

Structural basis for pH-regulated amino acid transport via SLC7A4.

Dimitrios Kolokouris^{1,2,3}, Anuja Bothra^{1,2}, Takafumi Kato^{1,2}, Yi C. Zeng^{1,2}, Simon Lichtinger^{1,3}, Joanne L. Parker^{1,2}, Philip Biggin^{1,3}, Simon Newstead^{1,2*}.

¹ Department of Biochemistry, University of Oxford, Oxford, OX1 3QU, UK.

² Kavli Institute for Nanoscience Discovery, University of Oxford, Oxford, UK.

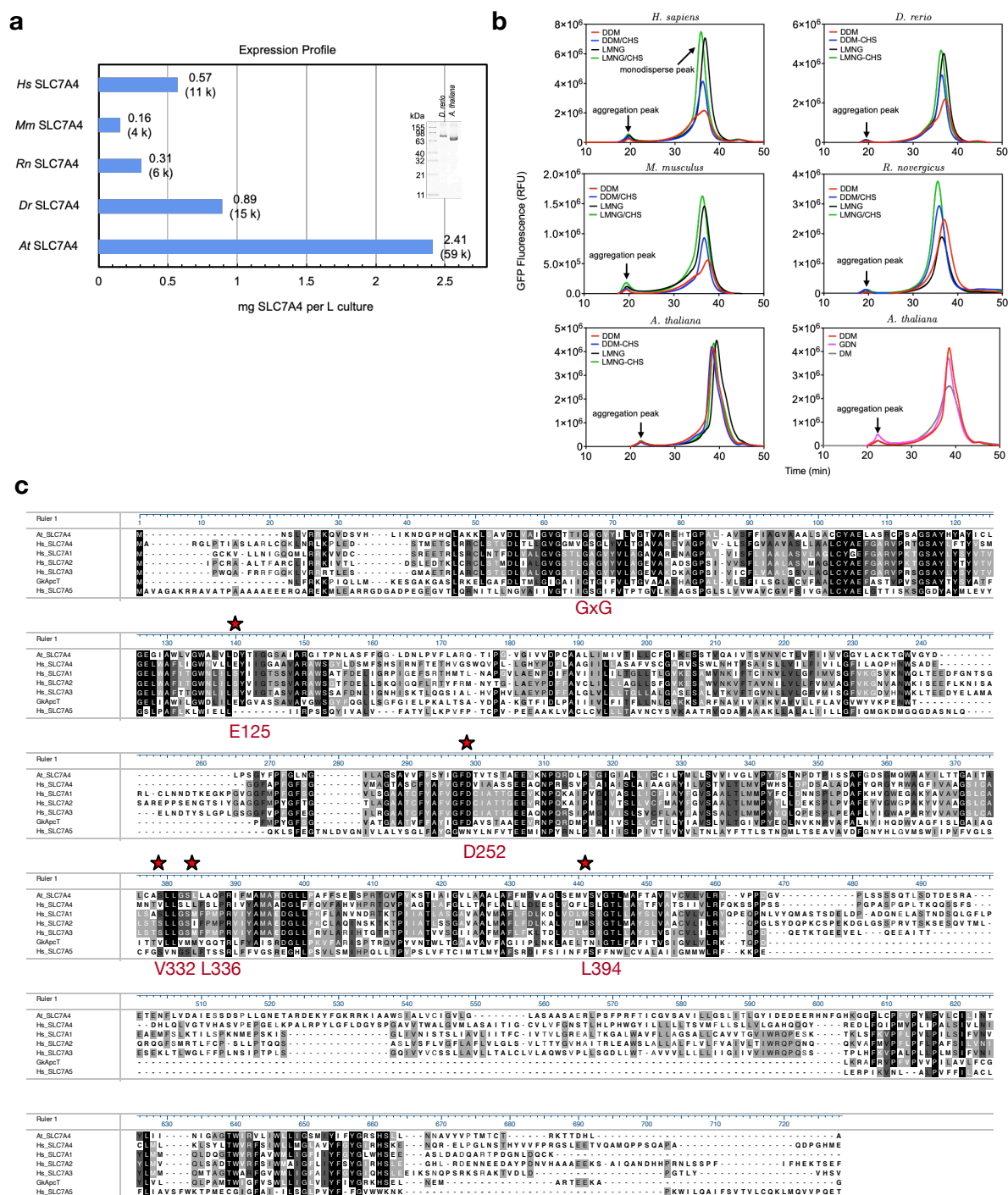
³ Structural Bioinformatics and Computational Biochemistry, Department of Biochemistry, University of Oxford, Oxford, UK.

* simon.newstead@bioch.ox.ac.uk

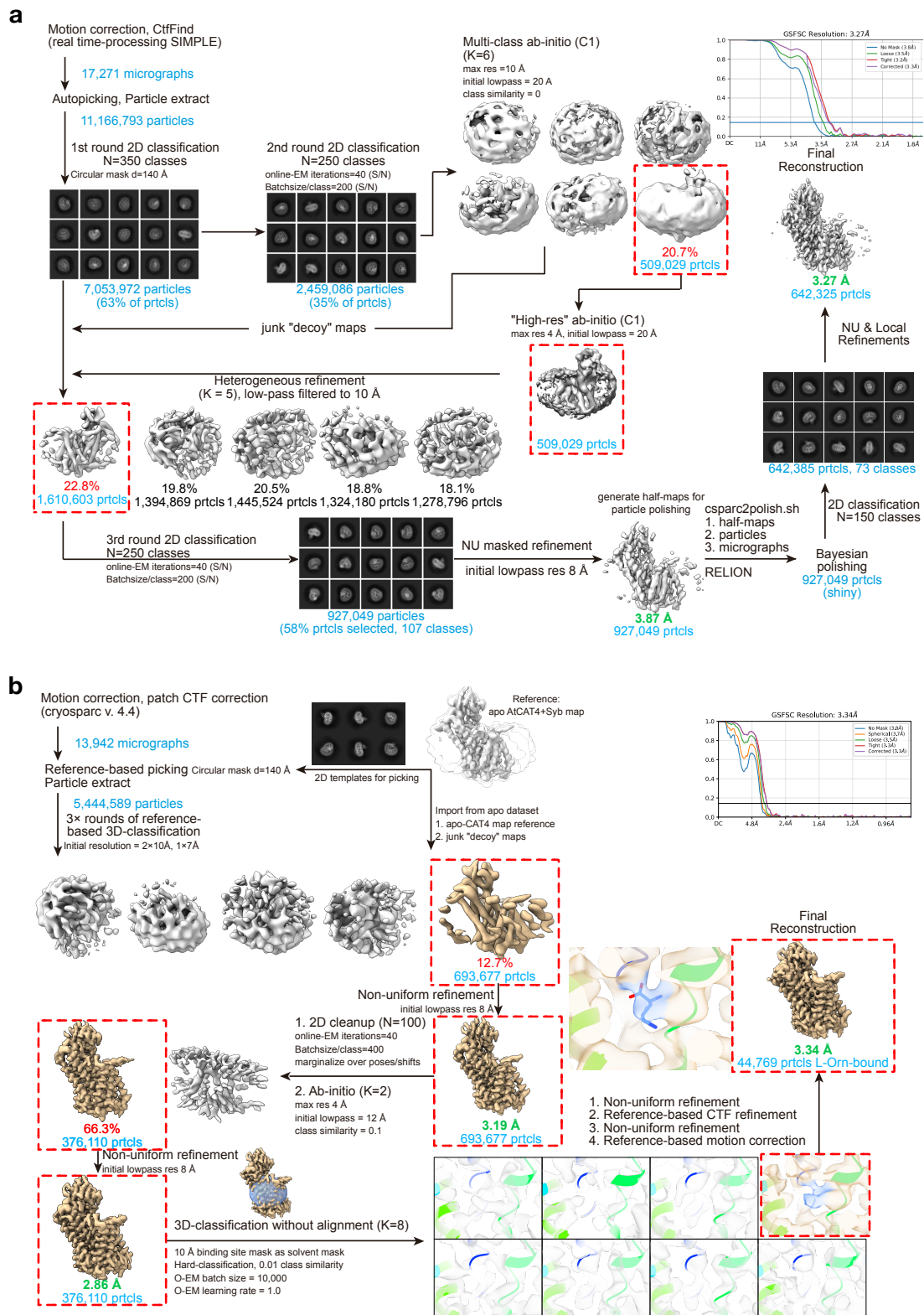
Extended Data Figures 1-9

Extended Data Table 1: Cryo-EM Data Collection, Processing and Refinement Statistics.

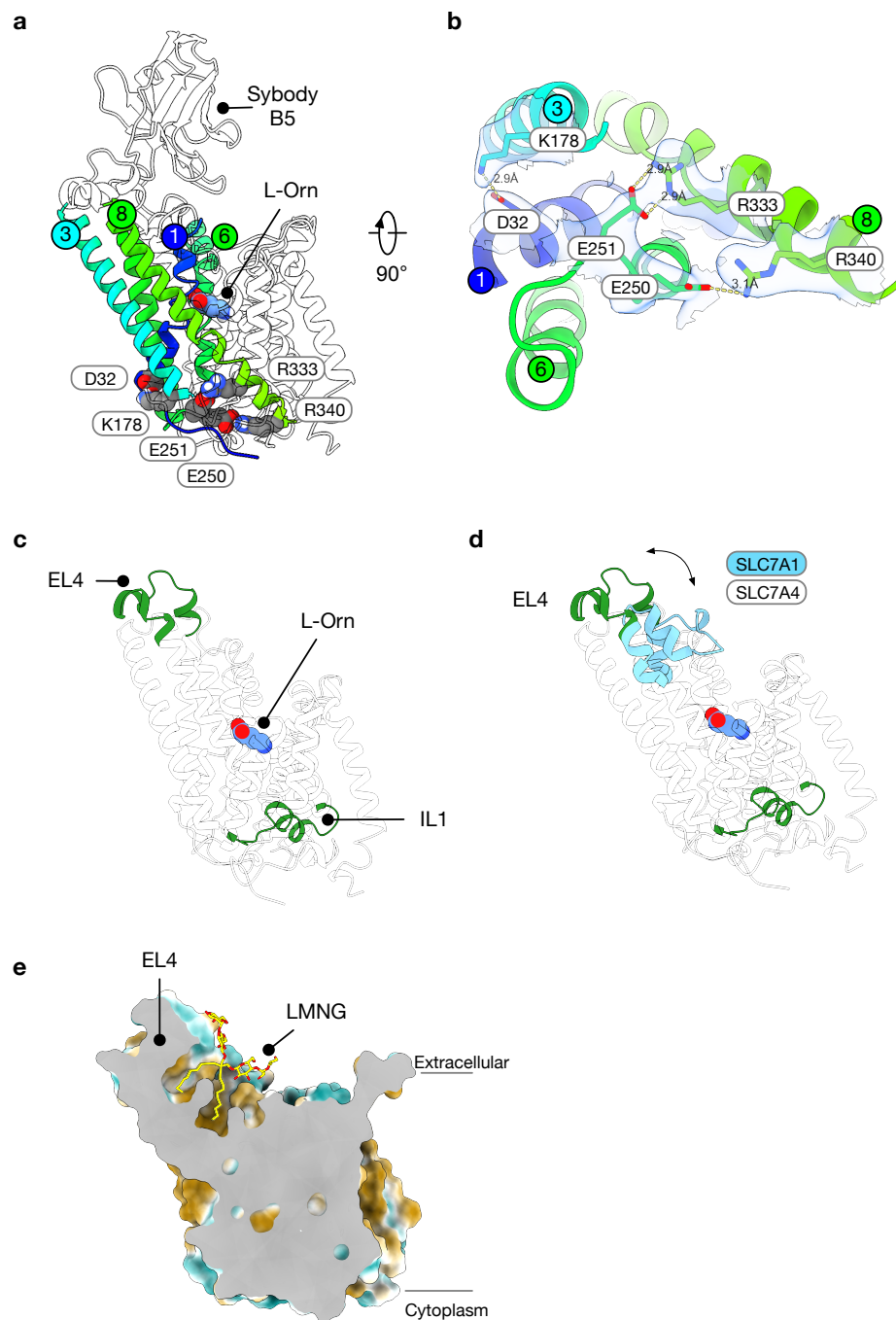
Extended Data Table 2: Molecular Dynamics Data Table.



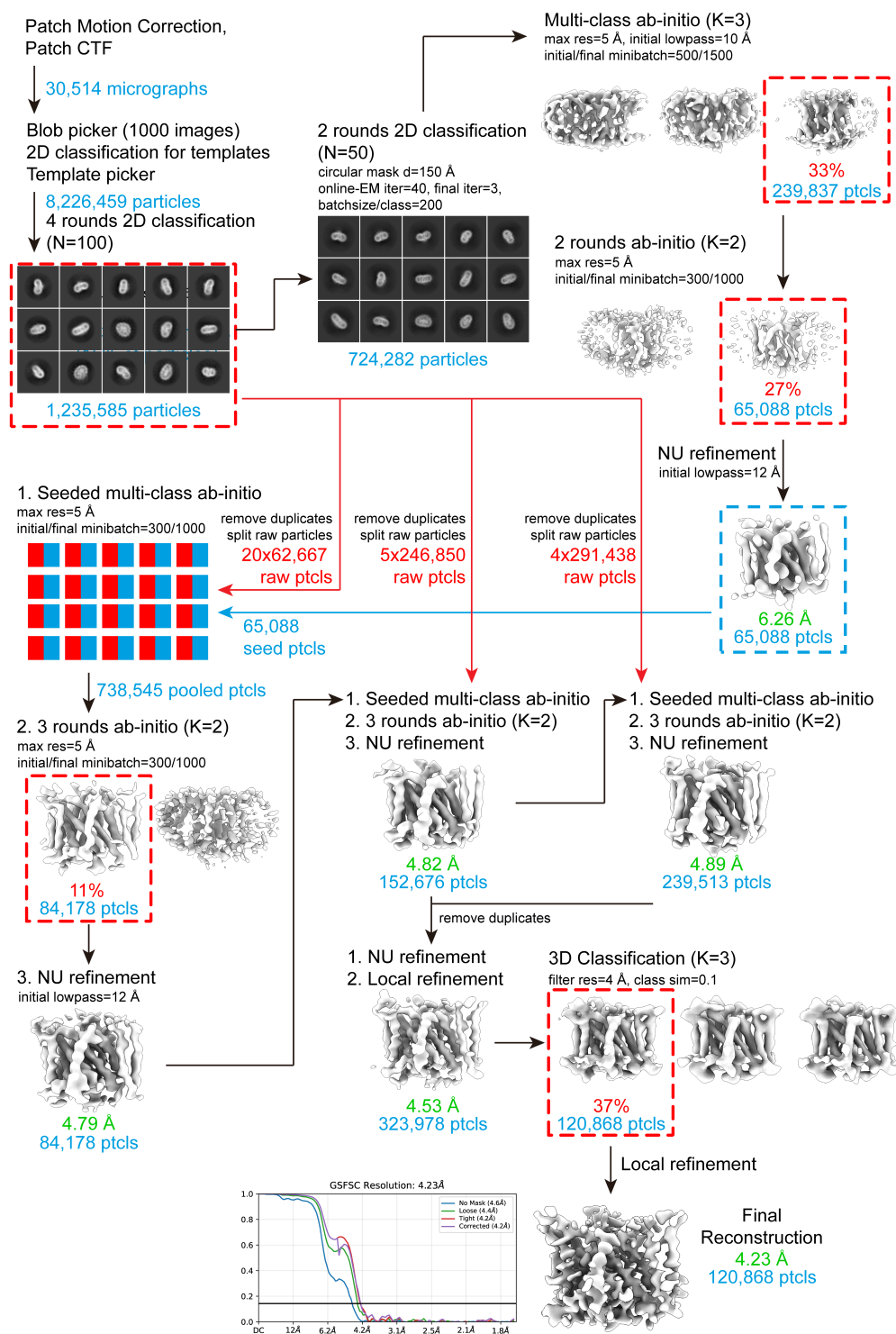
Extended Data Figure 1. Homologue screening for SLC7A4. **a.** Expression data for recombinant production of SLC7A4 homologues in *S. cerevisiae*. Inset – SDS PAGE gel of purified At SLC7A4 used for the structural and biophysical studies. **b.** Fluorescent SEC profiles for the SLC7A4 homologues displayed in panel a. **c.** Sequence alignment of SLC7A4 homologues (At_SLC7A4 Uniprot: Q8W4K3; Hs_SLC7A4: O43246; Hs_SLC7A1: P30825; Hs_SLC7A2: P52569; Hs_SLC7A3: Q8WY07; GkApcT: Q5L1G5; Hs_SLC7A5: Q01650).



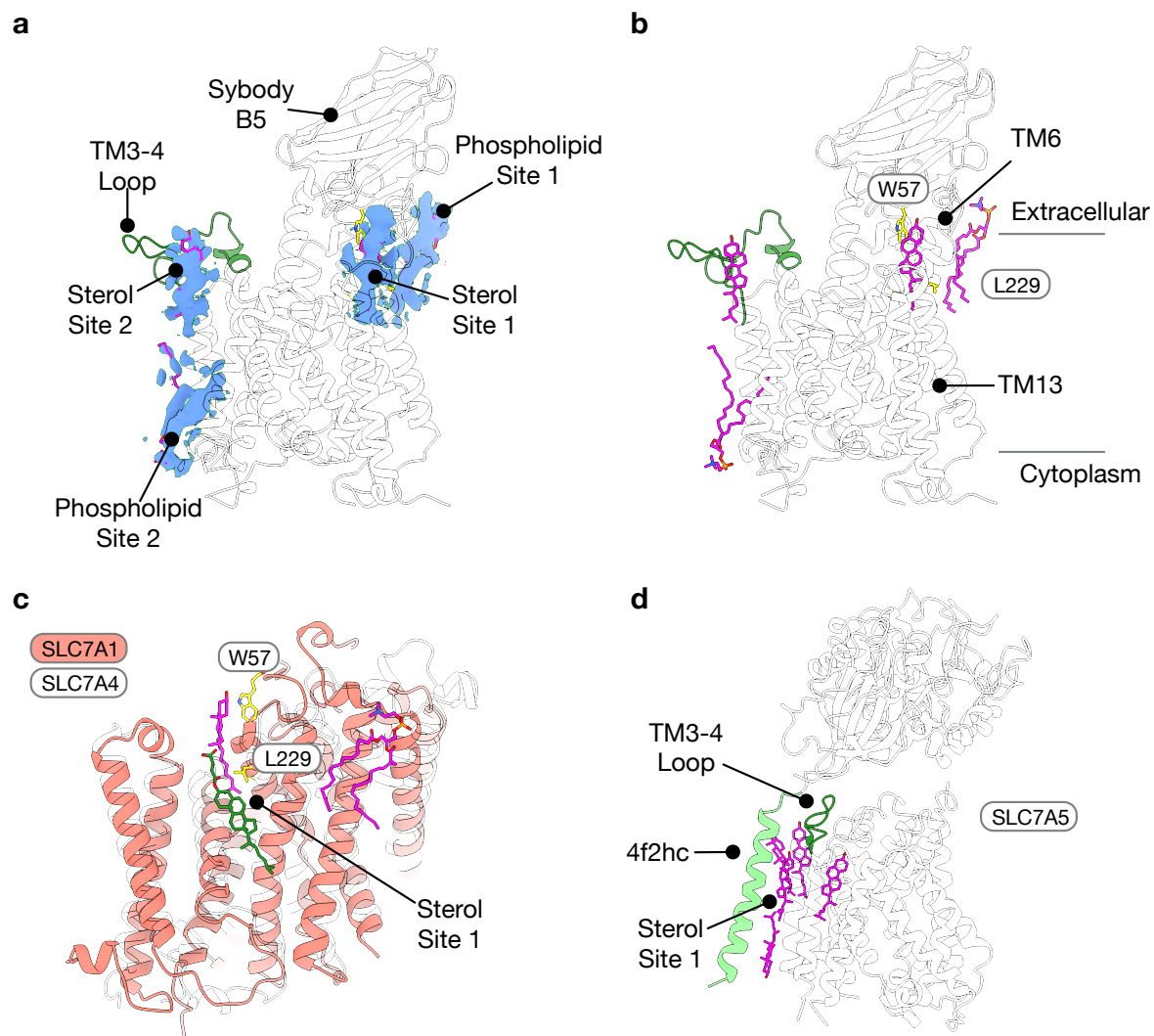
Extended Data Fig. 2. Cryo-EM processing pipelines. a. Cryo-EM processing pipeline for Apo SLC7A4. **b.** Cryo-EM processing pipeline for L-Orn bound SLC7A4.



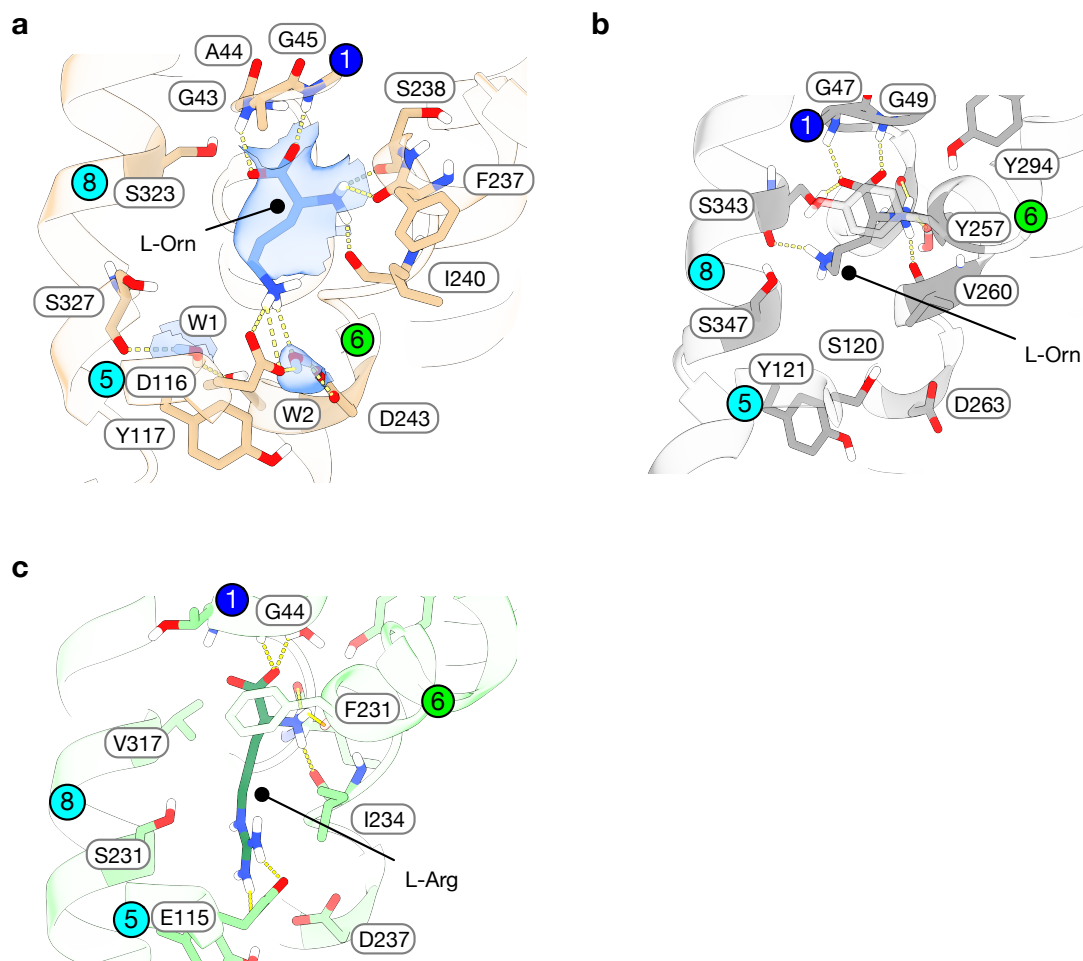
Extended Data Figure 3. Structural Analysis of SLC7A4. **a.** Cryo-EM structure of SLC7A4 showing the cytoplasmic salt bridge network stabilising the closed state of the transporter. **b.** Zoomed in view of the cytoplasmic salt bridge network showing the cryo-EM density (blue and threshold 0.01). **c.** Structure of SLC7A4 bound to L-Orn showing the location of EL4 and IL1. **d.** Structural comparison of EL4 between SLC7A4 (green) and SLC7A1 (blue; PDB: 9FQT). The arrow indicates the positional difference between EL4 between SLC7A1 and SLC7A4. **e.** Hydrophobic surface representation of At SLC7A4 showing the location of the bound LMNG detergent molecule.



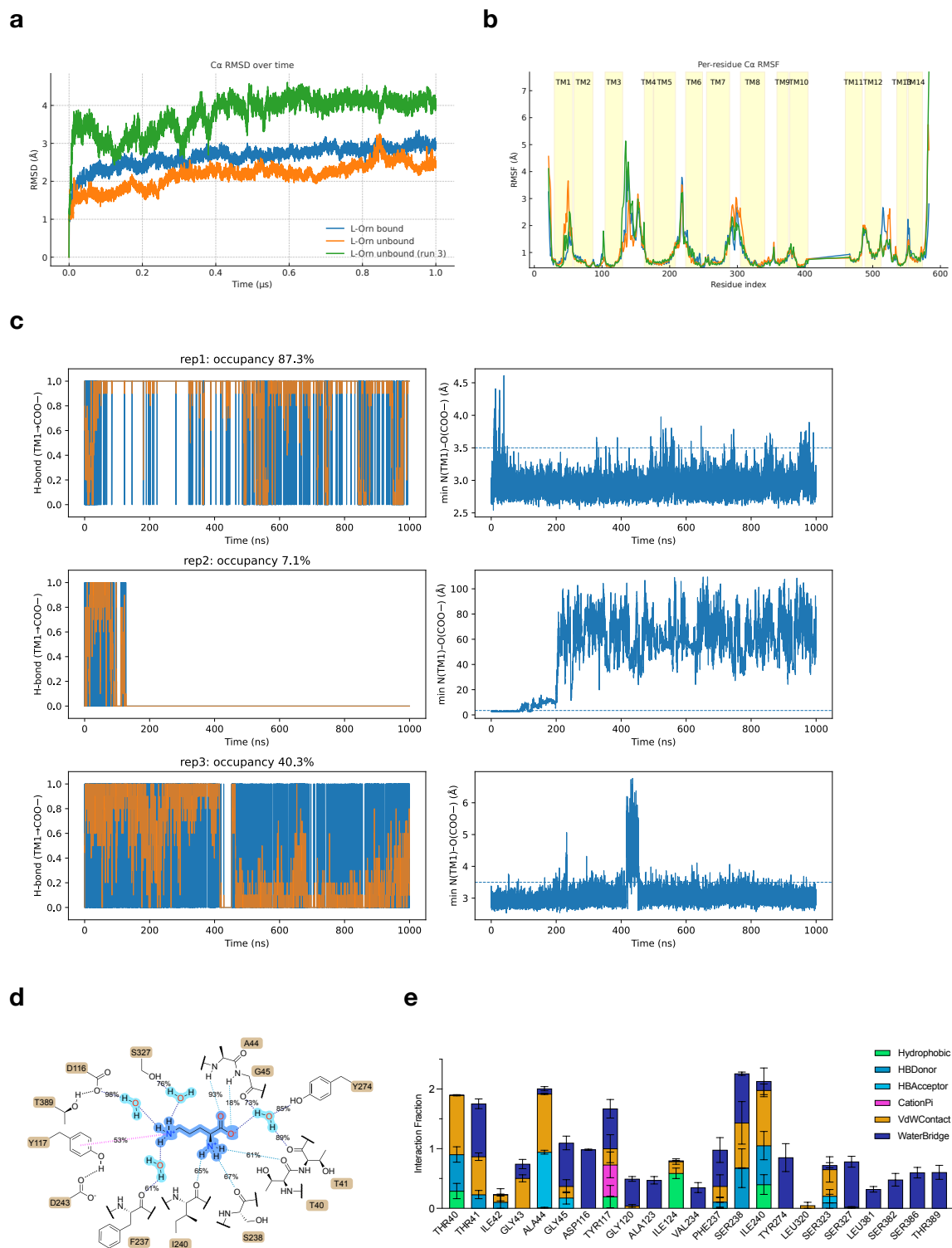
Extended Data Fig. 4. Cryo-EM processing pipeline for Apo SLC7A4 without sybody B5. Seeded multi-class ab-initio was performed using particles from NU refinement as “seeds” (blue) into split stacks of “raw” particles (red).



Extended Data Fig. 5. Analysis of lipid binding sites in SLC7A4. **a.** Cryo-EM structure of SLC7A4 with sharpened cryo-EM density for the bound lipid molecules shown (blue and threshold 0.025). **b.** Equivalent view as panel b, but without the cryo-EM density. The main interactions with the sterol in site 1 are labelled. **c.** Overlay of the mouse SLC7A1 structure (PDB: 9FQT) onto the At SLC7A4 structure, revealing the conservation of the site 1 sterol binding pocket. **d.** Cryo-EM structure of SLC7A5 (LAT1; PDB: 6JMQ) showing the sterol binding sites in a similar position to site 2 in SLC7A4.

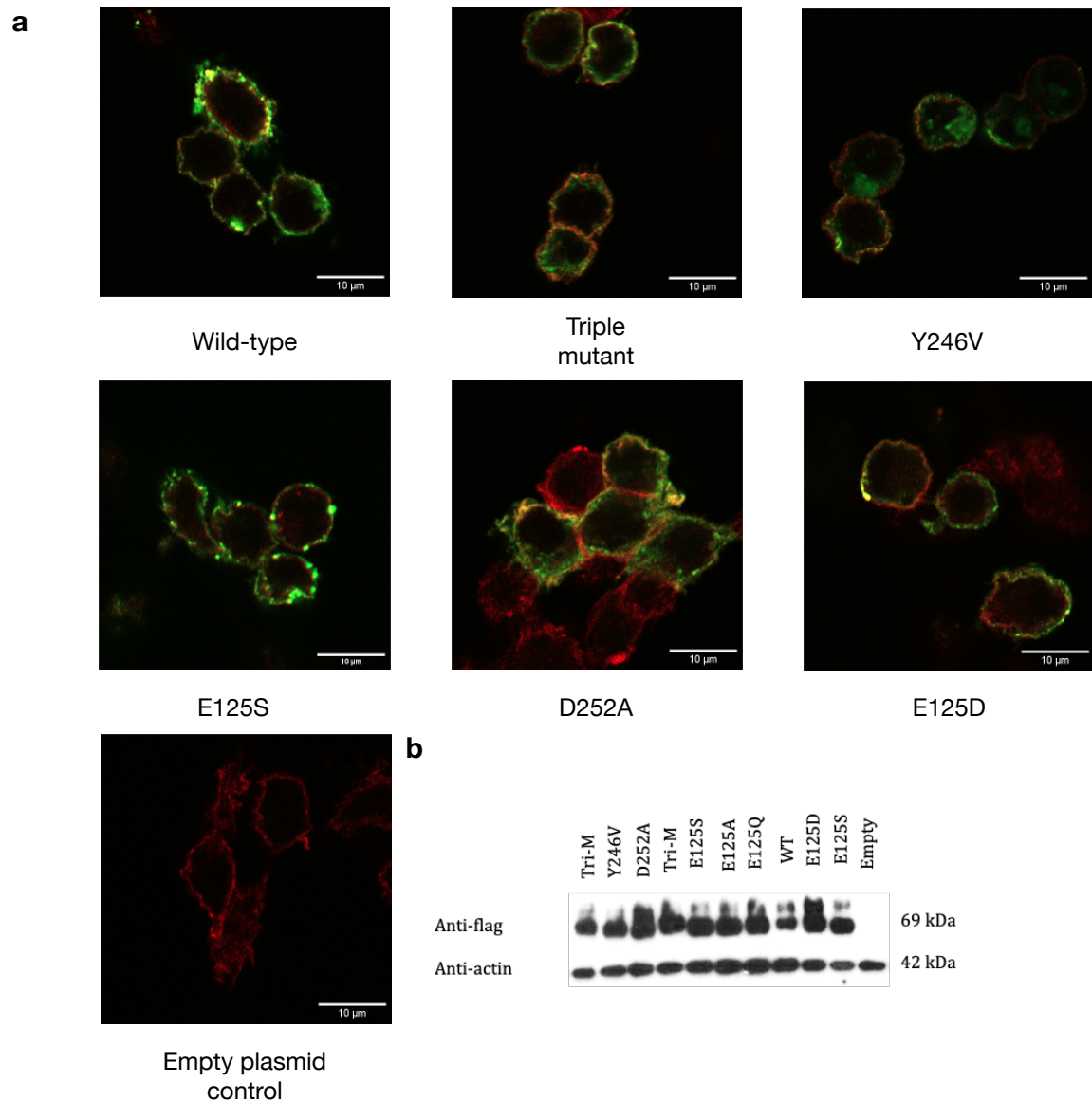


Extended Data Fig. 6. Structural comparison of amino acid binding in SLC7A4 to GkApcT and SLC7A1. **a.** The binding site showing the interaction with L-Orn. Key residues interacting with the substrate are shown as sticks, and hydrogen bonds are represented with dashed lines. Cryo-EM density for the L-Orn is shown (blue), contoured at 0.008. Replicated from Fig. 2d for comparison. **b.** Equivalent view of the amino acid binding site in SLC7A1 (PDB: 9FQW) bound to L-Ornithine (L-Orn). **c.** Equivalent view of the amino acid binding site in GkApcT (PDB: 6F34) bound to L-Arginine (L-Arg).

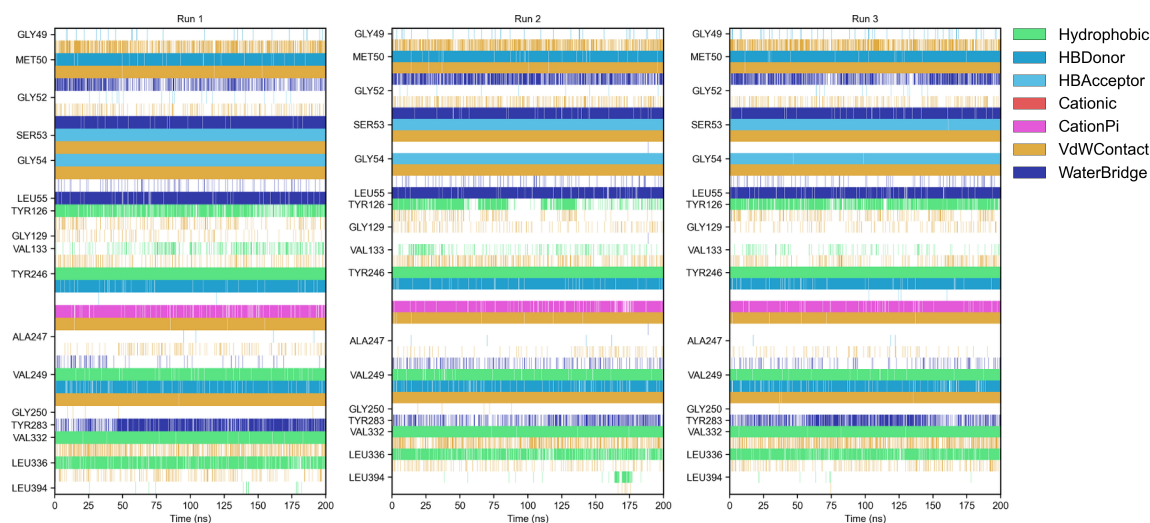
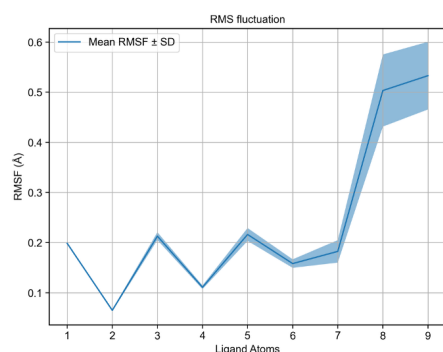
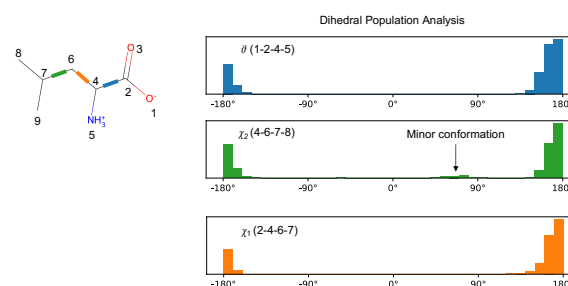


Extended Data Fig. 7. Molecular Dynamics simulations (3×1 μ s and 5×200 ns) of L-Orn bound SLC7A4. **a.** Ca root mean square deviation (RMSD) calculated over the whole protein sequence, quantifying At SLC7A4 deviation from initial conformation (in Å) over simulation time [excluding IL5 (residues 406-464), which is an unstructured 58-residue domain built as a loop for MD]. Protein least squares were fitted on the stable hash and gate domains (residues 87-224, 282-405, 465-

579). **b.** Per-residue Ca root mean square fluctuation (RMSF) calculated over the whole protein sequence, excluding IL5 and fitting as in **a**. Residues belonging to transmembrane helices 1-14 are shown. **c.** Time-series analysis of the L-Orn carboxylate engagement with TM1 (anchor residue ALA44) for each 1 μ s MD replicate (rep1-3). Left panel (per rep): binary trace indicating presence (1) or absence (0) of a direct H-bond where the ligand acts as acceptor (HBAcceptor) from the TM1 backbone N-H (blue line). The 10-window rolling average of the binary trace is shown (orange line). The H-bond occupancy over simulation time is shown per run. Right panel (per rep): minimum heavy-atom distance between ALA44 backbone N and one of the L-Orn carboxylate O atoms (whichever is closer as they are chemically degenerate); the dashed line at $y=3.5$ Å is a reference donor-acceptor distance consistent with typical N-H $\cdots$ O hydrogen-bond geometry used throughout in this work's H-bond analyses. **d.** 2d interaction schematic quantifying (in % total simulation time) pairwise interactions between L-Orn and AtSLC7A4 averaged from 5 $\times$ 200 ns independent repeats (dashed line colours as in **e**). The 5 reps were seeded from the most representative binding site configuration from 1 μ s rep 1 (residues within 10 Å of L-Orn, cutoff 1 Å with the linkage method). **e.** Bar chart showing the per-residue interaction profile averaged from the simulation replicates.



Extended Data Fig. 8. Functional characterisation of Hs SLC7A4. a. Immunofluorescence staining of Hs SLC7A4 overexpressed in HCT-116 SLC7A5 KO cells. The green channel shows the localisation of C-terminally FLAG-tagged Hs SLC7A4 imaged with anti-FLAG antibody; the red channel shows the plasma membrane Na^+/K^+ ATPase. Images are representative of three independent experiments and taken using a 40x objective. **b.** Western blot analysis of the WT and mutant SLC7A4 used in the cell-based uptake assays using anti-FLAG antibody for Hs SLC7A4 and anti-actin as a loading control.

a**b****c**

Extended Data Fig. 9. Molecular dynamics simulations on the L-Leu bound Hs SLC7A4 homology model. **a.** Time-resolved per residue interaction analysis; all recorded interactions shown for each independent run. **b.** L-leu RMSF plot showing stable ligand binding over 3 replicates. A modest increase in RMSF is seen for atoms 8 and 9, attributed to a small population of 90° rotation around the dihedral (4-6-7-8; marked in green), shown in panel c. **c.** L-Leu dihedral conformational analysis sampled from the 3 200 ns replicates with θ referring to the backbone dihedral formed by atoms 1-2-4-5 in the L-Leu structure, χ_1 and χ_2 reporting the dihedrals for the isopropyl sidechain atoms ($\chi_1=2-4-6-7$, $\chi_2=4,6,7,8$). A relatively minor $\chi_2 \sim 60^\circ$ versus the major 180° explains the $\text{RMSF} > 0.5 \text{ \AA}$ for atoms 8 and 9 in panel b.

Table 1. Cryo-EM data collection, refinement, and validation statistics

	At SLC7A4 apo + SybB5 (EMDB-52217) (PDB 9HJK)	At SLC7A4 L- Orn bound + SybB5 (EMDB-55065) (PDB 9SP8)	At SLC7A4 apo (EMDB- 55110) (PDB 9SQH)
Data collection and processing			
Magnification	105,000	105,000	105,000
Voltage (kV)	300	300	300
Electron exposure (e-/Å ²)	58.0	50.0	39.9
Defocus range (µm)	-2.0 to -1.0	-2.0 to -0.8	-2.0 to -0.8
Pixel size (Å)	0.832	0.4175	0.832
Symmetry imposed	C1	C1	C1
Initial particle images (no.)	1,610,603	5,444,589	8,226,459
Final particle images (no.)	642,325	44,769	120,868
Map resolution (Å)	3.3	3.3	4.2
FSC threshold	0.143	0.143	0.143
Refinement			
Initial model used (PDB code)	At SLC7A4/CAT4 AlphaFold2 entry (Q8W4K3)	9HJK	9HJK
Model composition in the asymmetric unit			
Non-hydrogen atoms	4818	4965	3193
Protein residues	631	633	433
Ligands	98	240	0
Average B factors (Å ²)			
Protein	90.83	98.74	77.76
Ligand	–	111.87	–
Waters	58.98	70.15	–
R.m.s. deviations			
Bond lengths (Å)	0.005	0.004	0.002
Bond angles (°)	0.793	0.615	0.583
Validation			
MolProbity score	1.58	1.46	1.56
Clashscore	7.68	5.33	6.86
Poor rotamers (%)	0.81	0.61	0.00

Table 2. Molecular dynamics data table

	System	Starting Model	Replicates × Time	Referenced in paper as	Deposited Files
1	At SLC7A4 L-Apo	9HJK	3 × 1000 ns	rep 1-3 referenced as "Apo" in Fig 2 rep 3 data used in Fig 2b	• Apo_1000ns/initial_config.gro • Apo_1000ns/rep1_whole_noju mp_center_rotxyz-transxyz.pdb • Apo_1000ns/rep2_whole_noju mp_center_rotxyz-transxyz.pdb • Apo_1000ns/rep3_whole_noju mp_center_rotxyz-transxyz.pdb • Associated .tpr, .topol and .mdp files
2	At SLC7A4 L-Orn bound	9SP8	3 × 1000 ns	rep 1: L-Orn remains bound - data used in Fig 2d, ED Fig 7a-c rep 2: L-Orn diffuses from binding site - data used in ED Fig 7a-c rep 3: site L-Orn disengages TM1 - data used in Fig 2c, ED Fig 7a-c	• Orn_1000ns/initial_config.gro • Orn_1000ns/rep1_whole_noju mp_center_rotxyz-transxyz.pdb • Orn_1000ns/rep2_whole_noju mp_center_rotxyz-transxyz.pdb • Orn_1000ns/rep3_whole_noju mp_center_rotxyz-transxyz.pdb • Associated .tpr, .topol and .mdp files • Zwitterionic L-ornithine (ORN) GAFF2/RESP parameterization files
3	At SLC7A4 L-Orn bound	Seeded from median binding site configura tion of system 2	5 × 200 ns	rep 1-5 data pooled and used in ED Fig d-e	• Orn_200ns/initial_config.gro • Orn_200ns/rep1_whole_noju mp_center_rotxyz-transxyz.pdb • Orn_200ns/rep2_whole_noju mp_center_rotxyz-transxyz.pdb • Orn_200ns/rep3_whole_noju mp_center_rotxyz-transxyz.pdb

					• Orn_200ns/rep4_whole_nojump_center_rotxyz-transxyz.pdb • Orn_200ns/rep5_whole_nojump_center_rotxyz-transxyz.pdb • Associated .tpr, .topol and .mdp files • Zwitterionic L-ornithine (ORN) GAFF2/RESP parameterization files
4	Hs SLC7A4 L-Leu bound	Homology model (HM) of 9SP8	1 × 1000 ns	HM and system equilibration; unrestrained – used to equilibrate L-leu-bound HM to seed system 5	• HM_1000ns/initial_config.gro • HM_1000ns/pr1000ns_whole_nojump_center_rotxyz-transxyz.pdb • Associated .tpr and .mdp files • Zwitterionic L-leucine (ZL) parm99sb “Horn” parameterization files (AMBER and Gromacs formats)
5	Hs SLC7A4 L-Leu bound	Seeded from median binding site configuration of system 4	3 × 200 ns	rep 1-3 data pooled and used in ED Fig 8 a-c	• HM_200ns/initial_config.gro • HM_200ns/rep1_whole_nojump_center_rotxyz-transxyz.pdb • HM_200ns/rep2_whole_nojump_center_rotxyz-transxyz.pdb • HM_200ns/rep3_whole_nojump_center_rotxyz-transxyz.pdb • Associated .tpr, .topol and .mdp files • Zwitterionic L-leucine (ZL) parm99sb “Horn” parameterization files (AMBER and Gromacs formats)
			8.6 μs total simulation time		All forcefield, topology and parameter files are deposited in Zenodo [DOI: 10.5281/zenodo.17184930].