



## A COMPLETE HIV-1 *Env* CODING SEQUENCES FROM HIV ISOLATED IN CENTRAL JAVA, INDONESIA

Irvan Raharjo<sup>1</sup>, AfionoAgung Prasetyo<sup>1,2,3\*</sup>

<sup>1</sup> Biomedical Laboratory Faculty of Medicine, Sebelas Maret University, Jl. Ir. Sutami 36A, Surakarta, Indonesia.

<sup>2</sup> Department of Microbiology Faculty of Medicine, Sebelas Maret University,  
Jl. Ir. Sutami 36A, Surakarta, Indonesia.

<sup>3</sup> Center of Biotechnology and Biodiversity Research and Development, Sebelas Maret University,  
Jl. Ir. Sutami 36A, Surakarta, Indonesia.

\* Correspondence author: AfionoAgungPrasetyo, M.D., Ph.D. Department of Microbiology Faculty of Medicine, Sebelas Maret University Jl. Ir. Sutami no 36A, Surakarta, Indonesia, Telp. 62-271-632489, Fax. 62-271-632489, E-mail: afieagp@yahoo.com; afie.agp.la@gmail.com

### ABSTRACT

Background: The HIV envelope glycoprotein (*env*) plays a central role in viral transmission to target cells. The genetic diversity of *env* gene contributes a great mechanism on how HIV-1 could survive. Aims: To determine the genotype and genetic relationship of our isolate compared to the other HIV-1 isolates, especially from Indonesia based on the *env* gene. Also, for further use, the molecular properties of our gene had been analyzed. Methods: A complete coding sequence of HIV-1 *env* gene was cloned from 09IDSKA-6 (HIV isolated in Central Java, Indonesia), inserted into an *Escherichia coli* expression plasmid, and sequenced. The sequencing results of the *env* gene then subjected for virus subtyping and genotyping. The phylogenetic analysis was performed to know the relationship between the isolate with all HIV isolates in Indonesia based on the *env* gene. The sequencing results were also subjected for bioinformatics analysis. Results: Based on the molecular analysis of HIV-1 *env* complete coding sequences, the 09IDSKA-6 belong to HIV CRF01\_AE which is quite common in Southeast Asia. The signal peptide, protein charge, transmembrane helix, antigenicity, hydrophobicity, secondary structure, and proteolytic cleavage of our gene also had been discussed.

Key words: HIV-1, *env* gene, Indonesia.

### INTRODUCTION

The envelope glycoproteins (*env*), gp120 and gp41, of HIV-1 play a central role in viral transmission to target cells, and mediate attachment and incorporation of the virus into the cells through specific interaction with CD4 receptor and chemokine receptors. A detailed understanding of the morphology of the HIV-1 *env* spike is key to understanding viral pathogenesis and for vaccine design (Zhu, *et al.*, 2008). *Env* is a major target of humoral immune responses (Pantophlet and Burton, 2006). Even so, *env* gp120 and gp41 are known to be the most variable HIV-1 proteins with typical intersubtype and intrasubtype differences soaring to 35% and 20%, respectively (Gaschen *et al.*, 2002). The persistence of HIV-1 infection in the presence of robust host immunity has been associated with variation in viral envelope proteins leading to antigenic variation and escape from neutralizing antibodies (Kalia, *et al.*, 2005). In present study, we tried to analyze our isolate sequence, performed phylogenetic analysis comparing it with the other HIV-1 *env* coding sequences from Indonesia, and bioinformatics analysis for the predictive protein.

## **MATERIALS AND METHODS**

### **Cloning and Sequencing**

A complete coding sequence of *env* gene was cloned from 09IDSKA-6 (HIV isolated in Central Java, Indonesia) by nested PCR as described previously (Kemal *et al.*, 2009). The PCR product was inserted into pETBlue-1 (an *Escherichia coli* expression plasmid). The recombinant plasmid was then transformed into competent cells, propagated, purified, and sequenced. The sequencing step was performed in both orientations for confirmation, three times for each orientation.

### **Nucleotide Analysis**

Our isolate was genotyped using Human immunodeficiency virus 1 (HIV-1) – 2009 RefSet, followed by BLAST analysis. HIV isolate with highest BLAST score and all HIV-1 genome from Indonesia were retrieved and aligned by ClustalW using MEGA version 5.2.1 (Tamura *et al.*, 2011). Phylogenetic tree analysis was generated using the neighbor-joining method based on Kimura 2-parameter model with 1000 bootstrap replicates.

### **Predicted Protein Analysis**

SignalP 4.1 software was applied for signal peptide prediction (Petersen *et al.*, 2011). ProtParam Tool software was used in order to find isoelectric point (pI) first (Gasteiger *et al.*, 2005). TMHMM v.2.0 software was used to predict transmembrane helical domain. The antigenic regions and hydrophobicity of our isolate were analyzed based on EMBOSS Antigenic (Kolaskar and Tongaonkar, 1990) and ProtScale (Gasteiger *et al.*, 2005) in Kyte and Doolittle Scale softwares. PSIPRED v3.3 was applied to spot the secondary structure (Jones, 1999). Prediction of the cleavage sites of multiple protease families was made using PROSPER software (Song *et al.*, 2012).

## **RESULTS AND DISCUSSION**

### **Genotype Determination and Phylogenetic Analysis**

There were only four HIV-1 *env* complete genome coding sequences originated from Indonesia in GenBank (Accession number: AB485653, AB485652, AB547463, and AB547464), and all of them were retrieved for sequences alignment. Subtype reference 2010, which represents the majority of HIV-1 subtypes, was also aligned. AB547464 has been listed in this subtype reference 2010 as a CRF33\_01B reference. The BLAST highest score was HIV-1 clone pCM235-2 from Thailand (Accession number AF259954), followed by HIV-1 clone pCM235-4 from USA (Accession number AF259955). Based on genotyping and phylogenetic analysis of *env* gene, our isolate belongs to CRF01\_AE and closely related to AB485653, AB485652, AF259954, and AF259955 (Figure 1).

### **Protein Characteristic**

Based on SignalP 4.1 analyzed, our predicted HIV-1 *env* protein had no signal peptide. C-score predicts the signal peptide cleavage site, immediately will be high after possible cleavage site. S-score is signal peptide score, which will recognize possible signal peptide and will be high right before the cleavage sites. Y-score is the combination of the C-score

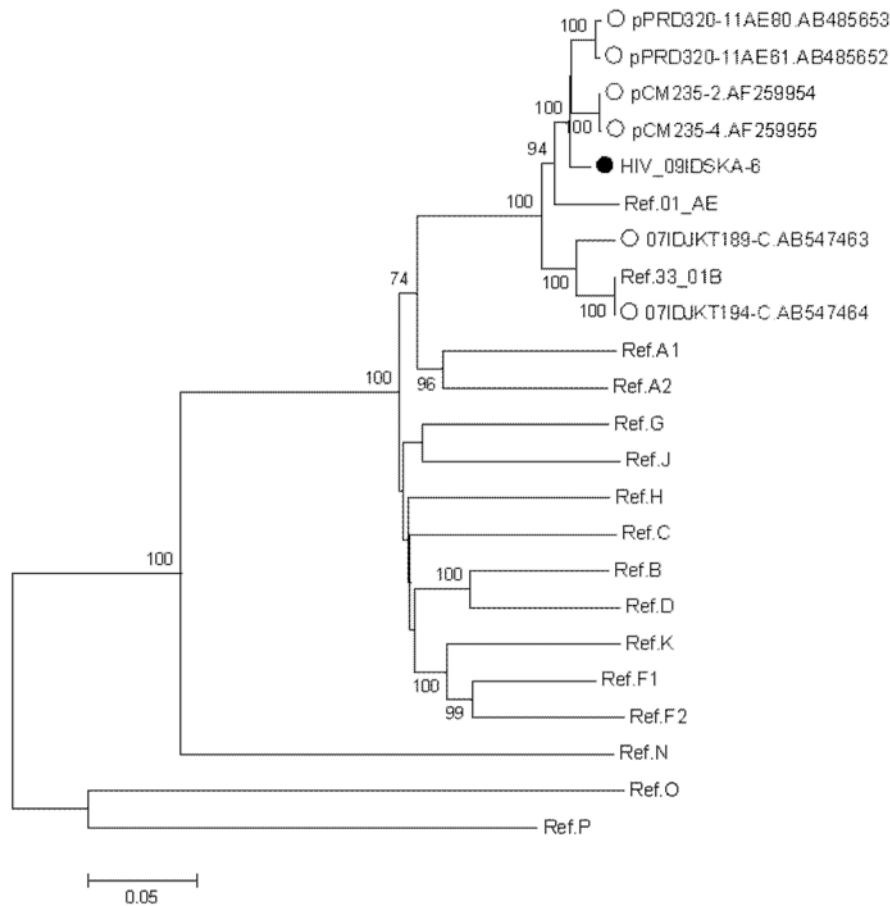


Figure 1. Phylogenetic analysis based on HIV-1 *env* gene of the HIV-1 isolated in Central Java, Indonesia (09IDSKA-06)(●) with all HIV-1 isolated from Indonesia (○). The scale bar indicates 5% nucleotide sequence divergence. Values on the branches represent the percentage of 1000 bootstrap replicates and bootstrap values over 60% were shown in the tree. GenBank accession numbers of the reference isolates in the phylogenetic tree were as follows: Ref.01\_AE: GQ477441, Ref.33\_01B: AB547464, Ref.A1: DQ676872, Ref.A2: AF286238, Ref.G: AF084936, Ref.J: EF614151, Ref.H: AF190127, Ref.C: U52953, Ref.B: K03455, Ref.D: K03454, Ref.K: AJ249235, Ref.F1: AF077336, Ref.F2: AY371158, Ref.N: AY532635, Ref.O: L20587, Ref.P: HQ179987; complete genome coding sequences originated from Indonesia: pP RD 320-11AE80: AB485653, pP RD 320-11AE61: AB485652, 07IDJKT189-C: AB547463, 07IDJKT194-C: AB547464; and highest score from BLAST analysis: pCM235-2: AF259954, pCM235-4: AF259955.

and the slope of the S-score, predicts cleavage site more accurately. Mean S is the average S-score of the possible signal peptide, from position 1 to position 29, a point before possible cleavage site. It could be predicted that signal peptide was located between position 1 to 29. The last score that will be used is D-score, the discrimination score. It is the score used to discriminate signal peptide from non-signal peptides, obtained from the average of the mean S and maximal Y-score (Petersen *et al.*, 2011). The D-score obtained was 0.422, which did not surpass the cut off score of 0.500; therefore the cleavage site found in our predictive protein was not the cleavage site of a signal peptide, since our predictive protein should be a mature protein.

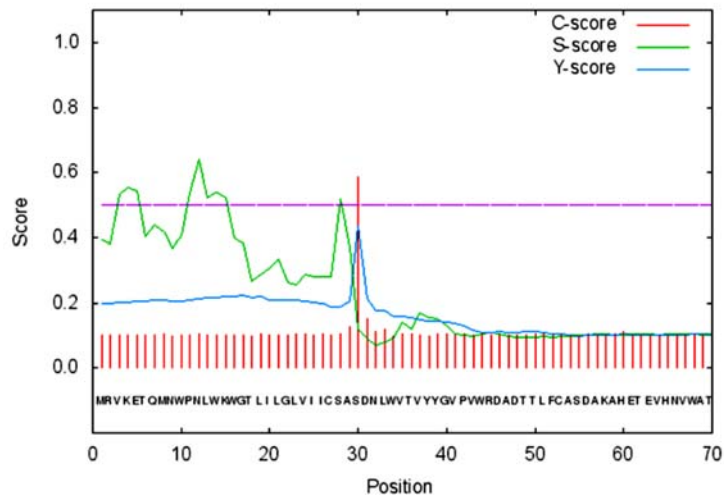


Figure 2. SignalP 4.1 prediction of a predicted protein from a complete *env* gene of HIV isolated in Central Java, Indonesia (09IDSKA-06). The graph indicates three output scores: raw cleavage site (C) score, signal peptide (S) score, and combined cleavage site score (Y) for each position in the sequence. Measures that will be used to do the analysis were as follows: Maximal C-score: 0.585 in position 30, Maximal Y-score: 0.435 in position 30, Maximal S-score: 0.639 in position 12, Mean S: 0.404 in position 1 to 29, D-score: 0.422 in position 1 to 29. The cut off point was 0.500.

The theoretical isoelectric point (pI) of our predicted HIV-1 *env* protein was 9.23, based on ProtParam Tool analysis. At a pH below 9.23, our predicted protein had a net positive charge and at a pH above 9.23, it had a net negative charge. Net charge could possibly influence the characteristic of gp120 and gp41, especially the net charge in the outer domain of gp120, of which majority of interactions with other cells happened. V3 loop net charge, which is located in the outer domain of gp120, is responsible for the conformation of the loops involved in binding to CD4. Virus carrying the gp120 with V3 net charge of +3 was resistant to neutralization by anti-CD4 binding site monoclonal antibodies, while the one carrying V3 net charge of +7 was not resistant (Yokoyama *et al.*, 2012). Moreover, there were three transmembrane helical domains predicted by TMHMM v2.0 (Figure 3).

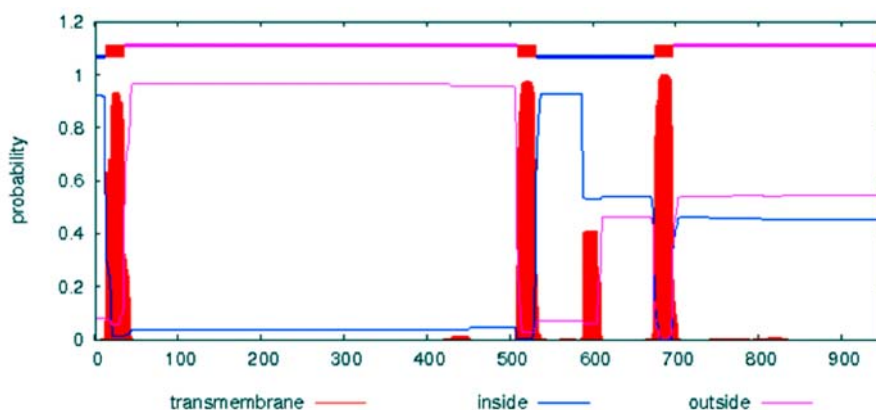


Figure 3. Transmembrane helical domain probabilities of a predicted protein from a complete *env* gene of HIV isolated in Central Java, Indonesia (09IDSKA-06).

In our predicted HIV-1 *env* protein, Isoleucine (Ile) was the most hydrophobic amino acid, based on Kyte and Doolittle Scale. There were two positions of which hydrophobicity score surpassed score of 3 (Figure 4). Amino acids with high level of hydrophobicity are generally buried within the hydrophobic core of the protein or within the lipid portion of the membrane (Betts & Russel, 2003). Our predictive protein also contains 35 possible antigenic regions (data not shown). The hydrophobic residues Cysteine (Cys), Leucine (Leu), and Valine (Val), if they occur on the surface of a protein, are more likely to be a part of antigenic sites (Kolaskar & Tongaonkar, 1990). Antigenic sites are important in *env* glycoprotein, as they interact with surface cells (Van Regenmortel, 2001). Amount of antigenic sites will greatly affect the antigenicity of the envelope glycoprotein (Kwong *et al.*, 1998).

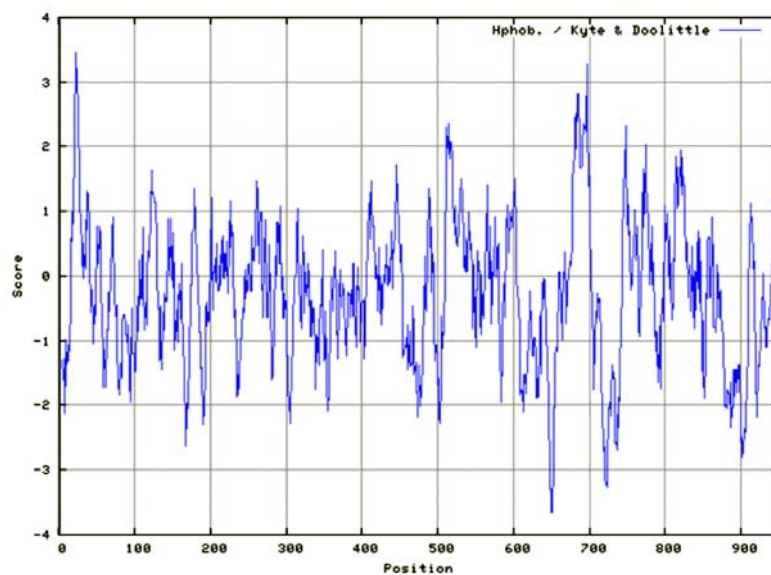


Figure 4. Hydrophobicity score of a predicted protein from a complete *env* gene of HIV isolated in Central Java, Indonesia (09IDSKA-06).

Our predicted HIV-1 *env* protein had both  $\alpha$ -helix and  $\beta$ -sheet secondary structures. The  $\alpha$ -helix secondary structure was predictably more common in our isolate, followed by  $\beta$ -sheet secondary structure. Secondary structures contribute to the constitution of the overall three-dimensional configuration of the protein chain. The  $\alpha$ -helix was more frequent in half-end of the sequence. Probably, that part constitutes outer domain of the glycoprotein, since  $\alpha$ -helix is quite common in surface of a protein (Richardson, 1981).

Our predicted HIV-1 *env* protein was found may be cleaved by some different proteases, with Serine protease as the most frequent. Proteolytic cleavage serves in many purposes, for an instance, it could be the system that regulates some physiological and cellular processes, or even the process that is needed in maturation. Proteolysis of the precursor envelope glycoprotein gp160 results in the formation of gp120 and gp41, which are important in viral infectivity (Rodriguez *et al.*, 1995).

Finally, we concluded that our isolate was CRF01\_AE. Our predictive protein should be a mature protein, no signal peptide, had a tendency to be positively charged and diverse hydrophobicity, with three transmembrane helical domains, had both  $\alpha$ -helix and  $\beta$ -sheet secondary structures, 35 possible antigenic regions, and may be cleaved by proteases. Taken

all data together, it will increase our understanding of our isolate characteristics for further use.

## ACKNOWLEDGMENT

This work was supported partially by grants from APBN/DIPA UNS (No. 2342/UN27.16/PN/2012 and No. 267a/UN27.16/PN/2013).

## REFERENCES

- Betts, M.J., and R.B. Russel. 2003. Amino acid properties and consequences of substitutions. In: *Bioinformatics for Geneticists*. John Wiley & Sons Ltd., Chicester.
- Gaschen, B., J. Taylor, K. Yusim, B. Foley, F. Gao, D. Lang, V. Novitsky, B. Haynes, B.H. Hahn, T. Bhattacharya, and B. Korber. 2002. Diversity considerations in HIV-1 vaccine selection. *Science*, vol. 296:2354-2360.
- Gasteiger, E., C. Hoogland, A. Gattiker, S. Duvaud, M.R. Wilkins, R.D. Appel, and A. Bairoch. 2005. Protein identification and analysis tools on the ExPASy server. In: *The Proteomics Protocols Handbook*. Humana Press., New York.
- Jones, D.T. 1999. Protein secondary structure prediction based on position-specific scoring matrices. *J Mol Biol*, vol.292:195-202.
- Kalia, V., S. Sarkar, P. Gupta, and R.C. Montelaro. 2005. Antibody neutralization escape mediated by point mutations in the intracytoplasmic tail of human immunodeficiency virus type 1 gp41. *J Virol*, vol.79:2097-2107.
- Kemal, K.S., M. Reinis, B. Weiser, and H. Burger. 2009. Methods for viral RNA isolation and PCR amplification for sequencing of near full-length HIV-1 genomes. In: *HIV Protocols*. Humana Press., New York.
- Kolaskar, A.S., and P.C. Tongaonkar. 1990. A semi-empirical method for prediction of antigenic determinants on protein antigens. *FEBS Lett*, vol.276:172-174.
- Kwong, P.D., R. Wyatt, J. Robinson, R.W. Sweet, J. Sodroski, and W.A. Hendrickson. 1998. Structure of an HIV gp120 envelope glycoprotein in complex with the CD4 receptor and a neutralizing human antibody. *Nature*, vol. 393:648-659.
- Pantophlet, R., and D.R. Burton. 2006. GP120: target for neutralizing HIV-1 antibodies. *Annu Rev Immunol*, vol.24:739-769.
- Petersen, T.N., S. Brunak, G. Von Heijne, and H. Nielsen. 2011. SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat Methods*, vol. 8:785-786.
- Richardson, J.S. 1981. The anatomy and taxonomy of protein structure. *Adv Protein Chem*, vol. 34:339.
- Rodriguez, D., J.R. Rodriguez, and M. Esteban. 1995. Enhanced proteolytic processing of the human immunodeficiency virus type 1 envelope protein in murine Ltk(-) cells. *AIDS Res Hum Retroviruses*, vol. 11:81-85.
- Song, J., H. Tan, A.J. Perry, T. Akutsu, G.I. Webb, J.C. Whisstock, and R.N. Pike. 2012. PROSPER: an integrated feature-based tool for predicting protease substrate cleavage sites. *PLoS One*, vol. 7:e50300.
- Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. 2011. MEGA5: Molecular Evolutionary Genetics Analysis using Maximum Likelihood, Evolutionary Distance, and

- Maximum Parsimony Methods. *Mol BiolEvol*, vol. 28:2731-2739.
- Van Regenmortel MH. 2001. Antigenicity and immunogenicity of synthetic peptides. *Biologicals*, vol. 29: 209-213.
- Yokoyama M, Naganawa S, Yoshimura K, Matsushita S, Sato H. 2012. Structural dynamic of HIV-1 envelope gp120 outer domain with V3 loop. *PLoS One*, vol. 7:e37530.
- Zhu P, Winkler H, Chertova E, Taylor KA, Roux KH. 2008. Cryoelectron tomography of HIV-1 envelope spikes: further evidence for tripod-like legs. *PLoSPathog*, vol.11:e1000203.