

# Substrate binding plasticity revealed by Cryo-EM structures of SLC26A2

Received: 28 September 2023

Wenxin Hu <sup>1</sup>, Alex Song<sup>1</sup> & Hongjin Zheng <sup>1</sup> ✉

Accepted: 17 April 2024

Published online: 29 April 2024

 Check for updates

SLC26A2 is a vital solute carrier responsible for transporting essential nutritional ions, including sulfate, within the human body. Pathogenic mutations within SLC26A2 give rise to a spectrum of human diseases, ranging from lethal to mild symptoms. The molecular details regarding the versatile substrate-transporter interactions and the impact of pathogenic mutations on SLC26A2 transporter function remain unclear. Here, using cryo-electron microscopy, we determine three high-resolution structures of SLC26A2 in complexes with different substrates. These structures unveil valuable insights, including the distinct features of the homodimer assembly, the dynamic nature of substrate binding, and the potential ramifications of pathogenic mutations. This structural-functional information regarding SLC26A2 will advance our understanding of cellular sulfate transport mechanisms and provide foundations for future therapeutic development against various human diseases.

Sulfate ion ( $\text{SO}_4^{2-}$ ) is the fourth most abundant anion in human plasma and plays a crucial role in various biological processes<sup>1</sup>. For example, one of its essential functions is as a key component in cellular growth and development through sulfate conjugation<sup>2</sup>. This conjugation is a crucial step for the biotransformation and detoxication of various foreign substances, known as xenobiotics<sup>3</sup>. In addition, through sulfation,  $\text{SO}_4^{2-}$  acts as a regulator of the bioactivity of endogenous molecules like steroids, peptide hormones, and neurotransmitters, influencing their functions and interactions<sup>4</sup>. Moreover,  $\text{SO}_4^{2-}$  is responsible for a post-translational modification called tyrosine sulfation, which occurs in various soluble and membrane proteins<sup>5</sup>. In the human body, the primary source of inorganic  $\text{SO}_4^{2-}$  stems from the active uptake of water and food, which is facilitated by specific solute carriers (SLCs) within the lipid bilayers. These transporters include sodium-dependent SLC13 and sodium-independent SLC26 members<sup>2</sup>. The active transport of  $\text{SO}_4^{2-}$  is essential for maintaining the required sulfate levels in various tissues and organs, contributing to the proper functioning of the body's physiological processes.

The SLC26 family consists of eleven multifunctional anion exchangers and channels that transport a broad range of substrates, including chloride ( $\text{Cl}^-$ ), bicarbonate ( $\text{HCO}_3^-$ ), oxalate ( $\text{C}_2\text{O}_4^{2-}$ ), iodide ( $\text{I}^-$ ), formate ( $\text{HCO}_2^-$ ), and  $\text{SO}_4^{2-}$ <sup>6</sup>. SLC26A2 is highly expressed in many tissues during development and postnatally prominently in

enterocytes and chondrocytes<sup>7,8</sup>. The significance of SLC26 members is evidenced by their association with genetic disorders and clinical conditions<sup>9</sup>. Pathogenic mutations in SLC26A2 result in a spectrum of phenotypes, ranging from severe and lethal diseases, such as achondrogenesis type 1B and atelosteogenesis type 2, to milder disorders like diastrophic dysplasia and multiple epiphyseal dysplasia<sup>10</sup>. Interestingly, the expression of SLC26A2 seems context-dependent and linked to various diseases. For instance, SLC26A2 is upregulated in Crohn's disease<sup>11</sup>. In contrast, SLC26A2 is downregulated in colon cancer as the repression of SLC26A2 in colon cancer cells in vitro increases proliferation<sup>12</sup>. Moreover, a recent study reported that SLC26A2 acts as an unusual mediator of TRAIL resistance, as the loss of SLC26A2 resensitizes cancer cells to TRAIL-mediated cell death<sup>13</sup>. These findings make SLC26A2 an important target for further investigation in understanding and potentially treating related diseases.

SLC26A2 is known to function as an electroneutral anion exchanger, facilitating the exchange of various substrates such as  $\text{SO}_4^{2-}$  and  $\text{C}_2\text{O}_4^{2-}$  for  $\text{Cl}^-$ , hydroxide ( $\text{OH}^-$ ), or even with each other<sup>14,15</sup>. The transport mode is promiscuous, as it is bidirectional and depends on the  $\text{Cl}^-$  and  $\text{OH}^-$  gradient across the membrane<sup>16</sup>. To understand the molecular mechanism of SLC26A2, it is crucial to have high-resolution structural information. However, the available structural information about SLC26 members has primarily focused on human SLC26A5

<sup>1</sup>Department of Biochemistry and Molecular Genetics, University of Colorado Anschutz Medical Campus, School of Medicine, Aurora, US.

✉ e-mail: [hongjin.zheng@cuanschutz.edu](mailto:hongjin.zheng@cuanschutz.edu)

(prestin) and human SLC26A9, along with their close homologs from other species<sup>17–23</sup>.

In this work, using cryo-electron microscopy (cryo-EM) single particle reconstruction, we determine three high-resolution structures of human SLC26A2 in complex with various substrates ( $\text{Cl}^-$ ,  $\text{C}_2\text{O}_4^{2-}$ ,  $\text{SO}_4^{2-}$ ) at 3–3.6 Å resolution. By combining molecular dynamics simulations, mutagenesis, and transport assays, we illustrate the molecular details of how SLC26A2 interacts with substrates that bound to distinct sites. We also discover a unique feature for the dimerization of SLC26A2, which involves a specific stretch of residues in the N-terminal loop (residues 53–61). This unique feature, distinct from the typical  $\beta$ -sheet arrangement in STAS domains of well-known SLC26 members, adds a different dimension to our understanding of its structural dynamics. Furthermore, we map pathogenic mutations identified in patients onto our high-resolution structures and provide logical explanations regarding their potential impact on the transporter function.

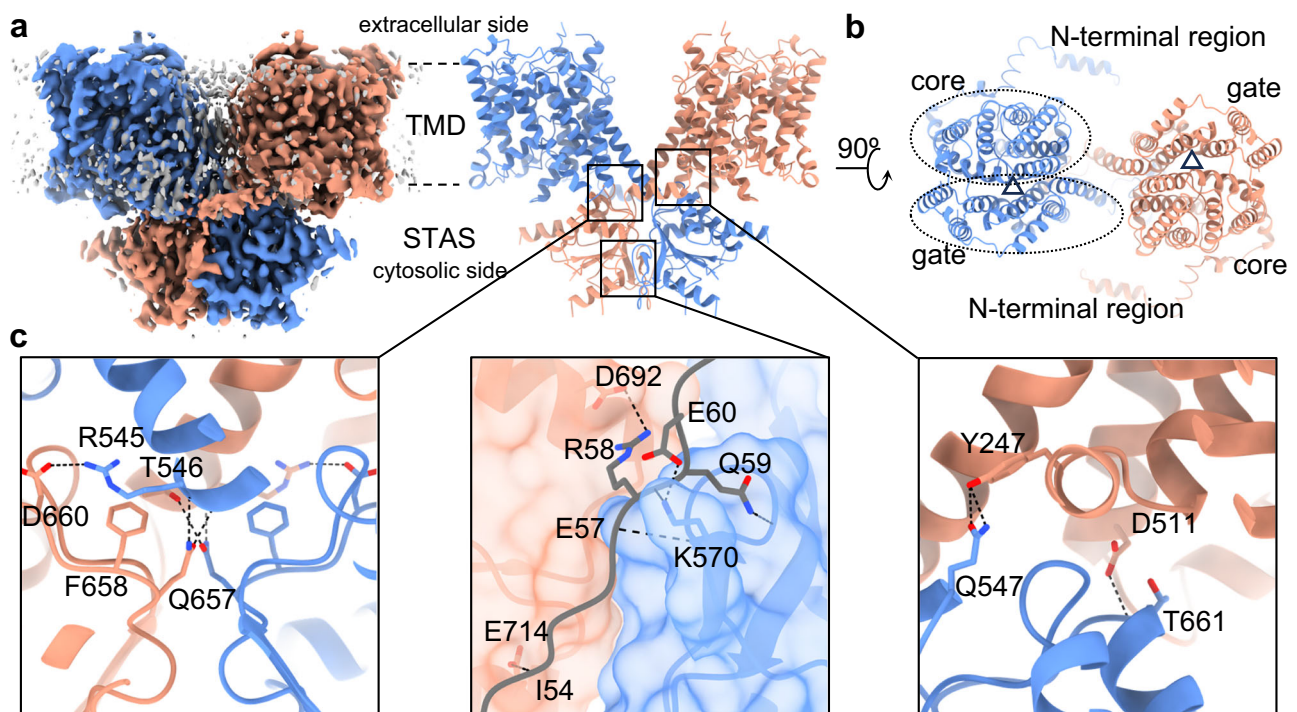
## Results

### The overall structure of SLC26A2

For structural and functional research described in this study, we expressed full-length human SLC26A2 with an N-terminal His-tag in HEK293 cells. We purified the overexpressed SLC26A2 in Lauryl Maltose Neopentyl Glycol (LMNG) detergent with 150 mM NaCl at pH 7.5 and then determined the structure, termed SLC26A2-Cl<sup>-</sup>, by cryo-EM (Fig. 1). The final reconstruction achieved a resolution of ~3.2 Å, with sufficient quality for model building (Supplementary Fig. 1, Supplementary Movie 1). The resulting model covers residues 52–724 without several flexible loops containing residues 186–210, 319–333, and 617–644. Like its homologs<sup>17–24</sup>, SLC26A2 forms a domain-swapped homodimer (Fig. 1a). Each protomer of SLC26A2 can be divided into three regions: the N-terminal region (residues 1–104) comprising a few flexible cytosolic helices and loops (Supplementary Movie 2), the transmembrane domain (TMD, residues 105–540) with 14

transmembrane helices (TMs) (Supplementary Movie 3), and the C-terminal cytosolic sulfate transporter and anti-sigma factor antagonist domain (STAS, residues 541–739) (Supplementary Movie 4). The TMD adopts a typical uracil transporter (UraA) fold, where the first seven TMs invertedly relate to the last seven TMs with a pseudo two-fold symmetry<sup>25</sup>. Furthermore, these 14 TMs are organized into two distinct sub-domains in TMD: the core domain formed by TMs 1–4, 8–11, and the gate domain formed by TMs 5–7, 12–14 connecting to the STAS domain (Fig. 1b). The substrate translocation pathway is located within the interface between the core and gate domains. The two sub-domains undergo an elevator-like rigid-body movement to facilitate substrate transport, enabling the transporter to switch between inward-facing and outward-facing conformations (Supplementary Fig. 2).

In the SLC26A2 dimer, the TMDs are physically separated, with apparent lipid densities observed in between (Supplementary Fig. 3a). Due to the lower local resolution, we could not unambiguously identify the specific lipids. However, we performed thin-layer chromatography (TLC) to confirm the presence of the bound lipids, which are likely phosphatidylcholine (PC), phosphatidylserine (PS), phosphatidylethanolamine (PE), and cholesterol, without precise identification (Supplementary Fig. 3b). Although these lipids likely contribute to the dimer assembly, the primary mediator of dimerization is the cytosolic region of SLC26A2. The buried interface between the STAS domain in one protomer and the TMD in the other protomer spans ~600 Å<sup>2</sup>, as calculated using ChimeraX<sup>26</sup>, and involves strong hydrogen bonds (H-bonds) between specific residues, such as Y247/Q547 and D511/T661, with bond lengths ranging from 2.1 to 3.3 Å (Fig. 1c). In addition, the two STAS domains are attached through hydrophobic interactions, with a buried surface area of ~1100 Å<sup>2</sup>. Specific interactions found include H-bonds between Q657 and T546 and the backbone of R545, H-bond between R545 and D660, as well as the cation- $\pi$  interaction between F658 and R545 (Fig. 1c). Interestingly, in previous studies of SLC26A5 and SLC26A9, a universal while critical dimerization factor



**Fig. 1 | The overall structure of SLC26A2 homodimer.** **a** The front view of the 3 Å resolution map of SLC26A2-Cl<sup>-</sup>. **b** The final model of SLC26A2 is a domain-swapped homodimer. The two protomers are dark salmon and blue. The lipid/detergent belt is gray. The extracellular side is at the top, and the cytosolic side is at the bottom.

The substrate translocation pathways are marked by triangles. **c** The dimerization is mediated by detailed interactions between domains. Specific H-bonds are highlighted. The N-terminal region from the dark salmon protomer is shown in gray for better illustration.

was identified as an anti-parallel  $\beta$ -sheet formed by joining a  $\beta$ -strand from each protomer's N-termini at the bottom of STAS domains<sup>18,20,24</sup>. However, in SLC26A2, such a  $\beta$ -sheet was not observed, suggesting a different dimerization mode within the SLC26 family. In SLC26A2, the N-terminal region plays a vital role in the dimeric assembly through a different mechanism. Specifically, residues 53–61 in the N-terminal loop traverse through the shallow groove formed by the two STAS domains, resulting in a buried interface of  $\sim 600 \text{ \AA}^2$ . Despite this seemingly shallow interaction, it is remarkably stable, with numerous H-bond interactions involving the residues I54, E57, R58, Q59, and E60 (Fig. 1c). To further demonstrate the significance of the N-terminal loop, we introduced truncation mutations  $\Delta 1$ –45 and  $\Delta 1$ –65. The expression level of  $\Delta 1$ –45 remained at approximately 55% compared to the wild-type, while  $\Delta 1$ –65 was nearly undetectable (Supplementary Fig. 4). This suggests that N-termini-mediated dimerization, especially involving residues 53–61, is critical for the stability of SLC26A2.

### Inward-facing conformation

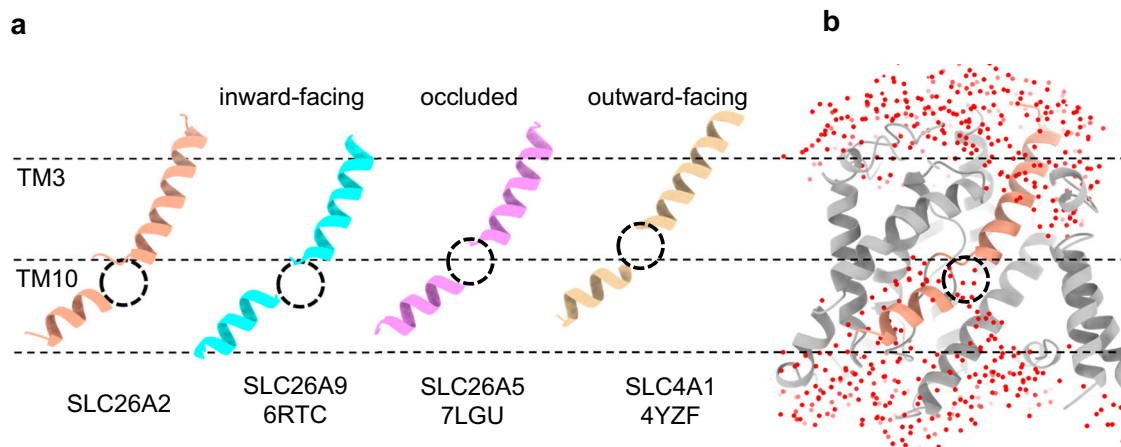
The substrate transport mechanism of SLC26A2 follows the classic “alternating access” model, where the protein alternates between major conformations, including inward-facing, occluded, and outward-facing. In each conformation, the substrate-binding pocket is solvent accessible to only one side of the membrane bilayer<sup>27</sup>. To determine the specific conformation of SLC26A2, we compared our SLC26A2-Cl<sup>-</sup> structure with known homologous structures, namely inward-facing SLC26A9<sup>24</sup>, occluded SLC26A5<sup>20</sup>, and outward-facing SLC4A1<sup>28</sup>. Previous studies have shown that an elevator-like motion between the core and gate domains in TMDs governs the conformational changes in these transporters. This motion is reflected in the vertical position of TM3/TM10 from the core within the lipid bilayer, as it corresponds to the location of the substrate-binding pocket (Supplementary Fig. 2). Here, we aligned the gate domains in all structures and compared their core domains. We found that the vertical position of TM3/TM10 in SLC26A2 closely resembles that of inward-facing SLC26A9 (Fig. 2a & Supplementary Fig. 5a). Using MOLEonline<sup>29</sup>, we calculated the pore size in available SLC homologs (Supplementary Fig. 5b). We considered the pore to be open when its radius is larger than the theoretical radius of a Cl<sup>-</sup> ion without hydration, which is  $\sim 1.8 \text{ \AA}$ . The result shows that our SLC26A2 structure is open to the cytosolic side and closed to the extracellular side. In addition, we performed molecular dynamics simulations using our SLC26A2 models. All simulations, lasting 1  $\mu$ s, showed that the overall structure remained unchanged, and the substrate-binding pocket between TM3/TM10 became filled with water

from the cytosolic side (Fig. 2b). These results collectively suggest that SLC26A2 adopts an inward-facing conformation.

### Multiple substrate-binding sites

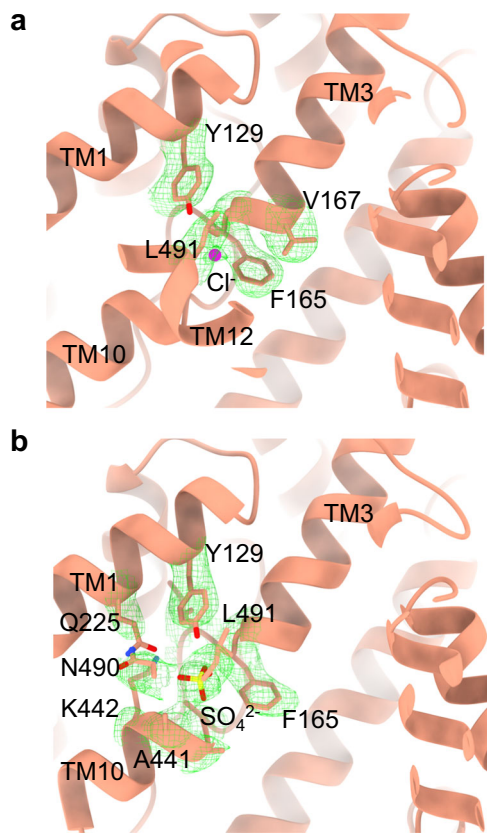
Like other known SLC homologs, the canonical substrate-binding pocket in SLC26A2 is located within the cleft between TM3 and TM10, where the two helices face each other with their N-terminal ends head-to-head (Fig. 2a). Notably, in the structure of SLC26A2-Cl<sup>-</sup>, we identified a distinct density corresponding to the Cl<sup>-</sup> ion in this cleft (Fig. 3a). Interestingly, this Cl<sup>-</sup> does not interact strongly with any surrounding residues. Within 5  $\text{Å}$  distance from the Cl<sup>-</sup>, there are five hydrophobic residues: Y129 from TM1, V167, G166, F165 from the end of TM3, and L491 from TM12. The primary binding force for Cl<sup>-</sup> appears to be the weak TM3 and TM10 helical dipoles with positive charges pointing towards the substrate-binding pocket. To further understand the Cl<sup>-</sup> binding, we performed molecular dynamics simulations, each lasting 1  $\mu$ s. These simulations revealed that the Cl<sup>-</sup> ion frequently moves in and out of the substrate-binding pocket, as evidenced by its distance to the end of TM3 (residue G166) (Supplementary Fig. 6). When such distance is between 3.5 and 8  $\text{Å}$ , the Cl<sup>-</sup> ion is within the TM3/TM10 cleft and considered bound. Combining data from three individual runs, we estimated that the substrate-binding pocket is occupied by Cl<sup>-</sup>  $82 \pm 4\%$  of the time.

To understand how SLC26A2 interacts with its primary substrate SO<sub>4</sub><sup>2-</sup>, we determined the structure of SLC26A2-SO<sub>4</sub><sup>2-</sup>. Specifically, in the last purification step, SLC26A2 was eluted in the buffer containing 50 mM Na<sub>2</sub>SO<sub>4</sub> without NaCl. The final structure of SLC26A2-SO<sub>4</sub><sup>2-</sup> reached an overall resolution of  $\sim 3.5 \text{ \AA}$ , with a clear and high signal-to-noise ratio density of SO<sub>4</sub><sup>2-</sup> in the substrate-binding pocket (Fig. 3b & Supplementary Fig. 7). It is noteworthy that, despite the proximity of Cl<sup>-</sup> to TM3 in SLC26A2-Cl<sup>-</sup>, the SO<sub>4</sub><sup>2-</sup> in SLC26A2-SO<sub>4</sub><sup>2-</sup> is instead closer to TM10, approximately 2–3  $\text{Å}$  away from the Cl<sup>-</sup> binding site. Since SO<sub>4</sub><sup>2-</sup> is still in the TM3/TM10 cleft, the substrate is stabilized by weak helical dipoles just as Cl<sup>-</sup>. However, this apparent shift in the binding site creates a different local environment surrounding the substrates. Within the 5  $\text{Å}$  radius of SO<sub>4</sub><sup>2-</sup>, we found several residues including Y129, Q225 from TM1, F165, I164 from the end of TM3, N490, L491 from TM12, and A439, L440, A441, K442 from TM10. These residues, particularly Y129, K442, Q225, and N490, are likely to interact with SO<sub>4</sub><sup>2-</sup>, providing additional stabilization forces. Indeed, our molecular dynamics simulations on SLC26A2-SO<sub>4</sub><sup>2-</sup> demonstrated that SO<sub>4</sub><sup>2-</sup> remains tightly bound to the binding site  $100 \pm 0\%$  of the time without any significant movements (Supplementary Fig. 6).



**Fig. 2 | Inward-facing SLC26A2.** **a** The comparison of vertical positions of TM3/TM10 in the membrane bilayer from representative homologous structures. **b** During simulation, the substrate-binding pocket in SLC26A2 (slabbed front view)

is filled with water (red ball) from the cytosolic side. Dashed lines indicate the boundaries and center of the lipid bilayer. Dashed circles mark the substrate-binding pockets. SLC26A2 is in gray with TM3/TM10 in dark salmon.



**Fig. 3 | Substrates binding.** **a**  $\text{Cl}^-$  in the binding pocket between TM3 and TM10, the same view as in Fig. 2. **b**  $\text{SO}_4^{2-}$  in the binding pocket.

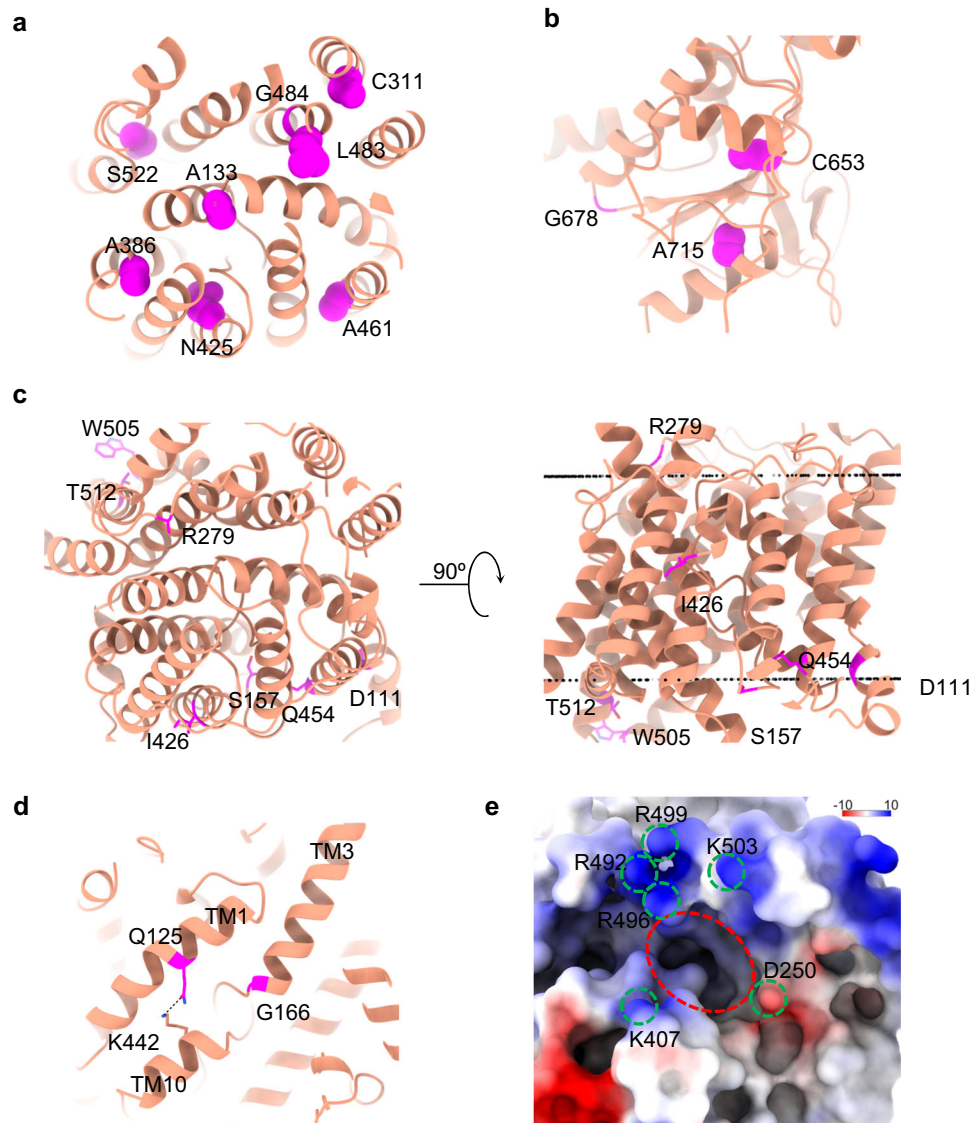
In addition to its known transport of  $\text{SO}_4^{2-}$  and  $\text{Cl}^-$ , SLC26A2 is also capable of  $\text{C}_2\text{O}_4^{2-}$  transport. We determined the structure of SLC26A2- $\text{C}_2\text{O}_4^{2-}$  without NaCl to an overall resolution of  $\sim 3 \text{ \AA}$  (Supplementary Fig. 8). Surprisingly, we found no experimental density for  $\text{C}_2\text{O}_4^{2-}$  in the canonical substrate-binding pocket between TM3 and TM10. Instead, we found weak densities outside the pocket, in the core and gate domains interface between TM5 and TM12, which could accommodate a  $\text{C}_2\text{O}_4^{2-}$  molecule (Supplementary Fig. 9). To assess the agreement between the modeled  $\text{C}_2\text{O}_4^{2-}$  and experimental densities, we calculated the Q-scores using ChimeraX, which should be  $\sim 0.6$  at the reported resolution of  $3\text{--}3.5 \text{ \AA}$ <sup>30</sup>. The result shows that  $\text{C}_2\text{O}_4^{2-}$  scored 0.77, confirming the model's reliability. Our model shows that the  $\text{C}_2\text{O}_4^{2-}$  substrate is likely stabilized via interactions with the side chains of adjacent Ser residues, as well as backbone carbonyls from L253. We mutated the Ser residues and carried out the cellular transport assay. The result shows a  $\sim 40\%$  reduction in transport efficiency, confirming the importance of these Ser residues for  $\text{C}_2\text{O}_4^{2-}$  translocation (Supplementary Fig. 10a). In addition, we performed molecular dynamics simulations to probe the dynamics between SLC26A2 and the  $\text{C}_2\text{O}_4^{2-}$ . The result showed that  $\text{C}_2\text{O}_4^{2-}$  remains bound for  $\sim 65 \pm 4\%$  of the time. Notably, in all simulations, the  $\text{C}_2\text{O}_4^{2-}$  molecule does not enter the exact binding pocket that accommodates  $\text{Cl}^-$  and  $\text{SO}_4^{2-}$  (Supplementary Fig. 6d), implying that the substrates' transport pathways might not be exactly the same.

### Mapping the pathogenic mutations

According to the ClinVar database (accessed August 2023), nearly 200 mutations in the SLC26A2 gene have been documented with a clinical significance of either pathogenic or likely pathogenic. Among these mutations, 25 are missense mutations capable of producing proteins. We successfully mapped all these missense mutations, involving 21 residues, onto our high-resolution structures (Supplementary Table 2).

How do these mutations cause functional problems for SLC26A2? After careful analysis, we could classify them into several categories based on their potential effects. The first category of mutations most likely affects the overall folding of the transporter, resulting in unstable proteins (Fig. 4a). For example, mutations such as A133V, C311R, A386V, A461V, G484D, and S522F in the TMD replace small hydrophobic side chains with larger and even hydrophilic ones, leading to potential disruption in helical packing. For instance, A386V expression in *Xenopus* oocytes was reported to be significantly reduced compared to the wild-type<sup>14</sup>. Moreover, residues like Gly and Pro, found in A386G (in TM8) and L483P (in TM12) mutants, have been known to disrupt  $\alpha$ -helices, contributing to their functional problems<sup>31</sup>. The substitution of N425 with negatively charged Asp, found in the N425D mutant, leads to H-bond loss with T435, T436, and Q233, destabilizing the loop containing T435 and T436 (Supplementary Fig. 11a). All these changes are energetically unfavorable as the whole region is within the membrane bilayer. In the STAS domains (Fig. 4b), mutations like G678V, A715T, C653S, C653Y, and C653G may impact protein functionality. G678 sits at the very short turn between an  $\alpha$ -helix and  $\beta$ -strand, which is apparently required for such an arrangement of secondary structures and should favor Gly over G678V mutation. The side chain of A715 disallows hydrophilic mutations like A715T as it points towards a highly hydrophobic environment formed by residues from the  $\beta$ -sheet and adjacent loop (V650, L682, V712, and F709) (Supplementary Fig. 11b). C653 is at the end of a  $\beta$ -strand, closely surrounded by hydrophobic residues I651, I656, L683, and C686. The side chain of C653 forms a weak H-bond with the backbone of I651 (Supplementary Fig. 11c). C653S is pathogenic because Ser is inherently more hydrophilic than Cys and is harder to tolerate the hydrophobic environment surrounding C653. C653Y is problematic because of the large size of the side chain, while C653G is unfavorable because of the intrinsic destabilizing nature of Gly in the  $\beta$ -sheets<sup>32</sup>. Although Gly destabilization can be 'rescued' by specific cross-strand pairing with aromatic residues, no aromatic residues are present around this region.

The second category of mutations in the SLC26A2 gene most likely impacts protein-lipid interactions (Fig. 4c). For instance, I426 is located at the midpoint of the lipid bilayer, faces toward the surrounding environment, and thus hydrophobically interacts with lipid tails. Mutations like I426T and I426N would be unfavorable because the side chains are much more hydrophilic. Mutations like D111Y, S157P, R279W, Q454P, W505R, and T512K have reversed charge properties compared to the wild-type, which could affect their interactions with the lipid headgroups because these residues are all predicted to be close to the lipid boundaries by the PPM3 server based on our structures<sup>33</sup>. Among these mutations, R279W is the most common pathogenic variant found outside of Finland<sup>34</sup>. In *Xenopus* oocytes, R279W shows a similar surface abundance as the wild-type but exhibits a reduced substrate transport rate, suggesting it is a partially dysfunctional mutant<sup>14</sup>. In contrast, another study using HEK293 cells demonstrated that R279W is expressed  $\sim 50\%$  less on the cell surface, and these cells have  $\sim 50\%$  reduced rates of sulfate transport compared to cells expressing the wild-type, suggesting that the transport efficiency of R279W mutant after normalization is roughly at the same level of wild-type SLC26A2<sup>8</sup>. In our structural model, R279 is located on an extracellular loop away from the potential substrate translocation path, making it difficult to impact the substrate translocation directly. Furthermore, in our cellular transport assay, we observed a  $\sim 50\%$  reduced transport and  $\sim 55\%$  reduced expression level of the R279W mutant compared to the wild-type, which translates to indistinguishable transport efficiency after normalization for the mutant protein compared to the wild-type protein (Supplementary Fig. 10b). This finding aligns with the latter study, suggesting that the R279W mutant may retain its normal substrate transport function. Instead, the mutant may lose its natural lipid anchoring ability, making it harder to insert



**Fig. 4 | Pathogenic mutations mapped on the SLC26A2 model. a** Residues in TMDs that affect the protein stability. **b** Residues in soluble STAS domains that affect the protein stability. **c** Residues that may be critical for lipid interactions. The dashed black line represents the predicted boundary of the lipid bilayer. **d** G166 and

Q125 around the TM3/TM10 cleft. **e** Electrostatic surface representation of the TMD domain viewing from the cytosolic side. The dashed red circle represents the putative transport pathway, while the green ones mark charged residues.

into the membrane correctly. This results in a reduced expression level and, thus, impaired substrate translocation.

The third type of mutation directly affects the interactions between substrates and the transporter, as they are around the substrate-binding pocket. For instance, Q125 is located in TM1, seemingly away from the TM3/TM10 cleft (Fig. 4d). However, it forms an essential H-bond with K442 in TM10, the only strong interaction stabilizing the K442 side chain. We hypothesized that the Q125L mutation would release the K442 side chain, allowing it to insert into the TM3/TM10 cleft easily. Similarly, G166 defines the edge of the substrate-binding pocket at the N-terminal end of TM3, and the G166R mutation would place its side chain into the substrate-binding cleft. Due to the positive charge of K442 and G166R, these residues may interact strongly with negatively charged substrates. Such strong interactions could impede substrate release and translocation, creating unfavorable conditions for the transport process. Such a hypothesis was tested by cellular transport assay using the following single mutations: Q125L, G166A, and G166R (Supplementary Fig. 6). The result shows that Q125L and G166R retain approximately

40–70% of sulfate translocation, while G166A has the same transport efficiency as the wild-type.

Notably, D250 is situated away from any lipids and is located at the end of TM5, near the cytosolic entry of the substrate translocation pathway (Fig. 4e). We analyzed the electrostatic distribution of the surface around the area using ChimeraX. As expected, the cytosolic entry is surrounded by positively charged residues, including K407, R492, R496, R499, and K503. This positive charge distribution seems essential to accommodate negatively charged substrates under neutral pH conditions. Interestingly, D250 is the only negatively charged residue in this area and the only one associated with pathogenicity identified so far (D250V). Our cellular transport assay shows that D250V has almost the same expression level and translocation ability as wild-type (Supplementary Fig. 10b). Thus, why D250V is pathogenic could not be explained in this study and needs further investigation.

## Discussion

In this study, we have determined three distinct structures of SLC26A2 transporter in complex with substrates  $\text{SO}_4^{2-}$ ,  $\text{Cl}^-$ , and  $\text{C}_2\text{O}_4^{2-}$ . Despite

adopting a similar inward-facing conformation, these structures showcase unique binding modes for different molecules. Prior studies have already established that the primary role of SLC26A2 is to transport  $\text{SO}_4^{2-}$ , even though it can effectively transport a variety of anions<sup>35,36</sup>. Our molecular dynamic simulations further underscore this idea. Among the small molecules studied,  $\text{SO}_4^{2-}$  is the only one consistently staying in its original binding site, indicating a relatively more robust substrate-transporter interaction (Supplementary Fig. 6). By analyzing the total time substrates spend in their binding site during simulations, we notice an order of binding affinities from high to low:  $\text{SO}_4^{2-} > \text{Cl}^- > \text{C}_2\text{O}_4^{2-}$ . This order aligns with the documented efflux rate:  $\text{SO}_4^{2-} \sim \text{Cl}^- > \text{C}_2\text{O}_4^{2-}$ , measured under consistent extracellular conditions in oocytes<sup>14</sup>.

It is well established that the anion exchange mediated by SLC26A2 is electroneutral<sup>14,15</sup>. However, the ongoing debate centers on whether the exchange mechanism operates as  $\text{SO}_4^{2-}/2\text{Cl}^-$  or  $\text{SO}_4^{2-}/\text{OH}^-/\text{Cl}^-$ . Do our structural studies shed light on this topic? First, we do not consider the exchange mode involving proton, such as  $\text{SO}_4^{2-}/\text{H}^+/\text{Cl}^-$ , because our structures show no protonatable residues (Glu and Asp) near the TM3/TM10 cleft region nor the potential translocation pathway. Second, in the context of extracellular  $\text{SO}_4^{2-}$  exchange with intracellular  $\text{Cl}^-$ , the stoichiometry of one  $\text{SO}_4^{2-}$  exchanging with two  $\text{Cl}^-$  would require simultaneous binding of two  $\text{Cl}^-$  ions in the TM3/TM10 cleft. This scenario appears highly unlikely for the following reasons: a) only a single  $\text{Cl}^-$  density was observed in the SLC26A2- $\text{Cl}^-$  structure; b) in the SLC26A2- $\text{Cl}^-$  simulations performed in the presence of 100 mM KCl, we observed only 0 or 1 ions binding in the cleft on the timescale of the trajectories. Thus, to remain electroneutral, the most plausible scenario should involve co-transporting one  $\text{OH}^-$  and one  $\text{Cl}^-$  for exchanging a single  $\text{SO}_4^{2-}$ . Given the resolution of our structures, it is challenging to discern specific densities for individual  $\text{OH}^-$ . However, hydrophilic residues in the binding pocket, such as Y129, K442, Q225, and N490, could likely play a role in  $\text{OH}^-$  translocation.

Upon comparing all three structures, we noticed that the size of the substrate-binding pocket within the TM3/TM10 cleft slightly changes in response to different substrates. To quantify that, we measured the distance between TM3 and TM10 using C $\alpha$  positions of three pairs of residues: F165/K442, G166/A441, and V167/L440 (Supplementary Fig. 12a). The analysis reveals that the cleft size is at its smallest when bound to  $\text{SO}_4^{2-}$ , intermediate when bound to  $\text{Cl}^-$ , and largest when  $\text{C}_2\text{O}_4^{2-}$ . However, the variation results in only a maximum change of approximately 1–2 Å in diameter. Recent studies on SLC26A5 have reported similar size changes within the TM3/TM10 region in response to the binding of small molecules like  $\text{SO}_4^{2-}$ ,  $\text{Cl}^-$ , or the inhibitor salicylate. The authors of those studies propose that these changes correlate with a larger-scale global conformational shift between expanded and contracted states<sup>19,20</sup>. Interestingly, the cleft change observed in SLC26A5 is at the same scale as observed in our SLC26A2 structures. It is also important to note that the distance measurements at the reported resolution (3–3.5 Å) are typically associated with a margin of error, which could be as large as 0.5 Å<sup>37</sup>. Furthermore, all our SLC26A2 structures align remarkably well with each other with an RMSD of ~0.8 Å (Supplementary Fig. 12b). Taken together; we believe such minute changes in the pocket size likely have no significant impact on SLC26A2. Such a disparity may arise from the dual nature of SLC26A5 as both a voltage-sensing motor and a transporter, in contrast to SLC26A2's sole transporter function<sup>38</sup>. After all, the precise mechanism governing the transition from the inward-facing to the occluded and outward-facing conformation of SLC26A2 members remains a topic for future research.

Here, with the high-resolution structures and related MD simulations, we could shed some light on the  $\text{SO}_4^{2-}/\text{Cl}^-$  exchange mechanism of SLC26A2. First,  $\text{SO}_4^{2-}$  and  $\text{Cl}^-$  ions bind to the same substrate-binding pocket within the TM3/TM10 cleft, though their interactions with surrounding residues are not the same. This phenomenon is not

unprecedented. For example, in the nitrate/nitrite exchanger NarK, both substrates occupy essentially the same pocket formed in between two Arg residues, R305 and R89, with R89 playing a more pivotal role in the binding of nitrite over nitrate<sup>39,40</sup>. Similarly, in the electroneutral  $\text{Na}^+/\text{H}^+$  antiporter NhaP, both  $\text{Na}^+$  and  $\text{H}^+$  bind to the same pocket around D159 but with distinct interactions with surrounding residues<sup>41</sup>. Second, the SLC26A2- $\text{SO}_4^{2-}$  structure in the inward-facing conformation most likely represents the state post-import of extracellular  $\text{SO}_4^{2-}$ . Thus, the stabilization of  $\text{SO}_4^{2-}$  in the binding pocket in all independent simulations suggests a mechanism that prevents premature leakage. The question then arises as to how the imported  $\text{SO}_4^{2-}$  is released from the substrate-binding pocket. Comparing the simulations of SLC26A2- $\text{Cl}^-$  and SLC26A2- $\text{SO}_4^{2-}$ , we found that  $\text{Cl}^-$  ion can intersect the  $\text{SO}_4^{2-}$  binding site when within 6–8 Å of G166, implying a possible competition between the two substrates. We previously noted that  $\text{SO}_4^{2-}$  appears to have a stronger binding affinity, given its longer residency within the binding pocket. Nevertheless, considering that transporter turnover rates are typically several to several hundred per second<sup>42,43</sup>, the 1  $\mu\text{s}$  duration of our simulations is likely too short and thus biased. We hypothesize that, over extended timeframes, intracellular  $\text{Cl}^-$  may competitively bind to the TM3/TM10 cleft, facilitating the release of the imported  $\text{SO}_4^{2-}$ . Third, our structures revealed that while  $\text{SO}_4^{2-}$  and  $\text{Cl}^-$  share a binding pocket, the  $\text{C}_2\text{O}_4^{2-}$  ion binds to a distinct site and does not interact with the  $\text{SO}_4^{2-}/\text{Cl}^-$  pocket during all simulations, implying alternative translocation pathways for different substrates. Such a phenomenon has been documented before. For example, in *Vibrio cholerae*, the multidrug and toxic compound extrusion transporter NorM uses two distinct ion-translocation pathways for  $\text{Na}^+$  and  $\text{H}^+$  transport<sup>44</sup>. Our finding suggests that the molecular mechanism governing  $\text{SO}_4^{2-}/\text{Cl}^-$  exchange may differ from that of the  $\text{C}_2\text{O}_4^{2-}/\text{Cl}^-$  exchange. To delineate the mechanistic details, future work will require additional structural determination of SLC26A2 in alternative conformations alongside more sophisticated MD simulations.

## Methods

### Protein expression and purification

The full-length human SLC26A2 was cloned into a pEZT-BM (Addgene) vector with an N-terminal His-tag and a thrombin digestion site. The resulting construct was overexpressed in Expi293F cells (ThermoFisher Scientific) using the BacMam Expression System. After 72 h of virus infection, the cells were harvested through centrifugation and lysed by being passed three times through a microfluidizer M110P (Microfluidics Corporation). The membrane fraction was collected by ultracentrifugation at 150,000 g for 1 h and resuspended in buffer A (20 mM HEPES pH 7.5 and 150 mM NaCl). To solubilize SLC26A2, the membrane was incubated with 1% lauryl maltose neopentyl glycol (LMNG) (Anatrace) at 4 °C for 2 hrs. The supernatant containing the target protein was isolated by ultracentrifugation at 1,50,000 g for 1 hr and then incubated with TALON IMAC resin (Clontech) in buffer A with 5 mM imidazole. After washing the resin with buffer B (20 mM HEPES pH 7.5, 150 mM NaCl, 0.003% LMNG) containing 10 mM imidazole, SLC26A2 was eluted in buffer B with 200 mM imidazole. The eluted protein was further digested with thrombin (Enzyme Research Laboratories) at a molar ratio 1:50 overnight at 4 °C. Subsequently, SLC26A2- $\text{Cl}^-$  was purified by gel filtration chromatography with a Superose 6 column (Sigma-Aldrich) in buffer B. To prepare SLC26A2- $\text{C}_2\text{O}_4^{2-}$ , NaCl in SLC26A2- $\text{Cl}^-$  was exchanged with 50 mM  $\text{K}_2\text{C}_2\text{O}_4$ . To prepare SLC26A2- $\text{SO}_4^{2-}$ , NaCl in SLC26A2- $\text{Cl}^-$  was exchanged with 50 mM  $\text{Na}_2\text{SO}_4$ .

### Cryo-EM data collection

The freshly purified SLC26A2 samples were concentrated to ~3.5 mg/ml. 3  $\mu\text{l}$  of each sample was applied to a plasma-cleaned UltraAuFoil gold grid (R 1.2/1.3, 300 mesh, Electron Microscopy Sciences) and prepared using a Vitrobot Mark IV (Thermo Fisher Scientific) with the

environmental chamber set at 100% humidity and 4 °C. The grid was blotted for -1.5 s and then flash-frozen in liquid ethane. The data was collected on a Titan Krios (Thermo Fisher Scientific) operated at 300 keV and equipped with a K3 direct detector (Gatan) in the Pacific Northwest Center for Cryo-EM (PNCC). The detailed parameters for data collection are shown in Supplementary Table 1.

### Image processing, model building, and model refinement

All cryo-EM data were processed using cryoSPARC v4<sup>45</sup>. Specifically, Patch Motion Correction was used to correct the beam-induced movement, and Patch CTF was used to estimate contrast transfer function parameters for each movie. Particles were picked with the help of Topaz and extracted with a box size of 320 × 320 pixels<sup>46</sup>. After several rounds of reference-free 2D classification, ab initio reconstruction, and heterogeneous refinement, the best particles were selected for further processing. The final reconstructions were obtained with 100k - 150k particles at the resolution of 3.2 - 3.6 Å. Model building of SLC26A2 was performed in Coot<sup>47</sup>. AlphaFold predicted model of SLC26A2, together with several homologous structures available in the Protein Data Bank (PDB), were used as a guide<sup>48</sup>. The final models were refined in PHENIX<sup>49</sup>. The quality of the models was assessed using MolProbity<sup>50</sup>. Statistical details are provided in Supplementary Table 1. All figures, movies, and charts were prepared using UCSF ChimeraX<sup>26</sup> and Microsoft Excel.

### Molecular dynamics simulations

The missing loops and side chains in all SLC26A2 models (residues 52–724) were built using the Molefactory plugin in VMD<sup>51</sup>. The protonation states of titratable residues were assumed to be those at pH7 by PROPKA<sup>52</sup>. The coordinates of the bound substrates, including Cl<sup>-</sup>, SO<sub>4</sub><sup>2-</sup>, and C<sub>2</sub>O<sub>4</sub><sup>2-</sup>, were based on the modeled ions in our cryo-EM structures. Molecular models for all MD simulations were further prepared using the online tool CHARMM-GUI<sup>53</sup>. Specifically, the proteins with substrates were embedded in a lipid bilayer made of POPC, whose orientation and position were calculated by the Orientations of Proteins in Membranes (OPM) server<sup>54</sup>. Parameters with all substrates were generated by the CHARMM general force field (CGenFF)<sup>55,56</sup>. The systems were solvated with TIP3P water, including 100 mM KCl, resulting in a periodic box dimension of -140 × 140 × 160 Å<sup>3</sup>, containing ~250,000 atoms.

All simulations were performed using GROMACS<sup>57</sup>. The force field parameters of the protein and lipids were CHARMM36m and CHARMM36, respectively<sup>58,59</sup>. The grid information for PME (Particle-Mesh Ewald) electrostatics was generated automatically. A constant number of particles (N), pressure (P) of 1 bar, and temperature of 310 K were used for the NPT ensemble. A 2 fs timestep was used. All simulations were carried out with the following steps: 1) 5000 steps of energy-minimization; 2) 20 ns of equilibration with gradually decreasing positional restraints; 3) 1 μs of production MD simulation with three repetitions starting from the same equilibrated systems after the Step 2.

### Thin-layer chromatography

SLC26A2 samples were subjected to two rounds of purification using the Superose 6 column to remove non-specifically associated lipids. The TLC experiment was performed following a previously described method<sup>60</sup>. Specifically, the lipids tightly associated with SLC26A2 were extracted using a mixed solution of chloroform and methanol (2:1). A TLC plate (Millipore Sigma) was washed with ethanol, air dried, and then activated at 100 °C for 30 min. SLC26A2 samples and lipid standards were dotted on the TLC plate. The experiment was performed with a first mobile phase of chloroform-methanol-water (65:25:4, v/v/v) till the solvent front was halfway through the plate, followed by a second mobile phase of hexane-acetone (100:1, v/v) till the solvent front reached the top. After the run, the TLC plate was air dried,

stained with 0.03% (w/v) G-250 in 20% methanol for 15 min, destained in 20% methanol for 10 min, and then air dried again.

### Cellular transport assay

Expi293F cells expressing wild-type and various mutations of SLC26A2 were used for the sulfate uptake experiments. Specifically, Expi293F cells were infected with corresponding viruses for 24 h to allow protein expression.  $0.2 \times 10^6$  cells were washed twice with Wash Buffer (240 mM mannitol, 2.5 mM potassium sulfate, 2.8 mM calcium gluconate, 1.2 mM magnesium sulfate, and 10 mM HEPES pH 7.5) and then incubated with 400 μl of transport buffer (Wash Buffer plus 3 μCi <sup>35</sup>S-sulfate) for 10 min at room temperature. The reaction was stopped by adding 400 μl of ice-cold wash buffer, quickly followed by filtration using pre-wetted cellulose nitrate filters. The filters were washed with 5 ml of wash buffer, dried for 1 h, and dissolved in 4 ml of Filter Count scintillation liquid (Perkin Elmer). The <sup>35</sup>S radioactivity within the cells was quantified by using a Tri-carb 2910 TR Scintillation counter (Perkin Elmer).

### Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

### Data availability

Pathogenic mutations in the *SLC26A2* gene are accessible from the ClinVar database. The cryo-EM maps of SLC26A2 were deposited in the Electron Microscopy Data Bank under the accession codes [EMD-41427](#), [EMD-41428](#), and [EMD-41429](#). Their corresponding coordinates of the atomic model were deposited in the Protein Data Bank under the accession codes [8TNW](#), [8TNX](#), and [8TNY](#), respectively. All other data are available from the corresponding author upon request. Source data are provided in this paper.

### References

1. Cole, D. E. & Evrovski, J. The clinical chemistry of inorganic sulfate. *Crit. Rev. Clin. Lab Sci.* **37**, 299–344 (2000).
2. Markovich, D. Physiological roles and regulation of mammalian sulfate transporters. *Physiol. Rev.* **81**, 1499–1533 (2001).
3. Mulder, G. J. & ProQuest. *Conjugation Reactions in Drug Metabolism: An Integrated Approach: Substrates, Co-substrates, Enzymes and Their Interactions In Vivo and In Vitro.* (Taylor & Francis, 1990).
4. Strott, C. A. Sulfonation and molecular action. *Endocr. Rev.* **23**, 703–732 (2002).
5. Woods, A. S., Wang, H. Y. & Jackson, S. N. Sulfation, the up-and-coming post-translational modification: its role and mechanism in protein-protein interaction. *J. Proteome Res.* **6**, 1176–1182 (2007).
6. Alper, S. L. & Sharma, A. K. The SLC26 gene family of anion transporters and channels. *Mol. Asp. Med.* **34**, 494–515 (2013).
7. Haila, S. et al. SLC26A2 (diastrophic dysplasia sulfate transporter) is expressed in developing and mature cartilage but also in other tissues and cell types. *J. Histochem Cytochem.* **49**, 973–982 (2001).
8. Karniski, L. P. Functional expression and cellular distribution of diastrophic dysplasia sulfate transporter (DTDST) gene mutations in HEK cells. *Hum. Mol. Genet.* **13**, 2165–2171 (2004).
9. Kere, J. Overview of the SLC26 family and associated diseases. *Novartis Found. Symp.* **273**, 2–11 (2006).
10. Silveira, C. et al. SLC26A2/DTDST Spectrum: a cohort of 12 patients associated with a comprehensive review of the genotype-phenotype correlation. *Mol. Syndromol.* **13**, 485–495 (2023).
11. Comelli, E. M. et al. Biomarkers of human gastrointestinal tract regions. *Mamm. Genome* **20**, 516–527 (2009).
12. Yusa, A., Miyazaki, K., Kimura, N., Izawa, M. & Kannagi, R. Epigenetic silencing of the sulfate transporter gene DTDST induces sialyl Lewisx expression and accelerates proliferation of colon cancer cells. *Cancer Res.* **70**, 4064–4073 (2010).

13. Dimberg, L. Y. et al. A Genome-Wide Loss-of-Function Screen Identifies SLC26A2 as a Novel Mediator of TRAIL Resistance. *Mol. Cancer Res* **15**, 382–394 (2017).
14. Heneghan, J. F. et al. Regulated transport of sulfate and oxalate by SLC26A2/DTDST. *Am. J. Physiol. Cell Physiol.* **298**, C1363–C1375 (2010).
15. Ohana, E., Shcheynikov, N., Park, M. & Muallem, S. Solute carrier family 26 member a2 (Slc26a2) protein functions as an electro-neutral SO<sub>4</sub><sup>2-</sup>/OH<sup>-</sup>/Cl<sup>-</sup> exchanger regulated by extracellular Cl<sup>-</sup>. *J. Biol. Chem.* **287**, 5122–5132 (2012).
16. Seidler, U. & Nikolovska, K. Slc26 family of anion transporters in the gastrointestinal tract: expression, function, regulation, and role in disease. *Compr. Physiol.* **9**, 839–872 (2019).
17. Geertsma, E. R. et al. Structure of a prokaryotic fumarate transporter reveals the architecture of the SLC26 family. *Nat. Struct. Mol. Biol.* **22**, 803–808 (2015).
18. Chi, X. et al. Structural insights into the gating mechanism of human SLC26A9 mediated by its C-terminal sequence. *Cell Discov.* **6**, 55 (2020).
19. Bavi, N. et al. The conformational cycle of prestin underlies outer-hair cell electromotility. *Nature* **600**, 553–558 (2021).
20. Ge, J. et al. Molecular mechanism of prestin electromotive signal amplification. *Cell* **184**, 4669–4679 (2021).
21. Wang, L., Chen, K. & Zhou, M. Structure and function of an Arabidopsis thaliana sulfate transporter. *Nat. Commun.* **12**, 4455 (2021).
22. Butan, C. et al. Single particle cryo-EM structure of the outer hair cell motor protein prestin. *Nat. Commun.* **13**, 290 (2022).
23. Futamata, H. et al. Cryo-EM structures of thermostabilized prestin provide mechanistic insights underlying outer hair cell electromotility. *Nat. Commun.* **13**, 6208 (2022).
24. Walter, J. D., Sawicka, M. & Dutzler, R. Cryo-EM structures and functional characterization of murine Slc26a9 reveal mechanism of uncoupled chloride transport. *Elife* **8**. <https://doi.org/10.7554/eLife.46986> (2019).
25. Lu, F. et al. Structure and mechanism of the uracil transporter UraA. *Nature* **472**, 243–246 (2011).
26. Pettersen, E. F. et al. UCSF ChimeraX: structure visualization for researchers, educators, and developers. *Protein Sci.* **30**, 70–82 (2021).
27. Drew, D. & Boudker, O. Shared molecular mechanisms of membrane transporters. *Annu Rev. Biochem.* **85**, 543–572 (2016).
28. Arakawa, T. et al. Crystal structure of the anion exchanger domain of human erythrocyte band 3. *Science* **350**, 680–684 (2015).
29. Sehnal, D. et al. MOLE 2.0: advanced approach for analysis of biomacromolecular channels. *J. Cheminform* **5**, 39 (2013).
30. Pintilie, G. et al. Measurement of atom resolvability in cryo-EM maps with Q-scores. *Nat. Methods* **17**, 328–334 (2020).
31. Imai, K. & Mitaku, S. Mechanisms of secondary structure breakers in soluble proteins. *Biophysics (Nagoya-shi)* **1**, 55–65 (2005).
32. Merkel, J. S. & Regan, L. Aromatic rescue of glycine in beta sheets. *Fold. Des.* **3**, 449–455 (1998).
33. Lomize, A. L., Todd, S. C. & Pogozheva, I. D. Spatial arrangement of proteins in planar and curved membranes by PPM 3.0. *Protein Sci.* **31**, 209–220 (2022).
34. Barbosa, M. et al. Clinical and molecular characterization of Diastrophic Dysplasia in the Portuguese population. *Clin. Genet.* **80**, 550–557 (2011).
35. Forlino, A. et al. A diastrophic dysplasia sulfate transporter (SLC26A2) mutant mouse: morphological and biochemical characterization of the resulting chondrodysplasia phenotype. *Hum. Mol. Genet.* **14**, 859–871 (2005).
36. Ballhausen, D. et al. Recessive multiple epiphyseal dysplasia (rMED): phenotype delineation in eighteen homozygotes for DTDST mutation R279W. *J. Med Genet* **40**, 65–71 (2003).
37. Ringe, D. & Petsko, G. A. in *Protein Engineering and Design* (ed Paul R. Carey) 205–229 (Academic Press, 1996).
38. Bai, J. P. et al. Prestin’s anion transport and voltage-sensing capabilities are independent. *Biophys. J.* **96**, 3179–3186 (2009).
39. Chon, N. L., Schultz, N. J., Zheng, H. & Lin, H. Anion pathways in the NarK Nitrate/Nitrite Exchanger. *J. Chem. Inf. Model* **63**, 5142–5152 (2023).
40. Zheng, H., Wisedchaisri, G. & Gonen, T. Crystal structure of a nitrate/nitrite exchanger. *Nature* **497**, 647–651 (2013).
41. Okazaki, K. I. et al. Mechanism of the electroneutral sodium/proton antiporter PaNhaP from transition-path shooting. *Nat. Commun.* **10**, 1742 (2019).
42. Zhang, X. & Wright, S. H. Transport turnover rates for human OCT2 and MATE1 expressed in Chinese hamster ovary cells. *Int. J. Mol. Sci.* **23**. <https://doi.org/10.3390/ijms23031472> (2022).
43. Gonzales, A. L. et al. Turnover rate of the gamma-aminobutyric acid transporter GAT1. *J. Membr. Biol.* **220**, 33–51 (2007).
44. Raturi, S. et al. Engineered MATE multidrug transporters reveal two functionally distinct ion-coupling pathways in NorM from *Vibrio cholerae*. *Commun. Biol.* **4**, 558 (2021).
45. Punjani, A., Rubinstein, J. L., Fleet, D. J. & Brubaker, M. A. cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nat. Methods* **14**, 290–296 (2017).
46. Bepler, T. et al. Positive-unlabeled convolutional neural networks for particle picking in cryo-electron micrographs. *Nat. Methods* **16**, 1153–1160 (2019).
47. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of coot. *Acta Crystallogr D. Biol. Crystallogr.* **66**, 486–501 (2010).
48. Jumper, J. et al. Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).
49. Liebschner, D. et al. Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix. *Acta Crystallogr D. Struct. Biol.* **75**, 861–877 (2019).
50. Chen, V. B. et al. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr D. Biol. Crystallogr.* **66**, 12–21 (2010).
51. Humphrey, W., Dalke, A. & Schulten, K. VMD: visual molecular dynamics. *J. Mol. Graph* **14**, 33–38 (1996).
52. Olsson, M. H., Sondergaard, C. R., Rostkowski, M. & Jensen, J. H. PROPKA3: consistent treatment of internal and surface residues in empirical pKa predictions. *J. Chem. Theory Comput.* **7**, 525–537 (2011).
53. Jo, S., Kim, T., Iyer, V. G. & Im, W. CHARMM-GUI: a web-based graphical user interface for CHARMM. *J. Comput Chem.* **29**, 1859–1865 (2008).
54. Lomize, M. A., Pogozheva, I. D., Joo, H., Mosberg, H. I. & Lomize, A. L. OPM database and PPM web server: resources for positioning of proteins in membranes. *Nucleic Acids Res.* **40**, D370–D376 (2012).
55. Yu, W., He, X., Vanommeslaeghe, K. & MacKerell, A. D. Jr. Extension of the CHARMM General Force Field to sulfonyl-containing compounds and its utility in biomolecular simulations. *J. Comput Chem.* **33**, 2451–2468 (2012).
56. Vanommeslaeghe, K. et al. CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. *J. Comput Chem.* **31**, 671–690 (2010).
57. Van Der Spoel, D. et al. GROMACS: fast, flexible, and free. *J. Comput Chem.* **26**, 1701–1718 (2005).
58. Huang, J. et al. CHARMM36m: an improved force field for folded and intrinsically disordered proteins. *Nat. Methods* **14**, 71–73 (2017).
59. Klauda, J. B. 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).



60. Kovacs, L., Zalka, A., Dobo, R. & Pucsok, J. One-dimensional thin-layer chromatographic separation of lipids into fourteen fractions by two successive developments on the same plate. *J. Chromatogr.* **382**, 308–313 (1986).

## Acknowledgements

A portion of this research was supported by NIH grant U24GM129547 and performed at the PNCC at OHSU and accessed through EMSL (grid.436923.9), a DOE Office of Science User Facility sponsored by the Office of Biological and Environmental Research. This work is partially supported by NIH (R01 GM126626, R01 HL168686, and R21 AI175646).

## Author contributions

W.H., A.S., and H.Z. designed the experiments, collected and analyzed the data, and wrote the manuscript.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41467-024-48028-3>.

**Correspondence** and requests for materials should be addressed to Hongjin Zheng.

**Peer review information** *Nature Communications* thanks Daniel Wacker, Chuangye Yan, and the other anonymous reviewer(s) for their contribution to the peer review of this work. A peer review file is available.

**Reprints and permissions information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2024