Role of Real-time Reverse Transcription - Polymerase Chain Reaction in the Detection of COVID - 19

Prof. Ranil Dassanayake and Mr Charitha Rajapakse

Coronavirus disease 2019 (COVID - 19) is caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). Since its first discovery in Wuhan, China in late 2019, it has escalated into a global pandemic within a few months, affecting more than 185 countries. As of 1st of May 2020, it has claimed more than 200,000 lives with more than three million confirmed cases worldwide. Due to the alarmingly rapid rate of the viral spread and the accompanying surge of COVID-19 cases, healthcare systems are challenged with the necessity to control the spread of the virus while treating and monitoring the patients. The presence of a considerable number of mild and asymptomatic cases renders the clinical characteristics alone insufficient for an accurate diagnosis. Furthermore, it has been observed that there is a time window between the viral infection and the manifestation of the clinical symptoms. However, these asymptomatic cases are capable of transmitting the virus. These features present an additional challenge in curbing the spread of the virus. Therefore, an accurate standard method for the detection of the virus is of utmost importance. Nucleic acid amplification tests (NAAT) have emerged as simple and reliable tests for the detection of viral pathogens. Among NAAT, the Polymerase Chain Reaction (PCR) based method is considered to be the gold standard for the detection and quantification of viral RNA.

SARS-CoV-2 belongs to a sub-family of the coronaviridae class of viruses which are capable of infecting mammals and birds. They are commonly found and are responsible for causing a wide array of symptoms ranging from mild to severe. In fact, about one third of the viruses that cause common cold are coronaviruses. They contain an envelope with protein spikes which appears as a crown in electron micrographs, thus inspiring the name 'coronavirus'. These protein spikes play an important role in the entry of the virus into host cells. A schematic representation of the SARS-CoV-2 viral structure is shown in Figure 1. SARS-CoV-2 is an RNA virus consisting of a single stranded positive sense RNA molecule. The term 'positive sense' indicates that the viral genome can be directly used for protein synthesis by translation in host cell ribosomes. This RNA molecule

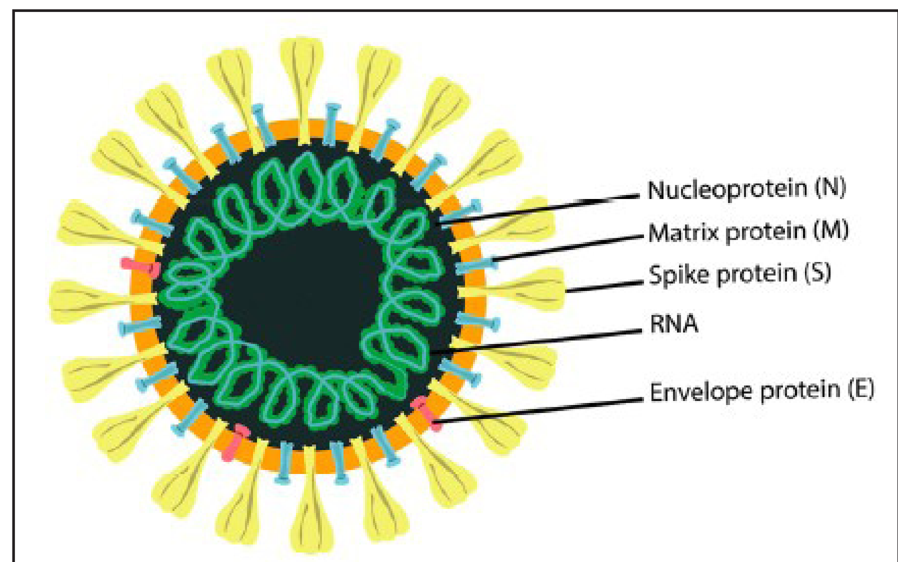

Figure 01 : A representation of the SARS-CoV-19 viral structure

is less than 30,000 bases long and codes for 16 proteins which are necessary for the survival and the propagation of the virus. The RNA genome of the virus contains sequences that are conserved among the virus family as well as sequences unique to the virus. Such sequences unique to the SARS-CoV-2 virus allow for the specific detection of the virus using nucleic acid based detection methods such as RT-PCR.

The detection of SARS-CoV-2 viral particles is based on the specific amplification of the unique regions of the viral genome. The amplification is carried out by a technique referred to as Polymerase Chain Reaction (PCR) in which a desired section of DNA is amplified to a detectable concentration. Since the SARS-CoV-2 genome is made up of RNA, an additional step has to be carried out to synthesize complementary DNA (cDNA). This reaction is called reverse transcription and thus the entire process is referred to as Reverse Transcription - Polymerase Chain Reaction (RT-PCR). This process allows for the real time detection of the amplification as well as the quantification of the viral load. As such it is also referred to as quantitative Real Time Reverse Transcription - Polymerase Chain Reaction (RT-qPCR).

In reverse transcription, an enzyme called reverse transcriptase is utilized. It is capable of synthesizing a complementary DNA (cDNA) molecule using an RNA molecule as the template. In addition to the template RNA and the enzyme, deoxyribonucleotide triphosphates (dNTPs) are used in the reaction as building blocks to synthesize the cDNA strand. Once the reaction mixture is composed, the reaction is facilitated by providing the optimal temperature required for the enzyme to function which is typically 42-48°C.

The composition of a PCR includes the template DNA to be amplified, oligonucleotide primers, a polymerase enzyme to catalyze the reaction, and dNTPs to synthesize new copies of DNA. In addition, magnesium ions which are required for the enzyme function are also added to the medium. A PCR consists of three repeating steps collectively referred to as a PCR cycle. The first step is the denaturation of the double stranded template DNA which is facilitated by the elevation of the temperature to around 94-98°C. In the second step, the temperature is decreased to facilitate the binding of oligonucleotide primers to the denatured template strands. The two primers will bind to the two separated strands of the template DNA flanking the desired region to be amplified. This binding, referred to as annealing, occurs due to the complementary base pairing between the template and the primers. Thus, it determines the specificity of the reaction. The annealing temperature can vary between 48-72°C and is determined by the lengths and the base compositions of the primers. In the third step, the primers are extended by adding nucleotides using the template strand as a guide. Extension is carried out by a polymerase enzyme which is typically active at 68-72°C. At the end of each PCR cycle, the number of copies of the template DNA is doubled. In order to obtain a considerable number of copies, the PCR is usually carried out for about 30-40 cycles. A PCR can be simplex where only one target is amplified, or multiplexed where more than one target is amplified simultaneously. Multiplexing is achieved via the utilization of several pairs of primers.

Both the above reactions are carried out in a buffered aqueous medium to provide the optimal chemical environment for the enzymatic reaction. An RT-PCR can be done either in a single step or in two steps. In a single step RT-PCR, the constituents required for both the reverse transcription and the PCR are added to the same vessel and the temperatures are provided such that the reverse transcription is followed by the PCR. In a two-step process, the two reactions are carried out separately.

In real time PCR, the amplification process is monitored at the end of each cycle. This is achieved by introducing a fluorescence-quencher probe to the reaction composition. The probe used is a single stranded oligonucleotide whose sequence is complementary to a region within the template to be amplified. Thus, it is capable of annealing the denatured template strand. The probe is tagged with a fluorescence reporter molecule at the 5' end and a quencher molecule at the 3' end. When the probe is intact, the quencher molecule captures the fluorescence emitted by the reporter molecule, thus no fluorescence can be detected. During primer extension, the annealed probe is hydrolyzed by the DNA polymerase so that the fluorophore and the quencher are separated. This results in the emission of a detectable fluorescence signal. A schematic representation of this principle

is shown in Figure 2. Since only the probe molecules bound to a template are cleaved, the intensity of the emitted fluorescence is proportional to the number of copies of the template DNA at the beginning of the cycle. Thus, by recording the emitted fluorescence at the end of each cycle, the progress of amplification of cDNA can be visualized in real time.

The quantitative endpoint for real-time PCR is the threshold cycle (Ct). The Ct is defined as the PCR cycle at which the fluorescent signal of the reporter molecule exceeds the chosen threshold above background fluorescence. There is an inverse relationship between the numerical value of Ct and the amount of amplicons in the reaction. That is, lower the Ct higher the amplicons in the reaction mixture. This Ct value contributes in generating a numerical relationship to analyze results. Moreover, both positive and negative controls should be in place in every RT-qPCR assay, as it is important to carry out an accurate interpretation of the end results. Negative control should not exhibit fluorescence signals more than the background to obtain a Ct value. A negative control which obtains a Ct value is considered as a false positive, thus the run should be invalidated and the assay should be repeated with utmost care.

It is essential for the success of the procedure to provide the precise temperatures at each step. This is difficult to achieve manually. Thus, an instrument called the thermocycler, which is capable of precisely providing the appropriate temperatures, is used. The apparatus may also include a mechanism to detect and record the emitted fluorescence.

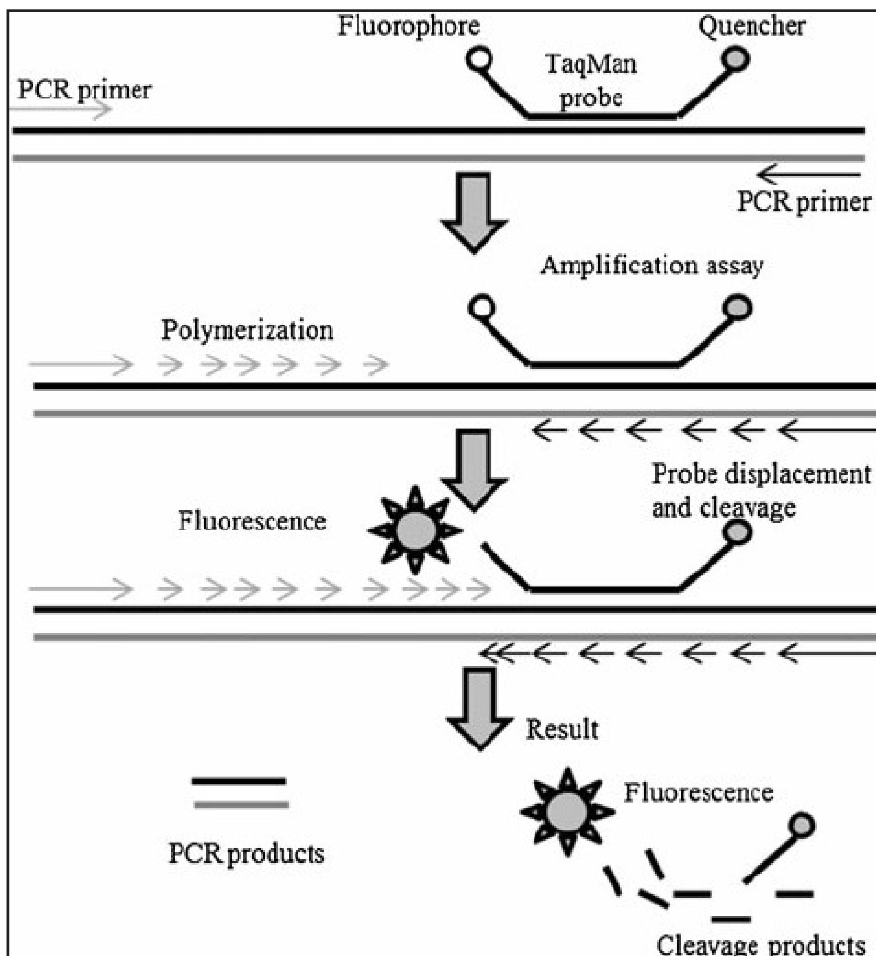

Figure 02 : Emission of fluorescence by the flourophore after the cleavage of the probe by DNA polymerase

The design and establishment of an RT-qPCR assay for the detection of a virus typically involves two steps. The first step involves the design of specific primers and probes. Since the specificity of the test is determined by the primers used, it is important to thoroughly and attentively analyze the genomic sequence of the target virus as well as those of related viruses. The primers should be designed in such a manner that they can selectively recognize the SARS-CoV-2 viral genomic sequences. The second step involves optimization of the reaction conditions and testing the assay.

In order to detect SARS-CoV-2, the viral RNA can be extracted from several types of samples. These include respiratory specimens such as sputum, bronchoalveolar lavage fluid (BALF), tracheal aspirates, as well as nasopharyngeal or oropharyngeal aspirates, washes or swabs. Among these, BALF is recommended for the monitoring of the virus in severe cases. However, its collection might be somewhat difficult as well as cause discomfort to the patients, rendering its use impractical for routine diagnosis. Thus samples which are more rapid, simpler and safer to collect, such as sputum,

nasopharyngeal and oropharyngeal swabs, are typically used in routine diagnosis. Furthermore, the sample used for viral RNA extraction as well as the time of its collection may have a significant impact on the accuracy of the test depending on the viral load they contain. It is also imperative that the collection of samples is done without introducing any contaminants or amplification inhibitors. Furthermore, the collected samples should be transported to the laboratories as soon as possible to test the viral RNA, otherwise RNA may get decomposed due to its inherent instability.

Once the RNA is extracted from a sample collected from a suspected patient, the reaction mixture is prepared by adding the required constituents according to the optimized volume of each component. The reaction vessels are placed inside the thermocycler and the assay is conducted under the optimized parameters. The thermocycler cycles through programmed temperatures triggering the processes described above ultimately resulting in the amplification of the targeted section of cDNA. Continuous amplification of the cDNA is detected via the fluorescence signal emitted due to the cleavage of the fluorophore-quencher probe. It is expected that in a well-designed and optimized RT-qPCR assay to detect SARS-CoV-2, the amplification and the subsequent emission of fluorescence will only occur if the virus is present.
Different organizations and companies have developed several RT-qPCR assays. These include both simplex and multiplex RT-qPCRs. They all employ the approach described above. However, different assays may achieve specificity by selectively amplifying different unique regions of the SARS-CoV-2 viral genome. This is determined by the primers and probes used in each assay.
The RT-qPCR assays developed by Centers for Disease Control and Prevention, United States (CDC, US), National Institute of Infectious Diseases, Japan and National Institute of Health, Thailand all target different regions of the N gene which encodes the nucleocapsid protein. Among these, the assay developed by CDC, US is a multiplex assay while the other two are simplex assays. Chinese Center for Disease Control and Prevention (China CDC) has developed a multiplex RT-qPCR assay targeting sequences in the N gene and the ORF1ab gene. Charite - Universitätsmedizin Berlin, Germany has developed a multiplex RT-qPCR assay which targets the RdRP gene (RNA-dependent RNA polymerase gene), the E gene (Envelope protein gene) and the N gene. The multiplex RT-qPCR assay developed by the Hong Kong University targets regions in the ORF1b-nsp14 and the N gene.

The RT-qPCR assay is characterized by its high sensitivity and specificity. Thus, it allows to achieve a highly accurate qualitative assessment about the presence of viral RNA in a sample. It can also be used to generate accurate quantitative data to assess viral load in a sample which might be helpful in monitoring the disease progression. In addition, this method allows early detection of the infected individuals compared to assays based on immunological detection. There are also a few limitations associated with this technique.
It can only be performed by trained personnel who are familiar with testing procedures and interpretation of results.
In addition, the requirement of sophisticated equipment such as thermocyclers confines the testing procedure only to laboratories with the aforementioned resources. Thus, in addition to the time spent for the procedure itself, additional time will be spent on sample transportation as well. Hence, the overall turn-around-time for the test is typically long. Furthermore, there is a possibility of the occurrence of false negative results due to improper sample collection, handling and transportation.

Prof. Ranil Dassanayake
Department of Chemistry
Faculty of Science
University of Colombo,
Colombo 3
vijitha@eng.pdn.ac.lk

Mr. Charitha Rajapakse
Department of Chemistry
Faculty of Science
University of Colombo,
Colombo 3
0714449244