On the Structure of the Scaffolding Core of Bacteriophage T4 and Its Role in Head Length Determination

Bonnie Berger* Gunnar W. Hoest† James R. Paulson‡
Peter W. Shor§

Abstract

The scaffolding core in bacteriophages is a temporary structure that plays a major role in determining the shape of the protein shell that encapsulates the viral DNA. In the currently accepted structure for the scaffolding core in bacteriophage T4, there is a symmetry mismatch between the protein shell, which has five-fold symmetry, and the scaffolding core, which is believed to consist of six helical chains. The analysis of T4 giant head data that was used to determine the six helical chains made an implicit assumption about the manner in which giant heads flatten during the preparation for electron microscopy, but reexamination of the experimental data shows that this assumption may be incorrect. Reanalysis of the data shows that it could be consistent with six, eight, or ten helical chains. The ten-helix core model is particularly attractive because it suggests a Vernier mechanism which is able to explain the process of length determination in giant head mutants of T4.

Subject classification: viruses and bacteriophages

Keywords: bacteriophage T4 / scaffolding proteins / self-assembly / morphogenesis / Vernier mechanism

*Author to whom all correspondence should be addressed, at Room 2-389, Mathematics Department and Lab. for Computer Science, Massachusetts Institute of Technology, Cambridge, MA 02139, U.S.A.

†Lab. for Computer Science, MIT, Cambridge, MA 02139 U.S.A.

‡Chemistry Department, University of Wisconsin, Oshkosh, WI 54901, U.S.A.

§AT&T Bell Laboratories, Murray Hill, NJ 07974, U.S.A.

1 Introduction

Bacteriophage T4 is one of the most extensively studied viruses. The protein capsid of this virus consists of a head and a tail which form separately and then join together. The head is a protein shell which contains the viral DNA in the mature form of the virus. Like many viruses, T4 has a shell that is essentially icosahedral. However, the T4 head is atypical in that it is elongated along a five-fold symmetry axis by the insertion of extra hexamers (Moody, 1965; Aebi *et al.*, 1974; Branton & Klug, 1975; Baschong *et al.*, 1988). Despite extensive study, the mechanism for the shape determination of this T4 protein shell is still quite poorly understood. Part of the difficulty is the relative complexity of the morphogenetic mechanism of T4. This mechanism is believed to be known for several viruses, *i.e.,* the $T = 3$ plant viruses, the picornaviruses, and tobacco mosaic virus, but these are comparatively simple structures (Harrison *et al.*, 1978; Hogle *et al.*, 1986; Silva & Rossmann, 1987; Berger *et al.*, 1994). Whereas these viruses have only one to three proteins that play a role in shape determination, the shape-determining mechanism of the T4 head appears to involve approximately ten proteins, including three proteins (two coat proteins gp23 and gp24 and the portal protein gp20) which remain in the mature virus coat, and approximately seven proteins (gp22, gp21, IPI, IPII, IPIII, gp67 and gp68), which are temporarily incorporated into an internal scaffolding core that is subsequently destroyed during the maturation process (Kellenberger, 1990). The focus of this paper is on the structure of the scaffolding core of T4 and its role in the length determination of the shell.

The structure of the T4 scaffolding core was first studied by Paulson & Laemmli (1977), who concluded that the scaffolding core contains six helical chains wrapped around a hollow core. These chains are believed to be composed of the protein gp22, a predominant component of the core which comprises 38% of the mass of the core according to current estimates of molecular masses and copy numbers for the core proteins (Black *et al.*, 1995). The protein gp22 will spontaneously polymerize into long filaments (van Driel, 1980b), and is thus a good candidate for the principal component of these helices.

A core consisting of six helical chains is surprising, because it implies a symmetry mismatch between the virus shell, which has five-fold rotational symmetry, and the scaffolding core, which could have three-fold or six-fold symmetry but not five-fold symmetry. Symmetry mismatches are known to occur elsewhere in viruses. For example, many bacteriophages, including T4, have a portal complex with twelve-fold rotational symmetry occupying a vertex of the shell with five-fold rotational symmetry (Bazinet & King, 1985). It has been suggested that this symmetry mismatch might allow the portal complex to rotate with respect to the shell, and thus could play a role in DNA packaging (Hendrix, 1978). In the case of bacteriophage T4, it has been suggested that the portal complex could also serve as a connector between shell and scaffold, the place at which pro-head assembly is initiated, and the site at which the tail (which has six-fold symmetry and would thus match a core structure with six helical chains) is later attached (Paulson & Laemmli, 1977). This would explain the observation that mutations in gene 22 can lead to many abnormally attached tails (Paulson *et al.*, 1976) and would be in keeping with the proposal that gene 22 is involved in initiation of head formation (Laemmli *et al.*, 1970; Showe & Black, 1973).

On the other hand, the core is definitely known to aid in the assembly of the bacteriophage shell (Paulson *et al.*, 1976; Kellenberger, 1990). It seems strange that this core should not then have the same symmetries as the structure for which it appears to act as a template. While six-fold symmetry of the core would match the symmetry of the portal complex, the shell is a larger and more complicated structure and the formation of the complete five-fold symmetric shell on a six-fold symmetric core would appear a more difficult task than the connection of a five- or ten-fold symmetric core to the twelve-fold symmetric portal complex.

This research was undertaken to discover whether an alternative structure, without the symmetry mismatch, would be consistent with the data. A possible flaw was found in Paulson and Laemmli's analysis. This analysis depended critically on an assumption about the manner in which the core and shell flatten on the grid during preparation for electron microscopy. Reanalysis of the original data showed that the simple model of flattening upon which the analysis depends is inconsistent with the data. Using a more general model for the flattening process, it was discovered that structures for the core consisting of six, eight, or ten helices, are all indeed plausible.

The alternative ten-helix structure also suggests a hypothesis for the length-determining mechanism of T4 which agrees with much of the experimental evidence. This hypothesis, which is a two-dimensional Vernier mechanism, is incomplete in that it does not explain how the width of the T4 shell is determined and needs to be augmented with some other mechanism to fully explain how the length is determined. However, it does explain the observed behavior of the length determination process in the giant head mutants of T4 (Lane *et al.*, 1990; Lane & Eiserling, 1990) and could plausibly form a major component of the length determination mechanism in wild type virus.

1.1 Background on previous work

The currently accepted T4 scaffolding core structure was proposed by Paulson & Laemmli (1977) based on studies of elongated aberrant T4 head particles called polyheads. Several types of polyheads were studied, but the most important ones were those induced by treating wild-type T4-infected cells with L-canavanine.

In this paper, canavanine-induced polyheads and other particles of similar structure will be referred to as *giant heads,* because evidence indicates that they have the same structure as an elongated version of a true pro-head, or head precursor.

The term *polyheads* will be reserved for other types of elongated structures of the T4 head proteins.

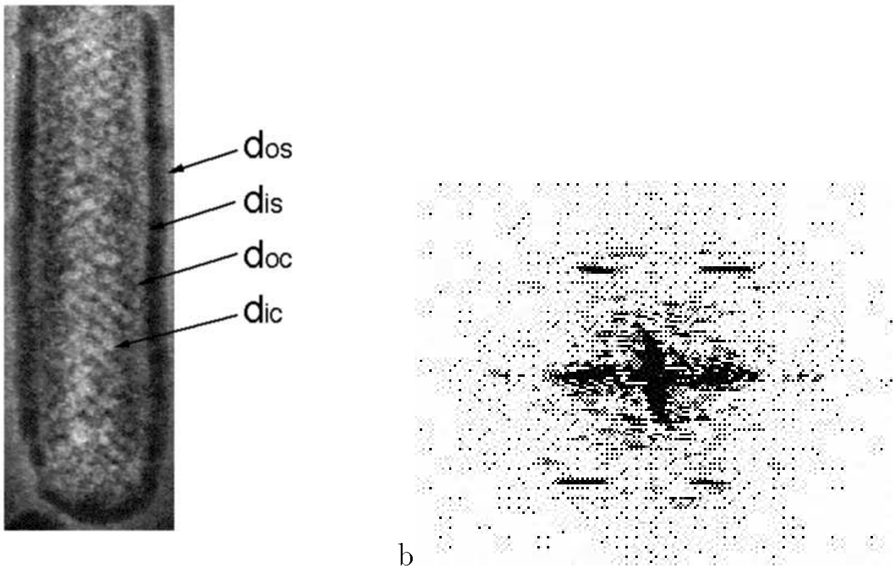
Paulson and Laemmli found a way to preserve and stain these particles so that the interior core could be observed in the electron microscope. Diagonal stripes were seen on the core, and these were interpreted as helical chains. From their analysis, Paulson and Laemmli concluded that the core in giant heads is composed of six helical chains wrapped around a hollow center (or lumen).

Since the protein gp22 is the predominant component of the core, and is known to polymerize into ribbons, it is believed that the observed helical chains in the core are composed primarily of gp22 (Engel *et al.*, 1982). The protein gp22 comprises approximately 38% of the mass of the core, according to current estimates of molecular masses and copy numbers for the core proteins (Black *et al.*, 1995). The filaments resulting from gp22 polymerization are approximately 95Å wide and 25Å thick, with subunits spaced approximately 15Å apart along the filaments. The gp22 molecules themselves have dimensions approximately $15 \times 25 \times 95$Å (van Driel, 1980a; Engel & van Driel, 1981; Engel *et al.*, 1982; Mesyanzhinov *et al.*, 1990).

Paulson and Laemmli used optical transforms and computer Fourier transforms on images of the T4 giant heads to measure the structural parameters of these helical chains (Figure 1). The four strong, symmetrically placed reflections observed at an approximate axial spacing of 115Å were found to correlate with the core striations seen in the micrographs (Paulson & Laemmli, 1977). In general, the transform of a projection of unflattened, concentric, vertically aligned helices of radius r is confined to horizontal layer lines in Fourier space. Suppose there are N such helices in the core. Then the magnitude of the transform on the first layer line is proportional to $J_N(2\pi Rr)$, where J_N is the Nth order Bessel function, R is the horizontal coordinate of the transform, and r is the effective radius of the core (that is, the radius at which the mass of the core helices is concentrated) (Moody, 1967). The horizontal coordinate that maximizes this expression is denoted by R_{max}. The four horizontal diffraction spots along the layer line at 115Å axial spacing were thus interpreted to correspond to the first peaks on the first layer line in the transform of the helical chains in the core, and R_{max} values could be measured directly by reading off the horizontal coordinates of these spots.

Note that R_{max} very roughly corresponds to the inverse of the lateral distance between helices in the core. The number of core helices was estimated by first using the measured values for R_{max} to estimate the effective radius of giant head cores for various numbers of core helices, and then deducing the number of core helices by assuming that the effective radius of the giant heads would be between the inner and outer radii of the core.

The original analysis was complicated by the fact that most of the giant heads used had undergone some degree of flattening. This would not only tend to increase the measured core and shell diameters, but would also decrease the observed values



a b

Figure 1: (a) Micrograph of a T4 canavanine giant head; d_{os}, d_{is}, d_{oc}, and d_{ic} indicate the outer edge of the shell, the inner edge of the shell, the outer edge of the core, and the inner edge of the core, respectively. (b) Picture of the computed Fourier transform of this giant head. The four horizontal oval-shaped diffraction spots correspond to the Fourier transform of the core helices. The upper right and lower left ovals correspond to the top surface of the core, and the other two ovals to the bottom surface. The horizontal distance from the y-axis to each of these ovals is defined to be R_{max}, and the vertical distance from the x-axis to each of these ovals is the reciprocal of the axial repeat.

for R_{max} (Moody, 1967). Paulson and Laemmli assumed that the degree of flattening of the core equals that of the shell. They also used an average value for R_{max} in their calculations. For seemingly reasonable estimates of the effective radius, the data obtained were found to be consistent with anywhere between five and seven helical chains. Since a phase measurement showed that the number of chains was even, it was concluded that the T4 core contained six helical chains.

1.2 Motivation for this work

We suggest that the analysis described above, which led to the conclusion that the core contains six helices, may have been flawed in two respects. First, there is a significant, roughly linear relationship between the shell diameter and $1/R_{max}$, a relationship which was not addressed in the original analysis, in the sense that average values were used for these parameters. This may have led to incorrect assumptions about the relationship between the helices in the core, their degree of flattening, and the corresponding Fourier transforms. As a consequence, the computed effective radii obtained, both in the cases of flattened and unflattened giant heads, may have been incorrect.

Secondly, all of Paulson and Laemmli's measurements were taken on cylindrical giant heads which had been more or less flattened by the preparation process, as discussed above. The flattening process may have been more complicated than was assumed in their analysis. This analysis was based on the assumption that the helices in the core flatten independently of the shell when lying on the grid; that is, when flattened, the core keeps its structure in that the bonding mechanisms that connect the core helices to each other are not stretched or broken. Thus, it was assumed that the distance between the core helical chains measured circumferentially around the core would be preserved. It is possible, however, that the core helices are connected to specific sites on the inside of the shell. Experiments consistent with this hypothesis are detailed by Doherty (1982a, 1982b) , where mutation studies were used to show that specific interactions between the core protein gp22 and the coat protein gp23 are involved in the length determination mechanism.

The geometry of flattened concentric cylinders is such that both of the bonding interactions discussed above cannot be preserved; either the connections between neighboring core helices are broken or stretched, or the connections from the core helices to the shell are broken or distorted. (See Figure 2.) It is thus plausible that, instead of the structure of the core being preserved, the helices in the core would remain fastened to the inner surface of the shell of the giant head. This would cause an increase in the lateral spacing of the core helices and a decrease in the observed value for R_{max} (Figure 2).

A direct consequence of the increased lateral spacing of the core helices would be that, in the case of a flattened core, the outer core radius is no longer an upper bound

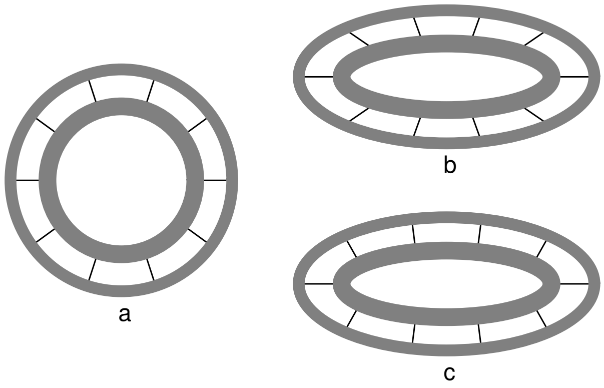


Figure 2: Schematic representation of the flattening of a giant head. The shell and the core form cylinders which are shown here in cross-section, although the dimensions and shapes are not drawn accurately. (a) The unflattened giant head. Ten evenly-spaced connections join the shell and the core. (b) The flattened giant head, assuming (as in Paulson & Laemmli (1977)) that the shell and the core both kept their structures. In this drawing, the binding sites are equally spaced in both the shell and the core after flattening. Note that some of the connections have now lengthened considerably and are no longer perpendicular to the shell and the core. (c) The flattened giant head, assuming as in this paper that the shell and the connections kept their structures. Note that in the flat section of the core the binding sites are spaced considerably farther apart than in (b).

on the calculated value of the effective radius of the core helices. The circumferential distance between the core helices on the flattened part of the core would grow larger, so that they would appear to have come from a larger core. In the extreme situation, where both the core and shell are completely flattened and the core is maximally stretched, the calculated value of the effective radius of the core could become the radius of the shell. In the remainder of this paper, this calculated value will be called the *apparent effective radius*.

From phase measurements on the Fourier transform of the giant heads, Paulson and Laemmli also concluded that there are an even number of helices in the core. This phase measurement, in contrast, would seem to be unaffected by the flattening process. If there were an odd number of helices in the core, it is difficult to see how the flattening process could produce phase measurements clustered around $0°$ as observed (Paulson & Laemmli, 1977). If we accept this phase measurement, there must then be an even number of helices in the core.

2 Materials and methods

For the data analysis, electron micrographs of T4 canavanine induced giant heads were obtained from the batch used in Paulson and Laemmli's original experiments. The micrographs were made at an original magnification of approximately 50000×. For a description of their preparation, see Paulson & Laemmli (1977). Thirty-eight giant heads suitable for analysis were chosen by inspection of photographic enlargements of the micrographs. In particular, only heads with at least one discernible cap were included in the analysis, since these are the ones that are presumed to have the same structure as elongated proheads (Paulson & Laemmli, 1977).

The image analysis and data collection for this paper were carried out at the Rosenstiel Basic Medical Sciences Research Center at Brandeis University in Waltham, MA, USA. The micrographs were scanned using an Eikonix model #1412 linear CCD based densitometer device. The CCD is 4096 pixels long. The light source was a Gordon S45 Plannar light source.

The giant head images were rotated and aligned horizontally so that the inner and outer core and shell diameters could be measured. The alignment of the giant heads was done by eye, and we believe the values of the measured diameters were accurate to within 5%. Some giant heads had variations in some or all of these parameters along their lengths. If any of these parameters could not be determined unambiguously, either due to an unclear image or due to excessive variance, the giant head in question was excluded from the analysis. When there were only slight variations in some of these parameters, average values were used. Polyheads with a diameter of less than 655Å, the diameter of an unflattened giant head, were not used.

Fourier transforms of the images were computed using an FFT program based on the FOURT.FOR subroutine, an implementation of the Cooley-Tukey FFT algorithm. The images and their transforms were displayed on a Lexidata 90 display device. For each giant head, the horizontal coordinates of each of the four, strong diffraction spots were recorded by eye. The spots were usually between $0.001\mathring{A}^{-1}$ and $0.002\mathring{A}^{-1}$ wide, and $0.0002\mathring{A}^{-1}$ thick. Generally, the centers of the spots were picked for the measurements. We believe these measurements were accurate to within 5%. In the cases where the spots were unclear, the giant head in question was discarded from the analysis. Four giant heads with anomalously high values of R_{max} were also discarded from the analysis.

Coordinate systems for the Fourier transforms were determined as follows. The wide, oval shapes of the four strong diffraction spots allowed for unambiguous determination of the orientation of the coordinate axes in Fourier space. The x-axis was determined by constructing a line through the origin, from which the four strong diffraction spots were equally distant. This distance corresponds to the inverse of the axial repeat of the core helices. The y-axis was determined by constructing a line perpendicular to the ordinate axis passing through the origin.

This construction seems reasonable, since the axial repeat (or pitch) of the helices is likely to remain uniform, unaffected by the flattening of the giant head. However, the lateral spacing of the helices, and hence the measured R_{max}, is indeed affected by the flattening, as discussed above. The degree of flattening of the bottom part of the core adjacent to the grid will tend to be larger than that of the topmost part (Moody, 1967). The bottom part of the core corresponds to the upper left and lower right quadrants in Fourier space, while the topmost part of the core corresponds to the remaining two quadrants. There are thus two values of R_{max}, one corresponding to the top part of the core, and one corresponding to the bottom part. The R_{max} value for the bottom part will tend to be less than that for the upper part of the core. While the top and bottom parts of the core gave similar results, in the rest of this paper, R_{max} will be taken to mean the value corresponding to the bottom part of the core, since this value was observed to be more sensitive to differences in the degree of flattening, and hence will give a more accurate picture of the effects of flattening.

The final analysis of the data was carried out using Matlab v4.2c on a cluster of Sun SPARC5 workstations.

3 Results

3.1 Effects of flattening on the core structure

In order to investigate the relationship between the degree of flattening of the giant heads and R_{max}, a plot was made of $1/R_{max}$ versus the outer shell diameter D_S, and a least-squares regression analysis was used to construct a best-fit straight line for the data (Figure 3). The resulting best-fit line is $1/R_{max} = 0.286D_S - 64.1\mathring{A}$, with a correlation coefficient R^2 of 0.51, indicating a statistically significant, near-linear relationship between the two parameters. The endpoints of the line are $(660\mathring{A}, 125\mathring{A})$ and $(860\mathring{A}, 182\mathring{A})$. The large difference between the $1/R_{max}$ values at these endpoints indicates that the effect of flattening on the Fourier transforms of the giant heads is quite significant. Therefore, using an average value for $1/R_{max}$ may not be sufficiently accurate, as it does not take this variation into account.

An indication of the effect of flattening on the Fourier transforms of the giant heads is the ratio of the $1/R_{max}$ value for a flattened giant head to the corresponding value for an unflattened one. In the original analysis of Paulson & Laemmli (1977), it was assumed that the helices in the core flatten independently of any interactions of the core with the shell, and thus that the apparent effective radius of the core remains the same after flattening. In other words, it was assumed that the distortion of the core helices during the flattening process was essentially the same as it would be for a cylinder made of helical chains that was not contained inside a shell. If this

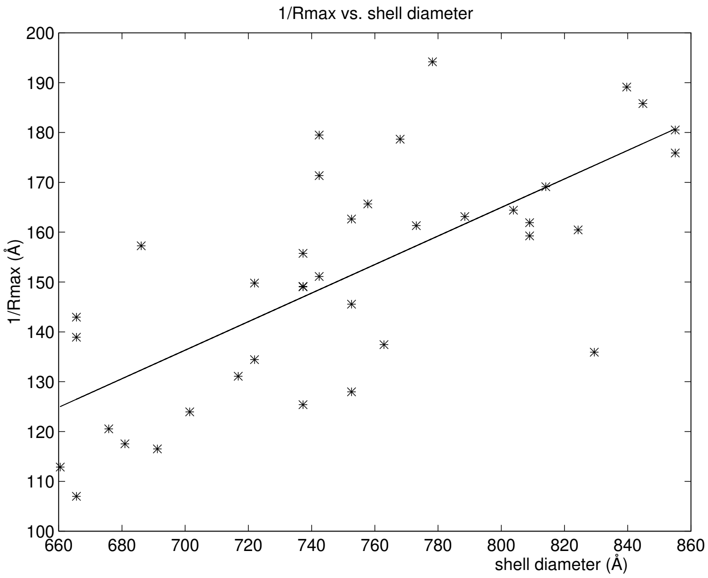


Figure 3: The effect of flattening on the structure of the T4 core helices. Plot of $1/R_{max}$ versus shell diameter (in Ångstroms), together with the least-squares best-fit straight line for the data.

were indeed correct, the ratio of $1/R_{max}$ for a maximally flattened giant head to a minimally flattened one would be at most 1.25 for any number of helices greater than one (Moody, 1967). In contrast, the ratio of the $1/R_{max}$ value at the rightmost endpoint of the least-squares line to that corresponding to a shell diameter of $660\AA$, gives a value of 1.46 on the $1/R_{max}$ values of the most to least flattened giant heads in the data set. The use of the shell diameter as a measure of flattening seems reasonable, since the exact shell diameter of an unflattened T4 giant head is known to be $655\AA$ (Paulson & Laemmli, 1977).

A reasonable explanation for this large ratio of values of $1/R_{max}$ could be the alternative model of flattening proposed in section 1.2; that is, when the giant head is flattened, the core helices are pulled apart laterally due to interactions with the shell, causing a decrease in the measured value of R_{max}, and hence an increase in the apparent effective radius of the helices.

Plots of $1/R_{max}$ versus two different estimates for the core diameter were also made, namely for the outer core diameter and inner shell diameter, and least-squares regression analyses were used to construct best-fit straight lines for the data (Figure 4). Two different estimates were used since it is not entirely clear where the outer boundary of the core lies. In particular, the outer boundary of the main mass of the core and the inner boundary of the shell are usually separated by a gap of width $10 - 50\AA$ between the main mass and the inner shell, which in many of the

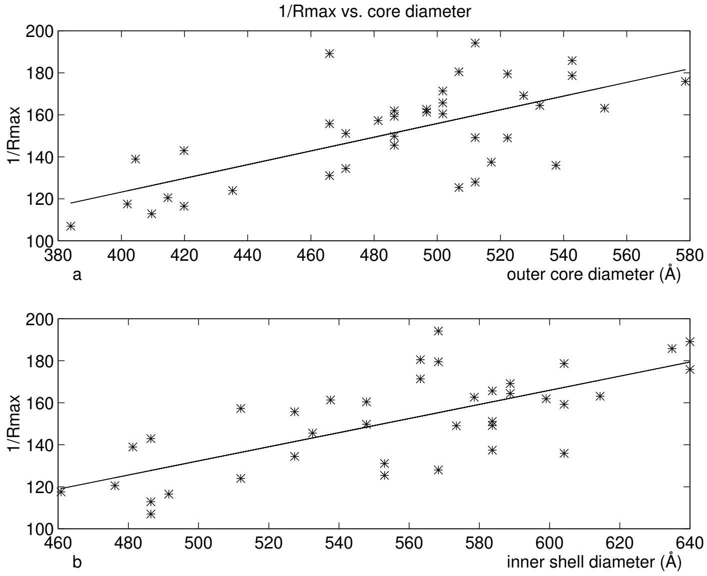


Figure 4: The effect of flattening on the structure of the T4 core helices. (a) Plot of $1/R_{max}$ versus core diameter (in Ångstroms), together with the least-squares best-fit straight line for the data. This plot uses the diameter of the main core mass. (b) Plot of $1/R_{max}$ versus inner shell diameter (in Ångstroms), together with the least-squares best-fit straight line for the data. This plot uses the diameter of the inside of the shell.

micrographs appeared not to be completely empty; indeed it appears that the helical chains span this gap (DeRosier, 1995). The first estimate uses the outer diameter of the main mass of the core, the other the diameter of the inside of the shell. The resulting best-fit lines are $1/R_{max} = 0.326D_{OC} - 7.26\mathring{A}$, with a correlation coefficient R^2 of 0.45, and $1/R_{max} = 0.336D_{IS} - 35.68\mathring{A}$, with an R^2 of 0.50, indicating a reasonably significant, near-linear relationship between the two parameters. The endpoints of the lines are $(380\mathring{A}, 117\mathring{A})$ and $(580\mathring{A}, 182\mathring{A})$ and $(460\mathring{A}, 117\mathring{A})$ and $(640\mathring{A}, 182\mathring{A})$, respectively. The data analysis presented here thus shows that there is a clear relationship between $1/R_{max}$ and the core and shell diameters.

In order to further test the hypothesis of independent flattening of the core helices used in Paulson and Laemmli's analysis, a series of theoretical plots of $1/R_{max}$ versus the outer shell diameter were generated for this independent flattening model, assuming various effective radii r of unflattened helices within a reasonable range ($150\mathring{A}$–$200\mathring{A}$). As described above, since all the heads with caps are believed to have the same diameter when unflattened, the observed shell diameters provide a way of computing a measure of the degree of flattening of the giant heads. The model of independent flattening assumes that the degree of flattening of the core, and hence of the core helices, is the same as that of the shell. One of the plots is shown in

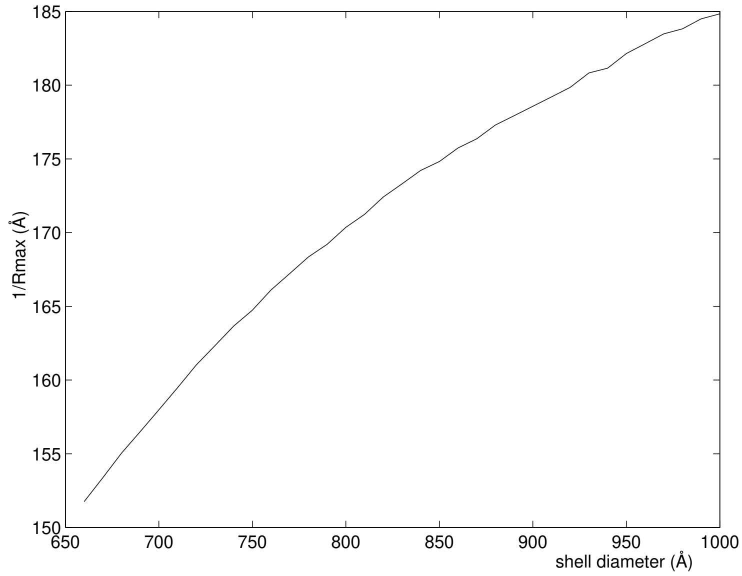


Figure 5: Theoretical plot of $1/R_{max}$ versus shell diameter for the six-helix model, assuming the hypothesis of independent flattening of the helices in the core and in this case an effective radius of $180\mathring{A}$.

Figure 5.

This model of flattening is not consistent with experimental data. As discussed above, it predicts a maximum ratio of the values $1/R_{max}$ for flattened helices to $1/R_{max}$ for circular helices of 1.25. For the observed range of values of the shell diameter (roughly 655Å to 860Å), this ratio in the theoretical plot decreases to 1.17. The experimental data give a corresponding ratio of 1.46. Moreover, while the slopes of the theoretical plots average approximately 0.1 (declining with increasing shell diameter), the experimental data give a slope of 0.3. These results seem to support the idea that the apparent effective radius of the helices in the core increases with an increased degree of flattening; that is, the lateral distance between helices increases when the observed shell diameter increases.

In producing the theoretical plots, the core was modeled as a set of N concentric, continuous helices of infinite length wrapped around a hollow core. Suppose the degree of flattening is δ, where $0 \leq \delta \leq 1$. It is perhaps easiest to conceptualize the shape of the flattened helices by considering a cross-section of the hollow core. In the case $\delta = 0$, this cross-section is a circle of radius r, in accordance with the analysis developed by Moody (1967). In the case $\delta = 1$, this cross-section is nothing but a line of length πr, also in accordance with Moody's analysis. In the general case, the cross-section was modeled as two parallel lines of length s, separated by a distance $2p$, connected by a semi-circle of radius p at each end, where $p = r(1 - \delta)$. The circumference of the cross-section is constrained by the

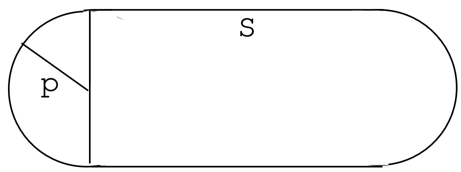


Figure 6: Schematic cross-section of the core as used in modeling of the core under the hypothesis of independent flattening of the helices. The semi-circles at the ends have radius p, and the length of the middle rectangle is s, where p and s are constrained by $2\pi r = 2\pi p + 2s$.

requirement that $2\pi r = 2\pi p + 2s$ (Figure 6).

When computing Fourier transforms of the helices, the analysis can be simplified by using the fact that the radial coordinate of the first peak on the Nth layer line in the transform of a single helix is identical to the radial coordinate of the first peak on the first layer line in the transform of N helices (Moody, 1967). This radial coordinate is R_{max}. Hence, it suffices to consider the transform of a single, continuous helix of radius r. Thus, the Fourier transforms of single continuous helices and their corresponding R_{max} values were computed for different degrees of flattening for the six-helix model (Figure 5).

3.2 The structure of T4 giant head cores

Particular attention was paid to the giant heads with observed outer shell diameter near 655$\mathring{A}$, since it seems reasonable to assume that these giant heads are nearly completely unflattened, and hence the analysis developed by Cochran, Crick and Vand (1952) for the Fourier transforms of completely unflattened helices is applicable. The best-fit line for the giant head data in Figure 3 evaluated at a shell diameter of 655$\mathring{A}$, corresponding to the left endpoint of the line, gives a value for $1/R_{max}$ of 123$\mathring{A}$. Using this value, the apparent effective radius of the core was computed for a range of different values for N, where N is the number of helices in the core. The results are found in the first column of Table 1. Note that only even values for N are included, since it is assumed that the number of helices in the giant head core must be even, as was shown by the phase measurements of Paulson and Laemmli (1977).

In order to establish upper bounds on the effective radius of the giant heads, plots were made of two outer core diameter measures, D_{OC} and D_{IS}, representing the outer diameter of the main mass of the core and the inner diameter of the shell, respectively, versus the shell diameter, D_S (Figure 7). Both these estimates were used since in the micrographs of heads showing a relatively small amount of stain,

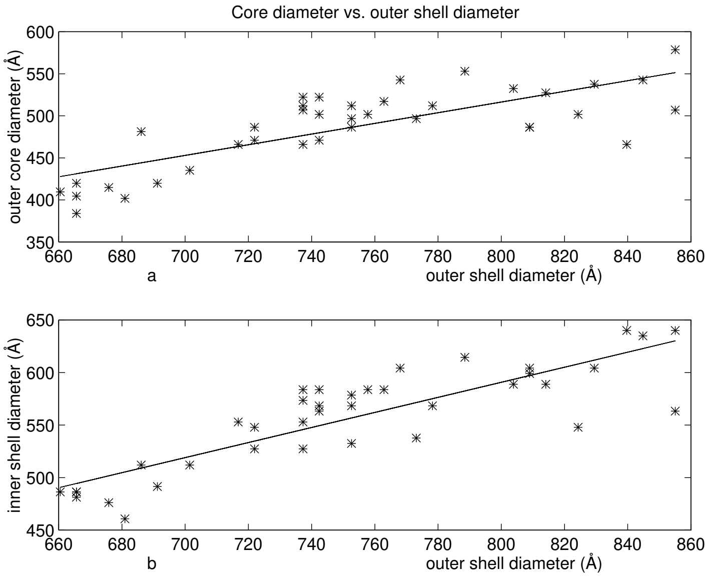


Figure 7: Plot of the outer core diameter and inner shell diameter versus the outer shell diameter. The least squares fit straight lines for these curves is given.

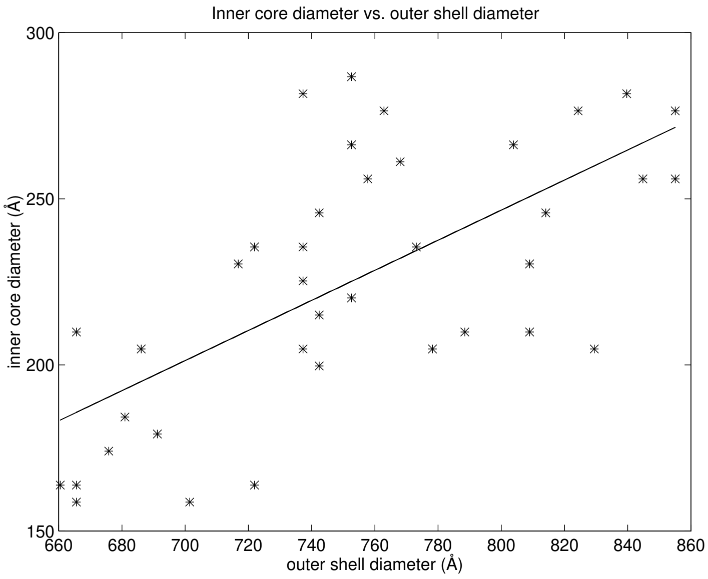


Figure 8: Plot of the inner core diameter versus the outer shell diameter. The least squares fit straight line for this curves is given.

N	circular core effective radius	flattened core effective radius
4	104	120
6	147	177
8	189	234
10	230	291
12	272	349

Table 1: Predicted effective radii of giant head core helices based on positions of diffraction maxima.

some structure was visible between the inner shell diameter and the outer diameter of the main mass of the core, and this may contribute to the Fourier transform. Thus, the inner shell diameter may be a better bound on the effective radius of the core helices than the outer core diameter. There is a significant relationship between both of these core diameters and the shell diameter. A lower bound on the effective radius was established from a plot of the inner core diameter, D_{IC}, versus the shell diameter, D_S (Figure 8.

Least-squares regression analysis was used to construct best-fit straight lines for each of the data sets. The resulting best-fit lines are $D_{IC} = 0.45D_S - 115.6\mathring{A}$, with a correlation coefficient R^2 of .45, $D_{OC} = 0.63D_S + 9.0\mathring{A}$, with a R^2 value of .60, and $D_{IS} = 0.72D_S + 17.4\mathring{A}$, with an R^2 value of .73, respectively. Hence, for an unflattened giant head, assuming the shell diameter is $D_S = 655$Å, the outer and inner core diameters will be approximately $D_{OC} = 422\mathring{A}$ or $D_{IS} = 489\mathring{A}$ and $D_{IC} = 179\mathring{A}$, respectively, corresponding to core radii of $211\mathring{A}$, $245\mathring{A}$, and $90\mathring{A}$. It then follows that N = 4, 6, 8, and 10 helical chains correspond to values for the effective radii of the helical chains between the inner core diameter and inner shell diameter (see Table 1, column 1). The four helix model seems implausible, however, since an effective radius of only 104Å would not appear to be consistent with visible striations being visible to nearly the outer edge of the core. If an effective radius of $211\mathring{A}$ is used for the outer core diameter, then ten helices is only possible with an 8% inaccuracy. Allowing this inaccuracy is plausible, since the core may not be completely unflattened even for giant heads with diameters near $655\mathring{A}$.

A similar analysis was carried out on giant heads with observed shell diameters around $860\mathring{A}$, since these giant heads were the most flattened of the ones observed. Hence, the analysis developed by Moody (1967) for the Fourier transforms of completely flattened helices was used. The best-fit line for the giant head data evaluated at a shell diameter of $860\mathring{A}$ gives a corresponding $1/R_{max}$ value of $182\mathring{A}$ (see Figure 3). Using this value, the apparent effective radius of the core was computed for the same values of N as above. The results are tabulated in the second column of Table 1.

As discussed in section 1.2 above, the problem of establishing an upper bound on the apparent effective radius of the core helices is even more complicated in the case of flattened giant heads, and it will depend on the model of flattening chosen. Using the alternative model of flattening proposed in section 1.2, a conservative upper bound must be applied, namely that of the unflattened shell radius, *i.e.*, $327\AA$. This analysis leaves possible values of $N = 6$, 8, and 10 for the number of helical chains.

It is believed that the helices constitute the outermost part of the core (Paulson & Laemmli, 1977). The data indicate an outer core radius of $211\AA$ or possibly $245\AA$ and an inner core radius of $90\AA$, which gives a core thickness of 121 to 155Å for unflattened helices. Previous studies have determined that negatively stained gp22 filaments have a width of approximately $95\AA$ and a thickness of approximately $25\AA$ (Engel *et al.*, 1982).

Depending on the orientation of the core helices, the contribution of these helices to the width of the core should lie between their width and their thickness, *i.e.*, between 25Å and 95Å. If the outer edge of the filaments lies at the outer edge of the core at 211Å radius, then the inner edge would lie between 116Å and 186Å, and the average diameter of the filaments (which might be a good estimate of the effective diameter) lies between 163Å and 199Å. This is still nearly consistent with the data of 147Å for the effective radius of a six-helix core, especially if the effective radius is further towards the center than the average position of the filaments. If the core helices extend to the inner edge of the shell at 245Å, then the inner edge of the filaments would lie between 150Å and 220Å, and the average diameter of the filaments falls between 197Å and 233Å. This case would be consistent with 10 helices.

A filtration was made of the core from one of the micrographs of a giant head (see Figure 9). This filtration appears to show that there are more than four helices in the core. This is to be expected, as there are visible striations on the polyhead cores that appear to extend most of the way to the edge of the core, and this is inconsistent with an effective radius for the chains very close to the inner edge of the core, as would be required for the four helix structure. Depending on how sharply the helical chains turn, and whether there is some helical structure which has been mostly obscured by stain between the outer edge of the main mass of the core and the inner shell, the filtration appears consistent with six, eight, or ten helices in the core.

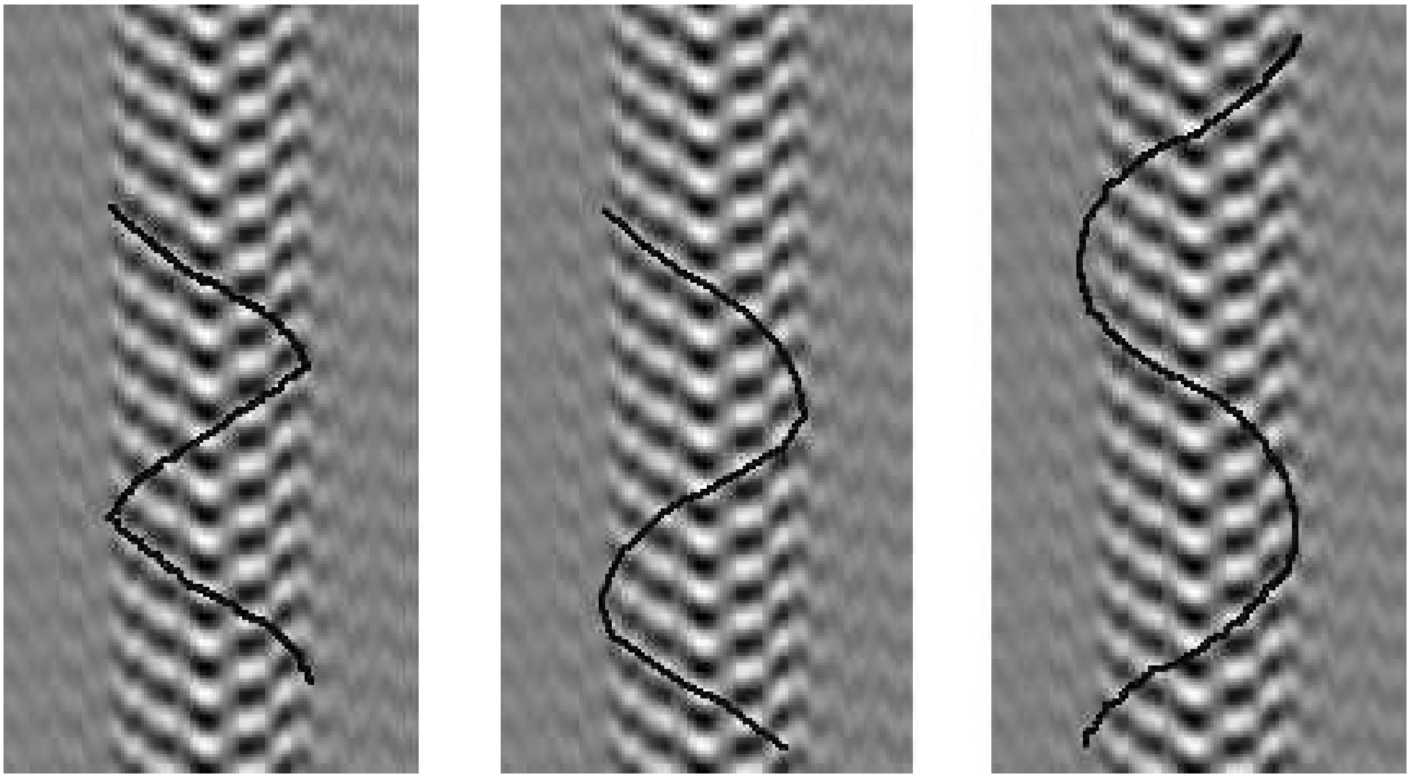


Figure 9: A filtration of a micrograph of a giant T4 head showing the core striations. Curves are drawn to show how one of the core helices might fit in a six-helix (left), an eight-helix (center), and a ten-helix (right) model.

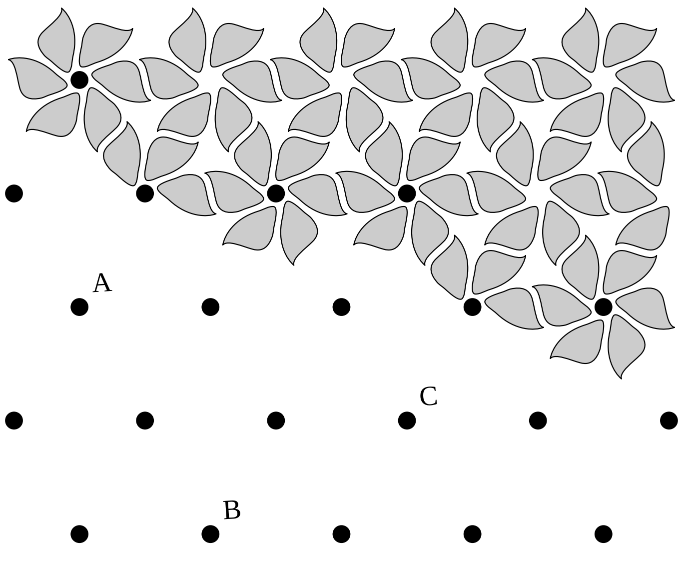


Figure 10: A hexagonal lattice. The large dots represent the centers of hexamers. In the top half of the figure, the individual coat protein subunits forming the hexamers are shown represented by blobs. The lattice vector between A and B is $(2, 0)$, between A and C is $(2, 1)d$, and between B and C is $(1, 1)$.

4 Theory

4.1 Notation and definitions

Hexagonal lattices will be discussed extensively in this section, and some notation will now be introduced which facilitates this discussion. Like the coat proteins of many viruses, the coat protein of T4 (gp23) tends to form hexagonal lattices. More specifically, the protein forms hexamers which are arranged in a hexagonal lattice. In virus capsids, this hexagonal lattice is "folded" to form a closed polyhedral structure. Closed structures cannot be formed solely by a hexagonal lattice; in order to form a closed structure, these viruses have 12 pentamers which replace hexamers at the vertices of the polyhedron. While in many viruses, the pentamers are formed from the same protein as the hexamers, in T4 the pentamers are believed to be formed from gp24 (Müller-Salamin *et al.*, 1977). A capsomere will mean either a hexamer or a pentamer in the shell.

Consider two hexamers, H_1 and H_2, in a hexagonal lattice. The lattice vector between them will be denoted (a, b) if starting from H_1, one can traverse a hexamers in a straight line, make a 120° angle (left or right), traverse b more hexamers and reach H_2 (see Figure 10). Two hexamers can be related by a lattice vector (a, b) in two different ways, which are mirror images. For clarity, a vector with $a \geq b$

will always be used, and $(a, b)d$ will be used for a 120° right turn and $(a, b)l$ for a 120° left turn. The vectors (a, a) and $(a, 0)$ do not have a handedness associated with them (see Figure 10). This definition of right- and left-handedness agrees with the accepted definition of handedness for both icosahedral shells and cylindrical polyheads.

Many viruses in which the coat protein forms a hexagonal lattice are icosahedral; *i.e.,* they have icosahedral symmetry (Caspar & Klug, 1962). The structure of these viruses is classified by T-number. If two vertices of the corresponding icosahedron are related by a lattice vector (a, b), then the T-number is $a^2 + ab + b^2$, and there are a total of $60T$ coat proteins in the shell. The shell is left- or right-handed depending on whether the lattice vector is left- or right-handed.

Some viruses, including T4, are not spherical but prolate in shape. Typically, these have a structure in which an icosahedron has been extended by adding extra rings of hexamers around its diameter, forming a cylindrical shell with icosahedral caps. The structure of these shells is classified by both the T-number and Q-number (Moody, 1965). If a virus is icosahedral, then the Q number is the same as the T number; for every five hexamers added around the diameter, the Q number increases by one. Note that five is the minimum number of hexamers that can be added, since such prolate viruses still have five-fold symmetry about their long axis. A virus shell is thus composed of $60T + 30(Q - T)$ copies of its coat proteins altogether. The T4 shell has $T = 13$ and $Q = 21$ (Baschong *et al.*, 1988).

A hexagonal lattice can be rolled up into a cylinder without introducing any pentamers. This is a common type of malformation seen in many viruses, including T4 (Yanagida *et al.*, 1972; Steven *et al.*, 1976; Hendrix, 1985). To describe the combinatorial structure of a hexagonal lattice folded into a cylinder, all that is necessary is the equatorial lattice vector; that is, the lattice vector that describes the path around the cylinder from any hexamer back to itself. The handedness of this vector agrees with the standard definition of handedness of these cylindrical shells; namely, a shell is left-handed if the least-steep lattice lines form a left-handed helix.

Much of the knowledge of the assembly mechanism in T4 comes from the study of malformations in its assembly. In most malformations, the coat protein still binds in a hexagonal lattice, but this lattice is folded incorrectly so as to form an improperly shaped virus shell. Two classes of malformations that have received much study are giant heads and polyheads (Yanagida *et al.*, 1972; Steven *et al.*, 1976; Lane *et al.*, 1990; Lane & Eiserling, 1990). In these types of mutation, the lattice is folded into a long cylinder. Other shape-altering mutations produce isometric heads and shorter-than-normal prolate heads called intermediate length heads (Mosig *et al.*, 1972; Doermann & Pao, 1987; Keller *et al.*, 1988; Lane *et al.*, 1990).

In giant heads, this cylinder has spherical caps which are of the normal size, *i.e.,* $T = 13$, but the Q number, instead of being 21, is much larger. Giant heads appear to begin assembly in the normal manner, but the head is never terminated or is terminated much later than usual, producing a head that is longer than normal. Some giant heads still close, yet with a larger Q number than normal, while other giant heads have caps on only one end (Paulson & Laemmli, 1977). Because of the caps, it is reasonably certain that the equatorial vector in giant heads is still $(15{,}5)l$ and that the structure of the core, in particular the number of helices in the core, is also normal. Indeed, the (15,5) equatorial vector has been verified by experiment (Aebi *et al.*, 1977; Paulson & Laemmli, 1977). Giant heads can be formed by the addition of the amino acid analogue L-canavanine to the T4-infected bacterial cells (Couse *et al.*, 1970; Paulson & Laemmli, 1977) as well as mutations in gp23 and 24 (Bijlenga *et al.*, 1976; Lane *et al.*, 1990; Lane & Eiserling, 1990; Black *et al.*, 1995).

In polyheads, the shell lattice forms a long cylinder which does not have caps on the ends and thus may not have the normal equatorial vector. For several mutations, the distributions of the equatorial vector of polyheads has been measured, and they vary depending on the mutation. The polyheads caused by the three mutations 20^-, 24 (tsB86), and 40 (tsL84) (Paulson & Laemmli, 1977) all seem to have the same structural parameters, including the same distribution of equatorial lattice vectors, which is clustered around the (11,6) equatorial vector (Steven *et al.*, 1976; Yanagida *et al.*, 1970). The 22^- and IPIII$^-$ mutations, which lack a core or have a defective core, form polyheads with the equatorial vectors clustering around (9,6) and (10,6) (Steven *et al.*, 1976; Yanagida *et al.*, 1970).

4.2 Ten helical chains and a possible Vernier mechanism

One of the mechanisms proposed for head length determination is a Vernier mechanism involving the core and the shell (Paulson & Laemmli, 1977; Lane *et al.*, 1990; Lane & Eiserling, 1990). In such a mechanism, the elongated part of the shell can be terminated only when elements in the head and the core lie in a specific relation to each other. If the shell and the core have repeating elements of different lengths, a Vernier mechanism can accurately measure a length much longer than either component repeating length. While the simplest Vernier mechanisms are one-dimensional, those discussed in this paper are two-dimensional, with the repeating elements forming a lattice in the plane. In this section, it is shown that the hypothesis of a core composed of ten helical chains aligned with a $(3,1)d$ lattice vector in the shell provides a possible mechanism for the previously unexplained behavior of the length determination process in giant head mutants of T4.

Experimental evidence for the existence of a Vernier mechanism is given in Lane *et al.* (1990) and Lane & Eiserling (1990). These papers detail experiments where the exact Q-number was determined for heads produced by several mutations of T4 that produce aberrant head lengths. Because T4 packs DNA using a head-full mechanism, *i.e.,* as much DNA is packed into a T4 head as will fit, the length of the DNA, which can be accurately determined, is proportional to the volume of the bacteriophage shell. This in turn can be used to deduce the Q number. Lane *et al.* observed the values of Q for various mutations that altered head length. These mutations did not produce uniform head lengths; rather, each mutation gave a set of possible Q numbers, and different mutations could produce different sets of observed Q numbers. With the mutations that produced giant heads, the possible Q numbers of giant heads with a single mutation formed a sequence that increased alternately by increments of 3 and 4. Together, these form a repeating pattern of period 7, which suggests the possibility of a Vernier mechanism for head length determination (Lane & Eiserling, 1990).

An alternative explanation for why T4 favors certain head lengths has also been proposed. Kellenberger (1990) suggested that shell geometry could determine the head lengths seen in Lane *et al.* (1990), so that the head length determination mechanism might produce Q numbers that are "more icosahedral" in that each of the five vertices terminating the central part of the core lies roughly half-way between two of the vertices starting the core. While this mechanism seems a plausible explanation of the length determination process for normal and short head lengths, it seems difficult to see how such a mechanism could give the strong signal seen in giant heads, where the alternating increments of 3 and 4 is still seen in heads with Q numbers near 70 (Lane *et al.*, 1990). At these lengths, the central section of the core is long enough that it might be expected that the elongated portion of the head should be nearly cylindrical in shape; if this is indeed the case, it is difficult to see how the shell geometry could markedly affect the head length.

We give a specific potential Vernier mechanism which could explain the observed alternating increments of three and four in giant heads. In Figure 11, the assumption of ten helices in the core was used to plot the relation of the core helices to the shell lattice on an unrolled representation of the giant head. The core helices and the shell lattice are plotted left-handed, based on experimental data (Engel *et al.*, 1982). Because the core helices are projected outward onto the shell lattice, the only data needed to draw this figure are the equatorial vector for the giant heads, the axial repeat (pitch) of the core helices, the lattice constant for the surface lattice, and the relative position of the core helices to the surface lattice at some point. The exact relative position of the core helices and the surface lattice is not relevant for the determination of the existence of a possible Vernier mechanism, as the only important issue here are the intervals at which the termination signal repeats, and these intervals are independent of this relative position. This relative position was chosen arbitrarily by drawing the core helices through the center of a pentamer (at the bottom of the figure). For comparison, the relation of the shell and the core given the assumption of six helical chains is shown in Figure 12.

For Figure 11, the values used for the core axial repeat of 115Å and the shell lattice constant of 114Å were those measured by Paulson & Laemmli (1977). The equatorial vector $(15,5)l$ follows from the $T = 13$ structure of T4. We also assume

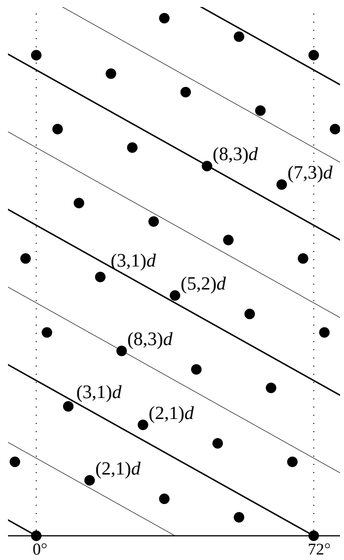


Figure 11: Schematic representation of one fifth of a giant head, unrolled onto a plane, assuming the core is composed of ten helical chains. The whole circumference of the giant head is formed by repeating the section between the vertical dotted lines five times. Large dots represent hexamers in the shell lattice. Diagonal lines are core helices. The heavy lines represent core helices which go through a pentamer, while the light ones do not. Certain hexamers are labeled by lattice vectors; these are the lattice vectors which would result if the nearest core helix went through the exact center of that lattice point. Thus, to find the small lattice vectors, it is generally necessary to look at dots near helices represented by heavy lines. Note that the two types of helices need not differ structurally, but only in their spatial relationship with the shell lattice.

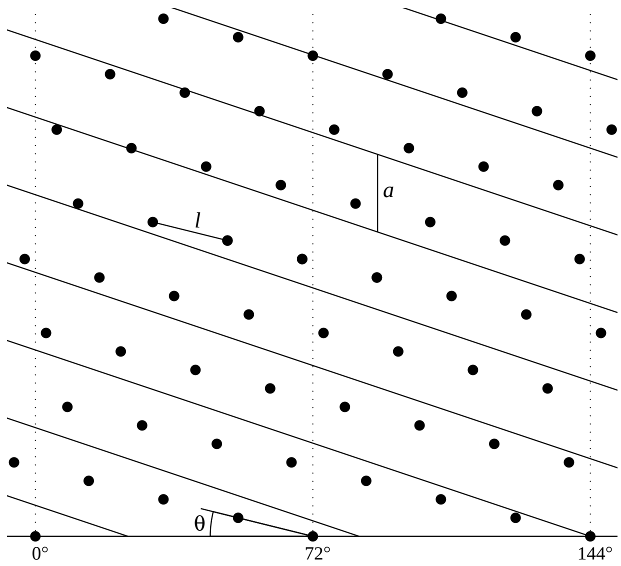


Figure 12: Two-fifths of an unrolled giant head with the currently accepted structure (*i.e.,* six helical chains in the core.) The large dots represent the centers of hexamers and the diagonal lines represent the projection of the core helical ribbons onto the shell lattice. The arrangement of hexamers between pairs of vertical dotted lines repeats to form the shell lattice. This structure is determined, up to the phase relation between the core and the shell, by the lattice constant of the shell (labeled l), the axial repeat or pitch of the core (labeled a), the pitch angle of the shell lattice (labeled θ), and the assumption that the core is six helical chains. The pitch angle of 13.9° for the shell lattice is determined by the (15,5)l equatorial vector. The values for axial repeat and lattice constant are from experimental measurements (Paulson & Laemmli, 1977). The phase of the core ribbons with respect to the shell lattice was decided arbitrarily for this illustration by putting a core ribbon through the center of the pentamer at 144°. The pitch angle of the helical ribbons in the core is actually steeper than it appears in the figure because these ribbons have been projected outward onto the shell lattice so as to depict the interaction between the core and the shell faithfully.

lattice vector	ratio of axial repeat to lattice constant	resulting increments of Q-number	number of non-equivalent positions of coat protein
$(4,1)d$	0.83	4	12
$(7,2)d$	0.88	15	45
$(10,3)d$	0.90	11	33
$(3,1)d$	0.95	7	21
$(8,3)d$	1.01	10	30
$(5,2)d$	1.04	13	39
$(7,3)d$	1.08	16	48
$(2,1)d$	1.17	3	9

Table 2: Comparison of various alignments of core helices with lattice vectors. The ratio of axial repeat to lattice constant observed by Paulson and Laemmli (1977) was $1.01 \pm .07$.

that the lattice constant for the surface lattice is the same in all three directions; if this is true, then the exact pitch angle of the shell lattice is known to be 13.9° because the equatorial vector of the giant head is $(15, 5)l$. While in some T4 giant heads, the lattice constant has been observed to vary by direction (Steven *et al.*, 1976), it is assumed for this analysis that it is the same in all directions; evidence that this assumption is not far off is that the theoretical pitch angle of 13.9° is close to the measured figure of $13° \pm 1.5°$ (Paulson & Laemmli, 1977) for canavanine induced giant heads.

A Vernier mechanism requires a specific relationship between the core helices and the shell lattice. The regular increments of Q number seen in the lengths of giant heads suggests that this relationship must repeat for long heads. If the shell lattice and the core helices have a regular repeating relation in the giant heads, then the angle formed by projecting the core helices out to the surface lattice must match some integer vector in the shell lattice, as otherwise the two patterns would never recur in the exact same relative positions. For this regular repeating relation to have a relatively small unit cell, the angle must match a small integer vector.

Assuming the hypothesis of ten helical chains, there are a large number of possible alignments of shell and core; however, only a relatively small number of these alignments result in a simple repeating pattern formed by the shell and the core. In Figure 11, the smallest lattice vectors that are close to the angle of the core helix vector are $(3,1)d$, $(5,2)d$, and $(8,3)d$. More relatively small lattice vectors close to this angle are detailed in Table 2. Of these, however, only the alignment corresponding to the $(3,1)d$ lattice vector results in a repeating pattern with a period of 7, as would seem to be required by the alternate increments of 3 and 4 of the Q number observed in Lane and Eiserling (1990) and Lane *et al.* (1990). More complicated alignments of the core and shell than are given in Table 2 are possible, but would result in larger increments before the pattern starts repeating.

The vectors $(5,2)d$ and $(8,3)d$ give a better fit to the measured value of 1.01 $\pm$ .07 for the ratio of the axial repeat to the lattice constant than the value of .95 given by $(3,1)d$ (see Table 2). However, the vector $(3,1)d$ is in some sense a simpler hypothesis since it is smaller. This is reflected in Table 2 by the minimum number of non-equivalent positions the coat protein would need to take in relation to the core ribbons.

As noted above, evidence for the $(3,1)d$ angle for the core helices comes from experiments on T4 head length regulation (Lane *et al.*, 1990; Lane & Eiserling, 1990). In these experiments, the exact Q-number was determined for several mutants of T4 producing aberrant head lengths. Because as much DNA is packed into a T4 head as fits, the length of the DNA, which can be measured accurately, is proportional to the volume of the bacteriophage shell. This in turn can be used to deduce the Q number. Lane *et al.* observed the values of Q for various mutations that altered head length. These mutations did not produce uniform head lengths; rather, each mutation gave a set of possible Q numbers, where different mutations could produce different sets of observed Q numbers.

What is observed in giant head mutants is that the Q number can increase alternately by increments of 3 and 4 (Lane *et al.*, 1990; Lane & Eiserling, 1990). Together, these form a repeating pattern of period 7, which is the period associated with a $(3,1)d$ lattice vector for the alignment of the core helices. Examination of the relation of the core and the shell, assuming the $(3,1)d$ alignment for the core helices, shows that there are two types of core helices that lie in different positions relative to the shell; in Figure 13, one type is drawn to go through the center of the hexamers, while the other just misses the hexamers.

If there were a small tolerance in the relative positions of the helices and the shell lattice required for the termination signal, the termination signal in Figure 13 could be given when a core helix passes either through the center of a capsomere or slightly below the center of a capsomere. Consequently, alternating increments of 3 and 4 could indeed be obtained as possible Q numbers. We have chosen to draw the core helices through hexamers corresponding to $Q = 7i$ since $Q = 21$ is known to be a favored stopping point. While the Q values of $7i + 3$ and $7i + 4$ come equally close to hitting the centers of hexamers Figure 13, there is no need to assume that the termination mechanism is equally likely to be triggered by a near-miss when the core helices go above or below the center of a hexamer. Also note that while we

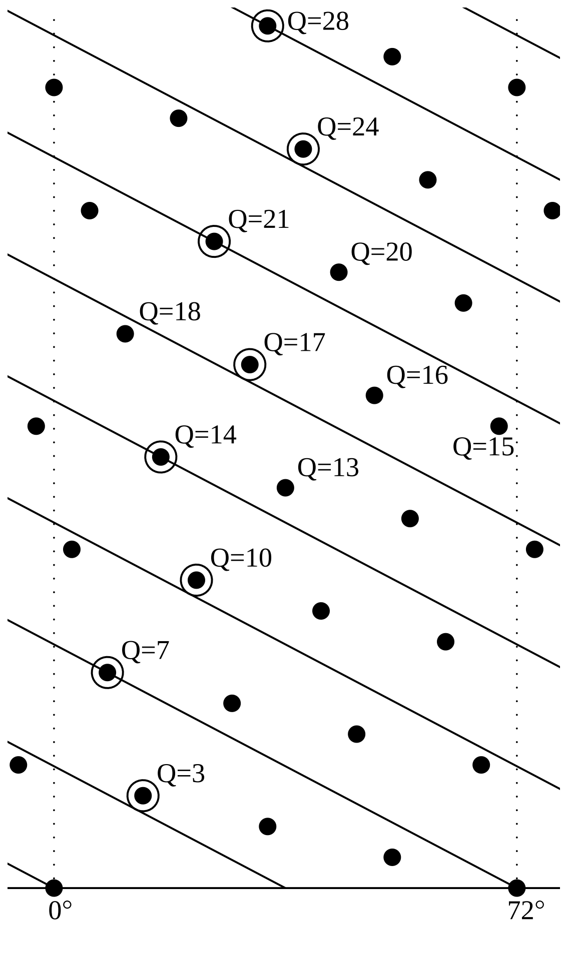


Figure 13: Schematic representation of one fifth of a giant head, unrolled onto a plane. The whole giant head is formed by repeating the section between the vertical dotted lines five times. Large dots represent the centers of hexamers in the shell lattice. Diagonal lines are core helices aligned with $(3,1)d$ lattice vectors. Note that the pitch angle and axial repeat are close to those in Figure 11. The circled hexamers are those with centers lying on or just above a core helix. The Q-numbers of the circled hexamers increase in alternating increments of 3 and 4.

have circled hexamers corresponding to $Q = 7i+3$ in Figure 13, there is no evidence to favor these over $Q = 7i + 4$, and this choice for this example was completely arbitrary.

For the moment it will be supposed that the core helices are located at the outer radius of the core at 211Å, and the structural parameters of the core will be calculated. Our analysis shows that if there are ten helices, the effective radius must be close to the outer radius, so this is a reasonable assumption. At this radius, the pitch angle of the core helices would be 39.3°. The distance between core helices would be 83Å. This should be compared with the 95Å width of gp22 filaments. If the filaments are oriented so their widest dimension is parallel to the shell, they would thus overlap slightly; however, if they were angled inward somewhat, they could avoid any overlap. Finally, the distance between two Vernier points measured along the core helices is 303Å. This should be a multiple of the approximately 15Å distance between individual gp22 molecules along the core helices; the ratio between these numbers is large enough that this is easily possible.

The Vernier mechanism proposed in this section could also play a part in length determination of the wild type T4 head, but the evidence for this is less clear. Furthermore, the complete length-determination mechanism must be at least somewhat more complicated. This issue is further addressed in Section 5.

4.3 Other Vernier mechanisms for six and eight helices

In light of the relatively good match between the Vernier mechanism discussed in the previous section and the biological evidence, an investigation of possible Vernier mechanisms on structures with six and eight helices in the core was undertaken to see whether any of them provided as good a match. Cores with six and eight helices cannot match the five-fold symmetry of the shell, and thus any Vernier mechanisms operating on a core containing six or eight helices must trigger cap formation by having the core helices come into register with the shell lattice at only one vertex of the shell, instead of at all five. These Vernier mechanisms must also operate with smaller tolerance than Vernier mechanisms on a five-helix model, as there will be many more hexamers which must be approached closely by core helices without triggering cap formation. However, while these arguments may make a Vernier mechanism with six or eight helices in the core somewhat less plausible, they should not be considered a barrier to the existence of such a mechanism. Several possible mechanisms were indeed found, although none was quite as good a fit with the data as the ten-helix mechanism.

With eight helices, using the axial repeat and the lattice constant from Paulson and Laemmli (1977), the match between the shell and the core is shown in Figure 14. If we wish our Vernier mechanism to operate with increments of period 7 in Q number for giant heads, the core helices must go through one of the five lettered

number of core helices	hexamer passed through	ratio of axial repeat to lattice constant	lattice vector aligned with core helices	closest approach to centers of lattice points
6	C	0.90	(17,1)d	$7i + 3$, $7i + 4$
6	D	1.01	(10,1)d	$7i + 2$, $7i + 5$
6	E	1.15	(13,2)d	$7i + 2$, $7i + 5$
8	C	0.93	(16,3)d	$7i + 1$, $7i + 6$
8	E	1.04	(4,1)d	not applicable
10		0.95	(3,1)d	$7i + 3$, $7i + 4$

Table 3: Comparison of various alignments of core helices with lattice vectors, under the assumptions of six and eight core helices. The second column refers to which labeled hexamer (C, D, or E) the core helices pass through in Figures 14 and 17. The corresponding values for the ten-helix mechanism proposed in Section 4.2 are also given. The ratio of axial repeat to lattice constant observed by Paulson and Laemmli (1977) was $1.01 \pm .07$.

hexamers. Of these, only the hexamers labeled C and E approach the core helices closely. Setting the pitch of the helices so that they go through the centers of one of these hexamers gives two candidate Vernier mechanisms. These are shown in Figures 15 and 16.

With six helices, the experimentally measured axial repeat and lattice constant fit very closely with a Vernier mechanism; this is presented in Figure 17. This mechanism, however, is not completely satisfactory for the same reasons as in Figure 15; the closest the core helices come to the centers of hexamers they do not directly pass through occurs at hexamers corresponding to Q numbers other than $7i + 3$ or $7i + 4$. This might be corrected by introducing a Vernier measure along the core helices. It is also possible to adjust the pitch of the core helices slightly and have them pass through the hexamers labeled C and E in Figure 17. Letting them pass through the hexamers labeled C gives a fairly satisfactory Vernier mechanism (Figure 18).

In general, Vernier mechanisms with six or eight helices do not seem as attractive as the Vernier mechanism with ten helices because the termination signal can only be triggered in one hexamer, as opposed to in five symmetric hexamers; this would appear to require a more complicated mechanism for triggering assembly of the distal cap. Further, none of the Vernier mechanisms investigated in this section fits the data quite as well as the ten-helix Vernier mechanism of Section 4.2; this is summarized in Table 3.

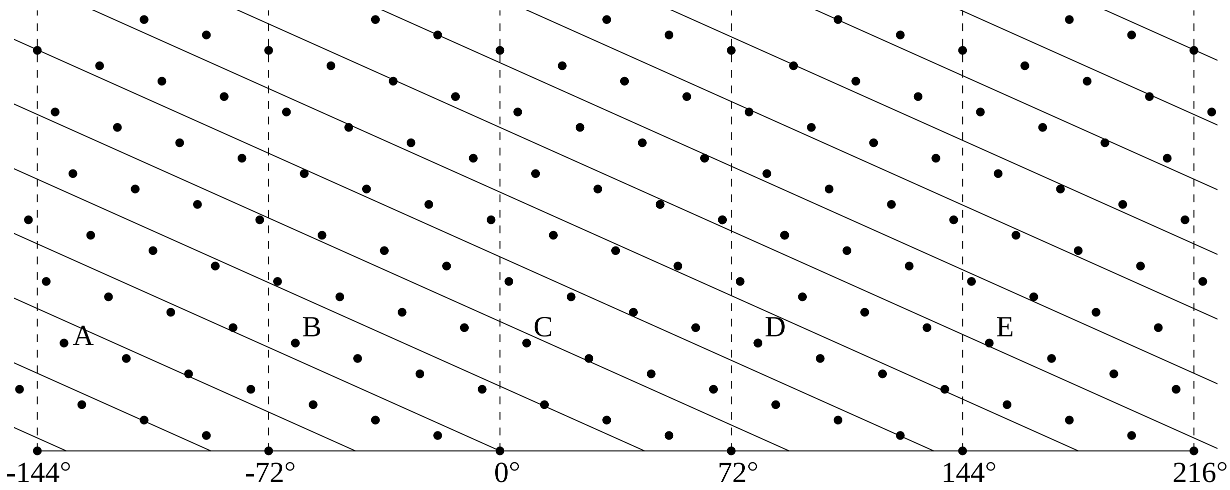


Figure 14: The relation of the core helices to the surface lattice obtained using the axial repeat and lattice constant measured in Paulson & Laemmli (1977), with the assumption of eight helices in the core. To obtain a Vernier mechanism having a period in Q-number of 7, the core helices must go through the center of one of the hexamers labeled A–E below. The hexamers C and E are closest to the core helices, and so are the best candidates.

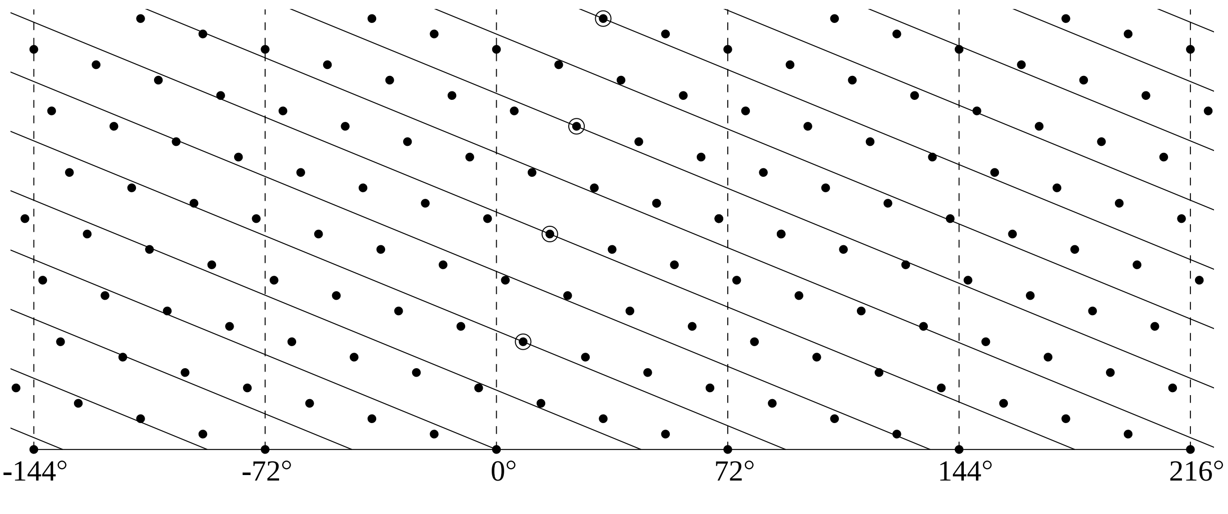


Figure 15: The relation of the core helices to the surface lattice obtained assuming the helices go through the hexamer labeled C in Figure 14. This gives a potential Vernier mechanism that works on eight helices, although it is not entirely satisfactorily since the closest the core helices pass to hexamers they do not go through the centers of occurs at Q numbers of $7i+1$ and $7i+6$, and not as $7i+3$ and $7i+4$, as might be suggested by experiments of Lane *et al.* (1990). Again, this may be rectifiable by assuming a Vernier length that operates along the core helices.

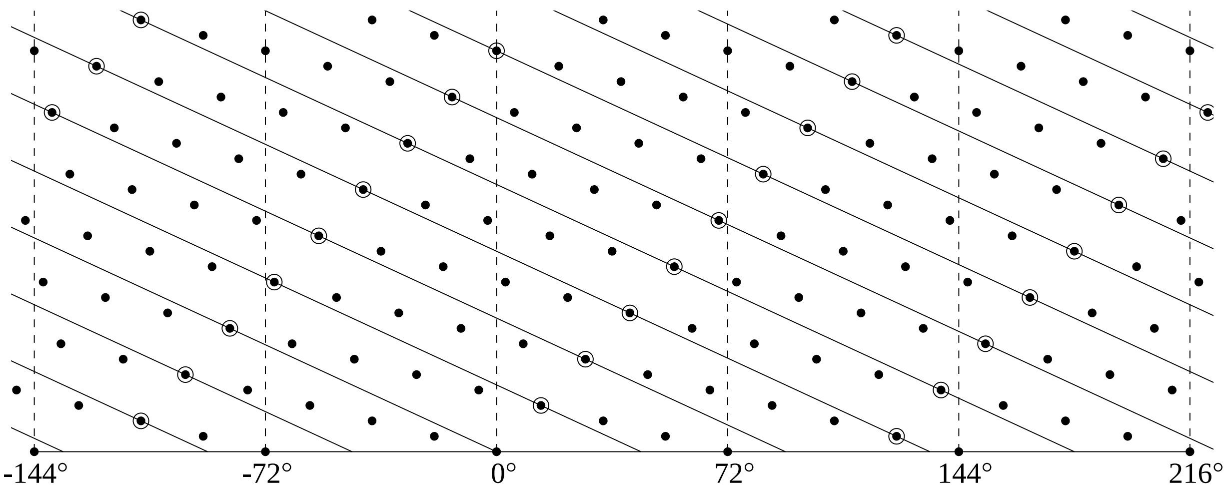


Figure 16: The relation of the core helices to the surface lattice obtained assuming the helices go through the hexamer labeled E in Figure 14. This does not result in a candidate Vernier mechanism, as the core helices pass through hexamers corresponding to all possible Q numbers. This difficulty might be rectifiable by also assuming a Vernier length that operates along the core helices.

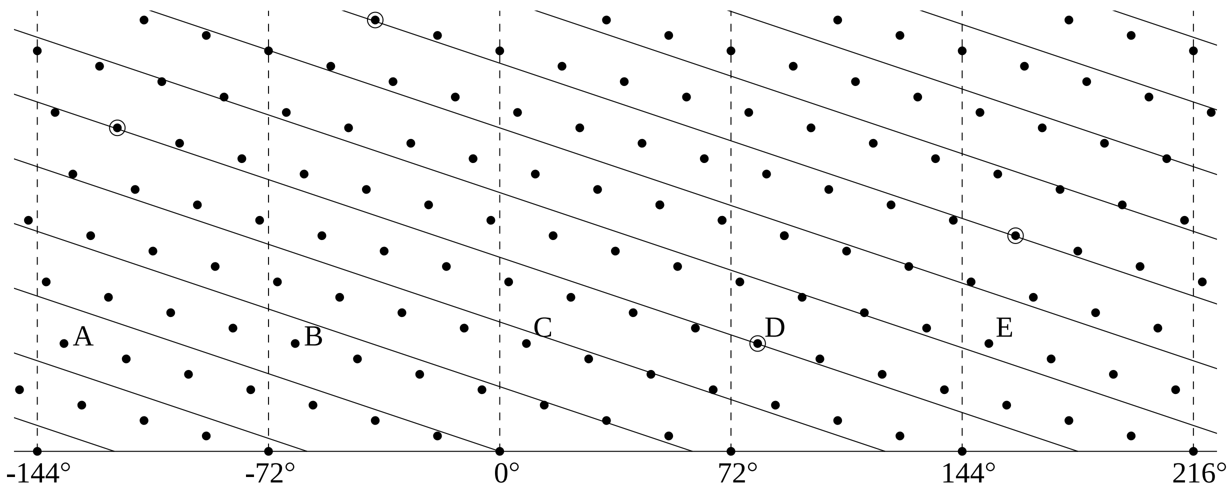


Figure 17: The values of axial repeat and lattice constant measured by Paulson & Laemmli (1977) are almost exactly right to give a potential Vernier mechanism on cores with six helices. The Vernier mechanism will be triggered at values of $Q = 7i$ by the core helices passing through the circled hexamers. This structure comes closes to triggering the Vernier mechanism at $Q = 7i + 2$ and $Q = 7i + 5$. Again, these values might be ruled out and values of $Q = 7i + 3$ or $Q = 7i + 4$ allowed by a Vernier length operating along the core helices. Of the hexamers corresponding to $Q = 7$, the ones closest to the core helices are those at C, D, and E.

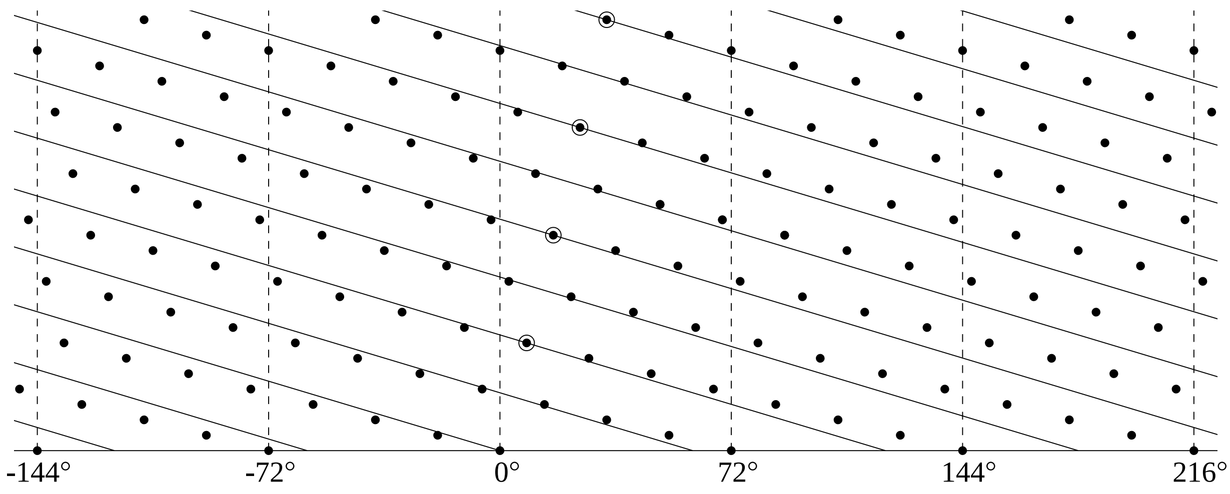


Figure 18: Structure obtained by adjusting structural parameters of Figure 17 to allow the core helices to pass through the center of hexamer C. This gives a Vernier mechanism which will come closest to triggering at hexamers with Q numbers of $7i + 3$ and $7i + 4$.

5 Discussion

We now have a simple model of the core which can explain the observed lengths of giant heads through a two-dimensional Vernier mechanism. A natural question is whether this mechanism can also explain the observed head lengths of mutations with short head lengths. This mechanism is unlikely to be the only one involved in head length determination, since if it were, one would expect to observe shells with $Q = 14$, and possibly even $Q = 7$, as well as $Q = 21$. However, it may be that this mechanism operates in tandem with another mechanism to produce $Q = 21$ length shells in wild-type virus.

In the work of Lane & Eiserling (1990), a number of mutations affecting the length-determining mechanism that resulted in intermediate-sized heads were examined, and the set of Q numbers that resulted from each mutation was determined. Most of the mutations resulted in a set of possible Q numbers that was a subset of the numbers 13, 17, 18, and 21. The numbers 17, 18, and 21 can easily be explained through the Vernier mechanism. Recall that the Vernier mechanism has a period of 7 but also comes close to the triggering position at offsets of 3 and 4. If the $Q = 21$ length is triggered by the Vernier mechanism, the 17 and 18 lengths can be explained through a loss of specificity in the Vernier mechanism induced by the mutation, so that this mechanism is now triggered at $Q = 17$ or $Q = 18$.

On the other hand, the Vernier mechanism would produce $Q = 14$ lengths rather than $Q = 13$. Experimental evidence, however, seems to show clearly that the isometric shells have $Q = 13$ and not 14 (Lane *et al.*, 1990; Lane & Eiserling, 1990). This may indicate that whatever mechanism is responsible for setting $T = 13$ is also operating here and trying to set $Q = 13$. It might normally be "overruled" by the Vernier mechanism, but in these mutations it is able to produce some isometric shells. It may be that the Vernier stopping point at $Q = 14$ is also "overruled" by the mechanism that is trying to produce isometric shells, and thus $Q = 21$ shells are produced in wild-type virus because they are the shortest length allowed by both mechanisms. Evidence consistent with this is that several mutations produce only isometric shells, without producing any intermediate-length shells (Doermann & Pao, 1987). These results can be explained if the mechanism "trying" to produce isometric shells is the only one operating properly in these mutations. This hypothesis is also consistent with the behavior of the gp22 mutant 22tsA74, which seems to cause instability in the core (Paulson & Laemmli, 1977) and which produces mainly (85%) isometric heads when grown at an intermediate temperature (Paulson *et al.*, 1976), indicating that in this case, isometric heads are indeed the default size with faulty cores.

Lane & Eiserling (1990) also observed some mutations that produced other sets of lengths. Two mutations produced possible Q numbers of 13, 16, 17, and 20. These can be explained using the Vernier mechanism through a shift in the position of the core helices. If, instead of triggering the stop mechanism at 17, 18 and 21, the position of the core helices were offset, the Vernier mechanism could trigger the stop mechanism at values which were less by 1, namely 16, 17 and 20.

Finally, in two of the mutations observed, a small percentage of the heads appeared to have length $Q = 15$ in mutants where the sets of stop lengths were $(13, 15, 17, 18, 21)$ and $(13, 15, 17, 21)$. The value of 15 is difficult to explain using the Vernier mechanism; it may result from a different length-determining mechanism.

Because the Vernier mechanism works through an interaction between the shell and the core, it can operate only if the shell and the core polymerize simultaneously. However, experiments show that cores of at least approximately the right size and shape will sometimes form in the absence of the main coat protein gp23 (Traub *et al.*, 1984; van Driel & Couture, 1978). One possible explanation for this is that the termination of the core is a coordinated process which is normally triggered by the Vernier mechanism, but which can also be triggered by other factors in the absence of the coat protein. Another possible explanation is that the core contains a redundant length determination mechanism that normally operates in tandem with the Vernier mechanism to produce $Q = 21$ shells.

Given the hypothesis that in the giant heads the shell lattice and the core helices lie in a regular relation to each other, it is reasonable to ask whether the same relation could also be present in the 20^- polyheads, which have different equatorial vectors but which also contain helical cores. These polyheads do not have caps, possibly because essential vertex proteins are missing or because the shell does not have the right circumference vector to accomodate a $T = 13$ cap. Paulson & Laemmli (1977) measured the values of the axial repeat and pitch angle of the shell lattice in both giant heads and polyheads. From the measured pitch angles of the shell lattice in the polyheads ($13 \pm 1.5°$ in the giant heads and $22 \pm 1.5°$ in the 20^- polyheads), if it is assumed that the 20^- polyheads are left-handed, it can be estimated that the lattice of the 20^- polyheads is angled 9° more steeply than the giant heads. The corresponding axial repeat this would give to the core is 127Å, which is close to the measured value of 131 ± 5Å. If the 20^- polyheads are right-handed, then it seems impossible for the core helices to have the same relation to the shell lattice as they do in giant heads. The 20^- polyheads have generally been assumed to be right-handed. One justification for this seems to have been via the analogy with 22^- polyheads, which shadowing studies showed are right-handed (Yanagida *et al.*, 1970). Since 22^- polyheads lack cores and appear to have different structural parameters than 20^- polyheads (Steven *et al.*, 1976), the assumption that they have a different handedness than the 20^- polyheads is not unreasonable. It was also reported in Steven *et al.* (1976) that preliminary shadowing studies showed 40^- polyheads to be right-handed. (The details of these experiments do not seem to have been published.) Since 40^- and 20^- polyheads are very similar, they presumably share the same handedness, which the ten-helix Vernier model proposed in this paper would then predict to be predominantly left-handed, in contradiction with the Steven *et al.* (1976) results.

In the model proposed in this paper for the structure of the T4 core, the shell and the core both form repeating lattice patterns that provide a two-dimensional Vernier mechanism for length determination, at least in giant heads. It has not been resolved in the literature how the shell and core relate. One hypothesis set forth is that the shell sees the core as a surfaceless template on which to polymerize (Engel *et al.*, 1982). For the length determination mechanism proposed in this paper to work properly in giant heads, the relationship between the shell and the core must be maintained in the relatively long giant heads investigated by Lane *et al.* (1990).

If the surface and core were actually behaving like featureless surfaces and only interacted to trigger the Vernier mechanism, it would not seem likely that the spatial relationship between them would be maintained for large distances, as this would imply that the angle between them formed during the initial cap formation is relatively accurate, and does not change during the elongation process. It would thus be expected that for the Vernier mechanism to work, deviations from this angle should be corrected during the elongation process. One way that this angle might be maintained would be if regular repeating links between the shell and the core existed. These links would also seem to be necessary for the core to be stretched laterally by the flattening of giant heads, rather than simply being flattened. However, a tendency for gp23 to polymerize on substrates while treating them as a featureless surface may still be important in the polymerization process, as these specific core-shell links might be spaced relatively far apart in the head.

Evidence is given by Doherty (1982a, 1982b) that specific interactions between the shell and the core are important in T4 head length determination, as second-site revertants of ptg mutations in gene 23 were found to be specific to certain alleles in gp22. This suggests that interactions at specific sites between the main core protein gp22 and the coat protein gp23 are important in the length determination mechanism. The interactions observed in (Doherty 1982a, 1982b) could be either regular repeating links connecting the shell and the core or they could be the interactions triggering cap formation through the Vernier mechanism.

As the combination shell-core lattice repeats only at relatively large intervals, only certain equatorial vectors in the polyhead will allow both the shell and the core to be in unison all the way around the cylindrical polyhead. In the 20^- lattices, the measurements of Yanagida *et al.* (1970) show that the equatorial lattice vectors are centered around (12,6) and (11,7), while the measurements of Steven *et al.* show they are clustered around (11,6). If it is assumed that they are in fact left-handed, and so have the same relative orientation of the core helices to shell lattice, the only close equatorial vector that will fit the unit cell of the core-shell combination is $(11,6)l$. This occurs quite often, although both Yanagida *et al.* and Steven *et al.*

seem to show that there are also many polyheads with other equatorial vectors. An equatorial vector different from $(11, 6)l$ would lead to a mismatch between the shell and the core somewhere in the polyhead, much like a defect in a crystal. Possibly this could explain the observation (Paulson & Laemmli, 1977) that the core and shell are more loosely bound in this type of polyhead.

It is worthwhile to compare this research with the local rule theories of shape determination (Berger *et al.*, 1994). In both cases, information about the structure of the shell is propagated through the shell using only local interactions between proteins. In a two-dimensional Vernier mechanism the information is carried by the relative positions of the core and shell proteins, while in a local rule theory the information is carried by the different conformations taken by the shell (and possibly the scaffolding) proteins.

Given the uncertainty introduced by the reanalysis of this data, it would be advisable to reexamine the structure of giant head cores to conclusively determine the number of helices they contain. While cryogenically prepared giant heads would not tend to undergo as much flattening and thus would be better suited to this analysis, they will still have some of the same problems as our analysis of the nearly unflattened heads in that uncertainties in the effective diameter of the helices may not allow a conclusive determination of the number of helices. Depending on the method of preparation, it may be possible to avoid the difficulty of stain obscuring possible structure in the space between the shell and the core. If it is indeed determined that the core contains 10 helices, the Vernier mechanism proposed in this paper appears to be a good candidate for a component of the length determination process. If this Vernier mechanism could be confirmed experimentally, this would be the first biological evidence for the participation of a Vernier mechanism in morphogenesis.

Acknowledgements

B.B. was supported in part by MIT grants AF #78181, ARPA #N00014-92-J-1799, and an NSF Career Award. G.W.H. was supported by the MIT Lab. for Computer Science Undergraduate Research Opportunities Program. We are indebted to David DeRosier and his laboratory for the use of their equipment and especially to David Morgan and Dennis Thomas for all their assistance. We would like to thank David DeRosier for his extensive comments on a draft of our paper. We would also like to thank Jonathan King, M.F. Moody, and Alasdair Steven for most helpful suggestions.

References

Aebi, U., Bijlenga, R., Broek, J. V. D., Eiserling, F., Kellenberger, C., Kellenberger, E., Mesyanzhinov, L., Showe, M., Smith, R., & Steven, A. (1974). The transformation of τ particles into T4 heads. *J. Supramol. Struct.* **2**, 253–275.

Aebi, U., Bijlenga, R. K. L., ten Heggeler, B., Kistler, J., Steven, A. C., & Smith, P. R. (1977). Comparison of the structural and chemical composition of giant T-even phage heads. *J. Supramol. Struct.* **5**, 475–495.

Baschong, W., Aebi, U., Baschong-Prescianotto, C., Dubochet, J., Landmann, L., Kellenberger, E., & Wurtz, M. (1988). Head structure of bacteriophages T2 and T4. *J. Ultrastructure Molec. Struct. Research,* **99**, 189–202.

Bazinet, C. & King, J. (1985). The DNA translocating vertex of dsDNA bacteriophage. *Ann. Rev. Microbiol.* **39**, 109–129.

Berger, B., Shor, P. W., Tucker-Kellogg, L., & King, J. (1994). Local rule-based theory of virus shell assembly. *Proc. of the Natl. Academy of Sci.* **91** (16).

Bijlenga, R. K. L., Aebi, U., & Kellenberger, E. (1976). Properties and structure of a gene 24-controlled T4 giant phage. *J. Mol. Biol.* **103**, 469–498.

Black, L. W., Showe, M. K., & Steven, A. C. (1995). Morphogenesis of the T4 head. In: *Bacteriophage T4*, (Mathews, C. K., ed) pp. 218–258. American Society for Microbiology Washington, D. C.

Branton, D. & Klug, A. (1975). Capsid geometry of bacteriophage T2: a freeze-etching study. *J. Mol. Biol.* **92**, 559–565.

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

Cochran, W., Crick, F. H. C., & Vand, V. (1952). The structure of synthetic polypeptides I. The transform of atoms on a helix. *Acta Cryst.* **5**, 581–586.

Couse, N. L., Cummings, D. J., Chapman, V. A., & DeLong, S. S. (1970). Structural aberrations in T-even bacteriophage 1: Specificity of induction of aberrations. *Virology,* **42**, 590–602.

DeRosier, D. J. (1995). personal communication.

Doermann, A. H. & Pao, A. (1987). Genetic control of capsid length in bacteriophage T4: Phenotypes displayed by *ptg* mutants. *J. of Virology,* **61** (9), 2835–2842.

Doherty, D. H. (1982a). Genetic studies on capsid-length determination in bacteriophage T4. I. Isolation and partial characterization of second-site revertants of a gene 23 mutation affecting capsid length. *J. of Virology,* **43** (2), 641–654.

Doherty, D. H. (1982b). Genetic studies on capsid-length determination in bacteriophage T4. II. Genetic evidence that specific protein-protein interactions are involved. *J. of Virology,* **43** (2), 641–654.

Engel, A. & van Driel, R. (1981). Structure of the prolate phage T4 prehead core: Its implication for head-shape determination. In: *Structure Aspects of Recognition and Assembly in Biological Macromolecules*, (Balaban, M., Sussman, J. L., Traub, W., & Yonath, A., eds) pp. 921–935. Balaban ISS.

Engel, A., van Driel, R., & Driedonks, R. (1982). A proposed structure of the prolate phage T4 prehead core: An electron microscopic study. *J. Ultrastructure Research,* **80**, 12–22.

Harrison, S. C., Olson, A. J., Schutt, C. E., Winkler, F. K., & Bricogne, G. (1978). Tomato bushy stunt virus at 2.9 Å resolution. *Nature,* **276**, 368–373.

Hendrix, R. W. (1978). *Proc. Natl. Acad. Sci. USA,* **75**, 4779–4783.

Hendrix, R. W. (1985). Shape determination in virus assembly: The bacteriophage example. In: *Virus Structure and Assembly*, (Casjens, S., ed) pp. 169–204. Jones and Bartlett Boston, MA.

Hogle, J. M., Maeda, A., & Harrison, S. C. (1986). Structure and assembly of turnip crinkle virus I: X-ray crystallographic structure analysis at 3.2 Å resolution. *J. Mol. Biol.* **191**, 625–638.

Kellenberger, E. (1990). Form determination of the heads of bacteriophages. *Eur. J. Biochem.* **190**, 233–248.

Keller, B., Dubochet, J., Adrian, M., Maeder, M., Wurtz, M., & Kellenberger, E. (1988). Length and shape variants of the bacteriophage T4 head: Mutations in the scaffolding core genes 68 and 22. *J. of Virology,* **62** (8), 2960–2969.

Laemmli, U. K., Molbert, E., Showe, M., & Kellenberger, E. (1970). Form-determining function of the genes required for the assembly of the head of bacteriophage T4. *J. Mol. Biol.* **49**, 99–113.

Lane, T. & Eiserling, F. (1990). Genetic control of capsid length in bacteriophage T4 VII: A model of length regulation based on DNA size. *J. Struct. Biol.* **104**, 9–23.

Lane, T., Serwer, P., Hayes, S. J., & Eiserling, F. (1990). Quantized viral DNA packaging revealed by rotating gel electrophoresis. *Virology,* **174**, 472–478.

Mesyanzhinov, V. V., Sobolev, B. N., Marusich, E. I., Prilipov, A. G., & Efimov, V. P. (1990). A proposed structure of baceriophage T4 gene product 22—a major prohead scaffolding core protein. *J. Struct. Biol.* **104**, 24–31.

Moody, M. F. (1965). The shape of the T-even bacteriophage head. *Virology,* **26**, 567–576.

Moody, M. F. (1967). Structure of the sheath of bacteriophage T4. I: Structure of the contracted sheath and polysheath. *J. Mol. Biol.* **25**, 167–200.

Mosig, G., Carnighan, J., Bibring, J., Cole, R., Bock, H.-G. O., & Bock, S. (1972). Coordinate variation in lengths of deoxyribonucleic acid molecules and head lengths in morphological variants of bacteriophage T4. *J. of Virology,* **9** (5), 857–871.

Müller-Salamin, L., Onorato, L., & Showe, L. (1977). Localization of minor protein components of the head of bacteriophage T4. *J. of Virology,* **24**, 121–134.

Paulson, J. R. & Laemmli, U. K. (1977). Morphogenetic core of the bacteriophage T4 head: Structure of the core in polyheads. *J. Mol. Biol.* **111**, 459–485.

Paulson, J. R., Lazaroff, S., & Laemmli, U. K. (1976). Head length determination in bacteriophage T4: the role of the core protein P22. *J. Mol. Biol.* **103**, 155–174.

Showe, M. & Black, L. (1973). Assembly core of bacteriophage T4: an intermediate in head formation. *Nature New Biology,* **242**, 70–75.

Silva, A. M. & Rossmann, M. G. (1987). Refined structure of southern bean mosaic virus at 2.9 Å resolution. *J. Mol. Biol.* **197**, 69–87.

Steven, A. C., Aebi, U., & Showe, M. K. (1976). Folding and capsomere morphology of the P23 surface shell of bacteriophage T4 polyheads from mutants in five different head genes. *J. Mol. Biol.* **102**, 373–407.

Traub, F., Keller, B., Kuhn, A., & Maeder, M. (1984). Isolation of the prohead core of bacteriophage T4 after cross-linking and determination of protein composition. *J. Virology,* **49**, 902–908.

van Driel, R. (1980a). The role of a scaffolding core in the assembly of the bacteriophage T4 head shell lattice. In: *Electron Microscopy at Molecular Dimension*, (Baumeister, W. & Vogell, W., eds) pp. 129–136. Springer-Verlag Berlin, Germany.

van Driel, R. (1980b). Assembly of bacteriophage T4 head-related structures IV. Isolation and association properties of T4 prehead proteins. *J. Mol. Biol.* **138**, 27–42.

van Driel, R. & Couture, E. (1978). Assembly of the scaffolding core of bacteriophage T4 preheads. *J. Mol. Biol.* **123**, 713–719.

Yanagida, M., Boy de la Tour, E., Alff-Steinberger, C., & Kellenberger, E. (1970). Studies in the morphopoiesis of the head of bacteriophage T-even. *J. Mol. Biol.* **50**, 35–58.

Yanagida, M., DeRosier, D. J., & Klug, A. (1972). Structure of the tubular variants of the head of bacteriophage T4 (polyheads). *J. Mol. Biol.* **65**, 489–499.