Cluster Analysis of Viral Proteins

J. H. Tremaine and Edward Argyle

Research Scientist, Canada Department of Agriculture Research Station, Vancouver 8, British Columbia; and Research Scientist, Dominion Radio Astrophysical Observatory, Penticton, British Columbia, respectively. The second author is currently on leave to Department of Computer Science, University of British Columbia.

Contribution No. 162, Research Station, Research Branch, CDA, Vancouver, British Columbia.

Accepted for publication 3 November 1969.

ABSTRACT

The amino acid compositions of the protein coats of 54 plant viruses and strains were compared by a computer program for cluster analysis. Strains of a single virus clustered tightly, showing a close relationship. Rod-shaped and spherical viruses did not have distinctive compositions. The results of the cluster analysis were compared with serological and physical properties of the viruses, and the relevance of cluster analysis to virus classification was discussed. Phytopathology 60:654-659.

An attempt by Tremaine & Goldsack (30) to correlate the amino acid composition of regular viruses with their structure was based on Waugh's (37) concept of protein structure, i.e., a nonpolar inner volume surrounded by a polar outer volume. The parameter used in that analysis, the ratio of the surface area of the polar side chains to the total protein volume, was similar for all viral proteins regardless of shape. Indeed, the parameter has a range of only 0.04 A^{-1} to 0.08 A^{-1} for all proteins. This small variation, the arbitrary assignment of polarity to certain amino acids (11), and the crude model of protein structure, limited the success of this attempt to relate amino acid composition to virus shape.

In the work reported here, a straightforward cluster analysis of viral proteins was carried out in the expectation that the results would bear some relationship to the size and shape of viruses or to their taxonomic classification.

Amino acid composition.—Data for the computer program consisted of the published amino acid compositions of 26 plant viruses and 27 strains (Table 1) and one other virus, TNV, analysed at this laboratory by a procedure described elsewhere (32). In the few cases where data for cysteine and tryptophan were missing, average values for the other proteins were inserted.

Cluster analysis.—In cluster analysis, each of the 54 proteins is represented by the relative molar ratios of its 18 amino acids. Consequently, if the 18 numbers are regarded as coordinate values in a hyperspace, each protein will be represented as a point in an amino acid space of 18 dimensions. It is then possible to compute the Euclidian distance (d) between each pair of points, and then to proceed with a search for clusters. However, such a procedure has some disadvantages. Firstly, since all mole ratios are positive, all proteins are forced into one hyperquadrant of the space. Secondly, if published compositions are used directly, their heterogeneity of scale will lead to large differences in the spatial representation. Thirdly, the correlation coefficient, defined below, is quicker to compute than the distance, and has none of its disadvantages.

A definition of the correlation coefficient that is suitable for computation is

Equation 1

$$r_{ij} = \frac{n \Sigma P_{ik} \cdot P_{jk} - \Sigma P_{ik} \cdot \Sigma P_{jk}}{\sqrt{[n \Sigma P_{ik}^2 - (\Sigma P_{ik})^2]\,[n \Sigma P_{jk}^2 - (\Sigma P_{jk})^2]}}$$

where r_{ij} is the correlation coefficient between the i-th and the j-th proteins, $j > i$; P_{ik} is the k-th coordinate of the i-th protein, $n = 18$; and the summation is over $k = 1$ to 18.

For purposes of illustration, the relationship between r and d is shown in Fig. 1 for a space of two dimensions (i.e., for two amino acids, A_1 and A_2). In this representation, it is convenient to think of each amino acid value also as a component of a protein vector based at the origin of the coordinate system. Proteins such as P_1 and P_2 can then be represented either as the points P_1 and P_2 or as the vectors OP_1 and OP_2. If, further, the vector representation is normalized to unit length, the vectors OQ_1 and OQ_2 are obtained. Finally, if the mean value of the components is made zero (by subtracting the mean from each component), all quadrants of a hypersphere may be occupied by representative vectors.

Both of these normalizations are included in the usual definition of the correlation coefficient, and when r_{ij} is computed according to equation 1, the normalizations fall outside of the computing loop for the onerous first term in the numerator (the only summation containing both i and j). The product $P_{ik} \cdot P_{jk}$ can be computed faster than the corresponding square of a difference, $(P_{ik} - P_{jk})^2$, that would require evaluation during computation of Euclidian distances.

Referring to Fig. 1, d is the Euclidian distance between proteins P_1 and P_2, and s is their separation after normalization. The correlation coefficient can be seen to satisfy the relations $r = \cos\theta$ and $-1 \leqslant r \leqslant 1$ where θ is the angle between the vectors. All the work reported here is described in terms of r, but r may be readily converted to the normalized distance between proteins by the relation

$$s = 2(1 - r)^{\frac{1}{2}}$$

and

$$0 \leqslant s \leqslant 2$$

The values of r for the 54 viruses were stored in a 54×54 correlation matrix. In order to obtain clusters

TABLE 1. Plant viruses and strains[a], abbreviations, and references to amino acid composition

Abbreviation	Common name	Strain		Reference
AMV	Alfalfa mosaic virus	AMV-1	from Canada	(32)
		AMV-T	top component from USA	(15)
		AMV-B	bottom component from USA	(15)
BBMV	Broad bean mottle virus	BBMV-A	from USA	(17)
		BBMV-G	from Germany	(39)
BMV	Bromegrass mosaic virus			(27)
BPMV	Bean pod mottle virus			(22)
BSMV	Barley stripe mosaic virus			(9)
CCMV	Cowpea chlorotic mottle virus			(3)
CMtV	Carnation mottle virus			(30)
CMV	Cucumber mosaic virus			(34)
CNV	Cucumber necrosis virus			(30)
CRSV	Carnation ringspot virus			(13)
EAMV	Echte Acherbohnenmosaik-Virus			(39)
PEMV	Pea enation mosaic virus			(25)
PVS	Potato virus S			(30)
PVX	Potato virus X	PVX-C	from Canada	(24)
		PVX-R	from Canada	(24)
		PVX-A	from USA	(19)
SBMV	Southern bean mosaic virus	SBMV-B1	bean strain	(29)
		SBMV-B2	bean strain	(7)
		SBMV-B3	bean strain	(7)
		SBMV-C1	cowpea strain	(29)
		SBMV-C2	cowpea strain	(7)
SoMV	Sowbane mosaic virus			(12)
STNV	Satellite of tobacco necrosis virus			(21)
SqMV	Squash mosaic virus			(16)
TBSV	Tomato bushy stunt virus			(6)
TbRSV	Tobacco ringspot virus			(26)
TCV	Turnip crinkle virus			(28)
TmRSV	Tomato ringspot virus			(31)
TMV	Tobacco mosaic virus	TMV-T	type strain	(33)
		TMV-J	J14D1 strain	(33)
		TMV-GA	green aucuba strain	(33)
		TMV-YA	yellow aucuba strain	(33)
		TMV-D	dahlemense strain	(33)
		TMV-YT	yellow tomato atypical strain	(33)
		TMV-GR	green tomato atypical strain	(33)
		TMV-N	chemical mutant NBSI 223	(33)
		TMV-O	Odontoglossum strain	(20)
		TMV-HR	Holmes ribgrass strain	(33)
		TMV-CV4	cucumber mosaic virus 4	(35)
TNV[b]	Tobacco necrosis virus			
TRV	Tobacco rattle virus	TRV-B		(23)
		TRV-C		(23)
TYMV	Turnip yellow mosaic virus	TYMV-1A	type strain	(28)
		TYMV-1B	Rademacher strain	(28)
		TYMV-1C	Honesty strain	(28)
		TYMV-2A	cauliflower strain	(28)
		TYMV-2B	Rothamsted strain	(28)
		TYMV-2C	Denmark strain	(28)
		TYMV-WC	wild cucumber mosaic strain	(28)
WCMV	White clover mosaic virus			(18)
WWMV	Winter wheat mosaic virus			(1)

[a] The term strain is restricted to those viruses closely related serologically and showing no differences or only slight differences in their morphology.

[b] The relative molar ratios of amino acids in TNV are: ala, 33; arg, 13; asp, 31; cys, 4; glu, 23; gly, 24; his, 1; ilu, 18; leu, 18; lys, 10; met, 3; phe, 8; pro, 20; ser, 17; thr, 18; trp, 3; tyr, 12; val. 12 (*unpublished results*).

from this information, it is necessary first to adopt a threshold value of r. Then any two proteins whose correlation exceeds this value are regarded as forming a cluster. Membership in this cluster is then extended to any protein having an above-threshold correlation to any already existing member.

This single-link method of determining cluster membership permitted the formation of "stringy" clusters,

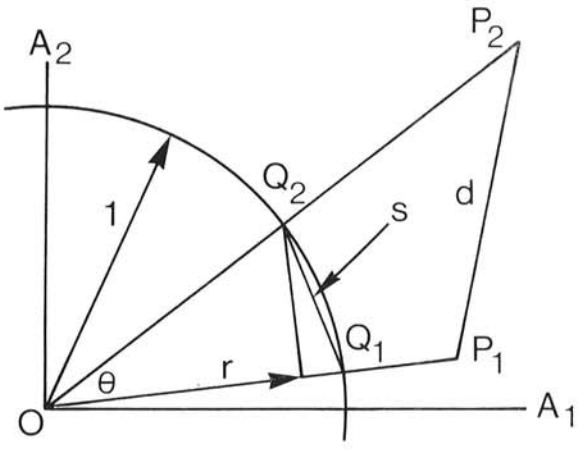

Fig. 1. In this two-dimensional representation of amino acid space, two proteins have composition denoted by points P_1 and P_2. Their distance apart is d. Alternately, these proteins may be regarded as vectors OP_1 and OP_2, which may be normalized to vectors OQ_1 and OQ_2 of unit length. The normalized separation of the proteins is then s, and their correlation coefficient r is equal to the cosine of θ, the angle between them.

but stringy clusters would not appear in the computer output if they were not inherent in the correlation matrix.

The problem of choosing the best threshold value of r was avoided by choosing all values from 0.99 down in steps of 0.01 in successive "sweeps" of the correlation matrix. Each sweep was programmed to insert a 1 in a Boolean matrix whenever a super-threshold value of r was encountered in the corresponding element of the correlation matrix. Then clusters were extracted from the Boolean matrix by a simple, fast algorithm due to Baker (2). Cluster extraction was continued at successively lower threshold values of r until all or most of the proteins appeared in a single large cluster.

A Fortran program for cluster analysis was written by one of the authors (EA). It usually takes less than 5 min on an IBM 7044, and has a number of simple but valuable properties. It begins to produce a few small clusters at high threshold values of r. As the threshold is lowered, the clusters grow in membership; new clusters appear and then grow. Nearby clusters coalesce. At some intermediate value of r there will be a maximum number of clusters. At lower values, clusters coalesce faster than new ones are formed, and the number of clusters falls towards unity.

There are several ways of displaying the program output. They range from a simple listing of clusters at each level of r to a contour diagram resembling a topographic map of mountainous terrain. In this paper, the results are presented in the dendrogram of Fig. 2.

Classification of plant viruses.—The dendrogram of virus clustering will be described with reference to a plant virus classification scheme given in Table 2. This table has been constructed from proposals by Brandes (4), Haselkorn (10), and Kaper (14). The use of proposals by these workers does not necessarily imply their agreement with the taxonomic scheme in Table 2. Indeed, Brandes & Bercks (5) state, "We cannot unite these groups [of elongated viruses] into higher taxonomic units because such steps would be merely speculative. We do not wish to imply that all elongated viruses belong to one taxonomic unit and therefore have a common ancestry." Nevertheless, Table 2 is useful as a background for consideration of the results of the cluster analysis.

TABLE 2. Proposed classification of 27 plant viruses

A. Nonspherical		
1. Rod-shaped		
Group 1	TRV BSMV	Brandes (4)
Group 2	TMV	Brandes (4)
Group 3	(PVX, WCMV)[a]	Brandes (4)
Group 4	PVS	Brandes (4)
2. Bacilliform	AMV	
B. Spherical		
1. Viruses containing 25% RNA or over and with multiple components involving RNA complement		Kaper (14)
a)	TbRSV, TmRSV	Haselkorn (10)
b)	(BPMV, SqMV) EAMV[b]	Haselkorn (10)
c)	(TYMV, TYMV-WC)	Haselkorn (10)
d)	PEMV	Kaper (14)
2. Viruses containing less than 25% RNA		
a)	Viruses with cores (TCV, SoMV, CMtV) TBSV, CNV[c]	Haselkorn (10)
b)	TNV-STNV system	Haselkorn (10)
c)	BBMV, BMV, CCMV, CMV	Kaper (14)
d)	SBMV, CRSV[d]	
C. Unknown shape	WWMV	

[a] Viruses within parentheses may be distantly serologically related, or co-related to another virus.

[b] EAMV was added to this group because of similarities in size, RNA content, and host range (38).

[c] CNV was added to this group because of similarities in size, the presence of cores, and RNA content (*unpublished results*).

[d] These viruses have similar sizes, RNA contents, and stabilities to temperature, pH values, and high molarity buffers (13, 29).

Discussion of cluster analysis results.—In this discussion it is assumed that the data, amino acid compositions, are exact. In fact, some of the determinations are of uncertain accuracy, owing to experimental difficulties. The general effect of random data errors in cluster analysis is to lower the correlation coefficients and weaken any tendency to clustering that might be present.

The most obvious result of the cluster analysis program is the tight clustering of strains of a virus. In most cases strains not only form compact clusters but also lie in widely separate regions of amino acid space. It is not uncommon, however, for an outlying member to form a bridge to a nearby cluster.

These features of the clustering are displayed by the large (11 member) tobacco mosaic virus (TMV) group (A.1. Group 2 in the classification of Table 2). Four TMV strains form a very tight cluster at a correlation threshold of 0.99, as do two other pairs. One pair joins

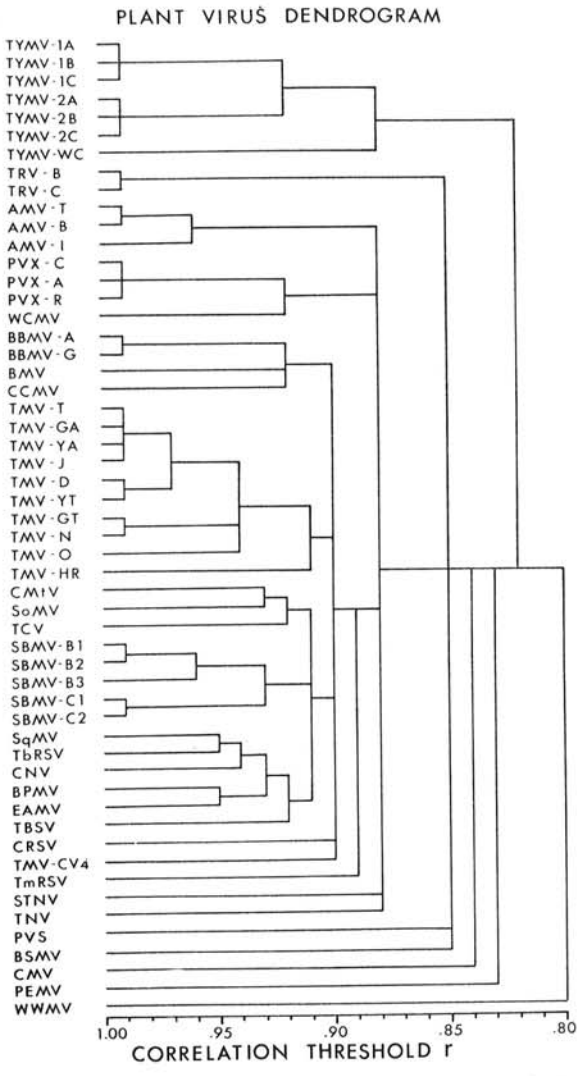

Fig. 2. The dendrogram of plant viruses from cluster analysis of amino acid composition of virus coat proteins. The horizontal lines from virus strains are joined by vertical lines at the correlation threshold, r.

the larger group at the 0.97 level, and the second pair comes in at 0.94, where the growing cluster is also joined by TMV-O. At 0.91, TMV-HR becomes a member. At this point, the 10-member cluster is isolated from all other viruses in the study. However, the eleventh member (TMV-CV4) is outlying, and only enters the cluster when a bridge is formed between the TMV cluster and other virus clusters.

There is wide variation in the tightness of clustering of virus strains. Strains of alfalfa mosaic virus (AMV), broad bean mottle virus (BBMV), potato virus X (PVX), and tobacco rattle virus (TRV) form very tight clusters, but there are outlying strains of TMV, Southern bean mosaic virus (SBMV), and turnip yellow mosaic virus (TYMV). The outlying strains of SBMV and TMV do not share a host in common with the members of the main cluster. This fact indicates a considerable evolutionary distance between the outlying strains and the members of the main cluster.

A less compact clustering of viruses that have a distant serological relationship is also evident in Fig. 2. The 3 PVX strains (A.1. Group 3) correlate at 0.99; the distantly related white clover mosaic virus (WCMV) joins them at 0.92. At this point, the four members are well isolated from all other viruses. Two viruses of the B.2.a group, carnation mottle virus (CMtV) and sowbane mosaic virus (SoMV), cluster at the 0.93 level, and are joined by turnip crinkle virus (TCV) at the 0.92 level, but the cluster is not well isolated. TCV is serologically related to CMtV, which in turn is related to SoMV (10).

Within the major divisions in Table 2 there is considerable agreement between the classification and the cluster results. The rod-shaped A.1. Groups 3 and 4 and bacilliform A.2. remain well isolated from other viruses, and only join other species clusters at the low correlation levels of 0.88, 0.85, and 0.88, respectively.

In some cases, a low level of correlation may indicate an incorrect grouping of viruses in Table 2. Brandes & Bercks (5) regarded the placing of tobacco rattle virus (TRV) and barley stripe mosaic virus (BSMV) in A.1. Group 1 as speculative and probably incorrect, because the normal length of BSMV is only 70% of TRV, and serological relationship has not been demonstrated. Although these viruses join with clusters of other viruses at the 0.85 level, they have a correlation coefficient of only 0.77.

Within the "spherical" viruses, the B.1.c or TYMV group forms a cluster with two condensations of TYMV and a more weakly coupled TYMV-WC. This group and the one virus in the B.1.d group are well separated from each other and the rest.

The B.1.b and B.2.a groups belong to overlapping clusters and include tomato ringspot virus (TmRSV) from group B.1.a. The two viruses in the B.2.b. group are not closely associated with any group or with each other. However, the common feature of these viruses is the dependence of satellite tobacco necrosis virus (STNV) on tobacco necrosis virus (TNV) for multiplication, and a close correlation was not expected.

The B.2.c group clusters well at the 0.92 level, except for cucumber mosaic virus (CMV), which is very outlying and not associated with any other virus. The molecular wt of CMV and of the protein subunit is 60% greater than that of the other members of the B.2.c group. For this reason, inclusion of CMV seems unwarranted. The three viruses of B.2.c, cowpea chlorotic mottle virus (CCMV), brome mosaic virus (BMV), and broad bean mottle virus (BBMV) have similar proteins which will mix in reassembly (36).

The SBMV strains of group B.2.d form a compact cluster, but the carnation ringspot virus (CRSV) of this group is some distance away and not linked closely to any cluster. Finally, winter wheat mosaic virus (WWMV), of unknown shape, has no correlations above 0.81, and lies well away from all the clusters. Although this protein is found only in virus-infected plants, it has not been proven to be a viral protein (1).

The presence of several structural proteins has been demonstrated in some small animal viruses and some bacteriophages (14). G-j. Wu and G. Bruening (*personal communication*) have detected two structural proteins in cowpea mosaic virus which are serologically related to bean pod mottle virus (BPMV) in the B.1.a group of Table 2. If this structure is common to members of this group, the amino acid compositions used for this group are mean compositions of two proteins and are not comparable to compositions of single proteins. Removal of BPMV, Squash mosaic virus (SqMV), and Echte acherbohnen mosaik virus (EAMV) from the cluster analysis does not affect the clustering of other viruses shown in Fig. 2.

Inspection of the correlation matrix (not shown here) reveals that carnation mottle virus (CMtV) lies near the center of gravity of the entire lot of 54 viruses. It correlates at $r = 0.69$ or better with all of them, and at 0.81 or better with the near half. In other words, these 54 virus proteins occupy a hyperspherical cap of angular radius $\cos^{-1}$ (0.69) or 46 degrees. In 3-dimensional space, this surface would occupy about one-sixth the area of the sphere, but in 18-space the corresponding fraction is only 5×10^{-4}. In an unpublished study, 100 miscellaneous proteins occupied 0.05 of the amino acid hypersphere. If this group of proteins had purely random amino acid compositions, they would occupy the entire surface of the hypersphere. In contrast with these results, the virus proteins may be regarded as forming a single compact cluster.

A principal coordinate analysis of most of the amino acid compositions of plant viruses used in our study was reported recently by Gibbs (8). Gibbs recognized that about half of the input information was lost by this method of analysis. The clustering of viruses in his diagrams is loose and not easily interpreted, making comparison with our results difficult. Gibbs (8) discussed the qualitative and quantitative characteristics of plant viruses and used these in a computer classification of 140 viruses. The results of his study are of great value to plant virus taxonomy.

DISCUSSION.—Low correlation coefficients between virus coat proteins signify large differences in composition, and these differences imply considerable evolutionary distance between the viruses. Even when the proteins being compared have an approximately equal number of amino acids, it is difficult to establish more than a rough relationship between the number of evolutionary exchanges of RNA bases and the correlation coefficient between the type and the evolved strain. In the comparison of proteins of substantially different sizes, the number of possible evolutionary paths between them is even greater. Nevertheless, it is probable that correlation coefficients will diminish as evolutionary distance increases.

The cluster dendrogram gives no support to the main subdivisions in the classification system of Table 2 that are based on shape of the virus. These subdivisions imply that rod-shaped viruses arose from one provirus, and that spherical viruses arose from another. We agree with Brandes & Bercks (5) that this implication is merely speculative. Apparently, viruses have evolved a considerable distance from their origins, and these evolutionary distances are too great to allow the deduction of origins from amino acid composition.

Since the dendrogram gives no support to the subdivision of spherical viruses based on RNA content and presence of multiple components involving RNA complement, these considerations also apply to the ancestry of the two groups of spherical viruses. There has been a lack of systematic serological studies of spherical viruses comparable to the work with rod-shaped viruses (5) requiring precise experimental conditions to detect low degrees of serological relationship. Such studies are critical for assessing the subdivisions of spherical viruses (Table 2).

The quality and quantity of virus protein analyses available for the present study undoubtedly had an effect on the assessment of the classification. The values of one-third of the amino acids (cysteine, tryptophan, serine, threonine, valine, and isoleucine) are undoubtedly inaccurate in many analyses. Of 47 viruses and strains of rod-shaped viruses listed by Brandes & Bercks (5), compositions were available for only 7. A greater number of more accurate analyses would undoubtedly show relationships or lack of relationships, of value in taxonomy, and would serve as a guide to future serological studies.

In addition to serology and composition there are other criteria of importance in virus classification. Some criteria involve structural information and have been discussed by other workers (5, 14). The sequence of amino acids in virus proteins may yield information on distant relationships, but technical difficulties preclude rapid increase of knowledge in this area. It must be stressed that knowledge of the structure of virus coats and of the sequence of amino acids in virus proteins still provides information on only one of the ten to twenty cistrons in the gene of a plant virus.

With the rapid increase in the number and accuracy of publications on analyses of viral proteins, it becomes important to assess the relevance of such information to virus taxonomy. Cluster analysis of proteins by amino acid composition offers a simple, inexpensive, and rapid procedure for gauging protein similarities. This study shows that closely related viruses and virus strains form tight clusters in amino acid space, and suggests that virus shape should not be the overriding criterion for virus classification.

LITERATURE CITED

1. ATABEHOV, J. G., G. A. POPOVA, M. A. KISELEV, A. S. KAFTANOVA, & G. V. PETROVSKY. 1968. In vitro polymerization of winter wheat mosaic virus antigen. Virology 35:458-472.
2. BAKER, J. J. 1962. A note on multiplying boolean matrices. Assoc. for computing machinery. Communications 5:102.
3. BANCROFT, J. B., E. HIEBERT, M. W. REES, & R. MARKHAM. 1968. Properties of cowpea chlorotic mottle virus, its protein and nucleic acid. Virology 34:224-229.
4. BRANDES, J. 1964. Identifizierung von gestreckten pflanzenpathogenen Viren auf morphologischer Grundlage, zugleich eine zusammenstellung von grundlegen

Daten für die Klassifierung dieser Viren. Mitt. Biol. Bundesanstalt Land-U. Forstwirtsch. Berlin-Dahlem 110:5-130.

5. Brandes, J., & R. Bercks. 1965. Gross morphology and serology as a basis for classification of elongated plant viruses. Advances Virus Res. 11:1-24.
6. De Fremery, D., & C. A. Knight. 1955. A chemical comparison of three strains of tomato bushy stunt virus. J. Biol. Chem. 214:559-566.
7. Ghabrial, S. A., R. J. Shepherd, & R. G. Grogan. 1967. Chemical properties of three strains of Southern bean mosaic virus. Virology 33:17-25.
8. Gibbs, A. 1969. Plant virus classification. Advances Virus Res. 14:263-328.
9. Gumpf, D. J., & R. I. Hamilton. 1968. Isolation and characterization of barley stripe mosaic virus protein. Virology 35:87-93.
10. Haselkorn, R. 1966. Physical and chemical properties of plant viruses. Annu. Rev. Plant Physiol. 17:137-154.
11. Hatch, F. T. 1965. Correlation of amino acid composition with certain characteristics of proteins. Nature 206:777-779.
12. Kado, C. I. 1967. Biological and biochemical characterization of sowbane mosaic virus. Virology 31: 217-229.
13. Kalmakoff, J., & J. H. Tremaine. 1967. Some physical and chemical properties of carnation ringspot virus. Virology 33:10-16.
14. Kaper, J. M. 1968. I. The small RNA viruses of plants, animals and bacteria. A. Physical properties, p. 1-133. *In* H. Fraenkel-Conrat [ed.] Molecular Basis of Virology. Reinhold, New York.
15. Kelley, J. J., & P. Kaesberg. 1962. Biophysical and biochemical properties of top component a and bottom component of alfalfa mosaic virus. Biochim. Biophys. Acta 61:865-871.
16. Mazzone, H. M., N. L. Incardona, & P. Kaesberg. 1962. Biochemical and biophysical properties of squash mosaic virus and related macromolecules. Biochim. Biophys. Acta 55:164-175.
17. Miki, T., & C. A. Knight. 1965. Preparation of broad bean mottle virus protein. Virology 25:478-481.
18. Miki, T., & C. A. Knight. 1967. Some chemical studies on a strain of white clover mosaic virus. Virology 31:55-63.
19. Miki, T., & C. A. Knight. 1968. The protein subunit of potato virus X. Virology 36:168-173.
20. Paul, H. C., C. Wetter, H. G. Wittmann, & J. Brandes. 1967. Untersuchungen am *Odontoglossum* ringspot virus, einem Verwandten des TMV. Z. Vererbunglehre 97:186-191.
21. Reichmann, M. E. 1964. The satellite tobacco necrosis virus: A single protein and its genetic code. U.S. Natl. Acad. Sci. Proc. 52:1009-1017.
22. Semancik, J. S. 1966. Studies on electrophoretic heterogeneity in isometric plant viruses. Virology 30:698-704.
23. Semancik, J. S. 1966. Purification and properties of two isolates of tobacco rattle virus from pepper in California. Phytopathology 56:1190-1193.
24. Shaw, J. G., M. E. Reichmann, & D. L. Hatt. 1962. Amino acid analysis of the proteins of two strains of potato virus X. Virology 18:79-88.
25. Shepherd, R. J., R. J. Wakeman, & S. A. Ghabrial. 1968. Preparation and properties of the protein and nucleic acid components of pea enation mosaic virus. Virology 35:255-267.
26. Stace-Smith, R., M. E. Reichmann, & N. S. Wright. 1965. Purification and properties of tobacco ringspot virus and two RNA-deficient components. Virology 25:487-494.
27. Stubbs, J. D., & P. Kaesberg. 1964. A protein subunit of bromegrass mosaic virus. J. Mol. Biol. 8:314-323.
28. Symons, R. H., M. W. Rees, M. M. Short, & R. Markham. 1963. Relationship between the ribonucleic acid and protein of some plant viruses. J. Mol. Biol. 6:1-15.
29. Tremaine, J. H. 1966. The amino acid and nucleotide composition of the bean and cowpea strains of Southern bean mosaic virus. Virology 30:348-354.
30. Tremaine, J. H., & D. E. Goldsack. 1968. The structure of regular viruses in relation to their subunit amino acid composition. Virology 35:227-237.
31. Tremaine, J. H., & R. Stace-Smith. 1968. Chemical composition and biophysical properties of tomato ringspot virus. Virology 35:102-107.
32. Tremaine, J. H., & R. Stace-Smith. 1969. Amino acid analyses of two strains of alfalfa mosaic virus. Phytopathology 59:521-522.
33. Tsugita, A. 1962. The proteins of mutants of TMV: Classification of spontaneous and chemically evoked strains. J. Mol. Biol. 5:293-300.
34. Van Regenmortel, M. H. V. 1967. Biochemical and biophysical properties of cucumber mosaic virus. Virology 31:391-396.
35. Van Regenmortel, M. H. V. 1967. Serological studies on naturally occurring strains and chemically induced mutants of tobacco mosaic virus. Virology 31:467-480.
36. Wagner, G. W., & J. B. Bancroft. 1968. The self-assembly of spherical viruses with mixed coat proteins. Virology 34:748-756.
37. Waugh, D. F. 1959. Proteins and their interactions. Rev. Mod. Phys. 31:84-93.
38. Wetter, C., H. L. Paul, J. Brandes, & L. Quantz. 1960. Vergleich zwischen Eigenschaften des echte Ackerbohnenmosaik-Virus und des broad bean mottle-Virus. Z. Naturforschg. 15b, 444-447.
39. Wittmann, H. G., & H. L. Paul. 1961. Vergleich der Aminosaurenzusammensetzung der Proteine des Echten Acherbohnenmosaik-Virus, des broad bean mottle-Virus, und des Tabakmosaik-Virus. Phytopathol. Z. 41:74-78.