nature portfolio

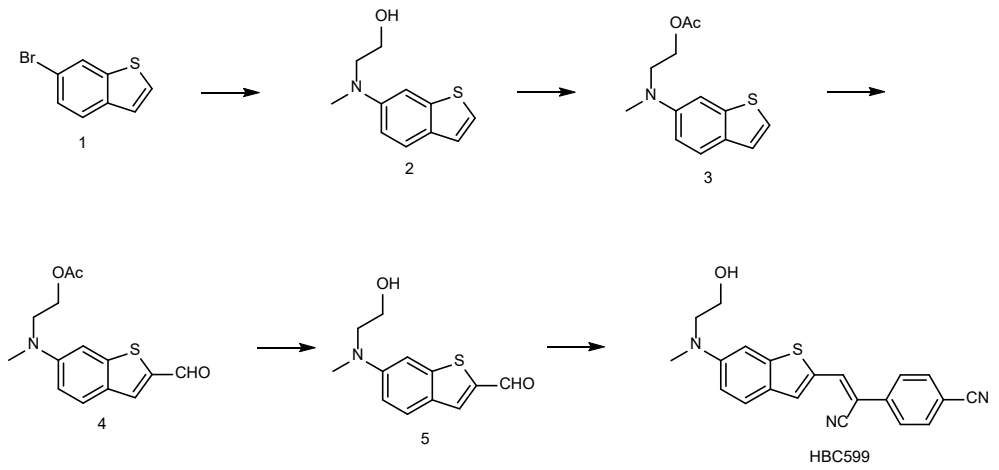
https://doi.org/10.1038/s41586-025-08598-8

Supplementary information

De novo design of transmembrane fluorescence-activating proteins

In the format provided by the authors and unedited

Supplementary Methods: Synthetic route of HBC599.



Compound 2: Compound 1(1.02 g, 4.8 mmol), CuI (183 mg, 0.96 mmol), K_3PO_4 (3.05 g, 14.4 mmol), (L)-proline (221 mg, 1.92 mmol), and 20 mL 2-methylaminoethanol were stirred at 100 °C overnight under Ar. After cooling to room temperature, 100 mL water was added to the mixture. The organic compounds were extracted with CH_2Cl_2, washed with saturated aqueous NaCl, and then dried over Na_2SO_4. The solution was filtered, and the solvent was removed under reduced pressure to obtain a yellow oil. It was then purified by silica gel column chromatography to yield **compound 2** (0.78 g, 79%). MS (ESI): m/z calcd. for $C_{11}H_{14}NOS$ $[M+H]^+$: 207.1; found 207.1.

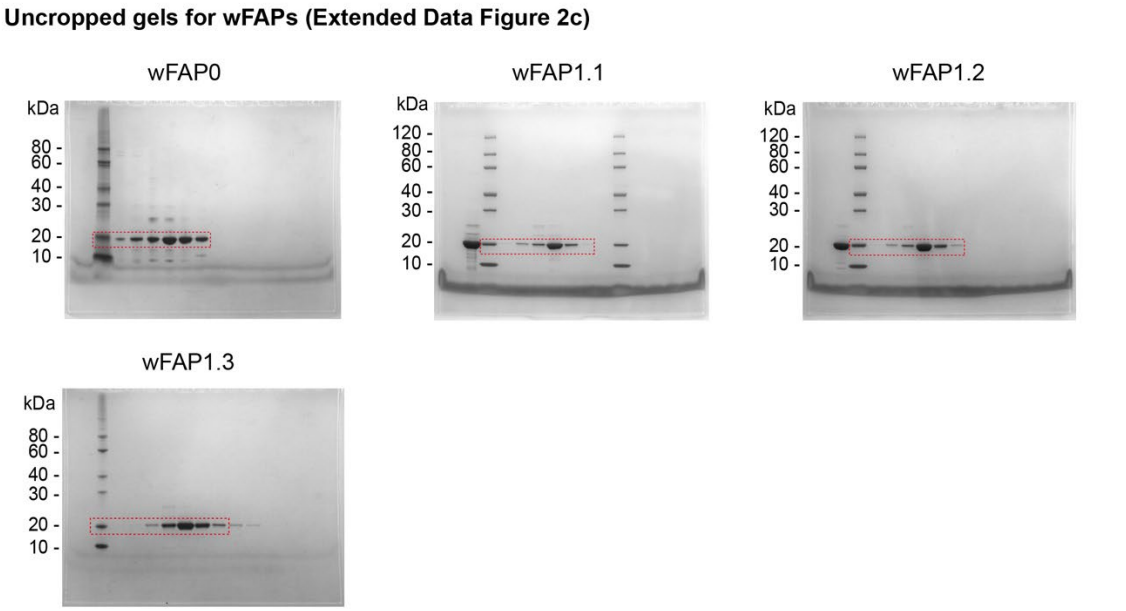
Compound 3: Acetic anhydride (203 μL, 2 mmol) was added to a solution of compound 2 (0.42 g, 1.0 mmol) and 4-dimethylaminopyidine (293 mg, 2.4 mmol) in dry CH_2Cl_2 at 0 °C. Then the mixture was warmed to room temperature and stirred for another 1 h. The reaction mixture was quenched with 2 mL water, extracted with CH_2Cl_2, washed by saturated aqueous NaCl, and then dried with Na_2SO_4. The solvent was removed under reduced pressure to give the crude product, then it was purified by silica gel column chromatography to yield **compound 3** (0.36 g, 72%). ^1H NMR (500 MHz, Chloroform-*d*) δ 7.66 (d, J = 8.9 Hz, 1H), 7.18 (d, J = 5.5 Hz, 2H), 7.12 (d, J = 5.5 Hz, 1H), 6.92 (dd, J = 8.9, 2.4 Hz, 1H), 4.29 (t, J = 6.1 Hz, 2H), 3.65 (t, J = 6.1 Hz, 2H), 3.03 (s, 3H), 2.01 (s, 3H). MS (ESI): m/z calcd. for $C_{13}H_{16}NO_2S$ $[M+H]^+$: 250.1; found 250.1.

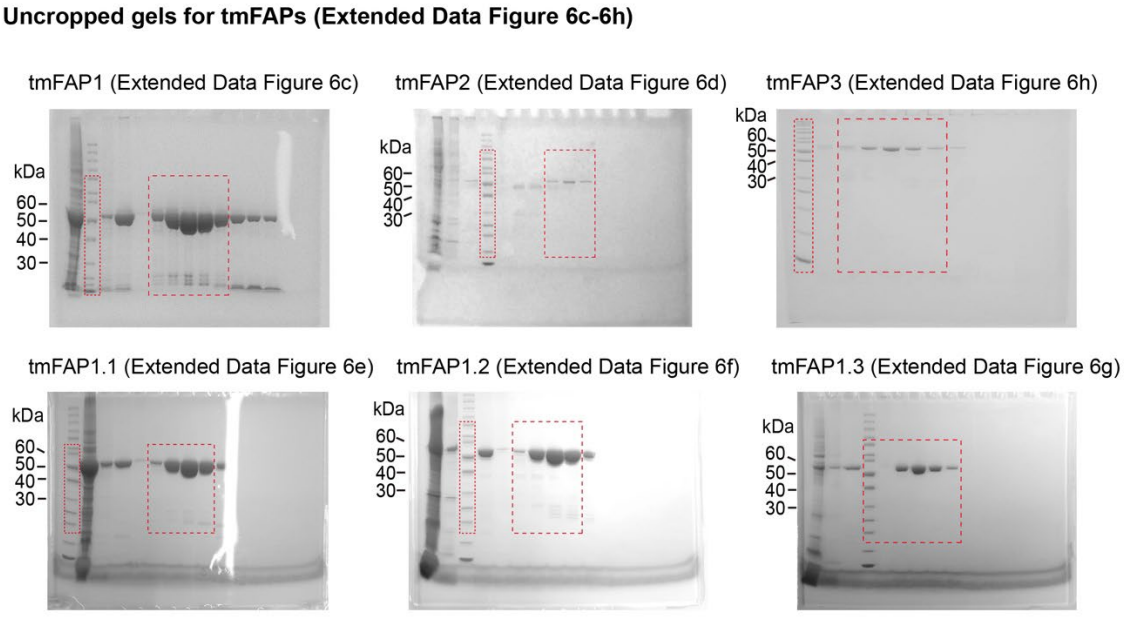
Compound 4: Compound 3 (0.18 g, 0.72 mmol) was dissolved in 10 mL dry CH_2Cl_2 and 2 mL DMF at 0 °C under the protection of Ar. Phosphorous oxychloride (100 μL, 1.08 mmol) was added dropwise with stirring. The resulting mixture was warmed to room temperature and stirred for 5 h. The mixture was quenched with saturated aqueous sodium carbonate and extracted with CH_2Cl_2, washed by saturated aqueous NaCl, and dried over Na_2SO_4. Then the solvent was removed under reduced pressure to give the crude product which was purified by silica gel column chromatography to yield compound 4 (10 mg, 5%). ^{1}H NMR (500 MHz, Chloroform-*d*) δ 9.95 (s, 1H), 7.84 (s, 1H), 7.73 (d, *J* = 9.0 Hz, 1H), 7.05 (d, *J* = 2.4 Hz, 1H), 6.93 (dd, *J* = 9.0, 2.4 Hz, 1H), 4.30 (d, *J* = 5.9 Hz, 2H), 3.73 – 3.69 (m, 2H), 3.09 (s, 3H), 2.00 (s, 3H). MS (ESI): m/z calcd. for $C_{14}H_{16}NO_3S$ $[M+H]^+$: 278.1; found 278.1.

Compound 5: Two mL saturated solution of sodium carbonate was added to a stirred solution of compound 4 (10 mg, 0.036 mmol) in 1 mL methanol. The resulting mixture was stirred and kept at room temperature for 2 h under Ar. The solvent was removed under reduced pressure, then the residue was re-dissolved in 50 mL CH_2Cl_2 and washed by saturated aqueous NaCl. The organic layer was dried over Na_2SO_4 and filtered. The solvent was removed under reduced pressure to give the crude product (10 mg, 99%) which was used immediately for next step. MS (ESI): m/z calcd. for $C_{12}H_{14}NO_2S$, $[M+H]^+$: 235.1; found 236.1.

HBC599: To a stirring solution of compound 5 (10 mg, 0.044 mmol) and 4-cyano-benzeneacetonitrile (6.3 mg, 0.044 mmol) in 3 mL dry methanol, 1 drop of piperidine was added. After stirring at ambient temperature for 1 h, the mixture was cooled to room temperature, filtrated, and washed by cold MeOH to yield compound HBC599 as a red solid (10 mg, 66%). ^{1}H NMR (500 MHz, DMSO-d_6): δ 8.52 – 8.40 (m, 1H), 7.94 (dt, *J* = 8.9, 3.0 Hz, 2H), 7.88 (td, *J* = 4.8, 2.7 Hz, 3H), 7.76 (dt, *J* = 9.0, 2.7 Hz, 1H), 7.23 (q, *J* = 2.7 Hz, 1H), 6.98 (dq, *J* = 8.9, 3.6, 3.0 Hz, 1H), 3.59 (t, *J* = 6.3 Hz, 2H), 3.53 (d, *J* = 6.1 Hz, 2H), 3.06 (d, *J* = 1.7 Hz, 3H).

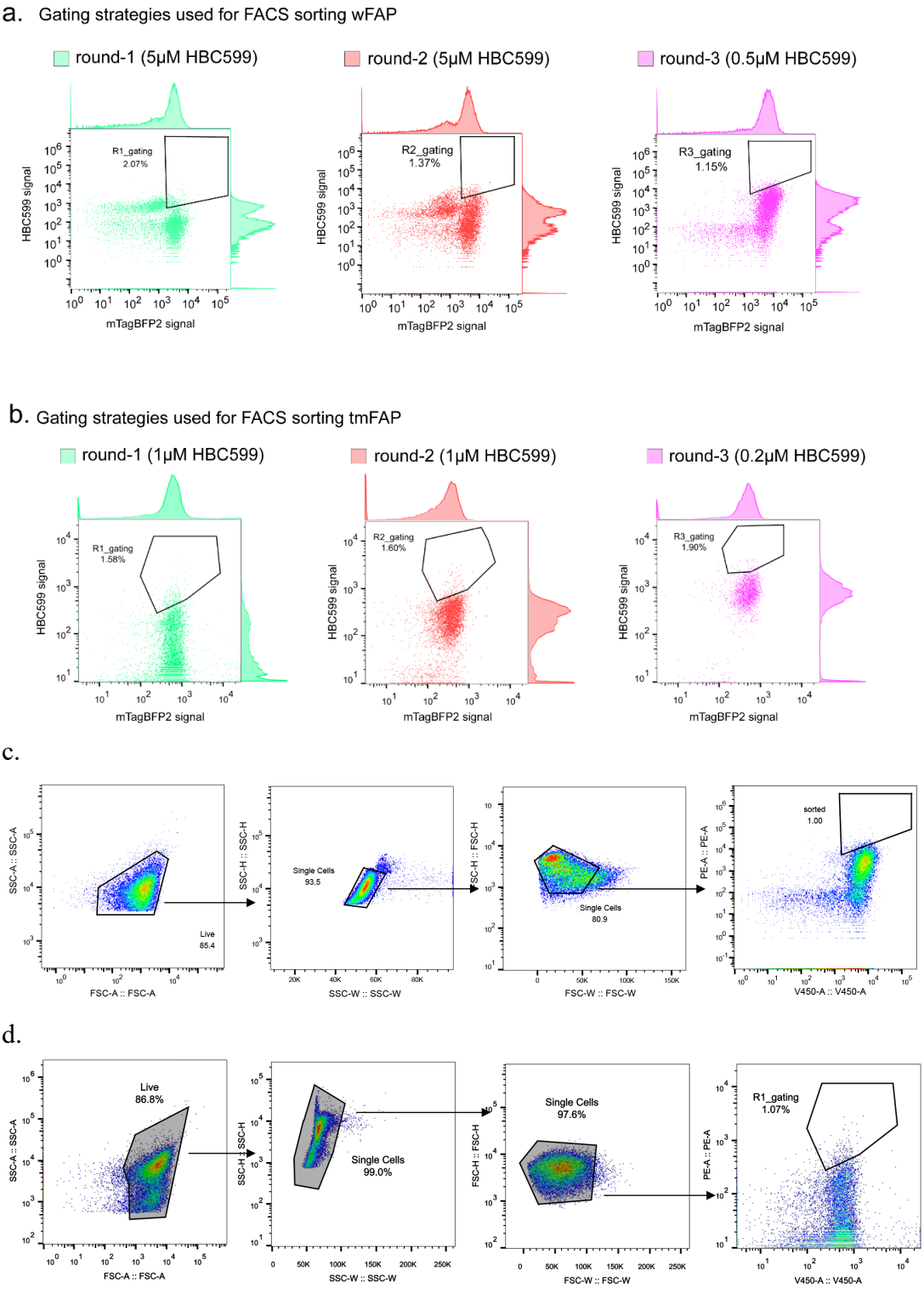
Supplementary Figures





Supplementary Figure 1 | Uncropped gel images.

Full scan of uncropped gels used in Extended Data Figures 2 and Extended Data Figures 6. Red boxes indicate the area shown in the final figures.



Supplementary Figure 2 | Flow cytometry gating strategy

Gating strategy for the flow cytometry experiments. **a-b**, Gating strategies for selecting the most active mutants in wFAP (**a**) and tmFAP (**b**) directed evolution. **c-d**, Gating strategies for selecting *E. coli* single cells in wFAP (**c**) and tmFAP (**d**) flow cytometry experiments.

Supplementary Table 1 | Fluorescence characteristics.

Fluorophore	Protein/ RNA	Ex max* (nm)	Em max* (nm)	EC† (M^{-1} cm^{-1})	QY†	Brightness†	Fold-activation TBS‡	Fold-activation TBS+1.1% OG‡	Kd (μM)
HBC599	wFAP0	495	570	34350	0.74	25.4	280	-	0.40±0.07
	wFAP1.1	510	575	50033	0.88	44.0	833	-	0.058±0.004
	wFAP1.2	495	555	58450	0.70	40.9	1150	-	0.05±0.016
	wFAP1.3	480	545	31300	0.77	24.1	420	-	0.23±0.08
	tmFAP1	500	565	38300	0.66	25.3	433	27	2.7±0.2
	tmFAP1.1	510	570	42767	0.84	35.9	957	26	0.23±0.05
	tmFAP1.2	490	555	59600	0.84	50.1	1675	44	0.018±0.004
	tmFAP1.3	495	565	55467	0.87	48.3	1111	41	0.066±0.006
HBC620	wFAP1.1	540	590	49300	0.71	35.0	700	-	0.09±0.02
	tmFAP1	540	590	-	-	-	723	27	0.52±0.05
	tmFAP1.2	515	575	47400	0.42	19.9	1041	27	0.11±0.007
	tmFAP1.3	525	585	50533	0.45	22.7	871	38	0.19±0.02
HBC599	Pepper§	515	599	54400	0.43	23.4	708	-	0.018
HBC620	Pepper	577	620	100000	0.58	58.0	12600	-	0.0061
-	EGFP‖	488	509	55000	0.6	33.0	-	-	-

*Ex max is Excitation maximum. Em max is Emission maximum. †EC is extinction coefficient measured at the maximum absorbance. QY is quantum yield. Brightness is calculated brightness, product of extinction coefficient and quantum yield divided by 1000. ‡Fold-activation TBS is fluorescence intensity relative to free fluorophore in TRIS-buffered saline (TBS). Fold-activation TBS+1.1% OG is fluorescence intensity relative to free fluorophore in TBS buffer with 1.1% (w/v) n-Octyl-β-D-Glucopyranoside (OG). §data were from Chen, et. al. Nature biotech (2019). ‖data were from Sarkisyan, et. al. Biophs J (2015).

Supplementary Table 2 | List of designer transmembrane protein sequences

name	amino acid sequence
tm-C2_1	TRTEIIRELERSLRLQFFLALFLLVLLLFLLVLIQQLKELLRELERLQREGSS DEDVRELLREIKELVERIVMLIVMIIILVLFIILLALAQKYLVEELKRQD
tm-C2_2	TRTEIIRELERSLRVQLVLALLLFILLLFLLILIQQLKELLRELERLQREGSSD EDVRELLREIKELVEIIVMLIIAILILVLGIILLALAQKYLVEELKRQD
tm-C2_3	TRTEIIRELERSLRMQLFLALFLLALLFALLELVQQLKELLRELERLQREGS SDEDVRELLREIKELVEAILILMFMIIMLVIFIIMLAIGQKYLVEELKRQD
tm-C2_4	TRTEIIRELERSLRIQFILALVLLFLLIILLELIQQLKELLRELERLQREGSSDE DVRELLREIKELVEGILMLLLGIIMLVLLIIMLALFQKYLVEELKRQD
tm-C2_5	TRTEIIRELERSLRLQFFLAIMLLVLLLFLLGLIQQLKELLRELERLQREGSS DEDVRELLREIKELVEEIIMLIMMIILLVLFILLLALFQKYLVEELKRQD
tm-C2_6	TRTEIIRELERSLRFQFALAVMLLFLLLGLLELLQQLKELLRELERLQREGS SDEDVRELLREIKELVEVILMLVLFIIMLVLFIILLAIFQKYLVEELKRQD
tm-C2_7	TRTEIIRELERSLRIQILLALFLLMLLLFLFILLQQLKELLRELERLQREGSSD

	EDVRELLREIKELVEIILDLLVFIILLVMFIIFLAFVQKYLVEELKRQD
tm-C2_8	TRTEIIRELERSLRFQIFLALVLLMLLLILLGLIQQLKELLRELERLQREGSS DEDVRELLREIKELVEMIIALLMAIIFLVIFIILLALGQKYLVEELKRQD
tm-C2_9	TRTEIIRELERSLRLQLLLAIFLLLLLLFLLVLIQQLKELLRELERLQREGSS DEDVRELLREIKELVERILLLLAAIIMLVVFIILLAIAQKYLVEELKRQD
tm-mFAP0_1	GKPIIPQSSVWIVFMFDLEGHFSIGLMVFKRQNPNVLAIIALGFIPNGEFIM GIGFVVVTSPGFLIVLIIYPNLQGIIVFGVIIMASPAIALWISVTSDGHFLFGV LIRQP
tm-mFAP0_2	KVKPPQPISVWAVIMLSSHGQASIGIMIFRPIRPNVLFIIALGMIEGGEAILGI GFVVVRSPRLLFVVIIYNNEPPIMVVGLIFMKSPNVAFWVSFTSNGQVLFG VLAKIK
tm-mFAP0_3	NFKAPIKSSIWIVIMIDLSGQVSIGAMIFKNVKPNILLIIAFGFIPEGEFIMGIG FVIVTSPTFLFVFIFYPDLNNIMVIGFIIMSTPSLAFWFSVTTQGNFLIGVLIK SL
tm-mFAP0_4	KFLQRIGFSVWLVIMINIESGISIGVMVFRPLNPNQLLILAFGLIPSGELIFGI GVVFVRNPNILFVIIIYPFEVPIVVIGVIIMASPNVAVWISVTSTQQILIGVLI KIL
tm-mFAP0_5	SGINHLANQAWLVIMLDDQGDFSIGIMLFRPLPGKKLFIIAIGMIPTGEFIIG FGFVFVASPHFLMVLIIYPQLEHIIVFGVILMASPSVALWASLTSDGHFLIG VLIRLK
tm-CLF4_1	MNEEQRRELLAIAAGLLLAALILLVAAYNTPIEIDLPISGGVKAILYNGKV YLIYENGKVEEIEIPEDDILYPIYNKYIETLVFALFTVAFLAGALFGLLEANE DGELSEEERLEKLAVLAAILFITAIILLGSIEEFSKFLEELKKKLPKNIKLNIN YSSINLAFIAAIMALMAAMLLFMVYES
tm-CLF4_2	MNEEQRTELLIFAAILLAIALILLIMAYNTPIEIDLPISGGVKAILYNGKVYLI YENGKVEEIEIPEDDILYPIYNKYIETLIMALGTVGMLGGMLGALLESNED GELSEEERLEKLGVLAIVLGFTAGMLLASIEEFSKFLEELKKKLPKNIKLNI NYSSINLAFFAAIMALFAAILLGIVYES
tm-CLF4_3	MNEENRTELLGIAAILLGIALILLIVAYNTPIEIDLPISGGVKAILYNGKVYLI YENGKVEEIEIPEDDILYPIYNKYIETLMGALVTVGGLAAILGGLLESREDG ELSEEERLEKLGLLAFILFITAMALLLSIEEFSKFLEELKKKLPKNIKLNINY SSINLAMIAALAALGAAILLFMVYES
tm-GLF18_1	SLPEEIFGAALALMGMAGAALKFLKTKKINSTNNIVLGNMLKAIAVAIIMI AGAVALLLIDEDLETVAIALIIIAFAWAFALAGIFNLLFVKKIEMKHGGGA LSPQKKRAKIKKAKMFAGVIFFFFIILMFAVLQDGSSSIQEHVLMLLMFFLI FLVIIMAMAMKLAKHSQTPHLAQNIFMAAVFLFIAFMALIQDLESESSSLL VLAMAVLVAAAVLFFAIVNPALQQLKH
tm-GLF18_2	EVVEAILAGAVGLFIGAFGAIKFLSARRQQGATPSPLGNMLRAIAIAIIMIA VAVALLLIPEPEEQVELALVIIAFAWAVALAAIAQLLAVKQFEIALPAFNQ HPRIKAQAIKQARAVAVGILIFFMIVFMIVMDDPSIDIQDHVLILLGAFLIV LVIIIGFAFSLAQSRQTSPLAKRIFVAAVILFVLFVGLAAELEHTEDSLLVLA VFVLAAAAMALFFIIVKPALNNFSP
tm-GLF18_3	EVIEAIGIFAFILFFMAGGAIKFLKTRRASGQANTPLGRMLKAIAIAIMFIAF AVELLLFPESDEDVDFALIIIAFAWAIALAAILRLLLVKAFEAKNGGPGFPR

	SNRKKAIKNAKGVAFLLFAFFIAIMFMVFNDPSLSVQEHVLILLGAFLIVL LIIIAMAARLASQPSVNKLARQILIAAVLLLFFFGVLINELQDPPQNLLVLA VVVLIAAFVLFGFIVAPALNGFKF
tm-GLF18_4	DISSAIFIGAVVLGLMAAGAILFLRAVKLPSPSARLLGNMLKGIAVAILMIA LAVALLLHPESPEAVEIALFIIAFAWAFALVGIFKLLLVKQFEANHPQAAA NPQNLRKAIRSAKAAAGLIFVAFLMVAIAVFADPSIEVGDHVLILLFAFIIA LIIIVLLAFQLAQQNPNHFLAQSIFIAAVFLLVVFIMLIEDLFADGGNLLVL AGGVLVAAAFLFFFIVAPALANFAQ
tm-GLF18_5	FLNQAIIAAAMILFVFAFFAIQFLNQNPPFVQQNPILGQMLKGIAIAIMMIAI AVQLLLGEPDEQEVDIALIIIAFAWAIALAAIARLLVVKNMEVANAGKGV SPKARKKNIARANMVAAGMMAVFAGVLGAVMNDPSGSQLEHVLVLLM VFAAFLMIIIMMAFRLAQINPANILAKQIFFAAIILFILFMLLIQSLIQTGFNL LVLAFFVLVAAAVLFFFIVAPALAHIHF
tm-GLF18_6	SVEAEIGFFAAFLMVMALFAMRFLNAHQFKNPSQQLLGNMLKGIAIAIILI AIAVELLLFGHDEEEVSLALIIIAFAWAMALAFILRLLLVKSFEAANPGKS VSPSARSQNIKNAQIVAMGIFLAFIAVMLIVMNDPNGSIIEHVLVLLVAFL VVLAVIIVMAFRLAAANPGNTLANRIFVAAVFLLVIFVALAEELATGEDSL LVLAAFVLFAAFILFFFIVAPALAAIIK
tm-GLF32_1	SPGETLAFGAAVLAVAAVAMFGILNAHQPKNAGQIQLYRALNGMLLSIIG LAVALEDLLHAHTPEAHLAALIALLMSMVGILVGMFNIAVYRYMVAAA VKQGIPIALFPNLPAQATFLAALIVVIFFALVVFLMTNPTISLEEGLQILLFF LFMVAILTFFAAANYAAHVPKHPLAKGLVVAAVFIAIALMIALASMADG EDSDLAIAVAAVIIAGAAFAFVIQPVVRNLLK
tm-GLF32_2	SPQEELVAVAGVLAGMAVIMMQILNANRAQNPGRAHLYSALNFMLASII MLALALDDLLHASEEQDHLFALMFLLMSMVGILGAMFRIAFYRYLVMK NKKHHGNLAINPALPQQAQIFAGFFFVAFIVIAIVLFQDPTVDVGQGLQVL LFFLLFLAVVTFFFAAKYAKFAPNHPLARGLVVAAGIIVIALFAAFDAMEL NQEPDLAIAVAAVVIAAAAVFFVIKPVVQHLAK
tm-GLF32_3	FILAALVAIAMGLAILALFMASILNKHQAQNPRHAGLYQALNVMLMSIV MLALALDALLFTGDENTHLGALFALLASMVMILAAMARIAAYRYLVAA AKLTGIGPQAIPQLPAQANAIAGIFIIMFIIFGFVLFSDPSVNPQAGLQALLM MLLVAAVITMLLAMSYAAITPNSHLAKGLLAAAILIVMALLFAMESMAV GVNPDLSIAIAAVGIAFLAMVFVIRPVVQQFAK
tm-GLF32_4	FLAQALVIAAGGLAVMAVAMMKILQTAAAQNPSKAHLYRALNAMLVSI VGLAVALENLLNSHDEENHLFALFGLLISMVGILAAMLQIALYRYFVAAV NSAGKSPQQHPNLPQAAQGIAAGFIFLFVMLIVALINDPSIDPAQGLELLL AFLLIVAVITFLFAVAYANSNAGSSLAKGLAFAALIIVLVLLAALEAMHDP EINDLTIAAVAVIIAAAAMVGVIQPVVQVNFP

Sequences of transmembrane proteins designed in addition to tmFAPs are listed here. tmFAP sequences are provided in the extended data figure.