STRUCTURE PREDICTION OF TB RPOβ AND ITS MUTATIONS BINDING ANALYSIS

ERÇİN DİNÇER
20091109001

STRUCTURE PREDICTION OF TB RPOβ
AND ITS MUTATIONS BINDING ANALYSIS

ERÇİN DİNÇER

B.S.Computer Engineer, Istanbul University, 2009

Submitted to the Graduate School of Kadir Has University

in partial fulfillment of the requirements for the degree of

Master of Science

Graduate in Computational Biology and Bioinformatics

KADIR HAS UNIVERSITY

Haziran 2012

KADIR HAS UNIVERSITY

GRADUATE SCHOOL OF SCIENCE AND ENGINEERING

STRUCTURE PREDICTION OF TB RPOβ
AND ITS MUTATIONS BINDING ANALYSIS

ERÇİN DİNÇER

APPROVED BY:

Prof. Dr. Kemal Yelekçi (Kadir Has University)   _____
(Thesis Supervisor)

Doç. Dr. Mehmet Vezir Kahraman (Marmara University)_____

Yrd. Doç Dr. Demet Akten (Kadir Has University) _____

APPROVAL DATE:

STRUCTURE PREDICTION OF TB RPOβ AND ITS MUTATIONS BINDING ANALYSIS

Abstract

Today Tuberculosis is a disease that is still a high-risk categories. Rifampicin is a drug that's used common in the treatment of TB. We know that the effect of this drug in the region of RNA polymerase on TB. Unfortunately, there isn't any three-dimensional crystal structure in the rpoβ. In this study, three-dimensional model was created from DNA sequence and applied the resistance mutations of TB for computing resistance .

There are many online tools for three-dimensional modeling with using DNA or amino acid sequences. And the best result of the modeling was used in studying that's more same with the experimental results.

After finding best model, the mutations were applied for computing binding energy of mutations.

Tuberküloz Rpoβ protein yapısal modellemesi ve mutasyon dirençlerinin ölçülmesi.

Özet

Tüberküloz bugün hala yüksek risk sınıfında bir hastalıktır. Verem tedavisinde en yaygın kullanılan ilaç rifampisindir. Günümüzde biliyoruz ki bu ilacın TB üzerinde ki etki bölgesi RNA Polimerazdır. Yapılan araştırmalarda ne yazık ki henüz Rpoβ için üç boyutlu bir kristal yapısı elde edilememiştir. Bu çalışmada DNA sekansından yola çıkılarak; üç boyutlu model oluşturulması, bu model üzerinden mutasyon dirençlerinin ölçülmesi ve sonuçlarının değerlendirilmesi hedeflenmiştir.

DNA veya Amino Asit sekanslarından yararlanarak üç boyutlu modelleme yapan bir çok online tool arasından yapılan modellemeler sonucunda deneysel sonuçlara en uygun çıkan model kullanılmıştır.

En iyi model bulunduktan sonra, bu model üzerinde TB mutasyonları uygulanarak yeni durumda RIF'in bağlanma enerjileri ölçülmüştür.

Acknowledgements

All thanks for Prof. Dr. Kemal Yelekçi, my dissertation supervisor. Having the opportunity to work with him over the years was intellectually rewarding and fulfilling. Thanks to Yrd. Doç Dr. Demet Akten for helping with suggestion when I started my thesis.

Thanks to my family for placing in my life. And thanks for my family and friend who supported me for finishing my thesis.

Table of Contents

List of Tables

List of Figures

List of Symbols

ΔG_{bind}: Estimation of Binding Affinity

m: sequence of length

n: sequence of length

m+1: the matrix dimension

n+1: the matrix dimension

S(i,j): score

i-1: the score from the cell at position

j-1: the score from the cell at position

s[*i,j*]: the new score at position

s[*i,j*-1]: the score one cell to the left

s[*i*-1,*j*]: the score immediately above the new cell

K,λ: constants

m: length of query sequence

n: length of the entire database

S: score of the alignment

E: expect value

K_i: binding constant

V: pair-wise evaluations

$\rho(\vec{x}_i)(\vec{x}_j)$: probability density function

$\rho(\vec{x}_i)$: the single body distribution function for atom I and is a constant for a given protein

ΔS_{conf}: entropy lost upon binding

L: ligand

P: protein

List of Abbreviations

TB: Tuberculosis

MTB: Mycobacterium Tuberculosis

RIF: Rifampicin

MD: Molecular Dynamics

3D: Three dimension

BLAST: Basic local alignment search

BLOSUM: Blocks of aminoacid substitution matrix

DSSP: Dictionary of Protein Secondary Structure

DOPE: Discrete Optimized Protein Energy

PDF: Probability Density Function

PDB: Protein data bank

RMSD: Root-mean-square deviation

MC: Monte carlo

GA: Genetic algorithm

FF: Force field

SFs: Scoring functions

DS: Discovery studio

gi: Query sequence

E: Expect value

rDAT: rat dopamine transporter

gpf: Grid parameter file

glf: Grid log file

dlf: Docking log file

dpf: Docking parameter file

MC: Monte Carlo

HB: Hydrogen bonding

Chapter 1

Introduction

This thesis is about computational methods of determining the drug resistance of mycobacterium tuberculosis (TB) on it's a First-Line drug; rifampicin (RIF). For examining RIF binding onto the active site of Mycobacterium tuberculosis, this study intends to utilize molecular docking and molecular dynamics (MD) simulations.

Drug	Cellular function inhibited	Target
First-line drugs		
INH	Mycolic acid synthesis	Enoyl reductase
RIF	RNA synthesis	RNA polymerase
Ethambuto (EMB)	Arabinogalactan synthesis	Arabinosyl transferase
Pyrazinamide (PZA)	Unclear	Unclear
Quinolones	DNA supercoiling	DNA gyrase
Ethionamide	Mycolic acid synthesis	Enoyl reductase
STM	Protein synthesis	30S ribosomal subunit
KAN, AMK	Protein synthesis	30S ribosomal subunit
Capreomycin	Protein synthesis	30S/50S ribosomal subuni

Table 1.1 TB Drug Table

RIF affects many bacteria by interacting with the RNA polymerase β- subunit and preventing transcription although it is not specific for mycobacteria. Clinical RIF causes distinct mutations in rpoβ because its resistance is mostly high- level.[1]
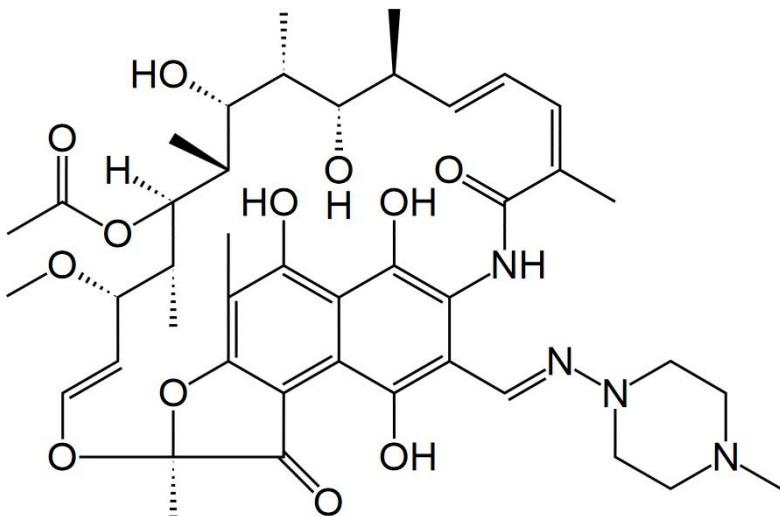
Figure 1.1: Chemical structure of RIF

1.1 Tuberculosis (TB)

TB usually infects the human lower respiratory system as a microbial disease, which has affected human beings for several millennia [2]. M. tuberculosis, is the aetiological agent of TB which was extracted 125 years ago by Koch [3]. Although there are many progress in the prevention and treatment of the disease, this ancient chastise still remains as a major pathogen of human and a global tragedy with immense public health and economic implications. World Health Organisation (WHO) identifies a need of US$47 billion to implement countrywide programmes to stop TB while another US$9 billion for the research and development of new diagnostics and treatments for TB according to Migliori et al's report (2007).

7% of all deaths in developing countries and 26% of avoidable matuıre deaths worldwide is based on TB for [4]. 3 million people dying from TB every year in average [5] and of all the infectious bacteria, it is currently the leading killer of adults in the world. WHO has appraise a astounding 8 million new cases globally and has projected about 30 million deaths from TB in this decade according to Manca's report et al. (1997).

There has been a constant rise in notification of TB in Malaysia for over the past 10 years. In year 2000, a terrible 15,057 cases of TB was reported where the incidence rate is 64.7 per 100,000 populations. The TB and HIV co-infection numbers has also escalated from 6 cases in 1990 to 734 cases in 2000. Advanced TB is seen in most patients with TB-HIV co-

2

infections, therefore the number of deaths due to TB-HIV has also increased. The fast growing numbers of immigrant workers from high TB burden neighbouring countries which might add to the problem of multi-drug resistant TB caused the worries to be further blended [6].

RIF is the prime drug for the treatment of TB [7] since 1952. Anyway its use has been restricted by up to 30% increase of RIF resistant streches [8]. By the increase of multi drug resistant M. tuberculosis strains especially amongst HIV infected individuals the problem has further been complicated [9]. A considerable amount of non-TB mycobacteria have been isolated from acquired immune deficiency syndrome (AIDS) patients [10]. In the AIDS patients such opportunistic mycobacteria include the member of Mycobacterium avium complex (MAC) caused the prevalent "TB-like" infection. These pathogens are mostly naturally resistant to RIF . When compared, the survival rates of AIDS patients who are not infected are much higher than one's with MAC infection [11].

The search for alternatives to RIF has prompted by the above declarations. However, the comprehending of drug-receptor interactions is required in order to develop strategies for the design of novel and potent drugs against M. tuberculosis. Therefore, there is a compelling need to understand resistance development at the molecular level that remains an enigma until today.

1.2 The significance of molecular modeling in TB research

The efficiency of TB treatment, control and prevention programs have been complicated by the limitation number of efficacious therapeutic agents to treat patients infected with MAC and rise of multi-drug resistant strains these days. The enduring worldwide threat of TB accentuates the importance of the urgent need in more effective diagnosis therapies, which is currently at a very slow process. It's main cause is the lack of detailed structural features regarding the drug-receptor interactions. Thus, to facilitate the rational development and improvement of anti-TB medications , the comprehending and insights of the molecular events that lead to drug action or resistant in M. tuberculosis are important [12].

The pharmacophore hypotheses which derived from those inhibitors with known structures are the top influencer of drug design . The pharmacophore model did not provide the details of the drug-receptor interactions, despite the fact that these hypotheses were able to discover some new inhibitors. Thus, molecular modelling method can predict the binding modes of RIF as well as its derivatives onto rpoβ. The search for lead/potential inhibitor(s) and the strategies for the design of new anti-TB compounds can be formulated with the comprehending of binding modes at the molecular level between RIF and rpoβ [13].

Acceptor-receptor binding mode predictions have led to a faster discoveries of new lead compounds with impressive improvements in the accuracy and speed of molecular docking. However, there are still many difficulties to over helm. Initially, the binding modes accuracy relies on the correct assessment of acceptor-receptor interaction energies and scoring functions which are simplified for computational efficiency. Secondly, the sampling of acceptor in flexible binding pocket has not been achieved. Third, the involvement of solvent molecules is yet to be addressed in molecular docking method. Fourth because the acceptor might be more mobile in the bound state, the molecular docking method is not able to provide the dynamics data of acceptor within the receptor binding site [14].

Thus, to elucidate a more refined and complete understanding of the binding properties of RIF within the Rpoβ binding pocket, MD simulation is also another alternative. MD simulation technique is wasting much time and expensive (regarding large computational storage compared to molecular docking method). However, MD simulation is able to provide the dynamics of RIF-Rpoβ complex as well as the detailed insights of intermolecular relationships. MD simulation also able to deal with the docking problems mentioned above, because it uses more accurate force field (a whole atom approach instead of the united atom method in molecular docking simulation). Generally, water molecules play a critical role in determining the conformation of an acceptor in a binding pocket because they are mediating interactions with the protein. Thus, MD simulation allows the involvement of explicit waters (instead of the simplified solvation parameter used in docking simulation) unlike molecular docking. Finally, as RIF and Rpoβ‛s flexibilities might be critical for recognition between each other, MD simulation methodology will also permit flexibility of both.

Chapter 2

2.1 Molecular modelling

X-ray crystallography and nuclear magnetic resonance (NMR) spectroscopy have arrives to high resolution at the molecular level of three dimensional structures for a numerous proteins with the ease of progresses in molecular biology. However, some proteins cannot easily crystallized, because NMR experiment does not give complete atomic structures. It only gives overviews for proteins with amino acids further than the number of 200 residues. Thus, molecular docking and MD simulation that are computer-based molecular modelling techniques are clearly fascinating methods to predict and derive a molecular insight of the acceptor-protein complex structure[15].

For describing the study of molecules and molecular systems, the molecular modelling term is used generally. It is also a technic that uses theoretical methods to investigate and predict chemical entities and processes. In order to study systems ranging from small chemical molecules to large biological molecules and material assemblies this technique has been used in many ranges such as chemistry, biology, or materials science . The atomistic level description of the studied systems can be gathered with molecular modelling. Inevitably computers are required to study reasonably sized systems while the simplest system can be performed by hand using theoretical calculations. Today's molecular modelling is invariably associated with computer modelling so it's known as computational chemistry. Developments in speed and memory ability of computational power have allowed extensive range of models and up to millions of atoms to be included in the computation[16].

Computer-based molecular modelling simulates the chemical structures and reactions based on the elemental laws of physics in terms of numbers usually. Researchers to study chemical phenomena by running computations instead of wet-lab experiment with this method. Not only stable molecules, but also unstable intermediates and transition states of chemical structures and reactions can be modeled by some simulations. With wet-lab experiment, it is almost impossible to observe molecules and reactions but these hypothetical methods can provide information about them easily. Therefore, molecular modelling is an autonomous research area which could be critical adjunct and alternative to exploratory studies[17].

Chapter 3

Generate 3D model for rpoβ protein and its ligand interaction with mutations

Introduction

We are working on a protein (rpoβ).

The sequence of   rpoβ protein (1079 amino acids) was downloaded for structural modeling from NCBI. Multiple alignments of the related sequences were performed using the online available ClustalW program accessible through the European Bioinformatics Institute. There isn't X-ray crystallographic or NMR structure of this protein. Tertiary structures of rpoβ protein were modeled on the basis of different template structures from different web based tertiary structures area. Each result of 3D structures was docked with RIF and the results were compared with experimental result. At sum which model is able to get more same result with experimental result, we could accept it for a good research model.[18]

3.1 Working on web based tertiary structures program

First try on swiss-model; An automated knowledge-based protein modelling server. It used 3tiB for based on template.
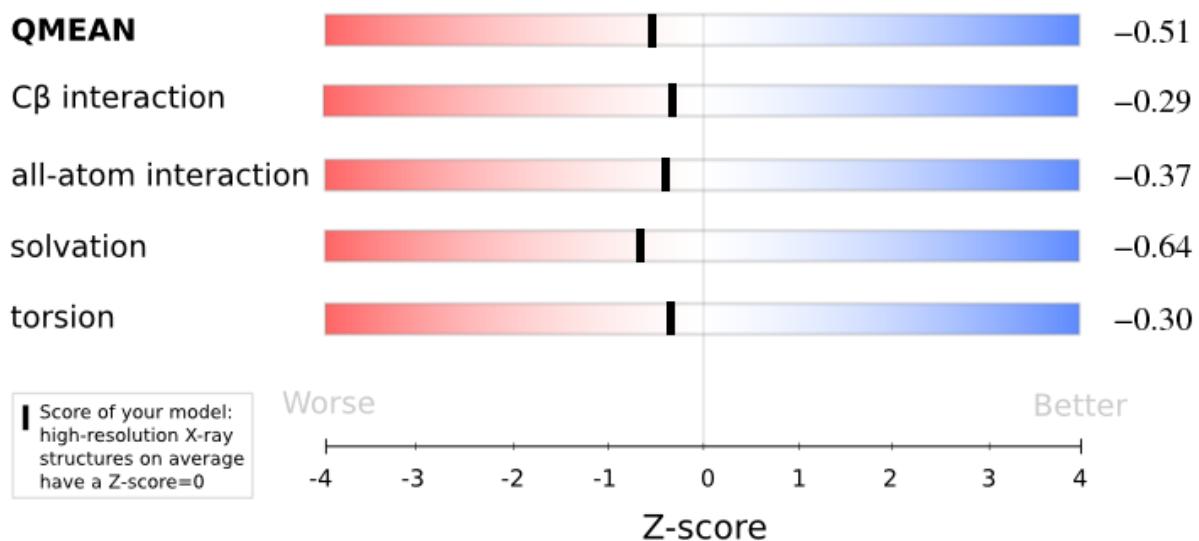
Figure 3.1 Swiss-Model Alligment
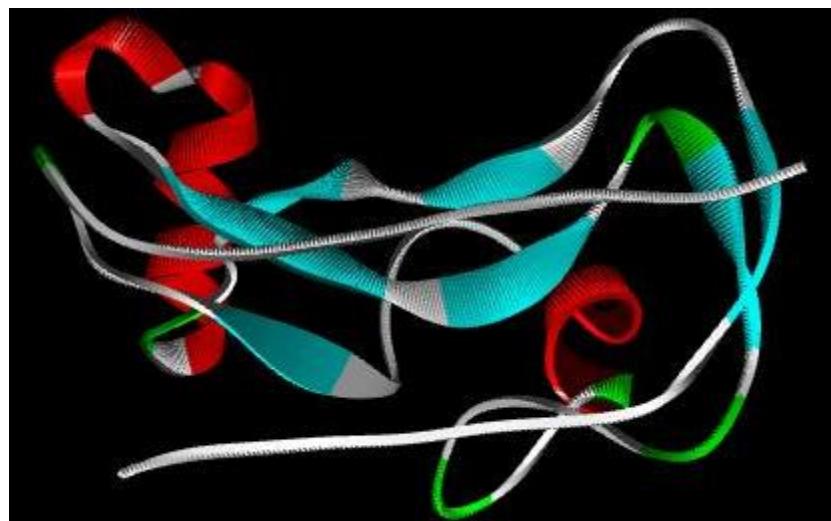
Figure 3.2 Swiss-Model rpoβ model
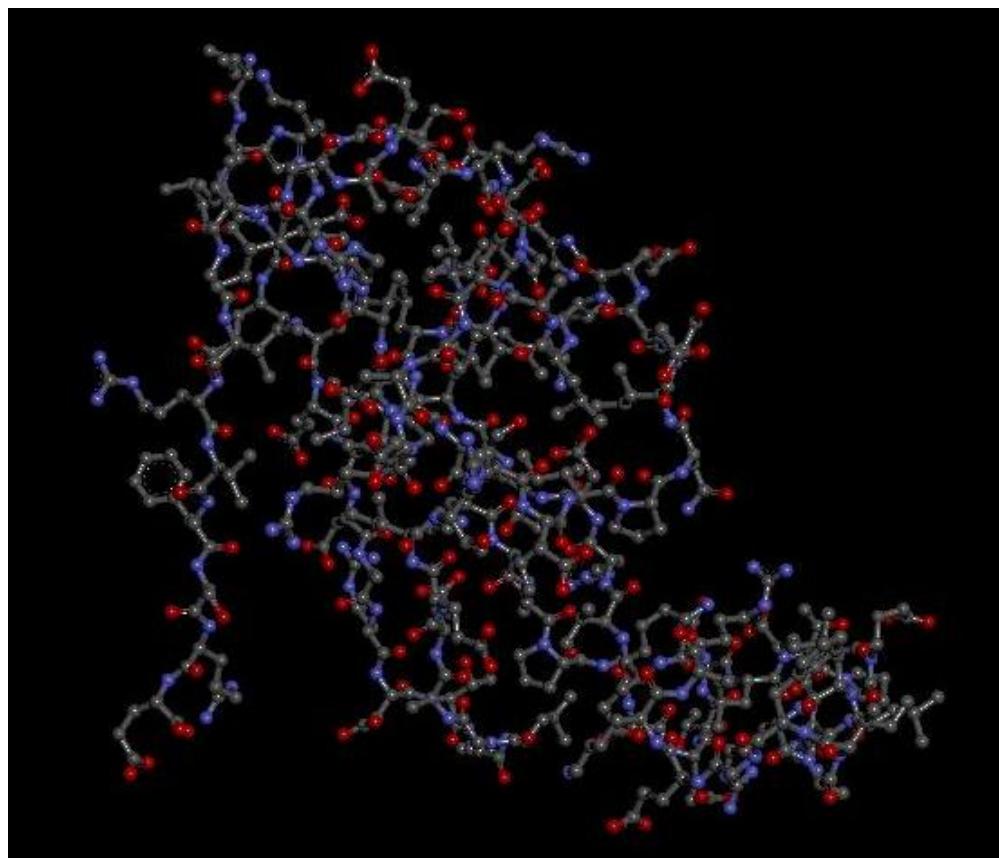
Figure 3.3 Swiss-Model rpoβ Model with atomic view

Secand try on CPHmodels 3.2 Server; CPHmodels 3.2 is a protein homology modeling server. The template recognition is based on profile-profile alignment guided by secondary structure and exposure predictions.

Query sequence:

>gi_15607807_ref_NP_215181.1_
MADSRQSKTAASPSPSRPQSSSNNSVPGAPNRVSFAKLREPLEVPGLLDVQTDSFEWLIG
SPRWRESAAERGDVNPVGGLEEVLYELSPIEDFSGSMSLSFSDPRFDDVKAPVDECKDKD
MTYAAPLFVTAEFINNNTGEIKSQTVFMGDFPMMTEKGTFIINGTERVVVSQLVRSPGVY
FDETIDKSTDKTLHSVKVIPSRGAWLEFDVDKRDTVGVRIDRKRRQPVTVLLKALGWTSE
QIVERFGFSEIMRSTLEKDNTVGTDEALLDIYRKLRPGEPPTKESAQTLLENLFFKEKRY
DLARVGRYKVNKKLGLHVGEPITSSTLTEEDVVATIEYLVRLHEGQTTMTVPGGVEVPVE
TDDIDHFGNRRLRTVGELIQNQIRVGMSRMERVVRERMTTQDVEAITPQTLINIRPVVAA
IKEFFGTSQLSQFMDQNNPLSGLTHKRRLSALGPGGLSRERAGLEVRDVHPSHYGRMCPI
ETPEGPNIGLIGSLSVYARVNPFGFIETPYRKVVDGVVSDEIVYLTADEEDRHVVAQANS
PIDADGRFVEPRVLVRRKAGEVEYVPSSEVDYMDVSPRQMVSVATAMIPFLEHDDANRAL
MGANMQRQAVPLVRSEAPLVGTGMELRAAIDAGDVVVAEESGVIEEVSADYITVMHDNGT
RRTYRMRKFARSNHGTCANQCPIVDAGDRVEAGQVIADGPCTDDGEMALGKNLLVAIMPW
EGHNYEDAIILSNRLVEEDVLTSIHIEEHEIDARDTKLGAEEITRDIPNISDEVLADLDE
RGIVRIGAEVRDGDILVGKVTPKGETELTPEERLLRAIFGEKAREVRDTSLKVPHGESGK
VIGIRVFSREDEDELPAGVNELVRVYVAQKRKISDGDKLAGRHGNKGVIGKILPVEDMPF
LADGTPVDIILNTHGVPRRMNIGQILETHLGWCAHSGWKVDAAKGVPDWAARLPDELLEA
QPNAIVSTPVFDGAQEAELQGLLSCTLPNRDGDVLVDADGKAMLFDGRSGEPFPYPVTVG
YMYIMKLHHLVDDKIHARSTGPYSMITQQPLGGKAQFGGQRFGEMECWAMQAYGAAYTLQ
ELLTIKSDDTVGRVKVYEAIVKGENIPEPGIPESFKVLLKELQSLCLNVEVLSSDGAAIE
LREGEDEDLERAAANLGINLSRNESASVEDLA

Query Mw: 129235 (1172 aa)
Searching for template ...

Round   0. Hits better than threshold: 0.000010:
entry: 3IYD chain: C score: 1175 E: 0.0
entry: 1IW7 chain: C score: 1088 E: 0.0
entry: 1IW7 chain: M score: 1088 E: 0.0
entry: 2A6E chain: C score: 1088 E: 0.0
entry: 2A6E chain: M score: 1088 E: 0.0
entry: 2A6H chain: C score: 1088 E: 0.0
entry: 2A6H chain: M score: 1088 E: 0.0
entry: 2A68 chain: C score: 1088 E: 0.0
entry: 2A68 chain: M score: 1088 E: 0.0
entry: 2A69 chain: C score: 1088 E: 0.0
entry: 2A69 chain: M score: 1088 E: 0.0
entry: 2BE5 chain: C score: 1088 E: 0.0
entry: 2BE5 chain: M score: 1088 E: 0.0
entry: 2CW0 chain: C score: 1088 E: 0.0
entry: 2CW0 chain: M score: 1088 E: 0.0
entry: 2O5I chain: C score: 1088 E: 0.0
entry: 2O5I chain: M score: 1088 E: 0.0
entry: 2O5J chain: C score: 1088 E: 0.0
entry: 2O5J chain: M score: 1088 E: 0.0
entry: 2PPB chain: C score: 1088 E: 0.0
entry: 2PPB chain: M score: 1088 E: 0.0
entry: 1SMY chain: C score: 1088 E: 0.0
entry: 1SMY chain: M score: 1088 E: 0.0
entry: 1ZYR chain: C score: 1088 E: 0.0
entry: 1ZYR chain: M score: 1088 E: 0.0
entry: 3DXJ chain: C score: 1088 E: 0.0
entry: 3DXJ chain: M score: 1088 E: 0.0
entry: 3EQL chain: C score: 1088 E: 0.0
entry: 3EQL chain: M score: 1088 E: 0.0
entry: 3AOH chain: C score: 1084 E: 0.0
entry: 3AOH chain: M score: 1084 E: 0.0
entry: 3AOH chain: H score: 1083 E: 0.0
entry: 2GHO chain: C score: 1083 E: 0.0
entry: 1YNJ chain: C score: 1083 E: 0.0
entry: 1YNN chain: C score: 1083 E: 0.0
entry: 1HQM chain: C score: 1082 E: 0.0
entry: 1I6V chain: C score: 1082 E: 0.0
entry: 1L9Z chain: C score: 1057 E: 0.0
entry: 1L9U chain: C score: 1057 E: 0.0
entry: 1L9U chain: L score: 1057 E: 0.0
entry: 3AOI chain: M score: 1039 E: 0.0

8

```
entry: 3AOI chain: C score: 1034 E: 0.0
entry: 3AOI chain: H score: 1032 E: 0.0
entry: 3LU0 chain: C score: 703 E: 0.0
entry: 2Y0S chain: B score: 179 E: 2e-44
entry: 2Y0S chain: R score: 179 E: 2e-44
entry: 2WAQ chain: B score: 179 E: 2e-44
entry: 2WB1 chain: B score: 179 E: 2e-44
entry: 2WB1 chain: R score: 179 E: 2e-44
entry: 2PMZ chain: B score: 178 E: 4e-44
entry: 2PMZ chain: R score: 178 E: 4e-44
entry: 3HKZ chain: B score: 178 E: 4e-44
entry: 3HKZ chain: J score: 178 E: 4e-44
entry: 3K1F chain: B score: 151 E: 5e-36
entry: 3H0G chain: B score: 150 E: 8e-36
entry: 3H0G chain: N score: 150 E: 8e-36
entry: 2NVX chain: B score: 146 E: 2e-34
entry: 3QT1 chain: B score: 145 E: 4e-34
entry: 3TBI chain: B score: 145 E: 4e-34
entry: 3HOU chain: B score: 145 E: 4e-34
entry: 3HOU chain: N score: 145 E: 4e-34
entry: 3HOV chain: B score: 145 E: 4e-34
entry: 2R92 chain: B score: 145 E: 5e-34
entry: 2R93 chain: B score: 145 E: 5e-34
entry: 2B8K chain: B score: 145 E: 5e-34
entry: 3FKI chain: B score: 145 E: 5e-34
entry: 3HOW chain: B score: 145 E: 5e-34
entry: 3HOX chain: B score: 145 E: 5e-34
entry: 2B63 chain: B score: 145 E: 6e-34
entry: 2JA7 chain: B score: 145 E: 6e-34
entry: 2JA7 chain: N score: 145 E: 6e-34
entry: 2JA8 chain: B score: 145 E: 6e-34
entry: 2JA5 chain: B score: 145 E: 6e-34
entry: 2JA6 chain: B score: 145 E: 6e-34
entry: 1PQV chain: B score: 145 E: 6e-34
entry: 1Y1V chain: B score: 145 E: 6e-34
entry: 1Y1W chain: B score: 145 E: 6e-34
entry: 1Y1Y chain: B score: 145 E: 6e-34
entry: 1Y77 chain: B score: 145 E: 6e-34
entry: 2R7Z chain: B score: 145 E: 6e-34
entry: 3H3V chain: C score: 145 E: 6e-34
entry: 3PO3 chain: B score: 145 E: 6e-34
entry: 3J0K chain: B score: 145 E: 6e-34
entry: 2E2I chain: B score: 145 E: 6e-34
entry: 3HOY chain: B score: 145 E: 6e-34
entry: 4A3C chain: B score: 145 E: 6e-34
entry: 4A3B chain: B score: 145 E: 6e-34
entry: 4A3D chain: B score: 145 E: 6e-34
entry: 4A3E chain: B score: 145 E: 6e-34
entry: 4A3F chain: B score: 145 E: 6e-34
entry: 4A3J chain: B score: 145 E: 6e-34
entry: 4A3K chain: B score: 145 E: 6e-34
entry: 4A3L chain: B score: 145 E: 6e-34
entry: 4A3M chain: B score: 145 E: 6e-34
entry: 4A3G chain: B score: 145 E: 6e-34
entry: 4A3I chain: B score: 145 E: 6e-34
entry: 3HOZ chain: B score: 145 E: 6e-34
entry: 3GTM chain: B score: 144 E: 6e-34
entry: 2NVQ chain: B score: 144 E: 6e-34
entry: 2NVT chain: B score: 144 E: 6e-34
entry: 2YU9 chain: B score: 144 E: 6e-34
entry: 3RZD chain: B score: 144 E: 6e-34
entry: 3RZO chain: B score: 144 E: 6e-34
entry: 3S14 chain: B score: 144 E: 6e-34
entry: 3S15 chain: B score: 144 E: 6e-34
entry: 3S16 chain: B score: 144 E: 6e-34
entry: 3S17 chain: B score: 144 E: 6e-34
entry: 3S1M chain: B score: 144 E: 6e-34
entry: 3S1N chain: B score: 144 E: 6e-34
entry: 3S1Q chain: B score: 144 E: 6e-34
entry: 3S1R chain: B score: 144 E: 6e-34
entry: 3S2D chain: B score: 144 E: 6e-34
entry: 3S2H chain: B score: 144 E: 6e-34
```

```
entry: 3I4M chain: B score: 144 E: 6e-34
entry: 3I4N chain: B score: 144 E: 6e-34
entry: 3PO2 chain: B score: 144 E: 6e-34
entry: 1I6H chain: B score: 144 E: 6e-34
entry: 1NIK chain: B score: 144 E: 6e-34
entry: 1R5U chain: B score: 144 E: 6e-34
entry: 1R9S chain: B score: 144 E: 6e-34
entry: 1WCM chain: B score: 144 E: 6e-34
entry: 1R9T chain: B score: 144 E: 7e-34
entry: 1SFO chain: B score: 144 E: 7e-34
entry: 2VUM chain: B score: 144 E: 7e-34
entry: 3GTL chain: B score: 144 E: 7e-34
entry: 3GTO chain: B score: 144 E: 7e-34
entry: 3GTP chain: B score: 144 E: 7e-34
entry: 3GTQ chain: B score: 144 E: 7e-34
entry: 3M3Y chain: B score: 144 E: 7e-34
entry: 3M4O chain: B score: 144 E: 7e-34
entry: 2E2J chain: B score: 144 E: 7e-34
entry: 2E2H chain: B score: 144 E: 7e-34
entry: 2NVZ chain: B score: 144 E: 7e-34
entry: 3GTG chain: B score: 144 E: 7e-34
entry: 3GTJ chain: B score: 144 E: 7e-34
entry: 3GTK chain: B score: 144 E: 7e-34
entry: 4A93 chain: B score: 144 E: 1e-33
entry: 3K7A chain: B score: 143 E: 2e-33
entry: 1NT9 chain: B score: 131 E: 6e-30
entry: 1I50 chain: B score: 123 E: 2e-27
entry: 2NVY chain: B score: 123 E: 2e-27
entry: 1TWF chain: B score: 123 E: 2e-27
entry: 1I3Q chain: B score: 123 E: 2e-27
entry: 1K83 chain: B score: 123 E: 2e-27
entry: 1TWA chain: B score: 122 E: 4e-27
entry: 1TWC chain: B score: 122 E: 4e-27
entry: 1TWG chain: B score: 122 E: 4e-27
entry: 1TWH chain: B score: 122 E: 4e-27
entry: 3CQZ chain: B score: 100 E: 2e-20
entry: 3MLQ chain: D score: 92 E: 7e-18
entry: 3MLQ chain: B score: 91 E: 8e-18
entry: 3MLQ chain: C score: 91 E: 9e-18
entry: 3MLQ chain: A score: 89 E: 4e-17
entry: 3LTI chain: A score: 77 E: 2e-13

Retrieving template ...

Entry: 3iyd
Chain: C

Making profile-profile alignment ...

Score: 1493.0 bits
Identity: 53.5 %

Query: 32  RVSFAKLREPLEVPGLLDVQTDSFEWLIGSPRWRESAAERGDVNPVG--GLEEVLYELSP 89
            R  F K  + L+VP LL +Q DSF+  I              + +P G  GLE    + P
Templ: 3   RKDFGKRPQVLDVPYLLSIQLDSFQKFI-------------EQDPEGQYGLEAAFRSVFP 49


Query: 90  IEDFSGSMSLSFSDPRFDDVKAPVDECKDKDMTYAAPLFVTAEFI-------NNNTGEIK 142
            I+ +SG+  L +  R +    V EC+ + +TY+APL V  +         +IK
Templ: 50  IQSYSGNSELQYVSYRLGEPVFDVQECQIRGVTYSAPLRVKLRLVIYEREAPEGTVKDIK 109


Query: 143 SQTVFMGDFPMMTEKGTFIINGTERVVVSQLVRSPGVYFDETIDK--STDKTLHSVKVIP 200
             Q V+MG+ P+MT+ GTF+INGTERV+VSQL RSPGV+FD    K    S+ K L++ ++IP
Templ: 110 EQEVYMGEIPLMTDNGTFVINGTERVIVSQLHRSPGVFFDSDKGKTHSSGKVLYNARIIP 169


Query: 201 SRGAWLEFDVDKRDTVGVRIDRKRRQPVTVLLKALGWTSEQIVERF-------------G 247
            RG+WL+F+ D +D + VRIDR+R+ P T++L+AL +T+EQI++  F             G
Templ: 170 YRGSWLDFEFDPKDNLFVRIDRRRKLPATIILRALNYTTEQILDLFFEKVDLLAKLSQSG 229


Query: 248 FSEI-------------MRSTLEKDNTVGTDEALLDIYRKLRPGEPPTKESAQTLLENLF 294
             I                + TL  D T    AL++IYR +RPGEPPT+E+A++L ENLF
Templ: 230 HKRIETLFTNDLDHGPYISETLRVDPTNDRLSALVEIYRMMRPGEPPTREAAESLFENLF 289
```

10

```
Query:  295 FKEKRYDLARVGRYKVNKKLGLHVGEPITSS-TLTEEDVVATIEYLVRLHEGQTTMTVPG 353
             F E RYDL+ VGR K N+ L   + E I  S  L+++D++   ++ L+ +  G+
Templ:  290 FSEDRYDLSAVGRMKFNRSL---LREEIEGSGILSKDDIIDVMKKLIDIRNGKG------ 340


Query:  354 GVEVPVETDDIDHFGNRRLRTVGELIQNQIRVGMSRMERVVRERMTTQDVEAITPQTLIN 413
                    E DDIDH GNRR+R+VGE+ +NQ RVG+ R+ER V+ER++  D++ + PQ +IN
Templ:  340 ------EVDDIDHLGNRRIRSVGEMAENQFRVGLVRVERAVKERLSLGDLDTLMPQDMIN 394


Query:  414 IRPVVAAIKEFFGTSQLSQFMDQNNPLSGLTHKRRLSALGPGGLSRERAGLEVRDVHPSH 473
             +P+ AA+KEFFG+SQLSQFMDQNNPLS +THKRR+SALGPGGL+RERAG EVRDVHP+H
Templ:  395 AKPISAAVKEFFGSSQLSQFMDQNNPLSEITHKRRISALGPGGLTRERAGFEVRDVHPTH 454


Query:  474 YGRMCPIETPEGPNIGLIGSLSVYARVNPFGFIETPYRKVVDGVVSDEIVYLTADEEDRH 533
             YGR+CPIETPEGPNIGLI SLSVYA+ N +GF+ETPYRKV DGVV+DEI YL+A EE  +
Templ:  455 YGRVCPIETPEGPNIGLINSLSVYAQTNEYGFLETPYRKVTDGVVTDEIHYLSAIEEGNY 514


Query:  534 VVAQANSPIDADGRFVEPRVLVRRKAGEVEYVPSSEVDYMDVSPRQMVSVATAMIPFLEH 593
             V+AQANS +D +G FVE  V  R K GE        +VDYMDVS +Q+VSV  ++IPFLEH
Templ:  515 VIAQANSNLDEEGHFVEDLVTCRSK-GESSLFSRDQVDYMDVSTQQVVSVGASLIPFLEH 573


Query:  594 DDANRALMGANMQRQAVPLVRSEAPLVGTGMELRAAIDAGDVVVAEESGVIEEVSADYIT 653
             DDANRALMGANMQRQAVP +R++ PLVGTGME    A+D+G   VA+  GV++ V A  I
Templ:  574 DDANRALMGANMQRQAVPTLRADKPLVGTGMERAVAVDSGVTAVAKRGGVVQYVDASRIV 633


Query:  654 VMHDNGTRRT-------YRMRKFARSNHGTCANQCPIVDAGDRVEAGQVIADGPCTDDGE 706
             + +                Y +K+ RSN  TC NQ P V  G+ VE G V+ADGP TD GE
Templ:  634 IKVNEDEMYPGEAGIDIYNLTKYTRSNQNTCINQMPCVSLGEPVERGDVLADGPSTDLGE 693


Query:  707 MALGKNLLVAIMPWEGHNYEDAIILSNRLVEEDVLTSIHIEEHEIDARDTKLGAEEITRD 766
             +ALG+N+ VA MPW G+N+ED+I++S R+V+ED  T+IHI+E     +RDTKLG EEIT D
Templ:  694 LALGQNMRVAFMPWNGYNFEDSILVSERVVQEDRFTTIHIQELACVSRDTKLGPEEITAD 753


Query:  767 IPNISDEVLADLDERGIVRIGAEVRDGDILVGKVTPKGETELTPEERLLRAIFGEKAREV 826
             IPN+ +  L+ LDE GIV IGAEV  GDILVGKVTPKGET+LTPEE+LLRAIFGEKA +V
Templ:  754 IPNVGEAALSKLDESGIVYIGAEVTGGDILVGKVTPKGETQLTPEEKLLRAIFGEKASDV 813


Query:  827 RDTSLKVPHGESGKVIGIRVFSRED-EDELPAGVNELVRVYVAQKRKISDGDKLAGRHGN 885
             +D+SL+VP+G SG VI ++VF+R+  E +L GV ++V+VY+A KR+I  GDK+AGRHGN
Templ:  814 KDSSLRVPNGVSGTVIDVQVFTRDGVEKDLAPGVLKIVKVYLAVKRRIQPGDKMAGRHGN 873


Query:  886 KGVIGKILPVEDMPFLADGTPVDIILNTHGVPRRMNIGQILETHLGWCAHSGWKVDAAKG 945
             KGVI KI P+EDMP+  +GTPVDI+LN  GVP RMNIGQILETHLG         AAKG
Templ:  874 KGVISKINPIEDMPYDENGTPVDIVLNPLGVPSRMNIGQILETHLGM---------AAKG 924


Query:  946 VPDWAARLPDELLEAQPNAIVSTPVFDGAQEAELQGLLSCTLPNRDGDVLVDADGKAMLF 1005
             +          +P            ++TPVFDGA+EAE++ LL        GD+    G+ L+
Templ:  925 IG-----MP-----------IATPVFDGAKEAEIKELLKL------GDL--PTSGQIRLY 960


Query: 1006DGRSGEPFPYPVTVGYMYIMKLHHLVDDKIHARSTGPYSMITQQPLGGKAQFGGQRFGEM 1065
             DGR+GE F  PVTVGYMY++KL+HLVDDK+HARSTG YS++TQQPLGGKAQFGGQRFGEM
Templ:  961DGRTGEQFERPVTVGYMYMLKLNHLVDDKMHARSTGSYSLVTQQPLGGKAQFGGQRFGEM 1020


Query: 1066ECWAMQAYGAAYTLQELLTIKSDDTVGRVKVYEAIVKGENIPEPGIPESFKVLLKELQSL 1125
             E WA++AYGAAYTLQE+LT+KSDD  GR K+Y+ IV G +  EPG+PESF VLLKE++SL
Templ: 1021EVWALEAYGAAYTLQEMLTVKSDDVNGRTKMYKNIVDGNHQMEPGMPESFNVLLKEIRSL 1080


Query: 1126CLNVEV 1131
             +N+E+
Templ: 1081GINIEL 1086


Modeling ...

Summary: Query= gi_15607807_ref_NP_215181.1_ Template= 3IYD.C Id=  53.5 Qlen= 1172
Model_len= 1100 Coverage=  93.9 Q_Mw= 129235 Model_Mw= 121855 Method= 'PDB Blast' E-
value= 0.0
```

Table 3.1 CPHmodels 3.2 Server Alligment Results
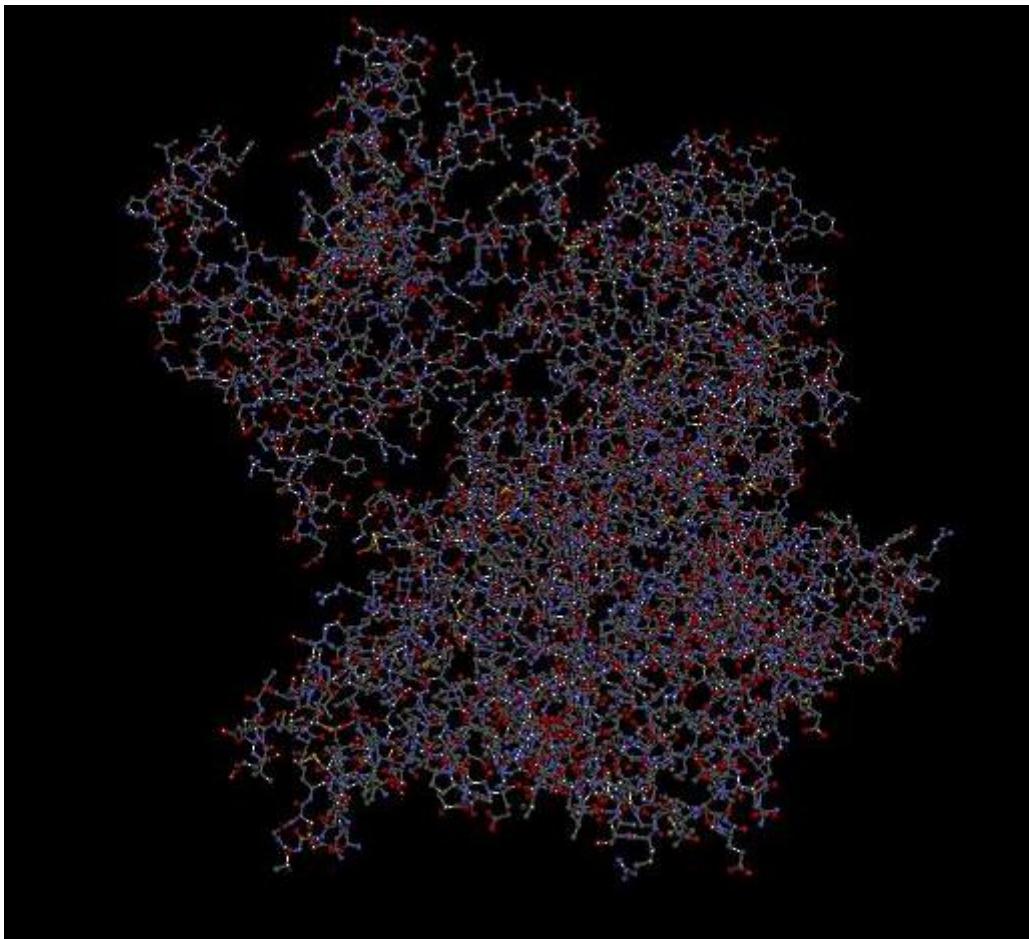
Figure 3.4 CPHmodels 3.2 Server rpoβ model atomic view
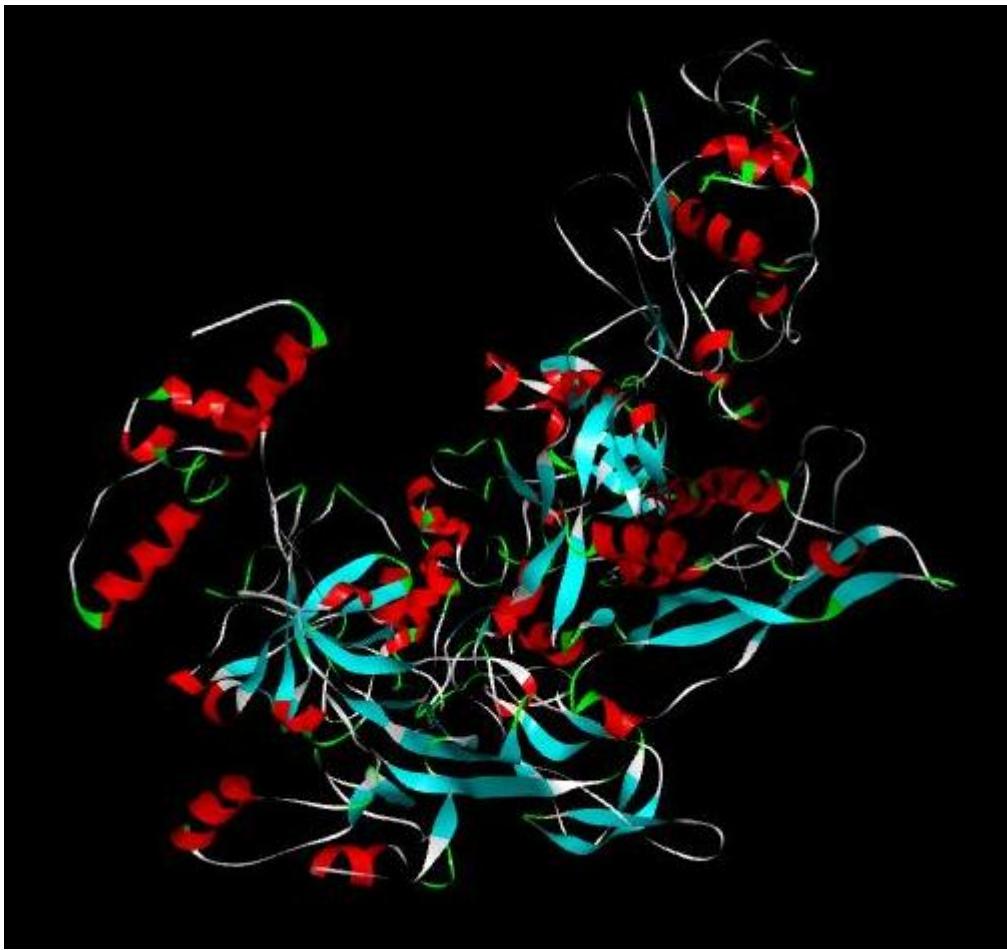
Figure 3.5 CPHmodels 3.2 Server Protein Chain View

3.2 Protein structure prediction on the Web: a case study using the Phyre server

Phyre server is an Automated homology modeling program using neural networks.

Query gi|15607807|ref|NP_215181.1| (seq=MADSRQSKTA...NESASVEDLA Len=1172 Neff=5.9 Nseqs=213)
Parameters score SS:yes search:local realign with MAP:no

```
   No Hit                            Prob E-value P-value  Score    SS Cols Query HMM
Template HMM
1 2a6h_C DNA-directed RNA polyme 100.0  5E-272  2E-276 2556.4  60.2 1059   31-1145    2-
1115(1119)
```

13

```
  2 3lu0_C DNA-directed RNA polyme 100.0  1E-268  4E-273 2551.1  45.5 1065   29-1134    9-
1342(1342)
  3 2waq_B DNA-directed RNA polyme 100.0  6E-221  2E-225 2091.2  49.5  940   31-1134   10-
1119(1131)
  4 1twf_B DNA-directed RNA polyme 100.0  7E-219  3E-223 2081.0  55.3  940   35-1136   31-
1221(1224)
  5 3h0g_B DNA-directed RNA polyme 100.0  3E-215  1E-219 2048.7  13.9  941   35-1137   18-
1210(1210)
  6 3mlq_A DNA-directed RNA polyme 100.0 3.6E-40 1.3E-44  345.1  19.6  184   46-434     2-
188 (188)
  7 3lti_A DNA-directed RNA polyme 100.0 1.4E-28 5.3E-33  274.2  19.2  177  172-362     1-
296 (296)
  8 3tbi_B DNA-directed RNA polyme  99.8 2.8E-20   1E-24  199.2  13.2  128  743-870     1-
228 (228)
  9 3qqc_A DNA-directed RNA polyme  97.2  0.0001 3.8E-09   86.7   3.2   57 1097-1153   25-
86  (436)
 10 2lmc_B DNA-directed RNA polyme  92.2    0.12 4.4E-06   47.1   4.5   57  633-699
23-84   (84)
 11 2xha_A NUSG, transcription ant  81.9    0.95 3.5E-05   47.2   4.2   57  633-699
100-159 (193)
 12 2xhc_A Transcription antitermi  75.4     1.7 6.3E-05   49.4   4.2   58  633-700
140-200 (352)
 13 3it5_A Protease LASA; metallop  60.0     6.9 0.00025   40.4   4.5   53  633-697
48-100 (182)
 14 2xha_A NUSG, transcription ant  56.0 1.4E+02   0.005   31.1  13.4  142  636-873
43-191 (193)
 15 1ax3_A Iiaglc, glucose permeas  49.2      17 0.00063   36.7   5.3   82  632-718
48-132 (162)
 16 1f3z_A EIIA-GLC, glucose-speci  37.5      37  0.0014   34.2   5.6   65  632-698
48-114 (161)
 17 3our_B EIIA, phosphotransferas  34.5      44  0.0016   34.4   5.6   65  632-698
70-136 (183)
 18 2gpr_A Glucose-permease IIA co  34.2      42  0.0015   33.5   5.3   65  632-698
43-109 (154)
 19 3nyy_A Putative glycyl-glycine  30.3      44  0.0016   36.0   5.1   58  627-697
129-197 (252)
 20 2xhc_A Transcription antitermi  30.0    2E+02  0.0075   32.3  10.7  109  636-797
83-197 (352)
 21 2hsi_A Putative peptidase M23;  25.7      66  0.0024   35.2   5.5   77  627-718
184-265 (282)
 22 2gu1_A Zinc peptidase; alpha/b  20.8 1.1E+02  0.0042   34.2   6.4   77  627-718
236-317 (361)
```

2a6h_C DNA-directed RNA polymerase beta chain; RNA polymerase holoenzyme,
streptolydigin, antibiotic,
  transcription regulation; HET: STD; 2.40A {Thermus thermophilus} SCOP: e.29.1.1 PDB:
1smy _C* 1zyr _C* 1iw7 _C* 2a69 _C*
  2a6e _C 2a68 _C* 2be5 _C* 2cw0 _C 2o5i _C 2o5j _C* 2ppb _C* 3aoh _C* 3aoi _C*
3dxj _C* 3eql _C* 1ynj _C* 1ynn _C* 2gho _C 1hqm _C
  1l9u _C ...
  Probab=100.00 E-value=5.5e-272 Score=2556.40 Aligned_cols=1059 Identities=53%
Similarity=0.908 Sum_probs=0.0

```
  Q ss_pred            eeeehhcccccccCCCCHHHHHHHHHHHHHHhCcccccccccccccccchhHHHHHHhcCCEEe--
-cCCcEEEEEEEEEEc
  Q gi|15607807|re   31 NRVSFAKLREPLEVPGLLDVQTDSFEWLIGSPRWRESAAERGDVNPVGGLEEVLYELSPIED--
-FSGSMSLSFSDPRFD  107 (1172)
  Q Consensus        31 ~r~~~~~~i~~~~~~p~Lv~~qi~SFn~Fl~~~~~~~~~~~~~~~~~~~GL~~ii~~~~pI~~~-
-~~~~~~l~f~~i~i~  107 (1172)
                        +|++|++++++|++|+|+++|++|||+|||.+++      |++|+++||++++++++||++
.+++++|+|++++|+
  T Consensus        2 ~r~~~~~~~~~~~~~Lv~~qi~SFn~Fl~~~~~-------
~~~~~~~GL~~i~~~~~pI~~~~~~~~~~L~f~~i~i~   74 (1119)
  T 2a6h_C           2 EIKRFGRIREVIPLPPLTEIQVESYRRALQADVP-------
PEKRENVGIQAAFRETFPIEEEDKGKGGLVLDFLEYRLG   74 (1119)
  T ss_dssp          EEEECCCCCCCSCCCCTTHHHHHHHHHHSCTTSC-------
TTSSCCCHHHHHHHHHHCSEEECCSSSCCEEEEECCCCBC
  T ss_pred          cceecccccccccCcCHHHHHHHHHHHHHHccCCc-------
ccccHHHHHHHHHHhcCCEeccCCCCCeEEEEEEEEEEc
```

14

```
   Q ss_pred
CCCCCHHHHHhcCCccCccEEEEEEEEEECCCceeeeEEEEEecCCEECCCcEEEEeCEEEEEEEEEecCCcEEEEecccc
  Q gi|15607807|re  108
DVKAPVDECKDKDMTYAAPLFVTAEFINNNTGEIKSQTVFMGDFPMMTEKGTFIINGTERVVVSQLVRSPGVYFDETIDK  187
(1172)
  Q Consensus      108
~P~~~P~EcR~r~lTYsapl~v~v~~~~~~~~~~~~~~v~iG~iPIM~~GGYFIING~ERVII~Q~~~sp~~~~~~~~~k  187
(1172)


+|+++|+|||+|++||+|||+|++++..++++++++++|++|+|||||+|||||||||+|||||+|++++||++++.++++
  T Consensus       75
~P~~~P~EcR~r~lTYsa~L~v~v~~~~~~~~~~i~~~~v~lG~IPIMv~GGYFIING~ERVII~Q~~~sp~~~~~~~~~k  154
(1119)
  T 2a6h_C          75
EPPFPQDECREKDLTYQAPLYARLQLIHKDTGLIKEDEVFLGHIPLMTEDGSFIINGADRVIVSQIHRSPGVYFTPDPAR  154
(1119)
  T ss_dssp
CCSSCHHHHHHTTCCCEEEBCCCEEECCSSSCCEECCCCCCCEEECCCTTSCCCSSSSCEEECEEEEECSCEEEECCSSC
  T ss_pred
CCCCCHHHHHhcCCccccceEEEEEEEEECCCceeeeeEEEecCCCeECCCcEEEEeCeEEEEEEEEecCCcEEEEeeccC


   Q ss_pred
cCCceEEEEEEEEecceEEEEEEeCCCeEEEEEcccCceehhhhHHHhCCCHHHHHHHhcCCHHHHHHHHhcc--ccCHH
  Q gi|15607807|re  188
STDKTLHSVKVIPSRGAWLEFDVDKRDTVGVRIDRKRRQPVTVLLKALGWTSEQIVERFGFSEIMRSTLEKDN--TVGTD  265
(1172)
  Q Consensus      188
~~~~~~~t~~ii~~rg~~~~l~~~~~~~i~v~i~~~~~IPl~ilLkALG~sD~eI~~~i~~~~~~~~~l~~~~-~~~~~  265
(1172)
                      .++..|+++++|.||+|+.+++++++.+|+++++
.+||+++|||||||+||+||++.|+.++.+.++|.++.  ..+++
  T Consensus      155 -~g~~~yt~~ii~~rg~~l~~~~d~~~~i~vri~k-
~~IPi~ilLkALG~sD~eI~~~i~~~~~~~~~l~~~~~~~~~~  232 (1119)
  T 2a6h_C         155 -PGRYIASIIPLPKRGPWIDLEVEPNGVVSMKVNK-
RKFPLVLLLRVLGYDQETLARELGAYGELVQGLMDESVFAMRPE  232 (1119)
  T ss_dssp                -SSCCEEEECCCSSSSCCEEEEEETTTEEEEESSS-
SEECHHHHHTHHHTCCHHHHHHHHHHSSCTTHHHHSSCHHHHTCHH
  T ss_pred                -CCceEEEEEEEeccCceEEEEEcCCCEEEEEEEeC-
cceeHHHHHHHHCCCHHHHHHHHhccCHHHHHHHHHhhhcccCCHH


   Q ss_pred               HHHHHHHHhcCCCCCCHHHHHHHHHHHHhCCccccchhhcChhhhhhhcCccccccc-----
cccccccHHHHHHHHHHH
  Q gi|15607807|re  266 EALLDIYRKLRPGEPPTKESAQTLLENLFFKEKRYDLARVGRYKVNKKLGLHVGEP------
ITSSTLTEEDVVATIEYL  339 (1172)
  Q Consensus      266 ~AL~~i~~~l~~~~~~~~~~a~~~L~~~f~~~~~y~l~~vGr~~ln~kl~l~~~~~------
~~k~~l~~~dl~~mi~kL  339 (1172)
                            +||+++++++++++..+.+.|+++|++.||+|+||+|+|+|++|+++|+++|++.  +
...++|+++||++|++||
  T Consensus      233 ~al~~i~k~l~~~~~~~~~~a~~~L~~~f~~p~~y~l~~vGr~~ln~~l~l~~~--
~~~~~~~~k~~~l~~~dl~~mi~kL  310 (1119)
  T 2a6h_C         233 EALIRLFTLLRPGDPPKRDKAVAYVYGLIADPRRYDLGEAGRYKAEEKLGIRL--
SGRTLARFEDGEFKDEVFLPTLRYL  310 (1119)
  T ss_dssp              HHHHHHHHHSSSCCCSCCCSSHHHHHTSSSSSSSCCSCTTTSSHHHHHTTSCSCC--
STTTTCCCSSSCCCCTTHHHHHHHH
  T ss_pred              HHHHHHHHHhcCCCCCcHHHHHHHHHHHhhCCcccccchhcchhhhhhhhccCC--
ccccccccccccccHHHHHHHHHH


   Q ss_pred
HHHhcCCcccccccccccccccccccchhhccEeccHHHHHHHHHHHHHHHHHHHHHHHHHHHhhhccccccCHHHhccCccHHH
  Q gi|15607807|re  340
VRLHEGQTTMTVPGGVEVPVETDDIDHFGNRRLRTVGELIQNQIRVGMSRMERVVRERMTTQDVEAITPQTLINIRPVVA  419
(1172)
  Q Consensus      340
l~l~~g~~~~~~~~~~~~~~~~~DD~Dhl~NKRv~l~GeLl~~~fr~~l~r~~~~i~~~l~~~~~~~~~~~~li~~~~It~  419
(1172)
                         +++..|.+
.+.+||+|||+|||||+++|+||+.+||.+|+++++.+++++.+.+.+.+++..+++++.||+
  T Consensus      311 l~l~~g~~-----------
~~~~~DD~D~l~NkRv~l~G~Ll~~~fr~~l~~~~~~i~~~l~~~~~~~~~~~~~~~~~It~   380 (1119)
```

```
    T 2a6h_C         311 FALTAGVP----------
GHEVDDIDHLGNRRIRTVGELMTDQFRVGLARLARGVRERMLMGSEDSLTPAKLVNSRPLEA  380 (1119)
    T ss_dssp            HHHHTSCS----------
SCCCCCTTSTTTEEEECHHHHHHHHHHHHHHHHHHHHHHHHHHSCSSCCSSTTTCCSHHHHH
    T ss_pred           HHhhcCCC----------
CCCCcCcchhhceEehhhhHHHHHHHHHHHHHHHHHHHhhhcccccchHHhhccchhHH


    Q ss_pred
HHHHHhccCCcccccccccchhHhhhhhhheecccCCcccccccccccCccccccChhhCceecccccCCCCccccchhhhhhhhcc
    Q gi|15607807|re  420
AIKEFFGTSQLSQFMDQNNPLSGLTHKRRLSALGPGGLSRERAGLEVRDVHPSHYGRMCPIETPEGPNIGLIGSLSVYAR  499
(1172)
    Q Consensus       420
~i~~ff~tg~lSQ~Ldq~N~Ls~LSH~RRiss~gpGgl~re~k~~~vR~LHpShwGrICPiETPEG~ncGLVknLA~~a~  499
(1172)


+|++||+||+||+|||+|||+||||+|||+|||+||+||||+++|||+++++|+++||+|||||||||+||+|||||+||||||||||||++|+
    T Consensus       381
~i~~ff~Tg~lsQ~Ldq~N~ls~lsh~Rrv~~~~~g~~~r~~k~~~vR~Lhps~wGriCPveTPEG~~cGLv~~LA~~a~  460
(1119)
    T 2a6h_C         381
AIREFFSRSQLSQFKDETNPLSSLRHKRRISALGPGGLTRERAGFDVRDVHRTHYGRICPVETPEGANIGLITSLAAYAR  460
(1119)
    T ss_dssp
HHHHHHHTCSSEEECCCSSTHHHHHHHHEEESSSSSSSCCSSCCHHHHSCCGGGTTTBCSSCSCSSSSCSSEEEBCSSCE
    T ss_pred
HHHHHhhccCcccccccccchHhhhhhhheecccCCCcccccccccCcccccCHhhCcccCCCcCCCCCccccchhhhhhhee


    Q ss_pred
hhhcCCccCceEEEECCEEeeeEEEcccceEeeeEEeccccccccccCCcccCCceEEEecCCceEeecccceeEEEEeccccc
    Q gi|15607807|re  500
VNPFGFIETPYRKVVDGVVSDEIVYLTADEEDRHVVAQANSPIDADGRFVEPRVLVRRKAGEVEYVPSSEVDYMDVSPRQ  579
(1172)
    Q Consensus       500
I~~~G~~~~p~~~v~nG~i~~~i~~l~~~~e~~~~ia~~~~~~~~~g~~~rp~i~~r~~~~~~~~~~~~~~~h~ei~p~~  579
(1172)
                                 |+.+|++++||.+|+||++++.++|+++++|+.+.|++.++.+++ ||++||++.+|+.
+++..+.++++||+||+|.+
    T Consensus       461 I~~~g~~~~~~~~v~nG~v~~~i~~l~~~~~~~~~i~~~~~~~~~-gr~~rp~~~~~~~~
~e~~~~~~~~i~~~ei~p~~  538 (1119)
    T 2a6h_C         461 VDELGFIRTPYRRVVGGVVTDEVVYMTATEEDRYTIAQANTPLEG-NRIAAERVVARRK-
GEPVIVSPEEVEFMDVSPKQ  538 (1119)
    T ss_dssp            ECSSSCEEEEEEEEEETTEEEEEEEEEECHHHHHHSCEECTTSCBSS-SBBCCSSEEEEESS-
SSEEEECTTTCCEEECCTTT
    T ss_pred            eccccCcccCCcEEEECCEEeeeeEEeChhcccceEEecCceeecC-CccccCceeEEec-
cceeeechhHeEEEEecccc


    Q ss_pred
eeccccccccCCcccCCcccchhhhhhhhhccccccccccCCeeecCcccceeecccCceeEeccCCEEEEecCcEEEEEecCC
    Q gi|15607807|re  580
MVSVATAMIPFLEHDDANRALMGANMQRQAVPLVRSEAPLVGTGMELRAAIDAGDVVVAEESGVIEEVSADYITVMHDNG  659
(1172)
    Q Consensus       580
ilsv~aslIPF~~hNds~R~l~~s~M~kQAv~l~~~~~p~VgTg~e~~~~~~s~~~~~a~~~G~v~~vd~~~i~~r~d~~  659
(1172)


|||++||||||+||||||||||||||||||||||||++.+|+|+||||+|+++++||+++++|+++|+|.|||+++|.+|+|++
    T Consensus       539
ilSv~aslIPF~eHNdspR~l~~s~M~KQAvgl~~~~~p~vgTg~e~~~~~~~~~~~a~~~g~v~~v~~~~i~~r~~~~  618
(1119)
    T 2a6h_C         539
VFSVNTNLIPFLEHDDANRALMGSNMQTQAVPLIRAQAPVVMTGLEERVVRDSLAALYAEEDGEVAKVDGNRIVVRYEDG  618
(1119)
    T ss_dssp
TSCHHHHTCTTGGGBCHHHHHHHHHHHHTTBCCBSSCCCCSEECSCHHHHHHHTTCSEECSSSEEEEEECSSBEEEEETTT
    T ss_pred
cccccccccCCCcccCCCcccccccccccccccccccccCCCeeeccccceeeeccCceeEeccCCEEEEecCceEEEEccCC
```

```
   Q ss_pred
CcceEEeeccccccccccCCcCceEecCceeeccceecccccccccccccCceeeEEEEecccCChhhhhhhhhhhhhhcC
   Q gi|15607807|re   660
TRRTYRMRKFARSNHGTCANQCPIVDAGDRVEAGQVIADGPCTDDGEMALGKNLLVAIMPWEGHNYEDAIILSNRLVEED  739
(1172)
   Q Consensus       660
~~~~y~L~~~~~snq~~~~~QkPiV~t~~~v~~g~~~ad~~~~~~~el~~G~N~~VAvm~y~GYn~EDaiiink~~i~rg  739
(1172)

+.+.|.|.+|++|||+|+|||+|||+++++|++||+|||++++.++|+|+|+|++|||||||+||||||||||||++++||
   T Consensus       619
~~~~y~L~~~~~snq~~~~~q~PlV~~~~~~~~~~~la~~~~~~~~~e~~~G~N~~VA~~~~~GYn~EDaiiin~~~v~rg  698
(1119)
   T 2a6h_C          619
RLVEYPLRRFYRSNQGTALDQRPRVVVGQRVRKGDLLADGPASENGFLALGQNVLVAIMPFDGYNFEDAIVISEELLKRD  698
(1119)
   T ss_dssp
EEECCBCCCSEECTTSCEECCEECCCSSCCBCTTCEEEECTTBSSSSBCCSEEEEEECSCCSSTTSSSEEEEETHHHHTT
   T ss_pred
cceeEEeeccccccccccccccceEeeCceeeccceecccccCCCccccCceeEEEEEecccCCcchhhhhhhhHHhcC


   Q ss_pred
CceEEEEEEEEeeeeecCCCCceeeeccCCCCchhhhhccCCCcCcCCCCEECCCCEEEEEecCCCcccCChhhhhhhhhh
   Q gi|15607807|re   740
VLTSIHIEEHEIDARDTKLGAEEITRDIPNISDEVLADLDERGIVRIGAEVRDGDILVGKVTPKGETELTPEERLLRAIF  819
(1172)
   Q Consensus       740
~~~s~~~~~~~~~~~~~~g~e~~~~~~p~~~~~~~~~LD~dGii~iG~~v~~gDilvgk~~p~~~~~~~~~~~l~~~if  819
(1172)

+|||+|+++|+++.+.+++|.|++|+++|++++..+++||+||+|++|++|++|||||||++|+.+.+.++++++++++|
   T Consensus       699
~~~s~~~~~~~~~~~~~~~e~~t~~~p~~~~~~~~~Ld~~Gi~~~g~~v~~gdiligk~~p~~~~~~~~~~~l~~~~~  778
(1119)
   T 2a6h_C          699
FYTSIHIERYEIEARDTKLGPERITRDIPHLSEAALRDLDEEGVVRIGAEVKPGDILVGRTSFKGESEPTPEERLLRSIF  778
(1119)
   T ss_dssp
CSEEEEEEEEEEEEEECCTTCCCBCCSCCSSSCSGGGSSCCSSSBCCTTCBCCTTSEEECCEEESSSSSCCHHHHHHHHHH
   T ss_pred
CceEEEEEEEEeeeeecCCCceeeeccCCCCchhhhhccCCCccccCCCCEECCCCEEEEEecCCccccCChhhhhhhhhh


   Q ss_pred               cccCccceeEEEEecCCCcEEEEEEEEEeCCC-
CCccCCCcceEEEEEEcccCcccccChhhcccccCCceeeeeeccccC
   Q gi|15607807|re   820 GEKAREVRDTSLKVPHGESGKVIGIRVFSRED-
EDELPAGVNELVRVYVAQKRKISDGDKLAGRHGNKGVIGKILPVEDM  898 (1172)
   Q Consensus       820 ~~~~~~~~d~s~~~~~~~~g~V~~v~~~~~~~-
~~~l~~~~~~~vkv~ir~~R~p~iGDKfssRHGqKGVis~i~~~eDM  898 (1172)
                             |++.+.++|+|+++++++.|+|++|.++.+++
|+.+++++.+.|+|++|+.|+|++|||||||||||||||||+|||+|||
   T Consensus       779
~~~~~~~~d~s~~~~~~~~g~V~~v~~~~~~~~g~~~~~~~~~~v~v~i~~~r~~~iGDK~ssRHGqKGvvs~i~~~~Dm  858
(1119)
   T 2a6h_C          779
GEKARDVKDTSLRVPPGEGGIVVRTVRLRRGDPGVELKPGVREVVRVYVAQKRKLQVGDKLANRHGNKGVVAKILPVEDM  858
(1119)
   T ss_dssp
HSCCCCEEECCEECCSSCCCEEEEEEEEECSSCSSCCCCTTEEEEEEEEEEEEEECCTTCEEECTTSCEEEEEEEECTTTS
   T ss_pred
cccccccceeeEEccCCCccEEEEEEEEEccCCCCCcccCCCCceEEEEEEcccCCcccCchhhhccCCCceEEeeccCCC


   Q ss_pred               CcCCCCCCCCcEEECCCCCccccccchhHHHHhhhhhhC-
CceeccchhhhhhhhhHHHHHhccCCCccccccCCCCCcHH
   Q gi|15607807|re   899 PFLADGTPVDIILNTHGVPRRMNIGQILETHLGWCAHS-
GWKVDAAKGVPDWAARLPDELLEAQPNAIVSTPVFDGAQEA  977 (1172)
   Q Consensus       899 Pf~~dG~~pDIIiNPhg~PSRMtIGqllE~~~Gka~~~-
g~~~d~~~~~~~~~~~~~~~~~~~~~~~tp~F~~~~~~  977 (1172)
                             ||++||++||||||||||||||||||||||+|+++||+|++ |.++
+||+|++.+.+
```

```
  T Consensus      859 Pf~~dG~~pDiI~NP~g~PSRMtiGql~E~~~gk~~~~~g~~~~~~~~~~~~~~~~~~~~~~
-----~tp~F~~~~~~  912 (1119)
  T 2a6h_C         859 PHLPDGTPVDVILNPLGVPSRMNLGQILETHLGLAGYFLGQRY------------------
-----ISPIFDGAKEP  912 (1119)
  T ss_dssp            CBCSSSCBCSEECCSTTTTTTTBTHHHHHHHHHHHHHHHHTTEEE-----------------
-----ECCTTTSCCHH
  T ss_pred            ccccCCCCccEEeCCCcCcccchhhhhHHHhhhHHhhcCCce-------------------
-----EecccCCCCHH


  Q ss pred            HHHHHHHhhhHHHc--------------------------CCccc-------------
CCCCceEeecCCCCCEece
  Q gi|15607807|re  978 ELQGLLSCTLPNRD--------------------------GDVLV-------------
DADGKAMLFDGRSGEPFPY 1015 (1172)
  Q Consensus       978 ~i~~~L~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~g~~~~~~~~~~~~~~
~~~Gke~lydG~TG~~~~~ 1015 (1172)
                         ++++.|       .++                               |   ||
+++|+++||||+||++|++
  T Consensus       913 ~i~~~L-----~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~g--
~~~~~~~~~~~~~~~~~G~~~lydG~TG~~~~~   985 (1119)
  T 2a6h_C          913 EIKELL-----AQAFEVYFGKRKGEGFGVDKREVEVLRRAEKLG--
LVTPGKTPEEQLKELFLQGKVVLYDGRTGEPIEG   985 (1119)
  T ss_dssp             HHHHHH-----HHHHHHHHHHHHHHHTCCCBHHHHHHHHHHHHHTTT--
SSCTTSCHHHHHHHHHHHTTEECCBCSSSCCBCSS
  T ss_pred             HHHHHH-----HHhhhhccccccccccccccchhhhhhhhhhhcCC--
ccccccchhhhcccCCCCceEeecCCCCCCccc


  Q ss_pred
eEEEehHHhhcccccccccCceEeeccccceeeecCCCCccccCCCeeeehhhhhhHHhCCHHHHHHHHHHhcCCcceeeeee
  Q gi|15607807|re 1016
PVTVGYMYIMKLHHLVDDKIHARSTGPYSMITQQPLGGKAQFGGQRFGEMECWAMQAYGAAYTLQELLTIKSDDTVGRVK 1095
(1172)
  Q Consensus      1016
~If~G~~Yy~kL~HmV~DKihaRs~Gp~~~lTrQP~~Gr~r~GG~RfGEMErdaL~a~GAs~~L~ErL~~~SD~~~gr~~ 1095
(1172)


+||+|++|||||+|||+||+|||++|||+.|||||++||||+|||||||||||||||+|||||++|+|||+++||+++||.+
  T Consensus       986
~i~~G~~yy~kL~HmV~DK~h~Rs~Gp~~~lT~QP~~Gr~~~GG~RfGEME~daL~a~Gaa~~L~E~l~~~SD~~~~~~~ 1065
(1119)
  T 2a6h_C          986
PIVVGQMFIMKLYHMVEDKMHARSTGPYSLITQQPLGGKAQFGGQRFGEMEVWALEAYGAAHTLQEMLTLKSDDIEGRNA 1065
(1119)
  T ss_dssp
CEEEEEEEEEEEBCCCTTTSCEEEESSCCBCSSSCSBCCCSSSCCCEEECHHHHHHHHHTTCSHHHHHHHTTTTTCHHHHHH
  T ss_pred
eEEEEhHHhhcchhhcccCcEEeeccccceeecCCCccccccCCCeeeehhehhhhhhccHHHHHHHHhccCCcccccccc


  Q ss_pred            eEEEeecCCccCccCCCHHHHHHHHHHHHhCCCceEEEeCCCCeecccccc
  Q gi|15607807|re 1096 VYEAIVKGENIPEPGIPESFKVLLKELQSLCLNVEVLSSDGAAIELREGE 1145 (1172)
  Q Consensus      1096 ~~~~~~~~~~~~~~~iPysfKlL~~EL~sm~i~~~~~~~~~~~~~~~~~~~ 1145 (1172)
                          .||++|++.+++++.+||||||||++||+||||++++.++++.++|+.++.
  T Consensus      1066 ~~~~~~~~~~~~~~~ip~sfk~L~~EL~sm~i~~~~~~~~~~~~~~~~~~~ 1115 (1119)
  T 2a6h_C         1066 AYEAIIKGEDVPEPSVPESFRVLVKELQALALDVQTLDEKDNPVDIFEGL 1115 (1119)
  T ss_dssp             HHHHHHTTCCCCCCCCCCHHHHHHHHHHHHHSSCEECCBCSSSCBCCSSCSS
  T ss_pred             eEEEeecCCccCccCCCHHHHHHHHHHHHhCCCceEEEecCCceeehhhhh
```

Table 3.2 Alignment result of Phyre Server

18

Figure 3.6 Amino Acid chain base viewing by Phyre Server

Figure 3.7 Atomic base viewing by Phyre Server

3.3 Docking result for rpoβ

We know that experimental docking energy is −13.3 kcal/mol.

We used autodock for docking process. We used it with that's configuration:

x-y-z coordinates: -41.936; 176.086; -3.12

maximum number of energy evaluations : 5000000,

num.grid points in xyz: 120 120 120.

Source	ΔG
Experimental	−13.3 kcal/mol.
swiss-model	−6.27 kcal/mol.
Phyre server	−12.64 kcal/mol.
CPHmodels 3.2 Server	−9.43 kcal/mol.

Table 3.3 Docking result for each modeling server

So Phyre server is given better result for docking. We used it for mutation resistance calculation.

Enzyme	AA	Mut	∆G	Ki
rpoβ			-12.64 kcal/mol	542.08 pM
rpoβ	S 456 (Ser)	L (Leu)	-11.44 kcal/mol	4.08 nM
rpoβ	D 441 (Asp)	V (Val)	-10.89 kcal/mol	12.10 nM
rpoβ	H 451 (His)	D (Asp)	-10.19 kcal/mol	1.16 nM
rpoβ	H 451 (His)	R (Arg)	-12.22 kcal/mol	1.11 nM
rpoβ	H 452 (His)	Y (Tyr)	-12.43 kcal/mol	552.11 pM
rpoβ	Q 438 (Gln)	K (Lys)	-10.84 kcal/mol	11.40 nM
rpoβ	S 447 (Ser)	Q (Gln)	-11.69 kcal/mol	3.87 nM
rpoβ	S 456 (Ser)	W (Trp)	-11.61 kcal/mol	3.07 nM

Table 3.4 Docking result for each mutation for Phyre Server Result
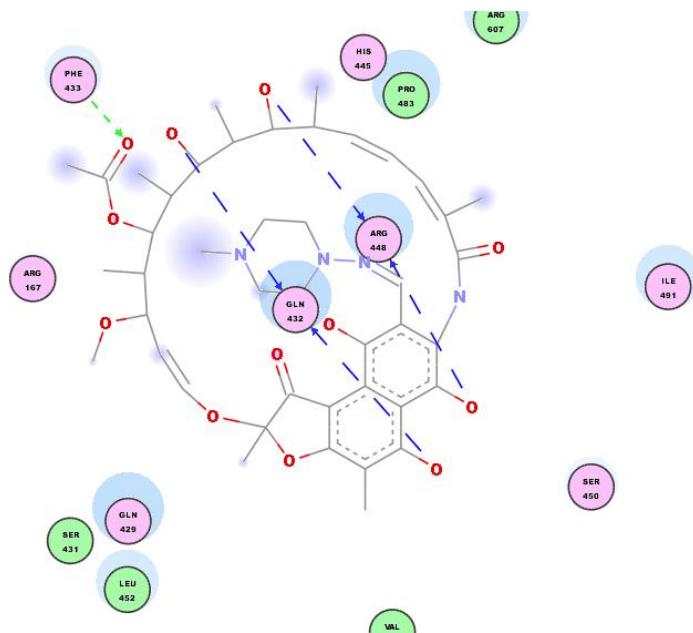
Figure 3.8 2D viewing RIF docking side with no mutation
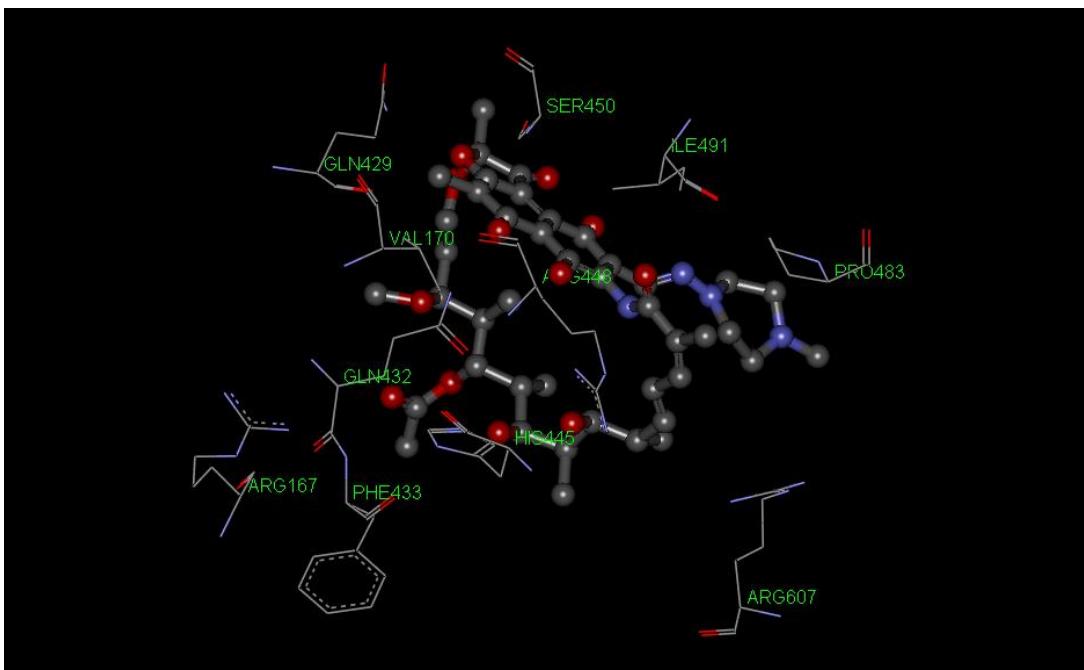
Figure 3.9 3D viewing RIF docking side with no mutation
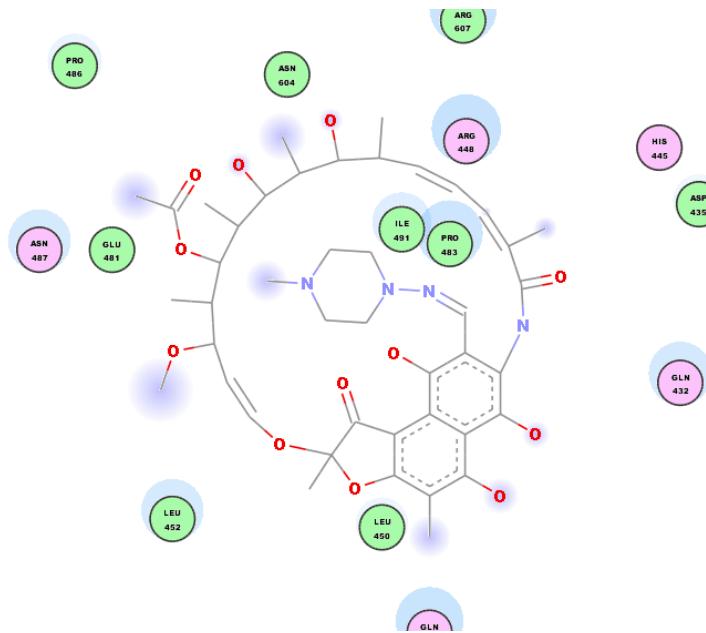
Figure 3.10 2D viewing RIF docking side with 456 S – L mutation
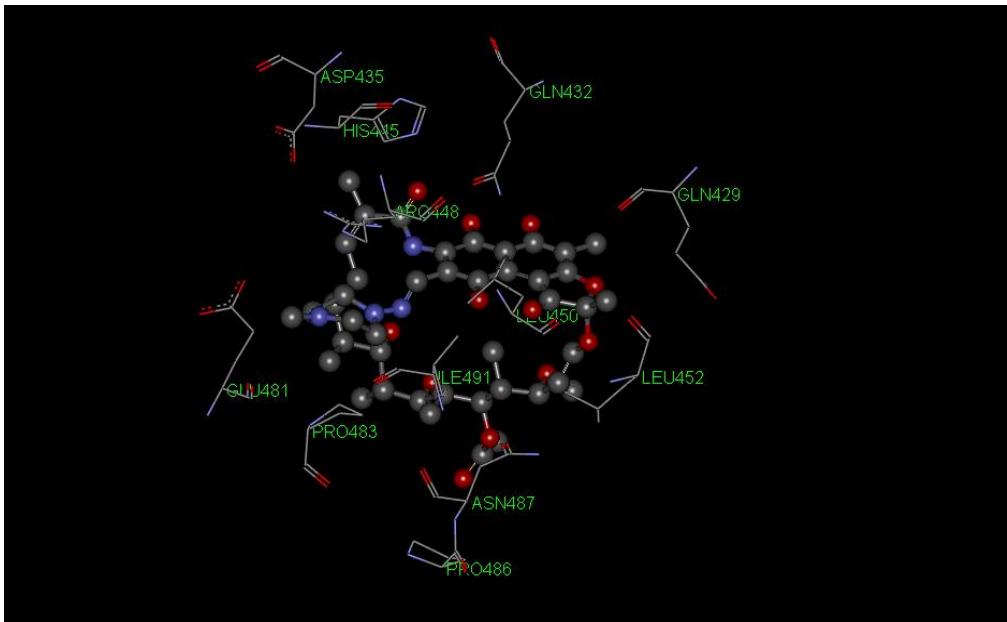
**Figure 3.11 3D viewing RIF docking side with 456 S – L mutation**
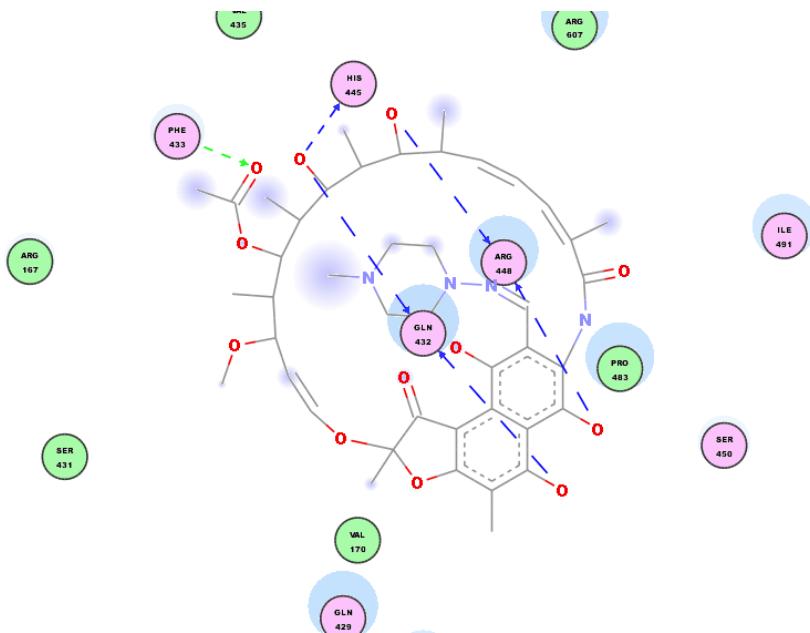
Figure 3.12 2D viewing RIF docking side with 441 D – V mutation
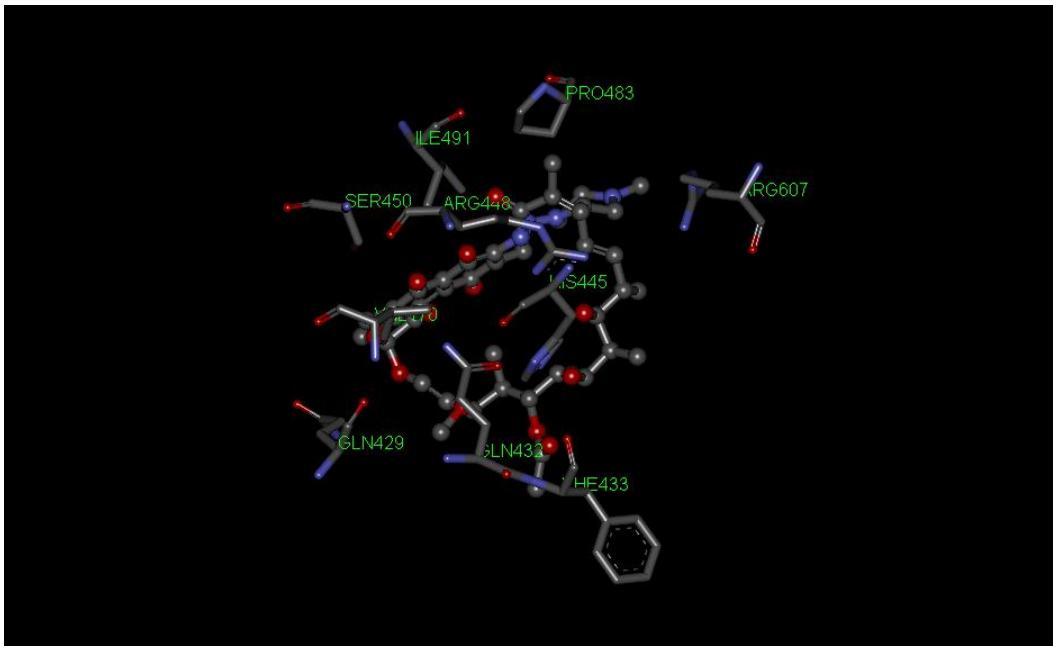
**Figure 3.13 3D viewing RIF docking side with 441 D – V mutation**
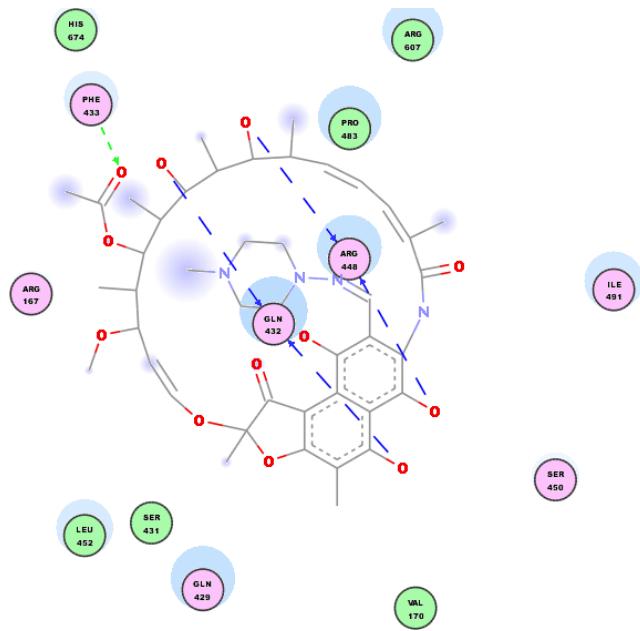
Figure 3.14 2D viewing RIF docking side with 451 H – D mutation
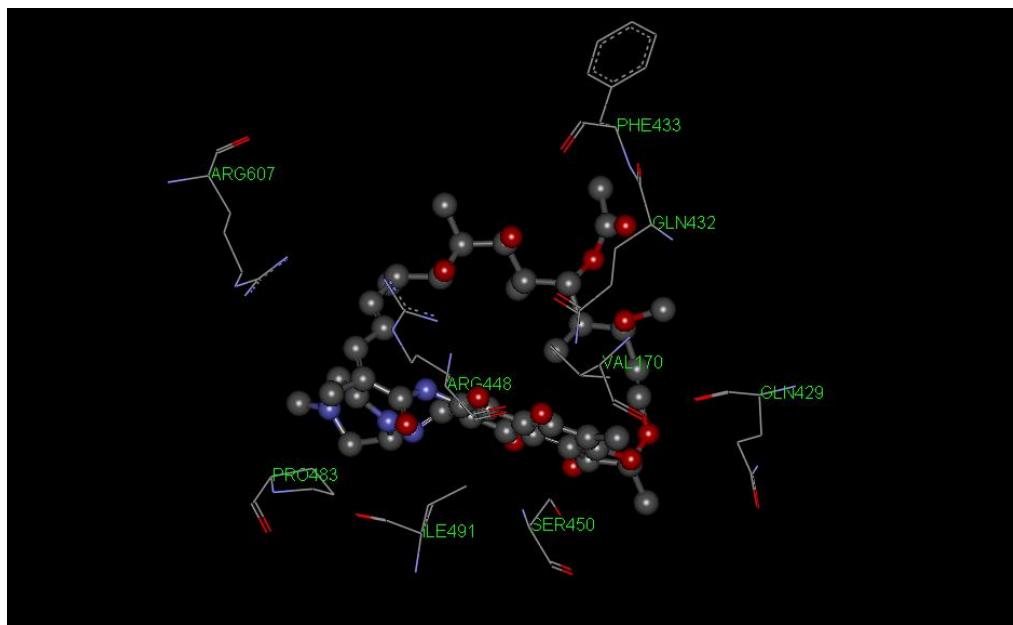
Figure 3.15 3D viewing RIF docking side with 451 H – D mutation
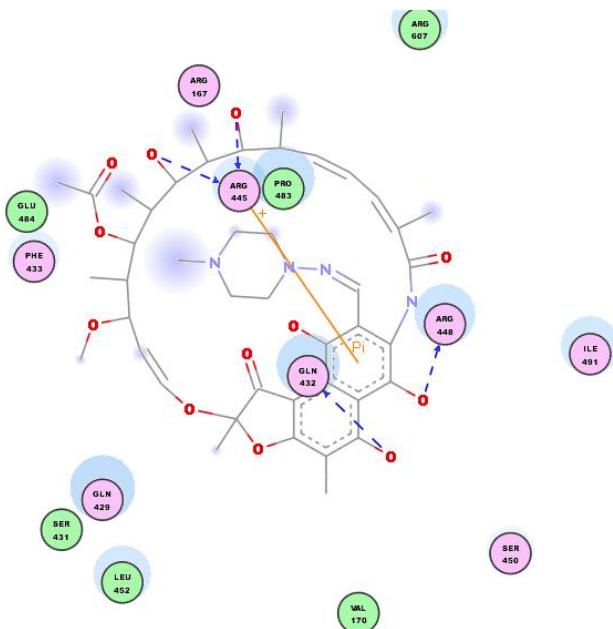
Figure 3.16 2D viewing RIF docking side with 451 H – R mutation
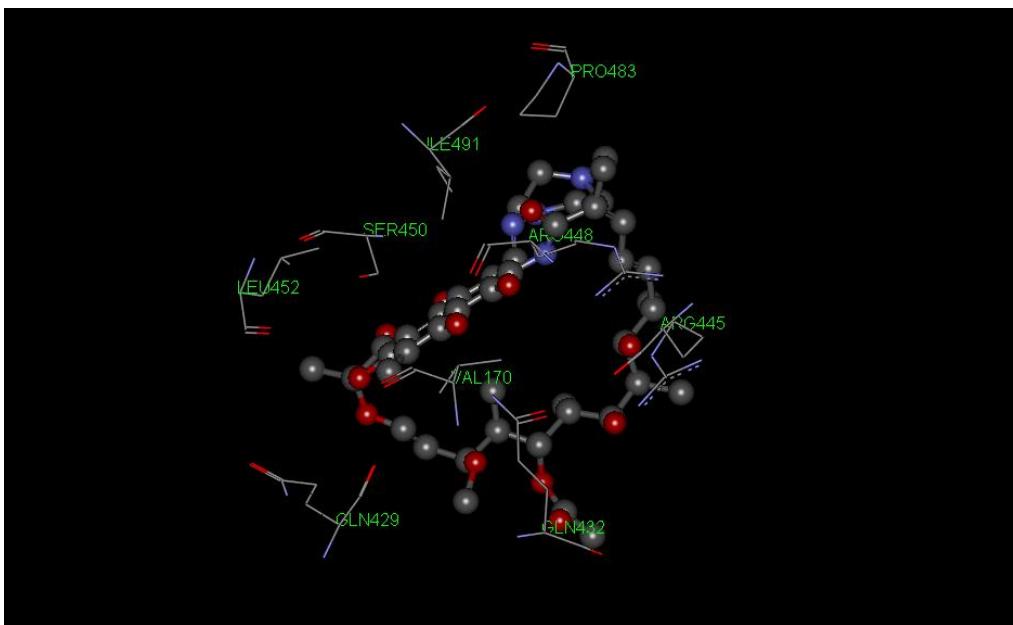
Figure 3.17 3D viewing RIF docking side with 451 H – R mutation
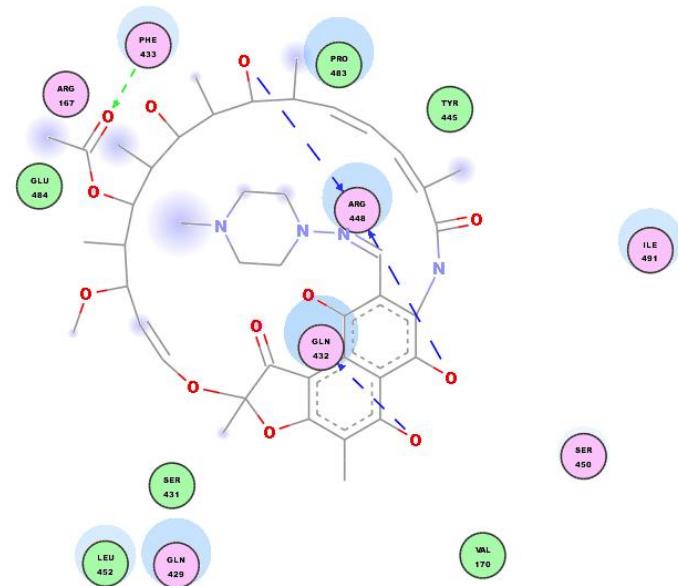
**Figure 3.18 2D viewing RIF docking side with 452 H – Y mutation**
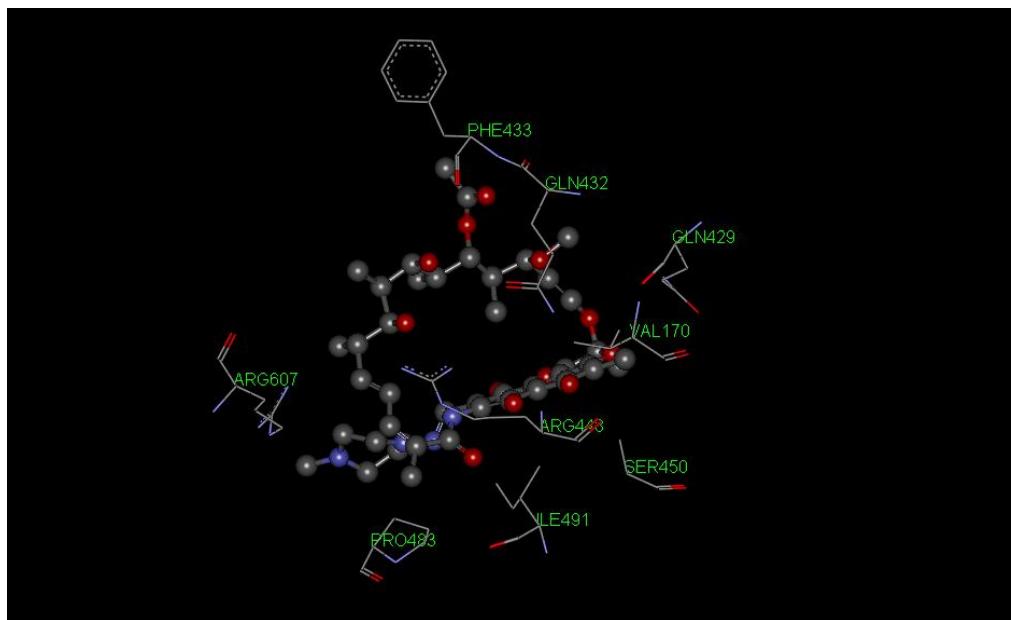
Figure 3.19 3D viewing RIF docking side with 452 H – Y mutation
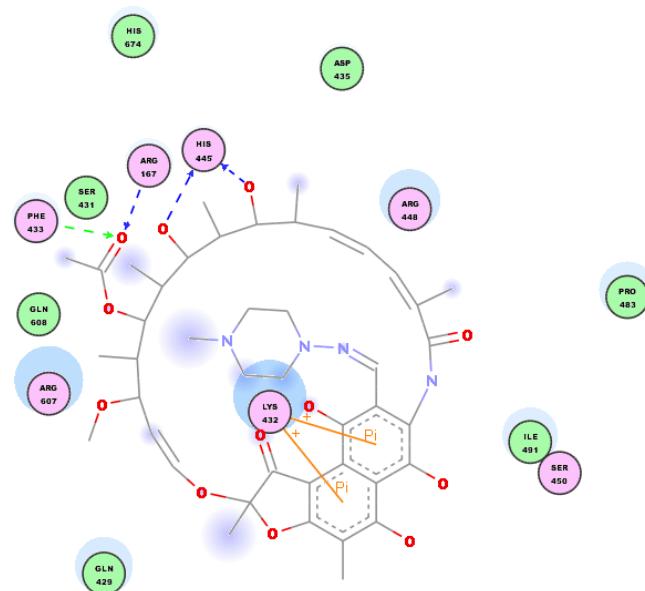
Figure 3.20 2D viewing RIF docking side with 438 Q – K mutation
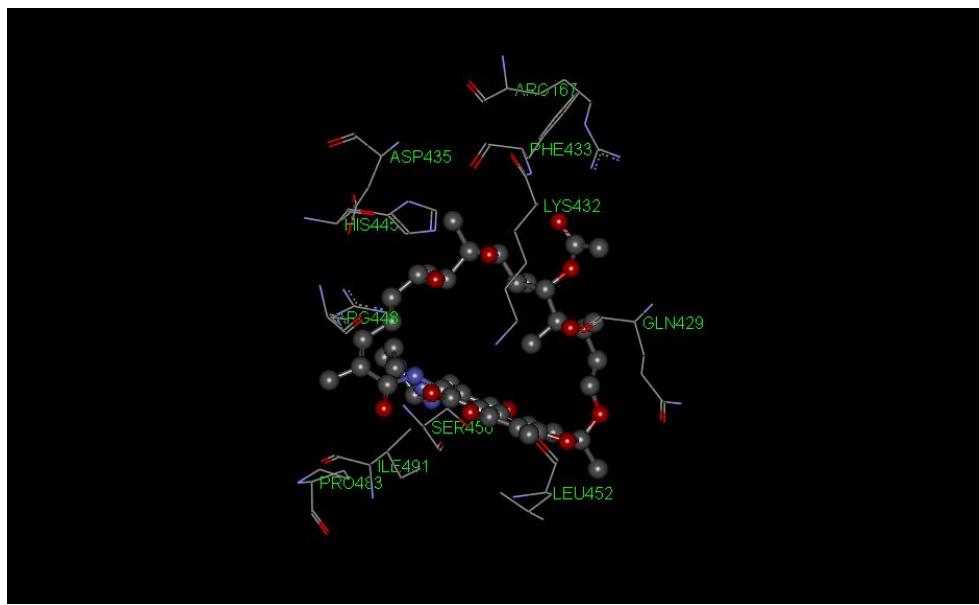
Figure 3.21 3D viewing RIF docking side with 438 Q – K mutation
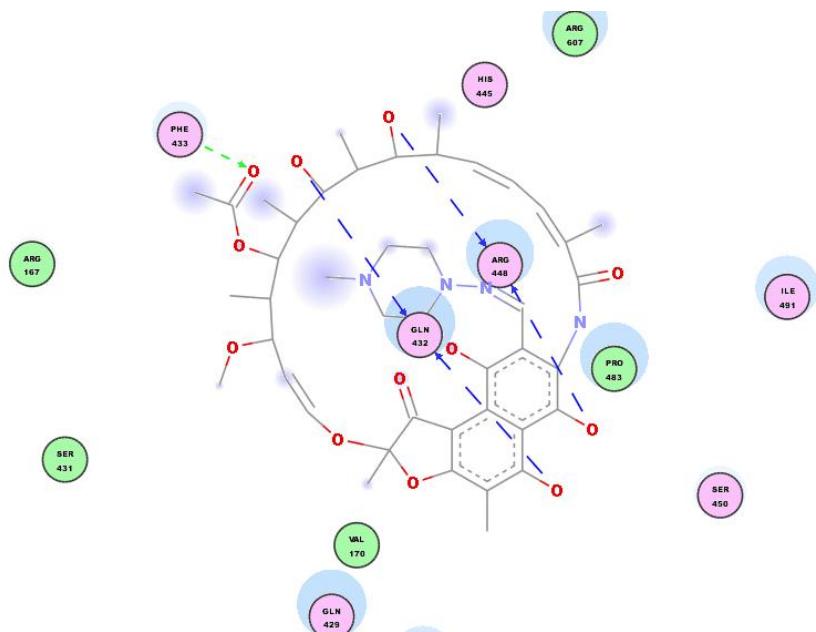
Figure 3.22 2D viewing RIF docking side with 447 S – Q mutation
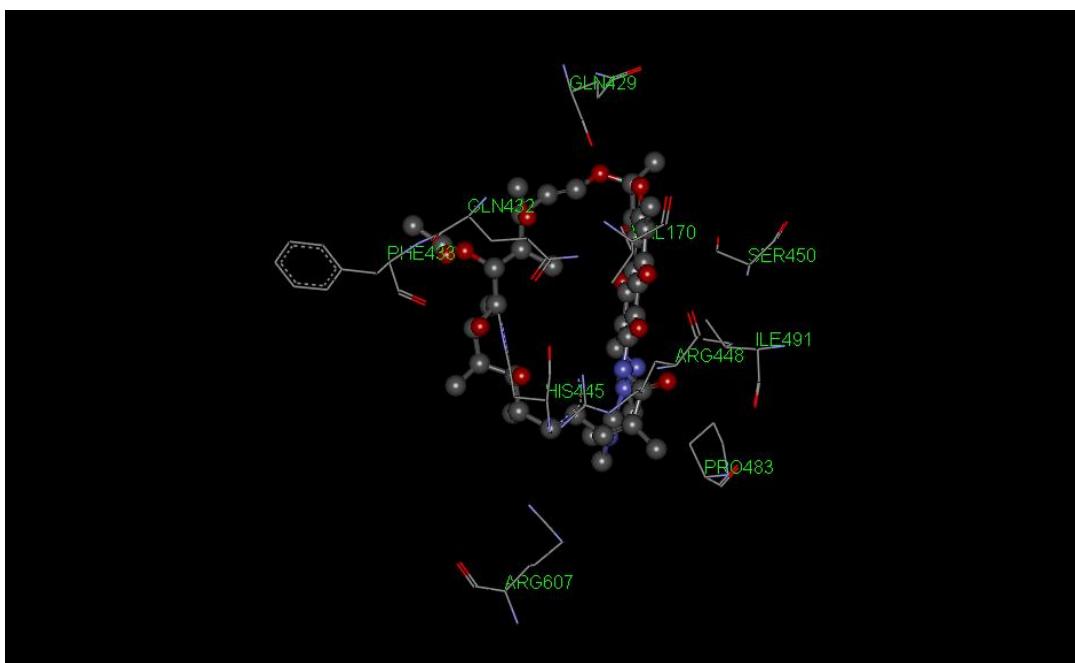
Figure 3.23 3D viewing RIF docking side with 447 S – Q mutation
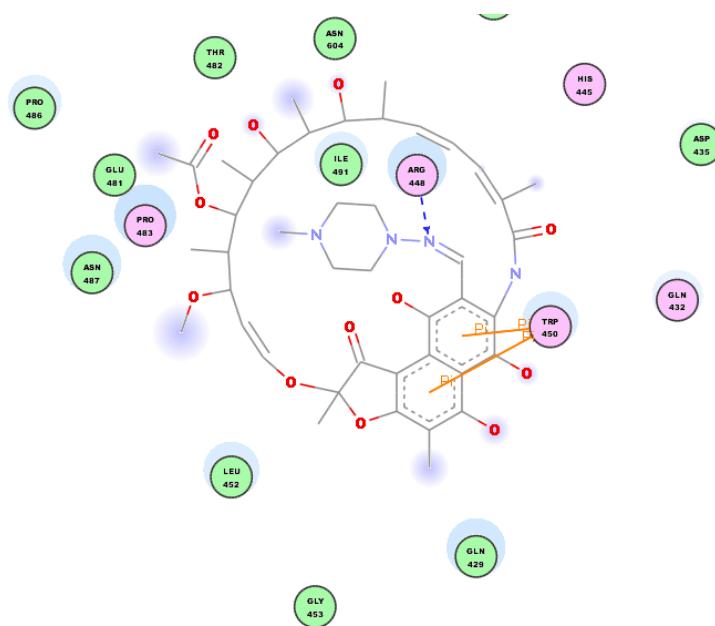
Figure 3.24 2D viewing RIF docking side with 456 S – W mutation
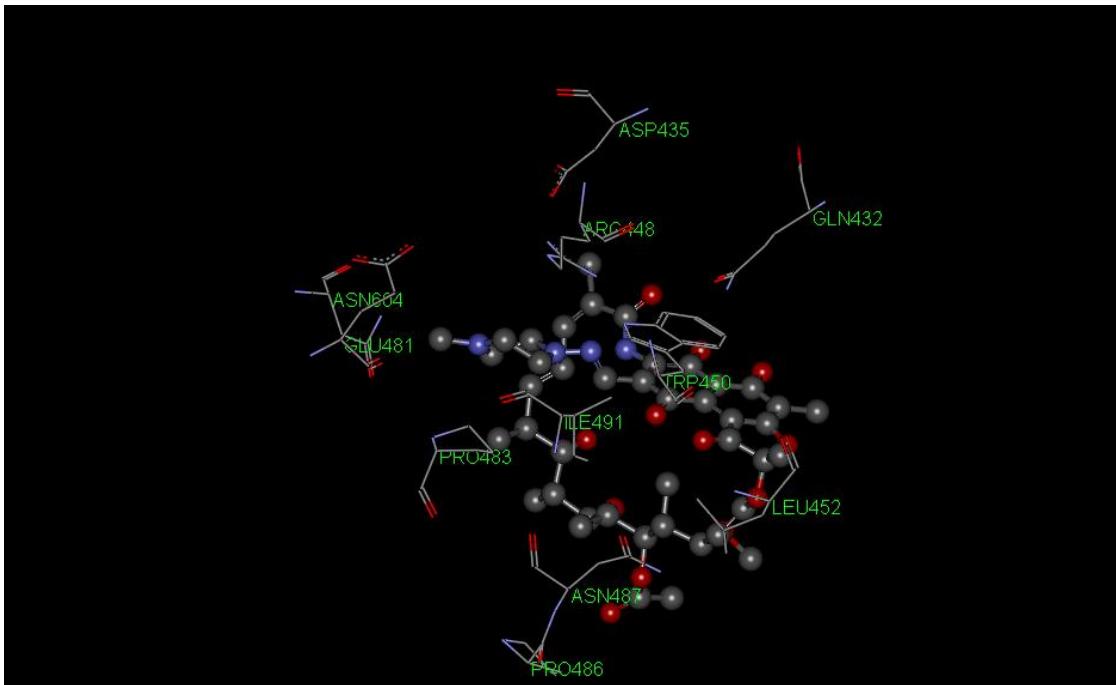
Figure 3.25 3D viewing RIF docking side with 456 S – W

Chapter 4

Conclusion

We understand that computational science method is good starting for computing mutation resistance to drug for TB with RIF. The chancing amino acid characters with mutation reflect to the effect of drug. Next step cloud be tried of changing RIF for breaking resistance of mutations.

About rpoβ protein modelin we can say that I couldn't locate its co-factor (Mg++). We know that rpoβ is need Mg++ for starting working. But it is not possible located an atom on modeling.

REFERENCES

[1] http://en.wikipedia.org/wiki/Rifampicin

[2] Tortora et al., **1995**; 329-339

[3] Koch, **1882**; 16: 73-93 **Rothschild, Bruce M. & Martin, Larry D. 2006.** Did ice-age bovids spread tuberculosis?. Naturwissenschaften 93: 565-569.

[4] Gomez JE, McKinney JD. M. tuberculosis persistence, latency, and drug tolerance. **2004**

[5] Stokes & Doxsee, Interaction of isoniazid with mycobacterium tuberculosis enoyl-acyl carrier protein reductase (inha):from molecular perspectives **1999**

[6] Iyawoo, A survey on tuberculosis cases in penang **2004**

[7] Middlebrook, The Middlebrook-Dubos Hemagglutination Test for Tuberculosis in Children and Adults **1952.**

[8] Miesel et al., **1998**

[9] Rodríguez, Unrestrained caspase-dependent cell death caused by loss of Diap1 function requires the Drosophila Apaf-1 homolog, Dark **2002**

[10] Shrivastava A, Radziejewski C, Campbell E, Kovac L, McGlynn M, Ryan TE, Davis S, Goldfarb MP, Glass DJ, Lemke G, Yancopoulos GD. An orphan receptor tyrosine kinase family whose members serve as nonintegrin collagen receptors. **1997**

[11] Ursula Pieper, Narayanan Eswar, Fred P. Davis, Hannes Braberg, M. S. Madhusudhan, Andrea Rossi, Marc Marti-Renom, Rachel Karchin, Ben M. Webb, David Eramian, Min-Yi Shen, Libusha Kelly, Francisco Melo, and Andrej Sali,., MODBASE: a database of annotated comparative protein structure models and associated resources. *Nucleic Acids Res. 2006 January 1*

[12] Marketa Zvelebil, Jeremy O. Baum, Understanding Bioinformatics, *Garland Science, Taylor&Francis Group,* LLC,**2008**

[13] Srinivas Aluru ; Handbook of Computational Molecular Biology, *Chapman&Hall/CRC,Taylor&Francis Group*, **2006**

[14] Jonathan Pevsner; Bioinformatics and Functional Genomics Second Edition, *Wiley Blackwell, A John Wiley&Sons, Inc, Publication*, **2009**

[15] Steven Henikoff and Jorja G. Henikoff ; Amino acid substitution matrices from protein blocks; *Proc. Natl. Acad. Sci. USA,***1992;** 10915-10919

[16] Albert Bray Hopkin Johnson Lewis Raff Roberts Walter; Essential Cell Biology Third Edition; *Garland Science, Taylor&Francis Group;* **2009**

[17] A.Storch, A.C.Wdolph and J.Schwarz; Dopamine transporter: Involvement in selective dopaminergic neurotoxicity and degeneration, *Journal of Neural Transmission,* **2004;** 111: 1267-1286

[18] Philip E. Bourne and Helge Weissig; Structural Bioinformatics, *Wiley Liss, A John Wiley&Sons Publication,* **2003**

[19] Tutorial of DS

CV

Name Surname: ERÇİN DİNÇER

E-mail: ercindincer@gmail.com

Birth: 02.03.1982 Akyazı

Status: Single

Gender: Male

EDUCATION

- B.S, Computer Engineer (2009)

İstanbul University, İstanbul, Turkey

TECHINICAL AND LANGUAGE SKILLS

Computer Skills

- Microsoft Systems, Linux Shell Scripting, Parallel Programing (MPI), C/C++, MATLAB, VMD, Autodock, Discovery Studio, NAMD.

Languages Skills

- Turkish(Native), English(Upper-Intermediate)