JOURNAL OF PURE & APPLIED MICROBIOLOGY, Nov. 2014. Vol. **8**(Spl. Edn. 2), p. 739-742

Gag Polyprotein Isolate JH32 of HIV1 group M subtype B 3D Structure Determination by Multithreading Method - A Bioinformatics Approach

Undety Benjamin Jason[1] and Daniel Alex Anand[2]

[1]Department of Bioinformatics, Sathyabama University, Chennai, India,
[2]Department of Biomedical Engineering, Sathyabama University, Chennai, India.

(Received: 06 September 2014; accepted: 13 October 2014)

The discovery of potential drug to arrest the replication of HIV-1 even in these modern days is a growing task. Drugs in market are targets the enzymes reverse transcriptase, integrase and protease. The encapsulation gene of HIV genome that is GAG is found to be more conserve. So the functional and structural conservation of HIV-1 Gag implies structure based drug design. Now HIV-1 replication can be successfully blocked by targeting gag polyprotein, a new promising avenue for the drug class. Gag's role is helping in forming the encapsulation of the virus if that is disturbed the whole Virus is vulnerable. The challenge is to determine the structure of Gag poly-protein as its structure is still unknown. The crystal structure is unavailable in the structural databases like PDB. The sequence of gag polyprotein of HIV1 group M subtype B (isolate JH32) (HIV-1) was downloaded to determine its structure. As there is no template close to the target a method called refinement through multi threading was used. The percentage of favored regions by Ramchandran diagram is 89.6%, allowed region is 5.0% and amino acids in outlier region is 5.4%.

Key words: Conservation, template , gag polyprotein, outlier region.

HIV-1, which causes acquired immune deficiency syndrome (AIDS), is a retrovirus in genus Lentiviridae. HIV-1 is an enveloped virus which encodes two envelope (Env) glycoproteins - the surface (SU) glycoprotein gp120 and a transmembrane (TM) glycoprotein gp41, Gag has four major proteins, they are matrix (MA), capsid (CA), nucleocapsid (NC), and p6—and the pol-encoded enzymes protease (PR), reverse transcriptase (RT), and integrase (IN). HIV-1 also encodes two regulatory proteins,they are Tat and Rev and several accessory proteins,they are Vpu, Vif, Nef, and Vpr. The genome is pseudodiploid which is composed of two single strands of RNA linked in dimer. The HIV-1 infection initiates with the attachment of gp120 to target cell plasma membrane.[1–4] The principal attachment of the receptor for HIV-1 and other pri- mate lentiviruses is CD4. Productive infection also requires the presence of a co-receptor,they are typically CXCR4 or CCR5. The binding of gp120 to CD4 and co-receptor initiates conformational changes in gp41, which in turn directs to fusion of the viral envelope and the target cell membrane and entry of the viral core into the host cell cytoplasm. Recent evidence suggests that HIV-1 entry can also occur in a low-pH endosomal compartment after receptor-mediated endocytosis.[5] Upon entry of the virion into the cytosol, the Env glycoproteins and the lipid-associated MA protein dissociate from the incoming particle at the membrane, and the poorly understood process of uncoating is initiated. The enzymes RT and IN, together with the NC protein, remain in close association with the viral RNA as it is converted to double-stranded DNA by RT-catalyzed reverse transcription.[6] NC acts as a

---

* To whom all correspondence should be addressed.
Mob.: +91-9941850363;
E-mail: ubjason@gmail.com

nucleic acid chaperone at several steps during reverse transcription to facilitate the conversion of RNA to DNA.[7] Vpr is also a component of the reverse transcription complex (RTC). The extent to which CA remains associated with the incoming RTC has been a topic of debate. However, reverse transcription and uncoating appear to be temporally linked,[8] and it is clear that some host restriction factors that block early post entry steps in the viral replication cycle target CA.[9,10] The newly reverse transcribed viral DNA is translocated to the nucleus in a structure known as the preintegration complex (PIC). The nuclear import process remains incompletely understood; however, a role for CA in this process[11,12] implies that some CA protein may remain associated with the viral nucleoprotein complex as it traffics to the nuclear pore. Once inside the nucleus, the double-stran- ded viral DNA integrates into the target cell genome through the action of the IN enzyme.[13] The integrated viral DNA serves as the template for transcription from the viral pro moter in the 5' long terminal repeat (LTR) to generate the spliced viral mRNAs and full-length genomic RNAs; these are transported out of the nucleus via the action of the Rev protein.[3]The Gag proteins are translated from full-length message as a polyprotein precursor containing MA, CA, NC, and p6 domains as well as two spacer peptides, SP1 and SP2.[14] During translation of the Gag precursor,known as $Pr55^{Gag}$, an occasional 1 ribosomal frameshift leads to the production of a GagPol precursor protein ($Pr160^{GagPol}$), the abundance of which is approximately 5% that of $Pr55^{Gag}$. The Gag and GagPol precursor polyproteins are transported to the plasma membrane, where they assemble and incorporate the viral Env glycoproteins. The membrane targeting of Gag and GagPol is regulated by the MA domain, which also plays an important role in the incorporation of the viral Env glyco-proteins. Assembly takes place in cholesterol-rich membrane microdomains (lipid rafts) through direct interactions be- tween MA and the phospholipid phosphatidylinositol-4,5- bisphosphate [PI(4,5)P2].[15] Interactions within the CA do- main of Gag initiate the Gag assembly process.

The Gag gene of HIV-1 expressed MA (p17),CA (p24),SP1 (p2),NC (p7), SP2 (p1) and P6. HIV p6 is a 6 kDa polypeptide at the N-terminus of the Gag polyprotein. It recruits cellular proteins Tsg101 (a component of ESCRT-1) and Alix to initiate virus particle budding from the plasma membrane. p6 has no known function in the mature virus.P6 protein was taken to study its variance in sequence and structural levels. Gag proteins are plays vital role in virus assembly, release, maturation, function in the establishment of a productive HIV 1 and which as also Viral structural proteins. Inspite of their vital role throughout the replication cycle, there are currently no approved antiretroviral therapies that target the Gag precursor protein or any of the mature Gag proteins. Recent progress in understanding the structural and cell biology of HIV-1 Gag function has revealed a number of potential Gag-related targets for possible therapeutic intervention. In this study, we emphasize that our current understanding of HIV-1 Gag P6 protein suggest some approaches to be as a target for novel antiretroviral agents.

MATERIALS AND METHODS

Sequence of Gag Poly Protein: The gag polyprotein of Human immunodeficiency virus type 1 group M subtype B (isolate JH32) (HIV-1) is downloaded for three dimensional structure determinations. Its length is 500 amino acids. The Gag polyprotein sequence is retrieved from the Uniprot database,

>sp|P12494|GAG_HV1J3 Gag polyprotein OS=Human immunodeficiency virus type 1 group M subtype B (isolate JH32) GN=gag, PE=3,SV=3,MGA RASVLSGGEL DRWEK
IRLRPGGKKKYKLKHIVWASR ELERFAVN
PSLLETSEGCRQILGQLQPSLQTGSEELKSL
FNTVATLYCVHQ RIEVKDTKEALEKIEEE
QNKSKKKAQQAAADTG NSSKVSQNYP
IVQNIQGQMV HQAISPRTLNA WVKVVEEKAFS
PEVIP MFSALSEGATPQD LNTMLNTVGGHQAA
MQMLKETINEEAA EWDRLHPAQAGPIAP
GQMREPRGSDIAGTTST LQEQIGW MTSNPP
IPVGEIYKRWIILGLNKIVRMYSPSSI LDIRQGP
KEPFRDYVDRFYKTLRAEQAS
QEVKNWMTETLL VQNANPDCKTILKA
LGPAATLEEM MTACQGVG GPGHKARVLAEA
MSQVTNSTTIM MQRGNFRNQRKIIKCFNCGK
EGHLARNCRAPRKKGCWKCGKEGH
QMKDCNERQANFLG KIWPSSKGR PGNFLQ
SRPEPTAPPEESFRFGEETTTPSQK QEPRDKEL
YPL ASLRSLFGNDPSSP

A hierarchical protein structure modeling approach is used on the secondary-structure enhanced profile-profile threading alignment and the iterative implementation of the threading assembly refinement program[3]. Critical Assessment of Structure Prediction (CASP) experiment implemented. CASP (or Critical Assessment of Techniques for Protein Structure Prediction) is a community-wide experiment for testing the state-of-the-art of protein structure predictions which takes place every two years since 1994. The experiment (often referred as a competition) is strictly blind because the structures of testing proteins are unknown to the predictors.

RESULTS AND DISCUSSION

The software first retrieves template proteins of similar folds (or super-secondary structures) from the PDB library by LOMETS, a locally installed meta-threading approach. Then the continuous fragments excised from the PDB templates are reassembled into full-length models by replica-exchange Monte Carlo simulations with the threading unaligned regions (mainly loops) built by ab initio modeling. In cases where no appropriate

Fig. 1. Constructed three dimensional structure of Gag Polyprotein isolate JH32

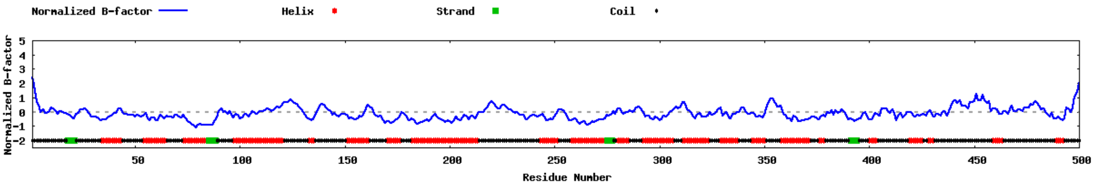
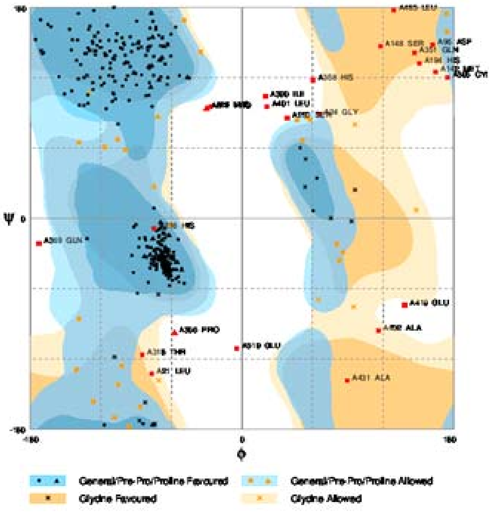


Fig. 2. Representation of each residue accuracy forming secondary structure



Number of residues in favoured region (~98.0% expected): 443 (89.0%)
Number of residues in allowed region (~2.0% expected) : 32 (6.4%)
Number of residues in outlier region : 23 (4.6%)

Fig. 3.

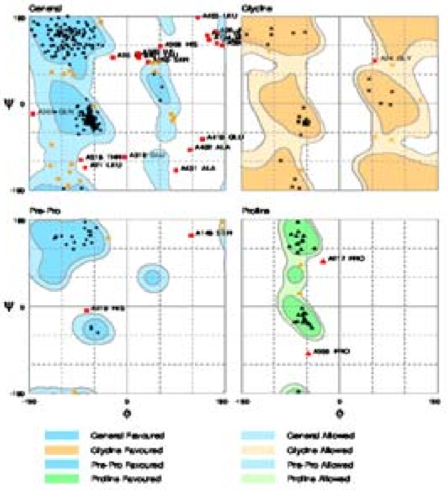


Fig. 4. Residues in the allowed and outlier regions

template is identified by LOMETS, I-TASSER will build the whole structures by ab initio modeling. The low free-energy states are identified by SPICKER through clustering the simulation decoys. In the third step, the fragment assembly simulation is performed again starting from the SPICKER cluster centroids, where the spatial restrains collected from both the LOMETS templates and the PDB structures by TM-align are used to guide the simulations. The purpose of the second iteration is to remove the steric clash as well as to refine the global topology of the cluster centroids. The decoys generated in the second simulations are then clustered and the lowest energy structures are selected. The final full-atomic models are obtained by REMO which builds the atomic details from the selected I-TASSER decoys through the optimization of the hydrogen-bonding network The three dimensional structure is determined with the 89.6% of amino acids in favored regions by Rampage.

The best predicted model evaluated by Rampage, University of Cambridge is shown in *Fig3&4* which was found to have best score being validated out of the five models. Fig 2 Shows the accuracy of each residues forming a helix, beta strand or a coil.

CONCLUSION

The drug resistance to Anti-retroviral therapy targeting mostly RT and IN has directed to study Gag poly protein. The Gag polyprotein was understood to be conserved and implicates rational drug design. As it forms the encapsulation of HIV, if blocked it allows core viral ingredients to fatal end. The non availability of PDB structure and similar template with higher number of residues caused to use threading assembly refinement. This work further carries in finding active site, lead screening and docking studies of Gag poly protein.

REFERENCES

1. Guangdi Li,Jens Verheyen,Soo-Yon Rhee, Arnout Voet, Anne-Mieke Vandamme,and Kristof heys:Functional conservation of HIV-1 Gag:implications, *Retrovirology*, 2013; **10**:126,
2. Wu ST, Zhang Y: LOMETS: A local meta-threading-server for protein structure prediction.
3. Zhang Y, Skolnick J: Automated structure prediction of weakly homologous proteins on a genomic scale.
4. Melikyan GB: Common principles and intermediates of viral protein-mediated fusion: The HIV-1 paradigm. Retrovirology 2008;5:111.
5. Miyauchi K, Kim Y, Latinovic O, Morozov V, and Melikyan GB: HIV enters cells via endocytosis and dynamin-dependent fusion with endosomes. *Cell* 2009; **137**(3): 433-444.
6. Telesnitsky A and Goff SP: Reverse transcriptase and the generation of retroviral DNA. In: Retroviruses (Coffin J,Varmus H, Hughes S, eds.). Cold Spring Harbor Laboratory,Cold Spring Harbor, New York, 1997; 121-160.
7. Levin J G, Guo J, Rouzina I, and Musier-Forsyth K: Nucleic acid chaperone activity of HIV-1 nucleocapsid protein:Critical role in reverse transcription and molecular mechanism.*Prog Nucleic Acid Res Mol Biol* 2005; **80**: 217-286.
8. Hulme AE, Perez O, and Hope TJ: Complementary assays reveal a relationship between HIV-1 uncoating and reverse transcription. Proc Natl Acad SciUSA 2011; **108**(24): 9975–9980.
9. Song B: TRIM5alpha. *Curr Top Microbiol Immunol* 2009; **339**: 47-66.
10. Towers GJ: The control of viral infection by tripartite motif proteins and cyclophilin A. Retrovirology 2007; **4**: 40.
11. Dismuke DJ and Aiken C: Evidence for a functional link between uncoating of the human immunodeficiency virus type 1 core and nuclear import of the viral preintegration complex. *J Virol* 2006; **80**(8): 3712–3720.
12. Yamashita M and Emerman M: Capsid is a dominant determinant of retrovirus infectivity in nondividing cells. *J Virol* 2004;**78**(11):5670–5678.
13. Coffin J, Varmus H,Hughes S, eds. ,Brown PO: Integration. In: Retroviruses ,Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, 1997, pp. 161–205.
14. J. A. G. Briggsa,1, J. D. Richesa, B. Glassb, V. Bartonovab, G. Zanettia,c and H.-G. Kräusslichb,+ Author Affiliations ,Structure and assembly of immature HIV.
15. Dettenhofer M, Yu XF.Source ,Proline residues in human immunodeficiency virus type 1 p6(Gag) exert a cell type-dependent effect on viral replication and virion incorporation of Pol proteins., Department of Molecular Microbiology and Immunology, Johns Hopkins University School of Hygiene and Public Health, Baltimore, Maryland 21205, USA.