PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

Macromolecular Order in Biology [and Discussion]

A. Klug, F. C. Frank and E. L. Thomas

Phil. Trans. R. Soc. Lond. A 1994 **348**, 167-178
doi: 10.1098/rsta.1994.0087

Macromolecular order in biology†

By A. Klug

MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, U.K.

Within a living cell there take place a large number and variety of biochemical processes, almost all of which involve large molecules, particularly proteins and nucleic acids. These macromolecules often interact to form ordered aggregates or specific complexes. A number of examples are discussed which show how different kinds of order develop on grounds of geometrical or physical necessity or for reasons of functional efficiency. Examples are taken from the structure and assembly of simple viruses and the higher-order organization of the DNA double helix in chromosomes.

1. Introduction: necessity and contingency

One of the many differences between biology and the physical sciences lies in the uniqueness of biological entities and the fact that these are the products of a long history. Max Delbrück has described (1949) how a mature physicist acquainting himself for the first time with the problems of biology might be puzzled by the fact that there are no absolute phenomena in biology. Everything is time-bound and space-bound.

Another outstanding feature of all organisms is their well nigh unlimited structural and dynamical complexity. Every biological system is so involved in multiple interactions and pathways, so rich in feedback devices, that one wonders whether a complete description is possible. As one goes to higher levels of organization not all the properties of the new entity are predictable consequences of the properties of the components, no more than chemistry is, in practice, predictable from physics, whatever is, in principle, contained in the Schrödinger equation.

However, in considering the complex structures one finds in biology, one must make the old distinction, going back to Plato's *Timaeus*, between contingency and necessity. In the inorganic world this distinction may be illustrated by the problem of the shape of snow crystals, a problem which exercised Kepler and which he discussed in his book *De nive sexangula* (Kepler 1966). Descartes, who had read this, explained the hexagonal form of the crystals as produced by the close packing in a plane of spherical water globules and we now know that this is, in essence, correct – the internal structure involves puckered hexagonal layers of molecules. The hexagonal symmetry is thus a necessity and follows from what Kepler called the demands of matter. But what of the external shape? Many different individual shapes are found and each is contingent on the particular history of its formation. How the symmetry of the external shape is maintained during growth remains an unsolved problem.

† This paper was produced from the author's disk by using the TEX typesetting system.

Phil. Trans. R. Soc. Lond. A (1994) **348**, 167–178
Printed in Great Britain

In the living world we might also ask the same question. Does a starfish have to have a five-fold symmetric shape? Since one can imagine a six-fold starfish, I would say it is not a necessity, but was arrived at early in evolution which fixed the form. Does a spherical shell of a virus particle have to have, as it does, five-fold symmetry? Here I would say yes, since, as I shall try to illustrate, it is a necessity once such shells are built as self-assembling objects constructed out of identical units. Thus the subject of my lecture is really that of order in biology on the molecular scale, and how much of it is imposed by physical necessity, purposeful as it may appear.

In the natural world around us, plants and animals often exhibit symmetry in their external forms. This geometrical regularity of living things has always been a source of wonder which over the centuries has continued to excite speculation about the invisible forces that guide the development of living organisms. From the *Facultas formatrix* of Johann Kepler in the early 16th century to the *Science of form* of D'Arcy Thompson in the early 20th century (Thompson 1942), there have been attempts to make general theories but these have always foundered for lack of detail or in unfruitful abstractions. Since then embryologists have sought for more objective explanations of the patterns on which living things develop. There is a good deal of progress here, but we are really only at the beginning of finding out how the genetic program is written and implemented to tell cells how to differentiate and where to go in a whole organism.

In this paper, therefore, I am going to limit myself to what, in another context, Galileo called the lesser phenomena – the kind of restriction which has in the history of science often proved more fruitful than grander questions about the universe at large. I shall deal with systems where the units of action are not cells but large biological molecules, macromolecules, that is, proteins or nucleic acids which form complexes or assemblies, and make patterns about which one can get precise information, by methods developed over the past 30 or 40 years. This is not to say that the patterns will all turn out to be regular in a geometrical sense, though indeed some are. As one proceeds to higher scales of structure, higher levels of organization, perfect regularity tends to get lost, but the assemblies are nevertheless ordered, according to rules we can try to grasp.

I shall use specific examples, almost all from problems I have been involved in over the years, and try to draw a more general picture out of these.

2. Specificity and self-assembly

Before one begins to discuss the arrangement of biological macromolecules it is necessary to say a word about the molecules themselves. The most important macromolecules are, of course, polymers, but built out of more than one element. Thus a protein consists of a linear polypeptide chain whose sequence is made up of choices among 20 different amino acids and a nucleic acid is a polymer made up of a sequence composed of four different nucleotides. Thus biological polymers are much more complex than synthetic polymers and of course even more so than small chemical compounds. The particular sequence within a biological polymer determines its properties, in other words, it carries the information content in that molecule. Thus, the order of amino acids in the polypeptide chain composing a protein carries implicit instructions for the folding up of that polymer into a

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

particular three-dimensional shape, and experiments show that in many cases this can be carried out a test tube. It is true that very large proteins sometimes need some assistance in establishing the right fold and there are classes of molecules called chaperones which can either assist the folding or, in the case of a misfolded molecule, kick it out of an energy minimum to enable the chain to fold correctly. Indeed, inside the cell there are further mechanisms for eliminating wrongly folded molecules.

Many large biological structures, such as the protein coats of virus particles, filaments of muscle, or the microtubules of nerve cells, consist of ordered arrays of protein subunits, but the number of different types of units is, in general, small. Some of these can be dissociated into their components, and these can in turn be reassembled *in vitro* under appropriate conditions to produce structures which are the same as, or very similar to, the original without the need of external instructions.

These large organized structures are formed by making use of the relatively weak, but specific, non-covalent interactions between molecules. Specificity is the key word here. An interaction between biological molecules must be highly selective in that the components must recognize each other to the exclusion of other potential partners. The typical energy of a 'bond', that is a contact between two such macromolecular surfaces, is of the order of 3–10 kcal (i.e. 5–15 kT), but the strength of the whole structure comes from the fact that there are many interactions in the assembly, and these are often cooperative. Moreover, I should stress that many of these assemblies (as I shall illustrate later) are polymorphic, and this polymorphism is vital to their dynamical role. This happens because the conformation of a protein molecule can change subtly but significantly during interactions. As a physicist would say, a protein molecule has many internal degrees of freedom and these are harnessed to develop changing interactions with other molecules of the same or different type. (It is, for example, the changes in the structure of the haemoglobin molecule when it binds oxygen that constitute the molecular aspect of breathing – it is a molecular lung, as Perutz puts it.)

Thus, the most significant feature of organized structures built in this way is that their design and stability can be determined completely, or almost completely, by the specific bonding properties of their constituent units. Thus, once the component parts are made they may assemble themselves – self-assemble – without a template or other specific external control, although in the case of certain more complex assemblies, such as T4 bacteriophage, various switches in state are necessary before the assembly can proceed to the next stage. A biological advantage of a self-assembly design for any large structure is that it can be completed specified by the genetic information required to direct the synthesis of the component molecules. The one-dimensional message contained in the DNA or RNA specifies the linear sequence of the protein for which it codes, and this in turn specifies how it will fold into a three-dimensional structure and consequently the shape and bonding properties of the protein. Economical use of the genetic information carried by the nucleic acid of a gene will require that identical copies of some basic molecule or group of molecules will be used to build any large structure. These large structures built of subunits can also be built efficiently and with great accuracy because of the possibility of checking; i.e. any bad or wrong copy can be rejected during the assembly process.

These general ideas just mentioned derive largely from investigations begun in

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

the late 1950s on the molecular structure of simple viruses. In the electron microscope, the virus particles are seen to have surprisingly simple shapes, cylindrical rods or spheres (really polyhedra). Biochemically they consist of long nucleic acid molecules, which carry the genetic information, encased in a shell of protein, a miniature parallel of spermatozoa.

3. The design and assembly of a simple virus: tobacco mosaic virus

The classical example of a rod-shaped virus is tobacco mosaic virus (TMV) whose rod shape results from its basic design, namely a regular helical array of identical protein subunits, in which framework is embedded a single molecule of RNA wound as a helix (figure 1, right). It is, of course, the RNA which carries the genetic information, i.e. the capacity to instruct the host cell to make many copies of the virus. This general picture was already complete by 1958, and it seemed easy to comprehend how a structure of this type might be built out of identical subunits: the subunits might assemble themselves by repeated identical interactions, like steps in a spiral staircase, enclosing the RNA as a corkscrew-like thread as the rod extends. In other words, the assembly might be likened to growth at a screw dislocation in a crystal. We now know that this simple picture of assembly is wrong in all essentials. The virus assembles in a much more complex way, for what, in hindsight, we can see to be good physical and biological reasons. The story of how my colleagues and I came to suspect that the simple scheme was deficient and how the path of assembly was found has been told elsewhere (Klug 1979) and I can only summarize the results here.

When the RNA and the protein of the virus are taken apart, the protein molecules alone under physiological conditions aggregate, not into a long helix, but into the 'disc' (figure 1*a*) – a two-layer cylindrical structure, each layer consisting of a ring of 17 molecules, compared with the $16\frac{1}{3}$ molecules present in each turn of the assembled helix. The disc can be crystallized, but because of the large molecular mass of the disc (600 000) the determination of its exact structure by X-ray methods posed formidable technical and analytical problems. These were overcome and after a dozen years it was posible to construct an atomic model showing the detailed structure of the protein subunit and how it interacts with its neighbours.

This study was pursued to the end, not merely for the sake of knowing the protein structure, but because we had earlier shown that the protein disc plays a crucial role in the assembly of the virus from its RNA and protein. The disc combines with a specific initiation tract on the single stranded viral RNA, and then dislocates to begin helix growth. Foreign RNAs which do not have this tract are rejected. When the sequence of bases of this initiation tract was determined it was seen that it could be folded so that the initial binding site is exposed at the apex of a 'hairpin' structure (figure 1*a*). Now the X-ray studies had shown that the two layers of the disc are so arranged as to leave a gap between them at the central hole of the disc, rather like a pair of 'jaws' waiting to 'bite' the RNA. So we were able to develop the picture of initiation, or nucleation as the physicist would say, of assembly shown in figure 1. The RNA hairpin loop inserts through the central hole of the disc, and the stem of the loop opens up and binds in the 'jaws' formed by the two layers of protein. The disc then dislocates into

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

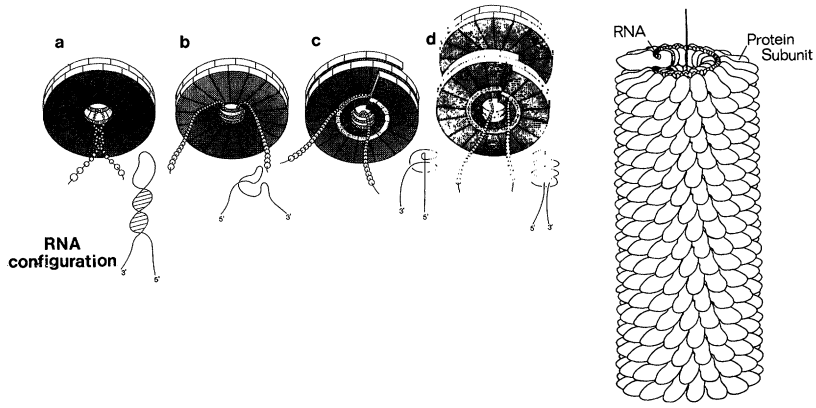


Figure 1. Nucleation of the assembly of tobacco mosaic virus (at right) begins with: (*a*) the insertion of the hairpin loop formed by the initiation region of the viral RNA into the central hole of the protein disc; (*b*) the loop intercalates between the two layers of subunits and binds around the first turn of the disc, opening up its base-paired stem as it does so; (*c*) some feature of the interaction causes the disc to dislocate into the helical lock-washer form; (*d*) this structural transformation closes the jaws made by the rings of subunits, trapping the viral RNA inside. The lock-washer-RNA complex provides the start of the helix. Additional discs then add rapidly to the nucleating complex. The special configuration of the RNA generated during the initial stages is perpetuated as the rod grows, by pulling further lengths of RNA up through the central hole, and the helix elongates to a minimum stable length.

a helical structure, entrapping the RNA, after which elongation proceeds by the addition of further discs, pulling up more RNA through the central hole. There is still some controversy surrounding this picture of growth after the initial stages, but there is no doubt about the role of the disc in initiating assembly.

The disc is thus an obligatory intermediate in the assembly of the virus which simultaneously fulfils both the physical requirement for nucleating the growth of the helical particle and the biological requirement for specific recognition of the viral RNA. A most intricate structural mechanism has been evolved to give the process an efficiency and purposefulness, whose basis we now understand. TMV is self-assembling, self-nucleating and self-checking. The general conclusion derived from the story of TMV assembly is that one must distinguish between the design of a structure and the construction process used to achieve it. In the TMV structure all protein subunits (except the few at the ends of the particle) make the same non-covalent contacts with each other, and this specific bonding pattern repeated many times leads to a symmetrical final structure. There is nothing in the design of the completed structure which gives a hint that different bonding patterns, and non-equivalent ones at that, are required during the process of assembly. This is unlike the case of the spherical viruses, where the design itself calls for departures from the precise identity of subunit packing. We turn to this next.

4. The architecture of spherical viruses: generalizations of icosahedral symmetry

The first attempt to understand how spherical viruses are built was made by Crick & Watson (1956). They argued that if identical units arranged themselves

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

regularly (or equivalently, as crystallographers say), as they do in TMV, but are designed to form a spherical rather than a cylindrical shell, then they would have to adopt the symmetry of one of the classical cubic point groups, namely tetrahedral, octahedral or icosahedral, in which cases the number of subunits would be restricted respectively to 12, 24 or 60; no regular arrangement of more than 60 is possible. The first experimental evidence from X-ray studies by Caspar on tomato bushy stunt virus, and by Finch and myself on turnip yellow mosaic virus and on polio virus, all pointed to icosahedral symmetry, a gratifying result, yet suggesting that there was somehow another special principle at work. Moreover, the first electron microscope and chemical data did not apparently agree with the X-ray ones: turnip yellow mosaic virus had 32 morphological units on its outside and contained about 150 protein units, whereas the X-ray results required 60, or a multiple of 60.

After much puzzling Caspar and I (Caspar & Klug 1962) were able to reconcile these apparently divergent results (figure 2). We asked how could one build a spherical shell out of large numbers of units, abandoning the assumption that the symmetry had to be perfect. Molecular structures are, after all, not built to conform to exact mathematical concepts but rather to satisfy the condition that the system be in a minimum energy condition. Dropping the requirement of mathematical equivalence, but allowing identical units to be quasi-equivalently related, either through slight adjustments of the bonding contacts between units or through a measure of internal flexibility, we showed that the optimum design for such a shell required icosahedral symmetry, since the distortions in the specific bonds are then minimized. Certain, but not all, multiples of 60 subunits could be accommodated in such designs, which Caspar and I enumerated. Many of the hemi-spherical, 'geodesic' domes designed by the architect Buckminster Fuller are built according to a similar geometry, but whereas these have to be assembled by following a fairly elaborate code, the virus shell, because of the flexibility of its units, can, so to speak, build itself.

In the 1960s John Finch and I and some of our colleagues, using electron microscopy combined with X-ray diffraction, verified that many spherical viruses – indeed every one we investigated – have their subunits arranged according to one or other of the predicted patterns, a result which holds to the present day (Klug 1983). It turns out that viruses with apparently quite different external morphologies (e.g. figure 3*b*, *c*) belong to the same structural class. Since the construction of the icosahedron in the cube is said to be the crown of Greek geometry (it is the last theorem in Euclid), I like to think it would have delighted Plato to know that fundamental forms lay beneath the variety of appearances. To quote F. M. Cornford (Cornford 1967) on Pythagoreanism, 'the key to intelligible order lies in the notion of limited quantity defining unlimited quantity, as the key to harmony lies in a few definite intervals marked out in the indefinite range of sound'.

5. Chromatin: an ordered hierachy of foldings

The work on viruses has given results not only of intrinsic interest, but has also influenced general ideas on biological structure by focusing attention on the interactions between large macromolecules. Equally important, however, but not

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

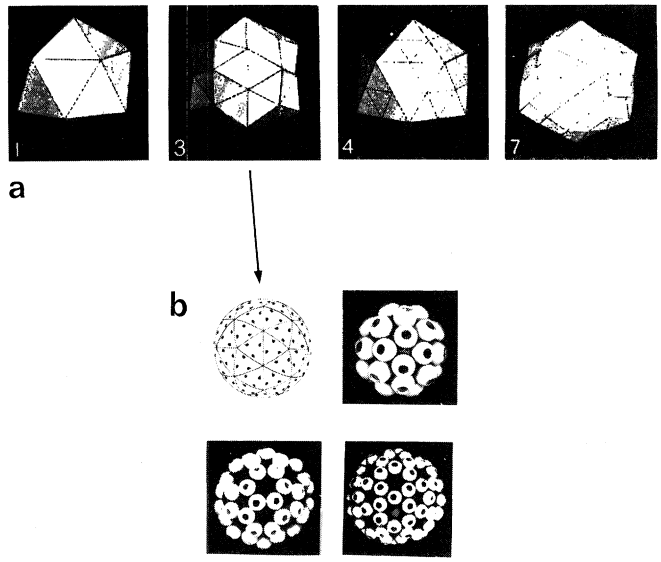


Figure 2. (*a*) Icosahedral surface lattices, labelled by their T numbers [6]. (*b*) The $T = 3$ icosahedral lattice and its possible clustering patterns. Reading clockwise from top left are shown: the lattice showing the $3 \times 60 = 180$ 'molecules' in general positions; the patterns of morphological units produced by clustering the molecules in 12 pentamers and 20 hexamers; in 60 trimers; and in 90 dimers. The photographs are of models and show only one side of the 'virus' particle viewed down a two-fold axis.

directly relevant here, the difficulties in tackling such large assemblies have led to the development of methods and techniques which could be applied to other systems besides viruses. Electron microscopy combined with three-dimensional image reconstruction, supplemented wherever possible by X-ray studies on wet materials, has provided what are now generally accepted models of the structural organization of a large number of biological systems, such as haemocyanin, microtubules, filaments of muscle, and sickle cell haemoglobin, to name just a few diverse applications (Crowther & Klug 1975).

A more recent example of this approach is that of chromatin. Chromatin is the name given to the chromosomal material when extracted from the nucleus of a cell. It consists mainly of DNA tightly associated with an equal weight of a small set of rather basic proteins called histones, as well as many other minor components which can vary from cell to cell. We took up the study of chromatin in Cambridge in the early 1970s, when protein chemists had shown that there were only five main types of histones.

The DNA in a chromosome of an animal or plant is a single molecule, stretching several centimetres if laid out straight. It must be highly folded to make the compact structure one sees in a chromosome (see figure 4). At the same time DNA is organized into separate functional units – the genes. All cells, whatever their type, contain the same total DNA complement of the organism but derive their special character by expressing certain genes and not others. The major problem to understand is what controls whether a particular gene is transcribed into RNA, enabling the cell ultimately to synthesize the protein product encoded in it. The way the chromatin is folded influences such 'gene expression'.

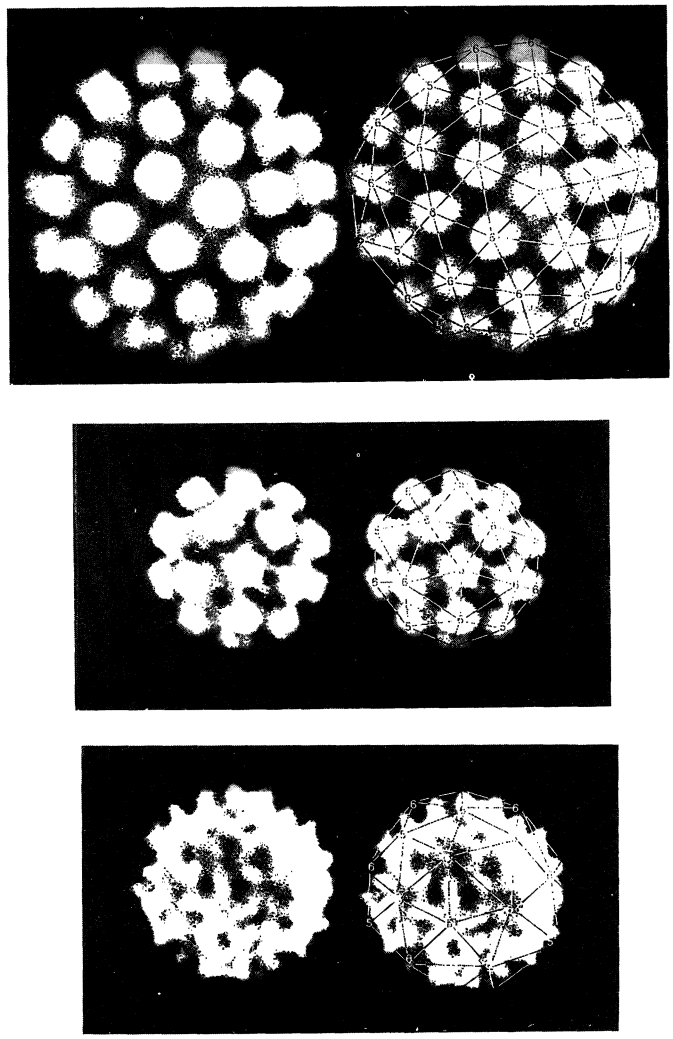


Figure 3. Three-dimensional image reconstructions from electron micrographs of some spherical viruses. Alongside are shown the underlying icosahedral surface lattices, with the five-fold and six-fold vertexes marked. (*a*) Human wart virus (about 550 Å in diameter; class $T = 7$); (*b*) turnip yellow mosaic virus (about 300 Å in diameter); (*c*) tomato bushy stunt virus; (*b*) and (*c*) both belong to the same class, $T = 3$, which has 180 units, organized around 12 strict five-fold axes and 20 local six-fold axes. However, the units are clustered at the surface quite differently in the two cases (cf. figure 2*b*): in (*b*) they are grouped into pentamers and hexamers around the five-fold and six-fold positions to form 32 morphological units, whereas in (*c*) they are clustered into 90 dimers about two-fold positions.

We have, therefore, sought to discover the structure of chromatin. When Roger Kornberg came to Cambridge in 1972, we began using X-ray diffraction to analyse the relationships between histones and DNA. These X-ray studies showed that the native structure could readily re-form if the four histones were kept together in pairs, but not once they had been taken apart. This, together with chemical studies on the histones, eventually led Kornberg to the discovery that chromatin

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

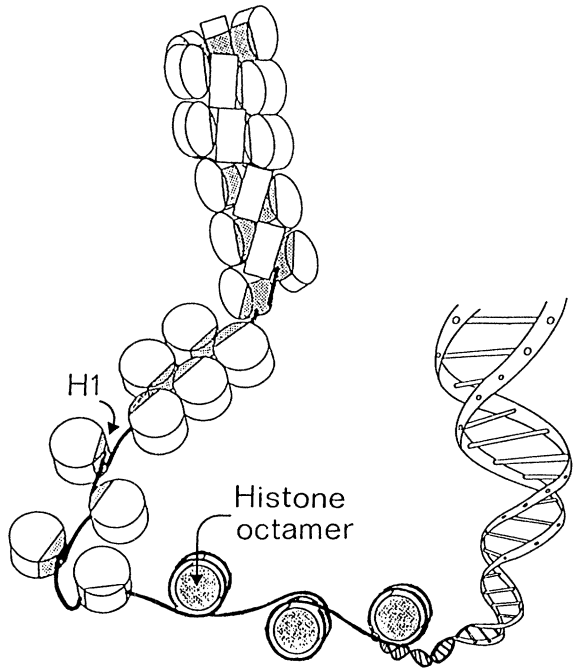


Figure 4. Schematic diagram showing the first two levels of folding of the DNA double helix (at right) in chromatin. In the first level (bottom) the DNA is wound as two superhelical turns around a flat core or spool made of eight histone protein molecules, forming structural units called nucleosomes. The next level of folding is mediated by another type of histone molecule, called H1, which attaches to the exit and entry points of the DNA on the nucleosome. The H1 molecules then aggregate into a helical polymer, as shown at left, culminating in a solenoidal arrangement of nucleosomes (top left) in which the H1 polymer runs along the centre. The drawing is idealized: the solenoidal structure is ordered, not perfectly regular.

consists of a succession of structural subunits, called nucleosomes, containing two each of the four main histones, combined with 200 base-pairs of DNA (Kornberg 1977). We were later able to crystallize and solve the structure of a form of the nucleosome called a core particle, which had been trimmed with an enzyme. This crystallization showed that almost all the DNA in the nucleus is organized at a fine structural level in a highly regular manner.

Over the years the work has continued in two main directions. We have worked down, to examine the internal structure of the nucleosome (Richmond *et al.* 1984), and worked up, to try to understand how the filament of a nucleosome is further folded in the nucleus, giving chromatin its next higher order of structure (Widom & Klug 1985).

The outcome of a wide range of structural and chemical studies is a model of the nucleosome in which the DNA double helix is wrapped in two superhelical turns around a spool (or core) formed by the eight histone molecules (figure 4). This DNA is sealed at the point where it enters and leaves the spool by the fifth type of histone, called H1. The H1 also mediates the coiling of the nucleosome filament into a helical or 'solenoidal' structure to form the next level of structure.

Indeed, the wrapping of DNA on a protein core is only the first step in a hierachical series of foldings which eventually results in a 10 000-fold linear con-

densation of the DNA into the compact chromosomes seen at the metaphase stage of cell division. The way in which this happens is not understood in detail, but there is good evidence, from the work of U Laemmli, that lengths of the solenoidal filament, containing perhaps 50–100 000 bases of DNA, are folded back to form loops 'tied' at their ends by special proteins of a 'scaffold' (Marsden & Laemmli 1979). These loops might correspond to units of gene expression or transcription: when a given transcriptional unit becomes active, the solenoid in a loop would unwind concomitantly with chemical modification of the nucleosomes. This picture of loops is also compatible with the further compaction of chromatin into the condensed metaphase chromosomes seen in cell division: the 'ties' could interact together, either directly or through other non-histone proteins, to give a helical array along the axis of the chromosome.

6. Active chromatin: specific recognition of DNA

The hierarchy of levels described above for packaging the DNA double helix applies to the bulk of the chromatin, containing the inert genes. A current question concerns the structure of regions of active chromatin containing genes poised for expression or actually being expressed. Genes are activated and begin transcribing the DNA into RNA through the binding of regulatory proteins ('transcription factors') to a control region ('promoter') of a gene, which event, as it were, switches on the enzymatic machinery. There are a number of different designs used by proteins to bind specifically to particular short sequences of DNA (Klug 1993) but I have space to describe only one particularly interesting example from our Laboratory. Some years ago I chose to work on a particular gene that codes for 5S-RNA, a component of the ribosome, since the ovaries of immature frogs contain large amounts of a transcription factor for this gene. My colleagues and I purified this protein, called TFIIIA, and showed that it has a remarkable repeating structure (Klug & Rhodes 1987; Rhodes & Klug 1993) Each structural unit, or domain, consists of a small loop of about 30 amino acids folded around a zinc ion (figure 5). We have called these units 'DNA-binding fingers', and have further shown that each finger binds to, and thereby recognizes the sequence on, about half a turn of the DNA double helix. The fingers are, so to speak, reading heads for identifying a specific control region on the DNA.

As we predicted, these DNA-binding fingers have turned up in many other regulatory proteins. TFIIIA thus represents a novel class of proteins. Their modular design offers a large number of combinatorial possibilities for specifically recognizing, and combining with, many different DNA sequences. It is likely that this design evolved by repeated duplications and mutations of an ancestral gene which coded for a single stable structural protein unit which bound to DNA. Presumably it was selected during the course of evolution because of its functional efficiency.

7. Concluding remarks

With these various examples, I hope I have been able to convey to you some glimpses of how molecular biology has enabled us to understand how ordered biological systems are built, and how they function, in terms of their three-dimensional structures and of the specific interactions between the molecules

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

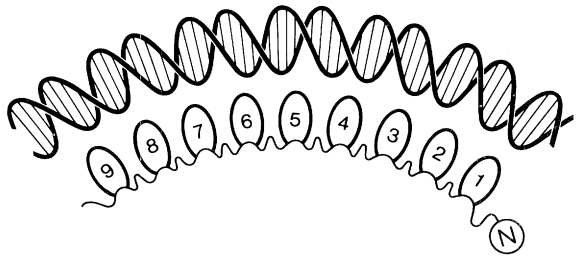


Figure 5. Schematic drawing of the structural features of the protein TFIIIA and its interaction with DNA. The bulk of the protein is organized into nine small domains, tandemly repeated. Each domain, or 'zinc finger', consists of a length of about 30 amino acids folded compactly about a zinc ion, and is of a suitable size to bind into the major groove of half a double helical turn of DNA. The fingers all have a common structural framework, but derive their chemical distinctiveness from variations in a set of amino acids and residues located at the tip and on one side of the 'finger'. In this way the protein 'reads' the varying sequence of base pairs along the DNA, and a specific interaction takes place when the two are correctly matched.

that comprise them. We also have seen that some of the patterns produced by the forces of evolution – as in cylindrical and spherical viruses – are shaped by the same spatial restrictions that apply to mathematical objects. Others have been freely chosen to give structures and processes which have no counterpart in the macroscopic natural world.

References

Caspar, D. L. D. & Klug, A. 1962 Physical principles in the construction of regular viruses. *Cold Spring Harb. Symp. Quant. Biol.* **27**, 1–24.

Cornford, F. M. 1967 *The unwritten philosophy and other essays*, p. 24. Cambridge University Press.

Crick, F. H. C. & Watson, J. D. 1956 Structure of small viruses. *Nature, Lond.* **117**, 473–475.

Crowther, R. A & Klug, A. 1985 Structural analysis of macromolecular assemblies by image reconstruction from electron micrographs. *A. Rev. Biochem.* **44**, 161–182.

Delbrück, M. 1949 A physicist looks at biology. *Trans. Connecticut Acad. Sci.* **38**, 173–190. (Reprinted in *Phage and the origins of molecular biology* (ed. J. Cairns *et al.*), *Cold Spring Harb. Quant. Biol.*, 1966.)

Kepler, J. 1966 *The six-cornered snowflake* (ed./transl. C. Hardie). Oxford: Clarendon Press.

Klug, A. 1979 The assembly of tobacco mosaic virus: structure and specificity. *The Harvey Lectures* **74**, 141–162.

Klug, A. 1983 Architectural design of spherical viruses. *Nature, Lond.* **303**, 378–379.

Klug, A. 1993 Protein dresigns for the specific recognition of DNA. *Gene.* **135**, 83–92.

Klug, A. & Rhodes, D. 1987 'Zinc fingers', a novel protein motif for nucleic acid recognition. *Trends Biochem. Sci.* **12**, 464–469.

Kornberg, R. D. 1977 Structure of chromatin. *A. Rev. Biochem.* **46**, 931–954.

Marsden, M. P. F. & Laemmli, U. K. 1979 Metaphase chromosome structure: evidence for a radial loop model. *Cell* **17**, 849–858.

Rhodes, D. & Klug, A. 1993 Zinc fingers. *Sci. Am.* **268**, 56–65.

Richmond, R. J., Finch, J. T., Rushton, B., Rhodes, D. & Klug, A. 1984 Structure of the nucleosome core particle at 7 Å resolution. *Nature, Lond.* **311**, 532–537.

Thompson, D'A. W. 1942 *On growth and form.* Cambridge University Press.

Widom, J. & Klug, A. 1985 Structure of the 300 Å chromatin filament: X-ray diffraction from oriented samples. *Cell* **43**, 207–213.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

Discussion

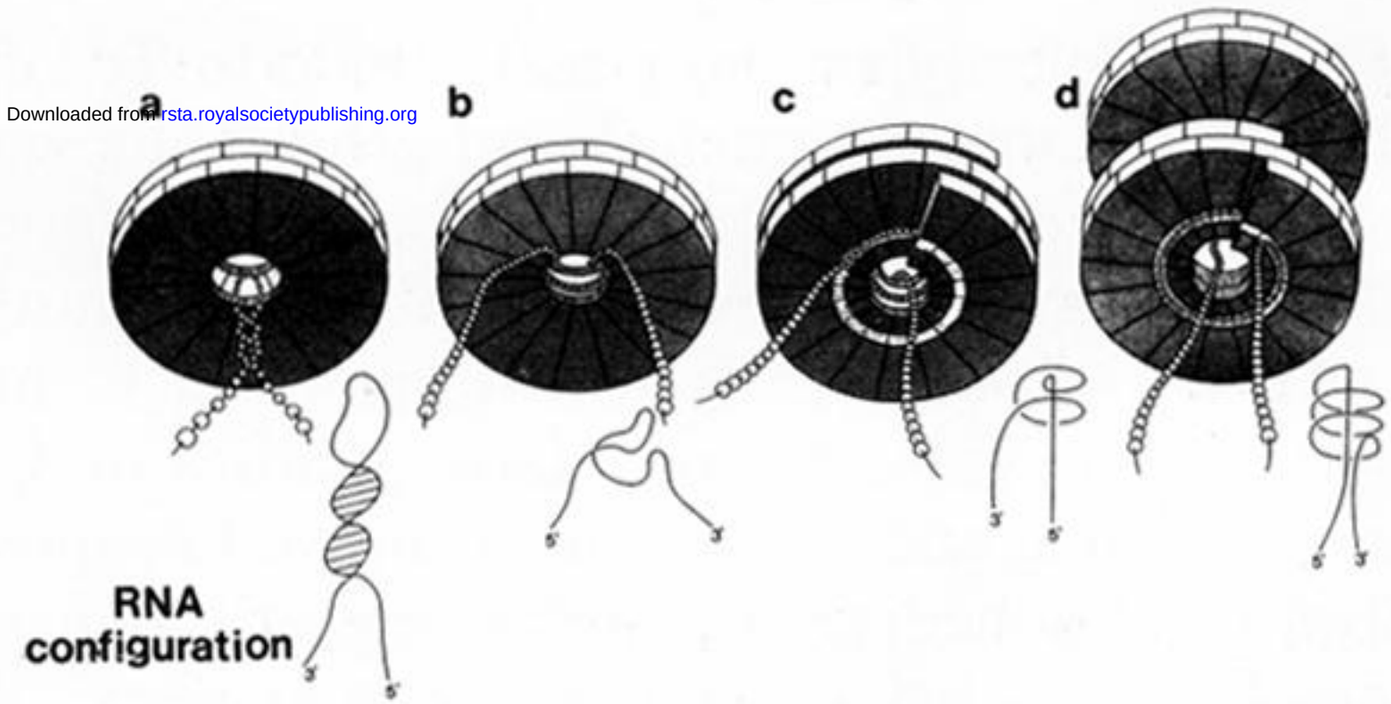
F. C. Frank (University of Bristol, U.K.). I understood you to say that in the assembly of TMV, protein molecules are first pre-assembled into rings of 17 which are threaded on to the hair-pin of RNA chain feeding up the axial hole: but they have to be selected and arranged to be in the right order to match a sequence on the next stretch of RNA chain: I cannot see how that can be a speedy process.

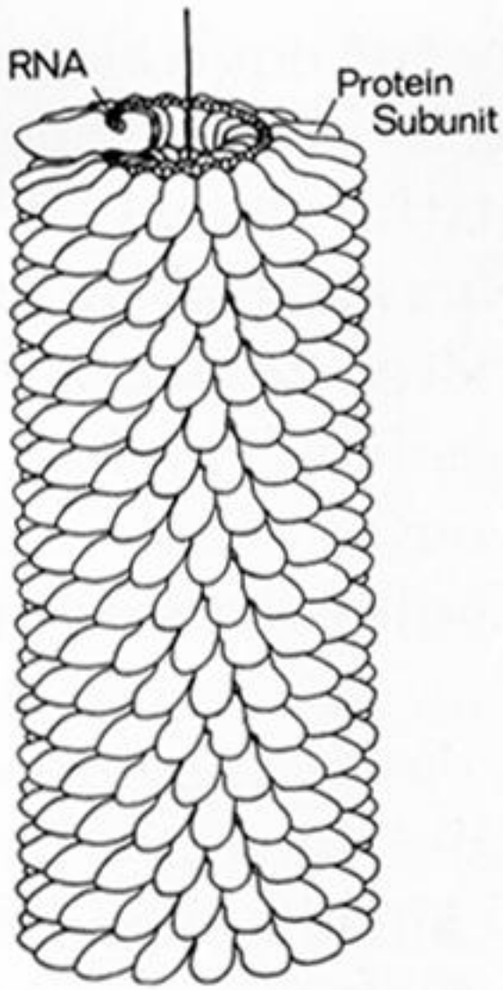
A. Klug. You are correct in saying that there has to be some matching of the sequence on the RNA chain to the sites on the protein subunits to achieve specificity. However, this specific recognition of RNA by protein is achieved at the stage of nucleation of the virus assembly, and here only a few hundred nucleotides out of a total of 6000 are involved, and even among these few hundred there is a particular subset which are involved at the very first stages: there is a stem-loop structure which has guanine residues in every third position on the RNA chain which fit into sites on successive protein subunits in the disc (remember that each protein subunit accommodates three nucleotides so that repeats in every third position of the RNA sequence match the protein structure). In fact there is a similar such pattern about 50 nucleotides further along the initiation sequence of the RNA and these presumably are used in incorporating the second ring of 17 protein subunits which dislocates into the 16 1/3 units per turn of the virus helix. After about three discs worth of protein have interacted with the RNA chain a stable complex is formed and then there is no requirement for specific sequences in the RNA, since now the cooperativity of the protein–protein interaction ensures that the helix structure is propagated. In the effect a random sequence of RNA can be incorporated into the growing particle once the nucleation stage is paused.

E. L. Thomas. Can you comment on work in the early 1970s by W. Harris concerning the role of disclinations in the structure of surface crystals?

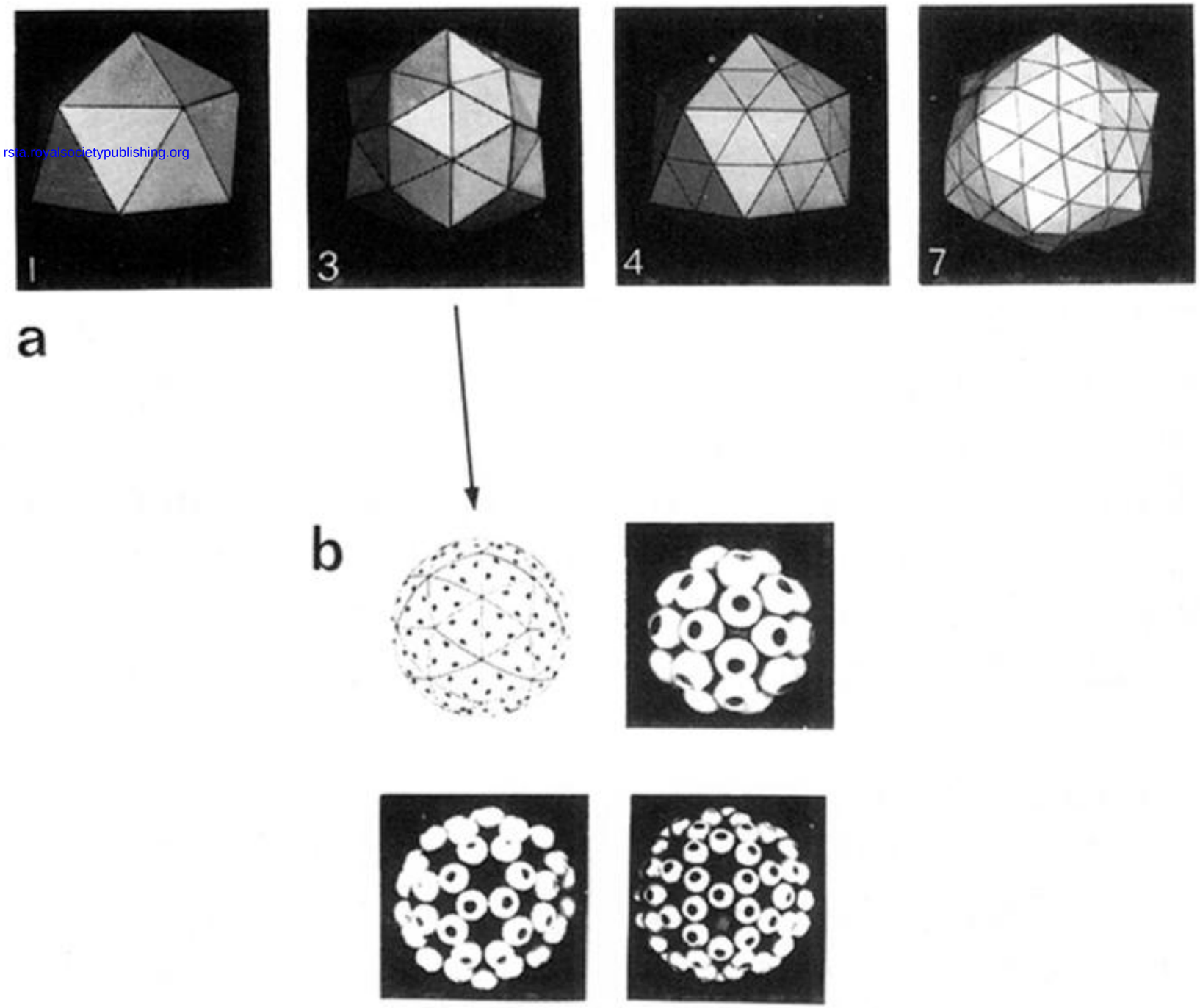
A. Klug. Harris pointed out that some of the discontinuities in patterns found for the arrangement of macromolecular units on the outer surface of certain bacterial cell walls were more accurately described as ‘disclinations’, rather than as ‘dislocations’, the term in common usage. Harris also suggested that disclinations played a part during the initial stages of the assembly of tobacco mosaic virus, but here, since the transition from disc to protohelix cannot be pictured in detail, there is little point in abandoning the general term ‘dislocation’.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

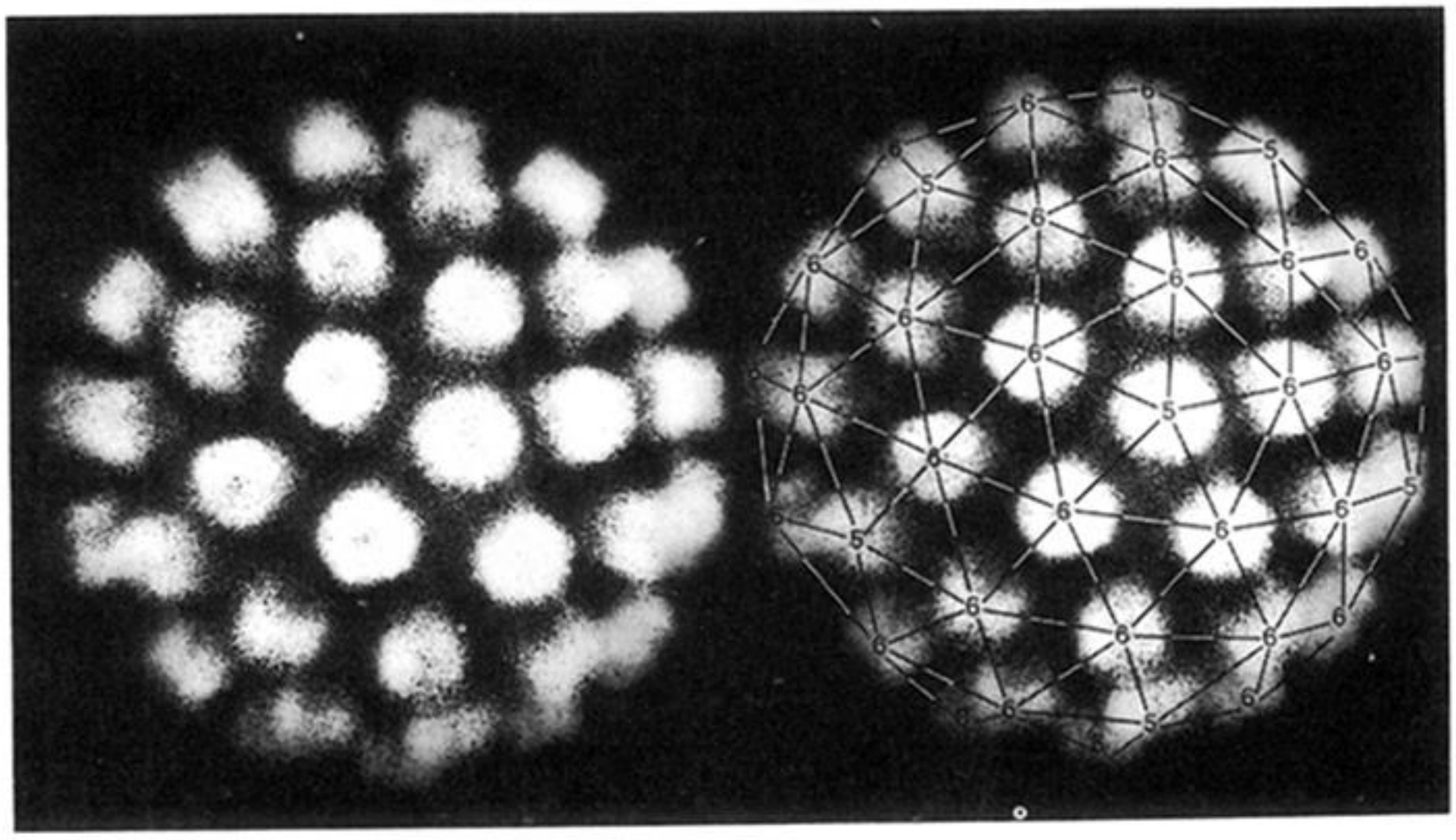


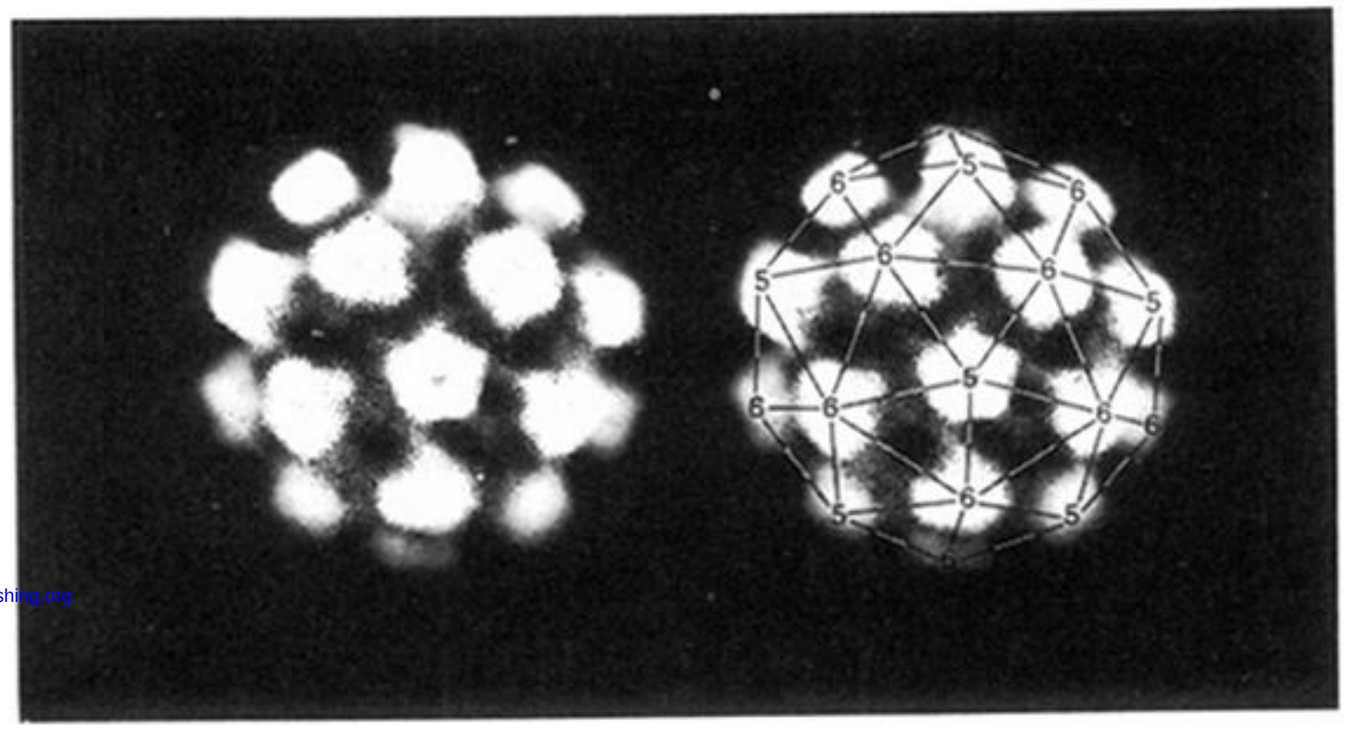


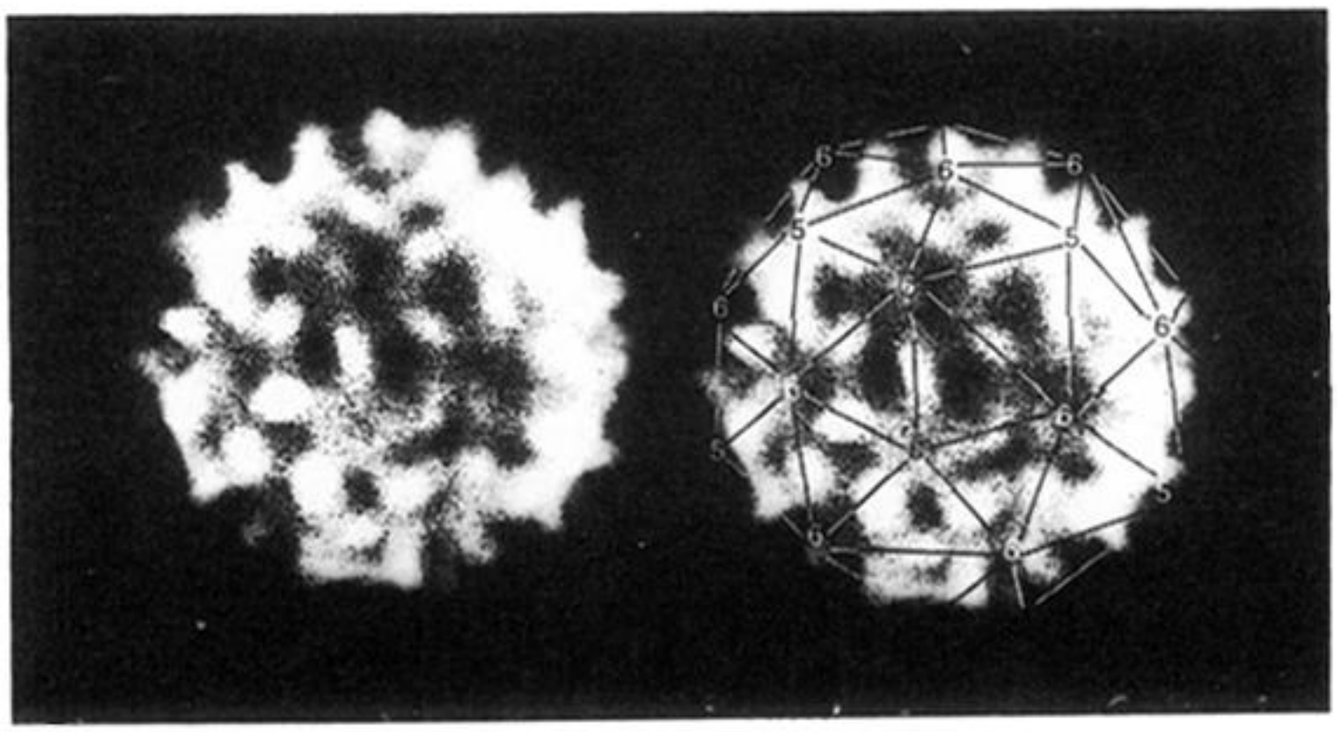
igure 1. Nucleation of the assembly of tobacco mosaic virus (at right) begins with: (*a*) the sertion of the hairpin loop formed by the initiation region of the viral RNA into the central ole of the protein disc; (*b*) the loop intercalates between the two layers of subunits and binds round the first turn of the disc, opening up its base-paired stem as it does so; (*c*) some feature of he interaction causes the disc to dislocate into the helical lock-washer form; (*d*) this structural ansformation closes the jaws made by the rings of subunits, trapping the viral RNA inside. he lock-washer-RNA complex provides the start of the helix. Additional discs then add rapidly the nucleating complex. The special configuration of the RNA generated during the initial ages is perpetuated as the rod grows, by pulling further lengths of RNA up through the central ole, and the helix elongates to a minimum stable length.



igure 2. (*a*) Icosahedral surface lattices, labelled by their T numbers [6]. (*b*) The $T = 3$ osahedral lattice and its possible clustering patterns. Reading clockwise from top left are nown: the lattice showing the $3 \times 60 = 180$ 'molecules' in general positions; the patterns of orphological units produced by clustering the molecules in 12 pentamers and 20 hexamers; 60 trimers; and in 90 dimers. The photographs are of models and show only one side of the irus' particle viewed down a two-fold axis.







igure 3. Three-dimensional image reconstructions from electron micrographs of some spherical iruses. Alongside are shown the underlying icosahedral surface lattices, with the five-fold and ix-fold vertexes marked. (*a*) Human wart virus (about 550 Å in diameter; class $T = 7$); (*b*) urnip yellow mosaic virus (about 300 Å in diameter); (*c*) tomato bushy stunt virus; (*b*) and (*c*) oth belong to the same class, $T = 3$, which has 180 units, organized around 12 strict five-fold xes and 20 local six-fold axes. However, the units are clustered at the surface quite differently n the two cases (cf. figure 2*b*): in (*b*) they are grouped into pentamers and hexamers around the ve-fold and six-fold positions to form 32 morphological units, whereas in (*c*) they are clustered nto 90 dimers about two-fold positions.