STRUCTURAL STUDIES OF ASPARTIC ENDOPEPTIDASE pep2 FROM NEOSARTORYA FISHERICA USING HOMOLGY MODELING TECHNIQUES

Maryam Jehangir[1] and Syed Farhan Ahmad[2]

[1]Department of Bio-informatics, Shaheed Benazir Bhutto Women University, Canal Road university town, Peshawar

maryam.bioinfo@yahoo.com

[2]Department of Biotechnology & Bioinformatics, Islamic International University, Islamabad

zaid_khan68@yahoo.com

ABSTRACT

Aspartic endo peptidase pep2 protein of neosartorya fisherica belongs from aspartic acid proteases that depend on aspartic acid residues for their catalytic activity. Catalytic mechanism has been proposed for the hydrolysis of proteins to peptides or free amino acids. It is accomplished the putative proteolytic activity of the conidial surface. Therefore, if it is not act like an allergen itself but on the conidial surface may well play an important role in the processing or signaling to allergens and causing different rare human infections such as (lung's aspergillosis and mycotic keratitis). To understand its different aspects, 3D structure model of (AEP pep2) was generated according to the structural coordinates of template (Proteinase A) with the help of MODELLER (9v8) software. To validate the energy and geometry in the model used PROCHECK and web server ProSA which determines that the predicted model to be stable and flexible.

KEYWORDS

Aspartic ebdopeptidase(pep2), Neosartorya fisherica, Modeller, Homology Modeling,proSA

1. INTRODUCTION

Neosartorya fischeris belongs from a filamentous fungi [1] and is a mold of the order Eurotiales, belonging to the family *Trichocomaceae*. The food-borne fungus *Neosartorya fischeri*, which is phenotypically associated to the human opportunistic pathogen Aspergillus fumigatus and mainly causes spoilage of food, which can pass through heat-processes techniques[2]. It produces mycotoxins of the fumitremorgin group (verrucologen, fumitremorgin A, B, C). When these mycotoxins are uptaken by an organism, it damages the central nervous system of pigs and sheep, whereby the mycotoxin verrucologen is the most effective one. Mostly caused rarely infections of humans and animals such as mycotic keratitis, mycotic endocarditis. Further cases were also seen such as an infection in a bone marrow transplant patient and as well as lung aspergillosis in a man after getting a liver transplant [3]. Some recent work done on *Neosartorya fischeri* proteins are peptide (NFAP) cysteine-rich protein acted as an antifungal and good inhibitor action against other filamentous fungi [4] as well as fungus *Neosartorya fischeri* strain 1008F produce two new compounds named fischeacid and fischexanthone can inhibit of the rapid growth of human gastric cancer cell and hepatic cancer cells [18]. The genome of *Neosartorya fischeri* consists of 80% intron and 47% coding gene have been studied [5].

Focusing towards the protein of interest *(Aspartic endopeptidase pep2)* belongs from aspartic proteinases. Aspartic proteinases are broadly dispersed among vertebrates, fungi, plants and retroviruses [19]. Aspartic proteinases are a group of proteolytic enzymes within which the peptide bond is attacked by a nucleophilic water molecule stimulate by two aspartic residues in a DTG motif by the side of the active site. The catalytic aspartic acid residues consist of the corners of two extended loops in the C- and N-terminal domains [20].The amino acid sequence of PEP2 consists of 398 amino acids. A signal sequence of 18 amino acids and a proregion of another 52 amino acids were analyzed. While the mature protein consists of 328 amino acids [6]. The endoprotease PEP2 is the protein which is capable of putative proteolytic activity of the conidial surface. Thus, if it is not an allergen itself but on the conidial surface may well play an important role in the processing or signaling to allergens. [7]

The method through which the 3-D model of *Aspartic endopeptidase pep* has been designed is referred as Homology modeling. This method calculates the three-dimensional structure of a given protein sequence based primarily on its sequence similarity to one or more proteins of defined structures. By determining distant homologues, aligning sequences with template structures, modeling of loops and side chains, as well as detecting errors in a model [21].

2. MATERIAL AND METHODS

2.1. Sequence Analysis

The primary amino acid sequence of target (*Neosartorya Fisherica* protein aspartic endopeptidase pep2) is access from the SWISSPROT (http://www.us.expasy.org) [8] or Unit Prot (http://www.uniprot.org/) by accession ID (A1D6T2).

2.2. Search the Template

 The homologous sequence of the target is searched via BLAST. In BLAST the Psi-BLAST algorithm is then chosen [9] against the Protein Data Bank (PDB) (10). Due to this procedure, the related sequences to target are extracted and choose that one sequence (template) which has high identity to target.

2.3. Multiple Sequences Alignment

Multiple sequence alignment was used with the help of program CLUSTAL X [11] with default parameters. Sequence's related to the target sequences were extracted from SWISS-PROT [8].

2.4. Phylogenatic Analysis

After the multiple sequence alignment, the tree was constructed which showed structural homology among the sequence in the family by the help of a matrix of pairwise similarity score. Phylogenetic relationship was studied through phylogeny Inference program Phylip (Version 3.6b) [12].

2.5. Secondary Structure Prediction

The Secondary structure prediction of target sequences was approved by submitting the sequences to the Consensus Secondary Structure Prediction Sever an http://pbil.ibcp.fr/NPSA/npsa_npsa.html [13] and PDB sum [14].

2.6. Model building and Refinement

Three-dimensional homology models of *aspartic endopeptidase pep 2* were constructed using the crystal coordinates of templates (proteinase A) assuming alignment between target and template

sequences. All steps of homology modeling and refinement were accomplished by the program MODELLER (Version 9 (9v8)) [15]. The model building process was completed in two steps.
a. Target-template alignment.
b. Model Building.

2.7. Model evaluation and Model visualization

Assessment of the predicted model concerned analysis of geometry, stereochemistry, and energy distributions in the models. Reliability of the homology model was carried out by the programs PROCHECK [16] (Version: 3.4) and web server ProSA. To examine the reliability of the alignment and modeling of variable surface loops, structural investigations on the graphics screen using 3D visualization programs, Ds-Viewer [17] be performed. Root Mean Square Deviation (RMSD) values were calculated between the set of targets and template protein to know how much modeled target deviates from the template protein structure.

2.8. Protein-Ligand interactions

Program Ligand Explorer (http://www.kukool.com/ligand) [22] was used to study the Protein–Ligand interactions.

2.8.1. LigPLOT

The program LigPLOT is used for analysis and automatically generates schematic 2-D representations of protein-ligand complexes from standard Protein Data Bank file input. The output is a color, or black-and-white [23].

2.9. ERRAT

ERRAT is a protein structure verification algorithm that is especially used for evaluating the progress of crystallographic model building and refinement. It can be retrieved online through http://nihserver.mbi.ucla.edu/ERRATv2/.[24]

3. RESULTS AND DISCUSSIONS

3.1. Sequence Analysis

A sequence homology search for the query aspartic endopetidase pep2 was carried out by using BLAST algorithm against Protein Data Bank (PDB) [10]. Crystal structures co-ordinates of Proteinase A complex with a IA3 mutant (1GOV.pdb) was chosen as a template as the basis of highest sequence similarity score and lowest E-value for building the predicted 3D structure of saccaropepsin (*aspartic endopeptidase pep2)*.1GOV (proteinase A complex with A IA3 mutant) of saccaromyces select as a template has 329 residues, and aspartic endopeptidase pep 2(target) have 398. Sequence similarity searches of AEP pep2 (SwissProt AC: A1D6T2) showed 71% identity with (proteinase A complex with A IA3 mutant) of saccaromyces cereviae (Swiss Prot AC: P07267). 230 amino acids (71%) out of 326 residues are identical and 282 out of 326(87%) are positives, which are physiochemical similar. While less E-value $2e^{-177}$ as compared to other sequences and used its coordinates for constructing the target 3D structure.BLAST result is represented in Figure 1.

```
> ☐ pdb|1GOV|A  S  Chain A, The Structure Of Proteinase A Complexed With A Ia3 Mutant,
Mvv
Length=329

 Score =  496 bits (1278),  Expect = 2e-177, Method: Compositional matrix adjust.
 Identities = 230/326 (71%), Positives = 282/326 (87%), Gaps = 1/326 (0%)

Query  73   HDVLVDNFLNAQYFSEISLGTPPQKFKVVLDTGSSNLWVPGSDCSSIACFLHNKYDSSAS  132
            HDV + N+LNAQY+++I+LGTPPQ FKV+LDTGSSNLWVP ++C S+ACFLH+KYD  AS
Sbjct  3    HDVPLTNYLNAQYYTDITLGTPPQNFKVILDTGSSNLWVPSNECGSLACFLHSKYDHEAS   62

Query  133  STYKANGTEFAIKYGSGELSGFVSQDTLQIGDLKVVKQDFAEATNEPGLAFAFGRFDGIL  192
            S+YKANGTEFAI+YG+G L G++SQDTL IGDL + KQDFAEAT+EPGL FAFG+FDGIL
Sbjct  63   SSYKANGTEFAIQYGTGSLEGYISQDTLSIGDLTIPKQDFAEATSEPGLTFAFGKFDGIL  122

Query  193  GLGYDTISVNKIVPPFYNMLEQGLLDEPVFAFYLGDTNKEGDN-SEASFGGVDKNHYTGE  251
            GLGYDTISV+K+VPPFYN ++Q LLDE  FAFYLGDT+K+ +N  EA+FGG+D++ + G+
Sbjct  123  GLGYDTISVDKVVPPFYNAIQQDLLDEKRFAFYLGDTSKDTENGGEATFGGIDESKFKGD  182

Query  252  LTKIPLRRKAYWEVDFDAIALGDNVAELENTGVILDTGTSLIALPSTLADLLNKEIGAKK  311
            +T +P+RRKAYWEV F+ I LGD  AELE+ G  +DTGTSLI LPS LA+++N EIGAKK
Sbjct  183  ITWLPVRRKAYWEVKFEGIGLGDEYAELESHGAAIDTGTSLITLPSGLAEMINAEIGAKK  242

Query  312  GFTGQYSIECDKRDSLPDLTFTLAGHNFTIGPYDYTLEVQGSCISSFMGMDFPEPVGPLA  371
            G TGQY+++C+ RD+LPDL F   G+NFTIGPYDYTLEV GSCIS+   MDFPEPVGPLA
Sbjct  243  GSTGQYTLDCNTRDNLPDLIFNFNGYNFTIGPYDYTLEVSGSCISAITPMDFPEPVGPLA  302

Query  372  ILGDAFLRKWYSVYDLGNNAVGLAEA   397
            I+GDAFLRK+YS+YDLGNNAVGLA+A
Sbjct  303  IVGDAFLRKYYSIYDLGNNAVGLAKA   328
```

Figure 1.Blast result of the selected template (1GOV) with Target (3DC4).

These two sequences are then aligned with the help of MODELLER, which works upon the principle of dynamic programming by MODELLER command align2d. After the execution of this command, insertions generate in the form of gaps at the start of the template (1GOV) sequence represented in the Figure 2. The (AEP pep2) structure has been modeled on the basis of (1GOV) template.

```
Identities = 230/326 (71%), Positives = 282/326 (87%), Gaps = 1/326
(0%)

1GOV   --------------------------------------------------------------
3DC4   MKSTSLLTASVLLGSASAAVHKLKLNKVPLDEQLYTHNIDAHVRALGQKYMGIRPNVHQELLEENSLN

1GOV   --GGHDVPLTNYLNAQYYTDITLGTPPQNFKVILDTGSSNLWVPSNECGSLACFLHSKYDHEASSSYK
3DC4   DMSRHDVLVDNFLNAQYFSEISLGTPPQKFKVVLDTGSSNLWVPGSDCSSIACFLHNKYDSSASSTYK
            ***   * *****   * ***** *** ***********   * * ***** ***   *** **

1GOV   ANGTEFAIQYGTGSLEGYISQDTLSIGDLTIPKQDFAEATSEPGLTFAFGKFDGILGLGYDTISVDKV
3DC4   ANGTEFAIKYGSGELSGFVSQDTLQIGDLKVVKQDFAEATNEPGLAFAFGRFDGILGLGYDTISVNKI
       ******** ** * *   ***** ****   ******** **** **** ************** *

1GOV   VPPFYNAIQQDLLDEKRFAFYLGDTSKDTENGGEATFGGIDESKFKGDITWLPVRRKAYWEVKFEGIG
3DC4   VPPFYNMLEQGLLDEPVFAFYLGDTNKEGDNS-EASFGGVDKNHYTGELTKIPLRRKAYWEVDFDAIA
       ******   * **** ******** *   * ** *** *    * *   *   *  ********* *  *

1GOV   LGDEYAELESHGAAIDTGTSLITLPSGLAEMINAEIGAKKGSTGQYTLDCNTRDNLPDLIFNFNGYNF
3DC4   LGDNVAELENTGVILDTGTSLIALPSTLADLLNKEIGAKKGFTGQYSIECDKRDSLPDLTFTLAGHNF
       ***   ****   *   ******* ***  **   * ******* ****   *   ** **** *   * **

1GOV   TIGPYDYTLEVSGSCISAITPMDFPEPVGPLAIVGDAFLRKYYSIYDLGNNAVGLAKAI
3DC4   TIGPYDYTLEVQGSCISSFMGMDFPEPVGPLAILGDAFLRKWYSVYDLGNNAVGLAEAK
       ***********  *****   ************  ******* ** *** *********** *
```

Figure 2. Target (AEP pep2)-Template (PA) alignment by using ALIGN2D command of Modeller (9v8). * and blue represents the conserved residues between the two sequences.

3.2. Multiple Sequence Alignment and Phylogenatic Analysis

For multiple sequence alignment, a result showed that the entire closely related homologous have less sequence identity but more physiochemical similar (positives). Most of the region is conserved in all of these homologues. To examine the phylogenatic analysis of aspartic endopeptidase pep2 to its peptidases A1 family,The phlogenatic tree and gram clearly indicates that AEP pep2 from *Neosartorya fischeri* (AID6T2 NEOFI) and CARP protein from Aspergillus fumigatus(CARP ASPFU) have sister taxas and arise from the same node mean closely evolutionary relation or originate from the same ancestor. While the template proteinase A from saccaromyces (CARP YEAST) have not more divergence to the target(AEP pep2) but have a same ancestor shown in the Figure 3.

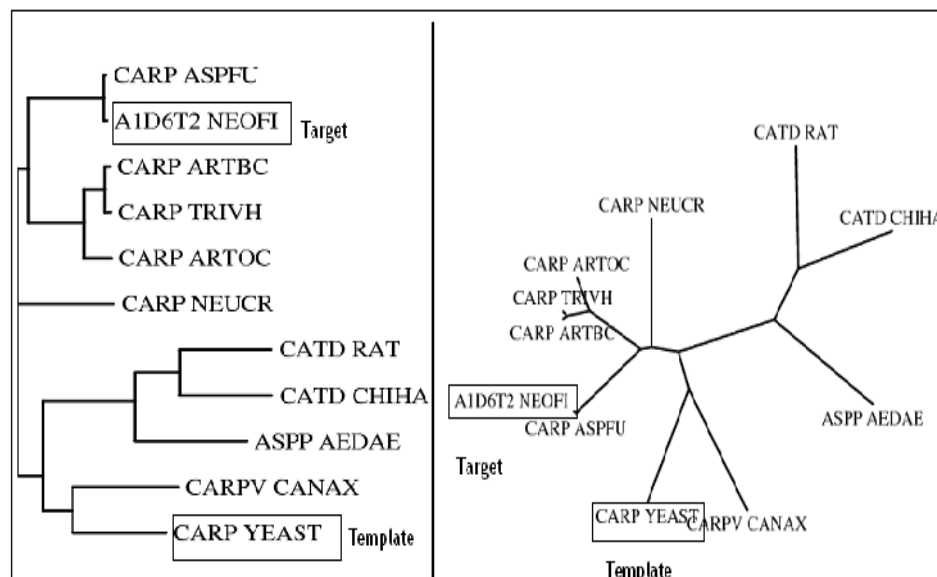Figure 4. Phylogenatic result thought draw gram and draw tree of *AEP* pep2 with the mostly homologous member of its family.

3.3. Topology of Predicted Homology of AEP pep2

The structure of AEP pep2 is bilobal which carry two lobes. Each of the lobe shows ∕ structure topology, and have five alpha helixes, which formed sandwich type structure with parallel and anti parallel beta sheets both in the C and N terminals clearly indicates in Figure 5 and 6.
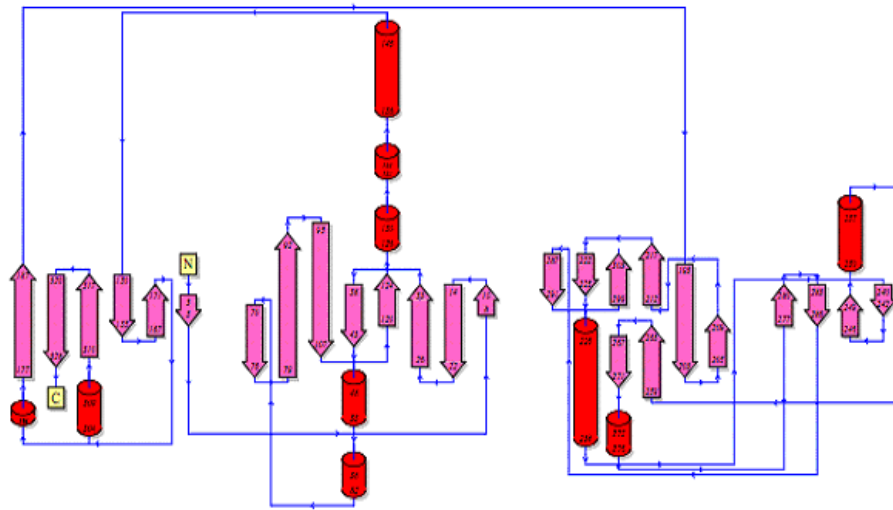
Figure 5 . ／ topology secondary structure of AEP pep2. The N terminal and C terminal are indicted as N and C.The red cylinders are alpha helixes while pink arrows are beta sheets .Blue lines shows the coils n turns.
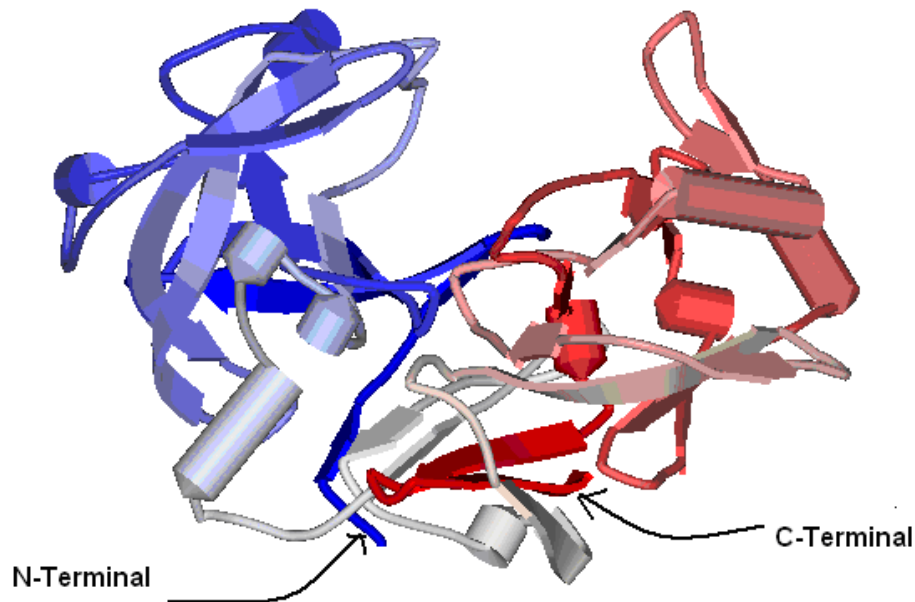

Figure 6. The schematic representation of the built homology model (3DC4) target showing arrangement of secondary structural elements. Blue region shows N-terminal domain and red regions indicates C-terminal domain, b-sheets by arrows and a-helices shows by cylinders.

It is characterized by pre-dominant 26 beta strands contents with 10 alpha helix's segments that form two similar domains. It consisted of 34 beta turns or tight loop,which often support the formation of antiparallel beta strands in this structure. It is stabilized by disulphide bonds between residues 46-51 and 251-284. The structure consisted of 2 gamma turns at position of 150 and 193 residues and 10 beta hairpins at the different position between the beta strands. The total number of sheets in the structure is A, B, C, D which topology is antiparallel and mixed indicates in Figure 7.
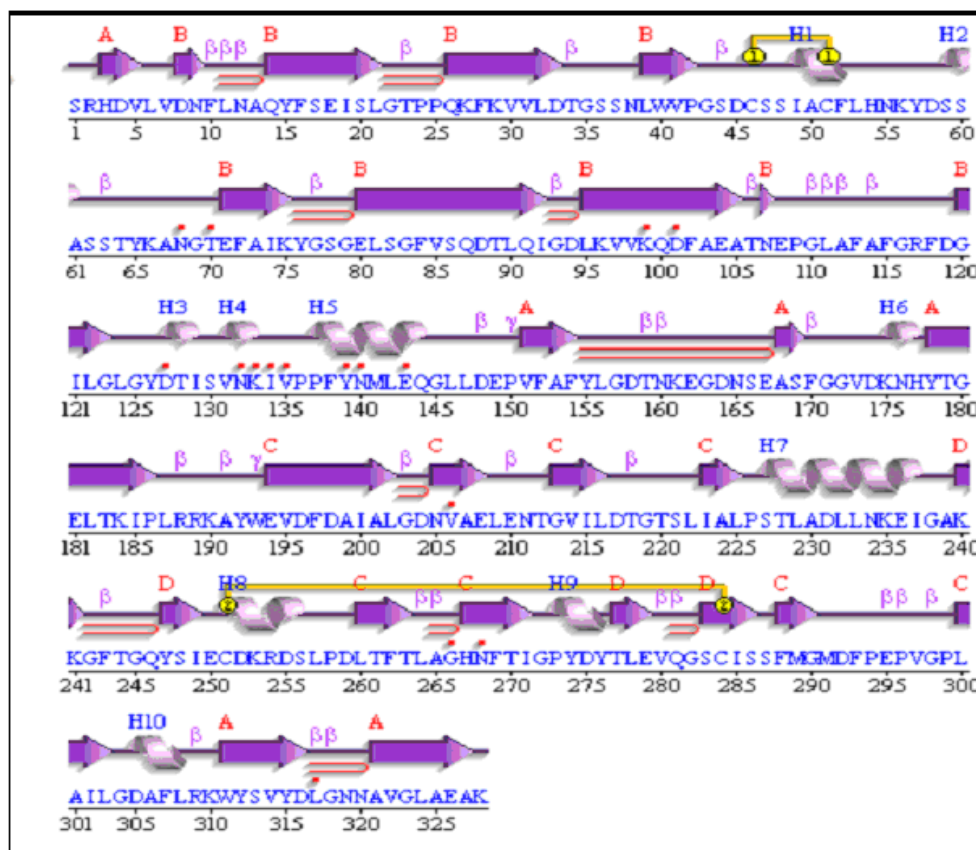
Figure.7 Overall secondary structure information of AEP pep2. (beta turn) (Gamma turns) (Helix and strand) (beta hairpin) (disulphide bond)

3.4. Model Assessment

Different validation tools are available for assessment of newly build model accessed from online source such as PROSA or the PROCHECK program was used to allow the model based on geometric examinations.

3.4.1. PROCHECK Results

The homology modeling of AEP pep2 was refined by $2.0A^0$ resolution as same as 1GOV (target) and passed over all the implementation steps of PROCHECK and give a result in the form of Ramachandran plot. The plot for AEP pep2 (3DC4) indicates that 94.3% of the main-chain dihedral angles are found in the most-favored regions, 4.3% in the additionally allowed, 1.4% in the generously allowed while there is no residue found in the disallowed regions indicates in Table 1 and Figure 8.

Table 1. Table shows Ramachandran plot statistics of the AEP pep2 models obtained with PROCHECK

```
+----------<<< P R O C H E C K    S U M M A R Y >>>----------+
|                                                            |
| 3DC419   2.0                                 328 residues |
|                                                            |
-| Ramachandran plot:   94.3% core   4.3% allow   1.4% gener    .0% disall |
|                                                            |
-| Gly & Pro Ramach:      1 labelled residues (out of  47)   |
-| Chi1-chi2 plots:       3 labelled residues (out of 187)   |
|                                                            |
| Main-chain params:    6 better    0 inside     0 worse    |
| Side-chain params:    5 better    0 inside     0 worse    |
|                                                            |
'| Residue properties: Max.deviation:    3.4        Bad contacts:    11 |
'|                     Bond len/angle:  20.2    Morris et al class:  1  1  2 |
-|     3 cis-peptides                                         |
| G-factors           Dihedrals:  -.07  Covalent:  -.44    Overall:  -.20 |
|                                                            |
'| M/c bond lengths: 83.1% within limits  16.9% highlighted    3 off graph |
'| M/c bond angles:  73.1% within limits  26.9% highlighted    5 off graph |
| Planar groups:   100.0% within limits   .0% highlighted   |
|                                                            |
+------------------------------------------------------------+
   + May be worth investigating further.  * Worth investigating further.
```
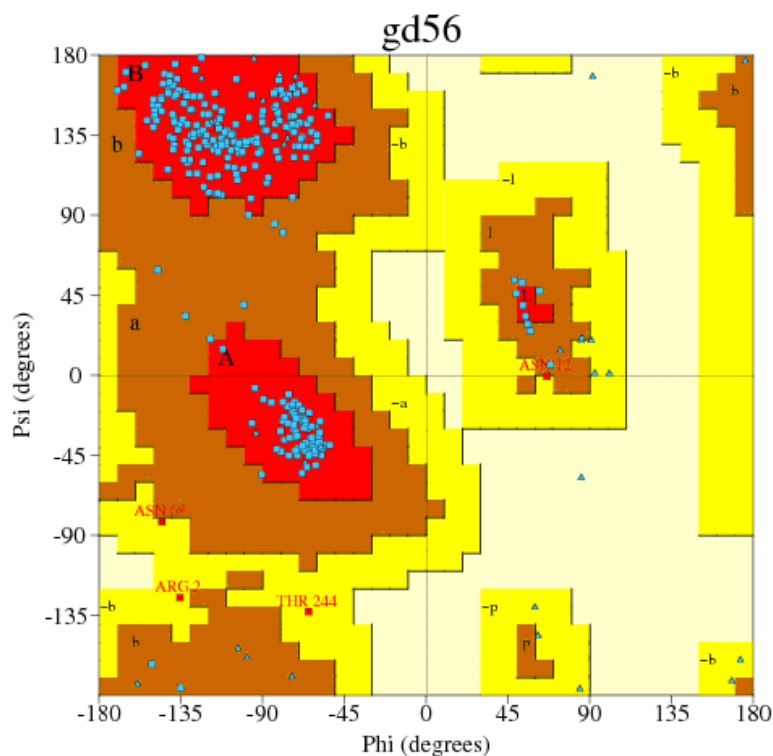

Fig 8. Ramachandran plot statistics of the target (AEP pep2) obtained through PROCHECK.

3.4.2 ProSA Results

While ProSA calculates an overall quality score of the newly build structure and demonstrate the result into a form of the plot that indicated the scores of all experimentally determined protein chains. The Z-score was used for the measurement of the energy as it indicated overall quality of the model. The positive Z-score shows that the structure is not stabilized while zero and negative score is indication of one of the best ideal structure. In the plot, groups of structures from

different sources (X-ray, NMR) are shown by different colors(light blue and dark blue).Z-score of the template sequence (1GOV) is -7.21 shown in Figure 9, whereas target sequence (3DC4) Z-score was -7.3. From this negative score of target demonstrated the model is structurally become reliable. While the graph in Figure 10 for target is mostly below from the zero on x-axis and on the basis of this graph, the modeled 3-d structure for target can be constructed.
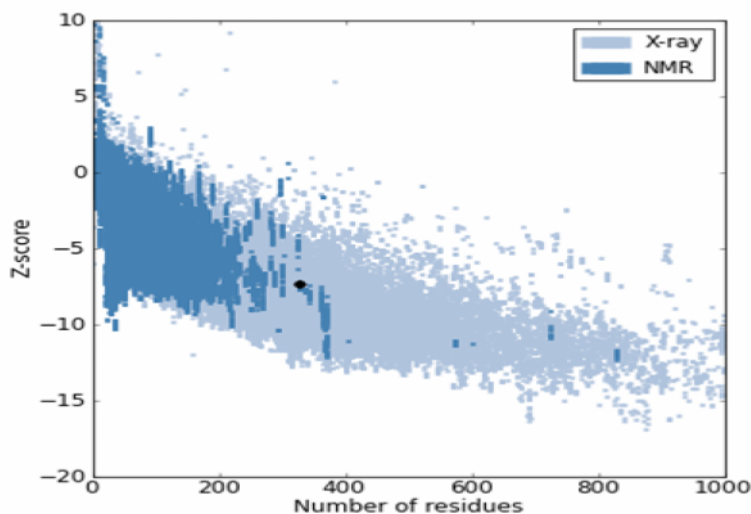

Figure 9. Screenshot of ProSA z-score plot of AEP pep2 showing the Z value < 0.neagative value -7.3
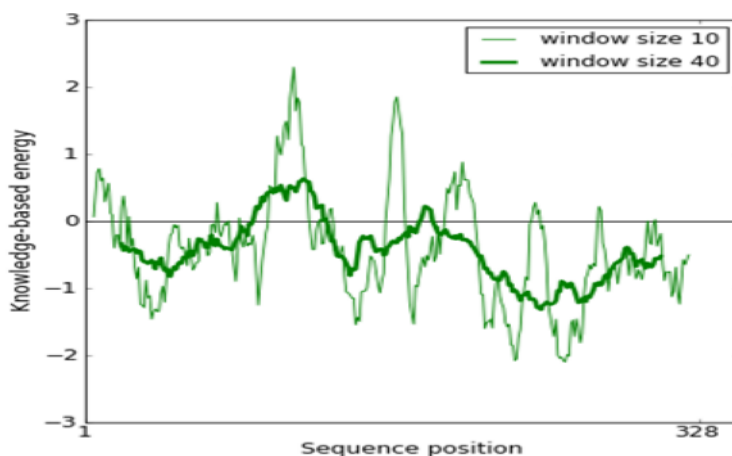
Figure 10. Screenshot of ProSA plot of AEP pep2 showing the energy graph of residue scores of a native protein structure.

3.4.3. Superposition

In order to find the backbone differences and similarities between the 3DC4 (target) and 1GOV(template)both the models were superposed on each other. The RMSD value of the superposed models was $3.3628A^{0}$. Due to higher DRSM value, the two structures are not well superposed and the distances between the two models are more than aspect range but have almost identical folding topology indicates in Figure 11. All the residues are mostly similar to each other.
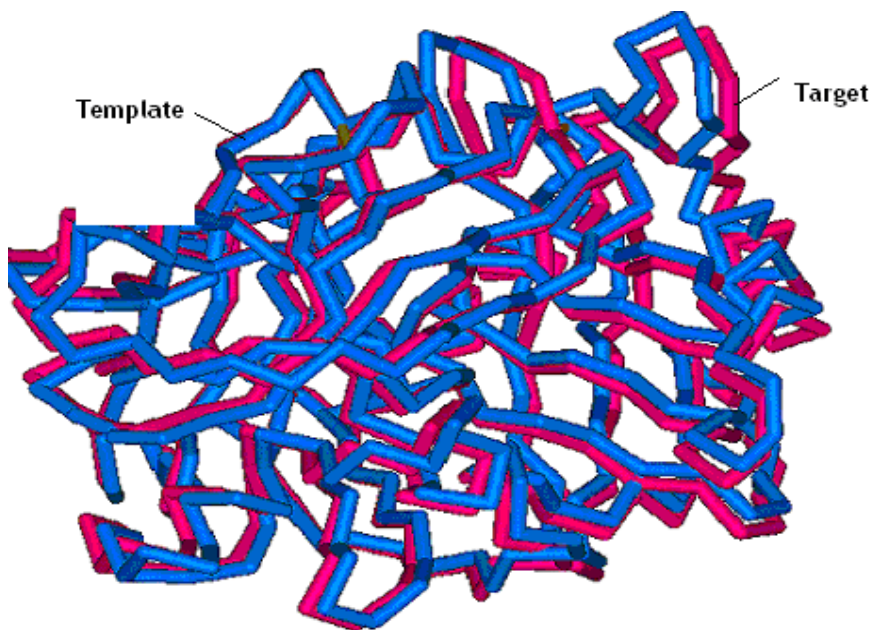
Figure 11. Structural superposition of atoms of AEP pep2 (pink) onto the crystal structure of 1GOV (blue).

3.5. Active site And Protein Ligand Interactions

In the target structure, there is a lot of ligands, which clearly shown by program LIGPLOT from Pdbsum. The ligands consist of carbohydrates such as BMA-MAN-MAN-BMA-NAG-NAG-MAN-MAN located at the different position in the structure. In NAG 333 bound with amino acid Asn132 through hydrogen bond of bond length 2.95 A^o. While MAN 336 bound with amino acid Lys 99 at 3.00 A^o and Lys 133, Val 135, Thr 70, etc are non-ligands residues involved in hydrophobic contacts represented by the eyelashes but located at the different position in the structure shown in Figure 12.
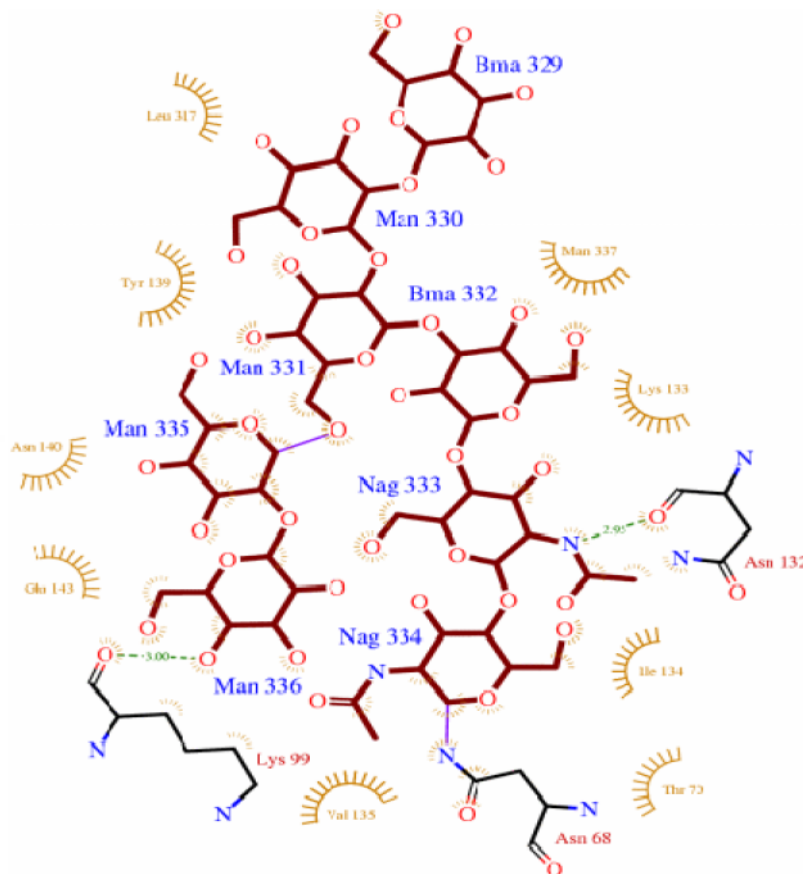
Figure 12. The ligplot result of interactions of different carbohydrates ligands such as NAG, MAN and BMA shows to interconnected with each other as well as to amino acids of target sequences. Hydrogen bonds are indicated by dashed lines between the atoms involved, while hydrophobic contacts are illustrated by spokes.

Active sites study to indicate that each domain which is characterized by two catalytically active aspartic acid's rasidues are,Asp 103 and Asp 287, one in each domain of protein.

3.6. ERRAT Model

In Errat model shows that the overall quality factor of target model is 71.384.which is identification of good model.
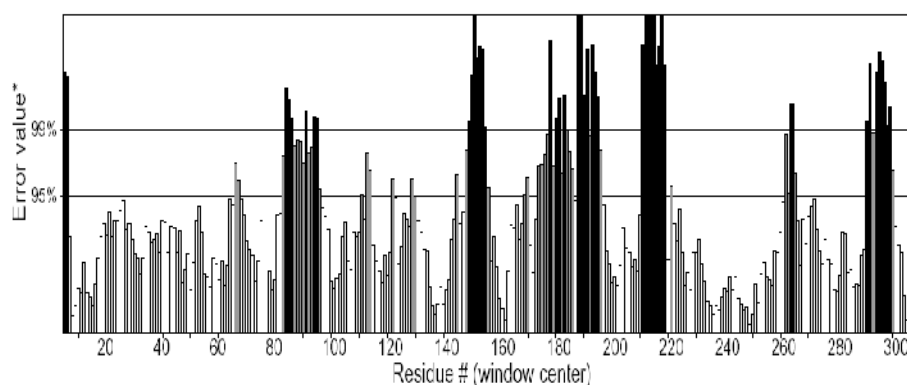
Figure.13 Errat model of AEP pep2 shows overall quality factor**: 71.384 and comparing it to the amount of errors (99%) allowed in a model. The *white* and *gray bars* indicate range of good refinement in terms of favorable nonbonded contacts.While black coloumns identify problem regions.

4. CONCLUSION

The primary structure of both *Neosartorya fisherica* proteins indicates strong amino acid homologies in the regions that are thought to be the active site such as aspergine,threonine,glycine(DTG) motif. Multiple sequences' comparison of (10 sequences) shows considerable sequence similarity among primary structures of all known Aspartic proteases of *A1peptidase* family. The modeled structure is characterized by pre-dominant beta sheets with 10 alpha helixe's segments that form two similar lobes, each providing one of the two aspartic acid residues, which are catalytically important. The overall topology and secondary structural elements are entirely conserved throughout the A1 family of peptidase**.** The tertiary structure shows that each lobe has   /   structural topology, mainly (mixed) parallel, antiparallel   - sheets sandwich between alpha helices providing characteristic functional geometry to the protein. A phylogenetic tree results show that *Aspartic endopeptidase pep2* and CARP *ASPFU (Asperigellus fumigatus)* is evolved from common ancestors and are homologous to each other according to evolutionary point of view. The predicted model of aspartic endopeptidase pep2 will be helpful better to understand the active site and ligand binding site for control of diseases causing from this protein for future drug analogs with more efficacies as well as less side effects.

ACKNOWLEDEMENT

REFERENCE

1. Johns, M.J., (2007) "The first filamentous fungal genome sequences: Aspergillus leads the way for essential everyday resources or dusty museum specimens?", *Microbiology,* Vol. 153, pp1-16.
2. Girardin, H. & Monod, M. & Latgé, J.P., (1995) "Molecular Characterization of the Food-Borne Fungus", *Appl Environ Microbio,* Vol. 61, No. 4, pp1378-1383.
3. Bundesamt Fur Verbraucherschutz und Lebensmittelsichereit. (2006) "Position statement of the ZKBS on risk assessment of *Neosartorya fischeri*,*Aspergillus amstelodami* and *Penicillium commune* according to the Genetic Engineering Safety Regulations (GenTSV)", Ref No. 6790-05-03-32.

4. Kovács, L. & Virágh, M. & Takó, M. & Papp, T. & Vágvölgyi, C. & Galgóczy, L., (2011) "Isolation and characterization of Neosartorya fischeri antifungal protein (NFAP)", *Peptides*. Vol. 32, No. 8, pp1724-1731.

5. Fedorova, N.D. & Khaldi, N. & Joardar, V.S. & Maiti, R. & Amedeo, P., (2008) "Genomic Islands in the Pathogenic Filamentous Fungus Aspergillus fumigatus", *PloS Genet,* Vol. 4, No. 4, ppe1000046.

6. Reichard, U. & Cole, G.T. & Rüchel, R. & Monod, M., (2000) *"*Molecular cloning and targeted deletion of PEP2 which encodes a novel aspartic proteinase from Aspergillus fumigatus", *Int J Med Microbiol*, Vol. 290, No. 1, pp85-96.

7. Abdul, R. & Michael, A.O. & Victor, W. & BirgitRiemenschneider, A. & Monod, A,. & Reichard, U., (2006) "Proteome of Conidial Surface Associated Proteins of Aspergillus fumigatus Reflecting Potential Vaccine Candidates and Allergens" *,* Vol. 5, No.4, pp954-962.

8. Bairoch, A. & Apweiler, R. & Wu, C.H. & Barker, W.C. & Boeckmann, B. & Ferro, S. & Gasteiger, E. & Huang, H. & Lopez, R. & Magrane, M. & Martin, M.J. & Natale, D.A. & O'Donovan, C. & Redaschi, N. & Yeh, L.S., (2005) "The Universal Protein Resource(UniProt)", *Nucleic Acids Res,* Vol. 33, ppD154-159.

9. McGinnis, S. & Madden, L.T., (2004) "BLAST: at the core of a powerful and diverse set of sequence analysis tools", *Nucleic Acids Res,* Vol. 32, ppW20-25.

10. Berman Helen, M. & Westbrook, J. & Feng, Z. & Gilliland, G. & Bhat, T.N. & Weissig, H. & Shindyalov, I.N. & Bourne, P.E., (1997) "The Protein Data Bank", *NucleicAcids Res,* Vol. 8, pp235-242.

11. Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG, (1997) "The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools", *Nucleic Acids Res,* Vol. 25, pp4876-4882.

12. Felsenstein, J., (1992) "Estimating effective population size from samples of sequences:inefficiency of pairwise and segregating sites as compared to phylogeneticestimates", *Genet Res*, Vol. 59, pp139-147.

13. UniProt Consortium, (2008) "The universal protein resource (UniProt)", *Nucleic AcidsRes,* Vol. 36,ppD190-195.

14. Laskowski, R.A. & Chistyakov, V.V. & Thornton, J.M., (2005) "PDBsum more: new summaries and analyses of the known 3D structures of proteins and nucleic acids" *,* Vol. 33, pp266-268.

15. Eswar, N. & John, B. & Mirkovic, N. & Fiser, A. & Ilyin, V.A. & Pieper, U. & Stuart, A.C. & Marti-Renom, M.A. & Madhusudhan, M.S. & Yerkovich, B. & Sali, A., (2003) "Tools for comparative protein structure modeling and analysis", *Nucleic Acids Res*, Vol. 31, pp3375-3380.

16. Laskowski, R.A. & Rullmann, J.A. & MacArthur, M.W. & Kaptein, R. & Thornton J.M., (1996) "AQUA and PROCHECK-NMR: programs for checking the quality ofprotein structures solved by NMR", *J Biomol NMR*, Vol. 4, pp477-486.

17. Kelly, B., (2004) "Structure Validation", *Crystallography Journal Club*.

18. Tan, Q.W. & Ouyang, M.A. & Shen, S. & Li,W., (2011)"Bioactive metabolites from a marine-derived strain of the fungus Neosartorya fischeri",.*Nat Prod Res*, Vol. 26, pp1402-1407. doi: 10.1080/14786419.2011.592834.

19. Charity, L.P. & Robert, A.B.K. & Brian, C.B. & Masahiro, O. & Rickey, Y., (2007) "The structure and function of Saccharomyces cereviae proteinase A", *Yeast,* VOL. 24: 467–480.

20. Sang, W.C. & Kim, N. & Choi, M. & Shin, W., (2001)"Structure of aspergillopepsin I from Aspergillusphoenicis: variations of the S1'-S2 subsite inaspartic proteinases", *Biological Crystallography,* VOL.57, pp.948-956.

21. Xiang, Z., (2006)"Advances in Homology Protein Structure Modeling," *Curr Protein Pept Sci*. VOL. 7, pp. 217-227.

22. Ligand Explore Home Page. http://www.kukool.com/ligand.

23. Laskowski, R.A. & Swindells, M.B.,(2011)"LigPlot: multiple ligand-protein interaction diagrams for drug discovery", *J Chem Inf Mode*l., VO. 51, pp.2778-2786.

24. http://nihserver.mbi.ucla.edu/ERRATv2

Authors

Maryam Jehangir   I have completed my BS honors in bioinformatics from Shaheed Benazir Bhutto Women University Peshawer, Pakistan. I am strongly determined to perform research and creative tasks in the field of bioinformatics and wish to pursue higher education in the same field which will assist me to serve humanity specifically my country in a positive way.

Syed Farhan Ahmad   I have done BS in biotechnology from University of Malakand, Pakistan and recently completed MS in biotechnology from Islamic International University Islamabad, Pakistan. I have always enjoyed biology and specifically human molecular genetics, but my inspiration to pursue research came from my internship at Institute of biomedical and genetic engineering, Dr. Abdul Qadeer Khan Research laboratories Islamabad, Pakistan. Over the past 6 years of my academic career, I have had a wide array of biotechnology courses which have expanded my knowledge of molecular biology, genetics, microbiology and biochemistry. I am looking forward to pursue doctorate degree in biotechnology.