European Journal of Biomedical Research
Vol 3 | Issue 4 | July 2024
ISSN 2796-0056

RESEARCH ARTICLE

Bioinformatics Techniques for Developing Molecular Detection Methods for the HIV-1 Gag Gene

Asryadin[1], Nilasari Indah Yuniati[2,*], Nur Aini Hidayah Khasanah[2], Adhi Aqwam[1], Rizka Khairunnisa[1], Hetti Koes Endang[1], Jumratul Nurhidayah[1], Daniel Djoko Wahyono[3], and Alice Yuniaty[3]

ABSTRACT

The HIV-1 Gag gene, which plays an essential role in HIV replication, can be detected accurately using qRT-PCR. The quality of qRT-PCR analysis is determined by the primers and probes used for DNA amplification. This research aims to use bioinformatics techniques to design primer pair sequences and qRT-PCR probes for HIV detection using the HIV-1 Gag gene. HIV-1 Gag gene sequences were obtained from HIV-1 isolates and serotypes, downloaded from the National Center for Biotechnology Information (NCBI) GenPeptd nucleotide database. Sequences were then examined using the ClustalW algorithm of the Bioedit sequence alignment editor version 7.2.5.0. through gene alignment using multiple sequence alignment (MSA) with conserved regions. The primer pair sequences of the Gag-HIV 1 gene were obtained, namely, forward 5′-CAGTACAATGTGCTTCCACAGGG-3 and reverse 3′-CGGGATAGAGATTCAGTCTAGG-5′ with the probe sequence 5′-GGATCACCAGCAATATTTCAGGGAACG-3′. The primer sequence has a length of 23 bases (forward), 22 bases (reverse), GC content of 52% (reverse), 50% (forward), and the same forward and reverse melting temperature (Tm) of 66°C. The probe sequence is 27 bases long, with a GC content of 48% and a Tm of 67.3°C. No hairpin loops and dimers were formed in the primer pair or probe, and the gag gene had 100% homology with HIV-1. It was concluded that the primer and probe pair sequences met the requirements and could be used to amplify the HIV-1 Gag gene using qRT-PCR.

Submitted: May 24, 2024

Published: July 24, 2024

10.24018/ejbiomed.2024.3.4.97

[1] *Regional Research and Innovation Agency, Indonesia.*
[2] *Bina Cipta Husada Institute of Health Science, Indonesia.*
[3] *Biology Faculty of Jenderal Soedirman University, Indonesia.*

****Corresponding Author:***
e-mail: nila@stikesbch.ac.id

Keywords: Bioinformatics, HIV-1 Gag Gene, qRT-PCR, sequence.

1. Introduction

Human Immunodeficiency Virus (HIV) is a virus that offences the immune system, in particular targeting CD4 cells (T cells), which support the immune system's defence against infections. If left untreated, HIV can lead to Acquired Immuno Deficiency Syndrome (AIDS), which is the final stage of HIV infection. By the end of 2022, an estimated 39,000,000 people worldwide were HIV positive, with 1,500,000 of those individuals being children (0–14 years old). Several countries reported an increasing trend in new infections, even though they had previously decreased [1].

HIV-1 is the most common and pathogenic strain of the virus that is separated further into several subtypes (group M, N, O, and P), with group M being the most common and the primary cause of HIV infections globally. The genome of HIV-1 is roughly 9.7 kb in length and contains nine genes encoding 15 proteins. These genes are organized into three central structural genes (Gag, Pol, and Env) and six accessory genes (Tat, Rev, Nef, Vif, Vpr, and Vpu) [2].

The Gag proteins play a crucial role in assembling new virus particles. They interact with the viral RNA genome and other viral and cellular components to facilitate the formation of the immature virus particle [3]. The Gag gene is the most conserved gene of HIV-1. It is thought to play an essential role in determining the replication capacity of the virus, especially in the assembly, release, and maturation of infectious particles [4].

Early and accurate diagnosis of HIV-1 is essential for initiating timely antiretroviral therapy and preventing disease progression. Molecular detection methods, such as quantitative reverse transcriptase PCR (qRT-PCR), offer high sensitivity and specificity for detecting HIV-1 RNA in patient samples. This assay for quantifying viruses using oligo (dT)-containing primers has a high sensitivity and specificity to identify HIV-1 RNA. It is applicable to measure viral load in the patient's plasma and laboratory studies of virus production or viral gene transcription [5].

The quality of qRT-PCR analysis is primarily determined by the primers and probes used for DNA amplification. Primers and probes can be designed according to the agent sequence. Bioinformatics plays a crucial role in developing qRT-PCR assays targeting the Gag gene by facilitating the identification of conserved regions suitable for primer and probe design. By leveraging computational tools and approaches, researchers can analyse genetic sequences, predict potential targets for PCR primers and probes, design assays, and optimize detection protocols. Through bioinformatics programs, the sequence design of primer pairs and probes qRT-PCR for HIV detection using the HIV-1 Gag gene can be designed to be efficient [6].

A set of primers that match the target DNA will be needed for the target fragment. To amplify the target gene segment, specifically the Gag gene, the primer needs to satisfy certain requirements. Primer length, GC (guanine-cytosine) content, melting temperature (Tm), primer interactions (dimers and hairpins), primer stability, repeats, runs, and other factors are among those used in primer design. Next, a set of primers that match the target DNA will be needed for the target fragment. To amplify the target gene segment, specifically the Gag gene, the primer needs to satisfy certain requirements. Primer length, GC (guanine-cytosine) content, melting temperature (Tm), primer interactions (dimers and hairpins), primer stability, repeats, runs, and other factors are among those used in primer design [7].

This research aims to develop a molecular detection method for HIV-1 using the Gag gene through bioinformatics techniques to design primers and probes for effective qRT-PCR assays.

2. Methods

The software used in this study was the National Center for Biotechnology Information (NCBI) Genbank database, Bioedit sequences alignment editor version 7.2.5.0. and PerlPrimer.

The HIV-1 Gag gene sequence was obtained from HIV-1 isolates and serotypes, downloaded from the NCBI nucleotide database GenPeptd (https://www.ncbi.nlm.nih.gov/nucleotide/), then saved in FASTA format.

Next, the sequences were examined using the Bioedit sequences alignment editor version 7.2.5.0 ClustalW algorithm. Through gene alignment using multiple sequence alignment (MSA) with conserved regions. The probe sequence was selected from the HIV-1 Gag gene's region involving the forward and reverse primer sequences.

TABLE I: The Primer Oligonucleotide Sequence

Primer	Oligonucleotide sequence
F	5′-CAGTACAATGTGCTTCCACAGGG-3′
R	3′-GGATCTGACTTAGAGATAGGGC-5′

Note: F: Forward; R: Reverse.

TABLE II: The Outcomes of the Primer Analysis

Type	Length (base)	GC (%)	Tm (°C)	Run and repeat	Secondary structure
F	23	52	66	No	No
R	22	50	66	No	No

Note: F: Forward; R: Reverse.

TABLE III: The Probe Oligonucleotide Sequence

Type	Oligonucleotide sequence
Probe	5′-GGATCACCAGCAATATTTCAGGGAACG-3′

TABLE IV: The Outcomes of the Probe Analysis

Type	Length (base)	GC (%)	Tm (°C)	Run and repeat	Secondary structure
Probe	27	48	67.3	No	No

The obtained primer and probe pairs underwent quality control measures, such as measuring Tm, annealing temperature (Ta), GC content, GC clamps, repeats, runs, and tests to detect secondary structures (self, cross, heterodimer). Next, utilizing PerlPrimer, observations were made regarding the formation of hairpin loops and the adequacy of the primer pair's Tm. Using NCBI primer BLAST nucleotide sequences, primer pairings with HIV-1 were analyzed for compatibility.

3. Result

Two sets of primers were designed and used in this study. All primers designed were finally checked using NCBI primer BLAST nucleotide sequences to ensure the specific detection of all HIV-1 subtypes. To avoid self-complementary and secondary structure (hairpin) formation, primers were checked using the PerlPrimer application.

The oligonucleotide sequence of the primer is shown in Table I. The forward primer sequences are at bases 721 to 743, while the reverse primer sequences are at bases 856 to 877.

The results of the primer pair analysis using the PerlPrimer application shown in Table II.

The probe sequence is between the primer pair sequences at base positions 751 to 777 (Gag gene HIV-1 accession number KT805893.2). The oligonucleotide sequence of the probe is shown in Table III.

The results of the probe pair analysis using the PerlPrimer application are shown in Table IV.

4. Discussion

The Gag gene plays a crucial role in the diagnosis of HIV-1 using RT-PCR. The Gag gene is a highly conserved region of the HIV-1 genome, encoding the viral core proteins necessary for viral particle formation. Due to its conservation, the Gag gene is often targeted in RT-PCR assays for HIV-1 detection. Primers specific to the Gag gene are used to amplify viral RNA, and the amplified products are then detected using fluorescent probes or other detection methods. The use of the Gag gene in RT-PCR for HIV-1 diagnosis offers several advantages. First, the high conservation of the Gag gene ensures that the assay can detect a wide range of HIV-1 strains, including newly emerging variants. Second, the Gag gene is present in high copy numbers in infected cells, increasing the assay's sensitivity. Finally, targeting the Gag gene allows for the detection of early HIV-1 infection, even before the development of detectable antibodies.

Two sets of primers were designed and used in this study. The forward primer oligonucleotide sequence is 5′-CAGTACAATGTGCTTCCACAGGG-3′ while the reverse primer oligonucleotide is 3′-GGATCTGACTTAGAGATAGGGC-5′ (Table I).

The primer in Table II demonstrates the qualities of an excellent primer for amplification. The primer length obtained 23 bases for forward and 22 bases for reverse. This length has ideal primer criteria: between 15 to 30 nucleotide bases. Primers are too brief and will be less specific [8]. Short PCR primer lengths will be susceptible to mispriming (pasting errors). At the same time, too long primers can undergo hybridisation to inhibit the DNA polymerisation process [9]. Otherwise, primers that are more than 30 bases may cause secondary structures in the primer [10].

GC content is the proportion of guanine and cytosine bases in a primer. The GC content plays a significant role in the PCR primer design. The percentage of GC is associated with the binding involving DNA strands. The primer pair in this study have GC percentages of 52% for forward and 50% for reverse. The ideal range for the proportion of GC content is between 40%–60% [7]. GC less than 30%, or higher than 70%, is complex due to secondary structures that block the DNA polymerase as well as mispriming and mis-annealing of the DNA. This complexity often generates incomplete or nonspecific products that hamper downstream applications [11]. The GC content as high as about 84% affects the optimal annealing temperature and primer specificity, making PCR amplification more difficult [12].

The primer pair shares the same Tm, which is 66°C. Melting temperature (Tm) is the temperature at which one-half of the DNA duplex dissociates to become single-stranded and indicates the duplex stability. To optimize for q-PCR, find primers of minimal length with melting temperatures (Tm) between 59°C and 68°C. The Tm of the primer pair should be within one °C of each other [13]. The primer with the higher Tm tends to secondary annealing, while the primer with the lower Tm would have difficulty binding at an annealing temperature chosen for these primers. This can drastically reduce the yield and specificity of PCR and even cause PCR to fail [7]. The two primer sets are designed with similar Tm so that both primers can quickly bond [14].

The DNA target length depends on what the copied DNA fragment (the 'amplicon' or 'PCR product') will be used for. Obtaining large amplicons may be difficult because of poor amplification efficiency. The PCR amplification cycle may need to be redesigned [14]. The amplicon size should be optimised for the specific application and the PCR conditions. In general, amplicon sizes between 50 and 300 bp are commonly used in qRT-PCR, although smaller amplicons can also be used [15]. Shorter amplicons amplify more efficiently than longer amplicons and are more tolerant of suboptimal reaction conditions [16]. Holm *et al*. also states that extremely effective qPCRs utilize amplicons less extended than 200 bp because the possibility of polymerisation mistakes is decreased, facilitating prompt, accurate, and efficient quantification [17].

The probe sequence is between the primer pair sequences. The oligonucleotide sequence of the probe is 5′-GGATCACCAGCAATATTTCAGGGAACG-3′. The quality assay of the Gag gene probe sequence also shows good results obtained by the requirements for qRT-PCR (Table IV).

The probe is designed to bind the one of the target strands and is close to the forward and reverse primers but does not overlap the primer binding site on the same strand. The probe sequence length was 27 bases, with a GC content of 48% and a Tm of 67.3°C, which was higher than the Tm of the primer pair so that it met the standard as a probe. In the probe sequence, there is no repetition of 2 bases more than four times, and no secondary structure is formed, which can affect the annealing process and qRT-PCR amplification results [16].

No excessive base sequence repetitions (repeats and runs) were found in the primer pair and probe sequence. The repeat of G or C bases three or more times in a DNA sequence can lead to false priming in the amplification process of PCR. These repeats can form stable secondary structures, such as hairpins, that may serve as priming sites for the PCR reaction. When these structures form, they can lead to non-specific binding of primers, resulting in amplification of non-target sequences.

The results analysis for the possibility of secondary structure formation using Perl Primers showed that hairpin loops and dimers were not formed in the primer pair and probe sequences as indicated by the energy to break down the hairpin structure (ΔG) at the 3′ end being smaller than −3 kcal/mol. The formation of secondary structures in the primer and probe can prevent the primer from attaching to the DNA template [16]. Furthermore, the dimers at the 3′ end can prevent the process of amplification so the PCR product may be reduced or not formed [18]. A 3′ dimer shall be stated hybridisation involving equivalent primer bases because of the complementary sequence at the 3′ end [10].

Good primer and probe sequences can be used to identify commonalities with other organisms. These similarities can then be utilized as templates to create primer and probe gene candidate maps and to facilitate the amplification step of the qRT-PCR method using the NCBI's "Nucleotide-Blast" application. The primer pair and probe Gag gene had 100% similarity with HIV-1, according to the homology test results. This means there is no possibility of cross-reaction (avoid cross-homology) between the primer and probe with organisms other than HIV-1, so the primer and probe meet the requirements.

5. Conclusion

Based on the primer and probe analysis results, it is known that the primer and probe pair meet various best criteria. This means that the primers and probes designed using this bioinformatics method are suitable for the qRT-PCR process and can produce products according to the desired range of areas.

Acknowledgment

The facilities for this study were provided by The Regional Research and Innovation Agency of Bima, West Nusa Tenggara, and the Bina Cipta Husada Institute of Health, Purwokerto, Central Java, which the authors would like to thank.

Funding

The Regional Research and Innovation Agency of Bima, West Nusa Tenggara, and The Bina Cipta Husada Institute of Health, Purwokerto, Central Java, funded this project.

Conflict of Interest

Authors declare no conflict of interest.

References

[1] Wolrld Health Organization. *The global health observatory HIV: global situation and trends*. 2023. Available from: https://www.who.int/data/gho/data/themes/hiv-aids.

[2] Marie V, Gordon ML. The HIV-1 gag protein displays extensive functional and structural roles in virus replication and infectivity. *Int J Mol Sci*. 2022;23(14):1–15.

[3] Freed EO. HIV-1 assembly, release and maturation. *Nat Rev Microbiol*. 2015;13(8):484–96.

[4] Klingler J, Anton H, Réal E, Zeiger M, Moog C, Mély Y, *et al.* How HIV-1 gag manipulates its host cell proteins: a focus on interactors of the nucleocapsid domain. *Viruses*. 2020;12(8):888.

[5] Shan L, Rabi SA, Laird GM, Eisele EE, Zhang H, Margolick JB, *et al.* A novel PCR assay for quantification of HIV-1 RNA. *J Virol*. 2013;87(11):6521–5.

[6] Asryadin, Khasanah NAH, Yuniati NI. Desain primer dan probe melalui pendekatan bioinformatika untuk deteksi gen gag HIV-1 menggunakan qRT-PCR. *Pros Asos Institusi Pendidik Tinggi Teknol Lab Med Indones*. 2023;2:118–28.

[7] Borah P. Primer designing for PCR. *Sci Vis*. 2011;11(3):134–6.

[8] Hung JH, Weng Z. Sequence alignment and homology search with BLAST and ClustalW. *Cold Spring Harb Protoc*. 2016;11:1–7. doi: 10.1101/pdb.prot093088.

[9] Atifah Y, Achyar A. Design of specific primer for methallothionein gene of tor fish (Tor tambra). *Nat Sci J Sci Technol*. 2023;11(2): 42–8.

[10] Syamsidi A, Aanisah N, Fiqram R, Al JI. Primer design and analysis for detection of mecA gene. *J Trop Pharm Chem*. 2021;5(3):245–53.

[11] Guido N, Starostina E, Leake D, Saaem I. Improved PCR amplification of broad spectrum GC DNA templates. *PLoS One*. 2016;11(6):1–11.

[12] Yang Z, Yang J, Yue L, Shen B, Wang J, Miao Y, *et al.* Enhancement effects and mechanism studies of two bismuth-based materials assisted by DMSO and glycerol in GC-rich PCR. *Molecules*. 2023;28(11):1–19.

[13] Thornton B, Basu C. Real-time PCR (qPCR) primer design using free online software. *Biochem Mol Biol Educ*. 2011;39(2):145–54.

[14] Briefs AMC. PCR-the polymerase chain reaction. *Anal Methods*. 2014;6(2):333–6.

[15] Kozina V, Cappallo-Obermann H, Gromoll J, Spiess AN. A one-step real-time multiplex PCR for screening Y-chromosomal microdeletions without downstream amplicon size analysis. *PLoS One*. 2011;6(8):1–10.

[16] Rodríguez A, Rodríguez M, Córdoba JJ, Andrade MJ. Design of primers and probes for quantitative real-time PCR methods. *Methods Mol Biol*. 2015;1275:31–56.

[17] Van Holm W, Ghesquière J, Boon N, Verspecht T, Bernaerts K, Zayed N, *et al.* A viability quantitative PCR dilemma: are longer amplicons better? *Appl Environ Microbiol*. 2021;87(5):1–11.

[18] Kumar A, Chordia N. In silico PCR primer designing and validation. In *PCR Primer Design: Methods in Molecular Biology Springer Protocols*. 2nd ed. Basu C, Ed. Humana Press Inc.; 2015, pp. 143–51.