

## VIRUS STRUCTURE: GENERAL PRINCIPLES

F. H. C. CRICK AND J. D. WATSON\*

*The Medical Research Council Unit for the Study of the Molecular Structure of Biological Systems, Cavendish Laboratory, Cambridge*

### Introduction

IN this article we shall discuss some general ideas about the structure of viruses. This is a hazardous undertaking. We know of no principles so compelling that we can be certain that they must be true; or, more correctly, those that must be true—the rules for inter-atomic distances, for instance—do not lead directly to any interesting conclusions. However, there are certain ideas suggested by experience in related fields (such as the study of protein crystals) which we might well expect to apply to viruses, or at any rate to small viruses. Moreover we can make some use of that powerful but dangerous weapon, the principle of simplicity.

Our ideas fall into two groups. There is good evidence in the case of three plant viruses, and indirect evidence for certain animal viruses, that the protein component of a virus is made up of sub-units. Our first set of ideas concerns the question: why does a virus have protein sub-units? We have not previously published this argument. Our second deals with the problem: if there are sub-units, how are they arranged? This we have recently put forward elsewhere, so that we shall only deal with it briefly. This paper should therefore be read in conjunction with our previous one (Crick and Watson, 1956).

We shall restrict our discussion in the first place to those small viruses which contain only protein and ribonucleic acid (RNA): that is, the majority of known plant viruses, and

\* On leave from the Biology Department, Harvard University, and supported by a grant from the National Science Foundation (U.S.A.).

certain animal viruses such as poliomyelitis and the various encephalitic viruses.

### The reason for protein sub-units

Our basic argument is that the protein component of a virus is unlikely to be either one large molecule or, alternatively, an assembly of small molecules, each of which is quite different from all the others. More precisely, we mean by "different" that the sequence of amino acids in any two such small molecules is quite unrelated.

Our first assumption is that an essential requirement for a virus of this type is that it should consist of a packet of RNA protected by a coat of protein. It is found experimentally that the molecular weight of the RNA is of the order of, say,  $2 \times 10^6$ . Imagine that this amount of RNA is folded as compactly as possible, so that it forms a rather dense sphere. Such a sphere could hardly be less than 150 Å diameter, and is more likely to be nearer 200 Å. We next surround this with a layer of protein, which we shall assume is more or less continuous. There must be a minimum thickness for such a layer; 1 Å, for example, would be impossibly small. A more reasonable minimum value would be 10 Å. Actually no protein crystal is known with a unit cell dimension of less than 24 Å, so that perhaps 20 Å would be a more realistic lower limit. This would require a volume of protein of about  $10^7 \text{ Å}^3$ , or a molecular weight near 7 million. The details of the calculation are unimportant; the point is that we require a large amount of protein. Notice that the ratio of protein to RNA *increases* as we make the virus smaller; that is, if we had considered a smaller amount of RNA we should not reduce the amount of protein required by very much. We can only have a much smaller *proportion* of protein if the virus is considerably bigger.

The model we have described must not be taken as a detailed model of a virus. It is used purely for illustration. If we follow through the argument for a rod-shaped virus of small diameter we reach a similar conclusion.

Thus, if our assumption that a small virus has to have a reasonably continuous protein coat is correct, we can conclude that a relatively large amount of protein will be required for it. Whatever the reason, the experimental evidence shows clearly that a considerable amount of protein is always present. In Table I we have set out the figures for all the

Table I  
AMINO ACID AND NUCLEOTIDE RESIDUES IN RNA VIRUSES

	Molecular Weight	% RNA	% Protein	Nucleotides per Particle	Amino Acids per Particle
Tobacco mosaic virus	$40 \times 10^6$	6	94	7,800	840,000
Potato virus X	$\approx 30 \times 10^6$	6	94	5,400	260,000
Potato virus Y	$75 \times 10^6$	5	95	11,800	650,000
Bushy stunt virus	$9 \times 10^6$	16	84	4,400	69,000
Turnip yellow mosaic virus	$5 \times 10^6$	40	60	6,000	27,000
Southern bean virus	$6 \times 10^6$	21	79	3,800	43,000
Tobacco ringspot virus	$\approx 6 \times 10^6$	40	60	7,300	33,000
Tobacco necrosis virus	$\approx 6 \times 10^6$	18	82	3,300	45,000
Poliomyelitis	$10 \times 10^6$	24	76	7,300	69,000
Influenza*	$100 \times 10^6$	2	75	6,000	680,000
Fowl plague*	$100 \times 10^6$	2	75	6,000	680,000

The figures in this Table are only approximate

\* These viruses may contain material from their host cell

small viruses for which data are available. It can be seen that in every case the *total* number of amino acids always greatly exceeds the *total* number of nucleotides.

Our next assumption is more difficult to justify. It really falls into three parts. We assume (a) that the amino acid sequence of the protein component of the progeny is determined wholly, or at least to a large extent, by the infecting virus; (b) that this amino acid sequence is determined by the molecular structure of the RNA of the infecting virus, and not at all by its protein component; (c) that the "coding" implied in (b) is relatively simple.

Of course none of these assumptions is new, though we believe that our argument as a whole is original.

We wish, from these assumptions, to make a crude estimate of how much protein can be "coded" by a given amount of RNA. To fix ideas consider a scheme in which the first three bases of the RNA chain is a code for the first amino acid in the polypeptide chain, the next three bases for the second amino acid, and so on. For an RNA chain of molecular weight  $2 \times 10^6$ , which has about 6,000 bases, this implies that we can code for a polypeptide chain (or chains) of total length 2,000 amino acids, or about 230,000 molecular weight. To form a protein coat, however, we need at least 10 times as much as this, and probably 20 or 30 times as much.

At the moment we know practically nothing about the "code", so we cannot tell whether the estimate of three bases to one amino acid is a good one. However, this ratio, which we have taken as 3 : 1, can scarcely be less than 1 : 1, and a careful study of known amino acid sequences in (non-viral) proteins (Gamow, Rich and Ycas, 1956) suggests that a 1 : 1 ratio is unlikely, since such a code necessarily puts considerable restrictions on the possible amino acid sequences, and these are not apparent.

The way out of this difficulty is obvious. There seems to be no reason why the virus-cell system should not produce a large number of copies of the protein molecules for which the RNA of the virus is a code. The protein coat of the virus would then consist not merely of *one* of these molecules, but of a number of them, and in this way one can obtain a large "molecular weight" from a system which can only produce relatively small protein molecules.

We can now summarize our argument in slightly different terms. The information required to synthesize the virus protein is contained in the RNA. As there is only a limited amount of RNA it can only carry a limited amount of information. Thus the protein molecules of the virus can only be of limited size. Rough numerical estimates show that this amount, used once, is not enough to produce a shell to cover the RNA. Thus the coat must be built up of identical sub-units.

### The Coding Ratio

If, as we surmise, the "coding ratio"—the number of nucleotides which code, on the average, for one amino acid—is a constant throughout Nature, the plant viruses present very attractive material from which to obtain this ratio experimentally. When we come to grips with the problem, however, we run into formidable difficulties. Only for tobacco mosaic virus (TMV) can we be even approximately sure of the sub-unit size, and here the number of nucleotides in the virus is approximately 50 times the number of amino acids in the protein sub-unit. This figure seems much too large to have a direct bearing on the coding problem. There are several possible explanations for the large value. One, which we must consider most seriously, is that the RNA of TMV controls the synthesis of other proteins in addition to that which forms its outer shell. Another is that the virus may contain several or many copies of the fundamental RNA chain. Still another possibility is that the protein sub-units are not all completely identical, so that the RNA is responsible for the synthesis of several closely related proteins. At first glance bushy stunt virus looks more hopeful, as we find that there are only four nucleotides per amino acid residue in the crystallographic sub-unit. This answer, however, is probably misleading, for the chemical data (de Fremery and Knight, 1955; quoted in Caspar, 1956) hints that each crystallographic sub-unit may contain perhaps five chemical sub-sub-units, thus giving 20 nucleotides per amino acid residue, which seems unreasonably large. It is clear that much further work, from many approaches (including that of genetics), will be required before a reliable figure can be obtained.

### Experimental Evidence

It would be impossible in a short article to review in detail the evidence in favour of each of our assumptions, but we can mention some of it briefly. We should state straightaway that it is, at the moment, inadequate in almost every case.

The assumption that the RNA is protected by a coat of protein is supported to some extent by the fact that these viruses are not attacked by ribonuclease. The evidence for the protein coat is good for tobacco mosaic virus (TMV) (see Williams, 1956; Franklin, Klug and Holmes, 1956) and for turnip yellow mosaic virus (Markham, 1951; Bernal and Carlisle, 1951; Schmidt, Kaesburg and Beeman, 1954). Even if this assumption is wrong, Table I shows clearly that, whatever the reason, the protein/RNA ratio is large for all the viruses studied.

For our assumption (a)—that the amino acid sequence of the progeny virus is determined by the genetic specificity of the infecting virus and not to any significant extent by the host cell—the evidence is again only suggestive. It rests largely on the existence of the mutant virus strains. These strains breed true within the usual plant hosts and have been shown (see Knight, 1956) to possess characteristic differences in their amino acid compositions. However a recent paper (Bawden, 1956) suggests that the host cell may play a far more significant rôle than commonly supposed.

Assumption (b)—that the amino acid composition of a virus is determined by the RNA of the infecting virus—is supported by the classic experiment of Harris and Knight (1952; 1955) in which they removed the terminal threonine from the protein sub-units of TMV and found them again in the progeny, and more strongly by the very recent work of Fraenkel-Conrat (1956). In a typical experiment protein from standard TMV was combined with RNA from the Holmes ribgrass strain. After infection the resulting progeny had a protein component closely resembling that of the Holmes ribgrass strain. Notice that it is immaterial to our argument whether the infective unit is RNA alone\* or the recombined RNA plus protein. The essential experimental requirement is that no protein from the Holmes ribgrass strain should get into the plant. This experiment of Fraenkel-Conrat's is one of the

\* Added in proof: see the important recent paper of Gierer and Schramm (1956) in which it is reported that the RNA alone is infective.

utmost importance, not only for virus work, but for the whole field of protein synthesis, and we look forward to it being repeated and extended by other workers\* so that its apparent conclusions can be established beyond doubt.

We know of no experimental evidence which directly supports our third postulate (c).

### The arrangement of protein sub-units

We have very recently discussed this (Crick and Watson, 1956) in an attempt to answer the question, "Why are all small viruses either rods or spheres?" Here we will put the question the other way round and ask, "Given that the protein component is made of sub-units, how will they be arranged?"

In our view it is not very likely, though not impossible, that the sub-units will aggregate in a random manner. On general grounds we would expect that the preferred arrangement will be one in which every sub-unit has the same environment as every other sub-unit. This can only be done if the sub-units are related by symmetry elements. The two most likely arrangements are a spherical shell, having cubic symmetry, or a cylindrical shell, having a screw axis, though other symmetries could occur. There are three possible arrangements of symmetry elements which can give cubic symmetry, having 12, 24 and 60 asymmetric sub-units respectively. Thus we predict that many small spherical viruses will have  $12n$  protein sub-units, where  $n$  is an integer. Our original article should be consulted for details.

The X-ray evidence has established the existence of sub-units for TMV (see Franklin, Klug and Holmes, 1956), for tomato bushy stunt (Caspar, 1956) and for turnip yellow mosaic (Franklin and co-workers, unpublished). Note that the X-rays can only show that the sub-units are structurally similar, not structurally identical, let alone chemically identical. Also note that the crystallographic sub-unit may consist

\* Added in proof: see the important recent paper of Gierer and Schramm (1956) in which it is reported that the RNA alone is infective.

of several chemical sub-units, which may or may not be identical. The X-rays are also incapable of showing whether the sub-units are held together by physical bonds only, or also by chemical bonds, such as S-S bridges, which might be formed during or after the aggregation process.

Very little is known about the detailed arrangement of the RNA in these viruses. It is a reasonable hypothesis that the RNA has the symmetry, or at least some of the symmetry elements, of the protein component. This symmetry may not apply to the precise sequence of the bases of the RNA but only to the phosphate-sugar backbone.

It is rather striking that the absolute content of RNA, for both the rod-shaped and the spherical viruses, varies over rather narrow limits compared with the highly varying amounts of protein. Almost all these small viruses contain between 3,000 and 7,500 nucleotides and we wonder whether there may be an effective lower limit to the size of a virus corresponding to the amount necessary to make a moderately sized protein molecule. There are reports of smaller viruses, in particular the Rothamsted strain of tobacco necrosis virus (Bawden and Pirie, 1950), but there seem to be strong doubts whether these small particles of 150 Å diameter are infective. If they are we should have a virus of  $1.5 \times 10^6$  molecular weight, containing about 900 nucleotides. According to our above arguments, this amount of RNA might be expected to control the synthesis of a protein of 20–30,000 molecular weight.

The possibility that the *arrangement* of the RNA may be practically the same in all spherical viruses should not be overlooked, since the ways of folding a fibrous molecule so that it has cubic symmetry may be rather limited. For this reason it is quite possible that microsomes, which also appear spherical and are of a similar size to the smaller viruses (actually a little smaller, but the percentage of RNA is probably greater), may also have cubic symmetry, and perhaps possess protein sub-units. In any case, since microsomes contain only a limited amount of RNA, we would predict, by

an extension of our first argument, that no very large protein molecule will be found that is not an aggregate. The fragmentary evidence available on large protein molecules (such as myosin) is compatible with this idea.

### Other Viruses

We shall only discuss this very briefly. We think it likely that these ideas will apply to all viruses (both those containing DNA as well as those containing RNA) which have a precise size and shape, and in particular to all small viruses. The ideas may have to be adapted somewhat in special cases, such as the influenza virus, the outer shell of which may not be so well ordered as the inner part and probably consists in part of components of the host. Bacteriophages, such as T2, present a very challenging problem. We feel that our ideas may apply to the separate parts, but even so the shape of the head requires some special explanation.

### REFERENCES

- BAWDEN, F. C. (1956). *Nature, Lond.*, **177**, 302.  
 BAWDEN, F. C., and PIRIE, N. W. (1950). *J. gen. Microbiol.*, **4**, 464.  
 BERNAL, J. D., and CARLISLE, C. H. (1951). *Disc. Faraday Soc.*, **11**, 227.  
 CASPAR, D. L. D. (1956). *Nature, Lond.*, **177**, 475.  
 CRICK, F. H. C., and WATSON, J. D. (1956). *Nature, Lond.*, **177**, 473.  
 FRAENKEL-CONRAT, H. (1956). *J. Amer. chem. Soc.*, **78**, 882.  
 FRANKLIN, R. E., KLUG, A., and HOLMES, K. C. (1956). This Symposium, p. 39.  
 DE FREMERY, D., and KNIGHT, C. A. (1955). *J. biol. Chem.*, **214**, 559.  
 GAMOW, G., RICH, A., and YCAS, M. (1956). *Advanc. biol. med. Phys.*, **4**.  
 GIERER, A., and SCHRAMM, G. (1956). *Nature, Lond.*, **177**, 702.  
 HARRIS, J. I., and KNIGHT, C. A. (1952). *Nature, Lond.*, **170**, 613.  
 HARRIS, J. I., and KNIGHT, C. A. (1955). *J. biol. Chem.*, **214**, 215.  
 KNIGHT, C. A. (1956). This Symposium, p. 69.  
 MARKHAM, R. (1951). *Disc. Faraday Soc.*, **11**, 221.  
 SCHMIDT, P., KAESBURG, P., and BEEMAN, W. W. (1954). *Biochim. biophys. Acta*, **14**, 1.  
 WILLIAMS, R. C. (1956). This Symposium, p. 19.

## DISCUSSION

(Dr. Caspar's remarks were illustrated with a variety of models)

*Caspar:* I should like to make some comments to indicate how symmetry can lead to spherical virus particles. First, Dr. Crick has pointed out that you cannot have mirror symmetry in a virus because it is built up of asymmetric molecules. Therefore there are only two types of symmetry relationships that you can have; either rotational or translational. As an example of rotational symmetry, consider a wheel with four spokes; if such a wheel is rotated through  $90^\circ$  about its axis it looks the same as it did before. This is the characteristic of the symmetry operation on a structure—that afterwards it looks the same as it did before. In this case we would say that the wheel has a fourfold rotation axis. Translational symmetry applies to a repeating linear. In biological structures the most common and interesting example of translational symmetry is when it is combined with rotational symmetry to give a screw axis, and this generates a helix, as in the case of TMV.

What I particularly want to discuss are the spherical viruses. Assuming that they do have sub-units and that these are arranged in a regular way, the question is: what kind of symmetry elements can lead to a spherical particle? The only symmetry elements there can be in this case are rotational symmetry. If there were just one axis of rotational symmetry, it would in general, with most shapes of sub-unit, lead to a disc-shaped, or a cylindrical-shaped, particle. It would be possible to have sub-units shaped like sectors of an orange, and related by rotational symmetry, and this would lead to a spherical particle; but this is actually a rather arbitrary model. The only general way to put sub-units of any shape together to get a spherical particle is by using cubic symmetry.

There are only three kinds of cubic symmetry of interest to us, and these can be illustrated by the five platonic solids. The tetrahedron with 2- and 3-fold axes represents the lowest type of cubic symmetry. The characteristic of cubic symmetry is that there are four 3-fold rotation axes, so related that they point along the diagonals of the cube. The second type is shown by the cube and the octahedron, which have threefold, twofold, and in addition, fourfold rotation axes. The third type is the symmetry represented by the regular dodecahedron or the icosahedron, in which there is fivefold, threefold, and twofold symmetry. (The platonic solids actually have mirror planes and a centre of symmetry as well as rotation axes, but for illustration we can ignore these.)

Symmetry does not tell us what a particle is going to look like; it just indicates how the parts are related. We have experimental evidence from X-ray work in the case of bushy stunt virus that it has the symmetry of the regular dodecahedron. Dr. Klug and Mr. Finch have also found that turnip yellow virus probably has this same type of symmetry.

Let us go back to the tetrahedron to indicate how you get sub-units from symmetry. Consider one of the threefold rotation axes. If I arbitrarily put a sub-unit in any place, I have to put another two,

related by this threefold rotation axis, on the same face. Since there are four identical faces I end up with twelve sub-units, arranged in a very definite way. Likewise for the cube or the octahedron I shall end up with twenty-four sub-units. In the dodecahedron there are twelve fivefold faces, so that this will lead to sixty sub-units, arranged in a regular way.

We have built models to show these 8 types of symmetry using ping-pong balls. If, by any chance, the virus has spherical sub-units it might look something like this. The characteristic of this model is that every ping-pong ball is related to all the others in exactly the same way. In this particular model there are sixty ping-pong balls, so it means that by just a combination of rotation I can twist it around in sixty different positions and yet the model will always look exactly the same in each of these positions. As Dr. Crick pointed out, these sub-units are not necessarily chemically identical; they are structural units which look similar to the X-rays.

You will notice that all the models I have here are approximately spherical, since they are made using rather compact sub-units. You will also notice that they all look different in detail, so that we cannot say exactly what the virus looks like merely from a knowledge of its symmetry.

What we have to do now is find out what the sub-units look like, and how they fit together to make up the complete virus.

*Klug:* On the question of packing of protein sub-units, I should like to report that we have recently obtained X-ray diffraction pictures of crystals of turnip yellow mosaic virus and we find that they can be interpreted in terms of the suggestions put forward by Crick and Watson, that is, that the particles themselves have cubic symmetry. It happens to be 52 symmetry in this case, just as Caspar found for bushy stunt. This implies again that the "spherical" virus is built up of sub-units. I think this is a case where the hypothesis has proved very useful, because it so happens that in the crystal the actual symmetry of the individual virus particles is masked by the higher symmetry of the particular lattice in which they are arranged, and the diffraction pattern would have been even more difficult to interpret had there not been this hypothesis.

*Williams:* Dr. Crick, is there any particular specification to the coding that makes a coding ratio of something like 1 to 10 and 1 to 20 quite unreasonable?

*Crick:* We know so little about the coding that one would not like to say that it is quite unreasonable, but one would certainly think that on general grounds it was more like 1 to 2 or 3. This is one of the subjects on which there is a lot of unpublished speculation. At the moment we are only considering the simplest possible models, and I should be surprised if it was 10 or 20 bases to one amino acid, but I do not think one can say that it is absolutely impossible. With ingenuity we could probably think of a scheme of that sort, but the one that you would think of first is that, since you have four possibilities in the RNA and 20 possibilities in protein, you would expect a ratio of a little more than 2 to 1.

*Watson:* The X-ray evidence of Caspar indicated that the protein coat of bushy stunt virus is constructed from 60 similar sub-units. Each sub-unit has a molecular weight around 125,000 and contains approximately 1,100 amino acids. The RNA content of bushy stunt is 16 per cent which corresponds to about 4,400 nucleotides. We thus find four nucleotides per amino acid in the crystallographic sub-unit. But there is a complication that the end-group analysis which Dr. Knight may tell us about, suggests a sub-unit smaller than the crystallographic sub-unit, maybe by a factor of 2 or 3. However, even here we cannot be at all sure that a similarity in end groups means equivalent amino acid sequences.

*Crick:* You should in fact be able to put an upper limit on an argument of this type to the value of the coding ratio. It seems unlikely that in bushy stunt it would be as much as 20 to 1.

*Williams:* You would say then that the nucleic acid directly codes a certain number of amino acids which would be perhaps something of a 1 to 10 or a 1 to 20 ratio; as much as is necessary. After that coding is established for one set of amino acids, the entire sub-unit is built up of a repetition of that set of coded amino acids. Is that essentially your argument?

*Watson:* No, I mean the protein shell is made up of a series of sub-units, and the RNA is sufficient to code to provide the information to determine the sequence in one of the protein sub-units.

*Crick:* And, it is not enough to code for a single protein having the molecular weight of the whole virus. Unfortunately the calculation is not precise enough, for all the reasons we have just been mentioning, to say exactly how much you can code. That is what we should like to know. In fact, I think the real point of this idea is to bring home to people that there is possibly something here to look for, rather than to put it forward as something very definite.

*Williams:* But are you not going to get into geometrical difficulties if you say that the RNA codes all of one sub-unit? How does the RNA expose itself to the whole of one sub-unit?

*Crick:* What we assume is that the RNA takes some extended form and that the coding is done when the polypeptide chain is this extended form, and that it subsequently folds up. That is the basic answer. How this is done, and whether there are geometrical restrictions, we do not know, though we can think of ideas of getting away from some of the geometrical restrictions. The essential point is that you cannot carry an unlimited amount of information in a limited amount of RNA. I think this makes the influenza case particularly interesting because the amount of RNA is small and you would therefore think that there were not a great many protein molecules which you could code for.

*Hoyle:* Why do you select the molecular weight of RNA as over a million?

*Crick:* Perhaps I should say particle weight. We use that figure for illustration because that is approximately what is found. It could be that the RNA was in sub-units itself, and the sub-units could be identical or different. However, I would be cautious about arguments that

say, for example, that the length of an RNA chain in TMV is only 50 residues, as was once claimed.

*Hoyle:* In the case of influenza virus, we have a particle weight of 800 million and less than 1 per cent of RNA. That allows only one or two of your molecules of RNA in the whole virus. Now Dr. Burnet wants more than one genetic unit in influenza virus.

*Burnet:* I should feel not that it is actually necessary to have more than two proteins though.

*Crick:* I don't want to give the impression that one long piece of RNA is necessarily just for one protein. It may code for several proteins, but it can only be for a limited amount of distinct amino acid sequences. We know, for example, to take the DNA case, that in transforming principle it appears as if one molecule of DNA carries more than one genetic factor.

*Hoyle:* There is another point about influenza virus; by disintegrating the virus with ether we got out a soluble antigen which was apparently a ribonucleoprotein and there appeared to be about 70 particles of soluble antigen in each virus. Now this brings down the molecular weight of the RNA in a shocking manner.

*Crick:* You have got the same amount of RNA there even if you take the outside part off, haven't you?

*Hoyle:* We have 70 individual units to contend with. Does it mean that one of them has got all the RNA, and all the rest of them are protein units? Or is your estimate of the molecular weight of RNA too high? You were suggesting a molecular weight of a million.

*Crick:* What are the experimental results on the amount and distribution of RNA in these smaller units?

*Hoyle:* We only know that their overall RNA content is 5.3 per cent, and the particle size is of the order of 12  $\mu$ . That gives a molecular weight of 82,000 if you assume that there is one molecule of RNA in each particle of soluble antigen. Is it then that my particles of soluble antigen are fragments of a larger unit?

*Crick:* That is what one would guess, though there are other alternatives. But the first guess I would make would be that you have split up the RNA in the process of treatment.

*Knight:* Dr. Crick, in your idea, is the whole of the RNA concerned with the synthesis of one sub-unit?

*Crick:* Not necessarily. It is very possible that only half the RNA makes the sub-unit and that the remainder makes another protein which is important inside the cell but is not incorporated into the final virus. There is no reason why that should not be so.

*Knight:* Yes, but sticking with one particular virus and taking the sub-unit versus the whole virus, then the whole of the RNA is concerned mainly with the synthesis of the sub-unit, and we are faced with the problem of getting the sub-units lined up in the virus molecules.

*Crick:* Since you can, up to a point, line-up the sub-units in solution, there surely isn't any difficulty. The process of aggregation is one which you might reasonably call crystallization.

*Knight:* Do you think that this perhaps explains why we have some

so-called X-proteins and other things in tobacco plants infected with tobacco mosaic virus; that these are perhaps proteins which were synthesized but didn't get crystallized with a nucleic acid core?

*Crick:* There are a number of hypotheses, all of which have been enumerated before; either they are precursors, in the way that, say, chromotrypsinogen is a precursor of chromotrypsin, and it needs some active enzyme to turn X-protein into A-protein, for example, or alternatively, they are proteins which have been damaged—some degradation product, or the one which you have just mentioned. I don't think the ideas which we have throw much light on this problem; it is a matter of experiment to decide the exact relationships between the small units you find and the actual small units you get when you break up the finished virus.

*Dulbecco:* The idea of the crystallization of protein units is useful also to explain some experimental data, like phenotypic mixing of viruses. It is known that the protein coat of a phage particle produced in mixedly infected cells can have variable proportions of the adsorption properties of the two parents; moreover, the composition of the protein, as determined by this property, is independent of the genetic composition of the DNA of the particle. So that it seems that the phage coat is made up of units which are built independently and then assembled together with DNA.