

# Molecular Evolution: HIV Drug Targets and Resistance

Brian Rybarczyk

Published online: 19 January 2008  
© Springer Science + Business Media, LLC 2008

**Keywords** Molecular evolution · Sequence alignment · Drug resistance · HIV · Inquiry · Active learning

## HIV Evolution: Drug Targets and Resistance

### Student Activity

#### *Background*

Once the human immunodeficiency virus (HIV) binds to a host immune cell through the interactions of the envelope protein with CD4 and a coreceptor, the host and virus membranes fuse and the HIV core, called the matrix/capsid, containing its RNA genome and related proteins enter the cell. The matrix/capsid core breaks down and releases HIV's two single-stranded RNA genome molecules into the cell's cytoplasm. Reverse transcriptase then synthesizes a DNA copy of HIV's RNA genome. The double-stranded DNA copy is then transported to the nucleus where it is integrated into the host's genome by the enzyme integrase. This integrated copy is called a provirus. Once these provirus copies are integrated, there is no way to remove them from the host's genome.

Using the host cell's protein machinery, including ribosomes and other proteins involved in the translation process, the provirus is transcribed into RNA molecules and translated into HIV proteins. Once synthesized, the HIV components assemble near the cell membrane. During this process, the immature polypeptide chains are cleaved and processed by protease, another HIV enzyme, resulting in mature, functional HIV proteins. Millions of new virus

particles, each potentially capable of infecting other cells, can be produced from one infected CD4-positive cell. Scientists are trying to develop drugs that can interfere with each step in the HIV lifecycle to stop the release of new viral progeny.

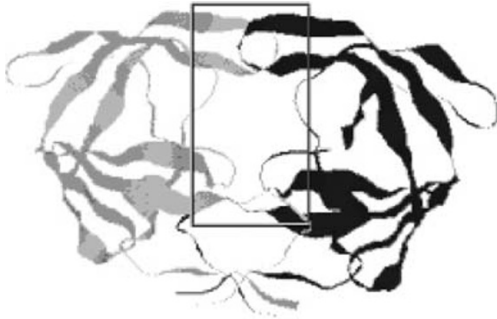
A potential target for antiretroviral drugs is the HIV protein, protease. When HIV progeny are released, they are 'immature' and noninfectious until protease cleaves certain proteins 'maturing' the virion. The structure of the protease enzyme is a homodimer, composed of two identical protein subunits that come together to form an active site (see Fig. 1). Each protein chain subunit is 99 amino acids long. Once folded properly, amino acid residues in the active site of the enzyme include the Asp–Thr–Gly sequence at positions 25 to 27, which are important for its enzymatic action, and Val82 and Ile84, which are accessible in the active site. If protease activity is blocked, 'new' viruses remain immature (Boden and Markowitz 1998; Resch et al. 2005).

Currently, nine protease inhibitors are available to treat HIV-infected individuals (<http://www.aidsmeds.com>). Ritonavir is one of these inhibitors. Research has shown that amino acid residues Val82 and Ile84 among others are important for ritonavir to bind to protease.

For another example, the first antiretroviral drug developed for HIV, AZT, is a chemical compound similar in structure to DNA nucleotides. AZT blocks the action of the reverse transcriptase enzyme. Early testing of AZT in patients resulted in decreased viral loads and decreased HIV replication. However, this inhibition lasted only a short time as HIV evolved so that reverse transcriptase could function even in the presence of the inhibitor. HIV can evolve resistance to any monotherapy currently in use. The appearance of these drug-resistant strains are a growing concern for researchers and HIV-infected patients, as these strains can be passed on to others, making the fight against HIV even more difficult (HIV Drug Resistance Database, <http://hivdb.stanford.edu>). Today, numerous antiretroviral

---

B. Rybarczyk (✉)  
University of North Carolina at Chapel Hill,  
Chapel Hill, NC 27599, USA  
e-mail: brybar@unc.edu



**Fig. 1** Structure of HIV protease: The 3-D structure shows protease as a homodimer. The *black outline box* indicates the enzyme's active site. Imagine that a long polypeptide chain is perpendicular to this flat surface going through the open active site of the enzyme

drugs have been developed and approved for use by HIV-infected individuals. These drugs are often taken as a cocktail, different types of inhibitors taken at the same time. A typical cocktail or drug combination will include at least one reverse transcriptase inhibitor and a protease inhibitor. Recently, integrase inhibitors have been developed and are in clinical trials; however, no compound in this class has been approved for therapy.

#### *Preactivity Questions*

1. If a polypeptide contains 99 amino acids, how many nucleotides does this represent?
2. Briefly describe the steps of how HIV evolution might occur in the presence of antiretroviral drugs.
3. Where in the host genome would it be most advantageous to insert a copy of the HIV genome? Where would it be most disadvantageous?
4. In terms of HIV evolution, why is it more effective to take multiple HIV drugs as opposed to only taking one drug?

#### **Protease Sequence Analysis Activity**

##### Case

John, an HIV positive individual, was given ritonavir monotherapy over a period of 1.5 years. Blood samples were taken at regular intervals to monitor his virus load, CD4 cell count, and for genotype analysis. John began to experience some minor side effects from the drug treatment last week.

##### Hypotheses

Generate several testable hypotheses to predict the evolutionary pattern of amino acid changes in HIV protease over time in this patient.

##### Data

Viral RNA was isolated from the blood samples; the HIV protease gene was amplified by PCR and sequenced. Each sequence is given a unique name and can be found in the file “protease sequences.txt”.

##### Sequence Analysis

Quick guide for Biology Workbench  
(<http://workbench.sdsc.edu/>)

1. Log into Biology Workbench, click the “**Session Tools**” button, select **New**, give your session a name, click **Start New Session**.
2. Upload sequences into the session. Batch uploading as described below will upload all the sequences from the one file provided and then separate each sequence so you can choose specific ones to work with.
  - Click the **Protein Tools** button since the data are amino acid sequences.
  - Click **Add** button. Browse your desktop to upload the sequence file “protease sequences.txt” into your session. Click the **Upload File** button, then click the **Save** button at the bottom of your sequence text boxes.
3. Now each sequence is uploaded separately into your session and you are ready to do more advanced manipulations.
4. Alignments:
  - Select the sequences you want to align by clicking the box next to each sequence name.
  - Select appropriate alignment algorithm (CLUSTALW—Multiple Sequence Alignment, will suffice for this exercise).
  - The next window will allow you to change various parameters. No need to change anything for this analysis so click the **Submit** button. Your sequences will be aligned.
  - You can save screen shots of your alignments (Print Screen button on the PC) and paste into a blank document for further analysis or just highlight the text and copy and paste into a document. Make sure you include the reference sequence named “ref\_NL43” in each of your alignments as the comparison sequence.

##### *Analysis Questions*

1. What are your observations and conclusions from the sequence alignment data? Why would some amino acids be conserved and others change over time? Does the data support or refute your proposed hypotheses?

2. What are the advantages and disadvantages of inhibiting protease function? Why would inhibiting integrase be considered a more effective approach to block HIV replication?
3. If you are a researcher investigating HIV, propose two additional molecular approaches to inhibit HIV replication and infection.

## HIV Evolution Drug Targets and Resistance

### Instructor Guide

#### Introduction

Currently, there are nearly 40 million people worldwide affected by HIV/AIDS (UNAIDS/WHO, [www.unaids.org](http://www.unaids.org)). The human immunodeficiency virus (HIV) has become a major health epidemic. Although basic scientific research has uncovered many facets of the biology of the virus, many questions remain unanswered. Although research has led to many treatment options in the form of antiretroviral drugs, there is currently no cure for HIV infection.

HIV uses three main enzymes during its life cycle. Reverse transcriptase converts HIV's RNA genome into DNA, integrase integrates multiple copies of the HIV genome into the host genome, and protease cleaves immature polypeptide chains into shorter chains for proper folding into mature proteins.

This module begins to address a common misconception that during antiretroviral treatment, the host, in this case a human, evolves immunity against the drugs that become ineffective. It is not the host that evolves but rather it is the virus itself that can alter its molecular composition to evade the effects of the drugs making them ineffective.

#### Learning Objectives

- Describe the process of microevolution of the protease enzyme.
- Propose possible selective pressures involved in virus evolution.
- Analyze molecular evidence for the evolution of the HIV protease enzyme.
- Propose approaches to more effectively treat HIV infection with antiretroviral drugs.

#### Implementation

- This lesson is primarily targeted to undergraduate students who have previous knowledge of basic cell biology and genetics.

- Students should have basic knowledge of the HIV lifecycle and the immune system. Students and instructors should visit the Centers for Disease Control and Prevention website (<http://www.cdc.gov/hiv>) and consult other resources for background information about HIV infection.
- Students can answer the preactivity questions using reliable internet resources and cite the references they used. Answers can be shared at the start of the sequence analysis activity (5 min).
- A computer with internet connectivity is needed for each pair of students to perform the sequence analysis activity outlined below. As an alternative, instructors can provide students the sequence alignment in the Answer Key below and then ask students to analyze this data to answer the questions.
- Students will use actual amino acid sequences to determine specific changes that occur during the course of infection as evidence of microevolution.
- Any program that has sequence alignment tools and uses sequences in FASTA format will be sufficient. One suggested resource is Biology Workbench (<http://workbench.sdsc.edu/>), a free, student-friendly bioinformatics resource.
- Once students are logged in, they will need to upload the protein sequence data file “protease sequences.txt” provided in this module. The sequences can be uploaded in batch. Students will use the CLUSTALW—Multiple Sequence Alignment function to compare sequences.
- A reference sequence is provided in the data named “ref\_NL43”
- Sequence alignment and data collection can be performed in a 50-min period followed by a 30-min debriefing session to discuss the results and “Analysis Questions”.

#### Answer Key

#### Preactivity questions:

1. If a polypeptide contains 99 amino acids, how many nucleotides does this represent?

1 amino acid represents 3 nucleotides, thus 99 amino acids=297 nucleotides.

2. Briefly describe the steps of how HIV evolution might occur in the presence of antiretroviral drugs.

Random mutations occur in the RNA genome when reverse transcriptase converts the RNA into DNA, these mutations are not corrected and are passed on in subsequent replication cycles. Selective pressures will change the frequency of certain variants in a population. Those virions adapted to a specific environment are considered “fit” and will survive. In this case, the

selective pressures are the antiretroviral drugs used to treat HIV infection.

3. Where would it be most advantageous to insert a copy of the HIV genome into the host genome? Where would it be most disadvantageous?

It would be most advantageous to insert a provirus into areas where there is a high rate of transcription (e.g., near ribosomal genes), where lots of genes are transcribed all the time. This will ensure that the HIV genome is copied when the host genome is copied. Other advantageous areas would be near active promoter regions. Areas of insertion that would be disadvantageous include areas where essential genes are located. Disrupting a host's essential genes which are necessary for survival would decrease HIV's chances of replicating and being passed on.

4. In terms of HIV evolution, why is it more effective to take multiple HIV drugs as opposed to only taking one drug?

As demonstrated with AZT monotherapy, HIV positive individuals can develop drug resistance more quickly as compared to those individuals taking many types of antiretroviral drugs simultaneously. Multiple drugs targeting reverse transcriptase and protease make it more difficult for HIV to develop resistance. Of course, the disadvantage would be that multidrug-resistant strains of HIV develop over time.

Suggested hypotheses:

Students should propose ideas related to the development of mutation in protease over time that would decrease the affinity of the protease inhibitor yet still ensuring the enzyme works. Thus the active amino acids between 25 and 27 would be conserved over time, but other amino acids involved in binding ritonavir would change over time.

Analysis questions:

Sample alignment:

	10	20	30	40	50	60
ref_NL43	PQITLWQRPL	VTIKIGGQLK	EALLDTGADD	TVLEEMNLPG	RWKPKMIGGI	GGFIKVRQYD
patientA_day1	-----	-----	-----	-----S---	-----	-----
patientA_day175	-----	-----	-----	-----S---	-----	-----V-----
patientA_day399	-----	-----	-----	--FQ--S---	-----	-----V-----
	70	80	90			
ref_NL43	QILIEICGHK	AIGTVLVGPT	PVNIIGRNLL	TQIGCTLNF		
patientA_day1	-VP-----	-----I---	-----	--L-----		
patientA_day175	--P-----	-----I---	-T-----	--L-----S		
patientA_day399	--P-----	V-----I---	-A-----M	--L-----		

1. What are your observations and conclusions from the sequence alignment data? Why would some amino acids be conserved and others change over time? Do the data support or refute your proposed hypotheses?

Thinking back to the important amino acids in the active site, positions 25–27 (Asp–Thr–Gly) are conserved throughout the entire sampling time. These positions are conserved, as these are the important amino acids involved in the enzyme's function as an aspartyl protease. However, amino acid 82 does change over time suggesting that this amino acid is involved in the binding of the inhibitor, ritonavir. Thus position 82 changes over time, providing evidence of drug resistance. By day 399, ritonavir may not bind to protease as efficiently because of changes at position 82 (and probably position 54 and 71), but the enzyme is still active because of the conservation of amino acids 25–27.

2. What are the advantages and disadvantages of inhibiting protease function? Why would inhibiting integrase be considered a more effective approach to block HIV replication?

The advantages would be that inhibiting protease blocks the cycle of HIV replication. Developing specific inhibitors just for HIV protease would directly target this enzyme and stop its function and production of new HIV virions. This would decrease the amount of virus in the host. The disadvantage is that protease is similar to other human protease enzymes. Thus, taking inhibitors commonly results in side effects caused by inhibiting other human enzymes needed for normal functioning. Inhibiting integrase would be considered more effective, as it would prevent the integration of HIV into the host's genome. Preventing integrase activity would lead to a decrease in chronic HIV infections.

3. If you are a researcher investigating HIV, propose two additional molecular approaches to inhibit HIV replication and infection.

Inhibiting any of the three HIV enzymes would be key targets. Inhibiting the binding of HIV to the host immune cell is also another approach. These approaches would block the envelope protein (gp120), CD4, and the two coreceptors, CCR5 and CXCR4. Other approaches would include prevention via a vaccine. Vaccines based on gp120 and other HIV antigens are being investigated.

>patientA\_day399

PQITLWQRPLVTIKIGGQLKEALLDTGADDTVLEEM  
SLPGRWPKPMIGGIGGFVKVRQYDQIPIEICGHKVI  
GTVLIGPTPANIIGRNLMTQLGCTLNF

>patientA\_day175

PQITLWQRPLVTIKIGGQLKEALLDTGADDTVLEEM  
SLPGRWPKPMIGGIGGFVKVRQYDQIPIEICGHKAI  
GTVLIGPTPTNIIGRNLLTQLGCTLNS

>patientA\_day1

PQITLWQRPLVTIKIGGQLKEALLDTGADDTVLEEM  
SLPGRWPKPMIGGIGGFVKVRQYDQVPIEICGHKAI  
GTVLIGPTPVNIIGRNLLTQLGCTLNF

>ref\_NL43

PQITLWQRPLVTIKIGGQLKEALLDTGADDTVLEEM  
NLPGRWPKPMIGGIGGFVKVRQYDQILIEICGHKAIG  
TVLVGPTPVNIIGRNLLTQIGCTLNF

**Acknowledgements** This work was supported by the Seeding Postdoctoral Innovators in Research and Education (SPIRE) program, grant K12GM000678 from the National Institute of General Medical Sciences, a division of the National Institutes of Health.

## References

- Boden D, Markowitz M. Resistance to human immunodeficiency virus type 1 protease inhibitors. *Antimicrob Agents Chemother* 1998;42(11):2775–83.
- Resch W, Parkin N, Watkins T, Harris J, Swanstrom R. Evolution of human immunodeficiency virus type 1 protease genotypes and phenotypes in vivo under selective pressure of the protease inhibitor ritonavir. *J Virol* 2005;79(16):10638–49.