Supporting information for

Coupling a Live Cell Directed Evolution Assay with Coevolutionary Landscapes to Engineer an Improved Fluorescent Rhodopsin Chloride Sensor

Hsichuan Chi[†‡§], Qin Zhou[‡§], Jasmine N. Tutol[†§], Shelby M. Phelps[†§], Jessica Lee[†‡], Paarth Kapadia[‡], Faruck Morcos[‡^]*, Sheel C. Dodani[†]*

Departments of [†]Chemistry and Biochemistry, [‡]Biological Sciences, and [^]Bioengineering, The University of Texas at Dallas, Richardson, TX 75080

[§]H.C., Q.Z., J.N.T., and S.M.P. contributed equally to this work.

*sheel.dodani@utdallas.edu, *faruckm@utdallas.edu

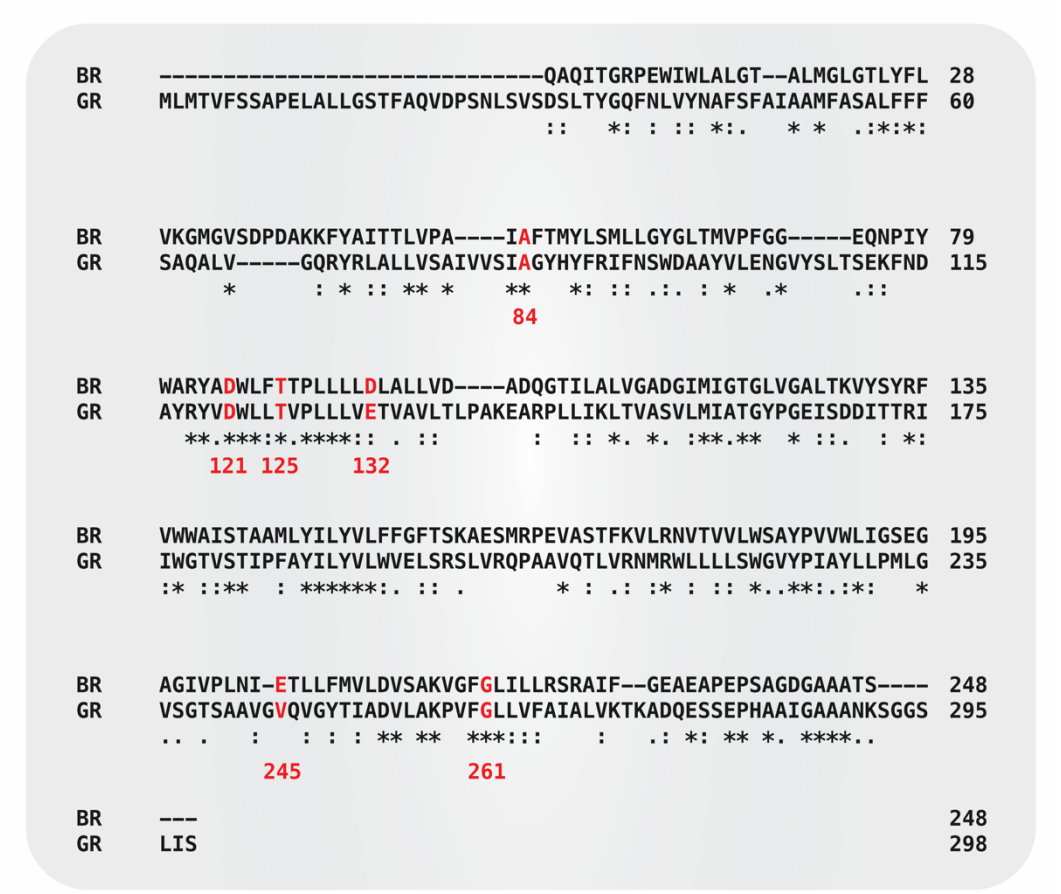

Figure S1. Multiple sequence alignment of bacteriorhodopsin (BR, PDB ID: 1FBB) and *Gloeobacter* rhodopsin (*wt*GR, PDB ID: 6NWD) generated using Clustal Omega.[1] Sequences were derived from the respective Protein Data Bank files. Residues along the proton-pumping pathway selected for this study are shown in red and labeled with the position number from *wt*GR.

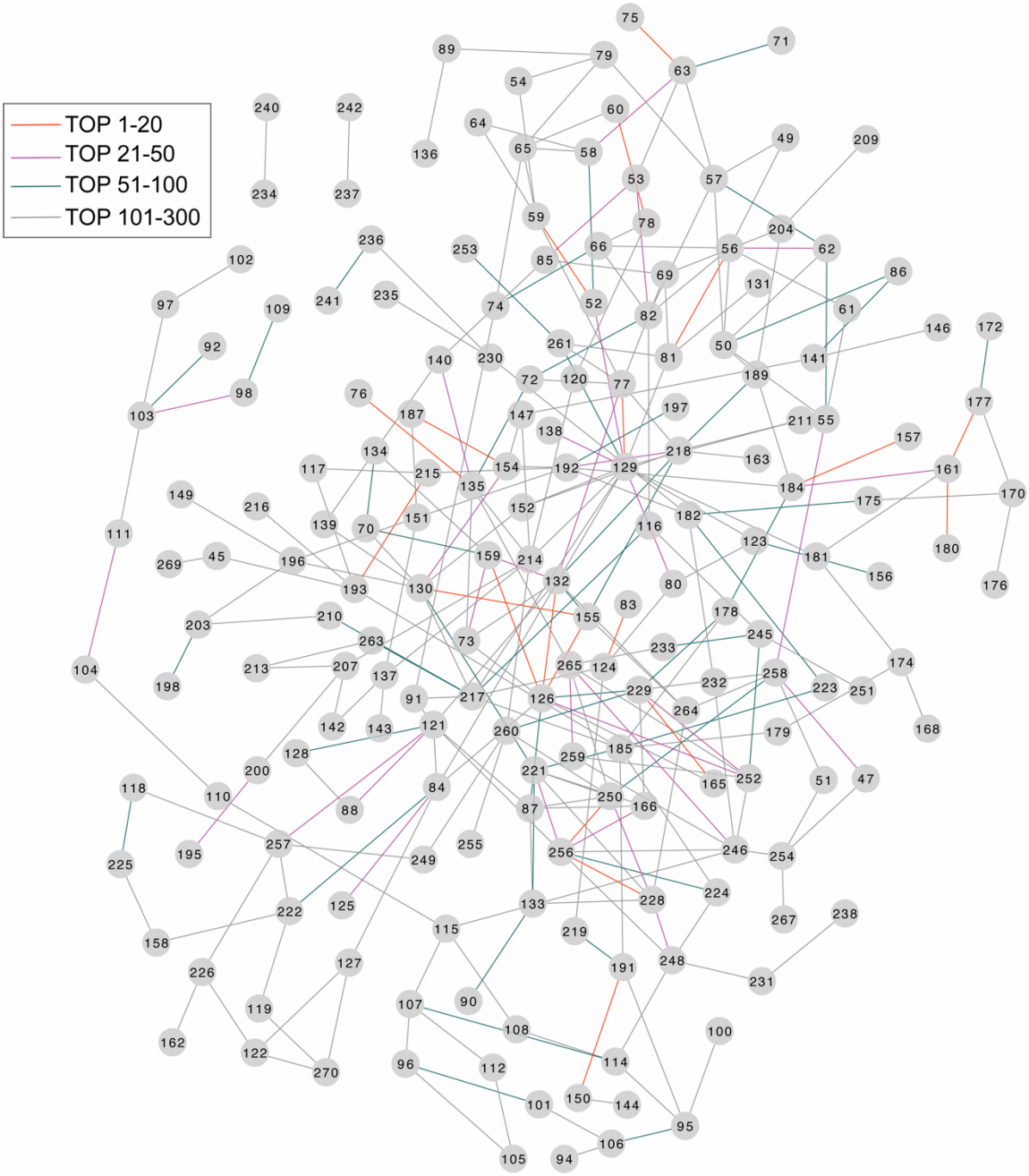

Figure S2. Connectivity map for the top 300 direct information (DI) pairs. Residues that maintain strong direct evolutionary couplings are shown in the map with position numbers mapped to *wt*GR in gray circles. Each DI pair is linked with a line. The colors of the lines represent the rank of the strength of the couplings for the DI pairs.

Supplemental File 1. Table of the top 300 DI pairs.

Table S1. Distances for the top DI pairs for the selected proton-pumping pathway residues. The shortest distances between the sidechains are reported.

Residue	Amino acid	Paired residue	Amino acid	Distance (Å)
84	ALA	121	VAL	5.2
		125	THR	4.6
		126	VAL	8.9
		127	PRO	11.3
		129	LEU	8.4
		222	TRP	12.9
121	VAL	88	TYR	5.6
		84	ALA	5.2
		91	ILE	4.8
		87	HIS	2.9
		128	LEU	7.5
		132	GLU	11.3
		256	ALA	10.1
		257	LYS	5.7
125	THR	84	ALA	4.6
132	GLU	73	ALA	4.1
		77	SER	2.8
		121	VAL	11.3
		126	VAL	6.9
		155	SER	11.0
		159	ILE	14.5
		218	LEU	9.4
		264	VAL	3.8
245	VAL	116	ALA	11.2
		233	MET	8.8
		251	ILE	7.0
		252	ALA	7.8
261	GLY	77	SER	4.4
		81	VAL	8.3
		129	LEU	6.9
		253	ASP	8.2

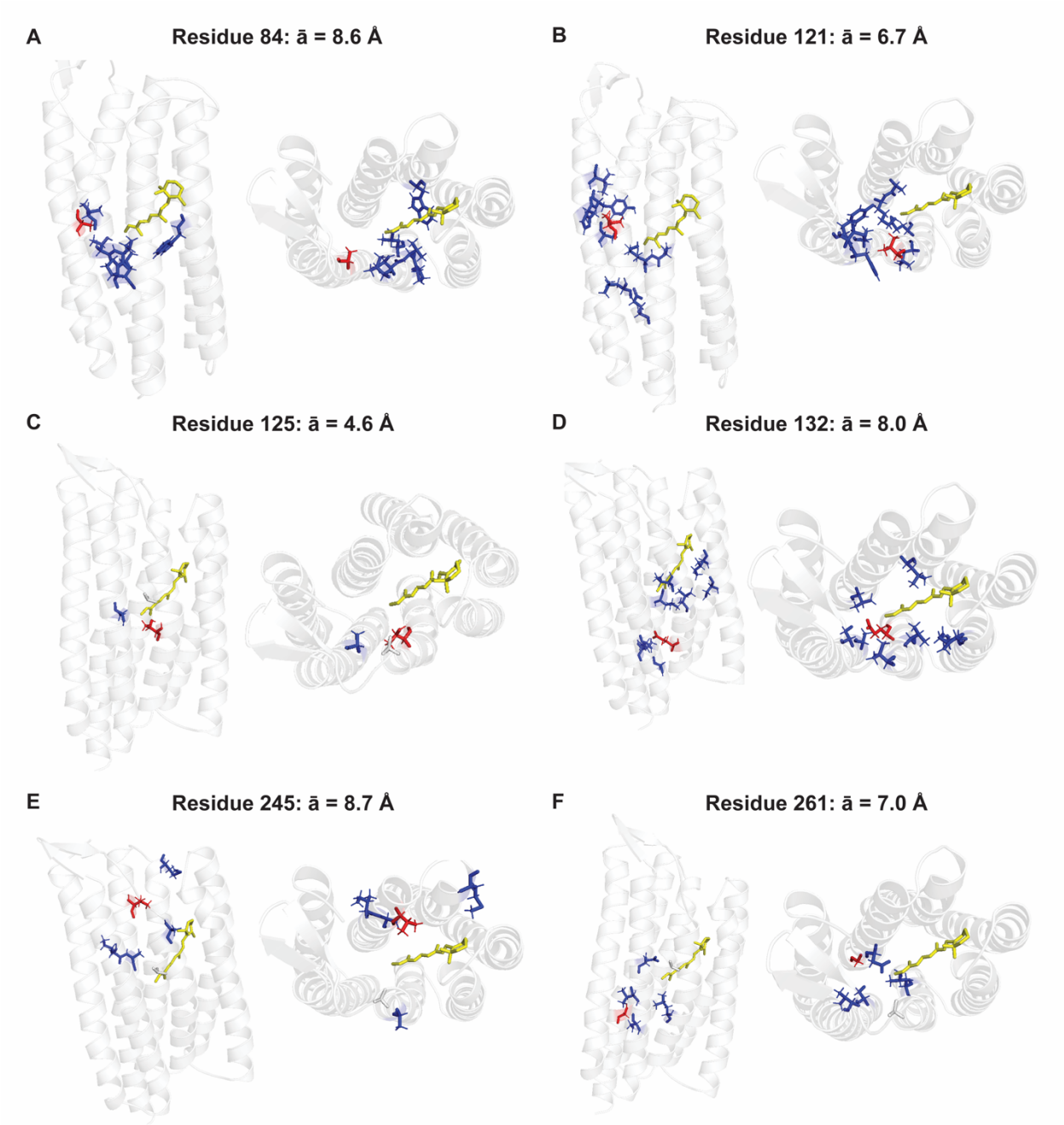

Figure S3. (A–F) Comparison of the residues along the proton-pumping pathway (red sticks) with the corresponding coevolved residues (blue sticks) for the top DCA pairs shown in Figure 2. In each panel, the side (left) and top (right) views are shown for the homology model of the rhodopsin chloride sensor GR1 only with the Schiff base chromophore (yellow sticks). The average distances ($\bar{a}$) from the selected residues along the proton-pumping pathway (red sticks) to the corresponding coevolved residues (blue sticks) are listed in Table S1.

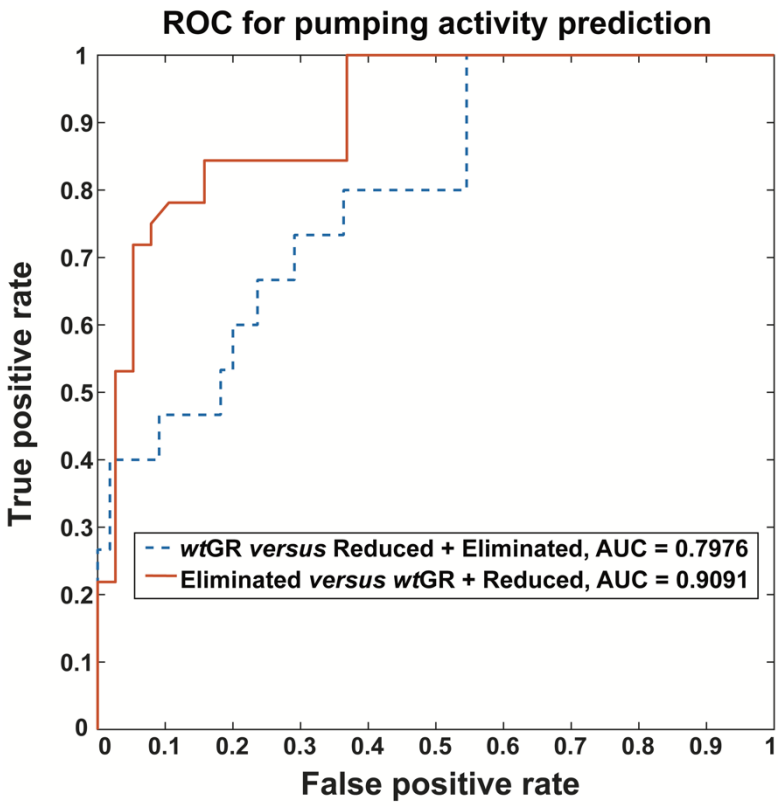

Figure S4. Receiver operating characteristic (ROC) curve for the Hamiltonian score prediction. Based on the proton-pumping activity, two one-vs.-rest (OVR) classifiers were used to create the ROC curve: *wt*GR-like (*wt*GR-like *versus* Reduced + Eliminated) and eliminated (Eliminated *versus* *wt*GR-like + Reduced). For the *wt*GR-like classifier, the curve is shown as a blue dashed line. The area under this ROC curve is 0.7976. For the eliminated classifier, the curve is shown as a solid red line. The area under this ROC curve is 0.9091.

Table S2. List of the primers used to generate the site-saturation mutagenesis (SSM) library at each position. The mutation site is in red.

Description	Primer (5' to 3')
132 SSM Forward	GTGCCTCTGTTGCTGGTGNDTACAGTGGCAGTGCTGA
	GTGCCTCTGTTGCTGGTGVHGACAGTGGCAGTGCTGA
	GTGCCTCTGTTGCTGGTGTGGACAGTGGCAGTGCTGA
132 SSM Reverse	CACCAGCAACAGAGGCACGGTCAACAGCCAAACC
84 SSM Forward	GCAATTGTTGTGAGTATCNDTGGGTACCACTACTTTC
	GCAATTGTTGTGAGTATCVHGGGGTACCACTACTTTC
	GCAATTGTTGTGAGTATCTGGGGGTACCACTACTTTC
84 SSM Reverse	GATACTCACAACAATTGCTGAAACAAGCAAGGCC
125 SSM Forward	TATGTGGTTTGGCTGTTGNDTGTGCCTCTGTTGCTGG
	TATGTGGTTTGGCTGTTGVHGGTGCCTCTGTTGCTGG
	TATGTGGTTTGGCTGTTGTGGGTGCCTCTGTTGCTGG
125 SSM Reverse	CAACAGCCAAACCACATAGCGGTAGGCGTCGTTG
245 SSM Forward	ACGTCCGCGGCTGTCGGCNDTCAGGTTGGCTATACGA
	ACGTCCGCGGCTGTCGGCVHGCAGGTTGGCTATACGA
	ACGTCCGCGGCTGTCGGCTGGCAGGTTGGCTATACGA
245 SSM Reverse	GCCGACAGCCGCGGACGTACCGGATACTCCAAGC
261 SSM Forward	CTGGCGAAGCCTGTATTTNDTCTTCTAGTCTTCGCGA
	CTGGCGAAGCCTGTATTTVHGCTTCTAGTCTTCGCGA
	CTGGCGAAGCCTGTATTTTGGCTTCTAGTCTTCGCGA
261 SSM Reverse	AAATACAGGCTTCGCCAGCACGTCTGCGATCGTA

CATATGTTGATGACCGTATTTTCTTCTGCACCTGAACTTGCCCTTCTCGGATCAACCTTTGC
CCAGGTCGATCCTTCAAACTTATCGGTCTCAGATTCGCTGACCTATGGTCAGTTCAATCTG
GTTTACAACGCTTTCTCGTTTGCCATCGCGGCAATGTTCGCATCTGCCCTCTTCTTCTTCAG
CGCTCAGGCACTCGTCGGTCAACGATACCGGTTGGCCTTGCTTGTTTCAGCAATTGTTGTG
AGTATCAAGGGGTACCACTACTTTCGGATCTTCAATAGTTGGGATGCTGCCTACGTTCTGG
AGAATGGCGTGTATTCCCTGACTAGCGAAAAATTCAACGACGCCTACCGCTATGTGGTTTG
GCTGTTGTGTGTGCCTCTGTTGCTGGTGAAGACAGTGGCAGTGCTGACGTTGCCTGCAAA
GGAGGCAAGACCCTTGCTGATCAAACTGACGGTGGCTTCAGTTCTGATGATTGCCACGGG
CTACCCCGGCGAGATTTCTGACGACATTACGACTCGCATCATCTGGGGTACGGTCAGCAC
GATTCCCTTCGCCTACATCCTCTATGTGTTGTGGGTCGAACTGTCCAGGTCCCTTGTCCGC
CAGCCCGCTGCTGTACAAACCCTGGTCCGCAACATGCGGTGGCTGCTGTTGCTCTCCTGG
GGTGTTTACCCGATCGCATACCTTCTACCCATGCTTGGAGTATCCGGTACGTCCGCGGCT
GTCGGCATTCAGGTTGGCTATACGATCGCAGACGTGCTGGCGAAGCCTGTATTTGGTCTT
CTAGTCTTCGCGATTGCACTCGTGAAAACAAAAGCAGATCAAGAAAGCAGTGAACCACATG
CCGCAATAGGTGCTGCTGCAAATAAATCGGGAGGCAGTCTTATCTCCGCGGCCGCAATGG
TGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGC
GACGTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGG
CAAGCTGACCCTGAAGTTCATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCT
CGTGACCACCCTGACCTGGGGCGTGCAGTGCTTCAGCCGCTACCCCGACCACATGAAGC
AGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGCGCACCATCTTCT
TCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTG
GTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCA
CAAGCTGGAGTACAACTACATCAGCCACAACGTCTATATCACCGCCGACAAGCAGAAGAAC
GGCATCAAGGCCAACTTCAAGATCCGCCACAACATCGAGGACGGCAGCGTGCAGCTCGC
CGACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCTGCTGCCCGACAACC
ACTACCTGAGCACCCAGTCCGCCCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGG
TCCTGCTGGAGTTCGTGACCGCCGCCCTCGAGCACCACCACCACCACCACTGA

Figure S5. Nucleotide sequence encoding GR2-CFP in the pET-21a(+) vector. The following regions are highlighted: restriction sites and the *A* nucleotide (orange), GR2 (black), mutation sites in GR2 (red), CFP (blue), His-tag (purple), and stop codon (green).

Figure S6. Amino acid sequence of GR2-CFP. Colors correspond to the description in Figure S5.

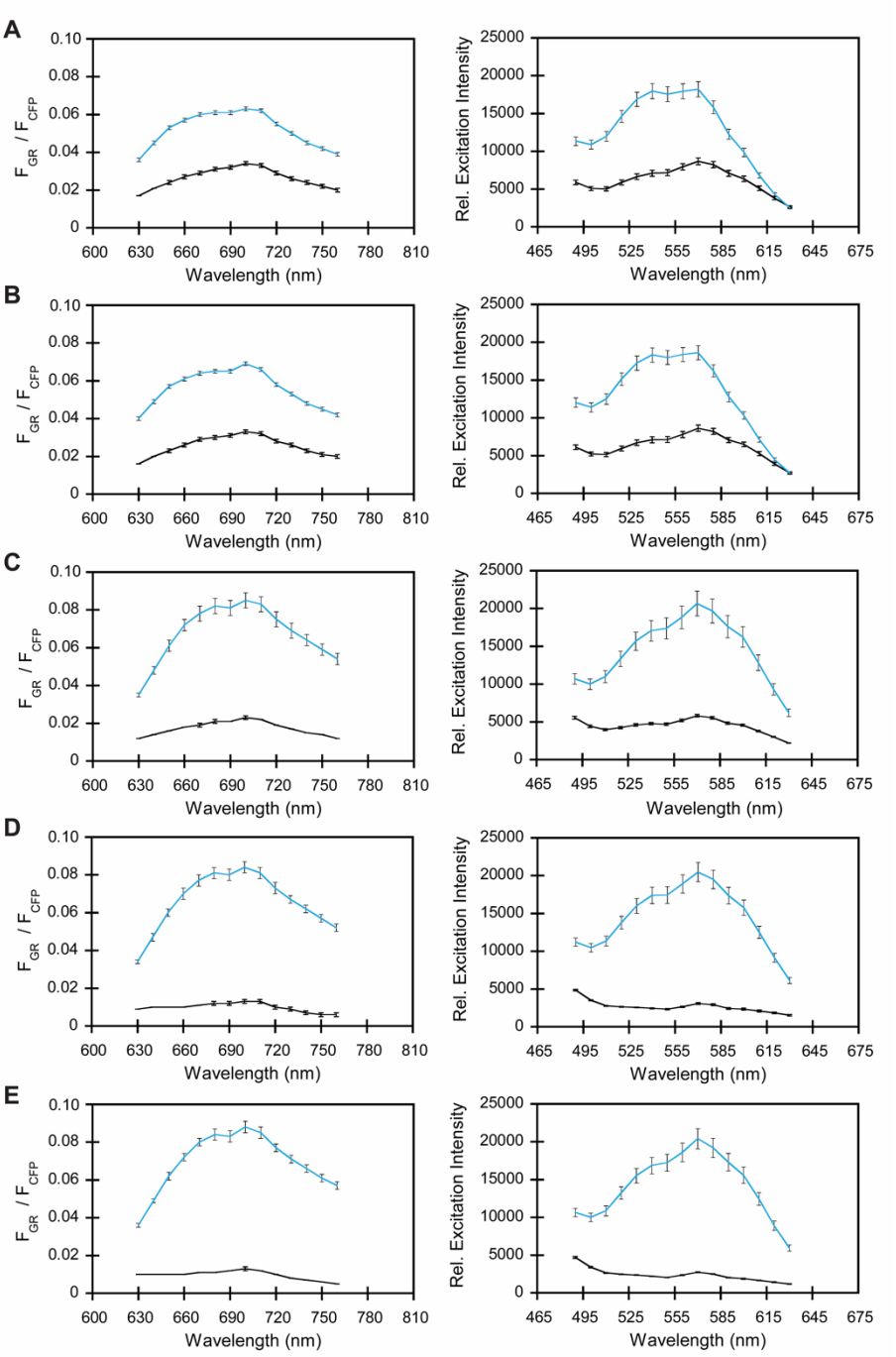

Figure S7. Characterization of the engineered GR1-CFP variants in *Escherichia coli* treated with 0 (black) and 400 mM (blue) sodium chloride in 50 mM sodium acetate buffer at pH 5. The emission spectra of the rhodopsin normalized by the CFP emission at 485 nm (left) and excitation spectra of the rhodopsin emission at 710 nm (right) are shown for (A) GR1-CFP, (B) GR1 E132K-CFP, (C) GR1 E132K/A84K-CFP, (D) GR1 E132K/A84K/T125C-CFP, and (E) GR2-CFP.

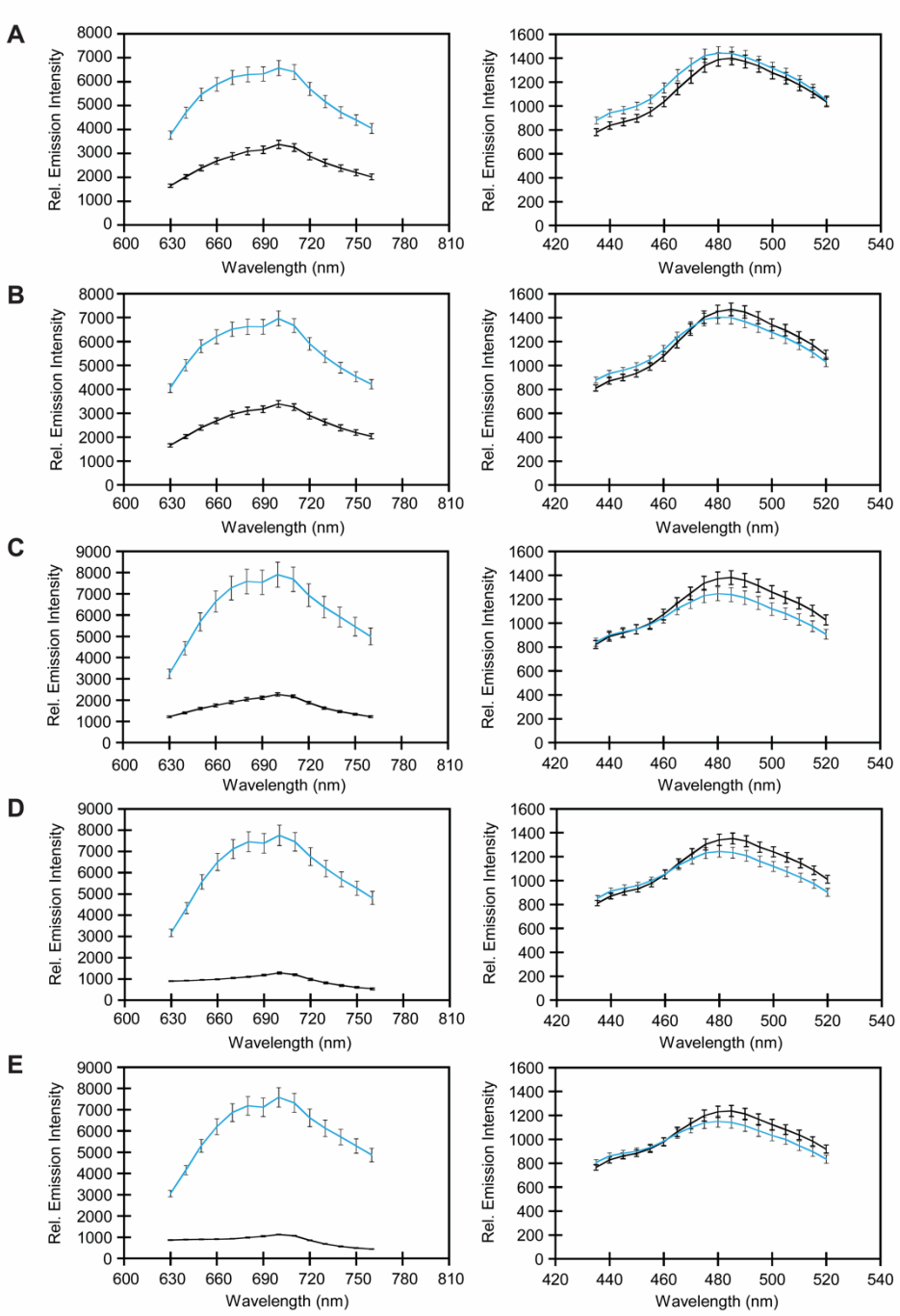

Figure S8. The emission spectra of the rhodopsin from 630–760 nm (left) and CFP from 435–520 nm (right) are shown for (A) GR1-CFP, (B) GR1 E132K-CFP, (C) GR1 E132K/A84K-CFP, (D) GR1 E132K/A84K/T125C-CFP, and (E) GR2-CFP from Figure S7.

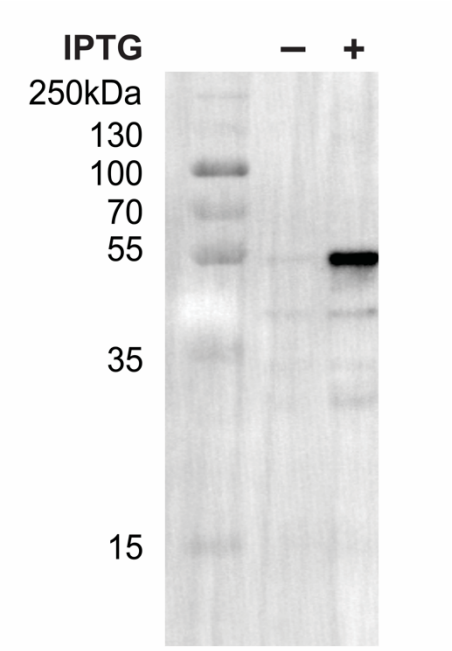

Figure S9. Representative western blot analysis using an anti-6x-His tag antibody shows that GR2-CFP is expressed in the isopropyl β-D-1-thiogalactopyranoside (IPTG) induced sample. The expected molecular weight for the monomer is ~59.3 kDa.

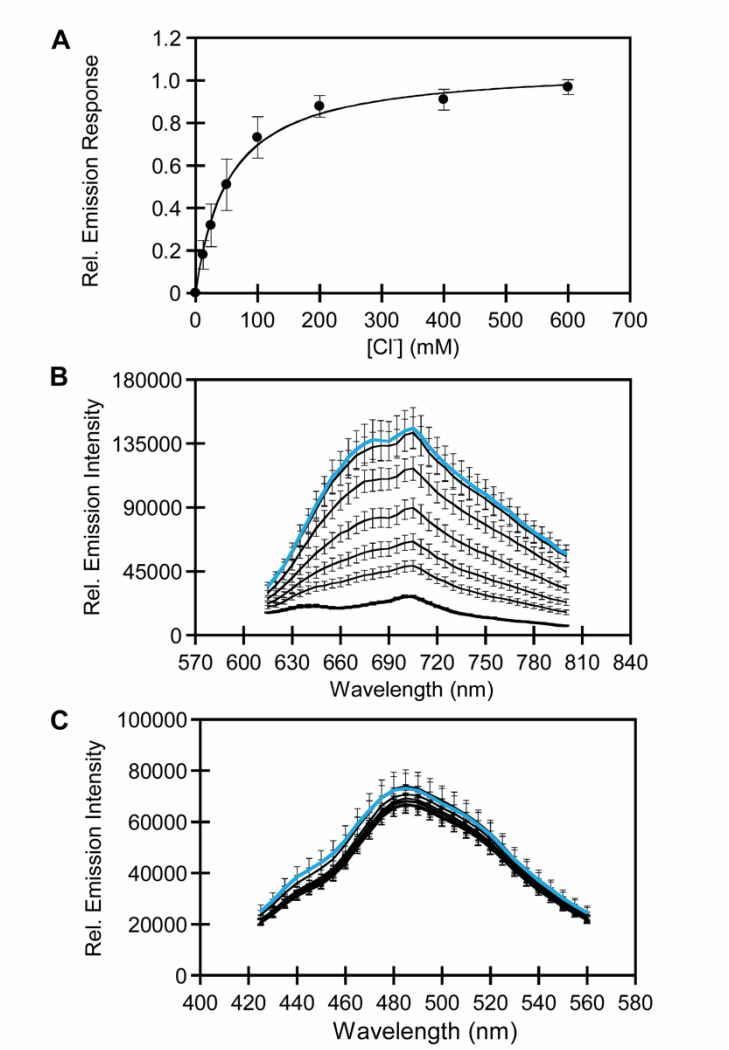

Figure S10. (A) Normalized integrated emission response (F_{GR2}/F_{CFP}) of GR2-CFP expressed in *E. coli* to 0, 12.5, 25, 50, 100, 200, 400, and 600 mM sodium chloride in 50 mM sodium acetate buffer at pH 5 to determine the apparent dissociation constant (K_d) from Figure 4A. Note: the additional data point for 600 mM sodium chloride is not shown in Figure 4A. The K_d for Cl^- is 53 ± 14 mM. (B) Emission spectra of GR2 for GR2-CFP from Figure 4A. (C) Emission spectra of CFP for GR2-CFP from Figure 4A. For the rhodopsin, excitation was provided at 570 nm, and the emission was collected and integrated from 615–800 nm (F_{GR2}). For CFP, excitation was provided at 390 nm, and the emission was collected from 425–560 nm. The emission spectra of GR2 at each point were normalized by the CFP emission intensity at 485 nm (F_{CFP}). The average of nine biological replicates from three technical replicates with standard error of the mean is shown.

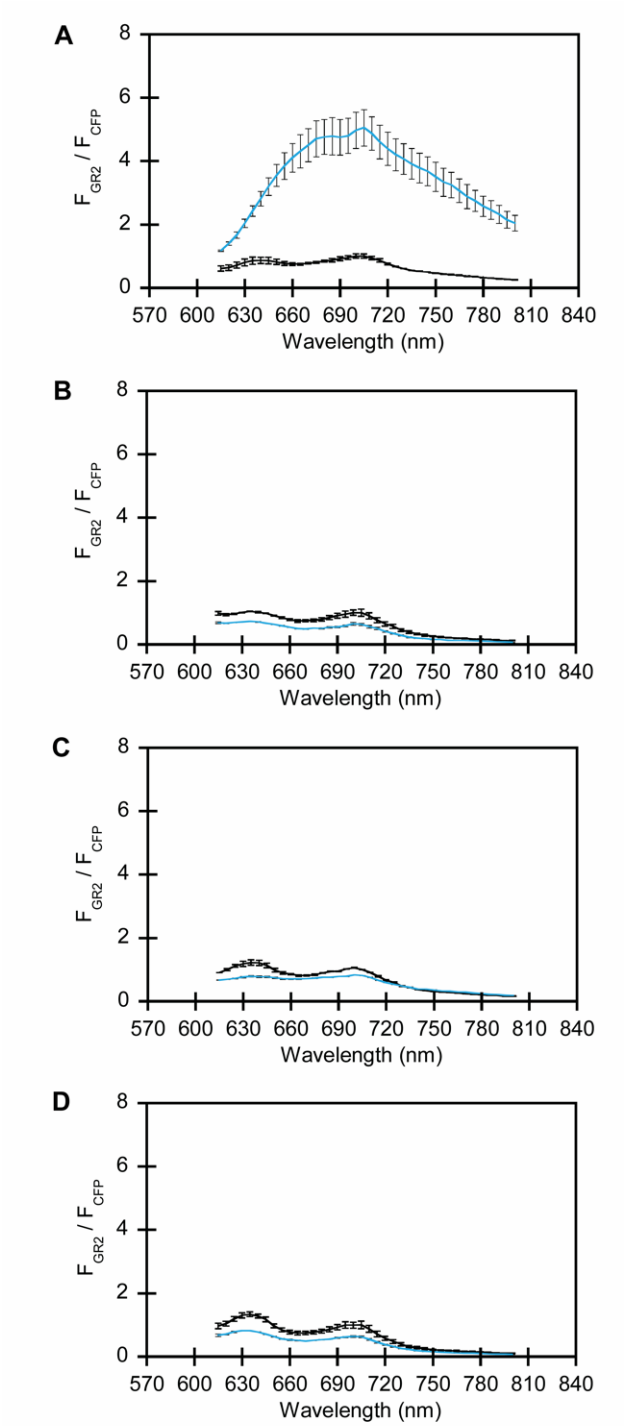

Figure S11. Normalized emission spectra of GR2-CFP in *E. coli* treated with 0 (black) or 400 mM (blue) sodium chloride in 50 mM sodium acetate buffer at pH 5 following expression (A) with IPTG and all-*trans*-retinal (ATR), (B) with IPTG only, (C) with ATR only, and (D) without IPTG and ATR. For the rhodopsin, excitation was provided at 570 nm, and the emission was collected from 615–800 nm. For CFP, the excitation was provided at 390 nm, and the emission was collected from 425–560 nm. The emission spectra of GR2 at each point were normalized by the CFP emission intensity at 485 nm (F_{CFP}). The average of nine biological replicates from three technical replicates with standard error of the mean is shown.

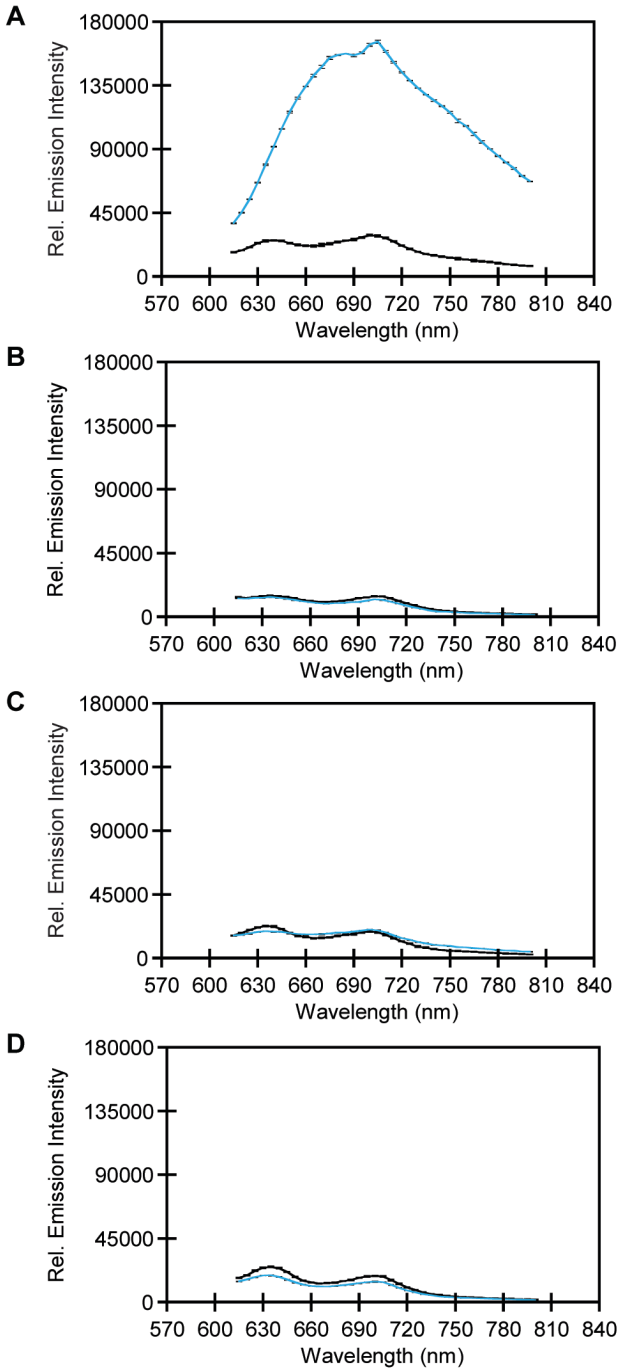

Figure S12. Emission spectra of GR2 for GR2-CFP in *E. coli* treated with 0 (black) or 400 mM (blue) sodium chloride following expression (A) with IPTG and ATR, (B) with IPTG only, (C) with ATR only, and (D) without IPTG and ATR from Figure S11.

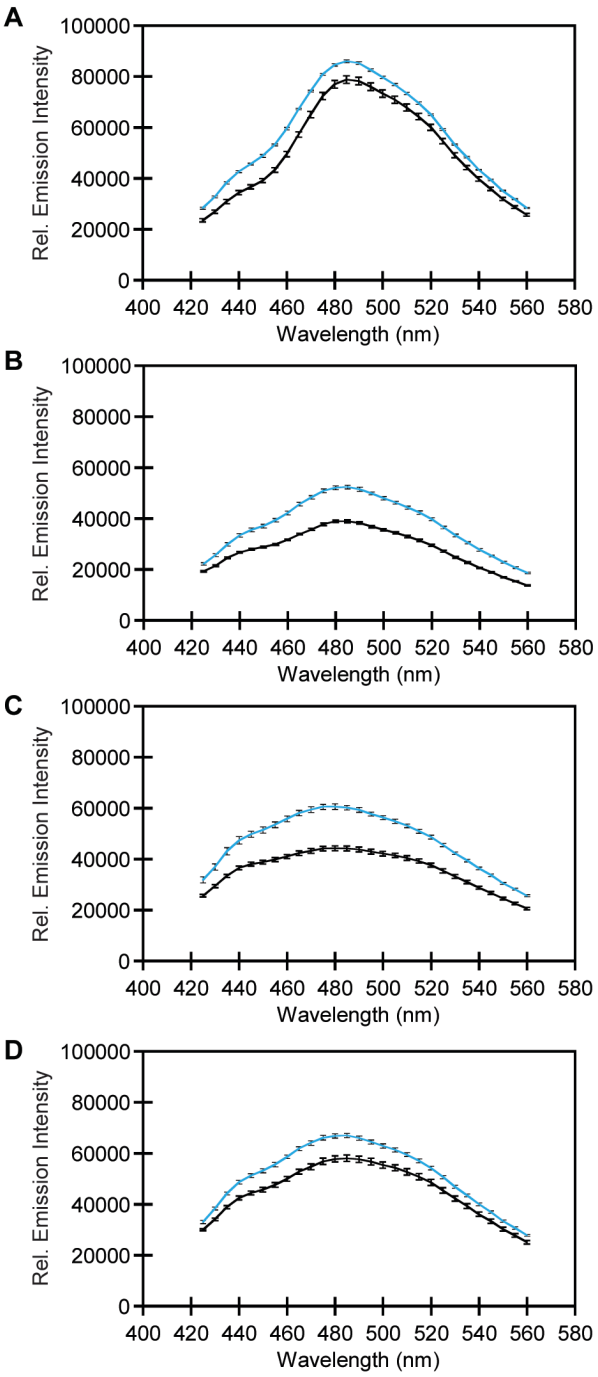

Figure S13. Emission spectra of CFP for GR2-CFP in *E. coli* treated with 0 (black) or 400 mM (blue) sodium chloride following expression (A) with IPTG and ATR, (B) with IPTG only, (C) with ATR only, and (D) without IPTG and ATR from Figure S11.

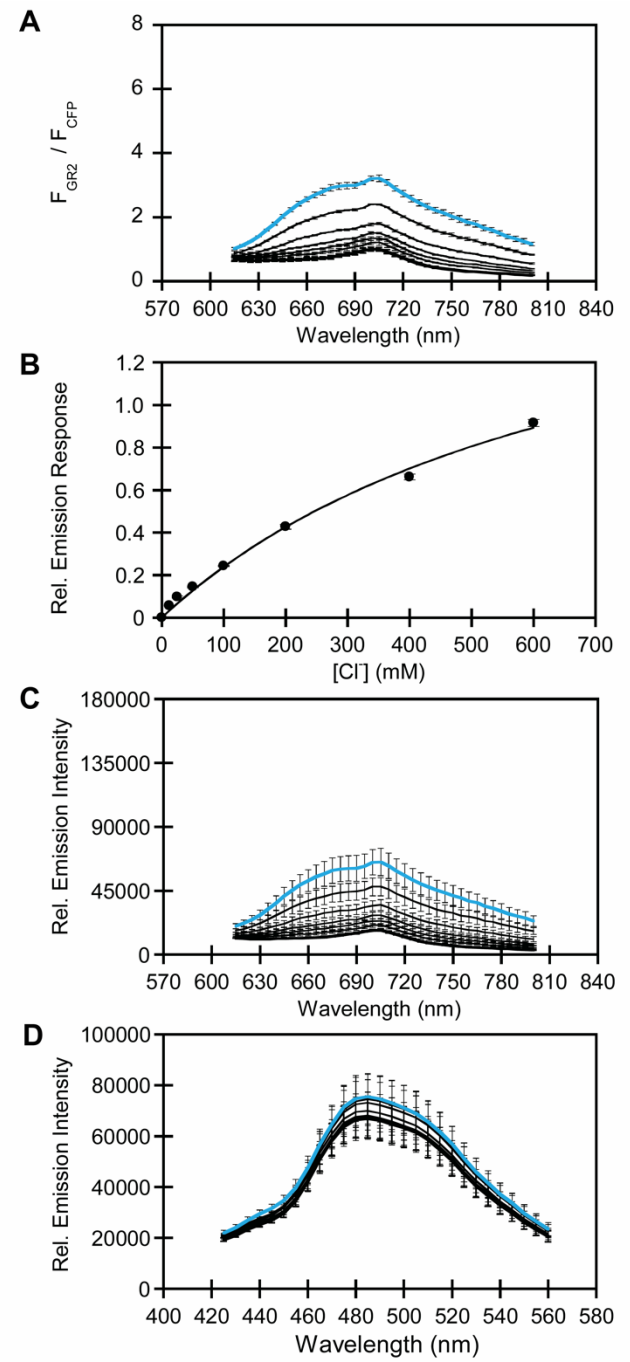

Figure S14. (A) Normalized emission spectra of GR2-CFP in *E. coli* treated with 0 (black and bold), 12.5, 25, 50, 100, 200, and 400 mM (blue) sodium chloride in 50 mM sodium phosphate buffer at pH 6. (B) Normalized integrated emission response (F_{GR2}/F_{CFP}) from *A* to determine the K_d. Note: the additional data point for 600 mM sodium chloride is not shown in *A*. The K_d for Cl^- is 731 ± 31 mM. (C) Emission spectra of GR2 from *A*. (D) Emission spectra of CFP from *A*. For the rhodopsin, excitation was provided at 570 nm, and the emission was collected and integrated from 615–800 nm (F_{GR2}). For CFP, excitation was provided at 390 nm, and the emission was collected from 425–560 nm (F_{CFP}). The emission spectra of GR2 at each point were normalized by the CFP emission intensity at 485 nm (F_{CFP}). The average of nine biological replicates from three technical replicates with standard error of the mean is shown.

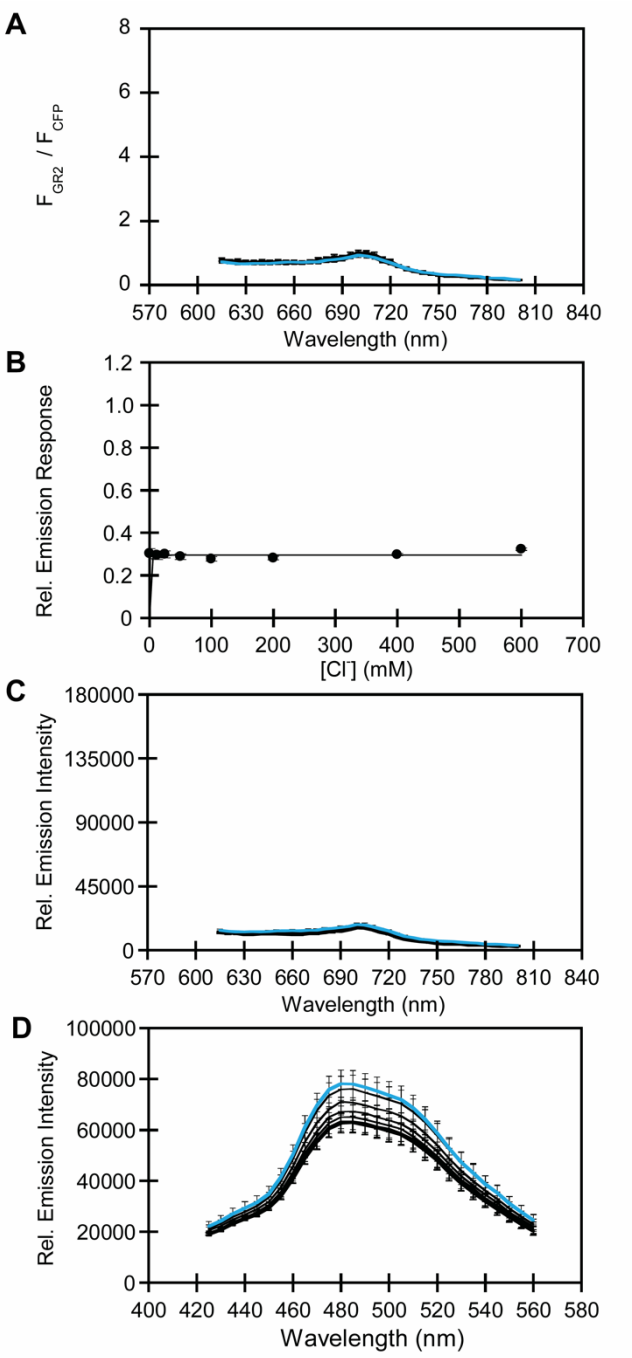

Figure S15. (A) Normalized emission spectra of GR2-CFP in *E. coli* treated with 0 (black and bold), 12.5, 25, 50, 100, 200, and 400 mM (blue) sodium chloride in 50 mM sodium phosphate buffer at pH 7. (B) Normalized integrated emission response (F_{GR2}/F_{CFP}) from *A.* Note: the additional data point for 600 mM sodium chloride is not shown in *A*. The K_d could not be determined. (C) Emission spectra of GR2 from *A*. (D) Emission spectra of CFP from *A*. For the rhodopsin, excitation was provided at 570 nm, and the emission was collected and integrated from 615–800 nm (F_{GR2}). For CFP, the excitation was provided at 390 nm, and the emission was collected from 425–560 nm (F_{CFP}). The emission spectra of GR2 at each point were normalized by the CFP emission intensity at 485 nm (F_{CFP}). The average of nine biological replicates from three technical replicates with standard error of the mean is shown.

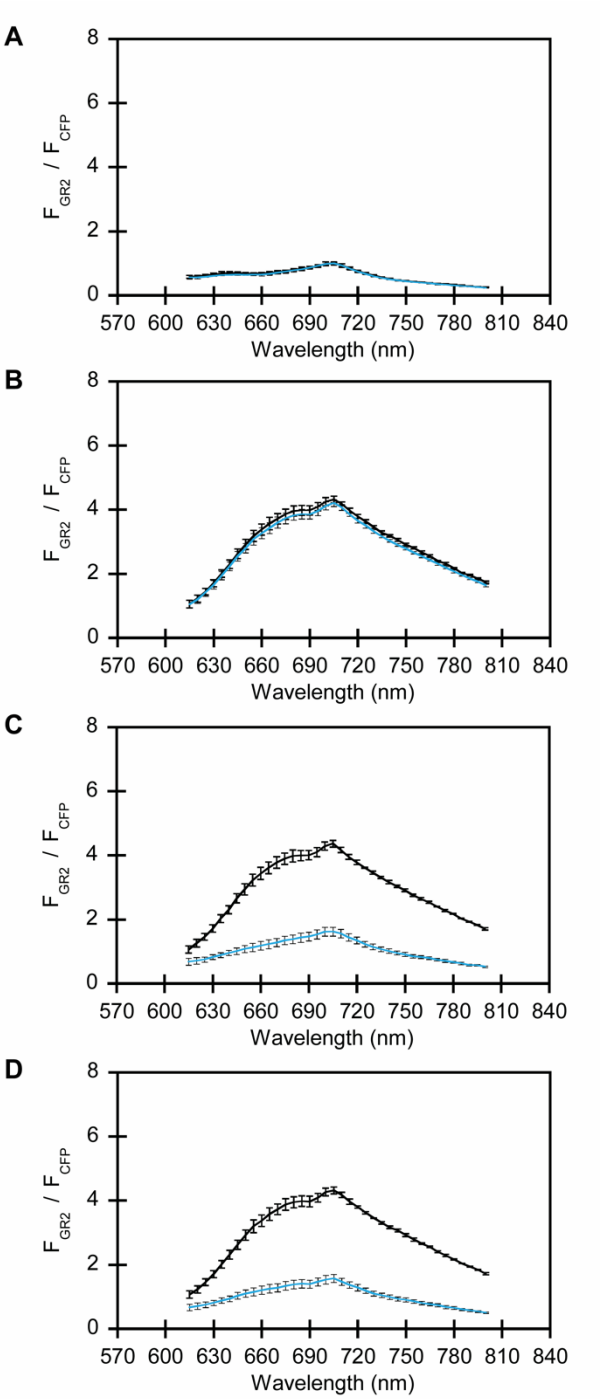

Figure S16. Normalized emission spectra of GR2-CFP in *E. coli* treated first in the presence of buffer only (control) or 400 mM sodium chloride (black) in 50 mM sodium acetate buffer at pH 5, followed by washing with (A) buffer only for the control sample, and (B) 400 mM sodium chloride, (C) buffer, or (D) 400 mM sodium gluconate for the samples treated with sodium chloride (blue) for Figure 4B. For the rhodopsin, excitation was provided at 570 nm, and the emission was collected from 615–800 nm. For CFP, the excitation was provided at 390 nm, and the emission was collected from 425–560 nm. The emission spectra of GR2 at each point were normalized by the CFP emission intensity at 485 nm (F_{CFP}). The average of nine biological replicates from three technical replicates with standard error of the mean is shown.

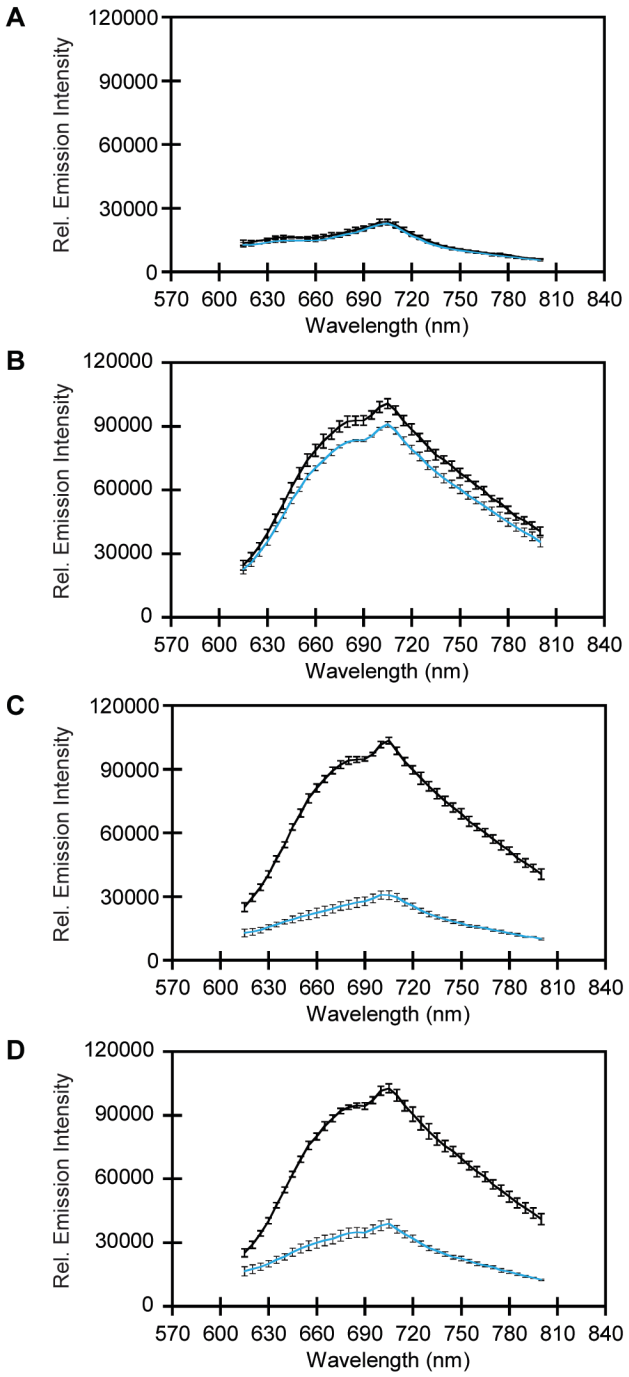

Figure S17. Emission spectra of GR2 for GR2-CFP in *E. coli* treated first in the presence of buffer only (control) or 400 mM sodium chloride (black), followed by washing with (A) buffer only for the control sample, and (B) 400 mM sodium chloride, (C) buffer, or (D) 400 mM sodium gluconate for the samples treated with sodium chloride (blue) from Figure S16.

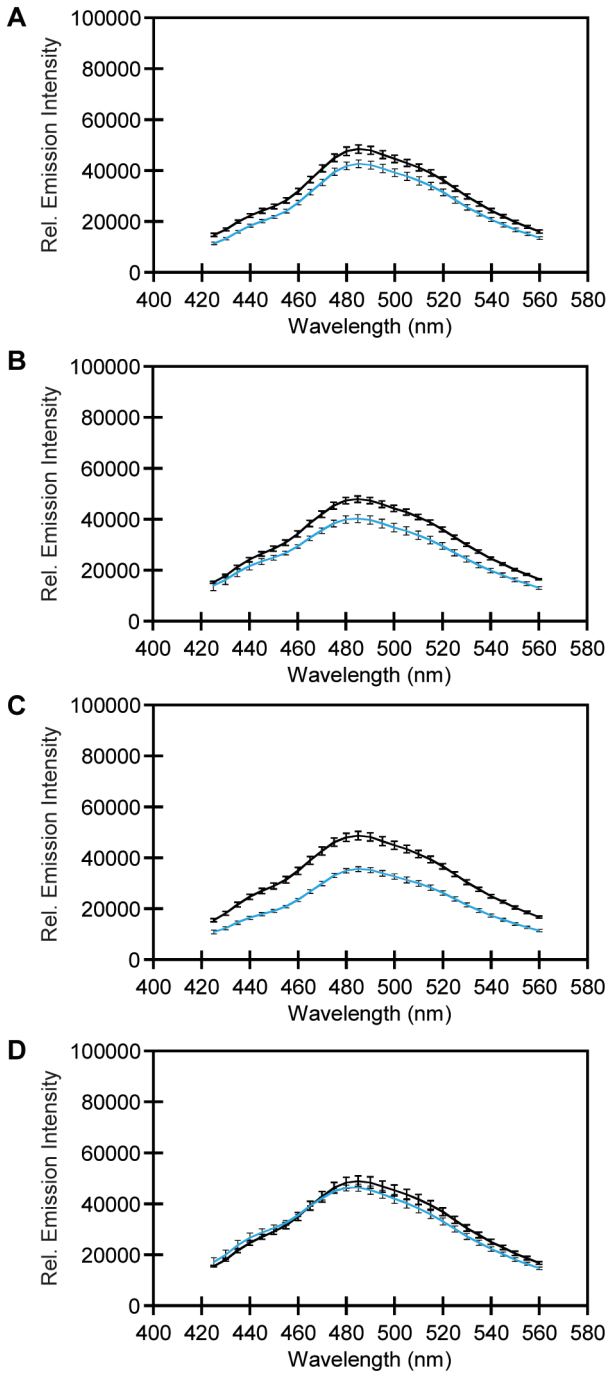

Figure S18. Emission spectra of CFP for GR2-CFP in *E. coli* treated first with buffer only (control) or 400 mM sodium chloride (black), followed by washing with (A) buffer only for the control sample, and (B) 400 mM sodium chloride, (C) buffer, or (D) 400 mM sodium gluconate for the samples treated with sodium chloride (blue) from Figure S16.

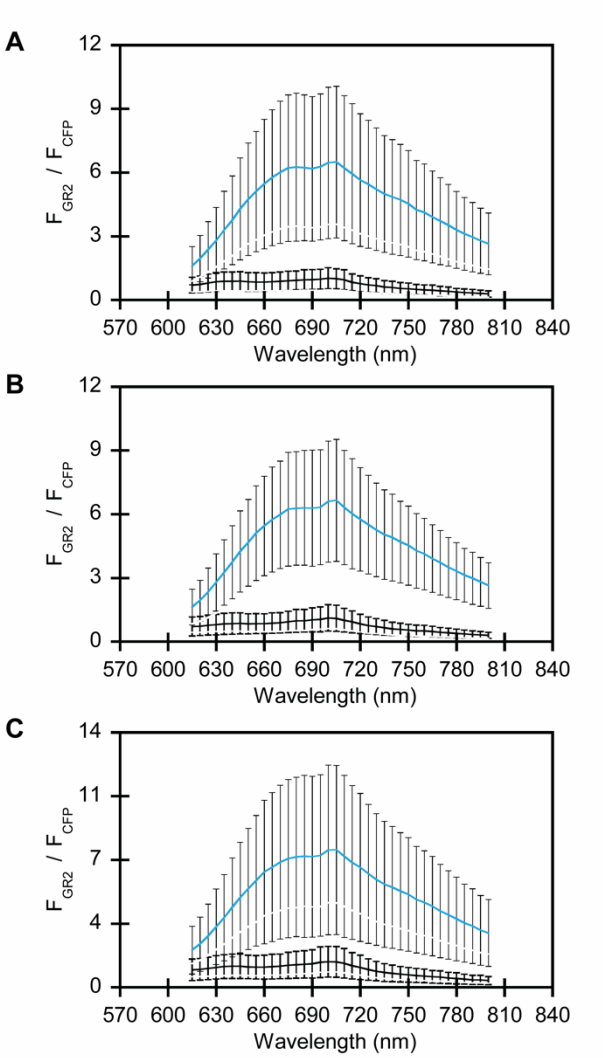

Figure S19. Normalized emission spectra of GR2-CFP in *E. coli* pre-treated with (A) 50 mM sodium acetate buffer at pH 5, (B) 0.3% DMSO, or (C) 30 μM CCCP with 0 (black) and 400 (blue) mM sodium chloride from Figure 4C. For the rhodopsin, excitation was provided at 570 nm, and the emission was collected from 615–800 nm. For CFP, excitation was provided at 390 nm, and the emission was collected from 425–560 nm. The emission spectra of GR2 at each point were normalized by the CFP emission intensity at 485 nm (F_{CFP}). The average of nine biological replicates from three technical replicates with standard error of the mean is shown.

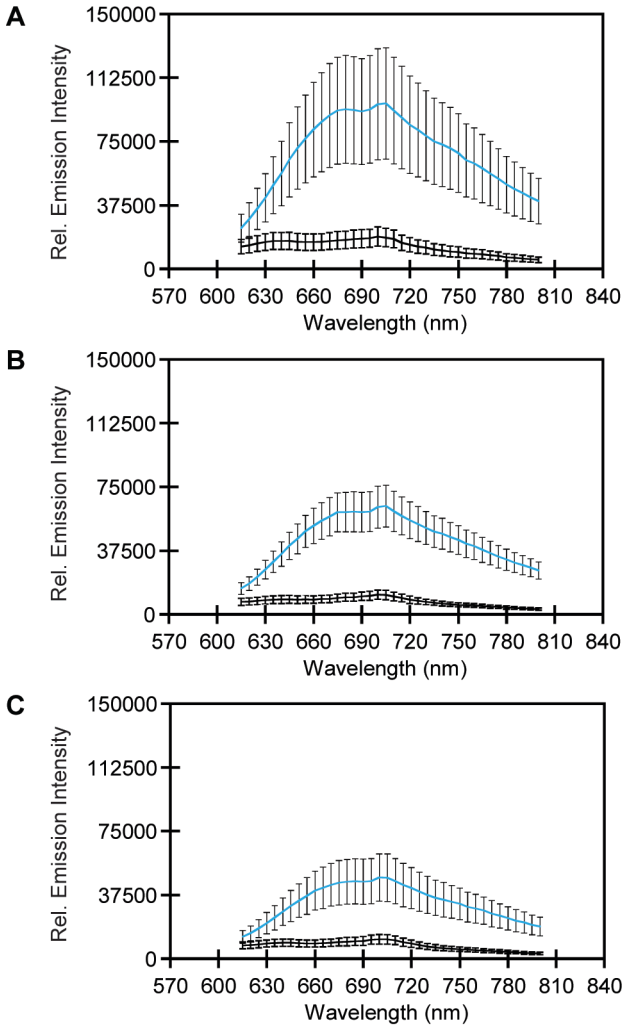

Figure S20. Emission spectra of GR2 for GR2-CFP in *E. coli* pre-treated with (A) 50 mM sodium acetate buffer at pH 5, (B) 0.3% DMSO, or (C) 30 µM CCCP with 0 (black) and 400 (blue) mM sodium chloride from Figure S19.

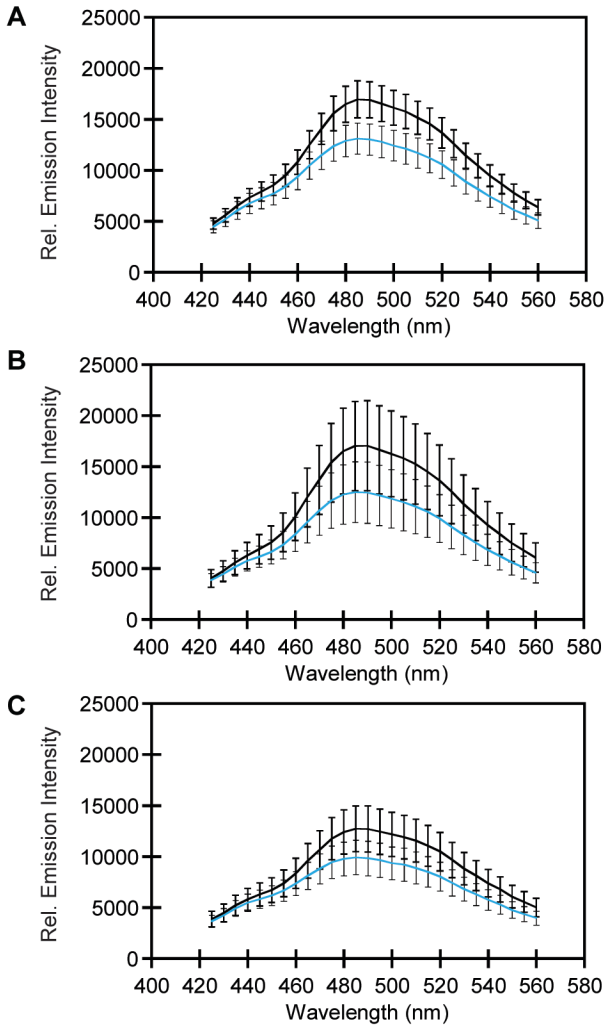

Figure S21. Emission spectra of CFP for GR2-CFP in *E. coli* pre-treated with (A) 50 mM sodium acetate buffer at pH 5, (B) 0.3% DMSO or (C) 30 μM CCCP with 0 (black) and 400 (blue) mM sodium chloride from Figure S19.

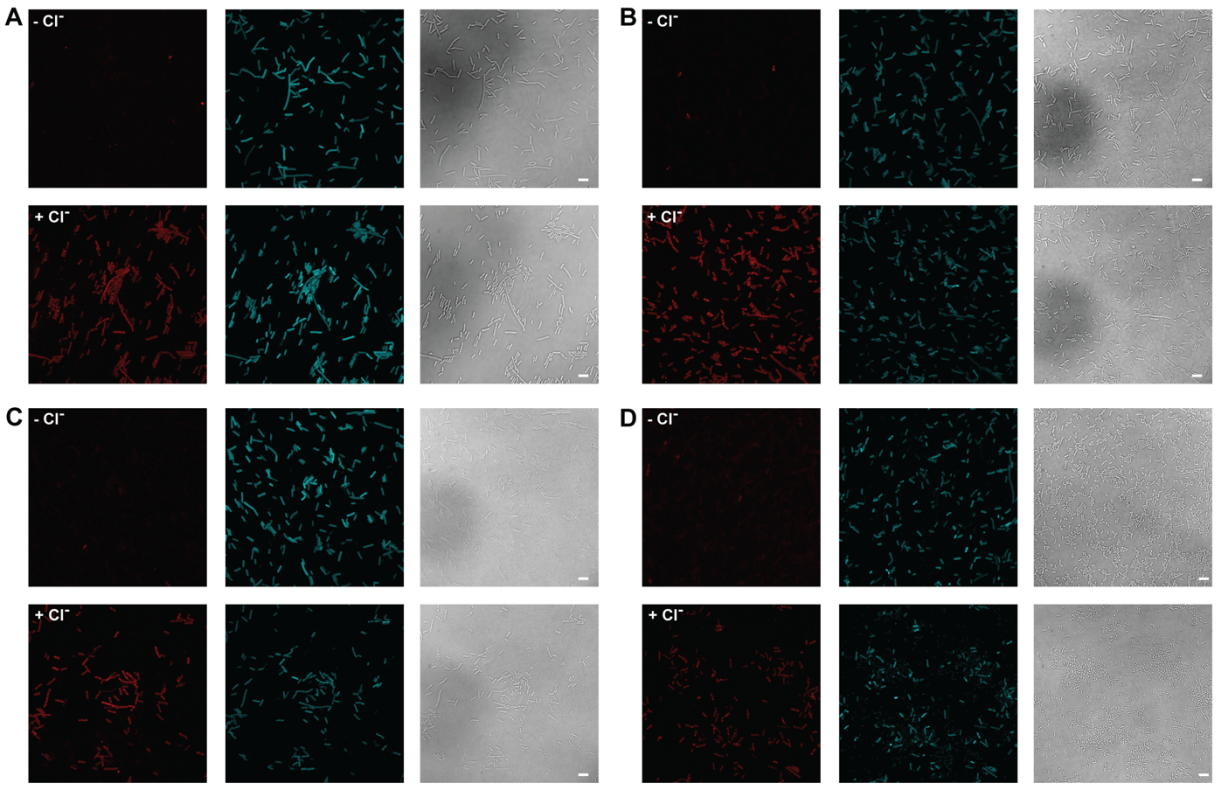

Figure S22. Representative confocal fluorescence microscopy images from four different biological replicates (A–D) are shown for *E. coli* expressing GR2-CFP on agarose pads with 0 mM and 400 mM sodium chloride in 50 mM acetate buffer at pH 5. At least five fields were imaged for each biological replicate. In each panel, the emission from GR2 (red) is on the left, emission from CFP (cyan) is in the middle, and differential interference contrast (DIC) image is on the right (scale bar = 5 μm). Analysis is shown in Figure 5 and Figure S23.

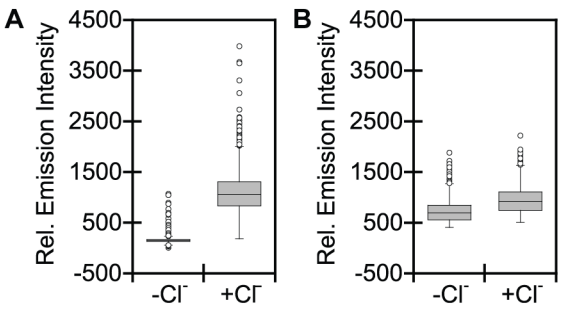

Figure S23. Boxplots show the emission response of (A) GR2 and (B) CFP for each cell analyzed from four biological replicates (n = 3,352 regions of interest (ROIs) for 0 mM sodium chloride; n = 3,062 ROIs for 400 mM sodium chloride). The gray boxes correspond to the lower and upper quartile data with the minimum and maximum values extending below and above the box. The median values are indicated by the black lines in the gray boxes, and outliers are shown as open circles.

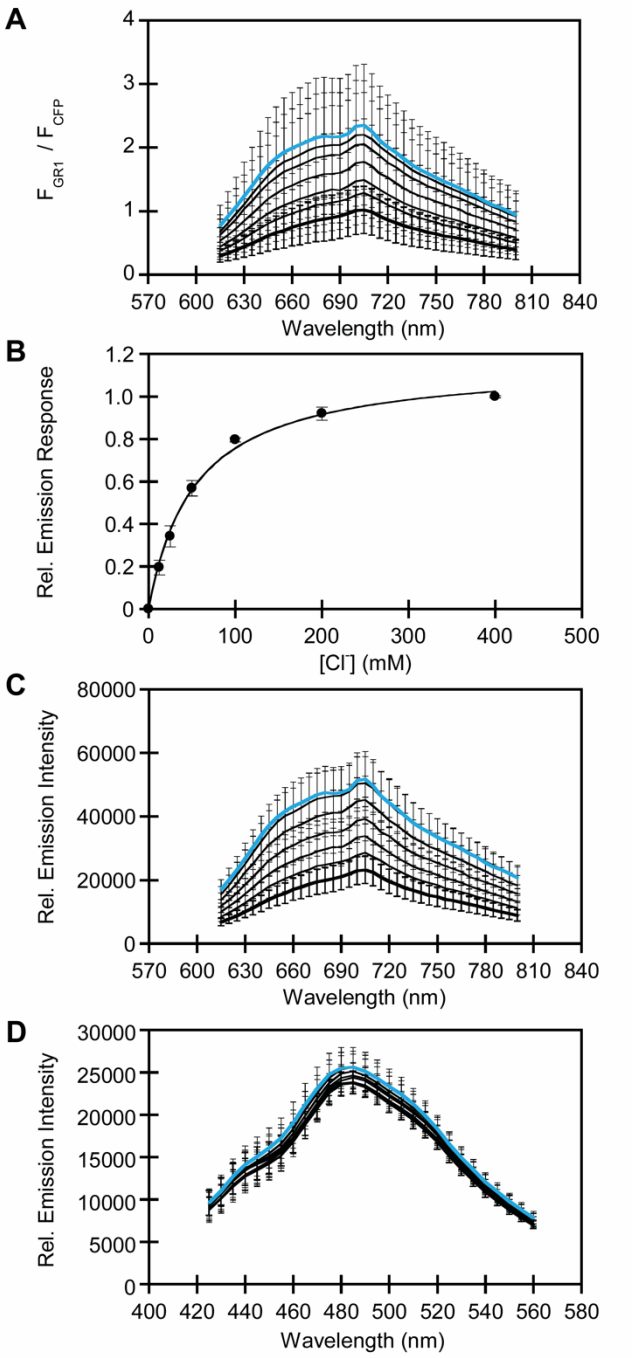

Figure S24. (A) Normalized emission spectra of GR1-CFP in *E. coli* treated with 0 (black and bold), 12.5, 25, 50, 100, 200, and 400 mM (blue) sodium chloride in 50 mM sodium acetate buffer at pH 5. (B) Normalized integrated emission response (F_{GR1}/F_{CFP}) from *A* to determine the K_d. The K_d for Cl^- is 54 ± 9 mM. (C) Emission spectra of GR1 from *A*. (D) Emission spectra of CFP from *A*. For the rhodopsin, excitation was provided at 570 nm, and the emission was collected and integrated from 615–800 nm (F_{GR1}). For CFP, the excitation was provided at 390 nm, and the emission was collected from 425–560 nm (F_{CFP}). The emission spectra of GR1 at each point were normalized by the CFP emission intensity at 485 nm (F_{CFP}). The average of nine biological replicates from three technical replicates with standard error of the mean is shown.

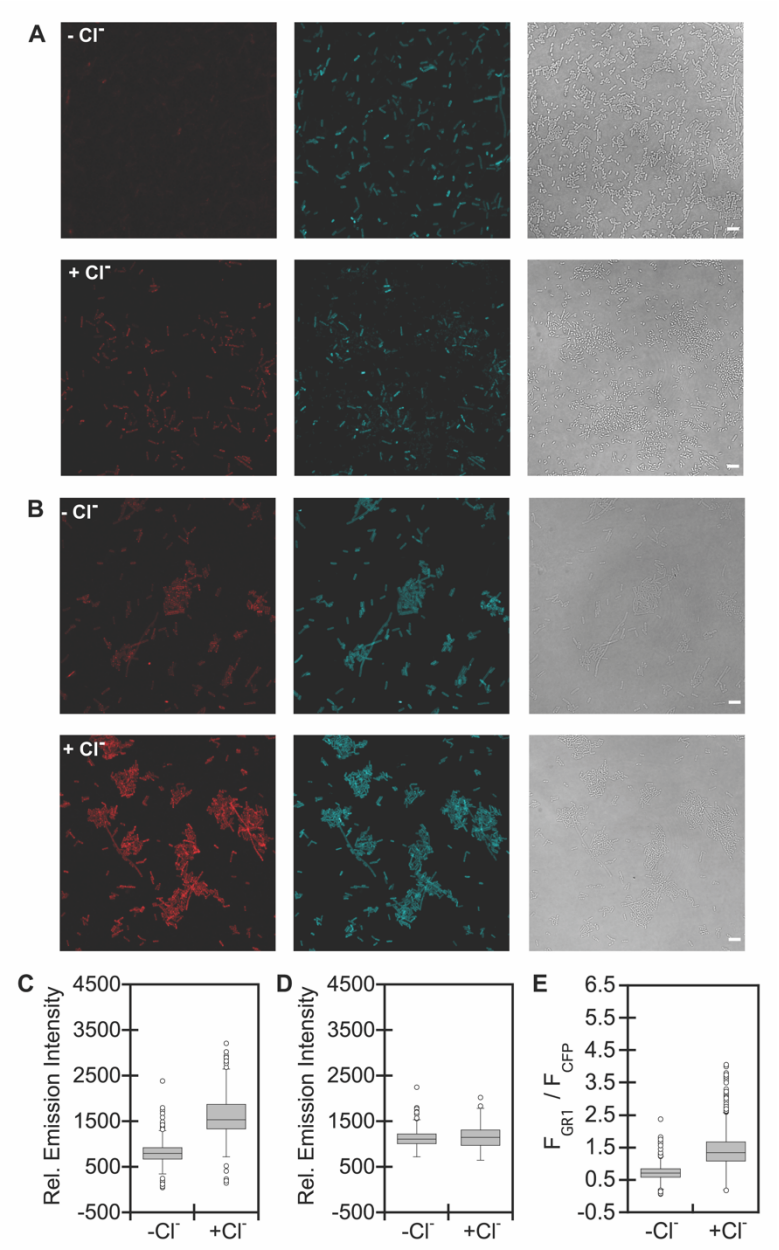

Figure S25. Comparison of (A) panel *D* from Figure S22 for GR2-CFP with (B) representative confocal fluorescence microscopy images of *E. coli* expressing GR1-CFP on agarose pads with 0 mM and 400 mM sodium chloride in 50 mM acetate buffer at pH 5. For GR1-CFP, at least five fields were imaged for one biological replicate. In each panel, the emission from GR1 (red) is on the left, emission from CFP (cyan) is in the middle, and differential interference contrast (DIC) image is on the right (scale bar = 5 μm). Boxplots show the emission response of (C) GR1, (D) CFP, and (E) the normalized emission response (F_{GR1}/F_{CFP}) for each cell analyzed from one biological replicate (n = 437 ROIs for 0 mM sodium chloride; n = 472 ROIs for 400 mM sodium chloride). The gray boxes correspond to the lower and upper quartile data with the minimum and maximum values extending below and above the box. The median values are indicated by the black lines in the gray boxes, and outliers are shown as open circles.

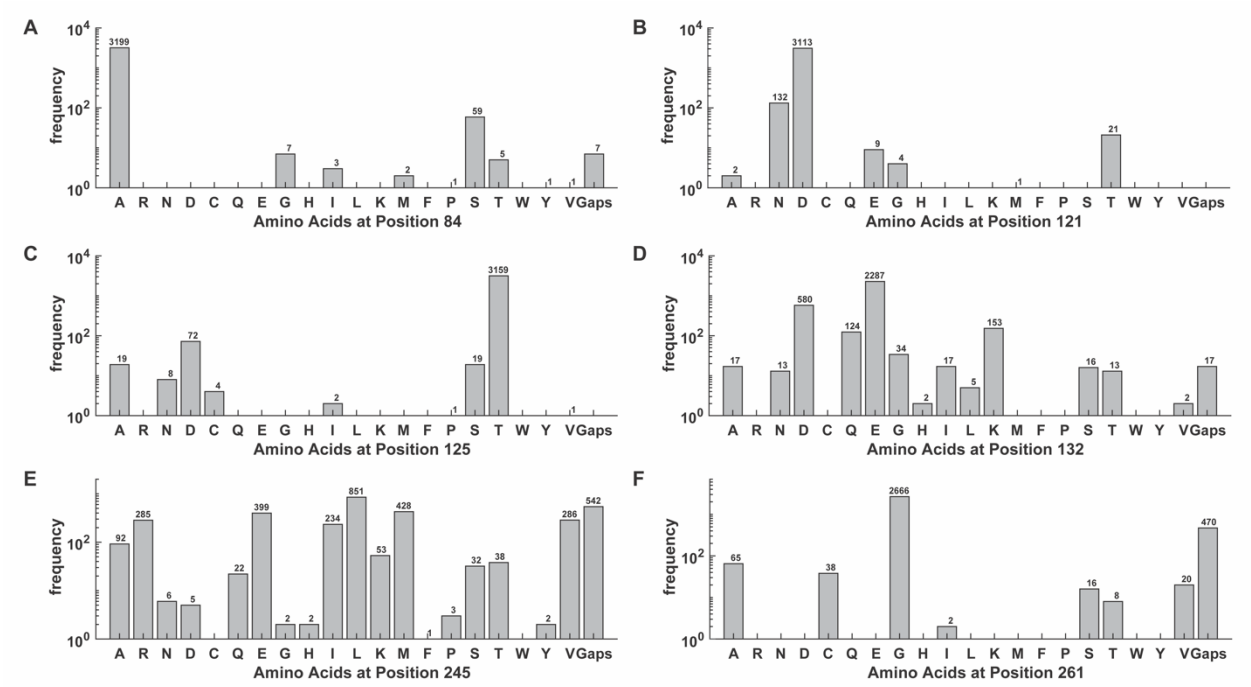

Figure S26. Composition of amino acids at each position along the proton-pumping pathway for all proteins in the rhodopsin family. The amino acid is shown on the x-axis with the number of proteins on the y-axis. If there is no amino acid at a given position in the alignment, it is defined as a gap.

References

1. McWilliam, H.; Li, W.; Uludag, M.; Squizzato, S.; Park, Y. M.; Buso, N.; Cowley, A. P.; Lopez, R. Analysis Tool Web Services from the EMBL-EBI. *Nucleic Acids Res.* **2013**, *41*, W597–600.