Biomolecular imaging and electronic damage using X-ray free-electron lasers

Harry M. Quiney* and Keith A. Nugent

The challenges involved in determining the structures of molecules to atomic resolution in non-crystalline samples using X-ray free-electron laser pulses are formidable[1]. Proposals to determine biomolecular structures from diffraction experiments using femtosecond X-ray free-electron laser pulses involve a conflict between the incident brightness required to achieve diffraction-limited atomic resolution and the electronic and structural damage induced by the illumination. Significant advances have already been made, however, in the design and preparation of experiments using fourth-generation sources[2] and the corresponding structural analysis of diffraction data. Here we show that previous estimates of the conditions under which biomolecular structures may be obtained in this manner are unduly restrictive, because they are based on a coherent diffraction model that is not appropriate to the proposed interaction conditions. A more detailed imaging model derived from optical coherence theory and quantum electrodynamics is shown to be far more tolerant of electronic damage. The nuclear density is employed as the principal descriptor of molecular structure. The foundations of the approach may also be used to characterize electrodynamical processes by carrying out scattering experiments on complex molecules of known structure.

The difficulties posed by diffraction data collected from single molecules dropped with random orientations into the path of a freely propagating X-ray free-electron laser (XFEL) pulse have led to the development of schemes by which two-dimensional diffraction patterns may be assembled into complete three-dimensional diffraction sets[3,4]. Such approaches enable the accumulation of diffraction data corresponding to identical targets, increasing the signal-to-noise ratio, particularly in the large-angle scattering sector of the data that carries the high-resolution information about molecular structure. The inevitable Coulomb explosion of molecules subjected to such pulses has also been addressed by detailed computational modelling of the interaction dynamics[1,5,6]. This has led to proposals for the incorporation of sacrificial tampers to delay the Coulomb explosion of the scattering target[7] or temporal gating of data acquisition to reduce the degradation of the structural information due to the molecular disintegration[8].

The feasibility of successful molecular structure determination from diffraction data using XFEL sources has been assessed in each of these studies by calculating some variant of the crystallographic R-factor,

$$R = \frac{\sum_{\mathbf{u}} \left| \sqrt{I_{\text{real}}(\mathbf{u})} - \sqrt{I_{\text{ideal}}(\mathbf{u})} \right|}{\sum_{\mathbf{u}} \sqrt{I_{\text{ideal}}(\mathbf{u})}}$$

in which $I_{\text{real}}(\mathbf{u})$ is the simulated intensity at spatial frequency $\mathbf{u}$ including the effects of damage to the scattering target, $I_{\text{ideal}}(\mathbf{u})$ is the corresponding intensity distribution in the absence of damage,

and the summations over $\mathbf{u}$ include all discrete samples included in the three-dimensional set of diffraction data.

The measure of data quality may be further refined, but its principal role is to provide a quantitative estimate of how closely the crystallographic model matches the experimental data. The requirement that $R \leq 0.15$ has been suggested[1] as a 'rule of thumb' by which diffraction data obtained from femtosecond XFEL experiments on isolated molecules should be regarded as possessing sufficient information to obtain a molecular structure with a spatial resolution determined by the maximum measured scattering angle. The R-factor has guided all computational studies regarding the pulse requirements for X-ray diffraction imaging of single biological molecules[1,5].

Implicit in the R-factor are the assumptions that the incident illumination and the diffracted wave possess full spatial and temporal coherence, independent of the nature of the matter–radiation interaction; these are implicit components of the crystallographic model against which the data are assessed. The validity of these assumptions is not supported, however, by a detailed consideration of the interaction physics that describes an encounter between a molecule and an intense XFEL pulse. This critical assessment pertains to the proposed experiments to determine molecular structures, even if the illumination exhibits full spatial and temporal coherence or if the scattering interaction takes place over so short a time that the nuclear framework is completely unaffected. A key component of the formulation in this Letter is the observation that the time-dependent evolution of the electron density imparts on the scattered wave the statistical characteristics of a partially coherent wavefield, which we incorporate explicitly in our analysis.

Two critical assumptions are made in our approach. The first is that the collection of diffraction data originates from a matter–radiation interaction such that we adopt the Born–Oppenheimer approximation and the nuclei are fixed in space throughout the encounter. Detailed simulations[1,6] put an upper limit on this period of 5 fs, which we assume is to be achieved either by pulse shaping or data gating. The second assumption is that a three-dimensional diffraction set has been assembled[3,4]. Each distinct molecular orientation must be associated with a sufficient number of two-dimensional molecular projections that the stochastic distribution of molecular electronic vacancies created by photoionization and Auger emission may be represented by on-site statistical atomic averages. This is also an implicit assumption in existing simulations[6] on which we have based our electyrodynamical model.

As we are concerned with reproducing detailed molecular scattering properties rather than dynamical averages, we depart from the earlier treatment[8] by using a model atomic basis of spherically averaged orbital electron densities, $\rho_{Z\gamma}(\mathbf{r} - \mathbf{R}^Z_{m_Z})$, located at nuclear positions $\mathbf{R}^Z_{m_Z}$, where Z labels the atomic species and

ARC Centre of Excellence for Coherent X-ray Science, School of Physics, The University of Melbourne, Victoria, 3010, Australia.
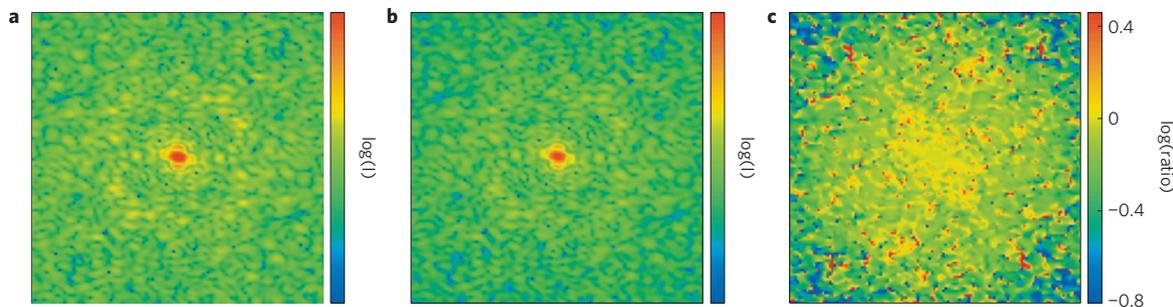*e-mail: quiney@unimelb.edu.au.

Figure 1 | Diffraction simulations and the impact of electronic damage. A full three-dimensional diffraction pattern from the bacteriorhodopsin molecule is used in this work, simulating the form of data that would be recovered after a successful alignment of the data from XFEL interactions with many randomly orientated molecules. **a**, The diffraction pattern from the molecule with the damage mechanisms turned off. **b**, The diffraction pattern obtained when the molecule is illuminated by a 5 fs pulse containing 10^{12} photons and based on the damage mechanisms described in the text. **c**, The logarithm of the ratio of the damaged to undamaged pattern, normalized so that the intensities at $|\mathbf{q}| = 0$ are of equal value. The reduction in the scattered intensity at high diffraction angles is apparent. Note that the ratio is highly non-uniform, so that a simple re-scaling is not able to recover the correct pattern. Adjacent red and blue pixels indicate regions where the ratio diverges indicating that a zero in the diffraction pattern has been removed owing to the effects of partial coherence; there are many such regions throughout the data. The width of the array in these calculations corresponds to a spatial resolution of 1.04 Å.

γ labels the orbital shell. Note that this re-states the problem of recovering structure to one of determining the nuclear positions. The impact of damage, then, only plays a role in the manner in which the desired information is transferred from the sample to the detector plane. We show here that this viewpoint offers considerable advantages. In our formalism, the coefficients $a_{Z\gamma}(t)$ define the time variation in the occupancy of $\rho_{Z\gamma}(\mathbf{r} - \mathbf{R}^Z_{m_Z})$, averaged over all equivalent sites. The time-dependent electron density is represented in the form

$$\rho(\mathbf{r}, t) = \sum_{Z\gamma} a_{Z\gamma}(t) \sum_{m_Z} \rho_{Z\gamma}(\mathbf{r} - \mathbf{R}^Z_{m_Z}) \qquad (1)$$

where m_Z denotes an atom of atomic weight Z located at $\mathbf{R}^Z_{m_Z}$. This representation is augmented by continuum approximation for the density of electrons trapped by the residual ionic charge of the molecule, which is treated as a separate species. The essence of optical coherence theory is to recognize that the field may vary on timescales well within the observation time. Equation (1) suggests that, even in the case of illuminating the sample with a coherent field, dynamical effects within the sample will result in a scattered field that is best described using the theory of partial coherence. Here, we make this connection explicit.

The coherence properties of the scattered wavefield are incorporated in a matrix, $\mathbf{A}$, the elements of which are defined by the statistical averages

$$\mathbf{A}_{Z\gamma, Z'\gamma'} = \left\langle I(t) a_{Z\gamma}(t) a_{Z'\gamma'}(t) \right\rangle$$

where $I(t)$ represents the time-dependent intensity of the incident pulse. The integrated intensity, $I(\mathbf{q})$, corresponding to momentum transfer, $\mathbf{q}$, may be written in the form

$$I(\mathbf{q}) = \sum_Z \sum_{Z'} T_Z(\mathbf{q}) A_{Z,Z'}(q) T^*_{Z'}(\mathbf{q})$$

where

$$T_Z(\mathbf{q}) = \sum_{m_Z} \exp\left(-i\mathbf{q} \bullet \mathbf{R}^Z_{m_Z}\right) \qquad (2)$$

$q = |\mathbf{q}|$, $A_{Z,Z'}(q) = \sum_{\gamma, \gamma'} \mathbf{A}_{Z\gamma, Z'\gamma'} f_{Z\gamma}(q) f_{Z'\gamma'}(q)$ and $f_{Z\gamma}(q)$ is the spherically symmetric elastic X-ray scattering factor for $\rho_{Z\gamma}(\mathbf{r})$. We

note that $T_Z(\mathbf{q})$ is readily generalized to include thermal effects on nuclear motion by convolving it with Gaussian ellipsoids. One can foresee further elaborations in which the effects of molecular dissociation precipitated by interaction with an XFEL pulse are incorporated by a similar modification of $T_Z(\mathbf{q})$. As foreshadowed, in this model, all of the structural information is contained within the vector $T_Z(\mathbf{q})$ that is specified purely by nuclear positions, and all electrodynamical information is contained within $A_{Z,Z'}(q)$; one may determine one of these quantities from measurements of $I(\mathbf{q})$ if the other is known to sufficient accuracy. Our attention here is restricted to the determination of molecular structures, but we note that $A_{Z,Z'}(q)$ is a smooth, continuous function of q, so that electronic damage may also be characterized from measurements of $I(\mathbf{q})$ in systems for which $T_Z(\mathbf{q})$ is known. This feature may facilitate the incorporation of recently reported experimental information on the femtosecond electronic response of atoms as part of an integrated imaging strategy[9].

Under the interaction conditions of interest, $\mathbf{A}_{Z\gamma, Z'\gamma'}$ is poorly approximated by $\mathbf{A}_{Z\gamma, Z'\gamma'} \simeq \bar{a}_{Z\gamma} \bar{a}_{Z'\gamma'}$, for fixed constants $\bar{a}_{Z\gamma}$ and $\bar{a}_{Z'\gamma'}$. As a consequence, $I(\mathbf{q})$ possesses the statistical characteristics of a partially coherent diffraction pattern and the R-factor provides an inappropriate measure of its information content.

It has recently been demonstrated[10] that explicit incorporation of models of partial coherence into the solution of inverse problems may markedly improve the quality of reconstructions using iterative, propagation-based techniques. To relate the optical properties of the molecule with a far-field scattered intensity, it proves convenient to write $I(\mathbf{q})$ as the modal expansion[11]

$$I(\mathbf{q}) = \sum_{k=1}^M \eta_k \psi_k(\mathbf{q}) \psi^*_k(\mathbf{q}) \qquad (3)$$

where the real, non-negative parameters, η_k, and the corresponding modes, $\psi_k(\mathbf{q})$, are solutions of an M-dimensional integral equation defined by the mutual optical intensity of optical coherence theory[11], making explicit the connection foreshadowed after equation (1). Each mode assumes the general form

$$\psi_k(\mathbf{q}) = \sum_Z T_Z(\mathbf{q}) \sum_\gamma c^k_{Z\gamma} f_{Z\gamma}(q) \qquad (4)$$

where $c^k_{Z\gamma}$ represents the effective occupancy of shell γ in atom of type Z in mode $\psi_k(\mathbf{q})$. A scheme to determine η_k and $c^k_{Z\gamma}$ under typical experimental conditions appears in the Methods section.
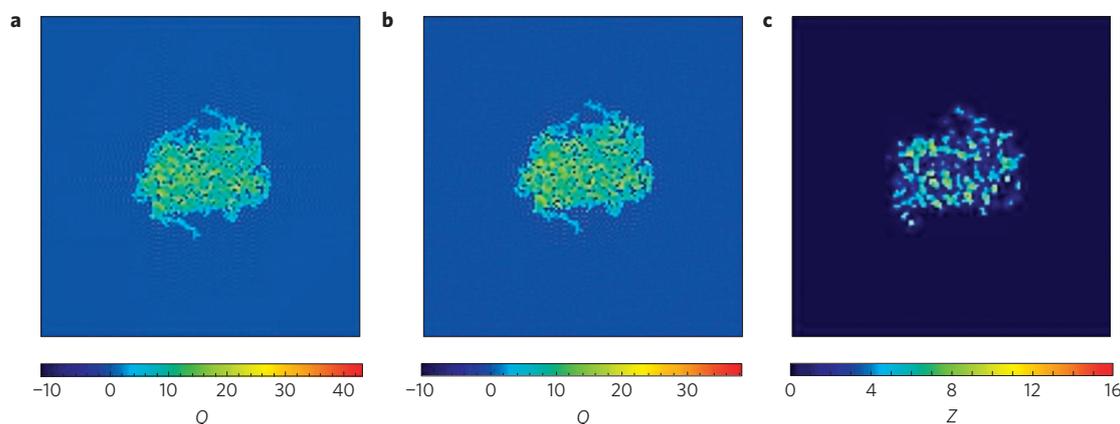
Figure 2 | Original and recovered structures of bacteriorhodopsin. The methods described here allow the direct recovery of the distribution of atomic species within a molecule using data from a molecule containing electronic damage. This figure shows the information returned by the method. **a**, Two-dimensional projection, Q, of the nuclear charge (in atomic units) through bacteriorhodopsin. The data in **a** are evaluated as the Fourier transform of $T(\mathbf{q})$, which is constructed directly from its definition and that of $T_Z(\mathbf{q})$, equation (2), using the nuclear positions of bacteriorhodopsin obtained from the protein data bank. The visible structure outside the molecular envelope carries super-resolution information that is available for recovery from high-resolution XFEL diffraction data and reflects the fact that the nuclear positions do not generally sit at the central pixel position. For comparison purposes, each nuclear position is assigned the value of the nuclear charge at that point. **b**, The reconstruction of the projected charge, Q, from simulated diffraction data corresponding to the interaction of bacteriorhodopsin with a 5 fs XFEL pulse of 10^{12} 5 keV photons. The weak attenuation of the super-resolution fringes by the spherical support function is visible; the reconstruction corresponds to the input structure in all significant structural details and confirms that the approach described here recovers the full structural information. **c**, A detail of a slice through the full three-dimensional reconstruction illustrating the recovery of the details of the molecular structure. In this case, the nuclear charge in the slice is plotted. One can directly associate a particular atomic species with a given site in the molecule by simply reading off the scale. To aid the visualization, the fringes were eliminated from this plot by convolving the reconstruction with a three-dimensional Gaussian function.

Without any detailed calculation, however, one may immediately deduce important qualitative features of the charge distribution from the structure of $\psi_k(\mathbf{q})$ that lead to the partially coherent character of $I(\mathbf{q})$. Each coefficient $c_{Z\gamma}^k$ is necessarily real, but may be either positive or negative, because the elements of the set $\psi_k(\mathbf{q})$ are orthonormal. The partially coherent character of $I(\mathbf{q})$, which is electrodynamical in origin, is reproduced precisely in this approach by a small number of electrostatic charge distributions that represent multi-centre shell polarizations of the various atomic types. Each of these modal density distributions, $\rho_k(\mathbf{r})$, is obtained as the inverse Fourier transform of the corresponding optical mode, $\psi_k(\mathbf{q})$.

The explicit inclusion of the electronic properties in the diffraction process has the interesting advantage that one is able to remove the intermediate step of recovering the electronic density distribution[12]. We are, instead, able to formulate the problem in such a manner that the nuclear structure is revealed directly. In common with a recent study of diffractive imaging using partially coherent X-ray sources[13], we here recast the partially coherent scattering information within an equivalent single-mode model inverse problem that carries the required information. It is clear from equation (4) that any single optical mode, $\psi_k(\mathbf{q})$, carries all of this information, because the unknown parameters, $\mathbf{R}_{m_Z}^Z$, appear within $T_Z(\mathbf{q})$. The inverse problem to be solved is of the form $T(\mathbf{q})T^*(\mathbf{q}) = I'(\mathbf{q})$, where $T(\mathbf{q})$ and $I'(\mathbf{q})$ are defined in the Methods section. The Fourier transform of $T(\mathbf{q})$ is subject to the constraint that it must represent a nuclear distribution function of known spatial extent; the solution reveals the positions and identities of the constituent atoms without constructing an intermediate electron density.

Three-dimensional diffraction data for bacteriorhodopsin were generated using the electrodynamical model[6] described in the Methods section, assuming that 10^{12} 10 keV photons were incident on the target in each 5 fs pulse. The number of photons in the pulse determines the damage levels in the simulations, but we should make clear that we do not include the statistical uncertainties that would be associated with an experimental diffraction pattern.

Compared with a similar calculation in the absence of any electronic damage, this resulted in an R-factor of 0.17. Figure 1 shows samples of the undamaged diffraction pattern (Fig. 1a) and damaged pattern (Fig. 1b). These data display the reduction in the high-q scattering resulting from depletion of the core electrons through the pulse. Figure 1c shows the ratio of these two patterns and illustrates that the impact of the electronic damage imposes an uneven and structure-dependent modification to the diffraction data indicating that a simple re-scaling of the pattern[12], although appealing, is not adequate but may be usefully regarded as a first approximation to a full solution.

A projection through the reconstruction is shown in Fig. 2, where it can be seen that the atomic number and location of each atomic species are recovered essentially perfectly. Indeed, each atomic position may be located beyond the formal resolution of the diffraction pattern using a centroiding process on the fringes around each nucleus; if the nucleus is centrally located in each pixel then these fringes are absent and their relative amplitudes encode the location of the nucleus in the form of the fringes. The concept of super-resolving crystallographic data has recently emerged in a rather different but related context[14]. There is no need to construct a representation of $\rho(\mathbf{r},0)$, or to determine $\mathbf{R}_{m_Z}^Z$ from it by crystallographic model-building. This appears to be the first reconstruction of its type; the molecular structure of a complex biomolecule has been recovered directly under interaction conditions in which the electron density is so comprehensively damaged that the usual working rules of protein crystallography are no longer valid. To underline this point, Fig. 2c shows a detail from a slice through the reconstruction, demonstrating that the atomic species information is automatically recovered using this algorithm.

The application of the XFEL to the determination of molecular structure is an extremely exciting endeavour. However, it is also one that offers very considerable scientific and technical challenges. A very fundamental challenge is the need to understand the role of the interaction of the molecule with the incident field. This Letter demonstrates that recovery of molecular structure in the presence

of damage is most readily achieved by adopting a model that reflects the detailed interaction physics, rather than the usual crystallographic assumption of full coherence. Our approach is predicated on the need for the damage mechanisms to be well characterized, either by kinetic modelling, or by inversion of diffraction data obtained from scattering targets of known structure. As such, we submit an important obstacle in the realization of single-molecule structural biology using XFEL sources has been removed.

Methods

Data simulation. The three-dimensional intensity distribution of X-ray photons scattered from a molecule of bacteriorhodopsin, $I(\mathbf{q})$, was simulated using methods previously published[6].

The modal distributions defined by equation (3) are determined in the following manner. A non-orthogonal basis of orbital densities with which to expand the modes is

$$\Phi_{Z\gamma}(\mathbf{r}) = \sum_{m_Z} \rho_{Z\gamma}(\mathbf{r} - \mathbf{R}_{m_Z}^Z)$$

so that $\psi_k(\mathbf{r}) = \sum_{Z\gamma} c_{Z\gamma}{}^k \Phi_{Z\gamma}(\mathbf{r})$, and $c_{Z\gamma}{}^k$ is an associated expansion coefficient that, in the present context, must be a real quantity. These coefficients and the corresponding real and non-negative eigenvalues, η_k, that appear in equation (3) are determined as the solutions of a generalized matrix eigenvalue equation of the form $\mathbf{JC} = \boldsymbol{\eta}\mathbf{SC}$. The diagonal matrix, $\boldsymbol{\eta}$, contains the eigenvalues, the columns of $\mathbf{C}$ contain the eigenvectors and the matrix elements of $\mathbf{J}$ and $\mathbf{S}$ are defined by

$$J_{Z_1\gamma_1, Z_2\gamma_2} = N_{Z_1} N_{Z_2} \sum \Psi_{\gamma_1\gamma_3}^{Z_1} A_{Z_1\gamma_3, Z_2\gamma_4} \Psi_{\gamma_4, \gamma_2}^{Z_2}$$
$$S_{Z_1\gamma_1, Z_2\gamma_2} = N_{Z_1} \Psi_{\gamma_1\gamma_2}^{Z_1} \delta_{Z_1 Z_2}$$

where N_Z is the number of atoms of nuclear charge Z, $\delta_{ZZ'}$ is the Kroneker delta and

$$\Psi_{\gamma_1\gamma_2}^Z = \int \rho_{Z\gamma_1}(\mathbf{r}) \rho_{Z\gamma_2}(\mathbf{r}) d\mathbf{r}$$

For consistency with the diffraction simulations, the orbital density integrals were evaluated using a spherical-atom model derived from Slater-type orbitals[15] within the tight-binding approximation. We note, however, that the formalism is readily extended to more detailed numerical treatments of the electron density simply by interpreting γ to be a label identifying N-electron electronic state functions including the effects of electronic relaxation, correlation and relativistic corrections. The three most significant modes contribute in the ratio 0.93:0.05:0.01 in the present model of bacteriorhodopsin.

Structure recovery. We have introduced the modes $\psi_k(\mathbf{q})$ through equation (4). The coherent mode formulation of optical coherence theory[11] then writes the coherence function, known in this context as the mutual optical intensity, in the form

$$J(\mathbf{q}_1, \mathbf{q}_2) = \sum_k \eta_k \psi_k(\mathbf{q}_1) \psi_k^*(\mathbf{q}_2) \qquad (5)$$

where the measured intensity is given by $I(\mathbf{q}) = J(\mathbf{q}, \mathbf{q})$, which is equation (3). If we substitute equations (4) into (5) and re-organize a little it adopts the form

$$J(\mathbf{q}_1, \mathbf{q}_2) = \sum_k \eta_k \sum_Z Z T_Z(\mathbf{q}_1) \sum_{Z'} Z' T_{Z'}^*(\mathbf{q}_2) \sum_\gamma \frac{1}{Z} c_{Z\gamma}^k f_{Z\gamma}(q_1) \sum_{\gamma'} \frac{1}{Z'} c_{Z'\gamma'}^k f_{Z'\gamma'}(q_2) \qquad (6)$$

We introduce the functions $T(\mathbf{q}) = \sum_Z Z T_Z(\mathbf{q}_2)$ and $\mu(q, Z) = \sum_\gamma 1/Z(c_{Z\gamma}^k f_{Z\gamma}(q))$. The sum in the function $\mu(q, Z)$ is weighted by the inverse of the atomic number so that it is approximately independent of the atomic number,

$$\mu(q, Z) \approx \mu(q) \qquad (7)$$

allowing us to write the mutual optical intensity in the approximate form

$$J(\mathbf{q}_1, \mathbf{q}_2) \simeq T(\mathbf{q}_1) T^*(\mathbf{q}_2) \mu(q_1)\mu(q_2) \sum_k \eta_k \qquad (8)$$

We may now use the expression for the intensity to conclude that to reconstruct molecular structure from diffraction data including electronic damage, we initially investigate the existence of a solution, $T(\mathbf{q})$, of the equation

$$I(\mathbf{q}) = B(q) T(\mathbf{q}) T^*(\mathbf{q}) \qquad (9)$$

for the function $B(q)$ defined in our theory by $B(q) = |\mu(q)|^2 \sum_k \eta_k$. The relatively simple modification suggested by equation (9) is a remarkable result and shows

that the effects of electronic damage may be incorporated into a scheme designed to solve directly for the nuclear positions, removing all reference to the electronic coordinates. As shown further on, the assumption that leads to equations (8), (7), can be removed at the last stage of the solution but we have found it to be unnecessary; the approximation is excellent for a complex biomolecule.

The left-hand side of equation (9), which represents the measured intensities, $I(\mathbf{q})$, and the product $T(\mathbf{q})T^*(\mathbf{q})$ are both non-negative definite functions. As we consider $I(\mathbf{q})$ to be constructed from modes that carry the structural information, any solution of the inverse problem $T(\mathbf{q})T^*(\mathbf{q}) = I(\mathbf{q})/B(\mathbf{q}) = I'(\mathbf{q})$ must satisfy the nuclear distribution constraints.

To a good approximation, we note that $A_{ZZ'}(q) \simeq ZZ' \bar{f}_{ZZ'}(q)$, where $\bar{f}_{ZZ'}(q)$ is a smooth, slowly varying function of q, for which $\bar{f}_{ZZ'}(0) = 1$. Defining $B(q)$ by

$$B(q) = \frac{\sum_{ZZ'} T_Z(\mathbf{q}) A_{ZZ'}(q) T_{Z'}^*(\mathbf{q})}{T(\mathbf{q}) T^*(\mathbf{q})}$$

we see that for $q \to 0$, $B(q) \simeq 1$.

Our focus on biomolecules suggests the use of a particularly simple model to specify a functional form for $B(q)$ that possess the required characteristics, including in the neighbourhood of the zeros of $I(\mathbf{q})$. The discrete nuclear positions in equation (2) are replaced by continuous spherical uniform nuclear charge distributions of finite radius, R, so that equation (2) is readily evaluated using a standard integral, leading to the approximation

$$T_Z(\mathbf{q}) T_{Z'}^*(\mathbf{q}) = \frac{9 N_Z N_{Z'}}{(qR)^6} \left\{ \sin(qR) - qR\cos(qR) \right\}^2 \qquad (10)$$

where N_Z is the number of atoms of charge Z. For $q \to 0$, this function has the limit $N_Z N_{Z'}$ and vanishes for $q \to \infty$ because no two atoms of different charge may share the same spatial position. In the case $Z = Z'$, we require that the function should possess the limit N_Z for $q \to \infty$ as there are N_Z occurrences of unity appearing in the summation that defines $T_Z(\mathbf{q})T_{Z'}^*(\mathbf{q})$ that are not present in the continuum approximation leading to equation (10). These terms are restored by adding a correction of the form $N_Z[1 - \exp(-\kappa q^2)]$. The selection of the value of κ proves to be not very critical provided that the function rises rapidly to the value N_Z for small q; we use $\kappa \simeq 0.1 R^2$.

Having specified $B(q)$, the solution of the resulting inverse problem for $T(\mathbf{q})$ is essentially conventional, and employs the iterative hybrid input–output and error-reduction algorithms[16], supplemented by the judicious use of 'charge-flipping'[17] in the early stages of the structure recovery to establish the effective exterior support surface of the molecule. The iterations are initialized by a uniform charge distribution with a radius of 30 Å, and typically converge within 1,000 iterations.

In view of the approximations involved in the specification of $B(q)$, we note that a final refinement of the structure may be carried out directly from the measured intensities, $I_0(\mathbf{q})$. This may be achieved by using the structure derived from $T(\mathbf{q})$ to initialize a set of trial parameters using a conventional, gradient-driven error-minimization scheme, such as is commonly used in crystallography, to fit $I_0(\mathbf{q})$, using the modal amplitudes to propagate the trial intensity, $I(\mathbf{q})$. All modal amplitudes and their gradients with respect to the nuclear coordinates are readily evaluated using elementary analytical methods. We have confirmed that such an error-minimization scheme recovers the target structure if, but only if, the procedure is initialized sufficiently close to the solution that the global minimum is located; in practice this means that each nuclear position in the trial solution should lie within the Wigner–Seitz sphere centred at the target position. Using simulated noiseless data, this subsequent refinement proves to be unnecessary, but it may provide a measure of additional robustness against the effects of measurement noise in experimental data, when it becomes available.

References

1. Neutze, R., Wouts, R., van der Spoel, D., Weckert, E. & Hajdu, J. Potential for biomolecular imaging with femtosecond X-ray pulses. *Nature* **406,** 752–757 (2000).
2. Chapman, H. N. *et al.* Femtosecond diffractive imaging with a soft-X-ray free-electron laser. *Nature Phys.* **2,** 839–843 (2006).
3. Fung, R., Shneerson, V., Saldin, D. K. & Ourmazd, A. Structure from fleeting illumination of faint spinning objects in flight. *Nature Phys.* **5,** 64–67 (2009).
4. Elser, V. Noise limits on reconstructing diffraction signals from random tomographs. *IEEE Trans. Inf. Theor.* **55,** 4715–4722 (2009).
5. Hau-Riege, S. P., London, R. A., Huldt, G. & Chapman, H. N. Pulse requirements for X-ray diffraction imaging of single biological molecules. *Phys. Rev. E* **71,** 061919 (2005).
6. Hau-Riege, S. P., London, R. A. & Szoke, A. Dynamics of biological molecules irradiated by short X-ray pulses. *Phys. Rev. E* **69,** 051906 (2004).

7. Jurek, Z. & Faigel, G. The effect of tamper layer on the explosion dynamics of atom clusters. *Eur. Phys. J. D* **50,** 35–43 (2008).
8. Jurek, Z., Faigel, G. & Tegze, M. Dynamics in a cluster under the influence of intense femtosecond hard X-ray pulses. *Eur. Phys. J. D* **29,** 217–229 (2004).
9. Young, L. *et al.* Femtosecond electronic response of atoms to ultra-intense X-rays. *Nature* **466,** 56–61 (2010).
10. Whitehead, L. W. *et al.* Diffractive imaging using partially coherent X rays. *Phys. Rev. Lett.* **103,** 243902 (2009).
11. Wolf, E. New theory of partial coherence in the space-frequency domain. 1. Spectra and cross spectra of steady-state sources. *J. Opt. Soc. Am.* **72,** 343–351 (1982).
12. Hau-Riege, S. P., London, R. A., Chapman, H. N., Szoke, A. & Timneanu, N. Encapsulation and diffraction-pattern-correction methods to reduce the effect of damage in X-ray diffraction imaging of single biological molecules. *Phys. Rev. Lett.* **98,** 198302 (2007).
13. Dilanian, R. A. *et al.* Diffractive imaging using a polychromatic high-harmonic generation soft-X-ray source. *J. Appl. Phys.* **106,** 023110 (2009).
14. Schroder, G. F., Levitt, M. & Brunger, A. T. Super-resolution biomolecular crystallography with low-resolution data. *Nature* **464,** 1218–1222 (2010).
15. Slater, J. C. Atomic shielding constants. *Phys. Rev.* **36,** 57–64 (1930).
16. Fienup, J. R. Phase retrieval algorithms—a comparison. *Appl. Opt.* **21,** 2758–2769 (1982).
17. Oszlanyi, G. & Suto, A. The charge flipping algorithm. *Acta Crystallogr. A* **64,** 123–134 (2008).

Author contributions

H.M.Q. carried out most of the theoretical analysis and computational implementation. Both authors contributed to the conceptual formulation of the research and to the writing of the Letter.

Additional information

The authors declare no competing financial interests. Reprints and permissions information is available online at http://npg.nature.com/reprintsandpermissions. Correspondence and requests for materials should be addressed to H.M.Q.