CHAPTER 11

Physical and Chemical Manipulation of the Structural Components of Viruses

VIRAL STRUCTURAL PROTEINS

Although viruses are nucleic acid genomes surrounded by a capsid (and sometimes by membrane-associated viral proteins), a large number of other virus-encoded nucleic acids and proteins are expressed during infection of the host cell and eventual formation of new virus particles. If these nucleic acids and proteins do *not* end up in the structure of the virus itself, they are termed **nonstructural**. Thus, proteins involved in, for example, replication of herpesvirus DNA during its infection are nonstructural proteins. Indeed, during replication of a DNA virus, all the viral mRNA expressed and encoding viral proteins will be nonstructural components because this mRNA remains in the host cell when new virus particles form and exit the cell.

It is conceptually simple to differentiate structural and nonstructural proteins. Any protein found in purified virions (complete virus particles, i.e., genomes, capsid proteins, and any envelope and membrane-associated proteins *in* the virion) is structural. If a protein is viral encoded but not found in the virion, it is nonstructural. In practice, this differentiation can be somewhat difficult owing to problems with isolation of absolutely pure virus. Some enveloped viruses are almost impossible to isolate completely free of infected cellular debris or extracellular proteins. Many viruses have the irritating ability to include small amounts of cellular and viral material in their maturation that are not necessary for virus viability or replication.

The ability to isolate pure (or nearly pure) viral structural and nonstructural components is very important in research and medicine. Some of the uses for such material are as follows:

1 sources for preparation of pure immunological reagents such as monoclonal or polyclonal monospecific antibodies, as well as prophylactic vaccines;

2 enzymes that can be studied to develop specific antiviral drugs targeted against specific features of an enzyme's mechanism of action;
3 pure "genes" encoding specific proteins that can be selectively modified to determine: (i) how modifications in either DNA (or RNA) sequences that control expression of a specific mRNA affect such expression; or (ii) how modifications to specific amino acid codons within the gene affect activity of the encoded protein;
4 proteins that can be modified and adapted for use in biotechnology and genetic engineering;
5 proteins for structural and assembly studies;
6 regulatory proteins with defined effects on the host cell so that the mechanism of the interaction between such viral proteins and host cell regulatory pathways can be studied; and
7 nucleic acid "probes" that can be used to identify cellular genes that have similar nucleic acid sequences and thus, can be inferred to have similar functions. They can also be used to monitor the virus load in patients following chemotherapy.

Isolation of structural proteins of the virus

A large number of techniques are available for fractionation of biological molecules and subcellular particles according to their size, density, or charge. **Buoyant density** differences are useful in fractionating enveloped viruses. Each subcellular particle has differences in buoyant density in aqueous solution. Those with large membrane components are "lighter" than those composed of only proteins and nucleic acids.

Virus particles also can be separated from cellular components of different density. This is accomplished by generating an equilibrium density gradient of sucrose or other material in an ultracentrifugal field. Virus particles will "band" or "float" at a specific location within the gradient corresponding to their equilibrium buoyant density (1.18 gm/cm^3 in the example shown in Fig. 11.1.).

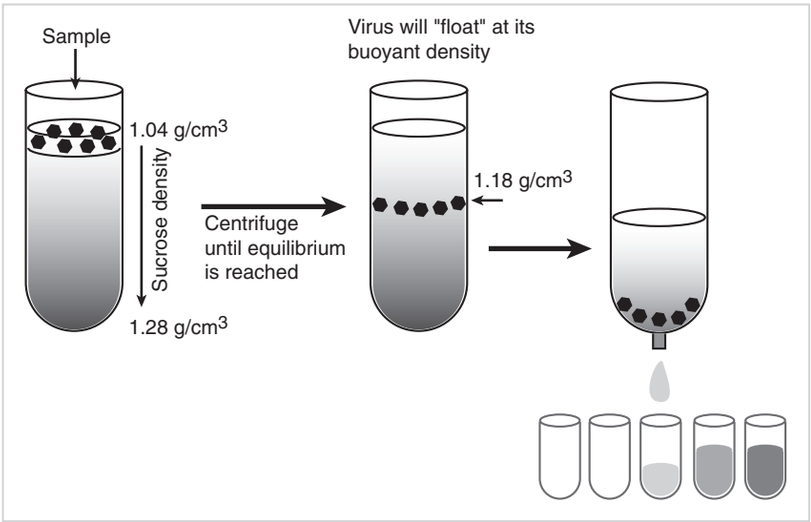


Fig. 11.1 Equilibrium density gradient centrifugation of virus-infected cell components to isolate virus particles. A preformed sucrose density gradient is layered with a solution of infected cell material and subjected to centrifugation at high *g* force at 4°C for several days. Virus particles sediment downward until they reach a layer with a density equivalent to their own. At this density, the virus particles will "float" and careful handling of the gradient in a clear plastic tube will reveal a turbid band of virions that can be removed. In the figure shown, the virus was collected by careful drop-wise fractionation of the gradient through a hole in the tube bottom into small tubes. The presence of virus in the appropriate fractions could be confirmed by plaque assay.

This position represents a balance of forces on the particle: the buoyant force trying to cause the particle to float and the centrifugal force working to cause the particle to sediment lower in the gradient.

Size fractionation is widely used, especially for nonenveloped viruses. For subcellular particles, organelles, and virions, differential sedimentation under a centrifugal field (**rate zonal centrifugation**) allows rapid fractionation and purification. In essence, one takes advantage of the difference in size of these components in the centrifugal field where the largest (the ones with the greatest sedimentation coefficient) will sediment most rapidly or under the least force. The practical aspects of such differential centrifugation can be complex. The basic approach is readily seen in Fig. 11.2. Since most viruses are smaller than mitochondria and larger than ribosomes, further fractionation could be obtained by taking the 100,000 g supernatant material and carrying out further differential centrifugation or more careful size fractionation.

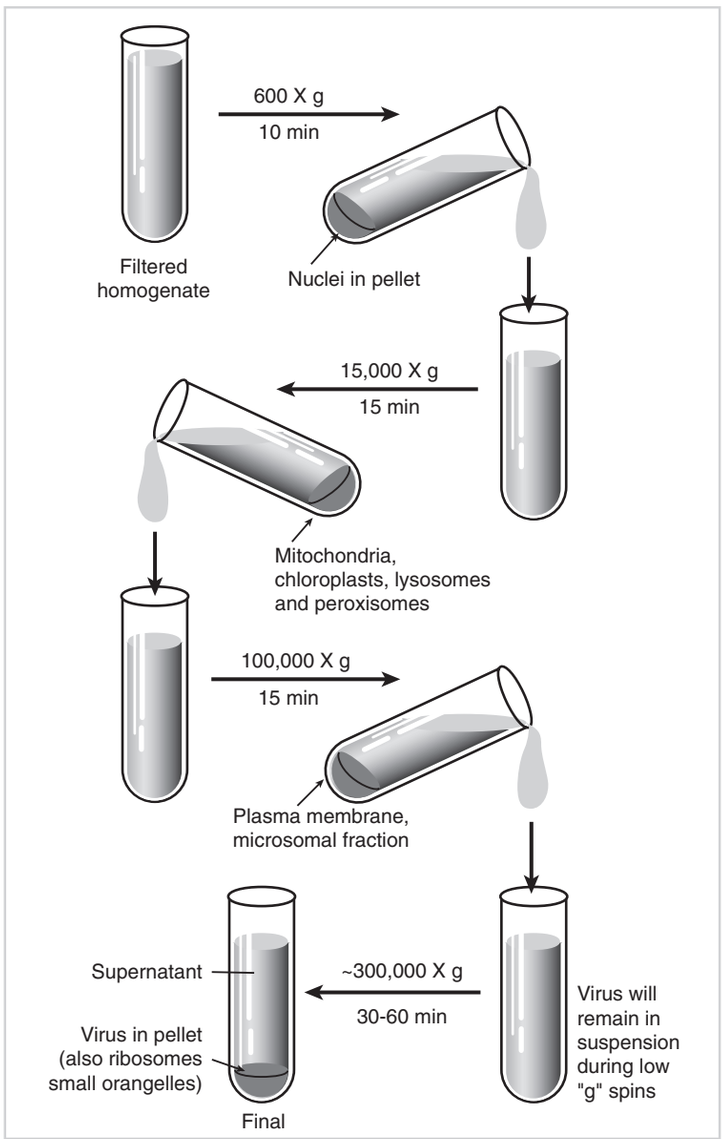


Fig. 11.2 Differential centrifugation to purify virions. Infected cells are homogenized and then subjected to varying steps of centrifugation at increasing *g* forces. At low speeds, large cellular components pellet and can be removed. At the proper speed, viral particles sediment to the bottom of the tube.

Size fractionation of viral structural proteins

Once pure virus is obtained, gentle disruption of the virions with mild detergents or appropriate salt treatments can lead to disruption of the particle and solubilization of the components. Proteins and nucleic acids can be separated from each other by a variety of extraction or differential degradation regimens. For example, small amounts of nuclease could be used to digest nucleic acid into nucleotides, or proteases could be used to digest proteins. These macromolecular components then can be separated according to size or charge, or a combination of both.

It can be shown using physical chemical analysis that the sedimentation rate of a macromolecule is a function of its molecular size and its hydrodynamic volume. Thus, a globular macromolecule (such as most proteins) will migrate at a different rate than an extended (linear) macromolecule of the same size. Further, the same parameters apply to the rate of migration of a similarly charged macromolecule of equivalent shape when subjected to an electrical field, provided the molecules are suspended in a medium of high viscosity that discourages diffusion, such as an acrylamide gel. This is the principle of gel electrophoresis.

In electrophoresis, the rate of migration is *inversely proportional* to sedimentation rate (**s value**). Two macromolecules of equivalent hydrodynamic shape and unit charge will migrate so that the molecule with the larger molecular size will migrate more slowly than the smaller molecule.

These principles are incorporated into a very powerful technique for the size fractionation of proteins. It involves mild denaturation (disruption) of protein structure with the detergent sodium dodecyl sulfate (SDS), which associates with denatured protein to give it a uniform net negative charge. Such proteins can then be size fractionated by electrophoresis on acrylamide gels where the *larger* proteins move more slowly through the gel network, and smaller proteins migrate more rapidly. If the procedure is properly done, such a gel provides good fractionation of viral structural proteins according to size.

Such gels can be stained with color reagents that provide a quantitative measure of the amount of protein of each size as the color reaction is based on reactions with amino acids in the proteins. A small protein has fewer amino acids per polypeptide than a large one; therefore, a sample of, say, 1000 small protein molecules will stain less intensely than will a sample of 1000 larger protein molecules.

A hypothetical example of protein size fractionation and a method of estimating molar ratios is shown in Fig. 11.3 where the fractionation of protein mixtures in a denaturing SDS-containing gel is representative. In this experiment, a solution of an equimolar mixture of four proteins of significantly different sizes (i.e., different number of amino acids in the peptide chain) was fractionated in lane 2. Another sample of three proteins of different sizes in variable amounts (with the smallest protein being present in higher molar concentration than the mid-sized one, and both present in higher concentration than the largest) was fractionated in lane 1. The staining pattern of the gel is shown in lanes 3 and 4 where the staining intensity is represented by band thickness.

The pattern of staining intensity shown in lane 3 makes it clear that the proteins are not present in equimolar amounts. Since staining intensity of the most rapidly migrating band is greater than that of the mid-sized and large bands, there must be more amino acids in the band of small protein. This can only happen if there are more *copies* of the small protein chains. The staining pattern of lane 4 shows a monotonically decreasing intensity of staining with size. Although a precise measure would be required, the band intensity appears to be (roughly, at least) *proportional to protein size*. This is the result expected for an equimolar mixture of proteins of different size, as one small protein polypeptide chain will have fewer amino acids than a single peptide chain of a larger protein.

Determining the stoichiometry of capsid proteins

The molar ratio of different structural proteins can be determined for a given virion, or component of the virion (such as the capsid of an enveloped virus). This is possible because the relative amount

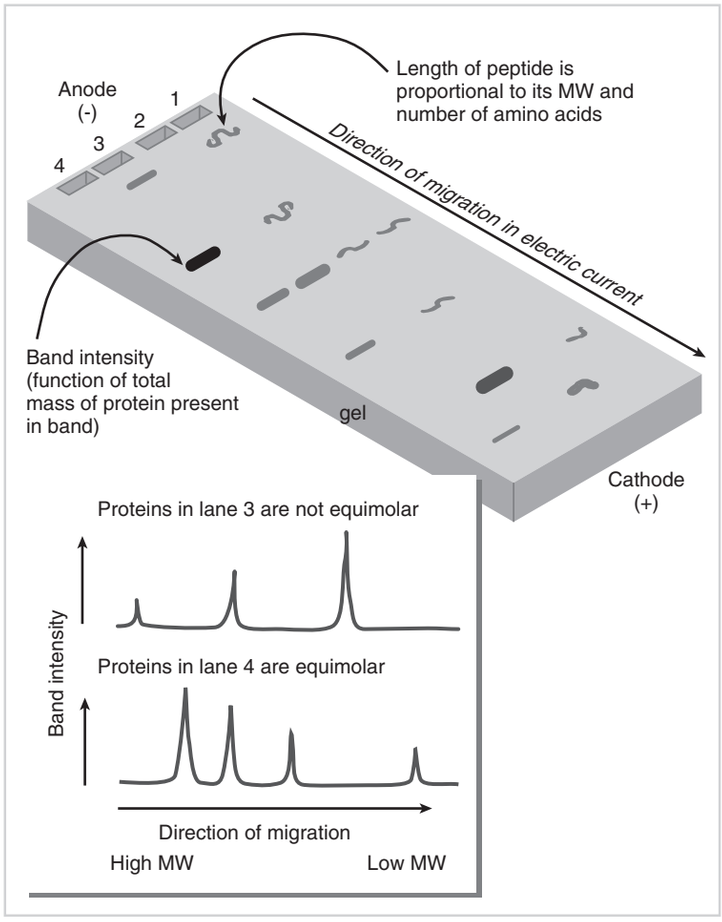


Fig. 11.3 Denaturing gel electrophoresis of proteins. If proteins are gently denatured in a detergent solution such as sodium dodecyl sulfate (SDS), they will assume globular shapes and a net negative charge due to interaction with the detergent molecules. The proteins then can be fractionated by size on acrylamide gels. The proteins migrate in specific bands, and the amount of mass in each band can be determined with a color reaction that measures protein mass. The intensity of banding is a function of the *total* amount of amino acids (a direct correlate with the total mass) in the band, *not* the number of protein molecules per se. (MW, molecular weight.)

of each protein can be measured by staining intensity, or by other means, and because each capsid will yield only the number of capsomer copies present in it when isolated. Full stoichiometric analysis of the capsid's protein composition also requires knowledge of how many capsids are being analyzed. While the ratio of capsid proteins will be constant for different preparations, the absolute amount of protein must be related to the number of capsids to determine how many copies of each protein are present in each capsid.

There are important caveats to the application of this analysis. The most important is that the preparation of virions or capsids must be homogeneous. If a preparation is made up of partial capsids, or truncated helical capsids, the analysis will not be valid. Second, except for a few small enveloped viruses, such as togaviruses and flaviviruses, the number of glycoproteins in the envelope is not stoichiometric. One virion may be enveloped with a virus-modified cellular membrane that has significantly more or less of one glycoprotein than another.

The poliovirus capsid—a virion with equimolar capsid proteins

It is relatively easy to determine that the poliovirus capsid is made up of just four proteins, and that the four capsid proteins (VP1, VP2, VP3, and VP4) are present in equimolar amounts in the capsid. Groups of five copies of each protein are arranged at each of the 12 vertices of the icosahedral

capsid (see Chapters 5 and 15). If the proteins are uniformly labeled with radioactive amino acids, more radioactivity will be in each large polypeptide chain than in each small one. A gel fractionation of the radiolabeled proteins extracted from purified capsids of poliovirus is shown in Fig. 11.4.

There is much less radioactivity in the small VP4 band than in the larger protein bands; however, comparison of the bands' molecular weight with the amount of radioactivity in each reveals the equal numbers of protein molecules. Quantitative analysis of the results of a similar gel fractionation is shown in Table 11.1. Note that the ratio of sizes of VP1 to VP4, for example, is 4.5, while the ratio of radioactivity between them is also 4.5.

Analysis of viral capsids that do not contain equimolar numbers of proteins

Most viruses that encode more than a few proteins in their genomes (e.g., adenovirus and herpesviruses) have capsids that contain proteins in vastly different molar amounts. The adenovirus capsid is shown in Fig. 11.5. A number of proteins within the capsid are not visible in the figure; these include core proteins and hexon-associated proteins. An example of an SDS gel fractionation for adenovirus is also shown in Fig. 11.5. The penton base protein, which is only found at the 12 vertices of the icosahedral capsid, is present in much *smaller* molar amounts (i.e., *fewer copies per capsid*) than is the hexon protein. Conversely, the 24 kd core protein is present in many more copies

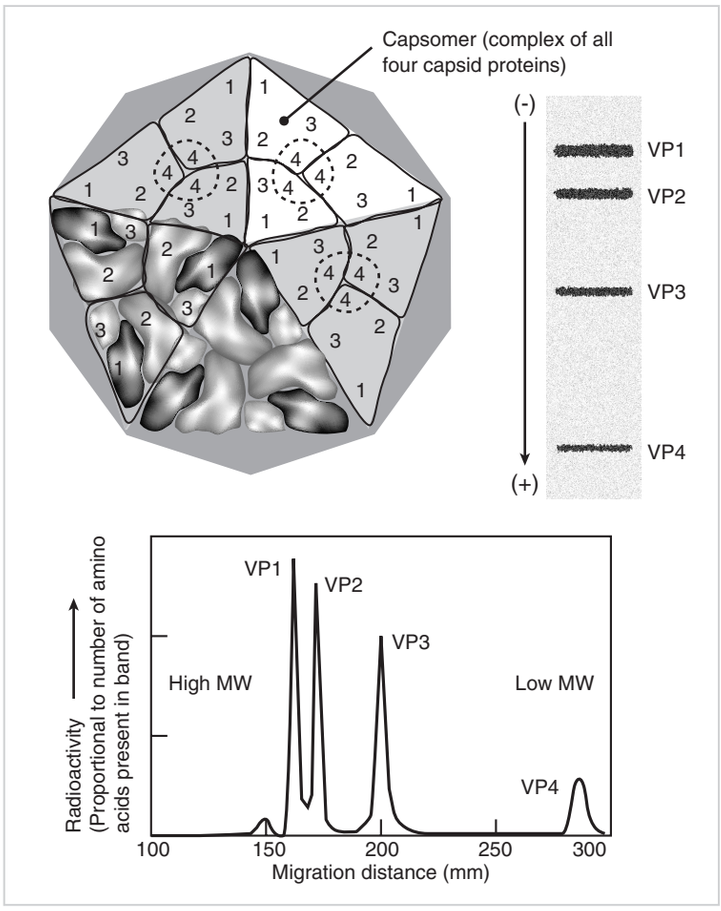


Fig. 11.4 Electrophoretic fractionation of the capsid proteins isolated from purified poliovirus virions. The icosahedral capsid is made up of 60 capsomers, each containing one copy of each of the four viral proteins. The arrangement is shown schematically. Proteins from purified virions were solubilized in buffer and loaded onto a denaturing SDS-containing acrylamide gel. According to their size, viral proteins migrate in denaturing gel electrophoresis, and the amount of total mass in each band can be measured. The ratio of band intensity demonstrates that all four proteins are present in equimolar amounts. (MW, molecular weight.)

Table 11.1 Gel fractionation of the poliovirus four capsid proteins.

Protein	Molecular weight	Radioactivity (cpm)
VP1	33,521	563,153
VP2	29,985	515,742
VP3	26,410	437,743
VP4	7,385	124,806

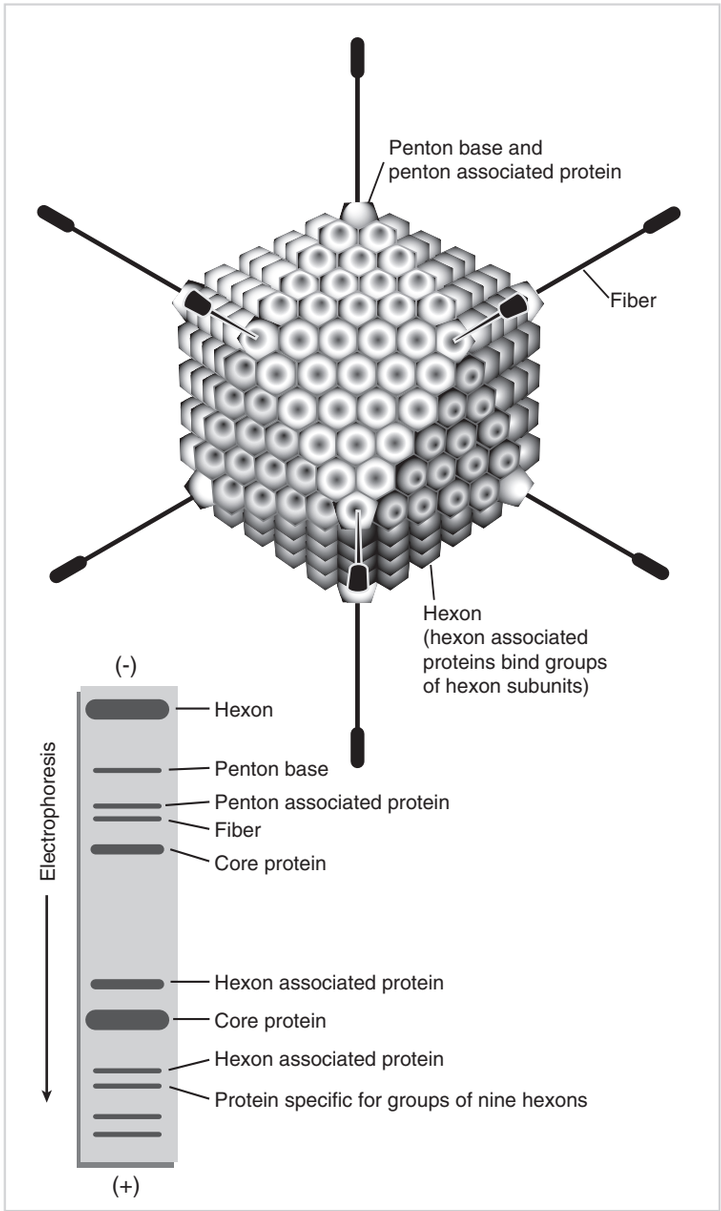


Fig. 11.5 Electrophoretic fractionation of the capsid proteins isolated from purified adenovirus virions. This complex virion contains many different structural proteins that can be fractionated by denaturing gel electrophoresis. The different band intensities do not correlate with protein size. This result demonstrates that the structural proteins are not present in equimolar amounts.

Table 11.2 Protein composition of the HSV-1 capsid.

Protein	Gene	Molecular weight	Copies per capsid	Location in capsid
VP5	UL19	149,075	960	Capsomers
VP19c	UL38	50,260	375	Triplexes
VP21	UL26	45,000	87	Inside capsid
VP23	UL18	34,268	572	Triplexes
VP24	UL26.5	26,618	47	Inside capsid
VP26	UL35	12,095	952	Capsomer tips

per capsid than is the hexon protein. This conclusion comes from the fact that the core protein is considerably smaller than the hexon protein, yet it stains to an equivalent density, while the large penton base protein stains only faintly. Similarly, the capsid of herpesviruses contains proteins in varying molar amounts. The number of copies of the six HSV capsid proteins is tabulated in Table 11.2.

CHARACTERIZING VIRAL GENOMES

Isolation of purified virions provides a primary source of viral genomes. Isolating viral genomes from purified virions is relatively simple. All that is necessary is a mild disruption of the capsid proteins, and the nucleic acid can be isolated by phenol extraction. A famous electron micrograph of a partially disrupted capsid of bacteriophage T4 with its DNA genome extruded is shown in Fig. 11.6.

An accurate determination of the viral genome's nature and molecular size is one of the first things that must be done when working with a newly isolated virus. Such information is important in establishing a basic idea of the virus's genetic complexity. This information, taken together with general characteristics of the virion (i.e., enveloped or not, icosahedral, helical, or complex shape), can be used to make a preliminary assignment of the relationship between the new virus and known virus families using criteria outlined in Chapter 5. Ultimately, of course, a full determination of nucleotide sequence of the viral genome will provide information as to the number and specific amino acid sequences of the proteins it encodes, as well as a precise measure of its degree of relatedness to other viruses.

Sequence analysis of viral genomes

The determination of a DNA virus genome sequence provides the ultimate physical description. While there are methods for sequencing RNA molecules, these methods are not applicable to determining the sequence of extremely large molecules such as those that are the genomes of RNA viruses. However, this problem is readily overcome in the study of RNA virus genomes because RNA can be conveniently converted to DNA using appropriate oligodeoxyribonucleotide primers and retrovirus reverse transcriptase. Enzymatic details of the conversion of RNA to cDNA and then double-stranded (ds) DNA are outlined in Chapter 20.

DNA sequence analysis requires only a few things: (i) pure DNA; (ii) a method for creating a "nested" set of overlapping fragments, all having one end at the same base and each terminating randomly at different bases in the sequence in question; (iii) a method for labeling these nested fragments at the same site; and (iv) a method of separating the fragments with high-enough resolution so that each fragment can be separately resolved.

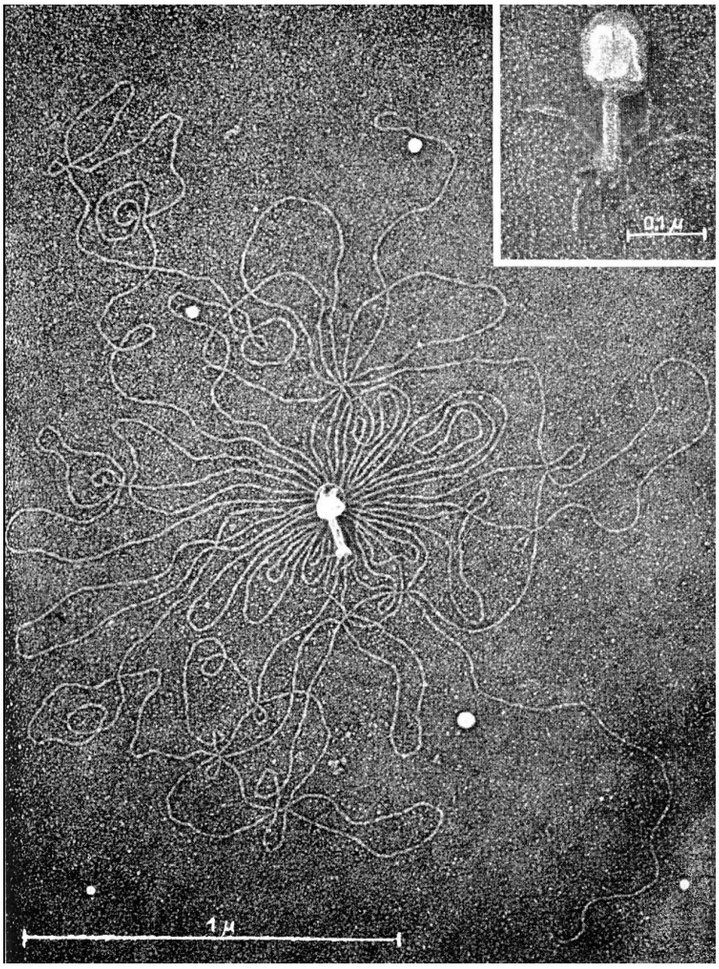


Fig. 11.6 The famous Kleinschmidt electron micrograph of phage T4 DNA extruded from the capsid. Before this photograph was made, there was controversy about whether the viral genome was a single piece of DNA or multiple pieces – the fragility of large DNA molecules made them difficult to isolate without shearing. Kleinschmidt took purified bacteriophages and very carefully exposed them to low osmotic pressure. Under the proper conditions, viral DNA was gently released from the capsid and visualized in the electron microscope. Note the presence of two ends, showing that the DNA is linear. (Reprinted with the kind permission of the publishers from Kleinschmidt, A. K., Lang, D. J., Jacherts, D., and Zahn, R. K. Darstellung und Längenmessungen des Gesamten Desoxyribonleinsäure-1 haltes von T2-Bakteriopahgen. *Biochimica et Biophysica Acta* 1961;61:857–64.)

All the necessary requirements are readily met with the repertoire of techniques available to molecular biologists. Thus, pure DNA can be generated by *cloning* specific fragments (some of the more basic cloning methods are described in Chapter 14). Labeling the fragments can be accomplished easily by use of one of a number of enzymatic methods to incorporate a nucleotide labeled with a radioisotope (usually phosphorus 32 [^{32}P] or sulfur 35 [^{35}S]) or a fluorescent-tagged nucleotide derivative. Separation of deoxyribonucleotides under denaturing conditions on thin acrylamide gels by high-voltage electrophoresis is sufficiently precise to resolve fragments differing in length by a single nucleotide. More recently, the technique of **capillary electrophoresis** using a polymer instead of a gel and very small sample sizes has provided high enough resolution to allow the separation of fragments ranging from ca. 10 to greater than 1,000 bases.

Chemical methods for cleaving DNA at specific bases were originally described by Russian biochemists and perfected for use in DNA sequence analysis by Alan Maxam and Walter Gilbert. Chemical sequencing methods are somewhat laborious, and involve the use of toxic chemicals. They have some advantages, however, and are used for a number of specific applications most notably at this time for determining the sequence and location vis-à-vis a defined restriction site or point on DNA, which interacts with specific DNA-binding proteins.

While chemical sequencing of DNA offers some particular technical advantages and is still occasionally used in almost all laboratories that routinely analyze DNA sequences, enzymatic methods for sequencing are more convenient and are the most frequently used approaches. These methods take advantage of the fact that DNA polymerase will generate a complementary copy of DNA onto a primer annealed to the template strand.

If a small amount of a dideoxynucleoside triphosphate (which causes chain termination due to lack of a 3′-OH group) is added to the primed synthesis reaction, the synthesis of the new DNA strand will terminate wherever this base is incorporated. The fact that strand synthesis can only proceed from the primer provides a convenient method for generating overlapping, nested sets of oligonucleotides complementary to any DNA sequence 5′ of the primer in question.

The enzymatic method was originally perfected by Sanger and collaborators, and has been modified in many ways. For example and as described a bit later, the method has been automated so that analysis can be carried out and directly entered into computer databases with little human interfacing. The rapid progress in the human genome project, as well as the increasingly frequent publication of sequences of the entire genomes of free-living organisms, is due to the ease and speed of enzymatic methods. Indeed, where it took several years to determine the complete sequence of HSV-1 (152 kbp) a decade ago, the same problem can now be solved in weeks! Complete sequence analysis of any virus of interest can be carried out essentially as soon as the virus is isolated and the genome purified.

To generate overlapping oligonucleotides with the same 5′ end, all that is needed is a primer sequence that will anneal to a region that is located 3′ to the sequence of interest. This is often a region in the vector used to clone the DNA in the first place. Annealing of the primer, which can either be labeled with a radioactive or fluorescent marker, or unlabeled, is followed by enzymatic synthesis of the complementary strand of the DNA template in the presence of a labeled base or bases. After synthesis is allowed to proceed for a short time to ensure the formation of highly labeled material, the reaction is broken into four aliquots and a small amount of a single di-deoxy-base-triphosphate is added to generate oligonucleotides with random stops at a given base. This is shown in Fig. 11.7a and below for T (remember, lowercase nucleotides signify the complementary base on the antiparallel strand, and DNAY is the region of DNA to which the labeled primer, dnay*, binds):

5′-DNAX-ATACCGATCGTG-DNAY-3′
tagcac-dnay*--5′
5′-DNAX-ATACCGATCGTG-DNAY-3′
tggctagcac-dnay*--5′
5′-DNAX-ATACCGATCGTG-DNAY-3′
tatggctagcac-dnay*--5′
etc.

Once generated, the oligonucleotides can then be fractionated on denaturing sequence gels as shown in Fig. 11.7, a schematic representation and examples of actual experimental data. While the separation method is essentially the same as for the chemical method, much less DNA can be loaded, as the labeling can be tailored to the fragment size range to be resolved. This allows higher resolution of the gels.

Automated sequencing takes advantage of the fact that laser light of a given wavelength can excite specific dye molecules to fluoresce at specific frequencies. Different dye molecules fluorescing at different wavelengths can be chemically linked to each of the four di-deoxy-base-triphosphates in the reaction mixes described above. These can be used all together in the polymerase reaction to generate nested products terminating at every base in the sequence. This mixture is then loaded onto a capillary electrophoresis apparatus and subjected to a high voltage. The shortest fragments will, of course, migrate most rapidly through the capillary and past a laser activated detector where the presence of the terminating, dye-containing fragment will fluoresce at a wavelength character-

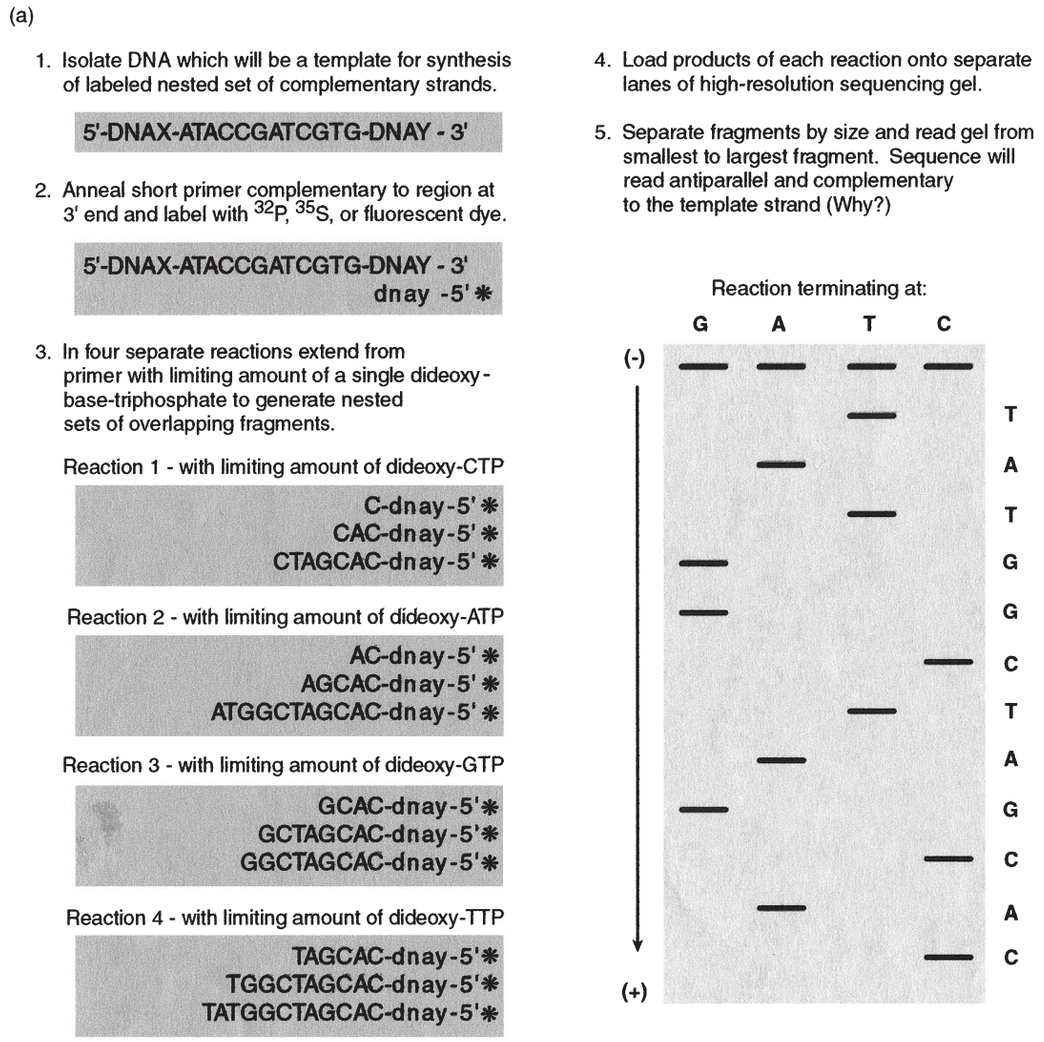


Fig. 11.7 Enzymatic sequencing of DNA. The generation of overlapping oligonucleotide sets complementary to a template strand of DNA for sequence analysis was developed by Sanger and colleagues and is described in the text. *a*. An outline of the basic method. One major advantage of the method is that it can be used to generate very long sequences with reactions using a single primer site. *b*. For example, the gel on the left shows the sequence of a cloned fragment of HSV-1 DNA and the plasmid it is cloned into about 100 bases 3′ of the primer site. The sequence can be read as follows:

5′-ACGTC$_2$T$_2$A$_2$GCTAG$_2$C$_2$G$_2$C$_2$TCGC$_2$ATCG$_2$AG$_5$C$_2$TAGT$_2$CGA$_2$TAGCTA-3′

The right gel shows a comparative analysis of the sequence of a wild-type and mutant promoter region for an HSV-1 capsid protein mRNA. This region is about 300 bases 3′ of the location of the sequencing primer and shows that high resolution is still readily obtainable as long as the reaction products are fractionated under proper conditions, which in this case are long fractionation times under denaturing conditions. The regions of the two sequences that are different are indicated; the sequences read as follows:

Wild type: 5′-TCACAGGGTTGTCTGGGCCCTGC-3′

Mutant: 5′-TCACAGGACCGGCTGACCGCCTGC-3′

Just above (i.e., 3′ of) this region is an example of a typical experimental artifact of this type of sequencing: a spot where there is termination in all reactions due to a structural feature of the sequence in question. Note that the sequence again can be read accurately beyond this point.

istic of the terminating deoxynuceleotide. A computer is used to keep track of order of appearance of the various colored signal peaks, and the sequence is automatically recorded. An example of this methodology is shown in Fig. 11.8.

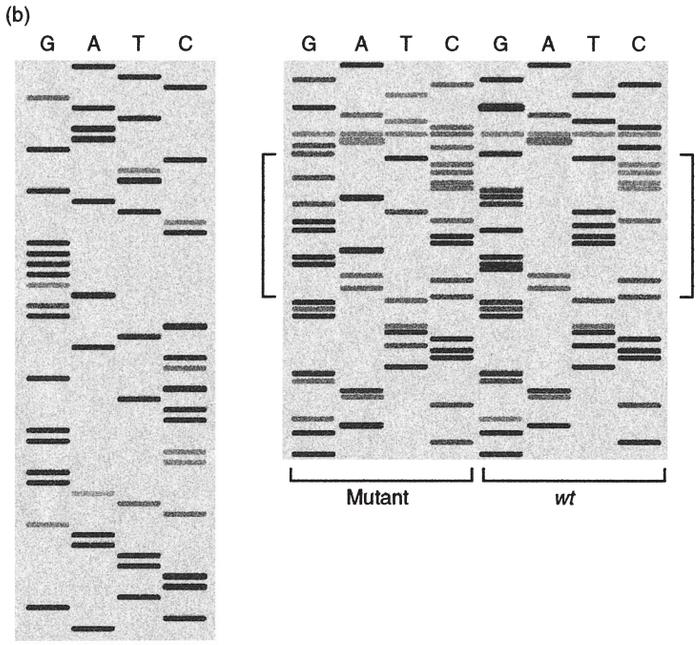


Fig. 11.7 *Continued*

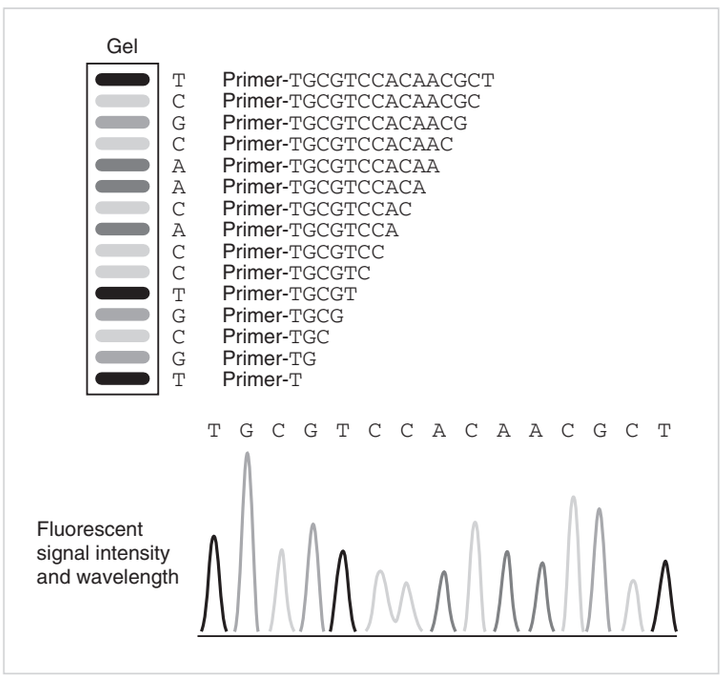


Fig. 11.8 Automated DNA sequencing. See Plate 5 for a color image.

Measuring the size of viral genomes

While nucleotide sequence of a viral genome automatically determines the genome's molecular size, a number of very accurate physical and biochemical methods for measuring genome size were developed well in advance of sequencing. These methods are still in occasional use and many important publications refer to them; therefore, a brief description of a few that were most widely used is valuable.

Direct measure of DNA genome lengths in the electron microscope

The entire size range of dsDNA genomes found in viruses is within the range of DNA sizes that can be visualized in the electron microscope using appropriate shadowing and spreading methods to ensure that the very long, flexible DNA strand is not so tangled as to be unmeasurable. DNA is chemically quite stable, and molecules up to 50,000 base pairs (50 kbp) can be isolated with relative ease with no particular precautions other than normal laboratory care. The biggest problem with isolating larger DNA molecules is mechanical shear, because of their relative stiffness. With proper experimental techniques, viral genomes as large as 250 to 300 kbp can be isolated without degradation. Under proper spreading and shadowing conditions, the length of DNA molecules is a direct function of their size in base pairs (about 3 μ/kbp). If a large viral genome, such as HSV, is spread along with an appropriate internal size standard (such as a small circular DNA molecule of known molecular size), the ratio of lengths can be used to calculate genome size quite accurately. As described in Chapter 9, inclusion of an internal standard is an important control against the inevitable variation in conditions and magnification inherent in any electron microscopic technique.

A spread of HSV DNA, along with the dsDNA replicative intermediate of the single-stranded ΦX174 bacteriophage, is shown in Fig. 11.9. The ratio of sizes for different strains of HSV ranges from 25.7 to 28.1 times the size of the bacteriophage marker. This calculates to a range of HSV genome sizes between 138 and 151 kbp, while the size measured by sequence analysis of a single strain of the virus is 152.6 kbp. It is not known at this time whether the variation in genome lengths found with different strains of the virus is due to actual differences in genome size or to experimental uncertainties.

Rate zonal sedimentation and gel electrophoresis for measuring viral genome size

In contrast to DNA molecules, molecules of single-stranded (ss) RNA, as found in the genomes of many RNA viruses, are susceptible to chemical degradation at relatively mild pH ranges (<3 and >9). Further, ribonucleases that readily degrade RNA are notoriously difficult to inactivate and are often excreted by bacteria and fungi, which can contaminate laboratory reagents. Indeed, a very

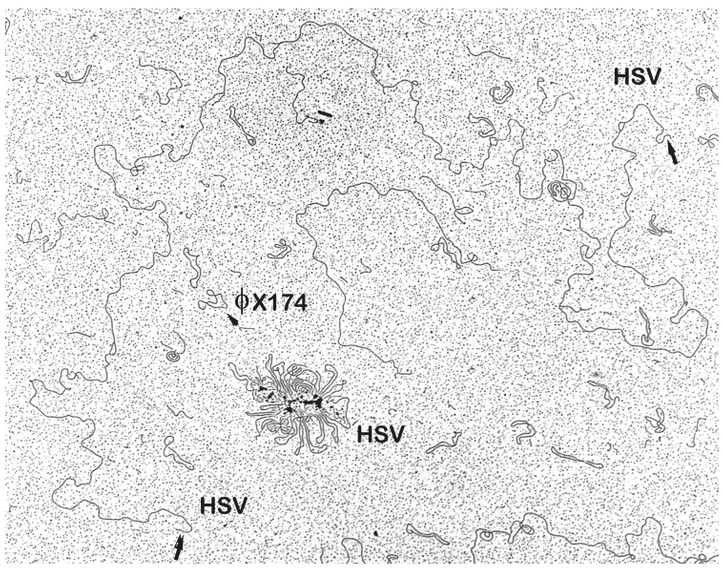


Fig. 11.9 Use of a method similar to that shown in Fig. 11.6 to spread HSV DNA for comparative contour length measurement. One full-length DNA molecule is extended and its length can be measured and compared to the length of the circular ΦX174 replicative form (RF) DNA molecules included as size standards. A second DNA molecule (or molecules) has formed a tangle around a contaminating protein fragment in the solution.

potent ribonuclease ("finger nuclease") is found in sweat and is carried on the hands. Despite the difficulties in working with RNA, with appropriate care, RNA molecules as large as 15,000 bases can be isolated with relative convenience. This means that the size range of ssRNA virus genomes (2–15 kb) is well within the bounds of experimental manipulation.

Unfortunately, ssRNA is not very easy to work with in the electron microscope because its relatively great flexibility makes it very difficult to spread for accurate measures of length. The size of viral genomes, however, is just the size that is convenient for rate zonal (velocity) sedimentation in sucrose gradients, and conversely, for size fractionation on low-density acrylamide gels using electrophoresis.

The same principles applied to protein fractionation generally can be applied to nucleic acid fractionation. Since all ssRNA and ssDNA molecules will have essentially the same shape in solution, their sedimentation rate in a centrifugal field will be only a function of their molecular size. Thus, each specific size of ssRNA or ssDNA macromolecule will sediment at a specific rate in a centrifugal field under standard conditions. The sedimentation rate under such standard conditions is termed the *sedimentation constant* (s value) for that macromolecule. This s value also determines the rate of migration under standard conditions in acrylamide gel electrophoresis. The s value is related to molecular size by a logarithmic function. For example, prokaryotic 16s ribosomal RNA (rRNA) and 23s rRNA are 1.5 kb and 2.3 kb, while eukaryotic 18s rRNA and 28s rRNA are 2 kb and 5.2 kb, respectively.

Interestingly, the size range of RNA (and ssDNA) molecules found as viral genomes is just that range that is readily separable by gel electrophoresis or rate zonal centrifugation. Further, bacterial and eukaryotic rRNAs provide readily available internal size standards of just the right general values for measuring the sizes of mRNA and viral genomes. For this reason, many scientific reports define species of RNA by s value, which is just a shorthand way of listing its molecular size.

The principles of rate zonal centrifugation for measure of molecular size also can be applied to dsDNA molecules, but the inflexibility of dsDNA in solution and its—generally—larger size require considerably different experimental techniques. Often determining the size of dsDNA molecules requires using analytical ultracentrifuges that generate very high centrifugal fields and very sophisticated optical methods for measuring sedimentation. It is very important to remember that the mathematical relationship between the s value of a dsDNA molecule and its size is quite different from one relating the size and sedimentation rates of ssRNA or ssDNA molecules.

While most double-stranded viral genomes are too large for easy gel electrophoresis, the ability to cut DNA molecules into specific pieces using restriction endonucleases (see Chapters 8 and 14) allows one to partially avoid this problem. If purified virion DNA is digested with a restriction endonuclease that does not cut it too often, fragments of a convenient size can be produced and fractionated on high-porosity agarose gels.

An example is shown in Fig. 11.10, which shows the electrophoretic separation of DNA fragments produced when the 48 kbp bacteriophage λ genome was cut with the restriction endonuclease *Bst*EII, which cleaves DNA at locations where the 7-base sequence GGTNACC occurs (N represents any base). Mobility of the fragments is roughly proportional to a logarithmic function, but this function is not constant throughout the whole size range of fragments produced. This can be seen by looking at the semilog plot of migration versus log of fragment size, which is also shown in Fig. 11.10. The only way that gel electrophoresis can be used to measure accurately the total size of all fragments produced by digestion of a mid-sized to large DNA genome is to include numerous different size markers and use several different agarose concentrations in the gel. Still, the method is very convenient, and is often used to compare the size of specific restriction fragments produced by digesting related viruses.

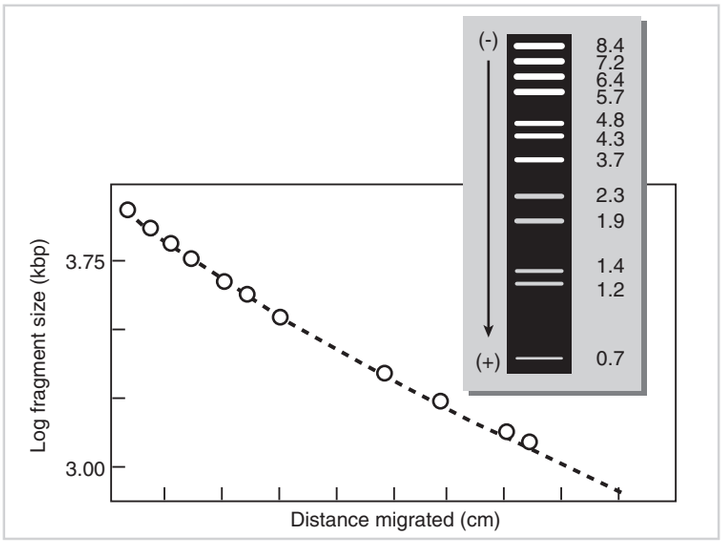


Fig. 11.10 Electrophoretic separation of bacteriophage λ restriction fragments. Bacteriophage DNA was digested with the restriction enzyme *Bst*EII, which is so named because it was derived from *Bacillus stearothermophilus* (a hot springs-loving organism or **extremophile**). The DNA fragments were fractionated by electrophoresis on 1% agarose gel, and visualized by viewing under ultraviolet light following the addition of ethidium bromide, which specifically binds dsDNA and produces orange florescence under ultraviolet light. The migration rate of individual fragments, whose sizes are shown, is plotted against a log of fragment size.

Use of renaturation rates to measure nucleic acid size and complexity

The **Watson-Crick base-pairing rules** for DNA are a shorthand way of describing the structure of nucleic acids made up of two **complementary strands**. The two strands are *antiparallel*, and A in one strand bonds by hydrogen to T in the other, and the same for G and C in the two strands. Thus, once the sequence of a strand of nucleic acid is known, the sequence of the complementary strand is also known.

If a dsDNA or dsRNA molecule or a DNA—RNA hybrid molecule is *denatured* or *melted* by heat, the two complementary strands will separate and become single strands. The temperature at which this denaturation takes place is called the **denaturation temperature (T_m)** for the duplex. The T_m of a DNA, RNA, or DNA—RNA duplex is primarily a function of the percentage G + C in the duplex and the salt concentration in which the melting occurs. It is not a function of the information encoded in the nucleic acid molecules' sequence at all.

If the two denatured strands of a duplex molecule are heated at moderate salt concentrations to a temperature about 10° to 25° below the T_m, the strands will **reanneal** and the precise double-strand duplex will re-form. This method, which is also the basis for generating a nucleic acid **hybrid**, provides a very powerful method for isolating RNA molecules encoded by a given DNA sequence, and for detecting RNA or DNA molecules that are complementary to a given one. Some uses of RNA—DNA hybridization to detect viral mRNA are briefly described in Chapter 13.

One use to which reannealing of duplex nucleic acids can be put is to measure the *complexity* of a nucleic acid sequence, which can be related to the size under certain conditions. This comes from the fact that if a piece of DNA 1000 base pairs long is made up of, for example, 10 identical repeats of a 100 base unique sequence, it will denature essentially 10 times faster under ideal conditions than the same length of DNA made up of one unique 1000 base sequence.

The relationship between denaturation rate and genome complexity of a given double-stranded nucleic acid is given by the following formula:

$$C/C_0 = 1/(1 + KC_0t)$$

In this formula, C/C_o is the proportion of nucleic acid that has reannealed at a given time (t), C_o is the initial concentration of the nucleic acid, and K is a second-order rate constant that depends on

the physical size of the reannealing nucleic acid fragments, the salt concentration, and the kinetic complexity of the nucleic acid.

Since viral genomes are essentially unique sequences, it follows that if the rates of renaturation of two different viral genomes are measured under standard conditions (ones that control for base composition, physical size of reannealing fragment, and other variables), the larger genome will renature more slowly than the smaller one. This observation allows measurement of relative sizes of dsDNA genomes by renaturation rate; if the size of one genome is known, that of the other can be estimated.

This method has some distinct advantages for working with large DNA molecules. First, the molecules need not be intact. Indeed, they must be fragmented to a uniform small size for the best results. Second, if a pure radioactive probe is used, the bulk of the DNA need not be pure. Contaminating DNA sequences will not affect the results.

As an example, this method was used to determine that the size of the HSV genome is about 80% of the size of bacteriophage T4. Comparative reannealing curves are shown in Fig. 11.11. Since bacteriophage T4 has a 168 kbp genome, while HSV has a genome of 152 kbp, the actual ratio of sizes is 89%, which is acceptably close to the measured value of 80%, and shows that the method is fairly accurate.

While in theory this method can also be used to measure renaturation rates of dsRNA molecules, it is not practical to do so, for the following reasons. Double-stranded RNA has a very high denaturation temperature, and renaturation rates are technically difficult to measure. Also, the number of known dsRNA viral genomes is relatively small, so there is little need to make such measurements.

The method is most useful for estimating the *abundance* of mRNA molecules expressed from different genes encoded by a virus or a cell. For example, by measuring the rate of annealing (hybridizing) viral DNA with RNA isolated from an infected cell, one could readily determine whether a given virus expresses more mRNA encoding a capsid protein than mRNA encoding a DNA replication enzyme. Clearly, the more abundant mRNA would hybridize more rapidly, which can be measured or estimated in a number of ways.

With viruses infecting eukaryotic cells, the most accurate way of doing such an experiment is to first convert the mRNA into cDNA using retrovirus reverse transcriptase. This conversion can be done readily using a short length of oligodeoxythymidine as a primer, since the mRNA expressed is

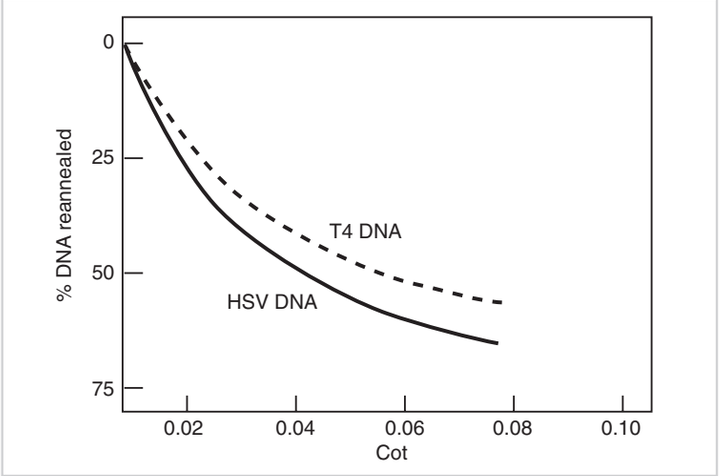


Fig. 11.11 Comparative reannealing of two different viral DNA genomes showing their relative size differences. DNA was denatured by heat and allowed to reanneal by incubation under appropriate conditions. DNA reannealing was measured by virtue of the fact that dsDNA absorbs less ultraviolet light than does the equivalent amount of ssDNA. The proportion of renatured DNA at any given time for a specific concentration of DNA was measured, and the results of a number of experiments combined and plotted here. The slower renaturation of the larger T4 DNA molecule is clearly evident. (Cot = the initial concentration of DNA (Co) multiplied by the time of reannealing (t).

polyadenylated at its 3′ end. This method has a great advantage in that only short (100–200) base lengths of cDNA need to be synthesized because the number of any one will be a function of the number of mRNA templates present. Thus, no great care needs to be taken to ensure that the mRNA is fully intact or that the cDNA reaction goes to completion. Such cDNA can also be used as a hybridization probe for detecting the RNA molecules from which it was derived. Some methods for this are described in Chapter 12.

One of the most important things to remember about nucleic acid hybridization or annealing is that *under the proper conditions, the presence of large (even overwhelming) amounts of RNA or DNA of a sequence even only slightly different from that of the test DNA or RNA has no effect on the rate or amount of hybrid formed.* The rate of hybrid formation is a function only of the amount of the test sequence present in the mix. This makes nucleic acid hybridization an exquisitely sensitive method for detecting and quantitating RNA and DNA, and much understanding of the regulatory processes involved in viral and cellular gene expression has a basis in the ability to precisely measure the amount of such expression taking place at any given time.

The polymerase chain reaction – detection and characterization of extremely small quantities of viral genomes or transcripts

The ability to characterize, work with, and control many viruses is limited by the fact that they are present in very small quantities in a given cell, tissue, or host. The use of a fluorescent stain such as ethidium bromide allows the ready detection of 100 ng or less of dsDNA. For a viral genome of, for example, 30 kbp, this works out to be approximately 5×10^{11} molecules. Radioactive labeling can greatly increase the sensitivity of detection, but it is not always possible to specifically label the DNA fragment of interest in the tissue being studied.

The problem of visualizing and manipulating extremely small quantities of DNA was overcome in larger part by developments of the *polymerase chain reaction (PCR)* initiated and commercialized by scientists at the Cetus Corporation in the mid-1980s. The principle, illustrated in Fig. 11.12a, is quite simple. Consider a fragment of dsDNA present as even a single copy in a cell or animal. If this DNA is denatured and short oligonucleotide primers can be found to anneal to the opposite strands at positions not too far away from each other (e.g., within a thousand bases or so), a strand of cDNA can be synthesized using DNA polymerase. The new product will be double stranded in the presence of the nonprimed denatured DNA.

Now, if the newly synthesized dsDNA is itself denatured, and the priming and DNA synthesis step is repeated, this short stretch of DNA will be amplified as compared to the strands of DNA that did not bind primer. This process can be repeated many times in a chain reaction to amplify the desired strand of DNA to useful amounts.

To work properly, the oligonucleotide primers must be long enough to be highly specific, but short enough to allow frequent priming. The appropriate length works out to be about 20 to 30 bases. The technology for synthesis of 20- to 30-base oligodeoxynucleotides is well established and can be chemically performed relatively inexpensively. Indeed, numerous large and small biotechnology companies make oligonucleotides commercially.

Also important is the ability to do the reaction, denaturation, and reannealing in a single tube many times over. This is accomplished by using the heat-stable DNA polymerases isolated from organisms such as *Thermophilis aquaticus* (*Taq*), which live in hot springs, and the use of computer-controlled thermal cyclers that can repeat the annealing, synthesis, and denaturation steps rapidly and repeatedly over 12 to 24 hours.

In practice, the method can be used to detect the presence of extremely small amounts (less than a single copy/cell) of a known viral genome by selection of appropriate primer pairs based on the

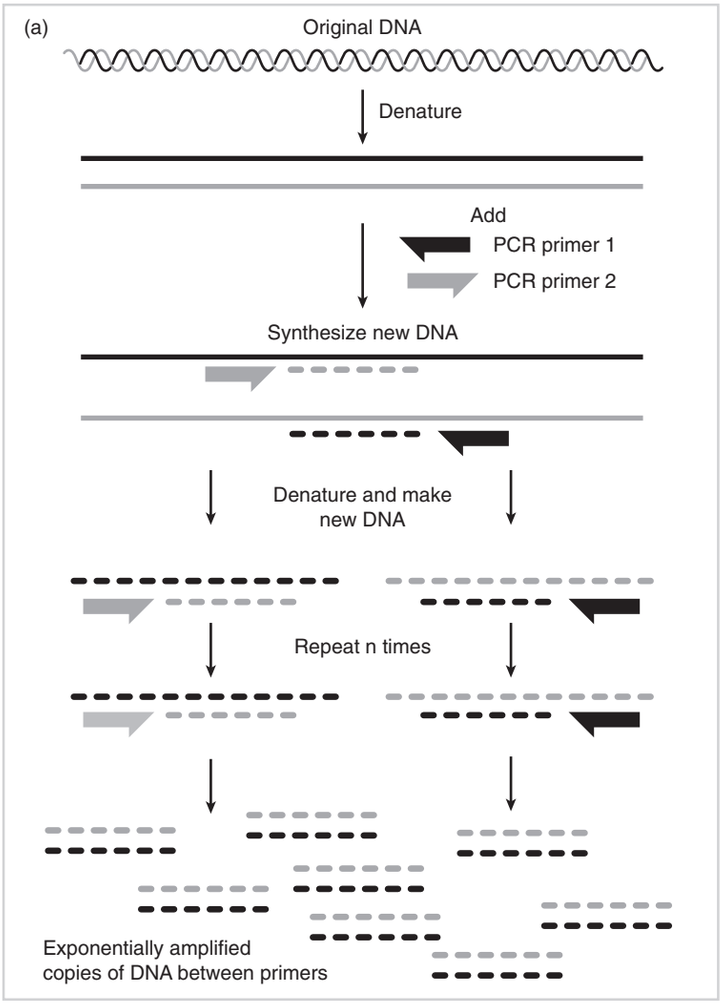


Fig. 11.12 Amplification of DNA with the polymerase chain reaction (PCR). *a.* The basic method requires specific primer sets that can anneal to opposite strands of the DNA of interest at sites relatively close to each other. After denaturation, the primers are annealed, and DNA is then synthesized from them. All other DNA in the sample will not serve as a template. Following synthesis, the reaction products are denatured, and more primer is annealed and the process repeated for a number of cycles. The use of heat-stable DNA polymerase allows the reaction to be cycled many times in the same tube. A single copy of a DNA segment of interest could be amplified to 10^9 copies in 30 cycles of amplification. Can you demonstrate this mathematically? *b.* The amplified DNA products from a segment of HSV DNA. A total of 1 μg of nonspecific DNA was added to each of a series of tubes, and viral DNA corresponding to the copy numbers shown was added. Following this, primers, heat-stable DNA polymerase, and nucleoside triphosphates were added, and 30 cycles of amplification were carried out in an automated machine. The reaction products were fractionated on a denaturing gel and visualized by autoradiography. The lower gel shows the results of amplification under identical conditions of DNA isolated from two rabbit trigeminal ganglia. One was taken from a control rabbit, and the other was taken from a rabbit that had been infected in the eye with HSV followed by establishment of a latent infection. The use of rabbits to establish HSV latency is shown in Fig. 3.5. Amplified DNA from each sample was fractionated in the lanes shown; in addition to the amplification products, a sample with PCR-amplified HSV DNA as a standard (std) as well as some size markers were fractionated.

knowledge of the sequence of the genome. An example of the use of PCR to detect HSV genomes is illustrated in Fig. 11.12b.

In addition to its value for detecting vanishingly small amounts of viral genomes, PCR can also be used to make quantitative estimates of the amounts of viral genomes or transcripts present in different tissues, or under different conditions of infection. Such analyses have been vital in formulating the present models for understanding the replication of HIV and its pathogenesis leading to AIDS. It has also been very useful in studying the latent phase of infection of herpesviruses.

In the experiment shown in Fig. 11.12b, a series of dilutions of a fragment of HSV DNA corresponding to the copy numbers shown were carried out and subjected to PCR amplification. The gel shown was used to fractionate the reaction products that were made radioactive by the addition of a small amount of radiolabeled nucleoside triphosphate to the reaction mix. An amplified signal from 1000 copies of the genome provided a detectable signal with a short exposure of the gel to x-ray film. While the sensitivity could have been greatly increased by increasing the rounds of ampli-

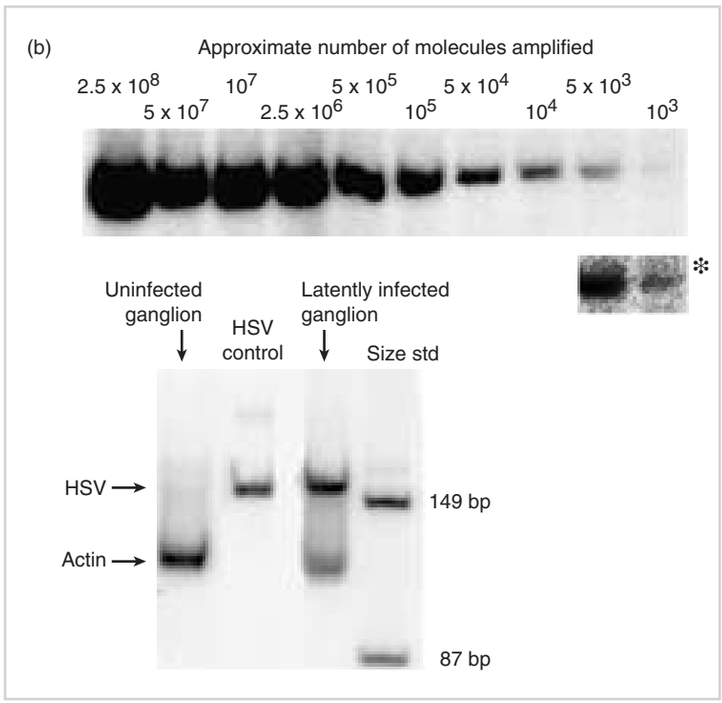


Fig. 11.12 *Continued*

fication, and altering other conditions, this would have made precise quantitation impossible, so each experiment must be designed to optimize the most desirable property of the reaction.

For Fig. 11.12b, the quantitative conditions used were applied to DNA isolated from rabbit ganglia latently infected with HSV DNA. The strength of the HSV-specific signal allows a measure of the amount of viral genomes present in the sample. The figure also illustrates the fact that two (or more) PCRs can be carried out in the same sample as long as the appropriate primers are included. Here, a primer set specific for the cellular actin gene was used to allow measurement of the recovery of tissue as an internal control.

Application of such quantitative analysis, along with knowledge of the number of neural cells in the sample, allows one to calculate that a typical latently infected neuron in an experimentally infected rabbit might harbor between 10 and 100 viral genomes. Of course, such results are average values, but the nucleic acid from a single cell can be subjected to PCR amplification and analysis.

PCR can also be used to look for the presence of genes related to a known gene. Such detection is based on the assumption that regions of a DNA sequence encoding a gene related to the one in hand will contain some stretches of identical sequences in their genomes. Detection can be accomplished by amplifying the DNA in question with a series of potential primer sets. If one or several of these yield products of a size within the range of those seen with the known gene, these products can be isolated and sequenced. If necessary, this can be done after the amplified fragment or fragments of interest are cloned using methods outlined in Chapter 14.

Finally, PCR also can be used to viral RNA (either genomes or transcripts) present in very low amounts. Detection is accomplished by generating a cDNA copy of the RNA by use of retrovirus reverse transcriptase, followed by PCR amplification using a known primer set. If oligodeoxythymidine is used as a primer, it will anneal to the polyA tails of mRNA for the generation of cDNA. If the correct primers are used, PCR can detect vanishingly small numbers of transcripts. An example of such a use in the analysis of HSV gene expression during reactivation is shown in Fig. 18.9.

The very high sensitivity of PCR, along with the ability to sequence the amplified products of PCR, also can be applied to determining splicing patterns of RNA expressed in cells. The application to analysis of viral transcription is briefly outlined in Chapter 13, and is illustrated in Fig. 13.7b.

Finally, PCR is invaluable for epidemiology and forensics. For example, it was used to amplify traces of influenza virus genomes still present in frozen cadavers of victims of the 1918–20 influenza pandemic. Study of the sequence of such material has allowed scientists to establish some relationships between that virus and modern strains. Its use in forensics is somewhat outside the scope of this text, but it should be clear that the ability to amplify traces of DNA along with rapid sequencing methodology allows the identification of any genome present in more than a very few copies from viral to human.

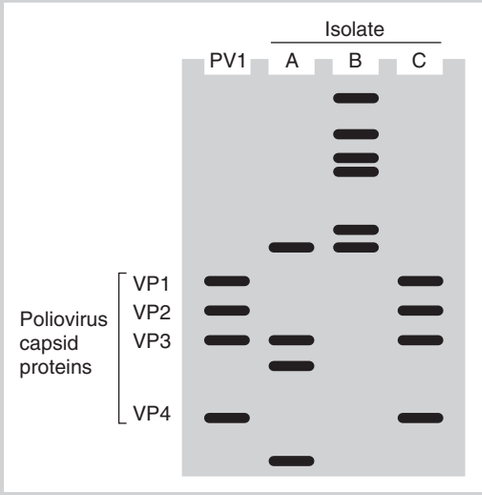
QUESTIONS FOR CHAPTER 11

1 You have encountered a virus named hotvirus with three capsid proteins, E, K, and W. After gel fractionation of a purified stock of pure viral capsids that were uniformly radiolabeled with radioactive amino acids, you obtain the following results:

Protein	Molecular weight	Radioactivity (cpm)
E	5,280	29,348
K	18,795	101,185
W	10,776	122,674

What are the best values for the ratios of the proteins E to K to W?

2 Your laboratory has isolated a number of possible enteric viruses from samples of contaminated water. You have grown these viruses in appropriate cell cultures and have labeled the proteins with ^{35}S-methionine. You have purified virus particles from these cultures and separated the capsid proteins by sodium dodecyl sulfate (SDS) polyacrylamide gel electrophoresis. Below is an autoradiogram of this experiment with poliovirus type 1 (PV1) included as a control:



a Which of these isolates is potentially a virus identical to or very closely related to poliovirus type 1?
b Which of these isolates may be another member of the same family as poliovirus type 1?
c From which family of viruses might isolate B come? (Note: you will probably have to do some searching in Chapter 16 to find properties of enteric viruses in order to answer this question.)

Continued

3 How is SDS polyacrylamide gel electrophoresis used for the analysis of proteins? What is the basis for this technique?

4 Besides the molar ratio of the proteins, what would you need to know to determine the amount of specific proteins per capsid in a particular virus?

5 While analyzing the structural proteins of a pure stock of adenovirus by SDS polyacrylamide gel electrophoresis, you find, among others, two bands of equal intensity that migrate at 30 kd and 60 kd, respectively. What conclusion can you draw from this observation?