Amino acids play essential roles in cell biology as regulators of metabolic pathways. Arginine in particular is a major signalling molecule inside the cell, being a precursor for both L-ornithine and nitric oxide (NO) synthesis and a key regulator of the mTORC1 pathway. In mammals, cellular arginine availability is determined by members of the solute carrier (SLC) 7 family of cationic amino acid transporters. Whereas CAT-1 functions to supply cationic amino acids for cellular metabolism, CAT-2A and -2B are required for macrophage activation and play important roles in regulating inflammation. Here, we present the crystal structure of a close homologue of the mammalian CAT transporters that reveals how these proteins specifically recognise arginine. Our structural and functional data provide a model for cationic amino acid transport in mammalian cells and reveals mechanistic insights into proton-coupled, sodium-independent amino acid transport in the wider APC superfamily.
Arginine plays an important role in regulating cellular metabolism through its ability to activate the mTORC1 kinase, a central signalling hub that integrates diverse environmental inputs to coordinate anabolic and catabolic processes in cells, such as insulin release and intestinal stem cell migration. In immune cells, arginine can also be shunted towards the production of nitric oxide (NO) via the action of nitric oxide synthase 2, where it functions in the production of reactive oxygen species. The intracellular concentration of arginine is therefore an important factor in determining key aspects of mammalian physiology.

Cellular arginine availability is determined in part by the efficiency and capacity of arginine transporters expressed in the plasma membrane, and in mammals is mostly mediated by members of the solute carrier 7 family (SLC7, pfam00324). The SLC7 family forms part of the much larger amino acid polycation (APC) superfamily, which is found in all organisms and represents a large family of secondary active transport proteins responsible for uniport, symport and antiport of a broad range of cellular metabolites into and out of the cell.

Several inherited disorders are linked to mutations in members of the SLC7 family, including cystinuria and lysinuric protein intolerance, both of which lead to abnormal amino acid transport, growth defects and kidney disease. Several prescription drugs, including L-DOPA and gabapentin are also recognised by members of the SLC7 family, which are expressed in the blood brain barrier and influence the transport and retention of drug molecules into the central nervous system.

The SLC7 family is divided into two subgroups, the cationic amino acid transporters (CATs, SLC7A1–4 and SLC7A14) and the L-type amino acid transporters (LATs, SLC7A5–13 and SLC7A15). Unlike the CATs, which function as monomers in the plasma membrane, the LATs are obligate heterodimers, forming a disulphide-linked dimer with a single transmembrane spanning glycoprotein, designated SLC3, which functions to traffic the transporter to the plasma membrane and aid in protein stability. Despite arising from the same ancestral transporter, the LAT and CAT families show various differences, not just in the interaction with the SLC3 family, but also with respect to their substrate specificity and transport mechanism. HATs transport a wide range of zwitterionic amino acids and for high affinity transport couple uptake to the Na+ electrochemical gradient. CATs in contrast are specific for cationic amino acids. Originally designated as system y+, the CATs mediate Na+-independent uptake with high affinity (K_M in the micromolar range) and are believed to operate as either exchangers or facilitators.

To regulate the uptake of arginine in different tissue and cell types the CAT-2 transporter is also found in two isoforms, CAT-2A and -2B, which differ in their affinity for arginine. Whereas CAT-2B, which is expressed in macrophages and T-cells, exhibits high affinity (µM) for arginine, CAT-2A, which is expressed in liver and muscle cells, displays low affinity (mM). Although the difference between CAT-2A and -2B has been localised to two amino acids that reside on an intracellular loop, there exists no molecular description for how this seemingly minor change can affect transport function in these proteins.

Although in mammals CATs operate as exchanger or facilitators, in plants they are pH dependent, raising the question of whether in certain species these systems are proton coupled. Indeed, the SLC7 family forms part of the much larger APC superfamily of secondary active transporters and are distantly related to the SLC36 family of proton coupled amino acid transporters. A key question in the field of transport biology is how superfamilies, such as the APC, have evolved to switch between passive facilitator, antiporter and coupled symporter mechanisms.

Crystal structures of APC superfamily transporters have revealed a conserved ‘5+5 inverted topology’ fold, wherein the first five TM helices are related to the second five helices via a pseudo two-fold symmetry axis running parallel to the plane of the membrane. Although originally discovered in Na+ coupled symporters, the same fold was also found associated with a sodium independent amino acid symporter ApcT and the amino acid exchangers AdiC and GadC. AdiC, an arginine-arginine exchanger, has been used as a model system for understanding amino acid recognition in the large-neutral amino acid transporter, LAT-1 (SLC7A5) and -2 (SLC7A8). However, neither AdiC nor the other currently reported structures from APC family members share high sequence identity with the mammalian CATs, leaving fundamental questions concerning amino acid recognition, selectivity and transport mechanism unanswered.

To address these questions, we determined the crystal structure of a close prokaryotic SLC7 homologue in complex with bound L-alanine and L-arginine ligands and studied its transport mechanism using in vitro biochemical assays. We exploited the close homology to the human proteins to identify a single amino acid substitution that dictates arginine recognition in the CATs. Site directed mutagenesis combined with liposome-based transport assays further revealed a structural explanation for the kinetic differences between CAT-2A and CAT-2B, and suggest a model for cationic amino acid recognition in eukaryotic cells.

## Results

### Structure of GkApcT

Following an extensive screen of prokaryotic homologues with high (>40%) amino acid sequence identity to eukaryotic SLC7 transport proteins, we identified a suitable candidate from the thermophilic bacterium *Geobacillus Kaustophilus*, GkApcT (Supplementary Fig. 1A & B). Initial attempts to crystallise GkApcT in the lipidic cubic phase were unsuccessful. Crystals were eventually obtained in the presence of L-Alanine and cholesterol doped monoolein, and diffracted X-ray to 2.86 Å. Phases were determined using molecular replacement, employing a CHAINSAW model of the distantly related APC homologue MjaApCt from *Methanococcus Jannaschii* (PDB 3GIA). The side chain register was further assigned and validated using the anomalous signal from a sulphur-SAD data set collected at 2.7 Å wavelength (Supplementary Fig. 2A & B) and the final structure refined to an R_free of 21.5/27.0 % (Table 1).

GkApcT consists of 12 transmembrane helices (TMs), which adopt the canonical APC superfamily fold, wherein transmembrane helices TM1–TM5 are related to helices TM6–TM10 by a pseudo two-fold symmetry axis located in the plane of the membrane (Fig. 1). Both TM1 and TM6 are broken in the centre of the transporter to form two discontinuous helices, termed here 1A, 1B and 6A and 6B respectively (Supplementary Fig. 1C). In GkApcT the two additional helices, TM11 and TM12, wrap around and to one side of the transporter. The N-terminus folds into a lateral helix that packs tightly on the cytoplasmic side of the protein before folding into the first TM helix. Our sequence alignment suggests a similar helix may exist in the mammalian proteins (Supplementary Fig. 1).

The electron density maps also revealed the presence of an additional single transmembrane helix that packs against TM5, and orientated with its amino-terminus facing the cytoplasm. The maps were of sufficient quality to allow us to identify this helix as YneM, which has recently been renamed as MgtS (Supplementary Fig. 2C & D), a single spanning transmembrane protein from *Escherichia coli*. MgtS packs tightly against GkApcT creating a hydrophobic pocket between TM5 of the transporter. In our structure, we observed density for a cholesterol molecule in this
We hypothesised that the combination of eukaryotic counterparts would be regulated by hopanoid binding in a similar way to their chromosomal counterparts in vivo and suggests that bacterial transport function may be observed in GkApcT may be a surrogate for a hopanoid binding site in vivo. We show that MgtS knock out strain of E. coli C43 (DE3) transports less substrate than the C43 (DE3) cells, despite the initial rates being similar (Supplementary Fig. 3A). The melting temperature of the GkApcT-MgtS heterodimer we observed no change in the presence of CHS. However, unlike for the GkApcT-MgtS heterodimer we observed no change in the presence of CHS. To test the effect of the MgtS interaction we constructed a ΔMgtS knock out strain of E. coli C43 (DE3) and used this for expression. The melting temperature of the GkApcT in the absence of MgtS was 58 °C, similar to that observed for the GkApcT-MgtS heterodimer (Supplementary Fig. 3B). However, unlike for the GkApcT-MgtS heterodimer we observed no change in the presence of CHS.

We also observed that the uptake of 3H-L-Alanine (L-Ala) in the protein purified from the ΔMgtS knock out strain transported less substrate than the C43 (DE3) cells, despite the initial rates being similar (Supplementary Fig. 3C & D). It is possible that the stability provided through the MgtS interaction enables GkApcT to operate over a longer period of time. Why we observe cholesterol in our crystal structure is also an interesting question. It is possible that the cholesterol we observe in GkApcT may be a surrogate for a hopanoid binding site in vivo and suggests that bacterial transport function may be regulated by hopanoid binding in a similar way to their eukaryotic counterparts.

Table 1 Data collection and refinement statistics for GkApcT

| Data collection | GkApcT WT + l-Ala PDB: 5OQT | GkApcT WT + l-Ala | GkApcT-M321S + l-Arg PDB: 6F34 |
|----------------|----------------------------|-----------------|-----------------------------|
| Space group    | P212121                    | P212121         | P212121                     |
| Cell dimensions|                            |                 |                             |
| a, b, c (Å)    | 76.9, 83.15, 119.74         | 77.19, 83.41, 119.64 | 77.04, 82.70, 118.78        |
| Resolution (Å) | 68.3-2.86 (2.93-2.86)       | 48.28-3.30 (3.56) | 64.63-3.13 (3.21-3.13)      |
| Rmerge (%)     | 15.2 (178.9)               | 34.2 (377.9)    | 17.3 (145.2)                |
| Rfree (%)      | 7.7 (84.2)                 | 7.9 (85.6)      | 7.4 (85.3)                  |
| I/σi            | 6.3 (1.1)                  | 11.1 (1.1)      | 5.5 (1.1)                   |
| CC1/2 (%)      | 86.3 (51.8)                | 99.6 (45.8)     | 99.8 (67.9)                 |
| Completeness (%) | 99.0 (100)       | 98.9 (97.7)     | 99.9 (99.8)                 |
| Multiplicity   | 5.2 (5.4)                  | 20.1 (21.4)     | 6.4 (6.3)                   |
| Anom. completeness (%) | 99.0 (97.8) | 99.9 (99.8)     | 99.9 (99.8)                 |
| Anom. multiplicity | 10.9 (11.3)   | 20.1 (21.4)     | 6.4 (6.3)                   |
| Mid-slope      | 1.12                       | 11.0            | 11.0                        |

Data collection and refinement statistics for GkApcT

- **GkApcT WT + l-Ala PDB: 5OQT**
  - Space group: P212121
  - Cell dimensions: a, b, c (Å) = 76.9, 83.15, 119.74
  - Resolution (Å): 68.3-2.86 (2.93-2.86)
  - Rmerge (%): 15.2 (178.9)
  - Rfree (%): 7.7 (84.2)
  - Completeness (%): 99.0 (100)
  - Multiplicity: 5.2 (5.4)
  - Anom. completeness (%) = 99.0 (97.8)
  - Anom. multiplicity: 10.9 (11.3)
  - Mid-slope: 1.12

- **GkApcT WT + l-Ala**
  - Space group: P212121
  - Cell dimensions: a, b, c (Å) = 77.19, 83.41, 119.64
  - Resolution (Å): 48.28-3.30 (3.56)
  - Rmerge (%): 34.2 (377.9)
  - Rfree (%): 7.9 (85.6)
  - Completeness (%): 98.9 (97.7)
  - Multiplicity: 20.1 (21.4)
  - Anom. completeness (%) = 99.9 (99.8)
  - Anom. multiplicity: 6.4 (6.3)
  - Mid-slope: 1.112

- **GkApcT-M321S + l-Arg PDB: 6F34**
  - Space group: P212121
  - Cell dimensions: a, b, c (Å) = 77.04, 82.70, 118.78
  - Resolution (Å): 64.63-3.13 (3.21-3.13)
  - Rmerge (%): 17.3 (145.2)
  - Rfree (%): 7.4 (85.3)
  - Completeness (%): 99.9 (99.8)
  - Multiplicity: 6.4 (6.3)
  - Anom. completeness (%) = 99.9 (99.8)
  - Anom. multiplicity: 6.4 (6.3)
  - Mid-slope: 1.112

*Highest resolution shell is shown in parentheses*
interaction made to a conserved tyrosine on TM7, Tyr268 (Fig. 2e). Interestingly, the hydroxyl group of Tyr268 sits in the same position as the sodium ion in the Na1 site in LeuT (Supplementary Fig. 4). This raised the question of whether GkApcT was coupled to either the sodium or proton electrochemical gradients.

GkApcT is a proton coupled amino acid transporter. Although mammalian CATs are facilitators, plant members of the SLC7 family are pH-dependent and members of the SLC36 family of amino acid transporters are proton-coupled. To deepen our understanding of amino acid transport and identify molecular commonalities between GkApcT and the wider APC superfamily, we sought to further understand the transport mechanism by studying [3H]-L-Ala uptake under the influence of inwardly directed proton or sodium gradients. GkApcT was determined to be proton-coupled, with uptake of L-[3H]-Ala observed in the presence of a membrane potential, negative inside and a pH gradient, alkaline inside (Fig. 3a). Transport was sodium-independent and abolished in the presence of the proton ionophore CCCP, indicating GkApcT is a proton-coupled amino acid transporter.

A previous study on the mechanism of a sodium-independent pH activated amino acid facilitator, MjApcT, suggested a semi-conserved lysine on TM5 as a site of proton binding. GkApcT contains an equivalent lysine on TM5, Lys191 (Fig. 3b). Similar to MjApcT, the amine group of Lys191 occupies the same position as the Na2 site in LeuT coordinating the discontinuous region of TM1 with TM8 (Supplementary Fig. 4). The pKa of Lys191 is 7.3 in the current structure, suggesting the amine group is in equilibrium between the neutral and positively charged form. To test whether Lys191 was responsible for proton coupling in GkApcT, we mutated this residue to alanine, arginine and asparagine (Fig. 3c). Under electrogenic uptake, which allows substrate accumulation, and counter flow uptake, which allows substrate equilibration, activity was abolished with the alanine and arginine substitutions. However, proton-driven transport was detectable, albeit at a lower level, in the Lys191Asn variant. This result suggests that Lys191 is not required for proton-coupled transport in GkApcT. However, when no transport is observed for a variant this may indicate disruption to substrate recognition. To test whether this was the case here, we used a thermal stability assay to investigate the effect of these variants on the ability of GkApcT to recognise L-Ala. We observed that in the presence of 10 mM L-Ala the WT protein showed a noticeable increase in melting temperature from ~58 to ~69 °C (Supplementary Fig. 5A). A similar stabilising effect was also observed in the Lys191Ala, Arg and Asn variants (Supplementary Fig. 5B), indicating that the loss of transport function we observed is not due to an inability to bind ligand but rather indicates that Lys191 is required for another aspect of the transport mechanism.
It is highly likely that TM1 undergoes significant conformational change during transport, as discussed below, and the neutral form of the lysine side chain would facilitate a dynamic interaction with the carbonyl groups in the unwound region of TM1. We therefore suggest that Lys191 has an important role in stabilising the unwound region of TM1 during transport, acting as a surrogate for the sodium ion observed in the Na2 site in structural homologues such as LeuT. The asparagine substitution in GkApcT at this position is able to partially substitute for this interaction. Taken together our data suggests that Lys191 may undergo protonation and deprotonation during the transport cycle; however, this is not involved in coupling transport to the proton electrochemical gradient but rather is important to facilitate the conformational changes in this region of the transporter.

To identify the proton-coupling sites, we next focused on two acidic residues, Glu115 (TM3) and Asp237 (TM6) that sit directly below the bound L-Ala ligand (Fig. 2b, d). As discussed above, both Glu115 and Asp237 coordinate an interaction between TM3 and TM6 via a water molecule. Asp237 is highly conserved within the SLC7 family (Supplementary Fig. 1) including those members that are not proton coupled. This suggests Asp237 is not the site of proton binding in GkApcT. However, its mutation to alanine did result in complete loss of function (Fig. 3d), highlighting an important role in transport. Mutation of Glu115 to alanine also resulted in loss of transport activity; however, a glutamine at this position demonstrated counterflow activity only, indicating a role in proton coupling. Such a role is consistent with its pKa value of 8.22 in the crystal structure, whereas Asp237 on the other hand has a pKa value of 5.48. Interestingly, the Glu115Gln variant can still recognise alanine to the same extent as wild-type protein (Supplementary Fig. 5C), suggesting the reduction of activity observed in the counterflow assay is not the result of the loss of the ability to bind ligand, but indicates an important secondary role in the transport cycle. It is likely given the observed interaction between Glu115 and Asp237 that this side chain coordinates interactions between TM3 and TM6A, which as discussed below forms part of the intracellular gate of the transporter.

Amino acid selectivity mechanism. CATs are highly specialised for L-arginine, L-lysine and L-ornithine transport unlike their close homologues the LAT transporters, which can also recognise neutral and anionic amino acids. Given the important role that CATs play in NO signalling and macrophage activation, we wished to understand the molecular basis of arginine selectivity within the CAT family. During our attempts to crystallise GkApcT, we used a counterflow assay to establish the protein as a relatively broad specificity amino acid transporter with a preference for small hydrophobic and polar amino acids (Fig. 4a). It was surprising, given the high sequence identity to the CATs, that GkApcT showed poor recognition of the cationic amino acids L-arginine and L-lysine. Analysis of the binding site in GkApcT revealed that several side chains that sit close to the bound L-Ala are considerably smaller in the human CAT homologues.
In the presence of the proton ionophore CCCP. Inset—transport is independent of sodium.

**Fig. 3** Functional characterisation of proton coupling residues in GkApcT. a Proton-coupled uptake assays showing that accumulation of L-Ala is driven by an inwardly directed proton electrochemical gradient generated using valinomycin (+v) and an inwardly directed pH gradient (−v). Transport was abolished in the presence of the proton ionophore CCCP. Inset—transport is independent of sodium. b View of Lys191 in the crystal structure, highlighting the interactions made to TMs 1 and 8. The bound L-Ala ligand is shown in magenta. c Mutational analysis of Lys191 using electrogenic and counterflow accumulation of 3H L-Ala. d Mutational analysis of Glu115 and Asp237 using electrogenic and counterflow accumulation assays. n = 3 independent experiments, error bars s.d.

In particular, Met321, Ala119, Val317 and Ile234 all adopt smaller side chains in the human proteins, being serine, glycine, serine and valine respectively. Given that amino acid recognition in GkApcT is mediated via backbone interactions to the α-carboxy and amino groups, we speculated that the substrate specificity, was due to steric constraints within the binding site, which effectively acted to exclude larger amino acids. To assess whether these side chains had any impact on specificity we generated a series of point mutants to assess their relative contribution to substrate specificity (Fig. 4c). Ile234 is positioned in the unwound region of TM6 and forms part of the cytoplasmic gate (Fig. 2b). Mutation to valine severely reduced transport in this variant. Ala119 and Val317 on TM3 and 8 respectively are positioned close to the bound L-Ala ligand, forming one side of the binding pocket. In a similar manner to Ile234, these mutants also showed reduced transport. However, we discovered that Met321Ser was able to recognise and transport both L-Arg and L-Ala uptake.

Relative transport (%)

|        | Electrogenic | Counterflow |
|--------|--------------|-------------|
| WT     | 100          | 50          |
| K191A  | 80           | 40          |
| K191R  | 60           | 30          |
| K191N  | 40           | 20          |

**Isoform differences between CAT-2A and -2B.** An intriguing aspect of CAT biochemistry is the observed kinetic difference between the two isoforms of CAT2. While CAT-2B displays the key features that characterise the system γ+ transporters, of μM affinity and trans stimulation, CAT-2A has ten-fold lower substrate affinity and shows no trans stimulation effect. The mechanistic basis underpinning these differences in transport kinetics however is currently not well understood, owing in part to the lack of an appropriate structural model. Previous studies have identified the intracellular region between TM8 and 9 as being important for the transport differences observed between CAT-2A and -2B and identified two amino acids responsible for the low affinity of CAT-2A, Arg369 and Asn381, which was inserted into this region extending the sequence by one amino acid. Arg369 in CAT-2A is equivalent to Arg334 in GkApcT, which in CAT-2B, CAT-1 and CAT-3 is a glutamate–aspartate pair (Fig. 5a). To further understand the molecular basis for the affinity switch we mutated the equivalent residue in GkApcT, with the α-carboxy and amino termini making identical interactions to those observed with L-Ala, with the side chain extending down into the space opened up by removal of the methionine. In our structure, the guanidinium group sits close to both Glu115 and Asp237. In the mammalian CATs Asp237 is strictly conserved, suggesting that this interaction may be important for cationic amino acid selectivity, whereas Glu115 is replaced with a serine, which would still provide a favourable polar interaction.
Arg334 to glutamate. Remarkably we observed that the IC$_{50}$ value for l-Ala uptake reduced from 2 $\mu$M in the WT protein to 0.4 $\mu$M in the Arg334Glu variant, mimicking the affinity increase observed for the CAT-2B isoform (Fig. 5b). The crystal structure provides a rational explanation for this affinity change. As noted above, the binding site is closed to the cytoplasmic side of the membrane through the packing of TM6B with TM3 and TM8. Arg334 makes a salt bridge interaction to the lateral helix through Glu15, which would stabilise TM8 and through side chain interactions TM6B and TM3. We suggest that changing this residue to glutamate, as occurs in the high-affinity CAT-2B isoform, causes a structural change in TM8, which in turn alters the packing of TM6B and TM3 affecting the affinity of amino acids in the binding site.

**Discussion**

The mammalian CAT proteins share only ~20% amino acid sequence identity with their counterparts, the L-type amino acid transporters. Unlike the LATs, the CATs function as amino acid facilitators and are sodium-independent$^{14}$, raising the question of how this subfamily of SLC7 transporters preferentially recognise and transport cationic amino acids into the cell. Previous studies on amino acid selectivity in AdiC, a bacterial homologue of the L-type proteins$^{26}$, suggests that arginine selectivity is driven by general electrostatic interactions in the binding site, rather than through specific interactions to side chain residues$^{42}$. Indeed, an overlay of the arginine-bound AdiC structure reveals several common features in the binding site between AdiC and GkApcT (Supplementary Fig. 6A). In particular, the positively charged $\alpha$-amino group interacts similarly with carbonyl groups in the discontinuous region of TM1 and TM6. The aromatic side chain of Phe231 also sits in a similar position to a conserved tryptophan in AdiC, Trp202, which acts to close the binding site from the extracellular side of the membrane during transport$^{43}$. A notable difference however is observed for the negatively charged $\alpha$-carboxylate group. In AdiC the $\alpha$-carboxylate group accepts two H-bonds from the side chain Ser26 (TM1) and the amide nitrogen of Gly27. In contrast, in GkApcT we observe the $\alpha$-carboxylate group of l-Ala interacting with a well-ordered water...
molecule, in addition to the backbone amide groups of Thr43 and Gly44 (Fig. 2b, d). However, the most striking feature of this arrangement is the water molecule sits in a similar position to the Na⁺ ion in LeuT, coordinating the interaction between the substrate and TM1 (Supplementary Fig. 4)\(^3\). We modelled this density as water, due to the sodium independence of GkApT in our transport assays. The use of water as a coordinating molecule is a feature often associated with transiently formed complexes, and suggests movement of TM1 would cause disruption to the hydrogen bond network and weakening of the interactions to the amino acid. Dynamic water networks have been proposed to play a role in coordinating TM3 and TM10 in AdiC\(^4\). The tethering of TM1A to the lateral helix would likely severely reduce the movement of this helix, supporting a more dynamic mechanism, whereby the protein switches between outward facing and inward facing states to move ligands across the membrane (Fig. 6a)\(^4\). Within the APC superfamily several hypotheses have emerged regarding the structural rearrangements\(^5\). Certainly, individual families have their own idiosyncratic mechanisms for transport, but within the ‘LeuT-like’ fold it is generally considered that TM1 and 6 act as toggle switches, which orchestrate the conformational changes during the alternating access cycle\(^2\). In our case, GkApT is the first proton-coupled amino acid transporter to be structurally and biochemically studied, revealing important insights into the transport mechanism in this subfamily of SLC7 transporters. In the current inward occluded state the bound l-Ala ligand is trapped inside the binding site through the formation of a thick extracellular gate and thin intracellular gate (Fig. 6b). For the bound ligand to exit into the cell, either TM1A or TM6B must swing away from TM8, breaking the cytoplasmic thin gate (Figs. 2a, 6a).

Secondary active transporters operate via an alternating access mechanism, whereby the protein switches between outward facing and inward facing states to move ligands across the membrane (Fig. 6a)\(^4\). Within the APC superfamily several hypotheses have emerged regarding the structural rearrangements\(^5\). Certainly, individual families have their own idiosyncratic mechanisms for transport, but within the ‘LeuT-like’ fold it is generally considered that TM1 and 6 act as toggle switches, which orchestrate the conformational changes during the alternating access cycle\(^2\). In our case, GkApT is the first proton-coupled amino acid transporter to be structurally and biochemically studied, revealing important insights into the transport mechanism in this subfamily of SLC7 transporters. In the current inward occluded state the bound l-Ala ligand is trapped inside the binding site through the formation of a thick extracellular gate and thin intracellular gate (Fig. 6b). For the bound ligand to exit into the cell, either TM1A or TM6B must swing away from TM8, breaking the cytoplasmic thin gate (Figs. 2a, 6a). The tethering of TM1A to the lateral helix would likely severely reduce the movement of this helix, supporting a more dynamic role for TM6B. In the crystal structure, we see that TM6B makes several interactions that help close the cytoplasmic gate. In particular a water-mediated interaction between Asp237 (TM6B) and Glu155 (TM3), and Thr241 (TM6B) and Arg327 (TM8). Our data suggest that Glu115 plays important roles in proton binding in GkApT. Given the pKa of this residue, ~8, in the current structure, we suggest that Glu115 is protonated, and interacting with Asp237 via the observed water molecule (Fig. 2c). Proton release from Glu115 would facilitate the movement of TM6B away from TM3 and TM8, driven by the repulsion between the two acidic residues. This movement would serve to break the interactions between the carbonyl groups of TM6 with the α-amino group of the ligand, thus dissolving the binding site, a process no doubt aided through the use of water as a coordinating molecule to the ligand and TM1. The importance of the interaction between TM6 and 8 is further supported by our finding that the isoform differences between CAT-2A and -2B are likely due, in part, to the stability of TM8, which may have an effect on the kinetics of transport by modulating this movement.

**Fig. 5** Structural basis for isoform difference between CAT-2A and -2B. **a** Zoomed-in view of the cytoplasmic region of the transporter, highlighting the interactions between TM6B, TM8 and the lateral helix. A sequence alignment of TM8 highlighting the difference (*) between CAT-2A and CAT-1, -2B and -3 is shown. **b** Representative I₅₀ values determined for l-Ala binding to either the WT or R334E variant of GkApT using electric accumulation assays. I₅₀ values were calculated from three independent replicates.
A long-standing question in the SLC7 field is the position of the heavy subunit of HATs (rBAT and 4F2hc; SLC3A1-A2) relative to the LAT domain (SLC7A5-13)\(^1\). The current structure of MgtS in complex with GkApcT reveals how a single TM helix can pack against the transport domain of an APC family member. The heteromeric complex between GkApcT and MgtS is consistent with a previous homology model of the SLC7A6 complex, which places the single TM of the heavy subunit (4F2hc) in close proximity to TM5 in LAT-2\(^2\) and may therefore represent a viable structural model for this interaction. A recent report on the modulation of LAT-1 transporter activity by cholesterol suggests a role for the sterol molecule in stabilising the interaction with SLC3A2\(^4\). Our observation that the interaction between MgtS and GkApcT creates a cholesterol binding site (Fig. 1) and that this also serves to stabilise the heterodimer, suggests a similar structural arrangement may also occur between the heavy and light subunits of the HATs.

In summary, we present crystal structures of the first close homologue of the mammalian CATs. The complexes with bound alanine and arginine reveal the molecular basis for amino acid selectivity is based on steric restraints rather than charge complementarity, providing insight into how the SLC7 family has evolved to select specific sub-sets of amino acids. An important question in transport biology is how similar protein folds have evolved different transport mechanisms. The structural and functional data reported here reveal how the '5 + 5 inverted topology' fold, which is widespread in many SLC families, has adapted to couple amino acid uptake to the proton electrochemical gradient within the APC superfamily.

### Methods

**Protein expression and purification.** The gene encoding GkApcT (Uniprot Q5L1G5) was amplified from *Geobacillus kaustophilus* genomic DNA and cloned into the pWaldo-GFP vector using primers detailed in Supplementary Table 1\(^4\).

---

**Fig. 6** A model for alternating access transport in GkApcT. **a** Schematic representation of alternating access transport in GkApcT. Starting from an outward facing state the transporter must reside in a conformation that facilitates access of protons, water and amino acid to the binding site. Binding of amino acid causes conformational changes in TM1 and 6 that act to close the extracellular gate. This may be facilitated by water binding to TM1 and movement of F231 that facilitates formation of the occluded state. Further conformational changes are predicted to occur that move the transporter into a fully occluded state, before the intracellular gate starts to open\(^4\). Our crystal structure likely represents this ‘inward occluded’ state, where the intracellular gate is still present, but held shut through side interactions rather than main chain packing. In GkApcT E115 is likely protonated in this state, holding TM6 in close approximation to TM8, where M321 forms part of the intracellular gate. Proton release from E115 causes the water mediated interaction to D237 to break, and TM6 to swing away from TM3 and 8, opening the binding site to the inside of the cell and releasing the bound amino acid and possibly a water molecule. **b** Structure of GkApcT shown in molecular surface representation with the predicted gating helices shown. Our data suggest that movement of TM6 away from TM3 and 8 opens the binding site to the interior of the cell, and facilitates release of the bound amino acid.
Variants forms of GkApcT were made through site directed mutagenesis using standard methods. Wild-type and variant GkApcT proteins were expressed and purified using standard NADH-oxidase protocols.26,27,28 Proteins were thawed and solubilised in purification buffer which consisted of, 1x PBS containing an additional 150 mM NaCl and 1% n-decyl-β-D-maltopyranoside (DDM) with stirring for 1 h. The solubilised material was recovered through ultracentrifugation at > 200,000 x g for 1 h. A final concentration of 15 mM imidazole was added and the protein was bound to nickel resin (GE Healthcare) in batch for between 1 and 2 h, or when the level of binding had plateaued, as measured using the GFP counts. The resin was washed with purification buffer containing 25 mM imidazole and 0.1% DDM for 20–30 column volumes. GkApcT was eluted from the resin with 250 mM imidazole. TEV protease was added and the protein was dialysed overnight in gel filtration buffer containing 0.03% DDM (20 mM Tris pH 7.5, 150 mM NaCl). After dialysis, the protein was passed through a HiTrap column to remove the TEV protease and the GFP tag. The pure protein was concentrated and applied to a superdex 200 10/300 gel filtration column (GE Healthcare). For crystallisation, size exclusion chromatography was performed using a buffer consisting of 20 mM Tris-HCl and 150 mM NaCl with 0.3% n-decyl-β-D-maltopyranoside (DDM).

Crystallisation. Crystallisation was performed using protein at 10 mg mL\(^{-1}\) final concentration, determined using absorption at 280 nm. 1-Alanine or l-arginine were added to a final concentration of 10 mM and the protein was left on ice for at least 2 h prior to LCP set up. Protein-laden mesophases was obtained by mixing cholesterol doped monolein (10% CS)25 with protein in a 60:40 (w/w) ratio using a coupled syringe device (Art Robbins, USA). Initial crystals appeared at 20 °C in a condition of the MembOleo screen (Molecular Dimensions Ltd, UK), which was further optimised to 28–34% PEG 400, 0.1 mM sodium acetate pH 4.0 and 0.1 mM potassium fluoride, containing 10 mM of either alanine or arginine. Diamond or rod-shaped crystals appeared after 3 days and reached their full size (~30 µm in the longest dimension) after 2 weeks. Only the rod-shaped crystals diffracted to high resolution; all the diamond-shaped crystals were noticeably less well ordered. Wells were opened using a tungsten glasscutter and the crystals were harvested using 30 µm micromotors (MiTeGen). Crystals were cryo-cooled directly in liquid nitrogen and stored in unipacks.

Structure determination. X-ray diffraction data collected at beamline I24, Diamond Light Source, UK resulted in a complete data set of GkApcT to 2.86 Å. Structure determination using a Superdex 200 10/300 column into 20 mM Tris-HCl, 150 mM NaCl and 0.3% n-decyl-β-D-maltopyranoside (DDM).

Proteoliposomes were harvested by ultracentrifugation and dialysed overnight in lipid buffer. The proteoliposomes were stored at ~80 °C.

Proton-coupled and counterflow transport assays. For proton-driven uptake, proteoliposomes were thawed, centrifuged at 90,000 x g for 30 min at 4 °C and re-suspended in internal buffer (20 mM potassium phosphate pH 6.50, 100 mM potassium acetate, 2 mM magnesium sulphate) followed by five freeze-thaw cycles and extrusion through a 400 nm polycarbonate filter. Proteoliposomes were subsequently diluted 1:25 (v/v) into external buffer containing 250 mM 3H-Ala with 100 µM unlabelled alanine (20 mM sodium phosphate pH 6.50, 100 mM sodium chloride, 2 mM magnesium sulphate) with 10 µM valinomycin. Assays were performed at 25 °C and stopped through rapid filtration onto 0.22 µm nitrocellulose filters. The filters were washed under vacuum with 2 x 2 ml of 0.1 M lithium chloride prior to scintillation counting. The 3H signal was converted to molar concentrations of peptide using standard curves for each substrate. For ligand-driven counterflow transport26,27, proteoliposomes were loaded through repeated freeze thaw cycles with 25 mM citrate phosphate, pH 6.0, 100 mM sodium chloride, 2 mM magnesium sulphate and 2 mM of ligand for determining the substrate specificity (Fig. 4a) or 10 µM l-Ala or Arg or Lys for subsequent counterflow studies (Fig. 4c).

Transport assays to determine IC\(_{50}\) values. Transport was carried out at 25 °C by diluting GkApcT proteoliposomes 1:25 into external buffer (20 mM sodium phosphate pH 6.50, 100 mM sodium chloride, 2 mM magnesium sulphate with 10 µM valinomycin) containing 250 mM 3H-Ala, and increasing concentrations of competitor. Reactions were terminated after 1 min by transferring 50-µl aliquots of each reaction mixture into 2 ml cold lithium chloride. Reaction mixtures were filtered and analysed as previously described. The entire experiment was conducted in triplicate to obtain a mean and standard deviation.

Thermal stability measurements. To evaluate thermal stability of protein samples two methods were employed. The first used CD spectroscopy to calculate melting temperatures.67 Concentrated (>10 mg mL\(^{-1}\)) protein was diluted to ~0.1 mg mL\(^{-1}\) into buffer consisting of 10 mM potassium phosphate pH 6.5, 50 mM sodium sulphate and 3x CMC of the desired detergent. CD spectra were obtained between 20 and 95 °C at 2 °C increments on a JASCO 815 spectrophotometer. Stability curves were obtained by plotting the absorption value at λ = 220 nm (% helicity) against the temperature. The resultant data were plotted in Prism v7 (GraphPad Software) and analysed using a sigmoidal curve fit. In addition, thermal stability in the presence and absence of 10 mM ligand was also analysed using a Prometheus NT.48 (NanoTemper Technologies). The proteins were diluted to a final concentration of 0.5 mg mL\(^{-1}\) into buffer containing 25 mM citrate phosphate, pH 6.0, 100 mM sodium chloride, 2 mM magnesium sulphate and 0.03% (w/v) DDM. Thermal measurements were carried out in a range from 20 to 95 °C with 1 °C per min steps. The resulting melting curves were generated by plotting the first derivative of the fluorescence ratio at 330 nm/350 nm against temperature.

Data availability. Atomic coordinates for the crystal structures have been deposited in the Protein Data Bank under accession numbers 5OQT (WT, L-Ala co-crystal complex) and 6F34 (M321S, l-Arginine co-crystal complex). Other data are available from the corresponding author upon reasonable request.

References
1. Saxton, R. A., Chantranupong, L., Knockenhauer, K. E., Schwartz, T. U. & Sabatini, D. M. Mechanism of arginine sensing by CASTOR1 upstream of mTORC1. Nat. Metab. 5, 536–547 (2016).
2. Chantranupong, L. et al. The CASTOR proteins are arginine sensors for the mTORC1 pathway. Cell. 165, 153–164 (2016).
3. Matthews, D. R. & Boland, O. The stimulation of insulin secretion in non-insulin-dependent diabetic patients by amino acids and glidazone in the basal and hyperglycemic state. Metabolism 46, 5–9 (1997).
4. Rhoads, J. M., Niu, X., Odle, J. & Graves, L. M. Role of mTOR signaling in intestinal cell migration. Am. J. Physiol. Gastrointest. Liver Physiol. 251, G150–G157 (2001).
5. Comalada, M., Yeramian, A., Modolell, M., Lloberas, J. & Celada, A. Arginine and macrophage activation. Methods Mol. Biol. 844, 223–235 (2012).
6. Morris, S. M. Jr. Arginine metabolism revisited. J. Nutr. 146, 2579S–2586S (2016).
Acknowledgements
We thank the staff of beamlines I24 and I23 Diamond Light Source, UK, Proxima-2A, Soleil, France and BL-1A, Photon Factory, Japan for assistance. We thank Professor O. Nureki and Professor R. Ishitani, University of Tokyo, Japan, for assistance in accessing BL-1A under the approval of the Photon Factory Program Advisory Committee (Proposal 2016-R01) and Dr. Armin Wagner, Diamond Light Source, UK, for assistance in accessing I23 during commissioning. K.E.J.J. wishes to thank Professor Elspeth Garman and Dr. Ed Lowe for training and assistance in crystallographic work. This work was supported by Wellcome (102899/Z/13/Z), the European commission Marie Curie Training Network NanoMem and Molecular Dimensions Ltd.

Author contributions
S.N. conceived the study. K.E.J.J., J.L.P. and S.N. designed the experiments and interpreted the data. J.L.P. and S.N. wrote the paper with assistance from K.E.J.J.

Additional information
Supplementary Information accompanies this paper at https://doi.org/10.1038/s41467-018-03066-6.

Competing interests: The authors declare no competing financial interests.

Reprints and permission information is available online at http://npg.nature.com/reprintsandpermissions/

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 license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons license 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 license, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2018