



Construction of A Preliminary Three-Dimensional Structure *Simian betaretrovirus* Serotype-2 (SRV-2) Reverse Transcriptase Isolated from Indonesian Cynomolgus Monkey

Authors:

Uus Saepuloh*, Diah Iskandriati, Joko Pamungkas, Dedy Duryadi Solihin, Sela Septima Mariya and Dondin Sajuthi

*Correspondence: uussaepuloh@yahoo.com

DOI: <https://doi.org/10.21315/tlsr2020.31.3.4>

Highlights

- Gene encoding reverse transcriptase enzyme of *Simian betaretrovirus serotype-2* (RT SRV-2) infected Indonesian long-tailed monkey was isolated and molecular characterised.
- We predicted the 3D structure model of this RT SRV-2 enzyme using computational program.
- This 3D model provides domains function, enzymatic site active and hybrid DNA/ RNA-enzyme interaction.

Construction of A Preliminary Three-Dimensional Structure *Simian betaretrovirus* Serotype-2 (SRV-2) Reverse Transcriptase Isolated from Indonesian Cynomolgus Monkey

¹Uus Saepuloh*, ¹Diah Iskandriati, ^{1,3}Joko Pamungkas, ²Dedy Duryadi Solihin, ¹Sela Septima Mariya and ^{1,3}Dondin Sajuthi

¹Primate Research Centre, Bogor Agricultural University (PSSP LPPM IPB), Jalan Lodaya II/5 Bogor 16151, Indonesia

²Department of Biology, Faculty of Mathematics and Natural Sciences, Bogor Agricultural University, Kampus Darmaga, Bogor 16680, Indonesia

³Faculty of Veterinary Medicine, Bogor Agricultural University, Kampus Darmaga, Bogor 16680, Indonesia

Publication date: 15 October 2020

To cite this article: Uus Saepuloh, Diah Iskandriati, Joko Pamungkas, Dedy Duryadi Solihin, Sela Septima Mariya and Dondin Sajuthi. (2020). Construction of a preliminary three-dimensional structure *Simian betaretrovirus* serotype-2 (SRV-2) reverse transcriptase isolated from Indonesian cynomolgus monkey. *Tropical Life Sciences Research* 31(3): 47–61. <https://doi.org/10.21315/tlsr2020.31.3.4>

To link to this article: <https://doi.org/10.21315/tlsr2020.31.3.4>

Abstract: *Simian betaretrovirus* serotype-2 (SRV-2) is an important pathogenic agent in Asian macaques. It is a potential confounding variable in biomedical research. SRV-2 also provides a valuable viral model compared to other retroviruses which can be used for understanding many aspects of retroviral-host interactions and immunosuppression, infection mechanism, retroviral structure, antiretroviral and vaccine development. In this study, we isolated the gene encoding reverse transcriptase enzyme (RT) of SRV-2 that infected Indonesian cynomolgus monkey (Mf ET1006) and predicted the three dimensional structure model using the iterative threading assembly refinement (I-TASSER) computational programme. This SRV-2 RT Mf ET1006 consisted of 547 amino acids at nucleotide position 3284–4925 of whole genome SRV-2. The polymerase active site located in the finger/palm subdomain characterised by three conserved catalytic aspartates (Asp90, Asp165, Asp166), and has a highly conserved *YMDD* motif as Tyr163, Met164, Asp165 and Asp166. We estimated that this SRV-2 RT Mf ET1006 structure has the accuracy of template modelling score (TM-score 0.90 ± 0.06) and root mean square deviation (RMSD) $4.7 \pm 3.1\text{Å}$, indicating that this model can be trusted and the accuracy can be seen from the appearance of protein folding in tertiary structure. The superpositionings between SRV-2 RT Mf ET1006 and Human Immunodeficiency Virus-1 (HIV-1) RT were performed to predict the structural in details and to optimise the best fits for illustrations. This SRV-2 RT Mf ET1006 structure model has the highest homology to HIV-1 RT (2B6A.pdb) with estimated accuracy at TM-score 0.911, RMSD 1.85 Å, and coverage of 0.953. This preliminary study

*Corresponding author: uussaepuloh@yahoo.com

of SRV-2 RT Mf ET1006 structure modelling is intriguing and provide some information to explore the molecular characteristic and biochemical mechanism of this enzyme.

Keywords: Reverse Transcriptase, SRV-2 Indonesian Isolates, 3D Structure Model

INTRODUCTION

Retroviral reverse transcriptase (RT) is a multifunctional enzyme that catalyses the formation of a double-stranded DNA from the single stranded retroviral RNA genome. This complex process is called reverse transcription, a critical step in the life cycle of all retroviruses which is responsible for viral genome replication. The newly synthesized DNA is then integrated into the host genome by integrase, another retroviral enzyme (Telesnitsky & Goff 1997). The retroviral RT enzyme has multiple function, such as an RNA-dependent DNA polymerase, a DNA dependent DNA polymerase, DNA-directed RNA cleavage, strand transfer, and strand displacement synthesis (Herschhorn & Hizi 2010). The RT enzyme has two enzymatic activities, as a DNA polymerase and RNase H activities localised in two separate protein domains. The DNA polymerisation function is capable of using either DNA or RNA as a template and RNase H function serves to hydrolyse the RNA strand within an RNA/DNA hybrid. Both the polymerase and RNase H activities are essential for viral replication (Coté & Roth 2008; Telesnitsky & Goff 1997).

RTs from the different groups of retroviruses share similar functional catalytic activities but they significantly differ in particular parameters, such as structure and subunit composition, molecular weights, catalytic properties, biochemical and biophysical characteristics and sensitivity to different inhibitors (Herschhorn & Hizi 2010). The RT of type-1 Human Immunodeficiency Virus (HIV-1) is the most intensively studied enzymes in the last decades (Herschhorn & Hizi 2010). The HIV-1 RT is a major target for the development of antiretroviral drug therapy of HIV-1, the virus causing acquired immunodeficiency syndrome (AIDS), with over half of current FDA-approved therapeutics against HIV infection targeting this enzyme (Iliina *et al.* 2012; Sarafianos *et al.* 2009). As a major target for anti-HIV therapy, RT has been the subject of extensive research using crystal structure determinations and biochemical assays (Sarafianos *et al.* 2009). HIV-1 RT has been characterised structurally. The structures of complexes with dsDNA (Ding *et al.* 1998); dsDNA and the incoming nucleotide (Huang *et al.* 1998); and an RNA/DNA hybrid (Sarafianos *et al.* 2001) are available.

Simian betaretrovirus serotype-2 (SRV-2) is a causative agent of simian acquired immunodeficiency syndrome in Asian macaques it accounts for high morbidity and mortality in research and breeding facilities (Marx *et al.* 1984; Gardner *et al.* 1988; Lerche 2010; Montiel 2010; Lerche & Osborn 2003). The SRV-2 genome contains four genes organised in the order 5'-*gag-prt-pol-*

env-3' and encode identically in size of 8105-bp proviruses (Marracci *et al.* 1995; Marracci *et al.* 1999). Similar to other retroviruses, SRV-2 can transcribe its own RNA genome into a double-stranded DNA by the action of a reverse transcriptase enzyme. The resulted double-stranded DNA is inserted randomly into the chromosome of the host cell and can express the genetic information of the retrovirus (Marracci *et al.* 1999). The presence of gene encoding RT enzyme in SRV-2 genome may potentially be studied and utilised further as RT enzyme model for developing an anti RT drug of other retroviruses especially HIV/AIDS.

In this study, we have amplified the gene encoding the RT enzyme of the SRV-2 isolated from an infected Indonesian cynomolgus monkey and predicted the three dimensional (3D) structure model of this SRV-2 RT enzyme using I-TASSER (Iterative Threading ASSEmbly Refinement) computational program. I-TASSER is a computational method that has been successful in accurately modelling protein structures (Zhang 2008; Roy *et al.* 2010; Yang *et al.* 2015). I-TASSER uses a combinatorial approach, employing all three conventional methods for structure modelling: comparative modelling, threading, and *ab initio* modelling (Roy *et al.* 2010).

MATERIALS AND METHODS

PCR Amplification and Sequencing of SRV-2 Reverse Transcriptase Gene

The DNA from archived peripheral blood mononuclear cells (PBMCs) of Indonesian cynomolgus monkey (the ID was ET1006) that positive to SRV-2 was extracted using QiaAmp DNA blood minikit (Qiagen, Hilden, Germany). PCR amplification of 2.5 μ L DNA was carried out with 2U Pfx DNA polymerase (Thermo Fisher Scientific, Carlsbad, USA), in 25 μ L total reaction containing 2 mM MgCl₂, 1x PCR buffer, 1 mM dNTPs (Thermo Fisher Scientific, Carlsbad, USA). We developed the specific primer using Primer3 program (Untergasser *et al.* 2012) based on SRV-2 complete genome (AF126467.1) as SRV-2 RT 3284F (5'-CCTGTTTGGGTTGATC-3') and SRV-2 RT 4925R (5'-GTGATTACCTTGAGATAAAGGTCC-3'). PCR amplification was performed for 35 cycles in the following condition: denaturation at 94°C for 30 s, annealing at 55°C for 30 s, and extension at 68°C for 1.5 min. Amplicon (12.5 μ L) were visualised on 1.0% TAE agarose gel and analysed with GelDoc machine (BioRad, California, USA). PCR product was gel purified using Qiaquick gel extraction kit (Qiagen, Hilden, Germany) then sequenced in First Base Laboratories Sdn Bhd (Malaysia). The primer used for sequencing was similar to that used for PCR amplification and some internal specific primers: SRV-2 RT 3751R (5'-GCCCAAGATTTTCTGATTGG-3'), SRV-2 RT 3799F (5'-GATTGGCGAA CAAGTTTTGC-3'), SRV-2 RT 4247F (5'-TACTGGCCTCTTCTGG -3'), and SRV-2 RT 4604F (5'-GGCATAGCCGCATACACTTT-3').

Phylogenetic Tree Analysis

The obtained nucleotide sequences were analysed using BioEdit program (Hall 1999) and alignment was performed with computer software BLAST program (Altschul *et al.* 1997) (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>). Nucleotide encoded amino acid translation was performed using EMBOSS Transeq Program (Madeira *et al.* 2019) (<http://www.ebi.ac.uk/Tools/emboss/transeq>). Amino acid sequences were aligned using the ClustalW2 (Madeira *et al.* 2019) (<https://www.ebi.ac.uk/Tools/msa/clustalw2/>) then adjusted to phylogenetic tree construction using MEGA 6.0 program based on Neighbour-Joining method (Tamura *et al.* 2013). Genetic distances were estimated using Kimura's two parameter method (Kimura 1980) and the bootstrap analysis was performed 500 replicates to assign confidence to tree nodes. RT amino acid sequences of others retroviruses were used as references: Avian Myeloblastosis Virus (AMV, AAB31929.1), Baboon Endogenous Virus (BaEV, BAA89659.1), Feline Leukemia Virus (FeLV, BAL04139.1), Feline Immunodeficiency Virus (FIV, ABO69501.1), Human Immunodeficiency Virus-1 (HIV-1, ADN28092.1), HIV-2 (AAB01352.1), Moloney Murine Leukemia Virus (MMLV, NP_955591.1), Mason-Pfizer Monkey Virus (MPMV, AAA47711.1), Simian Foamy Virus (SPF, YP_001961122.1), Simian Endogenous Retrovirus (SERV, AAC97565.1), Simian Immunodeficiency Virus (SIV, AAA74707.1), Simian Retrovirus-1 (SRV-1, AAA47732.1), SRV-2 (AAD43256.1), SRV-2 (AAA47562.1), SRV-4 (YP_003864102.1), SRV-5 (BAM71050), SRV-8 (AOS48113.1), Human T-Lymphotropic Virus-2 (HTLV-2, AAB59885.1), Simian T-Lymphotropic Virus-1 (STLV-1NP_049560.1) and Mouse Mammary Tumour Virus (MMTV, BAA03767).

Functional Structure and Three Dimensional Model of SRV-2 RT Enzyme

Amino acid sequences were aligned using the ClustalW2 and the secondary-structure predictions were made based on sequences and structural motif using I-TASSER (<https://zhanglab.ccmb.med.umich.edu/I-TASSER/>). The conserve domain and amino acids position of active site were refined manually based on predicted secondary structure and multiple sequences alignment as previously defined of HIV-1 (1HYS A) (Sarafianos *et al.* 2001) and SRV-2 *pol* AAD43256 (Marracci *et al.* 1999). The I-TASSER program was used to determine the function of the query protein inferred by structurally matching the predicted 3D models against the proteins of known structure and function in the Protein Data Bank (PDB). The functional analogs from the global search results are ranked based on the conserved structural patterns present in the model, measured using a scoring scheme that combines template modelling (TM-score), root-mean-squared deviation (RMSD), sequence identity, and coverage of the structure alignment. TM-score was defined to assess the topological similarity of protein structure pairs with a value in the range of (0, 1), a higher score indicating better structural match. The results from the global and local search are combined to present a

comprehensive list of functional analogs (Zhang 2008; Roy *et al.* 2010). Matching structures were analysed using PyMol program (Delano 2002) (<http://www.pymol.org>) that allowed us to further evaluate possible structural and functional domains of the SRV-2 RT protein.

RESULTS AND DISCUSSIONS

PCR Amplification and Sequencing of SRV-2 Reverse Transcriptase Gene

Recently, we have isolated and characterised the SRV-2 from wild Indonesian cynomolgus monkeys with sequences analysis result of this virus envelope region revealed high homology (97%–98% similarities) to SRV-2 reported previously in captive macaques in US Primate Research Centre (Iskandriati *et al.* 2010). In this research, we analysed the RT encoding gene from Indonesian SRV-2 (SRV-2 RT Mf ET1006) since there has been no report focusing on SRV-2 RT study. Therefore, this preliminary research may provide a better understanding of structure and mechanism of retroviral SRV-2 RT. The DNA amplification to SRV-2 RT Mf ET1006 resulted in 1641 bp amplicon (Fig. 1) at the position of 3284–4925 of whole provirus SRV-2 genome consisting of 547 amino acids as enzymatic RT. These SRV-2 RT Mf ET1006 amino acids have 98%–99% similarities to others SRV-2 previously isolated from captive macaques in some primate research centre in US (AF126467, AF126468, and M16605) and have 79%–87% similarities to others SRVs (SRV 1, 3, 4, 5, 8 and SERV) (Table 1). It seemed that the amino acid sequences of RT enzyme among SRVs are conserved.

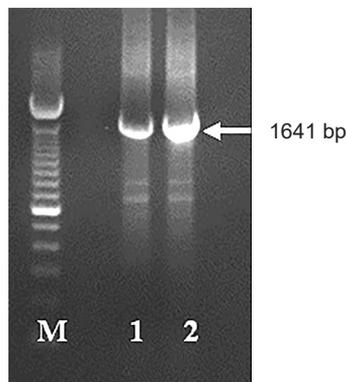


Figure 1: PCR amplification to SRV-2 RT gene isolated from Indonesian cynomolgus monkeys (SRV-2 RT Mf ET1006) resulting 1641 bp amplicon. (M) 100 bp DNA ladder (Invitrogen), (1) SRV-2 RT Mf ET1006, (2) SRV-2/A549 positive control (obtained from National Primate Research Centre, University of Washington, USA).

Table 1: Amino acid identities of SRV-2 RT Mf ET1006 compared with others SRVs.

		1	2	3	4	5	6	7	8	9	10
1	SRV2 RT Mf ET1006	–	99	99	98	86	86	83	82	84	82
2	SRV2 AF126467		–	99	97	85	85	83	82	83	81
3	SRV2 AF126468			–	97	85	85	83	82	83	81
4	SRV2 M16605				–	85	85	82	81	84	82
5	SRV1					–	96	85	84	86	79
6	SRV3						–	85	84	85	79
7	SRV4							–	82	83	77
8	SRV5								–	81	77
9	SRV8									–	79
10	SERV										–

Phylogenetic Tree Analysis of SRV-2 RT

We have multiple-aligned our amino acid sequences and constructed phylogenetic tree to determine genetic relationship among RT enzyme family retroviruses (Fig. 2). The phylogenetic result showed that the target sequences of SRV-2 RT Indonesian isolate (SRV-2 RT Mf ET1006) have the closes relationship with SRV-2 D2/RHE/OR (AF126467 and AF126468) first isolated from rhesus monkey at Oregon Regional Primate Research Centre US in 1999 (Marracci *et al.* 1999). SRV-2 RT Mf ET1006 was clustered with endogenous SERV, SRV-1, SRV-3, SRV-4, SRV-5, SRV-8 and MMTV as *betaretrovirus* subfamily. This subfamily has the closest relationship with *alpharetrovirus* (AMV), then both of this cluster branched with the *deltaretrovirus* subfamily (HTLV-1). *Alpha*, *beta* and *deltaretrovirus* subfamily then branched with *lentivirus* where HIV-1 as a member of this subfamily is the most investigated retrovirus in biomedical field. The furthest relation of SRV cluster was *spumavirus* and *gammaretrovirus* where the MMLV as a member of this subfamily has widely applied RT enzyme in RT-PCR technique. This data has the similar correlation with previously reported research by Li *et al.* (1995) on *pol* retrovirus phylogenetic from 55 retroelements where SRV-1 is closely related to MPMV (SRV-3) and evolved after SRV-2. Therefore, the highly conserved *pol* gene region has been the target in phylogenetic studies both in nucleotide and amino acids sequences (Li *et al.* 1995).

Functional Structure Prediction of SRV-2 RT Enzyme

In this study, we tried to compare the functional structure of SRV-2 RT Mf ET1006 with HIV-1 RT. The comparison was based on the HIV-1 RT crystal structure in PDB (2B6A and 1HYS) and SRV-2 *pol* previously reported in GenBank (AAD43256.1). Amino acid sequences of SRV-2 RT Mf ET1006 at the position of 1-547 and HIV-1

RT from 1-553 were aligned using ClustalW, STRAP2 (Gille & Frömmel 2001) and I-TASSER (Fig. 3). Comparison between HIV-1 RT p66 and SRV-2 RT Mf ET1006 showed that finger/palm subdomain has 32% similarity of 237 residues, the thumb subdomain has 25.6% similarity of 82 residues, the connection subdomain has 12% similarity of 133 residues, and the RNase H subdomain has 27% similarity of 118 residues. Overall, the full-length comparison between SRV-2 RT Mf ET1006 and HIV-1 RT has 25.5% of 553 residues. Meanwhile, if the very conservative amino acid (such as Ala vs. Val, Leu vs. Ile, etc.) was allowed for comparison, it resulted in 43% similarities.

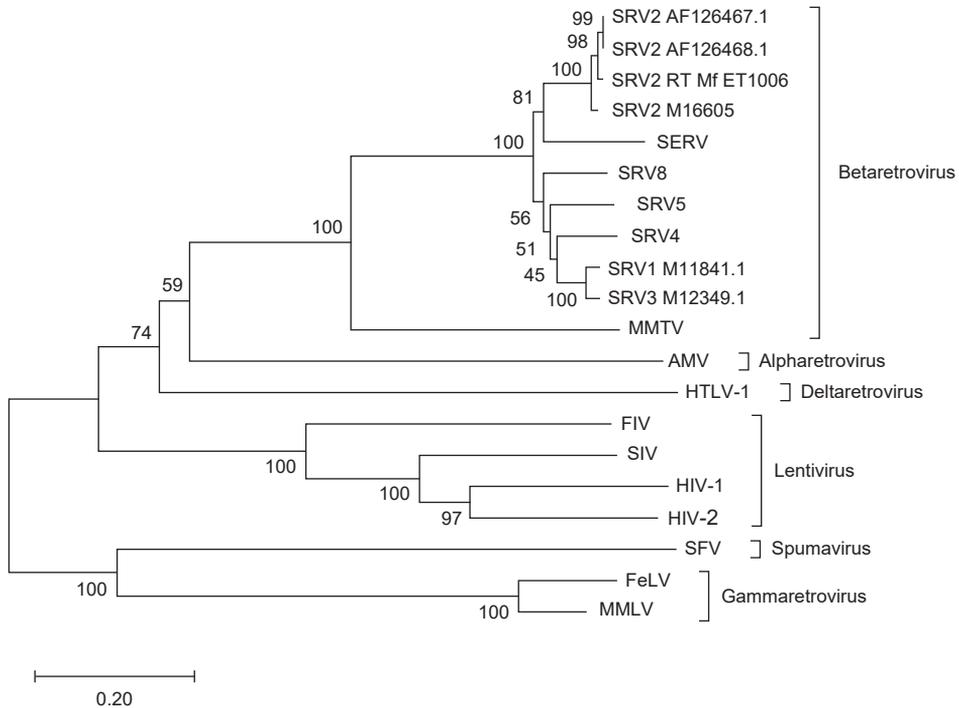


Figure 2: Phylogenetic tree of SRV-2 RT Mf ET1006 amino acid sequences relatedness to others retroviruses constructed using the neighbor-joining method. The bootstrap percentages of 500 replicate are shown at the branch nodes. The evolutionary distances are estimated by branch lengths of the tree indicated and drawn to scale, which represents the number of base substitutions per site.

The polymerase active site that is located in the finger/palm subdomain is characterised by three conserved catalytic aspartates at the position of Asp90, Asp165, Asp166 for SRV-2 RT Mf ET1006, and Asp110, Asp185, Asp186 for HIV-1 RT (Table 2). This triad aspartates (the last two residues are part of Y χ DD motif), play a significant role in the HIV-1 RT enzymatic catalytic through the binding to the divalent cation Mg²⁺ (Sharma *et al.* 2005). This leads to the

assumption that the catalytic active site of SRV-2 RT Mf ET1006 coordinate with two divalent ions, Mg²⁺. The YXDD motif, where X is a variable amino acid, is highly conserved and plays a critical role among various viral RNA dependent DNA polymerases. The conserved motif of SRV-2 RT Mf ET1006 and HIV-1 RT consists of four amino acids, YMDD for Tyr, Met, Asp, Asp at the position 163-166 for SRV-2 RT Mf ET1006 and 183-186 for HIV-1 RT. SRV-2 RT Mf ET1006 has a YMDD amino acid motif similar with HIV-1, SIV and FIV, however, it was different from YVDD motif that MMLV and FeLV have (Sharma *et al.* 2005). Recently, many studies focused on this YXDD motif due to its significance in NRTI resistance, RT fidelity, polymerase process, and viral replication (Sharma *et al.* 2005). The dNTPs binding sites is also located in the polymerase active site and have conserve residues which are Lys64, Arg72, Asp110, Asp113, Gln151, and Asp186 for HIV-1 RT. Meanwhile Lys45, Arg52, Asp90, Asp93, Gln131, and Asp166 for SRV-2 RT Mf ET1006. The RNase H active sites that interacted with DNA/RNA hybrid contain conserved DED motif such as Asp, Glu, and Asp at the position of 443, 478, and 498 for RNase H of HIV-1 RT and 436, 465, and 486 for RNase H of SRV-2 RT (Table 2 and Fig. 3).

Table 2: Amino acid comparison of selected structural motifs of SRV-2 RT Mf ET1006 compared to HIV-1 RT.

Domain/subdomain motif	SRV-2 RT Mf ET1006 amino acid ^a	HIV-1 RT amino acid ^b
Finger/palm	1–216	1–237
Thumb	217–297	238–316
Connection	298–430	317–437
RNase H	431–547	438–553
Polymerase active site	Trp7, Pro8, Phe42, Lys46, Arg52, Leu54, Gln55, Asp56, Arg58, Asn61, Ala69, Gln71, Pro72, Leu74, Asp90, Leu91, Lys92, Asp93, Cys94, Phe95, Gln131, Gly132, Ala134, Pro137, Tyr163, Asp165, Asp166, Leu209, Gly210	Trp24, Pro25, Phe61, Lys65, Arg72, Leu74, Val75, Asp76, Arg78, Asn81, Glu89, Gln91, Leu92, Ile94, Asp110, Val111, Gly112, Asp113, Ala114, Tyr115, Gln151, Gly152, Lys154, Pro157, Tyr183, Asp185, Asp186, Met230, Gly231
RNA/DNA hybrid	Ser438, Ser439, Thr440, Ala461, Ala488, His531, Arg533, Ala534	Ala446, Arg448, Glu449, Thr473, Asn474, Gln475, Gln500, His539
RNase active site	Asp436, Glu465, Asp486	Asp443, Glu478, Asp498
dNTPs binding site	Lys45, Arg52, Asp90, Leu91, Lys92, Asp93, Cys94, Phe95, Gln131, Asp166	Lys64, Arg72, Asp110, Val111, Gly112, Asp113, Ala114, Tyr115, Gln151, Asp186
YXDD	Tyr163, Met164, Asp165, Asp166	Tyr183, Met184, Asp185, Asp186

Notes: ^aSRV-2 RT Mf ET1006 was aligned based on I-TASSER result and referred to SRV-2 *pol* (GenBank AAD43256.1)

^bHIV-1 RT based on PDB of 2B6A, 1HYS and referred to HIV-1 *pol* (GenBank ADN28092.1)

Three Dimensional Structure Model of SRV-2 RT Mf ET1006

Three-dimensional structure prediction of the full-length SRV-2 RT Mf ET1006 is depicted in Fig. 4(A). The models were generated using I-TASSER method and visualised using Pymol program. Five models of SRV-2 RT Mf ET1006 were computationally generated based on pair-wise structure similarity and the Model I structure gave a highest C-score at 1.33 compared to Models 2 to 5 with C-score less than -0.1 . The C-score is a confidence score which ranges from -5 to 2 , with higher scores representing higher confidence in the model. The C-score has a strong correlation with the quality of the I-TASSER models (Roy *et al.* 2010). In general, models with C-score > -1.5 are expected to have a correct fold. The RMSD and TM-score are both well-known measures of topological similarity between the model and native structure. TM-score values range from 0 to 1 , where a higher score indicates a better structure match (Roy *et al.* 2010). The model I structure of SRV-1 RT Mf ET1006 has the estimated accuracy at TM-score 0.90 ± 0.06 with RMSD $4.7 \pm 3.1\text{\AA}$. Therefore, the Model I SRV-2 RT Mf ET1006 structure has the best C-score. So, it can be trusted, and the accuracy can be seen from the appearance of protein folding in tertiary structure (Fig. 4). The superpositionings between HIV-1 RT and SRV-2 RT Mf ET1006 were performed to predict the structural details and to optimise best fits for illustrations. Table 2, Fig. 3 and Fig. 4(B) showed the structural motifs used in the superposition comparisons between SRV-2 RT Mf ET1006 and HIV-1 RT. The HIV-1 RT structures possessing the most consistent and well-defined motifs for our study were 1HYS (Sarafianos *et al.* 2001), 2HMI (Ding *et al.* 1998) and 2B6A (Roth *et al.* 1997). This SRV-2 RT Mf ET1006 structure model has the highest homology to HIV-1 RT (2B6A) with estimated accuracy at TM-score 0.911 , RMSD 1.85 and coverage of 0.953 measured with TM-align from Zhang Lab (Zhang & Skolnick 2005).

The models of the SRV-2 RT Mf ET1006 in complex with 18/19-mer DNA/RNA nucleic acid substrates stretching from the polymerase active site to the RNase H active site are depicted in Fig. 5. The SRV-2 RT Mf ET1006 was superpositioning to HIV-1 RT (1N6Q) (Sarafianos *et al.* 2002) based on I-TASSER template proteins with similar binding site. The assessment of this model was carried out using TM-align server (<http://zhanglab.ccmb.med.umich.edu/TM-align>) with TM-score 0.838 , Identity 0.258 and BS-score 1.48 . BS-score is a measure of local similarity (sequence and structure) between template binding site and predicted binding site in the query structure. Based on large scale benchmarking analysis, BS-score > 1.1 reflects a very good local match between the predicted and template binding site (Roy *et al.* 2010).

The model of interaction between SRV-2 RT Mf ET1006 site actives with nucleic acids was developed based on the available crystal structure HIV-1 RT (Huang *et al.* 1998) (Fig. 6). The amino acids involved in interaction with incoming dNTP of HIV-1 RT such as Lys65, Arg72, Asp113, Ala114, Tyr115 and Gln151 (Huang *et al.* 1998), have a conserved amino acid with SRV-2 RT Mf ET1006 as Lys45, Arg52, Asp93, Gln131. Ala114 and Tyr115 in HIV-1 RT were replaced with Cys94 and Phe95, in SRV-2 RT Mf ET1006. The substitution of Tyr115 with Phe

caused a little deleterious effect in enzyme catalytic activities that interact directly with deoxyribose of the incoming triphosphate (Boyer *et al.* 2000). We predicted the first nucleotide overhang of the template that will pair with incoming dNTP and interact with side chain of Leu54 and against the backbone of Gly132 (Figs. 6(A) and (B)). The second and third 5'-template overhang will pack against the residues of Trp7, Pro8 and Phe42. Direct interaction of bases in minor groove occur with Pro137, Met164, Leu74 and Tyr163. The incoming dNTP pairs with the templating bases in which the triphosphate residues will interact with Lys45, Arg52, Asp93 and Cys94. Meanwhile, the 3'-OH of dNTP will line by side chain of Asp93, Phe95, Gln131 and the peptide backbone between 93 and 95.

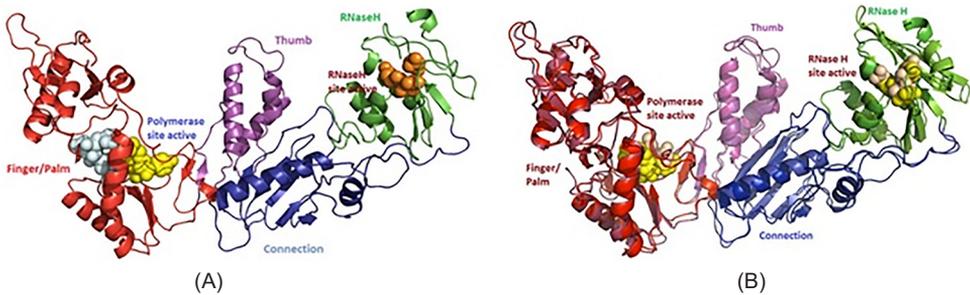


Figure 4: (A) Three-dimensional structure model of the full-length SRV-2 RT Mf ET1006. The models were generated using I-TASSER method and visualised using Pymol program. The estimated accuracy of model is 0.90 ± 0.06 (TM-score), $4.7 \pm 3.1\text{\AA}$ (RMSD) and C-score = 1.33. The fingers/palm, thumb, connection and RNaseH subdomains are coloured in dark red, purple, dark blue and green, respectively. Light blue spheres represent the residues of the polymerase active site and dNTPs binding site. Yellow spheres represent the YXDD motif in the finger/palm. Meanwhile, the brown spheres indicated the DED motif of RNaseH active site. (B) Superpositioning model between SRV-2 RT Mf ET1006. with HIV-1 RT (2B6A.pdb), estimated accuracy is 0.911 (TM-score), 1.85 Å (RMSD), and 0.953 (coverage).

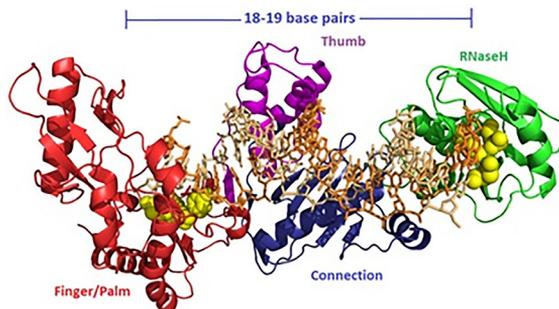


Figure 5: Ribbon diagram of the SRV-2 RT Mf ET1006 bound to a stick model of an 18–19-base RNA/DNA duplex oligonucleotide substrate. The various domains of the enzyme were colour-coded as presented in Fig. 4. The active site amino acids of the DNA polymerase and RNase H are shown in yellow spheres.

In this study, a predicted three-dimensional structure of SRV-2 RT Mf ET1006 enzyme should be very informative and useful for understanding its structure and functional characteristics, since no structure is currently available for this enzyme. This is because the structure of a protein provides precise molecular details that often facilitate experimental characterisation of an expected function. The development and validation of computational methods that can predict protein structure to a relatively high level of accuracy and thereby facilitate functional annotation, biochemical analyses, and biological characterisation are a high priority. Our model provides an initial structural framework for understanding the SRV-2 RT Mf ET1006 domains function, enzymatic site active and hybrid DNA/RNA-enzyme interaction.

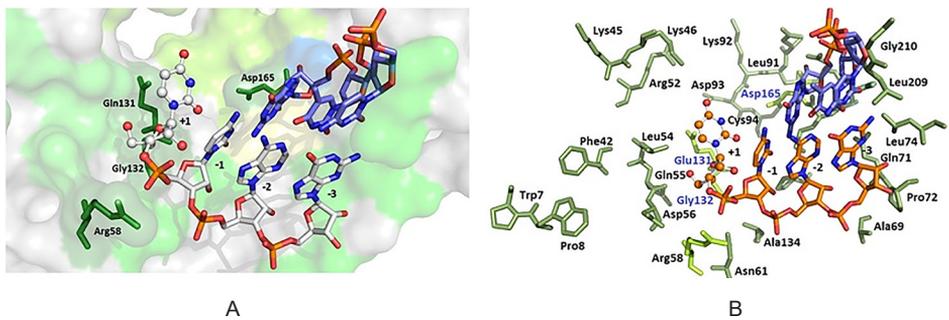


Figure 6: Interaction of amino acids that play roles in polymerase site active of SRV-2 RT Mf ET1006 with terminal portion of the RNA/DNA duplex. DNA was indicated with red and blue stick, while RNA in red, white and blue stick, signed with -1 , -2 and -3 as duplex RNA/DNA; overhang RNA was indicated by $+1$ as round edges stick. (A) SRV-2 RT Mf ET1006 site active is represented in green surface with conserved amino acids Asp165, Gln131, Gly132 and Arg58 (shown in green stick); (B) SRV-2 RT Mf ET1006 amino acids site active represented in green stick.

CONCLUSION

The preliminary study of three-dimensional structure modelling of monomeric SRV-2 RT Mf ET1006 compared to heterodimeric HIV-1 RT is intriguing, given some information about the similarities and differences of amino acid sequences between them to predict the same function of each subdomains. Using the structural model studies to explore the molecular characteristic and biochemical mechanism of SRV-2 RT Mf ET1006 and to understand the mechanism of enzymatic catalytic would provide valuable information. Recently, many RT structures have been explored and reported in PDB Data Bank. Therefore, these available RT structures have contributed and inspired us to explore the SRV-2 RT Mf ET1006 structure based on three-dimensional structure models. This will intrigue us to expand our knowledge about SRV-2 RT Mf ET1006 as a virus model of HIV and develop new improved therapeutics agents.

ACKNOWLEDGEMENTS

This study partially supported by National Innovation Research Incentives Programme 2016 (INSINAS) (Grant Number RT-2015-0612), Ministry of Research Technology and Higher Education Republic of Indonesia and Productive-Innovative Research Programme of Indonesian Endowment Fund for Education 2019 (RISPRO LPDP) (Grant Number PRJ-23/LPDP/2018), Ministry of Finance, Republic of Indonesia.

REFERENCES

- Altschul S F, Madden T L, Schaffer A A, Zhang J, Zhang Z, Miller W and Lipman D J. (1997). Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Research* 25(17): 3389–3402. <https://doi.org/10.1093/nar/25.17.3389>
- Boyer P L, Sarafianos S G, Arnold E and Hughes S H. (2000). Analysis of mutations at position 115 and 116 in the dNTPs binding site of HIV-1 reverse transcriptase. *PNAS USA* 97(7): 3056–3061. <https://doi.org/10.1073/pnas.97.7.3056>
- Coté M L and Roth M J. (2008). Murine leukemia virus reverse transcriptase: Structural comparison with HIV-1 reverse transcriptase. *Virus Research* 134(1–2): 186–202. <https://doi.org/10.1016/j.virusres.2008.01.001>
- Delano W L. (2002). PyMOL: An open-source molecular graphics tool. *CCP4 Newsletter On Protein Crystallography* 40: 82–92.
- Ding J, Das K, Hsiou Y, Sarafianos SG, Clark A D Jr, Jacobo-Molina A, Tantillo C, Hughes S H and Arnold E. (1998). Structure and functional implications of the polymerase active site region in a complex of HIV-1 RT with a double-stranded DNA template-primer and an antibody Fab fragment at 2.8 Å resolution. *Journal of Molecular Biology* 284(4): 1095–1111. <https://doi.org/10.1006/jmbi.1998.2208>
- Gardner M B, Luciw P, Lerche N and Marx P. (1988). Nonhuman primate retrovirus isolates and AIDS. *Advances Veterinary Sciences Comparative Medicine* 32: 171–190. <https://doi.org/10.1016/B978-0-12-039232-2.50011-6>
- Gille C and Frömmel C. (2001). STRAP: Editor for STRuctural Alignments of Proteins. *Bioinformatics* 17(4): 377–378. <https://doi.org/10.1093/bioinformatics/17.4.377>
- Hall T A. (1999). BioEdit: A user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids. Symposium Series* 41: 95–98.
- Herschhorn A and Hizi A. (2010). Retroviral reverse transcriptase. *Cell Molecular Life Science* 67: 2717–2747. <https://doi.org/10.1007/s00018-010-0346-2>
- Huang H, Chopra R, Verdine G L and Harrison S C. (1998). Structure of a covalently trapped catalytic complex of HIV-1 reverse transcriptase: Implications for drug resistance. *Science* 282(5394): 1669–1675. <https://doi.org/10.1126/science.282.5394.1669>
- Illina T, LaBarge K, Sarafianos S G, Ishima R and Parniak M A. (2012). Inhibitors of HIV-1 reverse transcriptase-associated Ribonuclease H activity. *Biology* 1: 521–541. <https://doi.org/10.3390/biology1030521>
- Iskandriati D, Saepuloh U, Mariya S, Grant R F, Solihin D D, Sajuthi D and Pamungkas J. (2010). Isolation and characterization of *simian retrovirus* type D from *Macaca fascicularis* and *M. nemestrina* in Indonesia. *Microbiologi Indonesia* 4(3): 132–136.

- Kimura M. (1980). A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *Journal of Molecular Evolution* 16: 111–120. <https://doi.org/10.1007/BF01731581>
- Lerche N W and Osborn K G. (2003). Simian retrovirus infections: Potential confounding variables in primate toxicology studies. *Toxicology Pathology* 31(Suppl.): 103–111. <https://doi.org/10.1080/01926230390174977>
- Lerche N W. (2010). Simian retroviruses: Infection and disease-implications for immunotoxicology research in primates. *Journal of Immunotoxicology* 7(2): 93–101. <https://doi.org/10.3109/15476911003657406>
- Li M D, Bronson D L, Lemke T D and Faras A J. (1995). Phylogenetic analysis of 55 retroelements on the basis of the nucleotide and product amino acid sequences of *pol* gene. *Molecular Biology Evolution* 12(4): 657–670.
- Madeira F, Park Y M, Lee J, Buso N, Gur T, Madhusoodanan N, Basutkar P, Tivey A R N, Potter S C, Finn R D and Lopez R. (2019). The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic Acids Research* 47: W636–W641. <https://doi.org/10.1093/nar/gkz268>
- Marracci G H, Kelley R D, Pilcher K Y, Crabtree L, Shiigi S M, Avery N, Leo G, Webb M C, Hallick L M and Axthelm M K. (1995). Simian AIDS type D serogroup 2 retrovirus: Isolation of an infectious molecular clone and sequence analyses of its envelope glycoprotein gene and 3' long terminal repeat. *Journal of Virology* 69(4): 2621–2628. <https://doi.org/10.1128/JVI.69.4.2621-2628.1995>
- Marracci G H, Avery N A, Shiigi S M, Couch G, Palmer H, Pilcher K Y, Nichols H, Hallick L M, Axthelm M K and Machida C A. (1999). Molecular cloning and cell-specific growth characterization of polymorphic variants of type D serogroup 2 simian retroviruses. *Virology* 261(1): 43–58. <https://doi.org/10.1006/viro.1999.9858>
- Marx P A, Maul D H and Osborne K G. (1984). Simian AIDS: Isolation of type D retrovirus and disease transmission. *Science* 223(4640): 1083–1086. <https://doi.org/10.1126/science.6695196>
- Montiel N A. (2010). An updated review of simian betaretrovirus (SRV) in macaque hosts. *Journal of Medical Primatology* 39(5): 303–314. <https://doi.org/10.1111/j.1600-0684.2010.00412.x>
- Roth T, Morningstar M L, Boyer P L, Hughes S H, Buckheit R W and Michejda C J. (1997). Synthesis and biological activity of novel non-nucleoside inhibitors of HIV-1 reverse transcriptase 2-Aryl-substituted benzimidazoles. *Journal of Medical Chemistry* 40(26): 4199–4207. <https://doi.org/10.1021/jm970096g>
- Roy A, Kucukural A and Zhang Y. (2010). I-TASSER: A unified platform for automated protein structure and function prediction. *Nature Protocol* 5: 725–738. <https://doi.org/10.1038/nprot.2010.5>
- Sarafianos S G, Das K, Tantillo C, Arthur D, Clark J, Jianping D, Whitcomb J M, Boyer P L, Hughes S H and Arnold E. (2001). Crystal structure of HIV-1 reverse transcriptase in complex with a polypurine tract RNA: DNA. *EMBO Journal* 20(6): 1449–1461. <https://doi.org/10.1093/emboj/20.6.1449>
- Sarafianos S G, Clark A D Jr, Das K, Tuske S, Birktoft J J, Ilankumaran P, Ramesha A R et al. (2002). Structures of HIV-1 reverse transcriptase with pre- and post-translocation AZTMP-terminated DNA. *EMBO Journal* 21(23): 6614–6624. <https://doi.org/10.1093/emboj/cdf637>
- Sarafianos S G, Marchand B, Das K, Himmel D, Parniak M A, Hughes S H and Arnold E. (2009). Structure and function of HIV-1 reverse transcriptase: Molecular mechanisms of polymerization and inhibition. *Journal of Molecular Biology* 385(3): 693–713. <https://doi.org/10.1016/j.jmb.2008.10.071>

- Sharma P L, Nurpeisov V and Schinazi R F. (2005). Retrovirus reverse transcriptase containing a modified YXDD motif. *Antiviral Chemical Chemotherapy* 16(3): 169–182. <https://doi.org/10.1177/095632020501600303>
- Tamura K, Stecher G, Peterson D, Filipski A and Kumar S. (2013). MEGA6: Molecular evolutionary genetics analysis version 6.0. *Molecular Biology and Evolution* 30(12): 2725–2729. <https://doi.org/10.1093/molbev/mst197>
- Telesnitsky A and Goff S P. (1997). Reverse transcriptase and the generation of retroviral DNA. In: Coffin J M, Hughes S H and Varmus H E (eds.). *Retroviruses*. Plainview, NY: Cold Spring Harbor Laboratory Press, 121–160.
- Untergasser A, Cutcutache I, Koressaar T, Ye J, Faircloth B C, Remm M and Rozen S G. (2012). Primer3-new capabilities and interfaces. *Nucleic Acids Research* 40(15): e115. <https://doi.org/10.1093/nar/gks596>
- Yang J, Yan R, Roy A, Xu D, Poisson J and Zhang Y. (2015). The I-TASSER suite: Protein structure and function prediction. *Nature Methods* 12(1): 7–8. <https://doi.org/10.1038/nmeth.3213>
- Zhang Y and Skolnick J. (2005). TM-alignment: A protein structure alignment algorithm based on the TM-score. *Nucleic Acid Research* 33(7): 2302–2309. <https://doi.org/10.1093/nar/gki524>
- Zhang Y. (2008). I-TASSER server for protein 3D structure prediction. *BMC Bioinformatics* 9(40). <https://doi.org/10.1186/1471-2105-9-40>