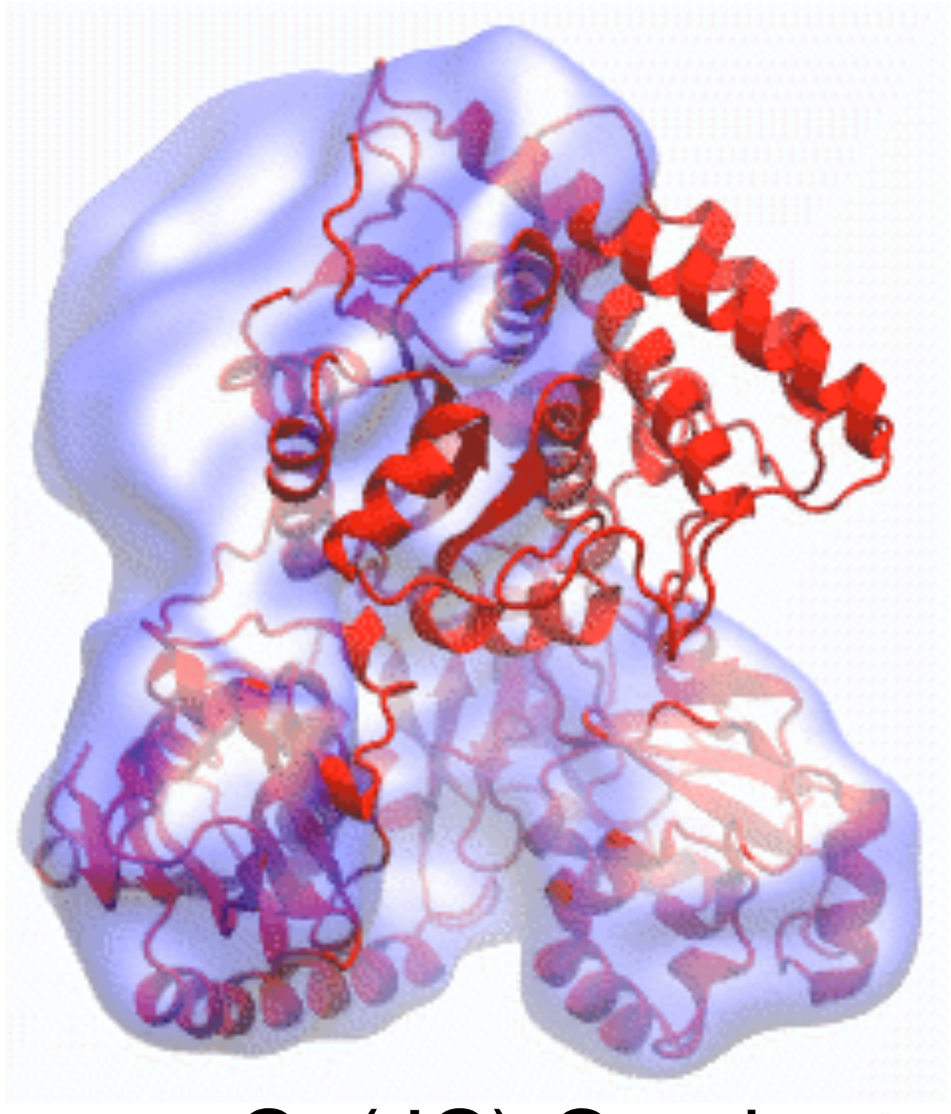
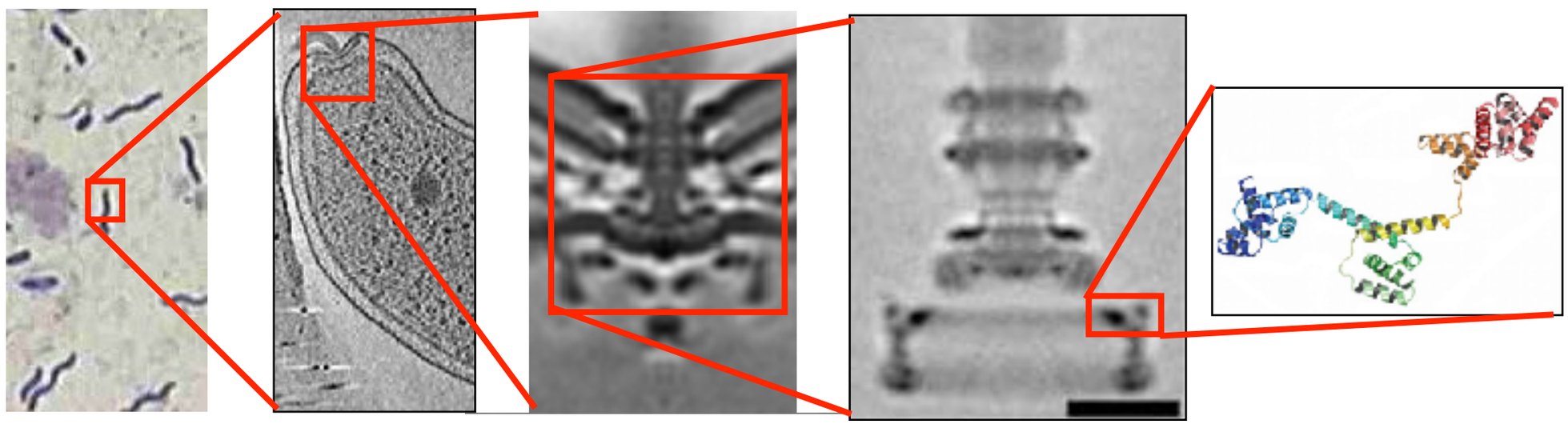
Molecular dynamics flexible fitting (MDFF)



James C. (JC) Gumbart
Georgia Institute of Technology, Atlanta

Computational Biophysics Workshop | DICP | July 12 2018

Structural biology continuum



1 cm | 1 mm | 1 µm | 1 nm | 1 Å

Light microscopy

Electron microscopy

X-ray, NMR

Modeling and molecular dynamics

resolution of the resulting image is limited* by the wavelength of light used

$$d = \frac{\lambda}{2(n \sin \theta)}$$

Abbe diffraction limit

**except in super-resolution imaging*

Single particle analysis (cryo-EM)

The Nobel Prize in Chemistry 2017 was awarded … "for developing cryo-electron microscopy for the high-resolution structure determination of biomolecules in solution".

The Nobel Prize in Chemistry 2017

Photo: A. Mahmoud

Jacques Dubochet

Prize share: 1/3

Photo: A. Mahmoud

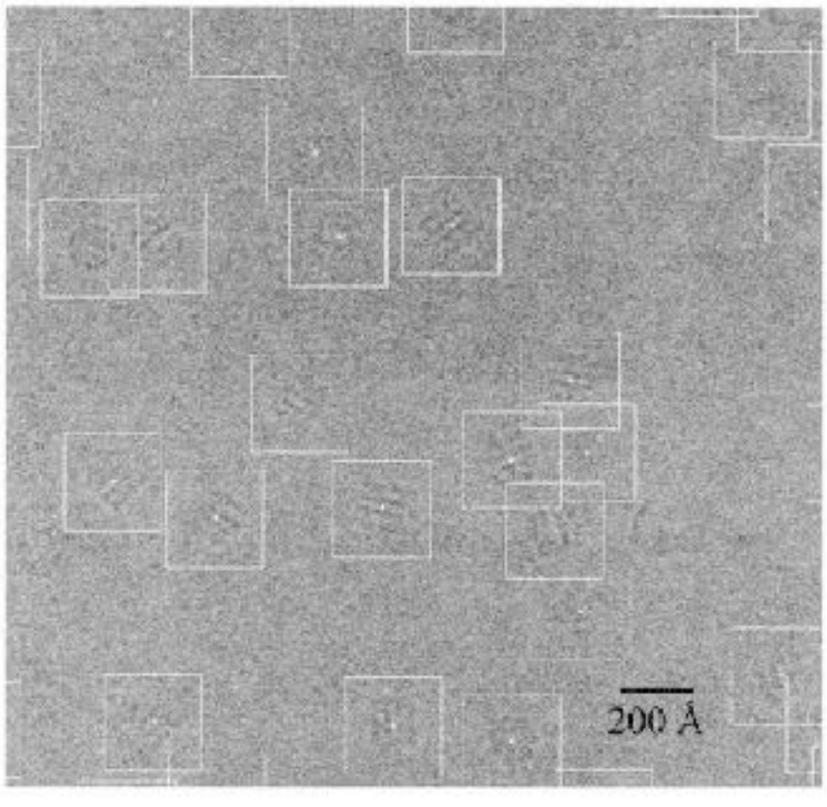
Joachim Frank

Prize share: 1/3

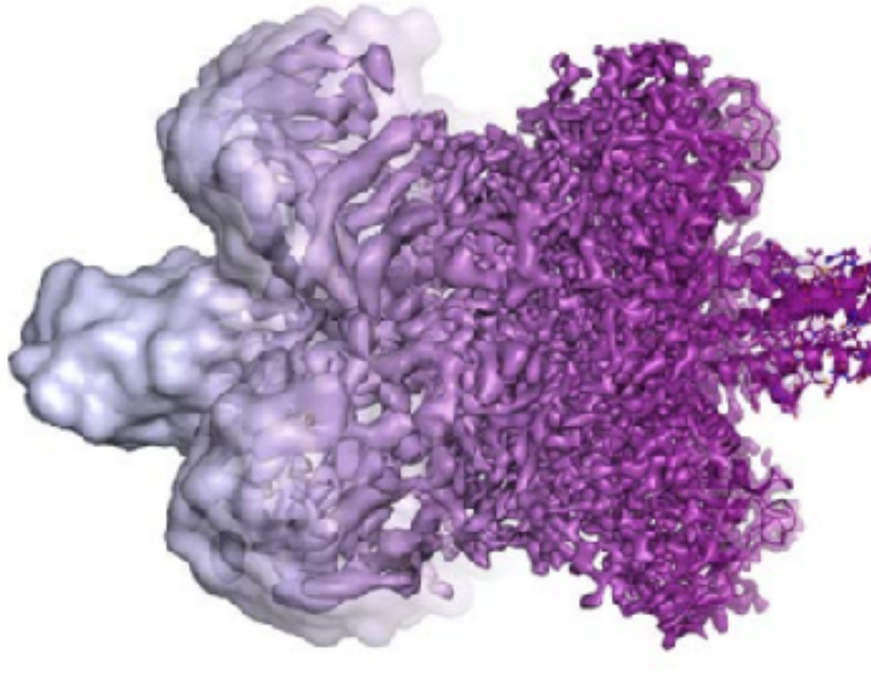
Photo: A. Mahmoud

Richard Henderson

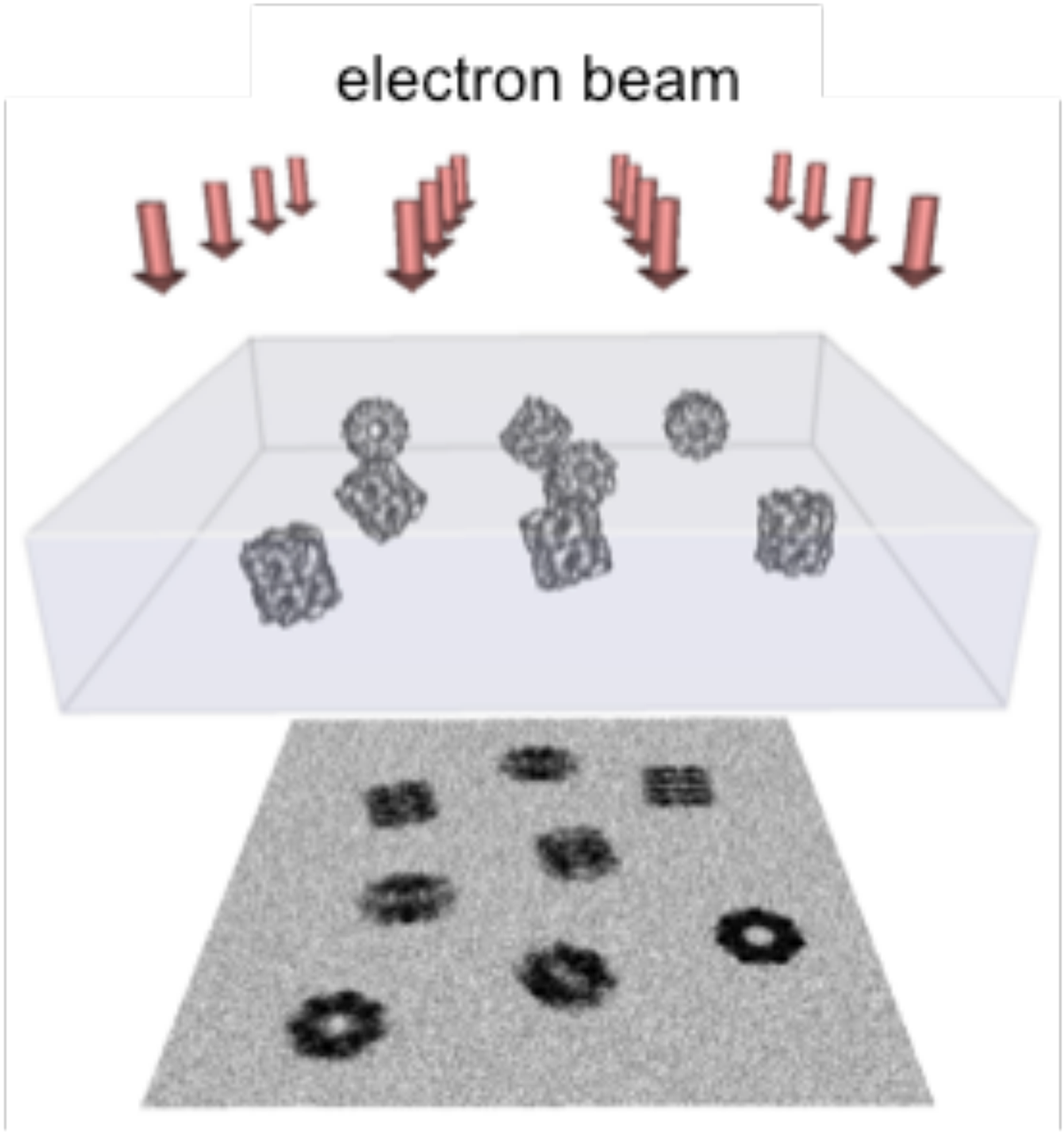
Prize share: 1/3

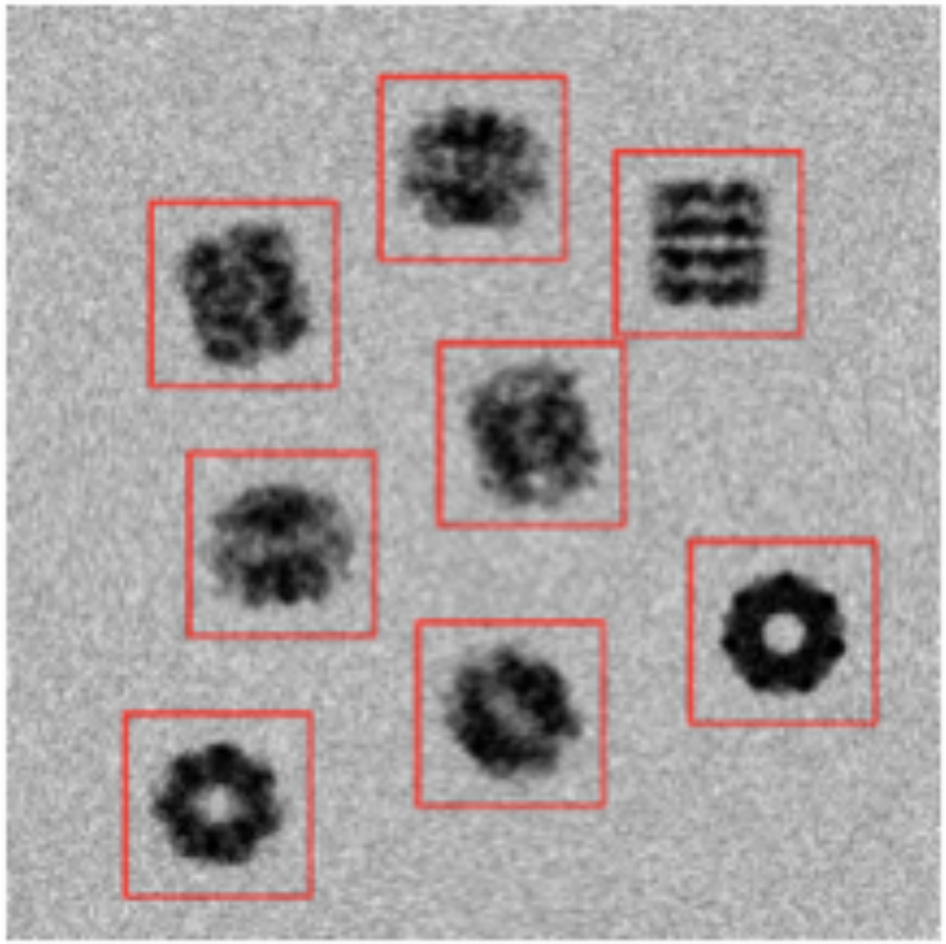


Align and average



sorting the data

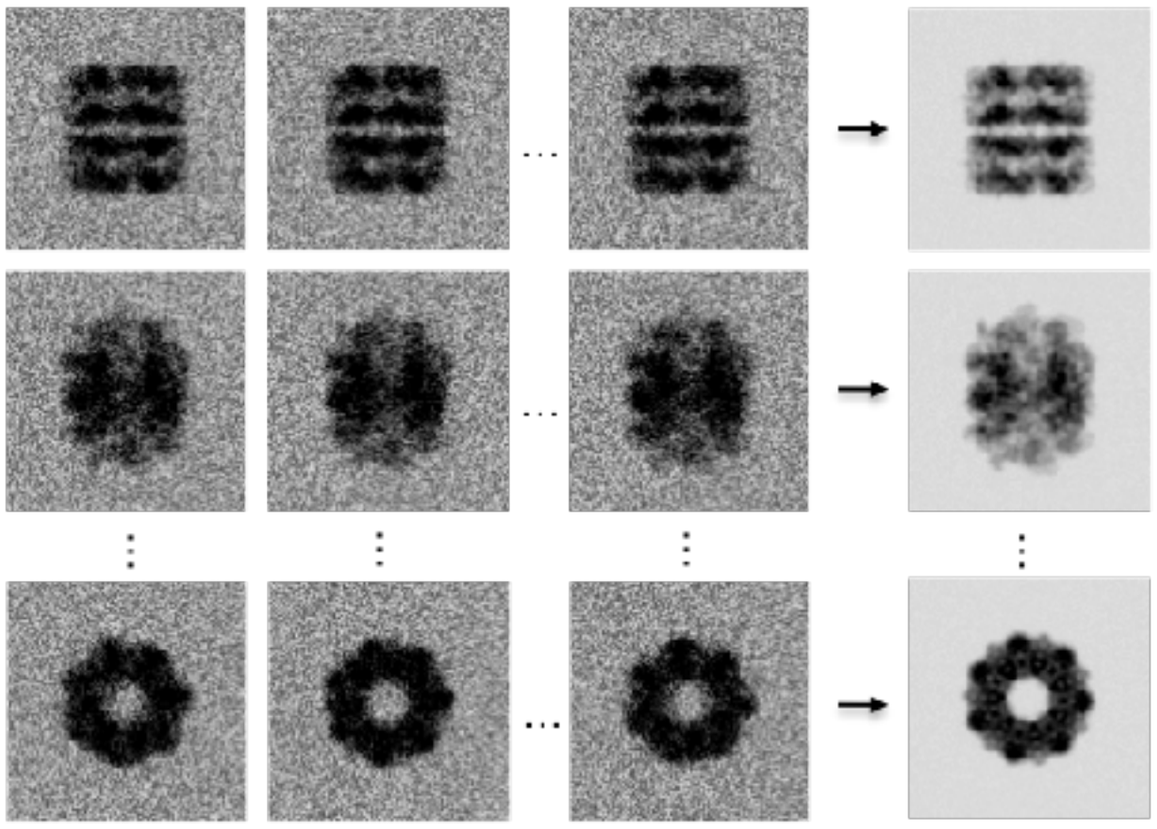


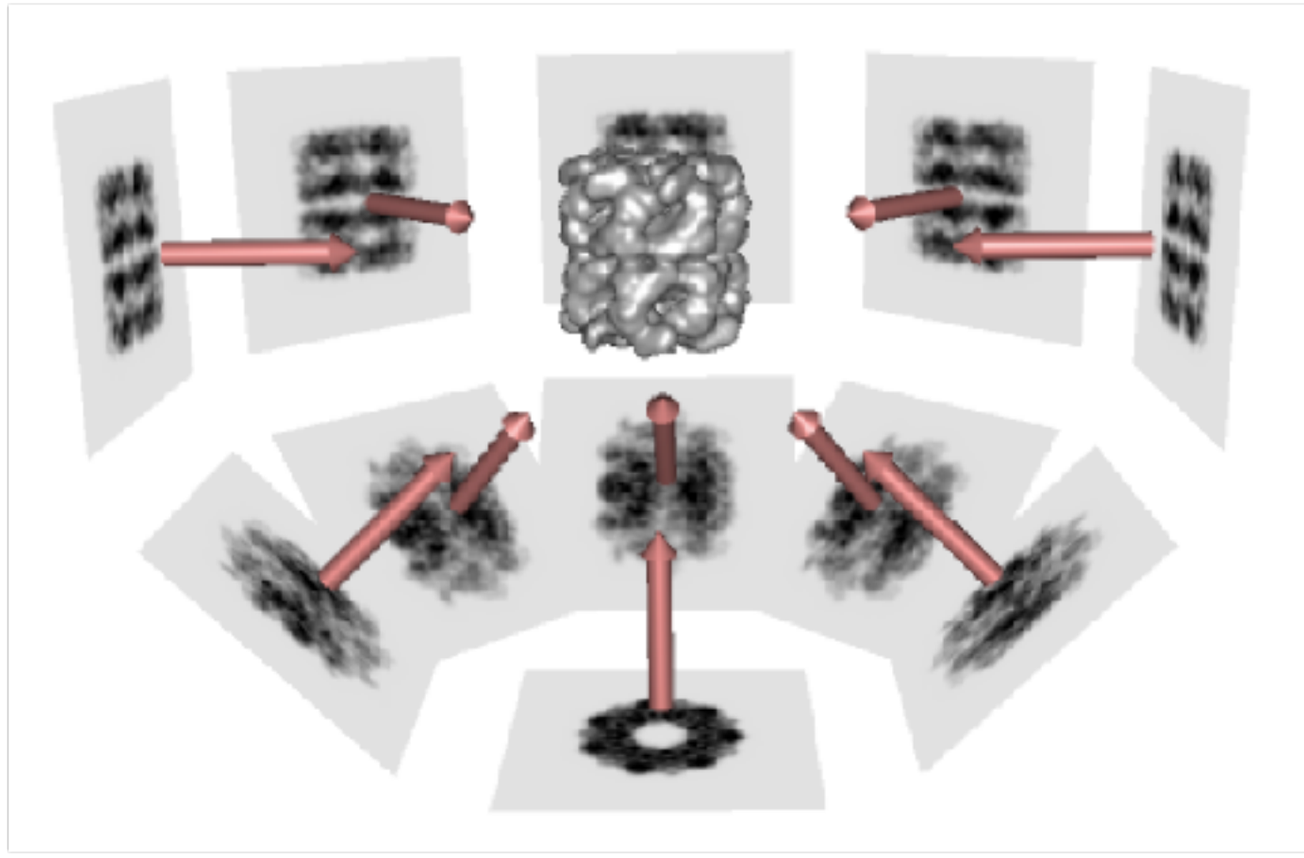


http://people.csail.mit.edu/gdp/cryoem.html

2D images are aligned and sorted computationally into classes representing homogeneous particles and perspectives

Class averages

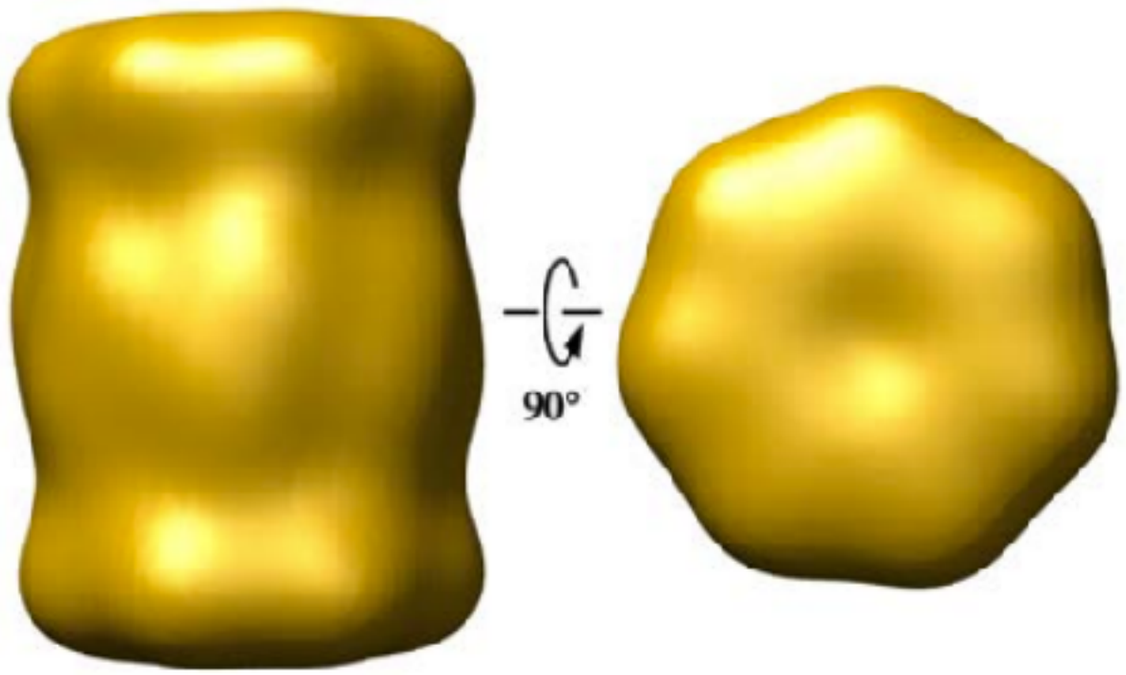


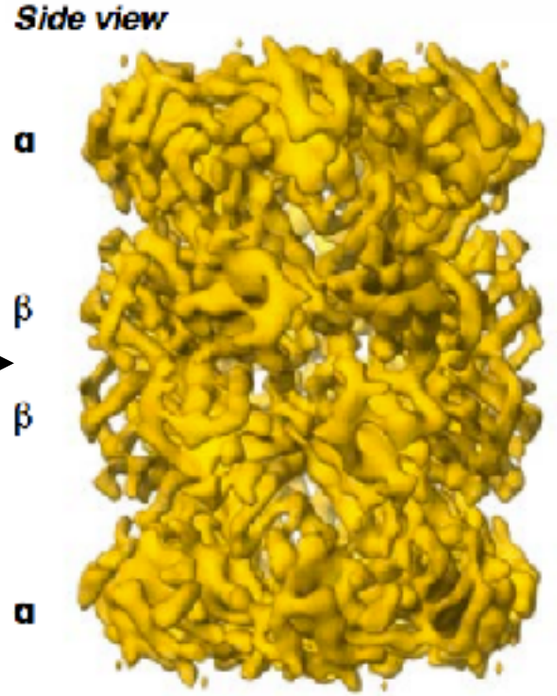


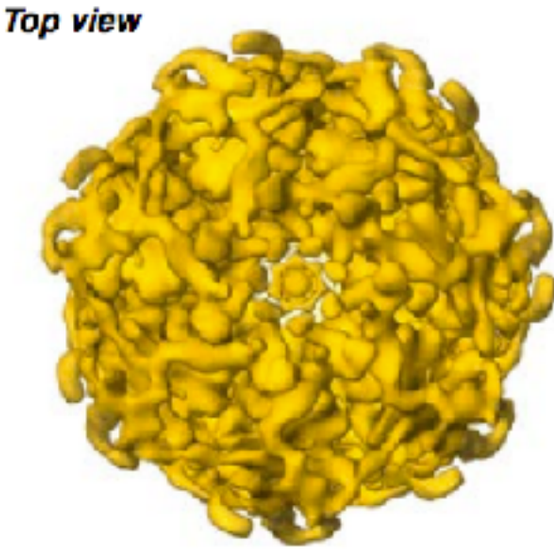
http://people.csail.mit.edu/gdp/cryoem.html

classes are then averaged and back-projected to produce 3D density map

iterative refinement





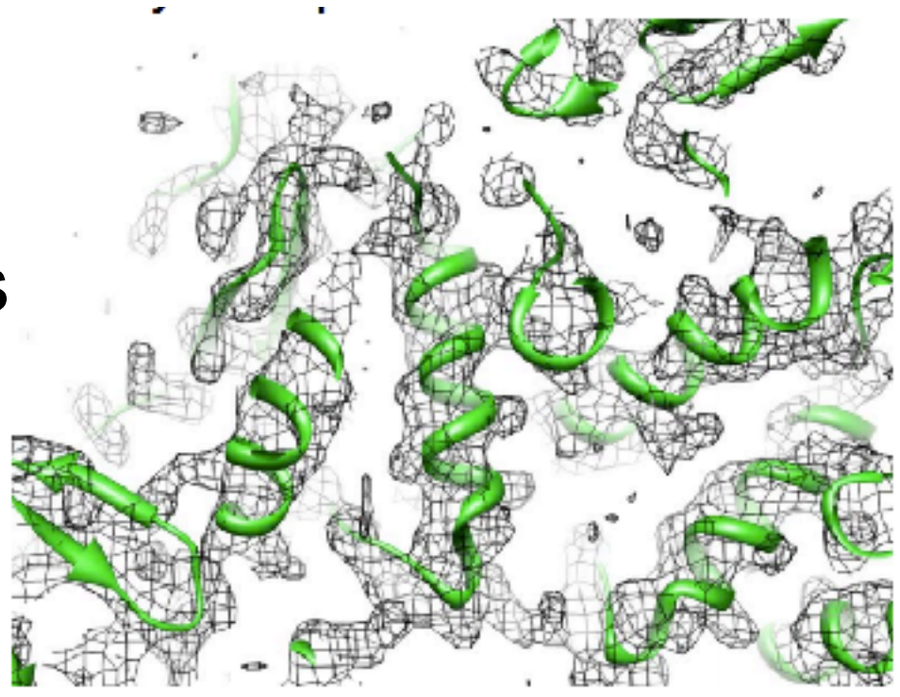


cryo-EM map of the proteosome (iteration 1)

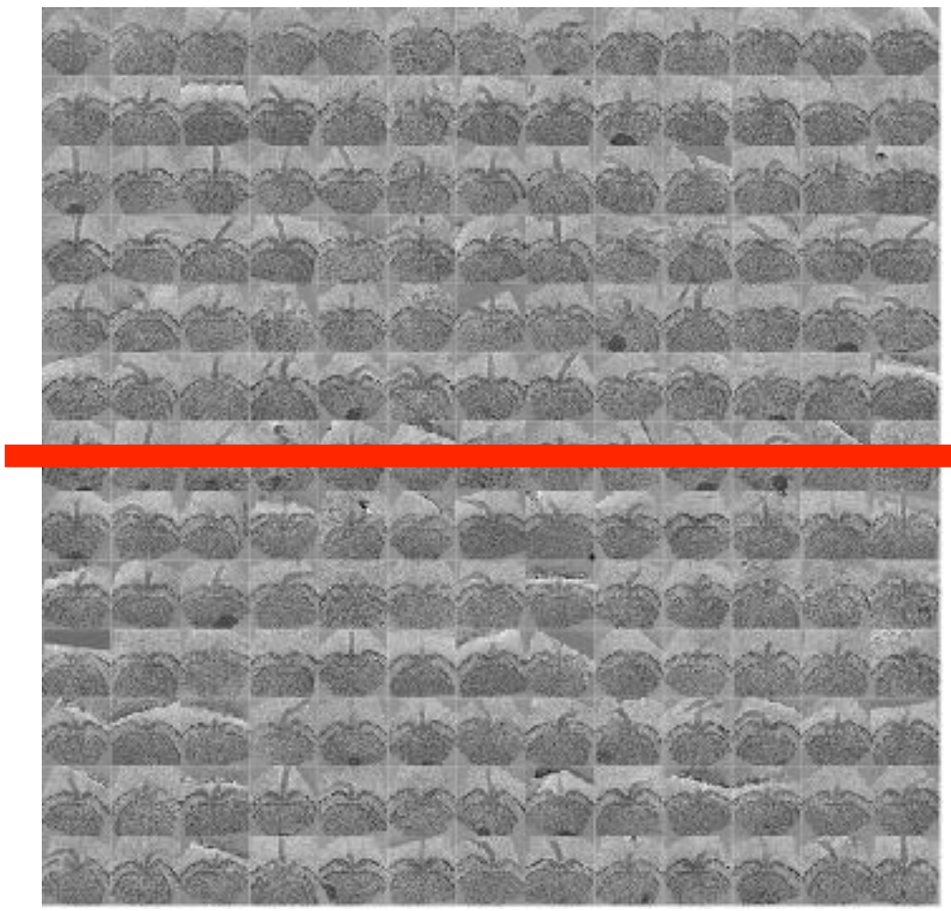
final map

back projection is iterative - need the model for **projection matching** with class averages

maps can have resolutions ranging from near-atomic (<5 Å) to 2-3 nm



map resolution



Split dataset in half, calculate two independent reconstructions

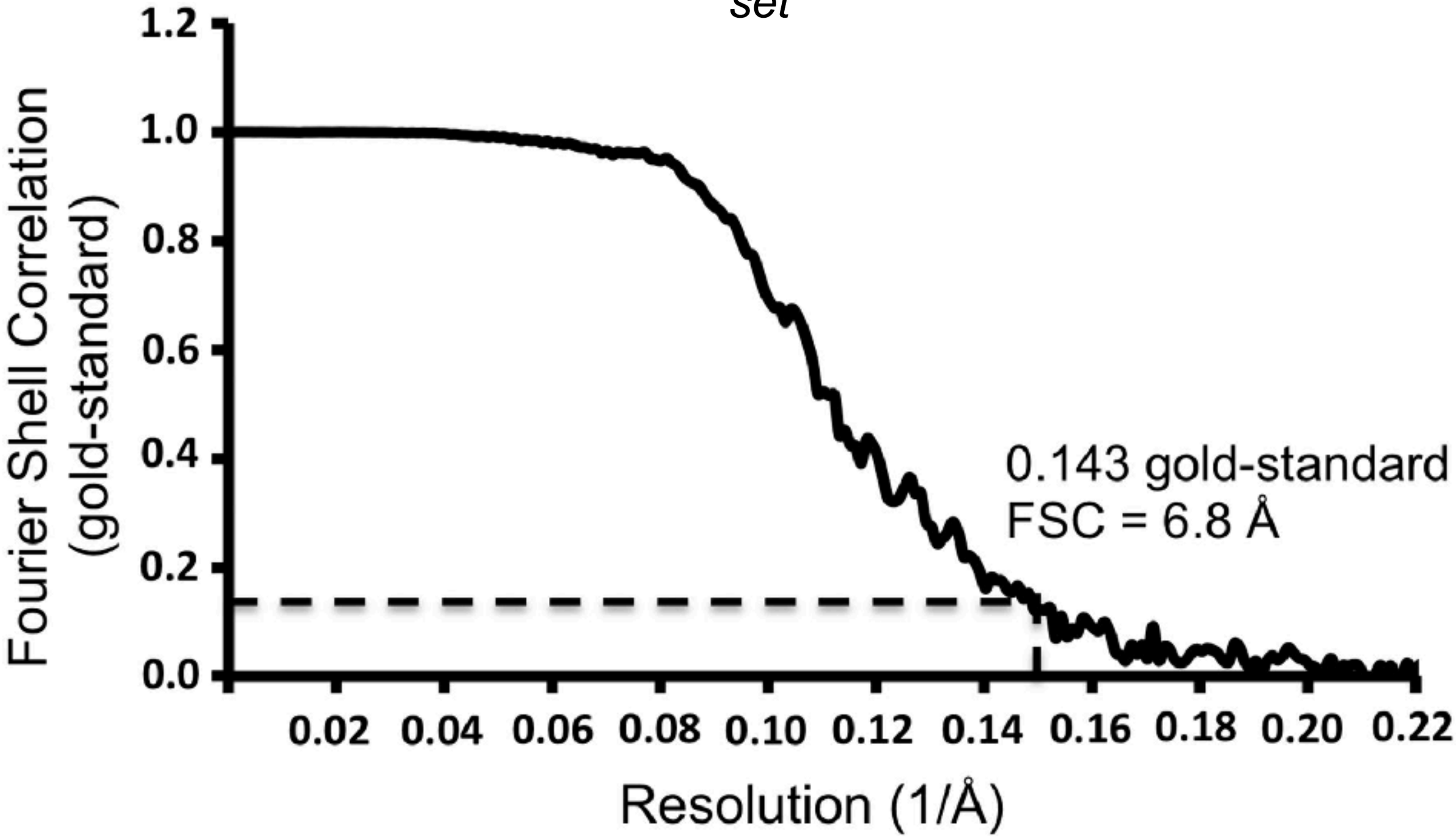
Align two structures, flip into reciprocal space (i.e., 3D FT), and calculate correlation co-efficients between bands of spatial frequency

Fourier shell correlation:

$$FSC(r) = \frac{\sum_{r_i \in r} F_1(r_i) \cdot F_2(r_i)^*}{\sqrt{\sum_{r_i \in r} |F_1(r_i)|^2 \cdot \sum_{r_i \in r} |F_2(r_i)|^2}}$$

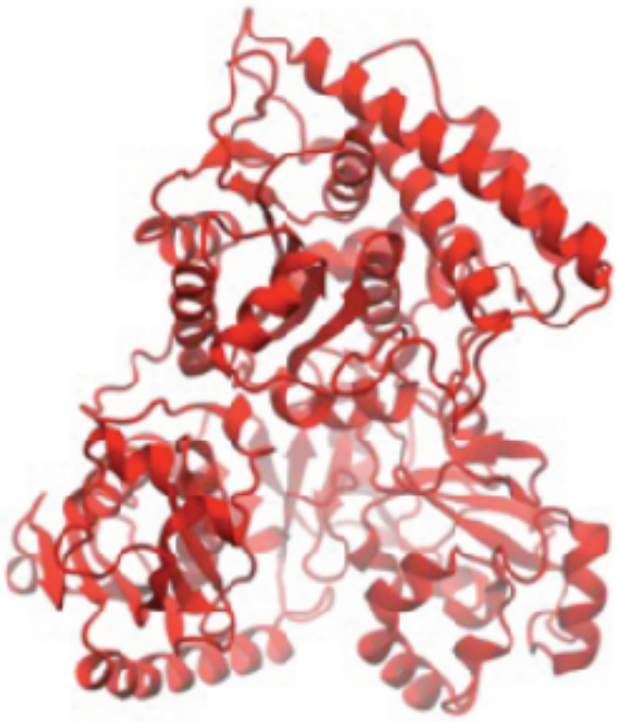
map resolution

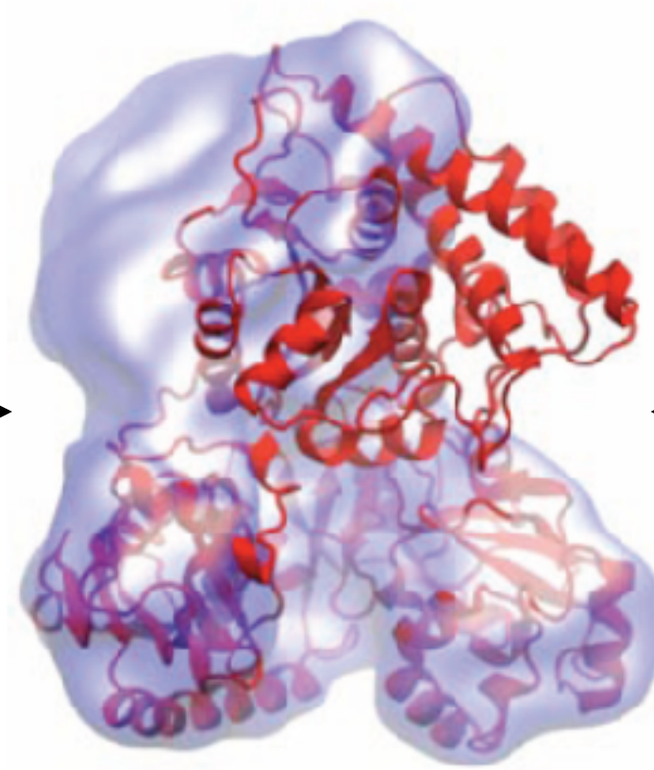
FSC between two halves of the data set



Multi-resolution modeling

high-resolution structure (X-ray/NMR/ modeling)



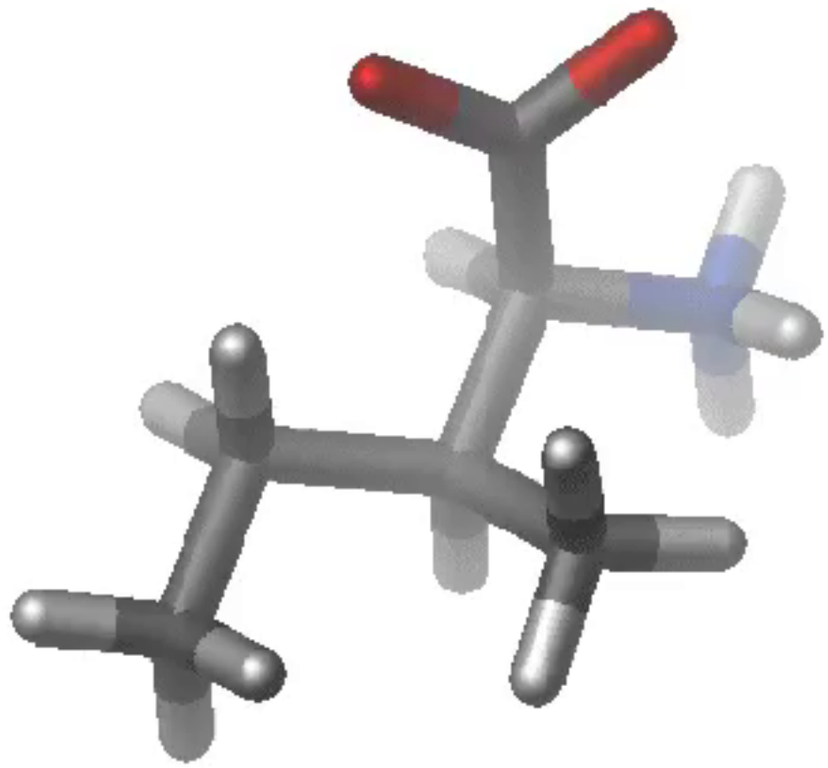


low-resolution density map (cryo-EM)

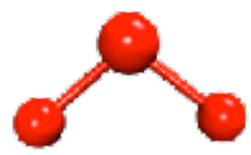
how to get these to meet in the middle?

The potential energy function

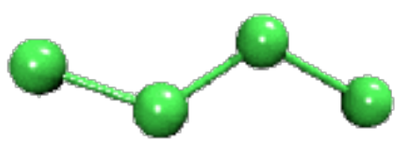
$$\mathbf{f}_i = -\frac{\partial}{\partial \mathbf{r}_i} U_{MD}(\mathbf{R}) + \mathbf{f}_i^{\text{ext}}$$



$$U_{MD} = \underbrace{\sum_{bonds} k_i^{bond}(r_i - r_0)^2}_{U_{bond}} + \underbrace{\sum_{angles} k_i^{angle}(\theta_i - \theta_0)^2}_{U_{angle}} +$$



$$\underbrace{\sum_{dihedrals} k_i^{dihe}[1 + \cos(n_i\phi_i + \delta_i)]}_{U_{dihedral}} +$$



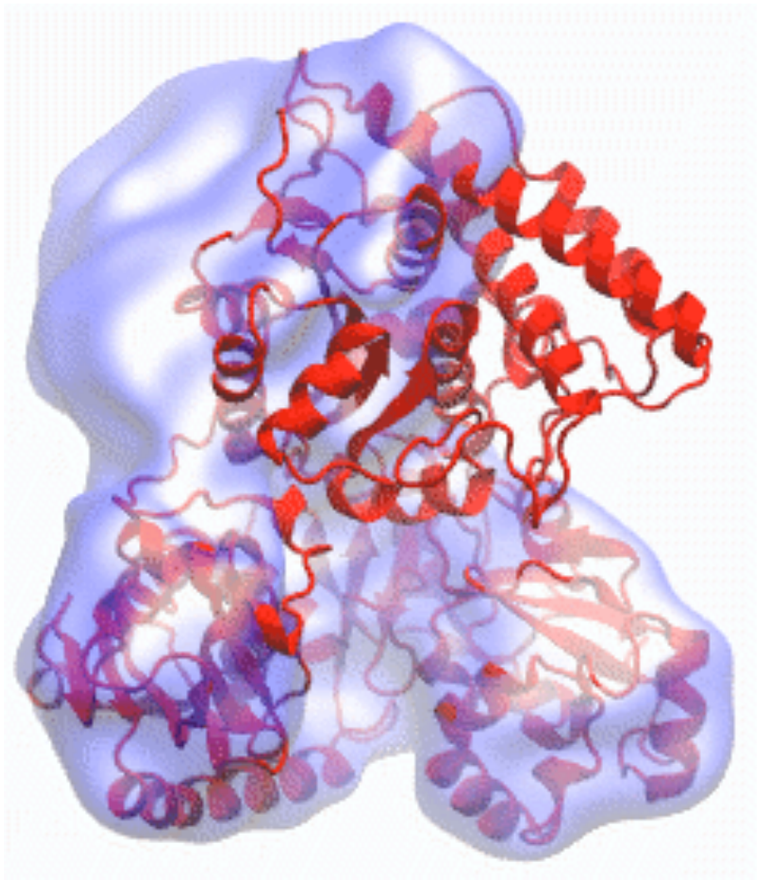
$$\underbrace{\sum_i \sum_{j \neq i} 4\epsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right] + \sum_i \sum_{j \neq i} \frac{q_i q_j}{\epsilon r_{ij}}}_{U_{nonbond}}$$

Merging cryo-EM data with atomic structures

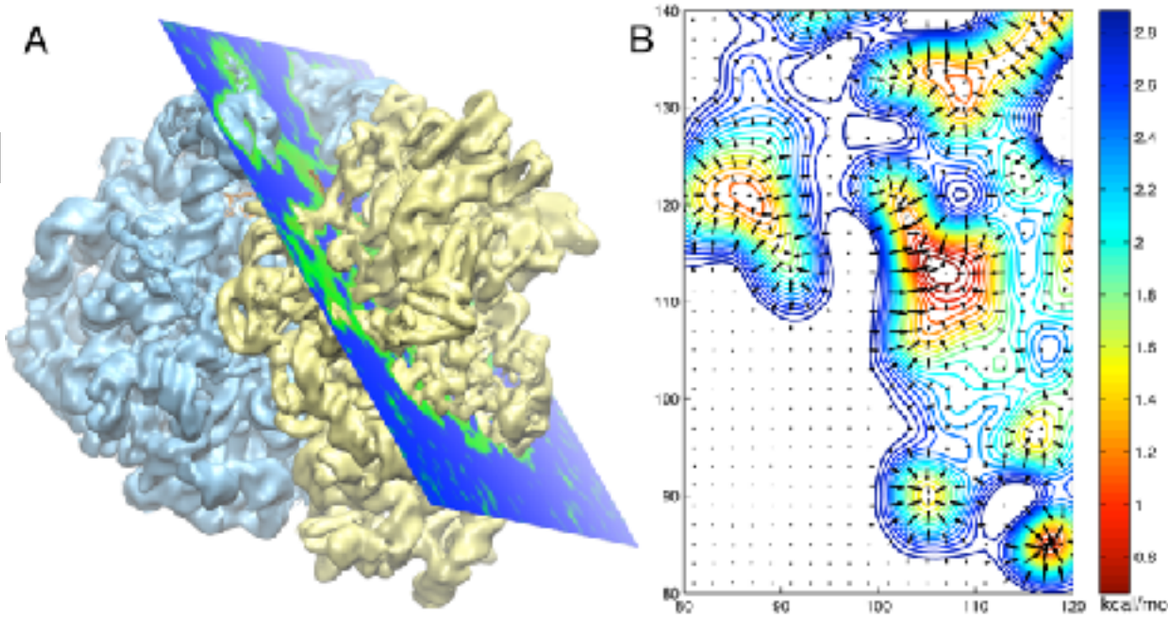
using Molecular Dynamics Flexible Fitting (MDFF)

Two terms are added to the MD potential

$$U_{total} = U_{MD} + U_{EM} + U_{SS}$$



Flexible fitting of atomic structures into electron microscopy maps using molecular dynamics. L G. Trabuco*, E Villa*, K Mitra, **J Frank**, K Schulten. *Structure*, **16**, 673-683, 2008.



An external potential derived from the EM map is defined on a grid as

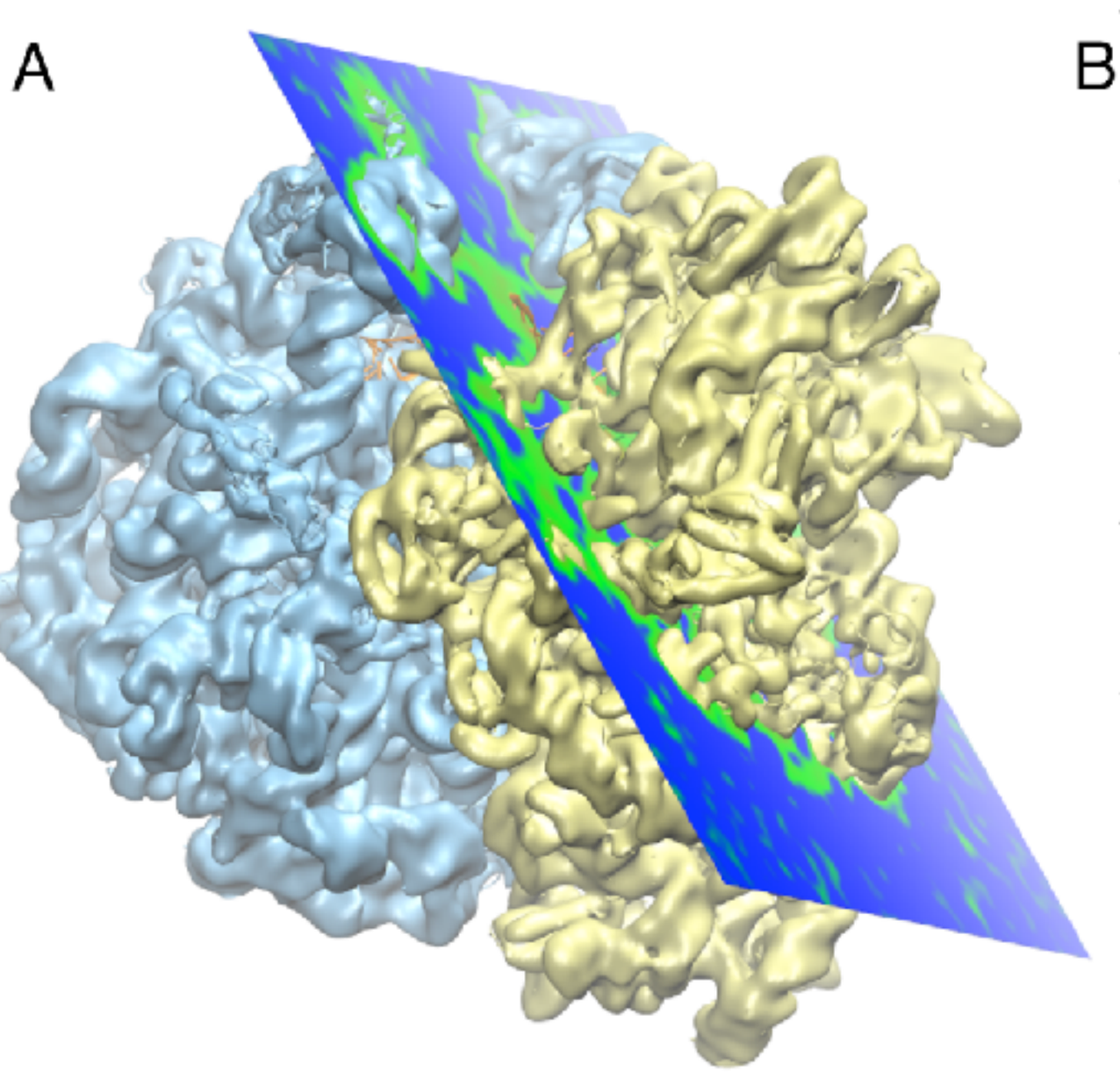
$$U_{EM}(\mathbf{R}) = \sum_j w_j V_{EM}(\mathbf{r}_j)$$

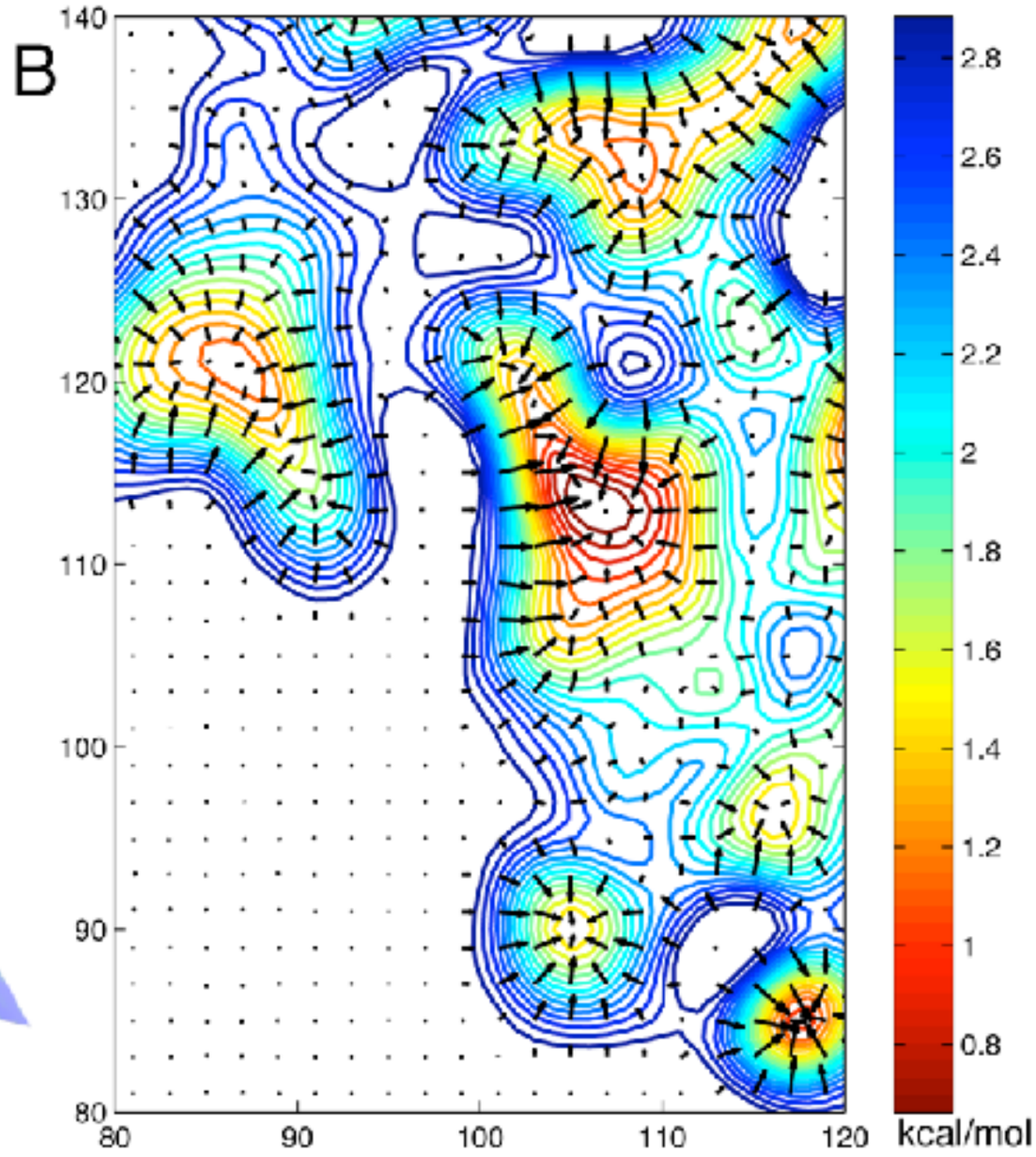
$$V_{EM}(\mathbf{r}) = \begin{cases} \xi \left(1 - \frac{\Phi(\mathbf{r}) - \Phi_{thr}}{\Phi_{max} - \Phi_{thr}}\right) & \text{if } \Phi(\mathbf{r}) \geq \Phi_{thr}, \\ \xi & \text{if } \Phi(\mathbf{r}) < \Phi_{thr}. \end{cases}$$

A mass-weighted force is then applied to each atom

$$\mathbf{f}_i^{EM} = -\nabla U_{EM}(\mathbf{R}) = -w_i \partial V_{EM}(\mathbf{r}_i)/\partial r_i$$

Map-derived potential and gradients





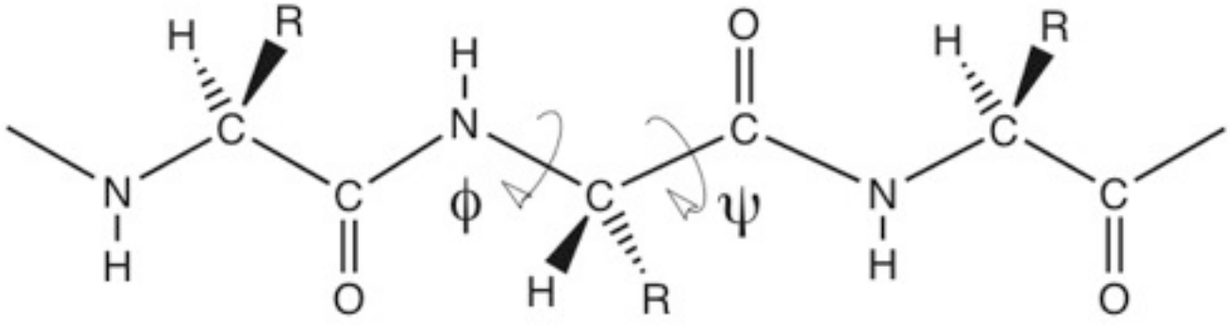
Flexible fitting of atomic structures into electron microscopy maps using molecular dynamics.
L G. Trabuco*, E Villa*, K Mitra, J Frank, K Schulten.
Structure, **16**, 673-683, 2008.

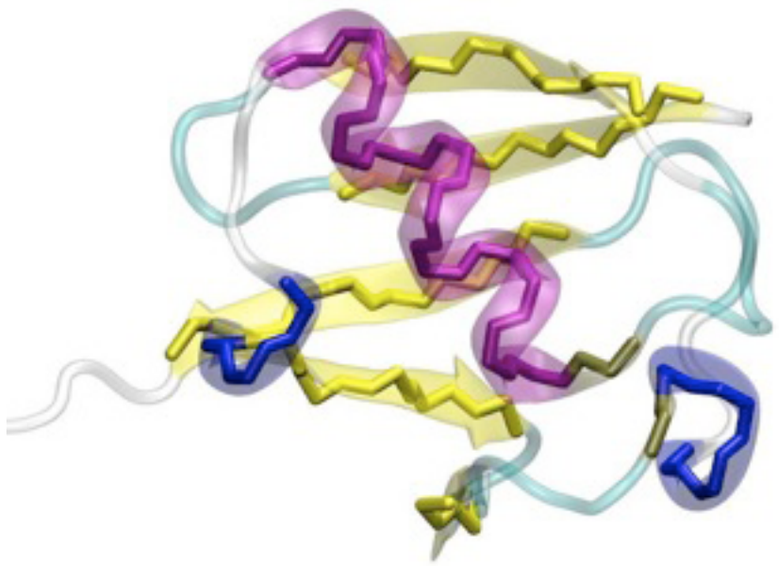
Arrows (representing forces) point to regions of higher density (lower energy)

MDFF: Secondary-structure restraints

Harmonic restraints are applied to preserve secondary structure of proteins and nucleic acids, avoiding "overfitting"

$$U_{\mathrm{SS}} = \sum_{\mathrm{restraints}} k_\mu (\mu - \mu_0)^2$$





For proteins, ϕ and ψ dihedral angles of residues within helices or β-strands are restrained. Hydrogen-bond restraints are also an option.

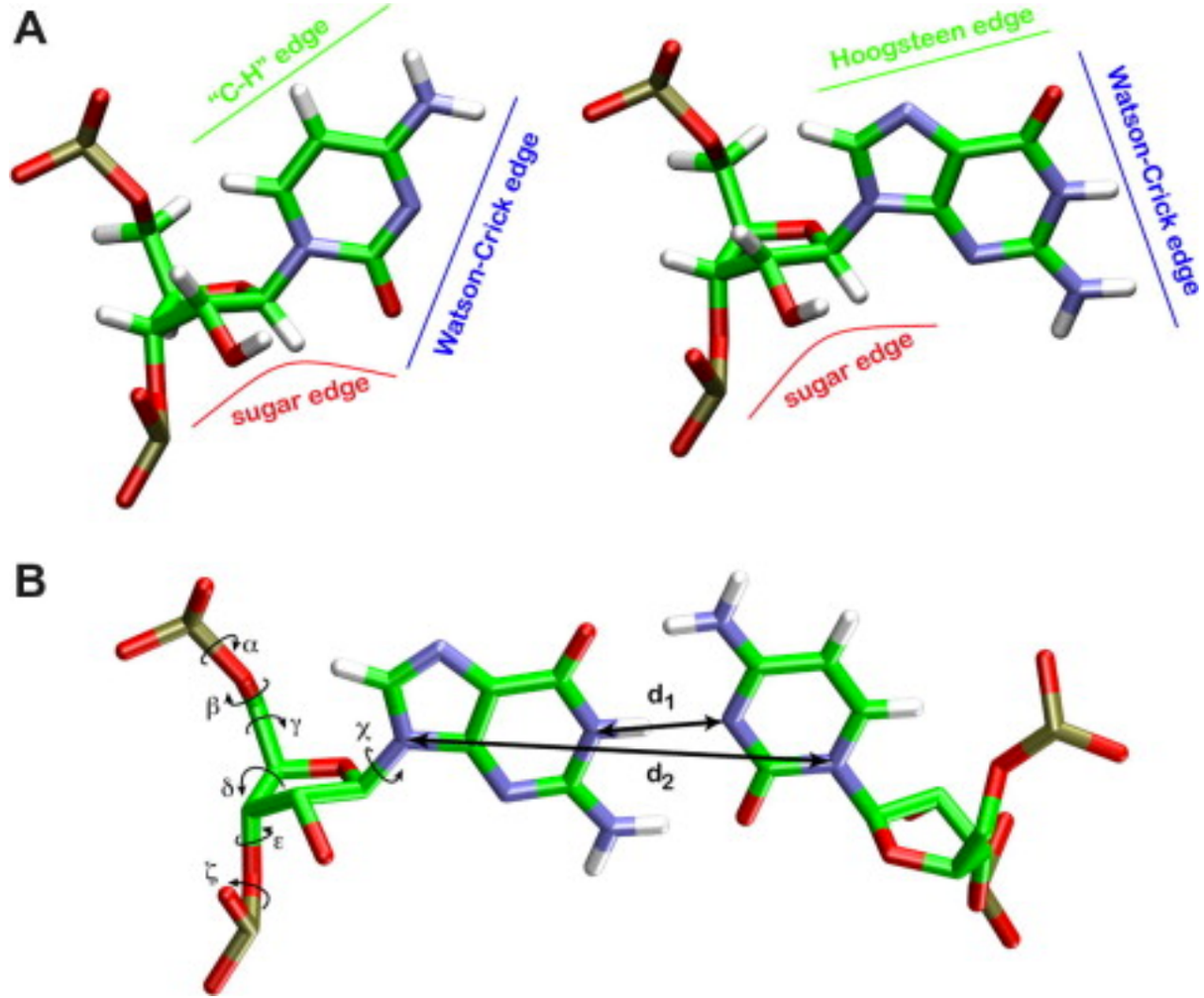
Flexible fitting of atomic structures into electron microscopy maps using molecular dynamics.
L G. Trabuco*, E Villa*, K Mitra, J Frank, K Schulten.
Structure, **16**, 673-683, 2008.

MDFF: Secondary-structure restraints

Harmonic restraints are applied to preserve secondary structure of proteins and nucleic acids, avoiding "overfitting"

$$U_{\mathrm{SS}} = \sum_{\mathrm{restraints}} k_\mu (\mu - \mu_0)^2$$

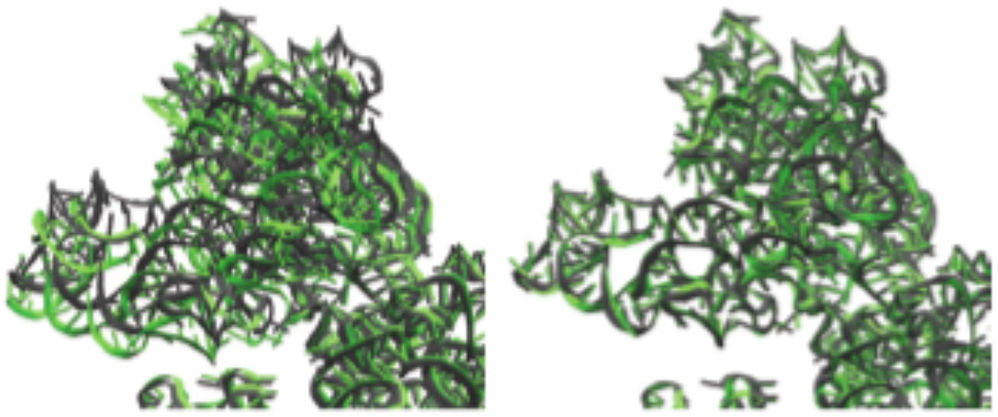
For nucleic acids, distance and dihedral restraints are applied to a selected set of base pairs.



Flexible fitting of atomic structures into electron microscopy maps using molecular dynamics.
L G. Trabuco*, E Villa*, K Mitra, J Frank, K Schulten.
Structure, **16**, 673-683, 2008.

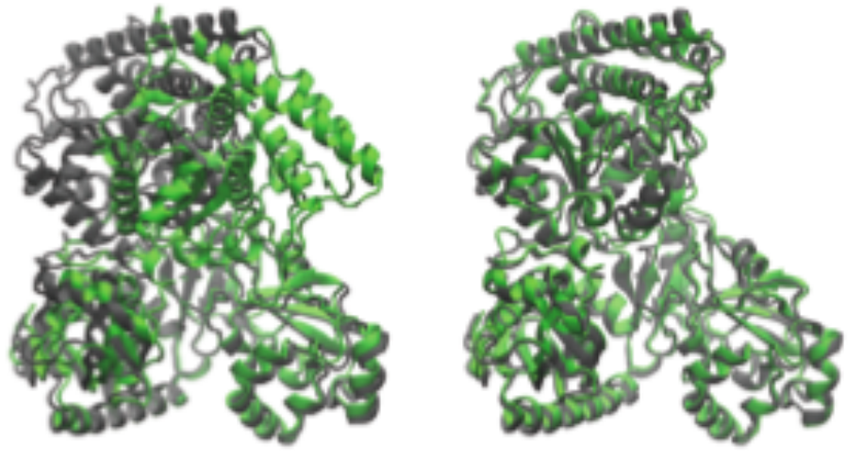
MDFF: Validation

E. coli 16S

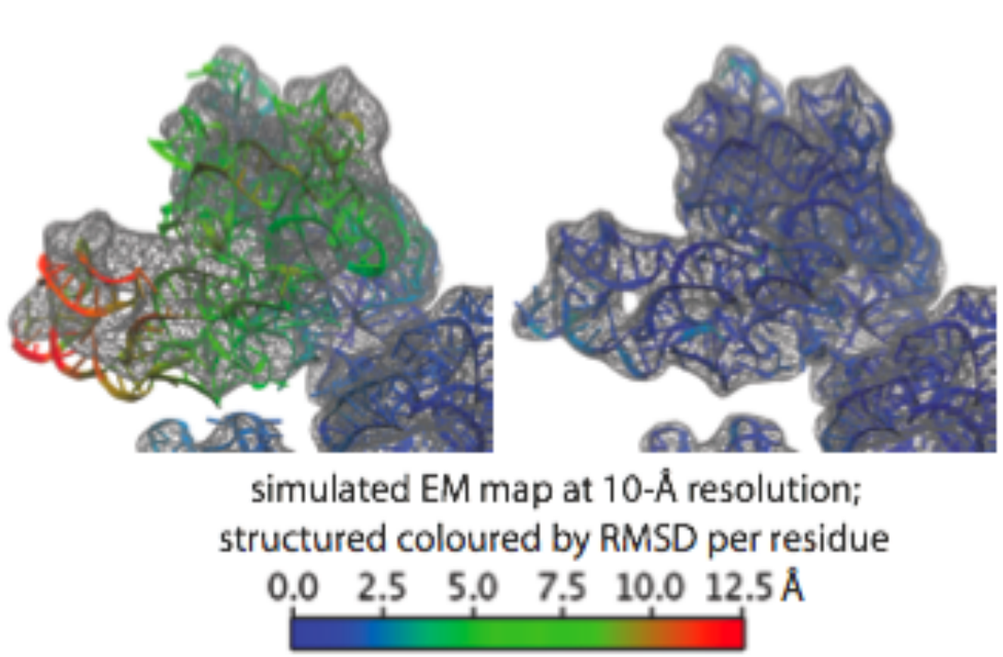


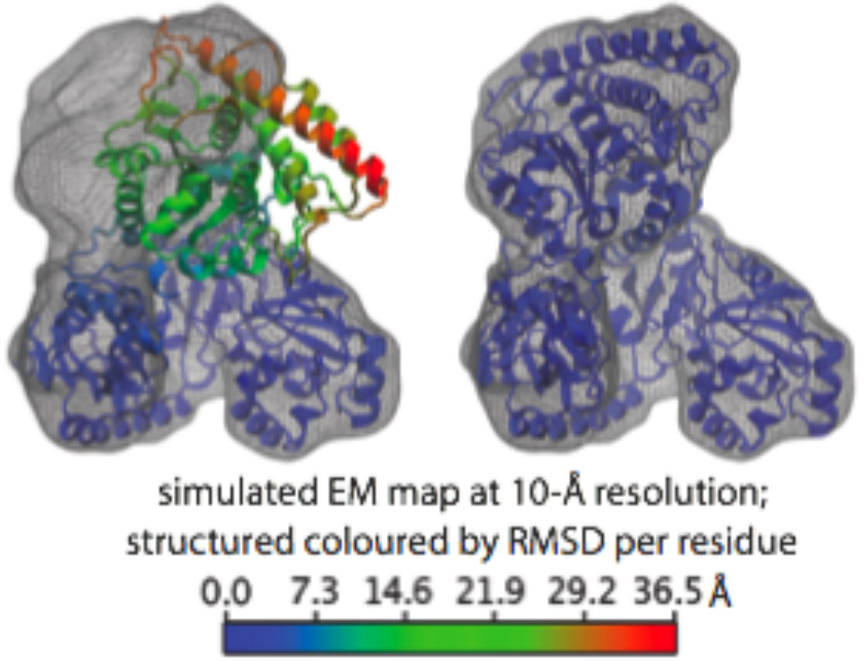
gray: 2AVY; green: 2AW7

Acetyl-CoA synthase



PDB 1OAO: gray open; green closed

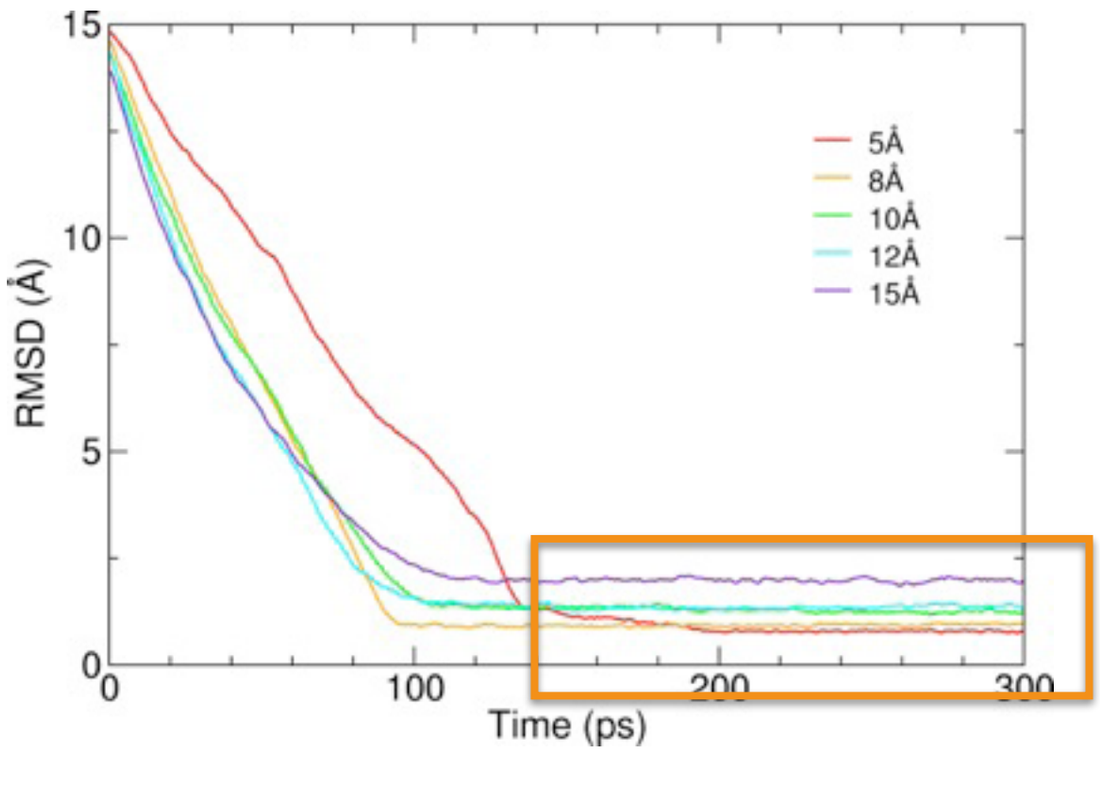


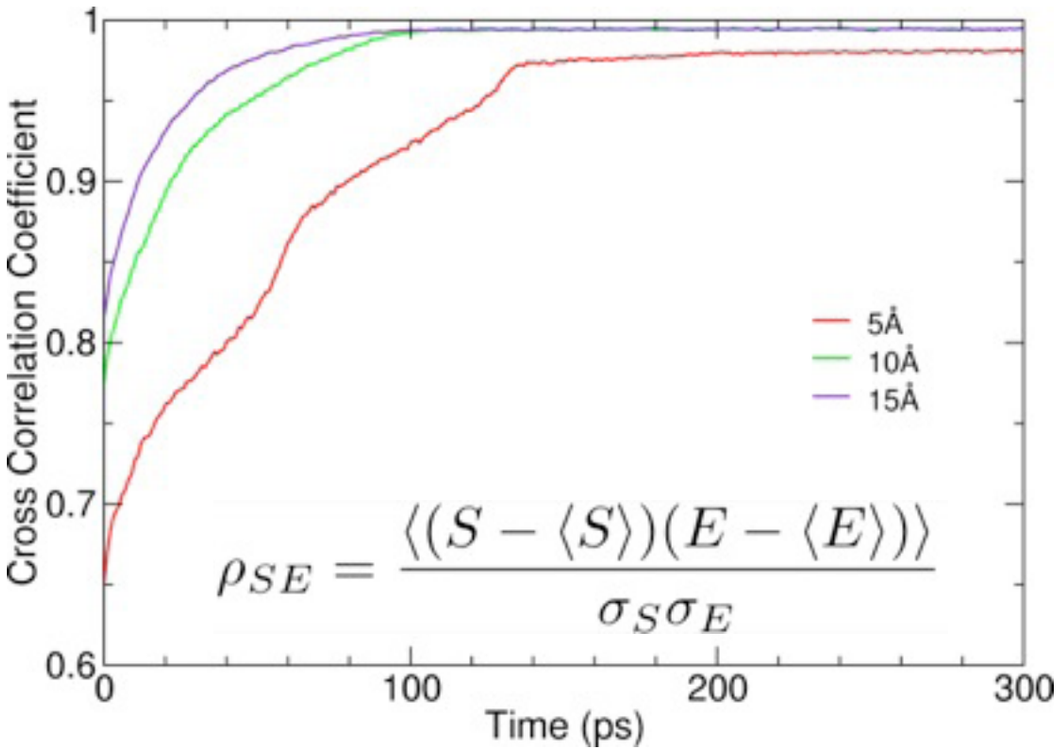


Simulated maps used as targets for proteins with crystal structures in two conformations

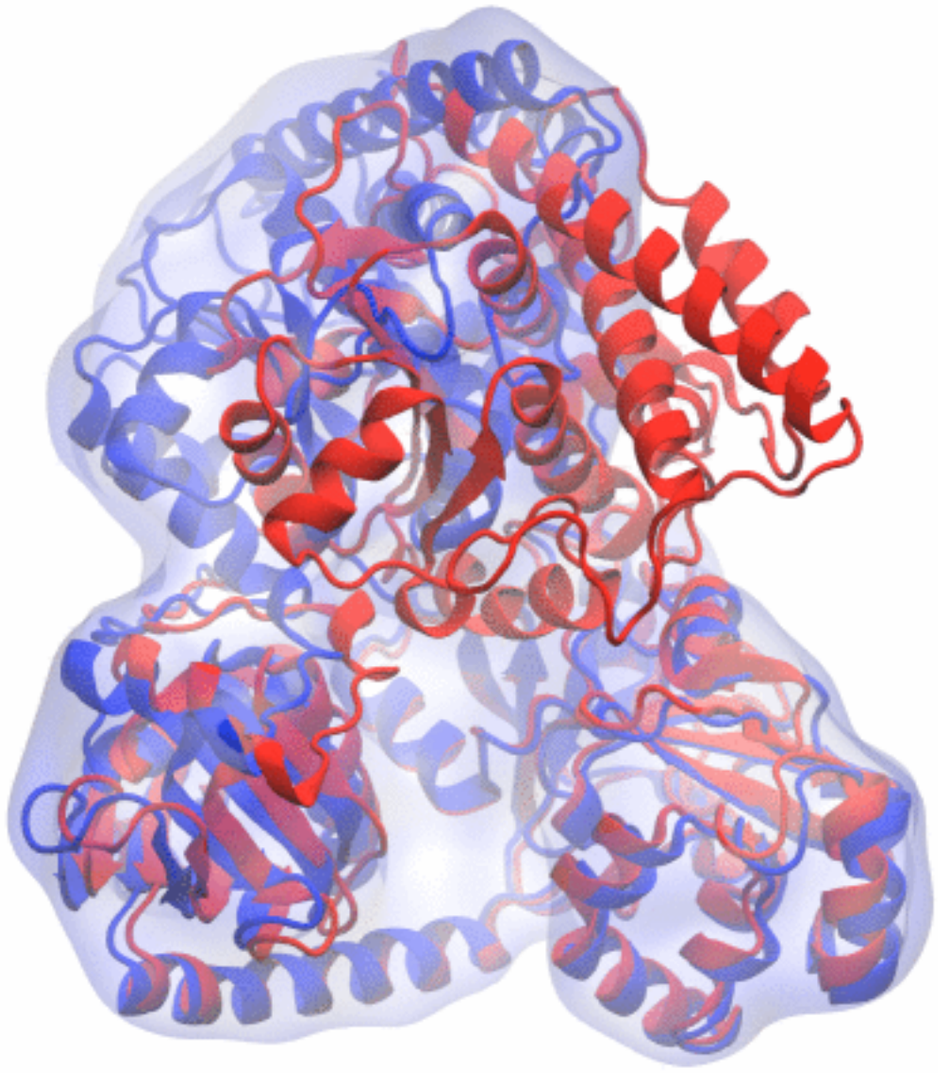
MDFF: Validation

Ways to evaluate the quality and convergence of the fit are to track RMSD and cross-correlation coefficient (CCC)

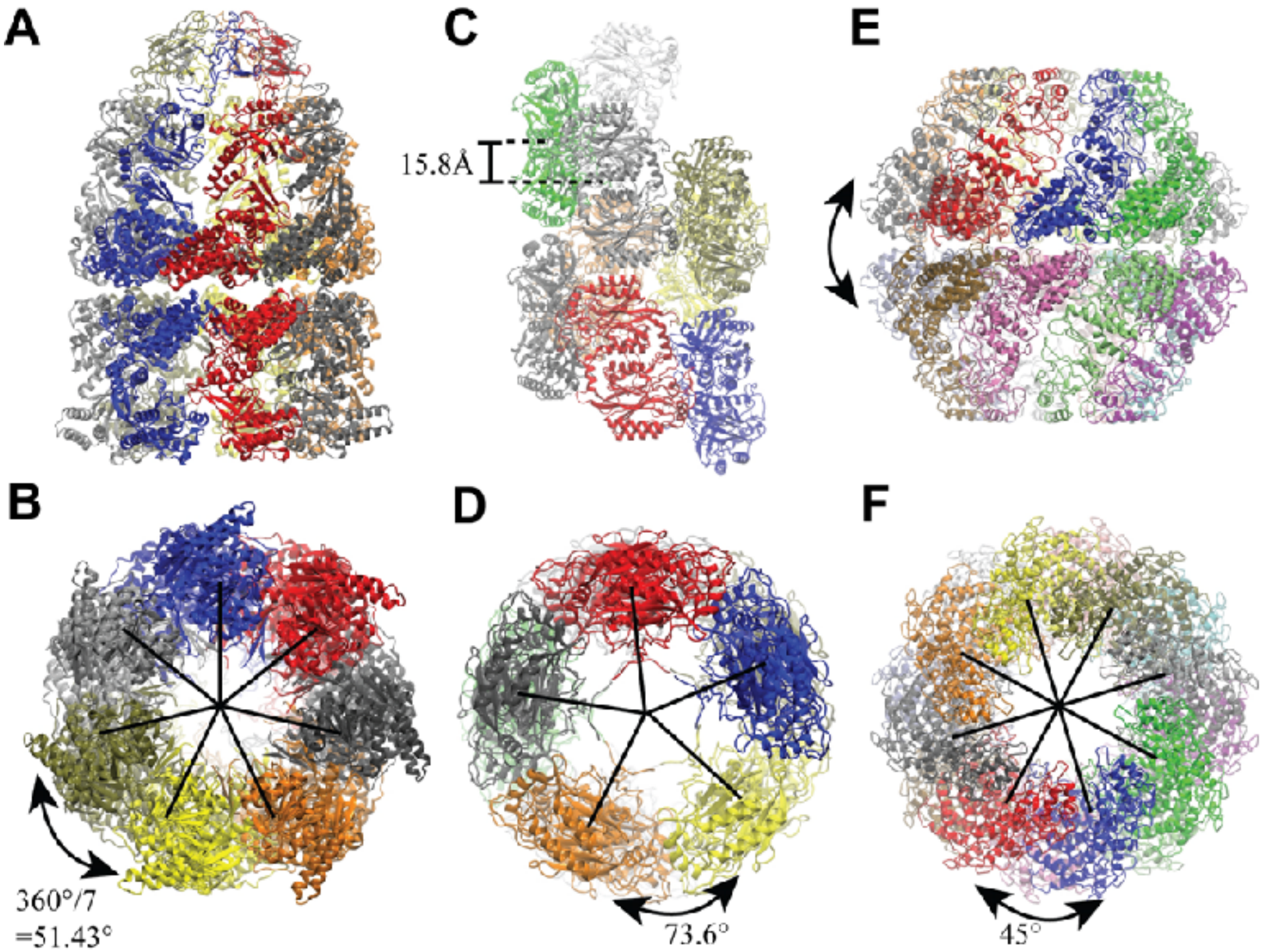




Fluctuations about the best fit
(“ensemble” of fitted structures)



Symmetry in biological molecules



GroEL-GroES
7-fold

Nitrilase
helical

Mm-cpn
16-fold

Symmetry as seen in cryo-EM maps

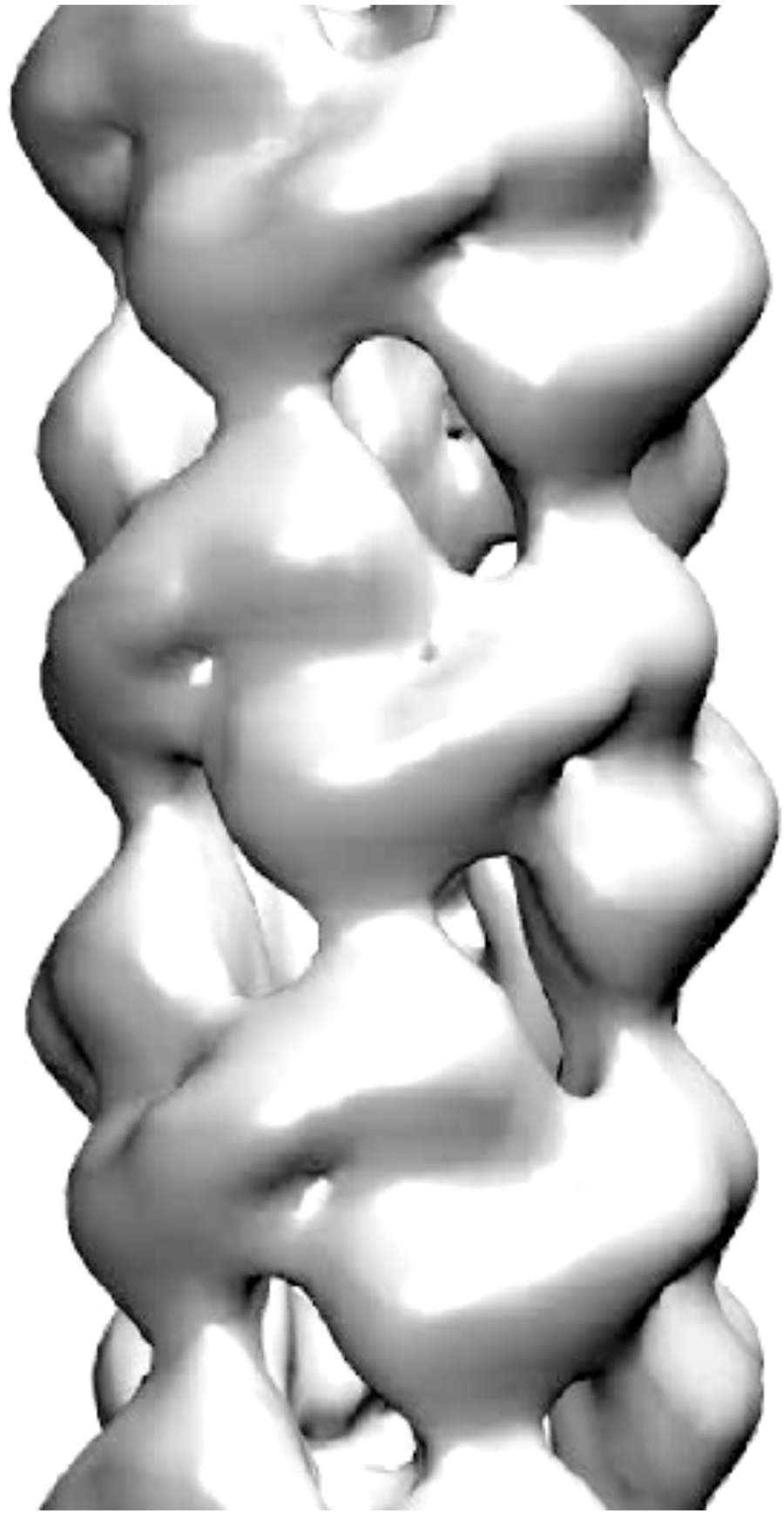
-helically symmetric nitrilase

-symmetry defined by two parameters, pitch (rotation about central axis) and rise

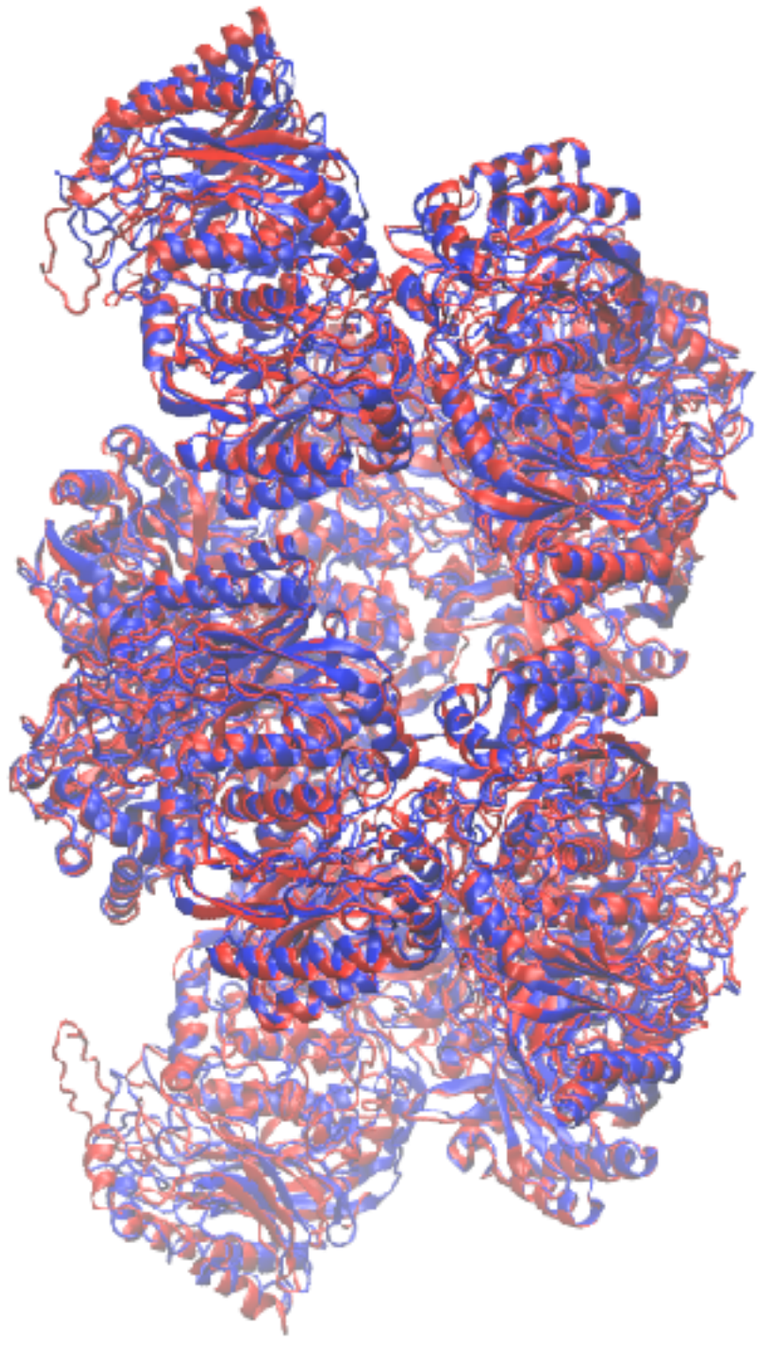
-can use parameters from cryo-EM map to define transformation matrices U_i

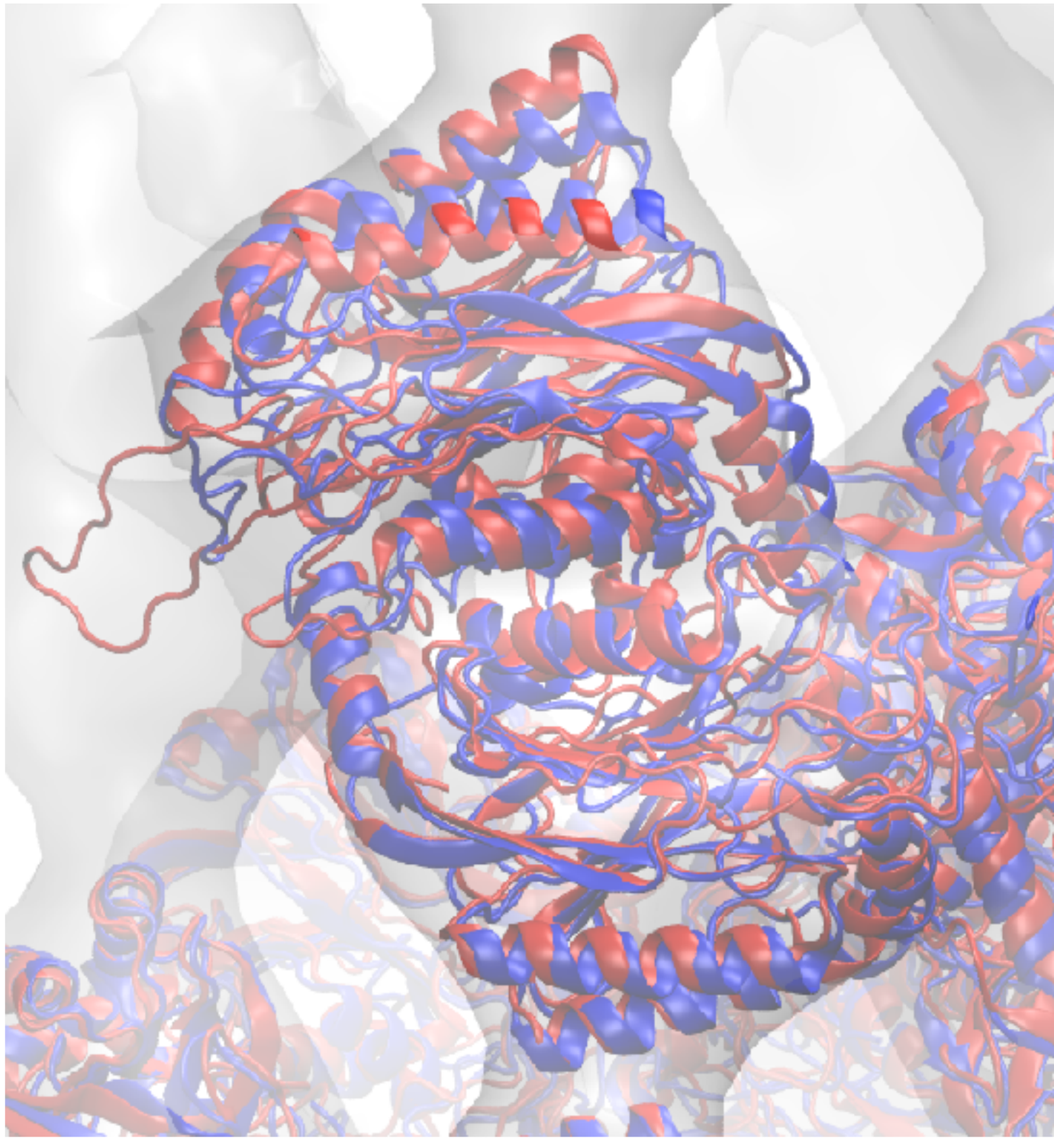
$$\overleftrightarrow{U} = \begin{pmatrix} \cos\theta & -\sin\theta & 0 & 0 \\ \sin\theta & \cos\theta & 0 & 0 \\ 0 & 0 & 1 & \Delta z \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

NAMD can also guess the parameters



Seeing the effects of symmetry restraints

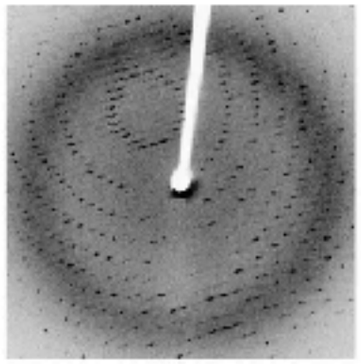


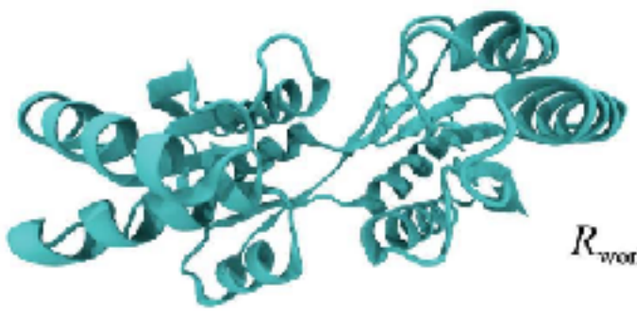


red - fit without symmetry
blue - fit with symmetry

Symmetry-restrained flexible fitting for symmetric EM maps. KY Chan, J Gumbart, R McGreevy, J M. Watermeyer, B. T Sewell, K Schulten. *Structure*, **19**, 1211-1218, 2011.

xMDFF: fitting for low resolution X-ray structures

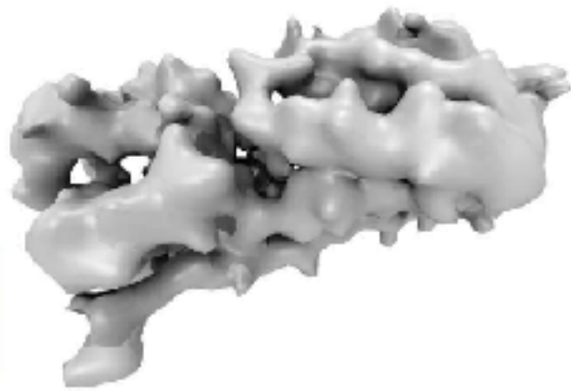




Amplitude $|F(q)|$ from X-ray diffraction intensity

Phase $\varphi(q)$ from tentative structural model

X-ray electron density $\Phi(\mathbf{r})$ using *PHENIX*



Compute steering forces f^{fit} from $\Phi(\mathbf{r})$

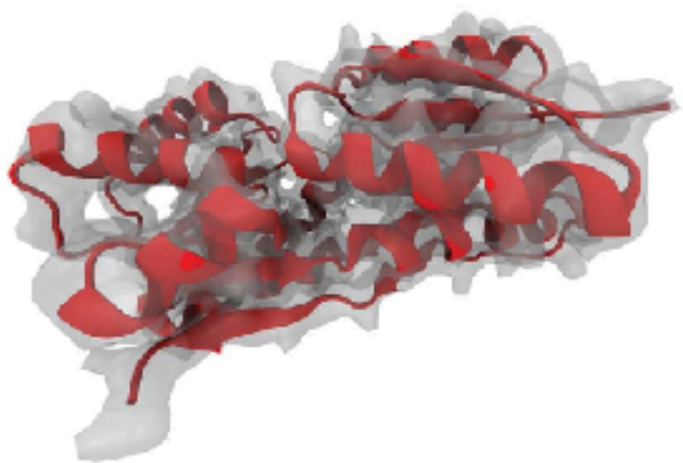
MDFF fitting of the tentative model to the density using f^{fit}

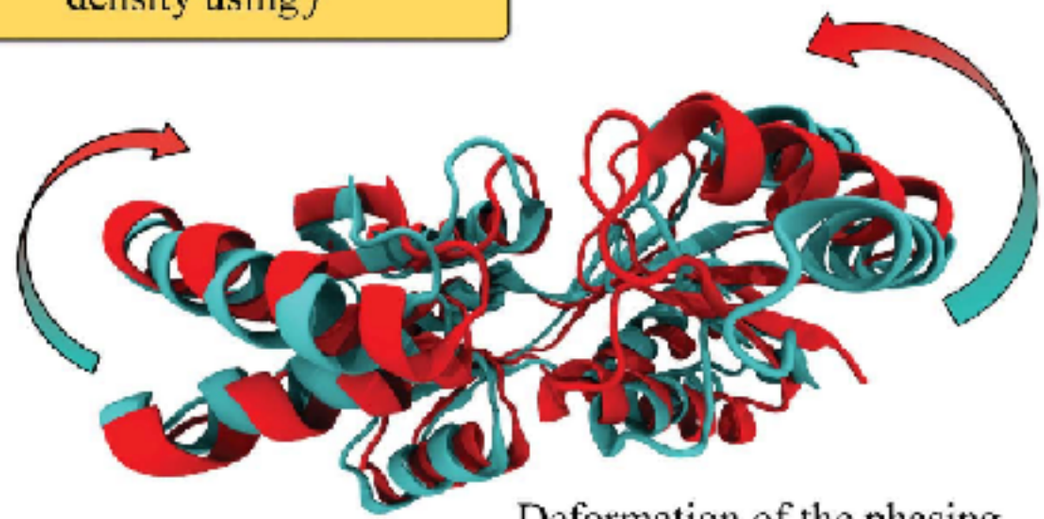
Convergence test

R_{work}/R_{free} > tolerance

R_{work}/R_{free} < tolerance

Refined model

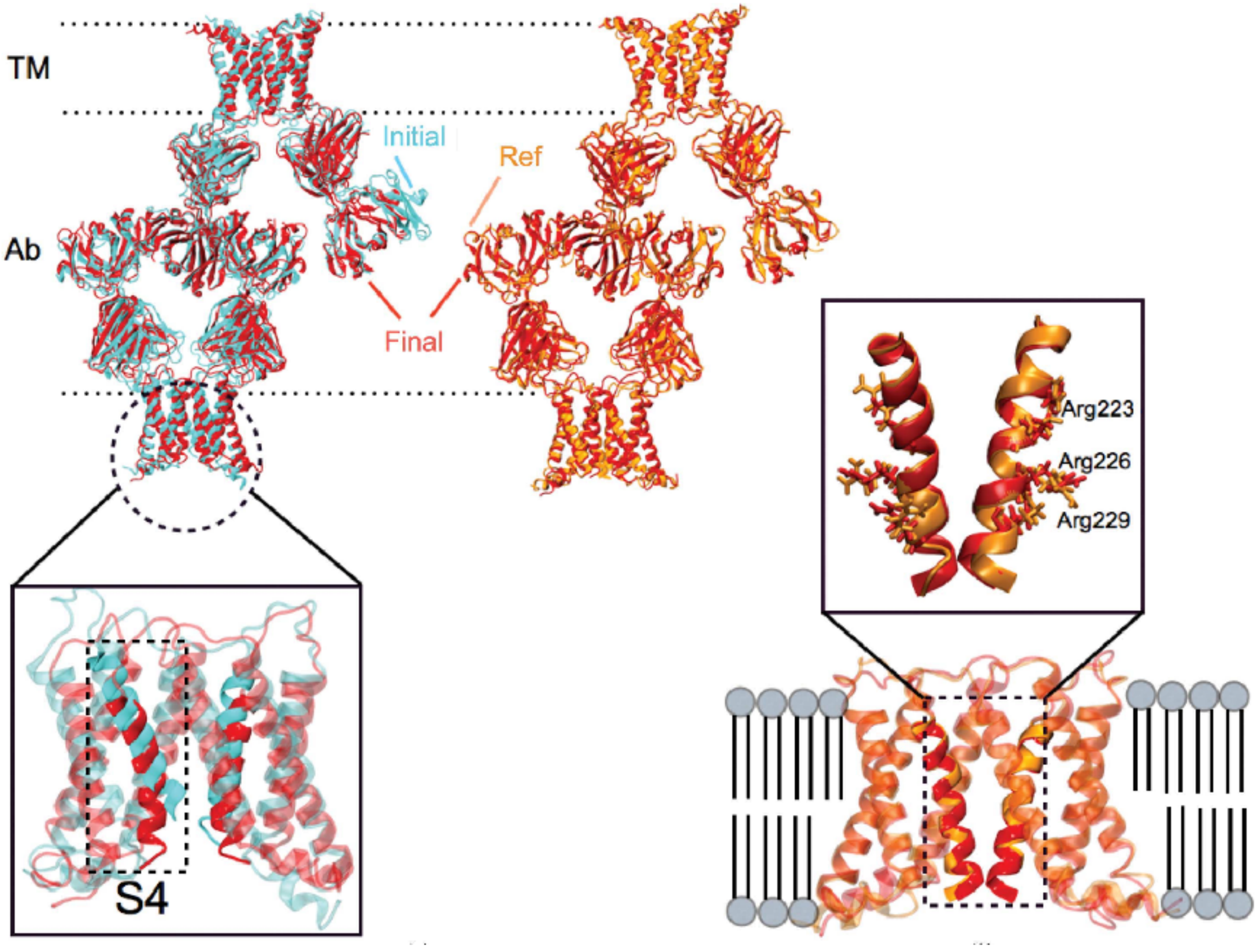




Deformation of the phasing model

xMDFF: Molecular dynamics flexible fitting of low-resolution X-Ray structures.R McGreevy*, A Singharoy*, Q Li, J Zhang, D Xu, E Perozo, K Schulten. *Acta Crystallographica D* **70** 2344-2355, 2014.

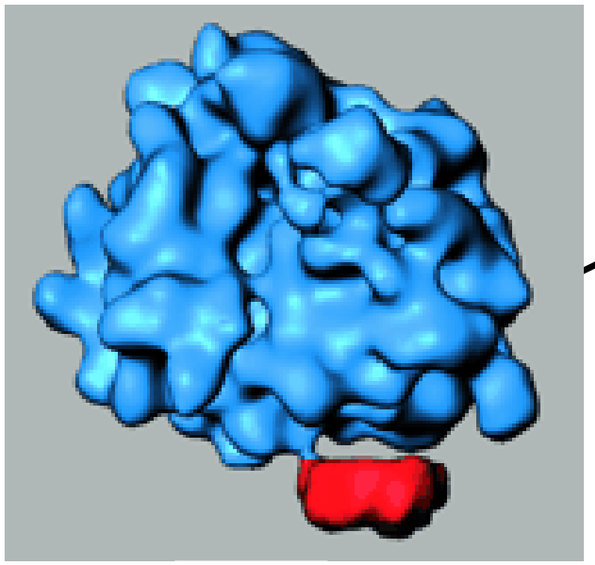
xMDFF: fitting for low resolution X-ray structures

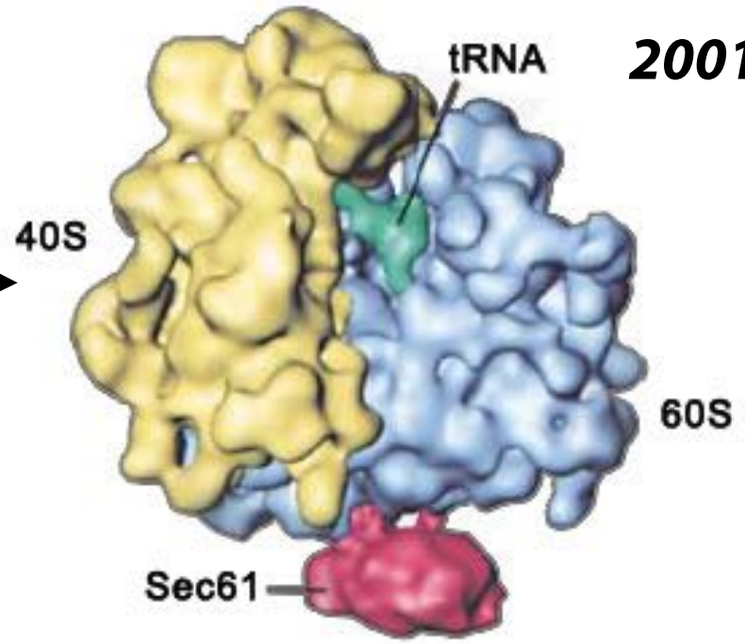


xMDFF: Molecular dynamics flexible fitting of low-resolution X-Ray structures.R McGreevy*, A Singharoy*, Q Li, J Zhang, D Xu, E Perozo, K Schulten. *Acta Crystallographica D* **70** 2344-2355, 2014.

Structures of ribosome-Sec complexes

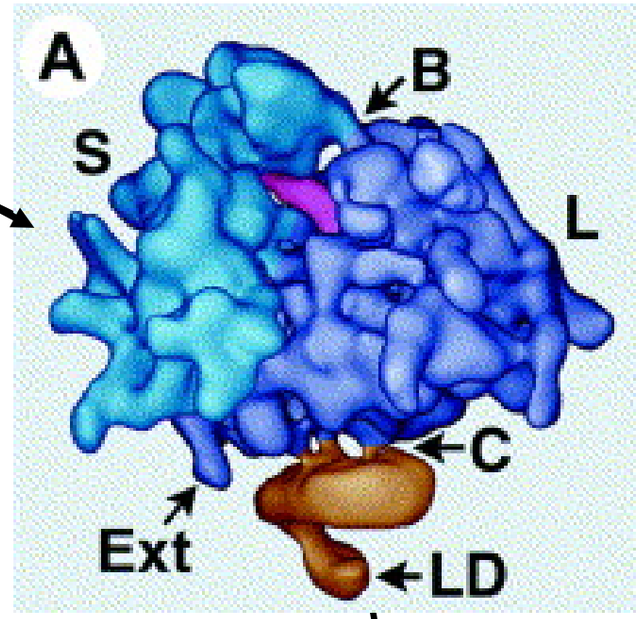
1997[1]





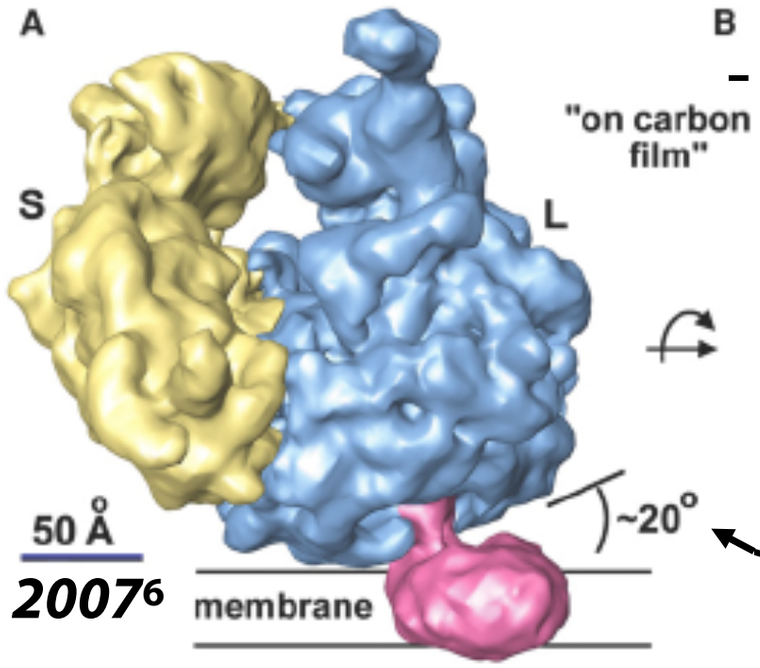
2001[2]

2002[3]

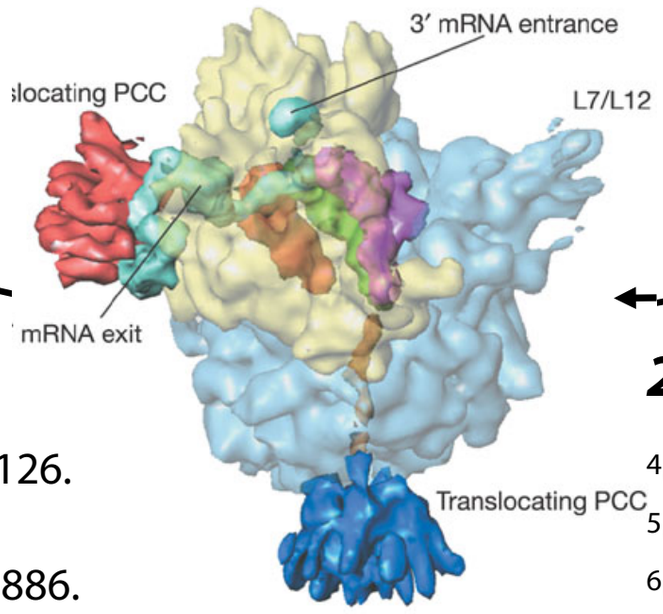


-cryo EM maps of complexes

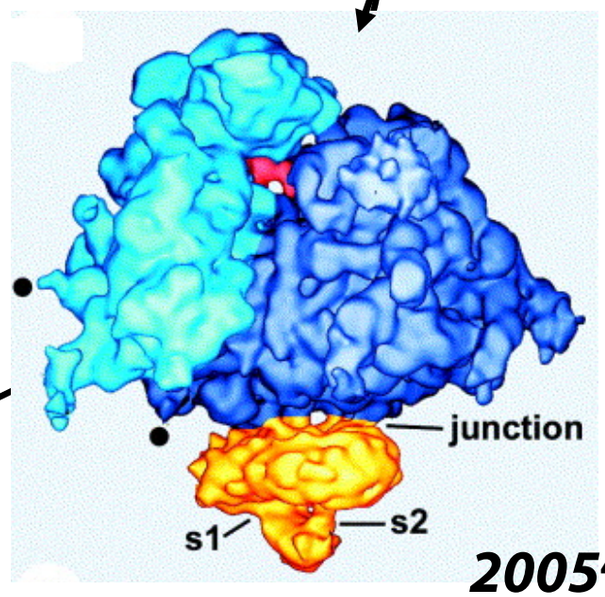
- How many copies of SecY/ Sec61 in each?



2007[6]



2005[5]

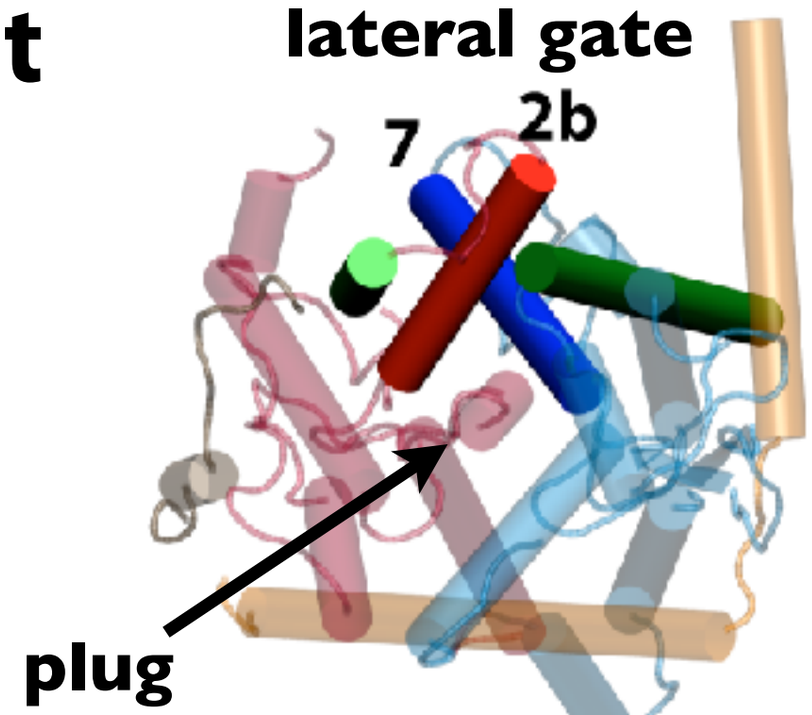


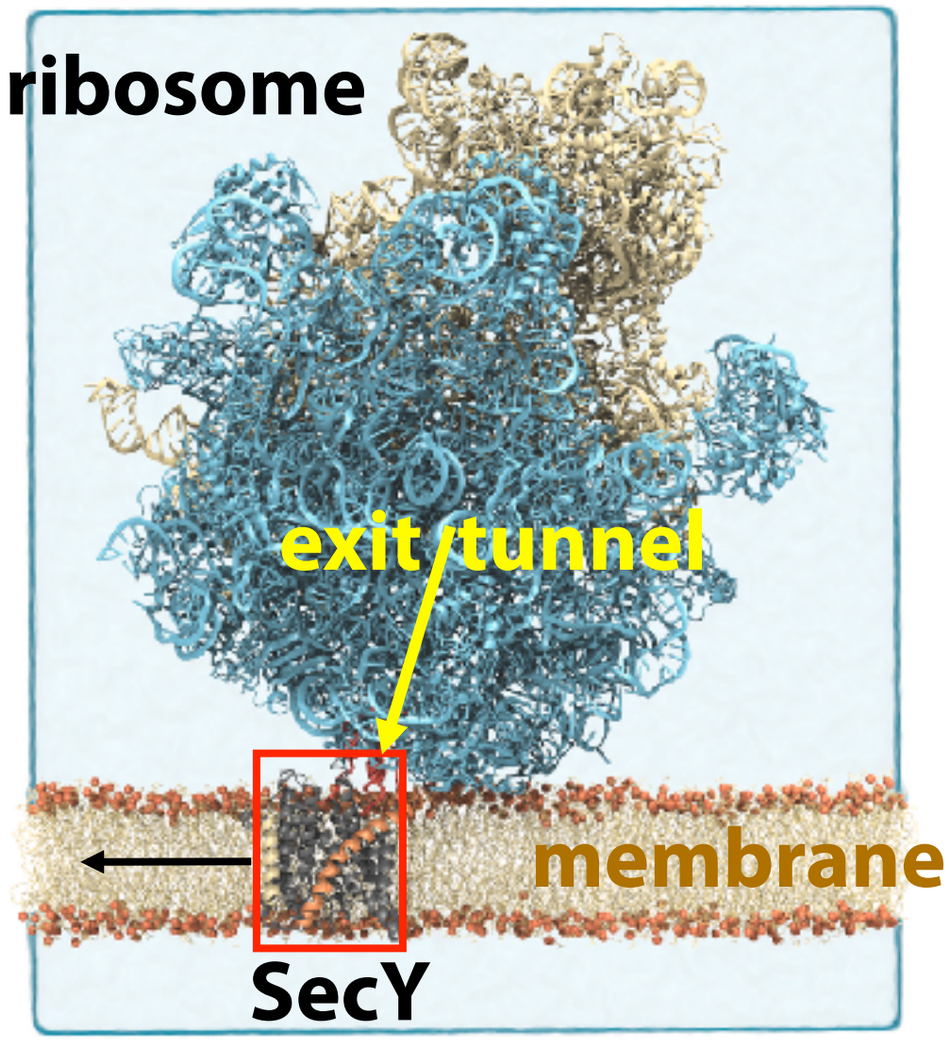
2005[4]

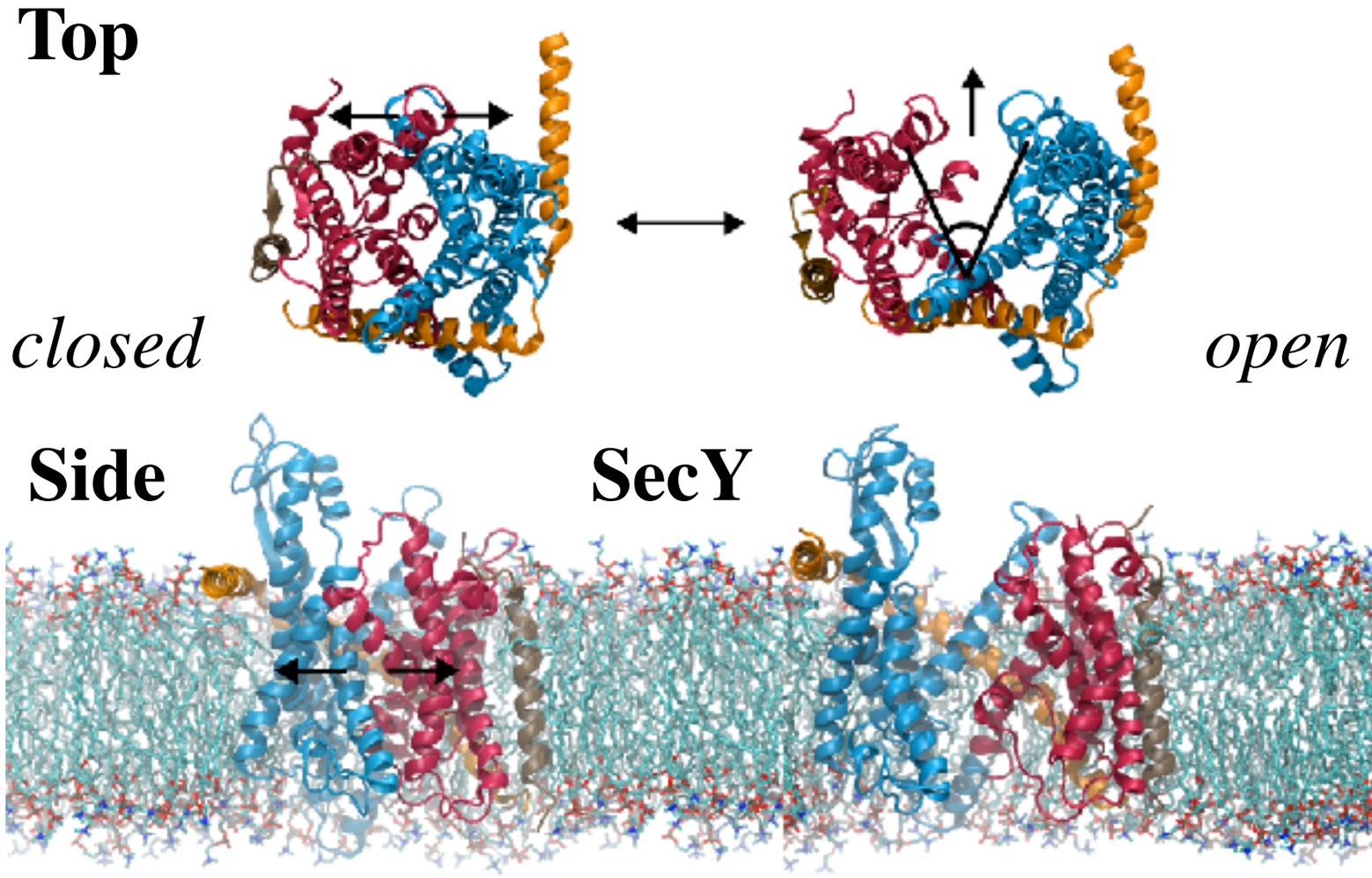
[1]Beckmann, R. *et al*. (1997) *Science* **278**:2123-2126.
[2]Beckmann, R. *et al*. (2001) *Cell* **107**:361-372.
[3]Morgan, D.G. *et al*. (2002) *J. Mol. Biol*. **324**:871-886.
[4]Menetret, J.F. et al. (2005) *J. Mol. Biol*. **348**:445-457.
[5]Mitra, K. *et al*. (2005) *Nature* **438**:318-324.
[6]Menetret, J.F. et al. (2007) *Mol. Cell* **28**:1083-1092.

(Inner) membrane proteins insert through SecY

-Ribosome feeds nascent protein into SecY (95% of all MPs in eukaryotes)
-Membrane protein segments exit through SecY's lateral gate

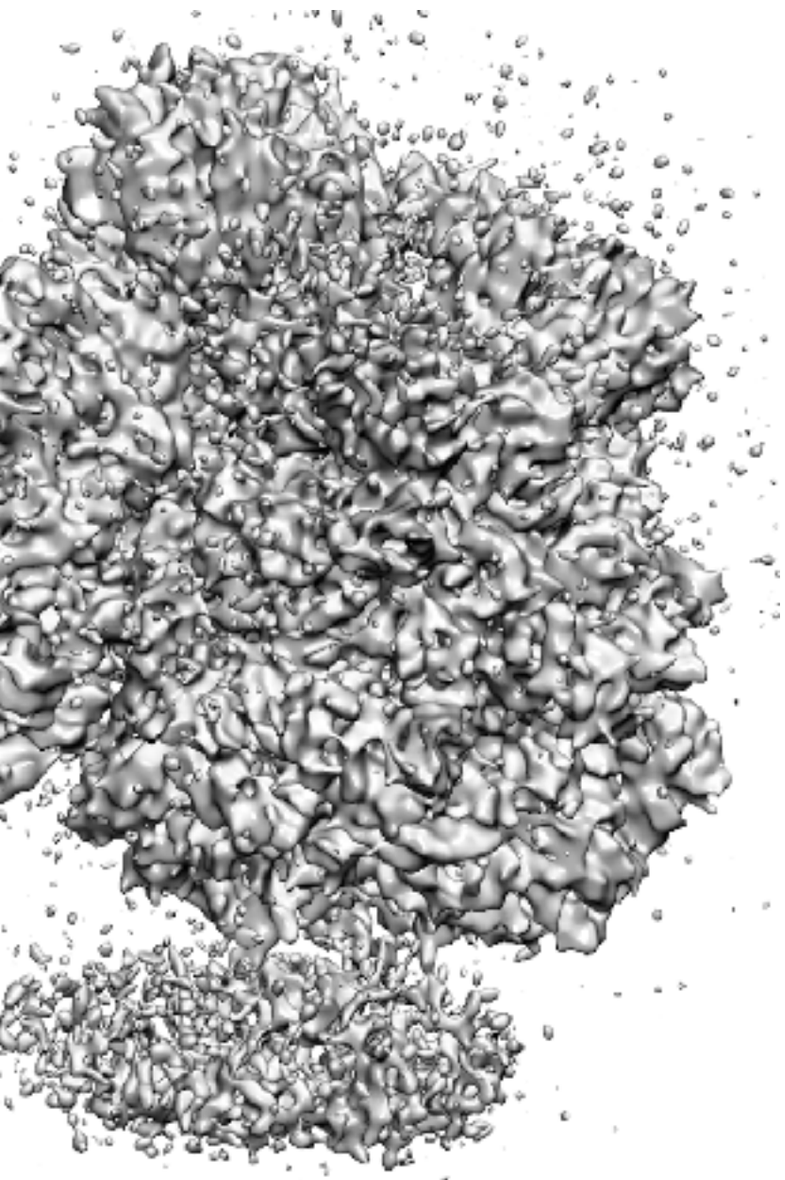


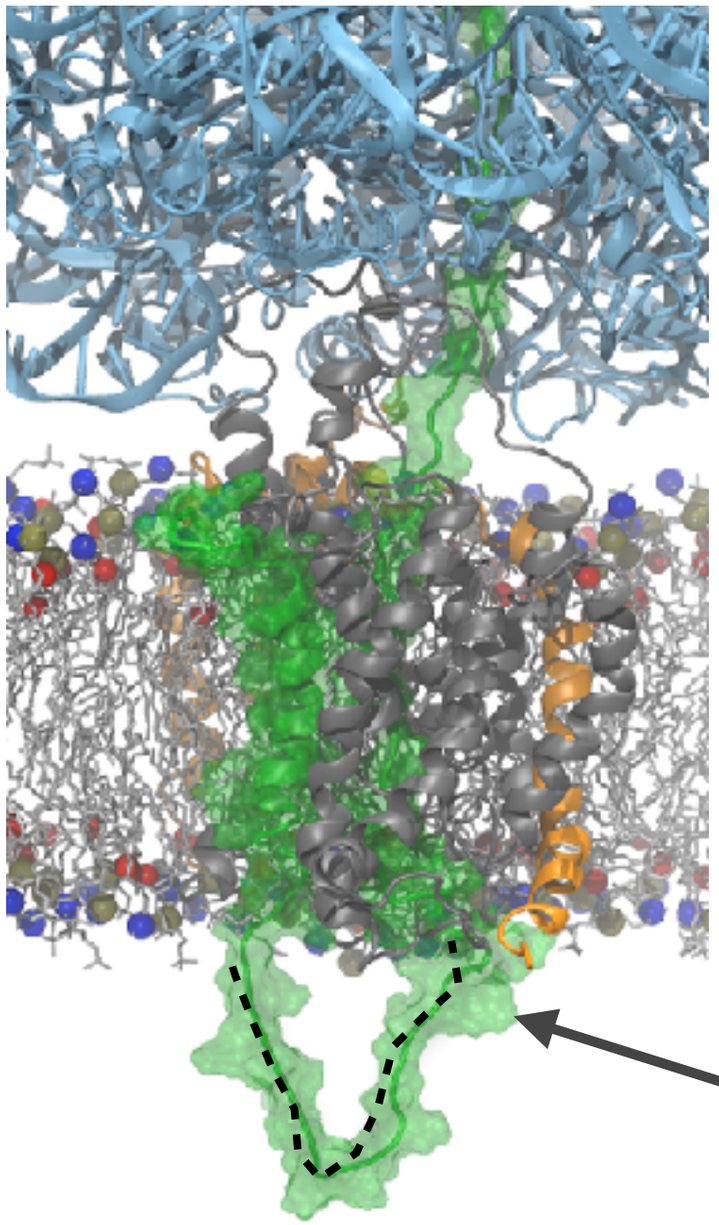




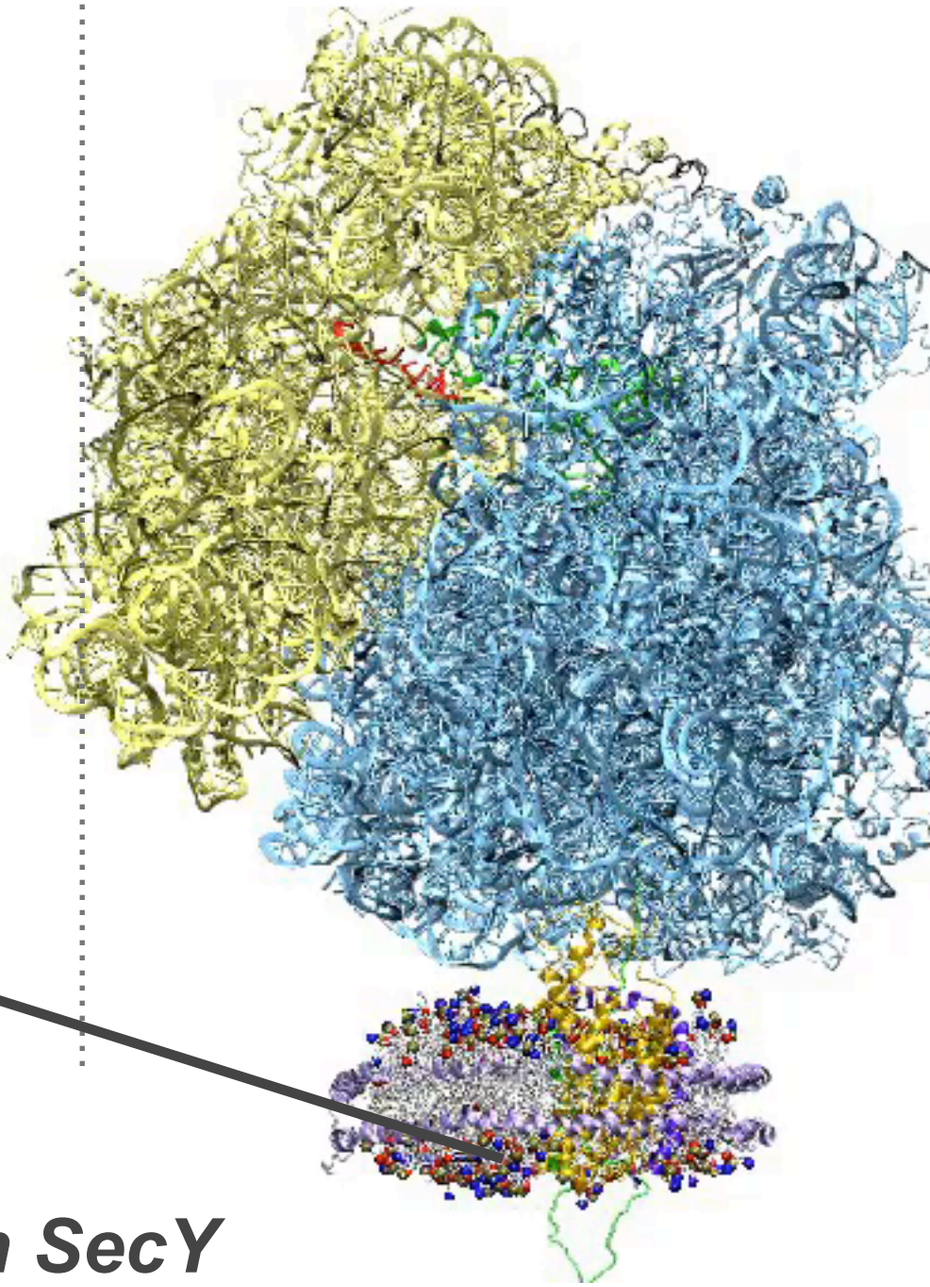
Membrane insertion seen at atomic resolution (circa 2011)

Low-resolution Data





High-resolution Structure

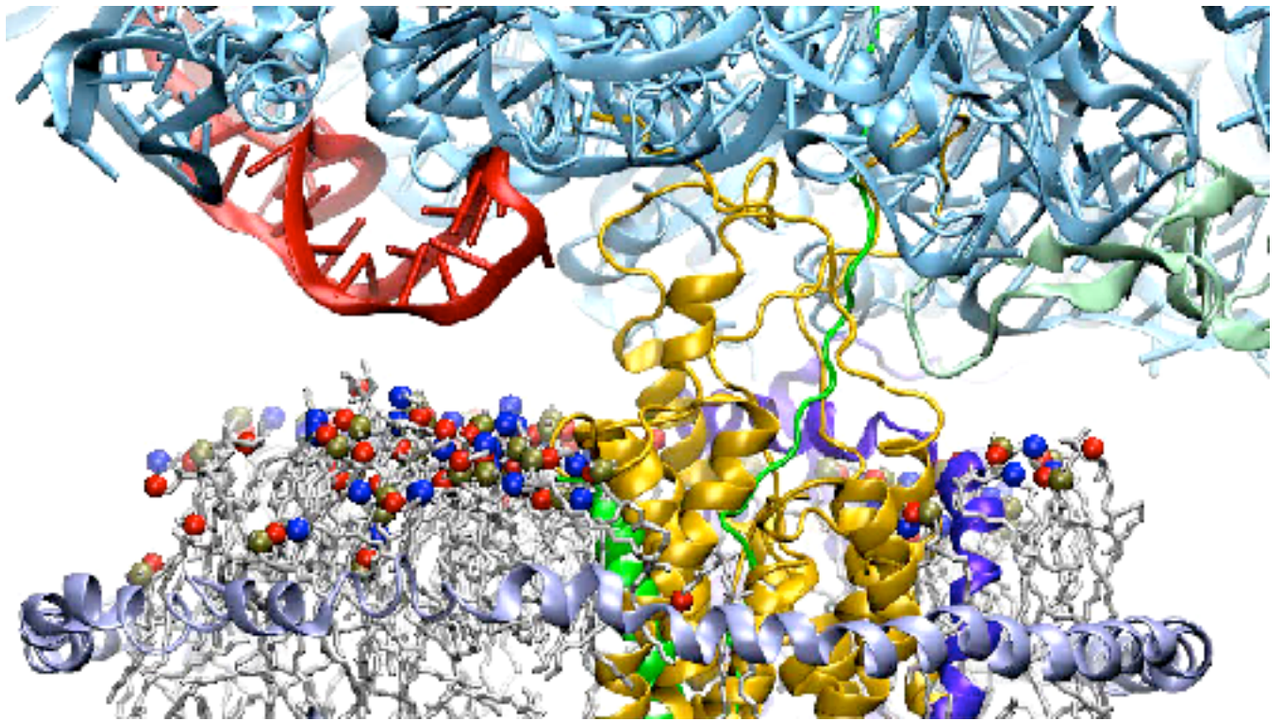


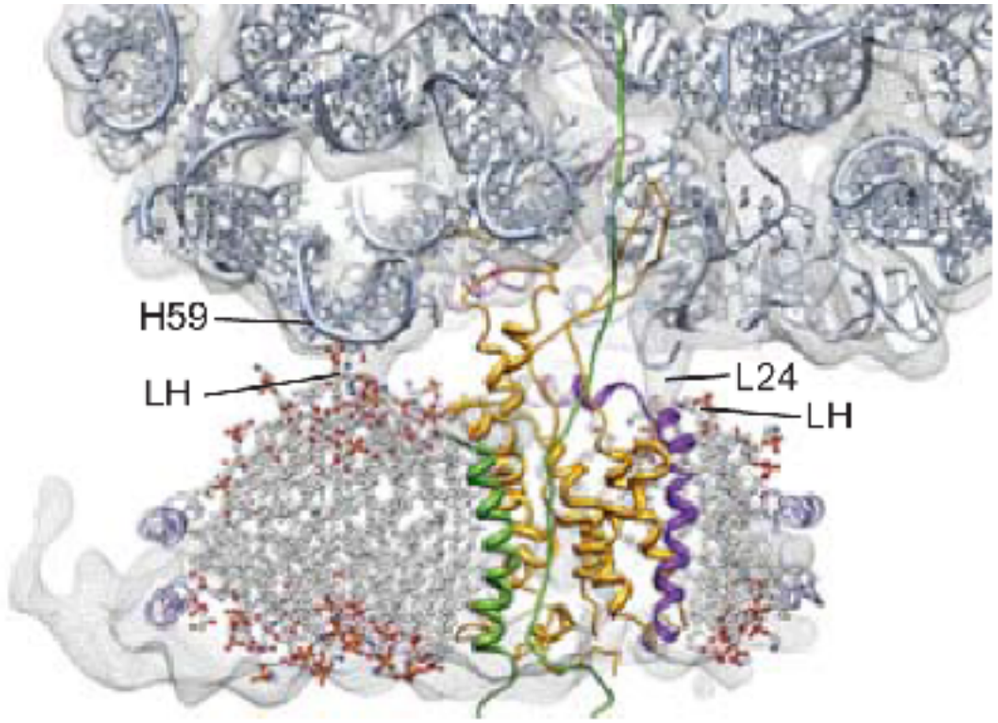
Close-up of Nascent Protein in SecY

J. Frauenfeld, J. Gumbart *et al*. (2011) *Nat. Struct. Mol. Bio.* 18:614-621.

***collaboration with cryo-EM lab of Roland Beckmann**

Structure of a ribosome-nascent-chain complex



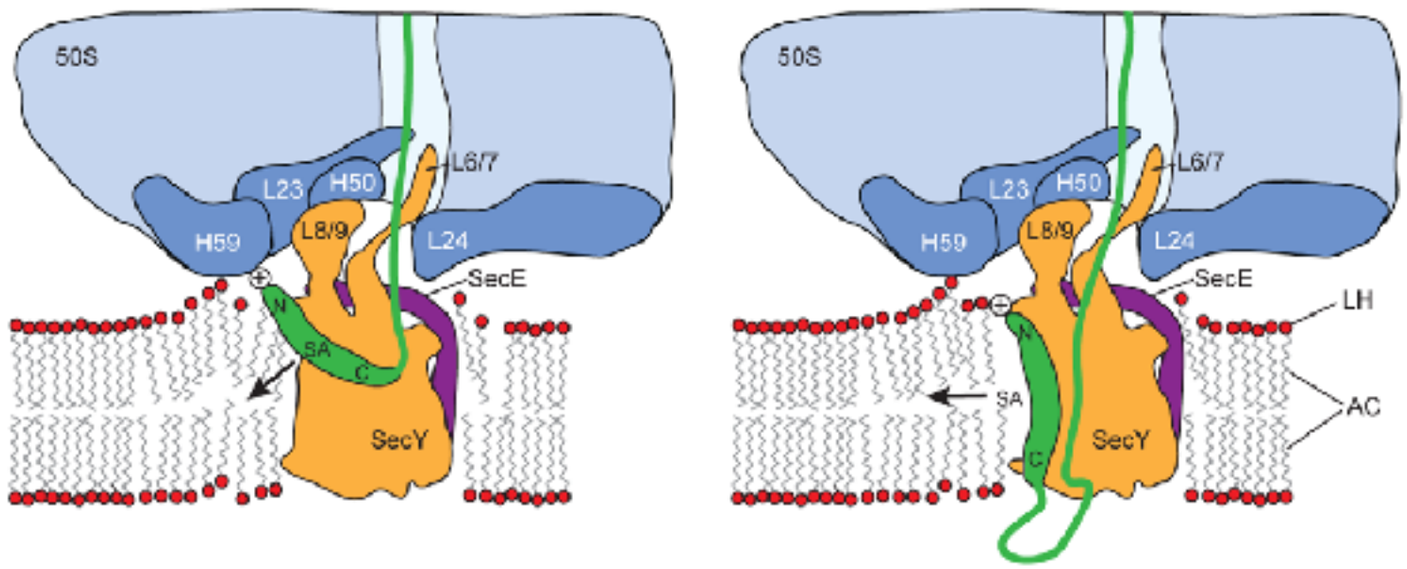


Unbiased simulation reveals spontaneous attraction of **lipids** to **H59** of ribosome's 23S...

...which explains ambiguous density observed in map!

-**H59** aids the proper orientation of the **signal anchor**

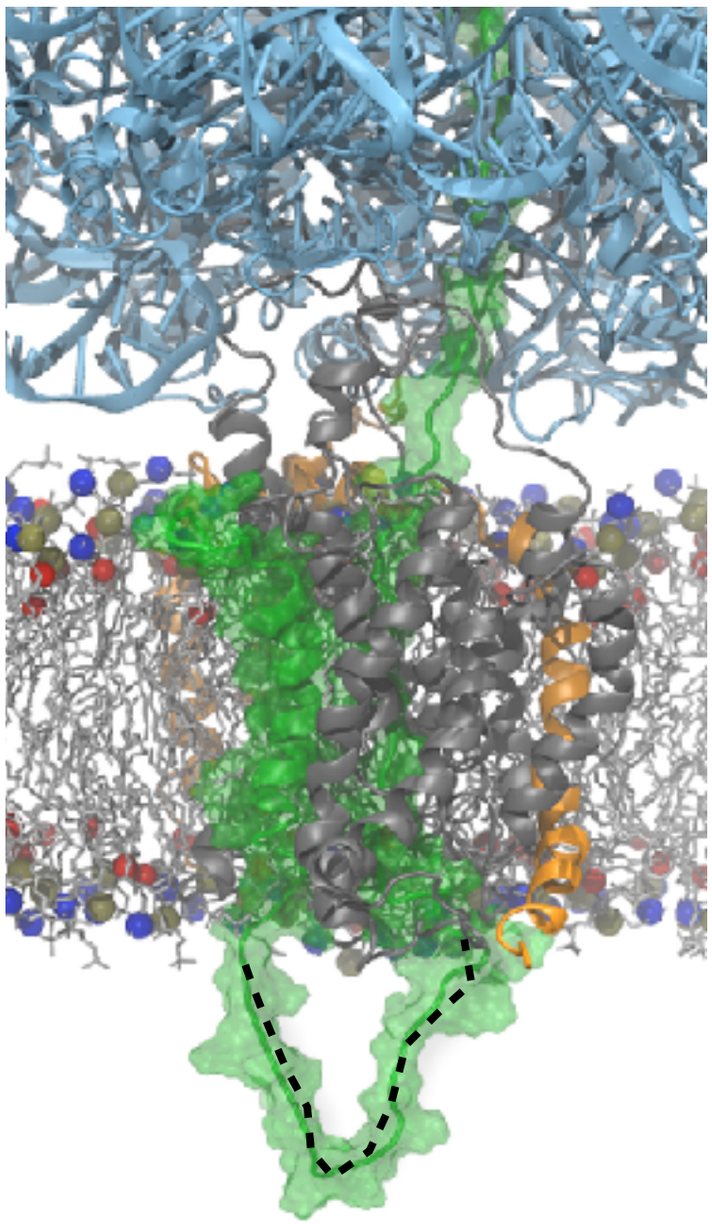
-**H59** also perturbs **lipids**, permitting easier access of the **signal anchor** to the bilayer

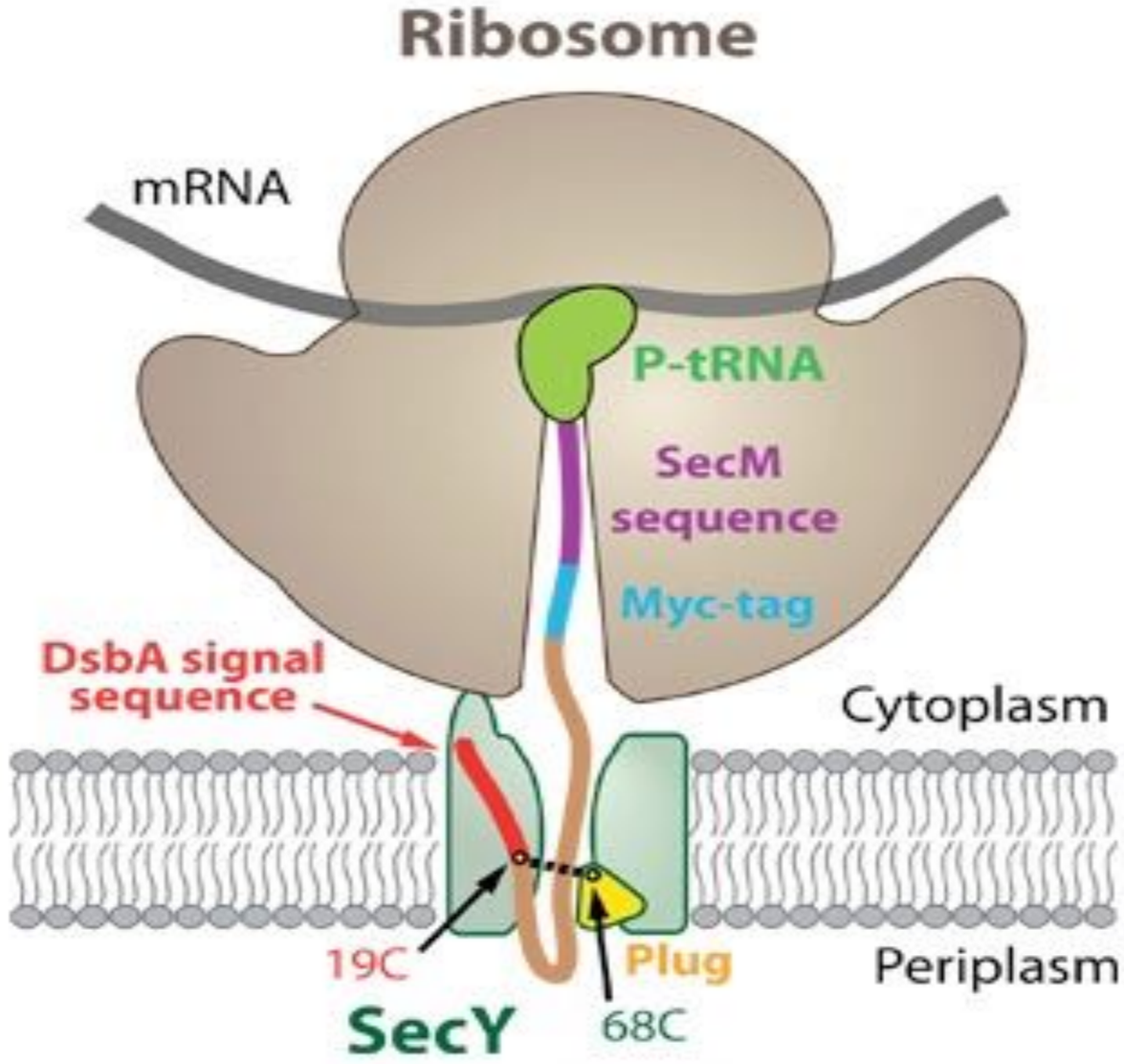


J. Frauenfeld, J. Gumbart *et al*. (2011) *Nat. Struct. Mol. Bio.* 18:614-621.

Membrane insertion seen at atomic resolution *(the remake)*

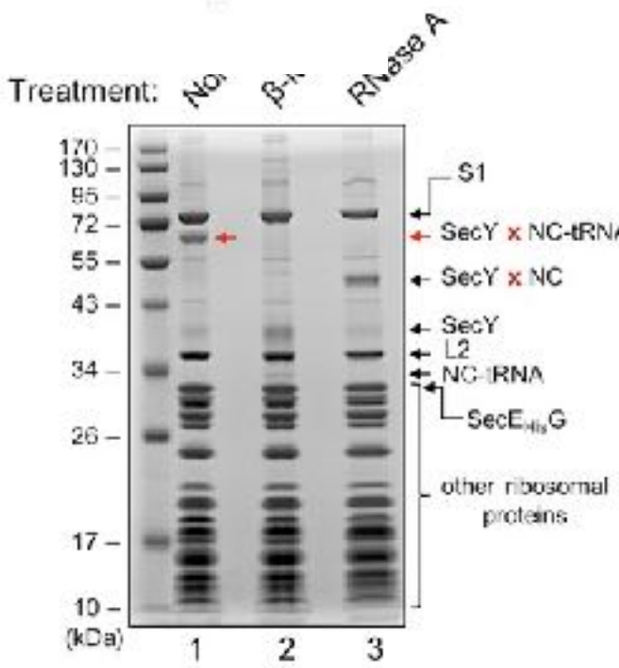
is the signal anchor where we think it is?





New construct made in **Rapoport lab** (Harvard U.) has disulfide cross-link between plug and nascent chain

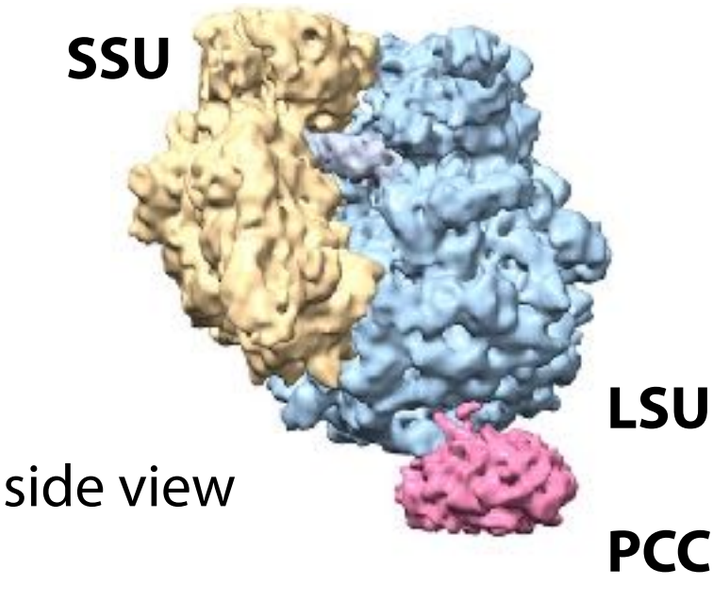
link confirmed *biochemically*



E. Park, J.F. Menetret, J. Gumbart S.J. Ludtke, W. Li, A. Whynot, T.A.Rapoport, C.W. Akey. (2014) *Nature.* **506**:102-106.

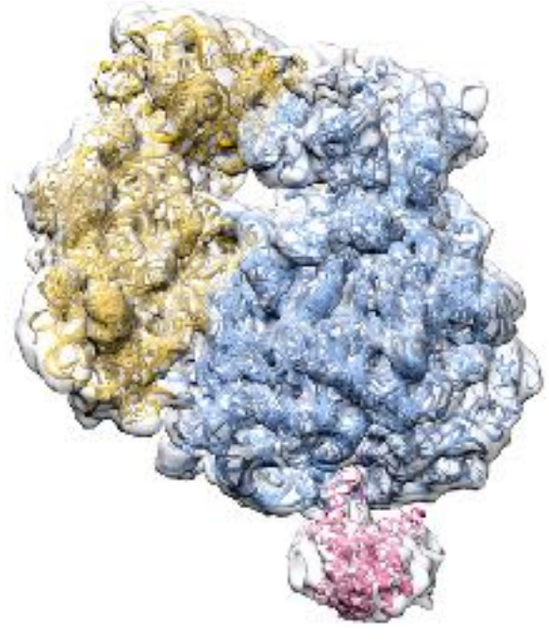
Cryo-EM visualization of cross-linked state

resolution 10.1 Å



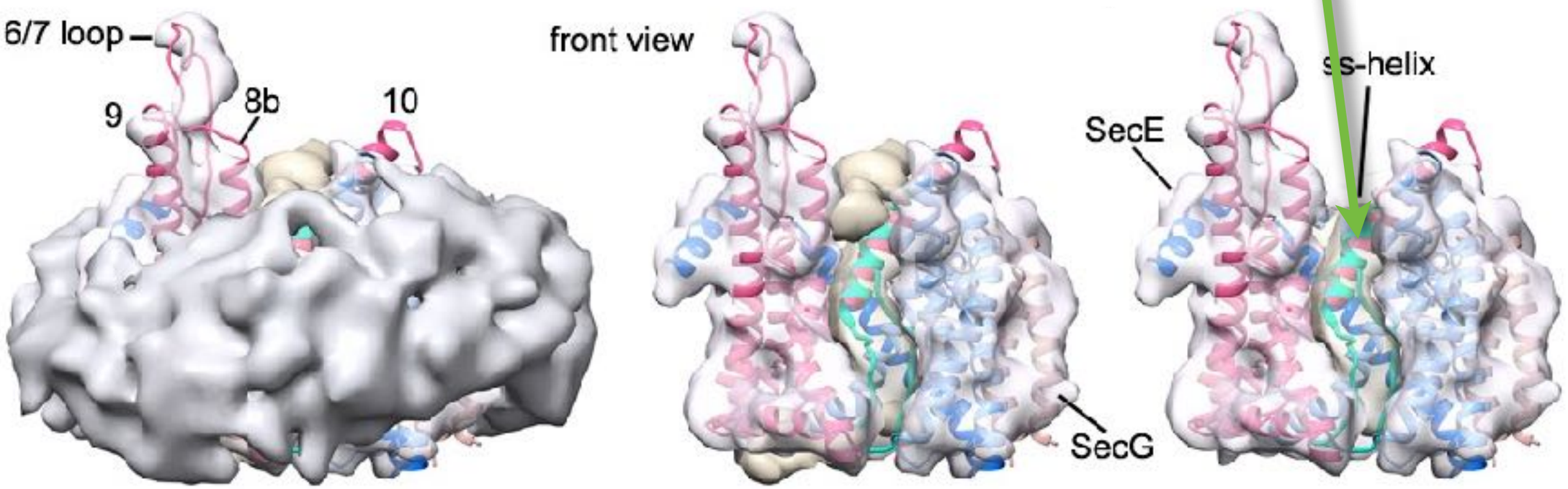
Empty complex also visualized

resolution 9.5 Å



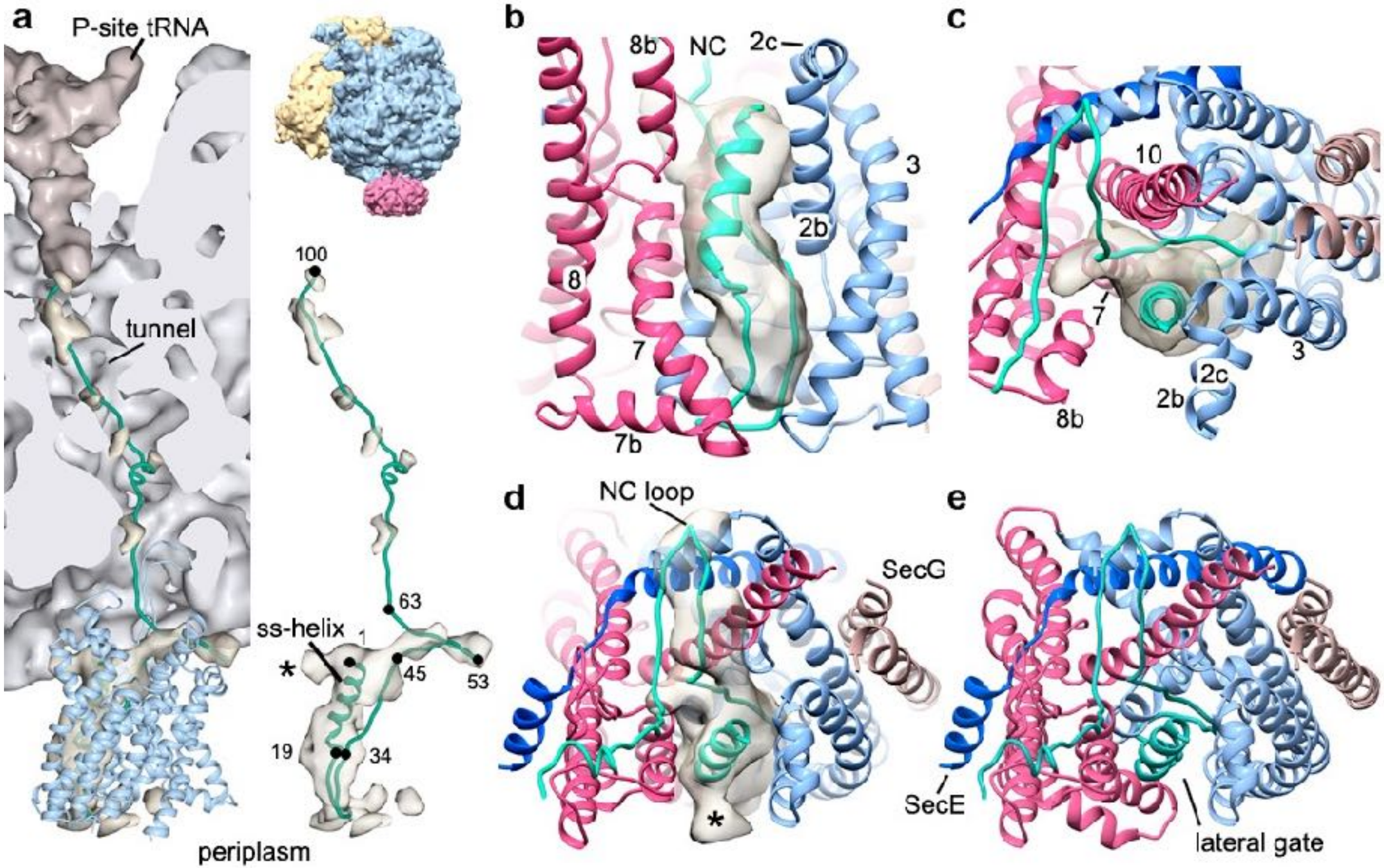
maps by Chris Akey, Boston U.

Segmented channel density with signal anchor



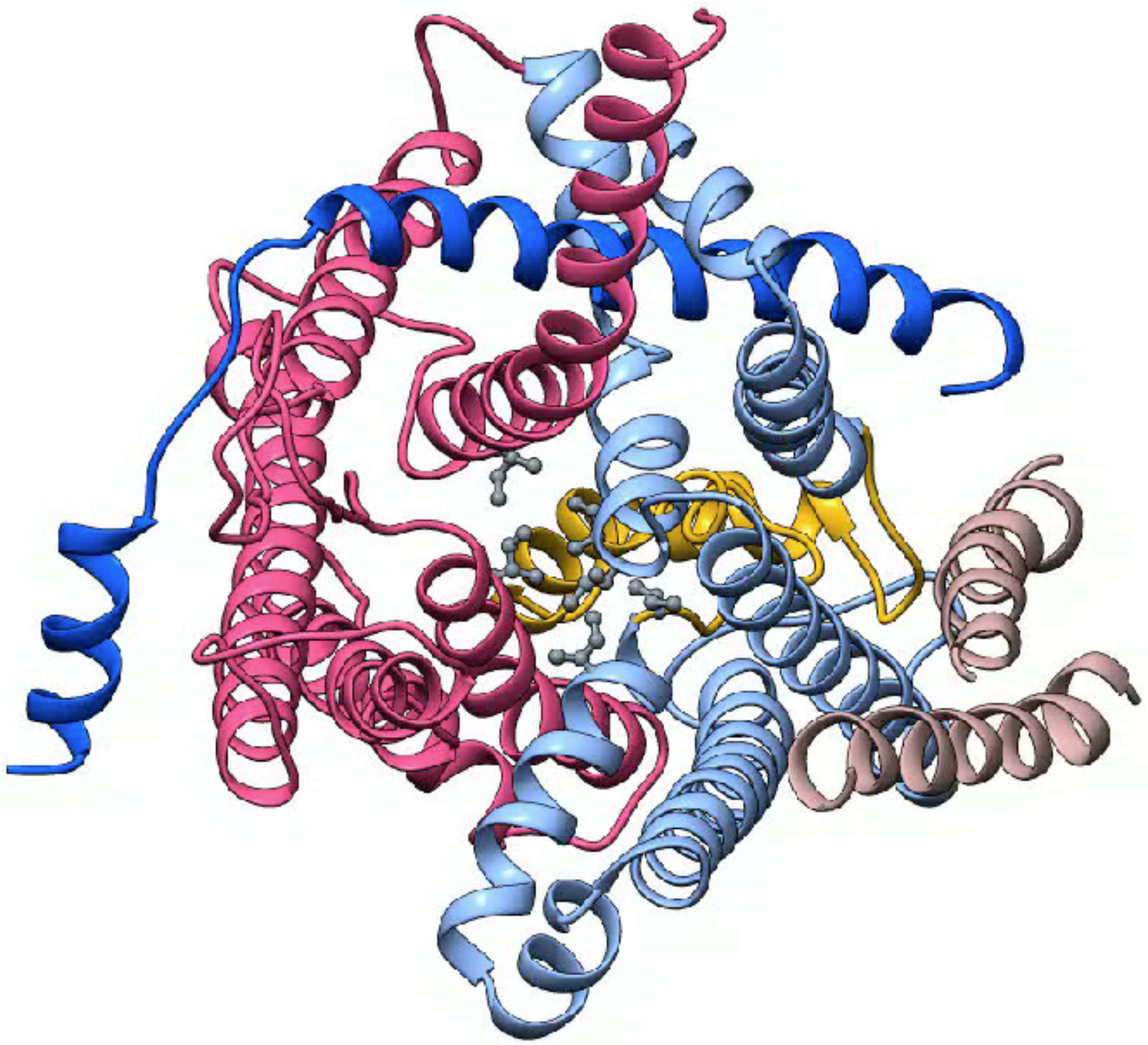
E. Park, J.F. Menetret, J. Gumbart S.J. Ludtke, W. Li, A. Whynot, T.A.Rapoport, C.W. Akey. (2014) *Nature.* **506**:102-106.

tracing the path of the nascent protein



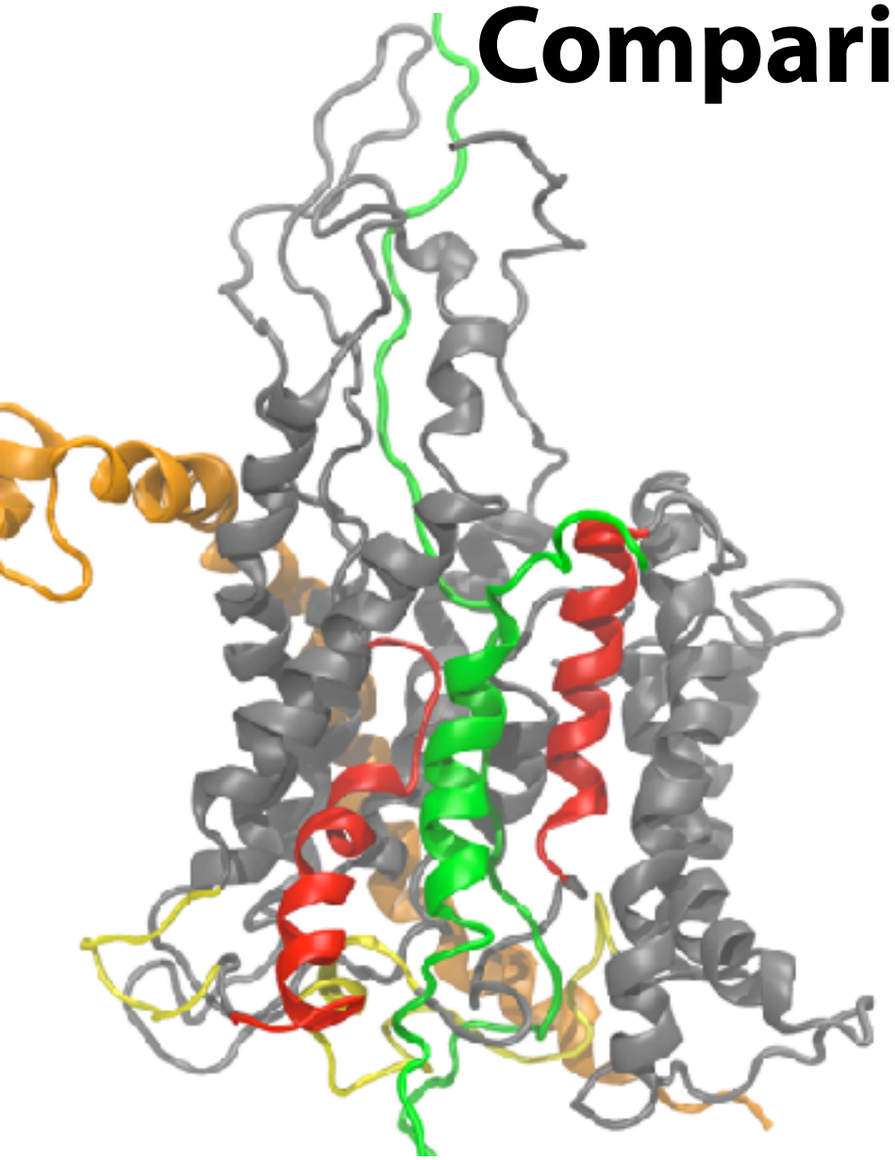
E. Park, J.F. Menetret, J. Gumbart S.J. Ludtke, W. Li, A. Whynot, T.A.Rapoport, C.W. Akey. (2014) *Nature.* **506**:102-106.

Gate opening requires only rigid-body motions

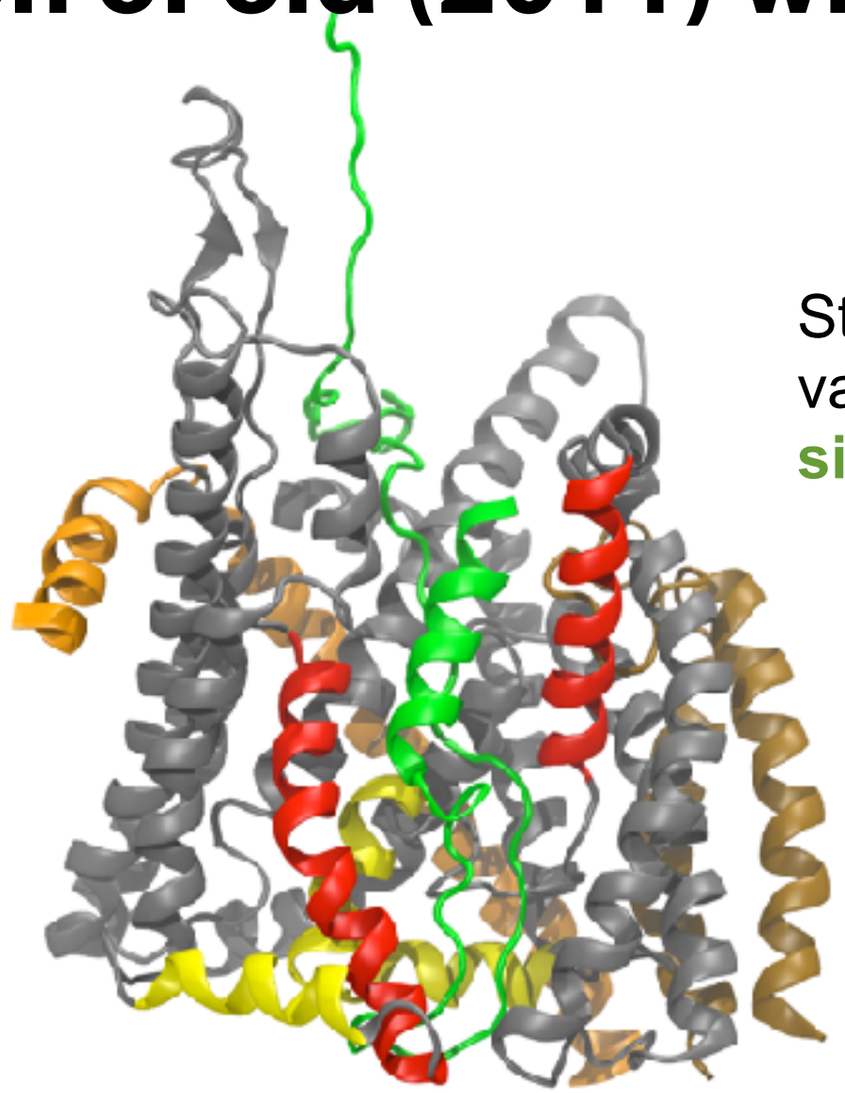


E. Park, J.F. Menetret, J. Gumbart S.J. Ludtke, W. Li, A. Whynot, T.A.Rapoport, C.W. Akey. (2014) *Nature.* **506**:102-106.

Comparison of old (2011) with new (2014)

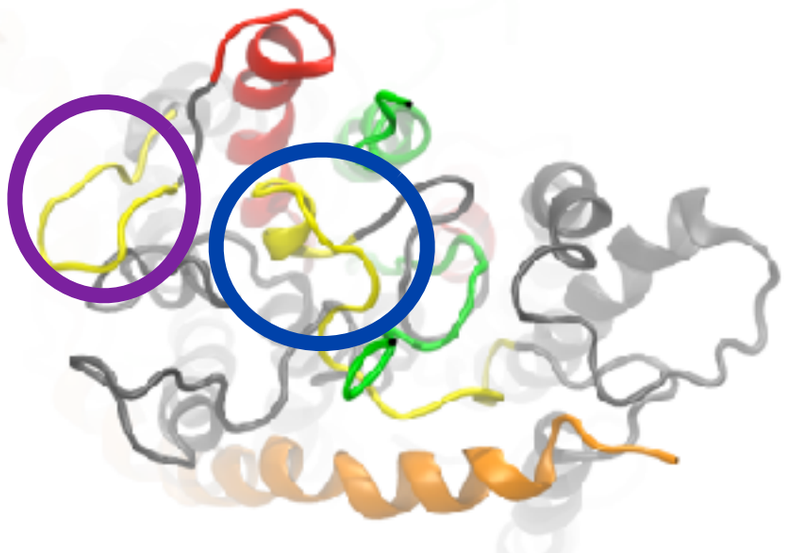


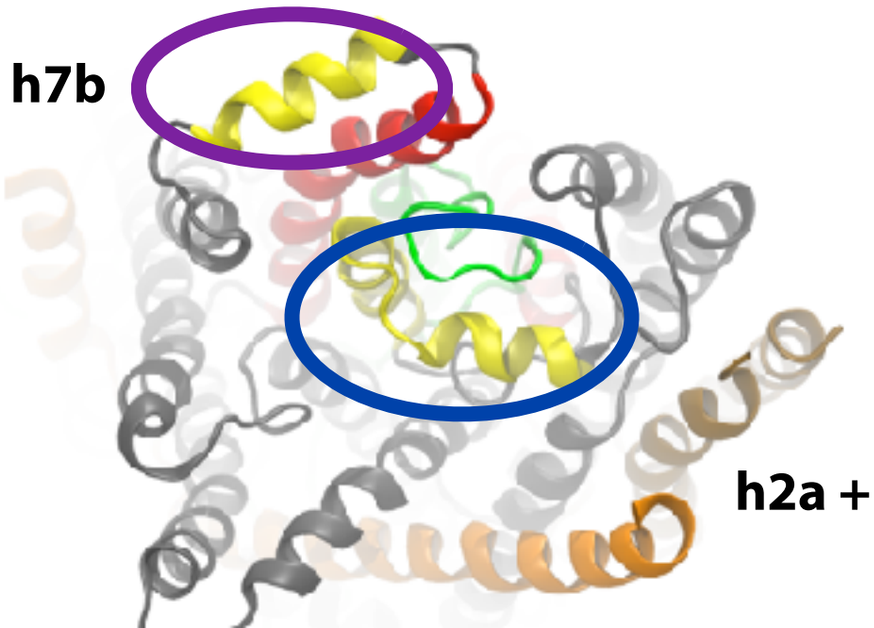
Frauenfeld et al. 2011



Park et al. 2014

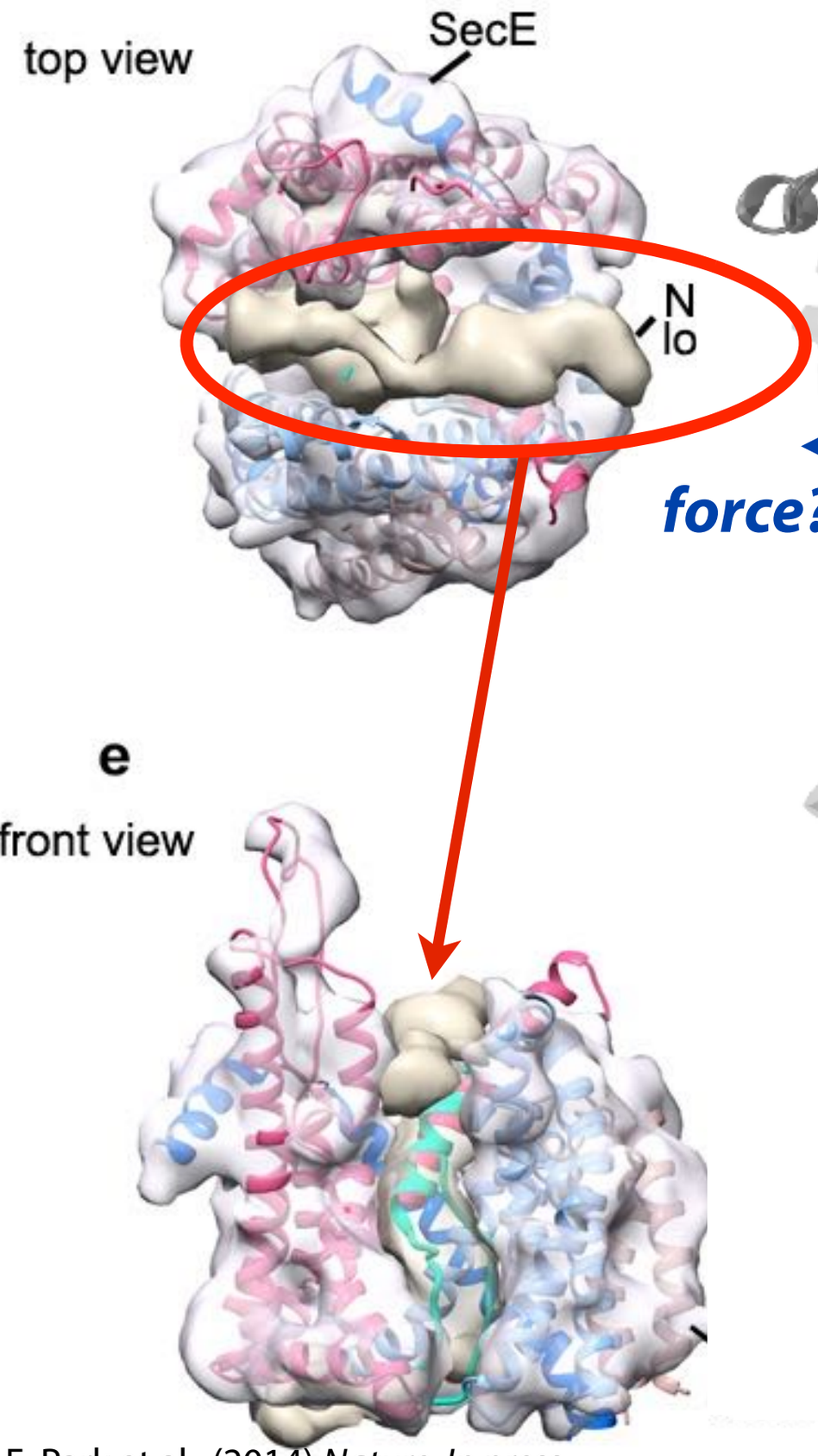
Structures roughly similar, validating placement of **signal anchor** in first one

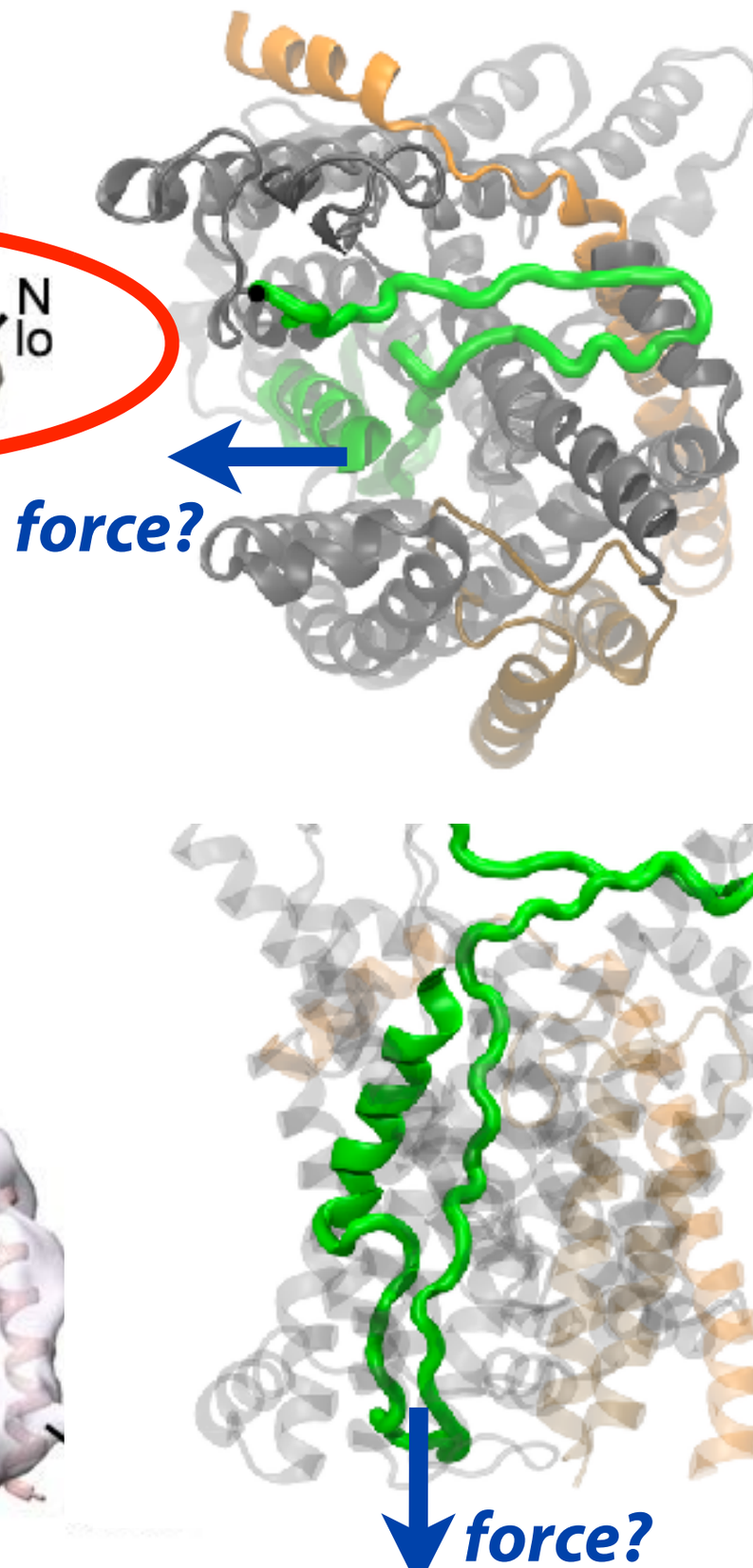




Largest differences are in improved modeling of *E. coli*-specific elements, e.g., **two-helix plug (h2a+h2z)** and **linker h7b** between TMs 7 and 8

The nascent chain is not fully extended





Nascent chain density is observed in the open pore and in a **V-shaped cleft** on top of SecY

force from peptide synthesis insufficient to drive translocation

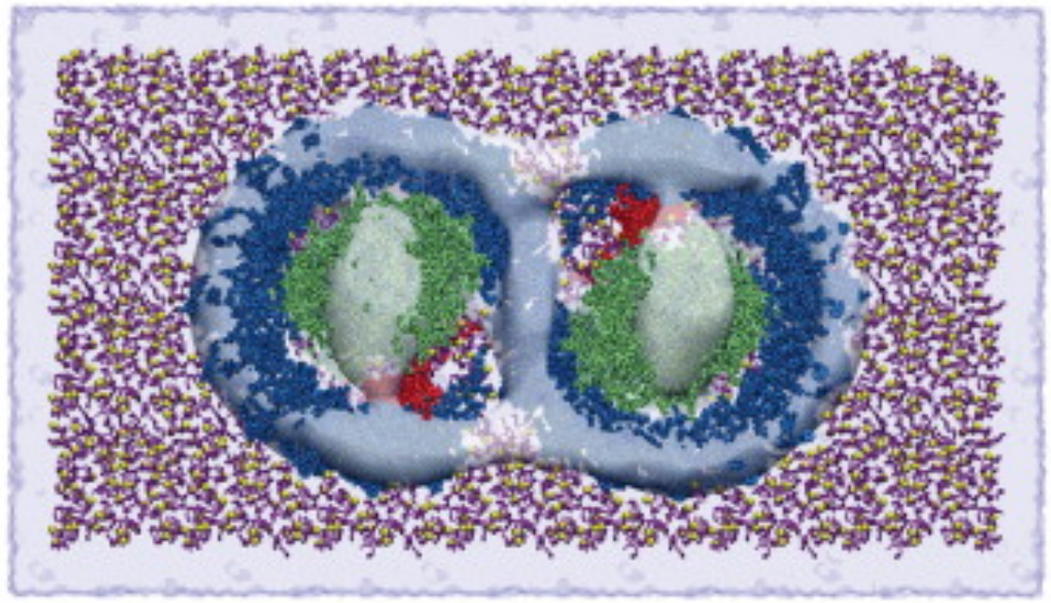
Translocation through the channel may require a **pulling force** from the other side (e.g., SecDF)

Membrane insertion has been demonstrated to exert a force on the nascent chain[1]

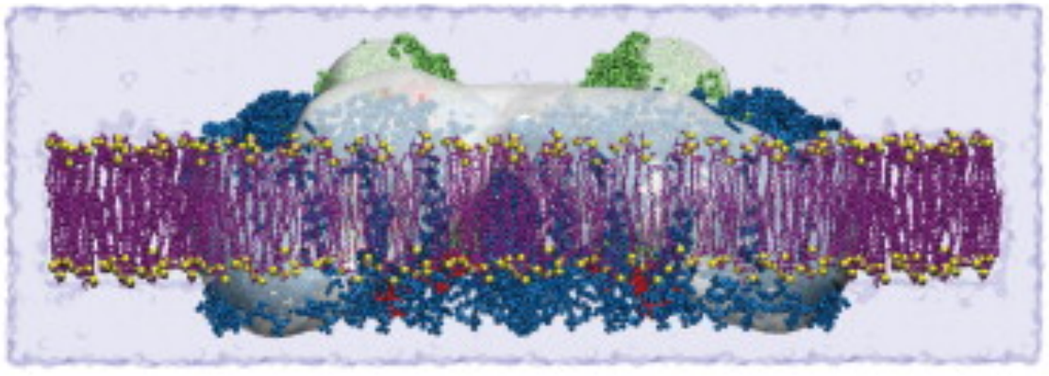
E. Park et al. (2014) *Nature. In press.*

[1] N. Ismail...G. von Heijne. A biphasic pulling force acts on transmembrane helices during translocon-mediated membrane insertion. (2012) *Nat. Struct. Mol. Bio.* 10:1018-1022.

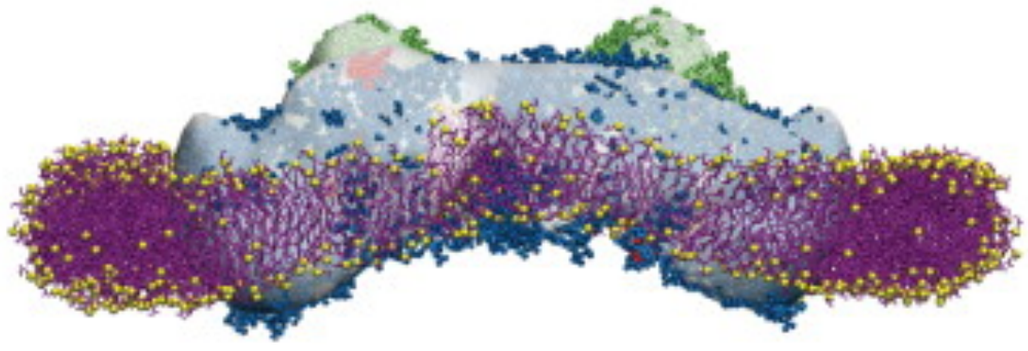
Membrane curvature induced by LH1



cryo-**EM** map (25-Å resolution) showed a bent structure (grey outline) in agreement with previous simulations



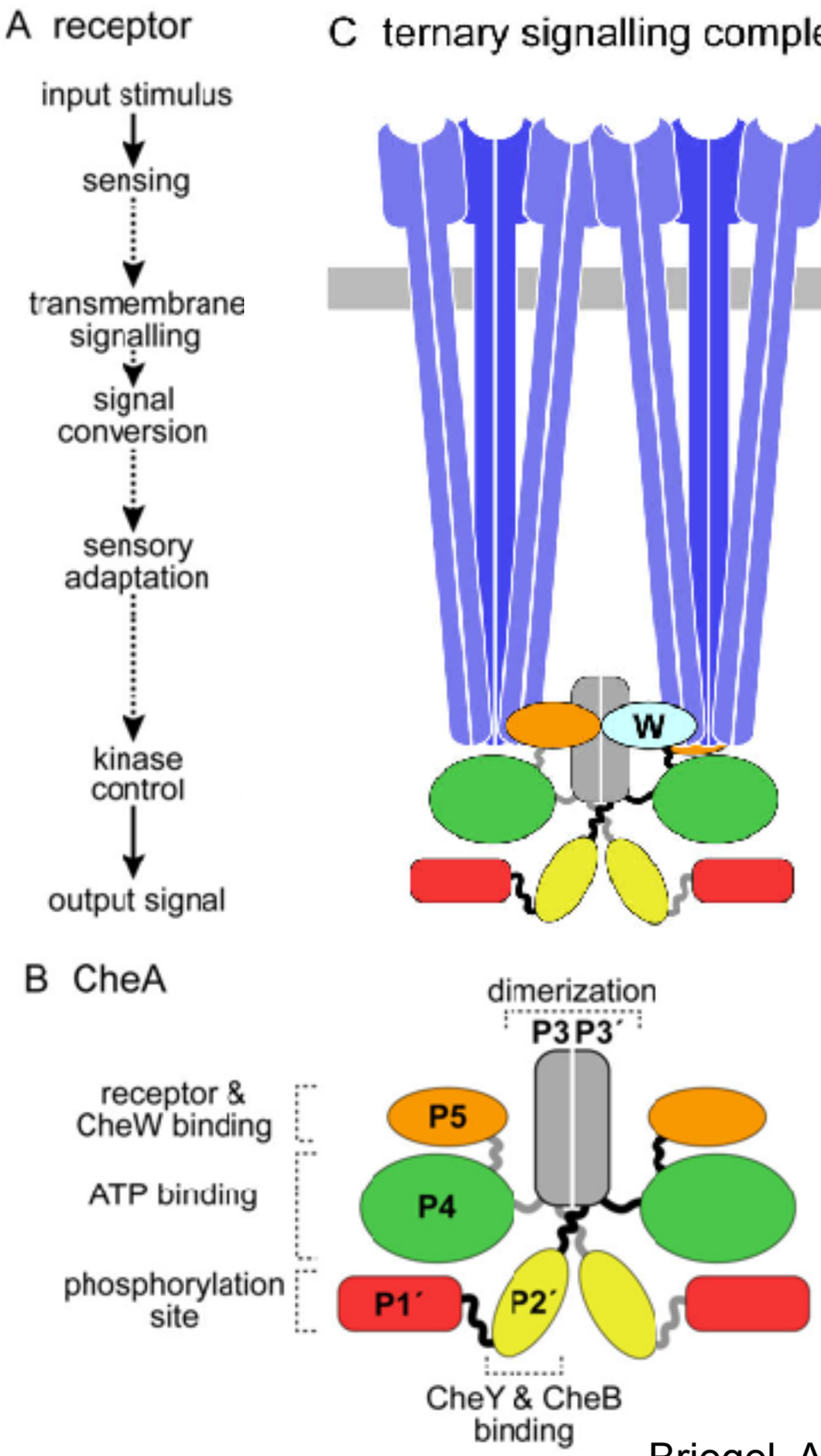
Molecular dynamics flexible fitting of LH1 to that map in the ***presence of a membrane*** induced membrane curvature

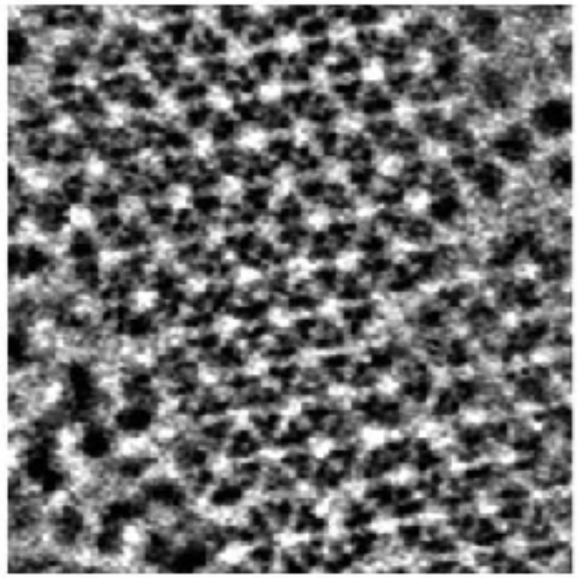


curvature was maintained even after fitting

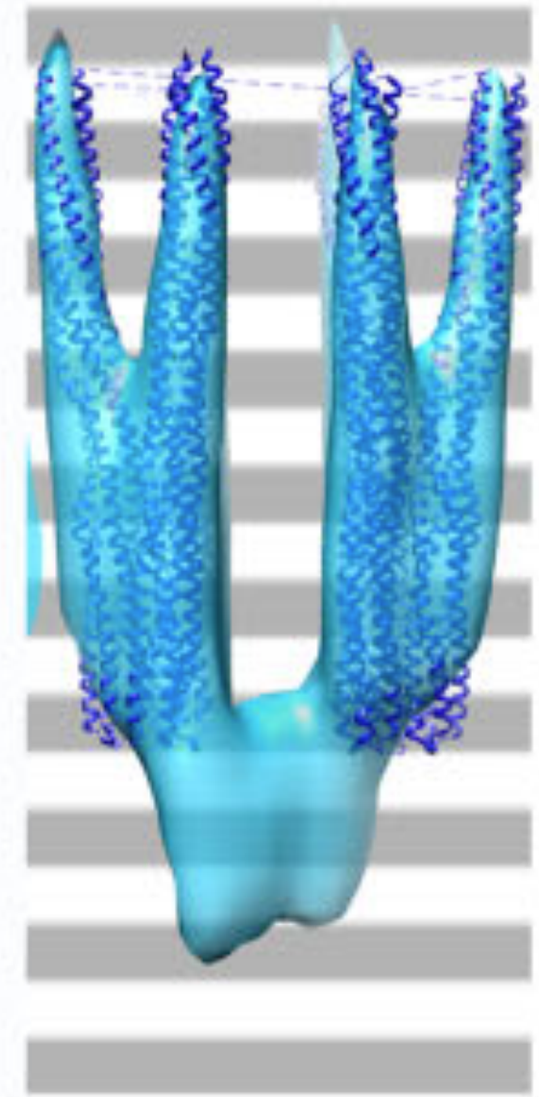
Jen Hsin, James Gumbart, Leonardo G. Trabuco, Elizabeth Villa, Pu Qian, C. Neil Hunter, and Klaus Schulten. Protein-induced membrane curvature investigated through molecular dynamics flexible fitting. *Biophys. J.*, 97:321-329, 2009.

MD fitting to *cryo-tomography* data





cryo tomogram of receptor array



MDFF fit of atomic structure to averaged map (>30 Å resolution)

Briegel, Ames, Gumbart *et al*. The mobility of two kinase domains in the *Escherichia coli* chemoreceptor array varies with signalling state (2013) *Mol Microbio.* **89**:831-841.

Where to get more information

The KS group maintains a detailed website with details, examples, and documentation

Over 100 published applications in last decade - look for ones similar to the problem or system you are interested in!

http://www.ks.uiuc.edu/Research/mdff/index.html