Metadata needed for the full exploitation of diffuse scattering data from protein crystals

Michael Wall
Los Alamos National Laboratory

ECM 29 Satellite Workshop on Metadata for Raw Data from X-ray Diffraction and Other Structural Techniques
Rovinj, Croatia
22-23 August 2015
Diffuse X-Ray Scattering to Model Protein Motions

Michael E. Wall,¹,* Paul D. Adams,²,³ James S. Fraser,⁴ and Nicholas K. Sauter²

¹Computer, Computational, and Statistical Sciences Division, Los Alamos National Laboratory, Los Alamos, NM 87545, USA
²Physical Biosciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA
³Department of Bioengineering, University of California, Berkeley, Berkeley, CA 94720, USA
⁴Department of Bioengineering and Therapeutic Sciences, University of California, San Francisco, San Francisco, CA 94143, USA
*Correspondence: mewall@lanl.gov
http://dx.doi.org/10.1016/j.str.2014.01.002

Problems in biology increasingly need models of protein flexibility to understand and control protein function. At the same time, as they improve, crystallographic methods are marching closer to the limits of what can be learned from Bragg data in isolation. It is thus inevitable that mainstream protein crystallography will turn to diffuse scattering to model protein motions and improve crystallographic models. The time is ripe to make it happen.
Kathleen Lonsdale
Diffuse Scattering Pioneer

• 1924, student of Lawrence Bragg
• 1928, Solved benzene structure
  – Ended 60 year debate about flat aromatic ring
• 1942, Champion of diffuse X-ray scattering
• 1945, Fellow of Royal Society
  – One of first two women (along with Marjory Stephenson)
• 1956, Dame Commander of OBE
• 1966, First woman president of IUCr
\[ I_D(q) = \sum_{g,s} \left[ I_0(q+g) + I_0(q-g) \right] |q \cdot v_s(g)|^2 \]

Bragg reflection
Diffuse reflection
Origin

Born & Sarginson, 1941
James, 1948

Thermal Diffuse Scattering (TDS)
\[ F = \sum_{n} f_{n} e^{i\mathbf{q} \cdot \mathbf{R}_{n}} \]

\[ I(t) = |F|^{2} = \sum_{n} \sum_{m} f_{n} f_{m}^{*} e^{i\mathbf{q} \cdot (\mathbf{R}_{n} - \mathbf{R}_{m})} \]

\[ I = \sum_{n} \sum_{m} \left[ \left( \langle |f_{n}|^{2} \rangle - \langle f_{n} \rangle^{2} \right) \delta_{nm} + \langle f_{n} \rangle^{2} \right] e^{i\mathbf{q} \cdot (\mathbf{R}_{n} - \mathbf{R}_{m})} \]

\[ I = I_{D} + I_{B} \]

\[ I_{B} = |\langle f \rangle|^{2} \sum_{n} \sum_{m} e^{i\mathbf{q} \cdot (\mathbf{R}_{n} - \mathbf{R}_{m})} \]

\[ I_{D}(\mathbf{q}) = N \left( \langle |f_{n} - \langle f_{n} \rangle|^{2} \rangle \right) \approx N \int d\mathbf{x} e^{i\mathbf{q} \cdot \mathbf{x}} \int d\mathbf{x}' \langle \Delta \rho_{n}(\mathbf{x}') \Delta \rho_{n}(\mathbf{x} + \mathbf{x}') \rangle_{n} \]

Sample using molecular dynamics simulations

Guinier 1956, 1963
Wall, 1996 (Ph.D. Thesis)
Wall, Ealick, and Gruner, *PNAS* 1997
http://github.com/mewall/lunus
\[ I_D(\vec{q}) = (1 - e^{-\vec{q} \cdot U \cdot \vec{q}}) e^{-\vec{q} \cdot U \cdot \vec{q}} I_0(\vec{q}) * \Gamma_G(\vec{q}) \]

\[ \Gamma_G(\vec{q}) = \frac{8 \pi \det|G|}{\left[ 1 + |Gq|^2 \right]^2} \]

Liquid-like motions (Caspar et al)

Snase: \( \gamma = 10 \text{ Å}; \sigma = 0.36 \text{ Å} \)
Calmodulin: \( \gamma = 4.8 \text{ Å}; \sigma = 0.38 \text{ Å} \)

\[ \Gamma_G(\vec{q}) = \frac{4 \pi \det|G|}{1 + |Gq|^2} \]

Acoustic modes

Correlations: Displacements:
\( \gamma_1 = 50 \text{ Å} \quad \sigma_1 = 0.0 \text{ Å} \)
\( \gamma_2 = 135 \text{ Å} \quad \sigma_2 = 0.4 \text{ Å} \)
\( \gamma_3 = 85 \text{ Å} \quad \sigma_3 = 0.0 \text{ Å} \)
\[ L = \frac{1}{2} \sum m_i \dot{\mathbf{R}}_i^2 - U(\mathbf{R}_1, \ldots, \mathbf{R}_N) \]

\[ \frac{d}{dt} \left( \frac{\partial L}{\partial \dot{\mathbf{R}}_k} \right) = - \frac{\partial L}{\partial R_k} \]

\[ \mathbf{f}_i = m_i \ddot{\mathbf{R}}_i = - \frac{\partial U}{\partial \mathbf{R}_i} \]
Clarage et al, *PNAS* 1995

Meinhold & Smith, *Biophys J* 2005
Wall et al, *PNAS* 2014
CC_{all} = 0.94
Data

14,800 independent data points

MD Model

Overlay

$CC_{\text{all}} = 0.49$
Predicting X-ray diffuse scattering from translation–libration–screw structural ensembles

Andrew H. Van Benschoten, A Pavel V. Afonine, B Thomas C. Terwilliger, C Michael E. Wall, D Colin J. Jackson, E Nicholas K. Sauter, B Paul D. Adams, B,F Alexandre Urzhumtsev G,H and James S. Fraser A*

Image processing and scripts in Lunus [http://github.com/mewall/lunus]

DIALS indexing methods from Nicholas Sauter and Aaron Brewster, LBNL

Experiments of James Fraser and Others
Raw Images

- No compression
- Simple layout of data
- Shared conventions for $r,c \rightarrow x,y$
- Human-readable header
  - *i.e.* SMV or the like
Beam Metadata

- Beam line
- Wavelength
  - Spectrum
- Polarization
  - Evidence
- Beam center
  - Evidence
Detector Metadata

- Detector model
- Detailed operating mode
- Relation of ADU to X-ray counts
- Distance
  - Evidence
- Detector face rotation
  - Evidence
- Pixel size
Crystal Metadata

- Space group
  - Evidence

- Unit cell
  - Evidence

- Chemical contents

- Light microscopy image for each exposure
  - Distinguish crystal from other scattering sources
  - Tomography model of specimen
Integrated Diffuse Data Deposition

- $D(hkl)$
  - Fractional $hkl$ possible
- Image processing parameters
  - Beam polarization
  - Solid-angle normalization
  - Bragg peak filtering
- Scale factors
- Frame-by-frame indexing information
- Flexible with respect to future needs for combined integration of Bragg and diffuse data
Diffuse Scattering Model Deposition

• Dynamical parameter values
  – Displacement correlations
  – Displacement amplitudes
  – Dispersion relation
• MD trajectories (large!)
• Calculated diffuse intensities
Acknowledgments

• Rice
  – George Phillips, Jr.
  – James Clarage (now at St Thomas)

• Cornell
  – Sol Gruner (formerly at Princeton)
  – Steven Ealick
  – CHESS staff

• FSU
  – Donald Caspar

• UCSF
  – James Fraser
  – Andrew Vanbenschoten

• LBNL
  – Paul Adams
  – Nicholas Sauter
  – Aaron Brewster

• LANL
  – Tom Terwilliger

• Global Phasing
  – Gérard Bricogne

• Funding
  – NSF, Welch Foundation, Keck Foundation, LANL LDRD

• ECM attendance
  – John Helliwell
  – Brian McMahon
Further Reading