Crystallography without Crystals

Breaking the Crystallization Paradigm

Dilano Saldin

dksaldin@uwm.edu

Main Collaborators

Valentin Shneerson – University of Wisconsin-Milwaukee

Hin-Cheuck Poon – University of Wisconsin-Milwaukee

John Spence – Arizona State University

Rick Kirian – Arizona State Univeristy

Kevin Schmidt – Arizona State University

Uwe Weierstall – Arizona State University

Henry Chapman – DESY

Malcolm Howells - LBNL

Statement of Problem

- The function of molecules follow their structure. Hence importance of structure determination
- Traditional workhorse, x-ray crystallography, requires crystals, but not all molecules are cystallizable
- Crystallography relies on amplification due to scattering by many identical copies in identical orientations
- One alternative is the single-molecule experiment discussed by other speakers
- The method we will describe may be used for structure solution for that problem
- There is also another alternative – scatter off many identical particles in random orientations and recover a single-particle diffraction pattern from the angular correlations

Protein Crystallography - Phase Problem

Fourier transform

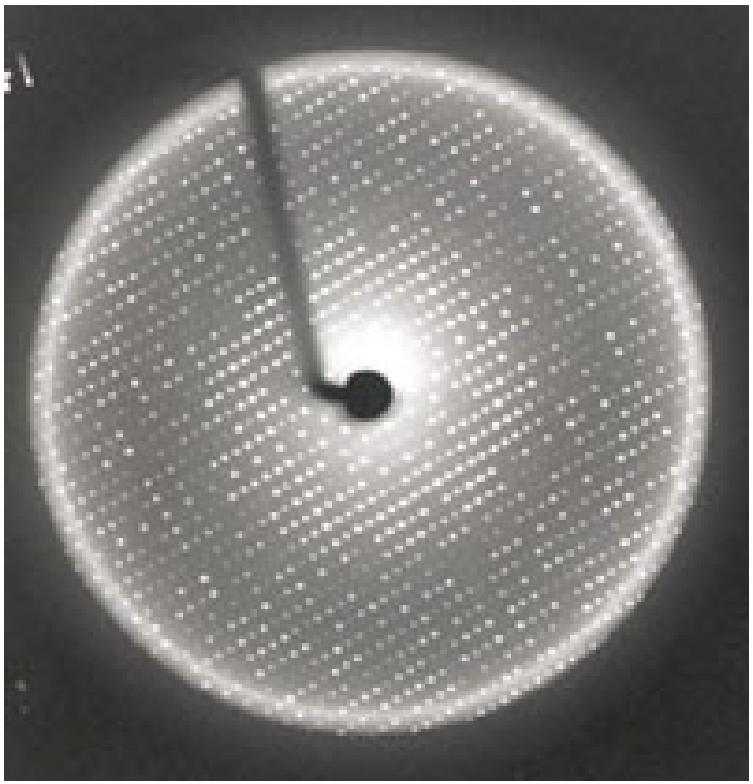
$$\left|F_{\mathbf{q}}\right|\exp\left(i\phi_{\mathbf{q}}\right)=\sum_{j}u_{j}\exp(i\mathbf{q}.\mathbf{r}_{j})$$

Measured $\left|F_{\mathbf{q}}\right|^{2}\Rightarrow\left|F_{\mathbf{q}}\right|$

Inverse transform

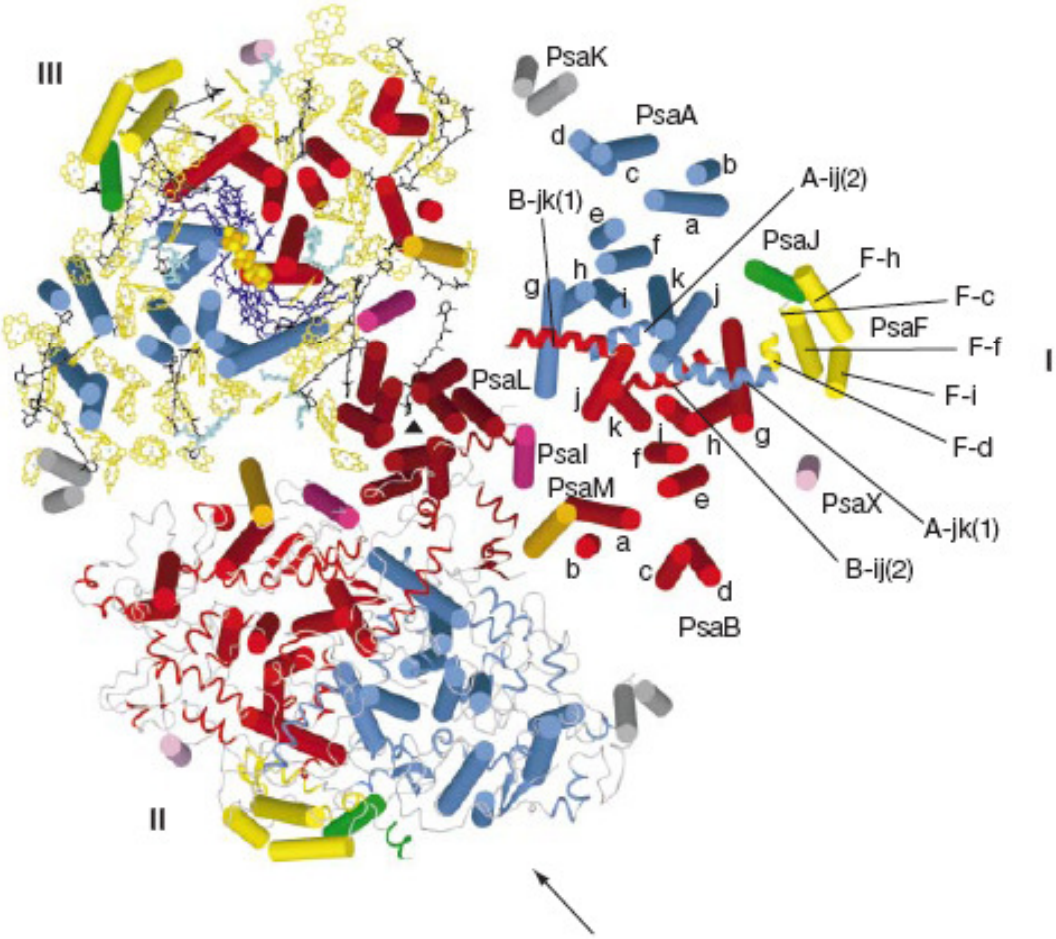
$$u_{j}=\frac{1}{N}\sum_{\mathbf{q}}\left|F_{\mathbf{q}}\right|\exp(i\phi_{\mathbf{q}})\exp(-i\mathbf{q}.\mathbf{r}_{j})$$

Not measured $\phi_{\mathbf{q}}$



Diffraction pattern from a protein crystal

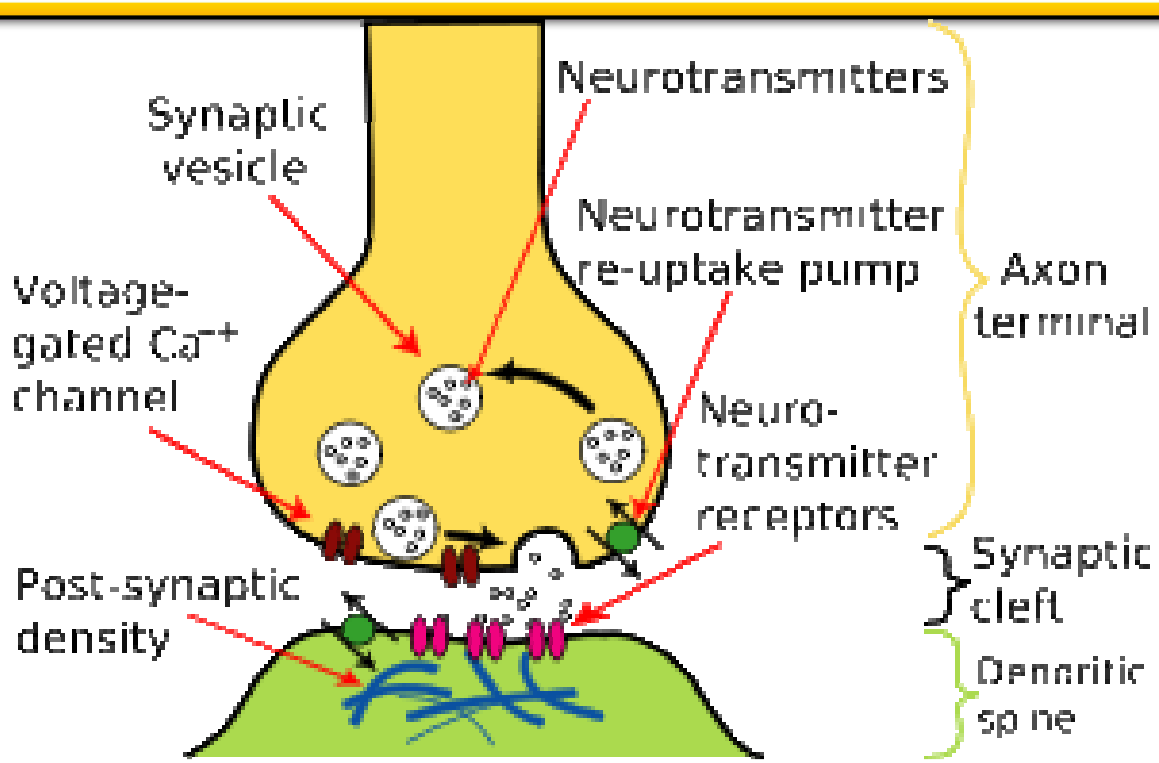
Cyanobacterial Photosystem I



P. Jordan et al.
Nature 411,
909-917 (2001)

2.5 Å
resolution

Neurotransmission



When an action potential arrives at the end of the pre-synaptic axon (yellow), it causes the release of neurotransmitter molecules that open ion channels in the post-synaptic neuron (green). The combined excitatory and inhibitory postsynaptic potentials of such inputs can begin a new action potential in the post-synaptic neuron.

Neurotoxins

Several neurotoxins, both natural and synthetic, are designed to block ion channels. Tetrodotoxin from the pufferfish block action potentials by inhibiting the voltage-sensitive sodium channel; similarly, dendrotoxin from the black mamba snake inhibits the voltage-sensitive potassium channel. Such inhibitors of ion channels make effective neurotoxins, and have been considered for use as chemical weapons.

Insecticides and Anaesthetics

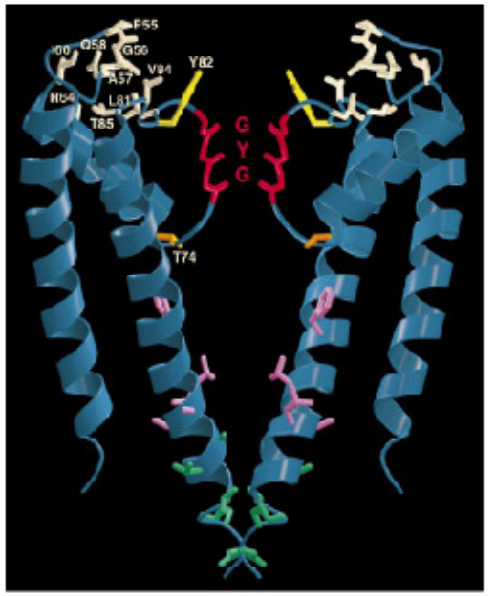
Neurotoxins aimed at the ion channels of insects have been effective insecticides; one example is the synthetic permethrin, which prolongs the activation of the sodium channels involved in action potentials. The ion channels of insects are sufficiently different from their human counterparts that there are few side effects in humans.

Many other neurotoxins interfere with the transmission of the action potential's effects at the synapses, especially at the neuromuscular junction.

Anesthetics work in a similar way, by blocking the transmission of nerve signals by blocking membrane protein ion channels.

Molecular Structure of Membrane Proteins

- Molecular structure of the membrane protein forming the K-channel was found in 1998 by Roderick MacKinnon and collaborators by x-ray crystallography
- Led to the 2003 Nobel Prize for Chemistry



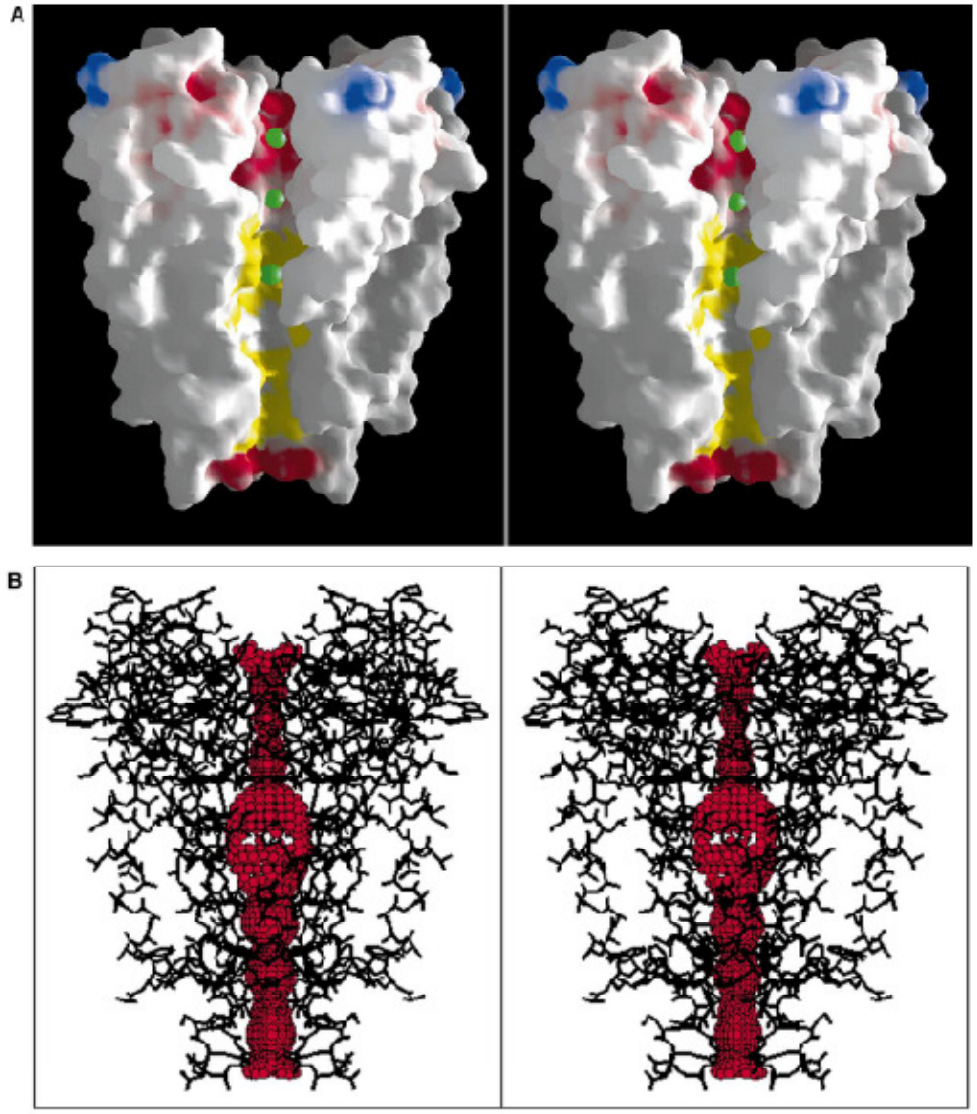
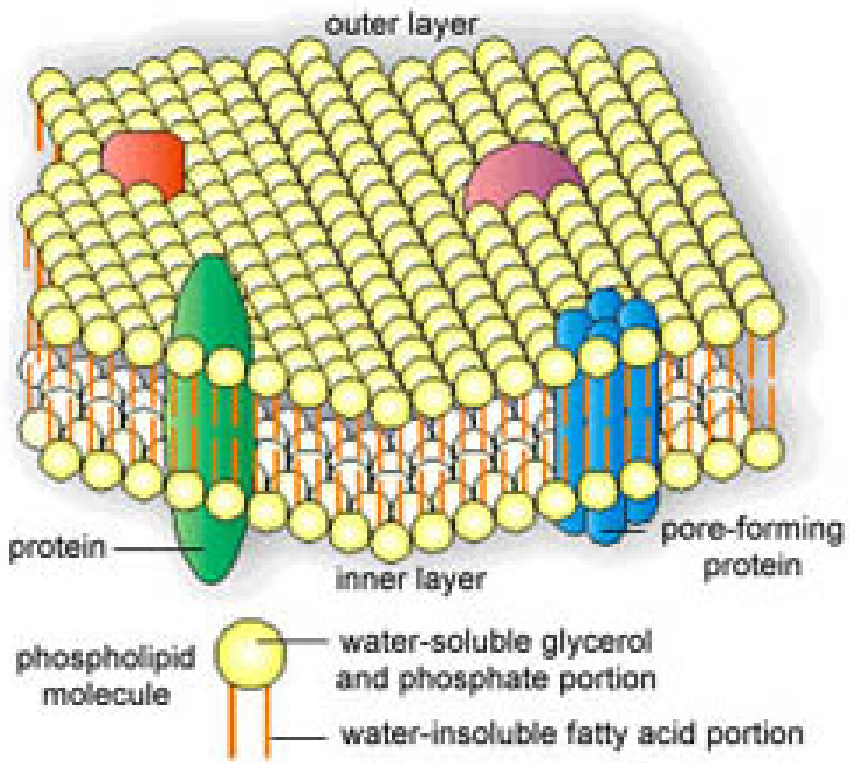


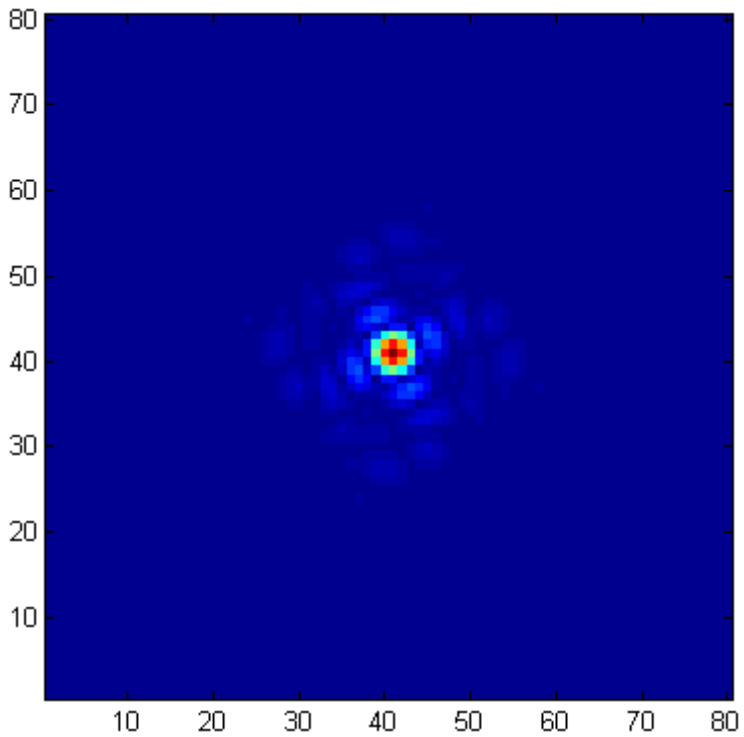
Fig. 4 (above). Mutagenesis studies on *Shaker*: Mapping onto the KcsA structure. Mutations in the voltage-gated *Shaker* K^+ channel that affect function are mapped to the equivalent positions in KcsA based on the sequence alignment. Two subunits of KcsA are shown. Mutation of any of the white side chains significantly alters the affinity of agitoxin2 or charybdotoxin for the *Shaker* K^+ channel (*12*). Changing the yellow side chain affects both agitoxin2 and TEA binding from the extracellular solution (*14*). This residue is the external TEA site. The mustard-colored side chain at the base of the selectivity filter affects TEA binding from the intracellular solution [the internal TEA site (*15*)]. The side chains colored green, when mutated to cysteine, are modified by cysteine-reactive agents whether or not the channel gate is open, whereas those colored pink react only when the channel is open (*16*). Finally, the residues colored red (GYG, main chain only) are absolutely required for K^+ selectivity (*4*). This figure was prepared with MOLSCRIPT and RASTER-3D. **Fig. 5 (right).** Molecular surface of KcsA and contour of the pore. (**A**) A cutaway stereoview displaying the solvent-accessible surface of the K^+ channel colored according to physical properties. Electrostatic potential was calculated with the program GRASP, assuming an ionic strength equivalent to 150 mM KCl and dielectric constants of 2 and 80 for protein and solvent, respectively. Side chains of Lys, Arg, Glu, and Asp residues were assigned single positive or negative charges as appropriate, and the surface coloration varies smoothly from blue in areas of high positive charge through white to red in negatively charged regions. The yellow areas of the surface are colored according to carbon atoms of the hydrophobic (or partly so) side chains of several semi-conserved residues in the inner vestibule (Thr^{75}, Ile^{100}, Phe^{103}, Thr^{107}, Ala^{108}, Ala^{111}, Val^{115}). The green CPK spheres represent K^+ ion positions in the conduction pathway. (**B**) Stereoview of the entire internal pore. Within a stick model of the channel structure is a three-dimensional representation of the minimum radial distance from the center of the channel pore to the nearest van der Waals protein contact. The display was created with the program HOLE (*34*).

Determine Structure of Membrane Proteins *In Situ* with X-Rays

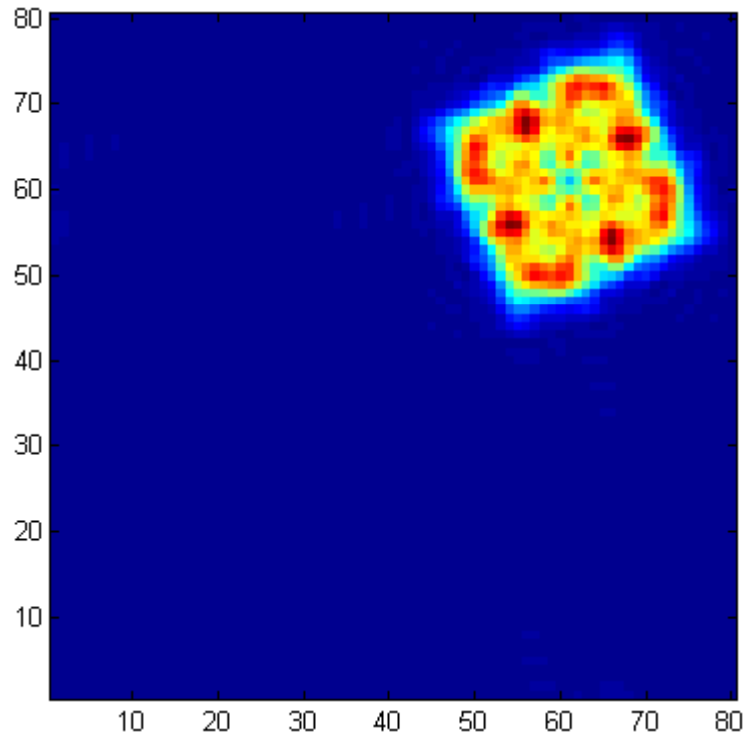


Proteins will be in random locations in membrane and in random orientations. Illuminate with x-rays part of membrane containing many proteins and record the diffraction pattern.

Projected Electron Density of K-Channel from Diffraction Pattern

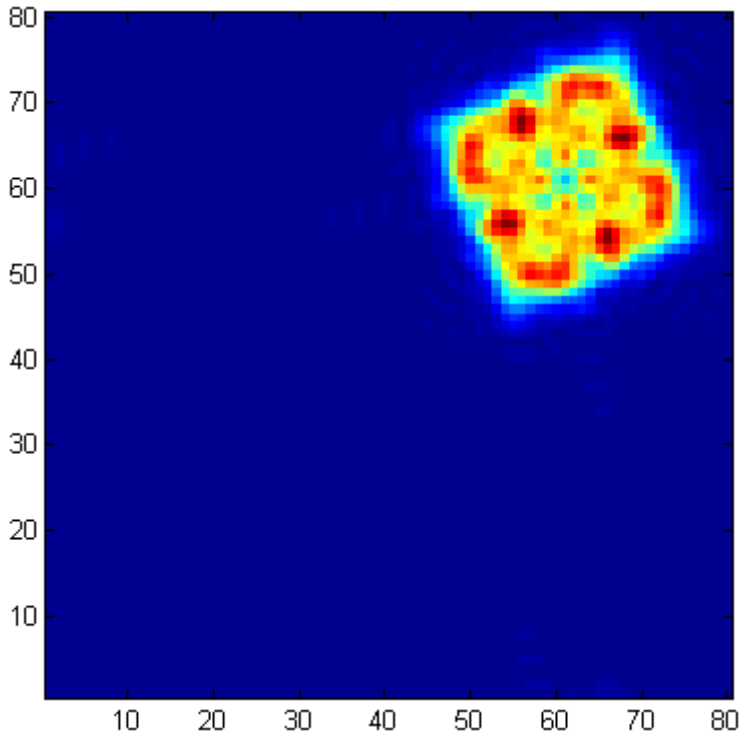


Simulated diffraction pattern

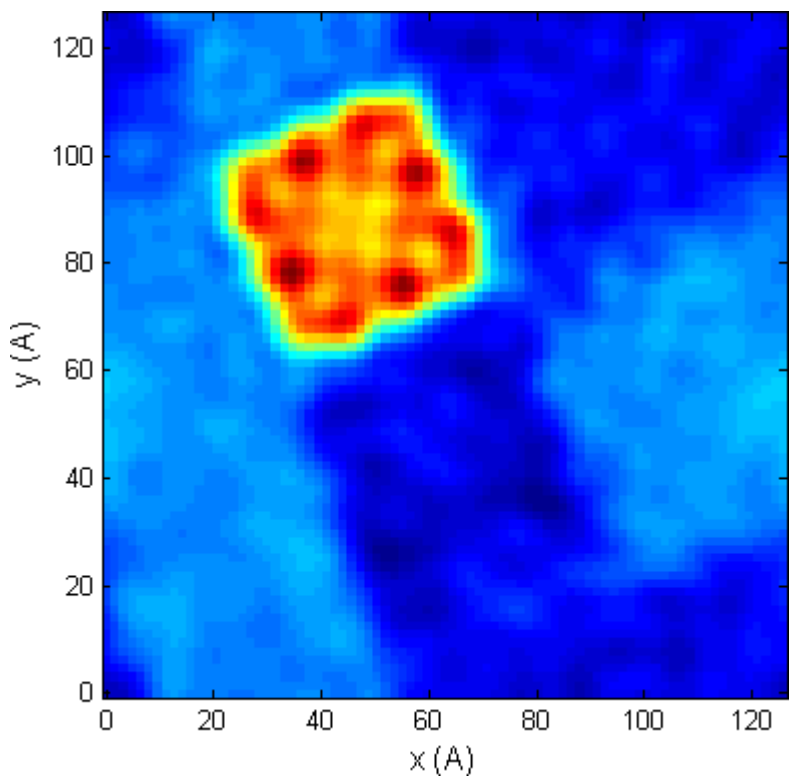


Ideal projected structure

Ideal and Recovered Projection of K-Channel Structure



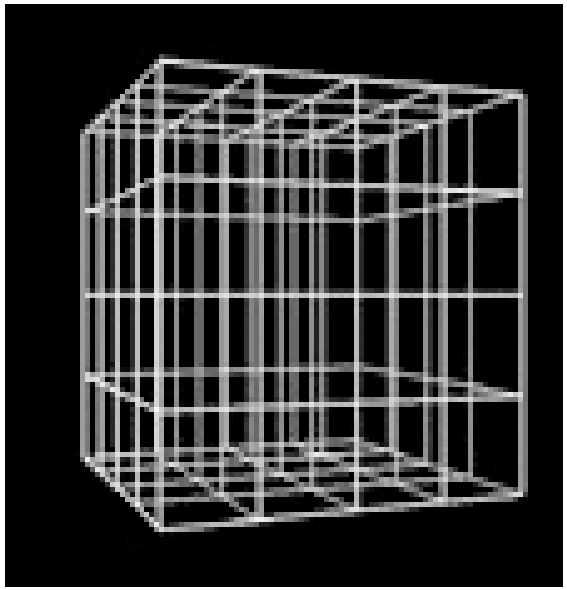
Ideal projected structure



From diffracted intensities, after 100 iterations of phasing algorithm

Iterative Phasing Algorithm

Fourier Space



Constrain to measured diffraction amplitudes

Start w/ random φ_q

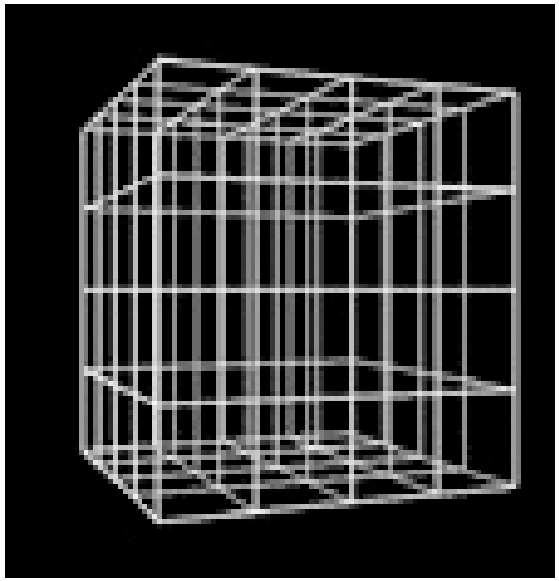
$$u_j = \sum_{\mathbf{q}} \left| F_{\mathbf{q}}^{(obs)} \right| e^{i\phi_{\mathbf{q}}} e^{-i\mathbf{q}.\mathbf{r}_j}$$

Retain $\left| \phi_{\mathbf{q}} \right|$

Replace $\left| F_{\mathbf{q}} \right|$ **by** $\left| F_{\mathbf{q}}^{(obs)} \right|$

$$\left| F_{\mathbf{q}} \right| e^{i\phi_{\mathbf{q}}} = \sum_j u_j e^{i\mathbf{q}.\mathbf{r}j}$$

Real Space



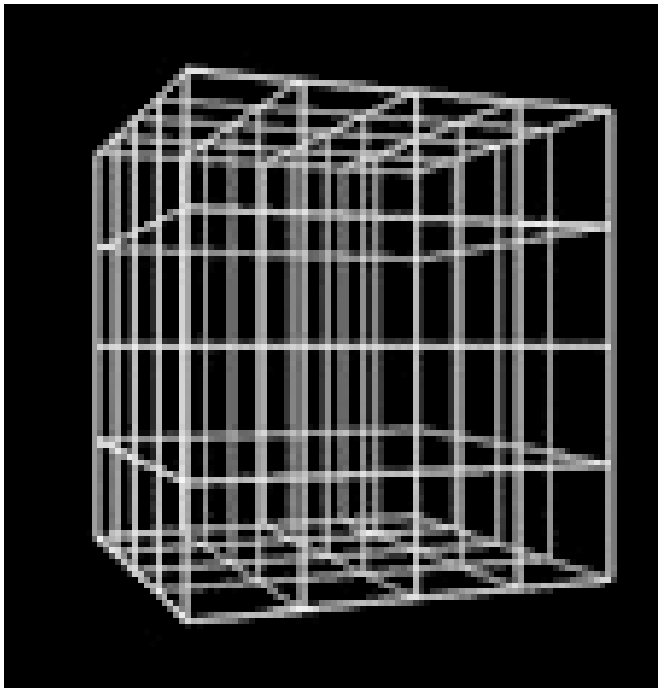
Constrain to e.g. expected object size

Provided $\{F_q\}$ is oversampled w.r.t. Nyquist criterion, at end of iterations, both sets of constraints satisfied. This tends to determine both the phases of the complex $\{F_q\}$, and the the real $\{u_j\}$

From 3D Diffaction Volume to 3D Electron Density

Evaluate 3D diffraction volume from

$$I(\mathbf{q}) = \sum_{lm} O_{lmm'} I_{lm'}(q) S_{lm'}(\hat{\mathbf{q}})$$



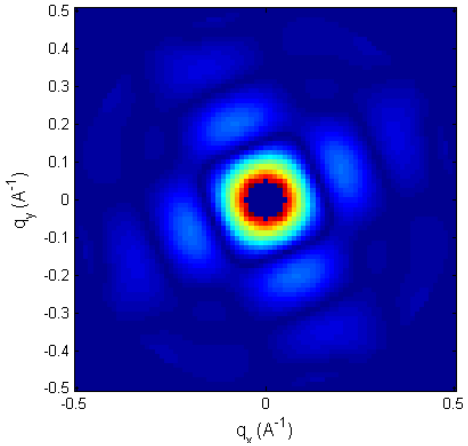
Here I_{lm} is found from the matrix square root of B, and $O_{lmm'}$ from the triple correlations, T, S is a real spherical harmonic. Evaluate $I(q_x,q_y,q_z)$ on a cubic grid. Amplitudes $|A(q_x,q_y,q_z)|=\sqrt{I(q_x,q_y,q_z)}$. The electron density can be found by a 3D inverse Fourier transform

$$\rho(\mathbf{r}_j) = \frac{1}{N} \sum_{\mathbf{q}} |A(\mathbf{q})| \exp(i\vartheta_{\mathbf{q}}) \exp(-i\mathbf{q}.\mathbf{r}_j)$$

The phases φ_q (and hence electron density of the molecule) are found by the iterative algorithm (described earlier) which alternately satisfies constraints in real and reciprocal space.

Reconstruction of DPs of K-channel protein from angular correlations

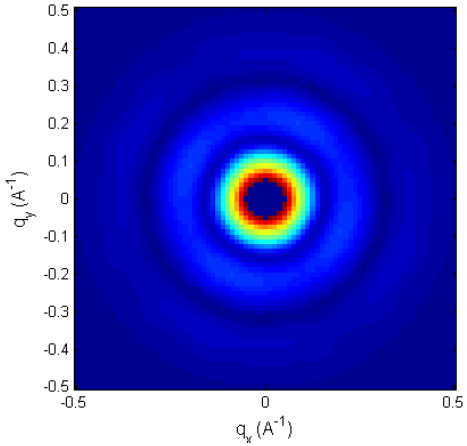
Model DP



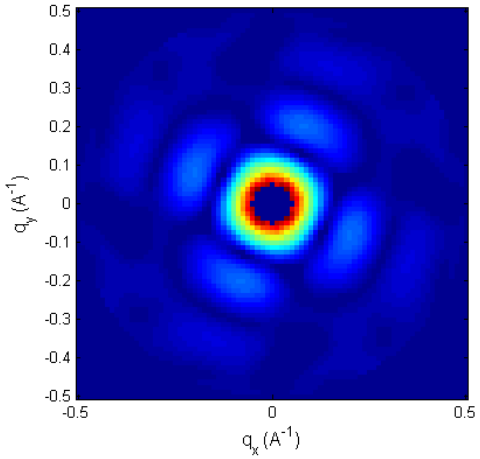
qmax=0.5 A^{-1}

Originals

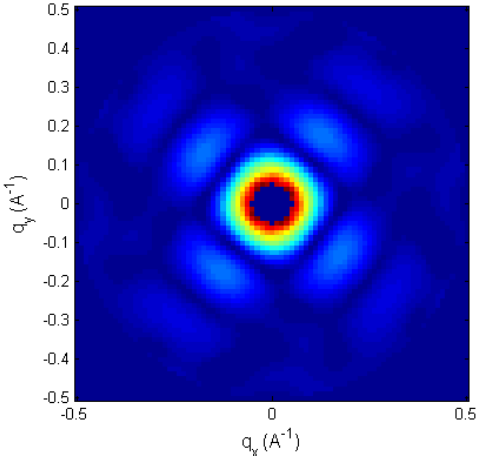
10 particles, random orientation



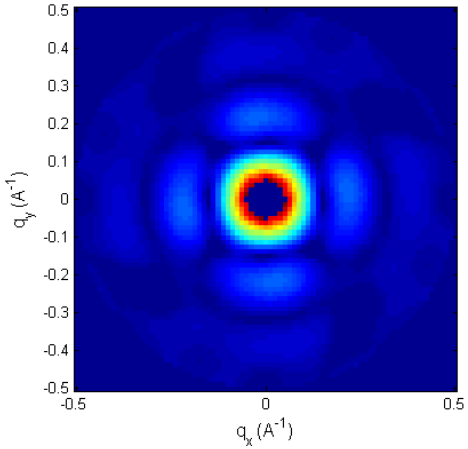
1 particle, 100 DPs



10 particles, 100 DPs



10 particles, 1000 DPs



Reconstructed from correlations

Alternative: Calculation of Autocorrelations (Z. Kam, 1982)

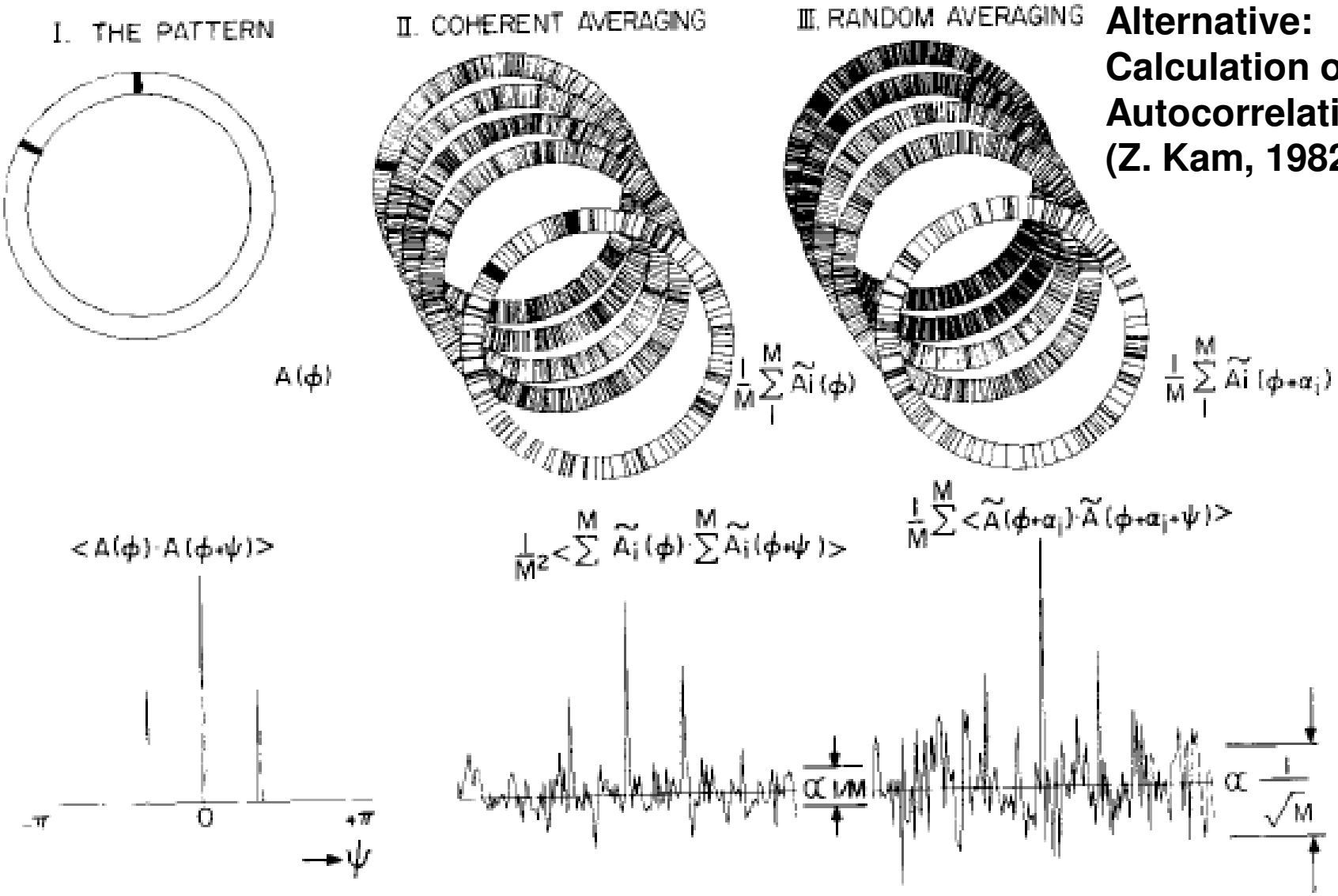
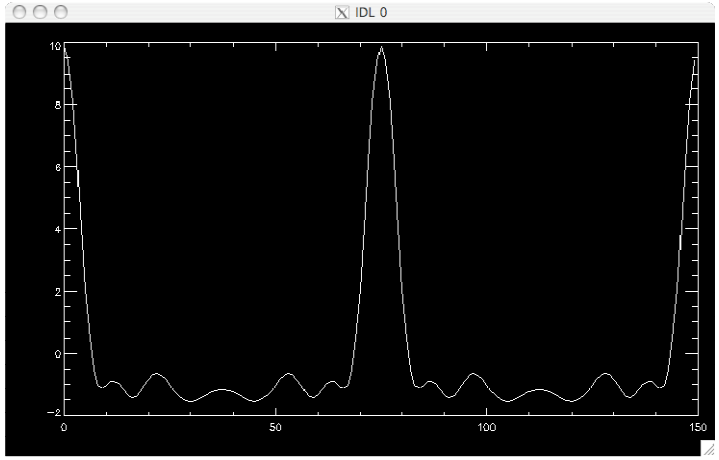
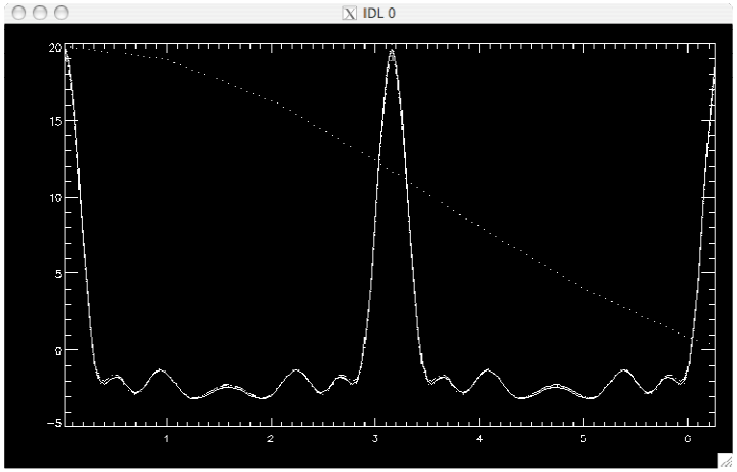


Fig. 1 A pictorial presentation of a coherent (synchroneous) enhancement of a pattern, as compared to enhancement of a randomly oriented pattern using correlation. (I) The ring pattern $A(\phi)$ and its autocorrelation $\langle A(\phi)A(\phi+\psi)\rangle$ (II) Aligned patterns with noise $A(\phi)$. By direct averaging the enhancement of the signal over the noise is proportional to $1/\sqrt{M}$. The autocorrelation of the averaged signal is enhanced over the noise as $1/M$. (III) Randomly oriented patterns. Direct averaging smears the signal but averaging of the auto-correlation enhances signal-to-noise as $1/\sqrt{M}$

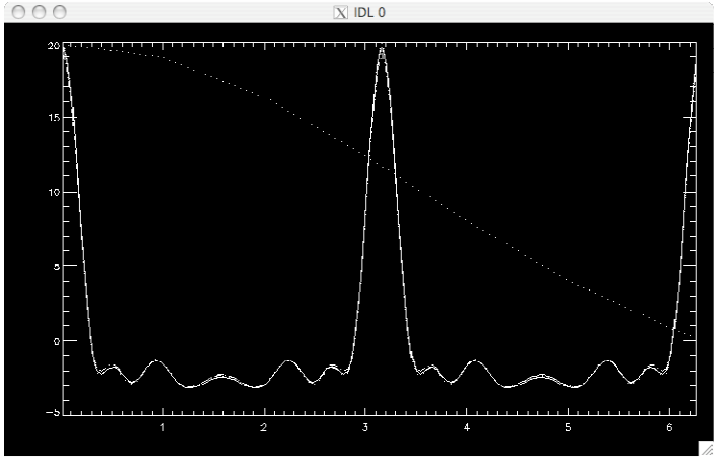
Simulated Pair Correlations



1 particle



2 particles, 1000 DPs



10 particles, 1000 DPs

Reconstructing the Single-Particle Diffraction Pattern

Circular Harmonic Expansion

$$I(q_x, q_y) = \sum_m I_m(q) \exp(im\phi)$$

where

$$q = \sqrt{q_x^2 + q_y^2} \qquad \phi = \tan^{-1}(q_y, q_x)$$

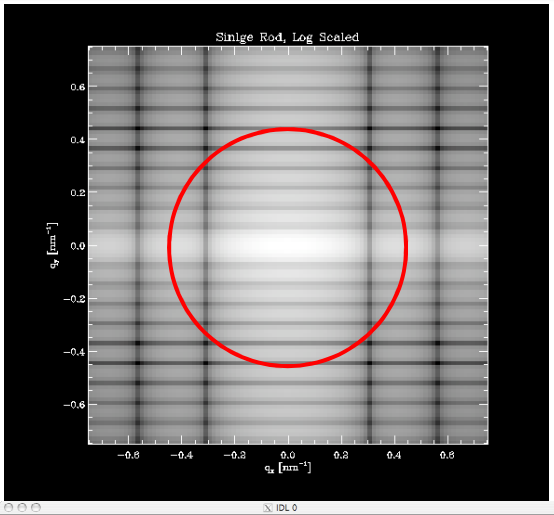
Can be done once the magnitudes of $I_m(q)$ are found from the pair correlations and their signs are found from the triple correlations

The reality of $I(\mathbf{q})$ ensures that $I_{-m}(q)=I_m(q)$. For a flat Ewald sphere, Friedel's rule, $I(-\mathbf{q})=I(\mathbf{q})$, will be satisfied if only even m's contribute.

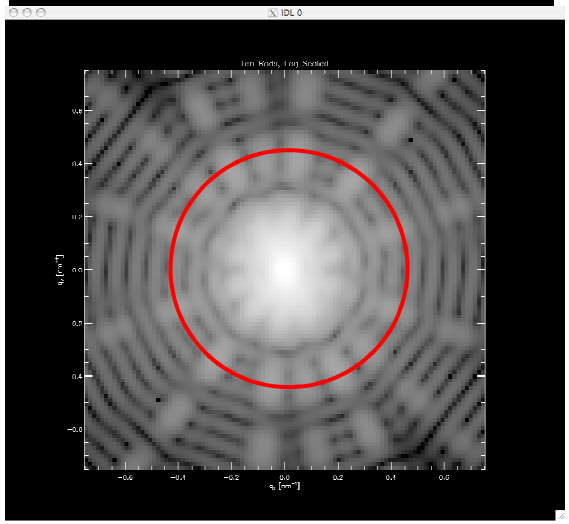
If the single particle diffraction pattern has a mirror line, can choose the $I_m(q)$ to be real (not necessary to assume this).

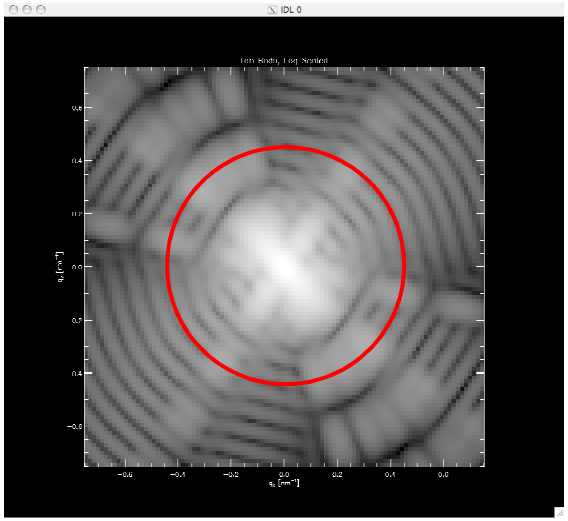
Saldin et al., New J. Phys., in press.

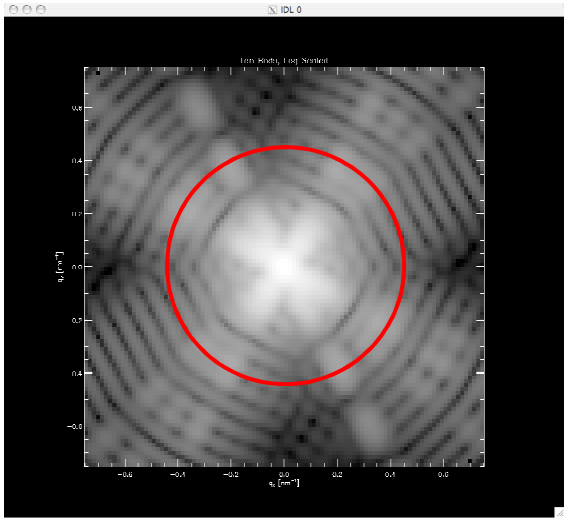
Key: Concentrate on Angular Correlations, Not Bare Intensities



Simulated single particle diffraction pattern







Sample 10 particle diffraction patterns, randomly oriented

Magnitude of Expansion Coeffs.

Pair Correlations (averaged over many short-pulse DPs)

$$C_2(q;q',\Delta\varphi_l) = \left\langle \frac{1}{N_\varphi}\sum_j \left\{I(q,\varphi_j) - I_{saxs}(q)\right\}\left\{I(q',\varphi_j+\Delta\varphi_l) - I_{saxs}(q')\right\}\right\rangle_t$$

$$= N_p \sum_{M\neq 0} I_M^*(q) I_M(q') \exp(iM\Delta\varphi_l)$$

FT of $C_2(q,q';\Delta\varphi_l)$

$$B_M(q,q') \equiv \frac{1}{N}\sum_{l=1}^{N} C_2(q,q;\Delta\varphi_l)\exp(-iM\Delta\varphi_l) = I_M(q) I_M^*(q')$$

Magnitude of expansion coefficients from the FT of the autocorrelations:

$$\left|I_M(q)\right| = \sqrt{B_M(q,q)}$$

The non-uniqueness of the square root is manifested by the unknown phases, which need to be determined by something which is sensitive to the phases.

Signs of Expansion Coeffs.

Triple Correlations

$$C_3(q,q;\Delta\phi) = \left\langle \frac{1}{N_\varphi} \sum_{j=1}^{N_\varphi} I(q,\varphi_j)^2 I(q,\varphi_j + \Delta\varphi_l) \right\rangle_t$$

FT of Triple Correlations

$$T_M(q,q) \equiv \frac{1}{N_\varphi} \sum_{l=1}^{N_\varphi} C_3(q,q;\Delta\phi)\exp(iM\Delta\phi)$$

$$= I_M(q) \sum_m I_{m-M}(q) I_m(q)$$

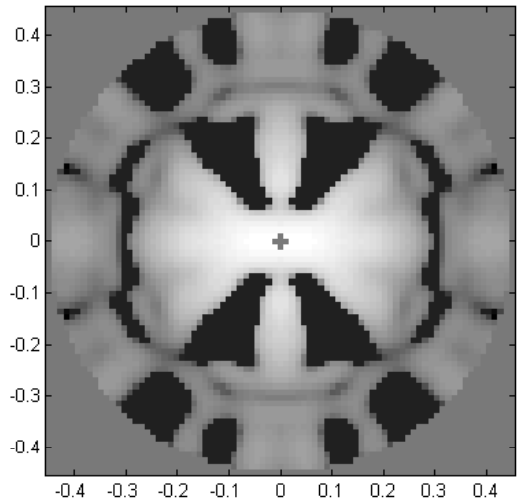
$$= \frac{B_M(q,q_1)}{|I_M(q_1)|e^{i\phi_M}} \sum_m \left\{ \frac{B_{m-M}(q,q_1)}{|I_{m-M}(q_1)|e^{i\phi_{M-m}}} \frac{B_m(q,q_1)}{|I_m(q_1)|e^{i\phi_m}} \right\}$$

The last step uses the FTs $B_M(q,q')$ of the *cross correlations* $C_2(q,q')$.
Needed to determine correct registry of the intensities on different rings q

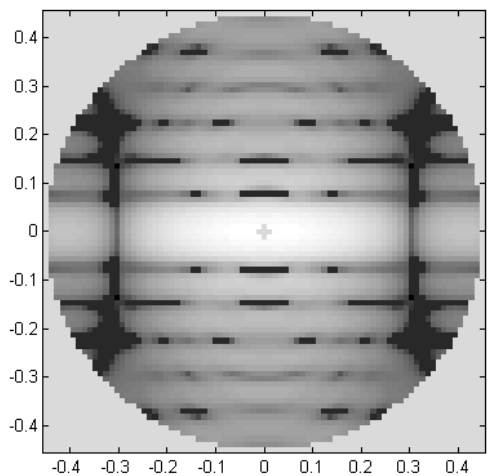
Unknown phases determined by comparison with $T_M^{(\exp)}(q,q)$

Can be done by an optimization routine

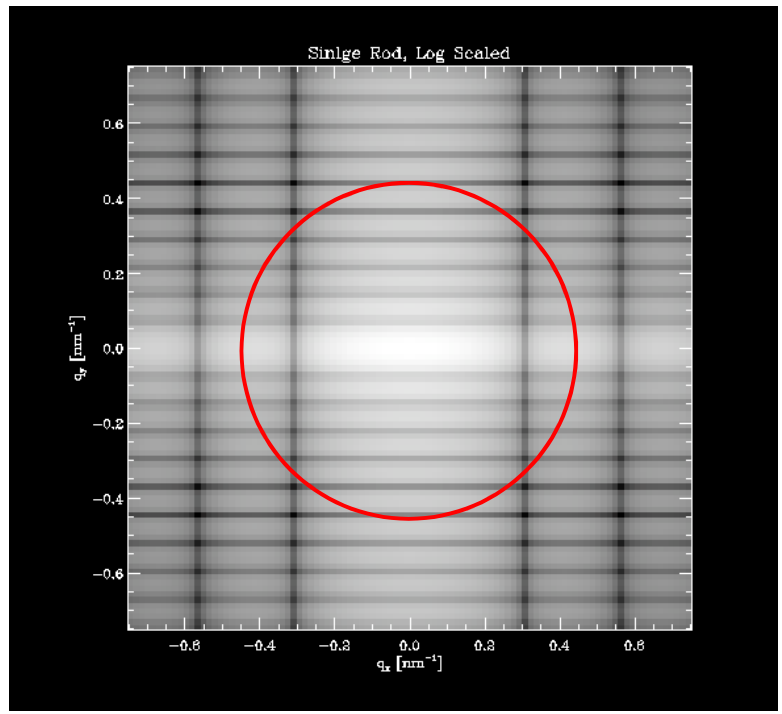
Triple Correlations for Phase Determination



Random phases of $I_m(q)$



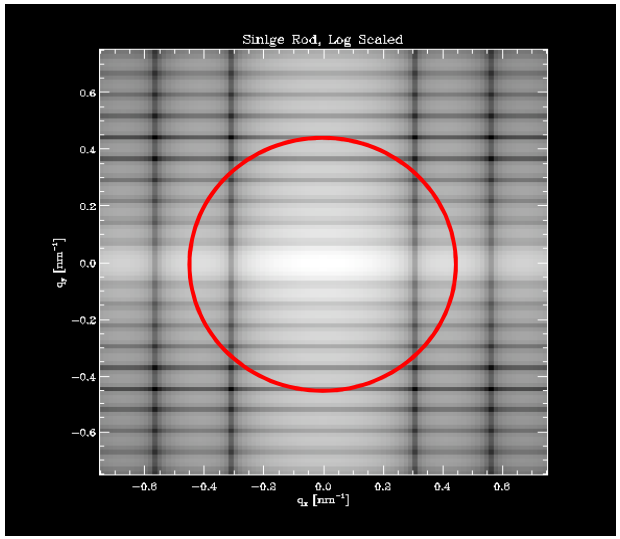
Correct phases from $T_M(q,q)$

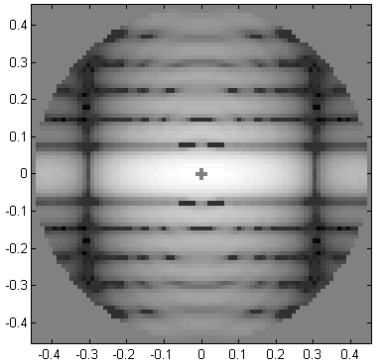


Diffraction Pattern from Model

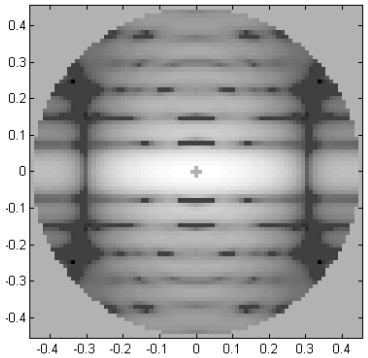
Directly Calculated & Reconstructed DPs

Diffraction Pattern from Model

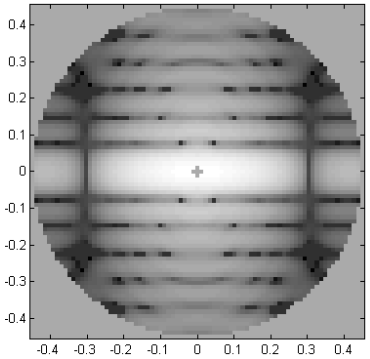




From correlations of single particle DP

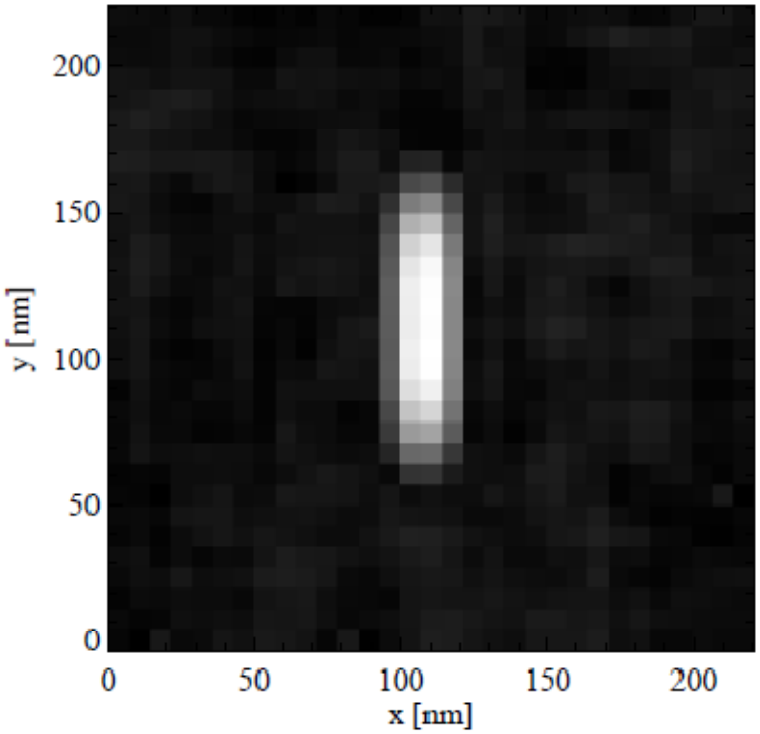


From 100 DPs 10 Particles per DP



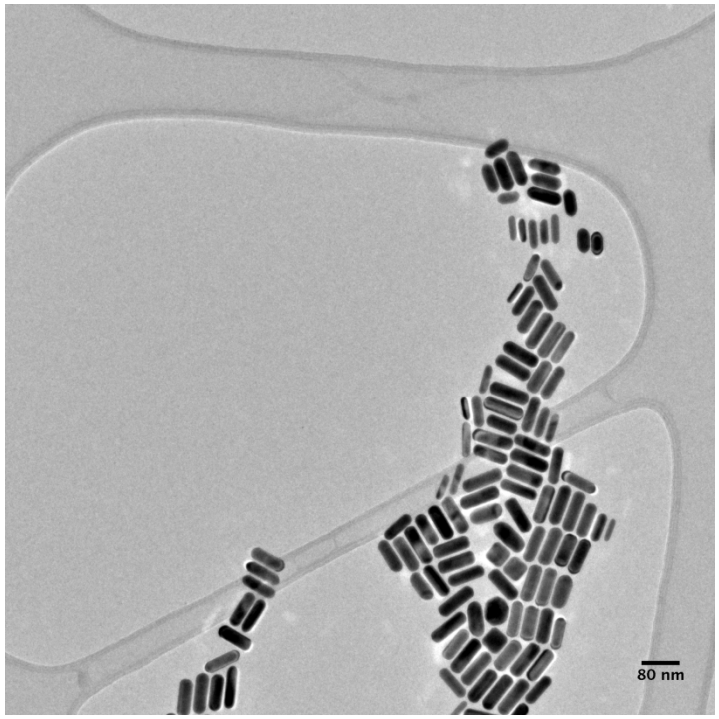
From 1000 DPs 100 particles per DP

Reconstructed Image



From the single particle diffraction pattern reconstructed from average correlations from 100 DPs of each of which contained 10 particles in different random orientations

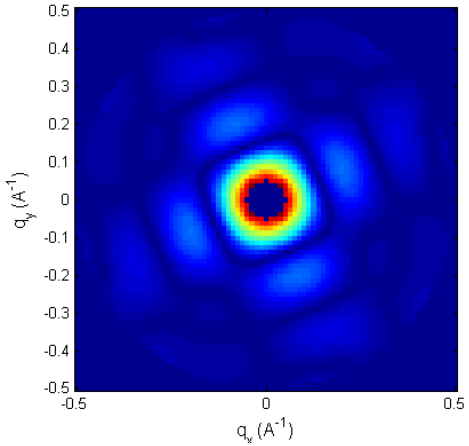
Experimental Test



EM image of metal rods which used for the test with soft x-rays. In the experiment they were not clumped as shown, but randomly oriented rods with random interparticle distances.

Reconstruction of DPs of K-channel protein from angular correlations

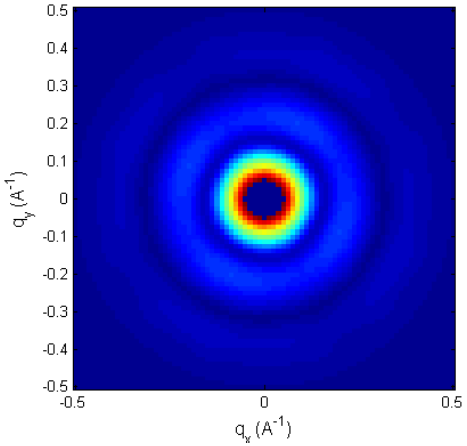
Model DP



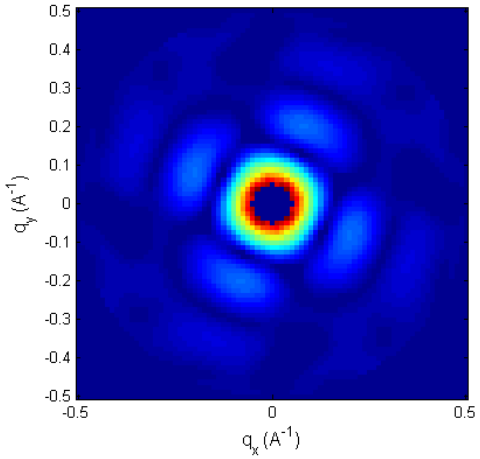
qmax=0.5 A^{-1}

Originals

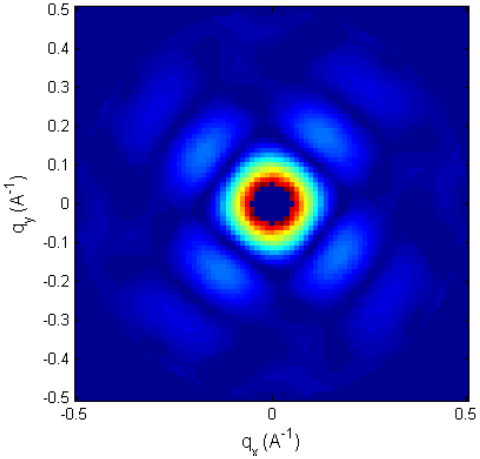
10 particles, random orientation



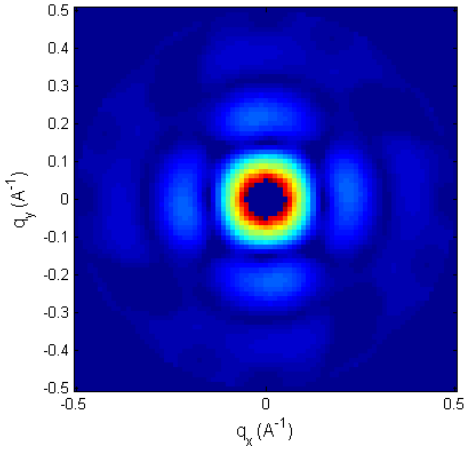
1 particle, 100 DPs



10 particles, 100 DPs



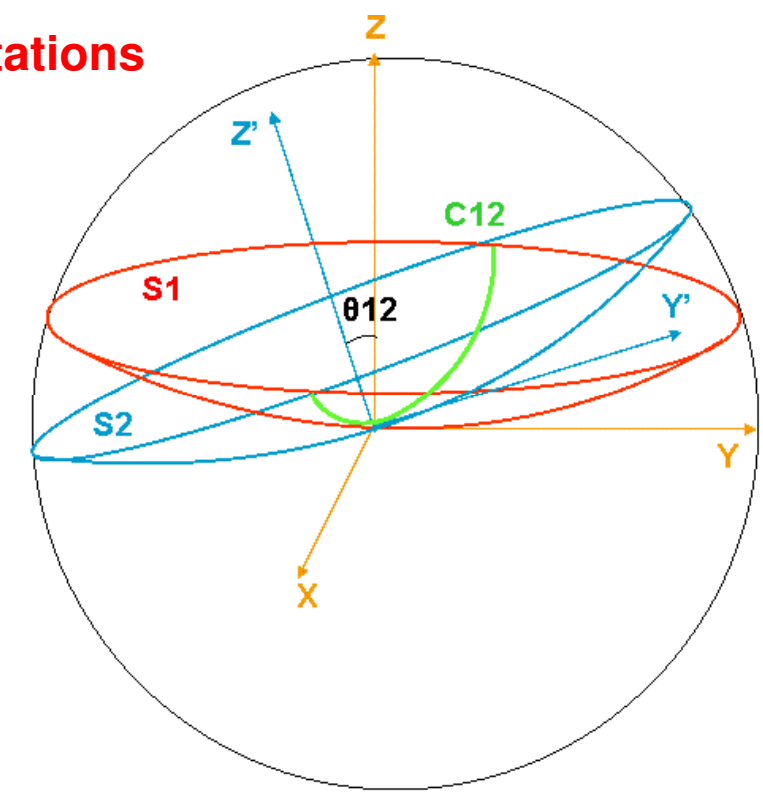
10 particles, 1000 DPs



Reconstructed from correlations

Measure angular correlations => No need to determine the orientations of the DPs

Exploit the only symmetry of the problem:
SO(3) symmetry of random molecular orientations
J. Phys.: Condens. Matter 21, 134014 (2009)



3D intensity in (X,Y,Z) coordinate system

$$I_{\mathbf{q}}^{(0)} = \sum_{lm} I_{lm}(q) S_{lm}(\hat{\mathbf{q}})$$

2D intensity on red Ewald sphere S1

$$I_{q\phi}^{(0)} = \sum_{l}\sum_{m} I_{lm}(q) S_{lm}\left(\pi/2 - \sin^{-1}(q/2\kappa), \phi\right)$$

Takes account of Ewald sphere curvature

2D intensity on blue Ewald sphere S2 (rotated by Euler angles Φ,θ,Ψ)

$$I_{q\phi}^{(p)} = \sum_{l}\sum_{mm'} \Delta_{lmm'}^{(p)} I_{lm}(q) S_{lm'}\left(\pi/2 - \sin^{-1}(q/2\kappa), \phi\right)$$

where the index p represents a DP specified by its orientation (Φ,θ,Ψ).

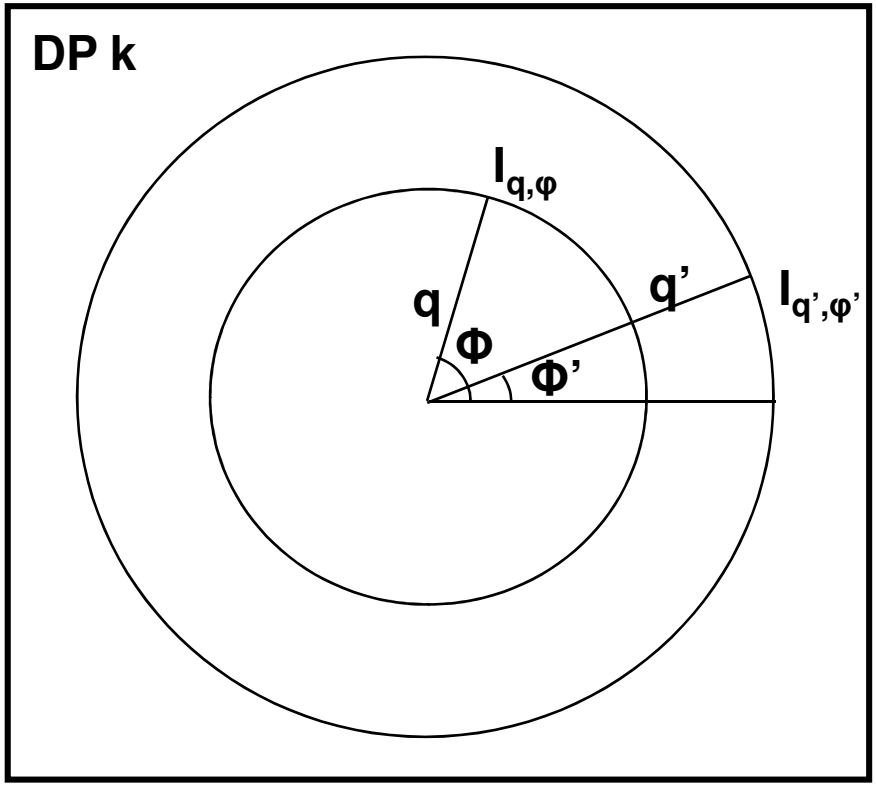
Input to algorithm

Angular correlations

$$J_{qq';\phi\phi'} = \frac{1}{N}\sum_{p}\left\{I_{q,\phi}^{(p)} - I_{SAXS}(q)\right\}\left\{I_{q',\phi'}^{(p)} - I_{SAXS}(q')\right\}$$

between two pixels on each DP p, but summed over all N DPs.

Note neither the magnitudes nor number of such quantities grows with the number of DPs measured. The correlations just becomemore accurate as N increases.



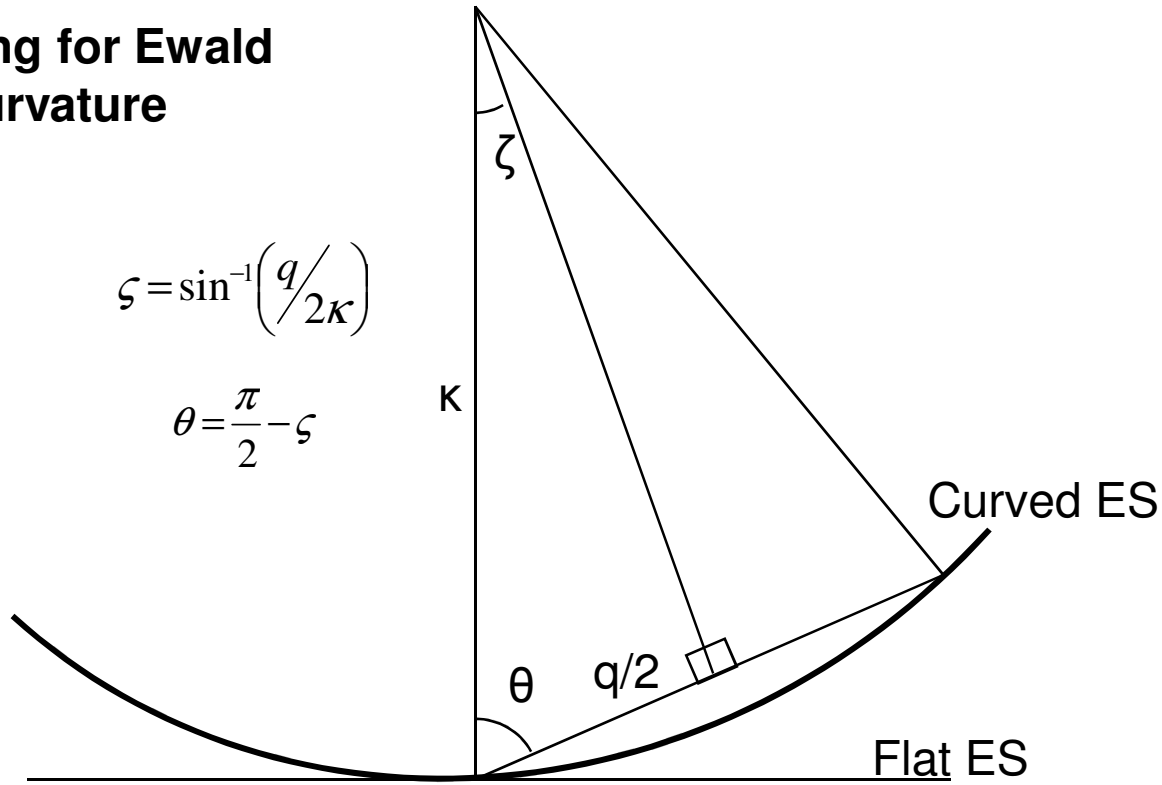
If the scattering angle is 2ζ, q=2κsinζ

Intensity in Lab Frame

Accounting for Ewald sphere curvature

$$\varsigma = \sin^{-1}\left(q/2\kappa \right)$$

$$\theta = \frac{\pi}{2} - \varsigma$$



$$I_{q\phi}^{(p)} = \sum_{l} \sum_{mm'} \Delta_{lmm'}^{(p)} I_{lm}(q) S_{lm'}\left(\pi/2 - \sin^{-1}(q/2\kappa), \phi\right)$$

Information from Angular Correlations

$$J_{qq';\phi\phi'} = \frac{1}{N}\sum_p \left\{I_{q\phi}^{(p)} - I_{SAXS}(q)\right\}\left\{I_{q'\phi'}^{(p)} - I_{SAXS}(q')\right\}$$

$$= \frac{1}{N}\sum_p \sum_{l\neq 0}\sum_{m'm} \Delta_{lm'm}^{(p)} I_{lm}(q) S_{lm'}[\theta(q),\phi] \sum_{l'\neq 0}\sum_{m'''m''} \Delta_{l'm'''m''}^{(p)} I_{lm''}(q') S_{l'm'''}[\theta'(q'),\phi']$$

$$\theta(q) = \frac{\pi}{2} - \sin^{-1}(q/2\kappa) \qquad \theta'(q') = \frac{\pi}{2} - \sin^{-1}(q'/2\kappa)$$

$$\frac{1}{N}\sum_p \Delta_{lm'm}^{(p)} \Delta_{l'm'''m''}^{(p)} = \frac{1}{(2l+1)}\delta_{ll'}\delta_{m'm'''}\delta_{mm''}$$

Great orthogonality theorem of representations of SO(3) group

$$J_{qq';\phi\phi'} = \sum_{l\neq 0} F_{qq';\phi\phi';l} B_{qq';l}$$

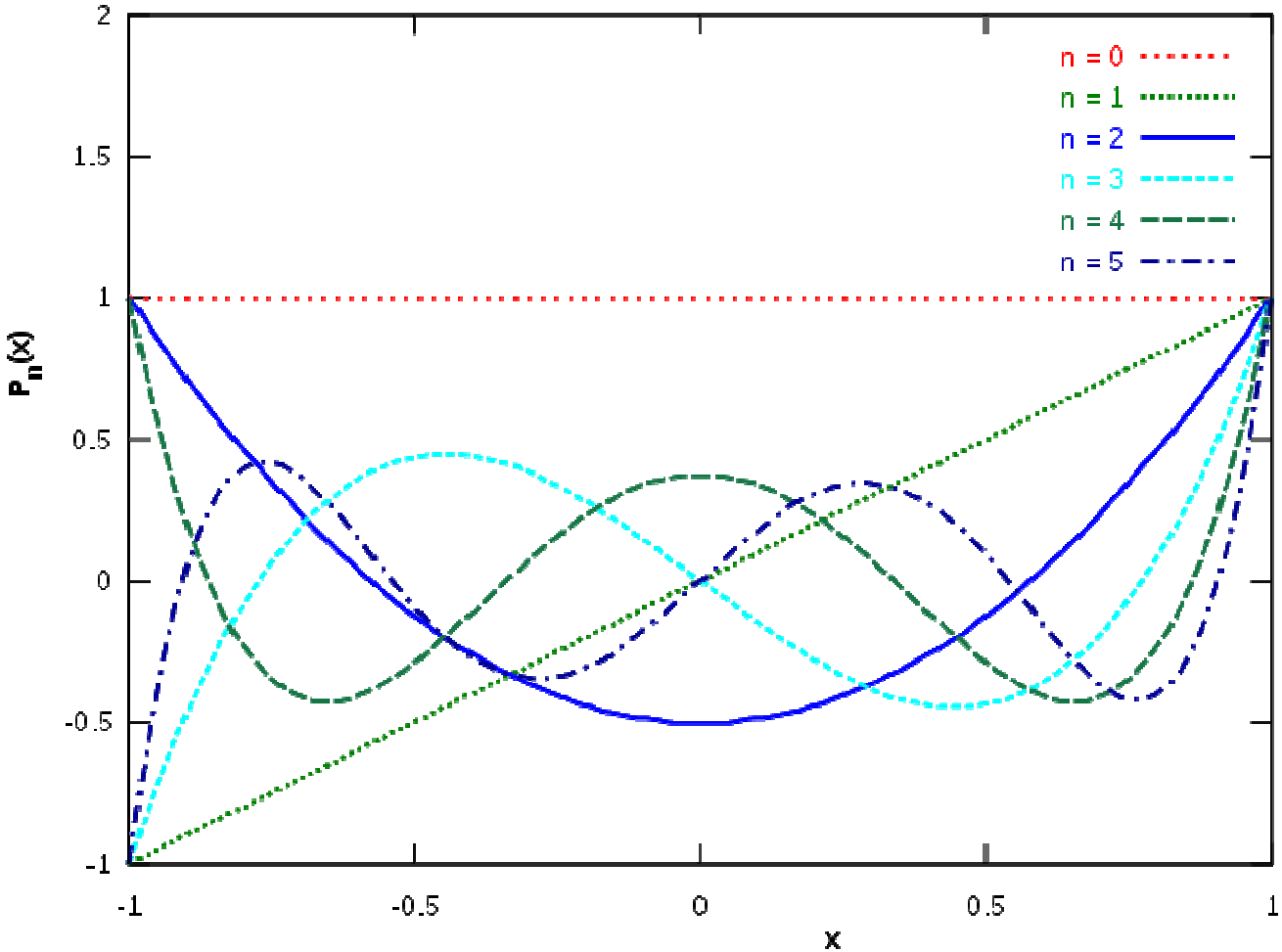
$$F_{qq';\phi\phi';l} = \frac{1}{4\pi} P_l[\cos\theta'']$$

θ'' **is the angle between (θ,φ) and (θ',φ')**

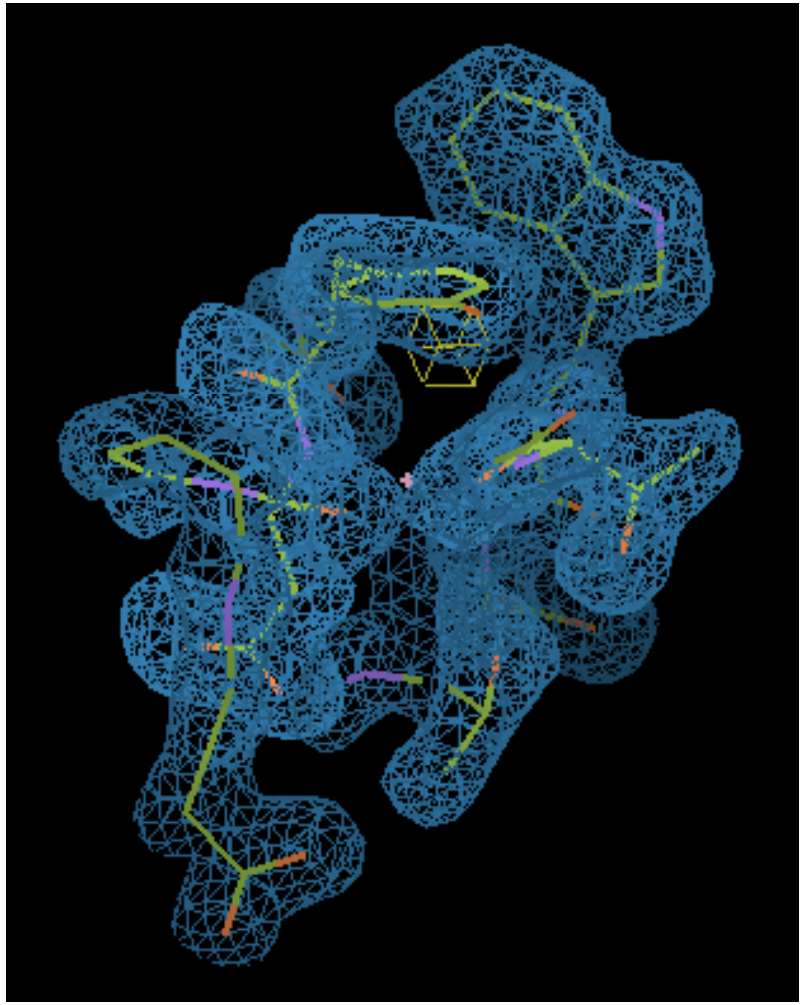
$$B_{qq';l} = \sum_m I_{lm}(q) I_{lm}(q')$$

This is an orientation-independent quantity characteristic of the "diffraction volume" of an individual particle

Legendre Polynomials



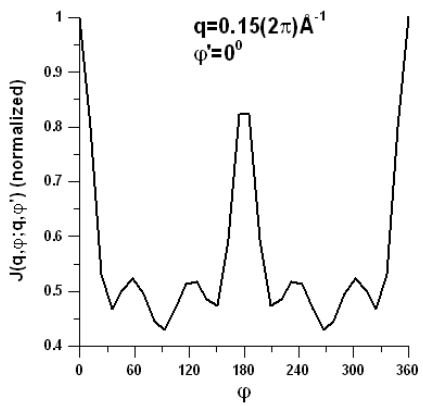
Chignolin (world's smallest protein)
10 residues - no symmetry

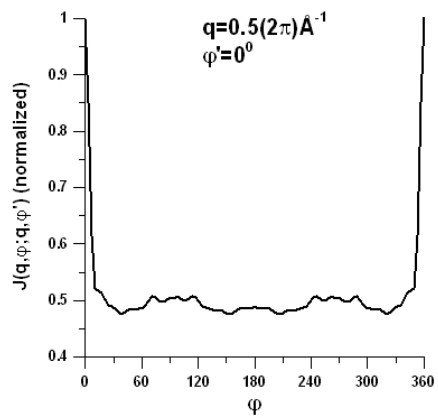


What do the corrrelations look like?

$$J_{qq';\phi\phi'} = \frac{1}{N}\sum_k I_{q\phi}^{(k)} I_{q'\phi'}^{(k)}$$

Put q=q', fix Φ', vary Φ





These plots were calculated from simulations of diffraction patterns from a small protein (chignolin) of no symmetry and random orientations. Yet they are remarkably simple and symmetric. They are consistent with the theoretical prediction $J_{qq';\phi\phi'} = \sum_l B_{qq';l} P_l[\cos\theta'']$ where

$$\cos\theta'' = \cos\theta(q)\cos\theta'(q') + \sin\theta(q)\sin\theta'(q')\cos(\phi-\phi') \qquad \theta(q) = \frac{\pi}{2} - \sin^{-1}(q/2\kappa)$$

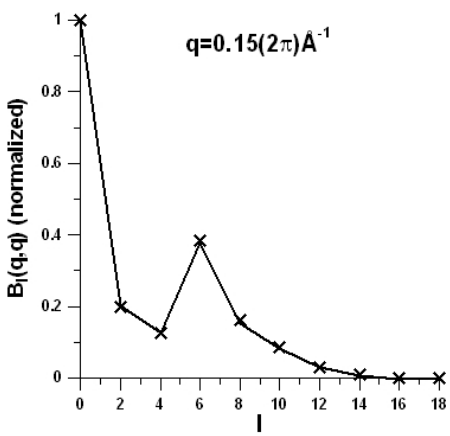
Extraction of Structural Information from Angular Correlations

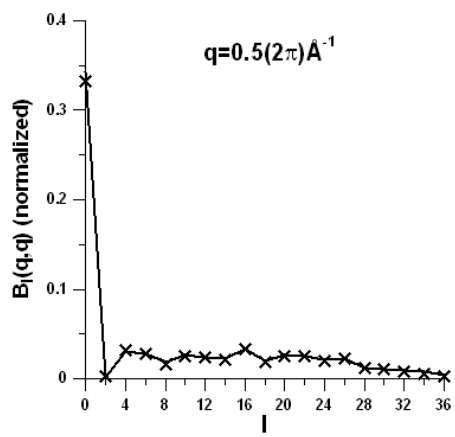
In the equation $J_{qq';\phi\phi'} = \sum_l P_l[\cos\theta'']B_{qq';l}$

The LHS (the angular correlations) may be extracted from the experimental data. On the RHS $P_l(\cos\theta'')$ is a known function, Therefore the coefficients $B_{qq';l}$ may be extracted by e.g. matrix inversion. The structural information resides in these coefficients, since

$$B_{qq';l} = \sum_m I_{lm}(q)I_{lm}(q')$$

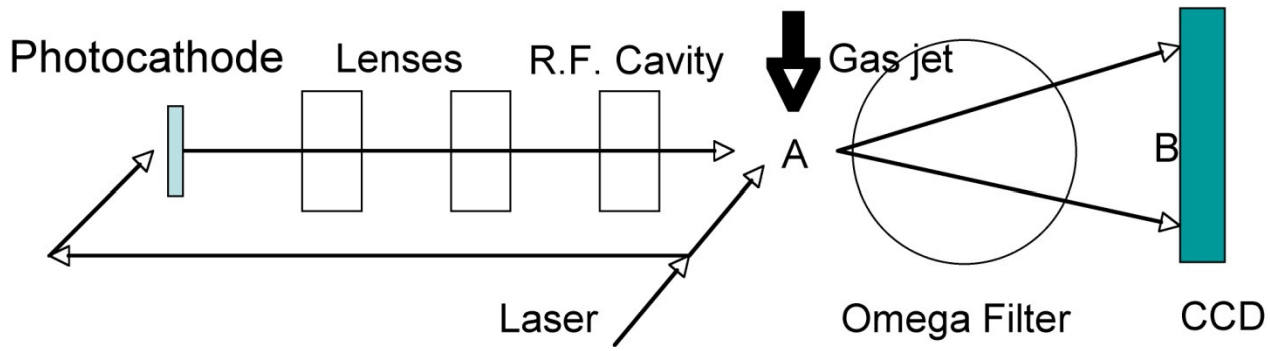
What do these look like (for q=q')?





Thus, the B's can be extracted accurately from experimental data. How can the l's be extracted from the B's?

Time-Evolution of a Molecular Structure
(Proposed Experiment: J. C. H. Spence)



Schematic arrangement for gas electron diffraction. At B a fast-readout CCD operating at 20 HZ is sychronized with the photocathode. A free-expansion nozzle at A injects gas at low temperature.

Finding the Change in the Scattered Intensities

Imagine a pump-probe experiment in which the DPs of individual molecules are measured before application of a laser excitation ("dark structure") and another DP of the same molecule is measured shortly after excitation (pump-probe Experiment). If the excited structure is a small perturbation of a known dark structure, we may take the variation of

$$B_{qq';l} = \sum_m I_{lm}(q) I_{lm}(q')$$

to find

$$\delta B_{qq';l} = \sum_m \{ I_{lm}(q) \delta I_{lm}(q') + \delta I_{lm}(q) I_{lm}(q') \}$$

LHS is measurable in an experiment. $I_{lm}(q)$ is Known. Thus $\delta I_{lm}(q)$ may be found by linear algebra. Then $\delta I(q)$ may be constructed from

$$\delta I(\mathbf{q}) = \sum_{lm} \delta I_{lm}(q') Y_{lm}(\hat{\mathbf{q}})$$

Change in the Electron Density

Scattered intensity $$I(\mathbf{q}) = A^*(\mathbf{q})A(\mathbf{q})$$

Intensity change $$\delta I(\mathbf{q}) = A^*(\mathbf{q})\delta A(\mathbf{q}) + \delta A^*(\mathbf{q})A(\mathbf{q})$$

Amplitude change $$\delta A(\mathbf{q}) = \sum_j \delta\rho(\mathbf{r}_j)\exp(i\mathbf{q}.\mathbf{r}_j)$$

$$\delta I(\mathbf{q}) = \sum_j \delta\rho(\mathbf{r}_j)\left[A^*(\mathbf{q})\exp(i\mathbf{q}.\mathbf{r}_j) + c.c.\right]$$

Once the LHS known from experiment, since A(q) is the structure factor of the known "dark structure", the only unknown is $\delta\rho(r_j)$, the change in the structure due to the excitation. This may be found by solving the above linear equation.

General Solution

For each value of l, the Eq: $B_{qq';l} = \sum_{lm} I_{lm}(q) I_{lm}(q')$

may be rewritten as $B_{qq'} = \sum_{m} I_{qm} I_{mq'}$ i.e. as the matrix Eq. $\mathbf{B} = \mathbf{I}^T \mathbf{I}$

which also may be written : $\mathbf{B} = \mathbf{C}^T [\lambda]_D C = \left(C^T \left[\sqrt{\lambda}\right]_D\right)\left(\left[\sqrt{\lambda}\right]_D C\right)$

where C is the matrix of the eigenvectors of the Hermitian matrix B and $[\lambda]_D$ the diagonal matrix of the real eigenvalues. Tempting to identify I with $([\sqrt{\lambda}]_D C)$, the "matrix square root". However, note that B may also be written;

$$\mathbf{B} = \mathbf{C}^T [\sqrt{\lambda}]_D \mathbf{O}^T \mathbf{O} [\sqrt{\lambda}]_D \mathbf{C}$$

Thus, in general: $\mathbf{I} = \mathbf{O}[\sqrt{\lambda}]_D \mathbf{C}$

Hence $I(\mathbf{q}) = \sum_{lm} O_{qm}^{(l)} I_{lm}^{(0)}(q) S_{lm}(\hat{\mathbf{q}}) = \sum_{lm} O_{qm}^{(l)} \sqrt{\lambda_m} C_{mm'} S_{lm'}(\hat{\mathbf{q}})$

Thus I(q) may be determined to within a set of orthogonal matrices $O^{(l)}$. How may these be found?

One solution:
consider also triple correlations

Definition:

$$T(q\phi;q'\phi') = \frac{1}{N}\sum_k \left\{I_{q\phi}^{(k)} - I_{SAXS}(q)\right\}^2 \left\{I_{q'\phi'}^{(k)} - I_{SAXS}(q')\right\}$$

$$I_{q\phi}^{(k)} = \sum_{l_1 m_1 m'} D_{lmm'}^{(k)}(\alpha,\beta,\gamma) I_{lm'}(q) Y_{lm}[\theta(q),\phi]$$

$$I_{q'\phi'}^{(k)} = \sum_{l_1 m_1 m'} D_{lmm'}^{(k)}(\alpha,\beta,\gamma) I_{lm'}(q') Y_{lm}[\theta'(q'),\phi']$$

Sum over k by assuming the diffraction patterns arise from all possible orientations in SO(3).

Triple correlations

$$T_{qq';\phi\phi'} = \sum_{l\neq 0} F_{\phi\phi';l} T_{qq';l}$$

Z. Kam, J. thoer. Biol. 82, 151 (1980)

where $$F_{\phi\phi';l} = \frac{1}{4\pi} P_l\left[\cos\theta''\right]$$

and $$T_{qq';l} \equiv T_l(q,q') = I_l(q,m)^* V_l(m,q')$$

$$V_l(m,q') = W(l_1m_1, l_2m_2, l-m) I_{l_1}(q',m_1) I_{l_2}(q',m_2)$$

Conventions:

(1) Superscript T denotes transpose

(2) Sum over repeated indices unless they appear on both sides of the equation.

$$W(l_1m_1, l_2m_2, l-m) = (-1)^m \begin{pmatrix} l_1 l_2 l \\ 000 \end{pmatrix} \begin{pmatrix} l_1 l_2 l \\ m_1 m_2 - m \end{pmatrix} \sqrt{\frac{(2l_1+1)(2l_2+1)(2l+1)}{4\pi}}$$

Summary

- One of the major justifications for the development of a billion dollar x-ray free electron laser (XFEL) is the promise of the determination of the structures of individual protein molecules – obviating the need for crystallization
- Aim to determine the high-resolution structure of a protein molecule from scattering by molecules in a water droplet injected into the x-ray beam
- Our proposal is to determine the 3D diffraction volume from the ensemble of diffraction data, without determining the orientations of the individual diffraction patterns in the reciprocal space of the molecule
- Needs reasonable structural homogeneity of different molecules (as do SAXS, NMR etc.)
- Method would work even if there were multiple particles in each droplet
- Would also work if x-rays focussed on a small number, e.g. 10, of molecules dissolved in a liquid.