THE UNIVERSITY of TEXAS
HEALTH SCIENCE CENTER AT HOUSTON
SCHOOL of HEALTH INFORMATION SCIENCES

Multi-Scale Modeling

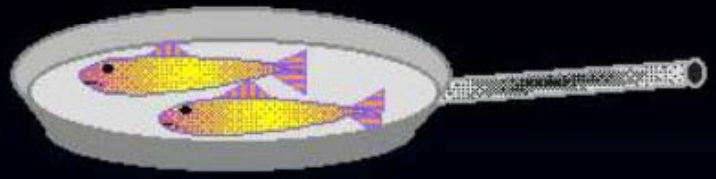
For students of HI 6327 “Biomolecular Modeling”

Willy Wriggers, Ph.D.
School of Health Information Sciences

http://biomachina.org/courses/modeling/13.html

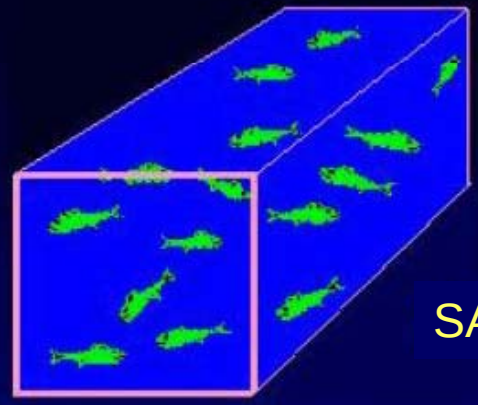
Structural Biology Techniques

crystallography



electron microscopy

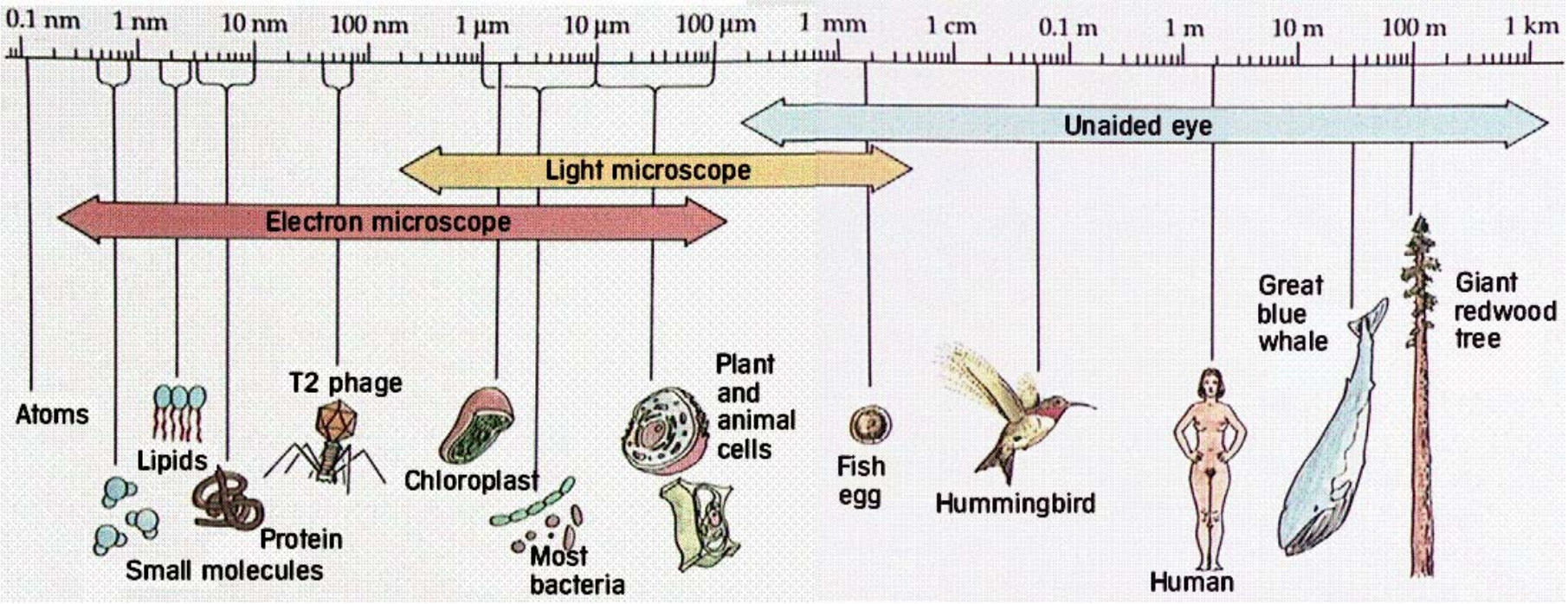
modelling

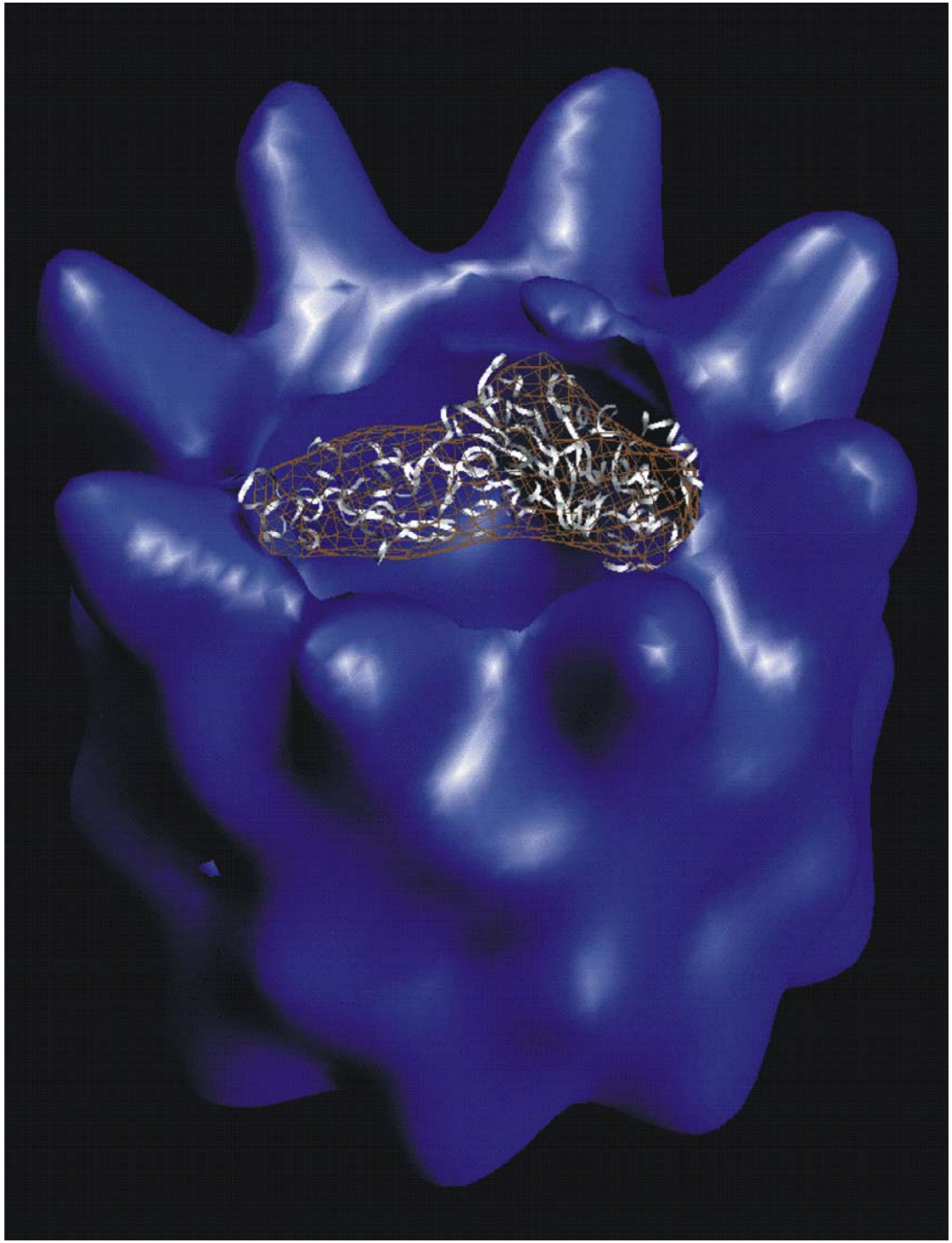


SAXS

ID Campbell 1999

Microscopy Techniques

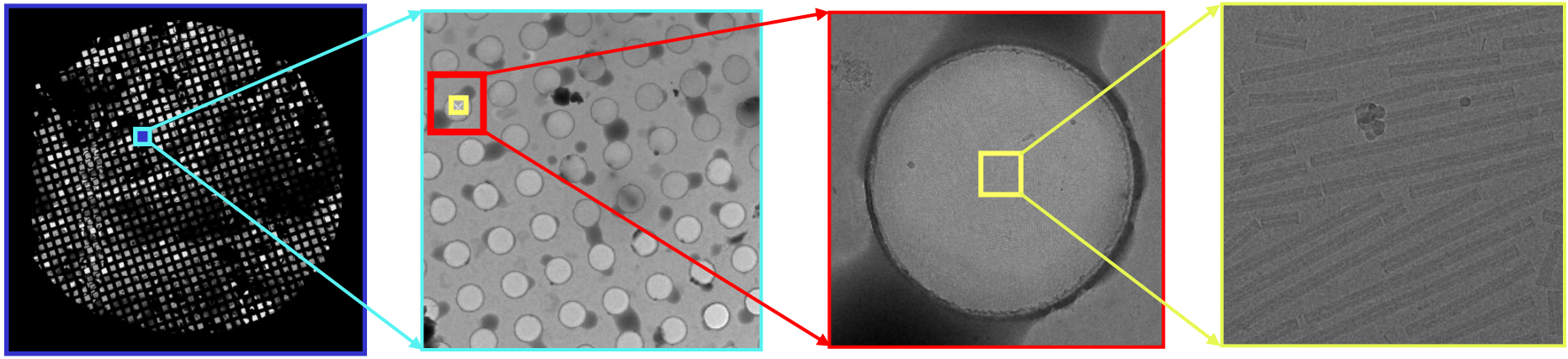




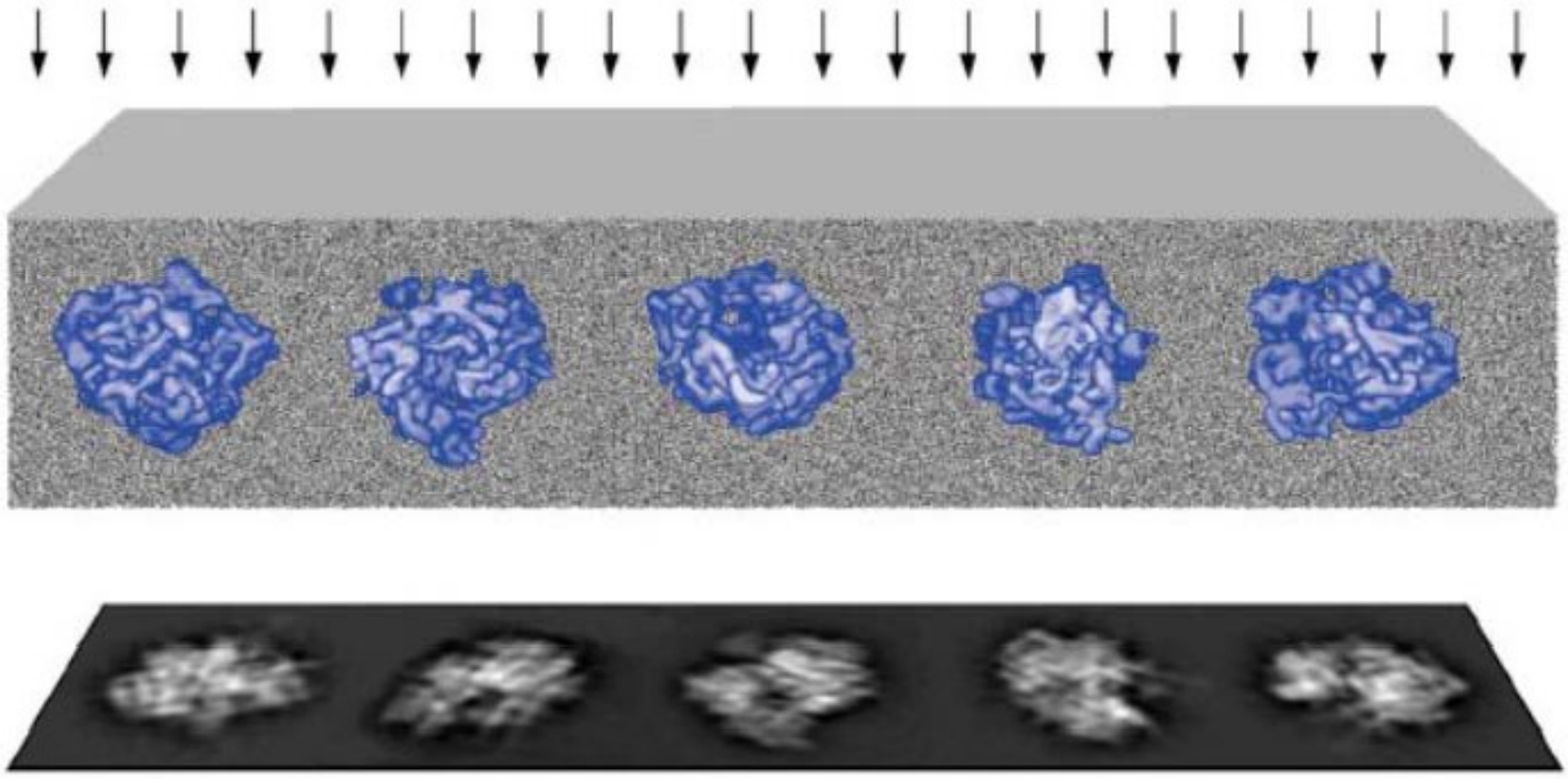
EM:
CCT Chaperonin and Actin

Valpuesta lab:
chaperonin CCT unfolds bound actin (Llorca et al., EMBO J. 19:5971, 2000)

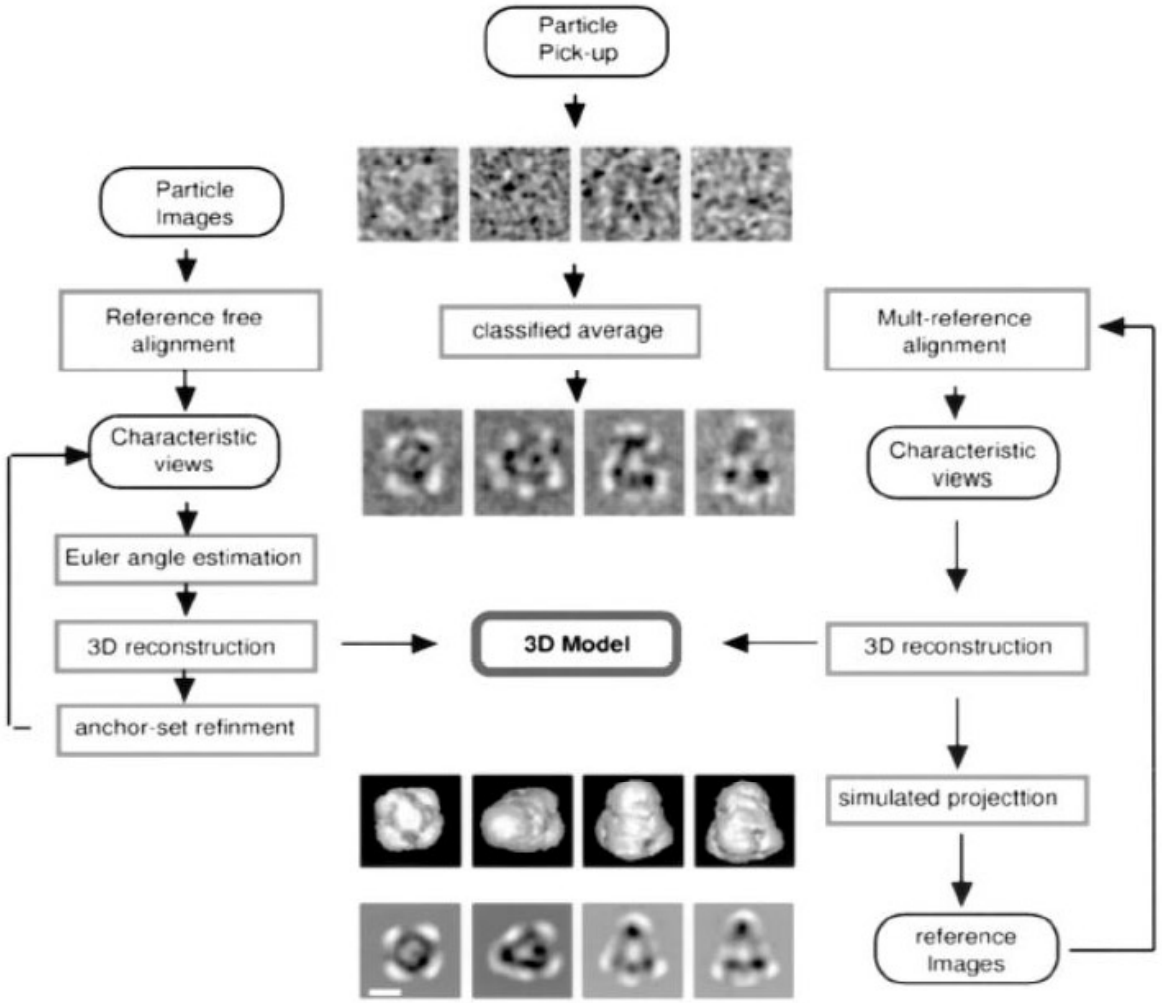
EM Specimen Preparation and Imaging



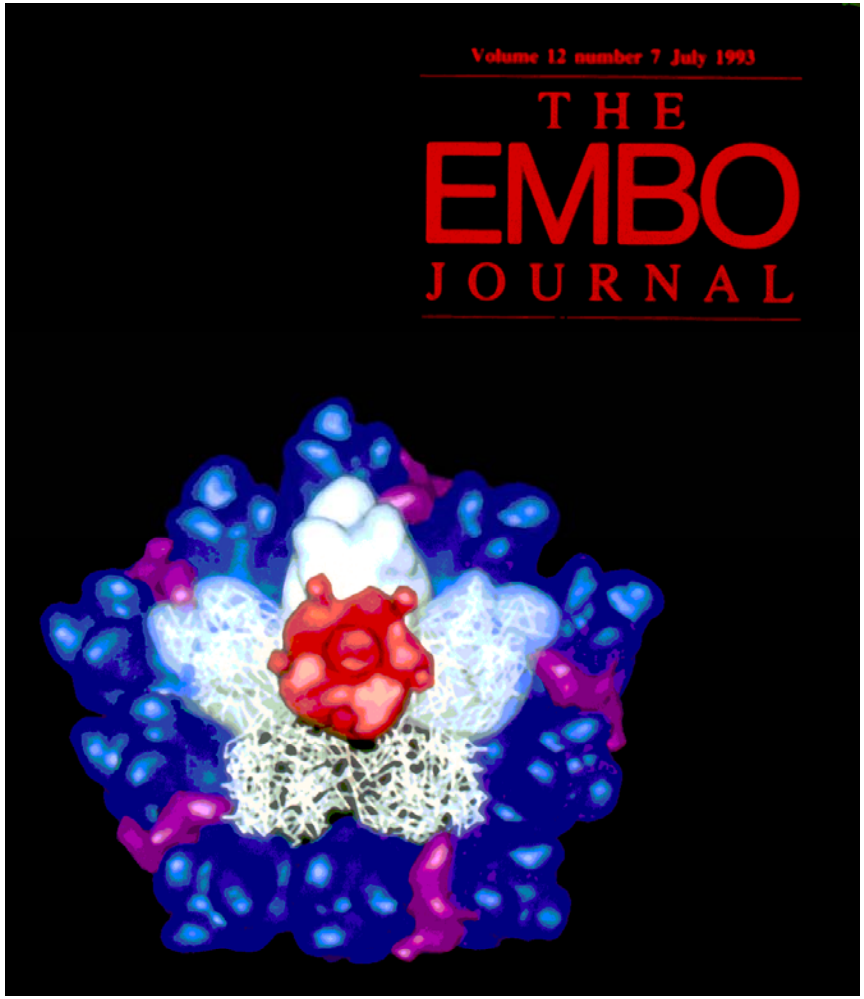
Single Particle Imaging



3D Reconstruction



“Quantitative” Electron Microscopy



Phoebe L. Stewart, Stephen D. Fuller, Roger M. Burnett:

Difference imaging of adenovirus: bridging the resolution gap between X-ray crystallography and electron microscopy.

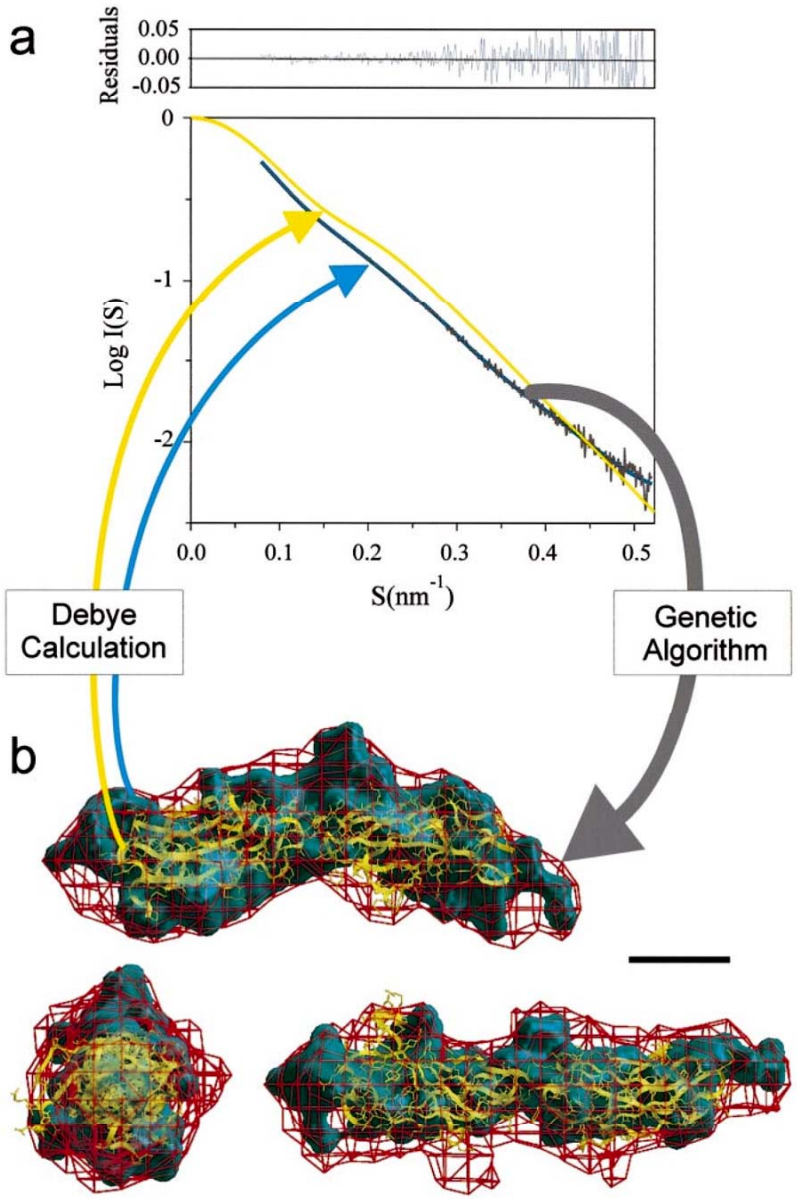
EMBO J., 12:2589, 1993.

“At that time, placing an atomic structure into an EM map seemed like a very dangerous idea…”

Phoebe Stewart, GRC 2003

Generating 3D Structures from 1D SAXS Data

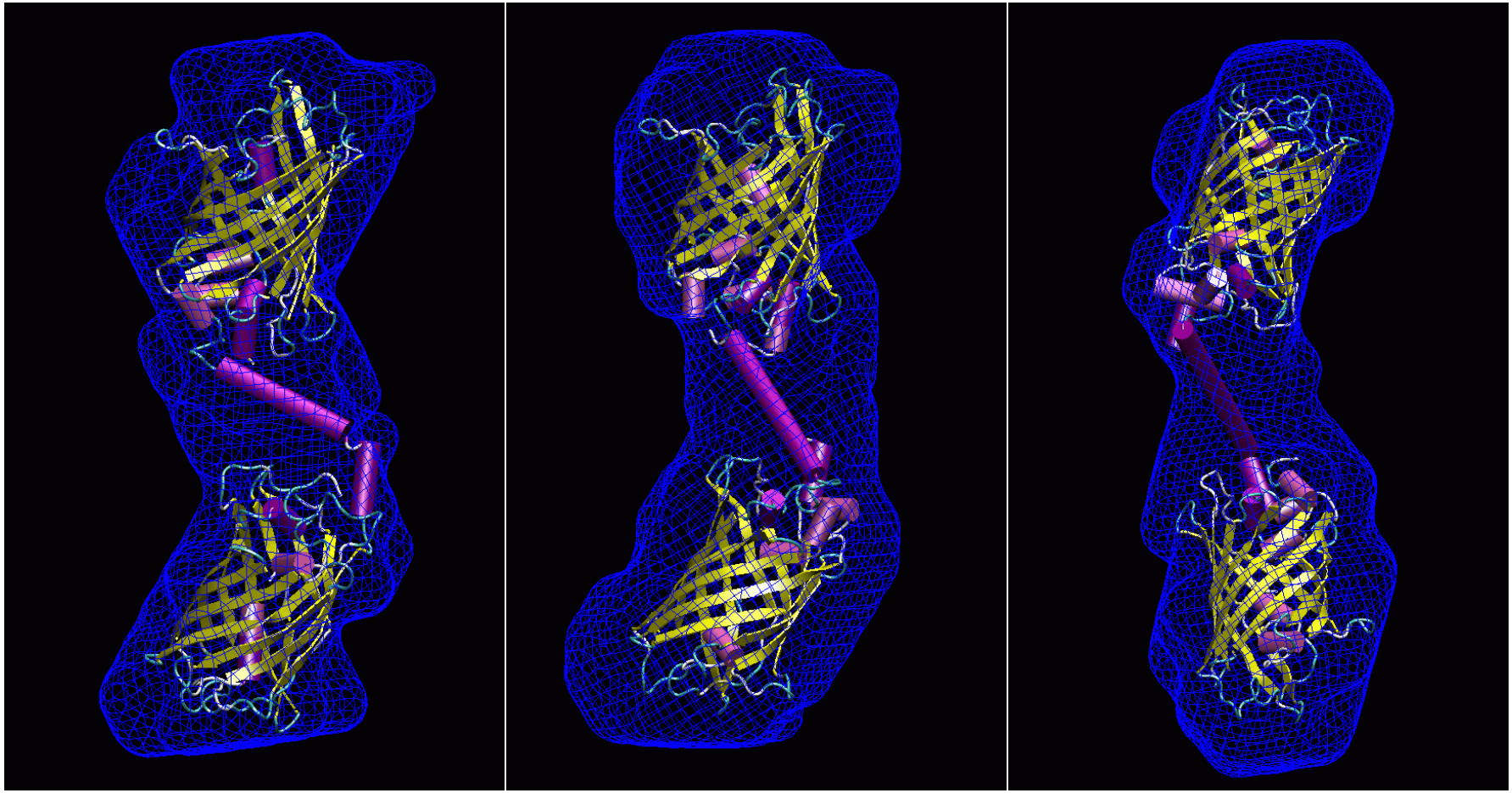
Low-resolution
3D shapes from
1D scattering profiles!



Chacon et al., JMB (2000) 299:1289

SAXS Application

EBFP-linker-EGFP fusion proteins
(Tetsuro Fujisawa)



LAEAAAKEAAAKEAAAKAAA (20)

LAEAAAKEAAAKEAAAKEAAAKAAA (25)

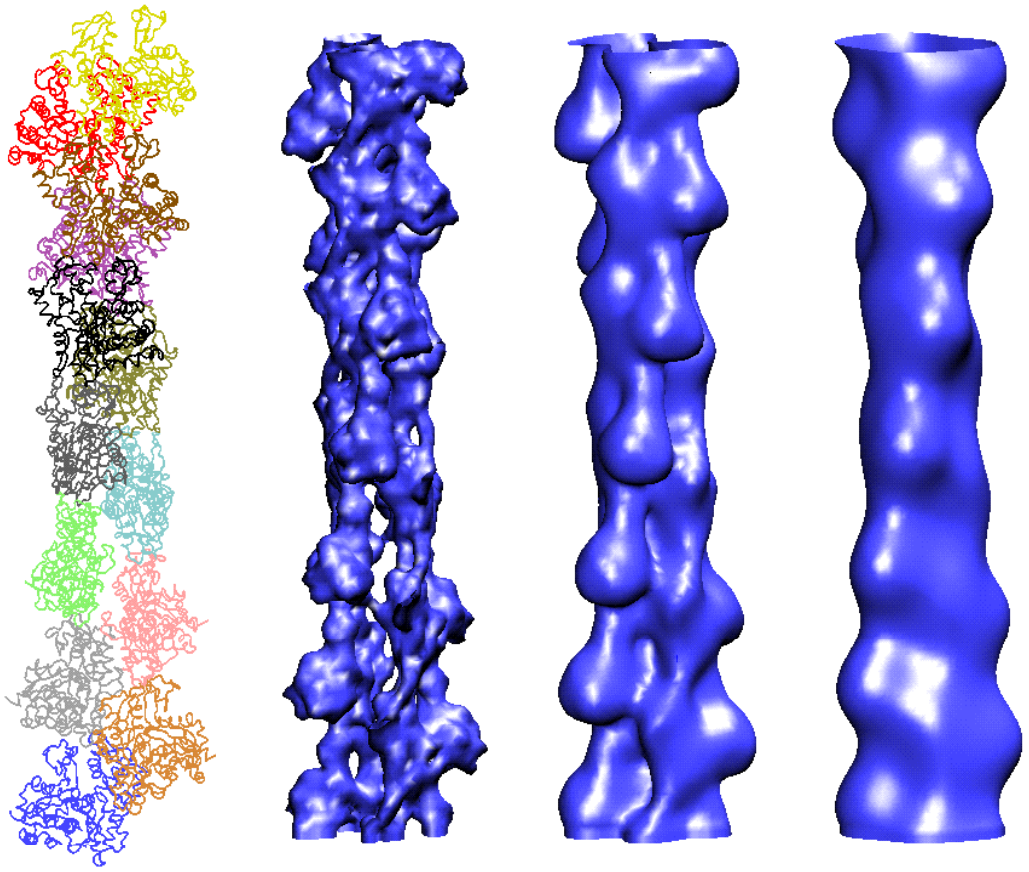
LAEAAAKEAAAKEAAAKEAAAKEAAAKAAA (30)

Combining Multi-Resolution Biophys. Data

Actin filament
(Holmes et al., 1990)

Convolution with Gaussian:

$$G(r) = \exp\left(\frac{-3r^2}{2\sigma^2}\right)$$

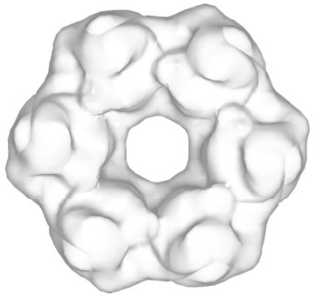


2σ=0Å 2σ=10Å 2σ=20Å 2σ=40Å

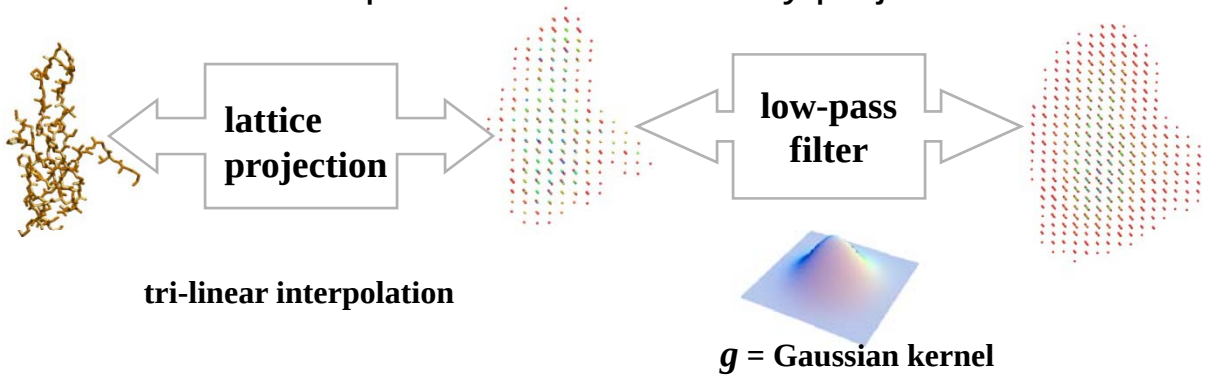
Q: We can lower the resolution of 3D data, but how can one increase it?
A: Combine low- with high-resolution data by flexible and rigid-body fitting.

Correlation-Based 'Interior' Docking

$\rho_{\mathrm{em}}(\mathbf{r})$ target density on lattice

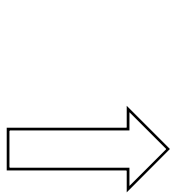


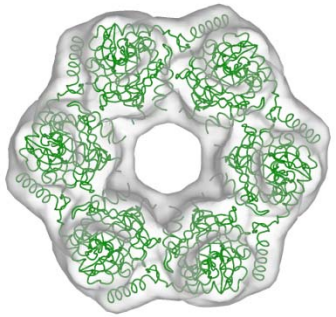
$\rho_{\mathrm{atomic}}(\mathbf{rR})$ rotated probe molecule density projected to the lattice:



$$g \otimes \rho_{\mathrm{atomic}}(\mathbf{rR}) \equiv \rho_{\mathrm{calc}}(\mathbf{r})$$

$$C(\mathbf{T}) = \int \rho_{\mathrm{em}}(\mathbf{r}) \times \rho_{\mathrm{calc}}(\mathbf{r}+\mathbf{T}) \, \mathrm{d}^3\mathbf{r}$$





Fitting criterion: e,g, linear cross-correlation, evaluate for every rotation **R** and translation **T**

FTM (Fast Translational Matching)

$$C(\mathbf{T}) = \int \left(e \otimes \rho_{\text{em}}\right)(\mathbf{r}) \times \left(e \otimes \rho_{\text{calc}}\right)(\mathbf{r}+\mathbf{T})\ \mathrm{d}^3\mathrm{r} = f^{-1}\left[f\left(e \otimes \rho_{\text{em}}\right)^{*} \times f\left(e \otimes \rho_{\text{calc}}\right) \right]$$

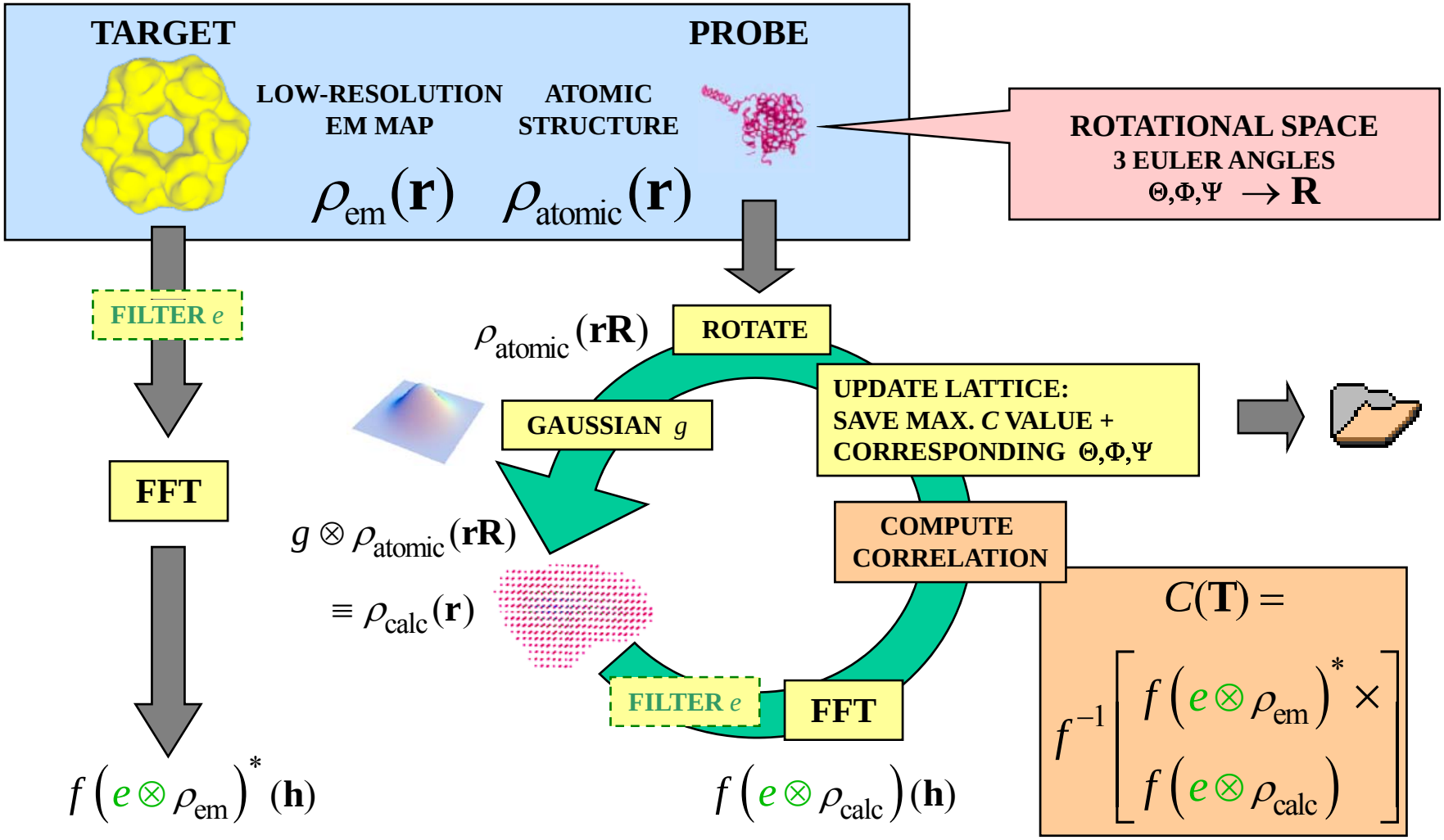
Fourier Convolution Theorem:

Direct Approach: N^2 multiplications
FFT Approach: N log N multiplications

N = number of voxels
e = optional filter

FFT Acceleration of the Translational Search!

6D Search with FTM



Off-Lattice Refinement

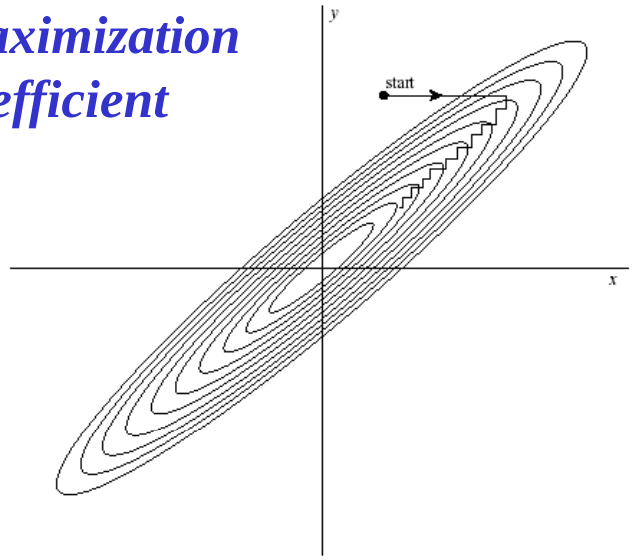
6D exhaustive search is limited:

- Rotational search →Angular sampling
- Translational search→ Grid size

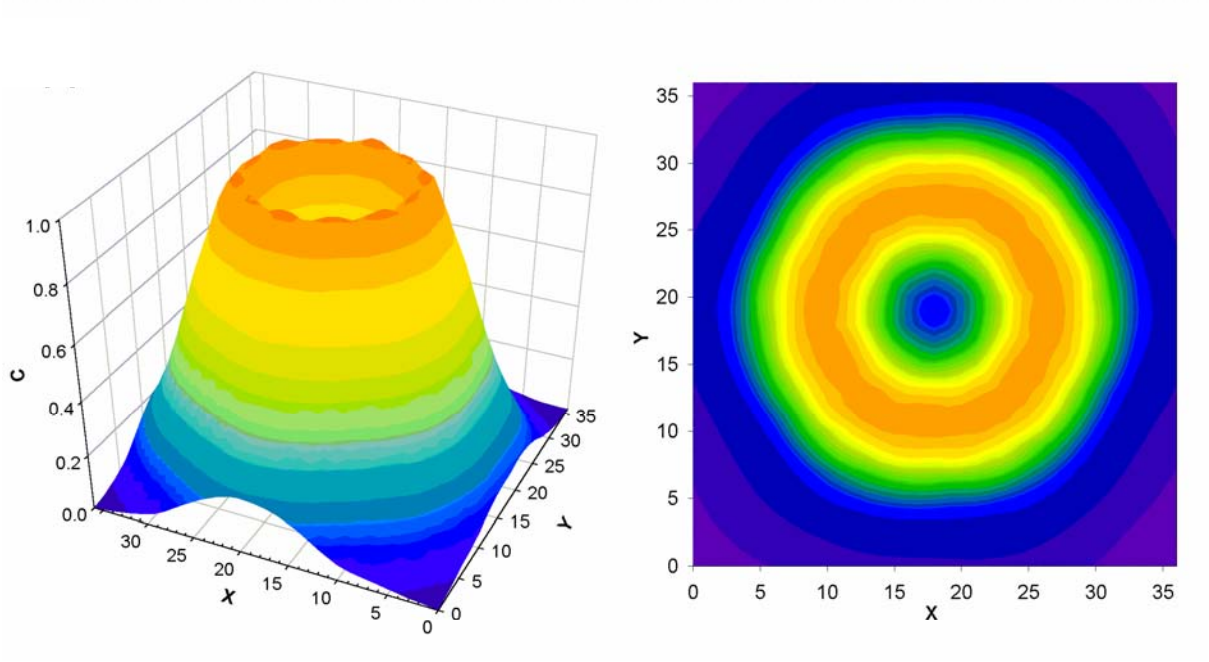
Improve the accuracy

Off-lattice (6D) local maximization of the correlation coefficient

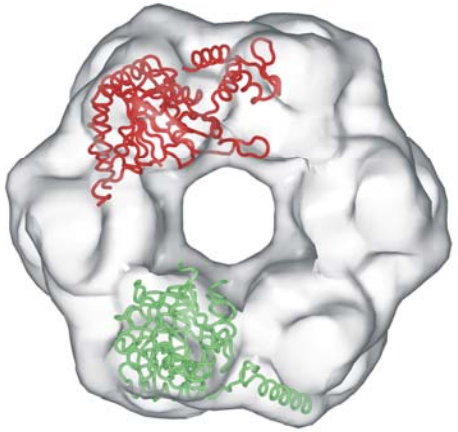
Powell's quadratically convergent maximization method can be used to perform a 6D search around the best fits found on the grid.



Correlation Landscape



With density cross-correlation we can not distinguish between correct and spurious fit

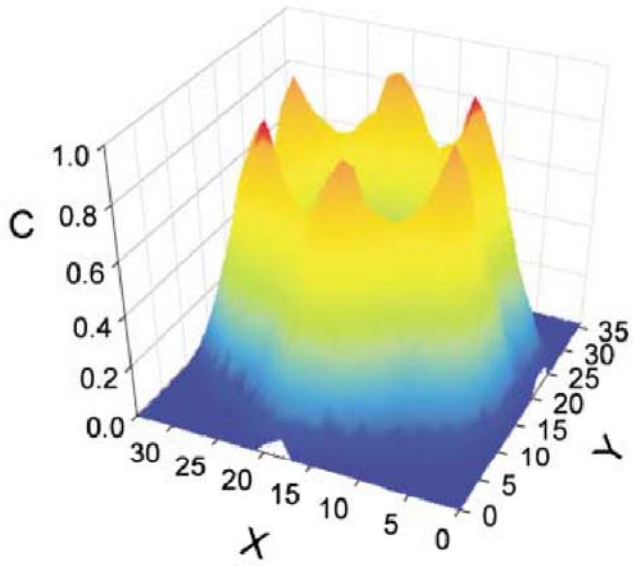


Density Masking

Renormalize (mask) the correlation locally

$$C(\mathbf{T}) = \frac{\int_{mask} \rho_{em}(\mathbf{r}) \times \rho_{calc}(\mathbf{r}+\mathbf{T})\ d^3r}{\sqrt{\int_{mask} \rho^2_{em}(\mathbf{r}) d^3r}\sqrt{\int_{mask} \rho^2_{calc}(\mathbf{r})\ d^3r}}$$

$$\text{mask} \rightarrow \rho_{\text{calc}\,l,m,n} > 0$$



→extends the reliability of correlation based docking (<15Å)

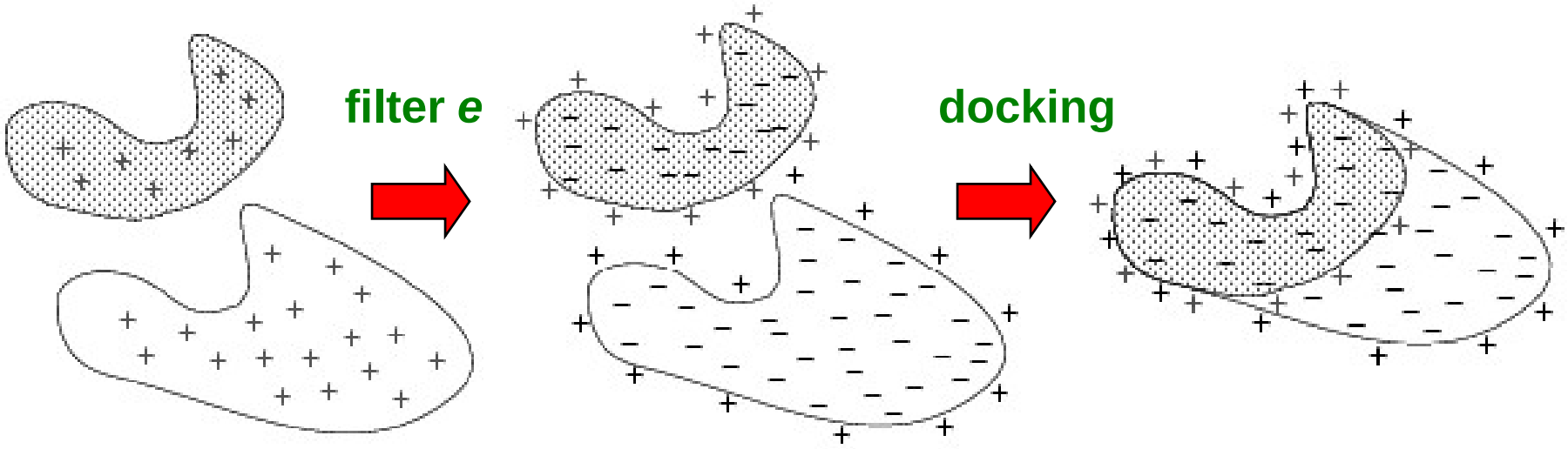
→Can not be easily FTT accelerated

DOCKEM
A.M. Roseman

Density Filtering

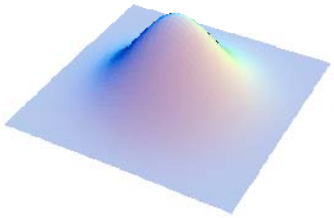
Adding surface/contour information

A suitable filter would assign negative values to the interior, positive values to the molecular contour. Both volume and contour matches would provide positive contributions to the correlation criterion:

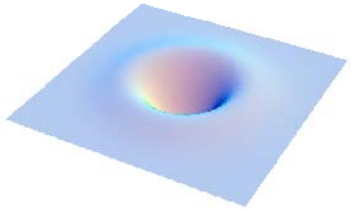


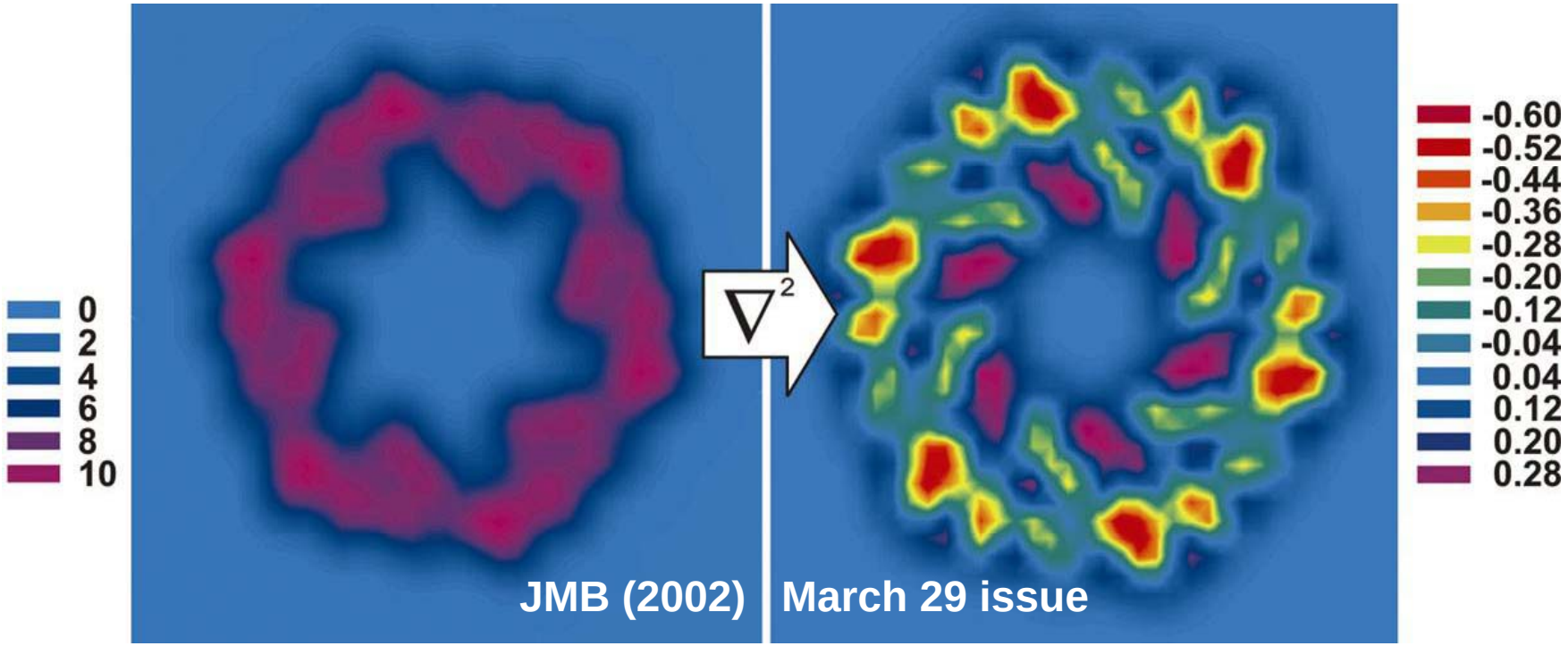
Contour Filter

Laplacian

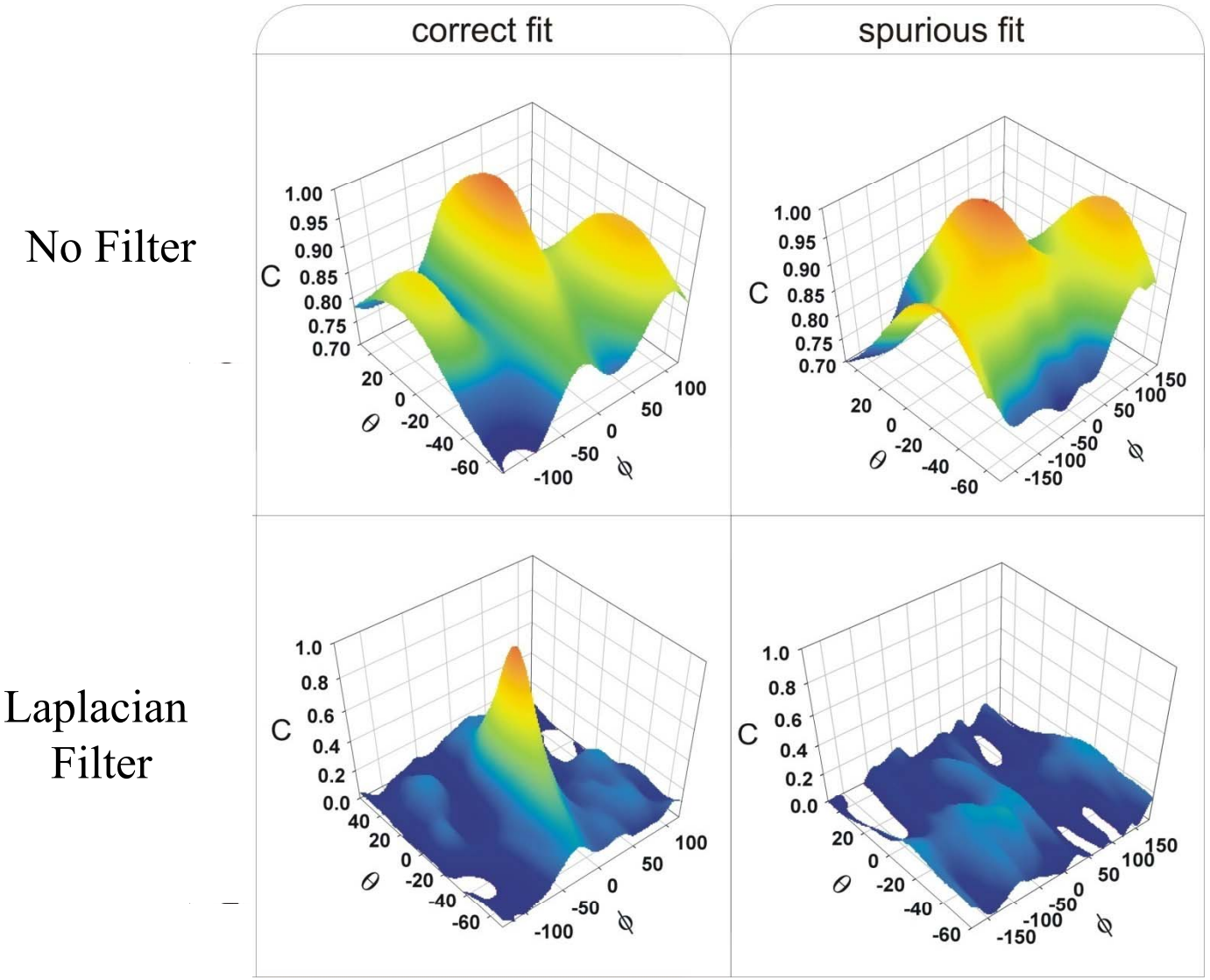


$$\nabla^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}$$

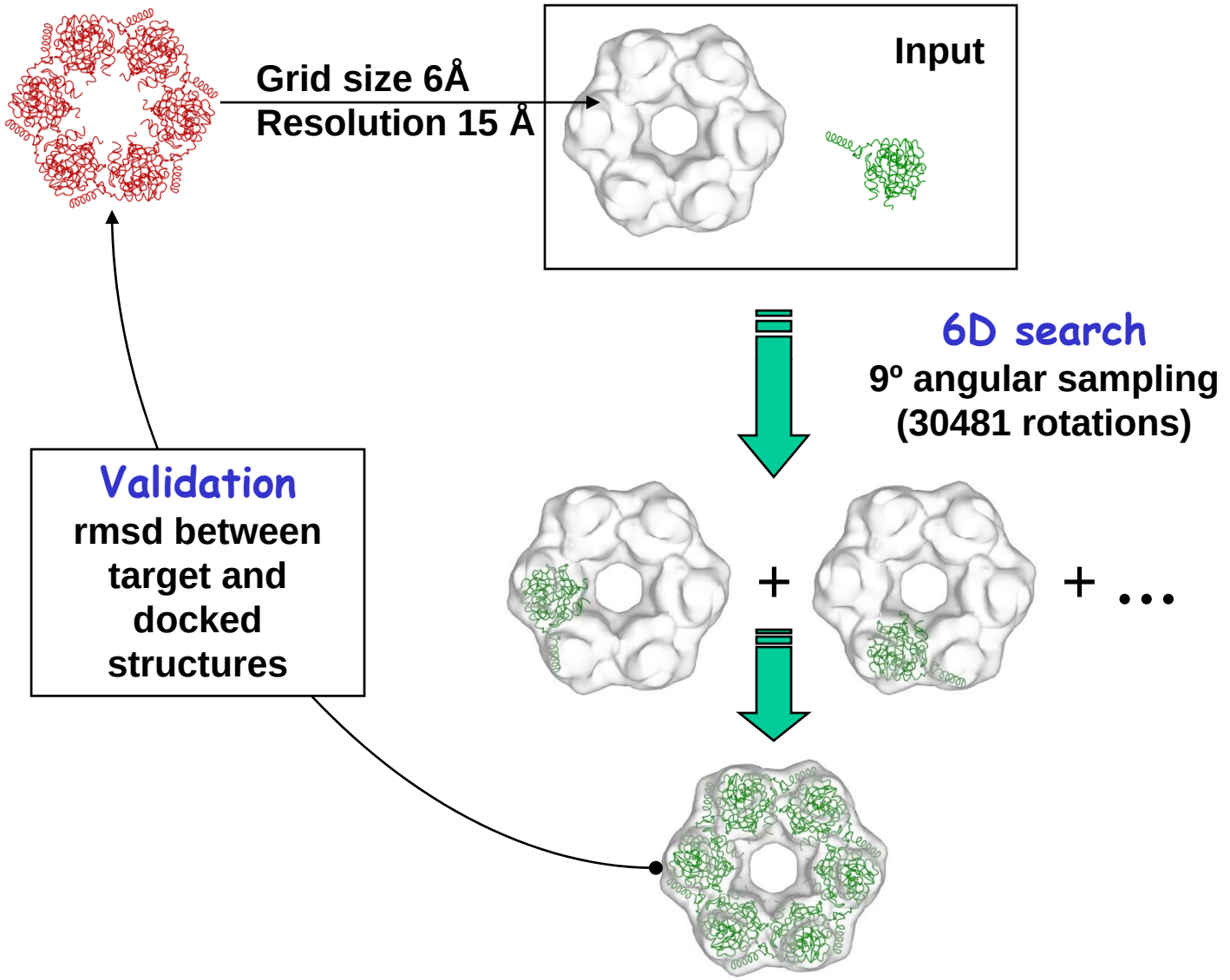




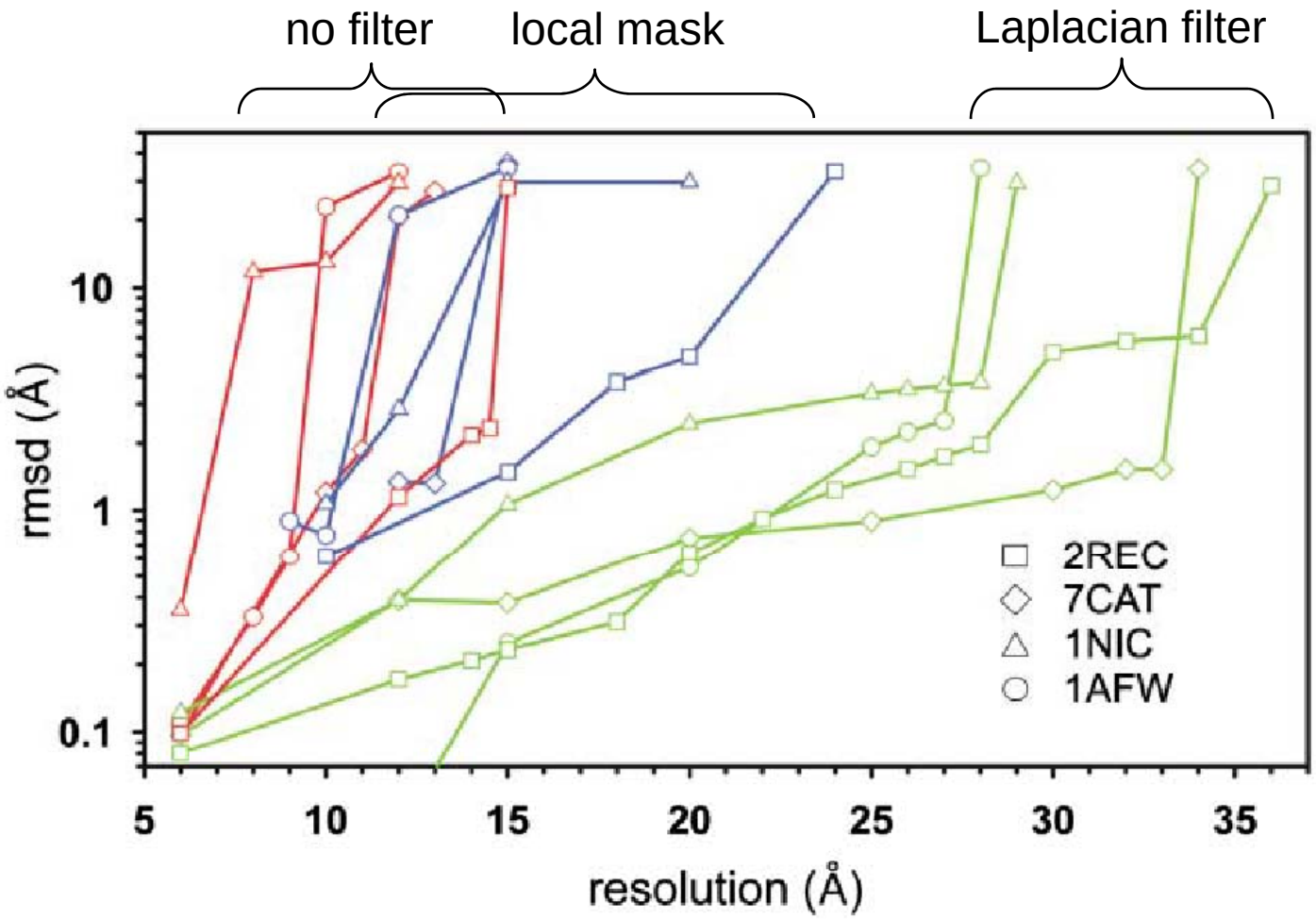
Effect of Filter on Orientation



Restoration Tests with Simulated Data



Restoring Various Oligomers



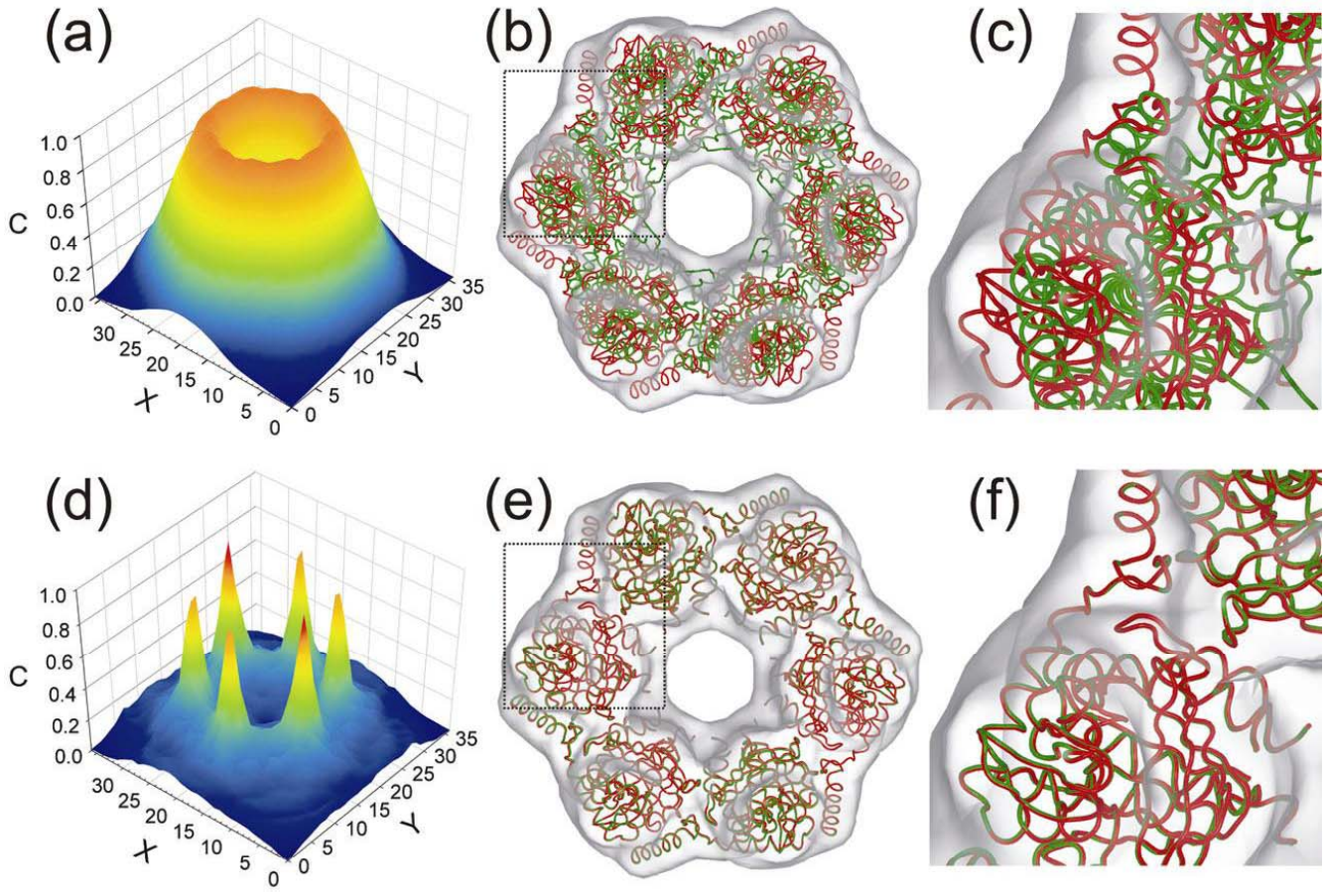
RecA (2REC), thiolase (1AFW), catalase (7CAT), and oxidoreductase (1NIC).

Example: RecA

Grid size 6Å
Resolution 15Å
9^0 steps (30481 rotations)

standard cross-correlation

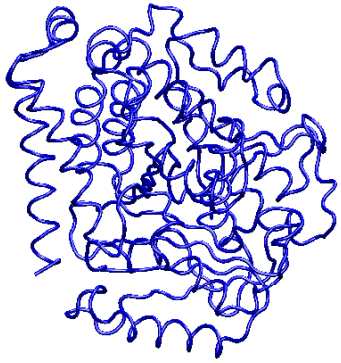
w/ Laplacian filtering



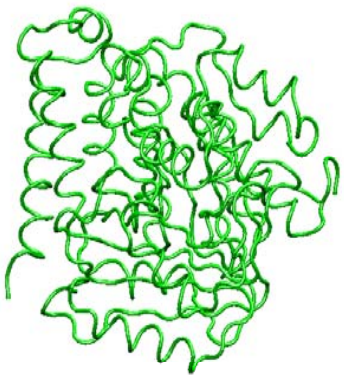
Only Laplacian filtering successfully restores the initial pose

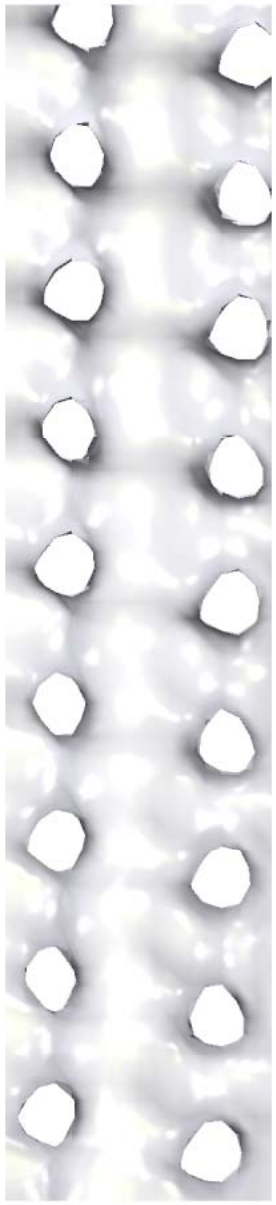
Application to Microtubule Data

β-tubulin

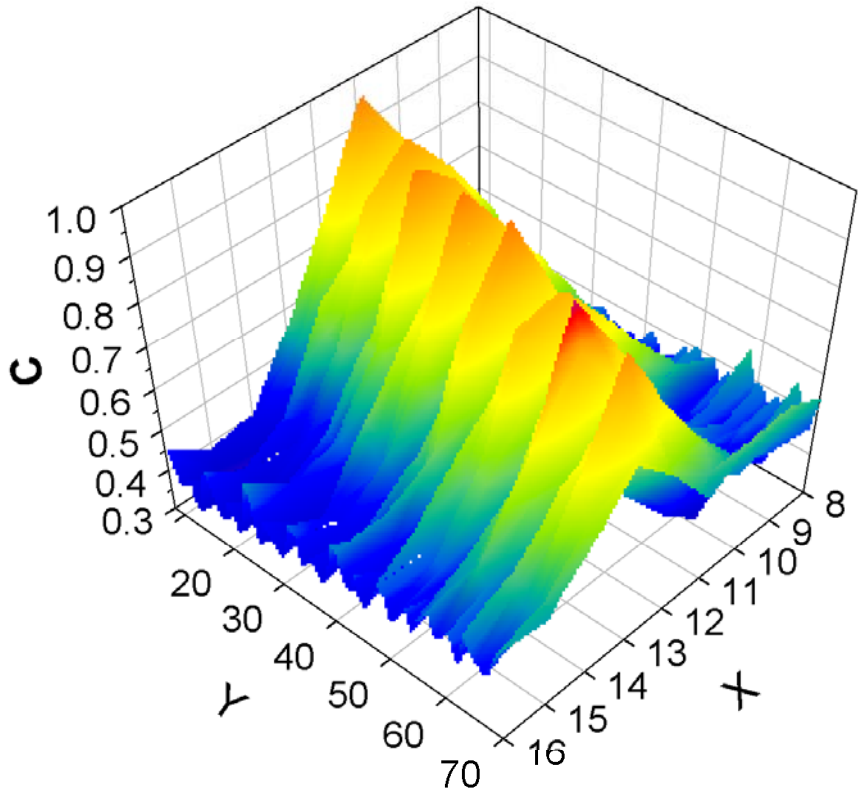


α-tubulin

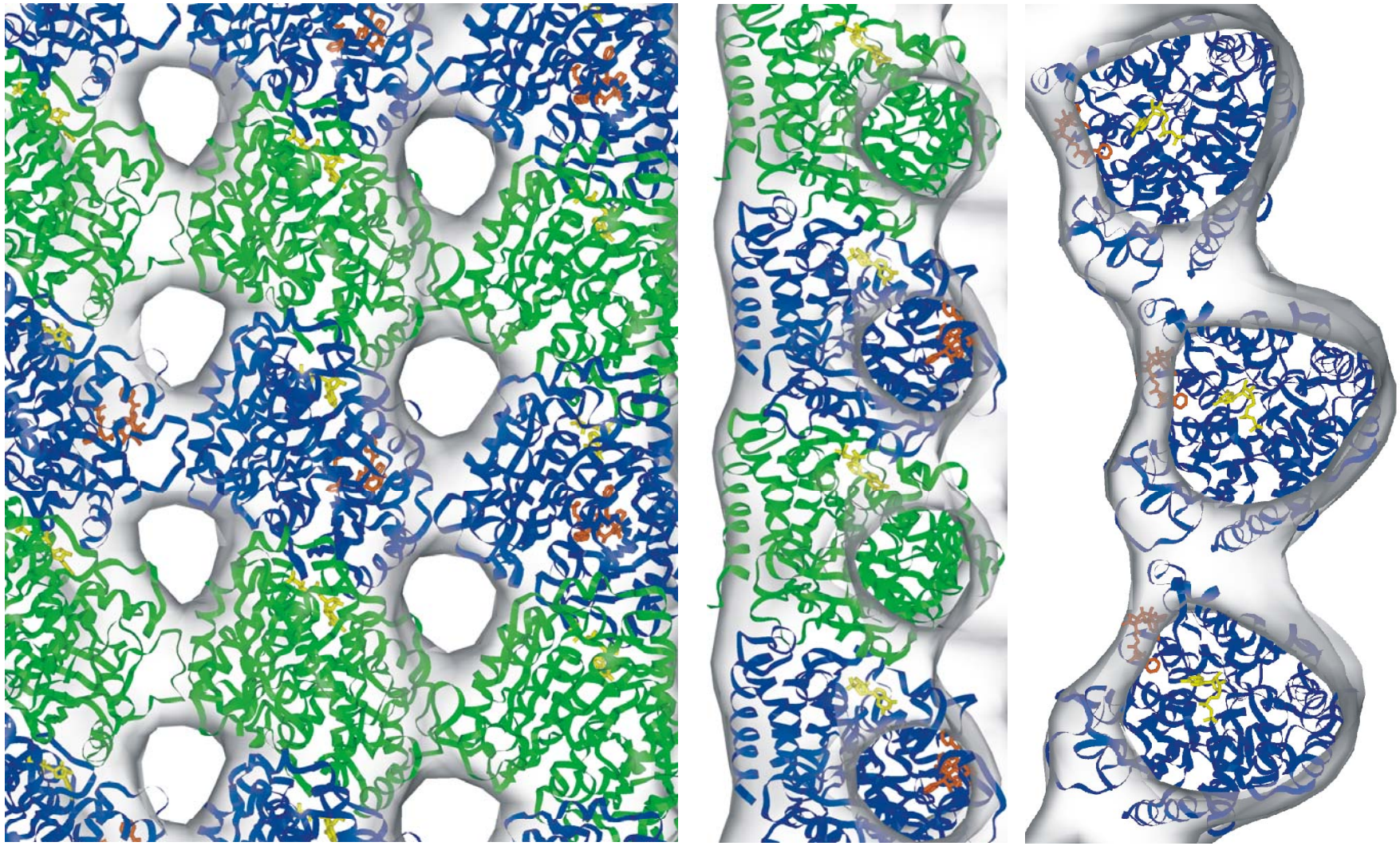




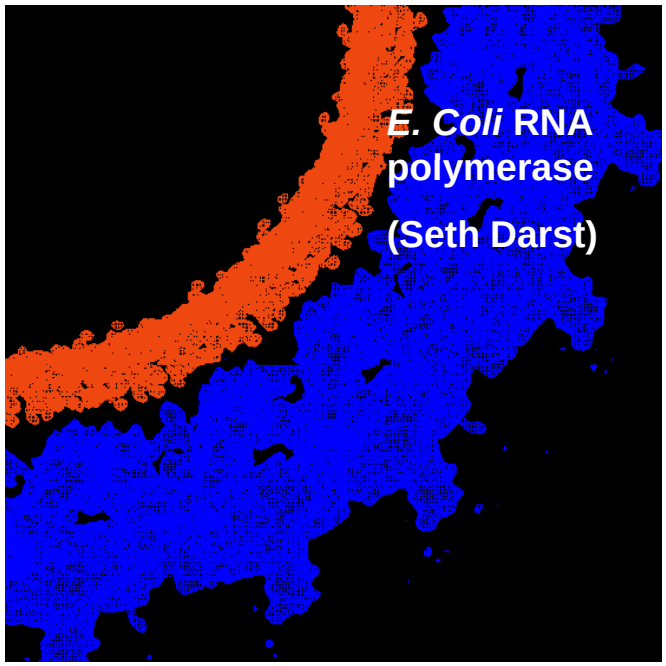
Resolution 20Å
Angular sampling 9º
Grid size 5Å

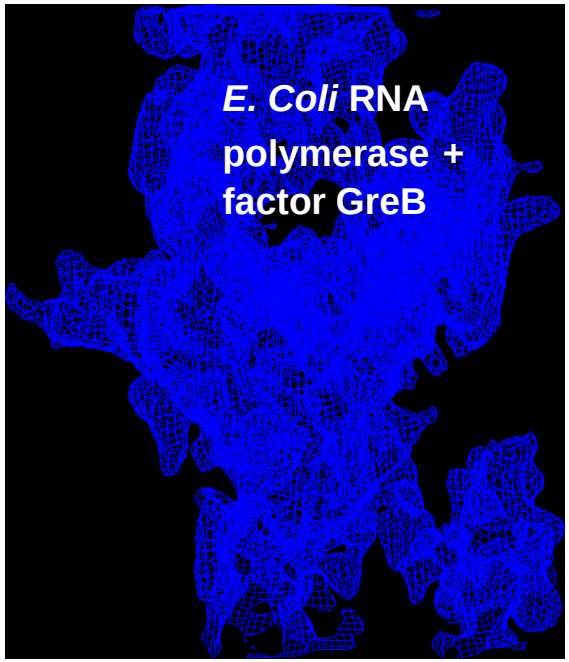


Microtubule Model



Registration of Two EM Maps

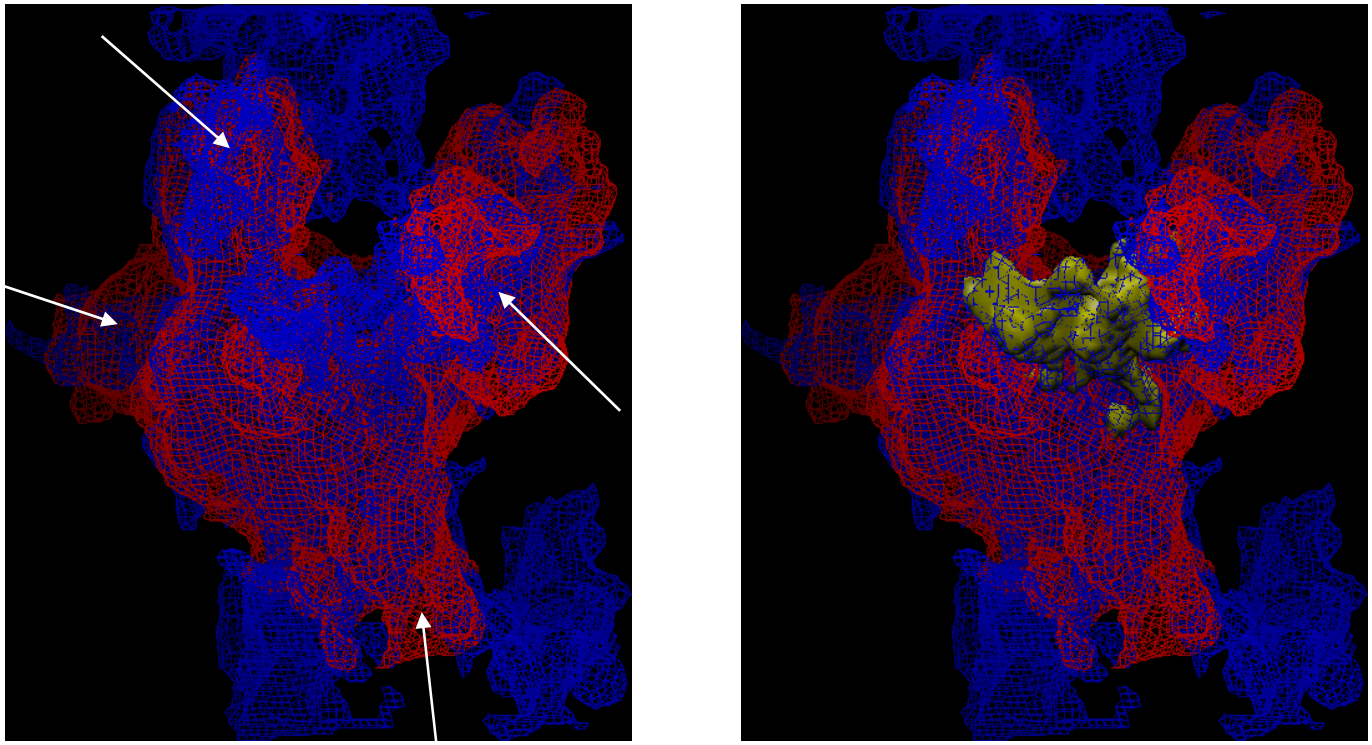




Problem: different helical arrays

Need to perform difference mapping to localize GreB (difficult at variable helical symmetry)

Registration and Difference Mapping



Rigid-body docking: The RNAP “jaws” are open in presence of GrepB factor, perform flexible map fitting

Map fitting will be available in Situs 2.2.

FRM: Fast *Rotational* Matching

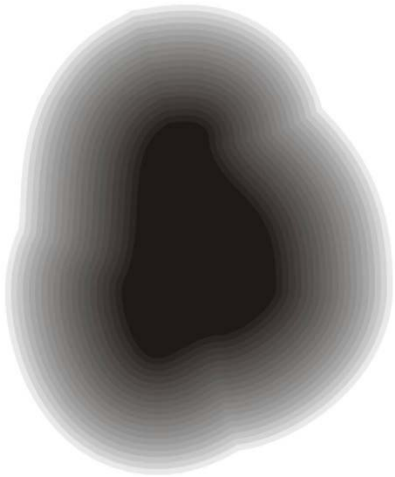
Euler angle search is expensive!

9º angular sampling (30481 rotations) requires > 10 minutes on standard workstation for rotations only.
Rotations + translations: 10-20 hours.

Our Goal: We seek to FFT-accelerate rotational search in addition to translational search.

To do this, we need to do the math in rotational space and take advantage of expressions similar to convolution theorem that are best described by group theory.

Expansion in Spherical Harmonics

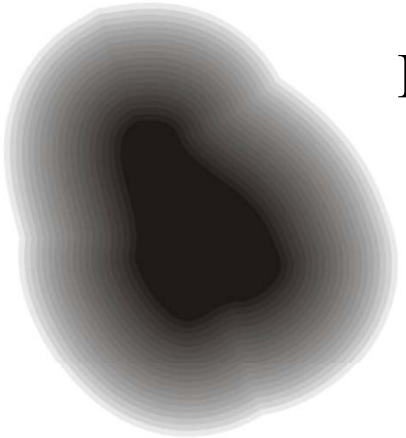


Target map

f

$$f(su) = \sum_{l=0}^{B-1} \sum_{m=-l}^{l} \hat{f}_{lm}(s) Y_{lm}(u)$$

Y_{lm} are the *spherical harmonic functions*



Probe map

g

$$g(su) = \sum_{l=0}^{B-1} \sum_{m=-l}^{l} \hat{g}_{lm}(s) Y_{lm}(u)$$

FRM_{3D} Method

The correlation function is:

$$c(R) = \int_{\mathbf{R}^3} f \cdot g(R) = c(\alpha, \beta, \gamma)$$

α,β,γ are specially chosen Euler angles (origin shift).

Expanding f and g in spherical harmonics as before

we arrive at: $\hat{c}(p,q,r) = \sum_l d^l_{pq}(\frac{\pi}{2})\ d^l_{qr}(\frac{\pi}{2})\ I^l_{pr}$

where: $I^l_{pr} = \int_0^\infty \hat{f}_{lp}(s)\,\overline{\hat{g}_{lr}(s)}\,s^2 ds$

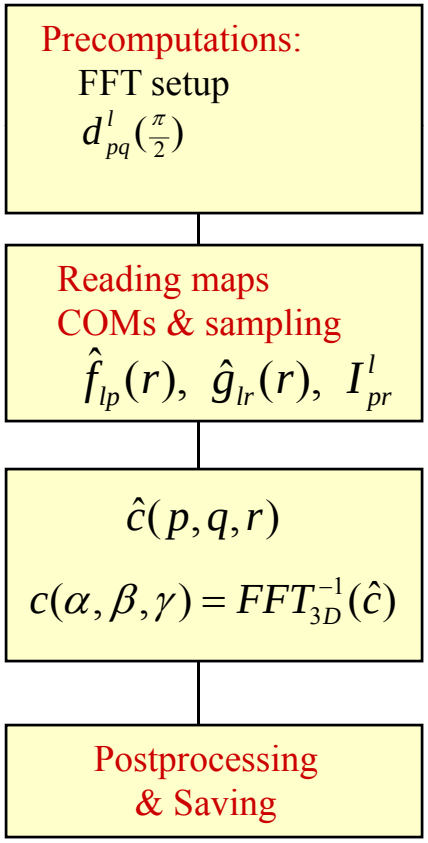
$$c(\alpha, \beta, \gamma) = FFT^{-1}_{3D}(\hat{c})$$

The quantities $d^l_{pq}(\frac{\pi}{2})$ (which come from rotation group theory) are precomputed using a recursive procedure.

Comparison between FRM_{3D} and Crowther

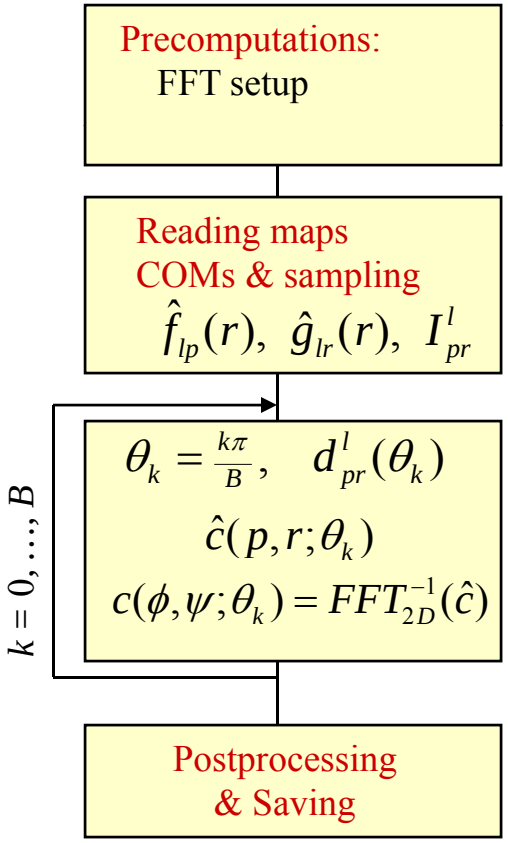
$$c(R) = c(\alpha, \beta, \gamma)$$
$$\hat{c}(p,q,r) = \sum_l d^l_{pq}(\tfrac{\pi}{2}) d^l_{qr}(\tfrac{\pi}{2}) I^l_{pr}$$

FRM



$$c(R) = c(\phi, \psi; \theta)$$
$$\hat{c}(p,r;\theta) = \sum_l d^l_{pr}(\theta) I^l_{pr}$$

Crowther



Timings of FRM_{3D} and Crowther

(seconds)

angular sampling	Crowther	FRM
6°	1.66	0.97
3°	19.3	3.75
1.4°	337	37.4

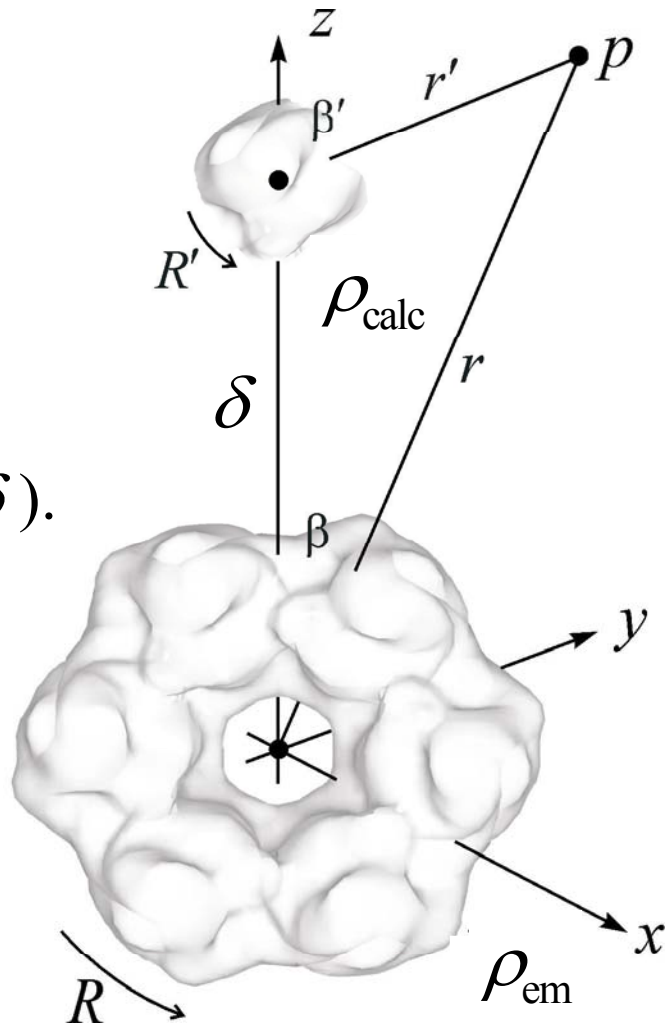
FRM_{6D} (Rigid-Body Matching)

5 angular parameters.

The correlation function is now:

$$c(R, R'; \delta) = \int_{\mathbf{R}^3} \left(e \otimes \rho_{\text{em}}\right)(R) \cdot \left(e \otimes \rho_{\text{calc}}\right)(R'; \delta).$$

1 linear parameter remains, distance δ of movement along the z axis.



Rigid-Body Search by 5D FFT

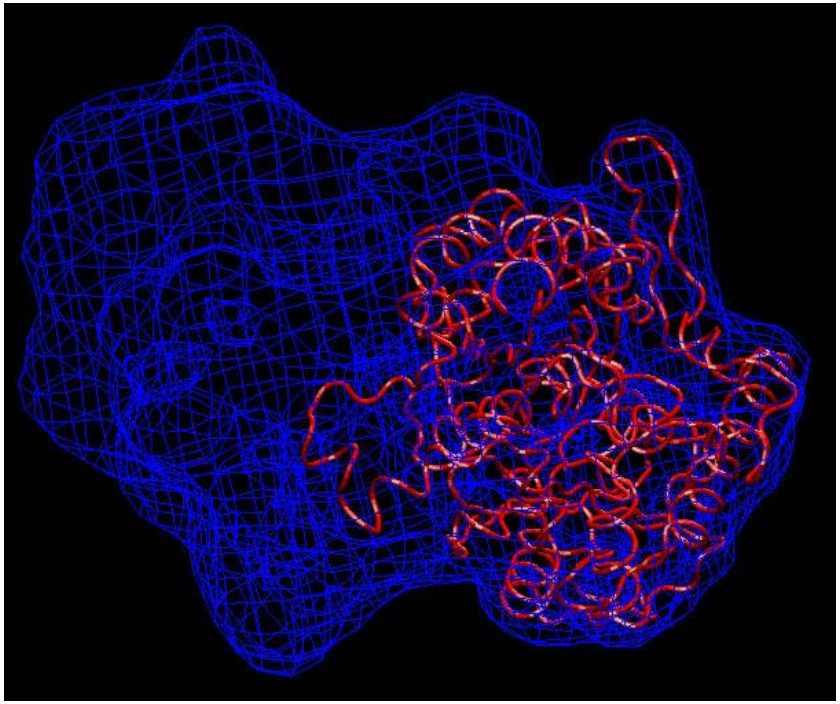
The 5D Fourier transform of the correlation function turns out to be:

$$\hat{c}(n,h,m,h',m';\delta) = (-1)^n \sum_{l,l'} d^l_{nh} d^l_{hm} d^{l'}_{-nh'} d^{l'}_{h'm'} I^{ll'}_{mnm'}(\delta).$$

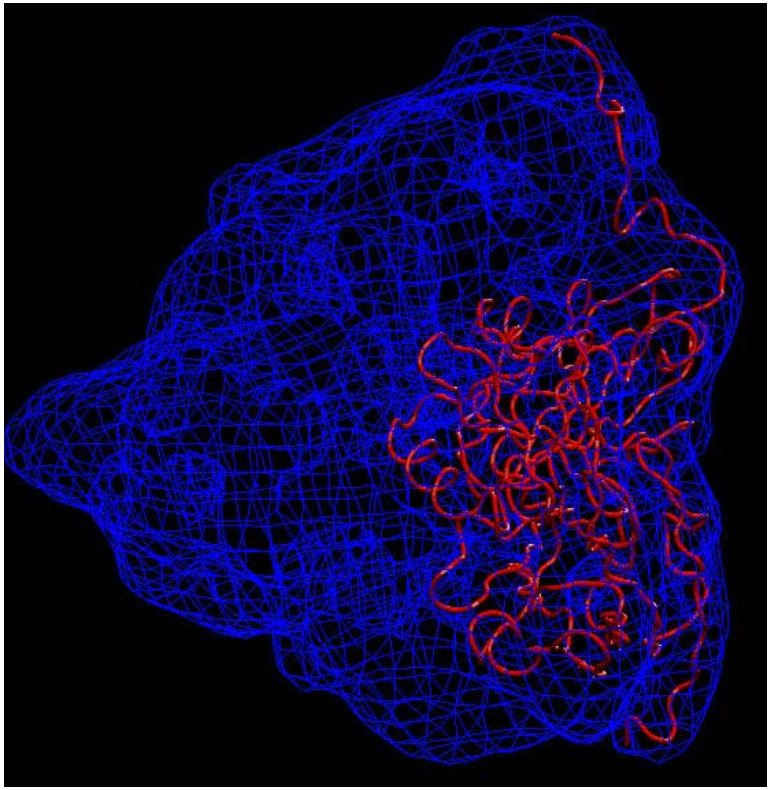
The quantities $I^{ll'}_{mnm'}(\delta)$ are the so called *two-center integrals*, corresponding to the spherical harmonic transforms of the two maps, at a distance δ of one another:

$$I^{ll'}_{mnm'}(\delta) = \sqrt{(l+\tfrac{1}{2})(l'+\tfrac{1}{2})} \int_0^{\pi} \left[\int_0^{\infty} \overline{\hat{\rho}^{lm}_{em}(r) \hat{\rho}^{l'm'}_{calc}(r')} \; d^{l'}_{n0}(\beta') r^2 dr \right] d^l_{n0}(\beta) \sin\beta \; d\beta$$

Test Cases

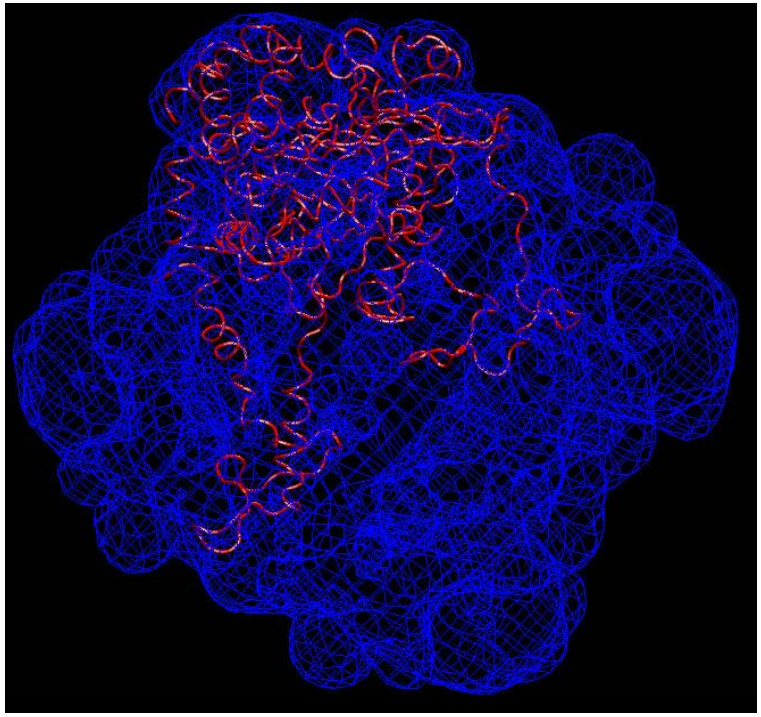


1afw
(peroxisomal thiolase)

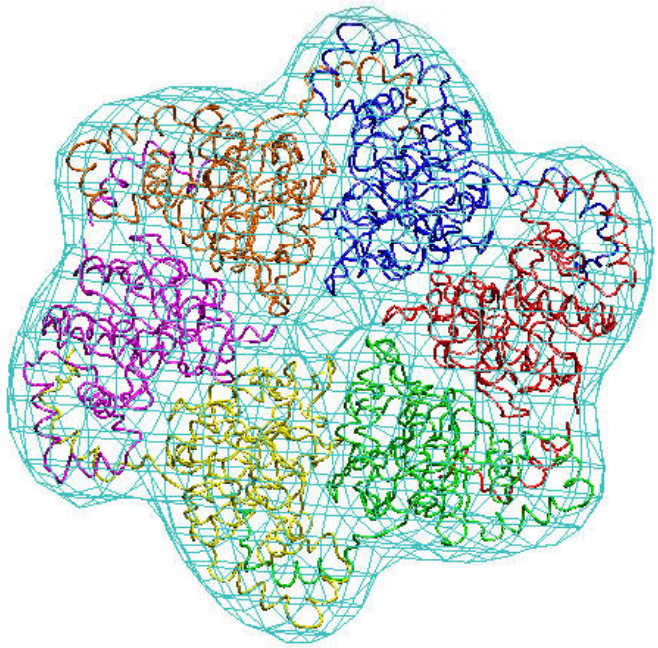


1nic (copper-
nitrite reductase)

Test Cases



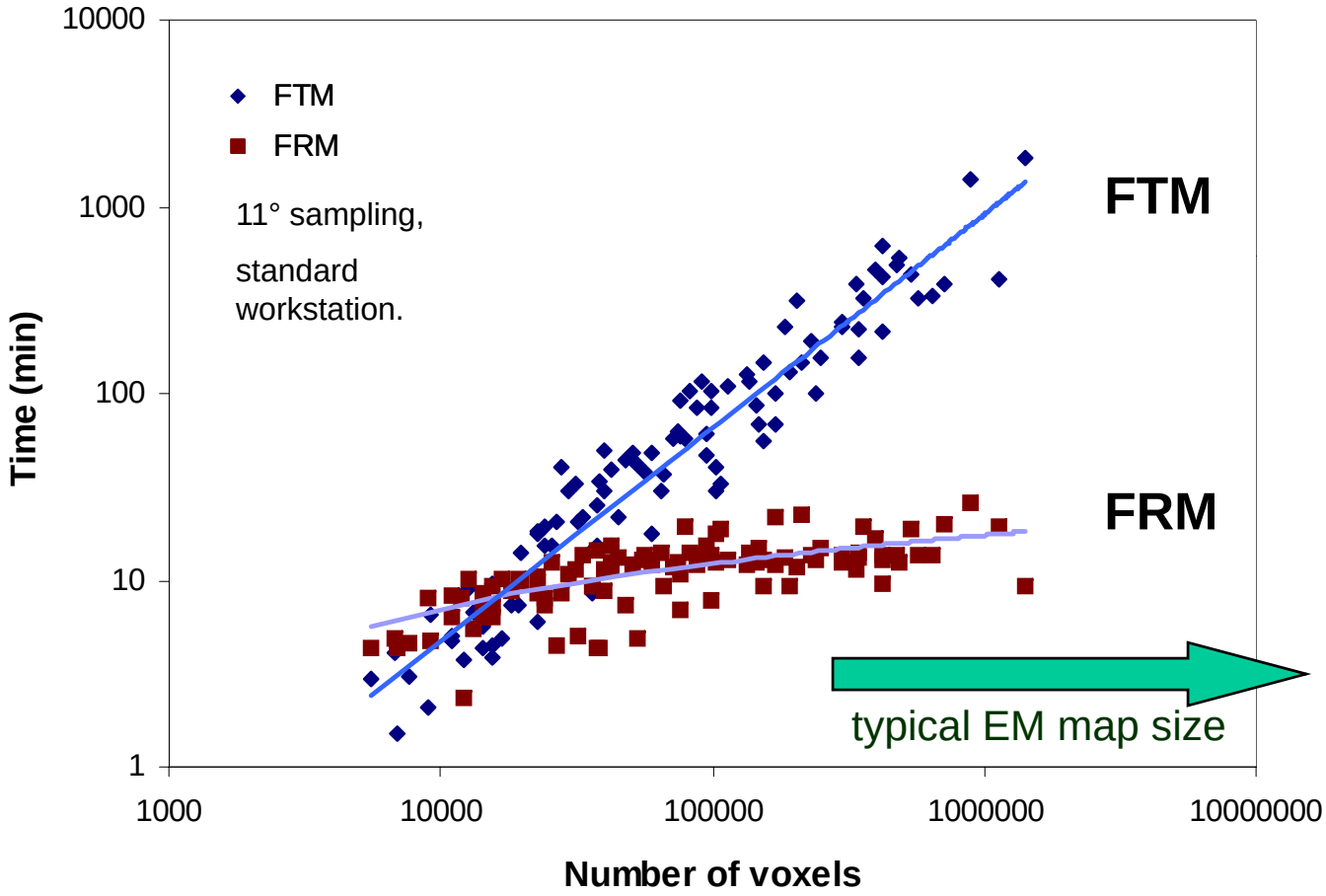
7cat
(catalase)



1e0j
(Gp4D helicase)

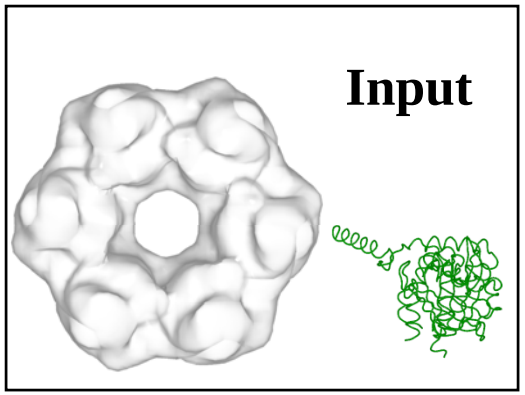
Efficiency: FRM vs. FTM

FTM: 3D FFT + 3D rot. search
FRM: 5D FFT + 1D trans. search



- Gain: 2-4 orders of magnitude for typical EM map

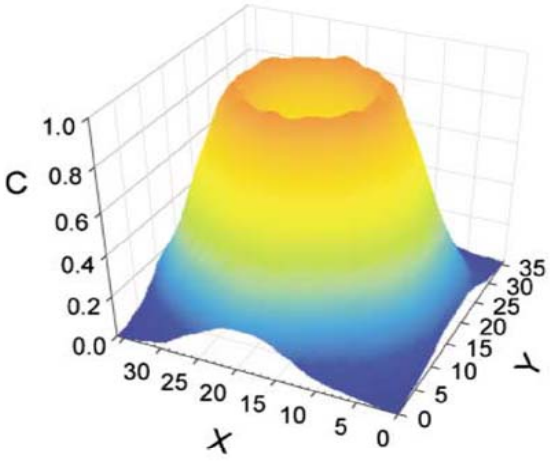
Summary: Correlation Based Matching



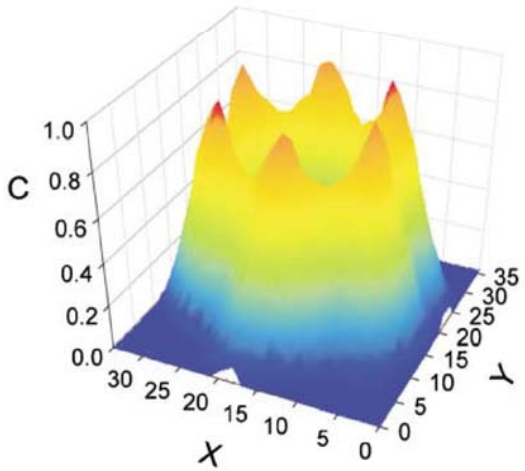
Situs 6D exhaustive searches:

- Rigid Body
- Fast Translational Matching
- Fast Rotational Matching
- Density Filtering

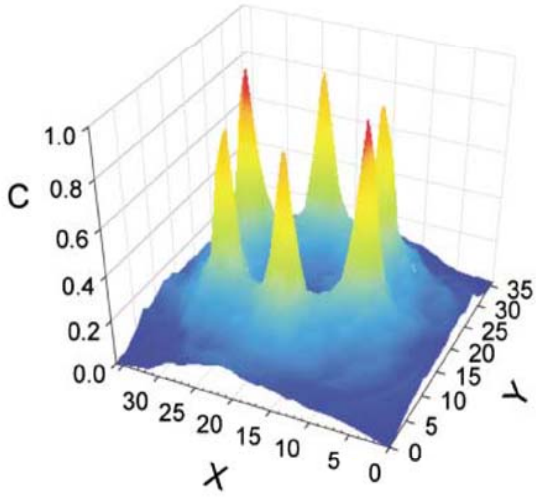
No filter



Local mask



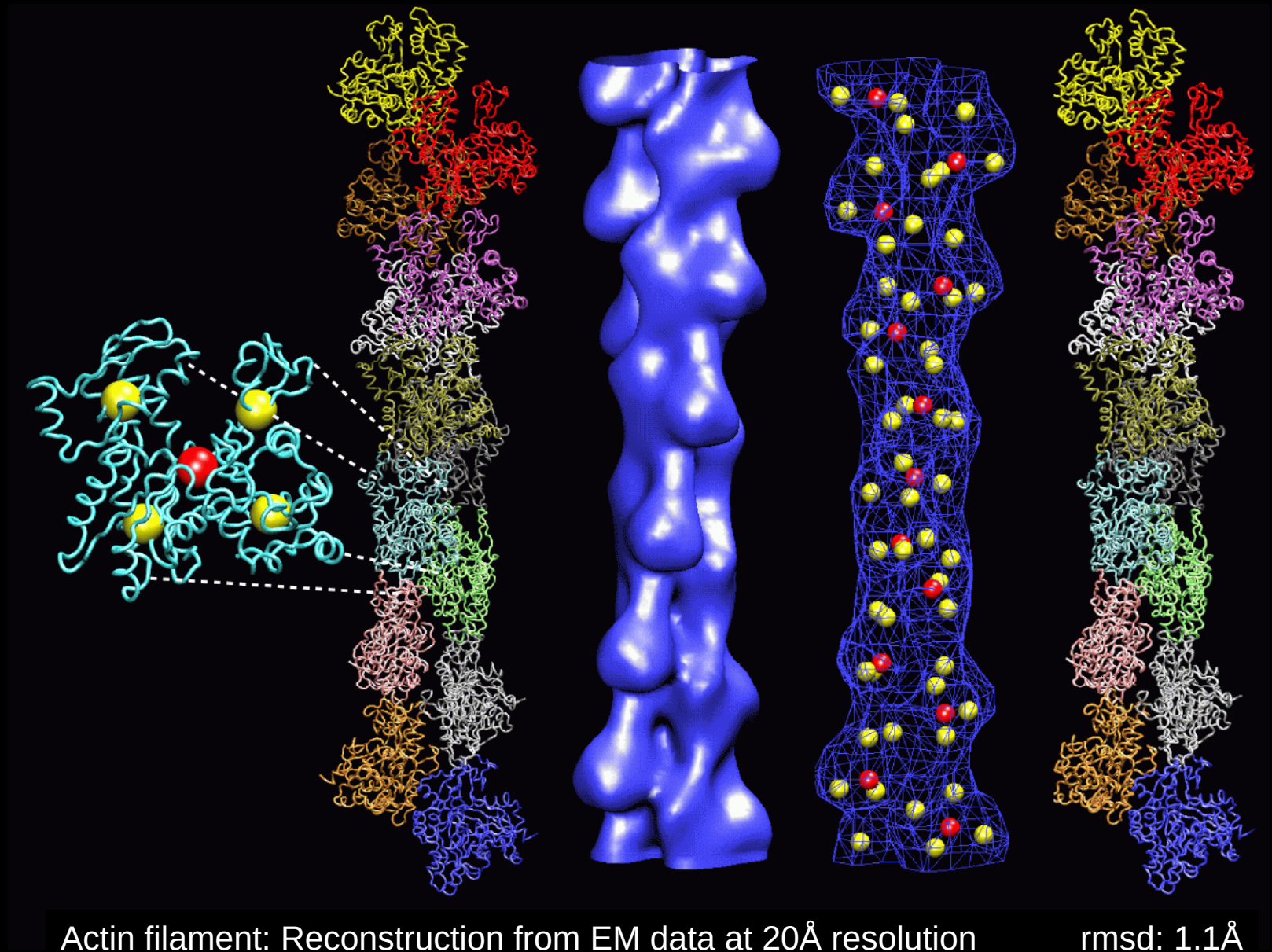
Laplacian filter



Increasing Fitting Contrast →

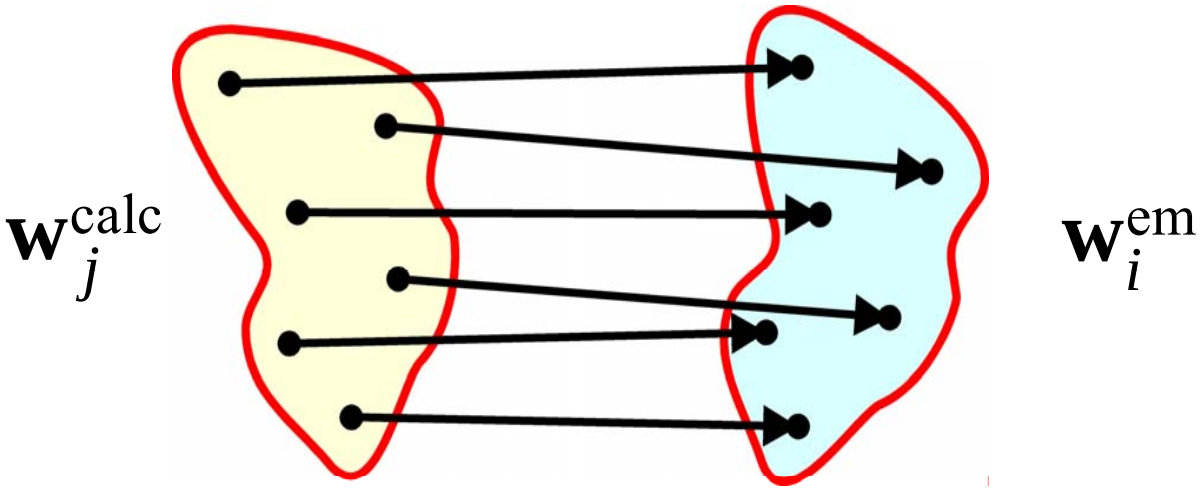
Questions?

Alternative: “Simulated Markers”



Actin filament: Reconstruction from EM data at 20Å resolution rmsd: 1.1Å

Reduced Representations of Biomolecular Structure



Feature points (fiducials, landmarks), reduce complexity of search space

Useful for:

- Rigid-body fitting
- Flexible fitting
- Interactive fitting / force feedback
- Building of deformable models

Vector Quantization

Lloyd (1957) } Digital Signal Processing,
Linde, Buzo, & Gray (1980) } Speech and Image Compression.
Martinetz & Schulten (1993) Topology-Representing Network.

Encode data (in $\Re^{d=3}$) using a finite set $\{w_j\}$ (j=1,…,k) of *codebook vectors*.

Delaunay triangulation divides $\Re^3$ into k *Voronoi polyhedra* ("receptive fields"):

$$V_i = \left\{ v \in \Re^3 \middle| \|v - w_i\| \leq \|v - w_j\| \forall j \right\}$$

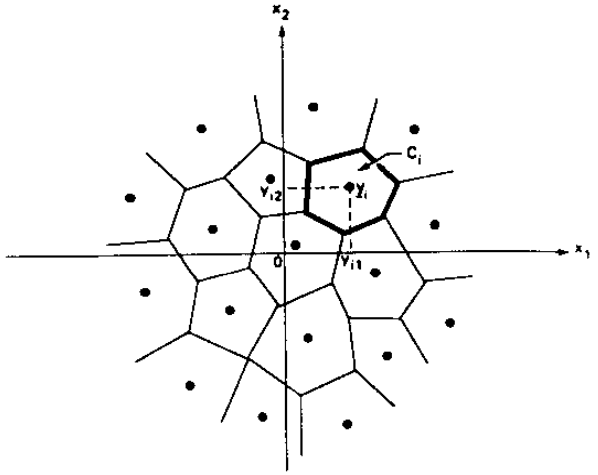
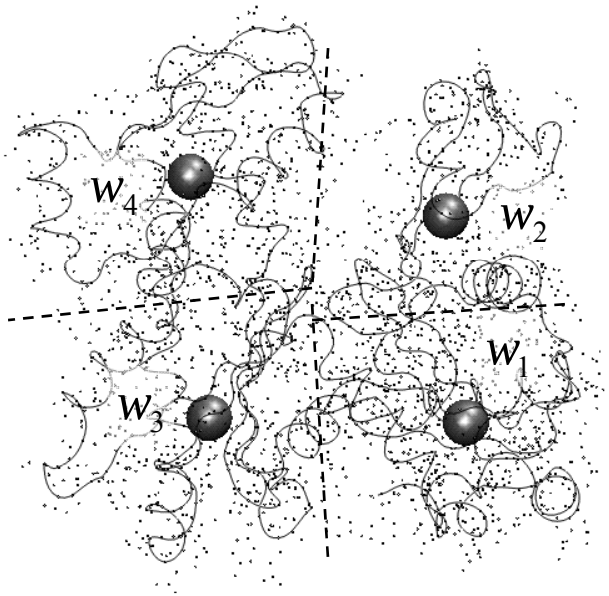
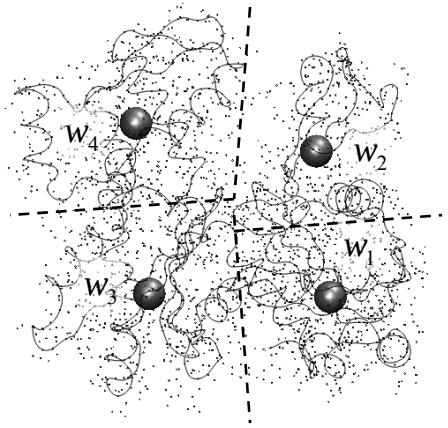


Fig. 3. Partitioning of two-dimensional space (N = 2) into L = 18 cells. All input vectors in cell C_i will be quantized as the code vector y_i. The shapes of the various cells can be very different.



Linde, Buzo, Gray (LBG) Algorithm



Encoding Distortion Error:

$$E = \sum_{\substack{i \text{ (atoms,} \\ \text{voxels)}}} \left\| v_i - w_{j(i)} \right\|^2 m_i$$

Lower $E\left(\{w_j(t)\}\right)$ iteratively: Gradient descent $\forall r$:

$$\Delta w_r(t) \equiv w_r(t) - w_r(t-1) = -\frac{\varepsilon}{2} \cdot \frac{\partial E}{\partial w_r} = \varepsilon \cdot \sum_i \delta_{rj(i)} (v_i - w_r) m_i \ .$$

Inline (Monte Carlo) approach for a sequence $v_i(t)$ selected at random according to weights m_i :

$$\Delta w_r(t) = \tilde{\varepsilon} \cdot \delta_{rj(i)} \cdot (v_i(t) - w_r).$$

How do we avoid getting trapped in the many local minima of E?

Soft-Max Adaptation

Avoid local minima by smoothing of energy function (here: TRN method):

$$\forall r: \ \Delta w_r(t) = \tilde{\varepsilon} \cdot e^{\frac{-s_r}{\lambda}} \cdot \left(v_i(t) - w_r\right),$$

Where $s_r\left(v_i(t), \{w_j\}\right)$ is the closeness rank:

$$\left\|v_i - w_{j0}\right\| \le \left\|v_i - w_{j1}\right\| \le \ldots \le \left\|v_i - w_{j(k-1)}\right\|$$

$$s_r = 0 \qquad s_r = 1 \qquad s_r = k-1$$

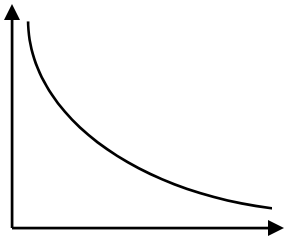
Note: $\lambda \to 0$: LBG algorithm.
$\lambda \neq 0$: not only "winner" $w_{j(i)}$ also second, third, ... closest are updated.

Can show that this corresponds to stochastic gradient descent on

$$\tilde{E}\left(\{w_j\}, \lambda\right) = \sum_{\mathrm{r}=1}^{\mathrm{k}} e^{\frac{-s_r}{\lambda}} \sum_i \left\|v_i - w_{j(i)}\right\|^2 m_i \, .$$

Note: $\lambda \to 0$: $\tilde{E} \to E$. LBG algorithm.
$\lambda \to \infty$: $\tilde{E}$ parabolic (single minimum).

$\Rightarrow \lambda(t)$



Convergence and Variability

Q: How do we know that we have found the global minimum of E?

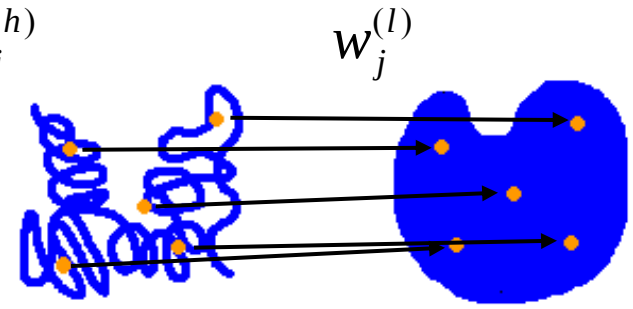
A: We don't (in general).

But we can compute the statistical variability of the $\{w_j\}$ by repeating the calculation with different seeds for random number generator.

Codebook vector variability arises due to:
- statistical uncertainty,
- spread of local minima.

A small variability indicates good convergence behavior.
Optimum choice of # of vectors k: variability is minimal.

Single-Molecule Rigid-Body Docking



Xtal structure

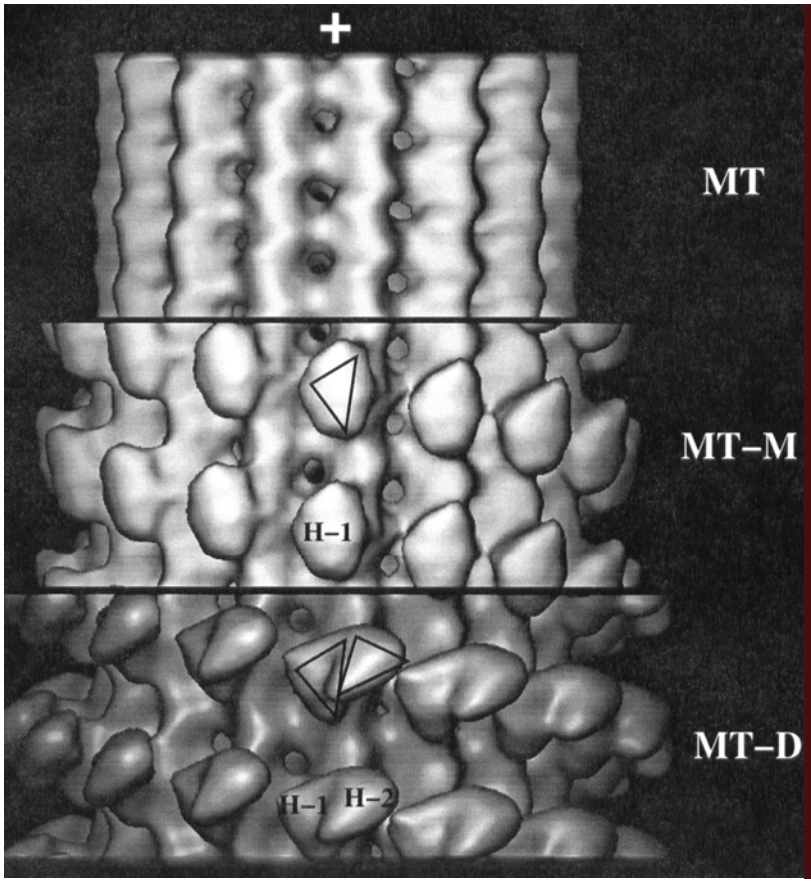
EM low res. data

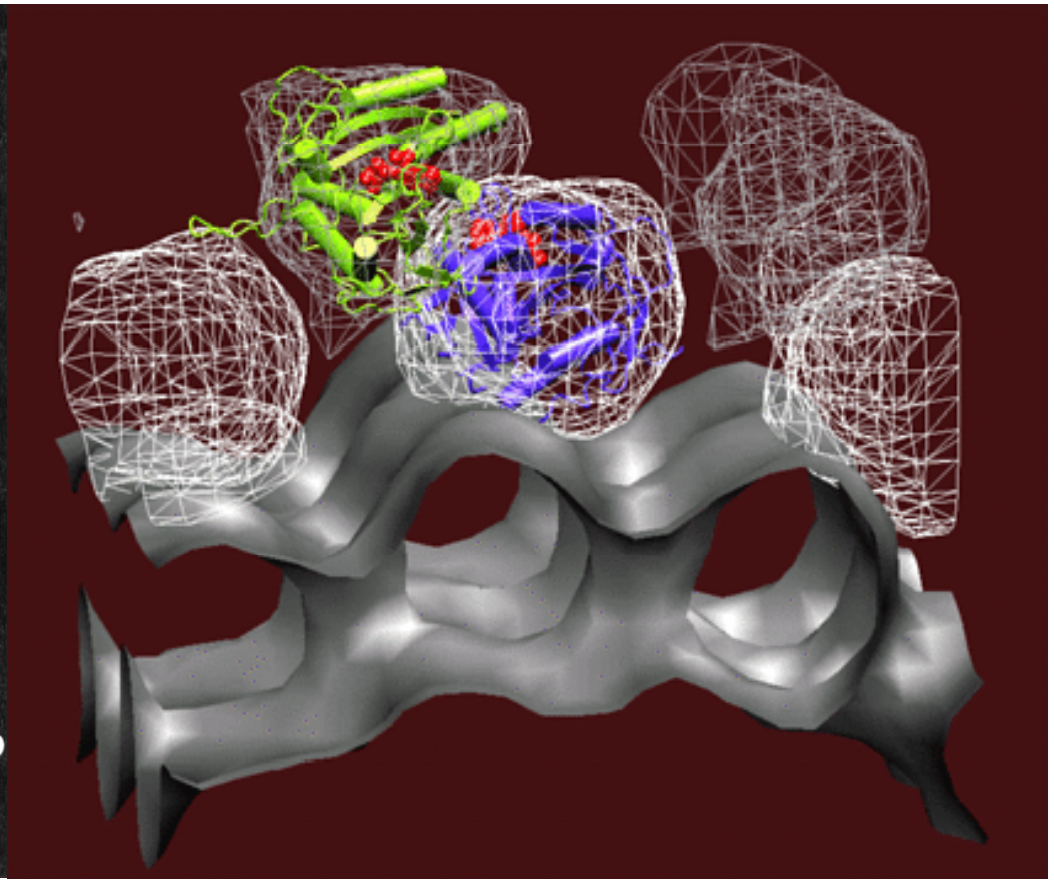
- Estimate optimum k with variability criterion.
- Index map I: $m \to n$ $(m, n = 1,\ldots,k)$.
- $k! = k\,(k\text{-}1)\ldots 2$ possible combinations.
- For each index map I perform a least squares fit of the $w_{I(j)}^{(h)}$ to the $w_j^{(l)}$.
- Quality of I: residual rms deviation

$$\Delta_I = \sqrt{\frac{1}{k}\sum_{j=1}^{k}\left\| w_{I(j)}^{(h)} - w_j^{(l)} \right\|^2}$$

- Find optimal I by direct enumeration of the $k!$ cases (minimum of Δ_I).

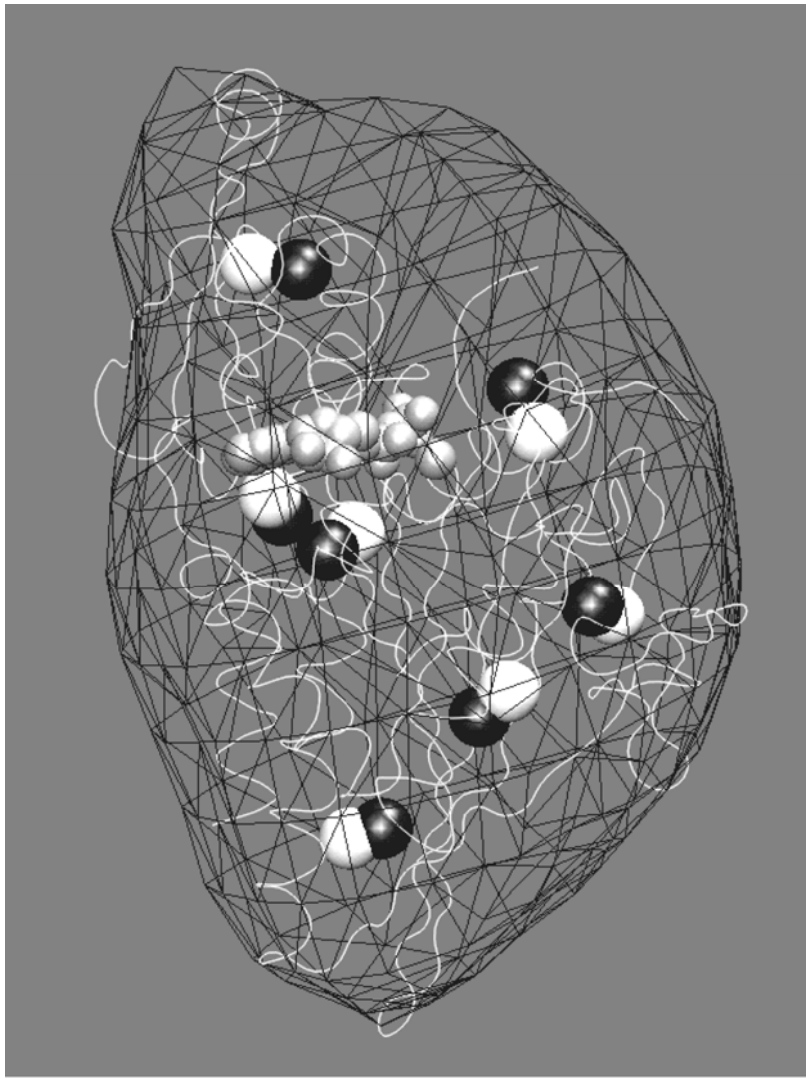
Application Example





ncd monomer and dimer-decorated microtubules (Milligan *et al.*, 1997)
ncd monomer crystal structure (Fletterick *et al.*, 1996,1998)

Search for Conformations



Two possible ranking criteria:

- Codebook vector rms deviation (Δ_I).
- Overlap between both data sets:

Voxel-Correlation coefficient:

$$C_{hl} = \frac{\sum_{x,y,z} h_{x,y,z} \cdot l_{x,y,z}}{\left(\sum_{x,y,z} h_{x,y,z}^2 \right)^{\frac{1}{2}} \left(\sum_{x,y,z} l_{x,y,z}^2 \right)^{\frac{1}{2}}}$$

ncd motor (white, shown with ATP nucleotide) docked to EM map (black) using k=7 codebook vectors

Reduced Search Features

Top 20, 7!=5040 possible pairs of codebook vectors.

	Δ_I	C_{hl}	I (permutation)
1.	3.115	0.913	(7,5,1,6,4,2,3)
2.	4.946	0.904	(2,3,5,7,4,6,1)
3.	5.455	0.897	(6,1,3,2,4,7,5)
4.	6.316	0.882	(5,7,4,3,1,2,6)
5.	7.612	0.867	(5,7,1,4,6,3,2)
6.	7.855	0.888	(3,2,4,1,5,6,7)
7.	7.994	0.884	(1,6,4,5,3,7,2)
8.	8.001	0.863	(6,1,4,3,5,2,7)
9.	8.192	0.888	(2,6,4,3,1,7,5)
10.	8.244	0.850	(7,5,6,2,1,3,4)
11.	8.298	0.881	(2,6,7,5,1,3,4)
12.	8.340	0.894	(6,2,4,1,3,5,7)
13.	8.481	0.867	(3,4,6,2,1,5,7)
14.	8.516	0.885	(2,3,4,5,1,7,6)
15.	8.532	0.857	(7,5,4,1,3,6,2)
16.	8.985	0.861	(6,1,5,7,4,3,2)
17.	8.988	0.838	(3,4,5,7,1,2,6)
18.	9.092	0.839	(3,2,5,4,7,1,6)
19.	9.124	0.858	(7,5,3,2,4,1,6)
20.	9.236	0.858	(1,6,5,7,4,2,3)

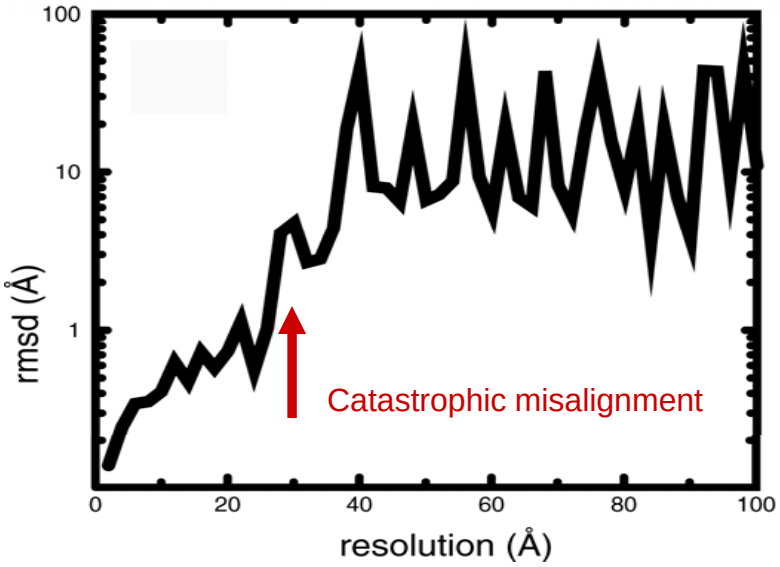
For a fixed k, codebook rmsd is more stringent criterion than correlation coefficient!

Performance (I)

Dependence on experimental EM density threshold (ncd, k=7):

orientations are stable:
+/- 5^{o} variability for +/-50% threshold density variation.
Threshold level can be optimized via radius of gyration of vectors.

Dependence on resolution (simulated EM map, automatic assignment of k from $3 \leq k \leq 9$):



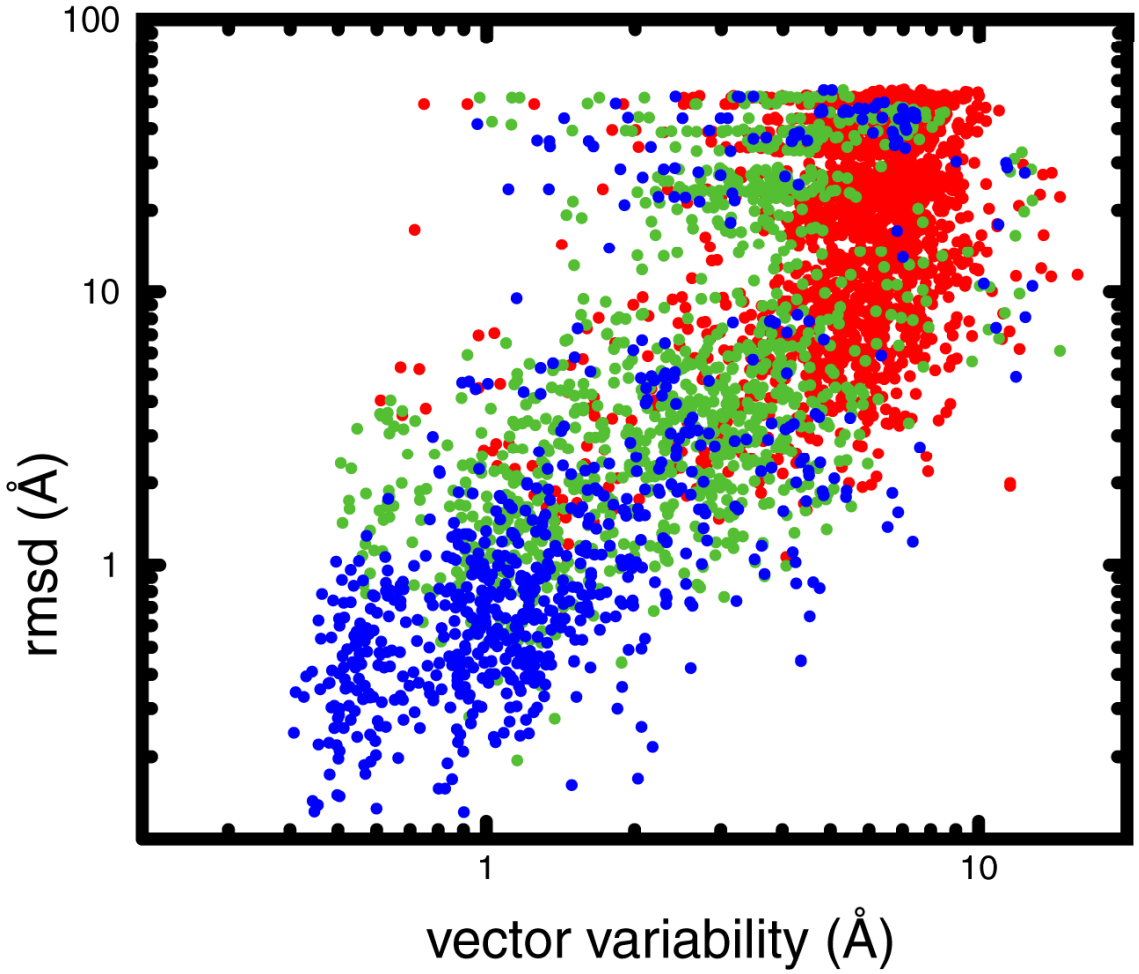
Deviation from start structure (PDB: 1TOP) used to generate simulated EM map.

Accurate matching up to ~30Å

Performance (II)

Is minimum vector variability a suitable choice for optimum k?

Wriggers & Birmanns, J. Struct. Biol 133, 193-202 (2001)

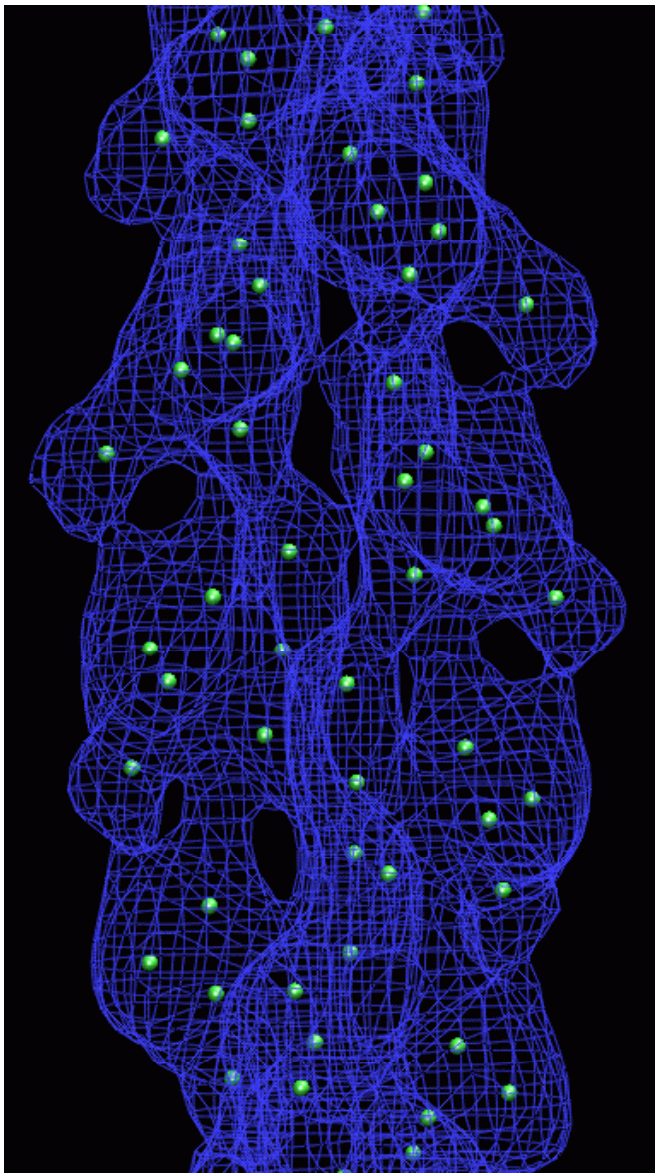


10 test systems, $3 \leq k \leq 9$ simulated EM densities from 2-100Å.

2-20Å (reliable fitting)
22-50Å (borderline)
52-100Å (mismatches)

Reasonable correlation with actual deviation

No “false positives” for resolution values < 20Å and variability < 1Å.



Performance (III)

Multiple Subunits

Egelman lab: High-resolution reconstructions of F-actin - plant ADF based on single-particle image processing.

Unrestrained vectors fail to distinguish between actin and ADF densities (poor segmentation)

Remedies:

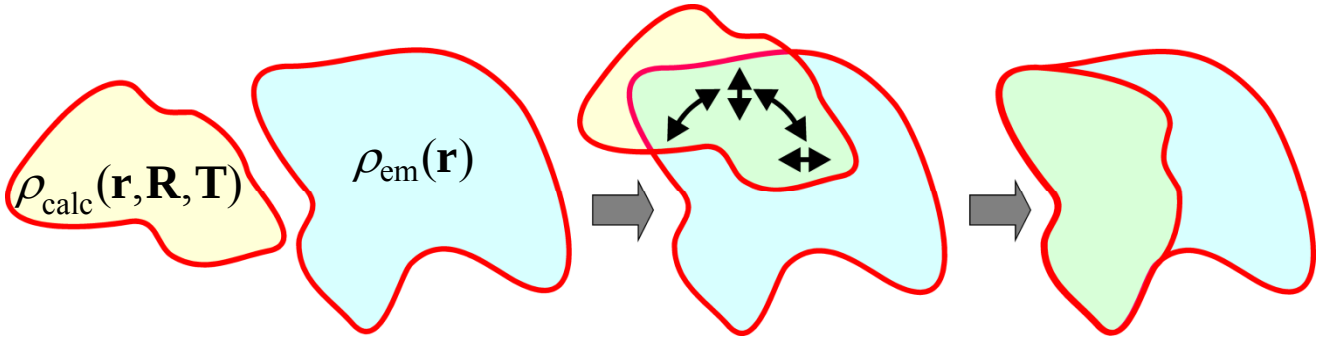
- Skeletons (later)
- Correlation-Based Search

Summary: Classic Situs (Versions 1.x)

Advantages of vector quantization:
•Fast (seconds of compute time).
•Reduced search is robust.

Limitations:
•No estimation of “fitting contrast” near optimum
•Works best for single molecules, not for matching subunits to larger densities.
•Largely superseded by FTM and FRM in Situs 2.x

Reduced Correlation Criterion



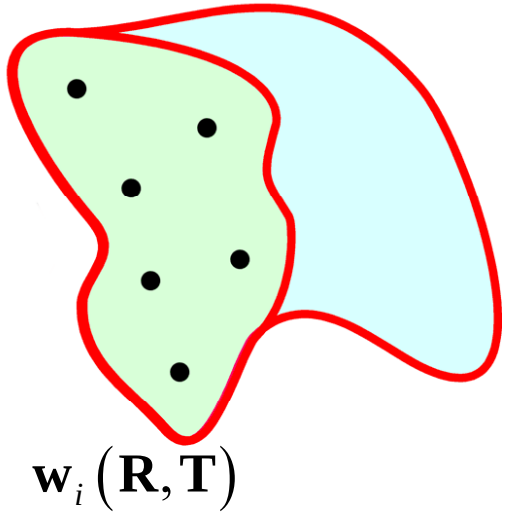
Correlation Coefficient

$$C(\mathbf{R},\mathbf{T}) \propto \int \rho_{\text{calc}}(\mathbf{r},\mathbf{R},\mathbf{T}) \cdot \rho_{\text{em}}(\mathbf{r})\, d^3r$$

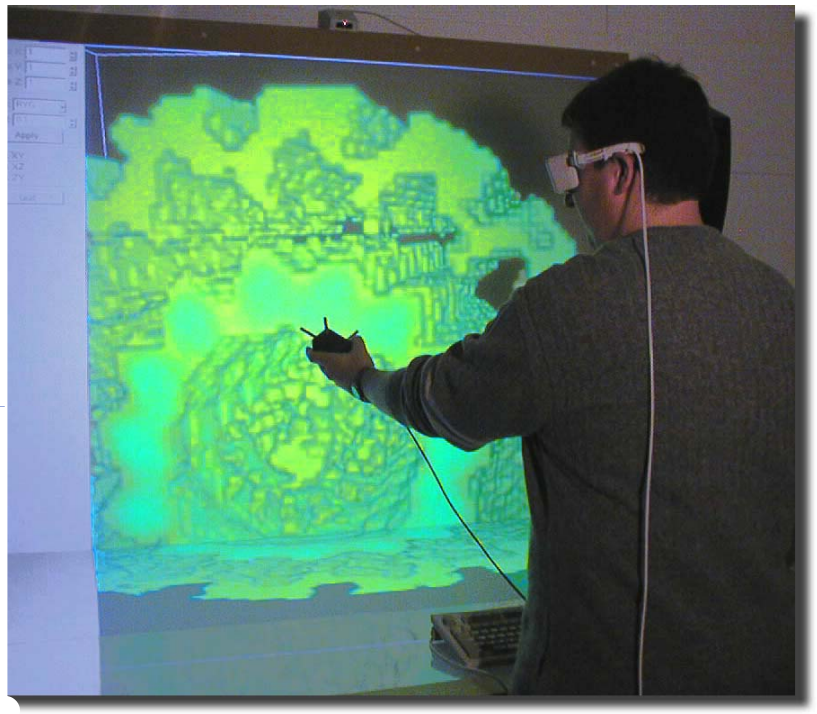
Reduced Model

$$\rho_{\text{calc}}(\mathbf{r}) \equiv \sum_{i=1}^{k} \delta(\mathbf{r} - \mathbf{w}_i)$$

$$C(\mathbf{R},\mathbf{T}) \propto \sum_{i=1}^{k} \rho_{\text{em}}\left(\mathbf{w}_i(\mathbf{R},\mathbf{T})\right)$$



Application in Haptic Rendering / VR

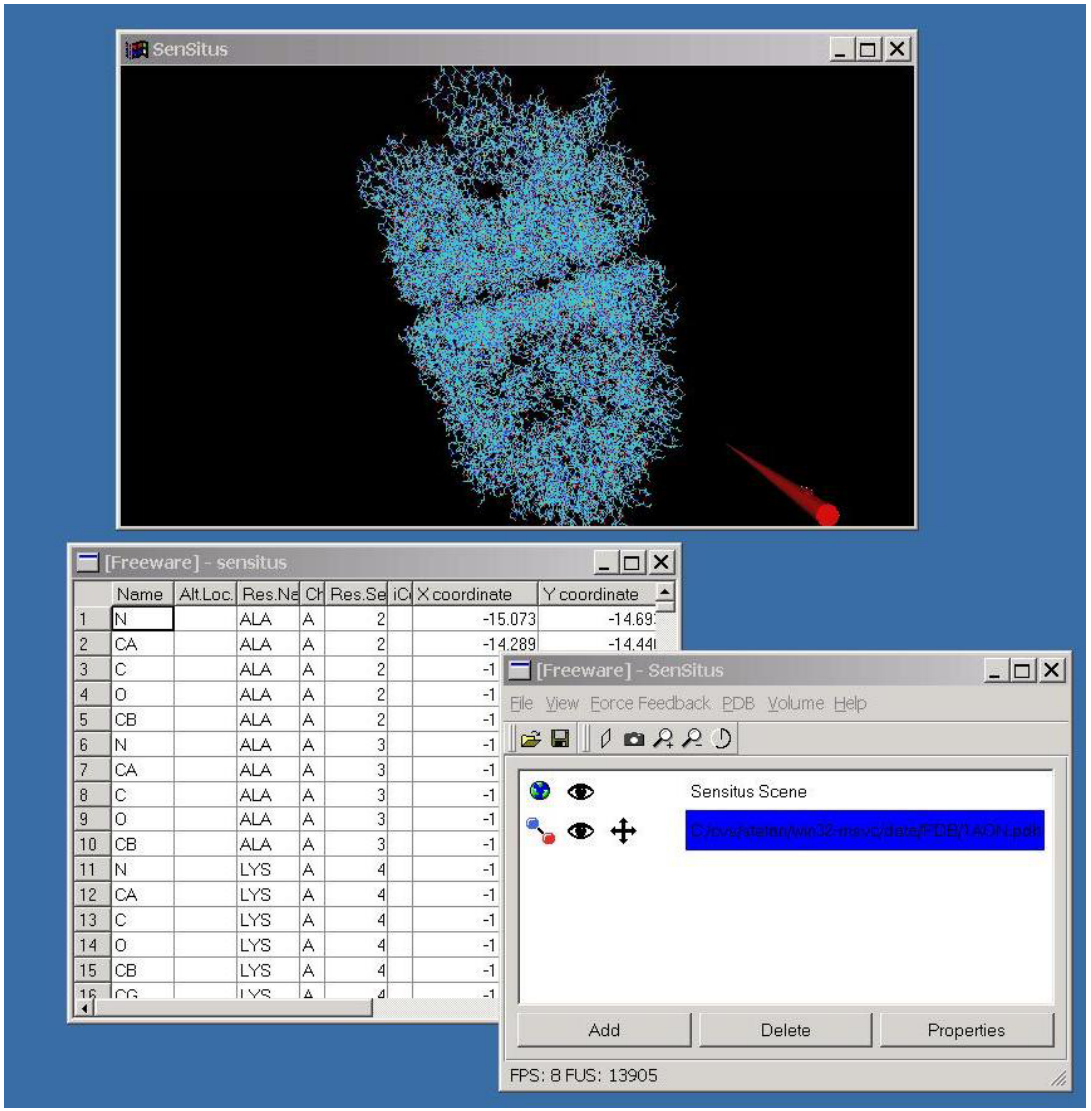


PHANTOM™ 1.5/6DOF
Haptic Device:
Force-Feedback in 6 DOF

Testing a Prototype of ***SenSitus***
in an immersive VR environment.

Force updates (100-1000 Hz) require a reduced model of probe structures.

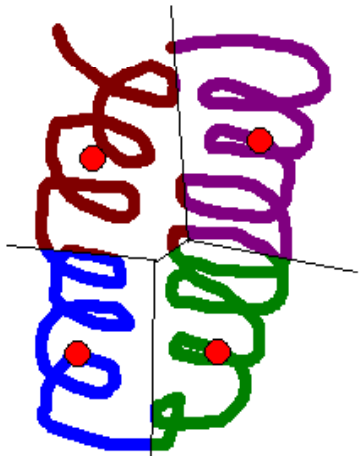
Interactive Modeling: *SenSitus*



(S. Birmanns)

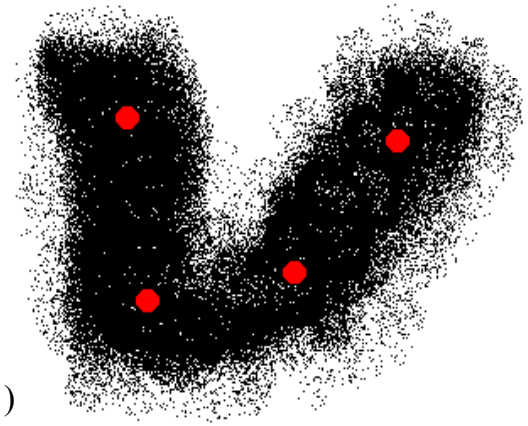
Application in Flexible Fitting

Xtal structure



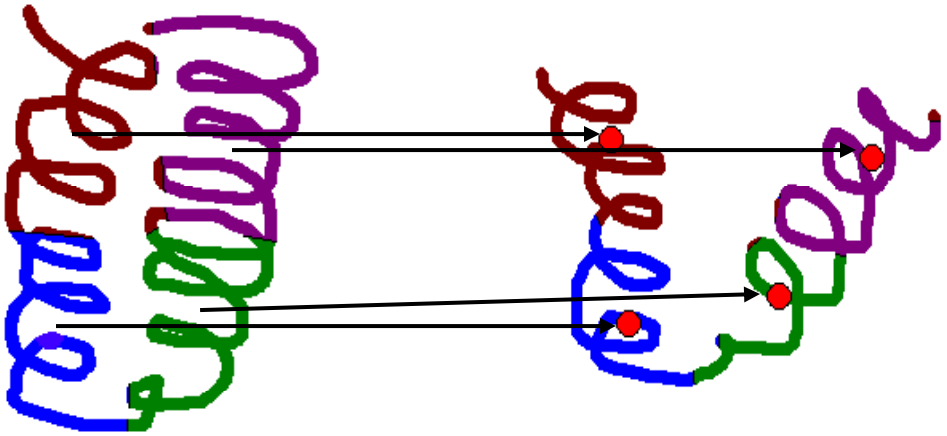
$w_j^{(h)}$

$w_j^{(l)}$



3D EM reconstruction

constrain centroids



molecular dynamics simulation

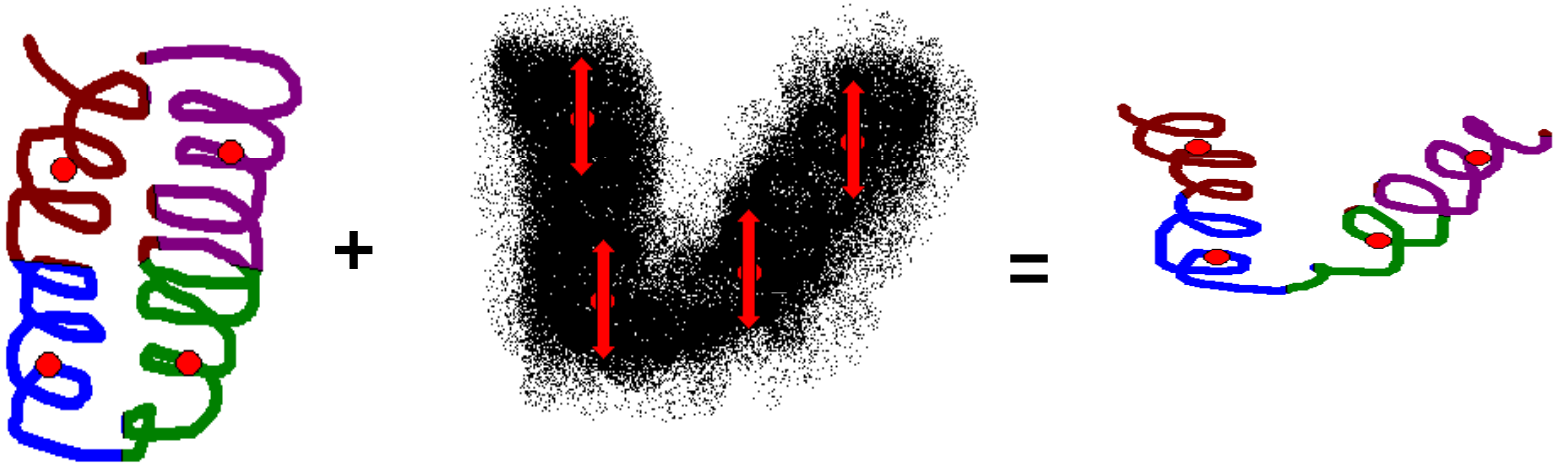
Stereochemical Quality of Flexible Fitting

The atomic model has many more degrees of freedom than there are independent pieces of information in the EM map. Hence, there is the danger that over fitting distorts the structure

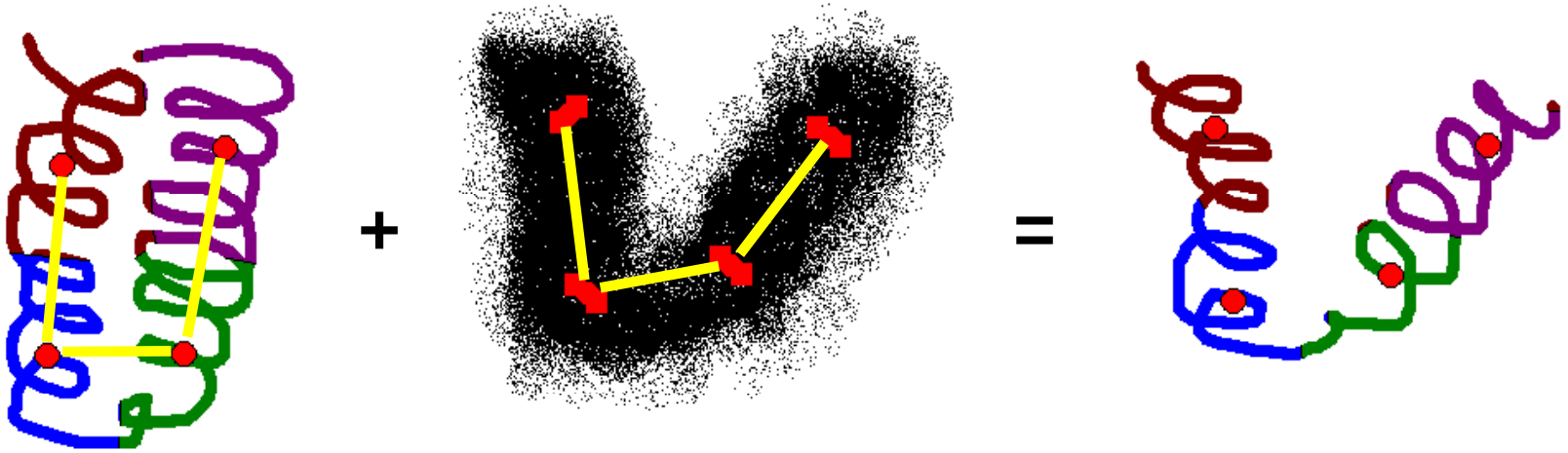
How can over fitting be avoided? Reduce noise by eliminating “inessential” degrees of freedom!…

Skeletons Limit the Effect of Noise:

freezing inessential degrees of freedom:

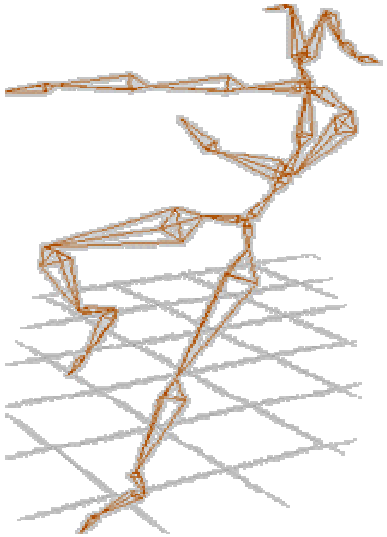


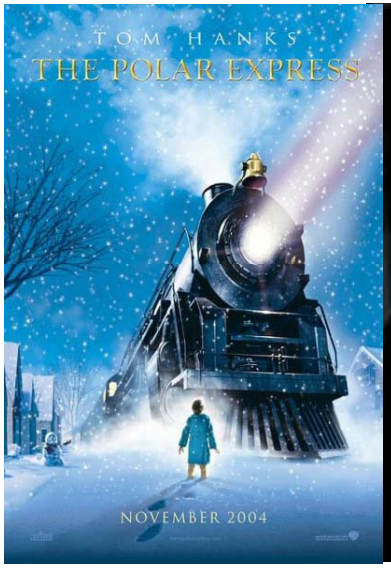
unrestrained markers exp. and meth. uncertainty distortion



skeleton distance constraints less distortion

Fitting Skeletons: Motion Capture



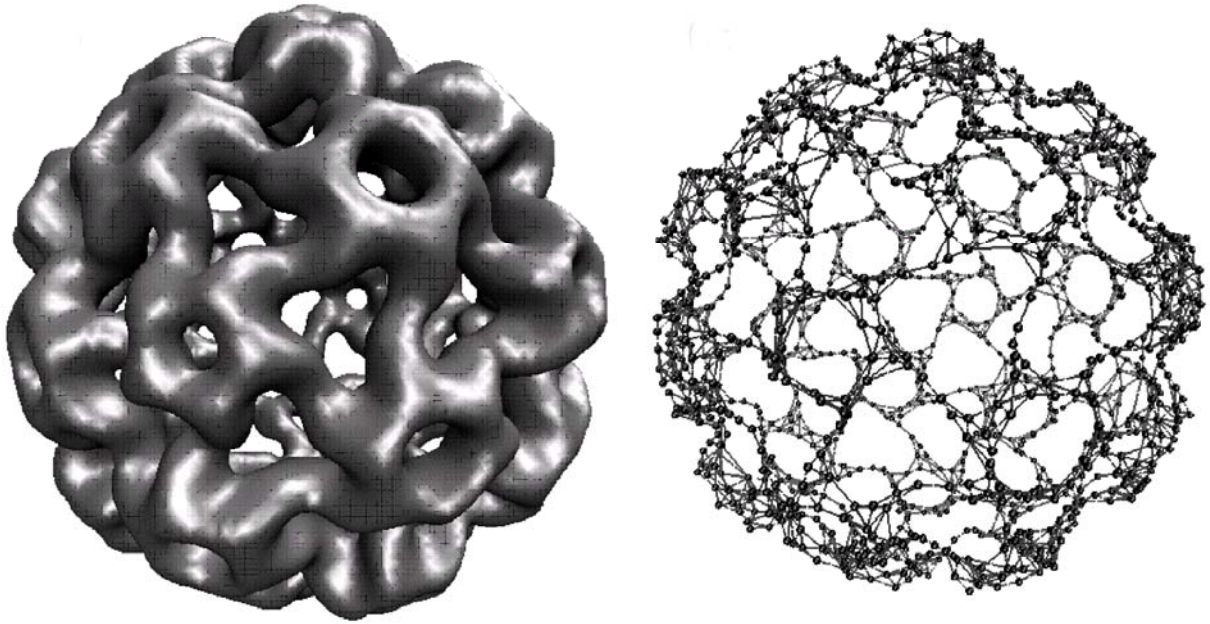


Motion Capture Network

Topology Representing Neural Network
(Martinetz and Schulten, 1993)

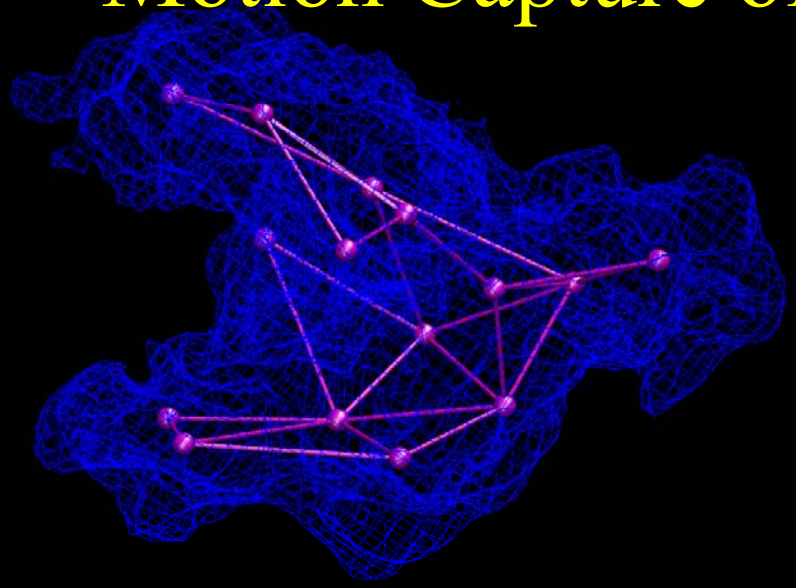
+

SHAKE Distance Constraints
(van Gunsteren, 1977)

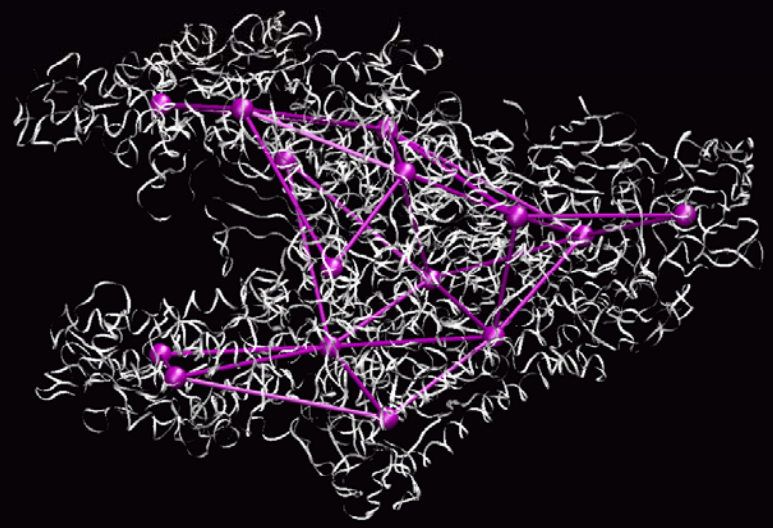


Neurocomputing (2004) 56:365

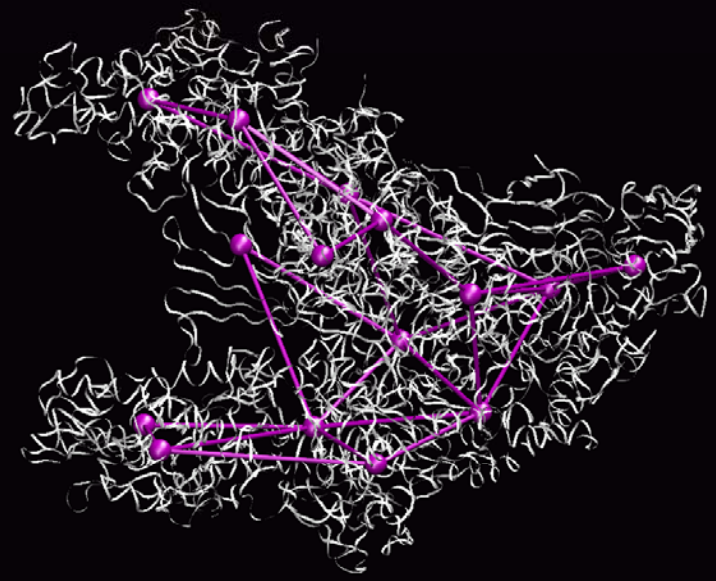
Motion Capture of RNA Polymerase



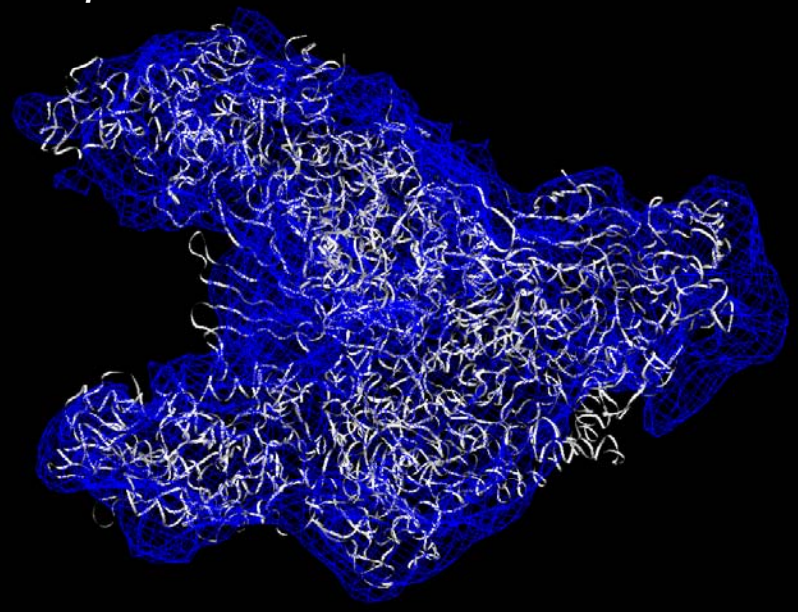
Taq-like single molecule map



Taq RNAP x-tal structure

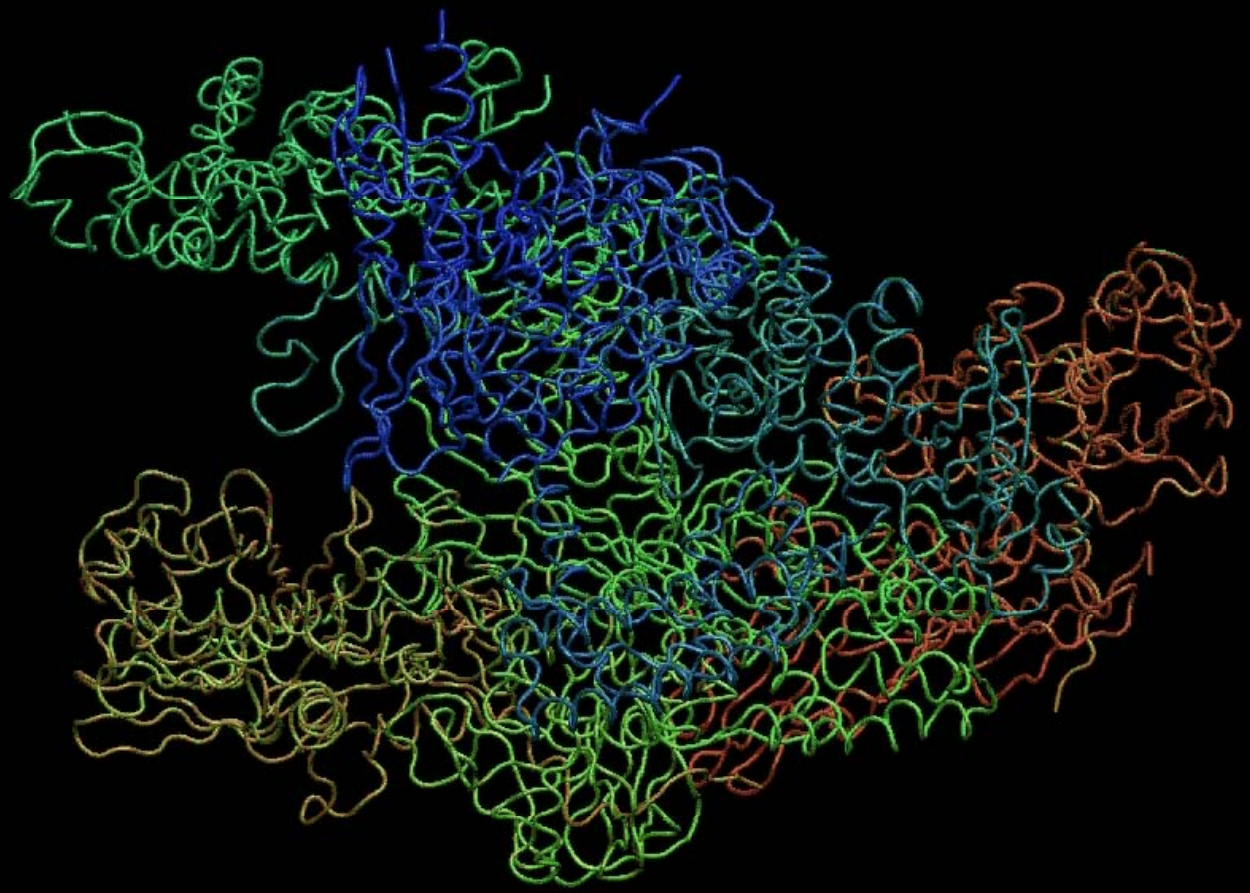


flexible fitting (15 vectors)



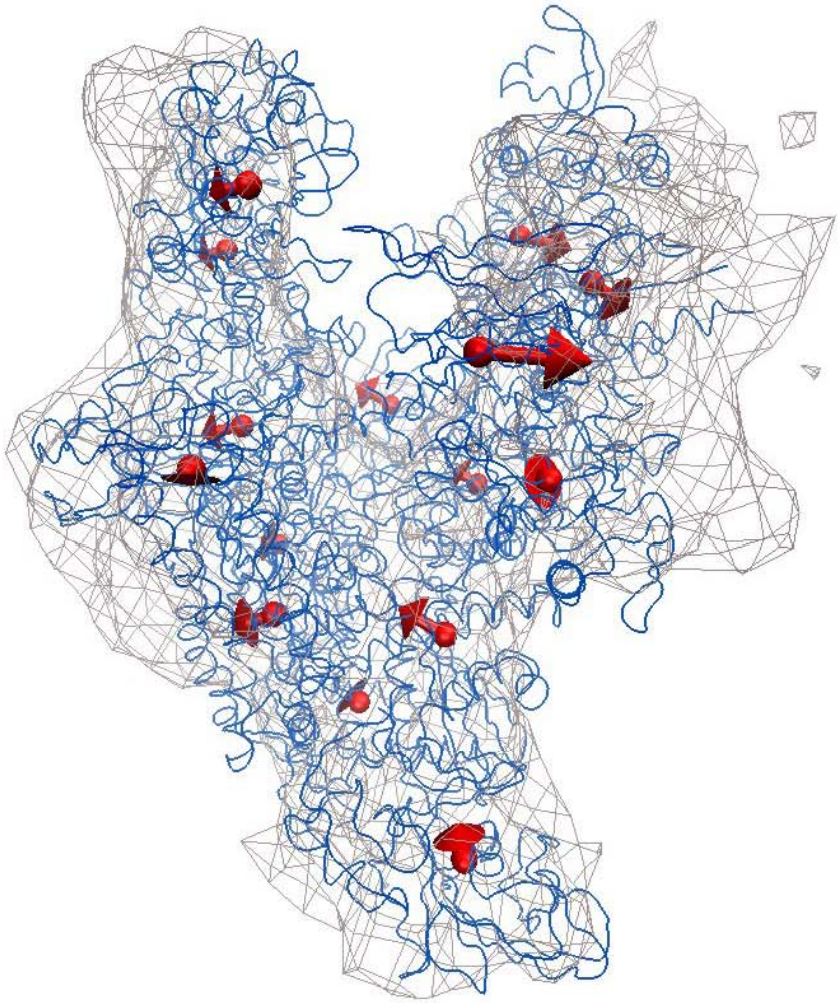
final result

Domain Motions

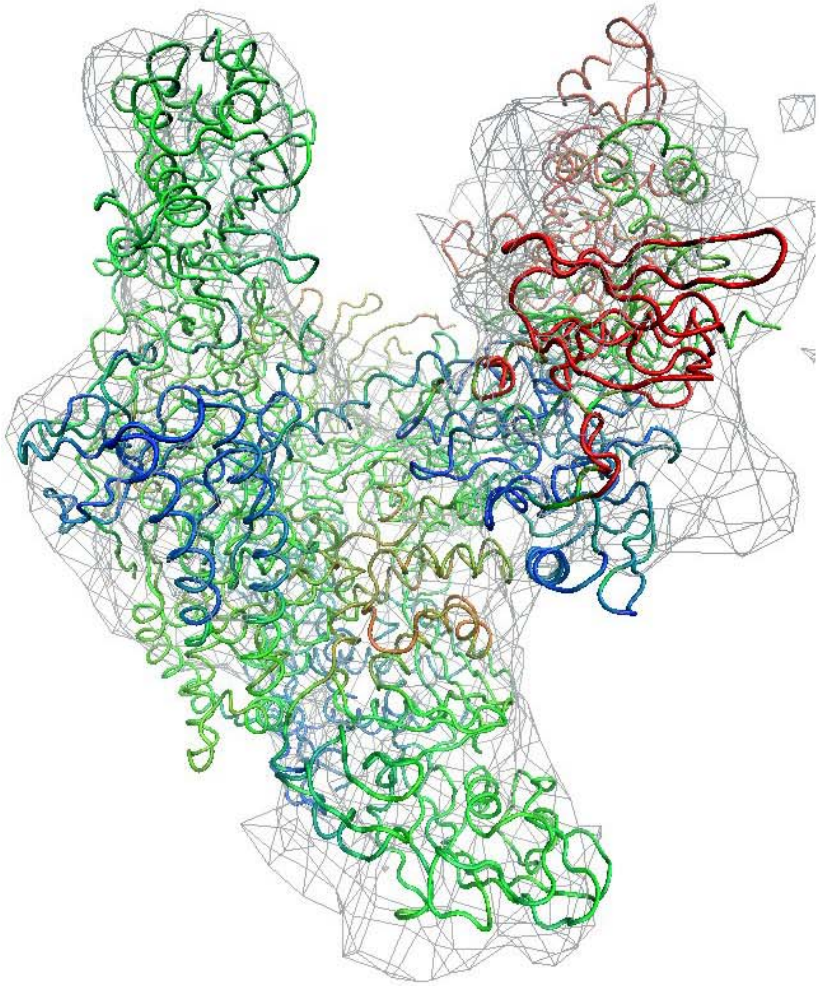


Flexing of the RNAP “jaws” suggests a jaw-closing in presence of DNA
PNAS (2002) 99:4296
Cell (2003) 114:335

What Information is Used?



Displacements



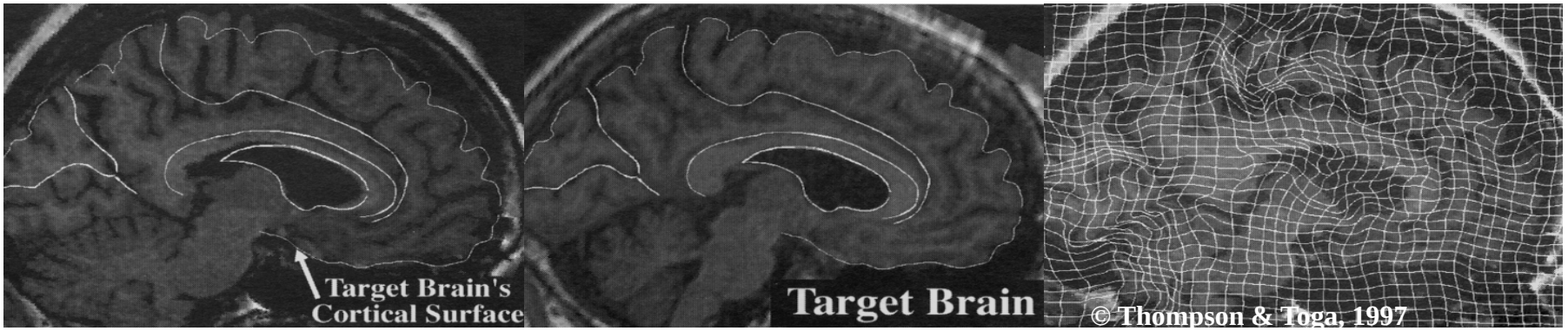
Molecular Dynamics

Molecular Dynamics vs. Interpolation

MD simulation requires an expert user and hours of preparation. We know a sparse estimation of the displacement field at markers. Can we extend the sparse estimate to the full space by an inexpensive interpolation?

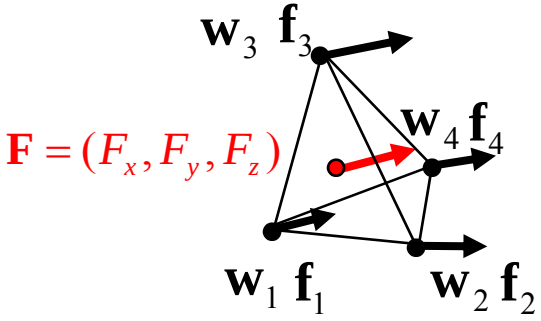
Interpolation Pros:

- Ease of use / implementation
- Detailed mass rearrangement plan.
- Linear or nonlinear registration of features
- Used in neuroscience and machine vision:



(i) Piecewise-Linear Inter- / Extrapolation

For each probe position find 4 closest vectors.



Ansatz: $F_x(x, y, z) = ax + by + xz + d$

$$F_x(\mathbf{w}_1) = f_{1,x},$$
$$F_x(\mathbf{w}_2) = f_{2,x},$$
$$F_x(\mathbf{w}_3) = f_{3,x},$$
$$F_x(\mathbf{w}_4) = f_{4,x} \quad \text{(similar for } F_y, F_z\text{)}.$$

Cramer's rule:

$$a = \frac{\begin{vmatrix} f_{1,x} & w_{1,y} & w_{1,z} & 1 \\ f_{2,x} & w_{2,y} & w_{2,z} & 1 \\ f_{3,x} & w_{3,y} & w_{3,z} & 1 \\ f_{4,x} & w_{4,y} & w_{4,z} & 1 \end{vmatrix}}{D}, \quad b = \frac{\begin{vmatrix} w_{1,x} & f_{1,y} & w_{1,z} & 1 \\ w_{2,x} & f_{2,y} & w_{2,z} & 1 \\ w_{3,x} & f_{3,y} & w_{3,z} & 1 \\ w_{4,x} & f_{4,y} & w_{4,z} & 1 \end{vmatrix}}{D}, \quad \cdots, \quad D = \begin{vmatrix} w_{1,x} & w_{1,y} & w_{1,z} & 1 \\ w_{2,x} & w_{2,y} & w_{2,z} & 1 \\ w_{3,x} & w_{3,y} & w_{3,z} & 1 \\ w_{4,x} & w_{4,y} & w_{4,z} & 1 \end{vmatrix}$$

(ii) Non-Linear Kernel Interpolation

Consider all k vectors and interpolation kernel function $U(r)$.

Ansatz:

$$F_x(x,y,z) = a_1 + a_x x + a_y y + a_z z + \sum_{k=1}^{k} b_i \cdot U\left(\left|\mathbf{w}_i - (x,y,z)\right|\right)$$

$$F_x(\mathbf{w}_i) = f_{i,x}, \ \forall i \quad (\text{similar for } F_y, F_z).$$

Solve :

$$\mathrm{L}^{-1}(f_{1,x},\cdots,f_{k,x},0,0,0,0) = (b_1,\cdots,b_k,a_1,a_x,a_y,a_z)^{\mathrm{T}},$$

$$\text{where} \quad \mathbf{L} = \left(\begin{array}{c|c} \mathbf{P} & \mathbf{Q} \\ \hline \mathbf{Q}^{\mathrm{T}} & \mathbf{0} \end{array}\right), \quad \mathbf{Q} = \begin{pmatrix} 1 & w_{1,x} & w_{1,y} & w_{1,z} \\ \cdots & \cdots & \cdots & \cdots \\ 1 & w_{k,x} & w_{k,y} & w_{k,z} \end{pmatrix}, \ k \times 4,$$

$$\mathbf{P} = \begin{pmatrix} 0 & U(w_{12}) & \cdots & U(w_{1k}) \\ U(w_{21}) & 0 & \cdots & U(w_{2k}) \\ \cdots & \cdots & \cdots & \cdots \\ U(w_{k1}) & U(w_{k2}) & \cdots & 0 \end{pmatrix}, \ k \times k.$$

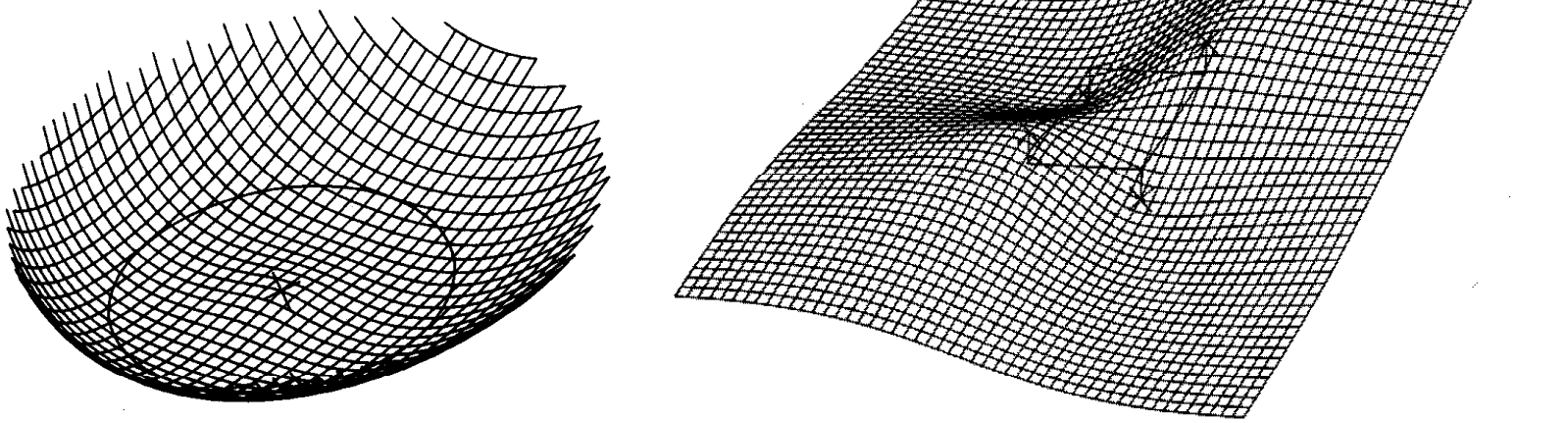
Bookstein "Thin-Plate" Splines

• kernel function $U(r)$ is principal solution of biharmonic equation that arises in elasticity theory of thin plates:

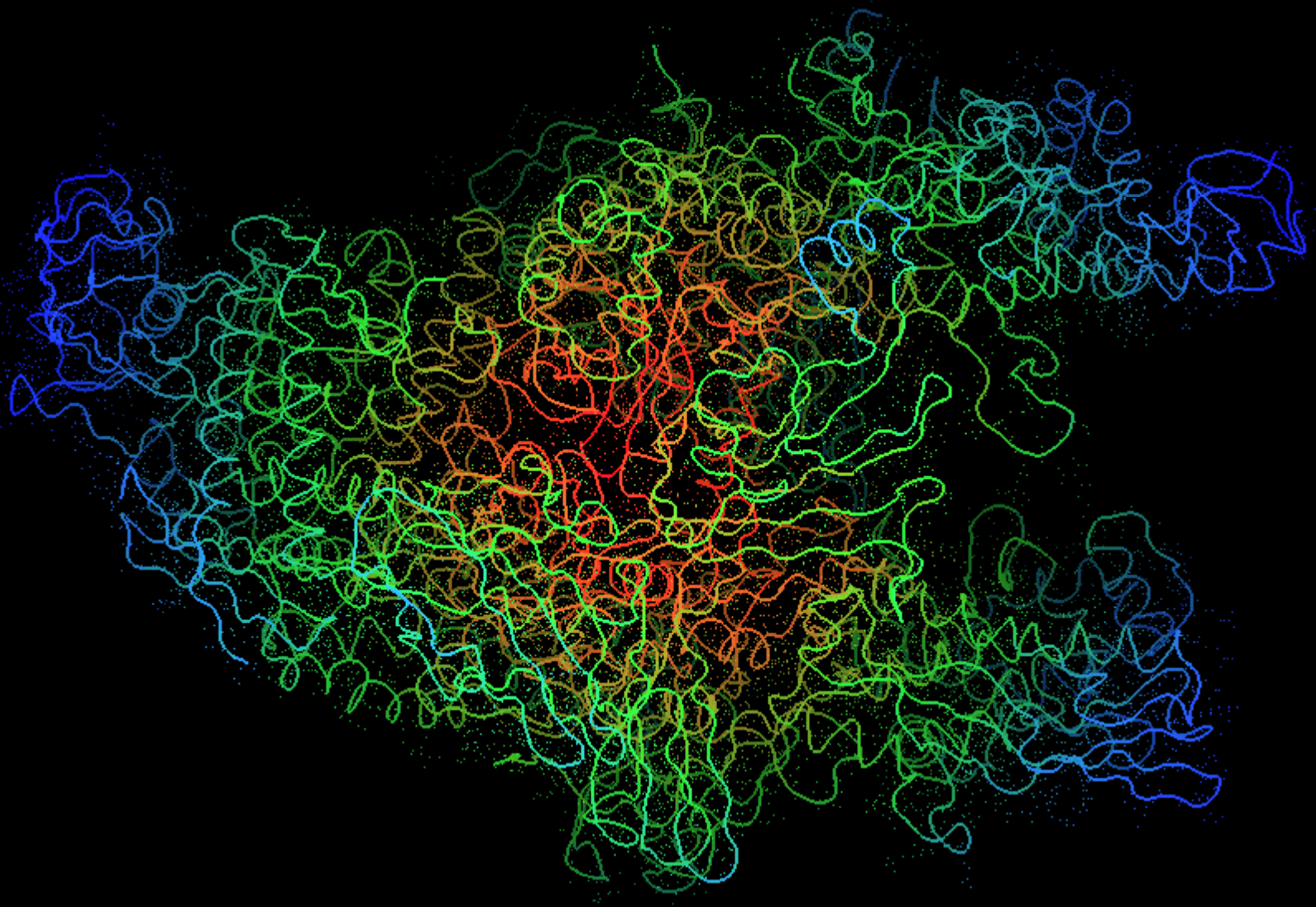
$$\Delta^2 U(r) = \nabla^4 U(r) = \delta(r).$$

• variational principle: $U(r)$ minimizes the bending energy (not shown).
• 1D: $U(r) = |r^3|$ (cubic spline)
• 2D: $U(r) = r^2 \log r^2$
• 3D: $U(r) = |r|$

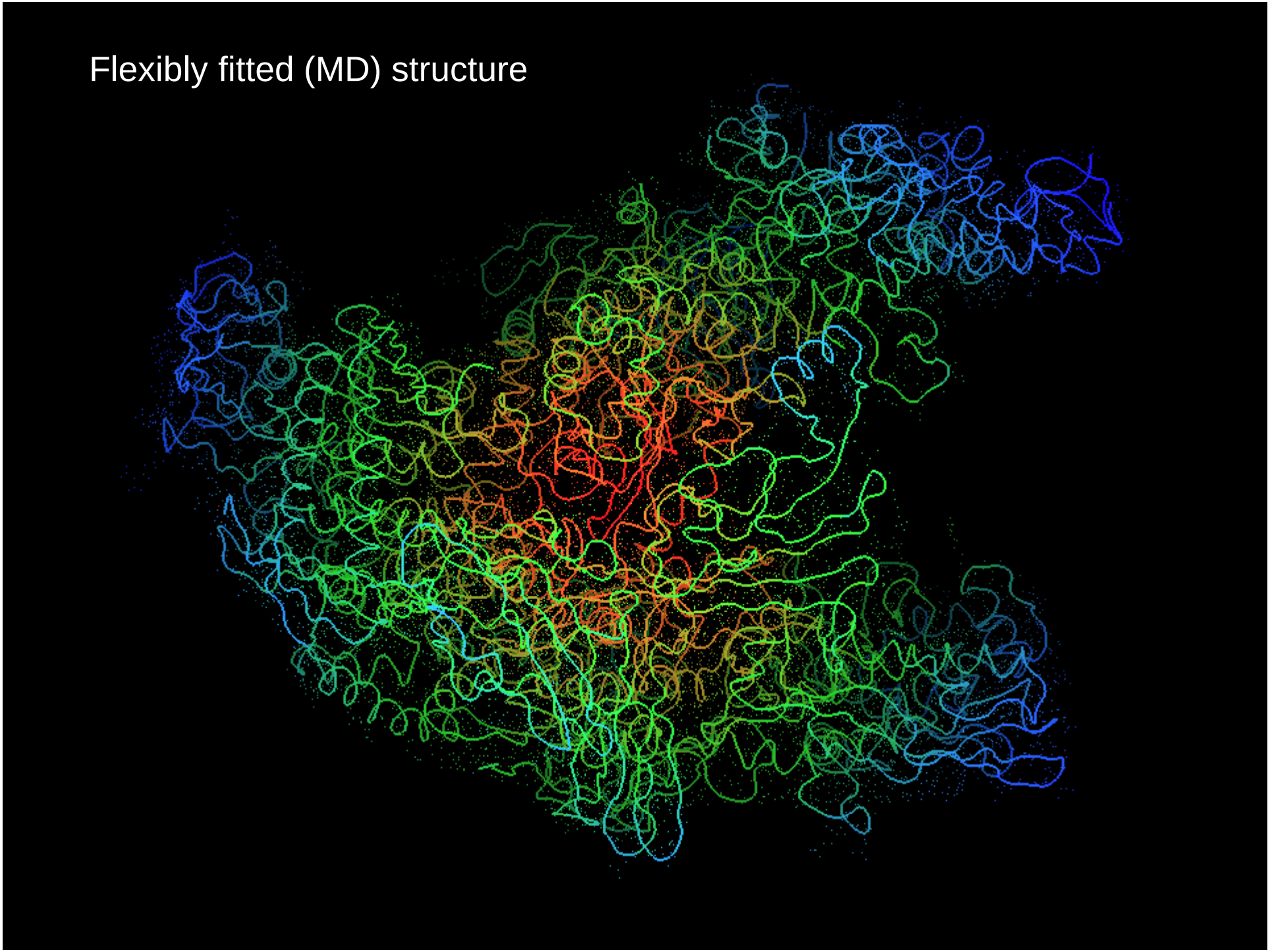
2D: $U(r)$ $\mathbf{F}(x, y)$



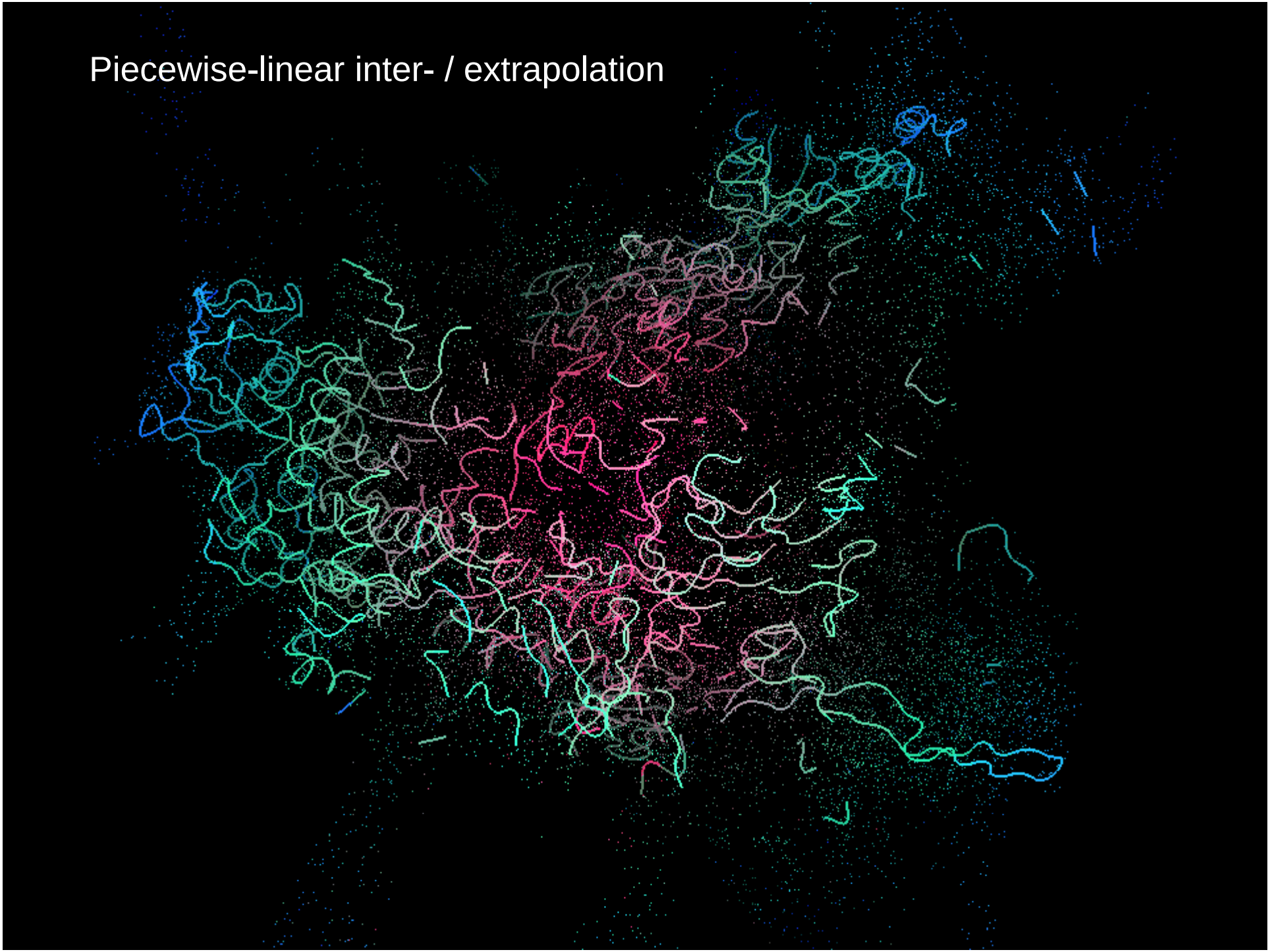
Taq RNAP x-tal structure



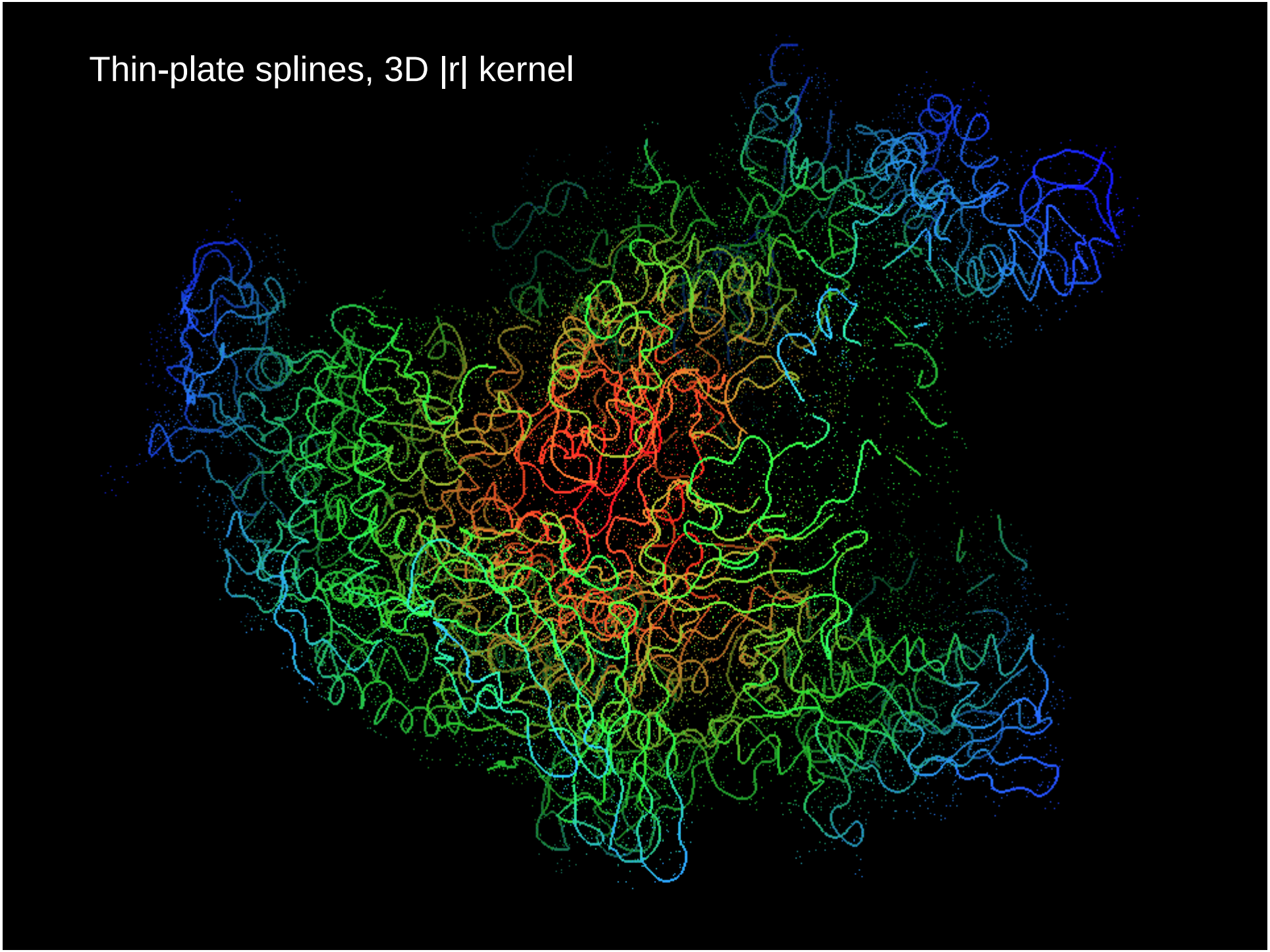
Flexibly fitted (MD) structure



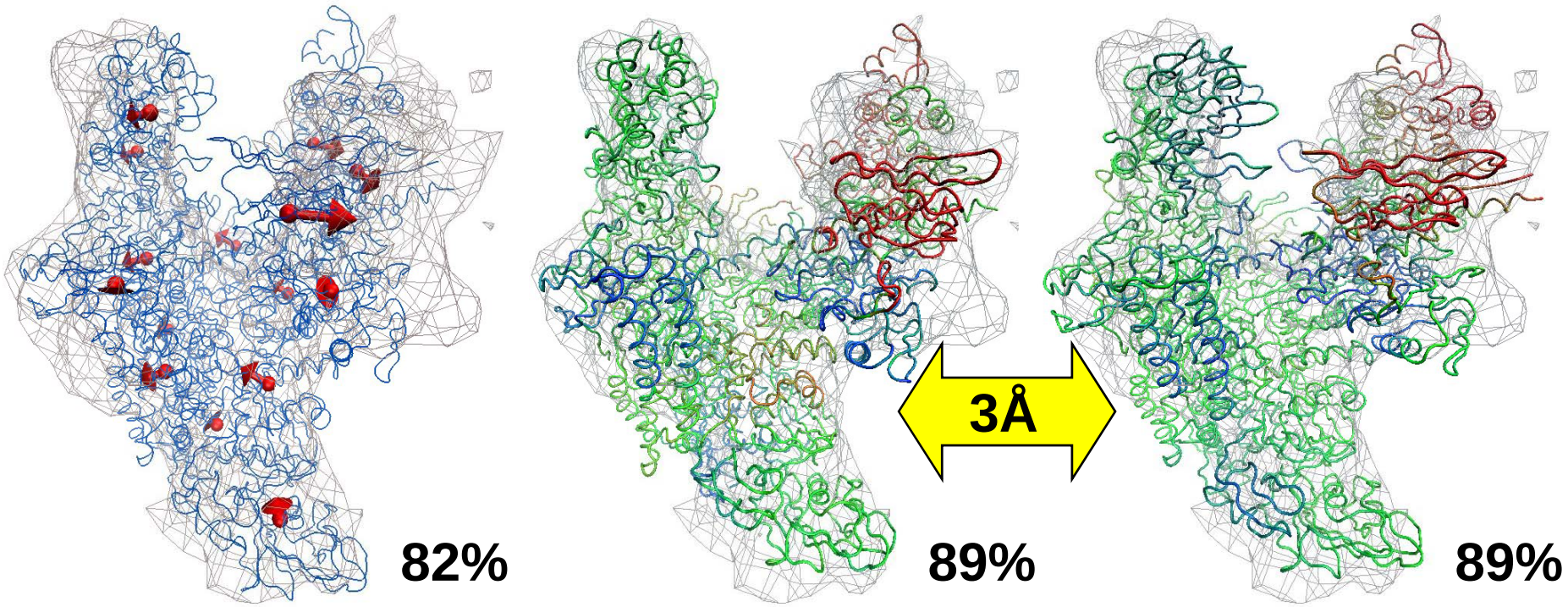
Piecewise-linear inter- / extrapolation



Thin-plate splines, 3D |r| kernel



MD vs. Thin Plate Splines

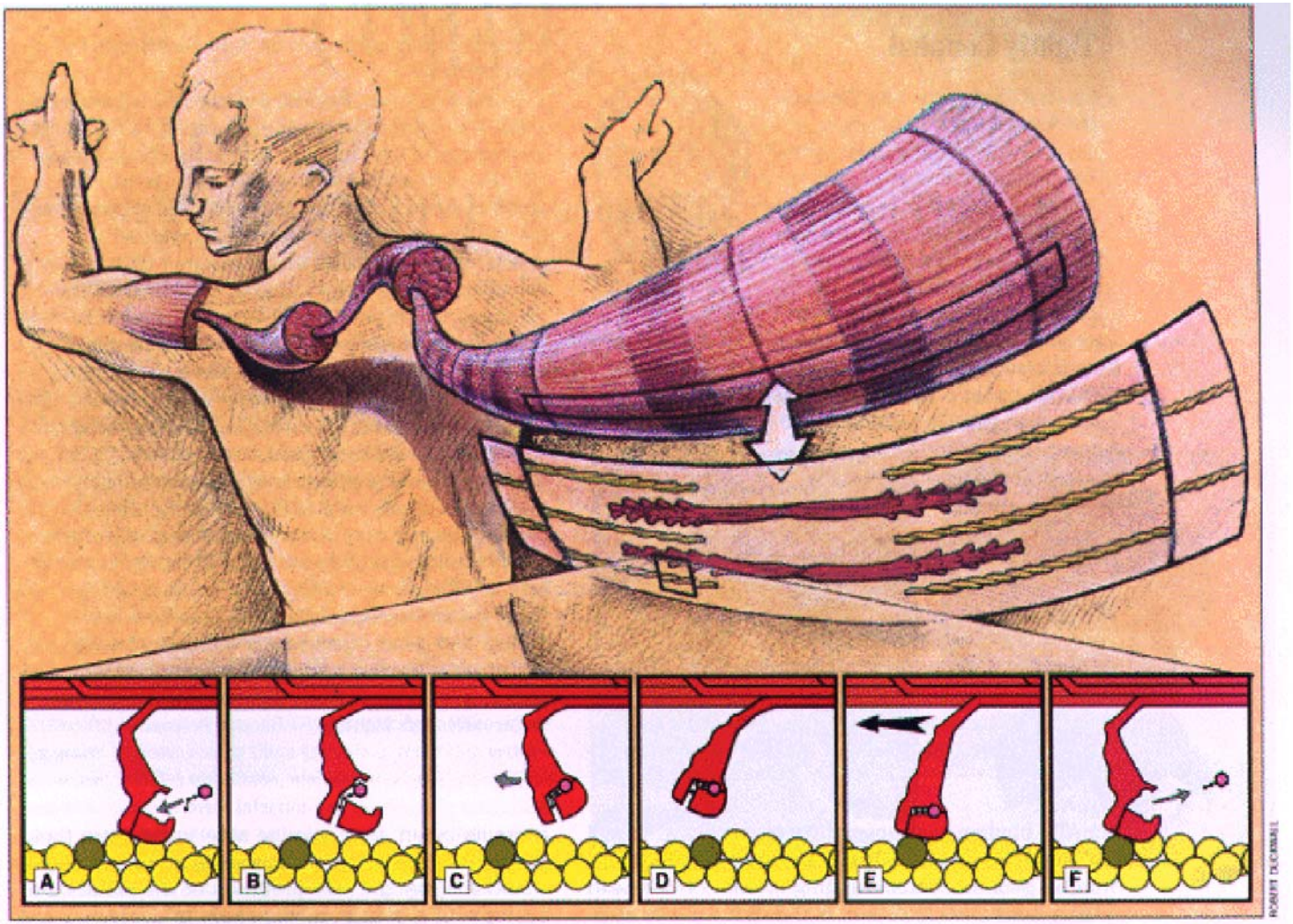


Displacements Molecular Dynamics Thin-Plate Splines

How do we know MD is really better?

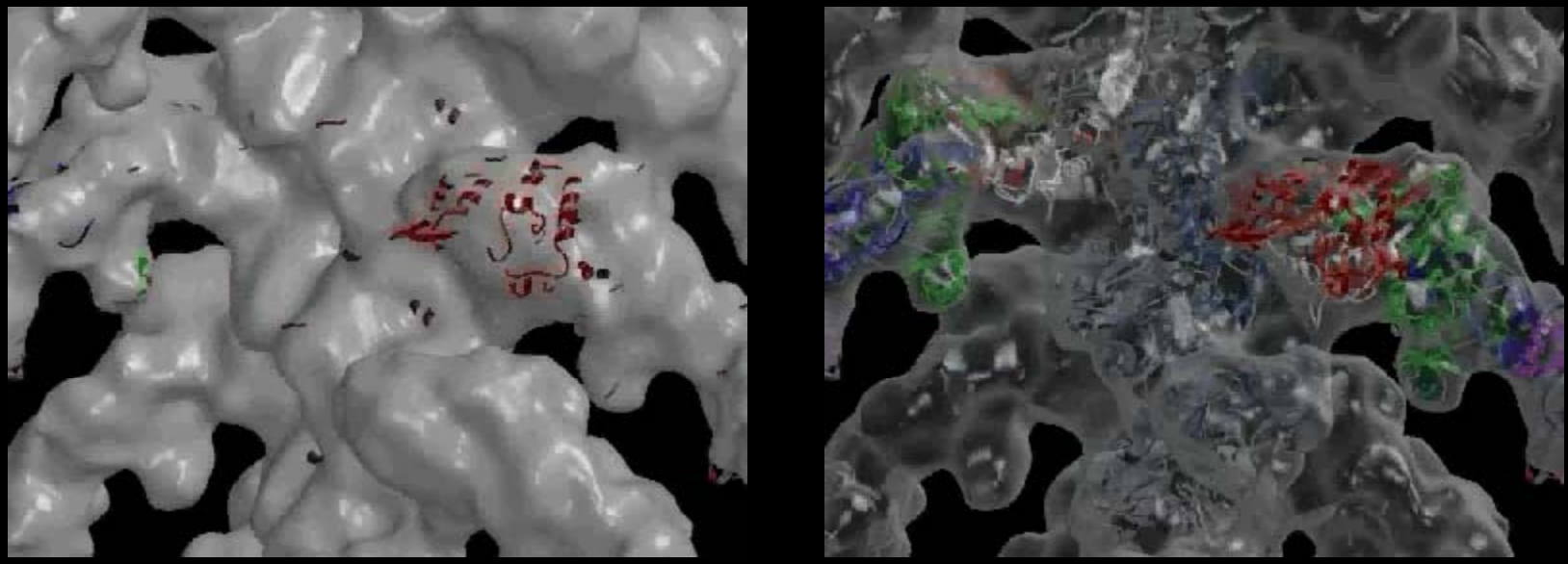
Structure (2004) 12:1

Validation Example: Muscle Contraction



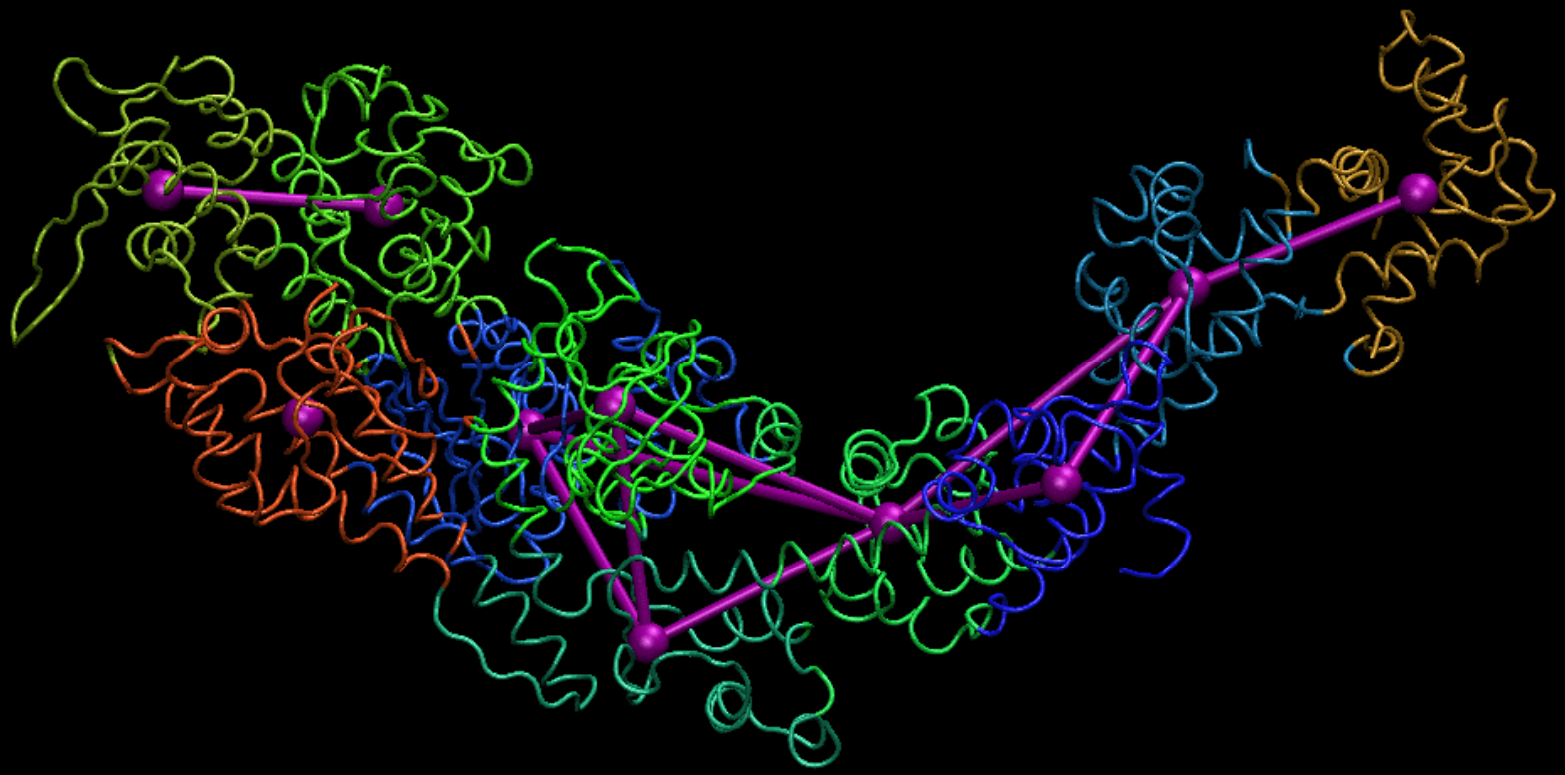
A Hierarchy of Muscle Structure, J. NIH Res. 1993

Acto-Myosin (II) Complex at 14Å

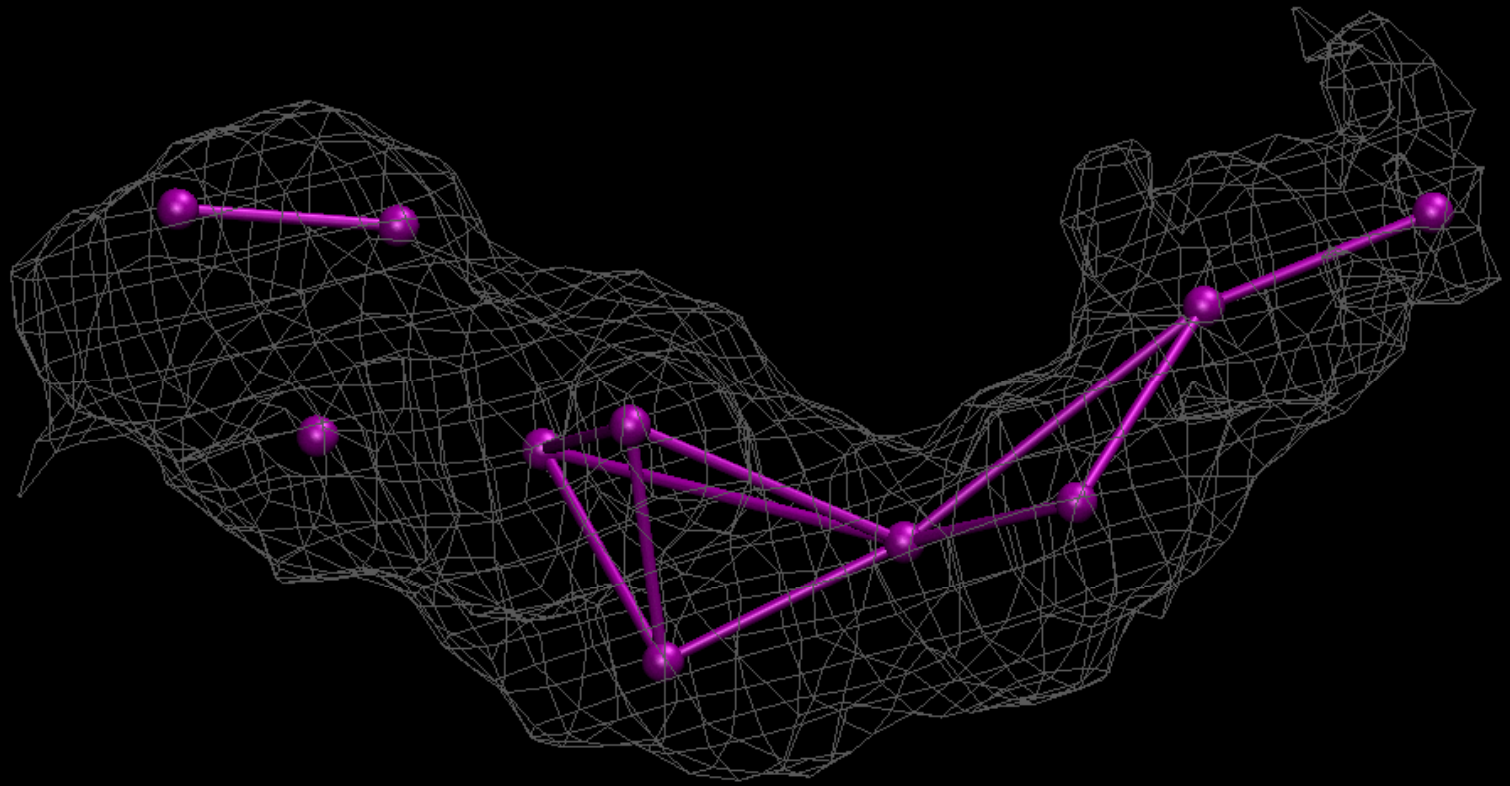


R. Schröder et al., *Nature* (2003) 425:423

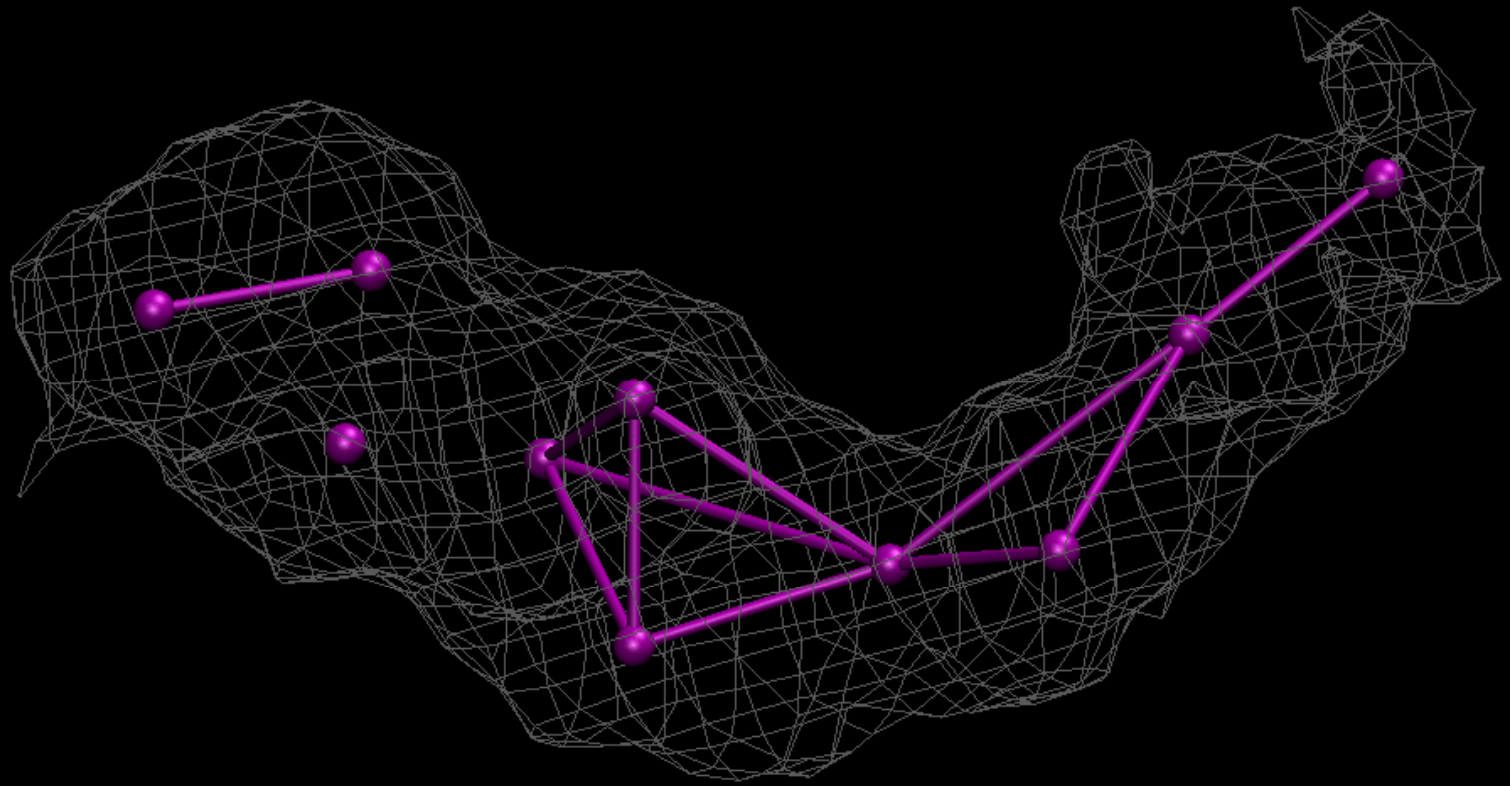
Flexing of Myosin II



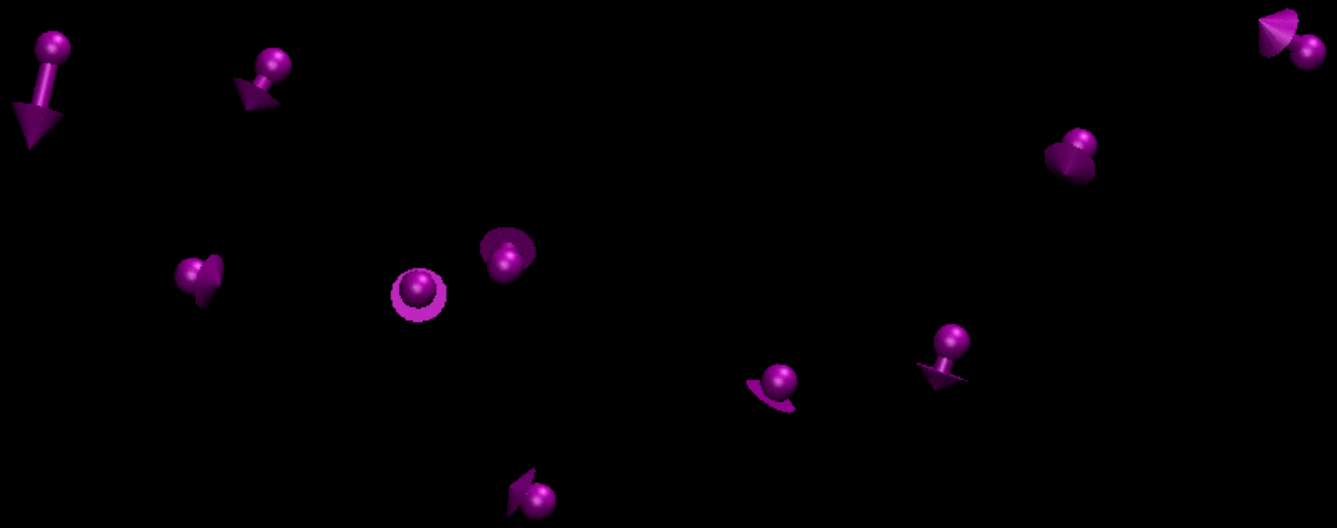
Flexing of Myosin II



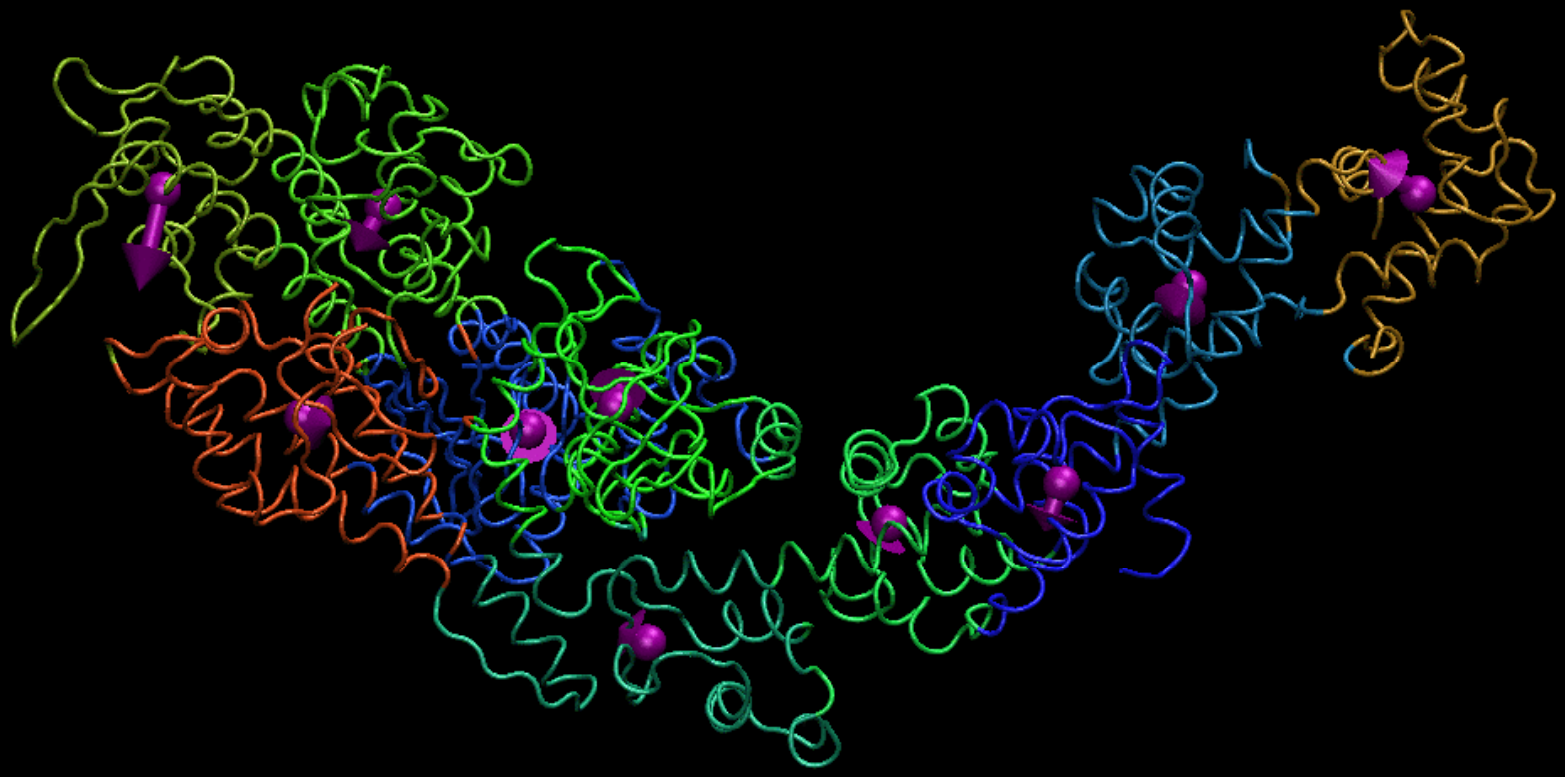
Flexing of Myosin II



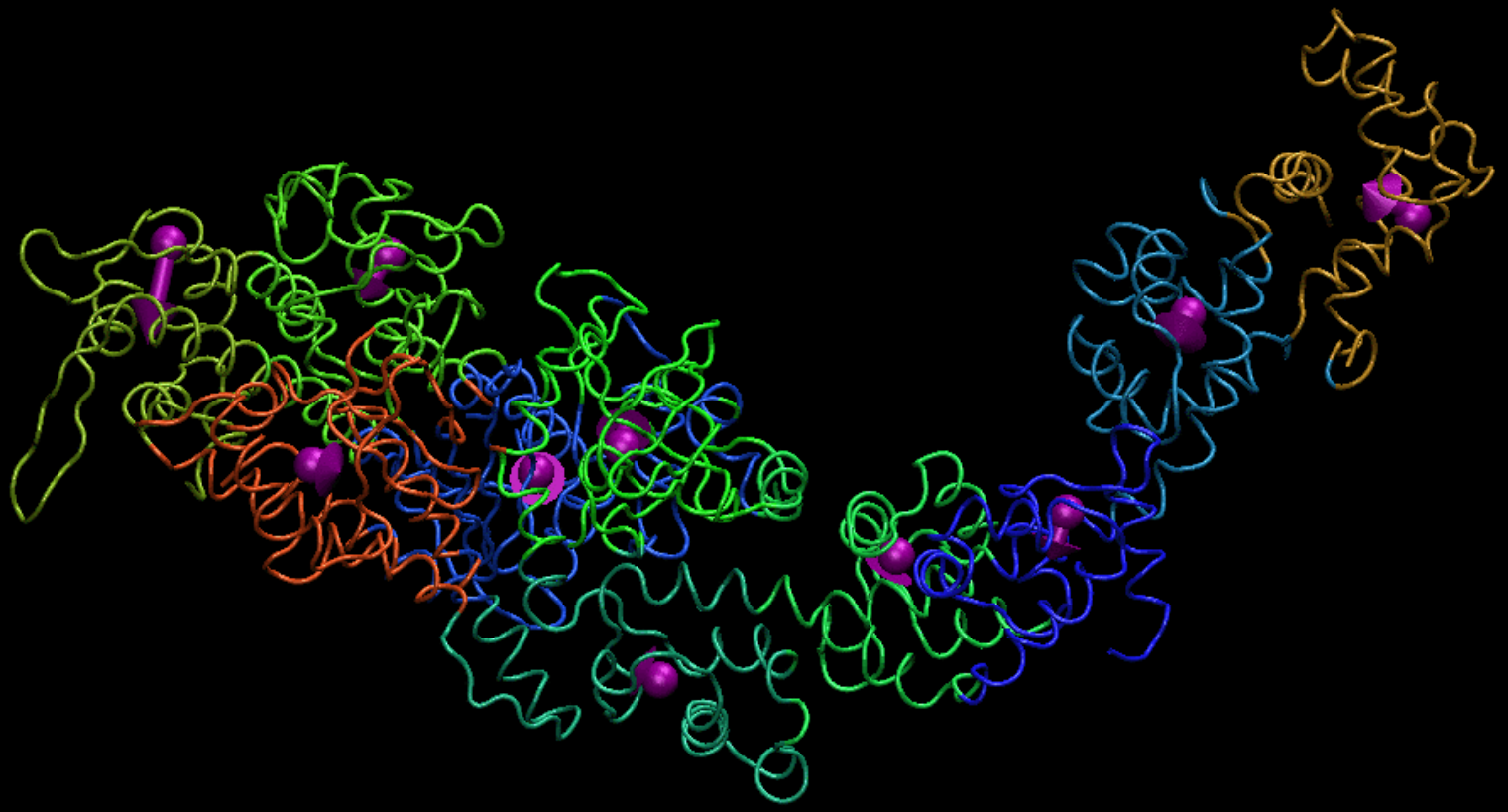
Flexing of Myosin II



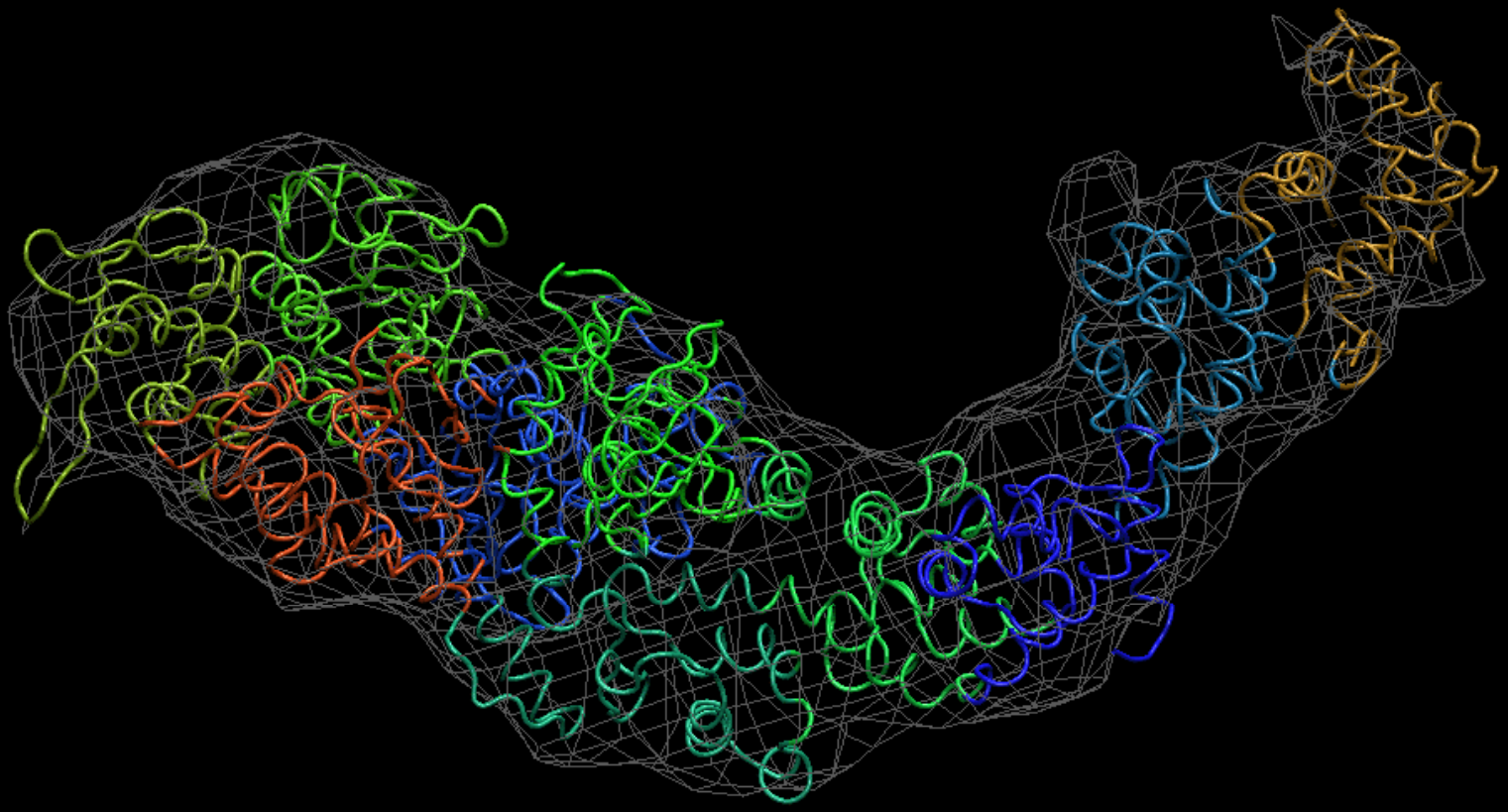
Flexing of Myosin II



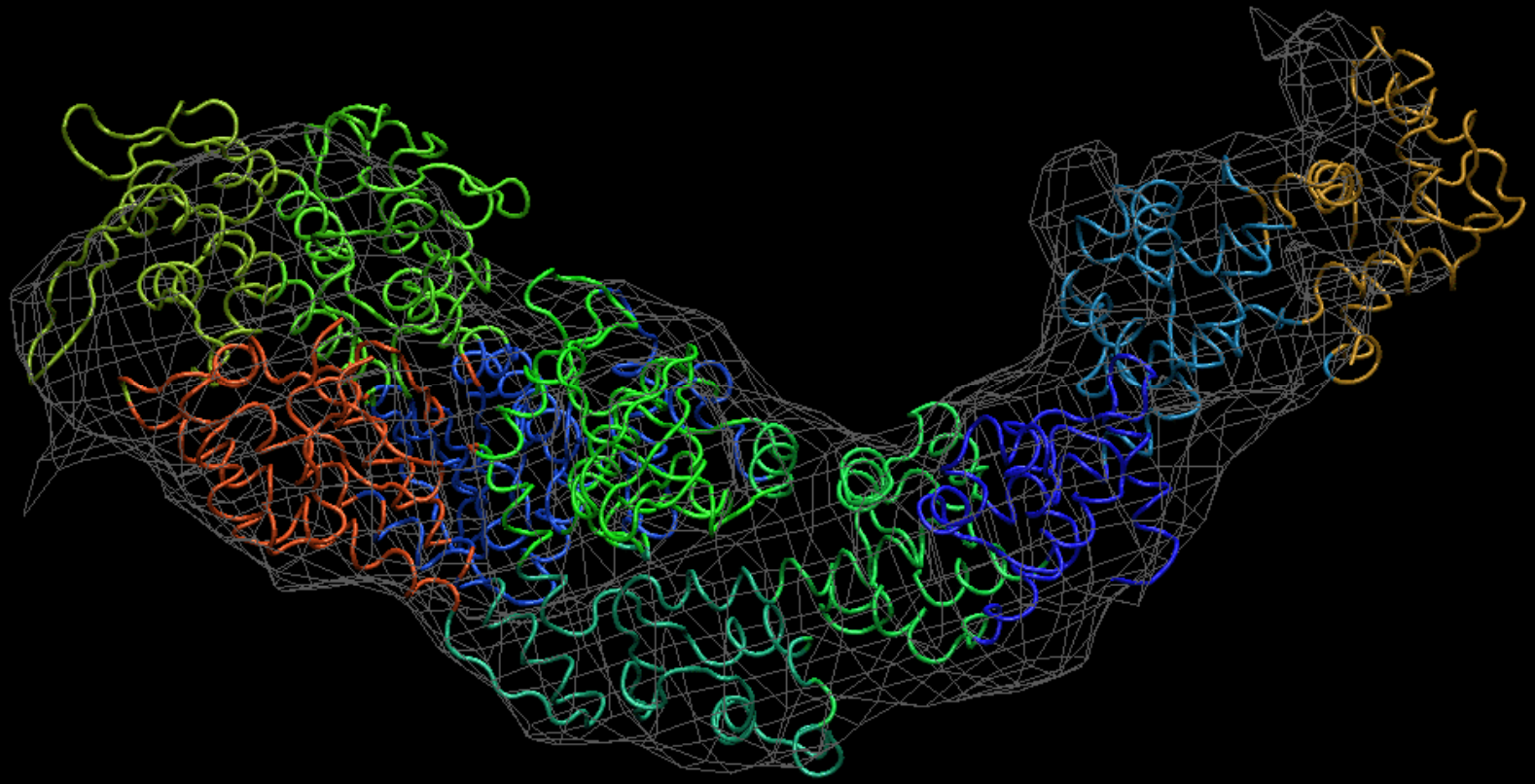
Flexing of Myosin II



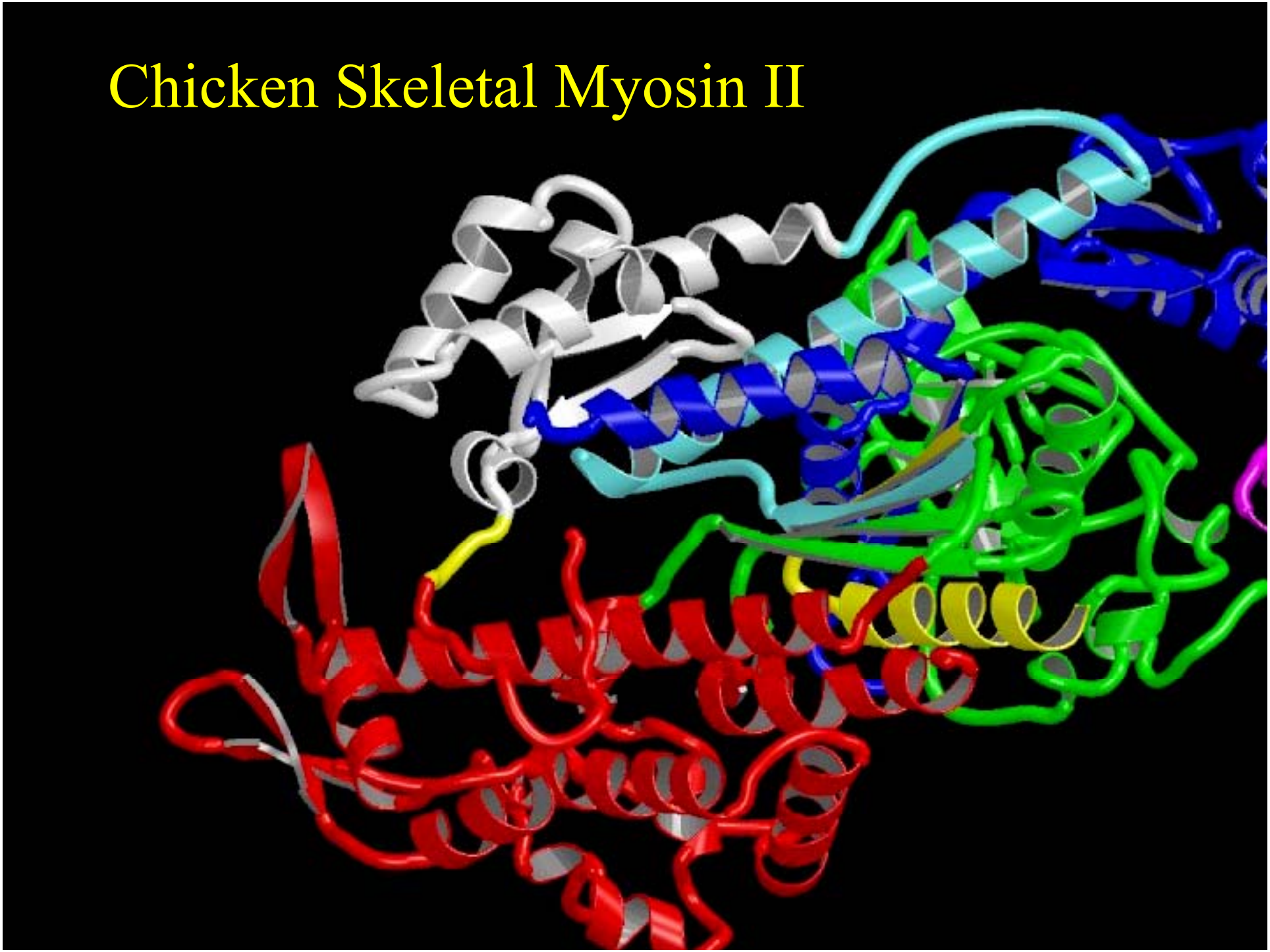
Flexing of Myosin II



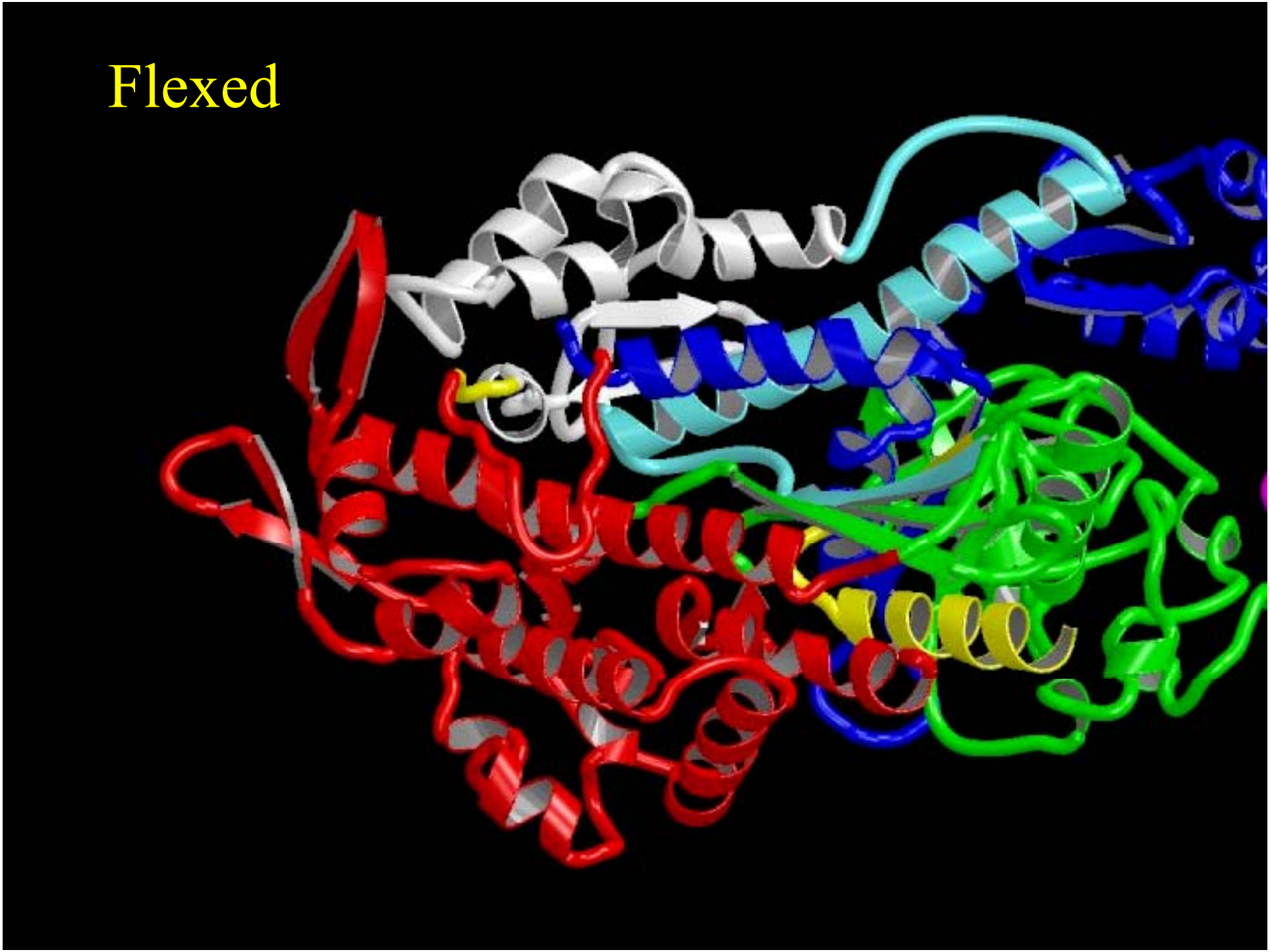
Flexing of Myosin II



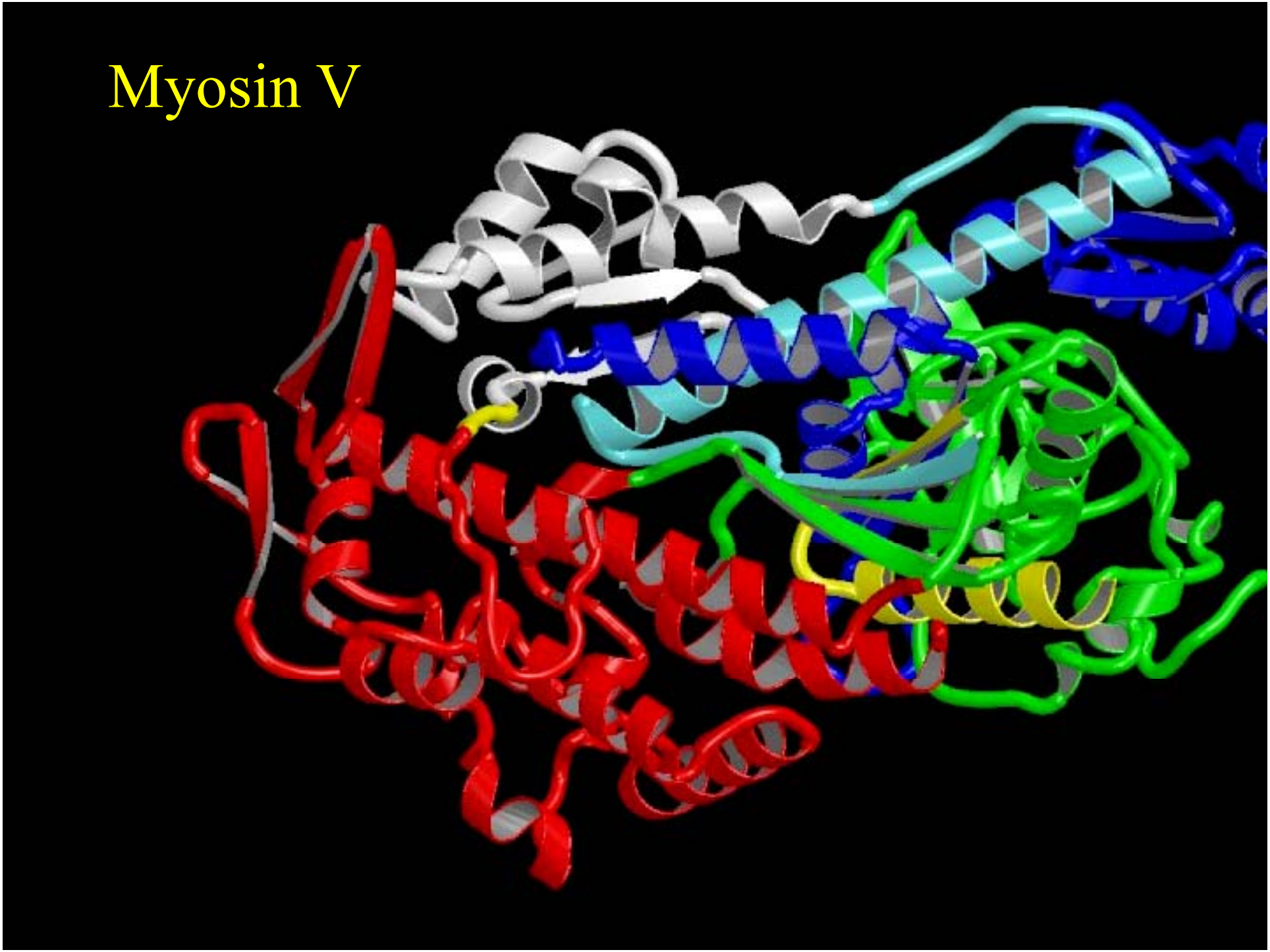
Chicken Skeletal Myosin II



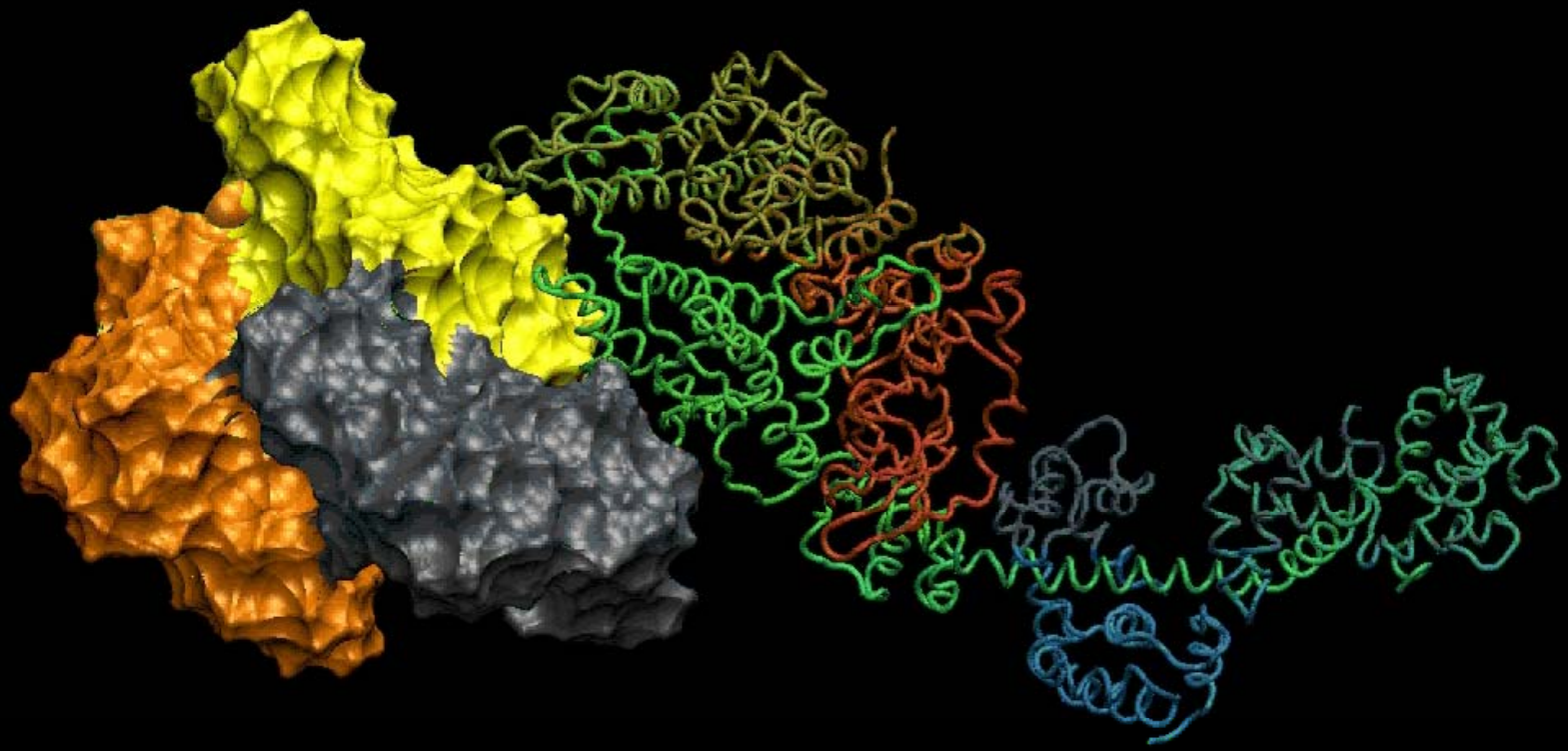
Flexed



Myosin V



Improved Actin Binding Surface



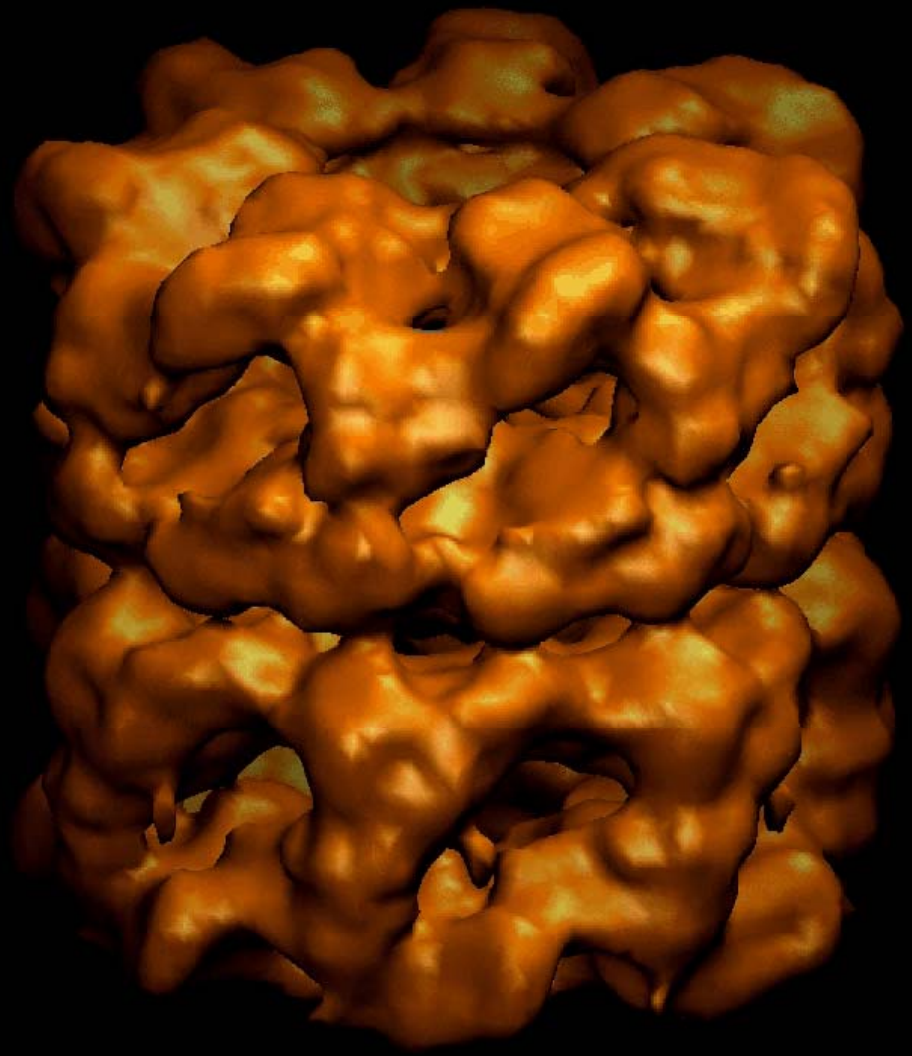
Cleft closure induced by actin binding

Myosin Flexing Validation Results:

- Agreement (~2Å rmsd) between flexed myosin II and myosin V too close to be coincidental.
- MD flexible fitting reproduces entire allosteric mechanism (cleft closure, beta sheet twist, etc).
- Mechanism only partially observed with rigid-body fitting.
- Since myosin V was not used for modeling, this validates technique.

GroEL Chaperonin

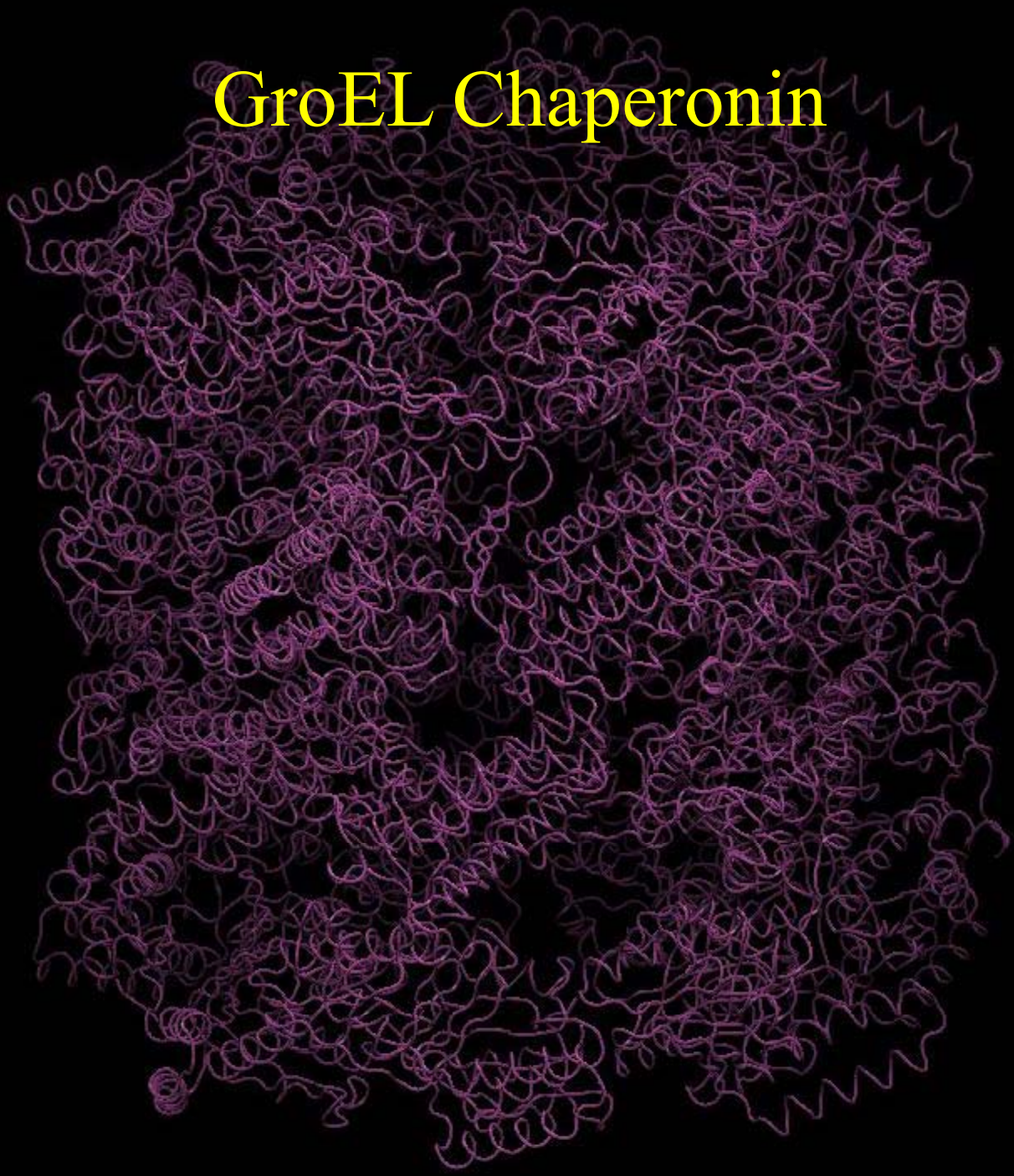
Dalia Segal,
Sharon Wolf,
Amnon Horovitz,
Weizmann
Institute, Israel



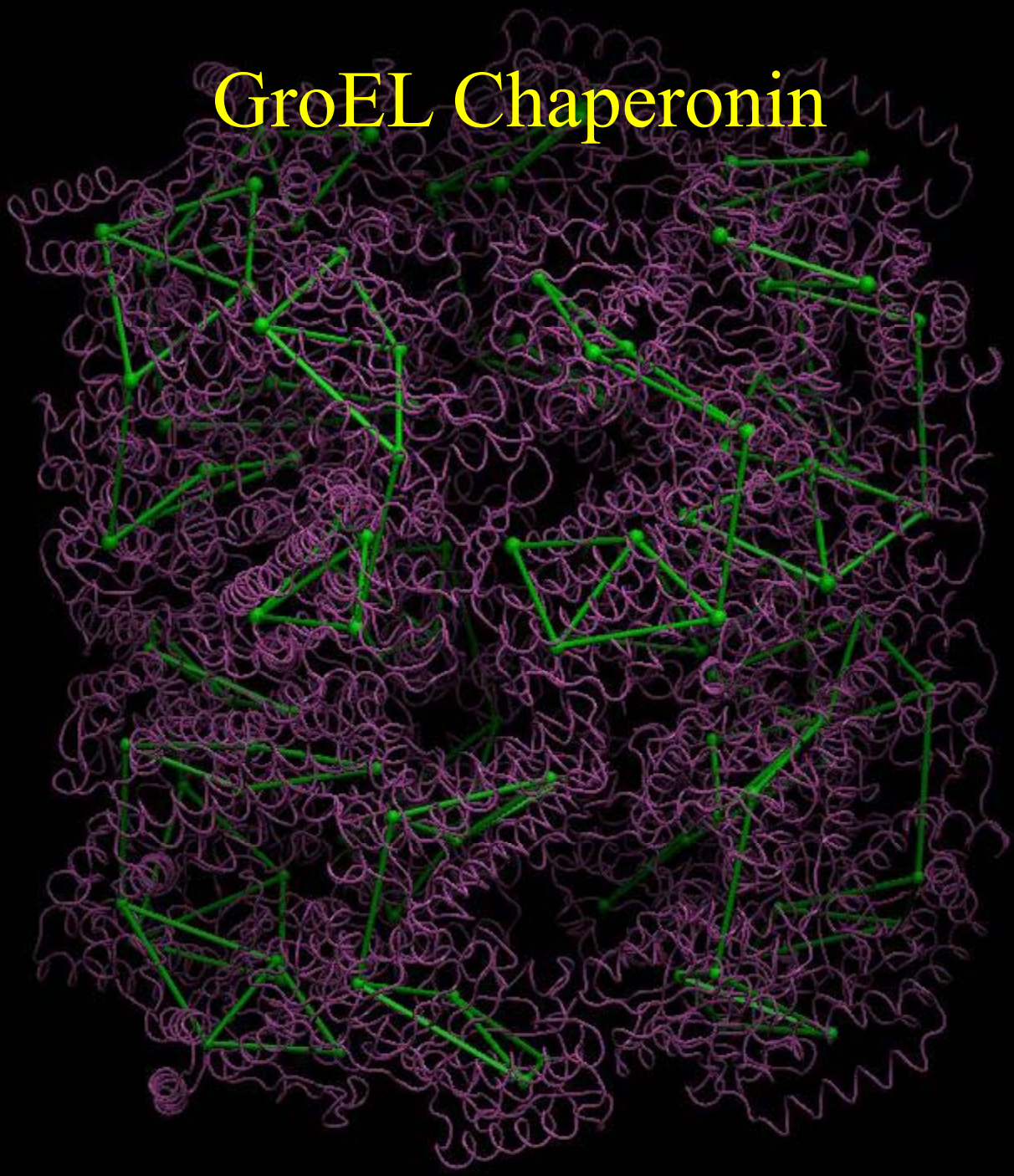
resolution ~14Å
wild type
(Sabil et al.)

& mutant

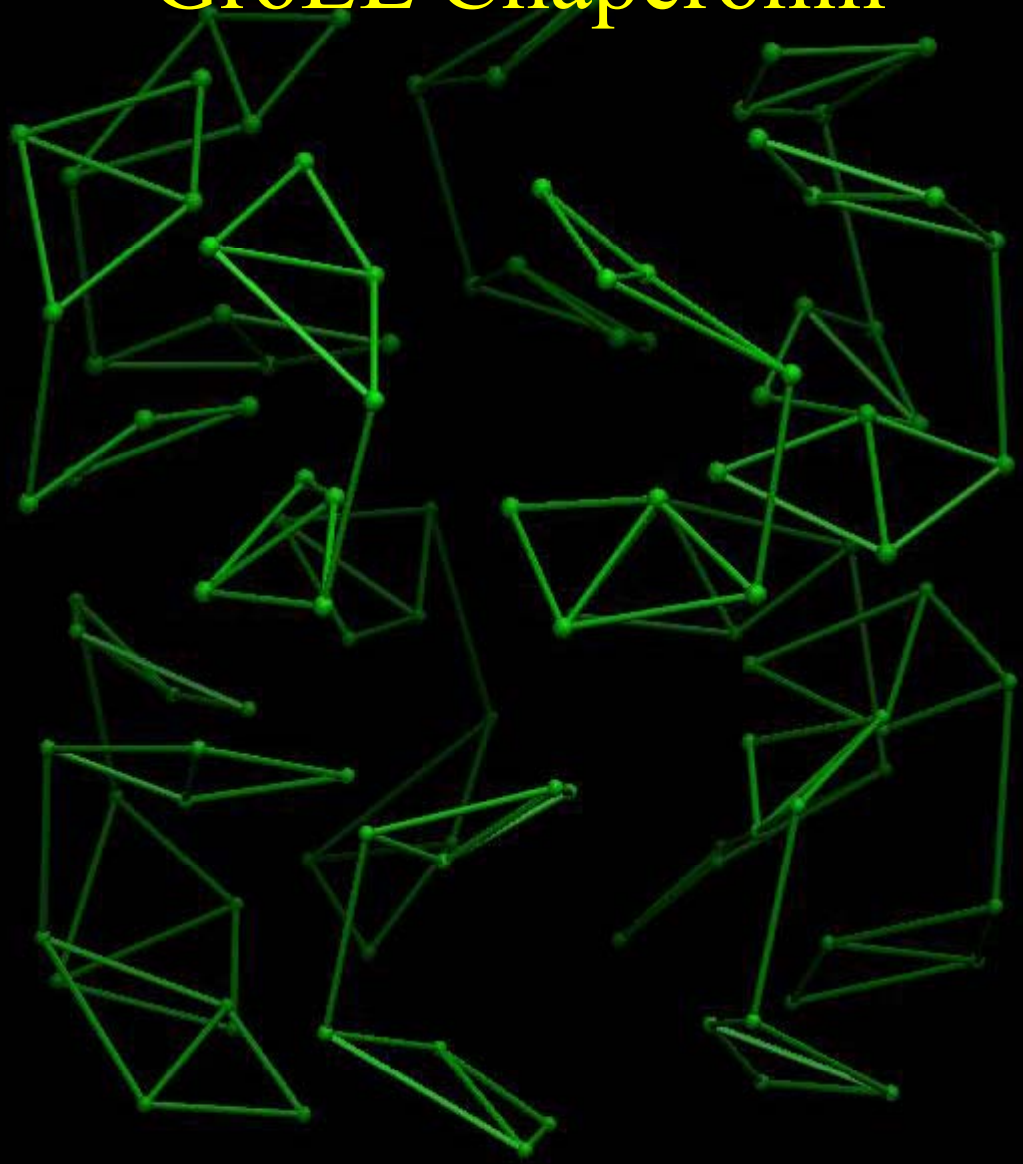
GroEL Chaperonin



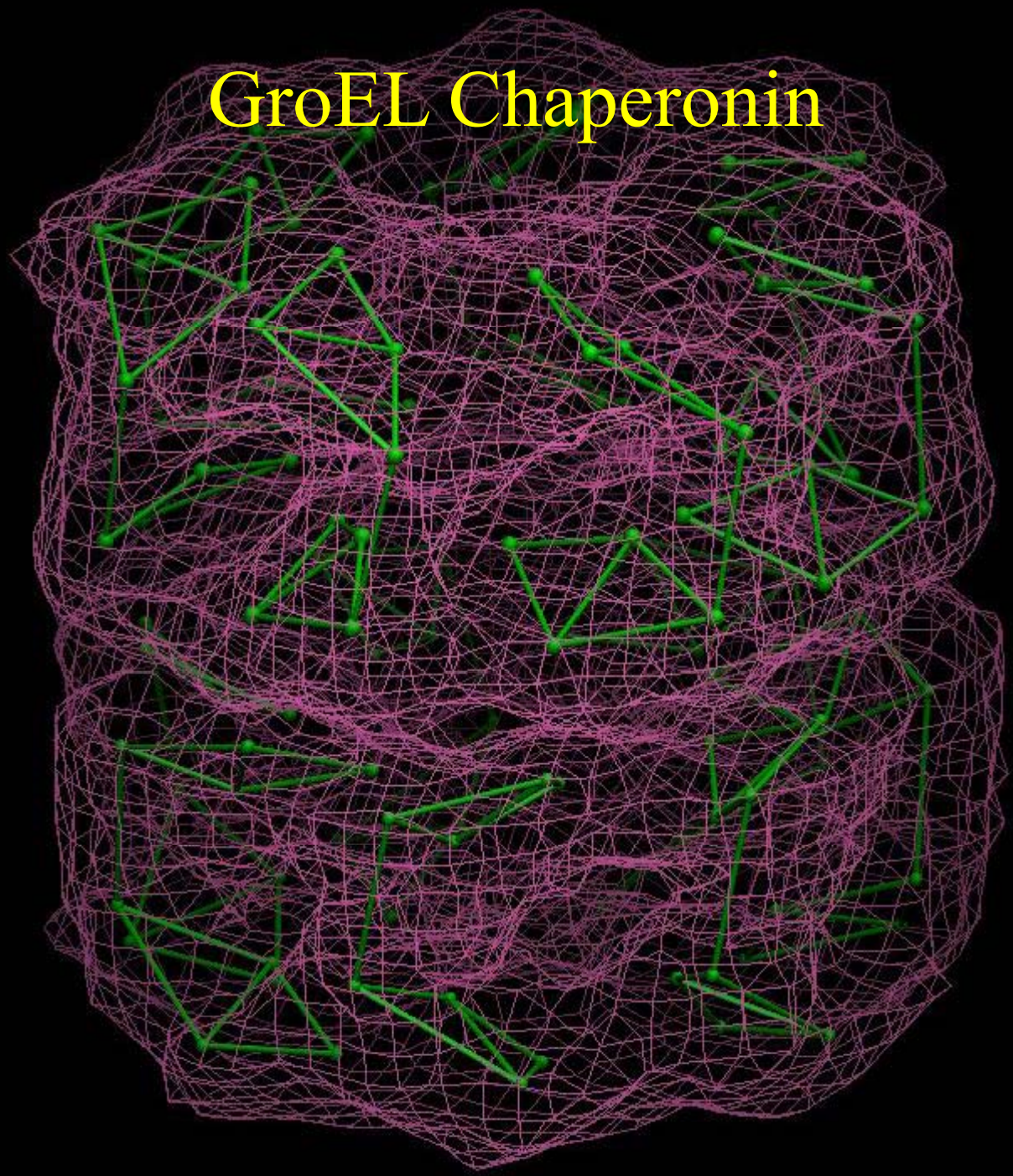
GroEL Chaperonin



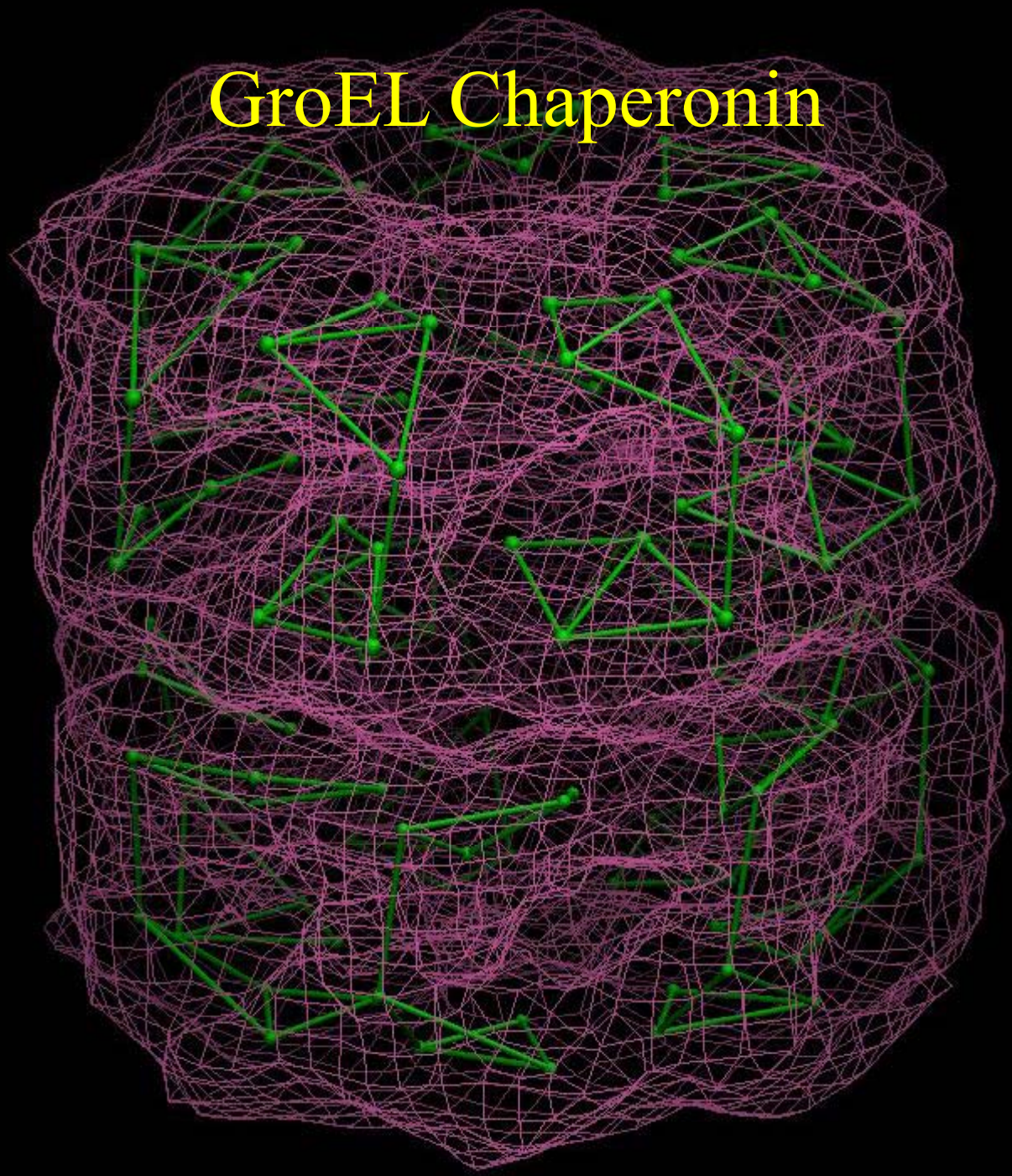
GroEL Chaperonin



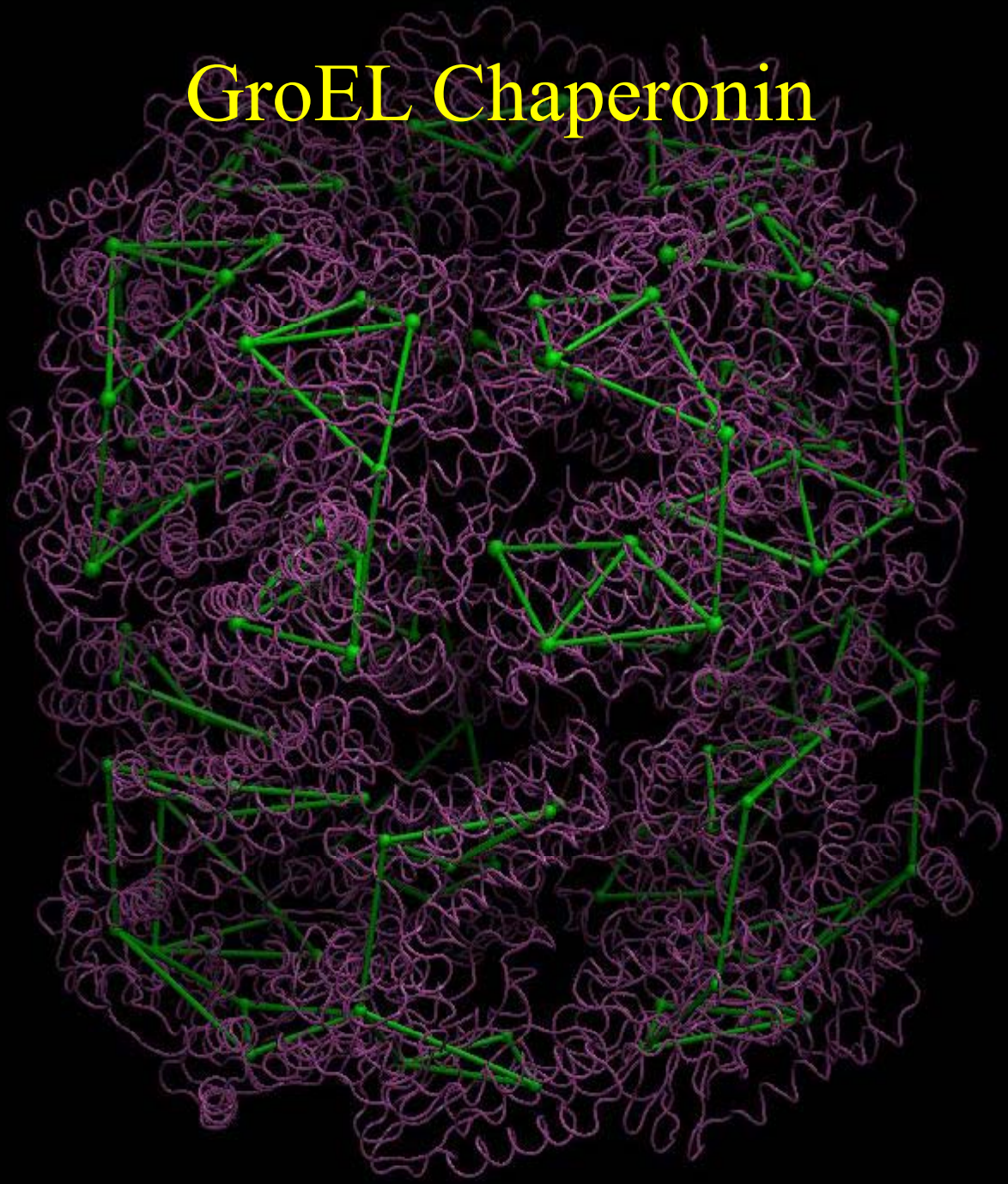
GroEL Chaperonin



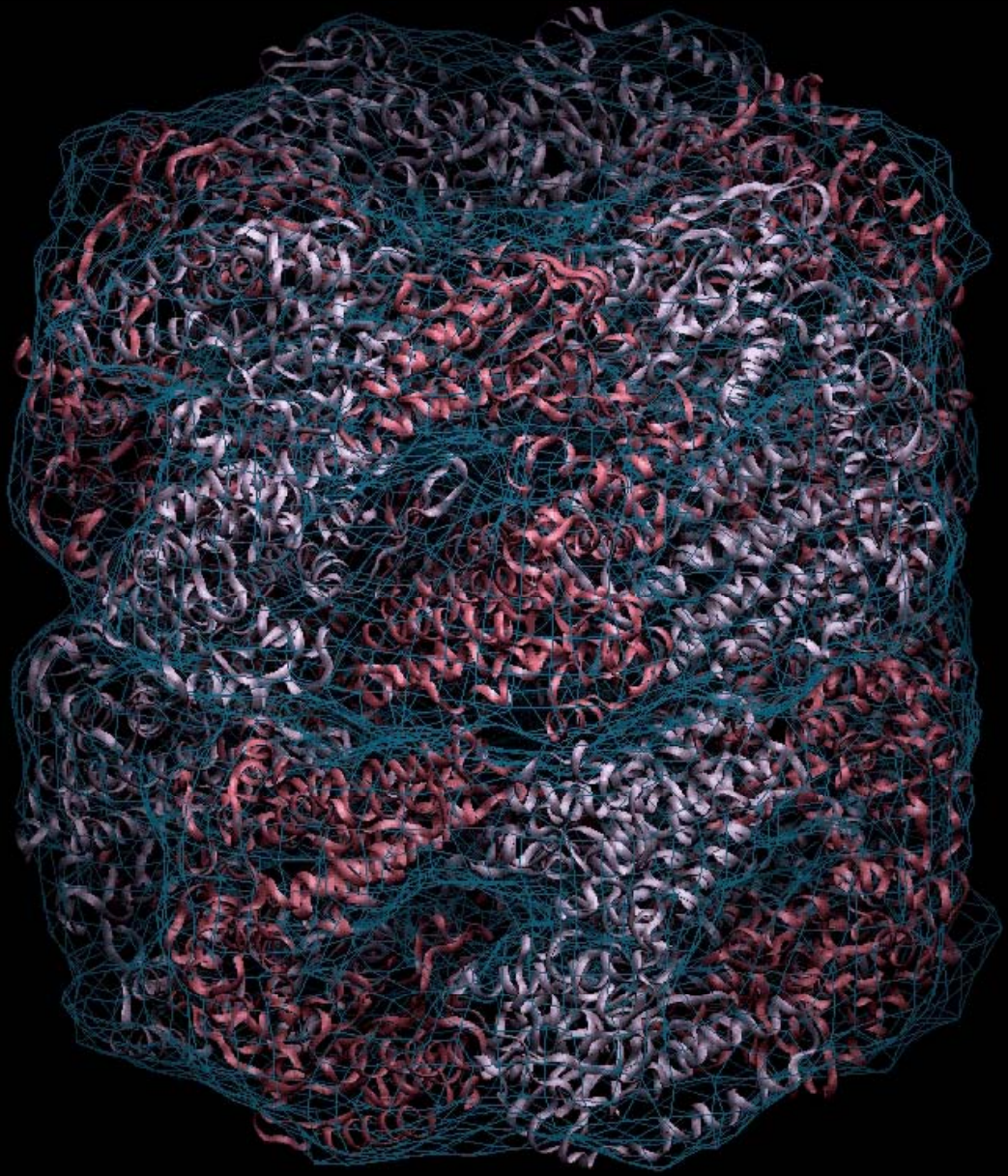
GroEL Chaperonin



GroEL Chaperonin



GroEL Chaperonin



Critical Assessment of EM Flexing

EM / Xtal Data	Resolution	Source	Precision (rmsd)
Myosin 2 Myosin 5	14Å	Schröder 2003	**2.0Å**
GroEL EM / Xtal WT	13Å	Saibil 2001	**3.0Å**
GroEL EM / Xtal WT	11A	Ludtke 2003	**2.5Å**
GroEL EM / Xtal WT	6Å	Ludtke 2004	**2.0Å**
simulated EM / Xtal WT	6-14Å	simulated	**<1.0Å**

Conclusion (Reduced Models)

Reduced (vector quantization) representations are useful for :

- Rigid-body docking.
- Flexible fitting with molecular dynamics.
- Estimation of displacement vector fields.
- Normal Mode Analysis (see earlier session).

Non-linear interpolation is a fast but less reliable alternative to MD in flexible fitting.

Interpolation allows displacements of markers to be interpolated to full space

Resources and Further Reading

WWW:
http://situs.biomachina.org
http://http://situs.biomachina.org/tutorial_colores.html
http://http://situs.biomachina.org/tutorial_flex.html

Papers:
http://situs.biomachina.org/fref.html

Acknowledgement:
Yao Cong, Stefan Birmanns, Julio Kovacs, Pablo Chacon (http://biomachina.org)