Fitting high resolution structures into low resolution EM maps

Michael Rossmann

Purdue University

Fitting Processes

1. Map scaling
2 Symmetry constraints
3. Fitting criteria
   a. fit of atoms into density
   b. avoiding negative density
   c. steric hindrance, inter atomic clashes
   d. restraints imposed by known structural features
4. Combining different criteria
   a. normalization of each measurement
5. The search process
   a. rotational search
   b.multi-dimensional “climb” or least squares
6. Verification
   a. hand of map
   b. subunit contacts
7. Problems
   a. symmetry missmatches
   b. unknown structural components
   c. uninterpreted density

Map Scaling

Minimize $\Sigma\ [\rho_1(x_1,y_1,z_1)\ \ - (a + b\ \rho_2(x_{2,}y_2,z_2))]^2$

where ρ_1 is the reference map (e.g. X-ray virus map)

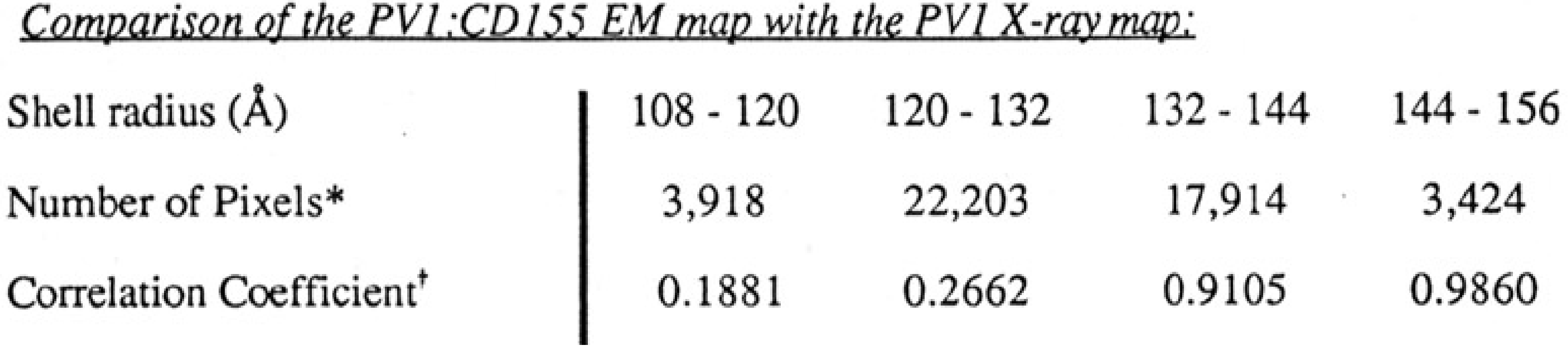
and ρ_2 is the map of interest (e.g. virus plus ligand complex)

And $x_1 = x_2 + \delta x_{2,},\ \ y_1 = y_2 + \delta y_2,\ \ \ z_1 = z_2 + \delta z_2$

Requiring interpolation for determinin ρ_2

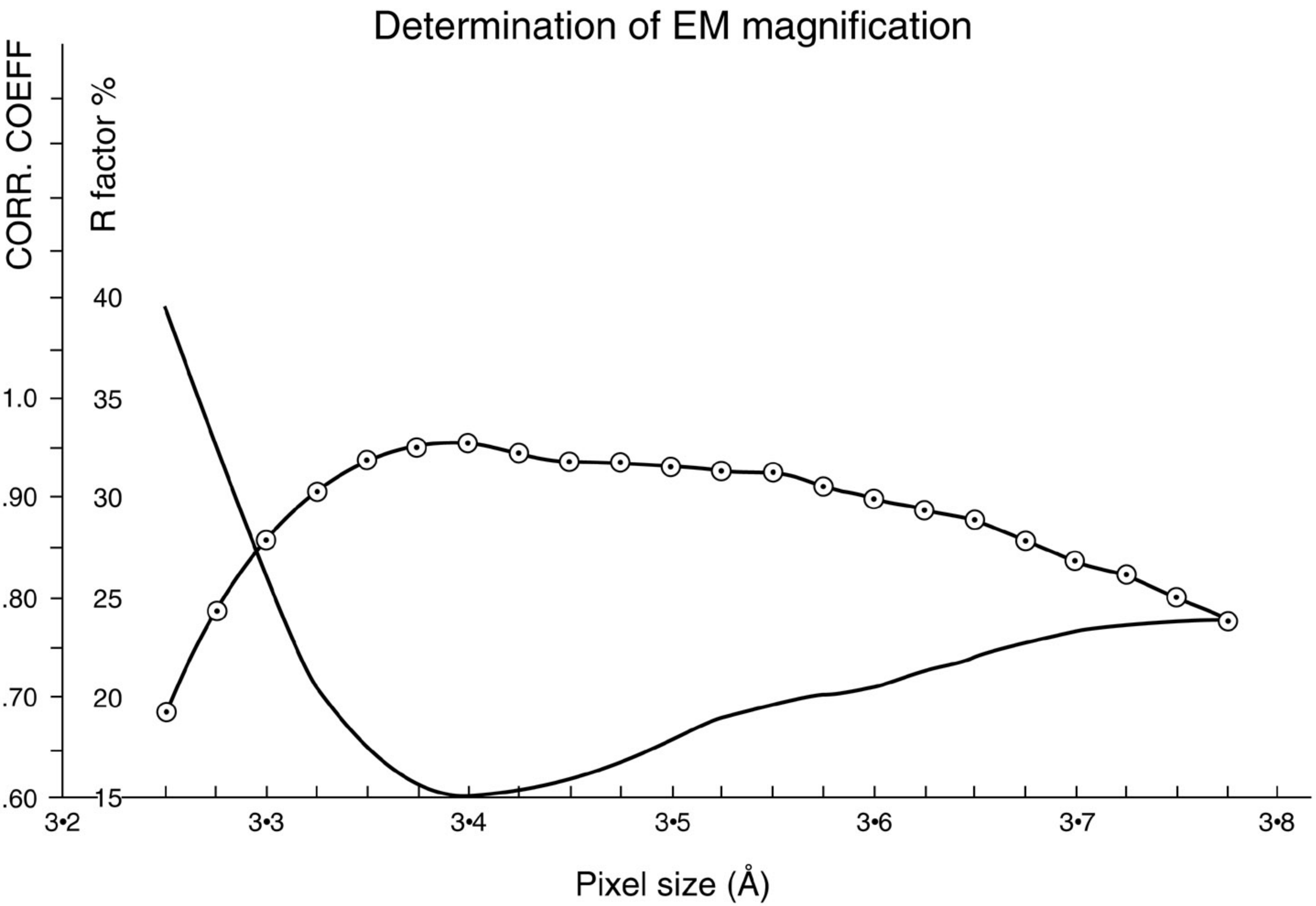
Or maximize the correlation C, where

$C= [\Sigma(<\rho_1> - \rho_1)(<\rho_2> - \rho_2)] / [\Sigma(<\rho_1> - \rho_1)^2].[\Sigma(<\rho_2> - \rho_2)^2]$



Comparison of the PV1:CD155 EM map with the PV1 X-ray map:

Shell radius (Å)	108 - 120	120 - 132	132 - 144	144 - 156
Number of Pixels*	3,918	22,203	17,914	3,424
Correlation Coefficient†	0.1881	0.2662	0.9105	0.9860


Determination of EM magnification
CORR. COEFF
R factor %
40
1.0
35
.90
30
.80
25
.70
20
.60
15
3•2
3•3
3•4
3•5
3•6
3•7
3•8
Pixel size (Å)

Symmetry Constraints

Let the atomic positions of a model be given by (X,Y,Z), or, in vector notation, by **X**, in an orthogonal coordinate system.

Let the origin of the model (defined by its center of mass) be at **S**.

Let the rotation matrix required to place the model into the "reference" EM density be $[E]$, Then

$$\mathbf{X'} = [E]\mathbf{X} + \mathbf{d},$$

where **X'** are the coordinates of the model atoms in the EM map and **d** is a translation vector.

Let **S'** be the approximate target position in the EM map for placement of the model's origin. Then

$$\mathbf{S'} = [E]\mathbf{S} + \mathbf{d}$$

and, hence,

$$\mathbf{d} = \mathbf{S'} - [E]\mathbf{S}$$

or

$$\mathbf{X'} = [E](\mathbf{X} - \mathbf{S}) + \mathbf{S'}\ .$$

Let the reference molecule be reproduced by *M* “crystallographic” and T “NCS” symmetry operations given by $[R_m]$ ($m = 1, M$ and t=1, T). Thus

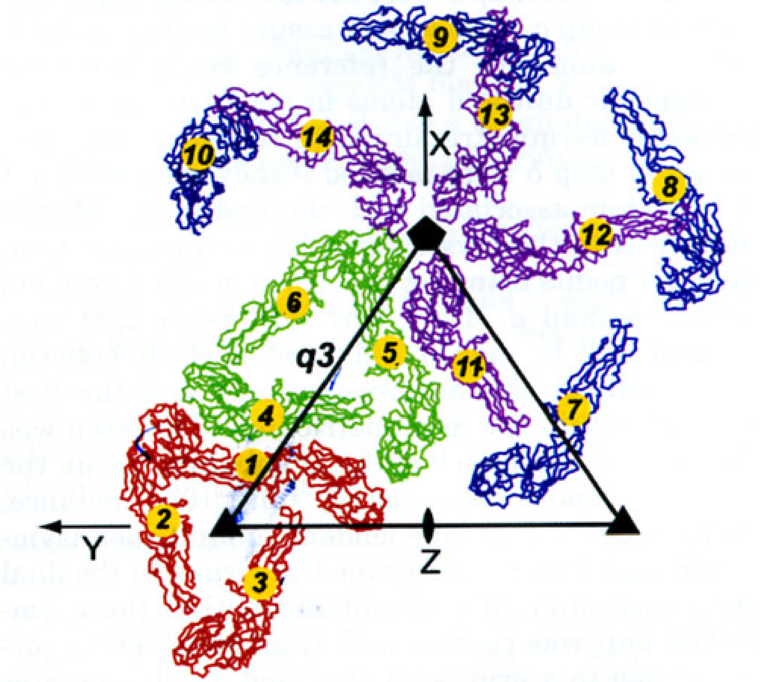
$$\mathbf{X}'' = [R_{m,t}]\mathbf{X}'$$

And hence using

$$\mathbf{X}' = [E](\mathbf{X} - \mathbf{S}) + \mathbf{S}'$$

It follows that

$$\mathbf{X}'' = [R_{m,t}]([E](\mathbf{X} - \mathbf{S}) + \mathbf{S}')$$



Sindbis Virus

M=60 icosahedral operators

T = 4 quasi symmetry NCS operators

Fitting criteria

a. fit over N atoms into density

$$sumf = 100.\sum_{T} (\sum_{N} \rho(X'') / TN\rho_{norm}$$

where ρ_{norm} is either the maximum or rms density

b. The number of atoms (N') in negative density, expressed as a %

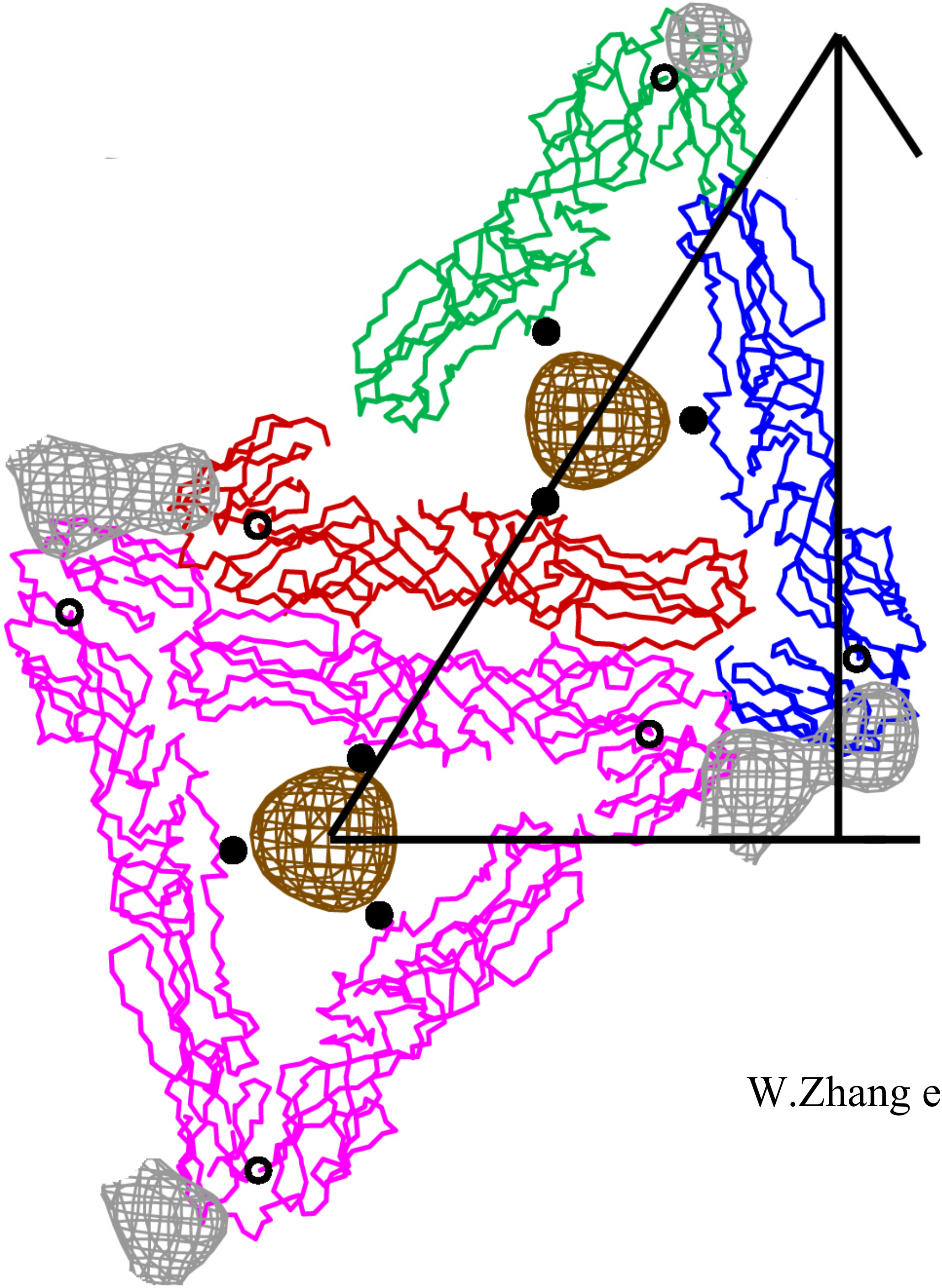
$$-den = 100.\sum_{T} (N') / TN$$

c. The number of atoms (N'') that approach atoms in another molecule to within 3.4A, expressed as a %.

$$clash = 100.\sum_{T} (N'') / TN$$

d. The average or rms distance between L specific fixed points (R_i) in the map and specific atoms on the molecule (X''_i) (e.g. Carbohydrate moities in the map and corresponding aas).

$$\text{avgdist} = \sum_{L} |(R_i - X''_i)| / L$$



Fitting the E1 protein of Sindbis virus : Using carbohydrate sites as restraints

W.Zhang et al, J.Virol, 2002, 76, 11645-11658

Use of Restraints

1. Minimizing the distance between recognizable features in the cryoEM map and the associated atomic Group of the molecule being fitted
2. Restraining the molecule being placed in a map to use a specific contact region to other parts of the structure
3. Keeping a short distance between the C-end of one domain and the N-end of the next, independently fitted, domain.

Combining different criteria

$$R_{crit} = \Sigma\ \omega_i s_i\ [(\nu_i - \langle\nu_i\rangle) / \sigma(\nu_i)] / \Sigma\omega_i$$

Where v_i is the value of the ith criterion,

$\langle\nu_i\rangle$ is the standard deviation of ν_i taken over a set of randomly oriented molecular fits into the density,

ω_i is the weight (usually 1.0) to be placed on the given criterion and

s_i is +1.0 if the criterion is to be maximized (e.g. *sumf*)
or -1.0 if the criterion is to be minimized (e.g. *–den*, *clash*
and *avgdist*.)

Fitting the E1 protein of Sindbis virus. The top 25 best fit converge to only 4 different fits on refinement

a. Values of criteria

Fit No	R_{crit}	sumf	clash	-den	avgdist Å
13	0.98	39.3	0.5	9.2	21.9
10	0.81	37.3	2.2	10.1	20.5
14	0.26	36.3	3.7	11.9	21.2
25	-2.37	39.2	17.5	10.1	28.7

b. Criteria expressed as the number of **σ** above mean

Fit No	R_{crit}	sumf	clash	-den	avgdist
13	0.98	2.38	0.19	1.52	0.93
10	0.81	1.40	-1.35	1.18	1.48
14	0.26	0.48	-2.78	0.42	1.22
25	-2.37	2.32	-23.10	1.15	-1.67

The search process

2. Explore all unique values of the three Eulerian angles that define the [E] rotation matrix, using fairly large angular intervals

   $0 \leq \theta_1 < 2\pi;$ $\quad 0 \leq \theta_2 \leq \pi,$ $\quad 0 \leq \theta_3 < 2\pi$

2. Rank according to sumf

3. Use results for determining the mean and standard deviation (σ) for each criterion required to calculate R_{crit}.

4. Refine the top n (e.g. 100) best fits by a six dimensional “climb” on R_{crit}, using fine angular and positional intervals.

5. Eliminate all but one of closely similar fits, leaving only distinctly different fits.

Note: fitting more than one rigid body at a time can be done sequentially and refined by least squares

Refine using “Climb”

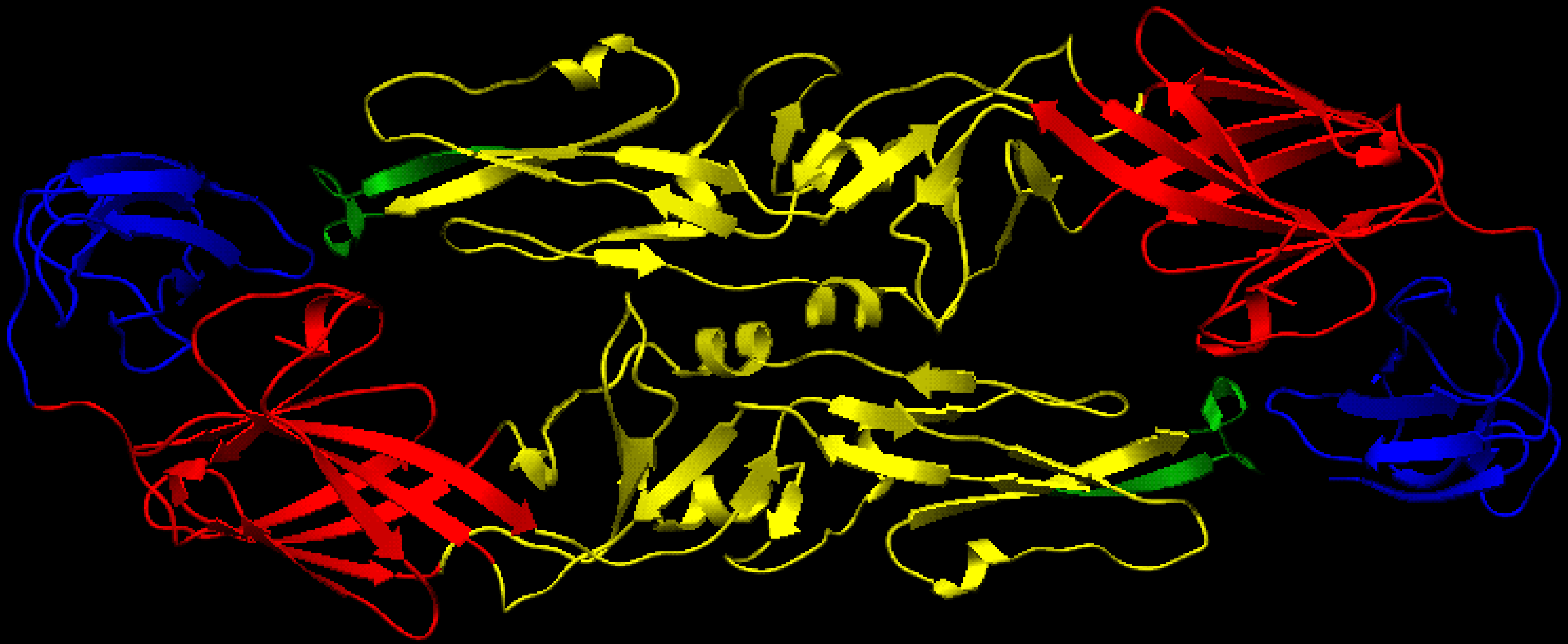
R_{crit} values at end of climb

Refining the placement of the E1 glycoprotein

Into Sindbis virus cryoEM density

param	$\xi - \Delta\xi$	ξ	$\xi + \Delta\xi$	ξ	$\Delta\xi$
θ_1	1.016	1.029	1.022	357.0	0.25
θ_2	1.008	1.029	1.026	40.5	0.25
θ_3	1.021	1.029	1.021	193.5	0.25
x	0.996	1.029	1.011	23.9	0.50
y	1.028	1.029	0.990	68.3	0.50
z	0.963	1.029	1.015	284.5	0.50

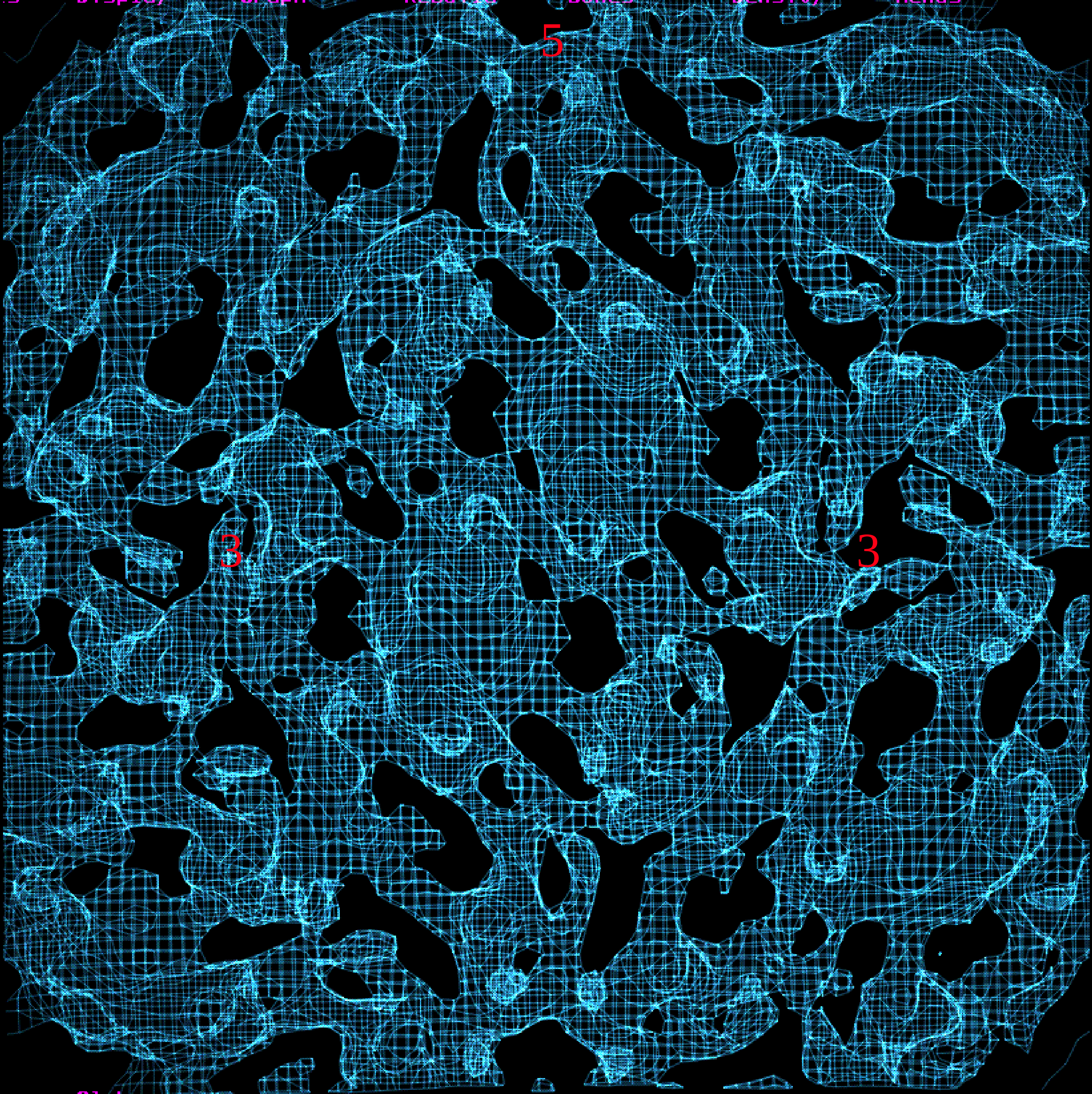
The E glycoprotein dimer of flaviviruses : Sequential fitting into the mature dengue EM map

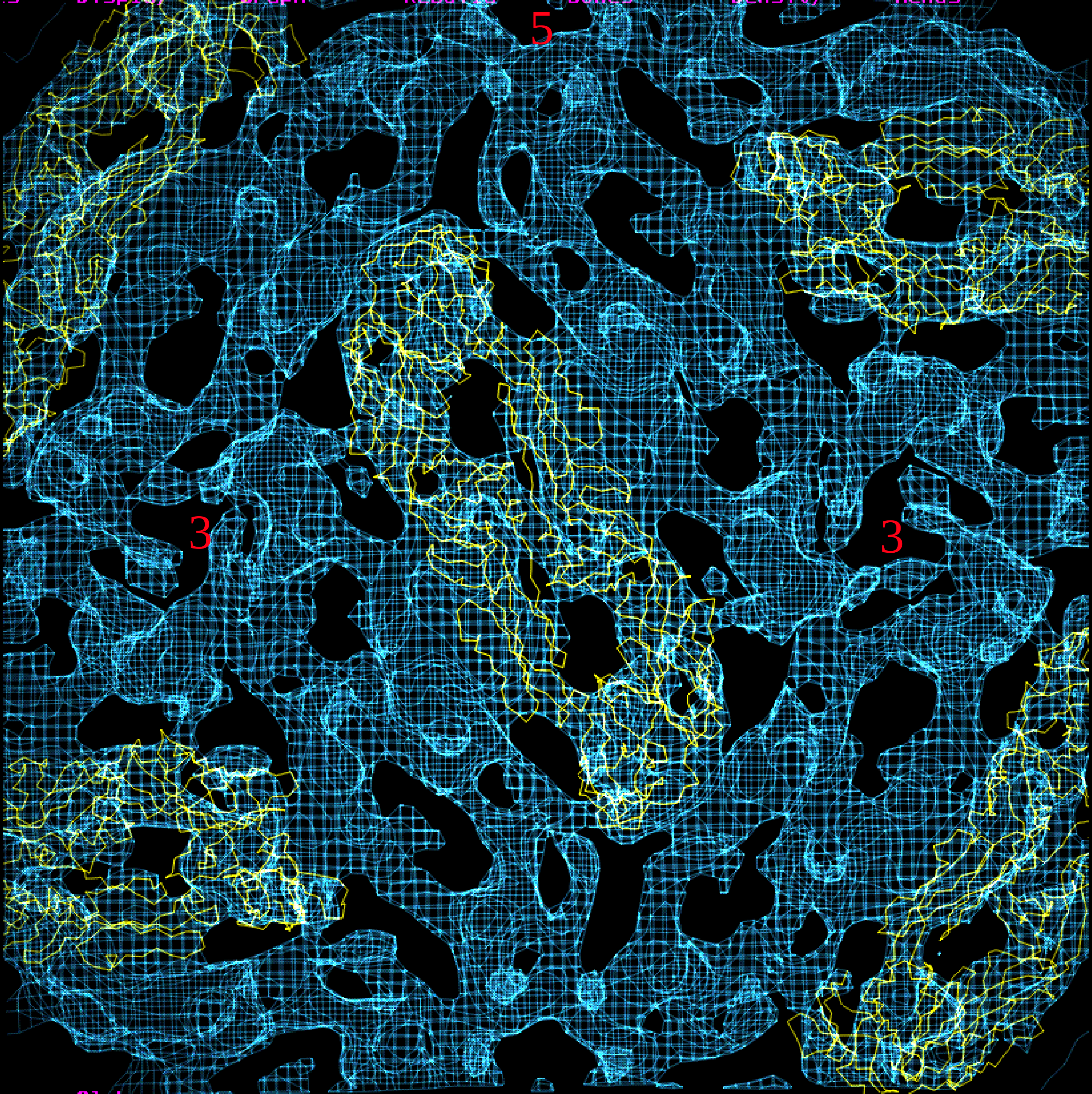


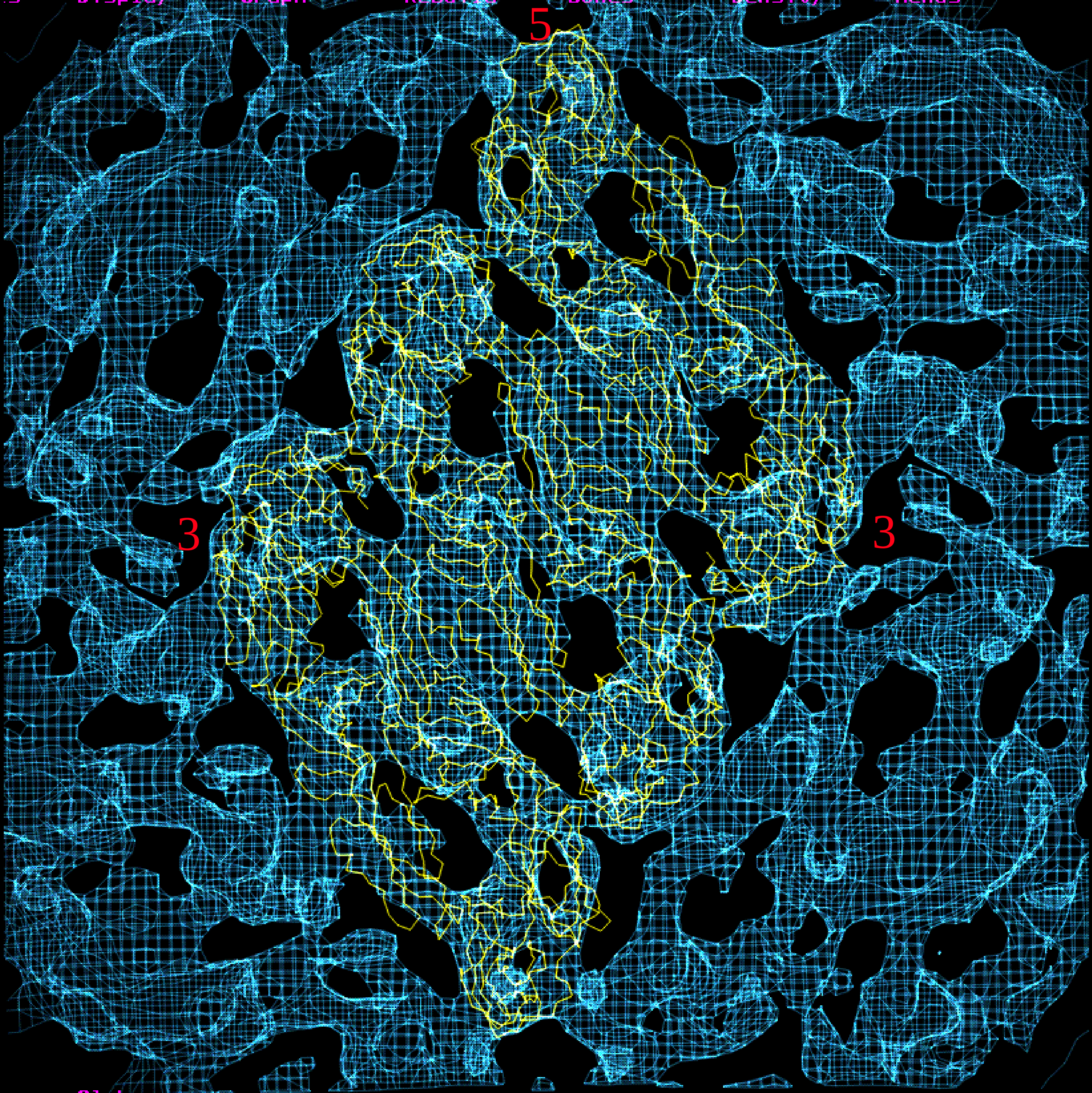
TBEV: F. Rey et al Nature, 1995, 375, 291-298

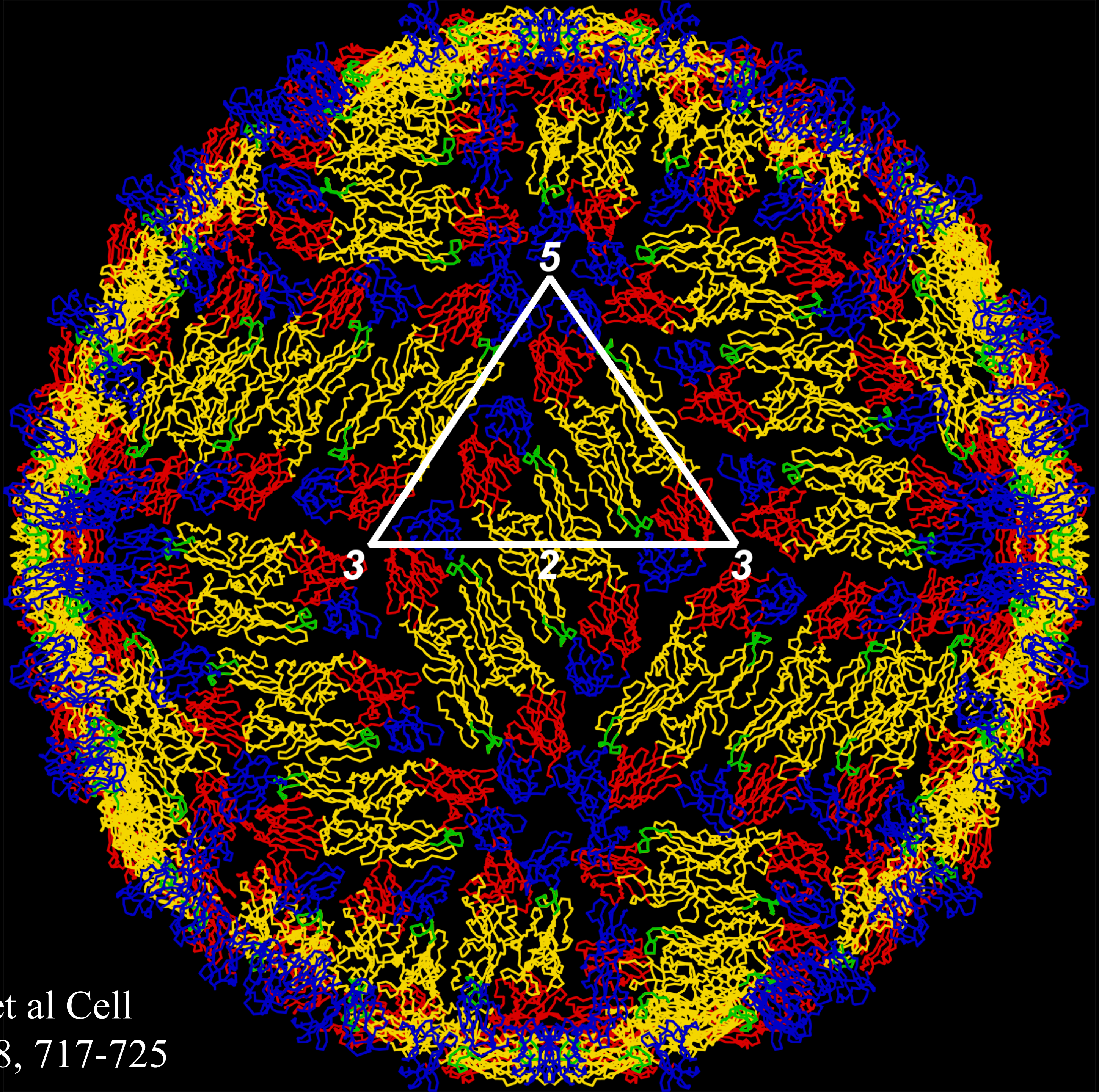
Dengue: Y. Modis et al PNAS, 2003, 100, 6986-6991

Y. Zhang et al, Structure 2004, 22, 2604-2613

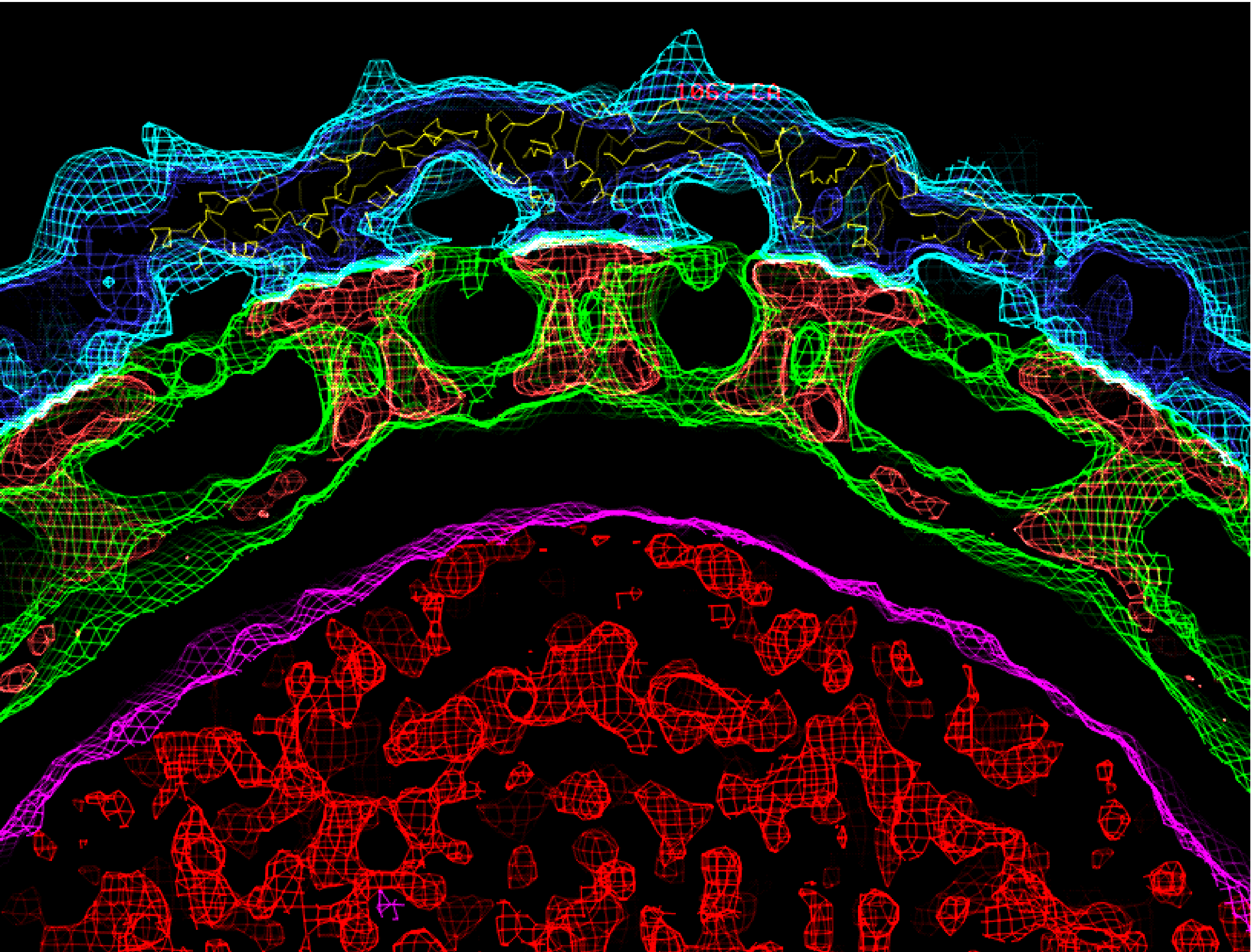

5
3
3


5
3
3

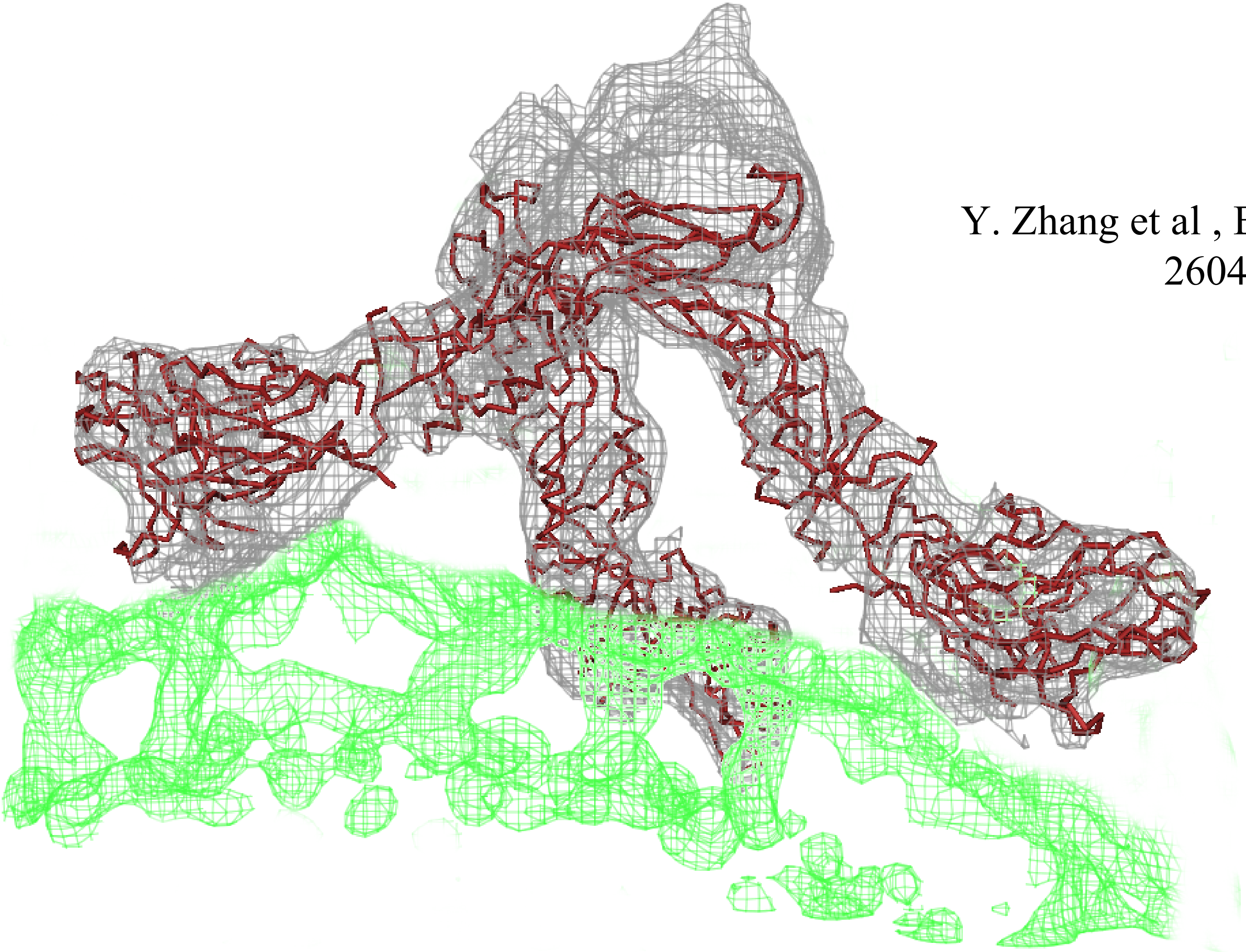

5
3
3



Kuhn et al Cell
2002, 108, 717-725


1067 CA

The E glcoprotein monomer of flaviviruses : Sequential fitting into the immature dengue virus map



Y. Zhang et al , EMBO J 2003, 22, 2604-2613

Sequential fitting of E monomer into the immature Dengue cryoEM map

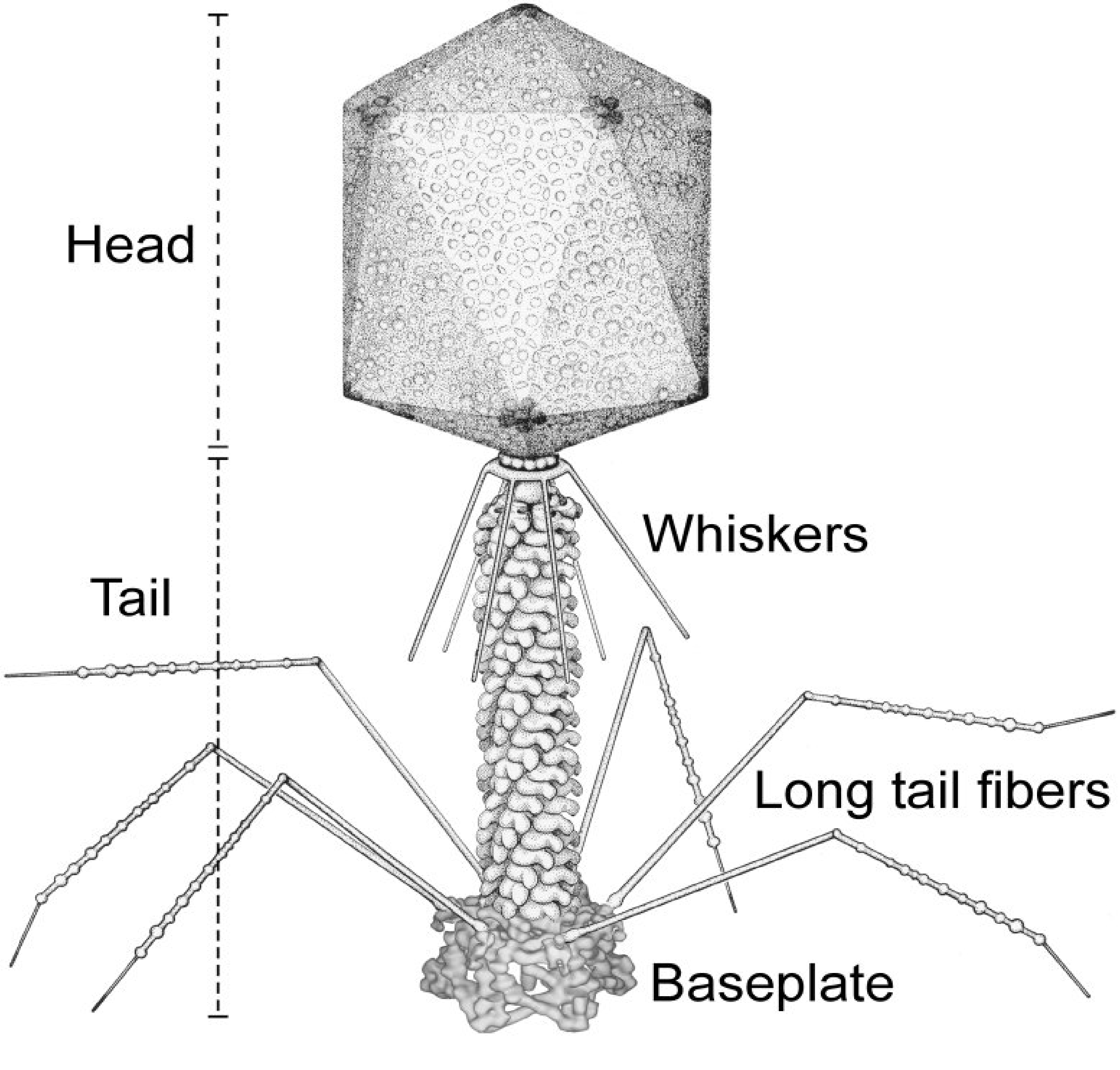
Results are independent of order of fitting

	sumf	*sumf*	*sumf*						
MOL	DI	DII	DIII	x	y	z	θ1	θ2	θ3
A 1st	50.8	55.8	42.3	32.0	-7.7	220.9	15.0	61.0	349.2
A 2nd	49.7	56.4	44.0	31.0	-6.7	221.4	11.0	61.5	345.0
A 3rd	50.9	56.0	40.5	31.5	-6.7	220.4	10.8	61.5	355.2
B 1st	48.4	57.6	42.9	72.1	8.2	210.6	38.0	64.5	162.5
B 2nd	49.8	57.4	41.8	71.6	8.2	210.6	34.8	63.5	164.8
B 3rd	49.7	57.4	41.9	72.1	7.7	210.6	37.5	64.0	163.5
C 1st	48.9	54.7	42.1	10.5	48.3	217.0	19.8	58.0	240.2
C 2nd	49.2	53.1	41.3	9.0	47.8	217.0	22.0	58.5	238.5
C 3rd	49.5	54.9	42.8	10.0	48.3	217.0	18.2	57.0	241,8

Validation

1. Is the hand consistent with each fitted protein?
2. Are distances between atoms in the interface reasonable?
3. Are the type of residues in the contact region appropriate? Look for:
   hydrophobic versus hydrophobic
   charge complimentarity
4. Have all the higher density regions been interpreted?
5. Do unexpected results make chemical sense?

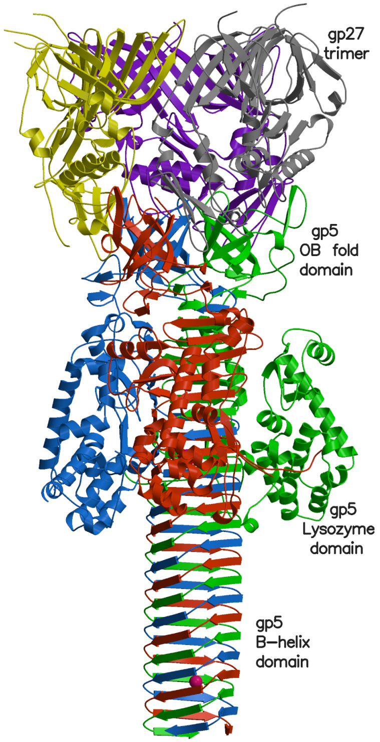
Validation: Consistent hand verification of the cryoEM map using T4 phage baseplate proteins



Hexagonal conformation (tube-baseplates)

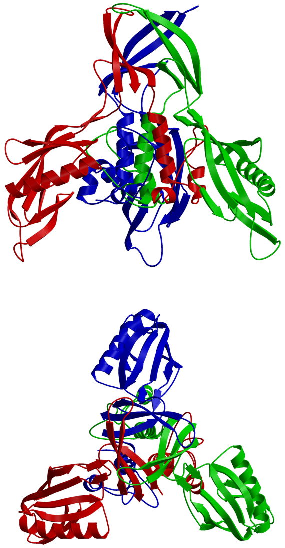
- Initial model – hexagonal prism connected to a tube
- Sixfold symmetry
- 945 particles used in the reconstruction
- Defoci 1.5 – 3.5 µm
- 12 Å resolution

Some crystal structures of the baseplate proteins

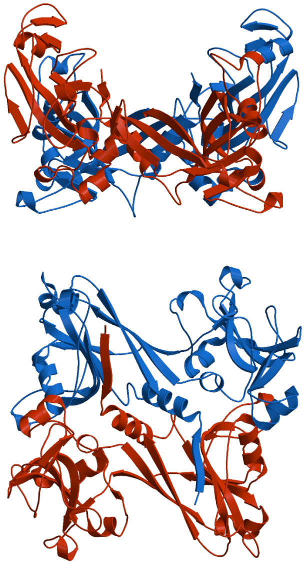


gp5-gp27

gp9,
LTF
attachment



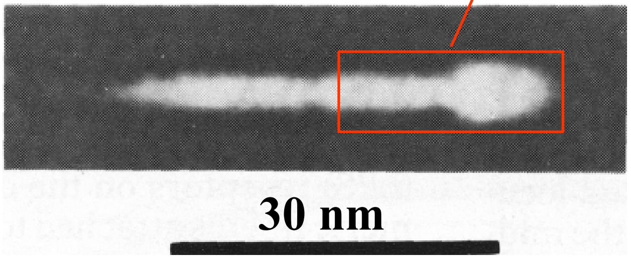
gp11,
STF attachment



gp8

gp12
C-term

gp12, STF

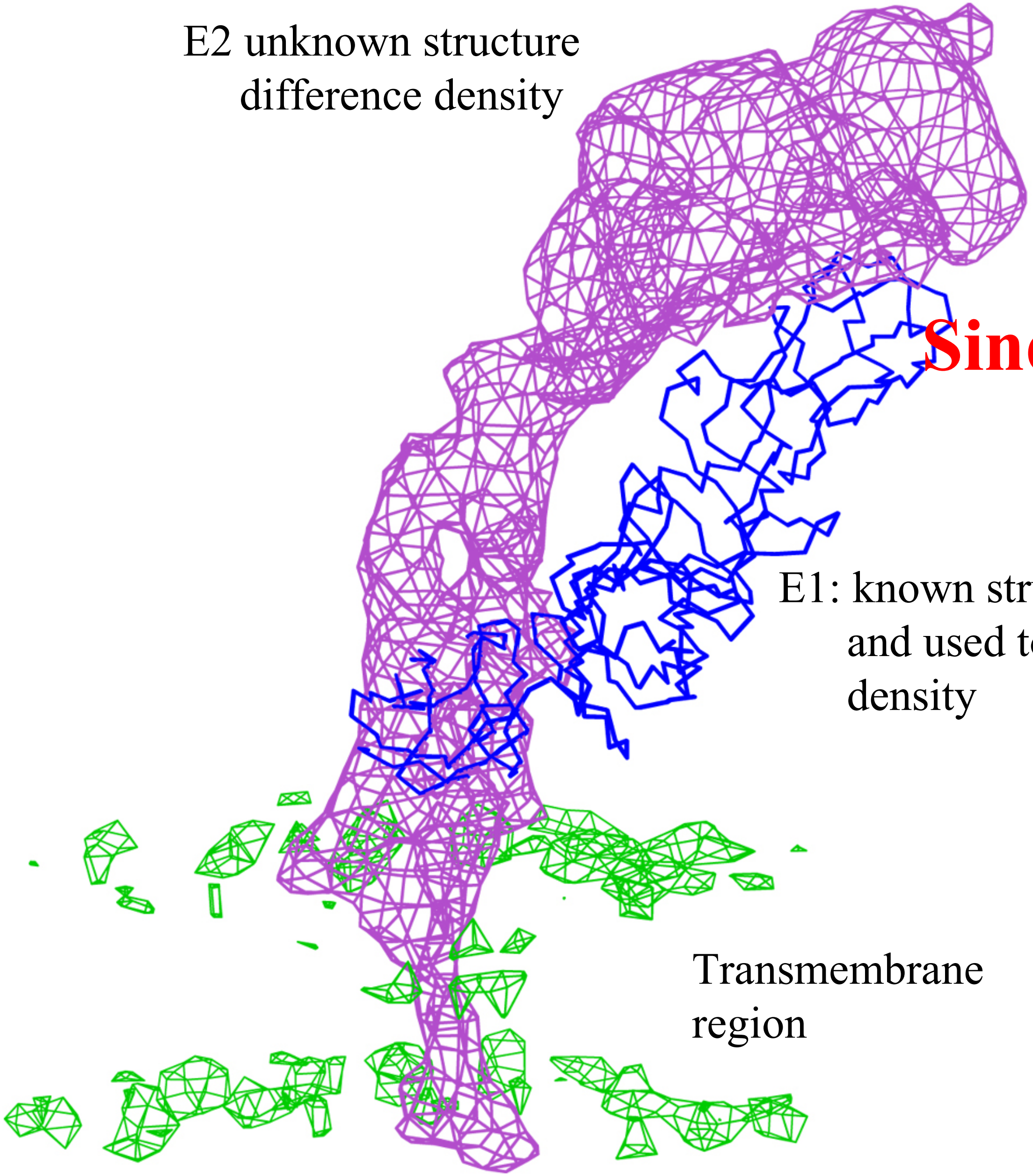


Kostyuchenko et al, Nat. Struct. Biol. 2003, **10**:688-693

T4 Hand determination: Un-normalized correlation coefficients

Baseplate protein	Correct hand	Incorrect hand
gp8	1.1	0.7
gp11	0.9	0.7
gp10	1.2	0.7
gp12	1.1	1.0

Validation : Interactions of the E1 and E2 Sindbis virus glycoproteins



E2 unknown structure difference density

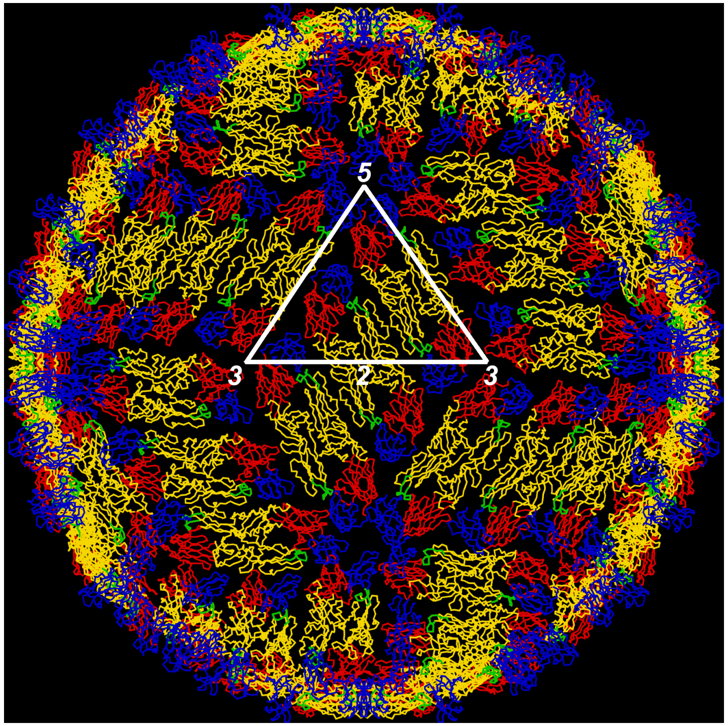
E1: known structure was fitted and used to zero out the EM density

Transmembrane region

Validation: Has all of the significant density been interpreted?

Original analysis of Dengue Virus Map at 26Å resolution

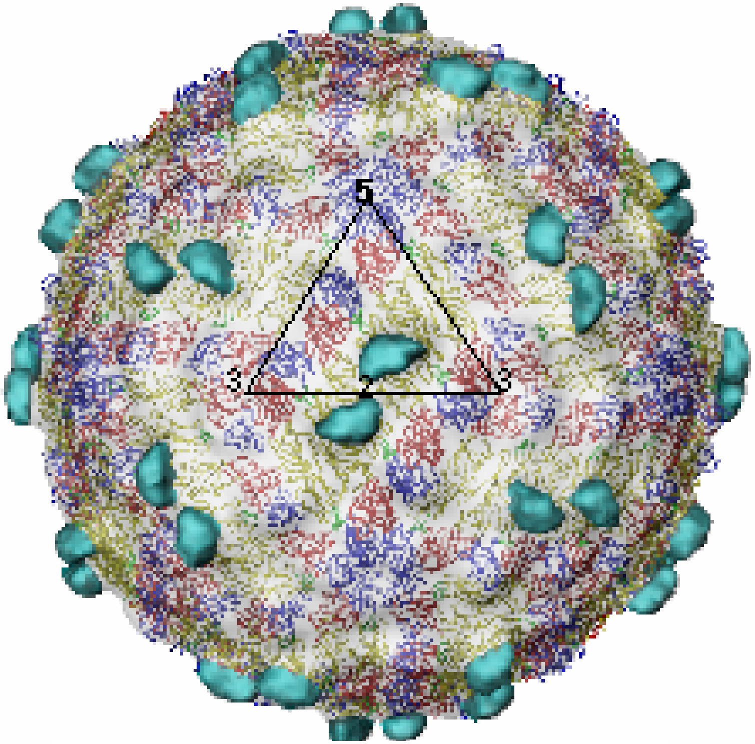
height	ratio1	ratio2
-7		147.0
-6		59.6
-5	129.8	21.7
-4	38.9	12.5
-3	17.9	4.9
-2	11.6	3.5
-1	6.6	1.9
0	5.1	2.7
1	3.4	0.8
2	2.7	0.5
3	2.4	0.3
4	1.8	0.2
5	2.0	0.1
6	1.8	0.1
7	1.9	0.0
8	0.9	0.0
9	2.0	0.0
10	2.3	0.0

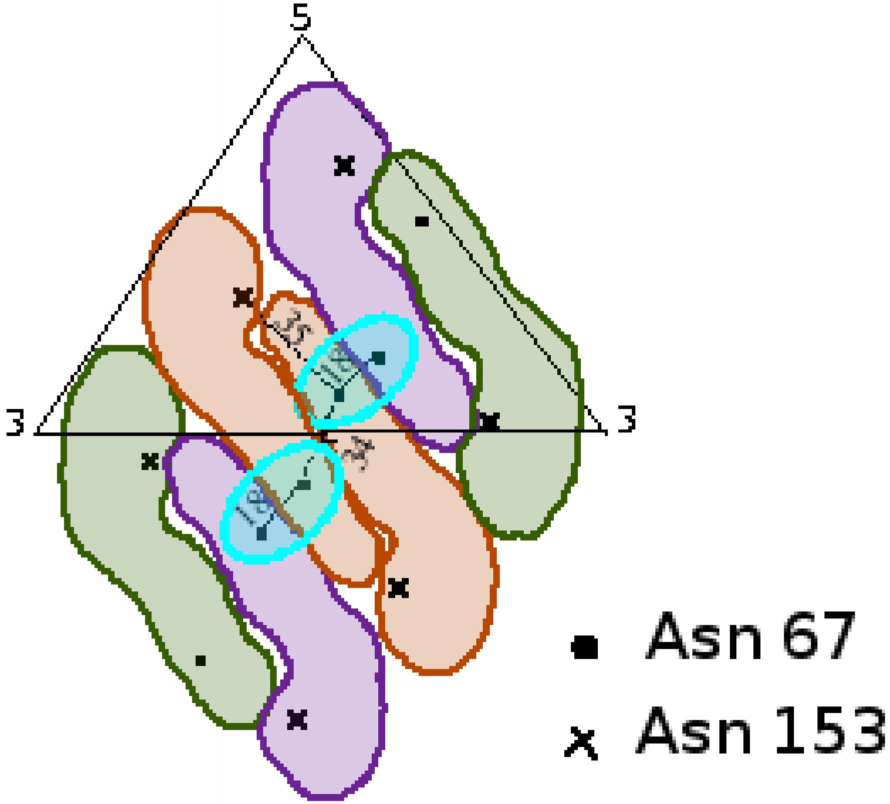


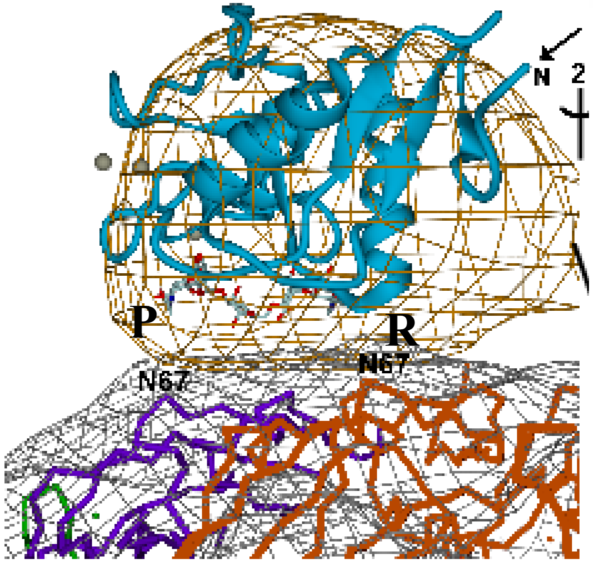
Ratio=unused/used pixels (between radii 230 & 250Å)

Ratio1: after fitting dimer on i2

Ratio2 : after fitting dimer on i2 and q2



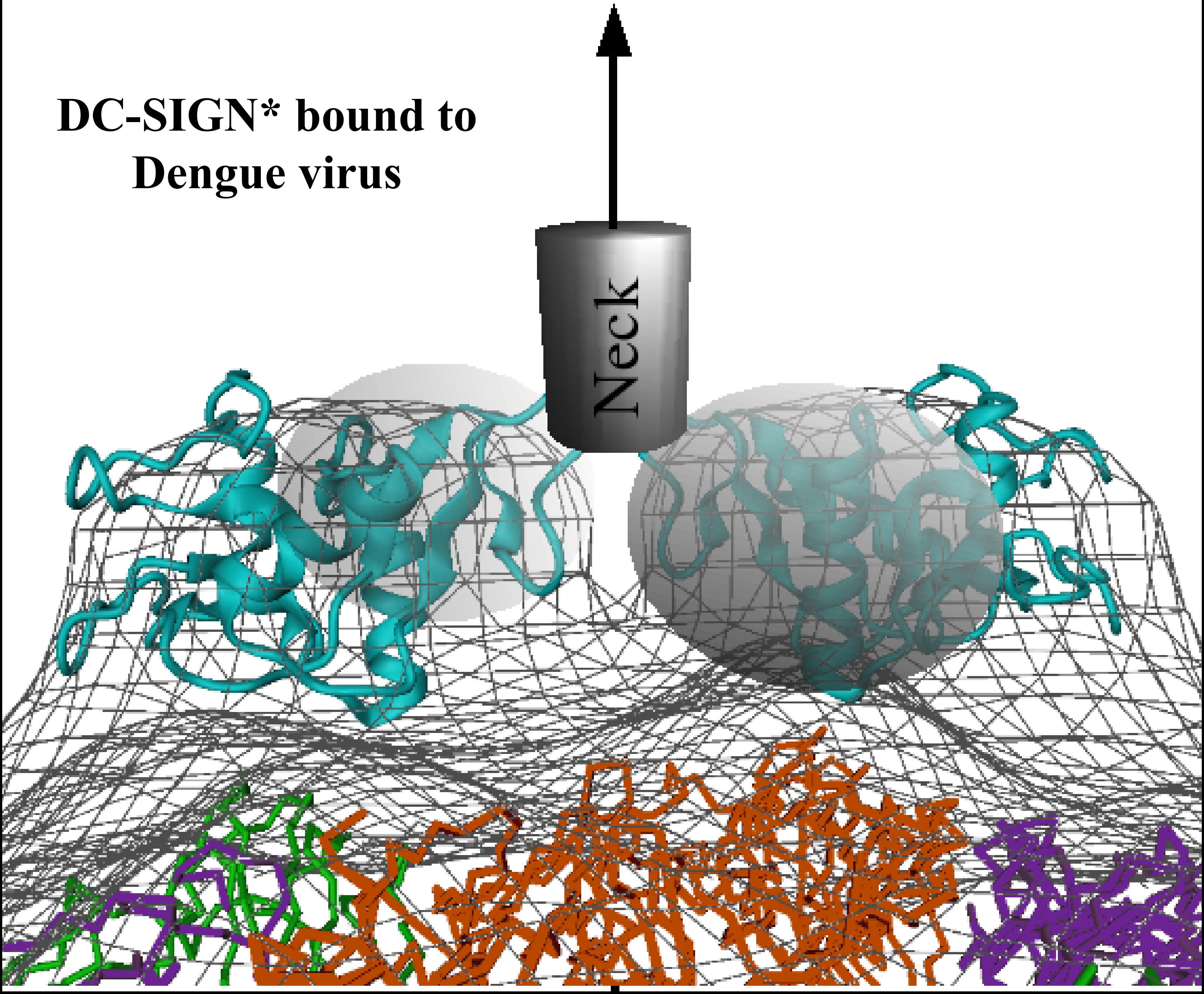




Validation: Chemical Reasonableness

Receptor recognition by Dengue virus

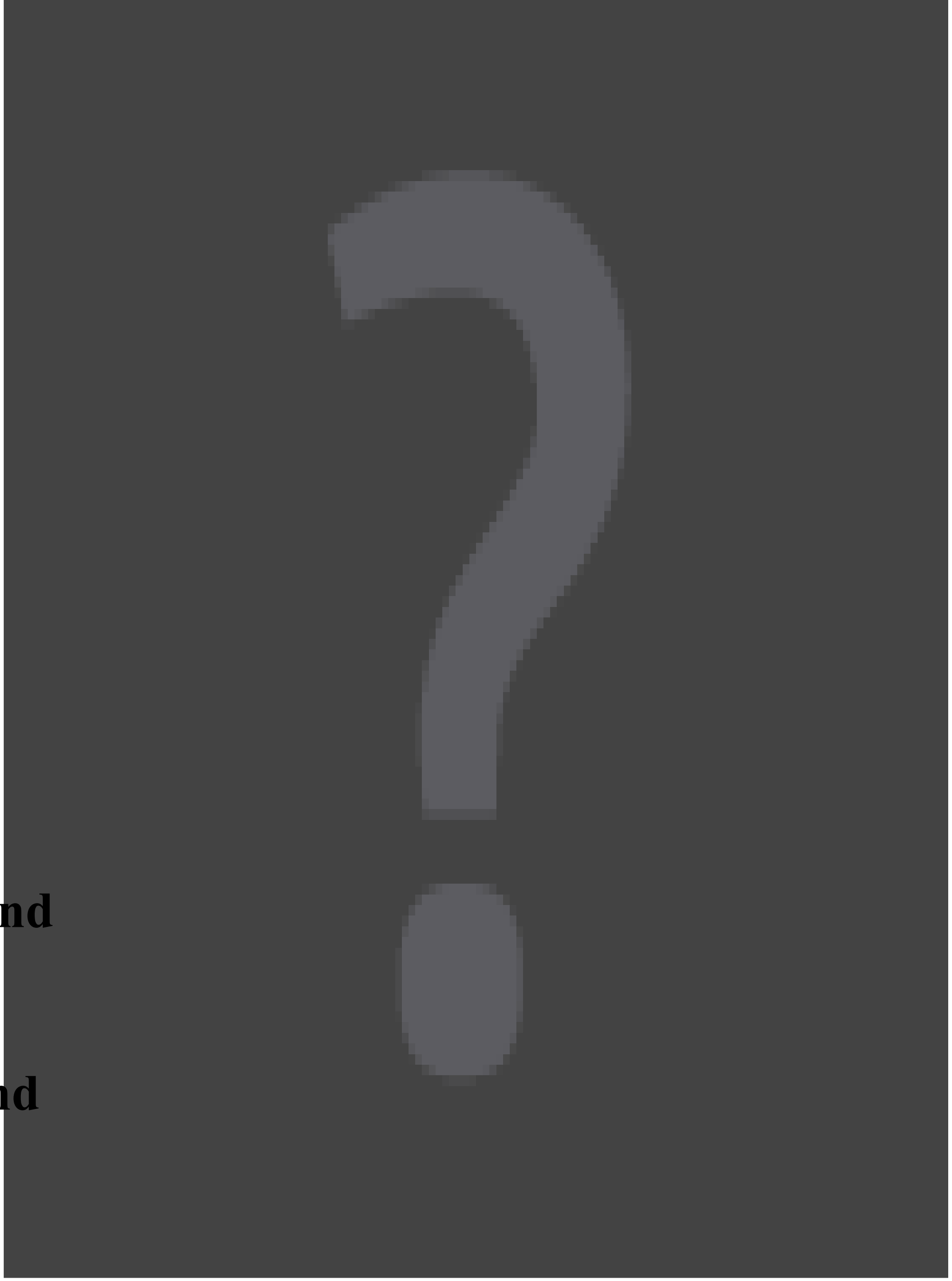
DC-SIGN* bound to Dengue virus



*Dendritic Cell Specific ICAM3 Grabbing Non-integrin;

Pokidysheva et al, Cell, submitted

Other problems:
1. Symmetry missmatches
2. Envelope of proteins whose structure is unknown



1. T4 phage 5-fold head symmetry, 6-fold tail symmetry

2. Yellow are the HOC molecules found by using a HOC$^-$ mutant

3. White are the SOC molecules found by using a HOC$^-$ SOC$^-$ mutant

Fokine et al, PNAS, 2004, **101**:6003-6008\

Relevant references

Gao et al, Structure, 2005, **13**, 401-406.

Hansen et al. Biophysics J., 2005, **88**, 818-827.

Navaza et al, Acta Cryst 2002, **D58**, 1820-1825.

Roseman et al, Acat Cryst 2000, **D56**, 1332-1340.

Rossmann et al, J. Struct Biol. 2001, **136**, 190-200.

Volkmann et al, J. Sruct Biol. 1999, **125**, 176-184.

Wriggers et al., Structure 2001, **9**, 779-788,

Wriggers et al, J. Struct Biol 1999, 125, 185-189.

Acknowledgements

T4

Petr Leiman, Victor Kostyuchenko, Paul Chipman, Shuji Kanamaru,
Mark van Raaij, Andrei Fukin, Fumio Arisaka, V. Rao, Vadim Mesynanzhinov,
Anthoni Battisti

Dengue Virus

Wei Zhang, Ying Zhang, Suchetana Mukhopadhyay, Elena Pokidysheva,
Glenn Gregorio, Shee-Mei Lok, Carol Bator-Kelly, Anthoni Battisti, Paul Chipman,
Tim Baker, Wayne Hendrickson, Jim Strauss, Richard Kuhn

Sindbis Virus

Wei Zhang, Suchetana Mukhopadhyay, Sergei Strelkov, Tim Baker, Richard Kuhn

Polio Virus

Yongning He, Steffen Mueller, Carol Bator-Kelly, Valorie Bowman, Paul Chipman,
Eckard Wimmer, Richard Kuhn

Program development

Chuan (River) Xiao, Ricardo Bernal