

ESCI - UPF

PROTEIN FUNCTION

## Structural Bioinformatics

*Marta Ortigas - Mònica Torner - Núria Mitjavila - Xavier Crespo*

Second year  
23/01/2023

## Contents

|                                         |          |
|-----------------------------------------|----------|
| <b>1 Understanding protein function</b> | <b>3</b> |
| 1.1 Protein function . . . . .          | 3        |
| <b>2 Bibliography</b>                   | <b>8</b> |

# 1 Understanding protein function

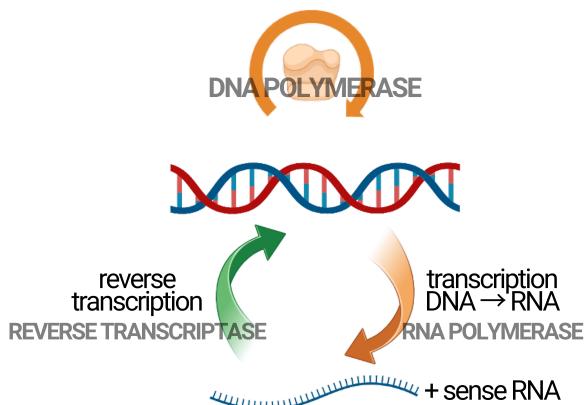
To answer the following questions, choose one member of the protein family you are working with. You will focus part of your project in this particular protein. Please, check that this protein has an available structure in the protein data bank. STRUCTURE: 1REV, COMPLEX WITH DNA: 3KJV

## 1.1 Protein function

### 1.- What is the function of your protein? Is the same for the whole protein family?

As indicated by the name, the main function of reverse transcriptase is to convert its RNA genomes into the viral DNA, to build DNA strands based on an RNA template. This allows the virus to persist and replicate within the host cell, being this the reason why the reverse transcription is an essential step in the retroviral replication and an important target for antiretroviral drugs used to treat HIV infection, as these drugs can inhibit the activity of the enzyme and prevent the virus from replicating.

The reverse transcriptase is part of the Eukaryotic DNA polymerase protein family, together with other proteins like the telomerase or different polymerases (alpha, beta, lambda, gamma...). Remember that, the main function of the DNA polymerase is to replicate the genome in order to ensure the maintenance of the genetic information and its faithful transmission through generations. Being more specific, we can find that the function of the telomerase is to replicate ends of linear chromosomes because the normal DNA polymerase cannot but, without using a template and for the other polymerases, they have more specific functions or this function is not clearly understood nowadays. For this reason, we can conclude that the function of the reverse transcriptase is similar to the protein family, but is not the same.



**Figure 1:** Reverse transcription diagram. Made in BioRender by Núria Mitjavila.

## **2.- How is able to carry out this function?**

Reverse transcriptase has two enzymatic activities, a DNA polymerase that can copy a DNA or RNA template, and an RNase H that cleaves RNA only if it is part of a duplex, cooperating together to convert the RNA into a double-stranded linear DNA. This conversion takes place in the cytoplasm of the infected cell; after DNA synthesis is completed, the linear double-stranded viral DNA is translocated to the nucleus, where the viral DNA is inserted into the host genome by IN.

The reverse transcriptases are enzymes capable of extracting the genetic information contained in an RNA molecule and transfer it to DNA and this reaction is performed in the polymerase active site, which is formed by two sets of arms that surround the RNA and DNA. After building the DNA strand, the enzyme then removes the original RNA strand by cleaving it into pieces. This is performed by a nuclease active site located at the opposite end of the enzyme. Finally, it builds a second DNA strand matched to the one that was just created to form the final DNA double helix. This reaction is also performed by the polymerase site.

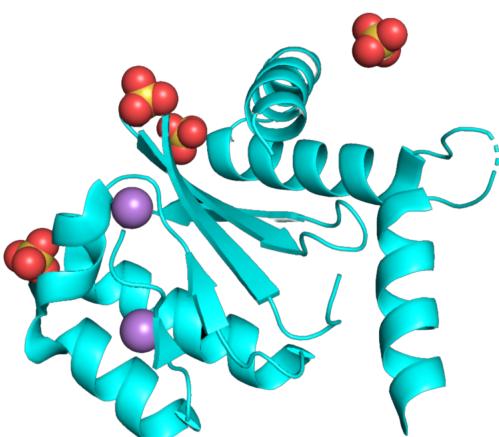
Although other viral proteins and some cellular factors help reverse transcriptase carry out the reactions that convert the viral RNA into DNA, reverse transcriptase contains all the necessary enzymatic activities for the conversion.

## **3.- Does your protein require the interaction with other proteins or molecules to carry out this function?**

There are several interactions that reverse transcriptase needs to have in order to perform reverse transcription. A protein that plays a remarkable role in this process is HIV integrase, an RT binding protein.

Other IN-binding proteins, including INI1, Sin3a complex components, and Gemin2, do also influence reverse transcription. Other proteins such as HuR and DNA topoisomerase may also interact with reverse transcriptase.

Moreover, eEF1, the cellular eukaryotic elongation factor 1 complex, has an important association with RT complex. This interaction, that happens by the binding of the subunit eEF1A to RT, is significant in the further stages of reverse transcription.



**Figure 2:** Structure of HIV-1 integrase (3LPT)

Also, reverse transcriptase is able to do its function thanks to other proteins or molecules already mentioned in other questions, such as DNA polymerase and a wide range of enzymes.

#### **4.- What is the fold of your protein? Is this the same fold for the other proteins of the family?**

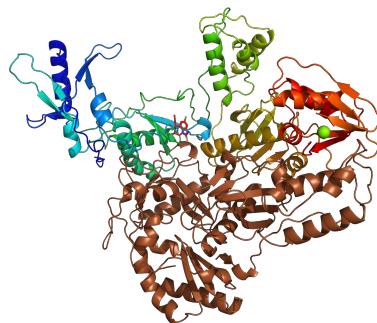
We found that the fold of our protein is Ferredoxin-like with alpha-helices and beta strand. Looking at the telomerases, a protein from the same family, we found that they have a different fold as the telomerases may have beta strands but they are proteins containing predominantly alpha-helices.

For answering this question, we follow the first seminar session about Blast. To search for proteins with available structure, we download the Fasta sequence from the PDB entry 1REV (the selection is explained in the following exercise) and we import it to the cluster to obtain templates inside the PDB which are homologues with our reverse transcriptase structure. From the Blast hits obtained, we searched them in the SCOP database, knowing that if these targets have some templates with a specific fold, we assume that the fold of the target is the same or similar. From the SCOP database, we get that the similar PDB entry is 2RKI and looking to the ancestry we observe that the family is reverse transcriptase-like, the superfamily is the DNA polymerase palm domain-like (common fold elaborated with additional secondary structures) and that the fold of the protein is Ferredoxin-like (beta-alpha-beta(2)-alpha-beta; antiparallel beta-sheet). Looking at the structural class, it comes from the alpha and beta proteins, proteins with segregated alpha-helices and beta-strands.

After looking if this was the same fold for the other proteins of the family, we decided to repeat the same process for the telomerases, downloading the fasta sequence from the PDB entry 3DU6. In this case we get from the SCOP that the similar PDB entry is 5C9H and looking to the ancestry we observe that the family is Telomerase RBD-like, the superfamily and the fold is TERT TR-binding domain-like (complex fold). Looking at the structural class, it comes from the all alpha proteins.

#### **5.- Are there available structures for your protein family? What are their PDB IDs?**

We have available structures in the PDB, most of them coming from the same scientific paper, that are crystal structures that are complex with inhibitors or mRNA sequences. The structure that we will take for the project is the one with the PDB IDs 1REV that is the HIV-1 Reverse Transcriptase structure without any complex. We would also take the same structure in complex with the DNA to better understand the function of the reverse transcriptase (the PDB ID is 3KJV).



**Figure 3:** Structure of HIV-1 reverse transcriptase (1REV)

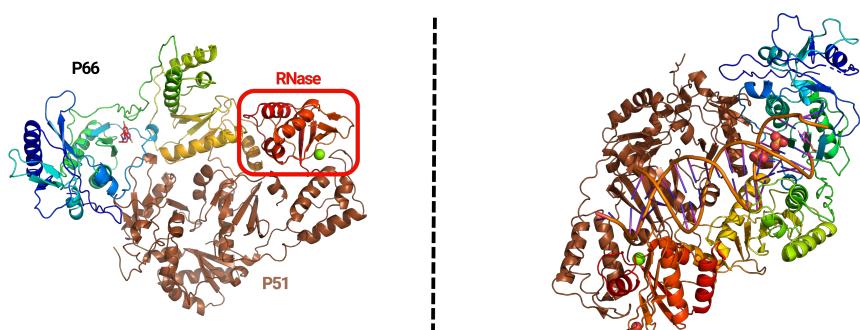
Also, reverse transcriptase is able to do its function thanks to other proteins or molecules already mentioned in other questions, such as DNA polymerase and a wide range of enzymes.

**6.- Does your protein have a region that is essential for its function? What is this region? Why is it essential to its function? Is this region also essential for the other proteins of the family?**

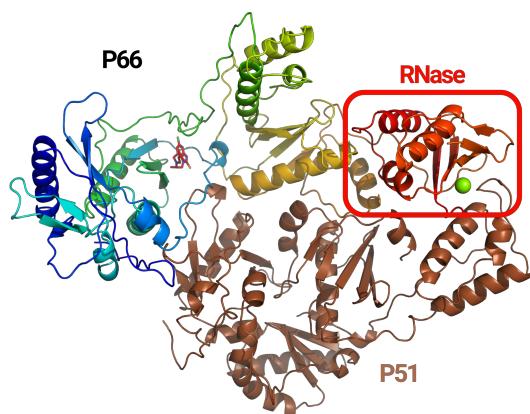
The reverse transcriptase has a catalytic domain formed by a Ribonuclease H (RNase H), an endonuclease that cleaves the RNA portion of RNA/DNA hybrids. In the case of HIV-1, RNase H non-specifically degrades the positive-strand RNA genome, while specifically removing the negative-strand tRNA primer and creating and removing the positive-strand or PPT primer during reverse transcription. We can not assure that this region is essential, but it seems important for the functionality of the protein because if these RNA/DNA hybrids that are formed during replication were not processed, it could lead to DNA instability.

This domain is found in the C-terminus of the P66 subunit of the reverse transcriptase.

For other proteins of the family, we could not assure that Ribonuclease H is essential, either due to the lack of information that exists regarding other proteins of the family such as telomerase and the RNase H.



**Figure 4:** At left side, RNase H in the reverse transcriptase. Next to it, reverse transcriptase in complex with DNA.



**Figure 5:** RNase H in the reverse transcriptase enlarged.

**7.- Use the UniProt database to choose a mutation that affects your protein. Try to find an interesting case, for example, a mutation that causes a disease. Describe the effects of this mutation at a molecular and phenotypical level.**

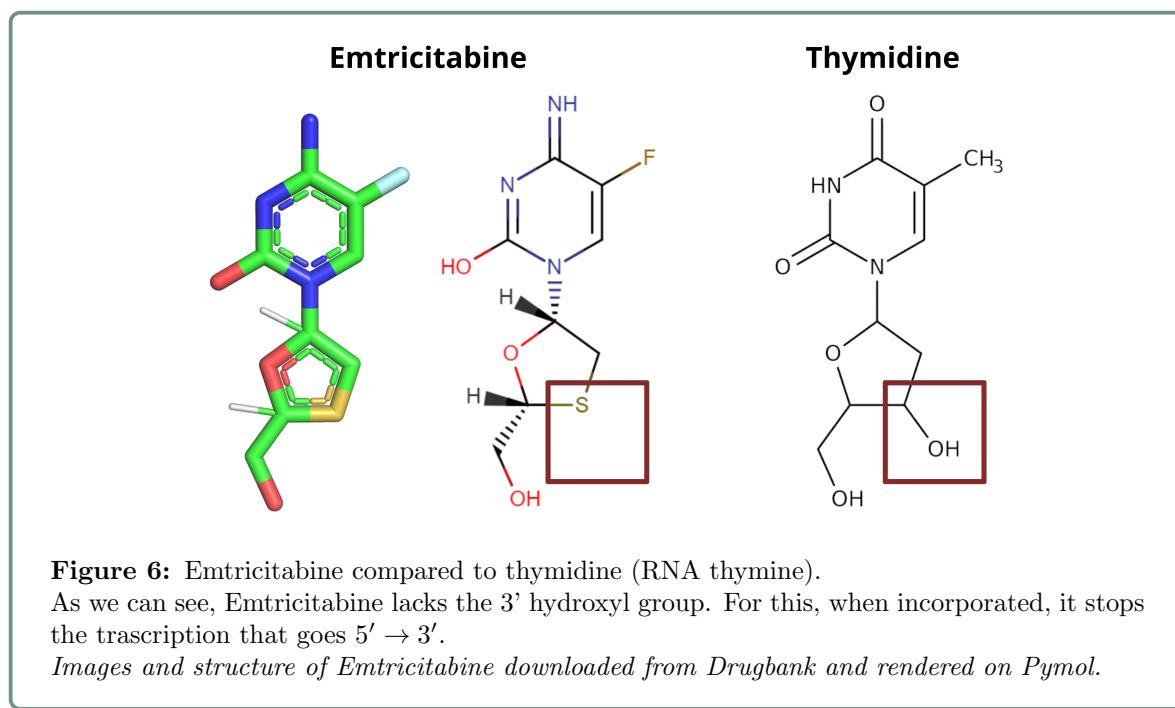
In our special case, as we are basing our project on a protein that leads to a disease, we cannot look for mutations that lead to a disease. Instead, we can look for some documentation on well known drugs and know more or less how do they work.

Based on prevention (PrEP: Pre-Exposure Prophylaxis), in this case, we chose Truvada (brand name).

Truvada is a prescription HIV nucleoside analog reverse transcriptase inhibitor (HIV NRTI) available in multiple countries. It contains two active ingredients: Emtricitabine and Tenofovir disoproxil. Emtricitabine is a medication used to treat and prevent HIV. Tenofovir disoproxil is a medication used to treat hepatitis B infection and to manage HIV infection.

We are going to stick with Emtricitabine as it is directly targeted to our protein. Briefly, Emtricitabine works by inhibiting HIV reverse transcriptase, preventing transcription of HIV RNA to DNA.

Emtricitabine is a cytidine analog that competes with the natural substrate of HIV-1 reverse transcriptase to be incorporated into newly formed DNA, terminating its transcription. More deeply into detail, Emtricitabine works by mimicking the structure of the nucleoside building blocks, such as thymidine, that are used by the reverse transcriptase enzyme to make a DNA copy of the viral RNA genome. Emtricitabine is incorporated into the growing viral DNA chain, and when the reverse transcriptase enzyme encounters it, it mistakes it for a natural nucleoside. However, the incorporated emtricitabine lacks the 3' hydroxyl group, which is necessary for the next nucleoside to be added to the chain. This terminates the DNA chain and prevents further replication of the virus. By inhibiting the reverse transcriptase, therefore stopping transcription of HIV RNA into DNA, the virus is unable to incorporate its viral DNA into host cells and reduces the viral load.



## 2 Bibliography

1. 1REV structure,  
<https://www.rcsb.org/structure/1REV>
2. 3KJV structure,  
<https://www.rcsb.org/structure/3KJV>
3. Visualizing the molecular interactions of a nucleotide analog, GS-9148, with HIV-1 reverse transcriptase-DNA complex,  
<https://pubmed.ncbi.nlm.nih.gov/20156454/>
4. UniProt page for P04585,  
<https://www.uniprot.org/uniprotkb/P04585/entry>
5. Truvada results at Drugbank,  
<https://go.drugbank.com/earth/qquery=truvada&button=&searcher=drugs>
6. Emtricitabine at Drugbank,  
<https://go.drugbank.com/drugs/DB00879>
7. Retroviral ribonuclease H at Wikipedia,  
[https://en.wikipedia.org/wiki/Retroviral\\_ribonuclease\\_H](https://en.wikipedia.org/wiki/Retroviral_ribonuclease_H)
8. Ribonuclease H: the enzymes in Eukaryotes,  
<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2746905/>
9. Design of Non-nucleoside Inhibitors of HIV-1 Reverse Transcriptase with Improved Drug Resistance Properties. 2.,  
<https://pubs.acs.org/doi/10.1021/jm040072r>
10. Evidence of Interactions between the Nucleocapsid Protein NCp7 and the Reverse Transcriptase of HIV-1,  
[https://www.jbc.org/article/S0021-9258\(19\)73636-9/pdf](https://www.jbc.org/article/S0021-9258(19)73636-9/pdf)
11. Specific Interaction between eEF1A and HIV RT Is Critical for HIV-1 Reverse Transcription and a Potential Anti-HIV Target,  
<https://journals.plos.org/plospathogens/article?id=10.1371/journal.ppat.1005289>
12. Binding of the eukaryotic translation elongation factor 1A with the 5'UTR of HIV-1 genomic RNA is important for reverse transcription,  
<https://virologyj.biomedcentral.com/articles/10.1186/s12985-015-0337-x>