# Extraction of Fast Changes in the Structure of a Disordered Ensemble of Photoexcited Biomolecules

H.-C. Poon,[1] M. Schmidt,[1] and D. K. Saldin[1]

[1]*Department of Physics, University of Wisconsin-Milwaukee, Milwaukee, WI 53211*
*E-mail addresses: phyphc@uwm.edu, m-schmidt@uwm.edu,*
*dksaldin@uwm.edu (Corresponding Author)*

## Abstract

The method of time-resolved crystallography [1] has given us insight into rapid variations of the structure of biomolecules via pump-probe experiments involving photoexcitation by a laser followed very quickly by measurement of x-ray diffraction patterns. Changes in the structure are then found by established methods [2, 3]. We propose here how these structural changes can be determined from individual biomolecules in solution.

## I.   INTRODUCTION

The x-ray free electron laser (XFEL) is a new instrument which promises to revolutionize our study of the atomic architecture of matter [4, 5]. The brightness of the x-rays produced by this instrument is some 10 billion times greater than any existing x-ray source (including present-day synchrotrons). This allows the possibility of measuring signals from scattered x-rays of even large single molecules, like proteins. The traditional limitation of x-ray flux for fragile biomolecles can be circumvented completely due to the fact that this very bright radiation is delivered in ultra-short pulses [6–8]. Although the molecules under study will undoubtedly suffer catastrophic radiation damage, the shortness of the pulse enables a signal to be measured from the particle before its disintegration. This enables structure determination of reprodicible biomolecules by essentially an unlimited x-ray flux, and is in one sense a complete solution of the radiation damage problem. It should also be possible to combine the ultra brightness [9] of the radiation with the ultra shortness of its duration [10] to enable the gathering of information never before possible, for example the changes in the structure of an uncystallized biomolecule as a result of some stimulus, such as photoexcitation, as a function of time since the photoexcitation. Since this time can be very short, the possibility then exists of experimentally following the course of rapid chemical reactions of such uncrystallized biomolecules, as with crystallized ones [10] by the technique of time-resolved diffraction [1, 11, 12]. This may allow for the first time the study of biochemical rections of molecules in aqueous solution in which they occur in nature. An idea proposed for sample delivery of hydrated molecules to an XFEL beam is to inject a continuous stream of a solutions containing the molecules into the sample chamber [13–15]. The incident x-rays then scatter off the protein solution. The design specification of the LCLS is to produce an x-ray beam of perhaps 0.1 microns in diameter at the sample. To maximize the protein scattering from such a solution stream (and to minimize scattering by the aqeous solution) it would be best to use a solution as concentrated as possible. Photoactive yellow protein (PYP) is a popular (15 kDa) protein for time-resolved structural studies. The reason is that this protein suffers a significant and reproducible structural change on illumination by laser light [16–19], where a chromophore swings a significant distance outward from the center of the molecule, and the nearby ARG 52 residue moves to accommodate the new position of the chromophore. PYP can be concentrated to 150 mg/ml or higher

[20]. This corresponds to about 10 mmol/L = 10 mole/$m^3$=150 mg/ml. Thus the volume of solution illuminated by the XFEL beam can be estimated by multiplying an estimated cross-sectional area of the beam based on the design estimate of 0.1 micron diameter, with an estimate of the thickness of the continuous stream of solvent. Our best estimate of this volume is thus $10^{-14} \times 3 \times 10^{-7} \sim 3 \times 10^{-21}$ $m^3$, where for a thickness we estimated the smallest reported diameter of a solvent droplet [**?** ]. Multiplying this by the number of moles per $m^3$, we estimate the number of moles illuminated by the XFEL beam to be $\sim 3 \times 10^{-20}$. Multiplying this by Avogadro's number, $\sim 6 \times 10^{23}$ molecules per mole, suggests a typical illuminated volume of solvent probably contains $\sim 10,000$ molecules of PYP at their optimum concentration. Even with a reduced concetration of 10 mg/ml, as used in small angle x-ray scattering (SAXS) work and a 500 kDa protein, when this number is reduced by a factor of 500 the number of molecules in a typical lluminated volume would still be expected to be about 20, *still a far cry from illumination of a single molecule per measured diffraction pattern* as assumed by many algorithms for the reconstruction of the real-space structure of a single molecule. Of course, methods have been developed [21] for automatically selecting single-particle X-ray diffraction snapshots from the entire collection of measured diffraction patterns for analysis by such techniques, which may be of particular use for large molecular ensembles or viruses. However, it should be pointed out that such methods depend on the rejection of diffraction patterns from multiple particles. The rejection of data from multiple particles reduces the particle/solvent ratio and thereby presumably reduces the signal to noise.

## II. STRUCTURAL INFORMATION FROM DISORDERED ENSEMBLES OF MOLECULES

The possibility of extracting structural information about an ensemble of molecules in a disordered ensemble as in solution from scattered x-ray signals was recognized as far back as the late 1970's [22]. The idea proposed was that, if the x-ray pulses were shorter than the rotational diffusion time of the molecules, the measured signal from a single x-ray pulse is of the ensemble of particles frozen in space and time. It was shown that, in this case, the average angular correlation functions of the scattered intensities (as defined below) are charateristic of the 3D structues of the individual biomolecules. This gives rise

to the exciting possibility of recovering the structure of a biomolecule from an a disordered ensemble of molecules as expected in their functional state in nature, rather than an ordered collection of the molecules as found in a crystal. A realistic possibility of advancing this idea is provided by the advent of the XFEL [23]. If the average over a set of diffraction patterns of angular pair correlation function between the intensities of a couple of resolution rings $q$ and $q'$ diffraction patterns from such a disordered entity is defined by

$$C_2(q, q', \Delta\phi) = < I(q, \phi)I(q', \phi + \Delta\phi) >_{DP} . \tag{1}$$

where $< \cdots >_{DP}$ represents an average over all the diffraction patterns, DP. Note that the orientational averaging of the particles implied by the reasonable assumption that all molecular orientations are equally likely suggests that the LHS of the above equation will be independent of the value of $\phi$ chosen on the RHS. If one makes the other reasonable assumptions of complete translational disorder (as expected for a dilute ensemble of particles, it is possible to show [23] that

$$C_2(q, q', \Delta\phi) = \sum_l B_l(q, q')P_l(\sin[\theta(q)]\sin[\theta(q')] + \cos[\theta(q)]\cos[\theta(q')]\cos[\Delta\phi]) \tag{2}$$

where

$$\theta(q) = \pi/2 - sin^{-1}(q/2\kappa) \tag{3}$$

and $\kappa$ is the wavenumber of the incident radiation. The angle $\theta(q)$ takes account of the curvature of the Ewald sphere, and is almost equal to $\pi/2$ radians for a small resolution ring or a flat Ewald sphere. In this expression, also

$$B_l(q, q') = \sum_m I_{lm}(q)I^*_{lm}(q') \tag{4}$$

where $I_{lm}(q)$ are a set of sperical harmonic expansion coefficients of the 3D diffraction volume of a single particle, and $P_l$ is a Legendre function. Recent developments of iterative phasing algorithms suggest that if an oversampled 3D diffraction volume were found via the coefficients $I_{lm}(q)$, the real-space structure of the particle may be recovered by means of an iterative phasing algorithm.

Although the quantities $B_l(q, q')$ may be found quite straightforwardly by inverting Eq.(2), finding the $I_{lm}(q)$ coefficients in general from these quantities are far from easy [24], although where the particle has a known symmetry, for example in the case of an icosahedral [26] or helical virus [25] this may be possible by exploitation of the known symmetry.

Another circumstance in which a solution may be possible is when trying to recover a small change in a structure from a known one, as in the case of pump-probe experiments on photoexcited biomolecules in studies of time-resolved structural changes. Such an experiment is of great significance not only because it allows the study of the course of fast chemical reactions, but also because they truly exploit not only the extreme brightness, but also the fast time structure of XFEL pulses.

A full calculation that shows this capability will be reported on in a future paper. We illustrate here the sensitivity of the quantities $B_l(q, q')$, to expected torsional angle changes on photoexcitation. In the case of PYP, earlier time-resolved studies have established that in 2ms after photoexcitation the primary change in the structure of PYP may be regarded as a cis-trans isomerization of its chromophore about its C2-C3 axis, and a change of the $\chi_3$ torsional angle of the side chain of the ARG 52 residue of PYP to make room for the structural changes due to the chromophore isomerization.

## III. PROPOSED TIME-RESOLVED EXPERIMENT ON DISSOLVED MOLECULES IN RANDOM ORIENTATIONS

Imagine an experiment where a continuous stream of a solution of photoexcitable molecules are injected into an XFEL sample chamber in the usual manner of so-called "diffract and destroy" experiments [4], but where a short distance $L$ before the intersection of the x-rays from the XFEL, the molecules are photoexcited by a powerful laser. If $v$ is the speed of the solvent stream (typically 10 m/s), the molecules will be illuminated by the laser a time $\Delta t = L/v$ before it is interrogated by the x-ray beam. Since the time $\Delta t$ is controllable by varying the distance $L$, this pump-probe experiment allows exquisite control of the time delay after photoexcitation. It is envisaged that a large number of diffraction patterns may be measured from different regions of the continuous solvent for a given time delay. It has been shown that structural information of the photoexcited molecules in the solution resides in the ensemble of such diffaction patterns.
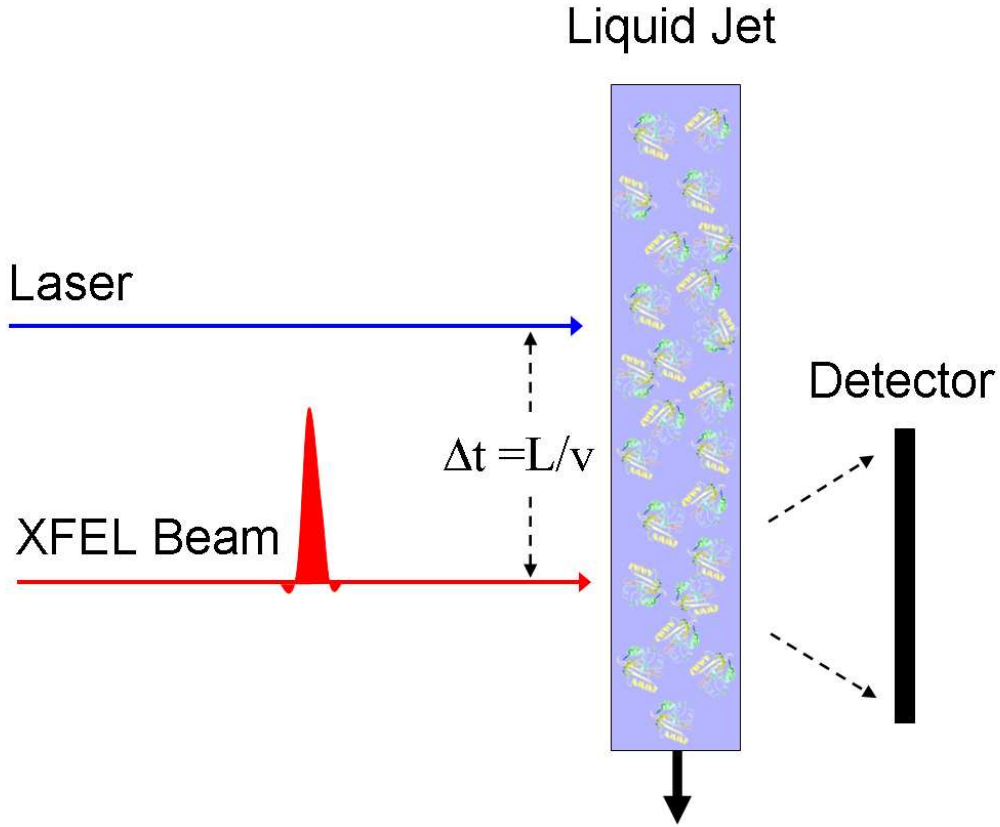
FIG. 1: Schematic diagram of the proposed pump-probe experiment. A continuous stream of a solution of identical molecules in random orientations is introduced to the path of the x-ray beam after photoexcitation by a laser a controllable time before interrogation by the x-rays. The time delay $\Delta t$ between photoexcitation and x-ray incidence is controlled by varying the distance $L$ between the positions of photoexcitation and x-ray incidence.

## IV. MODEL CALCULATION

Previous studies by time-resolved crstallography have established that the primary structural change of PYP 2ms after photoexcitation is a cis-trans isomerization of the chromophore which is a result of a $\theta$=180 degree rotation of the head of the chromophore about its C2-C3 bond as well as a 77 degree change of the $\chi_3$ torsional angle of the side chain of the ARG 52 residue of PYP [16–19]. Ignoring further small relaxations of other nearby atoms of the structure, such structural changes can be parametrized to a good approximation by just two torsional angles, $\theta$ and $\chi_3$. The two panels of Fig. 2 illustrate the dark and photoecited
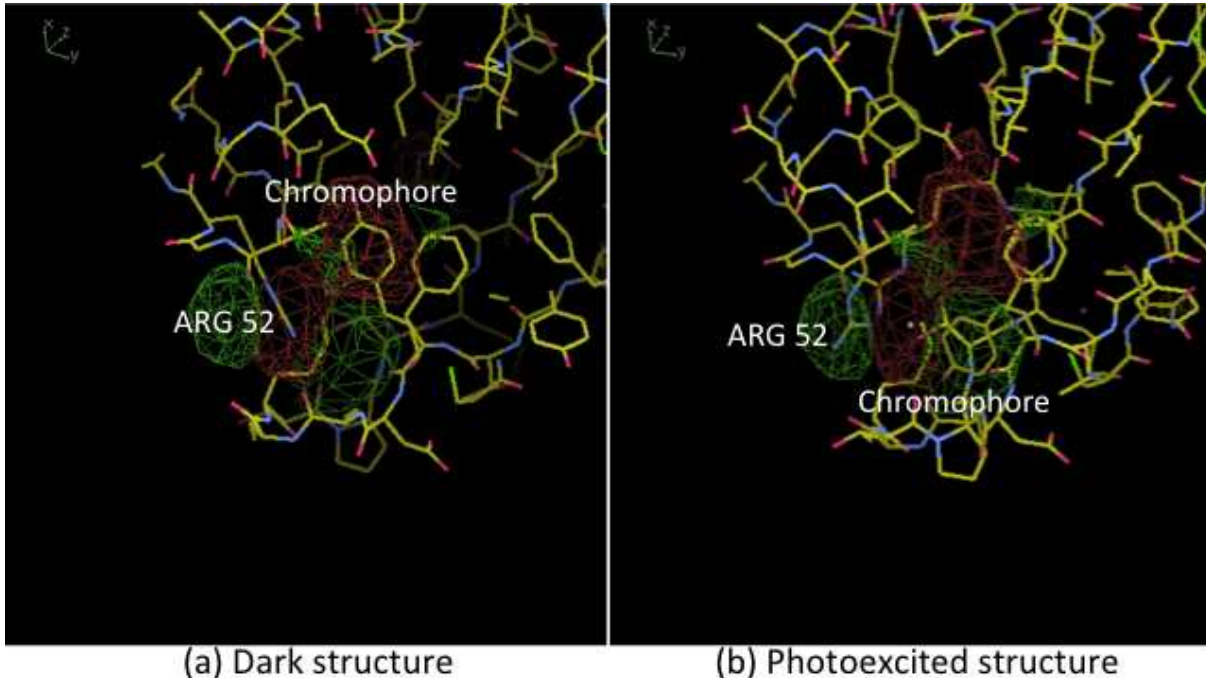
FIG. 2: Structural changes on photoexcitation. (a) The stick figure represents the dark structure, in which the chromophore and nearby ARG 52 redidue are associated with brown lobes of negative difference electron density, (b) the stick figure represents the structure 2ms after photoexcitation, where the chromophore is are seen to have swung outward from the center of the molecule by cis-trans isomerization, and the ARG 52 residue by a rotation of the side chain about its $\chi_3$ torsional angle. Both strucrural groups are now found in green lobes of positive difference electron density.

structures in the part of the molecule undergoing structural changes.

Starting from the atomic coordinate data for the dark strucrture of PYP from the Protein Data Bank entry 2PHY, we calculated a range of hypothetical excited state structures corresponding to $\theta$ and $\chi_3$ torsional angles at 5 degree intervals from the dark structure values of $\theta=0$ and $\chi_3=0$. For each of the hypothetical photoexcited structures, we calculated its structure factors from the formula:

$$A(\mathbf{q}) = \sum_j f_j \exp\left(i\mathbf{q} \cdot \mathbf{r}_j\right) \tag{5}$$

and hence its expected three-dimensional diffraction volume by evaluating $I(\mathbf{q}) = |A(\mathbf{q})|^2$. Then, using Gaussian quadrature on shells of radius $q$ we calculated the spherical harmonic expansion coefficients $I_{lm}(q)$ of this 3D diffraction volume. These were used to calculate the experimentally accessible quantities $B_l(q, q')$ via (4) for $l$ ranging from 0 to 25, and ranges of

$q$ and $q'$ from 0 to $q_{max} = 2\pi/5$ Å$^{-1}$ (i.e. up to 5 Å resolution) in intervals of $\Delta q = q_{max}/50$ for each of these hypothetical structures. These quantities were compared with the values of the same quantities for a reference structure corresponding to $\theta$=180 degrees and $\chi_3$=77 degrees (assumed to be the correct structure 2ms after photoexcitation) via a reliability factor (or R-factor) defined by

$$R(\theta, \chi_3) = \frac{\sum_{l,q,q'} |B_l^{model}(q,q') - B_l^{trial}(q,q',\theta,\chi_3))|}{\sum_{l,q,q'} |B_l^{model}(q,q')|} \tag{6}$$

where $\theta$ and $\chi_3$ are the torsional angles above of the chromophore and the $\chi_3$ side-chain torsional angle of the ARG 52 residue. The resulting 2D contour map of R as a function of $\theta$ and $\chi_3$ is shown in Fig. 3.

## V.  DISCUSSION AND CONCLUSIONS

The contour map reveals an unambigous minimum close to the assumed 2ms structure. Although there are about 15 atoms whose positions are different between the dark and 2ms photoexcited structures, the principle that bond lengths are unlikey to change on photoexcitation suggests that structural changes may be well parametrized by changes in only torsional angles. This parametrization allows the efficient determination of the changes of the positions of some 15 atoms between the dark and photoexcited structures. In the present case of just two varied structural parameters the R-factor contour map is seen to be simple enough that even a simple gradient descent algorithm that starts at the dark structure ($\theta = \chi_3$=0) is likely to find the correct photoexcited structure. It should be noted that such a parametrization may even be of use in established methods of time-resolved crystallography in order to determine directly even changes in atomic positions of a photoexcited structure rather that via the the fitting of a deduced difference electron density map with an atomic model. Indeed an input to our calculation is the PDB file of the dark structure and the output a PDB file of the excited structure. Of course, for photoexcited structures that differ from a dark structure by more structural parameters, a more sophisticated search algorithm like simulated annealing [27] may be necessary to find a global minimum.

The primary conclusion from this work is that a determination of a fast structural change of a molecule on photoexcitation may not require the formation of a crystal. What we have demonstrated is that quantities measurable in a pump-probe experiment on dark and pho-
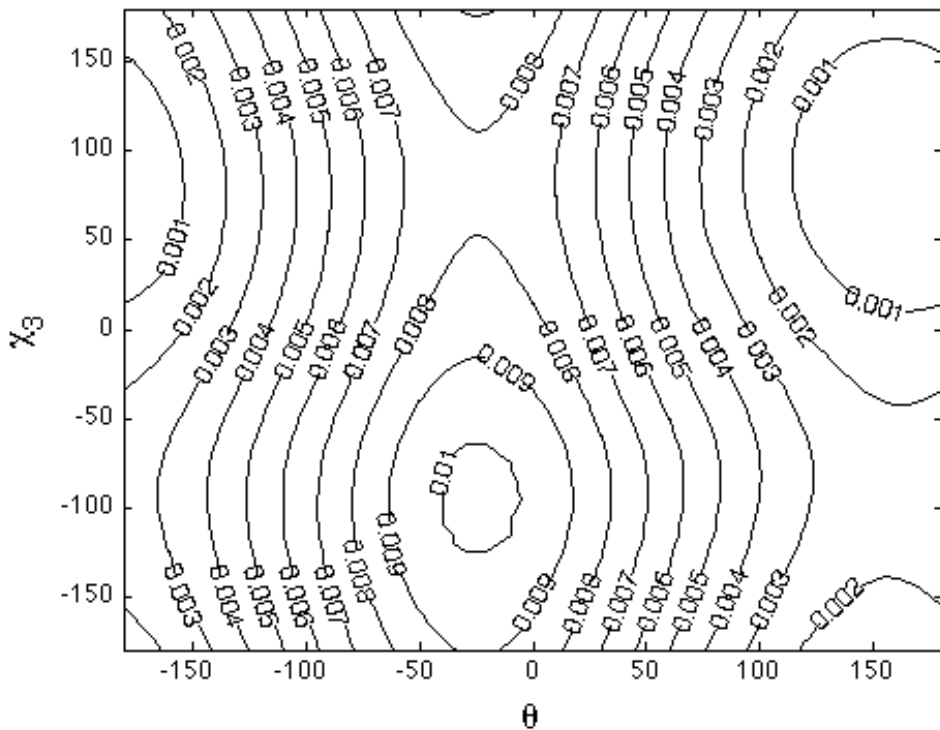
FIG. 3: Contour map of the R-factor comparing the quantites $B_l(q, q')$ calculated from an assumed excited structure (extractable from an ensemble of diffraction patterns of random molecular orientations of the excited structure) versus the same quantities calculated for trial values of the torsional angles $\theta$ and $\chi_3$ of the chromophore and ARG 52 residue, respectively. The minimum of this plot is consistent with the assumed photoexcited structure at $\theta = \pm 180$ degrees and $\chi_3 = 77$ degrees.

toexcited biomolecules in a disordered ensemble in solution such as may be particularly accessible nowadays with the advent of the x-ray free electron laser (XFEL), contain structural information from which small changes in the structure may be deduced. The initial demonstration is only for a case that can be characterized by the changes of just two structural parameters, namely two torsional angles of a protein. A more complete demonstration of the method will have to await the development of techniques for extracting from the same measurable data possible structural changes in any part of the molecule. Nevertheless, we feel the present demonstration is important in that it demonstrates that quantities measurable from disordered ensembles of biomolecules in solution may hold the key to unlocking such fast structural changes in uncrystallized biomolecules in solution.

## VI. ACKNOWLEDGEMENTS

[1] Moffat, K., Ann. Rev. Biophysics and Biophysical Chemistry **18**, 309-332 (1989).

[2] Schmidt, M., Rajagopal, S., Ren, Z. and Moffat, K., Biophys. J. **84**, 2112-2129 (2003).

[3] Schmidt, M. in: Ultrashort Laser Pulses in Medicine and Biology (Springer, Berlin, 2008)

[4] Neutze, R., Wouts, R., van der Spoel, D., Weckert, D. and Hajdu, J., Nature (London) **406**, 752-757 (2000).

[5] Neutze, R. and Moffat, K., Currr. Opin. Structural Biology **22**, 651-659 (2012).

[6] A. Cho, Science **330**, 470-471 (2010).

[7] Chapman, H. N., Fromme, P., Barty, A., White, T. A., Kirian, R. A., Aquila, A., Hunter, M. S., Schulz, J., DePonte, D. P., Weierstall, U., Doak, R. B., Maia, F. R. N. C., Martin, A. V., Schlichting, I., Lomb, L., Coppola, N., Shoeman, R. L., Epp, S. W., Hartmann, R., Rolles, D., Rudenko, A., Foucar, L., Kimmel, N., Weidenspointner, G., Holl, P., Liang, M. N., Barthelmess, M., Caleman, C., Boutet, S., Bogan, M. J., Krzywinski, J., Bostedt, C., Bajt, S., Gumprecht, L., Rudek, B., Erk, B., Schmidt, C., Homke, A., Reich, C., Pietschner, D., Struder, L., Hauser, G., Gorke, H., Ullrich, J., Herrmann, S., Schaller, G., Schopper, F., Soltau, H., Kuhnel, K. U., Messerschmidt, M., Bozek, J. D., Hau-Riege, S. P., Frank, M., Hampton, C. Y., Sierra, R. G., Starodub, D., Williams, G. J., Hajdu, J., Timneanu, N., Seibert, M. M., Andreasson, J., Rocker, A., Jonsson, O., Svenda, M., Stern, S., Nass, K., Andritschke, R., Schroter, C. D., Krasniqi, F., Bott, M., Schmidt, K. E., Wang, X. Y., Grotjohann, I., Holton, J. M., Barends, T. R. M., Neutze, R., Marchesini, S., Fromme, R., Schorb, S., Rupp, D., Adolph, M., Gorkhover, T., Andersson, I., Hirsemann, H., Potdevin, G., Graafsma, H., Nilsson, B., and Spence, J. C. H., Nature **470**, 73-81 (2011).

[8] Boutet, S., Lomb L., Williams, G. J., Barends, T. R. M., Aquila, A., Doak, R. B., Weierstall, U. DePonte, D. P., Steinbrener J., Shoeman, R. L., Messerschmidt, M., Barty, A., White, T. A., Kassemeyer, S., Kirian, R. A., Seibert, M.M., Montanez, P. A., Kenney, C., Herbst,

R., Hart, P., Pines, J., Haller, G., Gruner S.M. Philpp, H. T., Tate, M. W., Hromalik, M., Koerner, L. J., van Bakel, N., Morse, J., Ghonsalves, W., Arnlund, D., Bogan, M. J., Caleman, C., Fromme, R., Hampton, C. Y., Hunter, M. S., Johansson, L. C., Katona, G., Kupitz, C., Liang, M. N, Martin, A. V., Nass, K., Redecke, L., Stellato, F., Timneanu, N., Wang, D. J., Zatsepin, N. A., Schafer, D., Defever, J., Neutze, R., Fromme, P., Spence, J. C. H., Chapman, H. N., Schlichting, I., Science **337**, 362-364 (2012).

[9] We adopt here the definition in the paper: D. M. Mills, J. R. Heliwell, Å. Kwick, T. Ohta, I. A. Robinson, A. Authier, J. Synchrotron Rad. **12**, 385 (2005).

[10] Aquila, A., Hunter, M. S., Doak, R. B., Kirian, R. A., Fromme, P., White, T. A., Andreasson, J., Arnlund, D., Bajt, S., Barends, T. R., Barthelmess, M., Bogan, M. J., Bostedt, C., Bottin, H., Bozek, J. D., Caleman, C., Coppola, N., Davidsson, J., DePonte, D. P., Elser, V., Epp, S. W., Erk, B., Fleckenstein, H., Foucar, L., Frank, M., Fromme, R., Graafsma, H., Grotjohann, I., Gumprecht, L., Hajdu, J., Hampton, C. Y., Hartmann, A., Hartmann, R., Hau-Riege, S., Hauser, G., Hirsemann, H., Holl, P., Holton, J. M., Homke, A., Johansson, L., Kimmel, N., Kassemeyer, S., Krasniqi, F., Kuhnel, K. U., Liang, M., Lomb, L., Malmerberg, E., Marchesini, S., Martin, A. V., Maia, F. R., Messerschmidt, M., Nass, K., Reich, C., Neutze, R., Rolles, D., Rudek, B., Rudenko, A., Schlichting, I., Schmidt, C., Schmidt, K. E., Schulz, J., Seibert, M. M., Shoeman, R. L., Sierra, R., Soltau, H., Starodub, D., Stellato, F., Stern, S., Struder, L., Timneanu, N., Ullrich, J., Wang, X., Williams, G. J., Weidenspointner, G., Weierstall, U., Wunderer, C., Barty, A., Spence, J. C., and Chapman, H. N., Opt. Exp. **20**, 2706-2716 (2012).

[11] Moffat, K., Chem. Rev. **101**, 1569-1581 (2001).

[12] Schmidt, M., Ihee, H., Pahl, R. and Srajer, V., Meth. Mol. Biol. **305**, 115-154 (2005).

[13] Weierstall, U., Spence, J. C., and Doak, R. B., Rev. Sci. Instrum. **83**, 035108 (2012).

[14] Ganan-Castro, A. M., De Ponte, D. P., Herrada, M. A., Spence, J. C., Weierstall, U., and Doak, R. B., Small **6**, 822-824 (2010).

[15] Shapiro, D. A., Chapman, H. N., Deponte, D., Doak, R. B., Fromme, P., Hembree, G., Hunter, M., Marchesini, S., Schmidt, K., Spence, J., Starodub, D., and Weierstall, U., J. Synchrotron Radiat. **15**, 593-599 (2008).

[16] Ihee, H., Rajagopal, S., Srajer, V., Pahl, R., Anderson, S., Schmidt, M., Schotte, F., Anfinrud, P. A., Wulff, M., and Moffat, K., Proc. Nat. Acad. Sci. (USA) **102**, 7145-7150. (2005).

[17] Schmidt, M., Pahl, R., Srajer, V., Anderson, S., Ren, Z., Ihee, H., Rajagopal, S., and Moffat, K., Proc. Nat. Acad. Sci. (USA) **101**, 4799-4804 (2004).

[18] Schotte, F., Cho, H. S., Kaila, V. R., Kamikubo, H., Dashdorj, N., Henry, E. R., Graber, T. J., Henning, R., Wulff, M., Hummer, G., Kataoka, M., and Anfinrud, P. A., Proc. Nat. Acad. Sci. (USA) **109**, 19256-19261 (2012).

[19] Jung, Y. O., Lee, J. H., Kim, J., Schmidt, M., Moffat, K., Srajer, V., and Ihee, H., Nat. Chem., in press.

[20] M. Schmidt, unpublished.

[21] Yoon, C. Y., Schwander, P., Abergel, C., Anderssen J., Aquila, A., Bajt, S., Barthelmess, M., Barty, A., Bogan, M. J., Bostedt, C., Bozek, J., Chapman, H. N., Claverie, J.-M., Coppola, C., De Ponte, D., Ekeberg, T., Epp, S. W., Erk, B., Flekenstein, H., Foucar, L., Graafsma, H., Gumprecht, L., Hajdu, J., Hampton, C. Y., Hartmann, A., Hartmann, E., Hartmann, R., Hauser, G., Hirsemann, H., Holl, P., Kassemeyer, S., Kimmel, N., Kiskinova, M., Liang, M., Loh, N.-T. D. Loh, Lomb, L., Maia F. R. N. C., Martin, A., Nass, K., Pedersoli, E., Recich, C., Rolles, D., Rudek, B., Rudenko, A., Sclichting, I., Schulz, J., Siebert, M., Seltzer, V., Shoeman R. L., Sierra, R. G., Soltau, H. Starodub, D., Steinbrener, J., Stier, G., Strüder, L., Svenda, M., Ulrich, J., Weidenspointer, G., White, G. A., Wunderer, C., Ourmazd, A. Opt. Exp. **19**, 16542-16549 (2011).

[22] Z. Kam, Macromolecules **10**, 927-934 (1978).

[23] D. K. Saldin, V. L. Shneerson, R. Fung, and A. Ourmazd, J. Phys.: Condens. Matter **21**, 134014 (2009).

[24] V. Elser, Ultramicroscopy **111**, 788-792 (2011).

[25] H.-C. Poon, P. Schwander, M. Uddin, and D. K. Saldin, submitted for publication.

[26] D. K. Saldin, H.-C. Poon, P. Schwander, M. Uddin, and M. Schmidt, Opt. Exp. **19**, 17318-17335 (2011).

[27] S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi, Science **220**, 671-680 (1983).