Selection and orientation of different particles in single particle imaging

Miklós Tegze* and Gábor Bortel

Institute for Solid State Physics and Optics, Wigner Research Centre for Physics, Hungarian Academy of Sciences, H-1525 Budapest, P. O. Box 49, Hungary

Abstract. The short pulses of X-ray free electron lasers can produce diffraction patterns with structural information before radiation damage destroys the particle. The particles are injected into the beam in random orientations and they should be identical. However, in real experimental conditions it is not always possible to have identical particles. In this paper we show that the correlation maximization method, developed earlier, is able to select identical particles from a mixture and find their orientations simultaneously.

Keywords: X-ray free electron lasers; Single particle diffraction; Single molecule diffraction; Orientation

*Corresponding author. e-mail: tegze.miklos@wigner.mta.hu, tel: +36-1-3922222, fax:+36-1-3922219
Introduction

Determination of the structure of small particles by diffraction of short X-ray pulses before radiation damage destroys the particle was suggested first by Solem (1986). Neutze et al. (2000) predicted that high-intensity ultrashort pulses from X-ray free electron lasers (XFELs) could provide a new approach to structural determinations with X-rays. Model calculations have shown that sufficient information might be collected before the sample is ionized and explodes due to Coulomb forces (Huldt et al., 2003; Jurek et al., 2004a, 2004b; Jurek and Faigel, 2008; Hau-Riege et al., 2005; Lorenz et al., 2012). Several XFELs were built since then and most of the technical problems were overcome. In recent publications (Miao et al., 1999; Chapman et al., 2006; Seibert et al., 2011; Martin et al., 2011, 2012; Loh et al., 2012) successful 2D reconstructions of particles from experimental data were presented. Diffraction of XFEL beam on nanocrystals demonstrated that 3D atomic resolution structure of a protein can be determined from such experiments (Chapman et al., 2011; Boutet et al., 2012). However, experimental atomic resolution 3D structures of non-crystalline particles have not been published yet.

In a single particle diffraction experiment particles are injected into the XFEL beam. When an X-ray pulse hits a particle, a diffraction image is recorded on a 2D detector placed downstream. Since the total number of photons diffracted on a single particle and collected by the detector is rather small (in the order of a few thousands for a large biomolecule), it is necessary to record many diffraction patterns on identical particles to obtain detailed structural information. The particles are injected into the beam usually in random orientation. In order to assemble a consistent 3D intensity distribution, the relative orientation of the particles need be determined for all patterns. Several methods were developed recently to obtain the relative orientations from the patterns themselves. Some of the methods use the information in a common line or arc along the intersection of pairs of patterns (Shneerson et al., 2008; Bortel and Tegze, 2011). Other methods use more sophisticated algorithms like generative topographic mapping (GTM, Fung et al., 2009; Moths and Ourmazd, 2011) or expectation maximization (EMC, Loh and Elser, 2009, Loh et al., 2010) to solve the orientation problem. There are also suggestions to determine the orientation from independent experimental data, the angular distribution of the particle fragments after the Coulomb explosion (Jurek and Faigel, 2013). In this paper we use the simple and efficient method based on correlation maximization (CM), developed earlier by the present authors (Tegze and Bortel, 2012). This method has the advantage that besides being much faster even for relatively small objects, the computational cost also scales favourably with the size of the particle (Tegze and Bortel, 2012). Once the orientations of the particles are found and a consistent 3D intensity distribution is constructed from the diffraction patterns, the structure can be solved using one of several phase retrieval methods (Fienup, 1982; Miao et al. 2001; Marchesini et al., 2003; Oszlányi and Sütő, 2004, 2005; Martin et al., 2012).

In real experimental conditions it is not always possible to have identical particles. Biological objects are not easy to separate, large protein molecules may have several conformations, or contamination may introduce different objects into the beam. If the particles are not identical, orientation methods may not work and construction of a consistent 3D dataset could be difficult. Schwander et al. (2010) has shown that different molecular conformations can be separated by the generative topographic mapping (GTM) method. Here we show that the computationally more efficient CM method is also capable to select identical objects from a mixture. We study three cases. In the first, we mix patterns of lysozyme and cytochrome molecules. These molecules have nearly the same size, but their structure is completely different. In the second, lysozyme and its complex with arginine (Arg) are mixed. The
presence of arginine changes the structure of the lysozyme molecule only slightly. In this case we wanted to test whether these small changes in the structure, together with the scattering contribution of the relatively small arginine are sufficient to distinguish between the molecules. Finally, we attempt to select the components from a mixture of all three kinds of molecules.

**Methods**

**Measurement simulation**

Synthetic diffraction patterns were calculated for randomly oriented molecules as described in our earlier work (Tegze and Bortel, 2012). The structures of the cytochrome, lysozyme and Arg-lysozyme complex molecules (Fig. 1) were taken from the Protein Data Bank (PDB IDs: 2XL6, 3LZT, and 3AGI, respectively). Poisson noise was introduced according to realistic experimental conditions. The most important parameters of the diffraction patterns are shown in Table 1.

**Orientation**

We used the CM method to orient the diffraction patterns. It is a much simplified variant of the EMC method of Loh and Elser (2009), where the expectation maximization step is replaced by a simple search for the orientation with the highest correlation. The algorithm was described earlier in detail (Tegze and Bortel, 2012), here we give only an outline. The diffraction patterns represent spherical sections (parts of the Ewald-sphere) of a 3D intensity distribution in reciprocal space. The centres of the patterns are always at the origin of the reciprocal space, but their orientations are unknown. The essential part of the algorithm is the comparison of the measured (in this case simulated) diffraction patterns to a 3D intensity distribution in different orientations (defined on a grid, see Table 1 and Tegze and Bortel, 2012) and finding the best fitting orientation for each pattern. Then these orientations are used to construct a new 3D distribution from the measured intensities. Starting from a random distribution, convergence is usually reached after a few iterations.

The similarity between the measured patterns and cuts from the 3D distribution is expressed in terms of the Pearson correlation (Rodgers and Nicewander, 1988; Tegze and Bortel, 2012). We found, that convergence is faster if at beginning of the iteration process when correlations are low, we use only a certain fraction of the patterns (the ones with a correlation higher than a certain limit) to build the new intensity distribution. We used the following condition to select the patterns contributing to the new distribution:

\[ c_{m}^{\text{max}} > c_{\text{limit}}, \]

where \( c_{\text{limit}} = \min(\text{median}(\{c_{i}^{\text{max}}\}), c_{0}^{\text{limit}}) \).

Here \( c_{m}^{\text{max}} \) is correlation of pattern \( m \) in its best-fitting orientation. In other words, the best-fitting half of the patterns and patterns with correlations above a certain limit are used to construct the new intensity distribution. At the beginning of the iteration process all correlations are low, so \( \text{median}(\{c_{i}^{\text{max}}\}) < c_{0}^{\text{limit}} \), therefore only half of the patterns contribute to the new distribution. When the correlations increase, more and more correlations are above \( c_{0}^{\text{limit}} \). Eventually, when reaching convergence, all diffraction patterns are included if \( c_{\text{limit}} \) is chosen properly. The chosen value of \( c_{\text{limit}} \) should be slightly below the expected correlation values between the noisy measurement patterns and the perfect (almost noiseless) solution of the 3D intensity distribution. This value can be estimated from
the noise (number of counts) in the diffraction pattern or found from the actual distribution of the $c_m^{max}$ values after convergence.

**Particle selection**

The same or a similar rule can be used to select diffraction patterns of one kind of particle out of a mixture. The mixture may contain two or more kind of particles in any distribution. During iteration, those patterns will tend to contribute to the 3D intensity distribution, which represent the kind of particles with majority in the distribution. The correlations of patterns representing other particles will usually remain below $c_{limit}^0$ and (if the number of majority particles are more than half of the number of all particles) do not contribute to the 3D distribution. When convergence is reached, the 3D intensity distribution will represent only the majority particles.

In the case when the number of the majority particles is smaller than half of the number of all particles $N$, (this can happen only with more than two kinds of particles) the above selection rule must be slightly modified. The median is be replaced with a selection of the best (having the largest correlations) $n$ patterns, $n/N < 0.5$.

**Results and Discussion**

First we test the ability of our method to separate diffraction images of cytochrome and lysozyme molecules. They have similar size, but completely different structure. The structures are known from conventional X-ray diffraction measurements and are shown in Fig. 1a and b. A mixture of diffraction images of 12000 cytochrome and 8000 lysozyme molecules in random orientations were calculated. Poisson noise corresponding to an X-ray intensity of $2 \cdot 10^{13}$ photons/pulse focused to a 100nm x 100nm spot was added. The important parameters of the simulated patterns are listed in Table 1. For $c_{limit}^0$ the value of 0.32 was chosen. The CM method solved the orientation problem for the majority (cytochrome) molecules. The evolution of the distribution of the correlation values $c_m^{max}$ is shown in Fig. 2. The blue line shows the value of $c_{limit}^0$ during the iteration process. After convergence, the distribution of $c_m^{max}$ splits into two parts. The final distribution of $c_m^{max}$ is shown in Fig. 3. Since we know from our simulation that which diffraction pattern belongs to which molecule, as verification we can also plot the distribution of correlation separately for the two kinds of molecules. Almost all diffraction patterns in the higher correlation part above $c_{limit}$ belong to the cytochrome molecules and contribute to the 3D intensity distribution, while those below $c_{limit}$ belong to the lysozyme molecules, and do not contribute to it. Thus the resulting 3D intensity distribution represents only the structure of the cytochrome molecules. The angular deviation of the orientations of the patterns from the true orientations of the molecules was less than 2 degrees, corresponding to the distance between angular grid points (Tegze and Bortel, 2012). Further refinement of the orientations can reduce the deviation by an order of magnitude (Tegze and Bortel, 2012).

In our second study we test the ability of the method to distinguish between diffraction patterns of molecules with only small differences in their structure. For this study we have chosen pure lysozyme and its complex with arginine. Their respective structures are shown in Fig. 1b and c. Again, a mixture of diffraction images of 12000 lysozyme and 8000 Arg-lysozyme molecules in random orientations were calculated. It can be seen from Fig. 4 and 5 that the CM method can separate these molecules as well after 20 iterations. The convergence is slower than in the case of cytochrome-lysozyme mixture, due to the smaller differences between the structures of the molecules. The number of iterations necessary to solve the
orientation problem for one component of a mixture depends not only on the intensity of the X-ray pulse, the number of patterns and the differences in structure but also on the relative number of the components. For example, when mixing 16000 lysozyme patterns with 4000 Arg-lysozyme ones (but keeping all other parameters the same), the CM algorithm found the orientations of the majority molecules after 7 iterations.

In the third study, equal number (12000) of diffraction patterns of cytochrome, lysozyme and Arg-lysozyme were mixed. Since each component is less than 50% of the mixture, the median in the expression of $c_{\text{limit}}$ is replaced by the best 30% of the patterns. Starting again from a random 3D intensity distribution, after 26 iterations the CM method finds the orientations of the diffraction patterns of one component, which happens to be the cytochrome (Fig. 6 and 7a). The low-correlation peak in Fig. 7a now contains a mixture of lysozyme and Arg-lysozyme. Now we can remove the patterns with correlation above $c_{\text{limit}}$, and continue with the iteration. After further 30 iterations, the orientations of the Arg-lysozyme patterns are found (Fig. 6 and 7b). After removing again the patterns with correlation above $c_{\text{limit}}$ and continuing the iteration process, the orientation problem of the remaining component (lysozyme) can also be solved (Fig 6 and 7c). Note that since there are only small differences in structure between lysozyme and Arg-lysozyme, convergence was reached after a few iterations.

Conclusions

In single particle diffraction experiments the particle beam may contain more than one kind of particle. We have shown that the correlation maximization algorithm can select one kind of particle out of a mixture of two or more components based on their noisy diffraction patterns without any prior knowledge of their structure or concentration in the mixture. The method can find the orientation of these particles with great accuracy making possible the compilation of a consistent 3D diffraction intensity distribution and subsequent solution of the structure. Separation of the diffraction patterns of particles with small structural differences was also possible.

Acknowledgments

We are grateful to G. Oszlányi for discussions. This work was supported by the Hungarian OTKA grants K81348 and NK105691 and by the Bolyai János Scholarship of the Hungarian Academy of Sciences to GB.

References

Bortel, G. and Tegze, M., 2011. Common arc method for diffraction pattern orientation. Acta Cryst. A 67, 533–543.

Boutet, S., Lomb, L., Williams, G.J., Barends, T.R.M., Aquila A. et al., 2012. High-Resolution Protein Structure Determination by Serial Femtosecond Crystallography. Science 337, 362-364.

Chapman, H.N., Barty, A., Bogan, M.J., Boutet, S., Frank, M. et al., 2006. Femtosecond diffractive imaging with a soft-X-ray free-electron laser. Nature Phys. 2, 839–843.

Chapman, H. N. Fromme, P., Barty, A., White, T.A., Kirian, R.A. et al., 2011. Femtosecond X-ray protein nanocrystallography. Nature 470, 73-77.
Fienup, J.R., 1982. Phase retrieval algorithms: A comparison. Appl. Opt. 21, 2758–2769.

Fung, R., Shneerson, V., Saldin, D. K. and Ourmazd A., 2009. Structure from fleeting illumination of faint spinning objects in flight. Nature Phys. 5, 64–67.

Hau-Riege, S. P., London, R. A., Huldt, G. and Chapman, H. N., 2005. Pulse requirements for X-ray diffraction imaging of single biological molecules. Phys. Rev. E 71, 061919.

Huldt, G., Szoke, A. and Hajdu, J., 2003. Diffraction imaging of single particles and biomolecules. J. Struct. Biol. 144, 219–227.

Jurek, Z. and Faigel, G., 2008. The effect of tamper layer on the explosion dynamics of atom clusters. Eur. Phys. J. D 50, 35–43.

Jurek, Z. and Faigel, G., 2013. Orienting single-molecule diffraction patterns from XFELs using heavy-metal explosion fragments. EPL 101, 16007.

Jurek, Z., Faigel, G. and Tegze M., 2004a. Dynamics in a cluster under the influence of intense femtosecond hard X-ray pulses. Eur. Phys. J. D 29, 217–229

Jurek, Z., Oszlányi G., and Faigel G., 2004b. Imaging atom clusters by hard X-ray free-electron lasers. Europhys. Lett., 65, 491–497.

Loh, N.D., Bogan, M.J., Elser, V., Barty, A., Boutet, S. et al., 2010. Cryptotomography: Reconstructing 3D Fourier Intensities from Randomly Oriented Single-Shot Diffraction Patterns. Phys. Rev. Lett. 104, 225501.

Loh, N. D. and Elser V., 2009. Reconstruction algorithm for single-particle diffraction imaging. Phys. Rev. E 80, 026705.

Loh, N.D., Hampton C.Y., Martin A.V., Starodub D., Sierra R.G. et al., 2012. Fractal morphology, imaging and mass spectrometry of single aerosol particles in flight. Nature 486, 513-517.

Lorenz, U., Kabachnik, N.M., Weckert, E. and Vartanyants, I.A., 2012. Impact of ultrafast electronic damage in single-particle x-ray imaging experiments Phys. Rev. E 86, 051911.

Marchesini, S., He, H., Chapman, H.N., Hau-Riege, S.P., Noy, A., et al., 2003. X-ray image reconstruction from a diffraction pattern alone. Phys. Rev. B 68, 140101(R)

Martin, A.V., Andreasson, J., Aquila, A., Bajt, S., Barends, T.R.M. et al., 2011. Single particle imaging with soft X-rays at the Linac Coherent Light Source, in: Advances in Tschentscher, T. and Cocco, D. (Eds.), X-ray Free-Electron Lasers: Radiation Schemes, X-ray Optics and Instrumentation (Prague, Czech Republic, 18 April 2011). Proc. of SPIE vol. 8078, 807809.

Martin, A.V., Wang, F., Loh, N.D. Ekeberg, T., Maia, F.R.N.C. et al., 2012. Noise-robust coherent diffractive imaging with a single diffraction pattern. Optics Express 20, 16650.

Miao, J., Charalambous, P., Kirz, J. and Sayre, D., 1999. Extending the methodology of X-ray crystallography to allow imaging of micrometer-sized non-crystalline specimens. Nature 400, 342-344.
Miao, J., Hodgson, K.O. and Sayre D., 2001. An approach to three-dimensional structures of biomolecules by using single-molecule diffraction images. Proc. Natl. Acad. Sci. U.S.A. 98, 6641–6645.

Moths, B. and Ourmazd, A., 2011. Bayesian algorithms for recovering structure from single-particle diffraction snapshots of unknown orientation: a comparison. Acta Cryst. A 67, 481–486.

Neutze, R., Wouts, R., van der Spoel, D., Weckert, E. and Hajdu, J., 2000. Potential for biomolecular imaging with femtosecond X-ray pulses. Nature 406, 752–757.

Oszlányi, G. and Sütö, A., 2004, Ab initio structure solution by charge flipping. Acta Cryst. A 60, 134-141.

Oszlányi, G. and Sütö, A., 2005. Ab initio structure solution by charge flipping. II. Use of weak reflections. Acta Cryst. A 61, 147-152.

Rodgers, J.L. and Nicewander, W.A., 1988. Thirteen ways to look at the correlation coefficient. The American Statistician 42, 59–66.

Protein Data Bank, http://www.pdb.org/pdb.

Schwander, P., Fung, R., Phillips, G.N.Jr. and Ourmazd, A., 2010. Mapping the conformations of biological assemblies. New J. Phys. 12, 035007

Seibert, M.M., Ekeberg, T., Maia, F.R.N.C., Svenda, M., Andreasson, J. et al., 2011. Single mimivirus particles intercepted and imaged with an X-ray laser. Nature 470, 78-81.

Shneerson, V.L., Ourmazd, A. and Saldin, D.K., 2008. Crystallography without crystals. I. The common-line method for assembling a three-dimensional diffraction volume from single-particle scattering. Acta Crystallogr. A 64, 303-315.

Solem, J. C., 1986. Imaging biological specimens with high-intensity soft X-rays. J. Opt. Soc. Am. B 3, 1551-1565.

Tegze, M. and Bortel, G., 2012. Atomic structure of a single large biomolecule from diffraction patterns of random orientations. J. Struct. Biol. 179, 41–45.
Figure 1. Structure of the molecules used in this study: cytochrome (a), lysozyme (b) and arginine-lysozyme complex (c).
Figure 2. Evolution of the distribution of correlation maxima for the 3:2 mixture of cytochrome and lysozyme molecules. The blue line indicates the value of $c_{\text{limit}}$.

Figure 3. Distribution of correlation maxima for the 3:2 mixture of cytochrome and lysozyme molecules after convergence. The red (dotted) and blue (dashed) lines indicate the contribution of patterns of cytochrome and lysozyme, respectively. Since the separation is almost perfect, the black (solid) line indicating the full distribution is overlaid by the red (dotted) and blue (dashed) lines.
Figure 4. Evolution of the distribution of correlation maxima for the 3:2 mixture of lysozyme and Arg-lysozyme molecules. The blue line indicates the value of $c_{\text{limit}}$.

Figure 5. Distribution of correlation maxima for the 3:2 mixture of lysozyme and Arg-lysozyme molecules after convergence. The black (solid) line indicates the full distribution and the blue (dashed) and green (dash-dotted) lines indicate the contribution of patterns of lysozyme and Arg-lysozyme molecules, respectively.
Figure 6. Evolution of the distribution of correlation maxima for the 1:1:1 mixture of cytochrome, lysozyme and Arg-lysozyme molecules. Patterns of already oriented components were removed from the mixture after iterations 26 and 56. The blue line indicates the value of $c_{\text{limit}}$.

Figure 7. Distribution of correlation maxima for the 1:1:1 mixture of cytochrome, lysozyme and Arg-lysozyme molecules after iterations 26 (a), 56 (b) and 61 (c). The black (solid) line indicates the full distribution and the red (dotted), blue (dashed) and green (dash-dotted) lines indicate the contribution of patterns of cytochrome, lysozyme and Arg-lysozyme molecules, respectively.
### Table 1. Parameters of simulated diffraction patterns and grids.

| Parameter                                      | lysozyme | Arg-lysozyme | cytochrome |
|------------------------------------------------|----------|--------------|------------|
| Diameter of the molecule, $D$                  | 51.2 Å   | 53.5 Å       | 56.3 Å     |
| Wavelength of the radiation, $\lambda$        | 1 Å      |              |            |
| Pulse fluence                                  |          |              | $2 \cdot 10^{13}$ photons/(0.1μm×0.1μm) |
| Minimum scattering angle, $\theta_{\text{min}}$|          |              | 1°         |
| Maximum scattering angle, $\theta_{\text{max}}$|          |              | 24°        |
| Pattern pixel size (polar × azimuthal), $\Delta \theta \times \Delta \phi$ | 1°× 2°  |              |            |
| Average total photon counts in pattern         | 1702     | 1563         | 2149       |
| Average photon counts in outer pixels          | 0.105    | 0.100        | 0.123      |
| Average photon counts in outer Shannon–Nyquist pixels | 0.041    | 0.035        | 0.039      |
| (Huldt et al., 2003) of size ($\lambda/2D)^2$ |          |              |            |
| Number ($N_{\theta}$) and spacing of grid points in the ($\Theta,\Phi$) orientation subspace (Tegze and Bortel, 2012) | 5292 |            | ~3°         |
| Number of grid points (voxels) in the reciprocal space, $N_f$ | 129³ = 2,146,689 |          |            |