

What is the best crystal size for collection of X-ray data? Refinement of the structure of glycyl-L-serine based on data from a very large crystal

CARL HENRIK GÖRBITZ

Department of Chemistry, University of Oslo, PO Box 1033 Blindern, N-0315 Oslo, Norway.

E-mail: c.h.gorbitz@kjemi.uio.no

(Received 22 March 1999; accepted 28 June 1999)

Abstract

The dipeptide Gly-L-Ser was crystallized as part of a study on hydrogen-bonding patterns in the structures of dipeptides. Hydrogen-bond donors and acceptors have been assigned ranks (1 is best, 2 is next best *etc.*), and the observed hydrogen-bond connectivity is compared with the hypothetical pattern in which the rank n donor associates with the rank n acceptor ($n = 1, 2, \dots$), and with the pattern observed in the retroanalogue L-Ser-Gly, which contains the same functional groups. Crystallization of the title compound produced very bulky crystals. Rather than reducing the size of one of these before data collection, three data sets with different exposure times were collected with a Siemens SMART CCD diffractometer on a very large specimen ($2.2 \times 2.0 \times 0.8$ mm). The crystal was subsequently shaped into a 0.30 mm-diameter sphere for collection of two additional data sets. The discussion of the refinement results focus on the effect of absorption correction for the various data sets, and a comparison of geometrical and thermal parameters. One advantage of using a large crystal, the great speed with which data can be obtained, has been exemplified by collection of a complete data set of good quality in less than 25 min.

1. Introduction

One of the empirical 'hydrogen bond rules' (Etter, 1990) states that 'the best proton donor and acceptor remaining after intramolecular hydrogen-bond formation form intermolecular hydrogen bonds to one another'. This rule may tentatively be extended to say that the next best donor will form a hydrogen bond to the next best acceptor, and so on. However, the progressively tighter steric constraints mean that it is difficult for all acceptors and donors to associate in strict rank order. Still, identical hydrogen-bond pairing was observed for L-Val-L-Glu (Eggleston, 1984) and L-Glu-L-Val (Görbitz & Backe, 1996), even though the crystal-packing arrangements are different. Wondering if a similar observation could be made for a dipeptide without hydrophobic entities, we decided to study the crystal structure of the title compound, since the struc-

ture of the retroanalogue L-Ser-Gly had already been presented (Jones *et al.*, 1978).

The crystallization experiments with the title compound yielded some very large crystals. Regarding the preferred crystal size for data collection, the following advice is found in a classical textbook: 'In the usual apparatus for collecting diffraction data, a plateau of uniform intensity in the primary beam of dimensions 0.5×0.5 mm or a bit larger can be obtained. In single-crystal work, the specimen should not exceed this size, *i.e.* all parts of the crystal should be exposed to the same radiation intensity. The limit set by this plateau, however, is larger than the usually preferred crystal size of 0.1 to 0.3 mm, partially because of the difficulties in aligning the crystal precisely with a plateau of the same size' (Stout & Jensen, 1989). Most crystallographers adhere to these guidelines in their experimental work, and preferentially use small crystals (< 0.5 mm) for data collection, or reduce large ones in size if required. Examples of data collections with crystals larger than the incident X-ray beam are regarded with skepticism. However, this situation might change with the advent of diffractometers equipped with CCD detectors that render possible fast collection of high-redundancy data sets, and concomitant software, like *SADABS* (Sheldrick, 1996),[†] that perform empirical modification of highly redundant experimental data to correct not only for absorption effects but supposedly also for the effects of variable irradiated sample volume, beam non-uniformity and beam misalignment (which, if present, can seriously aggravate the non-uniformity). If these computer programs can really correct for systematic errors introduced when a large crystal is used, then why not use it as it is? The potential benefit is, apart from the elimination of the work associated with reduction in size, the high diffraction intensities and the possibility of performing very fast data collections. The current paper raises the question if this is indeed a viable and acceptable approach, illustrated by refinements of several data sets for the dipeptide Gly L-Ser obtained from both a very large crystal and a small spherical crystal.

[†] Although written specifically for the Siemens (now Bruker) CCD and multiwire detectors, *SADABS* can also be used for serial diffractometers provided that the redundancy is sufficiently high.

2. Experimental

2.1. Preparation

Large crystals (several mm) were grown by diffusion of 2-propanol into 30 ml of an aqueous solution containing approximately 2 mg of Gly-L-Ser. The smallest crystal, a distorted bipyramid with dimensions $2.20 \times 2.0 \times 0.80$ mm, was used in the initial data collections. The crystal was subsequently cut and then shaped into a 0.30 mm-diameter sphere by grinding for the final data collections.

2.2. Data collection and absorption correction

Five data sets were collected on a Siemens SMART CCD diffractometer, sets 1, 2 and 3 using the large crystal, sets 4 and 5 using the small spherical crystal. The data collections with *SMART* (Siemens, 1995) nominally covered over a hemisphere of reciprocal space, by a combination of six sets of exposures; two with the detector set at $2\theta = 35^\circ$ and four with $2\theta = 75^\circ$, except for set 3 which consisted of only three sets of exposures, all with $2\theta = 38^\circ$. Additionally, the first 30 frames of set 1 were always recollected at the end to check for intensity decay. For set 1, the exposure time for each frame was 10 s for a 0.6° ω scan, while 2 s 1.2° scans were used for set 2, corresponding to a 90% reduction in total exposure time. Speed was optimized for set 3 (which will not be discussed in detail) by using 2 s exposure time and 1.8° scan with no correlation check. For set 4, conditions were equivalent to set 1, while a tenfold increase in exposure time, to 100 s, was used for set 5. Coverage of the unique set for the four high-angle data sets is virtually complete up to the 110° limit for 2θ , while coverage is complete to $2\theta = 70^\circ$ for set 3.

Data sets obtained by modification of the original *Lp*-corrected data by *SADABS* (Sheldrick, 1996), using only default settings, have been denoted sets 1a, 2a, 3a, 4a and 5a. *SADABS* corrects for various absorption effects as well as for variable crystal volume in diffracting position, beam non-uniformity, crystal decomposition and primary beam intensity variation.

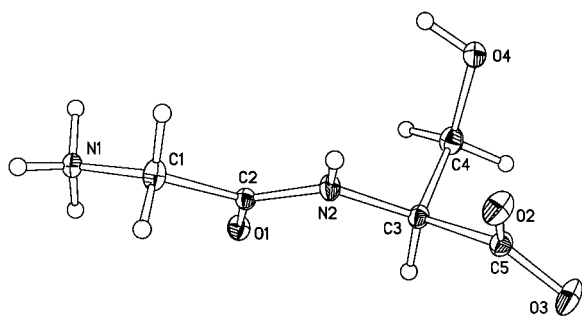


Fig. 1. The asymmetric unit with atomic numbering. Thermal ellipsoids are shown at the 50% probability level. H atoms are shown as spheres of arbitrary size.

For the present work there is no indication of the latter two effects being present. As far as absorption effects are concerned, they originate not only from the crystal but also from the air, the glass rod and the glue used to attach the crystal. Measures were taken to minimize these effects by using (i) a minimum crystal-to-detector distance of 3 cm, (ii) a thin boron silicate glass capillary for mounting and (iii) very little epoxy glue.

2.3. Structure determination and refinement

After integration with *SAINT* (Siemens, 1995), the structure of Gly-L-Ser was solved routinely with *SHELXTL* (Sheldrick, 1997) in the space group $P2_12_12_1$, and refined with anisotropic non-H atoms and isotropic H atoms. No constraints or restraints were used.

Crystal data, experimental conditions and refinement results are summarized in Table 1.† The molecular structure of Gly-L-Ser is shown in Fig. 1.

3. Results and discussion

3.1. Cell parameters

When comparing molecular geometry (and other) parameters derived from several data sets for the same compound, differences arise from both different fractional coordinates x , y and z , and different cell dimensions. Since all data sets used for the present study were collected from the same crystal, it is reasonable to assume that the second factor can be disregarded. Nevertheless, sets of cell dimensions from integration of the four high-angle data sets 1, 2, 4 and 5 are significantly different, judged by the very low s.u.'s, typically 0.0001 \AA . These s.u.'s result from a statistical analysis of a series of observations (reflections; type A evaluation of uncertainty; Schwarzenbach *et al.*, 1995) and do not take into account systematic errors. Systematic errors can be introduced for instance by absorption effects, but for integration of CCD data with *SAINT* (Siemens, 1995) it appears that the bias introduced by choosing, from the full experimental material, a specific subset of reflections for determination of cell parameters, and the sensibility to the initial orientation matrix are more important factors. This was established by a type B evaluation of uncertainty (by means other than the statistical analysis of a series of observations; Schwarzenbach *et al.*, 1995) based on the results from 12–15 integrations of each data set, with marginal differences in the initial orientation matrices and the maximum 2θ value for reflections included.‡ The s.u.'s for cell parameters derived this way are generally in the range 0.002–

† Supplementary data for this paper are available from the IUCr electronic archives (Reference: OS0029). Services for accessing these data are described at the back of the journal.

‡ An upper limit was used as calculated cell dimensions decrease slightly when very high angle reflections are included, presumably owing to $\alpha_1\alpha_2$ splitting.

Table 1. *Experimental details*

	Sets 1–5				
	1	1a	2a	3a	4
Crystal data					
Chemical formula	C ₅ H ₁₀ N ₂ O ₄				
Chemical formula weight	162.15				
Cell setting	Orthorhombic				
Space group	P2 ₁ 2 ₁ 2 ₁				
<i>a</i> (Å)	7.275 (2)†				
<i>b</i> (Å)	9.097 (3)				
<i>c</i> (Å)	10.507 (3)				
<i>V</i> (Å ³)	695.4 (4)				
<i>Z</i>	4				
<i>D_x</i> (Mg m ⁻³)	1.549				
Radiation type	Mo <i>K</i> α				
Wavelength (Å)	0.71073				
μ (mm ⁻¹)	0.134				
Temperature (K)	150 (2)				
Crystal form	Distorted bipyramid for 1, 2 and 3; sphere for 4 and 5				
Crystal size; radius (mm)	2.20 × 2.00 × 0.80; 0.15				
Crystal colour	Colourless				
Data collection					
Diffractionmeter	Siemens SMART CCD	Siemens SMART CCD	Siemens SMART CCD	Siemens SMART CCD	Siemens SMART CCD
Data collection method	Sets of exposures each taken over 0.6°ω rotation scans	Sets of exposures each taken over 0.6°ω rotation scans	Sets of exposures each taken over 1.2°ω rotation scans	Sets of exposures each taken over 1.8°ω rotation scans	Sets of exposures each taken over 0.6°ω rotation scans
Absorption correc- tion	None	Empirical	Empirical	Empirical	None
<i>T</i> _{min}	–	0.7570	0.7570	0.7570	–
<i>T</i> _{max}	–	0.9004	0.9004	0.9004	–
Exposure time per frame (s)	10	10	2	2	10
Total data collection time (h)	6.7	6.7	1.8‡	0.4	6.7
No. of measured reflections	28 701	28 701	28 444	9507	28 767
No. of independent reflections	8814	8814	8903	3467	8721
No. of observed reflections	8225	8233	7186	2948	6860
Criterion for observed reflections	<i>I</i> > 2σ(<i>I</i>)	<i>I</i> > 2σ(<i>I</i>)	<i>I</i> > 2σ(<i>I</i>)	<i>I</i> > 2σ(<i>I</i>)	<i>I</i> > 2σ(<i>I</i>)
<i>R</i> _{int}	0.1002	0.0276	0.0699	0.0457	0.0576
θ _{max} (°)	55.07	55.07	55.46	37.72	55.12
Range of <i>h</i> , <i>k</i> , <i>l</i>	–16 → <i>h</i> → 16 –17 → <i>k</i> → 20 –20 → <i>l</i> → 24	–16 → <i>h</i> → 16 –17 → <i>k</i> → 20 –20 → <i>l</i> → 24	–16 → <i>h</i> → 16 –17 → <i>k</i> → 20 –20 → <i>l</i> → 24	–12 → <i>h</i> → 11 –14 → <i>k</i> → 15 –17 → <i>l</i> → 17	–16 → <i>h</i> → 16 –19 → <i>k</i> → 20 –24 → <i>l</i> → 24
Refinement					
Refinement on	<i>F</i> ²	<i>F</i> ²	<i>F</i> ²	<i>F</i> ²	<i>F</i> ²
<i>R</i> [<i>F</i> ² > 2σ(<i>F</i> ²)]	0.0425	0.0315	0.0618§	0.0350	0.0533
<i>wR</i> (<i>F</i> ²)	0.1194	0.0848	0.1446	0.0878	0.1150
<i>S</i>	1.133	1.190	1.310	1.069	1.212
No. of reflections used in refinement	8814	8814	8903	3467	8721
No. of parameters used	141	141	141	141	140
H-atom treatment	All H-atom para- meters refined	All H-atom para- meters refined	All H-atom para- meters refined	All H-atom para- meters refined	All H-atom para- meters refined
Weighting scheme	$w = 1/[\sigma^2(F_o^2) + (0.0484P)^2 + 0.0508P]$, where $P = (F_o^2 + 2F_c^2)/3$	$w = 1/[\sigma^2(F_o^2) + (0.0405P)^2 + 0.0338P]$, where $P = (F_o^2 + 2F_c^2)/3$	$w = 1/[\sigma^2(F_o^2) + (0.0323P)^2 + 0.1795P]$, where $P = (F_o^2 + 2F_c^2)/3$	$w = 1/[\sigma^2(F_o^2) + (0.0500P)^2]$, where $P = (F_o^2 + 2F_c^2)/3$	$w = 1/[\sigma^2(F_o^2) + (0.0334P)^2 + 0.1007P]$, where $P = (F_o^2 + 2F_c^2)/3$

Table 1 (cont.)

	1	1a	2a	3a	4
$(\Delta/\sigma)_{\max}$	0.002	0.003	0.004	0.002	0.001
$\Delta\rho_{\max}$ ($e \text{ \AA}^{-3}$)	0.478	0.378	0.483	0.310	0.567
$\Delta\rho_{\min}$ ($e \text{ \AA}^{-3}$)	-0.434	-0.341	-0.464	-0.248	-0.352
Extinction method	<i>SHELXTL</i> (Sheldrick, 1997)	<i>SHELXTL</i> (Sheldrick, 1997)	<i>SHELXTL</i> (Sheldrick, 1997)	<i>SHELXTL</i> (Sheldrick, 1997)	None
Extinction coefficient	0.082 (11)	0.095 (8)	0.137 (10)	0.228 (14)	0
Source of atomic scattering factors	<i>International Tables for Crystallography</i> (1992, Vol. C)	<i>International Tables for Crystallography</i> (1992, Vol. C)	<i>International Tables for Crystallography</i> (1992, Vol. C)	<i>International Tables for Crystallography</i> (1992, Vol. C)	<i>International Tables for Crystallography</i> (1992, Vol. C)
Computer programs					
Data collection	<i>SMART</i> (Siemens, 1995)	<i>SMART</i> (Siemens, 1995)	<i>SMART</i> (Siemens, 1995)	<i>SMART</i> (Siemens, 1995)	<i>SMART</i> (Siemens, 1995)
Data reduction	<i>SAINT</i> (Siemens, 1995)	<i>SAINT</i> (Siemens, 1995)	<i>SAINT</i> (Siemens, 1995)	<i>SAINT</i> (Siemens, 1995)	<i>SAINT</i> (Siemens, 1995)
Structure solution	<i>SHELXTL</i> (Sheldrick, 1997)	<i>SHELXTL</i> (Sheldrick, 1997)	<i>SHELXTL</i> (Sheldrick, 1997)	<i>SHELXTL</i> (Sheldrick, 1997)	<i>SHELXTL</i> (Sheldrick, 1997)
Structure refinement	<i>SHELXTL</i> (Sheldrick, 1997)	<i>SHELXTL</i> (Sheldrick, 1997)	<i>SHELXTL</i> (Sheldrick, 1997)	<i>SHELXTL</i> (Sheldrick, 1997)	<i>SHELXTL</i> (Sheldrick, 1997)
Preparation of material for publication	<i>SHELXTL</i> (Sheldrick, 1997)	<i>SHELXTL</i> (Sheldrick, 1997)	<i>SHELXTL</i> (Sheldrick, 1997)	<i>SHELXTL</i> (Sheldrick, 1997)	<i>SHELXTL</i> (Sheldrick, 1997)
		4a		5	
Data collection					
Diffractometer		Siemens SMART CCD		Siemens SMART CCD	
Data collection method		Sets of exposures each taken over $0.6^\circ\omega$ rotation scans		Sets of exposures each taken over $0.6^\circ\omega$ rotation scans	
Absorption correction		Empirical		None	
T_{\min}		0.961		-	
T_{\max}		0.961		-	
Exposure time per frame (s)		10		100	
Total data collection time (h)		6.7		42.3	
No. of measured reflections		28 767		29 051	
No. of independent reflections		8721		8738	
No. of observed reflections		6864		8058	
Criterion for observed reflections		$I > 2\sigma(I)$		$I > 2\sigma(I)$	
R_{int}		0.0560		0.0255	
θ_{\max} ($^\circ$)		55.12		55.26	
Range of h, k, l		-16 \rightarrow h \rightarrow 16 -19 \rightarrow k \rightarrow 20 -24 \rightarrow l \rightarrow 24		-16 \rightarrow h \rightarrow 16 -19 \rightarrow k \rightarrow 20 -24 \rightarrow l \rightarrow 24	
Refinement					
Refinement on		F^2		F^2	
$R[F^2 > 2\sigma(F^2)]$		0.0539		0.0277	
$wR(F^2)$		0.1184		0.0736	
S		1.213		1.161	
No. of reflections used in refinement		8721		8738	
No. of parameters used		140		140	
H-atom treatment		All H-atom parameters refined		All H-atom parameters refined	
Weighting scheme		$w = 1/[\sigma^2(F_o^2) + (0.0363P)^2 + 0.1000P]$, where $P = (F_o^2 + 2F_c^2)/3$		$w = 1/[\sigma^2(F_o^2) + (0.0375P)^2 + 0.0200P]$, where $P = (F_o^2 + 2F_c^2)/3$	
$(\Delta/\sigma)_{\max}$		0.002		0.003	
$\Delta\rho_{\max}$ ($e \text{ \AA}^{-3}$)		0.598		0.399	
$\Delta\rho_{\min}$ ($e \text{ \AA}^{-3}$)		-0.340		-0.256	
Extinction method		None		None	
Source of atomic scattering factors		<i>International Tables for Crystallography</i> (1992, Vol. C)		<i>International Tables for Crystallography</i> (1992, Vol. C)	

Table 1 (cont.)

	4a	5
Computer programs		
Data collection	<i>SMART</i> (Siemens, 1995)	<i>SMART</i> (Siemens, 1995)
Data reduction	<i>SAINT</i> (Siemens, 1995)	<i>SAINT</i> (Siemens, 1995)
Structure solution	<i>SHELXTL</i> (Sheldrick, 1997)	<i>SHELXTL</i> (Sheldrick, 1997)
Structure refinement	<i>SHELXTL</i> (Sheldrick, 1997)	<i>SHELXTL</i> (Sheldrick, 1997)
Preparation of material for publication	<i>SHELXTL</i> (Sheldrick, 1997)	<i>SHELXTL</i> (Sheldrick, 1997)

† For discussion of the cell dimensions, see text. ‡ Owing to dead time between exposures the total data collection time for the combination 1 s exposure/0.6° scan would have been about 3.2 h. § This data set refines to $R(F) = 0.0311$ for reflections with $2\theta < 70^\circ$.

0.003 Å. It follows that there are no significant differences between the cell parameters obtained from the four data sets. The cell parameters given in Table 1, used throughout this paper, are the mean values from 12 different integrations of data set 5.

3.2. Absorption correction

The effects of using *SADABS* (Sheldrick, 1996) on data set 1 and data set 4 to produce the new data sets 1a and 4a, respectively, are very different (Table 1). For set 1, *SADABS* gives a dramatic reduction for R_{int} from 0.1002 to 0.0276 for set 1a with smaller, but still very significant, improvements for $R(F)$ and $wR(F^2)$ in subsequent refinements. In contrast, when *SADABS* operates on data set 4 no improvement is seen. In fact, despite a small decrease for R_{int} , from 0.0576 to 0.0560, the final values for $R(F)$ and $wR(F^2)$ are a little higher for set 4a than for set 4. Data for refinements of sets 2 and 5a have not been tabulated, but the effects of *SADABS* closely parallels those observed for sets 1/1a and 4/4a, respectively [2: $R_{\text{int}} = 0.1196$, $R(F) = 0.0683$, $wR(F^2) = 0.1788$; 5a: $R_{\text{int}} = 0.0220$, $R(F) = 0.0280$, $wR(F^2) = 0.0748$]. In general, results from sets 4 and 4a are very similar, as are results from sets 5 and 5a. The

discussion below will primarily describe results from sets 4 and 5.

It is noteworthy that even for the large crystal the true absorption effects are modest ($\mu = 0.134 \text{ mm}^{-1}$), so that the modifications made by *SADABS* can be attributed mainly to crystal size and beam non-uniformity effects. One might therefore argue that the term 'absorption correction' is not the most appropriate for the current structural study.

3.3. The R factor

The reduction in crystal size from sets 1a to 4 has radically decreased average diffraction intensities and consequently increased the number of unobserved reflections (Fig. 2), even though the same exposure time was used for both data sets. It follows that $R(F)$ and $wR(F^2)$ are much lower for set 1a than for set 4 (Table 1), an effect that is especially large at high 2θ values (Fig. 3). A rather similar effect is seen with the reduction in exposure time from sets 1a to 2a, and Fig. 3 shows that sets 2a and 4 in fact have rather similar R factors at all 2θ intervals, even though the number of unobserved reflections is higher for set 4 (Fig. 2). Not surprisingly, refinement results for data derived from the small crystal improve with the tenfold increase in exposure time from sets 4 to 5. The quality of data set 5 is exceptionally high, with refinements converging at

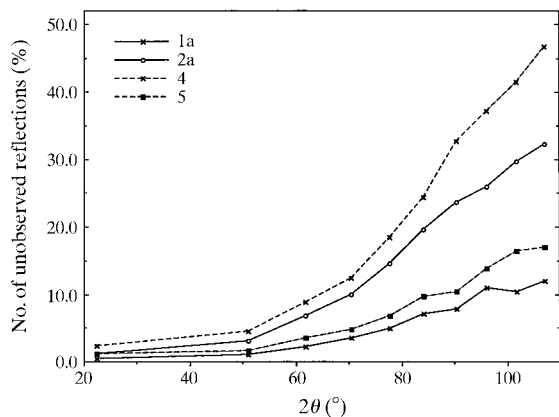


Fig. 2. Proportion of unobserved reflections for various data sets. For Figs. 2 and 3 each data point represents a set of reflections covering a range of 2θ values, e.g. $0\text{--}44.7^\circ$ for the first, $44.7\text{--}57.3^\circ$ for the second etc. About the same number of reflections are included for each data point.

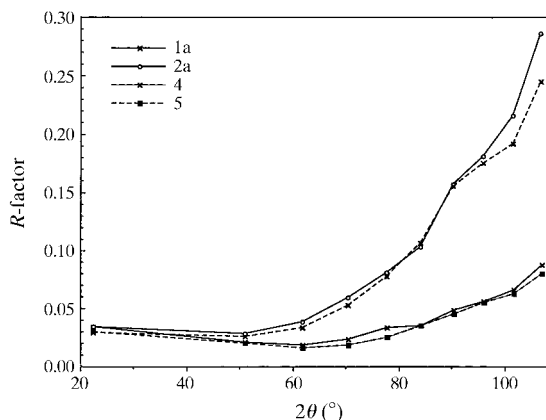


Fig. 3. Dependence of R factor on 2θ angle for various data sets.

Table 2. Selected bond lengths (Å), bond angles (°) and torsion angles (°)

	1	1a	2a	4	5
O1—C2	1.2427 (6)	1.2433 (5)	1.2432 (10)	1.2442 (8)	1.2436 (5)
O2—C5	1.2584 (7)	1.2589 (5)	1.2585 (11)	1.2584 (9)	1.2588 (5)
O3—C5	1.2591 (6)	1.2587 (5)	1.2592 (10)	1.2588 (9)	1.2585 (5)
O4—C4	1.4277 (8)	1.4279 (7)	1.4294 (13)	1.4301 (10)	1.4285 (6)
N1—C1	1.4788 (7)	1.4784 (6)	1.4776 (11)	1.4796 (10)	1.4795 (6)
N1—H1	0.889 (19)	0.861 (15)	0.89 (2)	0.878 (18)	0.851 (13)
N1—H2	0.97 (2)	0.942 (13)	0.93 (2)	0.927 (16)	0.950 (11)
N1—H3	0.837 (17)	0.892 (13)	0.89 (2)	0.889 (14)	0.897 (10)
N2—C2	1.3310 (6)	1.3304 (5)	1.3305 (10)	1.3298 (8)	1.3304 (5)
N2—C3	1.4532 (6)	1.4530 (5)	1.4541 (10)	1.4543 (9)	1.4525 (5)
C1—C2	1.5189 (7)	1.5187 (6)	1.5185 (11)	1.5182 (9)	1.5183 (5)
C3—C4	1.5342 (8)	1.5342 (6)	1.5345 (12)	1.5330 (10)	1.5344 (6)
C3—C5	1.5413 (7)	1.5413 (6)	1.5419 (11)	1.5415 (9)	1.5413 (6)
N1—C1—C2	111.46 (4)	111.50 (3)	111.47 (7)	111.45 (6)	111.47 (3)
C1—C2—N2	113.79 (4)	113.84 (3)	113.87 (7)	113.85 (6)	113.87 (3)
C2—N2—C3	122.86 (4)	122.91 (3)	122.92 (7)	122.97 (6)	122.95 (3)
N2—C3—C5	109.31 (4)	109.35 (3)	109.26 (6)	109.34 (5)	109.37 (3)
C3—C5—O2	118.15 (4)	118.13 (3)	118.15 (7)	118.14 (6)	118.11 (3)
C3—C5—O3	115.07 (4)	115.12 (3)	115.09 (7)	115.09 (6)	115.14 (3)
N1—C1—C2—N2 (ψ_1)	165.09 (4)	165.10 (3)	165.21 (7)	165.21 (6)	165.13 (3)
C1—C2—N2—C3 (ω_1)	171.99 (4)	172.01 (3)	172.12 (7)	172.08 (6)	172.04 (3)
C2—N2—C3—C5 (ϕ_1)	-151.73 (4)	-151.69 (3)	-151.74 (7)	-151.77 (6)	-151.73 (3)
N2—C2—C5—O2 (ψ_T)	-21.72 (6)	-21.73 (5)	-21.73 (10)	-21.79 (8)	-21.76 (4)
N2—C3—C4—O4 (χ_2^1)	61.06 (5)	61.13 (4)	61.19 (8)	61.25 (7)	61.16 (3)

$R(F) = 0.0277$ for the full range of data, $R(F) = 0.0224$ for 4332 unique reflection with $2\theta < 80^\circ$ and $R(F) = 0.0182$ for 3697 reflections in the 2θ interval ($40^\circ, 80^\circ$), which excludes reflections affected by the non-spherical distribution of the outer valence shell electrons causing the small R -factor increase at low angles seen in Fig. 3. Nevertheless, these values are only marginally better than those obtained for data set 1a, which were collected much more rapidly (42.3 h versus 6.7 h).

3.4. Molecular geometry and thermal parameters

A comparison of molecular geometries from various refinements given in Table 2 reveals no significant differences between the various refinements. This means that calculated geometry does not depend on either crystal size (e.g. sets 1a and 2a versus 4 and 5) or absorption correction being carried out or not (e.g. sets 1a versus 1). The latter observation is expected as long as the cross section of the crystal is reasonably centrosymmetric, in which case the atomic positions are not severely affected by lack of or inappropriate absorption correction (Stout & Jensen, 1989). The s.u.'s for geometric parameters are very low for all data sets, and in particular for sets 1a and 5 (note that 1a is better than 1 in this respect also).

Using the set of principal mean-square atomic displacements from refinement of data set 5 in Table 3 as the benchmark with which other sets are compared, it is immediately clear that the just slightly and uniformly larger values obtained from set 1a are just not very good,

but better than those obtained from set 4, that is the data set collected with the same exposure time on the small crystal. Calculated values from 1, however, are systematically larger than from sets 1a, 4 and 5, and also incorporate larger individual variations, as seen for instance for C4 and C5. It is interesting that set 4a (not tabulated) gives average principal mean-square atomic displacements 0.0194, 0.0123 and 0.0095 Å², meaning that thermal ellipsoids end up being uniformly larger using data set 4a in the refinement rather than set 4, but smaller for set 1a compared with set 1. Calculated values obtained from set 2a (not tabulated) are also quite satisfactory and very similar to those obtained from data set 4 with average values 0.0190, 0.0122 and 0.0093 Å².

3.5. Final data set comparison

All data sets are of very high quality judged by low R_{int} values, low $R(F)$ and $wR(F^2)$ values for the refined structures as well as unusually low s.u.'s for calculated thermal and geometrical parameters. There are no detectable liabilities associated with data sets 1a and 2a, collected from a very large crystal. The results from set 1a are clearly superior to those obtained from the small, spherical crystal with the same exposure time (10 s, 4), and are almost indistinguishable from those obtained for data collection on the small crystal even after a tenfold increase in exposure time (5). It thus appears that *SADABS* (Sheldrick, 1996), for this particular compound, does a very good job in modifying the original data obtained from the large crystal.

Table 3. Principal mean square atomic displacements U (\AA^2)Standard deviations are 00001–00002 \AA^2 .

	1			1a			4			5		
O1	0.0163	0.0131	0.0089	0.0156	0.0125	0.0084	0.0159	0.0128	0.0090	0.0153	0.0122	0.0081
O2	0.0255	0.0128	0.0104	0.0246	0.0125	0.0098	0.0254	0.0129	0.0103	0.0245	0.0121	0.0096
O3	0.0340	0.0157	0.0083	0.0332	0.0146	0.0083	0.0335	0.0155	0.0086	0.0327	0.0143	0.0079
O4	0.0239	0.0138	0.0108	0.0230	0.0130	0.0109	0.0240	0.0133	0.0111	0.0226	0.0128	0.0104
N1	0.0181	0.0125	0.0082	0.0174	0.0116	0.0081	0.0177	0.0122	0.0085	0.0172	0.0115	0.0079
N2	0.0163	0.0111	0.0077	0.0156	0.0102	0.0075	0.0160	0.0105	0.0080	0.0152	0.0100	0.0072
C1	0.0200	0.0127	0.0098	0.0189	0.0120	0.0097	0.0194	0.0124	0.0101	0.0185	0.0116	0.0096
C2	0.0114	0.0103	0.0078	0.0104	0.0094	0.0078	0.0112	0.0096	0.0083	0.0103	0.0093	0.0075
C3	0.0147	0.0108	0.0080	0.0141	0.0094	0.0080	0.0146	0.0098	0.0083	0.0139	0.0092	0.0077
C4	0.0191	0.0127	0.0118	0.0178	0.0126	0.0112	0.0179	0.0130	0.0118	0.0173	0.0121	0.0110
C5	0.0150	0.0118	0.0080	0.0143	0.0104	0.0082	0.0149	0.0111	0.0087	0.0142	0.0101	0.0078
Average	0.0195	0.0125	0.0091	0.0186	0.0117	0.0089	0.0191	0.0122	0.0094	0.0183	0.0114	0.0087

Data set 3/3a was collected in order to estimate the speed with which a data collection can be carried out with a large crystal. The total time for data acquisition was 24 min, with refinement of set 3a converging at $R(F) = 0.0350$ [$F^2 > 2\sigma(F^2)$] and $wR(F^2) = 0.0878$ (all). The average s.u. for bond lengths is 0.0011 \AA . The data are of good quality, with no significant differences in molecular geometry or thermal parameters compared with the results from the other refinements.

3.6. Hydrogen bonds and a comparison with the structure of L-Ser-Gly

A comparison of the crystal packing of Gly-L-Ser with that of L-Ser-Gly (Jones *et al.*, 1978), Fig. 4, shows that although the space group is $P2_12_12_1$ for both, the molecular packing arrangements are completely different, as are the hydrogen-bond interactions (Tables 4 and 5). L-Ser-Gly has a short (4.55 \AA) crystallographic axis with head-to-tail hydrogen-bonded chains along the b axis. The shortest axis for Gly-L-Ser is 9.6 \AA , and there are no obvious molecular layers as observed for L-Ser-Gly. The most prominent feature of the hydrogen-bond

network is a ribbon parallel to the b axis, which connects molecules related by a twofold screw axis (Fig. 5).

In an attempt to rationalize the observations of various hydrogen-bond patterns, it can be useful to assign ranks to the hydrogen-bond acceptors and donors, starting with the best as number 1, and then use the extended 'hydrogen-bond rule' (Etter, 1990) described in the introduction to include all donors and acceptors (donor 2 associates with acceptor 2 *etc.*).

The three best donors for the Gly-L-Ser/L-Ser-Gly structures, with ranks 1, 2 and 3, are the three amino N—H atoms, while the carboxylate group obtains acceptor ranks from 1 to 4. As far as the two additional donors ($>\text{N}-\text{H}$ and $-\text{CH}_2-\text{OH}$) are concerned, it is not obvious which should be assigned rank 4 and rank 5. This is also true for assigning ranks to the two acceptors $>\text{C}=\text{O}$ and $-\text{CH}_2-\text{OH}$, although judging by statistical values for the donor $\cdots\text{O}$ distances with carbonyl and water acceptors (Görbitz, 1989) one can tentatively assign rank 5 to the hydroxyl group and 6 to $>\text{C}=\text{O}$. With this set of ranks we find that hydrogen bonding in the L-Ser-Gly structure strictly follows the extended hydrogen-bond rule. Gly-L-Ser, on the other hand, does

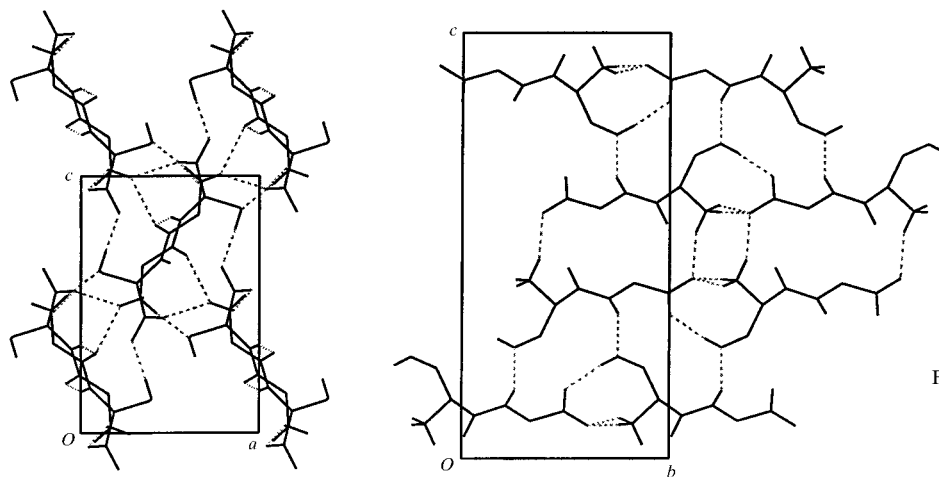


Fig. 4. Unit cell and molecular packing for Gly-L-Ser (left) and L-Ser-Gly (right; Jones *et al.*, 1978). Hydrogen bonds are shown as dotted lines in grey.

Table 4. Hydrogen-bond distances (Å) and angles (°) from refinement of data set 5

D—H...O	D—H	H...O	D...O	D—H...O
N1—H1...O1 ⁱ	0.851 (13)	2.268 (12)	2.8712 (7)	128.0 (11)
N1—H1...O2 ⁱⁱ	0.851 (13)	2.267 (13)	2.9991 (8)	144.3 (11)
N1—H2...O4 ⁱⁱⁱ	0.950 (11)	1.812 (11)	2.7610 (7)	176.3 (10)
N1—H3...O2 ^{iv}	0.897 (10)	2.045 (10)	2.8899 (8)	152.5 (9)
N2—H4...O1 ^v	0.843 (11)	2.137 (11)	2.9532 (9)	163.0 (9)
O4—H5...O3 ⁱⁱⁱ	0.884 (12)	1.716 (12)	2.5980 (8)	176.1 (12)

Symmetry codes: (i) $\frac{1}{2} + x, \frac{1}{2} - y, 2 - z$; (ii) $\frac{3}{2} - x, 1 - y, \frac{1}{2} + z$; (iii) $\frac{1}{2} - x, 1 - y, \frac{1}{2} + z$; (iv) $1 - x, y - \frac{1}{2}, \frac{3}{2} - z$; (v) $1 - x, \frac{1}{2} + y, \frac{3}{2} - z$.

not, most particularly because there are only two $-\text{NH}_3^+ \cdots \text{OOC}-$ interactions. In fact, a survey of dipeptide structures retrieved from the Cambridge Structural Database (CSD, April 1999 release; Allen & Kennard, 1993) revealed that only six out of 47 L-Xaa-L-Xaa dipeptides (Xaa can be Gly, but not L-Pro) have three interactions of this type.

Focusing on the relationship between calculated densities (D_x) in crystal structures and the nature of the side chains, a survey of dipeptide structures from the CSD (Allen & Kennard, 1993) reveals that they can be divided into the three following groups. (A) Hydrophilic structures with abundant hydrogen bonding. D_x typically from 1.40 to 1.60 g cm⁻³. (B) Dipeptides with one hydrophobic residue. D_x in the 1.25–1.40 g cm⁻³ range. (C) Strictly hydrophobic dipeptides. D_x between 1.05 and 1.20 g cm⁻³ (counting Gly as a hydrophilic residue). It is noteworthy that all observed examples of isomorphous dipeptide structures belong to group B or C; each group A structure studied possesses a unique packing arrangement and hydrogen-bonding pattern. Like Gly-L-Ser, several structures such as Gly-L-Asp.2H₂O (Eggleston & Hodgson, 1982), L-Arg-L-Glu.2H₂O (Pandit *et al.*, 1983), L-Arg-L-Asp.H₂O (Eggleston & Hodgson, 1985), L-Arg-L-Glu.H₂O (Eggleston & Hodgson, 1985) and Gly-D-Thr (Ho *et al.*, 1979) lack the short crystallographic axis which is typical for dipeptide

Table 5. Distribution of hydrogen-bond types in the L-Ser-Gly†/Gly-L-Ser structures

Acceptor	$-\text{NH}_3^-$	Donor >N—H	—OH
—COO—	3/2‡	0/0	1/1
>C=O	0/1‡	0/1	0/0
—OH	0/1	1/0	0/0

† Jones *et al.* (1978). ‡ One three-centre hydrogen bond.

structures (Görbitz & Etter, 1992). The coincident pairing of hydrogen-bond donors and hydrogen-bond acceptors for retroanalogues discussed above may be more difficult to realise within this group of compact structures than in the other two groups of dipeptides.

4. Conclusions

The hydrogen-bond pattern of Gly-L-Ser is different from that of the retroanalogue L-Ser-Gly and theoretical patterns in which donors and acceptors associate in strict rank order. This is attributed to steric constraints owing to the very compact molecular packing ($D_x > 1.5 \text{ Mg m}^{-1}$).

Several data sets were collected for the title compound, from a very large crystal and from the same crystal after reduction to a small sphere. There are no significant differences with respect to molecular geometry and thermal parameters in the refined structures. Apparently, empirical modification of the experimental data due to absorption, crystal size and beam non-uniformity effects works very satisfactorily. Since only Gly-L-Ser, which yields well diffracting and weakly absorbing crystals, and no other compounds have been studied, future experiments will establish if these results are representative for the bulk of organic and inorganic crystals.

It should be emphasized that one intention of this work is to raise the question of the best crystal size and

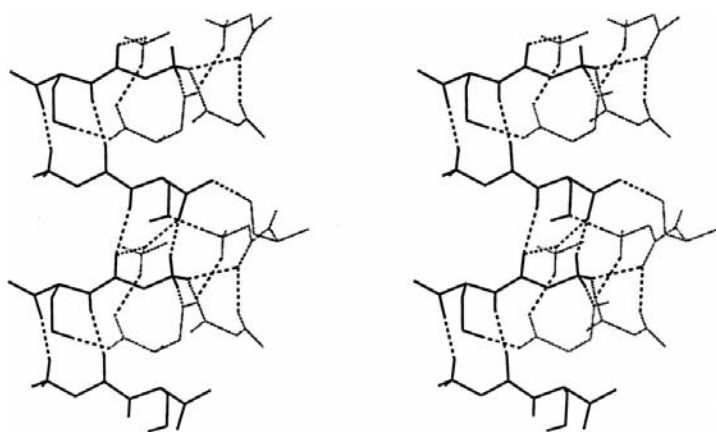


Fig. 5. Stereodrawing of a hydrogen-bonded ribbon in the Gly-L-Ser structure. Fragments of some molecules interacting with the ribbon are shown in grey. The *b* axis is approximately vertical.

not answering it. Furthermore, apart from the assessment of the scientific quality of data sets collected from a large crystal, there is clearly also a philosophical problem involved. Is it really acceptable to deliberately introduce systematic errors in the experimental data, on the assumption that they can be corrected for by a computer program at a later stage? Does the procedure impair the 'purity' of the experiment? Many crystallographers undoubtedly feel that one should always strive to collect the best data possible *without* having to make corrections. On the other hand, if accumulating evidence suggests that the convoluted effects of a large crystal and a non-uniform beam can be handled satisfactorily by programs like *SADABS* (Sheldrick, 1996), and that, at the end of the day, the same or even better structural results can be obtained in a shorter amount of time, then should the researcher abstain from using this option for philosophical reasons?

The purchase of the Siemens SMART CCD diffractometer was made possible through support from the Research Council of Norway (NFR).

References

- Allen, F. H. & Kennard, O. (1993). *Chem. Des. Autom. News*, **8**, 31–37.
- Eggleston, D. S. (1984). *Acta Cryst.* **C40**, 1250–1252.
- Eggleston, D. S. & Hodgson, D. J. (1982). *Int. J. Pept. Protein Res.* **21**, 206–211.
- Eggleston, D. S. & Hodgson, D. J. (1985). *Int. J. Pept. Protein Res.* **25**, 242–253.
- Etter, M. C. (1990). *Acc. Chem. Res.* **23**, 120–126.
- Görbitz, C. H. (1989). *Acta Cryst.* **B45**, 390–395.
- Görbitz, C. H. & Backe, P. H. (1996). *Acta Cryst.* **B52**, 999–1006.
- Görbitz, C. H. & Etter, M. C. (1992). *Int. J. Pept. Protein Res.* **39**, 93–110.
- Ho, P., Lai, T. F. & Marsh, R. E. (1979). *J. Cryst. Mol. Struct.* **9**, 181–188.
- Jones, P. G., Falvello, L. & Kennard, O. (1978). *Acta Cryst.* **B34**, 2379–2381.
- Pandit, J., Seshadri, T. P. & Viswamitra, M. A. (1983). *Acta Cryst.* **C39**, 1669–1672.
- Schwarzenbach, D., Abrahams, S. C., Flack, H. D., Prince, E. & Wilson, A. J. C. (1995). *Acta Cryst.* **A51**, 565–569.
- Sheldrick, G. M. (1996). *SADABS. Program for Absorption Correction*. University of Göttingen, Germany.
- Sheldrick, G. M. (1997). *SHELXTL*. Version 5.10. Bruker Analytical X-ray Systems, Madison, Wisconsin, USA.
- Siemens (1995). *SMART and SAINT Area-Detector Control and Integration Software*. Version 5.03. Siemens Analytical X-ray Instruments Inc., Madison, Wisconsin, USA.
- Stout, G. H. & Jensen, L. H. (1989). *X-ray Structure Determination*. New York: John Wiley and Sons.