Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

In-house phase determination of the lima bean trypsin inhibitor: a low-resolution sulfur-SAD case

Judit É. Debreczeni,[a] Gábor Bunkóczi,[a] Beatrix Girmann[b] and George M. Sheldrick[a]*

[a]Lehrstuhl für Strukturchemie, Georg-August Universität, Göttingen, Germany, and [b]Institut für Organische Chemie, Georg-August Universität, Göttingen, Germany

Correspondence e-mail: gsheldr@shelx.uni-ac.gwdg.de

Received 13 September 2002
Accepted 14 November 2002

PDB Reference: lima bean trypsin inhibitor, 1h34, r1h34sf.

SAD (single-wavelength anomalous diffraction) has enormous potential for phasing proteins using only the anomalous signal of the almost ubiquitous native sulfur, but requires extremely precise data. The previously unknown structure of the lima bean trypsin inhibitor (LBTI) was solved using highly redundant data collected to 3 Å using a CCD detector with a rotating-anode generator and three-circle goniometer. The seven 'super-S' atoms (disulfide bridges) were located by dual-space recycling with *SHELXD* and the high solvent content enabled the density-modification program *SHELXE* to generate high-quality maps despite the modest resolution. Subsequently, a 2.05 Å synchrotron data set was collected and used for further phase extension and structure refinement.

1. Introduction

The lima bean trypsin inhibitor (LBTI) belongs to the bifunctional Bowman–Birk-type protease inhibitors and consists of trypsin- and chymotrypsin-binding subdomains that are arranged symmetrically. It has a molecular weight of 9.0 kDa and contains seven disulfide bridges but no other S atoms. Because of its high thermostability and inhibitory activity, it is widely used in biochemical procedures (Birk, 1976). The members of the Bowman–Birk-type protease inhibitor family have been studied intensively because of their anti-carcinogenic activity (Flecker, 1993).

Here, we report the structure of LBTI determined by in-house measurement of the anomalous signal of sulfur. Samudzi *et al.* (1997) reported crystallization conditions but were unable to solve the structure by molecular replacement; the self-rotation map showed an apparent non-crystallographic twofold axis. Since LBTI is purified from a natural source (making selenomethionine MAD less attractive) and the high-symmetry space group should ensure a high data redundancy, it appeared to be a good challenge for SAD (single-wavelength anomalous diffraction) phase determination of an unknown structure using in-house data, even though the crystals did not diffract to better than 3 Å on a rotating anode.

After pioneering work 20 y ago on crambin (Hendrickson & Teeter, 1981) and the development of suitable computational methods (Wang, 1985), there was little progress in sulfur-SAD phasing until recent developments in high-intensity sources, cryocrystallography and detector technology enabled the small anomalous intensity differences (~1–1.5%) to be measured with the required precision. Most experiments in which the native sulfur was the only anomalous scatterer, including all the determinations of previously unknown structures (except crambin) that we are aware of, used synchrotron radiation at relatively long (>1.5 Å) wavelengths (Dauter *et al.*, 1999; Liu *et al.*, 2000; Bond *et al.*, 2001; Gordon *et al.*, 2001; Weiss *et al.*, 2001; Micossi *et al.*, 2002; Brown *et al.*, 2002) and in most cases involved crystals that diffracted to better than 2 Å. Some of these also included other fairly weak anomalous scatterers such as chlorine and calcium.

2. Experimental

2.1. Purification and crystallization

Crude LBTI powder (obtained from SIGMA) was purified by HPLC using a Nucleosil-100 C18 column with a 25–34% acetonitrile/water gradient. Three isoforms were detected, but only one of them could be separated with a reasonable yield. Therefore, two types of LBTI sample were used in crystallization trials: fraction *A*, containing one isoform, and fraction *B*, containing a mixture of the remaining two isoforms. Crystallization was carried out using the hanging-drop technique, with droplets containing equal volumes of the protein solution (10 mg ml^{-1} LBTI) and reservoir solution. Fraction *A* was crystallized from 30% MPD, 0.02 *M* $CaCl_2$ and 0.1 *M* NaOAc/HOAc pH 4.6. Fraction *B* was crystallized from 0.6 *M* potassium/sodium tartrate and 0.1 *M* HEPES pH 7.5. Since crystals obtained from fraction *A* had poor diffraction properties, only crystals from fraction *B* were used in subsequent procedures. The contradiction that crystals containing isoform impurity diffract better is accounted for by the

© 2003 International Union of Crystallography
Printed in Denmark – all rights reserved

Table 1
Data-collection statistics.

Values in parentheses are for the last resolution shell.

	In-house	Synchrotron
Space group	$I2_13$	$I2_13$
Unit-cell parameters (Å)	$a = b = c = 110.17$	$a = b = c = 109.20$
Data collected	393286	88491
Unique data	4608	13776
Resolution (Å)	3.00 (3.00–3.10)	2.05 (2.15–2.05)
Completeness (%)	99.7 (96.9)	98.5 (96.1)
Redundancy	85.3 (83.3)	6.3 (3.1)
R_{int} (%)	8.91 (42.27)	5.67 (29.75)
Average $I/\sigma(I)$	76.25 (16.65)	12.43 (3.58)

Table 2
Model-refinement statistics.

Resolution range	77.19–2.05
No. of residues	57
Total No. of non-H atoms	501
No. of water molecules	78
R_{factor}/R_{free} (%)	23.02/24.97
R.m.s.d. from ideal geometry	
Bond lengths (Å)	0.009
Angle distances (Å)	0.030
Ramachandran plot (%)	
Allowed region	98
Additionally allowed region	0
Disallowed region	2

biochemical finding that only isoform II and III together undergo self-association and form dimers and also explains the difficulties in the purification (Birk, 1976).

2.2. Data collection, processing and analysis

An LBTI crystal was mounted in an arbitrary orientation and flash-cooled to 100 K in a nitrogen stream. 25% glycerol was used as cryoprotectant. A highly redundant data set was collected to 3.0 Å with Cu $K\alpha$ radiation using a Bruker rotating-anode generator equipped with Osmic multilayer mirrors, a three-circle goniometer and a Bruker SMART6000 4K CCD detector. A 360° φ-scan and two 180° ω-scans were collected with fine slicing (0.2° oscillation) and an exposure time of 1 min per frame. After the structure had been solved using these in-house data, a higher resolution (2.05 Å) data set was collected at the BW7B beamline at EMBL/DESY, Hamburg and used for the final model building and refinement. Data-collection statistics are listed in Table 1.

The in-house data were collected and processed using the Bruker programs *PROTEUM*, *SAINT*, *SADABS* and *XPREP*. Firstly, the φ- and ω-scans were scaled separately and the correlation coefficient between the signed anomalous differences for the two scan modes was used to establish the resolution limit of the anomalous signal. All the data were then scaled together to give the final data set. For the purpose of heavy-atom location with *SHELXD*, the data were truncated at 3.7 Å, *i.e.* at the resolution where the correlation coefficient fell below 30% (Schneider & Sheldrick, 2002). The collection of independent φ- and ω-scan data required for this test and also partly responsible for the relatively high redundancy requires a three-circle or κ goniometer. The synchrotron data were processed with *DENZO* and *SCALEPACK* (Otwinowski & Minor, 1997).

2.3. 'Heavy-atom' substructure solution and phasing

Seven 'super-S' atoms were found by *SHELXD* (Sheldrick *et al.*, 2001) corresponding to the seven disulfide bridges (Fig. 1*a*). This also indicated that only one molecule is present in the asymmetric unit, giving an estimated solvent content of 73%. The correlation coefficients output by *SHELXD* were 42.13 (all data) and 24.37 (weak data). The substructure solution was input to *SHELXE* (Sheldrick, 2003) without any editing or heavy-atom parameter refinement. Two *SHELXE* jobs were started for the two possible heavy-atom enantiomorphs. The contrast and connectivity figures of merit unambiguously identified the correct hand (1.099 and 0.924, respectively, for the correct heavy-atom enantiomer and 0.201 and 0.750, respectively, for the wrong heavy-atom enantiomer). 20 cycles of density modification using the sphere-of-influence method implemented in *SHELXE* resulted in a very clear traceable electron-density map even though the resolution was limited to 3.0 Å (Fig. 2*a*). The high solvent content was undoubtedly a major factor in determining the map quality (Sheldrick, 2003).

2.4. Model building and refinement

When the synchrotron data became available, the phases were extended to 2.05 Å using *SHELXE*. The complete model was built by hand using *XtalView* (McRee, 1999) and refined with *SHELXL* (Sheldrick & Schneider, 1997). The stereochemistry of the model was assessed with *PROCHECK* (Laskowski *et al.*, 1993). A summary of the refinement statistics is given in Table 2. The density after model refinement with *SHELXL* is shown in Fig. 2(*b*).

3. Description of the structure

The LBTI molecule adopts a 'bow-tie' fold canonical to the Bowman–Birk protease inhibitor family (Fig. 1*b*). It consists of two pseudo-symmetry related domains consisting of two antiparallel β-strands, each connected by short loop regions without the formation of any α-helices. The structure is strongly constrained by the seven disulfide bridges and a number of intramolecular hydrogen bonds in such a manner that all the β-sheets and even the reactive-site loops lie in one slightly bent plane. This fold is rather conservative and the only difference between the members is the angle of the bending (Chen *et al.*, 1992; Song *et al.*, 1999; Li de la Sierra *et al.*, 1999; Koepke *et al.*, 2000). Least-squares superposition of the main-chain atoms in the two domains gives an r.m.s. deviation of 0.52 Å, showing the close structural relationship between them and accounting for the apparent twofold NCS in the self-rotation function.

Most of the molecule, even the solvent-

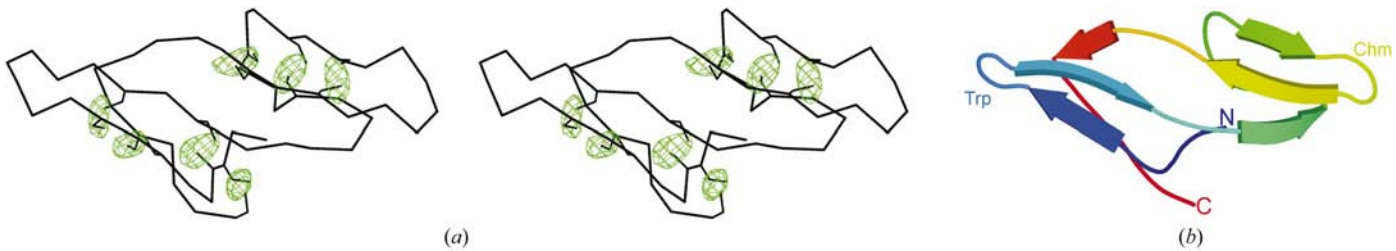

Figure 1
(*a*) Stereo backbone representation of LBTI showing the Cys residues and an anomalous Fourier map calculated with *SHELXE* phases at 3 Å resolution. Note the shape of the super-sulfur peaks caused by the disulfide bridges. (*b*) Overall fold of LBTI. Trp, trypsin-binding loop; Chm, chymotrypsin-binding loop.

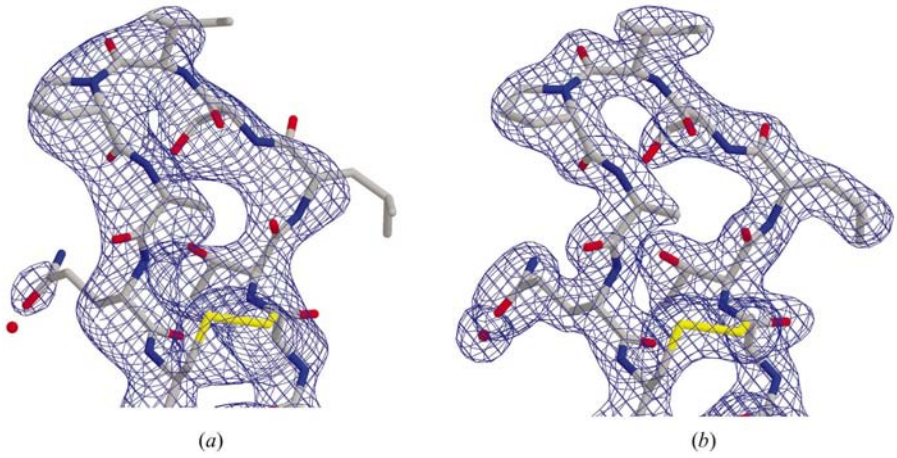

Figure 2
(*a*) Electron-density map around the chymotrypsin-binding loop produced by *SHELXE* using the in-house data at 3 Å resolution. (*b*) A $2F_o - F_c$ map of the same region after refinement against the 2 Å synchrotron data set.

exposed binding-site loops, is well defined in the electron-density map. The two terminal tails (residues 1–15 and 73–83) are not visible in the density owing to their extensive flexibility and the fact that they extend into the large solvent channel in the unit cell. There are two buried polar side chains (His43 and Asp65) in the interior of the molecule that are not involved in the intramolecular interactions. In the cavity around these residues there is an internal solvent channel filled with well defined water molecules.

Each binding-site loop of LBTI exhibits a heart-like shape determined by a *cis*-proline and fixed by a conserved hydrogen-bonding network (Suzuki *et al.*, 1993). The residues Thr25, Ser27 and Gln31 (in the trypsin-binding loop) and Thr52, Ser54 and Gln58 (in the chymotrypsin-binding loop) at the positions known as P2, P1′ and P5′, respectively, are involved in this characteristic pattern.

The molecules of LBTI form dimers in the crystal structure. Two molecules related by a crystallographic twofold axis form a compact globular-shaped dimer. It is possible that the two monomers are in fact two slightly different isoforms (see §1), but the resulting heterodimers are disordered about the twofold axis. The interface between the two monomers is formed by their β-sheets, with a rotation angle of 90° between the monomers, so that the binding-site loops are located on the surface of the dimer. The dimeric state is stabilized by a set of hydrogen bonds and a hydrophobic core. The residues His21, Arg33, Thr35, Glu58, Asp62 and four water molecules play essential roles in the hydrogen-bonding network. The hydrophilic core consists of the side chains of residues Ile50, Val60 and the methyl group of the residue Thr35. Since the peripheral hydrophobic residues are all shielded through dimer formation, the dimeric state is probably the physiological unit. However, it is not clear whether LBTI remains dimerized even when bound to trypsin and chymotrypsin and what conformational changes may occur upon binding.

4. Conclusions

S atoms can be employed to phase proteins, provided that their anomalous signal can be measured precisely, which in practice requires highly redundant data (Dauter & Adamiak, 2001). In this particular case, the high redundancy resulting from the cubic space group and three-circle goniometer as well as the high solvent content were crucial to success in the location of the super-sulfur sites and subsequent density modification at the modest resolution obtainable in-house. Subsequent tests indicated that about half the actual redundancy of 85.3 would in fact have sufficed. Nevertheless, a higher resolution synchrotron data set proved very useful for further phase improvement and refinement.

The authors are grateful to Ralph Krätzner for discussions and to the Fonds der Chemischen Industrie and the European Community (Access to Research Infrastructure Action of the Improving Human Potential Programme to the EMBL Hamburg Outstation, contract No. HPRI-1999-CT-00017) for support.

References

Birk, Y. (1976). *Methods Enzymol.* **45**, 707–709.
Bond, C. S., Shaw, M. P., Alphey, M. S. & Hunter, W. M. (2001). *Acta Cryst.* D**57**, 755–758.
Brown, J., Esnouf, R. M., Jones, M. A., Linnell, J., Harlos, K., Hassan, A. B. & Jones, E. Y. (2002). *EMBO J.* **21**, 1054–1062.
Chen, P., Rose, J., Love, R., Wei, C. H. & Wang, B.-C. (1992). *J. Biol. Chem.* **267**, 1990–1994.
Dauter, Z. & Adamiak, D. A. (2001). *Acta Cryst.* D**57**, 990–995.
Dauter, Z., Dauter, M., de La Fortelle, E., Bricogne, G. & Sheldrick, G. M. (1999). *J. Mol. Biol.* **289**, 83–92.
Flecker, P. (1993). *Protease Inhibitors as Cancer Chemopreventive Agents*, edited by W. Troll & A. R. Kennedy, pp. 161–176. New York: Plenum Press.
Gordon, E. J., Leonard, G. A., McSweeney, S. M. & Zagalsky, P. F. (2001). *Acta Cryst.* D**57**, 1230–1237.
Hendrickson, W. A. & Teeter, M. M. (1981). *Nature (London)*, **290**, 107–113.
Koepke, J., Ermler, U., Warkentin, E., Wenzl, G. & Flecker, P. (2000). *J. Mol. Biol.* **298**, 477–491.
Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). *J. Appl. Cryst.* **26**, 283–291.
Li de la Sierra, I., Quillien, L., Flecker, P., Gueguen, J. & Brunie, S. (1999). *J. Mol. Biol.* **285**, 1195–1207.
Liu, Z.-J., Vysotski, E. S., Chen, C.-J., Rose, J. P., Lee, J. & Wang, B.-C. (2000). *Protein Sci.* **9**, 2085–2093.
McRee, D. E. (1999). *J. Struct. Biol.* **125**, 156–165.
Micossi, E., Hunter, W. N. & Leonard, G. A. (2002). *Acta Cryst.* D**58**, 21–28.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
Samudzi, C., Schroeder, S., Griffith, S., Chen, X. & Quinn, T. P. (1997). *Proteins*, **27**, 311–314.
Schneider, T. R. & Sheldrick, G. M. (2002). *Acta Cryst.* D**58**, 1772–1779.
Sheldrick, G. M. (2003). In the press.
Sheldrick, G. M., Hauptman, H. A., Weeks, C. M., Miller, M. & Usón, I. (2001). *International Tables for Crystallography*, Vol. F, edited by E. Arnold & M. G. Rossmann, pp. 333–351. Dordrecht: Kluwer Academic Publishers.
Sheldrick, G. M. & Schneider, T. R. (1997). *Methods Enzymol.* **277**, 319–343.
Song, H. K., Kim, Y. S., Yang, J. K., Moon, J., Lee, J. Y. & Suh, S. W. (1999). *J. Mol. Biol.* **293**, 1133–1144.
Suzuki, A., Yamane, T., Ashida, T., Norioka, S., Haram & Ikenazka, T. (1993). *J. Mol. Biol.* **234**, 722–734.
Wang, B.-C. (1985). *Methods Enzymol.* **115**, 90–112.
Weiss, M. S., Sicker, T. & Hilgenfeld, R. (2001). *Structure*, **9**, 771–777.