Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Phasing in the presence of severe site-specific radiation damage through dose-dependent modelling of heavy atoms

M. Schiltz,[a,b]* P. Dumas,[c] E. Ennifar,[c] C. Flensburg,[b] W. Paciorek,[b] C. Vonrhein[b] and G. Bricogne[b]

[a]Laboratoire de Cristallographie, École Polytechnique Fédérale de Lausanne, EPFL–FSB–IPMC–LCR, Bâtiment BSP, CH-1015 Lausanne, Switzerland, [b]Global Phasing Ltd, Sheraton House, Castle Park, Cambridge CB3 0AX, England, and [c]Institut de Biologie Moléculaire et Cellulaire–IBMC, UPR 9002 du CNRS, 15 Rue R. Descartes, F-67084 Strasbourg CEDEX, France

Correspondence e-mail: marc.schiltz@epfl.ch

Received 2 February 2004
Accepted 18 March 2004

The case of a brominated RNA crystal structure determination in which standard three-wavelength MAD phasing was unsuccessful because of fast X-ray-induced debromination was reinvestigated [Ennifar *et al.* (2002), *Acta Cryst.* D**58**, 1262–1268]. It was found that if the data are kept unmerged and if a dose-stamp is associated with each reflection measurement, dose-dependent occupancies can be refined for the Br atoms. Such a parametrization has been implemented in the macromolecular phasing program *SHARP*. Refining such dose-dependent occupancies on an unmerged data set gave a dramatic improvement, even for SAD phases from only the first wavelength (peak), and resulted in a good electron-density map after solvent flattening. The adverse effect of radiation damage has been turned into a beneficial one. The crucial difference is made by the use of unmerged data: phasing power is generated through the intensity differences of symmetry-related reflections recorded at different doses, *i.e.* corresponding to different states of the X-ray-induced debromination. This approach should prove useful in all situations of experimental phasing where site-specific radiation damage occurs unavoidably and undesirably and not only in cases in which radiation damage is purposely being created in order to demonstrate its potential usefulness.

1. Introduction

From the very beginning of protein crystallography, radiation damage has been an important limitation on the data quality that can be achieved. Although the development of cryogenic freezing techniques has been a major step in reducing this problem, radiation damage is still a matter of growing concern as smaller and smaller samples with larger and larger unit cells are studied with ever higher photon fluxes and flux densities delivered by insertion devices on third-generation synchrotrons. A number of detailed studies (Burmeister, 2000; Ravelli & McSweeney, 2000; Weik *et al.*, 2000, 2002; Ennifar *et al.*, 2002) have shown that radiation damage first shows up at discrete well localized sites within a macromolecular crystal. Only in a second phase does radiation damage affect the whole of the crystal structure, causing long-range disorder with its characteristic consequences of overall intensity decay and loss of high-resolution reflections.

In protein crystals, site-specific damage occurs in the form of breakage of disulfide bonds, decarboxylation of aspartate and glutamate residues, loss of hydroxyl groups from tyrosines and loss of methylthio groups of methionines (Burmeister, 2000). Heavier atoms such as selenium in selenosubstituted proteins (Rice *et al.*, 2000), bromine in brominated nucleic acids (Ennifar *et al.*, 2002), metals in metalloproteins (Schlichting *et al.*, 2000; Berglund *et al.*, 2002) and heavy atoms in isomorphous derivatives (Evans *et al.*, 2003) are particularly

sensitive because of their high absorption cross-sections. The situation becomes even worse when the anomalous scattering of these atoms is to be exploited for MAD phasing because of the jumps in photoelectric absorption cross-sections at the edge energies. It has been suggested (Rice *et al.*, 2000) that collecting complete data at the peak wavelength should be a priority in order to at least obtain optimally measured SAD data. Yet even this strategy requires that at least one complete data set can be collected before substantial site-specific radiation damage shows up. Data at further wavelengths can be collected if the crystal does not exhibit particular sensitivity to radiation.

Radiation damage at anomalously scattering sites is particularly problematic because it has a deleterious effect on precisely those atoms from which one is relying to obtain the phasing signal. Unfortunately, the optical theorem (James, 1969) establishes a direct proportionality between atomic absorption coefficients (μ_a) and imaginary anomalous scattering factors (f''). Maximizing the magnitude of the phasing signal (*i.e.* maximizing f'') therefore inevitably also increases the site-specific radiation damage.

On the other hand, the site-specific radiation damage that occurs to heavier atoms (including sulfurs) can be used to phase macromolecular crystal structures in a pseudo-MIR fashion (Ravelli *et al.*, 2003; Evans *et al.*, 2003). This method essentially consists of collecting a first data set on a fresh crystal (the pseudo-native), followed by intense exposure to radiation and the subsequent collection of a second data set containing site-specific radiation damage (a pseudo-derivative). The feasibility of this approach has been established (Ravelli *et al.*, 2003) on model cases of both proteins (using the site-specific damage of disulfide bonds) and brominated nucleic acids. Again, for this method to work one needs to collect an essentially complete data set that is affected as little as possible by radiation damage.

There are, however, more sinister cases in which the onset of X-ray-induced structural changes is so rapid that it becomes impossible to collect even a single complete data set that is unaffected by radiation damage. Such a situation has been reported for a brominated RNA molecule for which MAD phasing was unsuccessful because of rapid X-ray-induced debromination (Ennifar *et al.*, 2002). In fact, debromination was so rapid that it was already impossible to extract SAD phases from just the first merged data set (peak wavelength). In this study, we demonstrate that even in such severe cases it is possible to recover phase information, provided that the data are kept unmerged and that the dose-dependence of site-specific radiation damage is adequately modelled.

2. Materials and methods

2.1. Sample preparation and X-ray data collection

Since full details of all experimental aspects and X-ray data collections have been presented by Ennifar *et al.* (2002), only a short summary will be given here.

The molecule studied is a chemically synthetized 23-nucleotide RNA fragment corresponding to the dimerization initiation site (DIS) of HIV-1(Lai) genomic RNA, with 5-bromouridine substituted for uridine at position 3. The molecule crystallizes in space group $C222_1$, with unit-cell parameters $a = 27.5$, $b = 116.4$, $c = 95.4$ Å and a dimer in the asymmetric unit. All X-ray diffraction data were collected from a single cryocooled (at 100 K) crystal at the bending-magnet beamline BM30 of the European Synchrotron Radiation Facility (ESRF), Grenoble, France using a MAR image-plate detector. A total of four data sets were collected at energies near the Br *K* absorption edge in the following order: (i) peak wavelength ($\lambda = 0.920184$ Å), (ii) inflection-point wavelength ($\lambda = 0.920374$ Å), (iii) fast low-resolution data collection at the peak wavelength ($\lambda = 0.920184$ Å) and (iv) high-energy remote wavelength ($\lambda = 0.917755$ Å). The raw diffraction data were integrated with the *HKL* package (Otwinowski & Minor, 1997).

Although the two bromine sites in the asymmetric unit could be located using the anomalous differences of the first data set (peak wavelength), all attempts to solve the structure by MAD phasing remained unsuccessful. Nothing unusual was apparent from the data-reduction statistics. Eventually, the structure was determined by molecular replacement using independent data from a crystal of non-brominated molecules (PDB code 1jjm). A *post mortem* analysis of the data from the brominated structure revealed that fast debromination had occurred during data collection. A careful analysis showed that significant debromination was already apparent during the first data collection. It was further shown that the evolution of debromination was consistent with an exponential decay model, where the occupancies (q) of the Br atoms vary as a function of X-ray dose (d) as

$$q(d) = \exp(-d/D). \quad (1)$$

The debromination decay constants (D) were estimated to be $D = 7.4\,(0.8)$ MGy at the peak wavelength and $D = 9.2\,(2.6)$ MGy at the inflection-point wavelength.

2.2. Data reduction, heavy-atom refinement and phasing

In the following, we only consider the first data set (peak wavelength). For this data set, a total of 153 raw 1° rotation data frames were collected in an uninterrupted fashion. The maximum resolution of the data is 2.46 Å. The data were scaled with the program *SCALA* (Evans, 1993) from the *CCP*4 program suite (Collaborative Computational Project, Number 4, 1994) in the usual way, *i.e.* by minimizing the disagreement between symmetry-equivalent reflections. Overall and resolution-dependent scale factors (pseudo-*B*-factors) were refined at intervals of 5° in spindle rotation, with a smooth interpolation in between. Relevant data-reduction statistics are summarized in Table 1. Parameters for the two Br atoms were refined in the program *SHARP* (La Fortelle & Bricogne, 1997; Bricogne *et al.*, 2003), starting from known atomic positions. Phase-improvement by solvent flattening was carried out using the *SOLOMON* (Abrahams & Leslie, 1996)

Table 1
Data-reduction statistics for the peak-wavelength data set.

X-ray data collection	
Beamline	ESRF BM30
Wavelength (Å)	0.920184
Detector	MAR IP
Exposure time per frame (s)	60
Angular increment per frame (°)	1
Total No. frames	153
Data reduction	
No. measured reflections	30712
No. unique reflections	5692
Resolution limits (Å)	22.1–2.46
Completeness (%)	96.7
Anomalous completeness (%)	94.4
R_{sym}	0.055
R_{meas}	0.067

Table 2
Data-reduction statistics for single merged data sets in scenarios II and III.

	Scenario II		Scenario III
	Set 1	Set 2	Average values
No. measured reflections	14722	15991	2007
No. unique reflections	5275	5410	1885
Completeness (%)	90.6	92.5	26.0
Anomalous completeness (%)	76.8	82.5	1.1

procedure. Four different scenarios for data reduction, heavy-atom refinement and phasing were explored.

2.2.1. Scenario I: merging all data. All data were merged in the usual way. Positional parameters, occupancies and isotropic atomic displacement parameters were refined for the two Br atoms in the usual way. A single parameter was refined for anomalous local non-isomorphism (*i.e.* model errors; La Fortelle & Bricogne, 1997). This scenario would thus correspond to standard SAD phasing.

2.2.2. Scenario II: splitting the data into two merged sets. The scaled data were split into two sets before merging. One set corresponds to the first 76 data frames and the second set corresponds to the final 77 data frames. The two data sets were merged separately (Table 2) and declared as individual 'crystals' within the hierarchical organization of data implemented in *SHARP* (La Fortelle & Bricogne, 1997). This means that for each of the two Br atoms separate occupancy factors were refined in each data set. However, the positional parameters and atomic displacement parameters were constrained to remain identical for both data sets. For each data set, a single parameter was refined for anomalous local non-isomorphism. In addition, two parameters were refined in the second dat set to model local and global non-isomorphism with respect to the first data set. This was the procedure used by Evans *et al.* (2003) to accommodate radiation damage in a triiodide derivative of elastase.

2.2.3. Scenario III: splitting up the data into 15 merged sets. In this scenario, the scaled data were split up into 15 sets before merging. Each set corresponds to ten consecutive data frames. The last three data frames were discarded in order to keep an identical number of frames in all 15 data sets. Including these last three data frames in the final data set did not produce any significant change in the results obtained. Each data set was merged separately and declared as an individual 'crystal' within the hierarchical organization of data implemented in *SHARP*. Anomalous pairs were kept separately as individual F_+ and F_- reflections, *i.e.* the usual transformation into $F_{mid} = (F_+ + F_-)/2$ and $\Delta_{ano} = F_+ - F_-$ (La Fortelle & Bricogne, 1997) was not applied. For each data set, a single parameter was refined to account for global non-isomorphism. There was in fact very little redundancy within each data set (Table 2), so that this procedure almost amounts to using unmerged data. For each of the two Br atoms, separate occupancy factors were refined in each data set. Positional parameters and atomic displacement parameters were constrained to remain identical across all data sets.

Table 3
Refined heavy-atom parameters and phasing statistics.

SF, solvent flattening. FOM, figure of merit. Correl., real-space correlation coefficient between electron-density maps from experimental phases and known structure. Standard uncertainties are indicated in parentheses for the parameters q_0 and β_d in scenario IV. No values for phasing powers are reported for scenario IV as these quantities are only meaningfully defined for merged data.

		Scenario II			
	Scenario I	Set 1	Set 2	Scenario III	Scenario IV
Site 1					
x	0.4576	0.4586		0.4583	0.4583
y	0.3152	0.3152		0.3150	0.3152
z	0.5491	0.5491		0.5492	0.5494
B (Å^2)	46.2	48.9		33.5	29.6
Occupancy	0.57	0.67	0.53	0.82–0.34	
q_0					0.85 (0.04)
β_d					0.0063 (0.0005)
Site 2					
x	0.9374	0.9365		0.9329	0.9327
y	0.6678	0.6676		0.6679	0.6682
z	0.5523	0.5521		0.5518	0.5516
B (Å^2)	49.3	46.9		38.6	42.9
Occupancy	0.50	0.56	0.43	0.75–0.33	
q_0					0.78 (0.05)
β_d					0.0056 (0.0007)
Phasing power					
Isomorphous			0.507	0.72–0.10	
Anomalous	0.98	1.004	0.655		
Before SF					
FOM	0.27	0.30		0.39	0.39
Correl.	0.30	0.35		0.39	0.40
After SF					
FOM	0.72	0.74		0.71	0.69
Correl.	0.37	0.45		0.66	0.78

2.2.4. Scenario IV: using unmerged data. Here, the data were not merged at all. Instead, a dose-stamp was associated with each single observation. In this mode, *SHARP* treats each observation as an independent measurement, *i.e.* each observation corresponds to one circle in the Harker construction with its own heavy-atom structure factor. The heavy-atom model corresponding to each single observation is then modeled by dose-varying occupancies of the form

$$q(d) = q_0 \exp(-\beta_d d), \qquad (2)$$

where d corresponds to the dose stamp of the given observation; the zero-dose occupancy q_0 and the decay factor β_d are refineable parameters that are common to all observations.

Thus, for each site, two occupancy parameters (q_0 and β_d), one isotropic atomic displacement parameter and a set of three coordinates were refined. For a given unique reflection, each observation was treated as a single factor in the likelihood function, which is a product of Rice functions similarly to the multiple isomorphous replacement case [equation (10) in La Fortelle & Bricogne, 1997]. A single parameter was refined to account for global non-isomorphism effects in the form of equations (19) and (20) of La Fortelle & Bricogne (1997), *i.e.* the non-isomorphism variances (V^{glo}) and Luzzati-attenuation factors (D) were constrained to remain identical across all observations of a given unique reflection. The use of such a likelihood function amounts to ignoring correlated non-isomorphism and model errors between the various observations of a given unique reflection, an approximation which is only of limited validity (see discussion below).

Owing to the limited resolution, it was not possible to place the data on an absolute scale. As a consequence, the refined occupancy factors and zero-dose occupancy factors (q_0) have no direct physical meaning as far as their absolute values are concerned. The data were put on the same scale as that used in the refinements of Ennifar *et al.* (2002) so that our results can be compared with theirs.

2.2.5. Estimation of X-ray dose. Estimating the X-ray dose received by the crystal at each given moment in time is not a trivial task and we have adopted the following pragmatic approach. From the data-collection parameters it was possible to exactly reconstruct the time history for all reflection measurements, *i.e.* to evaluate the total time t that the crystal had been exposed to the incident X-ray beam when a given observation was recorded on frame number k,

$$t = \left[(k-1) + \frac{\varphi - \varphi_k}{\Delta\varphi}\right]\Delta t, \tag{3}$$

where φ corresponds to the spindle-axis position of the centroid of the reflection, φ_k is the spindle-axis position at the start of the exposure of frame k, $\Delta\varphi$ is the rotation angle per frame (constant) and Δt is the exposure time per frame

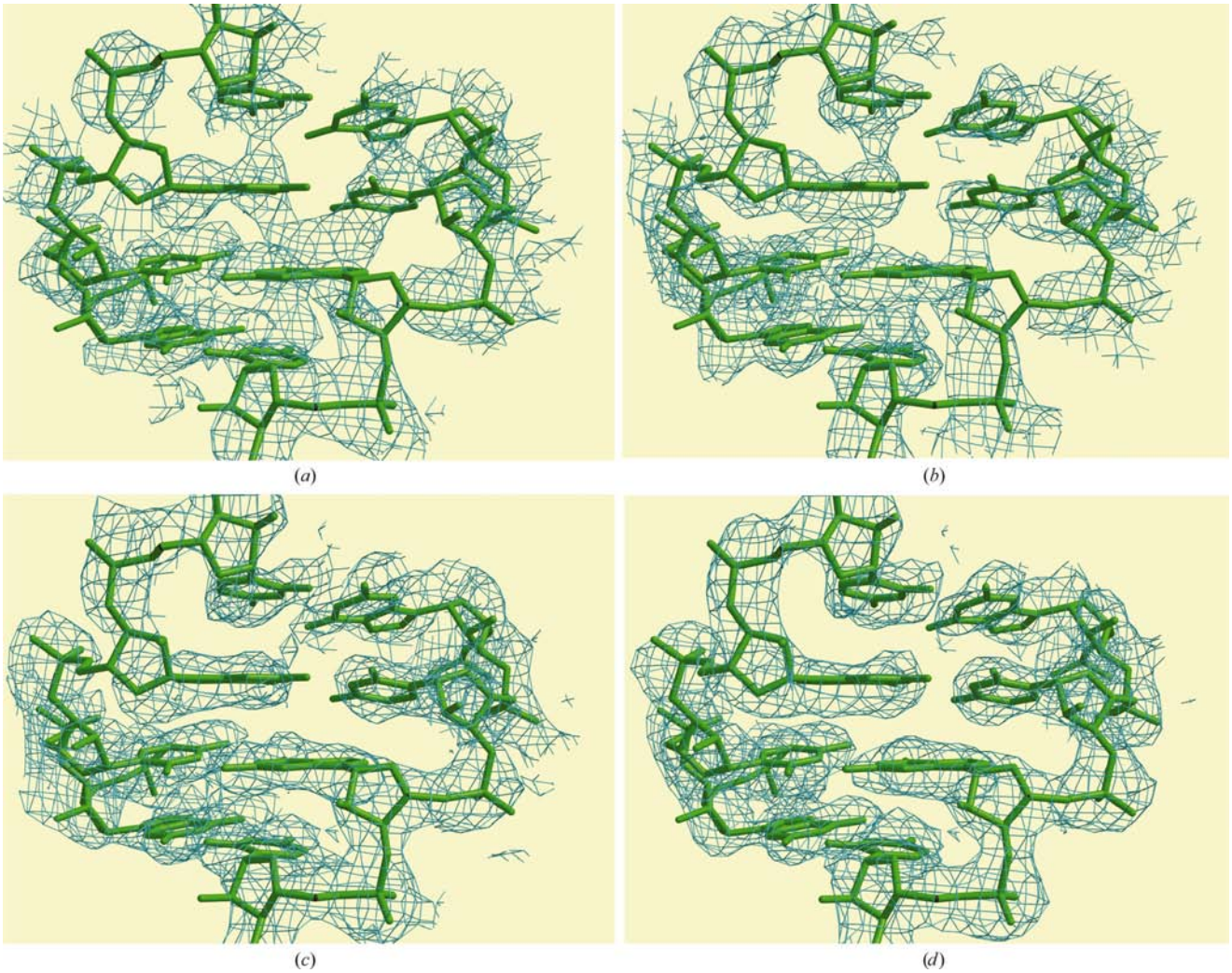

Figure 1
Representative views of the electron-density maps (2.46 Å resolution) after phasing with *SHARP* and solvent flattening with *SOLOMON*. (*a*) Scenario I: standard SAD, *i.e.* one merged data set. (*b*) Scenario II: splitting up the data into two merged data sets. *c* Scenario III: splitting up the data into 15 merged data sets. (*d*) Scenario IV: using unmerged data with an exponential decay model for occupancies. Countours are at a level of 1σ. The refined structural model of the brominated RNA molecule is superimposed on the maps (residues *A*10–*A*14 and *B*12–*B*15).

(constant). However, dose can only be considered to be proportional to the exposure time if (i) the incident beam has constant intensity, (ii) the volume of the crystal that is bathed in the incident beam is constant at all times and (iii) the variation of flux density in the incident beam is small on the scale of the crystal size (*i.e.* all parts of the crystal receive the same flux density irrespective of its orientation in the beam). These conditions were not fully satisfied in the present experiment (i) because of the natural decay of the stored synchrotron beam, (ii) because an automatic beam-intensity optimization procedure was periodically carried out during the experiment and (iii) because the crystal had a rather anisotropic shape with dimensions of 150 × 100 × 10 µm (Ennifar *et al.*, 2002). It is therefore not enough to simply evaluate the dose from ionization-chamber readings of the incident beam, as this would not necessarily yield the X-ray dose actually received by the crystal. However, the total number of intensity counts recorded on each data frame provides a good measure of the X-ray dose received by the crystal during the corresponding exposure. Provided that dark noise has been substracted, the intensity counts on a data frame largely arise from X-ray photons scattered by the crystal (both Bragg and diffuse scattering) and are thus proportional to the X-ray dose received by the crystal. It is important to note that the intensity counts in all pixels of the frames have to be taken into account (not just the Bragg reflections): as the crystal degrades, the Bragg peaks will lose intensity but the diffuse (background) scattering will increase correspondingly. Thus, the dose stamp for a reflection measured on data frame number k after a total exposure time t was evaluated as

$$d(t, k) = K\left(\sum_{i=1}^{k-1} I_i + \frac{t - t_{k-1}}{t_k - t_{k-1}} I_k\right), \tag{4}$$

where t_k is the total exposure time at the end of frame number k (with $t_0 = 0$) and I_i corresponds to the total number of intensity counts recorded on frame number i. The only unknown in this equation is the factor K which converts intensity counts on data frames to X-ray dose. No attempt was

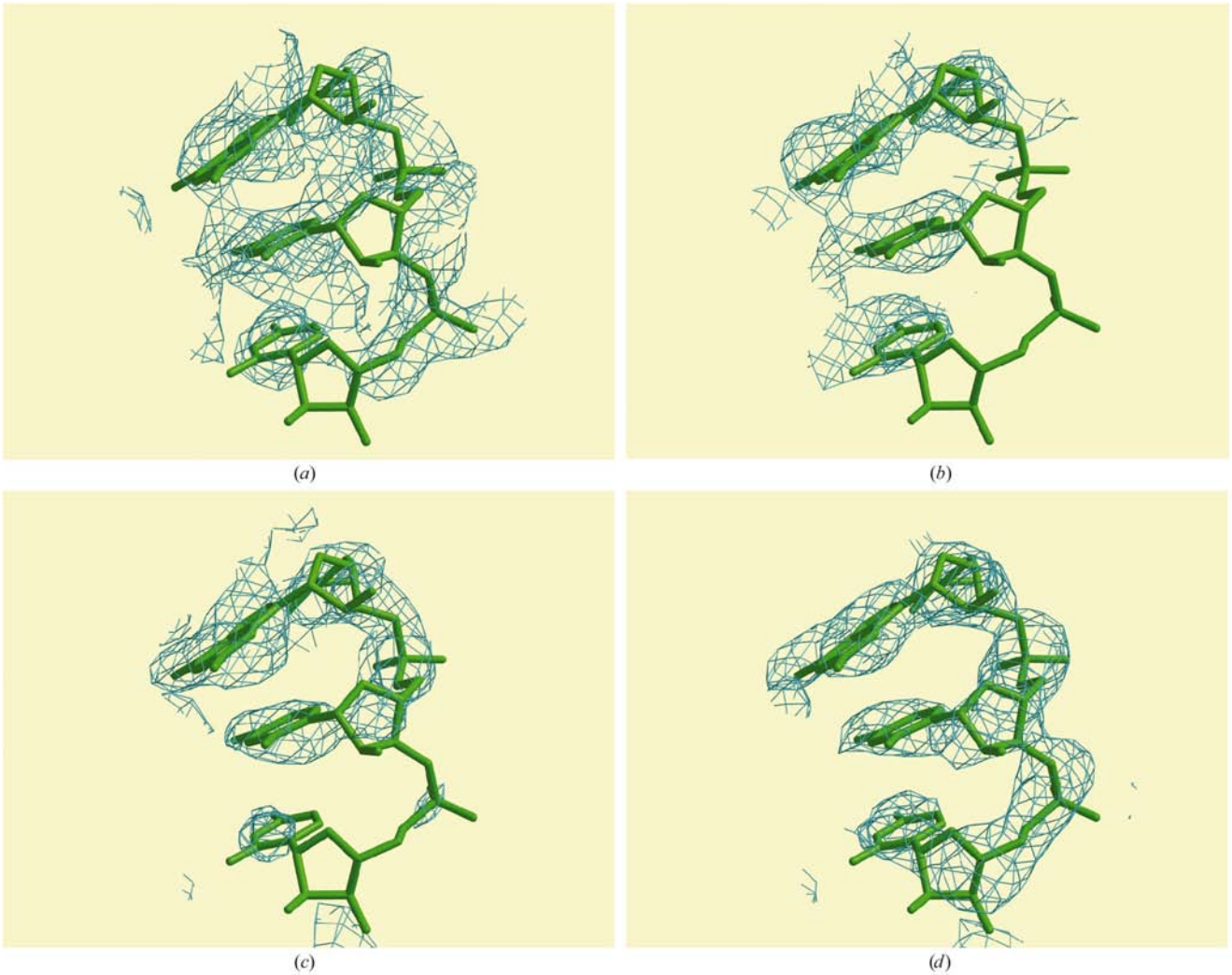

Figure 2
A different portion of the electron-density maps. (*a*) Scenario I. (*b*) Scenario II. (*c*) Scenario III. (*d*) Scenario IV. Countours are at a level of 1σ. The refined structural model of the brominated RNA molecule is superimposed on the maps (residues *A*4–*A*6).

made to estimate the absolute value of dose and so we simply set

$$K = \bar{I}^{-1}, \tag{5}$$

i.e. the dose stamps are normalized by the average dose per frame. This turns the factor β_d into a dimensionless quantity. However, we mention that Ennifar *et al.* (2002) have roughly estimated that the total dose during this data collection was 6.3 MGy. This would thus correspond to an average dose of 42 kGy per data frame and this conversion factor can be used to compute absolute values for β_d. The file containing both dose and time stamps for each observation was produced by a small *ad hoc* program that was run on the scaled but unmerged reflection file output by *SCALA*.

3. Results

Results from the various heavy-atom refinement and phasing computations are summarized in Table 3. The heavy-atom parameters are readily refined in all four scenarios. For scenario 1, the occupancies of the Br atoms correspond to average values over the duration of the data collection. This is seen if the refined values are compared with the corresponding values of scenario II. The phasing statistics are poor and the structure could not have been solved in this scenario (see Figs. 1 and 2). This supports the conclusions reached by Ennifar *et al.* (2002). In scenario II, each of the two data sets is essentially complete (Table 2) and this case would therefore correspond to the pseudo-MIR approach advocated by Ravelli *et al.* (2003). The differences in the refined occupancy values between the two data sets clearly reflect the decay of the bromine sites. Thus, in addition to the anomalous differences, phasing power is also generated through the isomorphous differences between the two data sets. The phasing statistics are thus improved with respect to scenario I but the electron-density map after solvent flattening remains noisy and large stretches of it are still poorly resolved (see Figs. 1 and 2).

In contrast, for scenario III and scenario IV the phases were of a quality such as to yield fairly good electron-density maps after density modification (see Figs. 1 and 2). In scenario III, a total of 15 occupancy factors were refined for each Br atom. The values are plotted in Fig. 3. It can be seen that the occupancies of the two bromine sites decrease substantially during the data collection. The decay is consistent with the simple exponential model that has been implemented in scenario IV. When the refined occupancy values of scenario III are fitted to an exponential decay model (see Fig. 3), we obtained decay constants that are in very good agreement with the β_d values refined by *SHARP* in scenario IV.

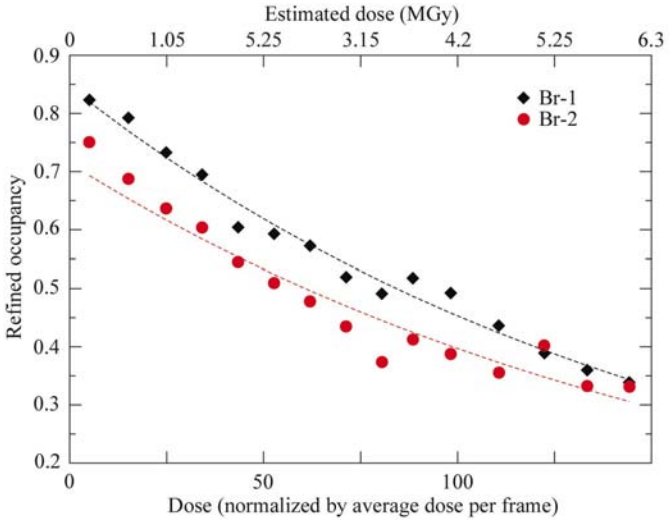

Figure 3
Refined values of the occupancies for each of the two Br atoms in scenario III. The dashed lines represent exponential fits with fitted parameters $0.846\exp(-0.00625d)$ for Br-1 and $0.714\exp(-0.00588d)$ for Br-2, in good agreement with the q_0 and β_d values refined in scenario IV (see Table 3). The absolute values for dose (upper horizontal axis) were obtained from data reported by Ennifar *et al.* (2002) and should only be considered as rough estimates. With these estimates, the exponential decay parameters for Br-1 and Br-2 are $\beta = (6.7\ \text{MGy})^{-1}$ and $\beta = (7.1\ \text{MGy})^{-1}$, respectively, in good agreement with the D_1 value of 7.40 (0.8) MGy reported by Ennifar *et al.* (2002).

4. Discussion

A more careful analysis of the data statistics gives a hint as to why scenario I (standard SAD) was unsuccessful in solving the structure. In Fig. 4, the cumulative overall and anomalous completeness of the data are plotted as a function of frame number (*i.e.* as a function of time or dose). This figure should be compared with Fig. 3. It can be seen that at around frame number 60, debromination was already very substantial (Fig. 3), whereas at the same time anomalous completeness barely exceeded 50%. In other words, the anomalous scatterers had already substantially decayed at a time when phase information (through anomalous differences) had been recorded for only about half the reflections. The phasing of a very substantial number of reflections was thus compromised.

When site-specific X-ray damage affects a crystal structure during data collection, one is faced with a special situation in

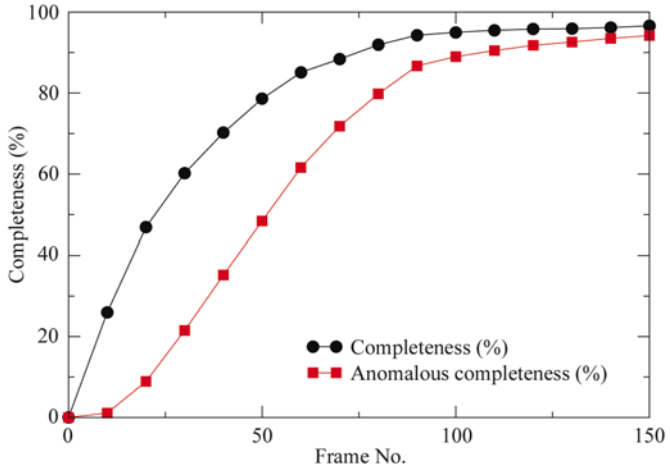

Figure 4
Completeness and anomalous completeness of the data set as a function of the number of recorded data frames. Anomalous completeness is the percentage of acentric reflections for which an anomalous differences (*i.e.* Bijvoet pairs) have been measured.

which the diffraction data do not correspond to a single structure. Instead, the data can be considered to arise from a mixture of several closely related structures. The balance between the different single structures in the mixture varies smoothly with dose. In the simple case reported here, the mixture consists of just two structures, one which is completely brominated and one which is completely debrominated. Since it is not possible to instantaneously collect a complete set of data (at least not with monochromatic X-ray techniques), the various reflections in the resulting data set sample the mixture of structures at different compositions. Thus, there is no such thing as a single structure pertaining to such a data set and it would be incorrect to refine a single structural model against these data. The same statement of course also holds for the refinement of the heavy-atom substructure. This emphasizes the necessity of implementing a dose-dependent parametrization of structural parameters. In the present case, we are in a particularly favourable situation because site-specific radiation damage essentially only affects the occupancy parameters of the Br atoms. The modelling of structural changes is thus especially simple. However, there are no fundamental difficulties in implementing more sophisticated parametrizations, for example to model displacements of atoms or groups of atoms as a function of time or dose.

With the current practice of recording diffraction data by the single-axis rotation method with an area detector, it is almost always the case that data sets with a certain degree of redundancy (multiplicity) are collected. Redundancy is often achieved before completeness, *i.e.* several symmetry-equivalent observations of certain reflections are recorded while for other reflections no observations have yet been measured. Thus, redundancy is usually a byproduct of striving to collect a complete data set. It can be seen from Figs. 4 and 5 that the present case is no exception in this regard. Redundancy has the advantage that for a given unique reflection observations are recorded at different times during the data collection, *i.e.* they sample different compostions of the mixture of structures. Merging the data therefore amounts roughly to concocting a data set that corresponds to a structure which is time-averaged over the duration of the experiment, although this is only true within certain limitations. If, for example, the first measurement of a given reflection is recorded only after substantial site-specific damage has already occurred in the crystal, its merged reflection intensity will correspond to a different time-average than that of a reflection for which observations were recorded early during data collection.

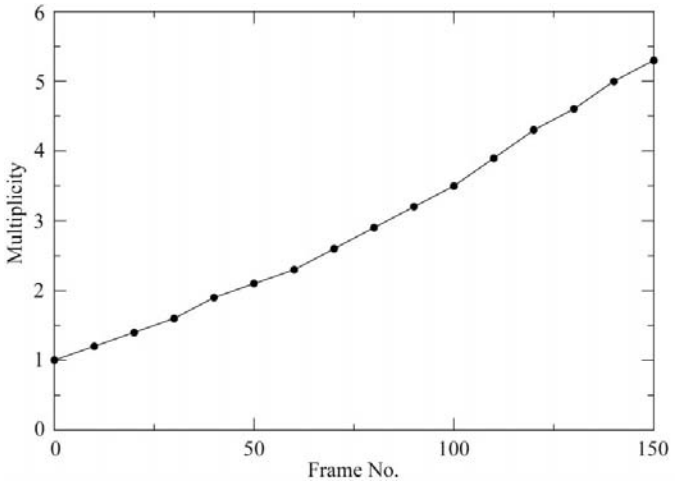

Figure 5
Average multiplicity (redundancy) of reflections in the data set as a function of the number of recorded data frames. Multiplicity is the number of measured Laue group equivalent observations per unique reflection.

We believe that the above considerations provide some explanations as to why scenario II was unsuccessful. Although the first data set (frames 1–76) in this scenario is essentially complete, the decay of the Br atoms between the beginning and the end of this data set is very substantial. This is therefore not a homogeneous data set and by merging the data one artificially produces a data set which corresponds to bromine occupancies that lie somewhat on the low side. Thus, the isomorphous differences with the second data set are levelled out and their potential phasing power is diminished. It should also be noted that the refined B factors for the Br atoms are systematically higher for scenarios I and II compared with scenarios III and IV. Again, this is most likely owing to the fact that the merged data sets in scenarios I and II are not homogeneous.

This highlights the fact that by merging all symmetry-equivalent observations, one in fact averages out some of the genuine 'isomorphous' differences that are present in the data. The conclusion is that symmetry-related observations which are measured at different times during the data collection are in fact no longer equivalent because they correspond to different structures (or mixtures of structures). These intensity differences can be used to model and refine the dose-dependence of site-specific structural modifications and thus to generate isomorphous phasing power. This is the key to why scenarios III and IV produce such dramatic improvements. Although anomalous differences have been collected for only about half the reflections at frame number 60 (Fig. 4), the completeness exceeds 80%. This implies that for most reflections at least one symmetry-related measurement has been recorded at an early stage (corresponding to a structure with high bromine occupancies). For most of these reflections another symmetry-related measurement was recorded at a late stage (corresponding to a structure with low bromine occupancies). Thus, even those reflections which are badly phased by the anomalous signal will have substantial isomorphous phasing power.

Scenario IV is clearly to be preferred over scenario III, because it has a much reduced number of parameters (four occupancy parameters *versus* 30 parameters in scenario III), while essentially reproducing the same variations for occupancies. It also yielded a better map. The fact that the phasing was successful with a minimum number of parameters gives strong support to the validity of the exponential decay model used by Ennifar *et al.* (2002). With the present internal organization of *SHARP* (Flensburg *et al.*, 2004), it would be trivial to implement alternative decay models, as long as they are parametrized functions.

A remark is in order concerning the refinement of non-isomorphism parameters in scenarios III and IV. The error model used in *SHARP* assumes that the effects of all sources of non-isomorphism are uncorrelated between different observations of a given reflection (La Fortelle & Bricogne, 1997; Bricogne *et al.*, 2003). In essence, a diagonal approximation is used for the non-isomorphism covariance matrix. Such an approximation may be grossly inadequate since that part of non-isomorphism which arises from radiation damage is expected to be strongly correlated across observations that are closely spaced in time. However, in the present case we are in a particular favourable situation where there is very little radiation damage apart from the debromination. As a consequence, the non-isomorphism parameters refine to rather small values. It thus appears that debromination is the main structural change and since this is explicitly modelled it does not have to be taken into account by non-isomorphism variances and covariances. In the particular case under discussion here, the diagonal approximation to the non-isomorphism covariance matrix is therefore justified. In less simple cases it may prove necessary to resort to multivariate likelihood functions which are capable of accommodating adequate patterns of covariances between the various observations (Bricogne, 2000; Pannu *et al.*, 2003).

5. Concluding remarks

It has been demonstrated in this study that a single data collection on a brominated derivative provides both anomalous and isomorphous phase information which, taken together, are sufficient to solve an RNA structure. Iodinated derivatives might even be more useful for radiation-damage-induced phasing because iodine offers a larger isomorphous signal compared with bromine. Since the rate of radiolysis of brominated and iodinated nucleotides is known to be highly wavelength-dependent (Furusawa *et al.*, 1996; Karnas *et al.*, 1999), it is tempting to speculate that there exists an optimum wavelength such that the combined anomalous and isomorphous phasing powers are maximized. This would amount to determining the f'' value which gives sufficient anomalous signal while at the same time keeping the rate of radiolysis low enough that the decay of bromine or iodine is progressive over the timespan of a single data collection. This approach should also be applicable to solving protein structures *via* incorporation of iodinated tyrosines (Leinala *et al.*, 2002).

It has been shown previously that the adverse effect of radiation damage can be turned into a beneficial one by using radiation-induced site-specific structural changes as a source of isomorphous phase information (Ravelli *et al.*, 2003; Evans *et al.*, 2003). However, these were benchmark cases in which experiments were carried out on well characterized samples of known molecular structures and where the explicit goal was to create and exploit site-specific radiation damage. In contrast, the case reported here is a real-life example in which an unknown structure could not be solved by standard techniques. The simple method that is advocated in the benchmark studies and which consists of recording two consecutive data sets (*i.e.* scenario II) would only have had a limited success here as opposed to the more sophisticated method that we have presented (*i.e.* scenario IV). This approach should prove to be useful in all situations of experimental phasing where site-specific radiation damage occurs unavoidably and undesirably and not only in cases where one is purposely creating radiation damage in order to demonstrate its potential usefulness.

We are grateful to the members of the Global Phasing Consortium for financial support and to Dr Gwyndaf Evans for helpful discussions. We also wish to acknowledge partial financial support from European Commission Grants No. HPRI-CT-1999-50015 within the EXMAD project and No. QLRT-CT-2000-00398 within the AUTOSTRUCT project.

References

Abrahams, J. P. & Leslie, A. G. W. (1996). *Acta Cryst.* D**52**, 30–42.
Berglund, G. I., Carlsson, G. H., Smith, A. T., Szöke, H., Henriksen, A. & Hajdu, J. (2002). *Nature (London)*, **417**, 463–468.
Bricogne, G. (2000). *Proceedings of the Workshop on Advanced Special Functions and Applications, Melfi (PZ), Italy, 9–12 May 1999*, edited by D. Cocolicchio, G. Dattoli & H. M. Srivastava, pp. 315–321. Rome: Arcane Editrice.
Bricogne, G., Vonrhein, C., Flensburg, C., Schiltz, M. & Paciorek, W. (2003). *Acta Cryst.* D**59**, 2023–2030.
Burmeister, W. P. (2000). *Acta Cryst.* D**56**, 328–341.
Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.
Ennifar, E., Carpentier, P., Ferrer, J.-L., Walter, P. & Dumas, P. (2002). *Acta Cryst.* D**58**, 1262–1268.
Evans, G., Polentarutti, M., Djinovic Carugo, K. & Bricogne, G. (2003). *Acta Cryst.* D**59**, 1429–1434.
Evans, P. R. (1993). *Proceedings of the CCP4 Study Weekend. Data Collection and Processing*, edited by L. Sawyer, N. Isaacs & S. Bailey, pp. 114–122. Warrington: Daresbury Laboratory.
Flensburg, C., Paciorek, W., Schiltz, M., Vonrhein, C. & Bricogne, G. (2004). In preparation.
Furusawa, Y., Maezawa, H., Takakura, K., Kobayashi, K. & Hieda, K. (1996). *Acta Oncol.* **35**, 877–882.
James, R. W. (1969). *The Optical Principles of the Diffraction of X-rays.* London: G. Bell & Sons.
Karnas, S. J., Yu, E., McGarry, R. C. & Battista, J. J. (1999). *Phys. Med. Biol.* **44**, 2537–2549.
La Fortelle, E. de & Bricogne, G. (1997). *Methods Enzymol.* **276**, 472–494.
Leinala, E. K., Davies, P. L. & Jia, Z. (2002). *Acta Cryst.* D**58**, 1081–1083.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
Pannu, N. S., McCoy, A. J. & Read, R. J. (2003). *Acta Cryst.* D**59**, 1801–1808.
Ravelli, R. B. G. & McSweeney, S. M. (2000). *Structure*, **8**, 315–328.
Ravelli, R. B. G., Schøder Leiros, H.-K., Pan, B., Caffrey, M. & McSweeney, S. M. (2003). *Structure*, **11**, 217–224.
Rice, L. M., Earnest, T. N. & Brunger, A. T. (2000). *Acta Cryst.* D**56**, 1413–1420.
Schlichting, I., Berendzen, J., Chu, K., Stock, A. M., Maves, S. A., Benson, D. E., Sweet, R. M., Ringe, D., Petsko, G. A. & Sligar, S. G. (2000). *Science*, **287**, 1615–1622.
Weik, M., Bergès, J., Raves, M. L., Gros, P., McSweeney, S., Silman, I., Sussman, J. L., Houée-Levin, C. & Ravelli, R. B. G. (2002). *J. Synchrotron Rad.* **9**, 342–346.
Weik, M., Ravelli, R. B. G., Kryger, G., McSweeney, S., Raves, M. L., Harel, M., Gros, P., Silman, I., Kroon, S. & Sussman, J. L. (2000). *Proc. Natl Acad. Sci. USA*, **97**, 623–628.