An environmentally ultrasensitive 19F NMR probe for monitoring protein conformational equilibria

Yun Huang[1,2,*], Krishna Reddy[1], Clay Bracken[3], Wenhu Zhang[4], David Eliezer[3,*], Olga Boudker[1,2,*]

[1]Department of Physiology & Biophysics, Weill Cornell Medicine, 1300 York Ave, New York, NY 10021; [2]Howard Hughes Medical Institute, Chevy Chase; [3]Department of Biochemistry, Weill Cornell Medicine, 1300 York Ave, New York, NY 10021; [4]Department of Microbiology & Immunology, Weill Cornell Medicine, 1300 York Ave, New York, NY 10065, United States

* To whom correspondence should be addressed

Abstract

Limited chemical shift dispersion represents a major barrier to studying multi-state equilibria of large membrane proteins by ^{19}F NMR. We describe a novel monofluoroethyl ^{19}F probe with a dramatically larger chemical shift dispersion than existing probes. The improved conformational sensitivity and line shape enabled the detection of hidden states of a critical membrane transporter, guiding subsequent structural elucidation by single-particle cryo-electron microscopy.

Main text

Many membrane proteins, such as transporters, channels, and receptors, require conformational transitions to function. Insights into their conformational dynamics are therefore important for mechanistic understanding and therapeutic development[1-3]. NMR is widely used to probe protein dynamics[4,5], and one-dimensional (1D) ^{19}F NMR is especially popular in membrane protein studies because of the exquisite sensitivity of ^{19}F chemical shifts to conformational changes, the absence of background signals, and the applicability of ^{19}F NMR even in the absence of expensive high-field instruments[6-8]. In contrast to multi-dimensional NMR, 1D ^{19}F NMR is a single-pulse experiment with no heteronuclear magnetization transfer and can therefore quantify state populations of even weak or broad resonances. Aromatic fluorine and trifluoromethyl (-CF_3) groups are widely used ^{19}F probes. Aromatic fluorine has unfavorable relaxation properties due to restricted motion and is less applicable for large proteins. The -CF_3 group shows faster longitudinal and slower transverse relaxation rates due to fast rotation, leading to higher sensitivity[9]. However, its chemical shift dispersion typically does not exceed 2 ppm[10-13], resulting in severe resonance overlap in large membrane proteins. Such poor dispersion is a significant barrier to studying conformational dynamics.

Here, we report a cysteine-conjugated monofluoroethyl (mFE) label that exhibits narrow linewidths and is ultrasensitive to conformational changes. In aliphatic monofluorides, ^{19}F chemical shifts depend on the local electric field, Van der Waals interactions[14], and X-C-C-F dihedral angle[15]. Noncovalent interactions and steric constraints can favor different dihedral angles in different protein conformations, resulting in chemical shift dispersion reaching 12 ppm[15]. However, this property of mFE has not yet been exploited in monitoring protein conformational dynamics. Notably, fluorine chemical shift anisotropy (CSA) is approximately two times lower in mFE than in the trifluoroethyl (tFE) group[16]. Because ^{19}F transverse relaxation depends quadratically on the CSA, mFE might be expected to feature slower relaxation and narrower linewidths.

We synthesized a cysteine-specific ^{19}F labeling compound S-(2-fluoroethyl-1,1,2,2-D_4) *p*-toluenethiosulfonate ($TsSCD_2CD_2F$) from commercially available reagents (**Figure 1a**).

Deuterium is used to reduce the ^{19}F-^{1}H dipole relaxation and peak splitting due to ^{19}F-^{1}H spin-spin coupling.

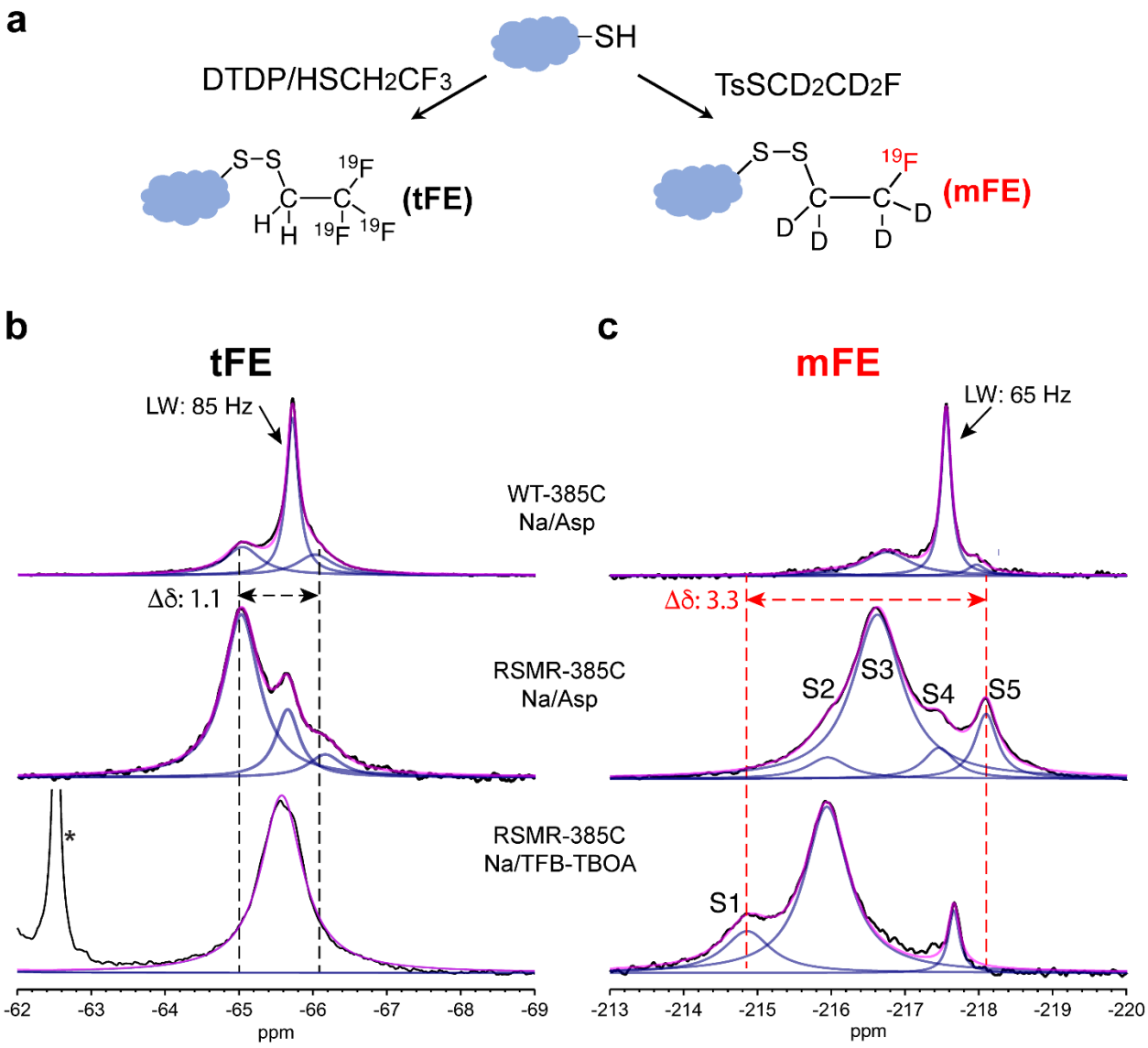


Figure 1. mFE-labeled Glt_{Ph} variants show wide chemical shift dispersion. (a) Site-specific Glt_{Ph} protein labeling (see Methods for details). (b, c) ^{19}F NMR spectra of Glt_{Ph}-M385C variants labeled with tFE (b) and mFE (c). Shown are from top to bottom: WT and RSMR mutant in 100 mM Na^+ and 2 mM aspartate (Asp), and RSMR mutant in 100 mM Na^+ and 0.6 mM TFB-TBOA. All spectra were recorded at 25 °C. Asterisk denotes signal from the CF_3 group of TFB-TBOA. S1-S5 denote resolved resonance peaks observed using the mFE probe, and $\Delta\delta$ is the largest observed chemical shift difference. All ^{19}F NMR spectra here and elsewhere were recorded at least twice on independently prepared protein samples, producing similar results.

We used Glt_{Ph} as a model membrane protein to test the sensitivity of the mFE probe chemical shift to conformational changes. Glt_{Ph} is an archaeal homolog of human glutamate transporters, which harnesses energy from sodium gradients to transport aspartate via an "elevator" mechanism[17]. It is a 134 kDa homotrimer with a protein-micelle particle size of about 300 kDa when solubilized in DDM[18]. During transport, Glt_{Ph} transitions between the outward- and inward-facing states (OFS and IFS, **Extended Data Fig. S1a**), where the substrate-binding site can open to the extracellular solution and cytoplasm, respectively. We compared 1D ^{19}F spectra of a cysteine-free Glt_{Ph} variant (termed wild-type, WT, for brevity; see Methods for details) mFE- or tFE-labeled at a single cysteine introduced by mutation (M385C) and a variant with additional gain-of-function mutations R276S/M395R (termed RSMR) that increase elevator dynamics[19] and speed up substrate transport ~ 8-fold [20].

Glt_{Ph} M385C retains aspartate uptake activity when labeled with tFE [13] or mFE and reconstituted into liposomes (**Extended Data Fig. S1b**). We recorded 1D ^{19}F NMR spectra in the presence of Na^+ ions and aspartate or a competitive blocker (3S)-3-[[3-[[4-(trifluoromethyl)benzoyl]amino]phenyl]methoxy]-L-aspartic acid (TFB-TBOA) (**Figure 1**). WT- and RSMR-M385C-tFE spectra exhibited 3 peaks with chemical shift dispersion of 1.1 ppm, as we previously reported[13]. In contrast, M385C-mFE spectra featured 5 peaks distributed over 3.3 ppm. Furthermore, when we lowered the temperature to slow conformational transitions, we observed 6 peaks distributed over 3.6 ppm for RSMR-M385C-mFE (**Extended Data Fig. S2**). In addition, the major sharp peak in WT-M385C-mFE (**Figure 1c**, upper panel), which appears to be less affected by conformational exchange, has a narrower linewidth of 65 Hz compared to 85 Hz for the corresponding peak in tFE-labeled protein. We also tested labeling at different positions, A380C and A381C. Both labeled proteins showed similar aspartate uptake as WT (**Extended Data Fig. S1**). The ^{19}F spectra of WT-A381C-mFE showed chemical shift dispersion of 3.3 ppm, wider than WT-A381C-tFE, and resolved one additional conformation (**Extended Data Fig. S3a**). Strikingly, WT-A380C-mFE featured chemical shift dispersion of 8.9 ppm with 7 well-separated peaks (**Extended Data Fig. S3b**), demonstrating the tremendous potential of the mFE probe. To our knowledge, this is the broadest chemical shift

dispersion observed and the highest number of conformations resolved using protein ^{19}F NMR.

We also observed that protonated mFE produced broader peaks than the deuterated mFE due to the two-bond ^{1}H-^{19}F coupling (47 Hz) (**Extended Date Fig. S4a, b**). However, deuteration significantly elongates the T_1 relaxation time (**Extended Date Fig. S4c**), increasing the recycle delay and reducing sensitivity. Therefore, protonated mFE may be preferable for applications where linewidth is not a limitation or if hardware for ^{1}H decoupling during ^{19}F detection is available.

Because of its environmental sensitivity, the mFE probe can report on changes of conformational equilibria due to ligand binding, mutations, temperature, pressure, and other factors. For example, we found that another competitive inhibitor DL-*threo*-β-benzyloxyaspartic acid (TBOA) stabilizes WT Glt_{Ph} and the RSMR mutant in different conformations (**Figure 2a**). The S4 resonance is dominant in TBOA-bound WT-M385C-mFE but is greatly diminished in the RSMR-M385C-mFE mutant, where the S2 resonance predominates (**Figure 2a**). Notably, solvent-exposed ^{19}F nuclei generally feature slower transverse relaxation than buried nuclei, resulting in narrower peaks. Crystal and cryo-electron microscopy (cryo-EM) structures show that residue M385 is solvent-exposed in the OFS and buried in the IFS [21,22] (**Extended Data Fig. S1**). Therefore, we tentatively assigned a prominent narrow S4 peak to an OFS and the broader S2 and S3 peaks to two different IFS conformations. This assignment of the S4 peak is consistent with the cryo-EM and crystal structures of TBOA-bound Glt_{Ph}, which is in an OFS. The excellent separation of the mFE resonances and their narrow linewidth allow assessment of the peak assignment by observing solvent paramagnetic relaxation enhancement (PRE)-mediated linewidth changes. Indeed, adding a paramagnetic reagent Gd-DTPT-BMA to the Na^+-bound WT transporter or TBOA-bound RSMR mutant significantly broadened the S4 but not the S2 and S3 peaks (**Extended Data Fig. S5**), supporting the proposed assignments.

Intrigued by the observation that the dominant state of TBOA-bound RSMR mutant (inward-facing S2) differs from the dominant state of the WT transporter (outward-facing S4), we imaged it by cryo-EM. Following particle alignment with imposed C3 symmetry,

we performed symmetry expansion and 3D classification without alignment to sort multiple conformations of the transporter protomers in the trimer[22,23]. We observed that the dominant conformation of the RSMR mutant is an IFS with a wide-open substrate-binding site occupied by TBOA (**Figure 2a**, **Extended Data Fig. S6,** and **Extended Data Table 1**; see **Methods** for data processing details). Only ~ 9 % of protomers were in an OFS similar to the previously resolved TBOA-bound WT [22] (RMSD of 0.646 Å). Because the dominant S2 and the minor S4 are the only peaks observed in the NMR spectrum, we ascribe them to the IFS and OFS structures revealed by cryo-EM, respectively. Notably, this S2 IFS conformation is also different from TBOA-bound WT constrained in IFS by crosslinking, where the substrate-binding site remained more closed[22] (**Extended Data Fig. S6c**). The new S2 IFS could therefore represent an intermediate in the substrate release process.

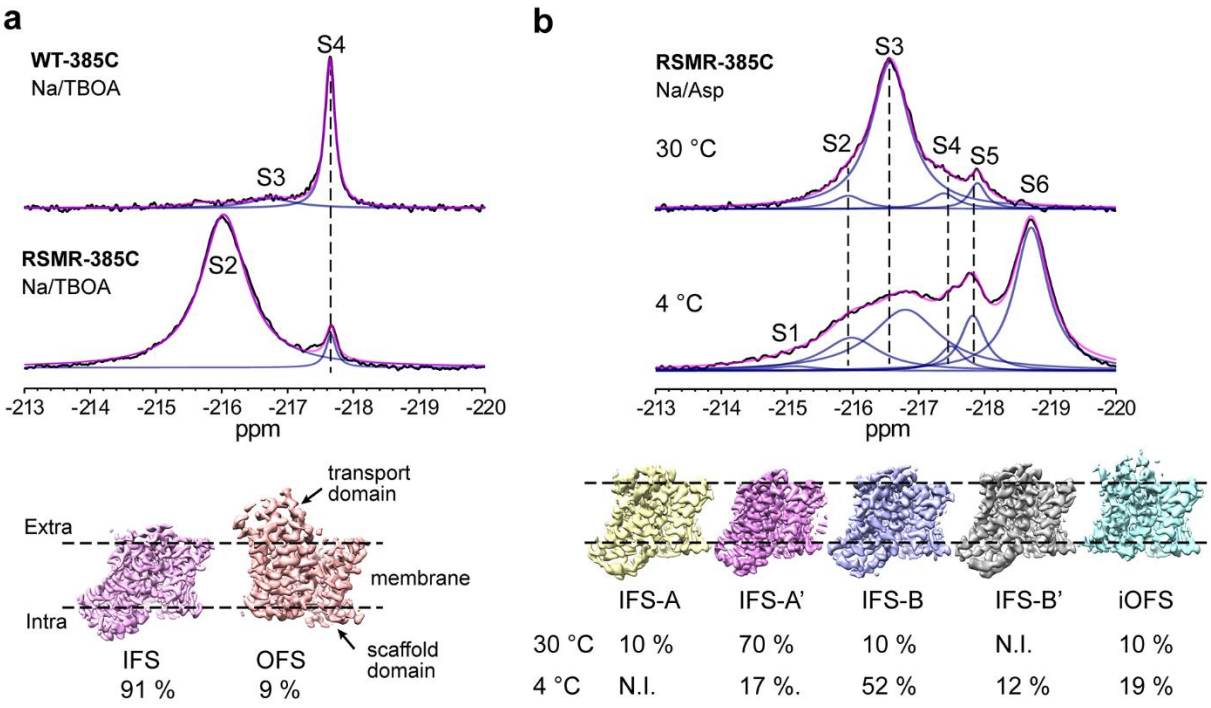


Figure 2. Identification and structural elucidation of new conformations of Glt_{Ph} by ^{19}F NMR and cryo-EM. (a) ^{19}F NMR spectra of mFE labeled WT-M385C (upper panel) and RSMR-M385C (middle panel) in the presence of 100 mM NaCl and 2 mM TBOA. The bottom panel shows maps of IFS (left) and OFS (right) cryo-EM structural classes of TBOA-bound RSMR and their populations. (b) ^{19}F NMR spectra of mFE-labeled RSMR-M385C in the presence of 500 mM NaCl and 2 mM aspartate recorded at 30 °C (upper panel) and 4 °C (lower panel). The bottom panel

shows maps of the representative cryo-EM structural classes and their populations at indicated temperatures. N.I. represents “not identified”.

^{19}F NMR using mFE labeling revealed that the RSMR mutant samples a surprising number of states, S1-S6, under transport conditions, in the presence of 100 mM NaCl and 2 mM aspartate (**Figure 2b**). Furthermore, temperature dramatically modulates the state populations (**Extended Data Fig. S2b**). Thus, above 25 °C, S3 is the dominant peak, and S6 is invisible. However, the S6 population increases at lower temperatures. Solvent PRE experiments suggest that S4 and S5 are OFS, and S2, S3, and S6 are IFSs (**Extended Data Fig. S7**). The S1 peak is too small to evaluate by PRE. To elucidate the temperature-dependent conformational changes suggested by NMR, we prepared cryo-EM grids of the RSMR mutant pre-incubated in 500 mM NaCl and 2 mM aspartate at 4 and 30 °C. 3D classification of protomers imaged at 30 °C (**Figure 2b** and **Extended Data Fig. S8, Extended Data Table 1**), revealed small protomer populations in the previously described intermediate OFS (iOFS, 10.4 % protomers)[13] and “unlocked” IFS (IFS-B, 9.5 % protomers), featuring a wide space between the scaffold and transport domains with intercalated detergent chains[19]. The majority of protomers formed an ensemble of states with varying positions of the transport domain similar to the “locked” IFS conformation (IFS-A and IFS-A’, 80 %)[24]. We observed mostly similar structural classes but significantly altered populations at 4 °C, consistent with changes in ^{19}F NMR spectra (**Figure 2b**). In particular, the “unlocked” IFS-B population increased to ~50 %. Approximately 10-20 % of protomers were in each IFS-A, iOFS, and another previously unobserved IFS-B’ conformation, intermediate between IFS-A and IFS-B (**Figure 2b** and **Extended Data Fig. S8**). Although the freezing process and data processing methods can affect population estimates[25], we observed marked correlations between cryo-EM and ^{19}F NMR results. The inward-facing S3 and S6 peaks correlate with the “locked”-like and “unlocked” conformations, respectively, and S5, the more populated of the two outward-facing resonances, likely corresponds to iOFS. The population of the OFS S4 resonance is likely too low to observe this state by cryo-EM. The behavior of the inward-facing S2 peak suggests it may correspond to IFS-B’ but this assignment may require additional data. Importantly, ^{19}F NMR visualizes the lowly populated states, typically invisible by classic NMR methods, providing a means to assess their populations and

thereby optimize conditions for their structural elucidation. Furthermore, our study suggests that despite the caveats associated with extracting population estimates from cryo-EM classifications, such estimates might correlate well with the populations present in solution prior to the freezing process.

In summary, mFE chemical shift dispersion significantly exceeds that of tFE, enabling a more detailed and quantitative description of protein conformational ensembles and guiding cryo-EM imaging. Improved peak separation should also enable measurements of chemical exchange rates by methods such as EXSY [13] and saturation transfer, including CEST [26]. Continuously expanding structural methodologies reveal that proteins populate diverse ensembles of conformations. For example, a G protein-coupled adenosine A_{2A} receptor samples at least 5 distinguishable states during activation[27]. Our model protein is similar in size to many physiologically important receptors, transporters, and ion channels, and the new ^{19}F probe should greatly facilitate mechanistic studies of such proteins.

Acknowledgements

The authors thank Dr. M. Goger and Dr. S. Bhattacharya for help with setting up NMR. We thank Dr. B. Wang, Dr. A. Paquette and Dr. B. Williams at the NYU cryo-EM facility for help with electron microscopy data collection. We also thank W. Eng and E. Wu for assistance with protein expression, Dr. B. Qiu for help with making cryo-EM samples and model building, and Dr. Z. Shen for assistance with chemical synthesis. This work was supported by NIH grants R37NS085318 and R01NS064357 (O.B.), R37AG019391 and R35 GM136686 (D.E.) and S10OD016320 (C.B.). O.B., D.E. and C.B. are members of the New York Structural Biology Center (NYSBC) which is supported in part by NIH Grant P41 GM118302 (CoMD/NMR: Center on Macromolecular Dynamics by NMR Spectroscopy), ORIP/NIH facility improvement grant CO6RR015495 and NIH grant S10OD018509.

Author contributions

Y.H., D.E. and O.B. designed the experiments; Y.H. performed the NMR experiments; Y.H. and O.B analyzed the data; C.B. assisted with NMR experimental data collection

and analysis; Y. H and W. Z conducted the chemical synthesis; Y. H and K. R. performed cryo-EM imaging, analyzed data, and refined molecular models; K. R. performed transport activity assays; Y.H., K. R., O.B. and D.E. wrote the manuscript.

Competing interests

The authors declare no competing interests.

References

1 Lee, Y., Lazim, R., Macalino, S. J. Y. & Choi, S. Importance of protein dynamics in the structure-based drug discovery of class A G protein-coupled receptors (GPCRs). *Curr. Opin. Struct. Biol.* **55**, 147-153 (2019).

2 Peng, J. W. Communication Breakdown: Protein Dynamics and Drug Design. *Structure* **17**, 319-320 (2009).

3 Sekhar, A. & Kay, L. E. An NMR View of Protein Dynamics in Health and Disease. *Annu. Rev. Biophys.* **48**, 297-319 (2019).

4 Lerner, E. *et al.* Toward dynamic structural biology: Two decades of single-molecule Förster resonance energy transfer. *Science* **359** (2018).

5 Palmer, A. G. NMR Characterization of the Dynamics of Biomacromolecules. *Chem. Rev.* **104**, 3623-3640 (2004).

6 Danielson, M. A. & Falke, J. J. Use of ^{19}F NMR to probe protein structure and conformational changes. *Annu. Rev. Biophys. Biomol. Struct.* **25**, 163-195 (1996).

7 Didenko, T., Liu, J. J., Horst, R., Stevens, R. C. & Wuthrich, K. Fluorine-19 NMR of integral membrane proteins illustrated with studies of GPCRs. *Curr. Opin. Struct. Biol.* **23**, 740-747 (2013).

8 Di Pietrantonio, C., Pandey, A., Gould, J., Hasabnis, A. & Prosser, R. S. Understanding Protein Function Through an Ensemble Description: Characterization of Functional States by ^{19}F NMR *Methods Enzymol.* **615**, 103-130 (2019).

9 Rashid, S., Lee, B. L., Wajda, B. & Spyracopoulos, L. Side-Chain Dynamics of the Trifluoroacetone Cysteine Derivative Characterized by ^{19}F NMR Relaxation and Molecular Dynamics Simulations. *J. Phys. Chem. B* **123**, 3665-3671 (2019).

10 Liu, J. J., Horst, R., Katritch, V., Stevens, R. C. & Wuthrich, K. Biased signaling pathways in beta2-adrenergic receptor characterized by ^{19}F-NMR. *Science* **335**, 1106-1110 (2012).

11 Manglik, A. *et al.* Structural Insights into the Dynamic Process of beta2-Adrenergic Receptor Signaling. *Cell* **161**, 1101-1111 (2015).

12 Frei, J. N. *et al.* Conformational plasticity of ligand-bound and ternary GPCR complexes studied by ^{19}F NMR of the beta1-adrenergic receptor. *Nat. Commun.* **11**, 669 (2020).

13 Huang, Y. *et al.* Use of paramagnetic ^{19}F NMR to monitor domain movement in a glutamate transporter homolog. *Nat. Chem. Biol.* **16**, 1006-1012 (2020).

14 de Dios, A. C. & Oldfield, E. Evaluating ^{19}F Chemical Shielding in Fluorobenzenes: Implications for Chemical Shifts in Proteins. *J. Am. Chem. Soc.* **116**, 7453-7454 (1994).

15 Feeney, J. *et al.* 19F Nuclear Magnetic Resonance Chemical Shifts of Fluorine Containing Aliphatic Amino Acids in Proteins: Studies on *Lactobacillus casei* Dihydrofolate Reductase Containing (2S,4S)-5-Fluoroleucine. *J. Am. Chem. Soc.* **118**, 8700-8706 (1996).

16 Grage, S. L. *et al.* Solid state ^{19}F NMR parameters of fluorine-labeled amino acids. Part II: aliphatic substituents. *J. Magn. Reson.* **191**, 16-23 (2008).

17 Drew, D. & Boudker, O. *Annual. Rev. Biochem.* **85**, 543-572 (2016).

18 Yernool, D., Boudker, O., Folta-Stogniew, E. & Gouaux, E. Trimeric Subunit Stoichiometry of the Glutamate Transporters from Bacillus caldotenax and Bacillus stearothermophilus. *Biochemistry* **42**, 12981-12988 (2003).

19 Akyuz, N. *et al.* Transport domain unlocking sets the uptake rate of an aspartate transporter. *Nature* **518**, 68-73 (2015).

20 Huysmans, G. H. M., Ciftci, D., Wang, X., Blanchard, S. C. & Boudker, O. The high-energy transition state of the glutamate transporter homologue GltPh. *EMBO J.* **40** (2021).

21 Yernool, D., Boudker, O., Jin, Y. & Gouaux, E. Structure of a glutamate transporter homologue from Pyrococcus horikoshii. *Nature* **431**, 811-818 (2004).

22 Wang, X. & Boudker, O. Large domain movements through the lipid bilayer mediate substrate release and inhibition of glutamate transporters. *eLife* **9**, e58417 (2020).

23 Reddy, K. D., Ciftci, D., Scopelliti, A. & Boudker, O. Heterogeneous substrate binding in a glutamate transporter homologue. *bioRxiv*, doi:10.1101/2021.07.08.451670 (2021).

24 Reyes, N., Ginter, C. & Boudker, O. Transport mechanism of a bacterial homologue of glutamate transporters. *Nature* **462**, 880-885 (2009).

25 Geraets, J. A., Pothula, K. R. & Schröder, G. F. Integrating cryo-EM and NMR data. *Curr. Opin. Struct. Biol.* **61**, 173-181 (2020).

26 Gao, J. *et al.* Fluorine Pseudocontact Shifts Used for Characterizing the Protein-Ligand Interaction Mode in the Limit of NMR Intermediate Exchange. *Angew. Chem. Int. Ed.* **56**, 12982-12986 (2017).

27 Huang, S. K. *et al.* Delineating the conformational landscape of the adenosine A_{2A} receptor during G protein coupling. *Cell* **184**, 1884-1894 (2021).

Methods

Synthesis of S-(2-fluoroethyl-1,1,2,2-D_4) *p*-toluenethiosulfonate

To a solution of potassium *p*-toluenethiosulfonate (678 mg, *TCI Chemical*) in dry acetonitrile (10 ml) was added 2-bromoethanol-1,1,2,2-d_4 (240 μl, *Cambridge Isotope Laboratories*) under argon. The mixture was stirred at 60 °C overnight. Solvent was evaporated under reduced pressure, and the residue was dissolved with 50 ml ethyl acetate and washed with 0.5 M HCl aqueous solution (2 × 20 ml) and brine. The organic layer was dried over anhydrous Na_2SO_4 and then evaporated under reduced pressure to give S-(2-hydroxyethyl-1,1,2,2-D_4) *p*-toluenethiosulfonate (600 mg, 85% yield) as an oil. The product was directly used for the next step without further purification.

To a solution of S-(2-hydroxyethyl-1,1,2,2-D_4) *p*-toluenethiosulfonate (480 mg, 2 mmol) in dry dichloromethane (5 ml) pre-cooled to -70 °C was added 1.5 equiv. diethylaminosulfur trifluoride (3 ml of 1 M solution in dichloromethane, *Sigma-Aldrich*) dropwise under argon. After 10 min of stirring at -70 °C, the flask was warmed to room temperature for 30 min. 0.5 ml of methanol was added to quench the reaction. The reaction mixture was evaporated to dryness and then purified by flash column (*RediSep Rf*) using 0-60% petroleum ether/ethyl acetate. The major fraction was collected and evaporated to obtain the pure title compound (220 mg, 45% yield). Identity of the compound was confirmed by NMR. ^{1}H NMR (500 MHz, $CDCl_3$) δ 7.82 (d, J = 8.4 Hz, 2H), δ 7.36 (d, J = 7.7 Hz, 2H), δ 2.46 (s, 3H); ^{13}C NMR (125 MHz, $CDCl_3$) δ 145.40, 141.85, 130.18, 127.24, 81.44-79.32 (J = 171.19 Hz), 34.88 (m), 31.84; ^{19}F NMR (470 MHz, $CDCl_3$) δ -215.25 (m).

Synthesis of S-(2-fluoroethyl) *p*-toluenethiosulfonate

To a solution of potassium *p*-toluenethiosulfonate (620 mg, 2.7 mmol, *TCI Chemical*) in dry acetonitrile (10 ml) was added 1-bromo-2-fluoroethane (203 μl, *Accela ChemBio Inc*.) under argon. The mixture was stirred at 60 °C overnight. Solvent was removed under reduced pressure and the residue was dissolved with 50 ml ethyl acetate and washed with 0.5 M HCl aqueous solution (2 × 20 ml) and brine. The organic layer was dried over anhydrous Na_2SO_4 and then evaporated under reduced pressure. The resulted residue was purified by flash column chromatography to give the pure title compound (503 mg, 86% yield) as an oil. Identity of the compound was confirmed by NMR. ^{1}H NMR (500 MHz, $CDCl_3$) δ 7.82 (d, *J* = 8.2 Hz, 1H), 7.36 (d, *J* = 8.1 Hz, 1H), 4.61 (t, *J* = 6.3 Hz, 1H), 4.52 (t, *J* = 6.3 Hz, 1H), 3.30 (t, *J* = 6.3 Hz, 1H), 3.26 (t, *J* = 6.3 Hz, 1H), 2.46 (s, 3H); ^{13}C NMR (126 MHz, $CDCl_3$) δ 145.22, 141.72, 130.01, 127.11, 81.01 (d, *J* = 172.6 Hz), 35.38 (d, *J* = 22.7 Hz), 21.70; ^{19}F NMR (471 MHz, $CDCl_3$) δ -213.51 (tt, *J* = 46.9, 20.4 Hz).

Protein expression, purification, and labeling

Glt_{Ph} constructs, Y215H/E219H/C321A/M385C (WT-M385C), Y215H/E219H/C321A/A380C (WT-A380C), Y215H/E219H/C321A/A381C (WT-A381C), and a gain-of-function mutant with additional R276S/M395R within Y215H/E219H/C321A/M385C background (RSMR-M385C) were expressed, purified, and labeled as described, with modifications[13,21]. In brief, pBAD plasmids encoding the corresponding constructs with C-terminal thrombin cleavage site followed by $(His)_8$-tag were transformed into *E. coli* DH10-B cells (*Invitrogen*). Cells were grown in LB media at 37 °C to OD_{600} of 1.1. At this point, protein expression was induced by 0.2 % arabinose

(*Goldbio*), the temperature was set to 30 °C, and cells were grown for additional 16 hours. The cells were harvested by centrifugation and re-suspended in 20 mM Hepes, pH7.4, 200 mM NaCl, 1 mM L-asp, 1 mM EDTA. The suspended cells were broken by Emulsiflex C3 high-pressure homogenizer (*Avestin Inc*.) in the presence of 0.5 mg/ml lysozyme (*Goldbio*) and 1 mM Phenylmethanesulfonyl fluoride (PMSF, *MP Biomedicals*). After centrifugation for 15 min at 5000 g, the supernatant was centrifuged at 125000 g for 50 min. For the mFE labeling, the membrane pellets were collected and solubilized in a buffer containing 20 mM HEPES, pH7.4, 200 mM NaCl, 1 mM L-asp, 10 mM 2-mercaptoethanol, 40 mM n-dodecyl-β-D-maltopyranoside (DDM, *Anatrace, Inc*.) for 2 hours at 4 °C. The mixture was centrifuged for 50 min at 125,000 g. The supernatant was diluted 3 times with buffer A (20 mM HEPES, pH7.4, 200 mM NaCl, 1 mM L-asp) and incubated with Ni-NTA resin (*Qiagen*) for 1 hour at 4 °C. The resin was washed with 6 volumes of Buffer A supplemented with 1 mM DDM and 25mM imidazole and proteins were eluted in buffer A supplemented with 1 mM DDM and 300 mM imidazole. EDTA was added to the collected protein fractions to a final concentration of 0.5 mM. The protein was concentrated to ca. 10 mg/ml using concentrators with 100 kDa MW cutoff (*Amicon*). Protein concentration was determined by UV absorbance at 280 nm using the extinction coefficient of 57400 M^{-1} cm^{-1} and MW of 44.7 kDa (protomer). 2 molar equivalents of $TsSCD_2CD_2F$ or $TsSCH_2CH_2F$, prepared as stock solutions in DMSO, were added to the protein samples followed by incubation at 4 °C for 2 hours or room temperature for 1 hour. Thrombin was added and incubated overnight at room temperature to cleave the $(His)_8$-tag. For tFE labeling, the membrane pellet was solubilized with the same buffer as mFE labeling but supplemented with 2mM 2,2'-dithiodipyridine (DTDP, *Sigma Aldrich*). The protein-resin

slurry was first washed with 5 volumes of Buffer A supplemented with 1 mM DDM. 2 mM trifluoroethanethiol (*Sigma Aldrich*) was added, and the slurry was incubated with mixing at 4 °C overnight. The resin was washed with buffer A supplemented with 1 mM DDM and 25mM imidazole and then eluted with buffer A supplemented with 1 mM DDM and 300 mM imidazole. The $(His)_8$-tag was cleaved using thrombin at room temperature overnight. Both mFE- and tFE- labeled proteins were further purified by size exclusion chromatography (SEC) using Superdex 200 Increase 10/300 GL column (*GE Healthcare Life Sciences*) in a buffer containing 20 mM HEPES, pH7.4, 50 mM KCl, 1 mM DDM. 100 mM NaCl was added to the protein fractions immediately. The protein was concentrated and supplemented with additional NaCl and ligands as needed.

Transport activity assay

Unlabeled wild-type and mFE-labeled Glt_{Ph} were concurrently reconstituted into liposomes, and initial rates of ^3H L-Asp uptake were measured as previously described [28]. Briefly, liposomes were prepared from a 3:1 (w/w) mixture of *E. coli* polar lipids and egg yolk phosphotidylcholine (*Avanti Polar Lipids*) in 20 mM HEPES/Tris, pH 7.4, containing 200 mM KCl and 100 mM choline chloride. Liposomes were destabilized by addition of Triton X-100 at detergent to lipid ratio of 0.5:1 (w/w). Glt_{Ph} proteins were added at final protein to lipid ratio of 1:2000 (w/w) and incubated for 30 min at room temperature. Detergents were removed with Bio-Beads™ SM-2 resin (*Bio-Rad*), with two incubations at room temperature, one overnight incubation at 4 °C, and two more incubations at room temperature. The proteoliposomes were concentrated to 50 mg/mL and flash-frozen in liquid N_2. On the day of the experiment, liposomes were thawed and freshly extruded through 400 nm filters. The uptake reaction was started by diluting the proteoliposomes

100-fold into reaction buffer containing 20 mM HEPES/Tris pH 7.4, 200 mM KCl, 100 mM NaCl, 1 μM ^{3}H-L-Asp (*PerkinElmer*), and 0.5 μM valinomycin. Uptake was performed at 2 minutes at 35 °C. Reactions were quenched 10-fold in cold buffer containing lithium (20 mM HEPES/Tris pH 7.4, 200 mM KCl, 100 mM LiCl). Data is composed of the results from three independent liposome reconstitutions. Within each set of reconstitutions, measurements of the mutants were normalized to wild-type protein. Each data point is the average of at least two technical replicates.

^{19}F NMR spectroscopy

^{19}F NMR spectra were collected on Bruker Avance III HD 500 MHz spectrometer equipped with a TCI ^{1}H-^{19}F/^{13}C/^{15}N triple resonance cryogenic probe (*Bruker Instruments*) tuned for ^{19}F. For mFE-labeled proteins, 50 μM 2-fluoroethanol and 10 % D_2O were added to the sample and used as a chemical shift reference (-224.22 ppm) and a lock signal, respectively. Typically, 160 μl sample was loaded into 4 mm Shigemi tube. 1D ^{19}F NMR spectra were recorded using the standard ZG pulse in the Bruker pulse library, with 2096 points recorded and the spectral width (SW) of 40 ppm. The carrier frequency was set at -220 ppm. The recycle delay was set to 1.5 and 0.9 s for deuterated and protonated mFE, respectively, except when otherwise indicated. For tFE-labeled proteins, 50 μM trifluoroacetic acid was added to the sample and used as chemical shift reference (-76.55 ppm), carrier frequency was set at -70 ppm, and the recycle delay was set to 0.6 s. All the spectra were recorded without decoupling.

^{19}F longitudinal relaxation times (T_1) were measured by the inversion recovery method. The recycle delays were set to 5 s and 2.5 s for deuterated and protonated mFE probes,

respectively. The peak intensities were obtained from spectral deconvolution. The T_1 relaxation times and standard errors were obtained by fitting data to a single exponential function $I = I_0\left(1 - 2exp(-1/T_1 t)\right)$, where *I* is the peak intensity, and *t* is the delay time. All experiments were repeated at least twice.

All 1D ^{19}F NMR spectra were processed using MestRaNova 12.0.0 software (*Mestrelab Research*) employing a 20 Hz exponential windows function except where otherwise indicated, zero filling to 8192 points. The spectra were baseline-corrected, and the peaks were fitted to Lorentzian peak shapes. The linewidth values were obtained by subtracting the line broadening value from the fitting linewidth.

Cryo-EM data collection

To prepare cryo-grids, 3.5 µL of labeled Glt_{Ph} protein (4.5 mg/mL) was applied to a glow-discharged QuantiFoil R1.2/1.3 300-mesh gold grid (*Electron Microscopy Sciences*) and incubated for 2 min under 100 % humidity at desired temperatures. Following incubation, grids were blotted for 3 s at 0 blot force and plunge frozen in liquid ethane using a Vitrobot Mark IV (FEI). Cryo-EM imaging data were acquired on a Titan Krios microscope (*Thermo Fisher Scientific*) operated at 300 kV with a K3 Summit direct electron detector (*Gatan*). Automated data collection was carried out in super-resolution counting mode with a magnification of 105,000 x, which corresponds to a calibrated pixel size of 0.852 Å on the specimen and 0.426 Å for super-resolution images. An energy slit width of 20 eV was used throughout the collection. For the TBOA-bound RSMR sample, movies were collected using Leginon [29] at a total dose of 52.88 $e^-/Å^2$ distributed over 48 frames (1.102 $e^-/Å^2$/frame) with an exposure time of 2.40 s (50 ms/frame) and a defocus range of 1.3 − 2.0 µm. A total of 8100 movies were collected. For the Asp-bound RSMR grids made

at 4 °C, movies had a total dose of 50.94 $e^-/Å^2$ distributed over 48 frames (1.061 $e^-/Å^2$/frame) with an exposure time of 2.40 s (50 ms/frame) and a defocus range of 1.3 – 1.5 µm. A total of 5267 movies were collected. For the Asp-bound RSMR grids prepared at 30 °C, movies had a total dose of 56.04 $e^-/Å^2$ distributed over 40 frames (1.401 $e^-/Å^2$/frame) with an exposure time of 1.6 s (40 ms/frame) and a defocus range of 1.3 – 1.6 µm. A total of 5823 movies were collected.

Image processing

The frame stacks were motion corrected using MotionCorr2 [30] with 2x binning, and contrast transfer function (CTF) estimation was performed using CTFFIND4.1 [31]. Further processing steps were carried out using RELION 3.0.8 or 3.1.0 and cryoSPARC 3.0 or 3.2 [32,33]. Particles were picked from micrographs using the Laplacian-of-Gaussian (LoG) picker, aiming for ~2000 picks per micrograph. These particles were extracted using a box size of 300 pixels with 4x binning and imported into cryoSPARC. Following one round of 2D classification to remove artifacts, the particles underwent 3 rounds of heterogeneous refinement in C1 using eight total classes. Seven of these were noisy volumes created by one iteration of *ab initio*, and one was an unmasked 3D model obtained from *ab initio*. The particles were converted back to Relion format via PyEM [34], and re-extracted at full box size. These particles were reimported into cryoSPARC, underwent one round of non-uniform (NU) refinement using C3 symmetry [35]. Dynamic mask threshold, dynamic mask near, and dynamic mask far parameters were set to 0.2, 12, and 28, respectively. These particles were converted back to Relion format and underwent Bayesian polishing with parameters obtained using 5,000 random particles [36]. The polished particles were reimported into cryoSPARC and subjected to one round of

NU refinement in C3 with both local and global CTF refinement options on. The particles were again polished in Relion using an expanded box size of 384 pixels, reimported into cryoSPARC, and subjected to one round of NU refinement with C3 symmetry and local and global CTF refinement options turned on. The particles were then C3 symmetry expanded and subjected to a focused 3D classification in Relion. The local mask was generated by UCSF Chimera [37] using a combination of OFS (chain A of PDB model 2NWX) and IFS (chain A of PDB model 3KBC). Population of a certain conformation was calculated by dividing the particle number of an individual class or similar classes by the total number of C3 symmetry expanded particles. The particles from the individual classes were imported separately into cryoSPARC and subjected to a local refinement, using the mask and map obtained from the most recent NU refinement.

Model building and refinement

Crystal or cryo-EM structures were used as initial models and docked into the protein densities using UCSF Chimera. The models were first real-space refined in PHENIX [38]. Mis-aligned regions were manually rebuilt, and missing side chains and residues were added in COOT[39]. Models were iteratively refined applying secondary structure restraints and validated using Molprobity[40]. To cross-validate models, refined models were randomly displaced by an average of 0.3 Å, and each resulting model was refined against the first half-map obtained from processing. FSC between the refined models and the first half-maps were calculated and compared to the FSC of the other half-maps. The resulting FSC curves were similar, showing no evidence of overfitting. The structural figures were prepared in PyMOL (DeLano Scientific).

Extended references

28 Boudker, O., Ryan, R. M., Yernool, D., Shimamoto, K. & Gouaux, E. Coupling substrate and ion binding to extracellular gate of a sodium-dependent aspartate transporter. *Nature* **445**, 387-393 (2007).

29 Suloway, C. *et al.* Automated molecular microscopy: the new Leginon system. *J. Struct. Biol.* **151**, 41-60 (2005).

30 Zheng, S. Q. *et al.* MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat. Methods* **14**, 331-332 (2017).

31 Rohou, A. & Grigorieff, N. CTFFIND4: Fast and accurate defocus estimation from electron micrographs. *J. Struct. Biol.* **192**, 216-221 (2015).

32 Zivanov, J. *et al.* New tools for automated high-resolution cryo-EM structure determination in RELION-3. *Elife* **7** (2018).

33 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).

34 Asarnow, D., Palovcak, E., Cheng, Y. UCSF pyem Zenodo https://doi.org/10.5281/zenodo.3576630 (2019).

35 Punjani, A., Zhang, H. & Fleet, D. J. Non-uniform refinement: adaptive regularization improves single-particle cryo-EM reconstruction. *Nat. Methods* **17**, 1214-1221 (2020).

36 Zivanov, J., Nakane, T. & Scheres, S. H. W. A Bayesian approach to beam-induced motion correction in cryo-EM single-particle analysis. *IUCrJ* **6**, 5-17 (2019).

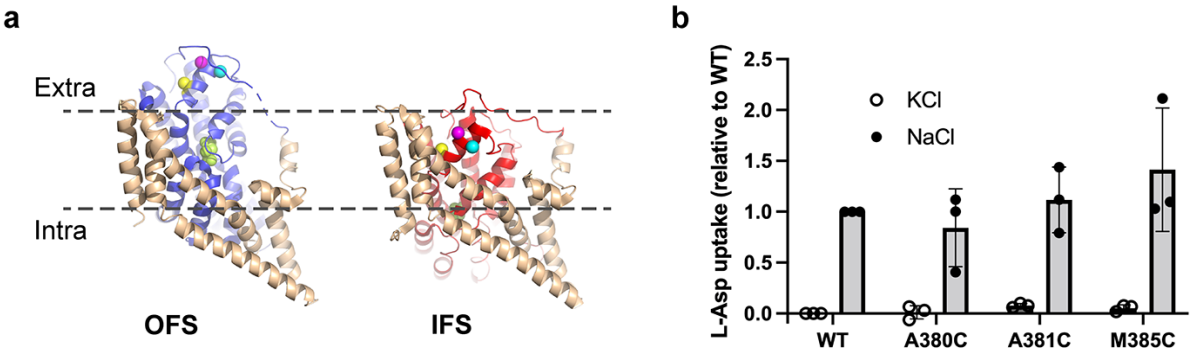
37 Pettersen, E. F. *et al.* UCSF Chimera—A visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605-1612 (2004).

38 Afonine, P. V. *et al.* phenix.model_vs_data: a high-level tool for the calculation of crystallographic model and data statistics. *J Appl. Crystallogr.* **43**, 669-676 (2010).

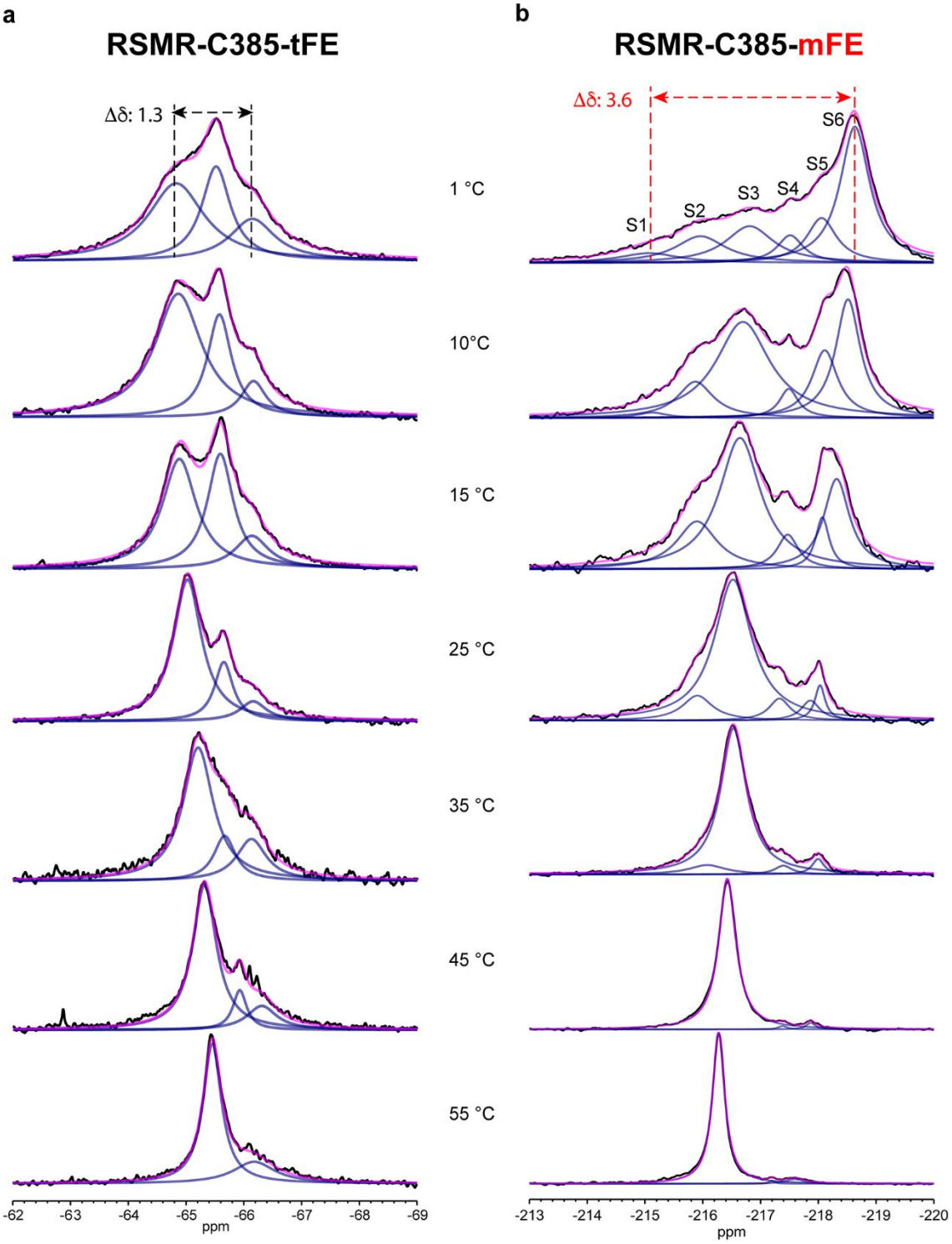
39 Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 486-501 (2010).

40 Chen, V. B. *et al.* MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 12-21 (2010).

Extended Data

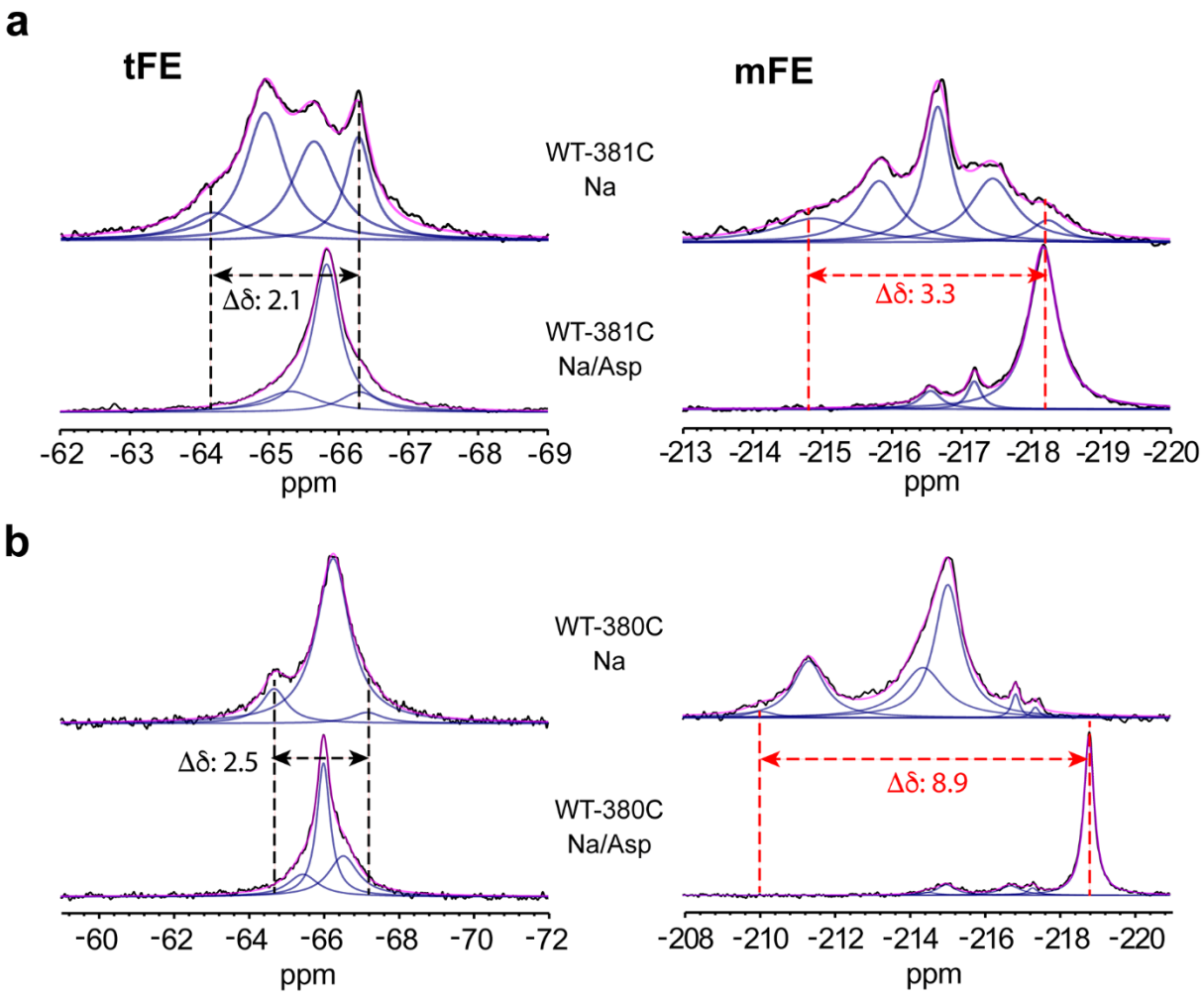


Extended Data Fig. S1. Labeling positions and ^{3}H-aspartate uptake assay. (a) Cysteine mutations introduced into Glt_{Ph}: A380C (magenta), A381C (cyan), and M385C (yellow). A single protomer in the outward-facing state (OFS, left) and inward-facing state (IFS, right) is shown in cartoon representation with scaffold domain colored wheat and transport domains blue and red, respectively. Bound substrate Asp is shown as lime-colored spheres. (b) ^{3}H-Asp uptake assay of WT Glt_{Ph} and mFE-labeled single-cysteine mutants (see Methods). Shown are initial ^{3}H-Asp uptake rates (measured at 2 min) of mFE-labeled Glt_{Ph} variants relative to an unlabeled cysteineless WT Glt_{Ph}. Data was normalized to WT uptake in the presence and absence of sodium gradients (closed circles and open circles, respectively). Experiments were repeated in triplicates on independently prepared protein samples.

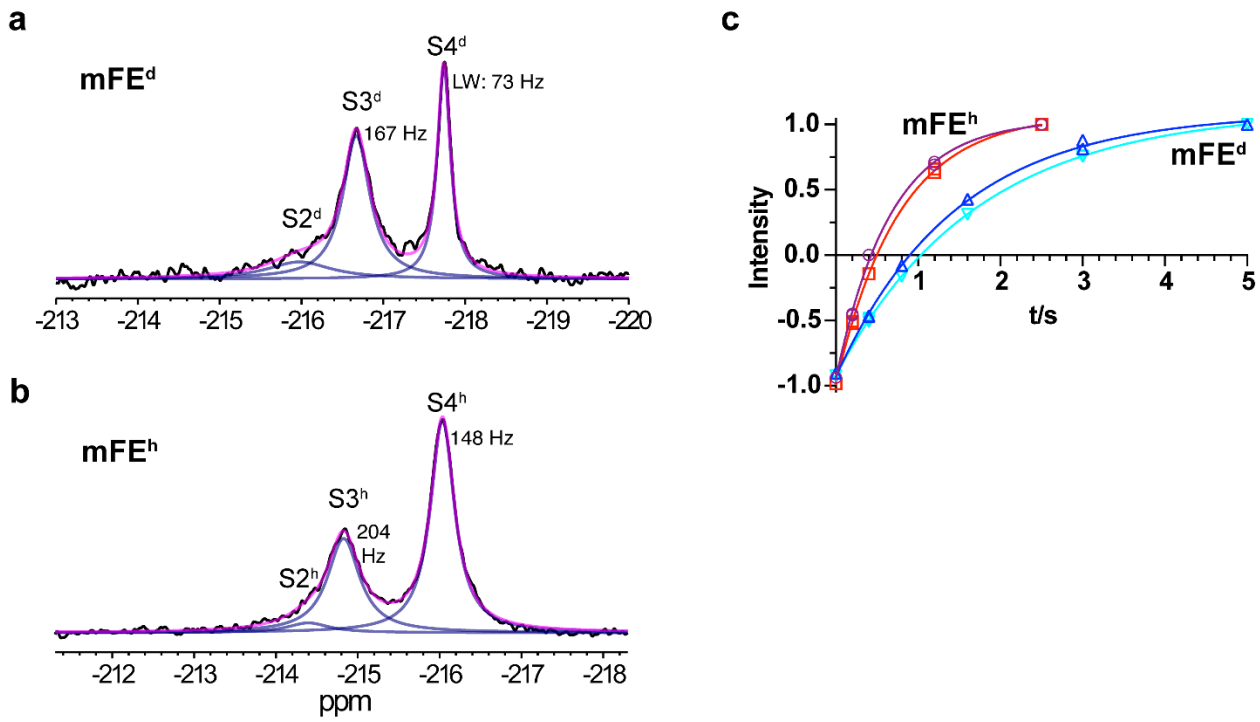


Extended Data Fig. S2. Temperature dependence of ^{19}F NMR spectra of tFE- (a) and mFE-labeled (b) RSMR-385C. Spectra were recorded in the presence of 100 mM NaCl

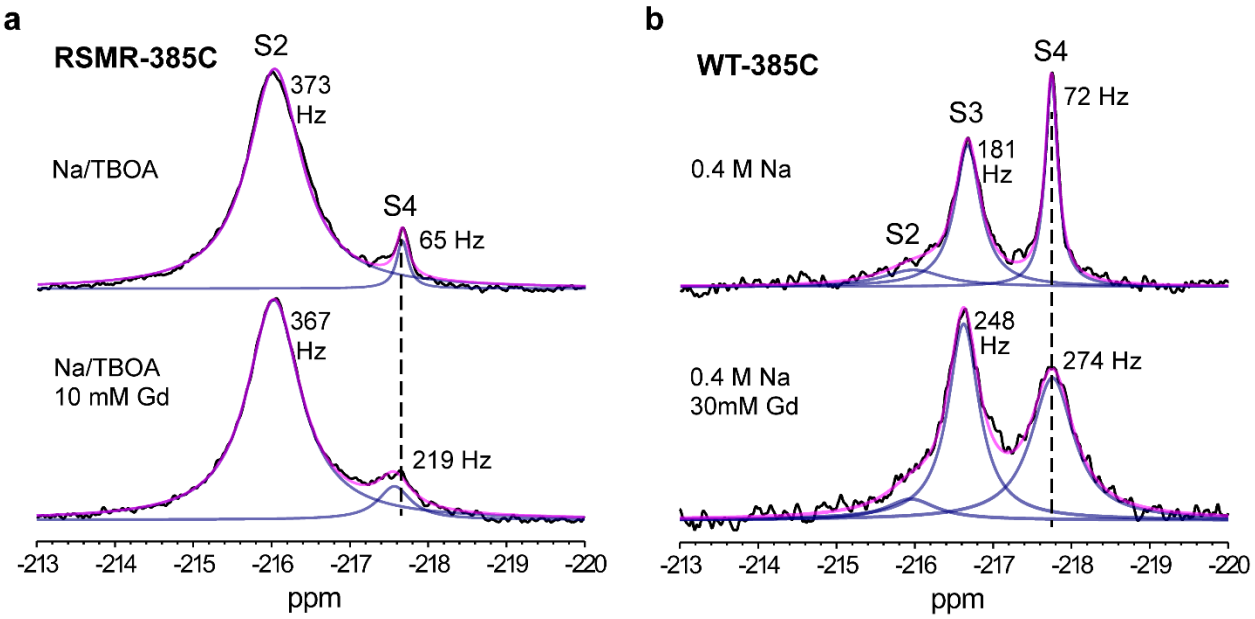
and 2 mM Asp at indicated temperatures. $\Delta\delta$ indicates the largest observed chemical shift difference in ppm.



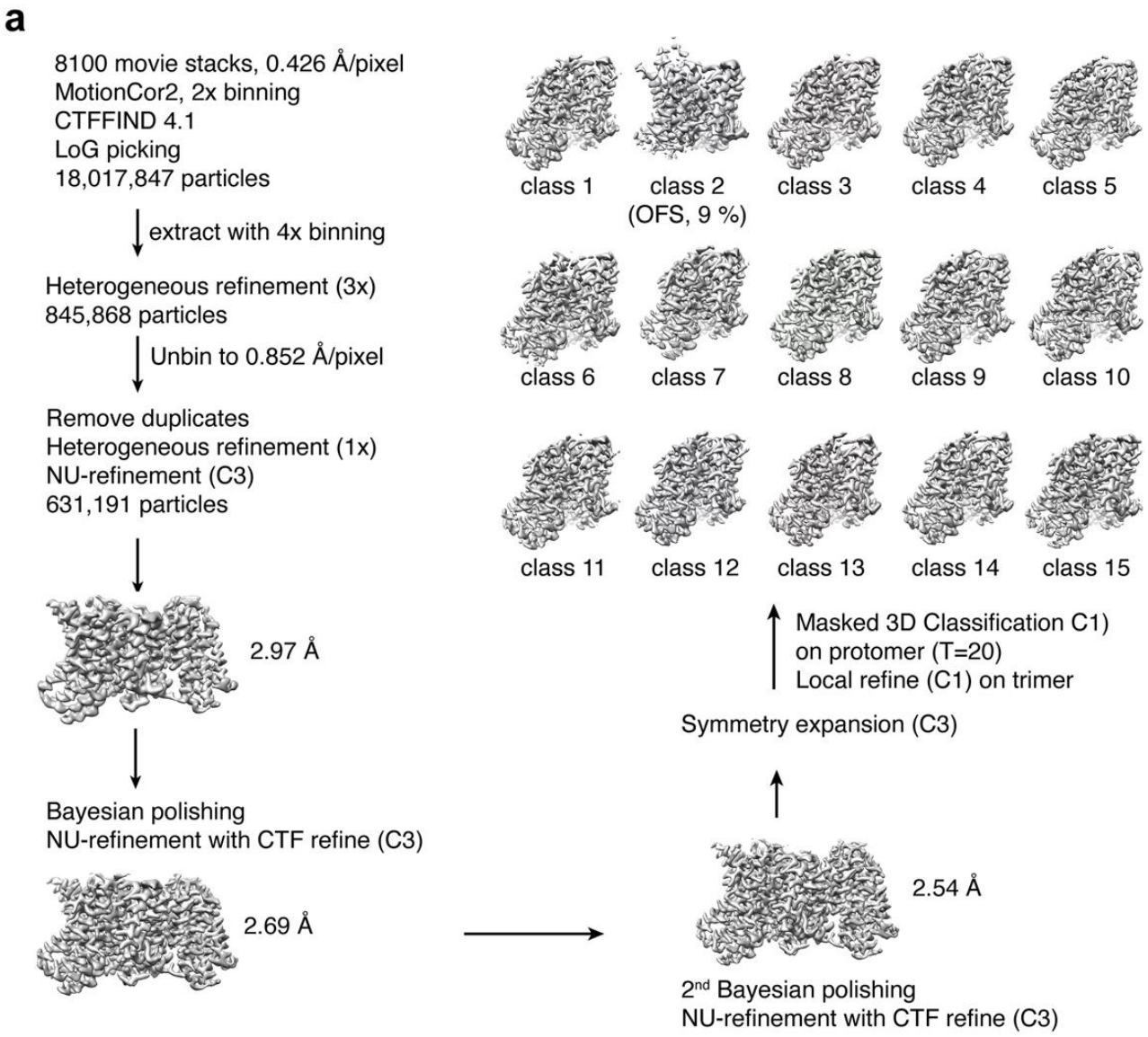
Extended Data Fig. S3. Comparison of chemical shift dispersion of tFE- and mFE-labeled Glt_{Ph} variants. ^{19}F NMR spectra of WT-A381C (a) and WT-A380C (b) labeled with tFE (left) and mFE (right). Shown are spectra in the presence of 400 mM Na^+ (upper panel) and 100 mM Na^+ and 2 mM Asp (lower panel). All spectra were recorded at 25 °C. $\Delta\delta$ indicates the largest observed chemical shift difference in ppm.



Extended Data Fig. S4. **Relaxation properties of deuterated (mFE^d) and protonated (mFE^h) mFE probes.** ^{19}F NMR spectra of mFE^d- (**a**) and mFE^h- (**b**) labeled WT-385C in the presence of 400 mM NaCl. (**c**) T_1 relaxation plots of the S3^d (purple circles) and S4^d (red squares) peaks of mFE^d-labeled protein and the corresponding S3^h (blue triangles) and S4^h (cyan invert triangles) peaks of mFE^h-labeled protein. Solid lines are least-square fits to mono-exponential equations, with time constants of 1.45 ± 0.08 s and 1.71 ± 0.03 s for the S3^d and S4^d peaks of mFE^d label, respectively, and 0.66 ± 0.04 s and 0.77 ± 0.02 s for the S3^h and S4^h peak of mFE^h label, respectively. All measurements were conducted at 25 °C.



Extended Data Fig. S5. ^{19}F peak assignment based on solvent PRE effects. (**a**) ^{19}F NMR spectra of RSMR-385C-mFE in the presence of 0.1 M NaCl and 2 mM TBOA without (top) and with (bottom) 10mM Gd-DPTA-BMA. (**b**) ^{19}F NMR spectra of WT-C385-mFE in the presence of 0.4M NaCl without (top) and with (bottom) 30mM Gd-DPTA-BMA. All spectra were recorded at 25 °C. The numbers next to the peaks correspond to the line widths.

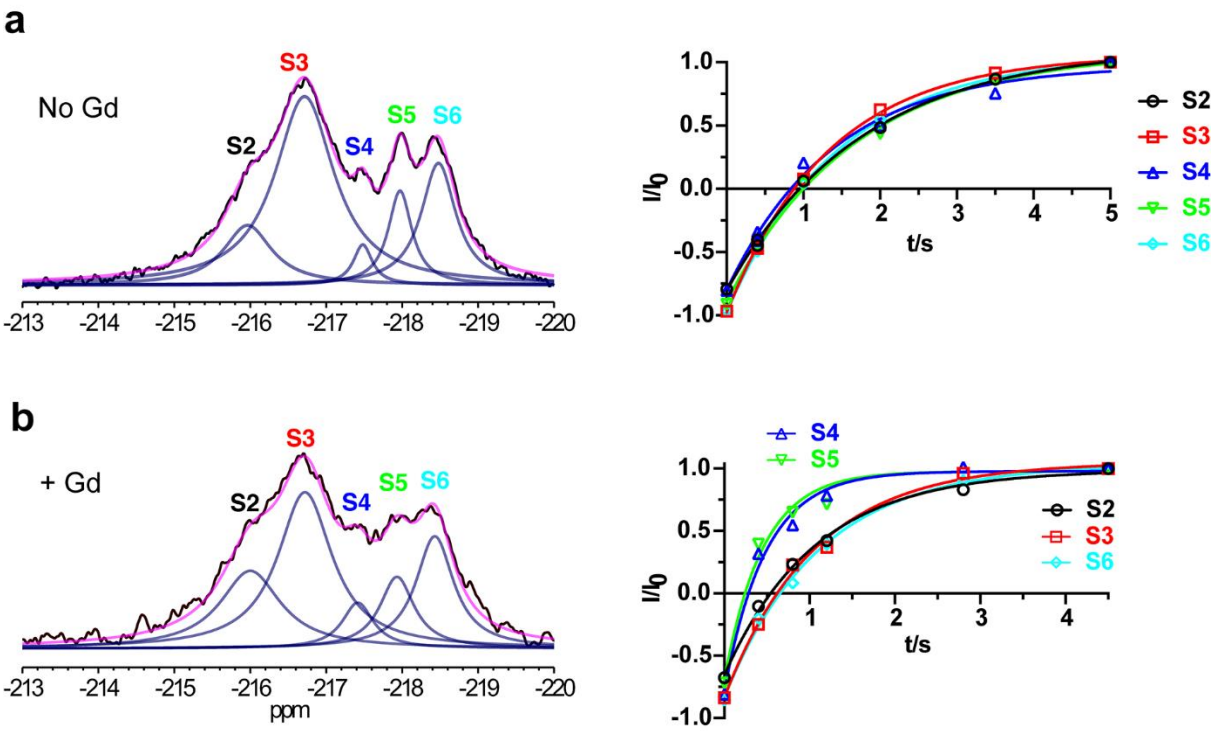


Extended Data Fig. S6. cryo-EM structures of TBOA-bound RSMR mutant. (a) Data processing scheme. Only Class 2 is OFS (9 %), and all other classes are IFSs. (b) Superposition of the TBOA-bound RSMR mutant in OFS (green) and WT Glt_{Ph} in OFS (silver, PDB code 6X17). The scaffold domain is dark gray. (c) Structural comparison of the TBOA-bound RSMR mutant refined from all IFS classes (blue) and WT Glt_{Ph} cross-linked in IFS (silver, PDB code 6X16). Left, superposition of the protomers aligned on scaffold trimerization region residues 140-215. Middle and right, isolated transport domains of the RSMR mutant and WT, respectively. TBOA is shown as sticks. The helical hairpins (HP) 1 and 2 are colored wheat and gold, respectively.

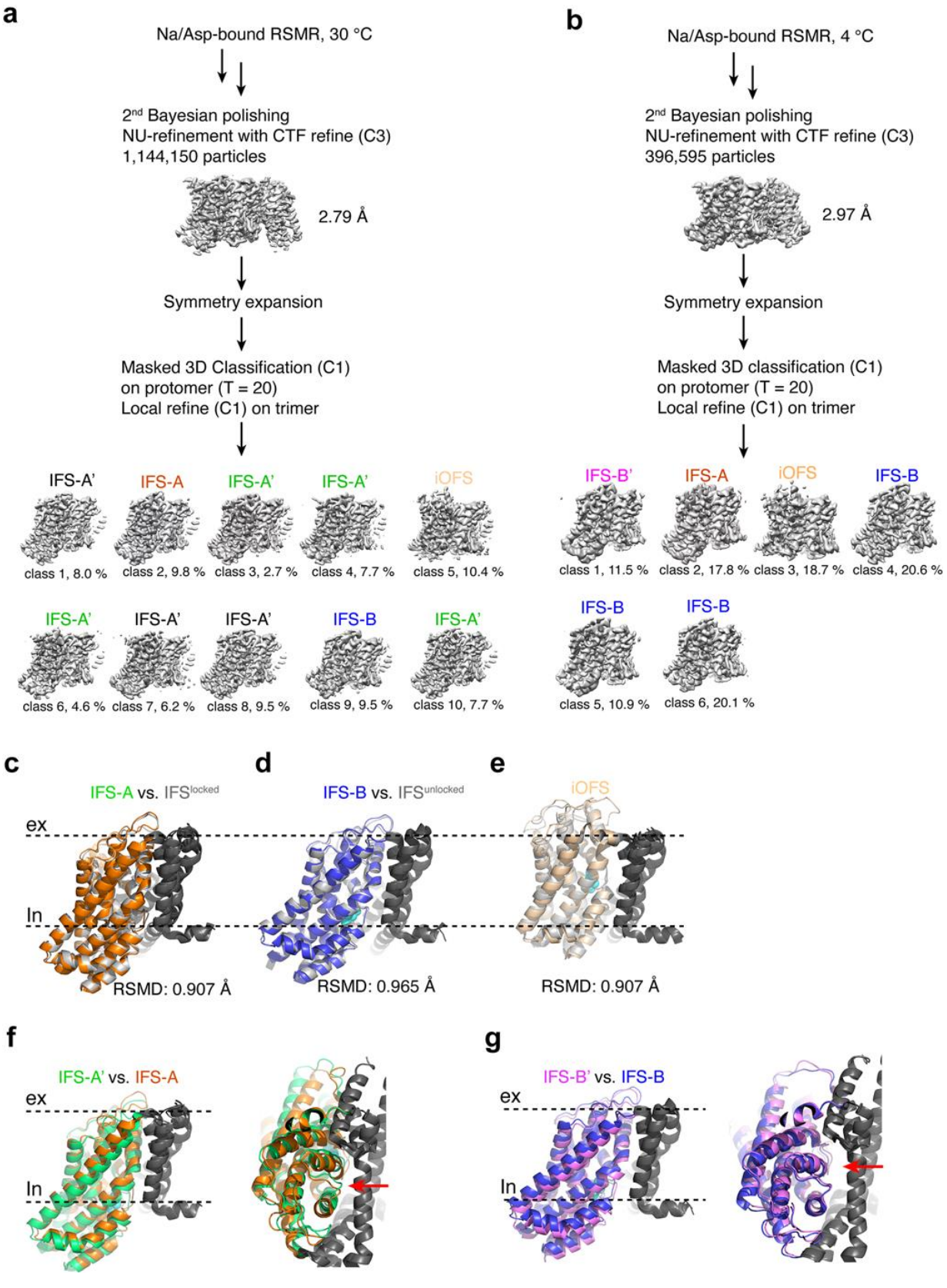
Extended Data Table 1. cryo-EM data collection, reconstruction, and model refinement statistics

	RSMR-TBOA (25 °C) [a]		RSMR-Asp (30 °C) [a]				RSMR-Asp (4 °C) [a]
Data collection and processing							
Magnification	105,000		105,000				105,000
Voltage (kV)	300		300				300
Electron exposure (e–/Å^2)	52.88		50.94				56.04
Defocus range (μm)	1.3-2.0		1.3-1.5				1.3-1.6
Pixel size (Å)	0.4260		0.4260				0.4260
Micrographs (no)	8,100		5267				5823
Initial particles (no.)	18,017,847		12,174,463				20,578,986
Final particles (no.)	631,191		396,595				1,144,150
Reconstruction							
	IFS (EMDB-26482) (PDB 7UG0)	OFS (EMDB-26489) (PDB 7UGJ)	IFS-A (EMDB-26497) (PDB 7UGV)	IFS-A' (EMDB-26487) (PDB 7UGD)	IFS-B (EMDB-26498) (PDB 7UGX)	iOFS (EMDB-26504) (PDB 7 UH3)	IFS-B' (EMDB-26505) (PDB 7UH6)
Particle no.	1,731,338	162,235	335,559	264,559	326,008	356,418	171,456
Symmetry imposed	C1	C1	C1	C1	C1	C1	C1
Map resolution (Å)	2.55	2.81	2.94	3.27	2.96	2.99	3.38
FSC threshold	0.143	0.143	0.143	0.143	0.143	0.143	0.143
Map resolution range (Å)	1.859 – 31.177	1.842 – 36.514	1.829 – 40.061	1.908 – 39.980	1.858 – 40.105	1.969 – 39.643	2.091 – 43.373
Refinement							
Initial model used (PDB code)	6X16	6X17	3KBC	3KBC	4X2S, chain B	6UWL	4X2S, chain B
Model resolution (Å)	2.8	3.1	3.3	3.4	3.4	3.4	3.7
FSC threshold	0.5	0.5	0.5	0.5	0.5	0.5	0.5
Map sharpening *B* Factor (Å^2)	59.1	68.5	94.8	91.9	92.7	98.3	88.8
Model composition							
Non-hydrogen atoms	3120	3087	3092	3087	3104	3092	3063
Protein residues	418	416	418	417	416	416	413
Ligands	3	3	3	3	3	3	3
B factors (Å^2)							
Protein	95.73	104.84	107.56	90.61	109.52	111.97	108.96
Ligand	108.34	112.85	109.15	90.31	125.77	127.68	118.51
R.m.s. deviations							
Bond lengths (Å)	0.002	0.002	0.002	0.003	0.002	0.003	0.002
Bond angles (°)	0.514	0.445	0.436	0.568	0.464	0.589	0.508
Validation							
MolProbity score	1.18	1.11	1.04	1.24	1.24	1.35	1.32
Clashscore	3.91	3.17	2.52	4.73	4.71	6.30	5.88
Poor rotamers (%)	0	0	0	0	0	0	0
Ramachandran plot							
Favored (%)	99.04	99.52	100	98.79	98.79	99.03	98.53
Allowed (%)	0.96	0.48	0	1.21	1.21	0.97	1.47
Disallowed (%)	0	0	0	0	0	0	0

[a] The chamber temperature for making grids.



Extended Data Fig. S7. Peak assignment of RSMR-385C-mFE by solvent PRE. ^{19}F NMR spectra in the presence of 300 mM NaCl and 2 mM Asp (left), and the T_1 measurements (right) in the absence (a) and presence (b) of 20mM Gd-DPTA-BMA. Solid lines are least-square fits using mono-exponential equation, giving T_1 values of 1.25 ± 0.12 s, 1.38 ± 0.01 s, 1.53 ± 0.18 s, 1.57 ± 0.28 s, and 1.44 ± 0.08 s for S2, S3, S4, S5 and S6, respectively, in (a), and 1.07 ± 0.09 s, 1.06 ± 0.11 s, 0.48 ± 0.07 s, 0.45 ± 0.09 s, and 1.11 ± 0.09 s for S1, S2, S3, S4, and S5, respectively, in (b). All measurements were conducted at 15 °C.



Extended Data Fig. S8. Temperature-dependent Na/Asp-bound RSMR structural ensemble. Cryo-EM data processing scheme for grids prepared at 30 °C (a) and 4 °C (b). (c-e) Superposition of the Asp-bound IFS-A (orange), IFS-B (blue), and intermediate OFS (iOFS, wheat) with the previously reported structures (silver): “locked” IFS (3KBC), “unlocked” IFS (PDB code 4X2S), and iOFS (6UWL), respectively. The scaffold domain is colored dark gray. Corresponding RMSDs are shown below the structure. (f) Superposition of IFS-A’ (green) with IFS-A (orange) from a side view (left) and bottom view (right), respectively. (g) Superposition of IFS-B’ (magenta) with IFS-B (blue) from a side view (left) and bottom view (right), respectively. Red arrows indicate the space between transport domain and scaffold domain. Structures were aligned on the scaffold domain.