KnE Life Sciences

ISSN 2413-0877 Volume 2 (2015) 34-40
The 3rd International Conference on Biological Science 2013
(The 3rd ICBS-2013)

MOLECULAR ANALYSIS OF A COMPLETE HIV-1 *Pol* GENE ISOLATED IN CENTRAL JAVA, INDONESIA

Ratna Sariyatun[1,2], Reviono[3], Afiono Agung Prasetyo[1,2,4*], Seiji Kageyama[5]

[1] Biomedical Laboratory Faculty of Medicine, Sebelas Maret University, Jl. Ir. Sutami 36A, Surakarta, Indonesia.
[2] Department of Microbiology Faculty of Medicine, Sebelas Maret University, Jl. Ir. Sutami 36A, Surakarta, Indonesia.
[3] Department of Pulmonology Dr. Moewardi General Hospital, Jl. KolonelSutarto 132 Surakarta, Indonesia.
[4] Center of Biotechnology and Biodiversity Research and Development, SebelasMaret University, Jl. Ir. Sutami 36A, Surakarta, Indonesia.
[5] Department of Microbiology and Immunology (Division of Virology) Faculty of Medicine, Tottori University, 86 Nishi choYonago, Japan.

[*] Correspondence author: AfionoAgungPrasetyo,M.D., Ph.D.Department of Microbiology Faculty of Medicine,

Sebelas Maret UniversityJl. 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: HIV-1 is an infectious agent causing global health problem. Molecular analysis of HIV-1 strains is crucial due to the association with viral fitness, drug resistance, serological failure, and more importantly, for vaccine development. However, HIV-1 molecular data in Indonesia is limited, including that of the HIV *pol* gene. Aims: The present study performed a molecular analysis of the *pol* gene of HIV isolated in Central Java, Indonesia, in order to enrich HIV-1 molecular data in Indonesia, particularly in Central Java.Methods: A complete coding sequence of *pol* 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 *pol* gene then subjected for virus subtyping, genotyping, and phylogenetic analysis. The drug resistance, identification of the HLA binding motifs and viral epitopes, proteomic motifs, and physicochemical analysis was performed.Results: The HIV-1 isolate analyzed in this study was CRF01_AE strain. The isolate remains in low evolutionary state due to the low level of nonsynonymous to synonymous substitutions ratio (dN/dS). None of mutations identified was related to ARV drug resistance. A total of 33 CTL/CD8+ epitopes, 6 T-helper/CD4+ epitopes and 2 antibody epitopes were identified in our isolate. Six distinct proteomic domains were found. Physicochemical analysis revealed the molecular weight (Mw), estimated half-life, instability index, aliphatic index, and hydrophilicity of each proteins encoded by the complete HIV-1 *pol* gene.

Key words: HIV-1, pol gene, Indonesia.

INTRODUCTION

Human immunodeficiency virus type 1 (HIV-1) is the main cause of the global pandemic of acquired immunodeficiency syndrome (AIDS).Previous studies inferred that theHIV geneticscould influence the disease progression (Kiwanuka *et al*., 2010), diagnostic assays (Bruzzone *et al*., 2010), virologic follow-up and therapeutic monitoring (Sire *et al*., 2011). HIV-1 *pol* gene is the one frequently analyzed, as this gene is the most conserved region within the genome and encodes proteins indispensably required in HIV replication (viral-specific enzymes of protease (PR), reverse transcriptase (RT), RNase H and integrase (IN)).However, the genetic information of HIV-1 *pol* gene in Indonesia remains limited. Indeed, only four complete HIV-1 *pol* genes from Indonesia, none from Central Java,aredeposited in GenBank. In this study, we aimed to isolate HIV strain from Central Java and analyze the virus genetics, particularly related to drug resistances, viral epitopes

ISSN 2413-0877 © 2015 The Authors.
Published by KnowledgeE Publishing Services This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0)
Selection and Peer-review under responsibility of the 3rd ICBS-2013
Doi http://dx.doi.org/10.18502/kls.v2i1.113

prediction and proteomic profile.Hence, our datawould contribute novel information about HIV geneticsin Indonesia, particularly in Central Java.

MATERIALS AND METHODS

Cloning and Sequencing

A complete coding sequence of *pol* 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.

Determination of Genotype/Subtype

We performed a BLAST analysis to confirm the sequencing results. Initial subtyping was conducted using the NCBI genotyping tool and REGA subtyping tool. A multiple alignment against reference sequences was reconstructed using ClustalW as implemented in MEGA5 software (Tamura *et al.*,2011).Phylogenetic tree was built using neighbor-joining method with Kimura two-parameter and bootstrap value of 1000 replicates.

Nucleotide and Amino Acid Substitution Analysis

Synonymous and non-synonymous amino acid substitution events were computed using the Nei-Gojoborimethod with the Jukes-Cantor correction (Nei and Gojobori, 1986) implemented in MEGA5 software and Synonymous Non-synonymous Analysis Program (SNAP).

Drug Resistance Analysis

Amino acid sequence of 09IDSKA-6 was submitted to the Stanford HIVdbserver for the identification of drug resistance-related mutations. References of each mutational event werelinked within the server web address.

Immunological Analysis

In order to estimate the possible interaction between viral sequence and human leukocyte antigens (HLAs) corresponding to both MHC class I (A, B, C) and class II (DP, DQ, DR), we performed an HLA binding motif analysis using Motif Scan tool linked inthe Los Alamos HIV database.

Proteomic analysis

The amino acid sequences of our isolate wasscanned against the PROSITE collection of motifs. Our analysis excluded motif with high probability of occurrence from the scan. Further, we computed the physical and chemical parameters (including molecular weight (Mw), estimated half-life, instability index, aliphatic index and grand average of hydropathicity (GRAVY)) of the isolate using ProtParamserver. Computational and representative analysis for hydrophobicity and hydrophilicity of our isolate was performed using ProtScaleserver.

RESULTS AND DISCUSSION

In present study, we conducted molecular analysis of a complete HIV-1 *pol* gene isolated in Central Java, Indonesia. The subtyping analysis using the NCBI genotyping tool, REGA HIV subtyping tool and phylogenetic treeshowed that our isolate, 09IDSKA-6, was a CRF01_AE strain. This was expected since CRF01_AE virus has been reported asthe predominant HIV-1 strain circulating in Indonesia (Sahbandar *et al.*, 2009; Merati *et al.*, 2012). Amino acid substitutions were assessed in order to estimate the viral evolution.We measured the nucleotide and amino acid substitution by counting the ratio of synonymous to non-synonymous substitutions (d_N/d_S). In SNAP analysis, the synonymous (d_S) and non-synonymous substitutions (d_N)were0.3871 and 0.0389, respectively (d_S/d_N= 9.9566).Whilst using MEGA5 software, the number of synonymous (d_S) and non-synonymous substitutions (d_N)were0.368 and 0.039, respectively (d_S/d_N= 9.4359).The nucleotide p-distance against consensus B sequence was 0.094. We calculated that the d_N/d_S value was 1.05×10^{-4} from SNAP analysis and 1.068×10^{-4} from MEGA5 analysis. In general, the d_N/d_S value greater than 1 is taken as evidence of positive Darwinian (diversifying) selection, while $d_N/d_S < 1$ refers an evidence of purifying selection (Hughes *et al.*, 2000).Thus, we referred that 09IDSKA-6 remained in a low evolutionary state, based on molecular analysis of *pol* gene. Moreover, the high synonymous substitution value (d_S) is supposed as an indication of conservation and essential region for biological function (Gordon *et al.*, 2003).

Certain mutations have been shown to influence antiretroviral drug susceptibility. In present study, we found thatour isolate was susceptible for all ARV regimens. However, we identified M36I, a minor tipranavir/r (TPV/r)-resistance mutation(Baxter *et al.*, 2006), which is also known as the consensus amino acid in CRF01_AE. Two polymorphic mutations were also identified, comprised one in the PR region (L89M) and another in RT region (K238R). Both L89M and K238R are not associated with decreased susceptibility to Protease Inhibitor(PI) and Non-Nucleotide Reverse Transcriptase Inhibitor (NNRTI), respectively.

Mutations also give direct impact tothe conformational changes of virus epitopes (antigenic determinants) which caninduce the failure of immune recognition by HLA class I (A, B, C) orII (DP, DQ, DR), a process required to initiate immune response by CTLs/CD8+ T cells(induced by HLA class I) andT-helper/CD4+ T cells (induced by HLA class II). Of all HIV-1 epitopes deposited in Los Alamos HIV Immunology Database, we detected a total of33, 12, and 2 epitopes recognized by CTL/CD8+, T-helper/CD4+, and human antibodies, respectively,within our isolate (Table 1).

We conducted motif analysis using PROSITE server and we found six domainsin our sequence.An aspartyl protease (retroviral-type family profile) was identified in aa 76-145 (score = 32.666), with a predicted active site in aa 81 (D/aspartate) of this domain. A reverse transcriptase (RT) catalytic domain profile was identified inaa 199-388 (score = 44.011), with three predicted metal domains (D/aspartate) positioned in aa 265, 340 and 341.RNase H domain profile lied from aa 588-711 (score = 19.593). Close by it, a zinc finger integrase-type profile was predicted to be in aa 717-758 (score = 16.269). Then, integrase catalytic domain profile was identified in aa 768-918 (score = 22.032), with two predicted metal domains (D/aspartate) in aa 778 and 830. The last domain identified was integrase DNA binding domain, lying in aa 937-984 (score = 17.218). In addition, one pat-

Table 1. List of human T-helper/CD4+ epitopes, human CTL/CD8+ epitopes and human antibody epitopes identified in HIV-1 09IDSKA-6*pol* predicted protein. Table position refers to the amino acid positions of HXB2 (Accession Number K03455).

Human T-Helper/CD4+ Epitopes			
Epitope	**Protein**	**Position**	**HLA**
FRKYTAFTIPSINNE	RT	124-138	DR, DR supermotif, I-A^{b}
SPAIFQSSMTKILEP	RT	156-170	DR, DR supermotif, I-A^{b}
KDSWTVNDIQKLVGK	RT	249-263	DR, DR5, DR5(11.01)
PLTEEAELELAENRE	RT	294-308	
TYQIYQEPFKNLKTG	RT	338-352	DP4
WEFVNTPPLVKLWYQ	RT	414-428	DR supermotif
LKTAVQMAVFIHNFK	Integrase	172-186	
KTAVQMAVFIHNFKR	Integrase	173-187	DR, DR supermotif
QKQITKIQNFRVYYR	Integrase	214-228	DR, DR supermotif
KQITKIQNFRVYY	Integrase	215-227	
LWKGEGAVVIQDNSDIKV	Integrase	242-259	
VIQDNSDIKVVPRRKAKI	Integrase	250-267	
Human CTL/CD8+ Epitopes			
Epitope	**Protein**	**Position**	**HLA**
ITLWQRPLV	Protease	3-11	A*6802, A*7401
RQYDQILIEI	Protease	57-66	B13
KAIGTVLV	Protease	70-77	B57
LVGPTPVNI	Protease	76-84	A*0201
TPVNIIGRNML	Protease	80-90	B81
GPKVKQWPL	RT	18-26	B*0801
KLVDFRELNK	RT	73-82	A*0301
GIPHPAGLK	RT	93-101	A*0301
TVLDVGDAY	RT	107-115	B*3501
TAFTIPSI	RT	128-135	B*5101
NETPGIRYQY	RT	137-146	B18
IRYQYNVL	RT	142-149	B*1401
LPQGWKGSPA	RT	149-158	B*5401
SPAIFQSSM	RT	156-164	B7
AIFQSSMTK	RT	158-166	A*0301, A*1101
VIYQYMDDL	RT	179-187	A*0201
LVGKLNWASQIY	RT	260-271	B*1501
KLNWASQIY	RT	263-271	A*3002
IYQEPFKNLK	RT	341-350	A*1101
PIQKETWETW	RT	392-401	A*3201
GAETFYVDGA	RT	436-445	A*6802
ETKLGKAGY	RT	449-457	A*2601
IVTDSQYAL	RT	495-503	Cw*0802
VTDSQYALGI	RT	496-505	B*1503
LFLDGIDKA	RT-Integrase	560-8	B81
LPPIVAKEI	Integrase	28-36	B*4201
IEAEVIPAET	Integrase	84-93	B*4002
HTDNGSNF	Integrase	114-121	Cw*05
KTAVQMAVF	Integrase	173-181	B*5701
IIATDIQTK	Integrase	203-211	A*11, A*1101
KIQNFRVYY	Integrase	219-227	A*3002
VPRRKAKII	Integrase	260-268	B42
RKAKIIRDY	Integrase	263-271	B*1503
Human Antibody Epitopes			

tern was identified on aa 78-89, and was predicted as eukaryotic and viral aspartyl proteases active site.

Protein physicochemical properties, like polarity, hydropathicity, charge, volume, aromaticity, aliphaticity and hydrogenation are shown to play a role in determining the rate and pattern of protein evolutions (Xia and Li, 1998).In present study, physicochemical profiling was perfomed using ProtParamserver to measure the molecular weight (Mw), estimated half-life, instability index (II), aliphatic index, and grand average of hydropathicity (GRAVY) of each protein encoded by our sequence (Table 2). Moreover, these datawere relevant for estimating the in-vivo ideal temperatures of function, solubility patterns in aqueous solution, and life-expectancies of the functional protein in vivo (Wayengera, 2011).

Table 2. Physicochemical profiling of each protein encoded by HIV-1 09IDSKA-06*pol* gene.

Protein	Molecular Weight (Mw)	Estimated Half-Life	Instability Index (II)	Aliphatic Index	Grand Average of Hydropathicity (GRAVY)
Protease (PR)	10754.7	>20 hours in mammalian reticulocytes, in vitro	32.42 (stable)	118.08	0.185
Reverse Transcrip-tase (RT)	51146.8	>20 hours in mammalian reticulocytes, in vitro	33.87 (stable)	82.12	-0.576
RNase	13144.8	2.8 hours in mammalian reticulocytes, in vitro	35.98 (stable)	99.83	-0.398
Integrase (IN)	32302.8	1.1 hours in mammalian reticulocytes, in vitro	34.36	85.00	-0.383

In present study, we identified the hydrophobicity and hydrophilicity values of each residue by using ProtScaleserver. Hydrophobicity value represents the non-polar residues in protein sequence and is commonly used to predict membrane helices (Chen and Rost, 2002). Otherwise, hydrophilicity implies a polar residue and shows the molecule tendency to be exposed tothe surface protein and may potentially be antigenic (Hopp and Woods, 1981).Putative membrane-spanning domain is defined for the site with high hydrophobicity (typically > 1.6 at the center based on the Kyte-Doolittle scale). Nonpolar amino acids generally perform greatest hydrophobicity and charged amino acids generally have lowest hydrophobicity.In all regions analyzed, we found that the maximum hydrophobicity value in Protease-, RT-, RNase and Integrase-encoding regions were located on aa 74 (2.486), aa 47 (-0.374), aa 48 (0.489) and aa 81 (1.763), respectively. Whilst the highest hydrophilicity value in those regions were aa 32 (1.186), aa 67 (2.129), aa 73 1.800 andaa 285 (1.671), respectively.

Overall, this study contributes the first molecular data of an HIV-1 *pol* sequence from Central Java, Indonesia, and thereby enriches the information of HIV-1 genetics in Indonesia. Our data are also beneficial for the analysis on the proteins expressed byour isolate. Further, these data are expected to be useful for antiretroviral and vaccine development.

ACKNOWLEDGMENT

This work was supported partially by grants from the Indonesian Directorate of Higher Education (No. 077/SP2H/PL/Dit.Litabmas/V/2013) and from APBN/DIPA UNS (No. 2342/UN27.16/PN/2012 and No. 267a/UN27.16/PN/2013).

REFERENCES

Baxter, J.D., J.M. Schapiro,C.A.B. Boucher, V.M. Kohlbrenner, D.B. Hall, J.R. Scherer, and D.L. Mayers. 2006. Genotypic changes in human immunodeficiency virus type 1 protease associated with reduced susceptibility and virologic response to the protease inhibitor tipranavir. *J Virol*, vol. 80:10794-10801.

Bruzzone, B., A. Ventura, F. Bisio, F.A. Mboungou, L.M. Miguel, F. Saladini, M. Zazzi, G. Icardi, A. De Maria, and C. Viscoli C. 2010. Impact of extensive HIV-1 variability on molecular diagnosis in the Congo basin.*J ClinVirol*, vol. 47:372-375.

Chen, C.P., and B. Rost. 2002. State-of-the-art in membrane protein prediction.*Appl Bioinformatics*, vol. 1:21-35.

Gordon, M., T. De Oliveira, K. Bishop, H.M. Coovadia, L. Madurai, S. Engelbrecht, E. Janse van Rensburg, A. Mosam, A. Smith, and S. Cassol. 2003. Molecular characteristics of human immunodeficiency virus type 1 subtype C viruses from KwaZulu-Natal, South Africa: implications for vaccine and antiretroviral control strategies. *J Virol*, vol. 77:2587-2599.

Hopp, T.P., and K.R. Woods. 1981. Prediction of protein antigenic determinants from amino acid sequences. *ProcNatlAcadSci USA*, vol.78:3824-3828.

Hughes, A.L., J.A. Green, J.M. Garbayo, and R.M. Roberts. 2000. Adaptive diversification within a large family of recently duplicated, placentally expressed genes.*ProcNatlAcadSci USA*, vol. 97:3319-3323.

Kemal, K.S., M. Reinis, B. Weiser, and H. Burger. 2009.Methods for Viral RNA Isolation and PCR Amplificationfor Sequencing of Near Full-Length HIV-1 Genomes. *In*: HIV Protocols. Humana Press. New York.

Kiwanuka, N., M. Robb, O. Laeyendecker, G. Kigozi, F. Wabwire-Mangen, F.E. Makumbi, F. Nalugoda, J. Kagaayi, M. Eller, L.A. Eller, D. Serwadda, N.K. Sewankambo, S.J. Reynolds, T.C. Quinn, R.H. Gray, M.J. Wawer, and C.C. Whalen. 2010. HIV-1 viral subtype differences in the rate of CD4+ T-cell decline among HIV seroincident antiretroviral naive persons in Rakai district, Uganda. *J Acquir Immune DeficSyndr*, vol. 54:180-184.

Merati, T.P., R.E. Ryan, T. Spelmen, D.N. Wirawan, I.M. Bakta, B. Otto, R.B. Oelrichs, and S.M. Crowe. 2012. CRF01_AE dominates the HIV-1 epidemic in Indonesia. *Sex Health*,vol. 9:414-421.

Nei, M., and T. Gojobori. 1986. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions.*MolBiolEvol*, vol. 3:418-426.

Sahbandar, I.N., K. Takahashi, Z. Djoerban, I. Firmansyah, S. Naganawa, K. Motomura, H. Sato, K. Kitamura, H.T. Pohan, and S. Sato. 2009. Current HIV type 1 molecular epidemiology profile and identification of unique recombinant forms in Jakarta, Indonesia.*AIDS Res Hum Retroviruses*,vol. 25:637-646.

Sire, J.M., M. Vray, M. Merzouk, J.C. Plantier, J. Pavie, S. Maylin, J. Timsit, C. Lascoux-Combe, J.M. Molina, F. Simon, and C. Delaugerre. 2011. Comparative RNA quantification of HIV-1 group M and non-M with the RoscheCobasAmpliPrep/Cobas TaqManHIV-1 v2/0 and Abbortt Real-Time HIV-1 PCR assays. *J Acquir Immune DeficSyndr*, vol. 56:239-243.

Tamura, K., D. Peterson, N. Peterson, G. Stecher, M. Nei, and S. Kumar. 2011. MEGA5: molecular evolutionary genetic analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol*,vol. 28:2731-2739.

Wayengera, M. 2011. Proviral HIV-genome-wide and pol-gene specific zinc finger nucleases: usability for targeted HIV gene therapy.*TheorBiol Med Model*, vol. 8:26.

Xia, X., W.H. Li. 1998. What amino acid properties affect protein evolution?*JMolEvol*, vol. 47:557-564.