Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

HHS Public Access

Author manuscript

ACS Synth Biol. Author manuscript; available in PMC 2023 April 15.

Published in final edited form as:

ACS Synth Biol. 2022 April 15; 11(4): 1627–1638. doi:10.1021/acssynbio.2c00033.

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**[†,*]

[†]Department of Chemistry and Biochemistry, The University of Texas at Dallas, Richardson TX 75080

[‡]Department of Biological Sciences, The University of Texas at Dallas, Richardson TX 75080

[^]Department of Bioengineering, The University of Texas at Dallas, Richardson TX 75080

Abstract

Our understanding of chloride in biology has been accelerated through the application of fluorescent protein-based sensors in living cells. These sensors can be generated and diversified to have a range of properties using laboratory-guided evolution. Recently, we established that the fluorescent proton-pumping rhodopsin *wt*GR from *Gloeobacter violaceus* can be converted into a fluorescent sensor for chloride. To unlock this non-natural function, a single point mutation at the Schiff counterion position (D121V) was introduced into *wt*GR fused to cyan fluorescent protein (CFP) resulting in GR1-CFP. Here, we have integrated coevolutionary analysis with directed evolution to understand how the rhodopsin sequence space can be explored and engineered to improve this starting point. We first show how evolutionary couplings are predictive of functional sites in the rhodopsin family and how a fitness metric based on sequence can be used to quantify known proton-pumping activities of GR-CFP variants. Then, we couple this ability to predict potential functional outcomes with a screening and selection assay in live *Escherichia coli* to reduce the mutational search space of five residues along the proton-pumping pathway in GR1-CFP. This iterative selection process results in GR2-CFP with four additional mutations: E132K, A84K, T125C, and V245I. Finally, bulk and single fluorescence measurements in live *E. coli* reveal that GR2-CFP is a reversible, ratiometric fluorescent sensor for extracellular chloride with an improved dynamic range. We anticipate that our framework will be applicable to other systems,

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

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

Author Contributions

H.C., Q.Z., J.N.T., and S.M.P. contributed equally to this work. F.M. and S.C.D. designed the research. H.C., J.N.T., S.M.P., J.L., and S.C.D. designed and carried out experimental assays and characterization. Q. Z., P.K., and F. M. designed and performed computational modeling and analysis. H.C., Q.Z., J.N.T., S.M.P., J.L., F.M., and S.C.D. analyzed and integrated results. H.C., Q.Z., J.N.T., S.M.P., F.M., and S.C.D. wrote the manuscript with input from all coauthors.

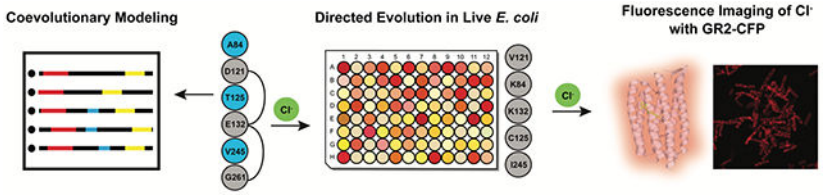
Supporting Information

Multiple sequence alignment, connectivity map, table of the top DI pairs, table and homology models of the distances for the top DI pairs with the select proton-pumping pathway residues, receiver operating characteristic curve, list of primers, nucleotide and amino acid sequences of GR2-CFP, fluorescence excitation and emission spectra, western blot, apparent dissociation constant curve fits, fluorescence microscopy images with analysis, and analysis of the select proton-pumping pathway residues in the rhodopsin family.

The authors declare no competing financial interest.

providing a more efficient methodology to engineer fluorescent protein-based sensors with desired properties.

Graphical Abstract



Keywords

chloride; coevolution; direct coupling analysis; fluorescent biosensor; protein engineering; rhodopsin

Introduction

Chloride is universally a vital ion for life.[1–3] Unicellular and multicellular organisms alike have evolved membrane-bound transporters to mobilize chloride for a wide range of biological functions.[1,3–5] Our ability to visualize and intercept these dynamic processes has been enabled by protein-based fluorescent sensors for chloride.[5–8] To date, these have in large part been derived from the green fluorescent protein (GFP) found in the jellyfish *Aequorea victoria*.[9] Early efforts to engineer GFP led to the discovery of chloride-sensitive variants, YFP, E^{1}GFP, and Topaz.[10–14] These three starting points have been extensively engineered, providing insights into how the sequence of GFP is connected to this unique function. Protein engineering workflows have primarily relied on site-directed/saturation mutagenesis coupled to selection in *Escherichia coli* extracts or cell-free systems.[5,6,8,13–18] As a complementary strategy, we have used bioinformatic and structure-guided approaches to identify chloride-sensitive fluorescent proteins with new properties, including the naturally occurring phiYFP from the jellyfish *Phialidium sp. SL-2003* and the engineered monomer mNeonGreen from the cephalochordate *Branchiostoma lanceolatum*.[19,20]

Expanding from the GFP family (Pfam: PF01353), we recently showcased a *de novo* design strategy to confer chloride-sensitivity into a fluorescent, outward proton-pumping rhodopsin from the cyanobacterium *Gloeobacter violaceus* (*wt*GR).[21] Like other microbial rhodopsins (Pfam: PF01036), *wt*GR is a membrane-spanning protein with a visible light-harvesting Schiff base chromophore (SBC) formed from the condensation of all-*trans*-retinal and a lysine residue (K257 in *wt*GR).[22–25] Outside of a natural context, the SBC can also absorb near-infrared light to access a fluorescent state.[23,26–31] Since the aspartate counterion (D121 in *wt*GR) can stabilize and tune the spectroscopic properties of the SBC, we hypothesized that a chloride ion could do the same.[21] To engineer this non-natural function, we first developed a high-throughput protein engineering workflow in live *E. coli*. A fusion between *wt*GR and cyan fluorescent protein (CFP) was used to generate a chloride-insensitive starting point with a red-shifted, ratiometric fluorescent output. Site-saturation mutagenesis at D121 resulted in the identification of GR1-CFP where mutation to valine not only resulted

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

in a turn-on fluorescence response at pH 5 but also eliminated pumping activity in live *E. coli*.[21]

Even though GR1-CFP does not pump protons nor chloride ions, the residues that define the parent proton pump are still present. In *wt*GR, the proton from the SBC is first donated to the aspartate counterion (D121), which is stabilized through hydrogen bonding interactions with T125 and released on the extracellular side.[25,32,33] Then, the glutamate donor (E132) can re-protonate the SBC on the cytoplasmic side.[25,32,33] This DTE motif is a characteristic sequence that defines a proton-pumping rhodopsin where the glutamate can be substituted with aspartate as found in the well-studied bacteriorhodopsin (BR) from the archaeon *Halobacterium salinarum*.[25,29,34–36] Along the proton-pumping pathway, residues A84 and G261 in *wt*GR are conserved in BR and related proton pumps, but the terminal proton acceptor position is substituted with V245 (Figure S1).[25,35] Since GR1-CFP has the D121V mutation, the context of its interactions with these and other residues that are essential for the parent function has changed. To better understand this and how the sensor capabilities of GR1-CFP could be improved through protein engineering, here, we have integrated coevolutionary analysis with our live cell directed evolution assay.

As proteins evolve to preserve or acquire new functions, their change in amino acid composition leaves a pattern that can be useful to identify critical sites relevant for function. Although amino acid conservation has been of use for many years, patterns of covariation can also provide important clues about epistatic interactions relevant to folding, catalytic activity, and interactions. Statistical modeling of coevolving sites at the amino acid level has become a useful technique to study structures, functions, and interactions of biomolecules. An efficient and accurate methodology to study residue coevolution in protein families is Direct Coupling Analysis (DCA).[37,38] DCA is a global statistical model that reliably captures evolutionary constraints on amino acid sites in multiple sequence alignments (MSAs) of biomolecular families. Coevolutionary information has been successfully used in diverse applications, including the determination of protein 3D structures,[39,40] discovery of protein conformational dynamics,[41] prediction of protein interaction partners and their specificities,[42–44] protein-protein interaction interface inference,[45–49] mutational landscapes for protein–nucleic acid recognition,[50] as well as functional protein design.[51,52] To our knowledge, this approach has yet to be explored for fluorescent protein-based sensors. In this study, we investigate how the use of evolutionary information and the determination of a sequence mutational landscape in the rhodopsin family provides a unique strategy to engineer GR1-CFP with potential extensions to fluorescent protein-based sensors in general. By integrating coevolutionary modeling, mutagenesis, screening, and an iterative selection strategy (Figure 1), we can effectively search the functional mutational space of the residues A84, T125, E132, V245, and G261 in GR1-CFP to generate the improved variant GR2-CFP.

Results

Coevolutionary model of the rhodopsin family

We first used DCA to capture the coevolutionary information from the rhodopsin family sequences with two types of metrics: i) direct information (DI) that quantifies the degree of coupling or coevolution among two residue positions in the MSA of the rhodopsin family

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

and ii) Hamiltonian scores that provide a global measure related to the probability of a given sequence to be a member of the family (Figure 1, step 2 and Figure S2). DI pairs were used to identify the key coevolutionary pairwise correlations between residues in *wt*GR. A Hamiltonian score is calculated as a sum of all the pairwise and single site parameters in the joint probability distribution of a family (see Methods). Since members of the rhodopsin family are light-activated ion pumps and not known to be fluorescent sensors for chloride, the Hamiltonian scores were used to select mutants whose parameters are not necessarily optimized for family likeness.

Identification of evolutionary coupled residues involved in the proton-pumping pathway of wtGR

In this model, we focused on position 121 and five residues along the proton-pumping pathway of *wt*GR described above (Figure 2A). We performed DCA for the protein sequences in the rhodopsin family to quantitatively investigate these residues and validate our modeling in identifying relevant interacting amino acids (see Methods). First, we generated a connectivity network based on the top 300 DI pairs (Figure S2, Supplemental File 1). The specific DI pairs that are related to the selected proton-pumping pathway residues are shown as a network in Figure 2B. We identified strong direct evolutionary interactions among 84, 125, 121, and 132. Some of these residue pairs are close in the GR1 homology model, which is based on the crystal structure of *wt*GR (PDB ID: 6NWD).[24] For instance, the DI pairs 84-121 and 84-125 may form physical contacts relevant for folding and function, as the contact distances are 5.2 Å and 4.7 Å, respectively (Table S1). For the 121-132 pair, the contact distance is 11.3 Å. This is in line with the fact that they are the proton donor/acceptor pair for the SBC. In addition to the coevolutionary interactions to the selected proton-pumping pathway residues, we found other residues that are highly coupled, suggesting additional relevant interactions that are essential for the natural function of *wt*GR. For example, residues 132 and 77 are key residues that allow for inward directed diffusion of protons under certain conditions.[33] The strong interaction between residues 121 and 87 is consistent with how these residues form a salt-bridge to functionally position residues for the light-induced proton transfer.[53,54] Interestingly, by labeling these residues in the structural model, we observed that they form a mesh that surrounds the proton-pumping pathway (Figure S3). The other two residues, 245 and 261, do not directly interact with the other selected proton-pumping pathway residues but are connected through intermediate residues. For example, both residues 261 and 84 are coupled to residue 129. This interaction is also reflected in the distance measurements. The 84-261 pair is far apart (~12 Å), while the 84-129 and 129-261 pairs are closer (8.4 Å and 6.9 Å, respectively). The latter pairs could potentially form physical contacts required for folding and function (Table S1). Other residues from our model such as 87, 88, 253, and 257 are known to be near the SBC and essential for the native proton release complex.[25] This analysis supports the idea that i) direct couplings do capture physical interactions relevant to the parent proton-pumping function and ii) evolutionary couplings can in principle be used to study changes that can affect functional phenotypes.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Coevolutionary modelling is predictive of proton-pumping activities of wtGR-CFP and variants

To test the accuracy of the coevolutionary model in predicting functional outcomes in the rhodopsin family, we first studied the effect of mutations in previously reported variants of *wt*GR-CFP.[23] For each variant, we correlated the mutations and proton-pumping activities (maintained, reduced, or eliminated) with the calculated Hamiltonian score. These results are illustrated in the heatmap of Figure 2C. With respect to the wildtype, we observed a gradual decrease in the relative Hamiltonian scores for variants that have no proton-pumping activity. With this information at hand, we created a Hamiltonian-based classifier that predicts if a variant will eliminate proton-pumping activity. We found that such classification task can be done with high accuracy (Area Under the Curve, AUC = 0.90; Figure S4) showcasing the predictive power of coevolutionary information for functional features in the rhodopsin family.

Iterative site-saturation mutagenesis of GR1-CFP guided by the Hamiltonian score

Having established the predictive power of the DI and Hamiltonian metrics, we utilized this methodology to guide the mutagenesis of GR1-CFP. To achieve this, we calculated i) the Hamiltonian score for every possible variant at position 121 and five residues selected along the proton-pumping pathway and ii) the mean of the relative Hamiltonian score changes for all the mutations at the same position (see Methods, Figure 3). Position 121 showed the most negative relative score change among all the six residues. This suggested that to increase the fluorescence response to chloride, we should select positions with more negative relative score changes for mutagenesis, thus deviating from the parent function.

In our experimental approach, site-saturation mutagenesis was used to generate all possible amino acid substitutions at each position with the largest Hamiltonian score change (Figure 3, Table S2, and Figure S5–S8). We first fixed position 121 as valine, and selected position 132 for mutagenesis. This residue had the largest relative score changes and co-evolved with many proton-pumping related residues (Figure 2B, Figure S2). The library was screened in live *E. coli* and variants with an improved turn-on, ratiometric response to 400 mM sodium chloride versus GR1-CFP ($\Delta F_{GR}/F_{CFP} \approx 2.0$ at pH 5) were rescreened prior to Sanger sequencing to identify the mutations. This initial screen generated GR1-E132K-CFP ($\Delta F_{GR}/F_{CFP} \approx 2.2$ at pH 5), which served as the new starting point. Next, position 125 was selected, but no improved variants were experimentally identified. We proceeded to position 84 that had the next largest score change and identified GR1-E132K-A84K-CFP ($\Delta F_{GR}/F_{CFP} \approx 3.9$ at pH 5). Continuing this cycle of modeling selection, mutagenesis, and screening, we identified the final variant with four mutations GR1-E132K-A84K-T125C-V245I-CFP or GR2-CFP ($\Delta F_{GR}/F_{CFP} \approx 7.5$ at pH 5). No improved variants were identified at position 261.

Characterization of GR2-CFP in live E. coli cells using bulk and single cell fluorescence measurements

To understand how the accumulated mutations altered sensing properties, bulk fluorescence measurements were directly carried out in live *E. coli* cells expressing GR2-CFP (Figure S9). In these experiments, the emission of GR2 was normalized to the emission of CFP, thus providing a ratiometric output for exogenous chloride supplementation. At pH 5, excitation

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

of apo GR2 at 570 nm results in no fluorescence signal above the background of the cells alone, but excitation of CFP at 390 nm results in an emission maximum centered at 485 nm, confirming the expression of GR2-CFP (Figure 4A, Figure S10–S13). However, in the presence of up to 400 mM sodium chloride, GR2-CFP has a broad emission centered at 705 nm with a $\Delta F_{GR}/F_{CFP} = 7.5 \pm 1.5$. This turn-on response arises from GR2 as the fluorescence intensity of CFP remains unchanged. Fitting of the titration data to a single site binding model results in an apparent dissociation constant (K_d) for chloride binding to GR2-CFP of 53 ± 14 mM. We next evaluated if GR2-CFP could operate at a more physiologically relevant pH (Figure S14 and S15). In the absence of chloride, GR2-CFP also has no measurable fluorescence at pH 6 and 7. However, surprisingly, the fluorescence ratio increases with up to 400 mM sodium chloride at pH 6 ($\Delta F_{GR}/F_{CFP} = 3.8 \pm 0.4$, $K_d = 731 \pm 31$ mM). No fluorescence response is observed at pH 7 (Figure S15).

Given the larger dynamic range, we further characterized GR2-CFP at pH 5. The chloride-dependent response can be reversed by washing cells with sodium acetate buffer (63% turn-off) or an equimolar amount sodium gluconate (65% turn-off) (Figure 4B, Figure S16–S18). This data suggests that GR2-CFP can dynamically sense changes in the extracellular and periplasmic chloride pool (Figure 4). Moreover, uncoupling of the cell membrane potential with the protonophore carbonyl cyanide 3-chlorophenylhydrazone (CCCP) does not attenuate the fluorescence response ($\Delta F_{GR}/F_{CFP} = 5.3 \pm 1.6$) in comparison to cells treated with a DMSO vehicle control ($\Delta F_{GR}/F_{CFP} = 6.0 \pm 1.2$) (Figure 4C, Figure S19–S21), thus confirming that GR2-CFP does not sense membrane potential. Encouraged by these data, we used confocal fluorescence microscopy to validate GR2-CFP at the single cell level. Immobilization of *E. coli* on agarose pads at pH 5 shows that GR2 is non-fluorescent even though CFP is expressed in the cytosol (Figure 5, Figure S22 and S23). Upon supplementation with 400 mM sodium chloride, the GR2 signal is localized to the plasma membrane, and the fluorescence ratio increases by *ca.* 5.4-fold.

Discussion

In this work, we show how the integration of evolutionary signals learned from extant sequences of the rhodopsin family can be used to guide the evolution of the engineered rhodopsin chloride sensor GR1-CFP into the improved variant GR2-CFP. Our integrated approach of computation and experiment allows a significant reduction of mutational search space along five proton-pumping pathway residues to accelerate the discovery process (Figure 1). To put this in perspective, a blind search of the possible variants would yield 19^5 ($\approx 2\times10^6$) combinations to be screened. Our modeling approach first suggested residue positions with the largest relative score changes to be mutated in GR1-CFP. Using this, experimental site-saturation mutagenesis coupled with a high-throughput *in vivo* screening and selection enabled the identification of the amino acid substitutions at each position. The whole process required the analysis of less than 95 (19×5) theoretical variants and 480 (96×5) experimental variants, which is significantly lower than sampling the complete sequence space (19^5) that would be needed to engineer GR2-CFP.

Using bulk and single cell fluorescence assays, we show that GR2-CFP is a reversible, ratiometric fluorescent sensor for chloride. In comparison to GR1-CFP, GR2-CFP has a

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

little to no background fluorescence from the rhodopsin (Figure 3, Figure 4A, Figure 5A, Figure S24 and S25).[21] This in combination with the larger dynamic range, translates into a *ca.* 3.3-fold higher ratio change for GR2-CFP ($\Delta F_{GR2}/F_{CFP}$ = 7.5 ± 1.5) versus GR1-CFP ($\Delta F_{GR1}/F_{CFP}$ = 2.3 ± 0.1) (Figure 4A, Figure S24).[21] However, the K_d for chloride binding to GR2-CFP (53 ± 14 mM) is comparable to that of GR1-CFP (54 ± 9 mM) (Figure S10 and S24).[21] Interestingly, even though we did not select for an expanded operational pH, the mutations along the proton-pumping pathway did result in measurable chloride sensitivity at pH 6 (Figure S14).

With respect to *wt*GR-CFP, GR2-CFP has five mutations: D121V, E132K, A84K, T125C, and V245I. In its native function, the D121 counterion position accepts a proton from the E132 donor and hydrogen bonds with T125, and residues A84, V245, and G261 line the proton-pumping pathway.[25] As we previously reported, the non-natural D121V mutation in GR1-CFP eliminates proton-pumping activity and confers the fluorescence response to chloride at pH 5.[21,23] In GR2-CFP, the enrichment of positively charged lysine residues at positions 132 and 84 provides a *ca.* 2-fold higher fluorescence ratio change relative to GR1-CFP (Figure 3, Figure S7, S8). Mutation from a polar threonine to a larger, non-polar cysteine at position 125 in combination with the isoelectronic mutation from a valine to a larger isoleucine at position 245 increases the dynamic range from GR1-E132/A84K-CFP by an additional *ca.* 1.9-fold (Figure 3, Figure S7 and S8). It is interesting to note that the A84K and T125C mutations influence the non-natural chloride sensing function to a larger extent, but the order of mutagenesis was key to identify these improvements. Moreover, the effect of these mutations is perhaps unsurprising since both residues are adjacent to the putative chloride binding pocket and SBC (Figure 2A).[24] Of the four mutations introduced into GR1-CFP, A84K is a non-natural substitution that has not been sampled in the natural evolution of the rhodopsin family (Figure S26).

Here, we only focused on a subset of residues along the proton-pumping pathway as an initial approach. However, examination of the top DI pairs reveals highly coupled residue pairs that merit further exploration to expand biosensor properties, including the fluorescence response, operational pH, and tuning of the excitation/emission wavelengths (Figure 2B). Thus, we are not necessarily limited to the selected residues, and in principle, additional relevant sites could be elucidated without prior experimental knowledge. Efforts to expand our search space in GR2-CFP and other rhodopsins are currently in progress, with a particular focus on mutations that could provide sensitivity to chloride in a physiological pH range. Accessing such a property would allow us to realize the potential of rhodopsin-based chloride sensors to monitor the recycling of chloride at the plasma and organelle membranes in mammalian cells. Beyond the rhodopsin family, we will extend our collaborative approach to other chromophore-containing fluorescent proteins to identify, evolve, and design chloride-sensitive fluorescent proteins with new properties. With the knowledge gained, we will not only shape our understanding of the sequence level determinants that govern anion recognition in biology but will also generate an expanded set of sensors for cellular applications. To close, the use of evolutionary signals to engineer proteins with new properties has broad utility as we can readily extract direct knowledge from evolutionary processes that have taken long periods of time to fine tune the desired properties we see in Nature. We envision that learning and utilizing these signals as a

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

guiding principle presents an exciting opportunity to intersect with existing and emerging protein engineering methods for fluorescent sensors and beyond.

Methods

Pre-processing of the rhodopsin family multiple sequence alignment (MSA).

The amino acid sequence of the wild-type proton pumping rhodopsin from *Gloeobacter violaceus* (*wt*GR) (UniProt ID: Q7NP59, residues 45–270) was used to search for homologous proteins in the rhodopsin family in UniProt Knowledgebase using the HMMER webserver v2.41.1.[55,56] The threshold of expectation value (E) was set to 0.01 to filter and obtain 6,906 protein sequences that are like *wt*GR. Next, we excluded protein sequences containing consecutive gaps longer than 20% of the protein length (> 45 consecutive gaps) to reduce the noise from incomplete coverage. The amino acid composition at each proton-pumping pathway position in the sequence family was plotted (Figure S26). Then, re-weighted sequences with greater than 80% identity were filtered as previously described, leaving 879 effective sequences.[38]

Direct coupling analysis of rhodopsin protein family.

Direct coupling analysis (DCA) was performed using the pre-processed sequences from the rhodopsin protein family to infer a joint probability distribution that satisfies the statistical observations. Two types of parameters are estimated for this distribution: pairwise couplings $e_{ij}(x_i, x_j)$ and local biases $h_i(x_i)$, where i, j indices refer to positions along the sequence and x refers to the amino acid at a specific position. Direct information (DI) values were computed from this distribution to quantify how the two sites are directly coupled.[41,57] The top 300 ranked DI pairs were selected (Figure S2, Supplemental File S1). DI pairs related to the proton-pumping residues were selected and plotted as a connectivity map (Figure 2B). For a given sequence L, the Hamiltonian score is defined as a sum of all the pairwise couplings $e_{ij}(x_i, x_j)$ and local biases $h_i(x_i)$: [50]

$$H(L) = -\sum_{i<j} e_{ij}(x_i, x_j) - \sum_i h_i(x_i)$$

The fitness (functional change with respect to the parent) of mutants σ_a was predicted by the relative change of Hamiltonian score compared to the wild-type sequence (*wt*) as $\Delta H(\sigma_a)$:

$$\Delta H(\sigma_a) = \frac{H(\sigma_a) - H(wt)}{H(wt)}$$

Relative changes of the average Hamiltonian.

To guide the iterative site-saturation mutagenesis, relative changes of the average Hamiltonian were calculated to evaluate the effect of changes for each proton-pumping pathway position (σ) involved.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

$$\Delta H_n(\sigma) = \frac{\frac{\sum_{a \in A} H_n(\sigma_a)}{19} - H_{n-1}(p)}{H_{n-1}(p)}$$

Where n is the number of mutation steps, a is the set of all the possible amino acid changes except the wild-type or parent amino acid. The variant with the most negative relative Hamiltonian score changes was selected as the parent ($H_{n-1}(p)$) for the next step. We used *wt*GR as the parent in the first step for the single mutations.

Homology model of GR1.

The homology model of GR1 in Figure 2 and Figure S2 was generated based on the crystal structure of *wt*GR (PDB ID: 6NWD) using the MODELLER software.[58]

General.

All supplies and chemicals were purchased from Bio-Rad Laboratories, Sigma-Aldrich, Thermo Fisher Scientific, or VWR, unless otherwise stated.

Construction, expression, and screening of site-saturation mutagenesis (SSM) libraries.

The pET-21a(+)-based plasmid encoding GR1-CFP was generated in our previous study.[21] Three forward primers and one reverse primer were synthesized using the 22c-trick (Sigma-Aldrich) to generate each SSM library as previously described (Table S2).[21,59] *E. cloni* EXPRESS BL21 (DE3) cells (Lucigen) were transformed with each library and expressed in a 96-well deep well format.[21] General methods for liquid handling and fluorescence plate reader screening at pH 5, 6, and 7 were adapted from the methods used to generate GR1-CFP.[21] Variants with an increased turn-on fluorescence response with respect to the parent were re-screened at least in duplicate.[21] If necessary, excitation and emission settings were altered for each experiment for optimal signal-to-noise. For the rescreens, variants were first selected based on an improved fluorescence response at pH 5, then the turn-on fluorescence response at pH 6 or 7 was used for selection. On the basis of this analysis, the plasmid DNA for each improved variant was isolated and sequenced as previously described.[21]

Validation of improved variants.

To compare the variants from each library, *E. cloni* EXPRESS BL21 (DE3) cells were transformed via electroporation and plated onto LB-Miller agar plates containing 100 μg/mL ampicillin (AMP). Three different colonies of each variant were tested for each of the three technical replicates. Single colonies were picked into 5 mL of LB media (10 g/L of tryptone and 5 g/L of yeast extract) containing 100 μg/mL AMP in 14 mL culture tubes and incubated overnight at 37 °C with shaking at 250 rpm. The next day, 150 μL of each overnight culture was diluted into 3 mL of fresh LB media containing 100 μg/mL AMP and incubated for 2 h and 15 min at 30 °C with shaking at 230 rpm, followed by addition of 15 μL of 100 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) in water and 15 μL of 2 mM all-*trans*-retinal (ATR) in ethanol to final concentrations of 500 μM and 10 μM, respectively. After 4 h, the cells were collected by centrifugation at 2,500*g* (5810 R, Eppendorf) for 5

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

min at room temperature. The resulting cell pellet was washed, resuspended, centrifuged, and resuspended in 500 μL of 50 mM sodium acetate buffer at pH 5. A 25 μL aliquot of the cell suspension was transferred to a well 96-well microtiter plate containing 175 μL 50 mM sodium acetate buffer at pH 5 or 175 μL 50 mM sodium acetate buffer at pH 5 with 457 mM sodium chloride. All spectra were collected on a Tecan plate reader. For the rhodopsins, excitation was set to 570 nm (20 nm bandwidth), and the emission was collected from 630–760 nm (10 nm bandwidth, 10 nm step size, 100 gain, 25 flashes). Based on these spectra, the emission for the excitation scan was set to 710 nm (20 nm bandwidth), and the excitation was collected from 490–630 nm (20 nm bandwidth, 100 gain, 10 nm step size, 25 flashes). For CFP, excitation was set to 390 nm (10 nm bandwidth), and the emission was collected from 435–520 nm (20 nm bandwidth, 5 nm step size, 55 gain, 10 flashes). For all rhodopsin emission spectra, the emission intensity at each wavelength was normalized to the emission intensity of CFP at 485 nm. The normalized emission spectra were then averaged across all 12 biological replicates. These spectra are reported with the standard error of the mean (Figure S7, S8). To determine the turn-on fluorescence response for each sample, the emission spectra of GR2 and CFP were first integrated in Kaleidagraph v4.5 (Synergy Software) from 630–760 nm and 435–520 nm, respectively. The integrated emission intensity from the rhodopsin (F_{GR}) was then normalized to the integrated emission intensity from CFP (F_{CFP}). The turn-on emission response was calculated by dividing the F_{GR}/F_{CFP} ratios for 400 mM sodium chloride by 0 mM sodium chloride. The average of all twelve biological replicates across three technical replicates with standard error of the mean is reported (Figure 3).

SDS-PAGE and Western blot analysis of GR2-CFP.

E. cloni EXPRESS BL21 (DE3) cells were transformed with the plasmid DNA encoding for GR2-CFP via electroporation and plated onto LB-Miller agar plates containing 100 μg/mL AMP. Single colonies were picked into 5 mL of LB media containing 100 μg/mL AMP in 14 mL culture tubes and incubated overnight at 37 °C with shaking at 250 rpm. The following morning, the overnight cultures were diluted 1:20 into 25 mL of LB media containing 100 μg/mL AMP in 125 mL baffled flasks. The cultures then incubated for 2 h at 30 °C with shaking at 230 rpm. Following this incubation period, protein expression was induced with 122 μL of 100 mM IPTG and 23 μL of 12 mM ATR to final concentrations of 500 μM and 10 μM, respectively. Non-induced cultures served as negative controls. The cultures incubated for 4 h at 30 °C with shaking at 230 rpm. After the final incubation period, a 6 mL portion of each sample was collected by centrifugation at 2,500*g* (5810 R, Eppendorf) for 5 min at 4 °C. Following this, the cells were resuspended in 500 μL of lysis buffer containing 20 mM Tris buffer at pH 7.5 with 200 mM sodium chloride, 5 mM magnesium chloride, 30 μg/mL deoxyribonuclease I, and two cOmplete protease inhibitor tablets. The resuspended cell pellets were lysed with a pellet mixer equipped with a cordless motor in 30 s on-off intervals for 3 min on ice. Following homogenization, the lysates were clarified via centrifugation at 5,000*g* (5424 R, Eppendorf) for 3 min at 4 °C.

The total protein concentration of each sample was determined using a commercial bicinchoninic acid protein assay. For each sample, 18 μL of 0.5 mg/mL protein was combined with 6 μL of 4X Laemmli sample buffer supplemented with 10% beta-

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

mercaptoethanol and 50 mM Tris buffer at pH 7.5 for a total volume of 24 μL. The protein-buffer mixtures were heated at 37 °C for 10 min, and then 15 μL of each sample were loaded onto a 12% Mini-PROTEAN TGX acrylamide gel alongside 5 μL of a prestained protein ladder.

The gel was run in 1X Tris-Glycine-SDS running buffer containing 25 mM Tris at pH 8.3 with 192 mM glycine and 0.1% SDS at 150 V for 1 h. The separated proteins were then transferred to a methanol-activated polyvinylidene difluoride (PVDF) membrane utilizing the Bio-Rad Trans-Blot Turbo Transfer system and kit with the 7 min mixed molecular weight transfer setting. All subsequent incubation and wash steps were performed with orbital shaking. Following the protein transfer, the membrane was blocked with 5% milk in 1X Tris-Buffered Saline containing 0.1% Tween20 (TBST) for 1 h at room temperature. After blocking, the membrane was incubated with 3 μL of 1 mg/mL 6x-His tag monoclonal antibody in 5% milk in TBST at 4 °C. The next day, the primary antibody was removed, and the membrane was washed with TBST (3 x 5 min) and incubated with 15 μL of 2 mg/mL goat anti-mouse horse radish peroxidase conjugated secondary antibody in 5% milk in TBST at 4 °C for 2 h. Following this, the secondary antibody solution was removed, and the membrane was washed with TBST (3 x 5 min). For visualization, the membrane incubated in 2 mL of the Immobilon Classico Western HRP substrate for 1 min and imaged on a Bio-Rad ChemiDoc imaging station with the colorimetric setting for the ladder and the chemiluminescent setting to visualize the protein. The resulting images were overlayed and analyzed with the Bio-Rad ImageLab software (Figure S9).

Characterization of GR2-CFP in *E. coli* using bulk fluorescence measurements.

E. cloni EXPRESS BL21 (DE3) cells were freshly transformed with plasmid DNA encoding for GR2-CFP via electroporation and plated onto LB-Miller agar plates containing 100 μg/mL AMP. For each biological replicate, a single colony was picked into 5 mL of LB media with 100 μg/mL AMP in a 14 mL culture tube and grown overnight at 37 °C with shaking at 250 rpm. The following day, the overnight culture was diluted 1:20 in 25 mL of LB media containing 100 μg/mL AMP in a 125 mL baffled flask and incubated for 2 h at 30 °C with shaking at 230 rpm. Following this, protein expression was induced with 122 μL of 100 mM IPTG and 23 μL of 12 mM ATR in ethanol to final concentrations of 500 μM and 10 μM, respectively. The induced culture was incubated for 4 h at 30 °C with shaking at 230 rpm, and then 6 mL of the culture was collected by centrifugation at 2,500*g* (5810 R, Eppendorf) for 5 min at room temperature.

Chloride titrations with GR2-CFP expressing cells were carried out to determine the dynamic ranges and apparent dissociation constants (K_d) for chloride. The resulting cell pellet was resuspended in 500 μL of 50 mM sodium acetate buffer at pH 5 or 50 mM sodium phosphate buffer at pH 6 and pH 7, and then re-collected by centrifugation and resuspended in 500 μL of the corresponding buffer. In a 96-well microtiter plate, a 25 μL portion of the cell suspension was further diluted with 175 μL of 50 mM sodium acetate buffer at pH 5 or sodium phosphate buffer at pH 6 and pH 7 and the corresponding buffer containing 14.3, 28.5, 57, 114.3, 228.5, 457, or 685.7 sodium chloride to final concentrations of 0, 12.5, 25, 50, 100, 200, 400, and 600 mM sodium chloride. All spectra were collected on

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

a Tecan plate reader. For the rhodopsin, excitation was set to 570 nm (10 nm bandwidth), and the emission was collected from 615–800 nm (10 nm bandwidth, 5 nm step size, 150 gain, 25 flashes). For CFP, excitation was set to 390 nm (10 nm bandwidth), and the emission was collected from 425–560 nm (20 nm bandwidth, 5 nm step size, 100 gain, 25 flashes). For all rhodopsin spectra, the emission intensity at each wavelength was normalized to the emission intensity of CFP at 485 nm. The normalized emission spectra were then averaged across all nine biological replicates. These averaged spectra are reported with the standard error of the mean (Figure 4A, Figure S10–S15). To determine the turn-on responses and K_ds, the emission spectra of GR2 and CFP for each sample were first integrated in Kaleidagraph v4.5 (Synergy Software) from 615–800 nm and 425–560 nm, respectively. The integrated emission intensity from the rhodopsin (F_{GR2}) was then normalized to the integrated emission intensity from CFP (F_{CFP}). The turn-on emission responses were calculated by dividing the F_{GR2}/F_{CFP} ratios for 400 mM sodium chloride by 0 mM sodium chloride. To determine the K_ds, the emission response of GR2-CFP was calculated with the following equation where F_{obs} is the observed F_{GR2}/F_{CFP} at a given concentration of sodium chloride, and F_{min} and F_{max} correspond to F_{GR2}/F_{CFP} at 0 mM and 600 mM sodium chloride, respectively: $F=(F_{obs}-F_{min})/(F_{max}-F_{min})$. These emission responses were plotted versus the concentration of sodium chloride and fitted to the following equation to determine the K_d where F is the observed F_{GR2}/F_{CFP} at a given concentration of sodium chloride and B_{max} is the fluorescence emission at saturation: $F=([Cl^-] \times B_{max})/(K_d + [Cl^-])$. The average of all nine biological replicates from three technical replicates with standard error of the mean is reported (Figure S10, Figure S14, Figure S15). This procedure was repeated in parallel on the same days with GR1-CFP as a positive control for comparison (Figure S24).

Control experiments were also performed with the following expression conditions. One culture tube was expressed using the conditions described above. The second culture tube was treated with IPTG and no ATR, the third culture tube was treated only with ATR and no IPTG, and the fourth culture tube was not treated with IPTG nor ATR. Each culture tube was then incubated at 30 °C with shaking at 230 rpm for 4 h. Following protein expression as described above, the cells were then harvested and analyzed on a plate reader using the conditions described above. For all rhodopsin spectra, the emission intensity at each wavelength was normalized to the emission intensity of CFP at 485 nm. The normalized spectra of all nine biological replicates from three technical replicates with standard error of the mean is reported (Figure S11–S13).

To test the reversibility of GR2-CFP, cultures of *E. cloni* EXPRESS BL21 (DE3) cells expressing GR2-CFP were prepared as described above. The resulting cell pellets were resuspended in 400 μL of 50 mM sodium acetate buffer at pH 5. In 1.5-mL centrifuge tubes, a 75 μL portion of each cell suspension was further diluted with 525 μL of 50 mM sodium acetate buffer at pH 5 (control) or buffer containing 457 mM sodium chloride to final concentrations of 0 mM or 400 mM sodium chloride. After mixing and incubating at room temperature for 10 min, a 200 μL portion of each cell suspension was transferred to a 96-well microtiter plate, and spectra were collected as described above. The remaining cell suspensions were collected by centrifugation at 500*g* (5424 R, Eppendorf) for 10 min at room temperature and resuspended as follows: the control cell pellet was resuspended in 400 μL of 50 mM sodium acetate buffer at pH 5, and the remaining cell pellets were

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

resuspended in 400 μL of buffer, buffer containing 400 mM sodium chloride, or 400 mM sodium gluconate. After mixing and incubating at room temperature for 10 min, a 200 μL portion of each cell suspension was transferred to a 96-well microtiter plate, and spectra were collected and analyzed as described above. The average of all nine biological replicates across three technical replicates with standard error of the mean is reported (Figure 4B, Figure S16–S18).

To validate that the turn-on response of GR2-CFP was not dependent on membrane potential, *E. cloni* EXPRESS BL21 (DE3) cells expressing GR2-CFP were treated with CCCP. Cultures of *E. cloni* expressing GR2-CFP were prepared as described above. The cells were collected by centrifugation at 2,500*g* (5810 R, Eppendorf) for 5 min at room temperature, and the resulting cell pellets were resuspended in 80 μL of 50 mM sodium acetate buffer at pH 5. In a 96-well microtiter plate, a 10 μL portion of each cell suspension was further diluted with 175 μL of 50 mM sodium acetate buffer at pH 5 or buffer containing 457 mM sodium chloride to final concentrations of 0 mM or 400 mM sodium chloride. Each well was treated with 15 μL of 50 mM sodium acetate buffer at pH 5, 4% DMSO (v/v), or 400 μM CCCP in DMSO in 50 mM sodium acetate buffer at pH 5 to a final concentration of 0.3% DMSO (v/v) or 30 μM CCCP. The microtiter plate was incubated in the plate reader at 30 °C for 30 min with shaking for 30 s every 5 min, and the spectra were collected and analyzed as described above. The average of all nine biological replicates across three technical replicates with standard error of the mean is reported (Figure 4C, Figure S19–S21).

Validation of GR2-CFP at the single cell level in *E. coli* with fluorescence microscopy.

Three-milliliter cultures of *E. cloni* EXPRESS BL21 (DE3) expressing GR2-CFP were prepared as described above for the 25 mL cultures. After expression, the cells were harvested by centrifugation at 2,500*g* (5810 R, Eppendorf) for 5 min at room temperature. The resulting cell pellet was resuspended in 200 μL of 50 mM sodium acetate buffer at pH 5 and 25 μL of the resuspension was further diluted into 175 μL of 50 mM sodium acetate buffer at pH 5 or buffer containing 457 mM sodium chloride to final concentrations of 0 mM or 400 mM sodium chloride. The cells were gently mixed, and 3–6 μL were transferred to 1.5% agarose pads containing 0 mM or 400 mM sodium chloride and imaged on 35 mm dishes with a 20 mm glass-bottom coverslip (No. 1.5, MatTek) as previously described.[21]

The fluorescence images were acquired as Z-stacks using a confocal laser scanning microscope (FV3000RS, Olympus) with a 100X silicon immersion objective.[21] Briefly, excitation for the rhodopsin was provided at 561 nm, and the emission was collected from 630–730 nm. Excitation for CFP was provided at 405 nm, and emission was collected from 450–550 nm. For each Z-stack, the maximum intensity Z-projection for the CFP channel was first duplicated and then the default threshold settings were applied using the software Fiji Is Just ImageJ (Fiji) v2.0.[60] Masks were created using the thresholded CFP image, and the Fiji Analyze Particles function was used to select regions of interest (ROIs) that were greater than 3 pixels with circularity between 0 and 1. Overlapping cells were considered as one ROI. For each sample, the ROIs were transferred to the unprocessed rhodopsin and CFP maximum intensity Z-projections, and the median fluorescence intensities were measured. ROIs that did not appear in both the rhodopsin and CFP channels were manually

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

excluded. The rhodopsin median fluorescence intensity was normalized to the CFP median fluorescence intensity (Figure 5, Figure S22 and S23). At least five different fields were analyzed for the four biological replicates. For GR2-CFP, n = 3,352 ROIs for 0 mM sodium chloride and n = 3,062 ROIs for 400 mM sodium chloride were analyzed. This procedure was repeated in parallel on the same day with one technical replicate of GR1-CFP as a positive control for comparison (Figure S25). For GR1-CFP, n = 437 ROIs for 0 mM sodium chloride; n = 472 ROIs for 400 mM sodium chloride were analyzed.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

We acknowledge Dr. Ved Prakash from the Olympus Discovery Center at UT Dallas for expert technical assistance with confocal microscopy experiments. F.M. acknowledges support from the National Science Foundation (MCB-1943442) and the National Institute of General Medical Sciences of the National Institutes of Health (R35GM133631). S.C.D. acknowledges support from The University of Texas at Dallas, the Welch Foundation (AT-1918-20170325, AT-2060-20210327), and the National Institute of General Medical Sciences of the National Institutes of Health (R35GM128923). This manuscript is the sole responsibility of the authors and does not represent the views of the funding sources.

References

(1). Berend K; Van Hulsteijn LH; Gans ROB Chloride: The queen of electrolytes? Eur. J. Intern. Med 2012, 23, 203–211. [PubMed: 22385875]

(2). Raven JA; Raines C Chloride: essential micronutrient and multifunctional beneficial ion. J. Exp. Bot 2017, 68, 359–367.

(3). Valdivieso ÁG; Santa-Coloma TA The chloride anion as a signalling effector. Biol. Rev 2019, 94, 1839–1856. [PubMed: 31231963]

(4). Jentsch TJ; Pusch M CLC chloride channels and transporters: Structure, function, physiology, and disease. Physiol. Rev 2018, 98, 1493–1590. [PubMed: 29845874]

(5). Zajac M; Chakraborty K; Saha S; Mahadevan V; Infield DT; Accardi A; Qiu Z; Krishnan Y What biologists want from their chloride reporters — a conversation between chemists and biologists. J. Cell Sci 2020, 133, 1–13.

(6). Arosio D; Ratto GM Twenty years of fluorescence imaging of intracellular chloride. Front. Cell. Neurosci 2014, 8, 1–12. [PubMed: 24478626]

(7). Baek K; Ji K; Peng W; Liyanaarachchi SM; Dodani SC The design and evolution of fluorescent protein-based sensors for monoatomic ions in biology. Protein Eng. Des. Sel 2021, 34, 1–9.

(8). Lodovichi C; Ratto GM; Trevelyan AJ; Arosio D Genetically encoded sensors for chloride concentration. J. Neurosci. Methods 2022, 368, 1–10.

(9). Prasher DC; Eckenrode VK; Ward WW; Prendergast FG; Cormier MJ Primary structure of the Aequorea victoria green-fluorescent protein. Gene 1992, 111, 229–233. [PubMed: 1347277]

(10). Ormö M; Cubitt AB; Kallio K; Gross LA; Tsien RY; Remington SJ; Ormo M; Remingtont SJ Crystal Structure of the Aequorea victoria green fluorescent protein. Science 1996, 273, 1392–1395. [PubMed: 8703075]

(11). Tsien RY The green fluorescent protein. Annu. Rev. Biochem 1998, 67, 509–544. [PubMed: 9759496]

(12). Wachter RM; Elsliger M-A; Kallio K; Hanson GT; Remington J Structural basis of spectral shifts in the yellow-emission variants of green fluorescent protein. Structure 1998, 6, 1267–1277. [PubMed: 9782051]

(13). Wachter RM; James Remington S Sensitivity of the yellow variant of green fluorescent protein to halides and nitrate. Curr. Biol 1999, 9, R628–R629. [PubMed: 10508593]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

(14). Arosio D; Garau G; Ricci F; Marchetti L; Bizzarri R; Nifosì R; Beltram F Spectroscopic and structural study of proton and halide ion cooperative binding to GFP. Biophys. J 2007, 93, 232–244. [PubMed: 17434942]

(15). Wachter RM; Yarbrough D; Kallio K; Remington SJ Crystallographic and energetic analysis of binding of selected anions to the yellow variants of green fluorescent protein. J. Mol. Biol 2000, 301, 157–171. [PubMed: 10926499]

(16). Galietta LJ .; Haggie PM; Verkman A . Green fluorescent protein-based halide indicators with improved chloride and iodide affinities. FEBS Lett. 2001, 499, 220–224. [PubMed: 11423120]

(17). Grimley JS; Li L; Wang W; Wen L; Beese LS; Hellinga HW; Augustine GJ Visualization of synaptic inhibition with an optogenetic sensor developed by cell-free protein engineering automation. J. Neurosci 2013, 33, 16297–16309. [PubMed: 24107961]

(18). Zhong S; Navaratnam D; Santos-Sacchi J A genetically-encoded YFP sensor with enhanced chloride sensitivity, photostability, and reduced pH interference demonstrates augmented transmembrane chloride movement by gerbil prestin (SLC26a5). PLoS One 2014, 9, 1–9.

(19). Tutol JN; Peng W; Dodani SC Discovery and Characterization of a Naturally Occurring, Turn-On Yellow Fluorescent Protein Sensor for Chloride. Biochemistry 2019, 58.

(20). Tutol JN; Kam HC; Dodani SC Identification of mNeonGreen as a pH-Dependent, Turn-On Fluorescent Protein Sensor for Chloride. ChemBioChem 2019, 20, 1759–1765. [PubMed: 30843313]

(21). Tutol JN; Lee J; Chi H; Faizuddin FN; Abeyrathna SS; Zhou Q; Morcos F; Meloni G; Dodani SC A single point mutation converts a proton-pumping rhodopsin into a red-shifted, turn-on fluorescent sensor for chloride. Chem. Sci 2021, 12, 5655–5663. [PubMed: 34163777]

(22). Choi AR; Shi L; Brown LS; Jung K-H Cyanobacterial light-driven proton pump, Gloeobacter rhodopsin: Complementarity between rhodopsin-based energy production and photosynthesis. PLoS One 2014, 9, 1–10.

(23). Engqvist MKM; McIsaac RS; Dollinger P; Flytzanis NC; Abrams M; Schor S; Arnold FH Directed evolution of Gloeobacter violaceus rhodopsin spectral properties. J. Mol. Biol. 2015, 427, 205–220. [PubMed: 24979679]

(24). Morizumi T; Ou W-L; Van Eps N; Inoue K; Kandori H; Brown LS; Ernst OP X-ray crystallographic structure and oligomerization of Gloeobacter rhodopsin. Sci. Rep 2019, 9, 1–14. [PubMed: 30626917]

(25). Zabelskii D; Dmitrieva N; Volkov O; Shevchenko V; Kovalev K; Balandin T; Soloviov D; Astashkin R; Zinovev E; Alekseev A; Round E; Polovinkin V; Chizhov I; Rogachev A; Okhrimenko I; Borshchevskiy V; Chupin V; Büldt G; Yutin N et al. Structure-based insights into evolution of rhodopsins. Commun. Biol 2021, 4, 1–12. [PubMed: 33398033]

(26). Imasheva ES; Balashov SP; Choi AR; Jung K-H; Lanyi JK Reconstitution of Gloeobacter violaceus rhodopsin with a light-harvesting carotenoid antenna. Biochemistry 2009, 48, 10948–10955. [PubMed: 19842712]

(27). Kralj JM; Hochbaum DR; Douglass AD; Cohen AE Electrical spiking in Escherichia coli probed with a fluorescent voltage-indicating protein. Science 2011, 333, 345–348. [PubMed: 21764748]

(28). Maclaurin D; Venkatachalam V; Lee H; Cohen AE Mechanism of voltage-sensitive fluorescence in a microbial rhodopsin. Proc. Natl. Acad. Sci. USA 2013, 110, 5939–5944. [PubMed: 23530193]

(29). Govorunova EG; Sineshchekov OA; Li H; Spudich JL Microbial rhodopsins: Diversity, mechanisms, and optogenetic applications. Annu. Rev. Biochem 2017, 86, 845–872. [PubMed: 28301742]

(30). Yu JK; Liang R; Liu F; Martínez TJ First-principles characterization of the elusive I fluorescent state and the structural evolution of retinal protonated Schiff Base in bacteriorhodopsin. J. Am. Chem. Soc 2019, 141, 18193–18203. [PubMed: 31621314]

(31). Rozenberg A; Inoue K; Kandori H; Béjà O Microbial rhodopsins: The last two decades. Annu. Rev. Microbiol 2021, 75, 427–447. [PubMed: 34343014]

(32). Miranda MRM; Choi AR; Shi L; Bezerra AG; Jung K-H; Brown LS The photocycle and proton translocation pathway in a cyanobacterial ion-pumping rhodopsin. Biophys. J 2009, 96, 1471–1481. [PubMed: 19217863]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

(33). Vogt A; Wietek J; Hegemann P Gloeobacter rhodopsin, limitation of proton pumping at high electrochemical load. Biophys. J 2013, 105, 2055–2063. [PubMed: 24209850]

(34). Besaw JE; Ou W-L; Morizumi T; Eger BT; Sanchez Vasquez JD; Chu JHY; Harris A; Brown LS; Dwayne Miller RJ; Ernst OP The crystal structures of a chloride-pumping microbial rhodopsin and its proton-pumping mutant illuminate proton transfer determinants. J. Biol. Chem 2020, 295, 14793–14804. [PubMed: 32703899]

(35). Muroda K; Nakashima K; Shibata M; Demura M; Kandori H Protein-bound water as the determinant of asymmetric functional conversion between light-driven proton and chloride pumps. Biochemistry 2012, 51, 4677–4684. [PubMed: 22583333]

(36). Inoue K; Kato Y; Kandori H Light-driven ion-translocating rhodopsins in marine bacteria. Trends Microbiol. 2015, 23, 91–98. [PubMed: 25432080]

(37). Weigt M; White RA; Szurmant H; Hoch JA; Hwa T Identification of direct residue contacts in protein-protein interaction by message passing. Proc. Natl. Acad. Sci. U.S.A 2009, 106, 67–72. [PubMed: 19116270]

(38). Morcos F; Pagnani A; Lunt B; Bertolino A; Marks DS; Sander C; Zecchina R; Onuchic JN; Hwa T; Weigt M Direct-coupling analysis of residue coevolution captures native contacts across many protein families. Proc. Natl. Acad. Sci. U.S.A 2011, 108, E1293–E1301. [PubMed: 22106262]

(39). Sułkowska JI; Morcos F; Weigt M; Hwa T; Onuchic JN Genomics-aided structure prediction. Proc. Natl. Acad. Sci. U.S.A 2012, 109, 10340–10345. [PubMed: 22691493]

(40). De Leonardis E; Lutz B; Ratz S; Cocco S; Monasson R; Schug A; Weigt M Direct-Coupling Analysis of nucleotide coevolution facilitates RNA secondary and tertiary structure prediction. Nucleic Acids Res. 2015, 43, 10444–10455. [PubMed: 26420827]

(41). Morcos F; Jana B; Hwa T; Onuchic JN Coevolutionary signals across protein lineages help capture multiple protein conformations. Proc. Natl. Acad. Sci. U.S.A 2013, 110, 20533–20538. [PubMed: 24297889]

(42). Cheng RR; Morcos F; Levine H; Onuchic JN Toward rationally redesigning bacterial two-component signaling systems using coevolutionary information. Proc. Natl. Acad. Sci. U.S.A 2014, 111, 1–9.

(43). Cheng RR; Nordesjö O; Hayes RL; Levine H; Flores SC; Onuchic JN; Morcos F Connecting the sequence-space of bacterial signaling proteins to phenotypes using coevolutionary landscapes. Mol. Biol. Evol 2016, 33, 3054–3064. [PubMed: 27604223]

(44). Bitbol AF; Dwyer RS; Colwell LJ; Wingreen NS Inferring interaction partners from protein sequences. Proc. Natl. Acad. Sci. U.S.A 2016, 113, 12180–12185. [PubMed: 27663738]

(45). Schug A; Weigt M; Onuchic JN; Hwa T; Szurmant H High-resolution protein complexes from integrating genomic information with molecular simulation. Proc. Natl. Acad. Sci. USA 2009, 106, 22124–22129. [PubMed: 20018738]

(46). Hopf TA; Schärfe CPI; Rodrigues JPGLM; Green AG; Kohlbacher O; Sander C; Bonvin AMJJ; Marks DS Sequence co-evolution gives 3D contacts and structures of protein complexes. eLife 2014, 3, 1–45.

(47). Ovchinnikov S; Kamisetty H; Baker D Robust and accurate prediction of residue-residue interactions across protein interfaces using evolutionary information. eLife 2014, 2014, 1–21.

(48). Dos Santos RN; Morcos F; Jana B; Andricopulo AD; Onuchic JN Dimeric interactions and complex formation using direct coevolutionary couplings. Sci. Rep 2015, 5, 1–10.

(49). Bai F; Morcos F; Cheng RR; Jiang H; Onuchic JN Elucidating the druggable interface of protein-protein interactions using fragment docking and coevolutionary analysis. Proc. Natl. Acad. Sci. U.S.A 2016, 113, E8051–E8058. [PubMed: 27911825]

(50). Zhou Q; Kunder N; De La Paz JA; Lasley AE; Bhat VD; Morcos F; Campbell ZT Global pairwise RNA interaction landscapes reveal core features of protein recognition. Nat. Commun 2018, 9, 1–10. [PubMed: 29317637]

(51). Russ WP; Figliuzzi M; Stocker C; Barrat-Charlaix P; Socolich M; Kast P; Hilvert D; Monasson R; Cocco S; Weigt M; Ranganathan R An evolution-based model for designing chorismate mutase enzymes. Science 2020, 369, 440–445. [PubMed: 32703877]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

(52). Jiang XL; Dimas RP; Chan CTY; Morcos F Coevolutionary methods enable robust design of modular repressors by reestablishing intra-protein interactions. Nat. Commun 2021, 12, 1–8. [PubMed: 33397941]

(53). Tsukamoto T; Kikukawa T; Kurata T; Jung K-H; Kamo N; Demura M Salt bridge in the conserved His-Asp cluster in Gloeobacter rhodopsin contributes to trimer formation. FEBS Lett. 2013, 587, 322–327. [PubMed: 23313943]

(54). Iizuka A; Kajimoto K; Fujisawa T; Tsukamoto T; Aizawa T; Kamo N; Jung KH; Unno M; Demura M; Kikukawa T Functional importance of the oligomer formation of the cyanobacterial H+ pump Gloeobacter rhodopsin. Sci. Rep 2019, 9, 1–12. [PubMed: 30626917]

(55). Eddy SR Accelerated profile HMM searches. PLoS Comput. Biol 2011, 7, 1–16.

(56). Potter SC; Luciani A; Eddy SR; Park Y; Lopez R; Finn RD HMMER web server: 2018 update. Nucleic Acids Res. 2018, 46, W200–W204. [PubMed: 29905871]

(57). Morcos F; Schafer NP; Cheng RR; Onuchic JN; Wolynes PG Coevolutionary information, protein folding landscapes, and the thermodynamics of natural selection. Proc. Natl. Acad. Sci. U.S.A 2014, 111, 12408–12413. [PubMed: 25114242]

(58). Šali A; Blundell TL Comparative protein modelling by satisfaction of spatial restraints. J. Mol. Biol 1993, 234, 779–815. [PubMed: 8254673]

(59). Kille S; Acevedo-Rocha CG; Parra LP; Zhang Z-G; Opperman DJ; Reetz MT; Acevedo JP Reducing codon redundancy and screening effort of combinatorial protein libraries created by saturation mutagenesis. ACS Synth. Biol 2013, 2, 83–92. [PubMed: 23656371]

(60). Schindelin J; Arganda-Carreras I; Frise E; Kaynig V; Longair M; Pietzsch T; Preibisch S; Rueden C; Saalfeld S; Schmid B; Tinevez J-Y; White DJ; Hartenstein V; Eliceiri K; Tomancak P; Cardona A Fiji: an open-source platform for biological-image analysis. Nat. Methods 2012, 9, 676–682. [PubMed: 22743772]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

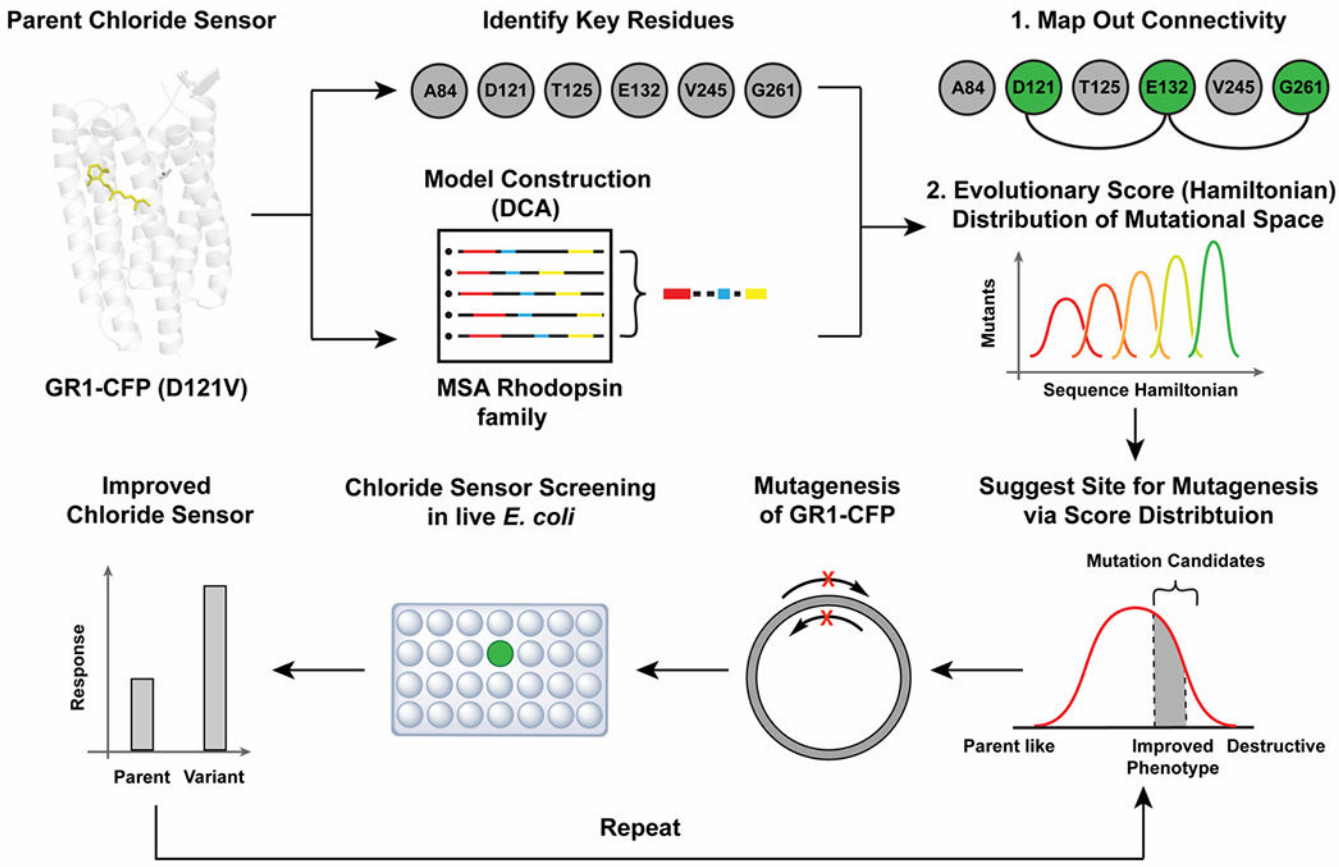


Figure 1.

The workflow combining directed evolution and coevolutionary landscapes to improve the properties of the fluorescent rhodopsin chloride sensor GR1-CFP was divided into three steps. First, residues along the proton-pumping pathway of wild-type *Gloeobacter violaceus* rhodopsin (*wt*GR) were identified from the literature and projected onto a direct coupling analysis of the rhodopsin family to build a connectivity map. Then, the available mutational space was refined based on the Hamiltonian distribution to guide the mutagenesis order. In parallel, an iterative site-saturation mutagenesis, screening, and selection strategy in live *E. coli* was developed to identify variants with an improved response to chloride.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

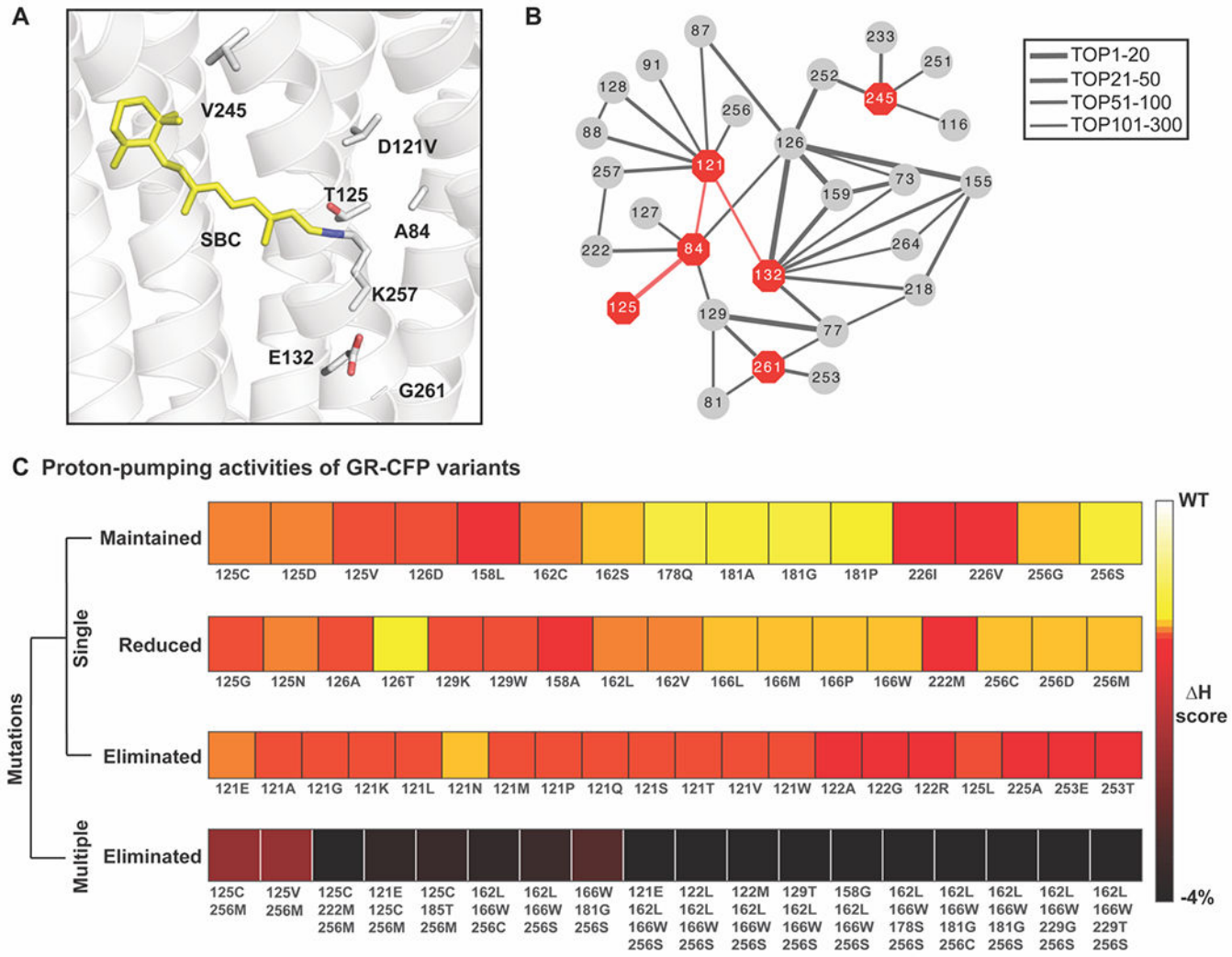


Figure 2.

Coevolutionary information reveals important functional residues in the rhodopsin family. (A) Homology model of the rhodopsin chloride biosensor GR1 engineered from the proton-pumping rhodopsin *Gloeobacter violaceus* (*wt*GR, PDB ID: 6NWD)[24] without CFP. The residues along the proton-pumping pathway selected for mutagenesis are shown. All residues are shown as gray sticks with the oxygen atoms in red and nitrogen atom in blue, each labeled with the single letter amino acid code and position number. The Schiff base chromophore (SBC) formed from all-*trans*-retinal (yellow sticks) and K257 is shown. (B) Connectivity map of the coupled pairs for the selected residues in the top 300 direction information (DI) pairs (Figure S2, Supplemental File S1). The position numbers of the residues along the proton-pumping pathway are shown as red octagons and the evolutionary coupled residues as gray circles. Each pair of coupled residues is connected with a line. The thickness of the lines correlates with a higher rank of the DI pair indicating a stronger evolutionary coupling strength. (C) Comparison of the proton-pumping activity levels with the Hamiltonian scores for previously reported GR-CFP variants used to validate the coevolution method.[23] The variants with the same level of proton-pumping activity

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

are grouped together in the same row. The relative changes of the Hamiltonian scores are illustrated in the heatmap from white (corresponding to a *wt*GR phenotype) to black with each mutation listed with the single letter amino acid code and position number.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

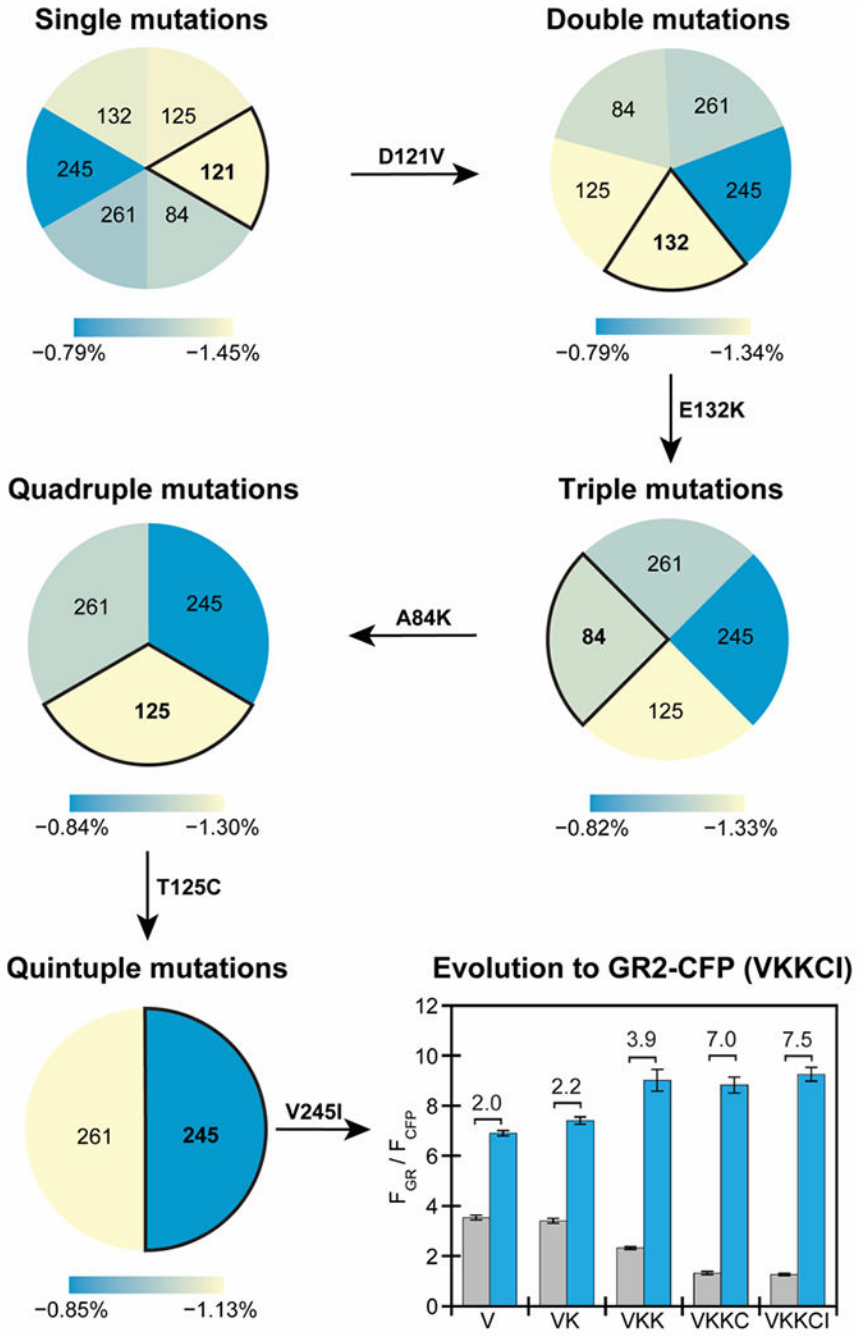


Figure 3.

Iterative site-saturation mutagenesis of residues along the proton-pumping pathway in GR1-CFP guided by a Hamiltonian score prediction results in GR2-CFP with an improved fluorescence response. From single mutations to quintuple mutations, the mutational space was refined by comparing the relative changes of the average Hamiltonian score from blue to yellow (largest change) between all the possible mutations at the positions listed with respect to the parent. In each step, the position with the largest change in the Hamiltonian score was selected for mutation. The identity of the variant with an improved fluorescence

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

response to chloride is shown above the arrow as the starting parent for the next site targeted for mutagenesis. The fluorescence response of GR1-CFP and each improved variant generated from site-saturation mutagenesis and screening in live *E. coli* is summarized in the bar graph. Each bar represents the ratio of the normalized integrated emission (F_{GR}/F_{CFP}) in the absence (gray) and presence (blue) of 400 mM sodium chloride in 50 mM sodium acetate at pH 5. For each bar, the average of twelve biological replicates from three technical replicates with standard error of the mean is shown. The fluorescence response to chloride ($\Delta F_{GR}/F_{CFP}$) is listed for each variant in the plot. For the rhodopsin, excitation was provided at 570 nm, and the emission was collected and integrated from 630–760 nm (F_{GR}). For CFP, excitation was provided at 390 nm, and the emission was collected from 435–520 nm (F_{CFP}). Emission spectra are shown in Figure S7 and S8.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

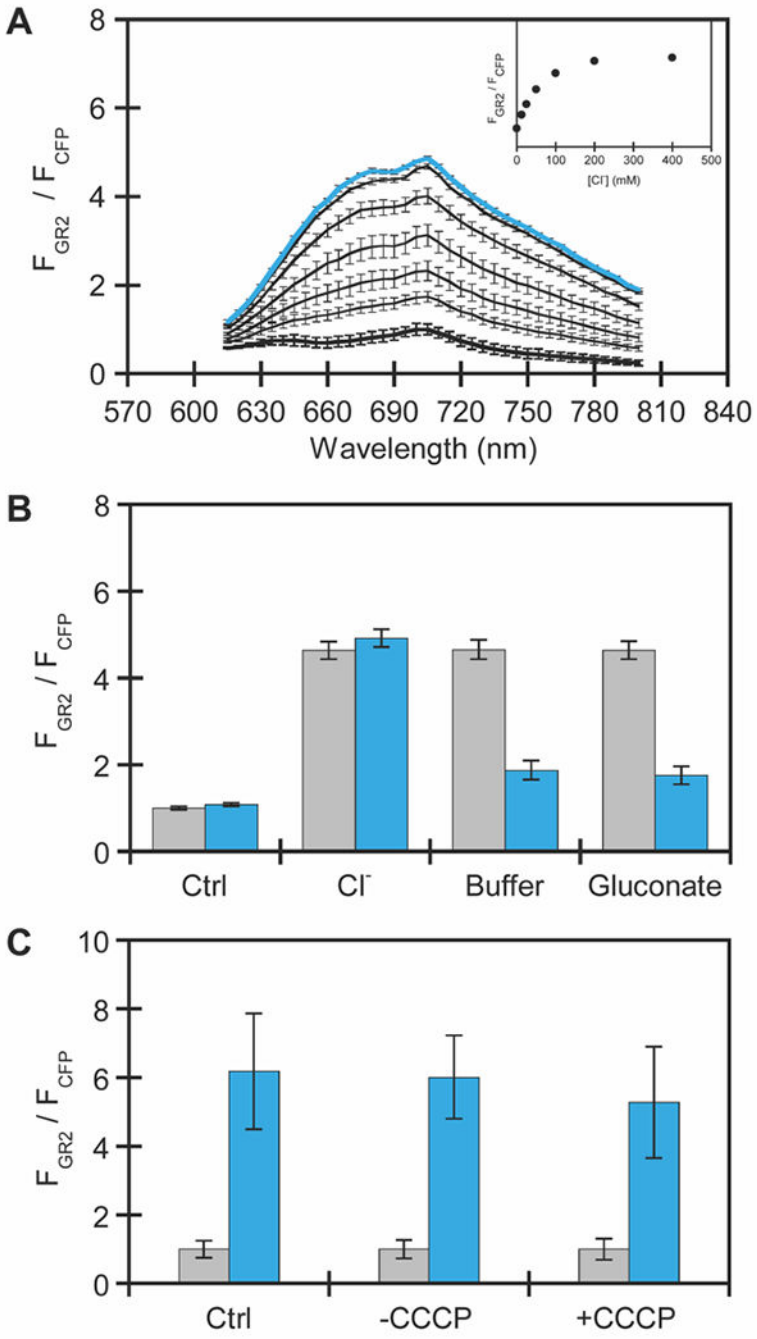


Figure 4.

Characterization of GR2-CFP expressed in *E. coli* using bulk fluorescence plate reader experiments. (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 and bold) sodium chloride. The normalized emission response (F_{GR2} at 710 nm/F_{CFP} at 485 nm) is shown in the inset. (B) Normalized integrated emission response of GR2-CFP in *E. coli* first treated with buffer (Ctrl) or buffer with 400 mM sodium chloride (gray bars), followed by washing with buffer for the control sample and buffer with 400 mM sodium chloride, buffer only, or buffer with

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

400 mM sodium gluconate for the sodium chloride treated samples (blue bars). Spectra are shown in Figure S16–S18. (C) Normalized integrated emission response of GR2-CFP in *E. coli* treated with buffer, DMSO, or 30 μM CCCP in the presence of 0 mM or 400 mM sodium chloride. Spectra are shown in Figure S19–S21. All experiments were carried out in 50 mM sodium acetate buffer at pH 5. 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 GR2 emission spectra were normalized at each wavelength by the CFP emission intensity at 485 nm (F_{CFP}). For each experiment, the average of nine biological replicates from three technical replicates with standard error of the mean is shown.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

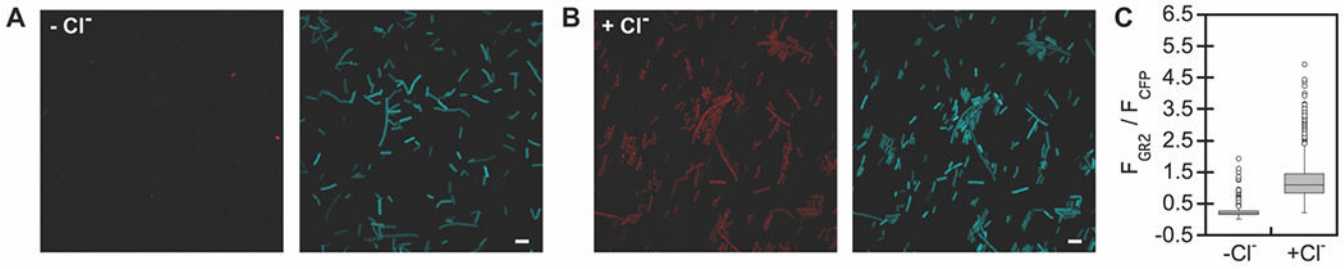


Figure 5.

Confocal fluorescence microscopy experiments demonstrate that GR2-CFP is a turn-on, ratiometric fluorescent sensor for chloride at the single cell level in exogenously supplemented *E. coli*. Representative images are shown for *E. coli* expressing GR2-CFP on agarose pads with (A) 0 mM and (B) 400 mM sodium chloride in 50 mM acetate buffer at pH 5. For each panel, the emission from GR2 (red) is on the left and emission from CFP (cyan) is on the right (scale bar = 5 μm). (C) Boxplots show the normalized emission response (F_{GR2}/F_{CFP}) of each cell analyzed from three biological replicates (n = 3,352 regions of interest (ROIs) for 0 mM sodium chloride; n = 3,062 ROIs for 400 mM sodium chloride, Figure S22 and S23). 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.