Supporting Information

Hydrogen/Deuterium Exchange Mass Spectrometry of Human Green Opsin Reveals a Conserved Pro-Pro Motif in Extracellular Loop 2 of Monostable Visual G-protein Coupled Receptors

Lukas Hofmann[1], Nathan S. Alexander[1], Wenyu Sun[1], Jianye Zhang[1], Tivadar Orban[1], and Krzysztof Palczewski*[1]*

AUTHOR ADDRESS: [1]Department of Pharmacology and Cleveland Center for Membrane and Structural Biology, School of Medicine, Case Western Reserve University, 10900 Euclid Avenue, Cleveland, OH 44106, USA

Sequence alignment of the M/LWS receptor and M/LWS construct. Sequences were aligned with the MUSCLE server provided by EMBL-EBI. [1-4] MUSCLE stands for MUltiple Sequence Comparison by Log-Expectation. The M/LWS construct starts with a haemagglutinin signal sequence, which is cleaved during protein maturation in *Sf*9 insect cells. The C-terminus of the M/LWS construct was replaced with the 1D4 sequence TETSQVAPA.

```
M/LWS               ----------------MAQQWSLQRLAGRHPQDSYEDSTQSSIFTYTNSNSTRGPFEGPN
M/LWS_construct     MKTIIALSYIFCLVFAMAQQWSLQRLAGRHPQDSYEDSTQSSIFTYTNSNSTRGPFEGPN
                                    ****************************************

M/LWS               YHIAPRWVYHLTSVWMIFVVIASVFTNGLVLAATMKFKKLRHPLNWILVNLAVADLAETV
M/LWS_construct     YHIAPRWVYHLTSVWMIFVVIASVFTNGLVLAATMKFKKLRHPLNWILVNLAVADLAETV
                    ************************************************************

M/LWS               IASTISVVNQVYGYFVLGHPMCVLEGYTVSLCGITGLWSLAIISWERWMVVCKPFGNVRF
M/LWS_construct     IASTISVVNQVYGYFVLGHPMCVLEGYTVSLCGITGLWSLAIISWERWMVVCKPFGNVRF
                    ************************************************************

M/LWS               DAKLAIVGIAFSWIWAAVWTAPPIFGWSRYWPHGLKTSCGPDVFSGSSYPGVQSYMIVLM
M/LWS_construct     DAKLAIVGIAFSWIWAAVWTAPPIFGWSRYWPHGLKTSCGPDVFSGSSYPGVQSYMIVLM
                    ************************************************************

M/LWS               VTCCITPLSIIVLCYLQVWLAIRAVAKQQKESESTQKAEKEVTRMVVVMVLAFCFCWGPY
M/LWS_construct     VTCCITPLSIIVLCYLQVWLAIRAVAKQQKESESTQKAEKEVTRMVVVMVLAFCFCWGPY
                    ************************************************************

M/LWS               AFFACFAAANPGYPFHPLMAALPAFFAKSATIYNPVIYVFMNRQFRNCILQLFGKKVDDG
M/LWS_construct     AFFACFAAANPGYPFHPLMAALPAFFAKSATIYNPVIYVFMNRQFRNCILQLFGKKVDDG
                    ************************************************************

M/LWS               SELSSASKTEVSSVSSVSPA
M/LWS_construct     SELSSASKTE---TSQVAPA
                    **********   .*.*:**
```

Regeneration, solubilization and purification of green cone opsin. Sf9 cell membranes from 5 liters of cell culture were disrupted with a Dounce homogenizer in a hypotonic buffer containing 25 mM HEPES, pH 7.0, and EDTA-free complete protease inhibitor cocktail (Roche, Indianapolis, IN). Cell lysis and washing of the membranes were performed by repeated passes in the Dounce homogenizer and centrifugation at 96,000 *g* for 20 min at 4 ˚C in the hypotonic buffer for a total of three times and then in a high osmotic buffer containing 1.0 M NaCl, 25 mM HEPES, pH 7.0, and EDTA-free complete protease inhibitor cocktail for three additional cycles. Cell membranes were centrifuged at 96,000 *g* for 20 min at 4 °C and stored at -80 °C. Prior to use, membranes were homogenized in 200 ml of 25 mM HEPES, pH 7.0, 50 mM NaCl. The following steps were enacted under dim red light. Green opsin was

regenerated with 40 µM 11-*cis*-retinal, and solubilized in ethanol for 1 h. Membrane pellets were solubilized by adding 1% (w/v) n-dodecyl β-D-maltopyranoside (DDM; Anatrace, Maumee, OH) to a final volume of 250 ml, and incubated for 3 h at 4 °C. The supernatant was collected after centrifugation for 45 min at 95,834 *g* and incubated with 7.5 ml of Rho1D4 monoclonal antibody linked to CNBr-activated Sepharose 4B (GE Healthcare, Waukesha, WI) in a batch for 3 h at 4 °C. After binding, the beads were collected by sedimentation and applied to a gravity flow column (BioRad, Hercules, CA). This step was followed by washing with 500 ml of buffer consisting of 25 mM HEPES, pH 7.0, 50 mM NaCl, and 0.02% (w/v) DDM. The green pigment was incubated with 10 ml of buffer and 1 mg/ml 1D4 peptide overnight and then was eluted with an additional 15 ml of buffer and 1 mg/ml 1D4 peptide. The eluted protein was concentrated to 500 µl with an Amicon 50 kDa NMWL cutoff filter (EMD Millipore, Billerica, MA). Finally, the pigment was passed through a Superdex 200 size-exclusion column (GE Healthcare) with a final buffer containing 20 mM HEPES, pH 7.0, 50 mM NaCl, and 0.02% DDM. The purified protein was concentrated to 1 mg/ml, flash frozen in 50 µl aliquots and stored at -80 °C for future use.

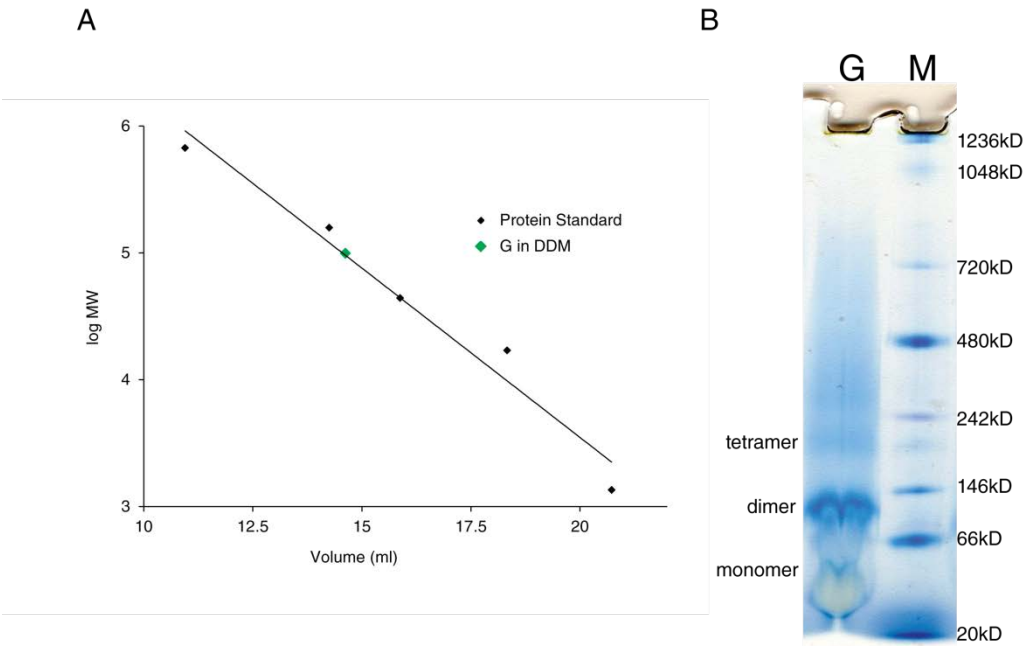

Figure S1. Analysis of oligomeric state by gel filtration and native PAGE of the green cone pigment in DDM. The gel filtration analysis (Panel A) was carried out with the Gel Filtration Standard (#151-1901) from BioRad. Standards were thyroglobulin (bovine), 670 kDa; γ-globulin (bovine), 158 kDa; ovalbumin (chicken), 44 kDa; myoglobin (horse), 17 kDa; vitamin B12, 1.35 kDa. Exponential plotting of the protein standards resulted in the following equation $y = 8*10^{8}*e^{-0.615*x}$ ($R^2 = 0.9666$). The experimentally determined MW of the green cone opsin was 99,284 Da. (B) Lane 1 displays the dark state green cone pigment solubilized in DDM. Lane 2 contains the protein marker NativeMark™, Novex (Life Technologies, Grand Island, NY). The NativePAGE™ 4-16% Bis-Tris Protein Gels were run according to the manufacturer's protocol (Novex) under dim red light conditions.

Immunoblotting of green cone pigment. 1D4 antibody was purified from mouse ascites fluid using a Pierce™ (Rockford, IL) Thiophilic Adsorption kit and conjugated to alkaline phosphatase with the Lightning-Link® Alkaline Phosphatase kit (Innova Bioscience, Cambridge, UK). The gel was transferred to an Immobilon®-P Transfer Membrane with the eBlot Protein Transfer Device (GenScript, Piscataway, NJ). The membrane was blocked with 5% nonfat-dry milk in TBST buffer for one hour. The membrane was then incubated with conjugated 1D4 antibody for three hours. After three washing steps with TBST for five min each, the membrane was developed for five min with Western Blue® Stabilized Substrate for Alkaline Phosphatase (Promega, Madison, WI).

Settings for peptide detection. Mass spectrometer settings were as follows: activation type, collision-induced dissociation; normalized collision energy, 35 kV; capillary temperature, 370 °C; source voltage, 5 kV; capillary voltage, 43 V; tube lens, 105 V. MS spectra were collected over a 200–2,000 m/z range. To avoid sample propagation from one run to the next,

each production run was followed with a mock injection of 10 µl of solution A (H_2O with 0.1 % (v/v) formic acid) and the same gradient program as described in Methods. This run then was followed by a 20-min equilibration run with 98% A and 2% B (acetonitrile 0.1 % (v/v) formic acid) for 20 min. Experiments were performed at least in triplicate for each time point.

H/D exchange analysis and normalization. Deuterium uptake was evaluated from the raw data, and the value for every peptide fragment was normalized to 80% of the theoretically maximum exchangeable sites to account for the 80% deuteration accomplished experimentally (Table S1, column 7). For calculations of the maximum exchangeable sites, only peptide bonds were used to account for the amide exchange; deuterium exchange from side chains was considered negligible and thus not included. The number of maximum exchangeable sites was decreased by one for each Pro residue in a peptide sequence under analysis (Table S1, column 6).

Table S1. Sequences of peptide fragments from the M/LWS construct sequence showing normalized deuterium uptake for dark M/LWS and bleached M/LWS receptor in the presence of DDM. Column 1 shows the peptide sequence from the M/LWS construct's primary sequence. Column 2 reveals the charge of the ion from Column 5. Column 3 reports the retention time (min) of the peptide. Column 4 shows the detected mass of the peptide (MW(obs)=m/z*charge-charge). Column 5 indicates the mass over charge (m/z) of the ion used to identify the peptide based on its MS/MS spectrum and Column 6 shows the maximum number of theoretically exchangeable sites of the peptide fragment (max = number of non-proline peptide bonds in the peptide fragment). Column 7 reveals the uptake normalized to 80% of the theoretical maximum exchangeable sites. The 80% normalization reflects the dilution percentage in D_2O (see Methods). Columns 8-9 and 10-11 indicate the uptake of the exchangeable sites in the dark state and in the bleached state, respectively. Column 12 shows the difference of uptake between the dark and bleached states. Columns 9 and 11 are colored and shaded according to a red (1) – yellow (25) – green (50) scale based on the value. Column 12 is colored and shaded in a green-red scale based on the value (red: negative, green: positive).

peptide	charge	RT	MW(obs)	m/z	max	realmax	dark	% exchange dark	bleached	% exchange bleached	ΔHDX
4-CLVFAMA	+2	15.55	754.55	377.78	6	4.8	1.56	32.50	1.28	26.67	5.83
4QWSLQRLAGR	+3	14.53	1219.71	407.24	9	7.2	2.50	34.72	1.48	20.56	14.17
12GRHPQDSYE	+2	13.18	1088.67	544.84	7	5.6	1.67	29.82	1.22	21.79	8.04
21DSTQSSIF	+1	17.14	884.36	884.36	7	5.6	2.01	35.89	1.69	30.18	5.71
29TYTDSDSTRGPFEGPNY	+2	16.2	1906.17	953.59	16	12.8	3.19	24.92	2.48	19.38	5.55

^{42}GPNYHI	+1	20.1	700.41	700.41	4	3.2	1.61	50.31	1.43	44.69	5.63
^{46}HIAPRWVYHL	+2	17.3	1291.49	646.25	8	6.4	1.17	18.28	0.75	11.72	6.56
^{56}TSVWMIFVVI	+3	13.76	1194.63	398.88	9	7.2	1.67	23.19	1.15	15.97	7.22
^{63}VVIASVFTNGLVL	+2	21.58	1331.55	666.28	12	9.6	0.39	4.06	2.02	21.04	-16.98
^{76}AATMKFKKLRHPLNWILVNL	+3	17.54	2393.58	798.53	18	14.4	1.84	12.78	1.84	12.78	0.00
^{83}KLRHPLNW	+2	13.75	1063.75	532.38	6	4.8	1.45	30.21	1.36	28.33	1.87
^{93}VNLAVAD	+1	16.08	701.39	701.38	6	4.8	1.06	22.08	1.18	24.58	-2.50
^{100}LAETVIA	+1	13.86	716.44	716.44	6	4.8	0.50	10.42	1.19	24.79	-14.38
^{107}STISVVNQVYGY	+2	18.54	1330.05	665.53	11	8.8	0.62	7.05	1.20	13.64	-6.59
^{119}FVLGHPMC	+2	16.75	903.42	452.37	6	4.8			0.62	12.92	
^{119}FVLGHPMCVL	+2	18.3	1115.87	558.44	8	6.4	0.36	5.63	0.67	10.47	-4.84
^{131}YTVSL	+1	17.77	582.29	582.29	4	3.2	0.27	8.44	0.81	25.31	-16.88
^{135}LCGITGLWSLA	+2	24.86	1133.31	567.16	10	8	0.26	3.25			
^{136}CGITGL	+1	18.18	563.31	563.3	5	4			0.79	19.75	
^{145}AIISWERW	+2	19.29	1060.85	530.93	7	5.6	0.19	3.39	0.43	7.68	-4.29
^{153}MVVCKPFGNVRF	+3	18.25	1398.63	466.88	10	8	1.83	22.88	1.21	15.13	7.75
^{164}FDAKLA	+2	19.66	661.51	331.26	5	4	0.49	12.25	0.30	7.50	4.75
^{169}AIVGIAF	+1	21.86	690.39	690.39	6	4.8	0.10	2.08	0.50	10.42	-8.33
^{173}IAFSWIWA	+1	22	993.51	993.51	7	5.6	0.11	1.96			
^{173}IAFSWIWAAV	+2	23.35	1164.05	582.53	9	7.2			0.77	10.69	

^{180}AAVWTAPPIFG	+2	21.3	1130.15	565.58	8	6.4	0.65	10.16	1.59	24.84	-14.69
^{183}WTAPPIFG	+1	18.68	888.46	888.46	5	4	0.49	12.25	1.48	37.00	-24.75
^{194}YWPHGLKT	+2	16.92	1089.87	545.44	8	6.4	0.54	8.44	0.35	5.47	2.97
^{210}GSSYPGVQSYM	+3	21.34	1170.06	390.69	9	7.2	0.26	3.61	0.32	4.44	-0.83
^{212}SYPGVQSYMIVL	+3	18.2	1351.59	451.2	10	8	0.33	4.13	0.04	0.50	3.63
^{224}MVTCCI	+1	18.12	669.32	669.32	5	4	0.61	15.25	0.55	13.75	1.50
^{229}ITPLSIIVL	+1	22.01	968.58	968.58	7	5.6			0.54	9.64	
^{231}PLSIIVLCYL	+2	23.03	1134.07	567.54	8	6.4	0.54	8.44			
^{239}YLQVW	+1	18.68	708.36	708.36	4	3.2	0.13	4.06	0.12	3.75	0.31
^{243}WLAIRAVAKQ	+3	19.62	1154.58	385.53	9	7.2	0.60	8.33	0.56	7.78	0.56
^{251}KQQKESESTQKAEKEVTRM	+3	20.91	2269.35	757.12	18	14.4			0.70	4.86	
^{252}QQKESESTQ	+2	11.84	1063.73	532.37	8	6.4	1.04	16.25	1.35	21.09	-4.84
^{261}KAEKEVTRM	+2	14.7	1091.91	546.96	8	6.4	1.23	19.22			
^{272}VMVLAFCFCWGPYAFFACFAAANPG	+3	17.24	2700.57	900.86	23	18.4	6.22	33.80	5.95	32.34	1.47
^{290}FAAANPGYPFHPLM	+3	19.08	1533.27	511.76	10	8	3.78	47.25	4.08	51.00	-3.75
^{304}AALPAF	+2	19.95	589.73	295.37	4	3.2	1.50	46.88	1.54	48.13	-1.25
^{310}FAKSATIYNPVIY	+2	18.2	1487.13	744.07	11	8.8	3.07	34.89	2.69	30.57	4.32
^{324}NRQFRNCILQL	+3	18.35	1406.37	469.46	10	8	1.09	13.63	1.21	15.13	-1.50
^{337}FGKKVDDGSELS	+2	12.82	1282.01	641.51	11	8.8	2.05	23.30	1.69	19.20	4.09
^{349}SASKTETSQVAPA	+2	11.38	1277.01	639.01	11	8.8	3.26	37.05	3.38	38.41	-1.36

The Pro-Pro motif in human opsins. Sequence alignment of the M/LWS construct, human long wavelength sensitive opsin (L/LWS), bovine rhodopsin, and human short wave sensitive opsin (SWS1). Conserved Pro residues are highlighted in green, the conserved Ala residues in cyan, and the GWSRY domain in yellow.

```
M/LWS_construct      MAQQWSLQRLAGRHPQDSYEDSTQSSIFT-YTNSNST-RGPFEGPNYHIAPRWVYHLTSV
L/LWS                MAQQWSLQRLAGRHPQDSYEDSTQSSIFT-YTNSNST-RGPFEGPNYHIAPRWVYHLTSV
rhodopsin            ------------------MNGTEGPNFYVPFSNKTGVVRSPFEAPQYYLAEPWQFSMLAA
SWS1                 ------------------MRKMSEEEFYL-FKNISSV--GPWDGPQYHIAPVWAFYLQAA
                                        .::  :.* ...  .*::.*:*::*  * : : :.

M/LWS_construct      WMIFVVIASVFTNGLVLAATMKFKKLRHPLNWILVNLAVADLAETVIASTISVVNQVYGY
L/LWS                WMIFVVTASVFTNGLVLAATMKFKKLRHPLNWILVNLAVADLAETVIASTISIVNQVSGY
rhodopsin            YMFLLIMLGFPINFLTLYVTVQHKKLRTPLNYILLNLAVADLFMVFGGFTTTLYTSLHGY
SWS1                 FMGTVFLIGFPLNAMVLVATLRYKKLRQPLNYILVNVSFGGFLLCIFSVFPVFVASCNGY
                     :*  :.  ..  * :.* .*:..**** ***:**:*::...:   . .    .  .  **

M/LWS_construct      FVLGHPMCVLEGYTVSLCGITGLWSLAIISWERWMVVCKPFGNVRFDAKLAIVGIAFSWI
L/LWS                FVLGHPMCVLEGYTVSLCGITGLWSLAIISWERWLVVCKPFGNVRFDAKLAIVGIAFSWI
rhodopsin            FVFGPTGCNLEGFFATLGGEIALWSLVVLAIERYVVVCKPMSNFRFGENHAIMGVAFTWV
SWS1                 FVFGRHVCALEGFLGTVAGLVTGWSLAFLAFERYIVICKPFGNFRFSSKHALTVVLATWT
                     **:*   * ***:  :: *    ***..:: **::*:***:.*.**. : *:  :  :*

M/LWS_construct      WAAVWTAPPIFGWSRYWPHGLKTSCGPDVFSGSSYPGVQSYMIVLMVTCCITPLSIIVLC
L/LWS                WSAVWTAPPIFGWSRYWPHGLKTSCGPDVFSGSSYPGVQSYMIVLMVTCCIIPLAIIMLC
rhodopsin            MALACAAPPLVGWSRYIPEGMQCSCGIDYYTPHEETNNESFVIYMFVVHFIIPLIVIFFC
SWS1                 IGIGVSIPPFFGWSRFIPEGLQCSCGPDWYTVGTKYRSESYTWFLFIFCFIVPLSLICFS
                      .   : **:.****: * *:: *** * ::       :*:   :::   * ** :* :.

M/LWS_construct      YLQVWLAIRAVAKQQKESESTQKAEKEVTRMVVVMVLAFCFCWGPYAFFACFAAANPGYP
L/LWS                YLQVWLAIRAVAKQQKESESTQKAEKEVTRMVVVMIFAYCVCWGPYTFFACFAAANPGYA
rhodopsin            YGQLVFTVKEAAAQQQESATTQKAEKEVTRMVIIMVIAFLICWLPYAGVAFYIFTHQGSD
SWS1                 YTQLLRALKAVAAQQQESATTQKAEREVSRMVVVMVGSFCVCYVPYAAFAMYMVNNRNHG
                     * *:  ::. .* **:** :*****.**:***::*: :: .*: **: .* :   : .

M/LWS_construct      FHPLMAALPAFFAKSATIYNPVIYVFMNRQFRNCILQ-LF-GKK-VDDGSELSSASKTET
L/LWS                FHPLMAALPAYFAKSATIYNPVIYVFMNRQFRNCILQ-LF-GKK-VDDGSELSSASKTEV
rhodopsin            FGPIFMTIPAFFAKTSAVYNPVIYIMMNKQFRNCMVTTLCCGKNPLGDDEASTTVSKTET
SWS1                 LDLRLVTIPSFFSKSACIYNPIIYCFMNKQFQACIMK-MVCGKA-MTDESDTCSSQKTEV
                     :   : ::*::*:*:: :***:** :**.**. *::  :  **  : * .   : .***.

M/LWS_construct      -----SQVAPA
L/LWS                SSV--SSVSPA
rhodopsin            -----SQVAPA
SWS1                 STVSSTQVGPN
                          :.*.*
```

Excerpt of the sequence alignment containing the Pro-Pro motif. Alignment between the *Human* M/LWS receptor, *Human* long wavelength sensitive opsin (L/LWS), *Bovine* rhodopsin, *Human* short wave sensitive opsin (SWS1), *Human* opsin-5 (Opn5), *Danio rerio* red sensitive opsin 1 (lw1_dr), *Danio rerio* short wavelength sensitive opsin 2 (SWS2_dr), *Danio rerio* medium wavelength sensitive opsin (MW1_dr), *Todarodes pacificus* rhodopsin (rhodopsin_squid), and *Drosophila melanogaster* opsins 1-6 (rh1-6_dm). The conserved Pro residues are highlighted in green, the conserved Ala residues in cyan, and the GWSRY domain in yellow.

```
LW1_dr              AKWASAGIIFSWVWAAAWCAPPIFG-WSRYWPHGLKTSCGPDVFSGSEDPGVQSYMVVLM
M/LWS               AKLAIVGIAFSWIWAAVWTAPPIFG-WSRYWPHGLKTSCGPDVFSGSSYPGVQSYMIVLM
L/LWS               AKLAIVGIAFSWIWSAVWTAPPIFG-WSRYWPHGLKTSCGPDVFSGSSYPGVQSYMIVLM
SWS1                SKHALTVVLATWTIGIGVSIPPFFG-WSRFIPEGLQCSCGPDWYTVGTKYRSESYTWFLF
SWS2_dr             TPHAIAGCILPWCMALAAGLPPLLG-WSRYIPEGLQCSCGPDWYTTNNKFNNESYVMFLF
rhodopsin_bovine    ENHAIMGVAFTWVMALACAAPPLAG-WSRYIPEGLQCSCGIDYYTLKPEVNNESFVIYMF
MW1_dr              ANHAMAGIAFTWFMACSCAVPPLFG-WSRYLPEGMQTSCGPDYYTLNPEYNNESYVMYMF
Opn5                RKHAYICLAAIWAYASFWTTMPLVG-LGDYVPEPFGTSCTLDWWLAQASVGGQVFILNIL
rhodopsin_squid     HRRAFIMIIFVWLWSVLWAIGPIFG-WGAYTLEGVLCNCSFDYIS--RDSTTRSNILCMF
rh6_dm              ATAAVLRLMVVWTICGAWALMPLFG-WNRYVPEGNMTACGTDYFA--KDWWNRSYIIVYS
rh1_dm              IPLALGKIAYIWFMSSIWCLAPAFG-WSRYVPEGNLTSCGIDYLE--RDWNPRSYLIFYS
rh2_dm              IKTSIMKILFIWMMAVFWTVMPLIG-WSAYVPEGNLTACSIDYMT--RMWNPRSYLITYS
rh5_dm              YGQIVLLILFTWLWATPFSVLPLFQIWGRYQPEGFLTTCSFDYLT--NTDENRLFVRTIF
rh3_dm              HGKAIAMIIFIYMYATPWVVACYTETWGRFVPEGYLTSCTFDYLT--DNFDTRLFVACIF
rh4_dm              FTKAVIMNIIIWLYCTPWVVLPLTQFWDRFVPEGYLTSCSFDYLS--DNFDTRLFVGTIF
```

Quantitative RMSD analysis. The RMSD of ten energetically favored models was averaged and plotted by residue.

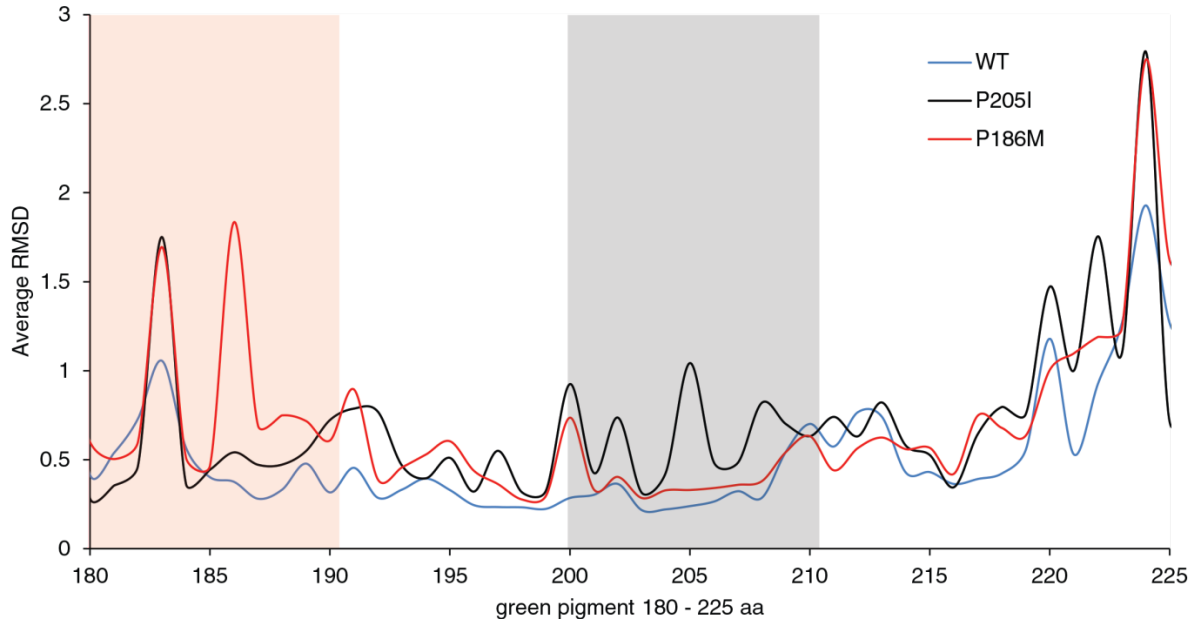

Figure S2. RMSD analysis by residue of green opsin and its P205I and P186M mutants. The RMSD of a residue is calculated relative to one structure in the ensemble of structures for a given opsin model. The RMSDs of a residue for all structures in an ensemble are then averaged. The average RMSD values are plotted against the amino acid residues 180 to 225. The average of the green cone pigment models is shown in blue, the mutant P205I and P186M are depicted in black and red, respectively. The highlighted area in red and black correspond to the red and black ovals in Figure 5.

Sequence alignment of M/LWS opsin and the 5-HT$_{1B}$, and 5-HT$_{2B}$ receptors. These sequences were aligned with the MUSCLE server provided by EMBL-EBI. [1-4] MUSCLE stands for MUltiple Sequence Comparison by Log-Expectation.

Conserved Pro residues are highlighted in green, the conserved Ala residues in cyan, and the GWSRY domain in yellow.

```
M/LWS           MAQQWSLQRLAGRHPQDSYEDS-TQSSIFTYTNSNSTRGPFEGPNYHIAPRWVYHLTSVW
5HT1B           ----MEEPGAQCAPPPPAGSETWVPQANLSSAPSQNCSAKDYIYQDSISLPWKVLLV-ML
5HT2B           MALSYRVSELQSTIPEHILQSTFVHVISSNWSGLQTESIPEEMKQIVEEQGNKLHWAALL
                             *    ..: .     . :  :.        :      .    . :

M/LWS           MIFVVIASVFTNGLVLAATMKFKKLRHPLNWILVNLAVADLAETVIASTISVVNQVY-GY
5HT1B           LALITLATTLSNAFVIATVYRTRKLHTPANYLIASLAVTDLLVSILVMPISTMYTVT-GR
5HT2B           ILMVIIPTIGGNTLVILAVSLEKKLQYATNYFLMSLAVADLLVGLFVMPIALLTIMFEAM
                : :: :.:   * :*: :.   .**. . *::: .***:**   ::. .*: :  :  .

M/LWS           FVLGHPMCVLEGYTVSLCGITGLWSLAIISWERWMVVCKPFGNVRFDA-KLAIVGIAFSW
5HT1B           WTLGQVVCDFWLSSDITCCTASILHLCVIALDRYWAITDAVEYSAKRTPKRAAVMIALVW
5HT2B           WPLPLVLCPAWLFLDVLFSTASIMHLCAISVDRYIAIKKPIQANQYNSRATAFIKITVVW
                : *   :*            :.:  *. *: :*: .: ...      :   * : *:. *

M/LWS           IWAAVWTAP-PIFGWSRYWPHGLKTSCGPDVFSGSSYPGVQSYMIVLMVTCCITPLSIIV
5HT1B           VFSISISLP-PFF-WRQAKAEEEVSEC---VVNTDHI----LYTVYSTVGAFYFP-TLLL
5HT2B           LISIGIAIPVPIKGIETDVDNPNNITC---VLTKERF---GDFMLFGSLAAFFTPLAIMI
                : :   : * *:             *   *.. .       : :   : .   * ::::

M/LWS           LCYL--------------------QVWLAIRAVAKQQ-----------------------
5HT1B           IALYGRIYVEARSRILKQTPNRTGKRLTRAQLITDSPGSTSSVTSINSRVPDVPSESGSP
5HT2B           VTYFLTIHALQKKAYLVKNKPPQRLTWLTVSTVFQRDETPCSSPEKVAMLDGSRKDKALP
                :                                 : .

M/LWS           ---------------KESESTQKAEKEVTRMVVVMVLAFCFCWGPYAFFA-CFAAANPGY
5HT1B           VYVNQVKVRVSDALLEKKKLMAARERKATKTLGIILGAFIVCWLPFFIISLVMPICKDAC
5HT2B           NSGDETLMRRTSTIGKKSVQTISNEQRASKVLGIVFFLFLLMWCPFFITNITLVLCDSCN
                               ::.      *. .:. : ::.  * . * *: :    :  ..

M/LWS           PFHP-LMAALPAFFAKSATIYNPVIYVFMNRQFRNCILQLFGKKVD--------------
5HT1B           WFHL-AIFDFFTWLGYLNSLINPIIYTMSNEDFKQA----FHKLI---------------
5HT2B           QTTLQMLLEIFVWIGYVSSGVNPLVYTLFNKTFRDA----FGRYITCNYRATKSVKTLRK
                      :  : .::.   :  **::*.: *  *.:.    * . :
```

```
M/LWS           --------------------------DGSELSSASKTEVSSVSSVSPA-----------
5HT1B           -----------------------------------RFKCTS-----------------
5HT2B           RSSKIYFRNPMAENSKFFKKHGIRNGINPAMYQSPMRLRSSTIQSSSIILLDTLLLTENE
                                                  .   ::

M/LWS           -----------
5HT1B           -----------
5HT2B           GDKTEEQVSYV
```

Non-overlapping local alignments of M/LWS opsin and 5-HT$_{1B}$ protein sequences. LALIGN finds internal duplications by calculating non-intersecting local alignments of protein sequences [3, 4]. The Pro residues 186-187 and 183-184, corresponding to the M/LWS opsin and 5HT$_{1B}$ receptor respectively, are highlighted in green.

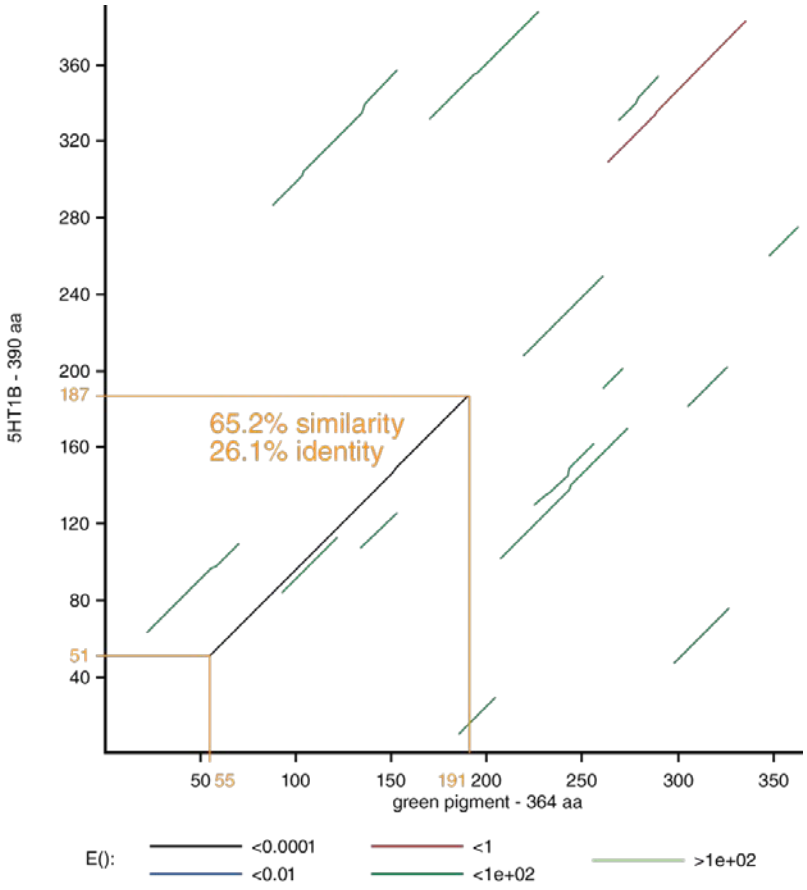

Figure S3. Non-overlapping local alignments of the green pigment and 5-HT$_{1B}$ protein sequences. The amino acid residues of interest, P186 and P187, are found in the local alignment of green opsin 55-191 aa and 5-HT$_{1B}$ 51-187 aa, highlighted in orange. The similarity score is 65.2% and the sequence shares 26.1% identity with an E(1) < 7.7e-13.

SI References

[1] Edgar, R. C. (2004) MUSCLE: a multiple sequence alignment method with reduced time and space complexity, *BMC Bioinf. 5*, 113.

[2] Edgar, R. C. (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput, *Nucleic Acids Res. 32*, 1792-1797.

[3] Li, W., Cowley, A., Uludag, M., Gur, T., McWilliam, H., Squizzato, S., Park, Y. M., Buso, N., and Lopez, R. (2015) The EMBL-EBI bioinformatics web and programmatic tools framework, *Nucleic Acids Res. 43*, W580-584.

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