CHAPTER 8

Molecular Modeling and Docking Studies of L-Glutaminase from *Bacillus Cereus* MTCC 1305

8.1. Introduction

The study of structure-function relationship of catabolic enzyme is important for facilitating the modification in enzyme specificity and design of novel metabolic pathways (Yadav et al., 2010). Docking study of regional specificity of enzyme for substrate is also requisite for structure-function relationship (Wang et al., 2009). The identification of the amino acid residues in active site of enzyme and their mechanism of action will help scientist to design a strategy for enhancing catalytic activity of the enzyme (Baig and Manickem, 2010). Knowledge of 3D structure of a protein is essential to understand their function and can be determined at high resolution by either computational analysis or using experimental methods such as X-ray crystallography, NMR spectroscopy. In absence of crystallographic structures, a variety of advanced homology modeling methods have been developed which can provide reliable models for protein that share ≥30% sequence identity with a known protein molecule (Bodade, 2010; Xiang, 2006; Nayeem et al., 2006; Khobragade et al., 2011; Beedkar et al., 2012). BLAST and FASTA have been widely used as alignment tools to provide evidence for homology by matching a new sequence against a database of previously annotated sequences (Senthil et al., 2012). However, these approaches can only detect homologous proteins that exhibit significant number of peptide sequence similarity. The availability of web based tools such as ExPASY's ProtParam, Swiss workspace, UCSF chimera, Swiss PDB Viewer etc. are used to characterize the physiochemical properties of desired protein as well as their primary, secondary and three dimensional structures. These tools have been used for homology modeling of L-asparaginase (Senthil et al., 2012; Oza et al., 2011), uricase (Beedkar et al., 2012) and RNase (Gundampati et al., 2012). There are only few reports

available for the determination of 3D structure of L-glutaminase using X-Ray crystallographic method (Brown et al., 2008; Yoshimune et al., 2006; Ziade et al., 2003). There is no report available for homology modeling of L-glutaminase produced from *Bacillus cereus* MTCC 1305 and study of its binding efficiency with L-glutamine substrate.

In this study, the peptide sequences of L-glutaminase from *Bacillus cereus* MTCC 1305 were obtained using MALDI-TOF/TOF analyzer. BLAST search of these peptide sequences against Protein Data Bank was performed to search the possible template model for homology modeling. The three dimensional structure for L-glutaminase from *Bacillus cereus* MTCC 1305 was predicted using the template model. ANOLEA, QMEAN4 global scores, Gromos 96 and PROCHEK tools was further applied to validate the predicted model structure. Docking analysis of the predicted model structure with L-glutamine as substrate was further studied using SWISS dock tool (http://swissdock.vital-it.ch/).

8.2. Materials and methods

8.2.1. Sequencing and alignment of L-glutaminase

MALDI-TOF/TOF analyzer (Applied Biosystems) was used to determine peptide sequences of purified L-glutaminase from *Bacillus cereus* MTCC 1305. After SDS-PAGE, gel pieces were excised, destained, washed, dehydrated in CH_3CN and dried in a vacuum centrifuge (Thermo Fischer scientific). Gel pieces were cooled on ice and soaked in buffer containing 50mM NH_4HCO_3, 5mM $CaCl_2$, and 12.5ng/µl of trypsin (Promega, sequencing grade) for overnight at 37˚C. The digested peptides were recovered from the gel by sonicating in water bath for 10min and the process was

repeated three times. The tryptic-digested sample was mixed with the matrix (saturated solution of α-cyano-4-hydroxy-cinnamic acid in 0.1%TFA and 50% acetonitrile) and spotted on MALDI plate. The mixture was allowed to dry at room temperature and used for the MALDI analysis. Masses obtained after tryptic digest were fed in general protein database (MASCOT Search) and compared with masses of known proteins in the database for the identification of the sample. The database contains a theoretical tryptic digest of all known proteins and the exact masses of the corresponding peptides. The peptide sequences of L-glutaminase from *Bacillus cereus* MTCC 1305 was used as query sequence.

Peptide sequences of L-glutaminase of different *Bacillus sp.* were retrieved randomly using BLAST as a tool provided by NCBI (http://blast.ncbi.nlm.nih.gov/Blast.cgi) in FASTA format. BLAST search against Protein Data Bank (http://www.rcsb.org) was performed to search the possible template for homology modeling.

8.2.2. Study of physicochemical properties of L-glutaminase

The source sequence was further aligned against PDB data base for sequences of L-glutaminase from different *Bacillus species* by using Clustal W 2.0.12 tool (Larkin et al., 2007). EXPASY's ProtParam tool is used to compute various physicochemical properties that can be deduced from a protein sequence (Gasteiger et al., 2005). The parameters computed by ProtParam include the molecular weight, theoretical pI, amino acid composition, atomic composition, extinction coefficient, estimated half-life, instability index, aliphatic index and grand average of hydropathicity (GRAVY). Extinction coefficient of native protein ($\varepsilon_{protein}$) at wavelength 280nm is computed using the following equation (Pace et al., 1995; Edelhoch, 1967):

$$\epsilon_{\text{Protein}_{280}} = (N_{Trp} \times \varepsilon_{Trp}) + (N_{Tyr} \times \varepsilon_{Tyr}) + (N_{Cys} \times \varepsilon_{Cys}) \ldots\ldots\ldots\ldots (8.1)$$

Where, $N_{Trp,}N_{Tyr}$ and N_{Cys} are the number of tryptophan, tyrosine and cystine and their extinction coefficient values at 280 nm were reported as follows:

$\varepsilon_{Trp} = 5500 M^{-1}cm^{-1}, \varepsilon_{Tyr} = 1490 M^{-1}cm^{-1}$ and $\varepsilon_{Cys} = 125 M^{-1}cm^{-1}$.

Two values have been reported for native proteins using above equation 8.1 in which, first value is based on the assumption that all cysteine residues appear as half cystines, and the second value assuming that no cysteine appears as half cystine (Gill and Von, 1989).

Half-life is a prediction of the time for disappearing of half of the amount of synthesized protein in a cell. ProtParam relies on the "N-end rule", which relates the half-life of a protein to the identity of its N-terminal residue (Bachmair et al., 1986). The identity of the N-terminal residue of a protein plays an important role in determining its stability in vivo (Gonda et al., 1989; Tobias et al., 1991; Varshavsky, 1997). The instability index (II) provides an estimate of the stability of native protein which is computed in terms of weight values by using this formula:

$$II = \left(\frac{10}{L}\right) \times \sum DIWV(X_i X_{i+1}) \ldots\ldots\ldots\ldots\ldots (8.2)$$

Where, L is the length of sequence, i=1 to L-1, $DIWV(X_i X_{i+1})$ is the instability weight value for the dipeptide starting in position 'i'.

A protein whose instability index is smaller than 40 is predicted as stable, a value above 40 predicts that the protein may be unstable (Guruprasad et al., 1990). The relative volume of a protein occupied by aliphatic side chains (alanine, valine, isoleucine, and leucine) is reported in term of aliphatic index (AI), which is regarded

as a positive factor for the increase of thermal stability of globular proteins (Ikai, 1980). The aliphatic index of a protein is calculated by using following formula:

$$AI = X(Ala) + a * X(Val) + b * X(Ile) + b * X(Leu) \ldots\ldots\ldots\ldots (8.3)$$

Where, $X(Ala), X(Val), X(Ile)$ and $X(Leu)$ are the amino acid compositional fractions of alanine, valine, isoleucine and leucine.

The relative volume of valine and leucine/isoleucine side chains compared to the side chain of alanine is reported in terms of constant values of a=2.9 and b=3.9 respectively. GRAVY value for a peptide or protein is calculated as the sum of hydropathy values of all the amino acids, divided by the number of residues in the sequence (Kyte and Doolittle, 1982).

8.2.3. Model refinement and structure validation

Swiss model was used to predict a suitable 3D model structure for this enzyme and model was refined by performing an optimized geometry calculation in mechanics including energy terms for bond stretch, bond angles, dihedral angle, improper torsion stretch, bond bend, Vander Waals electrostatic and hydrogen bond interactions (Arnold et al., 2006; Schwede et al., 2003; Guex and Peitsch, 1997).

The quality of refined model was assessed on the basis of both geometric and energetic aspects. The structure was evaluated first by QMEAN4 global scores, ANOLEA force field and GROMOS 96 (Melo and Feytmans, 1998). These tools provided the graphical representation of energy minimization of obtained protein model. The stereochemical property was assessed by Ramchandran Plot using PROCHECK. Ramchandran Plot provided the residue positions in particular segment based on phi (ϕ) and psi (ψ) angles between Cα–C and N–Cα respectively (Yadav et

al., 2010). PROSA test was employed to evaluate the quality of consistency between the native fold and the sequence to examine the energy of residue interactions in term of '*Z* score'. The model structure is considered within the expected quality range with Z score value deviates less than1.

UCSF Chimera (http://www.cgl.ucsf.edu/chimera) and Swiss PDB Viewer (http://www.expasy.org/spdbv) tools were used for visualization of predicted model structure after protein structure assessment.

8.2.4. Docking analysis of L-glutaminase with L- glutamine

Docking studies of predicted protein structure of glutaminase was performed using SWISS Dock tool (http://swissdock.vital-it.ch/). Swiss Dock is based on the docking software EA Dock DSS, whose algorithm consists of many binding modes. CHARMM energies (Grosdidier et al., 2009) of the binding modes were estimated on a grid and binding modes with the most favorable energies were clustered. Cluster of binding modes were evaluated using equation 8.4 in term of simple fitness scoring function "ΔG_{score}" which accounts for solvation of energy.

$$\Delta G_{score}=E_{intra,lig}+E_{intra,prot}+ E_{vdw} +E_{elec}+\Delta G_{solv,elec} + \Delta G_{solv,np} \quad \text{.....(8.4)}$$

Where, $E_{intra,lig}$ and $E_{intra,prot}$ were the internal energies of the ligand and the protein, respectively. E_{vdw} and E_{elec} were the Vander Waals and Coulomb electrostatic energies of the interaction between ligand and protein, respectively. $\Delta G_{solv,elec}$ and $\Delta G_{solv,np}$ were the electrostatic solvation energy and non-polar solvation energy of the complex respectively.

Fast Analytical Continuum Treatment of Solvation (FACTS) has been used as an efficient method for calculating the free energy of solvation of molecules embedded in a continuum solvent (Haberthur and Caflisch, 2008). This method is based on the fully analytical evaluation of the volume and spatial symmetry of the solvent that is

displaced around a solute atom by its neighboring atoms. The initial structure of L-glutamine substrate was obtained from NCBI Pubchem (http://pubchem.ncbi.nlm.nih.gov/) and its systematic analysis was done using KEGG tool (http://www.kegg.jp/kegg/kegg1.html).

8.3. Results and Discussion

Glutaminases (EC 3.5.1.2) belong to the large group of serine β-lactamases and penicillin-binding proteins, which have a common evolutionary origin and share the protein fold, structural motifs and catalytic mechanism (Brown et al., 2008). Purified L-glutaminase from *Bacillus cereus* MTCC 1305 was resolved into 326 peptide sequence length (Table8.1) and twenty three peaks were obtained after tryptic digestion in MALDI-TOF/TOF analyzer (Fig8.1). Among these, six peaks were obtained with good signal intensity and their peptide sequences are shown in Table8.2.

Table8.1. Total peptide sequences of purified L-glutaminase from *Bacillus cereus* MTCC 1305 after MALDI-TOF analysis

```
10         20         30         40         50         60
MIKDSSVQVE GQEKVCLDQW VAHYRTYAAK GRSASYIPAL GEINVSQLGI CIVKPDGTMI

        70         80         90        100        110        120
KSGDWEIPFT LQSISKVIGF IAACLSRGIS YVLERVDVEP TGDAFNSIIR LEIHKPGKPF

       130        140        150        160        170        180
NPMINAGAIT IASLLPGTSV QEKLESIYVL IEKMIEKRPA INEIVFQSEW ETAHRNRALA

       190        200        210        220        230        240
YYLKENGFLE SDVEETLEVY LKQCSIEINT EDIALIGLIL AHDGYHPIRK EQVLPKEVAR

       250        260        270        280        290        300
LTKALMLTCG MYNASGKFAA FIGLPAKSGV SGGIMTLVPS KSRKDLSFQD GCGIGIYGPA

       310        320
IDEYGNSLPG IMLLEHIAKE WDLSIF
```

Table8.2. Peptide sequences for six peaks of L-glutaminase from *B. cereus* MTCC 1305with their expected molecular weight

gi|49185938 **Mass:** 34926 **Score:** 93 **Expect:** 27 **Matches:** 6

Glutaminase [Bacillus cereus]

Observed	Mr(expt)	Mr(calc)	ppm	Start	End	Miss	Peptide
882.5263	881.5190	881.3735	165	87	94	0	RGISYVLE
1082.5691	1081.5618	1081.6179	-51.81	167	175	1	QSEWETAHR
1234.6384	1233.6311	1233.5733	46.8	95	105	0	RVDVEPTGDAF
1493.6940	1492.6867	1492.8296	-95.72	160	173	1	AINEIVFQSEWETA
1765.6999	1764.6927	1764.7712	-44.50	284	299	0	KDLSFQDGCGIGIYGPA
1889.9260	1888.9188	1888.8593	31.5	77	94	1	VIGFIAACLSRGISYVLE

No match to: 807.4183, 868.5084, 1037.5003, 1434.7248, 1470.6826, 1657.7503, 1707.7147, 1838.8842, 1851.8823, 1982.9109, 2047.0626, 2343.0359, 2351.2679, 2399.1007, 2717.2739, 2746.6724, 2872.8413

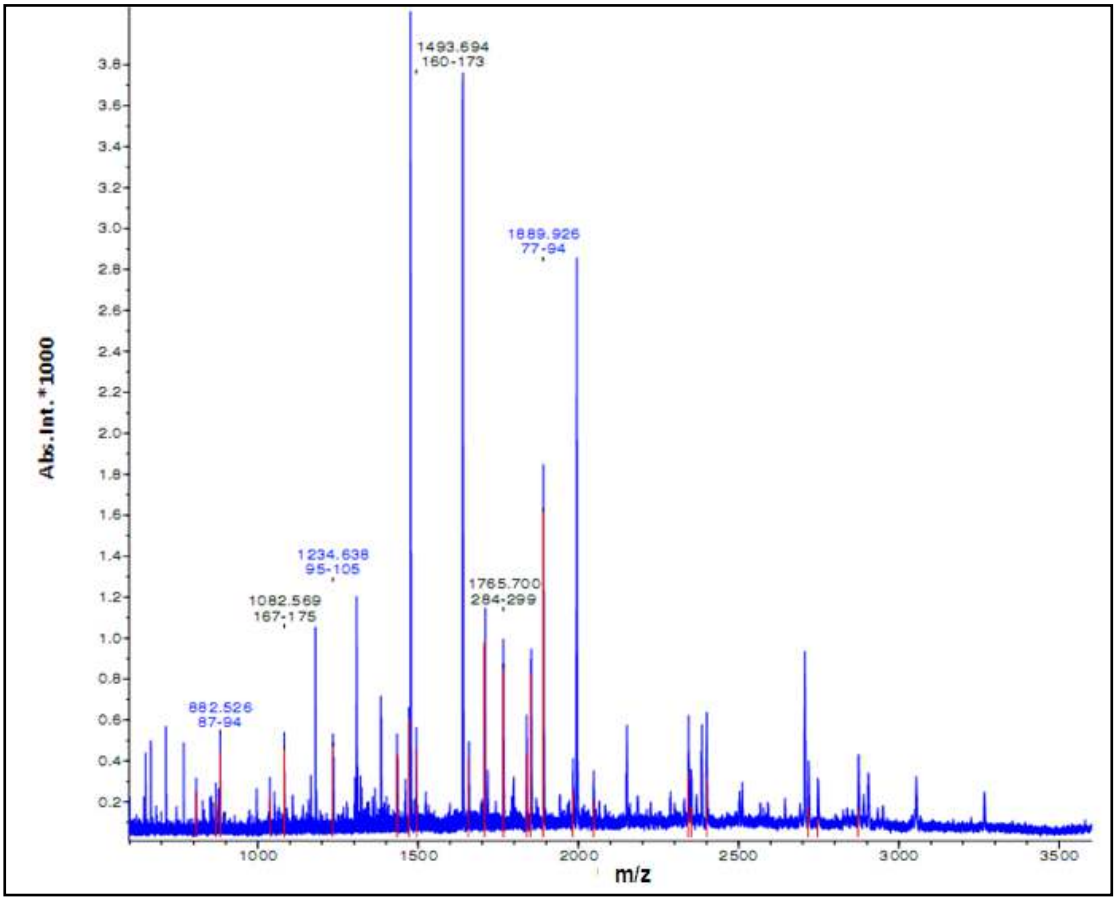


Fig.8.1. Resolution of purified L-glutaminase from *Bacillus cereus* MTCC 1305 into twenty three peaks

Absence of the three dimensional structures for L-glutaminase from *Bacillus cereus* MTCC 1305 prompted us to predict its homology model. L-glutaminase sequences for other *Bacillus* spp. were retrieved from NCBI database and results were saved in FASTA format. BLAST of theses sequences showed maximum similarity with the sequences of glutaminase of different *Bacillus species* as shown in Table8.3.

Table8.3. Multiple sequence alignments of L-glutaminase from different *Bacillus* spp. using clustal W 2.0.12 software tool

Bacillus sp.	Accession no.	Maximum score	Total score	Query Coverage	E value	Maximum identity
Bacillus thuringienis	YP_005566769.1	80.5	127	100%	6e-17	95%
	YP_895597.1	80.5	127	100%	6e-17	95%
	YP_037231.1	80.5	120	100%	6e-17	80%
	ZP_04121061.1	79.3	124	100%	1e-16	90%
	ZP_04105681.1	79.0	125	100%	2e-16	95%
	ZP_04079301.1	78.6	125	100%	2e-16	95%
	ZP_04072725.1	78.2	124	100%	3e-16	95%
	ZP_04084528.1	77.4	124	100%	7e-16	95%
	ZP_00742937.1	44.7	44.7	32%	8e-05	95%
Bacillus cereus	YP_002452114.1	80.5	117	100%	6e-17	75%
	ZP_00235321.1	80.5	125	100%	6e-17	90%
	ZP_04186829.1	80.1	126	100%	6e-17	95%
	EJR97850.1	79.7	125	100%	9e-17	90%
	ZP_04218082.1	79.3	122	100%	1e-16	85%
	NP_832857.1	79.3	125	100%	1e-16	95%
	YP_002339120.1	78.6	123	100%	2e-16	95%
	YP_005119710.1	78.6	122	100%	2e-16	90%
Bacillus subtilis	2OSU_A	61.6	98.2	100%	6e-10	68%
	1MKI_A	61.2	97.8	100%	6e-10	68%
Bacillus megaterium	YP_003596016.1	61.2	96.3	100%	7e-10	70%
Bacillus amyloliquefaciens	YP_005544011.1	62.0	98.6	100%	2e-10	64%
Bacillus pumilus	ZP_03056239.1	61.6	98.6	100%	5e-10	64%
Bacillus anthracis	NP_845472.1	80.1	126	100%	6e-17	95%
	ZP_02390686.1	80.1	126	100%	7e-17	95%

Multiple sequence alignment of three large peptide sequences of purified L-glutaminase from *Bacillus cereus* MTCC 1305 (query sequences) was studied against known peptide sequences of glutaminases from different *Bacillus* spp. and homology of sequences is sown in Fig8.2. These query peptide sequences showed strong homology with glutaminase sequences of *Bacillus cereus* (gi/52142384, gi/30021226, gi/218233997).

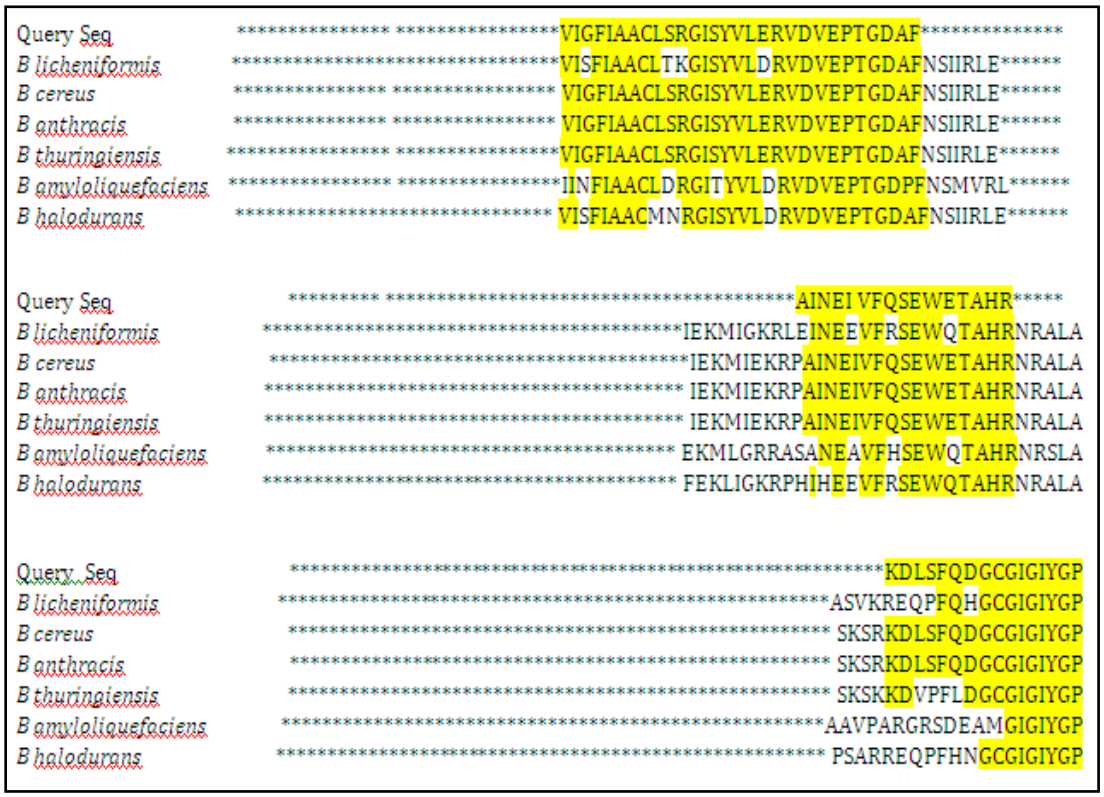


Fig.8.2. Multiple sequence alignment of query sequences of L-glutaminase (gi/49185938) with peptide sequences of glutaminase from other *Bacillus* species like *Bacillus licheniformis* (gi/52078728), *Bacillus cereus* (gi/52142384; gi/30021226; gi/218233997), *Bacillus anthracis* (gi/47528446), *Bacillus thuringiensis* (gi/49480086), *Bacillus amyloliquefaciens* (gi/154684749), *Bacillus halodurans* (gi/15615280)

Three dimensional structure of L-glutaminase from *Bacillus subtilis* (PDB ID: 1MKI_A) was already developed using X-ray crystallography (Brown et al., 2008) and this model structure was selected as a suitable template for homology modeling. The alignment of most promising identical L-glutaminase query sequences with peptide sequences of template model is shown in Table8.4.

Table8.4. Alignment of L-glutaminase sequences from *Bacillus cereus* MTCC 1305 with template sequences of *Bacillus subtilis* (PDB:1MKI_A). The conserved residues were represented by "h" and semi conserved substitution was represented by "s" in sequence alignment

```
TARGET    1                   LDQWVAH YRTYAAKGRS ASYIPALGEI NVSQLGICIV
userX     0     axkelinpal q--lhdwvey yrpfaangqs a--------- ndsqlgicvl

TARGET                         hhhhhh hhhhhh                   ssssss
userX              sss   hhh h  hhhhhhh hh   h                ssssss


TARGET    38    KPDGTMIKSG DWEIPFTLQS ISKVIGFIAA CLSRGISYVL ERVDVEPTGD
userX     55    epdgtxihag dwnvsftxqx iskvisfiaa cxsrgipyvl drvdveptgd

TARGET          s   ssssss s    sss   hhhhhhhhhh hhh   hhhh hh
userX           s   ssssss s    sss   hhhhhhhhhh hhh   hhhh hh


TARGET    88    AFNSIIRLEI HKPGKPFNPM INAGAITIAS LLPGTSVQEK LESIYVLIEK
userX     105   afnsiirlei nkpgkpfnpx inagaltias ilpgesayek leflysvxet

TARGET               hhhhh              hhhhhhhhh h     hhhh hhhhhhhhhh
userX                hhhhh              hhhhhhhhh h     hhhh hhhhhhhhhh


TARGET    138   MIEKRPAINE IVFQSEWETA HRNRALAYYL KENGFLESDV EETLEVYLKQ
userX     155   ligkrprihe evfrseweta hrnralayyl ketnfleaev eetlevylkq

TARGET          hh    sssh hhhhhhh    hhhhhhhhhh hhh      h hhhhhhhhhh
userX           hh    sssh hhhhhhh    hhhhhhhhhh hhh      h hhhhhhhhhh


TARGET    188   CSIEINTEDI ALIGLILAHD GYHPIRKEQV LPKEVARLTK ALMLTCGMYN
userX     205   caxesttedi aliglilahd gyhpirheqv ipkdvaklak alxltcgxyn

TARGET          hssss  hhh hhhhhhhh   sss    sss   hhhhhhhh hhhhh
userX           hssss  hhh hhhhhhhh   sss    sss   hhhhhhhh hhhhh


TARGET    238   ASGKFAAFIG LPAKSGVSGG IMTLVPSKSR KDLSFQDGCG IGIYGPAIDE
userX     255   asgkyaafvg vpaksgvsgg ixalvppsar reqpfqsgcg igiygpaide

TARGET           hhhhhh     sssss  ss sssss             ss sssss
userX            hhhhhh     sssss  ss sssss             ss sssss


TARGET    288   YGNSLPGIML LEHIAKEWDL SIF
userX     305   ygnsltggxl lkhxaqewel sif-

TARGET              hhhhhh hhhhhhhh
userX               hhhhhh hhhhhhhh
```

The sequence identity score and E-value for these query sequences were obtained as 74.194% and $8.07699e^{-137}$ respectively after homology modeling with selected template model structure using SWISS model workspace. The theoretical physicochemical properties (molecular weight, molecular formula, theoretical pI, extinction coefficient, estimated half-life, instability index, aliphatic index and GRAVY) for this predicted model structure was obtained after using ExPASY's ProtParam tool and results are summarized in Table8.5.

Table8.5. Physicochemical properties of predicted model structure of L-glutaminase from *Bacillus cereus* MTCC 1305 after using ExPASY's ProtParam tool

Number of amino acids: 326, **Molecular weight:** 35885.5 Dalton
Atomic composition:

Carbon	C	1618
Hydrogen	H	2572
Nitrogen	N	416
Oxygen	O	474
Sulfur	S	14

Formula: $C_{1618}H_{2572}N_{416}O_{474}S_{14,}$ **Total number of atoms:** 5094
Theoretical pI: 5.51
Extinction coefficients:
Extinction coefficients are in units of M^{-1} cm^{-1}, at 280 nm measured in water.

- Ext. coefficient 40255, Abs 0.1% (=1 g/l) 1.122, assuming all pairs of Cys residues form cystines
- Ext. coefficient 39880, Abs 0.1% (=1 g/l) 1.111, assuming all Cys residues are reduced

Estimated half-life:
The N-terminal of the sequence considered is M (Met).
The estimated half-life is: 30 hours (mammalian reticulocytes, in vitro).
>20 hours (yeast, in vivo).
>10 hours (Escherichia coli, in vivo).

Instability index: 38.40
Aliphatic index: 103.53
Grand average of hydropathicity (GRAVY): 0.060

The values of extinction coefficient for this protein was computed as 402555$M^{-1}cm^{-1}$ and 39880$M^{-1}cm^{-1}$ which may indicate the presence of disulfide bond in the structure of L-glutaminase. The instability index for this enzyme was computed as 38.40 which may indicate high stability for this enzyme (Guruprasad et al., 1990). Aliphatic index was obtained as 103.53 indicating stability of this enzyme for a wide temperature range. The lower value of GRAVY (0.060) for this enzyme indicates the possibility of better interaction of this enzyme with water. The molecular weight of this protein biomolecule was obtained as 35.88KD with molecular formulae $C_{1618}H_{2572}N_{416}O_{474}S_{14}$. The estimated half life of this enzyme (Table8.5) showed similarity to half life reported for beta-gal proteins, which may indicate the presence of methionine as N-terminal amino acid residue for this enzyme. The different composition of amino acid predicted for the purified L-glutaminase after using ExPASY's ProtParam tool is shown in Table8.6. The purified enzyme was found to be rich with Ile, Leu, Ala, Glu, Gly, Ser and Lys amino acids.

Table8.6. Amino acid composition for predicted model structure of L-glutaminase from *Bacillus cereus* MTCC 1305

Amino acids	Number of amino acid	% of amino acids
Ala (A)	25	7.7%
Arg (R)	11	3.4%
Asn (N)	10	3.1%
Asp (D)	13	4.0%
Cys (C)	6	1.8%
Gln (Q)	10	3.1%
Glu (E)	26	8.0%
Gly (G)	27	8.3%
His (H)	6	1.8%
Ile (I)	36	11.0%
Leu (L)	30	9.2%
Lys (K)	21	6.4%
Met (M)	8	2.5%
Phe (F)	10	3.1%
Pro (P)	15	4.6%
Ser (S)	25	7.7%
Thr (T)	12	3.7%
Trp (W)	4	1.2%
Tyr (Y)	12	3.7%
Val (V)	19	5.8%

The detailed secondary structure investigation of the L-glutaminase of *Bacillus cereus* MTCC 1305 revealed that 131 (40.18%) residues were in α helix, 47 (14.42%) residues in beta strand and 148 (45.40%) residues in random coil as shown in Fig8.3. A model structure for this enzyme was predicted by using protein 3D structure prediction tool of SWISS model as shown in Fig8.4. The stereo chemical quality of the predicted model was evaluated using Ramachandran plot map calculations based

on statistics values for plot quality, peptide bond planarity, Bad contacts of non-bonded interactions, alpha carbon tetrahedral distortion, hydrogen bond energies and overall G factor shown in Fig8.5.

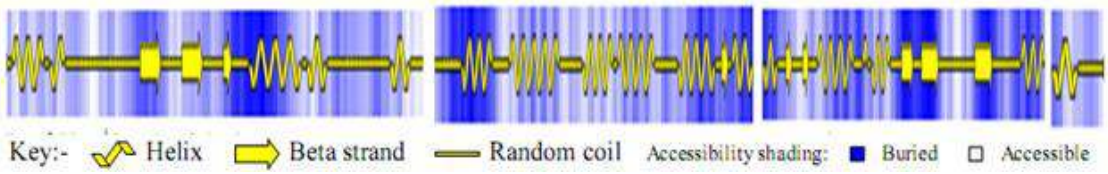


Fig.8.3. Secondary Structure of L-glutaminase from *Bacillus cereus* MTCC 1305 with estimated accessibility

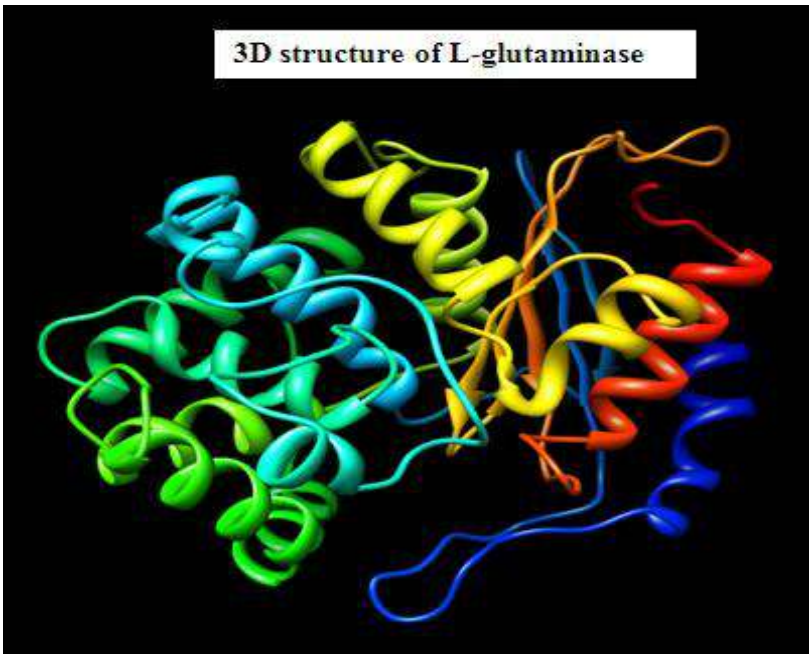
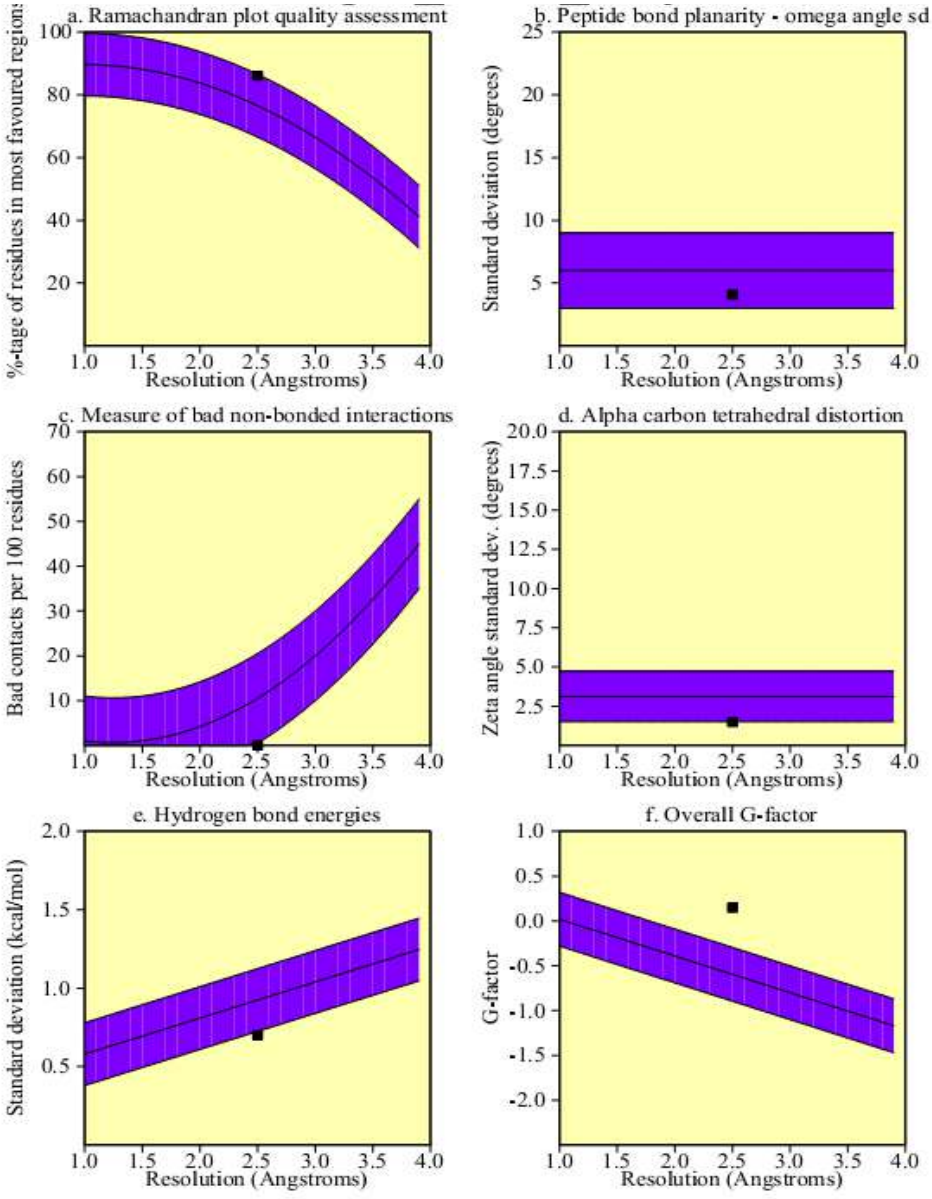


Fig.8.4. Predicted 3D structure for L-glutaminase from *Bacillus cereus* MTCC 1305 after homology modeling with α helix represented by yellow, green ribbon and β-turn represented by red, olive blue and dark blue ribbon.

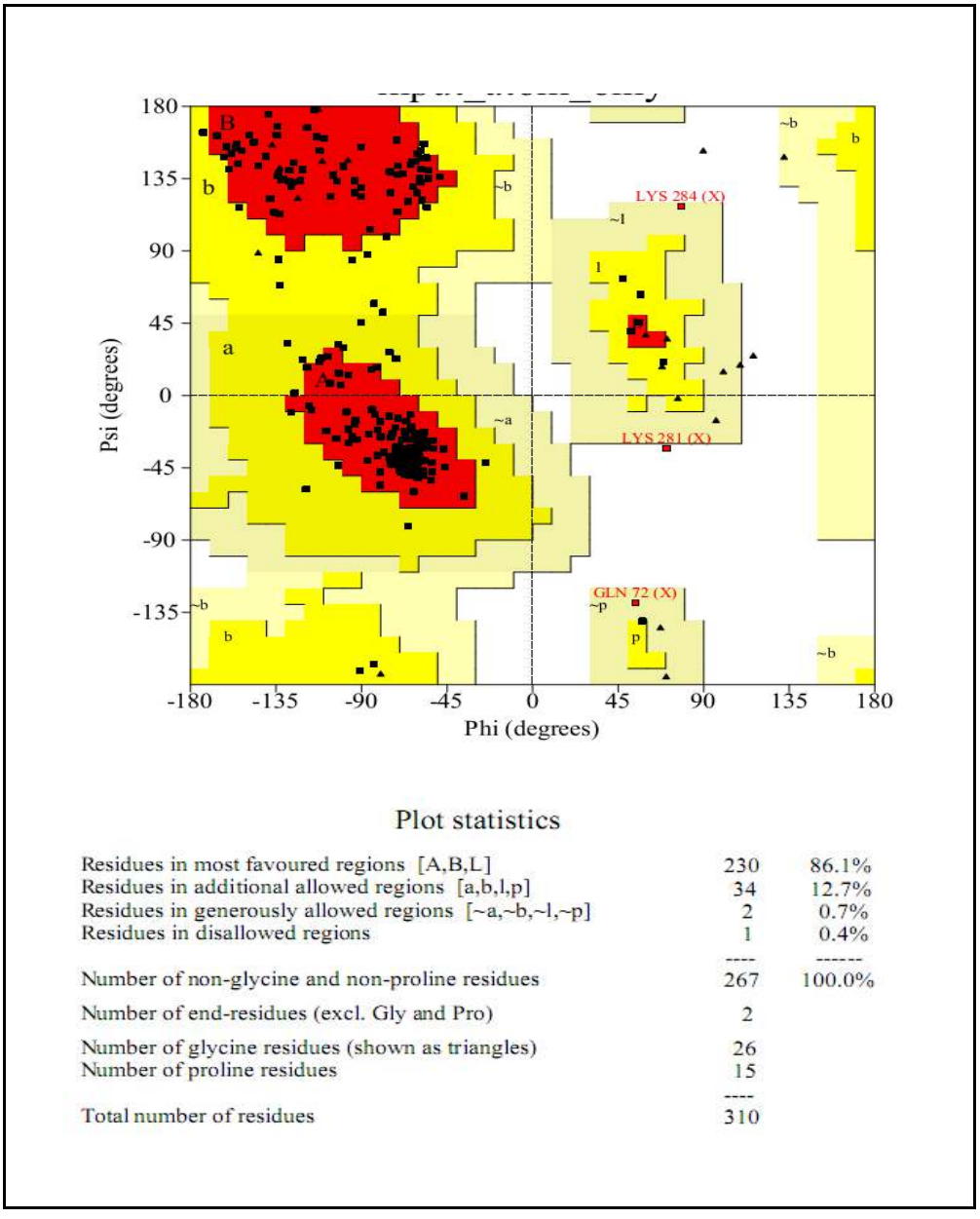


Plot statistics

Stereochemical parameter	No. of data pts	Parameter value	Comparison values Typical value	Band width	No. of band widths from mean	
a. %-tage residues in A, B, L	267	86.1	76.6	10.0	1.0	Inside
b. Omega angle st dev	308	4.1	6.0	3.0	-0.6	Inside
c. Bad contacts / 100 residues	0	0.0	10.5	10.0	-1.1	BETTER
d. Zeta angle st dev	284	1.5	3.1	1.6	-1.0	BETTER
e. H-bond energy st dev	205	0.7	0.9	0.2	-1.1	BETTER
f. Overall G-factor	310	0.1	-0.6	0.3	2.5	BETTER

Fig.8.5. Ramachandran map calculation on basis of different chain parameters (a) Plot quality assessment (b) peptide bond planarity (c) measure of bad non-bonded interaction (d) Alpha carbon tetrahedral distortion (e) hydrogen bond energies (f) Overall G factor

Ramachandran Plot for this model structure showing distribution of residues in different region is presented in Fig8.6. It was found that the phi/psi angles of 86.1% of residues fell in the most favored regions, 12.7% residues fell in allowed region, 0.7% residues fell in generously allowed region and 0.4% residues fell in disallowed regions.



Plot statistics

Residues in most favoured regions [A,B,L]	230	86.1%
Residues in additional allowed regions [a,b,l,p]	34	12.7%
Residues in generously allowed regions [~a,~b,~l,~p]	2	0.7%
Residues in disallowed regions	1	0.4%
Number of non-glycine and non-proline residues	267	100.0%
Number of end-residues (excl. Gly and Pro)	2	
Number of glycine residues (shown as triangles)	26	
Number of proline residues	15	
Total number of residues	310	

Fig.8.6. Ramachandran Plot for predicted model of L-glutaminase from *Bacillus cereus* MTCC 1305

The value of goodness factor (G) was obtained as 0.1 which lie in acceptable range of 0-0.5 for good quality model. The distribution of the main chain bond lengths and bond angles were found to be within the limits for these proteins. Asp56, Trp65, Glu66, Phe120, Glu142, Glu156, Tyr182, Asp212, Tyr225 and Phe261 were found as conserved residues in active site of this predicted model structure. The energy of each amino acid of refined protein structure was analyzed in term of ANOLEA, QMEAN score and GROMOS 96 as shown in Fig8.7. In these graphs, negative and positive values for given amino acid indicate the favorable and unfavorable energy environment respectively (Melo and Feytmans, 1998). Total energy of the refined protein structure was calculated as -11636.01KJ/mol which was minimum energy level and may indicate stable structure for this predicted model. Model quality estimates based on the QMEAN scoring function were normalized with respect to the number of interactions (Table8.7). The interaction energy of each residue with the remainder of a protein was computed to judge whether it fulfills certain energy criteria or not. Model quality score for individual model was then expressed as '*Z* score' in comparison to score obtained for high-resolution crystal structures. The average normalized QMEAN4 score was constant over a wide range of protein size and the variance of the distribution depends on protein size (Fig8.8). QMEAN *Z*-score for a given model was calculated from its normalized QMEAN score by subtracting the average normalized QMEAN score and divided by the standard deviation of the observed distribution. Z-score was calculated for all individual terms of the composite score. QMEAN *Z*-score of the predicted model structure was obtained as -1.47 indicating no significant deviation from typical native structure of similar size of template protein. Z score value for C_beta interaction energy, all-atom pairwise

energy, solvation energy and torsion angle energy were within the expected quality range as shown in Table8.7.

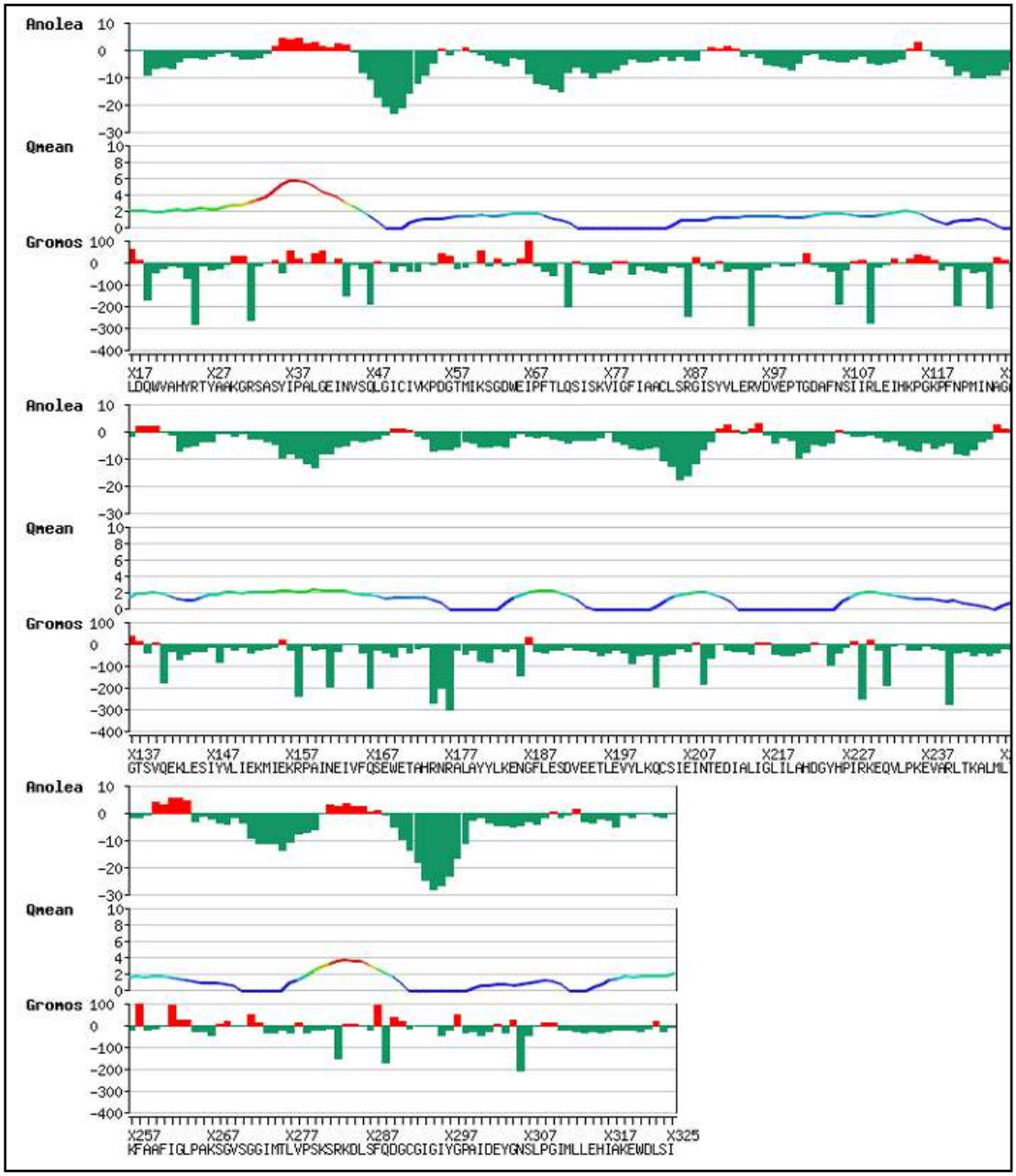


Fig.8.7. Validation of predicted model structure of L-glutaminase from *Bacillus cereus* MTCC 1305 in terms of ANOLEA and GROMOS force field

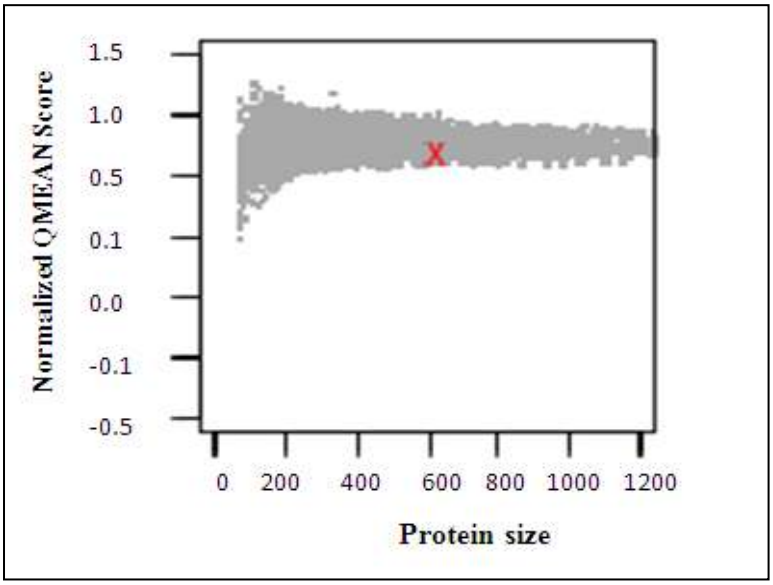


Fig8.8. Normalized QMEAN score composed of four statistical potential terms (QMEAN4) of high resolution structure

Table8.7. Z score analysis of protein structure in term of pseudo energies level and QMEAN 4 with respect to raw score obtained for high resolution experimental structure of similar size solved by X-Ray crystallography

Scoring function term	Raw score	Z-score
C_beta interaction energy	-102.13	-0.87
All-atom pairwise energy	-6718.77	-1.30
Solvation energy	-23.74	-0.77
Torsion angle energy	-69.83	-0.88
QMEAN4 score	0.685	-1.47

Where, QMEAN4 score is a composite score consisting of a linear combination of 4 statistical potential terms (estimated model quality between 0-1)

Docking study of all homology models was performed in order to elucidate its structural and functional relevance in terms of substrate binding and specificity. Enzyme-substrate complex represents a more detailed and accurate picture of the interactions of substrate in active site of predicted model structure. Docking analysis of L-glutaminase with L-glutamine substrate was performed and favorable docking score was observed for 4^{th} cluster and 16^{th} clusters. High value of energy of full fitness and low value of ΔG score for 16^{th} cluster indicates a good interaction between L-glutaminase and L-glutamine. The binding of the L-glutamine substrate with high affinity at the potential binding site of the predicted model structure of L-glutaminase is shown in Fig8.9 and Fig8.10.

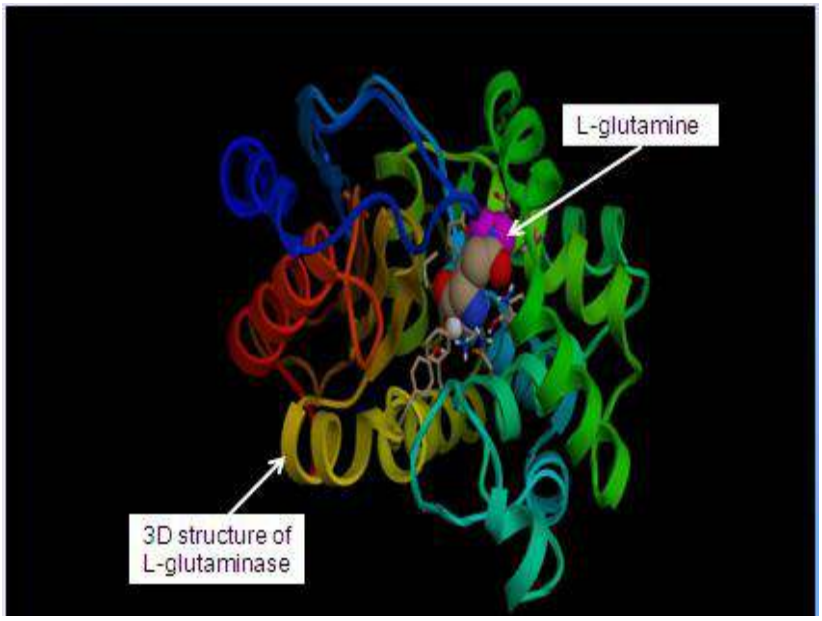


Fig.8.9. Structural view of binding of predicted 3D model structure of L-glutaminase with L-glutamine

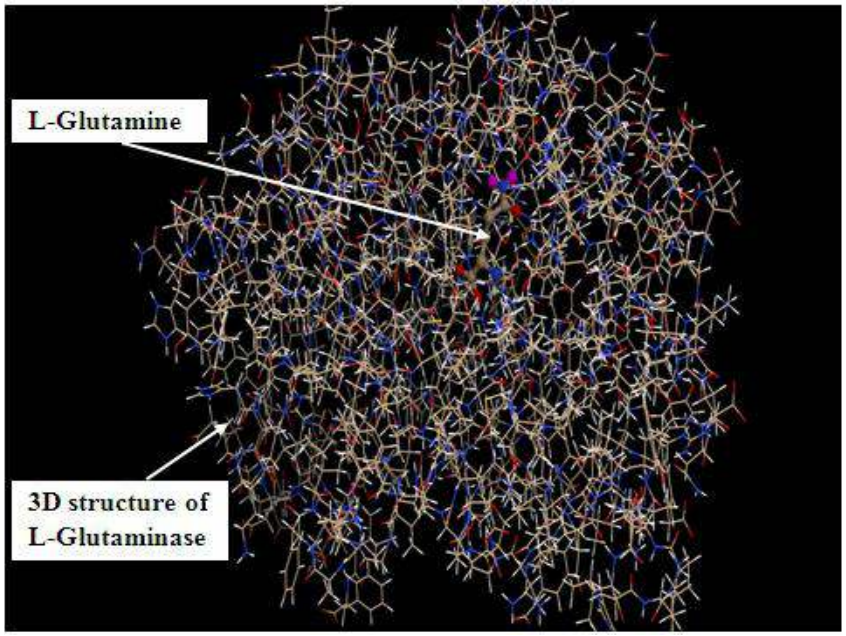


Fig.8.10.Validation of active site of 3D model structure of L-glutaminase from *Bacillus cereus* MTCC 1305 with L-glutamine

8.4. Conclusion

X-ray crystallographic structure of L-glutaminase from *Bacillus subtilis* (PDB: 1MKI_A) was used as template model to predict 3D model structure for L-glutaminase from *Bacillus cereus* MTCC 1305. The model structure was refined using ANOLEA, QMEAN, GROMOS 96, PROCHECK software tools. Ramachandran plot with distribution of 86.1% residue in favorable region shows that model structure possess considerably good geometry. Docking studies of the homology model showed amino acid viz., Asp56, Trp65, Glu66, Phe120, Glu142, Glu156, Tyr182, Asp212, Tyr225 and Phe261 as conserved residues in active site of this predicted model structure. Further analysis of docking energies indicated high affinity of this enzyme for substrate (L-glutamine) with favorable ΔG docking score.