

# Tracing the Evolution of Virus

Komal Sharma<sup>a</sup> and Dr. Rahul Roy<sup>b</sup>

<sup>a</sup> Indian Institute of Information Technology, Deoghat-Jhalwa, Allahabad, komalsharma.183@gmail.com  
<sup>b</sup> Department of Chemical Engineering, Indian Institute of Science, Bangalore, rahulfx16@nanobiophotonics.org

## ABSTRACT

Most RNA viruses undergo mutations inside the host at an incessant rate that allows them to elude the immune response. This constant evolution in RNA viruses is due to the low fidelity in replication carried out by the virus encoded RNA dependent RNA polymerases. This mis-incorporation of bases leads to establishment of a quasi-species of virus in the host. This makes it a challenge to combat viral diseases. To trace the evolution of RNA viruses, researchers resort to sequencing of virus genomes. However, even next generation sequencing techniques that rely on massively parallel sequencing of genomes generate processing and systemic errors. Here, we propose a method to distinguish errors incorporated during reverse transcription (RT), a key step in processing and sequencing of all RNA molecules including RNA viral genomes. We hypothesize that by generating multiple cDNA copies from every individual RNA molecule, we can distinguish RT induced errors from real sequence variants in our sequences. This approach relies on generation of cDNA by reverse transcriptase in presence of UvrD helicase that unwinds the RNA-DNA hybrid enabling repeated cDNA generation.

**Keywords:** RNA dependent RNA polymerase; UvrD helicase; single virus genomics; Next generation sequencing.

## 1. INTRODUCTION

Evolution is a phenomenon that the common ancestors of life on earth have undergone since time immortal to become a better and improved version of their selves. This concept is applicable to viruses where evolution is a consequence of several processes like spontaneous mutations (due to lack of proofreading activity in RNA dependent RNA polymerases) and recombination. The error rate of the viral RNA dependent RNA polymerase (RdRP) is estimated to be between  $1.5 \times 10^{-3}$  per bp (Q $\beta$  phage) and  $7.2 \times 10^{-5}$  per bp (influenza virus) [1]. Higher mutation rates explain why RNA viruses evolve rapidly. The ensemble population includes several variants which further get screened by natural selection. Such a population of individuals with different genomic sequences is called a quasi-species. Most viral infections proceed to generate a quasi-species upon infection. Few viral variants in this pool can evade drugs as the mutated sites can impair target-drug interactions (structurally or functionally). It therefore becomes important for us to understand how viruses evolve. Armed with this knowledge, we can think of disarming this pathogen of its virulence, and perhaps find a cure to many challenging viral diseases like AIDS, Dengue fever, Hepatitis, etc.

Next generation sequencing (NGS) has opened an avenue to trace viral lineages in a high-throughput fashion. Briefly, NGS workflow requires viral RNA to be reverse transcribed, PCR amplified, size selected, ligated to sequencing adaptors and sequenced base by base for several short segments of DNA in parallel. This extensive processing incorporates several errors and creates false outputs. For example, the first step of reverse transcription by reverse transcriptase incorporates errors during the process which are indistinguishable from those caused by RdRP in the sequence reads. This however, can be corrected by making multiple cDNA copies from each viral RNA. One approach to synthesize multiple cDNA relies on the activity of two different enzymes namely, Reverse transcriptase and UvrD Helicase. Reverse transcriptase is an enzyme which is responsible for the synthesis of a DNA strand which is complementary to the substrate template RNA strand. Its error rate is 1 error per  $10^6$  nucleotides polymerized. UvrD helicase is a member of the helicase SF1 super family which plays a crucial role in methyl directed mismatch repair and nucleotide excision repair, replication, and recombination [2-3]. This enzyme is capable of unwinding double stranded DNA (dsDNA) and proposed to unwind RNA-DNA hybrids [4]. It derives energy from the hydrolysis of ATP to break the hydrogen bonds linking the two strands. It initiates the unwinding from 3' end of the dsDNA tail and unwinds the duplex whilst moving along the strand in 3' to 5' direction [5]. UvrD helicase has strand switching property and it can translocate backward on the other strand which allows the DNA to reanneal in its wake [6]. The reannealing of separated strands can be prevented by using single stranded DNA binding protein. SSB exists as a stable homotetramer in solution; each tetramer has four ssDNA binding sites which can bind to the ssDNA. Because the activity of a thermostable version of UvrD helicase is optimum at 60°C, multiple cDNA copies can be produced without

denaturing reverse transcriptase which is sensitive to heat denaturation. We have tried to exploit aforementioned properties of these enzymes to make multiple cDNA copies in the following experiments.

## 2. MATERIALS AND METHODS

### 2.1 Substrates for UvrD helicase

Four single stranded oligonucleotides were annealed in following combinations: A-C, A-D, B-C, B-D in 1x Annealing buffer (10 mM KCl, 20 mM Tris-HCl, pH 8.8) at a concentration of 25uM by heating the reaction to 85°C and ramped down to 10°C at 1°C/min. The sequences of individual strands are as follows:

A	5' ATCAGGGTTGGAGTGAGTGGTGAC 3'
B	5' ACTGCATATCAGGGTTGGAGTGAGTGGTGAC 3'
C	5' CTCGTGACATCGCTACAGTCGACCAGGACTCTGTTCTGTATACTTGTCAACCACTCACTCCAA ACCCTGAT 3'
D	5' CTCGTGACATCGCTACAGTCGACCAGGACTCTGTTCTGTATACTTGTCAACCACTCACTCCAA ACCCTGATACGCTGCCAGCTGCAGTCA 3'

### 2.2 Optimal Substrate Geometry for Helicase Activity

To determine substrate specificity, helicase assays were performed with aforementioned four annealed substrates at 60°C for 60 minutes. Each of those four reactions were conducted with 3.5mM MgSO<sub>4</sub>, 1X of 5X FS buffer (250 mM Tris-HCl, 375 mM KCl, 15 mM MgCl<sub>2</sub>, pH 8.3), 5mM DTT, 3mM dATP or 3mM rATP, 6.6μM Helicase (BioHelix), 2.5μM Template, 7.5μM ET SSB (NEB). After 60 minutes of incubation, 2.5μl of Stop solution (0.25% Bromophenol blue, 25% Glycerol, 100mM EDTA) was added in each reaction to inactivate the enzyme. Each reaction was resolved on 12% native polyacrylamide gel and products were visualized by staining with ethidium bromide.

### 2.3 Helicase dependent isothermal reverse transcription

Reverse transcription was carried out with 1kb RNA (comprising 5' UTR and 3'UTR of dengue virus-2, TSV01) ligated to 70 bases DNA-oligo (for downstream processing) by T4 ssRNA ligase (NEB) as template in presence and absence of UvrD helicase. The reaction was incubated at 60°C for 70 minutes and aliquots were taken at respective time intervals, heat inactivated at 70°C and cDNA was quantified by real time PCR (Exicycler).

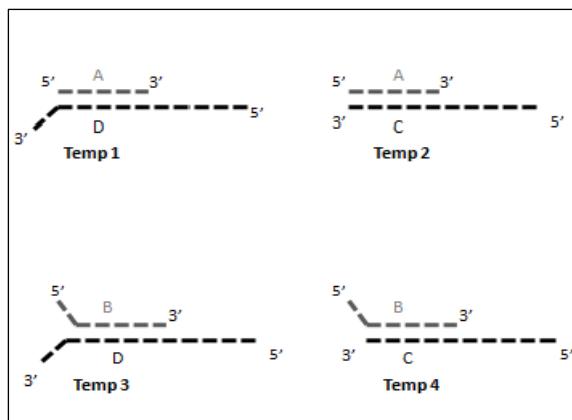
Composition of reaction mixture: 2.5μM of constructs, 10μM Gene specific primer, dNTP, 1X FS buffer, 3.5mM MgSO<sub>4</sub>, 3mM rATP, 7.5μM of ET-SSB, 1U of Maxima RT (Thermoscientific), 6.6μM Helicase (Biohelix).

## 3. RESULTS

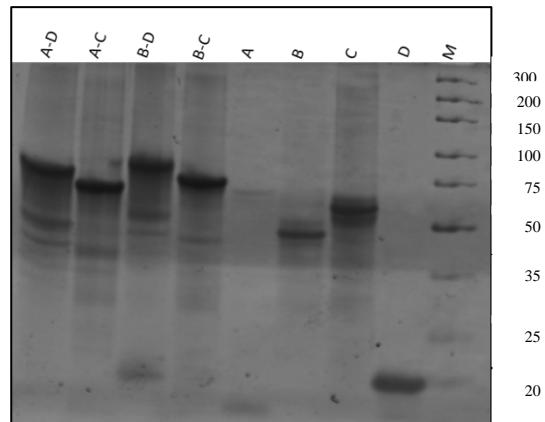
### 3.1 Optimal Substrate Geometry for Helicase Activity

Previous reports suggest that helicase is preferentially recruited on double stranded DNA (dsDNA) with 3' overhangs, however conflicting results with 5' overhangs are also reported [8]. To determine the optimal geometry of substrate required for recruitment of helicase and subsequent translocation, we designed four dsDNA constructs by annealing four oligonucleotides (A, B, C, and D: Fig. 1) of variable lengths (Fig. 2) which provide blunt ends, 5' and 3' overhangs. The unwinding reaction was carried out in a buffer which is compatible with reverse transcription and tested with UvrD helicase for substrate specificity. Comparing the gel profile of four templates, construct B-C, which has 5' overhangs on either end, shows complete loss of annealed band. However, reactions with helicase do not show unwound single strands presumably due to SSB binding that traps the DNA in the well.

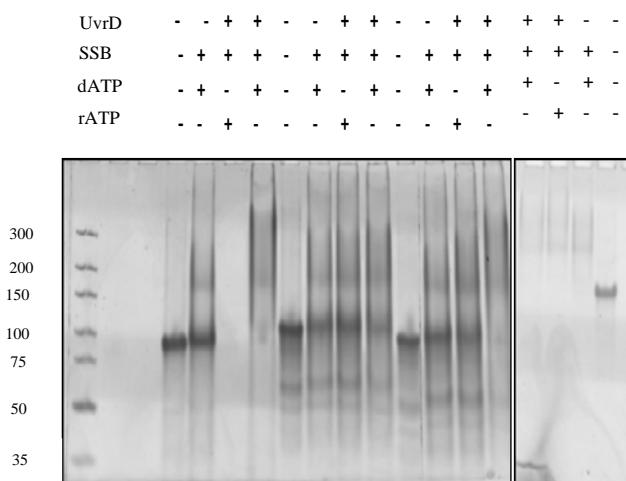
Furthermore, recent studies indicate that UvrD helicase can employ both dATP and rATP as a cofactor during catalysis [9]. We evaluated the suitable cofactor for aforementioned assay and found that dATP drives unwinding more efficiently than rATP (Figure 3). Further confirmatory experiments need to be conducted to validate these results. Though, our results that helicase unwinds dsDNA with 5' overhangs under given buffer conditions and this catalysis is more pronounced with dATP.



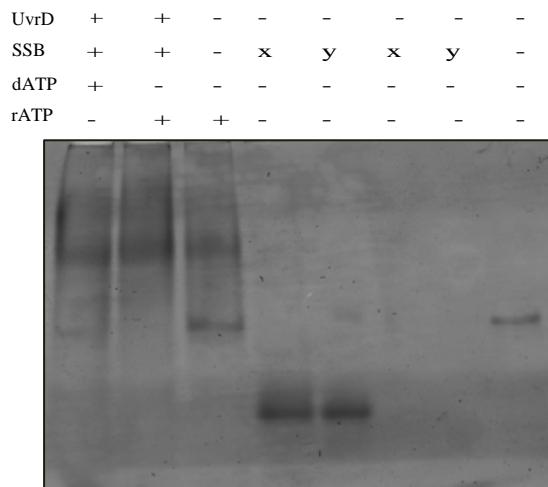
**Figure 1:** Helicase Substrate. Temp-1 has overhang at 3' and 5' end of the larger strand D, Temp-2 has overhang at 5' end of the larger strand C, Temp-3 has one overhang at 5' end of the smaller strand and two overhangs on the larger strand, Temp-4 has two overhangs and both are at the 5' ends of the template.



**Figure 2:** Gel image of the four constructs. First lane has Temp 1 (A-D), second lane has Temp 2 (A-C), third lane has Temp 3 (B-D), and fourth lane has Temp 4 (B-C). Lanes fifth to eighth have single stranded deoxyoligos A, B, C, D respectively.



**Figure 3:** Substrate unwinding reaction. Lane 1 has marker; Lanes 4 to 7 have Temp-4; Lanes 8 to 11 have Temp-3; Lanes 12 to 15 have Temp-1; Lanes 16 to 19 have Temp-2.

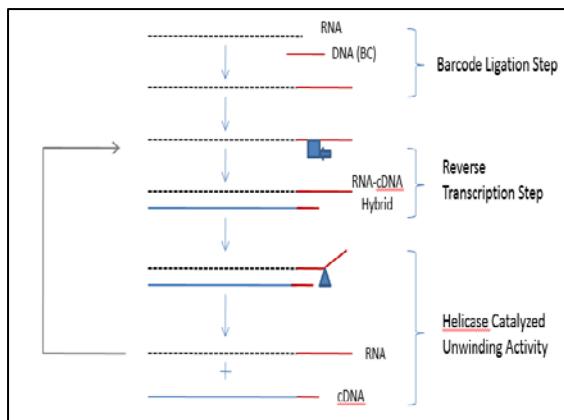


**Figure 4:** Substrate unwinding reaction. x is template and SSB in 1:3 ratio, and y is template and SSB in 1:6 ratio.

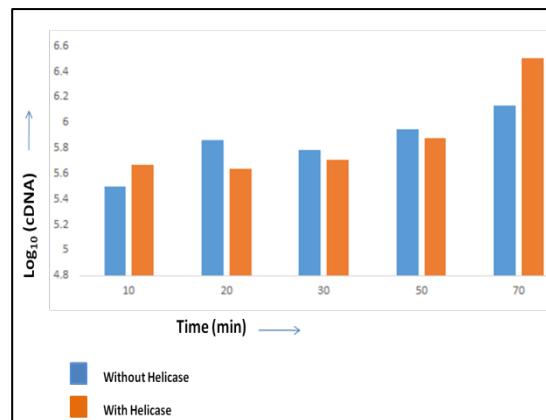
### 3.2 Helicase dependent isothermal reverse transcription

Helicase in coordination with reverse transcriptase was tested for multiple cDNA synthesis on an RNA template (1kb, 5' and 3' UTR Dengue type 2 virus RNA construct: *in vitro* transcribed from linearized plasmid by T7 polymerase, NEB) which is ligated to a 70 bases long DNA at its 3' end by T4 ssRNA ligase (NEB). This DNA segment acts as a barcode to assemble the final sequence of single RNA molecules. Each barcode uniquely tags to individual viral RNA molecule and the downstream experiment is designed in such a way that the barcode carried by the cDNA from the RNA will be flanked to each of the PCR generated segments of the virus constructs. This allows us to assemble the fragments of a single cDNA uniquely.

Upon reverse transcribing this template in the presence and absence of UvrD helicase and quantifying cDNA with real time PCR, we see an appreciable increase in cDNA levels at 70 minutes with UvrD helicase. However, the expected linear increase in cDNA amount was not observed. One reason for this could be the reverse transcriptase employed in this reaction has RNase H activity. Further attempts need to be made with reverse transcriptase proteins lacking RNase H activity to verify this.



**Figure 5:** Schematic of multiple cDNA synthesis. Where, rectangle represents reverse transcriptase; arrow is gene specific primer; and triangle is UvrD helicase.



**Figure 6:** Helicase dependent isothermal reverse transcription for multiple cDNA synthesis.

#### 4. DISCUSSION

Substrate specificity for UvrD helicase was investigated and the DNA constructs (B-C) which carry 5' overhangs at both the ends of the duplex display the highest unwinding. Even though the amount of substrate altered in presence of UvrD in the reaction mixture indicates unwinding activity of helicase, but we could not observe the single strands (generated by unwinding reaction) distinctively in the gel images. Therefore, we cannot strongly claim that helicase was necessary for the unwinding activity. However, UvrD helicase also displays strand switching property, it is difficult to ascertain if the two strands annealed back after unwinding [7]. Based on this, we have designed an experiment to confirm helicase activity by employing a trap strand instead of SSB (single stranded DNA binding protein). The trap is a hairpin looped deoxy-oligonucleotide which is designed such that it can pair with the longer strand, separated by UvrD helicase, thereby inhibiting the unwound strands from reannealing.

We also observed that in majority of the cases, reactions with UvrD helicase, SSB, and dATP showed relatively higher reduction in the band intensities. These results indicate that dATP is crucial for the activity of UvrD helicase. Since, helicases require ATP hydrolysis, we can measure the increase in the quantity of inorganic phosphate to measure helicase activity. This can allow us to standardize the buffer conditions for multiple cDNA synthesis. Furthermore, the ATPase activity results can be used to ascertain the extent to which helicase can unwind the DNA-RNA hybrid.

#### ACKNOWLEDGMENT

I want to thank Dr. Rahul Roy for his able guidance and valuable suggestions that helped improve this manuscript. I also want to express my gratitude to Vaseef Rizvi for his assistance in designing and analysing the experimental data.

#### REFERENCES

1. Drake, John W. "Rates of spontaneous mutation among RNA viruses." *Proceedings of the National Academy of Sciences* 90.9 (1993): 4171-4175.
2. Lohman, Timothy M, and Keith P Bjornson. "Mechanisms of helicase-catalyzed DNA unwinding." *Annual review of biochemistry* 65.1 (1996): 169-214.
3. Yamaguchi, Miyuki, Vivian Dao, and Paul Modrich. "MutS and MutL activate DNA helicase II in a mismatch-dependent manner." *Journal of Biological Chemistry* 273.15 (1998): 9197-9201.
4. Goldmeyer, James, Huimin Kong, and Wen Tang. "Development of a novel one-tube isothermal reverse transcription thermophilic helicase-dependent amplification platform for rapid RNA detection." *The Journal of Molecular Diagnostics* 9.5 (2007): 639-644.

5. Dessinges, Marie-Noëlle et al. "Single-molecule assay reveals strand switching and enhanced processivity of UvrD." *Proceedings of the National Academy of Sciences of the United States of America* 101.17 (2004): 6439-6444.
6. Lohman, Timothy M, and Marilyn E Ferrari. "Escherichia coli single-stranded DNA-binding protein: multiple DNA-binding modes and cooperativities." *Annual review of biochemistry* 63.1 (1994): 527-570.
7. Brosh, Robert M, and Steven W Matson. "A Point Mutation in Escherichia coli DNA Helicase II Renders the Enzyme Nonfunctional in Two Dna Repair Pathways Evidence For Initiation Of Unwinding From A Nick In Vivo." *Journal of Biological Chemistry* 272.1 (1997): 572-579.
8. Singleton, Martin R, and Dale B Wigley. "Modularity and specialization in superfamily 1 and 2 helicases." *Journal of bacteriology* 184.7 (2002): 1819-1826.
9. Curti, Elena, Stephen J Smerdon, and Elaine O Davis. "Characterization of the helicase activity and substrate specificity of Mycobacterium tuberculosis UvrD." *Journal of bacteriology* 189.5 (2007): 1542-1555.