Unveiling the molecular basis of selective fluorination of SAM-dependent fluorinases

Ravi Kumar Verma,‡[a] Wan Lin Yeo,‡[b] Elaine Tiong,‡[c] Ee Lui Ang,*[be] Yee Hwee Lim,*[de] Fong Tian Wong,*[cd] and Hao Fan*[aef]

a. Bioinformatics Institute, Agency for Science, Technology and Research, 30 Biopolis Street, #07-01 Matrix Building, Singapore 138671, Republic of Singapore.
b. Singapore Institute of Food and Biotechnology Innovation, Agency for Science, Technology and Research, 31 Biopolis Way, #04-01 Nanos, Singapore 138669, Republic of Singapore.
c. Institute of Molecular and Cell Biology, Agency for Science, Technology and Research, 61 Biopolis drive, Proteos, Singapore 138673, Republic of Singapore.
d. Institute of Sustainability for Chemicals, Energy and Environment, Agency for Science, Technology and Research, 8 Biomedical Grove, #07-01 Neuros Building, Singapore 138665, Republic of Singapore.
e. Synthetic Biology Translational Research Program, Yong Loo Lin School of Medicine, National University of Singapore, 10 Medical Drive, Singapore 117597, Republic of Singapore.
f. Department of Biochemistry, Yong Loo Lin School of Medicine, National University of Singapore, 8 Medical Drive, Singapore 117596, Republic of Singapore.

‡Authors contributed equally

*Corresponding authors:
Ee Lui Ang *email:* ang_ee_lui@sifbi.a-star.edu.sg
Yee Hwee Lim *email:* lim_yee_hwee@isce2.a-star.edu.sg
Fong Tian Wong *email:* wongft@imcb.a-star.edu.sg
Hao Fan *email:* hfan2006@gmail.com; fanh@bii.a-star.edu.sg

This PDF file includes:

SI Text

Additional details on molecular dynamics simulations:

In the current study, OPLS4 all-atom force field (Lu *et al*., 2022, *JCTC*) was used to model the entire system, including the protein, sodium, fluoride, chloride ions, and the cofactor S-Adenosylmethionine (SAM). OPLS4 improves accuracy for challenging chemistries, including drug-like chemical space with molecular ions and sulfur-containing moieties. It has been extensively benchmarked against experimental data, such as hydration free energies and ion-water interaction energies, showing strong agreement with observations. The choice of OPLS4 is further supported by its successful application in molecular dynamics (MD) simulations of the CLC family of chloride channels (McKiernan *et al.,* 2020, *PLoS Comput Biol.*) and Claudin-10a/-10b ion channels (Nagarajan *et al.,* 2024, *Int J Mol Sci.*). Consequently, we chose OPLS4 for its broad applicability to proteins and organic cofactors like SAM, and its reliable performance in modeling halide ions, ensuring robust representation of F^- and Cl^- interactions in the ion-binding site.

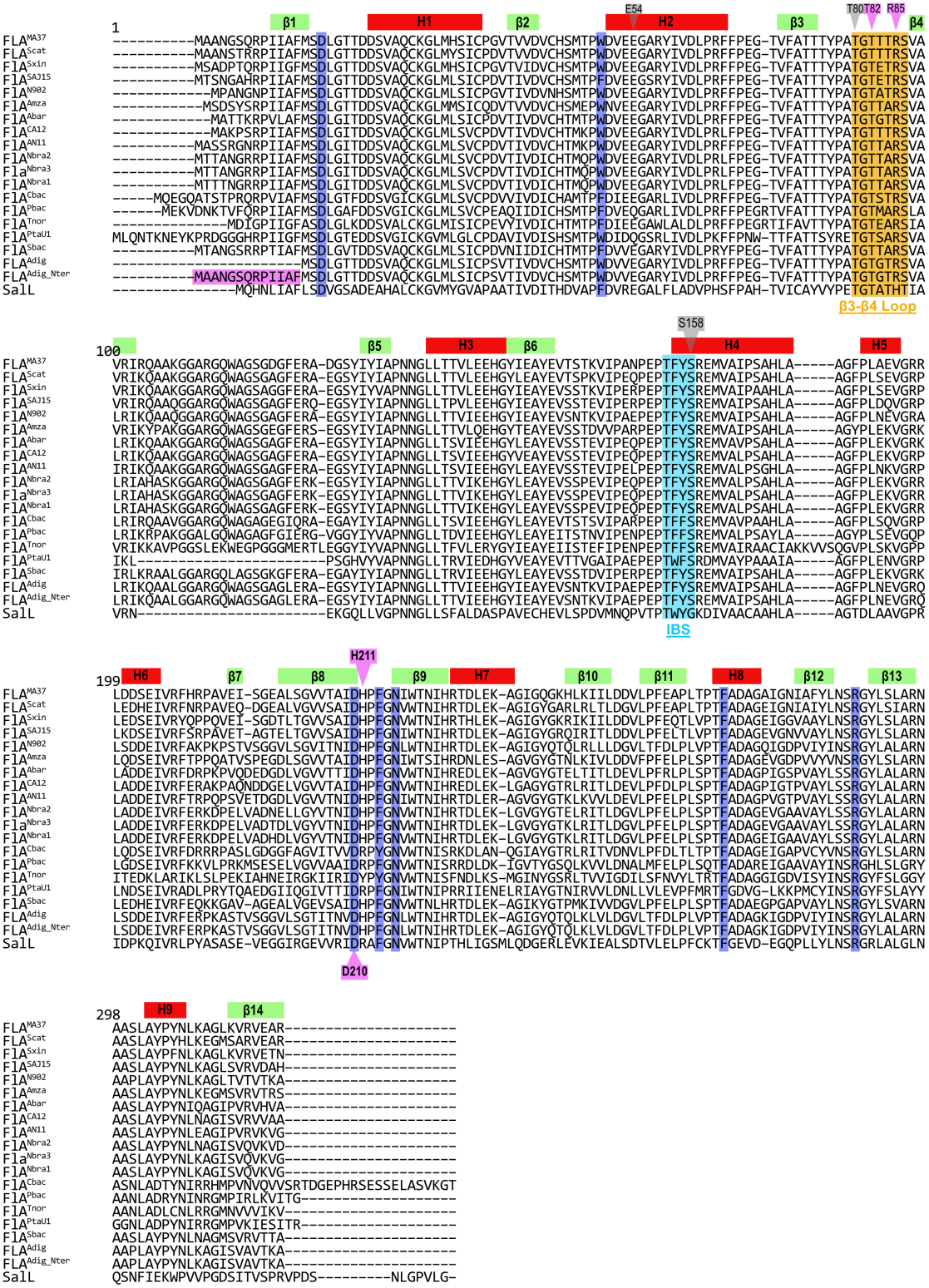

Fig. S1: Comparative sequence alignment of newly discovered fluorinases with previously characterized counterparts, with SalL chlorinase also included for reference. β-strand segments are shown in green, and helical segments in red. Residues constituting the IBS and IES are highlighted in cyan and orange boxes, respectively. Mutation sites are marked in magenta, while other key residues are indicated in grey. The seven SAM-interacting residues are indicated with a blue background.

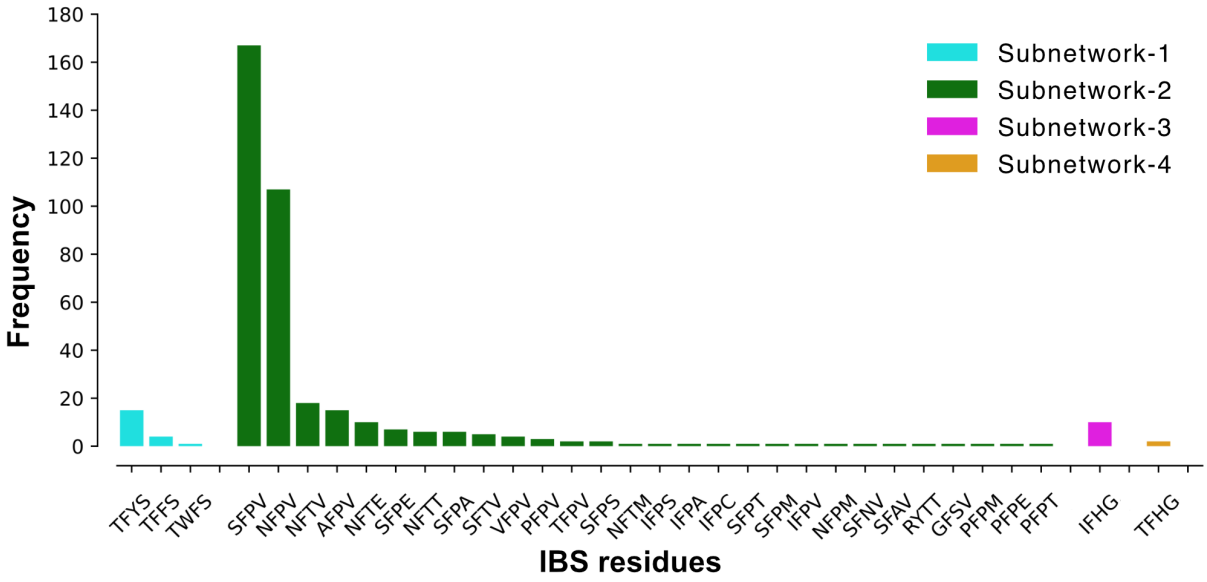

Fig. S2: Diversity and distribution of ion-binding site (IBS) across subnetwork 1-4 sequences.

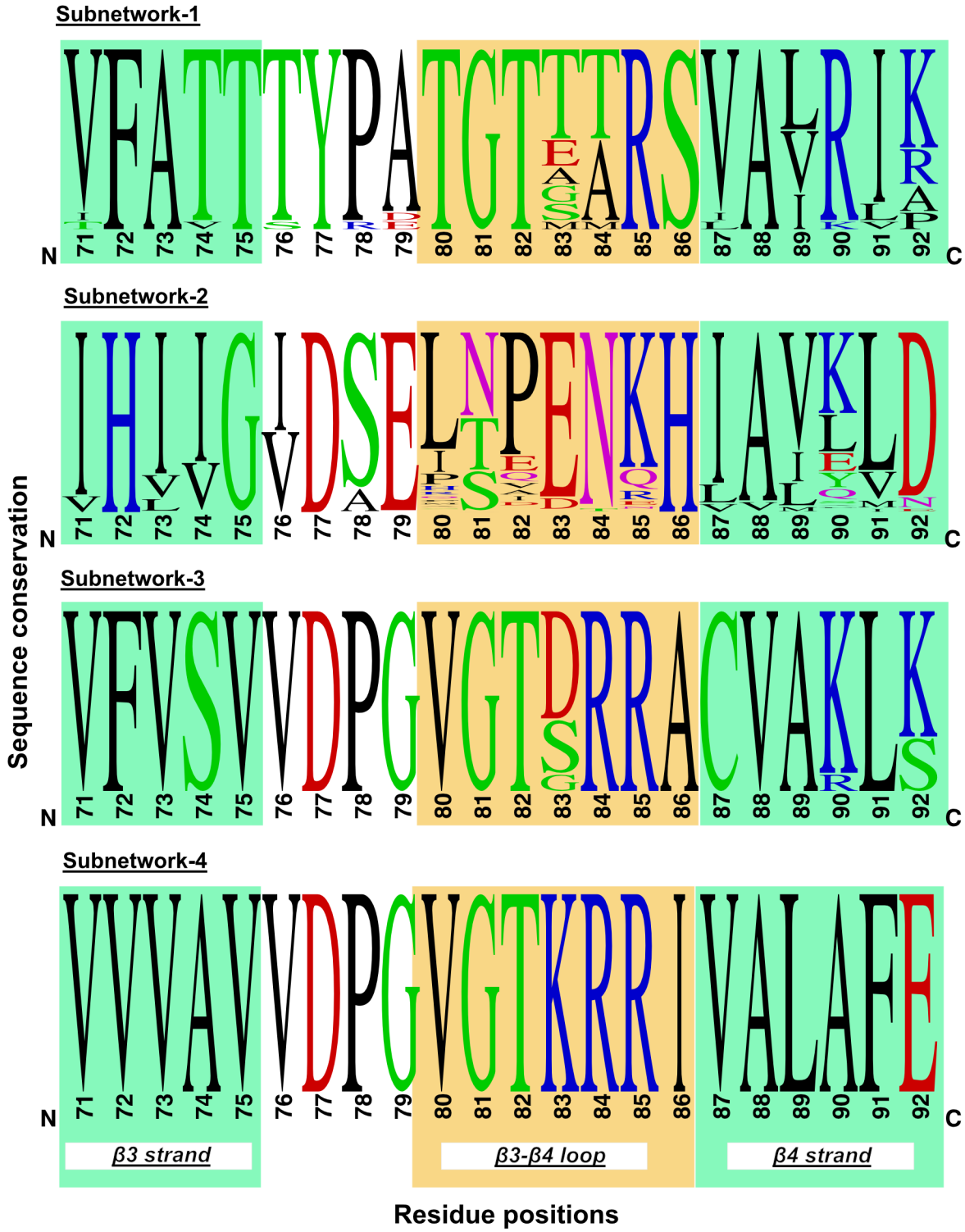

Fig. S3: Sequence conservation of region encompassing β3 and β4 stand in subnetworks 1-4.

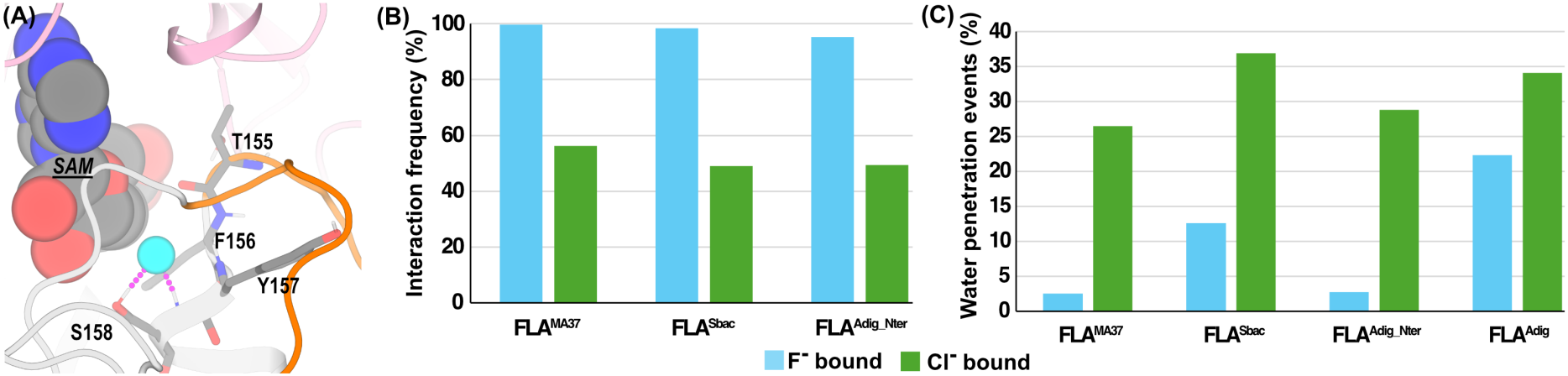

Fig. S4: (A) Interaction between the fluoride ion (cyan sphere) and the sidechain/backbone polar hydrogens of IBS residues in FLA^{Adig_Nter} and (B) Histogram showing interaction frequency between sidechain/backbone polar hydrogens of S158 and halide ion from MD simulation trajectories based on a minimum distance ≤ 2.5 Å criterion. (C) Histograms showing increased water penetration events in chloride bound compared to fluoride bound trajectories. Water penetration into the IBS was counted using on a minimum distance criterion of 3.0 Å between the sidechain/backbone polar hydrogens of S158 in the IBS and water molecules.

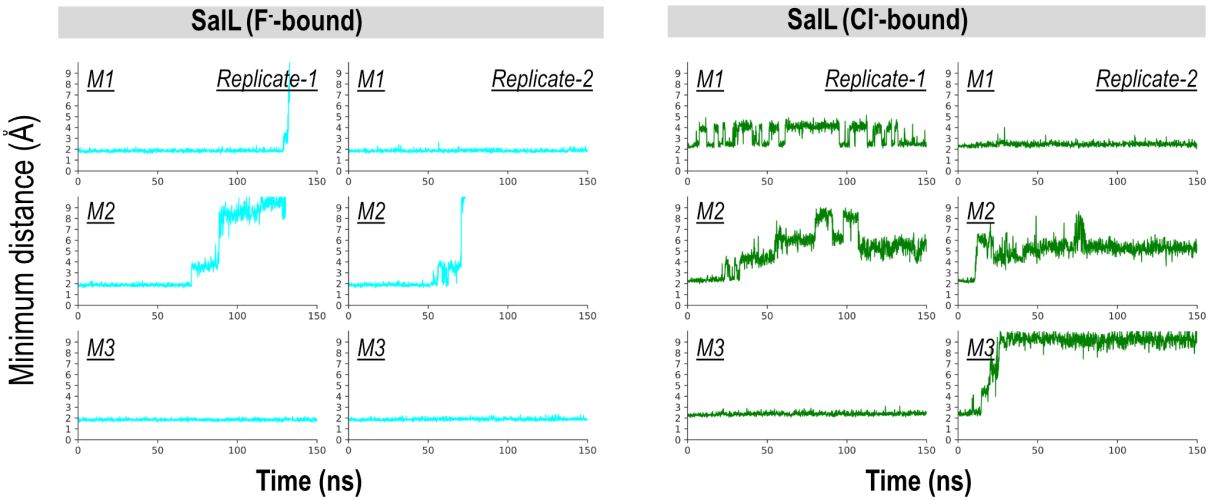

Fig. S5: Plots showing time-evolution of minimum distance between the sidechain and backbone polar hydrogens of S158 and the halide ion (F^- / Cl^-) in the IBS in SalL chlorinase. The distances were plotted separately for each monomer.

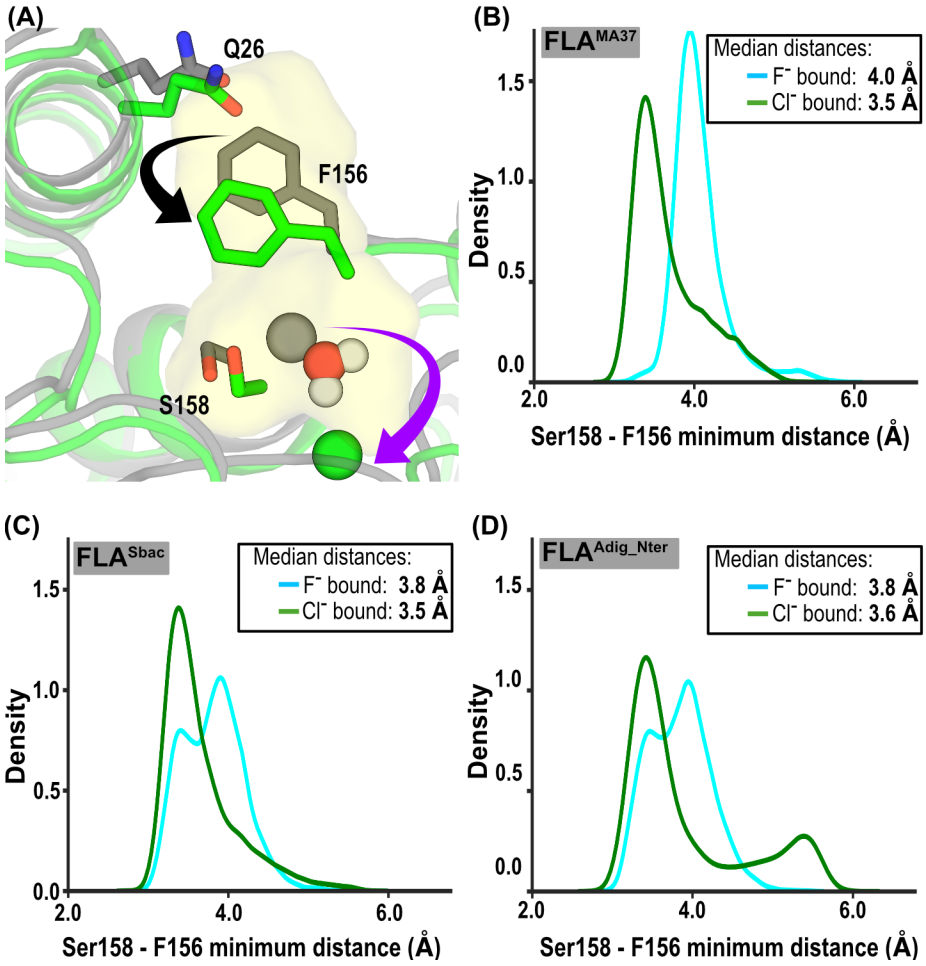

Fig. S6: (A) Representative frame from chloride bound FLA^{Sbac} MD trajectory (green) showing the inward movement of IBS F156, compared to the initial MD-relaxed (grey) structure. The Q26 is shown for reference. The inward movement of F156 coupled with the water penetration in the IBS leads to release of bound chloride, shown as spheres, from the IBS. (B-D) The distributions of minimum distance between F156 and S158 within the ion-binding site in fluoride (cyan) and chloride (green) bound trajectories. The median value for the distributions is given in the box.

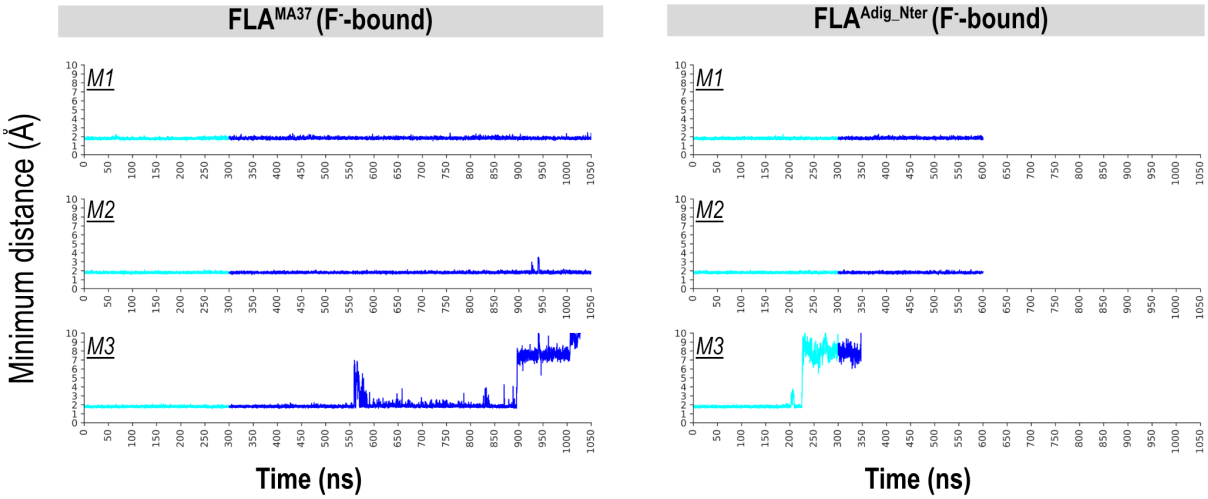

Fig. S7: Plots showing time-evolution of the minimum distance between the sidechain and backbone polar hydrogens of S158 in the IBS and the F^- ion in FLA^{MA37}, and FLA^{Adig_Nter} F^--bound trajectories. Distances were plotted separately for each monomer. The distances from the initial 300 ns of MD simulation trajectories are represented in cyan, with the extended simulation period shown in blue. Notably, the F^- ion completely dissociates from monomer M3 in both cases.

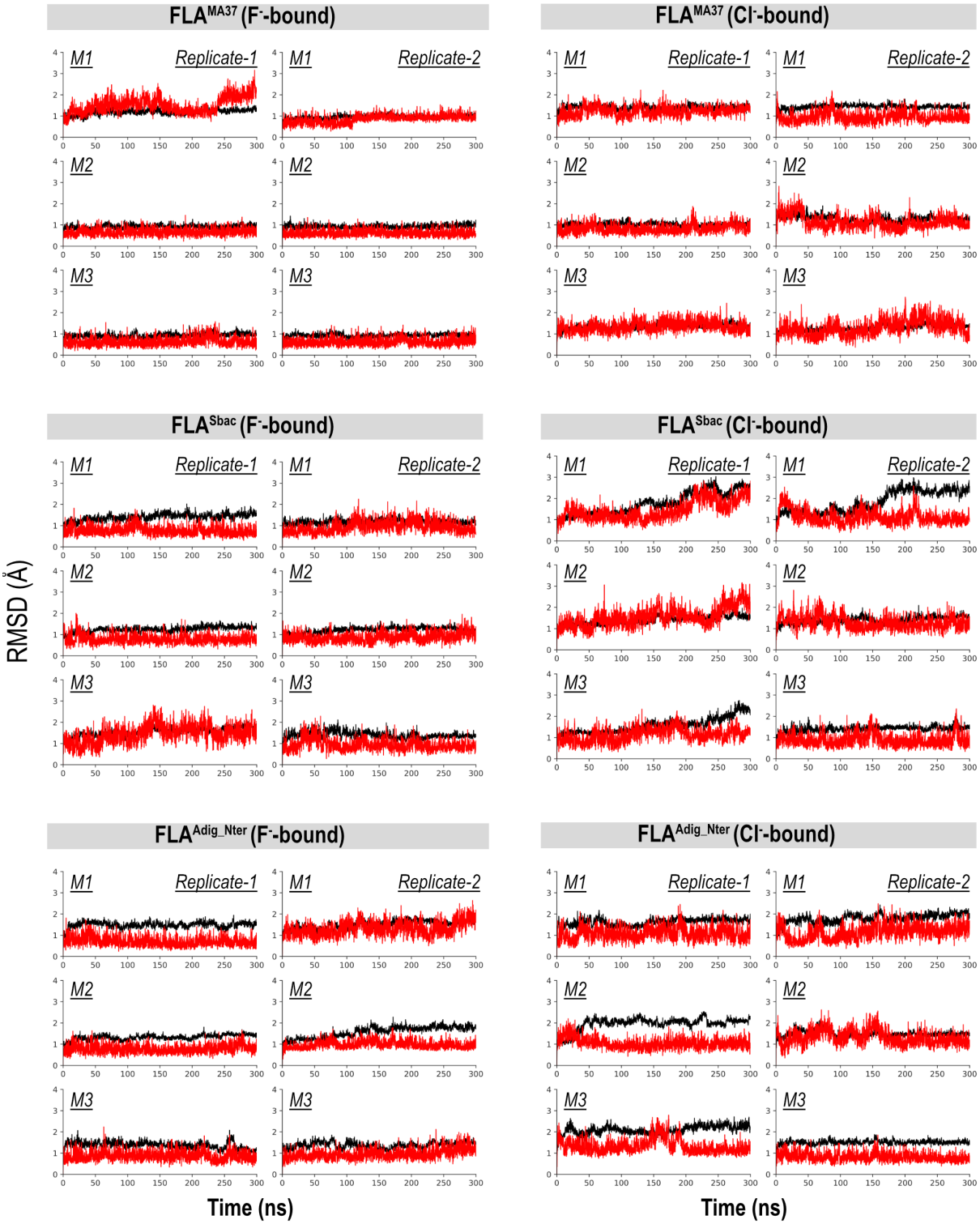

Fig. S8: RMSD plots for F^--bound and Cl^--bound trajectories of FLA^{MA37}, FLA^{Sbac}, and FLA^{Adig_Nter}. Two replicate MD simulations were conducted for each condition. The RMSDs of the protein Cα atom (black) and SAM heavy atoms (red) were plotted separately.

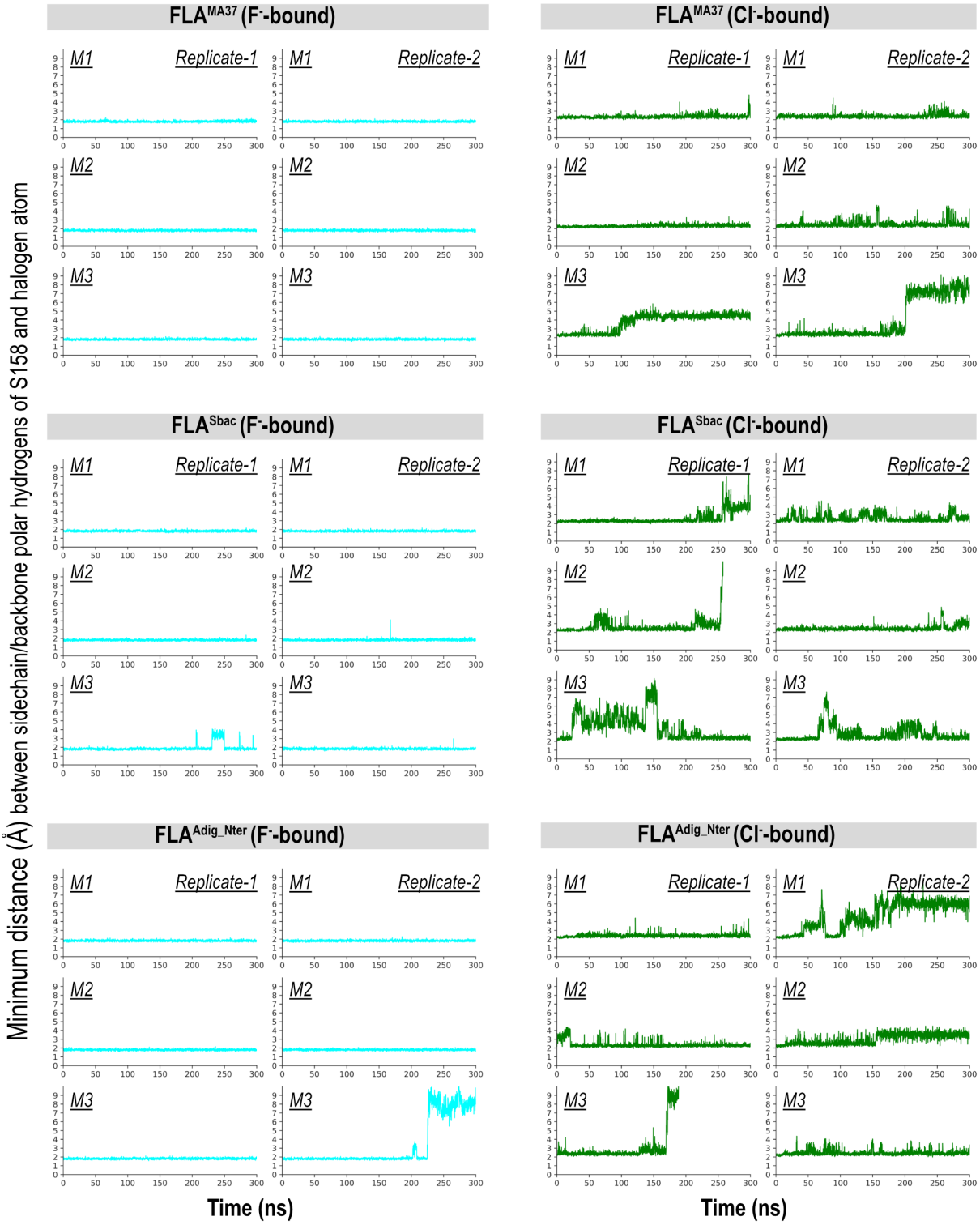

Fig. S9: Plots showing time-evolution of the minimum distance between the sidechain and backbone polar hydrogens of S158 in the IBS and the halide ion (F^- / Cl^-) in FLA^{MA37}, FLA^{Sbac}, and FLA^{Adig_Nter}. Two replicate MD simulations were conducted for each condition. The distances were plotted separately for each monomer.

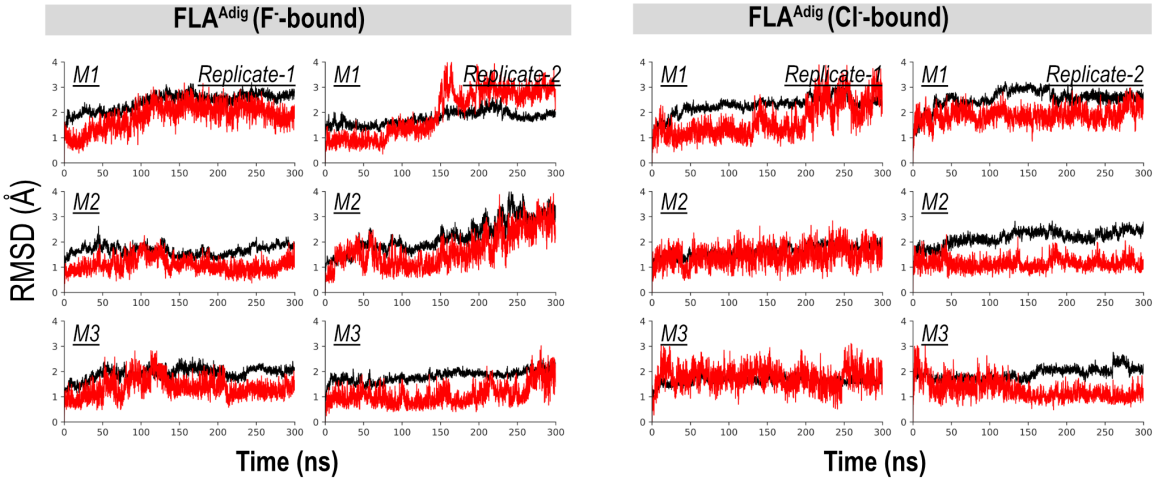

Fig. S10: RMSD plots for FLA^{Adig}. Two replicate MD simulations were conducted for each condition. The RMSDs of the protein Cα atom (black) and SAM heavy atoms (red) were plotted separately.

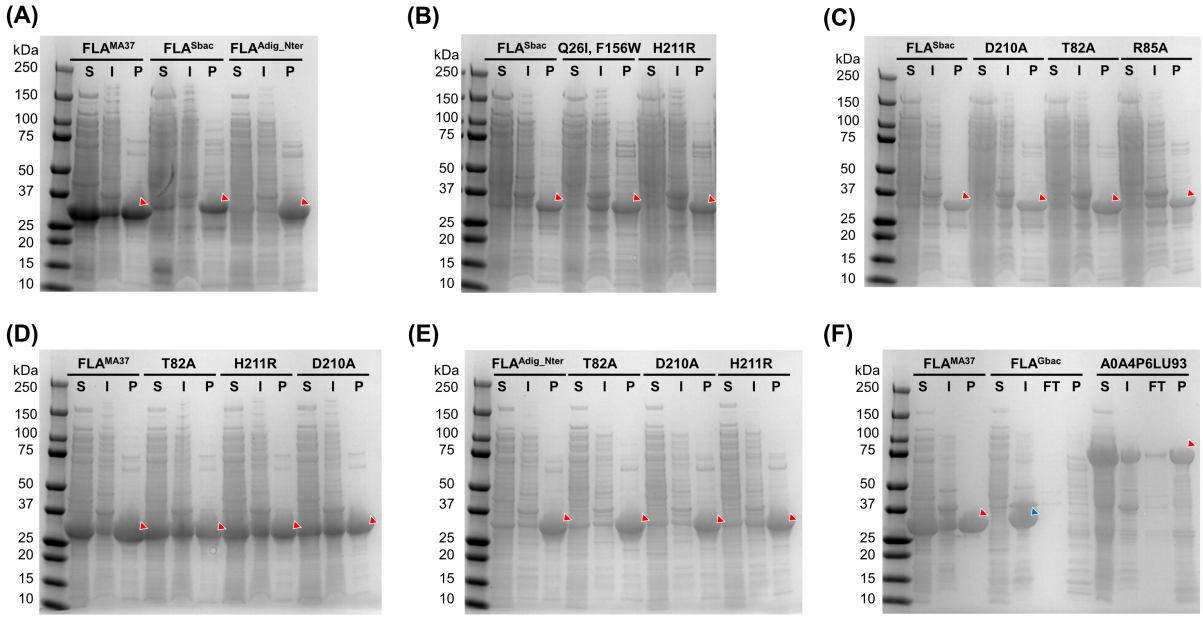

Fig. S11: Protein gels. (A) Wildtype fluorinases (FLA^{MA37}, FLA^{Sbac}, FLA^{Adig_Nter}), (B & C) FLA^{Sbac} and its mutants, (D) FLA^{MA37} and its mutants, (E) FLA^{Adig_Nter} and its mutants, (F) FLA^{MA37} and two wildtype fluorinases that do not show fluorination (FLA^{Gbac}, A0A4P6LU93).

S = soluble fractions of lysates, I = insoluble fractions of lysates (soluble and insoluble fractions' volume are normalized based on harvest OD), FT= Flow-through, P = purified proteins (25 µg). Red arrows indicated the purified proteins. Blue arrows indicated proteins stuck in insoluble fractions.

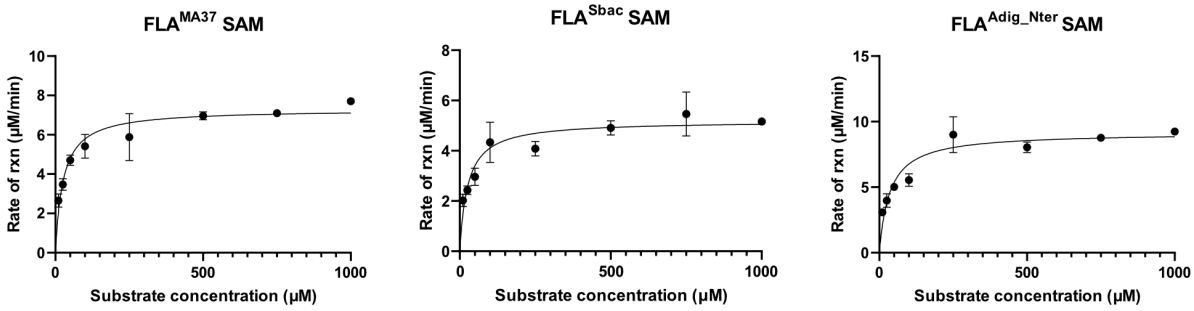

Fig. S12: Michaelis-Menten plots for FLA^{MA37}, FLA^{Sbac}, and FLA^{Adig_Nter} using SAM substrate, at 37 °C.

Table S1: List of subnetwork-1 members consisting of 16 previously identified known fluorinases and 4 newly identified putative fluorinases (highlighted in bold).

Name	Uniprot ID	Organism	Max seq-identity to known fluorinases (%)	IBS residues
FLA^{Gbac}	A0A7W1EMS4	*Gemmatimonadaceae bacterium*	68.9	TFFS
FLA^{Sbac}	A0A7J9ZI22	*Streptosporangiales bacterium*	78.8	TFYS
FLA^{Adig}	A0A7W7MQB5	*Actinoplanes digitatis*	93.7	TFYS
FLA^{Smor}	A0A7Y7B3E7	*Streptomyces morookaense*	97.9	TFYS
FLA^{MA37}	W0W999	*Streptomyces* sp. MA37	-	TFYS
FLA^{Scat}	Q70GK9	*Streptomyces cattleya*	-	TFYS
FLA^{Sxin}	A0A068VNW5	*Streptomyces xinghaiensis*	-	TFYS
FLA^{SAJ15}	A0A7M3LZW5	*Streptomyces* sp. SAJ15	-	TFYS
FLA^{N902}	R4LHX8	*Actinoplanes* sp. N902-109	-	TFYS
FLA^{Amza}	A0A1G9FQX8	*Actinopolyspora mzabensis*	-	TFYS
FLA^{Abar}	A0A8H9J0C4	*Amycolatopsis bartoniae*	-	TFYS
FLA^{CA12}	WP_103354124.1	*Amycolatopsis* sp. CA-128772	-	TFYS
FLA^{AN11}	A0A5B2WM47	*Goodfellowiella* sp. AN110305	-	TFYS
FLA^{Nbra2}	WP_029901962.1	*Nocardia brasiliensis* IFM 10847	-	TFYS
FLA^{Nbra3}	A0A379BKP5	*Nocardia brasiliensis* NCTC 11294	-	TFYS
FLA^{Nbra1}	W8JNL4	*Nocardia brasiliensis* ATCC 700358	-	TFYS
FLA^{Cbac}	A0A535N3L3	*Chloroflexi bacterium*	-	TFFS
FLA^{Pbac}	A0A0J1FI89	*Peptococcaceae bacterium* CEB3	-	TFFS
FLA^{Tnor}	A0A1I4QPY6	*Thermodesulforhabdus norvegica*	-	TFFS
FLA^{PtaU1}	A0A1V5AZT2	*Methanosaeta* sp. PtaU1.Bin055	-	TWFS

Table S2: HTP screening of fluorinase variants. Activity is given as fold comparison against FLA^{MA37} for both fluorination and chlorination. His: 6xHis tag, SUMO: Small ubiquitin-like modifier protein, MBP: maltose binding protein, ompA: signal peptide of outer membrane protein A, Nter: N-terminal linker from FLA^{MA37}. FLA^{Adig} without N-terminal linker from FLA^{MA37} has no activity.

Gene	ID	N-terminal	Fluorination	Chlorination
FLA^{Sbac}	121	His	0.3	0.07
	123	His	0.25	0.05
	230	His	0.25	0.06
	228	SUMO	0.11	0
	236	SUMO	0.17	0
	229	MBP	0.01	0
	232	MBP	0.01	0
	233	MBP	0.01	0
	122	OmpA	0.01	0
FLA^{Adig_Nter}	260	His+Nter	0.2	0.09
	261	His+Nter	0.21	0.1
	262	His+Nter	0.16	0.07

Table S3: List of putative fluorinases from subnetwork 2-4 selected for experimental testing.

Subnetwork ID	Uniprot ID	Organism	Max seq-identity to known fluorinases (%)	IBS residues
2	A0A1K1LRM2	Sinomicrobium oceani	28.68	IFPA
	A0A2U3B478	Flavobacteriaceae bacterium LYZ1037	25.77	SFPA
	A0A329N229	Sinomicrobium sp. N-1-3-6	27.91	IFPC
	A0A1M5AXG9	Arenibacter palladensis	27.13	SFPE
	A0A1M5XYQ6	Leeuwenhoekiella palythoae	27.52	NFTE
	A0A5C8V3H9	Flagellimonas hymeniacidonis	29.46	SFPM
	A0A1W6MHS0	Nonlabens spongiae	27.80	NFTM
	A0A7Y6XTT1	Flavobacteriaceae bacterium	26.61	SFPS
	A0A3N0ES88	Sinomicrobium pectinilyticum	27.91	IFPS
	A0A1Z8AQI3	Nonlabens dokdonensis	25.48	NFTT
	A0A3B0BU82	Ulvibacterium marinum	29.46	PFPT
	A0A7K0E2T2	Kriegella sp. EG-1	27.52	SFPT
	A0A432IJL1	Flavobacteriia bacterium	26.46	RYTT
	A0A497CI72	Bacteroidetes bacterium	26.56	SFAV
	A0A081DFQ2	Nonlabens ulvanivorans	27.13	NFTV
3	A0A1C5UCQ3	uncultured Clostridium sp.	27.21	IFHG
	A0A1C5YEJ9	uncultured Ruminococcus sp.	27.57	IFHG
	A0A4P6LU93	Blautia producta	29.30	IFHG
	C5EQ38	Clostridiales bacterium 1_7_47FAA	28.52	IFHG
4	A0A2H5XU78	Sinomicrobium oceani	27.48	TFHG

Table S4: Predicted mutations and the screening results of fold comparison against the respective wild-type (WT) of FLA^{MA37}, FLA^{Sbac}, and FLA^{Adig_Nter}, in terms of F^-/Cl^- selectivity at 1.5 h reaction for both fluorination and chlorination activity. F^-/Cl^- selectivity = % conversion of fluorination reaction / % conversion of chlorination reaction. N.T. = Not tested

	Mutation location	F^-/Cl^- Selectivity		
		FLA^{MA37}	FLA^{Sbac}	FLA^{Adig_Nter}
WT		11.53	10.43	8.65
H211R	Other	16.61	28.75	8.30
D210A	SAM-binding site	∞[#]	208.88	∞[#]
T82A	Ion-egress site	22.33	10.49	12.67
R85A	Ion-egress site	N.T.	∞[$]	N.T.

[#] Fluorination and chlorination activity: 0.03-fold and 0.00-fold, respectively.

[$] Fluorination and chlorination activity: 0.01-fold and 0.00-fold, respectively.

Table S5: Residue-wise binding free energies (kcal/mol) for IBS residues, computed form the MD-relaxed conformations using the Prime MMGBSA method. Atom-wise binding free energies were computed and subsequently aggregated to obtain residue-wise binding free energies.

IBS residue	FLA^{MA37}			FLA^{Sbac}			FLA^{Adig_Nter}		
	F^- bound	Cl^- bound	Difference	F^- bound	Cl^- bound	Difference	F^- bound	Cl^- bound	Difference
T155	-0.61	-0.50	-0.11	-0.28	-0.57	0.29	-0.26	-0.11	-0.15
F156	-2.74	-2.00	-0.74	-2.57	-1.19	-1.38	-1.29	-0.32	-0.97
Y157	-2.36	-1.64	-0.72	-2.38	-0.59	-1.79	-1.43	0.86	-2.29
S158	-2.87	2.09	-4.96	-0.33	2.79	-3.12	-3.21	2.19	-5.40

Table S6: Kinetic (k_{cat} and K_M) values for FLA^{MA37}, FLA^{Sbac}, and FLA^{Adig_Nter} using SAM substrate, at 37 °C.

Variant	K_M (μM)	k_{cat} (min^{-1})	k_{cat}/K_M (mM^{-1} min^{-1})
FLA^{MA37}	26.77 ± 4.06	0.37 ± 0.00	13.66 ± 2.15
FLA^{Sbac}	26.82 ± 7.41	0.26 ± 0.02	10.04 ± 1.90
FLA^{Adig_Nter}	35.85 ± 4.19	0.46 ± 0.02	12.93 ± 1.86

Table S7: Interaction frequency between SAM and protein residues. Residues positions where the interaction frequency difference >20% are shown in colour.

Residues	F^- bound trajectories			Cl^- bound trajectories		
	FLA^{MA37}	FLA^{MA37_D210A}	Difference	FLA^{MA37}	FLA^{MA37_D210A}	Difference
ASN215	63.4	40.8	22.6	60.6	47.5	13.1
ASP:21	98.0	95.8	2.2	92.7	91.1	1.6
ASP:210	100.0	0.0	100.0	99.7	0.0	99.7
ARG:270	73.6	28.4	45.2	75.3	20.6	54.7
TRP:50	98.1	98.8	-0.7	97.0	84.9	12.1
PHE:213	96.1	88.5	7.6	94.1	79.7	14.4
PHE:254	100.0	96.5	3.5	49.3	79.7	-30.4
	FLA^{Sbac}	FLA^{Sbac_D210A}	Difference	FLA^{Sbac}	FLA^{Sbac_D210A}	Difference
ASN215	70.0	55.5	14.5	64.8	47.7	17.1
ASP:21	100.0	93.4	6.6	97.3	80.3	17.0
ASP:210	100.0	0.0	100.0	98.3	0.0	98.3
ARG:270	98.0	14.9	83.1	79.3	36.7	42.6
PHE:50	62.4	71.5	-9.1	58.9	63.0	-4.1
PHE:213	92.2	77.3	14.9	93.5	71.7	21.8
PHE:254	98.2	89.6	8.6	83.0	88.9	-5.9
	FLA^{Adig_Nter}	$FLA^{Adig\_Nter\_D211A}$	Difference	FLA^{Adig_Nter}	$FLA^{Adig\_Nter\_D211A}$	Difference
ASN216	68.4	48.4	20.0	66.6	58.6	8.0
ASP:21	100.0	92.6	7.4	99.5	72.1	27.4
ASP:211	100.0	0.0	100.0	99.9	0.0	99.9
ARG:271	63.5	63.2	0.3	68.1	72.9	-4.8
TRP:50	98.4	98.4	0.0	90.2	51.6	38.6
PHE:214	96.0	81.5	14.5	92.5	66.3	26.2
PHE:255	100.0	97.9	2.1	96.3	97.6	-1.3

Table S8: Mutagenesis primers

Primer ID	Description	Sequence (5' – 3')
ppS130	FLA^{Sbac_H210R} (forward)	TTCCCAAACGGACGGTCAATCGCACTCACTT
ppS131	FLA^{Sbac_H210R} (reverse)	AAGTGAGTGCGATTGACCGTCCGTTTGGGAA
ppS315	FLA^{MA37_T82} (forward)	TATCCCGCGACCGGCGCGACGACCCGCTCCGTG
ppS316	FLA^{MA37_T82} (reverse)	CACGGAGCGGGTCGTCGCGCCGGTCGCGGGATA
ppS317	FLA^{MA37_H211R} (forward)	GTCACCGCGATCGACCGTCCCTTCGGCAACATC
ppS318	FLA^{MA37_H211R} (reverse)	GATGTTGCCGAAGGGACGGTCGATCGCGGTGAC
ppS319	FLA^{MA37_D210A} (forward)	GTCGTCACCGCGATCGCGCACCCCTTCGGCAAC
ppS320	FLA^{MA37_D210A} (reverse)	GATGTTGCCGAAGGGACGGTCGATCGCGGTGAC

Movie S1: Inter-monomer interaction between H211 and T20 in FLA^{Sbac}. The shorter side chain of histidine (H211) limits its interactions with threonine (T20), preventing it from interacting with E54. Salt-bridge or hydrogen bond interactions between these amino-acids are shown in dashed magenta line. SAM is shown in grey spheres.

Movie S2: Inter-monomer interaction between R211 and T20, and intra-monomer interaction between R211 and E54 in FLA^{Sbac_H211R}. The longer side chain of arginine (R211) in FLA^{Sbac_H211R} compared to histidine in FLA^{Sbac}, enables it to form additional intra-monomer interactions with E54. Salt-bridge or hydrogen bond interactions between these residues are depicted by dashed magenta lines. SAM is shown in grey spheres.

Movie S3: Interaction network extending from R211 to F50 in FLA^{Sbac_H211R}. Salt-bridge or hydrogen bond interactions between these residues are depicted by dashed magenta lines. SAM is shown in grey spheres.

Movie S4: Release of the fluoride ion bound FLA^{Adig_Nter} trajectory. The IES forming β3-β4 loop is shown in orange. Polar/charged residues in green. The IBS S158 is shown in grey. The fluoride ion (cyan) stabilizing interactions, with distances <2.5 Å, are shown in magenta.