# 4 RESULTS

## 4.1 SapC

### 4.1.1 Expression and purification of SapC

To overcome limitations of other expression systems such as missing glycosylation (*E. coli*), poor yield (natural sources), or high expenses (chemical synthesis), human saposin C was overexpressed in the methylotrophic yeast *P. pastoris.* The expression construct included a C-terminal $(His)_6$-tag which was not cleaved off prior to crystallization. It has been shown previously that human saposin C expressed with a C-terminal $(His)_6$-tag is fully functional as assayed by binding to phosphatidylserine and activation of β-glucocerebrosidase (Qi and Grabowski, 2001; Qi et al., 1994). As compared to *E. coli,* overexpression in *P. pastoris* offers the additional benefit of extremely facilitated downstream purification since the cells secrete practically only the overexpressed proteins. Thus, saposin C could be directly captured from the medium. Protein yields of human saposin C averaged ~ 40 mg/l medium which compares favorably with typical amounts obtained from *E. coli* expression.

The expectation that overexpression in yeast cells also results in glucosylation of Asn22 was not fulfilled. Analysing the sample by MALDI-TOF revealed a peak at 10,252.2 Da agreeing well with the length of the cloned sequence including the $(His)_6$-tag and the additional arginine residue at the C-terminus and tyrosine and valine residues at the N-terminus but without glucosylation. Glucosylation of SapC, however, is neither essential for binding to phospholipids nor for activation of lysosomal hydrolases (Hiraiwa et al., 1993b; Vielhaber et al., 1996). Four more peaks were found at masses incrementally decreasing by 137 Da indicating that the histidyl residues of the $(His)_6$-tag were partially cleaved off.

### 4.1.2 CD spectroscopy of SapC

The thermostability of SapC and the secondary structure content were investigated by CD spectroscopy. Figure 13 shows the far-UV circular dichroism (CD) spectrum of purified SapC at 25° C, 95° C, and after cooling to 25° C. The secondary structure estimation calculated using program CDSSTR (Sreerama and Woody, 2000) is presented in Table 1.
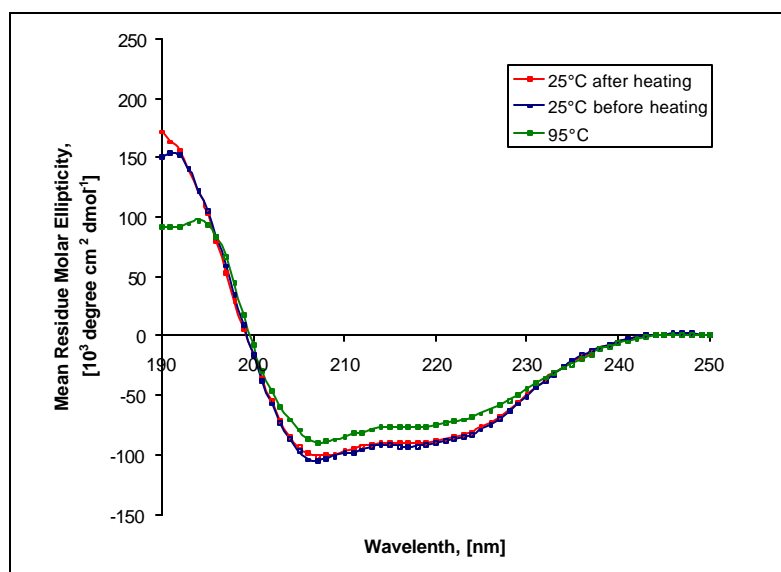
**Figure 13. Far-UV spectra of SapC at different temperatures.** Upon heating from 25° C to 95° C the total intensity of the spectrum slightly decreased and the original spectrum was obtained after cooling to 25° C.

**Table 1.** Relative secondary structure content (%) of SapC estimated by CDSSTR at different temperatures.

| Temperature, °C | Helix | Strand | Turn | Not ordered |
|---|---|---|---|---|
| 25 (before the heating step) | 84 (79)* | 6 | 6 | 5 |
| 95 | 79 | 5 | 7 | 9 |
| 25 (after the heating step) | 82 | 6 | 6 | 7 |

*Data in parentheses were derived from the crystal structure.

The CD spectrum with the negative maximum band at 208 nm and a shoulder at about 222 nm is typical for predominantly $\alpha$-helical proteins. Estimation of the $\alpha$-helical content using program CDSSTR yielded values of about 85%, in good agreement with the X-ray diffraction results given in the Table 1. Heating of SapC to 95°C decreased the intensity of the spectrum (Fig. 13) at 208 nm by about 20%, whereas the intensity of the spectrum as well as the SapC's secondary structure content at 95° C was calculated to be only slightly changed compared to the initial

data at 25° (Fig. 13 and Table 1). Both effects were reversible indicating the high thermostability of SapC.
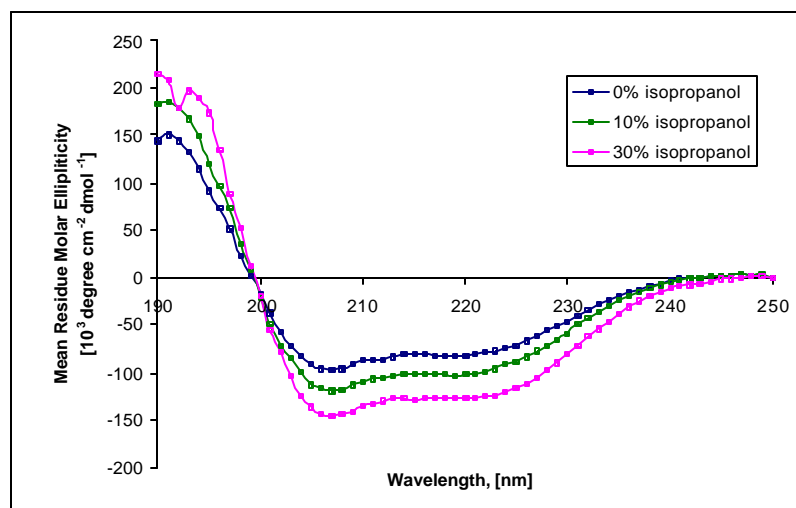


**Figure 14. Far-UV spectra of SapC in the presence of isopropanol.** Isopropanol increases the total intensity of the spectrum.

**Table 2.** Secondary structure content (%) of SapC estimated by CDSSTR at different isopropanol concentrations.

| Isopropanol, % | Helix | Strand | Turn | Not ordered |
|----------------|-------|--------|------|-------------|
| 0 | 82 | 4 | 6 | 7 |
| 10 | 84 | 5 | 5 | 7 |
| 30 | 80 | 8 | 5 | 8 |

To further investigate the flexibility of SapC, CD spectra were measured in the presence of isopropanol mimicking a hydrophobic environment and the helix promoter, 2,2,2-trifluoroethanol (TFE). Deconvolution results were obtained using the program CDSSTR and are presented in Tables 2 and 3. TFE seems to have a similar impact on the secondary structure of SapC as isopropanol. Notably, no significant changes in the secondary structure content were observed, although the

trifluorethanol is a helix promoter (Figs. 14-15, Tabs. 2-3). Only the total intensity of the spectrum was increased in the presence of both, isopropanol and TFE. Without changes in the secondary structure this effect could be probably explained by the solubilization of the SapC. SapC seems to aggregate or make oligomers in water at pH 4.5 whereas the oligomer formation causes some light-scattering effects. The oligomerization of SapC is decreased in the presence of TFE or isopropanol.
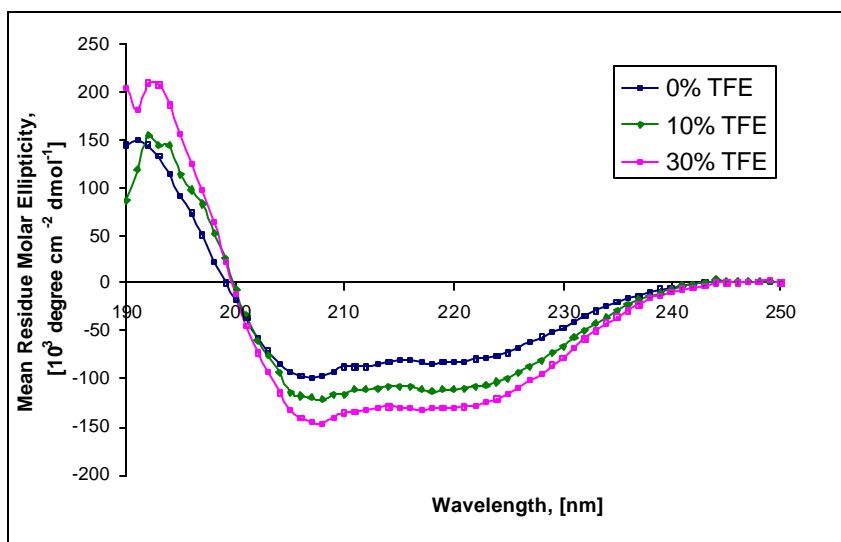


**Figure 15. Far-UV spectra of SapC in the presence of 2,2,2-trifluoroethanol (TFE).** TFE has the same effect on the CD spectra as isopropanol. i. e. it increases the intensity of the spectrum.

**Table 3.** Relative secondary structure content (%) of SapC estimated by CDSSTR at different TFE concentrations.

| Trifluoroethanol, % | Helix | Strand | Turn | Not ordered |
|---|---|---|---|---|
| 0 | 82 | 4 | 6 | 7 |
| 30 | 71 | 7 | 9 | 14 |
| 60 | 81 | 6 | 7 | 7 |

To investigate the effect of phospholipids on the SapC structure, CD spectra in the presence of phosphatidylserine (Sigma) were measured at molar protein:lipid proportion 1:10. According to the deconvolution made by CDSSTR, phosphatidylserine increased the percentage of $\alpha$-helical structure (Fig. 16, Table 4), whereas the intensity of the spectrum at 208 nm decreased by 50%.
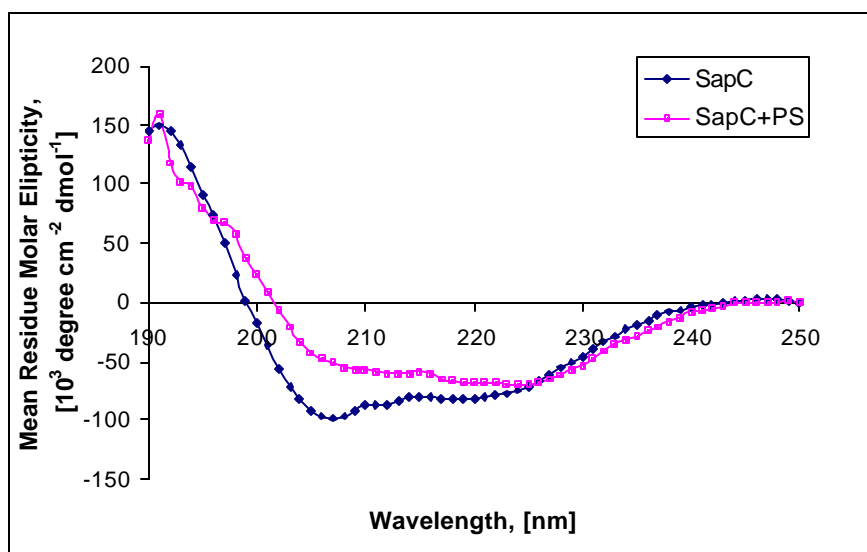


**Figure 16. Far-UV spectra of SapC in the presence of phosphatidylserine.**

**Table 4.** Relative secondary structure content (%) of SapC in the presence of phosphatidylserine estimated by CDSSTR.

| Phosphatidylserine, mmol | Helix | Strand | Turn | Not ordered |
|---|---|---|---|---|
| 0 | 82 | 4 | 6 | 7 |
| 0.15 | 92 | 7 | 0 | 0 |

The CD spectra indicated that the secondary structure elements of SapC are stable in different environments, although phospholipids were able to increase the helix content. SapC seems to form oligomers in water at acidic pH, and oligomerization shows light-absorption-related effects that decrease the spectrum's intensity, an

effect also observed by CD spectroscopy for other aggregating proteins (Arutyunyan et al., 2001). Organic solvents increase the CD spectrum intensity probably by solubilization of SapC.

### 4.1.3 Crystallization of SapC

Crystallization of saposin C was achieved by vapour diffusion using either the hanging or the sitting-drop method. Orthorhombic crystals grew in the presence of pentaerythritol ethoxylate 15/4. Typical crystals obtained from these conditions are shown in Fig.17a. The crystal dimensions were approximately 500 μm x 300 μm x 200 μm. Tetragonal crystals of SapC were obtained using the streak-seeding technique and are shown in Fig. 17b. The needle-like crystals reached a maximum length of 300 μm.
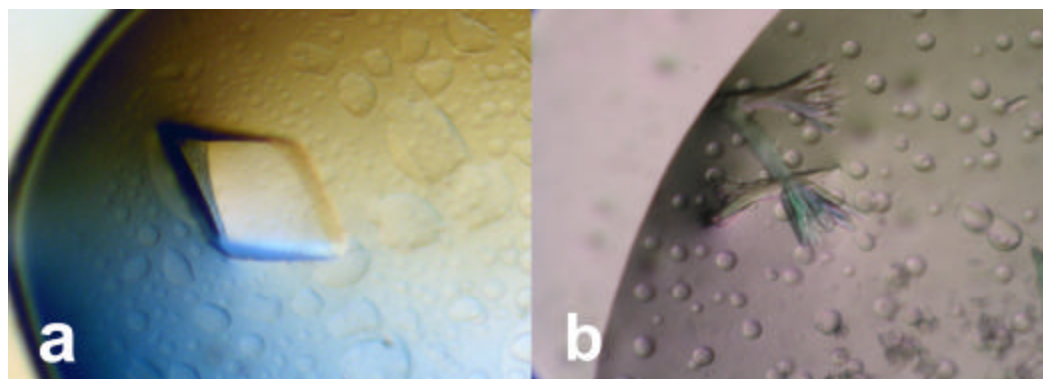


**Figure 17. SapC crystals. (a)** Orthorhombic SapC crystals with a maximum length of 500 μm **(b)** Tetragonal crystal of SapC with the length of 200 μm in the longest direction.

The orthorhombic crystal was measured at room temperature on a rotating anode source. The obtained cell dimensions were a=57.0 Å, b=88.3 Å and c=93.5 Å. The space group could be unambiguously determined as $C222_1$ from systematic extinctions.

X-ray diffraction data of tetragonal crystals of SapC were measured at ESRF, Grenoble (ID14-2). Using systematic absences the space group could be determined to be either $P4_12_12$ or $P4_32_12$ with two molecules in the asymmetric

unit and featured unit-cell dimensions of a=48.9 Å and c=154.3 Å. Data collection statistics of the two different crystal forms are compiled in Table 5.

**Table 5.** Data collection statistics of the two different SapC crystal forms.

|  | Orthorhombic SapC | Tetragonal SapC |
|---|---|---|
| Wavelength, (Å) | 1.54179 | 0.93300 |
| Space group | C $222_1$ | P $4_12_12$ |
| Unit-cell parameters, (Å) | a=57.0, b=88.3, c=93.5 | a=48.9, c=154.3 |
| Number of molecules in a.u. | 2 | 2 |
| Resolution, (Å) [*] | 40.0-2.45 (2.54-2.45) | 30.0-2.5 (2.59-2.5) |
| Mosaicity, (°) | 0.45 | 0.90 |
| Reflections |  |  |
| Total | 25914 | 45322 |
| Unique | 8953 | 7887 |
| Completeness, (%) | 98.9 (99.8) | 99.6 (99.6) |
| I/s(I) overall | 16.9 (2.1) | 13.0 (2.8) |
| $R_{sym}$ overall | 5.5 (46.5) | 8.3 (49.0) |

[*]Data in parentheses refer to highest resolution shell

### 4.1.4 Determination of the structure of SapC

The structure of tetragonal SapC in space group P $4_12_12$ was determined by molecular replacement with MOLREP (Vagin and Teplyakov, 1997) using data collected from tetragonal crystals (Schultz-Heienbrok et al., 2006) and searching with a poly-Ser model of SapC monomer in closed conformation. The poly-Serine model of SapC was generated by program MOLEMAN (1992-2004), Uppsala University, Uppsala, Sweden, unpublished program.) from PDB ID 2GTG (Ahn et al., 2006). The initial structure of SapC obtained by molecular replacement was further improved and refined with ARP/wARP (Perrakis et al., 1999) and multiple rounds of manual rebuilding with COOT (Emsley and Cowtan, 2004) followed by restrained refinement with REFMAC (Murshudov et al., 1997), CNS (Brunger et al., 1998), and PHENIX

(Adams et al., 2002). Solvent molecules were added using CNS. Since orthorhombic and tetragonal SapC crystallized under similar conditions, the refined model of the tetragonal SapC dimer in open conformation was used as search model in molecular replacement with MOLREP to determine the structure of SapC in the orthorhombic space group. The structure was refined as described for tetragonal SapC. Structure validation was performed using program PROCHECK (Vaguine et al., 1999). Details of refinement statistics are presented in Table 6.

**Table 6.** Refinement statistics of the two different SapC crystal forms.

|  | Orthorhombic SapC | Tetragonal SapC |
|---|---|---|
| Resolution (Å) | 40 -2.45 (2.45 - 2.51)* | 25 – 2.50 (2.56 – 2.50) |
| Unique reflections | 7876 | 7087 |
| R-factor (%) | 22.8 (30.2) | 25.9 (24.5) |
| $R_{free}$ factor (%) | 27.2 (32.3) | 28.5 (24.7) |
| Number of atoms |  |  |
| Protein | 1236 | 1212 |
| Ligand | - | 12 |
| Solvent | - | 29 |
| Average B-factors($Å^2$) |  |  |
| Protein | 70.1 | 46.4 |
| Ligand | - | 39.3 |
| Solvent | - | 46.4 |
| R.m.s. deviations |  |  |
| Bond lengths (Å) | 0.013 | 0.014 |
| Bond angles (°) | 1.43 | 1.49 |
| Ramachandran plot |  |  |
| Most favoured (%) | 94.6 | 95.2 |
| Additionally allowed (%) | 4.0 | 4.8 |
| Generously allowed (%) | 0 | 0 |
| Disallowed (%) | 0.7 | 0 |

*Data in parentheses refer to highest-resolution shell

## 4.1.5   Overall structure of SapC

SapC crystallized in two crystal forms – tetragonal and orthorhombic – from the precipitant pentaerythritol ethoxylate 15/4 using different pH and salts.
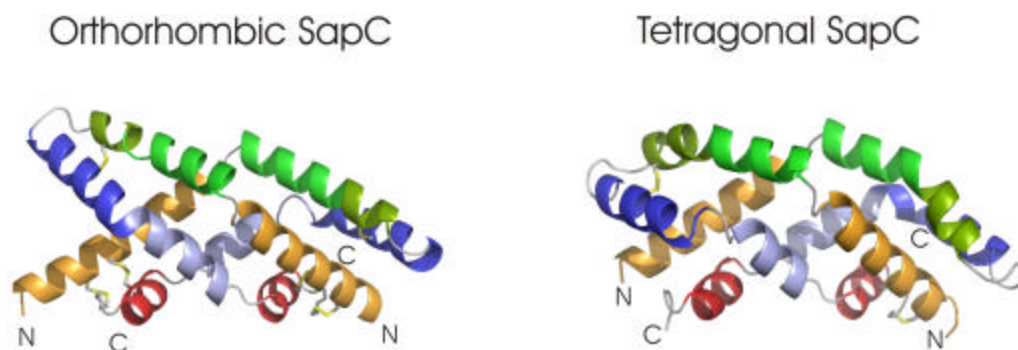


**Figure 18. Overall structure of SapC dimers.** The crystal structures of orthorhombic and tetragonal SapC were determined by molecular replacement at 2.45 Å and 2.50 Å resolution, respectively. Helices are coloured as in Fig 19c.

The crystal structures of tetragonal and orthorhombic SapC were determined by molecular replacement at 2.50 Å and 2.45 Å, respectively. Tetragonal SapC forms a homodimer with domain-swapped boomerang-shaped monomers in an extended open conformation that was also found in the orthorhombic space group (Fig. 18). SapC monomers form a distorted four-helix bundle. Helices a1 (residues 2-18), a2 (residues 25-38), a3 (residues 43-62), and a4 (residues 68-76) (Fig. 19) interact in a tight interface of hydrophobic interactions with contributions from a total of 39 out of the 60 helical residues.

Helix a2 is subdivided into helices a2´ (residues 25-32) and a2´´ (residues 33-38), and a kink divides helix a3 in a3´ (residues 43-54) and a3´´ (residues 55-62). The N- and C-terminal helices a1 and a4 are cross-linked by disulfide bridges as in other saposins.
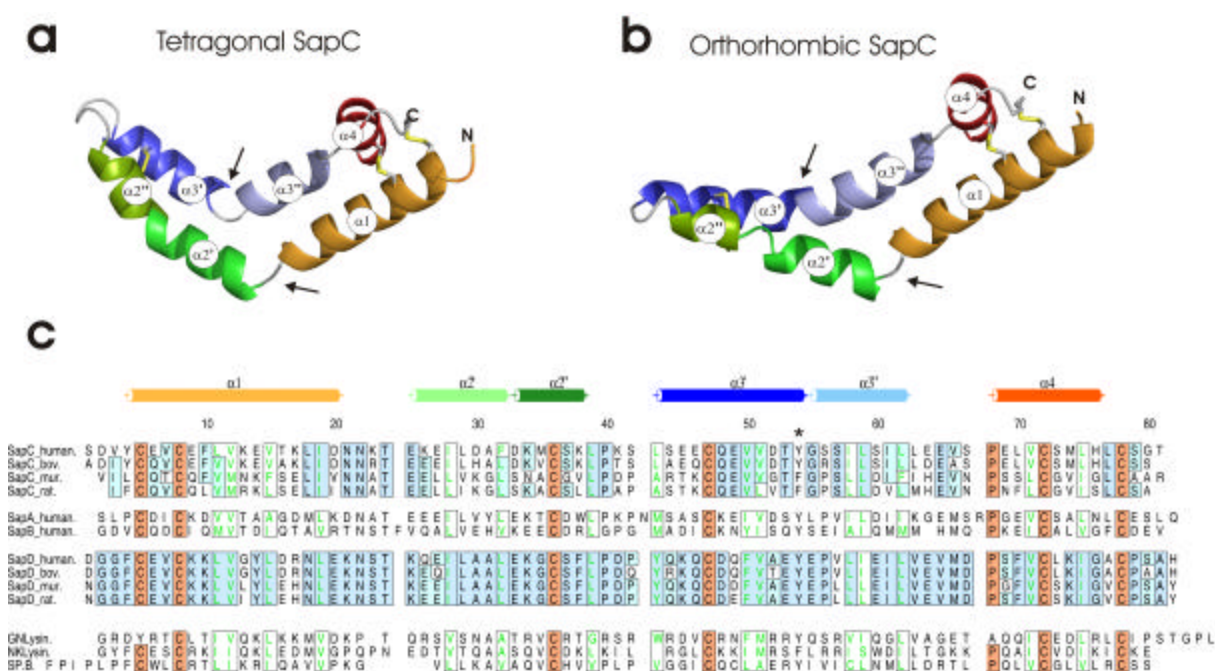
**Figure 19. Overall structure of SapC monomers.** *(a, b,)* **Ribbon representation of SapC monomers.** Helices coloured as in (*c*). Hinges between helices a1/a2' and a3'/a3" are indicated by arrows. Tyr54 with peptide torsion angles of f, ? at -123°, -15° for tetragonal SapC, -101°, -19° for orthorhombic SapC (Table 6) is located at the hinge a3'/a3". *(c)* **Sequence alignment of saposins and SAPLIPS generated with ALSCRIPT (Barton, 1993)**. Positions of a-helices a1 to a4 are shown on top. Conserved Cys and Pro are shaded orange, other conserved amino acids are shaded blue, type-conserved hydrophobic residues are shaded white and boxed. Conserved Tyr54/Phe54 is marked (*).

In both crystal forms, the two monomers of the domain swapped dimers adopt a virtually identical conformation, with r.m.s. deviations for the Ca atoms over residues 2-78 at 0.9 Å and 1.3 Å for the tetragonal and orthorhombic crystal form, respectively. However, between the two crystal forms the monomers differ in their torsion and opening angle (Fig. 20c), measured between helix pairs a1/a4 and a2/a3. Helix pairs a1/a4 and a2/a3 in orthorhombic SapC are oriented more parallel to each other than in the tetragonal form because the opening angle is wider (116º) (Figs. 19a-b, 20).The configuration of helix pair a1/a4 is comparable in both SapC, whereas helix a3 in orthorhombic SapC is less bent than in tetragonal SapC where it is broken (Figs. 19a-b, 20).
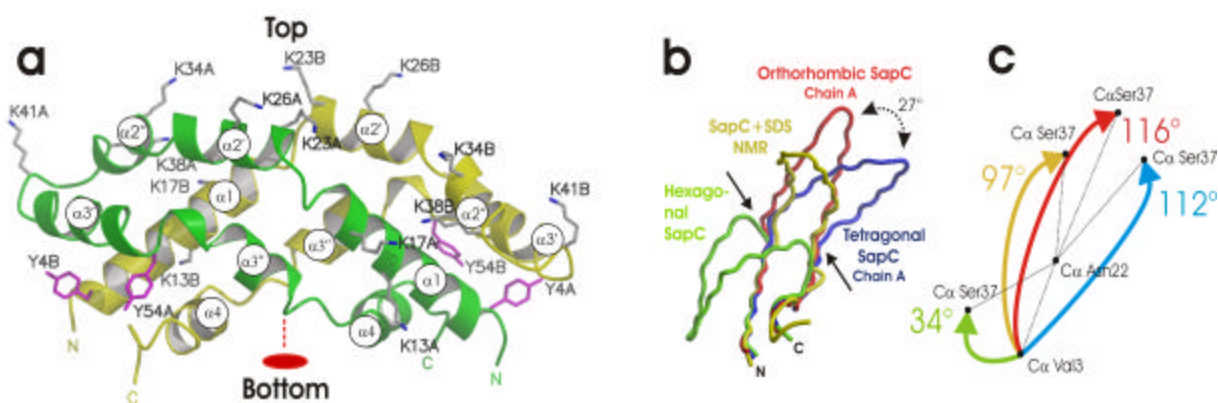
**Figure 20.** *(a)* **SapC dimer formation**. In the SapC dimer, domains a1/a2'/a2" and a3'/a3"/a4 are swapped, monomers intertwine to form a dimer, and the red ellipse indicates a local *C2* axis. *(b)* **Conformational flexibility of SapC**. Superimposition of the four SapC molecules, whereas atoms of residues 2-19 were used for superimposition (SapC+SDS: PDB 1SN6, hexagonal SapC: PDB 2GTG, tetragonal and orthorhombic SapC: this work). SapC undergoes remarkable bending at hinges (arrows) that transform it from a compact closed configuration with hydrophilic exterior (hexagonal SapC) to an open configuration exposing hydrophobic residues suitable to interact with lipids and membranes. **(c)** **Schematic presentation of the hinge-bending motion of SapC**. The hinge is located between a1 and a2' at Asn22. The angle is measured between Ca atoms of Val3, Asn22, and Ser37.

The dimerization interfaces of SapC bury 1721 $Å^2$ and 1965 $Å^2$ surface area per monomer, respectively. The most hydrophobic residues in the helical parts are involved in the formation of a tight dimer interface, resembling the packing of the helical bundle in the common closed saposin helical bundle. In tetragonal SapC, 35 residues on either molecule form the dimer through more than 100 intermolecular contacts, of which 51% are hydrophobic. The hairpin of molecule A in orthorhombic SapC is less strongly involved in crystal contacts than in the tetragonal space group and shows higher flexibility. Here, the dimer is formed by only 56 intermolecular contacts, although the amount of hydrophobic contacts remains about the same as in the tetragonal space group, stressing the hydrophobic nature of the hinge region.

The "top" of the SapC dimer is enriched in polar residues (Fig. 21), and positively charged lysines are located mostly on the "top" and "sides" of the dimer. The "bottom" surface of the dimer is hydrophobic and contains Tyr4 and Tyr54 (Fig. 20a), that act as gatekeepers of the hydrophobic inner cavity.
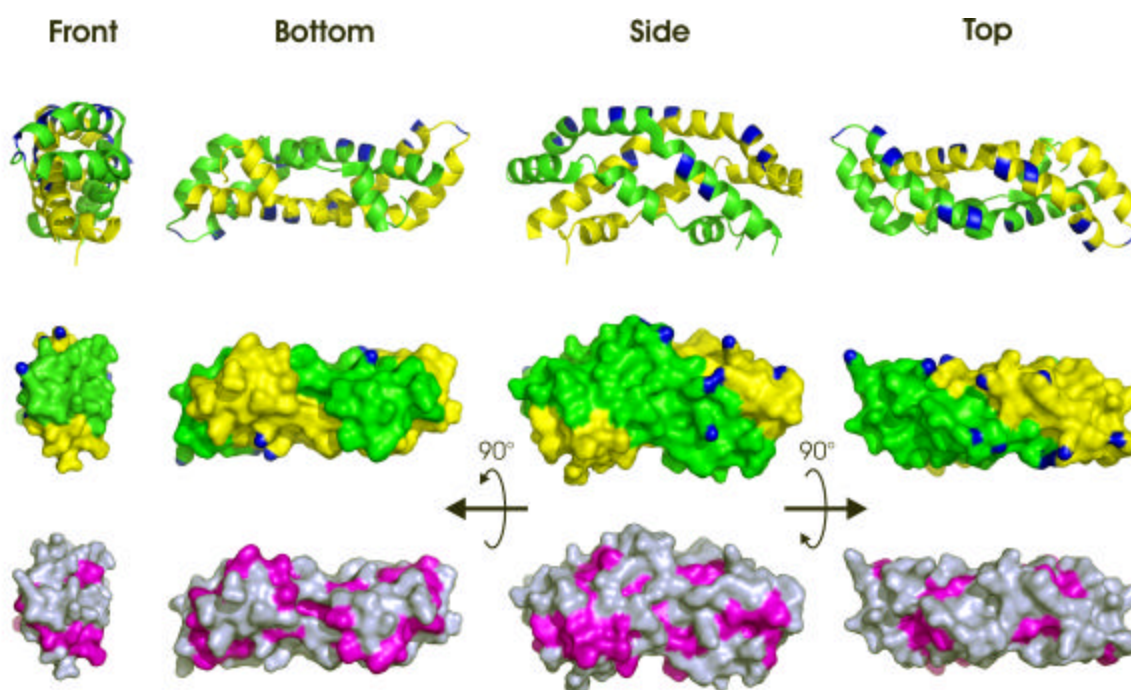
**Figure 21. Overall structure of SapC dimers in surface presentation.** Surface representation of the SapC dimers. Monomers are coloured yellow and green, respectively, positively charged lysines and arginines are coloured blue, hydrophilic amino acids are in grey, hydrophobic and aromatic amino acids and prolines in magenta. In the SapC dimer six of seven lysines are located on the surface on helices a1, a2' and a2'' and form the largest coherent hydrophilic surface ("top").

## 4.1.6 Conformational flexibility and oligomerization of SapC

Two conformations have been described previously for saposins: the canonical substrate-free closed 4-helix bundle and the V-shaped, ligand-bound conformations in the presence of lipids or SDS as observed for SapB and SapC, respectively. While the hydrophobic insides of SapB and SapC form the substrate-binding cavity in the open conformations, in the here-reported SapC structure they are involved in dimer formation in the domain-swapped monomers in the SapC dimer. This interface involves regions of the structure that are at the tips of both legs and separated farthest in the V-shaped or extended open conformations of other saposins. Thus, this contact must be disturbed upon ligand-induced opening of SapC, suggesting that the dimer represents a storage form of SapC that is stabilized against unspecific substrate binding by homophilic interaction.

The angle between helix pairs (legs) a1/a4 and a2/a3 is 97° in the SapC/SDS structure and widened to 112° and 116° in the orthorhombic and tetragonal crystal forms, respectively, of unliganded SapC presented here (Figs. 20 b,c and 22 ). This hinge-bending motion demonstrates an exceptional level of conformational flexibility that appears to be restrained only by the three disulfides, which may well be required for the interaction of SapC with membranes at various stages.
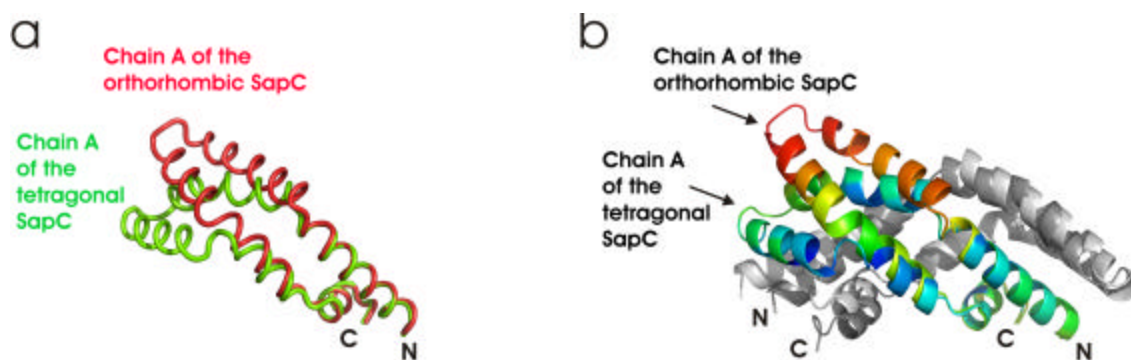


**Figure 22. SapC flexibility.** *(a)* Superimposition of A monomers of the tetragonal (green) and orthorhombic (red) SapC structures. Atoms of residues 2-19 of chains A were used for superimposition of the SapC molecules. Chains B are not shown. *(b)* Superimposition of SapC dimers. All atoms of residues 219 of chains A were used for superimposition of the SapC monomers. Both superimposed chains A are coloured according to their B-factor values. Chains B are coloured grey. Red and blue colouring indicates high and low B factors, respectively.

### 4.1.7  SAXS studies on SapC in solution

Bovine serum albumin (BSA, a 66.4 kDa monomeric protein), which is one of the most common standards for estimating the molecular mass of a studied protein, was used to evaluate the molecular mass of SapC by comparing the extrapolated forward scattering intensity I(0) of SapC with that of BSA in SAXS measurements. Unexpectedly, the molecular mass of SapC in solution at pH 4.5 was estimated to be ~120 kDa, indicating that SapC occurs as dodecameric protein in solution with ~10 kDa (90 residues, including a His-tag on the N-terminus) per monomer. The radius of gyration $R_g$ was determined to be ~88 Å using the Guinier approximation. $D_{max}$ was found to be 350 Å and thus about 10 times higher than the size of SapC monomer (25 Å) or dimer (55 Å) accordingly to their crystal structures. The shape of the curve of

the pair distribution function in real space P(r) corresponds to an anisometric particle (Fig. 23, inset) (Svergun, 1992).
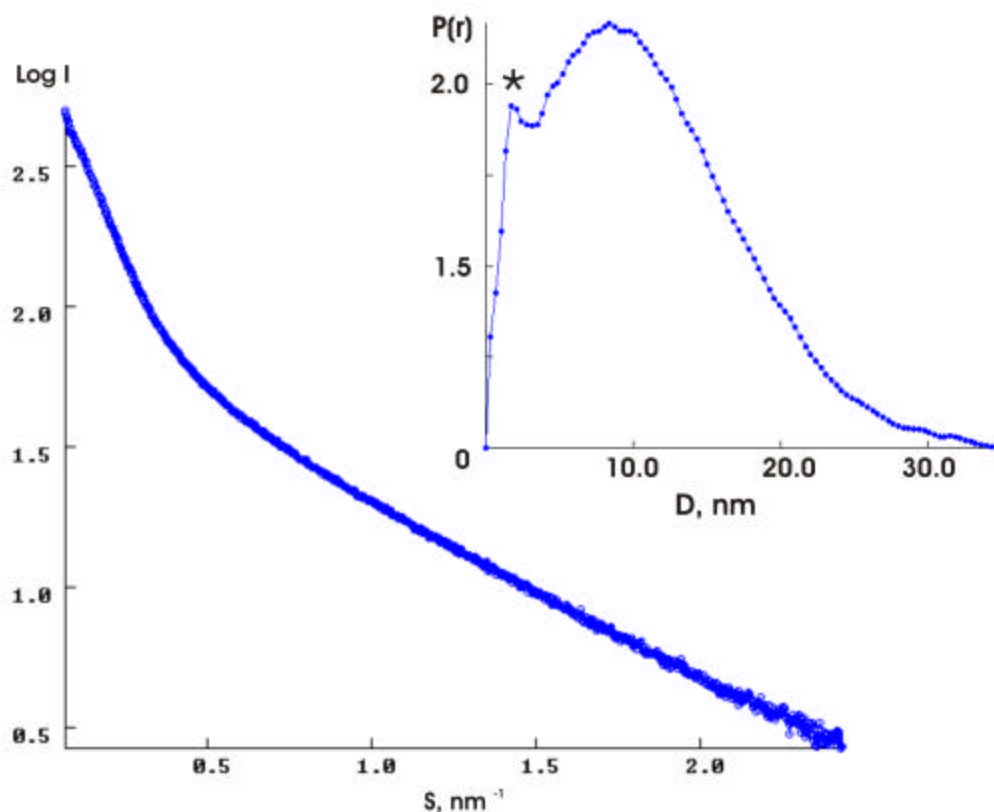


**Figure 23.** Experimental X-ray scattering pattern from SapC in solution at pH 4.5. Inset: pair distribution function of the experimental SAXS data. The maximum interatomic distance ($D_{max}$) is 35 nm or 350 Å, and the radius of gyration ($R_g$) is ~8.8 nm or 88 Å. Peak observed at 3 nm seems to correspond to the SapC monomer.

The excluded volume $V_p$ of the particle was computed to be 750 nm$^3$. For globular proteins, this value corresponds to a molecular mass of more than 100 kDa (personal note of Dr. Petukhov) and corroborates the formation of SapC oligomers.

### 4.1.8 Initial interactions of saposins with lipid head groups

Among the various SAPLIPs, the biological functions and details of interactions with membrane components are diverse. The most detailed model for initial interaction of SAPLIPs with membranes has been proposed on the basis of sulfate binding sites in crystalline granulysin (Anderson et al., 2003) and of biophysical studies on NK-Lysin

(Gutsmann et al., 2003). It involves the binding of a patch of positively charged amino acids on granulysin to a series of negatively charged phospholipid head groups. In agreement with this mechanism, the general importance of negatively charged lipids has been recognized for saposin-membrane interactions (Ciaffoni et al., 2001; de Alba et al., 2003; Liu et al., 2005). Of special importance is the anionic lipid bis(monoacyl)glycerophosphate (BMP), the marker lipid of intralysosomal vesicles (Möbius et al., 2003) where the physiological degradation of glycosphingolipid takes place (Fürst and Sandhoff, 1992).
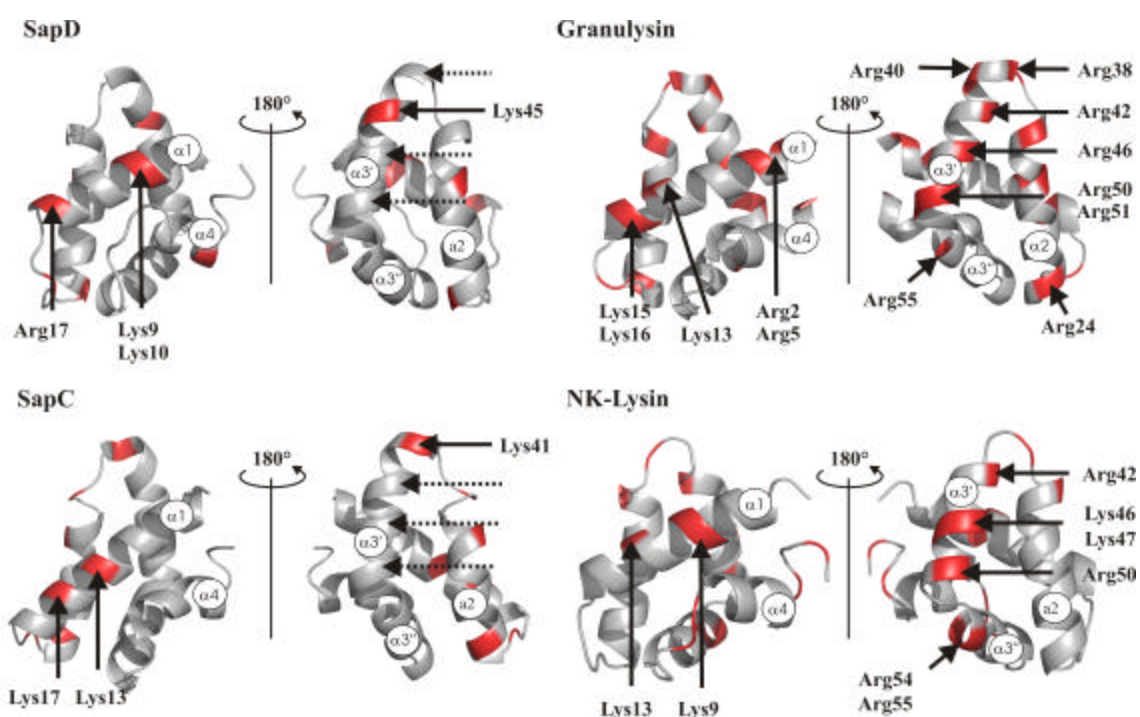


**Figure 24. Distribution of the positively charged amino acids on the surface of SAPLIPs.** Positions of positively charged residues in SapD, SapC, and homologous proteins granulysin and NK-Lysin are shown in red. The arrows indicate the presence of positively charged residues that are probably involved in binding phospholipid head groups. Dashed arrows show the lack of basic residues in SapD and SapC. In both saposins as well as in the homologous proteins, several positively charged amino acids are always located on helix a1. This contrasts granulysin and NK-Lysin where helix a3' and the N-terminus of helix a3'' contain a series of positively charged arginines and lysines while SapD and SapC contain only one Lys45 and Lys41, respectively, on helix a3', and there are no basic residues on helix a3''.

Positively charged surface clusters on the protein have been proposed to serve as initial membrane contacts for saposins (Liu et al., 2005; Qi and Chu, 2004). However, the crystal structure of SapC shows that the granulysin mechanism cannot apply for SapC, because the patch of positively charged amino acids for initial interaction with phospho- or sulfolipids is conserved between granulysin and NK-Lysin, but not in SapC (Fig. 24). In turn, positively charged lysines probably interacting with negatively charged membranes are neither conserved in granulysin nor in NK-Lysin. The anion-binding patch of positively charged residues in granulysin (Arg 38, 42, 46, 50, 51) and NK-Lysin (Arg 42, 46, 47, 50, 54, 55) is located on helix a3', oriented towards a2' and a2'' and towards the kink in helix a3 next at conserved Tyr54. In contrast, lysines in SapC point towards the loop between helices a2 and a1 at the tip of one of the legs of the open conformation of saposins. Notably, both sulfate- or phosphate-binding patches are orientated towards regions with particular importance for the hinge-bending motion, either towards the central hinge around Tyr54 in granulysin or towards the tip of one of the two legs, where helices separate farthest during opening.

Only three polar residues that are not cysteines are conserved between the four saposins: Asn22, Thr24, and Tyr54/Phe54 (Fig. 19c), and the conservation of the first two residues is only due to the presence of glycosylation sites N-X-T in prosaposin. As aromatic Tyr and Phe side chains were found to be involved in glycolipid binding in diverse proteins (Mahfoud et al., 2002; Malinina et al., 2004; Wright et al., 2005), they consequently are also candidates for glycolipid recognition centres in saposins. In fact, the solvent accessible and conserved Tyr54 and Phe4 are adjoined and seal the buried hydrophobic inside of SapC drawn in red in Fig. 20a. Conformational flexibility at the opening to a hydrophobic cavity has been suspected to be important for the recognition and extraction of lipids by saposins and other lipid-transport proteins (Ahn et al., 2003; Niere et al., 2001; Wright et al., 2005), and the kinking of helix a3 at the conserved residue Tyr54 was proposed to be a key for the plasticity of saposins (Ahn et al., 2006). The conservation of this kink in the novel extended open conformation of SapC (Table 7) emphasizes its importance.

**Table 7. Torsion angles $f$, $j$ of conserved Tyr54 (bold font) and adjacent two amino acids up- and downstream in SapA, B, C, and D.** Torsion angles $\phi$, $\varphi$ at positions 52, 55 and 56 are characteristic of α-helices. The torsion angles of residue Tys54 of SapB complexed with phospholipids differ from the values found in saposins without lipids. At position 54 there is a helix kink (indicated in black bold font) in all saposins except in chain A of SapB complexed with lipids (indicated in bold red font). In tetragonal SapC and SapC in closed conformation there is an additional helix break at position 53 (indicated in italic type). In orthorhombic SapC the helix at Thr53 is intact indicating the high flexibility of SapC at this position.

| Position | 52 | 53 | 54 | 55 | 56 |
|---|---|---|---|---|---|
| SapC (chain A), tetragonal. | (-90, 11) | *(-124, -42)* | **(-123, -15)** | (-45, -44) | (-80, -2) |
| SapC (chain A), orthorhombic | (-66, -41) | (-74, -47) | **(-101, -19)** | (-58, -50) | (-55, -47) |
| SapC [b] | (-63, -35) | *(-103, -34)* | **(-109, -12)** | (-60, -60) | (-70, -30) |
| SapD triclinic (chain A)[e] | (-65, -33) | (-88, -30) | **(-112, 2)** | (-51, -48) | (-54, -52) |
| SapA [c] | (-62, -42) | (-75, -36) | **(-109, -20)** | (-57, -47) | (-59, -39) |
| SapB (chain A) [d] (with lipids) | (-63, -44) | (-70, -42) | **(-64, -39)** | (-63, -39) | (-61, -51) |
| SapB (chain B) [d] | (-56, -31) | (-77, -44) | **(-105, +3)** | (-55, -50) | (-58, -43) |

($\phi$, $\varphi$) in degrees

[a] "n" is residue Tyr54 in the sequence DS**Y**LP in SapA, SQ**Y**SE in SapB, DT**Y**GS in SapC, and AE**Y**EP in SapD.

[b] Data from PDB entry 2GTG (Ahn et al. 2006)

[c] Data from PDB entry 2DOB (Ahn et al. 2006)

[d] Data from PDB entry 1N69 (Ahn et al. 2003)

[e] Data from PDB entry 2RB3 (Rossmann et al. 2007)

## 4.2 Conjugated bile salt hydrolase (CBAH)

### 4.2.1 Expression and purification of CBAH

CBAH was recombinantly expressed in *E. coli* strain BL21 (DE3). Wild-type CBAH and SeMet CBAH were purified at levels of 2 mg/l *E. coli* culture, whereas using the same protocol the purification rate of only ~0.01-0.02 mg/l was achieved for CBAH variants except for CBAH R18H with a purification rate of 0.1 mg/ml. A major part of CBAH variants was insoluble as shown by Western-Blots (Fig. 26). Additionally, a 40 kDa protein, named CBAH2, cross-reacted with anti-CBAH antibodies and was also detected in an untransformed BL21 (DE3) strain. The relationship of this protein to CBAH and its function are unknown.
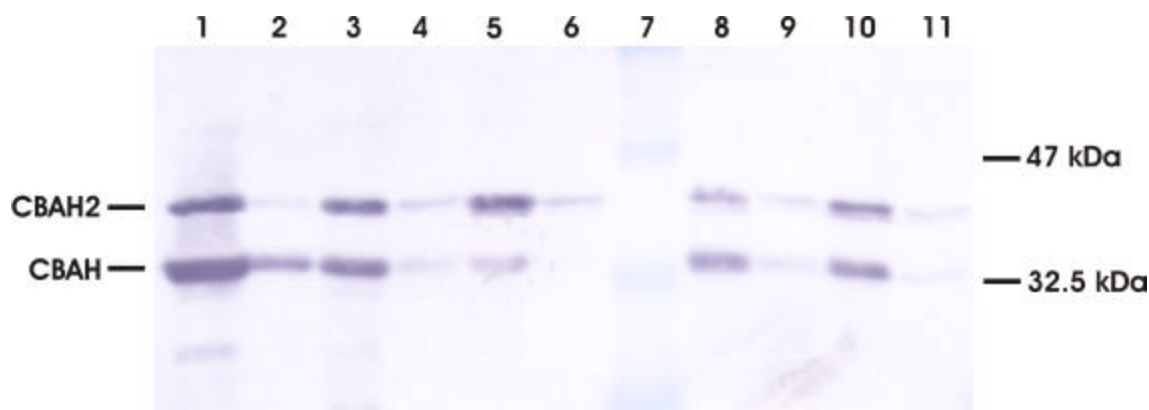


**Figure 26**. **Immunoblot analysis of CBAH expression in *E.coli*.** Lane 1-total cell extract of wtCBAH, lane 2-soluble fraction of wtCBAH, lane 3-total cell extract of CBAH R18H, lane 4-soluble fraction of CBAH R18H, lane 5-total cell extract of CBAH C2, lane 6-soluble fraction of CBAH C2S, lane 7-protein marker, 8-total extract of CBAH R18L, lane 9-soluble fraction of R18L, lane 10-total cell extract of CBAH C2A, lane 11-soluble fraction of CBAH C2A.

*CBAH2 is a protein with unknown function which is recognized by polyclonal anti-CBAH antibodies and can also be detected in untransformed BL21 (DE3) strain.

Wild-type CBAH (wtCBAH) and variants CBAH C2A, C2S, R18H, R18L were purified using a hydrophobic interaction column and gel filtration. Denaturating SDS-PAGE analysis of purified CBAH indicated a single polypeptide with molecular weight of ~34 kDa (Fig. 26), whereas chromatography of CBAH on a calibrated gel filtration column

provided only one peak with a retention time of 69 ml corresponding to a molecular weight of ~140 kDa, suggested that CBAH exists in its native state as tetramer (Fig. 27).
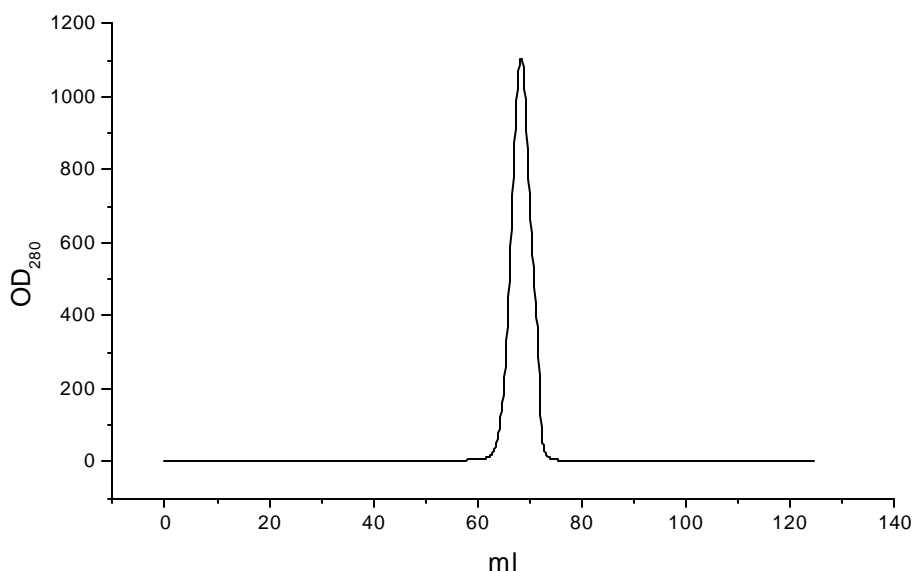


**Figure 27. Gel filtration of CBAH on a S200 column.** A retention volume of 69 ml corresponds to the apparent size of ~140 kDa indicating that CBAH occurs as tetramer.

### 4.2.2 Crystallization of CBAH and X-ray data collection

Crystallization of apoCBAH was achieved by vapor diffusion using either hanging or sitting drop methods. The sparse-matrix screens showed that the protein crystallized with different morphologies under a variety of conditions, whereas crystals appeared after 1-2 days incubation at 18° C. The best crystals of apoCBAH were obtained in hanging drop with 2.6 M ammonium sulfate, 100 mM sodium citrate pH 6.0 (Fig. 28a) and grew to approximate dimensions of $0.5 \times 0.4 \times 0.4$ mm and diffracted X-rays to 1.6 Å resolution. The apoCBAH crystals were found to belong to space group $C222_1$ with unit-cell parameters a = 90.13 Å, b = 189.30 Å, c = 86.80 Å. Detailed data-collection statistic for apoCBAH dataset is given in Table 9. Soaking of the crystals of apoCBAH with substrates led to rapid crystal destruction. The CBAH crystals complexed with different reaction products were obtained with 23% PEG 3350, 200 mM ammonium sulfate, 100 mM Bis-Tris pH 5.5 in sitting drop by co-crystallization with 1-10 mM substrate (Fig. 28 a-c,f).
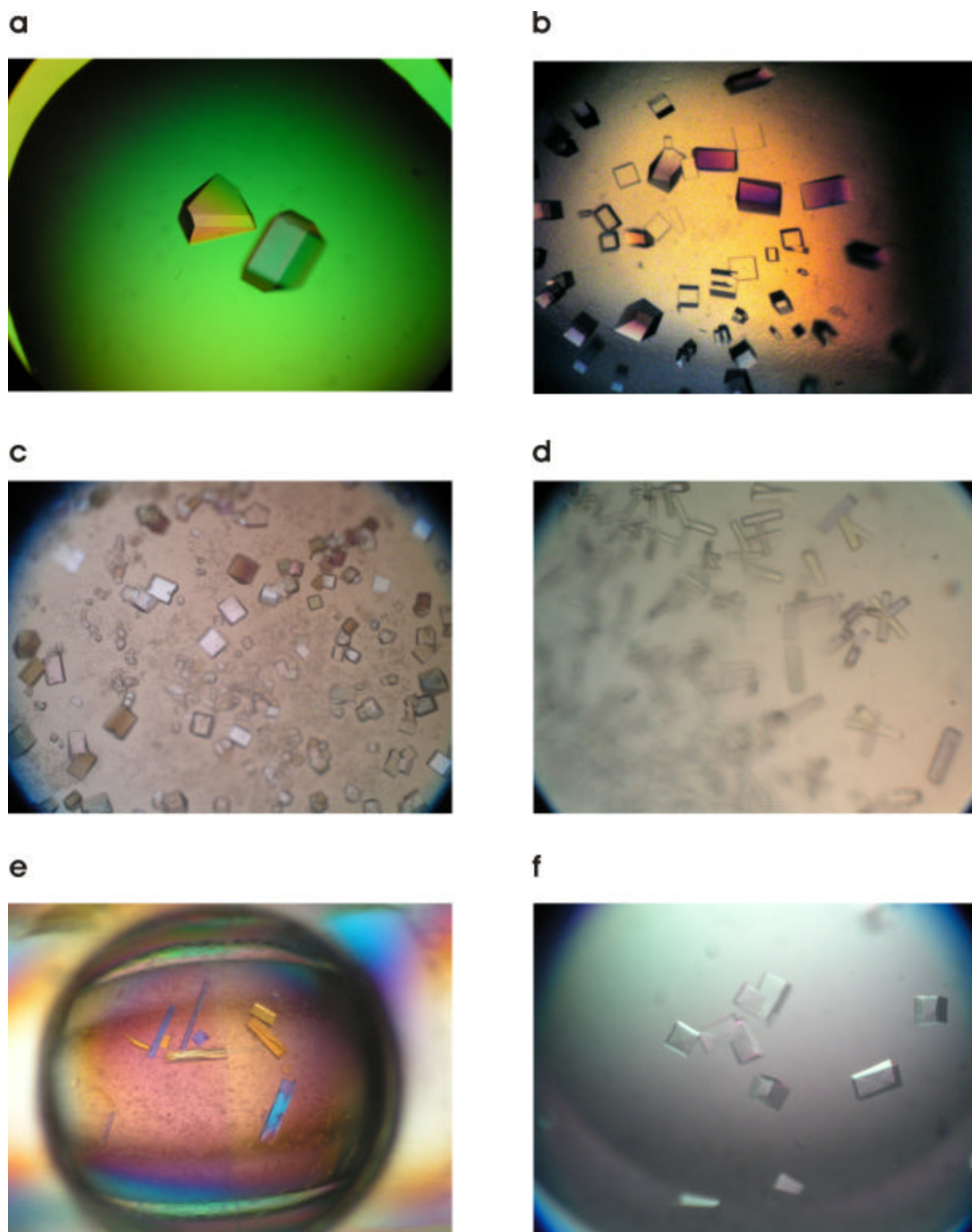
**Figure 28. CBAH crystals.** *(a)* Crystals of apoCBAH. *(b)* Crystals of CBAH co-crystallized with deoxycholoyltaurine. *(c)* Crystals of CBAH co-crystallized with choloylglycine. *(d)* Crystals of CBAH variant C2A. *(e)* Crystals of the CBAH variant R18L. *(f)* Crystals of CBAH co-crystallized with choloylsarcosine.

Crystals of the complexes diffracted to up to 1.7 Å resolution, whereas CBAH crystallized with both choloylglycine and choloylsarcosine in space group P4222 and taurocholate in space group F222. Detailed data collection statistics for CBAH-product complexes are given in Tables 9-10.

CBAH variant R18L crystallized in space group P4222 in sitting drop with 2.6 M and 100 mM HEPES pH 7.5 (Fig. 27e), whereas CBAH variant C2A was crystallized in space group $P4_122$ with 23% PEG 3350, 200 mM ammonium sulfate, 100 mM Bis-Tris pH 5.5 in sitting drop (Fig. 27d). CBAH mutants diffracted X-rays to 1.8 Å and 2.9 Å spacing, respectively. Detailed data collection statistics for both CBAH variants are given in Table 11.

### 4.2.3  Structure determination of CBAH

X-ray data of CBAH were first obtained with crystals of CBAH in complex with deoxycholoyltaurine, and consequently these data were used for structure determination. This structure was determined by molecular replacement using data between 3.5 and 30 Å resolution with program PHASER (Read, 2001; Storoni et al., 2004). A poly-serine model of penicillin V acylase monomer (PDB ID 3PVA, chain A) was taken as search model. The molecular replacement solution was improved with program ARP/WARP. Subsequent rounds of refinement were carried out by alternating between manual model building in COOT and refinement using maximum-likelihood method and a simulated annealing protocol from the CNS software. The products taurine and deoxycholate were built manually into the electron density using COOT. The topology files for taurine and deoxycholate were generated using the PRODRG sever (Schuttelkopf and van Aalten, 2004). The final model was refined and manually rebuilt until convergence of the conventional R-factor and of $R_{free}$ with 5% of the data reserved for cross-validation. Water molecules were added using the automated refinement program ARP in concert with REFMAC. Composite omit maps calculated with the CNS software package were used for the final model validation. The crystal structures of apoCBAH, CBAH in complex with choloylsarcosine, and CBAH variants C2A and R18L were determined using the refined model of CBAH as a search model in program MOLREP. Refinement and model-building steps were carried out as described above for the CBAH in complex with deoxycholoyltaurine.

**Table 9.** Data collection and refinement statistics for datasets of apoCBAH and CBAH with deoxycholoyltaurine (dctCBAH).

| | apoCBAH | dctCBAH |
|---|---|---|
| **Data collection** | | |
| Space group | C222$_1$ | F222 |
| Cell dimensions? | | |
| a, b, c (Å) | 90.13, 189.30, 86.80 | 63.16, 63.16, 339.38 |
| Resolution (Å) | 30 – 1.5 | 50 - 1.67 |
| Total number of reflections | 530059 | 148760 |
| Unique reflections | 105127 | 41324 |
| $R_{sym}$ (%) | 4.9 (37.7)* | 3.8 (11.7) |
| I / s(I) | 24.3 (2.8) | 29.4 (7.8) |
| Completeness (%) | 88.9 (95.7) | 97.1 (95.7) |
| Redundancy | 4.9 | 3.6 |
| | | |
| **Refinement** | | |
| Resolution (Å) | 30–1.60 (1.69-1.60) | 30-1.67 (1.71-1.67) |
| Unique reflections | 95,804 (13,708) | 40,736 (2,976) |
| $R_{work}$ / $R_{free}$ (%) | 19.3 / 22.6 (22.5 / 24.9) | 16.6 / 19.4 (18.8 / 23.6) |
| Number of atoms | | |
| Protein | 5,226 | 2,676 |
| Ligand | | |
| taurine | - | 7 |
| deoxycholate | - | 28 |
| Water | 936 | 419 |
| Average B-factors (Å$^2$) | | |
| Protein | 17.22 | 13.48 |
| Ligand | | |
| taurine | - | 46.56 |
| deoxycholate | - | 27.35 |
| Solvent | 32.01 | 29.58 |
| R.m.s. deviations | | |
| Bond lengths (Å) | 0.007 | 0.006 |
| Bond angles (°) | 1.078 | 1.026 |
| Ramachandran plot | | |
| Most favoured | 92.3 | 90.5 |
| Additionally allowed | 7.1 | 8.8 |
| Generously allowed | 0.5 | 0.7 |
| Disallowed | 0 | 0.3 |

*Values in parentheses are for highest-resolution shell.

**Table 10.** Data collection and refinement statistics for datasets of CBAH with choloylglycine (cgCBAH) and choloylsarcosine (csCBAH).

| | cgCBAH | csCBAH |
|---|---|---|
| **Data collection** | | |
| Space group | P4222 | P4222 |
| Cell dimensions? | | |
| a, b, c (Å) | 64.24, 64.24, 169.90 | 64.32, 64.32, 169.51 |
| Resolution (Å) | 20 - 1.8 | 50 – 1.7 |
| Total number of reflection | 320578 | 378241 |
| Unique reflections | 28678 | 40202 |
| $R_{sym}$ (%) | 7.1 (30.6.)* | 7.3 (37.0) |
| $I / \sigma(I)$ | 18.5 (5.3) | 22 (5.3) |
| Completeness (%) | 99.0 (97.6) | 100 (100) |
| Redundancy | 6.2 | 9.4 |
| | | |
| **Refinement** | | |
| Resolution (Å) | 50-1.70 (1.85-1.80) | 50-1.70 (1.79-1.70) |
| Unique reflections | 33,907 (2,294) | 38,137 (2,769) |
| $R_{work} / R_{free}$ (%) | 19.3 / 22.7 (22.0 / 25.2) | 19.7 / 22.4 (24.1 / 29.1) |
| Number of atoms | | |
| Protein | 2,626 | 2,655 |
| Ligand | | |
| glycine | 5 | - |
| sarcosine | - | 6 |
| cholate | 29 | 29 |
| Water | 379 | 326 |
| Average B-factors (Å$^2$) | | |
| Protein | 18.14 | 18.34 |
| Ligand | | |
| glycine | 41.20 | - |
| sarcosine | - | 44.35 |
| cholate | 29.19 | 30.78 |
| Solvent | 33.54 | 34.40 |
| R.m.s. deviations | | |
| Bond lengths (Å) | 0.014 | 0.007 |
| Bond angles (°) | 1.434 | 0.099 |
| Ramachandran plot | | |
| Most favoured | 89.4 | 90.5 |
| Additionally allowed | 9.6 | 8.5 |
| Generously allowed | 0.7 | 0.7 |
| Disallowed | 0.3 | 0.2 |

**Table 11.** Data collection and refinement statistics of datasets of CBAH variants R18L C2A.

| | CBAH R18L | CBAH C2A |
|---|---|---|
| **Data collection** | | |
| Space group | P42 2 2 | P4$_1$ 2 2 |
| Cell dimensions? | | |
| a, b, c (Å) | 64.24, 64.24, 170.02 | 64.64, 64.64, 338.83 |
| Resolution (Å) | 50 - 1.77 | 50 – 2.9 |
| Total number of reflections | 400844 | 141990 |
| Unique reflections | 35363 | 17112 |
| $R_{sym}$ (%) | 5.2 (24.3)* | 14.2 (46.2) |
| $I / s(I)$ | 44.9 (8.2) | 12.6 (3.3) |
| Completeness (%) | 98.9 (89.2) | 90.7 (90.3) |
| Redundancy | 11.3 | 8.3 |
| | | |
| **Refinement** | | |
| Resolution (Å) | 40-1.80 (1.85-1.80) | 30-2.90 (2.99-2.90) |
| Unique reflections | 32,302 (2,314) | 14661 (1,029) |
| $R_{work} / R_{free}$ (%) | 17.8 / 21.3 | 20.9 / 27.5 |
| | (18.7 / 27.2) | (25.2 / 36.6) |
| Number of atoms | | |
| Protein | 2,640 | 5,232 |
| Water | 392 | 120 |
| Average B-factors (Å2) | | |
| Protein | 19.82 | 26.60 |
| Solvent | 21.83 | 19.39 |
| R.m.s. deviations | | |
| Bond lengths (Å) | 0.011 | 0.006 |
| Bond angles (°) | 1.207 | 0.886 |
| Ramachandran plot | | |
| Most favoured | 91.2 | 88.1 |
| Additionally allowed | 8.2 | 11.5 |
| Generously allowed | 0.3 | 0.3 |
| Disallowed | 0.3 | 0.0 |

*Values in parentheses are for highest-resolution shell.

The crystal structure of CBAH with choloylglycine labeled with selenomethionine was determined by the SAD method. The dataset was collected close to the absorption edge of the selenium atom. Program autoSHARP (Vonrhein et al.,

2006) was used to locate the expected eight selenium atoms and to refine their atomic coordinates. Further improvement of phases by using solvent flattening with program DM assuming a solvent content of 48% revealed an interpretable electron density map in which an initial model of CBAH was built with program ARP/WARP. Further refinement of the model and manual model building were carried out as described above. Details of the refinement statistics are presented in Tables 9-11.

### 4.2.4  Overall structure of CBAH

Crystal structures of  *C. perfringens* CBAH, recombinantly expressed in *E. coli*, were determined at high resolution. The apoenzyme crystallized in the orthorhombic space group $C222_1$ with two monomers in the crystal asymmetric unit (AU). The apoCBAH monomer is made up of a single globular domain with approximate dimensions 40 Å x 50 Å x 55 Å excluding one extended loop (200s loop) from Gln188 to Pro225 which stretches out for about 40 Å from the globular domain (Figure 29a). Including the antiparallel ß-sheet of this extended loop, the domain has a six-layered structure of composition ßaßßaß (Fig. 29a, b). The core of the protein is composed of two sandwiched antiparallel ß-sheets that contain N- and C-termini and are covered by a layer of antiparallel a-helices. The N-terminal ß-sheet (I) is made up of five antiparallel  ß-strands with topography $NH_2$-ß1-ß2-ß16-ß17-ß18, and the C-terminal ß-sheet (II) is composed of eight antiparallel ß-strands  ß13-ß12-ß11-ß8-ß7-ß6-ß3-ß19-COOH,  the  last  five  ß-strands  being packed in parallel fashion against ß-sheet (I).

CBAH variants crystallized in the tetragonal space groups P4222 and $P4_122$, respectively. R18L contained one monomer and C2A contained two monomers per AU, respectively. CBAH in complex with deoxycholoyltaurine crystallized in the orthorhombic space group F222 with one monomer in the asymmetric unit. Before or during crystallization, deoxycholoyltaurine was hydrolyzed to the products taurine and deoxycholate that were both found in the crystal structure. In the structure of the CBAH co-crystallized with choloylglycine and choloylsarcosine in space group P4222, the substrates were also hydrolyzed by CBAH, and the reaction products remained bound to the crystallized protein.
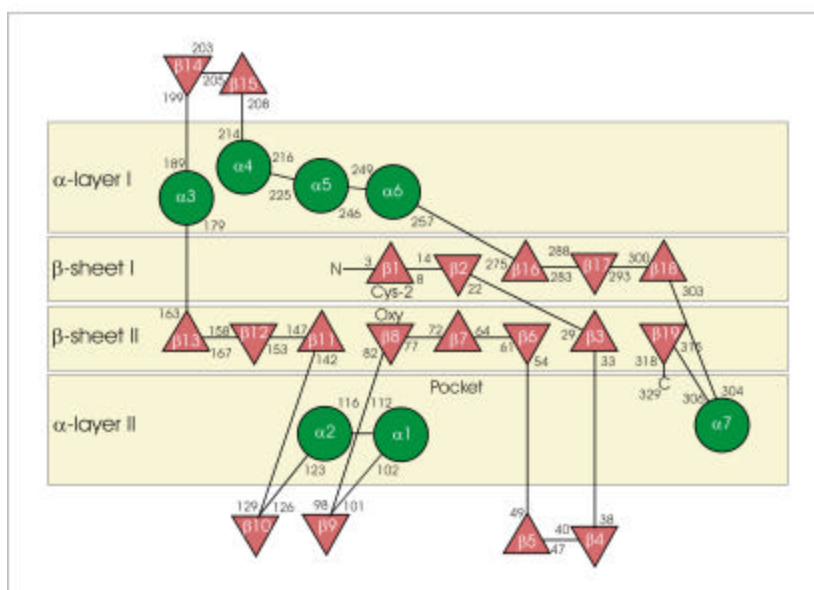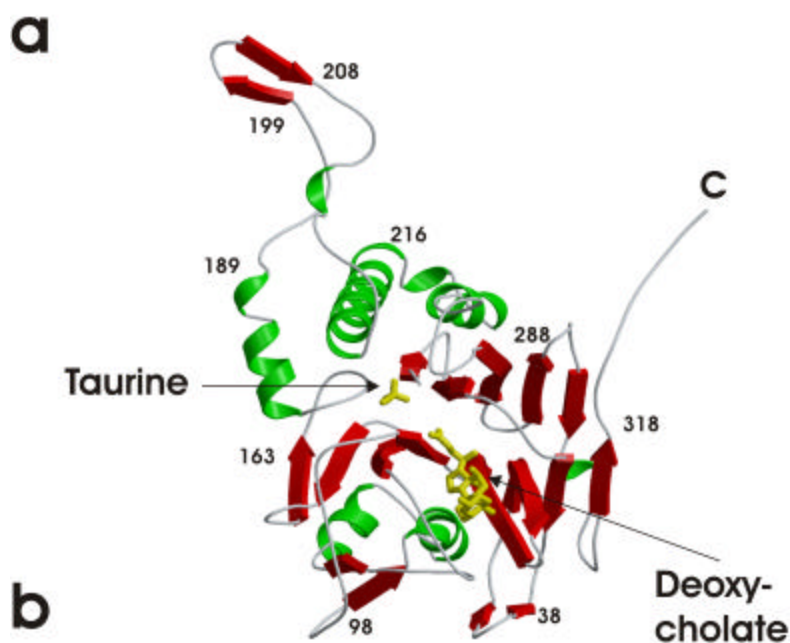
**Figure 29. Overall structural features of  C. perfringens CBAH**. *(a)* CBAH monomer. The reaction products taurine and deoxycholate are shown in magenta. *(b)* Topography of the CBAH monomer.

The reaction products deoxycholate or cholate are primarily bound to residues from ß-sheet (II) located on strands ß6, ß7, ß11. The second reaction product, the leaving

57

group (glycine, taurine or sarcosine), is bound by Asn82 only, which is an active site residue and located on strand β8. In apoCBAH this position is occupied by an ethylene glycol molecule, whereas a glycerol molecule was found at the position of deoxycholate/cholate (Fig. 34).
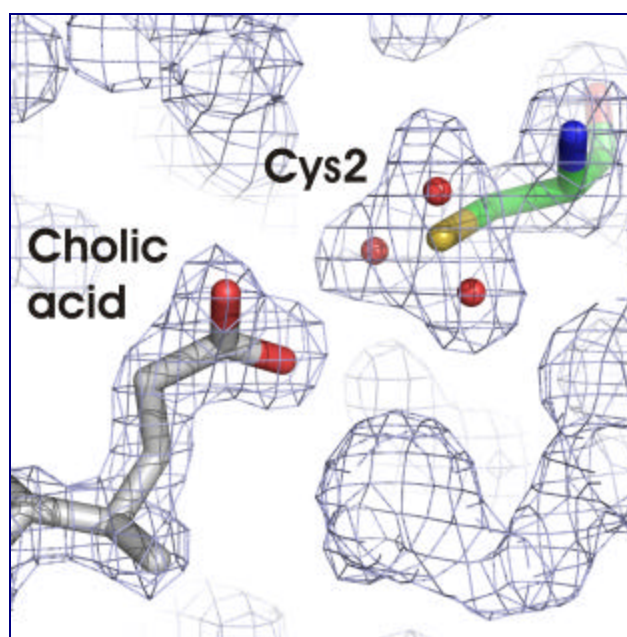


**Figure 30. Oxidation of the SH-group of the active site Cys2 to sulfonic acid.** In the CBAH crystals grown from protein preparations which were purified and crystallized without DTT, the SH-group of Cys2 was oxidized to sulfonic acid. Oxygen atoms of sulfonic acid are presented by red spheres. The $2F_o$-$F_c$ map is countered at 1σ.

Notably, in the CBAH crystals grown from protein preparations which were purified and crystallized without DTT, the SH-group of Cys2 was oxidized to sulfonic acid as shown in the electron density (Fig. 30). DTT protected Cys2 in CBAH against oxidation and CBAH crystallized in the native form when treated thoroughly under reducing conditions using DTT. Oxidation of the N-terminal cysteine was also observed in the crystal structure of PVA (Kumar et al., 2004) and indicates its strong nucleophilic character.

All the CBAH models except for chain B of the C2A variant contained the full polypeptide chain with only the initial formyl-methionine missing, in agreement with N-terminal sequencing and MALDI-TOF spectra yielding a molecular mass of 37060 Da

for the 328 amino-acid-long polypeptide. Chain B of the C2A variant contained the N-terminal methionine that was missing at the N-terminus of chain A.

In all CBAH models the main chain $f, j$ -torsion angles of about 90 % of residues lie in the most favoured regions. Remarkable, the torsions angles of Thr174 lie in the generously allowed or disallowed region in all structures. Notably, Thr174 is located next to the Asn175 in the active site, between ß13 and a3.

The aßßa fold together with the observation that the N-terminal amino acid is cleaved off to expose the potentially catalytic Cys2 residue, classified CBAH as a member of the N-terminal nucleophile (Ntn-) hydrolase superfamily (Brannigan et al., 1995; Oinonen and Rouvinen, 2000; Saarela et al., 2004). An additional characteristic of this group of enzymes is that they all display higher quaternary structural organization which has also been found in the here studied CBAH structure.

### 4.2.5 Quaternary structure of CBAH

Gel filtration and crystallization of CBAH were achieved under slightly acidic conditions close to the catalytic pH optimum of the enzyme (5.8 to 6.4) (Gopal-Srivastava and Hylemon, 1988). Gel filtration runs (Fig. 27) and crystallographic data showed that CBAH exists as a homotetramer under physiological conditions.

Besides the extended 200s loop with antiparallel ß-strands ß14, ß15, ßstrands ß9 and ß17-ß19 as well as the a-helices a3-a7, and the turns and loops connecting these strands and helices are the contact sites for the tetramer organization of CBAH (Figure 31a). This configuration leaves the substrate-binding pocket (formed by ß6, ß7, and ß11) freely accessible to solvent. The subunits in the CBAH tetramer are related by dihedral (D2 or 222) point group symmetry that is clearly different from a fourfold (C4) symmetry (Figure 31b). The dihedral organization results in different contacts between the individual protomers. The largest interface within the tetramer is the one observed between molecules A and C or B and D, respectively, that are related by the horizontal twofold rotation axis $2_H$ in Figure 31b. This dimer interface has a buried surface area of 3167 $Å^2$ per protomer with 38 amino acids per protomer involved in contact formation. The interface contains residues from the C-terminal tail, helices a4- a7, strands ß17- ß19, and part of the 200s loop.
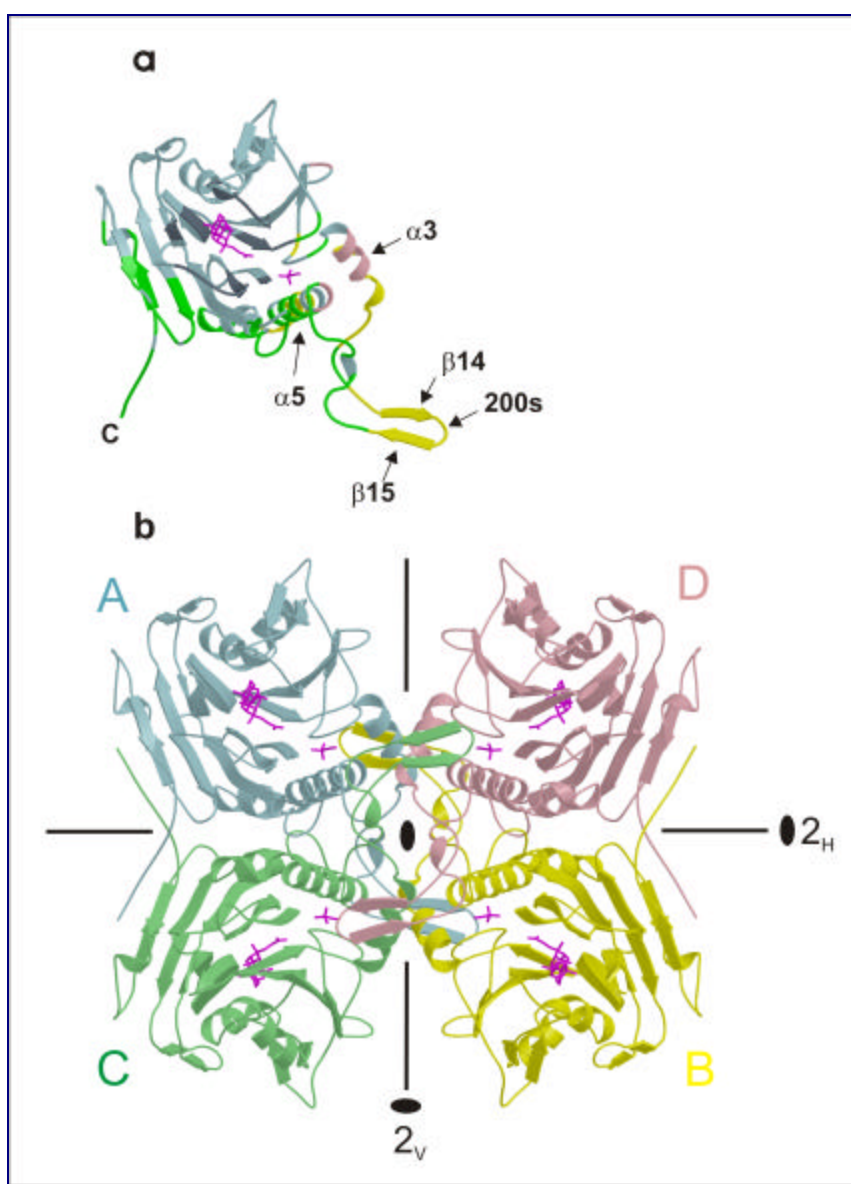
**Figure 31. Tetramer Formation of CBAH.** Reaction products taurine and deoxycholate are indicated in stick presentation. **(a) Ribbon diagram of CBAH monomer.** Contact surfaces with other molecules in the tetramer indicated by colour-coding as in (b). The shown monomer has the same orientation as protomer A in (b). **(b) CBAH homotetramer with monomers coloured differently.** Twofold rotation axes in the paper plane are marked $2_H$ (horizontal), $2_V$ (vertical), and the one perpendicular (normal) to the paper plane (2p) is indicated by the black ellipse in the center.

The second largest interface is formed between molecules A and B or C and D, respectively, which are related by the perpendicular twofold rotational symmetry axis $2_P$ in Fig. 31b. This contact involves helices a3-a5 and, in particular, ß-sheet

60

(ß14, ß15) of the 200s loop and involves 21 amino acids with a total buried surface area of 1405 $\text{Å}^2$ per protomer. In the tetramer, the four 200s loops cross over the symmetry centre leading to an extensively interlocked oligomer.

The third interface, formed by molecules A and D or B and C respectively, which are related by the vertical twofold rotation axis $2_V$ in Fig. 31b, is a minor contact site involving only 8 residues that are located mostly on a3 and on the N-terminus of a5 and buries a surface area of 490 $\text{Å}^2$ per protomer. The large surface between molecules A, C and B, D suggests that the tetramer may be considered as consisting of two homodimers.

Superimposition of crystal structures of apo- , productCBAH, and CBAH variants did not reveal any significant conformational differences between monomers, the overall root mean square deviations was not more than 0.40 Å between C$\alpha$ atoms of monomers as well as tetramers of all crystal forms. This agrees with kinetic data of the enzymes that fit the Michaelis-Menten equation (Gopal-Srivastava and Hylemon, 1988; Kirby et al., 1995; Nair et al., 1967), and suggest that CBAH may be described as an enzyme with four independent catalytic sites showing no cooperativity.

### 4.2.6 Substrate-binding pockets of CBAH

In the binding pocket of CBAH crystallized with deoxycholoyltaurine, the deoxycholate is sandwiched by Phe61 and Ile137 on the deoxycholat ring A and by Met20, Ala68 and Phe26 on the side of the isovaleric acid side chain of deoxycholate (Fig. 32-33). Additional hydrophobic interactions are formed between Ile133 and deoxycholat ring B and Leu142 and ring D. The only hydrogen bonds (below 3.4 Å) are donated by Cys2SH and Arg18NH2 to deoxycholate O26 and by water molecule to O25.
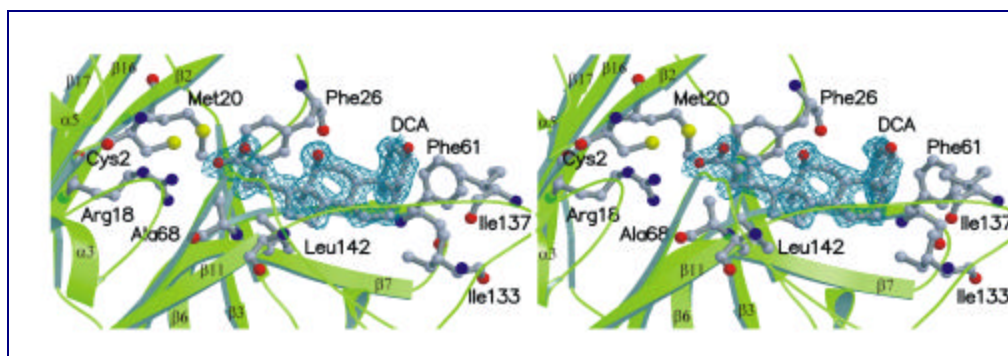
**Figure 32. Main contacts between deoxycholate and protein residues.** Stereo presentation of the density around deoxycholate from a 2Fo-Fc composite omit map contoured at 1s level shown in blue.
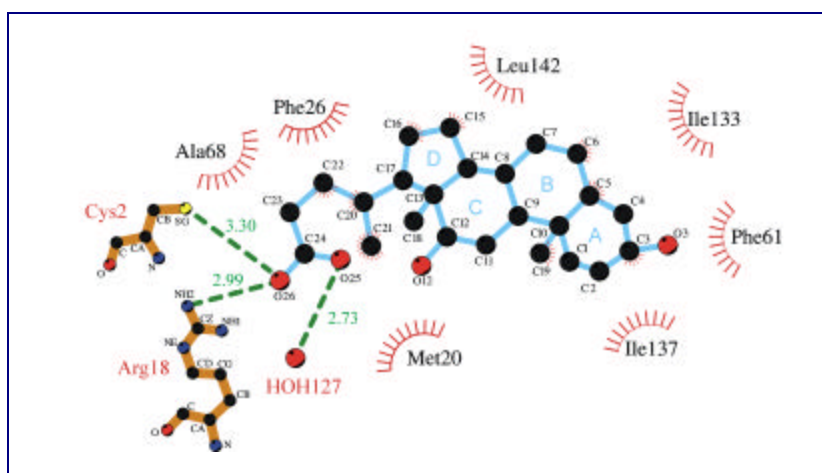


**Figure 33. Schematic drawing of the main contacts between deoxycholate and protein residues.**

In the crystal structures of CBAH crystallized with choloylglycine and choloylsarcosine additional hydrogen bond was found between the side chain of the Thr140 and hydroxyl group (this group is absent in deoxycholate) of the cholate (Fig. 35).

All the obtained CBAH structures revealed nearly identical conformations. The largest differences with r.m.s.d. of 1.4 Å calculated for Cα atoms were found between apoCBAH and CBAH complexes with reaction products in the loop connecting residues 128-140 (Fig. 34).
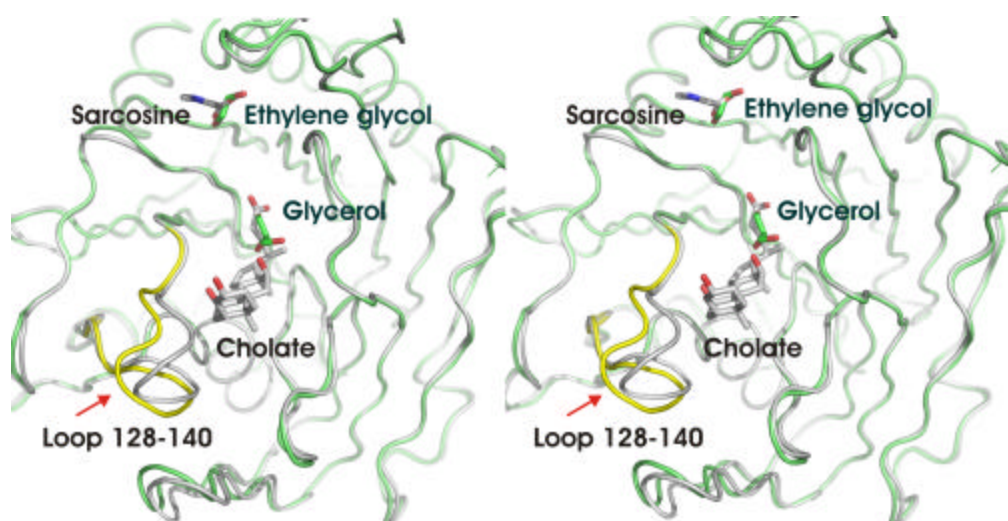
**Figure 34. Superimposition of crystal structures of apoCBAH- and CBAH with choloylsarcosine (stereo).** Atom positions in the loop connecting residues 128-140 (shown in yellow) shows r.m.s.d. of 1.4 Å. In the crystal structure of CBAH with choloylsarcosine this loop is more open (indicated by red arrow and colored grey). In the apoCBAH crystal structure glycerol and ethylene glycol molecules occupy positions equivalent to that of cholate or sarcosine, respectively, in CBAH with choloylsarcosine.

Taurine, sarcosine and glycine are bound to the protein through a one hydrogen bond. Interestingly, taurine in CBAH with deoxycholoyltaurine and sarcosine in CBAH with choloylsarcosine have a "reversed" orientation pointing with their sulfo or carboxyl group towards Cys2 and leaving the active site with their amino groups ahead (Figs. 35-36). The weak electron density and high B-values of glycine, taurine and sarcosine are indicative of a significant disorder or lower occupation of these ligands, in agreement with the finding that they form only one hydrogen bond to a protein side chain.

In the structure of the gcCBAH and scCBAH additional hydrogen bond was found between the side chain of the Thr140 and hydroxyl group (this group is absent in deoxycholate) of the cholate (Fig. 35).
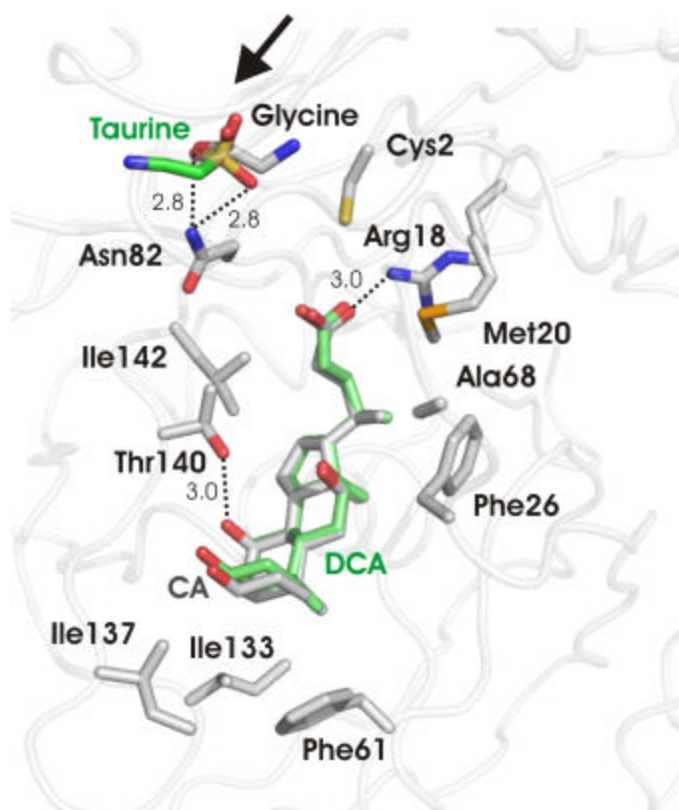
**Figure 35. Product binding in CBAH. Superimposition of crystal structures of CBAH- with choloylsarcosine (green) and CBAH with choloylglycine (grey).** Residues involved in the binding of the products are colored grey. Taurine and deoxycholate (DCA) in the crystal structure of CBAH with deoxycholoyltaurine are colored green. Glycine and cholate (CA) in the crystal structure of CBAH with choloylglycine are shown in grey. Taurine has a "reversed" orientation (indicated by black arrow) pointing with its sulfo group towards Cys2 and leaving the active site with its amino group ahead. Side chain hydroxyl group of Thr140 makes hydrogen bonds with the hydroxyl group of cholate.

## 4.2.7 Structural comparison of CBAH to other proteins

Ntn-hydrolases are characterized by a four-layered catalytically active aßßa-structure where the two-core ß-sheets are antiparallel and packed against each other (Artymiuk, 1995; Brannigan et al., 1995; Oinonen and Rouvinen, 2000). The crystal structure of *C. perfringens* CBAH reveals a close structural and catalytic similarity to Ntn-hydrolases.

**Table 12.** Comparison of CBAH with known protein structures using DALI[*]

| Protein | PDB | Z-score | rmsd | LALI[1] | LSEQ[2] | IDE[3] |
|---|---|---|---|---|---|---|
| penicillin V acylase | 3pva-A | 42.7 | 1.9 | 325 | 334 | 34 |
| glutaryl 7-aminocephalosporanic acid acylase | 1fm2 | 15.0 | 3.4 | 221 | 520 | 13 |
| penicillin amidohydrolase | 1ajq-B | 14.7 | 3.2 | 212 | 557 | 11 |
| proteasome | 1pma-P | 9.8 | 3.2 | 157 | 203 | 7 |
| structural genomics, unknown function | 1kuu-A | 8.4 | 3.9 | 163 | 202 | 7 |
| heat shock locus V | 1ned-A | 8.0 | 2.9 | 141 | 180 | 9 |
| glucosamine 6-phosphate synthase | 1gdo | 2.6 | 4.2 | 104 | 238 | 9 |
| aspartylglucosamidase | 1apy-B | 2.5 | 3.8 | 95 | 141 | 5 |
| glutamine amidotransferase | 1ecf-B | 2.1 | 5.9 | 116 | 500 | 8 |
| Ser/Thr Phosphatase 2C | 1a6q | 2.1 | 4.7 | 129 | 363 | 5 |

*Only the most similar proteasome subunit has been included in the table

[1] Length of aligned sequence

[2] Total length of sequence

[3] % of identical residues

Structural comparison to known structures deposited in the Protein Data Bank using DALI (Holm and Sander, 1993) underlines the relationship of *C. perfringens* CBAH to Ntn-hydrolases (Table 15). Among the 10 significantly similar structures (Z-score =2) the first 9 all classify as Ntn-hydrolases. Not surprisingly, *C. perfringens* CBAH

resembles penicillin acylase from *B. sphaericus* (PVA) most closely with which it shares the highest sequence homology.



**Figure 36. Sequence alignment.** CBAH from *C. perfringens* (CBH_CP), Choloylglycine Hydrolase from *Bacillus anthracis* (CGH_BA), CBAH from *Lactobacillus plantarum* (CBH_LP), Conjugated Bile Salt Hydrolase from *Lactobacillus johnsonii* (CBH_LJ), Conjugated Bile Salt Hydrolase from *Lactobacillus acidophilus* (CBH_LA), Bile Salt Hydolase from *Bifidobacterium bifidum* (BSH_BB), Choloylglycine Hydrolase from *Bifidobacterium longum* (CGH_BL), and penicillin V acylase from *Bacillus sphaericus* (PVA_BS). The alignment highlights active site residues (marked by black asterisk) and residues of the substrate-binding pocket (grey). Conserved residues of the contact sites are color coded according to Figure 31.

The similarities also extend to the quaternary structures of the proteins. For PVA a similar tetramer formation with dihedral symmetry has been described (Suresh et al.,

1999). The similarity in tetramer formation between CBAH and PVA is also reflected in the high number of strictly conserved residues found at the tetramer contact sites (see Fig. 36 and discussion below). There is overall high structural similarity with modest, low, or even absent sequence conservation among the members of the Ntn-hydrolase superfamily (Brannigan et al., 1995). An interesting observation is the structural similarity to Ser/Thr phosphatase C2 (Table 12) which is a member of the metallo-phosphoesterase superfamily that shares a similar fold to Ntn-hydrolases with an a/ß-aßßa-layered structure. As in the case of Ntn-hydrolases, the active site is located between the core ß-strands.

To uncover the most critical residues in tetramer formation, the monomer-monomer contact sites in *C. perfringens* CBAH were matched with a sequence alignment of BSH enzymes and penicillin V acylase (Fig. 36). The smallest contact interface between molecules A and B in the tetramer has only two highly conserved residues: Trp181 and Asn185 that form a four-side chains mini cluster where Trp181 contacts Asn185 and vice versa. Besides its contact to Asn185, Trp181 stacks hydrophobically with Trp181 of the neighbouring molecule. In some species Trp181 is replaced by phenylalanine or tyrosine which can form equivalent contacts. Most of the highly conserved residues of the interface AB are in the 200s loop, Leu186, Leu207, and Gly218 in molecule A contacting residues Leu186, Leu207, Gly218 in molecule B, respectively. The reciprocal nature of most of the contacts between the conserved residues of the dimer interface supports the view that conservation is due to contact formation. The large interface between molecules A and C comprises 10 strictly or highly conserved residues including a cluster from Gly218 to Asp222 making again reciprocal contacts to the same residues of the neighbouring molecule. Additional highly conserved residues that make reciprocal contacts are His254, Tyr294 and Ile299. The contact residues of Gly213 and Gly264 are not as well conserved as the glycines themselves. Since both glycines assumed phi/psi angles that cannot be easily formed by other residues (110.4°/13.1° and 103.5°/-28.8°, respectively), it is likely that conservation of these glycines is essential for the course of the main chain rather than for contact formation.

All Ntn-hydrolases catalyze amide bond cleavage through an autocatalytically exposed N-terminal nucleophile (Kim et al., 2003; Kim et al., 2002; Li et al., 1999; Saarela et al., 2004) that might be a Ser, Thr, or Cys, with the N-terminal amino group

functioning as base in this reaction (Brannigan et al., 1995; Kim et al., 1996; McVey et al., 2001). Also the N-terminal f-Met of CBAH was found to be cleaved off to expose Cys2 which resides in the active centre classifying CBAH as belonging to this group.

### 4.2.8 Catalytic mechanism of CBAH

The catalytically active residues of CBAH can be inferred from biochemical experiments and from analogy to the active sites of other Ntn-hydrolases (Duggleby et al., 1995; Kim et al., 2000; Oinonen et al., 1995; Tikkanen et al., 1996), showing that Cys2 is involved in catalysis (Kim et al., 2004). All residues within the active site and in the substrate-binding pocket occupy well defined positions in the crystal structure (Fig. 37). The catalytic Ntn-diad is formed by amino acids Cys2 and Asp21. The α-amino group makes a hydrogen bond to a water molecule that bridges to the nucleophilic cysteinyl sulphur. The oxyanion hole is formed by main chain NH of Asn82 on strand β8 opposite the catalytic centre and by Asn175NH2 in the loop connecting ß13 and a3. The N-H groups of Asn82 and Asn175 are 3.8 Å and 3.9 Å away from the catalytic S, respectively, and 3.4 Å apart from each other (Fig. 37).
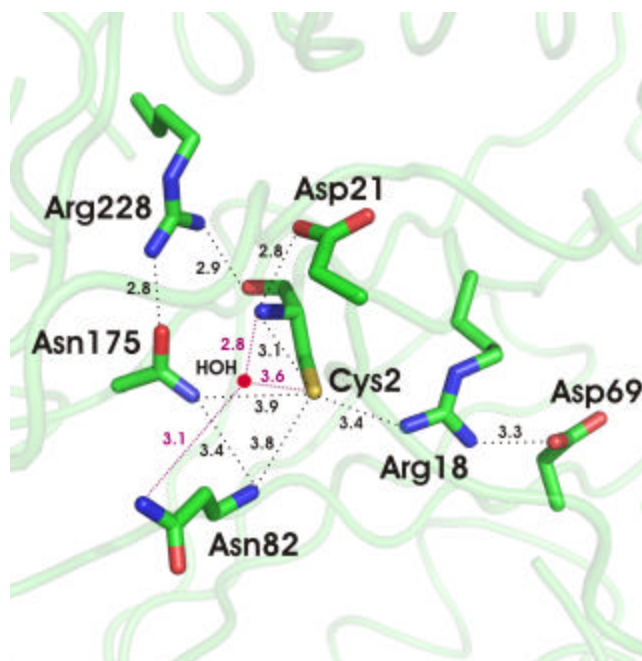


**Figure 37. Active site in apoCBAH.** Conserved active site residues occupy well-defined positions in the crystal structure of CBAH. The catalytic Ntn-diad is formed by amino acids Cys2 and Asp21. The α-amino nitrogen and the nucleophilic cysteinyl sulphur are bridged by a hydrogen-bonded water molecule (HOH, red sphere).
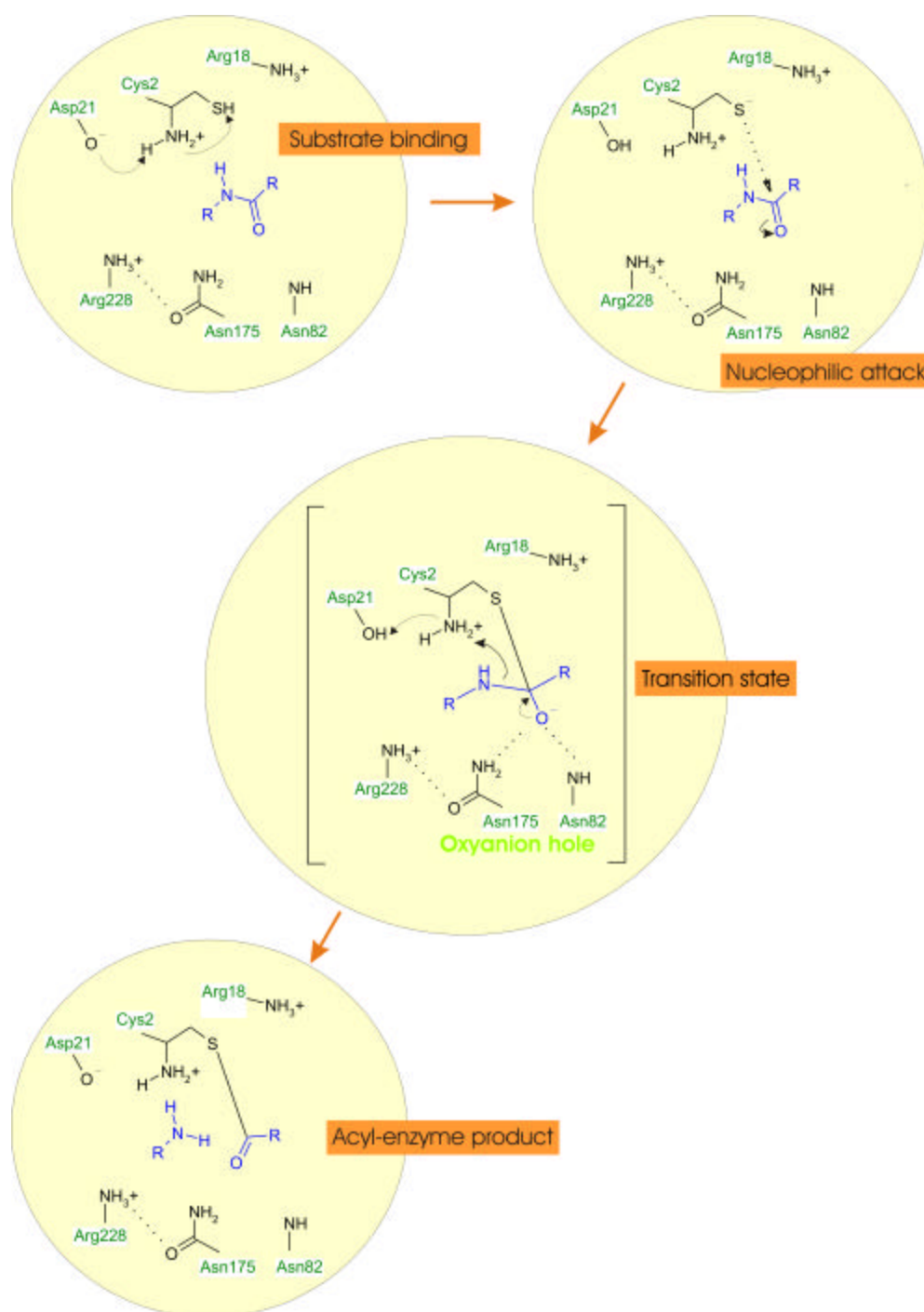
**Figure 38. Proposed reaction mechanism of CBAH, acylation step.** At the active site of the enzyme, Asp21 removes a H$^+$ from the ammonium group of Cys2, which in turn activates the sulfhydryl group of the same residue by accepting the S-H proton. The formed S$^-$ stabilized by Arg18 performs the nucleophilic attack on the carbon atom of the amide bond of the substrate. A tetrahedral intermediate is formed, which is stabilized by an oxyanion hole consisting of the side-chain amide of Asn175 and the main chain N-H of Asn82. The intermediate collapses after protonation of the R-NH- group, resulting in liberated R-NH$_2$ and the acyl-enzyme.
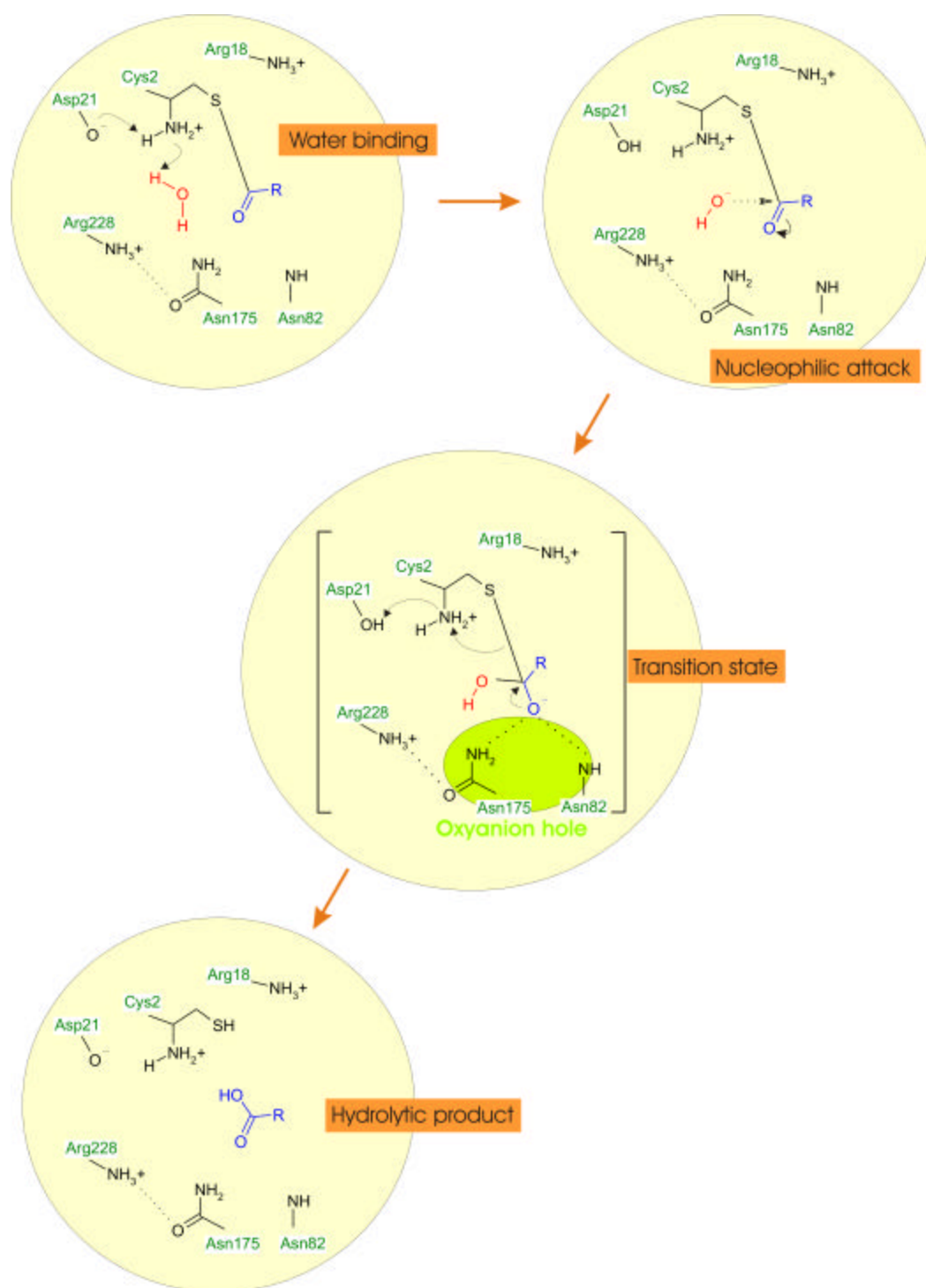
**Figure 39. Proposed reaction mechanism of CBAH, deacylation step.** Asp21 removes a H+ from the ammonium group of Cys2, which in turn deprotonates a water molecule that attacks the carbon atom of the thioester bond between Cys-S and the substrate. A tetrahedral intermediate is formed, which is stabilized by the oxyanion hole consisting of the side-chain amide of Asn175 and the main chain N-H of Asn82. The intermediate collapses, resulting in the regenerated enzyme and liberated carboxylic acid.

Based on the homologies of the catalytic centres (Fig. 40), the catalytic mechanism of CBAH can be proposed to follow that of Ntn-hydrolases. The $\alpha$-amino group of Cys2 acts as a base that deprotonates its own sulfhydryl group, which nucleophilically attacks the amide carbonyl group of the substrate (Figs. 38-39). This results in the formation of a covalent enzyme-substrate transition state that is stabilised by the oxyanion hole formed by the main chain NH of Asn82 and the side chain of Asn175. The acylation step is completed when the $\alpha$-amino group of Cys2 protonates the nitrogen of the scissile peptide bond, thereby forming an R-NH$_2$ leaving group that is liberated and a thioester is formed (Fig. 38). In the following deacylation step a water molecule is activated by the Cys2-amino group that removes a proton, and the formed OH $^-$ restores the enzyme by a nucleophilic attack to the acylenzyme (Fig. 39). The negatively charged intermediate is stabilized by the oxyanion hole, as in the previous acylation step. The reaction is complete when the $\alpha$-amino group donates the proton to the nucleophile to form S-H (Fig. 39).

### 4.2.9  Catalytic activity of CBAH variants

The activity of apoCBAH and CBAH variants was assayed colorimetrically by measuring the amount of glycine liberated from choloylglycine after 1 h at 37°C using the ninhydrin test. The results are summarized in the Table 13.

**Table 13.** CBAH activity at pH 5.5.

| Source | Spec. activity, unit* | Rel. activity, % |
|---|---|---|
| wtCBAH red. | 3.0 | 100 |
| wtCBAH ox. | 2.0 | 67 |
| CBAH C2S | 0 | 0 |
| CBAH C2A | 0 | 0 |
| CBAH R18H | 2.3 | 77 |
| CBAH R18L | 0 | 0 |

* 1 Unit = micromoles glycine formed per minute per mg protein at 37°C

The CBAH variants C2S, C2A and R18L were completely inactive, whereas the exchange of Arg18 by His18 showed 77% activity.

In crystals obtained with CBAH purified without DTT the active site sulfhydryl group was oxidized to sufonic acid (Fig. 30). The oxidized CBAH (wtCBAHox) was demonstrated to have 67% activity of the wild type (Table 13). Nevertheless, the substrates were bound to the enzyme in the crystals of CBAH crystallized with choloylglycine. It is not clear, whether the sulfur was oxidized prior or after the reaction. Theoretically, sulfonic acid can serve as a nucleophile and could attack the peptide bond, but it can also not be excluded that the activity of wtCBAHox is due to residual non-oxidized form.

## 4.2.10   Role of Arg18 in the active site of CBAH

It has long been known that cysteine is involved in the catalytic function of *C. perfringens* CBAH (Gopal-Srivastava and Hylemon, 1988; Nair et al., 1967). It was shown here that CBAH was inactivated by a single substitution Cys2? Ala2 (Table 16) and also the homologous bile salt hydrolase (BSH) from *Bifidobacterium longum* could be inactivated this way (Tanaka et al., 2000). The more conservative substitution Cys1? Ser1 inactivated CBAH (Table 13), and the BSH mutant C2S from *Bifidobacterium bifidum* was also shown to be inactive (Kim et al., 2004). Residues of the active site of CBAH were identified by sequence alignments of penicillin V acylase (PVA) with BSH (Tanaka et al., 2000), and active site residues in PVA were identified by comparison with penicillin G acylase (Suresh et al., 1999), of which enzyme-inhibitor complexes were described (Duggleby et al., 1995). Besides Cys2, residues Asp21, Asn82, Asn175, and Arg228 were identified as catalytically important in BSHs (Fig. 40). PVA uses the same catalytic residues except for Asn82 which is replaced by Tyr82. This, however, does not alter the nature of the active site since only the peptide NH atoms of both residues are important in catalysis by providing a hydrogen bond donor of the oxyanion hole as does Asn175ND2.
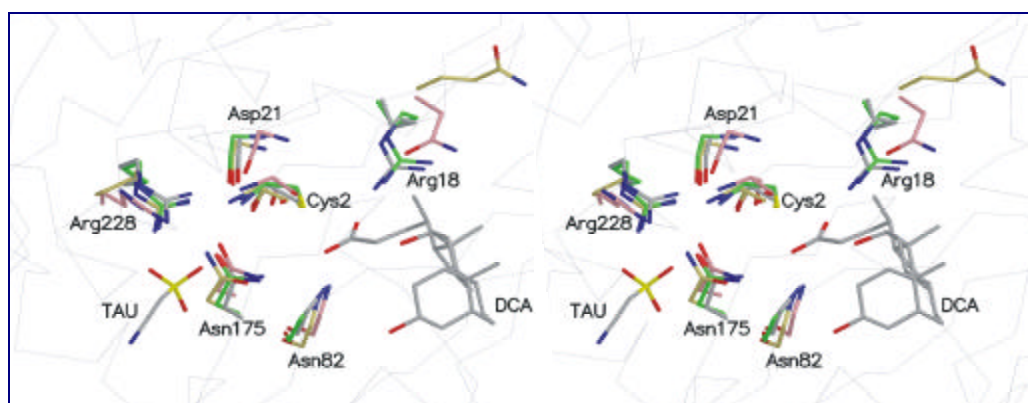
**Figure 40. Superimposition of active site residues (stereo).** The superimposition is based on the residues listed below with the exception of Arg18 in CBAH (or respective residues in the other proteins). Only main-chain atoms of Asn82 and Asp21 are shown.

| CBAH | Cys2 | Arg18 | Asp21 | Asn82 | Asn175 | Arg228 |
|------|------|-------|-------|-------|--------|--------|
| PVA | Cys1 | Arg17 | Asp20 | Tyr82 | Asn175 | Arg228 |
| CA | Ser170 | Gln189 | His192 | Val239 | Asn413 | Arg443 |
| PGA | Ser1 | Asn20 | Gln23 | Ala69 | Asn241 | Arg263 |

CA = Cephalosporin acylase from *Pseudomonas diminuta* (PDB ID 1fm2), PGA = Penicillin G acylase from *E. coli* (PDB ID 1gm9), PVA = penicillin V acylase from *Bacillus sphaericus* (PDB ID 3pva).

The geometry of the CBAH active sites is well conserved, and the residues of the active site in CBAH superimpose well on the active sites of other members of the Ntn-hydrolase superfamily (except for residue Arg18) including cephalosporin acylase (CA) and penicillin G acylase (PGA). Notably, Arg228 on $\alpha5$ is close to the catalytic Cys2 and forms hydrogen bonds to Asn175OD. From the structure it is not clear whether this apparent stabilization of the active site geometry is the only function of Arg228, but arginines are conserved in equivalent positions in all described Ntn-hydrolases and have been shown to be necessary for catalytic function (Isupov et al., 1996; McVey et al., 2001; Prabhune and Sivaraman, 1990).

It is of interest that another arginine, namely Arg18 in the *C. perfringens* CBAH structure, superimposes well with Arg18 from PVA (Fig. 40), but the side chains of the equivalent residues point in different directions in the more distantly related proteins

PGA and CA, which both feature N-terminal serine instead of cysteine in CBAH and in PVA. Since in all four proteins the hydrolysis of the respective substrates results in a carboxyl group to which Arg18 binds in CBAH models with products, it is likely that conservation of Arg18 is due to its involvement in catalysis rather than in conferring substrate specificity. Sequence alignments have suggested earlier that Arg18 of PVA might be involved in catalysis (Pei and Grishin, 2003). Interestingly, in glutamine amidotransferase (Smith et al., 1994) and in glucosamine 6-phosphate synthase (Isupov et al., 1996), two other members of the Ntn-hydrolase superfamily that act via a nucleophilic cysteine, the position of Arg18 is occupied by His70 and His71 respectively, which can fulfil a similar role as Arg18 if they are protonated. Remarkably, the CBAH variant R18H possessed 77% activity of the wild type, whereas the exchange of Arg18 by Leu18 deactivated CBAH completely (Table 13). CBAH R18L and apoCBAH are virtually identical proteins (r.m.s.d. for Cα - 0.4 Å), whereas Leu18 superimposes well with Arg18 in apoCBAH (Fig. 41).
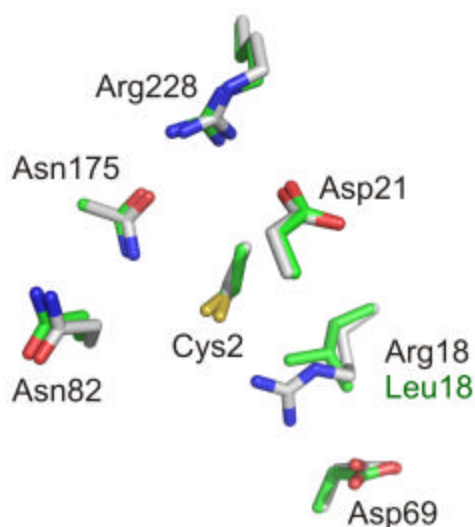


**Figure 41.** Superpimposition of active site residues of wild-type apoCBAH (C-atoms in grey) and CBAH R18L (C-atoms in green) based on all protein atoms. Residue Leu18 superimposes well with Arg18 in wild-type CBAH.

In CBAH and PVA that act at low pH (< 6), serine in the active site, typical nucleophile in Ntn-hydrolases, is replaced by cysteine, carrying more nucleophilic sulfhydryl

group. Arg18 being in the vicinity of Cys2SG seems to lower the $pK_a$ of the SH-group. Increasing the nucleophilicity of the sulfhydryl group is especially important for enzymes which act at low pH, to which CBAH and PVA belong. In contrast, CA and PGA act at high pH (7.8 and 8.2, respectively) and use the hydroxyl group of serine to attack the carbonyl group.

Using Monte-Carlo simulations, the theoretical $pK_a$ value of the SH-group of apoCBAH was calculated to be about 2 (personal notice from Dr. G. Kieseritzky) and mostly deprotonated at pH 5.5. The $\alpha$-amino group of Cys2, which forms the hydrogen bond to Asp21, was calculated to be protonated. In contrast, using the same program settings, which were calibrated by calculation of known $pK_a$ values of serine proteases, the N-terminus of the CA serine Ntn-hydrolase CA was calculated to be deprotonated. These results are in agreement with the catalytic mechanism proposed for CA and PGA (Duggleby et al., 1995), where serine in the Ntn-hydrolases CA and PGA require free uncharged N-terminal $\alpha$-amino group for their catalytic activity. The uncharged state of the amino group is maintained at high pH by two asparagine side chains. Thus, the mechanism of serine Ntn-hydrolases acting at high pH differs from that of cysteine Ntn-hydrolases.

### 4.2.11 Processing of CBAH

For a significant fraction of the bacterial proteins the N-terminal methionine is removed enzymatically by methionine amino-peptidase. Such a processing seems not to be that effective for CBAH, as the mutation of the active site residues resulted in mostly insoluble CBAH variants containing the N-terminal methionine (Fig. 26). Surprisingly, in crystals of of the CBAH C2A variant the initial methionine of monomer B was not removed from neither by native processing nor autocatalytically and provided insight into CBAH processing. In Figure 42 the active site of monomer B of the CBAH mutant C2A is shown. Superimposing the structures of apoCBAH and CBAH C2A variant provides insight into the processing of Met1 (Fig. 42). Cys2 would be well positioned for nucleophilic attack on the amide bond between Met1 and Cys2. As the $\alpha$-amino group of nucleophile Cys2 forms the peptide bond with Met1 and so is not available as a general base, it seems to be replaced by Arg18 which in turn forms a hydrogen bond with Asp69. The triad Cys2-Arg18-Asp69 seems to be a second catalytic centre responsible for the processing of the CBAH.
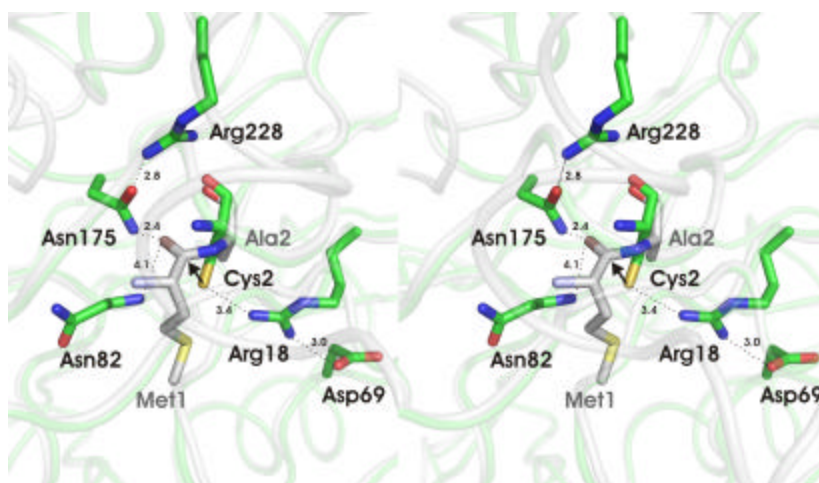
**Figure 42. Proposed second catalytic center in CBAH. Superposition of monomer B of CBAH C2A and wtCBAH (stereo).** In the C2A variant (grey) where Cys2 is replaced by Ala2, the N-terminal nitrogen is involved in the peptide bond with methionine and cannot serve as base. Cys2 is activated by the basic residue Arg18 and attacks amide bond between Met1 and Cys2. The $pK_a$ of Arg18 seems to be adjusted by Asp69. The catalytic triad Cys2-Arg18-Asp69 is assumed to be responsible for the processing of CBAH. The variant R18L was inactive. In contrast, R18H showed only reduced catalytic activity (Table 13). Arg18 seems to be involved in processing catalytic triad.

## 4.2.12   Initial substrate binding by CBAH

When CBAH was cocrystallized with choloylglycine, the reaction products cholate and glycine were bound to CBAH separated by a distance of 6.6 Å between the carboxylate group of cholate and the $\alpha$-amino group of glycine. These positions of the products suggest that the substrate choloylglycine will have a different position and orientation after the initial binding and, the reaction products will be disjoined after the reaction, probably accompanied by conformational changes in the protein. To investigate this process, the initial choloylglycine binding to apoCBAH was simulated using program AUTODOCK. The results indicated that the amide NH-group may form a hydrogen bond to the choloylglycine molecule oxygen of the Asn82 side chain (Fig. 43) and probably remains bound to it also after the reaction, as in the crystal structure of CBAH with different products the side chain of Asn82 was shown to bind the reaction product (taurine, glycine or sarcosine (not shown)) (Fig. 35) via

hydrogen bonds. At the same time the Asn82 main chain hydrogen is part of the proposed oxyanion hole in CBAH. It is tempting to speculate that Asn82 is involved in the guidance of the substrate to the active site.
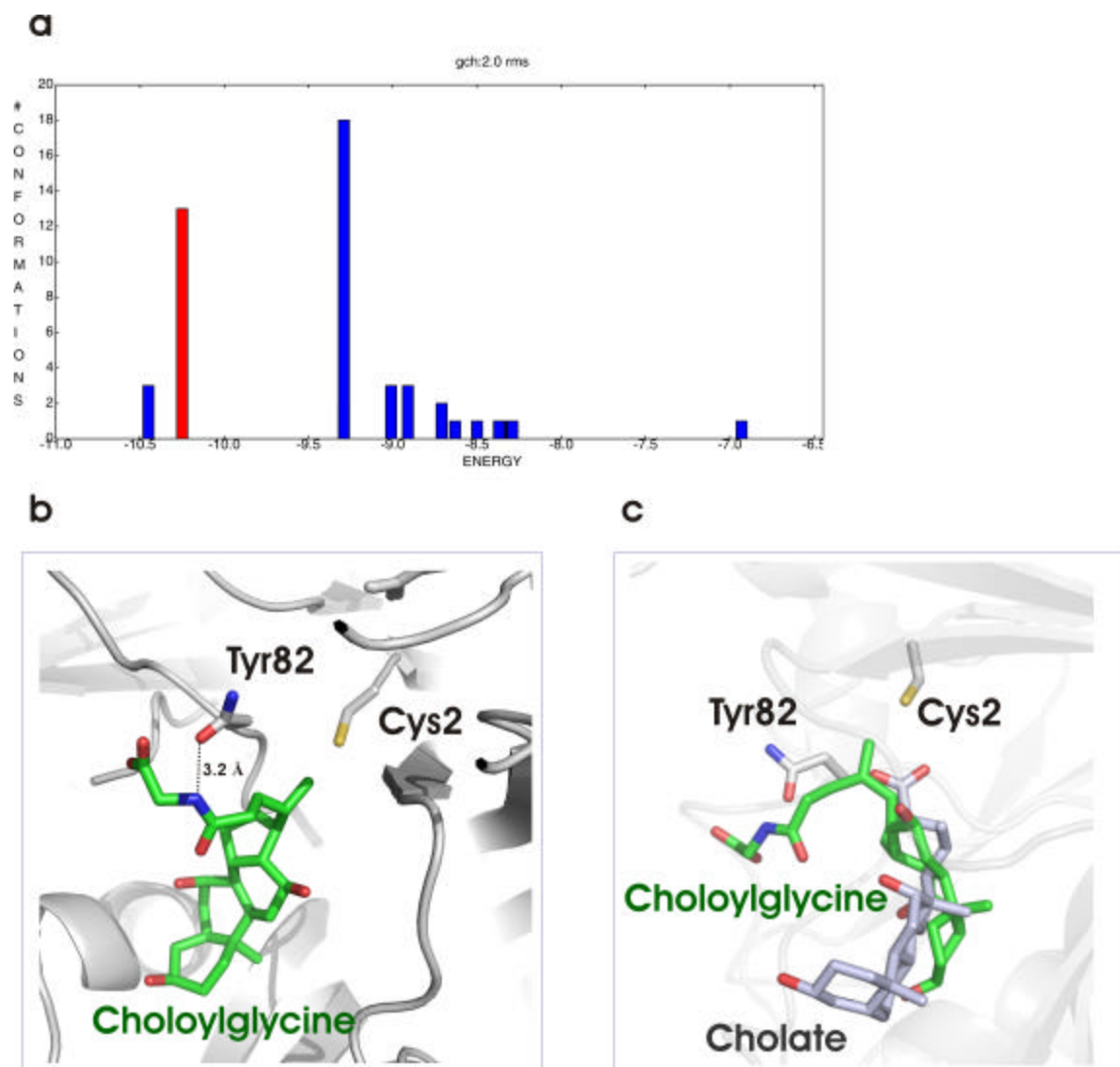


**Figure 43. Initial choloylglycine binding to CBAH simulated by AUTODOCK.** *(a)* Clustering of the solutions calculated by AUTODOCK. The most probable solution with lowest energy and the most populated cluster is indicated in red. *(b)* Position of choloylglycine (green) corresponding to the most probable solution from AUTODOCK. Residue Asn82 is involved in the initial substrate binding and is a part of the oxyanion hole. *(c)* Superimposing of the best solution from AUTODOCK and position of cholate (grey) found in the CBAH crystal structure cocrystallized with choloylglycine.

The binding of the leaving group to CBAH seems to be not that specific. As shown in this work it can bind taurine, glycine, sarcosine and glycerol at this position. Even larger leaving groups such as 6-APA (s. point 1.11) can be accepted, as CBAH was shown to have activity toward penicillin V (Kumar et al., 2006). Probably, the binding groove for the β-lactam ring could be further increased through site-directed mutagenesis and an exchange by Asn82 to Tyr82 could improve the specificity of CBAH towards 6-APA.
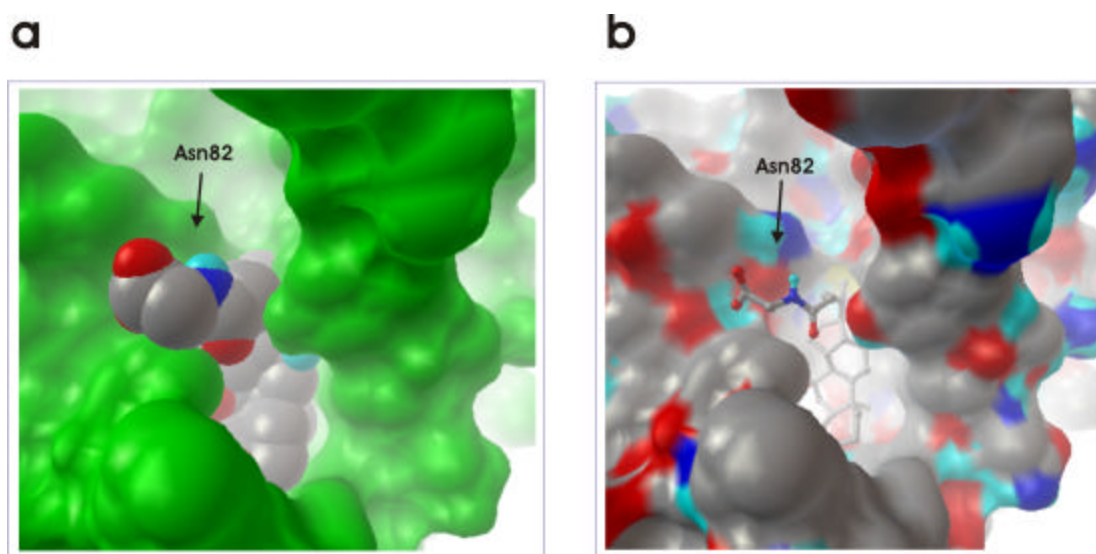


**Figure 44. Docking of the GCA into CBAH.** *(a)* CBAH surface is shown in green. The cholate core fits into the binding groove, whereas the hydrophilic side chain is more solvent exposed and recognized only by the side chain of Asn82 (indicated by arrow). *(b)* Same view as (a), the choloylglycine molecule is presented as a ball-and-stick model. The surface of CBAH is coloured according to electrostatic charge.

## 4.2.13   Product binding by PVA

No crystal structure of PVA in complex with its reaction products or substrates is available. AUTODOCK was used to calculate the probable mode of binding of 6-APA and phenyl acetate to the PVA. The coordinate and topology files of the reaction products were generated with the PRODRG server. The most probable conformations belonging to the most populated cluster with lowest energy were selected, and the results are presented in Figure 45.
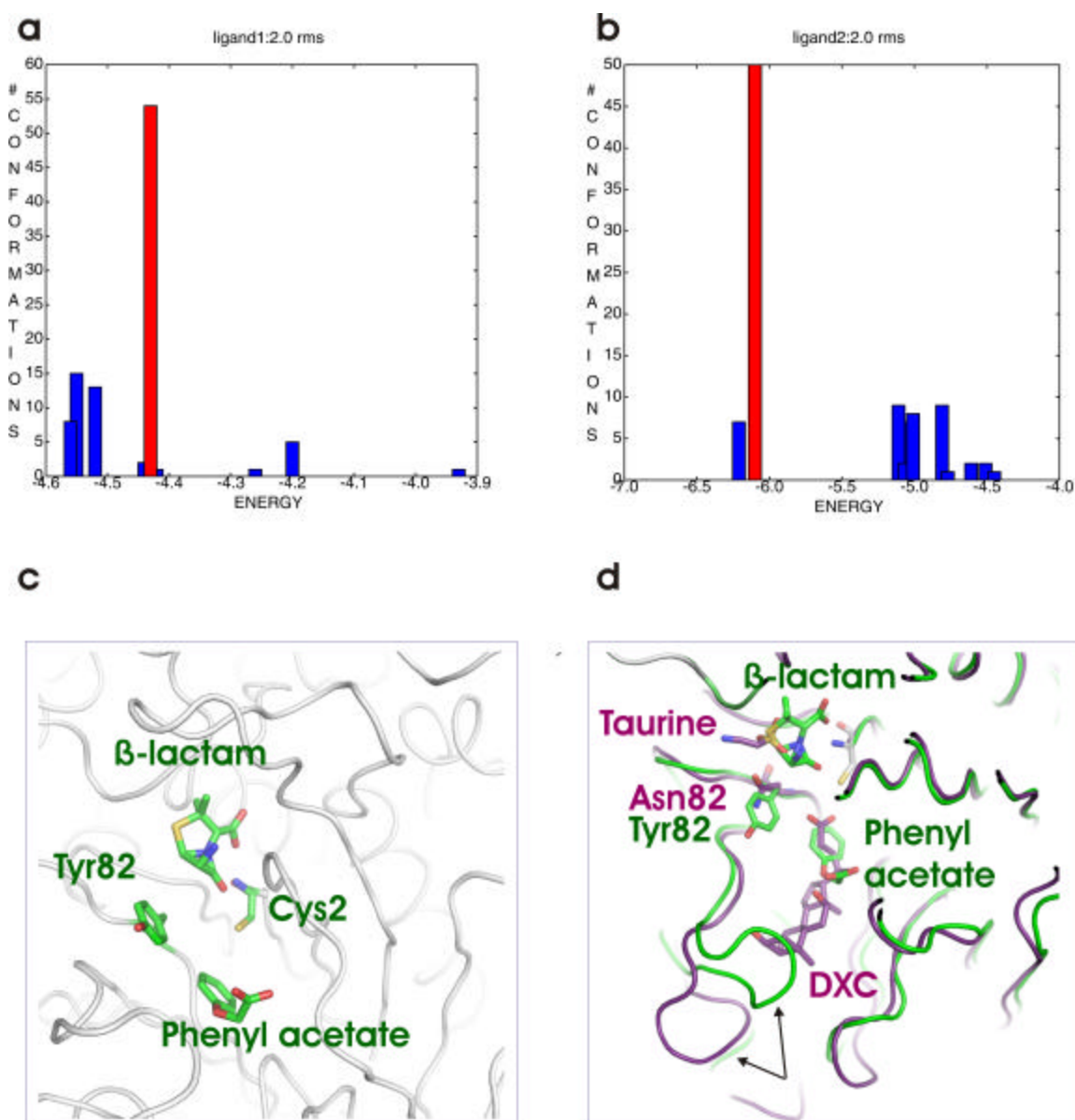
**Figure 45. Initial binding of the reaction products phenyl acetate and 6-APA to the PVA simulated by AUTODOCK. (a)** Clustering of the solutions for 6-APA and **(b)** clustering of the solutions for phenyl acetate calculated by AUTODOCK. The most probable solutions with lowest energy and the most populated clusters are indicated in red. **(c)** Positions of the reaction products in PVA corresponding to the most probable solutions. Residue Tyr82 seems to be involved in the initial substrate binding and is a part of the oxyanion hole. **(d)** Superimposition of the solutions from AUTODOCK (green) and positions of products found in the crystal structure of CBAH cocrystallized with choloyltaurine (magenta). Flexible loop between residues 129-141, which is closed in the structure of PVA, is indicated by arrows. DXC – deoxycholate.

RESULTS

The docking simulations performed with PVA and its reaction products revealed that the aromatic ring of Tyr82 could be (similar to Asn82 in CBAH) involved in the binding of the leaving group, which is a heterocyclic ring in the case of PVA. Thus, the side chain in position 82 seems to be important for substrate recognition both in CBAH and PVA. The phenyl acetate molecule was calculated to be buried in the small hydrophobic cavity close to the reaction centre (Fig. 45d).The size of this cavity is smaller compared to the binding cavity of CBAH for the cholate moiety. The reason for the different conformation of the loop is that in PVA the loop between residues 129-141 is more hydrophobic than in CBAH (Table 14) and it clasps into the hydrophobic substrate-binding groove. The phenyl acetate molecule was calculated to be buried in the small hydrophobic cavity close to the reaction centre (Fig. 45d). The size of the cavity is smaller compared to the binding groove of CBAH for the cholate moiety.

**Table 14.** Amino-acid composition of the loop between amino acids 129-141 in PVA and CBAH. Exchange of amino acids from polar in CBAH to hydrophobic in PVA and from hydrophobic in CBAH to polar in PVA are indicated by red and green boxes, respectively.

| Pos. | 129 | 130 | 131 | 132 | 133 | 134 | 135 | 136 | 137 | 138 | 139 | 140 | 141 | 142 |
|------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| CBAH | V | D | I | P | I | S | E | N | I | P | N | T | T | L |
| PVA | L | N | E | A | N | I | I | L | G | I | A | P | P | L |

## 4.2.14 Substrate selectivity of CBAH

Bile acids are conjugates between the amino acids taurine or glycine and a cholate derived from cholesterol. Deoxycholate and cholate are always hydroxylated on C3 in a βosition. In addition, the most common hydroxyl substituents are none at all (in lithocholic acid), 12a?(in deoxycholic acid),?7a (in chenodeoxycholic acid), 7β?in ursodeoxycholic acid), 7a, 12a?(in cholic acid) (Hofmann et al., 1992). Together, this results in ten combinations between differently hydroxylated cholates and the two

amino acids. It is conceivable that bile salt hydrolases have evolved to recognize bile acids on both the amino acid and cholate groups.

Deoxycholate is bound primarily by hydrophobic interactions, and the hydroxyl substituents are not recognized through hydrogen bonds formed with the enzyme. The only hydrogen bonds between protein and deoxycholate are to the carboxylate group of the isovaleric acid substituents, O25 and O26. This is in agreement with earlier experiments using epimerized hydroxyl substituents on cholate that did not influence substrate binding (Batta et al., 1984).

Residues of the active site are strictly conserved in bile acid hydrolases (asterisks in Fig. 32). By contrast, however, the residues for substrate recognition (grey in Fig. 32) are not particularly conserved although most amino-acid substitutions are conservative. The notable exception is Leu142 that is strictly conserved throughout the species and occurs even in penicillin V acylase which acts on completely different substrates.

Although the substrate specificity of bile salt hydrolases seems to be broad, small differences in $K_m$ values have been observed. For example, in CBAH the $K_m$ values for chenocholoylglycine and choloyltaurine are in the range of $10^{-2}$ M whereas that for deoxy-conjugates is generally in the $10^{-3}$ M range (Nair et al., 1967), suggesting that CBAH has a higher affinity towards deoxy-conjugates. A possible explanation for this difference could be an additional hydrogen bond formed between cholate and the side chain of Thr140, which could cause the more slowly release of the product from the enzyme and could gate the binding of the next substrate molecule.

To learn more about the substrate binding in CBAH and PVA, the channels leading from the active site to the surface were calculated using CAVER (Fig. 46). Positions of substrates determined in the CBAH crystal structure superimposed well with channels calculated by CAVER, whereas the volume of the channel for the cholate molecule was calculated to be larger than for the leaving group, approving the correctness of the calculations (Fig. 46a). Using the same input parameters, channels in PVA were calculated and revealed that the position of the docked phenyl acetate superimposed well with the by channel calculated by CAVER. The location of the leaving group (6-APA) in PVA was simulated not that exactly using AUTODOCK and superimposed only partially with the calculated channel (Fig. 46b).
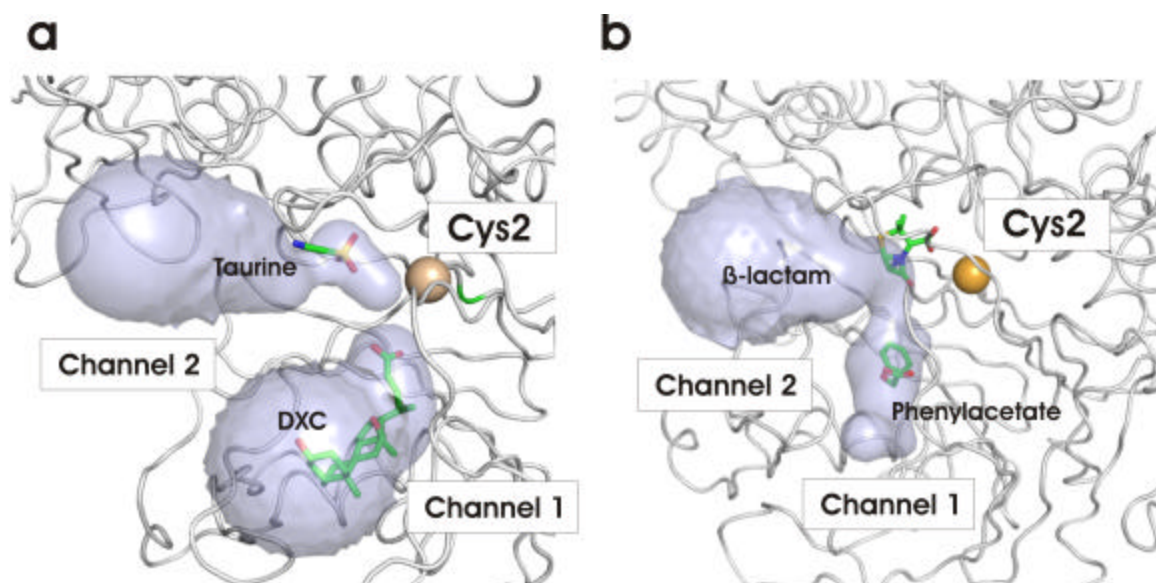
**Figure 46. Channels calculated with program CAVER.** *(a)* The positions of products taurine and deoxycholate (DXC) found in the crystal structure superimpose well with 2 channels calculated using CAVER. *(b)* Hydrolytic products phenyl acetate and 6-APA were docked to the PVA structure using AUTODOCK. Two channels leading from the active site to the surface were calculated with CAVER. The channel for the core (phenyl acetate) is smaller than in CBAH and superimposes well with the position of the phenyl acetate calculated by AUTODOCK. The position of 6-APA (β-lactam) seems to be not correctly calculated by AUTODOCK.

Since the binding pocket for the phenyl acetate in PVA is smaller than the pocket in CBAH, it cannot accept larger substrates. The reason for the small binding pocket in PVA is the loop 129-141, which is rich in hydrophobic residues that interact with the hydrophobic residues in the binding pocket, thereby covering the hydrophobic cavity of the binding pocket (Fig. 45d, Table 16). In CBAH the loop 129-141 contains polar residues which are exposed to solvent, rendering loop more open. Thus one could reduce the specificity of PVA by mutation of the loop residues to more polar amino acids as in CBAH.

With respect to the leaving group, its binding in CBAH is not very specific, as different amino acids and other polar molecules can be accepted that bind to the protein only through Asn82. The leaving group is also not recognized in most Ntn-hydrolases and thus offers good possibilities to design enzymes with wide specificities for leaving groups.

### 4.2.15 CBAH inhibitors

*N*-peptidyl-*O*-acylhydroxylamines were originally designed to inhibit serine proteases, in particular dipeptidyl peptidase IV (Fischer et al., 1983). It was later shown that this class of inhibitors strongly inactivates also cysteine proteases (Smith et al., 1988). The mechanism of inactivation is not yet fully understood and depends on the nature of the enzyme and on pH. In general, peptidyl hydroxamates are more potent inhibitors of cysteine proteases than of serine proteases. The differences can be attributed to the strength and reactivity of the cysteine thiole as nucleophile.

It is tempting to assume that the same inhibition principles can be assigned to the Ntn-hydrolases. *N*-deoxycholoyl-*O*-acylhydroxylamines could be suitable inhibitor of CBAH and could inhibit it in a similar manner proposed for cysteine proteases at low pH (Menard et al., 1991), Fig. 47.
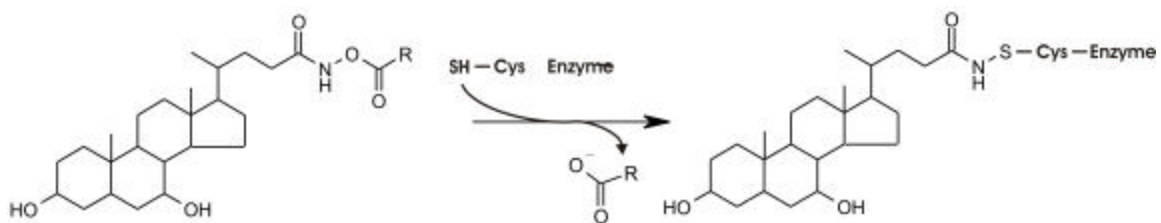


**Figure 47. Proposed mechanism for inhibition of CBAH by *N*-deoxycholoyl-*O*-acylhydroxylamines.** The proposed mechanism involves the formation of the thiohydroxylamine derivative covalently bound to the enzyme.

## 4.3  Acid ceramidase (ASAH)

### 4.3.1  CD spectroscopy of CBAH and ASAH

The secondary structure content of acid ceramidase and CBAH was investigated using CD spectroscopy. Figure 12 shows the far-UV circular dichroism (CD) spectrum of purified CBAH and ASAH at 25°C. The secondary structure estimations for CBAH and ASAH are presented in Table 15. The secondary structure calculation for CBAH using program CDSSTR agreed very well with the secondary structure derived from the crystallographic data. The same program and parameters were also used to

deconvolute the CD spectrum of ASAH. The CD data revealed that purified ASAH is a protein of the mixed $\alpha$-helix- and -$\beta$-strand type, and that the secondary structure of ASAH is distributed similar to that of the bacterial homologue CBAH (Table 15).
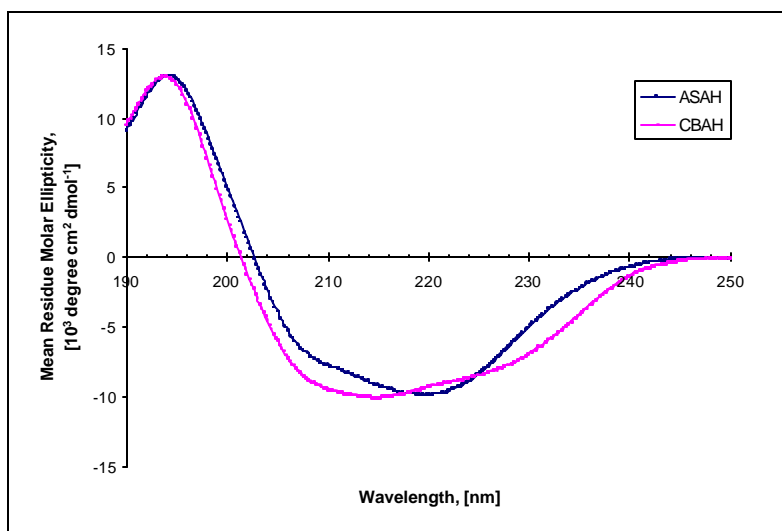


**Figure 48. Far-UV spectra of ASAH and CBAH.** The CD spectra of ASAH and CBAH were aquired in 0.1 mM sodium acetate, pH 4.5, 3 mM NaCl, 40 μg/ml, and 0.1 mM sodium acetate, pH 5.5, 8 mM NaCl, 0.1% glycerol, 37.5 μg/ml CBAH, respectively.

**Table 15.** Relative secondary structure content (%) of CBAH and ASAH estimated by program CDSSTR.

|  | Helix | Strand | Turn | Not ordered |
|---|---|---|---|---|
| CBAH | 18 (22)[1] | 29 (32) | 21 | 32 |
| ASAH | 19 | 29 | 27 | 25 |

[1]Data in parentheses refer to crystallographic data

**4.3.2 Crystallization of ASAH and X-ray data collection**

Crystallization of ASAH (3.3.2) yielded needle-like crystals in the presence of PEG 4000 (Fig. 49). The crystal dimensions were approximately 900 µm x 10 µm x 10 µm. X-ray diffraction image of ASAH crystals were measured at ESRF, Grenoble (ID14-2). The crystals were found to belong to space group C2 with unit-cell parameters a = 170 Å, b = 150 Å, c = 130 Å.
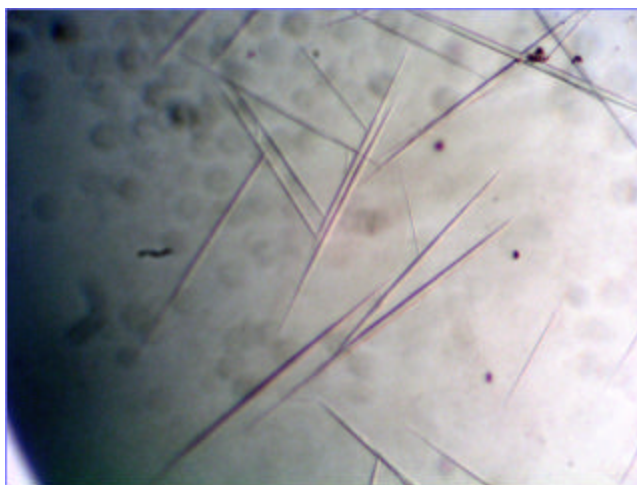


**Figure 49. ASAH crystals.** Typical needle-like ASAH crystals with dimensions of approximately 900 µm x 10 µm x 10 µm grew in the presence of PEG 400 and PEG 8000.

ASAH crystals diffracted X-rays very weak and anisotropically. Although in some orientations crystals diffracted X-rays to up to 4 - 8 Å resolution (Fig. 50), it was not possible to collect a complete data set and only several images were collected. Because of limited amounts of available protein, further crystallization trials were not performed.
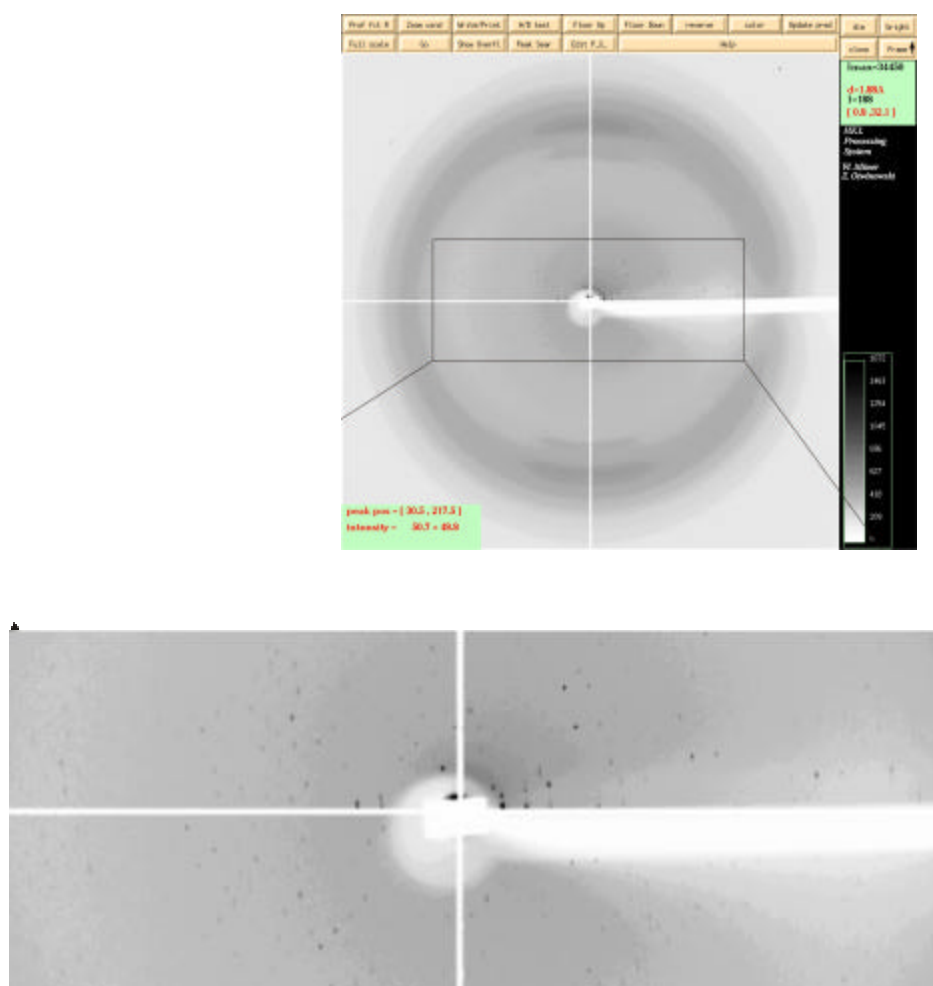
**Figure 50. Diffraction image of ASAH collected at ID 14-2, ESRF, Grenoble.**
ASAH crystals diffracted weak and anisotropically to max. resolution of 4 Å.

### 4.3.3 Homology modelling of the acid ceramidase subunits

The ASAH protein sequence was used to perform a search for homologous sequences using the automated fold-recognition method implemented in the program Phyre-3D-PSSM (http://www.sbg.bio.ic.ac.uk/3dpssm/). The program uses a combination of multiple-sequence profiles with structure-based profiles, like secondary structure and salvation potential. This feature is useful for alignments of proteins with remote homology. The $\alpha$-subunit of ASAH was found to be weakly homologous (12%) to the all-$\alpha$ helix HIV-1 surface protein. However, the primary and secondary structure homology was too low to perform the comparative modelling by MODELLER.

The best score provided by the 3D-PSSM server for the amino-acid sequence of the β-subunit of ASAH found for the structures of CBAH and PVA. The sequences of ASAH and CBAH possess an identity of about 20 %, and a reasonable alignment was obtained that was used for homology modelling using MODELLER. The calculated homology model of the β-subunit of ASAH is presented in Figure 51a.

### 4.3.4   Validation of the homology model of the b-subunit

The stereochemical quality of the ASAH model was validated by WHATIF and PROCHECK. The structure presented a relatively good percentage of residues in the "most favoured" and –"additionally" allowed regions of Ramachandran plot (82% and 16%, respectively), while 2% (4 AS) residues were found in the "generously" allowed and 1% in "disallowed" regions which are located in the loops where the modelling process is knowingly more difficult.

To further validate the quality of the obtained model, the interaction energy per residue was calculated using program PROSA II (Fig. 52). The PROSA *Z-score* of the ASAH model structure was (-2.3), indicating an acceptable quality of the model. The PROSA energy plot (Fig. 52) revealed that the amino acids in the region between residues 90 to 170 of ASAH showed elevated interaction energies, indicating that this part could be involved in the interface between ASAH subunits a and β. This region contains two cysteines, Cys150 and Cys198, which are solvent exposed in the model (Fig. 51). Cys198 was shown to form a disulfide bridge with Cys10, connecting herewith β- and α-subunits (Schulze et al., 2007) covalently with each other and supporting the notion that the region 90-170 is an interface between two ASAH subunits.
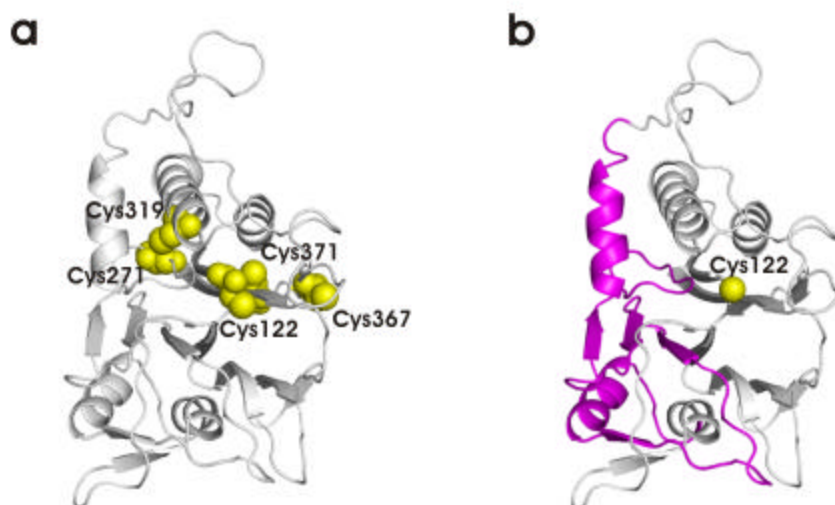
**Figure 51. Structure model of the b-subunit of ASAH**. *(a)* Positions of all 6 cysteines in the β-subunit model of ASAH shown to be involved in 3 disulfid bridges are located on the surface approving the correctness of the model. (side chains are presented as spheres and coloured yellow). *(b)* Regions of the ASAH model with higher energies (Fig. 52, right panel) calculated with PROZA are coloured pink. The N-terminal nucleophile Cys122 is shown as a yellow sphere.

ASAH contains two further cysteines (Cys246 and Cys250). They lie at the C-terminus and were shown to be connected with each other by a disulfide bridge (Schulze et al., 2007). Since homology-modelling programs do not allow a calculating of disulfide bridges not present in the basis model, this disulfide bridge was not modelled. This could be the reason why the C-terminus of the model shows higher interaction energies levels (Fig. 52).
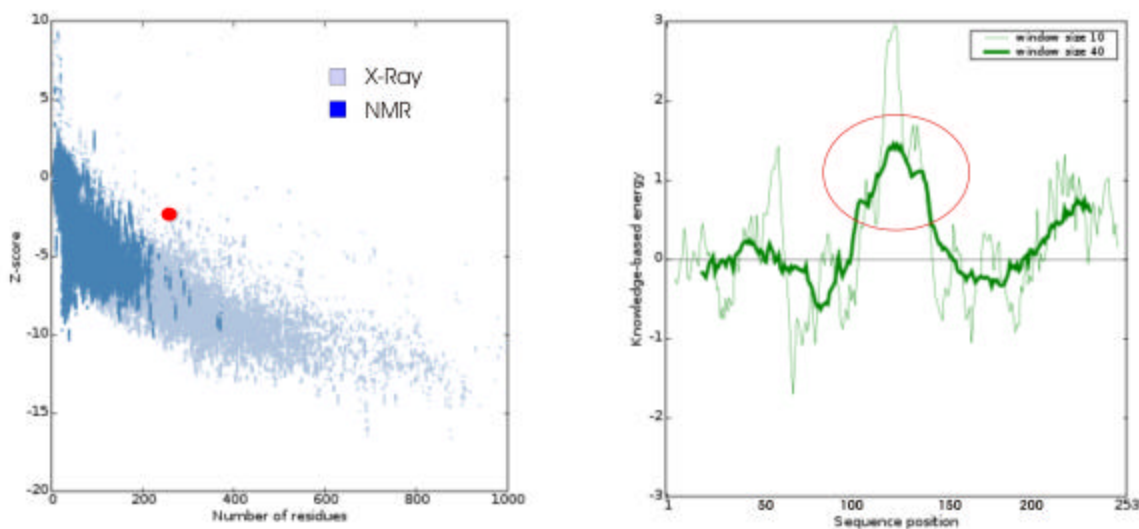
**Figure 52. PROSA Z-score and energy plot.** The PROSA Z-score of -2.3 (red dot) indicated a quite good model quality, as it is close to the values observed in crystal structures (left panel). The plot of the interaction energy versus residue (right panel) indicated elevated energies in the sequence region 90-170 (red circle), indicating that this part of ASAH could be the interface between ASAH subunits a and β or is involved in oligomerization.

To validate further the reliability of the homology model the positions of the six known glycosylation sites were analyzed. Most glycosylation sited were found to be solvent exposed in the model indicating their correct location (Fig. 53a).

### 4.3.5 Proposed active site of ASAH

Due to the nucleophilic cysteine at the N-terminus of the β-subunit, autoprocessing ability, sequence homology to the Ntn-hydrolases, and the fact, that ASAH cleaves amide bonds, support the notion that ASAH belongs to the family of the Ntn-hydrolases.
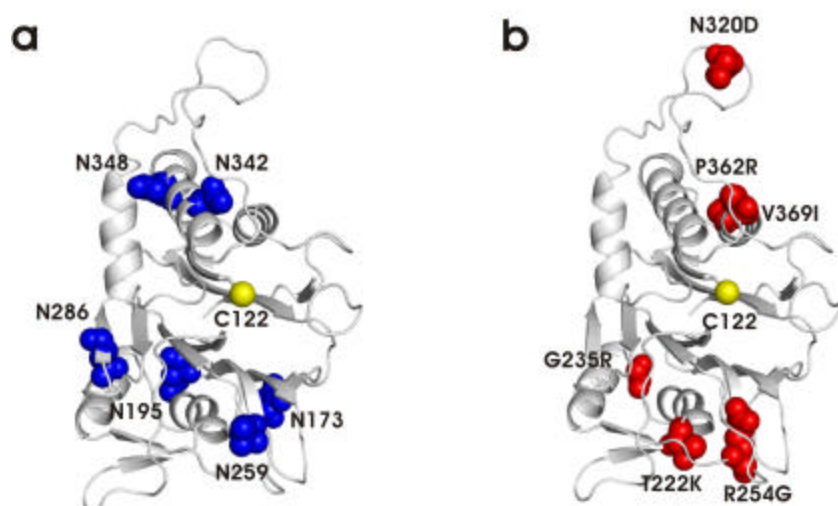
**Figure 53. Homology structure of the b-subunit of ASAH.** *(a)* Positions of the glycosylation sites in the homology structure. All six glycosylation sites in the β-subunit model of ASAH are located on the surface (side chains are presented as spheres and coloured blue) approving the correctness of the model. Side chains of asparagines are shown as blue balls. *(b)* Positions of the five known mutations in the β-subunit of the ASAH model are located on the protein surface. Side chains of the mutated amino acids are presented as red spheres.

However, Schulze et al. were not able prove the existence of the free N-terminal cysteine in the β-subunit using mass spectroscopy. The N-terminal cysteine Cys122 was shown to be involved in a disulfide bridge with Cys271 (Fig. 51a) in the catalytically active protein recombinantly expressed in Sf 9 cells. According to the homology model of ASAH formation of such a disulfide bridge is possible, although this will require a reorganization of the region 90-200, because the distance between Cys122 and Cys271 will be shorter than in the homology model. In this case the nucleophilic OH-group of the residue Ser3 is a possible candidate able to replace the sulfhydryl group of the Cys1, whereas propositions in respect other active side residues would be highly speculative.
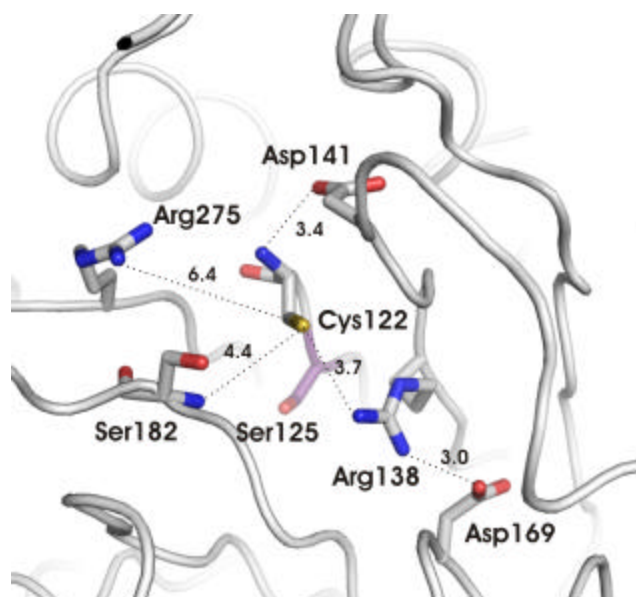
**Figure 54. Proposed active site residues in the b-subunit model of ASAH.** A catalytic Ntn-diad is proposed to be formed by amino acids Cys122 and Asp141. One of the amino acids forming the oxyanion hole assumed to be Ser182, the second residue forming the oxyanion hole could be Arg275. In analogy to CBAH, residues Arg138 and Asp169 are assumed to be a part of the second catalytic center responsible for the processing of the acid ceramidase.

The observation that reducing agents like DTT stimulate the ASAH activity (Schulze et al., 2007), suggests the participation of the conserved N-terminal Cys122 in the reaction. The disulfide bridge formation found in ASAH (Schulze et al., 2007) preparations seems to be characteristic for the standard purification protocol of ASAH without using DTT and could be a native mechanism protecting the active site cysteine against spontaneous oxidation observed in BSH (Kumar et al., 2006) and CBAH (Fig. 27).

The obtained homology model of the $\beta$-subunit was used to analyze the positions of the five known pathological mutations of the ASAH $\beta$-subunit (Park and Schuchman, 2006). As shown in Fig. 53a, none of the mutations affected the probable active site proposed in Fig. 54 based on the homology to the family of Ntn-hydrolases.