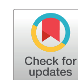
Available online at www.sciencedirect.com
ScienceDirect
Current Opinion in Structural Biology

Bringing diffuse X-ray scattering into focus☆



Michael E Wall[1], Alexander M Wolff[2,3] and James S Fraser[3]

X-ray crystallography is experiencing a renaissance as a method for probing the protein conformational ensemble. The inherent limitations of Bragg analysis, however, which only reveals the mean structure, have given way to a surge in interest in diffuse scattering, which is caused by structure variations. Diffuse scattering is present in all macromolecular crystallography experiments. Recent studies are shedding light on the origins of diffuse scattering in protein crystallography, and provide clues for leveraging diffuse scattering to model protein motions with atomic detail.

Addresses
[1] Computer, Computational, and Statistical Sciences Division, Los Alamos National Laboratory, Los Alamos, NM 87545, USA
[2] Graduate Group in Biophysics, University of California San Francisco, San Francisco, CA 94158, USA
[3] Department of Bioengineering and Therapeutic Sciences, University of California San Francisco, San Francisco, CA 94158, USA

Corresponding authors: Wall, Michael E (mewall@lanl.gov), Fraser, James S (jfraser@fraserlab.com)

Current Opinion in Structural Biology 2018, **50**:109–116

This review comes from a themed issue on **Sequences and topology**

Edited by **Joseph Marsh**

https://doi.org/10.1016/j.sbi.2018.01.009

0959-440X/© 2018 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

Introduction

With over 100 000 X-ray structures deposited in the wwPDB [1], improvements in data processing pipelines, and the advent of completely unattended data collection, it seems hard to imagine that there are any aspects of protein X-ray crystallography that remain to be optimized. However, only about half of the X-rays scattered by the crystalline sample are currently being analyzed — those in the Bragg peaks. The weaker, more smoothly varying features in diffraction images, known as diffuse scattering, are largely ignored by current practices. While the analysis of diffuse scattering is an established method in the fields of small molecule crystallography [2] and materials science [3], there are only very few foundational studies of diffuse scattering in macromolecular crystallography [4–15,16•,17,18]. However, the relative scarcity of diffuse scattering studies is poised to change as activity in the field has recently increased.

A small group of researchers (including MEW and JSF) met in 2014 to discuss the challenges and opportunities of investigating macromolecular diffuse scattering [20]. Our attention was drawn to several key developments in the field of macromolecular crystallography that motivated and enabled assessment of the diffuse signal. First, structural models obtained using traditional methods appear to be reaching a plateau in quality, as R factors remain relatively high compared to what can be achieved in small-molecule crystallography. The origin of this 'R-factor gap' is likely due to the underlying inadequacies of the structural models refined against crystallographic data [23]. These inadequacies can only be overcome if we can improve the modeling, including, for example, conformational heterogeneity (especially in data collected at room temperature [24]), solvation, and lattice imperfections that break the assumptions of 'perfect crystals' used in data reduction and refinement. Second, new detectors were enabling collection of data with lower noise, higher dynamic range, and highly localized signal. Third, new light sources were emerging with very bright, micro-focused beams (e.g. X-ray free-electron lasers). Collectively, these factors made us optimistic that diffuse scattering data both was needed and could be measured accurately enough to improve structural modeling. In early 2017, many of us met again to discuss the progress of the field with respect to each of these challenges identified in 2014 [25]. In this review, we provide our perspective on this progress and the status of the field, informed in part by our observations at that meeting and advances covered by Meisburger *et al.* [26••]. While there have been exciting developments in recent years, there are still major challenges ahead, include modeling atomic motions in protein crystals using diffuse scattering data with accuracy comparable to the Bragg analysis, and utilizing these models of protein motions to distinguish between competing biochemical mechanisms.

Data collection

Extraction of diffuse scattering data from conventional protein crystallography experiments is becoming straightforward thanks to the increased accessibility of photon-counting pixel array detectors (PADs, e.g. Pilatus detectors). These detectors have greater dynamic range and do not suffer from 'blooming' overloads that obscured diffuse signals near Bragg peaks on conventional charge-coupled device (CCD) detectors. (An early CCD detector

☆ Los Alamos National Laboratory Unclassified Release #LA-UR-17-30486.

was programmed to drain excess charge away from overflowing pixels to enable measurement of diffuse scattering data [18,27]; however, this feature was not implemented in commercial detectors.) Additionally the use of PADs has led to changes in data collection strategies, such as the use of fine phi angle scans, that facilitate analysis of Bragg peaks and diffuse features from the same set of images [19••]. A second major advance is the measurement of diffuse scattering using an X-ray free-electron laser (XFEL) in a serial femtosecond crystallography (SFX) experiment [28••]. Using an XFEL enables collection of radiation-damage-free room temperature data, as well the potential to examine time-resolved changes in the diffuse scattering signal.

Despite these advances in collection of diffuse scattering data, minimizing background scattering remains the most important obstacle to collecting high quality data. While it is possible to remove some background scattering during data processing, the cleanest separation requires one to remove scattering extraneous to the crystal during the experiment. Factors to consider during collection of single crystal datasets include the thickness and orientation of the loop (for relevant mounting schemes), the volume of liquid surrounding the crystal, and the amount of airspace between the crystal and the detector. Background air scatter can be also reduced by a Helium or vacuum path between sample and detector. Collection of SFX data adds additional complexity, as the injection stream and crystal size will vary. Ayyer *et al.* [28••] addressed this challenge by selecting only the frames with the strongest diffuse scattering signal, in which the size of the crystal was expected to be comparable to the width of the jet. As the landscape of sample delivery devices for SFX and conventional crystallography continues to evolve, mounted sample delivery on materials such as graphene [29•] provides a promising route for minimization of background scattering.

Data integration

Early studies of protein diffuse scattering focused on explaining features in individual diffraction images. The introduction of methods for three-dimensional diffuse data integration enabled quantitative validation of models of correlated motions [18]. Several approaches to 3D data integration now have been implemented [27,28••,30,31••,32]. These approaches differ in several key ways: the scaling of intensities when merging the data; the handling of intensities in the neighborhood of the Bragg peak; and the strategy for sampling of reciprocal space. In the *Lunus* software for diffuse scattering (https://github.com/mewall/lunus) we have chosen:

(1) To use the diffuse intensity itself to scale the diffuse data (as opposed to using the Bragg peaks, as in Ref. [31••]). This choice avoids artifacts due to potential differences in the way the Bragg and diffuse scattering vary with radiation damage and other confounding factors. The response of these signals to damage requires further study before a definitive scaling strategy can be chosen.
(2) To ignore or filter intensity values in regions where the variations are sharper than the 3D grid that will hold the integrated data. This can include masking halo intensities too close to a Bragg peak, and kernel-based image processing to remove Bragg peaks from diffraction images. These steps avoid the mixing of signal associated with sharp features into the signal associated with larger-scale, cloudy diffuse features. The sharply varying features (e.g. streaks) are an important component of the signal; however, to avoid artifacts in analysis, we prefer to measure them on a grid that is fine enough to resolve them [17]. If the sampling is finer than one measurement per integer Miller index, but still too coarse to resolve the halos, and if the halo intensity is nevertheless included (as in Ref. [31••]) then the measurements at integer Miller indices may be segregated from the rest of the data and analyzed separately.
(3) To sample the data on a grid that includes points at Miller indices (corresponding to where the Bragg peaks are located), and, for finer sampling, points corresponding to integer subdivisions of Miller indices. Sampling strategies that are not tied to the reciprocal lattice also are valid (as used in Refs. [28••,30]); however, on-lattice strategies enable leveraging of existing crystallographic analysis and modeling tools for diffuse scattering.

Efforts are now underway to decrease the burden of diffuse data integration and make diffuse data collection accessible for any protein crystallography lab. Recent algorithmic improvements have led to scalable, parallelized methods for real-time processing of single-crystal synchrotron data, decreasing the time required to extract a diffuse dataset from diffraction images. These improvements aim to keep pace with real-time analysis of Bragg data at high frame rates, such as those expected at LCLS-II and euXFEL. Initial tests mapped staphylococcal nuclease diffuse data onto a fine-grained reciprocal lattice, using two samples per Miller index [33•]. This implementation of the *Lunus* software is capable of processing thousands of diffraction images within a few minutes on a small computing cluster.

In addition to improving the scalability of diffuse scattering data processing, efforts are underway to create a push-button diffuse data processing pipeline. The *Sematura* pipeline (https://github.com/fraser-lab/diffuse_scattering) was inspired by the user-friendly environment provided by software for analyzing Bragg peaks, such as xia2 [34]. To ensure portability the project was built upon the CCTBX framework [35], with future work focusing on developing *Sematura* as a CCTBX module for ease of access.

Download English Version:

https://daneshyari.com/en/article/8319421

Download Persian Version:

https://daneshyari.com/article/8319421

Daneshyari.com