ISSN: 0975-8585

Research Journal of Pharmaceutical, Biological and Chemical Sciences

Molecular modeling and structure based validation of H. influenzae KW20 RD proteins

Kasturi K [*], Kiran Kumar Reddy L [1], Nawaz Babajaan[2] Pavithra V[1] and Sambasiva Rao K R S [1]

[1,] *Department of Biotechnology, Acharya Nagarjuna University, Guntur, A.P, INDIA.
[2] Bioinformatics facility center, Dept of Biochemistry, Sri Krishnadevaraya University, Anantapur, Andhra Pradesh.

ABSTRACT

Haemophilus influenzae, formerly was called Pfeiffer's bacillus or Bacillus influenzae. It was first described in 1892 by Richard Pfeiffer during influenza pandemic. It is generally aerobic, but can grow as a facultative anaerobe. This organism belongs to Pasteurellaceae family. It is a non-motile, rod shaped and fastidious Gram-negative organism. Haemophilus influenzae most commonly causes ear, eye, sinus infections, and pneumonia. A more dangerous strain of the bacteria called Haemophilus influenzae type b causes meningitis and a life-threatening infection called epiglottitis. Comparative modelling is utilized to predict the 3-dimensional conformation of a given protein (target) based on its sequence alignment to experimentally determined protein structure (template). The metabolic database KEGG was used to screen for the various pathogenic proteins, the peptidoglycan synthesis pathway enzymes were screened and Glutamine synthase A (glnA) was selected for the present study. The glnA was modeled using MODELLER 9v1taking crystal structure coordinates of 2GLS and 1F0K as templates, the generated model was validated using the programs PROCHECK and ProSA-web. The validated model can further be utilized for docking analysis for generation of better drugs.

Keywords: Comparative modeling, MODELLER 9v1, KEGG, PROCHECK and ProSA-web.

Corresponding author

ISSN: 0975-8585

INTRODUCTION

Haemophilus influenzae, formerly called Pfeiffer's bacillus or Bacillus influenzae. It was first described in 1892 by Richard Pfeiffer during an influenza pandemic. It is generally aerobic, but can grow as a facultative anaerobe. This organism belongs to Pasteurellaceae family. it is a non-motile, rod shaped and fastidious Gram-negative organism [5], a coccobacillus that affects only humans. Haemophilus influenzae represents a group of bacteria that may cause different types of infections in children. There are six generally recognized types of H. influenzae: a, b, c, d, e, and f. Haemophilus influenzae b most commonly causes ear, eye, sinus infections, and pneumonia. The increasing resistance of Hib to antibacterial agents, such as ampicillin, co-trimoxazole, chloramphenicol and, recently, also to cefalosporins, has been reported from many parts of the world. The disease now occurs in less than 2 in 100,000 children. It still causes 5% - 10% of bacterial meningitis cases in adults. So far, bacterial strain replacement has not been a prominent feature of large-scale Hib immunization. The Hib immunization has to a larger extent decreased the occurrence of the disease in developed countries and still 3, 86,000 deaths occur every year in the developing countries. This has leaded our study to identify certain key pathogenic metabolic enzymes which can be identified as potent targets for development of effective drugs against the disease. KEGG [4] pathway database was used as a source of metabolic pathway information. The peptidoglycan synthesis [KO: K01915] was identified to be an effective pathway with numerous enzymes which are potent drug targets. The present study was aimed to obtain evaluated 3D model for potent targets of glnA, using 1F0K and 2GLS as templates.

Peptidoglycan is a polymer of sugars and amino acids that forms a homogeneous layer outside the plasma membrane of eubacteria. It serves a structural role in the bacterial cell wall, giving it shape and strength, as well as counteracting the osmotic pressure of the cytoplasm. The peptidoglycan layer is a crystal lattice formed by linear chains of N-acetylglucosamine and N-acetylmuramic acid that are connected by short [4 –5-residue] amino acid chains. The nature of the bridge is to some extent genus-specific [10].Glutamine synthetase [GS] [EC 6.3.1.2] is an enzyme that plays an essential role in the metabolism of nitrogen by catalyzing the condensation of glutamate and ammonia to form glutamine.

$$\text{Glutamate} + \text{ATP} + NH_3 \rightarrow \text{Glutamine} + \text{ADP} + \text{phosphate} + H_2O$$

There are three different classes of GS: Class I enzymes [GSI] are specific to prokaryotes, and are oligomers of 12 identical subunits. The activity of GSI-type enzyme is controlled by the adenylation of a tyrosine residue. The adenylated enzyme is inactive.

METHODOLOGY

SEQUENCE RETRIEVAL AND ALIGNMENT

The amino acid sequence of the target protein, H. influenzae glnA [gi. No. 16272805] was retrieved from the NCBI database and is composed of 472 residues. The search for the sequence similarities with several members of the family within the PDB database was performed with the pBLAST program [1]. The coordinates of the crystal structure of H. influenzae glutamine synthetase with PDB id 2GLS and 1F0K was identified to possess highest similarity to H.influenzae-glnA, they were chosen as templates as to build the initial model. The pair-wise sequence alignment was performed using CLUSTAL-W.

HOMOLOGY MODELING AND REFINEMENT of H.Influenzae-glna

The three dimensional model of H.influenzae-glnA was built using the crystal structure coordinates of Haemophilus influenzae-glnA [PDB id 2GLS and 1F0K]. All steps of homology modeling were performed using MODELLER 9v1 software, a Silicon Graphics workstation. This program is fully automated and constructs energy minimized protein models by satisfaction of spatial restraints extracted from the template PDB file[s]. In order to construct the homology model of H.influenzae-glnA an atom file, alignment file and steering file were generated and run through MODELLER 9v1, 100 runs were set to obtain the apt homology model of H.influenzae-glnA and further the structural refinement was done using Swiss-PDB Viewer.

EVALUATION OF H.influenzae-glnA

The 3-D model derived from homology modeling techniques was evaluated using the programs PROCHECK [6, 7] and ProSA-web [11]. PROCHECK was used to obtain Ramachandran plots for H.influenzae-glnA model to understand the distribution of ø/Ψ angles of amino acids. ProSA-web analysis of H.influenzae-glnA was used to obtain Z scores for study of native conformations.

RESULTS AND DISCUSSION

The amino acid sequence of H.influenzae-glnA [gi. No. 16272805] was retrieved from NCBI in FASTA format. It was submitted for blast-P search against PDB at NCBI [www.ncbi.nlm.nih.gov/blast.html]. The sequence was obtained containing 472aa and is shown in FASTA format in Figure1. Among the obtained blast-P search results, the crystal structure of glutamine synthetase from Haemophilus influenzae with PDB id 2GLS and 1F0K were identified to possess highest similarity to H.influenzae-glnA, they were chosen as templates in Figure2.

Sequence alignment was done taking the target and template sequences using CLUSTAL W with appropriate parameters as per the specified instructions. It aligns the sequences based upon identities, similarities and differences. The quality of alignment between the target and template sequences is the most important factor determining the accuracy of the homology model. The sequence alignment of Haemophilus influenzae-glnA with the sequence of 2GLS has shown the overall identity as 64% and with the sequence of 1F0K has shown the overall

ISSN: 0975-8585

identity as 63%. There are few gaps and variations in alignment of sequences which correspond to the sequences at the loops of structures.

Figure 1. Amino acid sequence of *Haemophilus influenzae* Glutamine synthetase in FASTA format.

```
MPNANAIANVFKLIEENNVKFVLLRFTDIKGKEHGVSIPVSLVDEDMFEDGKMFDGSSVEGWKTINKADMLLMPMAETAIVDPFA
QIPTLSIRCSVYEPTTMQSYDRDPRSIAIRAENYMRSTGIADQAFFGPEPEFFLFDDVRFNVSMNKASFSIDDIEAAWNTNKKYEEGN
NAYRPLKKGGYCAVAPIDSAHDIRSEMCLILEEMGLVIEAHHHEVATAGQNEIATKFNTLTLKADETQIYKHVVQNVALEHGKTACF
MPKPITGDNGSGMHCNMSLSKDGKNIFQGDKYAGLSETALYYIGGIIKHAKALNAFTNPSTNSYKRLVPGYEAPVLLAYSASNRSASI
RIPAVTNPKAIRVEARFPDPLANPYLAFAALLMAGLDGVVNKIHPGDAMDKNLYDLPPEELKDIPAVASSLEEALNSLEKDYEFLTQG
GVFAKDFIDAFISIKRKEVERLNMAPHPVEFEMYYA
```

Figure 2. Haemophilus influenzae with PDB id 2GLS and 1F0K were chosen as templates in Multiple Sequence Alignment

```
CLUSTAL 2.0.11 multiple sequence alignment

ppdn        MPNANAIANVFKLIEENNVKFVLLRFTDIKGKEHGVSIPVSLVDEDMFEDGKMFDGSSVE 60
2GLS        -----SAEHVLTMLNEHEVKFVDLRFTDTKGKEQHVTIPAHQVNAEFFEEGKMFDGSSIG 55
            : :*:.::*::**** ***** ****: *:**. *: ::**:********:

ppdn        GWKTINKADMLLMPMAETAIVDPFAQIPTLSIRCSVYEPTTMQSYDRDPRSIAIRAENYM 120
2GLS        GWKGINESDMVLMPDASTAVIDPFFADSTLIIRCDILEPGTLQGYDRDPRSIAKRAEDYL 115
        *** **::**:*** *.**::***  .** ***.: ** *:*.********* ***:*:

ppdn        RSTGIADQAFFGPEPEFFLFDDVRFNVSMNKASFSIDDIEAAWNTNKKYEEGNNAYRPLK 180
2GLS        RATGIADTVLFGPEPEFFLFDDIRFGASISGSHVAIDDIEGAWNSSTKYEGGNKGHRPGV 175
        *:***** .:***********:**..*:.:.:***** .***:..*** **:.**

ppdn        KGGYCAVAPIDSAHDIRSEMCLILEEMGLVIEAHHHEVATAGQNEIATKFNTLTLKADET 240
2GLS        KGGYFPVPPVDSAQDIRSEMCLVMEQMGLVVEAHHHEVATAGQNEVATRFNTMTKKADEI 235
        **** .*.*:***:********::*:****:**************:**:***:* ****

ppdn        QIYKHVVQNVALEHGKTACFMPKPITGDNGSGMHCNMSLSKDGKNIFQGDKYAGLSETAL 300
2GLS        QIYKYVVHNVAHRFGKTATFMPKPMFGDNGSGMHCHMSLAKNGTNLFSGDKYAGLSEQAL 295
        ****:**:*** ..**** *****: *********:***:*:*.*:*.********* **

ppdn        YYIGGIIKHAKALNAFTNPSTNSYKRLVPGYEAPVLLAYSASNRSASIRIPAVTNPKAIR 360
2GLS        YYIGGVIKHAKAINALANPTTNSYKRLVPGYEAPVMLAYSARNRSASIRIPVVASPKARR 355
        *****:******:**::**:***************:***** ********.*:.*** *

ppdn        VEARFPDPLANPYLAFAALLMAGLDGVVNKIHPGDAMDKNLYDLPPEELKDIPAVASSLE 420
2GLS        IEVRFPDPAANPYLCFAALLMAGLDGIKNKIHPGEPMDKNLYDLPPEEAKEIPQVAGSLE 415
        :*.***** *****.***********: ******:.*********** *:** **.***

ppdn        EALNSLEKDYEFLTQGGVFAKDFIDAFISIKRKEVERLNMAPHPVEFEMYYA- 472
2GLS        EALNALDLDREFLKAGGVFTDEAIDAYIALRREEDDRVRMTPHPVEFELYYSV 468
```

ISSN: 0975-8585

```
CLUSTAL 2.0.11 multiple sequence alignment

e.coli      ---MKNKKLLVMAGGTGGHVFPAIAVAQTLQKQEWDICWLGTKDRMEAQLVPKYGIPIRF 57
1F0K        ------KRLMVMAGGTGGHVFPGLAVAHHLMAQGWQVRWLGTADRMEADLVPKHGIEIDF 54

e.coli      IQISGLRGKGIKALLNAPFAIFRAVLQAKKIIQEEKPDAVLGMGGYVSGPAGVAAKLCGV 117
1F0K        IRISGLRGKGIKALIAAPLRIFNAWRQARAIMKAYKPDVVLGMGGYVSGPGGLAAWSLGI 114

e.coli      PIILHEQNAIAGLTNKLLGKIATCVLQAFPTAFPHAEVVGNPVREDLFEMPNPDIRFSDR 177
1F0K        PVVLHEQNGIAGLTNKWLAKIATKVMQAFPGAFPNAEVVGNPVRTDVLALPLPQQRLAGR 174

e.coli      EEKLRVLVVGGSQGARVLNHTLPKVVAQLADKLEFRHQVGKGAVEEVSQLYGEN-LEQVK 236
1F0K        EGPVRVLVVGGSQGARILNQTMPQVAAKLGDSVTIWHQSGKGSQQSVEQAYAEAGQPQHK 234

e.coli      ITEFIDNMAEAYAWADVVICRSGALTVCEIAAVGAAAIFVPFQHKDRQQYLNAKYLSDVG 296
1F0K        VTEFIDDMAAAYAWADVVVCRSGALTVSEIAAAGLPALFVPFQHKDRQQYWNALPLEKAG 294

e.coli      AAKIIEQADLTPEILVNYLKNLTRENLLQMALKAKTMSMPNAAQRVAEVIKQYSN----- 351
1F0K        AAKIIEQPQLSVDAVANTLAGWSRETLLTMAERARAASIPDATERVANEVSRVARAL--- 351

e.coli      ----
1F0K        ----
```

HOMOLOGY MODELING OF Haemophilus influenzae –glnA

Many runs of model building were carried out in order to obtain the most reasonable homology model using MODELLER 9v.1 software and the built model was obtained using Swiss-PDB Viewer [3]. The Haemophilus influenzae -glnA model was built using MODELLER 9v.1 by taking crystal structure coordinates of 2GLS and 1F0K as templates [12]. The MODELLER 9v.1 program uses the spatial constraints, determined from the crystal structures of template protein, to build a 3-D model of the target protein with unknown tertiary structure. 100 runs were carried out to obtain the most reasonable model and the built 3-D model of Haemophilus influenzae -glnA [9].

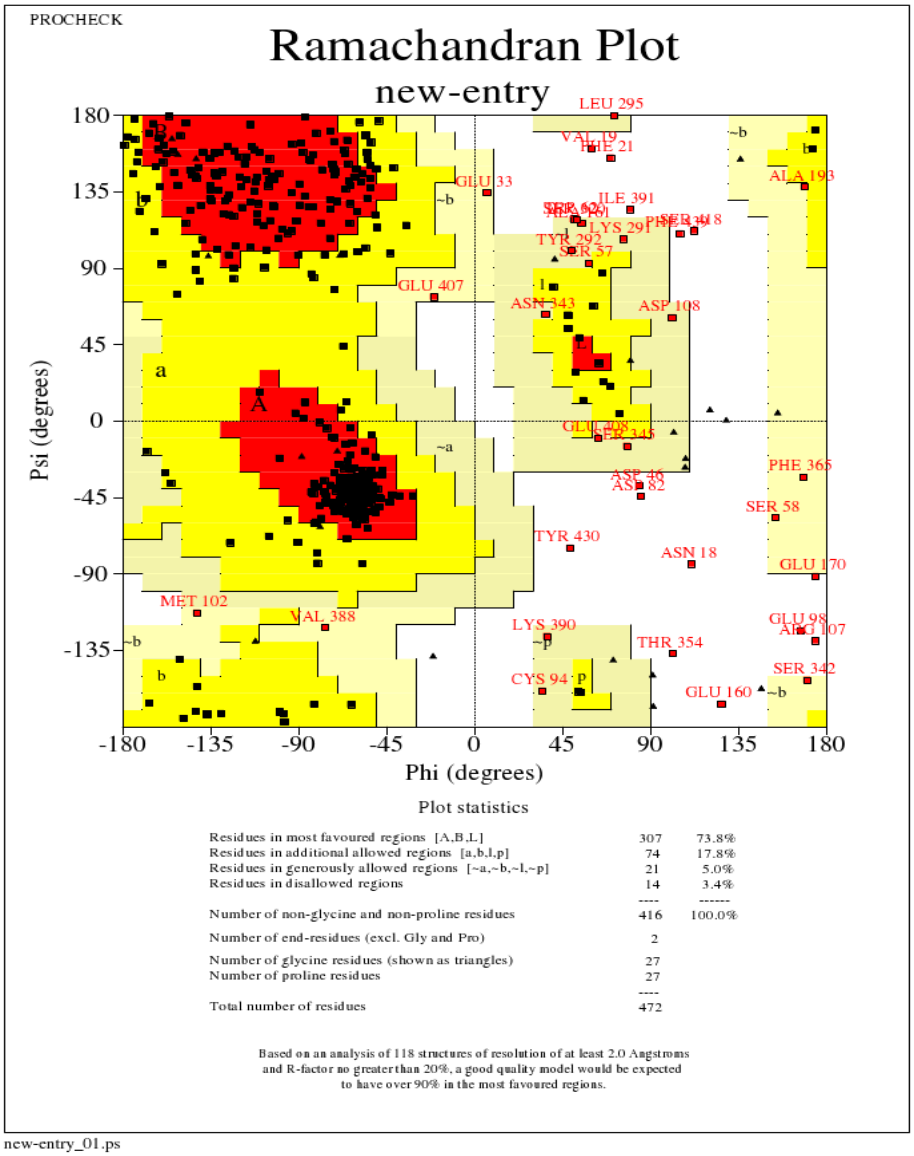
EVALUATION OF Haemophilus influenzae –glnA

The 3-D model derived from homology modeling techniques was evaluated using the programs PROCHECK and ProSA-web [8]. PROCHECK was used to obtain Ramachandran plots for Haemophilus influenzae -glnA model to understand the distribution of ø/Ψ angles of amino acids, the model generated using 2GLS has shown that 73.8% of the amino acids are in the most favored regions, 17.8% are in the additional allowed regions, 5.0% are in the generously

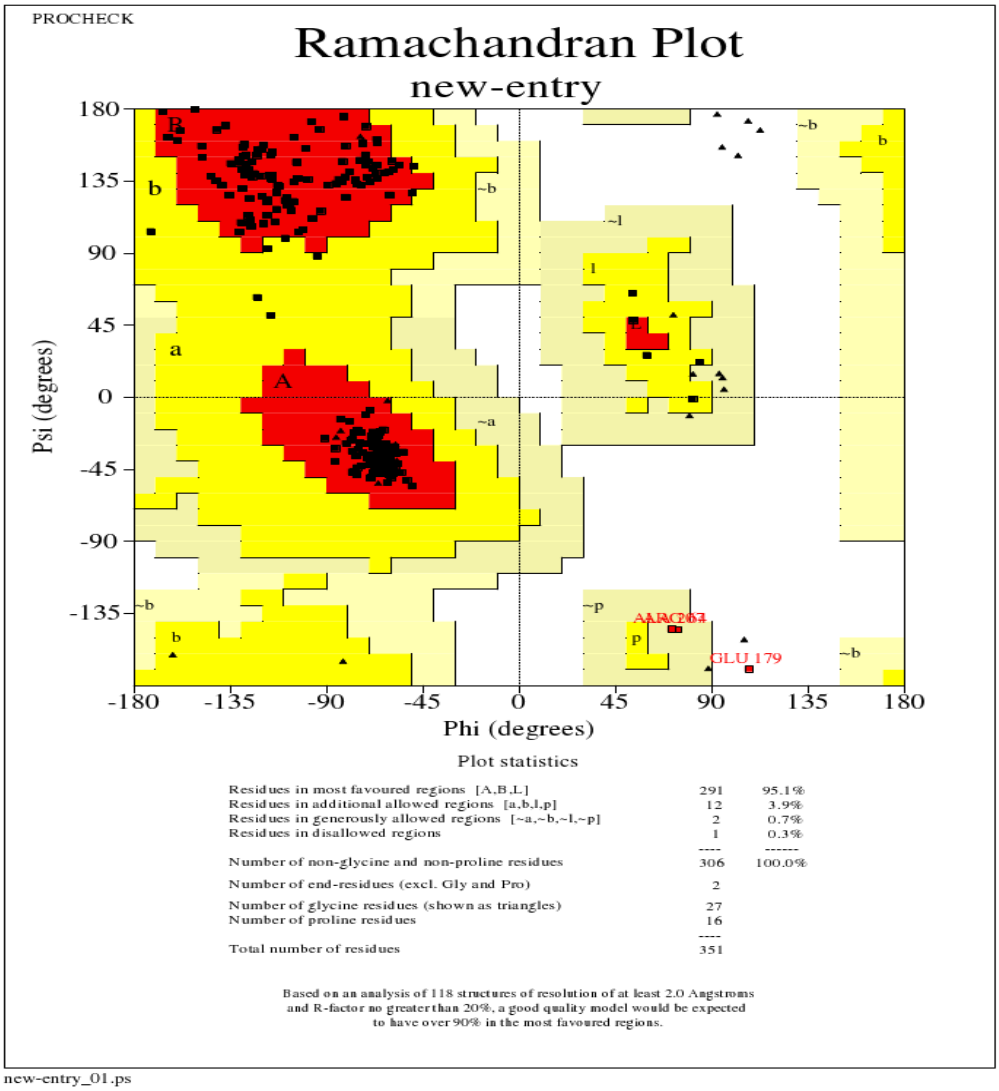
ISSN: 0975-8585

allowed region and 3.4% in the disallowed regions and angles of amino acids Figure. The model generated using 1F0K has shown that 95.1% of the amino acids are in the most favored regions, 3.9% are in the additional allowed regions, 0.7% are in the generously allowed region and 0.3% in the disallowed regions and angels of amino acids [Figures 3a and 3b], the results from the generated plots infer that further computational analysis can be carried out.

Figure 3a and 3b. Ramachandran plot of Haemophilus influenzae -glnA model showing all regions.



ISSN: 0975-8585



ProSA-web analyzes the energy distribution in protein structure as a function of sequence position to identity a structure as native-like or fault. The Figures 4a, 4b show the energy profile of the homology model of Haemophilus influenzae -glnA and a crystal structure. Overall quality or Z score of the 2GLS and Haemophilus influenzae -glnA reveals similar correlation with energy pattern between X-ray and NMR structures. The Z- score measures the deviation of the total energy of the structure with respect to an energy distribution derived from random conformations. The ProSA web analysis of Haemophilus influenzae -glnA model shows the Z-score of -7.92. This is within the range characteristic for native proteins indicating the good quality of the built model.

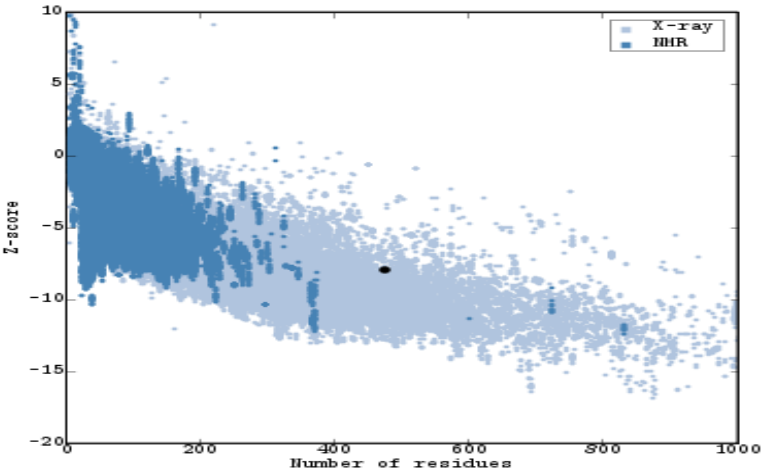
The Figure 4a, 4b shows the energy profile of the homology model of Haemophilus influenzae -glnA and a crystal structure 1F0K. Overall quality or Z score of the 1F0K and Haemophilus influenzae -glnA reveals similar correlation with energy pattern between X-ray and NMR structures. The ProSA web analysis of Haemophilus influenzae -glnA model shows the

ISSN: 0975-8585

Z-score of -7.92. This is within the range characteristic for native proteins indicating the good quality of the built model.

The Figure 4a, 4b shows the energy profile of the homology model of Haemophilus influenzae –glnA.
Results for ppdn.B99990001.pdb, chain blank (472 aa)

Overall model quality



Local model quality

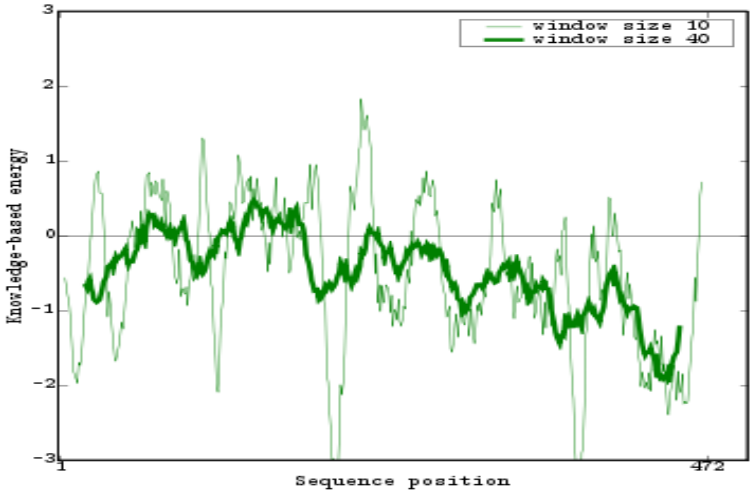
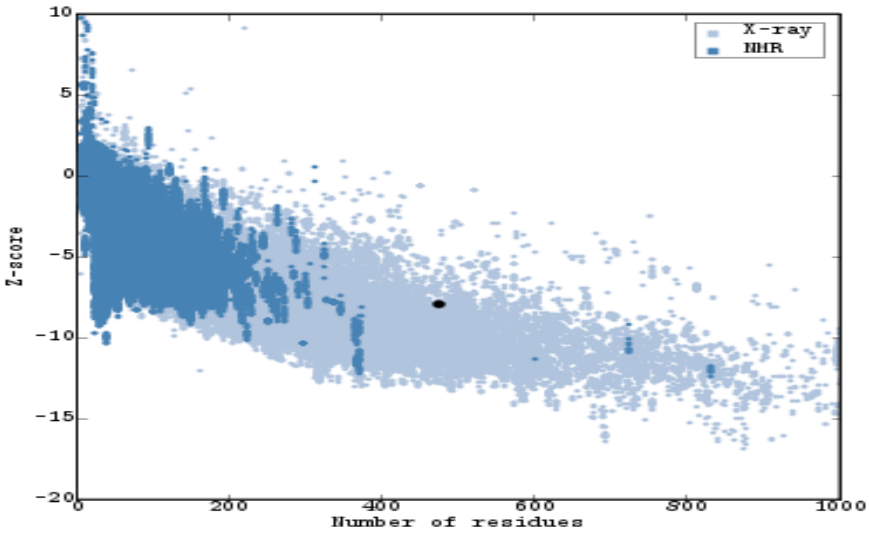


Figure 4a.

Overall model quality

ISSN: 0975-8585



Local model quality

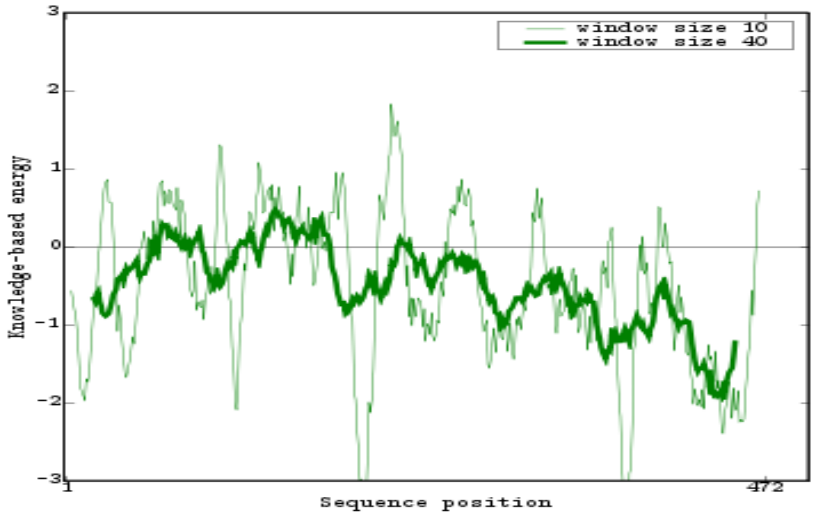


Figure 4b.

ACKNOWLEDGEMENTS

The authors are very much thankful to the Head of the Department, Department of Biotechnology, Acharya Nagarjuna University for providing the necessary facilities.

REFERENCES

[1] Altschul S F, Grisg W, Miller W, Myers E W and Lipman D J. J Bio 1990; 215: 403-410.
[2] Chenna R, Sugawana H, Koiket, Lopez R, Gibson T J, Higgins D G and Thompson J D. Nucleic Acids Res 2003; 31: 3497–3500.

[3] Guex N, Diemand A and Peitsch M C. Trends Biochem Sci 1999; 24: 364–367.
[4] Kanehisa M, Goto S, Kawashima S, Nakaya A. Nucleic Acids Res 2002; 30 (1): 42–46.
[5] Kuhnert P. Christensen H (editors). Pasteurellaceae: Biology, Genomics and Molecular Aspects. Caister Academic Press 2008.
[6] Laskowski R A, Chistyakov V V and Thornton J M. Nucleic Acids Res 2005; 33: D266-D268.
[7] Laskowski R A, MacArthur M W, Moss D S, J M and Thronton. J Appl Cryst 1993; 26: 283-291.
[8] Marti-Renom MA, Madhusudhan MS, Fiser A, Rost B, Sali A. Reliability of assessment of protein structure prediction methods, Structure (Camb) 2002; 10:435-440.
[9] May A C and Blundell T L. Curr Opin Biotech 1994; 5(4):355-360.
[10] Schon A, Bock A, Ott G, Sprinzl M and So¨ll D. Nucleic Acids Res 1989; 17: 7159–7165.
[11] Sippl MJ. 1993; Proteins 17: 355–362.
[12] Sali A, and Blundell TL. J Mol Biol 1993; 234 (3):779-815.