RNA PACKAGING IN VIRUSES

What are the differences between ssRNA and dsDNA?

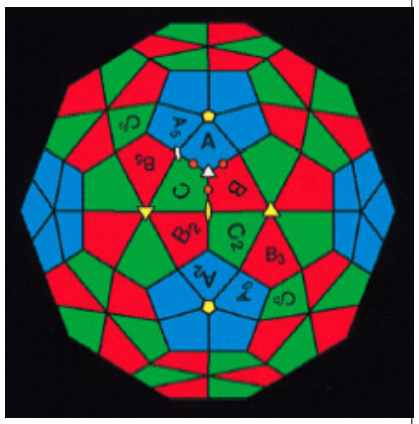
What are the consequent differences between self-assembled and motor-packaged viruses?

What determines the size of a ssRNA virus?
(I.e., the RNA, or the capsid proteins….?)

What is the size of a large (viral-genome-length) ssRNA molecule?

Let's consider a "typical" ssRNA virus….

cowpea chlorotic mottle virus (CCMV)

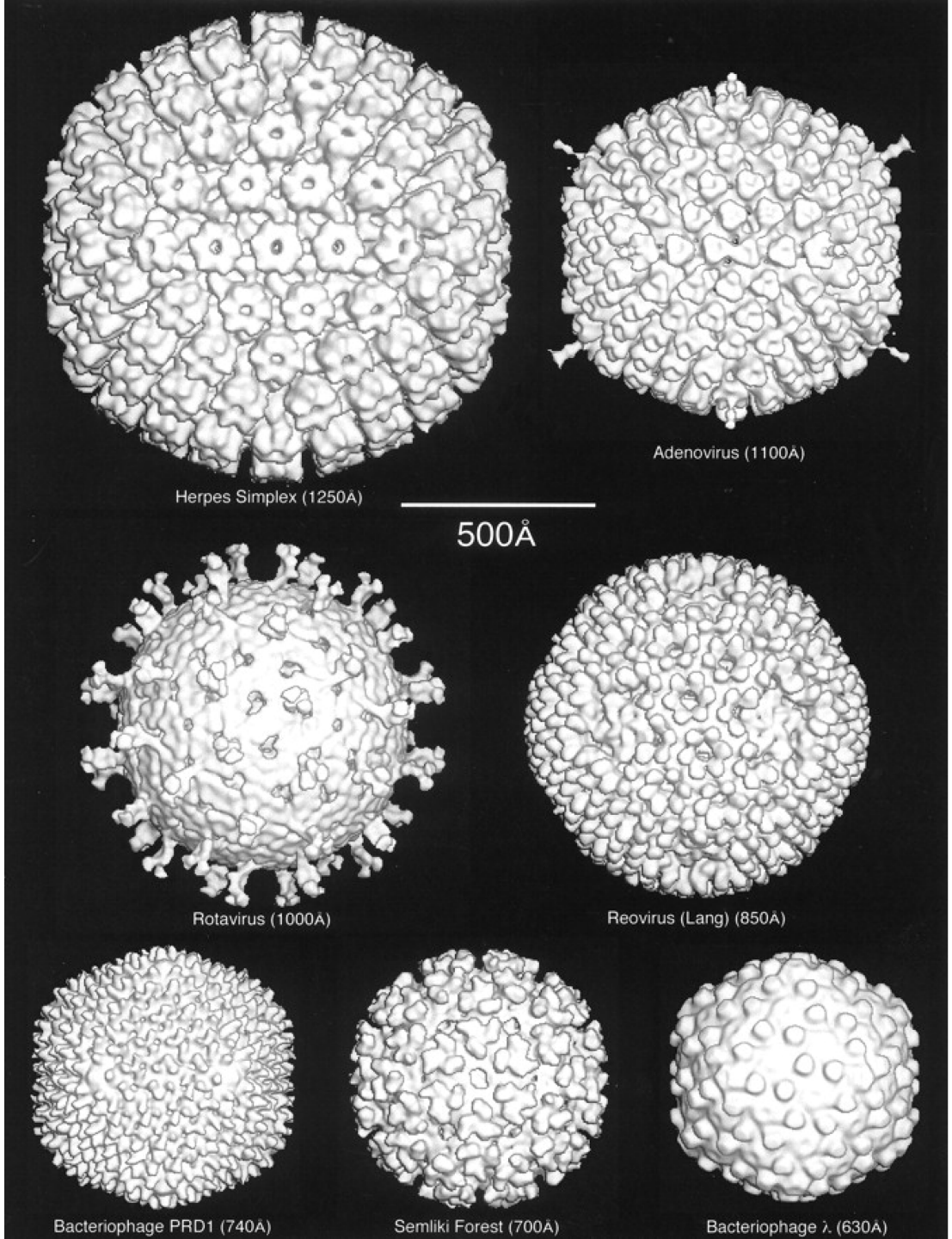


capsid is made up of many copies **(180, in this particular case)** of a single gene product

12 pentamers

20 hexamers

single molecule of ssRNA, inside the (28 nm) capsid -- **icosahedral!**



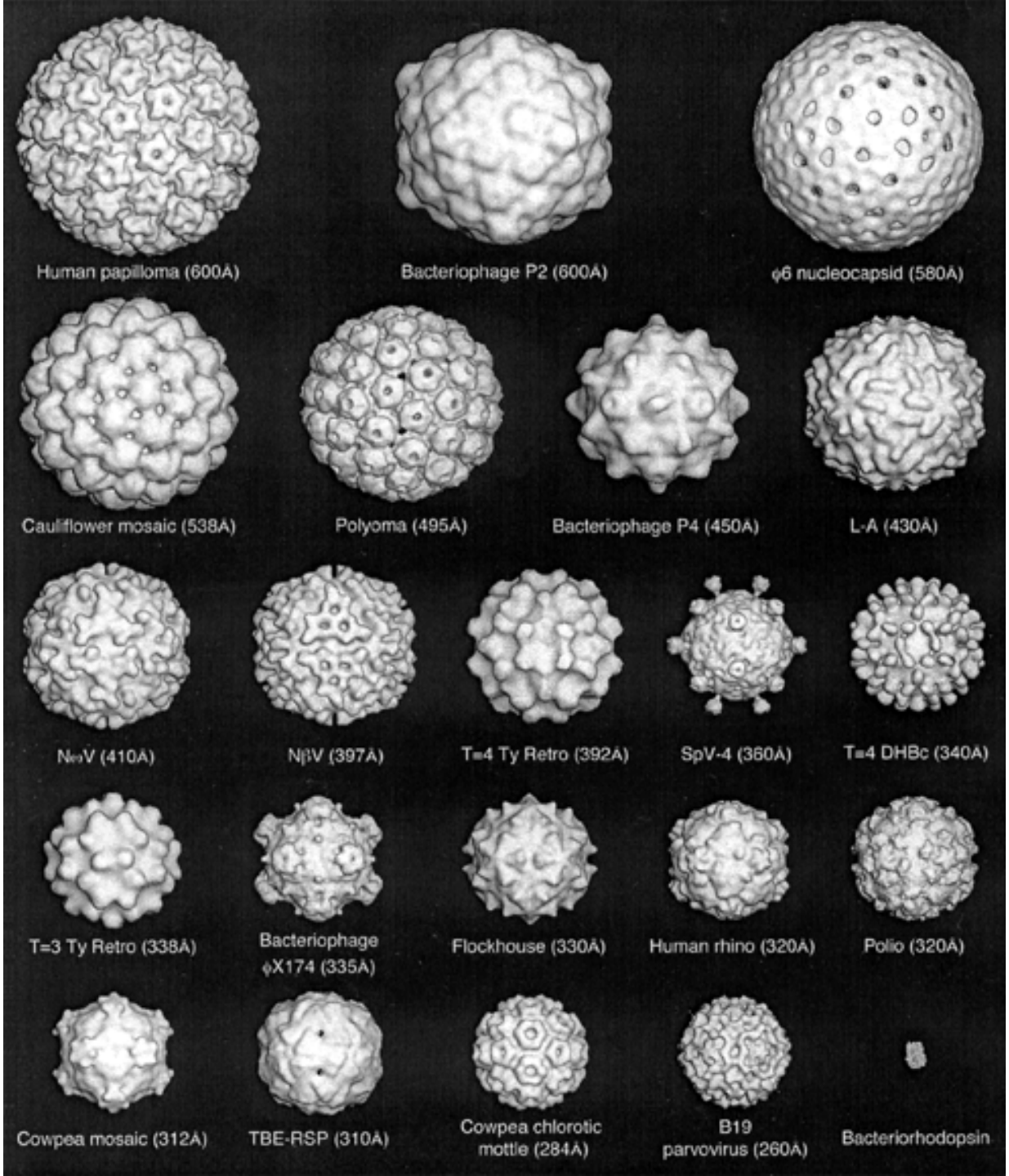
Icosahedral symmetry: ubiquitous among spherical viruses

Caspar & Klug:

Number of hexamers = 10(T-1),

T=1, 3, 4, 7, 9,…

Always 12 pentamers


Human papilloma (600Å)
Bacteriophage P2 (600Å)
φ6 nucleocapsid (580Å)
Cauliflower mosaic (538Å)
Polyoma (495Å)
Bacteriophage P4 (450Å)
L-A (430Å)
NωV (410Å)
NβV (397Å)
T=4 Ty Retro (392Å)
SpV-4 (360Å)
T=4 DHBc (340Å)
T=3 Ty Retro (338Å)
Bacteriophage φX174 (335Å)
Flockhouse (330Å)
Human rhino (320Å)
Polio (320Å)
Cowpea mosaic (312Å)
TBE-RSP (310Å)
Cowpea chlorotic mottle (284Å)
B19 parvovirus (260Å)
Bacteriorhodopsin

Conclusion

We can now describe our hypothesis in a more general manner. We assume that the basic structural requirement for a small virus is the provision of a shell of protein to protect its highly specific packet of ribonucleic acid. This shell is necessarily rather large, and the virus, when in the cell, finds it easier to control the production of a large number of identical small protein molecules rather than that of one or two very large molecules to act as its shell. These small protein molecules then aggregate around the ribonucleic acid in a regular manner, which they can only do in a limited number of ways if they are to use the same packing arrangement repeatedly. Hence small viruses are either rods or spheres. The number of sub-units in a rod-shaped virus is probably unrestricted, but for a spherical virus the number is likely to be a multiple of 12. Every small virus will contain symmetry elements and in favourable cases these can be discovered experimentally.

We believe that this hypothesis is likely to apply (in this form or a simple variant of it) to all small viruses which have a fixed size and shape.

F. H. C. Crick
J. D. Watson*

Physical Principles in the Construction of Regular Viruses

D. L. D. Caspar and A. Klug*
The Children's Cancer Research Foundation, The Children's Hospital Medical Center, and the Department of Biophysics, Harvard Medical School, Boston, Massachusetts;
**Medical Research Council Laboratory of Molecular Biology, University Postgraduate Medical School, Cambridge, England.*

THE FUNCTIONAL ORGANIZATION OF VIRUS PARTICLES

There are two key facts about viruses from which all consideration of their structure and functional organization must proceed. The first is that the essential infective agent of all viruses is a high molecular weight nucleic acid component—either deoxyribonucleic acid (DNA) or ribonucleic acid (RNA). Second, the nucleic acid molecule is contained in a protective package which serves to transmit this infectious agent in a functionally intact state through space and time to a susceptible host.

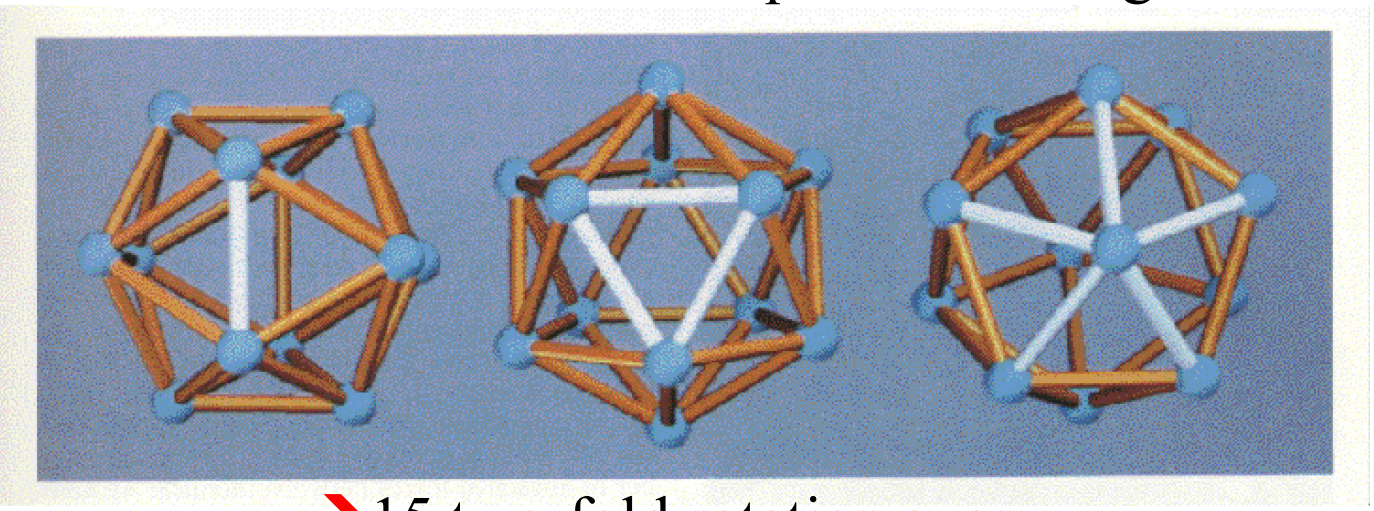
The virus nucleic acid has the capacity of re-directing the synthetic machinery of its host cell to the production of more virus. It is becoming increasingly clear that this control over the cell metabolism can be exerted at a number of different stages of normal biosynthesis. The DNA of large bacteriophages, for example, may pertinently be regarded as a transmissible piece of bacterial chromosome (Luria, 1959). In contrast, the RNA of tobacco mosaic virus and presumably of other RNA viruses, appears to be homologous to the normal messenger RNA of the cell (Matthaei et al., 1962). Indeed, the ultimate classification of many viruses may be primarily in terms of their relation to normal cell constituents.

It is not merely a matter of labeling viruses as DNA- or RNA-containing, but also of distinguishing them in terms of the amount of information carried by the nucleic acid. A complex DNA virus may be able to direct the synthesis of many new enzymes, as well as its own structure protein, whereas a simpler DNA virus may be able to ing ratio is 3:1, could specify at most only two or three different protein molecules. The comparably small tobacco necrosis virus particles (Kassanis and Nixon, 1960, 1961) do not appear to carry complete enough information for their own multiplication, and can only reproduce in association with another, larger tobacco necrosis virus. The molecular weight of the DNA content of vaccinia (Smadel and Hoagland, 1942) and *Tipula* iridescent virus (Thomas, 1961) are both about 150×10^6, which is considerably greater than the DNA content of the small living cells of pleuro-pneumonia-like organisms (PPLO) (Morowitz et al., 1962).

The infectivity of a virus must persist in a latent state outside the host cell. Isolated nucleic acid molecules are very labile, particularly in an intercellular environment containing nucleases. If the virus is to succeed in propagating itself, its nucleic acid must be contained in a protective package. This is achieved by the provision of a protein coat or framework which contains the nucleic acid. It may appear, at first sight, that there is an enormous variety in the ways in which this could be done, judging, for instance, only by the range of morphological variation found in viruses. On the contrary, it is the main thesis of this paper that this is not so. The important point is that there are only a limited number of efficient designs possible for a biological container which can be constructed from a large number of identical protein molecules (Caspar and Klug, 1963). The two basic designs are helical tubes and icosahedral shells. For this reason, the same kind of molecular architecture may turn up in RNA or DNA viruses infecting animals, plants, and bacteria.

Cold Spring Harbor Symposia on Qunatitative Biology, Vol. XXVII, 1962

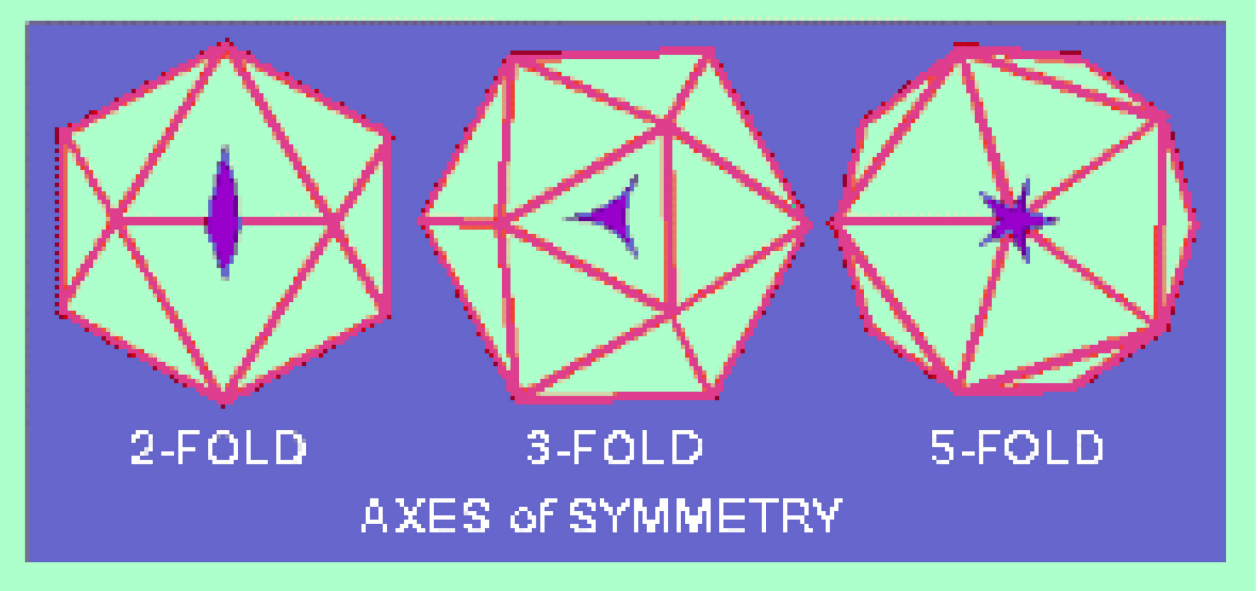
An **icosahedron** has 20 identical equilateral triangular faces



➔ 15 two-fold rotation axes

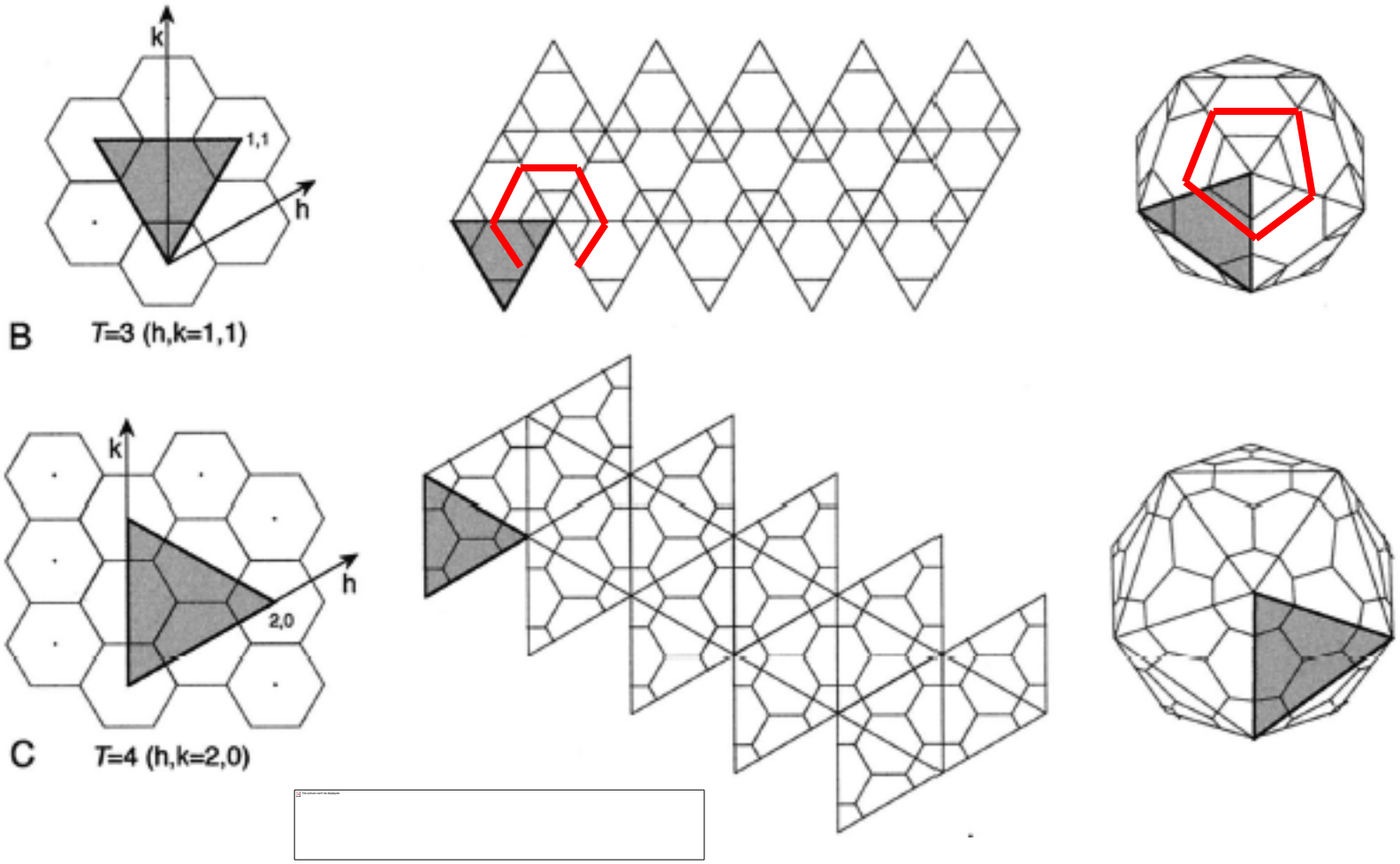
➔ 10 three-fold rotation axes

➔ 6 five-fold rotation axes



Caspar and Klug (1962), following Crick and Watson (1956):

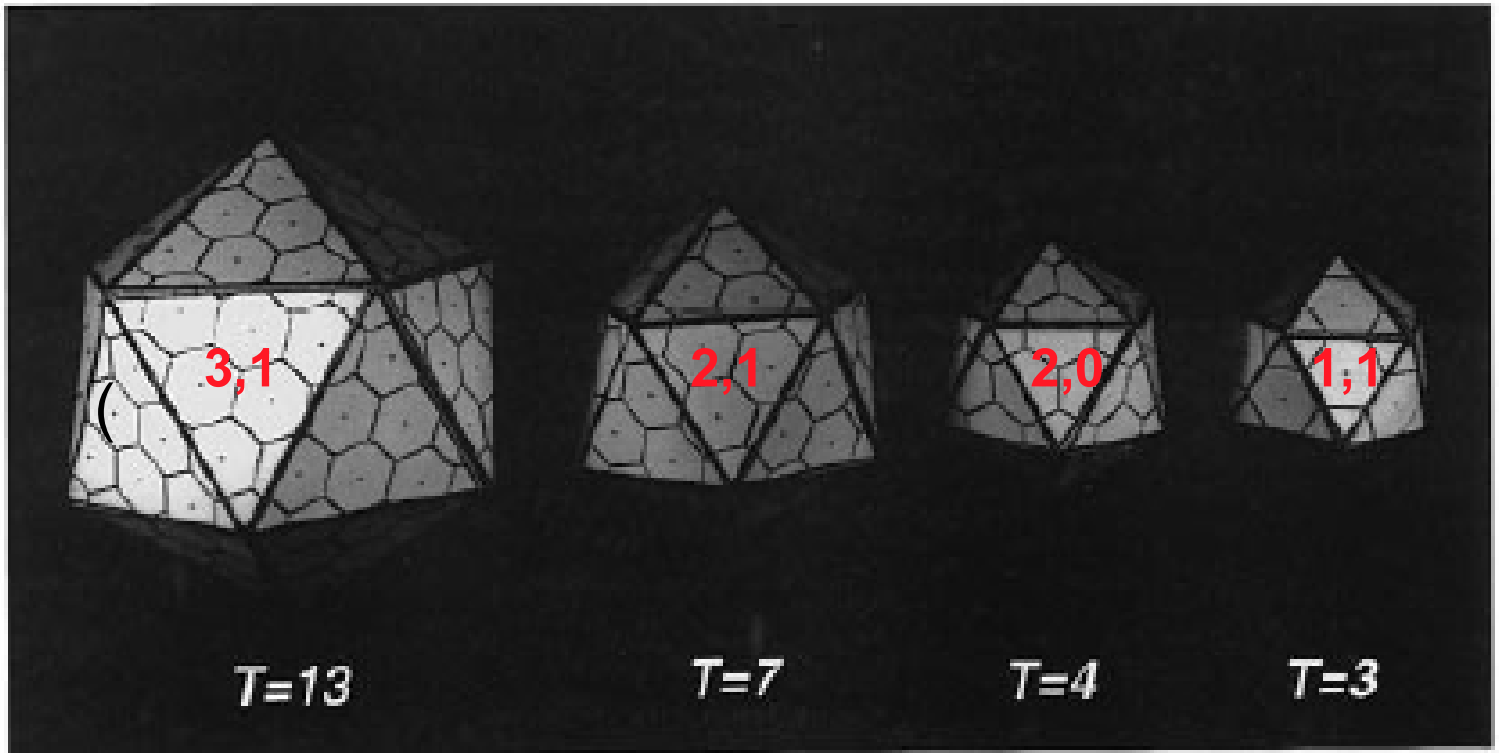
HIERARCHY OF ICOSAHEDRAL SHELLS



= # of inequivalent subunit (protein) positions

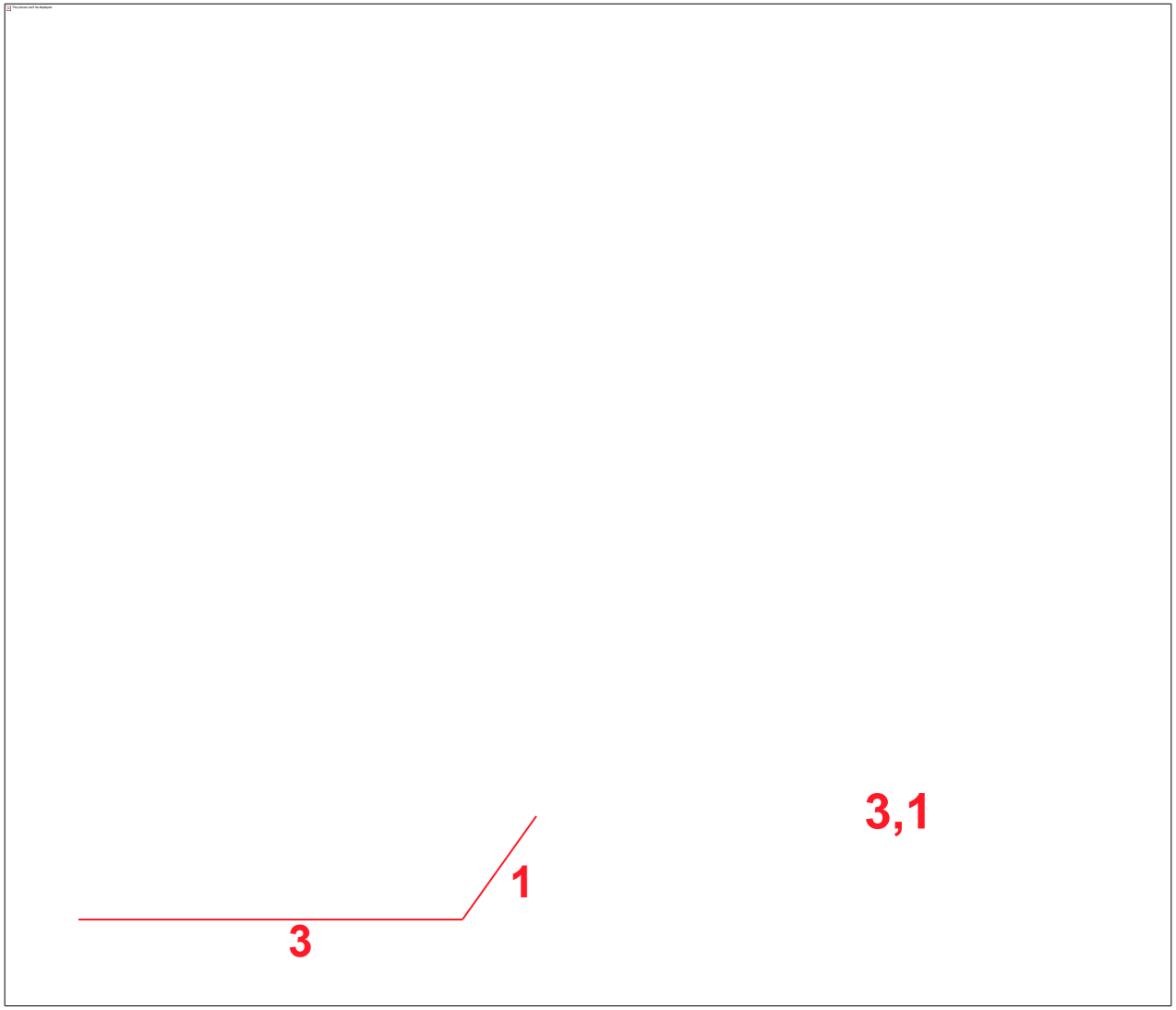
Q: What are the ways in which we can "wrap" a planar hexagonal lattice onto icosahedra of different sizes, **without straining any of the bonds**?

A: Fold it along equilateral triangles whose vertices are the centers of hexagons.



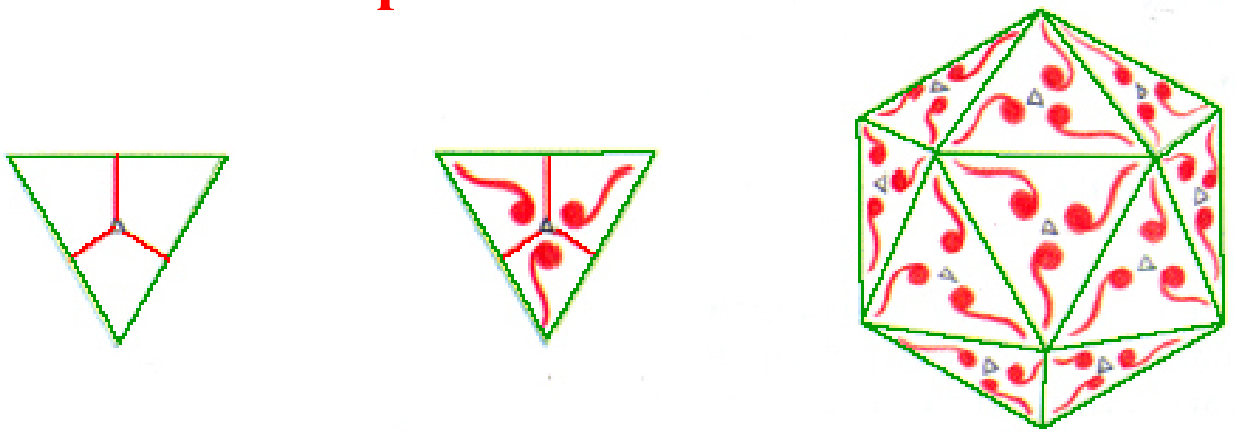
In this way we create shells with increasing values of the **minimum number (T) of inequivalent positions**.

Construction of a (h,k) = (3,1) Caspar-Klug structure:



The simplest virus has a shell of 60 protein subunits

Capsid of Human Papilloma Virus

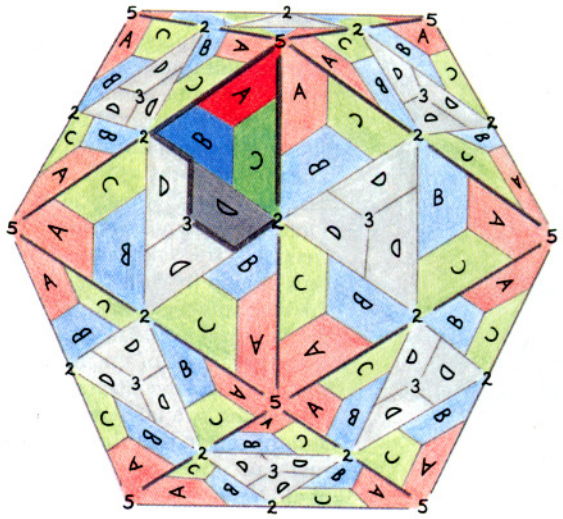


There are three asymmetric subunits -- proteins -- on each triangular face, and all of the 60 subunits are equivalent to one another (protein$_{60}$!)

Capsid contains 12 pentamer groupings

Complex viruses have more than one subunit in each icosahedral position

Here there are 4, no longer occupying equivalent sites on the shell



Number of *in*equivalent sites

= *T*riangulation number

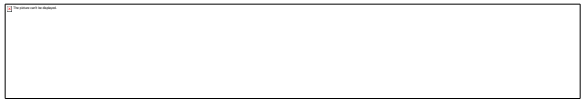
- The number of subunits is a multiple of 60
- Only certain multiples (3, 4, 7, …) of 60 are expected to occur

(Caspar and Klug):

with h and k both integers

herpes virus capsid

(pentons are colored in red)



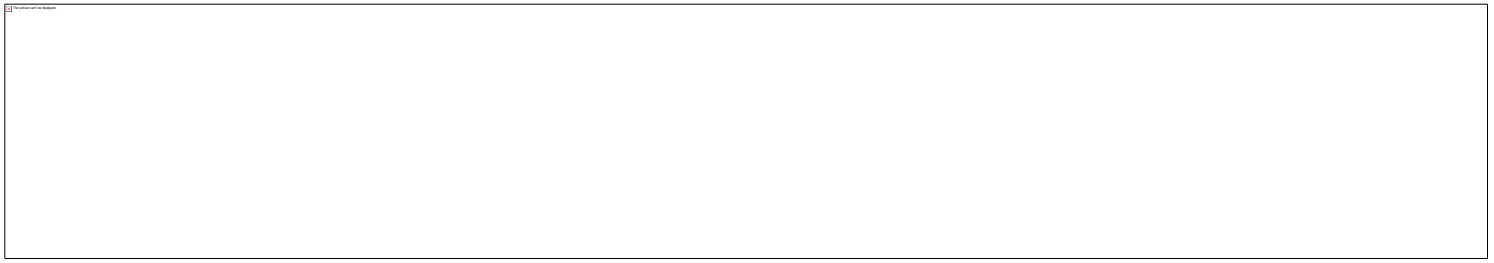
W. Chiu and F. J. Rixon, Virus Research **82**, 9 (2001)

N, number of protein subunits

$$= 5\ \#_{pentamers} + 6\ \#_{hexamers}$$

$$= 5 \times 12 + 6 \times [10(T-1)] = 60\ T$$

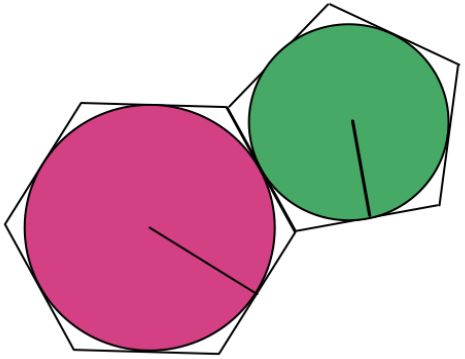
T = 1, 3, 4, 7, 9, ….



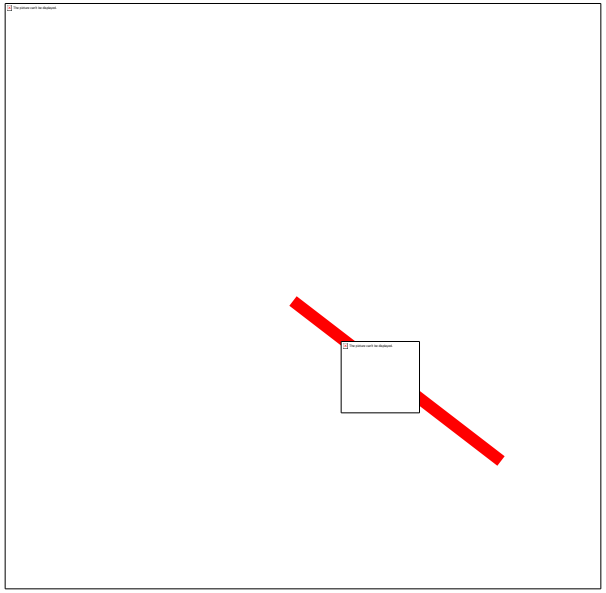
SIZE OF CAPSID (R_{capsid}) INCREASES AS 

and is discretized…!

Minimal model for self-assembly of isotropic multimers ("capsomers") of capsid proteins



Capsomers can adopt two different sizes, related by pentagon/hexagon geometry, and interact with energy

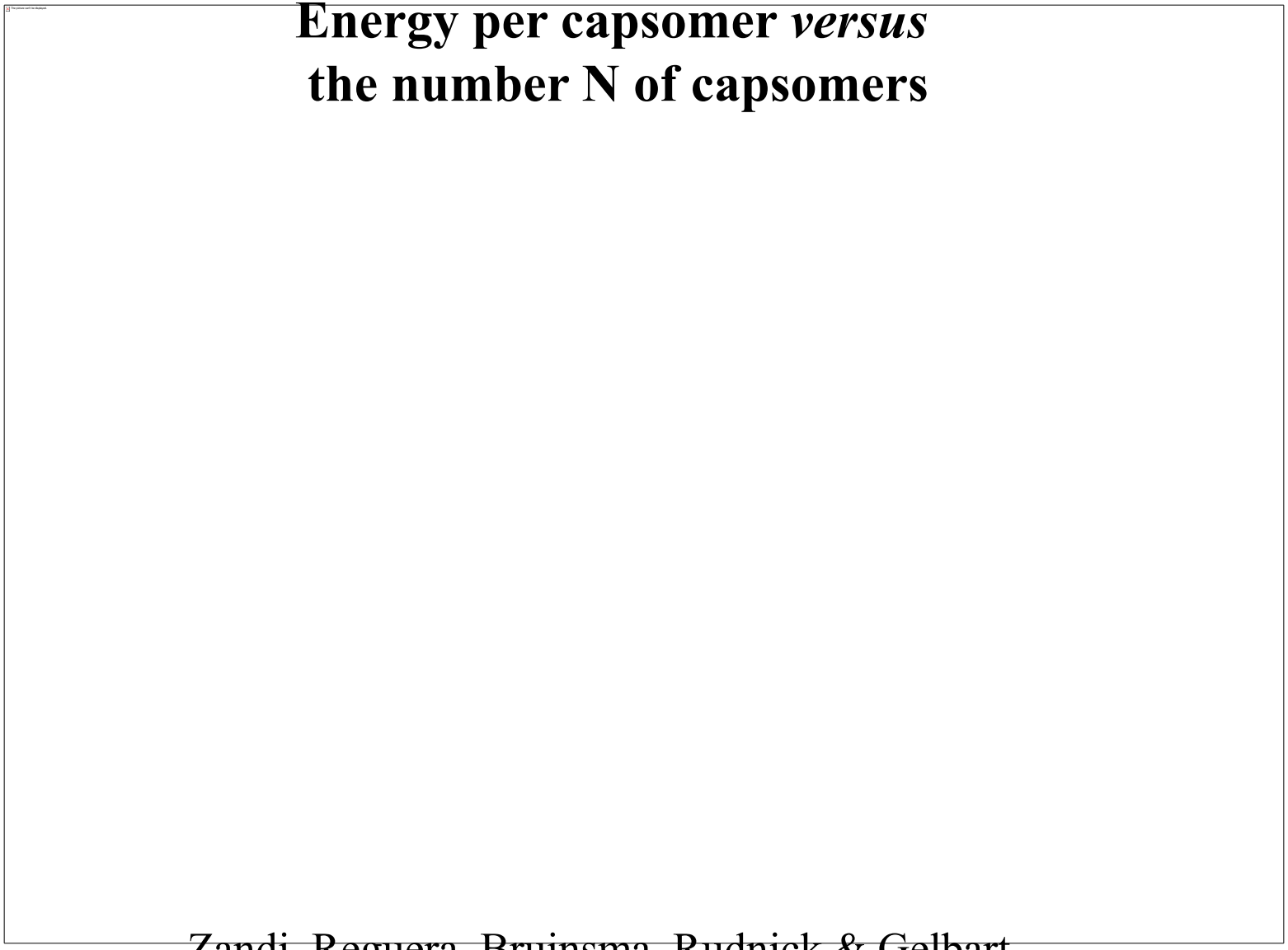


Monte Carlo simulation

- N interacting capsomers are allowed to range over a spherical surface (R) while switching between P and H states
- $N=N_P+N_H$ is total number of disks

Internal energy, E(R), is evaluated for each of a range of equilibrated sphere radii R and then minimized with respect to R

Energy per capsomer *versus* the number N of capsomers

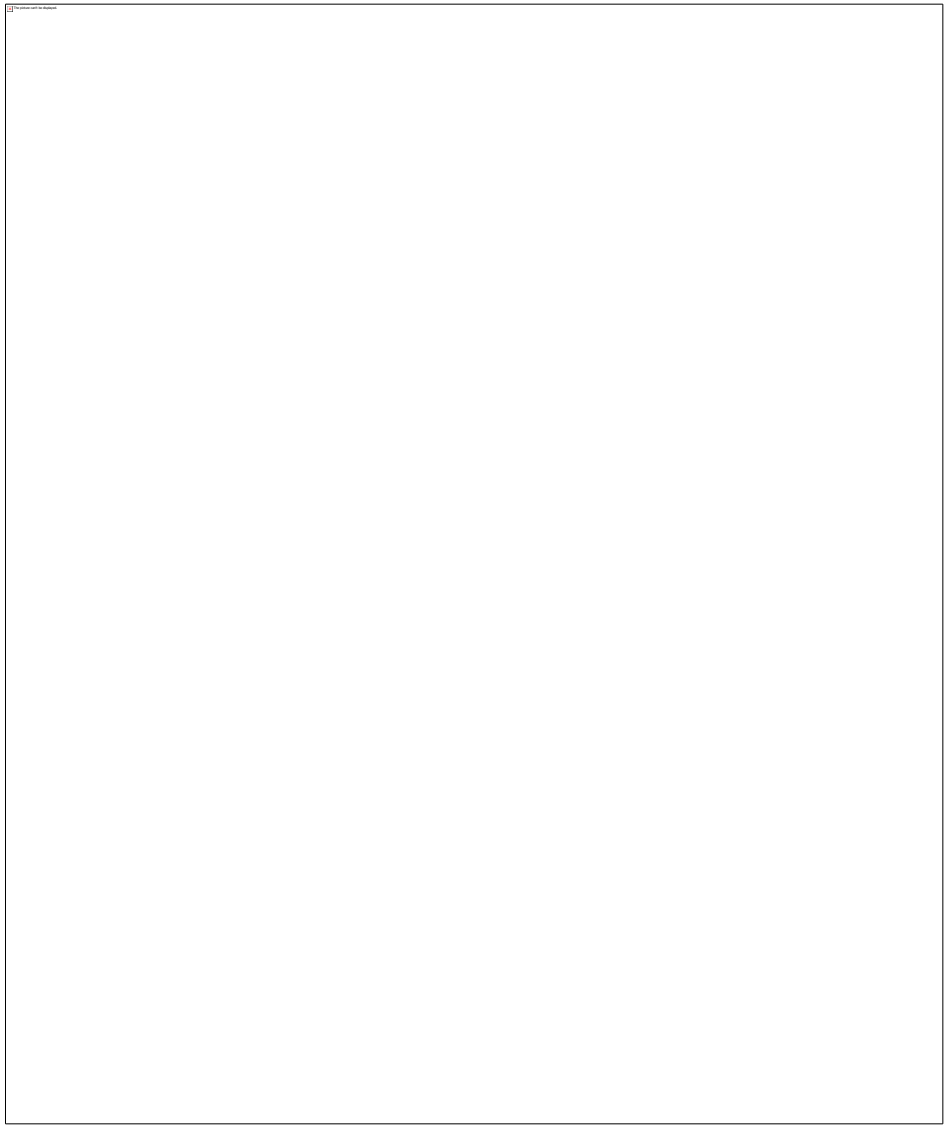


Zandi, Reguera, Bruinsma, Rudnick & Gelbart

Phys. Rev. Lett. **90**, 248101 (2003); Proc. Nat. Acad. Sci. USA **101**, 15556 (2004)

We recover the T-numbers as free energy minima

Capsid structures associated with the minima of energy correspond to those of Caspar-Klug



WHAT DETERMINES THE SIZE OF A VIRUS?

WHAT IS THE RELATIONSHIP BETWEEN

CAPSID SIZE AND GENOME SIZE?

(1) ***WHAT DETERMINES THE SIZE OF A CAPSID***?

(2) " " " " ***OF A GENOME***?

QUESTION (1) 

SELF-ASSEMBLY OF PROTEINS IN 2D CURVED SPACE

QUESTION (2) 

CONFIGURATIONAL STATISTICS OF DNA AND RNA

Recall that size of dsDNA bacterial virus is determined by close-packed volume of its genome; it follows that virus (virion) volume scales with MW of genome.

E.g., T4 genome is (169kbp/19.3kbp) = 8.8 x longer than ϕ29's and its capsid volume is $(92nm/44.1nm)^3$ = 8.9 x larger

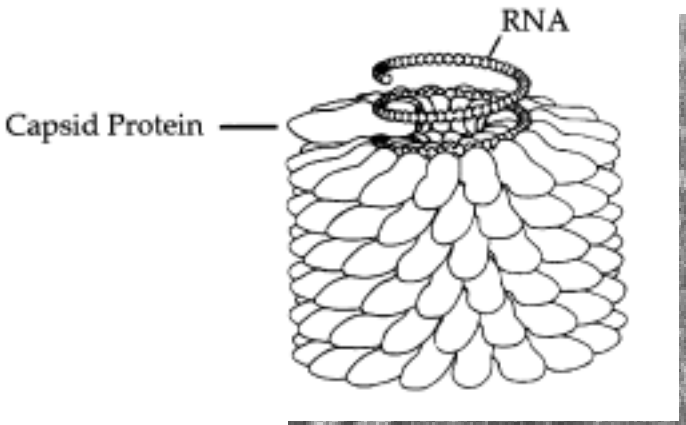
In fact, other dsDNA phage genomes show the same scaling….

Virus	Type	Diameter (nm)	Genome (kbp)	Efficiency (%)	Spacing (nm)	Rgyr (um)
lambda wt	dsDNA	55	48.5	59.5	2.47	0.91
lambda b221 cI26	dsDNA	55	37.4	45.9	2.81	0.80
T7	dsDNA	55	40	49.0	2.72	0.82
N4	dsDNA	66	72	51.1	2.66	1.11
PRD1	dsDNA	42	15	41.3	2.96	0.50

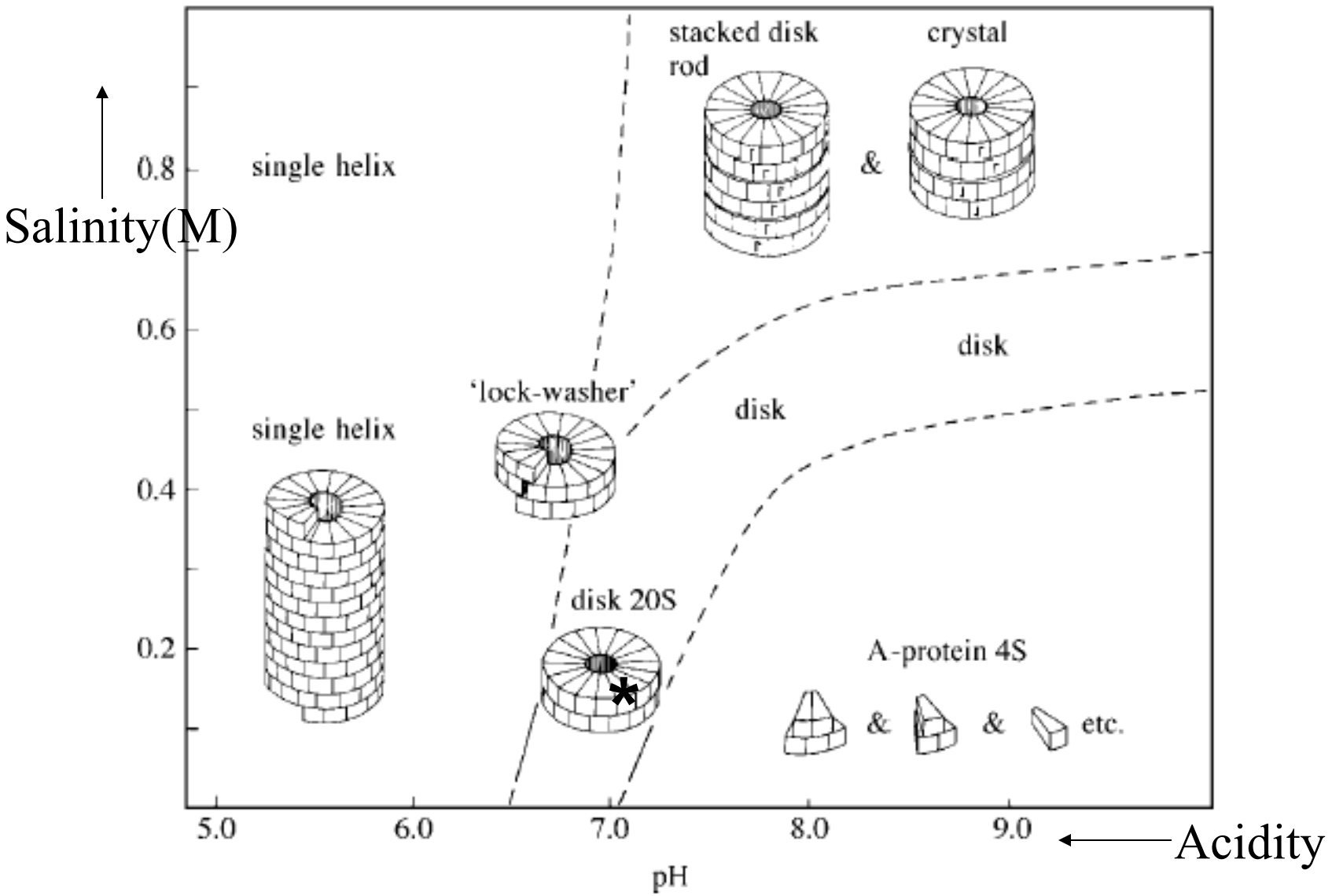
BUT ssRNA VIRUSES ARE VERY DIFFERENT

Let's consider *self-assembling* ssRNA viruses….

First, …, tobacco mosaic virus (TMV)



TMV capsid protein aggregates **with a single, cylindrical, curvature**



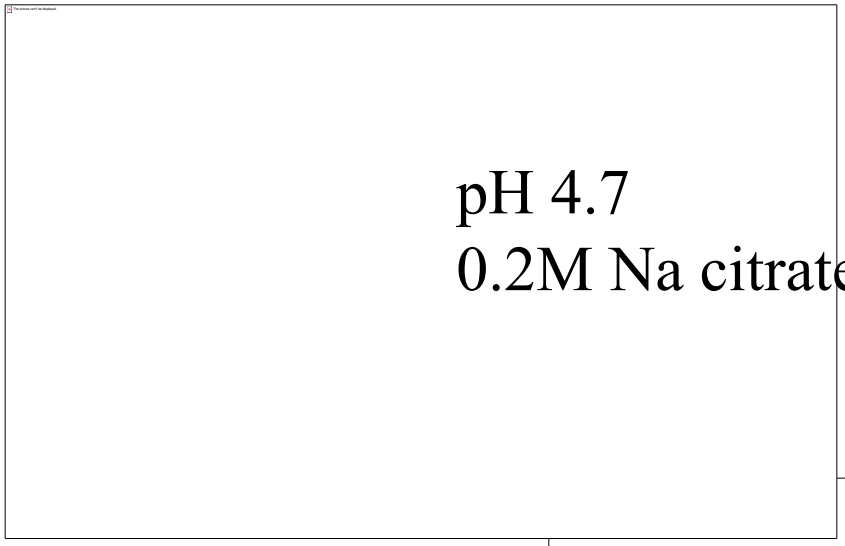
TMV capsid protein (CP) will form cylindrical virions (with same diameter -- 20 nm -- as TMV virus) when combined with ***any*** RNA, e.g., TYMV RNA whose nucleotide length is about the same (6400 nt) but which forms spherical capsids of diameter 30 nm with its own CP, or with CCMV RNA whose nucleotide length is half:
in each case the RNA length determines the length of the 20nm-diameter *cylindrical* protein shell

Also, TYMV CP will form 30nm-diameter *spherical* capsids when combined with TMV RNA!

COMPETITION BETWEEN CPs SELF-ASSEMBLING WITH PREFERRED CURVATURE, AND RNA BEING PACKAGED WITH PREFERRED SIZE AND SHAPE….

But **CCMV capsid protein alone (no RNA) forms *both*:**

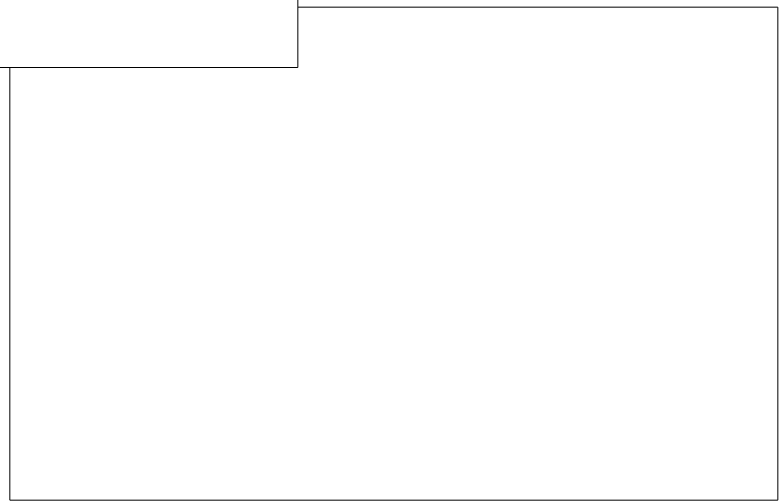
spherical capsids (empty)



pH 4.7
0.2M Na citrate

and other shapes…

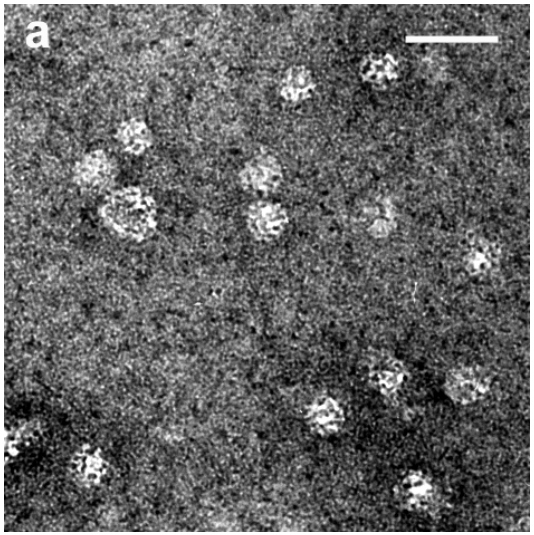
pH 6
0.01M Na cacodylate



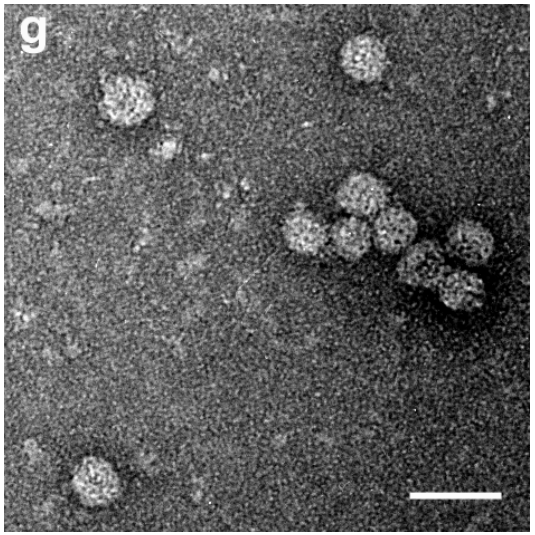
Can map out the "phase diagram" of a virus…

And control “sphere” sizes…

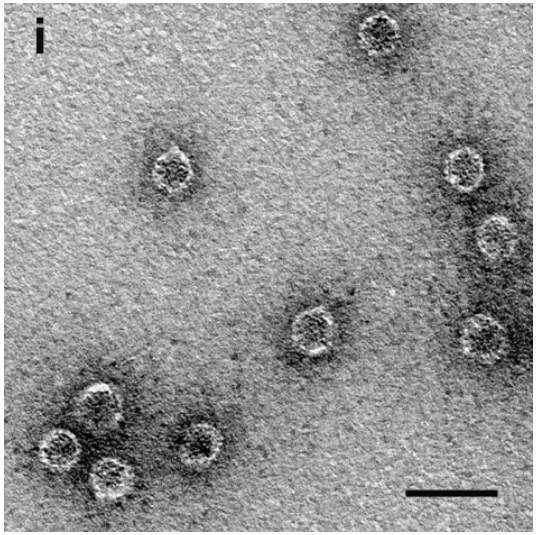
PSS 400 kDa



PSS 3.4 MDa



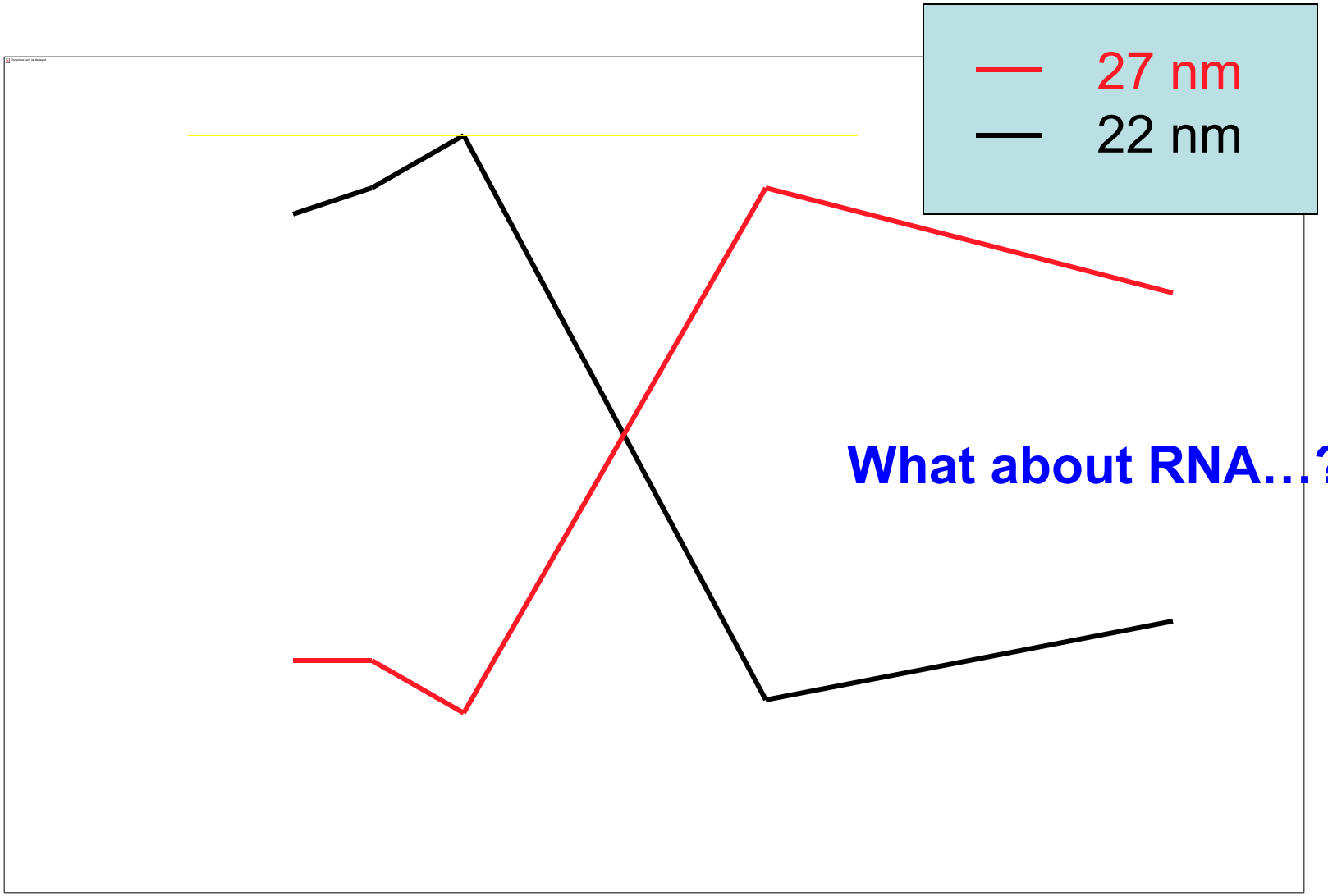
Empty capsids



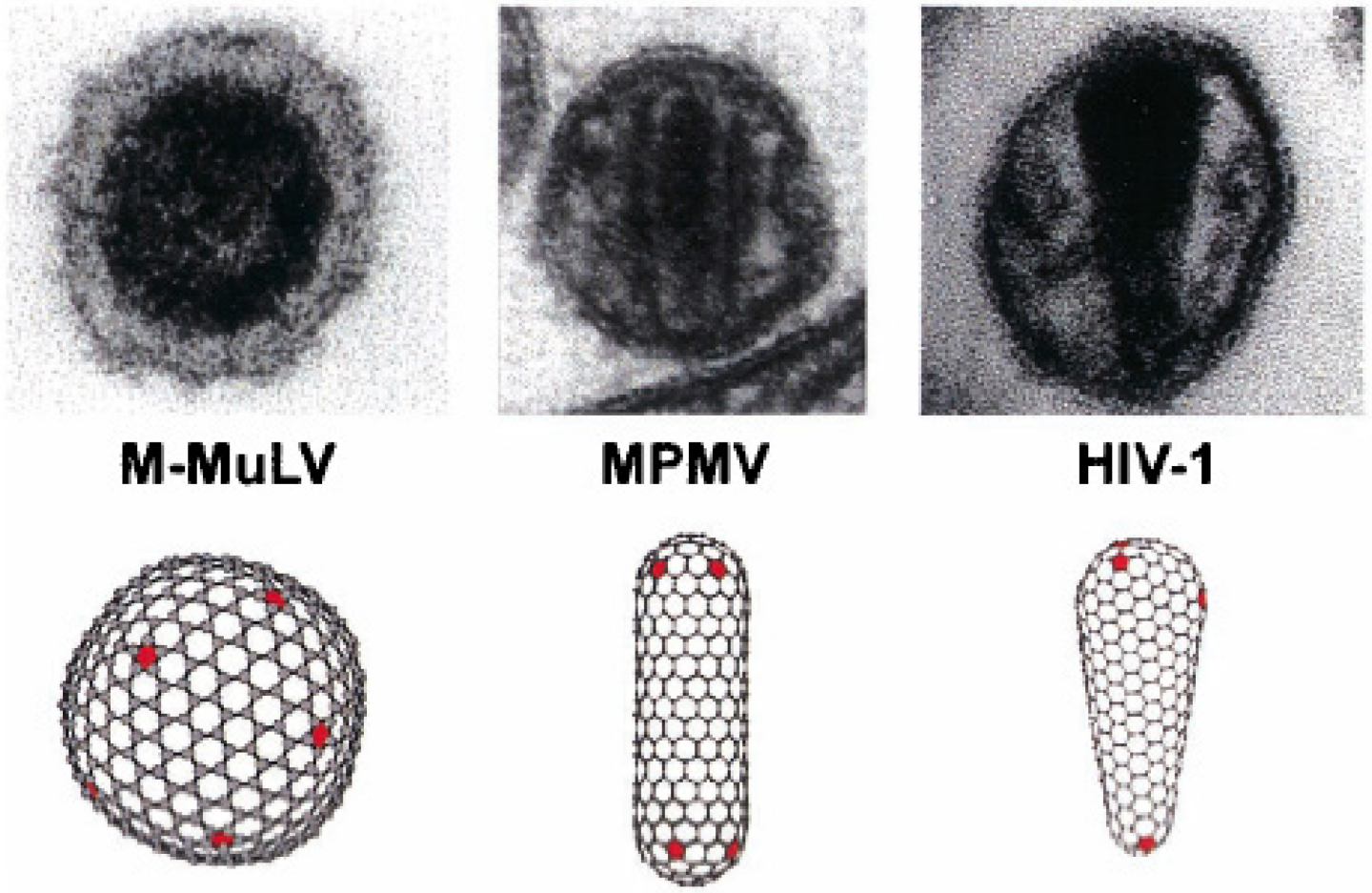
TEM images of Virus-like Particles (VLPs) formed from CCMV capsid protein and polystyrenesulfonate (PSS) of different molecular weights

Magnification is 66,000x; Scale bars are 50 nm

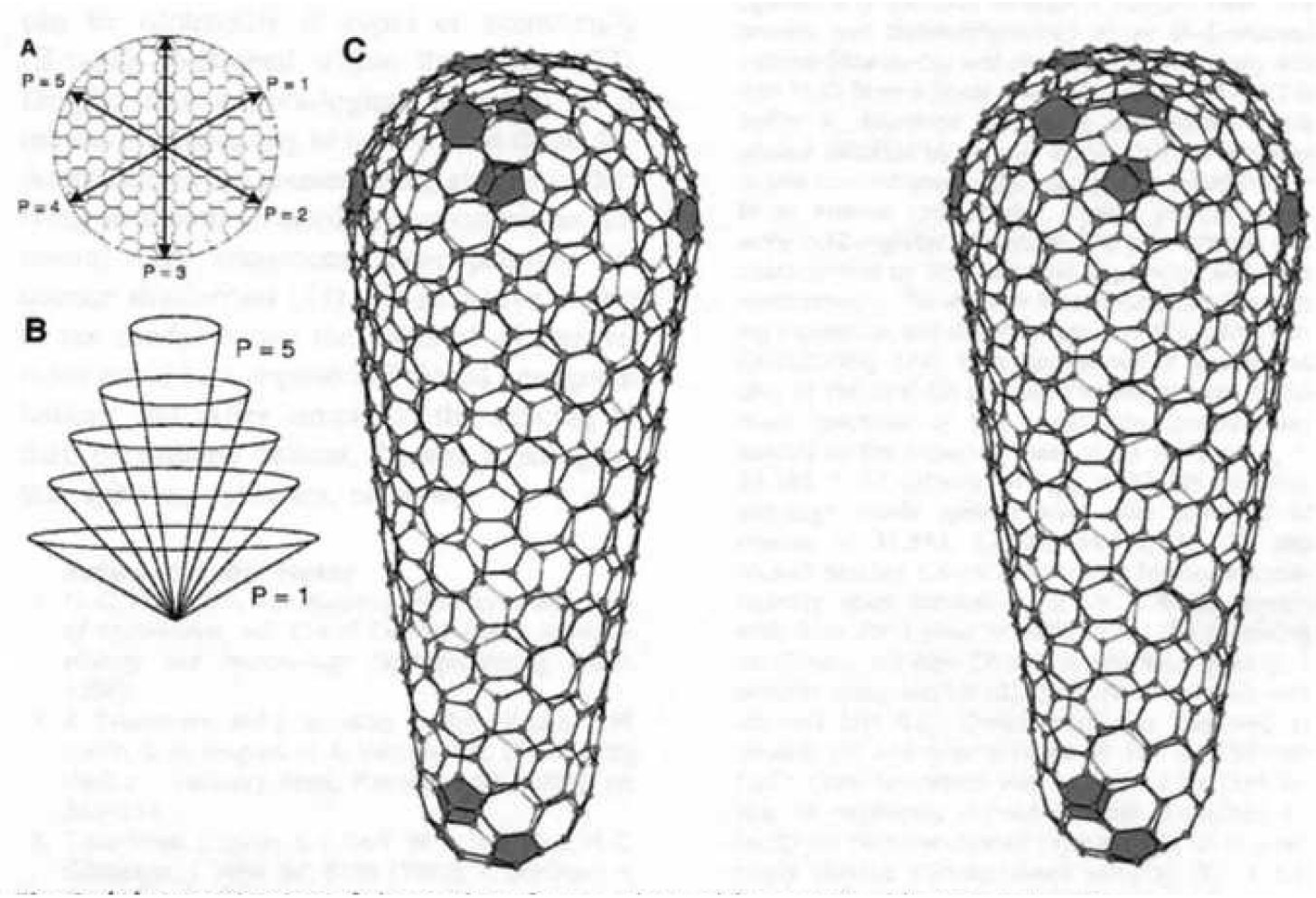
Higher MWs of polymer give larger protein shells…



Shapes can also be of lower symmetry…

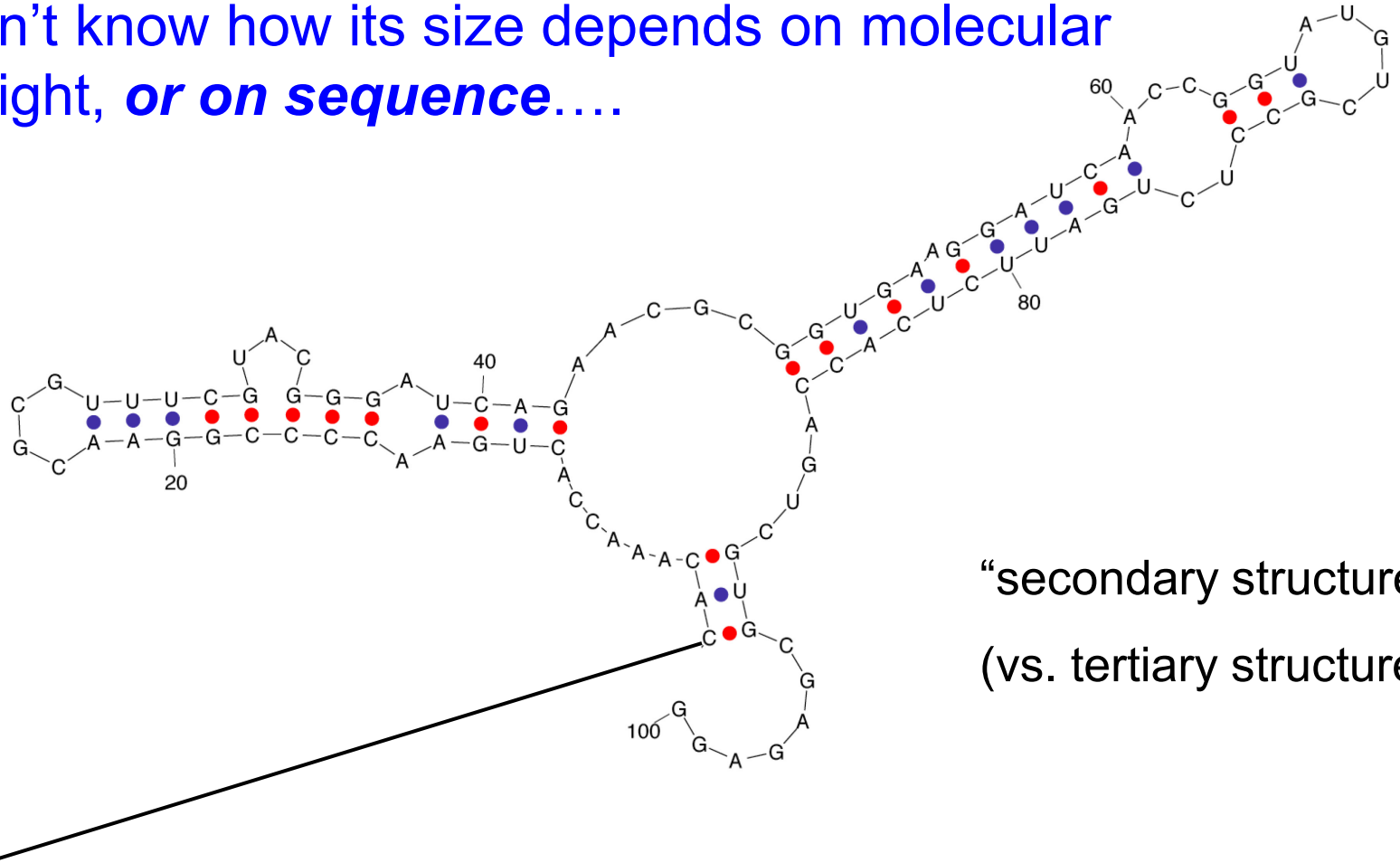


HIV capsid: CLOSED TRUNCATED CONE (!)



Problem: RNA is not a linear polymer, nor a simple branched one…

Don't know how its size depends on molecular weight, ***or on sequence***….



"secondary structure"

(vs. tertiary structure)

(5')CACAAACCACUGAACCCCGGAACGCGUUUCGUACGGGAUCAGAACGCGGUGAA
GGAUCAACCGGUAUGUCGCCUCUGAUUCUCACCAGUCGUGCGAGAGG(3')

Secondary structures can be determined for small enough RNA molecules (up to lengths of 100's of nucleotides), *via* both empirical energy minimizations and chemical analyses

Tertiary structures can also be determined for these small molecules (e.g., ribozymes), by X-ray and NMR…

But these approaches do not work for either secondary or tertiary structures of significantly larger RNA molecules (e.g., viral RNA genomes that are 1000's to 10000's of nt)

So we begin by asking more *coarse-grained* questions about the 3D size of a viral RNA molecule, and how it correlates with secondary structure of the nt sequence…

Work of Aron Yoffe and Ajay Gopal….

CCMV

3000 nt

2800 nt

2100 nt

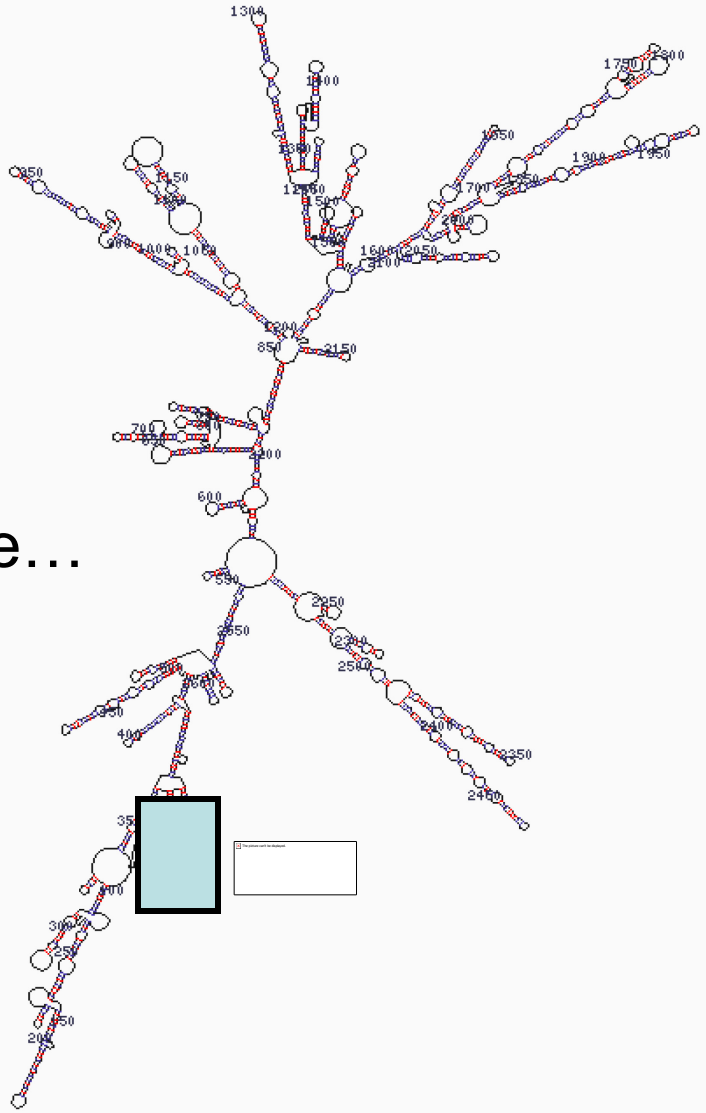
900 nt

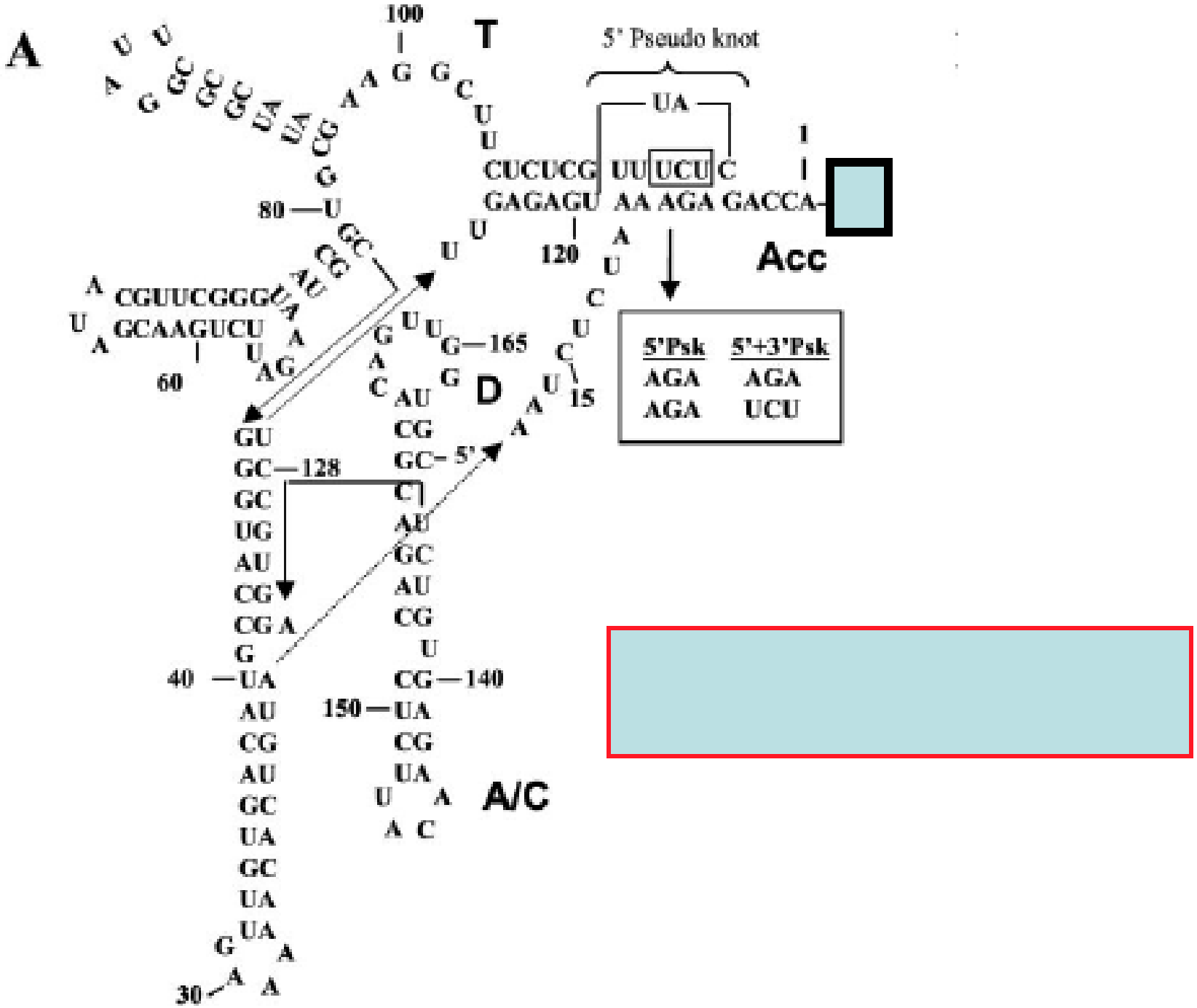
(about 3000 nt in each capsid: 1, 2, and 3+4…)

What is the 3D size and shape of a (viral-length) RNA molecule?

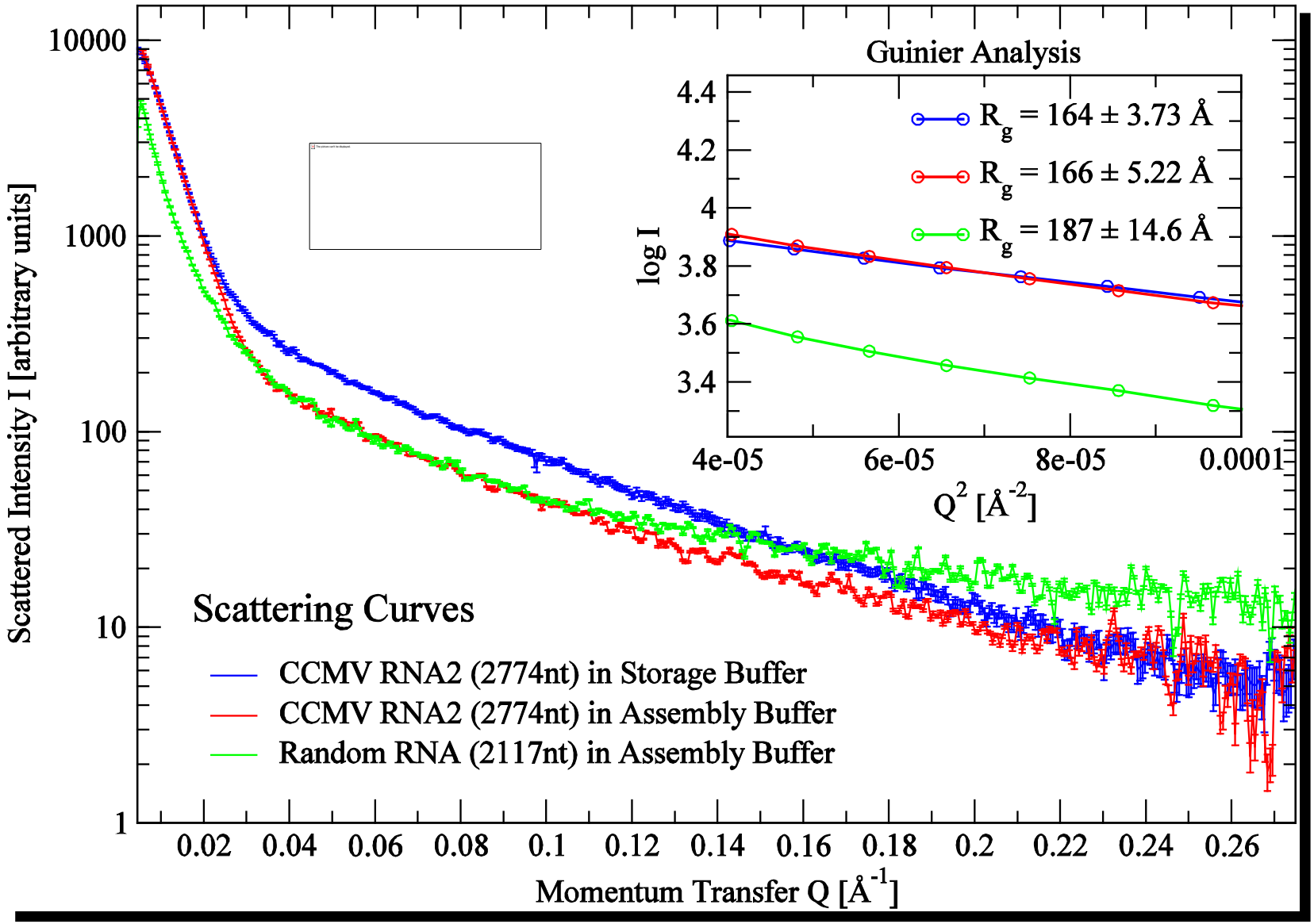
E.g., the second gene of CCMV, which is about 2800 nt long

Start with its secondary structure…

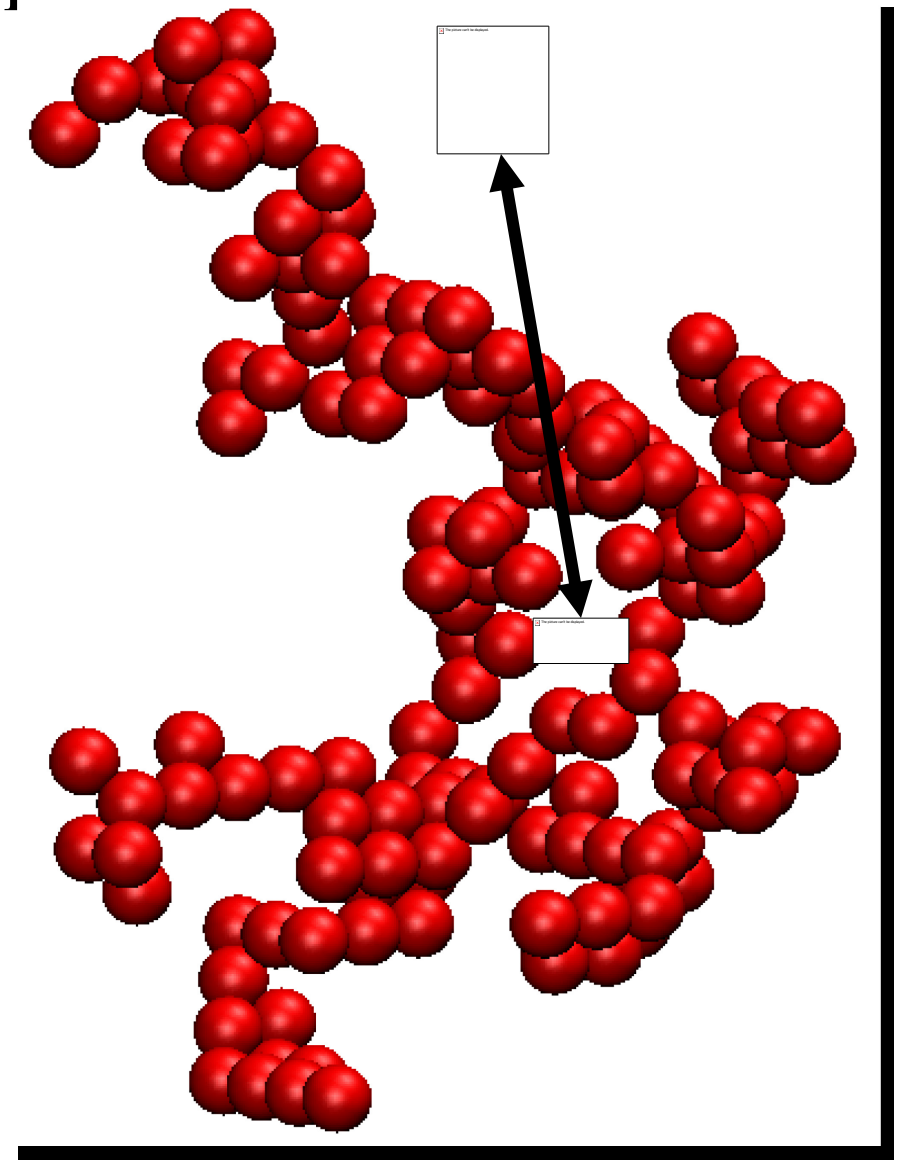



A
T
D
Acc
A/C
5' Pseudo knot
UA
1
100
80
120
60
165
15
128
5'
40
150
140
30
5'Psk
5'+3'Psk
AGA
AGA
AGA
UCU

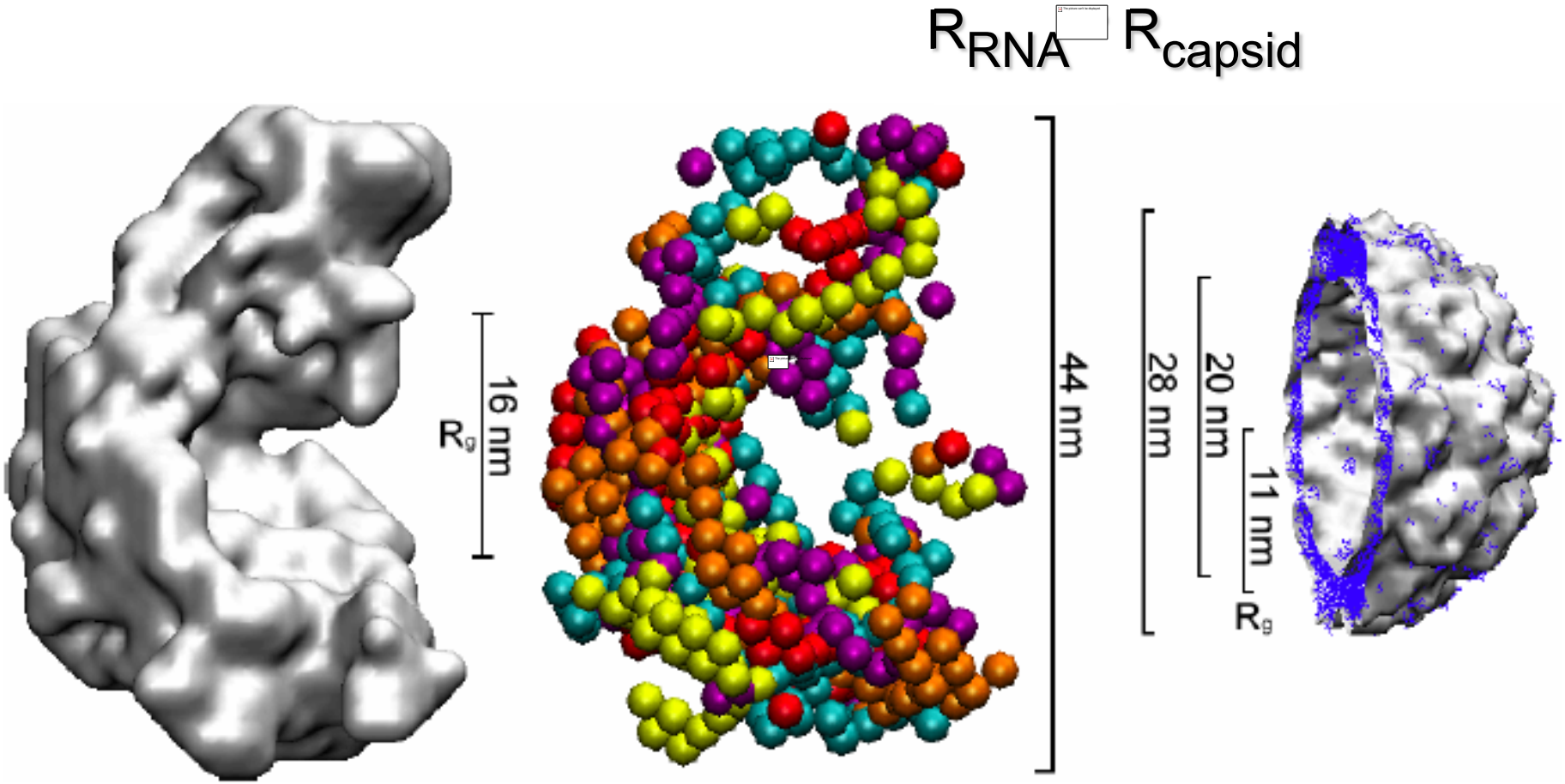
Need to go to classic experiment to determine 3D size & shape…

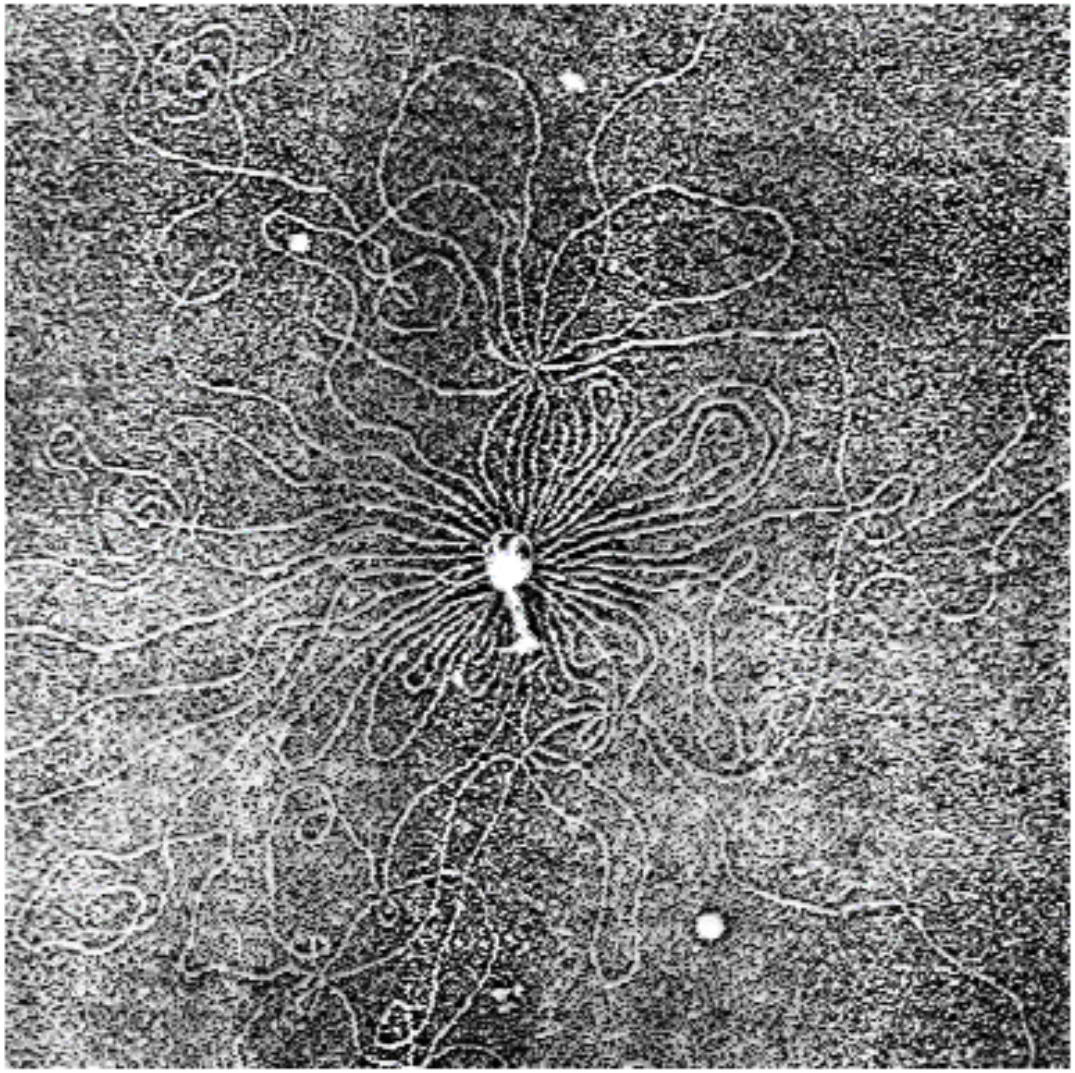


3D coarse-grained model fit to CCMV RNA2 scattering [I(Q)] data



Real-space image reconstruction of CCMV RNA2 from small-angle synchrotron X-ray scattering I(q)





Osmotically shocked
T2 bacteriophage

$R_{DNA} >> R_{capsid}$

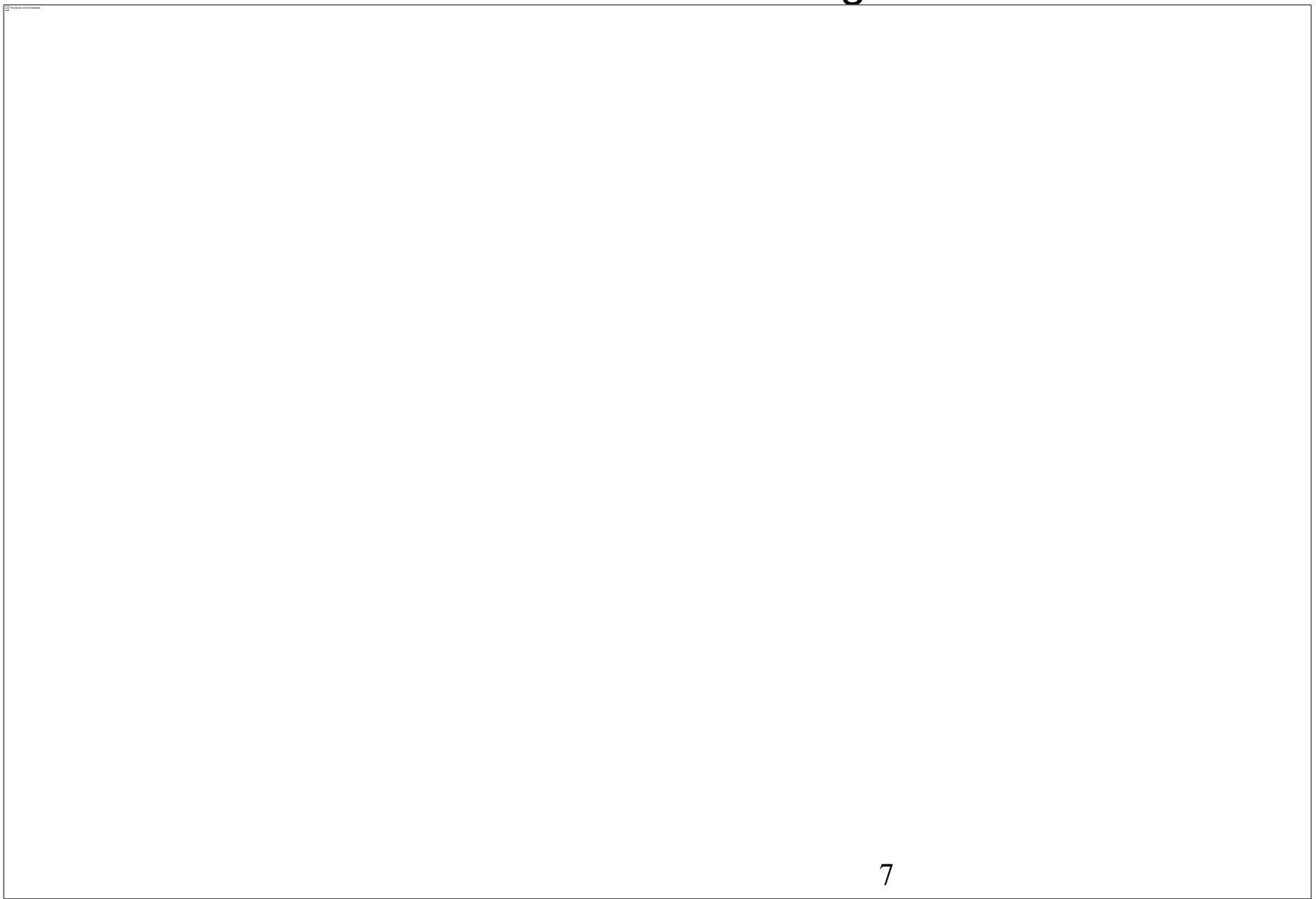
Kleinschmidt et al.

Comparison of Percent Base Pairing and Average Duplex Length in Viral, Random, and Ribosomal RNA

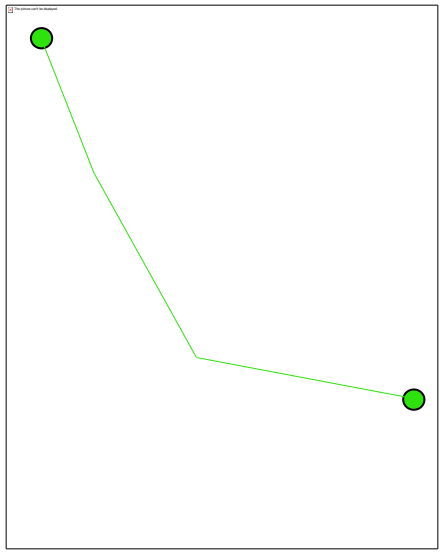
- Viral genome sequences have evolved ***not only*** to yield optimal gene products, ***but also*** to have a size and shape that ensures they will be encapsidated.
- Therefore, if the above coarse-grained characteristics are predictive for size, their values for random, viral, and ribosomal RNA should differ markedly. **BUT THEY DON'T.** **E.g., base pairing essentially always 60%.**
- **The distributions of duplex lengths are also the same for all three RNAs.**

CHARACTERISTIC	AVG. OF 25 2800-BASE RANDOM RNAs	AVG. OF 4 VIRAL RNAs (CCMV& BMV RNAs 1+2)	*E. Coli* ribosomal RNA, 2825 BASES
% OF BASES IN PAIRS	60.8 ± 1.1	63.5 ± 1.9	62.0
AVERAGE DUPLEX LENGTH	4.4 ± 0.1	4.7 ± 0.1	4.4

Distribution of Duplex Lengths Does Not Differ Markedly Between Random and Biological RNAs



There is a Large Difference in Maximum Ladder Distance (MLD) Between Random and Viral RNA

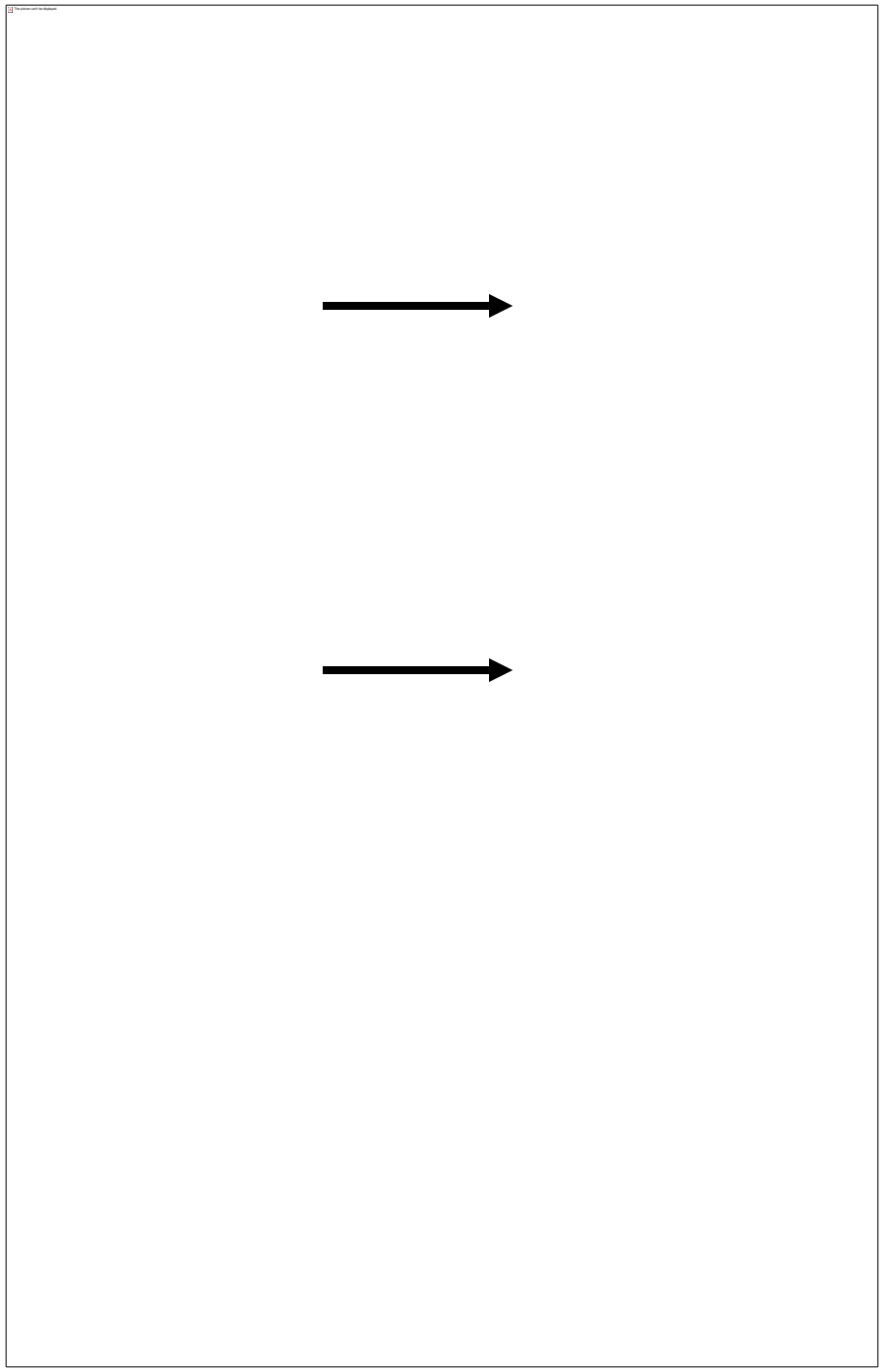


The MLD in this structure is 23, the number of base pairs crossed by the path connecting the two green dots.

- Bundschuh and Hwa have introduced the "ladder distance" as a quantitative measure of size in RNA secondary structures; h_{ij}, is the number of base pairs that are crossed by the shortest path connecting bases i and j.
- To characterize the size of RNA secondary structures using a single quantity, we calculate the "**maximum ladder distance**" -- the largest value of h_{ij} for all combinations of i and j -- **it is the longest path across the secondary structure**.

CHARACTERISTIC	AVG. OF 25 2800-BASE RANDOM RNAs	AVG. OF 4 VIRAL RNAs (CCMV RNAs 1+2, BMV RNAs 1+2)	*E. Coli* RIBOSOMAL RNA, 2825 BASES
MAX. LADDER DISTANCE	291.0 ± 38.1	206.4 ± 8.3	235.9

The picture can't be displayed.



maximum ladder distances (MLDs)
of selected 2117-nt sequences
from Yeast Chromosome 12

average yeast MLD

average random MLD

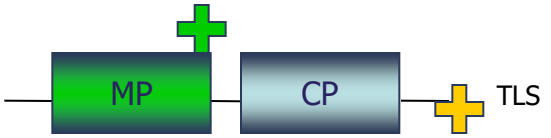
414th 2117-nt Sequence
from Yeast Chromosome 12
(MLD = 148, 58% base-paired)

326th 2117-nt Sequence
from Yeast Chromosome 12
(MLD = 368, 61% base-paired)

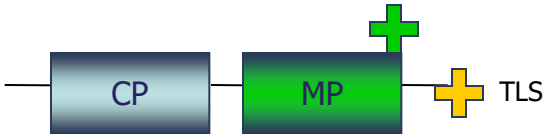

The picture can't be displayed.

Sequence-Dependence of RNA Size and Shape

BMV RNA3, 2117nt
(packages)



BMV RNA3Rev, 2117nt
(does not package)



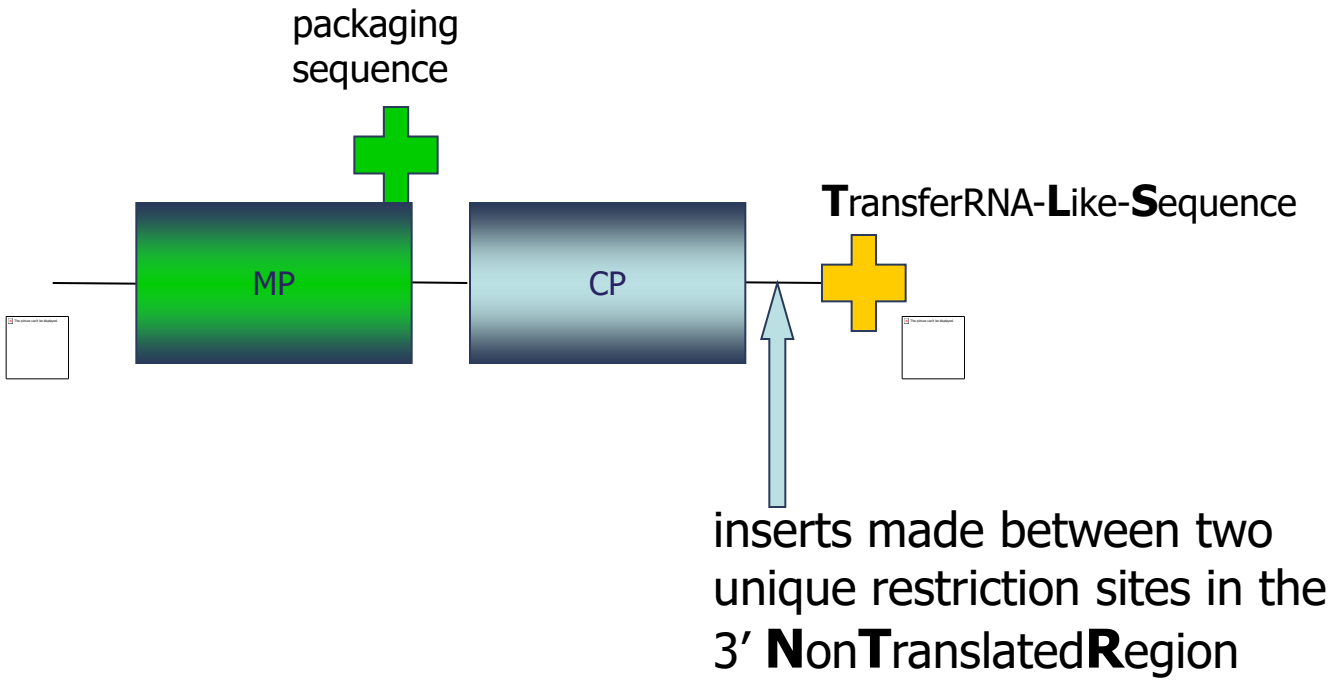
Control Sequence, 2117nt
("random" Yeast RPS22B)

COMPARE AND CONTRAST TO

CCMV RNA2, 2774nt

Length-(Nucleotide-)Dependence of RNA Size and Shape

Make "added-length" mutants of 2117 nt-long BMV genes 3 + 4



Measure R_g (and *in vitro* and *in vivo* packaging efficiencies…)

Experimental program for determining RNA genome size and its effect on packaging efficiency and viral infectivity

USING " PHYSICAL MUTANTS" OF 2117-nt, wild-type, BMV RNA (I.E., SEQUENCE-SCRAMBLED, AND LENGTHENED):

1) Measure RNA R_g as function of nucleotide sequence and of overall length

2) In buffer with purified capsid protein, measure *in vitro* packaging efficiency as function of same

3) Transfect host cells and measure *in vivo* yields of:

replicated RNA

infectious nucleocapsids

GELBART-KNOBLER GROUP

Aron Yoffe

Li Tai Fang

Odisse Azizgolshani

Ajay Gopal

Yufang Hu

COLLABORATORS

Avinoam Ben-Shaul (HU)

Laurence Lavelle (UCLA)

A. L. N. Rao (UCR)

Jim & Ellen Strauss (Caltech)

Marlene Biller (Caltech)