Journal of Computational Intelligence in Bioinformatics.
ISSN: 0973-385X Volume 4 Number 1 (2011) pp. 101—109

http://www.ripublication.com/jcib.htm

Comparative Protein Structure Analysis of HXB2=viral Protease from HIV-1 Genome

Nitin M. Atre*, Harshada H. Pilley, Sandip R. Nagmote, Amena Khan, Vishal Changole and Gauri S. Deshpande

P.G. Department of Bioinformatics,
Shri Shivaji Science College, Amravati (MS), India
**Corresponding Author E-mail: neet3atre@gmail.com*

Abstract

The HIV-1 genome was the first HIV genome to be sequenced in its entirety. A human immunodeficiency virus type 1 (HIV-1) variant with highly reduced susceptibility to Ro 31-8959, an inhibitor of the viral protease, has been selected by repeated passage of wild-type virus in CEM cells in the presence of increasing concentrations of the inhibitor. Directed mutagenesis of the HIV-1 HXB2=viral protease at positions 48 and 90 suggested that each mutation alone led to a moderate decrease in sensitivity of the recombinant virus to protease inhibitor. Imperfect DNA mirror repeats in the *gag* gene of HIV-1 (HXB2=viral) identify key functional domains and coincide with protein structural elements in each of the mature proteins. Comparison of different HIV-1 HXB2=viral proteins structure analysis by differences in the locations of secondary structures, by explicit superposition of structural elements, or by procedures that utilize a common symmetry element or geometrical feature of the structures provides important insight into the structure of proteins, which in turn greatly facilitates the understanding of its biochemical and cellular function.

Comparative amino acid sequence analysis was carried out to understand its biochemical nature. Comparative secondary structure predicted by GOR IV was carried out to understand the differences in locations of secondary structures. A homology modeling method was used for the prediction of the structures of HXB2=viral protease. For modeling, a template protein was obtained by Geno3D server. By comparing the template protein a rough model was constructed for the target protein using MODELLER, a program for comparative modeling. The predicted 3-D models are further characterized and studied for comparative structure analysis using other techniques.

Keywords: Comparative protein structure analysis, Homology modeling, HXB2=viral, HIV-1.

Introduction

Countless people have died from AIDS, a chronic infectious disease caused by the HIV. HIV is thought to have originated in non-human primates in sub-Saharan Africa and was transferred to humans late in the 19th or early in the 20th century [3]. Human immunodeficiency virus type 1 (HIV-1) is a member of the lentivirus subfamily of retroviruses [4]. HIV infection in humans is considered pandemic according to the World Health Organization (WHO). From its discovery in 1981 to 2006, AIDS killed more than 25 million people [6].

The viral strain HXB2=viral Location (GenBank [15] Accession Number K03455) was selected as the reference strain because so many studies use HXB2=viral protease and because crystal structures for HXB2=viral related proteins are available [12]. The precise positions of an epitope on the HXB2=viral strain can be readily obtained using the Sequence Locator Tool.

Proteins from different sources and sometimes diverse biological functions can have similar sequences, and it is generally accepted that high sequence similarity is reflected by distinct structure similarity [25]. Indeed, the relative mean square deviation (rmsd) of the alpha-carbon co-ordinates for protein cores sharing 50% residue identity is expected to be around 1Å. This fact served as the premise for the development of comparative protein modeling (also often called modeling by homology or knowledge-based modeling), which is presently the most reliable method. Comparative model building by using the method of homology modeling, consist of the extrapolation of the structure for a new (target) sequence from the known 3D-structure of related family members (templates).

Since there was no structure reported for these HXB2=viral protease proteins, the main aim of this study is to predict three dimensional structures, study the comparative sequence compositions and perform comparative protein structure study of these important proteins to establish some important analysis.

Material and Method

Retrieval of Target Sequence

The amino acid sequence of the HXB2=viral protease, was obtained from the sequence database of NCBI (Accession No. AAB33141.1, AAB33142.1, AAB33143.1, AAB33144.1, AAB33145.1 and GeneID: 913385, 913386, 913387, 913388, 913389) [19]. These proteins consist of 99, 99, 91, 91, 80 amino acids respectively. It was ascertained that the three dimensional structures of these proteins were not available in Protein Data Bank. Hence the present exercise of developing the 3D models of these HXB2=viral protease proteins and their comparative protein structure analysis were undertaken.

Sequence Composition Comparison

The sequences of all the five proteases were studied and analyzed for their compositions of amino acids [7]. Seven parameters were designed to apply to the composition data of each protease independently. The parameters dealt with

information regarding the amino acid contents such as Aliphatic, Aromatic, Sulphur, Basic, Acidic and Aliphatic hydroxyl types (*Table 1*).

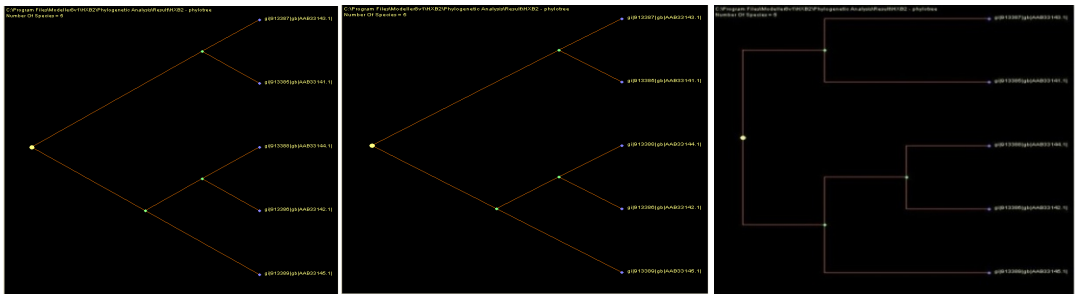
Table 1: Comparative Sequence Composition Table for HXB2=viral protease [AAB33141.1, AAB33142.1, AAB33143.1, AAB33144.1, AAB33145.1] of HIV-1 genome.

Sequence Pattern	AAB33141.1		AAB33142.1		AAB33143.1		AAB33144.1		AAB33145.1	
	Times found	%	Times found	%	Times found	%	Times found	%	Times found	%
A	3	3.03	3	3.03	3	3.30	3	3.30	3	3.75
B	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
C	2	2.02	2	2.02	1	1.10	1	1.10	1	1.25
D	4	4.04	4	4.04	4	4.40	4	4.40	4	5.00
E	4	4.04	4	4.04	4	4.40	4	4.40	4	5.00
F	2	2.02	2	2.02	1	1.10	1	1.10	1	1.25
G	11	11.11	12	12.12	11	12.09	11	12.09	11	13.75
H	1	1.01	1	1.01	2	2.20	1	1.10	1	1.25
I	12	12.12	12	12.12	12	13.19	12	13.19	10	12.50
K	6	6.06	6	6.06	6	6.59	6	6.59	6	7.50
L	11	11.11	10	10.10	8	8.79	8	8.79	8	10.00
M	2	2.02	2	2.02	2	2.20	2	2.20	2	2.50
N	4	4.04	4	4.04	3	3.30	3	3.30	3	3.75
P	6	6.06	6	6.06	6	6.59	6	6.59	4	5.00
Q	5	5.05	6	6.06	4	4.40	5	5.49	4	5.00
R	5	5.05	4	4.04	4	4.40	4	4.40	3	3.75
S	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
T	9	9.09	8	8.08	7	7.69	7	7.69	5	6.25
V	10	10.10	11	11.11	11	12.09	11	12.09	8	10.00
W	1	1.01	1	1.01	1	1.10	1	1.10	1	1.25
X	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
Y	1	1.01	1	1.01	1	1.10	1	1.10	1	1.25
Z	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
Aliphatic G, A,V, L, I	47	47.47	48	48.48	45	49.45	45	49.45	40	50.00
Aromatic F, W, Y	4	4.04	4	4.04	3	3.30	3	3.30	3	3.75
Sulphur C, M	4	4.04	4	4.04	3	3.30	3	3.30	3	3.75
Basic K, R, H	12	12.12	11	11.11	12	13.19	11	12.09	10	12.50
Acidic B, D, E, N, Q, Z	17	17.17	18	18.18	15	16.48	16	17.58	15	18.75

Aliphatic hydroxyl S, T	9	9.09	8	8.08	7	7.69	7	7.69	5	6.25
tRNA synthetase class I Z, E, Q, R, C, M, V, I, L, Y, W	53	53.54	53	53.54	48	52.75	49	53.85	42	52.50

Multiple Sequence Analysis

Multiple sequence alignment was performed to know about the conserved regions using Clustal-X tool [13] followed by phylogenetic analysis to understand the evolutionary relationship between the different HXB2=viral protease protein sequences *(Fig 1)*.



a b c

Figure 1: Phylogenetics Analysis reports of HXB2=viral protease proteins (A. Pair distance; B. Root distance; C. Rectangle distance).

Template Searching

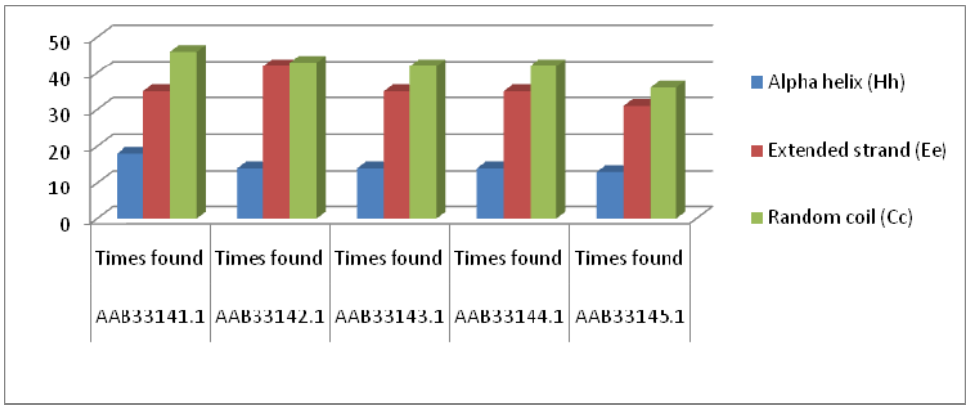
An attempt was made to find a suitable template protein for the modeling of the target proteins. The template protein was searched through Geno3D server [7], (an online tool for searching similar sequences, based on sequence and structure-wise similarity). The sequences of HXB2=viral protease proteins were found to resemble a single template protein named crystal structure of an in vivo HIV-1 protease mutant in complex with saquinavir: insights into the mechanisms of drug resistance from Human immunodeficiency virus 1 (HIV-1). [1FB7: A] Chain A.

Sequence Alignment

Amino acid sequence alignment of target and template proteins was derived using the Swiss-PdbViewer package [8]. Default parameters were applied and the aligned sequences were inspected and adjusted manually to minimize the number of gaps and insertions.

Secondary Structure Prediction and Comparison

Secondary structures were predicted from the sequences of all five proteases by using GOR – IV server (a method for predicting protein secondary structure from amino acid sequence). The resulting secondary structures were compared within themselves ***(Graph 1)***.



Graph 1: Comparative Secondary Structure Table for HXB2=viral protease [AAB33141.1, AAB33142.1, AAB33143.1, AAB33144.1, AAB33145.1] of HIV-1 genome

Generating Rough Models

Rough 3-D models (20 models) were constructed from the sequence alignment between *HXB2=viral protease* proteins and the template proteins using MODELLER 9v7 [3].

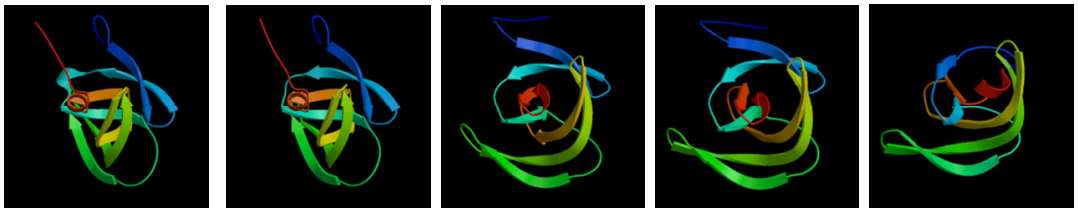


Figure 2: Predicted 3D structures of HXB2=viral protease [AAB33141.1, AAB33142.1, AAB33143.1, AAB33144.1, AAB33145.1] of HIV-1 genome.

Refinement

The rough models constructed were solvated and subjected to constraint energy minimization with a harmonic constraint of 100 kJ/mol/Å applied for all protein atoms, using the steepest descent and conjugate gradient technique to eliminate bad contacts between protein atoms and structural water molecules. Computations were carried out *in vacuo* with the GROMOS96 43B1 parameters set, implementation of Swiss-PdbViewer.

Evaluation of Refined Model

In the last step of HM the refined structures of the model was subjected to a series of tests for testing its internal consistency and reliability. Backbone conformations were evaluated by the inspection of the Psi/Phi Ramachandran plot obtained from PROCHECK analysis. The Swiss-PdbViewer energy minimization test was applied to check for energy criteria in comparison with the potential of mean force derived from a large set of known protein structures. Packing quality of the refined structures was investigated by the WHATIF [27]. A comparative analysis of Ramachandran plots of all the proteases was practiced in order to know the relative qualities of the predicted structures. This effort informed us that four out five structures were falling 100% in favored regions of Ramachandran plots *(Table 2)*.

Table 2: A study of comparison between Ramachandran plots of HXB2=viral protease of HIV-1 genome.

Accession	Favored Region	Additonal Allowed Region	Generously Allowed Region	Disallowed Region
AAB33141.1	96.3%	3.7%	0.0%	0.0%
AAB33142.1	100%	0.0%	0.0%	0.0%
AAB33143.1	100%	0.0%	0.0%	0.0%
AAB33144.1	100%	0.0%	0.0%	0.0%
AAB33145.1	100%	0.0%	0.0%	0.0%

Result and discussion

In general, 30% sequence homology is required for generating useful models. In our study, based on the results obtained from geno3D server [7], the crystal structure of an in vivo HIV-1 protease mutant in complex with saquinavir: insights into the mechanisms of drug resistance from Human immunodeficiency virus 1 (HIV-1) was selected as template for all the five HXB2=viral proteases. By using BLAST [1] [2], for template protein 1FB7 E- value between 0.0 to 1e-43, identity 92% to 98% and alignment score 168 to 177 was obtained. MODELLER was used for building the models and global energy minimization. With the help of MODELLER software 20 models were generated, stereochemical parameters of the proteins like main-and side chains data of HXB2=viral proteases was considered for determining the quality of the models which were generated by using PROCHECK 3.0 [14]. The main chain parameters like Ramachandran plot quality; peptide bond planarity, C-alpha Chirality, over-all G factor and the bad contacts per 80 to 99 residues are found to be within the limits for the models. After comparing all parameters best models were obtained. Best models were subjected to refinement by using energy minimization and it was found that models were same as refined model. Refined model was analyzed by different protein analysis programs including RAMPAGE Software for the evaluation of the Ramachandran plot quality, and WHATIF for the calculation of packing quality. The Ramachandran plots were plotted for HXB2=viral protease proteins using

PROCHECK software [14]. A comparison done in between Ramachandran plots of all the five proteases inferred that four out of five predicted models were falling 100% in favored regions describe by Ramachandran plot. Packing quality for these models was found in normal range with the help of WHATIF server [27]. From the tools VERIFY_3D & ERRAT which is available at Structural Analysis and Verification Server (SAVES) it was shown that 98% to 100% of the residues had an averaged 3D-1D score > 0.2 and overall quality factors were found in between 98.6 – 100%. The overall results provided the evidences that the predicted 3-Dimensional structure of HXB2=viral protease proteins are acceptable and of good quality. These structures (Fig 2; see PMDB ID - *PM0076397, PM0076398, PM0076399, PM0076400, PM0076395* for the corresponding coordinates in pdb format) were found to be satisfactory based on the above results [24].

The information gathered from all the above efforts strongly suggests that all the HXB2=viral proteases are closely related phylogenetically and must belong to a same family. The structural similarity that was observed after generating and comparing the 3D structures [28] of all five proteins was suggestive of their strong homology.

References

[1] Altschul S.F. (1993) A protein alignment scoring system sensitive at all evolutionary distances. *Journal of Molecular Evolution 36: 290-300.*

[2] Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ.: Basic local alignment search tool. *J Mol Biol*. (1990) Oct 5; 215(3):403-10.

[3] Ashish Gupta, Santosh Kumar Pal, Jitendra Narayan, Sharada M. Potukuchi. InSilico structure prediction of p27SJ, a novel protein in St John's Wort, that suppresses expression of HIV-1 genome. *Biofrontiers* (2009) Vol.1, Issue 1, 32-35.

[4] Barre-Sinoussi F, Chermann JC, Rey F, Nugeyre MT, Chamaret S, Gruest J, Dauguet C, Axler-Blin C, Vezinet-Brun F, Rouzioux C, *et al.*: Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS). *Science* (1983), 220:868-871.

[5] Brendel V., Bucher P., Nourbakhsh I., Blaisdell B.E., Karlin S. (1992) Methods and algorithms for statistical analysis of protein sequences. *Proceedings of the National Academy of Sciences USA 89: 2002-2006.*

[6] Broder S, Gallo RC: A pathogenic retrovirus (HTLV-III) linked to AIDS. *N Engl J Med (*1984), 311:1292-1297.

[7] Combet C, Jambon M, Deléage G, Geourjon C. (2002). Geno3D: Automatic comparative molecular modelling of protein. *Bioinformatics*.18:213-214.

[8] Cymerman I, Feder M et al.: Computational Method for Protein Structure Prediction and Fold Recognition. *Nucleic Acids and Molecular Biology (*2004), 15:02-109.

[9] Gish W. and States D.J. (1993) Identification of protein coding regions by database similarity search. *Nature Genetics 3: 266-272.*

[10] Guex N,Peitsch M.C. (1997). Swiss model and the Swiss-PdbViewer: An environment for comparative protein modeling. *Electrophoresis* 18:2714-2723.

[11] Hoon M.A., Ryba N.J.P.,(1997). Analysis and comparison of partial sequences of partial sequence of clone from a taste-bud-enriched cDNA library. *J DENT res.,* 76: 831

[12] Jacobsen H, Yasargil K, Winslow DL, Craig JC, Kröhn A, Duncan IB, Mous J.; Characterization of human immunodeficiency virus type 1 mutants with decreased sensitivity to proteinase inhibitor Ro 31-8959; Virology. 1995 Jan 10;206(1):527-34.

[13] Larkin M A, Blackshields G, Brown N P, Chenna R, McGettigan P A, McWilliam H, Valentin F, Wallace I M, Wilm A, Lopez R, Thompson J D, Gibson T J, Higgins D.G.(2007). ClustalW and ClustalX version 2. *Bioinformatics*. 23(21):2947-2948.

[14] Laskowski, R. A., MacArthur, M. W., Moss, D. S. and Thornton, J. M. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. J. Appl. Cryst. 26, 283-291.

[15] Lipman D.J. Ostell J. Benson D.A., Boguski M.S.(1994) Genbank. *Nucleic Acids Res*., pages 22: 3441-3444.

[16] Lüthy R., Bowie J.U., Eisenberg D.: Assessment of protein models with three-dimensional profiles. Nature (1992).356 (6364):83-85.

[17] Martí-Renom MA, Stuart AC, Fiser A, Sánchez R, Melo F, Sali A. (2000). Comparative protein structure modeling of genes and genomes. *Annu Rev. Biophys Biomol Struct*. 29:291-325.

[18] Pearson W.R. and Lipman D.J. (1988).: Improved Tools for Biological Sequence Comparison. *PNAS 85:2444- 2448.*

[19] Pruitt KD, Tatusova T, Klimke W, Maglott DR; NCBI Reference Sequences: current status, policy and new initiatives.; Nucleic Acids Res. 2009 Jan;37(Database issue):D32-6

[20] Sali A, Blundell T.L. (1993). Comparative protein modelling by satisfaction of spatial restraints. J Mol Biol. 234(3):779-815.

[21] Shpaer E.G., Robinson M., Yee D., Candlin J.D., Mines R., and Hunkapiller T. (1996) Sensitivity and Selectivity in Protein Similarity Searches: Comparison of Smith-Waterman in Hardware. *Genomics 38: 179-191.*

[22] Srinivasan, N. and Blundell, T.L. 1993. An evaluation of the performance of an automated procedure for comparative modelling of protein tertiary structure. *Protein Eng*. 6:501-512.

[23] Taylor, W.R., Flores, T.P., and Orengo, C.A. 1994. Multiple protein structure alignment. *Protein Sci.* 3:1858-1870.

[24] Tiziana Castrignano', Paolo D'Onorio De Meo, Domenico Cozzetto,Ivano Giuseppe Talamo, Anna Tramontano., The PMDB Protein Model Database. *Nucleic Acids Research* (2006)., 34: D306–D309.

[25] Unger, R., Harel, D., Wherland, S., and Sussman, J.L. 1989. A 3D building blocks approach to analyzing and predicting structure of proteins. *Proteins* 5:355-373.

[26] Von Ohsen, N., Sommer, I., and Zimmer, R. (2003). Profile-profile alignment: A powerful tool for protein structure prediction. *Pac. Symp. Biocomput.* 2003:252-263.

[27] Vriend, G. (1990). WHAT IF: A molecular modeling and drug design program. J. Mol. Graph. 8:52-56.

[28] Ye,Y., Jaroszewski, L., Li,W., and Godzik,A. (2003). A segment alignment approach to protein comparison. *Bioinformatics* 19:742-749.