Journal of Computational Intelligence in Bioinformatics
ISSN: 0973-385X Volume 2 Number 2 (2009) pp. 121–129

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

Structure Identification of Gp41 [HIV-1] Protein and Computational Analysis of its Binding Sites

Uday Bhan Sharma[1*], Ashish Patel[2*], Sheetal Khatri[2], Vinod Rishi[3], Shailendra Kr. Gupta[4] and M.K.Verma[2]

[1]Kanpur institute of Technology, UPSIDC, Rooma, Kanpur (U.P.) 208004, India
E-mail: udaysharma1@gmail.com
[2]Sub-DIC Bioinformatics Centre, NIT Raipur, Raipur (CG) 492010, India
E-mail:Ashish.dic@gmail.com
[3]Brahmanand College Kanpur (U.P.), India
E-mail: vinod.rishi25@gmail.com
[4]Indian Institute of Toxicology Research Lucknow (U.P.), India
E-mail:Shailendra15@gmail.com

Abstract

HIV (Human Immunodeficiency Virus) is a virus that causes AIDS (Acquired Immunodeficiency Syndrome), a health condition in which a person is affected by a series of diseases because of poor immunity. HIV by itself is not an illness and does not instantly lead to AIDS. Gp41 (HIV-1) protein is a glycoprotein which make the envelope of HIV-1 virus. It is very important in the infection of virus in to human body cells. So the Structure Prediction of Gp41 (HIV-1) protein can be used in the analysis of HIV virus infection in Human. Modeler is used for structure prediction of Gp41 [HIV-1 protein]. Predicted & validated structure can be used in the analysis of drug against AIDS disease, because by the analysis of structure we analyze conserve regions of Gp41 (HIV-1) Protein. These conserve regions sometimes defines as Binding Sites where appropriate ligands bind and inhibits the functions of Gp41 [HIV-1] protein. CASTp can be used to study surface features and functional regions of proteins. CASTp is used to analyze the binding pockets of Gp41 [HIV-1] protein.

Key Words: Acquired Immunodeficiency Syndrome (AIDS), Domain, Infection, Glycoprotein, Binding pockets.

Introduction

HIV (Human Immunodeficiency Virus) is a virus that causes AIDS (Acquired Immunodeficiency Syndrome). An HIV-infected person can lead a healthy life for several years before he/she develops AIDS. Infection with HIV occurs by the transfer of blood, semen, vaginal fluid, pre-ejaculate, or breast milk. Within these bodily fluids, HIV is present as both free virus particles and virus within infected immune cells. HIV belongs to a class of viruses called retroviruses, which have genes composed of ribonucleic acid (RNA) molecules. In contrast, the genes of humans and almost all other organisms are made of a related molecule, deoxyribonucleic acid (DNA). Like all viruses, HIV can replicate only inside cells commandeering the cell's machinery to reproduce. However, only HIV and other retroviruses use an enzyme called reverse transcriptase to convert their RNA into DNA, which can be incorporated into the host cells genes. Since only retroviruses use this enzyme (i.e., and it is not used by the host cell), treatment strategies have focused heavily on this enzyme.

The viral envelope: The outer coat of the virus, known as the viral envelope, is composed of two layers of fatty molecules called lipids. These lipids actually come from the membrane of a human cell when a newly formed virus particle buds from the cell.

The viral core (or capsid): Within the envelope of a mature HIV particle is a bullet-shaped core made of 2000 copies of another viral protein (p24.)[3]. The core surrounds two single strands of HIV RNA, each of which has a copy of the virus's nine genes. The core of HIV also includes another protein (p7), the HIV nucleo capsid protein; and three enzymes that carry out later steps in the virus's life cycle: reverse transcriptase, and protease.

HIV-1 envelope gp41 is a transmembrane protein that promotes fusion of the virus with the plasma membrane of the host cells required for virus entry. In addition, gp41 is an important target for the immune response and development of antiviral and vaccine strategies, especially when targeting the highly variable envelope gp120 has not met with resounding success. Mutations in gp41 may affect HIV-1 entry, replication, pathogenesis, and transmission. gp41, is present on the surface of HIV-1 non-covalently bound to gp120, is responsible for fusion of viral envelope to the plasma membrane of the host cell and is essential for HIV-1 entry and replication. The Envelop gp41 is comprised of an extraviral domain (ectodomain), a membrane spanning region and an unusually long endodomain within the virus. The ectodomain of gp41 consists of an amino-terminal fusion domain and N- and C-terminal heptads repeats [4] (HR-1 and HR-2, respectively. The gp41 amino terminus is a highly hydrophobic region bearing the FLG motif called fusion peptide (FP), which makes the initial contact with the target membrane and can fuse biological membranes by itself. The two heptads repeat regions self-assemble into a thermo stable six-helix bundle, consisting of a trimeric coiled-coil interior (HR-1) with three exterior helices (HR-2) packed in the grooves of the trimer in an antiparallel manner, which represents the fusion-active conformation of gp41.

The end domain of gp41 encodes a Tyr-based motif that interacts with the AP-2 clathrin adaptor protein and is required for optimal viral infectivity .Two lentivirus

lytic peptides (LLPs)[2] in this domain which are capable of binding and disturbing lipid bilayers, interact with calmodulin and inhibits Ca^{2+}-dependent T-cell activation. There are four sites in gp41 for N-linked glycosylation that promote efficient Env-mediated cell-to-cell fusion but are largely dispensable for viral replication.

The objective of this study is to predict the three dimensional structure from the Protein sequence and identify the binding site on protein surface.

Methodology

Structure prediction: Select the template for Gp41 (HIV-1) protein .For this purpose first copy the fasta format of the sequence Gp41 [HIV-1] from NCBI database. Perform BlastP by using NCBI-Blast. Finding approx 31 hits of Gp41 (HIV-1) sequence in whole PDB database. Select that sequence as template having low E-value. So 1ENV with accession gi|2781333|pdb|1ENV|A is selected as template for our query sequence Gp41 [HIV-1], then predict the structure by using MODELLER. Modeler works on the basis of 3 specialized files i.e. ATOM FILE (.atm), ALIGNMENT FILE (.ali), TOP FILE (.top), place these file in to Bin folder of Modeler, and run Modeler with command line prompt. MODELLER [7] generates the numbers of structures of query protein which we written in .top file. Modeler outputs comes in the form of 3 files i.e. Co-ordinate file, violation file, description file. Now only use the coordinate file extract from Modeler output, and save it as .pdb extension. This is structure of unknown GP41 (HIV-1) protein.

Identification of Binding Site: The pockets are identified by CASTp web-server using PDB co-ordiante file of Gp41 [HIV-1] as input. These PDB files are obtained as output of Modeler. These pockets shows some conserve part in different regions of structure of Gp41 [HIV-1] protein. Sometimes these pockets are used to analyze Binding Sites of Gp41 [HIV-1] Protein [5].

Results and Discussion

BLASTP Result: The blast result defines 31 hits for Gp41 [HIV-1] Protein in the whole PDB Database. So we select 1st hit of 1ENV with minimum E-value 2e-21 and score 96.3 Bits. (Table-1)

Table 1

Sequences producing significant alignments:		Score (Bits)	E Value	
gi|2781333|pdb|1ENV|A	Chain A, Atomic Structure Of The Ectodomai	96.3	2e-21	S
gi|3891827|pdb|2EZS|A	Chain A, Solution Nmr Structure Of Ecto...	80.5	1e-16	S
gi|118137444|pdb|2CMR|A	Chain A, Crystal Structure Of The Hiv...	70.1	2e-13	S
gi|14719634|pdb|1F23|A	Chain A, Contribution Of A Buried Hydr...	59.3	2e-10	S
gi|10120666|pdb|1FAV|A	Chain A, The Structure Of An Hiv-1 Spe...	53.1	2e-08	S
gi|2392144|pdb|1AIK|C	Chain C, Hiv Gp41 Core Structure	47.4	9e-07	S
gi|2781016|pdb|1SZT|	Chain , Atomic Structure Of A Thermost...	47.4	1e-06	S
gi|10120667|pdb|1FAV|C	Chain C, The Structure Of An Hiv-1 Spe...	47.0	1e-06	S
gi|6573368|pdb|1QR8|A	Chain A, Inhibition Of Hiv-1 Infectivit...	46.6	2e-06	S
gi|6573369|pdb|1QR9|A	Chain A, Inhibition Of Hiv-1 Infectivit...	46.2	2e-06	S
gi|16975347|pdb|1K33|A	Chain A, Crystal Structure Analysis Of Th	45.8	3e-06	S
gi|16975348|pdb|1K34|A	Chain A, Crystal Structure Analysis Of Gp	45.8	3e-06	S
gi|6730406|pdb|1DLB|A	Chain A, Helical Interactions In The Hi...	45.4	4e-06	S
gi|24158672|pdb|1I5X|A	Chain A, Hiv-1 Gp41 Core	45.4	4e-06	S
gi|24158673|pdb|1I5Y|A	Chain A, Hiv-1 Gp41 Core	45.4	4e-06	S
gi|2392143|pdb|1AIK|N	Chain N, Hiv Gp41 Core Structure	34.3	0.009	S
gi|6435769|pdb|1CZQ|A	Chain A, Crystal Structure Of The D10-P...	31.2	0.075	S
gi|20663943|pdb|1JPX|A	Chain A, Mutation That Destabilize The...	30.8	0.095	S
gi|20663948|pdb|1JQO|A	Chain A, Mutation That Destabilize The...	30.4	0.12	S
gi|24987349|pdb|1IM7|A	Chain A, Solution Structure Of Synthet...	27.7	0.69	S
gi|16975231|pdb|1JAU|A	Chain A, Nmr Solution Structure Of The...	27.7	0.72	S

Alignment Result

>gi|2781333|pdb|1ENV|A Chain A, Atomic Structure of the Ectodomain from Hiv-1 Gp41

Length=123

Score = 96.3 bits (238), Expect = 2e-21, Method: Composition-based stats.

Identities = 62/103 (60%), Positives = 68/103 (66%), Gaps = 33/103 (32%)

```
Query  1    QHMLQLTVWGIKQLQARVLAVERYLKDQQLLGIWGCSGKLICTTTVPWNASWSNKSLNDI  60
            QH+LQLTVWGIKQLQAR+LAVERYLKDQ
Sbjct  54   QHLLQLTVWGIKQLQARILAVERYLKDQ--------------------------------  81

Query  61   WNNMTWMEWEREIGNYTGLIYTLIEQSQNQQEKNEQELLELDK  103
             NNMTWMEW+REI NYT LI++LIE+SQNQQEKNEQELLELDK
Sbjct  82   -NNMTWMEWDREINNYTSLIHSLIEESQNQQEKNEQELLELDK  123
```

This results define the whole alignment between Gp41 [HIV-1] & 1ENV with sequence identities 60% with fully define X-Ray crystallographic[6] structure with appropriate parameters of Resolution and R-value.

Secondary Structure Result of Gp41 [HIV-1] Protein

```
        10        20        30        40        50        60        70        80
         |         |         |         |         |         |         |         |
QHMLQLTVWGIKQLQARVLAVERYLKDQQLLGIWGCSGKLICTTTVPWNASWSNKSLNDIWNNMTWMEWEREIGNYTGLI
cccHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHcccceeeeeeecccccccccccceeeeceeeeeeeeccccccch
YTLIEQSQNQQEKNEQELLELDKWASLWNWFDITKWLWYIKIF
hhhhhhhhhhhhhhhhhhhhhhhhhhhhceeeeeceeeeeeec

Sequence length :   123

GOR4 :
   Alpha helix      (Hh) :    61 is  49.59%
   3_10  helix      (Gg) :     0 is   0.00%
   Pi helix         (Ii) :     0 is   0.00%
   Beta bridge      (Bb) :     0 is   0.00%
   Extended strand  (Ee) :    32 is  26.02%
   Beta turn        (Tt) :     0 is   0.00%
   Bend region      (Ss) :     0 is   0.00%
   Random coil      (Cc) :    30 is  24.39%
   Ambigous states  (?)  :     0 is   0.00%
   Other states          :     0 is   0.00%
```

Energy results of Gp41 [HIV-1] models

Energy of model_1 = -5059.589 kj/mol
Energy of model_2 = -5199.621 kj/mol

Two models are generated and the energy results of 2 models of Gp41 [HIV-1] defines that the second model of our target sequence shows minimum energy. So considering second model is the correct model.

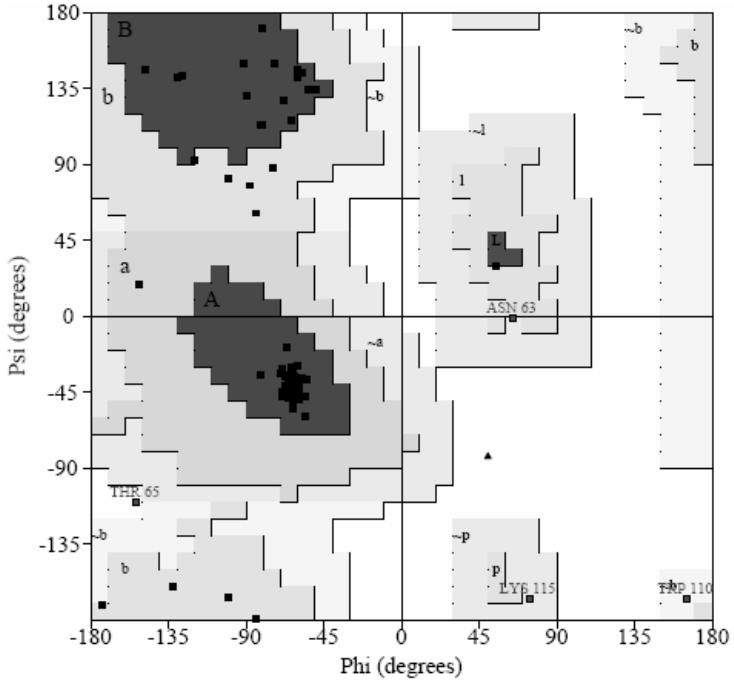
Visualization Results

The 3D structure of our query protein Gp41 [HIV-1] shown in the rasmol window (Figure-1).

Figure 1

Validation of Results

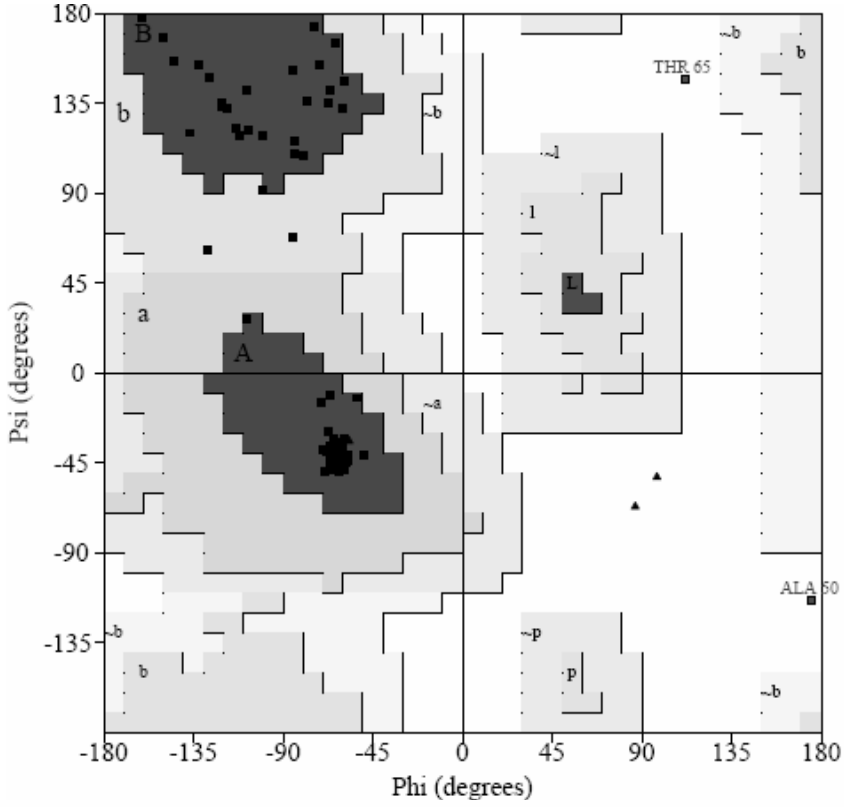
Ramachandran plot statistics defines that model_2 (Figure-2) of Gp41 [HIV-1] contains maximum residues in most favoured (allowed) regions in comparison to model_1 (Figure-3) of Gp41 [HIV-1].Because it contains 96.5% residues in favoured regions & model_1 contains 86.8% residues in favoured regions. As well as this model shows minimum energy in comparison to model_1 because in energy minimization steps model_1 energy result is -5059.589 Kj/mol and model_2 energy result is -5199.62 Kj/mol.So the analysis of these 2 models of Gp41[HIV-1] defines that the second model used as a validate model for our target sequence.



Plot statistics

Residues in most favoured regions [A,B,L]	99	86.8%
Residues in additional allowed regions [a,b,l,p]	11	9.6%
Residues in generously allowed regions [~a,~b,~l,~p]	3	2.6%
Residues in disallowed regions	1	0.9%
Number of non-glycine and non-proline residues	114	100.0%
Number of end-residues (excl. Gly and Pro)	2	
Number of glycine residues (shown as triangles)	6	
Number of proline residues	1	
Total number of residues	123	

Figure 2



Plot statistics

Residues in most favoured regions [A,B,L]	110	96.5%
Residues in additional allowed regions [a,b,l,p]	2	1.8%
Residues in generously allowed regions [~a,~b,~l,~p]	0	0.0%
Residues in disallowed regions	2	1.8%
Number of non-glycine and non-proline residues	114	100.0%
Number of end-residues (excl. Gly and Pro)	2	
Number of glycine residues (shown as triangles)	6	
Number of proline residues	1	
Total number of residues	123	

Figure 3

Pockets Information of Gp41 [HIV-1] By CASTp

The pockets are identified by CASTp web-server using PDB co-ordinate file of Gp41 [HIV-1] as input. These PDB files are obtained as output of Modeller. These pockets (Figure-4) shows some conserve part in different regions of structure of Gp41 [HIV-1] protein. Sometimes these pockets are used to analyze Binding Sites of Gp41 [HIV-1] Protein. Pockets are empty concavities on a protein surface into which solvent can gain access, i.e., these concavities have mouth openings connecting their interior with

the outside bulk solution, shallow depressions are excluded from the calculation. CASTp identifies all pockets and cavities on a protein structure and provides a detailed listing of all atoms participating in their formation. It also measures the volume and area of each pocket and cavity analytically, using both the solvent accessible surface and molecular surface models. In addition, it measures the size of mouth openings of individual pockets, which helps to assess the accessibility of binding sites to various ligands and substrates.

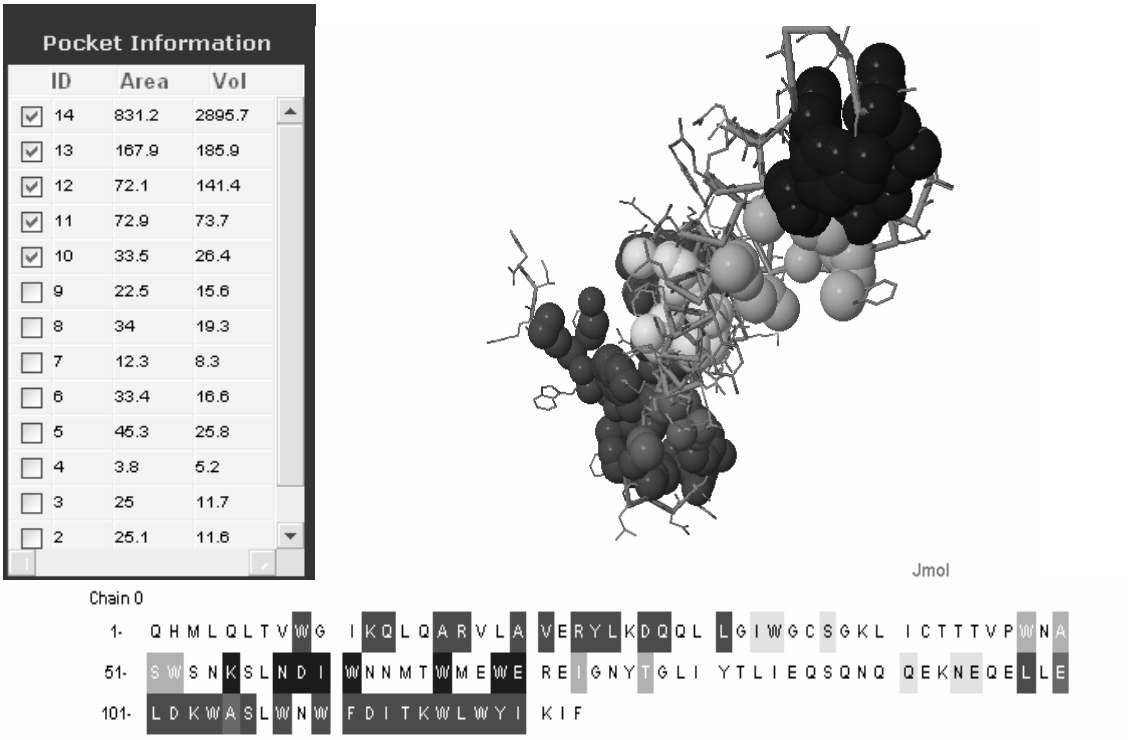


Figure 4

Conclusion

Structure Prediction by Modeler is very much successful because model is validated by both Ramachandran Plot calculations & Energy minimization results. Model generated by Modeler shows 96.5% residues in allowed or most favored regions as well as minimum energy -5199.62 kj./mol. So we used this structure in future study's like identification of conserve domains and analysis of binding sites of Gp41[HIV-1] Protein.

References

[1] Chan DC, Fass D, Berger JM, Kim PS: Core structure of gp41 from the HIV envelope glycoprotein. *Cell* 1997, 89:263-273

[2] Boge M, Wyss S, Bonifacino JS, Thali M: A membrane-proximal tyrosine-based signal mediates internalization of the HIV-1 envelope glycoprotein via interaction with the AP-2 clathrin adaptor. *J Biol Chem* 1998, 273:15773-15778

[3] Day JR, Munk C, Guatelli JC: The membrane-proximal tyrosine-based sorting signal of human immunodeficiency virus type 1 gp41 is required for optimal viral infectivity.*J Virol* 2004, 78:1069-1079

[4] Rajesh Ramakrishna, Roshni Mehta, Vasudha Sundaravaradan, Tiffany Davis and Nafees Ahmad, Characterization of HIV-1 envelope gp41 genetic diversity and functional domains following perinatal transmission. Department of Microbiology and Immunology, College of Medicine, The University of Arizona Health Sciences Center, Tucson, Arizona 85724, USA

[5] Gil Amitai, Arye Shemesh, Einat Sitbon, Maxim Shklar, Dvir Netanely, Ilya Venger and Shmuel Pietrokovski Network Analysis of Protein Structures Identifies Functional Residues.

[6] Jie Liang, Andrew Binkowski, Larisa Adamian, A Novel System for Discovery of Gene Function by Protein Surface Matching.

[7] Fiser A, Sali A, 2003, Modeller generation and refinement of homology based protein structure models Methods. Enzymole 374: 461-91.