EMfit

Michael Rossmann

Department of Biological Sciences

Purdue University

mr@purdue.edu

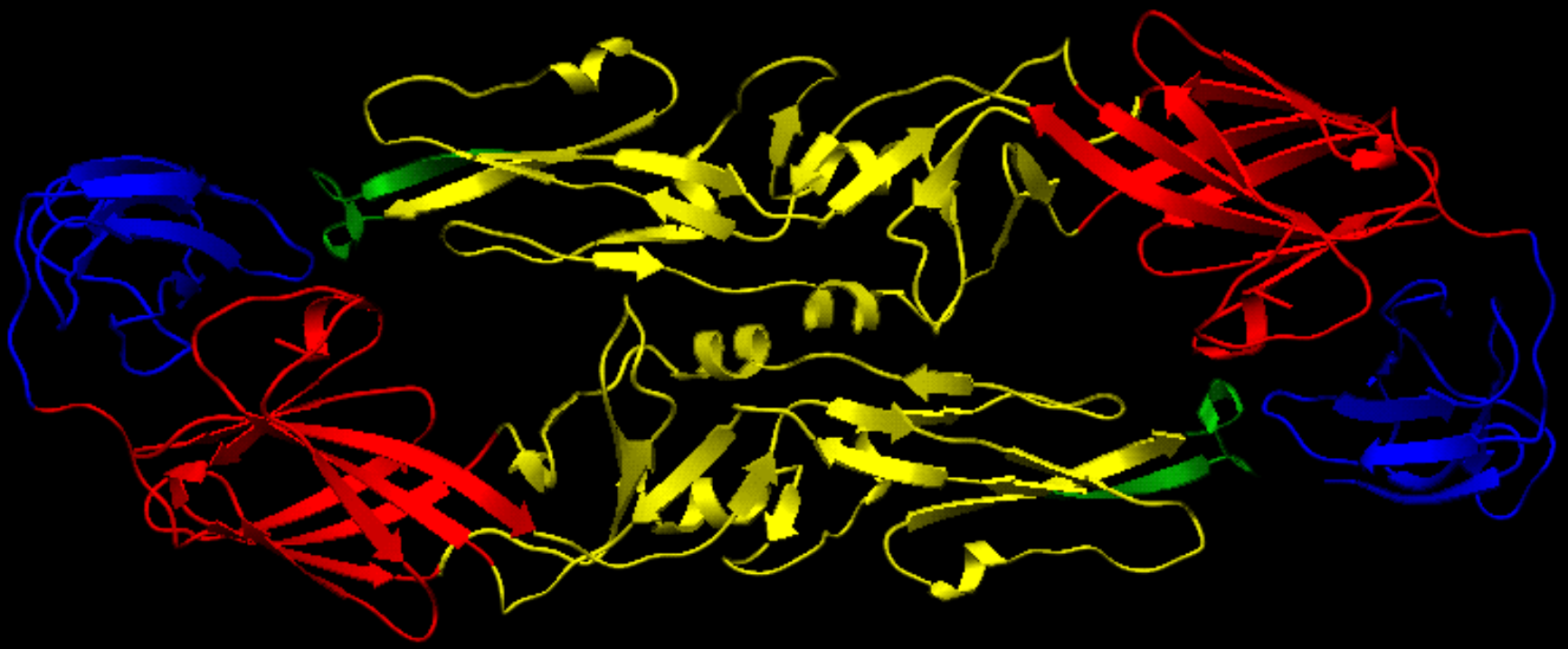
Combining CryoEM and Crystallography

1. Use cryoEM for lower resolution studies of large complexes

2. Use crystallography for higher resolution studies of component proteins

3. Use EMfit to combine these results

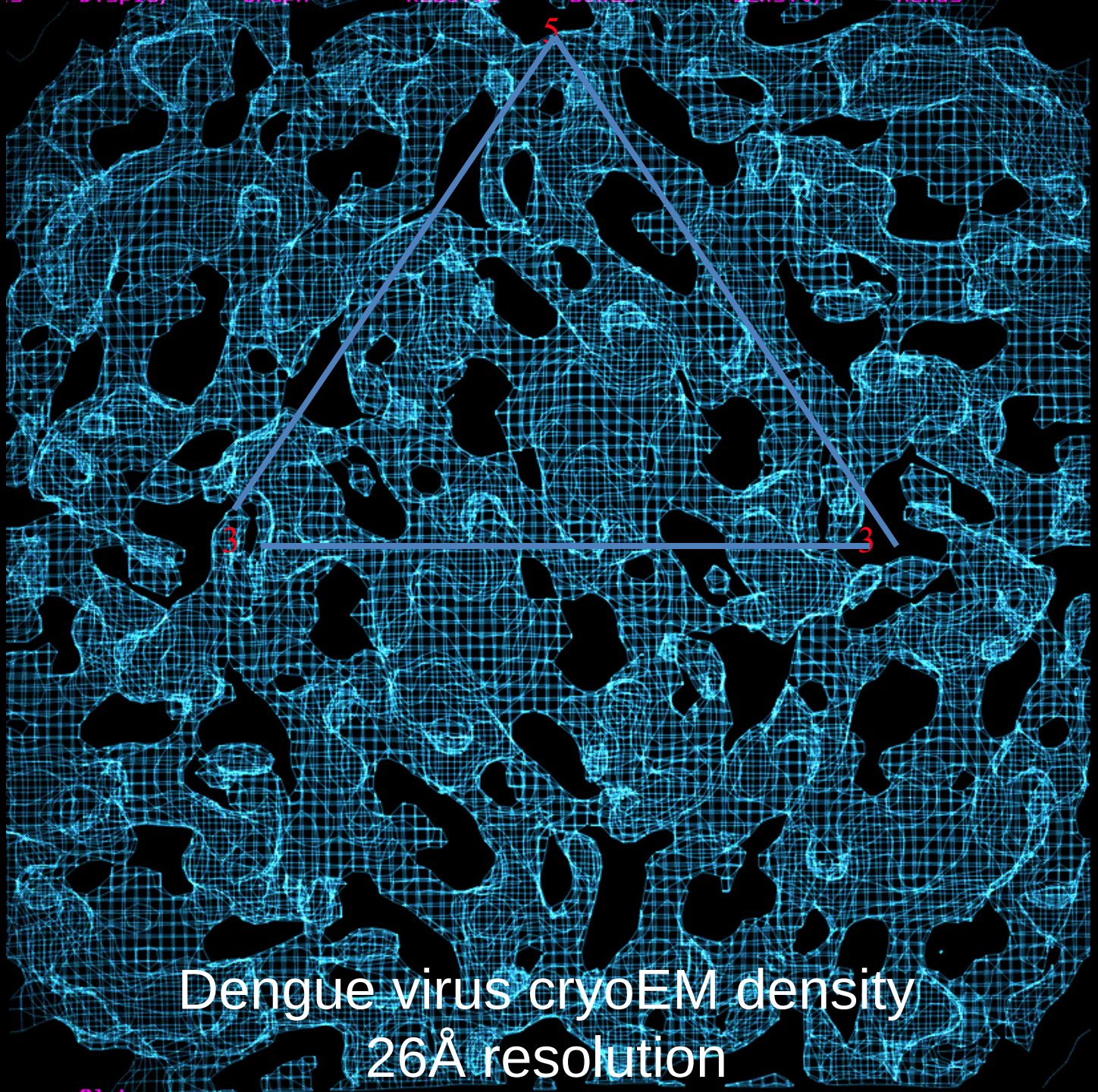
Flavivirus E glycoprotein dimer: neutral pH



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

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

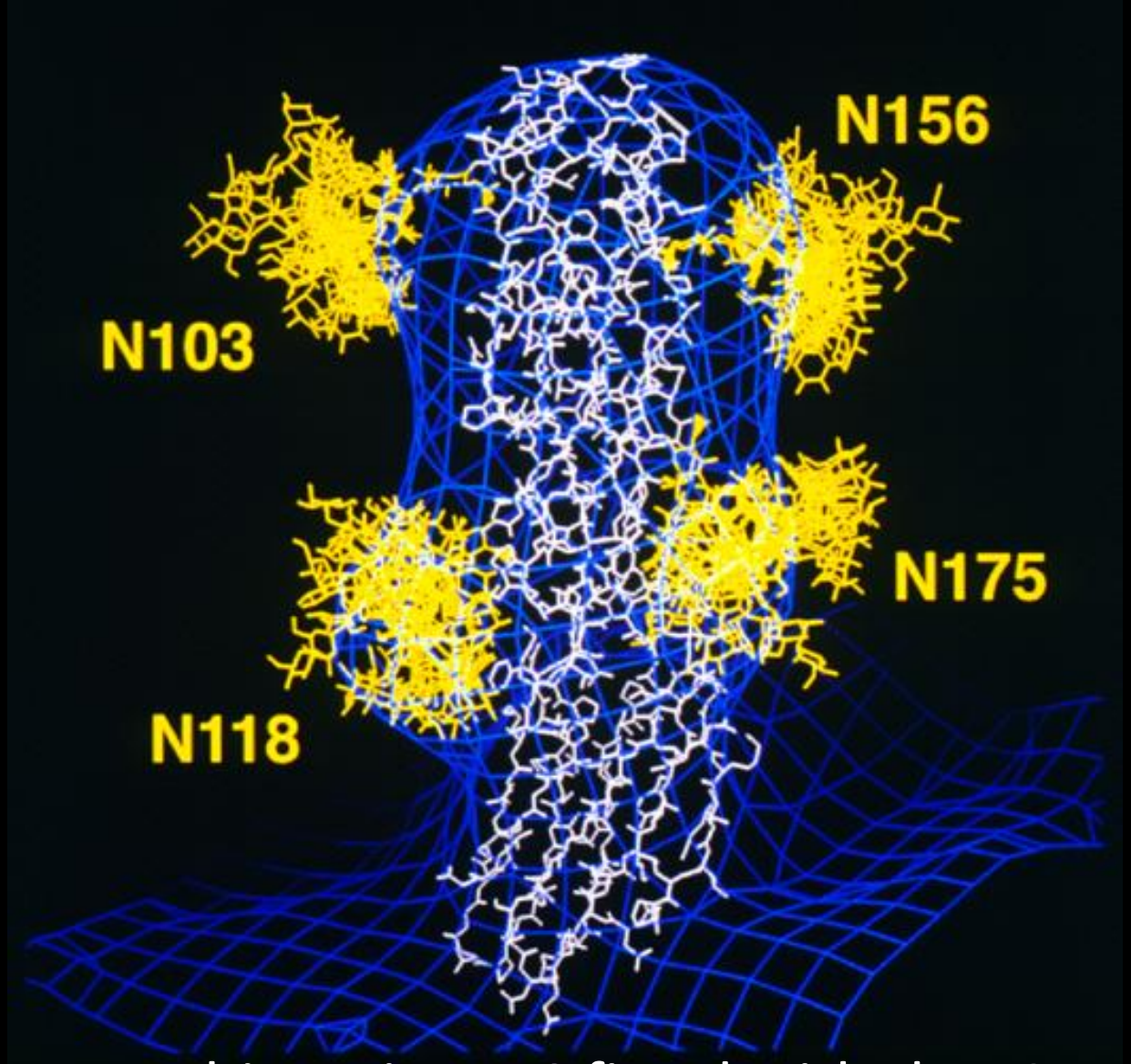
Dengue: Y. Zhang et al, Structure 2004, **12**, 1607-1608



Dengue virus cryoEM density
26Å resolution

Four ways of fitting the TBEV E dimer to the cryoEM dengue map

#	sumf	clash	-den	thet1	centx	centy	centz	D1	D2	D3
down+	53.0	0.	29.	24.5	.0	.0	232.5	48.6	56.5	49.4
down-	49.6	0.	40.	133.8	.0	.0	230.5	41.0	52.1	55.9
up+	44.8	0.	82.	119.8	.0	.0	233.5	38.1	48.9	59.9
up-	40.2	0.	116.	30.8	.0	.0	229.5	43.3	36.7	43.1



Human Rhino Virus 16 fitted with the ICAM1 receptor structure using
of the four gtlycan sites

Fitting Procedure

- 1. Determine accurate scale of EM map (distance between pixels) if possible.
- 2. Perform complete 3D angular and limited translational search . Use any chemical knowledge if available.
- 3. Refine results, by searching for the best fits, utilizing density fitting, steric overlap, symmetry clashes, and other chemical information.
- 4. Is the best fit at least 3 sigma above the next best fit?

Combining different search criteria

Crit(1)= Fit to density (sumf: average density height at atomic positions)

Crit (2)=Avoid fit to negative density (-den)

Crit(3)=Clash between symmetry related subunits (clash)

Crit(4)= Minimize distance between known sites in the map and on the model(dist)

$$rcrit = \Sigma\, w(i)\, crit(i)/sig(crit(i)) \quad i=1 \text{ to } 4$$

URL

http://bilbo.bio.purdue.edu/~viruswww/Rossmann_home//

Click on:

Software

EMfit

EMfit Manual

INTRODUCTION

Authors: Michael G. Rossmann with additions by Chuan Xiao (River) and Victor Kostyochenko

MAP INPUT, OUTPUT AND SCALING

Some keywords (**MAPIN**n, **ORIG**IN**n, MAPSCALEn,MAPPARAMn**) contain a number, **n= 1, 2 , 3** or **4** at their end (e.g. **MAPIN1**, **ORIGIN2)**. The number refers to the input map **1, 2** or **3** or output map **4**.

MAPINn filename and path [**ignore/purdue/write3d/ccp4/ascII/xplor**]
Map 1 should always be present. Several map formats are supported, and each map's format must be identified.

ORIGINn x, y, z

places the origin of the XYZ axes that define the positions of molecules in the input map **n** (**n=1,2 or 3**) in terms of pixels (**x, y, z**) counted from the top left-hand corner of the map

MAPPARAMn [APIX apix] [MIRROR, RMIRROR] [TURN jturn] [MOVE mvorig]

apix sets the pixel size in Å for map n.

MIRROR means the hand of the map is switched after reading the map

REMIRROR means the hand of the map is switched after reading and back again before writing out the resultant map.

jturn=0 do nothing (this is the default)

jturn=1 turn by 90 degrees (more…)

ATOM FITTING OPERATIONS

FIT icont [SEARCH isearch] [**CLIMB** iclimb] [**TURN** iturn] [**TOP** ntop] [**TYPE** itype]
icont .le. 0 Don't fit anything. This is used when EMfit is producing scaled maps or difference maps, defined by **MAPTYPE**. Or it is used to just look at the map . Then stop before attempting any fitting.
icont=1 use Calpha atoms only
icont=2 use all atoms. more

CENTER cx cy cz [**STEP** dx dy dz] [**LIMIT** fxyz(1) fxyz(2) fxyz(3)]

cx,cy,cz is the site in map1 where the model's center of mass is to be placed, relative to the origin, given in A unless crystcell=1. In the latter case these coordinates are given in fractional coordinates as is common for crystallographic calculations.

dx,dy,dz are the increments in the center position (in Å) to be used in climb. more

RDIFF rdiff

The number of CA atoms in symmetry related subunits that are less than **rdiff** Å from a CA atom in the reference molecule determines NCLASH. This is one of the criteria used to determine the fitting parameter RCRIT.

SYMMETRY

NCSFOLD kappa psi phi nopt1 nopt2 skew1 skew2

Can be repeated as many times as needed. The sequence of theses operators is critical for obtaining a complete, non repetitive closed point group. For instance a molecule with 222 point symmetry would require two lines representing two orthogonal 2-fold axes, a virus with icosahedral symmetry would require four lines more

NCSPDB n1 n2 ... ni (can be several lines)
The NCS operators defined by **NCSFOLD** produce a series of matrices $[R_1]$, $[R_2,]$, $[R_3]$,… .
The input coordinates, x, are rotated into the map by means of the rotation matrix [E], usually defined by the three Eulerian angles theta1, theta2, theta3. T
hus the original coordinates will be placed into the map at
$x_1=[R_1][E]x$, $x_2=[R_2][E]x$, $x_3=[R_3][E]x$, etc.

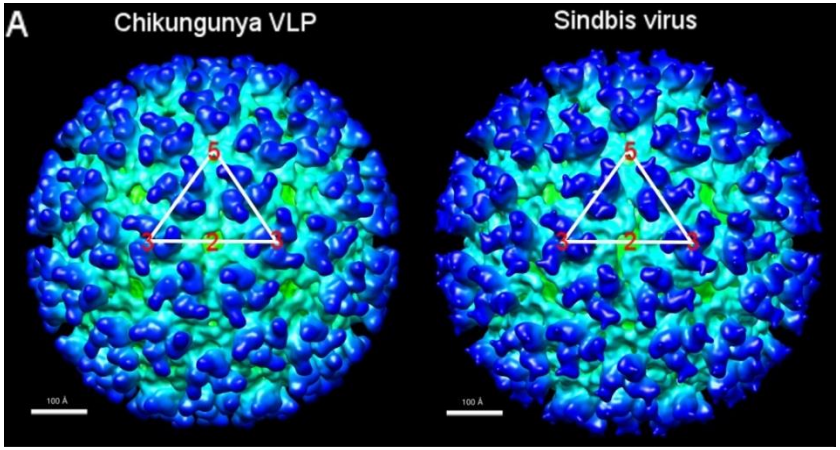
Interpretation of Chikungunya virus-like particles cryoEM maps using the EMfit program

Siyang Sun[1], Ye Xiang[1], Wataru Ahahata[2], Masaru Kanekiyo[2], Gary Nabel[2],
Michael G Rossmann[1]

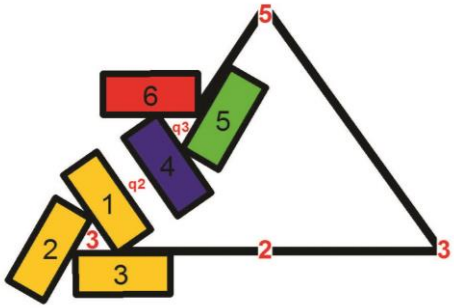
[1] Department of Biological Sciences, Purdue University, West Lafayette, IN, 47907
[2] Vaccine Research Center, National Institutes of Health, Bethesda, MD, 20817

Other participants formerly from the Rossmann lab:

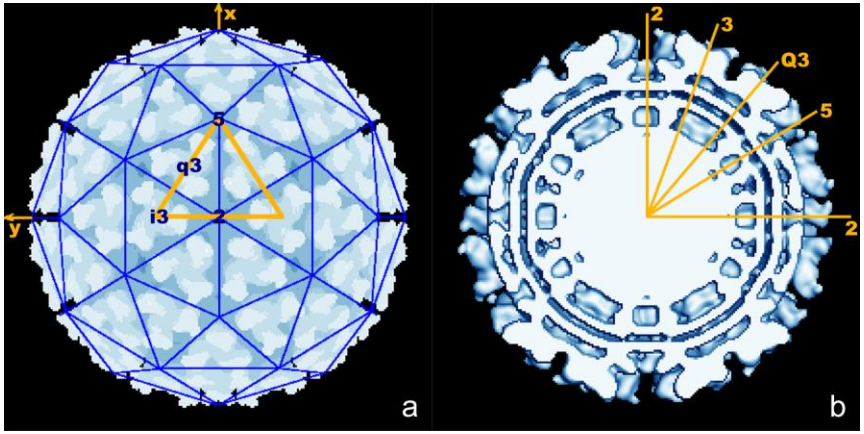
Chuan (River) Xiao, Now at University of Texas, El Paso
Victor Kostyuchenko, Now at Duke National University of Singapore



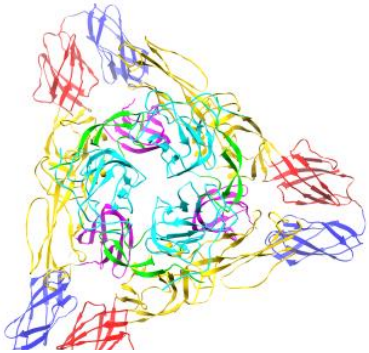
The Chikungunya virus like particles are closely similar to Sindbis virus



The icosahedral 5, 3 and 2 fold axes and the quasi q2 and q3 fold axes Generate T=4 symmetry



Sindbis virus surface view (left) and cross section (right)
A lipid membrane separates the inner nuclear capsid from the outer glycoproteins



(Top) Crystal structure of one trimeric spike on the viral surface. (Bottom) Color code for the Domains in the E1 and E2 glycoproteins

CryoEM maps:

1. 16A resolution of CHIKV VLPs with HIV gp120 outer domain insertion in E2 domain B
2. 4.4A resolution of CHIKV VLPs

Crystal structures:

1. Sindbis virus trimeric spike $(E1E2)_3$ (Rossmann et al Purdue)
2. CHIKV (E1E2) heterodimer (Felix Rey et al, Pasteur)

Fit E1 to 16Å resolution cryoEM map

Using all atoms except H Generalized rough search

Search results in descending values of sumf

id	theta1	theta2	theta3	centx	centy	centz	sumf	clash	-den	avgdst	near
1	260.00	90.00	200.00	21.9	83.3	267.0	55.11	11239.	1022.	0.00	0.
2	270.00	80.00	200.00	21.9	83.3	267.0	54.75	12053.	956.	0.00	0.
3	260.00	80.00	200.00	21.9	83.3	267.0	54.64	11392.	1110.	0.00	0.
4	260.00	100.00	200.00	21.9	83.3	267.0	54.51	11237.	1199.	0.00	0.
5	270.00	70.00	200.00	21.9	83.3	267.0	54.32	11252.	1133.	0.00	0.
6	280.00	80.00	250.00	21.9	83.3	267.0	54.02	2097.	711.	0.00	0.
7	250.00	100.00	200.00	21.9	83.3	267.0	53.80	11456.	1452.	0.00	0.
8	280.00	70.00	250.00	21.9	83.3	267.0	53.80	2934.	823.	0.00	0.
9	270.00	70.00	190.00	21.9	83.3	267.0	53.79	12353.	1445.	0.00	0.
10	250.00	90.00	200.00	21.9	83.3	267.0	53.61	11605.	1496.	0.00	0.
11	270.00	90.00	200.00	21.9	83.3	267.0	53.57	12337.	1165.	0.00	0.
12	260.00	80.00	190.00	21.9	83.3	267.0	53.56	11505.	1604.	0.00	0.
13	260.00	90.00	190.00	21.9	83.3	267.0	53.53	11273.	1455.	0.00	0.
14	270.00	80.00	190.00	21.9	83.3	267.0	53.48	12609.	1288.	0.00	0.
15	260.00	90.00	210.00	21.9	83.3	267.0	53.38	11505.	1232.	0.00	0.
16	250.00	100.00	190.00	21.9	83.3	267.0	53.36	10750.	1723.	0.00	0.
17	260.00	80.00	210.00	21.9	83.3	267.0	53.29	11476.	1236.	0.00	0.
18	260.00	70.00	200.00	21.9	83.3	267.0	53.28	11658.	1432.	0.00	0.
19	250.00	110.00	200.00	21.9	83.3	267.0	53.25	11384.	1581.	0.00	0.
20	270.00	80.00	210.00	21.9	83.3	267.0	53.15	12460.	1181.	0.00	0.
21	250.00	110.00	190.00	21.9	83.3	267.0	52.93	11269.	1837.	0.00	0.
22	280.00	60.00	190.00	21.9	83.3	267.0	52.87	12051.	1547.	0.00	0.
23	270.00	60.00	200.00	21.9	83.3	267.0	52.87	10979.	1560.	0.00	0.
24	270.00	90.00	250.00	21.9	83.3	267.0	52.81	2511.	1185.	0.00	0.
25	250.00	90.00	190.00	21.9	83.3	267.0	52.75	10712.	1977.	0.0	

Combining different search criteria

Crit(1)= Fit to density (sumf)
Crit (2)=Avoiding fit to negative density (-den)
Crit(3)=Clash between symmetry related subunits (clash)
Crit(4)= Minimizing distance between known sites in the map and on the model(dist)

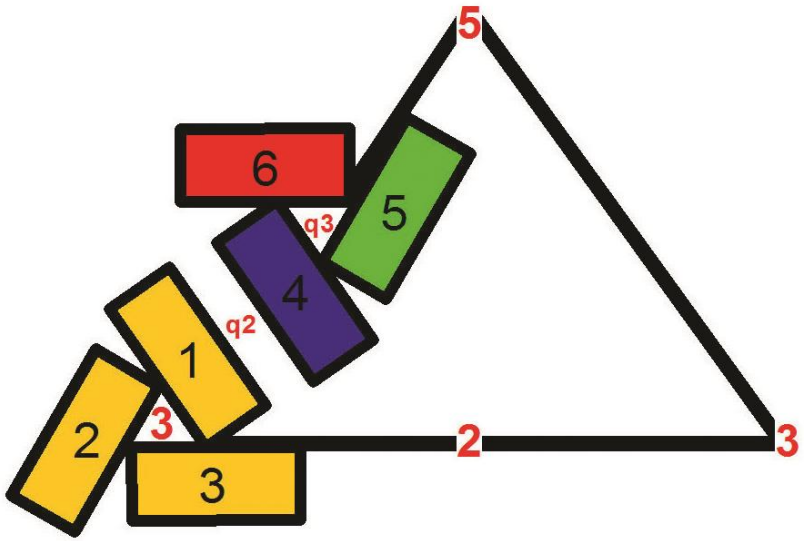
$$rcrit = \Sigma\ w(i)\ crit(i)/sig(crit(i))\quad i=1 \text{ to } 4$$

Six dimensional fine climb on top search results

```
variation in fitting restraints after general search
             sumf  clash   -den rmsdst avgdst   near
 average     53.610483.9 1334.0    0.0    0.0    0.0
 sigma        0.6 2985.3  298.0    0.0    0.0    0.0
 weights      1.0    5.0    1.0    0.0    0.0    0.0
 clash is the volume in A**3 of contact overlap per molecule in the au

% variation in fitting restraints after general search
             sumf  clash   -den rmsdst avgdst   near
 average %  53.62   2.83   0.36   0.00   0.00   0.00
 sigma   %   0.63   0.81   0.08   0.00   0.00   0.00
 weights     1.00   5.00   1.00   0.00   0.00   0.00


                                  A.        deg   deg   deg    A.    A.    A.
  #  rcrit  sumf clash  -den avgds  near thet1 thet2 thet3 centx centy centz
  6  3.280  56.6   7.9   1.0   0.0   0.0 278.0  80.0 251.0  21.4  86.8 265.5
  2  0.661  54.4  40.2   4.6   0.0   0.0 262.0  89.0 203.0  14.9  83.8 266.0
 23  0.366  53.7  41.1   5.4   0.0   0.0 270.0  75.0 200.0  16.9  83.8 265.0
 22 -0.395  52.1  43.6   7.7   0.0   0.0 279.3  61.3 191.3  14.9  84.3 267.0


 composition of rcrit in terms of # of sigmas
 rcrit is the weighted average of its components
  #   rcrit    sumf   clash    -den   rmsds   avgds    near
  6   3.280   0.671   2.083   0.526  -0.000  -0.000  -0.000
  2   0.661   0.188   0.336   0.137  -0.000  -0.000  -0.000
 23   0.366   0.020   0.291   0.055  -0.000  -0.000  -0.000
 22  -0.395  -0.353   0.155  -0.197  -0.000  -0.000  -0.000
 rcrit= sum [(x - mean x>/(rms from mean x)] over all criteria
```

Refinement of quasi symmetry

# ncs		capa	psi	phi
6	1	120.00	69.10	-90.00
6	2	180.00	74.35	278.80
6	3	72.00	90.00	-58.28
2	1	120.00	69.10	-90.00
2	2	180.00	74.55	278.65
2	3	72.00	90.00	-58.28
23	1	120.00	69.10	-90.00
23	2	180.00	74.55	278.00
23	3	72.00	90.00	-58.28
22	1	120.00	69.10	-90.00
22	2	180.00	74.50	278.10
22	3	72.00	90.00	-58.28



Fitting Results

domain	T num	sumf	num	miss
1	3	63.83	979	0
2	3	55.73	1015	0
3	3	58.70	656	0
4	3	47.44	951	0
5	3	57.06	1368	0
6	3	46.88	673	0

sumf as a function of domains

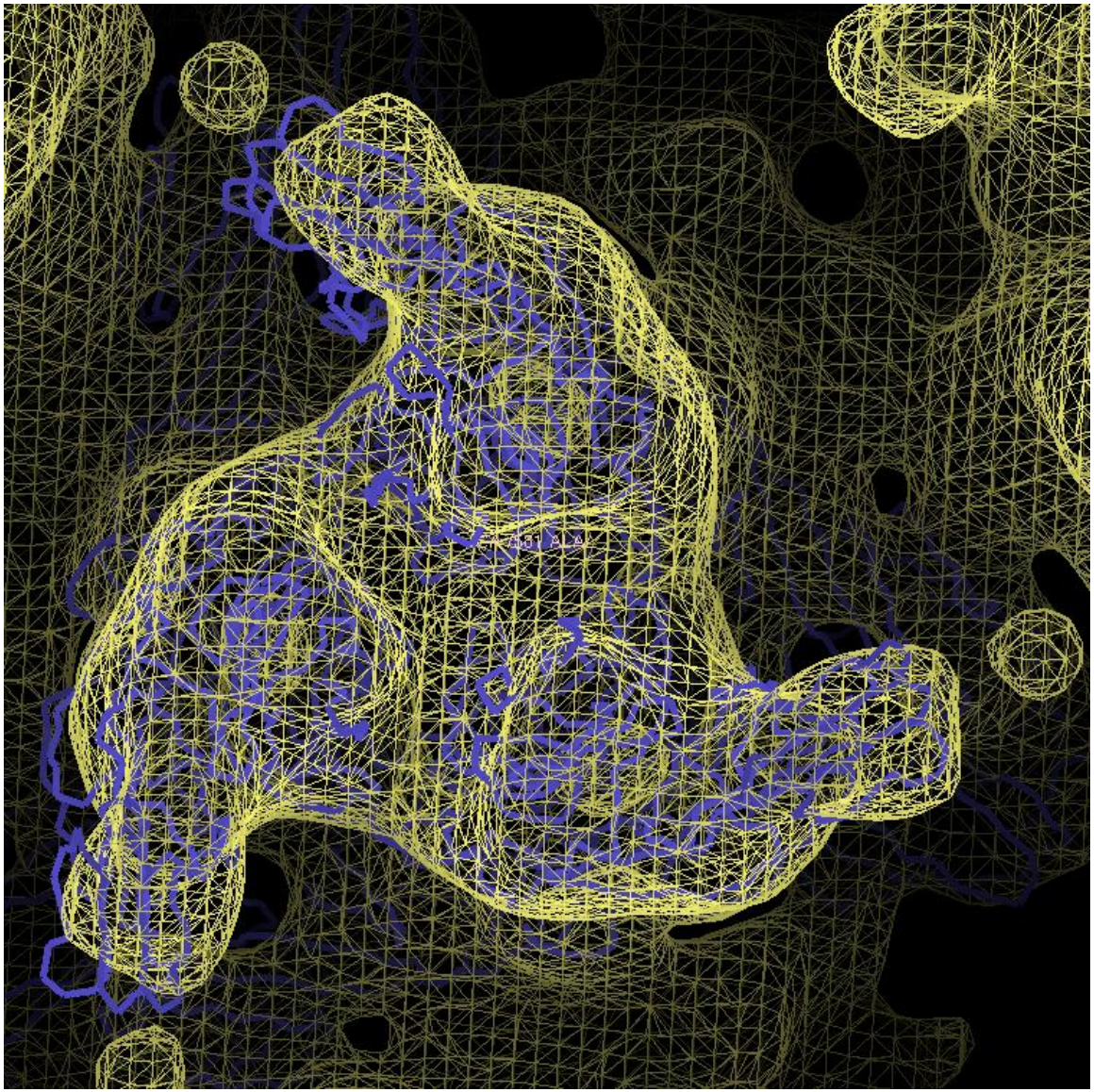
domain	T num	sumf	num	miss
1	4	56.74	979	0
2	4	57.63	1015	0
3	4	53.86	656	0
4	4	52.72	951	0
5	4	60.69	1368	0
6	4	54.73	673	0

domain	T num	sumf	num	miss
1	3	63.83	979	0
2	3	55.73	1015	0
3	3	58.70	656	0
4	3	47.44	951	0
5	3	57.06	1368	0
6	3	46.88	673	0

sumf as a function of domains

domain	T num	sumf	num	miss
1	4	56.74	979	0
2	4	57.63	1015	0
3	4	53.86	656	0
4	4	52.72	951	0
5	4	60.69	1368	0
6	4	54.73	673	0

The 16Å resolution CHIKV VLP EM map (yellow)
Top view of q3 spike fitted with EMfit determined Ca backbone (blue



Fitting of HIV gp120 outer domain into difference map

Without restraint

#	rcrit	sumf	clash	-den	A. avgds	near	deg thet1	deg thet2	deg thet3	A. centx	A. centy	A. centz
24	1.673	28.3	0.0	11.8	0.0	0.0	360.0	70.0	170.0	40.6	170.1	287.0
17	0.313	28.2	0.0	12.3	0.0	0.0	20.0	70.0	160.0	40.6	170.6	287.0
21	0.269	28.1	0.0	12.1	0.0	0.0	350.0	100.0	160.0	40.6	170.6	286.0
25	0.135	28.1	0.0	12.3	0.0	0.0	210.0	110.0	260.0	40.6	170.6	286.0
11	-0.031	28.2	0.0	12.8	0.0	0.0	30.0	50.0	160.0	40.6	170.6	286.0
14	-0.495	28.1	0.0	12.6	0.0	0.0	50.0	60.0	140.0	40.6	169.6	287.0
10	-1.443	28.1	0.0	13.4	0.0	0.0	190.0	120.0	259.0	40.6	170.6	286.5

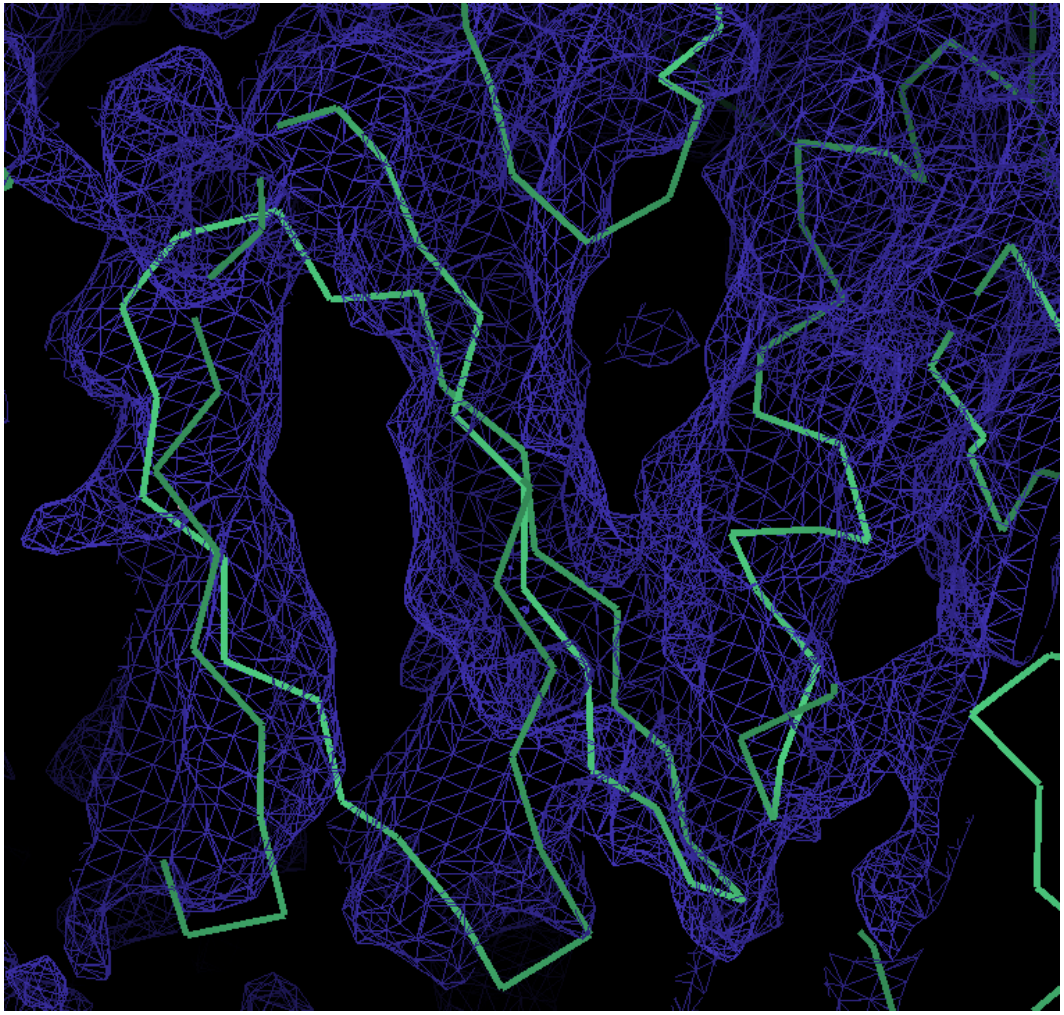
With restraints

C1 to N2 < 15Å and C2 to N3 < 15Å

N1--CHKV—C1---N2--HIV—C2---N3--CHKV—C3

#	rcrit	sumf	clash	-den	A. avgds	% near	deg thet1	deg thet2	deg thet3	A. centx	A. centy	A. centz
40	1.989	27.3	0.0	14.4	7.8	0.0	120.3	60.0	230.0	41.6	168.6	286.0
43	1.845	27.4	0.0	15.2	7.9	0.0	141.3	39.3	220.3	42.1	169.1	286.0
10	1.493	27.5	0.0	14.3	8.9	0.0	278.5	150.3	110.3	42.1	169.1	286.5
6	1.325	27.2	0.0	14.9	8.4	0.0	300.0	170.0	130.0	42.6	168.6	286.0
45	1.157	27.1	0.0	15.4	8.4	0.0	250.0	160.0	80.0	41.6	168.6	286.0

The 4.4Å resolution CHIKV VLP EM map (dark blue)
View of the Ca backbone of the EM fit determined
Domain C β-barrel (baby blue)



Fitting of E2+ E1 into the 4.4 Å resolution CHIKV VLP map

(using all atoms except H)

						A.	deg	deg	deg	A.	A.	A.
#	rcrit	sumf	clash	-den	avgds	near	thet1	thet2	thet3	centx	centy	centz
24	6.224	18.9	109.2	91.9	0.0	0.0	279.0	78.0	251.0	21.4	85.3	263.0
6	0.568	14.7	642.9	94.6	0.0	0.0	260.3	100.3	200.3	20.4	82.8	265.5
10	0.435	14.4	651.6	94.4	0.0	0.0	260.0	81.0	200.0	19.4	83.8	261.5
12	0.307	14.2	640.9	94.4	0.0	0.0	250.0	90.0	190.0	22.9	80.3	261.0
8	0.100	13.7	458.6	95.4	0.0	0.0	250.5	109.8	210.3	16.4	84.8	267.0
18	-0.371	13.2	588.7	94.8	0.0	0.0	271.3	81.3	221.3	22.9	85.3	262.0
13	-1.198	12.3	621.1	95.1	0.0	0.0	281.3	91.3	211.3	19.9	82.3	265.0

sumf as a function of domains

domain	T num	sumf	num	miss
1	1	18.17	979	0
2	1	13.85	1015	0
3	1	22.96	656	0
4	1	21.11	951	0
5	1	19.60	1368	0
6	1	18.68	673	0

sumf as a function of domains

domain	T num	sumf	num	miss
1	3	18.44	979	0
2	3	13.99	1015	0
3	3	23.20	656	0
4	3	19.77	951	0
5	3	18.78	1368	0
6	3	18.07	673	0

sumf as a function of domains

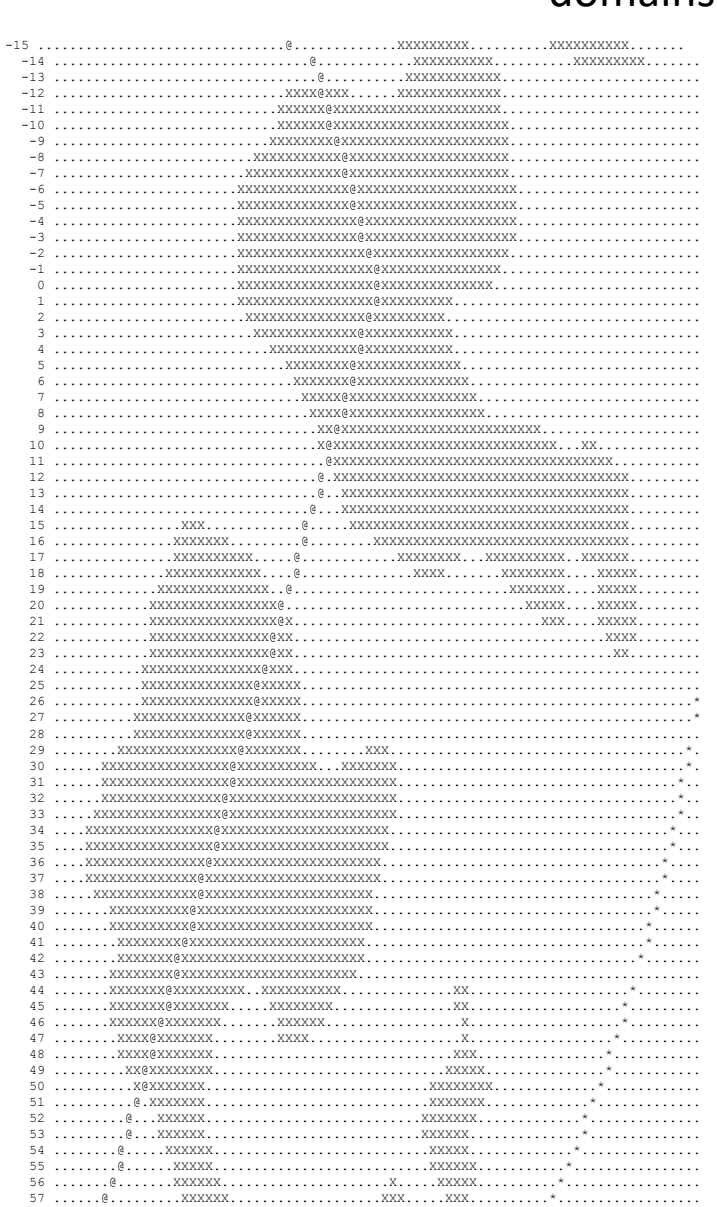
domain	T num	sumf	num	miss
1	2	16.58	979	0
2	2	14.72	1015	0
3	2	21.84	656	0
4	2	20.38	951	0
5	2	19.50	1368	0
6	2	19.13	673	0

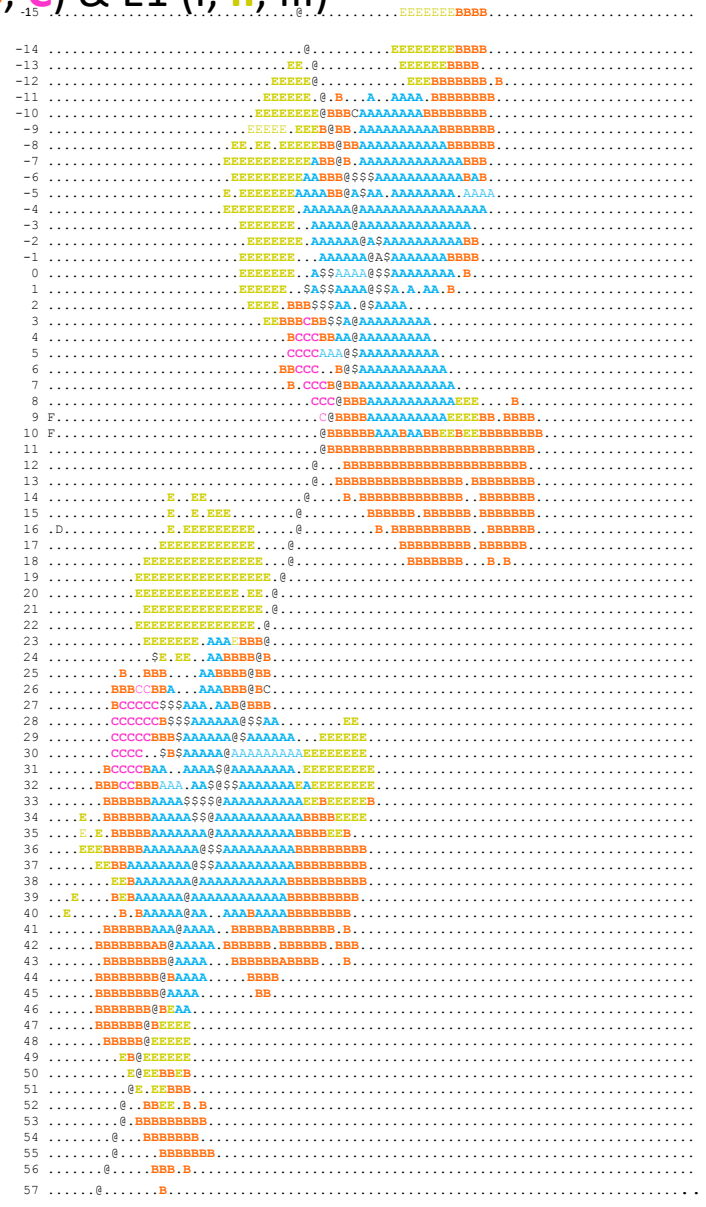
sumf as a function of domains

domain	T num	sumf	num	miss
1	4	18.49	979	0
2	4	14.41	1015	0
3	4	22.96	656	0
4	4	19.57	951	0
5	4	20.55	1368	0
6	4	19.50	673	0

Results of fitting the 16Å resolution map

Density > 4/26 of maximum height **Fitted atoms** indicated by domains E2 (A, B, C) & E1 (I, II, III)

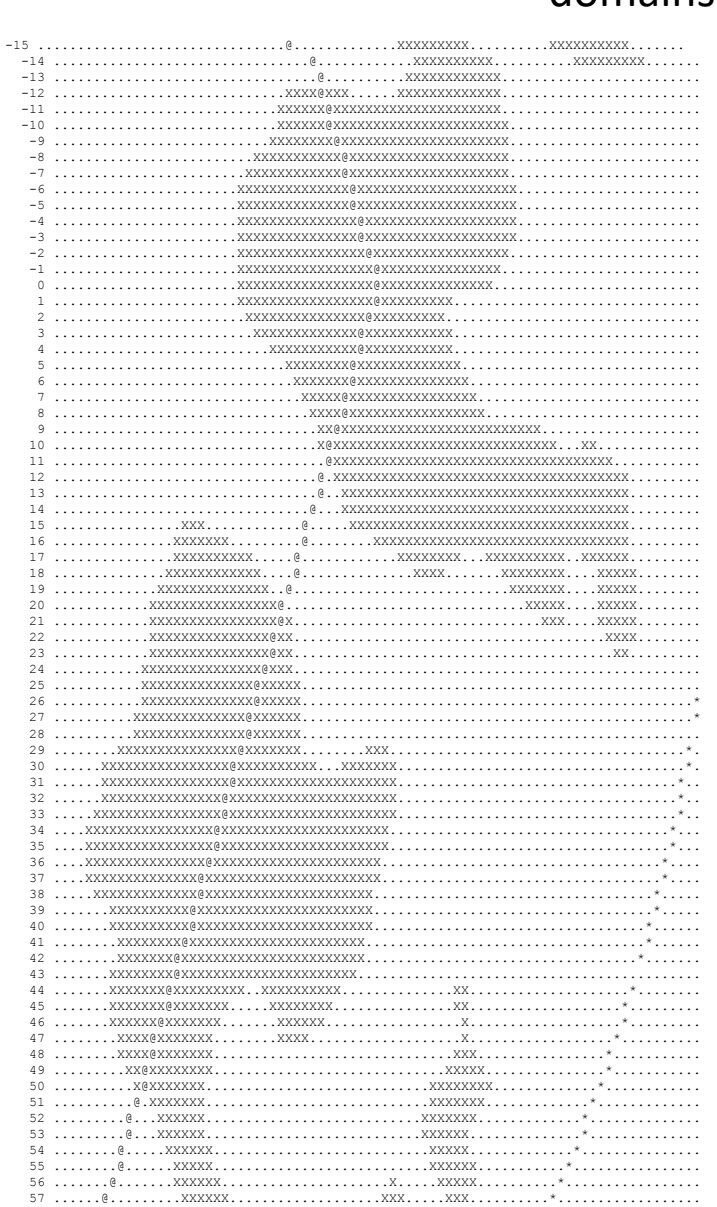


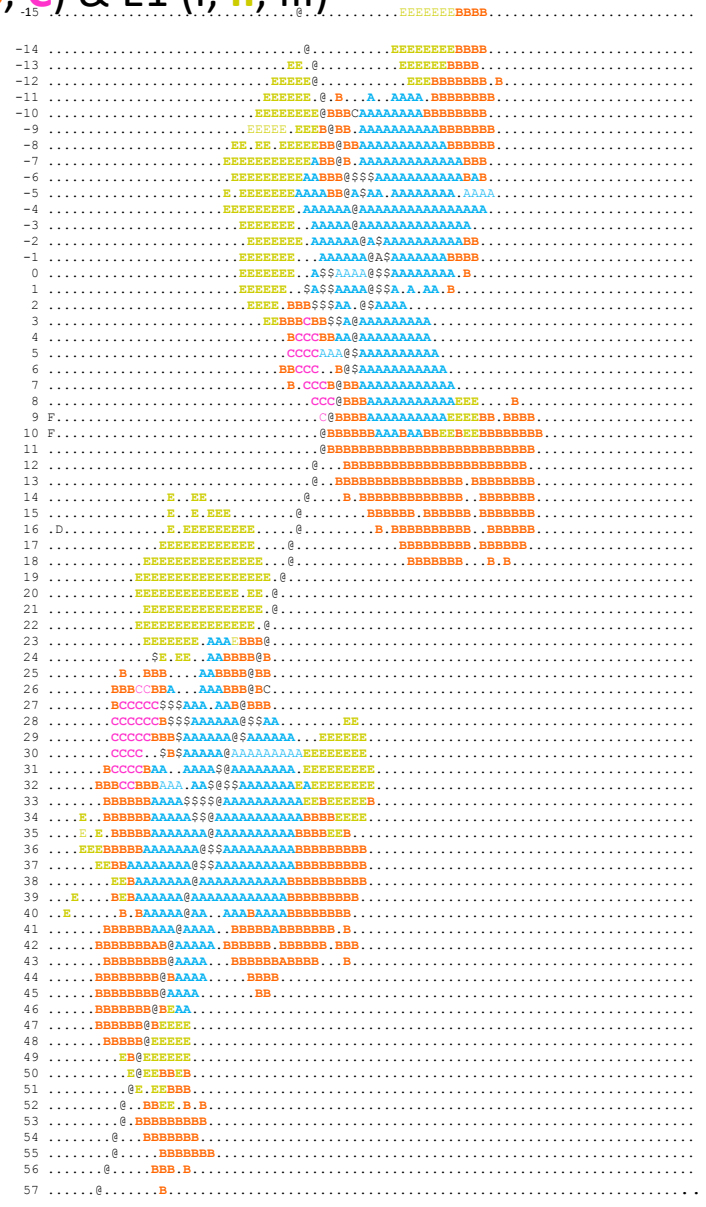


Section z=110 x 2.54Å from center

Results of fitting the 16Å resolution map

Density > 4/26 of maximum height **Fitted atoms** indicated by domains E2 (A, B, C) & E1 (I, II, III)





Section z=110 x 2.54Å from center

Other topics

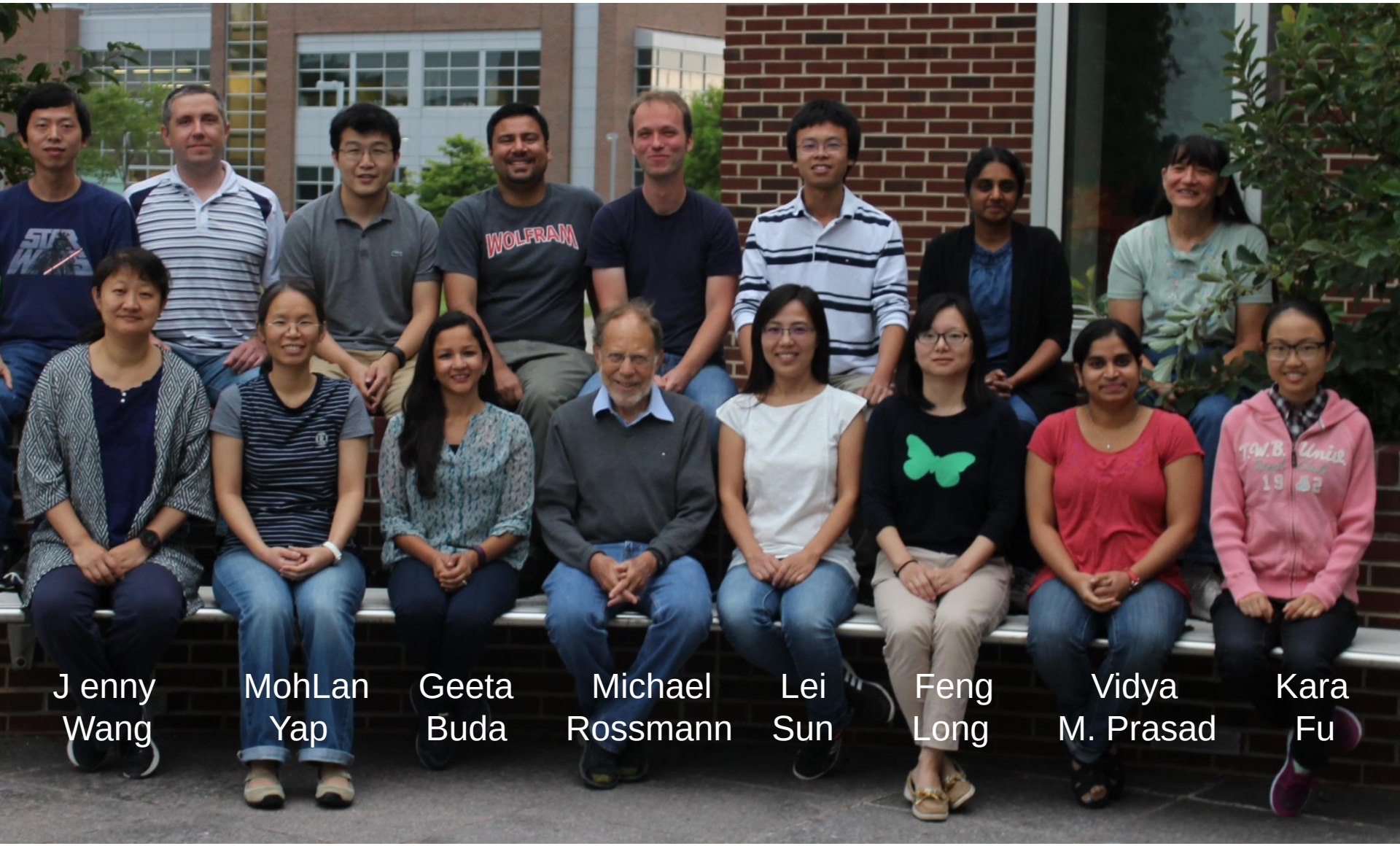
1. Sequential fitting of component parts

2. Accurate determination of EM magnification given precise knowledge of a relevant X-ray structures

3. Weighted correlation with atomic “valency”

4. Detection of groups of poorly fitting atoms that exceed the expected fitting error

5. Program Code and Manual freely available from me or my web site



Rossmann Lab September 2016