CURRICULUM ARTICLE

Molecular Evolution: The HIV Envelope Protein

Brian Rybarczyk

Student Activity

Background

Human immunodeficiency virus (HIV) particles, a.k.a. virions, can exist in the bloodstream of an infected individual. However, HIV replication or production of more HIV virions occurs within cells of the lymph nodes of an infected individual. HIV is capable of infecting multiple human cell types, but for simplicity's sake, we will focus on HIV infection of CD4-positive immune cells, a.k.a. T helper cells.

The HIV envelope protein (Env) is essential for binding to a human lymphocyte or T cell. Env is composed of two subunits, a transmembrane protein called gp41 and a surface protein gp120, which are attached to one another noncovalently (Fig. 1). This dimer associates with two other dimers forming the trimer that binds to CD4 receptors on the surface of the lymphocyte. CD4 receptors usually act as adhesion molecules to bind to other immune cells such as antigen-presenting cells, but HIV takes advantage of this receptor and uses it to bind to the T cell. A third type of receptor involved in this interaction, called a coreceptor, is also required for HIV infection. This coreceptor is located on the T cell. HIV also needs either the coreceptor CCR5 or

the coreceptor CXCR4 to enter the cell. Different types of T cells will express CCR5 or CXCR4 on their cell surface. Regions within the gp120 subunit of the envelope protein interact with both CD4 and the coreceptor, causing conformational changes in the envelope protein. This initiates the fusion of the HIV and the T cell lipid membranes, ultimately leading to the insertion of HIV's RNA genome into the host cell, subsequent copying of the RNA genome into double-stranded DNA, integration into the host's genome, and production of new virus particles.

Scientists have found that the fitness of an HIV virion is directly correlated with how fast it can infect a cell, so they are very interested in studying the evolution of gp120 in particular. The gp120 domain is important because it is a target for our immune system. The gp120 protein is composed of five hypervariable regions separated by constant regions.

The V1/V2 region is involved in modulating the interaction of the envelope receptor with CD4. Both the V1 region and V2 region consist of approximately 40 amino acids each. The V3 region is important for modulating interactions with the coreceptors CCR5 and CXCR4. Little is known about the function of V4/V5, but the region has been shown to contain epitopes for neutralizing antibodies.

The simian immunodeficiency virus (SIV) is very similar to HIV. In fact, HIV evolved from SIV. SIV is endogenous to nonhuman primate populations (chimpanzees, macaques, etc.). The *Rhesus macaque* serves as a model organism to study SIV infection and pathogenesis. There are many similarities between how SIV infects macaques and how HIV infects humans, although with some differences. Thus, scientists can learn much from experimenting with the SIV–macaque model to further determine molecular events during HIV infection of humans.

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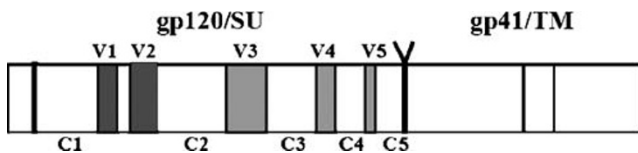Chapel Hill, NC 27599, USA
e-mail: brybar@unc.edu

Fig. 1 Envelope protein linear schematic. Constant regions are represented by C1, C2... and variable regions are represented by V1, V2.... *SU* surface domain, *TM* transmembrane domain, which anchors the protein into a lipid membrane. Other domains within gp41 are not indicated on this diagram

Preactivity Questions

1. If a polypeptide contains 40 amino acids, how many nucleotides does this represent?
2. Briefly describe how HIV evolution occurs.
3. What are the specific functions of the V1/V2 region of the envelope protein? What about the V3 region?

Envelope Sequence Analysis Activity

The Experiment

Ten *Rhesus macaques* were infected with the SIV via intravenous injection and ten macaques were infected via a mucosal route of infection, mimicking sexual transmission. Plasma samples were taken at time points 2, 4, 6, 8, 12, 24, and 32 weeks after infection. Virus RNA was isolated from the samples, RT-PCR was performed to specifically amplify the V1/V2 region, and the products were visualized using gel electrophoresis and sequenced.

Hypotheses

* Similar amino acid changes occur in the V1/V2 region comparing macaques infected via the two routes of infection, intravenous and mucosal.
* However, the route of infection, how the virus enters the host, will affect the *rate* of V1/V2 evolution.

Your goal is to describe the changes that occur in the V1/V2 hypervariable regions of the gp120 protein

during virus infection and determine if the data supports or refutes the hypotheses above.

Data

A) Table 1 indicates the number of SIV V1/V2 variants present in the sample at each time point after infection from three macaques infected through the intravenous route (IV) and three macaques infected through the mucosal (M) route.
B) Amino acid sequences obtained from the samples are provided in the file SIV_V1_V2 sequences.txt.

Each sequence is given a unique name. For example, the sequence name JJW_IV_wk2_variant1 indicates the name of the macaque subject (JJW) followed by the mode of infection (IV, intravenous or M, mucosal), time point after infection (week 2), and the variant found (clone 1).

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 later.

   * Click the **Protein Tools** button since the data is amino acid sequences.
   * Click **Add** button. Browse your desktop to upload the sequence file SIV_V1_V2 sequences.txt into your session. Click the **Upload File** button, then click the **Save** button at the bottom of your sequence text boxes.

Table 1 The number of SIV V1/V2 variants present in the sample at each time point after infection

Macaque	Route of infection	Time point after infection (weeks)						
		2 weeks	4 weeks	6 weeks	8 weeks	12 weeks	24 weeks	32 weeks
JJW	IV	2	2	2	2	8	7	6
PDW	IV	1	3	3	6	9	7	2
TDB	IV	3	3	3	4	11	9	8
AU86	M	2	2	2	2	2	4	7
AV87	M	4	4	4	4	4	7	8
AV95	M	1	1	1	1	1	2	4

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 the sequence name.
- Select appropriate alignment algorithm (CLUS-TALW–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_smh_4" in each of your alignments as the wild type comparison sequence.

Analysis Questions

1. What are your observations and conclusions from the data in Table 1? Does the data support or refute the hypotheses proposed?
2. What are your observations and conclusions from the sequence analysis? Does the data support or refute the hypotheses proposed?
3. Propose reasons why there are similarities and differences in both the timing and the specific amino acid changes that occur in HIV after infection in these two groups of subjects.
4. How is natural selection and microevolution illustrated in this experiment? In other words, identify the genetic basis for evolutionary change, possible selective pressures, and the resulting adaptation illustrated in the experiment. What happens to the HIV mutants that are able to escape detection by the immune system?
5. Would the envelope protein serve as an effective component of an anti-HIV vaccine? Why *and* why not?

Instructor Guide

Introduction

The human immunodeficiency virus (HIV) has become a major health epidemic and currently, there are nearly 40 million people worldwide affected by HIV/AIDS (UNAIDS/WHO, http://www.unaids.org). Basic scientific research has uncovered many facets to the biology of the virus, including how HIV evolves at a molecular level. This module focuses on the envelope protein, a receptor on the surface of the virus. The activity will provide students a hands-on experience working with actual scientific data to elucidate the molecular evolution of the virus during the infection of a host.

Learning Objectives

- Describe the process of microevolution of the envelope protein.
- Propose possible selective pressures involved in virus evolution.
- Analyze molecular evidence for evolution of the SIV envelope protein.
- Propose ideas of why the envelope protein may be an effective or ineffective component of an anti-HIV vaccine.

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 life cycle 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 from SIV 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 "SIV_V1_V2 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/wild-type SIV sequence is provided in the data named "ref_smh_4"
- Sequence alignment and data collection can be performed in a 50-min period followed by a 30-min debriefing session to discuss the analysis questions.

Answer Key

Preactivity Questions

1. If a polypeptide sequence has 40 amino acids, how many nucleotides does this represent?

   Each amino acid is encoded by 1 codon, 1 codon is composed of 3 nucleotides, then 40 amino acids×3 nucleotides=120 nucleotides.

2. Briefly describe the steps involved in how HIV evolution occurs.

   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.

3. What are the specific functions of the V1/V2 region of the envelope protein? What about the V3 region?

   V1/V2 modulates interactions with CD4 and it is an antigenic region containing epitopes where antibodies bind. The V3 region modulates interactions with CCR5 and CXCR4 coreceptors.

Analysis Questions

1. What are your observations and conclusions from the data in Table 1? Does the data support or refute the hypotheses proposed?

   The table shows that in general, for the macaque subjects infected via IV, the number of variants in each subject quickly increases between weeks 8 and 12 after infection. For macaque subjects infected via the mucosal route, the number of variants increases much later, at 24 weeks. Thus, this data supports the hypothesis that the route of infection does affect the rate of SIV evolution.

2. What are your observations and conclusions from the sequence analysis? Does the data support or refute the hypotheses proposed?

   Students should do multiple comparisons between the early time point and later time point within one subject (i.e., JJW_IV_wk2 and JJW_IV_wk12), and between the two subjects. Many more changes in the V1 region compared with the V2 region are evident. In JJW, these changes occur earlier compared to macaque AV95. There are no consistent sequence changes between JJW and AV95, although there are hot spots where amino acid changes occur more frequently in both subjects, especially in the V1 region (see green highlighted amino acids in the alignment below).

3. Propose reasons why there are similarities and differences in both the timing and the specific amino acid changes that occur in HIV after infection in these two subjects.

   Students should note that the time when sequence changes are evident occurs much earlier in the IV-infected macaques (12 weeks) compared to the mucosal-infected macaques. Possible reasons could be that there are different selective pressures influencing the timing of when the sequence changes occur in the two groups. Since the two groups were inoculated with the virus via different routes of transmission, the microenvironment of the initial establishment of infection could play a role in how and when the virus evolves. These microenvironments can also include different initial immune responses, for example, different antibody subtypes predominant in different parts of the host (IgA antibodies are more common near mucosal surfaces).

4. How is natural selection and microevolution illustrated in this experiment? In other words, identify the genetic basis for evolutionary change, possible selective pressures, and the resulting adaptation illustrated in the experiment. What happens to the HIV mutants that are able to escape detection by the immune system?

   The genetic basis for evolutionary change is the mutations that arise during the virus life cycle when its genome is being replicated. Selective pressures may include host T cell and B cell responses, specific subtype antibody responses, the microenvironment present when the virus enters the host, and the general state of the immune system during infection. The resulting adaptation is the development of variants which are not recognized by the immune system. The variants have a higher fitness compared to wild-type virus and, therefore, will go on to replicate and persist in the host.

5. Would the envelope protein serve as an effective component of an anti-HIV vaccine? Why *and* why not?

Researchers have logically chosen the envelope protein as a vaccine candidate since it is an outer surface, membrane protein easily accessible to the host's immune system. It also has constant regions which are generally not altered over the course of infection.

However, if this protein evolves over time, specifically in the hypervariable regions like V1/V2, it is a moving target for the immune system and, thus, does not make for an effective vaccine. For example, if the host generates antibodies which specifically recognize an epitope within the V1/V2 region at week 2, by week 12,

this same region could have evolved to escape antibody detection and the antibodies generated at week 2 will no longer recognize the epitope at week 12. Therefore, vaccines based solely on the envelope protein have not been shown to be 100% effective. The constant regions of the protein which do not evolve would be more effective antigenic components, while the hypervariable regions would not. Changes in glycosylation sites of gp120 can also modulate the antigenicity of HIV and, thus, can prevent antibodies from binding to certain regions of the protein allowing virus variants to go undetected by the immune system.

Sample Alignment

References

Centers for Disease Control and Prevention (CDC): http://www.cdc.gov/hiv/.

International AIDS Vaccine Initiative (IAVI): http://www.iavi.org.

World Health Organization (WHO) http://www.who.int/en/.

Rybarczyk BJ, Montefiori D, Johnson PR, West A, Johnston RE, Swanstrom R. Correlation between env V1/V2 region diversification and neutralizing antibodies during primary infection by simian immunodeficiency virus sm in rhesus macaques. J Virol 2004;78:73561–71.