Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

De novo structure determination of vancomycin aglycon using the anomalous scattering of chlorine

Patrick J. Loll

Department of Pharmacology, University of Pennsylvania, 3620 Hamilton Walk, Philadelphia, PA 19104-6084, USA

Correspondence e-mail: loll@pharm.med.upenn.edu

Received 20 February 2001
Accepted 2 May 2001

PDB Reference: vancomycin aglycon, 1ghg.

The crystal structure of vancomycin aglycon has been determined by exploiting the anomalous scattering of Cl atoms present within the molecule. Real-space–reciprocal-space cycling with *Shake-and-Bake* successfully located the chlorine positions from the Bijvoet differences, even though the anomalous difference Patterson map proved to be uninterpretable. The chlorine anomalous differences lacked sufficient phasing power to produce interpretable electron-density maps. However, when combined with high-resolution native data, the chlorine positions were sufficient to determine the structure using either *Shake-and-Bake* or a tangent-formula expansion.

1. Introduction

Once suitable crystals are obtained, estimating phases becomes the rate-limiting step in macromolecular crystal structure determination. Deriving phases is particularly vexing for so-called 'large small molecules', which are neither small enough for the application of classical direct methods nor amenable to phasing tools such as multiple isomorphous replacement. Novel direct-phasing methods have recently been developed which rely on alternating reciprocal-space phase refinement and real-space density modification (Hauptman, 1997*a*; Uson & Sheldrick, 1999). Such 'dual-space' methods have drastically increased the size of molecules for which *ab initio* phasing is possible (Deacon *et al.*, 1998; Frazao *et al.*, 1999; Uson *et al.*, 1999); however, they require atomic or near-atomic resolution data (Dauter *et al.*, 1997).

Anomalous scattering phasing methods have also become extremely important in recent years (Ealick, 2000; Helliwell, 2000). For the most part, such methods rely upon elements with relatively robust anomalous signals (*i.e.* $\delta f''$ values in excess of 5 electrons). However, at least in a few cases, phases have been successfully determined using the weak anomalous signal from sulfur (Dauter *et al.*, 1999; Hendrickson & Teeter, 1981).

Vancomycin is the most important member of the glycopeptide antibiotic family (Loll & Axelsen, 2000). The aglycon form of this antibiotic was crystallized in the author's laboratory (Kaplan *et al.*, 2001). Although weak diffraction could be observed from these crystals to near-atomic resolution, beyond 1.3 Å, intensities and data completeness fell precipitously. Attempts to determine this structure using the direct-methods program *Shake-and-Bake* (Hauptman, 1997*b*) failed, most likely because of the weakness of the high-resolution data. As is the case with many 'large small molecule' natural products, the crystals could not readily be

Table 1
Data-collection statistics.

	Data set No. 1 (Bijvoet difference data)	Data set No. 2 (native data)	Data set No. 3 (high-resolution native data)
Wavelength (Å) (energy, keV)	1.907 (6.5)	1.072 (11.6)	0.978 (12.7)
Maximum resolution (Å)	1.90	1.50	1.01
No. of observations	223534	93735	138541
No. of independent reflections	2591	5616	19044
Overall data set			
Completeness† (%)			
All data	90.6	96.7	92.8
Data for which $I > 2\sigma$	90.6	96.6	89.6
Mean $I/\sigma(I)$	44.7	48.2	29.8
$R_{\rm merge}$‡	0.126	0.051	0.052
Highest resolution shell			
Resolution range (Å)	1.98–1.90	1.55–1.50	1.05–1.01
Completeness† (%)			
All data	88.6	95.7	76.5
Data for which $I > 2\sigma$	88.3	95.7	51.7
Mean $I/\sigma(I)$	18.1	37.6	7.6
$R_{\rm merge}$‡	0.143	0.063	0.155

† For data set No. 1, the completeness figures refer to centric reflections plus acentric reflections for which both $F(+)$ and $F(-)$ have been measured. ‡ $R_{\rm merge} = \sum_h \sum_i |\langle I(\mathbf{h})\rangle - I(\mathbf{h})_i| / \sum_h \sum_i I(\mathbf{h})_i$.

derivatized with heavy atoms and vancomycin contains no S atoms. It does, however, contain two Cl atoms per monomer (Fig. 1). The anomalous scattering of these atoms was exploited, in combination with dual-space phasing methods, to determine the structure *de novo*. This appears to be the first reported macromolecular structure determination based principally upon chlorine anomalous scattering.

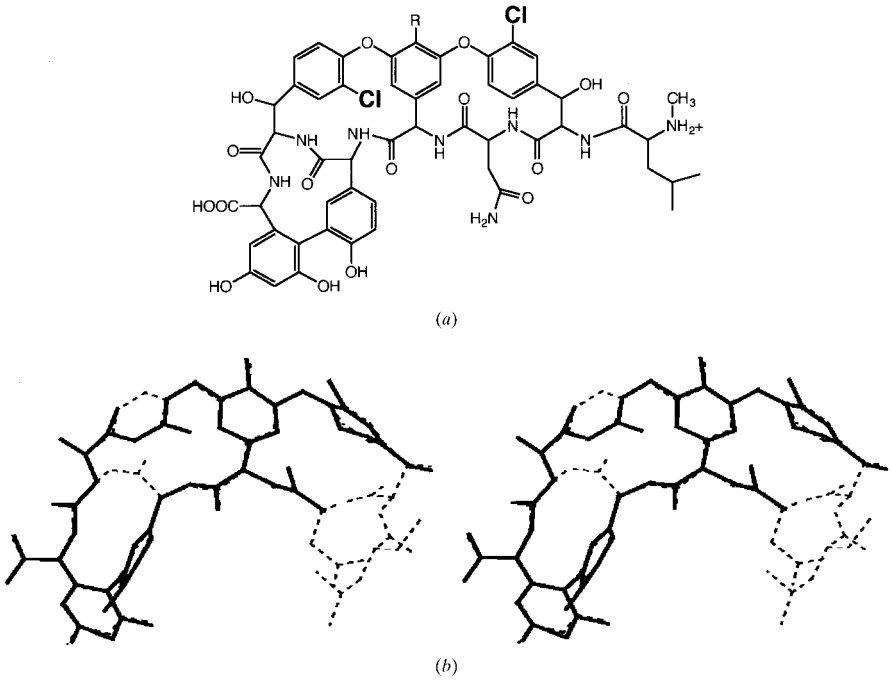

Figure 1
(*a*) Chemical structure of the vancomycin aglycon. (*b*) Divergent stereoview of the atomic positions predicted by *SnB*. One of the four vancomycin molecules in the asymmetric unit is shown. The thick lines show the positions predicted by *SnB*; the thin dashed lines show the final refined atomic model.

2. Methods

2.1. Crystallization and data collection

Crystals were prepared as described in Kaplan *et al.* (2001); they belong to space group $P4_12_12$, with unit-cell parameters $a = 29.48$, $c = 74.27$ Å. High-resolution native diffraction data were collected from a single flash-cooled crystal at beamline X12-B of the National Synchrotron Light Source (X-ray energy 12.7 keV). Highly redundant anomalous data were measured from a different flash-cooled crystal at beamline X8-C, NSLS (6.5 keV), collecting a total of 740 2° oscillations. After completion of the anomalous data set, the energy was changed to 11.6 keV and an additional 120 images were measured from the same crystal.

Images were processed using *HKL*2000 (Otwinowski & Minor, 1997), taking care to obtain accurate error estimates. The high-energy data sets yielded low merging R values; however, the 6.5 keV data gave rather higher R values (Table 1), which are likely to be a consequence of both the extremely high redundancy and of absorption effects. The high merging R values do not simply reflect crystal pathologies such as twinning, radiation damage *etc.*, because the same crystal was used for the 6.5 and 11.6 keV data sets.

2.2. Phasing

The anomalous scattering substructure was determined using *Shake-and-Bake* (*SnB*; Howell *et al.*, 2000; Smith *et al.*, 1998). Crystal density calculations suggested the asymmetric unit contained two vancomycin dimers; accordingly, eight Cl atoms were sought in each asymmetric unit. 240 reflections between 6 and 2.1 Å resolution were used to generate 2400 triplet invariants. 6000 trial solutions were examined, with 40 cycles of iterative phase and Fourier refinement per trial. Phase refinement was carried out using the 90°(2,3) strategy and the top 12 peaks were selected from the Fourier maps (Weeks & Miller, 1999). A bimodal distribution of minimal function values was observed, with roughly 7.5% of the trials producing apparent correct solutions. To check the consistency of the solutions produced, this procedure was repeated 14 times with different random number seeds. The top eight sites from all 14 jobs were found to be identical. Local scaling of the ΔF data proved essential (Blessing, 1997); no correct solutions were found without this procedure.

All attempts to generate phases using *MLPHARE* (Otwinowski, 1991) or *OASIS* (Hao *et al.*, 2000), either alone or in conjunction with *DM* (Cowtan, 1994), failed to yield interpretable maps. Therefore, the high-resolution native data set was phased from the eight

chlorine positions using *SnB*. 3200 reflections between 20 and 1.0 Å were used to generate 32 000 triplet invariants. Phases calculated from the chlorine positions were subjected to 160 cycles of iterative phase and Fourier refinement using the 90°(2,3) phase-shift strategy and picking 128 peaks from the Fourier maps. An additional 32 cycles of 'twice-baking' were added, choosing 352 peaks and phasing all reflections with $|E| > 0.75$. This procedure converged to a solution with a minimal function value of 0.48 and a crystallographic R value of 0.28. From this solution, 171 of the expected 320 non-H atom positions were easily and unambiguously identified (in fact more were correctly predicted, but this number was judged sufficient to proceed with the structure solution). These 171 atoms were used to calculate a tangent-formula expansion with *SHELXS* (Sheldrick, 1998), thereby identifying the bulk of the missing atoms. The remainder, as well as solvent molecules, were identified in difference maps during successive rounds of refinement and map inspection. The refinement converged to R and free R values for all data between 20 and 1 Å of 0.15 and 0.151, respectively (Kaplan *et al.*, 2001).

3. Results and discussion

3.1. Anomalous differences

The chlorine K edge lies at 4.39 Å and is not easily accessible. However, the anomalous scattering contribution is not trivial even far from the edge. In an effort to maximize this contribution, data were collected at a wavelength of 1.907 Å, where the chlorine $\delta f''$ value is estimated to be 1.0 electrons, compared with 0.7 electrons at the Cu $K\alpha$ wavelength (Cromer, 1983). The expected average Bijvoet difference ratio was calculated to be 0.034, using the following formula (Hendrickson *et al.*, 1985),

$$\frac{\langle |\Delta F|_{\rm anom} \rangle}{\langle F \rangle} = \left(\frac{2N_A}{N_P} \right)^{1/2} \frac{\delta f''}{Z_{\rm eff}},$$

where N_A and N_P are the number of anomalous scatterers and total number of non-H atoms, respectively, and $Z_{\rm eff}$ is the effective average atomic number (here, 6.7 electrons). As seen in Fig. 2, the observed ratio $\simeq 0.05$ up to 2 Å. Above 2 Å, the ratio rises and the mean $|\Delta F|/\sigma(\Delta F)$ ratio drops precipitously, suggesting unacceptable noise levels at high resolution. Some of this noise is certainly a consequence of absorption effects, which are exacerbated by the low-energy X-rays used. Indeed, it is possible that the increase in anomalous signal gained by moving to 1.9 Å is offset by systematic errors arising from absorption. The poor agreement statistics associated with this data set support this possibility.

3.2. Approaches to phasing

The dual-space strategy (as implemented here by *SnB*) proved vital at two stages of the structure determination. Firstly, it allowed the determination of the Cl-atom positions from the anomalous scattering data. This was not possible by manual or automated analysis of anomalous difference Patterson maps. Even in retrospect, knowing the final refined chlorine positions, these maps proved uninterpretable. In contrast, the positions predicted by *SnB* for the eight anomalous scatterers lay within 0.17 Å r.m.s. of their final refined positions. Interestingly, of these eight anomalous scatterers, only seven turned out to be Cl atoms; the eighth proved to be an S atom in an ordered solvent molecule. The eighth Cl atom did not appear among the 12 highest peaks predicted by *SnB* and has a significantly higher refined B value than the other seven Cl atoms.

Secondly, once the positions of the Cl atoms were known, the dual-space approach was able to use the high-resolution native data to calculate accurate phases, allowing the elucidation of the entire molecular structure. The only other successful phasing method proved to be a tangent expansion from the known chlorine positions (Karle, 1968), again using the full high-resolution native data set. Thus, both accurate ΔF measurements and atomic resolution native data were

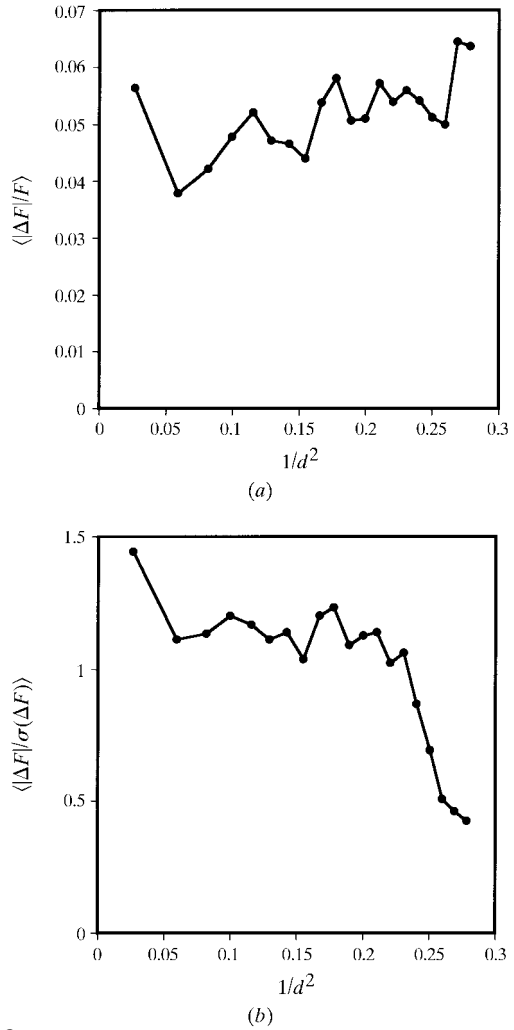

Figure 2
Dependence of the anomalous signal upon resolution. (*a*) Mean diffraction ratio, shown as the ratio of the Bijvoet difference to the structure-factor amplitude. The predicted value for the aglycon is 0.034 (see text). (*b*) Mean ratio of the Bijvoet difference to its estimated standard deviation.

required for the structure solution; neither the anomalous data nor the native data by themselves sufficed.

Interestingly, phasing methods utilizing simply the chlorine anomalous differences failed to yield acceptable phases. The average difference between *OASIS* experimental phases and the final refined phases is 52° and the experimental map is impossible to interpret, even in the context of the final refined model. Phasing with *MLPHARE* gave phase errors and maps that were, in general, somewhat worse. The correlation coefficient between the *OASIS* map and an $F_{obs}\varphi_{calc}$ map is 0.57. In contrast, phases obtained from the *SnB* phasing protocol differed on average from the refined phases by 20° and the resulting map had a correlation coefficient of 0.80 with the $F_{obs}\varphi_{calc}$ map.

Model Bijvoet difference data were calculated with *SOLVE* (Terwilliger & Berendzen, 1999), assuming no experimental error, a $\delta f''$ value of 1.0 electrons and using the final refined chlorine positions. When these data were subjected to the *OASIS* phasing protocol, only a modest improvement was observed, with the mean phase error decreasing to 47° and the map correlation coefficient increasing to 0.60. These results suggest that the phase information inherent in the chlorine anomalous signal is weak and will at best produce marginal phases.

Marginal phases are often improved drastically through the use of density modification. Unfortunately, in this case the crystals contain only approximately 30% solvent and efforts to improve the phases by solvent flattening failed. However, given that the phase errors for the experimental data are tantalizingly close to useful levels, it is not unreasonable to expect that if comparable experimental phases could be obtained with higher solvent content crystals, single-wavelength chlorine anomalous scattering plus density modification would suffice to produce interpretable maps.

Diffraction data for this study were collected at Brookhaven National Laboratory in the Biology Department single-crystal diffraction facility at beamlines X12-B and X8-C in the National Synchrotron Light Source. This facility is supported by the United States Department of Energy Offices of Health and Environmental Research and of Basic Energy Sciences, by the NSF and by the NIH. The author gratefully acknowledges expert assistance from Malcolm Capel and Leon Flaks.

References

Blessing, R. H. (1997). *J. Appl. Cryst.* **30**, 176–177.
Cowtan, K. (1994). *Jnt CCP4/ESF–EACBM Newsl. Protein Crystallogr.* **31**, 34–38.
Cromer, D. T. (1983). *J. Appl. Cryst.* **16**, 437–438.
Dauter, Z., Dauter, M., de La Fortelle, E., Bricogne, G. & Sheldrick, G. M. (1999). *J. Mol. Biol.* **289**, 83–92.
Dauter, Z., Lamzin, V. S. & Wilson, K. S. (1997). *Curr. Opin. Struct. Biol.* **7**, 681–688.
Deacon, A. M., Weeks, C. M., Miller, R. & Ealick, S. E. (1998). *Proc. Natl Acad. Sci. USA*, **95**, 9284–9289.
Ealick, S. E. (2000). *Curr. Opin. Chem. Biol.* **4**, 495–499.
Frazao, C., Sieker, L., Sheldrick, G. M., Lamzin, V. S., LeGall, J. & Carrondo, M. A. (1999). *J. Biol. Inorg. Chem.* **4**, 162–165.
Hao, Q., Gu, Y. X., Zheng, C. D. & Fan, H. F. (2000). *J. Appl. Cryst.* **33**, 980–981.
Hauptman, H. A. (1997*a*). *Curr. Opin. Struct. Biol.* **7**, 672–680.
Hauptman, H. A. (1997*b*). *Methods Enzymol.* **277**, 3–13.
Helliwell, M. (2000). *J. Synchrotron Rad.* **7**, 139–147.
Hendrickson, W. A., Smith, J. L. & Sheriff, S. (1985). *Methods Enzymol.* **115**, 41–55.
Hendrickson, W. A. & Teeter, M. M. (1981). *Nature (London)*, **290**, 107–113.
Howell, P. L., Blessing, R. H., Smith, G. D. & Weeks, C. M. (2000). *Acta Cryst.* D**56**, 604–617.
Kaplan, J., Korty, B. D., Axelsen, P. H. & Loll, P. J. (2001). *J. Med. Chem.* **44**, 1837–1840.
Karle, J. (1968). *Acta Cryst.* B**24**, 182–186.
Loll, P. J. & Axelsen, P. H. (2000). *Annu. Rev. Biophys. Biomol. Struct.* **29**, 265–289.
Otwinowski, Z. (1991). *Proceedings of the CCP4 Study Weekend*, edited by W. Wolf, P. R. Evans & A. G. W. Leslie, pp. 80–86. Warrington: Daresbury Laboratory.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
Sheldrick, G. M. (1998). *Direct Methods for Solving Macromolecular Structures*, edited by S. Fortier, pp. 401–411. Dordrecht: Kluwer Academic Publishers.
Smith, G. D., Nagar, B., Rini, J. M., Hauptman, H. A. & Blessing, R. H. (1998). *Acta Cryst.* D**54**, 799–804.
Terwilliger, T. C. & Berendzen, J. (1999). *Acta Cryst.* D**55**, 849–861.
Uson, I. & Sheldrick, G. M. (1999). *Curr. Opin. Struct. Biol.* **9**, 643–648.
Uson, I., Sheldrick, G. M., de La Fortelle, E., Bricogne, G., Di Marco, S., Priestle, J. P., Grutter, M. G. & Mittle, P. R. E. (1999). *Structure*, **7**, 55–63.
Weeks, C. M. & Miller, R. (1999). *Acta Cryst.* D**55**, 492–500.