

## Title

The mechanism of sphingolipid processing revealed by a GALC-SapA complex structure

## Authors

Chris H. Hill <sup>1,2</sup>, Georgia M. Cook <sup>1</sup>, Samantha J. Spratley <sup>1</sup>, Stephen C. Graham <sup>3</sup> and Janet E. Deane <sup>1\*</sup>

## Affiliations

<sup>1</sup> Cambridge Institute for Medical Research, Department of Pathology, University of Cambridge, Cambridge Biomedical Campus, Hills Rd, Cambridge CB2 0XY, UK

<sup>2</sup> Current address: MRC Laboratory of Molecular Biology, Francis Crick Avenue, Cambridge Biomedical Campus, Cambridge CB2 0QH, UK

<sup>3</sup> Department of Pathology, University of Cambridge, Tennis Court Road, Cambridge CB2 1QP, UK

\* Corresponding author: [jed55@cam.ac.uk](mailto:jed55@cam.ac.uk)

## Abstract

Sphingolipids are essential components of cellular membranes and defects in their synthesis or degradation cause severe human diseases. The efficient degradation of sphingolipids in the lysosome requires lipid-binding saposin proteins and hydrolytic enzymes. The glycosphingolipid galactocerebroside is the primary lipid component of the myelin sheath and is degraded by the hydrolase  $\beta$ -galactocerebrosidase (GALC). This enzyme requires the saposin SapA for lipid processing and defects in either of these proteins causes a severe neurodegenerative disorder, Krabbe disease. Here we present the structure of a glycosphingolipid-processing complex, revealing how SapA and GALC form a heterotetramer with an open channel connecting the enzyme active site to the SapA hydrophobic cavity. This structure defines how a soluble hydrolase can cleave the polar glycosyl headgroups of these essential lipids from their hydrophobic ceramide tails. Furthermore, the molecular details of this interaction reveal how specificity of saposin binding to hydrolases is encoded.

## Introduction

Sphingolipids are both essential membrane components and bioactive metabolites that regulate critical cell functions. Defects in sphingolipid metabolism underlie a range of diseases, including lysosomal storage diseases, and are implicated in a number of cancers (1-3). Sphingolipid degradation occurs in the lysosome and depends upon two families of proteins: glycosyl hydrolases, and lipid-transfer proteins including saposins and GM2 activator proteins (4-7). The hydrolases are water soluble while the substrates are embedded in lysosomal intraluminal vesicle membranes. The steric crowding of headgroups and lateral association of sphingolipids into clusters prevents hydrolases from accessing the scissile bonds of their target substrates. Saposins are required to facilitate sphingolipid processing by hydrolases, and extensive work provides evidence for two general models of saposin action (8-11). The ‘solubiliser’ model proposes the complete extraction of lipids from the membrane, while the ‘liftase’ model envisages saposin proteins binding directly to compatible membranes and improving lipid accessibility by membrane distortion, destabilisation or localised remodelling. The saposin proteins are produced as a polycistronic prosaposin protein that, upon delivery to the lysosome, is cleaved into the four saposins: A, B, C and D (12-14). The functions of these four saposins are distinct, as they cannot compensate for the loss of each other, and saposins appear to function specifically with their associated hydrolase (15, 16). Lysosomal storage diseases, and more specifically sphingolipidoses, are caused by mutations that inhibit degradation of sphingolipids. For example, Gaucher and Krabbe diseases are caused by loss of glucocerebrosidase (GluC) and galactocerebrosidase (GALC) activity, respectively, or by mutations in their associated saposins SapC and SapA, respectively (17-20).

Several high-resolution structures of saposins have been determined to date, revealing huge conformational variability and a propensity to form oligomeric assemblies (21-25). Saposins can broadly be described as existing in two states: a “closed” monomeric form where the helical protein folds back on itself, burying a large hydrophobic core; or a more “open” dimeric form, possessing a hydrophobic cavity into which lipids and detergents can bind. A recent structure of SapA revealed that this open conformation can form lipo-protein discs (21) and these discs have recently been exploited to aid the determination of challenging membrane protein structures by crystallography and CryoEM (26). However, it has remained unclear what form of saposin-lipid complex mediates binding to their cognate hydrolase. To remedy this deficit we have solved the structure of GALC in complex with SapA, defining how saposins solubilise lipids for processing by soluble hydrolases.

## Results

Murine GALC and SapA were expressed and purified from mammalian cells and *E.coli*, respectively. Pulldown assays reveal that SapA does not bind GALC at neutral pH, equivalent to that of the endoplasmic reticulum, but instead these proteins form a complex at low pH equivalent to that of the endolysosomal compartments where lipid degradation occurs (Fig. 1A). This interaction depends upon the presence of detergents, suggesting that the “open” (dimeric) lipid-bound form of SapA mediates the interaction (Fig. 1B). The pH-dependency of the interaction, combined with the low pI of SapA (pH 4.5) and the higher pI of GALC (pH 6.1), suggests that the interaction could be mediated by electrostatic interactions. In support of this, the interaction is sensitive to high salt concentrations as the presence of increasing concentration of NaCl reduces binding in these pulldown assays (Fig. 1B).

Based on these insights, the GALC-SapA complex was formed at pH 5.0 in the presence of the detergent LDAO and the X-ray crystal structure of this complex solved and refined to 3.6 Å resolution (Fig. 1C and Table 1). The structure reveals that the complex forms a 2:2 heterotetramer, comprised of a central SapA dimer with two molecules of GALC binding symmetrically on the surface of the SapA dimer. By peeling apart the complex structure the details of the surfaces involved in the interaction between GALC and SapA are revealed (Fig. 1D). Importantly, the active site of GALC is buried in the centre of the interaction interface. Not only does the SapA dimer sit over the active site, it makes contact with all three domains of GALC: the TIM barrel,  $\beta$ -sandwich and lectin domains (27). This extensive interface buries a total surface area of 1366 Å<sup>2</sup>, of which 70% is contributed by one SapA chain and 30% by the other (yellow and orange, respectively, in Fig. 1D).

Analysis of the SapA dimer identifies the presence of disordered detergent within the core. This is supported by three observations: the average electron density in the cavity is lower than the surrounding solvent, consistent with the presence of detergent (Fig. S1A); an ordered LDAO headgroup can be modelled at a point of contact with a SapA surface residue (Fig. S1B); and multiple exposed hydrophobic residues line the core, consistent with this surface being exposed to non-polar solvent (Fig. 2A, *right panel*). Further analysis of the hydrophobic surface of the SapA dimer reveals a clear hydrophobic patch at the point where the two SapA chains contact the GALC surface (Fig. 2A, *central panel*, and Fig. 1D). A cross-section through the structure at this position shows that this hydrophobic patch encircles an opening in the SapA dimer surface that lies directly opposite the GALC active site (Fig. 2B). This reveals a continuous open channel that stretches from the GALC active site through to the hydrophobic cavity buried in the SapA dimer. The hydrophobic acyl chains of glycosphingolipids such as galactocerebroside (GalCer) can thus be shielded from the polar solvent by the SapA dimer while the hydrophilic glycosyl head groups are presented to the GALC active site. Previous work from our group identified how *bona fide* substrates bind GALC and the position of this substrate is illustrated in the active site of the GALC-SapA complex (Fig. 2B). Substrate binding results in small but significant conformational changes of key residues at the GALC active site (28). Although the plasticity of the active site and the conformational flexibility of the GalCer acyl chains hindered accurate modelling of a GalCer substrate into the channel of the GALC-SapA complex, it is clear that the distance spanned by GalCer matches that of the channel dimensions (Fig. 2C). We therefore propose that the GALC-SapA structure presented here represents the functional complex that facilitates glycosphingolipid catabolism *in vivo*.

Analysis of the molecular details at the interface reveals that several sidechains of residues along the length of one SapA chain form critical hydrogen bonds and electrostatic interactions with residues in GALC (Fig. 3A and Fig. S2). These interactions span two stretches of this SapA chain: residues 19-32 and 49-63. The interactions contributed by the second SapA chain encompass residues 36-40 and are primarily backbone hydrogen bonds, indicating that this surface is unlikely to define the binding specificity. To confirm that the interface observed in our crystal structure mediates the interaction in solution, a series of point mutations were made in SapA and tested for their ability to bind GALC (Fig. 3B, 3C). Single charge inversions at residues K19, E25, E49, D52 or D60 each abolish binding to GALC, highlighting the electrostatic nature of the interface. Inverting the charge at residue 32 significantly reduces binding but does not completely abolish binding. To ensure that alteration of surface charge alone is not sufficient to abolish binding we mutated residue E65, which does not lie in the interface, to lysine and confirmed it does not interfere with SapA binding to GALC. In addition to electrostatic interactions, hydrophobic interactions at the interface are also critical for binding. The sidechain of

L28 is buried in a hydrophobic pocket of the GALC surface (Fig. S2C). Mutation of this residue to arginine completely knocks out binding to GALC. Residue N21 lies at the interaction interface and forms a backbone hydrogen bond with G144 and a number of hydrophobic interactions with the GALC surface. However, in the context of the cell this highly-conserved residue is post-translationally glycosylated. N-linked glycans often interact with hydrophobic patches on the surface of proteins (29) and mutation of N21 to the bulky hydrophobic sidechain tyrosine moderately increased SapA binding to GALC, suggesting that the glycosylation of N21 contributes to SapA-GALC binding *in vivo*. Residue 24 is a glutamine and forms hydrogen bonds with residues T523, R520 and A525 of GALC (Fig. 3A and S2). In human SapA the corresponding residue is a glutamate, which could form a salt bridge with R520. Mutation of Q24 to lysine blocks SapA binding to GALC, supporting its role in stabilisation of the interaction. Although the only hydrogen bond formed by K19 is via the backbone to GALC residue S146, the acyl region of the K19 sidechain forms hydrophobic interactions with the GALC surface. Disruption of this hydrophobic interaction by substitution of K19 for glutamate abolishes the GALC-SapA interaction. Interestingly, in human GALC, there is a glutamate residue at position 146 rather than a serine, raising the possibility that SapA K19 could form a salt bridge with residue 146 in the human GALC-SapA complex. Taken together, our mutational analysis confirms the critical importance of SapA residues K19, Q24, E25, L28, E32, E49, D52 and D60 for complex formation.

There is high structural similarity between the different saposin proteins (30), suggesting that the specific binding to their cognate enzyme is sequence-mediated. The identification here of specific residues essential for SapA binding to GALC provides a framework for understanding this specificity. Single charge inversions at specific residues are sufficient to abolish binding *in vitro* and thus highlight potentially critical specificity determinants. Sequence alignments of saposins A-D reveal charge inversions, equivalent to those used in this study, in the region spanning residues 19-32 (Fig. 4A and 4B). Based on our binding data, any one of these amino acid changes in saposins B, C and D would block an interaction with GALC. While we were unable to purify SapC, we observed that GALC binds specifically to SapA but not to SapB or SapD (Fig. 4C). This is consistent with the specificity for GALC being conferred by SapA residues 19-32, one of three stretches of SapA residues that mediate the interaction with GALC. As stated previously, the central region spanning SapA residues 36-40 primarily participates in backbone interactions so does not determine specificity. Although residues E49 and D60 of the third interacting region of SapA both form critical salt bridges at the GALC interface, the absence of charge inversions in the other saposin proteins suggests that these interactions contribute to affinity of hydrolase binding rather than specificity.

## Discussion

Here we reveal molecular details of how glycosphingolipids can be bound by SapA and presented to the active site of GALC for catabolic processing. The interaction of SapA with GALC is pH dependent, consistent with its activity in the late endosomal/lysosomal compartments. The interaction also requires the presence of detergents. Structures of SapA in the absence of detergents remain closed (23, 31) whereas in the presence of detergents SapA adopts an open conformation (this study and (21)), suggesting that detergent mediates the formation of open SapA dimers essential for the interaction. The crystal structure of GALC-SapA determined here reveals the complex to be a heterotetramer with an extensive interaction interface that involves residues from all domains of

GALC and both chains of the SapA dimer. There is an open channel running from the GALC active site through to the hydrophobic cavity of the SapA dimer, suggesting a clear mechanism for glycosphingolipid binding and processing.

The SapA dimer structure determined here differs from the previous structure of a SapA lipo-protein disc in isolation (21). Although the conformation of the SapA monomer is essentially identical, in our structure there are direct protein-protein contacts between the saposin monomers, whereas the previous structure contains no SapA-SapA contacts and the assembly is maintained via interactions between detergent molecules. Another difference between these SapA dimer structures is the relative orientations of the monomers. In our structure the monomers are head-to-head, with all termini together at one end, while in the lipo-protein disc structure the monomers adopt a head-to-tail orientation. Comparison of these different SapA dimers shows that the GALC-interacting surface identified in our structure is surface exposed in the lipo-protein disc structure. Docking of GALC onto the surface of the lipo-protein disc structure reveals this SapA dimer to be compatible with GALC binding (Fig. S3A). In this alternative arrangement the opening on the surface of the SapA dimer is less hydrophobic and, due to the loss of interactions with the second SapA chain, the substrate channel is more open and exposed to surrounding solvent (Fig. S3B and C). Although this alternative arrangement may be compatible with substrate binding, the tighter SapA dimer and enclosed hydrophobic channel identified in our complex provides a more protected environment from the hydrophilic solvent. We therefore conclude that our structure, possessing a large buried surface area, an enclosed hydrophobic channel and an obvious mechanism for substrate presentation to GALC, represents the *bona fide* interaction that stimulates SapA-mediated lipid catabolism in the lysosome.

The combination of structure analysis, pulldown assays and sequence comparison of the saposin family members identifies residues 19, 24, 25, 28, 32, 49, 52 and 60 of SapA as essential for binding to GALC. Within this subset, residues 19, 24, 25 and 32 are likely to be critical for determining binding specificity. The importance of the specific charges at these positions is highlighted by the loss of binding upon charge inversion and the conservation of these charges across species (30). The charge inversions between saposin proteins at these specific positions strongly suggests that the relevant patches on their cognate enzymes will possess the complementary charged residues. Despite the lack of structural similarity between glycosyl hydrolases, this identification and the observation that the saposins must sit over the active site provides the first framework for identifying the equivalent interaction interfaces on other saposin-dependent hydrolases.

While ours is the first saposin:enzyme structure, we can compare it to the recent structures of acid sphingomyelinase (ASM), which possesses an intra-molecular saposin-like domain (32-34). Although the enzymatic portion of ASM is not structurally similar to GALC the saposin-like region adopts a similar conformation to SapA in the GALC-SapA complex, allowing a reliable overlay to be made based on this portion (Fig. S4A). This comparison reveals that the catalytic domains are in similar positions relative to the saposin and that the enzyme active sites are adjacent to the saposin, although not directly overlaid (Fig. S4B). However, there are significant differences between these structures and this may reflect the different mechanisms by which the saposin domains facilitate lipid processing. Specifically, SapA is thought to function as a 'solubiliser', an assertion strongly supported by the GALC-SapA structure presented here. ASM, on the other hand, is likely to function via a 'liftase' mechanism. Specifically, ASM has been shown to bind tightly to negatively charged membranes and the saposin domain is likely to play a critical role in orienting the active site towards the membrane (33, 35).

Dimerisation of the saposin-like domain would not be required in this ‘liftase’ model of ASM activity, and indeed we observe that steric hindrance would prevent formation of a SapA-like dimer by the ASM saposin-like domain. We therefore conclude that the mechanisms used by saposins to facilitate lipid presentation will influence the stoichiometry and topology of their association with their cognate hydrolases.

The structure of GALC-SapA presented here defines how saposins can deliver sphingolipids to soluble hydrolases, presenting substrates to the hydrolase via a channel that shields the hydrophobic lipid tails from the surrounding polar solvent.

## Materials and Methods

**Protein expression and purification** His<sub>6</sub>-tagged murine GALC (mGALC) and untagged murine SapA (mSapA) were expressed and purified as described previously (28, 31). mSapA mutants were made by site-directed mutagenesis and verified by sequencing. Mutant proteins were expressed and purified exactly as for wild-type. Purified mGALC was concentrated to 5-15 mg/ml and stored in phosphate-buffered saline pH 7.4 at 4°C. Purified mSapA was concentrated to 8.0 mg/ml and stored in 50 mM Tris pH 7.4, 150 mM NaCl at 4°C.

**Pulldowns** mGALC was mixed with magnetic Ni-NTA agarose beads under saturating conditions (30 µl beads, 75 µg GALC per experiment) and incubated with mixing (90 min, 4 °C). Loaded beads were transferred to a flat bottomed 96-well plate and washed twice in pulldown buffer: 200 µl 100 mM sodium acetate pH 5.4, 20 mM NaCl, 0.05% Tween-20 or LDAO. For screening purposes the buffer pH, NaCl concentration and detergent were altered as appropriate. Concentrated SapA was pre-incubated with 0.1% Tween-20 or LDAO for 1 hour and 160 µg of SapA was added to GALC-loaded beads in 200 µl pulldown buffer and incubated with shaking (2 hr, 4 °C). Beads were then washed four times with 200 µl pulldown buffer. Proteins were eluted with 40 µl 500 mM imidazole, PBS pH 7.4, 0.05% Tween-20 and analysed by 4-12% gradient Bis-Tris PAGE. Following staining with Coomassie, gels were scanned on an Odyssey imaging system and band intensity determined using Image Studio Lite (LI-COR Biosciences).

**Protein complex crystallization** Purified mSapA was pre-incubated with LDAO in a final mix of 180µM mSapA with 50mM sodium acetate pH 5.0, 700mM NaCl and 0.1% LDAO. This SapA-LDAO complex was incubated with purified GALC to give a final SapA:GALC molar ratio of 2:1 for crystallisation trials. Crystallization experiments were performed in 48-well sitting drops (800 nL of complex as prepared above plus 800 nL of precipitant) equilibrated at 20°C against 200µL reservoirs of precipitant. Diffraction quality crystals grew against a reservoir of 75 mM sodium citrate pH 5.6 and 11% (w/v) PEG 3350. Crystals were cryoprotected in reservoir solution supplemented with 20% (v/v) glycerol and flash-cooled by plunging into liquid nitrogen.

**X-ray data collection, structure determination and refinement** Diffraction data were recorded at Diamond Light Source beamline I04-1 on a Pilatus 6M detector (Dectris). Diffraction data were collected at 100K. Data collection statistics are in Table 1. Diffraction data were indexed and integrated using DIALS and scaled and merged using AIMLESS *via* the xia2 automated data processing pipeline (36). Resolution cut-off was decided by CC<sub>1/2</sub> value of >0.5 and *I/σI* of 1.5 in the outer resolution shell.



The structure was solved by molecular replacement using Phaser with mouse GALC (PDB ID 4CCE (28)) and a model of mouse SapA, based on the structure of the human SapA in a lipo-protein disc (PDB ID 4DDJ (37)), as search models. Manual model building was performed using COOT (38). The structure was refined using autobuster (39), over-fitting of the data being minimised by the use of local structure similarity restraints (40) to the high-resolution structure of mouse GALC (PDB ID 4CCE (28)). Model geometry was evaluated with MolProbity throughout the refinement process (41). Final refinement statistics are presented in Table 1. Structural figures were rendered using PyMOL (Schrödinger LLC).

**Structure Analysis** The analysis of buried surface area and interface interactions in the GALC-SapA structure was carried out using the ePISA service at the European Bioinformatics Institute, EBI (42). Hydrogen bonding representations were created using LIGPLOT+ implementing DIMPLOT to generate schematic diagrams of protein-protein interactions (43). Structure-based alignments were carried out using SSM superposition implemented within COOT (44). Multiple sequence alignments were carried out using the MUSCLE service at the EBI (45). The atomic coordinates and structure factors for the GALC-SapA complex have been deposited in the Protein Data Bank (PDB ID 5N8K).

## Acknowledgements

We acknowledge Diamond Light Source for time on beamlines I04 and I04-1 under proposal MX11235. Remote access was supported in part by the EU FP7 infrastructure grant BIOSTRUCT-X (contract no. 283570). The Cambridge Institute for Medical Research is supported by Wellcome Trust Strategic Award 100140. J.E.D. is supported by a Royal Society University Research Fellowship (UF100371). This work is supported by MRC grant MR/N020626/1. S.C.G. is supported by a Sir Henry Dale Fellowship from the Royal Society and Wellcome Trust (098406/Z/12/Z).



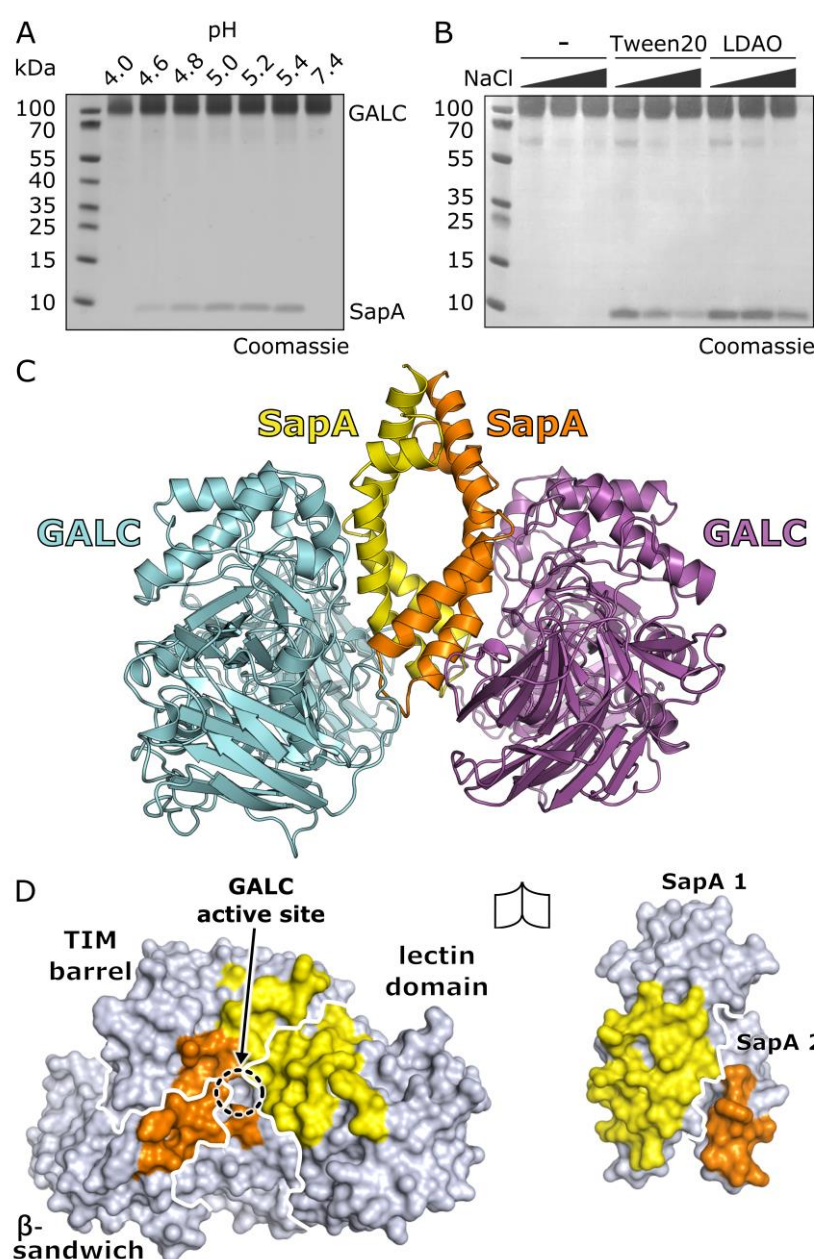
## References

1. Platt FM (2014) Sphingolipid lysosomal storage disorders. *Nature* 510:68-75.
2. Ogretmen B & Hannun YA (2004) Biologically active sphingolipids in cancer pathogenesis and treatment. *Nat. Rev. Cancer* 4:604-616.
3. Ryland LK, Fox TE, Liu X, Loughran TP, & Kester M (2011) Dysregulation of sphingolipid metabolism in cancer. *Cancer Biol. Ther.* 11:138-149.
4. Morimoto S, Yamamoto Y, O'Brien JS, & Kishimoto Y (1990) Distribution of saposin proteins (sphingolipid activator proteins) in lysosomal storage and other diseases. *Proc Natl Acad Sci U S A* 87:3493-3497.
5. Morimoto S, *et al.* (1989) Saposin A: second cerebrosidase activator protein. *Proc Natl Acad Sci U S A* 86:3389-3393.
6. Morimoto S, Martin BM, Kishimoto Y, & O'Brien JS (1988) Saposin D: a sphingomyelinase activator. *Biochem Biophys Res Commun* 156:403-410.
7. Sandhoff K & Harzer K (2013) Gangliosides and gangliosidoses: principles of molecular and metabolic pathogenesis. *J Neurosci* 33:10195-10208.
8. Locatelli-Hoops S, *et al.* (2006) Saposin A mobilizes lipids from low cholesterol and high bis(monoacylglycerol)phosphate-containing membranes: patient variant Saposin A lacks lipid extraction capacity. *J Biol Chem* 281:32451-32460.
9. Remmel N, Locatelli-Hoops S, Breiden B, Schwarzmann G, & Sandhoff K (2007) Saposin B mobilizes lipids from cholesterol-poor and bis(monoacylglycerol)phosphate-rich membranes at acidic pH. Unglycosylated patient variant saposin B lacks lipid-extraction capacity. *The FEBS journal* 274:3405-3420.
10. Alattia JR, Shaw JE, Yip CM, & Prive GG (2006) Direct visualization of saposin remodelling of lipid bilayers. *J Mol Biol* 362:943-953.
11. Alattia JR, Shaw JE, Yip CM, & Prive GG (2007) Molecular imaging of membrane interfaces reveals mode of beta-glucosidase activation by saposin C. *Proc Natl Acad Sci U S A* 104:17394-17399.
12. Hiraiwa M, *et al.* (1997) Lysosomal proteolysis of prosaposin, the precursor of saposins (sphingolipid activator proteins): its mechanism and inhibition by ganglioside. *Arch Biochem Biophys* 341:17-24.
13. Rorman EG & Grabowski GA (1989) Molecular cloning of a human co-beta-glucosidase cDNA: evidence that four sphingolipid hydrolase activator proteins are encoded by single genes in humans and rats. *Genomics* 5:486-492.
14. O'Brien JS & Kishimoto Y (1991) Saposin proteins: structure, function, and role in human lysosomal storage disorders. *FASEB J* 5:301-308.
15. Kolter T & Sandhoff K (2005) Principles of lysosomal membrane digestion: stimulation of sphingolipid degradation by sphingolipid activator proteins and anionic lysosomal lipids. *Annu Rev Cell Dev Biol* 21:81-103.
16. Xu YH, Barnes S, Sun Y, & Grabowski GA (2010) Multi-system disorders of glycosphingolipid and ganglioside metabolism. *J. Lipid Res.* 51:1643-1675.
17. Motta M, *et al.* (2014) Gaucher disease due to saposin C deficiency is an inherited lysosomal disease caused by rapidly degraded mutant proteins. *Hum Mol Genet* 23:5814-5826.
18. Vaccaro AM, *et al.* (2010) Saposin C mutations in Gaucher disease patients resulting in lysosomal lipid accumulation, saposin C deficiency, but normal prosaposin processing and sorting. *Hum Mol Genet* 19:2987-2997.
19. Spiegel R, *et al.* (2005) A mutation in the saposin A coding region of the prosaposin gene in an infant presenting as Krabbe disease: first report of saposin A deficiency in humans. *Mol Genet Metab* 84:160-166.
20. Matsuda J, Vanier MT, Saito Y, Tohyama J, & Suzuki K (2001) A mutation in the saposin A domain of the sphingolipid activator protein (prosaposin) gene results in a late-onset, chronic form of globoid cell leukodystrophy in the mouse. *Hum Mol Genet* 10:1191-1199.

21. Popovic K, Holyoake J, Pomes R, & Prive GG (2012) Structure of saposin A lipoprotein discs. *Proc Natl Acad Sci U S A* 109:2908-2912.
22. Popovic K & Prive GG (2008) Structures of the human ceramide activator protein saposin D. *Acta Crystallogr D Biol Crystallogr* 64:589-594.
23. Ahn VE, Leyko P, Alattia JR, Chen L, & Prive GG (2006) Crystal structures of saposins A and C. *Protein Sci* 15:1849-1857.
24. Ahn VE, Faull KF, Whitelegge JP, Fluharty AL, & Prive GG (2003) Crystal structure of saposin B reveals a dimeric shell for lipid binding. *Proc Natl Acad Sci U S A* 100:38-43.
25. de Alba E, Weiler S, & Tjandra N (2003) Solution structure of human saposin C: pH-dependent interaction with phospholipid vesicles. *Biochemistry* 42:14729-14740.
26. Frauenfeld J, *et al.* (2016) A saposin-lipoprotein nanoparticle system for membrane proteins. *Nat Methods* 13:345-351.
27. Deane JE, *et al.* (2011) Insights into Krabbe disease from structures of galactocerebrosidase. *Proc Natl Acad Sci U S A* 108:15169-15173.
28. Hill CH, Graham SC, Read RJ, & Deane JE (2013) Structural snapshots illustrate the catalytic cycle of beta-galactocerebrosidase, the defective enzyme in Krabbe disease. *Proc Natl Acad Sci U S A* 110:20479-20484.
29. Petrescu AJ, Milac AL, Petrescu SM, Dwek RA, & Wormald MR (2004) Statistical analysis of the protein environment of N-glycosylation sites: implications for occupancy, structure, and folding. *Glycobiology* 14:103-114.
30. Bruhn H (2005) A short guided tour through functional and structural features of saposin-like proteins. *Biochem. J.* 389:249-257.
31. Hill CH, Read RJ, & Deane JE (2015) Structure of human saposin A at lysosomal pH. *Acta Crystallogr F Struct Biol Commun* 71:895-900.
32. Gorelik A, Illes K, Heinz LX, Superti-Furga G, & Nagar B (2016) Crystal structure of mammalian acid sphingomyelinase. *Nat Commun* 7:12196.
33. Xiong ZJ, Huang J, Poda G, Pomes R, & Prive GG (2016) Structure of Human Acid Sphingomyelinase Reveals the Role of the Saposin Domain in Activating Substrate Hydrolysis. *J Mol Biol* 428:3026-3042.
34. Zhou YF, *et al.* (2016) Human acid sphingomyelinase structures provide insight to molecular basis of Niemann-Pick disease. *Nat Commun* 7:13082.
35. Oninla VO, Breiden B, Babalola JO, & Sandhoff K (2014) Acid sphingomyelinase activity is regulated by membrane lipids and facilitates cholesterol transfer by NPC2. *J. Lipid Res.* 55:2606-2619.
36. Winter G (2009) xia2: an expert system for macromolecular crystallography data reduction. *J. Appl. Cryst.* 43:186-190.
37. McCoy AJ, *et al.* (2007) Phaser crystallographic software. *J Appl Crystallogr* 40:658-674.
38. Emsley P, Lohkamp B, Scott WG, & Cowtan K (2010) Features and development of Coot. *Acta Crystallogr D Biol Crystallogr* 66:486-501.
39. Bricogne G, *et al.* (2016) BUSTER version 2.10.3. *Cambridge, UK.*
40. Smart OS, *et al.* (2012) Exploiting structure similarity in refinement: automated NCS and target-structure restraints in BUSTER. *Acta Crystallogr D Biol Crystallogr* 68:368-380.
41. Chen VB, *et al.* (2010) MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr D Biol Crystallogr* 66:12-21.
42. Krissinel E & Henrick K (2007) Inference of macromolecular assemblies from crystalline state. *J Mol Biol* 372:774-797.
43. Laskowski RA & Swindells MB (2011) LigPlot+: multiple ligand-protein interaction diagrams for drug discovery. *J Chem Inf Model* 51:2778-2786.
44. Krissinel E & Henrick K (2004) Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta Crystallogr D Biol Crystallogr* 60:2256-2268.

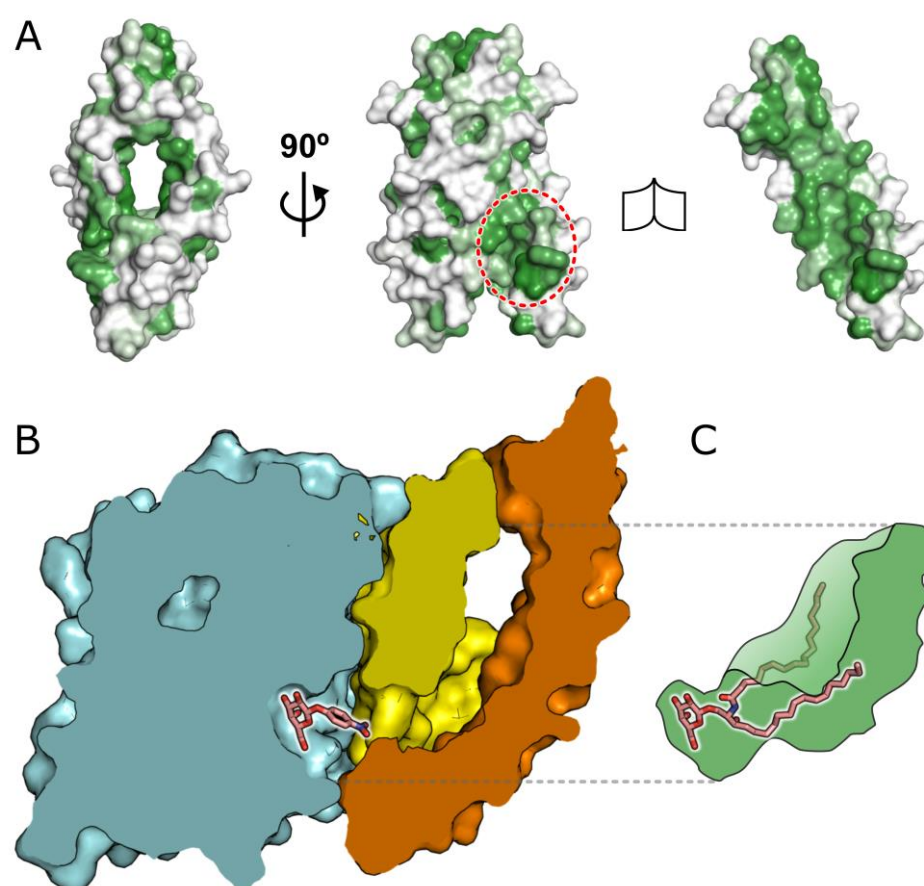
45. Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* 32:1792-1797.

**Figure 1**



**Figure 1. GALC-SapA complex formation and structure.** (A) Coomassie-stained SDS-PAGE of pull-down assays with immobilised GALC in the presence of the detergent, Tween-20, shows binding to SapA at pH values equivalent to the late-endosomal/lysosomal compartment. Glycosylation of GALC results in a smeared band on gradient gels. (B) Pull-down assays at pH 5.6 identify that the interaction requires the presence of detergents, such as Tween 20 or LDAO, and is reduced with increasing NaCl concentration (20, 150 and 300 mM NaCl). (C) The crystal structure of the GALC-SapA complex reveals a symmetric 2:2 heterotetramer of SapA (yellow and orange) with GALC (cyan and magenta). (D) The interacting surfaces of GALC (left) and the SapA dimer (right), rotated by 90° with respect to (C) to highlight the respective interaction surfaces. On the GALC surface the interacting regions contacted by SapA chain 1 (yellow) and SapA chain 2 (orange) are highlighted, the different domains of GALC are outlined (white lines) and the active site is identified. On the SapA surface the residues that interact with GALC are highlighted for each chain and the different chains of SapA are outlined (white line).

**Figure 2**

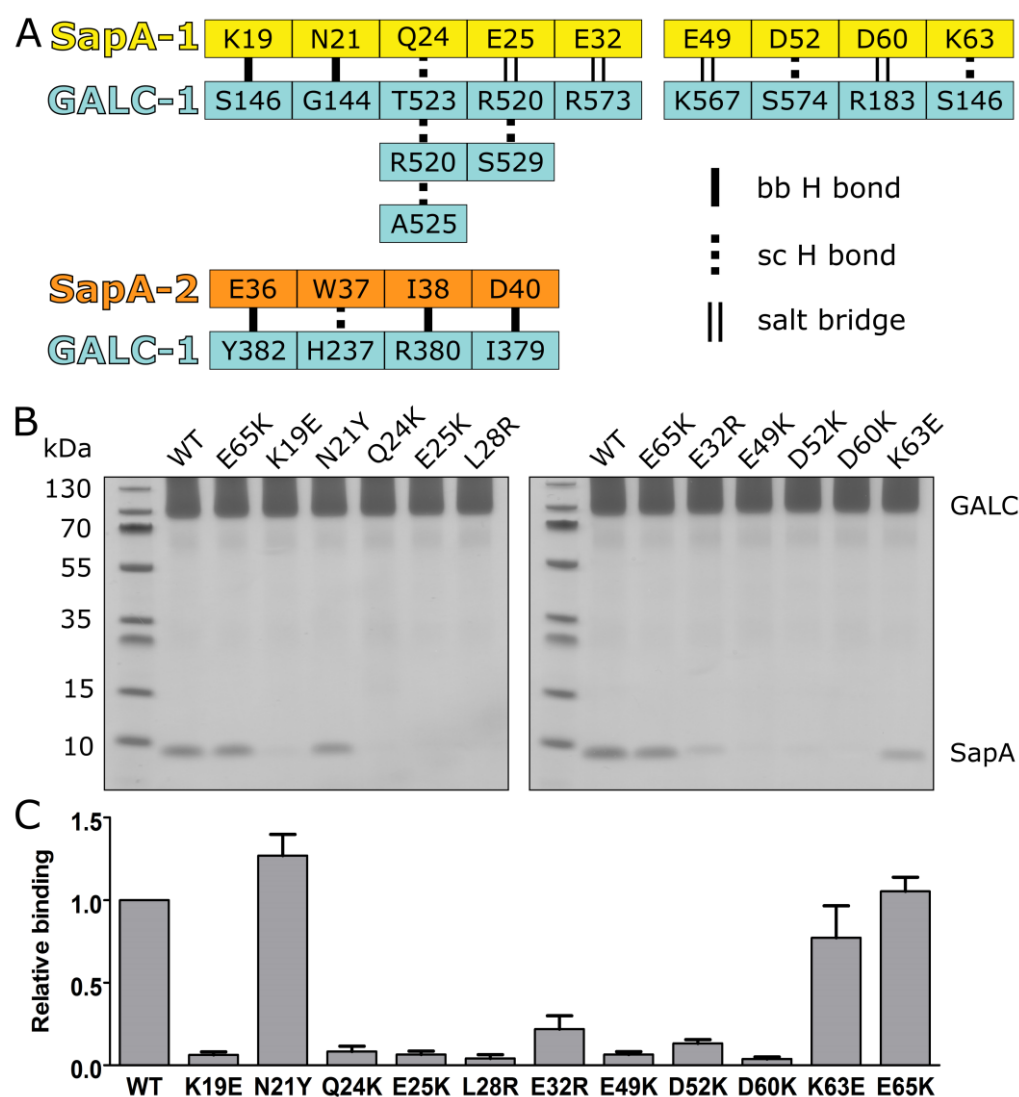


**Figure 2. The GALC-SapA structure reveals a hydrophobic channel buried in the core of the complex.**

(A) Residue hydrophobicity (green) is mapped onto the surface of the SapA dimer. The left panel is in the same orientation as shown in panel 1C, while the central panel is rotated by 90° to be in the same orientation as Fig. 1D. The region that binds over the GALC active site is circled (red dashed line). The right panel maintains the orientation of the central panel, with one SapA chain removed to reveal the highly hydrophobic inner surface of the SapA dimer. (B) Cross-section through the GALC-SapA structure reveals an open channel stretching from the GALC active site into the SapA hydrophobic cavity. The GALC surface and cut-through is shown in cyan with substrate (pink sticks) positioned in the active site based on the previous structure of GALC in complex with a soluble substrate (PDB ID 4CCC (28)). The SapA dimer surface is shown in yellow and orange. For clarity, the second GALC monomer is not shown. (C) Illustration of the hydrophobic cavity (green) showing that the lipidated substrate galactocerebroside (pink sticks) fits into this cavity, bridging the channel from the GALC active site to hydrophobic SapA core.

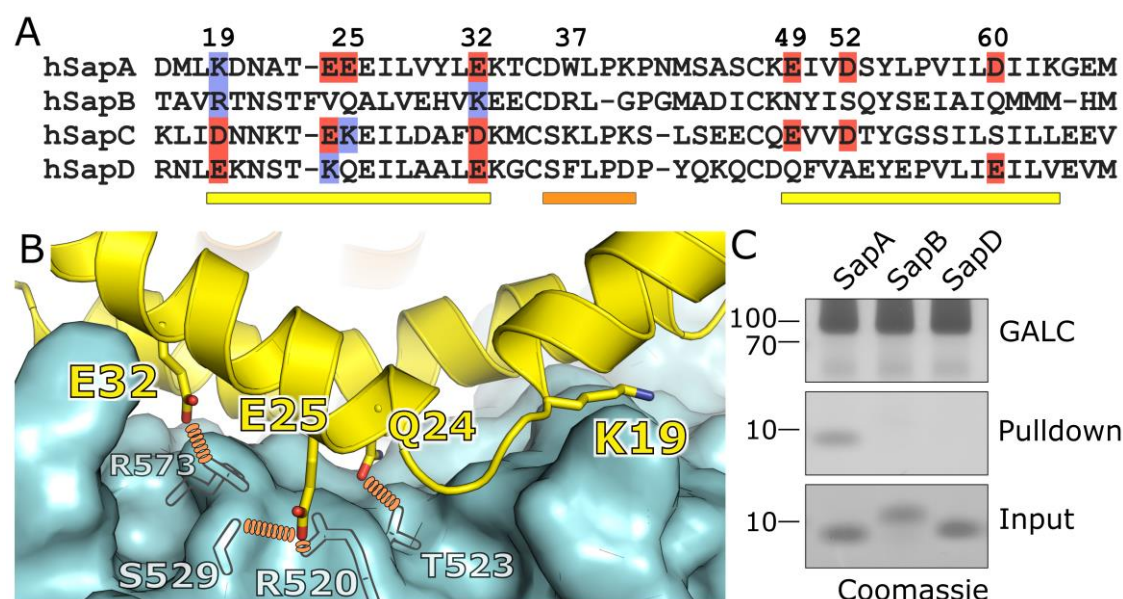


**Figure 3**



**Figure 3. Critical residues at the GALC-SapA interface.** (A) Schematic diagram illustrating the backbone (bb) hydrogen bonds, sidechain (sc) hydrogen bonds and salt bridges observed between GALC (cyan) and specific residues of SapA chains 1 (yellow) or 2 (orange). (B) Coomassie-stained SDS-PAGE of pull-down assays with immobilised GALC identifies critical SapA mutations at the interface that abolish GALC binding in solution. (C) Quantification of SapA binding in pull-down assays (n=3, error bars represent SEM).

**Figure 4**



**Figure 4. Saposin specificity.** (A) Sequence alignment of human saposins A-D highlighting charge inversions of critical interface residues. Acidic residues (red) and basic residues (blue) at critical positions are highlighted. Coloured bars below the sequence identify the regions that interact with GALC: two regions from SapA chain 1 (yellow) and the one from SapA chain 2 (orange) as detailed in Fig. 3A. (B) Critical residues that may determine binding specificity are shown (sticks) on SapA (yellow) at the interface with GALC (cyan surface, key residues depicted as transparent sticks). Hydrogen bonds between GALC and SapA residues are highlighted (orange dotted lines). (C) Coomassie-stained SDS-PAGE pulldowns with immobilised GALC illustrating specific binding of SapA, but not SapB or SapD.



**Table 1.** Data Collection and Refinement Statistics. Values in parentheses are for highest-resolution shell.

<b>Data collection</b>	
Space group	$P 6_2 2 2$
Cell dimensions	
$a, b, c$ (Å)	187.2, 187.2, 360.2
$\alpha, \beta, \gamma$ (°)	90, 90, 120
Resolution (Å)	180.12 – 3.60 (3.70 - 3.60)
$R_{\text{merge}}$	0.386 (2.946)
$CC_{1/2}$	0.945 (0.677)
$I/\sigma I$	9.3 (1.5)
Completeness (%)	100 (100)
Redundancy	25.6 (25.2)
<b>Refinement</b>	
Resolution (Å)	162.11-3.60
No. reflections	43,825
$R_{\text{work}}/R_{\text{free}}$	0.213/0.226
No. atoms	
Protein	11,522
Other <sup>1</sup>	252
Water	20
$B$ -factors	
Protein	100.9
Other <sup>1</sup>	139.3
Water	65.6
Ramachandran	
Favoured (%)	95.7
Outliers (%)	0.5
r.m.s. deviations	
Bond lengths (Å)	0.008
Bond angles (°)	0.94

<sup>1</sup> Includes carbohydrate, ions and LDAO molecules