of the compatible osmolyte trehalose. *J. Biol. Chem.*, 278(29), 26458-26465.
- [163] Wang W., Wang Y.J., Wang D.Q. (2008). Dual effects of Tween 80 on protein stability. *Int. J. Pharmaceutics*, 347, 31-38.
- [164] Grevesse C., Jijakli M.H., Duterme O., Colinet D., Lepoivre P. (1997). Preliminary study of exo-beta-1,3-glucanase encoding genes in relation to the protective activity of *Pichia anomala* (strain K) against *Botrytis cinerea* on postharvest apples. *Nucleotide Sequence/EMBL*.
- [165] Colovos C., Yeates T.O. (1993). Verification of protein structures: Patterns of nonbonded atomic interactions. *Protein Science*, 2, 1511-1519.
- [166] Luthy R., Bowie J.U., Eisenberg D. (1992). Assessment of protein models with three dimensional profiles. *Nature*, 356, 83-85.
- [167] Strub C., Alies C., Lougarre A., Ladurantie C., Czaplicki J., Fournier D. (2004). Mutation of exposed hydrophobic amino acids to arginine to increase protein stability. *BMC Biochemistry*, 5(9), doi:10.1186/1471-2091-5-9.

- [168] Sokalingam S., Raghunathan G., Soundrarajan N., Lee S.G. (2012). A study on the effect of surface lysine to arginine mutagenesis on protein stability and structure using green fluorescent protein. *PLoS ONE*, 7(7), e40410. doi:10.1371/journal.pone.0040410.
- [169] Mortazavi M., Hosseinkhani S. (2011). Design of thermostable luciferases through arginine saturation in solvent-exposed loops. *Prot. Eng., Des. & Selection*, 24(12), 893–903.
- [170] Zhou C., Zhang M., Wang Y., Guo W., Liu Z., Wang Y., Wang W. (2013). Enhancement of the thermo-alkali-stability of xylanase II from *Aspergillus usamii* E001 by site-directed mutagenesis. *Afr. J. Microbiol. Res.*, 7(16), 1535-1542.
- [171] Kumwenda B., Litthauer D., Bishop Ö.T., Reva O. (2013). Analysis of protein thermostability enhancing factors in industrially important *Thermus* bacteria species. *Evolutionary Bioinformatics*, 9, 327–342.

APPENDIX A

CHEMICALS AND THEIR SUPPLIERS

Acetic Acid (Sigma-Aldrich, Germany)
Acrylamide (Boehringer Mannheim, Germany)
Ammoniumpersulphate (Pharmacia Biotech, Sweden)
Arginine (Sigma, Japan)
Bacto-agar (Merck, Germany)
Bacto-peptone (Fluka, Denmark)
Bis-acrylamide (Pharmacia Biotech, Sweden)
Bromophenol Blue (Sigma, USA)
Butanol (Merck, Germany)
 β -mercaptoethanol (Sigma, USA)
Citric Acid (Merck, Germany)
Coomassie Brilliant Blue R-250 (ICN, USA)
Di-sodium Hydrogen Phosphate (Merck, Germany)
Ethanol (Merck, Germany)
Formaldehyde (Reidel-de Haen, Germany)
Glucose (Sigma-Aldrich, France)
Glutamic acid (Merck, Germany)
Glycerol (Sigma-Aldrich, Germany)
Glycine (Sigma, USA)
Hydrochloric Acid (Merck, Germany)
Histidine (Sigma-Aldrich, USA)
Lactose (Fluka, Netherlands)
Lysine (Aldrich, Switzerland)

Mannitol (Sigma-Aldrich, China)
Methanol (Sigma-Aldrich, Poland)
PEG 4000 (Merck, Germany)
PEG 6000 (Sigma-Aldrich, Germany)
Pluronic F-68 (Life Technologies, USA)
Potassium Dihydrogen Phosphate (Merck, Germany)
Proline (SAFC, USA)
Propanol (Merck, Germany)
Sodium Dodecyl Sulfate (Sigma, Japan)
Sodium Hydroxide (Merck, Germany)
Sodium Sulfate (Merck, Germany)
Sorbitol (Sigma, Germany)
Sucrose (Sigma, Switzerland)
TEMED (Pharmacia Biotech, Sweden)
Trehalose (Sigma, USA)
Trichloroacetic Acid (Merck, Germany)
Tris (Merck, Germany)
Tween 80 (Sigma-Aldrich, France)
Yeast extract (Fluka, India)

APPENDIX B

BUFFERS AND SOLUTIONS

Table B.1 SDS PAGE Gel Components and Staining Solutions.

Buffers/Solutions	Composition
<u>1.SDS-PAGE</u>	
Monomer Solution	30.8% T, 2.7% C _{bis}
4X Running Gel Buffer	1.5 M Tris-Cl at pH 8.8
4X Stacking Gel Buffer	0.5 M Tris-Cl at pH 6.8
SDS	10%
Initiator	10% Ammonium Persulfate
2X Treatment Buffer	0.125 M Tris-Cl, 4% SDS, 20% Glycerol, 10% β -mercaptoethanol and 0.02% Bromophenol blue at pH 6.8
Tank Buffer	0.025 M Tris, 0.192 M Glycine and 0.1% SDS
<u>2.COOMASSIAE BRILLIANT BLUE STAIN</u>	
Staining solution	0.025% Coomassiae brilliant blue R-250, 7% Acetic acid, 40% Methanol
Destain Solution I	7% Acetic Acid, 40% Methanol
Destain Solution II	7% Acetic Acid, 5% Methanol

APPENDIX C

PROGRAMS AND SERVERS

Table C.1 Programs and servers used.

Programs/Servers	Website
AUTO-MUTE	http://proteins.gmu.edu/automute/
BLAST	http://www.ncbi.nlm.nih.gov/blast/
CASTp	http://sts.bioe.uic.edu/castp/
CNA	http://cpclab.uni-duesseldorf.de/cna/
COACH	http://zhanglab.ccmb.med.umich.edu/COACH/
COFACTOR	http://zhanglab.ccmb.med.umich.edu/COFACTOR/
DoGSiteScorer	http://dogsite.zbh.uni-hamburg.de/
EasyModeller	http://modellergui.blogspot.com.tr/
Eris	http://troll.med.unc.edu/eris/
ERRAT	http://services.mbi.ucla.edu/ERRAT/
ExpASy	http://expasy.org/
FASTA	http://www.ebi.ac.uk/Tools/sss/fasta/
I-MUTANT2.0	http://folding.biofold.org/i-mutant/
I-TASSER	http://zhanglab.ccmb.med.umich.edu/I-TASSER/
MetaPocket	http://projects.biotec.tu-dresden.de/metapocket/
MODELLER	https://salilab.org/modeller/
MUpro	http://mupro.proteomics.ics.uci.edu/
NCBI	http://www.ncbi.nlm.nih.gov/
PDB	http://www.rcsb.org/pdb/
PrediSi	http://www.predisi.de/
PROSPER	https://prosper.erc.monash.edu.au/
Pymol	http://www.pymol.org/
Python	https://www.python.org/
SAVES	http://services.mbi.ucla.edu/SAVES/
Signal-3L	http://www.csbio.sjtu.edu.cn/bioinf/Signal-3L/
Signal-CF	http://www.csbio.sjtu.edu.cn/bioinf/Signal-CF/
SignalP 4.1	http://www.cbs.dtu.dk/services/SignalP/
SMART	http://smart.embl-heidelberg.de/
SPDB Viewer	http://spdbv.vital-it.ch/
TM Calculator	http://zhanglab.ccmb.med.umich.edu/TM-score/
UCSF Chimera	https://www.cgl.ucsf.edu/chimera/
UniProt	http://www.uniprot.org/
Verify 3D	http://services.mbi.ucla.edu/Verify_3D/
VMD	http://www.ks.uiuc.edu/Research/vmd/
WHATCHECK	http://swift.cmbi.ru.nl/gv/whatcheck/

