V18: Protein complexes – Density fitting

(1) We normally assume that various enzymes of a biochemical pathway „swim“ in the cytosol and randomly meet the substrate molecules one after another.

Yet, sometimes multiple enzymes of a pathway associate into large complexes and „hand over“ the substrates from one active site to the next one.

Advantage: avoids free diffusion, increases local substrate density.

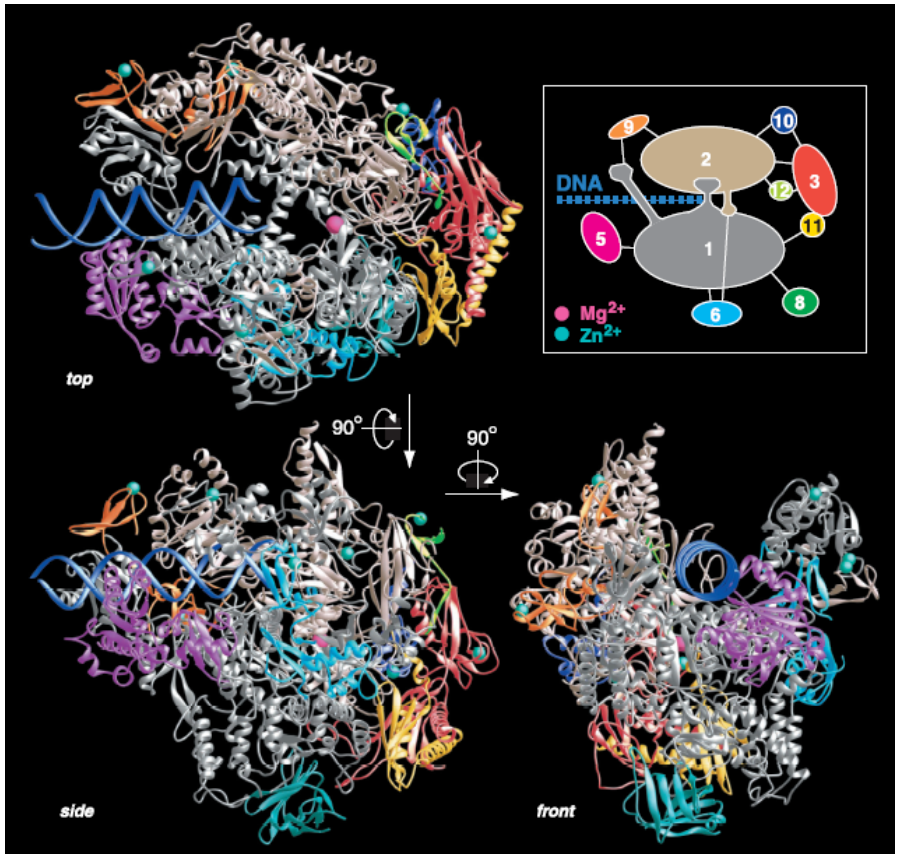
(2) Membrane transporters and receptors often form oligomers in the membrane.

Advantage: (i) large structures are built from small building blocks (simplicity)

(ii) Oligomer formation can be regulated separately from transcription.

(3) Also: complicated structural components of the cell (e.g. cytoskeleton) are built from many small components (e.g. actin)

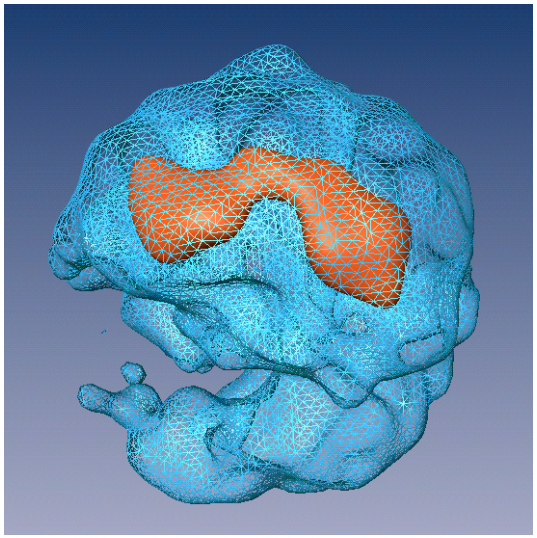
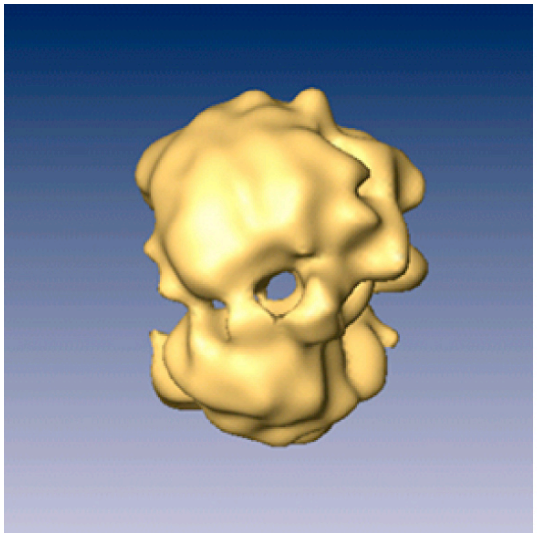
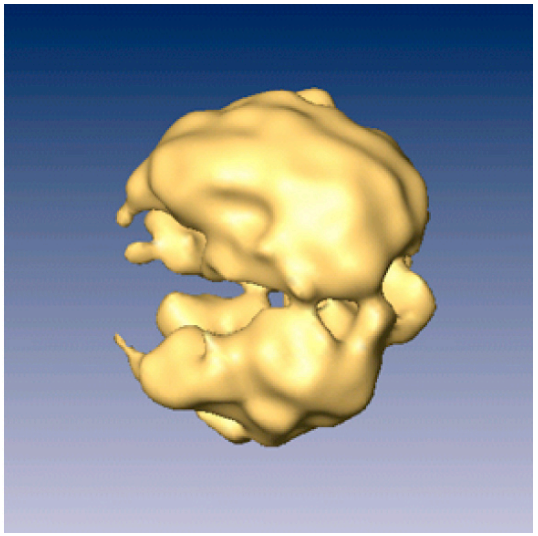
RNA Polymerase II



RNA polymerase II is the central enzyme of gene expression and synthesizes all messenger RNA in eukaryotes.

Cramer *et al.,* Science 288, 640 (2000)

RNA processing: splicesome

Structure of a cellular editor that "cuts and pastes" the first draft of RNA straight after it is formed from its DNA template.

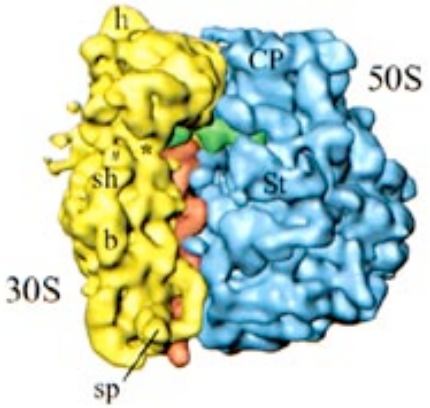
It has two distinct, unequal halves surrounding a tunnel.

Larger part: appears to contain proteins and the short segments of RNA, smaller half: is made up of proteins alone.

On one side, the tunnel opens up into a cavity, which is believed to function as a holding space for the fragile RNA waiting to be processed in the tunnel.

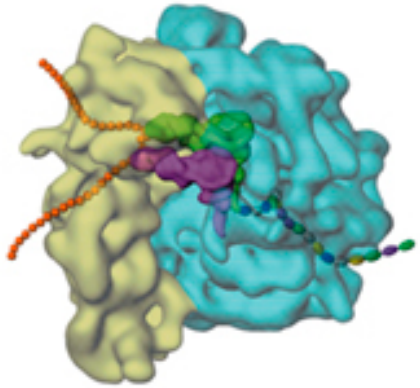
Profs. Ruth and Joseph Sperling
http://www.weizmann.ac.il/

Protein synthesis: ribosome



The ribosome is a complex subcellular particle composed of protein and RNA. It is the site of protein synthesis,

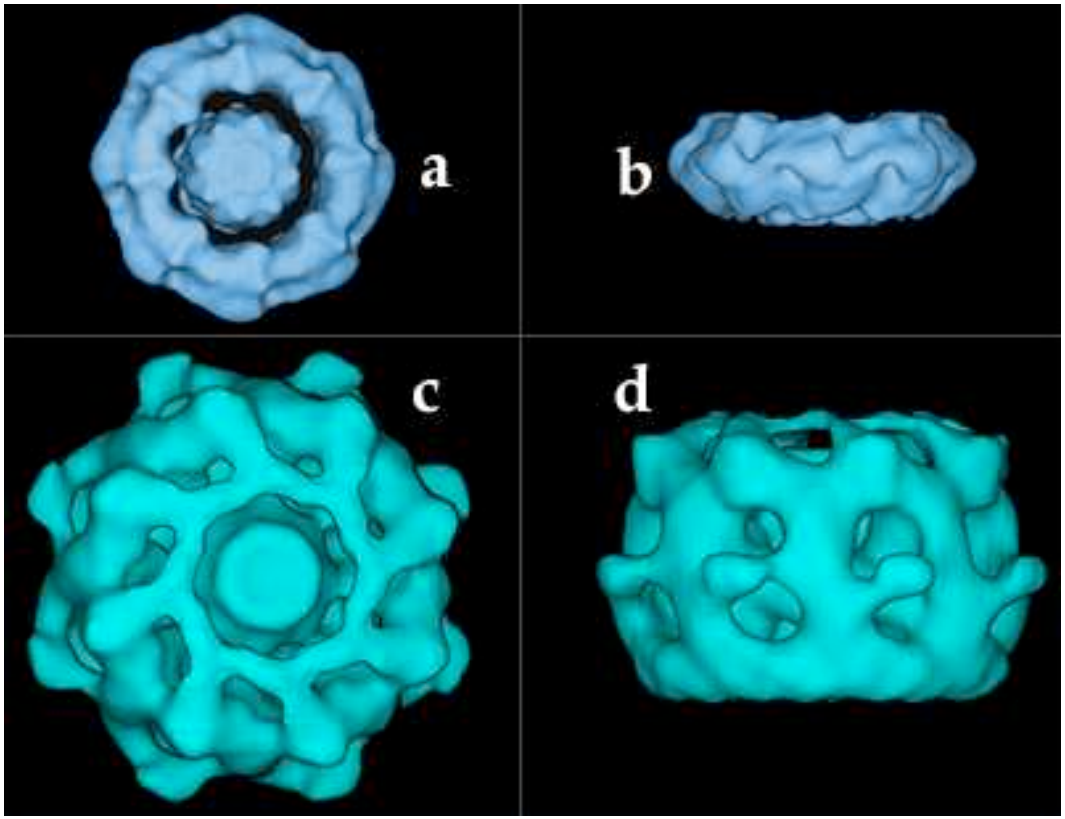
http://www.millerandlevine.com/chapter/12/cryo-em.html



Model of a ribosome with a newly manufactured protein (multicolored beads) exiting on the right.

Components of ribosome assemble spontaneously in vitro:
No helper proteins (assembly chaperones) needed

Nuclear Pore Complex



A three-dimensional image of the nuclear pore complex (NPC), revealed by electron microscopy.
A-B The NPC in yeast.
Figure A shows the NPC seen from the cytoplasm while figure B displays a side view.

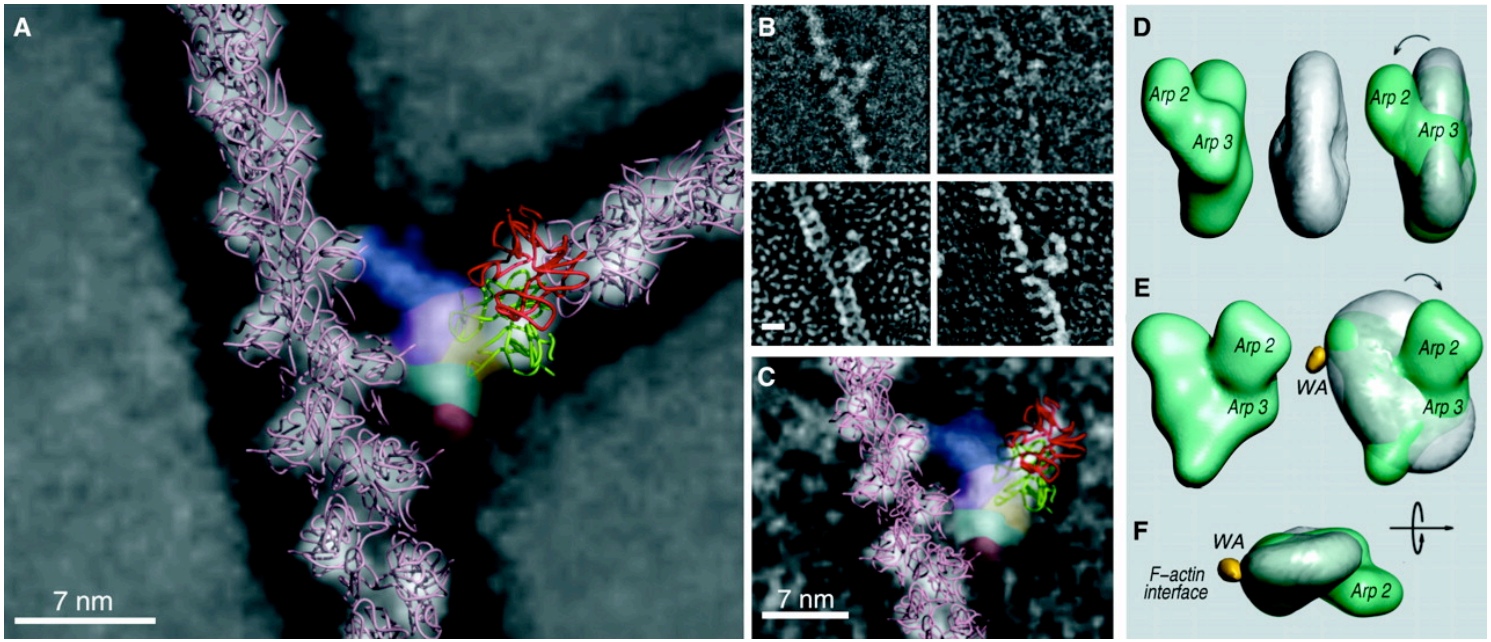
C-D The NPC in vertebrate (Xenopus).

http://www.nobel.se/medicine/educational/dna/a/transport/ncp_em1.html
Three-Dimensional Architecture of the Isolated Yeast Nuclear Pore Complex: Functional and Evolutionary Implications, Qing Yang, Michael P. Rout and Christopher W. Akey. Molecular Cell, 1:223-234, 1998

NPC is a 50-100 MDa protein assembly that regulates and controls trafficking of macromolecules through the nuclear envelope.

Molecular structure: lecture V22

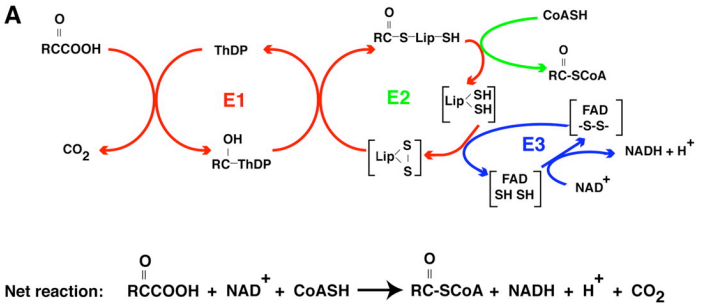
Arp2/3 complex

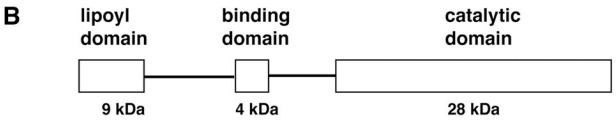


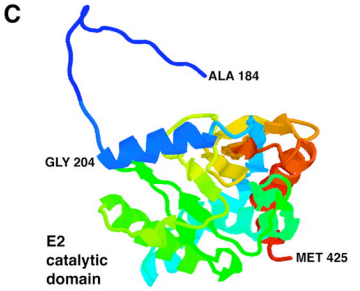
The seven-subunit Arp2/3 complex choreographs the formation of branched actin networks at the leading edge of migrating cells.
(A) Model of actin filament branches mediated by *Acanthamoeba* Arp2/3 complex. (D) Density representations of the models of actin-bound (green) and the free, WA-activated (as shown in Fig. 1D, gray) Arp2/3 complex.

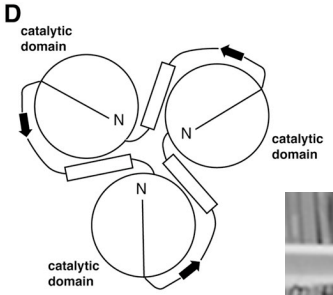
Volkmann *et al.,* Science 293, 2456 (2001)

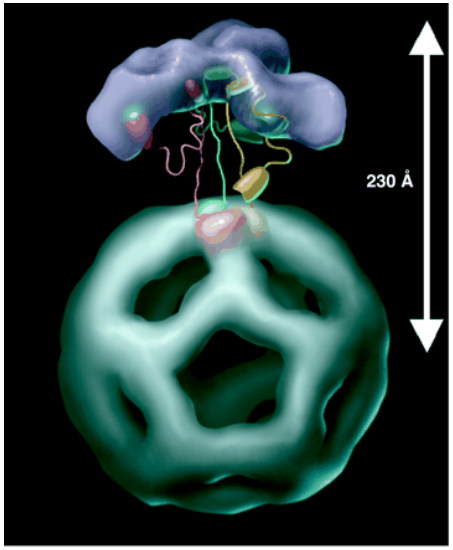
icosahedral pyruvate dehydrogenase complex: a multifunctional catalytic machine

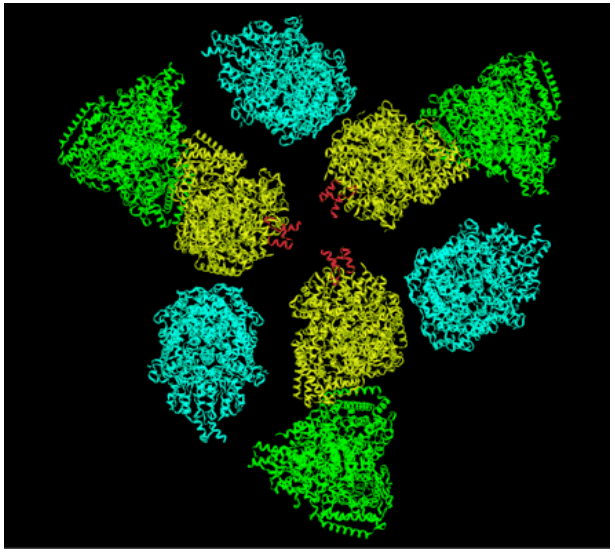








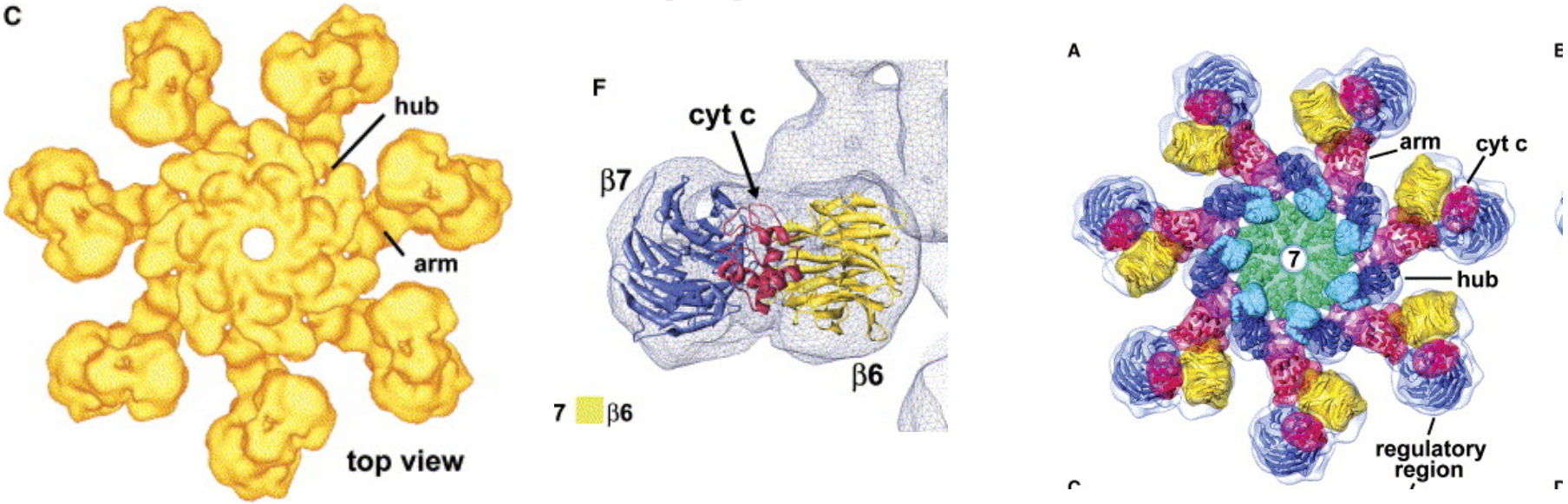




Model for active-site coupling in the E1E2 complex. 3 E1 tetramers (purple) are shown located above the corresponding trimer of E2 catalytic domains in the icosahedral core. Three full-length E2 molecules are shown, colored red, green and yellow. The lipoyl domain of each E2 molecule shuttles between the active sites of E1 and those of E2. The lipoyl domain of the red E2 is shown attached to an E1 active site. The yellow and green lipoyl domains of the other E2 molecules are shown in intermediate positions in the annular region between the core and the outer E1 layer. Selected E1 and E2 active sites are shown as white ovals, although the lipoyl domain can reach additional sites in the complex.

Milne *et al.,* EMBO J. 21, 5587 (2002)

Apoptosome



Apoptosis is the dominant form of programmed cell death during embryonic development and normal tissue turnover. In addition, apoptosis is upregulated in diseases such as AIDS, and neurodegenerative disorders, while it is downregulated in certain cancers. In apoptosis, death signals are transduced by biochemical pathways to activate caspases, a group of proteases that utilize cysteine at their active sites to cleave specific proteins at aspartate residues. The proteolysis of these critical proteins then initiates cellular events that include chromatin degradation into nucleosomes and organelle destruction. These steps prepare apoptotic cells for phagocytosis and result in the efficient recycling of biochemical resources.
In many cases, apoptotic signals are transmitted to mitochondria, which act as integrators of cell death because both effector and regulatory molecules converge at this organelle. Apoptosis mediated by mitochondria requires the release of cytochrome c into the cytosol through a process that may involve the formation of specific pores or rupture of the outer membrane. Cytochrome c binds to Apaf-1 and in the presence of dATP/ATP promotes assembly of the apoptosome. This large protein complex then binds and activates procaspase-9.

Determining molecular 3D structures

Experimental techniques:

Dimensions	proteins: 1 – 5 nm	atoms: 0.1 – 0.5 nm
bond stability	covalent ca. 300 kJ/mol	H-bonds: ca. 5 – 20 kJ/mol

X-ray crystallography
NMR
electron microscopy
FRET
AFM pulling

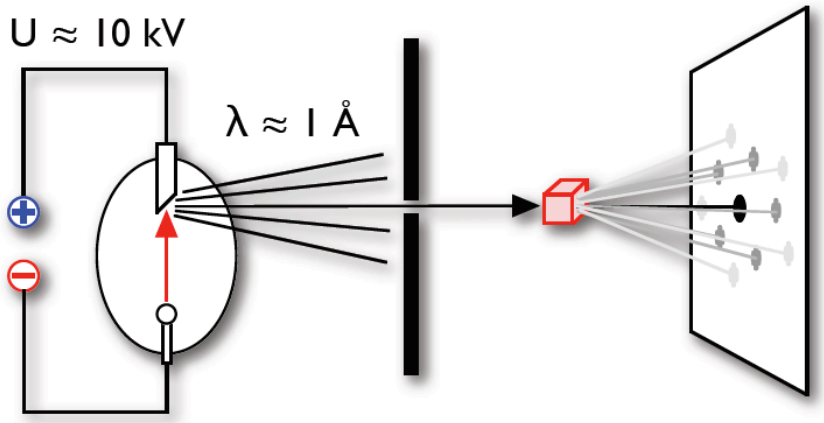
- applicability
- resulting information
- resolution
- distortions
- effort/cost ...

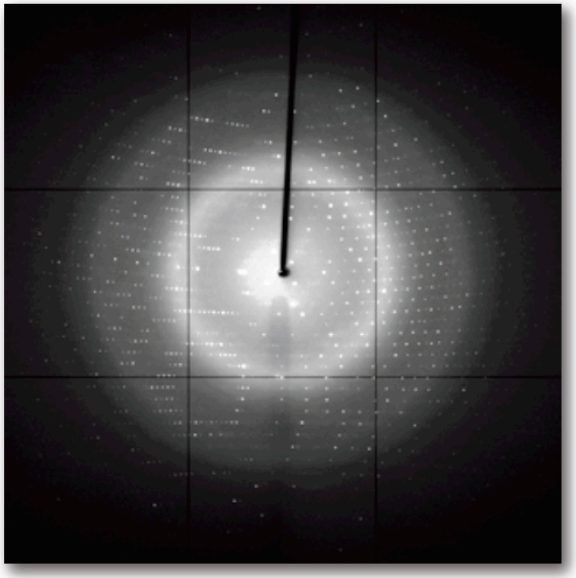
Prediction techniques:

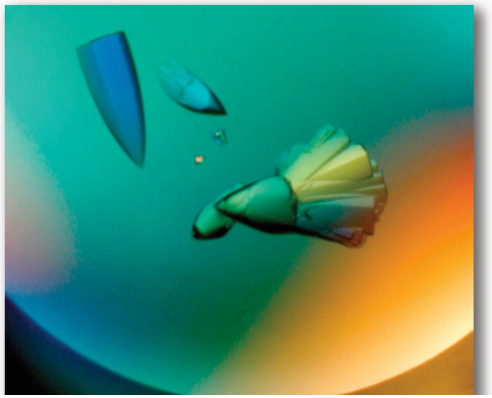
Homology modelling, correlation based fitting, ab-initio modelling

X-ray crystallography

Beam of photons (no mass),
need high energy, method
needs relatively large samples







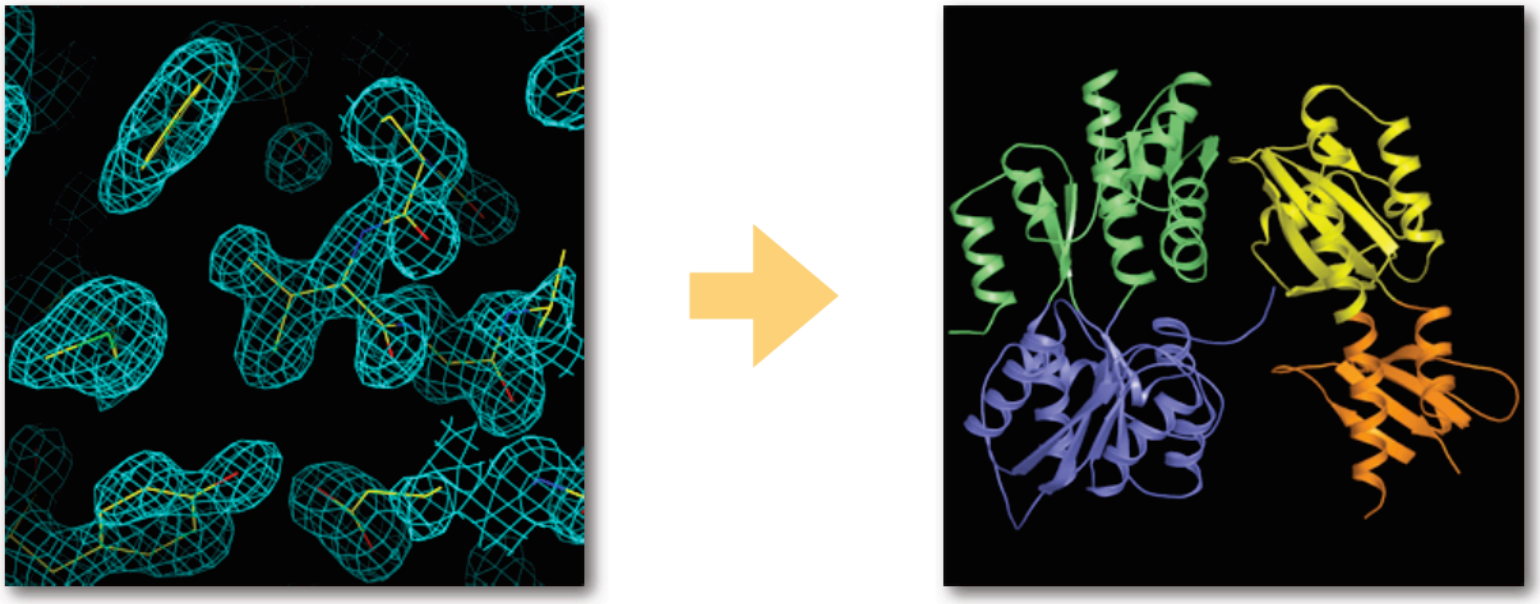
Crystallize proteins, record diffraction patterns
of X-rays (scatter at the electron clouds)
=> reconstruction

1 keV ≈ 10^5 kJ/mol

http://www.molbio1.princeton.edu/macro/about.html

V 20 – 14

X-ray reconstruction



Reconstruct electron density maps from diffraction patterns at multiple wavelengths and orientations, refine structure computationally

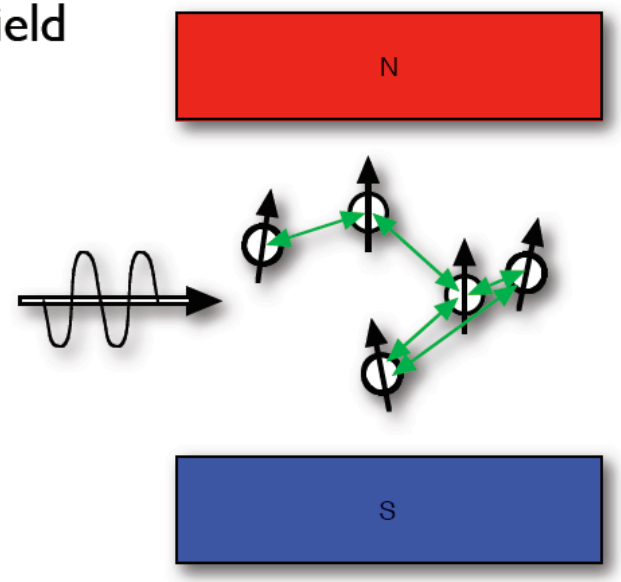
Main problem: proteins do not like to crystallize (especially membrane proteins)

PDB: experimental technique == X-Ray: 43138 results
X-Ray && "in the membrane": 1232 results

http://www.molbio1.princeton.edu/macro/about.html

Nuclear magnetic resonance

Place sample with the proteins into strong magnetic field
=> nuclei with non-zero magnetic moment
(^{1}H, ^{13}C, ^{15}N) align



Apply electromagnetic RF field
=> resonances of the nuclear spins depending on
- atom type
- chemical environment

=> extract distances between close by ^{1}H atoms (distance constraints)

\+ no crystallization required, proteins in physiological environment
\+ atomic resolution
– too much overlap for larger proteins (≥ 15 kDa)

Electron microscopy

Like X-ray crystallography, but with electrons (electrons have mass)

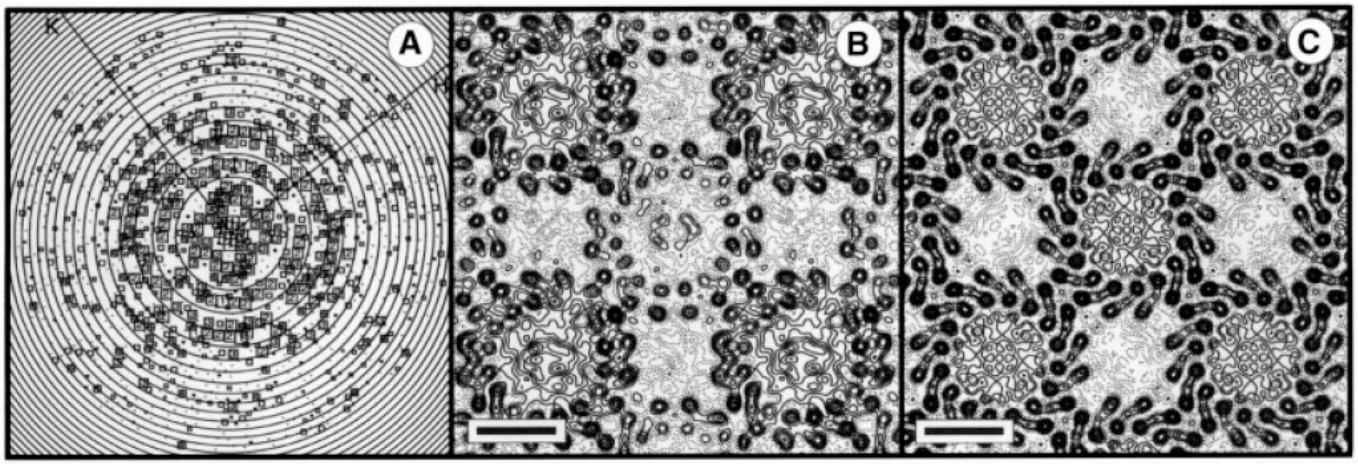
=> much stronger interaction with electron clouds

=> works already on 2D crystals (membrane proteins!)

=> strong radiation damage of the sample

=> resolution limited (keep electron energy low) (longer wavelength)

For 3D tomography: rotate sample

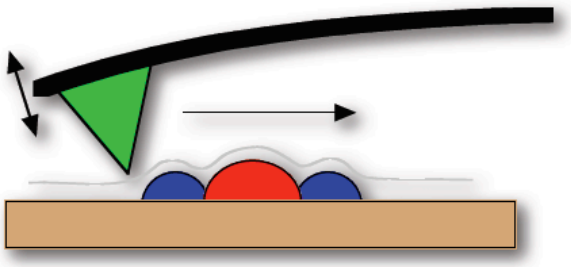


Cryo-EM images of the LH1-RC core complexes of *Rhodospirillum rubrum* at 8.5 Å resolution

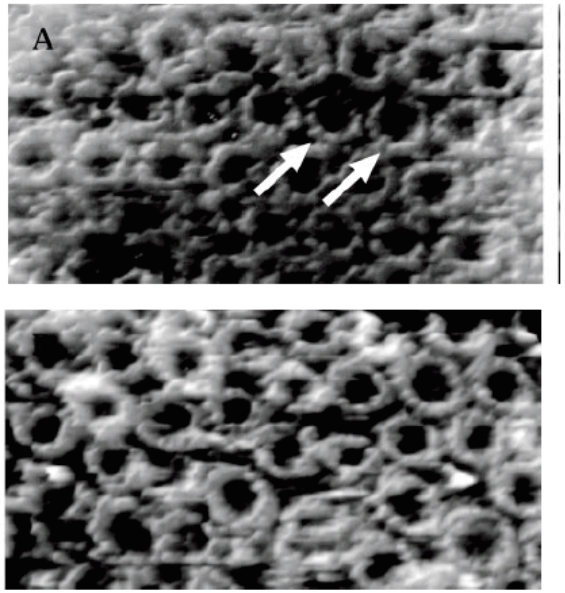
Fig. 3. (**A**) Representation of amplitudes of Fourier components calculated for one image of a glucose-embedded tetragonal crystal. Numbers and box sizes correspond to the spot IQ value, with spots of the highest signal-to-noise ratio having an IQ of 1 (Henderson *et al.*, 1986). Spots are shown to a resolution of 1/6 Å^{-1}. (**B**) 8.5 Å resolution projection map calculated from five averaged images of glucose-embedded tetragonal crystals, assuming *p*1 symmetry. Contouring is at 0.5 r.m.s. density with density above mean (protein) represented by solid contours. Scale bar: 50 Å. (**C**) As (B), with $p42_12$ symmetry applied.

Jamieson etal, *EMBO J* **21** (2003) 3927

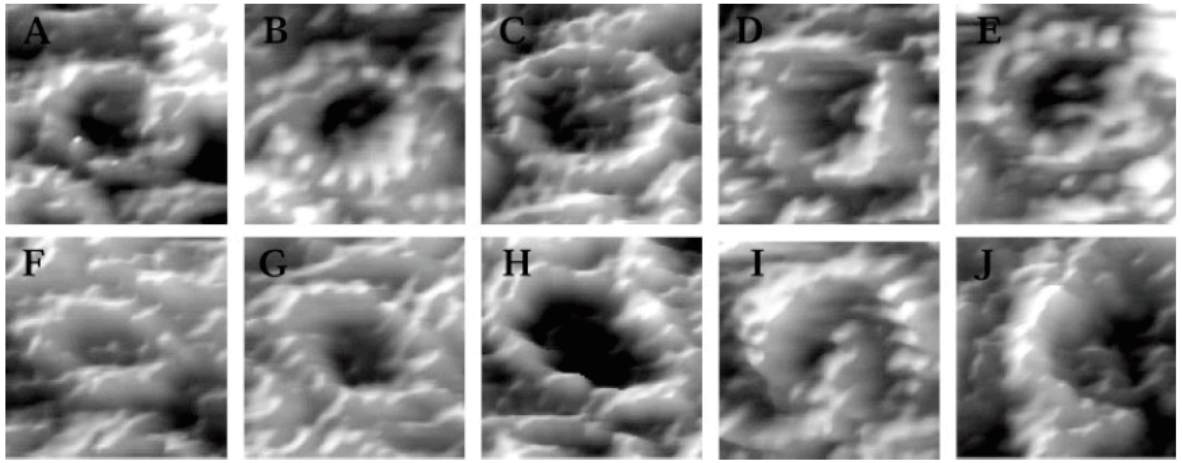
Atomic force microscopy



Scan membrane with proteins
(in physiological conditions)
=> protein arrangement (coarse view)
=> protein shape (high resolution)



Shapes and sizes of monomeric
LH1 from purple bacteria

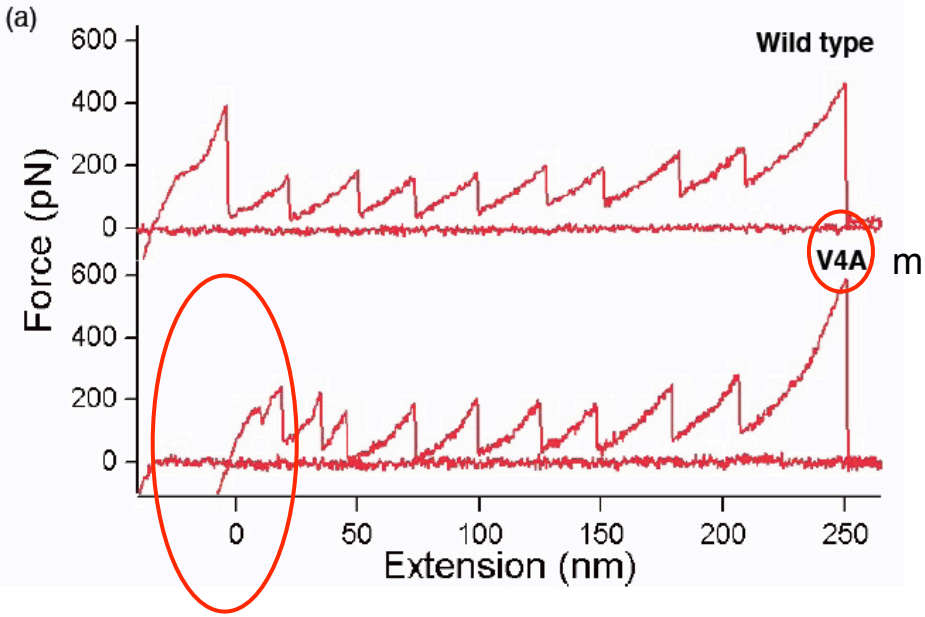


Bahatyrova etal, *J Biol Chem* **279** (2004) 21327

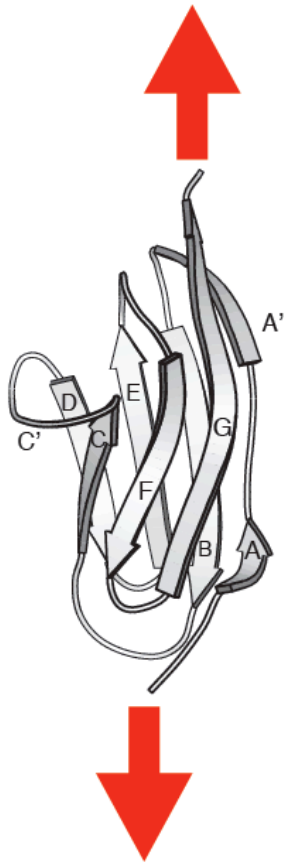
AFM pulling

Mechanical Unfolding of a Titin Ig Domain: Structure of Unfolding Intermediate Revealed by Combining AFM, Molecular Dynamics Simulations, NMR and Protein Engineering

Susan B. Fowler[1], Robert B. Best[1], José L. Toca Herrera[1]
Trevor J. Rutherford[1], Annette Steward[1], Emanuele Paci[2]
Martin Karplus[2,3] and Jane Clarke[1*]

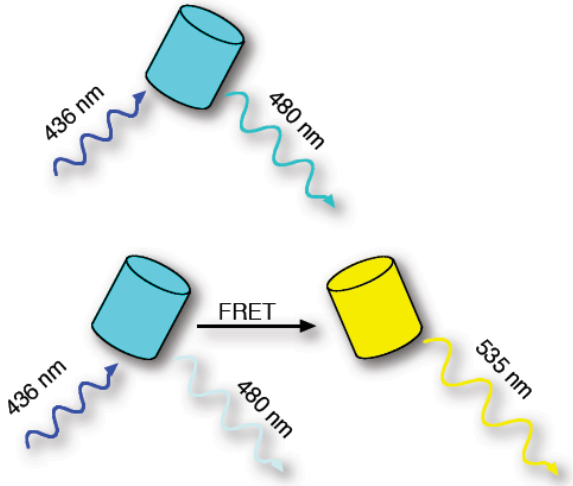


Can also be applied to protein complexes



J. Mol. Biol. **322** (2002) 841–849

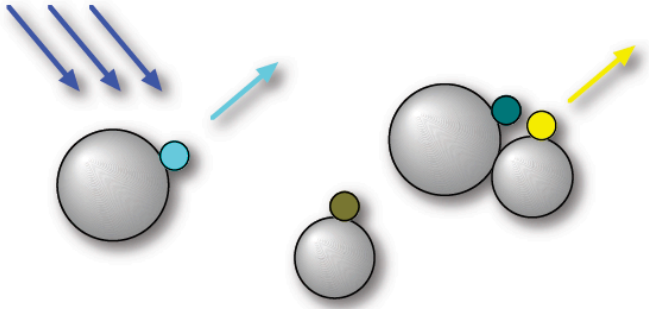
Fluorescence energy transfer

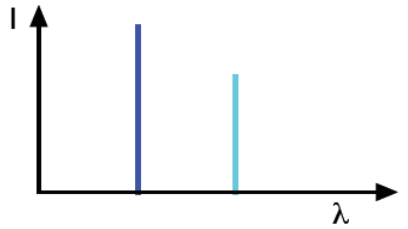


Chromophore of the cyan flourescent protein (CFP) absorbs at 436 nm and emits at 480 nm, YFP absorbs at 480 nm and emits at 535 nm.
=> **resonant** (non-radiative) energy **transfer** from CFP onto YFP when both are close enough

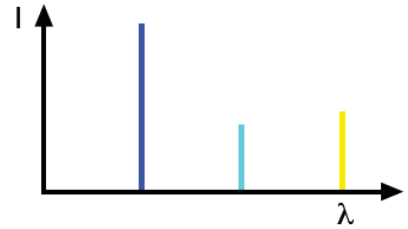
Resonant Förster transfer $\propto d^{-6}$

Tag two potential **complex partners** with CFP and YFP and measure **flourescence spectrum**:





Observed when CFP and YFP are far away



Observed when CFP and YFP are close

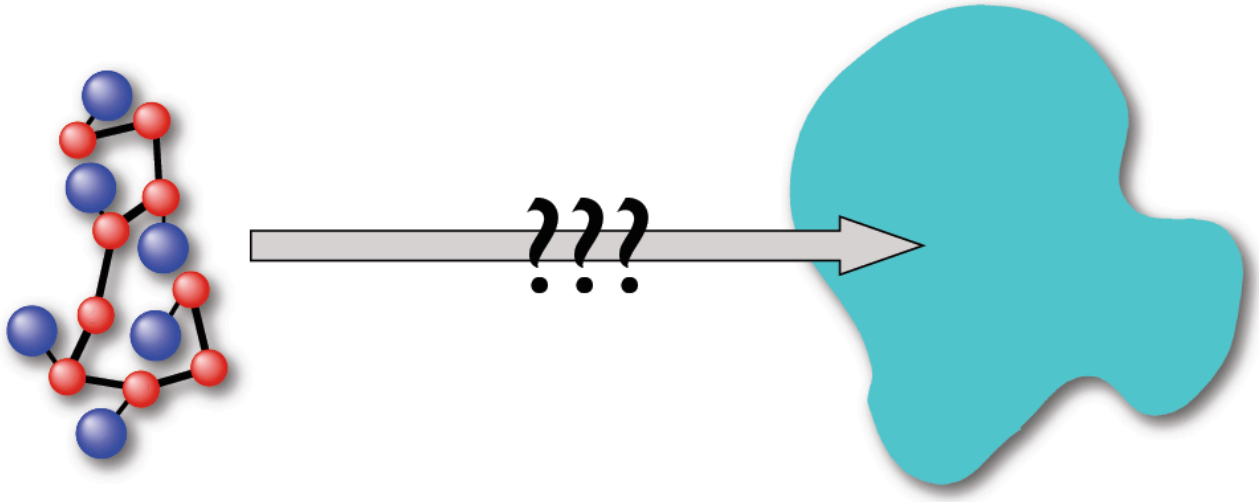
Structural techniques - overview

	X-ray crystallography	NMR	EM/ tomography	AFM	FRET	Y2H	TAP	MS
Structure ≤ 3Å	X	X						
structure ≥ 3Å	X	X	X	X				
contacts	X	X	X		X	X	X	X
proximity	X	X	X		X	X		
stoichiometry	X	X					X	X
complex symmetry	X	X	X	X				

Fitting atomistic structures into EM maps

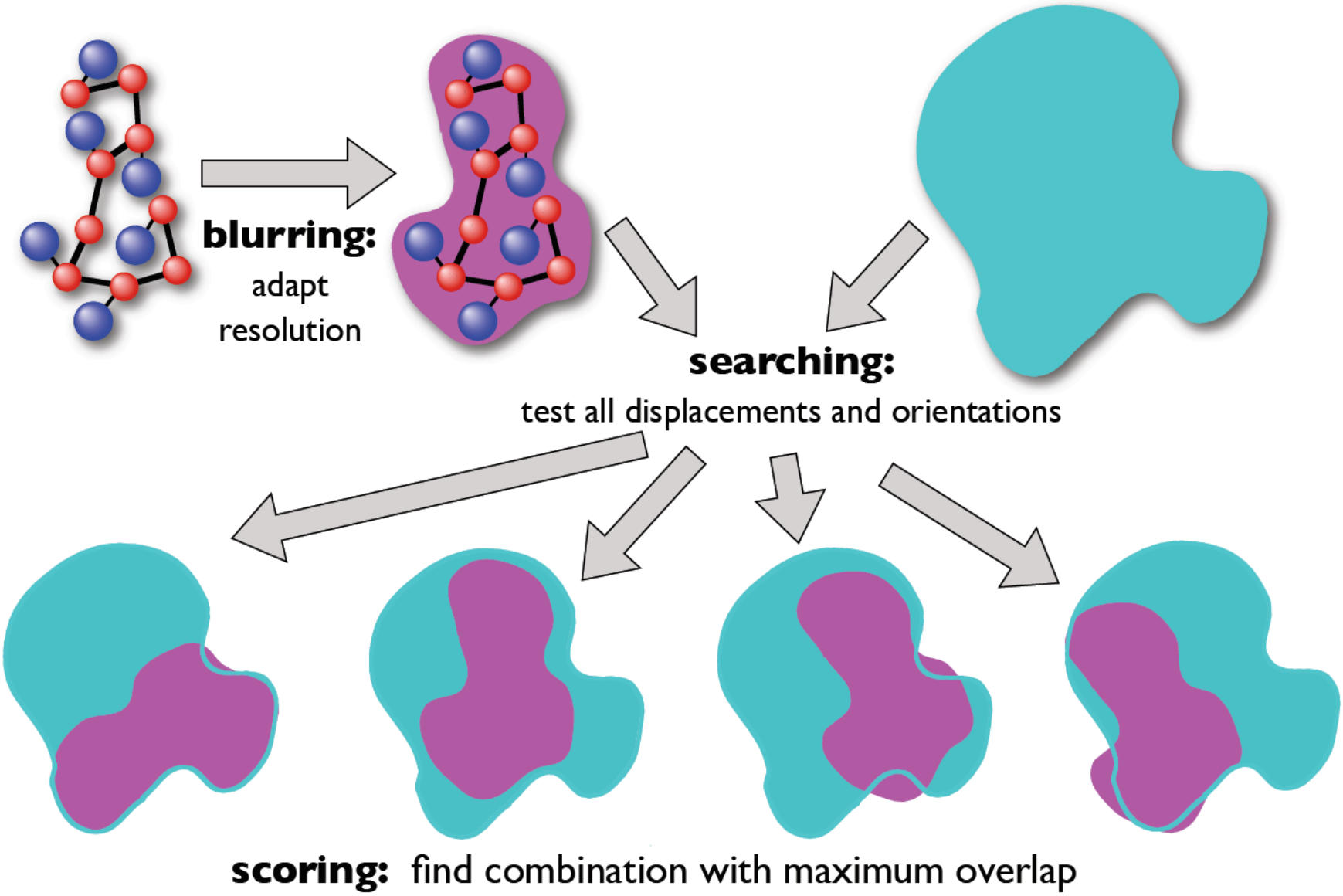
Atomistic structure of
a part of the complex

Coarse EM structure of
the whole complex



- same resolution for both structures
- exhaustive search with scoring
- choose best pose(s)

The procedure



Step 1: blurring the picture

Mathematically:
convolution of (exact) atomistic structure f(x)
with experimental resolution g(x)

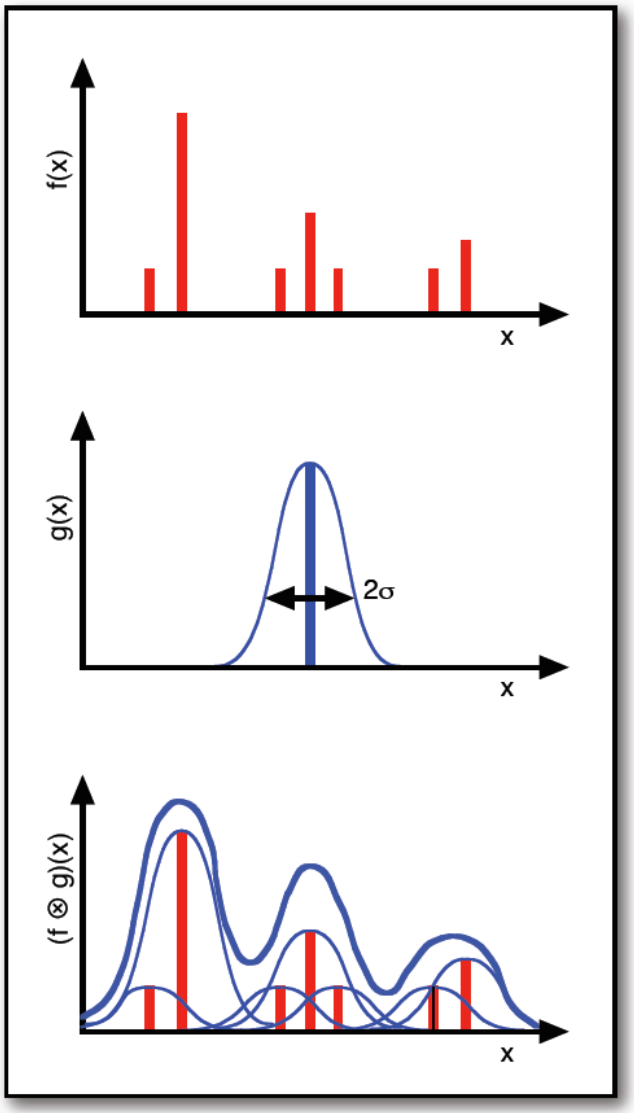
$$\tilde{f}(x) = (f \otimes g)(x) = \int dz\, f(z)\, g(x-z)$$

original data seen through the imaging apparatus

original signal

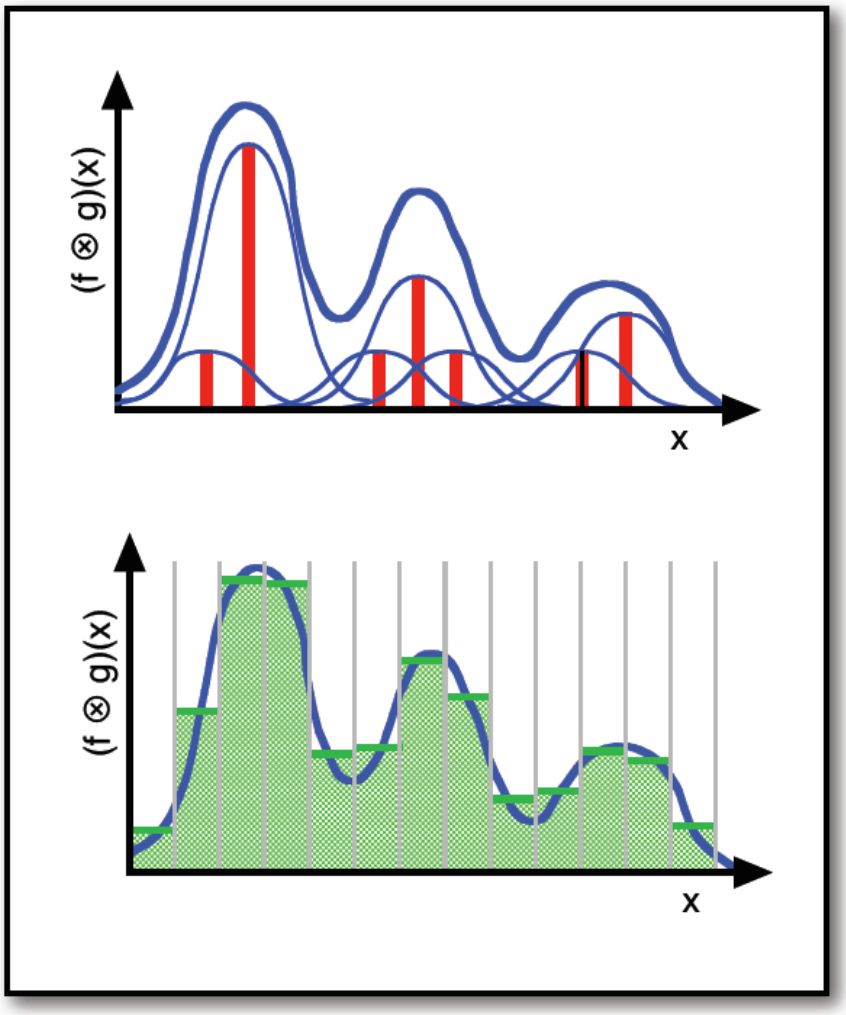
"kernel" => what is the image of a single point? (=delta signal)

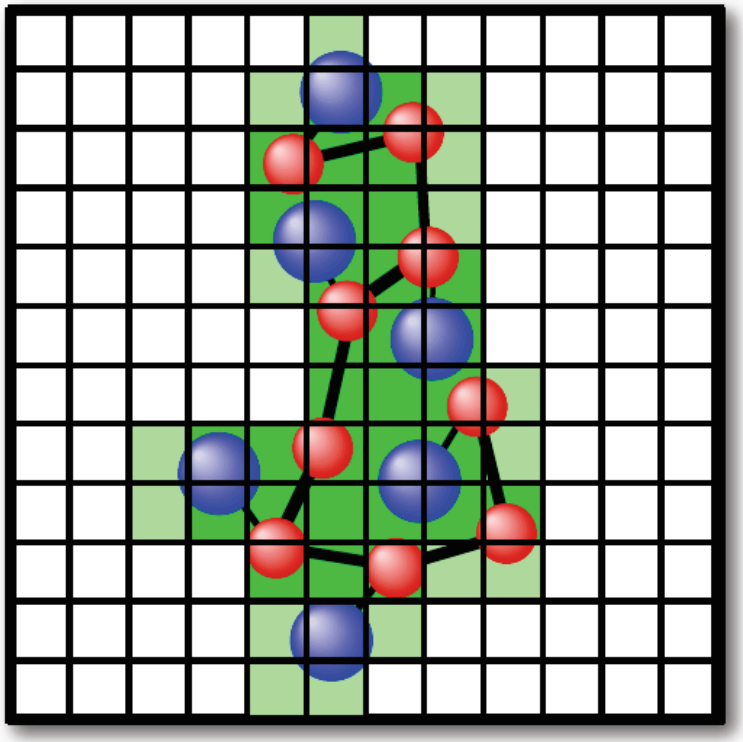
Often: blurring with gaussian



Put it on a grid

Discretize:





Fourier Transformation

Forward

$$F(k) = \int_{-\infty}^{\infty} dx\, e^{-ikx}\, f(x)$$

and inverse Fourier transform

$$f(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} dk\, e^{ikx}\, F(k)$$

with $e^{ikx} = \cos(kx) + i\sin(kx)$

=> convert between real and frequency (Fourier) space

short distances <=> high frequencies

long distances <=> low frequencies

Shift of the Argument

$$
\begin{aligned}
FT[f(x+\Delta x)] &= \int_{-\infty}^{\infty} dx\, e^{-ikx}\, f(x+\Delta x) \\
&= \int dy\, e^{-ik(y-\Delta x)}\, f(y) \\
&= e^{ik\,\Delta x} \int dy\, e^{-iky}\, f(y) \\
&= e^{ik\,\Delta x}\, FT[f(x)]
\end{aligned}
$$

Convolution

$$\tilde{f}(x) = (f \otimes g)(x) = \int dz\, f(z)\, g(x-z)$$

Apply FT on both sides:

$$FT[\tilde{f}(x)] = FT[(f \otimes g)(x)] =$$

$$=$$

$$=$$

$$= FT[f(x)]\ FT[g(x)]$$

Integration in real space is replaced by simple multiplication in Fourier space.

But FTs need to be computed.

What is more efficient?

If the same width g(x) is used for multiple displaced datasets
=> do FT[g(x)] only once

Fourier on a Grid

On a finite grid: => maximum wavelength = length of grid
=> minimum wavelength = grid spacing
=> sum instead of integral

$$F_k = \sum_{j=0}^{N-1} e^{-2i\pi\, j\, k/N}\, f_j \qquad f_k = \frac{1}{N} \sum_{j=0}^{N-1} e^{+2i\pi\, j\, k/N}\, F_j$$

Compute efficiently by (iteratively) splitting up into two sums
=> O(N log(N)) instead of $O(N^2)$ for direct summation for all N values of F_k
=> **F**ast **F**ourier **T**ransform (Cooley and Tukey, 1965)

$$F_k = \sum_{j=0}^{N-1} e^{-2i\pi\, j\, k/N}\, f_j$$

$$= \sum_{j=0}^{(N-1)/2} \left[e^{-2i\pi\, j\, k/(N/2)}\, f_j + e^{-2i\pi\, k/(N/2)}\, e^{-2i\pi\, j\, k/(N/2)}\, f_{j+1} \right]$$

And so forth

FFT by Danielson and Lanczos (1942)

Danielson and Lanczos showed that a discrete Fourier transform of length *N* can be rewritten as the sum of two discrete Fourier transforms, each of length *N/2.*

One of the two is formed from the even-numbered points of the original *N*, the other from the odd-numbered points.

$$
\begin{aligned}
F_k &= \sum_{j=0}^{N-1} e^{-2\pi i k \frac{j}{N}} f_j \\
&= \sum_{j=0}^{\frac{N}{2}-1} e^{-2\pi i k \frac{2j}{N}} f_{2j} + \sum_{j=0}^{\frac{N}{2}-1} e^{-2\pi i k \frac{2j+1}{N}} f_{2j+1} \\
&= \sum_{j=0}^{\frac{N}{2}-1} e^{-2\pi i k \frac{j}{\frac{N}{2}}} f_{2j} + e^{-2\pi i k \frac{1}{N}} \sum_{j=0}^{\frac{N}{2}-1} e^{-2\pi i k \frac{j}{\frac{N}{2}}} f_{2j+1} \\
&= F_k^e + e^{-2\pi i k \frac{1}{N}} F_k^o
\end{aligned}
$$

F_k^e : *k*-th component of the Fourier transform of length *N/2* formed from the even components of the original f_j 's

F_k^o : *k*-th component of the Fourier transform of length *N/2* formed from the odd components of the original f_j 's

FFT by Danielson and Lanczos (1942)

The wonderful property of the Danielson-Lanczos-Lemma is that it can be used recursively.

Having reduced the problem of computing F_k to that of computing F_k^e and F_k^o , we can do the same reduction of F_k^e to the problem of computing the transform of **its** *N/4* even-numbered input data and *N/4* odd-numbered data.

We can continue applying the DL-Lemma until we have subdivided the data all the way down to transforms of length 1.

What is the Fourier transform of length one? It is just the identity operation that copies its one input number into its one output slot.

For every pattern of $\log_2 N$ e‘s and o‘s, there is a one-point transform that is just one of the input numbers f_n

$$F_k^{eoeeoeo\ldots oee} = f_n \qquad \text{for some } n$$

FFT by Danielson and Lanczos (1942)

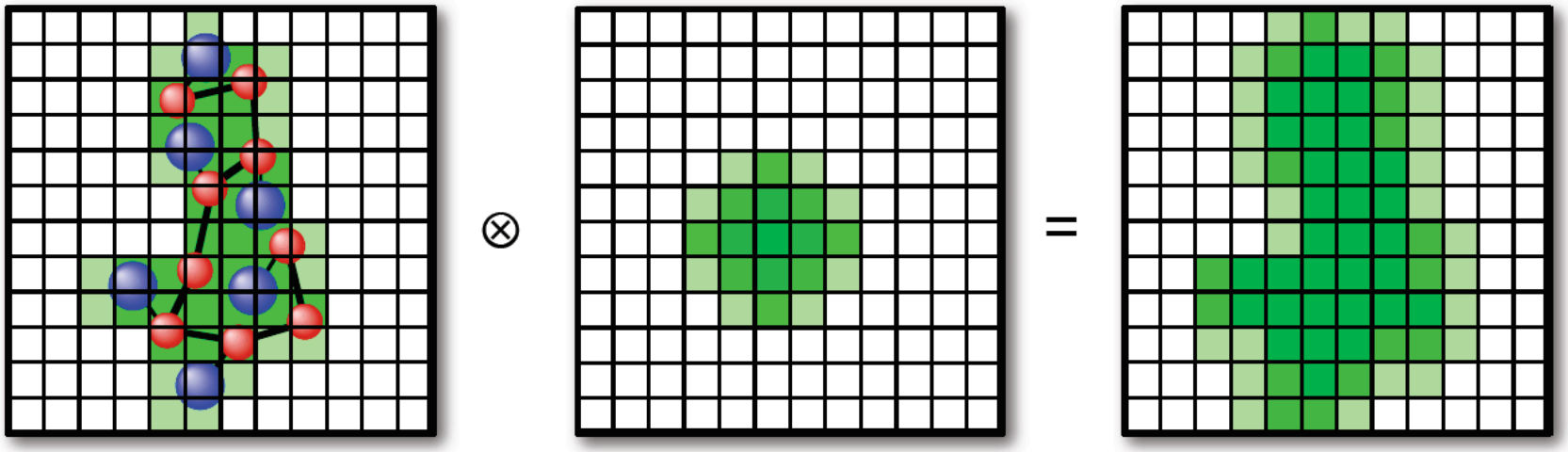
The next trick is to figure out which value of *n* corresponds to which pattern of e's and o's in $F_k^{eoeeoeo...oee} = f_n$

Answer: reverse the pattern of e's and o's, then let e = 0 and o = 1,
and you will have, in binary the value of n.

This works because the successive subdividisions of the data into even and odd are tests of successive low-order (least significant) bits of *n*.

Discretization and Convolution

For practical applications:
=> first put atomistic data onto the grid, then blur with FFT



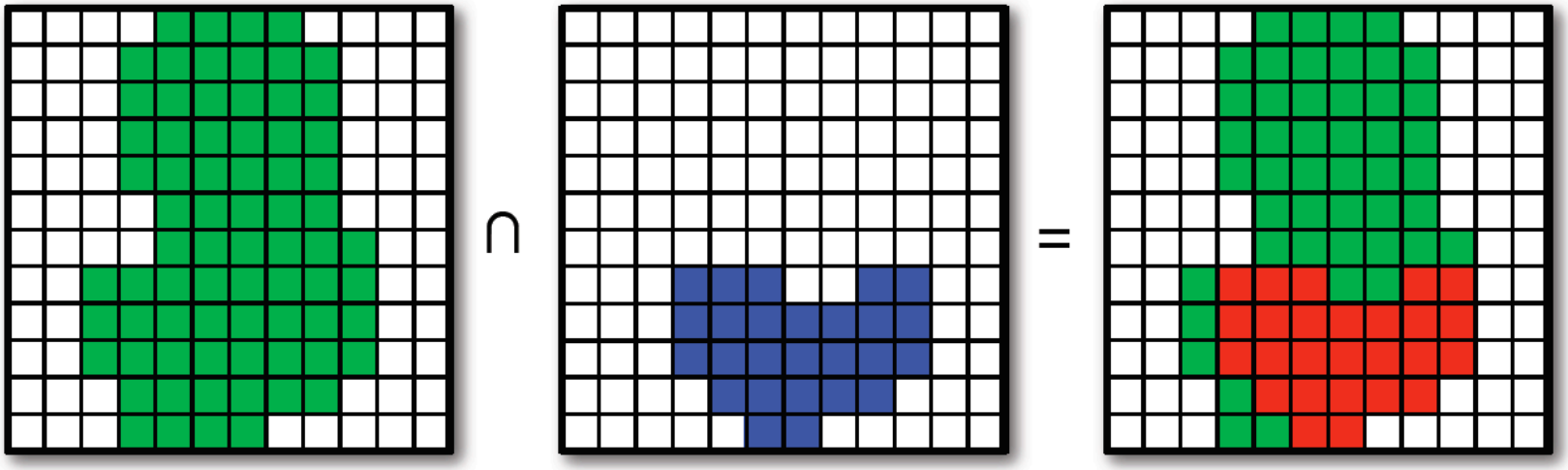
discretized hi-res data ⊗ blurring kernel = low-res image

Step 3: Scoring the Overlap

Most simple case:

- apply density threshold and count overlapping voxels
- displace images relative to each other, recount

=> find displacement with maximum overlap



In matrix form with displacements x, y:

$$c(x,y) = \sum_{l=1}^{N} \sum_{m=1}^{N} a_{l,m}\, b_{l+x,m+y}$$

Cross Correlation

Generalization: maximize cross correlation of grided densities with respect to displacement (and orientation)

$$C_{x,y,z} = \sum_{l=1}^{N} \sum_{m=1}^{N} \sum_{n=1}^{N} a_{l,m,n} \times b_{l+x,m+y,n+z}$$

Note: maximize the cross-correlation <=> minimize the squared difference

On a grid with N^3 gridpoints => N^3 possible displacements
=> runtime $O(N^6)$

Further complication: the convolution

$$C_{x,y,z} = \sum_{l=1}^{N} \sum_{m=1}^{N} \sum_{n=1}^{N} a_{l,m,n} \times (g \otimes b_{l+x,m+y,n+z})$$

Correlation and Fourier

Apply Fourier transformation to both sides of

$$C_{x,y,z} = \sum_{l=1}^{N} \sum_{m=1}^{N} \sum_{n=1}^{N} a_{l,m,n} \times b_{l+x,m+y,n+z}$$

=> matrix multiplication

$$FT[C] = FT[A]^* \times FT[B]$$

Runtime of 3D FFT = $O(N^3 \log^3(N)) << N^6$
=> all possible displacements tested simultaneously

Note: FT[A] only calculated once initially
=> two FFTs per orientation
=> scan orientation via Euler angles

<== Step 2: exhaustive search

Include convolution

Maximize

$$C_{x,y,z} = \sum_{l=1}^{N} \sum_{m=1}^{N} \sum_{n=1}^{N} a_{l,m,n} \times (g \otimes b_{l+x,m+y,n+z})$$

In Fourier space:

Insert convolution $FT[f \otimes g] = FT[f] \times FT[g]$

Into correlation:

$$\begin{aligned} FT[C] &= FT[A]^* \times FT[G \otimes B] \\ &= FT[A]^* \times (FT[G] \times FT[B]) \\ &= (FT[A]^* \times FT[G]) \times FT[B] \end{aligned}$$

can be precomputed

2 FFTs + 1 matrix multiplication

Katchalski-Kazir algorithm

Proc. Natl. Acad. Sci. USA
Vol. 89, pp. 2195–2199, March 1992
Biophysics

Molecular surface recognition: Determination of geometric fit between proteins and their ligands by correlation techniques

(protein–protein interaction/surface complementarity/macromolecular complex prediction/molecular docking)

EPHRAIM KATCHALSKI-KATZIR[†‡], ISAAC SHARIV[§], MIRIAM EISENSTEIN[¶], ASHER A. FRIESEM[§], CLAUDE AFLALO[‖], AND ILYA A. VAKSER[†]

Departments of [†]Membrane Research and Biophysics, [§]Electronics, [¶]Structural Biology, and [‖]Biochemistry, Weizmann Institute of Science, Rehovot 76100, Israel

Contributed by Ephraim Katchalski-Katzir, October 24, 1991

Developed for protein-ligand docking
<=> same techniques applicable for docking "on the inside"

Discretization for docking

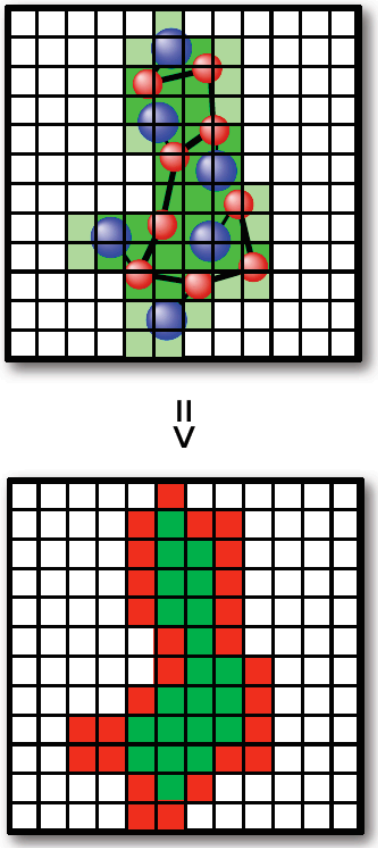
Next, to distinguish between the surface and the interior of each molecule, we retain the value of 1 for the grid points along a thin surface layer only and assign other values to the internal grid points. The resulting functions thus become

$$\bar{a}_{l,m,n} = \begin{cases} 1 & \text{on the surface of the molecule} \\ \rho & \text{inside the molecule} \\ 0 & \text{outside the molecule,} \end{cases} \qquad [2a]$$

and

$$\bar{b}_{l,m,n} = \begin{cases} 1 & \text{on the surface of the molecule} \\ \delta & \text{inside the molecule} \\ 0 & \text{outside the molecule,} \end{cases} \qquad [2b]$$

where the surface is defined here as a boundary layer of finite width between the inside and the outside of the molecule. The parameters ρ and δ describe the value of the points inside the molecules, and all points outside are set to zero. Two-

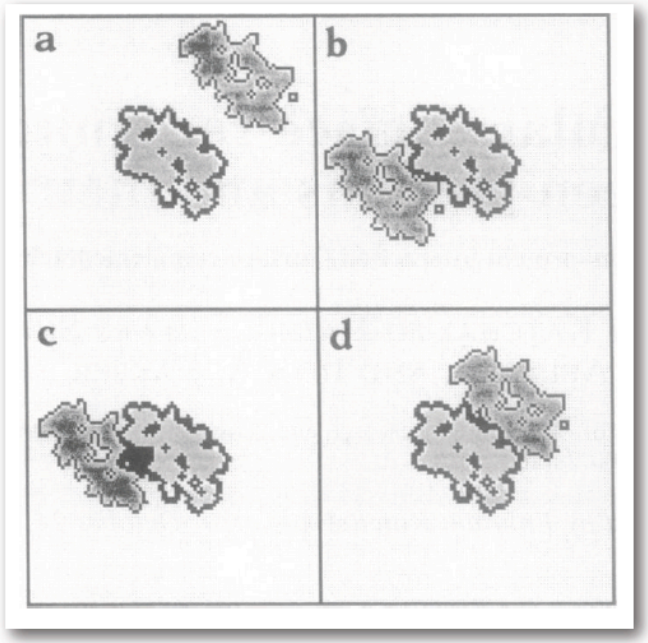


Typical values: $\rho = -15$, $\delta = 1$

=> penalty for overlap of volumes

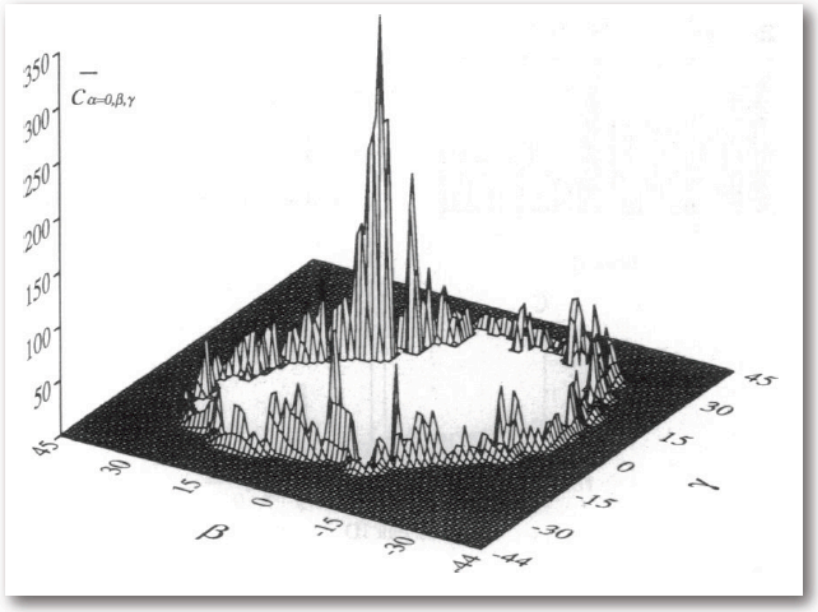
Docking the hemoglobin dimer

2D cross sections at l = 46 (N = 90)



a) no contact
b) limited contact
c) overlap (black area)
d) good geometric match

Correlation at α = 0



highest peak corresponds to native dimer arrangement

The algorithm

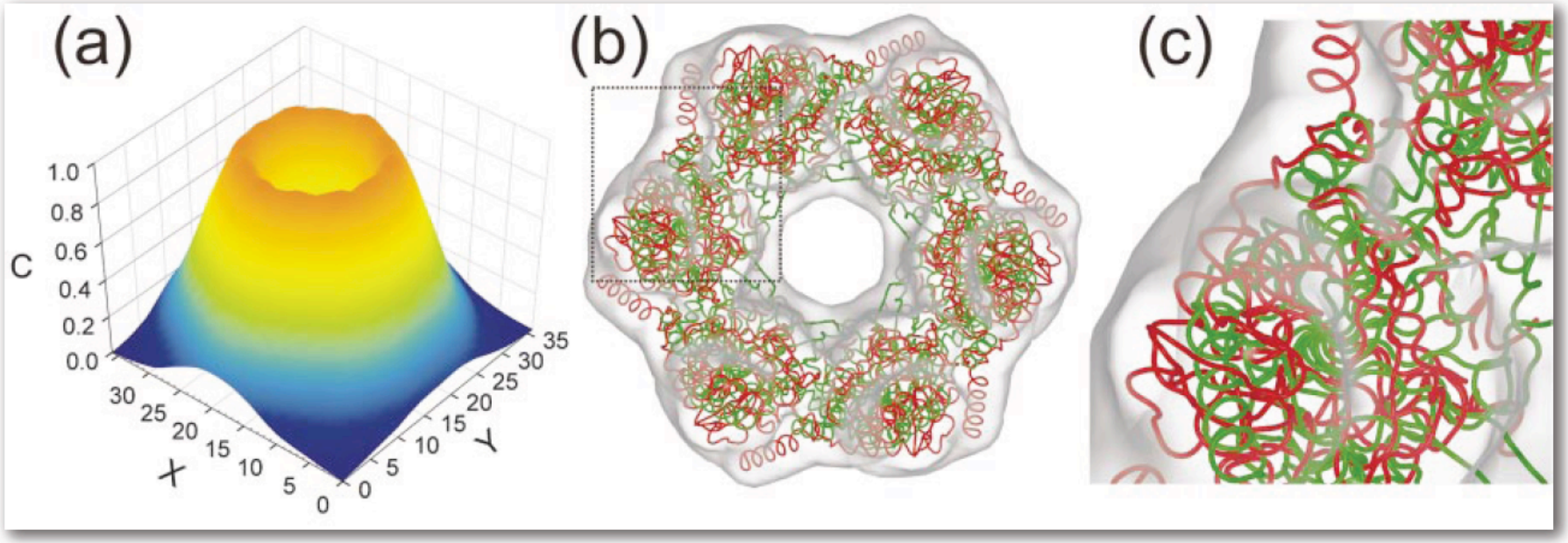
The entire procedure described above can be summarized by the following steps:
(*i*) derive $\bar{a}$ from atomic coordinates of molecule **a** (Eq. **2**),
(*ii*) $A^* = [\mathrm{DFT}(\bar{a})]^*$ (Eq. **4**),
(*iii*) derive $\bar{b}$ from atomic coordinates of molecule **b** (Eq. **2**),
(*iv*) $B = \mathrm{DFT}(\bar{b})$ (Eq. **4**),
(*v*) $C = A^* \cdot B$ (Eq. **5**),
(*vi*) $\bar{c} = \mathrm{IFT}(C)$ (Eq. **6**),
(*vii*) look for a sharp positive peak of $\bar{c}$,
(*viii*) rotate molecule **b** to a new orientation,
(*ix*) repeat steps *iii–viii* and end when the orientations scan is completed, and
(*x*) sort all of the peaks by their height.

Each high and sharp peak found by this procedure indicates geometric match and thus represents a potential complex. The relative position and orientation of the molecules within each such complex can readily be derived from the

Katchalski-Kazir et al. 1992

Algorithm has become a workhorse for docking and density fitting.

Problem I: limited contrast



Docking of the RecA helicase monomer into simulated EM density of the hexamer at 15 Å resolution
(exhaustive 6D search with 5Å / 9° steps plus off-lattice optimization)
=> multiple fits with similar correlations

Chacón, Wriggers, *J. Mol. Biol.* **317** (2002) 375

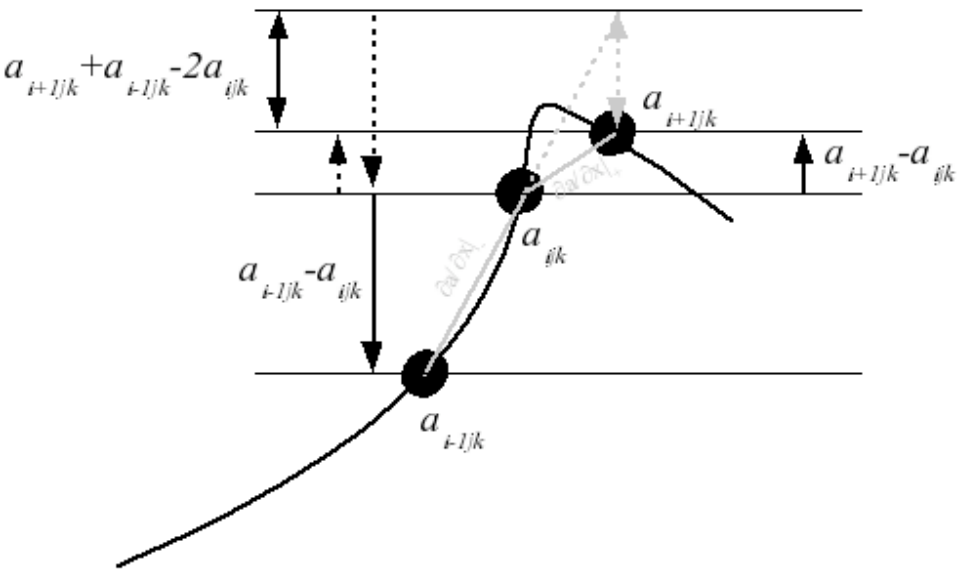
Laplace filter

Evaluate $\nabla^2 = \frac{d^2}{dx^2} + \frac{d^2}{dy^2} + \frac{d^2}{dz^2}$

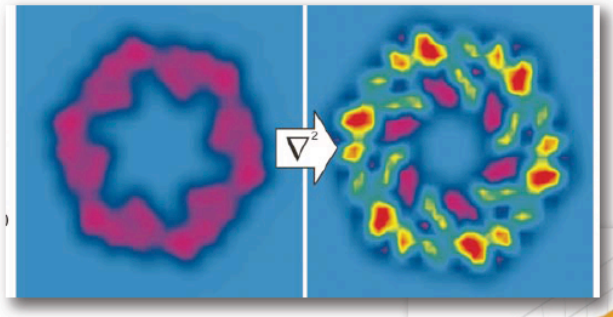
on a grid: $\nabla^2 a_{l,m,n} = -6a_{l,m,n} + a_{l+1,m,n} + a_{l-1,m,n} + a_{l,m+1,n} + a_{l,m-1,n} + a_{l,m,n+1} + a_{l,m,n-1}$

Correlation:

$$C_{x,y,z} = \sum_{l=1}^{N}\sum_{m=1}^{N}\sum_{n=1}^{N}(\nabla^2 \otimes a_{l,m,n}) \times (\nabla^2 \otimes g \otimes b_{l+x,m+y,n+z})$$

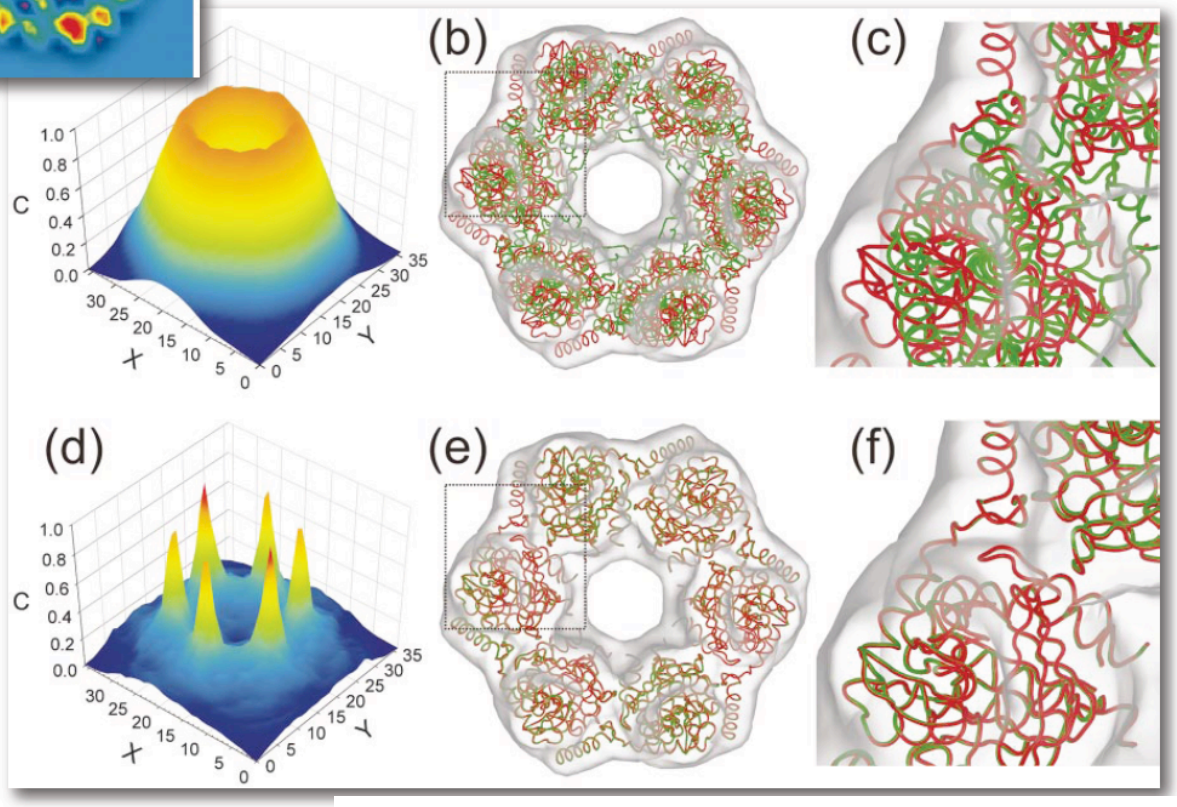


Enhanced contrast → better fit



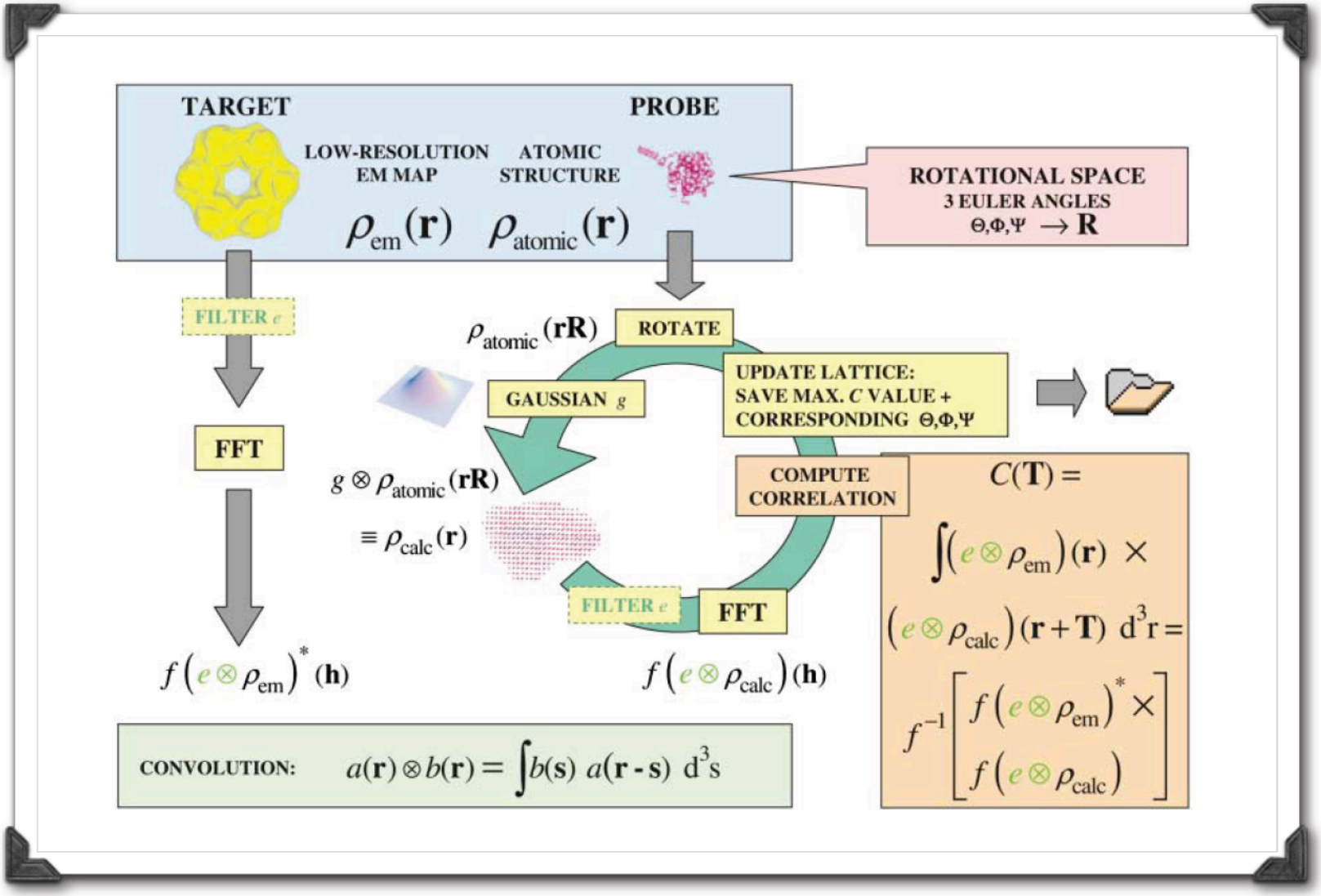
With the density alone:

With the Laplacian filter:



Chacón, Wriggers, *J. Mol. Biol.* **317** (2002) 375

The big picture



Wriggers, Chacón, *Structure* **9** (2001) 779

Problem 2: more efficient search

BIOINFORMATICS ORIGINAL PAPER

Vol. 23 no. 4 2007, pages 427–433
doi:10.1093/bioinformatics/btl625

Structural bioinformatics

ADP_EM: fast exhaustive multi-resolution docking for high-throughput coverage

José Ignacio Garzón, Julio Kovacs[1], Ruben Abagyan[1] and Pablo Chacón*

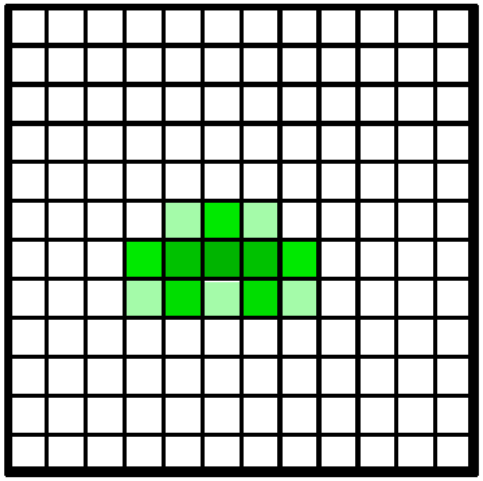
Centro de Investigaciones Biológicas, CSIC, Ramiro de Maeztu 9, 28040 Madrid, Spain and
[1]Department of Molecular Biology, The Scripps Research Institute La Jolla, CA 92037, USA

Received on September 28, 2006; revised on November 28, 2006; accepted on December 4, 2006
Advance Access publication December 6, 2006
Associate Editor: Alex Bateman

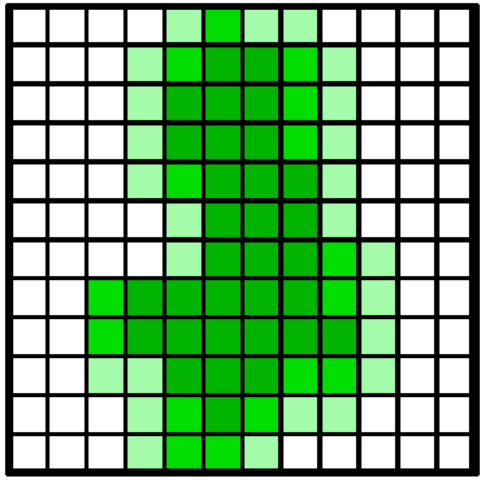
Observations:

- many displacements can be excluded a priori (FFT alg. calculates them all)
- FFT idea makes more sense for rotations (no simple limit on rotations)

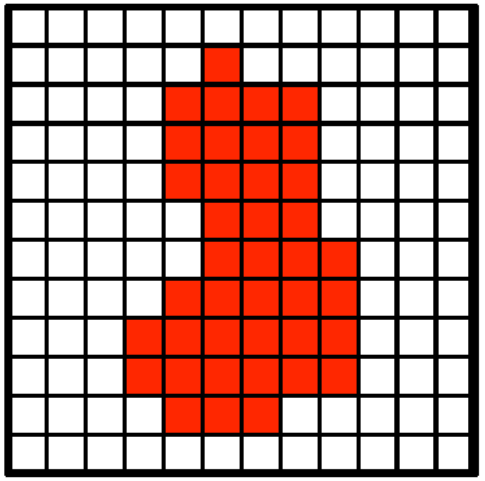
Masked displacements



Probe



Target



Mask =
potential hits

Search space for displacements =
(inside of the target molecule) – (extent of the probe)

Rotational search

Express densities in spherical harmonics on "onion shells"

$$\rho_{\text{low}}(r,\beta,\lambda) = \sum_{l=0}^{B-1} \sum_{m=-l}^{l} C_{lm}^{\text{low}}(r) Y_{lm}(\beta,\lambda) \qquad \rho_{\text{high}}(r,\beta,\lambda) = \sum_{l=0}^{B-1} \sum_{m=-l}^{l} C_{lm}^{\text{high}}(r) Y_{lm}(\beta,\lambda),$$

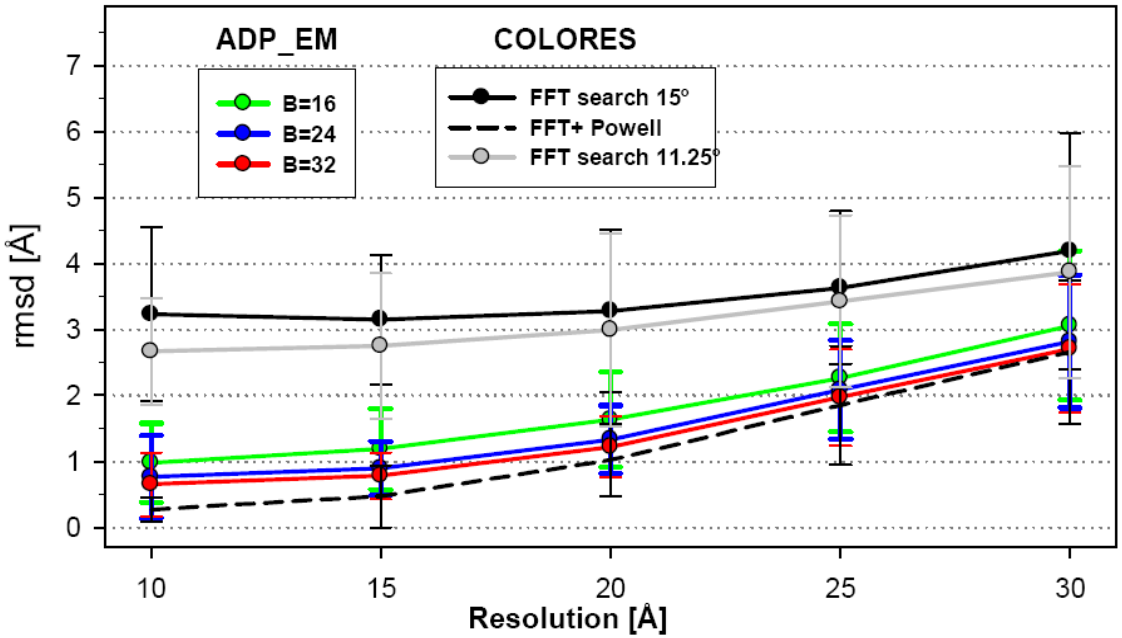
Y_{lm}	l = 0	l = 1	l = 2	l = 3
m = -3				$\sqrt{\frac{35}{64\pi}}\sin^3\vartheta\, e^{-3i\varphi}$
m = −2			$\sqrt{\frac{15}{32\pi}}\sin^2\vartheta\, e^{-2i\varphi}$	$\sqrt{\frac{105}{32\pi}}\sin^2\vartheta\cos\vartheta\, e^{-2i\varphi}$
m = −1		$\sqrt{\frac{3}{8\pi}}\sin\vartheta\, e^{-i\varphi}$	$\sqrt{\frac{15}{8\pi}}\sin\vartheta\,\cos\vartheta\, e^{-i\varphi}$	$\sqrt{\frac{21}{64\pi}}\sin\vartheta\,(5\cos^2\vartheta-1)\, e^{-i\varphi}$
m = 0	$\sqrt{\frac{1}{4\pi}}$	$\sqrt{\frac{3}{4\pi}}\cos\vartheta$	$\sqrt{\frac{5}{16\pi}}(3\cos^2\vartheta-1)$	$\sqrt{\frac{7}{16\pi}}(5\cos^3\vartheta-3\cos\vartheta)$
m = 1		$-\sqrt{\frac{3}{8\pi}}\sin\vartheta\, e^{i\varphi}$	$-\sqrt{\frac{15}{8\pi}}\sin\vartheta\,\cos\vartheta\, e^{i\varphi}$	$-\sqrt{\frac{21}{64\pi}}\sin\vartheta\,(5\cos^2\vartheta-1)\, e^{i\varphi}$
m = 2			$\sqrt{\frac{15}{32\pi}}\sin^2\vartheta\, e^{2i\varphi}$	$\sqrt{\frac{105}{32\pi}}\sin^2\vartheta\cos\vartheta\, e^{2i\varphi}$
m = 3				$-\sqrt{\frac{35}{64\pi}}\sin^3\vartheta\, e^{3i\varphi}$

Correlation for **all orientations** at a given displacement:

$$C(R) = FT^{-1}_{m,h,m'}\left[\sum_l d^l_{mh}\; d^l_{hm'} \int_0^\infty C_{lm}^{\text{low}}(r)\overline{C_{lm'}^{\text{high}}(r)} r^2 dr\right]$$

Known Fourier coefficients of spherical harmonics Y_{lm}.

Accuracy



rmsd with respect to known atomistic structure of target.

Registration accuracy on simulated EM maps of 28 structures
for bandwidths (number of angular sampling points) of
B = 16, 24 32 (11°, 8°, ~ 6°)
compared to Wriggers' COLORES (situs package – Katchalski-Katzir algorithm + local Powell optimization)

Performance

Table 1. Timing results, in seconds, obtained with the benchmark described in Figure 1

	Sampling B/°	Resolution 10Å	15Å	20Å	25Å	30Å
ADP_EM	16/11°	28	31	35	34	38
	24/8°	100	108	119	118	123
	32/6°	226	220	225	216	221
FFT search	–/15°	1697	1926	2341	5028	6681
Powell minim	–/15°	375	918	1747	3739	6597

ADP_EM (Another Docking Platform for EM) is much faster

- only limited spatial region is scanned
- fast evaluation of the orientational correlation via FFT
- spherical harmonics allow for better rotational representation
  => higher accuracy

Some examples

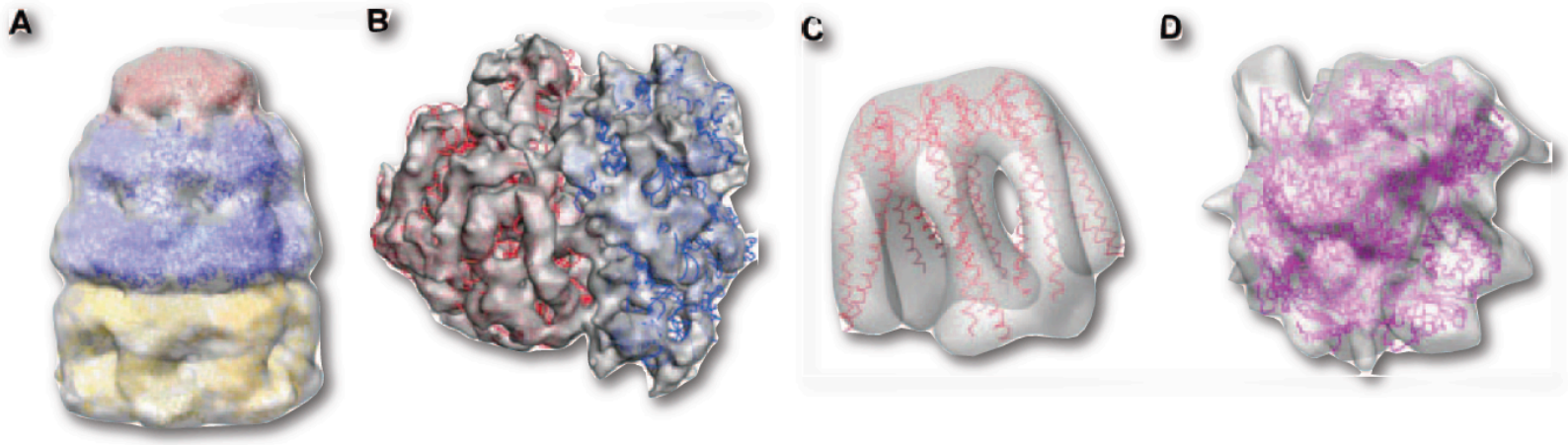


Fig. 2. Docking results with experimental EM data. (**A**) *E.coli* GroES-ADP7-GroEL-ATP7 from *E.coli* at 23.5 Å (EMD ID 1046, PDB: 1ml5); ADP and ATP GroEL subunits have been docked independently to reconstruct the cis and trans heptameric rings of the complex. For GroES the whole heptamer was used. (**B**) Docking of 30S and 50S subunits into *E.coli* ribosome map at 14 Å (EMD ID 1046, PDB: 1gix/1giy). Single-molecule docking of prefoldin (**C**) at 23 Å (Martin-Benito *et al.*, 2002), PDB: 1l6h, and of yeast RNA polymerase II (**D**) at 15 Å (Craighead *et al.*, 2002), PDB: 1fxk.

Summary

Today:

Docking into EM maps

- Discretization
- Correlation and blurring via FFT => Katchalski-Katzir algorithm
- Laplace filter => enhances contrast
- ADP_EM: FFT for rotations, scan displacements => better performance

Next lecture V19:

using connectivity information for complex assembly