

Sylvie Salamitou,^a Stéphane Réty,^{b,‡} Françoise Le Hégarat,^a Gérard Leblon^a and Anita Bentley^{b,*‡}

^a Institut de Génétique et Microbiologie, Bâtiment 409, Université Paris XI, 91405 Orsay, France, and ^bLURE, BP 34, 91898 Orsay CEDEX, France

[‡] Present address: LBPA, Ecole Normale Supérieure de Cachan, 61 Avenue du Président Wilson, 94235 Cachan, France.

Correspondence e-mail:
anita.bentley@lbpa.ens-cachan.fr

Received 10 November 2004

Accepted 22 December 2004

The use of high halide-ion concentrations and automated phasing procedures for the structural analysis of BclA, the major component of the exosporium of *Bacillus anthracis* spores

The structure determination of the recombinant form of BclA, the major protein component of *Bacillus anthracis* exosporium, involved soaking in a high concentration of potassium iodide as the means of obtaining a good-quality heavy-atom derivative. The data to 2 Å resolution collected on a laboratory source were of sufficient quality to allow successful phasing and chain tracing by automated methods.

1. Introduction

The lethal disease anthrax is most commonly spread by spores of *Bacillus anthracis*, a dormant cell form of certain bacilli characterized by high resistance to chemicals and to environmental factors such as heat and drought. The spore is encased in a thick multilayered structure, the coat, which is surrounded by the exosporium. The latter is composed of a paracrystalline basal layer with a hexagonal lattice structure and a hair-like outer region. Isolated *B. anthracis* exosporia contain at least 12 major protein components, the most abundant of which is BclA, a glycosylated protein with an internal collagen-like region (Sylvestre *et al.*, 2002).

BclA has only recently been identified and most studies have focused on immunological, genetic and electron-microscopy studies, with little work having been carried out on the biochemistry of this protein. In particular, no data are available on its structure, although it can be inferred from the presence of the collagen-like region that it will be unique. From recent fluorescent confocal microscopy studies it is known that upon contact with the epithelial layers of the lung, *B. anthracis* spores are rapidly phagocytosed by circulating macrophages (Guidi-Rontani, 2002). Since BclA is the most abundant component of the filaments of the exosporium of *B. anthracis* (Sylvestre *et al.*, 2003) and is the major antigen of the surface of the spore, it may play a central role in the invasion of lung alveolar macrophages. Thus, knowledge of its structure should have implications for understanding the function of the outermost layer of the spores, which in turn will advance our understanding of the first stages of infection by this lethal pathogen.

2. Materials and methods

2.1. Expression and purification

The *bclA* gene was PCR-amplified from *B. anthracis* strain CEB 9732 genomic DNA using the following primers: O1, 5'-GGAATT-CCATATGGCATTGACCCTAAT-3' carrying an *NdeI* site (underlined) and O2, 5'-TTAACCGGCGGCCGCTT**AGTGA**-TGGTGATGGTGATGAGCAACTTTTCAATAAT-3' carrying a *NotI* site (underlined), the stop codon (bold) and six histidine codons (italics). The first 18 codons of *bclA*, corresponding to the signal peptide of the protein, were omitted. The amplified fragment was cloned *NdeI/NotI* in pET9sn1, a derivative of pET9a (Novagen), in which a *NotI* site was inserted between the *Bpu1102* and the *NdeI* sites. The construct was checked by oligonucleotide sequencing (Genome Express, France).

The plasmid encoding the His-tagged BclA protein was freshly transformed into the Rosetta strain of *Escherichia coli* (Novabiogen). Protein expression was induced by adding 1 mM IPTG to exponentially growing cultures in LB medium plus 50 µg ml⁻¹ kanamycin (2 l)

at 310 K. The induced cultures were then grown overnight at 288 K in LB medium plus 50 $\mu\text{g ml}^{-1}$ kanamycin. The cells were washed with 50 mM NaH_2PO_4 pH 8, 0.3 M NaCl. The pellet was taken up in 20 ml of the same buffer plus 10 mM imidazole and the cells were broken by sonication. The crude extract was loaded onto a Ni-NTA column (5 ml, Hi-Trap Chelating HP, Amersham Bioscience) and purification at 277 K was carried out using Äkta Fast Protein Liquid Chromatography apparatus. Elution was performed with a linear gradient of imidazole (10–250 mM). The fractions containing BclA were then pooled and purified by gel-filtration chromatography on a Superdex 75 column (Amersham Biosciences) using 50 mM Tris-HCl pH 8, 1 M NaCl. The final purification step was carried out by Mono-Q ion-exchange chromatography in 50 mM Tris-HCl pH 8.5 using a linear gradient of NaCl (0–1 M) to elute the protein.

2.2. Collagenase activity

Lyophilized type VII collagenase (Sigma) was taken up in 50 mM Tris-HCl pH 7.5 to 1 mg ml^{-1} concentration. Activity was measured at 298 K in 50 mM Tris-HCl pH 7.5 containing 1 mM calcium for up to 1 h. Solubilized proteins were analysed by SDS-PAGE under reducing conditions (5% β -mercaptoethanol) in 12% polyacrylamide slab gels according to the method of Laemmli (1970). Molecular-weight (rainbow) markers (BioRad, Hercules, CA, USA) were run in parallel.

2.3. Crystallization

Crystallization conditions were screened with Crystal Screen 1 from Hampton Research using the hanging-drop vapour-diffusion technique. Hexagonal shaped crystals appeared using condition No. 41 [20% (w/v) PEG 4000, 20% (v/v) 2-propanol, 0.1 M Na HEPES pH 7.5]. The optimized conditions, 10 mg ml^{-1} protein in 50 mM Tris-HCl pH 8, 300 mM NaCl mixed with an equal volume of well solution containing 20% (w/v) PEG 4000, 20% (v/v) 2-propanol, 0.1 M sodium cacodylate pH 6.5, gave crystals of up to 200–300 μm across and 100–200 μm thick within 2–3 d.

2.4. Cryosolvent optimization

Initially, the crystals were transferred to a cryoprotectant solution corresponding to the crystallization solution containing 5% (v/v) glycerol (solution Cryo-1; Garman & Mitchell, 1996) for several seconds and then plunged into liquid nitrogen. The diffraction pattern extended to 2.4 Å resolution, but showed signs of crystal damage. A test at room temperature showed that the crystals could diffract to better than 1.8 Å resolution with low mosaic spread (0.3°). We therefore tried adding different concentrations of glycerol, PEG 400, 2-propanol and MPD to the crystallizing solution (McFerrin & Snell, 2002). The best results were obtained with a solution containing 0.1 M sodium cacodylate pH 6.5, 22.5% (w/v) PEG 4000 and 30% (v/v) 2-propanol (solution Cryo-2).

2.5. Data collection and processing

Data were collected on a Rigaku MicroMax-007 generator equipped with a MAR345 detector mounted on a MAR Research desktop beamline (dtb). Measurements were made both at room temperature on crystals mounted in capillaries and at 100 K. All data were processed using DENZO and SCALEPACK from the HKL data-processing system (v.1.97.1; Otwinowski & Minor, 1997).

Table 1

Data statistics for the crystals used.

Values in parentheses are for the highest resolution shell.

Crystal	Native, RT	Native, cryo	KI	KCl
Maximum resolution (Å)	2.19 (2.27–2.19)	2.47 (2.56–2.47)	2.0 (2.07–2.0)	2.0 (2.12–2.0)
Unit-cell parameters (Å)				
<i>a</i>	69.49	67.80	67.51	67.86
<i>c</i>	163.60	163.115	162.81	163.29
Mosaicity† (°)	0.288	0.561	1.194	0.443
$R_{\text{sym}}^{\ddagger}$	0.089 (0.268)	0.121 (0.339)	0.07 (0.162)	0.088 (0.314)
$\langle I \rangle / \langle \sigma(I) \rangle$	17.6 (5.8)	11.3 (3.6)	20.8 (7.3)	30.9 (9.1)
Completeness	0.965 (0.915)	0.924 (0.728)	0.576 (0.661)	0.996 (0.985)
Redundancy	7.5 (6.5)	5.5 (4.5)	7.5 (7.5)	15 (13)

† As determined by SCALEPACK. $\ddagger R_{\text{sym}} = \sum_h \sum_i |I(h_i) - \langle I(h) \rangle| / \sum_h \sum_i I(h)$.

2.6. Heavy-atom-derivative search: soaks in high concentrations of halide salts

The heavy-atom-derivative search was not aided by the peculiar amino-acid composition of the protein: there are neither cysteines nor methionines in the sequence and no histidines except for the C-terminal 6 \times His tag. After an unsuccessful search with standard heavy-atom compounds, we decided to soak the crystals in a high concentration of halide ions (Dauter *et al.*, 2000), choosing potassium iodide because of its higher anomalous signal at 1.54 Å (Evans & Brucogne, 2002). The crystals were soaked for 3–5 min in 1 M KI, 20% (w/v) PEG 4000, 20% (v/v) 2-propanol, 0.1 M sodium cacodylate pH 6.5, then rapidly passed through cryoprotectant solution Cryo-1 and flash-cooled in liquid nitrogen.

The crystals soaked in a high concentration of KI diffracted to higher resolution (better than 1.7 Å). We therefore also collected a complete data set on a crystal soaked first in 1 M KCl then in the optimized cryoprotectant solution (Cryo-2) and flash-cooled, and used that as the high-resolution native data set for the structure refinement.

2.7. Heavy-atom location and phasing

The SCALEPACK program (Otwinowski & Minor, 1997) was used to evaluate the anomalous signal from heavy-atom-derivative data sets. The data retained, together with a native data set, were introduced without prior scaling into the *autoSHARP* procedure (de La Fortelle & Brucogne, 1997). The results presented below were obtained using different runs of *autoSHARP* and different combinations of starting data sets. The complete cycle, including automatic model-building with *ARP/wARP* (Morris *et al.*, 2002), was requested systematically. The program *O* (Jones *et al.*, 1991) was used to build side chains into the *ARP/wARP* result and the program *REFMAC* from the *CCP4* suite (Collaborative Computational Project, Number 4, 1994) was used for refinement. In order to perform phasing comparisons, we also used *SHELXD* (Schneider & Sheldrick, 2002) for the location of anomalous scatterers and *SHELXE* (Sheldrick, 2002) for phasing. The search for atoms on special positions was included in *SHELXD* with a SIRAS protocol.

3. Results and discussion

The crystal data show sixfold symmetry and the space group was determined to be $P6_322$, with one molecule in the asymmetric unit and a solvent fraction of 0.54 and 0.515 for room-temperature and frozen crystals, respectively. Table 1 gives the statistics for all data sets used.

Consistent with the evaluation of the anomalous signal of the different derivatives tested by *SCALEPACK*, *autoSHARP* runs confirmed that only KI derivative data were useful for heavy-atom phasing. With a request for one site per heavy-atom data set, the automatic procedure determined its location, calculated and refined its phasing contribution, defined the correct hand of its coordinates and, upon calculating a solvent-flattened electron-density map, traced

a polypeptide chain corresponding to two-thirds of the expected sequence. At the end of the procedure, we obtained electron density of a quality sufficient to place all side-chains unambiguously (Fig. 1*a*).

What were the limitations to this success? We performed a number of test runs with different data sets and a different number of heavy-atom sites requested in order to evaluate the approach. The phasing statistics for these runs are summarized in Table 2. The phase difference with the refined model after solvent flattening was 34.91 and 23.96° for the initial 2.5 Å SIRAS(1) run and the 2.0 Å SIRAS(4) run, respectively. For a comparison of automatic phasing methods we decided to use the *SHELX* package as well. Both *SHARP* and *SHELXD* located the heavy atoms in the same positions, with a slightly different order for atoms with weak occupancies (Table 3). The phases obtained with *SHELXE* from atom positions from *SHELXD* gave an interpretable electron-density map with a map correlation coefficient of 70.6%. Density modification with either *DM* or *SOLOMON* did not improve this map and the phase difference calculated with the refined model was 50.2°. In this case, therefore, *SHARP* gave a better phase set as input to automatic chain tracing.

3.1. Halide-ion sites

The KI derivative is characterized by one strong and four weak sites (Table 3). The phasing quality is very good using just the first site; the supplementary sites only improve the statistics to a small extent. The anomalous signal from the iodide is sufficient when the SAD procedure is used (at 1.54 Å, f'' for iodine is 7 e). The isomorphous and anomalous signals are also very good in the case of KI *versus* KCl data. Since the high-resolution data used as 'native' were measured on a crystal soaked in a similar solution as the derivative, but with chloride instead of iodide, and the two data sets are highly isomorphous, we looked at solvent molecules in the refined structure. The main iodide site is unique to the heavy-atom derivative, lying in a shallow hydrophobic pocket of the protein (Fig. 2). This indicates that even a simple iodide ion has an affinity for hydrophobic sites and need not necessarily be combined with iodine to form a triiodide ion as suggested by Evans & Briconne (2002).

The room-temperature (RT) data were collected on a crystal that was not soaked in a halide solution and these data can be used to confirm the specificity of the iodide sites. One of the minor iodide sites lies close to the C-terminal His-tag tail, which is rather disordered. While the first of the six His residues is clearly visible in all three crystal forms, the next one is most ordered in the frozen KCl crystals and totally disordered in

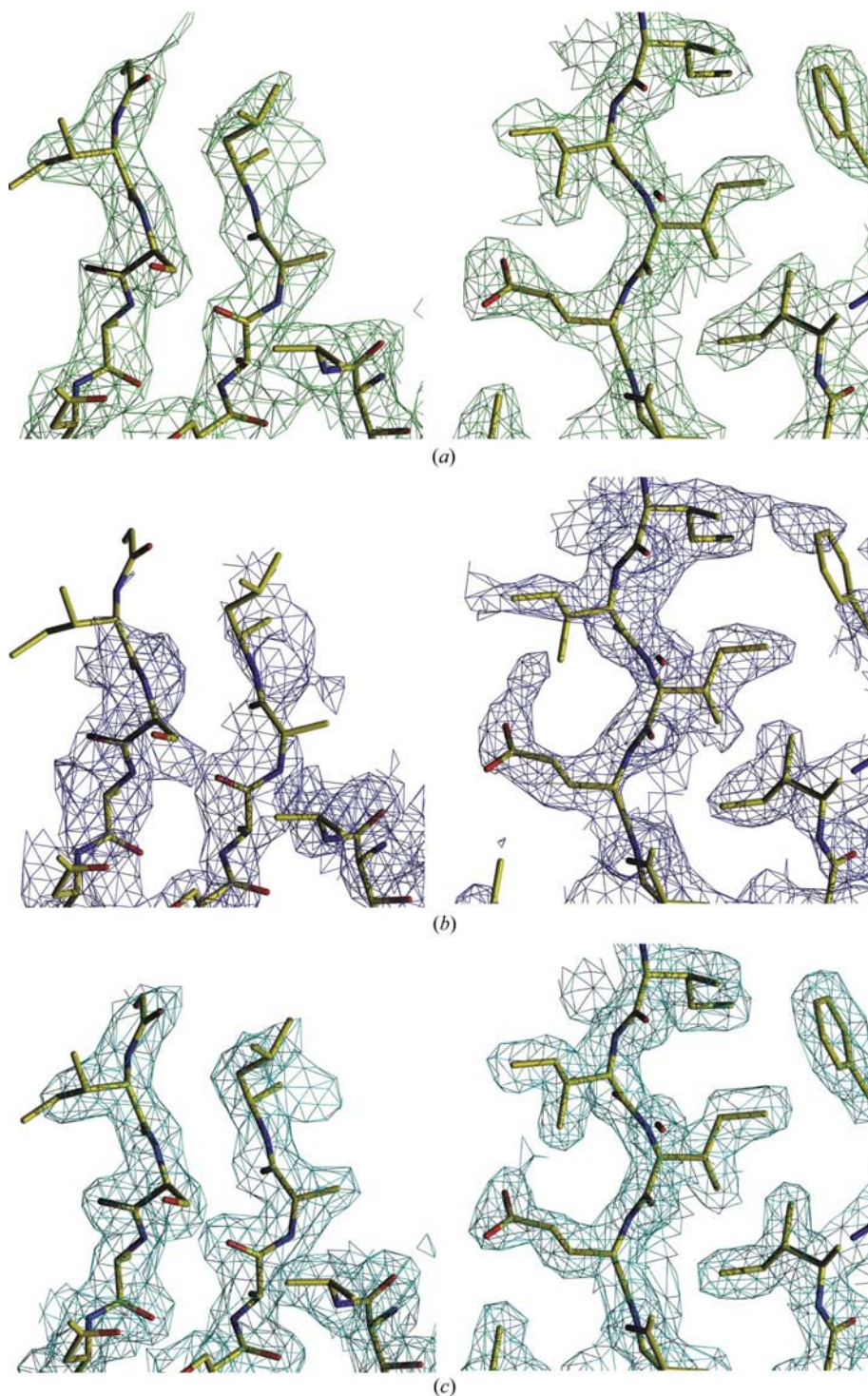


Figure 1
Examples of electron densities ($2F_o - F_c$ maps contoured at 1σ) in the worst (left) and best (right) regions of the map. (a) The result of the automatic phasing and tracing procedure; (b) SAD phasing; (c) the electron density after refinement. The refined structure is superimposed in all examples.

Table 2Phasing statistics for different *autoSHARP* runs.

PP, phasing power = F_H/E , where F_H is the r.m.s. mean heavy-atom contribution and E is the r.m.s. residual. FOM, figure of merit. CC, electron density correlation coefficient after *SHARP* phasing (SH) and after solvent flattening (Sol).

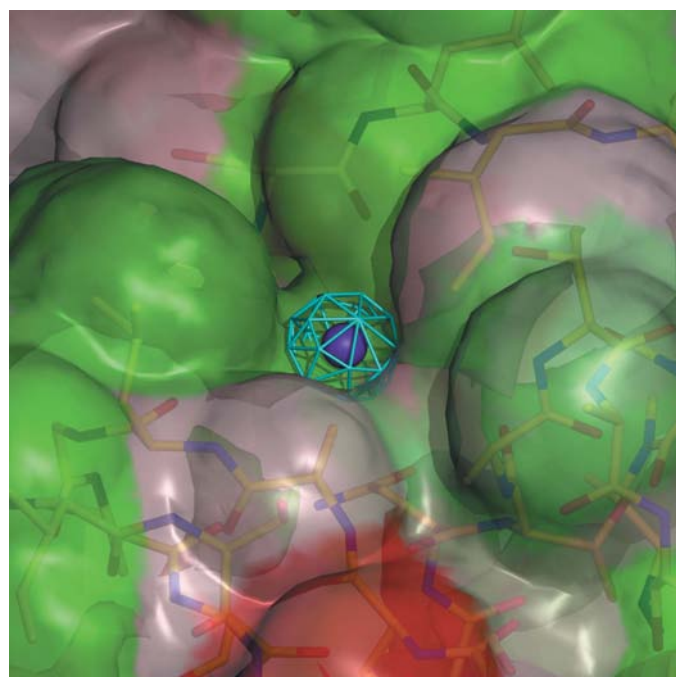
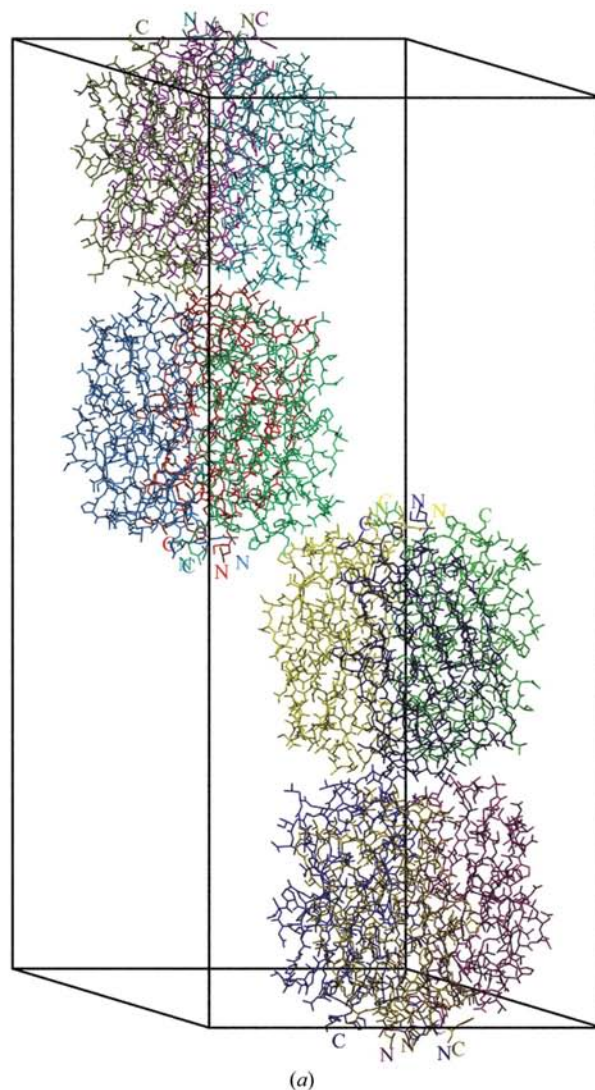
Data	Native/KI	Native/KI	KCl/KI	KCl/KI	KI
Resolution (Å)	2.5	2.5	2.0	2.0	2.0
No sites	1	5	1	5	5
Run	SIRAS(1)	SIRAS(2)	SIRAS(3)	SIRAS(4)	SAD (2.0 Å)
PP _{iso(centric)}	1.292	1.587	1.278	1.670	n.a.
PP _{iso(acentric)}	1.030	1.200	1.035	1.286	n.a.
PP _{ano}	1.116	1.349	0.909	1.121	1.160
FOM _{acentric}	0.2618	0.2981	0.2765	0.3317	0.3394
FOM _{centric}	0.2018	0.2631	0.2254	0.3000	0.1399
CC (SH)	0.2318	0.2701	0.2962	0.3582	0.2903
CC (Sol)	0.7055	0.6938	0.7362	0.7769	0.6825

the room-temperature crystal. It is only partially ordered in the frozen KI crystals, where its side-chain density is close to a minor iodide position.

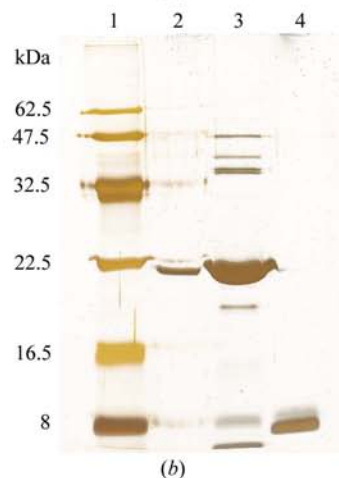
3.2. Data quality: effects of halides and cryoprotectant solutions

The automatic phasing procedure was initiated immediately, before all parameters had been optimized, in order to evaluate heavy-atom derivatives. The data quality was initially limited by our success in correctly freezing the crystals, which in turn limited the resolution and completeness of the data sets obtained (Table 1). We found that soaking the crystal in a high concentration of salt (KCl or KI) improved the diffraction limit of the crystals, while the composition of the cryoprotectant solution was crucial for reducing their mosaicity. The initial native data set was reasonably complete, although limited to 2.5 Å resolution. The KI-derivative crystal was measured before the optimal cryoprotectant solution had been defined and the data set, although extending to 2 Å resolution, was rather incomplete because of high mosaicity. This no doubt explains the poorer phasing

statistics for the SAD approach, yet the resulting solvent-flattened electron-density map was good enough for manual interpretation. A

**Figure 2**

The major iodide site. The iodide ion is represented as a purple ball surrounded by a 5σ contour from the anomalous difference map. The structure is visible beneath the surface, the properties of which are depicted from acidic (red) to hydrophobic (green).

**Figure 3**

(a) Packing of the molecules in the $P6_22$ crystal unit cell. Each monomer is drawn in a different colour and the C- and N-termini are labelled. (b) Silver-stained SDS-PAGE results for BclA analysis. Lane 1, molecular-weight markers; lane 2, purified BclA as control; lane 3, BclA from washed crystals; lane 4, BclA after collagenase treatment.

Table 3
Iodide sites located by *SHARP* and *SHELXD* (fractional coordinates).

Atom No.	x	y	z	Occupancy	B
<i>SHARP</i>					
1	-0.438	-0.230	-0.218	0.92	18.18
2	0.628	0.358	0.251	0.20	20.87
3	0.446	0.362	0.192	0.31	25.24
4	0.358	0.470	0.013	0.39	35.08
5	0.281	0.463	0.005	0.70	53.79
<i>SHELXD</i>					
1	-0.439	-0.231	-0.218	1.00	20.00
2	0.357	0.464	0.011	0.33	20.00
3	0.628	0.357	0.250	0.15	20.00
4	0.449	0.365	0.193	0.31	20.00
5	0.277	0.466	0.004	0.24	20.00

typical example of the best and worst portion of this electron-density map is shown in Fig. 1(b) and the same sections of the final refined electron density are shown in Fig. 1(c).

3.3. Crystal symmetry and missing density

Three molecules related by the crystallographic sixfold screw axis form a tight trimer with N- and C-termini coming together around the symmetry axis (Fig. 3a). The electron density accounts for only two-thirds of the sequence, with none visible for the collagen-like N-terminal 80 residues. The integrity of the protein was therefore ascertained using mass spectrometry on protein recovered from washed and dissolved crystals. A molecular weight of 21 270.8 Da was determined, which is compatible with the theoretical value of 21 258 Da after cleavage of the initial methionine (error 0.1–0.05%). The existence of a trimer in solution was confirmed by gel-filtration chromatography performed on a Superdex 75 column in 50 mM Tris–HCl, 100 mM NaCl buffer. The protein eluted as a single symmetrical peak with an apparent molecular weight of 70 kDa, which is consistent with a trimer.

The triple-helical collagen structure is very resistant to digestion by all proteases except collagenase. We therefore treated BclA and control proteins (BSA and lysozyme) with various amounts of *Clostridium histolyticum* collagenase, a protease specific for structured collagen, where it cleaves between the X and the Gly residues in the X-Gly-Pro-Y sequence. The digested BclA and the control proteins were then assessed by SDS–PAGE analysis. Whereas the untreated BclA migrated as a 29 kDa protein, BclA subjected to collagenase was present as a 14 kDa species. This molecular weight is in agreement with the digestion of the collagen-like amino-terminal domain of the protein, leaving the C-terminal domain (141 amino acids) intact. BSA and lysozyme were not degraded at all by the collagenase treatment (Fig. 3b). A time course of collagenase activity was then performed. The 14 kDa band was resistant to incubation with 0.01 µg collagenase for up to 60 min (data not shown). These data suggest that the N-terminal domain of recombinant BclA can adopt a collagen-like structure that is readily degraded by collagenase, whereas the C-terminal region is resistant.

We were initially hoping to observe a collagen-type triple helix at the N-terminus of the visible globular C-terminal trimer assembly. The symmetry of a collagen-type triple helix is, however, incompatible with the observed hexagonal crystal symmetry. We therefore examined the data closely to verify whether the true space group was in fact of a lower symmetry and the molecules of the trimer were related by non-crystallographic symmetry only. Table 4 summarizes the data statistics when different possible space groups are considered. It is clear that the highest hexagonal symmetry space group is the correct one. Since the protein is intact within the crystal, we

Table 4
Check for true crystal symmetry using the 2 Å KCl data (results from *SCALA* from the *CCP4* package; Collaborative Computational Project, Number 4, 1994).

Space group	R_{sym}^{\dagger}	$R_{\text{meas}}^{\ddagger}$	No. unique
P1	0.060	0.084	65541
C2	0.056	0.076	53422
C222 ₁	0.064	0.080	34873
P6 ₃ 22	0.076	0.083	14608

$\dagger R_{\text{sym}} = \sum_h \sum_l |I(h)_i - I(h)_l| / \sum_h \sum_l I(h)$. $\ddagger R_{\text{meas}}$ redundancy-independent (multiplicity-weighted) R_{sym} .

conclude that despite the excellent overall order of these crystals, one third of the polypeptide chain is disordered.

A similar situation was encountered in the crystal structure of collagen VIII, where the globular C-terminal part is well ordered but the N-terminal 42 residues that should form a triple helix are totally disordered (Kvansakul *et al.*, 2003). To date, only one crystal structure of a collagen triple helix linked to a non-collagenous C-terminal domain exists (Stetefeld *et al.*, 2002), showing for the first time the transition from a strictly threefold-related globular part to the helical symmetry of the collagen triple helix. Notably, the crystal symmetry was C2, with non-crystallographic threefold symmetry relating monomers within the globular part. Even in this structure the triple-helix part of the structure has much higher B values than the globular region, with the two Gly-Pro-Pro triplets most distant from it becoming quite disordered.

4. Conclusions

The recent developments in automatic structure-determination software have succeeded in providing some very powerful tools for the protein crystallographer. In this report, we show that in combination with a modern powerful laboratory X-ray source and even moderately complete data to moderate resolution, such software can produce extremely satisfactory results. We further show that in the context of a laboratory X-ray source equipped with a copper target, the use of a high concentration of iodide salts can provide an excellent heavy-atom derivative. Indeed, soaking in a high-salt solution can improve the quality of the data by improving the crystal and crystal-freezing properties.

We wish to thank Gilles Vergnaud (IGM, Orsay) for the gift of DNA from *B. anthracis* strain CEB973 and Robert Aufrère and Gilles Henckès from the IGM, Orsay for the use of the pET9sn1 plasmid. We are grateful to the group of Jean-Luc Popot at the IPBC, Paris, especially Ines Gallay and Daniel Picot, for allowing us to collect data using their X-ray equipment. Finally, we thank the referees for useful suggestions.

References

Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* **D50**, 760–763.
 Dauter, Z., Dauter, M. & Rajashankar, K. R. (2000). *Acta Cryst.* **D56**, 232–237.
 Evans, G. & Bricogne, G. (2002). *Acta Cryst.* **D58**, 976–991.
 Garman, E. F. & Mitchell, E. P. (1996). *J. Appl. Cryst.* **29**, 584–587.
 Guidi-Rontani, C. (2002). *Trends Microbiol.* **10**, 405–409.
 Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. (1991). *Acta Cryst.* **A47**, 110–119.
 Kvansakul, M., Bogin, O., Hohenester, E. & Yayon, A. (2003). *Matrix Biol.* **22**, 145–152.
 Laemmli, U. K. (1970). *Nature (London)*, **227**, 680–685.
 La Fortelle, E. de & Bricogne, G. (1997). *Methods Enzymol.* **276**, 472–494.
 McFerrin, M. B. & Snell, E. H. (2002). *J. Appl. Cryst.* **35**, 538–545.

- Morris, R. J., Perrakis, A. & Lamzin, V. S. (2002). *Acta Cryst.* **D58**, 968–975.
- Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
- Schneider, T. R. & Sheldrick, G. M. (2002). *Acta Cryst.* **D58**, 1772–1779.
- Sheldrick, G. M. (2002). *Z. Kristallogr.* **217**, 644–650.
- Stetefeld, J., Frank, S., Jenny, M., Schulthess, T., Kammerer, R. A., Boudko, S., Landwehr, R., Okuyama, K. & Engel, J. (2002). *Structure*, **11**, 339–346.
- Sylvestre, P., Couture-Tosi, E. & Mock, M. (2002). *Mol. Microbiol.* **45**, 169–178.
- Sylvestre, P., Couture-Tosi, E. & Mock, M. (2003). *J. Bacteriol.* **185**, 1555–1563.