



Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.

# The Fundamentals of Human Virology

JEFFREY WILUSZ

*Department of Microbiology, Immunology and Pathology, Colorado State University, Fort Collins, Colorado*

---

This chapter provides an overview of basic virology along with a primer on the numerous viruses that cause significant human disease. Viruses are grouped in terms of their genomic organization in order to facilitate a discussion of both the molecular and pathogenic aspects of their biology. In addition to increasing our understanding of mechanisms of pathogenesis and rational drug design, the detailed molecular characterization of these pathogens has provided the potential to identify molecular signatures that can be used to trace outbreaks of viral disease.

## INTRODUCTION

Viruses were established as agents that cause human disease at the beginning of the 20th century. Their small size (approximately 2- to 60-fold smaller than a standard Gram-positive *staphylococcus* or *streptococcus* bacterium) gave viruses the distinct property of being able to pass through the conventional filters of the day, allowing their identification as “filterable agents.” Today viruses have been identified that affect every kingdom—animals, plants, and bacteria. These ubiquitous agents therefore have the potential to influence the entire biota of the planet. While this chapter will focus on human viruses, the general principles discussed can be applied to other virus families.

Viruses are molecular pathogens. They possess no metabolism of their own and can be thought of as molecular parasites. A conventional virus is made up of two or three major components. A nucleic acid genome, which can be DNA or RNA (single- or double-stranded, contiguous or segmented) contains all of the genetic information and is responsible for encoding all of the

virus-specific macromolecules of the pathogen. Due to the aggressive application of molecular biological techniques, the sequences of many viruses are known and fully annotated. This nucleic acid genome is packaged in a protein shell called a capsid. The capsid proteins generally self-assemble to form the shell, which takes on an icosahedral, helical, or complex symmetry depending on the virus family. The functions of this capsid include packaging of the genome, protecting the genome from environmental insults, and effectively delivering the nucleic acid to the inside of living cells. Most RNA viruses and some DNA viruses are surrounded by a lipid envelope that is derived from the host cell. The membranes of enveloped viruses all contain viral-specific glycoproteins that aid in viral tropism.

It is interesting to note that the conventional virus described above is not the simplest molecular pathogen. RNA-only agents called viroids exist that can kill plants without ever making a protein. In addition, protein-only pathogens called prions “replicate” in the absence of a nucleic acid genome by inducing a conformational change in other normal prp proteins. Prions cause degenerative neurological diseases called spongiform encephalopathies in humans. The bottom line is that it is truly a molecular jungle out there.

There are eight basic steps in a viral infection of a eukaryotic cell. The first is adsorption of the virion to the cell surface through an interaction of viral proteins with specific cellular receptors and, in some cases, coreceptor molecules. Viruses can specifically and tightly interact with proteins (e.g., HIV and CD4), carbohydrates (e.g., influenza virus and sialic acid), and lipids (e.g., B19 parvovirus and globoside). This virus-receptor interaction in large part determines the tropism of many viruses. Preventing this interaction is the goal of neutralizing antibodies that are generated by the immune system. The second step is viral penetration of the cell surface. This can be done by direct membrane fusion or endocytosis. The third step is uncoating of the viral genome inside the cell. In many cases, uncoating is due to pH changes that occur inside endocytotic vesicles that result in structural rearrangements of viral surface proteins. The fourth step is primary transcription/gene expression. Many viruses are designed to actively express their genomes immediately upon uncoating. This is accomplished via several strategies, including the packaging of viral polymerases, packaging or production of strong transcriptional transactivation proteins, and the simple fact that some viral genomes can serve directly as mRNAs. Step five is replication of the viral genome. Since viral-encoded proteins play a key role in this step for many viruses, this step has been a prime target for the development of antivirals. The next step is called “secondary transcription,” or gene expression that occurs off of progeny genomes. This is an important step that viruses use to amplify their gene expression to obtain maximal yields of viral progeny. It should also be noted that many viruses make different sets of proteins at early and late times postinfection. Early proteins generally include replication and transcription factors

as well as proteins that allow the virus to usurp cellular metabolism. Late protein production focuses on virion structural components. The seventh step is the packaging of progeny viral genomes. Viral capsids in most cases self-assemble, and the genomic nucleic acid must be properly inserted. The final step is the release of progeny virions from the cell. For enveloped viruses, this involves budding and the acquisition of a lipid bilayer membrane.

There are several ways to determine whether a cell is infected with a virus. First, many viral infections elicit cytopathic effects such as the rounding of cells, shrinkage, aggregation, lysis, or cell-cell fusion. These effects are due in large part to the expression of specific viral proteins or the shut-down of host cell macromolecular synthesis that occurs during the infection. Second, focal points of viral replication and assembly can sometimes be observed. These are referred to as inclusion bodies; a classic example is the detection of Negri bodies (cytoplasmic acidophilic inclusion bodies) in rabies virus-infected samples. Third, viral proteins expressed on the surface of an infected cell can sometimes bind many red blood cells, a phenomenon called “hemadsorption” that can be very striking when observed in a light microscope. This technique can be very useful if a viral infection shows very little cytopathic effect. The fourth way to directly detect viral-specific gene products is through the use of antibodies in western blots, immunofluorescence, or enzyme-linked immunosorbent assays (ELISAs). Finally, polymerase chain reaction (PCR) with virus-specific primers can be used to directly identify viral-infected tissue. The use of PCR-based assays as a powerful molecular epidemiologic tool is discussed further below.

The quantitation of virus in a sample can be approached using either physical- or biological-based assays. Physical methods such as particle counts using an electron microscope, hemagglutination assays, and ELISAs/(RIAs) can provide an approximation of the total number of viral particles in a sample, but do not address the fact that for many viruses, the particle-to-biologically active particle ratio is rather low. Therefore biological assays that include serial dilutions in conjunction with plaque assays, focus-forming assays, or determination of the infectious dose needed to kill 50% of a target laboratory animal test group are often considered the gold standard.

There are several key definitions and unique aspects of viral genetics that should be stressed. First, a field or street isolate of a virus is obtained directly from the natural host. Viruses that are passaged in cell culture often adapt to growth under *in vitro* conditions and may lose or gain some properties relative to the field isolate. Second, co-infection with two or more viruses that contain a segmented genome can result in a reassortment of segments in the progeny virions. This can result in dramatic changes in the biologic and/or antigenic properties of the new viruses. Third, while DNA viruses have a mutation rate of  $10^{-8}$  to  $10^{-11}$  per nucleotide incorporated, RNA viruses have a million-fold greater mutation rate ( $10^{-3}$  to  $10^{-4}$ ). This is due to the presence

of error-prone polymerases that lack the ability to proofread. Fourth, phenotypic mixing or pseudotype formation can occur in an infection by two different viruses when progeny viruses are produced that contain the capsid of virus "A" surrounding the nucleic acid genome of virus "B." In other words, phenotypic mixing is the mixing of nonnucleic acid components between two viruses. Finally, many viruses generate defective progeny viruses as a normal part of their life cycle. While this is one reason for the low particle-to-plaque-forming-unit ratios that are observed in viral preparations, the production of defective particles can have biological consequences as well. It is well established for several RNA viruses that defective particles are naturally produced that interfere with the replication of the wild-type virus. These defective interfering particles moderate the course of an infection and can support the establishment of persistent infections.

One of the best ways to control viral infections in a susceptible population is to prevent them from occurring through the judicious use of vaccines. Vaccines have reduced the incidence of measles, mumps, and rubella by >99.7% in the U.S., and have completely eliminated natural transmission of smallpox and poliovirus. The three main types of vaccines currently in use are killed (i.e., formalin-fixed viruses), live attenuated vaccine strains, and subunit vaccines made from recombinant proteins. Additional types of vaccines, including DNA-based vaccines and use of single vectors for broad vaccination, may well be on the horizon.

Historically, the availability of effective antiviral drugs has lagged behind that of other antimicrobials. However due to advances in molecular virology and rational drug design, the arsenal of antiviral compounds is growing quickly. The prime targets of available antivirals include uncoating (e.g., amantadine prevents intraviral pH changes in influenza virions), viral polymerases to prevent replication (e.g., AZT, ribavirin, and acyclovir), viral proteases to prevent maturation of virions, neuraminidases to prevent viral release from cells, and cellular enzymes involved in protein synthesis (interferons). The application of antibodies for the treatment or prevention of viral diseases is also becoming more common. In summary, the antiviral arena is definitely an evolving field.

The International Committee on the Taxonomy of Viruses has classified the major human viral pathogens into approximately 21 families. Since more viruses remain to be discovered and the relationships of individual viruses to human disease continue to expand, this number is likely an underestimate. Viral classifications are based on a variety of factors. First is morphology—the size of the particle, its shape, and the presence or absence of membranes. Other factors include the properties of the viral genome, viral genomic organization and expression patterns, antigenic consideration, and biological relationships such as host range, etc.

## AN OVERVIEW OF THE DNA VIRUSES

Of the ~21 families of viruses that cause common human diseases, only six possess a DNA genome. Consistent with the location of cellular enzymes that act on DNA, all of these DNA viruses have a nuclear life cycle, with the exception of one. The Poxviruses remain cytoplasmic during infection, likely because they are too large to fit through the nuclear pore. Of the nuclear DNA viruses, only the herpesviruses are enveloped. Finally, epidemiological studies suggest that infection by many of the DNA viruses is very common in the human population. An overview of the six families of DNA viruses based on their molecular properties and life cycles is presented below.

Parvoviruses contain a small but unique 5–6-kb single-stranded DNA genome. The key human pathogenic parvovirus is called B19, which specifically infects red blood cell precursor cells via the P antigen as its receptor. This results in a transient, usually subclinical anemia that can be rather severe in hemolytic anemia patients (i.e., sickle cell disease) or the immune-compromised. Several weeks later, an immune-mediated rash (erythema infectiosum or fifth disease) or arthritis-like symptoms develop. Another parvovirus that is worth mentioning due to its utility in gene therapy is adeno-associated virus (AAV). As its name implies, the virus is dependent on co-infection with another DNA virus for efficient replication. Due to the facts that AAV causes no recognized human disease and integrates at a predominant location in the cellular genome, it is currently under development as a means of therapeutic gene delivery to human cells. In addition to therapeutic uses, the possibility clearly exists that AAV can be engineered as a biowarfare agent as well (i.e., potent toxin genes can be inserted into its genome and effectively delivered to human cells).

Adenoviruses were discovered about 50 years ago and generally cause respiratory and ocular infections. Two of the approximately 50 serotypes (types 40 and 41) are a major agent of gastrointestinal distress. Focal outbreaks of adenoviral-induced disease (e.g., pharyngoconjunctival fever associated with contaminated swimming pools or ponds) can occur in the U.S. The plethora of serotypes makes vaccination impractical given current vaccine technology. Guttated adenovirus vectors are also under intense development as a delivery system for gene therapy. As noted above for AAV, adenoviral vectors can also be used to potentially deliver detrimental genes as well as therapeutic ones.

There are eight major human herpesviruses, seven of which cause defined human disease. Individuals will maintain herpesvirus infections in a latent form following the initial acute infection. Reactivation can occur under a variety of circumstances. Herpes simplex viruses type I and II cause ulcerative lesions of the skin and remain latent in nerve ganglia. Cold sores are due to

herpes simplex virus reactivation from the trigeminal ganglia. Varicella zoster virus causes chicken pox during primary infection and zoster or shingles upon reactivation. Human cytomegalovirus is a major cause of congenital infections that result in a variety of symptoms including retardation and hearing loss. Epstein Barr virus (EBV) targets B cells and causes infectious mononucleosis in young adults. In addition, EBV infection is associated with cancers in some populations (Burkitt's lymphoma in Africa and nasopharyngeal carcinoma in Asia). Human herpes virus type 6 causes the common childhood rash roseola (exanthem subitum). HHV8 or Kaposi's Sarcoma Herpes Virus is a sexually transmitted agent that causes kaposi lesions in the immunosuppressed. In addition to vaccines for VZV, herpesvirus infections can be controlled by a highly effective antiviral acyclovir and its derivatives. The drug specifically targets the viral thymidine kinase gene and prevents replication during active infections. The widespread nature of most herpesviruses makes suspicious outbreaks less likely. It should be noted, however, that due to the sexually transmitted nature of HHV8, kaposi sarcoma was one of the defining lesions on homosexual AIDS patients noted early on in the HIV epidemic.

Papovaviruses contain a unique circular double-stranded DNA genome. The family consists of two human polyoma viruses, BK and JC, which are rather ubiquitous agents that often cause inapparent primary infections. Reactivation of these agents by immunosuppression, however, is associated with cystitis for BK and a severe neurological disorder, progressive multifocal leukoencephalopathy, for JC. The broad array of wart-causing agents, the human papillomaviruses (HPV), are also members of this family. In addition to warts and condylomas, it is important to note that the vast majority of cervical dysplasias observed in PAP smears are due to HPV infection. Infections by certain HPV subtypes, i.e. HPV16 and HPV18, are strongly associated with the development of cervical cancers and other malignancies.

The Hepadnavirus family contains a one key member Hepatitis B virus (HBV). HBV is a unique agent that replicates using an RNA intermediate and reverse transcriptase. It is a leading cause of chronic hepatitis outside of the U.S. An effective recombinant subunit-based vaccine is currently available.

Poxviruses are exceptionally large enveloped virions with a unique complex viral capsid structure. Their entirely cytoplasmic life cycle is unique among the human DNA virus families. The key human pathogens in this group are variola (smallpox) and mulluscum contagiosum (which causes wart-like lesions). Smallpox was eradicated many years ago due to the effective application of the smallpox vaccine (vaccinia virus) and the fact that humans are the only natural reservoir of the agent. The risk of smallpox as a biowarfare agent has received much recent attention due to the declining or total lack of immunity currently seen in the population. Finally, the 2003 outbreak of monkeypox, a nonfatal rash disease, in the upper Midwest that was previously

unseen in the western hemisphere serves as a reminder that other zoonotic viruses exist that can emerge in the human population.

## AN OVERVIEW OF THE RNA VIRUSES

Fifteen families of the 21 virus families that cause major human diseases contain an RNA genome. Four useful generalities can be applied to virus families that contain RNA genomes. All of the RNA virus families maintain an exclusively cytoplasmic life cycle, except for three [retroviruses (which go through a DNA intermediate) and the two RNA viruses that use the cellular splicing machinery (orthomyxo- and bornaviruses)]. All but four of the RNA viruses are enveloped—only the picornaviruses, astroviruses, caliciviruses, and reoviruses contain naked capsids. With the exception of the double-stranded RNA genomes of the reovirus family, all RNA virus families contain single-stranded genomes. These single-stranded RNA genomes can be of mRNA sense (i.e., positive sense), complimentary to mRNA (negative sense), or a combination of both (ambisense). Finally, most RNA viruses have nonsegmented genomes, with the exception of three families—orthomyxoviruses, bunyaviruses, and arenaviruses. An overview of the 15 families of RNA viruses based on their molecular properties and life cycles is presented below.

## POSITIVE-SENSE RNA VIRUSES

There are six families of RNA viruses that contain a single-stranded genomic RNA molecule of message-sense. The genomic nucleic acid of positive-sense RNA viruses is fully infectious when delivered to cells in the absence of viral proteins. The gene expression strategies of these agents fall into three groups. The least complex is illustrated by the picornaviruses and flaviviruses. These agents have a single large open reading frame that encodes a single large polyprotein that is subsequently cleaved in an ordered fashion by proteases to generate individual functional polypeptides. Cap-independent translation was discovered in picornaviral internal ribosome entry site elements located in the 5' noncoding region of the genome. Togaviruses, astroviruses, and caliciviruses generate two or more individual polyproteins that are encoded by the genome, as well as subgenomic mRNAs that are generated during infection. Coronaviruses use the most complex strategy. Instead of generating a polyprotein, numerous subgenomic mRNAs are made from their large genomes, each encoding an individual protein. Recombinant togaviruses that encode subgenomic mRNAs capable of producing large amounts of proteins of interest have been developed. Therefore in addition to naturally occurring positive-sense



RNA viruses, the possibility that designer viruses may be responsible for suspicious outbreaks cannot be overlooked.

Picornaviruses are a family of small stable viruses, most of which can survive the harsh environment of the gut and be recovered from stool samples. There are five key genera of picornaviruses. Poliovirus, a rather ineffective pathogen that can infect the central nervous system (CNS) and cause poliomyelitis, is eradicated in many parts of the world due to the use of highly effective killed and attenuated vaccines. Select members of the 30+ serotypes of echoviruses are among the most commonly isolated viruses from stools, and cause a variety of disease symptoms including meningitis, encephalitis, rashes, diarrhea, upper respiratory infections, and conjunctivitis. Coxsackieviruses (29 serotypes) are also rather commonly isolated from humans and cause hand, foot, and mouth disease (which is distinct from the animal virus that wreaked havoc on British livestock in 2000), herpangina, pleurodynia, and myocarditis in addition to a variety of nonspecific symptoms. Hepatitis A virus, the agent that causes acute infectious hepatitis, is also a member of the picornavirus family. The final branch of the human picornavirus family tree, rhinovirus, contains over 100 serotypes and causes the common cold. The biology of rhinoviruses is unique from other picornaviruses in that the virions are not stable to low pH, and the agent prefers growing at lower temperatures (33°C vs. 37°C). These properties are reflected in the restriction of rhinovirus infections to the upper respiratory tract.

Flaviviruses are enveloped viruses with icosahedral particles. There are six key members of this family, four of which are transmitted to humans via arthropod vectors. Yellow fever virus, the prototype of the group, was the first human virus isolated. It causes an *Aedes* mosquito-borne severe fever and myalgia, followed by jaundice and vomiting that can be prevented through the use of a live attenuated vaccine. The four serotypes of Dengue fever virus are also transmitted by *Aedes* species and cause a fever that can acquire hemorrhagic manifestations and be extremely severe. Dengue is probably the most significant mosquito-borne viral disease worldwide, with 50–100 million cases per year. St. Louis encephalitis virus is associated with *Culex*-borne epidemic encephalitis in the U.S. The last of the arthropod-borne flaviviruses, West Nile virus, is a key emerging pathogen in the U.S. since 1999. There are two hepatitis agents—hepatitis C virus (HCV) and hepatitis G virus (HGV)—that are classified as Flaviviridae that are not passed to human via an insect route. HCV is a major cause of chronic hepatitis in the U.S.

The four families of positive-sense RNA viruses that make subgenomic mRNAs contain species that cause a variety of human diseases. Key members of the Togaviridae include a variety of arthropod-borne encephalitis agents (Eastern equine encephalitis virus, Western equine encephalitis virus, and Venezuelan equine encephalitis virus) along with rubella virus (RV). RV is the

agent that causes the fever and rash associated with German measles. More importantly, the virus can cause severe fetal abnormalities if it crosses the placenta. RV is currently under control in the U.S. due to the wide use of an attenuated viral vaccine given as part of the MMR vaccination. The Astroviridae and Caliciviridae are difficult to culture and are associated with gastroenteritis outbreaks. In addition, the hepatitis agent HEV (hepatitis E virus) is classified as a calicivirus. The final family of viruses in this group—the Coronaviruses, contains exceptionally large genomes (~27 kb) in a helical nucleocapsid and are associated with common colds. The agent responsible for the SARS outbreak in 2003 is a member of the Coronaviridae.

## NEGATIVE-SENSE SINGLE-STRANDED RNA VIRUSES

All of the viruses in the four families of human pathogens classified in this group contain helical nucleocapsids within an enveloped particle. There are three general factors that govern gene expression in these viruses. First, the viral polymerase initiates transcription only at one end of the viral genome and reinitiates with approximately 50% efficiency following termination at the end of each gene. This property makes gene order an important means by which the relative levels of proteins are regulated. Second, the decision by the polymerase whether to transcribe mRNAs or replicate the entire genome is determined by the relative levels of the viral nucleocapsid protein. If levels are sufficient to package the nascent transcript, the polymerase ignores termination signals and proceeds copying the RNA template. Finally, cotranscriptional editing—the insertion of additional uncoded nucleotides by the viral polymerase—occurs at a select homopolymer run in all paramyxovirus genomes. This property is probably related to the tendency of the polymerase to stutter on poly(U) stretches to generate the poly(A) tail. In addition to the natural human viruses outlined below, it should also be noted that the technology to create recombinant negative-sense viruses has been developed.

The four families of negative-sense, single-stranded RNA viruses contain several classical and emerging human viruses of notoriety. The first family, the Paramyxoviridae, contains four major human viruses that all initiate infection in the upper respiratory tract. Measles and mumps viruses cause once-common childhood maladies (fever/rash and parotitis syndromes, respectively) that are preventable by live attenuated vaccines given as part of the MMR series. The four serotypes of human parainfluenza virus are responsible for numerous seasonal respiratory infections in children, most notably croup. The last member of the paramyxoviruses, respiratory syncytial virus, is the number one respiratory pathogen for children in terms of disease severity. No vaccine is

currently available to prevent infection by the last two members of the paramyxoviruses. Prophylactic doses of anti-RSV IgG is currently being given to reduce to the risk of respiratory syncytial virus infection in highly susceptible groups of children. The major member of the next family, the Rhabdoviridae, is rabies virus. This bullet-shaped virion is usually acquired through the bite of a rabid animal (most notably bats in recent years). Following replication at the site of infection, the virus travels to the CNS, where it invariably results in death. Because of the protracted time it takes for the virus to cross the neuromuscular junction to infect the nervous system, this is the only viral infection where post-exposure immunization is helpful. The Ebola virus is the prototypic member of the third family of negative-sense RNA viruses, the Filoviridae. As its name implies, this family consists of viruses with elongated nucleocapsids. Ebola causes outbreaks of severe hemorrhagic fever that is often fatal. Although the disease has not been seen in the U.S., several individuals in Reston, Virginia were infected by an Ebola strain with reduced virulence that was carried in a primate laboratory animal shipment. The bornaviridae make up the final family of this group. These viruses are unique among the negative-sense RNA viruses in that they replicate in the nucleus in order to utilize the cellular splicing machinery. It has been suggested that bornaviruses play a role in several neural pathologies, but their exact contribution to human disease is still being explored.

## SEGMENTED RNA VIRUSES AND THE RETROVIRIDAE

There are four families of human pathogenic RNA viruses that contain genomes consisting of two or more unique segments of nucleic acid. These genomes can take the form of negative-sense, ambisense, or double-stranded RNA. A key feature of these viruses is the rapid evolution that can occur due to reassortment of genomic segments in the progeny of infections by multiple viral species in the same cell. This is the reason that epidemiologists follow influenza virus very closely to insure that the current vaccine will provide coverage to any new variants that have recently emerged. One interesting molecular property of all of the single-stranded families in this group is that they “steal” the 5′ caps from cellular mRNAs to form the 5′ end of their own transcripts.

The segmented RNA virus families contain numerous important and interesting human pathogens. The Orthomyxoviridae consist of the enveloped negative-sense RNA viruses that cause influenza. Influenza A has eight segments. As noted above, due to reassortment, the nature of the two major surface antigens, hemagglutinin and neuraminidase, are closely monitored by the

U.S. Centers for Disease Control (CDC) in new influenza virus isolates. In addition to the killed influenza vaccine, infections can also be treated by antivirals that block uncoating (amantadine) and neuraminidase function. The Bunyaviridae family contains many arthropod-borne agents that cause fevers and encephalitis. These viruses have three genomic RNA segments that may be negative- or ambisense. In addition to these viruses, the rodent-borne hantavirus which emerged in the southwestern U.S. in 1993, causing numerous cases of severe pulmonary syndrome, is also a bunyavirus. The rodent-borne arenaviruses contain two ambisense genomic RNA segments and generally cause severe hemorrhagic fevers in humans (e.g., Lassa fever virus). The human members of the Reoviridae consist of 10–12 double-stranded RNA segments packaged in a double-walled protein shell. These viruses never fully uncoat their genome during infection, which helps avoid the action of double-strand activated enzymes of the interferon response. The major example of a human Reoviridae member is the common childhood diarrhea agent rotavirus. Over 80% of children in the U.S. will suffer diarrhea and vomiting due to infection by this agent.

The Retroviridae are included in this section not because they contain a segmented genome, but because they package two identical strands of positive-sense RNA. During infection, this genomic RNA is converted to DNA that then integrates into the cellular genome. Cellular DNA-dependent RNA polymerases are responsible for viral gene expression. The main examples of human retroviruses are the agent responsible for human acquired immune deficiency syndrome (AIDS: HIV) and human T-cell leukemia virus. Both of these agents are complex retroviruses that contain several regulatory genes (i.e., *tat* and *rev*) in addition to the *gag*, *pol*, and *env* genes of conventional retroviruses.

## THE APPLICATION OF MOLECULAR FORENSICS AND EPIDEMIOLOGY TO VIRAL INFECTIONS

Antigenic subtyping has been applied for many years to classify viral infections and outbreaks. Elaborate networks of surveillance teams, for example, are monitoring the antigenic characteristics and epidemiology of influenza virus strains isolated from infected patients, to determine the best vaccine cocktail to use in the coming year. The advent of rapid sequencing techniques has opened the door to obtaining detailed fingerprints of viruses that could provide important clues as to their source. In general, a selected region of the viral genome is sequenced from numerous isolates and subjected to

comparative phylogenetic analysis. Despite the increasing desire to rely on molecular analyses for all the answers, one should not overlook the fact that phenotypic features of the virus (tropisms, cytopathic effects, etc.) may provide important clues to facilitate identification.

As with other microbial species, the analysis of viral genomic segments can provide important clues as to the relatedness and origins of infections. Due to their small size and high rate of evolution, several considerations should be kept in mind when applying comparative molecular forensic analyses to viruses. First, for statistical reasons, it is usually advisable to look at as large a number of regions with variable sequence content as possible in the viruses that are isolated. Cost constraints, however, often make this impractical. Many viral genes, especially in RNA viruses, contain regions under rapid evolution and others that are under significant evolutionary constraint. In choosing a region of the viral genome to focus the analysis upon, it is important to consider that only rapidly evolving regions will provide enough useful variability when comparing agents from recently acquired infections. Second, if possible, it is extremely useful to include in the analysis a significant number of control viruses isolated from the surrounding population. This will allow for a full consideration of the background viruses in the local environment, and allow for a stronger statistical argument for relatedness to a predicted infection source. Finally, the overall strength of the argument for relatedness between two viruses requires detailed phylogenetic and statistical analyses that consider all alternative hypotheses. It will be important to get as close to “100%” certainty when one is performing the analysis for use in a legal rather than scientific context.

The tracking of HIV infections provides excellent examples of the successful application of molecular forensics to identify the source of a viral infection. In 1990, an HIV-positive dentist in Florida was suspected to be the source of HIV infection in six patients with no known risk factors. The sequencing of the *env* gene from viruses isolated from the doctor and patients strongly corroborated the epidemiological data that suggested the transmission route.<sup>1</sup> In more recent cases, polymerase chain reaction (PCR) amplification of the HIV genes followed by phylogenetic analyses have been used to suggest HIV transmission from a surgeon to a patient<sup>2</sup> and from a nurse to a patient.<sup>3</sup> Molecular forensics has been applied to other viral infections in addition to HIV. The nosocomial spread of specific strains of hepatitis C virus infections in hemodialysis units, for example, has been documented in several instances.<sup>4,5</sup>

The examples above involve molecular epidemiology of conventional viral pathogens. In order to rapidly identify viral agents involved in biocrimes, more work is needed to develop supportive resources. There is a clear need to establish an extensive sequence database of possible species. The identification of unique patterns and signatures beforehand will greatly facilitate the elabora-

tion of the strain and perhaps the source of the suspect agent. In the case of recombinant biowarfare agents, the bioengineered features along with the strain background should give reasonable clues to the source.

In closing, the goal of this chapter is to provide a background in the fundamentals of human virology and provide an overview of the utility and issues surrounding the use of molecular forensics and epidemiology to the world of virology. The rapid evolution of genomic technologies should continue to expand the capacity and impact of this exciting field. I hope that this chapter can be used as a stepping-stone to appreciate and give perspective to these advances.

## REFERENCES

1. Ou CY, Ciesielski CA, Myers G, Bandea CI, Luo CC, Korber BT, Mullins JL, Schochetman G, Berkelman RL, Economou AN, et al. Molecular epidemiology of HIV transmission in a dental practice. *Science* 1992;256:1165–1171.
2. Blanchard A, Ferris S, Chamaret S, Guetard D, Montagnier L. Molecular evidence for nosocomial transmission of human immunodeficiency virus from a surgeon to one of his patients. *J Virol* 1998;72:4537–4540.
3. Goujon CP, Schneider VM, Grofti J, Montigny J, Jeantils V, Astagneau P, Rozenbaum W, Lot F, Frocrain-Herchkovitch C, Delphin N, Le Gal F, Nicolas JC, Milinkovitch MC, Deny P. Phylogenetic analyses indicate an atypical nurse-to-patient transmission of human immunodeficiency virus type 1. *J Virol* 2000;74:2525–2532.
4. Seme K, Poljak M, Zuzec-Resek S, Debeljak M, Dovc P, Koren S. Molecular evidence for nosocomial spread of two different hepatitis C virus strains in one hemodialysis unit. *Nephron* 1997;77:273–278.
5. Halfon P, Roubicek C, Gerolami V, Quentin Y, Khiri H, Pepe G, Berland Y. Use of phylogenetic analysis of hepatitis C virus (HCV) hypervariable region 1 sequences to trace an outbreak of HCV in an autodialysis unit. *J Clin Micro* 2002;40:1541–1545.