Supplementary Materials for

Generation of bright monomeric red fluorescent proteins via computational design of enhanced chromophore packing

Sandrine Legault,[1] Derek P. Fraser-Halberg,[1] Ralph L. McAnelly,[2] Matthew G. Eason,[1] Michael C. Thompson[2] & Roberto A. Chica[1,*]

[1] Department of Chemistry and Biomolecular Sciences, University of Ottawa, 10 Marie-Curie, Ottawa, Ontario K1N 6N5, Canada

[2] Department of Chemistry and Biochemistry, University of California, Merced, Merced, California 95343, United States

* To whom correspondence should be addressed:

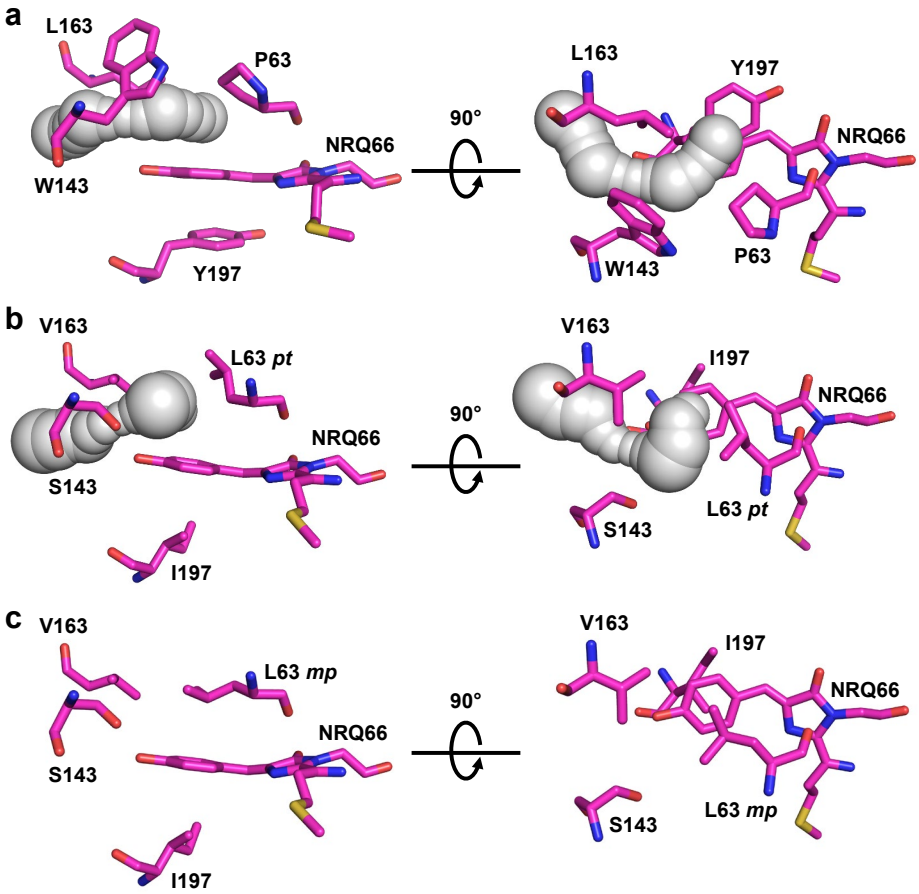
Roberto A. Chica, Email: rchica@uottawa.ca

This file includes:

Supplementary Table 1. Amino-acid sequences of red fluorescent proteins

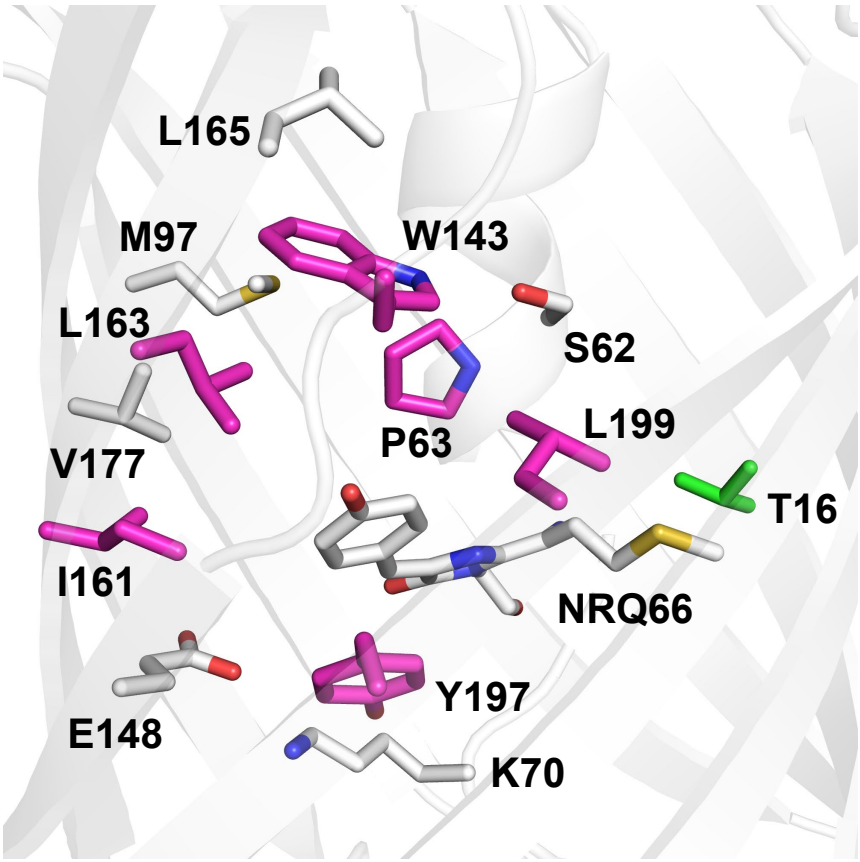
Protein	Sequence [a]
mRojoA	MGHHHHHHGVSKGEEDNMAIIKEFMRFKTHMEGSVNGHEFEIEGEGEGRPYEGTQTAKLKVTKG GPLPFAWDILSPQFMYGSKAYVKHPADIPDYLKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQD GEFIYKVKLHGTNFPSDGPVMQKKTMGWEASSERMYPEDGALKGEIKLRLKLKDGGHYDAEVKT TYKAKKPVQLPGAYNANYKLDITSHNEDYTIVEQYERCEGRHSTGGMDELYK
mRojo-Y197I	MGHHHHHHGVSKGEEDNMAIIKEFMRFKTHMEGSVNGHEFEIEGEGEGRPYEGTQTAKLKVTKG GPLPFAWDILSPQFMYGSKAYVKHPADIPDYLKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQD GEFIYKVKLHGTNFPSDGPVMQKKTMGWEASSERMYPEDGALKGEIKLRLKLKDGGHYDAEVKT TYKAKKPVQLPGAYNAN**I**KLDITSHNEDYTIVEQYERCEGRHSTGGMDELYK
mSandy0.1	MGHHHHHHGVSKGEEDNMAIIKEFMRFK**V**HMEGSVNGHEFEIEGEGEGRPYEGTQTAKLKVTKG GPLPFAWDILS**L**QFMYGSKAYVKHPADIPDYLKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQD GEFIYKVKL**R**GTNFPSDGPVMQKKTMG**S**EASSERMYPEDGALKGE**V**KLRLKLKDGGHYDAEVKT TYKAKKPVQLPGAYNAN**V**KLDITSHNEDYTIVEQYERCEGRHSTGGMDELYK
mSandy0.7	MGHHHHHHGVSKGEEDNMAIIKEFMRFK**V**HMEGSVNGHEFEIEGEGEGRPYEGTQTAKLKVTKG GPLPFAWDILS**L**QFMYGSKAYVKHPADIPDYLKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQD GEFIYKVKL**R**GTNFPSDGPVMQKKTMG**S**EASSERMYPEDGALKGEIKLRLKLKDGGHYDAEVKT TYKAKKPVQLPGAYNAN**V**KLDITSHNEDYTIVEQYERCEGRHSTGGMDELYK
mSandy0.9	MGHHHHHHGVSKGEEDNMAIIKEFMRFK**V**HMEGSVNGHEFEIEGEGEGRPYEGTQTAKLKVTKG GPLPFAWDILS**L**QFMYGSKAYVKHPADIPDYLKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQD GEFIYKVKL**R**GTNFPSDGPVMQKKTMG**S**EASSERMYPEDGALKGE**V**K**V**RLKLKDGGHYDAEVKT TYKAKKPVQLPGAYNAN**V**KLDITSHNEDYTIVEQYERCEGRHSTGGMDELYK
mSandy1	MGHHHHHHGVSKGEEDNMAIIKEFMRFK**V**HMEGSVNGHEFEIEGEGEGRPYEGTQTAKLKVTKG GPLPFAWDILS**L**QFMYGSKAYVKHPADIPDYLKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQD GEFIYKVKL**R**GTNFPSDGPVMQKKTMG**S**EASSERMYPEDGALKGE**V**KLRLKLKDGGHYDAEVKT TYKAKKPVQLPGAYNAN**I**KLDITSHNEDYTIVEQYERCEGRHSTGGMDELYK
mSandy2	MGHHHHHHGVSKGEEDNMAIIKEFMRFK**V**HMEGSVNGHEFEIEGEGEGRPYEGTQTAKLKVTKG GPLPFAWDILS**L**QFMYGSKAYVKHPADIPDYLKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQD GEFIYKVKLHGTNFPSDGPVMQKKTMG**S**EASSERMYPEDGALKGE**V**K**V**R**F**KLKDGGHY**V**AEVKT TYKAKKPVQLPGAY**Y**AN**I**K**M**DITSHNEDYTIVEQYERCEGRHSTGGMDELYK
mSandy1-L163V/L199M	MGHHHHHHGVSKGEEDNMAIIKEFMRFK**V**HMEGSVNGHEFEIEGEGEGRPYEGTQTAKLKVTKG GPLPFAWDILS**L**QFMYGSKAYVKHPADIPDYLKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQD GEFIYKVKL**R**GTNFPSDGPVMQKKTMG**S**EASSERMYPEDGALKGE**V**K**V**RLKLKDGGHYDAEVKT TYKAKKPVQLPGAYNAN**I**K**M**DITSHNEDYTIVEQYERCEGRHSTGGMDELYK
mCherry	MGHHHHHHGVSKGEEDNMAIIKEFMRFK**V**HMEGSVNGHEFEIEGEGEGRPYEGTQTAKLKVTKG GPLPFAWDILSPQFMYGSKAYVKHPADIPDYLKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQD GEFIYKVKL**R**GTNFPSDGPVMQKKTMGWEASSERMYPEDGALKGEIK**Q**RLKLKDGGHYDAEVKT TYKAKKPVQLPGAYN**V**N**I**KLDITSHNEDYTIVEQYER**A**EGRHSTGGMDELYK

[a] Mutations from mRojoA are underlined and highlighted in bold.



Supplementary Figure 1. Cavity within the chromophore pocket. (a) The chromophore (NRQ) pocket in mRojoA (chain A, PDB ID: 3NEZ) contains a small cavity above the chromophore *p*-hydroxybenzylidene moiety that is part of a tunnel extending to the protein surface (indicated by white spheres). (b) This cavity is also present in mSandy2 (chain H, PDB ID: 7RY2) when L63 adopts the *pt* (χ_1 = gauche$^{(+)}$, χ_2 = trans) rotamer, although the tunnel's shape is different. Tunnels and cavities were identified with the Caver3.0 software using a probe radius of 0.8 Å. (c) Caver3.0 was unable to detect a cavity or tunnel next to the chromophore *p*-hydroxybenzylidene in mSandy2 when L63 adopts the alternate *mp* (χ_1 = gauche$^{(-)}$, χ_2 = gauche$^{(+)}$) rotamer. All images generated using PyMOL 2.5.0.

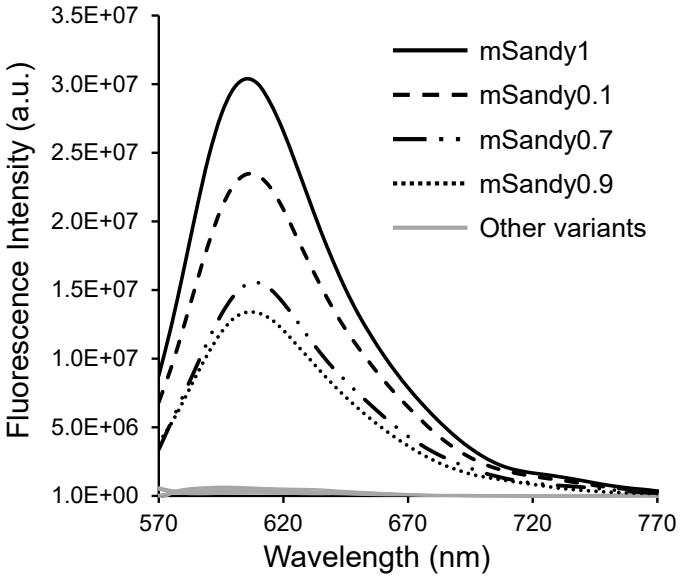
a



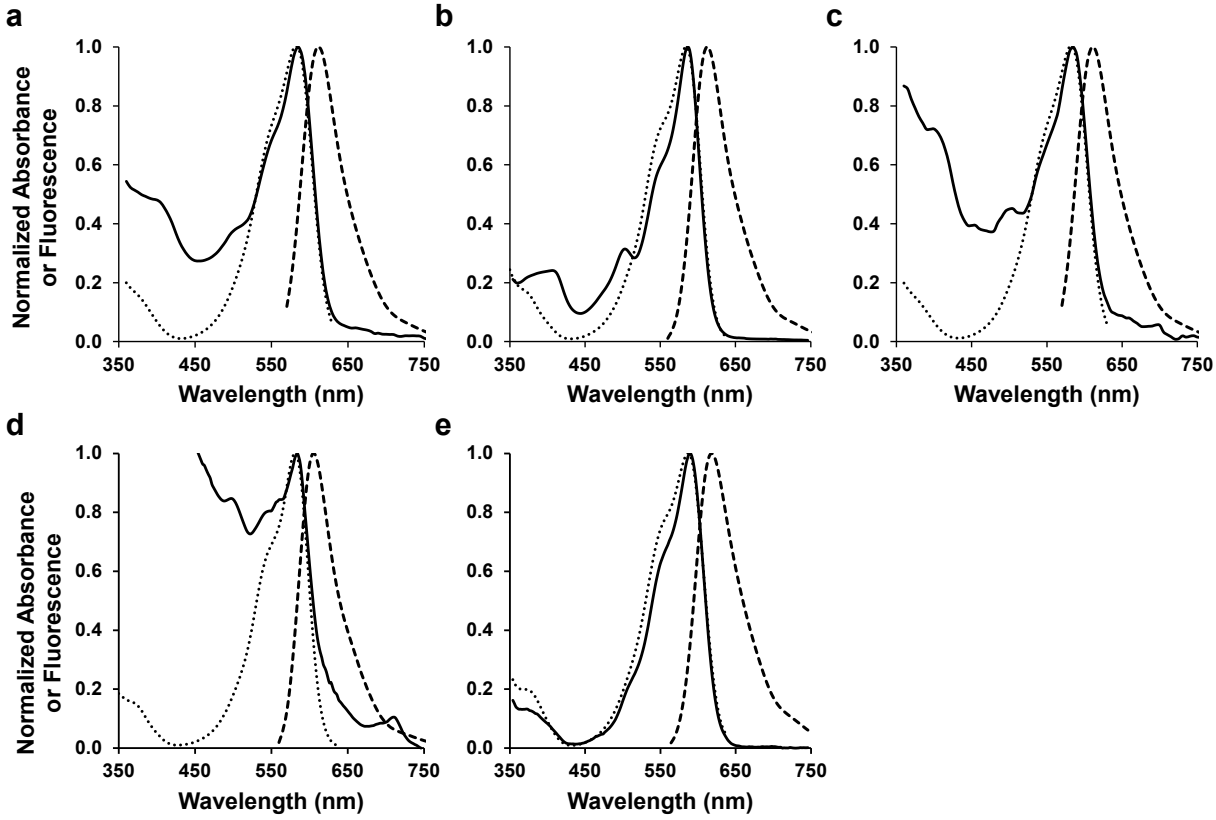
b

		Positions													
Seq	Energy	16	62	63	66	70	97	143	148	161	163	165	177	197	199
WT		T	S	P	NRQ	K	M	W	E	I	L	L	V	Y	L
0	-2342.436566	V	-	L	*	-	-	S	-	V	-	-	-	V	I
1	-2341.211358	V	-	L	*	-	-	S	-	V	-	-	-	V	-
2	-2340.200948	V	-	L	*	-	-	S	-	V	-	-	-	I	I
3	-2338.936610	V	-	L	*	-	-	S	-	V	-	-	-	I	-
4	-2338.836003	V	-	L	*	-	-	S	-	V	V	-	-	V	I
5	-2338.634281	V	-	L	*	-	-	S	-	V	-	-	-	V	V
6	-2338.172567	V	-	L	*	-	-	S	-	-	-	-	-	V	I
7	-2336.947382	V	-	L	*	-	-	S	-	-	-	-	-	V	-
8	-2336.675792	V	-	L	*	-	-	S	-	V	V	-	-	I	I
9	-2336.586179	V	-	L	*	-	-	S	-	V	V	-	-	V	-

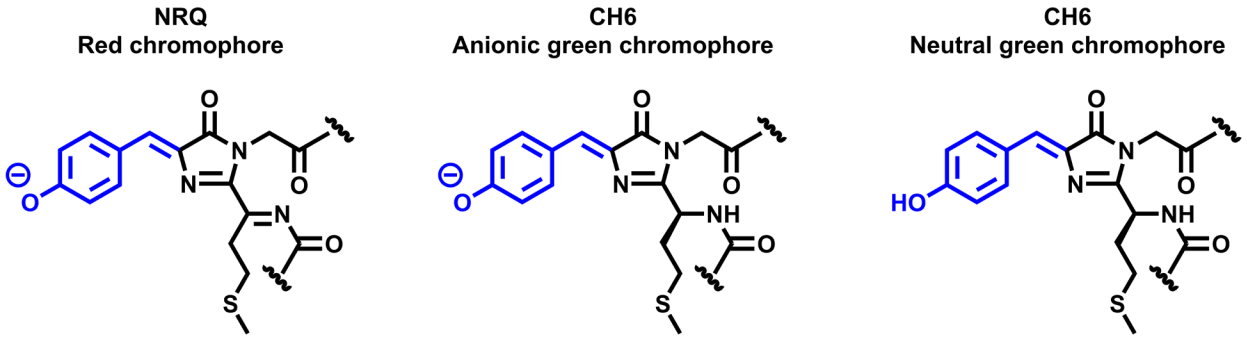
Supplementary Figure 2. Computational design of mSandy1. (a) Structure of mRojoA (PDB ID: 3NEZ) with designed residues shown as sticks and identified by their one-letter code. NRQ indicates the chromophore. Residues whose identify and conformation were allowed to change are colored magenta, while those whose conformation but not identity was allowed to change are colored white. The T16V mutation introduced to facilitate chromophore maturation is colored green. (b) Multistate design results. Asterisks, letters and dashes indicate the chromophore, mutations, or wild-type residues, respectively. The top 10 sequences, ranked according to their Boltzmann-weighted average energy (kcal mol^{-1}) across the four-template ensemble, were experimentally characterized. Sequence 3 corresponds to mSandy1.



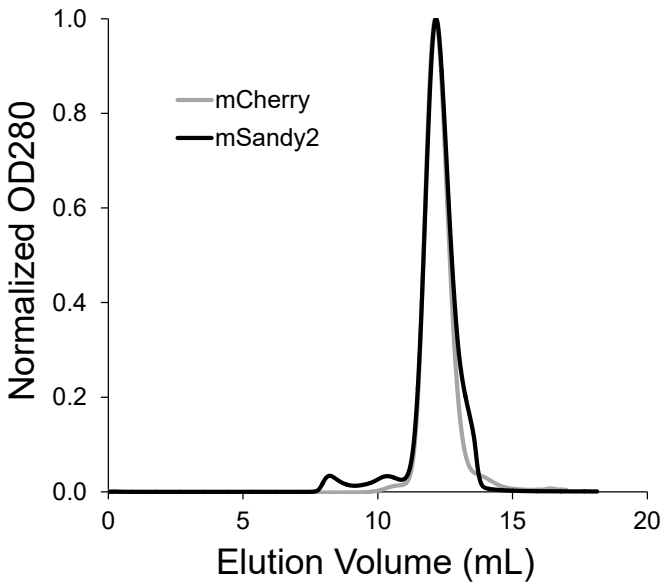
Supplementary Figure 3. Screening of designed sequences for bright fluorescence. Emission spectra (λ_{ex} = 535 nm) of purified samples of the top 10 designed sequences revealed that only four of these variants displayed high fluorescence.



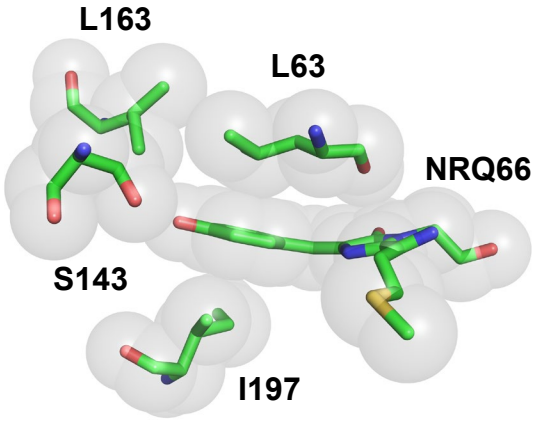
Supplementary Figure 4. Fluorescence and absorption spectra of RFP variants. (a) mSandy0.1, (b) mSandy0.7, (c) mSandy0.9, (d) mSandy1-L163V/L199M, and (e) mRojoA-Y197I. Absorption, excitation, and emission spectra are shown as solid, dotted, and dashed lines, respectively. Maturation of all mSandy variants shown here is inefficient, as evidenced by the presence of absorption peaks at approximately 390 nm and 510 nm corresponding to the neutral and anionic green chromophores, respectively, and/or relatively low absorption peaks at approximately 590 nm, which indicates that a low proportion of RFP molecules in the population of molecules in solution contain a mature red chromophore.



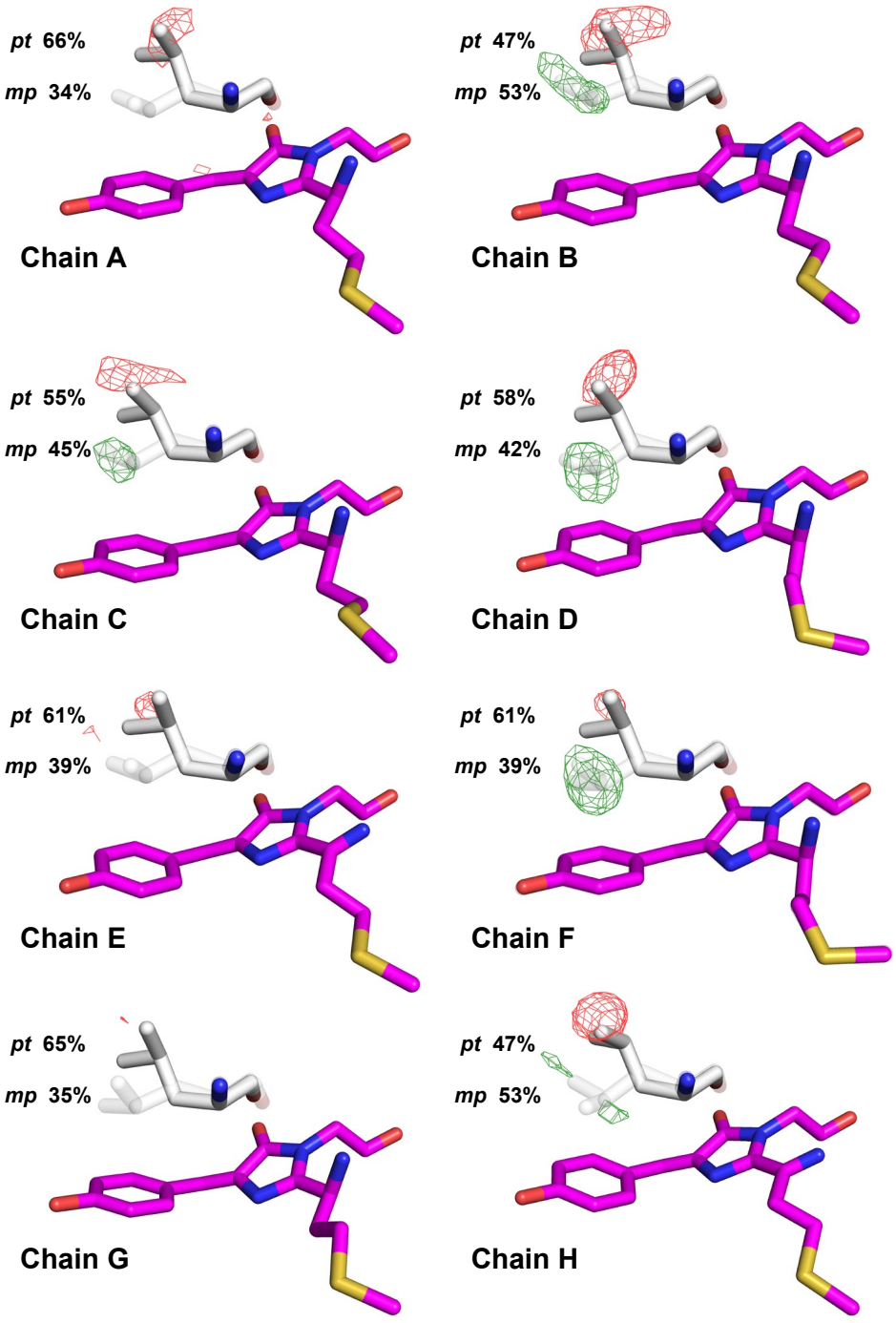
Supplementary Figure 5. Chromophore structure. The red chromophore differs from the green one by the presence on an *N*-acylimine group that extends the size of the conjugated system. Red, anionic green, and neutral green chromophores absorb at approximately 590 nm, 510 nm, and 390 nm, respectively. The *p*-hydroxybenzylidene moiety of each chromophore is colored blue.



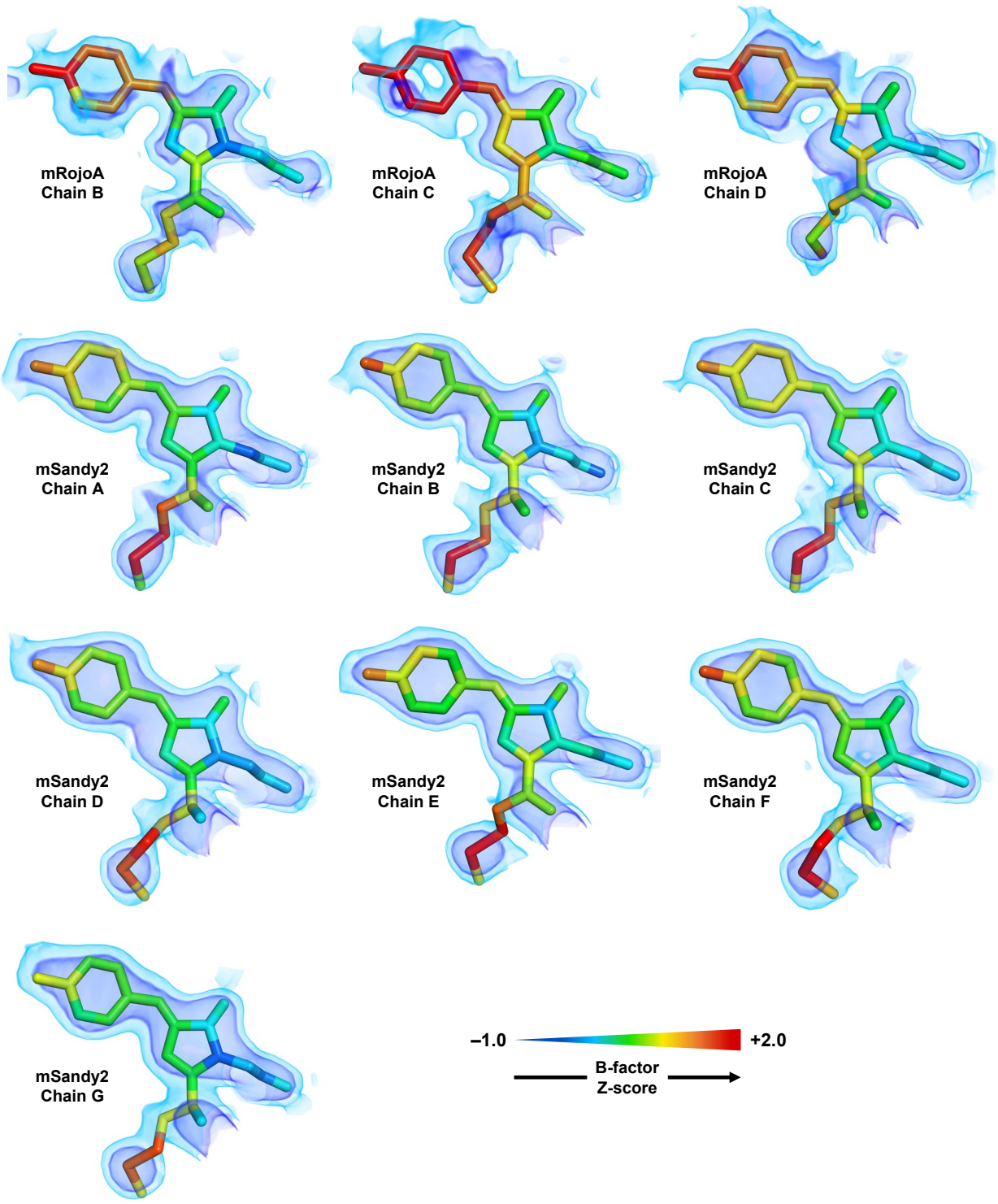
Supplementary Figure 6. mSandy2 is a monomeric red fluorescent protein. Size-exclusion chromatography elution profiles monitored by absorbance at 280 nm during the final purification step using an ÄKTA Pure system equipped with a Superdex 75 column (GE Healthcare) show a single predominant peak for mSandy2, which elutes at the same volume as mCherry, a known monomeric red fluorescent protein.



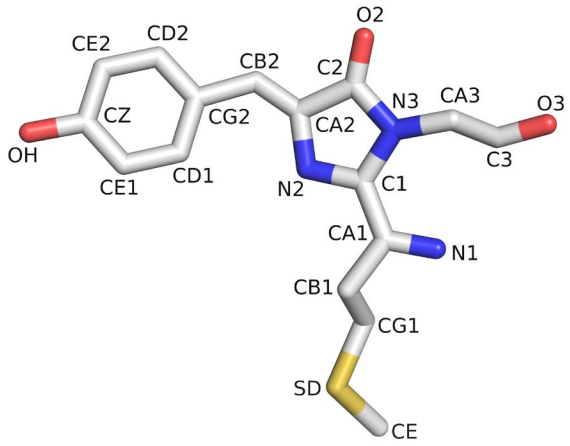
Supplementary Figure 7. Computational model of mSandy1. Key residues forming part the chromophore (NRQ66) pocket are shown as sticks and transparent spheres. The designed Leu63 residue adopts the *mp* rotamer (χ_1 = gauche$^{(-)}$, χ_2 = gauche$^{(+)}$) and tightly packs against the chromophore *p*-hydroxybenzylidene moiety. Image generated using PyMOL 2.5.0.



Supplementary Figure 8. Difference density increases around the Leu63 side chain after refinement in the absence of the *mp* conformer. The Leu63 side chain (white) adopts two conformations: *mp* (χ_1 = gauche$^{(-)}$, χ_2 = gauche$^{(+)}$) and *pt* (χ_1 = gauche$^{(+)}$, χ_2 = trans). To confirm the presence of the alternate conformation, the structure was re-refined in the absence of the *mp* conformer (transparent sticks), resulting in difference density (contoured at 3σ, positive and negative difference densities are colored green and red, respectively). The occupancies of each conformer are indicated in percentage. The chromophore is colored magenta.



Supplementary Figure 9. Chromophore rigidity. The chromophores of mRojoA (PDB ID: 3NEZ) and mSandy2 (PDB ID: 7RY2) are shown as sticks with atoms colored according to their B-factor Z-score. The 2Fo-Fc map is shown in volume representation at two contour levels: 0.5 and 1.5 eÅ^{-3} in light and dark blue, respectively. Chain A of mRojoA and chain H of mSandy2 are shown on Figure 4 (main text).



Supplementary Figure 10. Chromophore atom names. The methine bridge is formed by atoms CG2, CB2, and CA2.