

Original Article

DETERMINATION OF 3D STRUCTURE OF GAG POLY PROTEIN ISOLATE 90CF056 OF HIV TYPE 1 BY HIDDEN MARKOV MODEL AND NEURAL NETWORKS A POSSIBLE DRUG TARGET

UNDETY BENJAMIN JASON¹, DANIEL ALEX ANAND²

¹Department of Bioinformatics, Sathyabama University, Chennai, Tamilnadu, India, ²Department of Biomedical Engineering, Sathyabama University, Chennai, Tamilnadu, India.
Email: ubjason@gmail.com

Received: 01 Jul 2014 Revised and Accepted: 05 Aug 2014

ABSTRACT

Introduction: The study of understanding the structural and molecular conservation 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 poly protein suggest some approaches to be as a target for novel drugs.

Objective: The functional conservation of HIV-1 Gag indicates rational drug design taking Gag as he drug target¹. HIV-1 may be blocked by targeting gag poly protein. This could proffer new scheme for novel drug classes that could complement current HIV-1 treatment options.

Methods: The crystal structure of Gag poly-protein is unavailable. The templates similar are much smaller in size and thus ab-initio method is applied to determine the three dimensional structure of gag poly-protein. The value given in the program is an approximation of the probability as provided by the software with neural networks. The predictions are designed to be limited, to a score

≥ 1.8 which is actually an approximation of the probability. The predictor is an artificial neural network. NN: Inputs indicates inclusion of separation and sequence length, e-value statistic which are based on mutual information values, a statistic based on propensity of residues in contact with each other.

Results: The local structure predictions are performed with neural networks for several different local structure alphabets, and hidden Markov models are created.

Conclusion: The complete three-dimensional model of the Gag poly protein is constructed by fold recognition and alignment to proteins in the Protein Data Bank is done.

Keywords: HMM, Gag poly-protein, Neural networks.

INTRODUCTION

HIV-1 which causes AIDS is a retrovirus in genus Lentiviridae. HIV-1 is an enveloped virus and it encodes two envelope (Env) glycoproteins, they are surface (SU) glycoprotein gp120 and a transmembrane (TM) glycoprotein gp41. The gene of HIV-1, Gag encodes 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). The HIV-1 start off with the binding of gp120 to target cell plasma membrane [1-4]. The principal binding site of the receptor for HIV-1 is CD4. The binding of gp120 to CD4 and co-receptor commences conformational changes in gp41, thereby directs to the fusion of the viral envelope and the target cell membrane and entry of the viral core into the host cell cytoplasm. Latest works on HIV-1 suggests that HIV-1 entry is possible even in a low-pH endosomal compartment after receptor-mediated endocytosis [5]. Once the virion enters 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 gag-pol enzymes RT and IN, along with 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 nucleic acid chaperone at multiple repeated steps during reverse transcription for the conversion of RNA to DNA [7]. The protein Vpr is a component of the reverse transcription complex. The process of reverse transcription and uncoating looks to be temporally linked, [8] and it is obvious 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 process of nuclear import remains not understood completely; however, the role for CA in this

process [11, 12] indicates that some CA protein may remain associated with the viral nucleoprotein complex as it traffics to the nuclear pore. Reaching inside the nucleus, the double-stranded viral DNA integrates into the target cell genome through the action of the IN enzyme. [13] The integrated viral DNA works as the template for transcription from the viral promoter in the 5' long terminal repeat (LTR) to produce the spliced viral mRNAs and full-length genomic RNAs; these are transferred out of the nucleus via the action of the Rev protein. The Gag proteins are translated from full-length message as a polyprotein precursor which contains MA, CA, NC, and p6 domains along with other two spacer peptides, SP1 and SP2. [15] During the process of translation of Gag precursor, known as Pr55^{Gag}, an occasional 1 ribosomal frame shift 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 Gag and GagPol is regulated by the MA domain, plays an important role in the incorporation of the viral Env glyco proteins. This assembly identified in cholesterol rich membrane microdomains (lipid rafts) through direct interactions between MA and the phospholipid phosphatidylinositol-4, 5-bisphosphate [PI(4,5)P₂]. [17] Interactions within the CA domain of Gag regulate 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 on the N-terminal of the Gag polyprotein. It induces cellular proteins Tsg101 (a component of ESCRT-1) and Alix to initiate virus particle budding from the plasma membrane. Gag proteins are plays vital role in virus assembly, release, maturation and function in the establishment of a productive HIV 1. Though they play vital role throughout the replication cycle, there are no drugs targeting Gag poly protein.

Latest progress in understanding the structural and cell biology of HIV-1 Gag function has revealed a number of potential Gag-related targets for possible therapeutics.

MATERIALS AND METHODS

Sequence of Gag Poly Protein

The sequence of the Gag poly protein is retrieved from the Uniprot database whose structure is to be predicted. The complete sequence with 500 amino acids is given below:

>sp|O93182|GAG_HV190 Gag polyprotein OS=Human immune deficiency virus type 1 group M subtype H (isolate 90CF056) GN=gag PE=3 SV=3

MGARASVLSGGKLDAAWEKIRLRPGGKKKYRLKHLVWASRELERFALNP
GLETPEGCLQIEIQPAIKTGTEELKSLFNLVAVLYCVHRKIDVKDTKE
ALDKIEIQNKSQKTQAAADKEKDNKVSQNYPIVQNAQGMVHQAI
SPRTLNAWVKVVEEKAFSPEVPMFSALSEGATPQDLNAMLNTVGGHQ
AAMQMLKDTINEEAAEWDVHPVHAGPIPPGQMREPRGSDIAGTTSTL
QEQLAWMTGNPAIPVGDYKRWIILGLNKIVRMYSPVILDIKQGPKEPF
RDYVDRFFKTLRAEQATQDVKNWMTETLLVQNANPDCKTILRALGGQ
ASIEEMMTACQGVGSPSHKARVLAEMSQVTNTNTAIMMQKGNFKGQ
RKFVKCFNGKEGHIARNCRAPRKKGCWKCGRGHQMKDCTERQANF
LGKIWPSSKGRPGNLFQSRPEPTAPPAESFGFGEEMTPSPKQEQDKDE
PPLASLRSLFGSDPLLQ

This is a hidden markov model based methodology which is implemented by SAM_T08 in the CASP8 experiment. Primary results indicate that it is a good prediction method, but that meta servers combining results from several primary servers are likely to produce somewhat better results.

The SAM software finds similar protein sequences in NR and then aligns them, providing sequence logos that can show relative conservation of different positions. Local structure predictions are done with neural nets for several different local structure alphabets, and hidden Markov models are created. Fold recognition and alignment to proteins in the Protein Data Bank are done, and a full three-dimensional model is subsequently constructed.

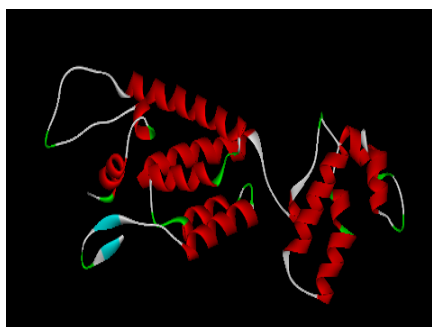


Fig. 1: Predicted structure of Gag polyprotein

RESULTS AND DISCUSSION

Models are predicted using HMM based protein structure prediction by Sequence Alignment and Modeling System, SAM-T08. By SAM-T08, E-values less than about $1.0E-5$ could be very good hits and are very likely to have a domain of the same fold as the target. The E values between $1.0E-5$ and 0.1 the goodness of the match will vary somewhat from target to target, but will often be a good match. When we get an extremely small E-value (say $1.e-10$ or smaller), then the alignments you get from SAM-T08 may not be any better than alignments that you get from sequence-sequence aligners like Smith-Waterman, FASTA, or BLAST. SAM-T08 is designed to do good fold recognition and alignment in the difficult cases, and it may give up some performance on the "easy" ones.

Five models are evaluated by the Rampage [18] with respect to Ramachandran plot to check the percentage residues in favored

regions and unfavored regions. All the five models have their residues more than 95% in favored region according to Table 1. But the model 4 has 98.1% residues in favored region. Fig1 shows the three dimensional structure of the best predicted model.

Table 1: Percentage of residues in favored regions by Rampage of five predicted models

S. No.	Predicted Model	Percentage of residues in favored regions by Rampage
1	Model 1	95.9%
2	Model 2	96.0%
3	Model 3	97.2%
4	Model 4	98.1%
5	Model 5	95.9%

The best predicted model evaluated by Rampage, University of Cambridge is shown in Fig2 which was found to have best score being validated out of the five models.

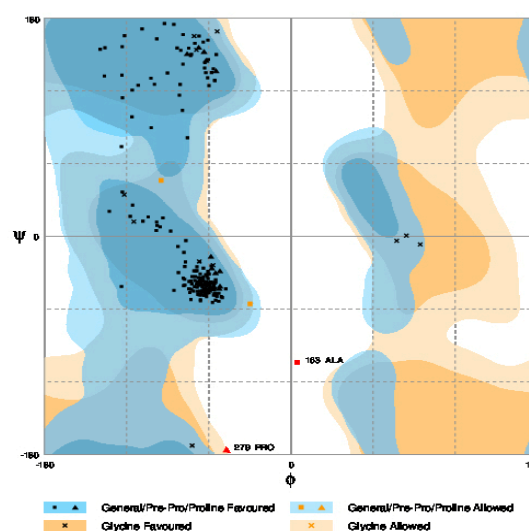


Fig 2: Gag polyprotein best model

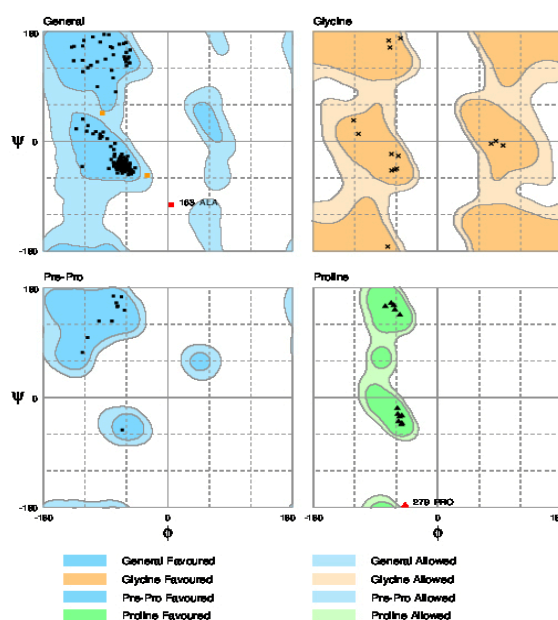


Fig 3: Residues in the unfavored region

Number of residues in favoured region (~98.0% expected): 209 (98.1%)

Number of residues in allowed region (~2.0% expected): 2 (0.9%)

Number of residues in outlier region: 2 (0.9%)

The SAM-T08 method builds models according to the size of the input sequence. Predicting domain boundaries when structure is unknown is an art that we have not attempted to automate. We have generally found it best to do a search primarily with the full-length protein and then remove any domains that are strongly predicted, and do the prediction again on what is left. A weaker prediction for a second domain may be masked by strong predictions for the more easily found domain in the full-length protein subsequently.

The two residues found in unfavored regions are Alanine and Proline in 163 & 279 position. These residues have steric hindrance because of Alanine without a side chain and Proline with no free side chain. Thus except these two residues all other amino acids in the predicted model fall in favored and slightly favored regions.

SUMMARY AND 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 HMM and neural networks. This work further carries in finding active site, lead screening and docking studies of Gag poly protein.

CONFLICT OF INTERESTS

Declared None

REFERENCES

- Guangdi Li, Jens Verheyen, Soo-Yon Rhee, Arnout Voet, Anne-Mieke Vandamme, Kristof heys:Functional conservation of HIV-1 Gag:implications. *J Retrovirology* 2013;10:126.
- Doms RW:Beyond receptor expression:The influence of receptor conformation, density, and affinity in HIV-1 infection. *J Virology* 2000;276(2):229-37.
- Freed EO, Martin MA, Knipe DM, Howley PM. HIVs and their replication. In:Fields Virology. Lippincott,Williams and Wilkins;Philadelphia:2007.p. 2107-85.
- Melikyan GB:Common principles and intermediates of viral protein-mediated fusion:The HIV-1 paradigm. *J Retrovirology* 2008;5:111.
- Miyauchi K, Kim Y, Latinovic O, Morozov V, Melikyan GB:HIV enters cells via endocytosis and dynamin-dependent fusion with endosomes. *J Cell* 2009;137(3):433-44.
- Telesnitsky A, Goff SP, Coffin J, Varmus H, Hughes S. Reverse transcriptase and the generation of retroviral DNA. In:Retroviruses. Cold Spring Harbor Laboratory, Cold Spring Harbor;New York:1997.p.121-60.
- Levin JG, Guo J, Rouzina I, Musier-Forsyth K:Nucleic acid chaperone activity of HIV-1 nucleocapsid protein:Critical role in reverse transcription and molecular mechanism. *J Prog Nucleic Acid Res Mol Biol* 2005;80:217-86.
- Hulme AE, Perez O, Hope TJ:Complementary assays reveal a relationship between HIV-1 uncoating and reverse transcription. *J Proc Natl Acad Sci USA* 2011;108(24):9975-80.
- Song B. TRIM5alpha. *J Curr Top Microbiol Immunol* 2009;339:47-66.
- Towers GJ. The control of viral infection by tripartite motif proteins and cyclophilin A. *J Retrovirology* 2007;4:40.
- Dismuke DJ, 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):3712-20.
- Yamashita M, Emerman M:Capsid is a dominant determinant of retrovirus infectivity in nondividing cells. *J Virol* 2004;78(11):5670-78.
- Brown PO, Coffin J, Varmus H, Hughes S. Integration In:Retroviruses Cold Spring Harbor Laboratory, Cold Spring Harbor;New York;1997.p.161-205.
- J Chiu HC, Wang FD, Chen YM, Wang CT, National M. *Virol. Jul*;87(Pt 7):.Effects of human immunodeficiency virus type 1 transframe protein p6* mutations on viral protease-mediated Gag processing. *J SourceInstitute of Clinical University Taipei Taiwan* 2006:2041-6.
- Briggs JD, Richesa B, Glassb V, Bartonovab G. Structure and assembly of immature HIV J. *J A G Zanettiac and HG Krusslichb Author Affiliations* 1.
- J Huang M, Orenstein JM, Martin MA, Freed EO, Diseases NIH, et al. Nov:;.p6Gag is required for particle production from full-length human immunodeficiency virus type 1 molecular clones expressing protease. *J SourceLaboratory of Molecular of Allergy and Infectious* 20460 USA 2009;69(11):6810-8.
- Dettenhofer M, Yu XF, Johns I, Maryland USA. 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. *J Source Department of Molecular Microbiology and University School of Hygiene and Public Health Baltimore*;21205.
- Lovell IW, Davis WB, Arendall, III, de Bakker JM, Word MG, Prisant JS. SC PIW Richardson and D. C Richardson Structure validation by Calpha geometry phipsi and Cbeta deviation. *J Proteins Structure Function Genetics* 2002;50:437-50.