Published in final edited form as:

Nature. 2022 January ; 601(7892): 274–279. doi:10.1038/s41586-021-04211-w.

Structure and mechanism of the SGLT family of glucose transporters

Lei Han[1,*], **Qianhui Qu**[1,2,†,*], **Deniz Aydin**[1,2,3,4,*], **Ouliana Panova**[1,2], **Michael J. Robertson**[1,2], **Yan Xu**[1], **Ron O. Dror**[1,2,3,4], **Georgios Skiniotis**[1,2,#], **Liang Feng**[1,2,#]

[1]Department of Molecular and Cellular Physiology, Stanford University School of Medicine, Stanford, CA 94305, USA

[2]Department of Structural Biology, Stanford University School of Medicine, Stanford, CA 94305, USA

[3]Department of Computer Science, Stanford University, Stanford, CA 94305, USA

[4]Institute for Computational and Mathematical Engineering, Stanford University, Stanford, CA 94305, USA

Abstract

Glucose is a primary energy source in living cells. The discovery in 1960s that a sodium gradient powers the active uptake of glucose in the intestine[1] heralded the concept of a secondary active transporter that can catalyze the uphill movement of a substrate by harnessing energy from another coupled substrate. Subsequently, coupled Na^+/glucose transport was found to be mediated by sodium-glucose cotransporters (SGLTs)[2,3]. SGLTs are responsible for active glucose and galactose absorption in the intestine and for glucose reabsorption in the kidney[4], and are targeted by multiple drugs to treat diabetes[5]. Intriguingly, several members within the SGLT family transport key metabolites other than glucose[2]. Here, we report near-atomic resolution structures of the prototypic sodium-glucose cotransporter human SGLT1 and a related monocarboxylate transporter SMCT1 within the same family. The structures, together with molecular dynamics simulations and functional studies, define the architecture of SGLTs, uncover the mechanism of substrate binding and selectivity, and shed light on water permeability of SGLT1. These results provide insights into the multifaceted functions of SGLTs.

[#]**Correspondence and requests for materials** should addressed to: Georgios Skiniotis (yiorgo@stanford.edu); Liang Feng (liangf@stanford.edu).

Author contributions. L.H. performed molecular biology, protein expression and purification, biochemistry and functional studies. Q.Q. prepared cryo-EM grids, collected and processed data, and reconstructed the maps of $SGLT1_{con}$, $SGLT1_{conHA}$ and SMCT1. L.H. built the model with the input from Q.Q. D.A. performed and analyzed MD simulations and contributed to preparation of the figures and manuscript. M.J.R. assisted with model building. Y.X. performed biochemistry and protein engineering. O.P. prepared cryo-EM grids, collected and processed data, and reconstructed the map of $SGLT1_{conHA}$. R.O.D. oversaw the MD simulations and contributed to the manuscript preparation. L.H., Q.Q., G.S. and L.F. wrote the manuscript with the input from all co-authors. G.S. and L.F. supervised the project.

[†]Present address: Shanghai Stomatological Hospital, Institutes of Biomedical Science, Department of Systems Biology for Medicine, Fudan University, Shanghai, China

[*]These authors contributed equally.

Competing interests: The authors declare no competing interests.

SGLT1, the first cloned transporter within the large sodium-solute symporter (SSS) family[2], is a high-affinity, low-capacity glucose transporter with an apparent 2 Na^+:1 sugar coupling stoichiometry[2,6]. SGLT1 mediates active glucose and galactose absorption in the intestine as well as renal glucose scavenging [4]. Mutations in SGLT1 cause glucose-galactose malabsorption (GGM) syndrome[7]. Due to its fundamental physiological significance, SGLT1 has been the focus of many functional and biophysical studies that have shaped our understanding of active transporters[2]. Moreover, SGLT1 underlies oral rehydration therapy to treat secretory diarrhea[2,8] and, together with SGLT2, is an important drug target to treat diabetes[5]. In addition to its Na^+/glucose transport function, SGLT1 can transport water and urea through a channel-like activity that is important for passive water transport in the small intestine[9,10].

Human SGLT family transporters can be phylogenetically grouped into two main branches: one that transports sugars and myo-inositol (e.g., SGLT1) and one that translocates key metabolites such as monocarboxylates, iodide, and biotin (e.g., SMCT1). It remains highly intriguing how the same structural scaffold in this family gives rise to vastly different substrate selectivity.

Substantial advances in our understanding of SSS transporters came from structural determination of two bacterial homologs, the sodium/galactose transporter vSGLT[11], and the sodium/sialic acid transporter SiaT[12]. Nonetheless, the modest sequence identity and lack of apparent inhibition by the classic SGLT inhibitor phlorizin[13] make these structures limited models for human SGLTs (Extended Data Fig. 1). Hence, high resolution structures of SGLTs are essential for understanding their transport mechanisms, physiological functions, pharmacology, and pathophysiology.

Here, we report cryo-electron microscopy (cryo-EM) structures of apo human SGLT1 and substrate-bound human SMCT1 (Extended Data Fig. 1, 2). Combined with molecular dynamics (MD) simulations and functional studies, our work provides a framework to understand key aspects of the eukaryotic SSS family.

Structural determination

Because wild-type (WT) SGLT1 is unstable and aggregates when extracted in detergent, we used the consensus-mutation approach[14] to engineer a thermostable variant, $SGLT1_{con}$ (>90% sequence identity with WT; Extended Data Fig. 1). $SGLT1_{con}$ showed clear Na^+-dependent and phlorizin-sensitive α-methyl-D-glucopyranoside (αMDG, a classic glucose congener) uptake in a cell-based assay when we reversed two point mutations near the C-terminus ($SGLT1_{conHA}$, Extended Data Fig. 3a). Despite reduced activity, the apparent substrate affinity is similar to WT (Extended Data Fig. 3e). We did not observe reliable activity above background for $SGLT1_{con}$. To facilitate cryo-EM reconstructions on SGLT1 (~74 kDa, lacking large soluble domains), we generated high affinity nanobodies to $SGLT1_{con}$ (Extended Data Fig. 4a). Although decorating $SGLT1_{con}$ with a ~14 kDa nanobody provides only a modest marker, we successfully determined a cryo-EM reconstruction of apo $SGLT1_{con}$ at 3.4 Å resolution and subsequently apo $SGLT1_{conHA}$ at 3.15 Å resolution (Fig. 1a, Extended Data Fig. 4, 5, and Extended Data Table 1). These two

structures overlaid well onto each other (Extended Data Fig. 4f), and we thus mainly focused our structural analysis on the higher-resolution $SGLT1_{conHA}$ structure.

In parallel, we determined the structure of SMCT1 (~66 kDa) in complex with substrate at 3.5 Å resolution using a similar nanobody-assisted strategy (Fig. 3a, Extended Data Fig. 5, 6, and Extended Data Table 1). These results highlight that nanobodies can effectively facilitate cryo-EM studies on relatively small membrane proteins, providing an important toolbox beyond conformation-stabilizing agents.

SGLT1 overall structure

SGLT1 contains 14 transmembrane helices (TM0-TM13) and adopts an APC-fold, with TM1-5 and TM6-10 forming an inverted repeat structure (Fig. 1b). The overall structure of SGLT1 resembles that of prokaryotic vSGLT, apart from the extracellular portion (~20% of structure). An elongated map density located in a furrow formed by TM1a, TM7, and TM13, matches well a cholesteryl hemisuccinate (CHS) packed closely to $Trp67^{TM1}$ (Fig. 1c, d, and Extended Data Fig. 7a). This location is compatible with a lipid in the membrane inner leaflet.

To validate the cholesterol/CHS binding site, we performed all-atom MD simulations of $SGLT1_{con}$, with cholesterol initially placed at this site. In three of five independent simulations (3 μs each), cholesterol remained stably bound at this site, and in a fourth, transiently moved away but then returned to its initial binding pose (Extended Data Fig. 7e). The presence of cholesterol is consistent with the cholesterol-dependent localization of SGLT1 to lipid rafts[15,16]. Mutating $Trp67^{TM1}$, which makes prominent contacts with the CHS, significantly impaired transport and CHS's thermostabilizing effect (Extended Data Fig. 3c, d). Mutating $Trp641^{TM13}$, which tangentially interacts with CHS, modestly affected activity and only slightly impacted the thermostabilizing effect (Extended Data Fig. 3c, d). These results suggest that cholesterol might be involved in regulating transport, conceivably through stabilizing certain SGLT1 conformations. This suggests an interesting parallel to the dopamine transporter, in which a cholesterol binds to an adjacent region and is thought to modulate transport[17] (Extended Data Fig. 7b).

Compared with prokaryotic SSS homologs, SGLT1's extracellular lid is unique. It is mainly composed of three long and ordered extracellular loops—EL3 (TM5-TM6), EL4 (TM7-TM8), and EL6 (TM11-TM12)—with several short α-helices (Fig. 1b, and Extended Data Fig. 7c). EL3 and EL4 interact extensively, forming the main body. The lid latches to the transmembrane region through a known disulfide bridge[18] between Cys255 (EL3) and Cys511 (EL6) (Extended Data Fig. 7c, d). All other extracellular cysteine residues are positioned to form three pairs of intra-loop disulfide bonds (Extended Data Fig. 7c, d), stabilizing the extended loops and thus the lid domain. Notably, C255W and C355S mutations have been linked to GGM syndrome[19], highlighting an important role of disulfide bridges in SGLT1 folding and/or function. The lid domain covers a significant portion of the extracellular surface and leaves a cavity-like opening, which connects to the extracellular gate. The electrostatic surface of the cavity is negative (Fig. 2f), and its relevance to Na^+ transport needs to be further investigated.

SGLT1 substrate translocation pathway

In the SGLT1 structure, a relatively narrow solvent-accessible vestibule extends from the cytosolic side toward the substrate-binding pocket, suggesting a partial inward-open conformation (Fig. 1c and Extended Data Fig. 9d). TM10's N-terminal end and its preceding loop pack against TM2 and TM1, forming the extracellular gate that closes by hydrophobic interactions of Phe453^{TM10}, L452^{TM10}, Val98^{TM2}, Phe101^{TM2}, and Gly86^{TM1} (Fig. 2b). Additionally, Asn363 on EL4 interacts with the backbone oxygen of Gly450 right before TM10's N-terminal end, contributing to extracellular-gate closure.

The vestibule is surrounded by TM1, TM3, TM5, TM6, and TM8, constituting a substrate-translocation pathway, and lined by highly conserved residues (Extended Data Fig. 7f). The vestibule is hydrophilic, which might facilitate glucose and Na^+ passage. The end of the vestibule contains the presumed glucose-binding pocket, which superimposes well onto that of vSGLT (Fig. 2d), indicating potentially similar substrate-binding poses.

SGLT1 couples the transport of two Na^+ with one glucose. The Na^+-binding site Na2, conserved among many APC superfamily transporters, is located near the helical break of TM1 on the opposite side of the glucose-binding site. Based on sequence and structural alignment with vSGLT, the backbone carbonyls of Ala76^{TM1}, Ile79^{TM1}, and Ser389Ala^{TM8} and side-chain oxygens of Ser392^{TM8} and Ser393^{TM8} would constitute Na2 in SGLT1 (Extended Data Fig. 8b). Previous studies showed that substituting Ser392^{TM8} or Ser393^{TM8} significantly impairs glucose uptake[20]. In the current SGLT1 conformation, tilting the C-terminus of TM8 loosened Na2 (Extended Data Fig. 8a). A clue about the other Na^+-binding site came from SiaT, which also shows a 2 Na^+:1 substrate stoichiometry. The Na3 site of SiaT[12] is conserved in SGLT1 based on sequence[12] and structural alignment. Na3 of SGLT1, located further from the glucose-binding pocket and closer to the cytosolic side, is composed of side-chain oxygens of Asp204^{TM5}, Ser396^{TM8}, Thr395^{TM8}, and Ser77^{TM1} and the backbone carbonyl of Ser392^{TM8} (Extended Data Fig. 8a, b). Mutating these residues individually decreases or abolishes glucose uptake[21] (Extended Data Fig. 8c). Interestingly, in SGLT2, Thr395^{TM8} is substituted by an alanine, which would cripple Na3; this might account for its 1 Na^+:1 glucose stoichiometry. In the current SGLT1 conformation, the TM5 shift loosened Na3 (Extended Data Fig. 8a).

To gain further insight into glucose coordination, we performed MD simulations of $SGLT1_{con}$ with glucose and two Na^+ ions placed into the proposed binding sites. In each of five simulations (2 μs each), SGLT1 maintained a stable structure (Extended Data Fig. 10a), and glucose remained within the substrate-binding pocket (Extended Data Fig. 10b, c). Glucose remained close to its initially modelled pose for at least several hundred nanoseconds, and frequently returned to this pose after repositioning. Glucose forms hydrogen bonds with Gln457^{TM10} and Glu102^{TM2}, while its pyranose ring stacks between Phe101^{TM2} and Tyr290^{TM6} (Fig. 2a and Extended Data Fig. 10d, e). Glucose also interacts frequently with Trp291^{TM6} and intermittently with Asn78^{TM1}, Thr460Ser^{TM10}, and Lys321^{TM7}. In the pocket, Lys321^{TM7} forms electrostatic interactions with Glu102^{TM2}, which may help position or stabilize the Glu102^{TM2} sidechain for interacting with glucose. In addition, His83^{TM1}, which helps maintain the TM1 helical break conformation, stabilizes

the Lys321^{TM7}– Glu102^{TM2} interaction. In simulations initiated without sodium in the binding pocket, glucose is generally more mobile (Extended Data Fig. 10g). The most populated pose is also close to the initially modelled pose, but glucose occasionally moves towards Lys157^{TM3}. These observations hint that sodium may stabilize glucose binding, as suggested previously[21]. The residues important for coordinating the substrate are highly conserved. When they are mutated, substrate affinity and transport activity are significantly impaired, as shown in published studies[20] and this study (Fig. 2c). Notably, Gln457 mutation is associated with GGM syndrome[19], highlighting its critical role in transport.

SGLTs and their homologs show various selectivities across closely related monosaccharides. Compared to vSGLT's selective transport of galactose, SGLT1 transports glucose and galactose nearly equally. Three residues differ between SGLT1 and vSGLT's substrate-binding sites: H83 (Q69 in vSGLT1), T287A (N260), and A105 (S91) (Fig. 2d). Of these, only Q69 and N260 in vSGLT are important for galactose transport[11]. To probe the role of these two positions in substrate selectivity, we replaced H83 or T287 of SGLT1 with their equivalent residue in vSGLT. H83Q mutation abolished SGLT1's αMDG and galactose uptake (Extended Data Fig. 7h), consistent with its important role in function[20]. Intriguingly, T287N mutation abolished αMDG uptake but retained galactose uptake (~15% of WT) (Extended Data Fig. 7h). In competition assays with WT SGLT1, we found that αMDG can effectively compete [^{14}C]-galactose uptake with an IC_{50} comparable to that of galactose (Fig. 2e). In contrast, for the T278N mutant, αMDG competes poorly against [^{14}C]-galactose for uptake while galactose competes effectively (Fig. 2e). Our results indicate that T287 position plays an important role in substrate selectivity and asparagine substitution confers the galactose selectivity to SGLT1. Notably, glucose and galactose differ only by the orientation of their 4th hydroxyl group. T287A of SGLT1 is the closest residue to this group of glucose based on the stable pose in our MD simulations (Fig. 2a); its equivalent residue in vSGLT (N260) interacts with galactose's 4th hydroxyl group. It is conceivable that the larger sidechain introduced by the T287N substitution in SGLT1 may sterically hinder the horizontal 4th hydroxyl group of glucose, with the upward orientation of this group of galactose alleviating such hindrance. In line with this, the T287-equivalent position is occupied by small amino acids (alanine, threonine, serine) in mammalian SGLT1 (Extended Data Fig. 1), and substituting alanine has no significant effect on glucose transport in SGLT1 (Extended Data Fig. 7g). These functional results further support our computational observations on substrate binding in the pocket.

SGLT1 water-permeation pore

In MD simulations of glucose-free and glucose-bound $SGLT1_{con}$, we observed water molecules traversing the cell membrane by flowing through the transporter. SGLT1's water-permeation pore runs through the substrate-translocation vestibule (Fig. 2h). The pore narrows above the substrate-binding pocket, with the constriction formed by residues on the N-terminal end of TM10 and on TM1, with contribution from EL4. In simulations, the N-terminal part of TM10, together with TM9, moves slightly away from TM1, loosening the extracellular gate (Fig. 2g) and opening a narrow pathway connecting to the extracellular solution. Concomitantly, the intracellular part of the substrate-translocation pathway narrows. We observed water flow through a similar route under glucose-free and

glucose-bound conditions (Extended Data Fig. 10f). To probe the water-permeation pore observed in simulation, we performed mutagenesis studies on the permeability of urea, a surrogate tracer for water permeation in SGLT1[9]. Mutating residues lining the narrow restriction, including L452, N363, and Q451, reduced permeation (Extended Data Fig. 10h, i). In contrast, mutating control residue D454, which faces away from the pore, showed no apparent reduction in permeation. Previous studies showed that cysteine substitutions for F453 or Q457 (which line the pore) increase permeation[9], consistent with important roles in the permeation pathway observed in simulation.

SMCT1 structure and substrate binding

To understand the diversity and common themes between SGLT family's two main branches, we also investigated SMCT1. SMCT1 mediates intestinal and renal uptake of physiologically important monocarboxylates (lactate, pyruvate, butyrate[22,23]) and monocarboxylate drugs (nicotinate, salicylates, benzoate, γ-hydroxybutyrate[24,25]), with a 2 Na^+:1 monocarboxylate stoichiometry[22,25]. SMCT1's transport activity also underlies its tumor-suppressor function in cancer[26,27].

Our structure captured SMCT1 in an inward-facing conformation (Extended Data Fig. 9d). Overall, its transmembrane domain adopts a similar architecture as SGLT1 (RMSD=2.5 Å), but lacks TM13 (Fig. 3b). Notably, SMCT1's extracellular domain shows significant differences in shape and orientation (Extended Data Fig. 9g), with a relatively short EL3, a long EL6, and only one disulfide bridge.

The map shows clear extra density within a central pocket near the end of the vestibule at a typical substrate-binding region in SSS transporters. The shape of the density matches well with butyrate, which was present during expression at high concentration (10 mM) and has the highest affinity among known substrates[27]. Therefore, we tentatively ascribe the density to butyrate, whose resulting pose fits well with its local environment (Fig. 3d). In the pocket, butyrate's alkyl tail points towards the extracellular side and carboxyl group towards the intracellular side (Fig. 3c). The tail mostly forms hydrophobic interactions with $Phe65^{TM1}$, $Try142^{TM3}$, $Trp253^{TM6}$, and $Phe417^{TM10}$ (Fig. 3c). Butyrate's carboxyl group forms hydrogen bonds with $Gln263^{TM6}$ (Extended Data Fig. 8d). Overall, the amphiphilic composition of the binding pocket matches well with chemical properties of monocarboxylate substrates (Fig. 3c). $Ser64^{TM1}$, $Ser67^{TM1}$, $Thr70^{TM1}$, $Ile256^{TM6}$, and $Try257^{TM6}$ are also within range to interact with butyrate. Substituting individual substrate-binding pocket residues with alanine substantially impaired monocarboxylate transport (Extended Data Fig. 6a, b, 8e). In contrast, mutating $Gln263^{TM6}$ to threonine, as found in SMCT2, retained most transport activity, indicating the importance of the hydroxyl group in interacting with monocarboxylate.

Structural comparison and sequence alignment suggest two conserved sodium-binding sites of SMCT1, similar to SiaT[12] (Extended Data Fig. 2, 8a, b). Na2 is composed of side-chain oxygens of $Ser351^{TM8}$ and $Thr352^{TM8}$ and backbone carbonyls of $Ala63^{TM1}$, $Met66^{TM1}$, and $Gly348^{TM8}$. Na3, located closer to the cytosol, is composed of side-chain oxygens of $Asp189^{TM5}$, $Ser354^{TM8}$, $Ser355^{TM8}$, and $Ser64^{TM1}$ and the backbone carbonyl

of Ser351^{TM8}. These Na^+-binding sites likely account for SMCT1's 2 Na^+:1 substrate stoichiometry[22,25]. Interestingly, Ser354^{TM8} is substituted by alanine in SMCT2, consistent with the loss of Na3 and its 1 Na^+:1 substrate stoichiometry. In the current inward-facing structure, Na2 and Na3 are distorted due to the outward movement of TM5 and TM8 at the cytosol side (Extended Data Fig. 8a), consistent with the typically unbound Na^+ in inward-open conformations of other APC transporters.

Comparison of SGLT1 and SMCT1

The substrate-translocating vestibule takes a comparable route in SMCT1 and SGLT1. Overall, substrate-binding pockets are located near the end of the vestibule (Extended Data Fig. 9a). However, the shape, location, and chemical environment of the binding pockets differ substantially. Glucose-coordinating residues in SGLT1 are located on TM1, TM2, TM6, TM7, and TM10, while butyrate-binding residues in SMCT1 are on TM1, TM3, TM6, and TM10 (Extended Data Fig. 9a). Butyrate locates towards TM3, while glucose locates towards TM2. Compared with the butyrate-binding site in SMCT1, the glucose-binding site in SGLT1 is located further into the vestibule.

In both SMCT1 and SGLT1, the substrate does not fully occupy the central pocket (Extended Data Fig. 9b). In SMCT1, the unoccupied region is highly hydrophobic, composed of Val69^{TM1}, Phe85^{TM2}, Phe88^{TM2}, Try92^{TM3}, Leu410^{TM10}, and Leu414^{TM10}, which can accommodate an alkyl chain with ~3-4 additional carbons, consistent with SMCT1 transporting monocarboxylates up to octanoate[22,23]. This provides a plausible explanation for why monocarboxylates with longer alkyl chains, such as ibuprofen, function as competitive inhibitors[24]; the steric hindrance of a long alkyl chain might impede extracellular gate closure. In MD simulations of SGLT1, the stable pose of glucose occupies one side of the binding pocket, with extra space next to the sugar's O1. This configuration is consistent with studies showing that various groups added to O1 make effective substrates[2,28].

Notably, compared to SMCT1, SGLT1 has an extra peripheral transmembrane helix, TM13, that directly contacts TM2 and TM7 and participates in cholesterol binding (Fig. 1d, Extended Data Fig. 9e). Mutating His660-Ala661 near the end of TM13 impairs transport (Extended Data Fig. 3b). TM13 might indirectly regulate transport through its interaction with cholesterol or with TM2 and TM7, both of which are involved in substrate-binding-pocket formation.

Discussion

Our studies reveal a conserved SGLT1 transport pathway that allows for multiple functions. In MD simulations, we observed a channel-like permeation pathway for water. The pathway coincides with the glucose translocation route when either gate is open. Consistent with this possibility, previous MD simulations of vSGLT suggested that water might translocate via the same path as sugar[9,29]. Nonetheless, the water channel activity is mechanistically distinct from active glucose transport. This is evident from mutations that selectively increase water channel activity while abolishing active glucose transport[9,20]. In simulations, SGLT1

undergoes slight conformational changes from the inward-facing conformation to open the water-permeation pathway, consistent with a trajectory *en route* to the outward-facing conformation, albeit to a much smaller extent. Perhaps, the extracellular-gate residues allow subtle rearrangement and relaxation, giving rise to slight conformational transitions. Passive water permeation has been functionally observed in various transporters[10,30]. A hydrophilic substrate-translocation pathway and relatively thin gate, as found in SGLT1, might set the stage for a transporter to gain water channel activity. In this scheme, the gate would only need to leave a proper opening for water to pass without leaking substrate. The water channel activity within the transporter, which is linked to a particular gate, would likely depend on the conformational state. Such dual functionality of the transport pathway blurs the boundary between transporters and channels. It remains to be investigated whether the proposed mechanism underlies the function of other transporters.

Studies on vSGLT and SiaT provided important insights into an alternating-access model of prokaryotic SSS transporters[12]. In our study, SGLT1 and SMCT1 were captured in an inward-facing conformation similar to that of vSGLT (C_α RMSD: 1.9 Å, SGLT1; 2.5 Å, SMCT1) (Extended Data Fig. 9f). Structural comparison with SiaT in an outward-open conformation provides clues about conformational transitions during transport cycles of human SGLTs (Extended Data Fig. 9e). Moreover, comparing the SGLT1 and SMCT1 structures revealed commonalities and divergence within the substrate-binding pocket. Both proteins' substrate-binding pockets occupy an overall comparable region but have distinct chemical environment and location. Sequence analyses revealed that the SGLT1 and SMCT1 substrate-binding residues are conserved only within their own respective branches (Extended Data Fig. 9c). These conserved residues potentially provide a compatible environment for binding either a sugar ring by the SGLT1-belonging branch or a substrate with a negatively charged group by the SMCT1-belonging branch. Intriguingly, the substrate does not occupy the whole binding pocket, leaving space on one side to allow a variety of related substrates. It is conceivable that the pocket may change shape during transitions between conformational states, providing a structural basis for designing high-affinity substrate analogues that can occupy the extra space as competitive inhibitors for modulating diseases.

METHODS

Protein expression and purification

The coding sequence of human SGLT1 was cloned into a modified pEGBacMam vector that contained a 3C-protease site, followed by an eGFP-His-tag with a transmembrane helix from glycophorin A[31,32]. To increase the stability of SGLT1, we carried out consensus mutagenesis. Human SGLTs are grouped into two main branches according to phylogenetic analysis. The branch that SGLT1 belongs to was used to generate consensus sequence. Residues in the TM regions of SGLT1 were substituted by consensus amino acids. The residues in the extracellular or intracellular regions remain unchanged. Multiple versions of constructs were generated using different levels of residue substitution. The resulted coding sequences were codon optimized. Their biochemical behaviors were evaluated by small scale purifications. A biochemically well-behaved construct, called $SGLT1_{con}$ (ED

Fig. 1) was chosen for large scale purification. Protein expression was carried out in mammalian HEK293S cells (ATCC #CRL-3022; no further authentications or mycoplasma contamination tests were performed for this study). HEK293S cells infected with BacMam virus were harvested approximately 48 hours after the addition of 10 mM sodium butyrate[32]. Crude cell membrane was prepared by lysing cells using a Dounce homogenizer in 20 mM HEPES (pH 7.0), 150 mM NaCl, and protease inhibitors, followed by centrifugation. The membrane was solubilized with 1% (w/v) lauryl maltose neopentyl glycol (LMNG, Anatrace) and 0.1% (w/v) CHS (Anatrace) at 4°C for 1 h. The solubilized fraction was incubated with cobalt resin at 4°C for 1 h. Impurities were removed by washing the resin with wash buffer [0.02% (w/v) LMNG, 0.002% (w/v) CHS, and 40 mM imidazole]. $SGLT1_{con}$ was released from the tags by incubating with 3C protease. Purified $SGLT1_{con}$ was mixed with nanobody (1:3 molar ratio) and incubated at 4°C for 1 h. Final purification was carried out on a size-exclusion column (Superdex 200 Increase 10/300 GL, GE Healthcare) in the presence of 20 mM HEPES (pH 7.0), 150 mM NaCl, 0.001% (w/v) LMNG, 0.00033% (w/v) glyco-diosgenin (GDN, Anatrace), and 0.00013% (w/v) CHS. Peak fractions were concentrated to about 30 mg/ml for structural studies. The $SGLT1_{conHA}$ and nanobody complex was prepared in the presence of 200 μM phlorizin or 1 mM indican during the purification process following the above procedure. The final purification was carried out by gel-filtration in the presence of 20 mM HEPES (pH 7.0), 150 mM NaCl, 0.001% (w/v) LMNG, 0.00033% (w/v) GDN, and 0.00013% (w/v) CHS. The peak fractions were concentrated to around 15 mg/ml for structural studies. For SMCT1, no additional helix was added to the C-terminal of the constructs. The SMCT1-nanobody complex was prepared similarly as in $SGLT1_{conHA}$. The final purification was carried out by gel-filtration in the presence of 20 mM HEPES (pH 7.0), 150 mM NaCl, 0.002% (w/v) LMNG, and 0.0002% (w/v) CHS. Peak fractions were pooled and concentrated to about 22 mg/ml for structural characterization.

Nanobody selection and purification

Nanobodies that bind to $SGLT1_{con}$ or SMCT1 were identified following published protocols[33,34]. After four rounds of selection from a phage display library, enriched binders were individually evaluated by pull-down assay using nanobodies as the immobilized protein. Nanobodies that specifically bound to $SGLT1_{con}$ or SMCT1 were expressed in *E. coli* BL21(DE3) at 37°C as His-tagged proteins. Nanobody purification followed a standard purification protocol and the protein was eluted by 300 mM imidazole [33]. The purified nanobodies were concentrated and frozen at −80°C for further use.

Uptake assay

The sugar uptake assay was performed following a published protocol[35]. In brief, *Xenopus laevis* oocytes (Ecocyte) injected with 28 nL mRNA or water were incubated at 18°C for 3-5 days. For uptake experiments, oocytes were first incubated in a Na^+-free buffer (120 mM choline chloride, 2 mM KCl, 1 mM $CaCl_2$, 1 mM $MgCl_2$, 10 mM HEPES, pH 7.4) for 30 min. The uptake buffer contained 100 μM αMDG and 4 μM [^{14}C]-αMDG (PerkinElmer) in the presence of Na^+ (120 mM NaCl) or in the absence of Na^+ (120 mM choline chloride). To assay for phlorizin inhibition, phlorizin (200 μM) was included in the presence of sodium. Uptake was carried out at room temperature. After incubation for 30-60 min, the incubation

buffer was removed and the oocytes were washed with 4x0.6 ml ice-cold Na^+-free buffer. Oocytes were solubilized using 10% (w/v) sodium dodecyl sulfate (SDS) and ^{14}C was quantified by scintillation.

The oocyte-based galactose or pyruvate or urea[36] uptake assay was performed similar as sugar uptake assay. Specifically, the uptake buffer contained 100 μM galactose and 9 μM [^{14}C]-galactose or 50 μM [^{14}C]-pyruvate (PerkinElmer) or 27.5 μM [^{14}C]-urea (PerkinElmer).

Cryo-EM sample preparation and image acquisition

A 3.5 μL aliquot of sample ($SGLT1_{con}$-Nb1, $SGLT1_{conHA}$-Nb1 or SMCT1-Nb2) supplemented with 0.05% (w/v) β-octyl glucoside was applied to glow-discharged 200 mesh grids (Quantifoil R1.2/1.3) and plunge-frozen using an FEI Vitrobot Mark IV (Thermo Fisher Scientific). Several sessions of data collection for $SGLT1_{con}$-Nb1, or SMCT1-Nb2 were conducted on the same Titan Krios equipped with an energy filter operated at 300 keV at a magnification of 47,000x, using a Gatan K2 Summit direct electron detector in counting mode, at a pixel size of 1.06 Å. Movie stacks were obtained with a defocus range of −1.0 to −2.0 μm, using SerialEM[37] with a set of customized scripts enabling automated low-dose image acquisition. Each movie stack was recorded for a total of 8 seconds with 0.2 seconds per frame. The exposure rate was 7 electrons per pixel per second. For $SGLT1_{conHA}$-Nb1 complex, cryo-EM imaging was performed on a Titan Krios electron microscope operated at 300 keV with a Gatan K3 direct electron detector, at a magnification of 58,000x (in super-resolution mode) and a pixel size of 0.426 Å. Movie stacks were collected for 2.5 s at a total dose of 67.88 electrons/$Å^2$ and a dose per frame of 1.08 e/$Å^2$/frame, with a defocus range from −0.8 to −1.8 μm.

Cryo-EM data processing

For the $SGLT1_{con}$ sample, a total of 15,039 image stacks were subjected to beam-induced motion correction using MotionCor2[38]. Contrast transfer function parameters for each micrograph were estimated from the exposure-weighted averages of all frames by Gctf v1.06[39]. The following processes were performed using RELION3[40], except those mentioned specifically. After 2D classification 1,587,680 particle projections were divided into four subsets for 3D classification. A reference map for 3D classification was generated by the “3D initial model” script in RELION3 using default stochastic gradient descent parameters. All of the stable classes were combined for a second round of 3D classification, leading to the selection of 262,587 projections for 3D auto-refinement. After Bayesian polishing of particles, the refinement using a soft mask that included the detergent micelle, produced a final map at 3.4 Å, as determined by the gold-standard measure of Fourier shell correlation (FSC) using a cutoff of 0.143. Local resolution was estimated with the Bsoft package[41].

Processing for SMCT1 sample was executed similarly as described above. Briefly, a total 8,823 image stacks were motion-corrected by MotionCor2. After 2D classification, 2,523,340 particles were divided into four subsets for 3D classification. Stable partitions were then combined for another round of 3D classification. A total of 173,859 good

particles were again masked to generate the final 3.5 Å map after Bayesian polishing. Local resolution was estimated with the Bsoft package.

The dataset for the $SGLT1_{conHA}$-Nb1 complex consisted of 7,026 total movie stacks, which were motion-corrected by MotionCor2. CTF parameters of the corrected micrographs were estimated using CTFfind and used to cull the dataset to 2,308 micrographs based on estimated resolution threshold of 3.4 Å. A total of 2,627,764 particles were extracted in Relion 3.1 and subject to two rounds of 2D classifications in Cryosparc. This yielded 878,859 particles that were used for three rounds of 3D classification and Heterogeneous Refinement. A final set of 94,369 particles was selected for the map reconstruction and ran through a Homogeneous and Non-Uniform refinements in Cryosparc before being imported back into Relion. In Relion, the particles' CTF parameters were refined for beam tilt and trefoil aberrations. The particles were then Bayesian polished before being imported back to Cryosparc for a set of Homogeneous and Non-Uniform refinements, resulting in the final map at 3.15 Å global estimated resolution as determined by the gold-standard measure of Fourier shell correlation (FSC) using a cutoff of 0.143.

Model building and refinement

The initial model of $SGLT1_{con}$ or SMCT1 was generated through SWISS-MODEL[42] using vSGLT (PDB: 3DH4) as a homology model. The initial model was manually rebuilt in COOT[43] and refined by real space refinement in PHENIX[44]. The clear side-chain densities in the map allowed for unambiguous sequence register. Stereochemistry and geometry were assessed by MolProbity[45]. Structural figures were generated using PyMOL[46], Chimera[47] and ChimeraX[48]. The interaction network of SMCT1-butyrate was present with LigPlot+[49].

System setup for molecular dynamics simulations

We performed simulations of $SGLT1_{con}$ in a hydrated lipid bilayer under four conditions: (1) simulations with cholesterol initially placed in the cholesterol/CHS binding site at the position proposed on the basis of the cryo-EM density (5 independent simulations, 3 μs each); (2) simulations with glucose and two sodium ions initially placed in the glucose and sodium binding sites, respectively (5 independent simulations, 2 μs each); (3) simulations with glucose initially placed in the glucose binding site and no sodium ions placed in the sodium binding sites (5 independent simulations, 2 μs each); and (4) simulations in the apo form (5 independent simulations, 4 μs each). We initiated all simulation conditions from the human $SGLT1_{con}$ structure reported in this paper, after removing the nanobody from the structure. For each simulation, initial atom velocities were assigned randomly and independently.

For simulations with glucose initially placed in the glucose binding site (conditions 2 and 3), we modeled glucose into $SGLT1_{con}$ by aligning the human $SGLT1_{con}$ structure reported in this paper to the previously published galactose-bound bacterial SGLT (vSGLT) structure (PDB ID: 3DH4)[11], and converting galactose to glucose using Maestro (Schrödinger). For simulations with sodium ions initially placed in the sodium binding sites (condition 2), the initial position of the sodium ions were determined by aligning the human $SGLT1_{con}$ structure reported in this paper to the previously published sodium

sialic acid symporter structure in complex with two sodium ions (PDB ID: 5NV9)[12]. For all simulation conditions, the protein structures were aligned to the Orientations of Proteins in Membranes[50] entry for 5NV9 (sodium sialic acid symporter) using PyMOL (Schrödinger)[46], and crystal waters from 5NV9[12] were incorporated. Prime (Schrödinger)[51] was used to model missing side chains, and to add capping groups to protein chain termini. Protonation states of all titratable residues were assigned at pH 7. Histidine residues were modeled as neutral, with a hydrogen atom bound to either the delta or epsilon nitrogen depending on which tautomeric state optimized the local hydrogen-bonding network. Using Dabble[52], the prepared protein structures were inserted into a pre-equilibrated palmitoyl-oleoyl-phosphatidylcholine (POPC) bilayer, the system was solvated, and sodium and chloride ions were added to neutralize the system and to obtain a final concentration of 150 mM. For all simulation conditions, the final systems comprised approximately 114,000 atoms and system dimensions were approximately 120 x 120 x 100 Å.

Molecular dynamics simulation and analysis protocols

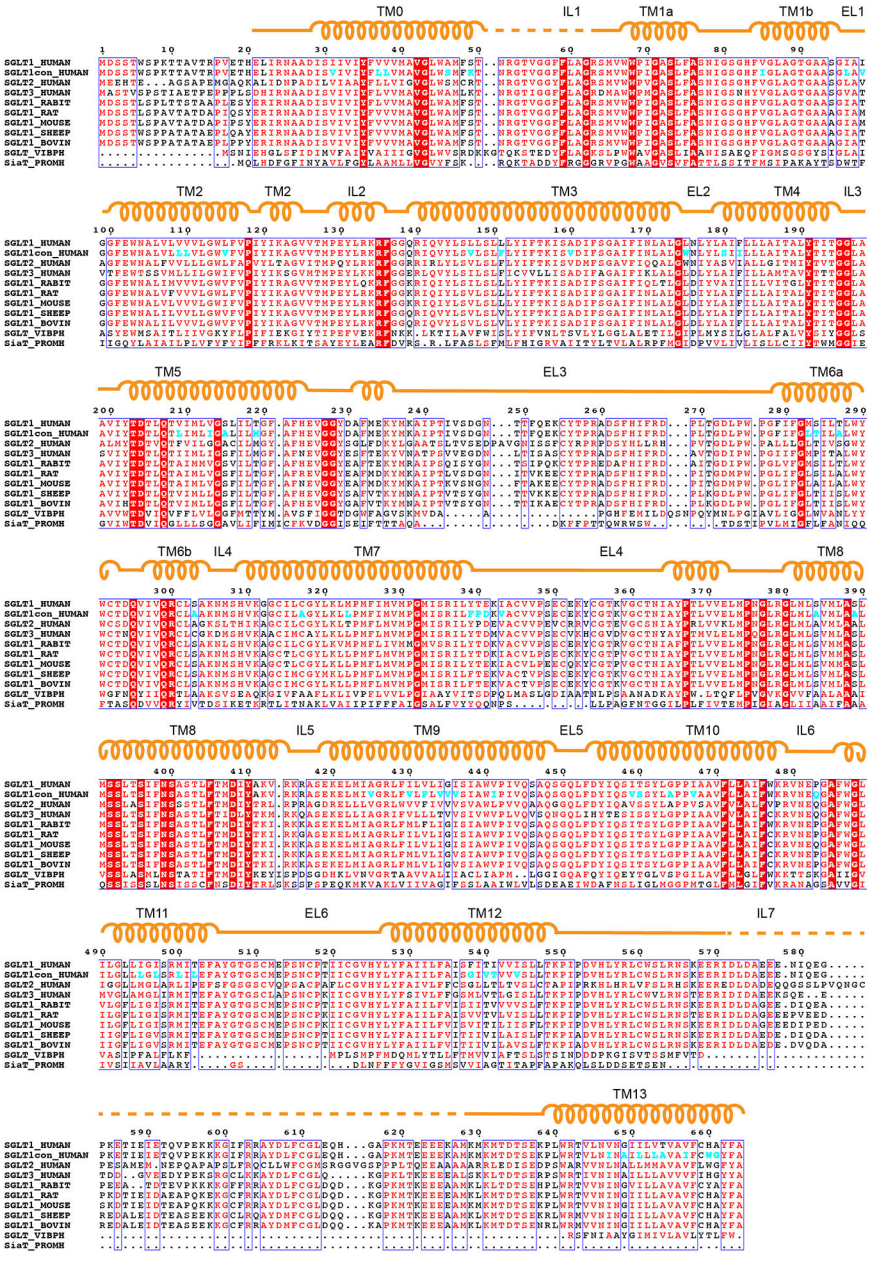
We used the CHARMM36m force field for proteins, the CHARMM36 force field for lipids, ions, and glucose, and the TIP3P model for water[53-55]. All simulations were performed using the Compute Unified Device Architecture (CUDA) version of particle-mesh Ewald molecular dynamics (PMEMD) in AMBER18[56] on graphics processing units (GPUs).

Systems were first minimized using three rounds of minimization, each consisting of 500 cycles of steepest descent followed by 500 cycles of conjugate gradient optimization. 10.0 and 5.0 kcal·mol^{-1}·Å^{-2} harmonic restraints were applied to protein, lipids, and glucose for the first and second rounds of minimization, respectively. 1 kcal·mol^{-1}·Å^{-2} harmonic restraints were applied to the protein, glucose and cholesterol non-hydrogen atoms for the third round of minimization. Systems were then heated from 0 K to 100 K in the NVT ensemble over 12.5 ps and then from 100 K to 310 K in the NPT ensemble over 125 ps, using 10.0 kcal·mol^{-1}·Å^{-2} harmonic restraints applied to protein, glucose, and cholesterol heavy atoms. Subsequently, systems were equilibrated at 310 K and 1 bar in the NPT ensemble, with harmonic restraints on the protein, glucose and cholesterol non-hydrogen atoms tapered off by 1.0 kcal·mol^{-1}·Å^{-2} starting at 5.0 kcal·mol^{-1}·Å^{-2} in a stepwise fashion every 2 ns for 10 ns, and then by 0.1 kcal·mol^{-1}·Å^{-2} every 2 ns for 20 ns. Production simulations were performed without restraints at 310 K and 1 bar in the NPT ensemble using the Langevin thermostat and the Monte Carlo barostat, and using a timestep of 4.0 fs with hydrogen mass repartitioning[57]. Bond lengths were constrained using the SHAKE algorithm[58]. Non-bonded interactions were cut off at 9.0 Å, and long-range electrostatic interactions were calculated using the particle-mesh Ewald (PME) method with an Ewald coefficient of approximately 0.31 Å, and 4th order B-splines. The PME grid size was chosen such that the width of a grid cell was approximately 1 Å. Trajectory frames were saved every 200 ps during the production simulations.

The AmberTools17 CPPTRAJ package was used to reimage trajectories[59]. Simulations were visualized and analyzed using Visual Molecular Dynamics (VMD)[60] and PyMOL (Schrödinger)[46]. Fig. 2a shows a representative frame from simulation no. 1 of condition 2. This frame was chosen such that the RMSD of glucose from its initial position closely

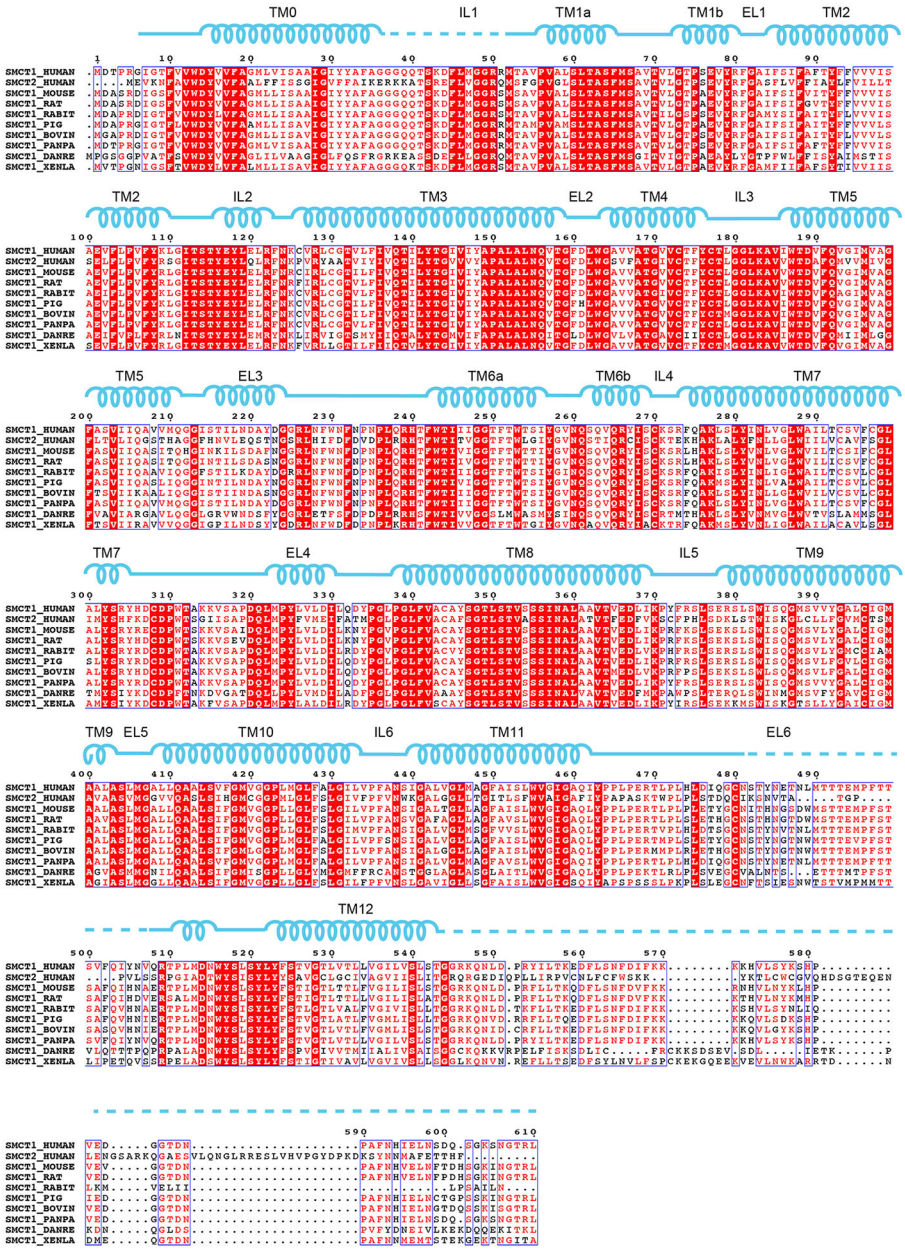
approximates the average glucose RMSD during the simulation, when frames are aligned on the protein transmembrane helices. In Extended Data Fig. 7e, 10a, 10b, 10g, unsmoothed traces (thin lines) and traces smoothed with a moving average (thick lines) are shown for 5 simulations. These time traces were smoothed using a moving average with a window size of 20 ns, and plots were visualized using the PyPlot package from Matplotlib. For Extended Data Fig. 10c, frames from simulation no. 1 of condition 2 were aligned on the protein transmembrane helices, and the position of glucose in the binding pocket was visualized every 200 ns. Hydrogen bonds between the glucose molecule and the protein in Extended Data Fig. 10d, 10e were detected using the VMD Hydrogen Bonds plug-in with a heavy atom donor-acceptor distance of 3.5 Å and an angle cutoff of 60°. Water occupancy maps in Extended Data Fig. 10f were generated using the GIST implementation in AmberTools17 CPPTRAJ[61,62]. Frames from every 2 ns of simulation were aligned to the initial structure and then used as input, with a grid size of 0.25 Å. The resulting map was smoothed using a Gaussian filter with a standard deviation of two grid cells.

Extended Data



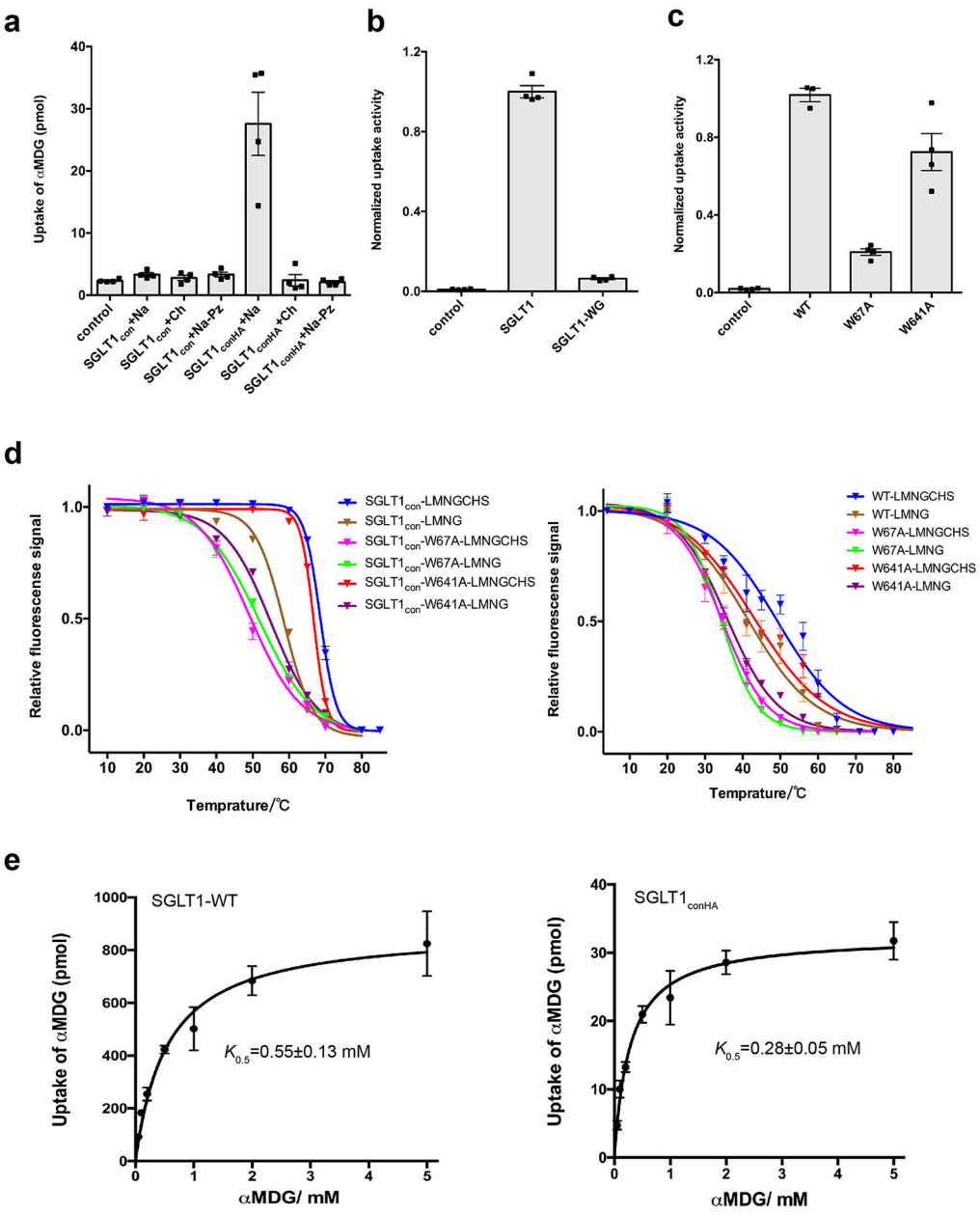
Extended Data Fig. 1 |. Sequence alignments of selected SSS transporters.

The sequences of selected SSS transporters were aligned using Clustal Omega (http://www.uniprot.org/) and adjusted manually. The secondary structural elements of SGLT1 are indicated above the sequence alignment. The mutations in $SGLT1_{con}$ are highlighted.



Extended Data Fig. 2 |. Sequence alignments of SMCT homologs.

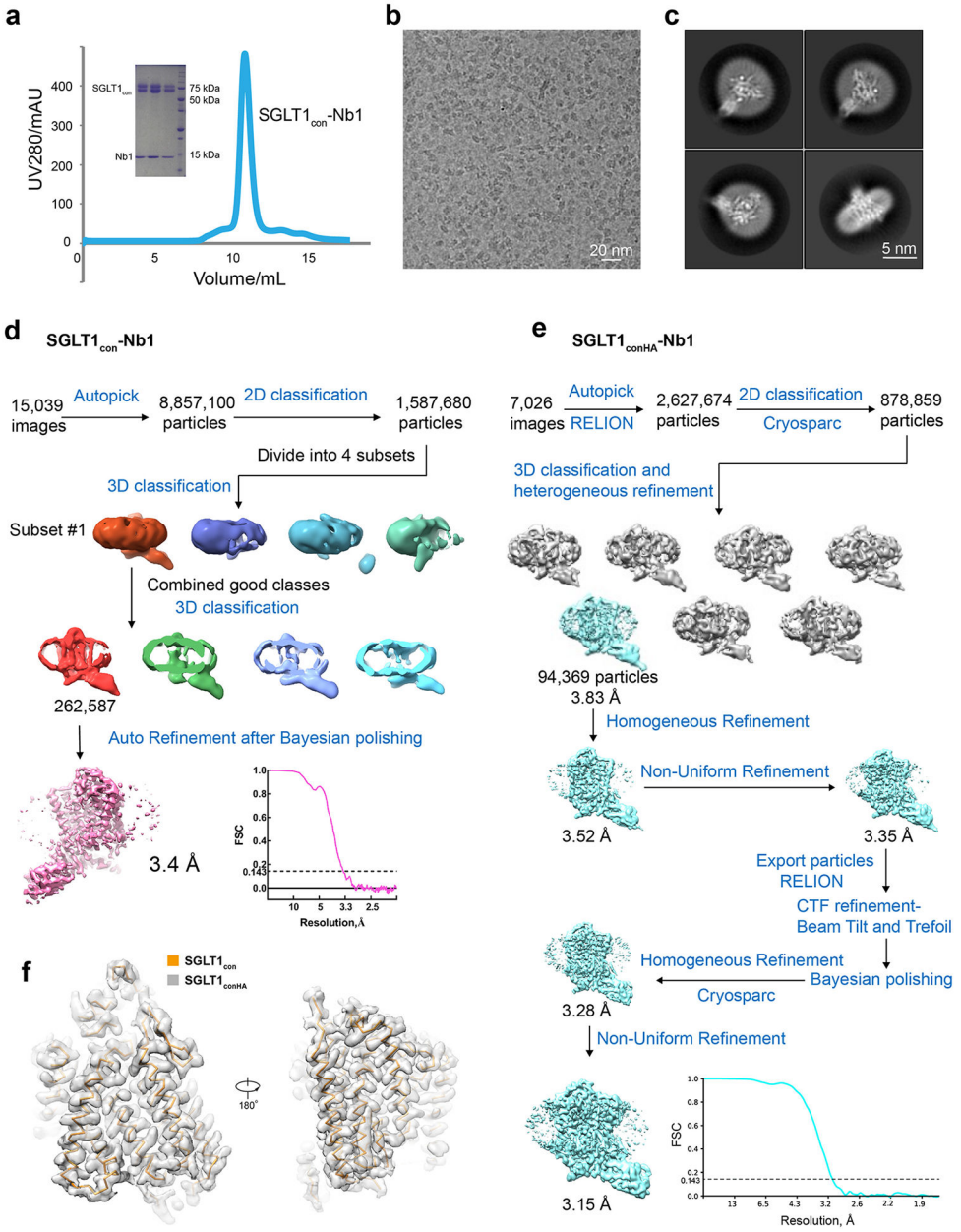
The sequences of selected SMCT homologs were aligned using Clustal Omega (http://www.uniprot.org/) and adjusted manually. The secondary structural elements of SMCT1 are indicated above the sequence alignment.



Extended Data Fig. 3 | Glucose uptake of SGLT1 mutants

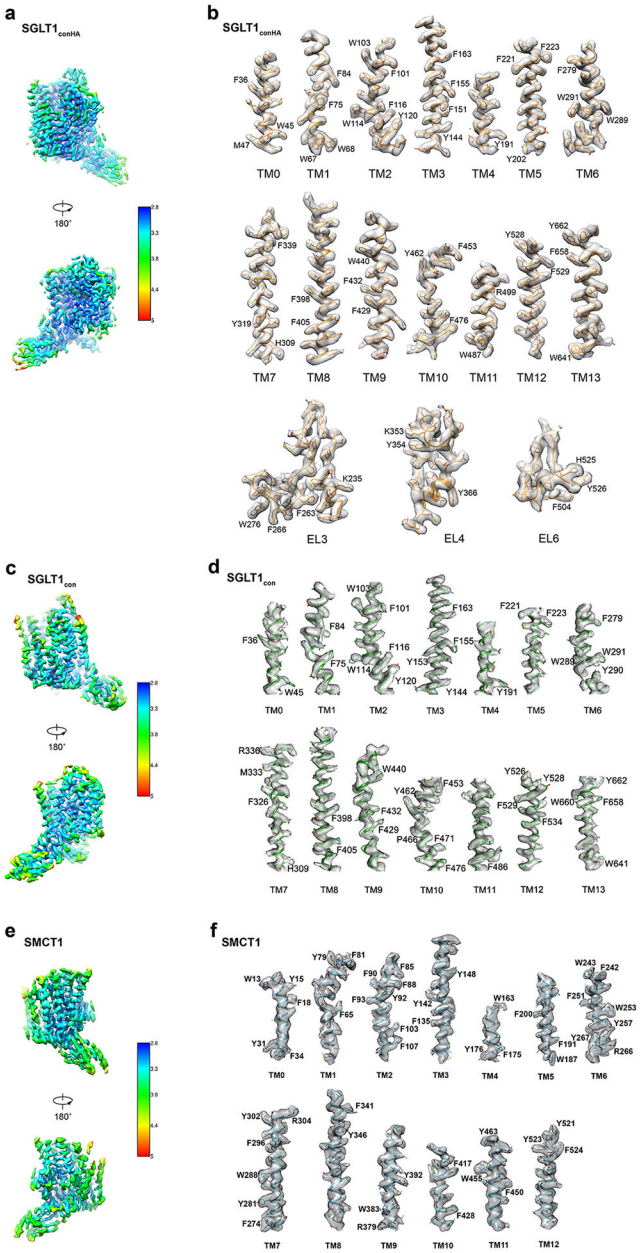
a, Uptake activities of $SGLT1_{con}$ and $SGLT1_{conHA}$. $SGLT1_{conHA}$ is the same as $SGLT1_{con}$, except that W660 and G661 of $SGLT1_{con}$ are reversed to H660 and A661. The uptake buffer contains different combinations of sodium (Na), choline (Ch), or phlorizin (Pz) as indicated. Uptake of αMDG is shown (mean ± SEM; n=4 biological replicates). **b**, Uptake activities of SGLT1 and SGLT1-WG mutant, in which H660 and A661 are substituted with W660 and G661 (mean ± SEM; n=4 biological replicates). **c**, Uptake activities of SGLT1 with mutations in the cholesterol-binding site (mean ± SEM; n=4 biological replicates). **d**, Thermostability of $SGLT1_{con}$ (left) or SGLT1 (right) and their variants under conditions with or without cholesteryl hemisuccinate (CHS) (mean ± SEM; n=3 biological replicates). **e**, The transport of αMDG by SGLT1 WT (left) and $SGLT1_{conHA}$ (right) in the presence of

various concentrations of αMDG. Data were plotted according to the equation, $U=U_{max}$ x [S]/ ($K_{0.5}$+[S]) (mean ± SEM; n=3 biological replicates).

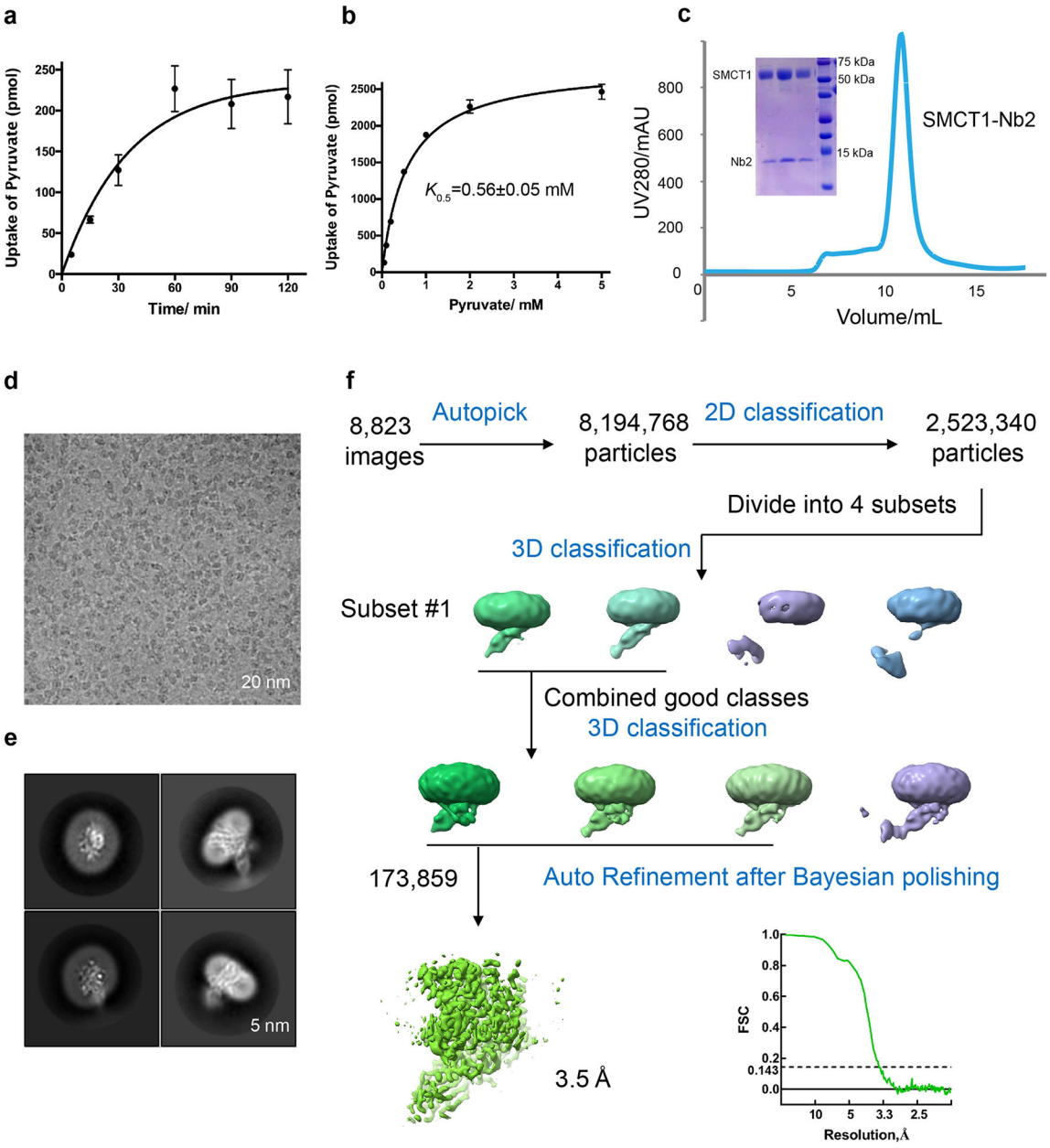


Extended Data Fig. 4 |. Cryo-EM sample preparation and data processing of SGLT1.
a, The elution profile of SGLT1$_{con}$-Nb1 on a size-exclusion column. The insert shows SDS-PAGE analysis of the purified sample. Data are representative of five independent experiments with similar results. **b**, Representative cryo-EM micrograph of SGLT1$_{con}$-Nb1 complex particles (from 15,039 micrographs with similar results). **c**, Selected 2D class averages of SGLT1$_{con}$-Nb1 complex (from 100 classes with similar results). **d**, **e**, The workflow of classification and refinement. The overall nominal resolutions of the SGLT1$_{con}$-Nb1 complex and the SGLT1$_{conHA}$-Nb1 complex were determined by the 'gold standard'

FSC curve using the FSC=0.143 criterion. **f**, Overlay of the $SGLT1_{con}$ model with the density map of $SGLT1_{conHA}$.



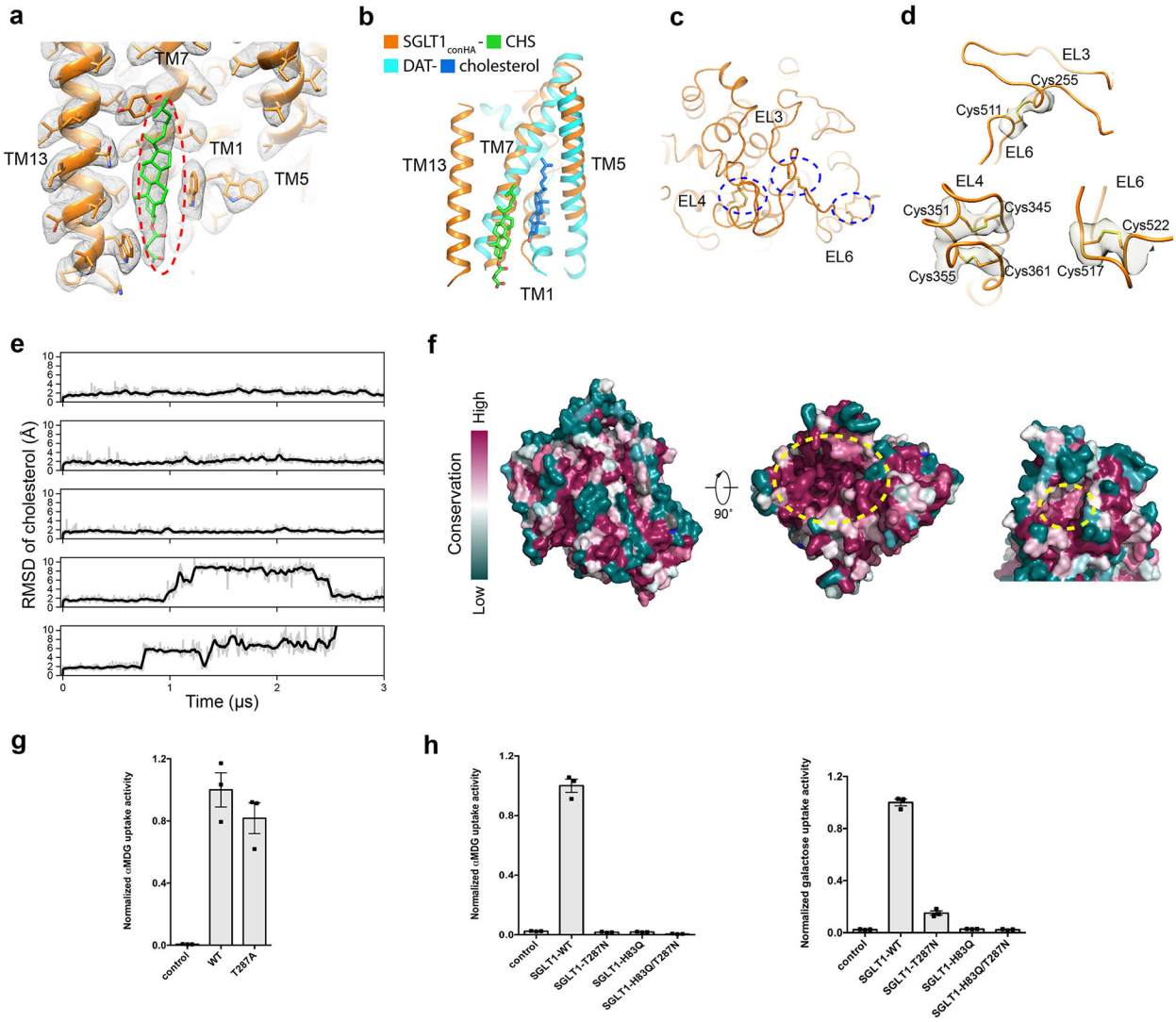
Extended Data Fig. 5 |. Cryo-EM densities and refined models.
a, Local resolution of the cryo-EM map of the $SGLT1_{conHA}$-Nb1 complex. **b,** Cryo-EM densities and model of $SGLT1_{conHA}$ transmembrane helices and extracellular loops. **c**, Local resolution of the cryo-EM map of the $SGLT1_{con}$-Nb1 complex. **d,** Cryo-EM densities and model of $SGLT1_{con}$ transmembrane helices. **e**, Local resolution of the cryo-EM map of the SMCT1-Nb2 complex. **f,** Cryo-EM densities and model of SMCT1 transmembrane helices.



Extended Data Fig. 6 | Functional characterization, cryo-EM sample preparation, and data processing of SMCT1.

a, Time course of pyruvate uptake by SMCT1 expressing oocytes (mean ± SEM; n=3 biological replicates). **b**, The transport of pyruvate by SMCT1 in the presence of various concentrations of pyruvate (mean ± SEM; n=3 biological replicates). **c**, The size-exclusion chromatography of SMCT1-Nb2 complex and SDS-PAGE analysis of the purified sample. Data are representative of five independent experiments with similar results. **d**, Representative cryo-EM micrograph of the SMCT1-Nb2 complex (from 8,823 micrographs with similar results). **e**, Selected 2D class averages of SMCT1-Nb2 complex particles (from 100 classes with similar results). **f**, The workflow of data processing on the SMCT1-Nb2

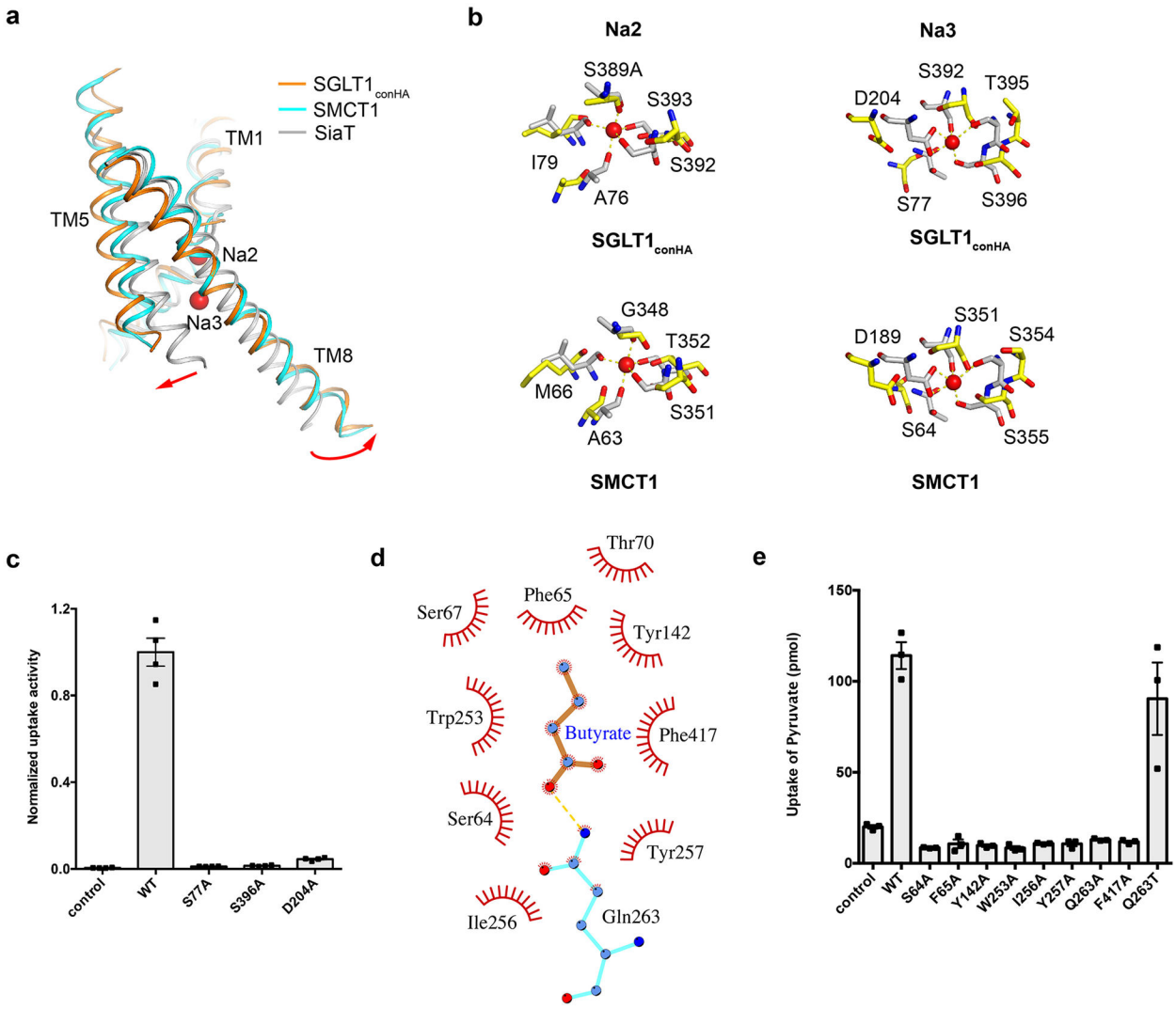
complex. The overall nominal resolutions of the SMCT1-Nb2 complex were determined by the ‘gold standard’ FSC curve using the FSC=0.143 criterion.



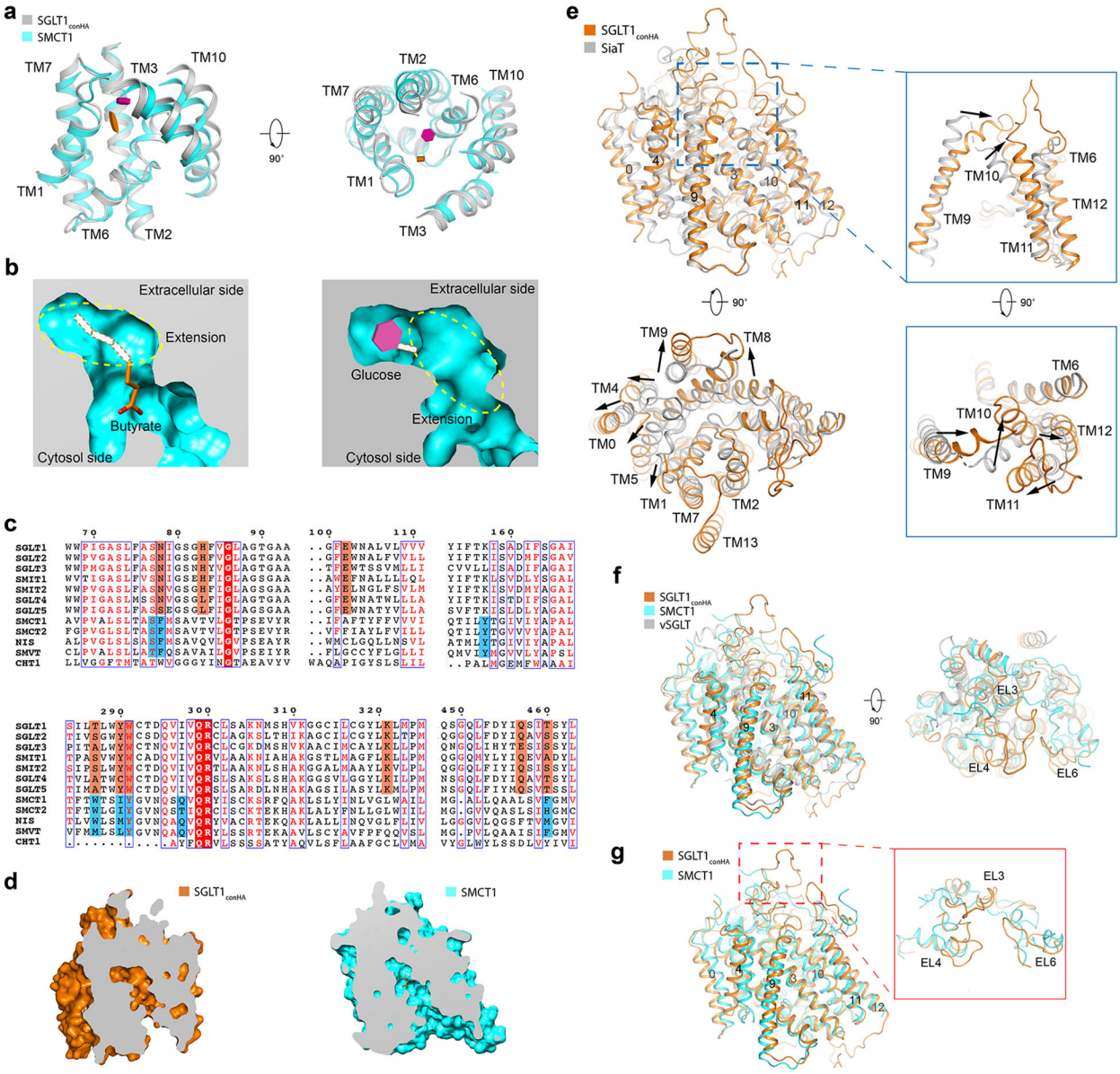
Extended Data Fig. 7 |. Structural features of SGLT1 and substrate selectivity.

a, The density map of the cholesteryl hemisuccinate binding site of $SGLT1_{conHA}$. The cholesteryl hemisuccinate density map is indicated by dashed red oval. **b**, Structural overlay of $SGLT1_{conHA}$ (orange) and a dopamine transporter DAT (cyan, PDB: 4M48). The helices near the cholesterol binding site are shown as ribbons. **c**, The overall organization of the lid domain of $SGLT1_{con}$. Disulfide bonds are indicated by dashed blue circles. **d**, Close view of the four disulfide bonds. **e**, For five simulations of cholesterol-bound $SGLT1_{con}$, the heavy atom RMSD of cholesterol from its initial position is plotted over time. **f**, Conservation surface mapping of SGLT1. The conservation scores are calculated from 200 SGLT1 sequences using ConSurf. The cytosolic vestibule (middle) and extracellular cavity (right) are indicated by dashed yellow oval. **g**, Uptake activities of SGLT1 T287A variant (mean ± SEM; n=3 biological replicates. WT and control are the same as in Fig. 2c). **h**,

αMDG (left) and galactose (right) uptake activities of SGLT1 variants with point mutations in the substrate binding pocket on residues that are different between SGLT1 and vSGLT (mean ± SEM; n=3 biological replicates).



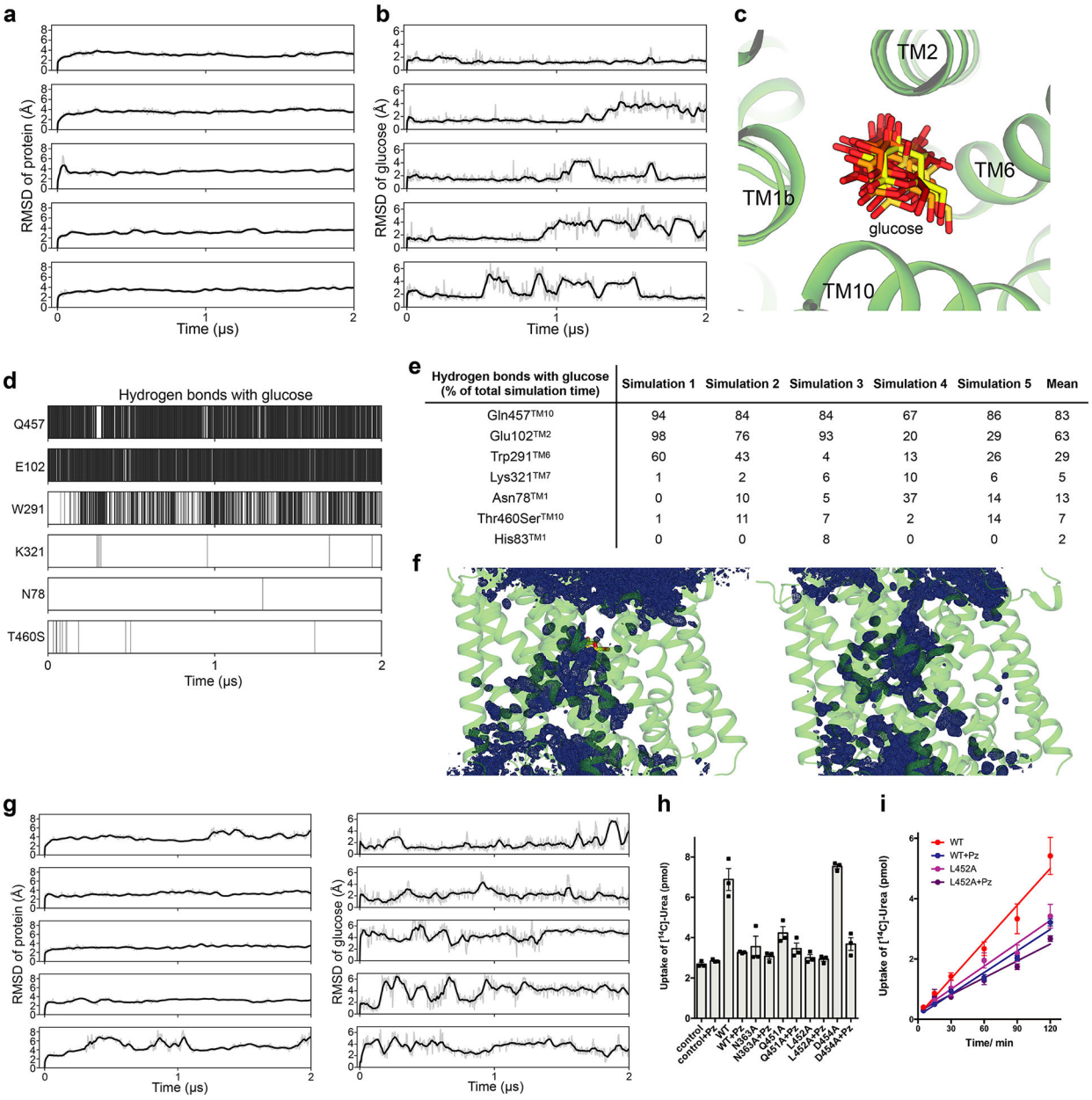
Extended Data Fig. 8 |. Sodium-binding sites and SMCT1's substrate binding site.
a, Structural overlay of $SGLT1_{conHA}$ (orange), SMCT1 (cyan), and SiaT (gray, PDB: 5NV9). The helices involved in sodium binding are shown as ribbons. The shift of TM helices from SiaT to $SGLT1_{conHA}$ or SMCT1-butyrate is depicted by red arrows. **b**, Sodium-binding sites compared between $SGLT1_{conHA}$ (yellow) or SMCT1 (yellow) and SiaT (gray). Residues in Na2 or Na3 sites are shown as sticks. **c**, Uptake activities of SGLT1 with mutations in the Na3-binding site. Uptake activities are normalized to WT (mean ± SEM; n=4 biological replicates). **d**, SMCT1-butyrate interaction shown as Ligplot^{+} diagram (yellow dashed lines, hydrogen bonds; spokes, hydrophobic interactions). **e,** Pyruvate uptake activities of SMCT1 mutants compared to WT (mean ± SEM; n=3 biological replicates).



Extended Data Fig. 9 |. Structural comparison among members of SSS family.

a, Superposition of the binding pockets of SMCT1 and apo SGLT1$_{conHA}$. Transmembrane helices involved in forming the central pockets are labeled. The relative positions of substrates are indicated by 3D shapes: pink hexagon for glucose and orange oval for butyrate. SMCT1 and SGLT1$_{conHA}$ are colored in cyan and gray, respectively. **b**, Surface representation of the central cavity of SMCT1 (left) and SGLT1$_{conHA}$ (right). Substrates are placed in the binding sites, the extension of which is indicated by dashed yellow ovals. **c**, The sequence comparison of the substrate binding site residues. For the sugar-transporting branch, the positions equivalent to SGLT1's sugar binding site residues are highlighted in orange. For the metabolite-transporting branch, the positions equivalent to SMCT1's substrate binding site residues are highlighted in blue. **d**, Sliced view of SGLT1$_{conHA}$ (left panel) and SMCT1-butyrate (right panel). **e**, The superimposed SGLT1 (inward-facing) and

SiaT (outward-facing). Zoomed-in view of regions that undergo considerable conformational changes are shown in blue boxes on the right (unrelated helices or loops are removed for clarity). The shift of TM helices between SiaT (gray) and SGLT1 (orange) is indicated by black arrows. From outward-open to inward-facing conformation, the N-terminal half of TM10 undergoes significant inward movement around a Gly-Pro-Pro motif at the center of the helix. Concomitantly, the short loop connecting TM9-TM10 and the C-terminal part of TM9 also moves inward. As a result, the N-terminal end of TM10 and the TM9-TM10 loop come into contact with TM2 and EL4 of the extracellular domain, which collapses the extracellular vestibule and stabilizes the closed conformation of the extracellular gate. $Phe453^{TM10}$, at the end of TM10, thus moves into a position to contact other extracellular gate residues to shield the substrate-binding pocket from the extracellular solution. In association, TM11 and TM12 tilt away to accommodate the movement of TM10. On the intracellular side, TM5, together with TM4, tilts outward while TM8 and TM9 tilt away from TM1, TM5 and TM6. These movements open the intracellular entrance of the vestibule and widen the permeation pathway to enable substrate release. The increased distance between TM8 and TM1/TM5 is linked with the disruption of both Na2 and Na3 sites. Thus, Na^+ binding is coupled to the conformational changes during state transitions and glucose transport. During state transitions, the extracellular lid domain is also expected to undergo significant conformational changes, which might help stabilize conformational states or give rise to distinct surface features to modulate Na^+ transport. **f**, Structural comparison of $SGLT1_{conHA}$ (orange), SMCT1 (cyan) and vSGLT (gray, PDB: 3DH4). **g**, Structural comparison of $SGLT1_{conHA}$ (orange), SMCT1 (cyan). The orientation difference of the extracellular domain is zoomed-in in red box.



Hydrogen bonds with glucose (% of total simulation time)	Simulation 1	Simulation 2	Simulation 3	Simulation 4	Simulation 5	Mean
Gln457^{TM10}	94	84	84	67	86	83
Glu102^{TM2}	98	76	93	20	29	63
Trp291^{TM6}	60	43	4	13	26	29
Lys321^{TM7}	1	2	6	10	6	5
Asn78^{TM1}	0	10	5	37	14	13
Thr460Ser^{TM10}	1	11	7	2	14	7
His83^{TM1}	0	0	8	0	0	2

Extended Data Fig. 10 |. Substrate binding and water permeation of SGLT1.

a, For five simulations of $SGLT1_{con}$ with glucose and two sodium ions initially placed in the glucose and sodium binding sites, the backbone RMSD of $SGLT1_{con}$ from the cryo-EM structure is plotted over time. **b**, For five simulations of $SGLT1_{con}$ with glucose and two sodium ions initially placed in the glucose and sodium binding sites, the heavy atom RMSD of glucose from its initial position is plotted over time. **c**, A frame from every 200 ns of the 2 μs molecular dynamics simulation (simulation no. 1 in panels (a) and (b)) shows the position of glucose in the binding pocket (red at t = 0, transitioning to yellow at t = 2 μs). **d**, Hydrogen bonds between glucose and $SGLT1_{con}$ binding pocket residues are shown in black lines for simulation no. 1 in panels (a) and (b). **e**, Hydrogen bonds between glucose and the $SGLT1_{con}$ binding pocket residues are shown as the percentage of total simulation

time for all simulations in panels (a) and (b). **f**, Water occupancy averaged over time for glucose-bound $SGLT1_{con}$ simulation no. 1 in panels (a) and (b) (left), and for a 4 μs apo $SGLT1_{con}$ simulation (right). Water density is shown in dark blue mesh (contoured at 0.0334 water molecules/$Å^3$, approximately the bulk density), and glucose is shown in yellow sticks. **g**, The backbone RMSD of $SGLT1_{con}$ from the cryo-EM structure (left), and the heavy atom RMSD of glucose from its initial position (right) for five simulations of $SGLT1_{con}$ with glucose initially placed in the glucose binding site and with no sodium initially placed in the sodium binding sites. Sodium did not enter the sodium binding sites during the timescale of these simulations. **h**, Uptake activities of SGLT1 mutants. Oocyte-based urea uptake activities with or without inhibitor (200 μM phlorizin, Pz) are compared to the wild-type (WT) transporter (mean ± SEM; n=3 biological replicates). **i**, Time course of urea uptake by SGLT1 WT and mutants with or without inhibitor phlorizin (Pz). Uptake data were plotted using linear regression (mean ± SEM; n=3 biological replicates).

Extended Data Table 1 |

Cryo-EM data collection, refinement, and validation statistics.

	$SGLTl_{con}$-Nb1 (EMDB-25194) (PDB 7SL8)	**S $GLT1_{conHA}$-Nb1 (EMDB-25196) (PDB 7SLA)**	**SMCT1-Nb2 (EMDB-25195) (PDB 7SL9)**
Data collection and processing			
Magnification	47,000	58,000	47,000
Voltage (kV)	300	300	300
Electron exposure (e−/$Å^2$)	56	67.88	56
Defocus range (μm)	−1.0~−2.0	−0.8~−1.8	−1.0~−2.0
Pixel size (Å)	1.06	super-resolution 0.426	1.06
Symmetry imposed	C1	C1	C1
Initial particle images (no.)	8,857,100	2,627,674	8,194,768
Final particle images (no.)	262,587	94,369	173,859
Map resolution (Å)	3.4	3.15	3.5
FSC threshold	0.143	0.143	0.143
Map resolution range (Å)	2.8-5	2.78-4.9	2.8-5
Refinement			
Initial model used (PDB code)			
Model resolution (Å)	3.84	3.31	3.55
FSC threshold	0.5	0.5	0.5
Model resolution range (Å)	3.65-212	3.11-258	3.25-212
Map sharpening *B* factor ($Å^2$)	−161	−94.1	−186
Model composition			
Non-hydrogen atoms	5,295	5458	4713
Protein residues	700	702	619
Ligands	1	1	1
B factors ($Å^2$)			
Protein	80.58	84.41	96.54
Ligand	85.02	94.69	61.49

	$SGLTl_{con}$-Nb1 (EMDB-25194) (PDB 7SL8)	$S GLT1_{conHA}$-Nb1 (EMDB-25196) (PDB 7SLA)	SMCT1-Nb2 (EMDB-25195) (PDB 7SL9)
R.m.s. deviations			
Bond lengths (Å)	0.008	0.009	0.005
Bond angles (°)	0.977	1.010	0.943
Validation			
MolProbity score	1.97	1.76	1.62
Clashscore	8.9	3.54	4.46
Poor rotamers (%)	0.92	1.74	0.4
Ramachandran plot			
Favored (%)	91.91	93.23	94.11
Allowed (%)	8.09	6.77	5.89
Disallowed (%)	0	0	0

Acknowledgements.

This work was made possible by support from Stanford University and the Harold and Leila Y. Mathers Charitable Foundation to L.F. and G.S., a Dean's Fellowship to L.H, a Stanford Bio-X seed grant to R.O.D. and G.S., the SNSF Early Postdoctoral Mobility fellowship P2ELP3_187989 and the EMBO Long-Term Fellowship ALTF 544-2019 to D.A. L.F. was an NIH Director's New Innovator awardee. Cryo-EM data for this work was collected at the Stanford-SLAC cryo-EM facility. We thank Feng lab and Skiniotis lab members for helpful discussions, Dror lab members J. Paggi, S. Eismann, M. Dämgen and M. Vögele for assistance with MD simulations and analysis, and Elizabeth Montabana for support with data collection.

Data availability

The cryo-EM maps have been deposited into the Electron Microscopy Data Bank (accession numbers EMD-25194, EMD-25195 and EMD-25196). The coordinates have been deposited into the Protein Data Bank (accession numbers 7SL8, 7SL9 and 7SLA).

REFERENCES

1. Kleinzeller A & Kotyk A Membrane transport and metabolism. (Praha: Publishing House of the Czechoslovak Academy of Sciences, 1961).
2. Wright EM, Loo DDF & Hirayama BA Biology of human sodium glucose transporters. Physiol. Rev 91, 733–794 (2011). [PubMed: 21527736]
3. Hopfer U, Nelson K & Isselbacher KJ Specific glucose transport in isolated brush border membranes from rat small-intestine. J. Biol. Chem 248, 25–32 (1973). [PubMed: 4692832]
4. Wright EM Renal Na^+-glucose cotransporters. Am. J. Physiol. Renal Physiol 280, F10–F18 (2001). [PubMed: 11133510]
5. Hsia DS, Grove O & Cefalu WT An update on sodium-glucose co-transporter-2 inhibitors for the treatment of diabetes mellitus. Curr. Opin. Endocrinol. Diabetes Obes 24, 73–79 (2017). [PubMed: 27898586]
6. Hummel CS et al. Glucose transport by human renal Na^+/D-glucose cotransporters SGLT1 and SGLT2. Am. J. Physiol. Cell Physiol 300, C14–21 (2011). [PubMed: 20980548]
7. Turk E, Zabel B, Mundlos S, Dyer J & Wright EM Glucose galactose malabsorption caused by a defect in the Na^+/glucose cotransporter. Nature 350, 354–356 (1991). [PubMed: 2008213]
8. Hirschhorn N et al. Decrease in net stool output in cholera during intestinal perfusion with glucose-containing solutions. N. Engl. J. Med 279, 176–181 (1968). [PubMed: 4968807]

9. Zeuthen T, Gorraitz E, Her K, Wright EM & Loo DDF Structural and functional significance of water permeation through cotransporters. Proc. Natl. Acad. Sci. U. S. A 113, E6887–E6894 (2016). [PubMed: 27791155]
10. Loo DD et al. Passive water and ion transport by cotransporters. J. Physiol 518, 195–202 (1999). [PubMed: 10373701]
11. Faham S et al. The crystal structure of a sodium galactose transporter reveals mechanistic insights into Na^+/sugar symport. Science 321, 810–814 (2008). [PubMed: 18599740]
12. Wahlgren WY et al. Substrate-bound outward-open structure of a Na^+-coupled sialic acid symporter reveals a new Na^+ site. Nat. Commun 9, 1753 (2018). [PubMed: 29717135]
13. Leung DW, Turk E, Kim O & Wright EM Functional expression of the Vibrio parahaemolyticus Na^+/galactose (vSGLT) cotransporter in Xenopus laevis oocytes. J. Membr. Biol 187, 65–70 (2002). [PubMed: 12029378]
14. Canul-Tec JC et al. Structure and allosteric inhibition of excitatory amino acid transporter 1. Nature 544, 446–451 (2017). [PubMed: 28424515]
15. Suzuki T et al. Apical localization of sodium-dependent glucose transporter SGLT1 is maintained by cholesterol and microtubules. Acta Histochem. Cytochem 39, 155–161 (2006). [PubMed: 17327902]
16. Ghezzi C, Calmettes G, Morand P, Ribalet B & John S Real-time imaging of sodium glucose transporter (SGLT1) trafficking and activity in single cells. Physiol. Rep 5 (2017).
17. Penmatsa A, Wang KH & Gouaux E X-ray structure of dopamine transporter elucidates antidepressant mechanism. Nature 503, 85–90 (2013). [PubMed: 24037379]
18. Gagnon DG, Bissonnette P & Lapointe JY Identification of a disulfide bridge linking the fourth and the seventh extracellular loops of the Na^+/glucose cotransporter. J. Gen. Physiol 127, 145–158 (2006). [PubMed: 16446504]
19. Wright EM Diseases of renal glucose handling. Genetic Diseases of the Kidney, 131–140 (2009).
20. Sala-Rabanal M et al. Bridging the gap between structure and kinetics of human SGLT1. Am. J. Physiol. Cell Physiol 302, C1293–1305 (2012). [PubMed: 22159082]
21. Bisignano P et al. Inhibitor binding mode and allosteric regulation of Na^+-glucose symporters. Nat. Commun 9, 5245 (2018). [PubMed: 30532032]
22. Gopal E et al. Expression of slc5a8 in kidney and its role in Na(+)-coupled transport of lactate. J. Biol. Chem 279, 44522–44532 (2004). [PubMed: 15322102]
23. Miyauchi S, Gopal E, Fei YJ & Ganapathy V Functional identification of SLC5A8, a tumor suppressor down-regulated in colon cancer, as a Na(+)-coupled transporter for short-chain fatty acids. J. Biol. Chem 279, 13293–13296 (2004). [PubMed: 14966140]
24. Cui D & Morris ME The drug of abuse gamma-hydroxybutyrate is a substrate for sodium-coupled monocarboxylate transporter (SMCT) 1 (SLC5A8): characterization of SMCT-mediated uptake and inhibition. Drug Metab. Dispos 37, 1404–1410 (2009). [PubMed: 19389857]
25. Gopal E et al. Transport of nicotinate and structurally related compounds by human SMCT1 (SLC5A8) and its relevance to drug transport in the mammalian intestinal tract. Pharm. Res 24, 575–584 (2007). [PubMed: 17245649]
26. Li H et al. SLC5A8, a sodium transporter, is a tumor suppressor gene silenced by methylation in human colon aberrant crypt foci and cancers. Proc. Natl. Acad. Sci. U. S. A 100, 8412–8417 (2003). [PubMed: 12829793]
27. Ganapathy V et al. Sodium-coupled monocarboxylate transporters in normal tissues and in cancer. AAPS J. 10, 193–199 (2008). [PubMed: 18446519]
28. Diez-Sampedro A, Lostao MP, Wright EM & Hirayama BA Glycoside binding and translocation in Na(+)-dependent glucose cotransporters: comparison of SGLT1 and SGLT3. J. Membr. Biol 176, 111–117 (2000). [PubMed: 10926676]
29. Choe S, Rosenberg JM, Abramson J, Wright EM & Grabe M Water permeation through the sodium-dependent galactose cotransporter vSGLT. Biophys. J 99, L56–58 (2010). [PubMed: 20923633]
30. Vandenberg RJ, Handford CA, Campbell EM, Ryan RM & Yool AJ Water and urea permeation pathways of the human excitatory amino acid transporter EAAT1. Biochem. J 439, 333–340 (2011). [PubMed: 21732909]

31. Turk E et al. Molecular characterization of Vibrio parahaemolyticus vSGLT: a model for sodium-coupled sugar cotransporters. J. Biol. Chem 275, 25711–25716 (2000). [PubMed: 10835424]
32. Goehring A et al. Screening and large-scale expression of membrane proteins in mammalian cells for structural studies. Nat. Protoc 9, 2574–2585 (2014). [PubMed: 25299155]
33. Pardon E et al. A general protocol for the generation of Nanobodies for structural biology. Nat. Protoc 9, 674–693 (2014). [PubMed: 24577359]
34. McMahon C et al. Yeast surface display platform for rapid discovery of conformationally selective nanobodies. Nat. Struct. Mol. Biol 25, 289–296 (2018). [PubMed: 29434346]
35. Hediger MA, Ikeda T, Coady M, Gundersen CB & Wright EM Expression of size-selected mRNA encoding the intestinal Na/glucose cotransporter in Xenopus laevis oocytes. Proc. Natl. Acad. Sci. U. S. A 84, 2634–2637 (1987). [PubMed: 3472228]
36. Leung DW, Loo DD, Hirayama BA, Zeuthen T & Wright EM Urea transport by cotransporters. J. Physiol 528 Pt 2, 251–257 (2000). [PubMed: 11034615]
37. Mastronarde DN Automated electron microscope tomography using robust prediction of specimen movements. J. Struct. Biol 152, 36–51 (2005). [PubMed: 16182563]
38. Zheng SQ et al. MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. Nat. Methods 14, 331–332 (2017). [PubMed: 28250466]
39. Zhang K Gctf: Real-time CTF determination and correction. J. Struct. Biol 193, 1–12 (2016). [PubMed: 26592709]
40. Zivanov J et al. New tools for automated high-resolution cryo-EM structure determination in RELION-3. Elife 7 (2018).
41. Heymann JB & Belnap DM Bsoft: image processing and molecular modeling for electron microscopy. J. Struct. Biol 157, 3–18 (2007). [PubMed: 17011211]
42. Waterhouse A et al. SWISS-MODEL: homology modelling of protein structures and complexes. Nucleic Acids Res. 46, W296–W303 (2018). [PubMed: 29788355]
43. Emsley P, Lohkamp B, Scott WG & Cowtan K Features and development of Coot. Acta Crystallogr. D Biol. Crystallogr 66, 486–501 (2010). [PubMed: 20383002]
44. Adams PD et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Crystallogr. D Biol. Crystallogr 66, 213–221 (2010). [PubMed: 20124702]
45. Chen VB et al. MolProbity: all-atom structure validation for macromolecular crystallography. Acta Crystallogr. D Biol. Crystallogr 66, 12–21 (2010). [PubMed: 20057044]
46. The PyMOL Molecular Graphics System v.2.0. (Schrödinger, 2017).
47. Pettersen EF et al. UCSF Chimera--a visualization system for exploratory research and analysis. J. Comput. Chem 25, 1605–1612 (2004). [PubMed: 15264254]
48. Goddard TD et al. UCSF ChimeraX: Meeting modern challenges in visualization and analysis. Protein Sci. 27, 14–25 (2018). [PubMed: 28710774]
49. Laskowski RA & Swindells MB LigPlot+: multiple ligand-protein interaction diagrams for drug discovery. J. Chem. Inf. Model 51, 2778–2786 (2011). [PubMed: 21919503]
50. Lomize MA, Lomize AL, Pogozheva ID & Mosberg HI OPM: orientations of proteins in membranes database. Bioinformatics 22, 623–625 (2006). [PubMed: 16397007]
51. Jacobson MP, Friesner RA, Xiang Z & Honig B On the role of the crystal environment in determining protein side-chain conformations. J. Mol. Biol 320, 597–608 (2002). [PubMed: 12096912]
52. Betz R Dabble (v2.6.3). Zenodo (2017).
53. Huang J et al. CHARMM36m: an improved force field for folded and intrinsically disordered proteins. Nat Methods 14, 71–73 (2017). [PubMed: 27819658]
54. Klauda JB et al. Update of the CHARMM all-atom additive force field for lipids: validation on six lipid types. J. Phys. Chem. B 114, 7830–7843 (2010). [PubMed: 20496934]
55. Guvench O, Hatcher ER, Venable RM, Pastor RW & Mackerell AD CHARMM Additive All-Atom Force Field for Glycosidic Linkages between Hexopyranoses. J. Chem. Theory Comput 5, 2353–2370 (2009). [PubMed: 20161005]
56. Case DA, B.-S. IY, Brozell SR, Cerutti DS, Cheatham TE III, Cruzeiro VWD, Darden TA, Duke RE, Ghoreishi D, Gilson MK, Gohlke H, Goetz AW, Greene D, Harris R, Homeyer N, Huang Y,

Izadi S, Kovalenko A, Kurtzman T, Lee TS, LeGrand S, Li P, Lin C, Liu J, Luchko T, Luo R, Mermelstein DJ, Merz KM, Miao Y, Monard G, Nguyen C, Nguyen H, Omelyan, Onufriev A, Pan F, Qi R, Roe DR, Roitberg A, Sagui C, Schott-Verdugo S, Shen J, Simmerling CL, Smith J, Salomon-Ferrer R, Swails J, Walker RC, Wang J, Wei H, Wolf RM, Wu X, Xiao L, York DM, Kollman PA. AMBER 2018. University of California, San Francisco (2018).

57. Hopkins CW, Le Grand S, Walker RC & Roitberg AE Long-Time-Step Molecular Dynamics through Hydrogen Mass Repartitioning. J. Chem. Theory Comput 11, 1864–1874 (2015). [PubMed: 26574392]
58. Ryckaert Jean-Paul, C. G, Berendsen Herman J.C. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. J. Comput. Phys 23, 327–341 (1977).
59. Roe DR & Cheatham TE 3rd. PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. J. Chem. Theory Comput 9, 3084–3095 (2013). [PubMed: 26583988]
60. Humphrey William, D. A, Schulten Klaus. VMD: Visual molecular dynamics. J. Mol. Graph 14, 33–38 (1996). [PubMed: 8744570]
61. Nguyen CN, Young TK & Gilson MK Grid inhomogeneous solvation theory: hydration structure and thermodynamics of the miniature receptor cucurbit[7]uril. J. Chem. Phys 137, 044101 (2012). [PubMed: 22852591]
62. Ramsey S et al. Solvation thermodynamic mapping of molecular surfaces in AmberTools: GIST. J. Comput. Chem 37, 2029–2037 (2016). [PubMed: 27317094]

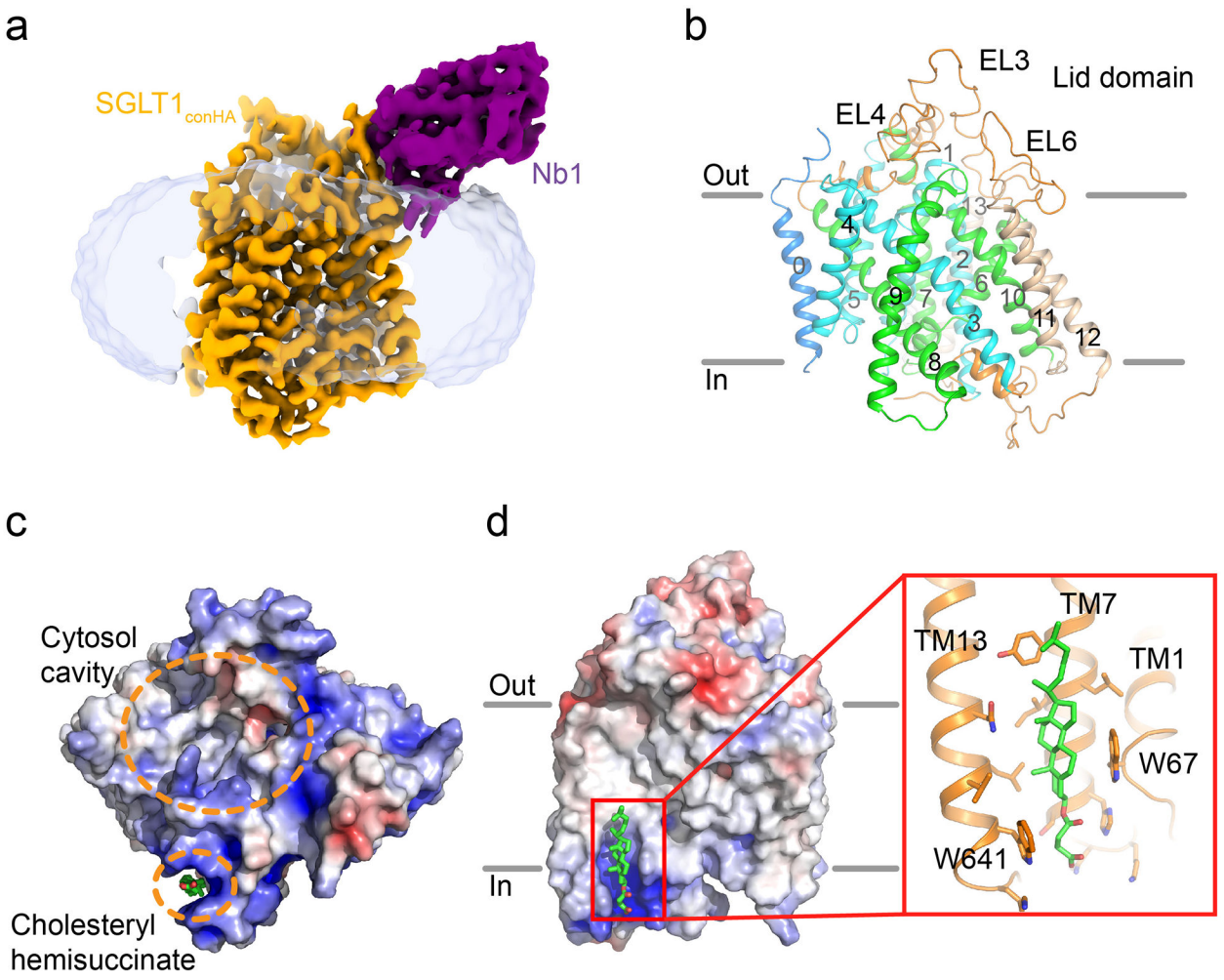


Fig. 1 | Overall structure of SGLT1.

a, Cryo-EM density map of $SGLT1_{conHA}$-nanobody complex [$SGLT1_{conHA}$, orange; nanobody (Nb1), purple]. **b**, Overall $SGLT1_{conHA}$ structure in partial inward-open conformation. Inverted repeat motifs TM1-5 and TM6-10 are colored in cyan and green, respectively, and N-terminal TM0, C-terminal TM11-13, and extra-membranous regions in marine, wheat, and orange. **c**, $SGLT1_{conHA}$ surface viewed from cytosol. Cytosol cavity and cholesterol hemisuccinate are indicated by orange dashed cycles. Surface is colored according to electrostatic potential (red, negative; blue, positive). **d**, Cholesterol hemisuccinate bound at $SGLT1_{conHA}$ surface. Residues in contact with cholesterol hemisuccinate are shown as sticks.

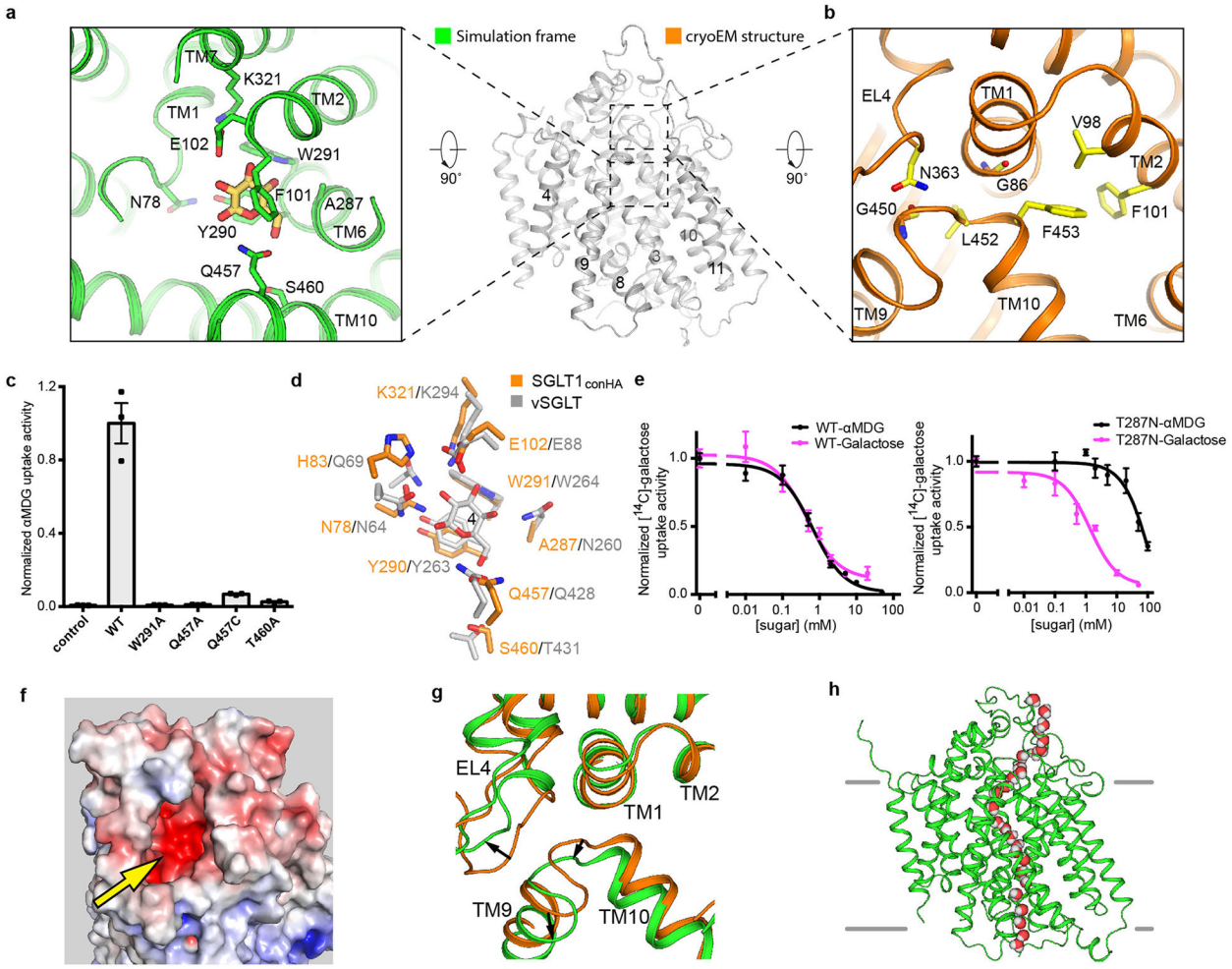


Fig. 2 | SGLT1's substrate binding and water permeation.

a, SGLT1 glucose-binding pocket. Representative simulation frame shows glucose as yellow sticks and residues involved in interactions as green sticks. **b**, SGLT1 extracellular gate. Residues forming extracellular gate are shown as yellow sticks. **c**, αMDG uptake activities of SGLT1 mutants. Activities are normalized to WT. **d**, Superposition of central binding pockets of $SGLT1_{conHA}$ (orange) and vSGLT (gray). Galactose and substrate binding residues are shown as sticks. Galactose's fourth hydroxyl group is indicated by "4". **e**, [^{14}C]-galactose uptake of SGLT1 WT (left) or T287N mutant (right) in the presence of αMDG or galactose at various concentrations. **f**, The negatively charged extracellular vestibule surface of SGLT1 (indicated by yellow arrow). The surface is colored according to electrostatic potential (red, negative; blue, positive). **g**, SGLT1 structure (orange) and representative MD simulation frame (green) showing extracellular-gate movement (black arrows). **h**, Water-permeation pathway, illustrated by a representative simulation frame, with water molecules in red and white. Gray lines represent membrane bilayer boundaries. **c,e,** mean ± SEM; n=3 biological replicates.

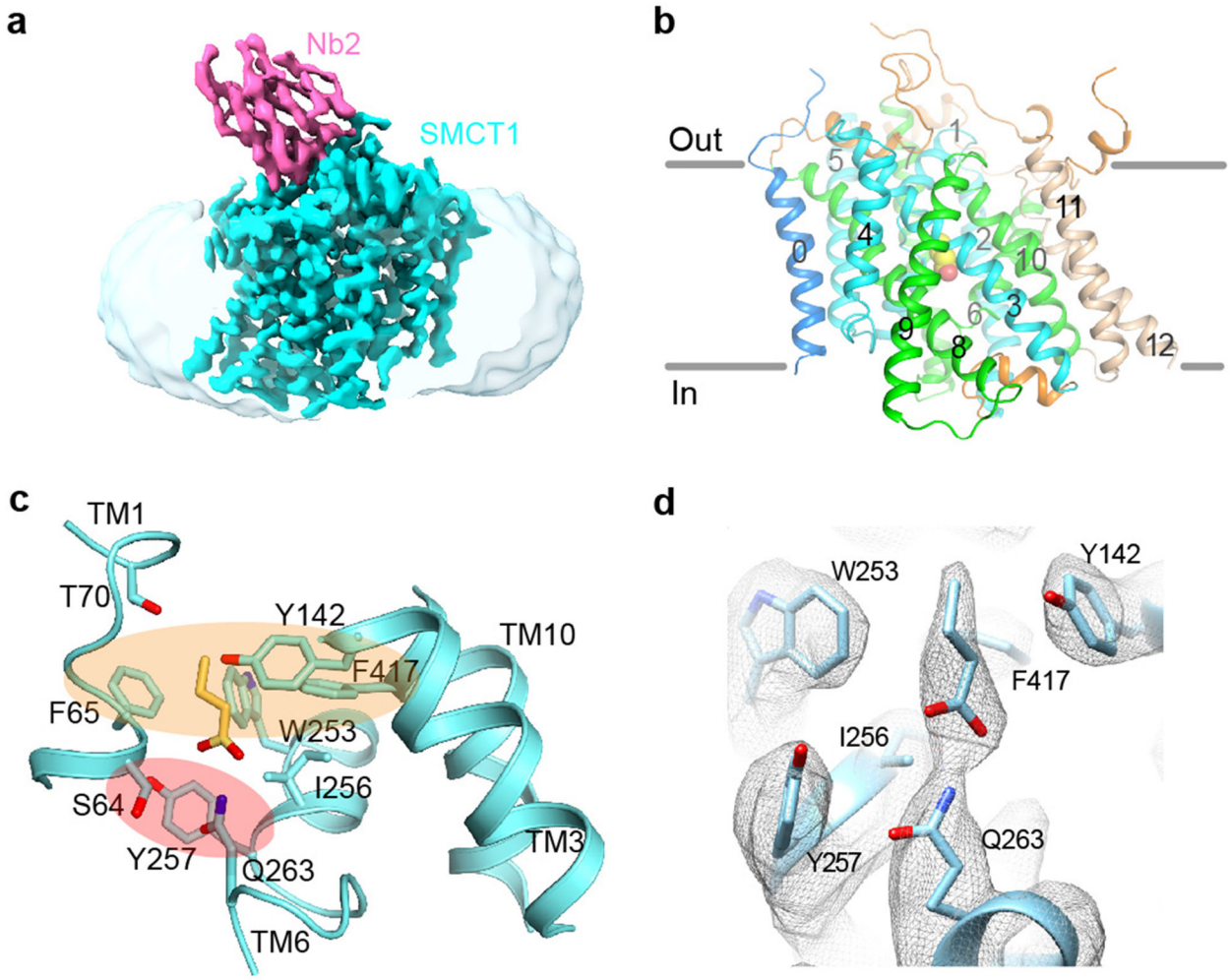


Fig. 3 | Overall structure and substrate-binding pocket of SMCT1.

a, Cryo-EM density map of SMCT1-nanobody complex [SMCT1, cyan; nanobody (Nb2), pink]. **b**, Overall structure of SMCT1. Ribbon representation is colored as in Fig. 1b. Butyrate is shown as spheres. **c**, Substrate-binding pocket. Butyrate and binding site residues are shown as sticks. Amphiphilic composition of binding pocket is indicated by ovals (hydrophobic, orange; hydrophilic, light red). **d**, Butyrate density and surrounding residues.