III

Composition of Viruses

Several hundred virus diseases are now recognized as such, but the chemical compositions of relatively few viruses have been reported. The main reasons for this situation are that many interesting experiments can be done without precise information about viral composition, and it is not easy to obtain all viruses in a state of purity adequate for analysis. Finally, it is somewhat tedious to perform complete and thorough analyses of viruses even when they can be obtained in adequate amounts and in sufficient purity. Consequently, it is common for investigators to make just those analyses most pertinent to a particular topic under investigation.

Tobacco mosaic virus proved to be a nucleoprotein, that is, a specific combination of nucleic acid and protein. The same characteristic nucleoprotein was obtained from a variety of hosts infected with tobacco mosaic virus (TMV). In each instance the virus nucleoprotein appears to be foreign to its host as judged from its absence in normal plants and its lack of serological relationship to the normal host constituents. This holds true for many other viruses, although among the more complex viruses, those that acquire envelope structures at and bud out through cell membranes typically contain some host cell constituents in their envelopes.

Most viruses give rise during multiplication to occasional mutants, or variants; when isolated, these are found to be of the same general composition as that of the parent virus. It now seems generally true that strains of a given virus have identical proportions of protein and nucleic acid, although exceptions to this rule have been observed in mutants of the more complex viruses.

All types of viruses, including plant, bacterial, higher animal, and insect viruses, contain either ribonucleic acid (RNA) or deoxyribonucleic acid (DNA) (see Table 2). At one time it was thought that some viruses, such as influenza virus, contained both types of nucleic acid, but, as a consequence of more refined methods for purification and analysis of viruses, no virus is presently known that contains both RNA and DNA. This may be contrasted with bacteria including *Rickettsia* and *Bedsonia*, which appear to have both DNA and RNA (see Allison and Burke 1962). Some other comparisons between viruses and bacteria are given in Table 3.

Almost all of the many plant viruses that have been obtained in a highly

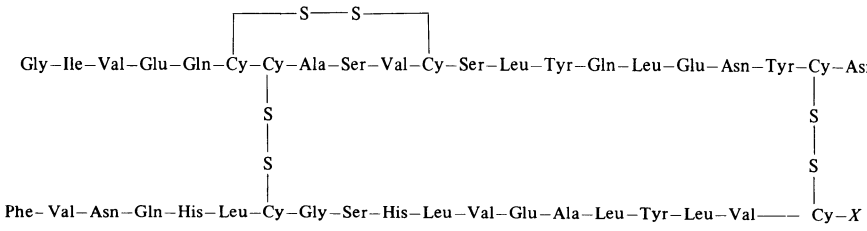
C. A. Knight, *Chemistry of Viruses*

purified state, as well as some small animal and bacterial viruses, have proved to be simple nucleoproteins. However, some of the larger animal and bacterial viruses also possess lipid, polysaccharide, and other components. It should be noted that a few plant and bacterial viruses have also been found to contain some lipid. The general compositions of some viruses are listed in Table 2.

A. Proteins

The structures of proteins, including those of viruses, are usually complex and hence require a series of studies to characterize them. A division of protein structure into four distinctive types can be made (see Kendrew 1959), each of which can be investigated experimentally:

1. *Primary structure*: the number of different peptide chains (determined mainly by end group analyses), interchain binding, if any (such as by S-S-bonds), and, most especially, the sequence of amino acid residues. These features of primary structure are illustrated by the formula for bovine insulin as determined by Sanger and associates (Ryle et al. 1955):



In order to save space, the last portion of the second peptide chain has been represented by *X*, in which *X* stands for -Gly-Glu-Arg-Gly-Phe-Phe-Tyr-Thr-Pro-Lys-Ala.

It can be seen from the formula that insulin has two peptide chains that are joined at two positions by disulfide bonds between cysteine residues, and the sequence of amino acid residues is as indicated. In this formula and in subsequent tables, the abbreviations commonly employed by protein chemists are used. A list of such abbreviations is given in Table 4.

2. *Secondary structure*: the geometric configuration of the peptide chain (or chains) with special reference to the presence or absence of helical structure. The commonest configuration of peptide chains is probably the so-called alpha helix of Pauling and associates (1951), which is illustrated in Figure 5. It will be noted that a spiral structure is favored by the formation of hydrogen bonds between adjacent -CO and -NH- groups.

3. *Tertiary structure*: the folding pattern of the peptide chain. Possibilities of folding (often the folding of a helical chain) are illustrated by

Table 2. Approximate Composition of Some Viruses.

Virus	% RNA	% DNA	Protein[a]	% Lipid	% Non-nucleic Acid Carbohydrate	Ref.[b]
Adenovirus		13	87			1
Alfalfa mosaic	19		81			2
Avian myeloblastosis	2		62	35	1[c]	3, 4
Blue-green algal LPP-1		48	52			5
Broad bean mottle	22		78			6
Brome mosaic	21		79			7
Carnation latent	6		94			8
Cauliflower mosaic		16	84			8
Coliphages f1, fd, M13		12	88			9
Coliphages f2, fr, M12, MS2, R17, QB, ZR	30		70			10
Coliphage ØX174, ØR, S13		26	74			9
Coliphages T2, T4, T6		55	40		5[d]	11, 12
Cowpea mosaic	33		67			8
Cucumber 3 (and 4)	5		95			13
Cucumber mosaic	18		82			8
Dasychira pudibunda L.	7		93			14
Encephalomyocarditis	30		70			15
Equine encephalomyelitis	4		42	54		16
Fowl plague	2		68	25	+	17
Herpes simplex		9	67	22	2	18, 18a
Influenza	1		74	19	6	19, 20
Mouse encephalitis (ME)	31		69			21
Pea enation mosaic	29		71			8
Poliomyelitis	26		74			22
Polyoma		16	84			23
Potato spindle tuber	100					24

Potato X	6		94			25
Reovirus	21		79		+	18
Rous sarcoma	2		62	35	1[c]	3, 4
Shope papilloma		18	82			23
Silkworm cytoplasmic polyhedrosis	23		77			26
Silkworm jaundice		8	77[e]			27
Simian virus 5	1		73	20	6	28
Southern bean mosaic	21		79			29
Tipula iridescent		13	82	5		30
Tobacco mosaic	5		95			31
Tobacco necrosis	19		81			32
Tobacco rattle	5		95			33
Tobacco ringspot	40		60			34
Tomato bushy stunt	17		83			35
Tomato spotted wilt	5		71	19	5	36
Turnip yellow mosaic	34		63			37
Vaccinia		5	88	4	3	38
Wild cucumber mosaic	35		65			39
Wound tumor	23		77			26

[a]Rounded figures, often obtained by difference between 100 percent and sum of other components.

[b](1) Green 1969; (2) Frisch-Niggemeyer and Steere 1961; (3) Bonar and Beard 1959; (4) Baluda and Nayak 1969; (5) Brown 1972; (6) Yamazaki et al. 1961; (7) Bockstahler and Kaesberg 1961; (8) Harrison et al. 1971; (9) Ray 1968; (10) Kaesberg 1967; (11) Kozloff 1968; (12) Thomas and MacHattie 1967; (13) Knight and Stanley 1941, Knight and Woody 1958; (14) Krieg 1956; (15) Faulkner et al. 1961; (16) Beard 1948; (17) Schäfer 1959; (18) Joklik and Smith 1972; (18a) Russell et al. 1963; (19) Ada and Perry 1954; (20) Frommhagen et al. 1959; (21) Rueckert and Schäfer 1965; (22) Schaffer and Schwerdt 1959; (23) Kass 1970; (24) Diener 1971; (25) Bawden and Pirie 1938; (26) Kalmakoff et al. 1969; (27) Bergold and Wellington 1954; (28) Klenk and Chopin 1969a; (29) Miller and Price 1946; (30) Thomas 1961; (31) Knight and Woody 1958; (32) Kassanis 1970; (33) Harrison and Nixon 1959a; (34) Stace-Smith 1970; (35) Stanley 1940, DeFremery and Knight 1955; (36) Best 1968; (37) Matthews 1970; (38) Fenner et al 1974; (39) Yamazaki and Kaesberg 1961.

[c]A rough estimate calculated from data of Baluda and Nyak (1969).

[d]Hydroxymethyl cytosine residues of the T-even DNAs are glucosylated but to different extents so that the glucose residues amount to about 4, 5, and 7 percent, respectively, of the DNAs of T2, T4, and T6 (Jesaitis 1956). An average value for DNA and carbohydrate is given here.

[e]About 15 percent of the virus was not accounted for as protein, nucleic acid, or lipid.

Table 3. Some Comparisons of Viruses and Bacteria.[a]

Microorganism	Size, nm (Approx. Diam.)	Chemical Composition	Multiplication	Inhibition by Antibiotics	Staining Characteristics
Bacteria	500–3,000	Complex: numerous proteins (including enzymes), carbohydrates, fats, etc.; DNA and RNA; cell wall contains mucopeptide	In fluids, artificial media, cell surfaces, or intracellularly, by binary fission	Inhibited	Stain with various dyes
Mycoplasmas or PPLOs[b]	150–1,000	Similar to other bacteria but generally possess no cell wall	In media similar to other bacteria but by budding rather than fission	Resistant to penicillins, sulfonamides; sensitive to tetracyclines, kanamycin, etc.	Stain with dyes but poorly
Rickettsia	250–400	Similar to other bacteria	Inside living cells by binary fission; major hosts: arthropods	Inhibited	Stain with various dyes
Chlamydia or *Bedsonia*	250–400	Similar to other bacteria	Inside living cells by binary fission; major hosts: birds and mammals	Inhibited	Stain with various dyes
Viruses	15–250	Mainly nucleic acid (one type) and protein. Some contain lipid and/or carbohydrate in addition	Inside living cells by synthesis from pools of constituent chemicals	Not inhibited	Stain for electron microscopy with salts of heavy metals

[a]Adapted from Knight 1974.
[b]PPLO is the abbreviation for pleuropneumonia-like organism, the first of this group of wall-less bacteria to be characterized.

Table 4. Names and Abbreviations of Common Amino Acids.

Alanine	Ala	Leucine	Leu
Arginine	Arg	Lysine	Lys
Asparagine	Asn	Methionine	Met
Aspartic acid	Asp	Phenylalanine	Phe
Cysteine	Cys	Proline	Pro
Glutamic acid	Glu	Serine	Ser
Glutamine	Gln	Threonine	Thr
Glycine	Gly	Tryptophan	Trp
Histidine	His	Tyrosine	Tyr
Isoleucine	Ile	Valine	Val

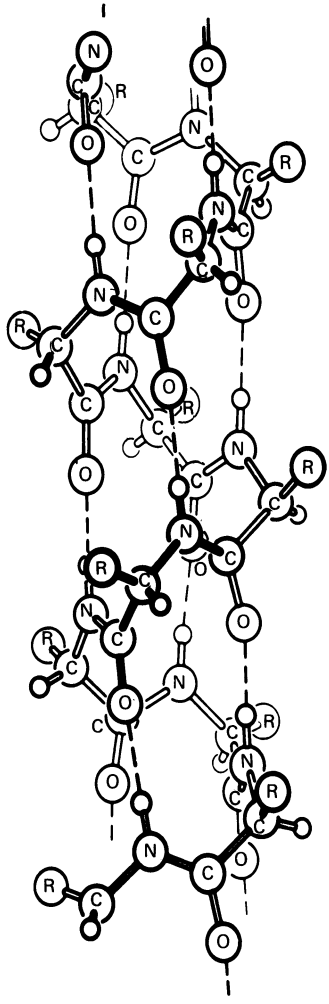


Fig. 5. The α-helix of Pauling et al. (1951). Note the formation of hydrogen bonds, indicated by dashed lines, between -CO and -NH groups. There are 3.7 amino acid residues per complete turn, and a unit residue translation of 1.47 Å giving a pitch of 5.44 Å. (From Kendrew 1959.)

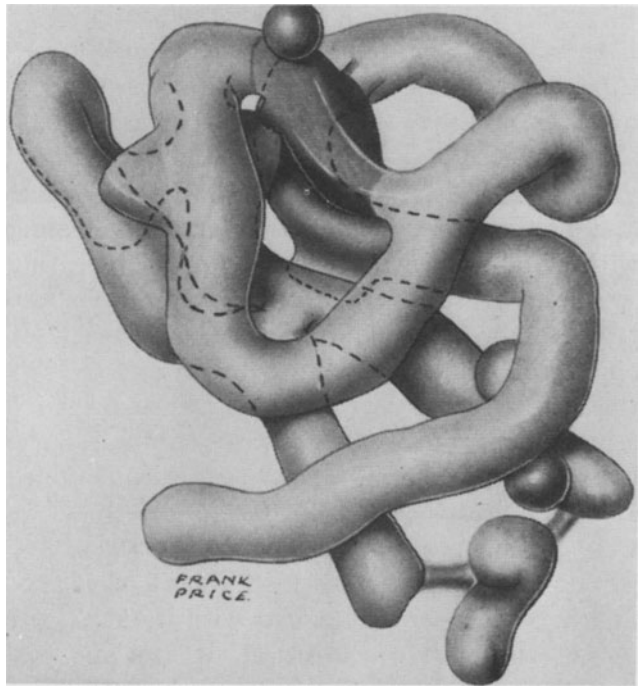


Fig. 6. Model of the myoglobin molecule derived from the three-dimensional electron density map based on x-ray data. The polypeptide chain is represented by solid rods; the side chains have been omitted but if present would fill the regions between the main chains. The small spheres are the heavy atoms used to determine the phases of the x-ray reflections. (From Kendrew 1959.)

the tertiary structure of myoglobin determined from crystallographic studies and represented by the model shown in Figure 6 (Kendrew 1959).

4. *Quaternary structure*: the number and spatial relationship of repeating subunits when these are present. An excellent example of quaternary structure is the architecture of the tobacco mosaic virus particle (see the model in Figure 7).

Primary structure is the stablest of the four types since this kind of structure involves covalent bonds, whereas the others depend on secondary attractions, that is, ionic and hydrophobic interactions and hydrogen bonds. The so-called "denaturation" of proteins results from disruption of secondary bonds and is manifested by the unfolding of the protein and changes in the solubility, charge, hydration, and serological and chemical reactivities. Rupture of secondary bonds affects quaternary as well as tertiary structure, and is the means for disaggregating virus particles, as will be described next.

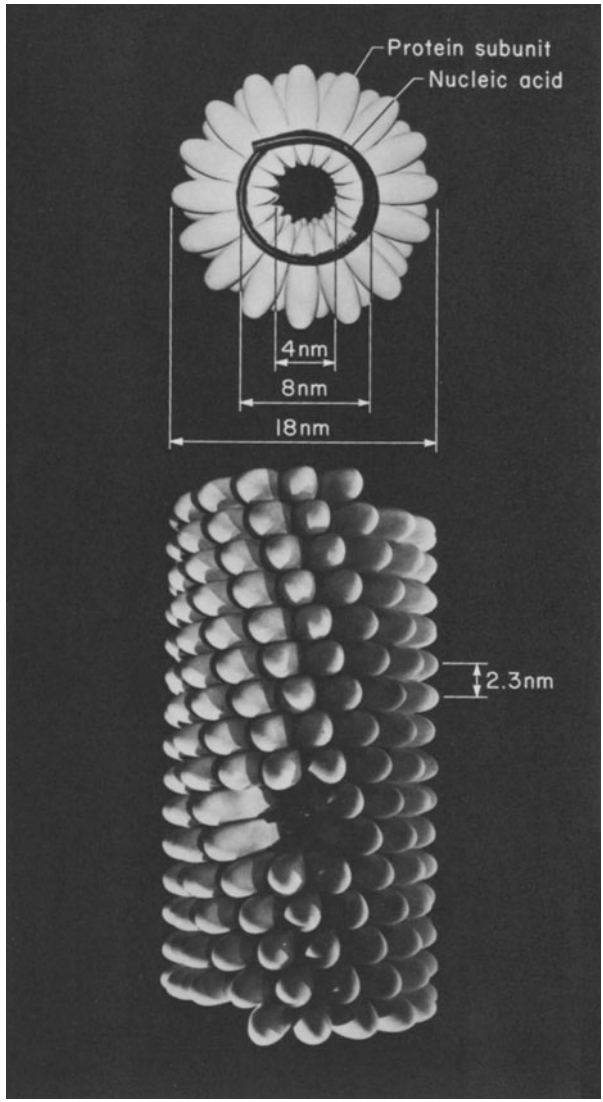


Fig. 7. Model of about one-tenth of the tobacco mosaic virus particle. The protein subunits are schematically illustrated in a helical array about the long axis of the particle. The structure repeats after 6.9 nm in the axial direction, and the repeat contains 49 subunits distributed over three turns of the helix of 2.3 nm pitch. Some of the subunits have been removed in order to show the deeply embedded location of the viral RNA, which, however, is better viewed in cross section as shown at the top of the illustration.

1. *Preparation of Viral Proteins*

A first step in determining primary structure is to obtain an accurate analysis of the amino acid content of the protein. This may be done with hydrolysates of the whole virus, but more reliable results are generally obtainable if the viral protein is first separated from the other viral constituents. This removes nonprotein materials whose acid-degraded products may interfere with the analytical methods, and eliminates spurious glycine that can arise from acid degradation of the nucleic acid (see Smith and Markham 1950). Isolation of proteins before analysis is mandatory for the many viruses that contain more than one species of protein.

Each virus presents a separate problem with respect to appropriate means for disrupting quaternary structure (without breaking covalent bonds) and subsequent isolation of homogeneous protein preparations. Obviously, the most complicated situations arise with such viruses as the myxoviruses (influenza, Newcastle disease, mumps) and bacterial viruses, which possess not one but several discrete protein components, as well as a variety of other chemical constituents. However, it has been found with several viruses that disruption of quaternary structure and resultant release of protein occur upon treatment of virus with acid, alkali, or/and detergent. Also, treatment with salts or with such reagents as urea, phenol, or guanidine hydrochloride have occasionally been used with success. Some of these methods will be summarized.

a. The Mild Alkali Method (Schramm et al. 1955; Fraenkel-Conrat and Williams 1955)

Exposure of TMV in the cold to dilute alkali at pH 10–10.5 results in release of protein, which, from its sedimentation and diffusion characteristics, appears to have a molecular weight of 90,000–100,000 (that is, it is either a pentamer or hexamer of the fundamental protein subunit). This material has been called "A protein" (Schramm et al. 1955). At the same time, the RNA of the virus undergoes alkaline hydrolysis to small fragments.

The quality of a good preparation of A protein is apparently so similar to that of the native protein in the virus particle that it can be used, together with viral RNA, to reconstitute (see section on Reconstitution of Viruses) virus rods that are virtually indistinguishable from undegraded virus.

The alkali employed appears not to be critical since sodium hydroxide, borate, carbonate, and glycine buffers have all been used with good results. The amino alcohols, such as ethanolamine, seem to be more effective than alkaline buffers in degrading the virus and hence achieve the desired result in a shorter time (Newmark and Myers 1957).

A convenient alkaline degradation procedure used successfully with TMV and many of its strains is as follows:

An aqueous solution of virus at 10 mg/ml is placed in a cellophane bag and dialyzed at about 4° against 2 liters of 0.1 M carbonate buffer (21.2 g Na_2CO_3 in 2 liters H_2O adjusted to pH 10.5 by addition of solid $NaHCO_3$) for 2–5 days. Alternatively, 0.02 M ethanolamine at the same pH and temperature can be used with a reduction of the dialysis time to 2–4 hr.

Separate undegraded virus by centrifugation of the contents of the dialysis bag at 60,000–100,000 *g* for 1 hr. Discard the pellet and add 1 vol of saturated ammonium sulfate to the supernatant fluid. Sediment the precipitated protein at about 5,000 *g* and redissolve in water. Precipitate twice more with 0.33 saturated ammonium sulfate and dialyze at 4° against several changes of distilled water. Adjust the pH of the dialyzed solution to 7–8 with dilute NaOH and centrifuge at 60,000–100,000 *g* for an hour to remove heavy particles. Store the final solution in the refrigerator, adding a drop or two of chloroform as preservative. The ultraviolet absorption ratio of the maximum to the minimum (280/250) should be about 2 or higher, depending mainly on the tryptophan and tyrosine contents of the protein.

The kinetics of the degradation of the quaternary structure of TMV in mildly alkaline solutions have been studied with the use of the ultracentrifuge and viscometry (Schramm 1947a; Schramm et al. 1955; Harrington and Schachman 1956). Some conclusions of general interest that Harrington and Schachman reached are:

1. The character of the alkaline degradation of TMV changes markedly between 0° and 25°, and different products are obtained at the two temperatures.
2. Contrary to earlier interpretations, some of the intermediates and some of the final products are the result not of degradation of larger components but rather of aggregation of smaller degradation products.
3. TMV seems to possess structural features of such a nature that the protein subunits are rapidly stripped at pH 9.8 from two-thirds of the particle leaving a relatively stable nucleoprotein fragment one-third the initial size of the particle.
4. Some particles of TMV seem to be completely resistant to degradation under conditions that lead to the breakdown of the bulk of the virus.

Conclusion 4 had also been reached earlier by Schramm et al. (1955), and remains a puzzle, along with the seemingly greater stability to alkali of one-third of the particle. Possibly the interaction between protein and nucleic acid is not uniform along the length of the rod, being stronger at one end than at the other.

b. The Cold 67 Percent Acetic Acid Method (Bawden and Pirie 1937; Fraenkel-Conrat 1957)

The proteins of some strains of TMV are too sensitive to alkali to permit their isolation from the virus at high pH values. In such cases the prepara-

tion of viral protein may be facilitated by use of the acid degradation procedure developed by Fraenkel-Conrat (1957), and described as follows:

To ice-cold virus solution of 10–30 mg/ml in water is added 2 vol of chilled glacial acetic acid, and the mixture is allowed to stand in an ice bath for 30–60 min, with occasional stirring. The nucleic acid separates out while protein stays in solution. The nucleic acid is removed by centrifuging in the cold and the protein is then dialyzed against several changes of distilled water at 4° for 2–3 days, by which time the protein may have reached its isoelectric range and have come out of solution. If the protein remains soluble, a series of precipitations with ammonium sulfate should be employed for completion of the preparation as described in the mild alkali method above. Otherwise, the isoelectrically precipitated protein is removed from the dialysis bag and pelleted by centrifugation; the pellet is dissolved in distilled water by adjusting to pH 8 with dilute NaOH and the solution is centrifuged at about 100,000 *g* for an hour to remove any undegraded virus or denatured protein. The water-clear supernatant, which contains the protein, may be used as such or dried from the frozen state and stored for use.

The acetic acid method did not work satisfactorily with turnip yellow mosaic virus (Harris and Hindley 1961) because insoluble protein aggregates were obtained. These aggregates were apparently formed by oxidation of SH groups to give S-S linkages between protein subunits. However, by converting the SH groups to carboxymethyl-SH with iodoacetic acid, subsequent aggregation was avoided, and the procedure could be carried out successfully as described above.

A useful modification of the acetic acid method is to substitute formic acid at 37° and to treat for 18 hr (Miki and Knight 1965).

c. The Guanidine Hydrochloride Method (Reichmann 1960; Miki and Knight 1968)

Some viruses are not readily dissociated by treatment with acid or alkali but do respond to protein-denaturing agents such as guanidine hydrochloride and urea. The former has proved more useful with both plant and animal viruses.

Dialyze virus at 10–20 mg/ml against 2 M guanidine HCl at room temperature overnight. During dialysis, the nucleic acid precipitates and the protein remains in solution. The nucleic acid is removed by centrifugation at 5,000 *g* and the supernatant is dialyzed against water for one or two days. The dialyzed material is centrifuged at 100,000 *g* for 2 hr in order to remove incompletely degraded virus. The yield of protein from potato virus X is about 90 percent.

A somewhat modified method is applied to poliovirus (Scharff et al. 1964) consisting of treatment with 6.5 M guanidine HCl at pH 8.3 for 3 hr at

37° followed by separation of protein and nucleic acid by centrifugation on a sucrose density gradient.

d. The Warm Salt Method (Kelley and Kaesberg 1962)

This procedure has been used successfully in preparing protein subunits from alfalfa mosaic virus, a small rodlike plant virus containing about 81 percent protein and 19 percent RNA.

Combine the virus at about 20 mg/ml in 0.01 M phosphate buffer at pH 7 with an equal volume of 2 M NaCl; hold the resultant mixture at 45° for 20 min. (During this time the solution becomes turbid, presumably because the released protein is less soluble than whole virus in M NaCl.) Cool immediately and centrifuge at low speed. The protein is sedimented while the nucleic acid and/or its degradation products remain in the supernatant fluid. Wash the protein pellet with M NaCl and centrifuge again. Dissolve the protein pellet in 0.01 M phosphate at pH 7 containing 0.005 M sodium dodecyl sulfate. Dialyze for 24 hr against distilled water and then precipitate by adding 0.66 vol of saturated ammonium sulfate. Resuspend the precipitate in phosphate buffer containing 0.05 M dodecyl sulfate, and then dialyze for 24 hr against phosphate buffer containing 0.005 M sodium dodecyl sulfate.

The molecular weight of the protein obtained by this method from alfalfa mosaic virus was estimated from sedimentation data to be about 34,000 (Kelley and Kaesberg 1962). The preparation was also found to be serologically active when tested with antiserum to whole alfalfa mosaic virus. This would not have been the case had the tertiary structure been extensively disrupted.

e. The Cold Salt Method (Yamazaki and Kaesberg 1963)

This procedure has been applied successfully to bromegrass mosaic and broad bean mottle viruses with recoveries of 60–80 percent of the viral protein.

The virus, at about 10–15 mg/ml in water, is dialyzed against 1 M $CaCl_2$ at pH 6–7 at 4° for 12 hr. During dialysis a white precipitate of nucleate is formed. The suspension is centrifuged at 5,000 *g* for 20 min. The supernatant is dialyzed against water to remove $CaCl_2$. The dialyzed material contains the soluble protein but practically no nucleic acid.

f. The Phenol Method (Anderer 1959a, 1959b)

Disruption of viruses with phenol is the basis for one of the commonest methods for isolation of viral nucleic acids (see section on Methods for Preparing Viral Nucleic Acids). In the preparation of nucleic acids the protein and other phenol-soluble components are usually discarded in the phenolic layer. However, it has been shown with TMV that protein can be

recovered from the phenolic layer, and, by suitable treatment, be restored to a condition which resembles the native state (Anderer 1959b).

To the phenolic layer remaining after separation of the aqueous, RNA-containing layer (see section on The Phenol Method for Preparing Nucleic Acid), add 5–10 vol of methanol and a couple of small crystals of sodium acetate. Remove the precipitate by centrifugation and wash three times with methanol and once with ether. Dry the product in air. To solubilize the air-dried protein, suspend 10 mg in 5 ml of water and heat at 60°–80°, adding enough 0.02 N NaOH to bring the pH to 7.5. The protein should dissolve and remain in solution upon cooling.

The TMV protein is denatured after extraction with phenol, precipitation with methanol, and so on, as described above. However, renaturation is assumed to occur to a large extent when the protein is warmed at about 60° at pH 7–7.5.

g. The Detergent Method and General Conclusions

It was early noted (Sreenivasaya and Pirie 1938) that 1 percent sodium dodecyl sulfate (SDS) disrupts TMV over a wide pH range. Later, when disruption by SDS was coupled with fractional precipitation with ammonium sulfate, Fraenkel-Conrat and Singer (1954) showed that the protein and nucleic acid of TMV could be rather cleanly separated. However, there are at least two disadvantages in the protein prepared by treatment of virus with SDS:

1. The protein strongly binds as much as 15 percent SDS, which seems to introduce only relatively small errors in ultracentrifuge studies (see Hersh and Schachman 1958, for example), but is more serious for other types of investigation. For example, trypsin is inhibited by anionic detergents such as SDS (Viswanatha et al. 1955), and TMV protein prepared by treatment with SDS does not appear to be satisfactory for structural studies dependent on a quantitative cleavage of protein by trypsin (Fraenkel-Conrat and Ramachandran 1959).

2. In addition to effecting the release of protein subunits, SDS tends to degrade the secondary and tertiary structures in a not readily reversed manner. Thus, TMV protein prepared with SDS is insoluble from pH 2 to 10 and does not participate in reconstitution (see section on Reconstitution of Viruses). However, the use of detergents, alone or in combination with other reagents such as phenol, has been invaluable in the disruption of virus particles for the isolation of viral nucleic acids (see Preparation of Nucleic Acids).

In general, all of the methods used to prepare viral proteins probably cause various denaturative changes, some of these reversible and others not. In the case of TMV, Anderer (1959b) has suggested the following criteria for distinguishing between native and denatured protein. Native TMV protein (1) is soluble in neutral aqueous media, (2) aggregates to

viruslike rods at pH 5–7, (3) reconstitutes to infectious virus with appropriate viral RNA, and (4) resembles the protein in the virus in amount of TMV antibody it binds. These criteria apply more or less to all viruses.

2. *Analysis of Viral Proteins*

The structure of proteins, as mentioned earlier, can be considered to fall into four main categories, each with its methods of analysis. Main consideration will be given here to the determination of primary structure because this is basic to the other types of structure. Thus the degree of helicity exhibited by the polypeptide chain (secondary structure), the nature of its folding (tertiary structure), and the assembly of protein subunits to form superstructures of characteristic morphology (quaternary structure) are virtually predestined by the sequence of amino acids in the protein chains. Some details of quaternary structure are considered in Sec. IV, Morphology of Viruses.

a. Amino Acid Analyses

In order to determine the composition of a viral protein, it is necessary, as with other proteins, to release the constituent amino acids by hydrolysis. This hydrolysis is usually accomplished by heating under vacuum 2–5 mg of protein in 1 ml of 6 N HCl in a thick-walled, sealed glass tube at 110° for 22–72 hr (see Knight 1964) (some proteins can be satisfactorily hydrolyzed at 120° for 6–24 hr; see Carpenter and Chramback 1962). Tryptophan and cysteine are largely destroyed by these conditions but can be preserved by modifying the hydrolysis medium (see Liu 1972; Liu and Inglis 1972). Alternatively, there are colorimetric procedures for determining tryptophan and cysteine (Anson 1942; Spies and Chambers 1949).

Customarily, two or more different times of hydrolysis are employed for evaluation of the release and recovery of individual amino acids. The highest value observed in a series or by extrapolation of observed values is generally accepted. Thus, as hydrolysis time increases, serine and threonine, and sometimes tyrosine, tend to be proportionately more destroyed; therefore, the contents of these amino acids are usually calculated by extrapolation back to zero time from the values observed at different times of hydrolysis. Conversely, when two or more residues of isoleucine or valine—and, to some extent, leucine—are contiguous, they are less readily released from peptide linkage than other amino acids. Thus even maximum periods of hydrolysis may yield somewhat low values. When the nature of the results suggests this, extrapolation to higher values is generally done by inspection and approximation. The uncertainties connected with such approximations are virtually eliminated if analyses can be made of the tryptic peptides of a protein (see Table 5) because the numbers of residues are

Table 5. Amino Acid Content of Tryptic Peptides of Tobacco Mosaic Virus.

Amino Acid	Tryptic Peptide[a] 1	2	3	4	5	6	7	8	9	10	11	12	Res.[b] Sum	Residue M. W. Sum
Ala	4	0	0	0	0	3	0	3	2	1	0	1	14	995.05
Arg	1	1	1	0	1	1	1	1	2	1	1	0	11	1,718.15
Asn	3	0	0	0	0	1	1	2	0	2	1	0	10	1,141.05
Asp	1	0	0	2	0	2	0	1	2	0	0	0	8	920.68
Cys	1	0	0	0	0	0	0	0	0	0	0	0	1	103.12
Gln	5	1	3	0	0	0	0	1	0	0	0	0	10	1,281.35
Glu	1	0	0	0	0	0	0	3	0	1	0	1	6	774.69
Gly	1	0	0	0	0	1	0	0	0	0	2	2	6	342.33
Ile	3	0	0	0	0	0	0	2	1	3	0	0	9	1,018.49
Leu	4	0	0	0	0	4	0	1	0	2	0	1	12	1,357.98
Lys	0	0	1	1	0	0	0	0	0	0	0	0	2	256.35
Phe	3	0	1	2	0	1	0	0	0	0	1	0	8	1,177.40
Pro	2	0	2	1	0	1	0	1	0	0	0	1	8	776.92
Ser	5	0	2	1	0	0	0	0	0	1	1	6	16	1,391.60
Thr	4	1	1	0	0	2	0	4	1	0	1	2	16	1,617.68
Trp	1	0	1	0	0	0	0	0	0	0	0	1	3	558.62
Tyr	1	0	0	0	1	1	0	0	0	0	1	0	4	652.70
Val	1	2	3	0	1	2	0	1	2	1	0	1	14	1,387.89
Totals													158	17,472.05[c]

[a]The 12 peptides resulting from treatment of tobacco mosaic virus coat protein with trypsin are numbered in order from the N-terminal to the C-terminal of the polypeptide chain.

[b]Res. = Residue. An amino acid residue is the molecular weight of the amino acid less one molecule of water.

[c]The N-terminal amino acid of TMV protein is acetylated. If the molecular weight of the acetyl group and a hydroxyl associated with the C-terminal threonine are added on to the sum of the residues, a value of 17,516.08 is obtained. For most practical purposes, a rounded value of 17,500 can be used for the TMV coat protein.

fewer and systematic errors are less significant. For example, if a peptide has 4 Ala residues, a 3 percent error in its analysis amounts to a negligible ± 0.1 Ala residue, whereas if the whole protein contains 20 Ala residues the error is ± 0.6 residue, which entails an uncertainty of ± 1 residue from the apparent value.

The amino acids resulting from hydrolysis of proteins are generally determined quantitatively in commercial, automatic amino acid analyzers such as that in Figure 8. These machines are an outgrowth of the laboratory models first developed and used effectively in amino acid analyses by S. Moore and W. H. Stein and associates at the Rockefeller University in New York (see Spackman et al. 1958; Spackman 1967).

In operation, something between a few microliters and a milliter or more of hydrolysate is applied at the top of a column of polysulfonic cation exchange resin and then appropriate buffers are pumped through resin

under high pressure. In passage through the resin, the various amino acids are repeatedly adsorbed and eluted at rates dependent upon their chemical composition. As the amino acid fractions emerge from the column, they undergo reaction with a ninhydrin solution and the resulting color is measured by passage through a colorimeter whose readings are recorded automatically on a chart such as that shown on the instrument in Figure 8.

The various amino acids are identified by the order in which they emerge from the column as indicated by the successive peaks observed on the recorder chart, and the quantity of each amino acid is obtained by comparison of the areas under the various peaks with the areas obtained with known quantities (for example, 0.1 μM) of standard amino acids. This comparison is made through a series of calculations as specified by the manufacturer of the analyzer, but is facilitated in some instruments by an integrator accessory that automatically integrates the areas under the respective peaks; there is also a system which translates recorder output to a

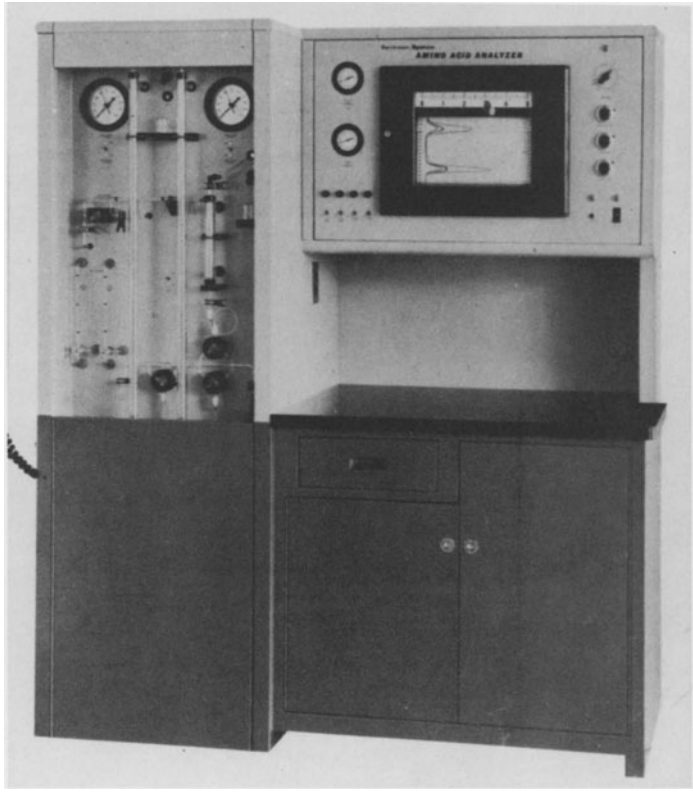


Fig. 8. Beckman amino acid analyzer. Note the cylindrical glass columns for the ion exchange resin on the left and the chart recorder on the right.

form suitable for subsequent processing in a digital computer (see Hirs 1967). Figure 9 illustrates the peaks obtained in a run made with a standard mixture containing 0.1 μM of each of the commonest amino acids.

The results from analyses made with amino acid analyzers are in terms of micromoles of each amino acid present in the applied sample. For maximum usefulness in structural and genetic analyses, these values are converted to numbers of each amino acid residue present per protein molecule (an amino acid residue is an amino acid minus the elements of water that are lost in the incorporation of the amino acid into a polypeptide chain). This conversion is done as follows.

First, a minimal molecular weight is calculated for the protein. In the absence of any estimate of the molecular size of the protein, this calculation involves a series of tentative assignments whose purpose is to find the lowest level at which the various amino acid residues appear in integral numbers. To start, a value of 1 can be assigned to the amino acid present in lowest amount (or the value of 1 can be assigned to any amino acid characterized by a consistently high recovery from protein hydrolysates) and the number of each of the other amino acid residues can be calculated on the basis of their relative micromolar values. Thus, by trial, a set of residue

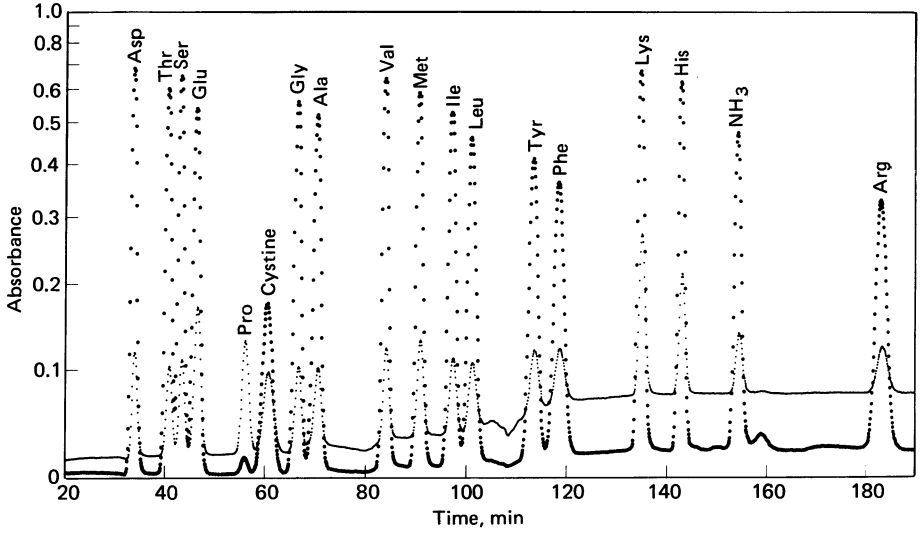


Fig. 9. Tracing from a Beckman amino acid analyzer chart showing the curves obtained with approximately 1 μM of each amino acid. The heavy dots represent the absorbance values at 570 nm and the small dots at 440 nm of the colors obtained in the reaction of each amino acid with ninhydrin. This color is purple for all common amino acids except proline, which yields a yellow product. Note that it took only about 3 hr for a complete analysis of the mixture of amino acids represented here.

numbers can be obtained in which the values are nearly whole numbers for all of the amino acids present. Summation of these residue weights yields a minimal molecular weight for the protein.

At this point, an independent estimate of the molecular size is needed in order to determine the factor by which the minimal numbers of residues must be multiplied to give the actual numbers present per molecule of viral protein. Some procedures to estimate molecular weights of proteins accurately enough for this use include (1) end group analysis (see Sec. IIIA, 2b), acrylamide gel electrophoresis (see Sec IIIA, 2c), agarose gel chromatography (see Sec. IIIA, 2c), and tryptic digestion followed by peptide mapping and counting (see Sec. IIIA, 2c). The molecular weight obtained by one or more of these methods divided by the minimal molecular weight yields a figure whose nearest integer is the factor by which the numbers of residues must be multiplied to give actual numbers per protein molecule and by which the minimal molecular weight must be multiplied to give the actual molecular weight.

In current practice, some estimate of the molecular weight of the protein is usually made early in the process by one of the methods listed above (for example, gel electrophoresis) and this figure is used with results of the amino acid analyses to make the approximations leading to assignment of residue numbers. This type of analysis can be illustrated by data obtained with the protein of tobacco mosaic virus (Table 6).

Similar procedures have been applied to the coat proteins of numerous plant viruses (see Tsugita and Hirashima 1972) and to some bacterial virus coat proteins; analysis of animal virus proteins in terms of amino acid residues has received almost no attention. Some examples of the results obtained with plant and bacterial viruses are given in Tables 7 and 8.

If the data in Tables 7 and 8 are compared with those for nonviral proteins (see, for example, Dayhoff 1972), it appears that viral proteins have ordinary quantities of the common amino acids and that no unusual amino acids have yet been observed in them. Furthermore, the general agreement between results of amino acid analyses made many years ago by microbiological assay (which detects only the L isomers, that is, the form generally present in proteins of all sorts) and more recent results obtained by methods that do not distinguish between optical isomers suggests that the amino acids of viral proteins are primarily if not exclusively the L isomers.

It is also apparent from the values shown in Tables 7 and 8 that viral coat proteins, while containing significant quantities of the basic amino acids arginine and lysine, do not have sufficient quantities of these to place them in the protamine or histone class of proteins. The latter often have been found in sperm and similar nucleoproteins. Actually, the isoelectric points reported for viruses are consistent with the idea that at least some viral proteins are acidic rather than basic. Some examples are given in Table 9.

Table 6. Amino Acid Analysis of Tobacco Mosaic Virus Coat Protein.

Amino Acid	Micromoles Amino Acid Found after Hydrolysis in 6N HCl at 120°C for				Maximum Micromoles	Number of Residues on Basis of			Best Estimate[c]
	6 hr	24 hr	48 hr	96 hr		Lys=2	Gly=6	Glu=16	
Asp	0.1773	0.1764	0.1755	0.1754	0.1773	18.96	18.21	18.31	18
Thr	0.1470	0.1337	0.1120	0.0953	0.1520[a]	16.26	15.61	15.70	16
Ser	0.1430	0.1119	0.0716	0.0493	0.1540[a]	16.47	15.82	15.91	16
Glu	0.1545	0.1547	0.1547	0.1549	0.1549	16.57	15.91	16.00	16
Pro	0.0799	0.0775	0.0827	0.0799	0.0799[b]	8.55	8.21	8.25	8
Gly	0.0584	0.0584	0.0582	0.0579	0.0584	6.25	6.00	6.03	6
Ala	0.1365	0.1349	0.1360	0.1348	0.1365	14.60	14.02	14.10	14
Val	0.1310	0.1386	0.1386	0.1385	0.1386	14.81	14.23	14.31	14
Ile	0.0744	0.0806	0.0822	0.0812	0.0822	8.79	8.44	8.49	9
Leu	0.1150	0.1185	0.1193	0.1184	0.1193	12.76	12.26	12.32	12
Tyr	0.0378	0.0384	0.0379	0.0354	0.0384	4.11	3.94	3.97	4
Phe	0.0763	0.0783	0.0789	0.0755	0.0789	8.44	8.11	8.15	8
Lys	0.0182	0.0176	0.0187	0.0187	0.0187	2.00	1.92	1.93	2
Arg	0.1045	0.1014	0.1028	0.0981	0.1045	11.18	10.74	10.79	11

[a]Values for Thr and Ser are obtained by extrapolation back to zero time of the plot of micromoles versus time of hydrolysis.

[b]The 48-hr value was discarded as anomalous.

[c]The best estimate is obtained by a series of test calculations. Such calculations were begun here by inspecting the list of maximum micromoles (column 5) of amino acid shown by the analyses and setting the amino acid occurring in smallest amount (Lys) equal to unity. When the residues of the other amino acids were calculated on the basis of Lys as 1, a value of 5.59 Arg was among those obtained. This gives a total of six or seven basic amino residues per protein subunit, and leads to the prediction that about this same number of peptides should result from tryptic digestion of the protein. In fact, nearly twice this number was observed. This suggests setting the Lys at 2, and the results of this assumption are shown in column 6. Inspection of these values shows that there are several that are far from integral and indicates that it might be more accurate to shift to an amino acid which is present in greater numbers of residues than Lys in order to reduce the effect of analytical error on the micromolar unit value. Glycine, which is known to be quantitatively released by and stable to acid hydrolysis conditions, was selected for the next test calculation (set at 6 from the first approximation) and the results yielded are shown in column 7. Values close to integral ones were obtained with almost all amino acids except isoleucine; when a shift was made to Glu as 16, the results were close to those obtained with Gly as 6 (column 8). Hence, the nearest integral values were assigned for each amino acid residue except isoleucine, which was assigned the next highest value on the basis of experience. The results are shown in the last column of the table. [It should be noted that neither histidine nor methionine is present in this protein but tryptophan and cysteine occur and as in proteins in general (see text) must be determined by independent analyses.]

b. Protein End Groups

A protein or peptide chain consists of a linear sequence of amino acid residues with two ends. In the usual manner of writing linear formulas for such structures, the amino acid residue on the extreme left is called the amino terminal, or N terminal, residue, whereas the amino acid residue on the extreme right is called the carboxyl terminal, or C terminal, residue. These terminal groups can be identified by cleaving them chemically or enzymatically from the protein chain followed by application of a method for amino acid analysis.

Knowledge about primary structure of viral proteins accumulated most rapidly with TMV, and the information obtained served as a model for the investigation of other viruses. For some years after Stanley's discovery of the nature of TMV, there was little interest in the primary structure of the virus protein, mainly because most investigators were overwhelmed with the idea of working with a protein whose apparent molecular weight was of the order of 38×10^6. However, the first step in primary structure work, namely, the determination of amino acid content of TMV and related viruses, was taken in Stanley's laboratory by Ross (Ross and Stanley 1939; Ross 1941) and was further developed in the same laboratory by Knight (Knight and Stanley 1941; Knight 1947b).

At the same time, evidence was gradually provided by physicochemical studies on the degradation products obtained from TMV with urea, alkali, and detergents (Stanley and Lauffer 1939; Lauffer and Stanley 1943; Wyckoff 1937; Schramm 1947a; Sreenivasaya and Pirie 1938) that the virus might possess a substructure. This idea was well supported by the classical x-ray diffraction studies that Bernal and Fankuchen (1941) made on TMV and other plant viruses. However, it was not clear from the results of the chemical degradation studies precisely what sorts of bonds were being broken, nor was it possible to define chemically the crystallographic subunits. Therefore the concept of viral protein subunits lay dormant until some years later.

A fresh approach to the question of viral subunits was launched with the attempt to determine the number and nature of the peptide chains in TMV by means of protein end group studies using the enzyme carboxypeptidase A.

Two pancreatic carboxypeptidases, A and B, carboxypeptidase C, found in citrus and a variety of other plants, and carboxypeptidase Y from yeast are known (see Ambler 1972; Hayashi et al. 1973). These differ in the rate at which they release specific terminal amino acids.

Carboxypeptidase A catalyzes the rapid release of terminal Ala, Gln, His, Ile, Leu, Met, Phe, Thr, Trp, Tyr, and Val; the slow release of Asn, Asp, Cys, Glu, Gly, Lys, and Ser; and generally fails to hydrolyze terminal Arg and Pro. In addition to their own refractory response to carboxypep-

Table 7. Amino Acid Residues Per Subunit of Some Plant Virus Coat Proteins.

Amino Acid	Virus									
	Brome Mosaic	Cowpea Chlorotic Mottle	Cucumber Mosaic	Cucumber 3 (Japan)	Cucumber 4	Potato X	Tobacco Mosaic	Tobacco Necrosis[d]	Tobacco Necrosis Satellite[e]	Turnip Yellow Mosaic
Ala	33	27	17	21	18	46	14	13	9	15
Arg	13	8	24	8	9	10	11	14	24	3
Asn							10			3
Asp	10[a]	11[a]	30	20[a]	17[a]	24[a]	8	18[a]	27[a]	7
Asx[b]										1
Cys	1	2	0	0	0	3	1	2	2	4
Gln							9			3
Glu	18[a]	17[a]	20[a]	10[a]	10[a]	19[a]	7	20	18	8
Glx[b]										3
Gly	10	10	16	9	5	13	6	8	8	8
His	4	2	4	1	0	2	0	1	6	3
Ile	8	7	16	7	5	12	9	11	13	15
Leu	15	16	26	18	13	10	12	10	20	17
Lys	13	13	18	4	3	12	2	12	11	7
Met	3	1	8	0	0	8	0	6	4	4

Phe	5	4	7	9	10	12	8	12	11	5
Pro	7	7	18	6	8	18	8	15	4	20
Ser	13	16	32	24	22	17	16	14	12	17
Thr	11	16	17	10	12	29	16	16	25	26
Trp	2	4	1	2	1	6	3	n.d.[c]	n.d.[c]	2
Tyr	5	5	11	4	4	2	4	11	6	3
Val	18	19	22	7	14	14	14	14	13	14
Total	189	185	287	160	151	257	158	197	213	188
Molecular weight	20,300	19,782	32,000	16,940	16,102	26,815	17,493	22,606	24,919	19,979
C-terminal	Arg	n.d.[c]	n.d.[c]	Ala	Ala	Pro	Thr	Ile	Leu	Thr
N-terminal	n.d.[c]	n.d.[c]	n.d.[c]	Acetyl-Ala	Acetyl-Ala	n.d.[c]	Acetyl-Ser	n.d.[c]	n.d.[c]	Acetyl-Met
References[f]	1	2	3	4,5	4	6,7	8-11	12	12	13,14

[a]Expressed as free acid but actually includes free acid and the relevant amide, since asparagine and glutamine were not determined as such.
[b]Uncertain whether residue is amide or free acid.
[c]n.d.: Not determined.
[d]American Type Culture 36 strain (AC 36 TNV).
[e]SV-C, Satellite associated with AC 36 TNV.
[f](1) Stubbs and Kaesberg 1964; (2) Chidlow and Tremaine 1971; (3) van Regenmortel 1967; (4) Tung and Knight 1972a; (5) Funatsu and Funatsu 1968; (6) Tung and Knight 1972b; (7) Miki and Knight 1968; (8) Anderer et al. 1960; (9) Anderer et al. 1965; (10) Tsugita et al. 1960; (11) Funatsu et al. 1964; (12) Uyemoto and Grogan 1969; (13) Harris and Hindley 1965; (14) Peter et al. 1972.

Table 8. Amino Acid Residues Per Subunit of Some Bacterial Virus Coat Proteins.

	Virus			
Amino Acid	Coliphage fd	Coliphage fr	Coliphage f2	Coliphage QB
Ala	9	16	14	15
Arg	0	4	4	7
Asn	0	10	11	8
Asp	3	4	3	7
Asx[a]		1		
Cys	0	2	2	2
Gln	1	5	6	8
Glu	2	6	5	5
Gly	4	9	9	7
His	0	0	0	0
Ile	4	6	8	4
Leu	2	5	8	12
Lys	5	7	6	7
Met	1	2	1	0
Phe	3	5	4	3
Pro	1	5	6	8
Ser	4	11	13	9
Thr	3	9	9	12
Trp	1	2	2	0
Tyr	2	4	4	4
Val	4	16	14	13
Total	49	129	129	131
Molecular weight	5,168	13,736	13,710	14,037
C-terminal	Ser	Tyr	Tyr	Tyr
N-terminal	Ala	Ala	Ala	Ala
References[b]	1	2	3	4

[a]Uncertain whether aspartic acid or asparagine.

[b](1) Asbeck et al. 1969; (2) Wittmann-Liebold and Wittmann 1967; (3) Weber and Konigsberg 1967; (4) Konigsberg et al. 1970.

tidase A when they are carboxylterminal amino acids, Arg, Asp, Cys, Glu, Gly, and Pro when in the penultimate position tend to decrease the rate of cleavage of C terminal amino acids that are normally readily released.

Carboxypeptidase B is effective mainly in the release of basic C terminal amino acids such as Arg and Lys.

Carboxypeptidase C catalyzes the hydrolytic cleavage of almost any C terminal residue, including Pro, although it is inefficient with such combinations as Pro-Pro and Pro-Gly, and also may not function on large polypeptides.

Carboxypeptidase Y catalyzes the release of most amino acids, including proline, from the C terminals of peptides and proteins (Hayashi et al. 1973). Glycine and aspartic acid may be released more slowly than other

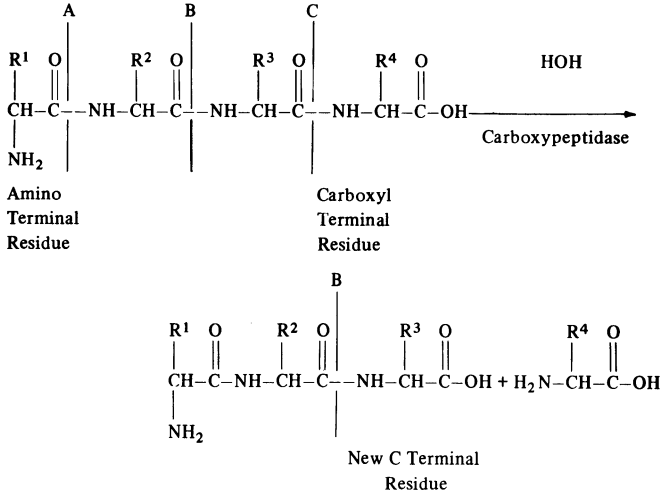
Table 9. Electrophoretic Isoelectric Points of Some Viruses.

Virus	Isoelectric Point pH	Reference
Alfalfa mosaic	4.6[a]	Lauffer and Ross 1940
Influenza A (PR8)	5.3	Miller et al. 1944
Shope papilloma	5.0	Beard and Wyckoff 1938 Sharp et al. 1942
Southern bean mosaic	5.9	MacDonald et al. 1949
Tobacco mosaic	3.5	Eriksson-Quensel and Svedberg 1936
Tomato bushy stunt	4.1	MacFarlane and Kekwick 1938
T_2 bacteriophage	4.2	Sharp et al. 1946
Turnip yellow mosaic	3.8	Markham and Smith 1949
Vaccinia	4.5	Beard et al. 1938
Wild cucumber mosaic	6.6	Sinclair et al. 1957

[a]Determined from minimum solubility rather than from electrophoretic measurements.

amino acids. The enzyme has the advantage of retaining activity in 6 M urea, which makes it possible to use it in structural studies on proteins whose C terminals are not exposed unless they are treated with chain-unfolding reagents such as urea.

The reaction of a peptide with carboxypeptidase A is illustrated by the following in which the R groups represent the side chains of common amino acids (such as H, CH_3, benzyl, and so on). The peptide bond at C is split and then, if conditions are favorable, the one at B.

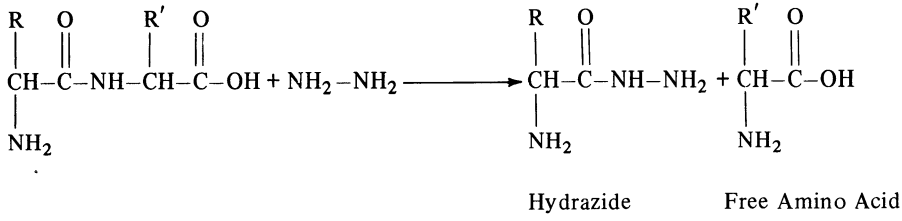


Another technique for determining the nature of C terminal amino acid residues is that employing hydrazinolysis (Akabori et al. 1956).

As shown by the following equation, in the hydrazinolysis reaction only the C terminal amino acid comes out as the free amino acid. All others are converted to hydrazides, which have different solubilities than the free amino acids and can be separated by extraction with appropriate organic solvents. The free amino acids can be identified by forming the dinitrophenyl derivatives and subjecting them to two-dimensional chromatography on paper. Elution of the spots and examination in the spectrophotometer provide quantitative values, which, after correction for the significant destruction that occurs during the hydrazinolysis step, are a measure of the C terminal end groups present.

Improved yields by use of a catalyst in the hydrazinolysis step and qualitative and quantitative determination of the free amino acids released by means of column chromatography enhance the usefulness of this technique (see Schroeder 1972).

The hydrazinolysis reaction may be illustrated with a dipeptide as follows:



The carboxypeptidase method has been singularly successful with TMV and strains, but has been disappointing with other viruses. For example, when applied to potato virus X, cucumber viruses 3 and 4, southern bean mosaic, tomato bushy stunt, and tobacco ringspot viruses, small, equivocal amounts of several amino acids were released from which no safe conclusions regarding the C terminals or numbers of subunits could be drawn (Knight 1955).

Three possible reasons can be suggested for this result: (1) The C terminal may contain sequences that are incompatible with the specificity of carboxypeptidase A. (2) More vigorous treatment (for example, more enzyme and/or higher temperature) may be needed. (3) The C terminal may be sterically unavailable to the enzyme. The following examples illustrate each of these situations.

C terminal proline is not cleaved by carboxypeptidase A. Potato virus X was found by the hydrazinolysis procedure (Niu et al. 1958; Miki and Knight 1968) to have a proline residue in the C terminal position, thus explaining the negative result with carboxypeptidase A.

Cucumber viruses 3 and 4 when treated with carboxypeptidase A at an enzyme:substrate ratio of about 1:400 at 25° yielded small amounts of several amino acids; and, while alanine seemed to be the major split product, even its concentration was so low as to make the result equivocal (Knight 1955). However, treatment at an enzyme:substrate ratio of 1:25 and at 37° clearly revealed alanine as the C terminal residue and threonine and serine as probable adjacent amino acids (Tung and Knight 1972a). The kinetics of the reaction are shown in Figure 10.

Two cases will serve to illustrate the steric hindrance possibility. No

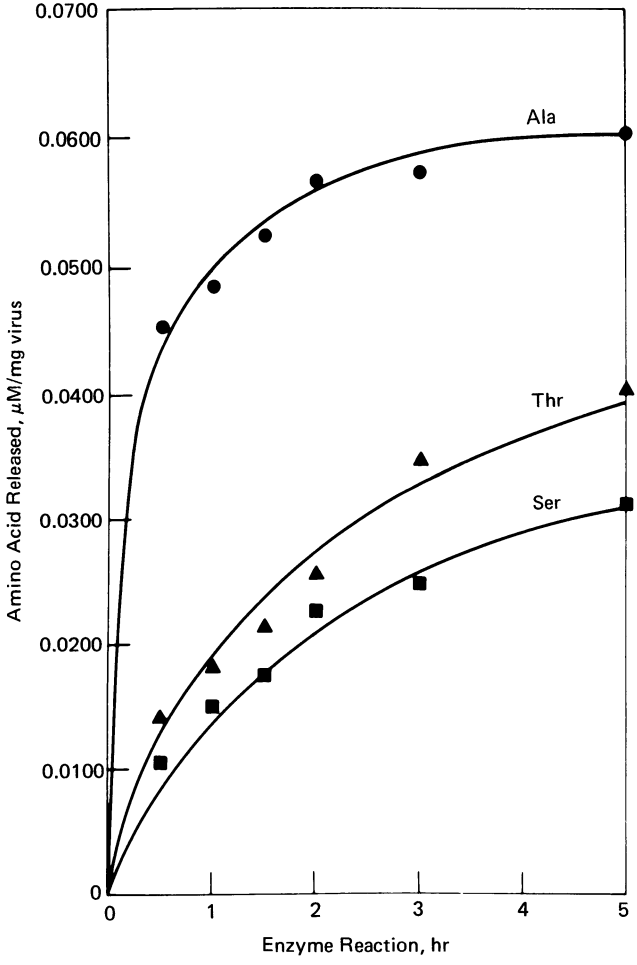
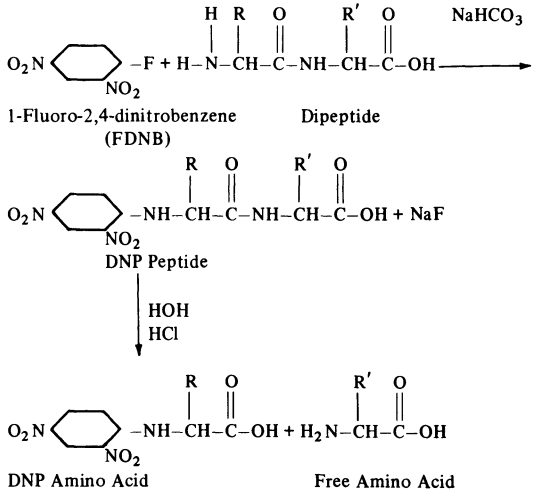


Fig. 10. Release of amino acids from Czech cucumber virus 4 by the action of carboxypeptidase A at an enzyme-substrate ratio of 1:25 at 37°. (From Tung and Knight 1972a.)

free amino acids were produced when intact turnip yellow mosaic virus was treated with carboxypeptidase, but when the isolated protein was employed, four different amino acids were released (Harris and Hindley 1961). From quantitative rate studies of the release, it was found that threonine was the C terminal amino acid, and that after removal of this, carboxypeptidase catalyzed the release from the peptide chain in succession—Ser, Thr, Val, and Asp. Another interesting case is that of the nitrous acid mutant TMV-171 (Tsugita and Fraenkel-Conrat 1960), from which carboxypeptidase caused the release of three amino acids, but from whose isolated protein the enzyme released about 15 residues.

There are several methods for determining the amino terminal (N terminal) residue of proteins and peptides, most of which, however, depend on the presence of an unsubstituted amino group. The two methods that have been used most in determining N terminal amino acid residues of viral proteins are (1) the Sanger fluorodinitrobenzene (FDNB) method (Sanger 1945, 1949; see also Knight 1964) and (2) the Edman phenylisothiocyanate (PTC) method (Edman 1950a, 1950b, 1956; see also Knight 1964).

The principle of the FDNB method is that the free amino group of the N terminal residue in a peptide or protein reacts with FDNB in mildly alkaline solution to give a *N*-dinitrophenyl (DNP) substituted residue. Upon acid hydrolysis, the DNP group remains attached to the N terminal amino acid in most cases, although some hydrolytic cleavage occurs with all DNP amino acids, being most severe with proline, glycine, and cysteine. The formation of the DNP derivative of a dipeptide and its subsequent hydrolysis can be illustrated as follows:



This general formulation becomes specific when R and R′ are replaced with H, CH_3, C_6H_5-CH_2, or other amino acid side chains.

The DNP amino acids are mainly extractable by ether and can thus be readily separated from nonterminal free amino acids, after which they are identified by paper chromatography and comparison with chromatograms of standard DNP amino acids. A standard chromatogram is given in Figure 11. The yellow color of DNP amino acids makes them easily located on chromatograms, and they can be quantitatively extracted and their concentrations determined from their light absorption in a spectrophotometer at 360 nm.

A procedure similar to the DNP method employs 1-dimethylaminonaphthalene-5-sulfonyl chloride ("dansyl" chloride) as the reagent. Dansyl (DNS) derivatives of the amino acids are very resistant to acid hydrolysis and show an intense yellow fluorescence that enables them to be detected at about 1 percent of the concentration needed for DNP amino acids. DNS derivatives can be identified by either electrophoresis or chromatography (see Bailey 1967; Gray 1972). The use of this technique is

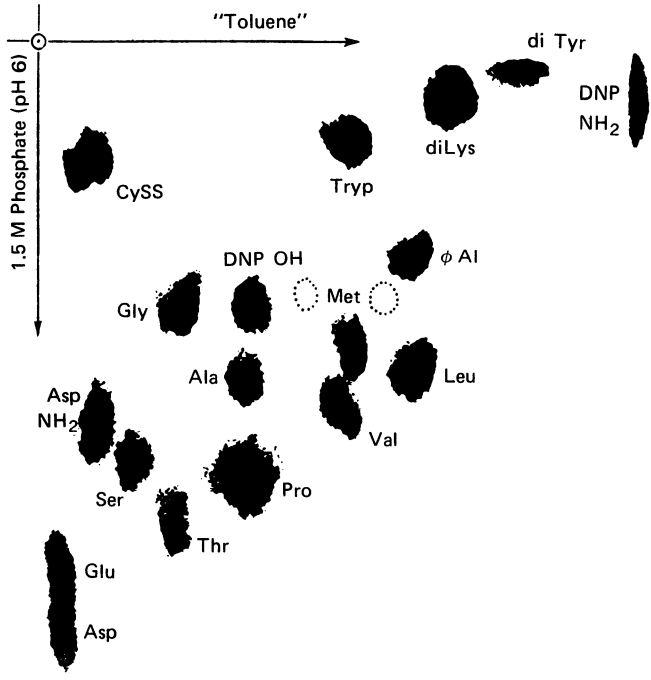
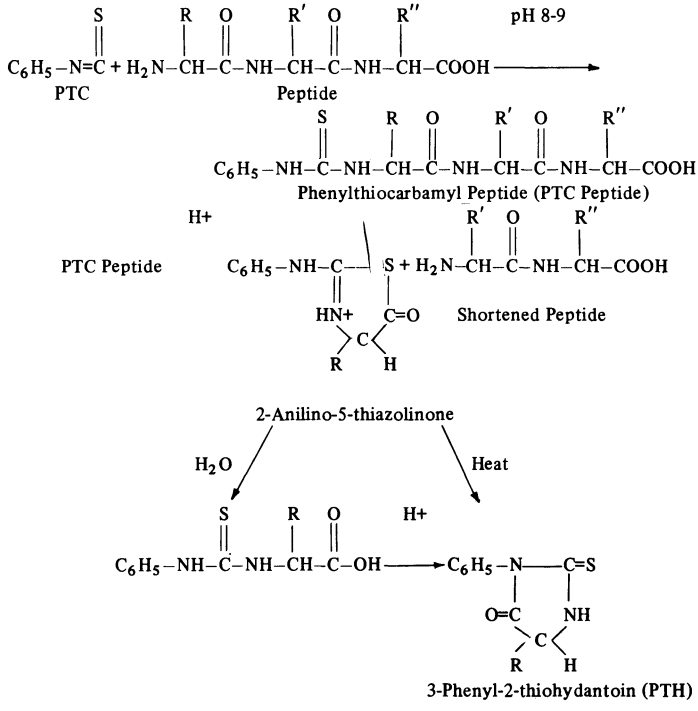


Fig. 11. Two-dimensional chromatogram of a mixture of 16 DNP-amino acids, dinitrophenol (DNPOH) and dinitroaniline (DNP NH_2). Approximately 0.02 μM of each was applied. (Fraenkel-Conrat et al. 1955.)

nicely illustrated by its application to the polypeptides of reovirus (Pett et al. 1973).

The major reactions of the Edman degradation, also called the phenylisothiocyanate (PTC) reaction, resemble the FDNB procedure in that an organic radical is first coupled to the protein at the N terminal in mildly alkaline solution and then the substituted N terminal amino acid is cleaved from the rest of the protein by treatment with acid. A milder acid treatment is used in the PTC method than in the FDNB procedure; hence, in contrast to the latter, the residual protein, shortened by one amino acid residue, is available for further stepwise degradative analysis. As in the case of the FDNB method, the procedure is applicable to peptides as well as to proteins. The chemical steps of the PTC reaction when applied to a peptide can be represented as follows:



As shown in the reaction scheme, the N terminal amino acid emerges as a 3-phenyl-2-thiohydantoin (PTH) derivative. The PTH derivative is extracted from the reaction mixture with an organic solvent in which the protein or peptide residue is insoluble. The quantity of PTH derivative can be estimated by reading its absorption in a spectrophotometer at the maximum (usually between 260 and 275 nm) and comparing the result with that

of an appropriate standard. Identification of the amino acid present as a PTH derivative can sometimes be made by paper chromatographic comparison with the standard PTH amino acids (Sjöquist 1953), or the PTH derivative can be hydrolyzed in acid to yield free amino acid, which then can be identified by any standard method for amino acid analysis. Various modifications of this procedure are used, including the substitution of potassium cyanate for phenylisothiocyanate (Stark and Smyth 1963; Stark 1972).

The methods just described for the determination of N terminal amino acid depend on the presence of a free amino group on the N terminal residue. However, it appears that many viral proteins have acetylated N terminal amino acids (see Table 10). Two methods have been employed in analyzing such N terminals:

1. The isolated viral protein is digested with pepsin or trypsin and the resulting peptides are fractionated on a strongly acidic cation exchange column. The peptide containing the acetylated amino acid as its N terminal will generally be least basic owing to the neutralization of its N terminal amino group by the acyl substituent. Therefore this peptide elutes from the column in one of the earliest fractions; it can be detected by the Folin colorimetric procedure and analyzed more or less readily by sequencing procedures (see Sec. IIIA, 2d). This approach was used successfully in identifying qualitatively the N terminal amino acids of TMV and cucumber virus 4 coat proteins (Narita 1958, 1959).

2. In the case of viral proteins acetylated at the N terminal, the amount of this terminal amino acid (and hence the molecular weight of the protein) can be estimated from the quantity of acetic acid released upon acid hydrolysis of the protein. The acetic acid can be determined readily by gas chromatography. Thus subunit molecular weights, but not the identity of the acetylated amino acids, were determined for tobacco mosaic and potato X virus proteins (Miki and Knight 1968). Some examples of the terminal amino acid residues found in some viruses are given in Table 9.

c. Protein Subunits

In 1952, Harris and Knight treated TMV with carboxypeptidase and found that more than 2,000 threonine residues were released from each mole of virus. The actual value from repeated determinations (Harris and Knight 1955) was about 2,320. (This figure is based on the now commonly used molecular weight for TMV of 40×10^6. The paper just cited gives a value of 2,900 based on a molecular weight for TMV of 50×10^6.) The key point, however, is that these results indicated that the TMV protein consists of over 2,000 polypeptide chains (protein subunits), and since these peptide chains all terminate in threonine it was correctly assumed, as subsequent data have shown, that they represent identical subunits. The alternative possibilities of a few chains ending in polythreonyl units or a

Table 10. The Protein Subunits of Some Viruses.

Virus	No. of Constituent Proteins	Approximate Molecular Weight of Subunits	N-Terminal	C-Terminal	Ref.[a]
Alfalfa (lucerne) mosaic	1	25,000	Ac-Ser	Arg	1,2
Broad bean mottle	1	20,500		Ala	3
Bromegrass mosaic	1	18,000		Arg	4
Cucumber 4	1	16,000	Ac-Ala	Ala	5,6,6a
Cucumber mosaic	1	24,000			7
Potato X	1	27,000	Ac-X[b]	Pro	8
Shope papilloma	1	40,000		Thr	9
Southern bean mosaic	1	30,000		Ser	10
Sowbane mosaic	1	19,000		Lys	11
Tobacco mosaic	1	17,500	Ac-Ser	Thr	12,13
Tobacco necrosis	1	30,000	Ala	Met	14
Tobacco necrosis satellite	1	20,000	Ala	Ala	15
Tobacco rattle	1	24,000			16
Tomato bushy stunt	1	41,000		Leu	17,18
Turnip yellow mosaic	1	20,000	Ac-Met	Thr	19
White clover mosaic	1	22,500			8

Adeno	9	7,500–120,000	20
Coliphage f2	2	14,000 and 39,000	21
Coliphage ØX174	4	5,000– 48,000	22
Coliphage T4	~28	11,000–120,000	23
Herpes simplex	~24	25,000–275,000	24
Influenza	~7	25,000– 94,000	25–27
Mouse Elberfeld	4	7,300– 33,000	28
Newcastle disease	~6	41,000– 74,000	29
Mouse mammary tumor	5	23,000– 90,000	30
Polio	4	6,000– 35,000	31
Polyoma	6	13,000– 43,000	32
Reovirus	~7	34,000–155,000	33
Rubella	~3	35,000– 62,000	34
Simian 40	6	9,300– 42,000	32
Vaccinia	~30	8,000–200,000	35

[a](1) Hull et al. 1969; (2) Kruseman et al. 1971; (3) Miki and Knight 1965; (4) Stubbs and Kaesberg 1964; (5) Narita 1959; (6) Niu et al. 1958; (6a) Tung and Knight 1972a; (7) Van Regenmortel 1967; (8) Tung and Knight 1972b; (9) Kass 1970; (10) Ghabrial et al. 1967; (11) Kado 1967; (12) Narita 1958; (13) Harris and Knight 1955; (14) Lesnaw and Reichmann 1969; (15) Reichmann 1964; (16) Offord and Harris 1965; (17) Tung and Knight, unpublished (from gel electrophoresis); (18) Niu et al. 1958; (19) Harris and Hindley 1961; (20) Maizel et al. 1968a, 1968b; (21) Hohn and Hohn 1970; (22) Burgess and Denhardt 1969; (23) Laemmli 1970; (24) Spear and Roizman 1972; (25) Compans et al. 1970; (26) Skehel and Schild 1971; (27) Schulze 1972; (28) Rueckert et al. 1969; Stoltzfus and Rueckert 1972; (29) Mountcastle et al. 1971; (30) Nowinski and Sarkar 1972; (31) Jacobson et al. 1970; (32) Hirt and Gesteland 1971; (33) Pett et al. 1973; (34) Vaheri and Hovi 1972; (35) Sarov and Joklik 1972.

[b]X = N-terminal amino acid residue unknown.

single huge polypeptide possessing threonyl side chains on ω carboxyl groups of aspartic or glutamic acid residues were eliminated by appropriate tests (Harris and Knight 1955). Also, application of the newly developed Akabori hydrazinolysis method (1956) confirmed the original conclusion that the virus protein consists of many peptide chains, each terminating in a single threonyl residue (Braunitzer 1954; Niu and Fraenkel-Conrat 1955).

There are several ways of calculating numbers of subunits in a viral protein from terminal amino acid data obtained as described in the preceding section. One of them is illustrated here, with the use of data from TMV analyses.

The molecular weights of TMV and threonine are 40×10^6 and 119, respectively. These molecular weights may be expressed in any units desired for the purposes of calculating relationships between TMV and its C terminal threonine. Thus the same end result will be obtained whether calculations are made using grams, milligrams, or micrograms. If micrograms are used, for example, the weight of one TMV particle (particle or molecular weight) is expressed as 40×10^6 μg, representing 1 μM of TMV. (The absolute weight of one TMV particle is, of course, the particle or molecular weight in grams divided by the number of particles in one mole, that is, by Avogadro's number. Thus the actual weight of one particle of TMV, assuming a molecular weight of 40×10^6, is 6.64×10^{-17} g.)

From 10^4 μg of TMV there were released by carboxypeptidase A 62 μg of threonine. Then

$$\text{Micromoles of threonine released} = \frac{\mu\text{g of threonine found}}{\text{molecular weight of threonine}}$$

$$= \frac{62}{119} = 0.52$$

$$\text{Micromoles of TMV used} = \frac{10^4}{40 \times 10^6} = 2.5 \times 10^{-4}$$

If 2.5×10^{-4} μM of TMV yielded 0.52 μM of threonine, 1 μM of TMV would yield $1 / (2.5 \times 10^{-4}) \times 0.52 = 2{,}080$ μM of threonine. Therefore, there must be 2,080 polypeptide chains (protein subunits) in the TMV particle. Since the virus is 95 percent protein, the molecular weight of each protein subunit is $(0.95 \times 40 \times 10^6) / 2080 =$ about 18,000.

There are many simple viruses like TMV that consist of a strand of nucleic acid ensheathed in a multisubunit protein coat. However, the large, tailed phages and many of the larger animal viruses have more than one species of protein in their particles. In these cases, it is common for each species of protein to be made up of identical subunits. Thus the major head protein of the T-even coliphages consists of about 2,000 identical subunits, each with a molecular weight of 40,000, while another protein, the tail sheath protein, is composed of 144 subunits, each of two species of

polypeptide, and so on (Mathews 1971). It is possible in some cases at least to isolate each species of protein from a virus particle and subject it to structural analysis, that is, analyze for amino acid content, do end group analyses, and so on. Often, however, it is useful just to determine the number of different polypeptide species present in the virus and their approximate molecular sizes. This can be done by applying the techniques of electrophoresis in acrylamide gel (Shapiro et al. 1967, 1969; Weber and

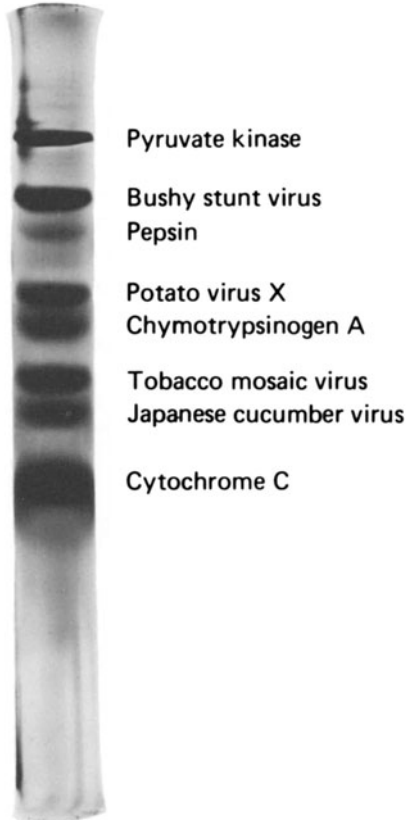


Fig. 12. Sodium dodecyl sulfate-polyacrylamide gel electrophoregram of some proteins ranging in molecular weight from 11,700 (cytochrome c) to 57,000 (pyruvate kinase). Electrophoretic migration, which was in 10 percent gel, is from top toward bottom in the illustration. The bands are visualized by staining in 0.25 percent Coomassie brilliant blue.

Osborn 1969; Dunker and Rueckert 1969) or gel chromatography (also called gel filtration) in agarose (Fish et al. 1969) to dissociated whole virus or to the protein fractions isolated from purified virus by one of the methods described in Sec. IIIA, 1.

In both techniques the molecular weights are determined by comparison of the migration rates of viral polypeptides with those of standard proteins whose molecular weights have been established by various means.

Figure 12 illustrates the migration of several plant virus proteins and standard proteins in SDS—10 percent acrylamide gels. The molecular weights assumed for the standard proteins and those calculated for the viral proteins shown in the gel of Figure 12 are pyruvate kinase, 57,000; tomato bushy stunt virus protein, 41,000; pepsin, 35,000; potato virus X protein, 27,000; chymotrypsinogen A, 25,700; tobacco mosaic virus protein, 17,500; Japanese cucumber virus 3 protein, 16,000; and cytochrome C, 11,700.

All of the plant viruses used in the illustration just given are characterized by a single species of polypeptide comprising the subunits of the viral coat protein. The electrophoresis of the polypeptides of influenza virus, which contains several different species, is illustrated in Figure 13. In this case, the whole virus was dissociated by treatment with SDS and the mixture was applied to the gel for electrophoresis. Duplicate gels are shown in order to illustrate how polypeptides and glycopolypeptides can

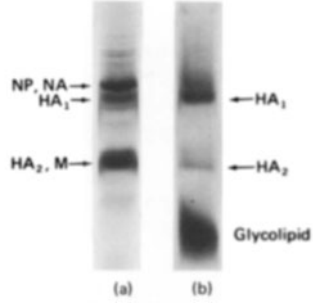


Fig. 13. Polyacrylamide gel electrophoregram of PR8 influenza virus dissociated in 1 percent sodium dodecyl sulfate and run on a 6 percent gel. Gel (*a*) was stained with Coomassie brilliant blue to reveal protein bands, and gel (*b*) was stained with *p*-rosanilin to detect carbohydrate (glycoprotein, two upper bands; glycolipid, lower band).

Abbreviations: NP, nucleoprotein (nucleocapsid subunit); NA, neuraminidase; HA_1, large hemagglutinin component; HA_2, small hemagglutinin component; M, membrane protein of viral envelope.

be distinguished after electrophoresis by application of different stains.

An extraordinarily useful modification of gel electrophoresis is the use of thin gel slabs that permit the side-by-side comparison of many samples (Reid and Bieleski 1968; Studier 1973). Location of protein bands (the procedure is also applicable with modifications to RNAs) is usually accomplished by autoradiography of the dried gels, appropriate isotopes having been previously introduced into the system under investigation. Quantitative distinctions between bands can be enhanced more readily in autoradiography than in the gel staining techniques simply by varying the length of time the recording film is exposed to the gel. Examples of such gels are shown in Figure 13A, which records the proteins extracted from *E. coli* cells infected with various combinations of coliphage P2 and its satellite P4 as well as certain mutants (Barrett and Calendar 1974; Lengyel et al. 1973, 1974). An excellent example of the application of the slab gel technique to the analysis of proteins in an animal virus system is the study by Honess and Roizman (1973) of herpes simplex proteins.

Some caution needs to be exercised with respect to molecular weight values for viral or other proteins obtained by SDS polyacrylamide gel electrophoresis. It has often been assumed that the rate of migration of polypeptides in SDS acrylamide gels depends solely on their molecular size. This assumption appears valid for many proteins. However, a rigorous application of the technique to plant virus proteins (Tung and Knight 1972a, 1972b, 1972c) indicates that the electrophoretic migrations of proteins of similar size in SDS polyacrylamide gels is a closely related function of their molecular weights only when the macromolecules under investigation have the same hydrodynamic shape and charge-to-mass ratio. This situation exists only when standard and test proteins react with SDS in a strictly comparable manner. The data summarized in Table 11 illustrate this point.

The values listed in the first column of the table, which were determined by amino acid analysis and peptide mapping as described earlier, represent the most accurate figures available. They probably deviate from the actual molecular weights by not more than one or two amino acid residues (± 100–200 daltons) and less than that for TMV whose complete amino acid sequence is known. A comparison of the molecular weight determined by other methods with that in the first column shows a good agreement for TMV (± 10 percent accuracy is usually ascribed to the gel electrophoresis and ± 7 percent accuracy is associated with gel chromatography results).

However, focusing on the SDS acrylamide gel values for all of the viruses listed in the table, it is apparent that close agreement with the actual values (column 1) was observed only with TMV and Japan CV3 proteins. The values indicated by gel electrophoresis are spuriously low for Berkeley CV3, Berkeley CV4, and Czech CV4 proteins. It seems likely that

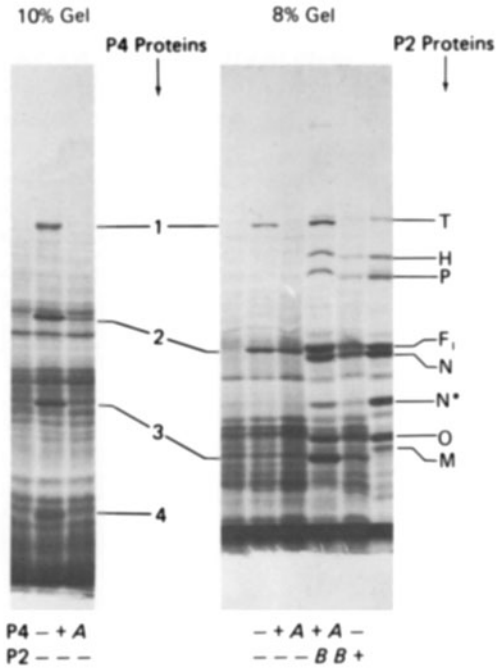


Fig. 13A. Use of slab gel electrophoresis and autoradiography to detect proteins made after infection of *Escherichia coli* bacteria with various combinations of coliphages P2 and P4 and mutants A and B. (Courtesy K. Barrett.)

E. coli C was irradiated with ultraviolet light in order to decrease the synthesis of host proteins. The cells were then infected with P2, P4, or with both together, and labeled with ^{14}C-reconstituted algal protein hydrolysate from 70 to 72 min after infection (*left panel*) or from 40 to 60 min after infection (*right panel*). The incorporation was stopped by adding an excess of cold amino acids and pouring the cells in ice. The cells were collected by centrifugation and lysed by holding in boiling water for 2 min in tris buffer at pH 6.8 containing SDS and mercaptoethanol (Laemmli 1970). The labeled proteins were analyzed in SDS polyacrylamide slab gels. The discontinuous gel system of Laemmli and Maizel as described by Laemmli (1970) was used with a 5 percent stacking gel and an 8 or 10 percent resolving gel. The slab gel apparatus was that of Studier (1973). The gels were dried for autoradiography according to Maizel (1971) and autoradiograms were made with Kodak No-Screen x-ray film.

Symbols: –, no phage; +, wild-type phage as indicated at the left; A or B, phage which contains mutations in genes A or B.

Table 11. Molecular Weights of Coat Protein Subunits of Tobacco Mosaic Virus and of Some Isolates of Cucumber Viruses 3 and 4 Determined by Different Methods.[a]

	Method of Determination			
Virus	Amino Acid Analyses and Peptide Mapping	C-terminal Analysis	SDS-Polyacrylamide Gel Electrophoresis	Agarose Gel Chromatography in Guanidine Hydrochloride
Tobacco mosaic	17,500	18,000	18,000	16,500
Berkeley CV3	17,100	10,100	14,200	16,000
Japan CV3	17,100	25,000	16,000	16,000
Berkeley CV4	16,100	13,300	14,200	16,000
Czech CV4	16,100	16,700	14,200	16,000

[a]Adapted from Tung and Knight 1972a.

these proteins retain enough tertiary structure in the presence of SDS to cause them to migrate at anomalous rates with respect to the standard proteins (and with respect to the proteins of TMV and Japan CV3). Consequently, as judged by SDS gel electrophoresis, the proteins of Berkeley CV3 and Japan CV3 appear to have significantly different molecular weights when, in fact, they are the same, and, conversely, the proteins of Berkeley CV3 and Berkeley CV4 appear to have the same molecular weight when, in fact, they are substantially different.

From these and other data, Tung and Knight have concluded that, as might be expected, the most accurate procedure for determining the molecular weights of viral and other polypeptides is to add up the weights of the constituent amino acid residues. In practice, this means that one determines the minimum number of residues in (and hence a minimum molecular weight of) the polypeptide from careful amino acid analyses and then determines the factor by which the minimum value must be multiplied to give the actual value. This factor is the integral number nearest to the quotient obtained by dividing the molecular weight estimated by gel electrophoresis or gel chromatography by the minimum molecular weight based on amino acid analysis. Alternatively, the factor can be deduced by comparing the number of peptides found on a map of the tryptic digest of the protein with the number expected from the arginine and lysine residues in the minimum molecular weight unit.

Obviously, the less accurate molecular weight values for viral polypeptides (protein subunits) obtainable by gel electrophoresis and gel chromatography are often sufficient and may be much more convenient to obtain since they do not require the chemically homogeneous product required for reliable amino acid analyses. Thus the choice of method for determining molecular weights of viral proteins will doubtless depend on conveni-

ence and degree of accuracy sought. In addition, it is clear that information other than molecular weight can be obtained by the various procedures employed, for example, numbers of different polypeptide species present, presence or absence of conjugated carbohydrate moieties, nature of terminal groups of the polypeptide chain, and so on.

Information about the numbers, size, and composition of polypeptides associated with each virus is still fragmentary. However, some data of this sort appear in Table 10.

It can be seen from the data in Table 10 that while there are some large viral polypeptides, most of them fall in the range of 14,000–50,000. As might be expected, the larger, morphologically complex viruses contain several species of polypeptides including some of the larger ones. The rather common occurrence of acylated N terminals is also illustrated.

d. Amino Acid Sequences

Most proteins contain basic amino acids such as arginine and lysine scattered along the length of the peptide chain. This fact, coupled with the marked specificity of the enzyme trypsin for bonds next to basic amino acids, provides a means for cleaving long peptide chains into more readily analyzed fragments.

A convenient survey of the peptides can be made by combining paper electrophoresis and chromatography in a "mapping" procedure. This is illustrated by the diagram of Figure 14, which shows a peptide map obtained after digestion of TMV protein with trypsin. However, the amounts of individual peptides in map spots seldom exceed 200 μg, whereas milligram amounts are usually required for sequential analyses. Therefore, countercurrent extraction or ion exchange chromatography is usually done to separate tryptic peptides for sequential analysis. The separation of the tryptic peptides from TMV protein by ion exchange chromatography is illustrated by the elution diagram shown in Figure 15.

The next step is to determine the amino acid sequences of each peptide. These steps can be illustrated by taking the peptide designated as 11 in Figure 15 and following the sequence determination made by Ramachandran and Gish (1959).

The purity of an aliquot of this fractionated peptide was checked by the mapping procedure. The spot labeled 11 in Figure 14 was the only major spot observed. N terminal analysis by the DNP method indicated that glycine was the N terminal residue. Analysis of the rest of the peptide by acid hydrolysis, formation of DNP derivatives, and so on (instead of the DNP procedure, it is also convenient to subject the hydrolysate to analysis directly in the automatic amino acid analyzer) showed amino acids present in the following molar proportions: Arg 1.00, Asp 1.14, Gly 0.98, Ser 1.00, Thr 1.00, Tyr 0.86. Hence the peptide is a heptapeptide with the formula Gly (Arg, Asp, Gly, Ser, Thr, Tyr). (In the formula the usual convention of

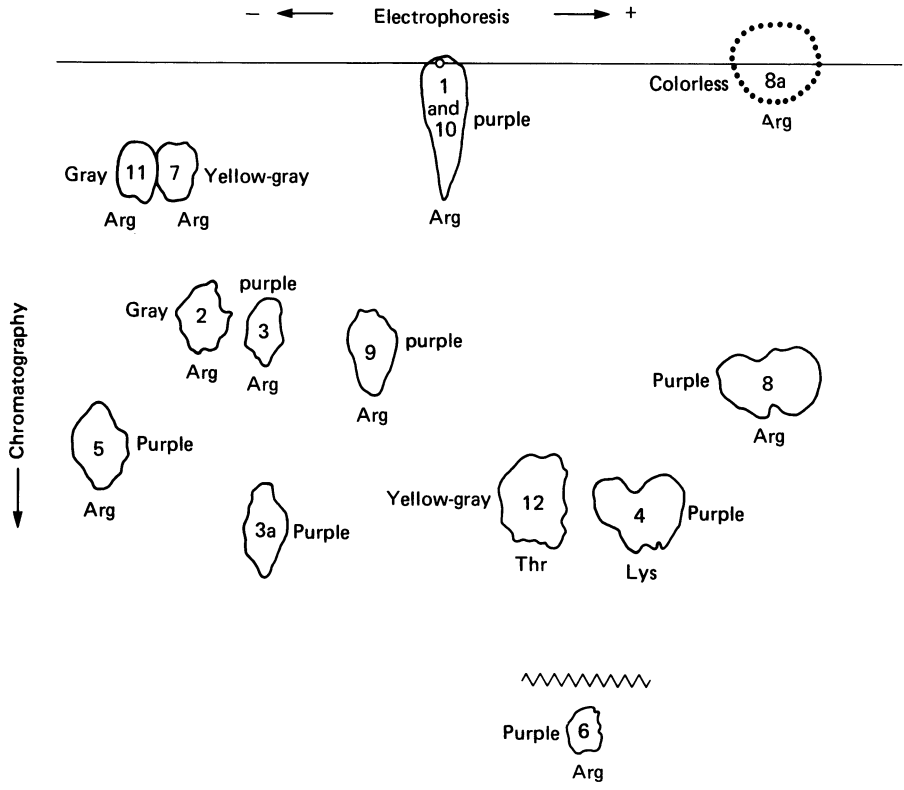


Fig. 14. Map of the peptides obtained by digestion of tobacco mosaic virus coat protein with trypsin. Adapted from Woody and Knight (1959) by deletion of four minor spots and addition of three major ones (tryptic peptides 1, 6, and 10) not shown on the original map. The latter were largely removed by precipitation at pH 4.5 prior to mapping in the experiments of Woody and Knight. Their migration in the mapping procedure was subsequently determined when the individual peptides became available. Peptide 6 travels much farther down the sheet than shown here, and the wavy line above it indicates that it has been brought up to a point more conveniently included in the diagram. Solid lines indicate ninhydrin-reactive spots that gave the colors indicated, and dotted lines signify a ninhydrin-negative but starch-iodine positive spot. The abbreviations for the amino acids—Arg, Lys, and Thr—indicate the nature of the C-terminal amino acid residue in a given peptide. See Table 12 for compositions of the peptides, which are numbered in order of their occurrence from the N-terminal to the C-terminal.

the protein chemists is used in which the amino acids whose sequences are unknown are placed in parentheses.)

Another portion of peptide 11 was treated with the enzyme leucine aminopeptidase, and aliquots were removed at various time intervals and analyzed by the DNP method. (Leucine aminopeptidase catalyzes hydrolytic cleavage of amino acids in a stepwise fashion from the N terminal end

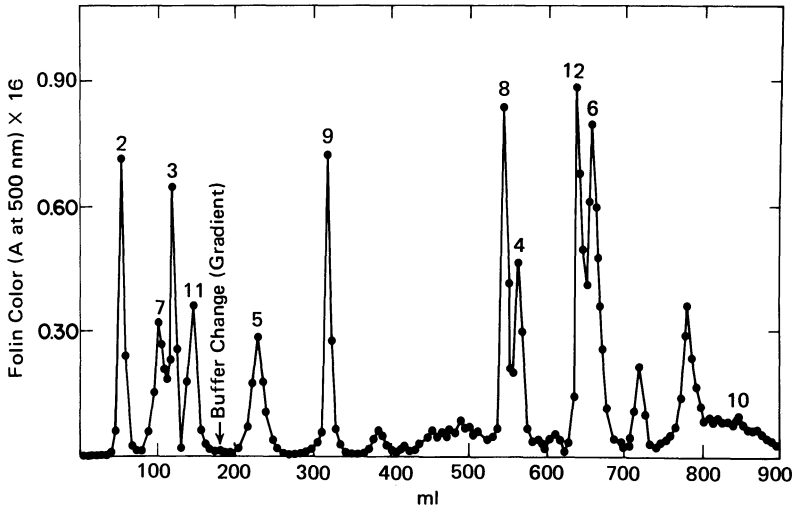


Fig. 15. Separation of tryptic peptides of TMV protein, after removal of material insoluble at pH 4.5 (mostly peptide 1) by passage through a column of Dowex 1-X2. The peptides were detected by their reaction with the Folin reagnet. The numbers are those of the tryptic peptides as listed in Table 12. (From Tsugita and Fraenkel-Conrat 1962.)

of peptide chains.) The results obtained are shown in Figure 16. These results indicate an N terminal order of Gly-Thr-Ser followed by asparagine and tyrosine in unknown order. It will be noted that the enzymatic degradation of the peptide revealed that one of the residues was asparagine rather than aspartic acid. Acid hydrolyses always yield the free acids rather than the amides, so that enzymatic hydrolyses are important in distinguishing between aspartic acid and asparagine and between glutamic acid and glutamine.

Another portion of peptide 11 was hydrolyzed with the enyzme chymotrypsin. Test analyses made by paper chromatography of portions of digest and by use of various indicator sprays indicated that the peptide was split rapidly into two peptides. One of these gave a positive test for arginine and the other a positive color reaction for tyrosine. Amounts of each peptide sufficient for analyses could be obtained by paper chromatography of the chymotryptic digest followed by elution of the separated peptides; a small strip of the chromatogram was reserved for spraying with ninhydrin in order to locate the spots. One of the peptides was found by application of the DNP method to contain equimolar amounts of aspartic acid (asparagine before hydrolysis) and arginine, the aspartic acid being N terminal. Hydrolysis of a portion of this peptide by leucine aminopeptidase yielded asparagine and arginine. Hence the peptide was Asn-Arg. A portion of the second peptide isolated from the chymotryptic digest of peptide 11 was

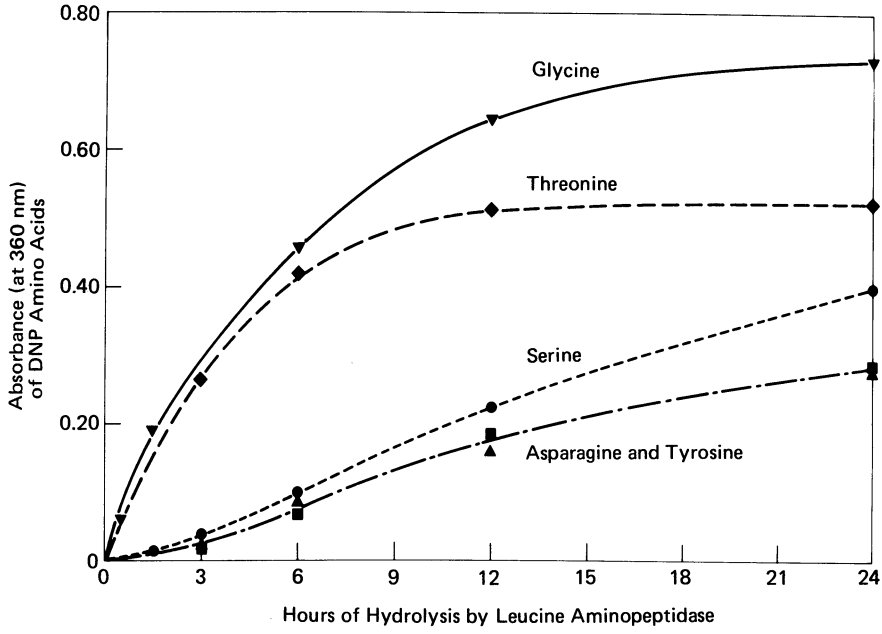


Fig. 16. Release of N-terminal amino acids from TMV tryptic peptide 11 by treatment with leucine aminopeptidase. The amino acids released at various times were identified as their DNP derivatives. (Adapted from Ramachandran and Gish 1959.)

treated with carboxypeptidase. Only tyrosine was released. Analysis of the remainder of the peptide by the DNP method revealed N terminal glycine and about equimolar proportions each of threonine, serine, and glycine. On the basis of these results and those obtained upon treating whole peptide 11 with leucine aminopeptidase, a sequence of Gly-Thr-Ser-Gly-Tyr could be assigned. Upon combination of the analytical results, the complete sequence for peptide 11 was found to be Gly-Thr-Ser-Gly-Tyr-Asn-Arg.

Twelve tryptic peptides were obtained from TMV protein and each was analyzed in a manner just outlined for peptide 11. The results are summarized in Table 12.

The final step in the primary structural analysis of TMV protein was to obtain "bridge peptides" whose sequences overlapped those of the peptides obtained by tryptic digestion. These were obtained by digesting portions of TMV protein with the chymotrypsin, pepsin, and subtilisin, and separating the resulting peptides by the same procedures used for the tryptic peptides. Analyses, or partial analyses, of the amino acid sequences of these new peptides provided information for linking the 12 tryptic peptides in the proper order to give the total sequence for the protein subunit. This procedure may be illustrated as follows.

Table 12. Peptides Obtained from TMV Protein by Tryptic Digestion.

Peptide Sequential Number	Number of Amino Acid Residues and Sequential Locations	Composition and Sequence	Color with Ninhydrin on Paper
1	41 (1-41)	Acetyl-Ser-Tyr-Ser-Ile-Thr-Thr-Pro-Ser-Gln-Phe-Val-Phe-Leu-Ser-Ser-Ala-Trp-Ala-Asp-Pro-Ile-Glu-Leu-Ile-Asn-Leu-Cys-Thr-Asn-Ala-Leu-Gly-Asn-Gln-Phe-Gln-Thr-Gln-Gln-Ala-Arg	Colorless
2	5 (42-46)	Thr-Val-Val-Gln-Arg	Gray
3	15 (47-61)	Gln-Phe-Ser-Gln-Val-Trp-Lys-Pro-Ser-Pro-Gln-Val-Thr-Val-Arg	Purple
3A	Same as 3, except for the preparative artifact of an N-terminal pyroglutamyl residue. See discussion by Gish (1960) of peptide K-0.66-A.		Purple
4	7 (62-68)	Phe-Pro-Asp-Ser-Asp-Phe-Lys	Purple
5	3 (69-71)	Val-Tyr-Arg	Purple

6	19 (72-90)	Tyr-Asn-Ala-Val-Leu-Asp-Pro-Leu-Val-Thr-Ala-Leu-Leu-Gly-Ala-Phe-Asp-Thr-Arg	Purple
7	2 (91-92)	Asn-Arg	Yellow-gray
8	20 (93-112)	Ile-Ile-Glu-Val-Glu-Asn-Gln-Ala-Asn-Pro-Thr-Thr-Ala-Glu-Thr-Leu-Asp-Ala-Thr-Arg	Purple
8A	Same amino acids as in 8. Exact structure unknown. Probably an N-terminal cyclic derivative of 8.		Colorless
9	10 (113-122)	Arg-Val-Asp-Asp-Ala-Thr-Val-Ala-Ile-Arg	Purple
10	12 (123-134)	Ser-Ala-Ile-Asn-Asn-Leu-Ile-Val-Glu-Leu-Ile-Arg	Purple
11	7 (135-141)	Gly-Thr-Gly-Ser-Tyr-Asn-Arg	Gray
12	17 (142-158)	Ser-Ser-Phe-Glu-Ser-Ser-Ser-Gly-Leu-Val-Trp-Thr-Ser-Gly-Pro-Ala-Thr	Yellow-gray

One of the peptides isolated from a chymotryptic digest of TMV protein was identified as Lys-Val-Tyr. There are two Lys residues in TMV protein, and one of them is in the tryptic peptide with a sequence -Lys-Pro-Ser (peptide 3 in Table 12). The other Lys is found in tryptic peptide 4 in a sequence ending in -Asp-Phe-Lys. This latter lysine must, therefore, be the one in the chymotryptic peptide, Lys-Val-Tyr. The only tryptic peptide with a Val-Tyr sequence is Val-Tyr-Arg. Hence the chymotryptic peptide bridges the two tryptic peptides listed in Table 12 as 4 and 5 and establishes their order in the TMV protein subunit.

By similar analyses, all of the tryptic peptides were located to give the total sequence of amino acids for the coat protein of common TMV and subsequently for other mutant strains. The sequences of four such strains are compared in Table 13.

The total number of amino acid residues in the coat proteins of many strains of TMV proved to be the same, 158. However, the coat proteins of some strains appear to reflect both additions and deletions in the viral genome (see Hennig and Wittmann 1972; Tung and Knight 1972a). Two amino acid deletions in the HR protein and their presumed location are indicated in Table 13. Admittedly, the relative evolutionary histories of these two strains are not known, and hence it can be argued that TMV was derived from an HR-like strain rather than the reverse. In that case, TMV protein would be viewed as possessing two additional amino acids rather than HR protein representing two deletions.

A comparison of the amino acid sequences of the fr, f2, and MS 2 bacteriophage coat proteins is shown in Table 14, and the sequence of the smallest known viral coat protein, that of fd phage, is given in Table 15 where it is compared with the protein of the closely related phage ZJ-2.

3. *Function of Viral Proteins*

The mass of most viruses is protein, much of which is located on the exterior of the virus particle. Thus situated, the protein comprises a coat or shell inside of which, or deeply embedded in which, lies the viral genetic material, the nucleic acid. Presumably, in the course of evolution, viruses whose genomes coded for a protein coat possessed survival value superior to those without such a structure. In any case, a noteworthy function of viral protein is the protection of the viral nucleic acid from destruction by nucleases or other degradative agents.

Another important function of viral protein is to mediate the process of infection, often determining the host specificity of a given virus. The basis for this action is that the first step in infection by many animal and bacterial viruses (but apparently not for plant viruses) is the attachment of a virus particle to a receptor site on a cell. A specific viral protein is involved in this attachment. In the case of tailed bacteriophages, the tail fibers, which

are protein, serve as specific attachment organs. Only phages whose tail fiber proteins have affinity for receptor sites on the bacterial envelope can attach and initiate infection. Spheroidal phages, such as the RNA-containing R17, f2, and fr phages, appear to possess a specific coat protein essential for initiation of infection even though they have no tails. Likewise, the coat proteins of animal viruses appear to be important in the capacity of these viruses to infect cells.

A striking example of the specificity that animal viral proteins can display is given by poliovirus. This virus when intact has a very restricted host range, namely, primate cells, and this restriction appears to be dependent on the specific affinity between poliovirus coat protein and receptor material on the primate cell surface (Holland 1964). However, when poliovirus RNA is used as the infectious agent, the host range of the virus is vastly expanded (for example, virus production occurs after intracerebral inoculations of mice, rabbits, guinea pigs, chicks, and hamsters) because the RNA gains entrance to cells by an inefficient, nonspecific mechanism. It should be noted that such "unnatural" infections are restricted because whole virus is produced in the cell initially infected by the RNA and whole virus can only attach productively to primate cells.

Viruses are good antigens. That is, when introduced into various animals either by injection or by infection they elicit the production of antibodies. These antibodies can react with viruses in a variety of immunologic and serologic ways (See vol. II, chap. 13 in Fenner 1968; Casals 1967; Matthews 1967). Viral proteins either alone or in some viruses as glycoproteins and lipoproteins are primarily responsible for such properties partly because they comprise a large part of the mass of virus particles and are exteriorly located, and especially because they are better antigens than other constituents of viruses.

Most of the larger and structurally more complex viruses contain enzyme constituents (these should be distinguished from the enzymes coded for by the viral genome but which do not become incorporated into the virus particles). Such enzymes are important protein constituents of viruses that have them. Functionally, they seem to fall mainly into two classes: enzymes that degrade cell envelope or membrane constituents (for example, phage lysozyme and influenzal neuraminidase) and those involved in viral nucleic acid synthesis (for example, the RNA transcriptase of reovirus and the RNA polymerase of Newcastle disease virus and the DNA polymerase called "reverse transcriptase" of Rous sarcoma virus) (Kozloff 1968; Webster 1970; Drzeniek 1972; Shatkin and Sipe 1968; Kingsbury 1972; Baltimore 1970; Temin 1970).

Protein kinases have been detected in the particles of several purified animal viruses, including numerous RNA tumor viruses, some viruses of the influenza and parainfluenza groups, vaccinia virus, and some herpes viruses (see Rosemond and Moss 1973). However, some viruses containing

Table 13. Sequence of Amino Acids in the Coat Proteins of Four Strains of Tobacco Mosaic Virus.[a]

	1	5	10	15
TMV[a]	Acetyl-Ser-Tyr-Ser-Ile-	Thr-Thr-Pro-Ser-	Gln-Phe-Val-Phe-Leu-	Ser-Ser-Ala-Trp-Ala-
D		-Ser-		-Val-
U2	Pro-	-Thr- -Asn-	-Tyr-	-Ala-Tyr-
HR	Acetyl-Ser-	-Asn- -Thr-Asn-Ser-Asn-	-Tyr-Gln-	-Phe-Ala-Ala-Val-Tyr-

	20	25	30	35
TMV	Asp-Pro-Ile-Glu-Leu-	Ile-Asn-Leu-Cys-	Thr-Asn-Ala-Leu-Gly-	Asn-Gln-Phe-Gln-Thr-
D		-Leu- -Val-	-Ser-Ser-	
U2	-Val-	-Ile- -Leu-	-Asn-Ala-	
HR	Glu- -Thr-Pro-Met-Leu-	-Gln-	—Val-Ser-	-Ser-Gln-Ser-Tyr-

	40	45	50	55
TMV	Gln-Gln-Ala-Arg-Thr-	Val-Val-Gln-Arg-	Gln-Phe-Ser-Gln-Val-	Trp-Lys-Pro-Ser-Pro-
D		-Thr- -Gln-	-Glu-	-Phe-
U2			-Ala-Asp-Ala-	-Ser-
HR	-Ala-Gly- -Asp-	-Arg-	-Asn-Leu-Leu-Ser-Thr-	Ile-Val-

	60	65	70	75
TMV	Gln-Val-Thr-Val-Arg-	Phe-Pro-Asp-Ser-	Asp-Phe-Lys-Val-Tyr-	Arg-Tyr-Asn-Ala-Val-
D		-Gly-Asp-Val-Tyr-		
U2	Val-Met-	-Ala-Ser-Asp-Phe-Tyr-		-Ser-Thr-
HR	Ala-Pro-Asp-Gln-	-Asp-Thr-Gly- -Arg-	-Val-Asn-Ser-Ala-Val-	

	80	85	90
TMV	Leu-Asp-Pro-Leu-Val-	Thr-Ala-Leu-Leu-Gly-	Ala-Phe-Asp-Thr-Arg-Asn-Arg-Ile-Ile-
D	-Ile-	-Thr-	
U2		-Asn-Ser-	
HR	Ile-Lys- -Tyr-Glu-	-Met-Lys-	

	95	100	105	110
TMV	Glu-Val-Glu-Asn-Gln-Ala-Asn-Pro-Thr-Thr-Ala-Glu-Thr-Leu-Asp-Ala-Thr-Arg-Arg-			
D		-Gln-Ser-		
U2	Glx- -Asx-Asx-Glx-Ala-Asx-		-(Asx,Thr,Glx,Glx,Pro,Val,Ile)-	
HR	Gln-Thr-Glu-Glu-Gln-Ser-Arg-	-Ser-Ala-Ser-Gln-Val-Ala-Asn-Ala-Thr-Gln-		

	115	120	125	130
TMV	Val-Asp-Asp-Ala-Thr-Val-Ala-Ile-Arg-Ser-Ala-Ile-Asn-Asn-Leu-Ile-Val-Glu-Leu-			
D				-Val-Asn-
U2		-Ala-Ser-		-Ala-
HR		-Ser-Gln-	-Gln-Leu-	-Leu-

	135	140	145	150
TMV	Ile-Arg-Gly-Thr-Gly-Ser-Tyr-Asn-Arg-Ser-Ser-Phe-Glu-Ser-Ser-Ser-Gly-Leu-Val-			
D	Val-	-Leu-	-Gln-Asn-Thr-	-Met-
U2		-Met-Phe-	-Ala-Gly-	-Thr-Ala-
HR	Ser-Asx-His-Gly-	-Tyr-Met-	-Arg- -Glu-	———Ala-Ile——— -Pro-

	155	158
TMV	Trp-Thr-Ser-Gly-Pro-Ala-Thr	
D	-Ala-	-Ser
U2	-Thr-Thr-	-Thr
HR	-Ala-	

[a] Adapted from Wittmann-Liebold and Wittmann 1967.

[b] TMV, common (vulgare) strain of tobacco mosaic virus; D, Dahlemense strain; U2, a mild strain; HR, Holmes' ribgrass strain.

Table 14. Sequence of Amino Acids in the Coat Proteins of Three Strains of Bacteriophage.[a]

```
     1                   5                  10                  15
fr   Ala-Ser-Asn-Phe-Glu-Glu-Phe-Val-Leu-Val-Asn-Asp-Gly-Gly-Thr-Gly-Asp-Val-
f2                    -Thr-Gln-                                       -Asn-
MS2                   -Thr-Gln-                 -Asp-Asn-

          20                  25                  30                  35
fr   Lys-Val-Ala-Pro-Ser-Asn-Phe-Ala-Asn-Gly-Val-Ala-Glu-Trp-Ile-Ser-Ser-Asn-
f2   Thr-
MS2  Thr-

                40                  45                  50
fr   Ser-Arg-Ser-Gln-Ala-Tyr-Lys-Val-Thr-Cys-Ser-Val-Arg-Gln-Ser-Ser-Ala-Asn-
f2                                                                    -Gln-
MS2                                                                   -Gln-

     55                  60                  65                  70
fr   Asn-Arg-Lys-Tyr-Thr-Val-Lys-Val-Glu-Val-Pro-Lys-Val-Ala-Thr-Gln-Val-Gln-
f2                       -Ile-                                        -Thr-Val-
MS2                      -Ile-                                        -Thr-Val-

          75                  80                  85                  90
fr   Gly-Gly-Val-Glu-Leu-Pro-Val-Ala-Ala-Trp-Arg-Ser-Tyr-Met-Asn-Met-Glu-Leu-
f2                                                      -Leu-    -Leu-
MS2                                                     -Leu-

               95                  100                 105
fr   Thr-Ile-Pro-Val-Phe-Ala-Thr-Asx-Asp-Asp-Cys-Ala-Leu-Ile-Val-Lys-Ala-Leu-
f2             -Ile-               -Asn-Ser-       -Glu-                -Met-
MS2            -Ile-               -Asn-Ser-       -Glu-                -Met-

         110                 115                 120                 125
fr   Gln-Gly-Thr-Phe-Lys-Thr-Gly-Ile-Ala-Pro-Asn-Thr-Ala-Ile-Ala-Ala-Asn-Ser-
f2          -Leu-Leu-   -Asp-    -Asn-Pro-Ile-Pro-Ser-
MS2         -Leu-Leu-   -Asp-    -Asn-Pro-Ile-Pro-Ser-

            129
fr   Gly-Ile-Tyr
f2
MS2
```

[a]Adapted from Wittmann-Liebold and Wittmann 1967; Min Jou et al. 1972.

a protein kinase, such as some of the RNA tumor viruses, do not have phosphate groups in their structural proteins, whereas the structural proteins of some viruses, such as simian virus 40, are all phosphoproteins despite the fact that the virus particles have no kinase (Tan and Sokol 1972). Therefore, the origin and function of these enzymes are unclear, although it has been suggested that production of phosphoproteins may be involved in the regulation of viral transcription.

Table 15. Amino Acid Sequences of The Coat Proteins of Bacteriophages ZJ-2 and fd.[a]

	1	5	10
ZJ-2	Ala-Glu-Gly-Asp-Asp-Pro-Ala-Lys-Ala-Ala		
fd			
	15		20
ZJ-2	Phe-Asp-Ser-Leu-Gln-Ala-Ser-Ala-Thr-Glu		
fd			
	25		30
ZJ-2	Tyr-Ile-Gly-Tyr-Ala-Trp-Ala-Met-Val-Val		
fd			
	35	40	
ZJ-2	Val-Ile-Val-Gly-Ala-Ala-Ile-Gly-Ile-Lys		
fd	-Thr-		
	45		50
ZJ-2	Leu-Phe-Lys-Lys-Phe-Thr-Ser-Lys-Ala-Ser		
fd			

[a]From Asbeck et al. 1969; Snell and Offord 1972.

B. Nucleic Acids

Nucleic acids are so named because they are acidic substances that were first isolated from the nuclei of cells.[1] It is now known that nucleic acids occur in both the nuclei and cytoplasm of all cells. The two major types of nucleic acid found in nature, ribonucleic acid (RNA) and deoxyribonucleic acid (DNA), both occur in viruses. However, in contrast to bacteria and other organisms, no virus appears to contain both RNA and DNA. The type of nucleic acid present can be determined by qualitative tests for sugar and pyrimidine components since it is only with respect to these constituents that RNA and DNA differ in composition. The detection of deoxyribose and thymine indicate DNA, whereas the presence of ribose and uracil denote RNA (procedures for analysis of these substances are

[1]The properties of nucleic acids in general apply to viral nucleic acids. The interested student may wish to refer to such comprehensive reference works as *Progress in Nucleic Acid Research and Molecular Biology*, J. N. Davidson and W. E. Cohn, editors, New York: Academic Press (published annually since 1963); *Procedures in Nucleic Acid Research*, Vol. 1 and 2, G. L. Cantoni and D. R. Davies, editors, New York: Harper and Row (Vol. 1 in 1966 and Vol. 2 in 1971); *The Chemistry of Nucleosides and Nucleotides*, A. M. Michelson, New York: Academic Press (1963); *Genetic Elements—Properties and Function*, D. Shugar, editor, New York: Academic Press (1967); *Methods in Enzymology—Nucleic Acids*, L. Grossman and K. Moldave, editors, New York: Academic Press (Vol. XII, 1967; Vol. XII, Part B, 1968; Vol. XX, Part C, 1971; Vol. XXI, Part D, 1971; Vol. XXIX, Part E, 1974; Vol. XXX, Part F, 1974.

described by Ashwell 1957; Schneider 1957; Lin and Maes 1967; Burton 1968; Hatcher and Goldstein 1969).

The quantity of nucleic acid, while fairly constant within a given group of viruses, varies considerably among different viruses. The extremes are represented by 0.8 percent RNA in influenza viruses and about 56 percent DNA in coliphage lambda. Viral nucleic acids, like those from other sources, are elongated, threadlike molecules. Some of them are single stranded, some double stranded, and some are cyclic. The amount and type of nucleic acid found in some viruses are given in Table 16.

1. Preparation of Viral Nucleic Acids

Viral nucleic acid is deeply embedded in the protein matrix of the virus particle. Despite this sheltered location, the nucleic acid is accessible to some chemical reagents such as mustards, nitrous acid, formaldehyde, and smaller molecular species in general. Nevertheless, for many experiments, it is desirable to isolate the nucleic acid from the rest of the material. No single procedure has proved universally successful for this purpose. However, reagents noted for an ability to break secondary valence bonds, such as salt linkages and hydrogen and hydrophobic bonds, have been most effective in disaggregating virus particles with release of the nucleic acid.

To best study the properties and function of viral nucleic acids, it has become ever more important to isolate the intact nucleic acid, to the extent that this exists, from virus particles. Three main factors work against this objective: (1) mechanical shearing of the nucleic acid during isolation, (2) chemical degradation at the extreme pH values that favor removal of protein coats, and (3) enzymatic degradation.

Mechanical shearing is a problem primarily with large DNA molecules such as those found in phages. Violent mixing or even forceful pipeting of solutions of phage DNA are sufficient to rupture the molecules (Hershey et al. 1962) and breakage may also be accompanied by denaturation, that is, strand separation, under certain conditions of temperature and salt concentration (Hershey et al. 1963). Therefore, gentle stirring procedures are recommended in the isolation of nucleic acids.

The sensitivities of RNA and DNA to extreme pH values differ somewhat but the structures of both types of nucleic acid may be irreversibly altered at pH values below 3 or above 10. Below pH 3, depurination (cleavage of adenine and guanine) tends to occur with double-stranded molecules. RNA is subject to alkaline hydrolysis above pH 10, and DNA, while resistant to alkaline hydrolysis, may be denatured above pH 12. Consequently, most nucleic acid isolation procedures are performed at intermediate pH values.

Probably the greatest hazard to intact viral nucleic acid is attack by nuclease enzymes, that is, by ribonucleases and deoxyribonucleases.

Traces of nucleases can often be detected even in the most highly purified preparations of viruses; and while several of the methods for isolating nucleic acid include provision for protecting the product from nucleases, none of the procedures is entirely satisfactory in this regard. In principle, use of strong protein denaturants in removing the viral coat protein and releasing nucleic acid will also eliminate accompanying nucleases. In practice, however, unless the removal of denatured protein is complete, traces of nuclease will remain with the nucleic acid and subsequently become renatured and active (Ralph and Berquist 1967).

Since complete removal of denatured protein is difficult to ensure, one needs to combat nuclease activity in preparing viral nucleic acids by starting with the most highly purified virus obtainable, avoiding prolonged procedures, and by adding nuclease inhibitors. Two such inhibitors are the acid clay bentonite (Fraenkel-Conrat, et al. 1961; Singer and Fraenkel-Conrat, 1961) and diethyl pyrocarbonate (Solymosy et al. 1968; Bagi et al. 1970). Some disadvantages of these nuclease inhibitors are that bentonite adsorbs some RNA (Fraenkel-Conrat 1966) and diethyl pyrocarbonate under some conditions seriously inhibits the separation of viral protein and nucleic acid (Bagi et al. 1970).

In addition to the above factors, the salt concentration and nature of cations present can affect the isolation and stability of viral nucleic acids (Ralph and Berquist 1967). For example, at very low ionic strengths (10^{-4} M) strand separations occur in double-stranded nucleic acids and in double-stranded regions of single-stranded nucleic acids (all single-stranded nucleic acids have the tendency to form some double-stranded loops). Such denaturation usually makes nucleic acids more susceptible to degradation by nucleases. Cations such as Mg^{2+} may cause aggregation and loss of nucleic acids, especially of RNA, as do also salt solutions stronger than 1 M.

In summary, the various procedures for preparing intact undenatured viral nucleic acids are generally successful in proportion to their ability to effect a thorough denaturation of viral coat protein and separation of it from nucleic acid under conditions that avoid extremes of mechanical treatment, pH, and salt concentration and that minimize contact with nucleases. Some examples of procedures that apply these principles follow.

a. The Hot Salt Method

This procedure is a modification of the method of Cohen and Stanley (1942) (see Knight 1957; Reddi 1958; Lippincott 1961) that has proved useful in preparing nucleic acid from TMV and strains, although the nucleic acid isolated in this manner is not consistently so infectious as that obtained by detergent treatment or phenol extraction. [Infectivities comparable to the highest obtained by any procedure have been reported by Boedtker (1959) and by Lippincott (1961), using the hot salt method, but

Table 16. Approximate Content and Type of Nucleic Acid in Some Viruses.

Virus	Nucleic Acid per Particle, Daltons	Type of Nucleic Acid[a]	References
Algal N-1	38×10^6	ds-DNA	Adolph and Haselkorn 1971
Adenovirus-2	23×10^6	ds-DNA	Joklik and Smith 1972
Avian myeloblastosis	10×10^6	ss-RNA	Joklik and Smith 1972
Broad bean mottle	1×10^6	ss-RNA	Yamazaki et al. 1961
Brome mosaic	1×10^6	ss-RNA	Bockstahler and Kaesberg 1961
Coliphages f1, fd, M13	1×10^6	ss-c-DNA	Hoffmann-Berling et al. 1966
Coliphages f2, MS2, R17	1×10^6	ss-RNA	Hohn and Hohn 1970
Coliphages ØX174, S13	2×10^6	ss-c-DNA	Thomas and MacHattie 1967
Coliphage lambda	32×10^6	ds-DNA	Thomas and MacHattie 1967
Coliphages T2, T4, T6	130×10^6	ds-DNA	Thomas and MacHattie 1967
Cucumber 3 (and 4)	5×10^6	ss-RNA	Knight and Stanley 1941
Cytoplasmic polyhedrosis	13×10^6	ds-RNA	Kalmakoff et al. 1969
Foot-and-mouth disease	2×10^6	ss-RNA	Rueckert 1971

Fowlpox	200×10^6	ds-DNA	Hyde et al. 1967
Herpes simplex	100×10^6	ds-DNA	Joklik and Smith 1972
Influenza	4×10^6	ss-RNA	Compans and Choppin 1973
Mouse encephalitis	2×10^6	ss-RNA	Rueckert 1971
Newcastle disease	6×10^6	ss-RNA	Blair and Duesberg 1970
Poliomyelitis	2×10^6	ss-RNA	Schaffer and Schwerdt 1959
Polyoma	4×10^6	ds-c-DNA	Kass 1970
Potato X	2×10^6	ss-RNA	Reichmann 1959
Reo, Type 3	15×10^6	ds-RNA	Joklik 1970
Rous sarcoma	10×10^6	ss-RNA	Robinson and Duesberg 1967
Shope papilloma	5×10^6	ds-c-DNA	Kass and Knight 1965
Silkworm jaundice	22×10^6	ds-DNA	Bergold 1953; Bergold and Wellington 1954
Simian 40	4×10^6	ds-c-DNA	Yoshiike 1968
Tobacco mosaic	2×10^6	ss-RNA	Knight and Woody 1958
Tomato bushy stunt	2×10^6	ss-RNA	DeFremery and Knight 1955
Turnip yellow mosaic	2×10^6	ss-RNA	Markham 1959
Wound tumor	16×10^6	ds-RNA	Kalmakoff et al. 1969

[a]ds = double-stranded; ss = single-stranded; c = cyclic.

with temperatures between 90° and 98.5°. However, recoveries of RNA are substantially lower under these conditions and the risk of contamination with undegraded virus, higher.]

1. *Hot salt procedure for preparing RNA from TMV and similar viruses.* To 0.3 M NaCl in a water bath at 100°C is added enough virus in aqueous solution to give a final concentration of 10–15 mg/ml. After mixing, the mixture is held at 100° for 1 min and then removed to an ice bath. After chilling, the mixture is centrifuged at 5,000–10,000 *g* to remove coagulated protein. The water-clear solution of sodium nucleate can be freed of salt by dialysis against water in the cold or by precipitation two or three times with ice-cold 67 percent alcohol, redissolving the nucleate in water each time. (A final step of high-speed centrifugation may be used, if desired, to pellet traces of insoluble matter.) Yields of about 80 percent are usually obtained.

2. *Modified hot salt method for preparation of RNA from tobacco ringspot and turnip yellow mosaic viruses.* For success with tobacco ringspot and turnip yellow mosaic viruses, Kaper and Steere (1959a, 1959b) found it necessary to modify the hot salt procedure by reducing the virus concentration and heating time and increasing the salt concentration. Thus, to 2 ml of M NaCl in a boiling water bath is added 1 ml of virus (at 5–10 mg/ml in 0.01 M phosphate buffer at pH 7) and heating is continued for 35 sec with constant mixing. The mixture is cooled immediately in an ice bath and the coagulated protein is removed by centrifugation. The nucleic acid, in the supernatant fluid, can be purified by two precipitations with cold alcohol and high-speed centrifugation as above.

3. *Modified hot salt procedure for preparing RNA from influenza and Rous sarcoma viruses.* Only about 0.8 percent of influenza virus is RNA and it is difficult to extricate the nucleic acid from the great mass of protein, lipid, and carbohydrate present. A hot salt procedure reported to give very good yields was developed by Ada and Perry (1954). The method has also been used to extract RNA from Rous sarcoma virus (Bather 1957).

The purified, frozen-dried (lyophilized) virus is defatted by extracting twice at room temperature with chloroform-methanol (2:1, v/v) followed by one extraction with *n*-butanol and two washes with ethyl ether. The RNA is obtained by extracting the defatted virus one to three times at 100° with 10 percent (w/v) NaCl, using 20-min extraction periods. The RNA can be freed of salt as above, with alcohol precipitation probably preferable. So far, the nucleic acid thus obtained from influenza virus has proved noninfectious.

A hot salt method has also been used to prepare DNA from phage ØX174 (Guthrie and Sinsheimer 1960; Sekiguchi et al. 1960).

b. Detergent Method

The following procedure, adapted from the method of Fraenkel-Conrat et al. (1957), gives good yields of RNA from TMV and related viruses. The RNA is infectious (Fraenkel-Conrat 1956) and reconstitutes well with pro-

tein obtained by acetic acid degradation of the virus (see section on Reconstitution of Viruses).

1. *Detergent procedure for isolating RNA from TMV and similar viruses.* Virus at 20 mg/ml in water is heated to 55° in a water bath, adjusted to pH 8.8 with dilute NaOH, and mixed with an equal volume of 2 percent sodium dodecyl sulfate (commercial preparations such as Duponol C are satisfactory also) that has also been adjusted to pH 8.8 at 55°. The mixture is allowed to remain in the water bath at 55° for 5 min during which the solution loses its characteristic opalescence, owing to degradation of the virus. After 5 min (greater or less time may be required for different strains of TMV) the mixture is rapidly cooled to room temperature (about 23°C) and 0.5 vol of saturated ammonium sulfate is added. After about 10 min the precipitated protein is removed by centrifugation at 5,000–10,000 *g* and the clear supernatant fluid is stored at 4° overnight. The RNA precipitates out under these conditions, and the precipitate is packed by centrifugation, redissolved in a small volume of water, and reprecipitated by adding 2 vol of cold alcohol. The alcohol precipitation may be repeated once or twice more and traces of insoluble material may be removed from the final solution of RNA by centrifuging at about 100,000 *g* for 2 hr with refrigeration. Yields of 60–90 percent are obtained.

2. *Modified detergent procedure for isolation of DNA from polyoma virus (Smith el al. 1960).* Polyoma virus was isolated from clarified extracts of a mouse embryo tissue culture by differential and density gradient centrifugation. Equal volumes of virus solution and 10 percent sodium dodecyl sulfate at pH 7 are mixed and heated at 65° for 2 hr. After adding enough ammonium acetate to give a final concentration of 0.1 M, the DNA is precipitated by adding 2 vol of ethanol. The precipitate is dissolved in 0.1 M ammonium acetate and reprecipitated by adding alcohol as before. This dissolving and precipitating procedure is repeated once more.

A modified detergent method has also been used in the isolation of DNA from Shope papilloma virus (Watson and Littlefield 1960).

c. Combined Detergent and Hot Salt Method

Some viruses from which low yields of nucleic acid are obtained by either the hot salt or detergent methods alone will give reasonable amounts of nucleic acid by a combined procedure (Dorner and Knight 1953). In this method, 1 vol of 10 percent Duponol C solution (or sodium dodecyl sulfate) is added to 4 vol of aqueous virus at about 10 mg/ml. The mixture is heated in a boiling water bath for 4 min and then cooled in an ice bath. Most of the free detergent is removed by dialysis and then enough 5 N NaCl is added to make the final concentration 1 N with respect to NaCl. This mixture is heated for 3 min at 100°, chilled in an ice bath, and the coagulated protein is removed by centrifugation. Salt is removed by dialysis and the preparation is concentrated by directing a stream of air at the dialysis bag (pervapora-

tion). The concentrate of nucleate can be clarified by centrifugation, or, if desired, the nucleate can be precipitated from the concentrate by addition of 2 vol of cold ethanol.

d. The Phenol Method

The phenol extraction method (Westphal et al. 1952) is perhaps the most generally useful procedure for obtaining nucleic acid from a wide variety of viruses (as well as from tissues). In operation, two layers—a phenolic and an aqueous layer—are obtained and protein is extracted into the phenolic layer (and some in the interface) while nucleic acid (and polysaccharide, if present) goes into the aqueous layer.

Phenol extraction was first used to prepare viral nucleic acid by Schuster et al. (1956), and very soon it was shown (Gierer and Schramm 1956a, 1956b) that RNA thus obtained from TMV is infectious. The initial procedure does not work satisfactorily on all viruses, but in several cases modifications have been developed that have successfully extended the usefulness of the technique. A convenient adaptation of the method for the preparation of RNA from TMV and strains is presented here together with some modifications extending the usefulness of the method.

1. Phenol procedure for preparing RNA from TMV and similar viruses. To the virus solution in 0.02 M phosphate buffer at pH 7 and at a virus concentration of 20–25 mg/ml is added an equal volume of water-saturated phenol. (This is about 80 percent phenol, and it is easily prepared by taking a fresh bottle of commercial reagent grade crystalline phenol and almost filling the bottle with distilled water. The mixture is liquefied by placing in a warm water bath and stirring occasionally. Two layers will be apparent: a large lower layer consisting of the water-saturated phenol and a small upper layer of excess water. If stored at about 4° in the dark glass bottle normally commercially available, the preparation keeps for weeks, and portions of the lower layer are used in the preparation of nucleic acid. Some workers redistill their phenol, add metal-chelating agents such as sodium versenate, and so on, but the author has not found these refinements to be generally necessary.)

The mixture of virus and phenol is stirred on a magnetic stirrer for 10–15 min at room temperature (about 23°C), after which the mixture is separated into two layers by centrifuging at 5,000–10,000 *g* for about 2 min. (The original procedure was carried out at low temperature, and 4° is still used in some cases; in other instances it has been found necessary to use temperatures as high as 50°–60°.) The aqueous (top) layer is drawn off, and about one-tenth its volume of water-saturated phenol is added, and the mixture is stirred again for 3–4 min followed by centrifugation. The aqueous layer is extracted once more with a tenth volume of phenol and then twice with equal volumes of ether (to remove the small amount of phenol which dissolves in the aqueous phase). Residual ether is removed from the aque-

ous nucleate by two to three precipitations of the RNA with ethanol, which is accomplished by chilling the nucleate solution and adding 2 vol of ice-cold ethanol. If difficulty in precipitating the material is experienced, a drop or two of 3 M sodium acetate at pH 5 can be added. The final precipitate, pelleted by centrifugation, is dissolved in a small volume of distilled water and centrifuged at about 100,000 *g* for 2 hr. The nucleic acid is not sedimentable under these conditions but a trace of insoluble material is usually removed as a tiny pellet. Yields of the order of 80 percent are commonly obtained.

Conditions similar to those described above (except that usually temperatures around 4° have been employed without evidence that such low temperatures are necessary) have been used successfully to prepare RNA from partially purified poliovirus (Alexander et al. 1958); from potato virus X (Bawden and Kleczkowski 1959); from tobacco rattle virus (Harrison and Nixon 1959b); from cucumber mosaic virus (Schlegel 1960b); from an RNA-containing insect virus, Smithia virus pudibundae (Krieg 1959); and others. The method has also been used to extract DNA from T2, T4, and ØX174 phages (Davison et al. 1961; Rubenstein et al. 1961; Guthrie and Sinsheimer 1960).

e. Phenol-Detergent Method

In some cases it has been found beneficial to make the phenol extraction after the virus structure has been opened up by a detergent such as sodium dodecyl sulfate. Rushizky and Knight (1959) used such a technique to obtain infectious RNA from tomato bushy stunt virus, Huppert and Rebeyrotte (1960) to extract DNA transforming principle from bacteria, Wahl et al. (1960) to prepare ØX174-DNA, and Bachrach (1960) to obtain infectious RNA from foot-and-mouth disease virus. Bachrach's method is given here.

> Virus concentrates are diluted six times in 0.02 M phosphate buffer at pH 7.6 which contains 0.01 percent sodium ethylenediaminetetraacetate (EDTA, or "Versene") and 0.1 percent sodium dodecyl sulfate. The diluted virus is twice extracted at 4° with water-saturated phenol containing 0.01 percent EDTA. Phenol is removed from the final aqueous phase by several extractions with ether. The ether is removed by a stream of nitrogen.

In the standard phenol procedure used for preparation of RNA from TMV and similar viruses, the extraction is now made at room temperature, which is about 20° warmer than used in the original procedure. Even this temperature, however, is not sufficient for some viruses, and it has been necessary to go to about 50°, for example, to extract the RNA from equine encephalomyelitis virus (Wecker 1959).

Another important modification of the phenol method involves the

addition of bentonite as an adsorbent for nucleases during the phenol extraction (Fraenkel-Conrat et al. 1961) The infectivity of RNA obtained from TMV by the phenol-bentonite procedure is stabler upon incubation in salt solutions than most preparations made without bentonite (Singer and Fraenkel-Conrat 1961).

f. Guanidine Hydrochloride Method

Bawden and Kleczkowski (1959) reported the preparation of infectious RNA from potato virus X by phenol extraction, but the reproducibility of the results was not good. Therefore Reichmann and Stace-Smith (1959) investigated other procedures and devised a method based on treatment with guanidine that gave consistently 70–90 percent yields of infectious RNA. Their method is as follows:

To virus solution at 5–10 mg/ml is slowly added a sufficient volume of concentrated, recrystallized guanidine hydrochloride and ethylenediaminetetraacetate at pH 8.4 to make a final concentration of 2.5 M in guanidine and 0.005 M in EDTA. After 1 hr the RNA, which is insoluble, and the protein, which is soluble, in this mixture are separated by centrifugation at about 4,500 *g*. The RNA pellet is washed twice with 2.5 M guanidine-EDTA solution and then dissolved in a small volume of water. Further purification of the RNA is accomplished with 2 vol of cold ethanol, a 90-min centrifugation at 75,000 *g*, and dialysis overnight against distilled water.

g. Alkaline Method

As discussed earlier, extremes of pH are generally to be avoided in the preparation of nucleic acids. However, DNA is fairly resistant to alkaline media up to pH 12, and this fact was taken advantage of in isolating the nucleic acid of a nuclear polyhedrosis virus of the silkworm (*Bombyx mori*) (Onodera et al. 1965).

Freshly prepared virus particles are suspended in a small amount of 0.1 M Na_2CO_3–$NaHCO_3$ buffer (pH 10) 0.1 M in sodium citrate. This is dialyzed against a large volume of the same buffer-citrate mixture at 4° for two days. The dialyzed material is centrifuged at 40,000 *g* for 30 min to remove insoluble material. Solid ammonium sulfate is added to the supernatant fluid to a concentration of about 25 percent (wt/vol). The resulting precipitate of protein is removed by centrifugation and the supernatant containing the nucleic acid is dialyzed at 4° against 0.15 M NaCl–0.15 M sodium citrate. The DNA can be further purified by precipitation with 2–3 vol of cold ethanol or by passage through a column of methylated albumin.

In general, it may be stated that isolated viral nucleic acids are stabler than at first supposed. The primary cause of the instability is apparently traces of nucleases, and if these are absent, the nucleic acids maintain their integrity on storage and withstand temperatures considerably above room

temperature. In short, nucleic acids are not intrinsically chemically unstable, nor are they particularly thermolabile.

2. *Analysis of Viral Nucleic Acids*

Nucleic acids may be considered to be polynucleotides, that is, chain-like molecules in which the links are nucleotides. This is illustrated in Figure 17, which depicts the essential features while showing only a very small segment of nucleic acid.

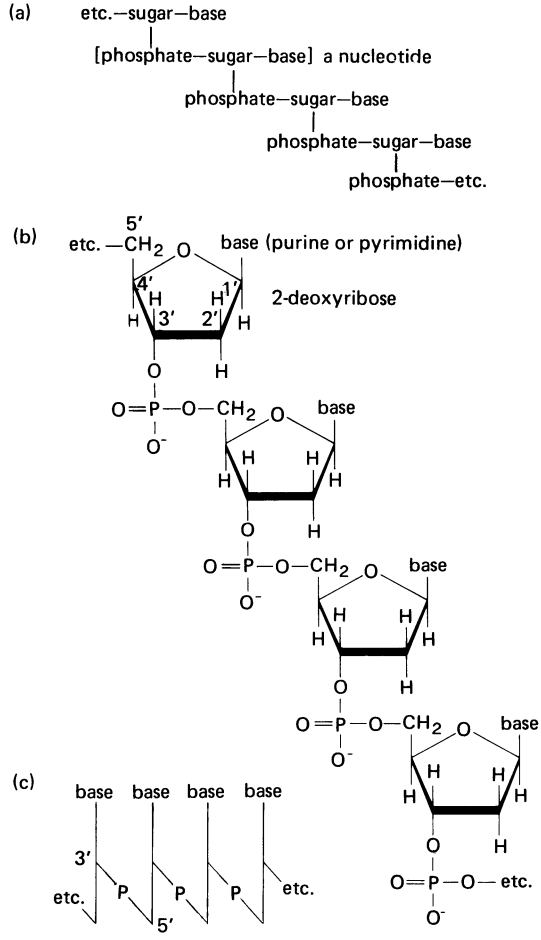
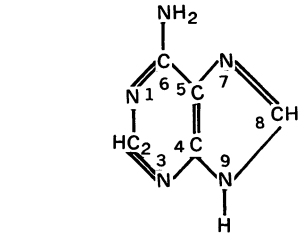


Fig. 17. Composition of a small segment of nucleic acid indicating the order of arrangement of the three major components (sugar, base, and phosphate). *a*. General scheme of sugar-phosphate backbone structure. *b*. Chemical structure of sugar-phosphate backbone of DNA showing numbering of atoms in the sugar (deoxyribose) and phosphodiester linkages between nucleotides. *c*. An abbreviated way to indicate oligonucleotides. (Adapted from Knight 1974.)

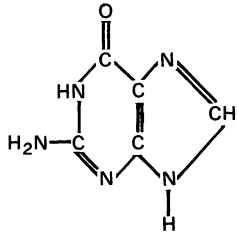
Nucelotides are named according to the purine or pyrimidine base they contain. In the case of DNA, which contains deoxyribose rather than ribose as in RNA, this is indicated in the naming of the nucleotides by appending the prefix deoxy. Thus, for RNA, the nucleotides are adenylic acid, guanylic acid, cytidylic acid, and uridylic acid; for DNA they are deoxyadenylic acid, deoxyguanylic acid, deoxycytidylic acid, and thymidylic acid (the deoxy prefix is not necessary for the thymine-containing nucleotide since there is no natural counterpart in the ribonucleic acid series).

As indicated in Figure 17, all nucleotides are built up from three simpler components: phosphate, sugar, and a purine or pyrimidine base. Nucleosides are made from sugar and a purine or pyrimidine base. Nucleosides are thus chemically closely related to nucleotides, and removal of phosphoric acid (by hydrolysis) from a nucleotide yields a nucleoside. Conversely, nucleotides can be viewed as nucleoside phosphates. The common ribonucleosides are adenosine, guanosine, cytidine, and uridine. The comparable deoxyribonucleosides are deoxyadenosine and so on, except that deoxy is commonly omitted from the name of the nucleoside consisting of deoxyribose and thymine (thymidine) since thymidine is characteristic of DNA only.

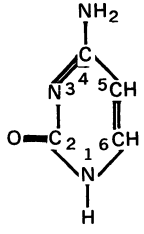
The purine bases commonly found in nucleic acids are adenine and guanine and the pyrimidines are cytosine, uracil, and thymine, the latter occurring only in DNA. In the DNA of certain phages, 5-hydroxymethylcytosine or 5-hydroxymethyluracil is found in place of cytosine. Formulas for some of these bases are as follows:



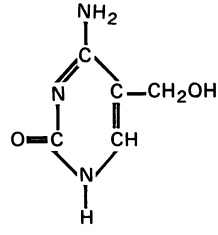
Adenine
(6-aminopurine)



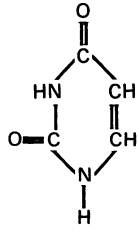
Guanine
(2-amino-6-oxypurine)



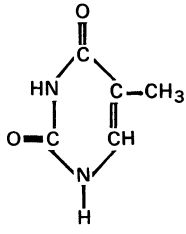
Cytosine
(2-oxy-4-amino-pyrimidine)



5-Hydroxymethyl-cytosine

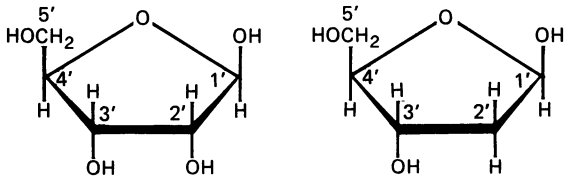


Uracil
(2,4-dioxypyrimidine)



Thymine
(5-methyluracil)

The sugar components of RNA and DNA are D-ribose and 2-deoxy-D-ribose, respectively. These sugars in the β configuration (see the following formula) are attached to purines and pyrimidines in nucleic acids (and also in nucleotides and nucleosides). Both sugars are found in nucleic acids (and in nucleotides) in the furanose ring form (oxygen ring between carbons 1′ and 4′), whereas the free sugars occur mainly in the pyranose form (oxygen ring between carbons 1′ and 5′). Structural and abbreviated formulas may be written for these sugars as follows:



It will be noted in the formulas for nucleic acid components that regular numbers were used to denote positions in the purine and pyrimidine rings, whereas prime numbers were used to indicate positions in the sugars. This convention applies to nucleosides, nucleotides, and nucleic acids in order to make the distinction between derivatives involving the sugar and those affecting the bases.

In nucleic acids, including those of viruses, the nucleotides are uniformly linked through 3′ → 5′ phosphate diester bonds between the sugar moieties as illustrated in the bit of DNA shown in Figure 17. This means that all DNA molecules share a common deoxyribose-phosphate structure; similarly all species of RNA have a common ribose-phosphate backbone. Consequently, aside from differences in molecular size, the biological specificity of nucleic acids depends entirely on the purine and pyrimidine bases, or, more precisely, on the sequence in which these bases occur along the sugar-phosphate chain.

An important step in characterizing nucleic acids is to determine the proportions of purines and pyrimidines present. This is commonly done by hydrolyzing the nucleic acid and separating the resultant components of the hydrolysate by either of three methods: paper chromatography or thin layer chromatography, paper electrophoresis, or column chromatography. Concentrations of discrete fractions separated by one of these methods are usually determined by ultraviolet spectrophotometry. Examples of some of the procedures are given below.

a. Determination of the Base Ratios in RNA by Acid Hydrolysis and Paper Chromatography

Numerous studies have been made on the hydrolysis of nucleic acids and the separation of nucleic acid constituents (see, for example, Fink and

Adams 1966; Lin and Maes 1967). An early and still useful procedure for determining base ratios in RNA is that of Smith and Markham (1950). Mild acid hydrolysis is used, which releases the purines in the form of free bases and the pyrimidines in the form of their nucleotides.

Viral RNA, isolated by one of the methods described in the previous section, is hydrolyzed at a concentration of 10 mg RNA/ml in 1 N HCl for 1 hr at 100° (boiling water bath). Twenty μl (equivalent to 200 μg of RNA) of hydrolysate is applied to Whatman No. 1 filter paper with a micropipette in such a way as to give a rectangular spot about 0.5 × 5 cm. Separation of the components is effected by either ascending or descending chromatography in 70 percent tert. butanol-water, 0.8 N with respect to hydrochloric acid (70 ml tert. butanol, 13 ml constant boiling HCl, and 17 ml H_2O). After chromatography, the papers are air-dried at room temperature, and the spots are located by examination with an ultraviolet light and marked with a pencil. Starting at the point of application and going in the direction of the solvent movement, the spots will be found in the following order: guanine, adenine, cytidylic acid, and uridylic acid. The sample spots are cut out, as well as paper blanks of approximately the same size next to the sample spots. Each paper cutout is placed in a test tube together with 5 ml of 0.1 N HCl and eluted by standing at room temperature overnight or by shaking for 2 hr. The absorption of each solution is read in a spectrophotometer at the wavelength of maximum (or near maximum) absorption for the compound in question. The amount of each component can be calculated by use of the proper extinction coefficients such as those given by Sober (1970). Some data of this sort are given in Table 17 for the compounds obtained from RNA as above, and for the compounds obtained by alkaline hydrolysis of RNA or acid hydrolysis of DNA.

During hydrolysis in 1 N HCl, about 5 percent of the pyrimidine

Table 17. Ultraviolet Absorption Data for Some Nucleic Acid Constituents Near Wavelengths of Maximum Absorption.[a]

Substance	pH	Wavelength, nm	Molar Extinction Coefficient, × 10^{-3}
Adenine	1	262.5	13.2
Guanine	1	248	11.4
Cytosine	1	276	10.0
5-Hydroxymethylcytosine	1	279	9.7
Thymine	4	265	7.9
Uracil	0	260	7.8
Adenylic acid	1	257	15.0
Guanylic acid	1	257	12.2
Cytidylic acid	2	279	13.0
Uridylic acid	1	262	10.0

[a]Compiled from Sober 1970.

nucleotides are hydrolyzed to nucleosides. Hence for more accurate values using this method, the cytidylic and uridylic acid figures can be corrected upward by 5 percent (Markham and Smith 1951) and the adenine value downward by the same amount (cytidine arising from partial hydrolysis of cytidylic acid migrates to the same area as adenine in tert. butanol-HCl. Uridine occupies an area between cytidylic and uridylic acids and hence does not affect values for other components). A more precise correction can be made by actual determination of the amount of cytidine in the adenine spot. This is done by using observed absorption values at two wavelengths, standard absorption values, and applying simultaneous equations (Loring and Ploeser 1949). Assuming a similar conversion of uridylic acid to uridine permits complete correction and accounts very well for the nucleic acid components in terms of phosphorus recovery (deFremery and Knight 1955).

The hydrolysis in 1 N HCl is capable of releasing the nucleic acid components from whole virus as well as from isolated nucleic acid (Dorner and Knight 1953). Hence it is not necessary to isolate nucleic acid in order to determine the proportions of purines and pyrimidines present. However, this analysis is affected somewhat by the relative proportions of protein and nucleic acid, and gives most accurate results on the viruses containing 10 percent or more RNA.

b. Determination of the Base Ratios in RNA by Alkaline Hydrolysis and Paper Electrophoresis

Another convenient method for determining the base composition of RNA is by hydrolizing the nucleic acid in dilute alkali at low temperature, separating the resulting four nucleotides by paper electrophoresis, and determining the quantity of each nucleotide by spectrophotometry on the material eluted from the paper (see Smith 1955; Crestfield and Allen 1955a, 1955b). A useful procedure is as follows.

Two mg of RNA is hydrolyzed in 0.1 ml of 0.4 N NaOH at 37° for 24 hr. Ten μl aliquots of this hydrolyzate is applied to buffer-moistened Whatman 3 MM paper in an approximately 2-cm streak. Any of a number of types of electrophoresis apparatus may be used (See Smith 1955; Crestfield and Allen 1955a; Rushizky and Knight 1960b). The buffer used is 0.05 M formate at pH 3.5 (prepared by adding 6.4 g ammonium formate and 10.3 g formic acid, 88–90 percent, to 6 liters of water). Electrophoresis is performed at a voltage gradient of 6 v/cm for about 15 hr. (With proper cooling, a higher voltage can be used and the separation accelerated.) After drying the paper in air, the nucleotides can be located by examination with an ultraviolet light, marked, cut out (with appropriate blanks), eluted in 5-ml portions of 0.01 N HCl, and measured in a spectrophotometer. At pH 3.5 the nucleotides are found in the following order, starting from the cathode side of the paper: cytidylic acid, adenylic acid, guanylic acid, and uridylic acid.

c. Determination of the Base Ratios in DNA by Acid Hydrolysis and Paper Chromatography

Analyses of DNA are based on methods of acid hydrolysis that release the purines and pyrimidines as free bases. Either 70 percent perchloric acid or 88 percent (or 98 percent) formic acid are usually employed. Wyatt (1955) suggests the use of formic acid for best recoveries of the various bases (including the somewhat labile 5-hydroxymethylcytosine), but hydrolyses must be made in sealed bomb (thick-walled) tubes and the pressure from decomposition of formic acid is conveniently released, after hydrolysis and cooling, by heating a small area at the top of the tube until a little hole blows open. The tube may then be safely and fully opened. If it is desired to get base analyses on whole virus without isolation of the nucleic acid, then 70 percent perchloric acid is recommended. The following is a possible procedure based on these observations.

DNA is placed in a pyrex glass bomb tube and enough 88 percent formic acid is added to give a concentration of 2 mg DNA/ml formic acid. The tube is sealed and heated at 175° for 30 min. After cooling, pressure is released as described above; the tube is opened and the hydrolysate is evaporated to dryness in vacuo. The residue is taken up in a small volume of N HCl to give a concentration equivalent to 10–20 mg/ml of the original DNA. Twenty μl of hydrolysate is placed on Whatman No. 1 paper and chromatographed in isopropanol-HCl-water (170 ml isopropanol, 41 ml concentrated HCl, 39 ml H_2O). The migration of the bases in increasing distance from the origin is in the order guanine, adenine, cytosine, and thymine. In cases where 5-hydroxymethylcytosine is present instead of cytosine, it will be found in the cytosine position. Location of the spots and elution and spectrophotometry are carried out as above.

d. Determination of the Nucleotide Ratios in ^{32}P-Labeled RNA by Alkaline Hydrolysis and Column Chromatography

An example of analysis of a phage RNA using column chromatography can be drawn from the studies of coliphage β (Nonoyama and Ikeda 1964). Coliphage β was grown in *E. coli* K12 bacteria in the presence of ^{32}P so that this was incorporated in the phage RNA. Radioactive phage RNA extracted from the virus by the phenol procedure was mixed with carrier yeast RNA (use of carrier enables analysis of minute amounts of viral nucleic acid) and the mixture was hydrolyzed to nucleotides by treatment with 0.3 N NaOH for 18 hr at 37°. The hydrolysate was neutralized with 0.3 N HCl and loaded on a Dowex column (formate type, 1 × 2). The nucleotides were separately eluted with a gradient of formic acid (0–4 N) and the radioactivities of the issuing fractions were measured in an automatic gas-flow counter. From these data and the assumption of equivalent labeling of the different nucleotides, the composition of the RNA could be calculated.

e. Determination of Base Ratios in DNA from Buoyant Density and Thermal Denaturation Values

By examination of many different samples it has been determined that the buoyant density of DNA in cesium chloride is directly proportional to its guanine plus cytosine content (the buoyant density of a substance is equivalent to the density of solution at the equilibrium position to which the substance sediments in a density gradient). Deviations from this linear relationship occur only if the purine or pyrimidine bases are substituted, that is, if the DNA contains bases other than adenine, thymine, guanine, or cytosine. Such cases are rare. Hence, by density-gradient centrifugation of viral DNA, together with a marker DNA of known density, data are obtained enabling the calculation of the viral DNA. These data are obtained by use of a centrifuge equipped to record the positions of the sedimenting species from their ultraviolet absorbancies. A detailed description of the technique and an illustrative calculation of density are given by Mandel et al. (1968). From a curve representing the best fit of measurements made on 51 DNA samples, Schildkraut et al. (1962) developed the relation

$$(\mathrm{GC}) = \frac{\rho - 1.660 \text{ g/ml}}{0.098}$$

where (GC) is the mole fraction of guanine plus cytosine and ρ is the buoyant density of the DNA in CsCl.

Similarly, a linear relationship exists between the molar percentage of guanine plus cytosine in DNA and the denaturation or "melting" temperature (T_m) of the nucleic acid. From observations on 41 samples of DNA, Marmur and Doty (1962) developed the relation: (GC) = (T_m-69.3) 2.44, where (GC) is the mole percentage (note that mole percentage = mole fraction × 100) and T_m is in degrees centigrade in a solvent containing 0.2M Na^+. The absorbance of the DNA solution at 260 nm is measured as a function of temperature, and T_m is taken at the midpoint of the increase in absorbance (hyperchromic rise). Details for the performance of such measurements are given by Mandel and Marmur (1968).

The errors associated with both the density and thermal denaturation procedures appear to be small and both procedures can be performed with microgram amounts of DNA. A comparison of some results obtained by these procedures and those obtained by chemical analysis are illustrated in Table 18.

An example of the use of such data as those in Table 18 is as follows. Using either the buoyant density or thermal denaturation temperature value for guanine plus cytosine listed for herpesvirus DNA in Table 18, and applying the molar equivalence rule that applies to the bases of double-stranded DNA (see Sec. f), one can readily calculate that the molar percentages of bases in this DNA are 16 percent adenine, 16 percent thymine, 34 percent guanine, and 34 percent cytosine.

Table 18. Guanine Plus Cytosine Content of Some Viral DNAs.

Virus	Molar Percentages of Guanine plus Cytosine			References[a]
	From Chemical Analysis	From Buoyant Density	From Thermal Denaturation Temp.	
Coliphage T3	50	53	49	1,2
Coliphage T7	48	51	48	1,2
Coliphage lambda	49	51	47	1,2
Adenovirus-2	58	57	57	3
Herpes	74	68	68	4,5
Shope papilloma	48	50	49	6

[a](1) Schildkraut et al. 1962. (2) Marmur and Doty 1962; (3) Piña and Green 1965; (4) Ben-Porat and Kaplan 1962; (5) Russell and Crawford 1963; (6) Watson and Littlefield 1960.

f. Proportions of Nucleotides in Some Viral Nucleic Acids

There are different ways of expressing the results of the base analyses made on nucleic acids. The commonest are (1) an arbitrary basis such as (a) moles of base per total of 4 moles, or (b) any one of the bases is set equal to 1 (or 10) and the values of the other bases are calculated accordingly; (2) moles percent, that is, moles base per 100 moles total bases; (3) moles base per mole phosphorus. Method 3 is probably to be preferred since it permits a ready evaluation of the recovery of the bases (there should be a total of 1 mole of bases per mole of phosphorus). However, this requires phosphorus analyses to be made, and sometimes there is not enough sample to make the desired replicate base determinations and phosphorus analyses too. Finally, since there is one base per nucleotide, results may obviously be expressed interchangeably in terms of moles base or moles of nucleotide. The compositions of some viral nucleic acids are summarized in Table 19 in terms of mole percent of nucleotides.

Several points about the compositions of viral nucleic acids as listed in Table 19 may be noted in passing. Nucleic acids of plant and mammalian tissues often contain small amounts of 5-methyldeoxycytidine (Hall 1971). However, such methylation rarely appears among viral nucleic acids. Two exceptions are noted in Table 19: 5-hydroxymethylcytosine found in the T-even coliphages and 5-hydroxymethyluracil observed in the DNA of *B. subtilis* phage SP8. In fact, on the basis of current information unusual purines or pyrimidines are quite uncommon in viral nucleic acids. One interesting variation is the occurrence of deoxyuridylic acid rather than thymidylic acid in the DNA of *B. subtilis* phage PBS2.

Inspection of the molar proportions of nucleotides for different viruses (Table 19) indicates considerable variation among viruses in this regard. Two examples, the DNAs of coliphage T3 and of *Salmonella* phage P22, appear to have equimolar proportions of all four constituent nucleotides. This is fortuitous and does not mean that the four nucleotides occur repeti-

tively in tandem (old tetranucleotide hypothesis). In contrast, the nucleic acids of potato virus X and of white clover mosaic virus have unusually large proportions of adenine; those of herpes simplex and pseudorabies, similarly big proportions of guanine; those of turnip yellow mosaic and wild cucumber mosaic viruses, extraordinary amounts of cytosine; and coliphages of the f1 group and influenza virus have lopsided proportions of thymine and uracil, respectively. The nucleic acids of most of the other viruses listed in Table 19 have undistinctive compositions.

While there are distinctive differences in the compositions of most of the nucleic acids of different viruses, such distinctions are seldom demonstrable with strains of the same virus. For example, the composition given for TMV suffices for its various strains, and one composition can be given for five strains of influenza A virus, one for the T-even coliphages, one for coliphages of the f1 series, one for three types of poliovirus, and so on. It should be remembered, however, that the present methods of analysis have an accuracy of about ± 3 percent, which, for example, is equivalent to about ± 50 nucleotides for any of the four nucleotides of TMV-RNA and of course proportionately larger for the bigger nucleic acids. This analytical situation should be viewed in the context that a difference in one nucleotide may be biologically significant.

The nucleic acids of viruses containing either double-stranded DNA or double-stranded RNA exhibit a molar equivalence of bases first noted in some DNAs by Chargaff and associates (1955):

$$\frac{\text{Adenine}}{\text{Thymine (or uracil)}} = \frac{\text{Guanine}}{\text{Cytosine (or 5-HMC)}} = \frac{\text{Purines}}{\text{Pyrimidines}} = 1$$

These regularities have definite implications concerning the structure of DNA (and of RNA) and were instrumental in the development of the Watson-Crick (1953a) double helix model.

g. Polynucleotide End Groups and Other Structural Features

In the case of proteins, primary structure analysis involves determination of numbers of chains in the protein molecule and the sequence of amino acid residues in the chain or chains. This concept, somewhat modified, can be carried over into nucleic acid structure, and, in the case of viruses, involves determination of the number of polynucleotide chains per virus particle and the sequence of nucleotides in the chain or chains.

With respect to these analyses, two points can be stated at the outset: (1) It is very common for viruses to have a single molecule or chain of nucleic acid per virus particle, but there are several instances, particularly in the case of viruses with double-stranded RNA in which the nucleic acid is segmented, that is, occurs in 10 to 15 discrete pieces per particle. (2) Sequencing of nucleic acids at present is more difficult than sequencing

Table 19. Amount of Nucleic Acid and Nucleotide Ratios of Some Viral Nucleic Acids.[a]

Virus	Type NA[b]	Daltons NA per particle, $\times 10^{-6}$	Approximate Moles Nucleotide per 100 moles					
			Ap[c] or dAp	Gp or dGp	Cp or dCp	Up or dUp[d]	Tp	5 HMdCp[e] or 5 HMdUp
Adenovirus 2	DNA	23	21	29	29		21	
Avian myeloblastosis	RNA	10	25	29	23	23		
B. subtilis PBS2	DNA	190	36	14	14	36[d]		
B. subtilis SP8	DNA	120	28	22	22			28[e]
Broadbean mottle	RNA	1	27	25	19	29		
Brome mosaic	RNA	1	27	28	21	24		
Coliphages f1, fd, M13	DNA	1	24	20	21			35
Coliphages f2, fr, M12, MS2, R17, β	RNA	1	23	26	26	25		
Coliphages ØX174, ØR, S13	DNA	2	24	25	19		32	
Coliphage Qβ	RNA	1	22	24	25	29		
Coliphages T2, T4, T6	DNA	120	33	17			33	17[e]
Coliphage T3	DNA	—[f]	25	25	25		25	
Coliphage lambda	DNA	30	26	24	24		26	
Coliphage T5	DNA	77	30	20	20		30	
Coliphage T7	DNA	24	26	24	24		26	
Cucumber 4	RNA	2	26	26	19	29		
Cytoplasmic polyhedrosis	RNA	13	29	21	21	29		
Foot-and-mouth disease	RNA	2	26	24	28	22		
Fowlpox	DNA	200	32	18	18		32	
Herpes simplex	DNA	81	16	34	34		16	
Influenza	RNA	4	23	20	24	33		
Mouse encephalitis	RNA	2	25	24	24	27		

Newcastle disease	RNA	6	24	24	23	29	
Poliomyelitis	RNA	2	29	24	22	25	
Polyoma	DNA	4	26	24	24		26
Potato X	RNA	2	32	22	24	22	
Pseudorabies	DNA	55	14	36	36		14
Reo Type 3	RNA	15	28	22	22	28	
Rous sarcoma	RNA	10	25	28	24	23	
Salmonella P22	DNA	28	25	25	25		25
Shope papilloma	DNA	5	26	24	24		26
Silkworm jaundice	DNA	22	30	20	20		30
Simian 40 (SV40)	DNA	5	26	24	24		26
Sindbis	RNA	—	29	26	25	20	
Tipula iridescent	DNA	156	34	16	16		34
Tobacco mosaic	RNA	2	28	24	22	28	
Tobacco necrosis	RNA	2	28	26	22	26	
Tobacco necrosis satellite	RNA	0.4	28	25	22	25	
Tobacco ringspot	RNA	2	24	25	23	28	
Tomato bushy stunt	RNA	2	26	28	21	26	
Turnip yellow mosaic	RNA	2	23	17	38	22	
Vaccinia	DNA	160	30	20	20		30
White clover mosaic	RNA	2	33	16	23	28	
Wild cucumber mosaic	RNA	3	18	16	40	26	
Wound tumor	RNA	16	31	19	19	31	

[a]Adapted from Knight 1974.
[b]NA is nucleic acid.
[c]Ap is adenylic acid, dAP is deoxyadenylic acid, and so on (see text for naming of nucleotides).
[d]*B. subtilis* phage PBS 2 is unusual in that it has deoxyuridylic acid in place of thymidylic acid, which is characteristic of DNA.
[e]5HMdCp is 5-hydroxymethyldeoxyytidylic acid which is found in coliphages T2, T4, and T6; and 5HMdUp is 5-hydroxymethyldeoxyuridylic acid found in *B. subtilis* phage SP8.
[f]–is unreported.

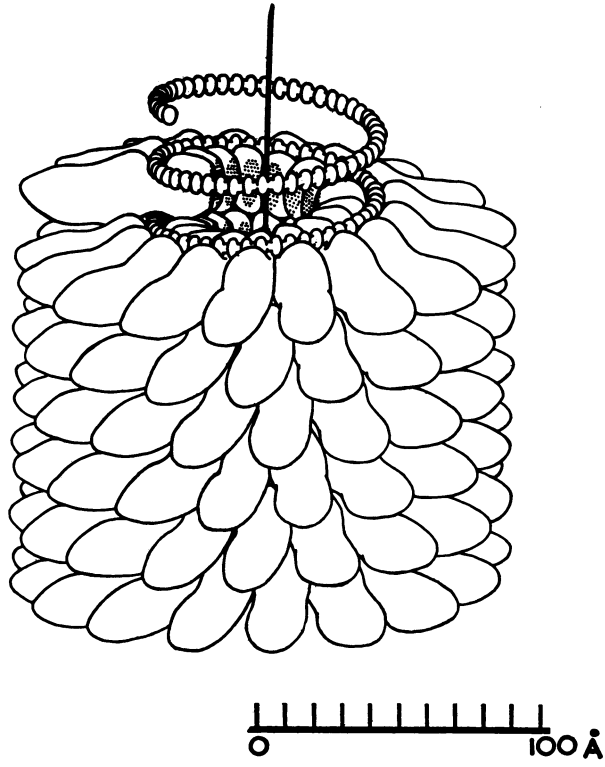


Fig. 18. Drawing of a segment of the tobacco mosaic virus particle with the protein subunits removed from the top two turns of the protein helix but maintaining the configuration of the RNA strand as it would be if the protein were there. The fit of the RNA in a helical groove of the protein subunits is indicated and individual nucleotides are denoted on the RNA strand as bead-like objects. It will be noted that there are about three nucleotides per protein subunit. (From Klug and Caspar 1960.)

proteins and, while extensive progress has been made, such sequencing is far behind that of proteins.

Progress in the chemical characterization of viral nucleic acids has been most pronounced with certain plant and phage nucleic acids. Therefore, these nucleic acids will be mainly used to illustrate some techniques and principles applicable to the determination of structure.

1. Number of nucleic acid molecules per virus particle. From the molecular weight of tobacco mosaic virus (about 40×10^6) and an RNA content of about 5 percent (Knight and Woody 1958), it can be calculated that each TMV particle contains 2×10^6 daltons of RNA. Likewise, it can be calculated that a polynucleotide chain of about this molecular weight would just occupy the length of a 300-nm rod if it were located at the radius shown in Figure 18 and followed the helical pitch of the protein subunits,

as it appears to do (Franklin et al. 1959; Hart 1958; Schuster 1960a). The total length of such a fiber would be 3,300 nm.

The critical question, of course, is whether there is a single fiber of RNA or several molecules, perhaps subunits, regularly spaced along the length of the TMV particle. Studies on hot salt preparations of TMV-RNA, using light-scattering measurements, yielded a molecular weight of 1.7×10^6 daltons for the isolated nucleic acid (Hopkins and Sinsheimer 1955). Similarly, Boedtker (1959), starting from highly monodisperse preparations of TMV, obtained fairly homogeneous preparations of RNA by a modified hot salt method (dilute virus, 90° heating for 1–3 min), and the molecular weight reported for this RNA, as determined from both light-scattering and sedimentation-viscosity measurements, was $1.94 \pm 0.16 \times 10^6$. Other light-scattering investigations were made by Friesen and Sinsheimer (1959) on TMV-RNA prepared by either the detergent or phenol procedures. A weight average molecular weight of 2×10^6 was found for both types of preparation, and infectivity was associated with this material and not with the smaller components that appeared upon storage of the RNA.

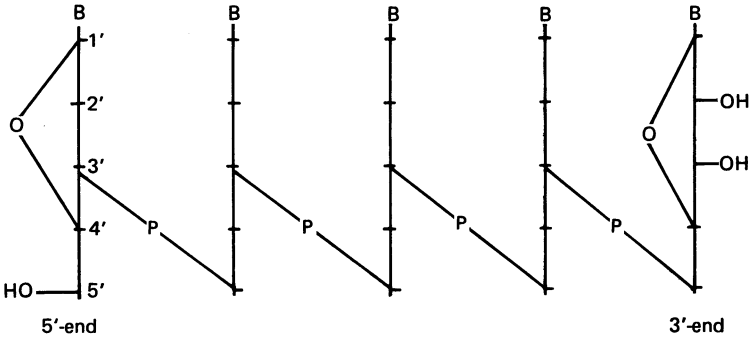
The physical properties of the RNA obtained by the phenol extraction procedure were also investigated extensively by Gierer (1957, 1958a, 1958b, 1958c). In sedimentation studies a well-defined, high molecular weight component was observed that accounted for the bulk of the RNA, the rest appearing on the sedimentation diagrams as smaller, polydisperse products. The observed sedimentation coefficient was about 31 Svedberg units and, after applying a viscosity correction, a molecular weight for the RNA of 2.1×10^6 was calculated. When such nucleic acid was treated with ribonuclease, the kinetics of degradation was found to be as expected for the random splitting of a single-stranded structure (Gierer 1957).

Thus, such data indicate that the RNA isolated from TMV has a molecular weight equivalent to the entire RNA content of a virus particle, and that it occurs in the form of a single strand. A question unanswered at this stage was: Is the RNA a uniform polynucleotide chain in which all the 6,400 nucleotides are linked by covalent bonds, or might there be polynucleotide subunits, joined perhaps by hydrogen bonds, to form a single strand? Gierer found (1959, 1960) that TMV-RNA could be heated at 70° for 10 min or at 40° in 36 percent urea for 30 min without the RNA strands breaking down. Since both of these treatments are known to be disruptive to hydrogen bonds, it may be concluded that these are not linking polynucleotide subunits together in TMV-RNA but TMV-RNA is rather a single, large, polynucleotide strand. More recently, studies of the electrophoretic migration of TMV-RNA in polyacrylamide gels yielded results consistent with a viral RNA strand of about 2×10^6 daltons (Bishop et al. 1967).

The RNAs of several phages as well as of animal viruses of the polio and mouse encephalitis types appear to occur in their respective virus particles in one piece, like TMV. The same holds true for DNA in a wide variety of viruses. However, there are some exceptions to the unitary genome struc-

ture. For example, the double-stranded RNA of reovirus occurs in 10 segments (Shatkin et al. 1968; Millward and Graham 1970). A similar situation is found with the double-stranded RNAs of cytoplasmic polyhedrosis and wound tumor viruses (Kalmakoff et al. 1969). In addition, the single-stranded RNA of influenza virus occurs in segments (Barry et al. 1970) and there may be other examples. Two consequences of a segmented genome are that it provides a basis for unusually high recombination in mixed infection (genotypic mixing) and it provides individual gene segments of nucleic acid that can probably be sequenced and the data subsequently related to specific gene products.

2. *End group determinations.* Names are given to the two ends of polynucleotide chains of nucleic acids just as they are for ends of the polypeptide chains of proteins. Thus, in formulas depicting nucleic acid structure, the left side of the linear array of nucleotides is customarily referred to as the 5′-end of the structure, while the right side is called the 3′-end. This nomenclature is based on the occurrence of a free hydroxyl on the 5′-carbon and 3′-carbon atoms, respectively, of the terminal nucleotides. This can be illustrated in the following abbreviated structure for RNA:



Inspection of this structure shows that the ribose residues (indicated by vertical lines topped with a purine or pyrimidine base, B) are joined by 3′-5′ linkage through phosphate (P). The 5′-OH is involved in this linkage in every case except for the terminal residue on the left. This terminal is therefore recognizable on the basis of its free 5′-OH as the 5′-end. Similarly, the 3′-OH is involved in the formation of phosphodiester linkages everywhere except in the ribose residue on the extreme right. The free 3′-OH there marks this residue as the 3′-end.

Two important features of the ends of nucleic acid chains are often investigated: the presence or absence of terminal phosphate groups and the nature of the purine or pyrimidine base on the terminal nucleotides.

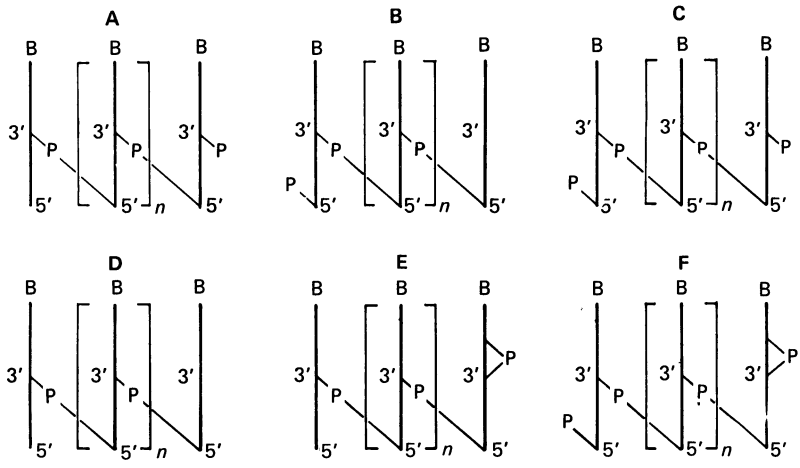


Fig. 19. Schematic representation of six chain end situations for RNA. Using the shorthand notation of Fig. 17*c* for the polynucleotide chain, B stands for purine or pyrimidine base, P for phosphate group, vertical lines for ribose residues, and *n* for number of nucleotides. (From Gordon et al. 1960.)

Some possibilities with respect to terminal phosphate groups, for example in TMV-RNA, are illustrated in Figure 19. It can be seen that treatment of RNA having a structure like that represented by A, B, or C, with an appropriate phosphatase enzyme (see Fraenkel-Conrat and Singer 1962), should release either one or two moles of inorganic phosphate per mole of RNA. By quantitative analysis of the released phosphate, which on a small scale is best done with ^{32}P-tagged RNA, the mean size of the RNA chains can be calculated. Subsequent mild alkaline hydrolysis, which breaks internucleotide links to the right of the phosphate groups as they are depicted in Figure 19, would yield one nucleoside per mole of RNA, and this can be separated from the nucleotides by paper electrophoresis and identified by comparison with standard compounds. Similarly, if the RNA should have one of the structures depicted by B and D, an estimate of the mean chain size and nature of the terminal residues can be obtained by determining the quantity and nature of nucleoside (and nucleoside diphosphate in the case of structure B) produced when the RNA is degraded with dilute alkali.

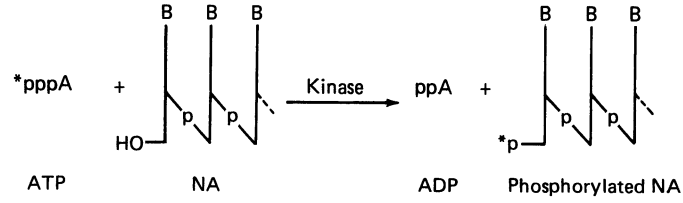
Other combinations of enzymatic and chemical treatments can provide information about the other end of the RNA chain, and such procedures can also be used to determine whether such structures as E and F are present. A structure similar to B except that it has a 5′-triphosphate group is common in certain phage RNAs. Alkaline hydrolysis of this structure would yield a nucleoside tetraphosphate that would have properties quite distinct from those of all the other digestion products and could be readily separated and identified.

The points just made regarding ends of RNA also apply in principle to DNA except that DNA is not subject to alkaline degradation and some of the enzymes active on DNA are different from those active on RNA. Also there is the problem of two strands in double-stranded DNA and these will need to be separated for some analyses. Finally, there is one structure not shown in Figure 19 that does occur with DNA, that is, a cyclic structure in which the two ends are joined. Essentially, in such a case, no split products are obtained when the nucleic acid is treated with phosphatases and exonucleases (exonucleases are phosphodiesterases that attack polynucleotide chains only or preferentially at ends and liberate mononucleotides therefrom. In contrast, endonuclease enzymes catalyze cleavage of mononucleotides from various locations in the middle of the chain). The cyclic single-stranded DNA of coliphage ØX174 constitutes such a structure (Fiers and Sinsheimer 1962a, 1962b) and cyclic double-stranded DNA is common in DNA-containing tumor viruses (Crawford 1968).

In the initial studies on TMV-RNA using ^{32}P-labeled RNA and treatment with prostatic phosphomonoesterase, Gordon et al. (1960) reported the release of one inorganic phosphate per 3,000–5,000 nucleotides. This was interpreted to mean that TMV-RNA had at the most one monoesterified end. Later results obtained with an *E. coli* phosphomonoesterase and specially purified RNA indicated considerably less than 1 mole of phosphorus per mole of RNA (Fraenkel-Conrat and Singer 1962; Gordon and Huff 1962). This combined with the evidence by Fraenkel-Conrat and Singer that snake venom diesterase (which catalyzes cleavage to the left of each internucleotide phosphate group in the formulations of Fig. 19) caused the release of considerably less than one mole of nucleotide diphosphate from TMV-RNA, leads to the conclusion that there is no monoesterified phosphate in TMV-RNA.

Assuming on the basis of the above that formula D of Fig. 19 represents the correct structure for TMV-RNA, it can be seen that mild alkaline hydrolysis should yield one nucleoside per mole of RNA from the 3′-end (right side of Fig. 19D), the rest of the RNA being converted to nucleoside 2′- or 3′-phosphates. Conversely, degradation of the RNA with venom diesterase should yield one nucleoside per mole of RNA from the 5′-end of the molecule, the rest of the RNA being converted to nucleoside 5′-phosphates.

However, the task of separating one nucleoside from approximately 6,400 nucleotides and identifying it is a formidable one. The key to the problem was the use of TMV-RNA highly labeled with ^{14}C (this is done by growing virus-infected plants in a chamber containing $^{14}CO_2$). When the degradative methods just outlined were applied to TMV-RNA, and the products were separated by paper electrophoresis and paper chromotography, it was found that in each case about 1 mole of adenosine was released per 2×10^6 daltons of RNA (Sugiyama and Fraenkel-Conrat 1961a,



1961b; Sugiyama 1962). Thus it appears that TMV-RNA has the structure represented by formula D of Figure 19 and that adenine is the base at both the 5′- and 3′-ends of the molecule.

However, it appears that the 5′-ends of TMV-RNA (and of brome mosaic virus RNA as well) may not be as uniform as the 3′-ends (Fraenkel-Conrat and Fowlks 1972). This was discovered by application of a neat labeling technique, that is, treatment of the nucleic acid with a polynucleotide phosphokinase (Richardson 1965), which transfers phosphorus from labeled adenosine triphosphate to the 5′-end of the nucleic acid:

Up to 1 mole of phosphate per mole of RNA was transferred in this manner to TMV-RNA, confirming the earlier conclusion that the 5′-end is unphosphorylated. The product was then degraded to nucleotides by alkali and the digest was analyzed for content of radioactive nucleoside 3′, 5′-diphosphates after separation of these by paper electrophoresis (or were analyzed for radioactive nucleoside 5′-phosphates separated from snake venom digests of the labeled RNA). Both methods indicated terminal heterogeneity with molar proportions of A:U:G:C = 54:17:18:11. As yet there is no evidence that this variability of the 5′-terminal of TMV-RNA has a serious effect on the biological activity, whereas oxidation of ribose in the 3′-terminal destroys most of the infectivity of the RNA (Steinschneider and Fraenkel-Conrat 1966).

The RNA of turnip yellow mosaic virus is another example of a viral nucleic acid with an unphosphorylated 5′-end (Suzuki and Haselkorn 1968). However, the 5′-ends of several phage RNAs have been found to terminate in triphosphate (Glitz 1968; de Wachter and Fiers 1969; Young and Fraenkel-Conrat 1970). Since nucleic acids appear to be synthesized from the 5′-end toward the 3′-end and nucleotide triphosphates are used in the syntheses, the presence of 5′-triphosphate ends as in the phage RNAs is expected. The absence of phosphate at the terminals of some of the plant virus nucleic acids may indicate that certain plant cells have more phosphatase activity than some other types of cells.

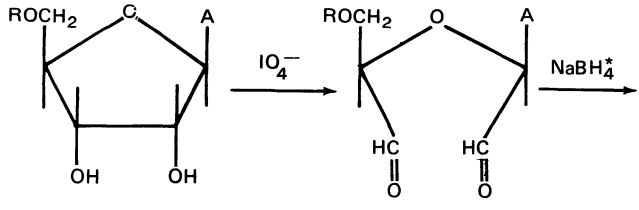
In any case, it appears that some viral nucleic acids as isolated from mature virus particles have phosphorylated ends and others do not. A special case of those that do not are the nucleic acids with cyclic structures.

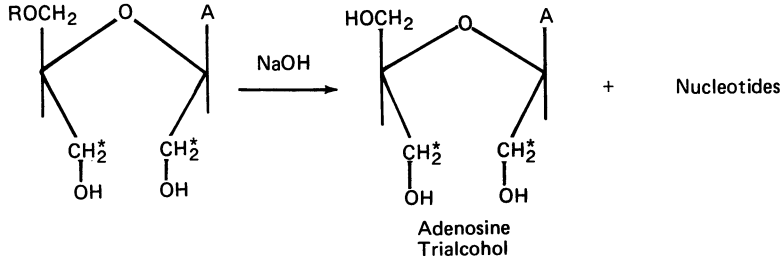
As just indicated, failure to detect release of significant amounts of phosphate when RNA is treated with *E. coli* phosphatase suggests that the terminals are not phosphorylated. This point can be checked with respect

to the 5′-terminal, and at the same time the nature of the terminal base can be established by phosphorylation of the 5′-OH of the terminal nucleoside using the Richardson kinase (Richardson 1965). Phosphorylation places a radioactive label on the terminal and subsequent alkaline hydrolysis yields one radioactive nucleoside diphosphate per RNA molecule and all the rest of the RNA as nucleoside monophosphates. The nucleoside diphosphate can be separated from the nucleotides and identified by electrophoresis (also called ionophoresis) in two dimensions (for details of methodology, see Dahlberg 1968; Brownlee 1972). Likewise, if the 5′-terminal is naturally phosphorylated, as those of several phage RNAs are, alkaline hydrolysis will yield a distinctive product that can be identified in the same manner. Uniformly ^{32}P-labeled RNA is highly desirable for this analysis because of the small amount of terminal compound compared with the bulk of the RNA. (Such compounds can be detected by ultraviolet spectrophotometric methods, but it should be noted that the sensitivity of the isotope technique employing ^{32}P is almost 1,000-fold greater than that with spectrophotometry.) Label can be introduced by growing virus in a medium containing ^{32}P-phosphate, or, if a specific replicase is available as it is for some phage RNAs, it is possible to synthesize a radioactive complementary copy by using unlabeled RNA as template in the presence of α-^{32}P-phosphate-labeled nucleotide triphosphates as substrates for the replicase (the three phosphate groups of nucleotide triphosphates are designated α, β, and γ, starting with α as the first one attached to the 5′-carbon).

Terminal 3′-groups can also be identified by special labeling procedures. An example of such a procedure for analyzing the 3′-end of a polynucleotide chain is the periodate oxidation-borohydride reduction method (Glitz et al. 1968; Leppla et al. 1968). This procedure requires an unphosphorylated terminal; if the end nucleoside is phosphorylated, the terminal phosphate is removed in a preliminary treatment with phosphatase. The adjacent 2′, 3′-hydroxyl groups on the ribose of the terminal nucleoside (these hydroxyls constitute the only such readily oxidized pair in the molecule) are oxidized to aldehyde groups by treatment with periodate. The aldehyde groups are next reduced and simultaneously tagged by treatment with tritiated borohydride. Finally, the terminal nucleoside derivative, conveniently referred to as a nucleoside trialcohol, is released by alkaline hydrolysis and separated from the nucleotides derived from the rest of the RNA molecule by paper or column chromatography procedures that permit its identification.

The main steps of the procedure can be illustrated as follows:





R stands for all of the RNA molecule except for the 3′-terminus, and A is adenine. The asterisk indicates location of the radioactive label (tritium). One atom of tritium is transferred from the borohydride to each alcohol group formed by reduction of the dialdehyde. A modification of this procedure that yields a somewhat less stable radioactive product is to substitute ^{14}C semicarbazide ($NH_2\text{-}NH\text{-}\overset{\overset{O}{\|}}{C}\text{-}NH_2$) for the tritiated borohydride (Steinschneider and Fraenkel-Conrat 1966), which then yields a radioactive semicarbazone.

Application of the method described above to TMV-RNA and to the RNAs of coliphages f2 and MS2 indicated that the 3′-terminal in all cases is adenosine (Glitz et al. 1968). The same result was obtained with the RNA of phage R17. But, as shown in Table 20, which illustrates the kind of data obtained, some distinctive situations were found with the double-stranded RNAs of three viruses with segmented genomes. Although analyses were made on individual segments of the genomes, the results were essentially the same within ± 2 percent; hence the table shows the results as though each virus contained only a single molecule of nucleic acid per particle as the R17-RNA does. A striking feature of the results shown in Table 20 is that the RNAs of cytoplasmic polyhedrosis and wound tumor viruses appear to have two different 3′-terminal nucleosides, uridine and cytidine. Since these appear in approximately equivalent amounts, it has been concluded that one strand of each RNA segment terminates in uridine while the complementary strand terminates in cytidine (Lewandowski and Leppla 1972).

Dahlberg (1968) has developed a procedure for analyzing the 3′-end of RNA that is uniformly labeled with ^{32}P. This procedure depends on the fact that in a complete T_1 ribonuclease (T_1 RNase) digest of RNA, the only product not susceptible to attack by alkaline phosphatase is the oligonucleotide derived from the 3′-terminal end of the RNA since this is unphosphorylated in most RNAs. (T_1 RNase catalyzes hydrolysis next to guanylic acid residues only. Oligonucleotide means a small polynucleotide segment.) Thus, a T_1 RNase digest of RNA is electrophoresed on DEAE paper providing a spread of oligonucleotide spots that are treated in situ with alkaline phosphatase. Charges will generally be different on the phos-

Table 20. The 3′-Terminals of some Viral RNAs as Determined from Incorporation of Label in Terminal Trialcohols Formed in the Periodate Oxidation-Borohydride Reduction Procedure.[a]

RNA from	Percent of Total Tritium Label in Nucleoside Trialcohols			
	U-triAlc[b]	G-triAlc	A-triAlc	C-triAlc
Reovirus	3		1	96
Cytoplasmic polyhedrosis virus	50		2	48
Wound tumor virus	52	1.5	6	40
R17 phage	3	1.5	94	1.5
Tobacco mosaic virus	1		99	

[a]Adapted from Lewandowski and Leppla 1972; Glitz et al. 1968.
[b]U-triAlc is uracil trialcohol, G-triAlc is guanine trialcohol, and so on.

phatase-treated oligonucleotides except for the terminal one. This will be the only one without any guanine, which migrates in the same way when subjected to electrophoresis at right angles to the first run but in the same solvent. The terminal oligonucleotide usually appears by itself on a diagonal drawn across the paper from the origin and can be eluted and digested separately with alkali, pancreatic RNase and venom nuclease. Electrophoresis in two dimensions of the digestion products provides data from which the nucleotide content and sequence can be deduced, including the 3′-terminal residue. Some data on terminal residues of a few viral RNAs are given in Table 21.

Table 21. Terminal Groups of Some Viral RNAs.

Virus	5′-End	3′-End	Reference[a]
f2, MS2, R17 coliphages	pppG. . .	. . .A_{OH}	1–3
Qβ coliphage	pppG. . .	. . .A_{OH}	1
Satellite necrosis	ppA. . .	. . .C_{OH}	4,5
Tobacco mosaic	A. . .	. . .A_{OH}	6,7
Turnip yellow mosaic	A. . .		8

[a](1) Dahlberg 1968; (2) Wachter and Fiers 1969; (3) Glitz 1968; (4) Wimmer et al. 1968; (5) Wimmer and Reichmann 1969; (6) Sugiyama and Fraenkel-Conrat 1961a; (7) Sugiyama and Fraenkel-Conrat 1962; (8) Suzuki and Haselkorn 1968.

h. Nucleotide Sequences

Viral nucleic acids, like viral proteins, are too large to analyze from end to end by stepwise degradation procedures. Methods of detecting struc-

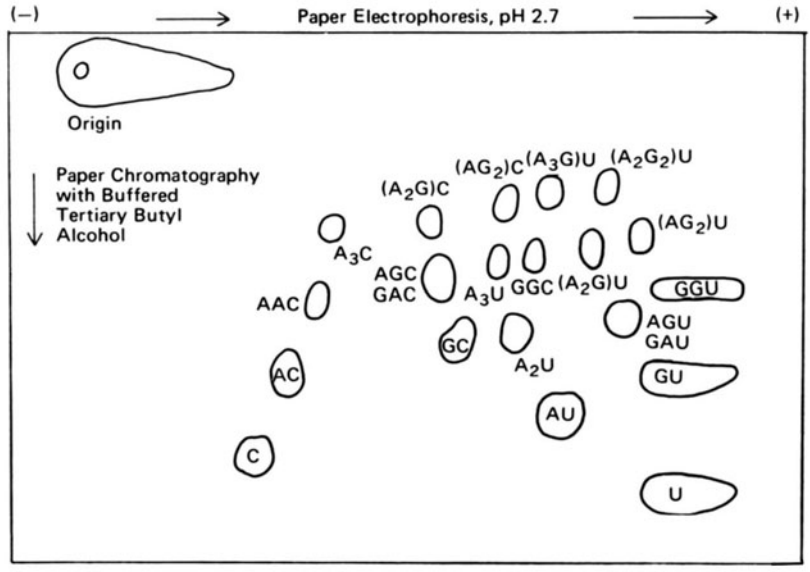
tural units (amino acids in the case of proteins and nucleotides in the case of nucleic acids) as they are split from the macromolecule are not sensitive enough to permit analysis of single molecules and hence thousands must be used. This requires substantial purity of the starting material and synchrony in the cleavage. Such synchrony can be achieved for a number of residues, but then begins to fail and is often accompanied by internal splits of the polynucleotide chain that yield spurious ends.

Consequently, the basic approach is to partially but specifically degrade the molecules into smaller fragments of various sizes. After purification, the sequences of these smaller fragments are established by further degradative procedures, often enzymatic, and identification of the products. By using enzymes of different specificities, and sometimes by partial rather than complete digestion, polynucleotide segments that overlap can be obtained. With these, progressively more of the oligonucleotides can be arranged in their correct order in larger segments of nucleic acid until the total sequence is deduced. Analyses of oligonucleotides for their nucleotide composition are made by alkaline and enzymatic digestion procedures.[2]

Some years ago, a two-dimensional procedure for separating small segments of RNA (one to ten nucleotides long) was developed (Rushizky and Knight 1960a, 1960b, 1960c; Rushizky et al. 1961; Rushizky 1967). The procedure was called two dimensional because it involves paper electrophoresis in one dimension on a large sheet of filter paper followed by chromatography in the second dimension on the same paper. The resulting reproducible spread of nucleotides and oligonucleotides on the paper was termed a "map" (some investigators subsequently called it a "fingerprint"). Such a map is shown in Figure 20. In the mapping procedure pancreatic ribonuclease (ribonuclease A) was employed at first to split the RNA into nucleotides and oligonucleotides. In subsequent experiments, a micrococcal nuclease (Reddi 1959; Rushizky et al. 1962a) and a fungal ribonuclease, ribonuclease T_1 (Miura and Egami 1960; Reddi 1960; Rushizky et al. 1962b), were used. An outline of the mapping procedure as it was applied to TMV-RNA is as follows.

To 1 ml of aqueous solution of TMV-RNA at about 8 mg/ml is added 0.3 ml of pancreatic ribonuclease (RNase) solution at 1 mg/ml and 0.02 ml of M sodium phosphate at pH 7.1. The mixture is allowed to stand at about 23° (room temperature) for 6–8 hr. To determine the precise amount of RNA being analyzed, two 25-μl aliquots are removed from the RNase digest, diluted with N KOH to 10 ml, and, after standing for 24 hr, read in the spectrophotometer at 260 nm. The initial concentration of RNA is calcu-

[2]Many procedures pertinent to analysis of oligonucleotides are described in detail in *Methods of Enzymology*, Vol. 12, Part A, Section II, L. Grossman and K. Moldave, editors, New York: Academic Press (1967).



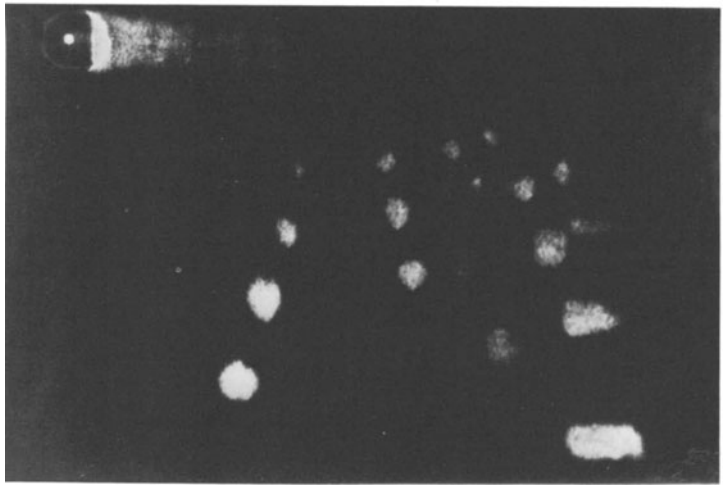


Fig. 20. Contact print map (*bottom*) and key to map (*top*) of pancreatic ribonuclease digestion products obtained from the RNA of the M strain of tobacco mosaic virus. (Maps for the RNAs of other strains and for many RNAs, including yeast RNA, show the same spots; however, they may differ in intensity, reflecting quantitative differences in runs of oligonucleotides). For convenience in labeling, the abbreviations used in the key omit the p normally used to designate phosphate groups. Thus Cp (cytidylic acid) becomes C, ApCp becomes AC, and so on. (From Rushizky and Knight 1960c.)

lated from the relationship that 1 mg of TMV-RNA/ml in N KOH has absorbance at 260 mn of 32.53 (Rushizky and Knight 1960c).

To fractionate the oligonucleotides in the RNase digest, 0.5-ml aliquots (equivalent to about 3 mg of RNA) are applied to buffer-moistened Whatman 3MM paper (46 × 57 cm), and a small application of picric acid (a visible electrophoresis marker that moves slightly faster than any component in the RNase digest) is also applied to the same edge but at the opposite corner of the paper. The buffer used is ammonium formate prepared by adding 7.0 ml of 98 percent formic acid to 2.5 liters of water and adjusting the pH to 2.7 with concentrated ammonium hydroxide. The paper electrophoresis is performed at 350 v (6 v/cm) for 17–20 hr with the point of application of sample near the cathode chamber. The run is finished when the picrate marker reaches the level of the buffer in the anode chamber. A Durrum-type electrophoresis apparatus was used by Rushizky and Knight (1960b). The paper is dried in a current of air at room temperature turned 90° from the direction used in electrophoresis, and serrated at the edge opposite the band of material in order to permit descending chromatography with runoff.

Chromatography is performed with a solvent consisting of equal parts of the electrophoresis buffer, adjusted to pH 3.8 with concentrated ammonium hydroxide, and tertiary butanol (the pH of the mixture at the glass electrode, without solvent correction, is about 4.8). For best results, chromatography should be carried out in a tank thoroughly saturated with solvent vapor for about 36 hr at approximately 25°.

After chromatography, the paper is dried in a current of air at room temperature, and the spots are located with an ultraviolet lamp. A record can also be secured by the contact printing method of Smith and Allen (1953), thus providing a map similar to that in Figure 20. Identification of the spots is by comparison with the positions of spots on a standard map on which the compounds had been identified by elution and analysis using enzymatic and chemical degradations (Rushizky and Knight 1960a).

For quantitative analysis, the spots, located under ultraviolet light, marked with a pencil, and cut out, are eluted in 5–10 ml of 0.01 N HCl at room temperature overnight. An appropriate paper blank is cut from each level of spots and treated in the same manner. The sample eluates are then read against an eluate of the proper paper blank in a spectrophotometer at 260 nm. The spectrophotometer readings are converted into quantities of compound by use of published extinction values for mononucleotides, or, in the case of the oligonucleotides, by use of extinction values calculated from composition using the extinctions of the component nucleotides. In such calculations, allowance is usually not made for the hypochromic state of the oligonucleotides since the error thus introduced is in most cases not great.

Some sequences found in TMV-RNA are summarized in Table 22. The

Table 22. Some Oligonucleotide Sequences Found in Tobacco Mosaic Virus Ribonucleic Acid.[a,b]

Dinucleotides	Trinucloetides	Tetranucleotides	Pentanucleotides
ApAp	ApApAp[c]	ApApApCp	(ApApApGp)Up
ApCp	ApApCp	ApApApGp	(ApApGpGp)Up
ApGp	ApApGp	ApApApUp	
ApUp	ApApUp	CpCpCpGp	
CpAp[c]	ApCpCp	UpUpUpGp	
CpCp	ApCpGp	(ApApCp)Gp	
CpGp	ApGpCp	(ApApGp)Cp	
CpUp	ApGpGp	(ApApGp)Up	
GpAp[c]	ApGpUp	(ApApUp)Gp	
GpCp	ApUpGp	(ApGpGp)Cp	
GpGp[c]	ApUpUp	(ApGpGp)Up	
GpUp	CpApGp	(ApCpCp)Gp	
UpAp	CpCpCp[c]	(ApCpUp)Gp	
UpCp	CpCpGp	(ApUpUp)Gp	
UpGp	CpUpGp	(CpCpUp)Gp	
UpUp	CpUpUp	(CpUpUp)Gp	
	GpApCp		
	GpApUp		
	GpGpCp		
	GpGpUp		
	GpUpUp		
	UpApGp		
	UpCpCp		
	UpCpGp		
	UpGpGp		
	UpUpGp		
	UpUpUp[c]		

[a]Data taken from Rushizky and Knight 1960c; Rushizky et al. 1961, 1962a, 1962b.

[b]The abbreviations are as indicated in the general section on nucleic acids. Where the composition is known but the sequence is not, parentheses are used.

[c]These sequences were deduced from higher oligonucleotides, whereas most of the compounds listed were actually isolated and identified as such after enzymatic digestion of TMV-RNA.

theoretical permutations (P) of the four common RNA nucleotides in which any nucleotide can occupy any position is given by $P = 4^n$ in which n is the number of nucleotide units in the oligonucleotide. Thus 4^2 dinucleotides are possible, 4^3 trinucleotides, and so on. As shown in Table 22, all 16 possible dinucleotide sequences have been found in TMV-RNA, as well as 27 of the possible 64 trinucleotide sequences. The compositions of relatively few tetranucleotide or higher fragments have been determined, although even now it is clear that a great variety occurs.

One of the findings arising from this early analysis of viral RNA sequences is that all RNAs yield the same pattern of oligonucleotides after complete digestion with pancreatic ribonuclease. However, definite quan-

titative differences are readily demonstrated, as will be described in the section on chemical differences between strains of a virus.

The determination of nucleotide sequences in a segment of nucleic acid larger than those just described for TMV can be illustrated by an example drawn from a study of the RNA of bacteriophage R17 (Jeppesen 1971). The nucleic acid of phage R17 was labeled by growing the bacteria infected with this virus in a radioactive medium, that is, a medium to which ^{32}P phosphate was supplied. The phage was isolated and purified by a combination of precipitation with ammonium sulfate and differential centrifugation. The RNA was isolated from the purified phage by extraction with sodium dodecyl sulfate and phenol, and analyzed by the following procedure.

About 20 μg of ^{32}P-labeled R17 RNA (approximately 20 μCi in radioactivity) is digested with 1 μg of ribonuclease T_1 for 30 min at 37° in 3 μl of 0.01M tris-HCl buffer, pH 7.4 containing 1 mM EDTA. The digestion mixture is then separated into oligonucleotides of various sizes and compositions by a two-dimensional procedure of electrophoresis and chromatography such as developed by Brownlee and Sanger and associates (see Brownlee 1972 for extensive details) and applied as follows.

The digest is applied to a 3 cm × 55 cm strip of cellulose acetate and subjected to electrophoresis in 7 M urea buffered with 5 percent (v/v) acetic acid and pyridine at pH 3.5 until the blue and pink marker dyes (xylene cyanol and acid fuchsin, respectively) separate by approximately 15 cm. The oligonucleotides, which can be detected by a portable Geiger counter, extend from the pink spot to about 3 cm behind the blue spot. The oligonucleotides are transferred from the cellulose acetate to a DEAE cellulose thin layer chromatography plate by placing the former on top of the latter and then a pad of Whatman 3 MM paper wet with water is placed on the cellulose acetate. The strips are pressed evenly together by placing a glass plate on top. Water from the paper pad passes through the cellulose acetate carrying the nucleotides with it into the DEAE cellulose where they are held by ion exchange.

Chromatography is performed on the thin layer plate by developing with a 3 percent mixture of partially hydrolyzed (10 min in 0.2 M NaOH at 37°) RNA dissolved in 7 M urea. (The use of carrier oligonucleotides in the chromatography of other oligonucleotides is called homochromatography.) The oligonucleotides from the partial digest of carrier RNA saturate the DEAE groups and displace the radioactive phage oligonucleotides. The latter then travel along in series of fronts with the nonradioactive oligonucleotides in accordance with size, which governs affinity for the DEAE groups. The smaller oligonucleotides are displaced by the larger ones and thus move more rapidly on the thin layer.

After the chromatography is completed, the thin layer is dried and the spots are located by autoradiography. The sort of separation achieved in

such a two-dimensional procedure is illustrated diagrammatically in Figure 21.

Oligonucleotides located by autoradiography can be individually removed from the thin layer and eluted from the DEAE cellulose with 30 percent (v/v) triethylamine carbonate at pH 10. The process of sequencing then involves treating aliquots of the isolated oligonucleotides with different enzymes, analyzing the resultant products, and deducing the sequence from the data obtained. The process of stepwise deduction of sequence can be illustrated for the oligonucleotide (a) (Figure 21) obtained from a ribonuclease T_1 digest of ^{32}P-labeled phage R17 RNA. Enzymes used to digest oligonucleotide (a) and the products found in each digest are summarized in Table 23.

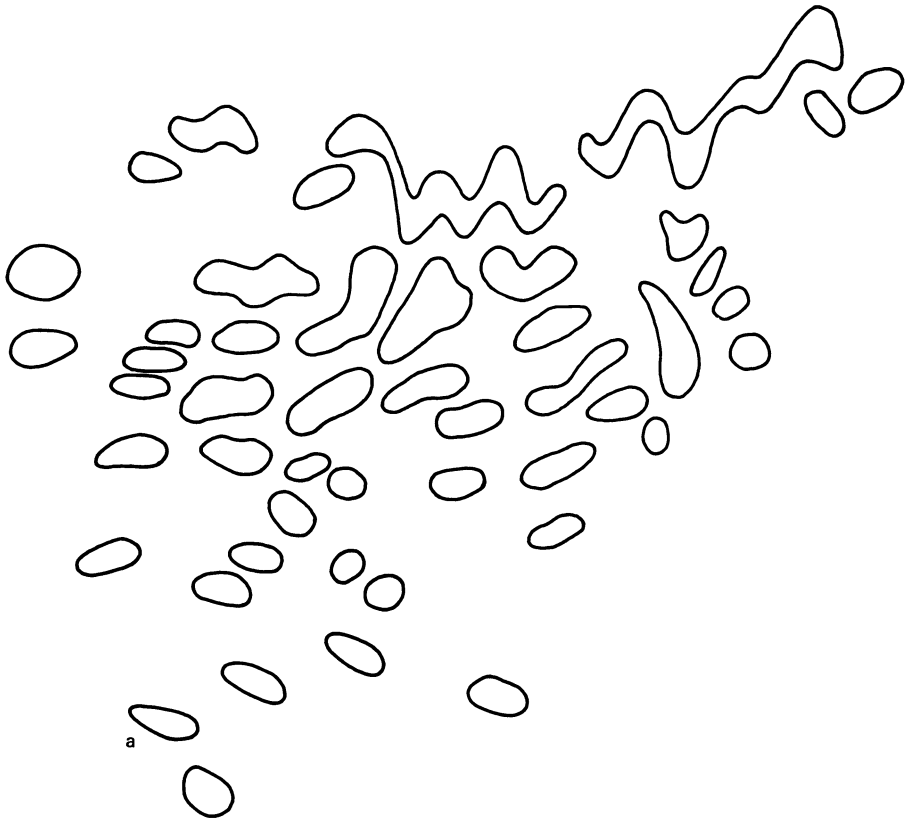


Fig. 21. Diagram of the spots revealed by autoradiography of a two-dimensional thin layer chromatogram. The material fractionated was a ribonuclease T_1 digest of ^{32}P-labeled phage R17 RNA. The oliognucleotide from the spot labeled (*a*) on the diagram is the one whose analysis is described in the text. (Adapted from Jeppesen 1971.)

Owing to the known specificity of ribonuclease T_1 (and confirmed by analysis) the 3′-terminal nucleotide of oligonucleotide (a) is Gp, which is the only guanylic acid residue in the oligonucleotide. Since Ġp (see Table 23 for explanation of Ġp) was found in the ribonuclease A digest of the CD oligonucleotide in which only bonds next to cytidylic acid are split, the 3′-terminal sequence is -C-Gp. A product of ribonuclease U_2 action is (U_4,C)Gp, which enables extension of the deduced 3′-terminal sequence to Pu-U-U-U-U-C-Gp, where Pu stands for purine. Since there is only one G in the oligonucleotide, all of the remaining purines must be A, and therefore the sequence is -A-U-U-U-U-C-Gp.

Among the ribonuclease A products of the CD oligonucleotide (Table 23, column 2) there is $(A\text{-}A\text{-}U,U_3)$Cp, which, because it was derived by ribonuclease action limited to C residues, must be preceded by a C to give $(\text{-}C\text{-}A\text{-}A\text{-}U,U_3)$Cp. This sequence appears to overlap the previously de-

Table 23. Products Obtained by Enzymatic Digestion of Oligonucleotide (a) from Phage R17 RNA.[a]

Ribonuclease A Products[b] (Molar Proportions)	Ribonuclease A Products from CD Oligonucleotide (Molar Proportions)	Ribonuclease U_2 Products (Molar Proportions)
2 A-A-Up	1 (A-A-U̇,U̇)A-A-Cp	1(U_4,C)Gp
1 A-A-Cp	1(A-U̇,U̇$_2$)Cp	<1(U_2,C_2)A-Ap
1 A-Up	1 (A-A-U̇,U̇$_3$)Cp	<1(U_2,C_2)Ap
1 Gp	1 Ġp	<1 U-U-A-Ap
3 Cp	1 Cp	<1 U-U-Ap
6 Up		1(C,U)Ap
		1 A-Ap
		1 Ap

[a]Adapted from Jeppesen 1971.

[b]The respective enzymatic specificities yielding from oligonucleotide (a) the products listed in the three columns of the table may be summarized as follows. Ribonuclease A: The oligonucleotide is cleaved to the right of each *cytidylic acid* and *uridylic acid* residue. Ribonuclease A on CD oligonucleotide: A particular carbodiimide reacts with uridylic and guanylic acid residues to form carbodiimide derivatives (CD products) indicated by U̇ and Ġ in the Table. U̇ residues are resistant to digestion by ribonuclease A. Therefore, cleavages of CD oligonucleotide (a) occur only to the right of *cytidylic acid* residues. Ribonuclease U_2: This enzyme cleaves to the right of purine (adenylic and guanylic acid) residues, but purine-pyrimidine sequences are split more readily than purine-purine sequences. Partial splitting of purine-purine bonds accounts for the less than molar yields of four compounds in column 3 of the table. In the formulas of the table, the nucleotides are indicated by the first letter of their names and terminal phosphoric acid residues by p. The linkage of nucleotides by the conventional 3′-5′ phosphodiester bond is represented by a hyphen when the sequence is known and a comma when the sequence is unknown. Unknown sequences adjacent to known sequences are placed in parentheses. (These arrangements are in accordance with recommendations of the IUPAC-IUB Commission on Biochemical Nomenclature as summarized in *Journal of Molecular Biology*, **55**, 299–310, 1971.)

duced one and combining them yields -C-A-A-U-U-U-U-C-Gp. Looking for oligonucleotides that might overlap this sequence, it is observed that there are three ribonuclease U_2 products terminating in -A-Ap. Since none of the ribonuclease A products contain A-A-A, it appears that the A-Ap observed must come from the 5′-terminal of oligonucleotide (a). Of the remaining two ribonuclease U_2 products with terminal -A-Ap, only (U_2,C_2)A-Ap fits compositional requirements for overlapping the nine-nucleotide segment above.

The U_2C_2 can be arranged in six different ways (U-U-C-C, C-U-U-C, U-C-U-C, C-C-U-U, C-U-C-U, U-C-C-U) but only the three with C on the 3′-end would fit requirements for the overlap. Two of these possibilities are eliminated because they would have produced U-Cp or U-U-Cp upon degradation of CD oligonucleotide (a) by ribonuclease A and these compounds are not among the products. This leaves U-U-C-C as the proper sequence for U_2,C_2 and since the oligonucleotide containing this was obtained by action of ribonuclease U_2, it must be preceded by an A. This brings the known sequence at the 3′-end to -A-U-U-C-C-A-A-U-U-U-U-C-Gp.

The only remaining CD product which can overlap with this is (A-U,U_2)Cp whose sequence must therefore be U-A-U-U-Cp, which adds a U to the 3′-sequence just given. The CD oligonucleotide not yet appearing in the sequence is (A-A-U̇,U̇)A-A-Cp, which, in order to account for the three remaining ribonuclease U_2 oligonucleotides—A-Ap, U-U-A-Ap, and (C,U)Ap (the latter is involved in an overlap)—must have the sequence A-A-U-U-A-A-Cp. Therefore, the total sequence of the 21 nucleotides of oligonucleotide (a) from the above is A-A-U-U-A-A-C-U-A-U-U-C-C-A-A-U-U-U-U-C-Gp.

From this example it should be apparent that sequencing viral RNAs ranging from 3,200 to over 20,000 nucleotides is a formidable task and not likely to be undertaken even for the smallest viral RNAs unless the information to be gained is highly important. However, sequencing of small segments of RNA in order to elucidate initiation and termination signals for translation or to characterize enzyme attachment sites are projects that may warrant the effort required.

Analyses similar to those used for RNA have been employed to determine the sequence of the single-stranded ends (cohesive ends) of the mainly double-stranded DNA of lambdoid phages (phages similar to coliphage lambda) such as Ø80 (Bambara et al. 1973). It appears that the cohesive ends, comprising 12 nucleotides, of the DNAs of phages lambda and Ø80 are identical. This structural feature permits through base pairing the formation of interesting mixed dimers between the two phage nucleic acids. The sequences of the complementary strands are

(5′-3′): -A-G-G-T-C-G-C-C-G-C-C-C-OH
(3′-5′): HO-T-C-C-A-G-C-G-G-C-G-G-G-

A fragment of bacteriophage ØX174 DNA(a single-stranded DNA), 48 nucleotides long, has also been sequenced by using enzymatic and electrophoretic techniques similar to those used on RNA (Ziff et al. 1973). In this work good use was made of the T4-induced enzyme, endonuclease IV, which cleaves DNA to yield cytidine 5′ phosphate terminals.

i. Two Ways to Compare Nucleotide Sequences Without Sequencing: Nearest Neighbor Analysis and Hybridization

If there is an *in vitro* method available for the synthesis of nucleic acids from radioactive substrates (specifically, nucleoside triphosphates containing $\alpha^{32}P$), it is possible to compare nucleic acids in terms of the frequencies with which various nucleotide pairs (doublets) occur (since most nucleic acids contain four different nucleotides, the total number of different nucleotide pairs is $4^2 = 16$). This procedure, which has been termed "nearest neighbor analysis," was initially proposed by Josse et al. (1961).

The basic plan of the procedure is to use the nucleic acid whose analysis is desired as a template ("primer") for synthesis by a polymerase enzyme of complementary strands containing radioactive phosphorus at specific points in accordance with which radioactive triphosphate was used in the substrate mixture. Thus four successive syntheses are performed in which all four of the usual nucleotides are present, but in each case a different deoxyribonucleoside triphosphate contains the ^{32}P marker:

Reaction 1: ppp*A, pppG, pppC, pppT + template + polymerase
Reaction 2: pppA, ppp*G, pppC, pppT + template + polymerase
Reaction 3: pppA, pppG, ppp*C, pppT + template + polymerase
Reaction 4: pppA, pppG, pppC, ppp*T + template + polymerase

The mechanism of synthesis with DNA polymerase is that the substrates are 5′-deoxyribonucleoside triphosphates and these are linked into polynucleotide chains by esterification with the 3′-OH of adjoining nucleotides (see Figure 22). After synthesis is complete, the product is digested enzymatically to yield 3′-nucleoside monophosphates. In effect, the ^{32}P goes into the product in one nucleotide and comes out in digested product attached to the neighboring nucleotide (see Figure 22).

The nucleotides liberated by enzymatic digestion are separated by paper electrophoresis and the ^{32}P content of each is estimated in an appropriate counter. From the data the frequency with which any nucleotide occurs next to any other can be calculated. This then provides a measure of the frequency with which each of the 16 possible dinucleotides occurs in the nucleic acid in question.

This type of analysis appears to give patterns of dinucleotide frequency that are reproducible and characteristic for different nucleic acids and hence are useful in comparing nucleic acids. The examples given in Table 24 are drawn from the more extensive compilations of Subak-Sharpe et al.

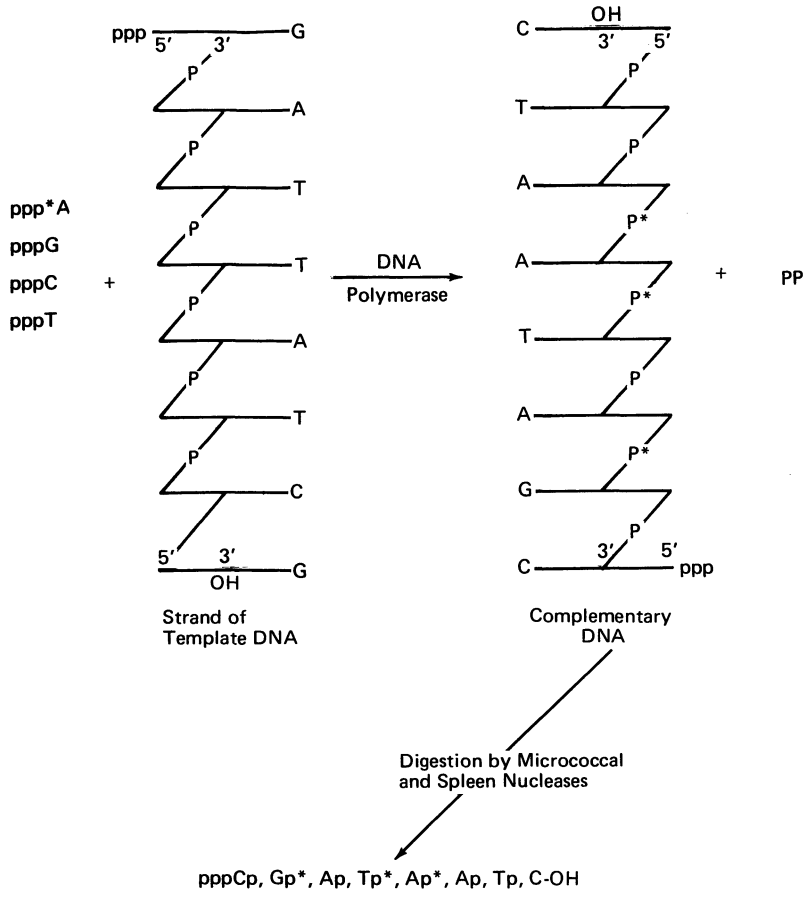


Fig. 22. Synthesis of new DNA from deoxyribonucleoside 5′ triphosphates. Radioactive phosphorus is indicated by p*. The new strand has a composition complementary to that of the template (when the template is double-stranded DNA, both strands serve as template for the polymerase). From the knowledge that each phosphate built into the new DNA is attached to the 5′ position of the entering nucleotide but leaves upon enzymatic digestion attached to the 3′ position of the neighboring nucleotide (the other partner of the phosphodiester linkage), the following dinucleotides can be deduced from the nature of the radioactive nucleotides in the enzymatic digest of the polymerase product shown above: GpA, TpA, and ApA.

Table 24. Some Doublet Frequencies[a] for DNA from Viruses and Other Sources.[b]

Doublet	Polyoma	Shope pap.	Vaccinia	Herpes	Phage λ	T1	T4	T5	ØX174	*E. coli* 2	Hamster BHK 21
ApT	56	57	69	68	70	64	64	65	59	72	53
TpA	50	45	58	68	47	48	55	62	46	47	50
GpC	64	70	54	65	72	76	70	78	80	72	75
CpG	22	30	72	65	64	70	53	58	59	64	13

[a]In parts per thousand normalized to correspond with DNA containing 50 percent (G + C).
[b]Adapted from Subak-Sharpe et al. 1966.

(1966). Even with only four of the 16 doublets represented in Table 23, it can be seen that there are substantial similarities between some of the T phages [Josse et al. (1961) found even closer correspondence, as would be expected on chemical, morphologic, and serologic grounds between the doublet frequencies for T2, T4, and T6], but also distinct differences among the several groups of viruses represented in the table. Comparison of the doublet frequencies of viral and host DNAs [see Table 24 and the more detailed results of Subak-Sharpe et al. (1966); and Bellett (1967)] leads to the conclusion that the smaller DNA-containing viruses exhibit doublet frequencies resembling those of mammalian animal cells (compare the values in Table 24 for the DNAs of polyoma and Shope papilloma viruses with those listed for hamster BHK 21 cells), whereas the DNAs of such large viruses as herpes and vaccinia, like those of phages, are more similar to bacterial DNA, such as that of *E. coli*. While the doublet frequencies do not provide adequate evidence to establish phylogenetic relationships, it has been suggested that the observed data are consistent with the idea that some bacterial and animal viruses evolved from their hosts, whereas others, such as herpes and vaccinia viruses may have had an external origin, perhaps from bacteria.

Limited nearest neighbor analyses have also been made on viral RNAs, including those of tobacco mosaic and turnip yellow mosaic viruses and of phage MS2 (Fox et al. 1964). Such analyses have also been especially useful in sequencing the RNA of some small phages (Bishop et al. 1968; Billeter et al. 1969).

In summary, comparisons of nucleotide doublets by the nearest neighbor technique can be very useful. However, such comparisons are in no way a substitute for the more arduous linear sequencing method outlined earlier. For example, as Josse et al. (1961) indicated, analytical errors as little as 1 percent, which are not unlikely, mean that even the moderate-size phage lambda DNA (molecular weight 20×10^6) could differ from a related DNA in 1,000 nucleotide sequences without any differences being detected by the nearest neighbor analyses.

The molecular hybridization technique is a method for obtaining infor-

mation about nucleic acid sequences that are very much longer than the doublets of the nearest neighbor method. The basis for this procedure is as follows.

Hydrogen bonds between complementary bases (A:T or U, and G:C) of the two strands of double-stranded nucleic acid (Figure 24) are broken by heating (and also by certain chemicals such as alkali or dimethyl sulfoxide), but the bonds are reformed by slow cooling ("annealing" or "renaturation") or removal of added chemical in the case of chemically induced destruction of hydrogen bonds. The separation of nucleic acid strands by heat is called molecular melting or denaturation, the latter being the more general term for strand separation by whatever cause. If strands of a second species of nucleic acid are added to those of the originally denatured one during the annealing process, strands of the second nucleic acid may compete with the original ones in the reforming of double stranded structures. This occurs only if substantial nucleotide sequences are the same or very similar in the two species of nucleic acid. If such homology exists, "hybrids" may be formed between complementary strands of DNA or between complementary strands of DNA and RNA; thus, the term hybridization is applied to the process.

Experimentally, either agar gels, or more commonly now, nitrocellulose filters, are used to immobilize denatured strands of nucleic acid. Hybrids can be formed if complementary structures are brought in contact with such fixed, denatured nucleic acid. (Appropriate nitrocellulose membranes bind denatured DNA, DNA-DNA, and DNA-RNA hybrids, but not free RNA or undenatured DNA.) In addition, there are procedures in which columns containing hydroxyapatite or other substances are employed. In almost all cases, quantitation is achieved by having one of the reacting species radioactively tagged. Details of various methods are given by Raskas and Green (1971) and Bøvre et al. (1971). An evaluation of the specificity of hybridization reactions is given by McCarthy and Church (1970).

Some examples of situations in which homologous sequences of nucleic acid or lack of them can be demonstrated by molecular hybridization include an evaluation of the degree of similarity of the nucleic acids of viruses that are thought to be similar, for example, comparison of the DNAs of the many types of adenovirus; investigation of the amount of replicating viral nucleic acid present at different times after infection; estimation of the number of viral genomes incorporated into host nucleic acid; determination of the presence or absence of virus-specific messenger RNA (mRNA); and discrimination between types of mRNA present at various times after infection.

An example involving simian virus 40 (SV40) can be cited here of the use of molecular hybridization to demonstrate the presence and number of copies of SV40 DNA integrated into the DNA of mouse cells (Westphal and Dulbecco 1968). Such integration of viral genome is thought to be as-

sociated with the transformation of normal cells to tumorous cells. In order to achieve greater sensitivity in the hybridization test, instead of seeking direct hybridization between viral DNA and mouse cell DNA, RNA complementary to SV40 DNA (cRNA) was employed. Two advantages of using cRNA rather than the direct approach with viral DNA are that higher specific radioactivities can be readily obtained in the cRNA and the use of cRNA eliminates the possibility of self-annealing of denatured viral DNA. Reconstruction experiments indicate that as few as three or four viral DNA molecules per cell can be detected by hybridization with cRNA (this is equivalent to less than two parts of viral DNA in a million parts of cellular DNA).

cRNA is synthesized *in vitro* from highly radioactive nucleoside triphosphates using SV40 DNA as primer and a DNA-dependent RNA polymerase. Hybridization is carried out essentially by the method of Gillespie and Spiegelman (1965). The circular, supercoiled SV40 DNA molecules are treated with deoxyribonuclease for 60 min at 30° to convert them into linear, circular strands with some breaks in one or another of the strands. This DNA is poured into 2 vol of boiling water and boiled for 15 min, thus effecting strand separation. The denatured DNA is chilled in ice and adjusted to contain 0.9 M NaCl–0.09 M Na citrate (called 6× SSC, 1 SSC being 0.15 M NaCl–0.015 M Na citrate). The DNA is then slowly passed through a Millipore membrane filter to which much of the denatured DNA attaches. In order to check the variability of results, replicate filters are employed with comparable aliquots of DNA.

The filters are next incubated at 66° for 22 hr in vials containing 1 ml 6X SSC, tritiated cRNA obtained as described above, 1 mg yeast RNA carrier, and sodium dodecyl sulfate at a concentration of 0.1 percent (the latter appears to reduce background counts by reducing nonspecific attachments of the radioactive cRNA). The filter-containing vials are gently shaken in a water bath during incubation. After incubation, during which hybridization is expected to occur when possible, the filters are removed from the vials and washed with 50 ml of 2X SSC using suction filtration. After this the filters are treated with 20 μg/ml of RNase A and 10 units/ml RNase T_1, for 60 min at 37°. Following this treatment, which is designed to remove RNA complexed to the DNA over short regions and hence of uncertain specificity, the filters are washed again, dried, and counted in a scintillation counter. The amount of DNA on each filter to which the counts need to be related is determined after counting by the colorimetric diphenylamine reaction.

In the investigation by Westphal and Dulbecco (1968) it was found that the DNAs of two different lines of mouse cells that had been transformed by SV40 (SV3T3-47 and SV3T3-56) fixed by hybridization different amounts of the SV40 specific cRNA. The counts per minute per 100 μg DNA above the backgrounds observed with the DNAs of untransformed

cells were, respectively, 800 and 264 (these are the means from counts of 10 to 12 filters each). From calibration tests in which known numbers of SV40 DNA molecules had been added to cell DNAs, it was determined that 40 cpm from attached cRNA was equivalent to one SV40 DNA; therefore, the counts per minute noted above represent about 20 and 7 SV40 DNA molecules, respectively, present in the DNAs of the two mouse cell lines that had been transformed by the virus. In these same experiments, it was demonstrated by hybridization that the SV40 DNA was in the nucleus rather than in the cytoplasm of transformed cells.

Another example may be cited, without giving the experimental details, to illustrate the many useful applications of molecular hybridization. Lacy and Green (1967) investigated the hybridization reactions between the DNAs of six members of the weakly oncogenic (oncogenic means tumor inducing) adenovirus group consisting of serological types 3, 7, 11, 14, 16, and 21. It was found that these viral DNAs are closely related, apparently sharing 70–100 percent of their nucleotide sequences. However, the DNAs of the weakly oncogenic adenoviruses apparently differ substantially from those of the strongly oncogenic types 12 and 18, for the results of the hybridization tests indicated that the two groups showed only 11–22 percent homology.

j. Secondary and Higher Structure of Nucleic Acids

The term "secondary structure" will be used here as it was in connection with proteins to mean geometric configuration, with special reference to the presence or absence of helical, hydrogen-bonded structures; similarly, folding of the polynucleotide chain can be considered "tertiary structure."

A secondary structure for DNA was proposed by Watson and Crick (1953a, 1953b) that seemed at once compatible with data on the composition of DNA, general chemical features of DNA, and x-ray diffraction data (Wilkins et al. 1953). The validity of this structure has been confirmed by many experiments over the succeeding years, and it is now widely accepted for DNA from many sources.

This DNA structure is briefly described as a dyad, or duplex, of right-handed helical chains each coiled around the same axis but with antiparallel nucleotide sequences (sequences running in opposite directions). Such an arrangement is shown diagrammatically in Figure 23. The two chains are held together by hydrogen bonding between complementary pairs of bases, one base of each pair being a purine and the other a pyrimidine. Thus, as shown in Figure 24, adenine pairs with thymine and guanine with cytosine. In terms of DNA composition, this should be reflected in A/T and G/C ratios of unity. It will be further noted (Figure 24) that two hydrogen bonds can readily form between adenine and thymine on adjacent strands, but three can form between guanine and cytosine.

One of the first examples of DNA to be shown to give an x-ray diagram consistent with the Watson-Crick structure was that of T2 coliphage (Wilkins et al. 1953). At first, the results of chemical analyses seemed not to support this conclusion. The double helix structure requires equimolar amounts of adenine and thymine and of guanine and cytosine for proper base pairing. Early analyses (Wyatt and Cohen 1952) showed T2 DNA to contain 33.2 moles adenine, 35.2 moles thymine, 17.9 moles guanine, and 13.6 moles 5-hydroxymethylcytosine per 100 moles of bases. The agreement between adenine and thymine was fair, but the guanine/5-HMC ratio was seriously off. With the development of the double helix theory, interest in analyses grew, which led to refinements in procedure, and in this case especially, a recognition of the lability of 5-HMC under common hydrolytic conditions. When these factors were adjusted for, both the A/T and G/5-HMC ratios were found to be close to unity (Wyatt and Cohen 1953). The molar equivalence of purines to pyrimidines in double-stranded viral DNAs is illustrated by many examples in Table 19.

The double helical secondary structure of DNA imparts noteworthy properties to the particle that distinguish it from a single-stranded structure. Primarily, the double-stranded structure has greater rigidity and order than the single-stranded one, and this is reflected in hydrodynamic behavior, optical properties, and chemical reactivity. Thus the transition from helical structure to the less ordered random coil structure assumed by separated strands (or vice versa, since the process is more or less reversible) can be followed by:

1. Sedimentation behavior (helical form sediments slower, that is, has lower sedimentation coefficient).
2. Viscosity (helical form is more viscous than random coil). The properties described in (1) and (2) apply to helical and denatured DNAs in dilute salt in the middle pH range; under these conditions, the separated strands of denatured DNA collapse to a globular form that sediments faster and has a lower viscosity than undenatured DNA. If conditions are employed to keep the separated strands of denatured DNA extended, they may sediment slower and have a higher viscosity than the helical structure (see Studier 1965).
3. Optical rotation (helical form has "handedness" so that it acquires optical activity above that inherent in its components, such as the sugar. Thus optical activity is proportional to helicity).
4. Ultraviolet absorption (helical form absorbs less than random coil at 260 nm because of stacking of bases. The state of reduced ultraviolet absorption characteristic of an hydrogen-bonded, ordered structure is called "hypochromicity." An increase in absorption is then called hyperchromy and a decrease, hypochromy).
5. X-ray diffraction (helical forms have distinctive x-ray diffraction characteristics, of which the absence of meridional reflections is outstanding).

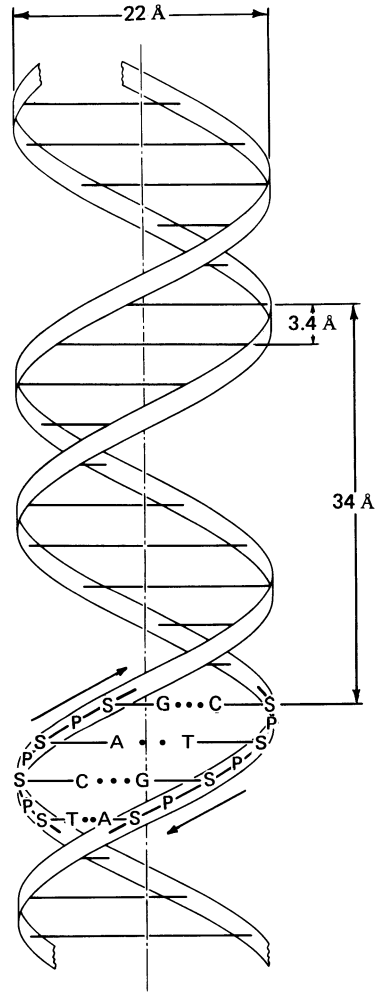


Fig. 23. Diagrammatic sketch of the structure of DNA (modified from Watson and Crick 1953b) by indication of the components in a segment: P, phosphate; S, sugar (2-deoxyribose); G, guanine; C, cytosine; A, adenine; T, thymine. The two ribbons represent the sugar-phosphate backbones of the two helical strands of DNA, which, as the arrows indicate, run in opposite directions, each strand making a complete turn every 34 Å. The horizontal rods symbolize the paired purine and pyrimidine bases. There are ten bases (and hence ten nucleotides) on each strand per turn of the helix. The nature of the base pairs and the number of hydrogen bonds between them are shown in the detailed central segment. The vertical line marks the fiber axis. (From Knight 1974.)

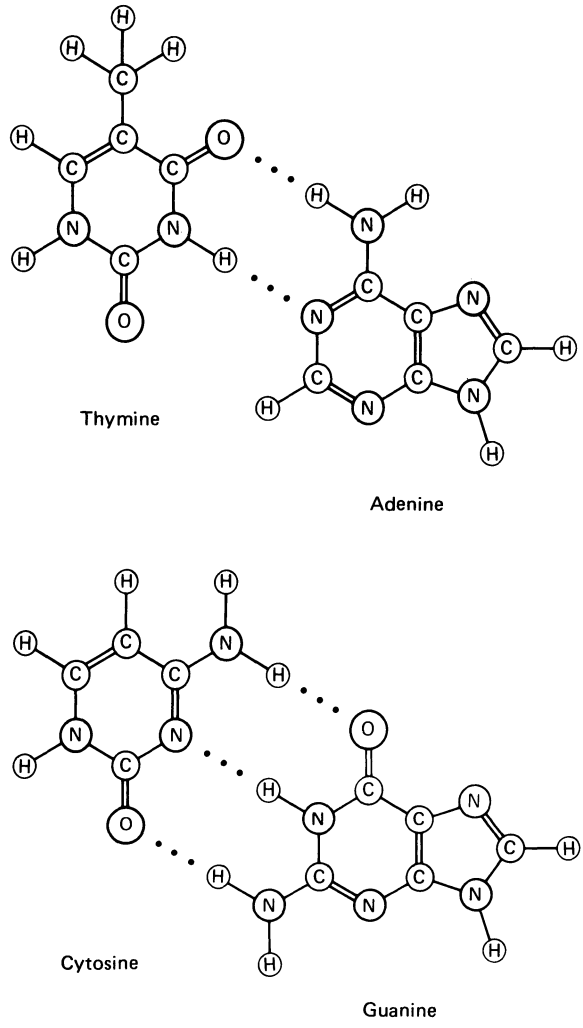


Fig. 24. Illustration of common hydrogen bonding that occurs between specific purine and pyrimidine bases in nucleic acids.

6. Chemical reactivity (keto and amino groups involved in hydrogen bonding do not titrate like free groups, and the reaction of amino groups with reagents like formaldehyde or nitrous acid is inhibited by hydrogen bonding).

7. Chromatography (double-stranded nucleic acid adheres to hydroxyapatite at salt concentrations and temperatures at which single-stranded molecules readily elute).

8. Electron microscopy (under appropriate conditions the thicker strands of double-stranded nucleic acid can be distinguished from those of single-stranded nucleic acid).

The nucleic acids of most viruses are linear structures but some, such as those of DNA tumor viruses and of some phages, are circular (see Table 16). The circular DNAs of animal tumor viruses tend to occur not only in circles but also in supercoiled forms or twisted circles. The significance of these unusual structures is not clear.

Turning now to RNA, it has been noted that the RNA of TMV is held in a helical configuration in the intact virus, and follows the pitch of the protein subunits at a radius of 40 Å from the long axis of the particle. In the TMV particle, the RNA therefore may be represented as a helix with a diameter of 80 Å and a pitch of 23 Å. Since this pitch is too large to permit hydrogen bonding between bases on successive turns of the helix, the helical structure must be stabilized simply by its position in the helical groove formed by the protein subunits (Klug and Caspar 1960). This situation is illustrated in Figure 18, a model drawing based on the results of the x-ray studies described earlier showing a segment of the TMV particle from which the protein subunits of the last two turns are removed. The nucleotides are indicated by the little discs in the RNA chain.

In contrast to the regular helical form assumed by TMV-RNA in the virus particle, the high molecular weight RNA isolated from the virus by the methods described earlier behaves in solution at low temperature and low salt concentration as though it were a flexible but tight random coil (Haschemeyer et al. 1959; Boedtker 1959; Gierer 1960) arising from a single polynucleotide strand. However, the strictly random coil concept is almost surely too simple to account for the observed physical properties of such isolated RNA in solution. The marked changes of properties of the RNA in different ionic media and at different temperatures strongly suggest the ready formation of secondary valence bonds under one set of conditions and the rupture of these bonds under another set (Boedtker 1959; Haschemeyer et al. 1959). Since there is no evidence for combination of separate RNA strands, it must be assumed that it is possible to form intramolecular bonds. To explore this possibility, Doty et al. (1959) applied to TMV-RNA some of the tests for helix-coil transition listed above.

In taking TMV-RNA in 0.1 M phosphate at pH 7 from 10° to 70°, Doty et al. (1959) found a 32 percent increase in absorption in the ultraviolet at 260 nm. This change was at least 95 percent reversible upon cooling. At room temperature, treatment of the RNA with 6 M urea caused half of the increase in absorbance observed in the thermal experiment.

In another test, the reaction of the RNA with formaldehyde at different temperatures was followed spectrophotometrically. This test was based on the observation by Fraenkel-Conrat (1954) that treatment of TMV-RNA with 1–2 percent formaldehyde at pH 6.8 caused a gradual increase in the ultraviolet absorbance at the maximum as well as a shift of 3–5 nm toward higher wavelengths. This effect seems to depend upon the presence of free amino groups in adenine, guanine, and cytosine, and is illustrated by the results shown in Table 25. In the experiments of Doty et al., an increase in reactivity of TMV-RNA with formaldehyde at 45° as compared with 25°

Table 25. Effect of Formaldehyde on the Ultraviolet Absorption of Viruses, Proteins, Nucleic Acids, and Some Nucleic Acid Constituents.[a]

Material	Approximate Increase of Maximum Absorption,[b] %	Approximate Shift in Wave Length of Maximum, nm
Nucleic acids and constituents:		
TMV-RNA (prepared by hot salt method)	+29	+3
TMV-RNA (prepared by detergent method)	+28	+3
Liver-RNA	+19	+4
Yeast-RNA (commercial)	+24	+3
Thymus-DNA	None	None
Adenine	+23	+5
Adenosine	+19	+5
Adenylic acid	+22	+5
Guanylic acid	+5	+5
Cytidylic acid	+16	+3
Thymine	None	+3
Uracil	None	+1
Uridine	None	None
Uridylic acid	None	None
RNA-Containing Viruses:		
TMV	+3	+3
TMV (after 24 hr in 1% sodium dodecyl sulfate)	+18	+3
Tomato bushy stunt virus	+15	+4
Turnip yellow mosaic virus	+15	+3
Tobacco ringspot virus	+23	+3
DNA-Containing Viruses and Proteins:		
Shope papilloma virus	None	None
T2-coliphage	−3	None
T2-coliphage (after 24 hr in sodium dodecyl sulfate)	None	None
TMV protein	None	None
Bovine serum albumin	None	None
Ovomucoid	None	None

[a]Adapted from Fraenkel-Conrat 1954.

[b]Solutions containing the equivalent of about 0.025–0.05 mg of nucleic acid per milliliter in 0.1 M phosphate buffer at pH 6–8 were treated with 1–2 percent formaldehyde for 12 hr at 40° or 48 hr at 23°. The same maximum was reached at the two time and temperature levels.

(after 50 min) was about 19-fold, whereas a control mixture of the appropriate nucleotides showed only a sixfold increase.

The results cited so far definitely support the concept that in dilute, neutral salt at low temperature, portions of the TMV-RNA chain are bound to other portions of the same molecule by hydrogen bonds, presumably of the base-pairing sort, but these results do not answer the question of whether the hydrogen bonding is random or occurs in definite regions in such a manner as to provide helical segments. Evidence for helicity was obtained by Doty et al. (1959) by studying the effect of temperature on the optical rotation of TMV-RNA. If the RNA were devoid of any regular secondary structure, that is, if it were a random coil, its only optical activity would be that of the ribose, which has asymmetric carbon atoms. However, if a significant portion of the RNA had helical structure, the contribution to optical rotation, as judged from known helical structures, might be substantial. Furthermore, such optical rotation should be largely abolished by treatments that convert helical structures to random coils. It was found that the specific rotation of TMV-RNA decreased about 160° in going from a temperature of 8° to about 75°. Most significantly, the optical rotation-temperature and optical density-temperature profiles can be shown to coincide by adjusting the ordinate scales as shown in Figure 25. Thus the decrease in optical rotation of TMV-RNA with rise in temperature is approximately congruent with the increase in optical density observed.

From these results, Doty et al. conclude that the hydrogen bonding in TMV-RNA occurs in definite areas and results in helical segments. Some support for such helical structure is provided by the x-ray pattern obtained by Rich and Watson (1954), which shows some of the characteristics of the patterns obtained with helical DNA. An estimate of the extent of these helical regions was made by Doty et al. by comparing the maximum variation with temperature of the specific rotation (or optical density) of TMV-RNA with the maximum variation observed with the completely helical model, polyadenylic acid-polyuridylic acid. The resulting conclusion is that about 50–60 percent of the nucleotides in TMV-RNA are involved in helical regions. It is further postulated, from experiments with polyribonucleotides and by analogy with DNA, that the predominant base-pairing is probably between adenine and uracil and between guanine and cytosine. Unmatched bases are also predicted and the resulting structure, illustrated in part in Figure 26, consists of a number of imperfect helical loops with randomly coiled regions interspersed. Such a structure would be compatible with the observed hydrodynamic properties of TMV-RNA and would especially account for variations in physical properties with changes in environment. A somewhat similar model has been deduced from the results of x-ray studies made on RNA from ascites tumor cells, *E. coli*, and yeast by Timasheff et al. (1961). They concluded that their RNA was represented by short, rigid, double helical rods about 50–150 Å long joined by small flexible single-stranded regions.

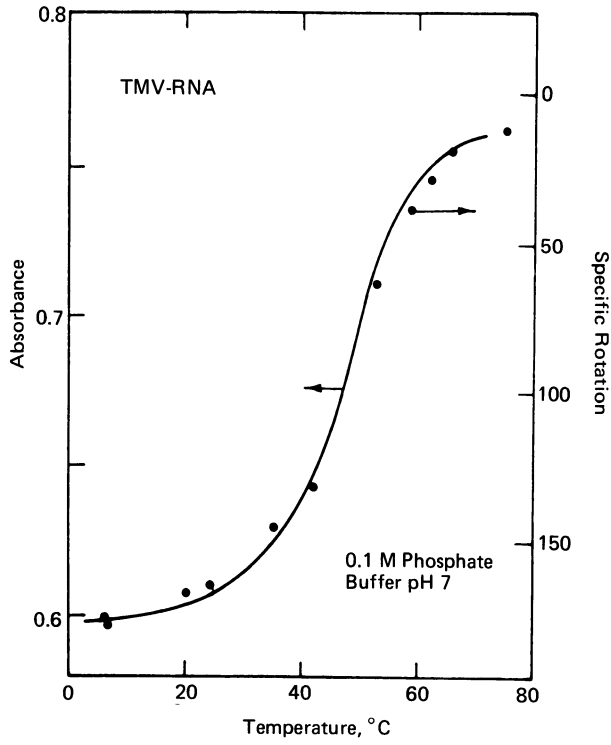


Fig. 25. The variation of specific rotation (*dots*) and adsorbance (*solid line*) of TMV-RNA with temperature. (From Doty et al. 1959.)

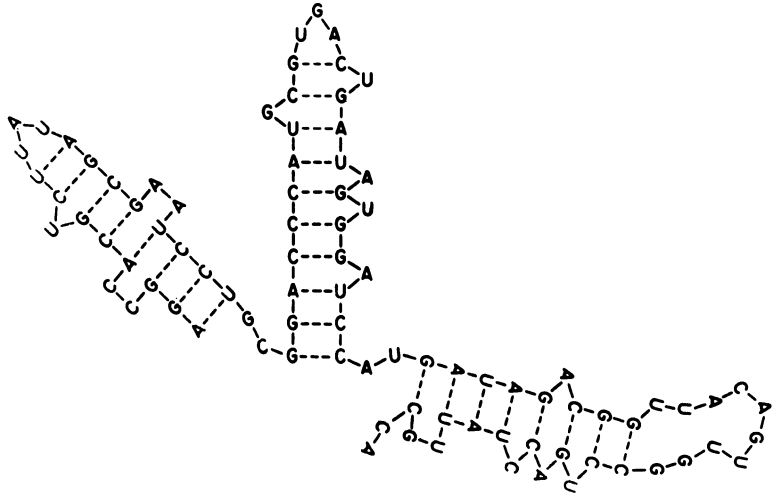


Fig. 26. A possible model for a segment of TMV-RNA in dilute neutral salt solution at low temperature. A, C, G, and U are abbreviations for the nucleotides: adenylic, cytidylic, guanylic, and uridylic acids, respectively. (From Doty 1961.)

Viral nucleic acids, whether single stranded, double stranded, or cyclic, generally appear to occur in unbroken strands. In contrast to this generality there are some RNA viruses whose nucleic acid appears to occur in segments. Thus the double-stranded RNA genomes of reovirus and of cytoplasmic polyhedrosis virus of the silkworm occur in ten segments (Shatkin et al. 1968; Millward and Graham 1970; Lewandowski and Millward 1971) and that of the wound tumor virus of clover in 12 segments (Kalmakoff et al. 1969; Reddy and Black 1973). (A viral genome may be defined as the total ensemble of genes associated with a virus.) Separation of the RNA segments of the cytoplasmic polyhedrosis virus by electrophoresis in acrylamide gel is illustrated in Figure 27. Similar segments but consisting of single-stranded RNA seem to characterize influenza viruses (Duesberg 1968; Pons and Hirst 1968), and possibly RNA tumor viruses (Vogt 1973).

It should be emphasized that, in general, the physical properties of nucleic acid within a virus particle may or may not be the same as those of the extracted nucleic acid treated as a hydrodynamic entity. This was demonstrated by Bonhoeffer and Schachman (1960) using four viruses, two of which contained DNA and two RNA. A comparison of the ultraviolet absorption spectra before and after degradation with sodium dodecyl sulfate with the spectra obtained by heating the degradation mixtures was used as a measure of the degree of hypochromicity (and hence of hydrogen-bonded structure) of the nucleic acid within the virus particles and upon release from them. In the case of the DNA viruses, Shope papilloma virus and T6 coliphage, no change in secondary structure upon release of the nucleic acid could be detected. Upon heating the degraded viruses, however, the ultraviolet absorptions increased 30–35 percent, from which it can be assumed that the DNA is present, in each of the cases, within the virus as well as upon release, in the form of the classical double helix.

On the other hand, a definite decrease in ultraviolet absorption occurred upon release of RNA from TMV, as shown in Figure 28. Heating the degradation mixture restored the absorption to the level of the undegraded virus. This result confirms the deduction made from the x-ray and other data (see above) that the spacing of RNA in the TMV particle precludes base-base interaction, which, however, occurs intramolecularly upon release of the nucleic acid. Thus the TMV-RNA appears to go from the protein-imposed helical configuration of the intact virus illustrated in Figure 18 to an intramolecularly, partially hydrogen-bonded structure such as shown in Figure 26.

Upon degradation of the bushy stunt virus, a slight decrease in absorbance was noted. Heating the degraded virus caused a 23 percent increase. From these facts it was concluded that the RNA within bushy stunt virus has some secondary structure and that more is acquired upon release of the RNA from the particle.

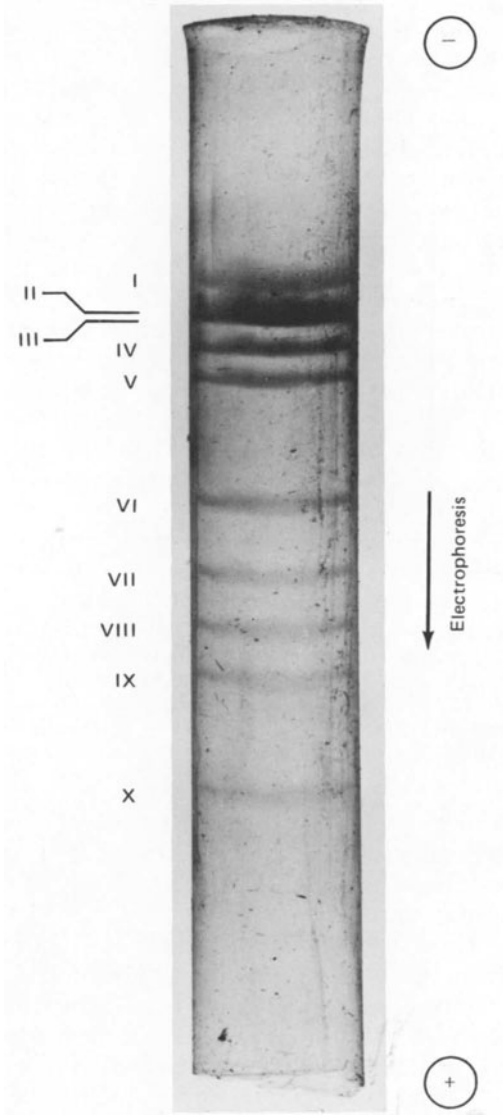


Fig. 27. Polyacrylamide gel electrophoresis of the double-stranded RNA genome of cytoplasmic polyhedrosis virus of the silkworm. There are ten segments of RNA, and all except segments II and III separated under the conditions of electrophoresis used for this experiment (3 percent polyacrylamide gel, pH 7.5, stained with methylene blue). (Courtesy B. L. Traynor.)

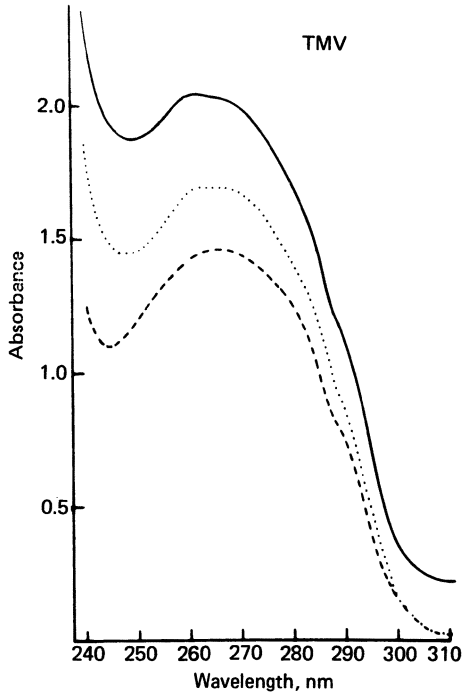


Fig. 28. Ultraviolet absorption spectra of TMV before and after degradation by sodium dodecyl sulfate into RNA and protein. Solid and dotted curves are for the intact nucleoprotein, the former giving the observed optical density values and the latter the values after application of light-scattering corrections. The dashed curve was obtained after degradation of the virus into RNA and protein. (From Bonhoeffer and Schachman 1960.)

Therefore, at present it appears that the nucleic acid within viruses can have (1) no independent secondary structure (TMV); (2) extensive secondary structure (Shope papilloma virus and T6); (3) partial secondary structure (bushy stunt virus).

Finally, it has been suggested by Fresco et al. (1960) that the unpaired bases in such semihelical RNA structures as illustrated in Figure 26 constitute a tertiary structure since they are held in loops or folds in fixed relation to the rest of the structure. It is further suggested that their conformation might provide new possibilities in coding and information transfer that are not inherent in a structureless single strand or in a perfect base-paired helix. These suggestions remain open for further investigation.

Investigations have also been made to determine whether the segmented state of certain RNA genomes mentioned above might be an artifact of preparation or represent the state of the nucleic acid in the viral particle as well. Application of the periodate oxidation-borohydride reduction pro-

cedure described in Sec. 2, g2 showed the same number of 3′ terminal groups for the RNA within the virus particles as in the isolated nucleic acid. Hence it appears that the nucleic acid in reovirus, cytoplasmic polyhedrosis virus, and influenza virus particles is segmented to the same extent as the nucleic acid isolated from such particles (Millward and Graham 1970; Lewandowski and Millward 1971; Lewandowski et al. 1971).

3. *Function of Viral Nucleic Acids*

That nucleic acids are the genetic material of viruses is now widely accepted. This view did not appear suddenly but evolved through a series of observations made over a period of years. The following are noteworthy examples.

a. A Suggestive Idea from Bacterial Transformation

In 1944 Avery et al. reported experiments in which it was shown that minute amounts of DNA extracted from Type III pneumococci were able, under appropriate cultural conditions, to induce the transformation of unencapsulated R variants (characterized by rough colonies) of pneumococcus Type II into fully encapsulated S cells (characterized by smooth colonies) of pneumococcus Type III. Evidence that the active transforming agent was actually DNA, unaided by protein, polysaccharide, or any other substance, mounted over the years, together with reports of transforming principles in other bacterial systems (see reviews by Zamenhof 1957; Hotchkiss 1957; Hotchkiss and Gabor 1970). Since the presence or absence of capsules was known to be genetially determined, the potential significance of DNA in bacterial genetics was made apparent by transformation phenomena. Moreover, other hereditary characteristics are now known to be transferred in this manner.

b. A Hint from the Chemical Analysis of Spontaneous Mutants of Tobacco Mosaic Virus

At a time when proteins were considered largely responsible for the biological properties of viruses, enzymes, and certain hormones, some mildly disturbing results were obtained upon analysis of the protein coats of spontaneous mutants of tobacco mosaic virus (Knight 1947a). At least one strain was found whose protein appeared to have the same composition as that of common TMV, although this strain caused markedly different symptoms in infected Turkish tobacco. Therefore, it was suggested that the primary change responsible for mutation might be in the nucleic acid of the virus. However, it was not certain that the amino acid analyses were accurate enough to have detected a small but significant difference between mutants (although this was later shown to be true), and there was the

possibility that the proteins of the two mutants were identical in composition but differed in sequence of amino acids. Consequently, these findings had less effect than might have been expected on the direction of thought about the role of viral nucleic acids.

c. RNA Shown Essential for Plant Virus Duplication

A few years after the work of Avery et al. (1944) on the pneumococcal transforming principle, Markham and Smith (1949) isolated and crystallized a new plant virus from turnip, which they called turnip yellow mosaic virus. This virus proved to be homogeneous in the Tiselius electrophoresis apparatus, but had two distinct components as judged from sedimentation studies. The major component, comprising 70–80 percent of the material by weight, and the minor component, comprising the balance of the material, had sedimentation coefficients of 106 S and 49 S, respectively. In other physicochemical properties, the two components were virtually identical save that the major component (called "bottom component" from sedimentation behavior) contained about 37 percent RNA, whereas the minor component ("top component") had essentially none. Significantly, the particles containing RNA were found to be highly infectious while those lacking RNA were noninfectious.

These findings could be interpreted to mean that only the combination of protein and nucleic acid is infectious, or that nucleic acid alone is essential for infectivity. Markham (1953) took the latter view in a paper presented at Oxford in April 1952, in which he said, "The role of the protein constituent of plant viruses is undoubtedly very important, but there is some evidence that the nucleic acid is in fact the substance directly controlling virus multiplication."

d. Role of DNA in Infection by T Phages

The idea of the hereditary primacy of viral nucleic acid received a great stimulus from Hershey and Chase's study (1952) of the process of infection by *E. coli* by coliphage T2. Using phage whose protein was labeled with ^{35}S and whose DNA contained ^{32}P, they showed that at least 80 percent of the phage sulfur (and hence most of the protein) remained on the outside of infected cells, whereas only 21–35 percent of the phosphorus (representing DNA) remained outside. The bulk of the protein was mechanically removed at this stage, and yet the cells went ahead and produced T2 phage. Furthermore, 30 percent or more of the parental phosphorus was found in the progeny phage in contrast to less than 1 percent of the sulfur. From these and other facts, it was proposed that the DNA probably exercises the genetic function of the phage, and the protein of a mature phage particle acts as a protective coat for the DNA and is responsible for the adsorption of the phage to the bacterium and the injection of the DNA into the cell.

e. Infectious Nucleic Acid from Tobacco Mosaic Virus

The crowning evidence that nucleic acid is the prime germinal substance of viruses was obtained when it was shown that TMV-RNA is infectious. This was demonstrated after the Hershey-Chase experiment, but a few years prior to the report that bacterial protoplasts (bacteria whose cell walls had been enzymically removed), in contrast to whole bacteria, could be directly infected with phage nucleic acid.

Fraenkel-Conrat (1956) reported that RNA preparations obtained from TMV by treatment of the virus with sodium dodecyl sulfate (see Detergent Procedure in section on Methods for Preparing Viral Nucleic Acids) was infectious though apparently devoid of characteristic virus particles, and that this infectivity was abolished by treatment of the preparation with ribonuclease. At about the same time, Gierer and Schramm (1956a, 1956b) described similarly infectious TMV-RNA preparations they had obtained by extracting the virus with phenol, according to the method of Schuster et al. (1956) (see The Phenol Method in the preparations section). The infectivity of the Gierer-Schramm preparations was also sensitive to ribonuclease; in addition it was shown to sediment much more slowly than virus, to be relatively unaffected by anti-TMV serum, and to be considerably more sensitive to elevated temperature than intact virus. Also, the level of protein in the infectious RNA preparations was found to be very low. These points were confirmed by Fraenkel-Conrat et al. (1957).

The indication from these pioneer experiments that viral nucleic acid is the genetic material of viruses has been repeatedly verified in many ways and is now taken for granted. Thus, in viruses, as well as in higher organisms such as protists, plants, and animals, a major function of nucleic acid is as a repository of genetic information. In addition, and in contrast with higher organisms, viral nucleic acid, when it is RNA, often acts as its own messenger RNA (mRNA). When viral nucleic acid is DNA, it cannot perform this function directly but serves as a template from which mRNA is transcribed. The nucleic acids found in mature virus particles, whether RNA or DNA, serve as templates for their own replication.

C. Lipids

Viruses containing lipid include representatives of all major types of viruses (see Table 26), although lipid components are much commoner among animal viruses than they are with bacterial or plant viruses.

Lipid-containing viruses share three common properties:

1. Particle morphology. The virus particles usually exhibit a nucleoprotein core surrounded by a membranous envelope composed of lipid, protein, and sometimes carbohydrate through which glycoprotein structures called spikes project (see Figure 31).

Table 26. Some Lipid-Containing Viruses.

Virus	Percent Lipid	Lipid Constituents	Reference[a]
Avian myeloblastosis	35	Partly phospholipid	1, 2
Equine encephalomyelitis	54	Phospholipids, cholesterol, triglycerides	3
Fowl plague	25	Phospholipid, cholesterol	4, 5
Fowlpox	27	Phospholipid, cholesterol, triglycerides, fatty acids	6
Herpes simplex	22	Phospholipid	7
Influenza	18	Phospholipids, tryglyceride, and cholesterol	8
Potato yellow dwarf	20	Phospholipids, sterol and possibly other lipids	9
Pseudomonas phage Ø6	25	Phospholipid	10
Pseudomonas phage PM2	10	Phospholipid	11
Rous sarcoma	35	Partly phospholipid	12
Simian virus 5 (SV5)	20	Phospholipids, cholesterol, triglyceride	13
Sindbis	29	Phospholipids, cholesterol	14
Tipula iridescent	5	Phospholipid	15
Tomato spotted wilt	19	Not yet known	16
Vaccinia	5	Phospholipid, cholesterol, triglycerides	17

[a](1)Bonar and Beard 1959; (2) Allison and Burke 1962; (3) Beard 1948; (4) Schäfer 1959; (5) Gierer 1957; (6) Randall et al. 1964; (7) Russell et al. 1963; (8) Frommhagen et al. 1959; (9) Ahmed et al. 1964; (10) Vidaver et al. 1973; (11) Espejo and Canelo 1968; (12) Vogt 1965; (13) Klenk and Choppin 1970; (14) Pfefferkorn and Hunter 1963; (15) Thomas 1961; (16) Best 1968; (17) Joklik 1966.

2. *Mechanism of maturation and release of virus particles.* Nascent virus particles mature at plasma, vesicular, or nuclear membranes through which they are then released by an extrusion or budding process.

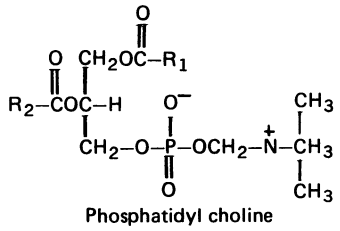
3. *Sensitivity to lipid-degrading agents.* Most lipid-containing viruses disintegrate and lose infectivity upon treatment with organic solvents (for example, ether or methanol-chloroform), certain detergents (for example, deoxycholate), or lipolytic enzymes (for example, phospholipase A). Vaccinia virus and certain iridescent insect viruses, which contain small amounts of lipid, are exceptions to this rule, but logical exceptions because their lipids serve little or no structural function, nor do they play a vital role in the infectious process.

Several different kinds of lipids have been identified among the fatty substances extracted from viruses and include cholesterol, triglycerides ("neutral fat"), and such phospholipids as sphingomyelin, phosphatidylcholine, phosphatidylserine, phosphatidylethanolamine, and phosphatidylinositol. The structures of some of these compounds are illustrated in Figure 29.

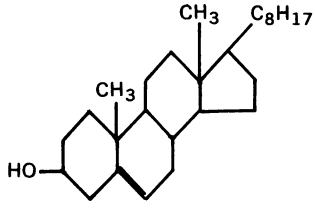
As noted above, most lipid-containing viruses have envelope structures that are acquired upon budding from a membrane. Abundant analytical evidence supports the assumption that much if not all of the viral lipid is obtained directly from the cell membrane in the budding process (Wecker 1957; Frommhagen et al. 1959; Kates et al. 1961; Franklin 1962; Klenk and Choppin 1969b, 1970). The chemical relationship between viral envelope lipids and cell membrane lipids is especially well illustrated by the studies of Klenk and Choppin (1969, 1970) with the paramyxovirus, simian virus 5 (SV5), and the membranes of different cells in which this virus was cultured. The data reproduced in Table 27 show how the quantities of different types of SV5 lipids parallel those of the membranes of two types of kidney cells in which the virus was grown.

Phospholipids predominate among the lipids found in cell membranes and this is reflected in the composition of viral lipids, including those of SV5 (Table 27). The comparison between cell membrane and viral lipids is sharpened by comparing the contents of individual phospholipids as is done for SV5 in Table 28. As shown in the table, there are marked differences in content of individual phospholipids between membranes from mouse kidney and hamster kidney cells. These differences are reflected in the compositions of SV5 lipids from virus grown in the two types of cells. Data of this sort mean that the same virus grown in different cell strains can have lipid of diverse compositions.

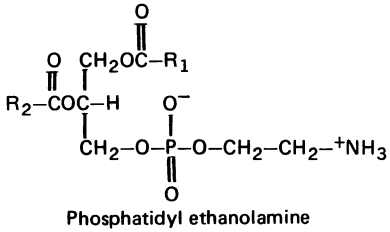
How are diverse lipids fabricated into viral envelopes? This is not yet clear. Cholesterol may be dissolved in the other fatty substances, but the various phospholipids appear to be coupled to protein and polysaccharide to form specific lipoprotein and glycolipid complexes. The precise nature of the linkages involved in these complexes and in their fabrication into an



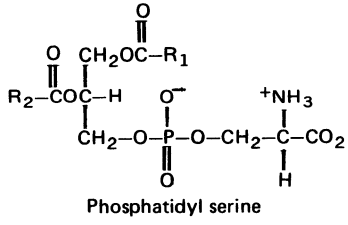
Phosphatidyl choline



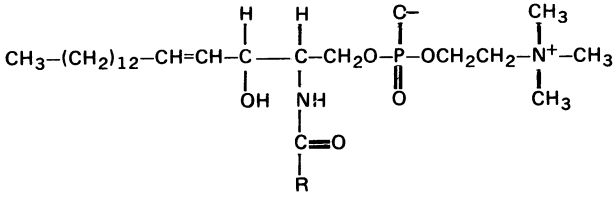
Cholesterol



Phosphatidyl ethanolamine



Phosphatidyl serine



Sphingomyelin

Fig. 29. Structural formulas for some phospholipids (phosphatides) and cholesterol. In these formulas R_1 is typically the hydrocarbon chain of a saturated fatty acid, while R_2 is a similar chain for an unsaturated fatty acid. Usually these fatty acids contain 16 or 18 carbon atoms.

Table 27. Lipid Content of Simian Virus 5 (SV5) and of Plasma Membranes from Monkey Kidney (MK) and Baby Hamster Kidney (BHK21-F) Cells in Which the Virus Was Grown [a]

Source	Total Lipid Percent of Dry Weight	Percent of Total Lipid			
		Phospholipid	Triglycerides	Cholesterol	Cholesterol Esters
MK membranes	28.5	55.0	5.1	23.0	0.3
SV5 from MK cells	20.0	50.9	5.0	29.0	0.9
BHK21-F membranes	30.7	60.0	2.8	20.5	1.3
SV5 from BHK21-F cells	21.0	57.0	3.0	18.6	2.7

[a]From Klenk and Choppin 1969b.

Table 28. A Comparison of the Phospholipid Contents of Simian Virus 5 (SV5) with Those of Plasma Membranes of Monkey Kidney (MK) and Baby Hamster Kidney (BHK21-F) Cells.[a]

Source	Percent of Total Phospholipid				
	Sphingomyelin	Phosphatidyl Choline	Phosphatidyl Inositol	Phosphatidyl Serine	Phosphatidyl Ethanolamine
MK membranes	11.8	32.1	—[b]	17.2	38.8
SV5 from MK cells	12.2	25.2	2.9	17.9	40.3
BHK21-F membranes	24.2	49.5	10.0	5.1	11.2
SV5 from BHK21-F cells	30.0	38.5	10.5	5.2	15.6

[a]From Klenk and Chopin 1969b.
[b]None detected.

envelope remains to be elucidated, but probably involves some of the same sorts of interactions that characterize enzyme-substrate complexes. In short, there is little evidence for primary covalent linkages between protein and lipid moieties in lipoproteins and the combination seems to depend on steric fit and upon interactions between nonpolar hydrophobic residues and between polar or charged groups. In addition there is probably significant hydrogen bonding in which water molecules have a bridging function.

1. *Preparation of Viral Lipids*

Lipids tend to be less soluble in aqueous media than the other constituents of viruses and more soluble in organic solvents. Generally, they are also more susceptible to air oxidation and to temperature effects. Consequently, lipids are extracted from frozen-dried (lyophilized) virus samples with organic solvents at moderate temperatures and often in an atmosphere of nitrogen. It should be noted that some lipolytic enzymes are solvent activated, an effect that increases with temperature. This suggests that extraction at room temperature is generally desirable.

Since in many cases lipids appear to occur in lipoprotein complexes and water plays some part in this union, it appears that dehydrating organic solvent should help to rupture the lipid-protein linkage. Hence such polar solvents as methanol and ethanol are usually included in the initial solvent of a several step procedure. However, since many lipids are not very soluble in such solvents, a more nonpolar solvent such as chloroform or diethyl ether is often included. A commonly employed solvent system of this sort is chloroform-methanol, approximately 2:1 (v/v). An example using such a mixture in the extraction of lipid from SV5 is as follows (Klenk and Choppin 1969a).

In order to extract total lipid, lyophilized virus is extracted with chloroform-methanol-water (65:25:5) (10 ml/50 mg dry weight) twice for 20 min at room temperature and once for 20 min under nitrogen using boiling solvent. To the combined extracts is added 1/6 vol of water, and the mixture is separated into aqueous and organic phases. If gangliosides (complex lipids composed of sphingosine, fatty acid, one or more sugars, and neuraminic acid) are present, they go into the aqueous phase while all other lipids remain in the organic phase. The latter includes virtually all of the SV5 lipid and the solvent in this fraction is removed under nitrogen in a rotary evaporator to give total lipids.

2. *Analysis of Viral Lipids*

The analysis of lipids obtained in the manner described in the previous section is fairly complex. The techniques employed include column chro-

matography, thin layer chromatography, gas-liquid chromatography, phosphorus analysis, and sometimes infrared spectroscopy. A detailed description of such methodology is given by Kritchevsky and Shapiro (1967); the details of the work on SV5 are described by Klenk and Choppin (1969a). A summary sketch of the SV5 analysis follows.

A sample of the total lipid fraction, extracted from SV5 as described in the previous section, was applied to a silicic acid column. Neutral lipids (a term applied to cholesterol and its esters, free fatty acids, and triglycerides) were separated from phospholipids by elution first with chloroform, which yielded neutral lipids, and then with methanol, which eluted phospholipids. The chloroform eluate was evaporated and the residue dissolved in hexane, which then was applied to a Florisil column (Florisil is a synthetic magnesium silicate). Chromatography on the Florisil column separated components of the neutral lipid, mainly cholesterol and its esters and triglycerides. The triglycerides were identified and quantitated as hydroxamic acids while cholesterol and its esters were also determined colorimetrically by another procedure. The methanol eluate from the silicic acid column was dried and dissolved in 2:1 chloroform-methanol. Aliquots of this solution were used for phosphorus determinations, and the different phospholipids were identified and quantitated by a combination of gas chromatography and quantitative two-dimensional thin layer chromatography. The results of such analyses are summarized in Tables 27 and 28.

3. *Function of Viral Lipids*

The lipids occurring in viral envelopes have been termed "peripheral structural lipids" (Franklin 1962). Extraction of these lipids with organic solvents or detergents or digestion of them with lipases results in considerable degradation of the viral particles. Clearly, such lipid components are essential for maintaining the structure of virus envelopes. The reason for loss of infectivity when viral lipids are removed is doubtless associated with the inability of disrupted virus to attach and penetrate because these steps in infection depend largely on viral surface structures. Specific attachment is especially important in infections by animal and bacterial viruses. How removal of lipid alters the infectivity of plant viruses is not yet apparent.

D. Carbohydrates

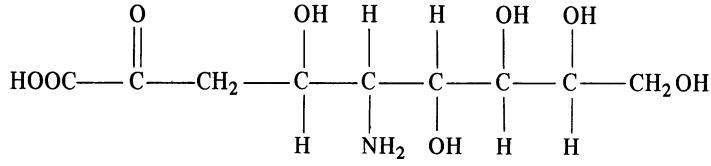
Carbohydrates are found in all viruses since all viruses have a nucleic acid component. Nucleic acids, as indicated in a previous section, contain one of two carbohydrates, ribose or deoxyribose. However, some viruses also contain nonnucleic acid carbohydrate. This has been observed in two

general situations: (1) glucose residues attached to pyrimidine in the DNA of certain bacterial viruses, and (2) polysaccharide coupled with protein (that is, glycoprotein) and lipid (glycolipid) in the envelope structures possessed by some animal and plant viruses. Quite a few enveloped viruses exhibit surface projections called spikes, and these are often glycoprotein in composition, especially in the case of viruses showing the capacity to agglutinate red cells.

Carbohydrate appears in the T-even (T2, T4, T6) bacterial viruses in the form of glucose or gentiobiose in O-glycosidic linkage with the 5-hydroxymethylcytosine (5-HMC) of the viral DNA (Sinsheimer 1960; Lehman and Pratt 1960; Kuno and Lehman 1962). The T-even phages contain essentially the same quantity of 5-HMC in their DNA components; yet the amount of glucoside is distinct for each. Glucose occurs in the proportions of about 0.8, 1.0, and 1.6 moles per mole of HMC for T2, T4 and T6, respectively (Jesaitis 1956; Lichtenstein and Cohen 1960). The glucose is uniformly distributed in the case of T4 where each HMC is glucosylated, but in T2 there is some unsubstituted HMC, some monoglucosylated, and a small amount of diglycosylated-HMC, the latter represented by the glucose disaccharide gentiobiose. In T6, an unsymmetrical distribution of glycosyl units is also found, but here about two-thirds of the glucose is present as gentiobiose.

There is evidence that the degree of glucosylation of the T phage DNA is an inherited trait, although in crosses the trait "glucose content" does not segregate symmetrically as a simple Mendelian character (Sinsheimer 1960). For example, the progeny of a T2 and T4 cross were all found to have the T4 glucose content, whereas the recombinants of a T2 × T6 cross were found to have the glucose content of either T2 or T6, although many more were found to have the glucose content of T2 and the host range of T6 than vice versa.

As indicated above, it has been found that viruses with an envelope structure often contain some of their proteins in the form of glycoproteins. Also, some of the lipid of viral envelopes may be present as glycolipid. The widespread occurrence of such nonnucleic acid carbohydrate constituents among viruses is illustrated by examples given in Table 29. In the table specific viruses are listed but in each case the group of viruses to which the example belongs is also indicated because the characteristics of the example are likely to apply throughout the group, and some of the groups are very large. The carbohydrate components of viral glycoproteins and glycolipids are complex polysaccharides usually fabricated from fucose, galactose, glucosamine, and mannose. Some viral glycoproteins, such as those from Sindbis and vesicular stomatitis viruses, contain sialic acid (Burge and Huang 1970). Sialic acid is the group name for a series of acylated derivatives of neuraminic acid:



Neuraminic Acid

(5-amino-3,5-dideoxy-D-glycero-D-galactononulesonic acid)

The simplest sialic acid is *N*-acetylneuraminic acid.

Although detailed structural analyses are yet to be made of viral glycoproteins, it seems likely, by analogy with other better studied glycoproteins, that the mode of linkage between protein and carbohydrate is by glycosidic bonds between carbohydrate chains and asparagine, serine, and threonine residues of the protein (Neuberger et al. 1972).

1. Preparation of Viral Carbohydrates

General methods have not been developed for the isolation of viral carbohydrates as they have for viral nucleic acids and proteins. Instead, it

Table 29. Some Viruses Containing Nonnucleic Acid Carbohydrate.

Virus	Carbohydrate-containing Constituent		Reference[a]
	Glycolipid	Glycoprotein	
Herpes simplex virus (a herpesvirus)	?	+	1
Influenza virus (an orthomyxovirus)	+	+	2
Murine leukemia virus (an oncornavirus)	?	+	3
OC 43 (a human coronavirus)	?	+	4
Potato yellow dwarf virus (a plant rhabdovirus)	?	+	5
Simian virus 5 (SV5) (a paramyxovirus)	+	+	2
Sindbis virus (a togavirus)	−	+	6
Vesicular stomatitis virus (an animal rhabdovirus)	+	+	7

[a](1) Roizman and Spear 1971; (2) Compans and Choppin 1971; (3) Nowinski et al. 1972; (4) Hierholzer et al. 1972; (5) Knudson and MacLeod 1972; (6) Schlesinger and Schlesinger 1972; (7) Knudson 1973.

has usually seemed sufficient to obtain qualitative and quantitative values for these constituents, and even such analyses have often been neglected. However, a growing realization of the importance of glycoproteins in many animal viruses may lead before long to the development of general procedures for isolating viral carbohydrates in undegraded forms.

The investigation of avian tumor virus glycopeptides represents a step in this direction. For example, Lai and Duesberg (1972) used the following technique to isolate avian tumor virus glycopeptides which were estimated to contain less than 10 percent protein: Purified tumor virus was disrupted by treating it at 37° for 30 min with 1 percent sodium dodecyl sulfate in the presence of 0.05 M mercaptoethanol. A precipitate of proteins and glycoproteins from the disaggregated virus was obtained by addition of 5 vol of ethanol; this precipitate was subsequently dissolved in a solution of 0.1 percent SDS–0.1 M tris buffer at pH 8. Alternatively, the viral proteins and glycoproteins were extracted from the virus by treatment with water-saturated phenol; they were recovered from the phenol phase by precipitation with 5 vol of ethanol in the presence of 2 M ammonium acetate. After washing twice with 75 percent ethanol, the precipitate was dissolved in the tris buffer–0.1 percent SDS noted above. By digesting the proteins in this mixture with pronase (a 48-hr treatment with pronase at 1 mg/ml followed by a second 48-hr treatment with pronase at 0.5 mg/ml) the proteins, including those of the glycoproteins, were largely degraded. The glycopeptides could be separated from this mixture by gel filtration chromatography on Sephadex G-50. When the viral protein had been labeled with ^{3}H amino acids before applying the procedures outlined above, it was found that over 90 percent of the label eluted from the Sephadex as amino acids or small peptides rather than in the glycopeptide fraction. This suggests that the procedure just described could be employed to prepare viral polysaccharides containing only a small amount of protein still attached.

2. *Analysis of Viral Carbohydrates*

The results of chemical and spectrophotometric analyses made by Taylor (1944) on influenza viruses provided the first indication that any highly purified virus contained nonnucleic acid carbohydrate. These early assays also indicated that the viral carbohydrate might contain galactose, mannose, and glucose. Later, more extensive studies based on colorimetric, chromatographic, and spectrophotometric analyses indicated that galactose, mannose, glucosamine, and fucose are constituents of the influenzal carbohydrates (Knight 1947b; Ada and Gottschalk 1956; Frommhagen et al. 1959). Since then, sialic acid and galactosamine have been added to the list of viral polysaccharide constituents (Strauss et al. 1970; McSharry and Wagner 1971).

The total carbohydrate of a virus can be estimated colorimetrically by the orcinol reaction as follows [based on Marshall and Neuberger (1972)]:

1. Dissolve 3–4 mg of dry virus in 1 ml of 0.1 N NaOH in a 10-ml glass-stoppered bylinder or test tube.
2. Add 8.5 ml of orcinol-H_2SO_4 reagent (a fresh mixture of 7.5 vol of 60 percent H_2SO_4 and 1 vol of 1.6 percent orcinol in H_2O) and mix well.
3. Place the loosely stoppered cylinder together with cylinders containing reagent and 1 ml of 0.1 N NaOH and other cylinders containing various total amounts (from 50 to 200 μg) of carbohydrates (standard solution containing equal amounts of fucose, galactose, and mannose) in a water bath at 80°C.
4. After 15 min, cool the tubes in tap water and take readings in a spectrophotometer at 505 nm.

This test will give only a crude approximation owing to small but variable contributions to the color by other constituents of the virus and by a failure to get an appropriate color yield from amino sugars. The best estimate of total nonnucleic acid carbohydrate is obtained by the summation of analyses for the individual carbohydrates.

The analysis of individual carbohydrates in viral glycoproteins and glycolipids involves a variety of methods ranging from colorimetric analyses through chromatography (paper, thin layer, column, and gas-liquid) (see Marshall and Neuberger 1972; Clamp et al. 1972). In any case, the analyses must be preceded by or include in them hydrolysis of the polysaccharides. Hydrolysis of carbohydrates has many of the features and cautions of protein hydrolysis, and these must be taken into consideration if incomplete hydrolysis is to be avoided, on one hand, and destructive hydrolysis is to be minimized, on the other. Such problems are thoroughly discussed by Marshall and Neuberger (1972).

Interference in the carbohydrate analysis can be reduced if glycopeptides are first isolated from the virus by gel electrophoresis or chromatography. Seven different carbohydrates (fucose, mannose, galactose, glucose, galactosamine, glucosamine, and neuraminic acid) were identified by gas-liquid chromatography as constituents of the nonnucleic acid carbohydrate of vesicular stomatitis virus (McSharry and Wagner 1971). Gas-liquid chromatography has attracted considerable interest for analyzing viral carbohydrates but there are technical difficulties in the method. One of them is in getting quantitative derivitization of the individual carbohydrates and another is in eliminating or reducing spurious peaks (background). Balanced against these problems are the exquisite sensitivity of the procedure (in the nanomole region) and the added specificity that can be achieved if the chromatography is coupled to a mass spectrometer or to a counter (gas-liquid radiochromatography).

3. *Function of Viral Carbohydrates*

Glucosylation of T-even phage DNAs appears to be essential for phage survival in certain bacterial strains where the glucosyl residues appear to confer resistance to degradation by nucleases. This resistance mechanism seems to be peculiar to the T-even phages and certain *E. coli* cells since other nonglucosylated phages multiply and produce infectious progeny in these bacteria.

The specific function of carbohydrates in the glycoproteins and glycolipids of enveloped viruses is not known. However, the spikes of ortho- and paramyxoviruses are glycoproteins and these are of two types, one of which constitutes the hemagglutinin of these viruses and the other a neuraminidase enzyme (see Compans and Choppin 1971). The potential importance of the carbohydrate moiety in hemagglutination is indicated by the loss of hemagglutination capacity concomitant with the cleavage of reducing sugar by a specific glycosidase enzyme (Bikel and Knight 1972). It can also be surmised because of their location in the surface of virus particles that the carbohydrates of viral envelopes play a role in the attachment and penetration of these viruses in the course of infecting cells and probably also in their exit from cells. It is not yet clear with regard to release of enveloped viruses from infected cells to what extent the viral envelope is determined by host genome and viral genome, respectively. An interesting aspect of this question is the observation that with avian tumor viruses, the glycopeptides of all viruses released from transformed cells are larger than those of viruses released from normal cells (Lai and Duesberg 1972).

Glycoproteins also are heavily involved in the immunological reactions of enveloped animal viruses. For example, antiserum to influenza hemagglutinin has potent virus-neutralizing capacity (Schild 1970), and the glycoprotein of vesicular stomatitis virus appears to be the specific antigen that induces the synthesis of and reacts with viral neutralizing antibody (Kelley et al. 1972).

E. Polyamines and Metals

In addition to protein, nucleic acid, lipid, and carbohydrate, some other substances are found in small amounts in highly purified preparations of certain viruses. Most of these minor components are probably adventitious elements. For example, many cells contain significant amounts of polyamines (Tabor et al. 1961; Cohen 1971), and these cations are strongly attracted to the phosphoryl anions of viral nucleic acids, where they remain to become a part of the mature virus particle in those cases in which low particle permeability and other relationships are favorable. Thus, putrescine, $H_2N\text{-}CH_2\text{-}CH_2\text{-}CH_2\text{-}CH_2\text{-}NH_2$, and spermidine, $H_2N(CH_2)_4NH$-

$(CH_2)_3NH_2$, were found by Ames and associates (1958, 1960) in T2 and T4 phages in amounts sufficient to neutralize about half of the DNA charge. A similar situation was reported by Kay (1959) for bacteriophage 3 of *E. coli* 518.

In the case of the T2 and T4 phages, which are not normally very permeable to cations, it was shown that the putrescine is associated with the DNA inside the phage head and that this internal putrescine could not be displaced with ^{14}C-containing putrescine on the outside nor by Mg^{++} and Ca^{++}. These closely adhering polyamines are thought to be the same as Hershey's A substances (1957), which are injected along with DNA in the course of infection by T2 and T4. The lack of specificity of putrescine and spermidine was shown by growing the host cells in a medium rich in spermine, $H_2N(CH_2)_3NH(CH_2)_4NH(CH_2)_3NH_2$, a polyamine not normally present in these bacteria. The mature phages isolated from these cells were found to contain spermine rather than putrescine or spermidine (Ames and Dubin 1960).

Hence the polyamines appear to play no specific role in phages, a view which is supported by the absence of polyamines in T3, T5, and P22 phages whose permeabilities to cations presumably allow displacement of polyamines by metallic cations during isolation and purification of the phages. Furthermore, it was shown that a permeable mutant of T4 could be isolated with or without spermidine with no change in biological properties. Nevertheless, it is possible that polyamines or metallic cations may assist in the folding of DNA in the process of phage assembly. Moreover, these cations may be essential to some other stage or stages of phage biosynthesis although the mechanism of such effects is presently obscure (Cohen and Dion 1971).

Polyamines occur in the virions of herpesvirus and of influenza and Newcastle disease viruses (Gibson and Roizman 1971; Bachrach et al. 1974). Traces have also been reported in several plant viruses while amounts sufficient to neutralize about a fifth of the charges of the viral RNA have been found in turnip yellow mosaic, turnip crinkle, and broad bean mottle viruses (Ames and Dubin 1960; Johnson and Markham 1962; Beer and Kosuge 1970).

As many as 14 metallic cations have been found in plant virus preparations, some of them loosely bound to protein and others more tightly bound to RNA (Pirie 1945; Loring et al. 1958; Wacker et al. 1963; Johnson 1964). These cations can be largely removed by treatment with a chelating agent without significantly reducing infectivity. It is doubtful if the remaining few atoms of tightly bound metal are of crucial importance.

The general conclusion about both organic and inorganic cations is that they bind randomly to protein and nucleic acid in amounts dependent on the environment and relative affinities of the ions involved. Such binding, especially to the nucleic acid, may well affect the conformation and function, but specific effects have yet to be elucidated.

F. Summary: Composition of Viruses

There are many viruses in nature whose mature particles consist solely of nucleic acid and protein. There are numerous other more complex viruses that contain, in addition to nucleic acid and protein, lipid, nonnucleic acid carbohydrate, and a variety of other minor constituents. Nucleic acid and protein are properly emphasized because these constituents play a predominant role in the structure and function of viruses, although some of the minor constituents may in specific cases, as indicated in the preceding sections, be very important. Finally, nucleic acid is recognized as the one indispensable constituent of all viruses (some viruses may consist of nucleic acid alone) because it is the genetic material and is capable of inducing infection by itself.