Crystal structure of the human sterol transporter ABCG5/ABCG8

Jyh-Yeuan Lee[1], **Lisa N. Kinch**[2,3,4], **Dominika M. Borek**[2,3], **Jin Wang**[1], **Junmei Wang**[5], **Ina L. Urbatsch**[6], **Xiao-Song Xie**[1], **Nikolai V. Grishin**[2,3,4], **Jonathan C. Cohen**[1], **Zbyszek Otwinowski**[2,3], **Helen H. Hobbs**[1,4,*], and **Daniel M. Rosenbaum**[2,3,*]

[1]Eugene McDermott Center for Human Growth and Development, University of Texas Southwestern Medical Center at Dallas, Dallas, Texas 75390, USA

[2]Department of Biophysics, University of Texas Southwestern Medical Center, Dallas, Texas 75390, USA

[3]Department of Biochemistry, University of Texas Southwestern Medical Center, Dallas, Texas 75390, USA

[4]Howard Hughes Medical Institute, University of Texas Southwestern Medical Center, Dallas, Texas 75390, USA

[5]Cecil & Ida Green Center for Molecular, Computational and Systems Biology, University of Texas Southwestern Medical Center, Dallas, Texas 75390, USA

[6]Department of Cell Biology and Biochemistry, Texas Tech University Health Sciences Center, Lubbock, Texas 79430, USA

Abstract

ATP binding cassette (ABC) transporters play critical roles in maintaining sterol balance in higher eukaryotes. The ABCG5/ABCG8 heterodimer (G5G8) mediates excretion of neutral sterols in liver and intestines[1–5]. Mutations disrupting G5G8 cause sitosterolaemia, a disorder characterized by sterol accumulation and premature atherosclerosis. Here we use crystallization in lipid bilayers to determine the X-ray structure of human G5G8 in a nucleotide-free state at 3.9 Å resolution, generating the first atomic model of an ABC sterol transporter. The structure reveals a new transmembrane fold that is present in a large and functionally diverse superfamily of ABC transporters. The transmembrane domains are coupled to the nucleotide-binding sites by networks of interactions that differ between the active and inactive ATPases, reflecting the catalytic

asymmetry of the transporter. The G5G8 structure provides a mechanistic framework for understanding sterol transport and the disruptive effects of mutations causing sitosterolaemia.

Cholesterol is an essential component of vertebrate cell membranes. Animals maintain sterol balance by limiting dietary sterol uptake from the gut and promoting sterol secretion from hepatocytes into bile. These physiological processes are mediated by a heterodimeric ABC transporter, consisting of G5 and G8 (refs 1–3) polypeptides, which is embedded in apical membranes of bile ducts and intestinal enterocytes[4,5]. Mutations in G5 or G8 that block sterol secretion into bile and the gut lumen cause sitosterolaemia, an autosomal recessive disorder in which sterol accumulation leads to premature coronary atherosclerosis.

ABC transporters constitute a ubiquitous protein superfamily that utilizes energy derived from ATP hydrolysis to translocate substrates across membranes[6]. Family members share a common architecture that comprises two transmembrane domains (TMDs) and two nucleotide-binding domains (NBD; specified herein as the contiguous polypeptide domain contributed from each subunit, whereas NBS denotes a composite nucleotide-binding site made up of elements from both subunits). Humans have 48 ABC transporters that are classified into seven subfamilies (A–G)[7]. In the G5 and G8 half-transporters, the NBD is amino (N)-terminal to the TMD, which consists of six transmembrane helices (TMHs) (Fig. 1a). The molecular mechanism by which G5G8 effluxes sterol from plasma membranes remains poorly defined.

Lipid-driven three-dimensional crystallization is a powerful method to determine structures of integral membrane proteins[8,9]. The only ABC transporters that have been crystallized in lipid bilayers are the bacterial polypeptide processing and secretion transporter (PCAT1, an ABCB homologue)[10] and the maltose transporter–EIIA complex[11]. No ABCG family member has been structurally characterized. To obtain diffraction-quality crystals, human G5 and G8 were coexpressed in *Pichia pastoris*[12], and tandem affinity chromatography was used to purify stable, monodisperse G5G8 heterodimers that retained ATPase activity (Extended Data Fig. 1). The protein was reconstituted into di-myristoyl-phosphatidylcholine (DMPC) bicelles[13], and growth of optimal bicelle crystals required the presence of cholesterol to obtain diffraction higher than 3.9 Å resolution (Extended Data Fig. 2).

The G5G8 structure (Fig. 1b) was solved using tungsten-derived single-wavelength anomalous dispersion (Extended Data Table 1 and Extended Data Fig. 3a). The G5G8 crystals comprise two-dimensional layers in which two heterodimers in the asymmetric unit pack in an anti-parallel fashion related by twofold non-crystallographic symmetry (Extended Data Fig. 3b). Native diffraction data were averaged from 19 crystals to 3.9 Å resolution, and the structure was refined to $R/R_{\text{free}} = 0.242/0.328$ (representative electron density maps for selected regions are shown in Extended Data Fig. 3c). G5 and G8 share 28% amino-acid identity, and show a high degree of structural conservation with a root mean squared deviation (r.m.s.d.) of 2.0 Å (Fig. 1c).

Our G5G8 structure adopts an inward-facing conformation, which is analogous to some other nucleotide-free ABC exporters[10,14,15] and importers[16,17]. The packing of the TMHs and interfacial contacts of G5 and G8 differ from other ABC transporter structures,

including the type I (for example, ModBC-A)[16] and type II (for example, BtuCD-F)[17] importers, and type I exporters (for example, TM287/288)[14] (Fig. 1d). No TMH from either subunit crosses over into the other half-transporter's TMD. In the extracellular domain (ECD), the regions between TMH5 and TMH6 form distinct α-helical structures (Fig. 1b). Three missense mutations causing sitosterolaemia, R419P and R419H in G5 and G574R in G8 (refs 2, 3), are located near the apices of TMH2 and TMH5, respectively (Extended Data Fig. 4a), and both residues are involved in contacts with the ECDs (for example, R419 forms hydrogen bonds with E578 on the G5 ECD). These mutations would be predicted to interfere with the native positions of the ECD helices, suggesting their importance for sterol exit from the TMDs.

The lack of electron density for nucleotide and the spatial separation between opposing Walker A and Signature motifs indicates that the G5G8 structure represents a nucleotide-free state (Extended Data Fig. 4b, c). Nonetheless, the two NBDs contact each other at the extreme cytoplasmic end to form a closed conformation through a pair of NPXDF motifs (G5: NPFDF; G8: NPADF; Extended Data Fig. 4d), which are conserved in the ABCG family and are required for cholesterol efflux by ABCG1 (ref. 18).

How does G5G8 move sterol out of the energetically favourable environment of the plasma membrane? We identified features in our electron density map that may represent cholesterol. These features are located at symmetrical 'vestibules' on opposing faces of the TMD dimer, which open to the bilayer and extend into the centre of the dimer interface (Extended Fig. Data 5a, b). Each vestibule is flanked by TMH1-2 of one TMD and TMH4-6 of the other TMD, with a 'ceiling' formed by an α-helix from the ECD extending into the membrane. Several residues in these vestibules are conserved throughout eukaryotic evolution and may represent binding surfaces or entryways for sterols to access the core of the heterodimer interface. To test this hypothesis, we performed an *in vivo* functional reconstitution assay using adenoviruses to express recombinant G5 and G8 in G5G8 knockout mice[5,19]. Expression of wild-type (WT) G5 together with WT G8 resulted in a ~30-fold increase in cholesterol transport into bile. In contrast, expression of G8 with G5 containing a substitution (A540F) that occludes the putative cholesterol-binding site failed to restore biliary cholesterol transport despite forming WT levels of the mature G5G8 heterodimer (Extended Data Fig. 5c).

ATP-dependent sterol translocation across membranes requires allosteric communication between catalytic NBSs and substrate-exporting TMDs. In models derived from bacterial transporters, TMD conformational changes result from engagement of the NBDs with 'coupling helices' (CpH) in the intracellular loops of the TMDs[20]. The TMDs of G5 and G8 each contain a prototypical CpH in the linker between TMH2 and TMH3 that is analogous to the position of coupling helices in other ABC exporters (Fig. 2). Each half-transporter also contains an orthogonal α-helix that we have named the 'connecting helix' (CnH), which is interfacial to the membrane bilayer and connects the NBD to the TMD. The CnH packs against a short cytoplasmic helix containing a conserved glutamate (designated the E-helix). The CpHs and CnHs are in proximity (~10–15 Å) to the consensus Signature motifs (orange) of the same half-transporters, which form the NBS with the Walker A (red) and Walker B (green) motifs of the opposing half-transporters (Fig. 2a, b). In G5, R374 on the

CnH forms a buried salt bridge with E452 on the CpH, while R381 interacts with the conserved glutamate (E146) that defines the E-helix. A disease-causing missense mutation at this residue (E146Q)[3] is predicted to alter crosstalk between the TMD and NBD. The CnH, CpH, and E-helix elements on G5 form a stabilized three-helix bundle. In contrast, the CpH in G8 is rotated such that E481 (corresponding to E452 in G5) points away from R405 (homologous to G5-R374 on the CnH), instead packing against the G8 carboxy (C) terminus that loops back into the heterodimer interface.

The distinct architectures for the two TMD/NBD interfaces reflect the asymmetry in catalytic sites of G5G8. The inactive G5 Signature motif (part of NBS1)[21], which binds but does not hydrolyse ATP, is adjacent to the CnH/CpH/E-helix bundle of G5, whereas the active G8 Signature motif (part of NBS2) is near the three-helix bundle of G8 (Fig. 2). We hypothesize that the stable three-helix bundle in G5 acts as a rigid body, whereas the three-helix bundle of G8 exhibits greater flexibility and transitions between different conformations. With the proximity of G8-CpH to the catalytically active NBS2, these conformational changes could allosterically link ATP hydrolysis to sterol transport.

What are the likely consequences of ATP binding and hydrolysis on TMD conformation and sterol binding? A network of conserved polar residues in both G5 and G8 forms hydrogen bonds and salt bridges that extend from the CnH and CpH to the proximal part of the TMD interface (Fig. 3). The hydrogen bonds connecting the TMD polar relay may render this network more deformable than a buried hydrophobic core, allowing this region to serve as a flexible hinge for subunit motions with a low energy barrier. Involvement of a TMD polar relay may be a general feature of ABC transporters: the bacterial PCAT1 (ref. 10) exporter and maltose importer[22] structures also contain clusters of conserved polar residues in the TMD, which rearrange in the transition to a nucleotide-bound state (Extended Data Fig. 6a, b). We performed a 100-ns molecular dynamics simulation and vibrational mode analysis of our G5G8 structure in an explicit POPC/cholesterol bilayer and water. In these calculations, we found that the CnH and CpH elements of both subunits move inwards, bringing opposing Walker A and Signature motifs into closer contact (Extended Data Fig. 6c). Inward movement of the NBDs was coupled to upward movement of a subset of TMD elements (Extended Data Fig. 6d). R543, which is embedded in the core of the G8 TMD (part of the TMD polar relay), interacts with E503; both residues participate in the upward movement during the molecular dynamics simulation. The sitosterolaemia-causing mutation R543S[3] would be predicted to disrupt interaction with E503 and thereby destabilize the TMD polar relay.

We performed coevolution analysis[23] on ABCG TMD sequences to predict positions that could form close contacts at the TMD interface during the transport process (Extended Data Fig. 7a, b). The highest scoring co-evolving positions highlighted three potential interactions between subunits. As these surfaces are distant (>8 Å apart) in the current structure, we infer that they may come into contact at another stage of the transport cycle with inward conformational changes of the TMDs. Y432 on G5 (TMH2) and N568 on G8 (TMH5) are among these pairs, and Y342 is connected to the NBDs through the TMD polar relay (Fig. 3). Accordingly, the mutation G5-Y432A disrupted cholesterol transport in our *in vivo* assay without affecting heterodimer maturation (Extended Data Fig. 7c). Together, these results

support a model in which recruitment of ATP stabilizes inward movement of the NBDs and inward/upward movement of the TMDs. We propose that these movements, which would reshape the TMD interface and outer membrane-facing surfaces of the transporter where we observe electron density for cholesterol (Extended Data Fig. 5), contribute to sterol transport across the membrane.

The TMDs of G5 and G8 adopt a tertiary fold that differs from those of known ABC transporter structures[20] (Fig. 1d), which have different topologies, longer cytoplasmic extensions of TMH segments and differing placement of coupling helices. Previous classification of ABC exporters suggests that TMDs have arisen at least three independent times, giving rise to the ABC1, ABC2, and ABC3 superfamilies[24]. G5G8 is a member of the ABC2 exporter superfamily[25], which includes the ABCA and ABCG eukaryotic subfamilies and a diverse group of prokaryotic transporters for substrates ranging from polysaccharide-containing teichoic acids and lipo-oligosaccharides (components of bacterial cell walls and outer membranes) to the antibiotic peptides mutacin and bacitracin (Fig. 4a). The ABCG subfamily includes the largest family of ABC transporters in the plant kingdom (Fig. 4b)[26] and the historically important pigment transporters that determine eye colour in *Drosophila* (white, brown, and scarlet)[27]. Using our G5G8 structure as a template, we used the program Modeller[28] to create a homology model of the white/brown heterodimer that confers brown eye colour to *Drosophila* (Extended Data Fig. 8). Residues at the extracellular end of TMH5 of white (G588, G589, and F590) have been identified as potentially interacting with the dye substrate[29,30]. These amino acids are conserved in G5G8 at the interface between the TMDs, at a site homologous to the G574R sitosterolaemia mutation in G8. This model indicates a possible conserved site of substrate exit for ABCG transporters and shows that the structure of G5G8 can serve as a powerful platform for structure–function studies of this large and functionally diverse set of membrane proteins.

METHODS

Cloning and expression of recombinant human ABCG5 and ABCG8

P. pastoris expression vectors (pSGP18 and pLIC) were derived from pPICZB (Invitrogen) as described[12,31]. The cDNAs for human ABCG5 (NCBI accession number NM_022436) and ABCG8 (NCBI accession number NM_022437) were obtained from the National Institutes of Health collection (M. Dean). A tag encoding a rhinovirus 3C protease site followed by a calmodulin binding peptide (CBP) was added to the C terminus of G8 (pSGP18-G8-3C-CBP). A tandem array of six histidines separated by glycine ($His_6GlyHis_6$) was added to the C terminus of G5 (pLIC-G5-H_{12}). The plasmids were linearized using PmeI and co-transformed into *Pichia* strain KM71H by electroporation. To select for plasmid integration, cells were plated on YPD plates containing sorbitol (1.0 M) and Zeocin (0.5 or 1 mg ml^{-1}), and incubated at 30 °C for 2–5 days. A total of 10–20 yeast colonies from each plate were selected and grown in minimal glycerol yeast nitrogen base (MGY) medium (10 ml), and then induced by adding methanol (0.5%) in minimal methanol (MM) medium. Crude microsomal membranes were prepared, and 30 µg of protein was resolved by SDS–PAGE. Protein expression was analysed by immunoblotting using monoclonal anti-RGSH$_4$ antibodies (Qiagen) to detect G5 and polyclonal anti-hABCG8 antibodies (see

below) to detect G8. The clones expressing the highest level for both G5 and G8 were selected and stored in 15–20% glycerol at −80 °C.

Cell culture and microsomal membrane preparation

A starter yeast culture was prepared by growing transformed yeast in MGY medium (10 ml) to an absorbance $A_{600\ nm}$ of 10. The culture was used to inoculate a litre of MGY medium in a 2.8-l Fernbach flask and grown in a refrigerated Innova shaker (New Brunswick) at 250 rpm for 24 h (28–30 °C). To maximize yield, the acidity of the culture was monitored with the pH maintained at 5–6 using 10% (w/w) ammonium hydroxide (NH_4OH). To induce protein expression, cells were incubated with 0.1% (v/v) methanol for 6–12 h. The methanol concentration was increased to 0.5% (v/v) by adding methanol every 12 h for 36–48 h. Cell pellets were collected and re-suspended in lysis buffer (0.33 M sucrose, 0.3 M TrisCl, pH 7.5, 0.1 M ε-aminocaproic acid, 1 mM EDTA, and 1 mM EGTA) to a concentration of 0.5 g ml^{-1}, and stored at −80 °C. Approximately 30 ± 5 g of cell mass was typically obtained from 1 l of cultured cells.

To prepare microsomal membranes, frozen cells were thawed and reducing agent and protease inhibitors were added (final concentration: DTT (10 mM), leupeptin (2 μg ml^{-1}), pepstatin A (2 μg ml^{-1}), and PMSF (2 mM)). Cells were passed through an ice-chilled microfluidizer (Microfluidics) three to five times at 25,000–30,000 psi. The unbroken cells, nuclei and organelles were spun down at 3,500–4,000g for 15 min followed by 15,000g for 30 min, all at 4 °C. Microsomal membrane vesicles were pelleted by ultracentrifugation using a Beckman 45Ti rotor at 40,000–45,000 rpm (maximum 200,000g) at 4 °C for 2 h and re-suspended in buffer A (50 mM Tris pH 8.0, 100 mM NaCl, and 10% glycerol) using a dounce homogenizer, and stored at −80 °C.

Protein purification and pre-crystallization treatment

Frozen microsomal membranes were thawed and the protein concentration was adjusted to 4–6 mg ml^{-1} using solubilization solution (50 mM Tris-HCl, pH 8.0, 100 mM NaCl, and 10% glycerol, 1% (w/v) β-dodecyl maltoside (β-DDM, Inalco Pharmaceuticals), 0.5% (w/v) sodium cholate (Sigma-Aldrich), 0.25% (w/v) cholesteryl hemisuccinate Tris (CHS-Tris, Anatrace), 5 mM imidazole, 5 mM β-mercaptoethanol (β-ME), 2 μg ml^{-1} leupeptin, 2 μg ml^{-1} pepstatin A, 2 mM PMSF). Insoluble membranes were removed by centrifugation using a Beckman 45Ti rotor at 30,000 rpm (100,000g) for 30 min at 4 °C, and a final concentration of 20 mM imidazole and 0.1 mM Tris (2-carboxylethyl) phosphine (TCEP) was added to the solubilized supernatant.

Tandem affinity chromatography was performed. First, the soluble membrane proteins were bound to a nickel-nitrilotriacetic acid (Ni-NTA) column (Qiagen; 1°Ni-NTA) that was pre-equilibrated with buffer A, and washed with 10 column volumes of buffer B (50 mM HEPES (4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid), pH 7.5, 100 mM NaCl, 0.1% (w/v) β-DDM, 0.05% (w/v) cholate, 0.01% (w/v) CHS (Steraloids), 0.1 mM TCEP) with 25 mM imidazole. The column was then washed with 10 column volumes of buffer B with 50 mM imidazole, and eluted with buffer C (buffer B with 200 mM imidazole, 1 mM CaCl$_2$, 1 mM MgCl$_2$). Peak fractions from the Ni-NTA eluates were mixed with equal volume of

buffer D1 (buffer B plus 1 mM $CaCl_2$, 1 mM $MgCl_2$), and loaded onto a CBP column (Agilent; 1°CBP) that was pre-equilibrated with buffer D1. To exchange the detergent, the CBP column was washed serially with buffer D1 and buffer D2 (buffer B plus 1 mM $CaCl_2$, 1 mM $MgCl_2$, 0.1% (w/v) decyl-maltose neopentyl glycol (DMNG, Anatrace), but no β-DDM) in a step-wise fashion: 3 column volumes of D1, 3 column volumes of D1:D2 (3:1, v/v), 3 column volumes of D1:D2 (1:1, v/v), 3 column volumes of D1:D2 (1:3, v/v), and 6–10 column volumes of D2. Finally, the G5G8 heterodimers were eluted with buffer E (50 mM HEPES, pH 7.5, 300 mM NaCl, 2 mM EGTA, 0.1% (w/v) DMNG, 0.05% (w/v) cholate, 0.01% (w/v) CHS, 1 mM TCEP). Divalent cations were added to the CBP eluates to a final concentration of 10 mM $MgCl_2$ and 10 mM $CaCl_2$ to quench residual EGTA. No detergent exchange was performed for proteins used for ATPase assays. The N-linked glycans and the CBP tag were cleaved by endoglycosidase H (Endo H, ~0.2 mg per 10–15 mg purified protein) and HRV-3C protease (~2 mg per 10–15 mg purified proteins) for 6–12 h at 4 °C. The CBP tag-free proteins were collected from the flow-through fraction of a second CBP column (2°CBP). G5G8 was concentrated to a volume of 1–2 ml and separated from aggregates, impurities, and enzymes by gel filtration chromatography using an ÄKTA Purifier and a Superdex 200 30/100 GL column (GE Healthcare Life Sciences) in buffer F (10 mM HEPES, pH 7.5, 100 mM NaCl, 0.1% (w/v) DMNG, 0.05% (w/v) cholate, 0.01% (w/v) CHS). The peak fractions were pooled together, and additional HEPES and TCEP were added to final concentrations of 50 mM and 1 mM, respectively.

Purified G5G8 dimers were treated by reductive methylation. Briefly, proteins were incubated twice with 20 mM dimethylamine borane (DMAB) and 40 mM formaldehyde for 2 h at 4 °C on an oscillatory shaker and then 10 mM DMAB was added. After 12 h, the reaction was stopped by 100 mM TrisCl, pH 7.5. For protein relipidation, the methylated proteins were loaded onto a second Ni-NTA column (2°Ni-NTA) that was pre-equilibrated with 100 mM TrisCl, pH 8.0, and 100 mM NaCl. The column was washed slowly with 10 column volumes of buffer G (10 mM HEPES, pH 7.5, 100 mM NaCl, 0.5 mg ml^{-1} DOPC:DOPE (3:1, w/w), 0.1% (w/v) DMNG, 0.05% (w/v) cholate, 0.01% (w/v) CHS), and the lipidated proteins were eluted using buffer H (10 mM HEPES, pH 7.5, 100 mM NaCl, 200 mM imidazole, 0.5 mg ml^{-1} DOPC:DOPE = 1:1 (w/w), 0.1% (w/v) DMNG, 0.05% (w/v) cholate, 0.01% (w/v) CHS). TCEP (1 mM) and $MgSO_4$ (10 mM) were added to the 2°Ni-NTA eluates, and the protein was treated with 5 mM and 2 mM iodoacetamide for 1 h on ice and then passed through a PD-10 desalting column (GE Healthcare Life Sciences) that was equilibrated with buffer I (10 mM HEPES, pH 7.5, 100 mM NaCl, 200 mM imidazole, 0.1% (w/v) DMNG, 0.05% (w/v) cholate, 0.01% (w/v) CHS). Finally, the precipitants were removed by ultracentrifugation using a Beckman TLA120.2 rotor (150,000g) for 10 min at 4 °C. The supernatants were concentrated to a final protein concentration of 25–50 mg ml^{-1} using a 100 kDa cutoff Vivaspin concentrator (Sartorius), and used within 1 week of purification for crystallization.

Protein crystallization and crystal sample preparation

All crystals were obtained by reconstituting G5G8 proteins into DMPC/cholesterol/CHAPSO or DMPC/cholesterol/DHPC (Anatrace) bicelles. First, 10% bicelle stock solution was prepared by mixing lipids and detergents (CHAPSO or DHPC) in a ratio of 3:1 (w/w),

where the lipids contained 5 mol % cholesterol (Sigma-Aldrich) and 95 mol % DMPC. Immediately before preparing the protein/bicelle mixture, the concentrated proteins were incubated with 10 mM ATP (sodium salt) for 30 min at 4 °C. The proteins and 10% bicelles were then gently mixed in a ratio of 1:4 (v/v), such that the final protein concentration was 5–10 mg ml^{-1}.

The protein/bicelle mixture was incubated on ice for 30 min. The crystallization was set up in a hanging-drop vapour diffusion format at 20–22 °C by using VDX48 trays and mixing protein/bicelle preparation with equal-volume crystallization reservoir solution containing 1.7–2.0 M ammonium sulfate [$(NH_4)_2SO_4$], 100 mM MES pH 6.5 (or 100 mM HEPES pH 7.0), 2–5% PEG400 (or 2–5% PEG350 MME), and 1 mM TCEP. Crystals suitable for data collection appeared in 3 days to 2 weeks and reached a maximum size of 75–150 μm × 40–60 μm × 10–20 μm in 1–2 months, and decayed after 3 months of crystallization. Crystals used for structural analysis were harvested within 1–2 months. Crystals of similar morphology were also obtained when nucleotide was omitted or when a non-hydrolysable analogue of ATP (either AMPPNP or TNP-ATP) was added. We also tried to grow crystals of a catalytically deficient mutant consisting of WT G5 and G8 mutation (G216D)[32], attempting to solve a nucleotide-bound structure. No crystal growth was observed, and further optimization in crystallization will be necessary.

For experimental phasing, crystals were derivatized by adding 1 mM sodium phosphotungstate ($PW_{12}O_{40}{}^{3-}$, Sigma-Aldrich) to the crystallization drops for 12 h before harvesting. Crystals were cryo-protected in a solution containing 25% glycerol, 2 M $(NH_4)_2SO_4$, 100 mM MES, pH 6.5, and 2–4% PEG400. The crystals were harvested in cryoloops (Mitegen), flash-frozen and stored in liquid nitrogen.

Data collection, structure determination and refinement, final model validation, and uncertainty

X-ray diffraction data sets were collected at the Advanced Photon Source beamlines 19-ID and 23-ID-D. HKL3000 was used to process both the partial and full diffraction data sets used for the ABCG5/ABCG8 heterodimer structure solution[33,34]. Computational corrections for absorption in a crystal and for imprecise calculations of the Lorentz factor resulting from a minor misalignment of the goniostat were applied[35,36]. Anisotropic diffraction was corrected to adjust the error model and to compensate for a radiation-induced increase of non-isomorphism within the crystal[37–39]. Typical, well-behaving native crystals diffracted anisotropically to a resolution of ~3.9 Å in the x direction, ~4.0 Å in the y direction, and ~4.5 Å in the z direction, while sodium phosphotungstate ($PW_{12}O_{40}Na_3$, Sigma-Aldrich) derivatized crystals diffracted to a resolution of ~5.5 Å in the x and y directions, and ~6.5 Å in the z direction. The data processing statistics are presented in Extended Data Table 1.

Initial phases were obtained in a single-wavelength anomalous diffraction experiment with two crystals derivatized with sodium phosphotungstate, with data collected at the L-III edge of tungstate ($\lambda = 1.21$ Å). The estimated level of anomalous signal was ~6.5% of the native intensity. The diffraction data set was processed to a resolution of 5.0 Å, while the search for heavy atom positions was performed to a resolution of 7.0 Å. The eight positions of the tungstate cluster were identified using SHELXC/D[40], run within HKL3000, with correlation

coefficients $CC_{All} = 45.88\%$, $CC_{Weak} = 22.52\%$, and PATFOM = 21.56. The handedness of the best solution was determined with SHELXE, in which the radius of the sphere of the variance map calculation was redefined from 2.42 to 4.84 Å. The eight tungstate cluster positions were refined anisotropically to 5.2 Å with MLPHARE[41], with the final FOM reaching 0.220 for all observations. Twofold NCS was identified by PROFESS from CCP4 (ref. 42) NCS-averaging and solvent flattening was performed by DM[43] and later with PARROT[44]. The procedure produced a clean electron-density map to a resolution of about ~6.5 Å which showed alpha helical features, but had insufficient resolution to for model building. Therefore, we attempted to improve the phases by combining the phasing signal of the tungstate derivative with the phasing signals of a lead derivative (trimethyl lead chloride, $Pb(CH_3)_3Cl$, Sigma-Aldrich) and tantalum derivative (hexatantalum tetradecabromide, Ta_6Br_{14}, Jena Bioscience), for which the estimated levels of anomalous signal were below 1% of the native intensity, but which diffracted to better resolution. Three positions of Pb^{2+} were identified using anomalous difference Fourier maps phased with a solvent-flattened tungstate derivative. They were then introduced to MLPHARE together with the previously identified tungstate clusters positions, and refined together to a resolution of 4.2 Å, while the tantalum derivative served as a native data set.

Although the multiple isomorphous replacement phase combination improved the maps, the quality and resolution were still insufficient to build and refine an atomic model automatically. Therefore, we first located the NBD domains using a homologous model (3D31.pdb) and positioned them manually. Then, we placed the alpha helices using the 'Place Helix Here' option in the 'Other Modeling Tools' in Coot[45]. The topology of the transmembrane domains was identified manually and that defined the directionality of the α-helices, which at this resolution were initially placed in two possible directions by Coot. The resulting assembly was used to redefine the solvent mask to improve NCS averaging and then the model was further rebuilt and corrected by iterative application of BUCCANEER[46], Coot and REFMAC[47]. To proceed with the model building, we had to improve the resolution of the data, which was done by merging together 19 full and partial native data sets to a final resolution of 3.9 Å. Although the data were essentially complete in terms of Bragg's law, the anisotropy of diffraction resulted in an uneven distribution of information in the reciprocal space. Correcting for anisotropy retained informative observations of intensity in ellipsoidal resolution shells; however, all downstream procedures reported completeness in the spherical shells. Therefore, the completeness in refinement is lower than in scaling. The merged native data set was combined with the PARROT-modified phases obtained during an earlier step and was used in the model building with BUCCANEER and in the refinement with REFMAC. The initial model served as the starting point for BUCCANEER, which rebuilt it to ~65% of completeness and partly assigned it to the sequence, with ~60% of its side chains docked. Intermediate models from different cycles of BUCCANEER were combined into a more complete model and rebuilt manually. The restrained, phase-stabilized refinement was performed with the Hendrickson–Lattman coefficients from PARROT down-weighted by factor of 0.5 and blurred with a B-factor of 200 $Å^2$. Additionally, the ProSMART[48] option 'to generate H-bond restraints (e.g. secondary structure restraints)' was used to stabilize the model geometry during refinement, together with REFMAC's local NCS restraints, jelly body refinement, and B-factor values restrained with a weight of 0.2

and to a range of allowable values 20–400. The resolution cutoff of 3.94 Å in the refinement was selected on the basis of multiple refinements in which the R_{free} values in the last resolution shell were inspected and the features of electron density maps were analysed. All these parameters were chosen to stabilize the refinement and reduce bias. The model quality was validated with Molprobity[49] and assessed as satisfactory with a Molprobity score of 3.47, which corresponds to the 71st percentile in comparison with the set of 342 Protein Data Bank deposits solved at resolutions from 3.25 to 4.18 Å.

Although nominal resolution of diffraction data extends to 3.94 Å, high anisotropy results in uneven quality of maps in real space, with some electron density features being very well defined while other features have less definition. Specifically, the amino-acid register of the NBDs, having predominantly β-sheet structure, is highly uncertain when based on electron density alone. We used Robetta models[50], sequence homology, and analysis of plausibility of chemical interactions to validate the amino-acid register, but we expect that registry errors are possible in these domains, in particular in regions that do not have reliable alignments to homologous structures. In the case of transmembrane domains, we do not expect major registry problems; however, minor mistracings of loops are possible owing to an uncertainty that is related to the C-caps of helices, which can have multiple alternative conformations that are not discernible at this resolution. Finally, although the fragment that connects the NBD with the transmembrane domains (from residues ~320 to ~395 both in ABCG5 and in ABCG8) can be traced for ABCG5 (chains A and C), it is less well defined in the ABCG8 subunits (chains B and D). For chain B, a part of this fragment is stabilized in the α-helical conformation by crystal lattice interactions, but the linker sequences connecting this fragment to the NBD on one side and the transmembrane domain on the other are not visible in the electron density, and therefore the register for this fragment is highly uncertain.

ATPase assay

The ATPase activity of purified G5G8 was determined as described[12,51]. Briefly, 4–10 μg proteins (final concentration 27–67 μg ml^{-1}) were mixed with 100 μg liver polar lipids (Avanti) and 20 mM DTT for 10 min at room temperature (22 °C), and then left on ice (used within 2 h). Reactions were performed in a final volume of 150 μl containing 50 mM Tris/MES (pH 7.0), 30 mM KCl, 5 mM $MgSO_4$, 2 mM ^{32}P-γ-ATP, 4.5 mM sodium azide, and 1% sodium cholate at 37 °C. Released inorganic phosphate was extracted by molybdate and monitored by ^{32}P radioactivity to measure specific activity in three independent experiments. As a negative control, we paired a catalytically deficient G8 mutation (G216D)[32] with WT G5 in the assay.

Generation of anti-human G5 and G8 antibodies

A DNA fragment encoding the N-terminal region of human G8 (residues 2–400) was PCR-amplified from the G8 cDNA and cloned into the pET30a^{+} vector (Novagen). The peptide was expressed in BL21 (DE3) competent *Escherichia coli* cells (Novagen) and then isolated from inclusion bodies, solubilized in 8 M urea, and purified using a Ni-NTA column in the presence of 8 M urea. Rabbits were injected every 2 weeks with 0.1 mg of the peptide to generate anti-G8 polyclonal antibodies. To generate monoclonal anti-human G5 antibodies, splenic B lymphocytes were isolated from female BALB/c mice ($n = 2$) that had been

immunized eight times with 50 μg of purified G5G8 (see above). Cells were incubated with SP2/mIL-6 mouse myeloma cells and the resultant hybridomas were screened using an ELISA assay. Positive clones that recognized denatured G5 by immunoblotting were selected for subcloning. Class 1 immunoglobulin-γ (IgG1) was purified from the supernatant of cultured hybridoma cells using gravity-flow affinity chromatography with Protein-G Sepharose-4 Fast Flow beads.

In vivo functional reconstitution cholesterol transport assay

Point mutations were introduced into the human G5 cDNAs using QuickChange II site-directed mutagenesis kits (Agilent). The recombinant adenoviruses expressing human WT or mutant G5 and G8 were generated using an AdenoVator adenoviral vector system (QBioGene). Total knockout (KO) (Abcg5/Abcg8$^{-/-}$)[19] and liver-specific KO (L-Abcg5/Abcg8$^{-/-}$)[52] mice were maintained on a regular chow diet. Adenoviral particles (5×10^{12} particles per kilogram), containing no external gene (RR5), WT, or mutant human G5G8, were injected into the tail veins of the mice. After 72 h, the mice were fasted for 4 h, anaesthetized with halothane, and killed by exsanguination. Bile was collected, and neutral sterol levels were measured using gas liquid chromatography and mass spectrometry as described[4]. Liver tissue was snap frozen in liquid nitrogen and stored at –80 °C.

G5G8 coevolution analysis

Information about the covariance of residue substitution patterns in multiple sequence alignments (MSAs) or co-evolution can provide a basis for predicting structure contacts[53,54], and has been successfully applied to predicting residue–residue interactions across protein interfaces[23]. We employed the GREMLIN server[53] to predict the top L/3 co-evolving residue pairs in the TMD of ABCG5 and ABCG8 (see Supplementary Information). We mapped the co-evolving pairs that were separated by at least four residues in the primary sequence to the corresponding structures and compared them with the ABCG5 and ABCG8 TMD structure residue–residue contact maps using CMview[55]. GREMLIN co-evolving residue pairs from the ABCG5 and ABCG8 TMD that were closer than an all-atom cutoff of 8 Å represent the majority of co-evolving residue pairs and are depicted as lines connecting Cα atoms of the residues in contact in the respective TMD structures. The remaining residues that are spatially distant within the G5 or G8 molecules are candidates for forming interfacial contacts between G5 and G8 (ref. 23). For those remaining Gremlin pairs in TMH1, TMH2, and TMH5, we used the Gremlin alignment to map residues from G5 (or G8) to the corresponding residues on the opposite molecule. The highest scoring interfacial Gremlin pairs (three pairs with >0.95 probability of co-evolving, corresponding to six potential symmetry-related pairs) were between residues in TMH2 (G8 L459, F461, and Y465 or the corresponding G5 T430, Y432, and L436) and residues in TMH5 (G5 A535, I539, S538 or the corresponding G8 N564, N568, and Y567, respectively) (Extended Data Fig. 7).

Molecular dynamics simulation

CHARMM-GUI (http://www.charm-gui.org) was applied to add DMPC lipid bilayer, cholesterol, counter ions, and water molecules. The entire system consisted of one copy of the G5G8 biological unit (chains A and B), 16 cholesterol (CHL), 304 DMPC, 128 Na$^+$, 180

Cl$^-$, and 47,789 TIP3P water molecules[56]. In total, there were 163,088 atoms in the simulation box for the whole system. For force field parameters, the partial atomic charges of CHL were derived by RESP[57] to fit the HF/6-31G electrostatic potentials generated using the GAUSSIAN 09 software package (revision D.01). The other force field parameters came from GAFF in AMBER12 (ref. 58). The residue topology of CHL was prepared using the ANTECHAMBER module in AMBER12 (ref. 59). The force fields of AMBER FF12SB[60] and LIPID14 (ref. 61) were used to model proteins and lipids, respectively.

All molecular dynamics simulations were performed with periodic boundary condition to produce isothermal–isobaric ensembles using the PMEMD.CUDA program in AMBER12. The particle mesh Ewald method[62] was used to calculate the full electrostatic energy of a unit cell in a macroscopic lattice of repeating images. All bonds were constrained using the SHAKE algorithm[63] in molecular dynamics simulations. Temperature was regulated using Langevin dynamics[64] with a 5 ps^{-1} collision frequency. Pressure was regulated using the isotropic position scaling algorithm with the pressure relaxation time set to 1.0 ps. The integration of the equations of motion was conducted at a time step of 1 fs for the relaxation and equilibrium phases and 2 fs for the sampling phases. Before molecular dynamics simulations, the systems were relaxed to remove any possible steric crashes by a set of 10 thousand-step minimizations with the main chain atoms restrained. The harmonic restraint force constants decreased from 20 to 10, 5, and 1 kcal mol^{-1} Å^{-2}, progressively. At last, the systems were further relaxed by a 10,000-step minimization without any constraint or restraint. There are three phases in a molecular dynamics simulation: the relaxation phase, the equilibrium phase, and the sampling phase. In the relaxation phase, the system was gradually heated up from 50 K to 300 K in steps of 50 K. At each temperature, molecular dynamics simulation was run for 1 ns. In the following equilibrium phase, the system was further equilibrated for 2 ns at 298 K. In the sampling phase, 50,000 snapshots were collected at an interval of 2 ps for post-analysis. All the post-analysis was performed using the CPPTRAJ module of AMBER12.

The stability of the molecular dynamics trajectory was monitored with the backbone r.m.s.d. along the simulation trajectory, showing that the current crystal structure was very stable along the whole simulation course with an overall r.m.s.d. of 3.0 Å, which is reasonable for a structure resolved at 3.9 Å. A subset of 10,000 snapshots was evenly selected for the solvent-accessible surface calculation and cluster analysis. We found no obvious trend for the solvent-accessible surface areas of TMHs. We then performed fixed radius clustering analysis (radius was set to 2.0 Å) for the subset using the clustering toolkit of MMTSB (http://www.mmtsb.org). The biggest cluster, 5,793 members and the shortest distance to the cluster centre, was selected to compare with the crystal structure. Only slight narrowing in the TMD could be observed, but the NBD had larger deviations, probably owing to the missing residues not resolved in this study.

The quasi-harmonic analysis was performed for the main chain atoms using all 50,000 snapshots collected every 2 ps. First, an average molecular dynamics structure was calculated by aligning the molecular dynamics snapshots to the crystal structure using the main chain atoms; each snapshot was then realigned to the average structure to generate the mass-weighted covariance matrix. Vibrational frequencies and modes, which are the

eigenvalues and eigenvectors of the covariance matrix, were then obtained. For each vibrational mode, the contribution of each residue was calculated by adding up the vectors of the main chain atoms. The low-frequency modes usually correspond to the global movements of a protein related to its biological function[65]. Among the lowest 20 vibrational modes, modes 9 and 10 were selected for further examination (Extended Data Fig. 6). Under current quasi-harmonic analysis, we were not able to observe a mode describing opening or closing at the TMD interface. Experimental structures of different catalytic or substrate-bound states will be necessary.

Homology modelling

The homology model of the white/brown heterodimer that confers brown eye colour to *Drosophila* was generated using Modeller[28] with the G5G8 crystal structure as the template. The sequence alignment for homology modelling was generated by PROMALS3D[66]. One hundred homology models were generated and the one with the best DOPE score was selected as the final model. The final model was further evaluated using the PROTABLE module of SYBYL software (https://www.certara.com) and no severe violation was identified in the Ramachandran plot.

G5G8 sequence analysis

To collect ABCG family sequences, we ran PSI-BLAST (three iterations, E value cutoff 0.001) against the NCBI NR database using the G5 TMD as a query sequence (gi|11692800, residue range 368–651). Collected sequences were clustered using CLANS[26]. Sequences that did not cluster with the eukaryotic ABCG family were excluded to make a subset of ABCG sequences. To establish residue conservations for the G5 and G8 subfamily subsets, sequences from our initial PSI-BLAST against the NR database that clustered together with the human G5 and G8 sequences were kept, and an MSA was generated for each subfamily using MAFFT[67]. MSA results of G5 and G8 are shown in the Supplementary Information. Redundant and incomplete sequences were manually removed from the MSA. The MSA was used to generate positional conservations for each subfamily and they were mapped to the last line of the MSA using a scale ranging from variable to conserved (scale 0–9), to the structure B-factors in G5 (scale −1.82 to1.52), and in G8 (scale −2.04 to 1.41) using the program AL2CO[68]. Residues with conservation values of 0.6 or greater in the B-factors (MSA conservation value ~7) were considered highly conserved.

To gain a broader view of all sequences related to the helical domain of TMD, we collected all related eukaryotic sequences by searching the SWISSPROT sequence database with the same G5 query. We also collected related prokaryotic sequences by initiating PSI-BLAST with representative queries from all COGS that belong to the ABC2 exporter superfamily clan (COG1277, COG1682, COG2386, COG559, COG3694, COG4200, and COG4587). To cut down on redundancy of the NR database, we searched a subset of 122 high-quality curated bacterial genomes designated by NCBI as reference genomes (http://www.ncbi.nlm.nih.gov/genome/browse/reference/). The sequences, identified in the SWISSPROT database, were combined with those identified in the prokaryotic reference genomes and were clustered in two dimensions using CLANS[26] to visualize the ABC2 exporter superfamily sequence relationships. Export substrates by prokaryotic transporters in

the ABC2 exporter superfamily include polysaccharide teichoic acids, which are found within cell walls of Gram-positive bacteria (TagG)[69]; lipo-oligosaccharides, which are located in outer membranes of Gram-negative bacteria (NodJ)[70]; haem, which is required for cytochrome C biogenesis (CcmB)[71]; and bacterial-killing lantibiotic peptides such as mutacin (MutG)[72] and bacitracin (BcrB)[73].

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Extended Data
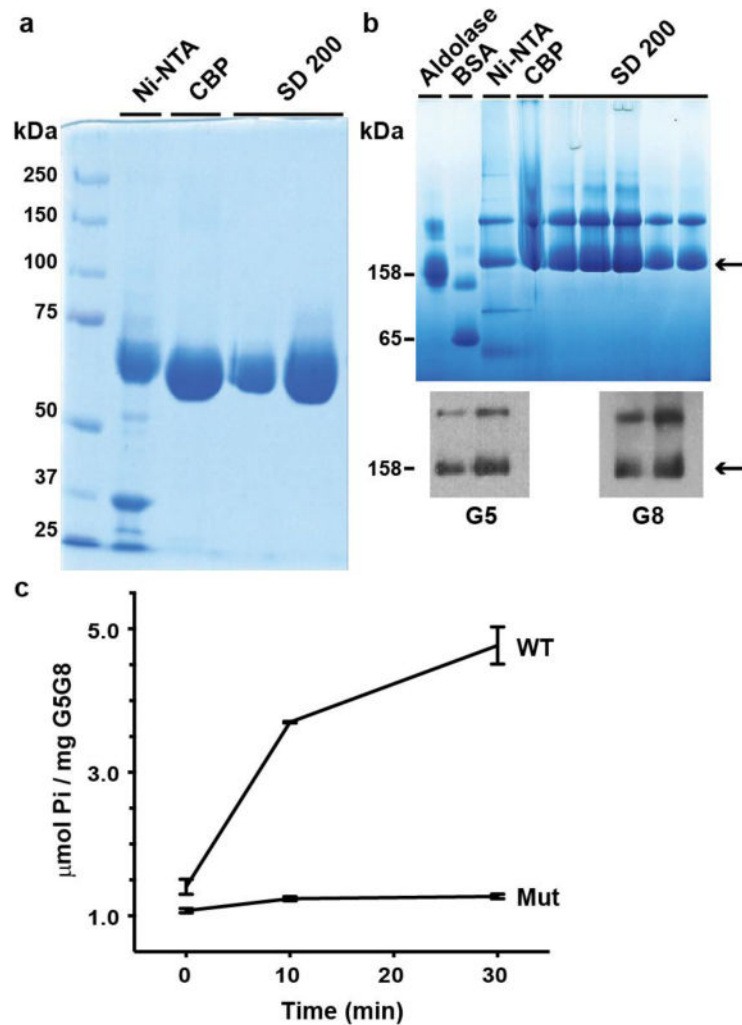
Extended Data Figure 1. Purification and ATPase activity of G5G8 heterodimers
a, G5G8 heterodimers were co-purified by sequential affinity chromatography (Ni-NTA and CBP) followed by gel-filtration chromatography using Superdex 200 (SD 200) as described in the Methods. Protein purity was analysed by sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS–PAGE) and Coomassie staining. **b**, Dimerization of detergent-soluble G5G8 was analysed by size fractionation using blue native PAGE (BN-PAGE) followed by

Coomassie staining of the gel (top). Proteins were transferred to nitrocellulose and immunoblotting was performed using an anti-RGSH$_4$ monoclonal antibody (Qiagen) and anti-G8 rabbit polyclonal antibodies (bottom). Arrows point to the G5G8 dimer. **c**, Bile acid stimulated ATPase activity of purified recombinant WT G5 and G8 (4 μg) and WT G5 and catalytic-deficient G8 (G8-G216D) (Mut, 8 μg). Each point represents the mean ± s.d. of three separate measurements.
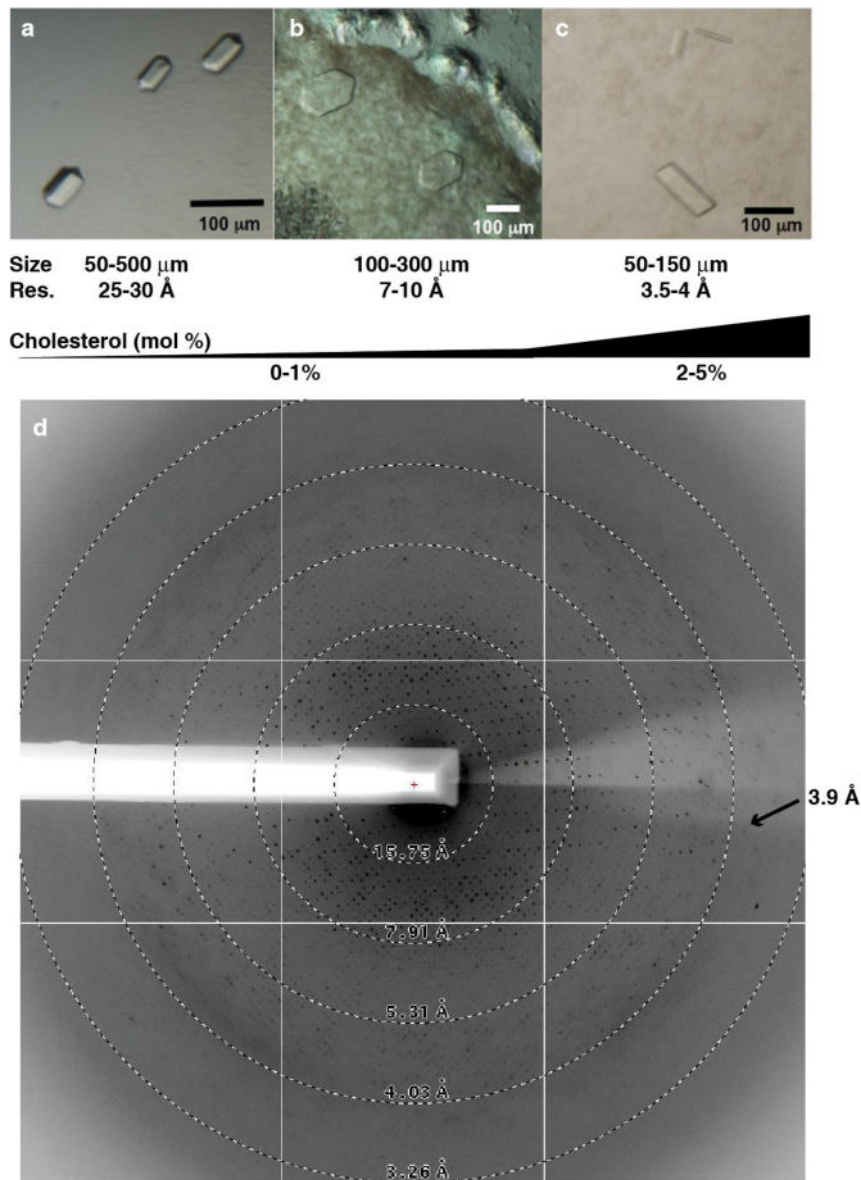
Extended Data Figure 2. Optimization of crystal growth and X-ray diffraction
Crystals of G5G8 were grown using bicelle crystallization (see Methods). Optimal diffraction-quality crystals were obtained by using cholesterol as additive to the bicelle matrices. Shown here are crystals observed under a light microscope without the polarizer. Scale bar, 100 μm. **a**, **b**, In the presence of 0–1% (mol) cholesterol, crystals grew up to 500 μm, but only diffracted to 7–10 Å. **c**, In the presence of 2–5% (mol) cholesterol, crystals

grew to 50–150 μm (longest edge) and diffracted to 3.5–4 Å. **d**, A diffraction frame shows structural information of G5G8 crystals out beyond 3.9 Å resolution.
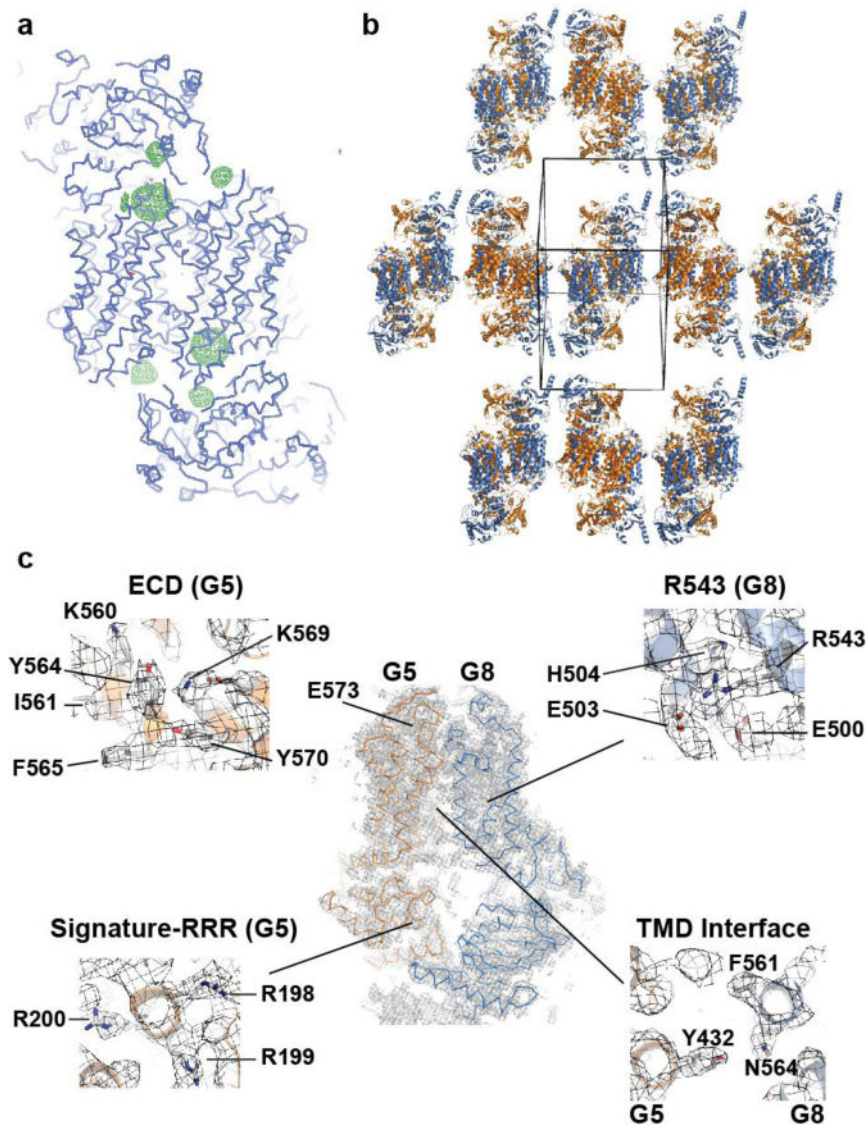

Extended Data Figure 3. Experimental electron density maps and crystal packing in G5G8 bicelle crystals

a, Side view of Cα backbone for two G5G8 dimers in an asymmetric unit (containing two transporters) with the anomalous Fourier map calculated from W-clusters plotted at 3σ. **b**, Orthorhombic G5G8 bicelle crystals with lateral crystal packing of TMDs indicative of G5G8 reconstitution into phospholipid bilayers (G5: orange; G8: blue). **c**, Electron density maps ($2F_o − F_c$) show the complete G5G8 heterodimer (centre, contoured at 1.5σ), selected residues at the G5 ECD (upper left, 1.5σ), conserved three arginines of G5 (R198/R199/R200) in proximity to the Signature motif (bottom left, 1.0σ), R543 of G8 where point mutation R543S causes sitosterolaemia (upper right, 1.5σ), and Y432 on TMH5 (G5) and F561 on TMH5 (G8) at the TMD interface (bottom right, 1.5σ).
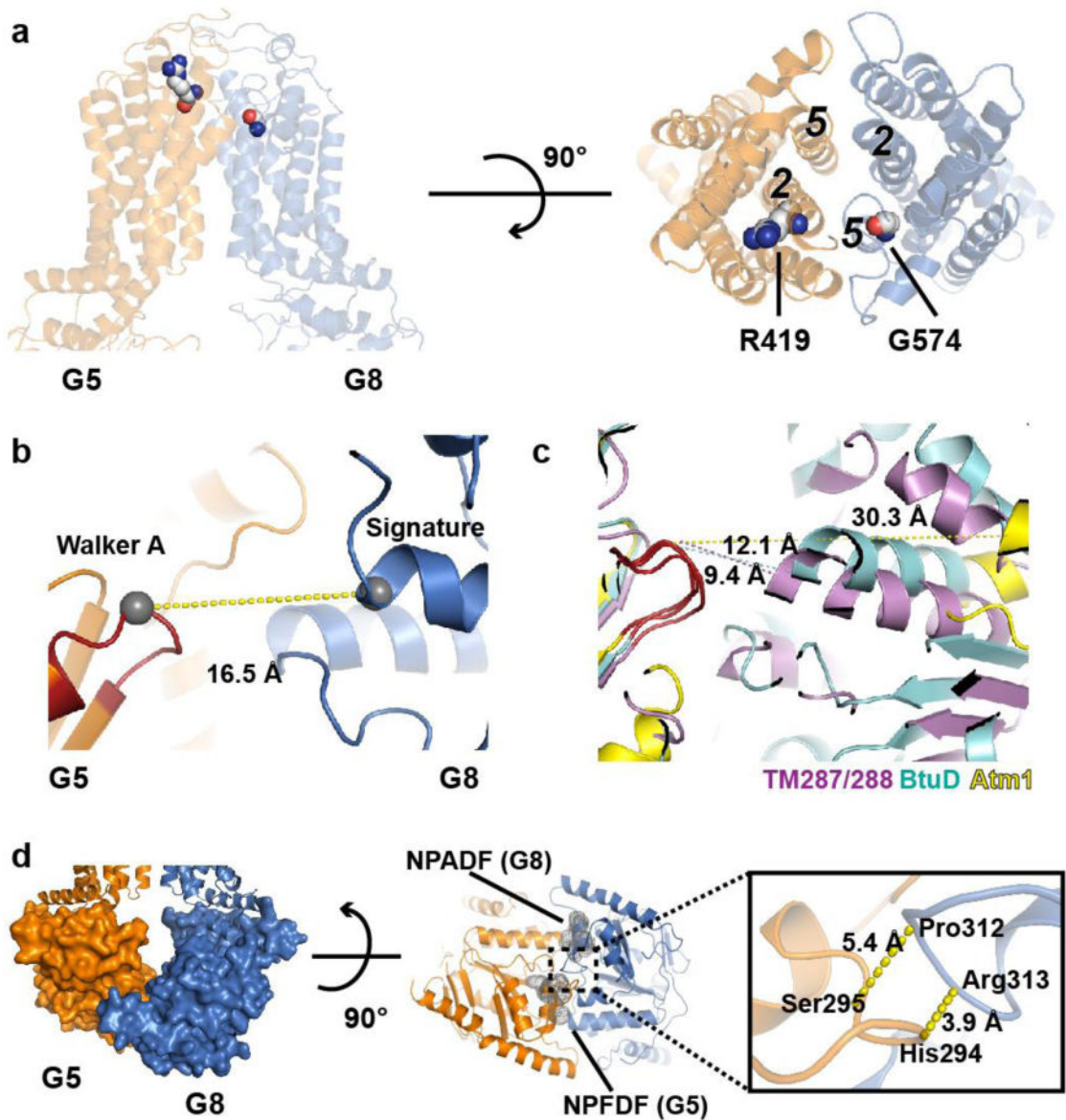
Extended Data Figure 4. G5G8 heterodimer interface
a, Sitosterolaemia mutation positions R419 (G5) and G574 (G8) at the G5-TMH2 and G8-TMH5 interface near the extracellular surface. **b**, Separation between the +4 glycine of G5 Walker A (G5, GSS*G*SGKT) and G8 Signature (LSG*GE*) motifs. **c**, Superposition of the nucleotide-free NBDs of TM287/288 (PDB accession number 4Q4H), BtuD (PDB accession number 2QI9), and Atm1 (PDB accession number 4MYC) with varying distances between NBDs. **d**, Left: surface-filled view of the NBDs of G5G8 showing contacts between the two half-transporters at the cytoplasmic apex. Middle: interaction between conserved NPFDF (G5) and NPADF (G8) sequences (grey dots). Right: closest distances between the two half-transporters.
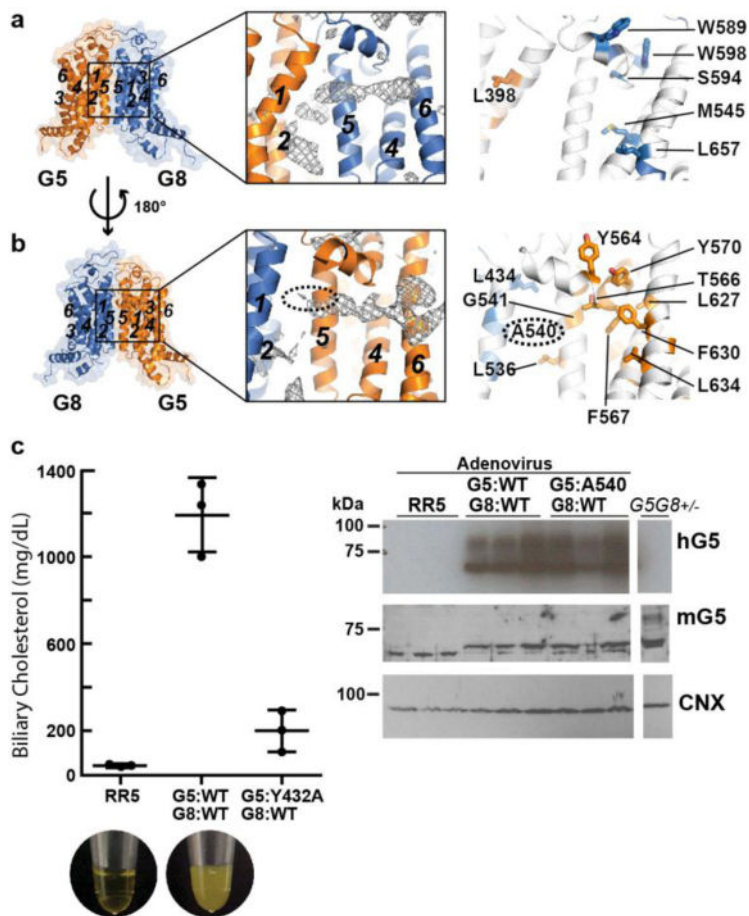
Extended Data Figure 5. Vestibules in the membrane spanning region

a, **b**, Left: cartoon of G5G8 TMDs in transparent surface (orange and blue). Vestibules on opposing faces of the TMDs are highlighted in boxes. Middle: $F_o - F_c$ difference electron density map contoured at 3.0σ showing extended features at the vestibules. TMHs are numbered. Right: highly conserved residues in G5 and G8 (sequence conservation 7, see Methods) are shown as orange (G5) and blue (G8) cartoon/sticks. **c**, Left: *in vivo* functional reconstitution assay using *G5G8* KO mice (*G5G8*−/− mice)[19]. Biliary cholesterol levels from mice infected with RR (empty adenovirus), or with adenoviruses expressing human (h)G5:WT and hG8:WT, or with adenoviruses expressing hG5:A540F and hG8:WT are shown. Pictures of the bile are below the graph. Each experiment represents the mean ± s.d. from three infected mice (n = 3) in each group. Right: expression of hG5 and hG8 detected by immunoblotting using anti-hG5 monoclonal antibodies (see Methods). Connexin was used as the gel-loading control. As a positive control for the mouse and a negative control for the human anti-G5 antibody, we also immunoblotted liver membranes from *G5G8*+/− mice.
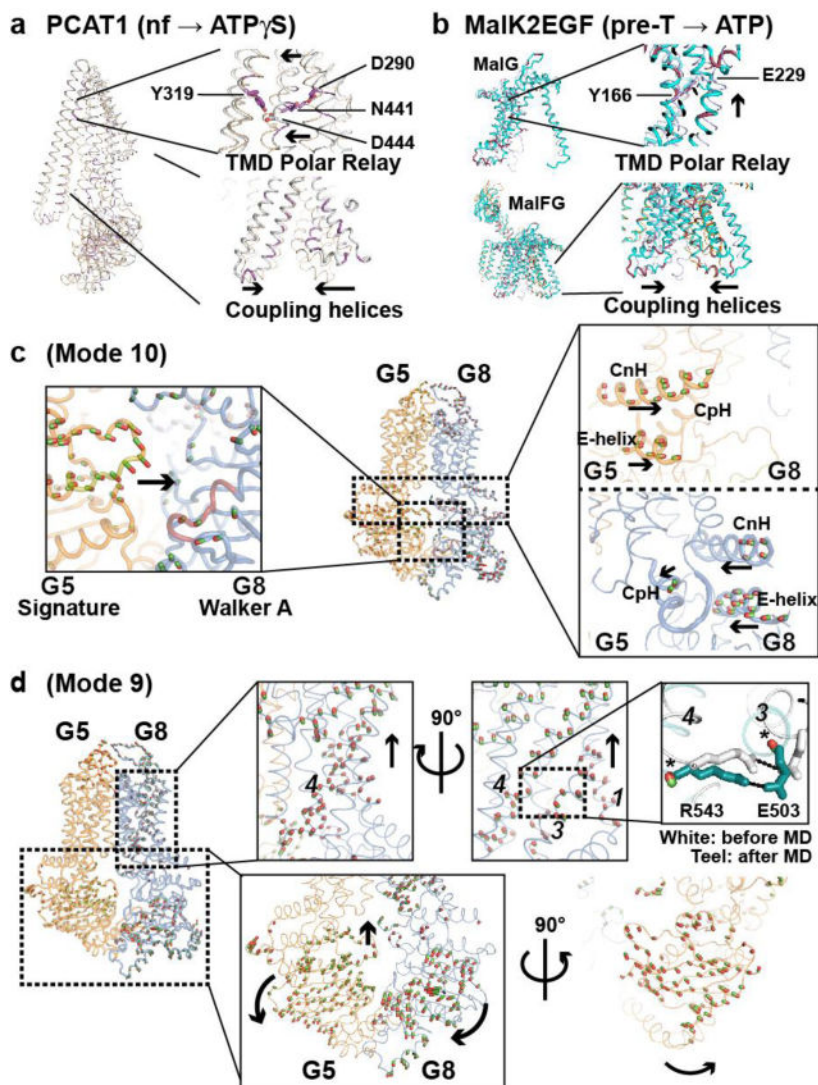
Extended Data Figure 6. TMD polar relay in other ABC transporters and molecular dynamics simulation of G5G8 structure

a, b, Analysis of bacterial PCAT1 (**a**) and maltose transporter (MalK₂EGF, **b**) structures reveals conserved polar residues in the TMD subunits. TMD superposition of a nucleotide-free structure with a nucleotide-bound structure shows breakage of hydrogen bonding in TMD polar relay (top right) and inward movement of coupling helices (bottom right). (Nucleotide-free PCAT1 (PDB accession number 4RY2), ATPγS-bound PCAT1 (PDB accession number 4s0f), nucleotide-free MalK₂EGF (PDB accession number 3PV0), and ATP-bound MalK₂EGF (PDB accession number 2R6G).) **c, d,** Molecular dynamics simulation (100 ns) and vibrational mode performed on the nucleotide-free G5G8 structure (see Methods). For each vibrational mode, the contribution of each residue was calculated by adding up the vectors of the main chain atoms. Among the lowest 20 vibrational modes, modes 9 and 10 are shown. Mode 10 (**c**) describes collective movements (green-to-red vectors on the Cα atoms) between CpH/CnH/E-helix bundle and the interface at ATP-binding cassette, indicating inward movement of the CpH/CnH/E-helix bundles from both G5 and G8. This is consistent with stereotypical inward motion of coupling helices in
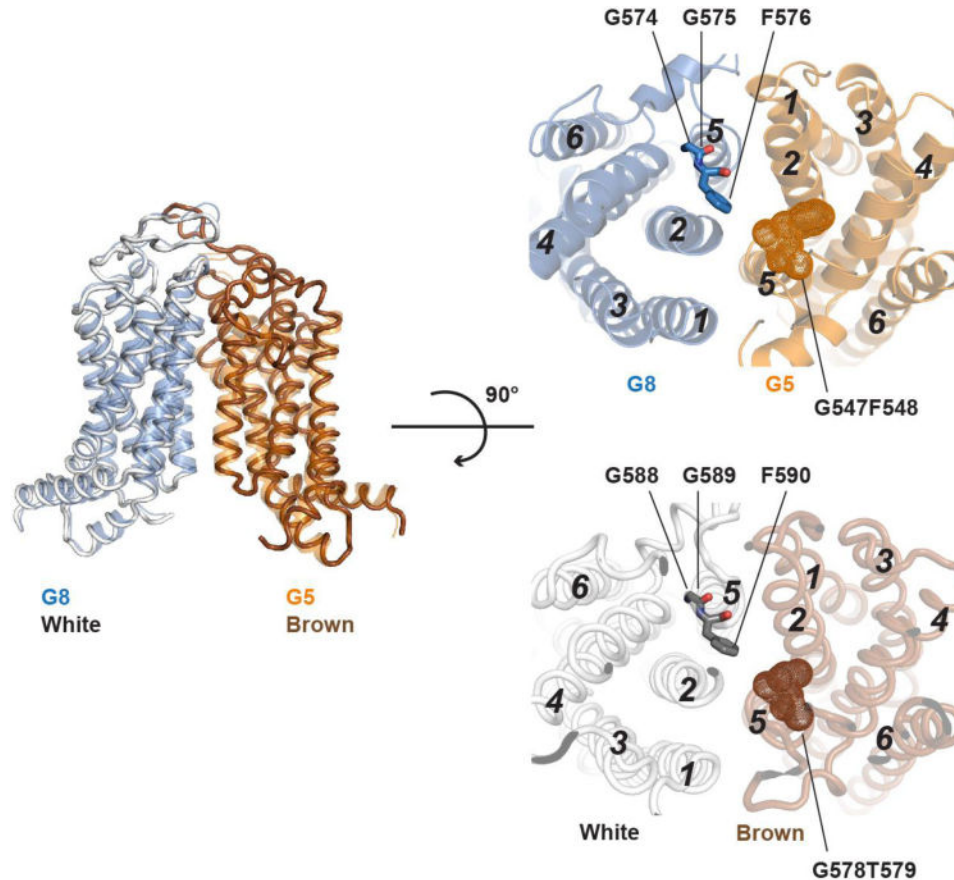
bacterial ABC transporters. Mode 9 (**d**) describes a collective movement (arrows) of NBD (bottom) and TMH1, 3, and 4 in G8 (top), indicating coupling between NBD and TMD. R543 of G8, a conserved polar residue in the TMD polar relay where a disease-causing mutation occurs, is highlighted together with its interacting partner, E503.
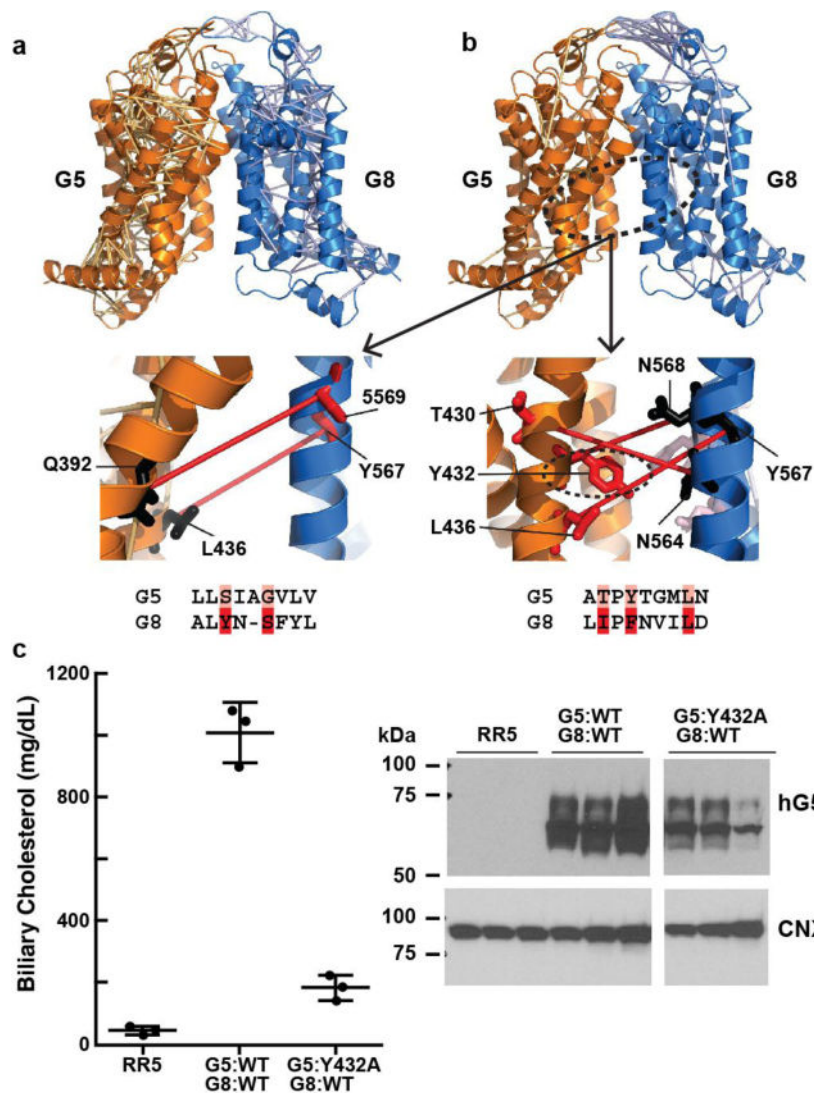
Extended Data Figure 7. Coevolution analysis of ABCG5 and ABCG8

a, Co-evolving residue pairs from the ABCG5 and ABCG8 TMD that were closer than an all-atom cutoff of 8 Å (8 Å) are depicted as lines connecting Cα atoms of the residues in contact in the respective TMD structures. **b**, Co-evolving pairs that are more spatially distant than the defined cutoff (>8 Å). Expanded views show potential interface contacts predicted by spatially distant G5 and G8 coevolving residue pairs (red lines connecting Cα atoms) (top). Cross contacts in G8 (red) were mapped from the distant G5 GREMLIN pairs near the interface (pink) using the GREMLIN MSA, and vice versa (bottom). For example, we mapped S538 from the predicted G5 pair L436/S538 to the corresponding residue in G8-Y567, and considered G5-L436 and G8-Y567 to be a potential interface contact. Predicted

interface contacts that are distant in the nucleotide-free structure could form contacts at another stage of the ABCG ATPase cycle. **c**, Left: *in vivo* functional reconstitution assay using liver-specific *G5G8*[−/−] mice[52]. Biliary cholesterol levels of mice infected with RR (empty adenovirus), or with adenoviruses expressing hG5:WT and hG8:WT, or with adenoviruses expressing hG5:Y432A and hG8:WT are shown. G5-Y432 is part of the TMD polar relay and predicted to interact with G8-N568 during the sterol transport cycle (**b**). Each experiment represents the mean ± s.d. from three infected mice (n = 3) in each group. Right: immunoblot analysis of hG5 levels in membranes isolated from the mice using an anti-hG5 monoclonal antibody (see Methods). Connexin was used as the gel-loading control.

Extended Data Figure 8. Homology model of *Drosophila* white/brown based on the G5G8 structure

A model of the *Drosophila melanogaster* white/brown heterodimeric guanine pigment transporter was made using the program Modeller[28]. On the basis of the sequence alignment and comparison of scoring metrics for the two possible heterodimer models (that is, white modelled on G5 versus G8, brown on the other), we selected G8 (blue cartoon) as the white template and G5 (orange cartoon) as the brown template. At right, top/extracellular views of the G5G8 structure and white/brown model are shown, with residues important for dye interaction highlighted. At bottom is an MSA for the C-terminal region of TMH5, showing the conservation of the functionally important amino acids.

Extended Data Table 1

Data processing and refinement statistics.

	Native[*]	$[PW_{12}O_{40}{}^{3-}]^{\dagger}$	$[(CH_3)_3Pb^-]$	$(Ta_6Br_{12}{}^{2+})$
Data collection				
Beamline	19-ID-D/23-ID-D[‡]	19-ID-D	19-ID-D	19-ID-D
Space group	I 222	I 222	I 222	I 222
Cell dimensions				
a, b, c (Å)	173.6, 224.8, 253.3	175.5, 227.5, 254.5	174.6, 225.9, 253.4[§] 173.6, 225.9, 252.7[‖]	176.0, 228.0, 253.7
Resolution (Å)	50–3.9 (3.93–3.9)	50–5.0 (5.04–5.0)	50–4.5 (4.54–4.5)	50–5.0 (5.04–5.0)
R_{sym} or R_{merge}	16.1 (NA)	13.5 (33.5)	8.7 (NA)[§] 7.1 (NA)[‖]	8.8 (NA)
$\langle I \rangle / \langle \sigma I \rangle$	8.8 (0.15)	5.1 (1.4)	8.0 (0.45)[§] 6.1 (0.18)[‖]	8.6 (0.50)
Completeness (%)	99.4 (84.2)	94.4 (47.3)	97.4 (55.9)[§] 94.7 (54.7)[‖]	81.1 (18.3)
Redundancy	18.9 (2.5)	3.1 (1.7)	6.0 (2.3)[§] 4.3 (2.5)[‖]	3.7 (1.3)
Refinement				
Resolution (Å)	25–3.94			
No. reflections	34889			
R_{work}/ R_{free}	24.5/32.9			
No. atoms				
Protein	18151			
R.m.s deviations				
Bond lengths (Å)	0.010			
Bond angles (°)	1.64			

[*] Number of crystals: 19.

[†] Number of crystals: 2.

[‡] MAR CCD detector was used.

[§] Data collected at peak wavelength: 13.053 keV.

[‖] Data collected at inflection wavelength: 13.043 keV.

Acknowledgments

References

1. Berge KE, et al. Accumulation of dietary cholesterol in sitosterolemia caused by mutations in adjacent ABC transporters. Science. 2000; 290:1771–1775. [PubMed: 11099417]

2. Lee MH, et al. Identification of a gene, ABCG5, important in the regulation of dietary cholesterol absorption. Nature Genet. 2001; 27:79–83. [PubMed: 11138003]

3. Lu K, et al. Two genes that map to the STSL locus cause sitosterolemia: genomic structure and spectrum of mutations involving sterolin-1 and sterolin-2, encoded by ABCG5 and ABCG8, respectively. Am J Hum Genet. 2001; 69:278–290. [PubMed: 11452359]

4. Graf GA, et al. Coexpression of ATP-binding cassette proteins ABCG5 and ABCG8 permits their transport to the apical surface. J Clin Invest. 2002; 110:659–669. [PubMed: 12208867]

5. Graf GA, et al. ABCG5 and ABCG8 are obligate heterodimers for protein trafficking and biliary cholesterol excretion. J Biol Chem. 2003; 278:48275–48282. [PubMed: 14504269]

6. Theodoulou FL, Kerr ID. ABC transporter research: going strong 40 years on. Biochem Soc Trans. 2015; 43:1033–1040. [PubMed: 26517919]

7. Dean M, Rzhetsky A, Allikmets R. The human ATP-binding cassette (ABC) transporter superfamily. Genome Res. 2001; 11:1156–1166. [PubMed: 11435397]

8. Caffrey M. Crystallizing membrane proteins for structure determination: use of lipidic mesophases. Annu Rev Biophys. 2009; 38:29–51. [PubMed: 19086821]

9. Rosenbaum DM, et al. GPCR engineering yields high-resolution structural insights into beta2-adrenergic receptor function. Science. 2007; 318:1266–1273. [PubMed: 17962519]

10. Lin DYW, Huang S, Chen J. Crystal structures of a polypeptide processing and secretion transporter. Nature. 2015; 523:425–430. [PubMed: 26201595]

11. Chen S, Oldham ML, Davidson AL, Chen J. Carbon catabolite repression of the maltose transporter revealed by X-ray crystallography. Nature. 2013; 499:364–368. [PubMed: 23770568]

12. Johnson BJH, Lee JY, Pickert A, Urbatsch IL. Bile acids stimulate ATP hydrolysis in the purified cholesterol transporter ABCG5/G8. Biochemistry. 2010; 49:3403–3411. [PubMed: 20210363]

13. Faham S, et al. Crystallization of bacteriorhodopsin from bicelle formulations at room temperature. Protein Sci. 2005; 14:836–840. [PubMed: 15689517]

14. Hohl M, Briand C, Grütter MG, Seeger MA. Crystal structure of a heterodimeric ABC transporter in its inward-facing conformation. Nature Struct Mol Biol. 2012; 19:395–402. [PubMed: 22447242]

15. Ward AB, et al. Structures of P-glycoprotein reveal its conformational flexibility and an epitope on the nucleotide-binding domain. Proc Natl Acad Sci USA. 2013; 110:13386–13391. [PubMed: 23901103]

16. Hollenstein K, Frei DC, Locher KP. Structure of an ABC transporter in complex with its binding protein. Nature. 2007; 446:213–216. [PubMed: 17322901]

17. Hvorup RN, et al. Asymmetry in the structure of the ABC transporter-binding protein complex BtuCD-BtuF. Science. 2007; 317:1387–1390. [PubMed: 17673622]

18. Wang F, Li G, Gu HM, Zhang DW. Characterization of the role of a highly conserved sequence in ATP binding cassette transporter G (ABCG) family in ABCG1 stability, oligomerization, and trafficking. Biochemistry. 2013; 52:9497–9509. [PubMed: 24320932]

19. Yu L, et al. Disruption of Abcg5 and Abcg8 in mice reveals their crucial role in biliary cholesterol secretion. Proc Natl Acad Sci USA. 2002; 99:16237–16242. [PubMed: 12444248]

20. Hollenstein K, Dawson RJP, Locher KP. Structure and mechanism of ABC transporter proteins. Curr Opin Struct Biol. 2007; 17:412–418. [PubMed: 17723295]

21. Wang J, et al. Sequences in the nonconsensus nucleotide-binding domain of ABCG5/ABCG8 required for sterol transport. J Biol Chem. 2011; 286:7308–7314. [PubMed: 21209088]

22. Oldham ML, Chen J. Crystal structure of the maltose transporter in a pretranslocation intermediate state. Science. 2011; 332:1202–1205. [PubMed: 21566157]

23. Ovchinnikov S, Kamisetty H, Baker D. Robust and accurate prediction of residue-residue interactions across protein interfaces using evolutionary information. eLife. 2014; 3:e02030. [PubMed: 24842992]

24. Wang B, Dukarevich M, Sun EI, Yen MR, Saier MH Jr. Membrane porters of ATP-binding cassette transport systems are polyphyletic. J Membr Biol. 2009; 231:1–10. [PubMed: 19806386]

25. Ren Q, Chen K, Paulsen IT. TransportDB: a comprehensive database resource for cytoplasmic membrane transport systems and outer membrane channels. Nucleic Acids Res. 2007; 35:D274–D279. [PubMed: 17135193]

26. Frickey T, Lupas A. CLANS: a Java application for visualizing protein families based on pairwise similarity. Bioinformatics. 2004; 20:3702–3704. [PubMed: 15284097]

27. Morgan TH. Sex limited inheritance in Drosophila. Science. 1910; 32:120–122. [PubMed: 17759620]

28. Webb B, Sali A. Protein structure modeling with MODELLER. Methods Mol Biol. 2014; 1137:1–15. [PubMed: 24573470]

29. Ewart GD, Cannell D, Cox GB, Howells AJ. Mutational analysis of the traffic ATPase (ABC) transporters involved in uptake of eye pigment precursors in Drosophila melanogaster. Implications for structure-function relationships. J Biol Chem. 1994; 269:10370–10377. [PubMed: 8144619]

30. Mackenzie SM, et al. Mutations in the white gene of Drosophila melanogaster affecting ABC transporters that determine eye colouration. Biochim Biophys Acta. 1999; 1419:173–185. [PubMed: 10407069]

31. Chloupková M, et al. Expression of 25 human ABC transporters in the yeast Pichia pastoris and characterization of the purified ABCC3 ATPase activity. Biochemistry. 2007; 46:7992–8003. [PubMed: 17569508]

32. Zhang DW, Graf GA, Gerard RD, Cohen JC, Hobbs HH. Functional asymmetry of nucleotide-binding domains in ABCG5 and ABCG8. J Biol Chem. 2006; 281:4507–4516. [PubMed: 16352607]

33. Otwinowski Z, Minor W. Processing of X-ray diffraction data. Methods Enzymol. 1997; 276:307–326.

34. Minor W, Cymborowski M, Otwinowski Z, Chruszcz M. HKL-3000: the integration of data reduction and structure solution – from diffraction images to an initial model in minutes. Acta Crystallogr D. 2006; 62:859–866. [PubMed: 16855301]

35. Otwinowski Z, Borek D, Majewski W, Minor W. Multiparametric scaling of diffraction intensities. Acta Crystallogr A. 2003; 59:228–234. [PubMed: 12714773]

36. Borek D, Minor W, Otwinowski Z. Measurement errors and their consequences in protein crystallography. Acta Crystallogr D. 2003; 59:2031–2038. [PubMed: 14573959]

37. Borek D, Ginell SL, Cymborowski M, Minor W, Otwinowski Z. The many faces of radiation-induced changes. J Synchrotron Radiat. 2007; 14:24–33. [PubMed: 17211069]

38. Borek D, Cymborowski M, Machius M, Minor W, Otwinowski Z. Diffraction data analysis in the presence of radiation damage. Acta Crystallogr D. 2010; 66:426–436. [PubMed: 20382996]

39. Borek D, Dauter Z, Otwinowski Z. Identification of patterns in diffraction intensities affected by radiation exposure. J Synchrotron Radiat. 2013; 20:37–48. [PubMed: 23254654]

40. Sheldrick GM. Experimental phasing with SHELXC/D/E: combining chain tracing with density modification. Acta Crystallogr D. 2010; 66:479–485. [PubMed: 20383001]

41. Otwinowski, Z. Maximum likelihood refinement of heavy atom parameters. In: Evans, PR.; Wolf, W.; Leslie, AGW., editors. Proc. CCP4 Study Weekend: Isomorphous Replacement and Anomalous Scattering. Daresbury Laboratory: Science and Engineering Research Council; 1991. p. 80-86.

42. Winn MD, et al. Overview of the CCP4 suite and current developments. Acta Crystallogr D. 2011; 67:235–242. [PubMed: 21460441]

43. Cowtan K, Main P. Miscellaneous algorithms for density modification. Acta Crystallogr D. 1998; 54:487–493. [PubMed: 9761844]

44. Zhang KY, Cowtan K, Main P. Combining constraints for electron-density modification. Methods Enzymol. 1997; 277:53–64. [PubMed: 18488305]

45. Emsley P, Lohkamp B, Scott WG, Cowtan K. Features and development of Coot. Acta Crystallogr D. 2010; 66:486–501. [PubMed: 20383002]

46. Cowtan K. The Buccaneer software for automated model building. 1. Tracing protein chains. Acta Crystallogr D. 2006; 62:1002–1011. [PubMed: 16929101]

47. Murshudov GN, Vagin AA, Dodson EJ. Refinement of macromolecular structures by the maximum-likelihood method. Acta Crystallogr D. 1997; 53:240–255. [PubMed: 15299926]

48. Nicholls RA, Fischer M, McNicholas S, Murshudov GN. Conformation-independent structural comparison of macromolecules with ProSMART. Acta Crystallogr D. 2014; 70:2487–2499. [PubMed: 25195761]

49. Davis IW, et al. MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. Nucleic Acids Res. 2007; 35:W375–W383. [PubMed: 17452350]

50. Chivian D, et al. Prediction of CASP6 structures using automated Robetta protocols. Proteins. 2005; 61(Suppl 7):157–166. [PubMed: 16187358]

51. Stone DK, Xie XS, Racker E. Inhibition of clathrin-coated vesicle acidification by duramycin. J Biol Chem. 1984; 259:2701–2703. [PubMed: 6142047]

52. Wang J, et al. Relative roles of ABCG5/ABCG8 in liver and intestine. J Lipid Res. 2015; 56:319–330. [PubMed: 25378657]

53. Kamisetty H, Ovchinnikov S, Baker D. Assessing the utility of coevolution-based residue-residue contact predictions in a sequence- and structure-rich era. Proc Natl Acad Sci USA. 2013; 110:15674–15679. [PubMed: 24009338]

54. Ovchinnikov S, et al. Large-scale determination of previously unsolved protein structures using evolutionary information. eLife. 2015; 4:e09248. [PubMed: 26335199]

55. Vehlow C, et al. CMView: interactive contact map visualization and analysis. Bioinformatics. 2011; 27:1573–1574. [PubMed: 21471016]

56. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. Comparison of simple potential functions for simulating liquid water. J Chem Phys. 1983; 79:926–935.

57. Bayly CI, Cieplak P, Cornell W, Kollman PA. A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: the RESP model. J Phys Chem. 1993; 97:10269–10280.

58. Case DA, et al. The Amber biomolecular simulation programs. J Comput Chem. 2005; 26:1668–1688. [PubMed: 16200636]

59. Wang J, Wang W, Kollman PA, Case DA. Automatic atom type and bond type perception in molecular mechanical calculations. J Mol Graph Model. 2006; 25:247–260. [PubMed: 16458552]

60. Wickstrom L, Okur A, Simmerling C. Evaluating the performance of the ff99SB force field based on NMR scalar coupling data. Biophys J. 2009; 97:853–856. [PubMed: 19651043]

61. Dickson CJ, et al. Lipid14: the amber lipid force field. J Chem Theory Comput. 2014; 10:865–879. [PubMed: 24803855]

62. Sagui C, Pedersen LG, Darden TA. Towards an accurate representation of electrostatics in classical force fields: efficient implementation of multipolar interactions in biomolecular simulations. J Chem Phys. 2004; 120:73–87. [PubMed: 15267263]

63. Ryckaert JP, Ciccotti G, Berendsen H. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. J Comput Phys. 1977; 23:321–341.

64. Uberuaga BP, Anghel M, Voter AF. Synchronization of trajectories in canonical molecular-dynamics simulations: observation, explanation, and exploitation. J Chem Phys. 2004; 120:6363–6374. [PubMed: 15267525]

65. Bahar I, Lezon TR, Bakan A, Shrivastava IH. Normal mode analysis of biomolecular structures: functional mechanisms of membrane proteins. Chem Rev. 2010; 110:1463–1497. [PubMed: 19785456]

66. Pei J, Grishin NV. PROMALS3D: multiple protein sequence alignment enhanced with evolutionary and three-dimensional structural information. Methods Mol Biol. 2014; 1079:263–271. [PubMed: 24170408]

67. Katoh K, Standley DM. MAFFT: iterative refinement and additional methods. Methods Mol Biol. 2014; 1079:131–146. [PubMed: 24170399]

68. Pei J, Grishin NV. AL2CO: calculation of positional conservation in a protein sequence alignment. Bioinformatics. 2001; 17:700–712. [PubMed: 11524371]

69. Lazarevic V, Karamata D. The tagGH operon of Bacillus subtilis 168 encodes a two-component ABC transporter involved in the metabolism of two wall teichoic acids. Mol Microbiol. 1995; 16:345–355. [PubMed: 7565096]

70. Evans IJ, Downie JA. The nodI gene product of Rhizobium leguminosarum is closely related to ATP-binding bacterial transport proteins; nucleotide sequence analysis of the nodI and nodJ genes. Gene. 1986; 43:95–101. [PubMed: 3019841]

71. Goldman BS, Beck DL, Monika EM, Kranz RG. Transmembrane heme delivery systems. Proc Natl Acad Sci USA. 1998; 95:5003–5008. [PubMed: 9560218]

72. Ajdi D, et al. Genome sequence of Streptococcus mutans UA159, a cariogenic dental pathogen. Proc Natl Acad Sci USA. 2002; 99:14434–14439. [PubMed: 12397186]

73. Han CS, et al. Pathogenomic sequence analysis of Bacillus cereus and Bacillus thuringiensis isolates closely related to Bacillus anthracis. J Bacteriol. 2006; 188:3382–3390. [PubMed: 16621833]

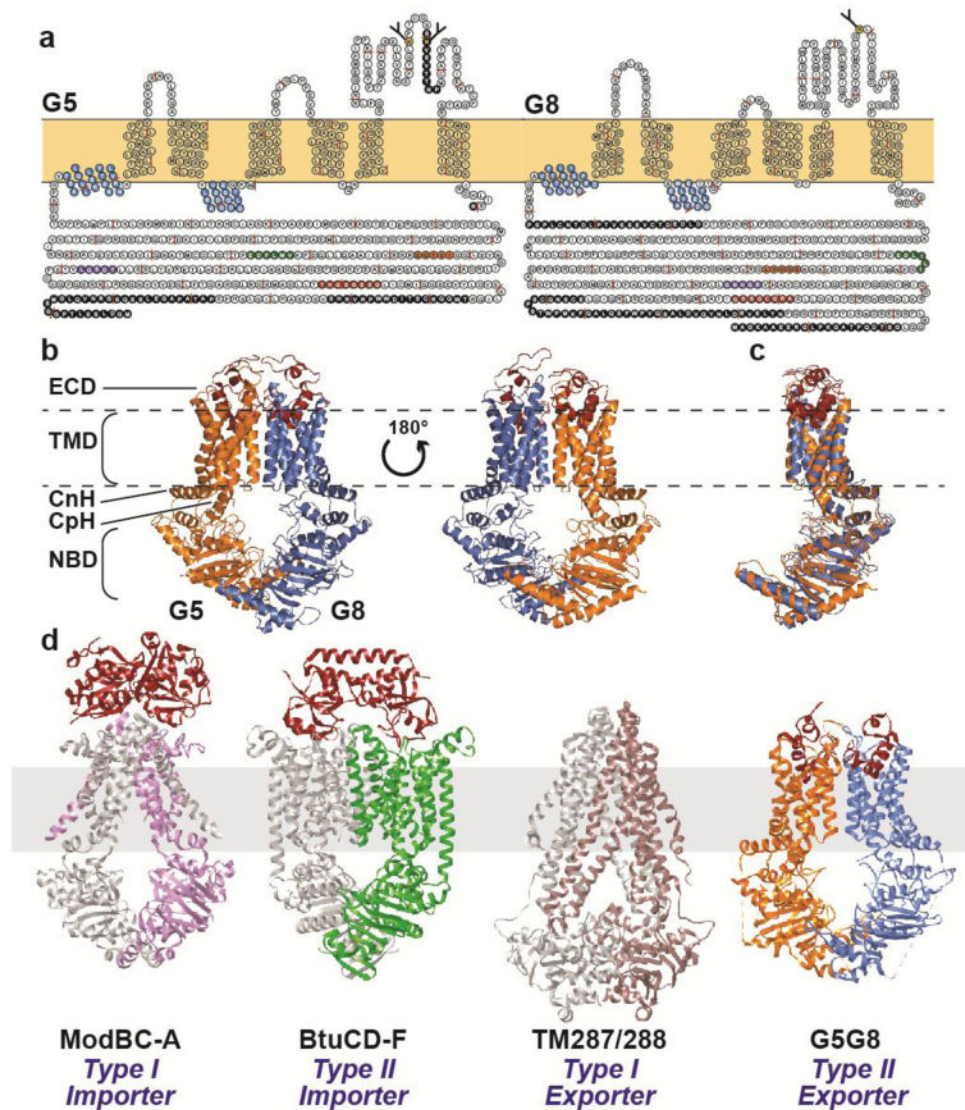**Figure 1. Structure of the G5G8 heterodimer and comparison with other ABC transporters**
a, Topological illustrations of G5 and G8: black filled circles for unresolved residues, blue filled circles for the connecting (CnH) and coupling helices (CpH), red filled circles for Walker A motif, green filled circles for Walker B motif, orange filled circles for ABC Signature motif, purple filled circles for the Q-loop. N-linked glycans indicated as branches from asparagine residues. **b**, Cartoon of the G5G8 heterodimer with G5 in orange, G8 in blue, and the ECDs in red. **c**, Superposition of Cα backbones of G5 and G8 (468 out of 571 (G5)/579 (G8) residues). **d**, Nucleotide-free structures of ABC transporter superfamilies: ModBC-A (PDB accession number 2ONK), BtuCD-F (PDB accession number 2QI9), TM287(grey)/TM288(pink) (PDB accession number 4Q4H), and G5G8. Substrate binding proteins of BtuF and ModA are coloured red.
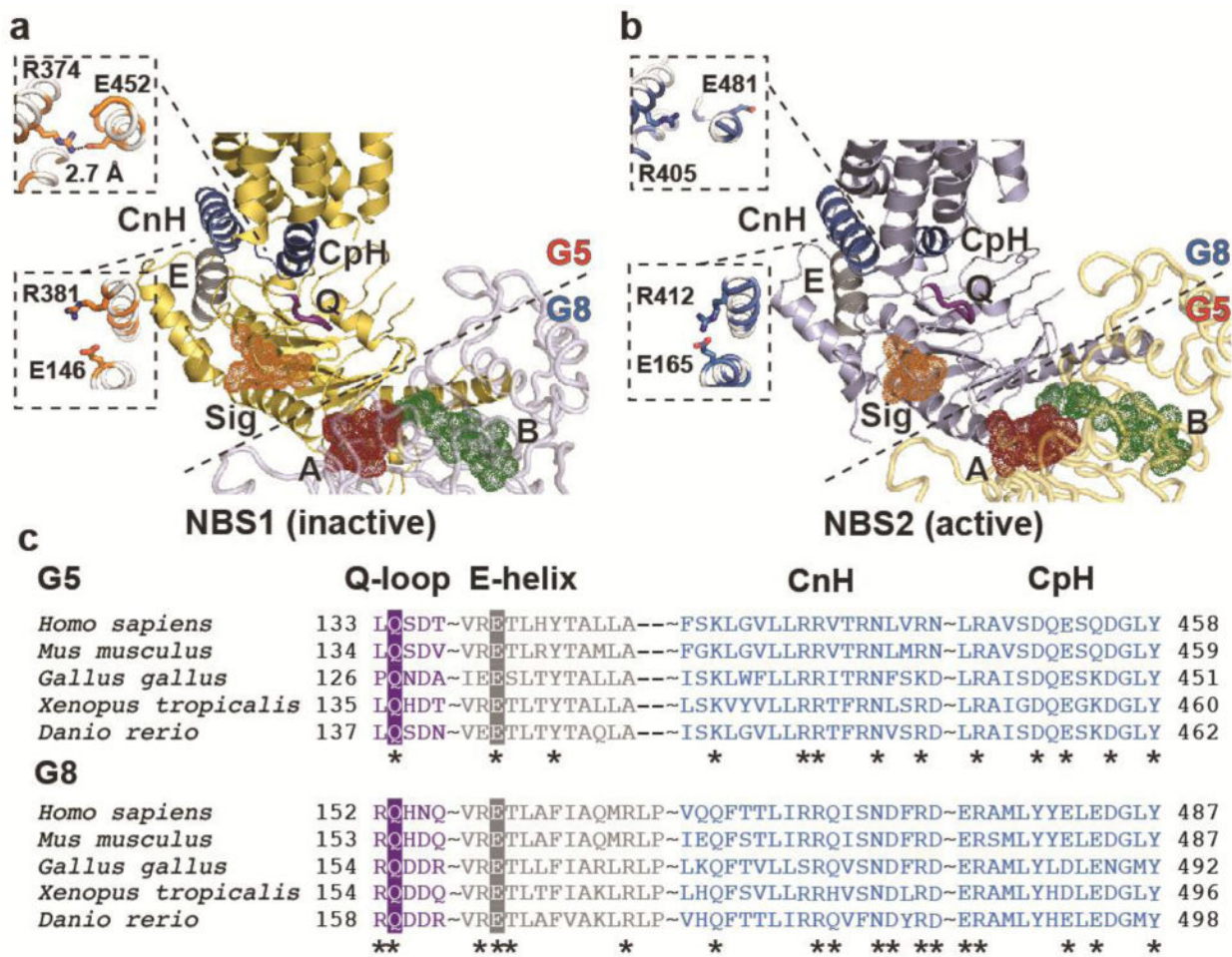
Figure 2. Interface between NBDs and TMDs of G5 and G8
a, Interface between TMD of G5 and inactive NBS1: connecting (CnH) and coupling (CpH) helices (blue), ABC Signature motifs (orange), Q-loops (purple) and E-helices (grey), Walker A motifs (red) and Walker B motifs (green) of G8. Insets show asymmetric interactions between CnH, CpH, and E-helix. **b**, Interface between TMD of G8 and active NBS2. **c**, Sequence alignments of Q-loop, E-helix, CnH, and CpH: the conserved glutamine of the Q-loop and glutamate in the E-helix are highlighted. Asterisks indicate highly conserved polar residues.
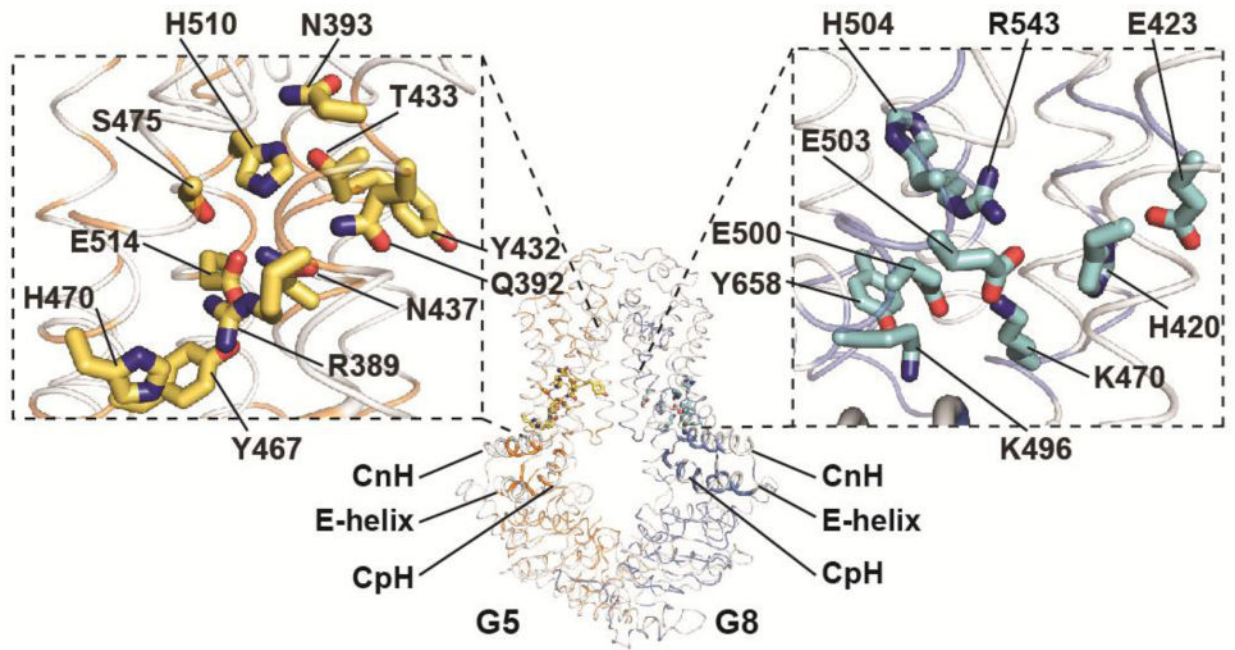
Figure 3. TMD polar relay of G5 and G8

Stick representation of the conserved polar residues that form the TMD polar relay proximal to CpH and CnH (left: G5; right: G8).
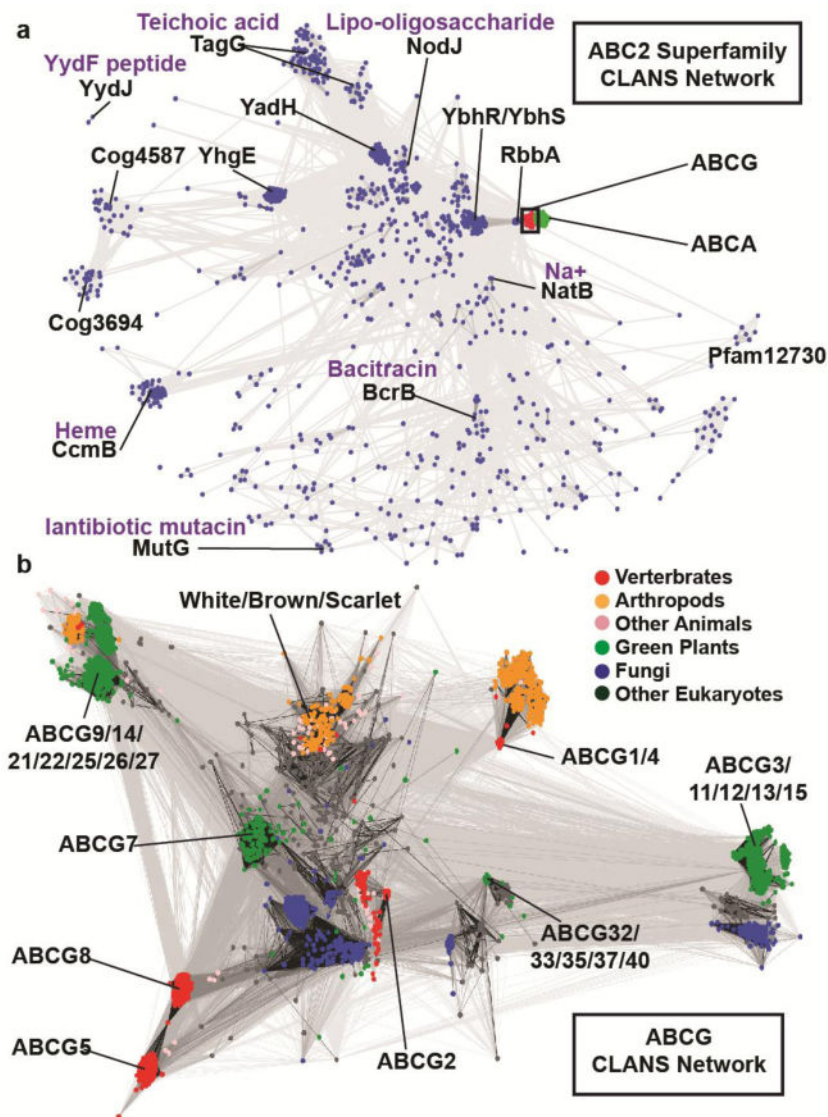
Figure 4. CLANS network of the ABC2 exporter superfamily and ABCG transporters
a, Sequence analysis of the TMD of ABC2 exporter superfamily transporters from both eukaryotes and prokaryotes is shown in a two-dimensional graph output. Lines indicate the similarities between protein sequences with an E value $<10^{-10}$. Export substrates by prokaryotic transporters in the ABC2 exporter superfamily include polysaccharide teichoic acids, which are found within cell walls of Gram-positive bacteria (TagG); lipo-oligosaccharides, which are located in outer membranes of Gram-negative bacteria (NodJ); haem, which is required for cytochrome C biogenesis (CcmB); and bacterial-killing lantibiotic peptides such as mutacin (MutG) and bacitracin (BcrB). See Methods for more information. **b**, CLANS[26] analysis of the TMDs from 4,400 ABCG subfamily members (filled circles) shows functionally diverse clusters.