“Non-Bragg” Scattering from Protein Crystals

George N. Phillips Jr.

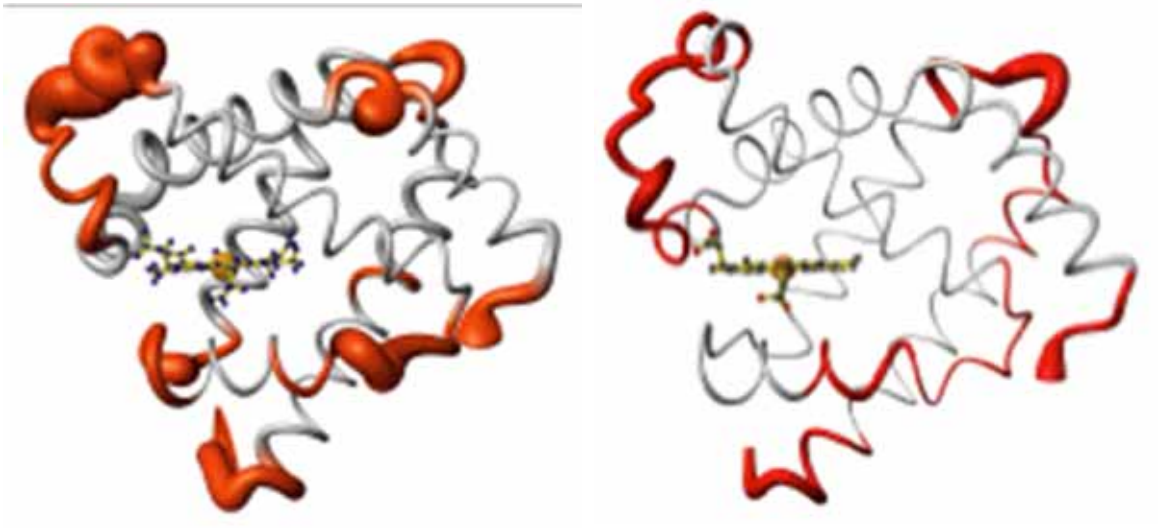
Center for Eukaryotic Structural Genomics
Departments of Biochemistry and Computer Sciences
University of Wisconsin, Madison

Introduction

- Crystallography (usually) confuses the space and time averages.
- Dynamic behavior remains--There IS temperature dependence, both kT-ish and energy landscapes more shallow
- The crystal lattice constrains the ‘dynamics’ to varying degrees
- Even when cooled, an ensemble of structures remains

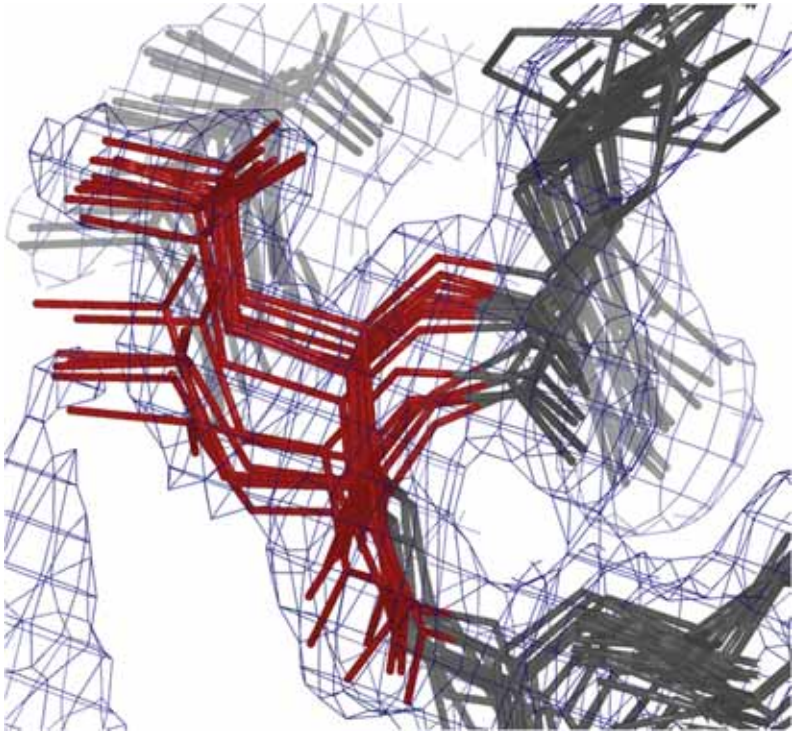
Molecules don’t do *exactly* the same thing in every unit cell--even having symmetry is an approximation

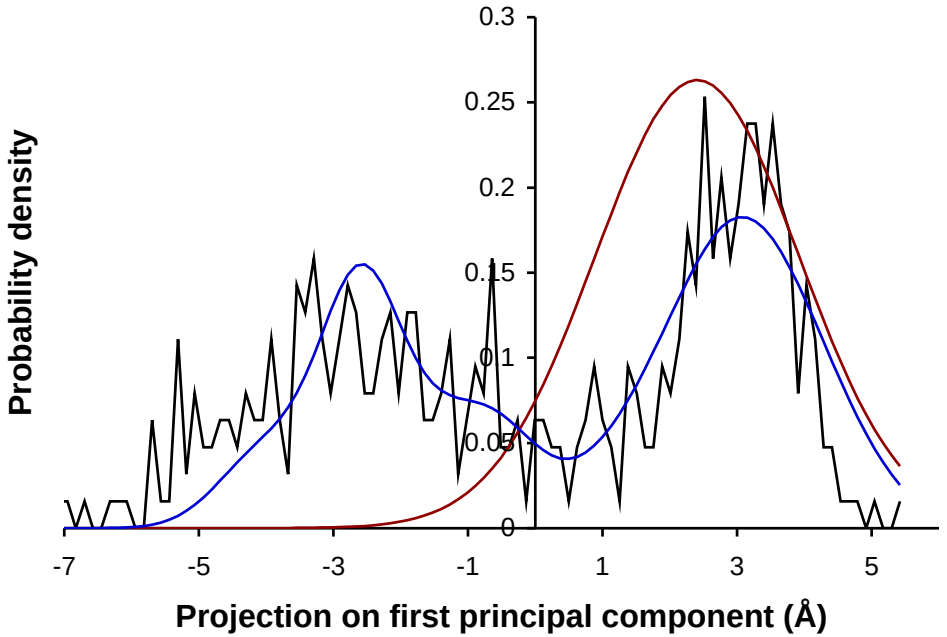
NMR and Crystallography: comparison of backbone dynamics



Main chain variations from NMR ensemble and various crystal forms of myoglobin.

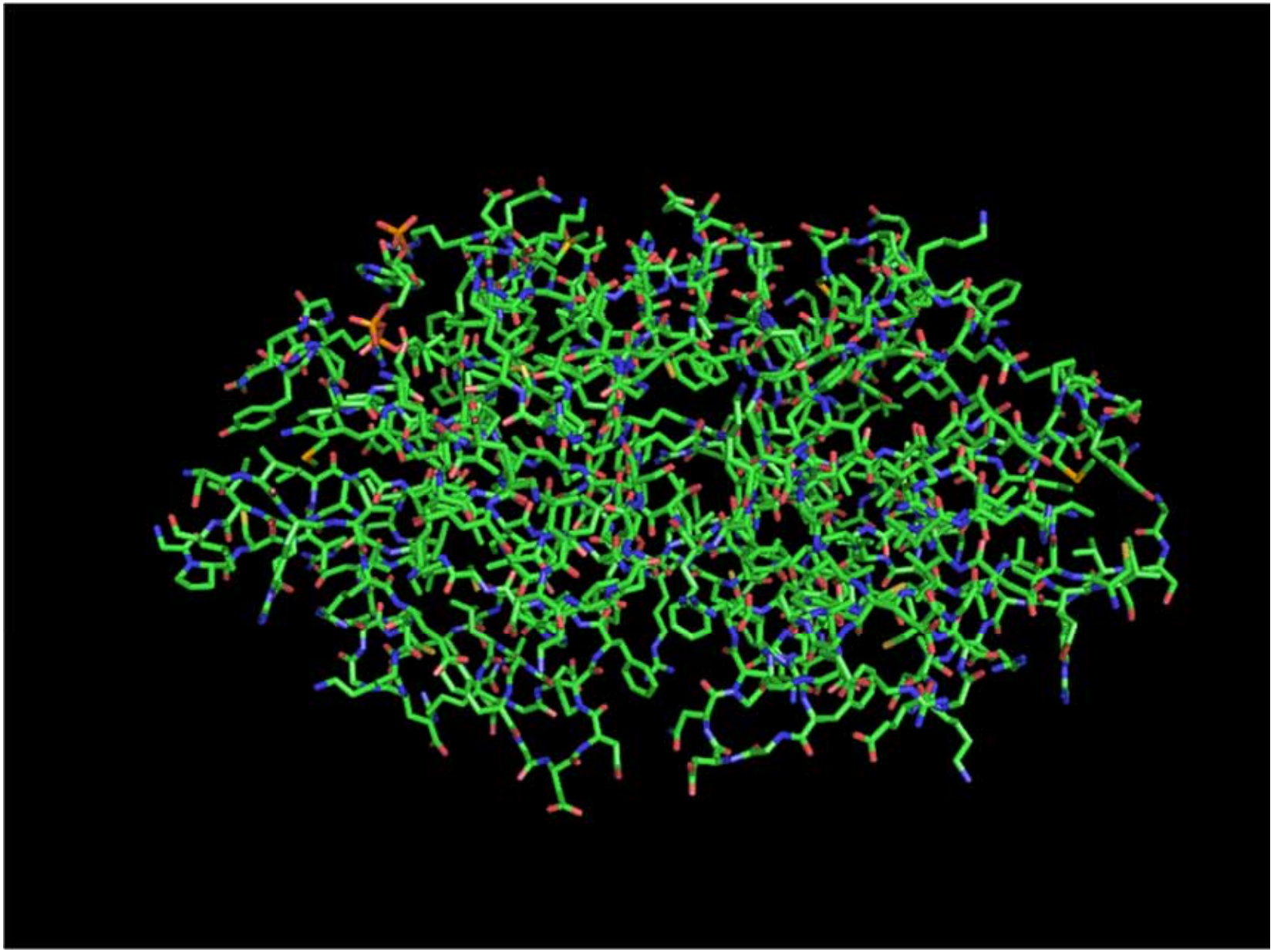
Kondrashov, Zhang, Aranda, Stec, and Phillips *Proteins* 2008





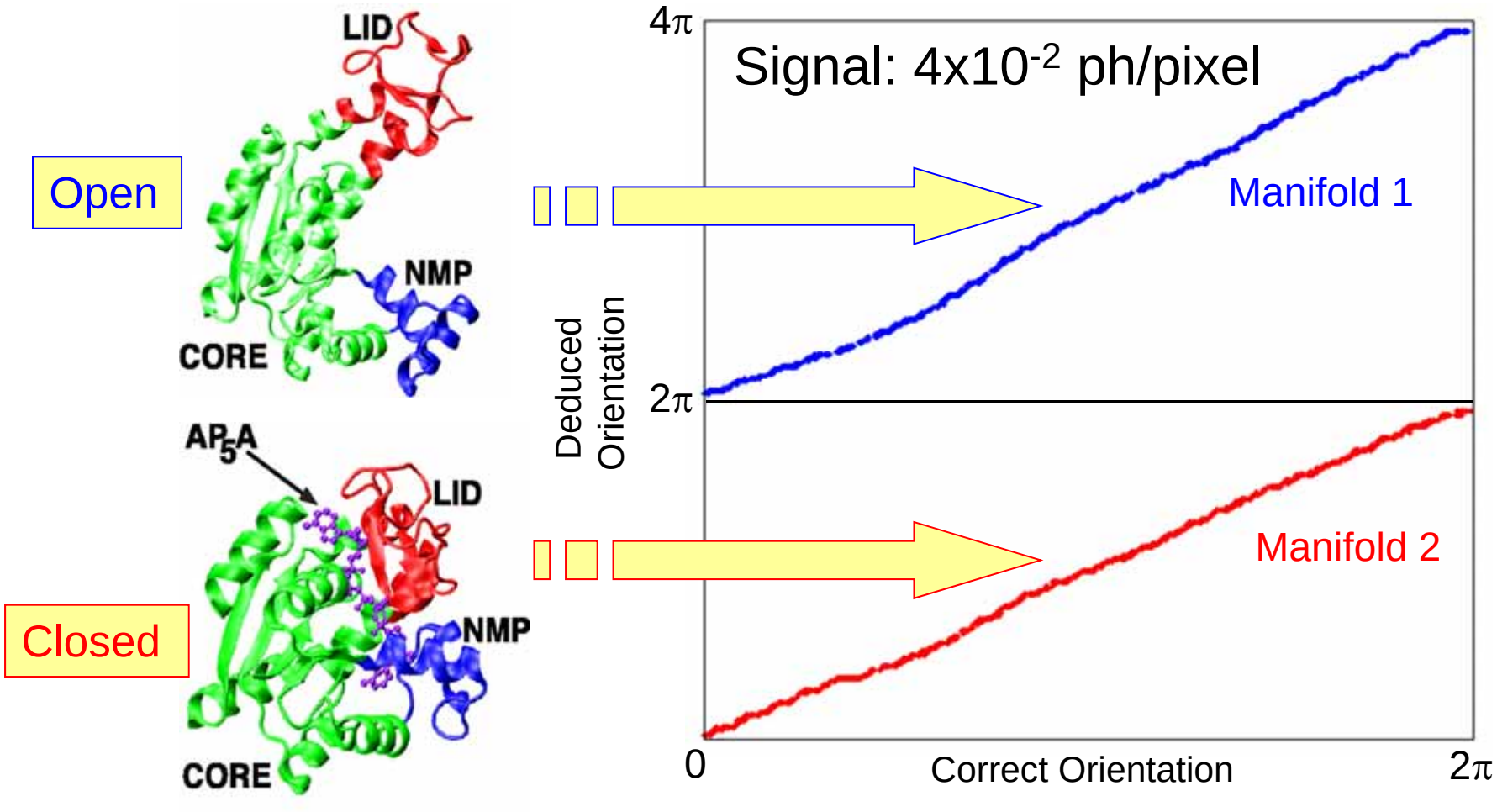
Levin, Kondrashov, Wesenberg, Phillips, *Structure* (2007)

Entire Dimeric Protein



Ensemble of Two Conformations Deconvolved from Simulated FELS data

Adenylate Kinase: Open and Closed



Schwander, Fung, Phillips, Ourmazd, NJP (2010)

What about Calculation/Modeling of Motions

- All-atom MD (slow)
- New coarse-grained models work surprisingly well!

Riccardi, Cui, Phillips Biophys J 2009

Define Diffuse Scatter?

- Water diffraction? Yes
- Capilllary diffraction? No, subtract it
- Air Scatter? No, subtract it
- Protein variational scattering? Yes

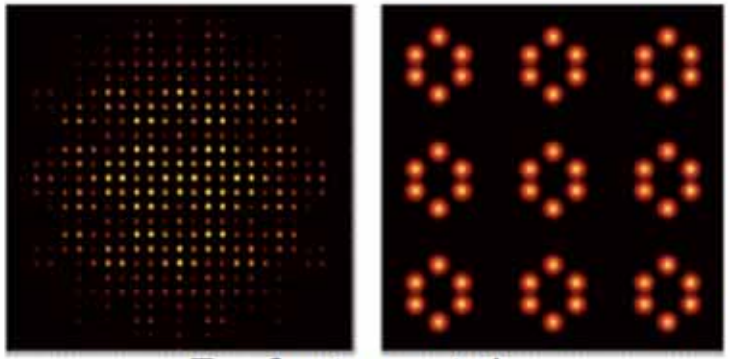
$$e^{-BS^2} versus \left(1 - e^{-BS^2}\right)$$

Diffuse X-ray Scattering

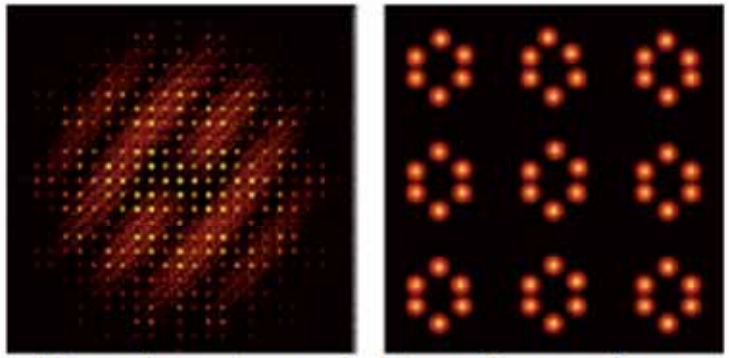
- Has information about displacements from the average structure
- Illustrates intrinsic mechanical properties of the macromolecule
- Couples with lattice motions
- Any one snapshot is ‘coherent’, averages over time or crystals are not.

Clarage and Phillips, *Methods Enz.* 1997

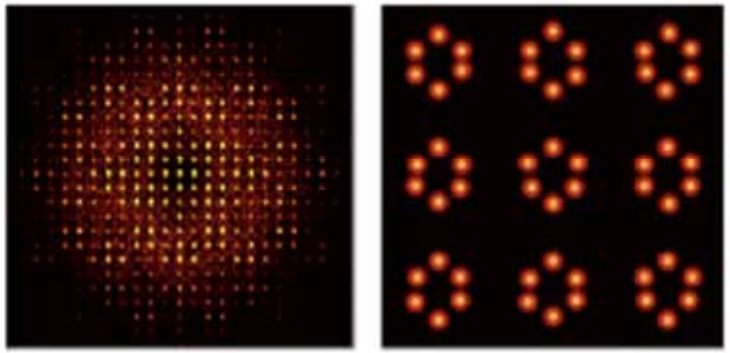
Adapted from Clarage and Phillips, Methods Enz, 277 (1996).



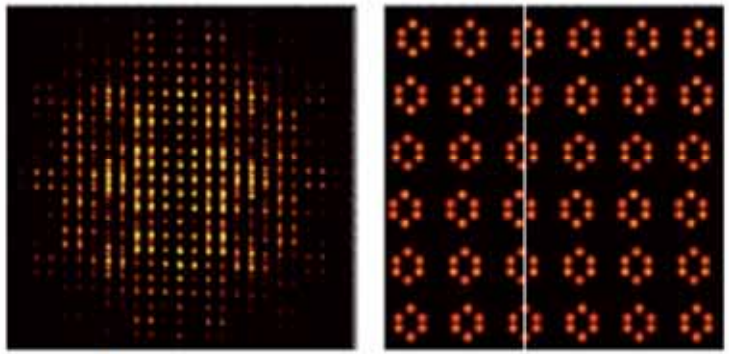
Perfect crystal



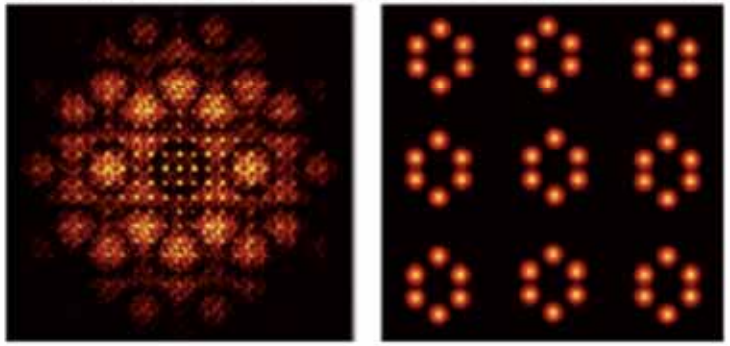
Random top-right pair motions



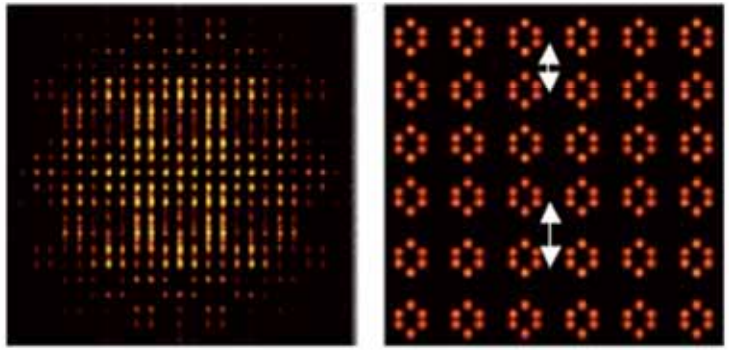
Independent random ‘atomic’ motions



Vertical transverse wave

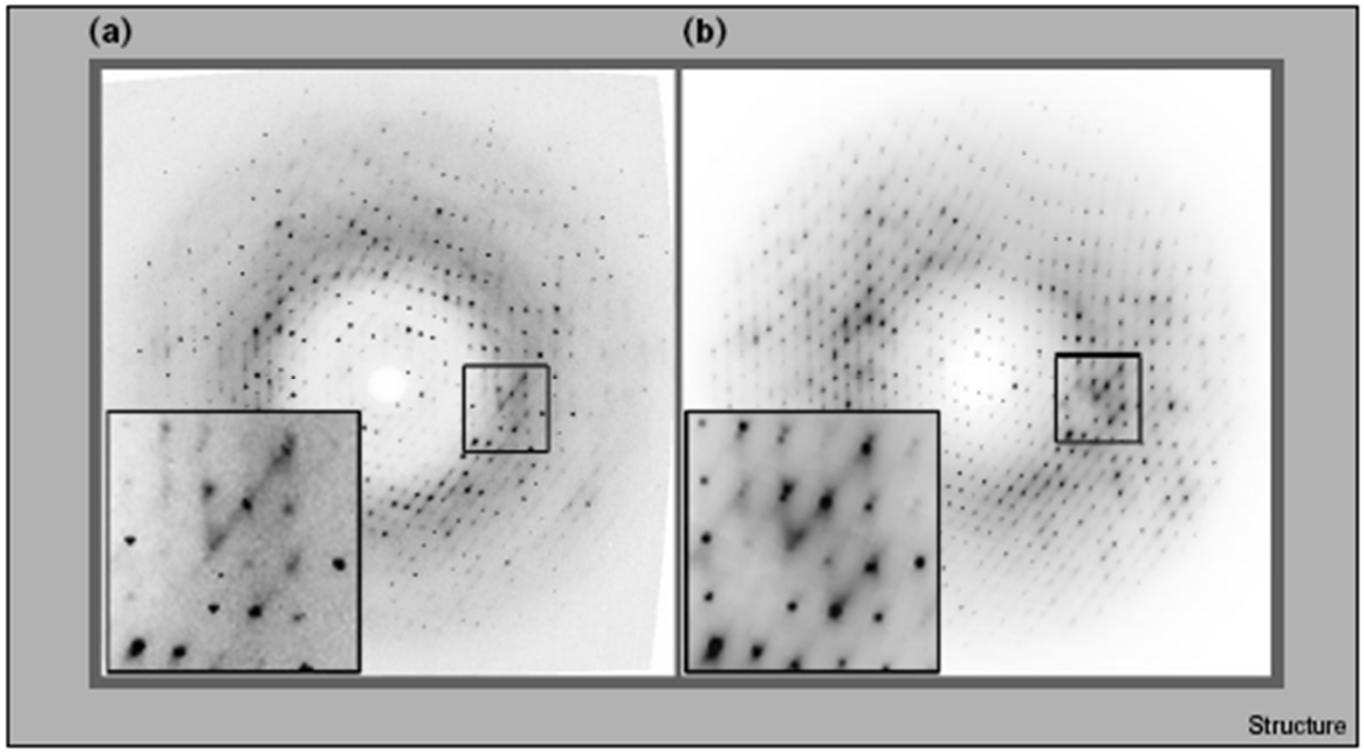


Random whole ‘molecule’ motions



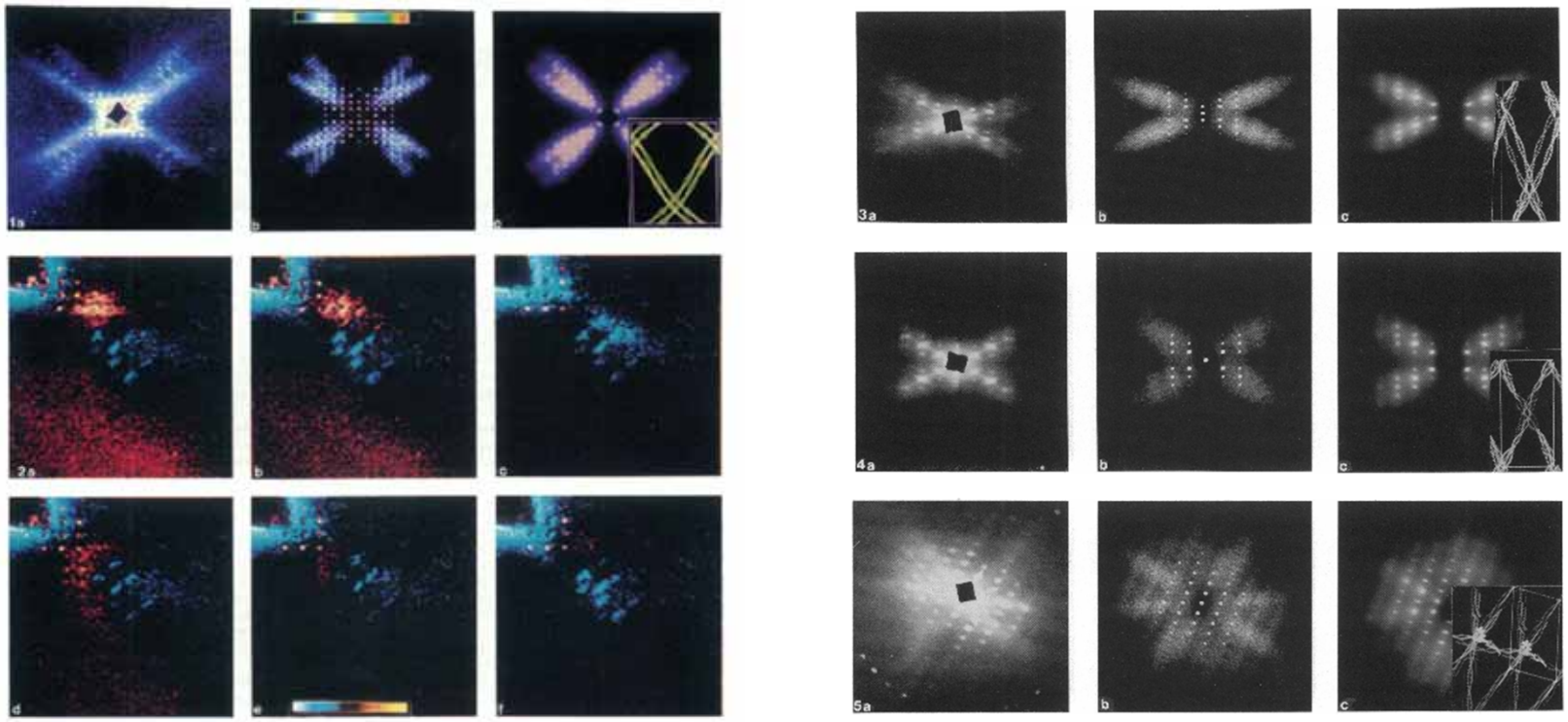
Vertical compression wave

Diffuse scattering depends on correlationed displacements



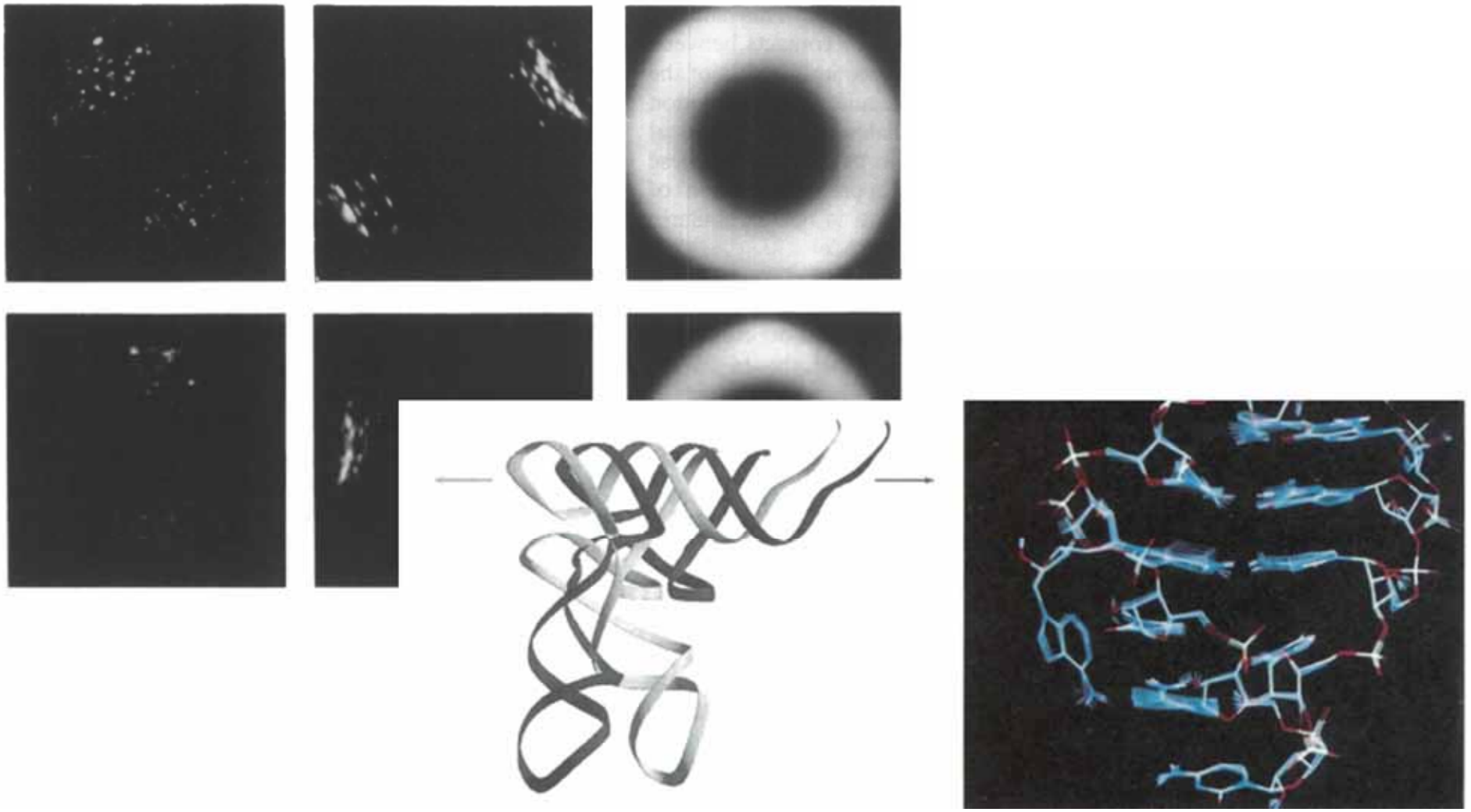
Wall, Clarage and Phillips *Structure*1997

Tropomyosin and its motions in the crystal



Chacko and Phillips, Biophys J.

tRNA motions



Kolatkar and Phillips, Acta D.

General statement about variational scatter

Whatever units are varying in non-correlated ways have their *intensity* transforms added as diffuse scatter.

In protein crystals, this will always be some atoms, some sidechains, and sometimes domains, whole molecules, or lattice coupled longer range assemblies.

How do coupled motions affect X-ray scattering?

- The Bragg diffraction is changed to some degree

- Diffuse scattering appears between (and underneath) the Bragg spots

Kinematic Theory (Thompson Scattering)

$$I(\mathbf{Q}) = FF^* = \left(\sum_{lk}^{N} f_{kQ} e^{i\mathbf{Q}\cdot\mathbf{R}_{lk}} \right) \left(\sum_{l'k'}^{N} f_{kQ}^{*} e^{-i\mathbf{Q}\cdot\mathbf{R}_{l'k'}} \right)$$

Assuming identical unit cells,

$$I(H) \propto FF^* = \left(\sum_{k}^{n} f_{kH} e^{iH\cdot\mathbf{R}_{k}} \right) \left(\sum_{k'}^{n} f_{kH}^{*} e^{-iH\cdot\mathbf{R}_{k'}} \right)$$

where *H* is restricted to integer triples

Time (and space) averaging

$$R_{lk} = r_{lk} + u_{lk}$$

Average position for atom in l^{th} unit cell and k^{th} atom in the cell

Displacement from average

$$I(\mathbf{Q}) = \sum_{lk} \sum_{l'k'} f_{kQ} f^{*}_{kQ} e^{i\mathbf{Q}\cdot(\mathbf{r}_{lk} - \mathbf{r}_{l'k'})} \left\langle e^{i\mathbf{Q}\cdot(\mathbf{u}_{lk} - \mathbf{u}_{l'k'})} \right\rangle$$

The usual approximation

$$\left\langle e^{i\mathbf{Q}\cdot(\mathbf{u}_{lk}-\mathbf{u}_{l'k'})}\right\rangle \cong e^{-\langle\{\mathbf{Q}\cdot(\mathbf{u}_{lk}-\mathbf{u}_{l'k'})\}^2\rangle/2}$$

Willis and Pryor, eqn 4.43

$$\langle\{\mathbf{Q}\cdot(\mathbf{u}_{lk}-\mathbf{u}_{l'k'})\}^2\rangle = \langle(\mathbf{Q}\cdot\mathbf{u}_{lk})^2\rangle + \langle(\mathbf{Q}\cdot\mathbf{u}_{l'k'})^2\rangle - 2\langle(\mathbf{Q}\cdot\mathbf{u}_{lk})(\mathbf{Q}\cdot\mathbf{u}_{l'k'})\rangle$$

James, pg 23.
The last term is usually dropped without mention.

Kinematic Treatment

General expression

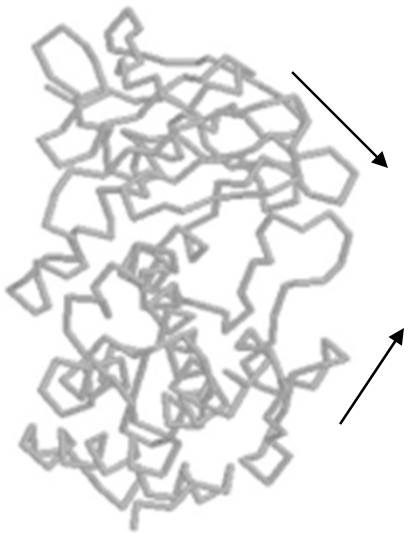
$$I(\mathrm{Q}) = \sum_{lk} \sum_{l'k'} f_{kQ} e^{-(\mathbf{Q}^T \langle \mathbf{u}_{lk} \mathbf{u}_{lk}^T \rangle \mathbf{Q})/2} f^*_{kQ} e^{-(\mathbf{Q}^T \langle \mathbf{u}_{l'k'} \mathbf{u}_{l'k'}^T \rangle \mathbf{Q})/2} e^{i\mathbf{Q}\cdot(\mathbf{r}_{lk} - \mathbf{r}_{l'k'})} e^{\mathbf{Q}^T \langle \mathbf{u}_{lk} \mathbf{u}_{l'k'}^T \rangle \mathbf{Q}}$$

For a crystal with identical unit cells and isotropic displacements

$$I(\mathrm{H}) = \sum_{k} \sum_{k'} f_{kH} e^{-(2\pi H)^2 \langle u_k^2 \rangle/2} f^*_{k'H} e^{-(2\pi H)^2 \langle u_{k'}^2 \rangle/2} e^{i2\pi\mathbf{H}\cdot(\mathbf{r}_k - \mathbf{r}_{k'})} \boxed{e^{(2\pi H)^2 \langle u_k u_{k'} \rangle}}$$

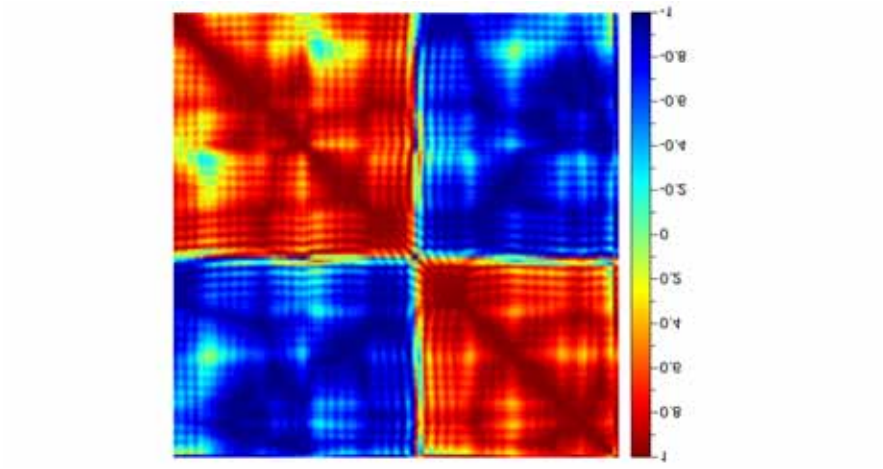
Where to get the correlation terms?

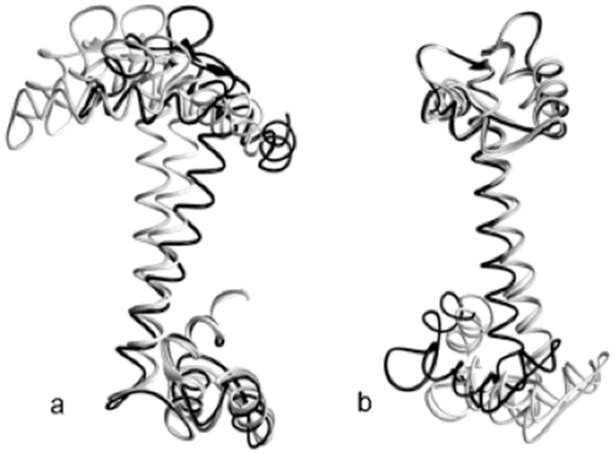
The inverse of the Kirchoff matrix or other Elastic Network models not only provide “B-factors”, but also covariance terms.

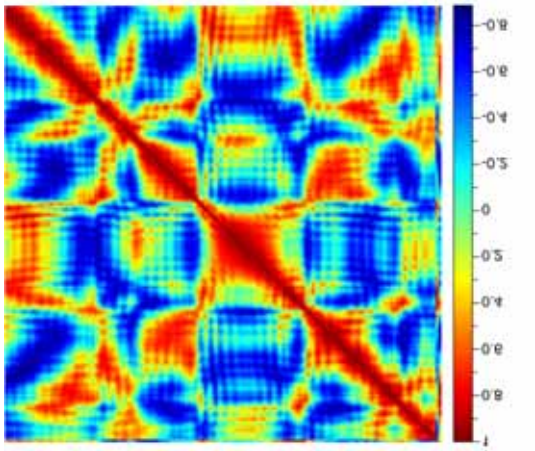


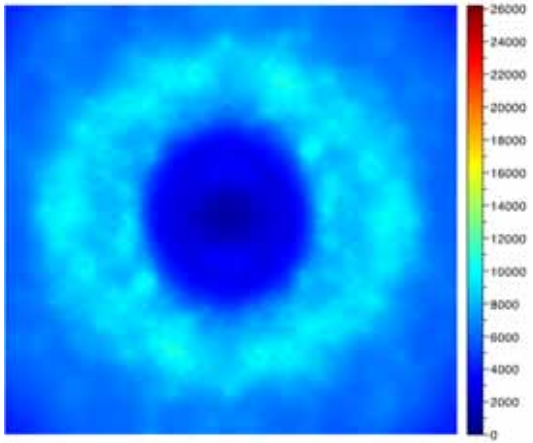
Compared to inorganic or small molecule structures, proteins have many more local (optical) modes, and they are all ignored at present in diffraction studies.

TnC rigid Covariance Matrix







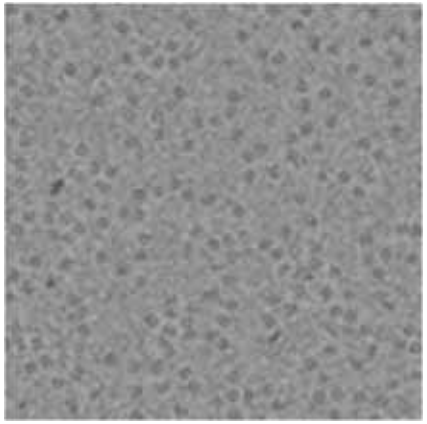


How do we get any structure right?

- “Lucky Larry” The diffraction the Bragg peak is miscalculated, but by subtracting the intensity around the peak, the oversight is corrected (to first order)
- The ‘richer’ the diffuse scatter, the worse is this approximation?

XFEL/ERL allows new approaches to biological structure

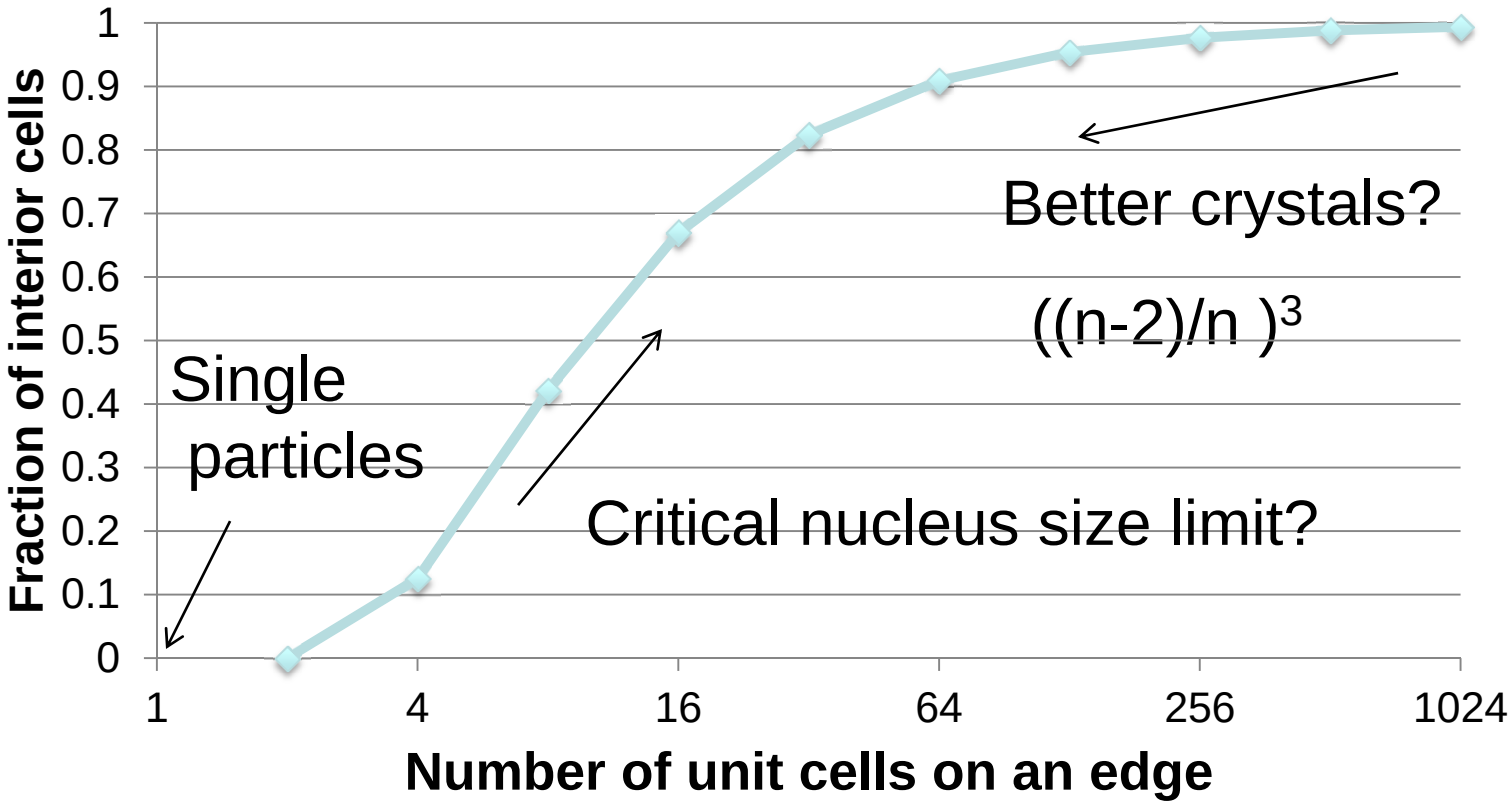
Hybrid culture
Some symmetry,
i.e. icosahedral or translational
Structures from “less hard to make”, *in situ* samples!



Single particle culture
Cryo-EM

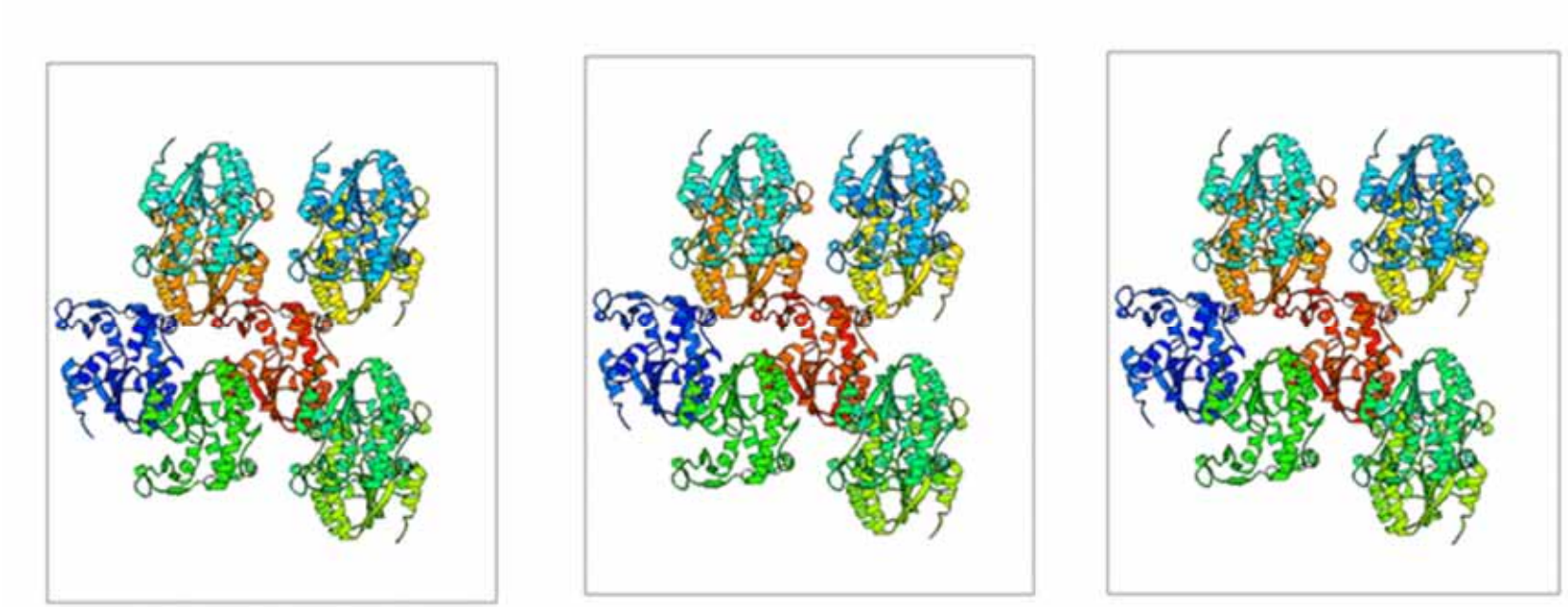
Crystallography culture
“Infinite”, but mosaic crystal

Fraction of interior unit cells



1 μm crystal with 100 Å cell = 100 unit cells on an edge

There exist an ensemble within nanocrystals AND the nanocrystals vary



Edge molecules will add significantly to the complication of the analysis. According to the buildup method with thousands of nanocrystals, intensities of different structures will be added ‘incoherently”.

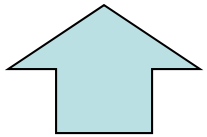
The “instantaneous” structure and scattering of the solvent around the protein will also be problematic.

Edge effects on Scattering/Data Processing

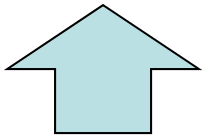
(after Spence, Lattman)

$$\rho_{1molecule} \otimes (L_{\inf} \times Cube_{finite}) \neq (\rho_{1molecule} \otimes L_{\inf}) \times Cube_{finite}$$

$$FT(\rho_{1molecule}) \times \left[(FT(L_{\inf}) \otimes FT(Cube_{finite}) \right] \neq \left[FT(\rho_{1molecule}) \times FT(L_{\inf}) \right] \otimes FT(Cube_{finite})$$

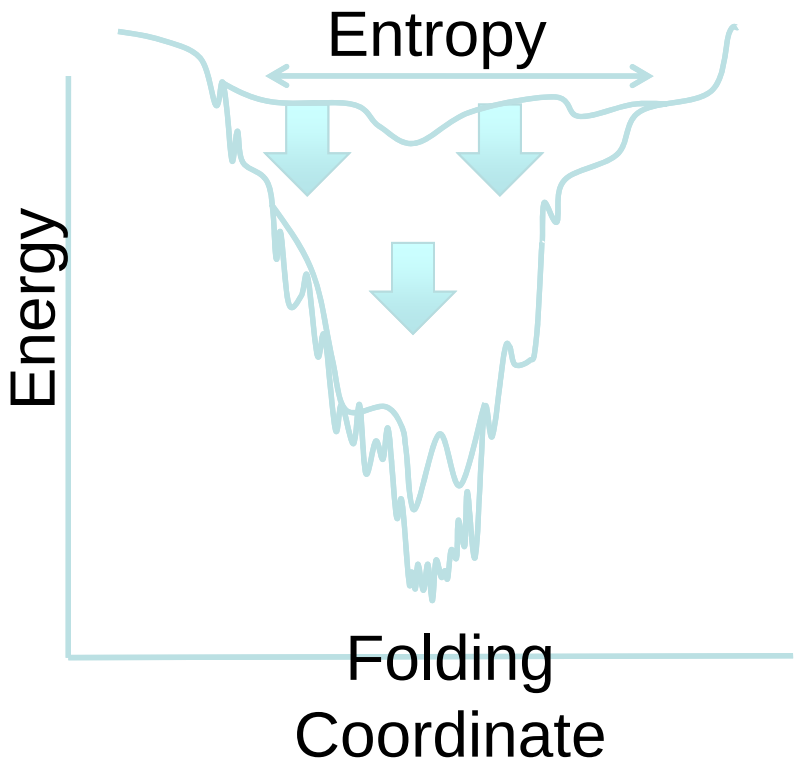


Fringes sample the molecular transform; adding fringes thus complicating integration of the "spot"



If edges are largely 'disordered' and don't contribute except at very low resolution, fringes will scale with Bragg spots instead of varying

One point: Variability in the samples will (eventually) limit what can be done. Need to manage the conformations of proteins for XFEL studies and learn to deal with variation, even learn from it!



Acknowledgements

NIH, NSF, DOE, WARF

- Elena Levin
- Dmitri Kondrashov
- Jason McCoy
- Ryan Bannen
- Roman Aranda
- Andre Francis
- Friedrich Schotte
- Philip Anfinrud
- Anand Kolatkar
- Gary Wesenberg
- Craig Bingman
- Ed Bitto
- Michael Wall
- Wei Zhang
- Bog Stec
- Sibsankar Kundu
- Euiyoung Bae
- Demian Riccardi
- All other members

PSI Protein Structure Initiative

NIGMS