

## Comparison of Different X-ray Data-Collection Systems using the Crystal Structure of Octreotide

BY EHMKE POHL, ANDREAS HEINE AND GEORGE M. SHELDRICK\*

*Institut für Anorganische Chemie der Universität Göttingen, Tammannstrasse 4, D-37077 Göttingen, Germany*

ZBIGNIEW DAUTER, THOMAS R. SCHNEIDER AND KEITH S. WILSON

*European Molecular Biology Laboratory (EMBL), c/o DESY, Notkestrasse 85, D-22603 Hamburg, Germany*

AND JÖRG KALLEN

*Preclinical Research Department, Sandoz Ltd, CH-4002 Basel, Switzerland*

(Received 1 July 1993; accepted 27 May 1994)

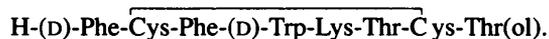
### Abstract

The octapeptide octreotide crystallizes with three peptide molecules and about 20% water in the asymmetric unit, and in many ways possesses diffraction properties similar to those of a 'mini-protein' consisting of 24 amino-acid residues. It diffracts to about 1.0 Å but data in the range 1.4–1.0 Å are weak. It provides a suitable test of different macromolecular X-ray data-collection techniques, especially of their ability to measure weak reflections accurately. In contrast to typical proteins it is possible to perform a full anisotropic refinement, that we believe provides a more objective test of the quality of the data than the internal consistency of equivalent reflections. We have collected a total of six data sets. The X-ray sources included synchrotron radiation, Cu  $K\alpha$  rotating anodes and Mo  $K\alpha$  sealed tubes; position-sensitive two-dimensional detectors from four manufacturers and a four-circle diffractometer with scintillation counter were employed. Two of the six data sets were collected at low temperature. Reasonable anisotropic refinement was possible with all area-detector data sets, although significant differences in the precision of the final model were observed. In addition we tested the ability of automated Patterson interpretation to solve the structure using the six independent data sets. The structure solution was only successful using the synchrotron or rotating-anode data sets, *i.e.* for the more intense sources. It appears that for structure solution the maximum resolution of the data is critical, whereas for refinement the accuracy of the data is more important.

### Introduction

In the course of our determination of the structure of octreotide (Pohl *et al.*, 1994) we had the opportunity to record X-ray data on six different systems. Octreotide

is an octapeptide with one disulfide bridge and the sequence,



We felt that octreotide would be a good test of macromolecular data-collection techniques, because it behaves in many ways like a 'mini-protein' with 24 residues that diffracts, albeit weakly, to about 1.0 Å. It is small enough for 'small-molecule' methods to be used for structure solution and refinement, thus providing a more objective test than simply the internal consistency of the data, but has a large enough unit-cell and diffracts sufficiently weakly to provide a realistic test of area detectors.

Data were collected using the following systems (the same numbering scheme has been employed in the tables and figures).

(1) Synchrotron radiation from the DORIS storage ring at DESY, Hamburg, with an EMBL imaging-plate scanner as detector (220 mm diameter).

(2) Cu  $K\alpha$  rotating anode with a X200B multiwire proportional chamber from Siemens (Blum, Metcalf, Harrison & Wiley, 1987).

(3) Cu  $K\alpha$  rotating anode with a FAST system from Enraf-Nonius.

(4) Mo  $K\alpha$  sealed tube with a Stoe imaging-plate scanner of 180 mm diameter (Stoe imaging-plate diffraction system, IPDS).

(5) Mo  $K\alpha$  sealed tube with an MAR Research imaging-plate scanner (180 mm diameter) with the crystal at 153 K.

(6) Mo  $K\alpha$  sealed tube, Stoe-Siemens four-circle diffractometer and a scintillation counter with the crystal at 193 K.

To compare the data sets we first calculated  $R_{\text{int}}$  values for the agreement of intensities of equivalent reflections (including Friedel opposites) within each data set. However these  $R_{\text{int}}$  values are highly dependent on the redundancy and on the scaling algorithms in the

\* Author to whom correspondence should be addressed.

Table 1. *Data-collection and data-processing parameters*

	EMBL IP (1)	Siemens (2)	FAST (3)	Stoe IP (4)	MAR IP (5)	Four-circle (6)
X-ray source	Synchrotron	Rotating anode	Rotating anode	Sealed tube	Sealed tube	Sealed tube
Wavelength (Å)	0.70	1.54178	1.54178	0.71073	0.71073	0.71073
Operating conditions (kV, mA)	n/a	50, 90	40, 70	50, 40	50, 50	55, 30
Focal spot (mm)	n/a	0.3	0.3	0.5	0.4	0.8
Temperature (K)	285	293	293	296	153	193
Crystal-detector (mm)	110, 220, 350	70	43	100, 190	115, 240	220
Total observations	52006	52183	81087	64338	68507	10521
Unique observations	18951	17225	12330	22925	16327	9388
Max. resolution (Å)	1.04	1.05	10.5	1.0	1.11	1.1
Mean $I/\sigma(I)$	12.8	18.5	13.2	14.9	7.6	6.2
Collected 1.1–1.2 Å (%)*	94.3	85.4	50.3	100	94.3	100
$I > 2\sigma(I)$ 1.1–1.2 Å (%)*	84.0	58.0	42.1	54.7	67.1	20.6
$R_{int}^\dagger$	0.024	0.037	0.024	0.073	0.078	0.064

\* Percentage calculated relative to total number of possible reflections in this range, Friedel opposites merged.

† Calculated with *SHELXS92*, Friedel opposites merged,  $R_{int} = \sum |F_o^2 - F_o^2(\text{mean})| / \sum [F_o^2]$ .

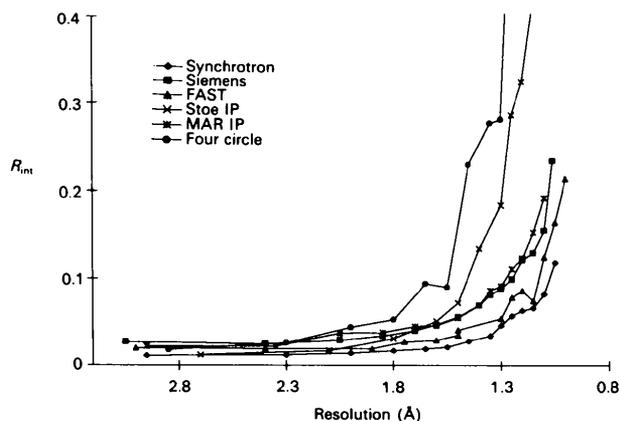
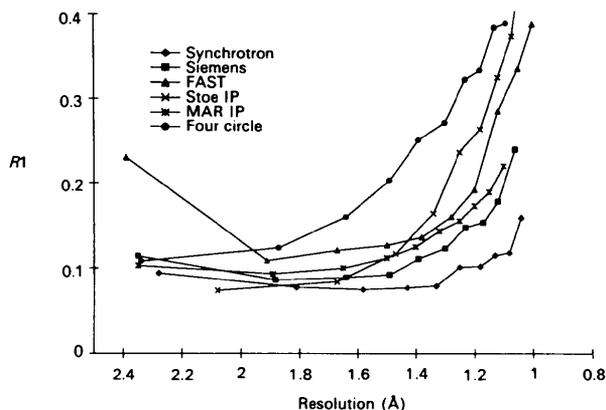
data-reduction software; systematic errors, for example  $2\theta$ -dependent bias, may well fail to be detected in this approach. On the other hand  $R_{merge}$  values, which measure the consistency of the intensities of equivalent reflections in different data sets after bringing them onto the same scale, are difficult to compare because there is no absolute standard. In view of the different data-collection temperatures and wavelengths we did not attempt to scale and merge the different data sets together. In previous tests (Tucker, 1990; Krause & Phillips, 1992) a four-circle diffractometer data set served as a reference. This was not possible here as the four-circle data in the critical resolution range 1.0–1.4 Å were clearly too weak. If the synchrotron data had been taken as reference, systematic errors of the other imaging-plate systems might well have been underestimated.

Therefore, we employed two other tests, which we believe to be more objective. Firstly the ability to solve the structure by automated Patterson interpretation using *SHELXS92* (Sheldrick, 1992) was investigated (successfully for three of the data sets). At the time we were not able to solve this structure by direct methods (Sheldrick, 1990) using any of the data sets. Secondly, the structure was refined independently against each data set by least-squares techniques using *SHELXL93* (Sheldrick, 1993), and the resulting structures and residuals were compared. The final residuals and chemical reasonableness of the refined model were used to assess the quality of the data. We also considered the accuracy of the unit-cell determination, for which it is reasonable to use the values obtained with a four-circle diffractometer as a reference.

### Experimental

The crystallization and crystal structure determination using the synchrotron data are described in the preceding paper (Pohl *et al.*, 1995). Data-collection and processing parameters for all six data sets are summarized in Table 1. Further details of each data set are given here.

For each data collection we attempted to use crystals of similar size and shape (2.0 × 0.3 × 0.4 mm). It should be noted that one crystal dimension was longer than the width of the beam but it proved impossible to cut the crystals as they cracked easily; this situation is not unusual in protein crystallography. For the room-temperature experiments a single crystal was sealed in

Fig. 1.  $R_{int}$  values versus resolution.Fig. 2.  $R_1$  values versus resolution.

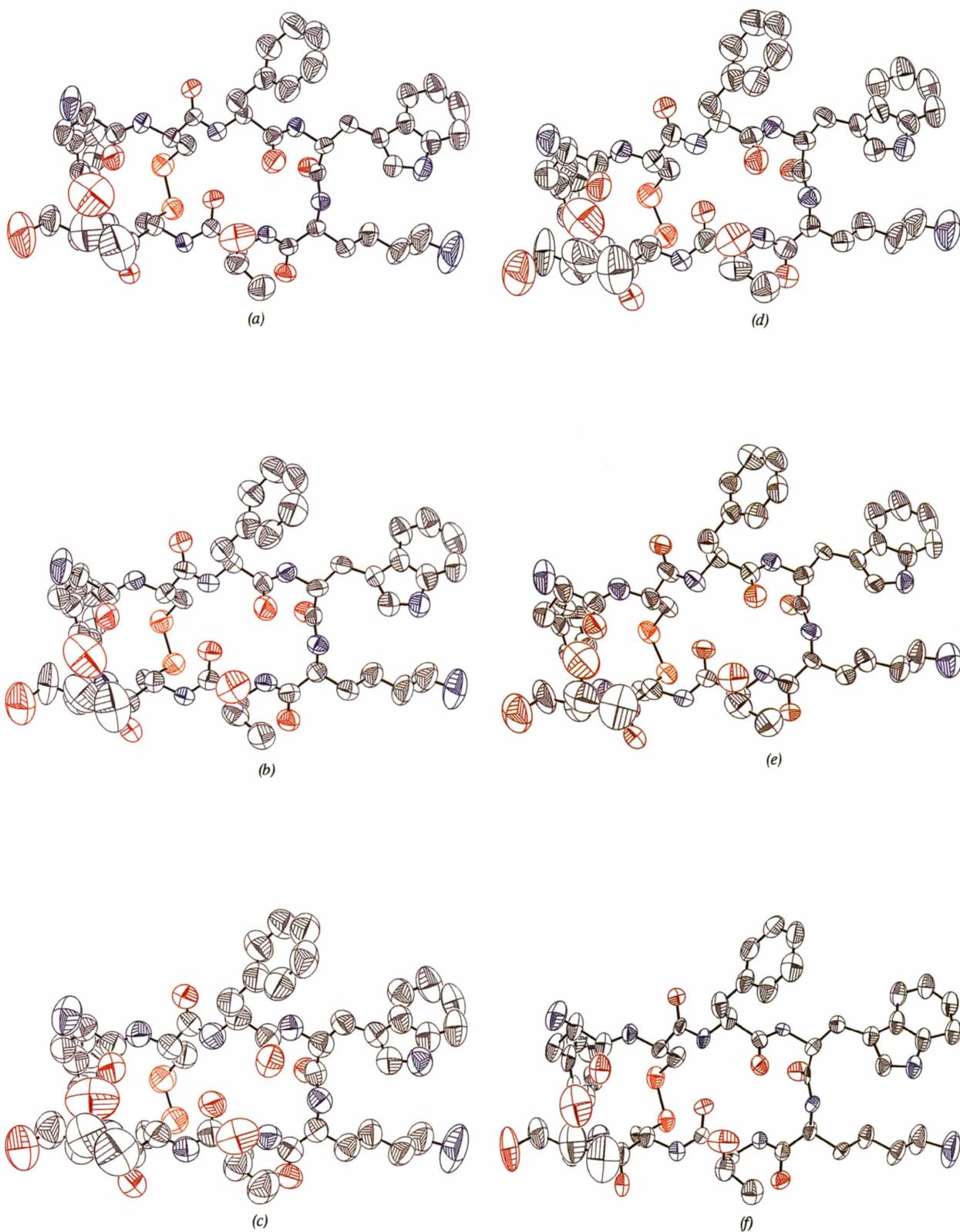


Fig. 3. 50% probability anisotropic displacement parameters plots of one molecule (designated molecule I) for each data set (a) Synchrotron (1), (b) Siemens (2), (c) FAST (3), (d) Stoe image plate (4), (e) MAR image plate (5), (f) Stoe four-circle diffractometer (6).

a Lindemann glass capillary with some mother liquor to prevent rapid crystal decomposition associated with solvent loss. The low-temperature experiments were performed using an oil-coated rapidly cooled crystal on the tip of a glass fibre (Hope, 1988; Kottke & Stalke, 1993).

#### (1) Synchrotron data

Data were collected on the X31 EMBL beamline located at the DORIS storage ring, DESY, Hamburg. The ring was operating at 5.7 GeV with electrons in a multi-bunch mode. The EMBL imaging-plate scanner (220 mm diameter) was used as detector. The data-collection strategy was based on the rotation method developed for photographic film (Arndt & Wonacott, 1977). Data were collected in three passes for the same crystal at high, medium and low resolution using different exposure times. High-resolution data were collected at a crystal-to-detector distance of 110 mm with  $90^\circ$   $\varphi$  range and a  $\varphi$  increment of  $2^\circ$ , medium resolution at 220 mm and  $\varphi$  increment  $4^\circ$ , low resolution at 350 mm and  $\varphi$  increment  $8^\circ$ . Three passes at high, medium and low resolution were needed as the stronger (low-resolution) reflections saturated the electronic scanner read-out and thus some pixels were unusable when the exposure time was long enough to obtain significant intensities for the weak (high-resolution) reflections. The images were subsequently processed using the program *DENZO* (Otwinowski, 1991). Only fully recorded reflections were accepted from each image; symmetry-equivalent reflections and Friedel opposites were not merged during data processing. No absorption or decay correction was applied.

#### (2) Siemens data

A Siemens three-axis diffractometer, Rigaku HB200 rotating anode as source of graphite-monochromated  $\text{Cu } K\alpha$  radiation and an X200B multiwire proportional chamber were employed for data collection. High-resolution data were collected at a  $60^\circ$  detector-swing angle with 200 s exposure per frame; the  $\varphi$  range was  $240^\circ$  with an increment of  $0.6^\circ$  per frame. Low-resolution data were collected at a  $15^\circ$  swing angle with 200 s exposure and  $1^\circ$   $\varphi$  increment. As some reflections were not recorded during the first  $\varphi$ -rotation at high detector-swing angle because the area of the detector did not cover the volume of reciprocal space swept out by the diffracting sphere, a second  $\varphi$ -rotation at a different crystal orientation was performed. Data processing was performed using the *XENGEN* program package (Howard *et al.*, 1987). All symmetry-related reflections, but not the Friedel opposites, were merged during data processing.

#### (3) Enraf–Nonius FAST data

An Enraf–Nonius FAST area detector with a VARIAN intensifier on an Enraf–Nonius FR571 rotating anode

( $\text{Cu } K\alpha$  radiation) was employed for data collection. The special asymmetric  $\theta$  arrangement allowed collection of data up to a resolution of  $1.05 \text{ \AA}$  (at  $\theta = -68^\circ$ , crystal-to-detector distance 43 mm). High- and low-resolution data were collected with different detector-swing angles and exposure times, ranging from 12 s per  $0.2^\circ$  frame at  $\theta = -30^\circ$  to 90 s per  $0.2^\circ$  frame at  $\theta = -68^\circ$ . A total of 13 batches were collected from two crystals. Data processing and merging were performed with the *MADNES* program package (Messerschmidt & Pflugrath, 1987) followed by *PROCOR* (Kabsch, 1988). Thus, absorption and decay corrections were indirectly applied by scaling together data from the frames. The cell dimensions were obtained using 250 strong reflections. Data merging was carried out using the *CCP4* program package (SERC Daresbury Laboratory, 1979). Again all symmetry-equivalent reflections, but not the Friedel opposites, were merged, which resulted in a set of  $F$  rather than  $F^2$  values (thus, negative or zero intensities were converted to positive  $F$  values).

#### (4) Stoe imaging-plate diffraction system

The Stoe system contains a 180 mm diameter imaging plate similar to the EMBL–MAR imaging plate. A sealed tube was employed as source of graphite-monochromated  $\text{Mo } K\alpha$  radiation. The preliminary orientation matrix and unit-cell dimensions were obtained from three images at  $\varphi = 0, 45$  and  $90^\circ$ . 175 strong reflections were used for the successful indexing. High-resolution data were collected at a 100 mm crystal-to-detector distance with 360 s exposure per image; the  $\varphi$  range was  $120^\circ$  and the  $\varphi$ -increment was  $1^\circ$ . The inner image-plate radius was 10 mm, the outer radius 90 mm. Low-resolution data were collected at a 190 mm distance with 180 s exposure; the  $\varphi$  range was  $60^\circ$  and the increment  $2^\circ$ . Integration of all recorded reflections was performed by a peak-minus-background algorithm. A linear decomposition correction showed virtually no decay, so the same crystal was subsequently used for data collection (2) (Siemens). Symmetry-equivalent reflections and Friedel opposites were not merged during data processing. All programs used for data collection and processing were supplied by the manufacturer.

#### (5) MAR Research imaging-plate scanner

The commercially available MAR imaging-plate scanner (180 mm diameter) was used as a detector. The data-collection and processing strategies were similar to those described for the synchrotron data, and the same programs were employed. A sealed tube was used as radiation source for graphite-monochromated  $\text{Mo } K\alpha$  radiation. High-resolution data were collected with a 115 mm crystal-to-detector distance and 1200 s exposure per image; the  $\varphi$  range was  $120^\circ$  and the increment  $1.5^\circ$ . Low-resolution data were collected with a 240 mm

distance and 360 s exposure time; the  $\varphi$  range was  $100^\circ$  and the increment  $4.5^\circ$ . The crystal was cooled to 153 K during data collection by a nitrogen gas stream from an Oxford cryostream cooler (Cosier & Glazer, 1986).

#### (6) Four-circle diffractometer

Data were collected on a Stoe-Siemens four-circle diffractometer with graphite-monochromated Mo  $K\alpha$  radiation from a sealed tube. Integrated intensities were obtained from  $\omega/2\theta$ -scans by real-time profile fitting with variable scan speed (Clegg, 1981). A scintillation counter was employed as detector. The software used for data collection and processing was supplied by Stoe. The cell dimensions were refined against the  $\omega$  values of 30 strong reflections in the range  $20 > 2\theta > 22^\circ$  centered at  $(2\theta, \omega, \chi, \varphi)$  and  $(-2\theta, -\omega, \chi, \varphi)$ . A locally built low-temperature device using a nitrogen gas stream was used to cool the crystal to a temperature of 193 K (Kottke, 1993). Three standard reflections measured every 90 min showed no crystal decay during data collection. As a result of the large number of unique reflections and long data-collection time, only some 1000 Friedel opposites were collected in addition to the unique data.

### Results and discussion

The overall  $R_{\text{int}}$  values are given in Table 1. The synchrotron data show the best internal agreement. For the Siemens and the FAST data symmetry-equivalent reflections (but not Friedel opposites) were merged during data processing. Although we cannot compare the  $R_{\text{int}}$  values directly, the internal agreement is clearly satisfactory, with  $R_{\text{int}}$  values well below 0.1 in all cases. Such low values, typical of those frequently quoted by manufacturers, might well have led to undue optimism concerning the quality of the data. The appreciably higher  $R$  indices for the refined structures indicate that  $R_{\text{int}}$  may be relatively insensitive to some systematic errors.

The  $R_{\text{int}}$  values for Friedel opposites as a function of resolution are shown in Fig. 1; dispersion effects would be expected to be very small for this structure. The synchrotron data possess the best internal agreement; even at the highest resolution the  $R_{\text{int}}$  value is below 0.2. The values are slightly higher for the FAST data, the Siemens data and the MAR image-plate data. For the Stoe image-plate and the four-circle data the  $R_{\text{int}}$  values increase rapidly at about 1.5 Å resolution, and exceed 0.5 beyond 1.3 Å, indicating that the data outside the 1.3 Å sphere are essentially noise.

#### Comparison of the data sets by attempted structure solution

We attempted to solve the structure independently from each data set by automated Patterson interpretation (Sheldrick, 1992; Sheldrick, Dauter, Wilson, Hope &

Sieker, 1993). In this approach, the highest unique Patterson peaks greater than a given distance from the nearest lattice point are used to generate Patterson vector-superposition minimum functions, which are then searched for peaks. This function possesses the apparent space group  $P\bar{1}$ , and in the ideal case in which all peaks are resolved and a single weight vector has been used for the superposition, it consists of two images of the structure, related by an inversion centre. The problem is reduced to finding the origin shift for which one of these two images possesses the symmetry required by the true space group. For each combination of superposition vector and origin shift, atom types were assigned to the potential atoms (*i.e.* atoms for which all symmetry equivalents were present as peaks in the origin-shifted superposition minimum function). The solutions with the highest correlation coefficients between  $E_o$  and  $E_c$  (Fujinaga & Read, 1987) were then examined by hand to see if disulfide bridges were present; up to this point the procedure was entirely automatic. Using the synchrotron data in Göttingen, the automatic selection of 20 superposition vectors longer than 6 Å led to two correct solutions for the three disulfide bridges. The solution with the highest correlation coefficient revealed three bridges with S-S distances 1.94, 2.02 and 1.89 Å. Starting from these positions a tangent phase expansion followed by *E*-Fourier recycling (Sheldrick, 1982) produced an essentially complete structure. Almost simultaneously the structure was solved independently at Sandoz using the FAST data and exactly the same procedure and programs. Two disulfide bridges (1.88 and 2.07 Å) and one single S atom were correctly located. With the Siemens area-detector data, two disulfide bridges with lengths of 1.94 and 1.91 Å could be found; comparison with the solution for the synchrotron data showed that the third disulfide bridge was also present but with a length of 2.34 Å, as a result of inaccurate S-atom positions. However, the first four S atoms were sufficient to generate the rest of the structure. For these three data sets, the correct solutions were those with the highest correlation coefficient.

For the other three data sets the procedure above did not lead to solution of the structure. Even when a 'good' superposition vector from one of the correct solutions was employed, the disulfide bridges could not be located. However, the partial structure expansion succeeded with all six data sets starting from the correct positions for the six S atoms. We also tested the effect of the resolution by truncating the synchrotron data, *i.e.* discarding all reflections with resolution  $d$  less than a specified numerical value; the limiting value of  $d$  for successful automatic generation of the rest of the structure by tangent expansion followed by *E*-Fourier recycling is about 1.2 Å.

It should be noted that the process of converting  $F^2$  values to  $E^2$  by fitting the mean  $E^2$  to unity in each resolution shell would compensate for any  $\theta$ -dependent

systematic errors and so make the 'structure-solution' test insensitive to such errors.

#### Comparison by anisotropic refinement

All refinements were performed against  $F^2$  using all data from 10.0 Å resolution to the highest resolution (lowest  $d$ ) collected. The anisotropic refinement was monitored using  $R_{\text{free}}$  (Brünger, 1992, 1993). All refinements started from an isotropic model that included 35 water molecules and had been refined to convergence against all the data in the appropriate data set. For the  $R_{\text{free}}$  tests 90% of all the reflections were used as a working set for refinement and the remaining 10% were used as a reference set only to evaluate the free  $R$  value. As the model should not be biased by the reference set, the  $R_{\text{free}}$  value offers an objective measure of the quality of the model, and hence the data.

For each data set, the isotropic refinement was first repeated against the working set alone to eradicate memory effects on  $R_{\text{free}}$ . The 1,2- and 1,3-distances were restrained so that equivalent distances in chemically equivalent residues were equal with e.s.d.'s of 0.03 Å. The environments of the carbonyl C atoms were restrained to be planar, as were the aromatic ring systems (e.s.d.'s of 0.1 Å<sup>3</sup>). 'Anti-bumping' restraints were employed for the solvent molecules to prevent close contacts. For the anisotropic refinement, rigid-bond restraints (Rollet, 1970; Hirshfeld, 1976; Trueblood & Dunitz, 1983) were applied to 1,2- and 1,3-atom pairs with e.s.d.'s of 0.01 Å<sup>2</sup>, and the  $U_{ij}$  displacement parameters of atoms closer than 1.7 Å to each other were restrained to be 'similar' (e.s.d.'s 0.05 and 0.10 Å<sup>2</sup> for terminal atoms). The anisotropic displacement parameters of the water molecules and the oxalate dianion were restrained to be approximately isotropic. In addition, diffuse solvent was modelled using 'Babinet's principle' (Langridge *et al.*, 1960; Driessen *et al.*, 1989) except for the FAST data where the single variable parameter was an order of magnitude larger than for the other data sets, suggesting a  $\theta$ -dependent systematic error in the data. For all restraints the standard program values for the standard deviations were used, except for the four-circle diffractometer data where harder restraints had to be applied to the anisotropic displacement parameters (0.008 Å<sup>2</sup> for the rigid-bond restraint and 0.04 Å<sup>2</sup> for 'similarity' restraints) because some atoms became 'non-positive definite' when the default values were used. For the oxalate dianion harder restraints on the anisotropic displacement parameters were used to give reasonable anisotropic displacement parameters (0.002 Å<sup>2</sup> for the rigid-bond restraint and 0.01 Å<sup>2</sup> for 'similarity' restraints). The standard deviations for all restraints were calibrated with the  $R_{\text{free}}$  test; the default values built into the program proved to be reasonable, except for the four-circle data. Each of these refinements was carried out to convergence using the

Table 2.  $R_{\text{free}}$  tests

	1	2	3	4	5	6
$R1^*$ isotropic	0.172	0.165	0.203	0.163	0.175	0.187
$R1_{\text{free}}$ isotropic	0.192	0.198	0.233	0.268	0.214	0.354
$R1$ anisotropic	0.103	0.120	0.151	0.122	0.136	0.146
$R1_{\text{free}}$ anisotropic	0.128	0.158	0.188	0.230	0.181	0.335

$$* R1 = \frac{\sum |F_o| - |F_c|}{\sum |F_o|}, \text{ [for } F > 4\sigma(F)\text{].}$$

conjugate-gradient method to solve the least-squares normal equations (Hendrickson & Konner, 1980).

The results tabulated in Table 2 show that the anisotropic refinement reduces the working set  $R$  value substantially in all cases. On the other hand the more objective  $R_{\text{free}}$  decreases significantly only for data sets 1–5. In the case of the diffractometer data, the improvement is marginal, showing that the decrease in  $R$  for the working set is almost entirely caused by the increased number of parameters. In all other cases  $R_{\text{free}}$  falls by between 3 and 6% on anisotropic refinement, the greatest improvement being observed for the synchrotron data. The synchrotron data gave the lowest  $R_{\text{free}}$  values for both isotropic and anisotropic refinement. For the Siemens and the MAR data the  $R_{\text{free}}$  values are only a little higher. For the FAST and the Stoe image-plate data the values are significantly higher and the highest  $R_{\text{free}}$  values of all were obtained for the four-circle diffractometer data.

Each data set was then refined further against the full data by blocked full-matrix least-squares techniques. All restraints described above were applied. As the cell dimensions obtained using the *DENZO* program with EMBL imaging-plate scanner data were relatively inaccurate, the cell parameters from the four-circle diffractometer were used in these refinements. The parameters determined in each refinement are summarized in Table 3.

The lowest overall conventional  $R$  value [on  $F$  for reflections with  $I > 2\sigma(I)$ ] was obtained with the synchrotron data ( $R1 = 0.084$ ). The  $R1$  values are a little higher for the Siemens (0.104) and Stoe data (0.100), and appreciably higher for the FAST data (0.158). The  $R1$  value for the low-temperature data collected with the MAR imaging-plate scanner (0.107) is similar to those for the room-temperature Siemens and Stoe data at the same wavelength, but the four-circle data gave a significantly higher value (0.132). Fig. 2 shows  $R1$  as a function of resolution. We can estimate the resolution of the data by applying an  $R1$  limit of 0.25, and discarding the remaining data. The synchrotron data show the lowest  $R1$  values over the whole range. Even at highest resolution the  $R1$  value is well below 0.2. The low-temperature MAR imaging-plate data show a similar curve with only slightly higher values, and an effective resolution of 1.1 Å. The room-temperature Siemens data are very comparable with about the same effective resolution of 1.1 Å. For all other data sets the  $R1$  values

Table 3. *Refinement parameters*

	1	2	3	4	5	6
Unique	18951	17225	12330	22925	16326	9388
$R_{\text{int}}^*$	0.026			0.077	0.073	0.083
No. of data	18930	17201	12318	22865	16326	9365
No. of restraints	3296	3292	3297	3287	3308	3305
No. of parameters	2441	2440	2422	2432	2450	2450
No. of water positions	52	52	50	51	53	53
Weight parameters $g1, g2^\dagger$	0.1794, 8.30	0.2152, 11.65	0.3895, 1.82	0.1489, 12.80	0.2215, 29.87	0.2295, 133.0
$R1$ value [ $I > 2\sigma(I)$ ]	0.0843	0.104	0.158	0.100	0.107	0.132
$wR2$ (for all data) $^\ddagger$	0.246	0.292	0.437	0.334	0.309	0.440
$S$ (on $F^2$ ) $^\S$	1.063	1.055	1.135	1.110	1.028	1.086
Abs. structure parameters $^\P$	0.41 (10)	0.09 (3)	0.15 (4)	0.05 (14)	0.1 (2)	0.1 (4)
Largest diff. peak ( $e \text{ \AA}^{-3}$ )	0.351	0.384	0.474	0.268	0.717	0.549
Lowest diff. hole ( $e \text{ \AA}^{-3}$ )	-0.590	-0.438	-0.540	-0.307	-0.391	-0.516

\* Calculated with *SHELXL93*, Friedel pairs not merged.  $R_{\text{int}} = \sum (F_o^2 - F_c^2(\text{mean})) / \sum (F_o^2)$ .

$^\ddagger wR2 = \{\sum [w(F_o^2 - F_c^2)^2] / \sum [w(F_o^2)^2]\}^{1/2}$ .

$^\dagger w^{-1} = \sigma^2(F_o^2) + (g1 \cdot P)^2 + g2 \cdot P$ , where  $P = (F_o^2 + 2F_c^2)/3$ .

$^\S S = \{\sum [w(F_o^2 - F_c^2)^2] / (n - p)\}^{1/2}$ .

$^\P$  Flack (1983).

increase more rapidly at higher resolution. For the FAST data we find an effective resolution of 1.2 Å, for the Stoe imaging-plate diffraction system 1.25 Å and for the low-temperature four-circle data 1.4 Å. The Stoe imaging-plate data show relatively low  $R1$  values at low resolution, whereas at higher resolution the  $R1$  increases dramatically.

The resulting structures are essentially the same. The bond lengths, angles and torsion angles are very comparable. The r.m.s. deviations of the  $C\alpha$ -atom positions relative to the refined structure for the synchrotron data are less than 0.1 Å for the low-temperature data sets and less than 0.05 Å for the room-temperature data sets. There are some differences in the solvent structure, but with every data set one oxalate dianion and between 48 and 52 water positions were found and refined. In addition the positions of the water molecules are very similar. Better criteria for the accuracy of a model are the bond lengths and angles and their estimated standard deviations. These values have been checked using the independent program *PROCHECK* (Laskowski, MacArthur, Moss & Thornton, 1993). Table 4 summarizes the mean, maximum and minimum values for CA—N, CA—C, CA—CB, C=O and C—N distances with the range of their e.s.d.'s. For all data sets the mean values agree within their e.s.d.'s with the standard values given by Engh & Huber (1991). As a result of the reduced librational effects the bond lengths tend to be slightly longer for the low-temperature than for the room-temperature data sets. The largest e.s.d.'s were obtained with the four-circle data and the FAST data, and the lowest e.s.d.'s with the synchrotron data. The e.s.d.'s from the Siemens data, the Stoe image-plate and the MAR image-plate data all lie in the same range.

50% probability displacement-parameter plots of one of the three unique peptide molecules (designated molecule I in the preceding paper) are shown in Fig. 3

for each data set. For the area-detector data sets the ellipsoids appear to make chemical sense. They are somewhat larger for the FAST data and slightly smaller for the EMBL/MAR Research low-temperature data. For the four-circle data the ellipsoids are significantly smaller and some atoms are close to 'non-positive-definite', although harder restraints were applied. We suggest that the four-circle data do not justify anisotropic refinement.

The experimental unit-cell dimensions are given in Table 5. The values are close to those obtained with the four-circle diffractometer, except for the two sets of values obtained at EMBL, Hamburg using the program *DENZO*, which differ appreciably. We suspect that these cell determinations are subject to systematic errors because the crystal-to-detector distance was not adequately calibrated. For example the  $c$  axis is almost 0.9 Å shorter for the low-temperature data collection using the MAR system and *DENZO* than the value determined at low temperature with the four-circle diffractometer

### Concluding remarks

A strict comparison of the data sets is not possible because there was too much variation in data-collection conditions and parameters. In addition, different data-processing strategies and programs were used. Nevertheless, some important conclusions can be drawn. Under all conditions investigated, data of adequate quality for anisotropic refinement were obtained, except using the diffractometer data. Structure solution succeeded only with data sets collected with stronger X-ray sources (synchrotron or rotating anode). The solution was most straightforward using data collected with the strongest X-ray source, synchrotron radiation.

The four-circle diffractometer data gave the second highest  $R$  values and the highest e.s.d.'s for bond lengths

Table 4. Average, minimum and maximum bond lengths and angles, and range of e.s.d.'s (Å, °)

	1	2	3	4	5	6	Standard†
C—N							
Average	1.317	1.320	1.337	1.313	1.338	1.314	1.329 (1.4)
Min.	1.267	1.285	1.308	1.272	1.306	1.283	
Max.	1.346	1.361	1.369	1.353	1.378	1.380	
C—O							
Average	1.229	1.224	1.235	1.233	1.255	1.249	1.231 (2.0)
Min.	1.200	1.186	1.194	1.182	1.220	1.212	
Max.	1.261	1.261	1.286	1.274	1.320	1.288	
C—CA							
Average	1.511	1.519	1.500	1.507	1.521	1.534	1.525 (2.1)
Min.	1.483	1.487	1.450	1.487	1.482	1.475	
Max.	1.531	1.553	1.561	1.548	1.587	1.575	
CA—N							
Average	1.448	1.468	1.477	1.458	1.467	1.458	1.458 (1.9)
Min.	1.397	1.422	1.405	1.408	1.434	1.420	
Max.	1.491	1.505	1.551	1.496	1.492	1.517	
CA—CB*							
Average	1.523	1.528	1.533	1.537	1.559	1.553	1.530 (2.0)
Min.	1.489	1.502	1.484	1.506	1.528	1.518	
Max.	1.556	1.560	1.575	1.560	1.591	1.607	
E.s.d.'s (Å)	0.004–0.013	0.009–0.018	0.015–0.025	0.010–0.019	0.010–0.021	0.014–0.024	
CA—C—N							
Average	116.9	116.0	115.8	117.1	117.3	116.3	116.2 (2.0)
Min.	113.8	112.4	111.1	114.1	113.8	112.2	
Max.	120.2	119.9	120.1	121.1	122.2	122.0	
N—C—O							
Average	122.7	123.3	122.5	122.2	122.0	123.0	123.0 (1.6)
Min.	120.5	120.2	119.4	120.5	120.1	118.1	
Max.	125.2	125.9	125.6	124.0	124.2	126.4	
C—N—CA							
Average	123.2	122.5	122.4	123.5	123.2	122.0	121.7 (1.8)
Min.	120.8	119.1	118.6	120.3	120.3	116.4	
Max.	127.1	126.4	125.4	127.7	127.4	128.1	
CA—C—O							
Average	120.4	120.7	121.7	120.6	120.7	120.7	120.8 (1.7)
Min.	115.8	116.9	116.5	116.2	117.3	116.2	
Max.	122.5	123.9	125.6	123.5	123.1	123.8	
N—CA—C							
Average	110.7	110.1	110.3	110.4	110.9	110.0	111.2 (2.8)
Min.	106.9	106.5	105.9	105.9	107.0	104.7	
Max.	113.7	113.0	113.5	114.3	114.8	113.8	
E.s.d.'s (°)	0.7–1.3	0.6–1.5	1.2–2.5	0.6–1.5	0.7–1.6	2.1–3.3	

\* Not including threonine and valine.

† Standard values from Engh &amp; Huber (1991).

Table 5. Unit-cell dimensions

	1	1*	2	3	4	5	6
Temperature (K)	285	293	293	293	293	153	193
<i>a</i> (Å)	18.57	18.458 (5)	18.54	18.55	18.49 (1)	18.20	18.328 (9)
<i>b</i> (Å)	30.34	30.009 (7)	30.11	30.18	30.12 (1)	30.19	30.130 (60)
<i>c</i> (Å)	39.74	39.705 (27)	39.80	39.70	39.73 (2)	38.87	39.424 (5)
<i>V</i> (Å <sup>3</sup> )	22390	21993	22126	22226	22126	21357	21770

\* Values obtained with a Stoe four-circle diffractometer at room temperature. These values were used in the refinement of the room-temperature synchrotron data set (1).

and angles. This is probably caused by the relative weakness of the data. Beyond 1.4 Å resolution less than half of the reflections are 'observed' [ $I > 2\sigma(I)$ ], and between 1.1 and 1.2 Å only 20% of the data were 'observed'. This data set has the lowest mean( $I$ )/mean[ $\sigma(I)$ ] value, and the lowest redundancy. As a result of the much longer data-collection time on the diffractometer, only about 1000 Friedel pairs and no symmetry-equivalent reflections were collected. Despite the weakness of the diffractometer data between 1.1 and 1.4 Å, it was possible to generate almost the complete structure starting

from the six S atoms, although this is no longer possible for the synchrotron data if they are truncated to (e.g.) 1.3 Å. Clearly for structure solution it is important to identify a small number of high  $E$  values at the highest possible resolution, because these contain a great deal of structural information, even if most of the data in this resolution range are almost pure noise.

The data collected with the MAR imaging-plate resulted in a slightly better model than the Stoe data, although the hardware configuration of these two systems is very similar. In particular the high-resolution

data seem to be more accurate. This can primarily be attributed to the lower temperature of data collection, but there may also be differences due to different data processing and, especially, integrating strategies. In addition the exposure time was about three times longer using the MAR detector, which must also improve the signal-to-noise ratio. The effective resolution is significantly higher for the low-temperature data.

The Siemens X200B detector performed appreciably better than the FAST system in all our tests, despite a similar X-ray source, wavelength and data-collection temperature. In particular the high-resolution data are clearly more accurate. The FAST data gave the highest  $R$  values of all area-detector data sets. The Siemens room-temperature data are a little more precise than the low-temperature data from the MAR imaging-plate scanner; especially when  $R_{\text{free}}(\text{anis})$  is regarded as the most reliable criterion.  $R_{\text{free}}(\text{iso})$  is less suitable because the atoms might be expected to behave more 'isotropically' at lower temperature. The effective resolution of these two data sets is, however, very similar, despite very different data-collection conditions, hardware and data-collection strategies.

The best data were obtained with synchrotron radiation and the EMBL imaging-plate scanner. The data have the highest effective resolution and the refinement produced the best model. In view of the other results we feel that this should almost entirely be attributed to the much more brilliant radiation source.

We wish to thank Dr F. Hahn (Stoe & Cie) and Dr N. P. C. Walker (BASF) for their considerable help in the data collections on the Stoe and the Siemens systems, respectively. We are grateful to the Deutsche Forschungsgemeinschaft (Leibniz-Programm) for financial support.

### References

- ARNDT, U. W. & WONACOTT, A. J. (1977). Editors. *The Rotation Method in Crystallography*. Amsterdam: North Holland.
- BLUM, M., METCALF, P., HARRISON, S. C. & WILEY, D. C. (1987). *J. Appl. Cryst.* **20**, 235–242.

- BRÜNGER, A. (1992). *Nature (London)*, **355**, 472–475.
- BRÜNGER, A. (1993). *Acta Cryst.* **D49**, 24–36.
- CLEGG, W. (1981). *Acta Cryst.* **A37**, 22–28.
- COSIER, J. & GLAZER, A. M. (1987). *J. Appl. Cryst.* **19**, 105–107.
- DRIESSEN, H., HANEEF, M. I. J., HARRIS, G. W., HOWLIN, B., KHAN, G. & MOSS, D. S. (1989). *J. Appl. Cryst.* **22**, 510–516.
- ENGH, R. A. & HUBER, R. (1991). *Acta Cryst.* **A47**, 392–400.
- FLACK, H. D. (1983). *Acta Cryst.* **A39**, 876–881.
- FUJINAGA, M. & READ, R. J. (1987). *J. Appl. Cryst.* **20**, 517–521.
- HENDRICKSON, W. A. & KONNERT, J. H. (1980). *Computing in Crystallography*, edited by R. DIAMOND, S. RAMASESHAN & K. VENKATESAN, pp. 31.01–31.23. Bangalore: Indian Academy of Sciences/IUCR.
- HIRSHFELD, F. L. (1976). *Acta Cryst.* **A32**, 239–244.
- HOPE, H. (1988). *Acta Cryst.* **B44**, 22–26.
- HOWARD, A. J., GILLILAND, G. L., FINZEL, B. C., POULOS, T. L., OHLENDORF, D. H. & SALEMME, F. R. (1987). *J. Appl. Cryst.* **20**, 383–387.
- KABSCH, W. (1988). *J. Appl. Cryst.* **21**, 916–924.
- KOTTKE, T. (1993). PhD thesis, Univ. of Göttingen, Germany.
- KOTTKE, T. & STALKE, D. (1993). *J. Appl. Cryst.* **26**, 615–619.
- KRAUSE, K. L. & PHILLIPS, G. N. JR (1992). *J. Appl. Cryst.* **25**, 146–154.
- LANGRIDGE, R., MARVIN, D. A., SEEDS, W. E., WILSON, H. R., HOOPER, C. W., WILKINS, M. H. F. & HAMILTON, L. D. (1960). *J. Mol. Biol.* **2**, 38–64.
- LASKOWSKI, R. A., MACARTHUR, M. W., MOSS, D. S. & THORNTON, J. M. (1993). *J. Appl. Cryst.* **26**, 283–291.
- MESSERSCHMIDT, A. & PFLUGRATH, J. W. (1987). *J. Appl. Cryst.* **20**, 306–315.
- OTWINOWSKI, Z. (1991). *DENZO. A Film Processing Program for Macromolecular Crystallography*. Yale Univ., New Haven, CT, USA.
- POHL, E., HEINE, A., SHELDRIK, G. M., DAUTER, Z., WILSON, K. S., KALLEN, J., HUBER, W. & PFÄFFLI, P. J. (1995). *Acta Cryst.* **D51**, 48–59.
- ROLLET, J. S. (1970). *Crystallographic Computing*, edited by F. R. AHMED, S. R. HALL & C. P. HUBER, pp. 167–181. Copenhagen: Munksgaard.
- SERC Daresbury Laboratory (1979). *CCP4. A Suite of Programs for Protein Crystallography*. SERC Daresbury Laboratory, Warrington WA4 4AD, England.
- SHELDRIK, G. M. (1982). *Computational Crystallography*, edited by D. SAYRE, pp. 506–514. Oxford: Clarendon Press.
- SHELDRIK, G. M. (1990). *Acta Cryst.* **A46**, 467–473.
- SHELDRIK, G. M. (1992). *Crystallographic Computing*, edited by D. MORAS, A. D. PODJARNY & J. C. THIERRY, pp. 145–147. Oxford: IUCr/Oxford Univ. Press.
- SHELDRIK, G. M. (1993). *SHELXL93. Program for Crystal Structure Refinement*. Univ. of Göttingen, Germany.
- SHELDRIK, G. M., DAUTER, Z., WILSON, K. S., HOPE, H. & SIEKER, L. C. (1993). *Acta Cryst.* **D49**, 18–22.
- TRUEBLOOD, K. N. & DUNITZ, J. P. (1983). *Acta Cryst.* **B39**, 120–144.
- TUCKER, P. A. (1990). Proceedings of the CCP4 Study Weekend, Accuracy and Reliability of Macromolecular Crystal Structures, 26–27 January 1990, SERC Daresbury Laboratory, Warrington, England.