

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 1

Amino Acid Sequences of Plant and Animal Viral Proteins

Stephen Oroszlan and Raymond V. Gilden

| | |
|---------------------------------------|----|
| 1. Introduction | 1 |
| 2. Plant Viruses | 2 |
| 2.1. Tobacco Mosaic Virus | 2 |
| 2.2. Alfalfa Mosaic Virus | 8 |
| 2.3. Bromoviruses | 9 |
| 2.4. Turnip Yellow Mosaic Virus | 11 |
| 3. Animal Viruses | 11 |
| 3.1. Influenza Virus | 11 |
| 3.2. Picornaviruses | 14 |
| 3.3. Reovirus | 16 |
| 3.4. Nuclear Polyhedrosis Virus | 16 |
| 3.5. Hepatitis B Virus | 17 |
| 3.6. Papovaviruses | 18 |
| 3.7. Adenovirus | 18 |
| 3.8. Retroviruses | 19 |
| 4. References | 29 |

Chapter 2

Structure of the RNA of Eukaryotic Viruses

H. Fraenkel-Conrat

| | |
|--|----|
| 1. Introduction | 37 |
| 2. General Features of RNA Sequencing | 38 |
| 2.1. End-Group Analysis and Stepwise Degradation of RNA | 38 |

| | | |
|------|--|----|
| 2.2. | Internal Sequences of Viral RNAs and Structural Considerations | 42 |
| 3. | Elucidation of Plant Virus RNA Sequences | 46 |
| 3.1. | Bromovirus and Cucumovirus Group | 46 |
| 3.2. | Alfalfa Mosaic Virus | 47 |
| 3.3. | Tobacco Mosaic Virus | 47 |
| 3.4. | Tymoviruses | 51 |
| 3.5. | Viroids | 52 |
| 3.6. | Other Plant Virus RNA Sequences | 53 |
| 4. | Elucidation of Animal Virus RNA Sequences | 54 |
| 4.1. | Picornaviridae | 54 |
| 4.2. | Toga- and Coronaviridae | 55 |
| 4.3. | Minus-Strand RNA Viruses | 56 |
| 4.4. | Double-Stranded Viruses | 57 |
| 4.5. | Retroviridae | 57 |
| 5. | References | 60 |

Chapter 3

Structure and Function of RNA Bacteriophages

Walter Fiers

| | | |
|------|--|-----|
| 1. | Introduction | 69 |
| 2. | Classification of RNA Phages | 70 |
| 2.1. | RNA Coliphages | 70 |
| 2.2. | RNA Bacteriophages of Other Genera | 75 |
| 3. | Architecture of the Virion | 79 |
| 3.1. | General Properties | 79 |
| 3.2. | Coat Protein | 84 |
| 3.3. | A Protein | 88 |
| 3.4. | Viral RNA | 94 |
| 4. | Structure of the Viral RNA | 99 |
| 4.1. | Terminal Sequences | 99 |
| 4.2. | Internal Oligonucleotides | 106 |
| 4.3. | Specifically Protected Regions | 111 |
| 4.4. | Approaches Leading to Partial Sequences of R17 and f2 RNA and to the Total Sequence of MS2 RNA | 129 |
| 4.5. | Regions of Q β RNA Sequenced by Partial Replication | 148 |

| | |
|--|-----|
| 4.6. Nucleotide Sequence of Small Q β Replicase Template RNAs | 149 |
| 4.7. 3'-Terminal Elongation, Reverse Transcription, and Cloning of Phage RNA | 151 |
| 5. Biological Functions of the Bacteriophage Genome | 153 |
| 5.1. The Untranslated Regions | 153 |
| 5.2. Viral Proteins | 155 |
| 5.3. Use of the Genetic Code | 157 |
| 5.4. Regulatory Systems Controlling the Expression of the Viral Genes | 161 |
| 5.5. Termination Regions | 168 |
| 5.6. RNA Phage Mutants | 169 |
| 5.7. General Properties of the Viral Nucleotide Sequence | 174 |
| 6. Concluding Remarks | 178 |
| 7. Literature | 179 |
| 7.1. Reviews | 179 |
| 7.2. Specific References | 179 |

Chapter 4

DNA Sequencing of Viral Genomes

Gillian M. Air

| | |
|--|-----|
| 1. Introduction | 205 |
| 1.1. The Sanger and Coulson Sequencing Method | 206 |
| 1.2. The Maxam and Gilbert Sequencing Method | 208 |
| 1.3. Chain-Termination Sequencing Method | 212 |
| 1.4. Other Methods of Sequencing DNA | 214 |
| 2. DNA Sequences from Bacteriophage Genomes | 219 |
| 2.1. The Bacteriophage λ | 219 |
| 2.2. Bacteriophages P2, 299, and 186 | 227 |
| 2.3. Bacteriophage T7 | 228 |
| 2.4. Bacteriophage ϕ X174 | 229 |
| 2.5. Bacteriophage G4 | 242 |
| 2.6. Bacteriophage S13 | 242 |
| 2.7. Bacteriophages f1, fd, M13, and ZJ-2 | 243 |
| 3. DNA Sequences from Animal Viruses | 253 |
| 3.1. Simian Virus 40 | 253 |
| 3.2. Polyoma Virus | 264 |

| | |
|-----------------------------------|-----|
| 3.3. Adenoviruses | 265 |
| 3.4. Adeno-Associated Virus | 272 |
| 4. General Conclusions | 272 |
| 5. References | 274 |

Chapter 5

Viral Membranes

Richard W. Compans and Hans-Dieter Klenk

| | |
|--|-----|
| 1. Introduction | 293 |
| 2. Methods for Analysis of Viral Membranes | 294 |
| 2.1. Proteins | 294 |
| 2.2. Lipid Bilayer | 295 |
| 2.3. Carbohydrates | 297 |
| 3. Viral Membrane Structure and Assembly | 298 |
| 3.1. Arenaviruses | 298 |
| 3.2. Bunyaviruses | 300 |
| 3.3. Coronaviruses | 302 |
| 3.4. Herpesviruses | 304 |
| 3.5. Myxoviruses | 308 |
| 3.6. Paramyxoviruses | 330 |
| 3.7. Poxviruses | 343 |
| 3.8. Rhabdoviruses | 347 |
| 3.9. Retroviruses: C-Type Particles | 356 |
| 3.10. Togaviruses | 364 |
| 4. Conclusions | 372 |
| 4.1. Viral Membrane Structure | 372 |
| 4.2. Biogenesis and Assembly of the Envelope | 374 |
| 5. References | 377 |

Chapter 6

Adenovirus Structural Proteins

Harold S. Ginsberg

| | |
|-------------------------|-----|
| 1. Introduction | 409 |
| 2. Classification | 410 |

| | | |
|------|---|-----|
| 3. | The Virion | 411 |
| 3.1. | Chemical Composition | 411 |
| 3.2. | Architecture | 413 |
| 4. | Synthesis of Viral Proteins | 417 |
| 5. | Purification of Adenovirus Structural Proteins: A Critique..... | 419 |
| 6. | The Hexon | 421 |
| 6.1. | Morphology | 421 |
| 6.2. | Immunological Characteristics | 423 |
| 6.3. | Chemical Characteristics | 425 |
| 6.4. | Physical Characteristics | 430 |
| 6.5. | Genes Affecting Hexon Structure and Function .. | 431 |
| 6.6. | Model of Hexon Structure | 432 |
| 7. | The Penton | 433 |
| 7.1. | General Description | 433 |
| 7.2. | Penton Base | 434 |
| 7.3. | Fiber | 434 |
| 7.4. | Immunological Characteristics of the Penton and Its Subunits | 437 |
| 7.5. | Biological Functions of the Penton and Its Subunits | 438 |
| 8. | Other Capsid Proteins | 441 |
| 9. | The Core | 445 |
| 10. | An Overview | 446 |
| 11. | References | 448 |

Chapter 7

The Adenovirus-SV40 Hybrid Viruses

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

| | | |
|------|--|-----|
| 1. | Introduction | 459 |
| 2. | SV40 and the Adenoviruses | 461 |
| 2.1. | SV40 | 461 |
| 2.2. | The Adenoviruses | 462 |
| 2.3. | SV40 and Adenovirus Transcription during Lytic Infection | 465 |
| 2.4. | Viral Transcription in Cells Transformed by SV40 or the Adenoviruses | 467 |
| 2.5. | Interactions between SV40 and the Adenoviruses during Lytic Infection | 468 |

| | |
|---|-----|
| 3. SV40 Enhancement of Adenovirus Replication in Monkey Cells | 469 |
| 3.1. Adenovirus DNA and RNA Synthesis in Unenhanced and SV40-Enhanced MKC | 470 |
| 3.2. Adenovirus Protein Synthesis in Unenhanced and SV40-Enhanced MKC | 470 |
| 3.3. Enhancement: An Early SV40 Function | 472 |
| 3.4. Summary of SV40 Enhancement of Adenovirus Replication in MKC | 472 |
| 3.5. SV40 Complementation of Adenovirus 5 Temperature-Sensitive Mutants | 474 |
| 4. Discovery of the Adenovirus 7-SV40 Hybrid Virus | 474 |
| 4.1. The Adenovirus 7-SV40 Population: A Mixture of Nonhybrid Adenovirus 7 and Defective Adenovirus 7-SV40 Hybrid Virus | 475 |
| 4.2. The Physical Structure of the Adenovirus 7-SV40 Hybrid Genome | 477 |
| 4.3. Heteroduplex Mapping of the Adenovirus 7-SV40 Hybrid Genome | 478 |
| 4.4. SV40 RNA Transcription in Cells Infected with Ad7 ⁺ | 480 |
| 4.5. Summary of Structure and Function of the Adenovirus 7-SV40 Hybrid Virus | 480 |
| 5. DNA-DNA Recombination in the Adenoviruses and the Adenovirus-SV40 Hybrid Viruses | 480 |
| 5.1. A Hypothetical Model for the Formation of the Adenovirus 7-SV40 Hybrid Genome | 482 |
| 5.2. Transcapsidation | 482 |
| 6. Adenovirus-SV40 Hybrid Viruses That Yield Infectious SV40 | 484 |
| 6.1. An Ad5 ⁺⁺ Hybrid Virus Population | 485 |
| 6.2. The Adenovirus 2-SV40 Hybrid Virus Population Ad2 ⁺⁺ | 485 |
| 6.3. Plaque Isolation from the Ad2 ⁺⁺ Population | 486 |
| 7. Biological Properties of Ad2 ⁺⁺ HEY and Ad2 ⁺⁺ LEY .. | 489 |
| 7.1. The Physical Structures of the Hybrid Viral Genomes in Ad2 ⁺⁺ HEY and Ad2 ⁺⁺ LEY | 490 |
| 7.2. Heteroduplex Mapping of the Hybrid Viral Genomes in Ad2 ⁺⁺ HEY and Ad2 ⁺⁺ LEY | 491 |
| 7.3. SV40 Transcription in Cells Infected with Ad2 ⁺⁺ HEY and Ad2 ⁺⁺ LEY | 494 |

| | | |
|-------|--|-----|
| 7.4. | Adenovirus 2-SV40 Hybrid Viruses Derived from the Ad2 ⁺ HEY Population | 495 |
| 8. | Isolation of the Nondefective Adenovirus 2-SV40 Hybrid Viruses | 496 |
| 8.1. | Early SV40 RNA in Cells Infected with the Nondefective Adenovirus 2-SV40 Hybrid Viruses | 497 |
| 8.2. | The Physical Structures of the Nondefective Adenovirus 2-SV40 Hybrid Genomes | 498 |
| 8.3. | The Origin of the Nondefective Adenovirus 2-SV40 Hybrid Viruses | 501 |
| 8.4. | SV40 RNA Transcription during Lytic Infection with Nondefective Adenovirus 2-SV40 Hybrid Viruses | 505 |
| 8.5. | Adenovirus 2 Transcription during Lytic Infection with Nondefective Adenovirus 2-SV40 Hybrid Viruses | 506 |
| 8.6. | SV40 Protein Synthesis in Cells Lytically Infected with the Nondefective Adenovirus 2-SV40 Hybrid Viruses | 508 |
| 9. | Host Range Mutants of Ad2 ⁺ ND ₁ | 513 |
| 10. | Revertants of Host Range Mutants of Ad2 ⁺ ND ₁ | 514 |
| 11. | Temperature-Sensitive Mutants of Ad2 ⁺ ND ₁ | 515 |
| 12. | Early SV40 Antigens and Functions Expressed by the Nondefective Adenovirus 2-SV40 Hybrid Viruses .. | 516 |
| 12.1. | Early SV40 Antigens | 517 |
| 12.2. | SV40 Enhancement of Adenovirus Replication in Monkey Kidney Cells | 518 |
| 12.3. | Host DNA Synthesis | 519 |
| 13. | Cell Transformation and Tumor Induction by SV40 and the Adenoviruses | 519 |
| 13.1. | Cell Transformation and Tumor Induction by Ad7 ⁺ | 522 |
| 13.2. | Oncogenicity of Transcapsidated Adenovirus 7-SV40 Hybrid Viruses | 523 |
| 13.3. | Cell Transformation and Tumor Induction by Defective Adenovirus-SV40 Hybrid Viruses That Yield Infectious SV40 | 525 |
| 13.4. | Cell Transformation and Tumor Induction by Adenovirus 2 and the Nondefective Adenovirus 2-SV40 Hybrid Viruses | 526 |
| 14. | References | 530 |

*Chapter 8***Bacteriophage Structure***Frederick A. Eiserling*

| | |
|---|-----|
| 1. Introduction | 543 |
| 1.1. Advantages of Bacteriophages for Structural Studies of Viruses | 543 |
| 1.2. Methods of Structure Analysis | 544 |
| 2. Structure of Bacteriophage Heads and Capsids | 546 |
| 2.1. Isometric Phage and Head Structure: λ , P22, T7, and P2 | 547 |
| 2.2. λ Head Structure | 549 |
| 2.3. P22 Head Structure | 549 |
| 2.4. T7 Head Structure | 551 |
| 2.5. P2 Head Structure | 551 |
| 2.6. Structure of T5 and SP01 Heads | 551 |
| 2.7. Elongated Head Structures: ϕ 29, T4, and CbK ... | 554 |
| 2.8. ϕ 29 Structure | 555 |
| 2.9. T4 Head Structure | 556 |
| 2.10. Phage ϕ CbK Head Structure | 563 |
| 3. Structure of Phage Tails | 565 |
| 3.1. Short Hexagonal Tail Structure | 565 |
| 3.2. Tubular Tail Structure | 567 |
| 3.3. Contractile Tail Structure | 568 |
| 4. Discussion | 571 |
| 4.1. Head Structure and Assembly | 571 |
| 4.2. Tail Structure | 573 |
| 5. Summation | 574 |
| 6. References | 575 |

*Chapter 9***Genetic Control of Complex Bacteriophage Assembly***William B. Wood and Jonathan King*

| | |
|---|-----|
| 1. Introduction | 581 |
| 1.1. Scope and Purposes of This Chapter | 581 |
| 1.2. Approaches to the Problem of Phage Assembly .. | 582 |

| | | |
|-------------|--|-----|
| 2. | Assembly of Phage Heads | 584 |
| 2.1. | General Problems of Head Assembly | 584 |
| 2.2. | λ Head Assembly | 586 |
| 2.3. | P22 Head Assembly | 589 |
| 2.4. | T7, T3, and ϕ 29 Head Assembly | 591 |
| 2.5. | T4 Head Assembly | 592 |
| 3. | Assembly of Phage Tails | 597 |
| 3.1. | General Problems of Tail Assembly | 597 |
| 3.2. | P22 and T7 Tail Assembly | 598 |
| 3.3. | λ Tail Assembly | 599 |
| 3.4. | T4 Tail Assembly | 602 |
| 3.5. | Assembly of Other Phage Tails | 605 |
| 4. | Head-Tail Joining | 605 |
| 5. | Assembly of Phage Tail Fibers | 606 |
| 6. | Role of the Host Cell in Phage Assembly | 609 |
| 7. | Discussion, Conclusions, and Remaining Problems | 610 |
| 7.1. | General Nature of Assembly Pathways' | 610 |
| 7.2. | Sequential Ordering | 612 |
| 7.3. | Nonstructural Accessory Proteins in Assembly ... | 613 |
| 7.4. | Specification of Dimensions | 616 |
| 7.5. | Gene Clustering and Control of Synthetic Rates .. | 618 |
| 7.6. | Lattice Transitions | 619 |
| 7.7. | The Problem of DNA Packaging | 620 |
| 7.8. | Remaining Problems and Future Applications ... | 623 |
| 8. | References | 624 |
| Index | | 635 |

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.*, 1975a; 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 hydrophilic 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 dahlmense, representing classes A and B, respectively) were found to show no apparent nucleotide sequence homology. A possible explanation is that