

Supporting Information

Bedbrook et al. 10.1073/pnas.1700269114

Parental ChR Constructs

Each of the three ChR library parent genes was built using a consistent vector backbone (pFCK) with the same promoter (CMV), trafficking signal (TS) sequence, and fluorescent protein (mKate). We used the pFCK vector from the construct FCK-CheRiff-eGFP [Addgene plasmid #51693 (41)]. A TS sequence (42) was inserted between the opsin and the fluorescent protein. The TS sequence has been shown to enhance opsin membrane trafficking (42). The GFP was replaced with mKate2.5 (43). Use of a red fluorescent protein as the marker for the opsin expression enabled use of SpyCatcher-GFP labeling for membrane-localized proteins. mKate2.5 is a monomeric far-red fluorescent protein that shows no aggregation. The mKate2.5 sequence was synthesized by IDT with overhangs for cloning into the desired vector system.

For the SpyTag/SpyCatcher membrane localization assay, it was necessary to add the SpyTag sequence close to the N terminus of each of the parental proteins and C-terminal to the signal peptide sequence cleavage site. For C1C2, an optimal position of the SpyTag had already been published. The SpyTag-C1C2 gene was amplified from the construct pLenti-CaMKIIa-SpyTag-C1C2-TS-mCherry (44) and inserted into the pFCK backbone. For CheRiff and CsChrimR, it was necessary to test various N-terminal SpyTag locations. The CheRiff gene was first amplified from FCK-CheRiff-eGFP [Addgene plasmid #51693 (41)], and the SpyTag sequence was added at different N-terminal positions by assembly PCR methods. The CsChrimR gene was built by assembly of the Cs N-terminal sequence (synthesized by IDT) with the C-terminal end of ChrimsonR amplified from the FCK-ChrimsonR-GFP construct [Addgene plasmid #59049 (39)]. The sequence of CsChrimR was designed to be identical to the previously published sequence (39). The SpyTag sequence was then inserted at different positions in the N-terminal region of the protein using assembly PCR methods. We tested three different pFCK-SpyTag-CheRiff-TS-mKate designs and three different pFCK-SpyTag-CsChrimR-TS-mKate designs and selected the design that showed expression and localization levels most similar to the nontagged parent.

Assembly-based methods and traditional cloning were used for vector construction and parental gene insertion. Annotated vector sequences of the three SpyTagged parental constructs are included as Datasets S3–S5.

Library Design

SCHEMA was used to design recombination libraries of the three parental ChRs to minimize the library-average disruption of the ChR structure (10, 25, 28). For the contiguous library, the SCHEMA-predicted block definitions were not modified. This 10-block library had roughly even-length blocks (14–43 residues), a relatively low average E value ($E = 25$), and whose sequences have an average of 73 mutations from the nearest parent. For the noncontiguous library, the SCHEMA-predicted block definitions were modified to group the N- or C-terminal domains into single blocks, maintain the presumptive dimer interface, and minimize the number of small blocks (less than five mutations). Specifically, a 13-block noncontiguous recombination library was generated for which two N-terminal blocks were combined, two C-terminal blocks were combined, two of four blocks in TM 5 were combined, and two residues of TM 3 were switched to the same block as TM 4 (where TM 3 and 4 make up the dimer interface observed for C1C2). The two loops that were not modeled in the C1C2 structure, between TM 1 and TM 2 and in the β -turn of the

C-terminal motif, were added to the block containing TM2 and the C-terminal block, respectively. The unmodeled residues of the N and C termini were added to the N- and C-terminal blocks. The resulting noncontiguous library had 10 blocks, an average E value of 23, an average of 71 mutations, and block size similar to the contiguous library (Fig. 2 C and D).

Among the three ChR parents, five unique N-linked glycosylation sites have been predicted by the NetNGlyc 1.0 (www.cbs.dtu.dk/services/NetNGlyc/) and GlycoEP servers (52). C1C2 harbors four of these sites with by far the highest confidence at each site. With one exception, the putative N-linked glycosylation sites do not overlap with recombination block borders. The exception site (SpyTag-C1C2 N95) is located in between the N-terminal domain and the first TM helix.

Contiguous recombination design was done using a software package for calculating SCHEMA energies and running the RASPP algorithm (23) openly available at cheme.che.caltech.edu/groups/fha/Software.htm (53). Noncontiguous recombination design was done using a software package for performing noncontiguous protein recombination (24) openly available at cheme.che.caltech.edu/groups/fha/Software.htm (54). Both software packages are written in the Python programming language.

Construction of Chimeras

The SCHEMA software outputs the amino acid sequences of all chimeras in a library. The amino acid sequence for each chimera chosen for experimental testing was converted into a nucleotide sequence using the following method to define codon use:

1. Align the amino acid sequence to the C1C2 parent.
2. Assign conserved amino acids in the alignment to the C1C2 parental codon.
3. Assign nonconserved amino acids to the parental codon from which the amino acid is derived.

This method was used for all chimeras to ensure that codon use was consistent. Once amino acid sequences were converted into nucleotide sequences, additional 3' and 5' sequences containing a BamHI and a NotI restriction enzyme cut site, respectively, were appended to the gene sequence. These sequences were necessary for cloning in the pFCK vector using either restriction ligation or homology-based cloning strategies. Gene sequences for the 223-chimera set were synthesized by Twist Bioscience, using its proprietary silicon-based DNA writing technology. After assembly, each fragment was cloned in the pFCK vector by homology-based cloning strategy and transformed into Stbl3 cells (Invitrogen) or Endura cells (Lucigen). Individual clones were picked and sequenced by NGS. Perfect clones were stored as individual glycerol stocks. Eight of the single-block swap sequences failed either the synthesis or cloning steps; these were not included in the chimera set.

Purified plasmid DNA of each chimera was prepared for HEK cell transfection. Each construct was streaked onto LB-amp plates from a glycerol stock, and an individual colony from each construct was picked and used to inoculate a 5-mL LB-ampicillin liquid media. Cultures were then grown overnight to reach saturation. Plasmid DNA for each construct was then purified using the QIAprep Spin Miniprep Kit. DNA concentrations for all constructs were measured and normalized before HEK cell transfection.

HEK Cell Maintenance and Transfection

HEK 293T cells were cultured at 37 °C and 5% CO₂ in D10 [DMEM supplemented with 10% (vol/vol) FBS, 1% sodium

bicarbonate, and 1% sodium pyruvate]. For 96-well transfections, HEK cells were plated on poly-D-lysine-coated glass-bottom 96-well plates at 20–30% confluency. Cells were left to divide until they reached 70–80% confluency. HEK cells were then transfected with one library variant per well at a pre-normalized DNA concentration using Eugene6 reagent according to the manufacturer's recommendations. Cells were given 48 h to express and then subjected to the SpyCatcher-GFP labeling assay and imaged.

Recombinant SpyCatcher-GFP Expression and Purification

The SpyCatcher-GFP was produced from a previously published construct—*pQE80l-T5::6xhis-SpyCatcher-Elp-GFP* [for details, see Bedbrook et al. (44)]. *E. coli* expression strain *BL21(DE3)* harboring the *pQE80l-T5::6xhis-SpyCatcher-Elp-GFP* plasmid was grown at 37 °C in TB medium to an optical density of 0.6–0.8 at 600 nm, and protein expression was induced using 1 mM isopropyl β -D-1-thiogalactopyranoside at 30 °C. After 4 h of induction, cells were harvested and frozen at –80 °C before protein purification. Protein purification was carried out using HisTrap columns (GE Healthcare) following the column manufacturer's recommendations. Protein was buffer exchanged into sterile PBS at 4 °C. Protein was stable through multiple freeze/thaws and over many months.

SpyCatcher Labeling of HEK Cells

HEK cells were subjected to SpyCatcher labeling 48 h post-transfection. Labeling was done in a 96-well format using multichannel pipettes. SpyCatcher-GFP was added directly into the D10 media of wells containing HEK cells at a final concentration of 30 μ M, and the cells were then incubated for 45 min at 25 °C. To avoid variability in labeling in the 96-well format screen, we used a saturating concentration of the SpyCatcher (30 μ M) for labeling experiments. After labeling, HEK cells were washed with D10 three times, and then cells were incubated at 37 °C for 1 h to allow any remaining SpyCatcher to diffuse off of the well surface. For cell imaging, D10 medium was replaced with extracellular buffer (in mM: 140 NaCl, 5 KCl, 10 Hepes, 2 MgCl₂, 2 CaCl₂, 10 glucose; pH 7.35) to avoid the high autofluorescence of the D10. Cells were washed two times with extracellular buffer to fully remove any residual D10 before imaging.

Imaging and Image Processing of ChR Expression and Localization

Imaging of ChR expression and localization was done using a Leica DMI 6000 microscope. Four positions in each well were imaged in all 96-well plates using a fully automated system with motorized stage and automated *z* focus. Three channels were imaged at each position (mKate, GFP, and bright field). Cell segmentation was done using CellProfiler (55), an open-source image-processing software, and whole population intensity measurements were done using custom image-processing scripts written using open-source packages in the SciPy ecosystem (56–58). Both processing methods require a series of filtering steps and background sub-

traction. Whole population intensity measurements required a thresholding step when defining a pixel mask for image processing. We used wells containing nontransfected HEK cell that went through the labeling experiment as a background for establishing a threshold. A threshold was set to 2 SDs above the mean intensity values calculated in these background wells for each channel (mKate and GFP). For each image, a mask was defined for each channel (mKate and GFP) as the pixels above a set threshold. The masks for the two channels were then combined so that the mask included any pixel that was above threshold in the GFP channel or the mKate channel. This combined pixel mask was used to calculate the mean mKate fluorescence intensity (expression) and mean GFP fluorescence intensity (localization) across the pixels in the mask. The ratio mean mKate intensity/mean GFP intensity is the localization efficiency.

Electrophysiology for ChR Photocurrents

Conventional whole-cell patch-clamp recordings were done in cultured HEK cells at 2 d posttransfection. Cells were continuously perfused with extracellular solution at room temperature (in mM: 140 NaCl, 5 KCl, 10 Hepes, 2 MgCl₂, 2 CaCl₂, 10 glucose; pH 7.35) while mounted on the microscope stage. Patch pipettes were fabricated from borosilicate capillary glass tubing (1B150-4; World Precision Instruments) using a model P-2000 laser puller (Sutter Instruments) to resistances of 2–5 M Ω . Pipettes were filled with intracellular solution containing the following (in mM): 134 K gluconate, 5 EGTA, 10 Hepes, 2 MgCl₂, 0.5 CaCl₂, 3 ATP, and 0.2 GTP. Whole-cell patch-clamp recordings were made using a Multiclamp 700B amplifier (Molecular Devices), a Digidata 1440 digitizer (Molecular Devices), and a PC running pClamp (version 10.4) software (Molecular Devices) to generate current injection waveforms and to record voltage and current traces.

Patch-clamp recordings were done with short light pulses to measure photocurrents. Photocurrents for each chimera were induced by three different wavelengths of light (473 ± 10 , 560 ± 25 , and 650 ± 13 nm) at 2 mW (~ 0.1 mW \cdot mm^{–2}). Photocurrents were recorded from cells in voltage clamp held at –50 mV with one light pulse for 1 s with each wavelength of light tested sequentially with 2 min between light exposures. Because ChRs show some level of desensitization to light after continued light exposure, we ran all colors in one direction (red \rightarrow green \rightarrow blue) and then again in the other direction (blue \rightarrow green \rightarrow red). The means of peak and steady-state currents were calculated for each color between the two trials for a given cell. Light wavelengths were produced using LED illumination using a Lumencor SPECTRAX light engine with quad band 387/485/559/649-nm excitation filter, quad band 410/504/582/669-nm dichroic mirror, and quad band 440/521/607/700-nm emission filter (all SEMROCK).

Electrophysiology data were analyzed using custom data-processing scripts written using open-source packages in the Python programming language to do baseline adjustments, find the peak inward currents, and find the steady-state currents.

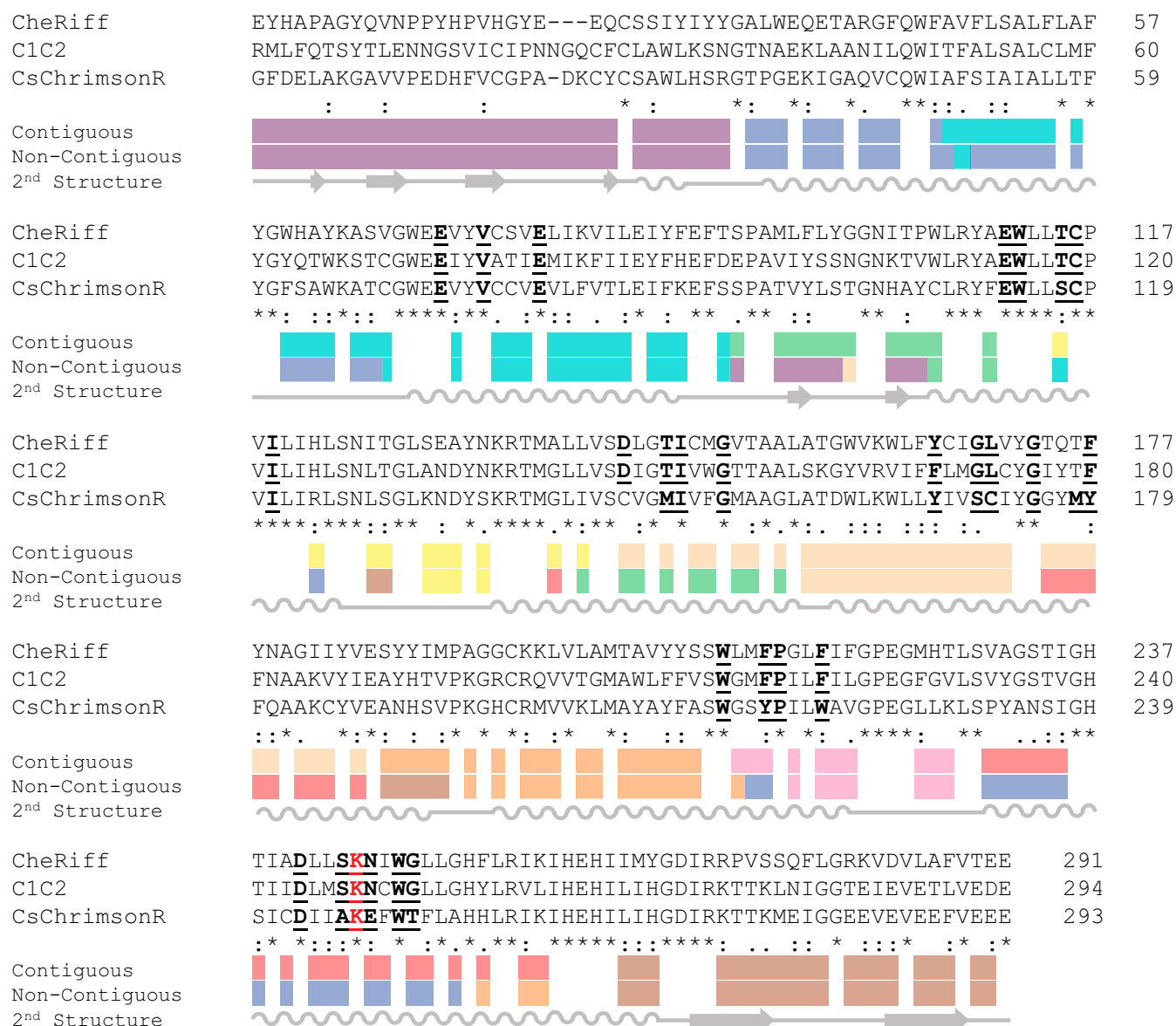


Fig. S1. Amino acid alignment of parental sequences and recombination block designs. Alignment showing the contiguous and noncontiguous block designs. Each color represents a different block, and white shows the conserved residues. Amino acids thought to be important for ChR spectral properties are bolded and underlined. The conserved lysine residue that participates in a Schiff base linkage with retinal is highlighted in red text. The secondary structure is shown below the alignment.

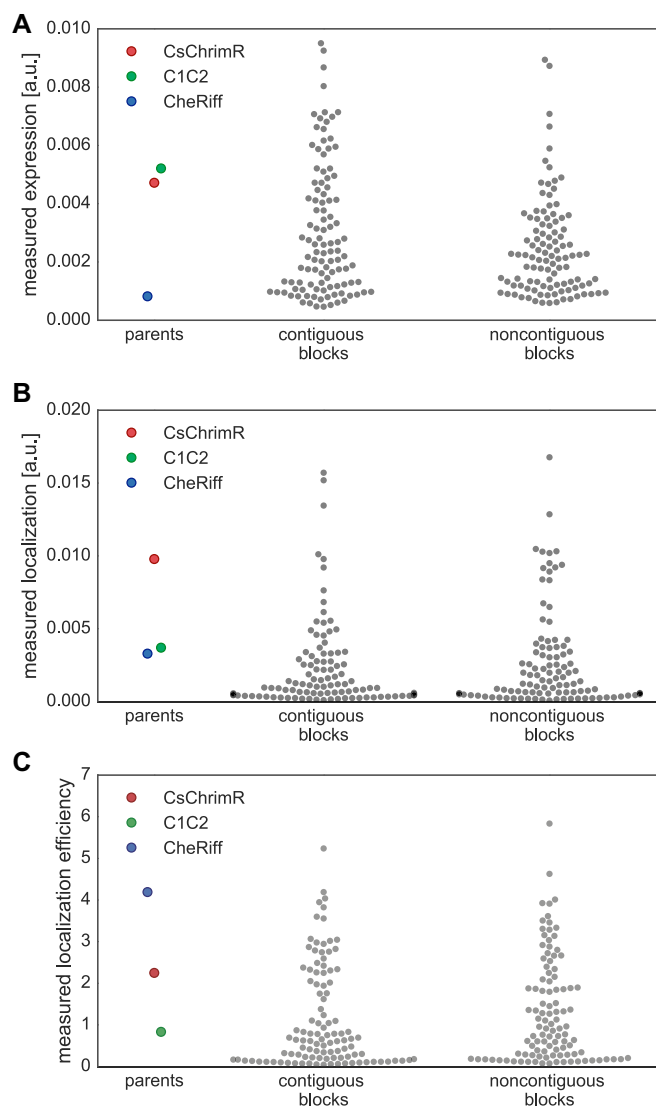


Fig. S4. Comparison of chimeras from the contiguous and noncontiguous recombination libraries. Swarm plot showing each chimera's expression [mean mKate fluorescence (in arbitrary units)] (A), localization [mean GFP fluorescence (in arbitrary units)] (B), and localization efficiency (mean mKate/GFP fluorescence) (C) for the contiguous and noncontiguous recombination libraries. Chimera data are plotted as gray points, and parental data points are highlighted in color (red, CsChrimR; green, C1C2; and blue, CheRiff).

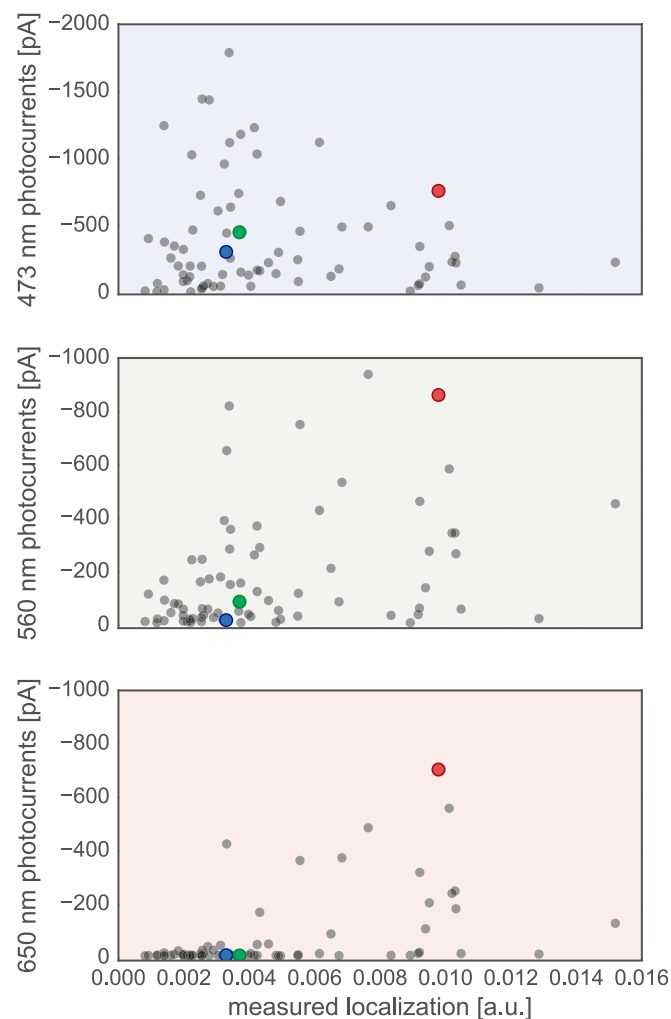


Fig. S6. Photocurrents vs. measured localization for all tested chimeras. Chimera data are plotted as gray points and parental data points are highlighted in color (red, CsChrimR; green, C1C2; and blue, CheRiff). Plot of measured photocurrents vs. measured localization [mean GFP fluorescence (in arbitrary units)] for three different wavelengths: 473 nm (*Top*, blue shading), 560 nm (*Middle*, green shading), and 650 nm (*Bottom*, red shading).

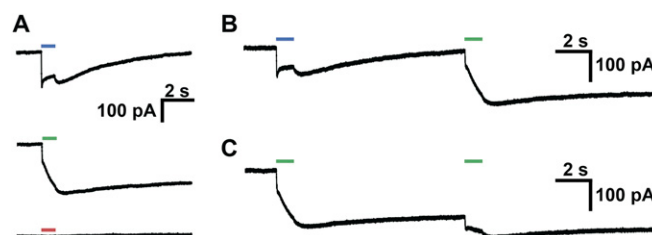


Fig. S7. One multiblock swap chimera with unique properties. (A) Chimera photocurrents upon 1-s exposure to 473-nm (*Top*), 560-nm (*Middle*), and 650-nm (*Bottom*) light. (B) Sequential activation of chimera with 473-nm and then 560-nm light. (C) Sequential activation of chimera with 560-nm and then 560-nm light.

PNAS

Measured localization and expression properties for each chimera tested and associated chimera name, chimera_block_ID, and sequence. Chimera names and chimera_block_ID begin with either "c" or "n" to indicate the contiguous or noncontiguous library. The following 10 digits in the chimera_block_ID indicate, in block order, the parent that contributes each of the 10 blocks ("0," CheRiff; "1," C1C2; and "2," CsChrimR). For the contiguous library, blocks in the chimera_block_ID are listed from N to C terminus; for the noncontiguous library, the block order is arbitrary. Sequences list only the ChR ORF; the C-terminal trafficking and mKate2.5 sequences have been removed, but are available parental sequences in annotated GenBank files. The table shows mean properties (mKate_mean, GFP_mean, and intensity_ratio_mean) and the SD of properties (mKate_SD, GFP_SD, and intensity_ratio_SD). ND, not detected, below the limit of detection for our assay.

Dataset S2

Functional characteristics of each tested chimera and associated chimera name and chimera_block_ID. Photocurrent was measured using patch-clamp electrophysiology in voltage-clamp mode upon exposure to 473-nm (cyan), 560-nm (green), or 650-nm (red) wavelength light. The table has mean peak and steady-state photocurrent (in picoamperes) and the SD of peak and steady-state photocurrent (in picoamperes) at each wavelength. The chimera_block_ID begins with either "c" or "n" to indicate the contiguous or noncontiguous library. The following 10 digits in the chimera_block_ID indicate, in block order, the parent that contributes each of the 10 blocks ("0," CheRiff; "1," C1C2; and "2," CsChrimR).

Dataset S3

Dataset S4. SpyTagged CheRiff sequence

Dataset S4

Dataset S5. SpyTagged CsChrimsonR sequence

Dataset S5