SANCHO SCIENCE

Journal of Bioinformatics & Research 1(1) 2012: 15-20

Research Article
Open Access

ISSN: 2277-3363

Influence of Blast, Fasta and Wu-Blast algorithms on sequence alignments and 3-D structure prediction of DPP-IV

P. Kanchanamala[1], Allam Appa Rao[2], P. Srinivasa Rao[3], G. R. Sridhar[4]

[1]*Department of Information Technology, Dadi Institute of Engineering & Technology, NH-5, Anakapalle, Visakhapatnam, India*
[2]*Vice-Chancellor, Jawaharlal Nehru Technological University Kakinada, Kakinada, India*
[3]*Prof, Department of CS & SE, College of Engineering, Andhra University, India*
[4]*MD, Endocrine & Diabetes Centre, Visakhapatnam, India*

ABSTRACT

CD26/Dipeptidyl Peptidase IV (DPPIV) is a 110-kDa glycoprotein that is expressed on numerous cell types and has multiple biological functions. DPPIV is highly expressed on endothelial cells, epithelial cells and lymphocytes. DPPIV is involved in the regulation of several important physiological processes such as immune system, cell adhesion and glucose homeostasis. DPPIV inhibitors are thought to improve glycemic control in early stage Type II diabetes by reducing DPPIV-mediated inactivation of GLP-1. Sequence analysis and homology modeling studies of DPP-IV was carried out using NCBI BLAST, Fasta and Wu-Blast programs. ClustalW multiple alignment was done to identify highly conserved residues in DPP-IV on one hand and functionally important residues or mutations are identified by comparing multiple alignments with 3-dimensional structures of proteins from pdbsum database. Homology modeling routine was employed using Modeller 8.1 software.

KEYWORDS: sequence analysis, alignments, blast, fasta, wu-blast, multiple alignments, DPP-IV

INTRODUCTION

Dipeptidyl peptidase (DPP) IV is a ubiquitous type II transmembrane glycoprotein and a serine protease of the S9 prolyl-oligopeptidase family. The ubiquitous DPPIV glycoprotein has proved interesting in the fields of immunology, endocrinology, haematology and endothelial cell and cancer biology and DPPIV has become a novel target for Type II diabetes therapy [1]. DPPIV also has its role in T-cell costimulation, chemokine biology, type-II diabetes and tumor biology [2].

The dipeptidyl peptidase IV gene family contains the four peptidases dipeptidyl peptidase IV, fibroblast activation protein, dipeptidyl peptidase and dipeptidyl peptidase. Dipeptidyl peptidase IV and fibroblast activation protein are involved in cell–extracellular matrix interactions and tissue remodeling. Dipeptidyl peptidase IV is dysregulated in chronic liver disease [3]. CD 26, or dipeptidyl peptidase 4 (DPP-4) is a membrane-associated peptidase of 766 amino acids that is widely distributed in numerous tissues. DPP-4 also exists as a soluble circulating form in plasma and significant DPP-4-like activity is detectable in plasma from humans and rodents [4]. In this paper, an attempt has been made to build the 3-D structure of DPP-IV using sequence analysis and homology modeling routines.

*Corresponding Author: *kanchanakanta@gmail.com*

MATERIALS AND METHODS

Sequence analysis

Sequence analysis was performed using pairwise and multiple sequence alignments. NCBI- BLAST [5], EBI-BLAST [6] and EBI-FASTA [6] tools are utilized for pairwise alignments and multiple alignments were done using the EBI-CLUSTAL W tool [7]. The emphasis of this work is to find regions of sequence similarity, which in other words allows us to yield functional and evolutionary relationships among proteins considered for the study. Based on the sequence similarities, influence of PAM and BLOSUM matrices were studied to reveal identities, similarities and dissimilarities between proteins under study.

Protein Sequence Selection

About six DPP IV protein sequences were extracted from SWISS PROT [8] database with accession numbers, P81425, Q9N2I7, P27487, P28843, P22411 and P14740. BLAST program, protein - protein blastp, from NCBI was selected to scan the query protein sequence against PDB structure database [9]. The query sequence in fasta format was subjected to blast, fasta and wu-blast analysis using default parameters. Further, based on the output, various matrices were changed and the results are tabulated. The scoring matrices for all the three sequence analysis tools viz. FASTA, WU-Blast2 and BLAST were compared and the common matrices among them were identified and selected. Five matrices that are common among the three tools like, Blosum 50, 62, 80 and PAM 120, 250 matrices were selected to

study FASTA and WU-Blast2 whereas Blosum 45, 62, 80 and PAM 30, 70 were selected for NCBI BLAST.

Clustal W

ClustalW multiple sequence analysis is performed to determine the number of proteins that share common structural and functional features. As an input to clustalw all sequences in fasta formats are supplied with default options. The output is analyzed for sequences that are aligned for the complete length, scores, alignment, conserved residues, substituted and semi-conserved substituted residue patterns.

3-D structure prediction

3-dimensional structure of DPP-IV was built using Modeler 9v1 [10] software with default parameters.

RESULTS AND DISCUSSION

Six proteins when subjected to blast analysis resulted in hits with 100% identities with DPP-IV structures from protein data bank except P81425 and Q9N2I7 (Table-1). Q9N2I7 was considered for further analysis as the overlap residues are more than P81425. A search against a protein database yielded several alignments using five scoring matrices and the scores along with amino acid overlaps accompanying these alignments are used to distinguish sequences related to degree of divergence. The outputs of each scoring matrix run with BLAST, FASTA and WU-Blast2 are given in Tables-2-5 respectively.

Table 1: NCBI-blast result showing pdb hits, %identities, score, and gaps for all the sequences.

S. No	SWISS PROT ID	% ID	% SIM	No. OF GAPS	E-VALUE	SCORE	PDB ID	OVERLAP
1	P81425	91	95	0	0.0	1378	10RV_A	38-764
2	Q9N2I7	87	94	0	0.0	1358	2BGR_A	29-765
3	P27487	100	100	0	0.0	1513	2BGR_A	29-766
4	P28843	90	94	0	0.0	1362	2GBC_A	38-759
5	P22411	100	100	0	0.0	1489	10RV_A	39-766
6	P14740	100	100	0	0.0	1490	2GBC_A	38-767

Table 2: NCBI-blast result showing %id, score, and gaps with different matrices.

S. No	MATRIX	% ID	% SIM	No. OF GAPS	E-VALUE	SCORE	PDB ID	OVERLAP
1	PAM 30	87	94	0	0.0	1997	2BGR_A	29-765
2	PAM 70	87	94	0	0.0	1764	2BGR _A	29-765
3	BLOSUM80	87	94	0	0.0	1580	2BGR _A	29-765
4	BLOSUM62	87	94	0	0.0	1358	2BGR _A	29-765
5	BLOSUM45	87	94	0	0.0	1162	2BGR _A	29-765

Table 3: WU-blast result showing the pdb hits, %identities, score and e-values for Q9N217

S. No	MATRIX	% ID	% SIM	No. OF GAPS	E-VALUE	SCORE	PDB ID	OVERLAP
1	PAM30	87	91	1	0	4960	2BGR_A	29 - 765
2	PAM70	87	93	1	0	4432	2BGR_A	29 - 765
3	PAM120	87	95	1	0	3666	2BGR_A	29 - 765
4	PAM250	87	95	1	0	3588	2BGR_A	29 - 765
5	BLOSUM50	87	94	1	0	4552	2BGR_A	29 - 765
6	BLOSUM62	87	94	1	0	3589	2BGR_A	29 - 765
7	BLOSUM80	87	94	1	0	5785	2BGR_A	29 - 765
8	BLOSUM45	87	94	1	0	4272	2BGR_A	29-765

By observing the above table, it can be concluded that almost all matrices resulted in similar identities and similarities but the differed in the score values. And hence, PAM30 matrix reported with high score and low e-values. On the other hand, this table indicates that all these proteins belonging to different species have great similarities with query sequence.

In this case, the percentage identities and similarities are matched to the maximum extent, BLOSUM-80 matrix with high score and low e-value was selected. From fasta result, PAM 250 was selected even though the score was lower than the remaining due to the absence of gaps in the alignment.

Table 4: FASTA result showing pdb hits, %identities, score and e-values for Q9N217

S. No	MATRIX	% ID	% SIM	No. OF GAPS	E-VALUE	SCORE	PDB ID	OVERLAP
1	BLOSUM50	87	96	1	0	4549	2BGR_A	29 – 765
2	BLOSUM62	87	96	1	0	3586	2BGR_A	29 – 765
3	BLOSUM80	87	96	1	0	3812	2BGR_A	29 – 765
4	PAM120	87	96	1	0	3661	2BGR_B	29 – 765
5	PAM250	87	96	0	0	3585	2BGR_B	29 – 765

Table 5: Comparative data of methods vs matrices

METHOD	MATRIX	% ID	% SIM	No. OF GAPS	E-VALUE	SCORE	PDB ID	OVERLAP
BLAST	PAM30	87	94	0	0	1997	2BGR_A	29 - 765
FASTA	PAM250	87	96	0	0	3585	2BGR_B	29 – 765
Wu blast	BLOSUM80	87	94	1	0	5785	2BGR_A	29 – 765

Therefore, from the final comparative result reported in table-5, PAM 250 matrix of fasta method was selected to study multiple alignments due to the absence of gaps in the alignment and high percent similarity than blast result.

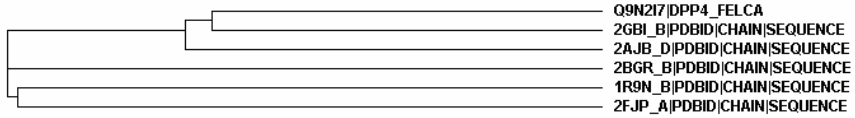
The default scoring matrix in ClustalW was Gonnet 250 that uses an iterative classical distance measures to estimate alignments of proteins [11]. A multiple sequence alignment was done with ClustalW for 6 hits resulted from FASTA with the matrix PAM250. All the structural sequences were downloaded from protein data bank. The program was run with default parameters like, a value of 10 for gap open penalty and a value of 0.05 and 8 for extending a gap and separating a gap respectively. The multiple alignments are given below.

Figure 1: Multiple alignments of DPP-IV query sequence vs 6 proteins

```
CLUSTAL W (1.83) multiple sequence alignment


1R9N_B|PDBID|CHAIN|SEQUENCE      --------------------------ADPGGSHHHHHSRKTYTLTDYLKN 24
2FJP_A|PDBID|CHAIN|SEQUENCE      -----------------------------------TRKTYTLTDYLKN 13
2BGR_B|PDBID|CHAIN|SEQUENCE      --------------------------NKGTDDATADSRKTYTLTDYLKN 23
2AJB_D|PDBID|CHAIN|SEQUENCE      -----------------------------------SRRTYTLTDYLKS 13
Q9N2I7|DPP4_FELCA                MKTPWKVLLGLLGLAALITIITVPVVLLNKGNDAAADSRRTYTLTDYLKN 50
2GBI_B|PDBID|CHAIN|SEQUENCE      -----------------------------------RRTYTLADYLKN 12
                                 :     ::: :: :  :: :: : ::           *:****:****.

1R9N_B|PDBID|CHAIN|SEQUENCE      TYRLKLYSLRWISDHEYLYKQENNILVFNAEYGNSSVFLENSTFDEFGHS 74
2FJP_A|PDBID|CHAIN|SEQUENCE      TYRLKLYSLRWISDHEYLYKQENNILVFNAEYGNSSVFLENSTFDEFGHS 63
2BGR_B|PDBID|CHAIN|SEQUENCE      TYRLKLYSLRWISDHEYLYKQENNILVFNAEYGNSSVFLENSTFDEFGHS 73
2AJB_D|PDBID|CHAIN|SEQUENCE      TFRVKFYTLQWISDHEYLYKQENNILLFNAEYGNSSIFLENSTFDELGYS 63
Q9N2I7|DPP4_FELCA                TFRVKFYSLRWVSDHDYLYKQDNNILLFNAEYGNSSIFLENSTFDEFEHS 100
2GBI_B|PDBID|CHAIN|SEQUENCE      TFRVKSYSLRWVSDSEYLYKQENNILLFNAEHGNSSIFLENSTFEIFGDS 62
                                 *:*:* *:*:*:** :*****:****:****:****:*******: :  *

1R9N_B|PDBID|CHAIN|SEQUENCE      INDYSISPDGQFILLEYNYVKQWRHSYTASYDIYDLNKRQLITEERIPNN 124
2FJP_A|PDBID|CHAIN|SEQUENCE      INDYSISPDGQFILLEYNYVKQWRHSYTASYDIYDLNKRQLITEERIPNN 113
2BGR_B|PDBID|CHAIN|SEQUENCE      INDYSISPDGQFILLEYNYVKQWRHSYTASYDIYDLNKRQLITEERIPNN 123
2AJB_D|PDBID|CHAIN|SEQUENCE      TNDYSVSPDRQFILFEYNYVKQWRHSYTASYDIYDLNKRQLITEERIPNN 113
Q9N2I7|DPP4_FELCA                INDYSVSPDGQFILLEYNYVKQWRHSYTASYDIYDLNKRQLITEEKIPNN 150
2GBI_B|PDBID|CHAIN|SEQUENCE      ISDYSVSPDRLFVLLEYNYVKQWRHSYTASYSIYDLNKRQLITEEKIPNN 112
                                  .***:***  *:*:****************.**************:****
```

```
1R9N_B|PDBID|CHAIN|SEQUENCE      GWVGRFRPSEPHFTLDGNSFYKIISNEEGYRHICYFQIDKKD---CTFIT 371
2FJP_A|PDBID|CHAIN|SEQUENCE      GWVGRFRPSEPHFTLDGNSFYKIISNEEGYRHICYFQIDKKD---CTFIT 360
2BGR_B|PDBID|CHAIN|SEQUENCE      GWVGRFRPSEPHFTLDGNSFYKIISNEEGYRHICYFQIDKKD---CTFIT 370
2AJB_D|PDBID|CHAIN|SEQUENCE      GWVGRFRPAEPHFTSDGNSFYKIISNEEGYKHICHFQTDKSN---CTFIT 360
Q9N2I7|DPP4_FELCA                GWVGRFRPAEPHFTSDGRNFYKIISNEDGYKHICRFQIDKKD---CTFIT 397
2GBI_B|PDBID|CHAIN|SEQUENCE      GWCGRFRPAEPHFTSDGSSFYKIVSDKDGYKHICQFQKDRKPEQVCTFIT 362
                                 ** *****:***** ** .****:*:::**:*** ** *:.    *****

1R9N_B|PDBID|CHAIN|SEQUENCE      KGTWEVIGIEALTSDYLYYISNEYKGMPGGRNLYKIQLSDYTKVTCLSCE 421
2FJP_A|PDBID|CHAIN|SEQUENCE      KGTWEVIGIEALTSDYLYYISNEYKGMPGGRNLYKIQLSDYTKVTCLSCE 410
2BGR_B|PDBID|CHAIN|SEQUENCE      KGTWEVIGIEALTSDYLYYISNEYKGMPGGRNLYKIQLSDYTKVTCLSCE 420
2AJB_D|PDBID|CHAIN|SEQUENCE      KGAWEVIGIEALTSDYLYYISNEHKGMPGGRNLYRIQLNDYTKVTCLSCE 410
Q9N2I7|DPP4_FELCA                KGAWEVIGIEALTTDYLYYISNEYKGMPGGRNLYKIQLNDYTKVACLSCE 447
2GBI_B|PDBID|CHAIN|SEQUENCE      KGAWEVISIEALTSDYLYYISNEYKEMPGGRNLYKIQLTDHTNKKCLSCD 412
                                 **:****.*****:*********:* ********:***.*:*:  ****:

1R9N_B|PDBID|CHAIN|SEQUENCE      LNPERCQYYSVSFSKEAKYYQLRCSGPGLPLYTLHSSVNDKGLRVLEDNS 471
2FJP_A|PDBID|CHAIN|SEQUENCE      LNPERCQYYSVSFSKEAKYYQLRCSGPGLPLYTLHSSVNDKGLRVLEDNS 460
2BGR_B|PDBID|CHAIN|SEQUENCE      LNPERCQYYSVSFSKEAKYYQLRCSGPGLPLYTLHSSVNDKGLRVLEDNS 470
2AJB_D|PDBID|CHAIN|SEQUENCE      LNPERCQYYSASFSNKAKYYQLRCFGPGLPLYTLHSSSSDKELRVLEDNS 460
Q9N2I7|DPP4_FELCA                LKPERCQYYSVSFSKEAKYYQLRCSGPGLPLYTLHRSSNDEELRVLEDNS 497
2GBI_B|PDBID|CHAIN|SEQUENCE      LNPERCQYYSVSLSKEAKYYQLGCRGPGLPLYTLHRSTDQKELRVLEDNS 462
                                 *:********.*:*::****** * ********** * .:: ********

1R9N_B|PDBID|CHAIN|SEQUENCE      ALDKMLQNVQMPSKKLDFIILNETKFWYQMILPPHFDKSKKYPLLLDVYA 521
2FJP_A|PDBID|CHAIN|SEQUENCE      ALDKMLQNVQMPSKKLDFIILNETKFWYQMILPPHFDKSKKYPLLLDVYA 510
2BGR_B|PDBID|CHAIN|SEQUENCE      ALDKMLQNVQMPSKKLDFIILNETKFWYQMILPPHFDKSKKYPLLLDVYA 520
2AJB_D|PDBID|CHAIN|SEQUENCE      ALDKMLQDVQMPSKKLDVINLHGTKFWYQMILPPHFDKSKKYPLLIEVYA 510
Q9N2I7|DPP4_FELCA                ALDKMLQEVQMPSKKLDFIILNETKFWYQMILPPHFDTSKKYPLLIDVYA 547
2GBI_B|PDBID|CHAIN|SEQUENCE      ALDKMLQDVQMPSKKLDFIVLNETRFWYQMILPPHFDKSKKYPLLIDVYA 512
                                 *******:*********.* *: *:*************.*******::***

1R9N_B|PDBID|CHAIN|SEQUENCE      GPCSQKADTVFRLNWATYLASTENIIVASFDGRGSGYQGDKIMHAINRRL 571
2FJP_A|PDBID|CHAIN|SEQUENCE      GPCSQKADTVFRLNWATYLASTENIIVASFDGRGSGYQGDKIMHAINRRL 560
2BGR_B|PDBID|CHAIN|SEQUENCE      GPCSQKADTVFRLNWATYLASTENIIVASFDGRGSGYQGDKIMHAINRRL 570
2AJB_D|PDBID|CHAIN|SEQUENCE      GPCSQKVDTVFRLSWATYLASTENIIVASFDGRGSGYQGDKIMHAINRRL 560
Q9N2I7|DPP4_FELCA                GPCSQKADAIFRLNWATYLASTENIIVASFDGRGSGYQGDKIMHAVNRRL 597
2GBI_B|PDBID|CHAIN|SEQUENCE      GPCSQKADAAFRLNWATYLASTENIIVASFDGRGSGYQGDKIMHAINKRL 562
                                 ******.*: ***.*********************************:*:**

1R9N_B|PDBID|CHAIN|SEQUENCE      GTFEVEDQIEAARQFSKMGFVDNKRIAIWGWSYGGYVTSMVLGSGSGVFK 621
2FJP_A|PDBID|CHAIN|SEQUENCE      GTFEVEDQIEAARQFSKMGFVDNKRIAIWGWSYGGYVTSMVLGSGSGVFK 610
2BGR_B|PDBID|CHAIN|SEQUENCE      GTFEVEDQIEAARQFSKMGFVDNKRIAIWGWSYGGYVTSMVLGSGSGVFK 620
2AJB_D|PDBID|CHAIN|SEQUENCE      GTFEVEDQIEATRQFSKMGFVDDKRIAIWGWSYGGYVTSMVLGAGSGVFK 610
Q9N2I7|DPP4_FELCA                GTFEVEDQIEAARQFSKMGFVDDKRIAIWGWSYGGYVTSMVLGAGSGVFK 647
2GBI_B|PDBID|CHAIN|SEQUENCE      GTLEVEDQIEAARQFLKMGFVDSKRVAIWGWSYGGYVTSMVLGSGSGVFK 612
                                 **:********:*** ******.**:******************:******

1R9N_B|PDBID|CHAIN|SEQUENCE      CGIAVAPVSRWEYYDSVYTERYMGLPTPEDNLDHYRNSTVMSRAENFKQV 671
2FJP_A|PDBID|CHAIN|SEQUENCE      CGIAVAPVSRWEYYDSVYTERYMGLPTPEDNLDHYRNSTVMSRAENFKQV 660
2BGR_B|PDBID|CHAIN|SEQUENCE      CGIAVAPVSRWEYYDSVYTERYMGLPTPEDNLDHYRNSTVMSRAENFKQV 670
2AJB_D|PDBID|CHAIN|SEQUENCE      CGIAVAPVSKWEYYDSVYTERYMGLPTPEDNLDYYRNSTVMSRAENFKQV 660
Q9N2I7|DPP4_FELCA                CGIAVAPVSRWEYYDSVYTERYMGLPTPQDNLDYYKNSTVMSRAENFKQV 697
2GBI_B|PDBID|CHAIN|SEQUENCE      CGIAVAPVSRWEYYDSVYTERYMGLPTPEDNLDHYRNSTVMSRAENFKQV 662
                                 *********:******************:****:*:***************

1R9N_B|PDBID|CHAIN|SEQUENCE      EYLLIHGTADDNVHFQQSAQISKALVDVGVDFQAMWYTDEDHGIASSTAH 721
2FJP_A|PDBID|CHAIN|SEQUENCE      EYLLIHGTADDNVHFQQSAQISKALVDVGVDFQAMWYTDEDHGIASSTAH 710
2BGR_B|PDBID|CHAIN|SEQUENCE      EYLLIHGTADDNVHFQQSAQISKALVDVGVDFQAMWYTDEDHGIASSTAH 720
2AJB_D|PDBID|CHAIN|SEQUENCE      EYLLIHGTADDNVHFQQSAQLSKALVDAGVDFQTMWYTDEDHGIASNMAH 710
Q9N2I7|DPP4_FELCA                EYLLIHGTADDNVHFQQSAQISKALVDAGVDFQAMWYTDEDHGIASGPAH 747
2GBI_B|PDBID|CHAIN|SEQUENCE      EYLLIHGTADDNVHFQQSAQISKALVDAGVDFQAMWYTDEDHGIASSTAH 712
                                 ********************:*******.******:************.  **

1R9N_B|PDBID|CHAIN|SEQUENCE      QHIYTHMSHFIKQCFSLP 739
2FJP_A|PDBID|CHAIN|SEQUENCE      QHIYTHMSHFIKQCFSLP 728
2BGR_B|PDBID|CHAIN|SEQUENCE      QHIYTHMSHFIKQCFSLP 738
2AJB_D|PDBID|CHAIN|SEQUENCE      QHIYTHMSHFLKQCFSLP 728
Q9N2I7|DPP4_FELCA                QHIYTHMSHFIKQCFSLP 765
2GBI_B|PDBID|CHAIN|SEQUENCE      QHIYSHMSHFLQQCFSLR 730
                                 ****:*****::*****
```

Figure 2: Phylogram tree DPP-IV query sequence vs 6 proteins



Homology Modeling

About five models are generated by taking into consideration the default parameters. In this step a sequence-structure alignment, number of atoms, topology and restraints were also constructed. The summary of the produced five models is given in the following table. It is evident from the table that the energy associated with the first model represents the lowest possible energy [2668.67 kcal/mol]. Of the five models, the optimized energy for first model-1 is much lower than the remaining; hence the model evaluation (ramachandran plot, Figures 2 and 3) was carried out with that model.

Table 6: Summary of models generated by Modeler and their respective minimized energies.

Model protein	Optimization energy (kcal/mol)	Superposition RMSD (Aº)
1	2668.6763	0.2632
2	2771.8875	0.1860
3	2773.2507	0.2174
4	2691.0183	0.2803
5	2745.4941	0.1710

Figure 3: Ramachandran plot of template protein 2BGR\

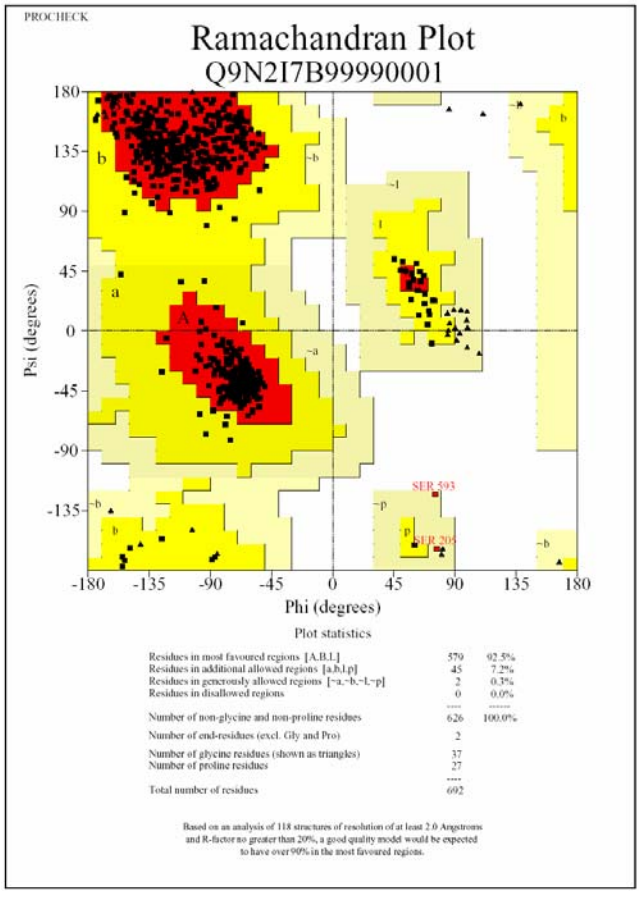
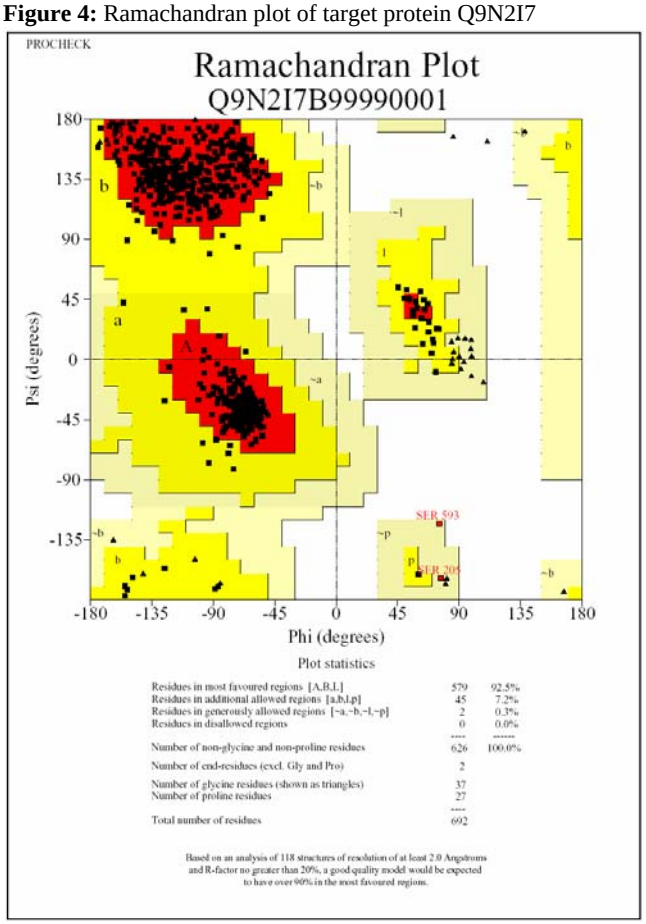


Figure 4: Ramachandran plot of target protein Q9N2I7



CONCLUSION

Pair wise sequence alignments performed by using three different programs like FASTA, BLAST, WUBLAST2 to study the influence of matrix on sequence alignment revealed quality alignments by these methods and FASTA (with PAM250) represented the method of choice when compared with BLAST and WUBLAST2 towards DPP-IV analysis. 3-D structure prediction strategies employed in the study were validated using ramachandran plot analysis and obtained model 1 with good stereo-chemical quality when compared with the other four generated models with low energy (2668.67 kcal/mol) and has low RMSD value (0.26 A°). The number of residues in the most favored region of ramachandran plot is 92.5% in Q9N2I7 model-1 protein when compared to 89.1% in 2BGR. There/fore from the study, it can be emphasized that utilizing such computational tools and methods would be advantageous in identifying sequence similarities by comparative sequence analysis and model building represents a fast and reliable source of detecting homologous sequences.

REFERENCES

1. Gupta R, Walunj SS, Tokala RK et al. (2009) Emerging drug candidates of dipeptidyl peptidase IV (DPP IV) inhibitor class for the treatment of Type 2 Diabetes. Curr Drug Targets. 10: 71-87.
2. Pratley RE and Salsali A. (2007) Inhibition of DPP-4: a new therapeutic approach for the treatment of type 2 diabetes. Curr Med Res Opin. 23: 919-931.

3. Gorrell MD. (2005) Dipeptidyl peptidase IV and related enzymes in cell biology and liver disorders. Clin Sci (Lond). 2005 108: 277-292.
4. Barnett A. (2006) DPP-4 inhibitors and their potential role in the management of type 2 diabetes. Int J Clin Pract. 60: 1454-1470
5. Altschul SF, Gish W, Miller W et al. (1990) Basic local alignment search tool. J Mol Biol. 215: 403-410
6. Pearson W. (2004) Finding protein and nucleotide similarities with FASTA. Curr Protoc Bioinformatics. Chapter 3:Unit3.9
7. Chenna R, Sugawara H, Koike T et al. (2003) Multiple sequence alignment with the Clustal series of programs. Nucleic Acids Res. 31: 3497-3500
8. Boutet E, Lieberherr D, Tognolli M et al. (2007) UniProtKB/Swiss-Prot. Methods Mol Biol. 406: 89-112.
9. http://www.rcsb.org/pdb
10. http://www.salilab.org
11. Gonnet GH, Cohen MA and Benner SA (1992) Exhaustive matching of the entire protein sequence database. Science 256: 1443-1445.

Received: 19 September 2011 Revised: 12 November 2011
Accepted: 12 November 2011 Online: 01 January 2012