Comprehensive Virology

Edited by Heinz Fraenkel-Conrat
University of California, Berkeley

and Robert R. Wagner
University of Virginia, Charlottesville

Volume 13

Structure and Assembly

Primary, Secondary, Tertiary, and Quaternary Structures

Virology

13

Structure and Assembly

Primary, Secondary, Tertiary, and Quaternary Structures

PLENUM PRESS • NEW YORK AND LONDON

Foreword

The time seems ripe for a critical compendium of that segment of the biological universe we call viruses. Virology, as a science, having passed only recently through its descriptive phase of naming and numbering, has probably reached that stage at which relatively few new—truly new—viruses will be discovered. Triggered by the intellectual probes and techniques of molecular biology, genetics, biochemical cytology, and high resolution microscopy and spectroscopy, the field has experienced a genuine information explosion.

Few serious attempts have been made to chronicle these events. This comprehensive series, which will comprise some 6000 pages in a total of about 18 volumes, represents a commitment by a large group of active investigators to analyze, digest, and expostulate on the great mass of data relating to viruses, much of which is now amorphous and disjointed, and scattered throughout a wide literature. In this way, we hope to place the entire field in perspective, and to develop an invaluable reference and sourcebook for researchers and students at all levels.

This series is designed as a continuum that can be entered anywhere, but which also provides a logical progression of developing facts and integrated concepts.

Volume 1 contains an alphabetical catalogue of almost all viruses of vertebrates, insects, plants, and protists, describing them in general terms. Volumes 2–4 deal primarily, but not exclusively, with the processes of infection and reproduction of the major groups of viruses in their hosts. Volume 2 deals with the simple RNA viruses of bacteria, plants, and animals; the togaviruses (formerly called arboviruses), which share with these only the feature that the virion's RNA is able to act as messenger RNA in the host cell; and the reoviruses of animals and plants, which all share several structurally singular features, the

most important being the double-strandedness of their multiple RNA molecules.

Volume 3 addresses itself to the reproduction of all DNA-containing viruses of vertebrates, encompassing the smallest and the largest viruses known. The reproduction of the larger and more complex RNA viruses is the subject matter of Volume 4. These viruses share the property of being enclosed in lipoprotein membranes, as do the togaviruses included in Volume 2. They share as a group, along with the reoviruses, the presence of polymerase enzymes in their virions to satisfy the need for their RNA to become transcribed before it can serve messenger functions.

Volumes 5 and 6 represent the first in a series that focuses primarily on the structure and assembly of virus particles. Volume 5 is devoted to general structural principles involving the relationship and specificity of interaction of viral capsid proteins and their nucleic acids, or host nucleic acids. It deals primarily with helical and the simpler isometric viruses, as well as with the relationship of nucleic acid to protein shell in the T-even phages. Volume 6 is concerned with the structure of the picornaviruses, and with the reconstitution of plant and bacterial RNA viruses.

Volumes 7 and 8 deal with the DNA bacteriophages. Volume 7 concludes the series of volumes on the reproduction of viruses (Volumes 2–4 and Volume 7) and deals particularly with the single- and double-stranded virulent bacteriophages.

Volume 8, the first of the series on regulation and genetics of viruses, covers the biological properties of the lysogenic and defective phages, the phage-satellite system P 2–P 4, and in-depth discussion of the regulatory principles governing the development of selected lytic phages.

Volume 9 provides a truly comprehensive analysis of the genetics of all animal viruses that have been studied to date. These chapters cover the principles and methodology of mutant selection, complementation analysis, gene mapping with restriction endonucleases, etc. Volume 10 also deals with animal cells, covering transcriptional and translational regulation of viral gene expression, defective virions, and integration of tumor virus genomes into host chromosomes.

Volume 11 covers the considerable advances in the molecular understanding of new aspects of virology which have been revealed in recent years through the study of plant viruses. It covers particularly the mode of replication and translation of the multicomponent viruses and others that carry or utilize subdivided genomes; the use of proto-

plasts in such studies is authoritatively reviewed, as well as the nature of viroids, the smallest replicatable pathogens. Volume 12 deals with special groups of viruses of protists and invertebrates which show properties that set them apart from the main virus families. These are the lipid-containing phages and the viruses of algae, fungi, and invertebrates. These groups will be followed in Volume 14 by special and/or newly characterized vertebrate virus groups (e.g., arena-, corona-, hepatitis, calici-, and bunyaviruses).

The present volume collects chapters on various topics related to the structure and assembly of viruses, dealing in detail with nucleotide and amino acid sequences, as well as with particle morphology and assembly, and the structure of virus membranes and hybrid viruses. The first complete sequence of a viral RNA is represented as a multicolored foldout.

Several subsequent volumes will deal with virus–host relationships and with methodological aspects of virus research.

Contents

Chapter 3

Structure and Function of RNA Bacteriophages

Walter Fiers

Chapter 4

DNA Sequencing of Viral Genomes

Gillian M. Air

Chapter 5

Viral Membranes

Richard W. Compans and Hans-Dieter Klenk

Chapter 6

Adenovirus Structural Proteins

Harold S. Ginsberg

Chapter 7

The Adenovirus-SV40 Hybrid Viruses

Cephas T. Patch, Arthur S. Levine, and Andrew M. Lewis, Jr.

Chapter 8

Bacteriophage Structure

Frederick A. Eiserling

Chapter 9

Genetic Control of Complex Bacteriophage Assembly

William B. Wood and Jonathan King

CHAPTER 1

Amino Acid Sequences of Plant and Animal Viral Proteins

Stephen Oroszlan and Raymond V. Gilden

Biological Carcinogenesis Program
Frederick Cancer Research Center
Frederick, Maryland 21701

1. INTRODUCTION

This chapter will present currently available primary structure data on the proteins of a variety of plant and animal viruses. Excellent descriptions of methods for the determination of amino acid sequences of proteins are found in several review articles and books (Hirs, 1967; Schroeder, 1968; Blackburn, 1970; Needleman, 1970; Starbuck, 1970; Hirs and Timasheff, 1972; Niall, 1973; Perham, 1975) and therefore will not be presented here. Since the pioneering studies on TMV in the early 1960s which contributed to establishing the principle of colinearity of gene and protein, sequencing efforts have contributed to an understanding of evolutionary relationships among other plant and animal viruses. Such data can also provide practical end points such as the design of synthetic vaccines; this may be especially crucial in cases where complete freedom from viral nucleic acids is essential. The continuing development of sensitive automated microsequencing procedures utilizing stepwise Edman degradation, high-sensitivity methods for identification of amino acids, radiolabeled reagents, and proteins with intrinsic radioactivity (Edman and Begg, 1967; Jacobs *et*

al., 1974; McKean *et al.*, 1974; Jacobs and Niall, 1975; Brauer *et al.*, 1975; Oroszlan *et al.*, 1975*a*; Silver and Hood, 1975; Ballou *et al.*, 1976; Bridgen, 1976; Henning *et al.*, 1976; Vitetta *et al.*, 1976; Zimmerman *et al.*, 1976) will allow data gathering on minor (by mass) virion polypeptides, for example, the reverse transcriptases of retroviruses, which until recently would have required the availability of enormous amounts of virus. The concluding section on retrovirus proteins, the major interest of the authors' laboratory, will summarize sequence data and recent evidence of viral groupings based on immunological data.

2. PLANT VIRUSES

2.1. Tobacco Mosaic Virus

Tobacco mosaic virus (TMV), which was the first virus to be identified and crystallized, is a 300-nm-long, rodlike particle with a diameter of 15 nm and an approximate particle weight of 4×10^7. The virion is composed of 5% RNA (single-stranded, single chain) and 95% protein. The protein component representing the viral coat consists of 2130 identical polypeptide subunits, each with an approximate molecular weight of 17,500, consisting of 158 amino acid residues (Fraenkel-Conrat, 1968; Dayhoff, 1972; Benjamini *et al.*, 1972; Smith, 1977). The coat protein of TMV vulgare (common strain) was the first viral protein whose complete amino acid sequence was determined (Tsugita *et al.*, 1960). Now protein sequences of several naturally occurring strains as well as artificially induced mutants are known.

The complete amino acid sequences of six TMV strains—vulgare (V), OM, dahlemense (D), cowpea (CP), U2, and Holmes ribgrass (HR)—are aligned in Fig. 1.

The N termini (serine and alanine) are acetylated in all strains except in U2, which has proline at the N terminus. As pointed out by Fraenkel-Conrat (1968), it appears probable that the acetyl group fulfills a blocking function, protecting the peptide chain against aminopeptidases, and that this function proves redundant in a mutant carrying an N-terminal proline which is not susceptible to cleavage by the known exopeptidases (Hill, 1965).

Those amino acids which are positionally different from the sequence of the common strain (V) protein are underlined in Fig. 1. In order to maintain positional homology (best fit) in the sequence of CP strain protein, which is three residues longer, an insert (glutamic acid)

is placed between residues 64 and 65 (Rees and Short, 1975). The HR strain is two residues shorter than the V protein and appears to have a gap consisting of residues 148 and 149. The OM strain has only three amino acid substitutions: in positions 50 (Glu→Gln), 129 (Ile→Val), and 153 (Thr→Asn). The single cysteine appears in position 27 in all strains except CP, where it is substituted by leucine. In general, most of the observed exchanges involve only the smaller and hydrophilic residues, although there are a large number of exceptions. The most variable regions of the protein chain seem to be those from residue 19 to 28, from 49 to 68, from 97 to 101, and from 138 to 158. These regions include a high proportion of hydrophylic residues; therefore, they may represent the positions of the folded protein in the intact virus which may be in contact with the solvent (Fraenkel-Conrat, 1968; Durham and Butler, 1975).

The segments from residue 87 to 94 and from 113 to 122 are identical in all strains except CP. These two regions were thought to be completely conserved sequences until the sequence of the CP strain protein became available just 4 years ago (Rees and Short, 1975). It has been assumed that these regions compose the RNA binding site. They contain four of the six conserved positive charges in the protein (arginine residues 90, 92, 113, 122) which could neutralize the negative charges of the RNA chain–phosphate backbone. The CP strain also has arginine in all the above positions except position 122, where it is substituted by histidine. Apart from the above regions only residues 2, 4, 17, 18, 36–38, 41, 61, 63, 82–83, 88–92, 94, 128, 132, 137, 140, 144–145, 150, and 156 are unchanged in all the six strains for which the complete sequences are given in Fig. 1.

The HR and CP strains appear to be evolutionarily the most distantly related to the other more closely related strains (Dayhoff, 1972). The HR protein has 80 amino acid changes from the common V strain of TMV and CP coat protein has a total of 98 changes (Rees and Short, 1975). The percent differences between the sequences of the above six TMV strains are shown in Table 1. These correspond very well with the previous ordering of TMV strains into groups based on amino acid compositional differences (Fraenkel-Conrat, 1974). The results of nucleic acid hybridization experiments, however, did not reveal a quantitative relationship similar to that shown by amino acid sequence data. Even those strains among the various classes which differ least in primary structure of coat protein (vulgare and dahlemense, representing classes A and B, respectively) were found to show no apparent nucleotide sequence homology. A possible explanation is that