crystals

MDPI

Article

Probabilistic Estimate of $|F_{oa}|$ from FEL Data

Carmelo Giacovazzo *, Benedetta Carrozzini and Giovanni Luca Cascarano

Istituto di Cristallografia, CNR, Via G. Amendola 122/o, I-70126 Bari, Italy; benedetta.carrozzini@ic.cnr.it (B.C.); gianluca.cascarano@ic.cnr.it (G.L.C.)

* Correspondence: carmelo.giacovazzo@ic.cnr.it; Tel.: +39-080-592-9140

Received: 28 February 2018; Accepted: 9 April 2018; Published: 18 April 2018

Abstract: The method of the joint probability distribution function was applied in order to estimate the normal structure factor amplitudes of the anomalous scatterer substructure in a *FEL* experiment. The two-wavelength case was examined. In this, the prior knowledge of the moduli $|F_1^+|, |F_1^-|, |F_2^+|, |F_2^-|$ was used to predict the value of $|F_{0a}|$, which is the structure factor amplitude arising from the normal scattering of the heavy atom anomalous scatterers. The mathematical treatment provides a solid theoretical basis for the RIP (Radiation-damage Induced Phasing) method, which was originally proposed in order to take the radiation damage induced by synchrotron radiation sources into account. This was further adapted to exploit *FEL* data, where the crystal damage is usually more massive.

Keywords: joint probability distribution function; *FEL* data; RIP method

1. Introduction

The properties of large tunability, high monochromaticity, small beam divergence and large power of the modern synchrotron beamlines have made *MIR*, *MAD* and *MIRAS* popular tools for determining the crystal structure solution of proteins. Such phasing procedures are usually organized into two steps. In the first step, the amplitudes of the structure factors corresponding to the anomalous scatterers or to the heavy atom substructure are estimated. After this, they are used to solve the substructure via the Direct or Patterson methods [1–6]. In the second step, the full protein structure is identified using the prior information of the substructure [7–13].

Despite the increased efficiency of the modern phasing methods, a bottleneck is still present, which is related to the capacity of the method in growing sufficiently large and well-ordered crystals for a large class of biologically important proteins (e.g., membrane proteins). We can use X-ray free-electron lasers (*FEL*) to overcome the difficulty. Furthermore, due to their unprecedented high X-ray fluence [14,15], the diffraction effects may be measured even in nanocrystals [15–19].

However, the extreme high fluence [20] causes severe radiation damage [21] although diffraction before destruction is allowed as an effect of the ultrashort X-ray pulses. During femtosecond pulses, an ionization process [15] occurs, with this effect usually described as diffraction during ionization [22].

Therefore, radiation damage is of major concern if *FEL* data are collected for phasing purposes. According to Blake and Phillips [23], the induced damage quickly reduces the diffractive power of the cryocooled crystals, increases the vibrational atomic parameters and consequently, increases the cell volume. Local variations of the electron density are also generated. Heavy-atom substructures seem to be particularly sensitive to damage and as a consequence, dispersive signals can be obscured in *SAD-MAD* experiments.

More recently, the radiation damage has been used to phase macromolecular structures [24,25]. This method is mainly focused on synchrotron radiation sources and has been named *RIP* (Radiation-damage Induced Phasing). The data collected before the damage and data collected from a damaged crystal are used as the isomorphous data.

The *RIP* approach has also been applied to *FEL* data, where the damage is usually more massive and of a different nature. Indeed, the damage generated by a synchrotron source is mainly constituted of the breaking of bonds and implies atomic movements. Due to the very short pulse-length, the *FEL* damage takes the ionization of the atoms into account and does not imply atomic movement.

A detailed description of the physical phenomena occurring during the ultrashort X-ray pulses has been described in recent papers [22,26–30]. These papers studied the evolution of different electron populations for some selected state charges during an X-ray pulse and demonstrated how the neutrality of heavy atoms may be depleted into high-charge states. The practical physical effects may be condensed in a short statement. In a *FEL* experiment, nanocrystals may be used to provide different sets of diffraction data, with one corresponding to the undamaged crystal (obtained via low fluence pulses) and one or more with high fluence pulses.

The interaction X-ray pulse-crystal affects both the coherent atomic scattering factor and the dispersion effects. Son et al. [22] suggested that a basic assumption may be made in order to ensure that the *FEL* anomalous dispersion experiment is highly informative. This assumption is that the contribution of the light atoms to the anomalous scattering may be neglected when the photon energy is near to the inner-shell ionization threshold of the heavy atoms. They also showed that the *K*-shell edge is shifted from the standard position by a certain quantity, which depends on the specific charge state of the atom (accordingly, different configurations of the same heavy atom species may be simultaneously present in the crystal). Furthermore, Δf_j and f_j'' amplitudes also change according to the configuration selected for the different charge states.

In this paper, we will consider a simplified model for both the damaged and undamaged crystals, in which diffraction data with anomalous scattering effects are available. In the state 1, where radiation damage is assumed to be absent, the *j*th scattering factor may be represented by:

$$[f_j]_1 = f_j^0 + [\Delta f_j']_1 + i[f_j'']_1 = [f_j']_1 + i[f_j'']_1$$

In the state 2, where radiation damage occurs, we obtain:

$$[f_j]_2 = f_j^0 + [\Delta f_j^0]_2 + [\Delta f_j']_2 + i[f_j'']_2 = f_j^0 + [\Delta f_j]_2 + i[f_j'']_2$$

In both cases, the scattering factor may be represented (see the definition at Section 1) by:

$$f_j = f_j^0 + \Delta f_j + i f_j'',$$

where Δf_j includes the variations of both the normal scattering factor and the real component of the anomalous scatterer. Thus, we can conclude that:

for the state 1 $\Delta f_j = [\Delta f_j']_1$ and $f_j'' = [f_j'']_1$;
for the state 2 $\Delta f_j = [\Delta f_j^0]_2 + [\Delta f_j']_2$ and $f_j'' = [f_j'']_2$.

The above two states are usually generated in a *FEL* experiment by two different fluences, although the same wavelength is used. This situation has no counterpart in a SAD experiment and is closer to a two-wavelength MAD experiment, where the two different states correspond to two different pairs $(\Delta f', f'')$. The substantial difference between *FEL* and the two-wavelength MAD case is not due to the use of a different number of wavelengths (the final MAD formulas do not depend on the number of wavelengths but on the states corresponding to them), but this is attributed to the fact that the crystal damage at high fluence involves the use of $f_j^0 + [\Delta f_j^0]_2$ rather than f_j^0 in a *FEL* experiment.

The continuous ionization of the heavy atoms during *FEL* experiments makes it difficult to directly apply MAD techniques to *FEL* data [22]. Indeed, the present theory cannot take into account the large number of electronic states accessible to heavy atoms during a high-fluence *FEL* experiment and the dynamics of these states during measurements. Our model is static. In other words, we assume that the structure factors do not change during measurements taken from the damaged crystal. The model

would be certainly more representative of the dynamic nature of the damage if it might include a time-dependent term, but this is outside of our present capacity.

However, the fact that many protein structures have been solved using *FEL* data obtained by Molecular Replacement techniques suggests that data collected from an undamaged crystal (low fluence) and data collected from a damaged crystal (high fluence) have sufficient internal consistency to be useful in attempts to determine the protein crystal structure. This was the main reason that encouraged us to describe a probabilistic theory integrating *FEL* damage into the framework of the anomalous dispersion techniques [12].

This work focuses on the first step of a phasing procedure, which involves finding the heavy atom positions according to the diffraction data.

2. The Mathematical Model

A necessary condition for finding the heavy atom substructure from the diffraction amplitudes of undamaged and damaged crystals involves having sufficiently accurate estimates of the R_{0H} amplitudes. Once available, the Direct or Patterson methods can use these estimates to locate the heavy atom positions. The most general approach for estimating R_{0H} involves the calculation of the joint probability distribution function:

$$P\left(E_{0H}, E^{+}, E_{d}^{+}, E^{-}, E_{d}^{-}\right). \tag{1}$$

From this (see below), we can derive the conditional estimate as follows:

$$< R_{0H} \left| \, R^{+}, R_{d}^{+}, R^{-}, R_{d}^{-} \right. > .$$

where the superscripts + and − indicate that the structure factors are calculated for the reflections **h** and −**h**, respectively.

The use of Equation (1) requires the preliminary modelling of the structure factors. Their algebraic form should reflect the physical phenomena occurring in a *FEL* experiment. According to Section 1, the isomorphous derivatives created in such experiments are similar, although they are not identical to those available in traditional *MAD* or *MIRAS* cases. Thus, it may be worthwhile to compare these derivatives in order to discover similarity and differences in addition to profiting from the mathematical theories already developed for treating the usual *MAD* and *MIRAS* cases [5,13,31].

Our *FEL* model involves the following assumptions:

(1) For light atoms, Δf_j is assumed to be negligible both for the damaged and for the undamaged crystals. In our model, we set $\Delta f_j = 0$ for $j = 1, \ldots, L$. Furthermore, f_j^0 is assumed to be the same for the jth atom independently of whether this refers to the undamaged or damaged crystal.

(2) For heavy atoms, the scattering factor of the jth atom f_j^0 is described by two different functions according to whether the atom is considered in the undamaged or in the damaged crystal. As a result, Δf_j will assume different values.

(3) The atomic positions in the damaged and in the undamaged crystals coincide [26,32]. Indeed, there is no evidence so far that the pulse duration (about a few femtoseconds) produces detectable changes in the heavy atom positions in femtosecond X-ray nanocrystallography even if the simulations of radiation dynamics in proteins suggests some correlated movement of the heavy atoms [33].

(4) The observed amplitudes $F^{+}, F_{d}^{+}, F^{-}, F_{d}^{-}$ are affected by errors, which are calculated as:

$$\mu^{+} = \left|\mu^{+}\right| exp\left(i\vartheta^{+}\right) \text{ and } \mu^{-} = \left|\mu^{-}\right| exp\left(i\vartheta^{-}\right),$$

$$\mu_{d}^{+} = \left|\mu_{d}^{+}\right| exp\left(i\vartheta_{d}^{+}\right) \text{ and } \mu_{d}^{-} = \left|\mu_{d}^{-}\right| exp\left(i\vartheta_{d}^{-}\right)$$

These represent the measurement errors relative to the undamaged and the damaged crystal, respectively. The errors are assumed to be complex quantities because they influence both the real and the imaginary components of the structure factors. Since R_{0H} is not measured and is instead estimated using the probabilistic approach, no error will be associated with it. Furthermore, the errors will be assumed to be uncorrelated and thus, we can obtain:

$$< \mu^+ >=< \mu_d^+ >=< \mu^- >=< \mu_d^- >= 0$$

and

$$< \mu^+\mu_d^+ >=< \mu^+\mu^- >=< \mu^-\mu_d^- >=< \mu_d^+\mu_d^- >= 0$$

According to Equations (1)–(4), the following mathematical model will be adopted:

$$A^+ = \left\{\sum_{j=1}^{L} f_j^o cos(2\pi \mathbf{hr}_j) + \sum_{j=1}^{H} f_j' cos(2\pi \mathbf{hr}_j) - \sum_{j=1}^{H} f_j'' sin(2\pi \mathbf{hr}_j) + \left|\mu^+\right| cos\vartheta^+ \right\} / \left(\varepsilon \sum_{N}\right)^{1/2},$$

$$B^+ = \left\{\sum_{j=1}^{L} f_j^o \sin(2\pi \mathbf{hr}_j) + \sum_{j=1}^{H} f_j' \sin(2\pi \mathbf{hr}_j) + \sum_{j=1}^{H} f_j'' \cos(2\pi \mathbf{hr}_j) + \left|\mu^+\right| \sin\vartheta^+ \right\} / \left(\varepsilon \sum_{N}\right)^{1/2},$$

$$A^- = \left\{\sum_{j=1}^{L} f_j^o cos(2\pi \mathbf{hr}_j) + \sum_{j=1}^{H} f_j' cos(2\pi \mathbf{hr}_j) + \sum_{j=1}^{H} f_j'' sin(2\pi \mathbf{hr}_j) + \left|\mu^-\right| cos\vartheta^- \right\} / \left(\varepsilon \sum_{N}\right)^{1/2}$$

$$B^- = \left\{-\sum_{j=1}^{L} f_j^o sin(2\pi \mathbf{hr}_j) - \sum_{j=1}^{H} f_j' sin(2\pi \mathbf{hr}_j) + \sum_{j=1}^{H} f_j'' cos(2\pi \mathbf{hr}_j) + \left|\mu^-\right| sin\vartheta^- \right\} / \left(\varepsilon \sum_{N}\right)^{1/2},$$

$$A_d^+ = \left\{\sum_{j=1}^{L} f_j^o cos(2\pi \mathbf{hr}_j) + \sum_{j=1}^{H} f_{dj}' cos(2\pi \mathbf{hr}_j) - \sum_{j=1}^{H} f_{dj}'' sin(2\pi \mathbf{hr}_j) + \left|\mu_d^+\right| cos\vartheta_d^+ \right\} / \left(\varepsilon \sum_{Nd}\right)^{1/2},$$

$$B_d^+ = \left\{\sum_{j=1}^{L} f_j^o \sin(2\pi \mathbf{hr}_j) + \sum_{j=1}^{H} f_{dj}' \sin(2\pi \mathbf{hr}_j) + \sum_{j=1}^{H} f_{dj}'' \cos(2\pi \mathbf{hr}_j) + \left|\mu_d^+\right| \sin\vartheta_d^+ \right\} / \left(\varepsilon \sum_{Nd}\right)^{1/2},$$

$$A_d^- = \left\{\sum_{j=1}^{L} f_j^o cos(2\pi \mathbf{hr}_j) + \sum_{j=1}^{H} f_{dj}' cos(2\pi \mathbf{hr}_j) + \sum_{j=1}^{H} f_{dj}'' sin(2\pi \mathbf{hr}_j) + \left|\mu_d^-\right| cos\vartheta_d^- \right\} / \left(\varepsilon \sum_{Nd}\right)^{1/2},$$

$$B_d^- = \left\{-\sum_{j=1}^{L} f_j^o sin(2\pi \mathbf{hr}_j) - \sum_{j=1}^{H} f_{dj}' sin(2\pi \mathbf{hr}_j) + \sum_{j=1}^{H} f_{dj}'' cos(2\pi \mathbf{hr}_j) + \left|\mu_d^-\right| sin\vartheta_d^- \right\} / \left(\varepsilon \sum_{Nd}\right)^{1/2}.$$

It is important to notice that for heavy atoms:

$$f_j' = f_j^0 + \Delta f_j', \tag{2a}$$

$$f_{dj}' = f_{dj}^0 + \Delta f_{dj}', \tag{2b}$$

where

$$\Delta f_{dj}' = \Delta f_{0j} + \Delta f_j'. \tag{2c}$$

$[\Delta f_j']_1$ and $[\Delta f_j']_2$ are the real components of the anomalous scattering of the jth heavy atom for the undamaged and damaged crystals, respectively.

We then considered the mathematical model adopted for the usual two-wavelength *MAD* experiment by Giacovazzo and Siliqi [12]. In order to facilitate an easier comparison with the above *FEL* model, we slightly changed the original Giacovazzo and Siliqi notation. We denoted $E^+ \equiv (A^+ + iB^+)$, $E^- \equiv (A^- + iB^-)$ as the normalized structure factors corresponding to the

first wavelength. $E_d^+ \equiv (A_d^+ + iB_d^+)$, $E_d^- \equiv (A_d^- + iB_d^-)$ is denoted as the normalized structure factor corresponding to the second *MAD* wavelength, while E_{0H} is denoted as the structure factor corresponding to the anomalous scatterer substructure (anomalous scattering is not included). L and H is the number of non-hydrogen light atoms and the number of the anomalous scatterers in the unit cell, respectively. Thus, the Giacovazzo and Siliqi model may be rewritten as follows:

$$A^+ = \left\{ \sum_{j=1}^{L} f_j^o cos(2\pi \mathbf{hr}_j) + \sum_{j=1}^{H} f_j' cos(2\pi \mathbf{hr}_j) - \sum_{j=1}^{H} f_j'' sin(2\pi \mathbf{hr}_j) + \left|\mu^+\right| cos\vartheta^+ \right\} / \left(\varepsilon \sum_{N} \right)^{1/2},$$

$$B^+ = \left\{ \sum_{j=1}^{L} f_j^o sin(2\pi \mathbf{hr}_j) + \sum_{j=1}^{H} f_j' sin(2\pi \mathbf{hr}_j) - \sum_{j=1}^{H} f_j'' cos(2\pi \mathbf{hr}_j) + \left|\mu^+\right| sin\vartheta^+ \right\} / \left(\varepsilon \sum_{N} \right)^{1/2},$$

$$A^- = \left\{ \sum_{j=1}^{L} f_j^o cos(2\pi \mathbf{hr}_j) + \sum_{j=1}^{H} f_j' cos(2\pi \mathbf{hr}_j) - \sum_{j=1}^{H} f_j'' sin(2\pi \mathbf{hr}_j) + \left|\mu^-\right| cos\vartheta^- \right\} / \left(\varepsilon \sum_{N} \right)^{1/2},$$

$$B^- = \left\{ \sum_{j=1}^{L} f_j^o sin(2\pi \mathbf{hr}_j) + \sum_{j=1}^{H} f_j' sin(2\pi \mathbf{hr}_j) - \sum_{j=1}^{H} f_j'' cos(2\pi \mathbf{hr}_j) + \left|\mu^-\right| sin\vartheta^- \right\} / \left(\varepsilon \sum_{N} \right)^{1/2},$$

$$A_d^+ = \left\{ \sum_{j=1}^{L} f_j^o cos(2\pi \mathbf{hr}_j) + \sum_{j=1}^{H} f_{dj}' cos(2\pi \mathbf{hr}_j) - \sum_{j=1}^{H} f_{dj}'' sin(2\pi \mathbf{hr}_j) + \left|\mu_d^+\right| cos\vartheta_d^+ \right\} / \left(\varepsilon \sum_{Nd} \right)^{1/2},$$

$$B_d^+ = \left\{ \sum_{j=1}^{L} f_j^o sin(2\pi \mathbf{hr}_j) + \sum_{j=1}^{H} f_{dj}' sin(2\pi \mathbf{hr}_j) - \sum_{j=1}^{H} f_{dj}'' cos(2\pi \mathbf{hr}_j) + \left|\mu_d^+\right| sin\vartheta_d^+ \right\} / \left(\varepsilon \sum_{Nd} \right)^{1/2},$$

$$A_d^- = \left\{ \sum_{j=1}^{L} f_j^o cos(2\pi \mathbf{hr}_j) + \sum_{j=1}^{H} f_{dj}' cos(2\pi \mathbf{hr}_j) - \sum_{j=1}^{H} f_{dj}'' sin(2\pi \mathbf{hr}_j) + \left|\mu_d^-\right| cos\vartheta_d^- \right\} / \left(\varepsilon \sum_{Nd} \right)^{1/2},$$

$$B_d^- = \left\{ \sum_{j=1}^{L} f_j^o sin(2\pi \mathbf{hr}_j) + \sum_{j=1}^{H} f_{dj}' sin(2\pi \mathbf{hr}_j) - \sum_{j=1}^{H} f_{dj}'' cos(2\pi \mathbf{hr}_j) + \left|\mu_d^-\right| sin\vartheta_d^- \right\} / \left(\varepsilon \sum_{Nd} \right)^{1/2}$$

where

$$f_j' = f_j^0 + \Delta f_j' \tag{3a}$$

and

$$f_{dj}' = f_j^0 + \Delta f_{dj}'. \tag{3b}$$

Now f_j^0 does not vary when considered for the first or for the second wavelength independently of whether Equation (3) refers to heavy or to light atoms.

Comparing Equation (3) with Equation (2) reveals the main difference between the *FEL* and the usual two-wavelength *MAD* mathematical model. The *FEL* data imply a supplementary decay of the scattering factor of the heavy atoms, which is generated by the crystal damage when passing from the first to the second wavelength. However, there is significantly similarity between the two models so thus, most of the mathematics developed for *MAD* by Giacovazzo and Siliqi [12] may also be used for *FEL* with small modifications, which is shown in Section 3.

3. The Joint Probability Distribution $P(E_{0H}, E^+, E_d^+, E^-, E_d^-)$

The characteristic function of

$$P\left(A_{0H}, A^+, A_d^+, A^-, A_d^-, B_{oH}, B^+, B_d^+, B^-, B_d^-\right)$$

may be written down as:

$$\begin{aligned} &C\left(u_{0H}, u^{+}, u_d^{+}, u^{-}, u_d^{-}, v_{0H}, v^{+}, v_d^{+}, v^{-}, v_d^{-}\right) \approx \\ &\quad exp\Big\{-\tfrac{1}{4}\big[k_{11}\left(u_{0H}^2+v_{0H}^2\right)+k_{22}\left(u^{+2}+v^{+2}\right)+k_{33}\left(u_d^{+2}+v_d^{+2}\right)+k_{44}\left(u^{-2}+v^{-2}\right) \\ &\qquad +k_{55}\left(u_d^{-2}+v_d^{-2}\right)+2k_{24}\left(u^{+}u^{-}-v^{+}v^{-}\right)+2k_{35}\left(u_d^{+}u_d^{-}-v_d^{+}v_d^{-}\right) \\ &\qquad +2k_{29}\left(u^{+}v^{-}+v^{+}u^{-}\right)+2k_{3,10}\left(u_d^{+}v_d^{-}+v_d^{+}u_d^{-}\right)+2k_{34}\left(u_d^{+}u^{-}-v_d^{+}v^{-}\right) \\ &\qquad +2k_{23}\left(u^{+}u_d^{+}+v^{+}v_d^{+}\right)+2k_{25}\left(u^{+}u_d^{-}-v^{+}v_d^{-}\right)+2k_{2,10}\left(u^{+}v_d^{-}+v^{+}u_d^{-}\right) \\ &\qquad +2k_{28}\left(u^{+}v_d^{+}-v^{+}u_d^{+}\right)+2k_{12}\left(u_{0H}u^{+}+v_{0H}v^{+}\right)+2k_{14}\left(u_{0H}u^{-}-v_{0H}v^{-}\right) \\ &\quad +2k_{17}\left(u_{0H}v^{+}-v_{0H}u^{+}\right)+2k_{19}\left(u_{0H}v^{-}+v_{0H}u^{-}\right)+2k_{13}\left(u_{0H}u_d^{+}+v_{0H}v_d^{+}\right) \\ &\qquad +2k_{15}\left(u_{0H}u_d^{-}-v_{0H}v_d^{-}\right)+2k_{18}\left(u_{0H}v_d^{+}-v_{0H}u_d^{+}\right)+2k_{1,10}\left(u_{0H}v_d^{-}+v_{0H}u_d^{-}\right) \end{aligned} \tag{4}$$

where $u_{0H}, u^{+}, u_d^{+}, u^{-}, u_d^{-}, v_{0H}, v^{+}, v_d^{+}, v^{-}, v_d^{-}$ are the carrying variables associated with $A_{0H}, A^{+}, A_d^{+}, A^{-}, A_d^{-}, B_{oH}, B^{+}, B_d^{+}, B^{-}, B_d^{-}$, respectively.

Equation (4) can also be written as:

$$C = exp\left\{-\frac{1}{4}\left(\overline{\mathbf{U}}\mathbf{K}\mathbf{U}\right)\right\} \tag{5}$$

where

$$\overline{U} = \left[u_{0H}, u^{+}, u_d^{+}, u^{-}, u_d^{-}, v_{0H}, v^{+}, v_d^{+}, v^{-}, v_d^{-}\right]$$

and **K** is the symmetric square matrix. The elements of **K** are specified as follows:

$$k_{11} = k_{66} = 1$$

$$k_{22} = k_{77} = e^{+} = 1 + \sigma^{+2} \text{ with } \sigma^{+2} = \ <\left|\mu^{+}\right|_2> / \left(\varepsilon \sum_{N}\right)$$

$$k_{33} = k_{88} = e_d^{+} = 1 + \sigma_d^{+2} \text{ with } \sigma_d^{+2} = \ <\left|\mu_d^{+}\right|_2> / \left(\varepsilon \sum_{Nd}\right)$$

$$k_{44} = k_{99} = e^{-} = 1 + \sigma^{-2} \text{ with } \sigma^{-2} = \ <\left|\mu^{-}\right|_2> / \left(\varepsilon \sum_{N}\right)$$

$$k_{55} = k_{10,10} = e_d^{-} = 1 + \sigma_d^{-2} \text{ with } \sigma_d^{-2} = \ <\left|\mu_d^{-}\right|_2> / \left(\varepsilon \sum_{Nd}\right)$$

$$k_{12} = k_{14} = k_{67} = -k_{69} = \ S_9 / \left(\sum_{H}^{0} \sum_{N}\right)^{1/2}$$

$$k_{13} = k_{15} = k_{68} = -k_{6,10} = \ S_{11} / \left(\sum_{H}^{0} \sum_{Nd}\right)^{1/2}$$

$$k_{16} = k_{27} = k_{38} = k_{49} = k_{5,10} = 0$$

$$k_{17} = k_{19} = k_{46} = -k_{26} = \ S_{10} / \left(\sum_{H}^{0} \sum_{N}\right)^{1/2}$$

$$k_{18} = k_{1,10} = -k_{36} = k_{56} = \ S_{12} / \left(\sum_{H}^{0} \sum_{Nd}\right)^{1/2}$$

$$k_{23} = k_{45} = k_{9,10} = k_{78} = \ \left(\sum_{L}^{0} + S_5 + S_6\right) / \left(\sum_{N} \sum_{Nd}\right)^{1/2}$$

$$k_{24} = -k_{79} = \left(\sum_{L}^{0} + S_1\right) / \sum_{N}$$

$$k_{25} = k_{34} = -k_{7,10} = -k_{89} = \left(\sum_{L}^{0} + S_5 - S_6\right) / \left(\sum_{N}\sum_{Nd}\right)^{1/2}$$

$$k_{28} = -k_{37} = +k_{4,10} = -k_{59} = (S_7 - S_8) / \left(\sum_{N}\sum_{Nd}\right)^{1/2}$$

$$k_{29} = k_{47} = S_3 / \sum_{N}$$

$$k_{2,10} = k_{39} = k_{48} = k_{57} = (S_7 + S_8) / \left(\sum_{N}\sum_{Nd}\right)^{1/2}$$

$$k_{35} = -k_{8,10} = \left(\sum_{L}^{0} + S_2\right) / \sum_{Nd}$$

$$k_{3,10} = k_{58} = S_4 / \sum_{Nd},$$

where

$$S_1 = \sum_{H}\left(f_j'^2 - f_j''^2\right),\ S_2 = \sum_{H}\left(f_{dj}'^2 - f_{dj}''^2\right),\ S_3 = 2\sum_{H} f_j' f_j''$$

$$S_4 = 2\sum_{H} f_{dj}' f_{dj}'',\ S_5 = \sum_{H} f_j' f_{dj}',\ S_6 = \sum_{H} f_j'' f_{dj}''$$

$$S_7 = \sum_{H} f_j' f_{dj}'',\ S_8 = \sum_{H} f_j'' f_{dj}',\ S_9 = \sum_{H} f_j^o f_j'$$

$$S_{10} = \sum_{H} f_j^o f_j'',\ S_{11} = \sum_{H} f_j^o f_{dj}',\ S_{12} = \sum_{H} f_j^o f_{dj}''$$

The formal expressions of the k and S parameters coincide with those obtained by Giacovazzo and Siliqi for the usual two-wavelength *MAD* experiments. However, the f_{dj}', according to Equation (2b), implies a change in the normal jth heavy atoms scattering factor, which will influence the values of all the S parameters that involves f_{dj}'.

In our mathematical treatment, Δf_{0j} is implicitly assumed to be negative (it describes the passage from a neutral to an ionic state). As $\Delta f_j'$ is also usually negative, the value of $\Delta f_{dj}' = \Delta f_{0j} + \Delta f_j'$ is expected to be largely negative, particularly for high ionization states. This effect may strongly reduce the f_{dj}' values. As a result, the S parameters f_{dj}' are expected to be smaller for *FEL* data than for standard *MAD* data.

In accordance with the above considerations, the joint probability distribution function $P\left(A_{oH}, A^+, A_d^+, A^-, A_d^-, B_{oH}, B^+, B_d^+, B^-, B_d^-\right)$ may be obtained by the Fourier inversion of the characteristic function of Equation (5). Through the change of variables, we can obtain:

$$A_{0H} = R_{0H} cos\varphi_{0H},\ B_{0H} = R_{0H} sin\overset{\leftrightarrow}{\varphi}_{0H}$$

$$A^+ = R^+ cos\varphi^+,\ B^+ = R^+ sin\varphi^+$$

$$A^- = R^- cos\varphi^-,\ B^- = R^- sin\varphi^-$$

$$A_d^+ = R_d^+ cos\varphi_d^+,\ B_d^+ = R_d^+ sin\varphi_d^+$$

$$A_d^- = R_d^- cos\varphi_d^-,\ B_d^- = R_d^- sin\varphi_d^-.$$

The following joint probability distribution function is obtained:

$$\begin{aligned}
&P\left(R_{0H}, R^{+}, R^{-}, R_d^{+}, R_d^{-}, \varphi_a, \varphi^{+}, \ldots, \varphi_d^{-}\right) \approx \pi^{-5} R_{0H} R^{+} R^{-} R_d^{+} R_d^{-} [\det \mathbf{K}]^{-1/2} \\
&exp\Big\{-\Big[\lambda_{11}(R_{0H})^2 + \lambda_{22}(R^{+})^2 + \lambda_{33}(R_d^{+})^2 + \lambda_{44}(R^{-})^2 + \lambda_{55}(R_d^{-})^2 \\
&\quad + 2\lambda_{23} R^{+} R_d^{+} cos\left(\varphi^{+} - \varphi_d^{+}\right) - 2\lambda_{28} R^{+} R_d^{+} \sin\left(\varphi^{+} - \varphi_d^{+}\right) \\
&\quad + 2\lambda_{24} R^{+} R^{-} cos\left(\varphi^{+} + \varphi^{-}\right) + 2\lambda_{29} R^{+} R^{-} \sin\left(\varphi^{+} + \varphi^{-}\right) \\
&\quad + 2\lambda_{25} R^{+} R_d^{-} cos\left(\varphi^{+} + \varphi_d^{-}\right) + 2\lambda_{2,10} R^{+} R_d^{-} \sin\left(\varphi^{+} + \varphi_d^{-}\right) \\
&\quad + 2\lambda_{34} R^{-} R_d^{+} cos\left(\varphi^{-} + \varphi_d^{+}\right) + 2\lambda_{39} R^{-} R_d^{+} \sin\left(\varphi^{-} + \varphi_d^{+}\right) \\
&\quad + 2\lambda_{35} R_d^{+} R_d^{-} cos\left(\varphi_d^{+} + \varphi_d^{-}\right) + 2\lambda_{3,10} R_d^{+} R_d^{-} \sin\left(\varphi_d^{+} + \varphi_d^{-}\right) \\
&\quad + 2\lambda_{45} R^{-} R_d^{-} cos\left(\varphi^{-} - \varphi_d^{-}\right) - 2\lambda_{4,10} R^{-} R_d^{-} \sin\left(\varphi^{-} - \varphi_d^{-}\right) \\
&\quad + 2\lambda_{12} R_{0H} R^{+} cos\left(\varphi_{0H} - \varphi^{+}\right) - 2\lambda_{17} R_{0H} R^{+} \sin\left(\varphi_{0H} - \varphi^{+}\right) \\
&\quad + 2\lambda_{13} R_{0H} R_d^{+} cos\left(\varphi_{0H} - \varphi_d^{+}\right) - 2\lambda_{18} R_{0H} R_d^{+} \sin\left(\varphi_{0H} - \varphi_d^{+}\right) \\
&\quad + 2\lambda_{14} R_{0H} R^{-} cos\left(\varphi_{0H} + \varphi^{-}\right) - 2\lambda_{19} R_{0H} R^{-} \sin\left(\varphi_{0H} + \varphi^{-}\right) \\
&\quad + 2\lambda_{15} R_{0H} R_d^{-} cos\left(\varphi_{0H} + \varphi_d^{-}\right) + 2\lambda_{1,10} R_{0H} R_d^{-} \sin\left(\varphi_{0H} + \varphi_d^{-}\right)]\Big\},
\end{aligned} \tag{6}$$

where λ_{pq} are the elements of the matrix $\mathbf{K}^{-1}$.

The distribution found in Equation (6) is the main result of this paper.

In accordance with Giacovazzo and Siliqi and based on Equation (6), the conditional probability may be obtained:

$$P\left(R_{0H} \middle| R^{+}, R^{-}, R_d^{+}, R_d^{-}\right) \tag{7}$$

From this, the conditional mean value is:

$$< \left(R_{0H} \middle| R^{+}, R^{-}, R_d^{+}, R_d^{-}\right) > \approx \frac{1}{2}(\pi/\lambda_{11})^{1/2}\left(1 + \frac{4}{\pi}\frac{X^2}{\lambda_{11}}\right)^{\frac{1}{2}}, \tag{8}$$

where:

$$X^2 = {Q_1}^2 + {Q_2}^2$$

$$Q_1 = \lambda_{12} R_1 + \lambda_{13} R_2 + \lambda_{14} G_1 + \lambda_{15} G_2$$

$$Q_2 = \lambda_{17} R_1 + \lambda_{18} R_2 - \lambda_{19} G_1 - \lambda_{1,10} G_2.$$

The details of the mathematical derivation may be found in the referenced paper. In this paper, the standard deviation associated with the estimate of Equation (8) was calculated to be:

$$\sigma_{R_{0a}} = \left[< R_{oa}^2 \middle| .. > - < R_{oa} \middle| .. >^2\right]^{1/2} = \left[\left(1 - \frac{\pi}{4}\right)\lambda_{11}^{-1}\right]^{1/2}$$

so that

$$\frac{< R_{oa} | .. >}{\sigma_{R_{oa}}} = \left[\frac{\pi/4 + X^2/\lambda_{11}}{1 - \pi/4}\right]^{1/2}. \tag{9}$$

Since λ_{11} is always expected to be positive, the expected values of $< R_{oa} >$ and $\sigma_{R_{oa}}$ are always positive. The reflections with the largest values of the ratio found in Equation (9) are likely to be the most useful ones.

4. Conclusions

We have described a probabilistic approach for modelling the crystal structure factors when the crystal is damaged due to the strong impact of *FEL* radiation on the crystal sample. This involves using corresponding data to solve the crystal structure. The calculations show that the mathematical formulation described by Giacovazzo and Siliqi [12] may also be applied to *FEL* data, provided that suitable modifications are introduced in the parameters depending on the scattering factors.

Author Contributions: Benedetta Carrozzini, Giovanni Luca Cascarano and Carmelo Giacovazzo conceived the theory; Carmelo Giacovazzo wrote the paper.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviation

ε	statistical Wilson coefficient (corrected for expected intensities in reciprocal lattice zones).
L , H	number of non-hydrogen light and heavy atoms in the unit cell, respectively. Their values do not vary when considered in the undamaged or damaged crystals.
$N = L + H$	number of non-hydrogen atomic positions in the unit cell, for the undamaged and the damaged crystals.
$f_j = f_j^0 + \Delta f_j + if_j'' = f_j' + if_j''$	scattering factor of the jth atom. f' is its real and f'' is its imaginary part. The thermal factor is included.
$\Sigma_N = \sum_{j=1}^{N} \left(f_j'^2 + f_j''^2\right)$,	the summation is calculated for the undamaged crystal and is extended to all the atoms in the unit cell.
$\Sigma_{Nd} = \sum_{j=1}^{N} \left(f_{dj}'^2 + f_{dj}''^2\right)$,	the summation is calculated for the damaged crystal and is extended to all the atoms in the unit cell.
$\Sigma_L^0 = \sum_{j=1}^{L} \left(f_j^o\right)^2$	the summation is extended to all the light atoms in the unit cell.
$\Sigma_H^0 = \sum_{j=1}^{H} \left(f_j^o\right)^2$	the summation is extended to all the heavy atoms in the unit cell.
$F = \lvert F\rvert exp(i\varphi) = \sum_{j=1}^{N} f_j exp\left(2\pi i \mathbf{hr}_j\right)$	structure factor of the undamaged crystal
$E = \left\lvert F\right\rvert exp(i\varphi)/(\varepsilon\Sigma_N)^{1/2} = Rexp(i\varphi) = A + iB$	normalized structure factor of the undamaged crystal.
$F_d = \lvert F_d\rvert exp(i\varphi_d) = \sum_{j=1}^{Nd} f_{dj} exp\left(2\pi i \mathbf{hr}_j\right)$	structure factor of the damaged crystal.
$E_d = \lvert F_d\rvert exp(i\varphi)/(\varepsilon\Sigma_{dN})^{1/2} = R_d exp\left(i\varphi_d\right) = A_d + iB_d$	normalized structure factor of the damaged crystal
$F_{0H} = \lvert F_{0H}\rvert \exp(i\varphi_{0H}) = \sum_{j=1}^{H} f_j^o \exp\left(2\pi \mathbf{hr}_j\right)$	normal structure factor for the heavy atom substructure (anomalous scattering excluded).
$E_{0H} = \lvert F_{0H}\rvert exp(i\varphi_{0H})/(\varepsilon\Sigma_{0H})^{1/2} = R_{0H} exp(i\varphi_{0H}) = A_{0H} + iB_{0H}$	normalized structure factor of the normal heavy atom substructure (anomalous scattering excluded).
$F_H = \lvert F_H\rvert \exp(i\varphi_H) = \sum_{j=1}^{H} f_j \exp\left(2\pi \mathbf{hr}_j\right)$	structure factor of the heavy atom substructure.
MIR, *MAD*, *MIRAS*	multiple isomorphous replacement, multiple anomalous dispersion, multiple isomorphous replacement combined with anomalous scattering techniques, respectively. For brevity, we include into the above definitions the particular cases of *SIR* (single isomorphous replacement), *SAD* (single anomalous dispersion) and *SIRAS* (single isomorphous replacement combined with anomalous scattering techniques).

References

1. Blow, D.M.; Crick, F.H.C. The treatment of errors in the isomorphous replacement method. *Acta Cryst.* **1959**, *12*, 794–802. [CrossRef]
2. Blow, D.M.; Rossmann, M.G. The single isomorphous replacement method. *Acta Cryst.* **1961**, *14*, 1195–1202. [CrossRef]
3. North, A.C.T. The combination of isomorphous replacement and anomalous scattering data in phase determination of non-centrosymmetric reflexions. *Acta Cryst.* **1965**, *18*, 212–216. [CrossRef]
4. Mathews, B.W. The extension of the isomorphous replacement method to include anomalous scattering measurements. *Acta Cryst.* **1966**, *20*, 82–86. [CrossRef]
5. Burla, M.C.; Carrozzini, B.; Cascarano, G.L.; Giacovazzo, C.; Polidori, G.; Siliqi, D. MAD phasing: Probabilistic estimate of $\lvert F_{oa}\rvert$. *Acta Cryst.* **2002**, *D58*, 928–935. [CrossRef]
6. Burla, M.C.; Carrozzini, B.; Cascarano, G.L.; Giacovazzo, C.; Polidori, G. SAD or MAD phasing: Location of the anomalous scatterers. *Acta Cryst.* **2003**, *D59*, 662–669. [CrossRef]
7. Terwilliger, T.C.; Eisenberg, D. Isomorphous replacement: Effects of errors on the phase probability distribution. *Acta Cryst.* **1987**, *A43*, 6–13. [CrossRef]

8. Miller, R.; Gallo, S.M.; Khala, M.G.; Weeks, C.M. *SnB*: Crystal structure determination via shake-and-bake. *J. Appl. Cryst.* **1994**, *27*, 613–621. [CrossRef]
9. Sheldrick, G.M.; Gould, R.G. Structure solution by iterative peaklist optimization and tangent expansion in space group *P*1. *Acta Cryst.* **1995**, *B51*, 423–431. [CrossRef]
10. Pähler, A.; Smith, J.L.; Hendrickson, W.A. A probability representation for phase information from multiwavelength anomalous dispersion. *Acta Cryst.* **1990**, *A46*, 537–540. [CrossRef]
11. Terwilliger, T.C. MAD phasing: Bayesian estimates of F_A. *Acta Cryst.* **1994**, *D50*, 11–16. [CrossRef]
12. Giacovazzo, C.; Siliqi, D. The method of joint probability distribution functions applied to SIR-MIR and to SIRAS-MIRAS cases. *Acta Cryst.* **2002**, *A58*, 590–597. [CrossRef]
13. Giacovazzo, C.; Siliqi, D. Phasing via SAD/MAD data: The method of the joint probability distribution functions. *Acta Cryst.* **2004**, *D60*, 73–82. [CrossRef]
14. Gaffney, K.J.; Chapman, H.N. Imaging atomic structure and dynamics with ultrafast X-ray scattering. *Science* **2007**, *316*, 1444–1448. [CrossRef] [PubMed]
15. Neutze, R.; Wouts, R.; van der Spoel, D.; Weckert, E.; Hajdu, J. Potential for biomolecular imaging with femtosecond X-ray pulses. *Nature* **2000**, *406*, 752–757. [CrossRef] [PubMed]
16. Chapman, H.N.; Barty, A.; Bogan, M.J.; Boutet, S.; Frank, M.; Hau-Riege, S.P.; Marchesini, S.; Woods, B.W.; Bajt, S.; Benner, W.H.; et al. Femtosecond Diffractive Imaging with a Soft-X-ray Free-Electron Laser. *Nat. Phys.* **2006**, *2*, 839–843. [CrossRef]
17. Chapman, H.N.; Nugent, K.A. Coherent lensless X-ray imaging. *Nat. Photonics* **2010**, *4*, 833–839. [CrossRef]
18. Mancuso, A.P.; Schropp, A.; Reime, B.; Stadler, L.M.; Singer, A.; Gulden, J.; Streit-Nierobisch, S.; Gutt, C.; Grübel, G.; Feldhaus, J.; et al. Coherent-Pulse 2D Crystallography Using a Free-Electron Laser X-ray Source. *Phys. Rev. Lett.* **2009**, *102*, 035502. [CrossRef] [PubMed]
19. 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.; et al. Femtosecond X-ray protein nanocrystallography. *Nature* **2011**, *470*, 73–77. [CrossRef] [PubMed]
20. Henderson, R. The potential and limitations of neutrons, electrons and X-rays for atomic resolution microscopy of unstained biological molecules. *Q. Rev. Biophys.* **1995**, *28*, 171–193. [CrossRef] [PubMed]
21. Howells, M.R.; Beetz, T.; Chapman, H.N.; Cui, C.; Holton, J.M.; Jacobsen, C.J.; Kirz, J.; Lima, E.; Marchesini, S.; Miao, H.; et al. An assessment of the resolution limitation due to radiation-damage in X-ray diffraction microscopy. *J. Electron Spectrosc. Relat. Phenom.* **2009**, *170*, 4–12. [CrossRef] [PubMed]
22. Son, S.-K.; Young, L.; Santra, R. Impact of hollow-atom formation on coherent X-ray scattering at high intensity. *Phys. Rev.* **2011**, *A83*, 033402. [CrossRef]
23. Blake, C.; Phillips, D.C. Effects of X-irradiation on single crystals of myoglobin. In Proceedings of the Symposium on the Biological Effects of Ionising radiation at the Molecular Level, Brno, Czechoslovakia, 2–6 July 1962; pp. 183–191.
24. Ravelli, R.B.-G.; Leiros, H.K.; Pan, B.; Caffrey, M.; McSweeney, S. Specific radiation damage can be used to solve macromolecular crystal structures. *Structure* **2003**, *11*, 217–224. [CrossRef]
25. Nanao, M.H.; Sheldrick, G.M.; Ravelli, R.B.G. Improving radiation-damage substructures for RIP. *Acta Cryst.* **2005**, *D61*, 1227–1237. [CrossRef] [PubMed]
26. Galli, L.; Son, S.K.; Barends, T.R.; White, T.A.; Barty, A.; Botha, S.; Boutet, S.; Caleman, C.; Doak, R.B.; Nanao, M.H.; et al. Towards phasing using high X-ray intensity. *IUCrJ* **2015**, *2*, 627–634. [CrossRef] [PubMed]
27. Galli, L.; Son, S.-K.; Klinge, M.; Bajt, S.; Barty, A.; Bean, R.; Betzel, C.; Beyerlein, K.R.; Caleman, C.; Doak, R.B.; et al. Electronic damage in S atoms in a native protein crystal induced by an intense X-ray free-electron laser pulse. *Struct. Dyn.* **2015**, *2*, 041703. [CrossRef] [PubMed]
28. Barty, A.; Caleman, C.; Aquila, A.; Timneanu, N.; Lomb, L.; White, T.A.; Andreasson, J.; Arnlund, D.; Bajt, S.; Barends, T.R.M.; et al. Self-terminating diffraction gates femtosecond X-ray nanocrystallography measurements. *Nat. Photonics* **2012**, *6*, 35–40. [CrossRef] [PubMed]
29. Lomb, L.; Barends, T.R.M.; Kassemeyer, S.; Aquila, A.; Epp, S.W.; Erk, B.; Foucar, L.; Hartmann, R.; Rudek, B.; Rolles, D.; et al. Radiation damage in protein serial femtosecond crystallography using an X-ray free-electron laser. *Phys. Rev. B Condens. Matter Mater. Phys.* **2011**, *84*, 21411. [CrossRef] [PubMed]
30. Nass, K.; Foucar, L.; Barends, T.R.; Hartmann, E.; Botha, S.; Shoeman, R.L.; Doak, R.B.; Alonso-Mori, R.; Aquila, A.; Bajt, S.; et al. Indications of radiation damage in ferredoxin microcrystals using high-intensity X-FEL beams. *J. Synchrotron Radiat.* **2015**, *22*, 225–238. [CrossRef] [PubMed]

31. Giacovazzo, C.; Ladisa, M.; Siliqi, D. Crystal structure solution of proteins by direct methods: An automatic procedure for SIR-MIR and SIRAS-MIRAS cases. *Acta Cryst.* **2002**, *A58*, 598–604. [CrossRef]
32. Lomb, L.; Barends, T.R.; Kassemeyer, S.; Aquila, A.; Epp, S.W.; Erk, B.; Foucar, L.; Hartmann, R.; Rudek, B.; Rolles, D.; et al. De novo protein crystal structure determination from X-ray free-electron laser data. *Nature* **2014**, *505*, 244–247. [CrossRef]
33. Hau-Riege, S.P.; Bennion, B.J. Reproducible radiation-damage processes in proteins irradiated by intense X-ray pulses. *Phys. Rev.* **2015**, *E91*, 022705. [CrossRef] [PubMed]