Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Characterization and crystallization of a novel *Sarcocystis muris* lectin, SML-2

Jürgen J. Müller,[a]* Eva-Christina Müller,[b,c] Thomas Montag,[d] Nadja Zyto,[d] Bettina Löschner,[d] Harald Klein,[d] Udo Heineman[a,e] and Albrecht Otto[b]

[a]Kristallographie, Max-Delbrück-Centrum für Molekulare Medizin, Robert-Rössle-Strasse 10, D-13092 Berlin, Germany, [b]Proteinchemie, Max-Delbrück-Centrum für Molekulare Medizin, Robert-Rössle-Strasse 10, D-13092 Berlin, Germany, [c]Hämatologie, Onkologie und Tumorimmunologie, Charité der Humboldt-Universität Berlin, Lindenberger Weg 80, D-13125 Berlin, Germany, [d]Paul-Ehrlich-Institut, Bundesanstalt für Seren und Impfstoffe, FG Parasitologie/Diagnostika, Paul-Ehrlich-Strasse 51–59, D-63225 Langen, Germany, and [e]Institut für Chemie – Kristallographie, Freie Universität, Takustrasse 6, D-14195 Berlin, Germany

Correspondence e-mail: jjm@mdc-berlin.de

© 2001 International Union of Crystallography
Printed in Denmark – all rights reserved

Received 17 March 2001
Accepted 1 May 2001

A novel lectin (SML-2) consisting of 138 amino acids was isolated from cyst merozoites of *Sarcocystis muris* and sequenced by Edman degradation and mass spectrometry. All 12 cysteinyl residues are involved in disulfide bridges, four of which are attributed to a characteristic pattern of cysteines as found in the so-called PAN-module superfamily. Crystals of SML-2 diffracting to 2.1 Å resolution at a synchrotron were grown by the hanging-drop vapour-diffusion technique. They belong to the space group $P2_12_12_1$, with unit-cell parameters $a = 53.6$, $b = 128.8$, $c = 158.2$ Å and eight molecules in the asymmetric unit. SML-2 cocrystallized with Au galactose results in two different crystal forms. The first form is isomorphous with the native crystals and the second form adopts space group $C222_1$, with unit-cell parameters $a = 74.7$, $b = 82.0$, $c = 131.0$ Å, and diffracts to 2.4 Å at a rotating-anode X-ray generator.

1. Introduction

S. muris is an obligate intracellular cyst-forming parasite which propagates in mice as the intermediate and cats as the final host. Cyst merozoites occur in the muscle tissue of the mouse. They become motile and penetrate the intestinal cells of cats after being ingested with raw meat (Mehlhorn & Heydorn, 1978). Invasive stages of apicomplexan parasites (*e.g. Plasmodium*, *Toxoplasma*, *Eimeria*, *Sarcocystis*) are characterized by specialized secretory organelles located in the apical cell region. Several proteins which have been proposed to interact with receptors on the surface of host cells are initially located within so-called micronemes, *e.g.* the thrombospondin-related anonymous protein (TRAP/SSP-2; Rogers *et al.*, 1992) and the erythrocyte-binding proteins (EBAs) of malaria parasites (Adams *et al.*, 1992). It may therefore be hypothesized that the microneme antigens contribute in an adhesin-like manner to the specific recognition of host cells during an early phase of host-cell invasion (Entzeroth *et al.*, 1992).

A lectin fraction (*S. muris* lectin pool, SML-p) has been isolated from *S. muris* cystmerozoites. The fraction contains at least three proteins, which could be further separated by reversed-phase chromatography into a mixture of SML-1 and SML-3 and the highly purified SML-2. As determined by haemagglutination assays, the functionality of the molecules includes high affinity towards *N*-acetyl-galactosamine and galactose. Gel-filtration chromatography showed that the *S. muris* lectins exist as dimers and/or higher oligomers of non-covalently associated subunits (Montag *et al.*, 1997) which temporarily inhibit *in vitro* the mitogen-induced proliferation of splenocytes of the natural intermediate host of *S. muris* (Zyto *et al.*, 1997). The cDNA of one microneme protein (SML-1 gene product; Eschenbacher *et al.*, 1993), as well as a second clone (SML-3 gene product) showing 87% identity to the cDNA clone mentioned above, could be isolated from a subgenomic library of *S. muris* (Klein *et al.*, 1996). Because SML-2 could be purified to homogeneity from the rest of the lectin pool, it has been characterized by mass spectrometry and by Edman sequencing and X-ray diffraction-quality crystals have been grown. This work should provide the first three-dimensional information about a lectin directly isolated from an intracellular protozoan and should provide further insight into the mechanisms by which apicomplexan parasites attach to and invade host cells.

2. Isolation and purification

S. muris cystozoites were harvested from the skeletal muscles of infected mice 3–9 months post-infection using trypsin treatment after homogenization of the muscle tissue. Debris and cystozoites were separated by filtration. Trypsin was removed by centrifugation. SML-2 was purified in a two-step procedure using a Pharmacia HPLC System from *S. muris* extract, which was prepared by sonication followed by ultracentrifugation of *S. muris* cystozoites (10^9 cyst merozoites, 2 ml 50 m*M* Tris, 150 m*M* NaCl buffer pH 8.3; sonication: 2×15 min in 0.5 s pulses of 40 W; centrifugation: 850 000*g*, 1 h, 277 K).

Firstly, sugar-affinity chromatography yielded the *S. muris* lectin-pool fraction (SML-p). After affinity binding of SML molecules to the lactose groups of an α-lactose gel (EY Laboratories Inc., San Mateo, USA), the proteins were eluted from the column using galactose as a competitive sugar (buffer: 50 m*M* Tris, 150 m*M* NaCl pH 8.3, galactose gradient 0–0.3 *M*). Since the preparation had to be switched to a sugar-free buffer for further investigation, the elimination of galactose residues from SML-p solution was performed using a Sephadex G-25 superfine column, fast desalting PC 3.1/10 (Pharmacia, Sweden). Secondly, purification of SML-2 from SML-p resulted from reversed-phase chromatography using a µRPC C2/C18 PC3.2/3 column (Pharmacia, Sweden) and a polar/non-polar buffer system with 0.1% TFA in doubly distilled water and 0.1% TFA in acetonitrile. After reversed-phase separation there were two distinct lectin fractions: a protein mixture of SML-1 and SML-3 and the highly purified SML-2.

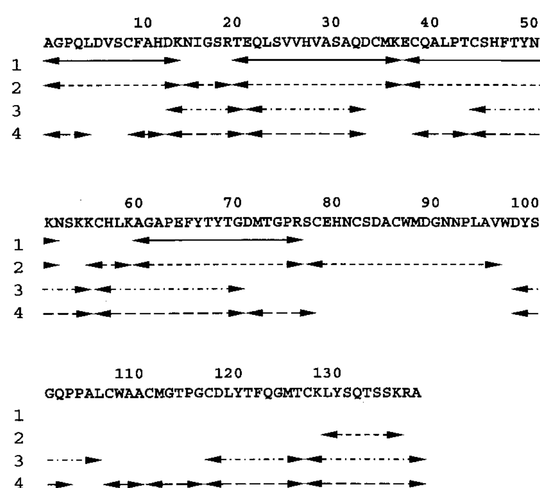

Figure 1
Primary structure of SML-2 derived from Edman and MS/MS sequencing. The peptides shown below the sequence result from tryptic digestion after reduction and treatment with iodoacetamide and MS/MS sequencing (1), tryptic digestion after reduction and pyridylethylation for Edman sequencing (2), Asp-N cleavage after cysteine oxydation with performic acid and MS/MS sequencing (3) and Asp-N cleavage after cysteine oxydation with performic acid and Edman sequencing (4). The complete sequence of SML-2 has been deposited in the SWISS-PROT database under entry MIA2_SARMU (P81860).

For Edman degradation and mass spectrometry, SML-2 was desalted, alkylated and tryptically digested. 1 µl of SML-2 from a crystallization experiment (see below) was dissolved in 50 µl 0.1% trifluoroacetic acid (TFA) and adsorbed to 20 mg of C4 reversed-phase material (Vydac, Hesperia, CA, USA). An aliquot of about 200 p*M* SML-2 was alkylated as described by Thomsen & Bayne (1988). The alkylated protein was desalted by reversed-phase chromatography using a Vydac C-4 precolumn and an increasing gradient of acetonitrile in water containing 0.1% TFA. The fraction containing the alkylated protein was taken to dryness in a Speedvac centrifuge and digested at 309 K for 15 h with 0.5 µg of sequencing-grade modified trypsin (Promega Corp., Madison, WI, USA) dissolved in 50 µl 0.1 *M* Tris–HCl buffer pH 8.2. The oxidation of an aliquot of 200 p*M* SML-2 was performed according to Hirs (1956) and Drapeau (1979). Asp-N cleavage was carried out in 50 µl 50 m*M* Tris–HCl buffer pH 8.0 at 309 K for 15 h. Tryptic peptides and peptides after cleavage with Asp-N were obtained by reversed-phase HPLC on a µRPC C2/C18 SC 2.1/10 column using the Smart system (Amersham Pharmacia Biotech, Uppsala, Sweden). Peptides of interest were loaded onto a Biobrene-coated glass filter fibre of a Procise sequencer (PE Applied Biosystems, Weiterstadt, Germany).

3. Mass spectrometry

The mass-spectrometric measurements were performed with a hybrid quadrupole time-of-flight mass spectrometer (Q-Tof, Micromass, Manchester, England) equipped with a nanoflow Z-spray ion source. The protein and peptide mixtures were dissolved in a methanol/water/formic acid mixture [50:49:1(*v*/*v*/*v*)]. Edman degradation and ESI mass spectrometry established the N-terminal sequence to be AGPQLDVSCFAHDK. According to ESI mass spectrometry, the molecular mass of SML-2, 15 066 Da, is similar to those of the *S. muris* gene products SML-1 (15 101 Da) and SML-3 (15 008 Da).

Different methods of cleavage are necessary to align the peptides of the protein in order to establish the full SML-2 sequence. Fig. 1 shows: (1) MS/MS results after reduction and treatment with iodoacetamide before the tryptic digest, (2) peptides identified with Edman degradation after reduction, pyridylethylation and tryptic digest, (3) MS/MS and (4) Edman results, both after cysteine oxidation with performic acid and Asp-N cleavage.

The MS spectra of the peptide mixture after tryptic digest and tryptic digest after reduction and treatment with iodoacetamide are shown in Fig. 2. All sequences (Table 1) as well as disulfide bridges were identified by MS/MS measurements. The disulfide bridges were found between C38 and C44 (peptide ECQALPTCSHFTYNK) and between C34 (peptide TEQLSVVHVASAQDCMK) and C56 (peptide CHLK). After reduction and treatment with iodoacetamide, two more peptides containing cysteinyl residues were identified (Table 1*b*), which could not be detected without peptide reduction and modification because of the disulfide bridges between two peptides. The Asp-N cleavage after oxidation with performic acid produced short peptides with an aspartate or cysteic acid and one with a glutamate at the N-terminal end (see Fig. 1, line 3 and line 4).

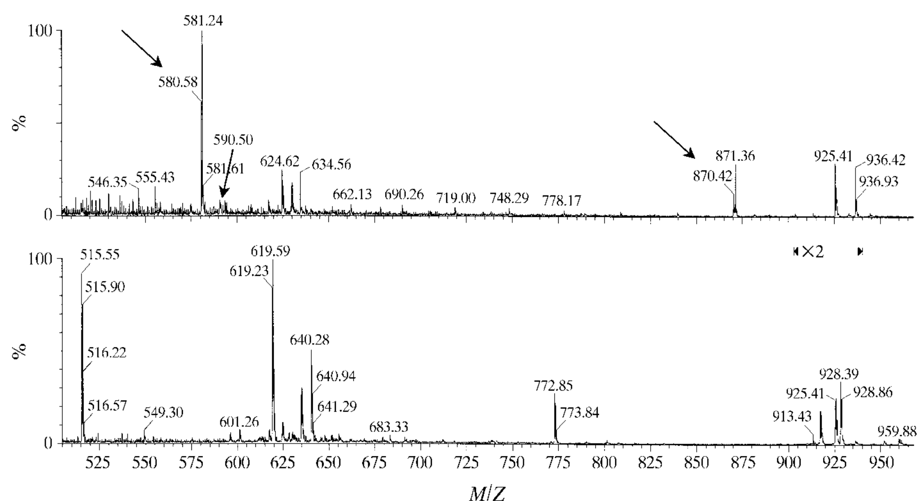

Figure 2
Nanoflow ESI spectra of SML-2 after tryptic digestion (*a*) and with tryptic digestion after reduction and carboxyamidomethylation (*b*). The peaks correspond to multiply charged ions and were subjected to MS/MS experiments (see Tables 1*a* and 1*b*). The arrows mark the peptides containing disulfide bridges (see Table 1*a*).

Table 1
Molecular masses and structural assignment of peptides of SML-2.

The masses of peptides *A*–*H* were determined by MS/MS measurements (spectra not shown); the others agree with the results of the Edman degradation. Cam, carboxyamidomethyl cysteine; Ca, cysteinic acid; Mox, oxidized methionine.

(*a*) Tryptic digest. Residues in bold are linked by disulfide bridges.

Mass/charge (Da)	Mass (Da)	Theoretical mass (Da)	Sequence
580.58, 870.42	1738.74, 1738.84	1740.77	ECQALPTCSHFTYNK
590.50	2358.00	2358.10	TEQLSVVHVASAQDCMoxK / CHLK
617.25, 925.41	1848.75, 1848.82	1848.81	AGAPEFYTYTGDMoxTGPR
624.62	1870.86	1870.81	AGAPEFYTYTGDMoxTGPR + Na

(*b*) Tryptic digest after reduction and treatment with iodoacetamide.

Mass/charge (Da)	Mass (Da)	Theoretical mass (Da)	Sequence
515.55, 772.85	1543.65, 1543.70	1543.72	AGPQLDVSCamFAHDK
619.26	1854.72	1854.81	ECamQALPTCamSHFTYNK
640.28	1917.84	1917.90	TEQLSVVHVASAQDCamMoxK
917.39	1832.78	1832.80	AGAPEFYTYTGDMTGPR
925.37	1848.74	1848.80	AGAPEFYTYTGDMoxTGPR
928.39	1854.78	1854.81	AGAPEFYTYTGDMTGPR + Na

(*c*) Asp-N cleavage after cysteine oxidation with performic acid.

Peak	Mass/charge	Mass (Da)	Theoretical mass (Da)	Sequence
A, *C*	297.48, 445.73	889.44, 889.46	889.46	DKNIGSRT
B, *E*	376.92, 502.23	1503.68, 1503.69	1503.68	CaSHFTYNKNSKK
	398.16	794.32	794.33	DMoxTGPRS
D, *G*	473.88, 710.32	1418.64, 1418.64	1418.68	CaKLYSQTSSKRA
	485.25	484.25	484.26	AGPQL
	525.20	524.20	524.17	CaFAH
F	634.31	1266.62	1266.66	EQLSVVHVASAQ
H	853.36	1704.72	1704.74	CaHLKAGAPEFYTYTG

The fragmentation of the peptides after the Asp-N cleavage was suitable for sequencing. HPLC fractions of two peptides were used to confirm or to support the Edman results. In case of the peptide D118LYTFQGMT126, the MS/MS found a modification of methionine to methionine sulfone caused by oxidation with performic acid. The peptide D98YSGQPPAL106 was completely identified by MS/MS, but only partially with Edman degradation (D98YSGQP103). If there were no overlapping regions, the position of the peptide in the sequence was deduced from alignment with SML-1 and SML-3 (Fig. 3).

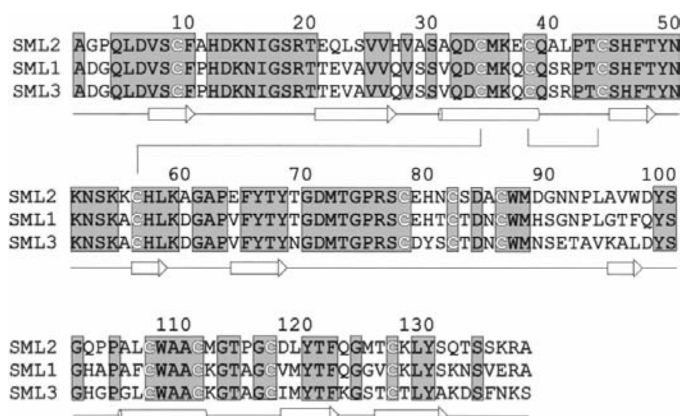

Figure 3
Alignment of lectin sequences SML-1, SML-2 and SML-3 of *S. muris*. Boxes enclose conserved residues. Disulfide bridges, determined by mass spectrometry, are linked by heavy lines. Arrows indicate β-strands and rods α-helices according to the protein secondary-structure prediction program *Jpred*^{2} at EBI (Cuff *et al.*, 1998).

SML-2 consists of 138 amino acids with a calculated molar mass of 15 078 Da and a total of 12 negatively and 11 positively charged residues. 12 cysteinyl residues are present in the molecule and, as determined from the experimentally measured mass of 15 066 $\pm$ 1 Da, six intramolecular disulfide bridges contribute to maintaining the functional structure. The sequence alignment revealed a high degree of similarity of the three gene products (Fig. 3). Most closely related to each other are SML-1 and SML-3 (79.7%), whereas SML-2 shares 72.5% of its residues with SML-1 and 65.2% with SML-3. The SMLs preferentially bind galactose and N-acetylgalactosamine, but they show only a low sequence identity of about 15% with the family of animal galectins (for a review, see Rini, 1995) or any other lectin and do not contain the eukaryotic galactoside-binding lectin signature (PROSITE motif PS00309). SML-2, like the two other *S. muris* lectins, twice contains the sequence motif -C-*X*(3)-C-*X*(5)-C-*X*(2)-[FYW]-*X*-[FYW]-*X*(4,6)-C-, which belongs to the PAN-module superfamily (Tordai *et al.*, 1999) and which is present also in plasma kallikrein of mouse, rat, pig and human, in human coagulation factor XI (Klein *et al.*, 1996), in the plasminogen/hepatocyte growth-factor (HGF) family and domains of various nematode proteins. The SMLs also share this sequence motif with several apicomplexan proteins and molecules involved in cell adhesion; it also occurs in several copies in microneme proteins of *Toxoplasma gondii* (Mic4) and *Eimeria tenella* (Mic5; Brown *et al.*, 2000). Sequence-based secondary-structure prediction using the server *Jpred*^{2} at EBI (Cuff *et al.*, 1998) results in a mixed-class protein with a minimum of 11% helical and 27% β-sheet structure as shown in Fig. 3 and is in qualitative agreement with CD results, which indicate 23% α-helix content, 43% β-sheet, 16% β-turn and 18% coil (spectra not shown).

4. Crystallization

For crystallization, 1 µl of buffer containing 50 m*M* Tris–HCl, 150 m*M* NaCl pH 8.3 and SML-2 at 12 mg ml^{-1} was mixed with 1 µl of reservoir solution (0.1 *M* HEPES buffer pH 7.5, 0.1 *M* NaCl, 1.7 *M* ammonium sulfate, 15% glycerol). SML-2 was then crystallized using the hanging-drop vapour-diffusion method at 291 K. After 6 d, crystals of

Table 2
Diffraction data statistics for SML-2 crystallized in space group $P2_12_12_1$.

Values given in parentheses refer to reflections in the outer resolution shell, 2.12–2.07 Å.

Observed/unique reflections	265151/68279
Completeness (%)	99.4 (96.2)
R_{sym}† (%)	5.3 (28.0)
$\langle I/\sigma(I)\rangle$	16.6 (3.8)
Average B factor‡ (Å^2)	21.8

† Defined as $R_{sym} = 100\% \times \sum |I_{i,h} - \langle I\rangle_h| / \sum \langle I\rangle_h$. $I_{i,h}$ is the intensity of the ith observation of reflection h, $\langle I\rangle_h$ is the mean intensity value of the reflection and the summation is over all reflections. ‡ Determined from Wilson plot.

0.28 × 0.2 × 0.2 mm were observed (Fig. 4) which were assigned to space group $P2_12_12_1$. Crystals of a second space group, $C222_1$, grew to 0.2 × 0.2 × 0.2 mm within one month when galactose was added to the lectin in a molar ratio of 100:1.

5. X-ray data collection and analysis

For diffraction experiments, crystals were mounted at 110 K in the nitrogen stream of a Cryostream Cooler (Oxford Cryosystems, England) on a 345 mm imaging-plate detector (MAR Research, Hamburg, Germany) at the EMBL beamline BW7A at DESY, Hamburg, Germany. The crystals diffracted to 2.07 Å resolution, with a mosaicity of 0.28°. Crystals of space group $C222_1$ were investigated under identical conditions but using a conventional Rigaku Denki X-ray generator RU H2B with a direct-driven copper anode.

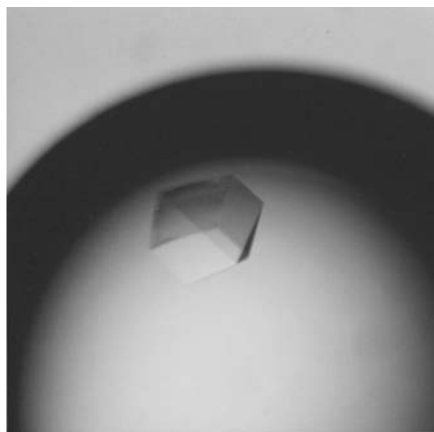

Figure 4
Crystal of the *S. muris* lectin SML-2. The crystals belong to space group $P2_12_12_1$ and have approximate dimensions 0.28 × 0.2 × 0.2 mm.

X-ray diffraction data were evaluated with *DENZO* and *SCALEPACK* (Otwinowski, 1993). The full data-reduction statistics are presented in Table 2. SML-2, which crystallized in the primitive orthorhombic space group $P2_12_12_1$, has unit-cell parameters a = 53.3, b = 129.9, c = 159.8 Å. The volume of the unit cell is 1.06 × 10^6 Å^3. Assuming eight molecules, possibly organized in four dimers or two tetramers, of SML-2 per asymmetric unit, the protein content is about 55% of the unit cell (V_M = 2.29 Å^3 Da^{-1}). The existence and the orientation of non-crystallographic symmetry axes are demonstrated by calculation of the self-rotation function. A twofold rotation axis exists in the (φ = 90, ψ = 45°) direction. A second crystal form was found for SML-2 co-crystallized with modified galactose in the molar ratio 1:100. Identical data quality was obtained to 2.4 Å resolution. The smaller unit-cell parameters a = 74.7, b = 82.0, c = 131.0 Å provide space for two to four molecules in the asymmetric unit (V_M = 3.33–1.66 Å^3 Da^{-1}). This crystal form may prove to be more convenient for phase determination. Knowledge of the existence of the disulfide bridges from mass spectrometry was decisive for the alignment of the PAN modules of SML-2 and HGF. Because of the low sequence identity and because the PAN module comprises only one third of a lectin molecule, all molecular-replacement attempts have so far failed to provide a solution for the SML-2 structure. As an intensive search for heavy-atom derivatives also failed, structure determination using the anomalous signal from S atoms (Dauter *et al.*, 1999; Bartunik, personal communication) has been initiated with both crystal forms.

We thank R. Misselwitz and B. Kannen for the CD measurements and A. Feske for crystallization experiments. The modified galactose was a gift from U. Pfüller, University of Witten-Herdecke, Germany. We are grateful to Y. A. Muller for helpful discussions and to the Fonds der Chemischen Industrie for financial support.

References

Adams, J. H., Sim, B. K. L., Dolan, S. A., Fang, X., Kaslow, D. C. & Miller, L. H. (1992). *Proc. Natl Acad. Sci. USA*, **89**, 7085–7089.
Brown, P. J., Billington, K. J., Bumstead, J. M., Clark, J. D. & Tomley, F. M. (2000). *Mol. Biochem. Parasitol.* **107**, 91–102.
Cuff, J. A., Clamp, M. E., Siddiqui, A. S., Finlay, M. & Barton, G. (1998). *Bioinformatics*, **14**, 892–893.
Dauter, Z., Dauter, M., de La Fortelle, E., Bricogne, G. & Sheldrick, G. M. (1999). *J. Mol. Biol.* **289**, 83–92.
Drapeau, G. R. (1979). *J. Biol. Chem.* **255**, 839–840.
Entzeroth, R., Kerckhoff, H. & König, A. (1992). *Eur. J. Cell Biol.* **59**, 405–413.
Eschenbacher, K. H., Klein, H., Sommer, I., Meyer, H. E., Entzeroth, R., Mehlhorn, H. & Rüger, W. (1993). *Mol. Biochem. Parasitol.* **62**, 27–36.
Hirs, C. H. W. (1956). *J. Biol. Chem.* **219**, 611–621.
Klein, H., Mehlhorn, H. & Rüger, W. (1996). *Parasitol. Res.* **82**, 230–237.
Mehlhorn, H. & Heydorn, A. O. (1978). *Adv. Parasitol.* **16**, 43–91.
Montag, T., Bornhak, M., Loeschner, B., Klein, H., Otto, A., Zyto, N. & Entzeroth, R. (1997). *Eur. J. Cell Biol.* **74**, (Suppl. 46), 21.
Otwinowski, Z. (1993). *DENZO: An Oscillation Data Processing Program for Macromolecular Crystallography.* New Haven, CT, USA: Yale University.
Rini, J. M. (1995). *Annu. Rev. Biophys. Biomol. Struct.* **24**, 551–577.
Rogers, W. O., Rogers, M. D., Hedstrom, R. C. & Hoffman, S. L. (1992). *Mol. Biochem. Parasitol.* **53**, 45–51.
Thomsen, J. & Bayne, S. (1988). *J. Protein Chem.* **7**, 295–296.
Tordai, H., Banyai, L. & Patthy, L. (1999). *FEBS Lett.* **461**, 63–67.
Zyto, N., Bornhak, M., Klein, H., Löschner, B. & Montag, T. (1997). *Eur. J. Cell Biol.* **74**, (Suppl. 46), 41.