I. What controls viral size?

II. Reconstituting an animal virus

Two not totally unrelated topics

GELBART-KNOBLER GROUP

Odisse Azizgolshani

Li Tai Fang

Aron Yoffe

Ajay Gopal

Yufang Hu

Adriana Anavitarte

Mauricio Comas

Serge Dmitrieff

Connie Chang (Tom Mason)

COLLABORATORS

Laurence Lavelle (UCLA)

A. L. N. Rao (UCR)

Jim & Ellen Strauss (Caltech)

Marlene Biller (Caltech)

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***?

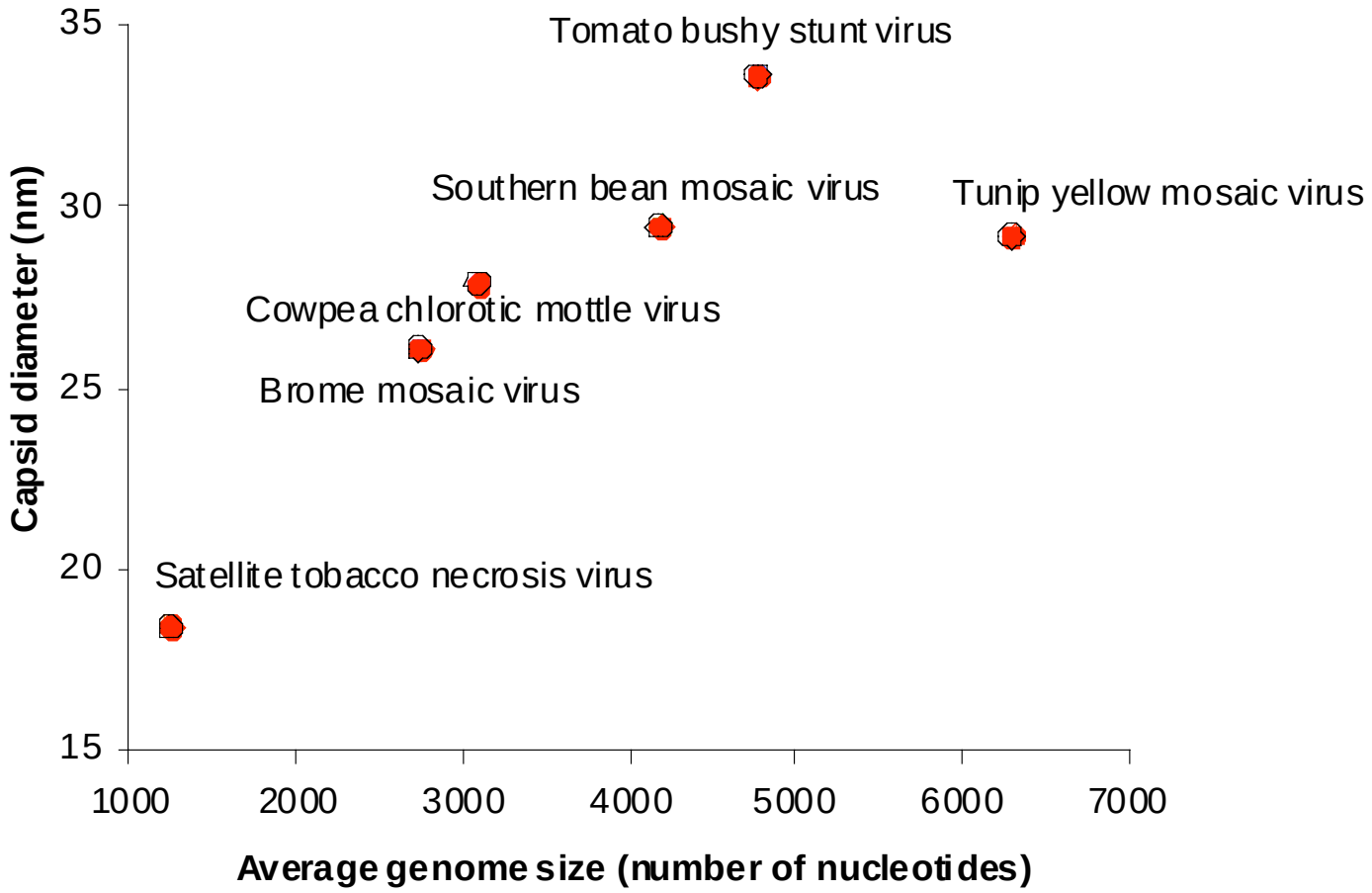
(2) ” “ “ “ ***OF A GENOME***?

PHYSICAL

QUESTION (1) ⇔ ***SELF-ASSEMBLY OF PROTEINS IN 2D CURVED SPACE***

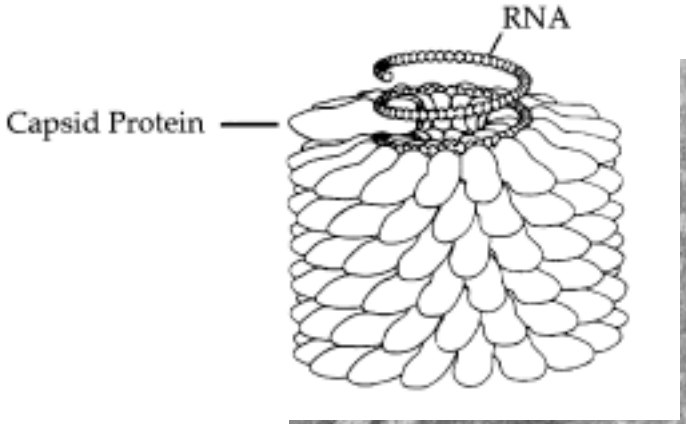
QUESTION (2) ⇔ ***CONFIGURATIONAL STATISTICS OF DNA AND RNA***

GENOME SIZE PLAYS A ROLE…



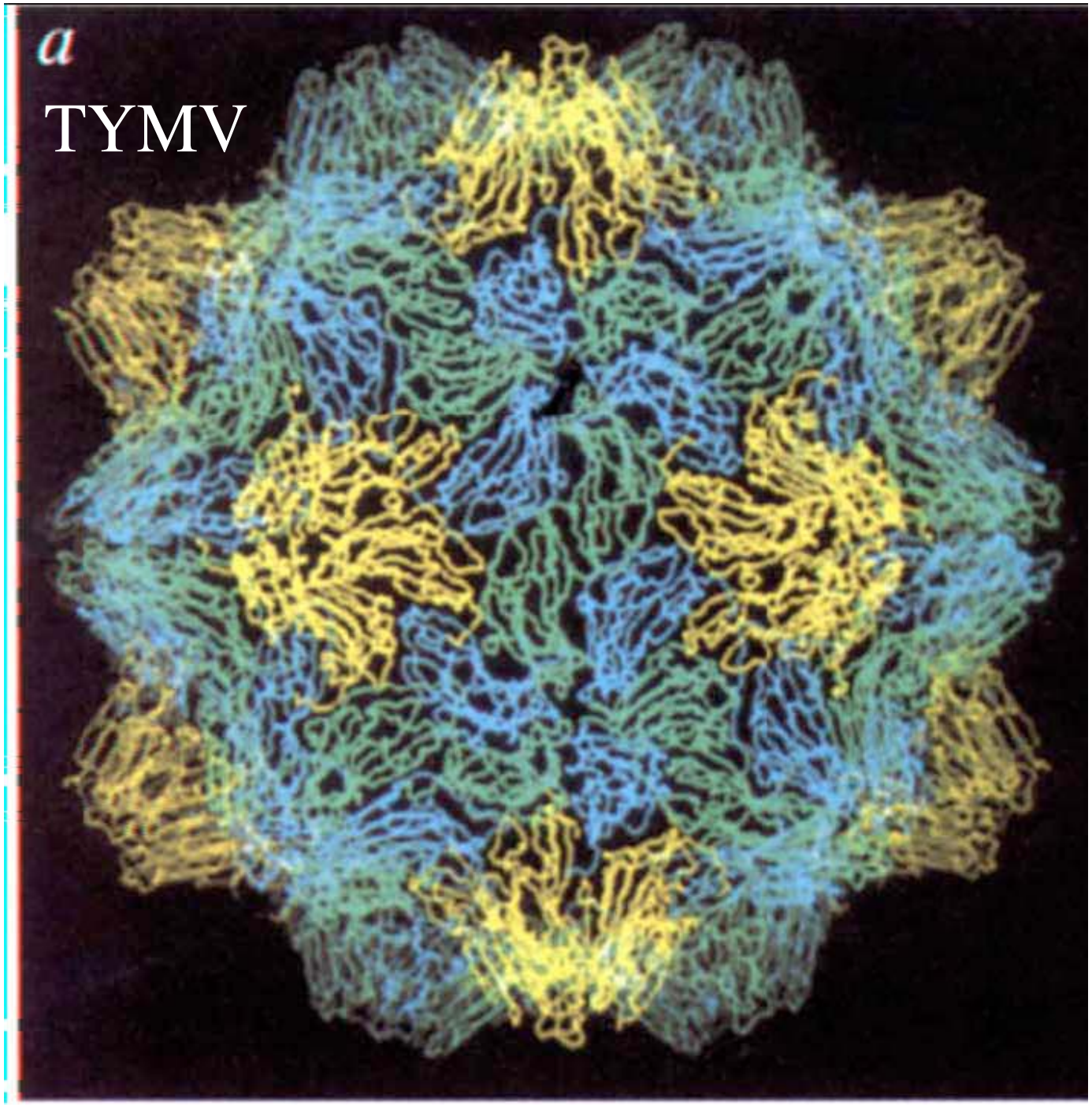
BUT…

Tobacco Mosaic Virus (TMV)



RNA 6395nt long

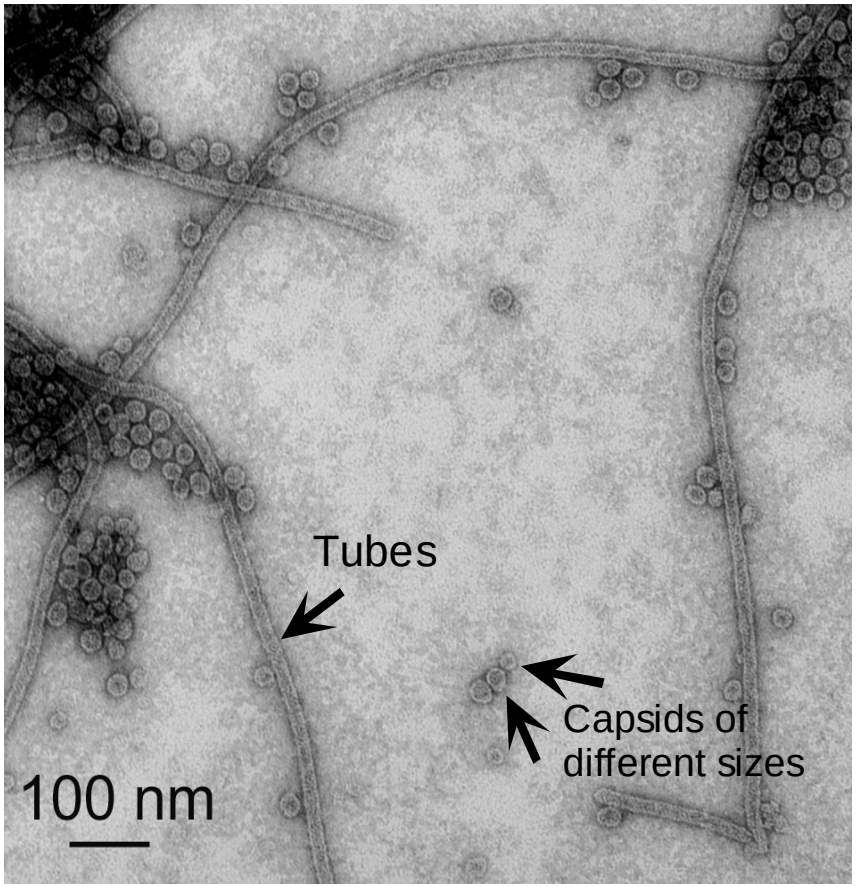
RNA 6318 nt long



THE CAPSID PROTEIN AND ITS SPONTANEOUS CURVATURE MUST PLAY A ROLE

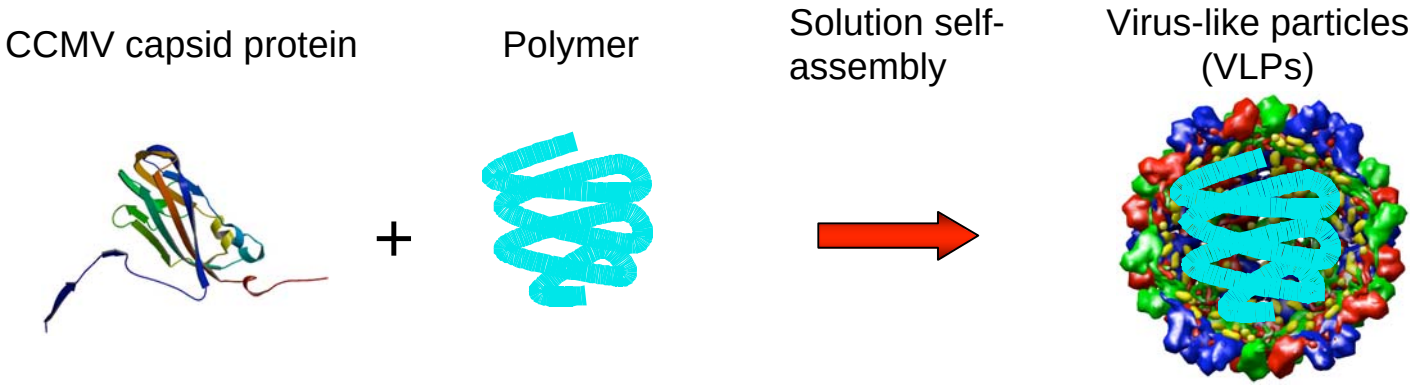
In the absence of RNA, CCMV capsid protein has many polymorphs

- Capsid protein can assemble into different structures
  - Icosahedral
  - Tubular
  - Planar hexagonal
  - Multishell



PACKAGING EXPERIMENTS

- Materials
  - Purified CCMV capsid protein
  - Sodium polystyrene sulfonate (PSS)



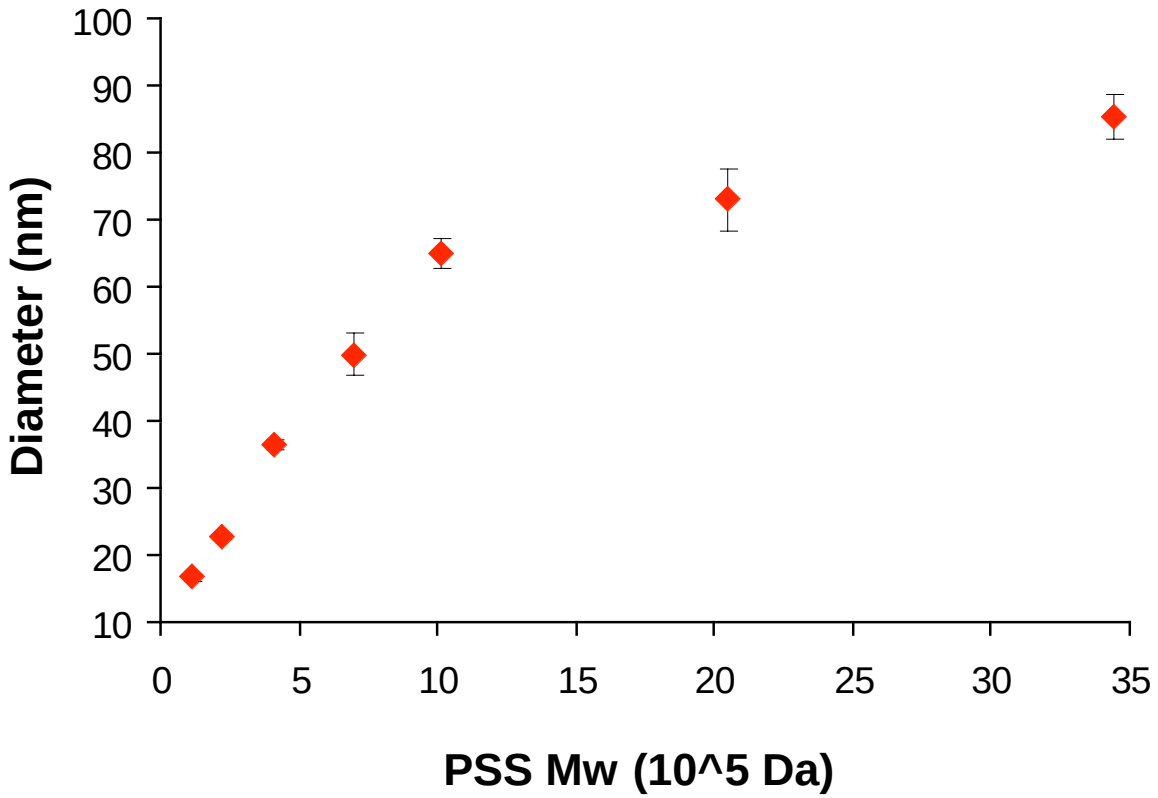
- Techniques
  - Negative staining transmission electron microscopy (TEM)
  - Analytical ultracentrifugation

Properties of PSS

- Highly negatively charged (>95%), flexible polymer
- Linear charge density is ~1e/3Å, comparable to RNA
- Wide range of molecular weights available:

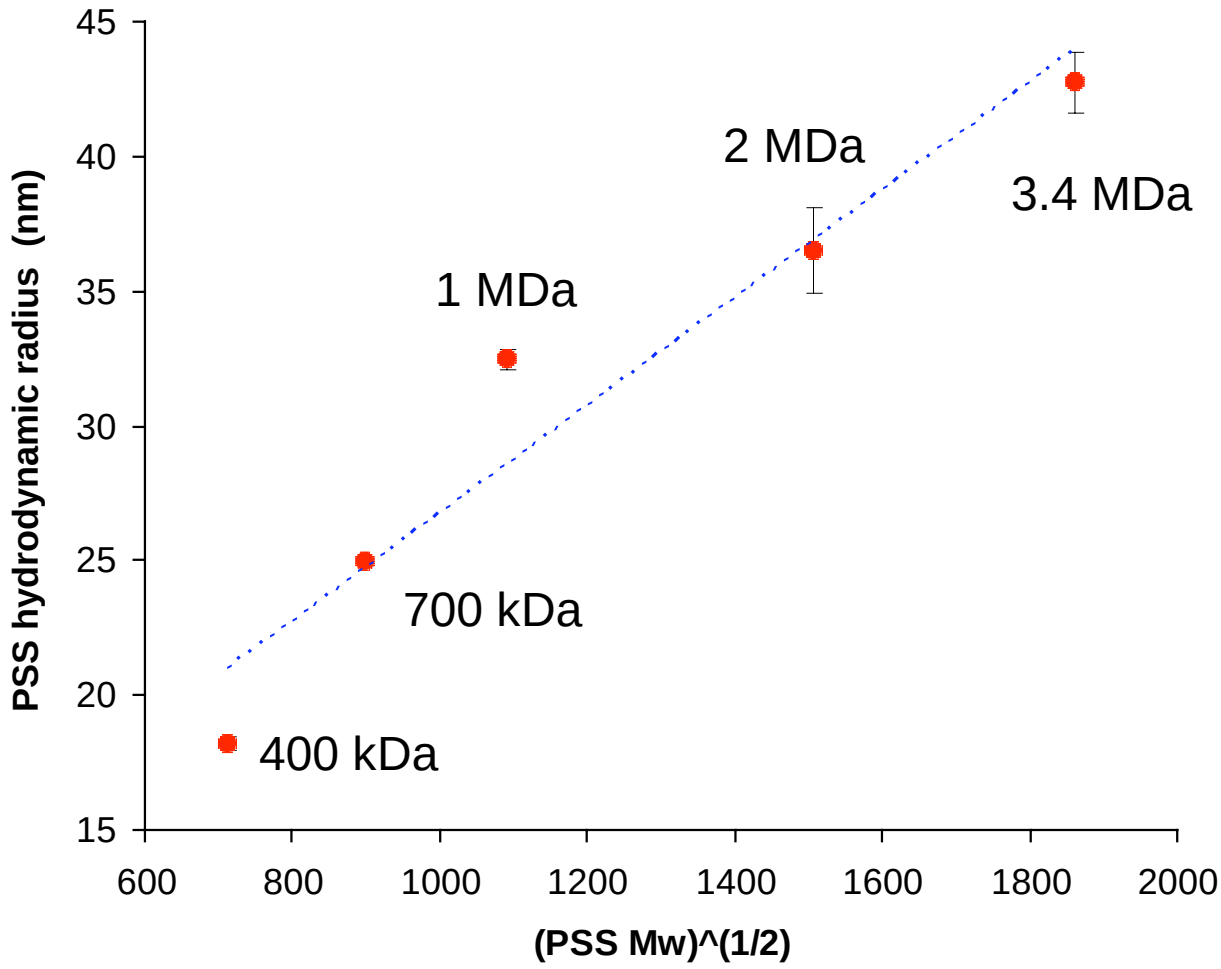
  400 k, 700 k, 1 M, 2 M, 3.4 MDa

Hydrodynamic radius of PSS
Laser light scattering



Assembly buffer: 50 mM NaCl, 50 mM Tris-HCl, pH 7.2, 10 mM KCl, 5 mM $MgCl_2$

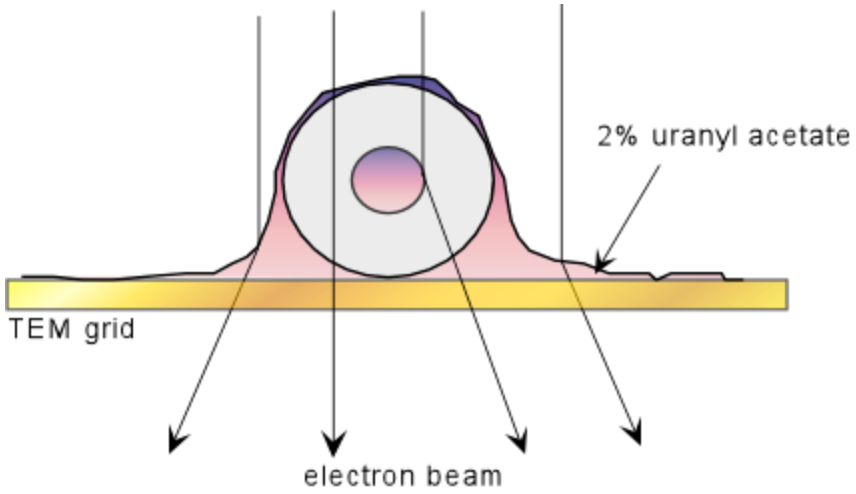
Hydrodynamic Radius of PSS

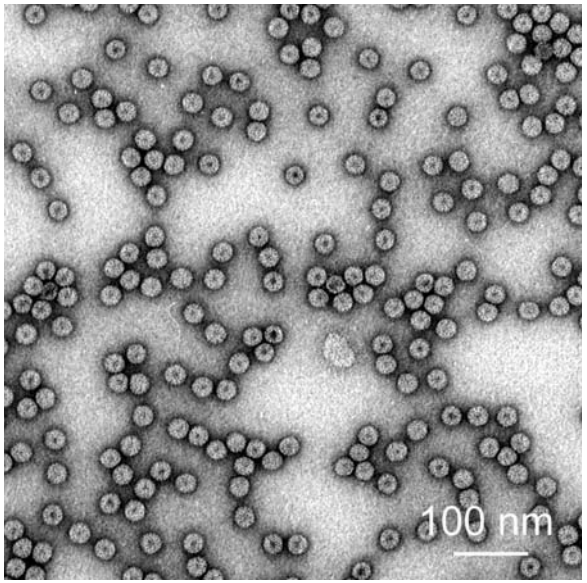


Assembly buffer: 50 mM NaCl, 50 mM Tris-HCl, pH 7.2, 10 mM KCl, 5 mM $MgCl_2$

Negative Staining TEM

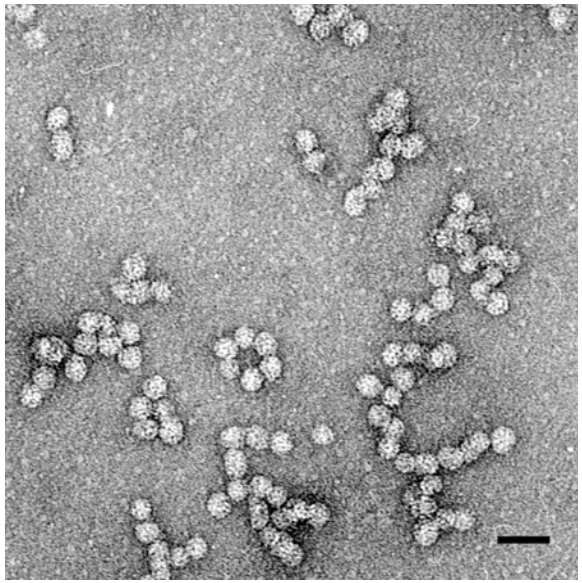
- Presence of sample excludes stain
- Contrast is provided by stain
- Direct visualization of VLP structure
- Sample is dehydrated - so dimensions differ from those in solution



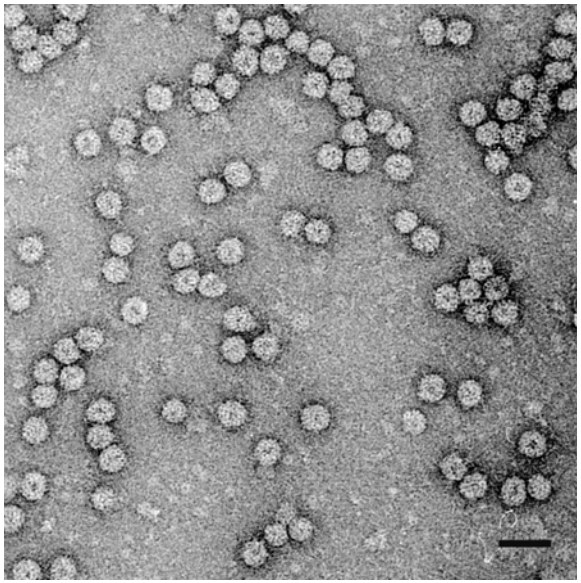


VLPs Formed in Assembly Reactions

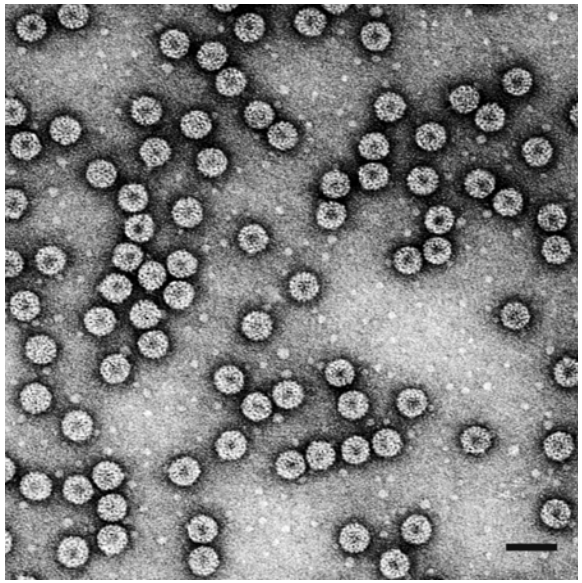
VLP 700k



VLP 2M



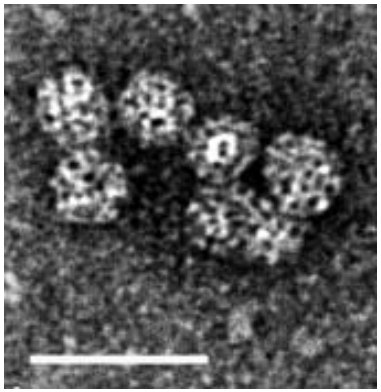
wt CCMV



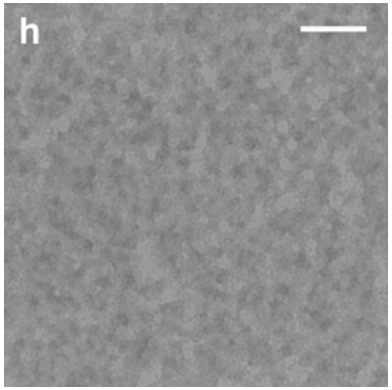
Scale bars are 50 nm

Morphology of VLPs

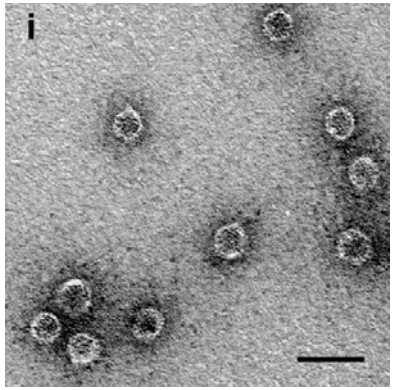
VLP 1 MDa



Protein only, no PSS

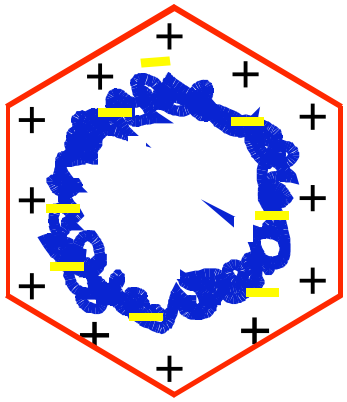


Empty capsid

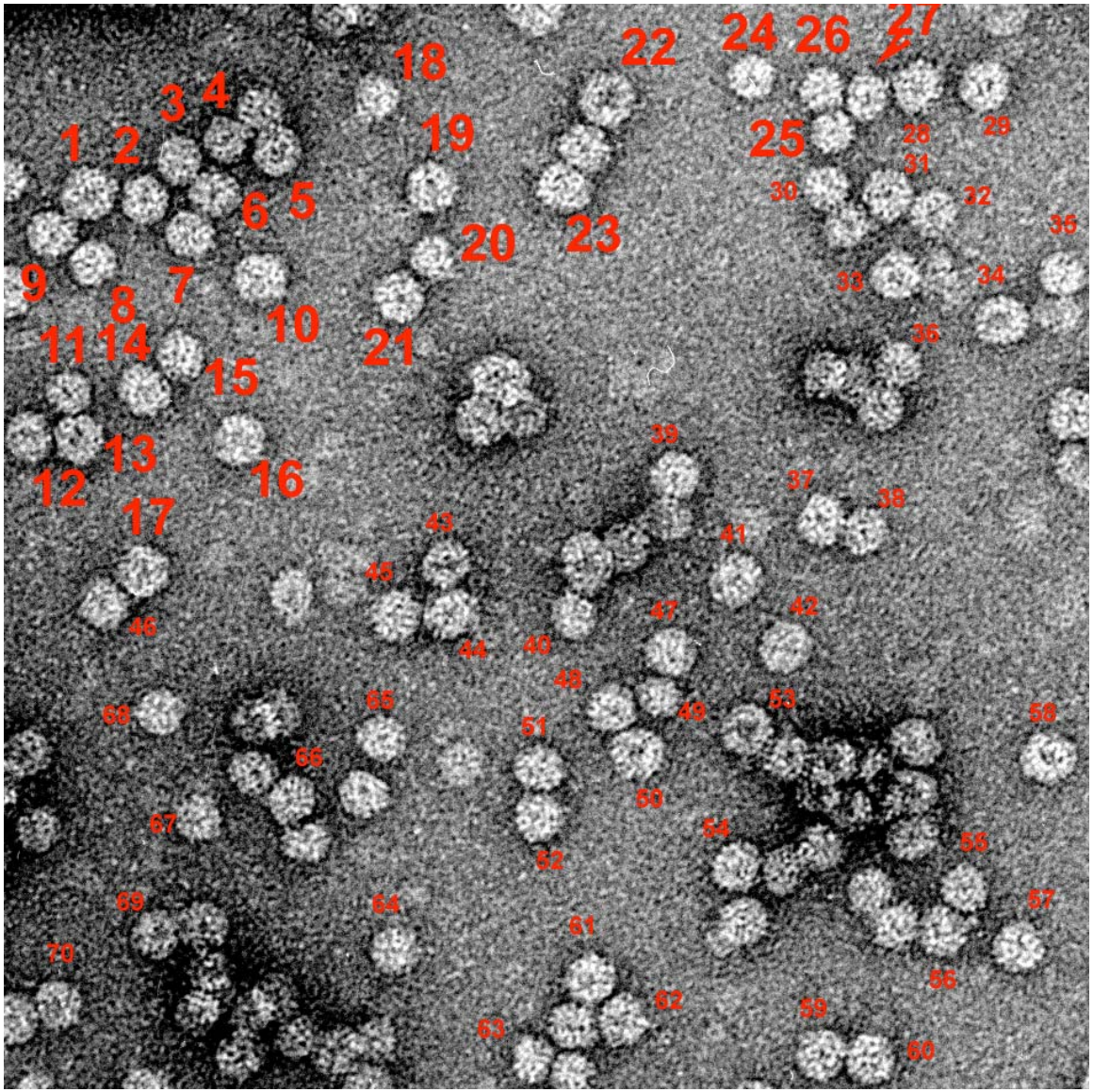


Scale bars are 50 nm

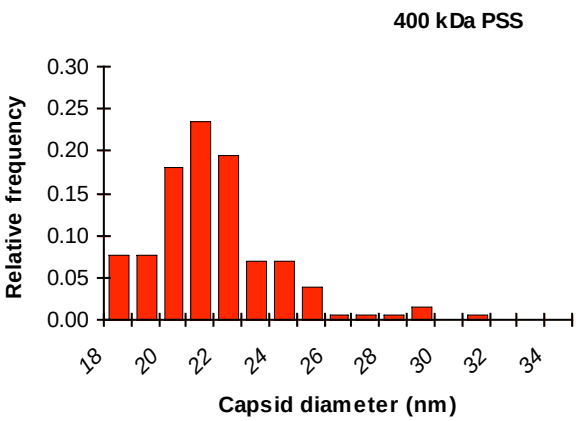
- No stain penetration into interior of VLPs
- Interior of VLPs are filled with material
- VLP formed via encapsidation of PSS

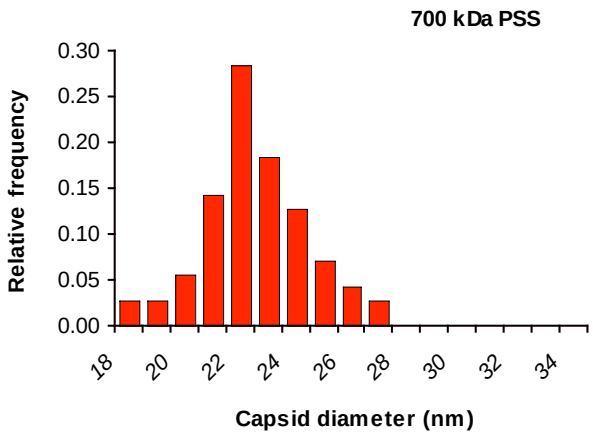


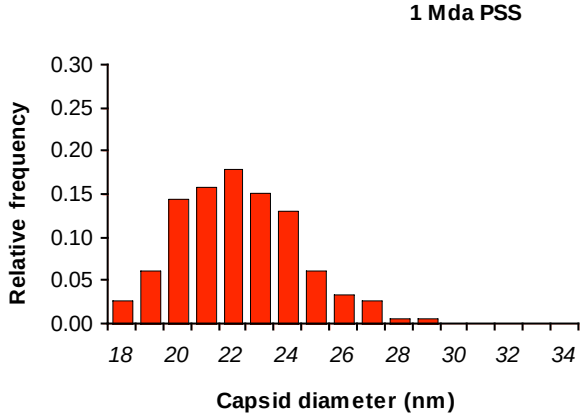
Sizing the VLPs

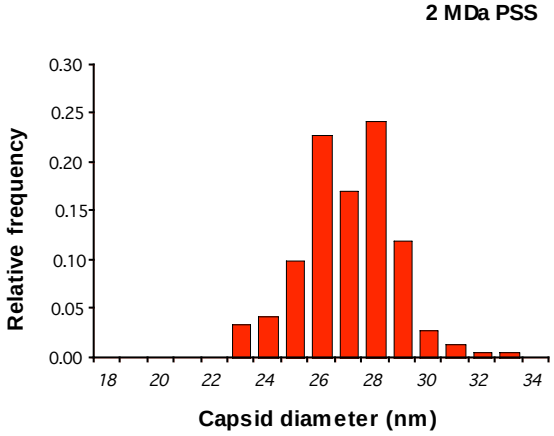


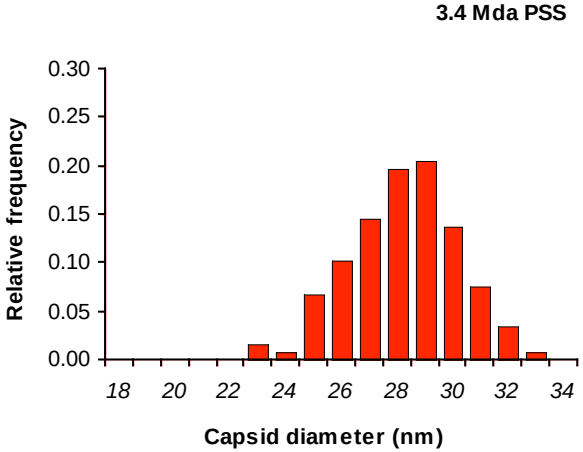
Capsid Size Distribution





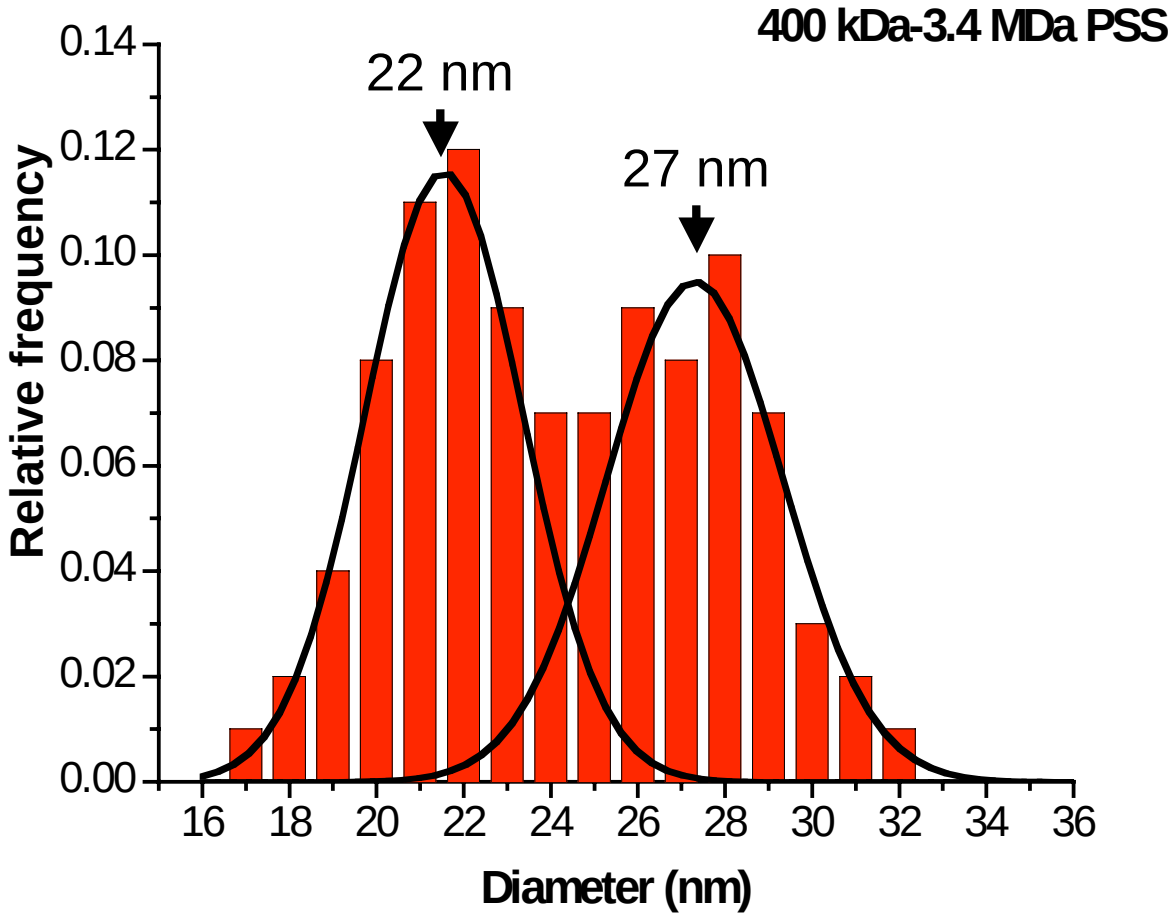






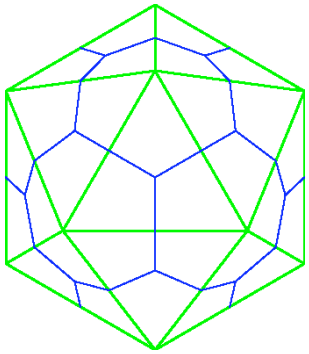
Capsid size converges to two discrete values: 22, 27nm

The Capsid Sizes are Discrete

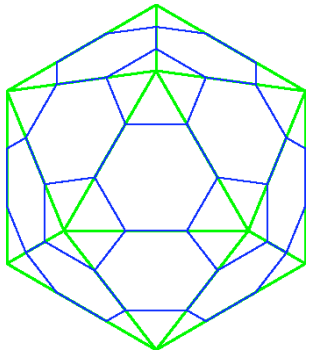


Estimation of T Numbers

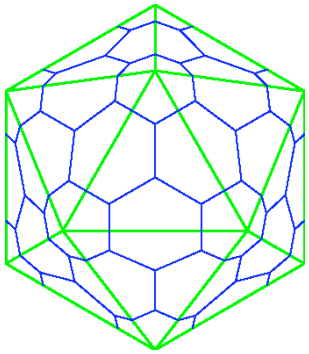
T=2, 120 proteins



T=3, 180 proteins



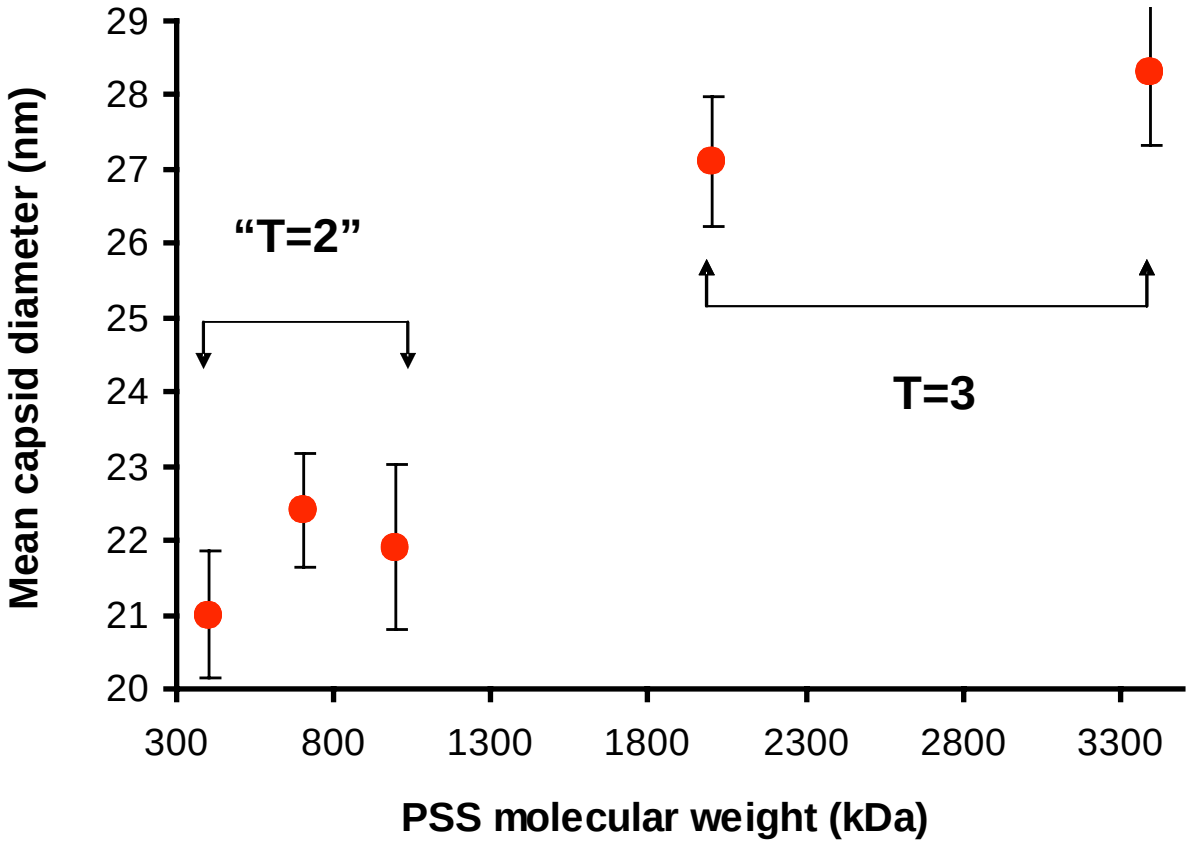
T=4, 240 proteins



Capsid surface area is equal to 60T x surface area per capsid protein

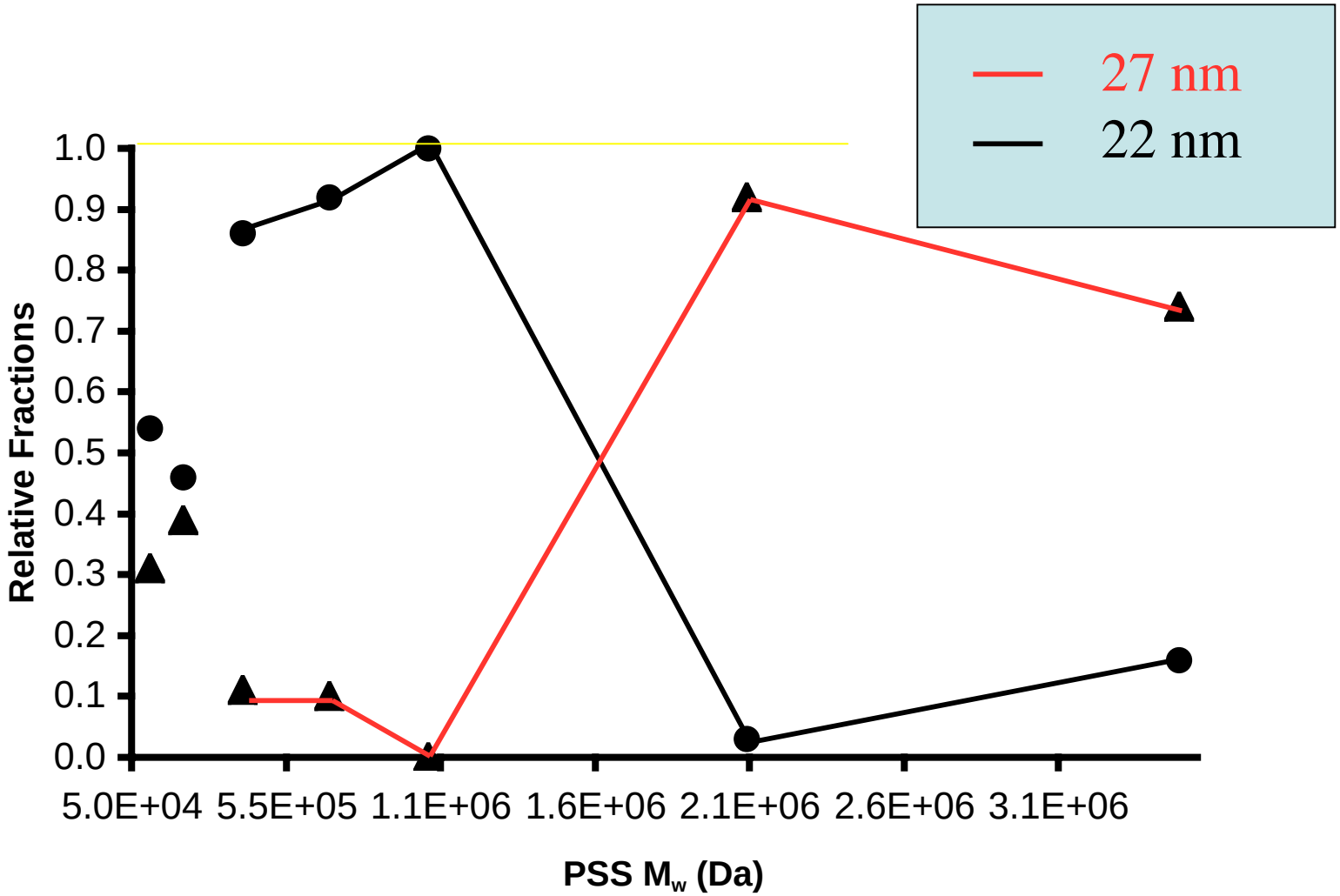
$$\frac{D_{T1}}{D_{T2}} = \sqrt{\frac{T_1}{T_2}} \quad \rightarrow \quad \frac{D_{T1}}{D_{T2}} = \frac{22}{27} = \sqrt{\frac{2}{3}}$$

Estimation of T Numbers

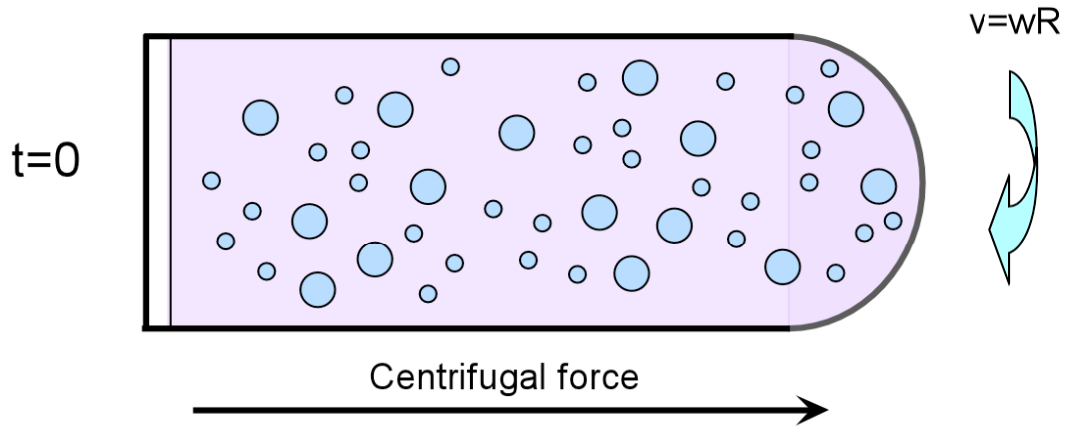


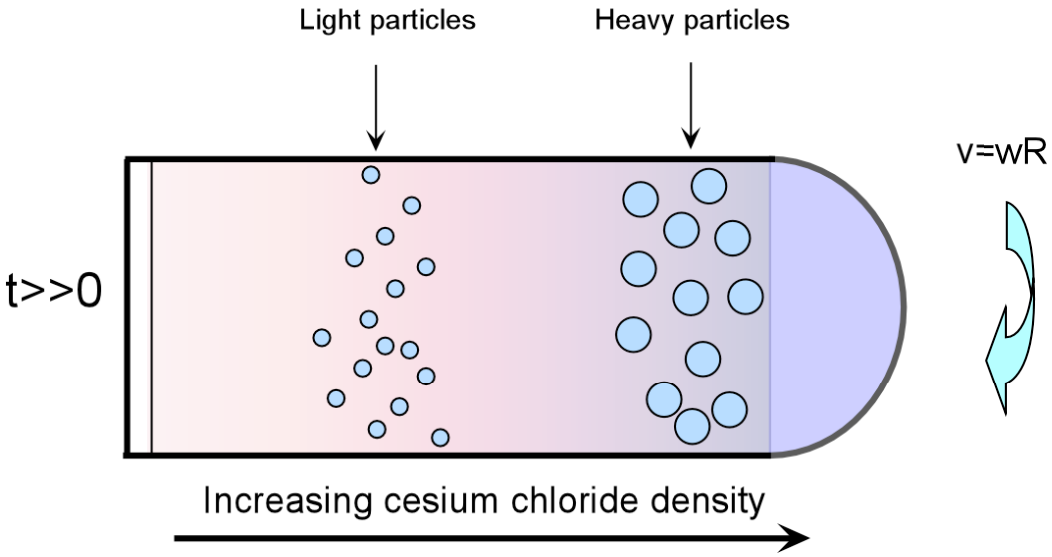
The observed capsid diameters of 22 and 27 nm correspond to T=2 and T=3 triangulation numbers

Higher MWs of polymer give larger protein shells…



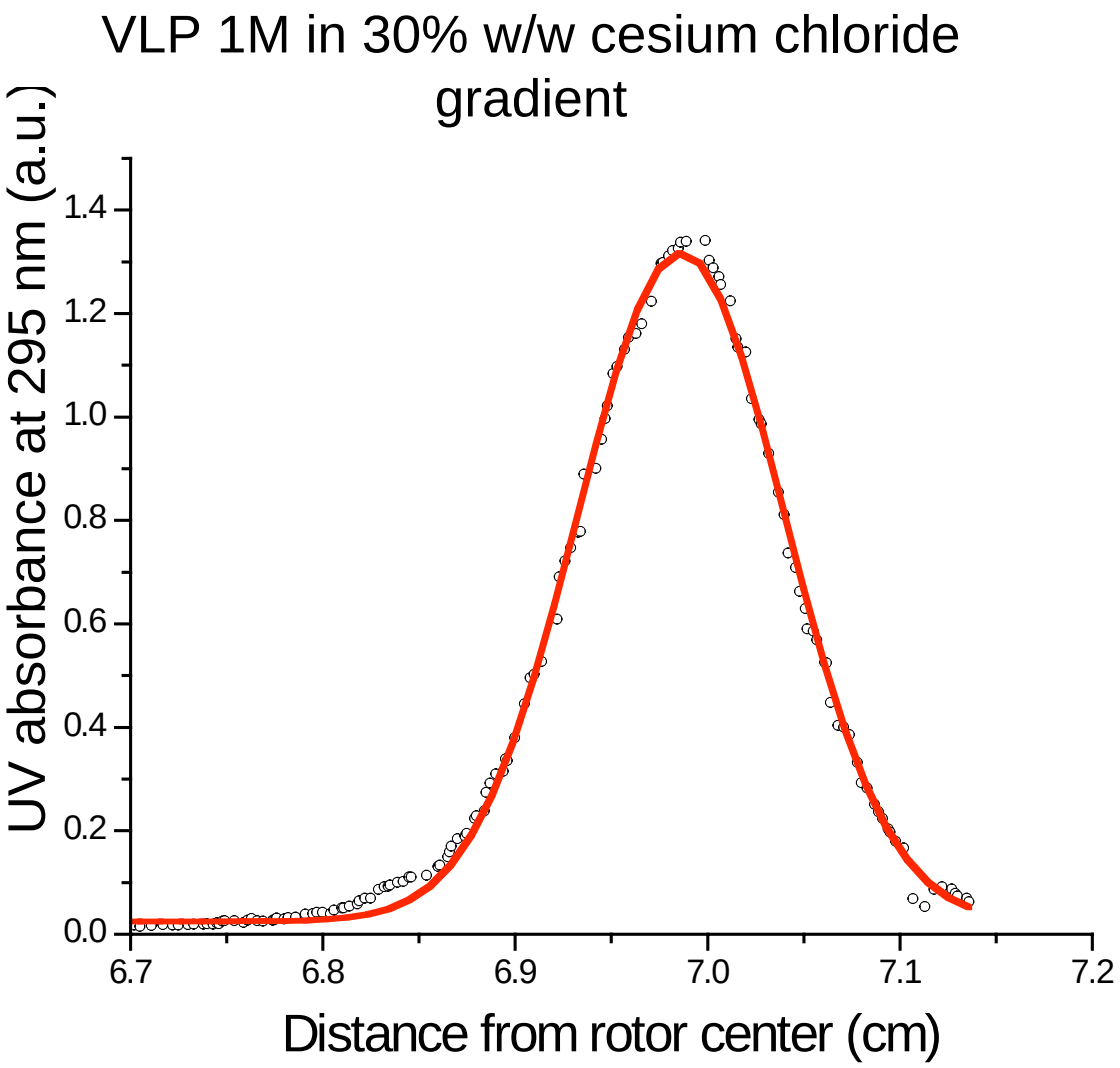
Equilibrium Centrifugation





- VLPs are loaded onto cesium chloride density gradient and centrifuged
- Particles are separated by density at equilibrium
- Particle molecular weight can be calculated from its equilibrium distribution profile in the gradient

Molecular Weight Estimation for VLP 1M



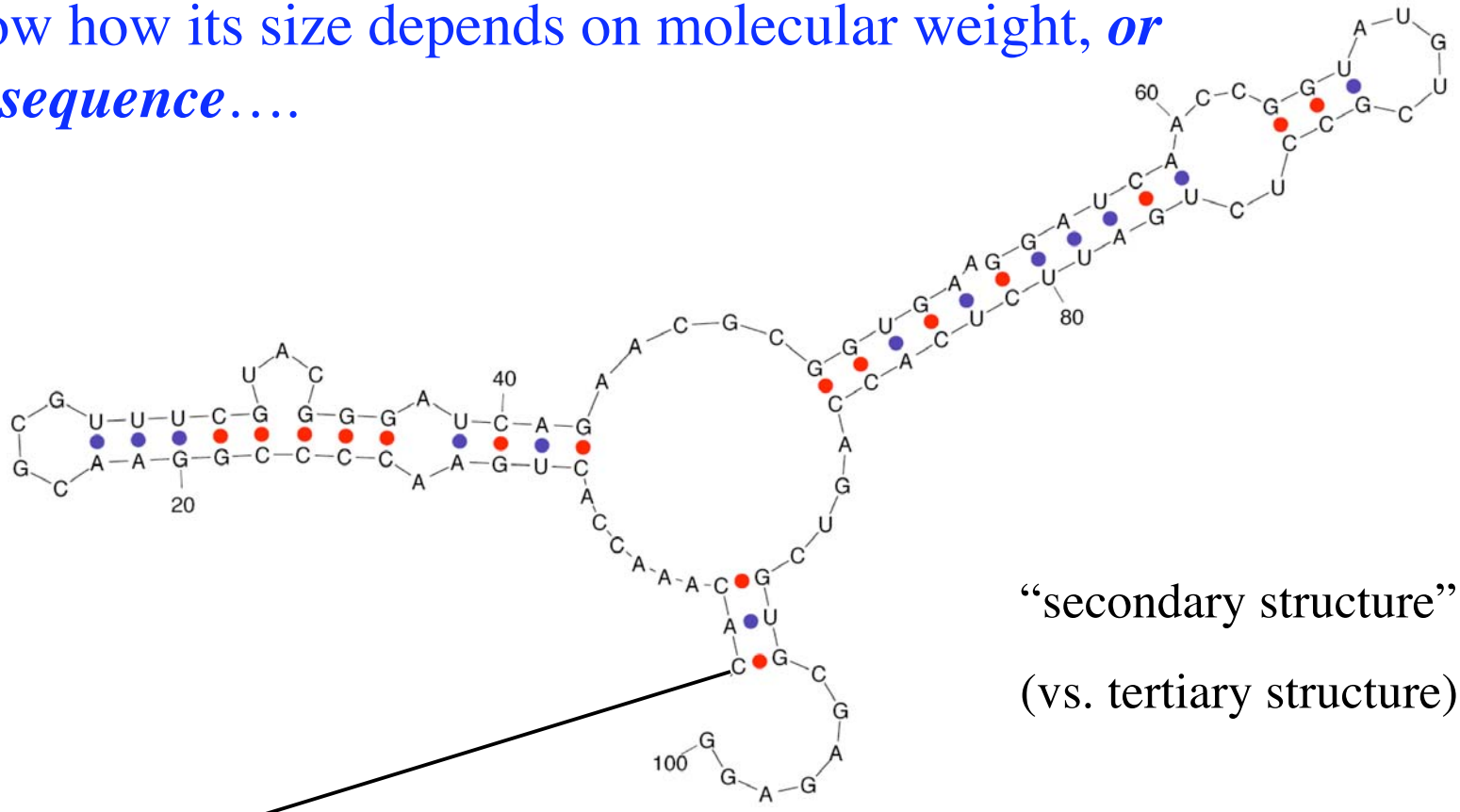
- The molecular weight of a Gaussian band is

$$M_w = \frac{RT}{\omega^2 r \bar{\upsilon} (d\rho/dr)\sigma^2}$$

- The estimated molecular weight of the species in the band is 3.7×10^6
- The molecular weight of anhydrous VLP 1M with 120 proteins and 1 PSS inside is 3.4×10^6
- VLP 1M has T=2 capsid

What determines the size and shape of an RNA in solution under the conditions of assembly?

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)

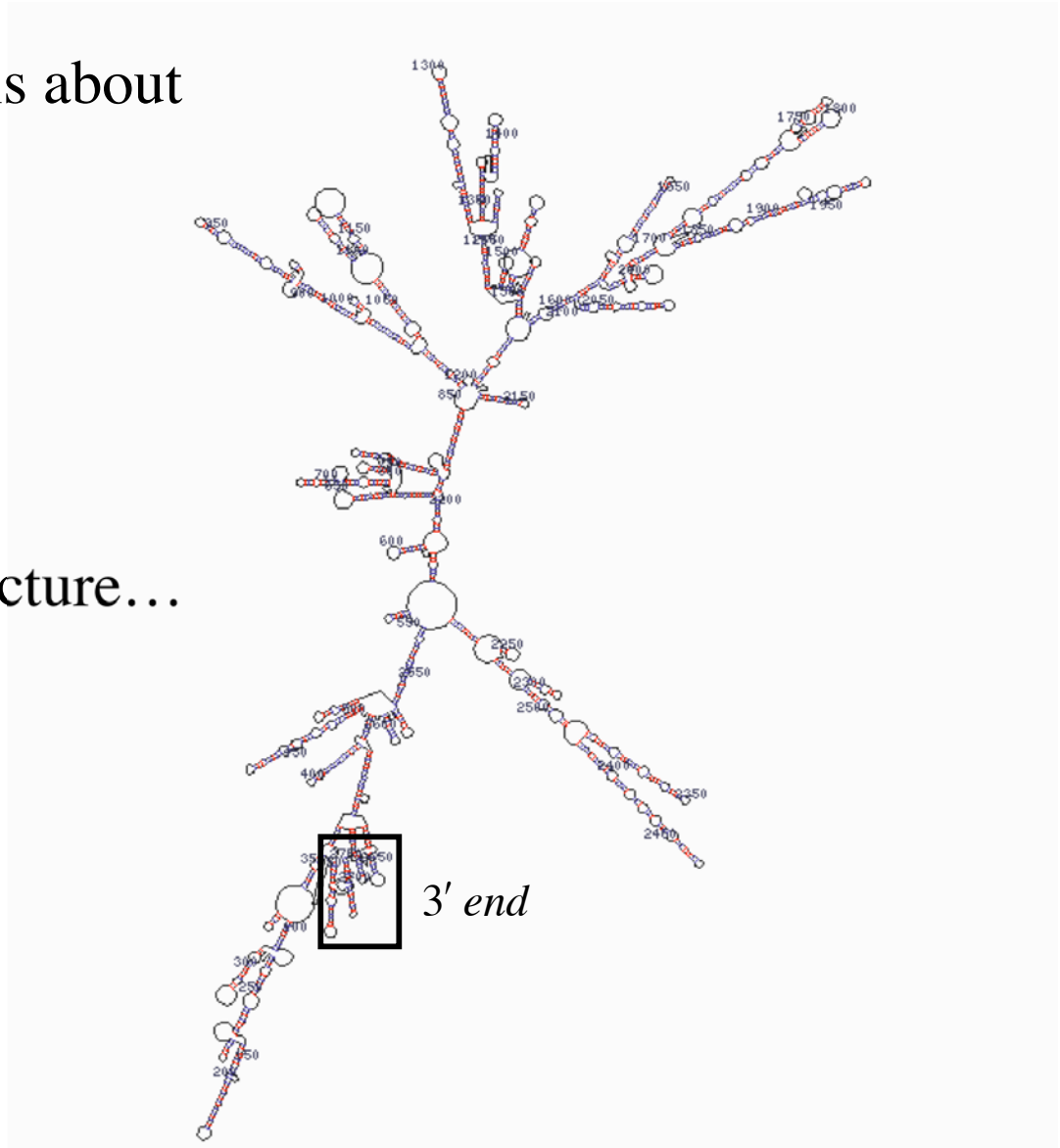
(5')CACAAACCACUGAACCCCGGAACGCGUUUCGUACGGGAUCAGAACGCGGU
GAAGGAUCAACCGGUAUGUCGCCUCUGAUUCUCACCAGUCGUGCGAGAGG(3')

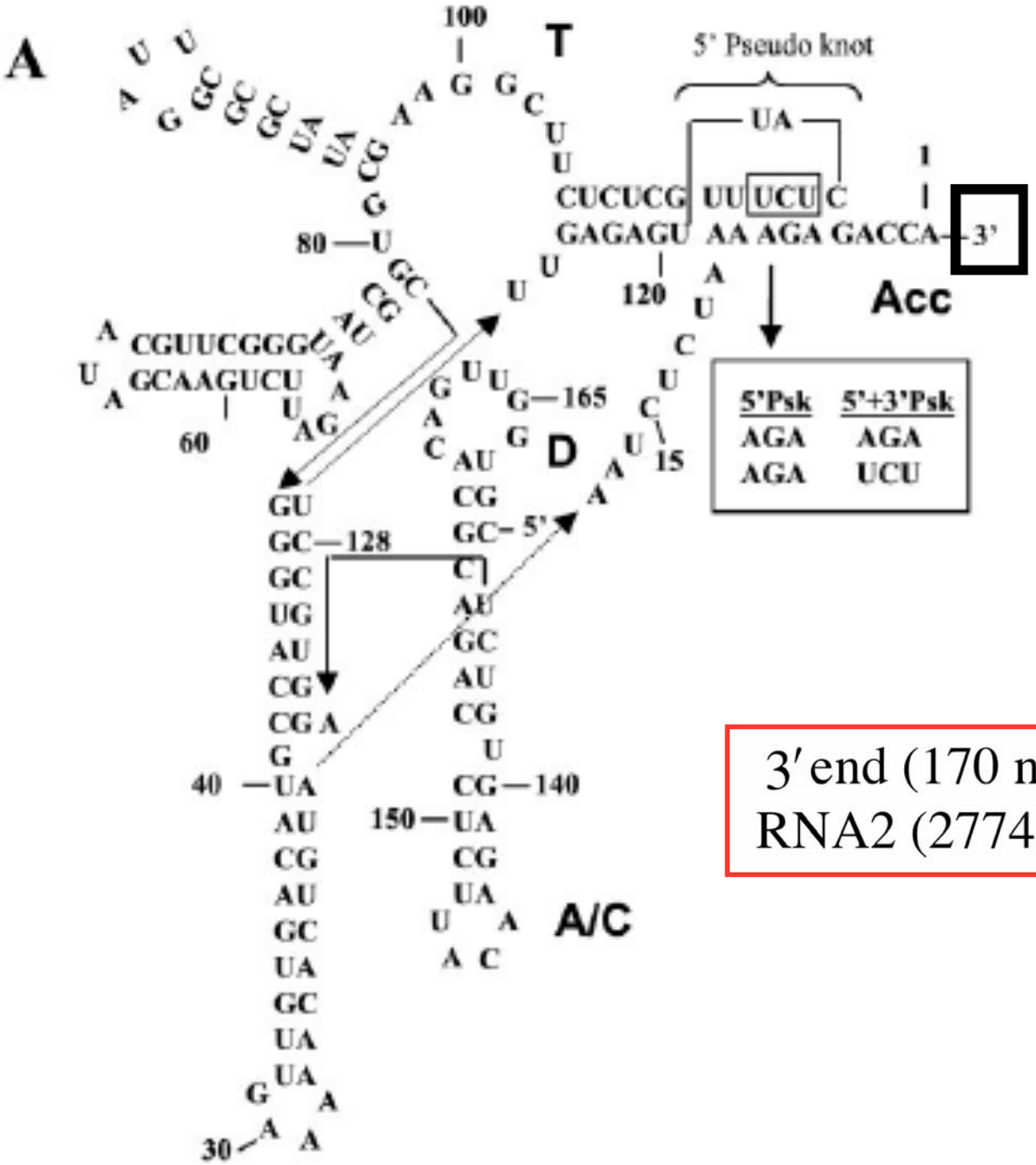
What is the 3D size and shape of a (viral-length) RNA molecule?

E.g., CCMV RNA2, which is about 2800 nt long

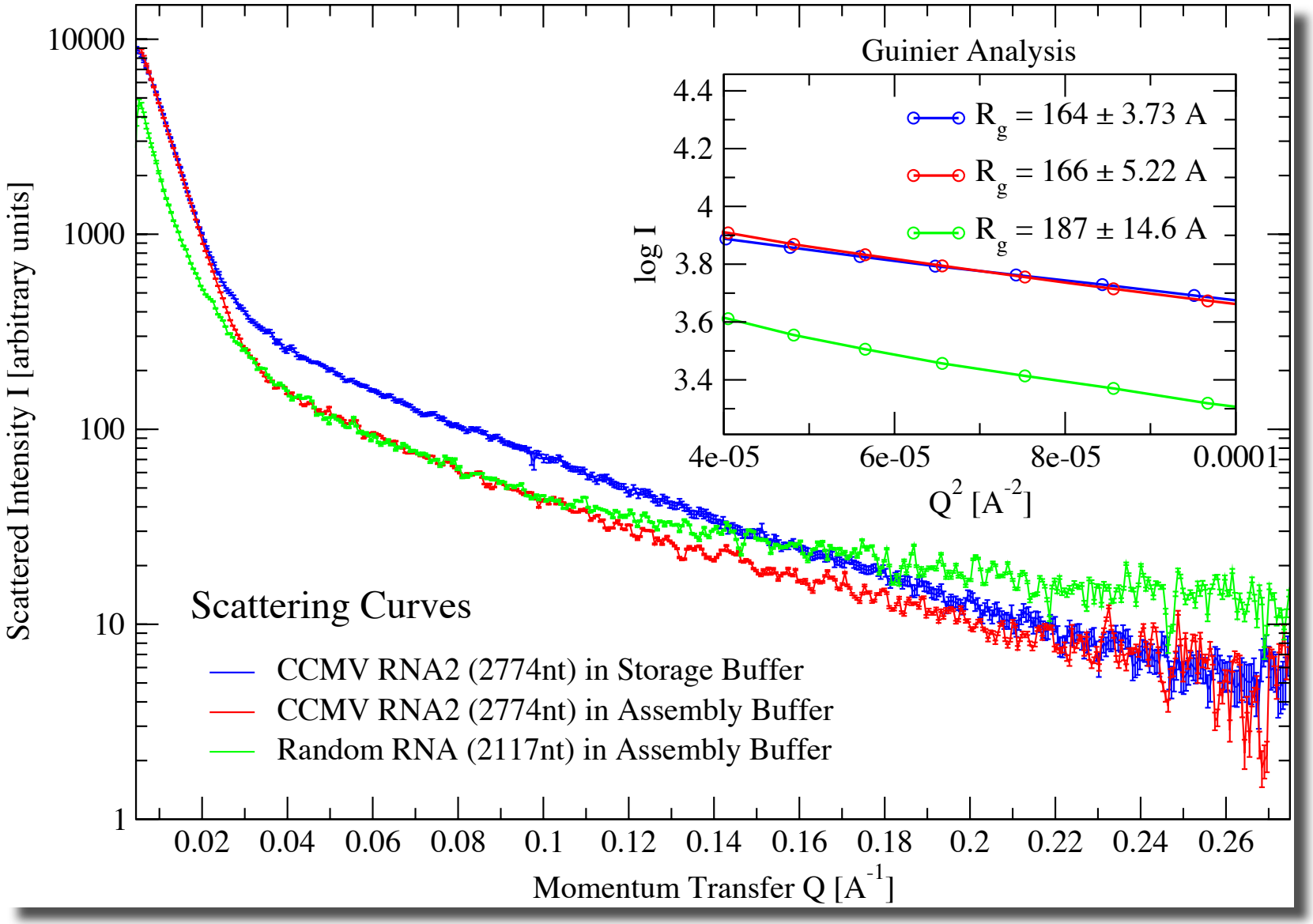
Start with its secondary structure…

(from MFOLD)



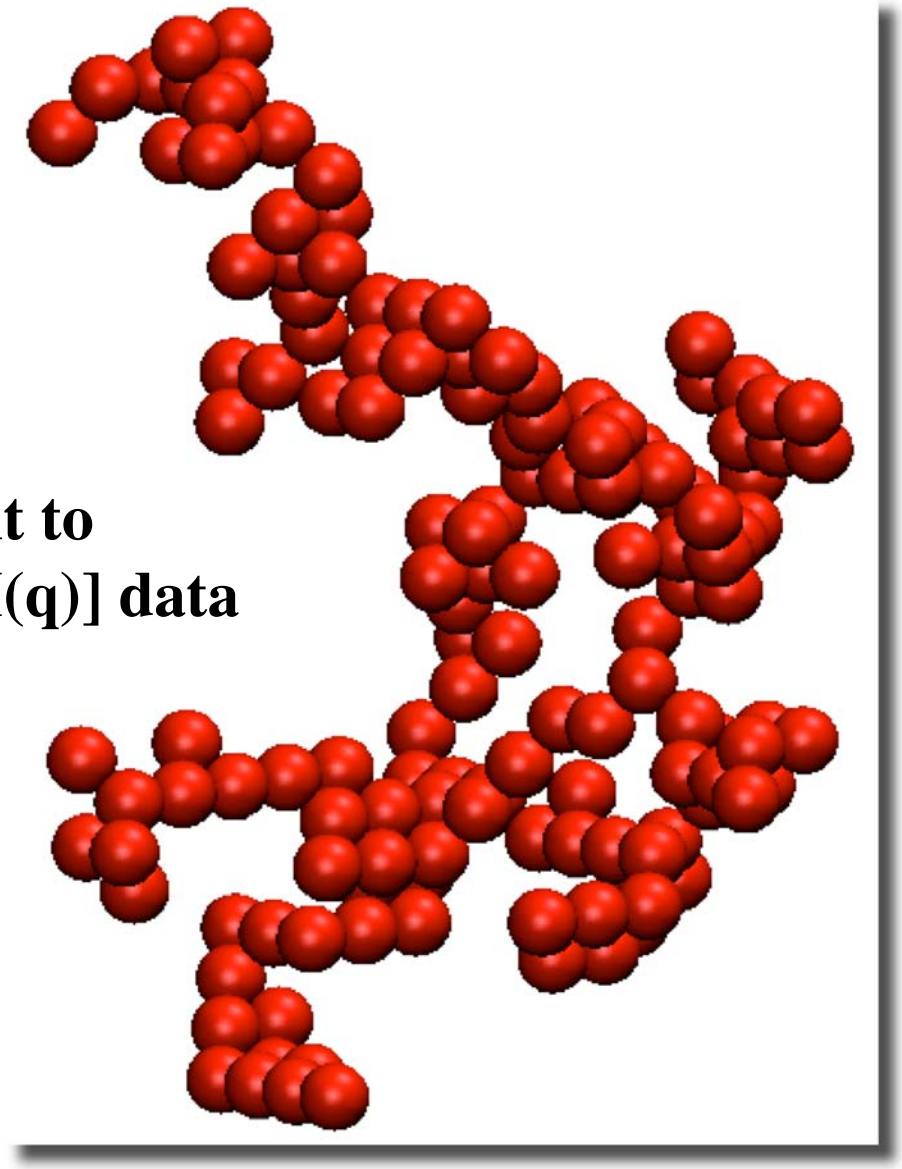

A
T
100
5' Pseudo knot
UA
1
CUCUCG UU UCU C
GAGAGU AA AGA GACCA 3'
80
120
Acc
5'Psk 5'+3'Psk
AGA AGA
AGA UCU
CGUUCGGG
GCAAGUCU
60
G U U G—165
D
15
GU
GC—128
GC
UG
AU
CG
CG A
G
40 —UA
AU
CG
AU
GC
UA
GC
UA
UA
G A
A A
30
GC— 5'
C
AU
GC
AU
CG
U
CG—140
150—UA
CG
UA
A/C
3′end (170 nt) of CCMV RNA2 (2774 nt)

Use SAXS to determine 3D size & shape…



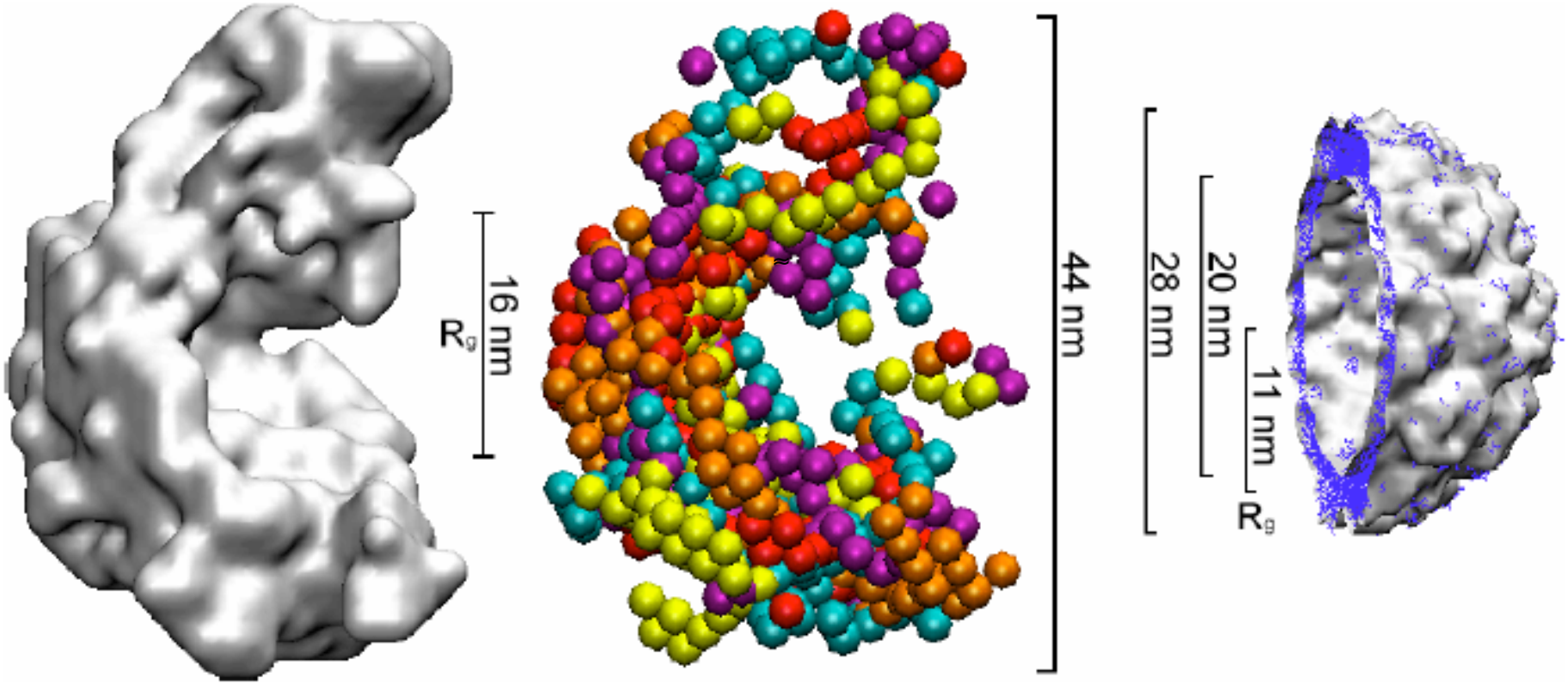
3D coarse-grained model fit to CCMV RNA2 scattering [I(q)] data

D. Svergun, S. Doniach



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

$R_{RNA} \gtrapprox R_{capsid}$

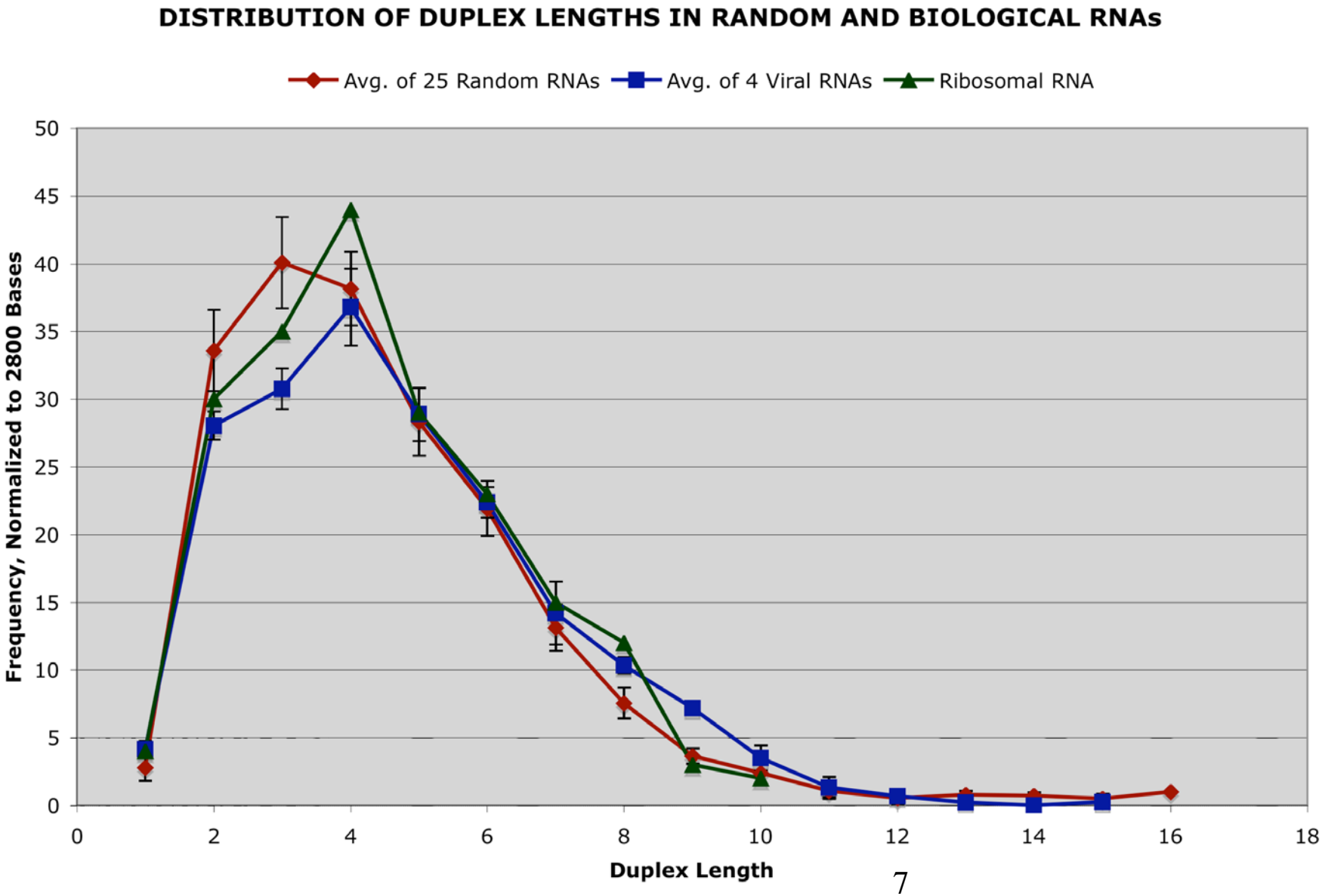


Comparison of Percent Base Pairing and Average Duplex Length in Viral, Random, and Ribosomal RNA

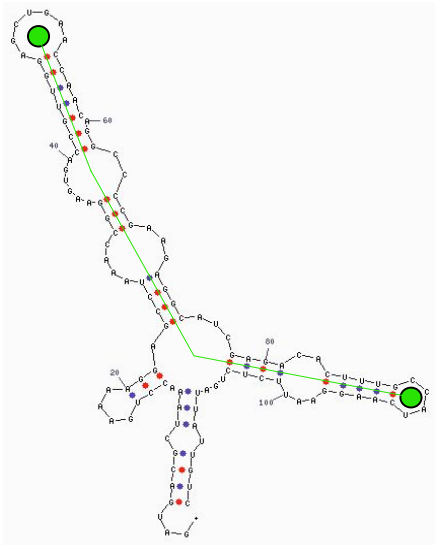
- Viral genome sequences have evolved ***not only*** to yield optimal gene products, ***but also*** to have a size and shape that ensures they will be encapsidated.
- Therefore, if the above coarse-grained characteristics are predictive for size, their values for random, viral, and ribosomal RNA should 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



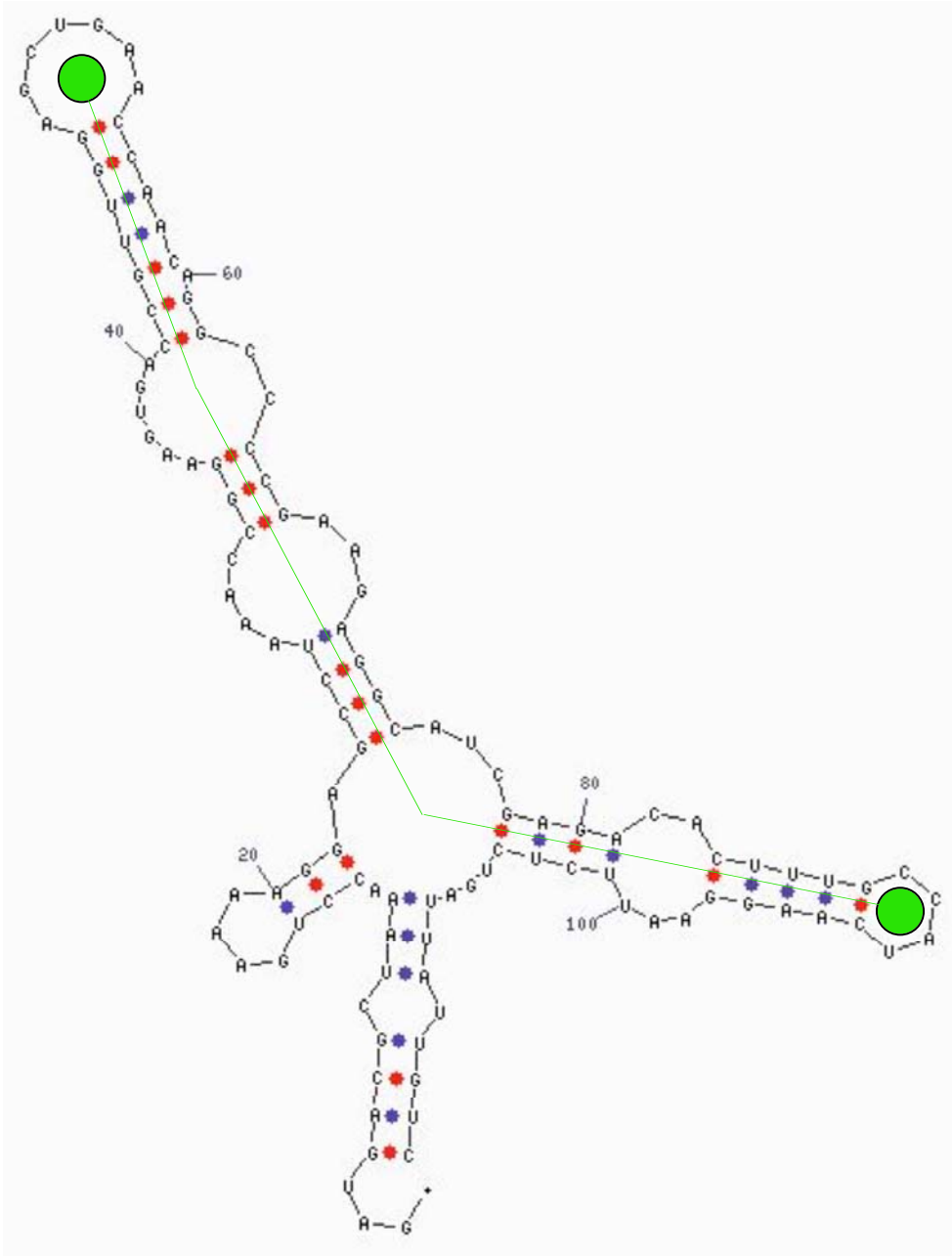
There is a Large Difference in Maximum Ladder Distance (MLD) Between These Random and Viral RNAs

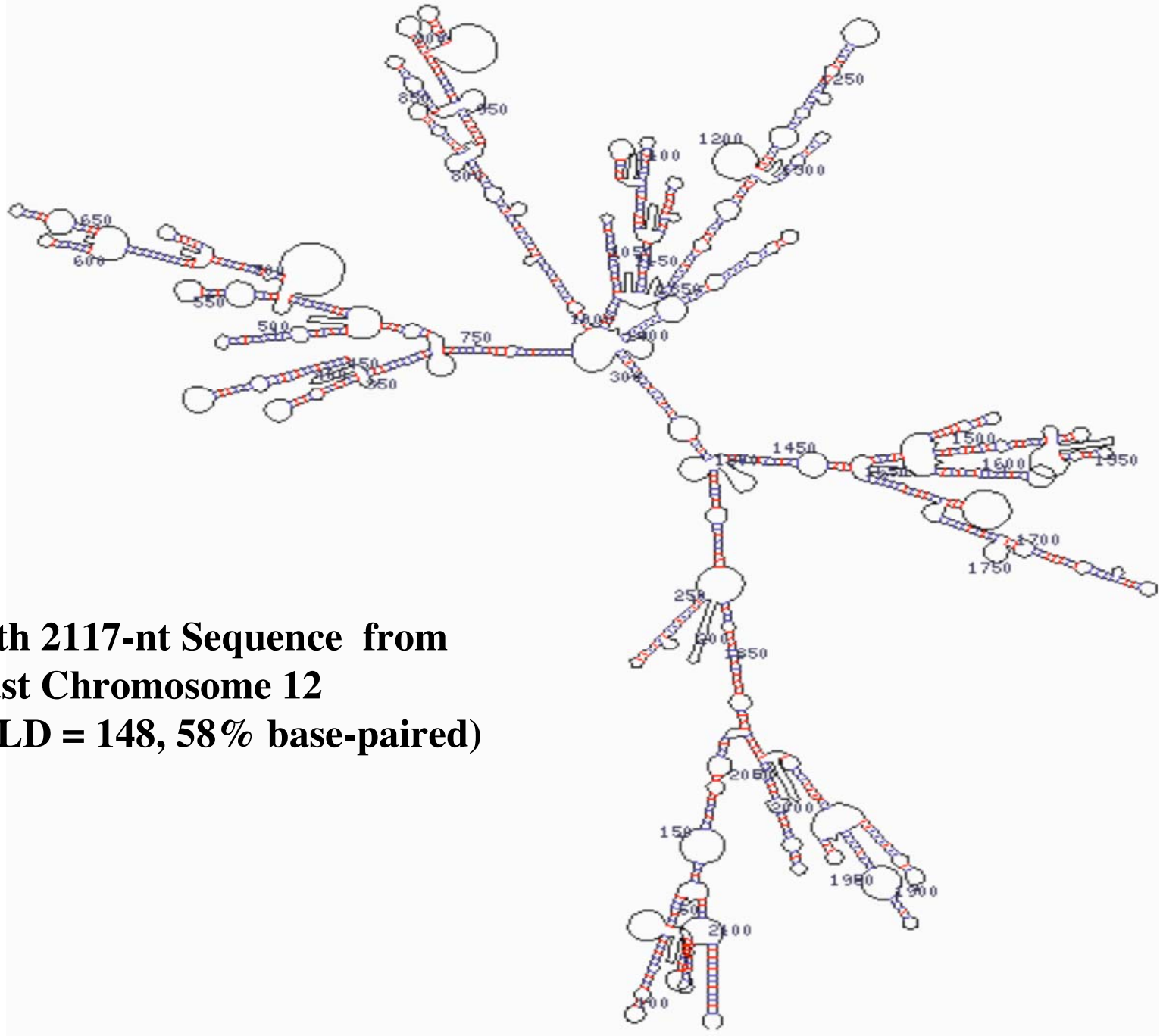


The MLD in this structure is 23, the number of base pairs crossed by the path connecting the two green dots.

- 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**.

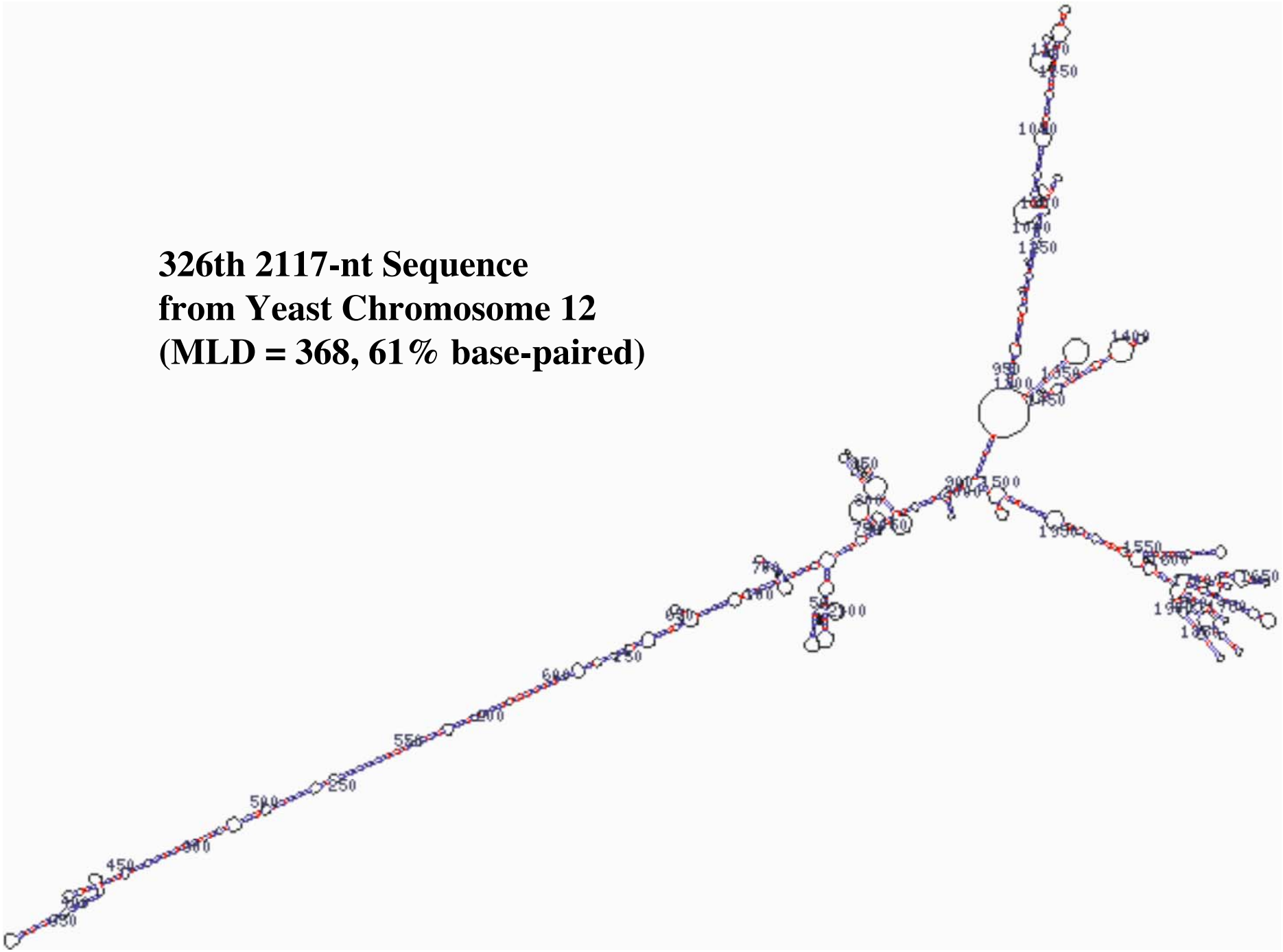
CHARACTERISTIC	AVG. OF 25 2800-BASE RANDOM RNAs	AVG. OF 4 VIRAL RNAs (CCMV RNAs 1+2, BMV RNAs 1+2)	*E. Coli* RIBOSOMAL RNA, 2825 BASES
MAX. LADDER DISTANCE	291.0 ± 38.1	206.4 ± 8.3	235.9


20
40
60
80
100



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)

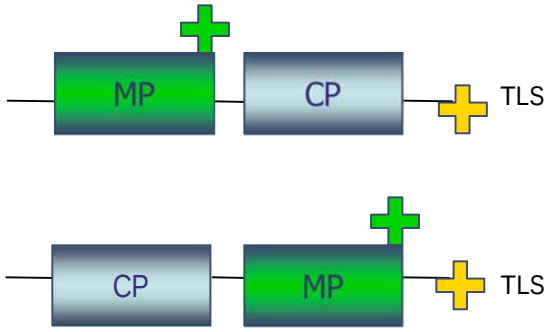


Determine Sequence-Dependence of RNA Size and Shape

BMV RNA3, 2117nt
(packages)

BMV RNA3Rev, 2117nt
(does not package)

BMV RNA3Neg-sense, 2117nt

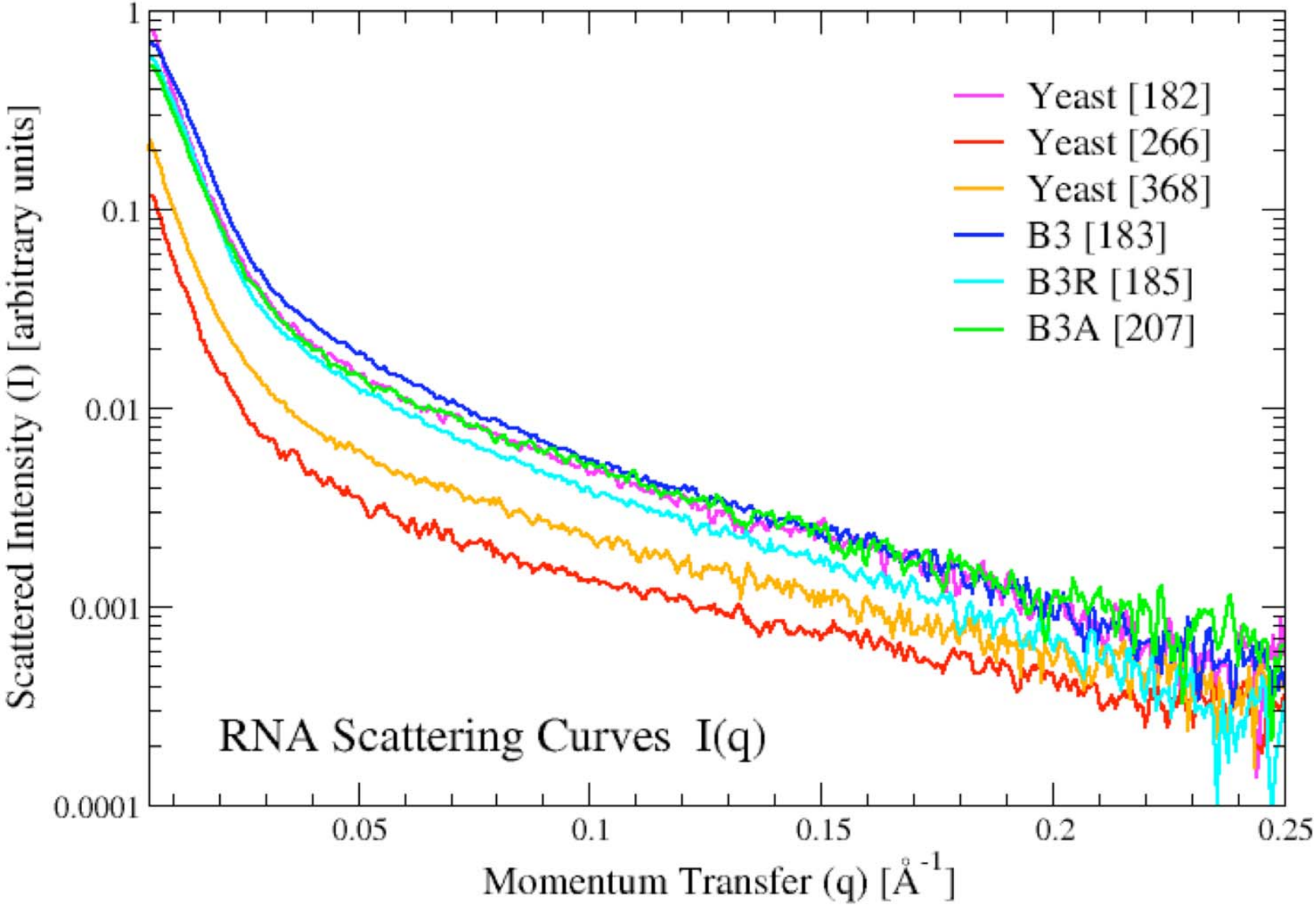


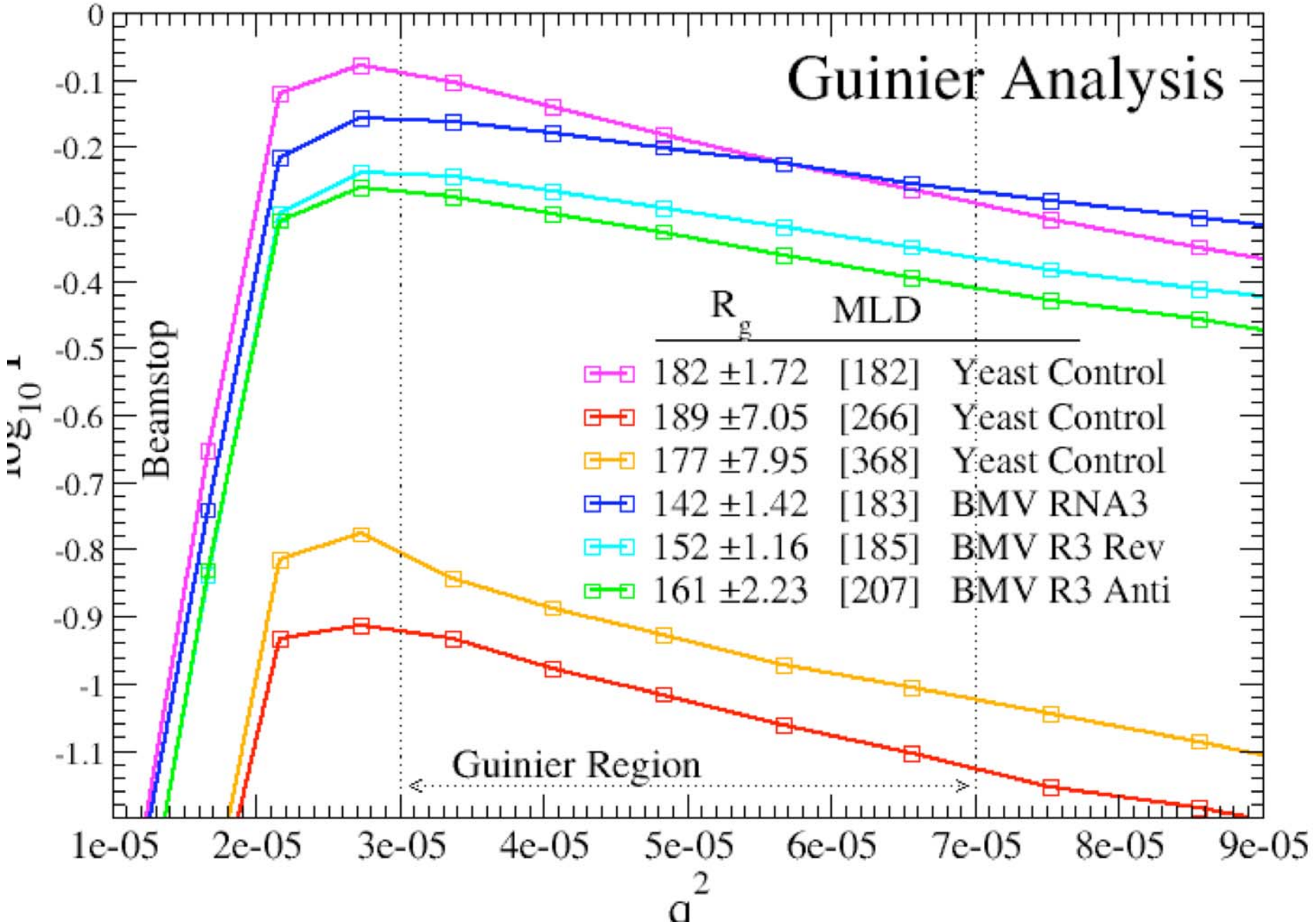
COMPARE AND CONTRAST TO

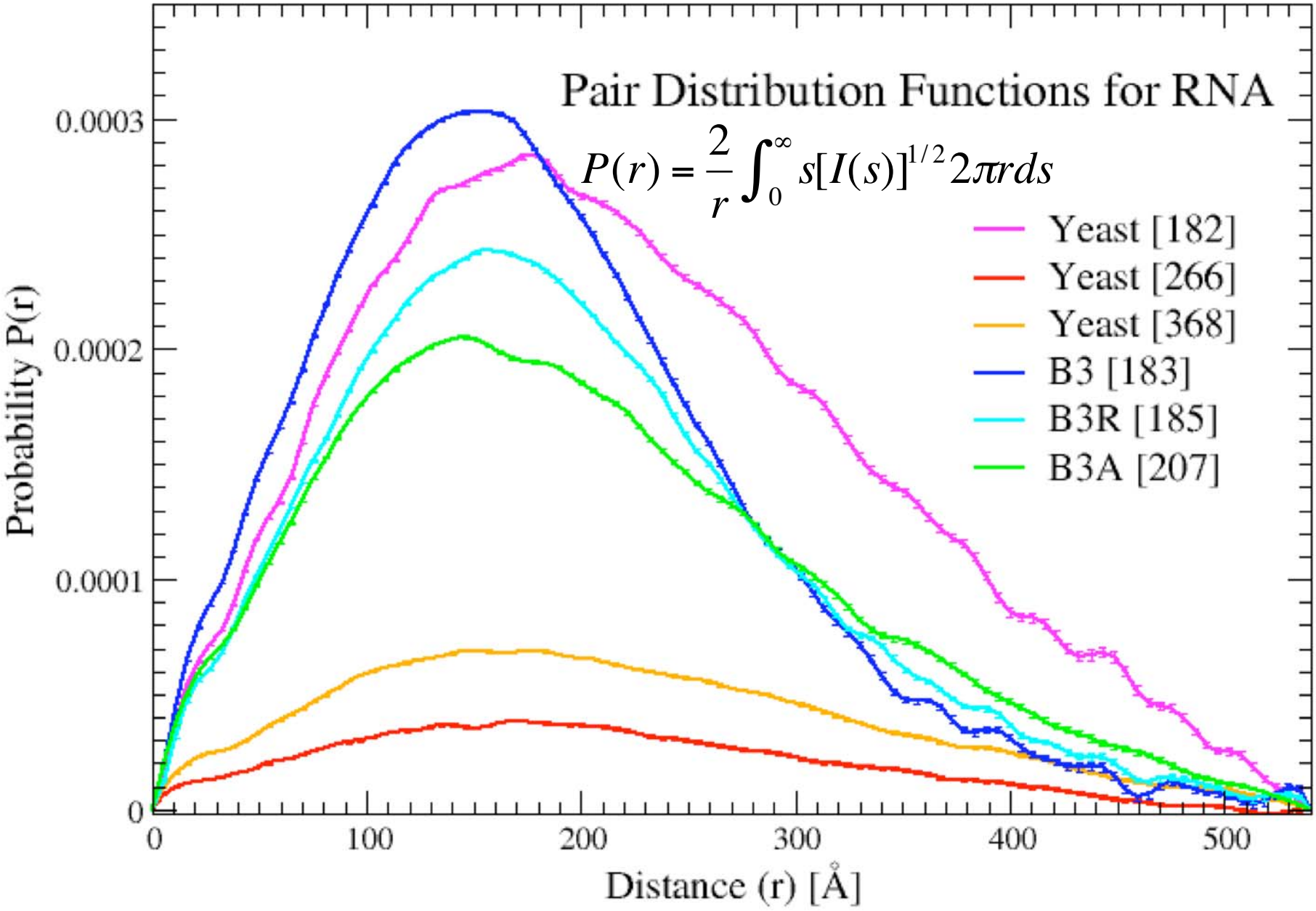
Control Sequences, 2117nt
(“random” Yeast RPS22B)

CHARACTERISTICS OF 2117 nt RNAs

	MLD	% Pairing	% GC
BMV3	183	62	48
BMV3REV	185	64	48
BMV3NG	207	59	48
YEAST182	182	59	43
YEAST266	266	61	35
YEAST368	368	62	32
AV. 509 YEAST RNAs	239	58	38
RANGE 509 YEAST RNAs	139 - 368	44 - 65	
AV. 509 RAND. RNAs	239		50
RANGE 509 RANDOM RNA	149- 358		50


Yeast [182]
Yeast [266]
Yeast [368]
B3 [183]
B3R [185]
B3A [207]
RNA Scattering Curves I(q)
Scattered Intensity (I) [arbitrary units]
Momentum Transfer (q) [Å^{-1}]
1
0.1
0.01
0.001
0.0001
0.05
0.1
0.15
0.2
0.25


Guinier Analysis
R_g MLD
182 ±1.72 [182] Yeast Control
189 ±7.05 [266] Yeast Control
177 ±7.95 [368] Yeast Control
142 ±1.42 [183] BMV RNA3
152 ±1.16 [185] BMV R3 Rev
161 ±2.23 [207] BMV R3 Anti
Beamstop
Guinier Region
$\log_{10} I$
q^2
0
-0.1
-0.2
-0.3
-0.4
-0.5
-0.6
-0.7
-0.8
-0.9
-1
-1.1
1e-05
2e-05
3e-05
4e-05
5e-05
6e-05
7e-05
8e-05
9e-05


Pair Distribution Functions for RNA
$P(r) = \frac{2}{r}\int_0^\infty s[I(s)]^{1/2} 2\pi r ds$
Yeast [182]
Yeast [266]
Yeast [368]
B3 [183]
B3R [185]
B3A [207]
Probability P(r)
0.0003
0.0002
0.0001
0
0
100
200
300
400
500
Distance (r) [Å]

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

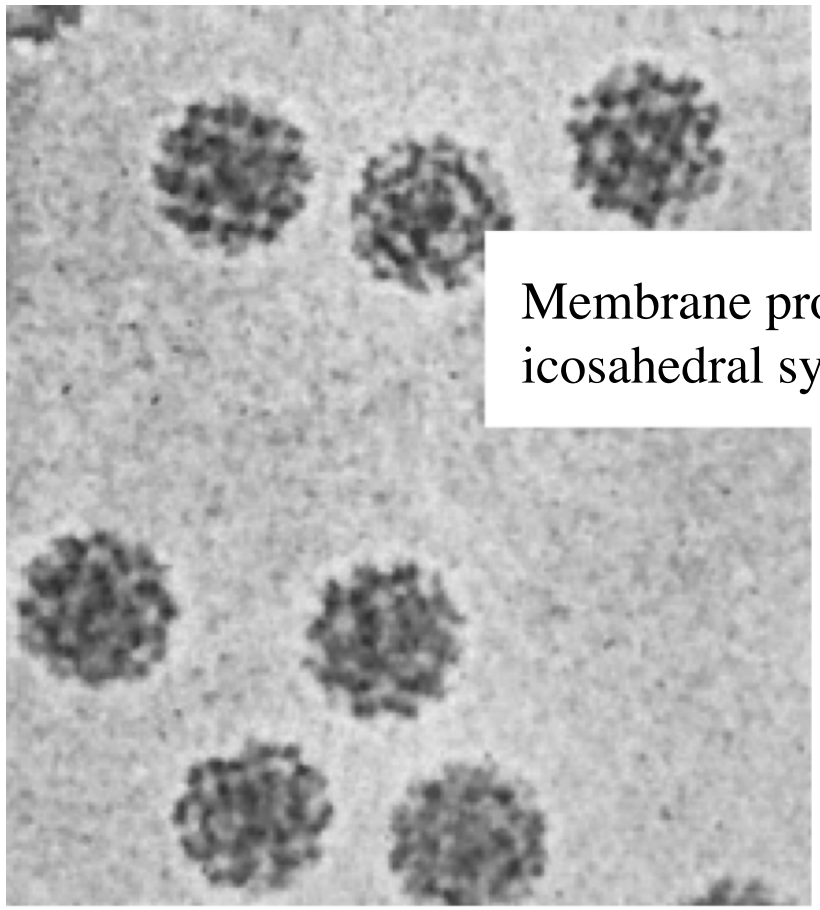
3) Transfect host cells and measure *in vivo* yields of:

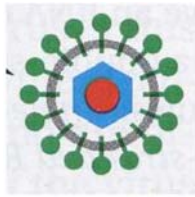
- **replicated RNA**
- **infectious nucleocapsids**

PUTTING IT TOGETHER

HOW TO MAKE AN ENVELOPED VIRUS

E.g., Semliki Forest Virus (SFV)



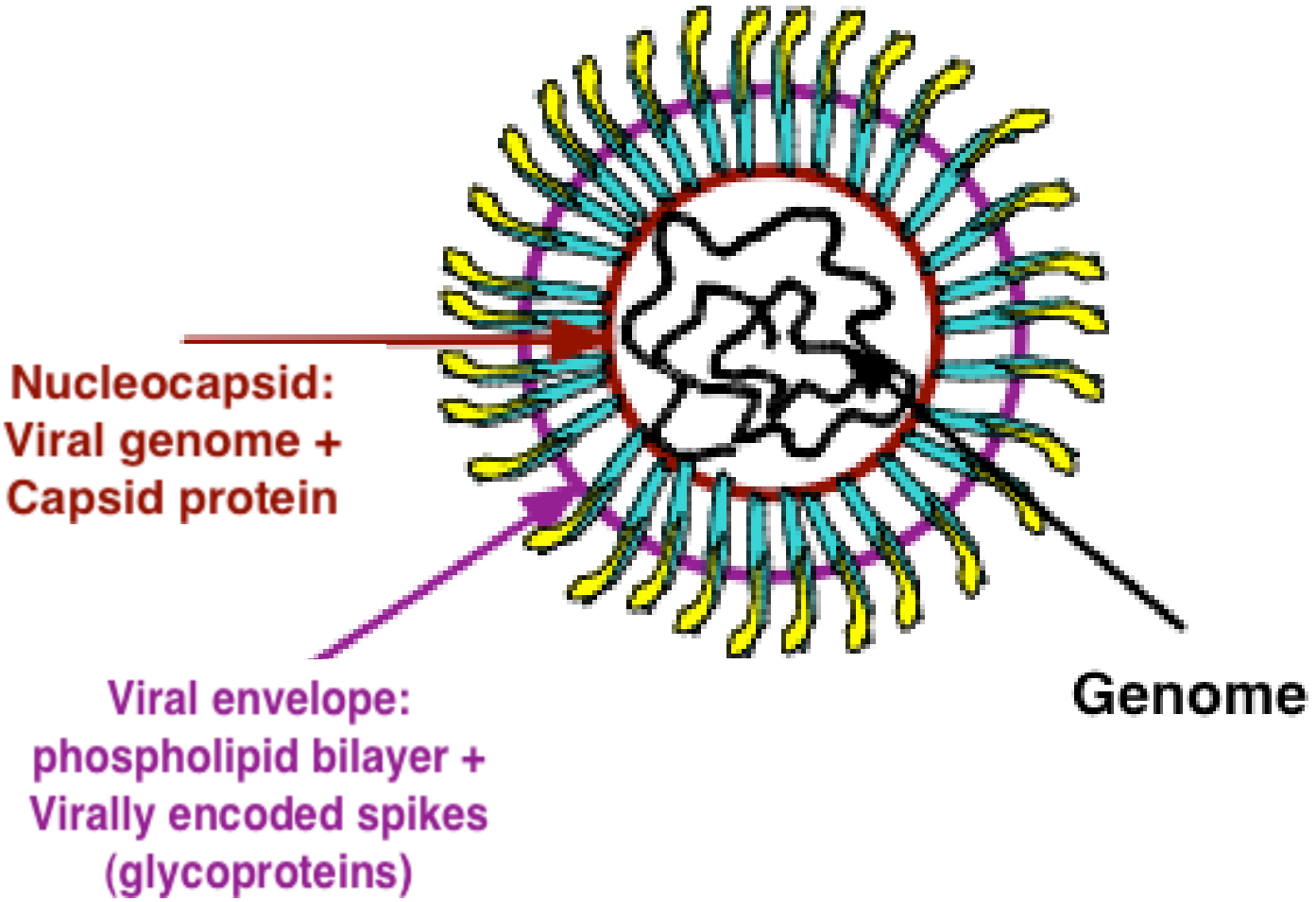


Membrane proteins have same icosahedral symmetry as capsid!

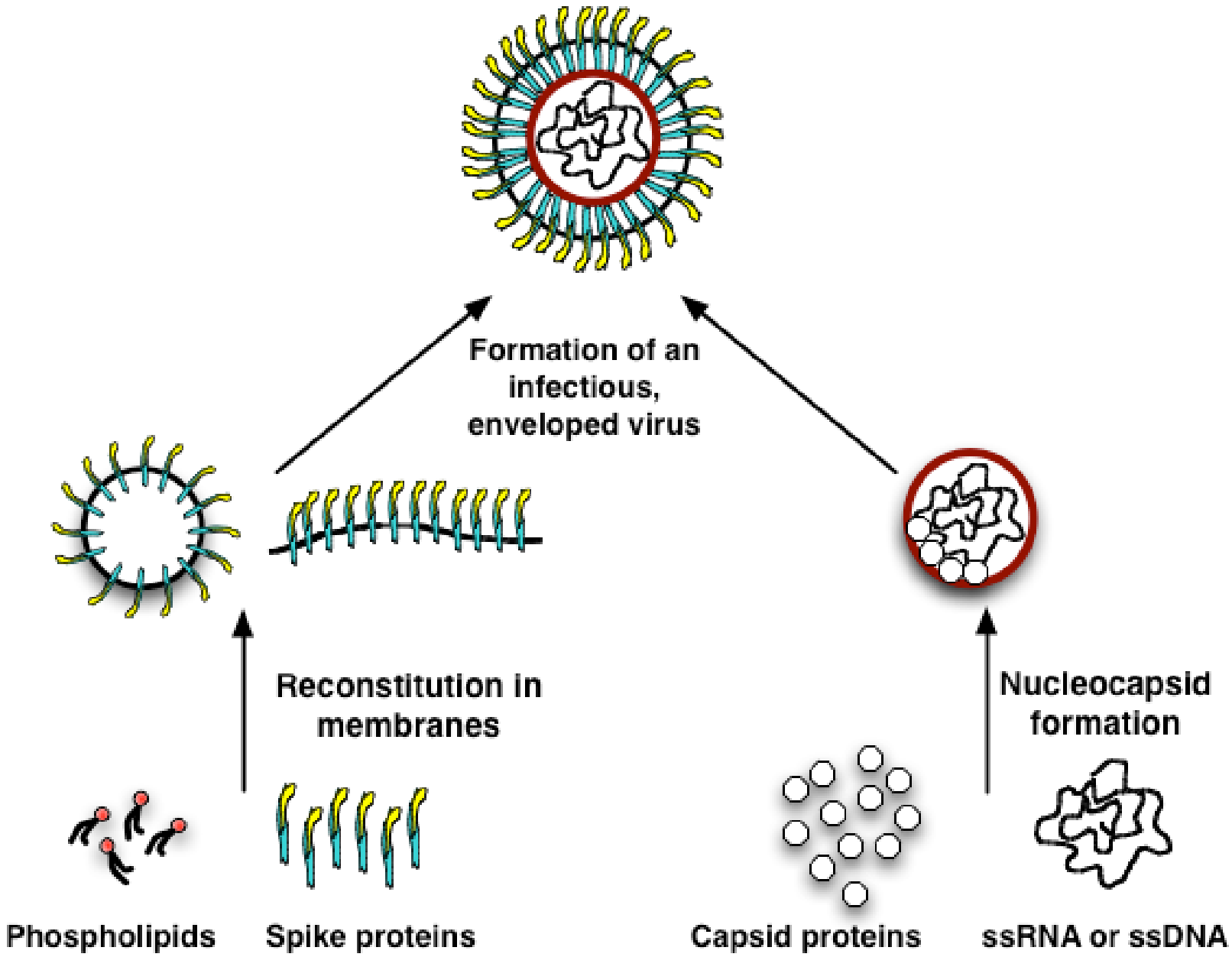
single molecule of ssRNA inside

Baker et al. (1999)

The “parts list” of an enveloped virus…

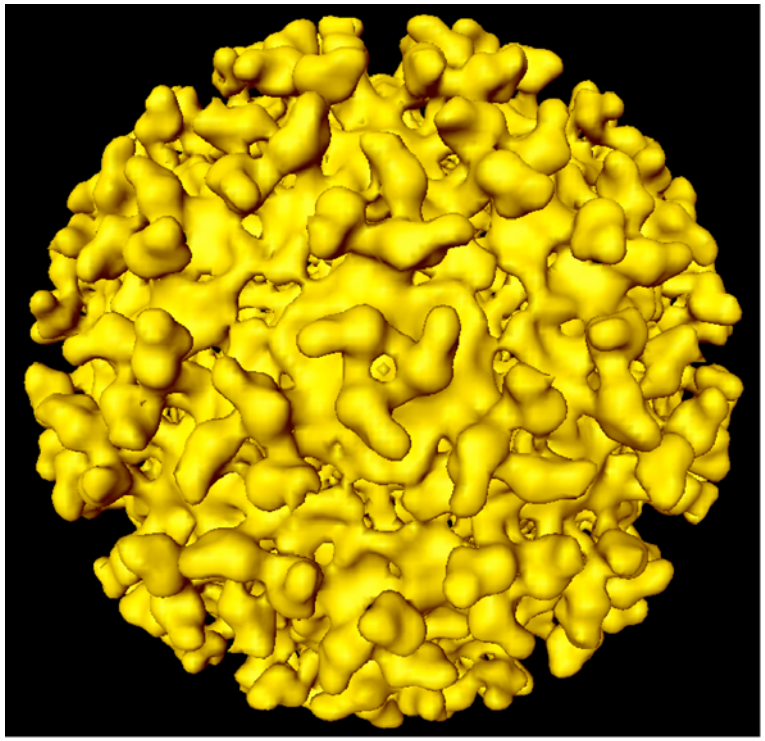


Putting the parts together…

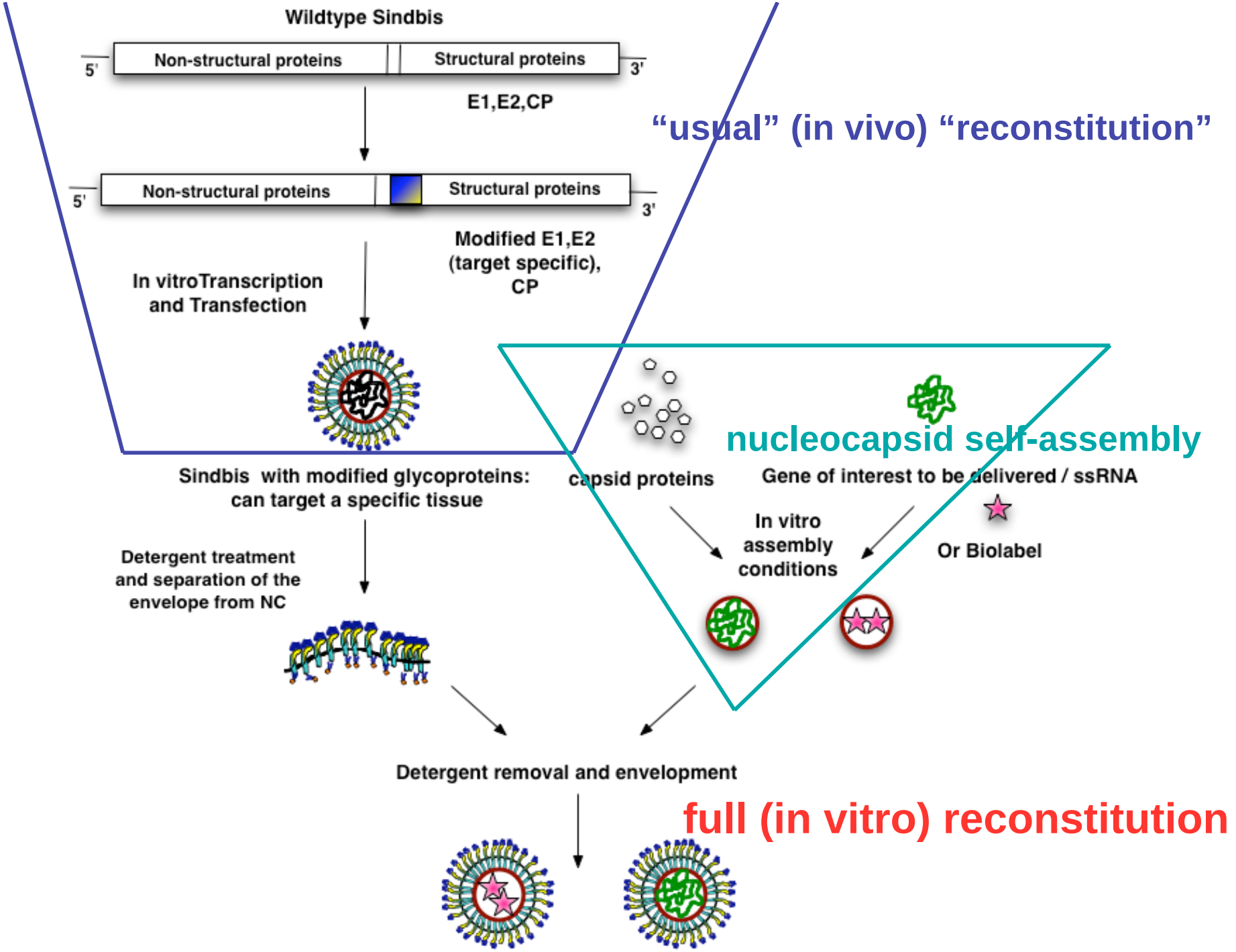


Sindbis: Virus of Choice

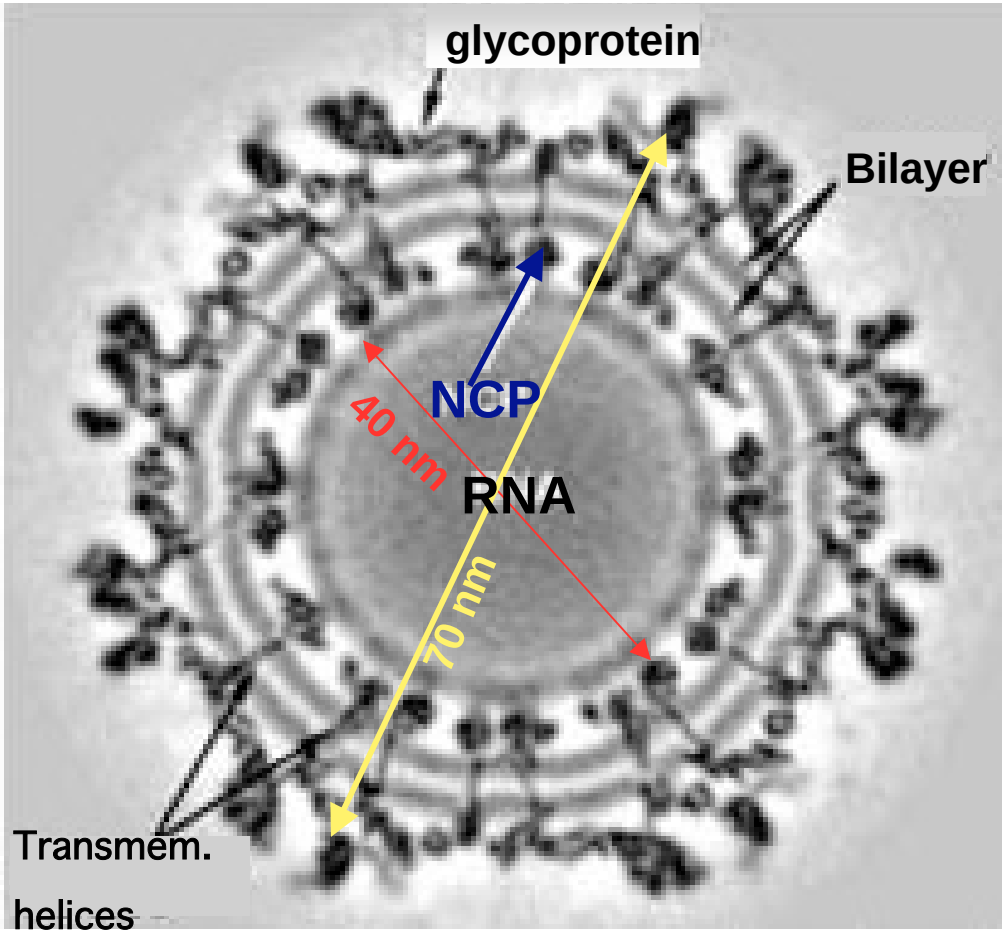
- Arguably the simplest animal virus: ***two-stage assembly*** life cycle
- Its nucleocapsid has been self-assembled *in vitro*
- One-to-one correspondence between each capsid protein and spike protein
- Strong interaction between capsid and spike proteins established *in vitro*
- Relatively benign



Sindbis reconstruction from cryo EM images (R. Kuhn)


Wildtype Sindbis
5'
Non-structural proteins
Structural proteins
3'
E1,E2,CP
5'
Non-structural proteins
Structural proteins
3'
Modified E1,E2
(target specific),
CP
In vitroTranscription
and Transfection
Sindbis with modified glycoproteins:
can target a specific tissue
capsid proteins
Gene of interest to be delivered / ssRNA
In vitro
assembly
conditions
Or Biolabel
Detergent treatment
and separation of the
envelope from NC
Detergent removal and envelopment
"usual" (in vivo) "reconstitution"
nucleocapsid self-assembly
full (in vitro) reconstitution

Sindbis cross-section



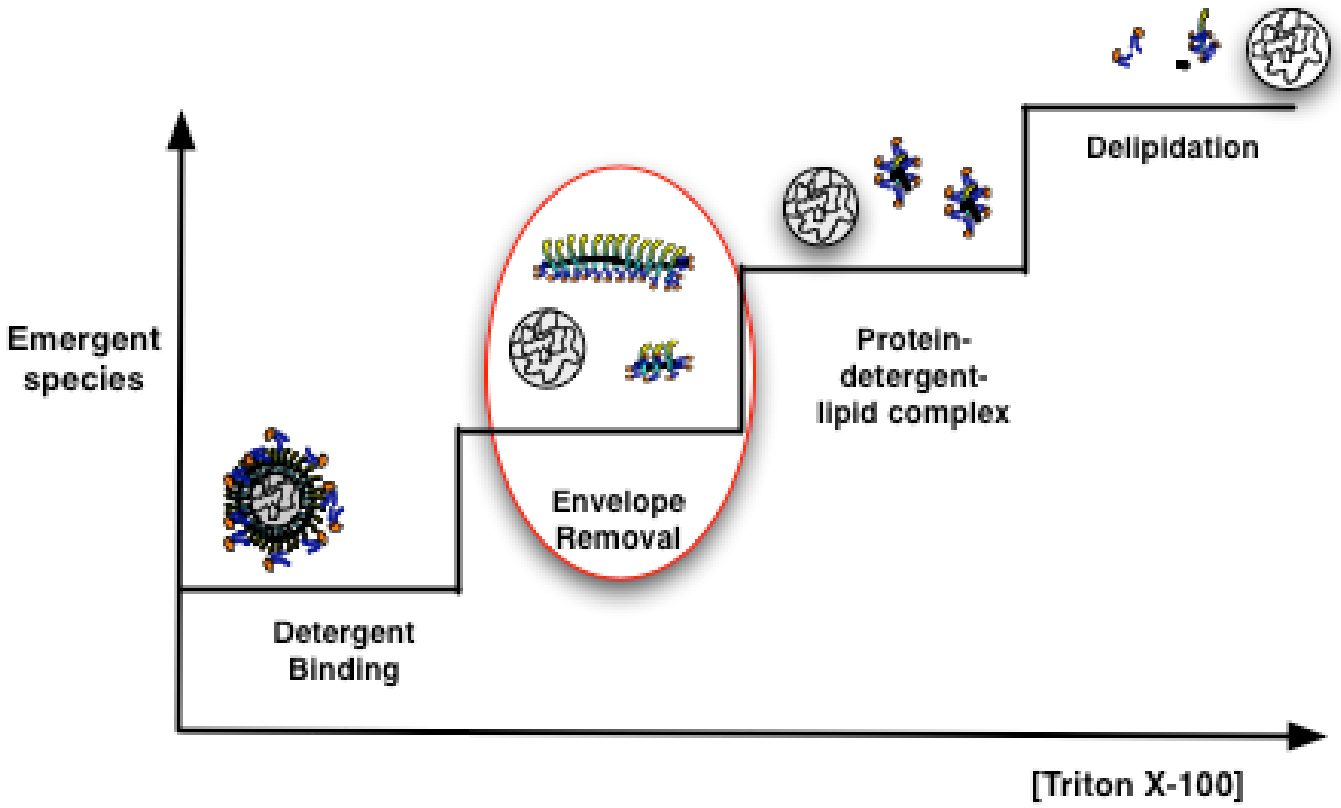
Zhang et al. (2002)

We “peel off” envelope from nucleocapsid using detergent; then re-assemble it around the NC by removing detergent

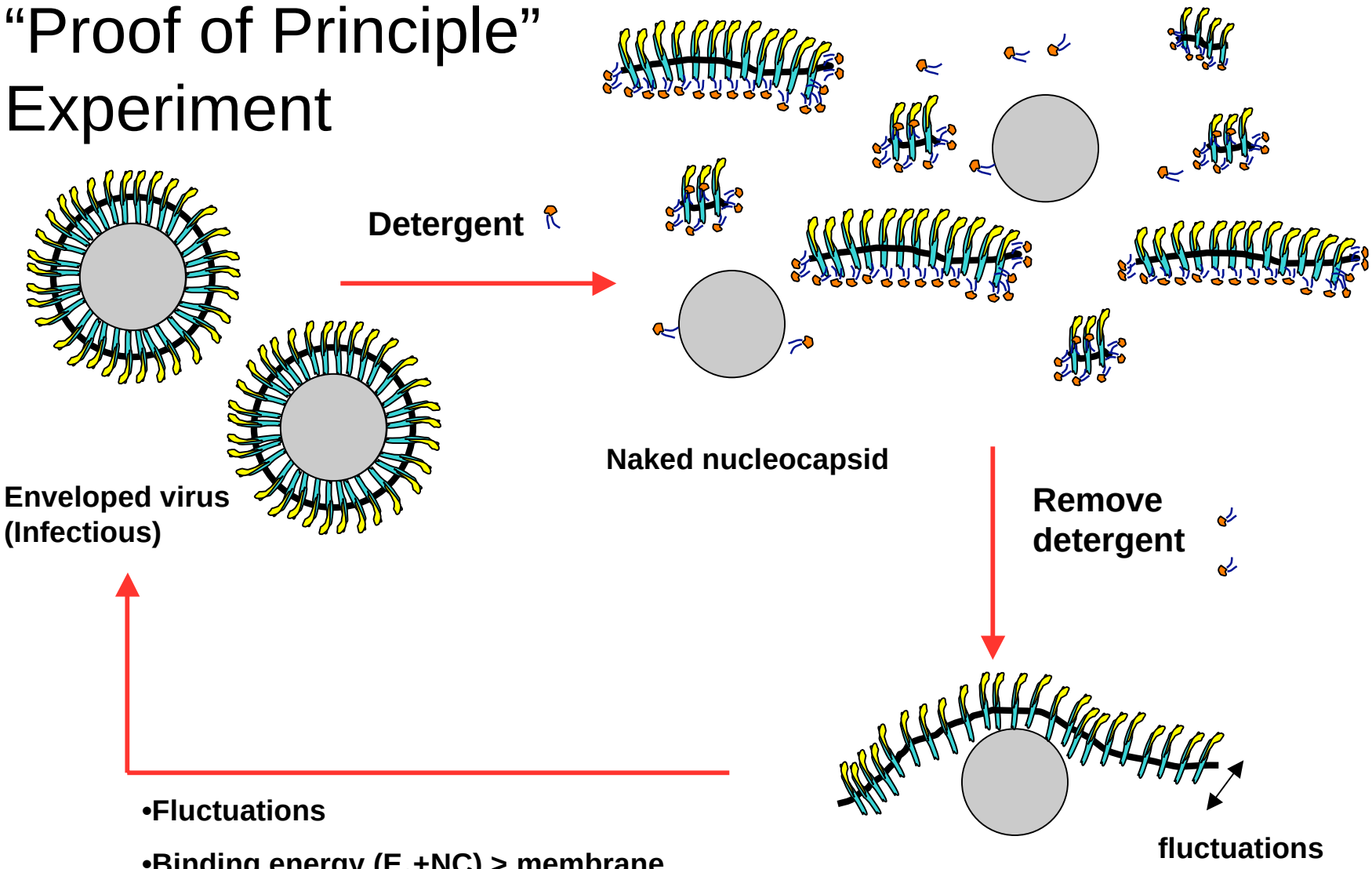
Peeling off the membrane is possible!

Dissociation of the viral envelope from the nucleocapsid is stepwise (for SFV and Triton X-100 system)

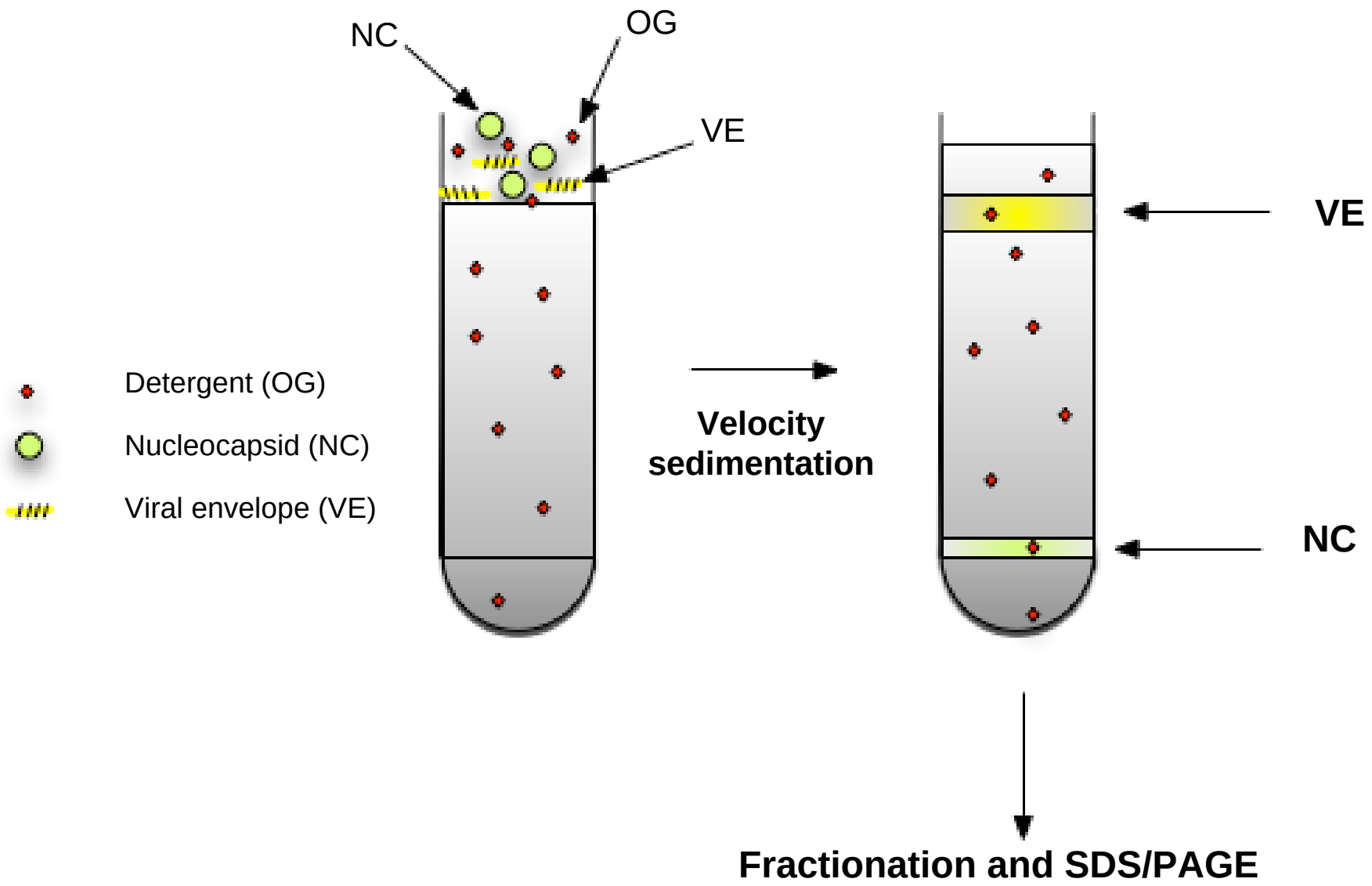
Helenius and Söderland, *Biochimica et Biophysica Acta*, 307 (1973), 287-300.



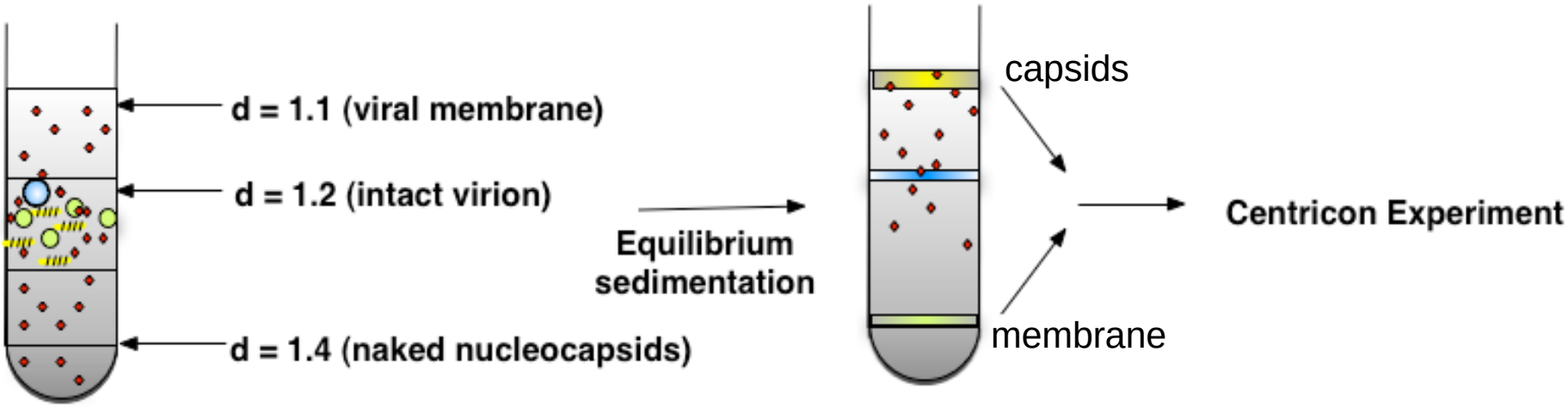
“Proof of Principle” Experiment



Physical separation of nucleocapsids and viral membranes after detergent treatment



- Can also separate the viral envelopes from nucleocapsids by equilibrium sedimentation
- Mix the purified components on centricons, to rid of detergent and allow intact virions to reform
- Show that infectious particles have been reconstituted from ***non***-infectious components



Determine number of infectious virions before and after detergent treatment, and then after detergent removal

Plaque assay: allows us to count ***individual*** infectious virions

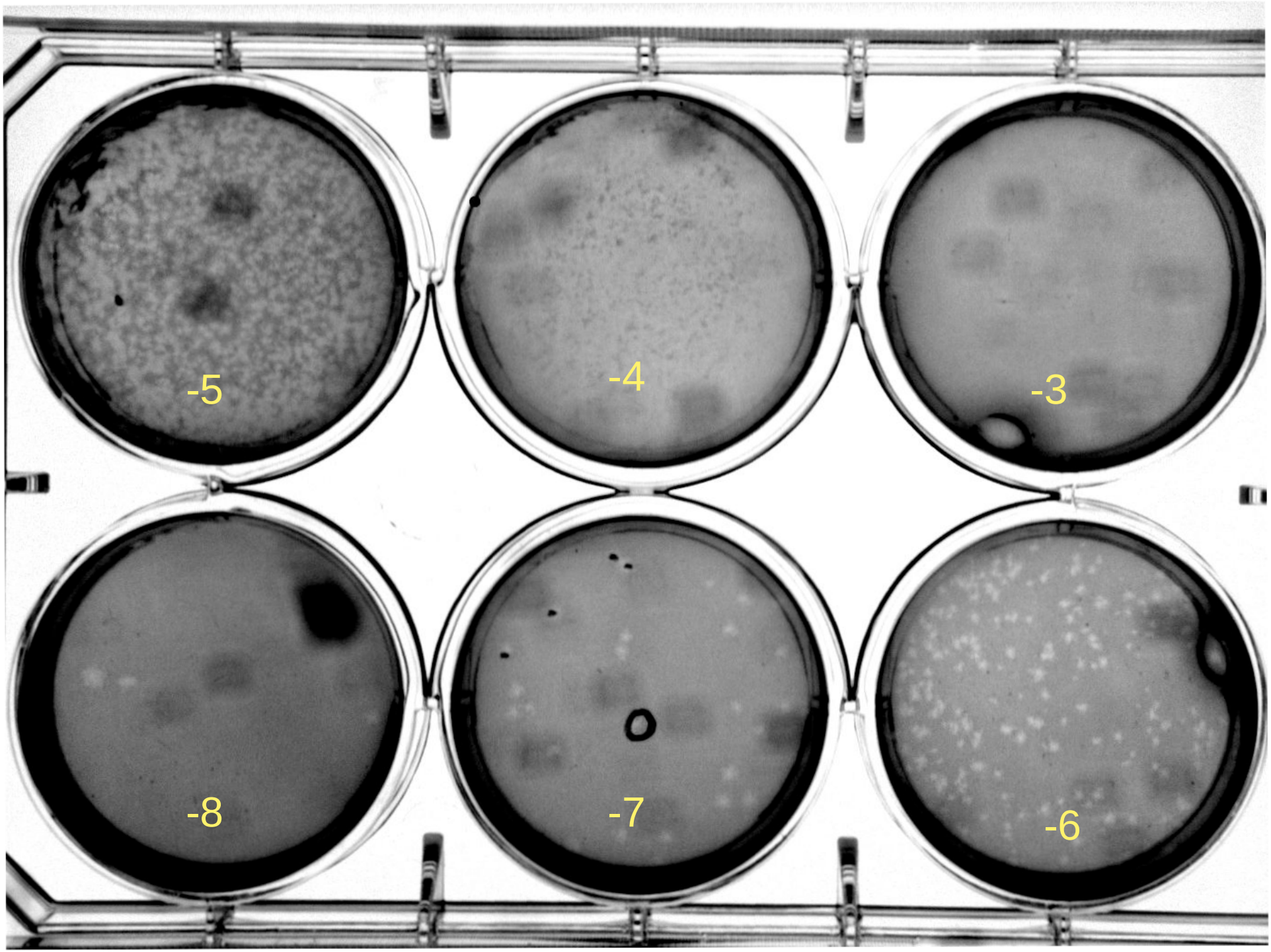
Plaque: a region of infected cells, detectable by naked eye

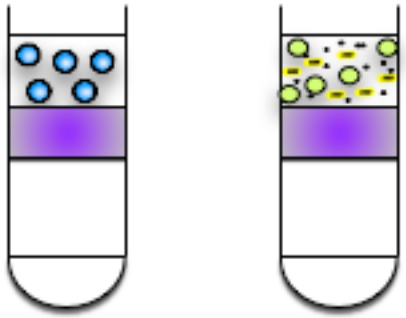
Substrate: a monolayer of host (2-day-old BHK) cells

Need to be able to count the cells that are infected: serial (powers of ten) dilutions lead to small (distinct, and hence countable) numbers (1's, 10's, 100's) of infectious virions (hence infected cells: plaques) per plate

Only interested in the number of infected cells that were *initially* infected by the virus: agarose is added to decrease the mobility of the virions in the medium

Plaques are visualized by a stain


-5
-4
-3
-8
-7
-6



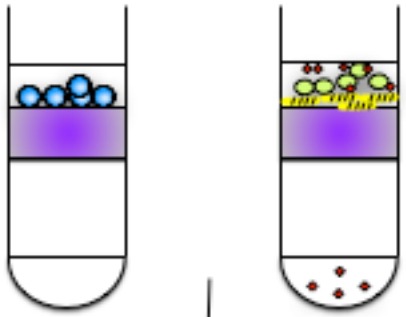
Control

Detergent-treated sample

Detergent-treated sample unfiltered

Unfiltered control

Concentrate and wash away the detergent x5



Incubate

Plaque assay:

~10^8 pfu/mL

~ 10^4 pfu/mL

Partial recovery !!!

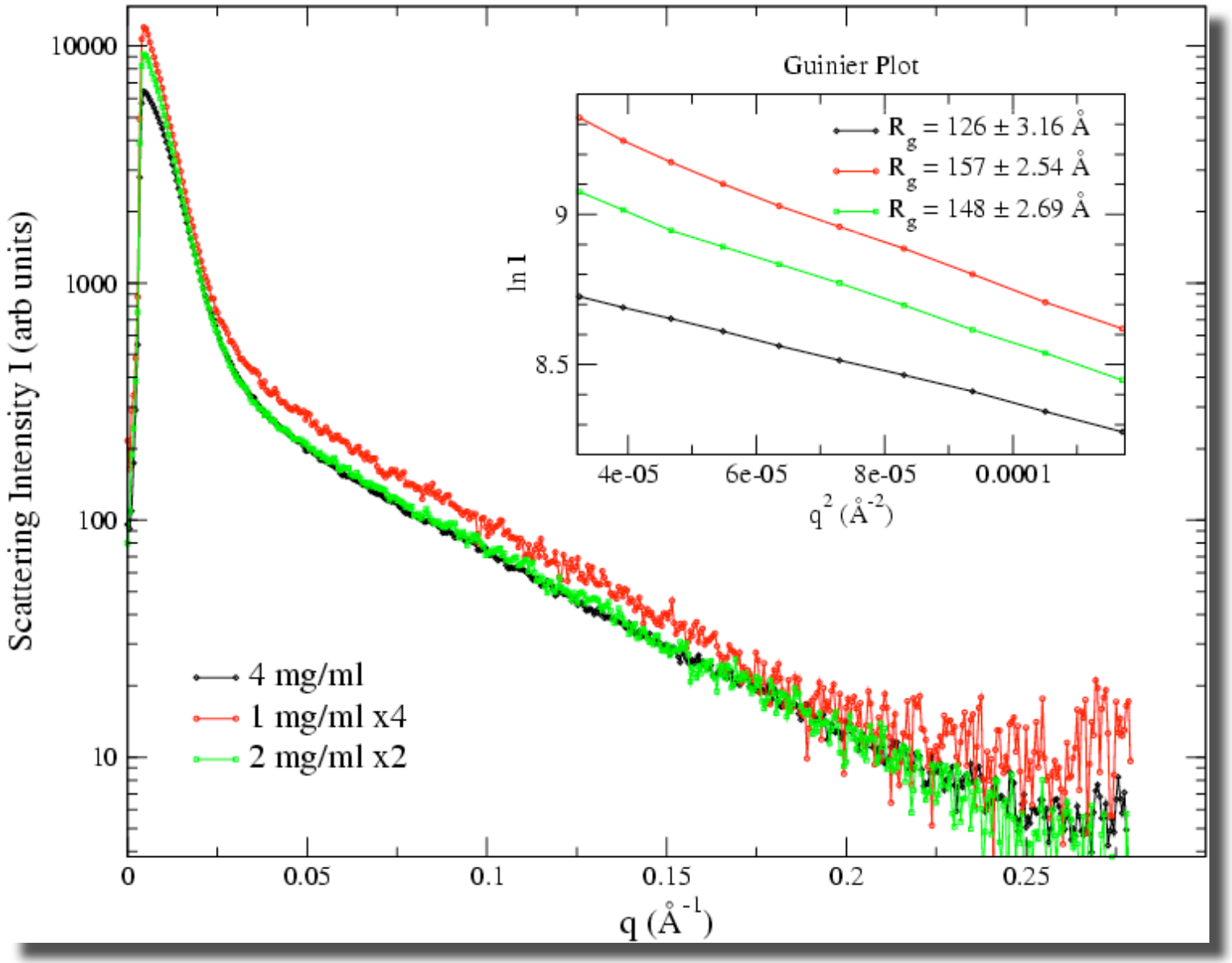
Zero

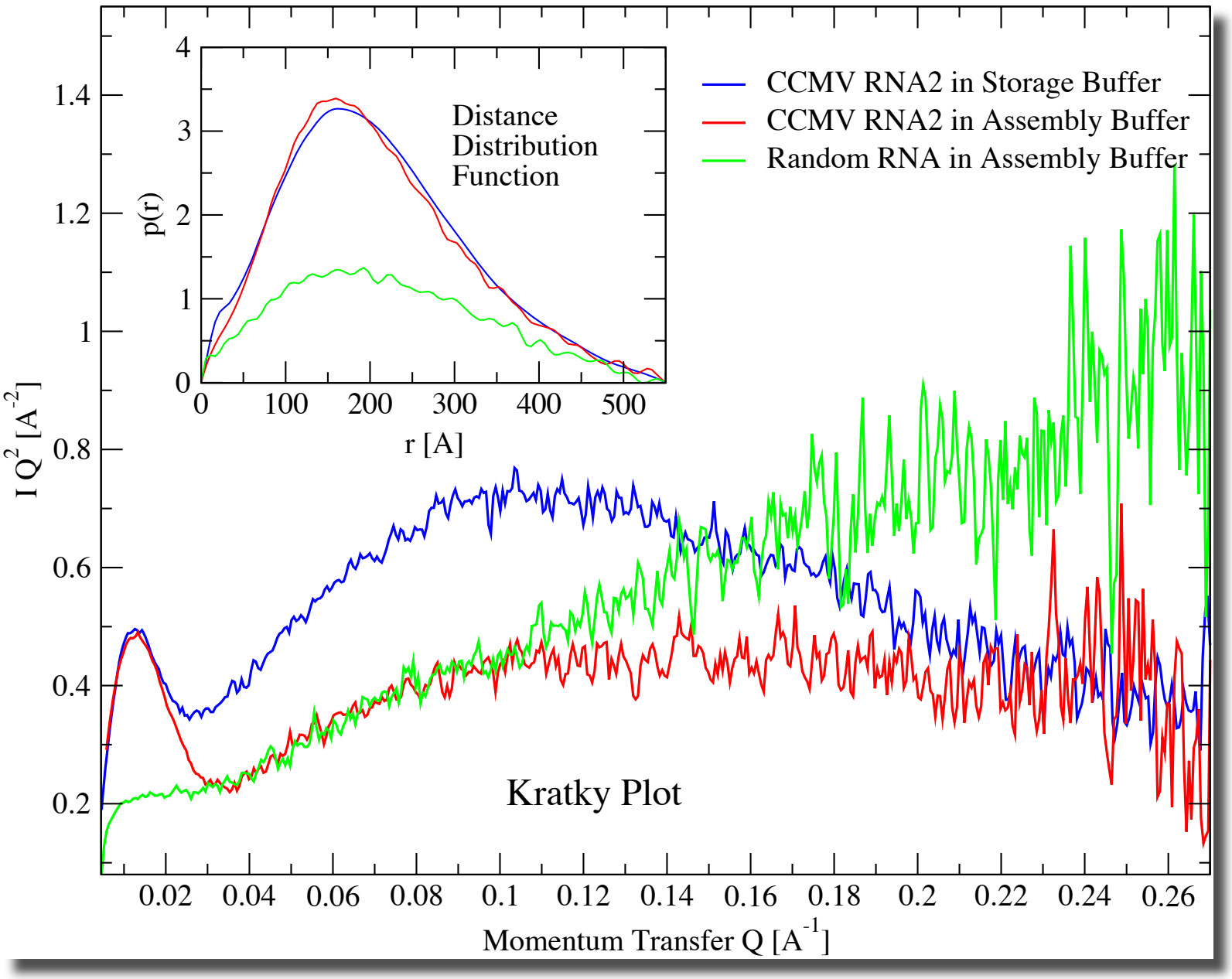
~ 10^8 pfu/mL

Future experiments

- Structural studies of the infectious particles by EM
- Sedimentation studies of infectious particles formed after Centricon experiment
- Increase the yield of recovery by concentrating the components:

Make NCs in vitro and add them to the reaction (taking advantage of the special Sindbis genome…)


Guinier Plot
R_g = 126 ± 3.16 Å
R_g = 157 ± 2.54 Å
R_g = 148 ± 2.69 Å
ln I
9
8.5
4e-05
6e-05
8e-05
0.0001
q^2 (Å^{-2})
Scattering Intensity I (arb units)
10000
1000
100
10
4 mg/ml
1 mg/ml x4
2 mg/ml x2
0
0.05
0.1
0.15
0.2
0.25
q (Å^{-1})

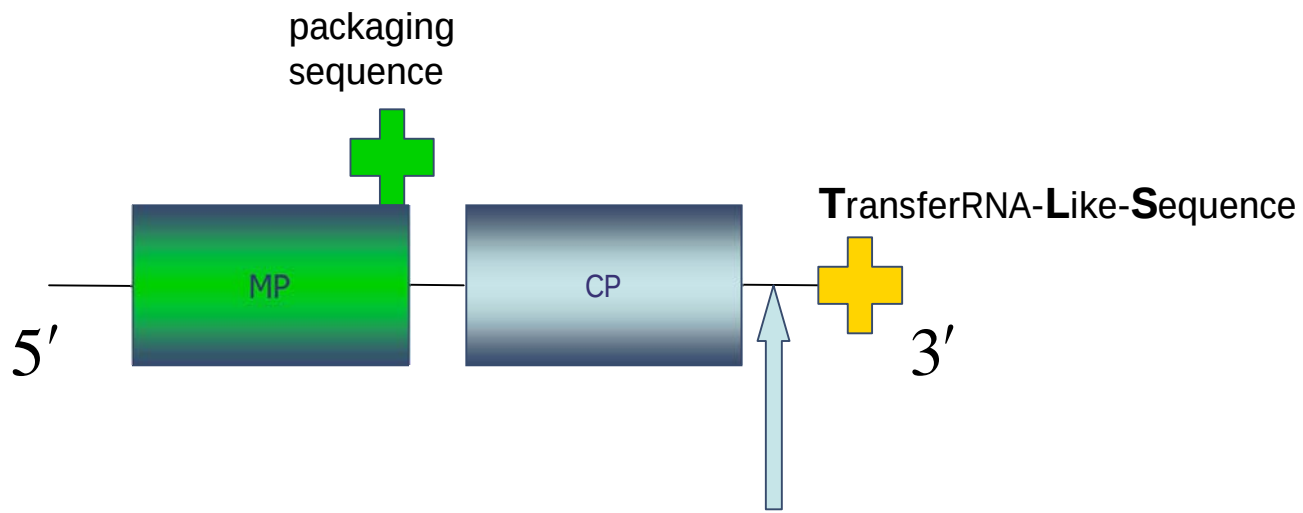

CCMV RNA2 in Storage Buffer
CCMV RNA2 in Assembly Buffer
Random RNA in Assembly Buffer
Distance
Distribution
Function
p(r)
r [A]
Kratky Plot
I Q^2 [A^{-2}]
Momentum Transfer Q [A^{-1}]

Selected 2117nt-long sequences from 500 chosen at random from 1M bps of the yeast (S. cerivisiae) Chromosome 12, some including no coding regions, some fully coding, and most mixed (i.e., introns + exons from different genes)

RNA	First nt# (ChrXII)	Last nt#	GC Fraction	Max. Ladder Dist.	%nt paired
R22	GG+855700	857814	0.351	266	60.56
LD326	688026	690142	0.324	368	61.88
SC326	687701	689817	0.326	347	63.01
LD414	874322	876438	0.427	148	58.10
SC414	874269	876385	0.427	182	59.04
NC1	449324	451440	0.342	236	56.49
NC2	390695	392811	0.402	197	59.23
FC1	352947	355063	0.368	215	57.81
B3E	na	na	0.478	183	61.69
B3R	na	na	0.478	185	64.45

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

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

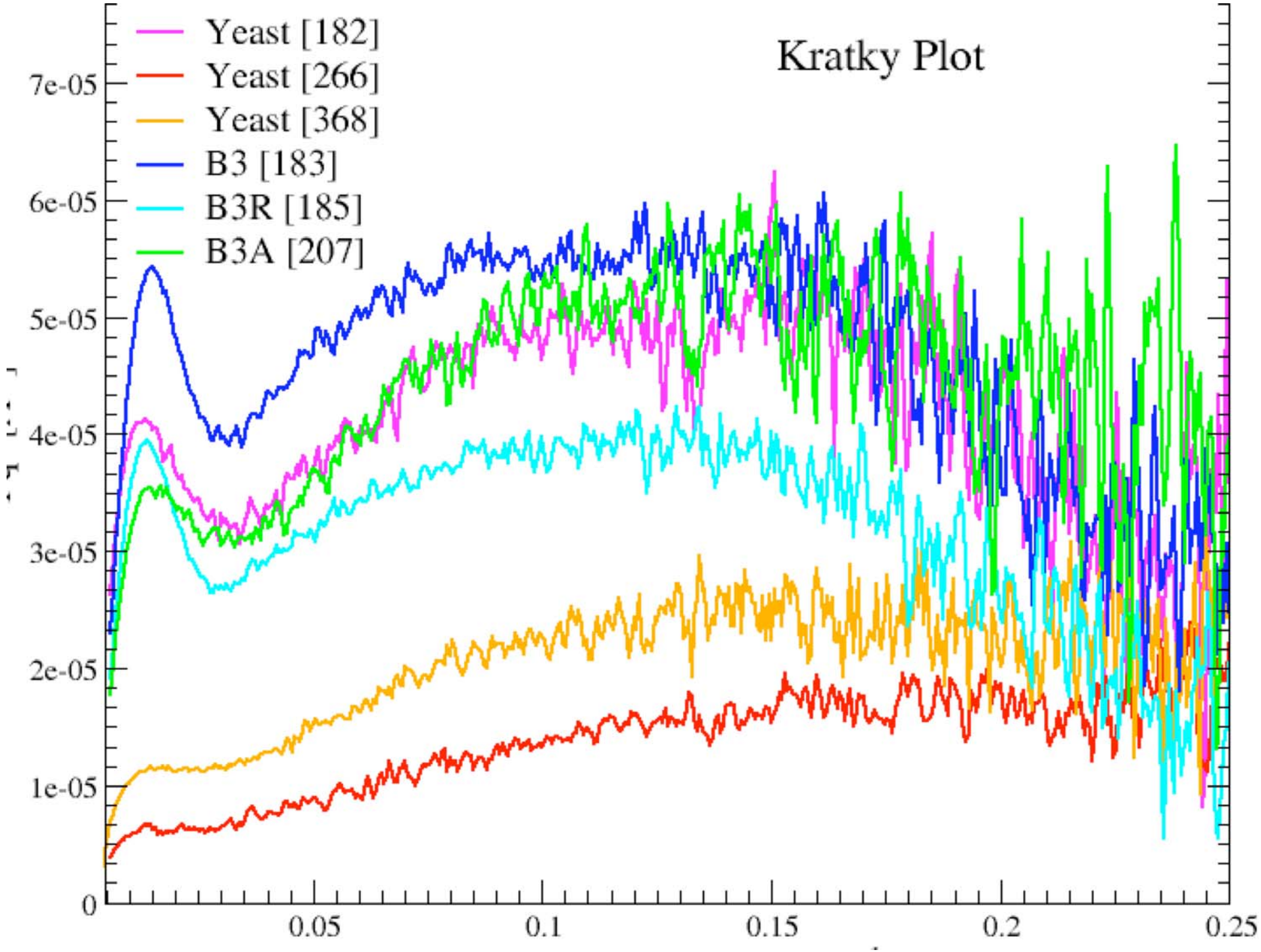


Measure R_g (and *in vitro* and *in vivo* packaging efficiencies…)

YEAST SEGMENTS WITH MIN. MAX LADDER DISTANCE		
position	max ladder	%pairing
325	139	49.60
378	141	52.05
219	142	54.61
414	148	58.10
210	153	55.08
274	154	51.68
6	156	44.40
YEAST SEGMENTS WITH MAXIMUM MAX. LADDER DISTANCE		
position	max ladder	%pairing
326	368	61.88
319	335	59.42
440	335	58.57
199	325	58.48
63	324	60.46
	max. ladder	**stdev**
AVERAGE YEAST	239.0	38.9
position	max ladder	%pairing
BMV3-ENGINEERED	183	61.69
BMV3-REVERSE	185	62.45
RANDOM-2117	266	60.56
AVERAGE RANDOM	236	~61
(calc'd)		

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

average yeast MLD $\approx$ average random MLD


Kratky Plot
Yeast [182]
Yeast [266]
Yeast [368]
B3 [183]
B3R [185]
B3A [207]
7e-05
6e-05
5e-05
4e-05
3e-05
2e-05
1e-05
0
0.05
0.1
0.15
0.2
0.25