Reconstructing 3D protein crystal intensity from unoriented sparse diffraction patterns

Ti-Yen Lan Cornell University

Ewald-sphere slice



 $\mathbf{s_0}, \mathbf{s'}$: wave vectors λ : wavelength

$$|\mathbf{s_0}| = |\mathbf{s}'| = 1/\lambda$$

Ewald-sphere slice



$$\begin{split} \mathbf{s_0}, \mathbf{s'}: \text{ wave vectors} & |\mathbf{s_0}| = |\mathbf{s'}| = 1/\lambda \\ \lambda: \text{ wavelength} & |\mathbf{q}| = |\mathbf{s'} - \mathbf{s_0}| = 2\sin(\theta/2)/\lambda \end{split}$$

Effect of periodicity

real space



reciprocal space



Effect of periodicity

reciprocal space real space ρ |F|

Effect of periodicity

real space



reciprocal space



Crystal diffraction



$$I_{hkl} \propto J_0 \Delta t \ V_{cryst} |F(\mathbf{q}_{hkl})|^2$$

 J_0 : beam fluence Δt : exposure time V_{cryst} : crystal volume

Radiation damage



Axford et al., Acta Cryst. D (2012)

Outrunning damage — X-ray free electron laser



S. Boutet et al. Science 337, 362 (2012)

Time resolved serial crystallography



Schmidt, Advances in Condensed Matter Physics (2013)

Orientation determination - indexing



Orientation determination - indexing



- 1. embed Bragg peaks onto the Ewald sphere
- 2. find crystal unit cell parameters & orientation using Fourier methods (requires at least 20-30 resolvable peaks)

XFEL vs. synchrotron

XFEL	beamlines	synchrotron	beamlines
LCLS	7		
SACLA	3		
European XFEL	6	> 40 facilities	> 10 each
PAL-XFEL	3		
SwissFEL	2		

Synchrotron-based serial crystallography



Stellato et al., IUCrJ (2014)

Sparse data frame



Roadmap for sparse serial crystallography

target	rotational parameters	source
aligned crystal	1	table-top
aligned crystal	2	table-top
unaligned microcrystal	3	synchrotron

Single-axis experiment



Wierman et al., IUCrJ (2016)

The expand-maximize-compress (EMC) algorithm



data frames

The expand (E) step



Conditional probability calculation

$$P_{jk} \equiv P(K_{ik}|W(\mathbf{p}), \Omega_j) = \frac{w_j \prod_i \exp(-W_{ij}) W_{ij}^{K_{ik}}}{\sum_j w_j \prod_i \exp(-W_{ij}) W_{ij}^{K_{ik}}}$$



The maximize (M) step

$$W'_{ij} = \arg \max_{W'_{ij}} \sum_{k} P_{jk} (K_{ik} \log W'_{ij} - W'_{ij})$$
$$= \sum_{k} P_{jk} K_{ik} / \sum_{k} P_{jk}$$



updated slices

Maximize likelihood (M)

conditional probability over orientations of each frame

$$P_{jk}$$

The compress (C) step



$$W'(\mathbf{p}) = \frac{\sum_{ij} f(\mathbf{p} - R(\Omega_j) \cdot \mathbf{q}_i) W'_{ij}}{\sum_{ij} f(\mathbf{p} - R(\Omega_j) \cdot \mathbf{q}_i)}$$

reconstruction





Two-axis experiment



Lan et al., J. Appl. Cryst. (2017)

Peak statistics



EMC - local update scheme





reconstruction

reference

28

serial micro-crystallography at synchrotrons



Martin-Garcia et al., IUCrJ (2017)

Goal for the reconstruction

data frames resolution

Martin-Garcia et al.,	strongest 50/	ιΩÅ
IUCrJ (2017)	strongest 5%	2.0 A

This	work

the rejected 95% by comparable Martin-Garcia et al. value Complicating factors

- data reduction
- background scatter
- varying crystal size

peak identification



peak identification



peak identification



Powder pattern → unit cell parameters



• data reduction \checkmark

- background scatter
- varying crystal size

Modified EMC algorithm



The expand (E) step



Conditional probability calculation

$$P_{jk} \equiv P(K_{ik}|W(\mathbf{p}), \Omega_j, \phi_k, b_{ik}) = \frac{w_j \prod_i \exp(-\tilde{W}_{ijk}) \tilde{W}_{ijk}^{K_{ik}}}{\sum_j w_j \prod_i \exp(-\tilde{W}_{ijk}) \tilde{W}_{ijk}^{K_{ik}}}$$
$$\tilde{W}_{ijk} = b_{ik} + p_i \phi_k W_{ij}$$



The maximize (M) step

$$\phi'_{k} = \arg \max_{\phi'_{k}} \sum_{ij} P_{jk} \left[K_{ik} \log(b_{ik} + p_{i}\phi'_{k}W_{ij}) - (b_{ik} + p_{i}\phi'_{k}W_{ij}) \right]$$

conditional probability over orientations of each frame

$$P_{jk}$$

The compress (C) step $W'(\mathbf{p}) \bigoplus_{\text{intensity model}} \mathbf{Compress (C)} \bigoplus_{\substack{\mathsf{updated slices}}} W'_{ij}$

$$W'(\mathbf{p}) = \frac{\sum_{ij} f(\mathbf{p} - R(\Omega_j) \cdot \mathbf{q}_i) (\sum_k P_{jk} \phi_k) W'_{ij}}{\sum_{ij} f(\mathbf{p} - R(\Omega_j) \cdot \mathbf{q}_i) (\sum_k P_{jk} \phi_k)}$$

Summary

- Indexability does not limit the smallest usable crystal size in serial crystallography.
- Further reduction of background scatter will enable protein structure determination from much smaller crystals at synchrotron sources.

Acknowledgements

- Cornell Veit Elser, Sol Gruner, Jennifer Wierman, Mark Tate, Hugh Philipp
 - ASU Jose Martin-Garcia, Nadia Zatsepin
 - APS Robert Fischetti



