“Non-Bragg” Scattering from Protein Crystals

George N. Phillips Jr.

Department of Biochemistry and Cell Biology
Rice University, Houston

Introduction

- Crystallography (usually) confuses the space and time averages.
- Dynamic behavior is ever present--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 doing the same thing in every unit cell or even space group symmetry, is only a first order approximation

Space (and time) averaging displacements

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

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

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

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

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

How do we get any structure right?

- “Lucky Larry” The diffraction at 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 this approximation?

The variance covariance matrix terms

$$\langle [\boldsymbol{q} \cdot (\boldsymbol{u}_{m,s} - \boldsymbol{u}_{m',s'})]^2 \rangle = \langle (\boldsymbol{q} \cdot \boldsymbol{u}_{m,s})^2 \rangle + \langle (\boldsymbol{q} \cdot \boldsymbol{u}_{m',s'})^2 \rangle - 2\langle (\boldsymbol{q} \cdot \boldsymbol{u}_{m,s})(\boldsymbol{q} \cdot \boldsymbol{u}_{m',s'}) \rangle$$

Last term can be written…

$$-2\boldsymbol{q}^T \langle \boldsymbol{u}_{m,s} \boldsymbol{u}_{m',s'}{}^T \rangle \boldsymbol{q}$$

Where the term in brackets is a 3x3 matrix for each atom in each unit cell (anisotropic)

Or

Collapsed to an isotropic term, one per atom, a projection onto the scattering vector reduces this to a scalar

Diffuse X-ray Scattering
Non-Bragg Scattering

- Has information about displacements from the average structure
- Illustrates intrinsic mechanical properties of the macromolecule
- Couples with lattice motions

Clarage and Phillips, *Methods Enz.* 1997

Scattering as a series of terms

$$I(\mathbf{q}) = I_0 + I_1 + I_2 + \ldots ,$$

$$I_0 = NI_e \sum_{m,s,s'} \{f_s f_{s'} \exp(-M_s - M_{s'}) \times \exp[-i\mathbf{q}\cdot(\mathbf{R}_m + \boldsymbol{\tau}_{s,s'})]\} ,$$

$$I_1 = NI_e \sum_{m,s,s'} \{f_s f_{s'} \exp(-M_s - M_{s'}) \times \exp[-i\mathbf{q}\cdot(\mathbf{R}_m + \boldsymbol{\tau}_{s,s'})]\, G_{m,s,s'}(\mathbf{q})\} ,$$

$$I_2 = \frac{NI_e}{2} \sum_{m,s,s'} \{f_s f_{s'} \exp(-M_s - M_{s'}) \times \exp[-i\mathbf{q}\cdot(\mathbf{R}_m + \boldsymbol{\tau}_{s,s'})]\, [G_{m,s,s'}(\mathbf{q})]^2\} .$$

Xu and Chiang, Z Cryst. 2005

Other manipulations…

$$I_{Total}(\boldsymbol{q}) = I_{Bragg}(\boldsymbol{q}) + I_{Diffuse}(\boldsymbol{q})$$

$$I_{Diffuse}(\boldsymbol{q}) = N(\langle|F_n|^2\rangle - |\langle F_n\rangle|^2)$$

$$I_{Diffuse}(\boldsymbol{q}) = N\sum_j \left[f^j(\boldsymbol{q})\right]^2 \left[1 - e^{-\frac{1}{2}\langle\left(\boldsymbol{q}\cdot\boldsymbol{u}_n^j\right)^2\rangle}\right]$$

This latter term supports the notion that as the Debye Waller factor takes photons from the Bragg spots, they appear elsewhere in the pattern, with no diffuse scatter near the origin

Comes down to the variance covariance matrix of the displacements

- Use a simple function to approximate
- Use molecular dynamics or NMA method
- Maybe refine in some restrained way? (Future)

Go et al., Proteins, 2004

Define Diffuse Scatter?

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

How do coupled displacements affect X-ray scattering?

- The Bragg diffraction is changed to some degree

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

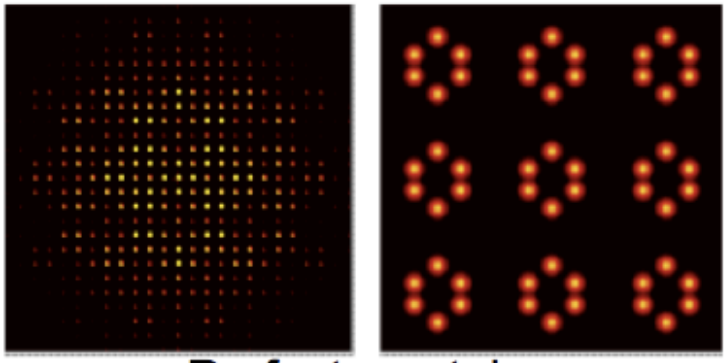
Analytical/Multi-cell Method

There are two common methods of calculating the diffuse scattering from structures

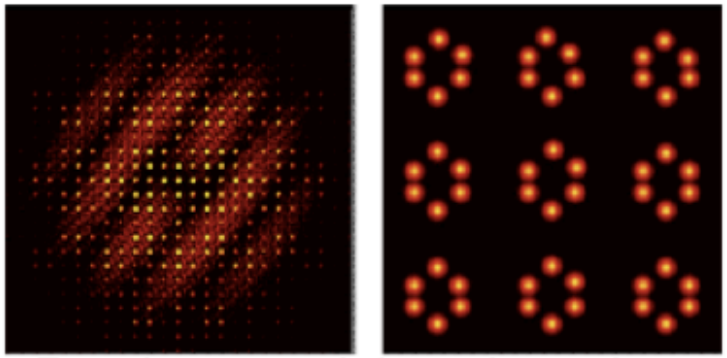
1. Analytical approximations

2. A Monte Carlo multi-cell simulation

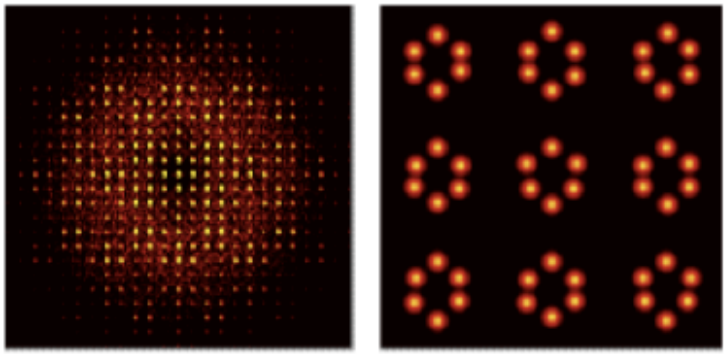
Adapted from Clarage and Phillips, Methods Enz, 277 (1996).



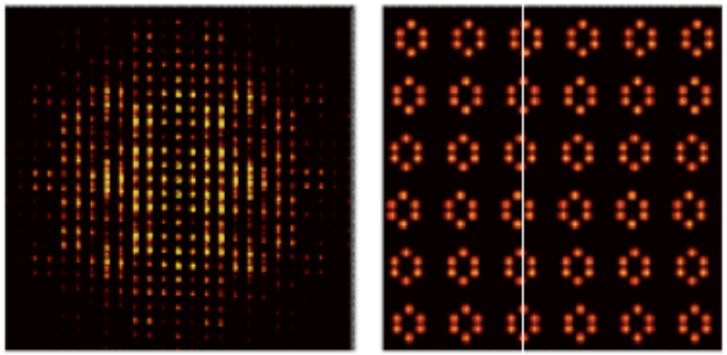
Perfect crystal



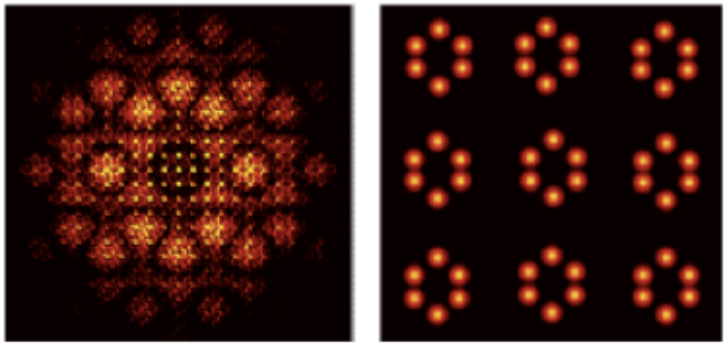
Random top-right pair motions



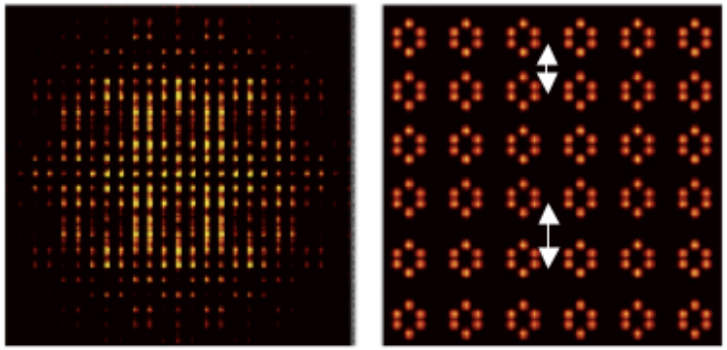
Independent random ‘atomic’ motions



Vertical transverse wave



Random whole ‘molecule’ motions



Vertical compression wave

General statements about diffuse scatter

Whatever units are varying in a concerted ways, but independent of other units, their *intensity* transforms add as diffuse scatter.

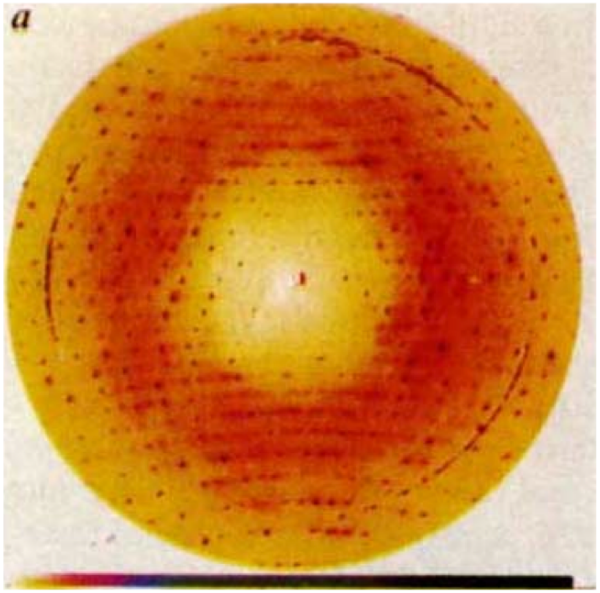
In protein crystals, this can be atoms, clusters of sidechains, or domains, whole molecules, or lattice coupled longer range assemblies.

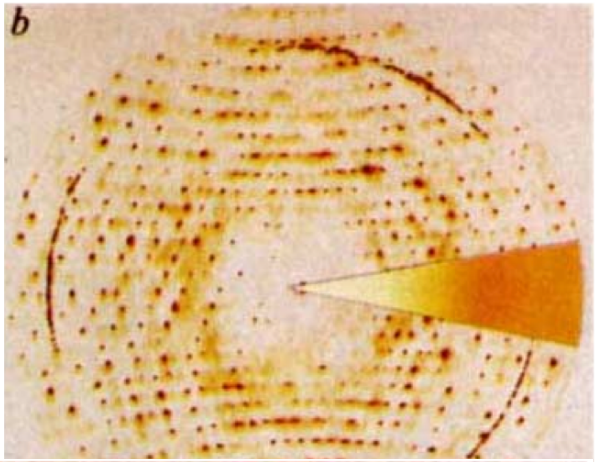
If the variation in structure is a displacement, the diffuse scatter will diminish at small q, if substitutional, it will be strong near the origin of reciprocal space.

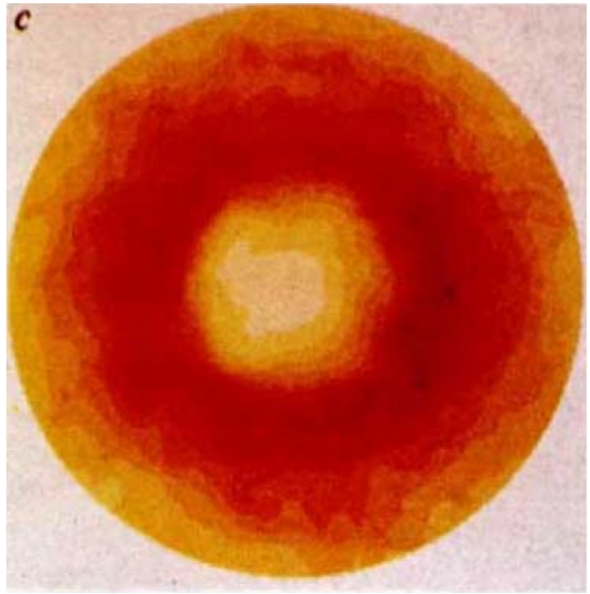
If there are streaks emmenating from the Bragg spots, this will likely arise from transverse waves of disorder with a propagation direction along the direction of the streak.

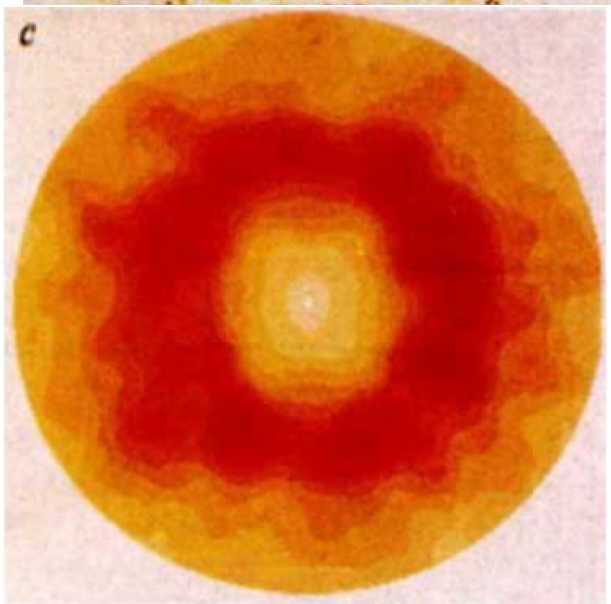
Liquid-like correlations?

Think up a reasonable potential function for the interactions and see what you get…









Caspar et al.
Nature, 1988

Insulin, a coupling constant and an amplitude

Use a better model

Use an molecular dynamics trajectory or at least a good number of the low frequency modes from a careful normal mode analysis…

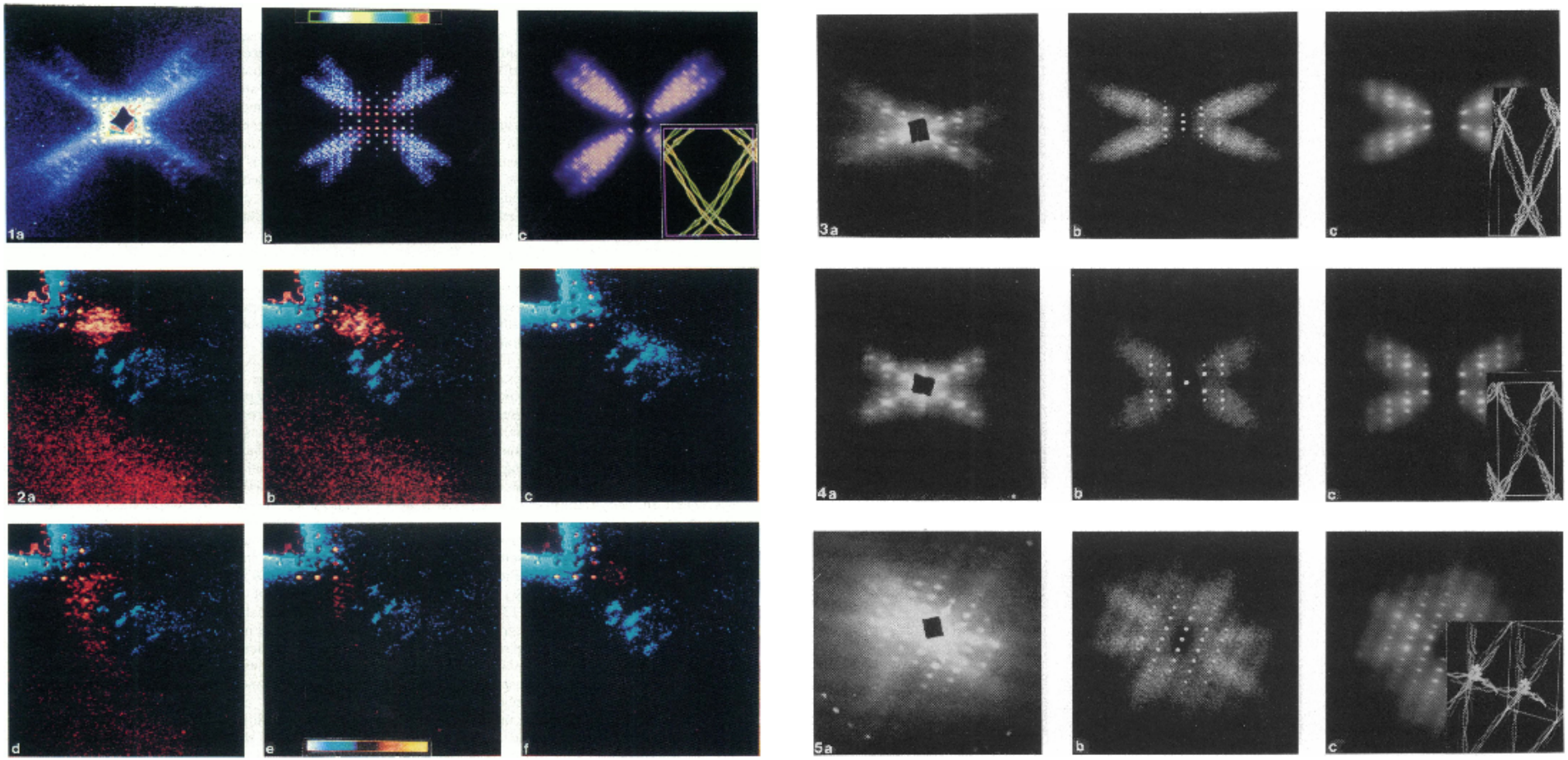
And a careful analysis…

Same staph nuclease data set.

Get a more satisfying fit.

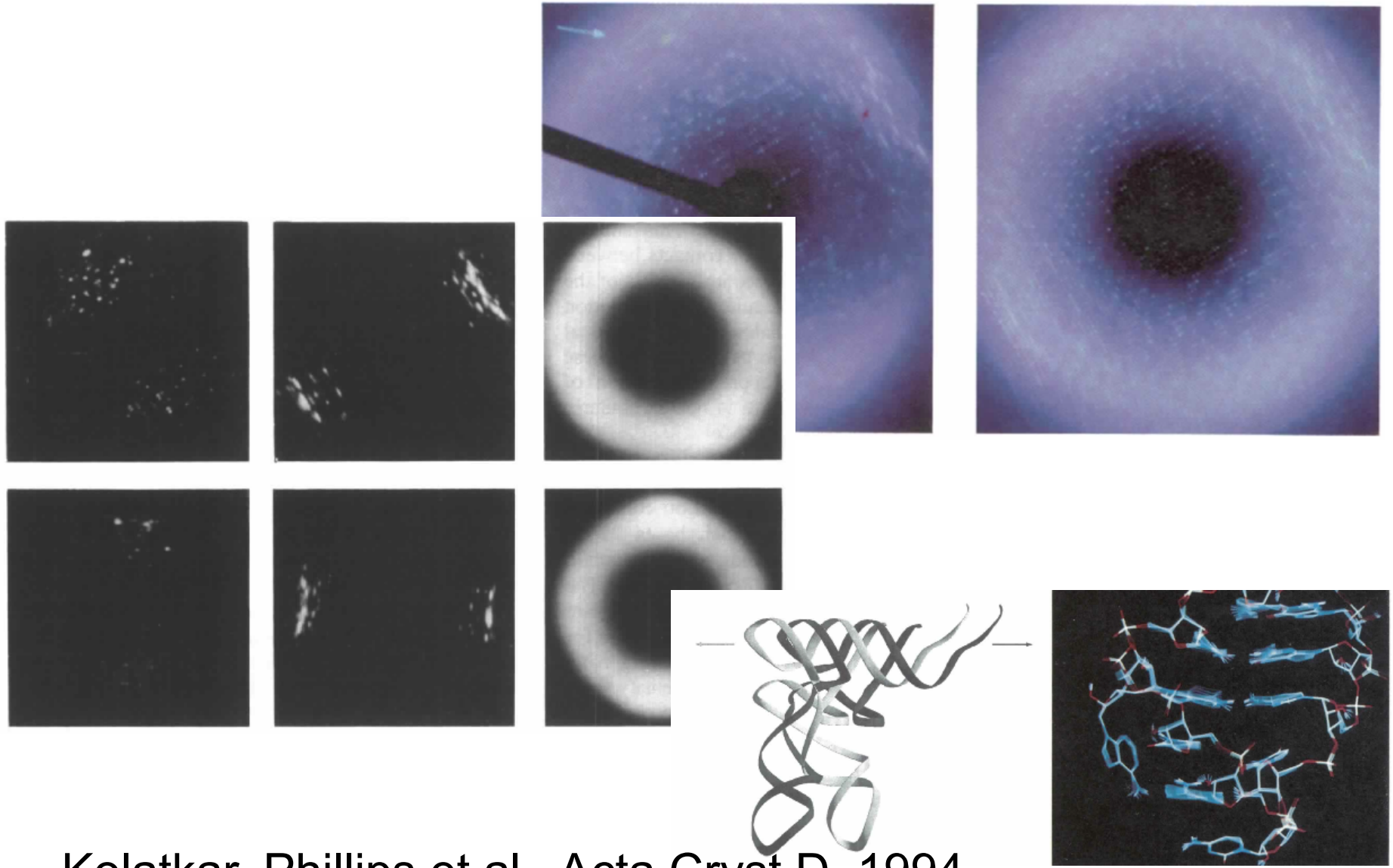
Meinhold and Smith, Proteins, 2007

Tropomyosin and its motions in the crystal



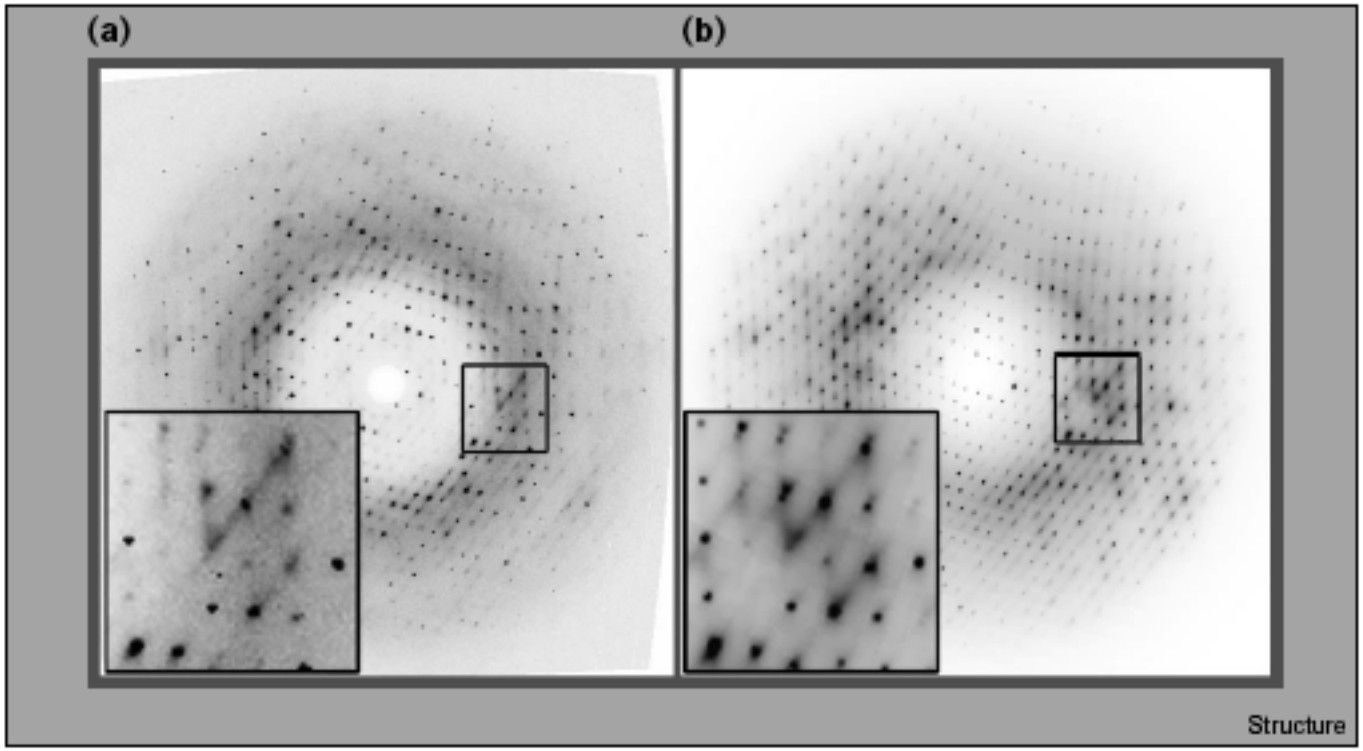
Chacko and Phillips, Biophys J.

tRNA motions



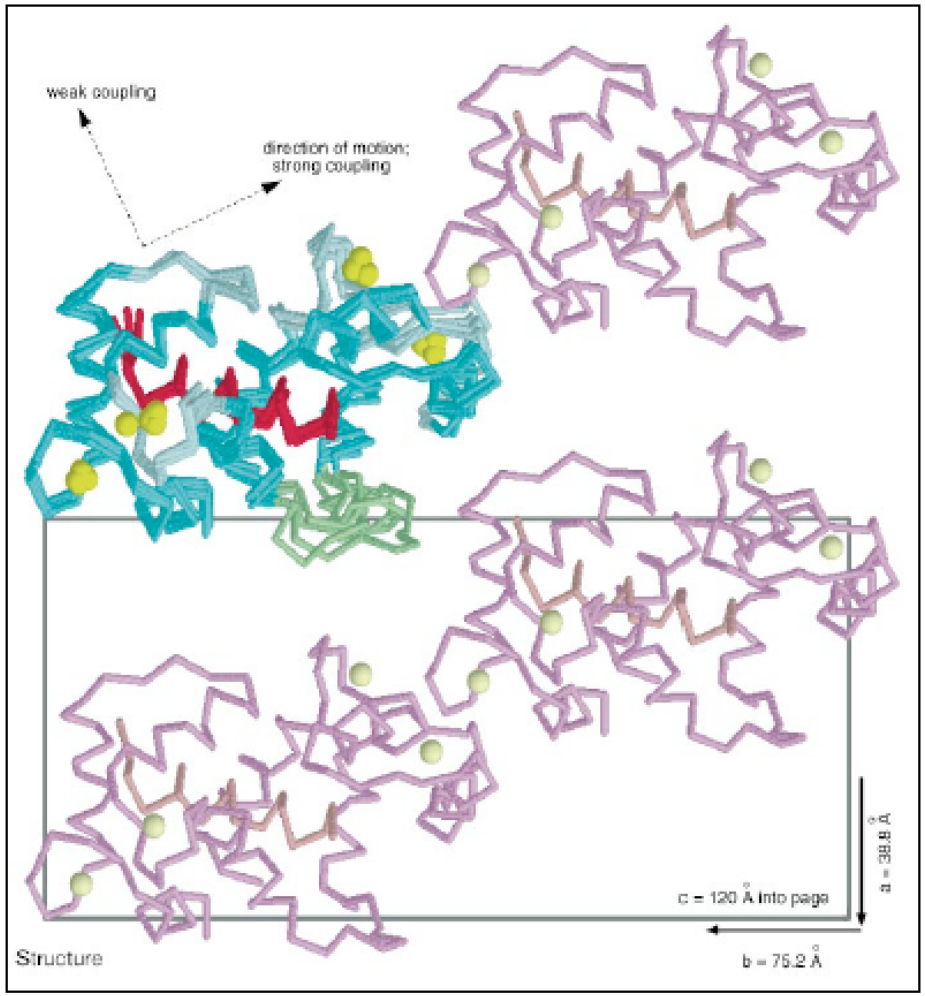
Kolatkar, Phillips et al., Acta Cryst D. 1994

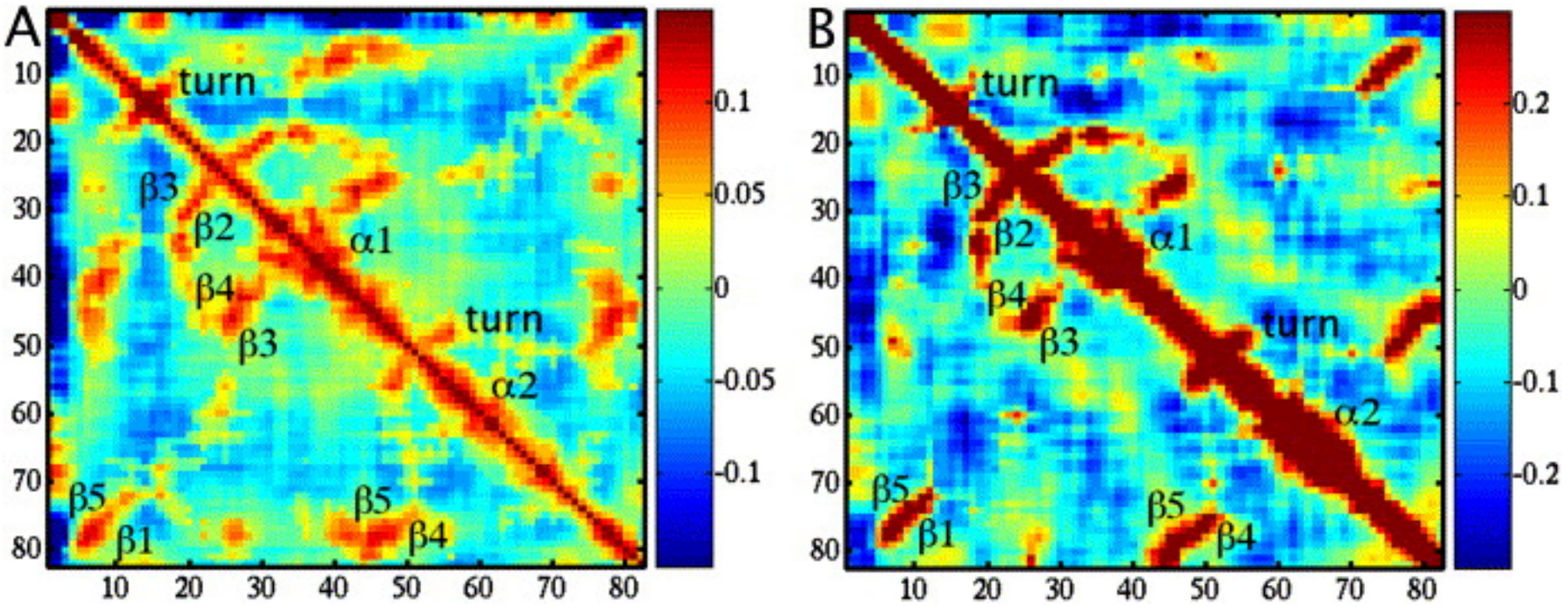
Diffuse scattering depends on correlationed displacements



Wall, Clarage and Phillips *Structure*1997

Lattice couplings





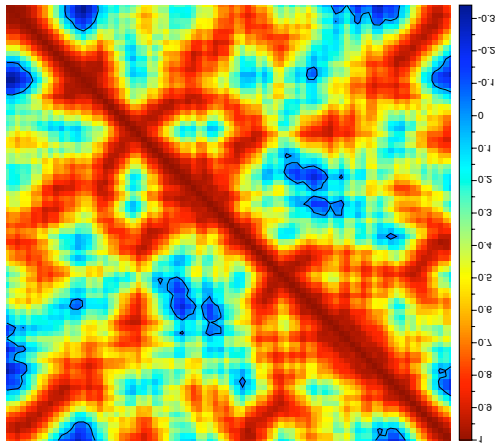
Correlation Matrices Generated from Normal Mode Analyses of a PDZ Domain

(A) Correlation from an anisotropic network model (DNM).

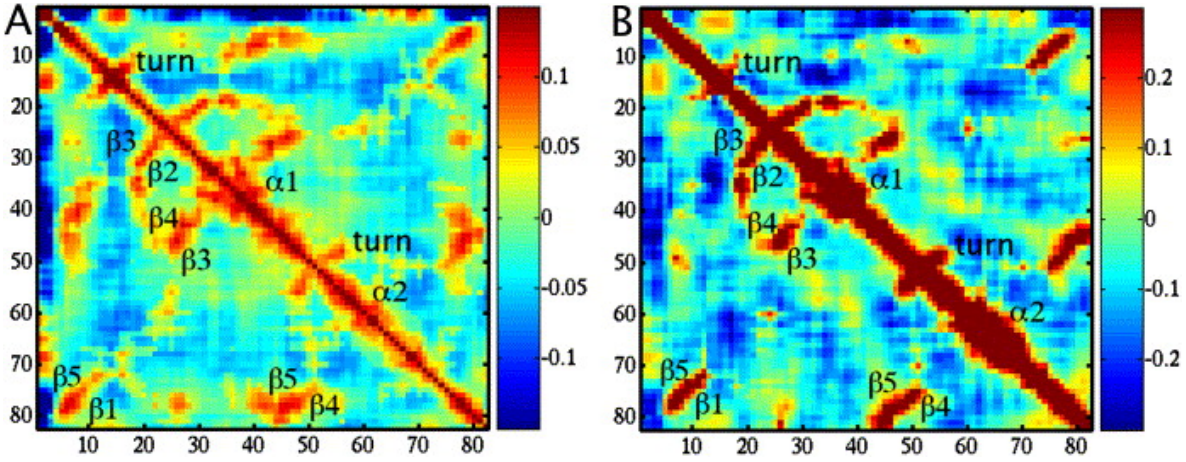
(B) Correlation from CHARMM-based block normal modes.

Kondrashov, Phillips et al., Protein Structural Variation in Computational Models and Crystallographic Data, *Structure* (2007).

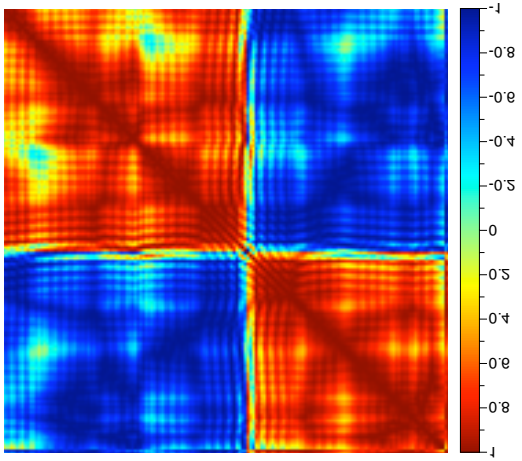
Rigid body rotation comparison

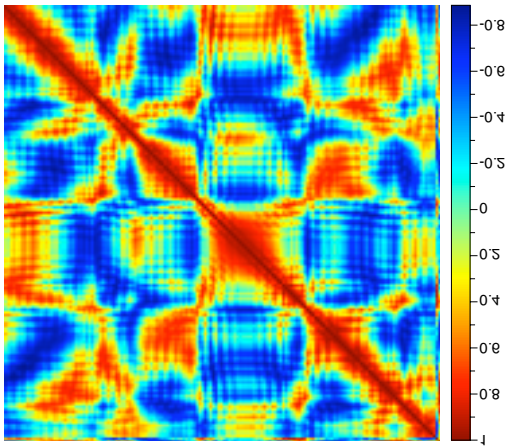


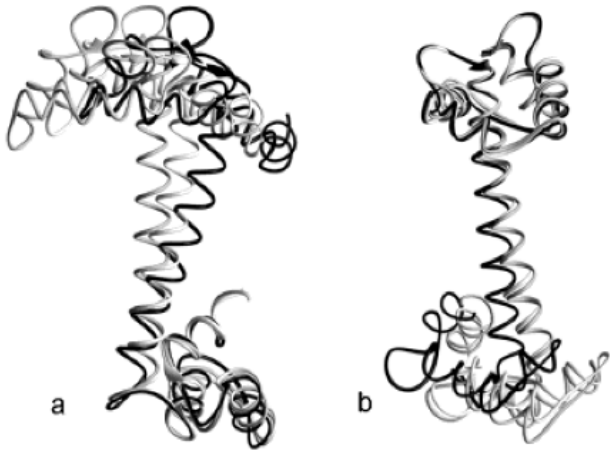
Has too much correlation, but similar diagonal

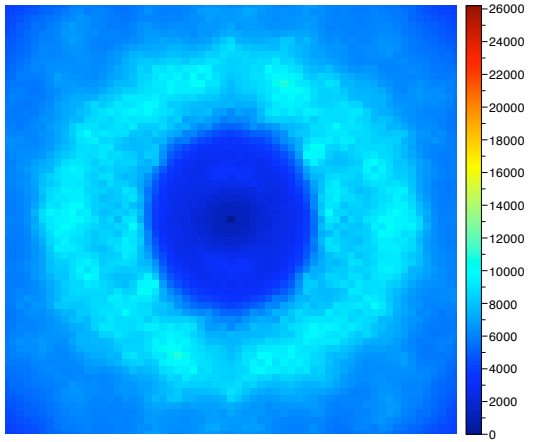


TnC rigid Covariance Matrix









What about Calculation/ Modeling of Motions

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

Riccardi paper…

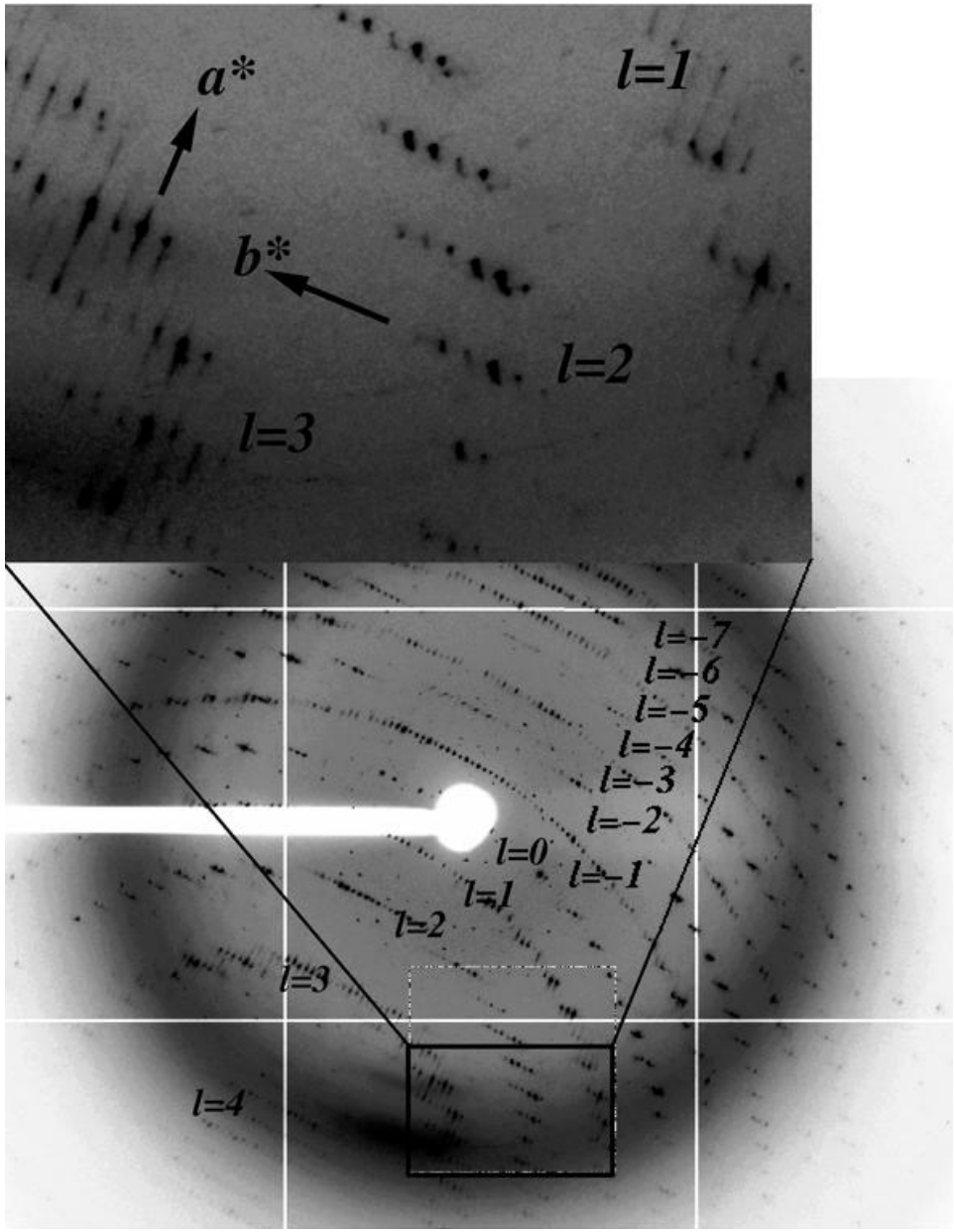
- Pointed out too many goofy models out there that don't have reasonable density of states or heat capacities….
- Its easy to fit the B-factors, but models are all over the place for the covariance terms of the matrix

Riccardi, Cui, and Phillips, Biophys J, 2009, 2010

Diffuse scatter comes from any variations

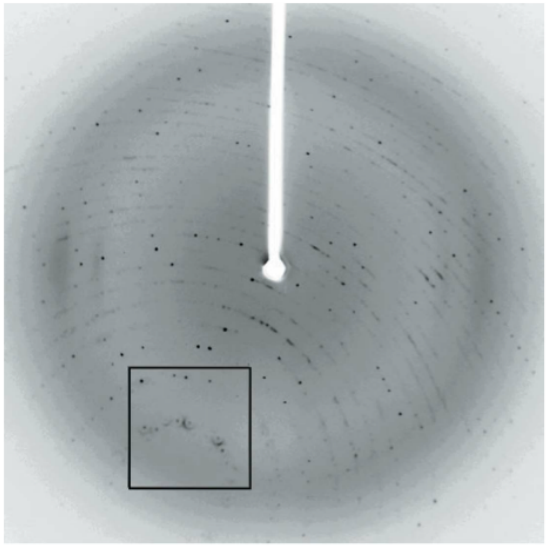
- From displacements
- From substitution disorder
- From changes in electron density

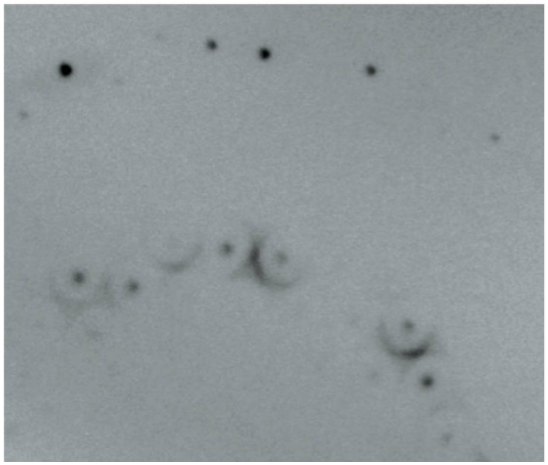
Modulation in layers (one kind of ‘substitution disorder’)



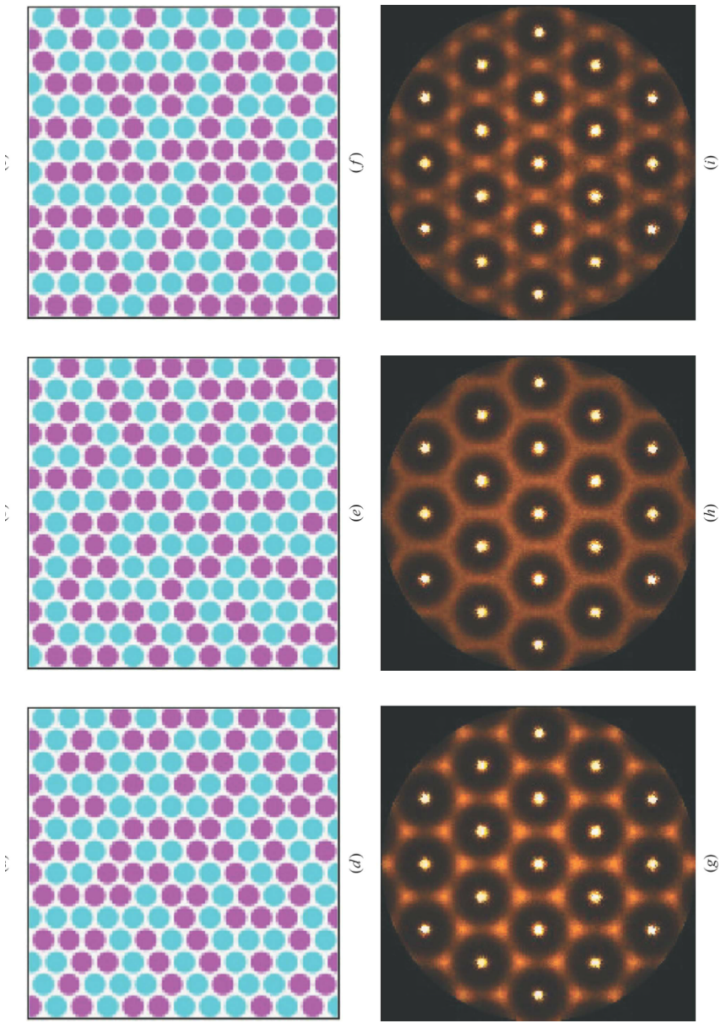
Wang, J., Kamtekar, S., Berman, A. J. & Steitz, T. A. (2005). Acta Cryst. D61, 67-74.

Frustration as source of diffuse scatter





Fragment of GAG protein from feline foamy virus



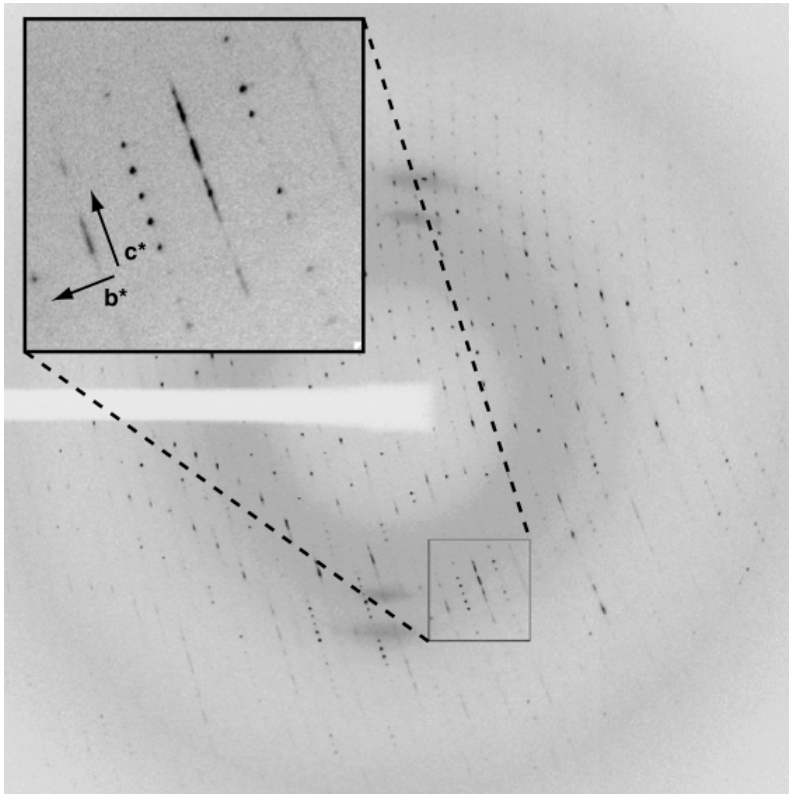
Wellberry et al., Acta Cryst B, 2011

Multi-cell method

Alternating layers as source of diffuse scatter

$$I_{\text{total}}(\mathbf{h}) = \frac{1}{N}\sum[(2\kappa^2 - 2\kappa + 1) + 2\kappa(1-\kappa)\cos(2\pi\mathbf{h}\mathbf{t}_d)]I_0$$

N is number of layers
K is fraction
H is recip lattice index
Td = fractional shift
Io is ‘measured’ I



Need more data sets or at least saved frames!

- Methods are needed to collect diffuse scattering data sets, free from air scatter and other artifacts (new detectors)
- Preferably with small incident beams so that Bragg spots don't dominate (third generation syncs, XFEL?)
- Like the staph nuclease data from Wall and Gruner, data should be shared (Who can curate/host data?)
- Need to match calculations and measured frames, not abstracted data (high performance computing?)

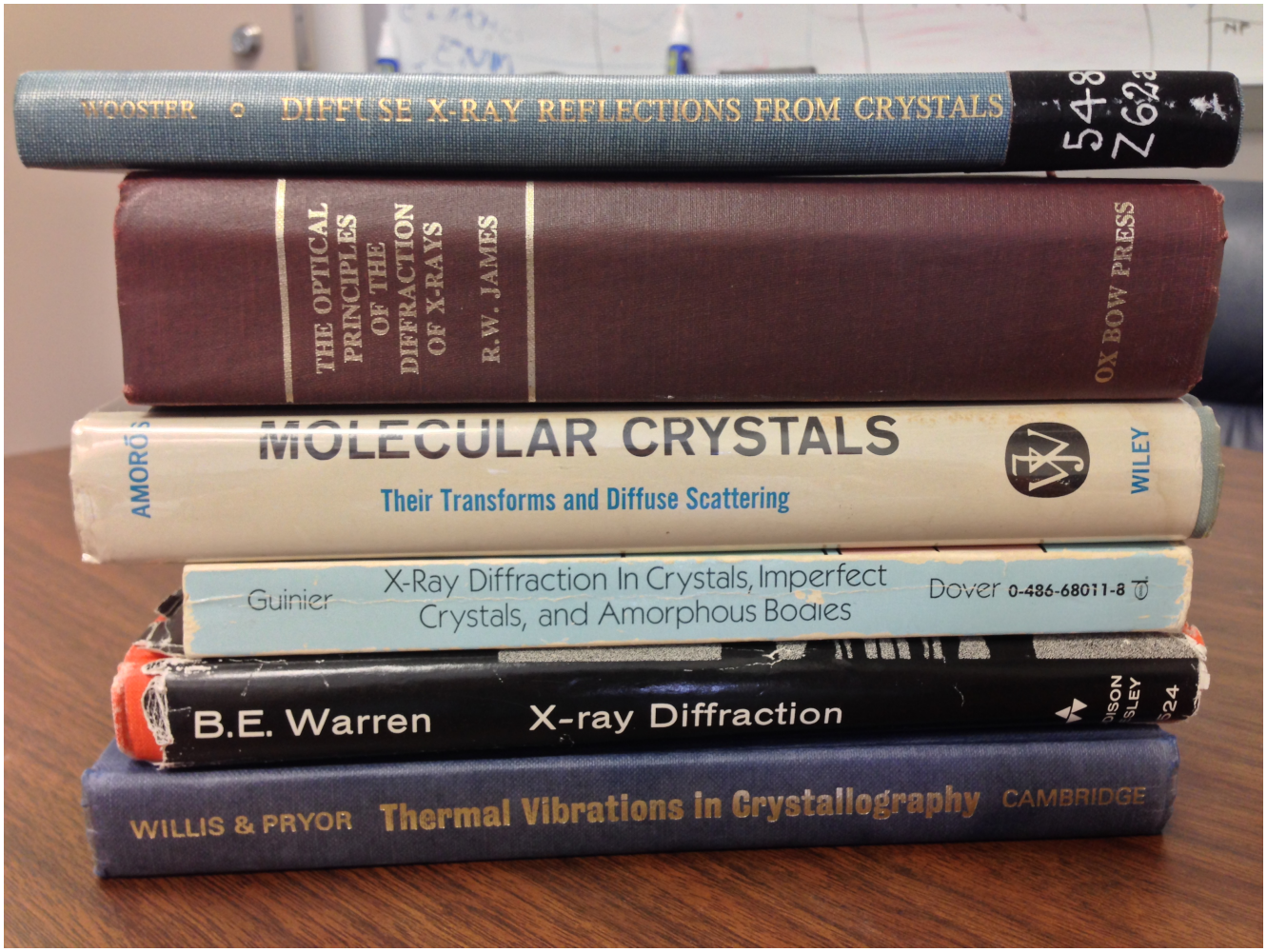
What can we learn?

- More about the ensembles of protein conformations
- Ideally a solid variance-covariance matrix
- Connections between structure, dynamics, and function
- Crystallographic science

Displacement Diffuse Scatter

- Has information about displacements from the average structure
- Illustrates intrinsic mechanical properties of the macromolecule
- Often couples with the lattice to give more complicated situation

The books


WOOSTER
DIFFUSE X-RAY REFLECTIONS FROM CRYSTALS
THE OPTICAL PRINCIPLES OF THE DIFFRACTION OF X-RAYS
R.W. JAMES
OX BOW PRESS
AMORÓS
MOLECULAR CRYSTALS
Their Transforms and Diffuse Scattering
WILEY
Guinier
X-Ray Diffraction In Crystals, Imperfect Crystals, and Amorphous Bodies
Dover 0-486-68011-8
B.E. Warren
X-ray Diffraction
WILLIS & PRYOR
Thermal Vibrations in Crystallography
CAMBRIDGE

Acknowledgements

NIH, NSF, DOE, WARF

- Elena Levin
- Dmitri Kondrashov
- James Clarage
- Anand Kolatkar
- Roman Aranda
- Andre Francis
- Friedrich Schotte
- Philip Anfinrud
- Gary Wesenberg
- Craig Bingman
- Jonathan Clinger
- Michael Wall
- Wei Zhang
- Bog Stec
- Sibsankar Kundu
- Demian Riccardi
- All other members

PSI
Protein Structure Initiative