

Specific Radiation Damage Can Be Used to Solve Macromolecular Crystal Structures

Raimond B.G. Ravelli,^{1,*}

Hanna-Kirsti Schröder Leiros,² Baocheng Pan,³

Martin Caffrey,³ and Sean McSweeney⁴

¹European Molecular Biology Laboratory (EMBL)

Grenoble Outstation

6 rue Jules Horowitz

B.P. 181

F38042 Grenoble Cedex 9

France

²Protein Crystallography Group

Department of Chemistry

Faculty of Science

University of Tromsø

N-9037 Tromsø

Norway

³Biochemistry, Biophysics, Chemistry

The Ohio State University

100 West 18th Avenue

Columbus, Ohio 43210

⁴European Synchrotron Radiation Facility (ESRF)

B.P. 220

F-38043 Grenoble Cedex

France

Summary

The use of third generation synchrotron sources has led to renewed concern about the effect of ionizing radiation on crystalline biological samples. In general, the problem is seen as one to be avoided. However, in this paper, it is shown that, far from being a hindrance to successful structure determination, radiation damage provides an opportunity for phasing macromolecular structures. This is successfully demonstrated for both a protein and an oligonucleotide, by way of which complete models were built automatically. The possibility that, through the exploitation of radiation damage, the phase problem could become less of a barrier to macromolecular crystal structure determination is discussed.

Introduction

With the genome sequence of an increasing number of organisms in hand, the attention of the biologist is shifting toward functional and structural genomics [1–5], with the ultimate aim of determining the three-dimensional structure of at least one representative member from each protein family. It is reasonable to assume that X-ray crystallography will remain the primary structure determination method for the foreseeable future [6].

Progress in the area of structural genomics based on X-ray crystallography is limited by crystal production [5] and the speed with which structures can be solved [1, 4].

A number of robotic systems are under development to allow full automation of large-scale crystallization experiments, with a view to circumventing the crystallization bottleneck. X-ray data collection is being speeded up as a result of improvements to dedicated synchrotron beamlines and the development of new systems for automatic mounting and alignment of pre-frozen crystals [7, 8].

Despite these improvements, the time required to get good starting phases needed for solving de novo structures still can be limiting. Despite recent successes with fast multiwavelength anomalous dispersion (MAD) experiments, where enough data could be collected for solving a new structure in less than 1 hr [9, 10], the current rate at which new structures are produced at the APS, the ESRF, and SPring-8 is still short, by at least an order of magnitude, of the goals set by the structural genomics initiatives [4]. Selenomethionine (SeMet)-derivatized crystals do not always behave identically to their native counterparts, in that crystals do not always grow to the same size or diffract to the same resolution. Often, it is necessary to reoptimize the crystallization and freezing conditions for the selenoprotein. Further, the phasing power from a MAD data set can be impaired by nonisomorphism introduced by the X-ray beam [6, 11, 12]. The isomorphous replacement method, using conventional heavy atoms such as Pt, Au, and Hg, halide soaks [13], or noble gasses [14], requires laborious screening. Ideally, one would like to solve new structures with a single native crystal.

For extremely well-diffracting crystals (resolution of better than 1.15 Å), direct methods can sometimes play a role in solving the structure ab initio or with the use of minor additional information. In particular, the program Acorn [15] can produce electron density maps of remarkable quality in a matter of minutes with, for example, knowledge of the kind and number of metal ions in a metalloprotein or the sulphur substructure of the protein. The single-wavelength anomalous dispersion (SAD) method exploits the weak anomalous signals of atoms like S and P present in native crystals. These anomalous signals can be obtained when data is collected at longer wavelengths (>1.5 Å). Thus, the sulphur SAD (S-SAD) method has been successfully applied to several cases of well-diffracting crystals, both for demonstration purposes and for new structure determination [16–19]. Investigations are in progress to determine the limitations of the SAD method, where signals as small as $\langle|\Delta F| \rangle / \langle F \rangle = 2\%$ must be collected accurately. These signals increase at longer wavelengths, although unavoidable and detrimental X-ray absorption effects rise in parallel.

The damaging effects of X-ray absorption are already severe at all wavelengths used at our brightest synchrotron radiation sources, even for cryocooled crystals. In the course of data collection at these sources, the dif-

*Correspondence: ravelli@embl-grenoble.fr

fractive power of the crystal is reduced, the mosaicity, overall B factor, and cell volume rise, and, eventually, all higher-order reflections are lost. In addition to these general effects, some highly specific changes can occur. These include disulphide bond cleavage and the loss of definition of carboxyl groups in proteins [11, 20, 21]. Both the general and the specific X-ray damage effects result in a loss of perfect isomorphism throughout data collection, thus hampering traditional phasing techniques like MAD and SAD.

In this paper, we present a novel method for phasing native crystals. It exploits specific structural changes caused by powerful X-ray beams from third generation synchrotron undulator beamlines. Accurate and redundant data were collected on a protein crystal with an attenuated X-ray beam. In between two such data sets, the crystal was exposed to the unattenuated X-ray beam to an accumulated dose of 1×10^7 Gy. Existing software packages were used to identify and refine a large number of specific X-ray-induced structural changes. Phases were obtained from which virtually the complete structure could be built automatically. We have named the method radiation damage-induced phasing (RIP). In order to further investigate the generality of the method, we performed a similar experiment on a crystal of an oligonucleotide. Some highly susceptible sites were found showing changes that were used to obtain high-quality phases that were essentially identical to those from the refined model. The possibility of RIP being incorporated into a general method for de novo determination of macromolecular crystal structures and/or aiding in the improvement of macromolecular phases is discussed.

Results

Data Collection and Phasing

We chose benzylamine-inhibited bovine trypsin as a model protein because crystals are readily obtainable and diffract to high resolution. The 23.5 kDa monomeric protein contains six disulphide bridges, one of which, Cys191-220, was shown to be the most radiation sensitive [22]. Two highly redundant 1.4 Å data sets were collected on the ESRF beamline ID14-4 with a 24-fold attenuated beam, resulting in a total absorbed dose of about 1.3×10^6 Gy per data set. In between the data sets, the rotating crystal was exposed to the unattenuated beam for a total of 2 min, corresponding to an absorbed dose of 1×10^7 Gy.

The integrated intensities were merged into two final data sets, hereafter referred to as “before” and “after” the X-ray burn. The R_{merge} was calculated to compare the differences in intensities between the two data sets. An R_{merge} value of 11.0% was obtained, which is much higher than the R_{sym} values of 4.7% and 5.0% obtained by comparing the intensities of symmetry-equivalent reflections within the before and after data sets, respectively. This, together with the increase in unit cell dimensions [11], crystal mosaicity, and the R_{sym} values in the highest resolution shell (Table 1), clearly indicated that the crystal had suffered from radiation damage during the X-ray burn, although the after crystal still diffracted to high resolution.

Table 1. Data Collection Statistics for Trypsin

	“Before” Burn	“After” Burn
Beamline	ID14-4 at ESRF	
Wavelength (Å)	0.9763	
Resolution (last shell) (Å)	45–1.4 (1.47–1.4)	
P2 ₁ ,2 ₁ ,2 ₁ cell (Å)		
a	54.209	54.241
b	56.658	56.752
c	66.126	66.204
Mosaicity (°)	0.385	0.410
Wilson B factor (Å)	12.45	12.31
Completeness (%)	99.7 (98.3)	99.6 (97.4)
R_{sym} on intensity (%)	4.7 (22.7)	5.0 (31.2)
Multiplicity	13.4 (12.5)	13.1 (11.7)
$\langle I/\sigma I \rangle$	11.3 (3.6)	11.2 (2.7)
R_{merge} on intensity (%)	11.0 (25.4)	
$\langle \Delta F/\sigma \Delta F \rangle$	4.03 (1.88)	
FOM (centrics/accentrics)	0.26/0.32	
Phasing power (acentrics)	1.31	
Unique reflection set	40,998	40,972
R_{cryst} (R_{free}) (%)	13.8 (16.7)	14.3 (17.3)

Data sets were scaled together with Scala and Scaleit [32], resulting in a similar Wilson B factor for both data sets.

In a similar experiment, a crystal of a DNA/RNA hybrid, 5'-RUp-D(Br)Gp-R(ApGpGpUp)-3', was used. Two data sets were collected on beamline ID14-2, which did not require the use of an attenuated beam. Instead, the X-ray burn was imposed by exposing the rotating crystal to the X-ray beam for 20 min, corresponding to an absorbed dose of about 8×10^6 Gy, in between the before and after data sets, which corresponded to an absorbed dose of 1.5×10^6 Gy each. The R_{merge} between the two data sets was 31.8%. This, plus the resulting loss of resolution, increased unit cell dimensions, and crystal mosaicity, indicates clearly that the hybrid crystals succumbed to radiation damage (Table 2).

In a manner analogous to that of Crick and Magdoff [23], the expected relative root-mean-square intensity change $\sqrt{\langle \Delta I^2 \rangle / \langle I(d_0) \rangle}$ for acentric reflections resulting

Table 2. Data Collection Statistics for the DNA/RNA Hybrid

	“Before” Burn	“After” Burn
Beamline	ID14-2 at ESRF	
Wavelength (Å)	0.933	
Resolution (last shell) (Å)	20–1.2 (1.26–1.2)	20–1.4 (1.48–1.4)
P2 ₁ ,2 ₁ ,2 ₁ cell (Å)		
a	31.233	31.247
b	51.102	51.154
c	70.651	70.834
Mosaicity (°)	0.69	0.73
Completeness (%)	99.5 (98.5)	99.3 (98.1)
R_{sym} on intensity (%)	8.7 (34.1)	9.5 (46.4)
Multiplicity	4.2 (4.2)	4.2 (4.2)
$\langle I/\sigma I \rangle$	10.8 (3.8)	11.0 (3.1)
R_{merge} on intensity (%)	31.7 (55.6)	
$\langle \Delta F/\sigma \Delta F \rangle$	3.05(1.00)	
FOM (centrics/accentrics)	0.38/0.39	
Phasing power (acentrics)	2.64	
Unique reflection set	35,974	22,939
R_{cryst} (R_{free}) (%)	12.4 (15.6)	15.2 (19.9)

No manual adjustments were made to the “after” model, resulting in high values for R_{cryst} and R_{free} .

from specific radiation damage can be estimated with the assumption that the occupancy and B factor, but not the position, of a limited group of atoms are affected by the X-rays:

$$\frac{\sqrt{\langle \Delta I^2 \rangle}}{\langle I(d_0) \rangle} = \sqrt{2} \frac{\sqrt{\langle I(\Delta d) \rangle}}{\sqrt{\langle I(d_0) \rangle}}$$

where $\langle I(d_0) \rangle$ is the average intensity of the reflections for the structure at zero X-ray dose, d_0 , and $\langle I(\Delta d) \rangle$ is the average intensity if the unit cell only contained atoms whose occupancies and B factors were X-ray sensitive. Defining the occupancy and B factors of this subset of n atoms as $q(d)$ and $B(d)$, we can write $\langle I(\Delta d) \rangle$ as follows:

$$\langle I(\Delta d) \rangle = \sum_{j=1}^n f_j^2(s) \{q_j(d_1) \exp[-B(d_1)s^2] - q_j(0) \exp[-B(0)s^2]\}^2$$

where $s = \sin \theta/\lambda$, f_j is the atomic-scattering factor, and d_1 corresponds to the dose absorbed by the crystal during the X-ray burn. Schröder Leiros [22] has reported that the occupancies of the six sulphurs in three disulphide bonds of bovine trypsin were reduced by 26% upon exposure to X-rays. If this was the only change induced by X-ray damage, an intensity change of 10% at $2\theta = 0$ would be expected. The intensity drop becomes larger at higher resolution if the increase in the atomic B factors upon X-ray exposure is taken into account. For comparative purposes, it is noted that the intensity differences that must be measured for successful S-SAD experiments can be as low as 2%. The differences that are commonly used in MAD experiments are normally greater than 2% but are still often less than 10%. Thus, it should be trivial to accurately measure the intensity change expected to accompany radiation damage in the above example.

Although illustrative, the formula given above tends to overestimate the available signal, since radiation damage affects the whole structure and not just a limited group of atoms. Furthermore, the atoms that seem to disappear in the electron density will not just evaporate, but will be redistributed over a number of new positions. Earlier work has shown that one-tenth to one-third of the observed intensity differences are caused by small rotations and translations of the molecules in the expanding unit cell [11]. The remaining differences are accounted for by changes in crystalline behavior of the sample and by specific damage. In practice, the distinction between these effects is not clear-cut. Consequently, although radiation damage can lead to measurable intensity differences, the interpretation of their origins is problematic.

However, powerful programs are available for finding subsets of atoms for SAD, MAD, and single- and multiple-isomorphous replacement (SIR/MIR) data. In order to locate specific sites of radiation damage, we treated our data as a case of SIR and used the program Shelxd [24] to find the sites of largest differences between the before and after data sets. For the bovine trypsin crystal structure, 14 solutions of six sites were found out of 500 trials, with a clear distinction between the correct and incorrect solutions. It was encouraging that, for the solu-

tions, two of the six sites were 2.0 Å apart, indicating a broken disulphide bond. For the DNA/RNA hybrid, 76 solutions of eight sites were found out of 100 trials. The sites were input to the phasing program Sharp [25], with the after data set used as a reference and the before data set used as a pseudoderivative with low-occupancy sulphur and phosphorous sites for the protein and DNA/RNA hybrid, respectively. For conventional SIR and MIR, additional electrons (atoms) are found in the derivatives. Thus, the RIP analysis treats the radiation damage-induced changes in a manner analogous to conventional SIR. Several cycles of refinement and identification of additional sites from the residual maps were performed, resulting in 28 sites in the asymmetric unit for the protein, including 4 new atomic positions that appeared after the X-ray burn, and 24 sites in the asymmetric unit for the oligonucleotide, including 8 new positions that were induced by the X-ray burn.

Additional phase information is available as a result of radiation damage, since the low-occupancy sites correspond to a partial structure that was likely to exist at 100% occupancy for the before data set. This additional partial structure information could be combined with the Sharp phases, after which solvent flattening was performed. Excellent experimental maps were obtained for the oligonucleotide, as shown in Figure 1. Unfortunately, the resulting protein phases were not sufficiently good to allow for straightforward manual interpretation of the electron density. However, automatic building of the entire protein model was possible with Arp/wArp after further density modification by the multimodel-averaging procedure of this program [26].

Radiation Damage-Induced Structural Changes

The sites used for phasing can be compared with the radiation damage structural changes, as calculated by means of difference Fourier between the two data sets with phases from the refined models.

The six sites found by Shelxd for the protein included the five strongest peaks (ranging between 25 and 39 σ) from the difference Fourier. All 12 sulphur atoms involved in disulphide bonds were found in the phasing process. One of the disulphide bonds is shown in Figure 2. In agreement with the observations of Schröder Leiros et al. [22], disulphide bonds Cys128-232, Cys168-182, and Cys191-220 were more susceptible than the other three. The Cys191-220 bond was most susceptible, where S_γ from Cys220 refined to the largest occupancy (0.34) during phasing. Three new sulphur positions were found during the iterations with Sharp, indicating new, radiation damage-induced conformations for the S_γ atoms of Cys42, Cys232, and Cys220. The new position of S_γ from Cys232 did not refine to the correct position compared with the difference Fourier map, probably due to correlations during refinement, considering the short distance between the new and old S_γ positions. The difference map also showed new conformations for the S_γ atoms from Cys22 (Figure 2B), Cys136, and Cys182, which were not found during phasing.

Apart from the sulphur positions involved in disulphide bonds, many other sites proved to be susceptible to X-ray damage. A lot of changes were observed around

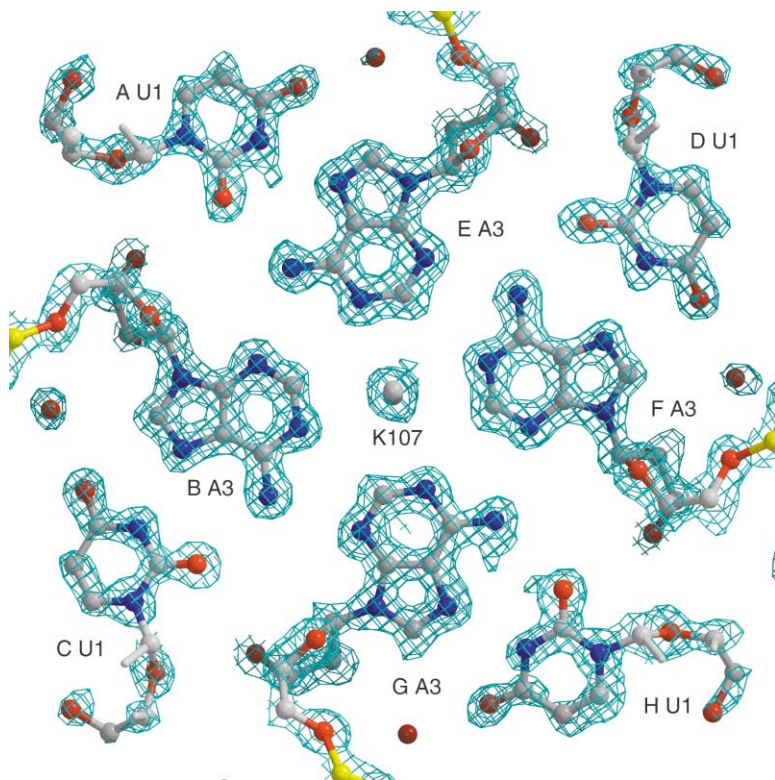


Figure 1. Section of the Experimental Electron Density Map for the DNA/RNA Hybrid
The map is shown after RIP phasing and solvent flattening. The correlation between this map and the final map after refinement is 0.901. The final model, refined against the before data set, has been superimposed. Eight nucleotides were found in the asymmetric unit and are labeled A–H in the figure.

the benzylamine inhibitor that was cocrystallised with the trypsin. As shown in Figure 2D, the inhibitor is highly susceptible to X-rays, as are oxygen atoms coordinating

the amine group of the inhibitor. Many of these sites were identified and used in phasing (Figure 2C). All sites were treated as low-occupancy sulphur sites in Sharp,

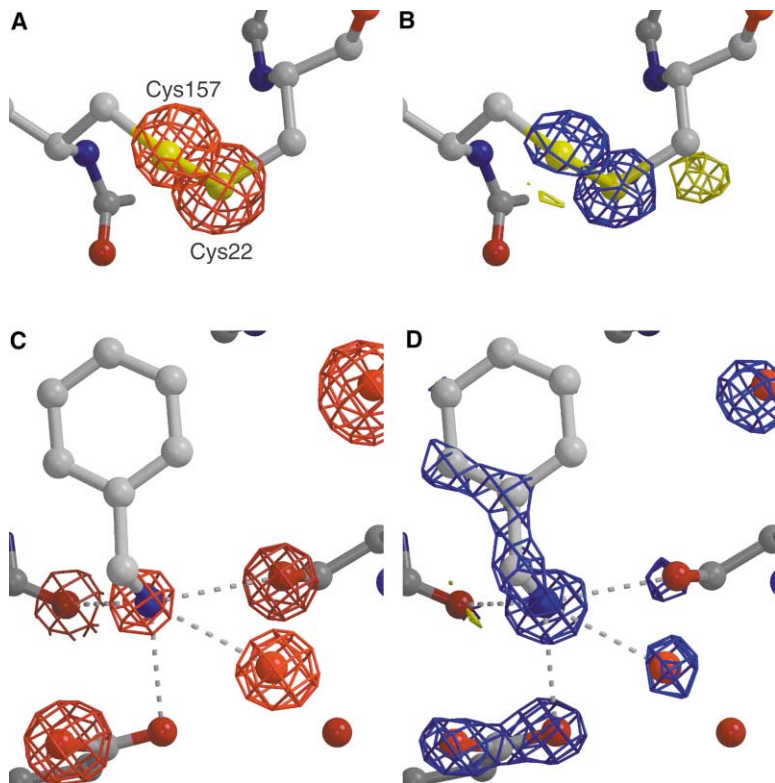


Figure 2. Refined Model of the Cys22-Cys157 Disulphide Bond (A and B) and the Inhibitor Benzylamine and Its Binding Pocket (C and D) in the Trypsin Structure

Hydrogen bonds to the amine group of the inhibitor are drawn as dashed lines. (A) and (C) show the corresponding part of the heavy-atom model that was found and used during the RIP iteration. Red indicates the heavy-atom model of the sites that disappeared from the before data set, contoured at 7σ . (B) and (D) show the difference Fourier map with phases that were obtained after structure determination and refinement. The X-ray susceptible part (blue) is contoured at 5σ ; new positions that were induced by the X-ray burn (yellow) are shown (contoured at -5σ). A comparison of Figures 2A with 2B and 2C with 2D shows that many, though not all, of the X-ray-induced differences in these parts of the structure were found during the phasing iteration.

though, in reality, some corresponded to oxygen or nitrogen atoms. It is encouraging to note that the occupancies of the latter, in general, refined to lower values than the true sulphur sites.

The difference Fourier map for trypsin contained a total of 115 peaks with an absolute height larger than 5σ , 23 of which were larger than 10σ . The positive peaks corresponded to sites that were particularly susceptible to X-ray damage, whereas the negative peaks indicated positions where the relative electron density increased upon the X-ray burn. Forty-three negative peaks were found, the strongest being -14.2σ . All peaks, other than those close to the inhibitor, correspond to specific changes that have been identified earlier in other proteins [11, 20, 21]. They include disulphide bonds, carboxyl groups, methionine, carbonyl oxygen atoms, and waters.

When we attempted to solve the DNA/RNA structure, no structural studies had been reported on X-ray-induced specific damage in oligonucleotides. So that this could be used as a real test case for the feasibility of phasing macromolecules with radiation damage alone, no information about the content of the unit cell was passed on to the experimenters, other than the information that the crystals contained RNA. Shelxd found clear solutions for eight sites, which were assumed to correspond to phosphorus atoms. Even though the occupancy of these sites refined to large values (around 0.52), this hypothesis was maintained throughout the phasing procedure. From the residual maps, it was clear that the susceptible atoms moved 1.9 \AA away from their original positions and that more atoms in the direct surroundings of the eight major sites were susceptible to the X-radiation. Inspection of the RIP electron density map revealed that the asymmetric unit contained eight copies of the DNA/RNA hybrid, and, for each copy, a Br atom was cleaved from the C8 of a brominated guanine (Figure 3). Recently, debromination was reported for C5-bromouridine during the course of a MAD data collection at a third generation bending magnet beamline [27]. In this study, a total dose similar to that in our experiment was used, amounting to $1.3 \times 10^7\text{ Gy}$ for the final data set, after which the bromine sites were no longer present in the σ_a -weighted ($F_o - F_c$) omit difference maps. Although it was suggested that the debromination might have provided additional phase information to the MAD phases that were used, no actual attempt was made to do so.

In parallel with the hypothesis that the breakage of disulphide bonds upon X-ray exposure is due to electron capture [11], we hypothesize that 8-Br-dGua loses its bromine ion while serving as an electron sink [28]. The difference Fourier map that was obtained after refinement showed the highest peaks ($42\text{--}49\sigma$) for the bromine ions. In total, there were 214 peaks with an absolute height larger than 5σ , 29 of which had an absolute height larger than 10σ . One hundred twenty-eight negative peaks were found, including the eight new positions of the cleaved Br atoms. The new Br positions appeared in the difference Fourier map at an absolute peak height larger than 10σ . They had all been identified and used during the phasing procedure.

Discussion

Although our brightest synchrotron sources have had an extraordinary impact in structural biology during the past few years, they are, in fact, also creating havoc with crystalline biological samples. Processes referred to as radiation damage militates against collecting highly redundant data to the highest possible resolution. The problem of radiation damage is even more of a threat to future generation synchrotron sources [29]. However, we have shown in the present paper that radiation damage, instead of hindering successful structure determination, provides an important opportunity for phasing macromolecular structures by using the RIP method. Although the RIP protocol has similarities to SIR, no chemical treatment of the sample is needed. The RIP method suffers from a large number of small, nonisomorphous differences that cannot be accurately modeled with the current software while treating RIP as SIR. Nonetheless, phases could be obtained that led to the correct structures.

Very high correlation coefficients (0.501 and 0.901, respectively) were obtained between the final DNA/RNA electron density map and the maps after initial phasing and solvent flattening, despite the use of a rather incomplete and partially incorrect model of the susceptible sites. The experimental phases obtained were of outstanding quality, as shown in Figure 1. For the protein, the correlation coefficients between the final map and the maps after initial phasing, phase combination, solvent flattening, multimodel averaging, and autobuilding were 0.342, 0.388, 0.509, 0.706, and 0.955, respectively. This indicates the importance of sophisticated density modification techniques, such as multimodel averaging in wArap for this example, where the solvent content was rather low. The quality of the initial phases was relatively independent of resolution, whereas, in SAD/MAD, the best phases are normally obtained at lower resolution. The large numbers of sites that can be found in RIP forms an important additional source of phase information and has been used by combining the initial phases with a partial model consisting of these sites at full occupancy.

The RIP method works for the DNA/RNA hybrid at any resolution down to 4 \AA , in that the starting eight sites were always found. The iteration with Sharp has been repeated with data to 2.5 \AA , and it yielded good phases. The question of resolution limit was not further investigated for the protein, for which one clear limitation could be formed by the program wArap, which generally does not work at resolutions lower than $2.0\text{--}2.3\text{ \AA}$ [26]. Both cases that we tested concern well-diffracting crystals. It is evident that further studies are required in order to make the RIP method more generally applicable for larger macromolecules with a smaller number of highly susceptible sites and for poorly diffracting crystals.

However, the wealth of additional data that can be collected during RIP experiments makes us optimistic about future applications of this method. In the present study, a dose of about $1 \times 10^7\text{ Gy}$ was used to irradiate the protein and oligonucleotide crystals between the two data sets. The data sets themselves were collected with a dose an order of magnitude less than was used

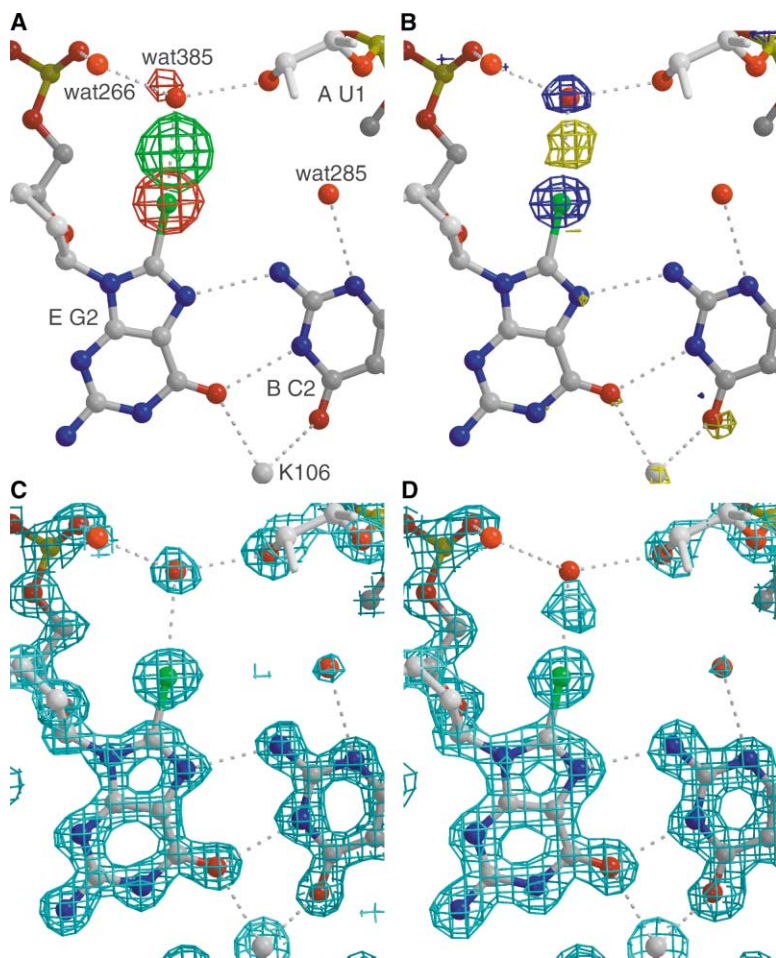


Figure 3. Refined Model of 8-Br-Gua in Chain E of the DNA/RNA Hybrid Structure

Hydrogen bonds are drawn as dashed lines. (A) shows the heavy-atom model as used for phasing, in red (7 σ) for sites that disappeared upon radiation damage and in green (-7 σ) for the new site that appeared upon X-ray damage. (B) shows the difference Fourier map with phases that were obtained after structure determination and refinement. The X-ray susceptible part (blue) is contoured at 5 σ ; the sites that showed a relative increase in electron density upon radiation damage (yellow) are contoured at -5 σ . (C) and (D) show the electron density as calculated after refinement for the before (C) and after (D) data sets. The main feature in these maps is the debromination of the guanine base. Part of the detached Br atoms moves 3.8 Å away from the C8 of guanine. Concurrently, a water molecule that was hydrogen bonded to the Br atom disappears.

in the damage phase of the experiment. If we assume that crystalline diffraction is lost beyond an accumulated dose of 2×10^7 Gy [30], then at least ten complete data sets of the type shown in Tables 1 and 2 could have been collected on these crystals. This would allow appropriate phasing programs to refine far more sophisticated models for the susceptible sites, thus producing better phases. These models would become dynamic rather than static, in that both position and occupancy of the susceptible sites would be allowed to change with accumulated dose. RIP data collection protocols would be facilitated greatly by faster detectors. For example, a total of 2 hr was required for data collection from the trypsin crystal used in the study. However, only 13% of this time was required for actual exposure of the sample. Ongoing hardware and software developments will make such experiments feasible in a not too distant future.

For proteins, combining the RIP and the S-SAD methods is likely to be important. If we want to make full use of the latter method at third generation synchrotron undulator beamlines, specific radiation damage must be taken into account. Specific damage could serve as an extra source of phase information, with the additional benefit that it could resolve the inherent phase ambiguity of SAD.

Biological Implications

Two major bottlenecks can be seen in the use of X-ray crystallography for the determination of three-dimensional structures of macromolecules, namely, crystallizing the macromolecule and solving its structure once suitable crystals have been obtained. The MAD method has become the method of choice for dealing with the latter bottleneck, particularly when MAD is used in combination with selenomethionine-derivatized proteins. However, it is not always possible to introduce selenomethionine into the protein of interest, and crystals of selenated proteins do not always grow or diffract as well as their native counterparts. Furthermore, the quality of the phases obtained by the MAD method can be degraded by radiation damage, a common feature when third generation synchrotron undulator beamlines are used with weakly diffracting crystals.

In this paper, we have shown that radiation damage, long considered the bane of the brightest synchrotron sources, can, in fact, facilitate structure determination. Radiation damage brings about specific changes in structure that enable reliable phasing of structure. The method introduced in this paper is called radiation damage-induced phasing (RIP). That the RIP method works has been demonstrated by its having been used

to determine the structure of a native protein and of a brominated oligonucleotide for which data were collected below the bromine absorption edge. The general applicability of the method is discussed. Further developments of the RIP method will benefit the more traditional phasing techniques like S-, Se-, or Br-SAD/MAD. We feel optimistic that the use of RIP, either as a primary phasing technique or as a secondary source of phase information, will render derivatization a considerably less significant barrier to macromolecular structure determination.

Experimental Procedures

Bovine trypsin (EC 3.4.21.4) was obtained from Sigma and used without further purification. Crystals were grown at room temperature by the sitting drop method in the presence of the inhibitor benzylamine. The reservoir solution consisted of 25% polyethylene glycol (PEG) 8000, 0.2 M ammonium sulphate, 0.1 M Tris buffer (pH 8), and 100 mM benzylamine [31]. The drops consisted of 5 μ l protein solution at 15 mg ml⁻¹ mixed with 5 μ l reservoir solution. One crystal measuring 150 \times 20 \times 40 μ m³ was used for data collection. The crystal was frozen at 100 K in an Oxford S6000 cryostream with paratone oil as cryoprotectant.

The DNA/RNA hybrid was synthesized by the phosphoramidite method with an Applied Biosystem DNA synthesizer 391. It was cleaved from the solid support with ammonia/ethanol (3:1 by volume) and incubated at room temperature for 24 hr. The 2'-hydroxyl groups were deprotected by triethylamine Tris (hydrofluoride) for 3 hr at 55°C and purified by ion exchange chromatography (FPLC). Crystallization was carried out by the hanging drop vapor diffusion method at room temperature (20°C). The crystallization conditions were 1 mM (single-strand) RNA, 20 mM sodium cacodylate buffer (pH 6.0), 5 mM spermine tetrachloride, 32 mM KCl, and 2% (v/v) methyl-2,4-pentanediol (MPD) against 25% MPD in the reservoir. The crystal for data collection grew in several days. The crystal was frozen at 100 K in an Oxford S6000 cryostream with 30% MPD buffered by 40 mM sodium cacodylate (pH 6.0) as cryoprotectant.

Data were collected at the undulator beamlines ID14-4 (trypsin) and ID14-2 (DNA/RNA hexamer) at the European Synchrotron Radiation Facility (ESRF). For the trypsin data collection, a beamspace of 100 \times 100 μ m² was used, and the wavelength was set at 0.976 Å, where a high photon flux at the sample position could be obtained. Two data sets were collected, consisting of 360 frames of 1° each. An attenuator of 0.84 mm Al was used, giving rise to a measured attenuation factor of 24. The exposure time was 1 s per frame, and the crystal diffracted, despite the strong attenuation, to about 1.4 Å. In between the data sets, the crystal was exposed to the unattenuated beam for 2 min, during which the crystal was rotated over 360°. For the DNA/RNA hybrid data collection, a beamspace of 20 \times 50 μ m² (horizontal by vertical) was used. The data were collected at a fixed energy beamline, with a wavelength of 0.933 Å. This is below the absorption edge of Br, and no anomalous signal was used. Two data sets were collected, consisting of 110 frames of 1° each, with an exposure time of 2 s per frame. No attenuators were used, and the crystal was exposed for 20 min in between the data sets, using a 100 \times 100 μ m² beam. The crystal was rotated over 360° during the X-ray burn. The data sets are referred to as before and after the burn.

The absorbed doses were calculated with the program RadDose (http://www.esrf.fr/exp_facilities/ID14-4/ID14-4.html) on the basis of measured crystal size, photon flux and energy, and calculated crystal absorption and density. The photon flux was determined after calibration of the beamline diodes with a calibrated reference diode. A flux of about 1.9 \times 10¹² photons/s was used for the burn in the trypsin experiment, whereas a flux of about 1.1 \times 10¹¹ photons/s was used for the burn in the oligonucleotide experiment. The overall errors in the given absorbed doses are estimated to be somewhat less than one order of magnitude given the uncertainties in (1) absolute photon flux measurements, (2) crystal density calculation, (3)

absorption coefficient calculations, and (4) assumptions regarding the beam profile and beam path through the crystal.

Data were indexed and integrated with Denzo and XDS and scaled with Scala from the CCP4 suite [32]. Data statistics are given in Tables 1 and 2. The data sets before and after the burn were scaled toward each other with Scala and Scaleit. Additional local scaling between the two data sets was performed with Xprep on the basis of unmerged original index data from Scala. The program Shelxd [24] was used to find the X-ray-susceptible sites. For the protein, 8268 normalized structure factors with an amplitude larger than 1.3 were used. Solutions were characterized by good Patterson figures of merit (10) but had very low scores for the correlation coefficient (8%). Better statistics were obtained for the DNA/RNA hybrid, where 3730 normalized structure factors with an amplitude larger than 1.4 were used. In this case, the scores of the Patterson figure of merit (13) and the correlation coefficient (35%) clearly indicated a good solution. The sites found by Shelxd were refined in Sharp [25] with all data, treating the two data sets as individual crystals. The before data set was modeled as the structure corresponding to the after data set plus low-occupancy sulphur (protein) or phosphorus (DNA/RNA) sites corresponding to the X-ray-susceptible sites. Peaks with absolute peak heights higher than 4.5 σ were selected from the residual maps and included as low-occupancy sites in subsequent rounds of refinement.

For the DNA/RNA hybrid, high-quality electron density maps (Figure 1) were obtained directly after solvent flattening with Solomon, as directed by Sharp. Eight hexamers were identified and placed in the asymmetric unit. For the protein, the phases as obtained with Sharp were combined with the program Sigmaa [33] with phases of a partial model that consisted of the X-ray-susceptible sites at full occupancy. Solvent flattening was subsequently performed with Dm [34] and a solvent content of 43%. The phases were further improved with Arp/wArp [26] with averaging of six free-atom models. These improved phases were used to autobuild the structure with the warpNtrace option of Arp/wArp, resulting in three chains, 213 residues, and a connectivity index of 0.97.

Where necessary, the models were adjusted upon visual inspection in O [35]. Solvent molecules were added to the model when the difference density exceeded 4 σ and were further refined in subsequent cycles with Refmac [36]. For the protein, a ligand was built, and double conformations were found for the residues Asn79, Ser122, Ser236, and Ile242. For the oligonucleotide, two spermine molecules were found, as well as nine K ions. A partial breakage of the Br-Gua bond was already observed in the before data set, and the C8-bonded Br atom was modeled at 0.7 occupancy for this data set. For both the protein and the oligonucleotide, atoms were refined with anisotropic B factors, with the exception of the solvent molecules that were kept isotropic. Five percent of the data was reserved for crossvalidation with the free R factor.

The final models were refined against both data sets before and after for 20 cycles of restrained maximum-likelihood refinement. The refined model was not additionally adjusted manually for the after data set, resulting in features in the σ_A -weighted $F_o - F_c$ map that show the X-ray-susceptible sites, such as, for the protein, negative density for disulphide bonds and positive density for new positions for some of the sulphurs. For the hybrid hexamer, both the $2F_o - F_c$ and the $F_o - F_c$ maps showed breakage of the Br-Gua bond for the after data set. The Fourier difference map was calculated between the before and after data sets as described [11]. The peaks from this difference map with an absolute peak height higher than 5 σ were sorted with Watpeak [32].

Acknowledgments

In memory of Professor Jan Kroon, who died on May 3, 2001. We thank all staff of the ESRF and EMBL Grenoble Outstation for maintenance of the JSBG beamlines. We gratefully acknowledge Mark Brooks, Gordon Leonard, and Edward Hough for critical reading of the manuscript and Professor Mutaiya Sundaralingam for his support of the DNA/RNA experiment. Research was supported, in part, by grants from the National Institutes of Health (GM 56969 and GM 61010) and the National Science Foundation (DIR 9016683 and DBI 9981990) to M.C.

Received: August 13, 2002
Revised: November 20, 2002
Accepted: November 25, 2002

References

1. Heinemann, U., Illing, G., and Oschkinat, H. (2001). High-throughput three-dimensional protein structure determination. *Curr. Opin. Biotechnol.* **12**, 348–354.
2. Sansom, C. (2000). Structural genomics. *Mol. Med. Today* **6**, 258.
3. Dean, P.M., Zanders, E.D., and Bailey, D.S. (2001). Industrial-scale, genomics-based drug design and discovery. *Trends Biotechnol.* **19**, 288–292.
4. Thornton, J. (2001). Structural genomics takes off. *Trends Biochem. Sci.* **26**, 88–89.
5. Mittl, P.R.E., and Grutter, M.G. (2001). Structural genomics: opportunities and challenges. *Curr. Opin. Chem. Biol.* **5**, 402–408.
6. Hendrickson, W.A. (2000). Synchrotron crystallography. *Trends Biochem. Sci.* **25**, 637–643.
7. Muchmore, S.W., Olson, J., Jones, R., Pan, J., Blum, M., Greer, J., Merrick, S.M., Magdalinos, P., and Nienaber, V.L. (2000). Automated crystal mounting and data collection for protein crystallography. *Structure* **8**, R243–R246.
8. Service, R.F. (2001). Structural biology: robots enter the race to analyze proteins. *Science* **292**, 187–188.
9. Walsh, M.A., Dementieva, I., Evans, G., Sanishvili, R., and Joachimiak, A. (1999). Taking MAD to the extreme: ultrafast protein structure determination. *Acta Crystallogr. D Biol. Crystallogr.* **55**, 1168–1173.
10. McSweeney, S., and Leonard, G.A. (1999). Multi-wavelength anomalous dispersion (MAD) studies. In *ESRF Highlights 1999*, D. Cornuejols, ed. (Grenoble, France: ESRF) pp. 14–15.
11. Ravelli, R.B.G., and McSweeney, S.M. (2000). The ‘fingerprint’ that X-rays can leave on structures. *Structure* **8**, 315–328.
12. Rice, L.M., Earnest, T.N., and Brunger, A.T. (2000). Single-wavelength anomalous diffraction phasing revisited. *Acta Crystallogr. D Biol. Crystallogr.* **56**, 1413–1420.
13. Dauter, Z., Li, M., and Wlodawer, A. (2001). Practical experience with the use of halides for phasing macromolecular structures: a powerful tool for structural genomics. *Acta Crystallogr. D Biol. Crystallogr.* **57**, 239–249.
14. Cohen, A., Ellis, P., Kresge, N., and Soltis, S.M. (2001). MAD phasing with krypton. *Acta Crystallogr. D Biol. Crystallogr.* **57**, 233–238.
15. McAuley, K.E., Jia-Xing, Y., Dodson, E.J., Lehmebeck, J., Ostergaard, P.R., and Wilson, K.S. (2001). A quick solution: ab initio structure determination of a 19 kDa metalloproteinase using ACORN. *Acta Crystallogr. D Biol. Crystallogr.* **57**, 1571–1578.
16. Wang, B.C. (1985). Resolution of phase ambiguity in macromolecular crystallography. *Methods Enzymol.* **115**, 90–112.
17. Dauter, Z., Dauter, M., de La Fortelle, E., Bricogne, G., and Sheldrick, G.M. (1999). Can anomalous signal of sulfur become a tool for solving protein crystal structures? *J. Mol. Biol.* **289**, 83–92.
18. Gordon, E.J., Leonard, G.A., McSweeney, S., and Zagalsky, P.F. (2001). The C1 subunit of alpha-crustacyanin: the de novo phasing of the crystal structure of a 40 kDa homodimeric protein using the anomalous scattering from S atoms combined with direct methods. *Acta Crystallogr. D Biol. Crystallogr.* **57**, 1230–1237.
19. Micossi, E., Hunter, W.N., and Leonard, G.A. (2002). De novo phasing of two crystal forms of tryparedoxin II using the anomalous scattering from S atoms: a combination of small signal and medium resolution reveals this to be a general tool for solving protein crystal structures. *Acta Crystallogr. D Biol. Crystallogr.* **58**, 21–28.
20. Weik, M., Ravelli, R.B., Kryger, G., McSweeney, S., Raves, M.L., Harel, M., Gros, P., Silman, I., Kroon, J., and Sussman, J.L. (2000). Specific chemical and structural damage to proteins produced by synchrotron radiation. *Proc. Natl. Acad. Sci. USA* **97**, 623–628.
21. Burmeister, W.P. (2000). Structural changes in a cryo-cooled protein crystal owing to radiation damage. *Acta Crystallogr. D Biol. Crystallogr.* **56**, 328–341.
22. Schröder Leiros, H.K., McSweeney, S.M., and Smalás, A.O. (2001). Atomic resolution structures of trypsin provide insight into structural radiation damage. *Acta Crystallogr. D Biol. Crystallogr.* **57**, 488–497.
23. Crick, F.H., and Magdoff, B.S. (1956). The theory of the method of isomorphous replacement for protein crystals. I. *Acta Crystallogr. A* **9**, 901–908.
24. Uson, I., and Sheldrick, G.M. (1999). Advances in direct methods for protein crystallography. *Curr. Opin. Struct. Biol.* **9**, 643–648.
25. de la Fortelle, E., and Bricogne, G. (1997). SHARP: a maximum-likelihood heavy-atom parameter refinement program for the MIR and MAD Methods. In *Methods in Enzymology*, Volume 276, R.M. Sweet and C.W. Carter, Jr., eds. (New York: Academic Press) 472–494.
26. Perrakis, A., Sixma, T.K., Wilson, K.S., and Lamzin, V.S. (1997). wARP: improvement and extension of crystallographic phases by weighted averaging of multiple refined dummy atomic models. *Acta Crystallogr. D Biol. Crystallogr.* **53**, 448–455.
27. Ennifar, E., Carpentier, P., Ferrer, J.L., Walter, P., and Dumas, P. (2002). X-ray-induced debromination of nucleic acids at the Br K absorption edge and implications for MAD phasing. *Acta Crystallogr. D Biol. Crystallogr.* **58**, 1262–1268.
28. Nese, C., Yuan, Z., Schuchmann, M.N., and Von Sonntag, C. (1992). Electron transfer from nucleobase electron adducts to 5-bromouracil. Is guanine an ultimate sink for the electron in irradiated DNA? *Int. J. Radiat. Biol.* **62**, 527–541.
29. Henderson, R. (2002). Excitement over X-ray lasers is excessive. *Nature* **415**, 833.
30. Henderson, R. (1990). Cryo-protection of protein crystals against radiation damage in electron and X-ray diffraction. *Proc. R. Soc. Lond. B Biol. Sci.* **248**, 6–8.
31. Kurinov, I.V., and Harrison, R.W. (1994). Prediction of new serine proteinase inhibitors. *Nat. Struct. Biol.* **1**, 735–743.
32. Collaborative Computational Project 4 (1994). The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D Biol. Crystallogr.* **50**, 760–763.
33. Read, R.J. (1986). Improved Fourier coefficients for maps using phases from partial structures with errors. *Acta Crystallogr. A* **42**, 140–149.
34. Cowtan, K., and Main, P. (1998). Miscellaneous algorithms for density modification. *Acta Crystallogr. D Biol. Crystallogr.* **54**, 487–493.
35. Jones, T.A., Zou, J.Y., Cowan, S.W., and Kjeldgaard, M. (1991). Improved methods for binding protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A* **47**, 110–119.
36. Murshudov, G.N., Vagin, A.A., Lebedev, A., Wilson, K.S., and Dodson, E.J. (1999). Efficient anisotropic refinement of macromolecular structures using FFT. *Acta Crystallogr. D Biol. Crystallogr.* **55**, 247–255.

Accession Numbers

Coordinates have been deposited with the Protein Data Bank under ID codes 1N6X, 1N6Y, 1N7A, and 1N7B.