The 1.8 Å crystal structure of human tear lipocalin reveals an extended branched cavity with capacity for multiple ligands

Daniel A. Breustedt¹, Ingo P. Korndörfer¹, Bernhard Redl² and Arne Skerra*¹

¹Lehrstuhl für Biologische Chemie, Technische Universität München, D-85350 Freising-Weihenstephan, Germany

²Institut für Molekularbiologie, Medizinische Universität Innsbruck, A-6020 Innsbruck, Austria

*Corresponding author:
Prof. Dr. Arne Skerra
Lehrstuhl für Biologische Chemie
Technische Universität München
D-85350 Freising-Weihenstephan
Germany
Phone: +49 8161 71 4351
Fax: +49 8161 71 4352
e-mail: skerra@wzw.tum.de

Short title: Crystal structure of tear lipocalin

Key words: beta-barrel, cystatin, Lcn1, ligand-binding pocket, retinoic acid, rifampin, von Ebner’s gland protein, zinc
SUMMARY

In contrast with earlier assumptions, which classified human tear lipocalin (Tlc) as an outlier member of the lipocalin protein family, the 1.8 Å resolution crystal structure of the recombinant apo-protein confirms the typical eight-stranded antiparallel β-barrel architecture with an α-helix attached to it. The fold of Tlc most closely resembles the bovine dander allergen Bos d 2, a well characterized prototypic lipocalin, but reveals also similarity with β-lactoglobulin. However, compared with other lipocalin structures Tlc exhibits an extremely wide ligand pocket, whose entrance is formed by four partially disordered loops. The cavity deeply extends into the β-barrel structure, where it ends in two distinct lobes. This unusual structural feature explains the known promiscuity of Tlc for various ligands, with chemical structures ranging from lipids and retinoids to the macrocyclic antibiotic rifampin and even to microbial siderophores. Notably, earlier findings of biological activity as a thiol protease inhibitor have no correspondence in the three-dimensional structure of Tlc, rather it appears that its proteolytic fragments could be responsible for this phenomenon. Hence, the present structural analysis sheds new light on the ligand-binding activity of this functionally obscure but abundant human lipocalin.

Abbreviations: Bda, bovine dander allergen Bos d 2; Blg, β-lactoglobulin; CSD, Cambridge Structural Database; PDB, Protein Data Bank; MAD, multiple wavelength anomalous diffraction; R.m.s., root mean square; SeMet, selenomethionine; Tlc, human tear lipocalin
INTRODUCTION

Tear lipocalin (Tlc, also called Lcn1 or von Ebner’s gland protein) was originally described as a unique major component of human tear fluid, comprising 15 to 33% of its protein content and secreted by the lacrimal glands (1,2). Later on it was also found to be produced by a number of other secretory glands and tissues, including lingual glands, nasal mucosal glands, secretory glands of the tracheobronchial tract, sweat glands, mammary gland, adrenal gland, prostate, thymus, testis, and by corticotrophs of the pituitary gland (3-6).

cDNA cloning and sequencing revealed that this protein component is a member of the lipocalin superfamily with significant amino acid similarity to β-lactoglobulin (7), one of the classical lipocalins. Since it lacks one of the three conserved amino acid motifs that are normally present in lipocalins, Tlc was initially classified as an outlier member of this protein family (8). However, a typical β-barrel fold was predicted from a hydropathy profile of the amino acid sequence (7), circular dichroism analysis (9), and also from an extended site-directed tryptophan substitution and fluorescence study (10), revealing β-sheet and α-helical composition similar to other lipocalins.

Nevertheless, Tlc possesses several biochemical activities that must be conferred by unique structural properties. As an ordinary feature of a lipocalin protein it binds hydrophobic compounds; yet, compared with other members of this family, it exhibits an unusually broad ligand specificity. A large number of lipophilic substances belonging to different chemical classes, including fatty acids, fatty alcohols, phospholipids, glycolipids, cholesterol, retinol, arachidonic acid as well as lipid peroxidation products, such as isoprostanes, were found as endogenous ligands of the protein isolated from tear fluid or from human NT2 precursor cells (5,11,12). Contrasting with other lipocalins, the ligand affinity correlates with the length of the
hydrocarbon chain both for alkyl amides and fatty acids. Thus, Tlc binds most strongly the least soluble lipids (13).

By transporting or scavenging lipid compounds Tlc may function in the integrity of the tear film with its peculiar aqueous-lipid interface and, more generally, in the protection or repair of epithelial surfaces (5). Recently, Tlc was also shown to bind bacterial and fungal iron siderophore complexes, thus suggesting antimicrobial function (14). Furthermore, binding activity for the antibiotic drug rifampin was reported (15). Finally, Tlc accepts several synthetic ligands, such as DAUDA (11-(((5-(dimethylamino)-1-naphthalenyl)sulfonyl)amino)undecanoic acid) (13) or the spin label compound 16-DSA (16-doxylstearic acid) (10).

The structural influence of the three Cys residues in Tlc has been investigated (9). Mass determination of tryptic fragments revealed that a single intramolecular disulphide bond joins Cys-61 with Cys-153 and that Cys-101 resides as free a cysteine (numbers corresponding to the mature polypeptide sequence after processing of the 18 residue signal peptide). The disulphide bridge is conserved in most lipocalins and connects the β-barrel, usually strand D, with the C-terminal peptide segment (16). Reduction of this disulphide bridge in Tlc did not significantly affect protein secondary structure but had an influence on the ligand-binding behaviour (9).

Apart from its ligand-binding properties, Tlc was shown in vitro to inhibit cysteine proteinases (17) and to act as a non-specific endonuclease (18), both activities that are unusual among lipocalins. So far, no high resolution structural information was available for this lipocalin. In order to understand the molecular basis of its possibly multiple physiological functions, we expressed Tlc in E. coli and crystallized the recombinant protein. Here, we report the 1.8 Å crystal structure of Tlc in its apo-state.
EXPERIMENTAL PROCEDURES

Vector construction

The coding sequence for human tear lipocalin, comprising residues 5 to 157 of the mature Lcn1 gene product (7) was amplified from plasmid pQE70/Lcn1 (19) via PCR according to a published procedure (20), using primers 5’-CCG CCT CAG ACG AGG AGA TTC AG-3’ and 5’-GCT CCC TGG AGA GCA GGT TTC GC-3’. The unique amplification product was purified by agarose gel electrophoresis, phosphorylated with T4 polynucleotide kinase (New England Biolabs, Beverly, MA), and ligated with the expression vector pASK75-strepII (21) that had been cut with StuI and Eco47III and dephosphorylated using shrimp alkaline phosphatase (USB, Cleveland, OH). After transformation of E. coli XL1-Blue the plasmid was isolated and its composition was confirmed by restriction digest as well as double-stranded dideoxy-sequencing (ABI PRISM 310 Genetic Analyzer, Applied Biosystems, Foster City, CA). On the resulting vector, which was designated pTlc2, the truncated Tlc was encoded as a fusion protein with the N-terminal OmpA signal peptide and with the C-terminal Strep-tag II of nine residues. The codon for the unpaired Cys-101 (numbering according to the mature full length protein) was subsequently replaced by a Ser codon via site-directed mutagenesis (22) with the oligodeoxynucleotide 5’-AGC TCG CCC TCA GAG TAA AAG ATG TAG TG-3’ and yielding the vector pTlc3, which was used throughout this study.

Protein production and purification

Recombinant Tlc was produced in the E. coli K-12 strain JM83 (23) harbouring pTlc3 via secretion into the bacterial periplasm. Cultures (2 l) were grown in LB medium (24) supplemented with 100 mg/l ampicillin at 22°C. Gene expression was induced at a cell density of OD550 = 0.5 by adding 0.2 mg/l anhydrotetracycline (21). After further shaking for 3 h the cells were harvested by centrifugation, resuspended in 0.5 M sucrose, 1 mM EDTA, 100 mM Tris-HCl pH 8.0, and kept on ice for 30 min. The resulting spheroplasts were
sedimented by centrifugation and the supernatant containing the recombinant protein was recovered. The protein extract was dialyzed against 150 mM NaCl, 1 mM EDTA, 100 mM Tris-HCl pH 8.0 and applied to a column with immobilized engineered streptavidin (25) using the same buffer. The recombinant Tlc was competitively eluted by application of 2.5 mM D-desthiobiotin in the chromatography buffer. Elution fractions were concentrated, applied to a Superdex 75 gel filtration column (Amersham Pharmacia, Uppsala, Sweden), equilibrated with 150 mM NaCl, 1 mM EDTA, 50 mM Tris-HCl pH 8.0 as running buffer, and eluted in a homogeneous peak corresponding to the monomeric protein. The yield was ca. 1.5 mg of purified protein per 1 l E. coli culture.

SeMet-labelled Tlc was produced in the same manner by shutting down endogenous Met biosynthesis of E. coli in the presence of the amino acid derivative (26,27). To this end the culture was grown in M9 minimal medium (24) supplemented with 0.4 % (w/v) glucose, 1 mM MgSO₄, 200 mg/l L-Pro, 10 mg/l thiamin, and 100 mg/l ampicillin at 22°C. L-Lys, L-Thr, L-Phe (100 mg/l each), L-Leu, L-Ile, L-Val (50 mg/l each), and L-SeMet (25 mg/l) were added at a cell density of OD₅₅₀ = 0.5. After 15 min incubation, gene expression was induced by adding 0.2 mg/l anhydrotetracycline. The cells were harvested after further shaking at 22 °C for 14 h, and recombinant Tlc was purified as above. Gel filtration proved to be dispensable due to the high purity after Strep-tag chromatography and to the higher yield of ca. 2.5 mg purified recombinant protein per 1 l E. coli culture, probably as a result of the longer induction period. SeMet incorporation was verified by MALDI-TOF mass spectrometry.

**Crystallization and data collection**

Crystals of native as well as SeMet-labelled Tlc were grown using the hanging drop vapour diffusion technique by mixing 3 µl protein solution (10 mg/ml, dialyzed against 10 mM Tris-HCl pH 8.0) with 1 µl precipitant solution on a siliconized glass cover slip, followed by
equilibration against 0.5 ml of the precipitant solution. Several crystals were obtained at 12°C in the presence of 60 % (w/v) ethylene glycol, 200 mM Zn-acetate, 100 mM Tris-HCl pH 7.0 – 7.9 after about six weeks. Protein crystals were harvested directly from the drop using Nylon loops (Hampton Research, Laguna Niguel, CA), cryo-protected by transfer to a drop of Paratone N (Hampton Research) – thereby removing excessive mother liquor – and frozen in a 100 K nitrogen stream (Oxford Cryosystems, Oxford, UK). Native diffraction data (Table 1) were collected at the SLS beamline PX1 (PSI, Villigen, Switzerland). A three-wavelength anomalous dataset of the SeMet derivative (Table 1) was collected at the synchroton beamline BM14 (ESRF, Grenoble, France). Data were processed with MOSFLM and SCALA (28). The space group was determined as C2 (a = 76.4 Å, b = 61.4 Å, c = 50.8 Å, β = 129.7° for the native data set) with one protein molecule in the asymmetric unit.

Structure determination, model building, and refinement

The selenium substructure of a crystal grown at pH 7.9 was solved with SnB (29) using anomalous peak data, whereby three of the four SeMet sites that had to be expected from the primary structure were identified. Phasing was carried out with SHARP (30,31) using the three SeMet sites and the anomalous data from all three wavelengths (Table 1) to a resolution of 1.8 Å. Solvent flattening was subsequently performed with SOLOMON (32) by using a solvent content of 45 %. A partial structural model, comprising the β-barrel and the α-helix but lacking residues in the loop regions and at both polypeptide termini, was calculated with ARPwARP (33) as implemented in SHARP. Missing residues were then added in several sessions of manual rebuilding with QUANTA (34), alternating with restrained refinement cycles using REFMAC 5 (28). During this process three coordinated zinc ions and two chloride ions were identified and added to the model. Final refinement was carried out with the native data set of a crystal grown at pH 7.0. At this pH, two previously missing residues (Asn-32 and Leu-33) in loop #1 appeared in the electron density map and were built
accordingly. Water molecules were added where stereochemically plausible and the $2F_o-F_C$ and $F_o-F_C$ difference Fourier maps revealed densities of more than 1.0 and 3.5 $\sigma$, respectively. The backbone dihedral angles fall mostly into the favoured regions of the Ramachandran plot. The only exception is residue Leu-33, which is situated C-terminal to the disordered part of loop #1 and has only weak electron density.

**Analysis of the crystal structure**

The few missing residues at the tips of loop #2 (Leu-56 and Ile-57) and loop #4 (Leu-105 and His-106) were inserted in QUANTA (34) with plausible stereochemistry (using the “regularize range” option). In order to model the stretch of seven residues in loop #1 without clear electron density (Asp-25 to Met-31), the Protein Data Bank was searched for a loop with identical length and matching conformation of the well resolved flanking segments (C$\alpha$ positions 21 to 23 and 33 to 35, respectively) using INSIGHT II (Accelrys, San Diego, CA). One nicely fitting peptide segment, corresponding to residues 117 to 125 of actophorin (PDB accession code 1CNU), was found and grafted onto the Tlc stump using least squares superposition of the flanking segments. Side chains were subsequently adjusted to match the Tlc sequence and oriented in standard conformation (35). The completed model was subjected to one round of refinement using REFMAC 5 with the occupancy of all three amended segments (residues 25 to 31, 56 and 57, 105 and 106) set to 0.01, resulting in an $R$-factor of 19.6 % and $R_{free}$ of 25.4 %. There were no structural clashes due to the new residues. Notably, the modelled loop #1 partly adopted a similar orientation as the corresponding loop of the lipocalin Bda, though shorter by two residues (see text).

Graphical representations were prepared using MOLSCRIPT (36), RASTER3D (37), and PyMOL (38). Secondary structure assignments were made with DSSP (39) and extracted with MAKEMOLS (40). For use with MOLSCRIPT, electron densities were converted from
CCP4 format (28) with MINIMAGE (41). Molecular surfaces were either prepared with PyMOL or calculated with GRASP (42) and exported for use in MOLSCRIPT with a modified version of UNGRASP (37). Superpositions were made with LSQMAN (43). Ligand complexes were modelled using INSIGHT II. The crystallographic coordinates of the apo-Tlc structure – without the modelled segments – have been deposited at the Protein Data Bank (44) under accession code 1XKI.

RESULTS

Crystallization, structure determination, and quality of the final model

Recombinant human Tlc, comprising residues 5 to 157 of the mature protein (7,19), with the free Cys-101 substituted by Ser, was fused to the OmpA signal peptide and secreted into the periplasm of *E. coli*, where processing, folding, and efficient disulphide bond formation can take place. One-step purification from the bacterial periplasmic extract was achieved by streptavidin affinity chromatography via the Strep-tag II (25), which had been fused to the C-terminus. Tlc was finally isolated as a homogeneous monomeric protein by gel filtration.

Crystallization of Tlc was achieved at pH 7.9 with ethylene glycol as precipitant in the presence of Zn-acetate. The crystal structure was determined by MAD phasing after *in vivo* labelling with selenomethionine (SeMet). Refinement against a data set for a crystal of the native protein, grown under similar conditions at pH 7.0, resulted in an *R*-factor of 18.9% and an *R*$_{free}$ of 25.4% with good stereochemistry (Table 2). One molecule of the apo-protein was found in the asymmetric unit. The quality of the electron density map at a resolution of 1.8 Å allowed to unambiguously build 128 of the 152 residues of the recombinant Tlc (neglecting the Strep-tag II).

The N-terminal residues Ala-5 to Gln-11 and the C-terminal residues starting from Glu-151, including the Strep-tag II (residues 158 – 166), were disordered. The disulphide bond joining
Cys-61 and Cys-153 was also not fully resolved in the electron density and hence omitted from the model. Residues Asp-25 to Met-31, which form the long first loop connecting β-strands A and B at the open end of the β-barrel and partially cover the calyx in other lipocalins (e.g. bovine lipocalin allergen Bos d 2 (45)), were largely disordered. Furthermore, small pieces of electron density were missing at the tips of loop #2 connecting strands C/D (Leu-56 and Ile-57) and of loop #4 connecting strands G/H (Leu-105 and His-106).

Adjacent residues had elevated temperature factors but could be built unambiguously. Hence, as in other lipocalins the β-barrel of Tlc appears rather rigid while extended loops and extraneous terminal segments of the polypeptide chain seem to be more flexible, probably also due to a lack of stabilizing crystal contacts. For purposes of illustration, the missing loop segments at the entrance to the ligand pocket were modelled with plausible stereochemistry (for details see Materials and Methods) and included in the Tlc structure described here.

**Structure of tear lipocalin**

Tlc exhibits the typical lipocalin fold, consisting of an eight-stranded, strictly antiparallel β-barrel that forms a central cavity. The cavity opens to one end of the barrel, where four loops (#1 – #4) connect neighbouring β-strands (dubbed A – H) in a pairwise manner (Fig. 1). A C-terminal α-helix (residues Leu-127 to Ala-136), which is followed by an almost extended peptide segment (strand l), packs against one side of the β-barrel. The N-terminal peptide segment of Tlc covers the bottom of the barrel between the loops connecting strands B/C and F/G and thereby buries Trp-17, one of the few residues conserved in all lipocalins (8).

Three of the loops at the open end of the β-barrel are rather long and partially disordered in the crystal structure, hence allowing access to the central cavity even for voluminous ligands. Especially the Ω-type loop #1 with its altogether 13 residues is one of the largest loops
among the prototypic lipocalins that have been structurally characterized so far (16). On its N-terminal side this loop is likely to form a short α-helix comparable to those observed in Bda and Blg (Fig. 2). Loops #2 und #4 appear as β-hairpins, whereas the tip of the rather short loop #3 connecting β-strands E/F (cf. Fig. 1A) locally assumes the conformation of a 3_10-helix (residues Asp-80 to Gly-82).

Even though Tlc represents one of the smallest members of the lipocalin family – with 158 residues for the native, mature polypeptide (cf. Fig. 2) – it exhibits an unusually large cavity inside its β-barrel. Nevertheless, the β-barrel itself, with the three loops that tie it up at its closed end and with the α-helix attached to it, appears conformationally rigid and is nicely superimposable to the supersecondary structure of other lipocalins.

Interestingly, apart from the wide mouth formed approximately at the middle between the loops #1 to #4 there is a second, smaller entry to the cavity of Tlc, which appears between β-strands A and H (Fig. 1D). The main entrance of the calyx-like pocket measures about 10 Å in diameter and is lined with negatively charged (Glu-34 and Asp-80) as well as hydrophobic residues. The cavity itself exhibits a bifurcated shape, is about 15 Å deep, and its inside is slightly positively charged (Fig. 3). This positive charge arises from the side chains of residues His-86 and Lys-114 together with the dipoles of main chain amide groups of up to ten residues in the vicinity.

The remaining side chains that line the pocket are hydrophobic in nature, particularly residues Phe-99 and Met-39, which separate the two lobes at the bottom of the calyx. Although the cavity is accessible to solvent, the lower end and, most notably, the two lobes at its bottom are free of ordered water molecules in the crystal structure. This may be due to the absence of suitable H-bond donor/acceptor groups in order to fix the polar molecule (46).
Implications for ligand binding

There are three bound zinc ions in the crystal structure of Tlc (Fig. 1B). Zn-H2 is complexed by His-96 and Glu-125 together with two chloride ions. Zn-H3 is coordinated by His-92 and Asp-129, also located in the region around the N-terminus of the α-helix, as well as by two water molecules. In contrast, Zn-H1 is involved in a crystal packing contact between two Tlc molecules and complexed by the side chains of Glu-73 and Lys-76 of one polypeptide and Glu-102 of a symmetry-related molecule. The fourth coordination site is occupied by a water molecule.

All three bound Zn(II) ions exhibit the typical tetrahedral coordination geometry, and their distances to the closest side chain atoms of Tlc (Table 3) are in accordance with the published parameters for Zn-binding sites (47). Notably, when the Zn(II) that had been used in the precipitant solution (see Experimental Procedures) was substituted with Mg(II), no crystal growth was observed. However, preliminary binding studies between the recombinant Tlc and Zn(II) revealed a rather weak dissociation constant of approximately 1.33 mM (determined as half-maximal metal ion concentration to achieve saturation during CD titration at 212 nm; Skerra, Cirl and Breustedt, to be published).

The unusual shape of the Tlc binding pocket prompted us to investigate possible modes of interaction with two of its known ligands, retinoic acid and rifampin (7,15). Unfortunately, corresponding co-crystallisation trials remained unsuccessful. However, molecular docking with rifampin (CSD entry RIFAMP) clearly revealed that the bulky molecule occupies almost the entire pocket and that it fits the cavity with the two lobes astonishingly well (Fig. 3B). In contrast, two possible binding modes were identified for retinoic acid (CSD entry VITAAC01), which is significantly smaller than rifampin and has a more extended shape. In both
orientations, the β-ionone ring of retinoic acid can be placed into the larger of the two lobes at the bottom of the cavity while the terminal carboxylate group is alternatively arranged at the two different exits (Fig. 3C,D).

**Structural comparison with other lipocalins**

A search through the PDB using the DALI server (48) revealed that Tlc (without the modelled loop regions) is structurally most closely related to the bovine dander allergen Bos d 2 (PDB accession code 1BJ7 (45); Z-score = 15.1), one of the prototypic lipocalins (16). Mouse major urinary protein (1MUP (49); Z-score = 14.2), rat epididymal retinoic acid-binding protein (1EPA (50); Z-score = 14.1), and bovine β-lactoglobulin (1BEB (51); Z-score = 13.7) showed lower similarity. Due to its comparably high sequence homology of 23 % the latter lipocalin was previously employed as a template structure for the homology modelling of Tlc (10). Even though Tlc and Bda display merely 17.8 % sequence identity (Fig. 2B), superposition of the 58 Cα atoms of the β-barrel (16) gave rise to a low R.m.s. deviation of 1.5 Å while a corresponding value of 2.3 Å was calculated for all 128 superimposable Cα atoms.

Despite the high structural similarity of the β-barrel and α-helix, major differences were found in the loop regions at the open end of the lipocalin calyx (Fig. 2A). Loop #1, which is largely disordered in Tlc but well defined in the Bda crystal structure (45), runs almost across the opening of the calyx, thereby controlling accessibility of the ligand-binding site. The spatial orientation of the segments flanking this loop, which could be built with confidence in the Tlc crystal structure, together with the comparable loop lengths in both lipocalins (with two residues less in Tlc than in Bda; cf. Fig. 2B), indicate that the loop might adopt a similar, though more flexible conformation in Tlc.
In addition, the Cα position of residue Gly-59 at the tip of loop #2 is located ca. 8 Å further away from the central axis of the β-barrel in Tlc than the homologous residue Gly-61 of Bda. Notably, the well resolved loop #3, which is flipped towards the cavity in Bda, points outward in Tlc, resulting in an 8 Å displacement of the Cα positions of the corresponding residues Gly-84 and Gly-82. As a consequence, the entrance to the cavity is wide open in Tlc whereas it is closed in Bda. Apart from the small gap in the electron density of Tlc, the conformation of loop #4 is similar in both lipocalins. Interestingly, at the bottom of the β-barrel the short loops connecting the β-strands B/C and F/G, respectively, are shifted by about 7 Å away from the central axis in Tlc compared with Bda. Hence, another surface region with concave shape appears opposite to the conventional ligand pocket in this lipocalin.

The structure of Blg, which has served as a homologue of Tlc so far (7,10), is highly similar to that of Bda. Again, the β-barrel and α-helix superimpose well with those of Tlc (Fig. 2A) whereas the loop regions of Blg resemble more closely those of Bda. Loops #1, #2, and #3 are found in almost the same conformation as in Bda while loop #4 is shifted away from the β-barrel axis in Blg compared with both Bda and Tlc. The N-terminal peptide segment, which is disordered in Tlc and in Bda, is well defined in Blg and runs across the bottom of the β-barrel as observed for most other lipocalins (16).

Comparison of tear lipocalin with protease inhibitors

Previous experimental evidence indicated that full length Tlc as well as a synthetic peptide comprising residues Leu-3 to Ala-21 exert inhibitory activity toward the cysteine proteinase papain (17,52). Three sequences were identified within the first 17 N-terminal residues of Tlc that are homologous to the motifs known to be essential for the inhibitory activity of cystatins and the closely related stefins (53). Although the sequence motifs are more than 40 residues apart in the primary structures of these well characterized inhibitors, they are spatially close...
in their three-dimensional fold, thus forming a complementary interface with the catalytic cleft of the proteinase.

To investigate any conformational similarities between the N-terminal segment of Tlc and the cystatin-type proteinase inhibitor fold, the responsible motifs of chicken egg-white cystatin (PDB accession code 1CEW (54)), stefin A (PDB accession code 1NB5 (55)), and stefin B (PDB accession code 1STF (56)) were superimposed onto each other and with the N-terminal segment of Tlc (Fig. 4). While mutual superposition of the different cystatin-type inhibitors revealed a clear pattern of conserved main chain and side chain orientation, the proposed inhibitory peptide of Tlc did not exhibit reasonable structural similarity. Especially Trp-17, which was considered functionally homologous to the central tryptophan residue of the cystatins in the third sequence motif and occurs there exposed at the tip of a hairpin loop, is buried in the hydrophobic core of Tlc. In fact, this amino acid is even part of the Gly-Xaa-Trp signature characteristic for all lipocalin structures (8) and therefore not likely to participate in an inhibitor-proteinase interaction, at least as long as the native fold is adopted.

DISCUSSION

As predicted based on its amino acid sequence (7), Tlc exhibits the typical lipocalin fold, comprising an eight-stranded, antiparallel β-barrel with a C-terminal α-helix. The crystal structure reveals a remarkably large ligand cavity inside the β-barrel and hence provides insight into some of the unusual features of Tlc as member of the lipocalin family, especially its broad ligand specificity. The high promiscuity for a range of different ligands is reflected by the unusual bifurcated shape of the central cavity, which is about 15 Å deep. With a mouth of roughly 10 Å in diameter and an even wider bottom, the binding site of Tlc cannot only
accommodate slim fatty acid molecules but also much bulkier compounds like cholesterol and rifampin.

A three-dimensional model of Tlc was previously constructed on the basis of an extended site-directed tryptophan substitution and fluorescence study (10), in which each native amino acid, except Gly-117, was sequentially exchanged. Measurement of the fluorescence emission maxima for each mutant revealed characteristic periodicities as expected for eight \( \beta \)-strands and altogether three \( \alpha \)-helical segments. These regions were subsequently aligned with the corresponding secondary structural elements of Blg and a homology model of Tlc was constructed. All in all, this homology model is in good agreement with the crystal structure described here, whereby the lengths of the \( \beta \)-strands and of the C-terminal \( \alpha \)-helix deviate by no more than 3 residues from the lengths determined on basis of the crystallographic data using DSSP (39).

Furthermore, perturbations in the periodicities of the fluorescence emission peaks within the secondary structure elements allowed the correct prediction of the position of the \( \alpha \)-helix with respect to the \( \beta \)-barrel. Residues Val-113 and Leu-115 of \( \beta \)-strand H as well as Phe-130 within the C-terminal \( \alpha \)-helix showed less exposure to solvent than expected and were therefore placed at the interface between the \( \alpha \)-helix and \( \beta \)-strand H. These findings could be confirmed in the crystal structure, where Phe-130 is located almost in the middle of the \( \alpha \)-helix and packs against the \( \beta \)-barrel, contacting the two aliphatic side chains in strand H. Its \( C_\alpha \) position lies in 7.5 Å and 6.0 Å distance to the \( C_\alpha \) positions of Val-113 and Leu-115, respectively.

Two other spectroscopic perturbations in the \( \beta \)-strands A (at Lys-20 and Ala-21) and F (at Ile-89 and Arg-90) were interpreted as \( \beta \)-bulges and, indeed, the side chains of these
neighbouring residues are all solvent-exposed in the crystal structure, with concomitant irregularities in the β-strand backbone. Interestingly, the three residues following the β-bulge in strand A (Met-22 to Val-24) form a short antiparallel β-sheet with the short strand I (Ser-143 to Leu-145) that is part of the extended C-terminal polypeptide segment, which is found in most lipocalins and attaches to one side of the β-barrel.

The predicted short loop #3 connecting β-strands E/F becomes also apparent from the crystal structure. Yet, its tip points outward from the β-barrel axis, in contrast with the situation in Blg (cf. Fig. 2A). For residues Pro-29 to Thr-37 as part of loop #1 and residues Glu-9 to Ser-14 close to the N-terminus of Tlc α-helical and 3_{10}-helical conformation, respectively, was predicted. These assumptions could not be confirmed in the crystal structure, even though it must be pointed out that some of the residues in these regions are not well resolved in the electron density.

Determination of the dissociation constants between various Tlc mutants and 16-doxylstearic acid was used to identify residues that play a role in ligand recognition and are therefore likely to be located inside the calyx (10). In these experiments, three alanine residues (Ala-51, Ala-66, and Ala-86) were found that could not be replaced by tryptophan without significantly diminishing ligand affinity. A fourth residue, Ala-79, was furthermore proposed to play a role in this context, but did not exhibit a similar experimental effect. In the crystal structure, all four Ala residues, as well as another two residues (Cys-101 and Gly-103) that were previously found to interact with a ligand of Tlc (57), are indeed located within the cavity.

Apart from that, energy minimization of the homology model for Tlc led to the assumption that Tyr-97 is exposed in the ligand pocket (10). Contrasting, in the crystal structure this
residue appears with its polar hydroxyl group on the outside of the β-barrel. Perturbation of the alternating periodicity around Met-22 in β-strand A was explained by its close distance to the side chain of Lys-114. The authors claimed that the positive charge of this residue induces the unexpected red shift in the fluorescence of the Met-22-Trp mutant and demonstrated that in the double mutant Met-22-Trp / Lys-114-Cys the expected periodic behaviour is reinstalled. Indeed, in the crystal structure Met-22 and Lys-114 directly neighbour each other on β-strands A and H, respectively, with their side chains lying almost in parallel.

Interestingly, in the Trp substitution study all but one of the Tlc mutants, i.e. Gly-117-Trp, were purified as dimeric proteins. Inspection of the crystal structure revealed that this residue is located within the hydrophobic core of the protein. The introduction of another bulky residue is probably not tolerable there because it would disturb the dense packing of the surrounding side chains. However, in the production and purification of the recombinant wild-type Tlc employed here for protein crystallisation no signs of homo-dimerization appeared. Also, no tight packing contacts with a neighbouring molecule were observed in the Tlc crystal so that an explanation for the previously described Tlc dimerization remains obscure. The unpaired Cys-101 residue does not seem to play a role in this respect because in contrast with other lipocalins that display a free Cys residue for covalent complex formation, such as NGAL (58), its side chain is buried sterically inaccessible inside the binding pocket of Tlc.

In the presently described crystal structure of the apo-form of human Tlc a stretch of seven residues in loop #1, the tips of loop #2 and loop #4 as well as the disulphide bond linking the C-terminus of the polypeptide chain to loop #2 were not defined in the electron density map. Adjacent residues showed high temperature factors, indicating that the whole loop region at the entrance to the cavity exhibits enhanced mobility. Previous investigation of the structural
changes in Tlc upon ligand binding via CD spectroscopy revealed that ligand complexation induces the formation of additional β-sheet structure and of a more rigid state of the protein (59). Thus, it is conceivable that the loops at the open end of the β-barrel can at least to some degree adapt to different ligands.

The structure of Tlc is most similar to Bda, a lipocalin with yet unknown biological function produced by the sweat glands of domestic animals. In this protein the loop regions are more rigid and some minor structural differences with Tlc are found. Loop #1 partially forms an α-helix in Bda and runs almost straight across the entrance to the ligand-binding site. Also, loops #2 and #3 are shifted towards the central axis of the β-barrel in Bda compared with Tlc. As result, the ligand pocket is essentially obstructed in this allergen, and in fact no endogenous ligands, except of two water molecules, were found inside its cavity (45).

Unexpectedly, two zinc-binding sites were identified in the crystal structure of Tlc, together with a third one that occurs at a crystal packing contact. Spectrophotometric titration experiments revealed that there is a measurable affinity for Zn(II) in the low millimolar range. Hence, the zinc-binding effect may not just be an artefact of crystallisation in the presence of 200 mM Zn(II), even though the apparent dissociation constant seems to be above physiological levels. While Zn(II)-binding activity was not yet described for Tlc, zinc-responsive promoter elements were identified in the Lcn1 gene (60). Furthermore, it was shown that gene expression is induced in knockout mice for MTF-1 (metal-responsive transcription factor 1), a highly conserved protein involved in the transcriptional response to heavy metal exposure (61).

Further to its broad ligand-binding function, which is a more or less typical feature of lipocalins, Tlc was reported to inhibit the proteolytic activity of papain. Van’t Hof et al. (17)
could show that Tlc inhibits papain to the same extent as cystatin S, an inhibitor also present at high concentration in tear fluid. However, the target proteinase of cystatin S has not yet been identified and its inhibitory activity towards papain is much lower than that of cystatin C. In following experiments with recombinant Tlc – and cystatin C serving as positive control – its inhibitory function was confirmed, but it also became evident that inhibition of papain occurs only in a narrow concentration range (52).

Still, a synthetic 18 amino acid peptide derived from the N-terminus of Tlc exhibited comparable proteinase inhibitory activity as a peptide that contained the three inhibitory motifs of cystatin S linked by single Ala residues, although both peptides inhibited papain only at much higher concentration than their respective parent proteins (17). In fact, at increased papain concentration Tlc is used as a substrate (52). N-terminal sequencing revealed that two cleavage sites lie within the N-terminal Leu-Leu-Ala-Ser motif (residues 3 to 6 of Tlc), but not in the conserved Gln-X-Val-X-Gly motif (residues 11 to 15 of Tlc) as it is typical for cystatin-type inhibitors.

Structural comparison revealed (Fig. 4B) that while in various cystatins the three inhibitory sequence motifs form a contiguous interface that binds to the catalytic cleft of the target proteinase (also reviewed in (53)) the N-terminal segment of Tlc cannot be superimposed onto this structural element in a meaningful way. Whereas the N-terminal Leu-Leu-Ala-Ser sequence and the first residue of the Gln-X-Val-X-Gly motif are not resolved in the electron density and probably not conformationally ordered in Tlc, the third sequence motif containing Trp-17 is buried within its hydrophobic core. Since the Trp motif seems to be of minor importance for the inhibitory activity of cystatin-derived peptides (62) it is possible that the flexible N-terminus of Tlc acts in a peptide-like manner. This would also explain why Tlc is cleaved at higher papain concentration within one of the inhibitory motifs, similarly as the cystatin-derived peptides but in contrast to the full-length cystatins.
Notably, there are several other proteinase inhibitors which rely on an eight-stranded β-barrel fold similar to Tlc and for which the mode of interaction with their respective target enzymes is known in detail (Fig. 4C,D). Triabin (PDB accession code 1AVG (63)), a 142 residue protein from the saliva of the blood-sucking insect Triatoma pallidipennis, prevents blood clot formation by inhibiting the serine proteinase thrombin via binding to its exosite I. Unlike the prototypic lipocalin fold, β-strands A to D are arranged in an up-up-down-down topology while the following β-strands are antiparallel.

Recently, the structure of staphostatin B of Staphylococcus aureus, the inhibitor of the cysteine proteinase staphopain B, was elucidated (PDB accession code 1PXV (64)). In contrast to Tlc with its eight strictly antiparallel β-strands, neighboring strands A and B of staphostatin B have parallel topology. The inhibitory site of staphostatin B, which was co-crystallized with its target proteinase, is located in the loop connecting the two C-terminal β-strands. Residues 97 to 101 of staphostatin B interact with the catalytic cleft of staphopain. Interestingly, this peptide stretch does not bind in a substrate-like manner but in an inverted (antiparallel) orientation.

Like Tlc, the metalloproteinase inhibitor Inh (PDB accession code 1SMP (65)) from Erwinia chrysanthemi folds into a strictly antiparallel β-barrel. This protein was shown to inhibit the 50 kDa proteinase from Serratia marcescens and its presumed physiological role is the protection of periplasmic proteins against proteolysis (66). In the crystallized complex with this proteinase, the five N-terminal residues of the inhibitor extend away from the β-barrel and bind to the catalytic cleft of the enzyme. Faint electron density for residue Ser-1 of the inhibitor, together with N-terminal sequencing, revealed that this residue is cleaved by ca. 50 % in the proteinase-inhibitor complex.
In contrast with Tlc, none of these proteinase inhibitors possesses a cavity that appears suitable for ligand binding, rather the interior of each β-barrel is densely packed with mostly hydrophobic side chains. Furthermore, they lack the C-terminal α-helix typical for lipocalins and, due to their varying β-barrel topologies, they exhibit structural differences in the loop regions. However, comparison with Tlc also reveals some similarities. The Cα positions of the β-barrel residues superimpose astonishingly well, resulting in R.m.s. deviations of 1.6 Å, 2.1 Å, and 2.2 Å for triabin, staphostatin B, and Inh, respectively.

Moreover, the Gly-Xaa-Trp motif, which is strictly conserved in the lipocalins, is also found in these three inhibitors at approximately the same position (Fig. 4C), close to the bottom of the β-barrel with the Trp side chain buried in the hydrophobic core. Indeed, triabin and Inh have recently been ascribed to the calycin protein superfamily (67). Nevertheless, a structural explanation for the inhibitory activity of Tlc towards a thiol proteinase is not evident from comparison with either of these three proteins (Fig. 4D).

Even though experimental evidence exists that the N-terminus of Tlc can act as a weak inhibitor of a thiol-proteinase in vitro, this probably does not correspond to a relevant physiological function. There are different, more potent cystatins present in the same secretory fluids and tissues, which, in contrast to Tlc, have also been isolated from tear samples in complex with various proteinases (68). Consequently, the major function of Tlc seems to lie in the binding of multiple physiological or pathological compounds. In order to gain further insight into the mode of interaction between this lipocalin and its presumed ligands, co-crystallization experiments are currently under way.
ACKNOWLEDGEMENTS

We gratefully acknowledge the help of G. Reil for mass-spectrometric analysis of the SeMet derivative and the EMBL Grenoble Outstation, in particular M. Walsh, for support of measurements at the ESRF under the European Union "Improving Human Potential Programme". B.R. was supported by FWF grant P14850.
REFERENCES

1. Josephson, A. S., and Lockwood, D. W. (1964) J Immunol 93, 532-539
2. Bonavida, B., Sapse, A. T., and Sercarz, E. E. (1969) Nature 221, 375-376
3. Scalfari, F., Castagna, M., Fattori, B., Andreini, I., Maremmani, C., and Pelosi, P. (1997) Comp Biochem Physiol 118B, 819-824
4. Lacazette, E., Gachon, A. M., and Pitiot, G. (2000) Hum Mol Genet 9, 289-301
5. Redl, B. (2000) Biochim Biophys Acta 1482, 241-248
6. Wojnar, P., Dirnhofer, S., Ladurner, P., Berger, P., and Redl, B. (2002) J Histochem Cytochem 50, 433-435
7. Redl, B., Holzfeind, P., and Lottspeich, F. (1992) J Biol Chem 267, 20282-20287
8. Flower, D. R. (1996) Biochem J 318, 1-14
9. Glasgow, B. J., Abduragimov, A. R., Yusifov, T. N., Gasymov, O. K., Horwitz, J., Hubbell, W. L., and Faull, K. F. (1998) Biochemistry 37, 2215-2225
10. Gasymov, O. K., Abduragimov, A. R., Yusifov, T. N., and Glasgow, B. J. (2001) Biochemistry 40, 14754-14762
11. Glasgow, B. J., Abduragimov, A. R., Farahbakhsh, Z. T., Faull, K. F., and Hubbell, W. L. (1995) Curr Eye Res 14, 363-372
12. Lechner, M., Wojnar, P., and Redl, B. (2001) Biochem J 356, 129-135
13. Gasymov, O. K., Abduragimov, A. R., Yusifov, T. N., and Glasgow, B. J. (1999) Biochim Biophys Acta 1433, 307-320
14. Fluckinger, M., Haas, H., Merschak, P., Glasgow, B. J., and Redl, B. (2004) Antimicrob Agents Chemother 48, 3367-3372
15. Gasymov, O. K., Abduragimov, A. R., Gasimov, E. O., Yusifov, T. N., Dooley, A. N., and Glasgow, B. J. (2004) Biochim Biophys Acta 1688, 102-111
16. Skerra, A. (2000) Biochim Biophys Acta 1482, 337-350
17. van't Hof, W., Blankenvoorde, M. F., Veerman, E. C., and Amerongen, A. V. (1997) J Biol Chem 272, 1837-1841
18. Yusifov, T. N., Abduragimov, A. R., Gasymov, O. K., and Glasgow, B. J. (2000) *Biochem J* **347**, 815-819

19. Holzfeind, P., Merschak, P., Rogatsch, H., Culig, Z., Feichtinger, H., Klocker, H., and Redl, B. (1996) *FEBS Lett* **395**, 95-98

20. Skerra, A. (1992) *Nucleic Acids Res* **20**, 3551-3554

21. Skerra, A. (1994) *Gene* **151**, 131-135

22. Geisselsoder, J., Witney, F., and Yuckenber, P. (1987) *Biotechniques* **5**, 786-790

23. Yanisch-Perron, C., Vieira, J., and Messing, J. (1985) *Gene* **33**, 103-119

24. Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbour Laboratory Press, Cold Spring Harbour, New York

25. Skerra, A., and Schmidt, T. G. (2000) *Methods Enzymol* **326**, 271-304

26. Van Duyne, G. D., Standaert, R. F., Karplus, P. A., Schreiber, S. L., and Clardy, J. (1993) *J Mol Biol* **229**, 105-124

27. Gassner, N. C. (1998) *PhD thesis, University of Oregon, U.S.A.*

28. CCP4 (1994) *Acta Crystallogr D* **50**, 760-763

29. Weeks, C. M., and Miller, R. (1999) *J Appl Crystallogr* **32**, 120-124

30. de la Fortelle, E., Irwin, J., and Bricogne, G. (1997) *SHARP: a maximum-likelihood heavy atom parameter refinement and phasing program for the MIR and MAD methods*. Crystallographic Computing (Bourne, P. E., and Watenpaugh, K. D., Eds.), Oxford University Press, Oxford, U.K.

31. de la Fortelle, E., Irwin, J., and Bricogne, G. (1997) *Advances in MIR and MAD phasing: Maximum-likelihood refinement in a graphical environment with SHARP*. Proceedings of the CCP4 Study Weekend (Wilson, K. S., Davies, G., Ashton, A. W., and Bailey, S., Eds.), CCLRC Daresbury Laboratory, York, U.K.

32. Abrahams, J. P., and Leslie, A. G. W. (1996) *Acta Crystallogr D* **52**, 30-42

33. Lamzin, V. S., and Wilson, K. S. (1993) *Acta Crystallogr D* **49**, 129-149

34. Oldfield, T. J. (2001) *Acta Crystallogr D* **57**, 82-94
35. Ponder, J. W., and Richards, F. M. (1987) *J Mol Biol* **193**, 775-791
36. Kraulis, P. J. (1991) *J Appl Crystallogr* **24**, 946-950
37. Merrit, E. A., and Bacon, D. J. (1997) *Methods Enzymol* **277**, 505-524
38. DeLano, W. L. (2002) http://www.pymol.org
39. Kabsch, W., and Sander, C. (1983) *Biopolymers* **22**, 2577-2637
40. Holm, L. (1995) *Protein Data Bank Newsletter* **71**, 6-9
41. Arnez, J. G. (1994) *J Appl Crystallogr* **27**, 649-653
42. Nicholls, A., Sharp, K. A., and Honig, B. (1991) *Proteins* **11**, 281-296
43. Kleywegt, G. J., Zou, J. Y., Kjeldgaard, M., and Jones, T. A. (2001) *Around O.*
   International Tables for Crystallography, Crystallography of Biological Macromolecules
   (Rossmann, M. G., and Arnold, E., Eds.), F, Kluwer Academic Publishers, Dordrecht,
   Netherlands
44. Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H.,
   Shindyalov, I. N., and Bourne, P. E. (2000) *Nucleic Acids Res* **28**, 235-242
45. Rouvinen, J., Rautiainen, J., Virtanen, T., Zeiler, T., Kauppinen, J., Taivainen, A., and
   Mantyjarvi, R. (1999) *J Biol Chem* **274**, 2337-2343
46. Levitt, M., and Park, B. H. (1993) *Structure* **1**, 223-226
47. Harding, M. M. (2001) *Acta Crystallogr D* **57**, 401-411
48. Holm, L., and Sander, C. (1993) *J Mol Biol* **233**, 123-138
49. Bocskei, Z., Groom, C. R., Flower, D. R., Wright, C. E., Phillips, S. E., Cavaggioni, A.,
   Findlay, J. B., and North, A. C. (1992) *Nature* **360**, 186-188
50. Newcomer, M. E. (1993) *Structure* **1**, 7-18
51. Brownlow, S., Morais Cabral, J. H., Cooper, R., Flower, D. R., Yewdall, S. J., Polikarpov,
   I., North, A. C., and Sawyer, L. (1997) *Structure* **5**, 481-495
52. Wojnar, P., van't Hof, W., Merschak, P., Lechner, M., and Redl, B. (2001) *Biol Chem*
   **382**, 1515-1520
53. Bode, W., and Huber, R. (2000) *Biochim Biophys Acta* **1477**, 241-252
54. Bode, W., Engh, R., Musil, D., Thiele, U., Huber, R., Karshikov, A., Brzin, J., Kos, J., and Turk, V. (1988) EMBO J 7, 2593-2599
55. Jenko, S., Dolenc, I., Guncar, G., Dobersek, A., Podobnik, M., and Turk, D. (2003) J Mol Biol 326, 875-885
56. Stubbs, M. T., Laber, B., Bode, W., Huber, R., Jerala, R., Lenarcic, B., and Turk, V. (1990) EMBO J 9, 1939-1947
57. Gasymov, O. K., Abduragimov, A. R., Yusifov, T. N., and Glasgow, B. J. (2000) Protein Sci 9, 325-331
58. Goetz, D. H., Willie, S. T., Armen, R. S., Bratt, T., Borregaard, N., and Strong, R. K. (2000) Biochemistry 39, 1935-1941
59. Gasymov, O. K., Abduragimov, A. R., Yusifov, T. N., and Glasgow, B. J. (1998) Biochim Biophys Acta 1386, 145-156
60. Holzfeind, P., and Redl, B. (1994) Gene 139, 177-183
61. Lichtlen, P., Wang, Y., Belser, T., Georgiev, O., Certa, U., Sack, R., and Schaffner, W. (2001) Nucleic Acids Res 29, 1514-1523
62. Lalmanach, G., Hoebeke, J., Moreau, T., Brillard-Bourdet, M., Ferrer-Ditt Martino, M., Borras-Cuesta, F., and Gauthier, F. (1993) J Protein Chem 12, 23-31
63. Fuentes-Prior, P., Noeske-Jungblut, C., Donner, P., Schleuning, W. D., Huber, R., and Bode, W. (1997) Proc Natl Acad Sci U S A 94, 11845-11850
64. Filipek, R., Rzychon, M., Oleksy, A., Gruca, M., Dubin, A., Potempa, J., and Bochtler, M. (2003) J Biol Chem 278, 40959-40966
65. Baumann, U., Bauer, M., Letoffe, S., Delepelaire, P., and Wandersman, C. (1995) J Mol Biol 248, 653-661
66. Letoffe, S., Delepelaire, P., and Wandersman, C. (1989) Mol Microbiol 3, 79-86
67. Flower, D. R., North, A. C., and Sansom, C. E. (2000) Biochim Biophys Acta 1482, 9-24
68. Sack, R. A., Sathe, S., Beaton, A., Kozinski, M., Bogart, B., Lew, G., Sharma, S., and Upsoni, A. (2004) Exp Eye Res 78, 371-378
Table 1: Data collection statistics

| Data set       | Peak             | Remote          | Inflection      | Native          |
|----------------|------------------|-----------------|-----------------|-----------------|
| Space group    | C2               |                 |                 |                 |
| a [Å], b [Å], c [Å], β [°] | 75.1, 60.9, 50.4, 130.0 |                 |                 | 76.4, 61.4, 50.8, 129.7 |
| Beamline       | BM14/ESRF        |                 |                 | PX1/SLS         |
| Wavelength [Å] | 0.97888          | 0.88556         | 0.97912         | 0.97960         |
| Resolution [Å] | 1.95             | 1.71            | 2.36            | 1.80            |
| Unique reflections | 12650           | 18695           | 7139            | 16397           |
| Completeness [%] (resolution shell [Å]) | 99.6 (38.63-1.95) | 98.7 (30.43-1.71) | 98.7 (41.89-2.36) | 97.9 (39.22-1.80) |
| Anomalous completeness† | 98.6 (38.63-1.95) | 93.9 (30.43-1.71) | 93.9 (41.89-2.36) | —               |
| Anomalous fraction‡ | 98.9 (38.63-1.95) | 94.6 (30.43-1.71) | 94.3 (41.89-2.36) | —               |
| I/σ(I) (resolution shell [Å]) | 7.4 (38.63-1.95) | 8.1 (30.43-1.71) | 9.6 (41.89-2.36) | 5.1 (39.22-1.80) |
| Redundancy (resolution shell [Å]) | 6.9 (38.63-1.95) | 3.8 (30.43-1.71) | 4.3 (41.89-2.36) | 2.1 (39.22-1.80) |
| Mosaicity [°]  | 0.31             | 0.45            | 0.35            | 0.59            |
| R_{mean}§ (resolution shell [Å]) | 0.059 (38.63-1.95) | 0.059 (30.43-1.71) | 0.061 (41.89-2.36) | 0.092 (39.22-1.80) |

All parameters as defined in MOSFLM and SCALA.

† percentage of acentrics measured
‡ percentage of measured acentric reflections for which anomalous difference has been measured
§ redundancy-independent (multiplicity-weighted) R_{sym}, relative to I^+ or I^−,

\[ R_{sym} = \frac{\sum_h \sum_k |I(hk0)|^2 <I_{hk0}>^2}{\sum_h \sum_k |I(hk0)|^2} \left| \frac{1}{\sum_h \sum_k |I(hk0)|^2} \right|, \]

where i is over the different measurements of reflections.
Table 2: Refinement statistics

|                         |                  |
|-------------------------|------------------|
| Resolution [Å]          | 42.64-1.80       |
| R/R<sub>free</sub> (resolution shell) | 18.9/25.4 (42.64-1.80) |
|                         | 25.6/30.2 (1.85-1.80) |
| Protein residues present| 162              |
| Protein residues visible | 128              |
| No. of solvent molecules | 146              |
| No. of zinc ions        | 3                |
| No. of chloride ions    | 2                |
| R.m.s. deviation from ideality of bond lengths [Å] | 0.015             |
| R.m.s. deviation from ideality of bond angles [°] | 2.215             |
| Average B-values [Å²]:  |                  |
| protein                 | 33.0             |
| solvent                 | 48.3             |
| zinc                    | 29.5             |
| chloride                | 27.0             |
### Table 3. Interatomic distances of the zinc-binding sites

| Atom 1     | Atom 2            | Measured distance | „Target“ distance \(^a\) |
|------------|-------------------|-------------------|--------------------------|
| Zn H1      | Lys 76, N\(\zeta\) | 1.89 Å            | 2.00 Å                   |
|            | Glu 73, O\(\varepsilon_1\) | 2.02 Å            | 2.04 Å                   |
|            | W 55, O           | 2.19 Å            | 2.09 Å                   |
|            | Glu 102, O\(\varepsilon\) Sym. | n.d. \(^b\)     | 2.04 Å                   |
| Zn H2      | Glu 125, O\(\varepsilon_2\) | 2.01 Å            | 2.04 Å                   |
|            | His 96, N\(\varepsilon_2\) | 2.10 Å            | 2.00 Å                   |
|            | Cl H4             | 2.20 Å            | n.d. \(^b\)             |
|            | Cl H5             | 2.27 Å            | n.d. \(^b\)             |
| Zn H3      | Asp 129, O\(\delta_1\) | 2.08 Å            | 2.04 Å                   |
|            | His 92, N\(\delta_1\) | 2.21 Å            | 2.00 Å                   |
|            | W 54, O           | 2.41 Å            | 2.09 Å                   |
|            | W 56, O           | 1.79 Å            | 2.09 Å                   |

\(^a\) from (47)  
\(^b\) n.d.: not determined
FIGURE LEGENDS

Fig. 1. **Three-dimensional structure of human Tlc.** A, ribbon diagram, side view: Tlc exhibits the typical lipocalin fold with an eight-stranded antiparallel β-barrel forming a central cavity at its open (upper) end and an α-helix attached to it. The N- and C-termini of the polypeptide chain, the eight β-strands A to H, the extraneous β-strand I, and the four loops at the open end of the β-barrel are labeled. Loop #1 connecting β-strands A/B (Asp-25 to Met-31), loop #2 connecting β-strands C/D (Leu-56 and Ile-57), and loop #4 connecting β-strands G/H (Leu-105 and His-106) are partially disordered in the crystal structure; the modelled regions are depicted in grey. A short stretch with local 3_{10}-helical conformation at the tip of loop #3 connects β-strands E/F. B, complexation of three Zn (II) ions (coloured magenta), in conjunction with two chloride ions (green), by Tlc side chains (magenta). C, top view into the calyx of Tlc (the molecule shown in A rotated by ca. 90° about a horizontal axis): the β-barrel is ellipsoid rather than circular and loop #1 partially covers the cavity. D, surface representation of Tlc, in a roughly similar orientation as in C, including the modelled loop segments (coloured according to the electrostatic potential: positive charge, blue; negative charge, red). The opening of the calyx measures about 10 Å in diameter and is lined with negative charges, most prominently Glu-34 and Asp-80. A second entry to the large cavity, located between loops #1 and #4, is indicated by an arrow on the left.

Fig. 2. **Superposition between Tlc, Bda, and Blg.** A, stereoview of the superposition of Tlc (red, modelled regions in magenta) with Bda (yellow, PDB accession code 1BJ7), and Blg (blue, PDB accession code 1BE8). Structurally conserved β-barrel residues (as defined in B) are coloured in different shades of gray. R.m.s. deviations relative to Tlc for the 58 Cα positions of these β-barrel residues are 1.5 Å and 1.7 Å for Bda and Blg, respectively. Loop #1 of Tlc, connecting β-strands A/B, is by two and three residues shorter than the corresponding loops of Bda and Blg, respectively. B, sequence alignment of Tlc, Bda, and
Blg. All three amino acid sequences were aligned based on structural similarities. Equivalent β-barrel residues (as defined for Bda and Blg in (16)) that were used for superposition are boxed. Secondary structure elements are indicated as interpreted by DSSP (39): β-strands A to I are depicted with green background, α-helices with orange background. In case of Tlc β-strands and α-helices are also marked as green arrows and orange cylinders, respectively. Residues His-1 to Leu-4 of Tlc, which were not present in the recombinant protein, as well as the Cys-101→Ser replacement and the C-terminal Strep-tag II are shown as lower-case characters. Residue Asp-158 of native Tlc was substituted by Ala as linker to the Strep-tag II.

Fig. 3. Illustration of the Tlc cavity and apparent shape complementarity with cognate ligands. A, inner accessible surface of the cavity, coloured according to the electrostatic potential (positive charge, blue; negative charge, red), with the polypeptide backbone shown in grey. The ligand pocket is about 15 Å deep, bifurcated, and positively charged at the bottom. The large cavity is lined with basic and hydrophobic residues, most notably Met-39 and Phe-99, which separate the two lobes. B, the antibiotic drug rifampin (stick model) can be nicely docked into the cavity (here shown as translucent inner surface) and almost fully occupies the pocket. C and D, retinoic acid docked into the cavity in two different orientations. While the β-ionone ring of the ligand resides approximately at the same position within one of the lobes at the bottom, its aliphatic side chain with the terminal carboxylate group can take two different directions, one pointing toward the small entry between strands A and H (C) and one pointing toward the central opening of the β-barrel (D).

Fig. 4. Structural comparison between Tlc and proteinase inhibitors. A, the first ordered residues at the N-terminus of Tlc, Asp-12 to Tyr-18, with 2F_o-F_c electron density map at 1.8 Å resolution contoured at 1.0 σ. B, superposition of the inhibitory motif at the N-terminus of Tlc (coloured red) with the inhibitory sites of chicken egg-white cystatin (PDB accession code
Crystal structure of tear lipocalin

Breustedt et al., p. 33

1CEW), stefin A (PDB accession code 1NB5), and stefin B (PDB accession code 1STF), all depicted in grey. Only five (Asp-12 to Gly-15 and Trp-17) of altogether ten residues that comprise the three inhibitory motifs of Tlc are structurally defined in the present crystal structure and were used for the comparison. In the three proteinase inhibitors the side chain orientations of Gln-53 to Gly-57 (same residue numbers in all three molecules) as well as of residue Trp/Gly/His-104 are closely similar while the presumed inhibitory peptide segment of Tlc clearly does not adopt a cystatin-like pattern. C, superposition of the N-terminus of Tlc, residues Val-13 to Tyr-18 (coloured red), with Phe-21 to Tyr-26 of triabin (yellow), Phe-27 to Ser-32 of staphostatin B (green), and Leu-11 to Val-16 of Inh (blue), i.e. the segments coloured light grey in panel D. The Gly-X-Trp motif of residues 15 to 17 of Tlc, which is conserved in the β-barrel structures of lipocalins, is also present in the three proteinase inhibitors. Yet, it is not involved in their differing inhibitory mechanisms. D, crystal structures of the three known proteinase inhibitors with an eight-stranded β-barrel fold: triabin (PDB accession code 1AVG, yellow), staphostatin B (PDB accession code 1PXV, green), and Inh of Erwinia chrysanthemi (PDB accession code 1SMP, blue). Side chains of residues interacting with the respective target proteinase are depicted in a ball-and-stick representation in dark grey. The conserved stretch of six residues close to the bottom of each β-barrel, shown in greater detail in panel C, is coloured light grey.
Breustedt et al.; Fig. 1
The 1.8 Å crystal structure of human tear lipocalin reveals an extended branched cavity with capacity for multiple ligands
Daniel A. Breustedt, Ingo P. Korndoefer, Bernhard Redl and Arne Skerra

J. Biol. Chem. published online October 15, 2004

Access the most updated version of this article at doi: 10.1074/jbc.M410466200

Alerts:
• When this article is cited
• When a correction for this article is posted

Click here to choose from all of JBC's e-mail alerts