Supplementary Information for:

Structural basis for substrate binding and specificity of a sodium/alanine symporter AgcS

Jinming Ma[1], Hsiang-Ting Lei[1], Francis E. Reyes[1], Silvia Sanchez-Martinez[1], Maen F. Sarhan[1], Johan Hattne[1,2] and Tamir Gonen[1,2$]

[1] Janelia Research Campus, Howard Hughes Medical Institute, 19700 Helix Drive Ashburn VA 20147 USA.
[2] Howard Hughes Medical Institute and Departments of Biological Chemistry and Physiology, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles CA 90095 USA.

[$] Correspondence should be addressed to T.G. (tgonen@ucla.edu)

This PDF file includes:
Supplementary text
Figures S1 to S9
Table S1
References for SI reference citations

Supplementary Information Text

Supplementary Methods

Protein expression and purification

The genes of wild-type AgcS (Gene ID: 2761075) from *Methanococcus maripaludis* and its site-directed mutants were produced by polymerase chain reaction (PCR) and then subcloned into a pET52b plasmid containing a thrombin-cleavage site and a C-terminal octa-histidine tag. All proteins were overexpressed in *E. coli* BL21 (DE3) C43 at 16°C for overnight with 0.2 mM isopropyl-β-D-thiogalactopyranoside (IPTG) as inducer. Then, the cells were harvested, homogenized in a lysis buffer containing 20mM Tris (pH 8.0) and 150mM NaCl, and disrupted using a Microfluidizer (Microfluidics Corporation) with 2 passes at 15,000 p.s.i., followed by a low-speed centrifugation for 10 min to remove cell debris. The supernatant was collected and ultra-centrifuged at 130,000g for 1 hour. The pellets containing membrane were resuspended in the same lysis buffer and frozen at -80 °C until use.

To purify AgcS and mutants, membrane fraction was thawed and solubilized with 2% n-decyl-β-D-maltoside (DM, Anatrace) for 3 hours at 4 °C. Following another ultra-centrifugation at 130,000g for 1 h, the supernatant was loaded onto Ni^{2+} nitrilotriacetate affinity resin (Ni-NTA, Qiagen). The resins were washed with 50 mM imidazole in the buffer containing 20 mM Tris-HCl (pH 8), 150 mM NaCl, 0.2% DM. The protein was eluted by 300 mM imidazole in the same buffer. The octa-histidine tag was removed by thrombin cleavage at enzyme:protein molar ratio of 1:1,000 at 4°C overnight. The protein was then applied to gel filtration (Superdex-200, GE Healthcare) in 20 mM Tris-HCl (pH 8), 150 mM NaCl and 0.2% DM. The peak fractions were collected for crystallization and further functional studies.

Selenomethionine-labeled AgcS protein was prepared as previously described(1) and purified like wild-type protein.

Fab fragments production

Fab fragments were produced at Monoclonal Antibody Core of Vaccine and Gene Therapy Institute, OHSU. Mouse monoclonal antibodies against AgcS wild-type were raised by standard protocol(2) using purified protein in detergent as antigen. Western blot and native-to-denature ELISA assays(3) were performed to assess the binding affinity and specificity of the antibodies generated from hybridoma cell lines. Several monoclonal antibodies showing high binding affinity and specificity to conformational epitope were then selected and purified from the hybridoma supernatants. Fab fragments were generated by Papain (Thermo Fisher Scientific) digestion and purified by Protein A affinity chromatography (GE Healthcare).

Purification of AgcS-Fab complexes for crystallization

Purified AgcS was mixed with excess Fab at a molar ration of 1:2 for 1 hour, and the complex was then subjected to gel filtration (Superdex-200, GE Healthcare) in the buffer containing 20 mM Tris-HCl (pH 8), 150 mM NaCl and 0.2% DM with different substrates. The purified protein complex was collected and concentrated to 3.0-3.5 mg/mL.

Crystallization

Crystallization was carried out by hanging-drop vapor-diffusion at 4 °C. Initial hits of purified AgcS were identified in multiple PEG conditions. However, these crystals gave anisotropic diffraction to ~6 Å. Well-diffracting crystals were only obtained when AgcS was co-crystallized with Fab fragments at 2.0-3.0 mg/mL mixed 1:1 with drop solution containing 0.1 M Sodium Citrate pH 5.6 and 2.5 M Ammonium Sulfate.

Data collection and structure determination

Before data collection, crystals were soaked in a cryoprotectant buffer containing 20% Ethylene Glycol in the same drop solution for 1 min, and rapidly frozen in liquid nitrogen. All diffraction data for AgcS-Fab complex and derivatives were collected at 100K using synchrotron radiation at the Advanced Light Source (Beamlines 8.2.1 and 5.0.2) and the Advanced Photon Source (NE-CAT 24-ID-C). Diffraction data indexing, integration and scaling were performed with autoPROC(4), the RAPD online server and the CCP4 suite(5). Statistics for data collection, phasing and refinement are given in Table S1. Molecular replacement using Phaser(6) was able to place the Fab monoclonal antibody (PDB ID: 4x80) in native datasets (TFZ score of 15). The Se-Met dataset was then phased by single anomalous dispersion in Phenix(7, 8) using differences from 27 Se atoms at $\lambda = 0.9791$ Å and the two Fab fragments previously placed as a partial model in Phaser (9). Helices of AgcS were manually placed in the density-modified map and extended within Coot(10) (Fig. S7), interleaving rounds of model building and refinement against the experimental phases, a reference model for the antibody, and non-crystallographic restraints using BUSTER-TNT(11, 12). Anomalous difference Fourier maps computed from a selenomethionine derivative were used for sequence assignment. L-alanine or D-alanine was modeled into sigma-weighted *mFo-DFc* maps using LeuT (PDB ID: 3tt1) as a guide (Fig. S9a). Sodium sites could be placed by strong positive density in sigma-weighted *mFo-DFc* maps but not refined due to the low resolution of the data (Fig. S9b). The Ramachandran analyses of the final structures were performed using Molprobity(13). The models have been deposited into the Protein Data Bank (PDB: 6cse, 6csf).

Isothermal titration calorimetry

Isothermal Titration Calorimetry (ITC) was employed to measure the binding affinities between wild-type AgcS with various amino acids. A solution of wild-type AgcS at 40-80 μM in 20 mM Tris-HCl pH 8.0, 150 mM NaCl, 10 mM KCl and 0.2% DM was loaded into the sample cell of a MicroCal iTC200 calorimeter. Amino acids at 400 μM for titrations were dissolved in the same buffer and loaded into the injection syringe. Before data collection, the system was equilibrated to 25 °C with the stirring speed set to 1,000 r.p.m. Titration curves for binding amino acids were generated by one 0.8 μL injection followed by nineteen 2.0 μL injections at 180 s intervals. Control injections of different amino acids into the buffer without protein were also determined for background corrections. The integrated heat from each injection were normalized to the amount of ligand per injection and then fitted to a single-site binding isotherm using ORIGIN 7.

Proteoliposomes reconstitution and radioligand uptake assays

Liposomes were prepared using a 3:1 ratio of *E. coli* total lipid extract (Avanti Polar Lipids) to chicken egg phosphatidylcholine (egg-PC) at 30 mg/mL in assay buffer (20 mM HEPES pH 7.5, 150 mM NaCl and 1mM DTT). An extruder with pore size of 0.4 μm was used to obtain unilamellar vesicles. Triton X-100 was then added to the extruded liposomes at 10:1 (w:w) lipid:detergent ratio. Purified wild-type AgcS and variants were reconstituted at a 1:400 (w/w) ratio in destabilized liposomes and excess detergent was removed by SM2 Bio-Beads (Bio-Rad) at 4 °C overnight. Next day, proteoliposomes were collected, aliquoted and frozen at -80°C for storage until needed.

Uptake assays were initiated by adding [^{3}H]-labeled amino acids (American Radiolabeled Chemicals) to 100 μL of proteoliposomes (total of 1μg protein) at room temperature. At different time points, reactions were stopped by quenching the samples with 20 mL assay buffer followed by rapid filtration through 0.22 μm membrane filter (MilliporeSigma, GSWP02500) to remove excess radioligands. The filter was then suspended in 10 mL of scintillation fluid and quantified by scintillation counting. As controls, non-specific uptake was assessed by using protein-free liposomes under identical conditions for background corrections. All the uptake assays were repeated three times.

Supplementary Figures and Table

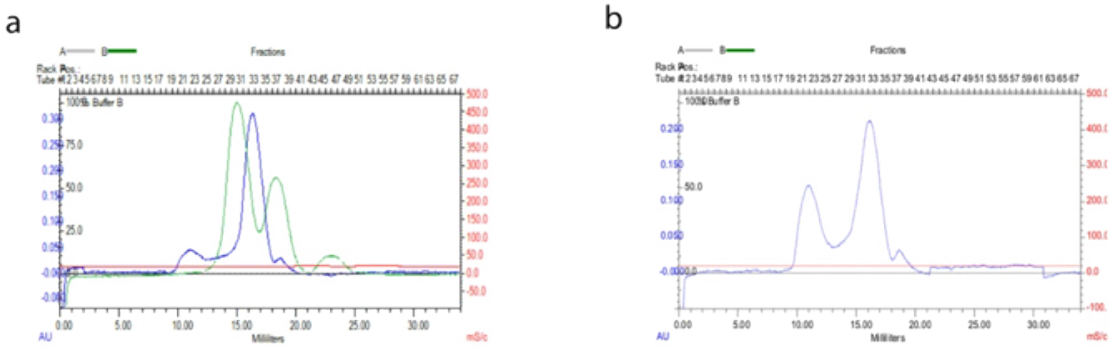

Fig. S1. Elution profile of purified AgcS and AgcS-Fab assembly. **(a)** Blue line is the elution trace of pure AgcS protein. Green line represents the elution trace of AgcS-Fab complex and unbound Fab. An apparent peak shift was observed for the formation of AgcS-Fab complex. **(b)** The elution trace for 3A mutation of AgcS. Only the well-folded proteins were taken to the reconstitution experiments in the liposomes.

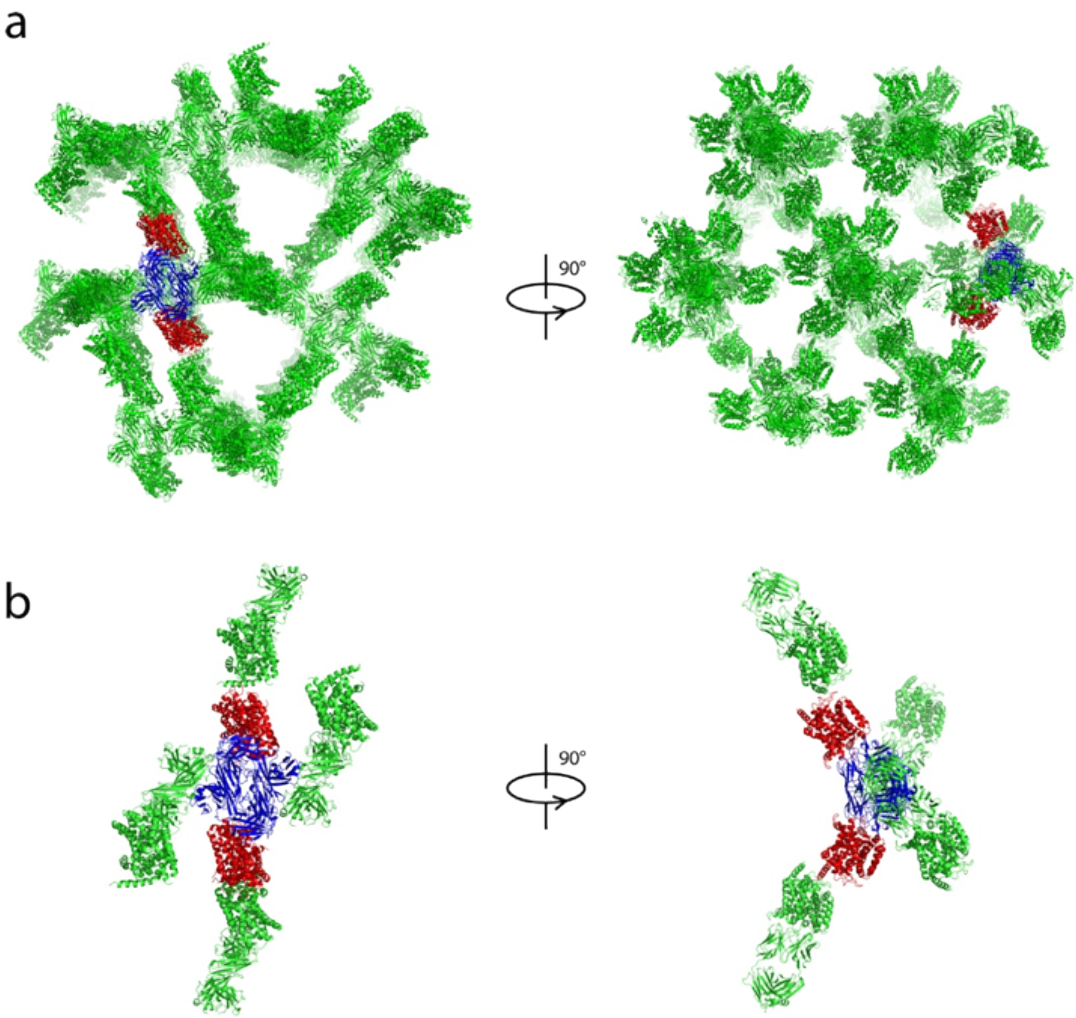

Fig. S2. Crystal packing and asymmetric unit of the AcgS-Fab complex. **(a)** Crystal packing showing AcgS-Fab complex lattice. One asymmetric unit is selected to show the building block that is composed of two Fab (blue) and two AgcS (red) molecules. **(b)** Interactions between AcgS and adjacent Fab fragments. The biologically functional contact is between the luminal loops of AgcS (red) and the complementary determining regions (CDRs) of the Fab (blue). One AcgS (red) also makes contacts with another AgcS molecules from next asymmetric unit. Fab fragments (blue) stacks tightly along the other axis, which are connected by AgcS molecules in a propeller-like head-to-side manner.

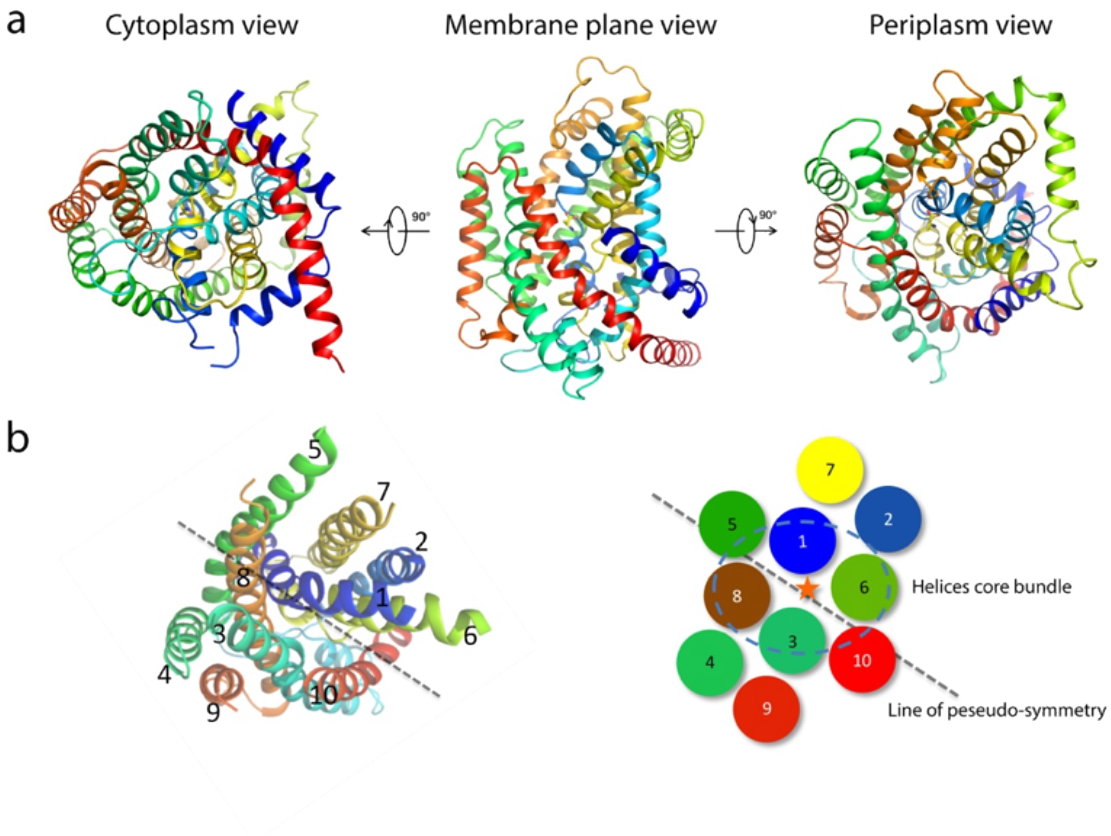

Fig. S3. (a) Overall structure of AgcS from three different views as indicated. The protein is colored by rainbow with the N- and C- termini in blue and red, respectively. **(b)** Schematics of helix packing of AgcS, viewed from the periplasm. The approximate binding site is indicated as an orange star and the rotation axis of pseudo-symmetry as a dotted gray line. The helices core bundle is highlighted by a blue dashed circle. TM1 is not shown in this figure.

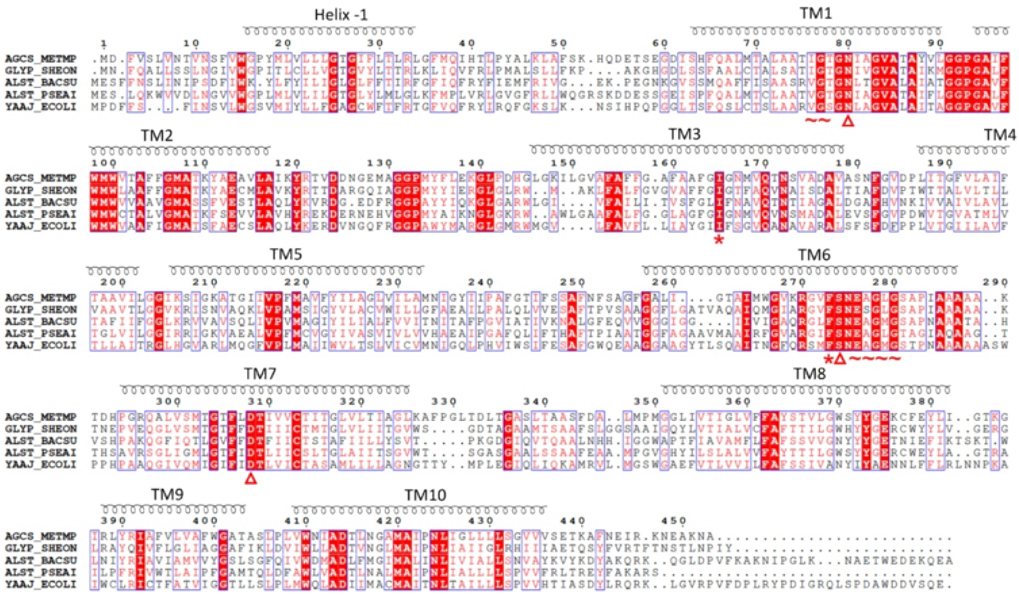

Fig. S4. Amino acid sequence alignment of *Methanococcus maripaludis* AgcS (AgcS_METMP) with other AgcS family members (GlyP from *Shewanella oneidensis*, AlsT from *Bacillus subtilis* and *Pseudomonas aeruginosa*, and YaaJ from *Escherichia coli*). Conserved residues are highlighted in red, and transmembrane helices are depicted as coils. The unwound regions in TM2 and TM7 are indicated by tildes. The open red triangles show residues involved in coordinating the sodium ion. The asterisks show the conserved residues forming hydrophobic face of binding pocket.

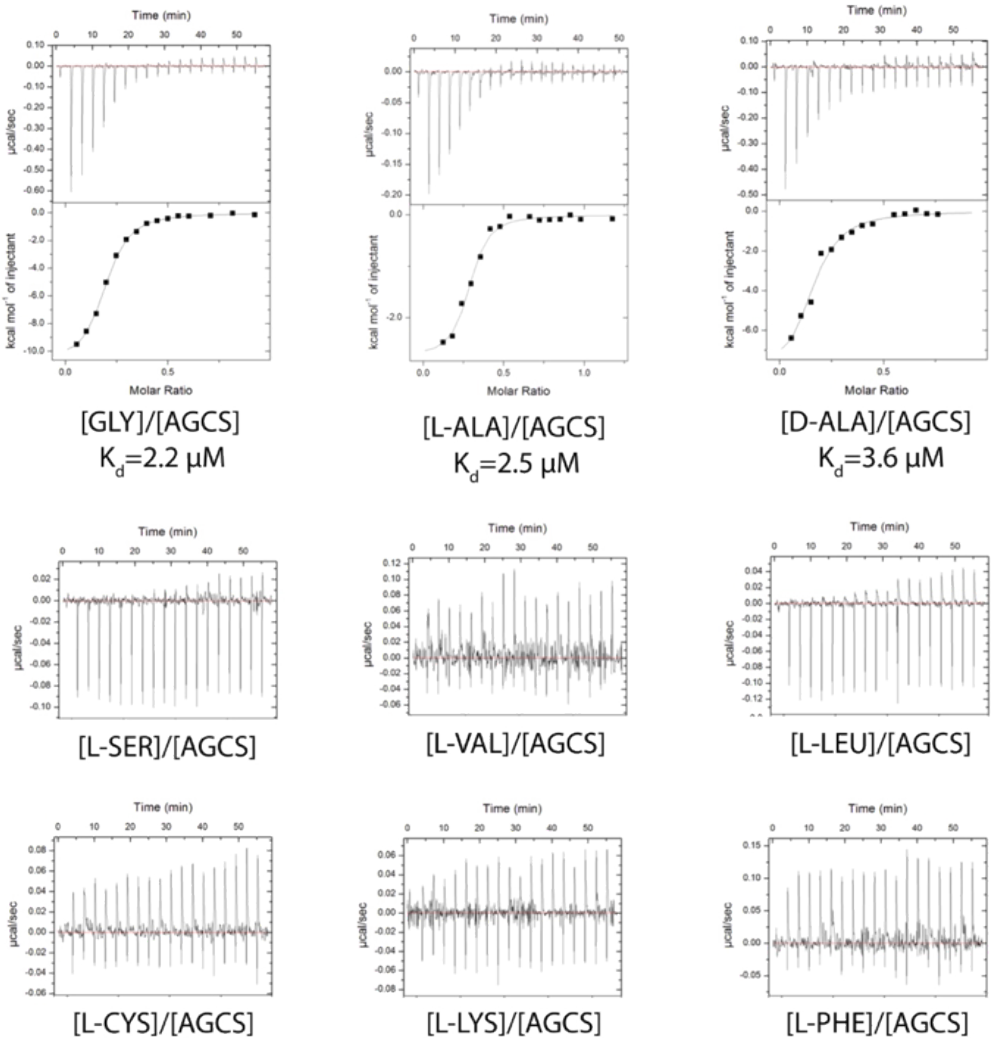

Fig. S5. Isothermal titration calorimetry of AgcS with a subset of amino acids at 25°C. The titration traces are shown in the top panels and the corresponding binding isotherms (calculated from the integrated injection heats and normalized to moles of injectant) are shown in the bottom panels.

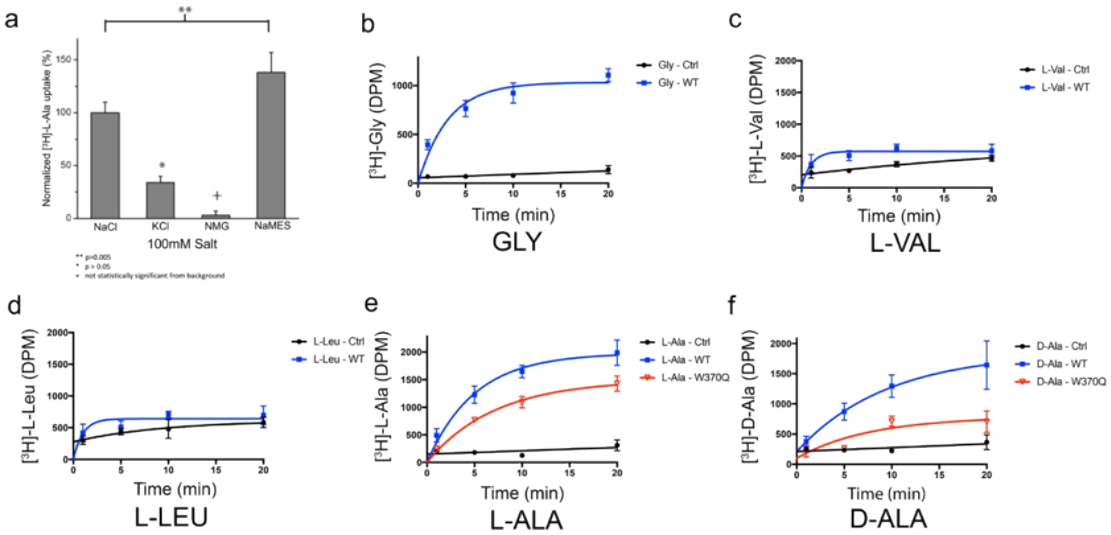

Fig. S6. (a) Ionic dependence of uptake to L-alanine of wild-type AgcS at 10 mins shows a preference for sodium. No uptake was apparent when cations were removed by the addition of NMG. **(b-f)** Time-dependent uptake of [^{3}H] labeled substrates by wild-type AgcS or W370Q mutant in proteinliposomes reconstituted with the purified protein fraction. Error bars represent standard error of the mean (s.e.m.) of triplicate experiments. Non-specific uptake was assessed by using protein-free liposomes under identical conditions.

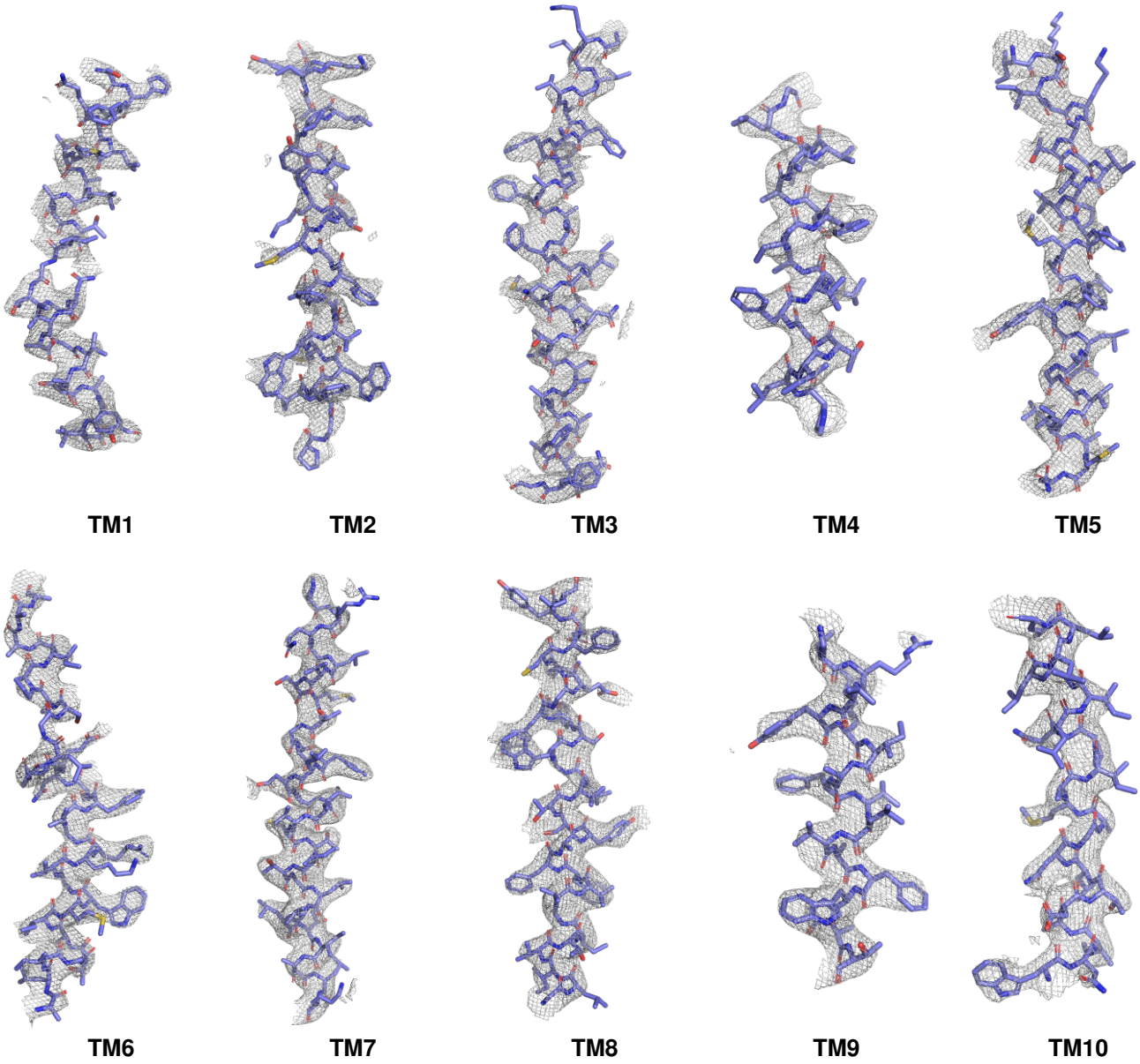

Fig. S7. Overall experimental density of AgcS membrane helices are shown with $2F_o$-F_c map contoured at 1.5 σ above the mean, carved at 2 Å around the molecule (gray mesh).

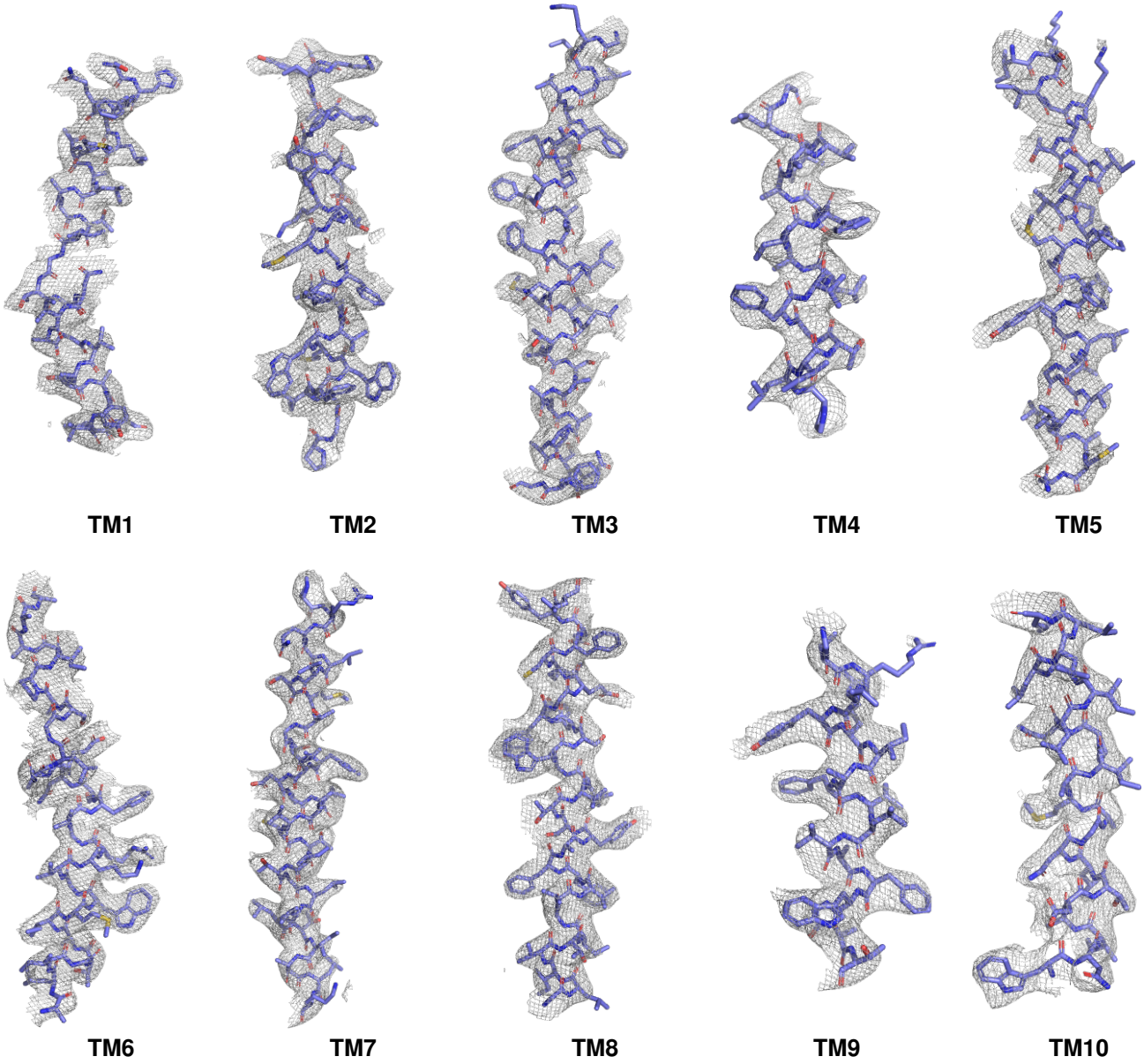

Fig. S8. Simulated annealing omit map of the AgcS membrane helices shown in Fig. S7, contoured at 1.5 σ above the mean, carved at 2 Å (gray mesh).

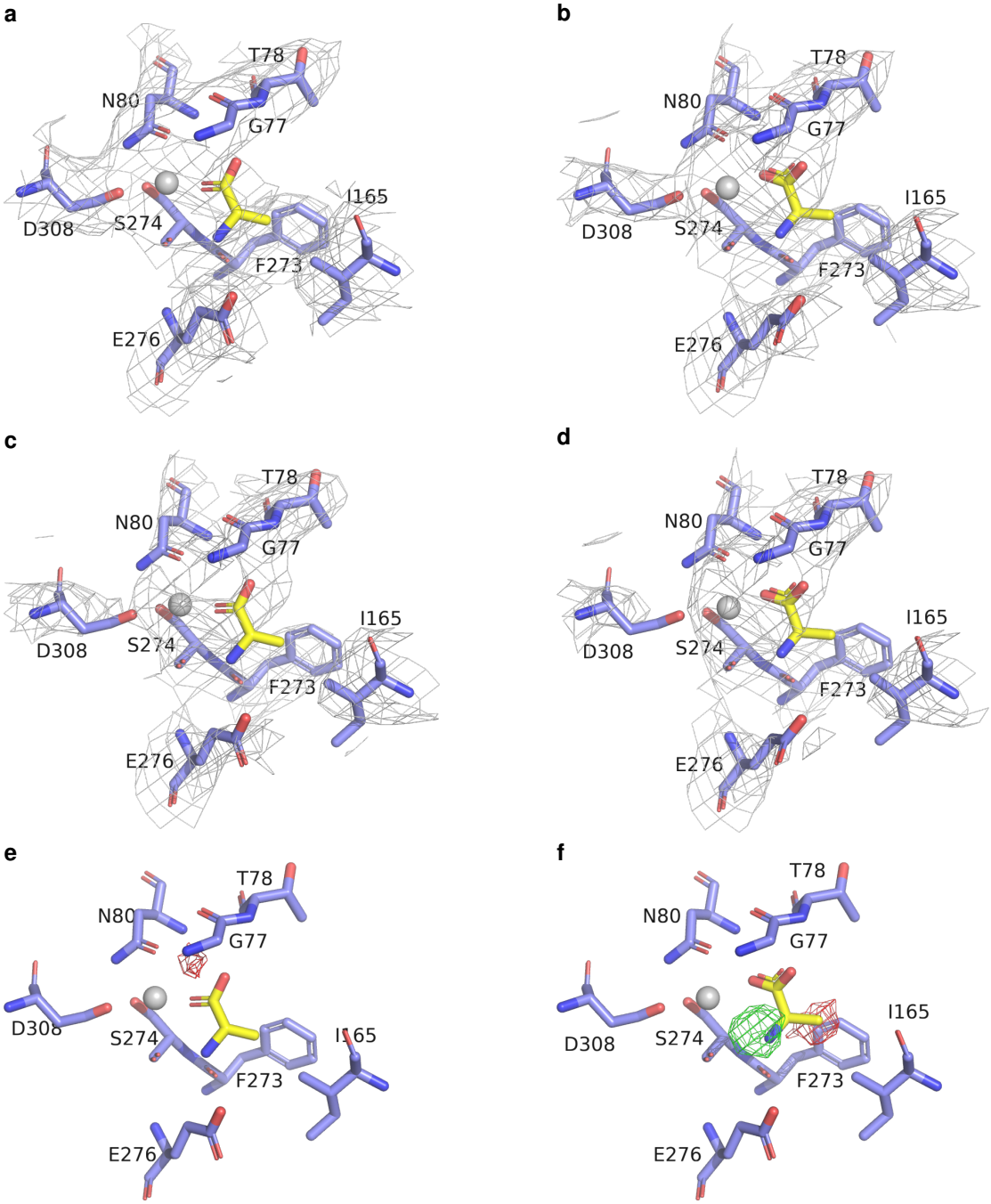

Fig. S9. L- and D-alanine in the AgcS binding pocket. 2*mF*o-*DF*c density in the binding pocket for L-alanine **(a)** and D-alanine **(b)**; the corresponding simulated annealing omit maps are shown in **(c)** and **(d)**. Densities are carved at 2.0 Å and contoured at 1.5 σ above the mean. The *mF*o-*DF*c density around the L-alanine is featureless **(e)**, but rigid-body refinement of D-alanine model against the L-alanine data yields approximately antisymmetric differences in the *mF*o-*DF*c density **(f)**. Densities in **(e)** and **(f)** are contoured at ±3.5 σ around the mean and carved at 3.5 Å from the alanine atoms.

Table S1. Crystallographic data collection, phasing and refinement statistics

	AgcS + Fab L-alanine	SeMet AgcS + Fab L-alanine	AgcS + Fab D-alanine
Data collection			
Space group	P 6_1	P 6_1	P 6_1
Cell dimensions			
a, *b*, *c* (Å)	183.5, 183.5, 349.1	184.9, 184.9, 351.5	183.1, 183.1, 349.8
α, β, γ (°)	90, 90, 120	90, 90, 120	90, 90, 120
Wavelength (Å)	0.9791	0.9791	0.9791
Resolution (Å)	158.90-3.24 (3.30-3.24)†	146.4-3.65 (3.71-3.65)	158.53-3.30 (3.36-3.30)
R_{pim}‡	0.099 (0.707)	0.063 (0.302)	0.099 (0.766)
$I/\sigma(I)$	8.7 (1.0)	10.9 (2.1)	6.9 (0.8)
$CC_{1/2}$	0.995 (0.363)	0.998 (0.740)	0.997 (0.354)
Completeness (%)	99.9 (100.0)	97.3 (98.2)	99.7 (100.0)
Redundancy	7.3 (7.8)	5.2 (5.1)	7.8 (8.3)
Phasing			
Sites		27	
Phasing Power			
Isomorphous		0.22	
Anomalous		0.45	
Refinement			
Resolution (Å)	29.91-3.24	29.91-3.65	29.91-3.30
No. reflections	104,943	72,913	99,167
R_{work} / R_{free}	0.234 / 0.254	0.226 / 0.231	0.242 / 0.256
No. of non-hydrogen atoms			
Protein	12,705	12,589	12,669
Ligand/ion	7	35	7
Average *B*-factors			
Protein	106.78	100.49	117.51
Ligand/ion	106.83	111.43	113.61
R.m.s. deviations			
Bond lengths (Å)	0.005	0.010	0.004
Bond angles (°)	0.83	1.30	0.80
Ramachandran			
favored (%)	94.80	90.40	94.39
outliers (%)	0.71	2.6	0.80

† Values in parentheses are for highest-resolution shell.

‡ Rpim is a measure of the quality of the data after averaging the multiple measurements and $R_{pim} = \Sigma_{hkl} [n/(n-1)]^{1/2} \Sigma_i |I_i(hkl) - \langle I(hkl) \rangle| / \Sigma_{hkl} \Sigma_i I_i(hkl)$, where n is the multiplicity, I_i is the intensity of the i^{th} observation, $\langle I \rangle$ is the mean intensity of the reflection and the summations extend over all unique reflections (hkl) and all equivalents (i), respectively(14).

Supplementary References

1. Hendrickson WA, Horton JR, LeMaster DM (1990) Selenomethionyl proteins produced for analysis by multiwavelength anomalous diffraction (MAD): a vehicle for direct determination of three-dimensional structure. *EMBO J* 9(5):1665–1672.
2. Harlow E, Lane D (1988) *Antibodies: A Laboratory Manual* (Cold Spring Harbor Laboratory Press).
3. Lim HH, Fang Y, Williams C (2011) High-efficiency screening of monoclonal antibodies for membrane protein crystallography. *PLoS One* 6(9):e24653.
4. Vonrhein C, et al. (2011) Data processing and analysis with the autoPROC toolbox. *Acta Crystallogr D Biol Crystallogr* 67(Pt 4):293–302.
5. Winn MD, et al. (2011) Overview of the *CCP* 4 suite and current developments. *Acta Crystallogr Sect D Biol Crystallogr* 67(4):235–242.
6. McCoy AJ, et al. (2007) Phaser crystallographic software. *J Appl Crystallogr* 40(Pt 4):658–674.
7. Zwart PH, et al. (2008) Automated structure solution with the PHENIX suite. *Methods Mol Biol* 426:419–435.
8. Adams PD, et al. (2010) PHENIX: A comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr Sect D Biol Crystallogr* 66(2):213–221.
9. Panjikar S, et al. (2009) On the combination of molecular replacement and single-wavelength anomalous diffraction phasing for automated structure determination. *Acta Crystallogr Sect D Biol Crystallogr* 65(10):1089–1097.
10. Emsley P, Lohkamp B, Scott WG, Cowtan K (2010) Features and development of Coot. *Acta Crystallogr D Biol Crystallogr* 66(Pt 4):486–501.
11. Blanc E, et al. (2004) Refinement of severely incomplete structures with maximum likelihood in *BUSTER–TNT*. *Acta Crystallogr Sect D Biol Crystallogr* 60(12):2210–2221.
12. G. B, et al. (2017) *BUSTER version 2.11.2* (Cambridge, United Kingdom: Global Phasing Ltd.).
13. Chen VB, et al. (2010) MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr D Biol Crystallogr* 66(Pt 1):12–21.
14. Weiss MS, IUCr (2001) Global indicators of X-ray data quality. *J Appl Crystallogr* 34(2):130–135.