

Structure–Function Dataset Reveals Environment Effects within a  
Fluorescent Protein Model System

Peer-reviewed author version

De Zitter, Elke; Hugelier, Siewert; DUWE, Sam; Vandenberg, Wim; Tebo, Alison G.;  
Van Meervelt, Luc & Dedecker, Peter (2021) Structure–Function Dataset Reveals  
Environment Effects within a Fluorescent Protein Model System. In: ANGEWANDTE  
CHEMIE-INTERNATIONAL EDITION, 60 (18), p. 10073-10081.

DOI: 10.1002/anie.202015201

Handle: <http://hdl.handle.net/1942/33890>

# Structure-Function Dataset Reveals Environment Effects within a Fluorescent Protein Model System

Elke De Zitter<sup>1‡</sup>, Siewert Hugelier<sup>1</sup>, Sam Duwé<sup>1†</sup>, Wim Vandenberg<sup>1</sup>,  
Alison G. Tebo<sup>2</sup>, Luc Van Meervelt<sup>1</sup>, Peter Dedecker<sup>1\*</sup>

<sup>1</sup>*Department of Chemistry, KU Leuven, Celestijnenlaan 200G - box 2403, 3001 Leuven, Belgium*

<sup>2</sup>*Janelia Research Campus, Howard Hughes Medical Institute, 19700 Helix Drive, Ashburn, Virginia 20147, USA*

<sup>‡</sup>*Present address: University Grenoble Alpes, CEA, CNRS, IBS, 71 Avenue des Martyrs, 38000 Grenoble, France.*

<sup>†</sup>*Present address: Advanced Optical Microscopy Centre, Hasselt University, Agoralaan building C, 3590 Diepenbeek, Belgium*

\*peter.dedecker@kuleuven.be, <https://www.chem.kuleuven.be/pd>

February 2, 2021

## Abstract

Anisotropic environments can drastically alter the spectroscopy and photochemistry of molecules, leading to complex structure-function relationships. We examined this using fluorescent proteins as easy-to-modify model systems. Starting from a single scaffold, we have developed a range of 27 photochromic fluorescent proteins that cover a broad range of spectroscopic properties, including the determination of 43 crystal structures. Correlation and principal component analysis confirmed the complex relationship between structure and spectroscopy, but also allowed us to identify consistent trends and to relate these to the spatial organization. We find that changes in spectroscopic properties can come about through multiple underlying mechanisms, of which polarity, hydrogen bonding and presence of water molecules are key modulators. We anticipate that our findings and rich structure/spectroscopy dataset can open opportunities for the development and evaluation of new and existing protein engineering methods.

## Introduction

Highly structured and anisotropic environments can profoundly alter the (photo-)chemistry of embedded molecules, which is crucial for many essential processes such as enzymatic catalysis.<sup>[1]</sup> However, the mechanistic link between environment and properties can be difficult to assess due to the intricate effects and the challenges in generating well-defined three-dimensional environments. Fluorescent proteins (FPs) present excellent model systems to examine this link on account of their well-defined structures and the ease with which they can be engineered. Indeed, a wide variety of fluorescent proteins with very different spectroscopic properties is readily available, which is particularly remarkable when considering that only a handful of different chromophore structures have been identified.<sup>[2]</sup> This variety is not just associated with changes in spectra or brightness, but also with the presence of photochemical reactions such as light-induced fluorescence activation or photochromism, the reversible switching between a fluorescent and a non-fluorescent state.<sup>[3]</sup> Beyond the many applications of fluorescent proteins in conventional imaging, such photochemical transformations form the basis of many forms of advanced fluorescence microscopy.

The many applications of these proteins have led to extensive investigations both to generate improved labels and to better understand their properties.<sup>[4–12]</sup> These studies have revealed a very rich photochemistry involving multiple spectroscopic states, reflecting their intricate structure-function

relationship in which mutating just a single residue can lead to profound changes in the spectroscopic properties.<sup>[13–15]</sup> While a fascinating range of insights has been gathered, it remains difficult to rationally develop variants with properties tailored to particular applications. This reflects not only the complex modulation of the properties by the anisotropic environment, but also the diversity of the fluorescent protein family that makes it difficult to apply insights uncovered for one variant to other labels. This broad variability also limits efforts to apply innovative new methodologies, such as machine learning, to FP engineering.

Photochromic fluorescent proteins are excellent candidates for investigations on the chromophore-environment relationship since they display both an intrinsic spectroscopic complexity as well as a pronounced and reversible phototransformation. It is known that this process involves a chromophore *cis/trans* isomerization<sup>[16]</sup> and (de-)protonation,<sup>[4,17,18]</sup> and requires an environment capable of both allowing this isomerization and of stabilizing the resulting states.<sup>[14,19–21]</sup>

In this contribution we aim to broaden our understanding of this structure-function relationship by generating a large structural and spectroscopic dataset of photochromic FPs derived from just a single protein scaffold, rsGreen0.7.<sup>[19]</sup> We selected 27 mutants that displayed a large variety in spectroscopic properties yet differed in only one or two amino acid mutations, aiming to realize a large diversity without focusing on applicability for a particular use case. We characterized these mutants with respect to 19 different spectroscopic parameters, and also obtained 43 crystal structures of the fluorescent and non-fluorescent states. We identified correlations between the different spectroscopic parameters and/or structural properties and established a structural picture of the spectroscopic variation. Overall, our results confirm that the spectroscopic properties of fluorescent proteins display a variety of correlations, making it difficult to *a priori* predict a mutational outcome. Nevertheless, we do identify several fundamental relationships that can ease the design process and that provide insights into the effect of the chromophore environment. We also identify additional insights such as the importance of hydrogen bonding in photochromism but also that the presence of water molecules in the hydrogen bonding network is highly important. Our extensive structure-function data not only leads to the conclusions presented here, but also provides a rich dataset that can be leveraged for future investigations into this and other classes of systems.

## Results and discussion

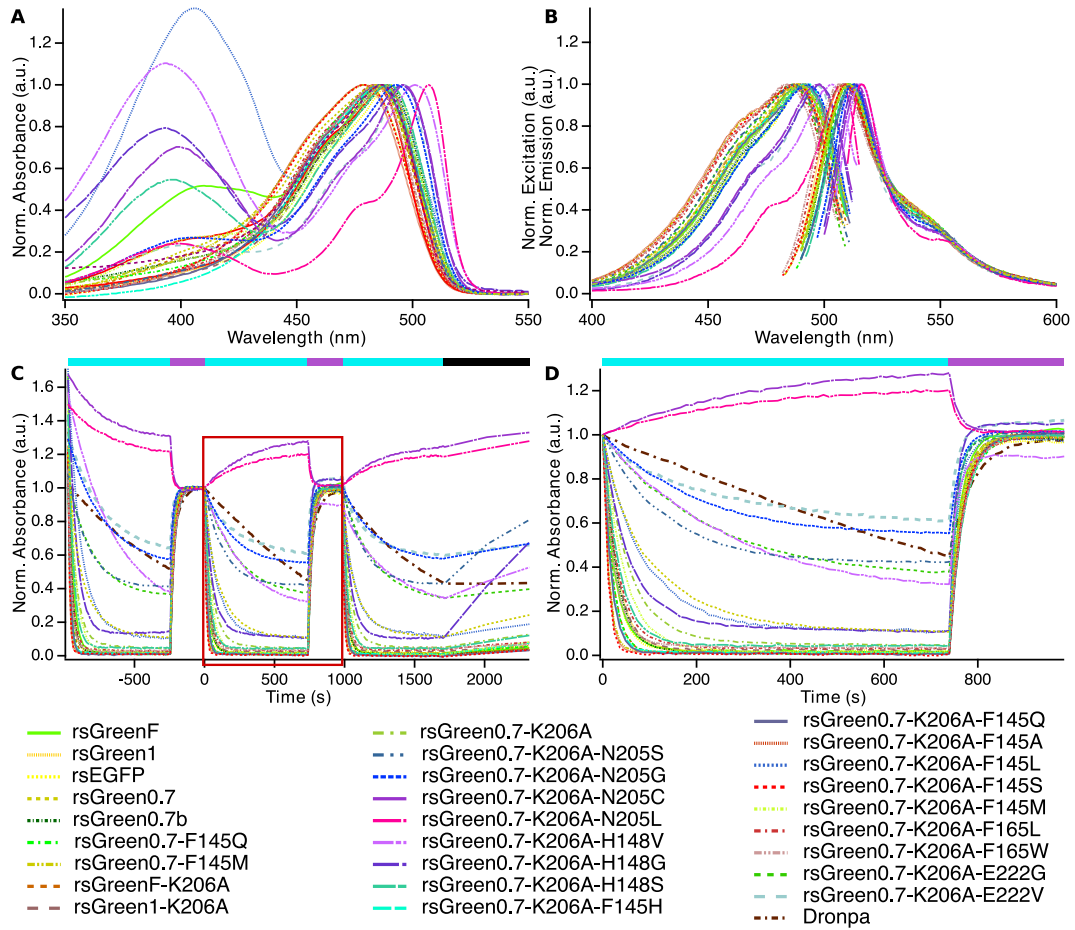
### Generation of a spectroscopically diverse set of mutants

We based our work on the photochromic green fluorescent rsGreen0.7, an EGFP mutant that has been used in a range of imaging experiments.<sup>[19]</sup> Using its crystal structures and our experience with rsGreen-mutagenesis, we identified a number of amino acids in the chromophore pocket (Phe145, His148, Phe165, Asn205 and Glu222) that were likely to give rise to a variation in spectroscopic properties, and generated mutants by subjecting these sites to saturation mutagenesis. We also introduced the K206A mutation, which enhances the tendency for protein dimerization,<sup>[22]</sup> to increase the likelihood of crystallogenesis. Mutants that clearly differed in spectroscopic properties from rsGreen0.7 and each other were retained for purification and *in vitro* assessment, without regards to usefulness for a particular purpose apart from the presence of a measurable fluorescence.

All purified fluorescent proteins underwent an extensive spectroscopic characterization (Table 1 and Figure S1). Figure 1 highlights a few of these measurements, including absorption and excitation/emission spectra at pH 7.4 (Figure 1A,B), and the evolution of absorption and emission during photoswitching (Figure 1C,D). The behavior of Dronpa,<sup>[23]</sup> one of the first photoswitchable proteins identified, is also shown for comparison.

### Exploratory analysis of the spectroscopic data

Correlation analysis can identify those parameters that vary in concert due to similar dependencies. We therefore calculated Pearson correlation coefficients over our data (Table 2), which allowed us to identify several correlated parameters that hint to shared underlying mechanisms, while providing



**Figure 1:** (A-B) Normalized absorption (A), excitation and emission (B) spectra of rsEGFP and the rsGreen-variants at pH 7.4. (C) Normalized time evolution of the anionic absorption band during illumination with 488-nm light (cyan bar), 405-nm light (violet bar) and a thermal recovery period (black bar). Dronpa was added as a reference and is depicted in bold. (D) Zoomed-in region from the second cycle that was used to calculate the dynamic spectroscopic properties except for  $A_{\text{unrecov}}$  and  $k_{\text{therm}}$ .



**Table 1:** Spectroscopic properties of rsEGFP, rsGreen1, rsGreenF, rsGreen0.7 and rsGreen0.7-mutants.

Protein	$\lambda_{\text{abs}}$ (nm)	$\lambda_{\text{ex}}$ (nm)	$\lambda_{\text{em}}$ (nm)	$\lambda_{\text{off}}^{\text{abs}}$ (nm)	$\delta_{\text{Stokes}}$ (nm)	$\lambda_{\text{ex}} - \lambda_{\text{abs}}$ (nm)	$\epsilon_{\text{overall}}$ (mM <sup>-1</sup> cm <sup>-1</sup> )	$\epsilon_{\text{anionic}}$ (mM <sup>-1</sup> cm <sup>-1</sup> )	$\epsilon_{\text{off}}$ (mM <sup>-1</sup> cm <sup>-1</sup> )
rsGreenF	486	489 [485]	511 [511]	391 [390]	22	3	35 [41]	65 [52]	30 [18]
rsGreen1	486	488 [486]	510 [509]	397 [396]	22	2	62 [58]	71 [61]	26 [20]
rsEGFP	489	491 [491]	510 [510]	397 [396]	19	2	58 [46,47]	72 [57]	28 [18]
rsGreen0.7	484	489 [487]	510 [511]	397 [396]	21	5	46 [48]	65 [62]	27 [21]
rsGreen0.7b	489	491 [489]	511 [511]	396 [396]	20	2	55 [44]	70 [53]	26 [20]
rsGreen0.7-F145Q	484	491	512	393	21	7	53	61	21
rsGreen0.7-F145M	479	488	510	396	22	9	50	59	25
rsGreenF-K206A	485	490	509	390	19	5	52	70	27
rsGreen1-K206A	487	487	508	396	21	0	75	72	33
rsGreen0.7-K206A	488	489	510	399	21	1	56	67	29
rsGreen0.7-K206A-N205S	486	487	513	414	26	1	50	59	19
rsGreen0.7-K206A-N205G	493	493	513	410	20	0	52	74	33
rsGreen0.7-K206A-N205C	496	498	514	496	16	2	21	50	NA
rsGreen0.7-K206A-N205L	507	508	516	507	8	1	64	104	NA
rsGreen0.7-K206A-H148V	501	503	516	394	13	2	25	67	38
rsGreen0.7-K206A-H148G	496	498	514	400	16	2	30	64	40
rsGreen0.7-K206A-H148S	490	491	511	399	20	1	29	62	29
rsGreen0.7-K206A-F145H	485	494	511	395	17	9	59	66	24
rsGreen0.7-K206A-F145Q	485	492	513	395	21	7	55	61	23
rsGreen0.7-K206A-F145A	478	487	509	397	22	9	53	58	23
rsGreen0.7-K206A-F145L	487	491	513	393	22	4	19	60	35
rsGreen0.7-K206A-F145S	479	488	509	394	21	9	53	58	22
rsGreen0.7-K206A-F145M	488	492	510	393	18	4	46	59	26
rsGreen0.7-K206A-F165L	483	486	509	393	23	3	55	64	23
rsGreen0.7-K206A-F165W	484	484	506	398	22	0	54	64	27
rsGreen0.7-K206A-E222G	487	487	511	412	24	0	68	71	29
rsGreen0.7-K206A-E222V	496	497	511	403	14	1	56	87	34

Protein	pK <sub>a,app</sub>	$\lambda_{\text{abs,pH10}} - \lambda_{\text{abs,pH6.5}}$ (nm)	Q <sub>fluo</sub> (%)	Molecular brightness	Q <sub>off</sub> (10 <sup>-1</sup> %)	Q <sub>on</sub> (%)	A <sub>res</sub> (%)	A <sub>res,∞</sub> (%)	k <sub>therm</sub> (10 <sup>-3</sup> s <sup>-1</sup> )	A <sub>unrecov</sub> (%)
rsGreenF	7.2 [6.7]	2	39 [39]	14	7.3 [8.7]	12.7 [1.8]	1.7	1.6	0.5	40.0
rsGreen1	6.2 [5.9]	0	43 [42]	27	3.7 [4.2]	14.0 [1.4]	1.2	1.0	0.4	8.6
rsEGFP	6.6 [6.7,6.5]	2	42 [42, 36]	24	4.5 [5.5]	14.6 [1.7]	2.0	1.9	0.6	15.8
rsGreen0.7	6.9 [6.7]	16	30 [40]	14	1.3 [2.6]	14.6 [1.2]	10.7	8.2	2.6	16.5
rsGreen0.7b	6.8 [6.7]	6	38 [43]	21	4.1 [5.3]	13.9 [1.6]	1.3	1.2	0.4	18.7
rsGreen0.7-F145Q	6.3	0	32	17	6.7	13.8	0.9	0.7	0.8	8.0
rsGreen0.7-F145M	6.5	2	23	12	12.8	14.0	1.5	1.4	0.8	16.8
rsGreenF-K206A	6.8	0	43	22	6.9	14.5	1.0	1.0	0.4	25.1
rsGreen1-K206A	6.3	0	47	35	3.0	11.6	3.0	2.9	0.5	9.5
rsGreen0.7-K206A	6.9	2	39	22	1.9	12.2	4.8	4.0	1.4	18.4
rsGreen0.7-K206A-N205S	6.6	0	49	25	1.7	17.9	42.4	25.5	17.8	7.7
rsGreen0.7-K206A-N205G	7.1	2	43	22	0.4	10.4	55.3	51.9	3.7	22.7
rsGreen0.7-K206A-N205C	7.7	0	22	5	NA	NA	NA	NA	NA	NA
rsGreen0.7-K206A-N205L	7.4	0	30	19	NA	NA	NA	NA	NA	NA
rsGreen0.7-K206A-H148V	7.7	2	33	8	1.1	13.4	32.3	18.1	6.7	36.3
rsGreen0.7-K206A-H148G	7.8	4	40	12	3.9	14.2	11.2	2.6	15.3	39.2
rsGreen0.7-K206A-H148S	7.5	2	51	15	6.1	16.9	4.4	3.9	1.4	30.5
rsGreen0.7-K206A-F145H	6.1	-2	72	42	6.3	13.4	1.0	0.8	1.1	8.1
rsGreen0.7-K206A-F145Q	6.2	0	39	21	6.9	13.8	0.6	0.5	0.5	10.8
rsGreen0.7-K206A-F145A	6.2	0	35	19	13.2	13.2	1.1	1.1	0.5	13.3
rsGreen0.7-K206A-F145L	7.8	2	50	9	3.6	13.4	11.1	9.9	1.5	41.5
rsGreen0.7-K206A-F145S	6.5	2	27	14	13.5	12.0	0.4	0.3	0.7	17.7
rsGreen0.7-K206A-F145M	7.1	0	56	26	3.3	14.0	1.7	1.1	1.2	31.5
rsGreen0.7-K206A-F165L	6.6	2	28	16	4.0	13.3	2.9	2.6	0.8	11.9
rsGreen0.7-K206A-F165W	6.5	2	40	22	5.0	10.3	3.3	3.2	0.4	9.9
rsGreen0.7-K206A-E222G	6.4	0	52	36	0.4	15.5	37.9	35.9	1.4	5.9
rsGreen0.7-K206A-E222V	7.5	2	38	21	0.3	12.5	61.2	59.2	2.4	24.5

Values between squared brackets are literature values [19, 24]. NA: Not Applicable.  $\lambda_{\text{abs}}$ : absorption peak of the anionic chromophore.  $\lambda_{\text{ex}}$ : excitation peak.  $\lambda_{\text{em}}$ : emission peak.  $\lambda_{\text{off}}^{\text{abs}}$ : absorption maximum of the off-state chromophore.  $\delta_{\text{Stokes}}$  Stokes shift.  $\lambda_{\text{ex}} - \lambda_{\text{abs}}$ : difference between  $\lambda_{\text{ex}}$  and  $\lambda_{\text{abs}}$ .  $\epsilon_{\text{overall}}$ : Excinction coefficient of the anionic chromophore at pH 7.4.  $\epsilon_{\text{anionic}}$ : extinction coefficient of the anionic chromophore at high pH (pH 9 or 10) in which the chromophore is completely in the anionic state.  $\epsilon_{\text{off}}$ : extinction coefficient of the off-state chromophore at pH 7.4. pK<sub>a,app</sub>: apparent pKa.  $\lambda_{\text{abs,pH10}} - \lambda_{\text{abs,pH6.5}}$ : difference between  $\lambda_{\text{abs}}$  at pH 10 and pH 6.5. Q<sub>fluo</sub>: quantum yield of fluorescence. Molecular brightness: brightness at pH 7.4 calculated as Q<sub>fluo</sub> ×  $\epsilon_{\text{overall}}$ . Q<sub>off</sub>: quantum yield of off-switching. Q<sub>on</sub>: quantum yield of on-switching. A<sub>res</sub>: residual absorbance of the anionic peak in the off-state. A<sub>res,∞</sub>: residual absorbance of the anionic peaks in the off-state in the theoretical case of an infinite illumination. k<sub>therm</sub>: rate of thermal recovery from the off-state. A<sub>unrecov</sub>: relative loss in absorbance after a first photoswitching cycle.

**Table 2:** Pearson correlation matrix of the spectroscopic properties. Color code from green ( $r=1$ ) over white ( $r=0$ ) to red ( $r=-1$ ). Values in bold display statistically significant ( $\alpha=0.05$ ) correlations. The positive switchers (rsGreen0.7-K206A-N205C and rsGreen0.7-K206A-N205L) were left out of the calculation.

	1. $\lambda_{abs}$	2. $\lambda_{ex}$	3. $\lambda_{em}$	4. $\lambda_{off}^{abs}$	5. $A_{res}$	6. $A_{res,\infty}$	7. $\epsilon_{off}$	8. $pK_{app}$	9. $A_{unrecov}$	10. $k_{therm}$	11. $\epsilon_{antonic}$	12. $Q_{flu}$	13. Molecular brightness	14. $\epsilon_{overall}$	15. $Q_{off}$	16. $\lambda_{ex} - \lambda_{abs}$	17. $\beta_{Stokes}$	18. $\lambda_{abs,pH10} - \lambda_{abs,pH6.5}$	19. $Q_{off}^*$
1.	1.00	<b>0.81</b>	<b>0.63</b>	0.24	<b>0.57</b>	<b>0.50</b>	<b>0.78</b>	<b>0.71</b>	<b>0.52</b>	<b>0.41</b>	<b>0.55</b>	0.19	-0.11	-0.36	<b>-0.70</b>	<b>-0.62</b>	<b>-0.70</b>	0.06	0.00
2.	<b>0.81</b>	1.00	<b>0.76</b>	-0.07	0.37	0.29	<b>0.61</b>	<b>0.61</b>	<b>0.56</b>	0.32	0.29	0.10	-0.25	<b>-0.48</b>	-0.30	-0.04	<b>-0.88</b>	0.03	0.01
3.	<b>0.63</b>	<b>0.76</b>	1.00	0.23	<b>0.45</b>	0.33	0.40	<b>0.52</b>	<b>0.44</b>	<b>0.56</b>	-0.01	0.08	-0.31	<b>-0.56</b>	-0.33	-0.07	-0.36	0.00	0.31
4.	0.24	-0.07	0.23	1.00	<b>0.76</b>	<b>0.72</b>	0.04	0.00	-0.29	<b>0.53</b>	0.27	0.22	0.32	0.23	<b>-0.47</b>	<b>-0.49</b>	0.26	-0.03	0.24
5.	<b>0.57</b>	0.37	<b>0.45</b>	<b>0.76</b>	1.00	<b>0.97</b>	0.36	0.36	0.08	<b>0.44</b>	<b>0.54</b>	0.08	0.05	-0.03	<b>-0.60</b>	<b>-0.48</b>	-0.20	0.01	0.01
6.	<b>0.50</b>	0.29	0.33	<b>0.72</b>	<b>0.97</b>	1.00	0.34	0.31	0.04	0.26	<b>0.63</b>	0.09	0.11	0.07	<b>-0.57</b>	<b>-0.46</b>	-0.18	0.00	-0.09
7.	<b>0.78</b>	<b>0.61</b>	0.40	0.04	0.36	0.34	1.00	<b>0.76</b>	<b>0.68</b>	0.21	<b>0.45</b>	0.07	-0.24	<b>-0.44</b>	<b>-0.46</b>	<b>-0.50</b>	<b>-0.59</b>	0.16	-0.26
8.	<b>0.71</b>	<b>0.61</b>	<b>0.52</b>	0.00	0.36	0.31	<b>0.76</b>	1.00	<b>0.90</b>	0.35	0.15	-0.02	<b>-0.59</b>	<b>-0.81</b>	-0.38	-0.38	<b>-0.50</b>	0.29	0.05
9.	-0.36	<b>-0.56</b>	<b>-0.44</b>	-0.29	0.08	0.04	<b>0.68</b>	<b>0.90</b>	1.00	0.18	-0.03	-0.01	<b>-0.61</b>	<b>-0.84</b>	-0.08	-0.13	<b>-0.48</b>	0.17	-0.03
10.	<b>0.41</b>	0.32	<b>0.56</b>	<b>0.53</b>	<b>0.44</b>	0.26	0.21	0.35	0.18	1.00	-0.13	0.08	-0.15	-0.33	-0.31	-0.26	-0.06	0.06	<b>0.42</b>
11.	<b>0.55</b>	0.29	-0.01	0.27	<b>0.54</b>	<b>0.63</b>	<b>0.45</b>	0.15	-0.03	-0.13	1.00	0.11	0.33	0.36	<b>-0.57</b>	<b>-0.54</b>	<b>-0.43</b>	0.04	-0.25
12.	0.19	0.10	0.08	0.22	0.08	0.09	0.07	-0.02	-0.01	0.08	0.11	1.00	<b>0.69</b>	0.07	-0.32	-0.20	-0.08	<b>-0.41</b>	0.22
13.	-0.11	-0.25	-0.31	0.32	0.05	0.11	-0.24	<b>-0.59</b>	<b>-0.61</b>	-0.15	0.33	<b>0.69</b>	1.00	<b>0.76</b>	-0.23	-0.14	0.12	<b>-0.42</b>	0.01
14.	-0.36	<b>-0.48</b>	<b>-0.56</b>	0.23	-0.03	0.07	<b>-0.44</b>	<b>-0.81</b>	<b>-0.84</b>	-0.33	0.36	0.07	<b>0.76</b>	1.00	-0.01	-0.02	0.29	-0.23	-0.20
15.	<b>-0.70</b>	-0.30	-0.33	<b>-0.47</b>	<b>-0.60</b>	<b>-0.57</b>	<b>-0.46</b>	-0.38	-0.08	-0.31	<b>-0.57</b>	-0.32	-0.23	-0.01	1.00	<b>0.79</b>	0.20	-0.21	-0.06
16.	<b>-0.62</b>	-0.04	-0.07	<b>-0.49</b>	<b>-0.48</b>	<b>-0.46</b>	<b>-0.50</b>	-0.38	-0.13	-0.26	<b>-0.54</b>	-0.20	-0.14	-0.02	<b>0.79</b>	1.00	0.02	-0.07	0.02
17.	<b>-0.70</b>	<b>-0.88</b>	<b>-0.36</b>	<b>-0.26</b>	-0.20	-0.18	<b>-0.59</b>	<b>-0.50</b>	<b>-0.48</b>	-0.06	<b>-0.43</b>	-0.08	0.12	0.29	0.20	0.02	1.00	-0.04	0.21
18.	0.06	0.03	0.00	-0.03	0.01	0.00	0.16	0.29	0.17	0.06	0.04	<b>-0.41</b>	<b>-0.42</b>	-0.23	-0.21	-0.07	-0.04	1.00	0.04
19.	0.00	0.01	0.31	0.24	0.01	-0.09	-0.26	0.05	-0.03	<b>0.42</b>	<b>-0.25</b>	0.22	0.01	-0.20	-0.06	0.02	0.21	0.04	1.00

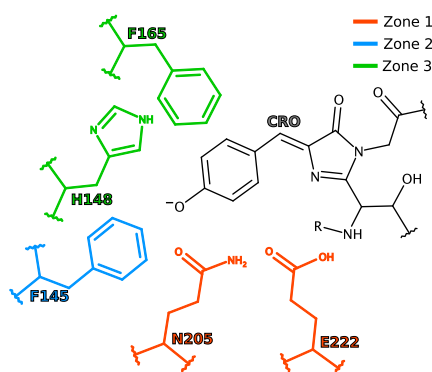
\* Though it is highlighted as significant, the positive correlation between  $Q_{on}$  and  $k_{therm}$  is not reliable because of the low spread in  $Q_{on}$  and the difficulty in measuring this parameter precisely.

information that can be used to accelerate future FP engineering efforts. Whereas strong correlations between  $\lambda_{abs}$ ,  $\lambda_{ex}$  and  $\lambda_{em}$  or between  $Q_{flu}$  and molecular brightness are expected, we also found several relationships that were less anticipated. For example, the difference between the maximal absorption and excitation wavelengths ( $\lambda_{ex} - \lambda_{abs}$ ) positively correlates with the off-switching quantum yield  $Q_{off}$ . This wavelength difference is usually seen as an indicator of ground state heterogeneity within an FP-ensemble, and is consistent with the presence of at least two populations with different photoswitching propensity and brightness in the ground state.<sup>[25,26]</sup> We will discuss particular insights enabled by these correlations further along in this contribution.

One limitation of this analysis is that it cannot reveal intricate dependencies between multiple parameters, nor whether our mutants can be separated into different groups depending on their properties and structures. We were interested in exploring this question using principal component analysis (PCA). PCA identifies linear combinations of properties (principal components, PCs) such that the variance of the projection of the protein properties along these PCs is maximized.<sup>[27]</sup> The obtained principal components are ranked in terms of the variance of the data that they capture, thereby allowing to draw conclusions about linear relationships between the properties. For example, if projecting the protein properties onto the first principal component captures all of the variance, then the sample variability is presumably due to a single mechanism in which the parameters change in the same coordinated manner.

We applied PCA to the spectroscopic dataset, excluding the two positive switchers (rsGreen0.7-K206A-N205C and rsGreen0.7-K206A-N205L) because their properties cannot be consistently grouped ( $Q_{off}$ , for example, must be defined differently for positive switchers). We found that 10 principal components were required to capture 98.2% of the observed variance, suggesting that the FP variability does not arise through just a single or a few different mechanisms. The results were subjected to a visual analysis detailed in Supporting Discussion 1.1. To connect the spectroscopic parameters to the structural modifications, we augmented this analysis by labeling each mutant according to where the corresponding mutations are situated with respect to the chromophore. We divided the chromophore environment into three different zones, consisting of zone 1 (amino acid 205 and 222), zone 2 (amino acid 145) and zone 3 (amino acid 148 and 165) (Figure 2, Table S1).

Ultimately, our methodology allowed us to construct Table 3. This table shows how mutants belonging to a chromophore zone differ from a theoretical FP that has properties equal to the averages of the properties of our mutants, and suggests the parameters that will be modified by introducing a mutation at that particular position. For example, introducing a mutation in zone 1 is likely to result in a change in  $\lambda_{ex} - \lambda_{abs}$  and  $Q_{off}$ . This table also presents a modest guide for further engineering: for



**Figure 2:** Classification of the FPs according to the position of their mutated amino acid in relation to rsGreen0.7 or rsGreen0.7-K206A.

**Table 3:** Principal component analysis on spectroscopic properties indicating the effect of mutation compared to the average.

Zone1 E222X and N205X	Zone 2 F145X	Zone3 H148X and F165X
$\lambda_{ex} - \lambda_{abs} \downarrow$	$\lambda_{ex} - \lambda_{abs} \uparrow^*$	No consistent changes identified
$\lambda_{abs,off} \uparrow$	$\lambda_{abs} \downarrow^*$	
$A_{res} \uparrow$	$\epsilon_{overall} \sim \uparrow$	
$A_{res,\infty} \uparrow$	$pka^{APP} \downarrow^*$	
$Q_{off} \downarrow$	$k_{therm} \downarrow$	
	$\lambda_{abs,off} \downarrow$	
	$Q_{off} \uparrow^*$	

↑: higher than average, ↓: lower than average, ~: equal to average, \*: exceptions present.

example, a faster switcher can be created by mutating in zone 2, though this is likely to also introduce additional spectroscopic heterogeneity as expressed by the  $\lambda_{ex} - \lambda_{abs}$  parameter. However, it includes only a subset of all spectroscopic properties, and no properties at all for zone 3 mutants, reflecting the difficulty in extracting clear conclusions from the overall complexity associated with FPs.

## Inclusion of structural information

We then solved the structures of our mutants in both the on- and off-state using X-ray crystallography, retaining only those samples that diffracted to a resolution of 2.5 Å or better (43 structures in total, including also the structures of rsGreen0.7 published previously;<sup>[19]</sup> Table S2, Supplementary discussion 1.2).

The characterized proteins manifested as 11-stranded  $\beta$ -barrels with a central helix bearing the chromophore (Figure S2). The conformations of the chromophores are highly similar, except for the three mutants at position His148, where the hydroxybenzylidene ring is almost coplanar with the imidazolinone in the off-state (Figure S3). Where it is present, His148 takes up the space created during the *cis*-to-*trans* isomerization of the chromophore and causes it to adopt a twisted conformation, indicating an impact on the photo-isomerization efficiency.

In order to quantify the structural information, we calculated a total of 95 different structural descriptors for each structure, detailed in the Materials and Methods section and Table S3. These properties can be divided into five groups, relating to (i) properties of the chromophore itself, (ii) distances between the chromophore and nearby atoms, (iii) distances between different amino acids surrounding the chromophore, (iv) angles between the planes of nearby aromatic amino acids and the chromophore hydroxybenzylidene ring, and (v) general descriptors of the chromophore environment.

We redid the correlation analysis and PCA using the spectroscopic and structural descriptors associated with the on-state crystal structures. We found that 13 PCs were required to obtain 90.67% explained variance (Figure S4 and S5, Table S4, S5 and S6). We were similarly able to identify specific changes associated with mutations at a particular zone of the chromophore, analogous to the information shown in Table 3. This is shown in Supporting Table S7 rather than in the main text due to the large size of this table.

Performing the same analysis on the off-states was complicated by the lower number of off-state structures available. Furthermore, we were forced to exclude the H148 mutants from consideration due to their different off-state chromophore conformation. The remaining off-state structures mainly belong to zone 2 (Table S1), rendering the previously used zone classification unusable for the off-state analysis. Instead we classified the FPs according to their off-switching quantum yield ( $Q_{\text{off}}$ ) resulting in ‘fast’, ‘medium’ and ‘slow’ switchers (Table S1; Figure S6, Table S8 and S9). This led to the insights summarized in Supporting Table S10.

## **Combined spectroscopic-structural analysis provides insights into FP functioning**

Our dataset and analysis provide several insights into the influence of the chromophore environment on FP functioning.

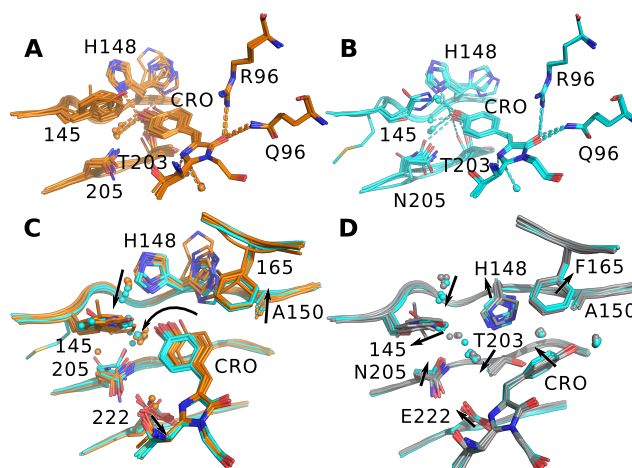
### **Hydrogen bonding to water molecules is a key determinant of the photochromism efficiency**

Our analysis reveals that fast switchers show more hydrogen bonds with the chromophore (Table S7), most of which involve water molecules (Figure 3A and B). These water molecules are present when the residues next to the hydroxybenzylidene moiety have smaller and more polar side chains. In particular, the presence of F145 leads to slow off-switching that is accelerated by its substitution with smaller amino acids. The opening in the  $\beta$ -barrel next to F145 and the hydroxybenzylidene oxygen allows water molecules to be in relatively fast exchange with surrounding water, as compared to other water molecules in the chromophore pocket.<sup>[28,29]</sup> Water mobility has indeed been linked to protein dynamics and interactions and enzymatic kinetics for a wide range of proteins.<sup>[30–32]</sup> The dynamic nature of the water molecules thus presumably accelerates the *cis/trans* isomerization by being sufficiently mobile to stabilize a broad range of different conformations involving the chromophore and/or the surrounding amino acids. The increased dynamics around the hydroxybenzylidene oxygen are confirmed by the higher crystallographic B-factors for the hydrogen bond interaction partners in the fast switchers as compared to the slow switchers. Recently, we also annotated a reduced chromophore anchoring as the cause for isomerisation in photoconvertible FPs.<sup>[21]</sup>

Intriguingly, a previous computational study<sup>[26]</sup> reported that photo-isomerization is precluded in the presence of multiple hydrogen bonds with the chromophore hydroxybenzylidene oxygen, since such bonding interferes with the charge transfer required for photo-isomerization. Our observations suggest that this restriction may be less absolute, at least for this FP scaffold, though it does align with the notion that highly mobile water molecules favor photochromism by facilitating the transient breaking of hydrogen bonds to the chromophore.

This observation also offers an explanation for the high  $\lambda_{\text{ex}} - \lambda_{\text{abs}}$  values observed for fast switchers, since the dynamic water molecules may accommodate a variety of ground state conformations with varying fluorescence brightnesses. The fast switchers also show a larger tilt angle and rotation of the chromophore around the imidazolinone nitrogen atom. This probably also is a consequence of the established hydrogen bonds (Figure 3C), since these bonds are largely oriented in a direction that favors this chromophore geometry.<sup>[33]</sup>

We also observe that the F145M mutant is a fast switcher, which is surprising because methionine is large and rather apolar. However, a detailed inspection of the crystal structure reveals that M145 is present in two conformations, one of which points away from the chromophore and makes room for the inclusion of water molecules (Figure 3B,C and Supplementary Discussion 1.3), and which is presumably responsible for the faster switching. Intriguingly, the F145M/K206A double mutant displays considerably slower switching compared to the F145M mutant. Residue 145 is located close



**Figure 3:** (A and B) Hydrogen bonds made by the on-state chromophore of the FPs belonging to the slow (A) and the fast switchers (B) (list of mutants provided in Table S1), indicating the correlations between the number of hydrogen bonds made by the hydroxybenzylidene oxygen and  $Q_{\text{off}}$  ( $r = 0.58$ ,  $p < 0.01$ ) and  $\lambda_{\text{ex}} - \lambda_{\text{abs}}$  ( $r = 0.72$ ,  $p < 0.001$ ). Hydrogen bonds are depicted by dashes (criterion: length  $\leq 3.2$  Å). If due to alternative conformations multiple hydrogen bond possibilities exist, then only the shortest interaction is shown. If the chromophore has alternative conformations, then only the hydrogen bonds made by a single conformation are drawn. (C and D) Structural superposition of the fast switchers with slow switchers in the on-state (C; cyan and orange, respectively) and off-state (D; cyan and grey, respectively). Arrows point to the main structural shifts.

to residue 206, suggesting that the K206A mutation interferes with the outward conformation of M145.

However, the simple presence of water molecules near the chromophore is not sufficient to confer fast switching: the FPs mutated at the zone 1 region switch more slowly than the other FPs. The N205G and E222G/V mutations both lead to the introduction of water molecules next to the chromophore imidazolinone, although these changes lead to even slower switching. This shows that the positioning of the water molecules is important, and that it is the water molecules situated around the hydroxybenzylidene oxygen of the chromophore that are responsible for enhancing the switching quantum yield.

Steric hindrance has also been considered to be a key determinant for photochromism.<sup>[14,34,35]</sup> We attempted to capture this effect by quantifying the chromophore pocket volume. We find that the on-state volume does not correlate with the off-switching quantum yield, indicating that steric hindrance may be less important, or that it may not be captured well by this descriptor. However, this conclusion needs to be mitigated somewhat when considering that the pocket volume descriptor is sensitive to the overall chromophore environment, including also positions 205 and 222, which should not cause steric hindrance during isomerisation. On the other hand, we do find a strong correlation between the off-switching quantum yield and the chromophore volume in the off-state. In the off-state, the positions of the mutated residues show much less variability (Figure 3D), and therefore the changes in the chromophore pocket volume are dominated by the variability at positions 145, 148, and 150, indicating that steric hindrance by these residues may play a role. While it was difficult to observe a clear effect of steric hindrance through analysis of the pocket volume, the steric hindrance at these residues does appear to be implicated in the off-switching quantum yield.

Overall, the off-switching efficiency appears to be largely determined by the hydrogen bonding to the hydroxybenzylidene oxygen, in line with previous work that highlighted the importance of electrostatic effects,<sup>[36]</sup> and whether these hydrogen bonds arise from dynamic bonding partners. We suggest that the reduced steric hindrance previously suggested as cause for altered photochromism originates from the increased probability for favorable water interactions and resulting stabilization of multiple chromophore conformations.

## The interplay of multiple structural parameters determines the absorption maximum

Our spectroscopic data shows pronounced differences in the peak absorption wavelength  $\lambda_{\text{abs}}$ . The H148 mutants (forming zone 3 together with the F165 mutants) and some zone 1 mutants have a red-shifted absorption peak compared to the other FPs (Table 1, Figure S1). The PCA cannot clearly isolate these red-shifted FPs from other FPs (Figure S7), while the correlation analysis identifies only a few significant, but weak, correlations between  $\lambda_{\text{abs}}$  and the structural descriptors (Table S6). This indicates that the absorption wavelength is determined by multiple structural aspects that are difficult to untangle. However, our analysis does allow us to identify a number of prominent contributors.

The correlation analysis reveals that the red shift correlates with a reduction in the number of hydrogen bonds that are made to the hydroxybenzylidene oxygen atom (Figure S8A). This also explains why fast switchers display a blue-shift in their absorption maximum, as indicated by the strong anticorrelation between  $\lambda_{\text{abs}}$  and  $Q_{\text{off}}$  (Table 2). Several studies, many involving QM/MM, have revealed that (de-)stabilization of the chromophore charge distribution, both in the ground and excited state, determines the absorption maximum.<sup>[37–41]</sup> The excited state of the FP chromophore displays charge transfer away from the hydroxybenzylidene moiety.<sup>[36,42–44]</sup> Increased hydrogen bonding to the hydroxybenzylidene oxygen stabilizes the ground state and destabilizes the excited state, resulting in a blue-shift.

The blue-shifted absorption of rsGreen0.7-K206A-F165W seems to be contradictory because of the low number of hydrogen bonds. The W165 indole can also engage in electron-donating  $\pi$ - $\pi$  interactions with the hydroxybenzylidene ring (Figure S8B), which are assumed to stabilize the excited state charge transfer and lead to a red-shift.<sup>[36,41,45–47]</sup> Instead, the observed blue-shift likely arises because the stacking perturbs the conjugation of the chromophore (tilt, twist and methylene bridge bond angles of  $-156^\circ$ ,  $-21^\circ$  and  $129^\circ$  as compared to the average of  $-164^\circ$ ,  $-9^\circ$  and  $127^\circ$ , respectively; see Figure S9 for angle definitions). This indicates that geometrical constraints on the chromophore can cancel the effects of hydrogen bonding and other chromophore interactions.<sup>[39,46]</sup>

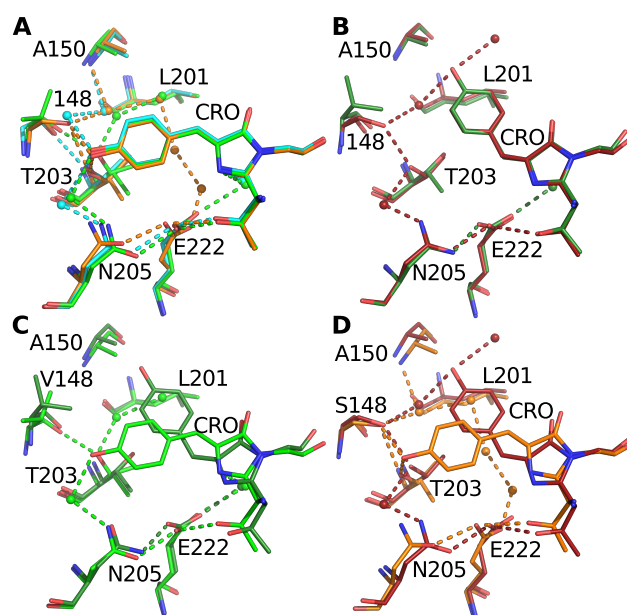
The variation within the F145 mutants (zone 2 mutants) is only marginally captured by these two parameters. Whereas direct interactions, such as hydrogen bonding, with the chromophore can be expected to have the highest influence on the absorption maximum, indirect interactions can also play a role, such as a blue-shifting through the presence of polar residues.<sup>[41,47]</sup> However, the net effect of these polar residues appears to be minor: proteins with Y145 (rsEGFP and rsGreen07.b) are actually red-shifted compared to their non-polar residue-containing counterparts (rsGreen0.7 and rsGreenF) while the introduction of H/Q145 does not introduce the blue-shift that might be expected from these polar residues. Other effects may also be too small to be captured in the X-ray structures, such as a possible dependence of  $\lambda_{\text{abs}}$  on bond-length variations.<sup>[37,48]</sup>

Taken together, our results show that  $\lambda_{\text{abs}}$  depends on multiple structural aspects and is difficult to explain using simple considerations, though hydrogen bonding and chromophore planarity play a key role.

## Hydrogen bonding networks and steric hindrance appear as main contributors to thermal recovery

Three FPs, two of them belonging to the H148 mutants, show rates of thermal recovery ( $k_{\text{therm}}$ ) that are much higher than those observed for other mutants (Table 1, Figure S1). The H148 mutants have an altered chromophore off-state conformation, making direct comparison difficult. However,  $k_{\text{therm}}$  shows considerable variation among the three H148 mutants, suggesting that the comparison of these three FPs can hint to key structural contributions to  $k_{\text{therm}}$ .

The fastest recovery ( $15.3 \times 10^{-4} \text{ s}^{-1} \text{ M}^{-1}$ ) is observed for the H148G mutant. This mutation leads to a chromophore pocket that is directly open to the surrounding bulk solvent, characterized by fewer structured water molecules and heteroatoms close to the hydroxybenzylidene. Unfortunately, no crystal structure of its off-state could be determined, though the unconstrained on-state chromophore pocket suggests a very low steric hindrance<sup>[49]</sup> and the absence of large conformational changes during spontaneous chromophore *trans*-to-*cis* isomerization. In contrast, the H148S mutant shows a slow thermal recovery ( $1.4 \times 10^{-4} \text{ s}^{-1} \text{ M}^{-1}$ ). We find that S148 forms part of an extended hydrogen



**Figure 4:** Superposition of the on-state (A) and off-state (B) chromophore and environment of rsGreen0.7-K206A-H148S (orange, dark red), rsGreen0.7-K206A-H148V (green, dark green) and rsGreen0.7-K206A-H148G (cyan), showing the interaction network around the chromophore. Interactions (hydrogen bonds or electrostatic) are depicted by dashes (criterion: length  $\leq 3.2$  Å). (C and D) Comparison of the on- and off-state interactions made by rsGreen0.7-K206A-H148V and rsGreen0.7-K206A-H148S.

bond network both in the on- and off-state (Figure 4), which has to be broken and rebuilt during *trans*-to-*cis* isomerization, requiring multiple conformational changes and thus bringing down the thermal recovery rate. Finally, the H148V mutant shows an intermediate thermal recovery ( $6.7 \times 10^{-4} \text{ s}^{-1} \text{ M}^{-1}$ ). Valine is slightly larger than serine but apolar and does not support hydrogen bonding, suggesting that its main effect is the result of steric hindrance. We therefore conclude that thermal recovery is affected by both steric hindrance and the hydrogen-bonding interaction network around the chromophore.

### Correlation analysis reveals different responses to the polarity of the chromophore environment

An intriguing finding from our data is the presence of the large green and red ‘blocks’ in the correlation table (Table 2, green: spectroscopic parameter 1-11; red: spectroscopic parameter 13-17). This suggests that the spectroscopic parameters divide into two groups that show opposite responses to mutations, though some of the correlations are comparatively weak. We propose that these two groups arise on account of their opposite sensitivity to the introduction of (a-)polar groups in the chromophore environment. Such polar groups can exert influence by engaging in direct hydrogen bonding and/or by creating a generally more polar environment which is less dependent on the actual orientation. The correlation of a particular spectroscopic parameter with the number of hydrogen bonds to the chromophore can offer insight into the origin of a particular change. For example, blue shifting of the absorption maximum appears to be mainly determined by the establishment of hydrogen bonds, while the pKa seems to be rather influenced by a global polar environment (correlation between  $\lambda_{\text{abs}}$ , pKa and number of H-bonds made by hydroxybenzylidene oxygen -0.44 and -0.29, being significant and non-significant ( $\alpha = 0.05$ ), respectively).

The spectroscopic responses are further modulated by the positioning of these polar/apolar groups relative to the chromophore. This explains why there may be no significant correlation between individual parameters, even though they are part of the same ‘block’ in the correlation table. Indeed, the Stokes shift appears to be less susceptible to mutating residues 145 and 165, but is highly affected by mutations on position 205 and 222 which are closer to the imidazolinone ring compared to the other

mutations done in this study. The impact of these mutation sites is reversed for  $Q_{\text{off}}$ , which is why no correlation is observed between the Stokes shift and  $Q_{\text{off}}$  even though both take part in the red block in the correlation table.

Our structural analysis supervises only the chromophore pocket, while mutations can have effects on much longer distances. For example, the altered properties of rsGreen1 compared to rsGreen0.7 must arise from structural differences further away from the chromophore, since they have the same amino acid composition in the chromophore pocket. This was also observed during the PCA in which rsEGFP was classified as a zone 2 FP based on its mutation in the chromophore pocket, but mostly behaved as an outlier. Finally, the spectroscopic properties can be affected by dynamic or electronic effects that are not observable in the crystal structure.

## Conclusion

We have generated a collection of 27 photochromic fluorescent proteins that show a broad spectroscopic diversity, yet differ in only one or two mutations. We characterized their spectroscopic properties using 19 different parameters and also obtained 43 crystal structures. Our work distinguishes itself from previous efforts in that we aimed to recruit a broad diversity of systems, without emphasizing fitness for a particular purpose, while restricting the number of mutations present.

Our study reinforces the notion that comparatively small biochemical changes in the molecular environment can trigger large changes in the spectroscopic properties of a fluorescent protein chromophore. Our correlation analysis was able to identify multiple correlations across our dataset, which can accelerate FP engineering by identifying parameters that are likely to be linked. A more detailed approach using principal component analysis allowed us to link these changes to mutations at particular positions. However, accounting for the dataset variability requires the inclusion of a large number of principal components, showing that the observed diversity cannot be considered to arise from a limited number of underlying mechanisms. These findings emphasize the high sensitivity of the chromophore to its anisotropic environment. Nevertheless, we were able to infer parameters that are likely to be changed by mutating at a particular position relative to the chromophore.

The inclusion of structural data allowed us to identify a number of parameters that are (some of) the main levers by which FP properties are determined. We find that hydrogen bonding to the hydroxybenzylidene oxygen by water molecules is a key determinant of the photochromism efficiency. Thermal recovery, in contrast, is controlled by restrictions to conformational changes due to hydrogen bonding but also contains clear contributions from steric hindrance. Other effects, such as shifts in the absorption wavelength, are less easy to untangle and appear to arise from multiple mechanisms. Our correlation analysis also suggests that the polarity of the chromophore pocket is a key determinant for many of the spectroscopic properties, though their response depends on whether they are mainly affected by hydrogen bonding or by the overall polarity of the pocket.

In addition to the insights developed here, our work presents a high-quality and consistent dataset of closely related systems including both spectroscopic and structural information, all of which is freely available via this publication and the Protein Data Bank. We anticipate that our dataset can open opportunities for the development and evaluation of new and existing protein engineering and theoretical methods, including computational modelling and machine learning.

## Acknowledgements

We are grateful to Gerrit Groenhof (University of Jyväskylä), Jeremy Harvey (KU Leuven), Raffaele Vitale (Univ. Lille), and Dominique Bourgeois (Institut de Biologie Structurale) for critical insights and discussion. E.D.Z. and L.V.M. thank the beamline staff from X06DA at the Swiss Light Source (Villigen, Switzerland), Proxima1 and Proxima2A at synchrotron Soleil (Gif-sur-Yvette, France), XRD1 at Elettra (Trieste, Italy) and I02 at Diamond Light Source (Oxfordshire, UK) for assistance during X-ray diffraction data collection. E.D.Z., S.H., and S.D. thank the Research Foundation Flanders (FWO) for a doctoral fellowship and postdoctoral fellowships. This work was supported through



funding from the Research Foundation Flanders through grants 1514319N, G090819N, G0B8817N, and the European Research Council through grant 714688 NanoCellActivity.

## Contributor roles

P.D. and E.D.Z. designed the project. E.D.Z. and L.V.M. performed structure determination. E.D.Z. and S.D. carried out mutagenesis and protein purification. E.D.Z., S.D. and W.V. carried out spectroscopic characterization and analysis. E.D.Z. and S.H. carried out statistical analysis with input from A.G.T.. P.D. and E.D.Z. wrote the manuscript with input from all authors.

## Keywords

Structure-Function Relationships, Photochromism, Biophysics, Fluorescent Proteins

## Supporting information

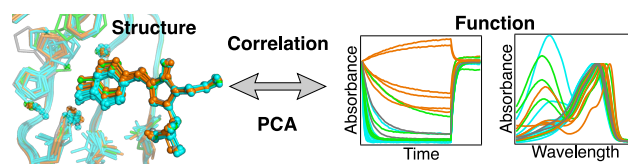
Experimental procedures, individual absorption and emission spectra, pH titrations, light titrations, details on the analysis procedures, crystallographic information.

## References

- [1] R. Silverman, *Organic Chemistry of Enzyme-Catalyzed Reactions*, Elsevier, **2002**.
- [2] D. M. Chudakov, M. V. Matz, S. Lukyanov, K. A. Lukyanov, *Physiological Reviews* **2010**, *90*, 1103–1163.
- [3] P. Dedecker, F. C. De Schryver, J. Hofkens, *J. Am. Chem. Soc.* **2013**, *135*, 2387–2402.
- [4] S. Habuchi, R. Ando, P. Dedecker, W. Verheijen, H. Mizuno, A. Miyawaki, J. Hofkens, *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 9511–9516.
- [5] A. Stiel, S. Trowitzsch, G. Weber, M. Andresen, C. Eggeling, S. Hell, S. Jakobs, M. Wahl, *Biochem. J.* **2007**, *402*, 35–42.
- [6] E. Fron, C. Flors, G. Schweitzer, S. Habuchi, H. Mizuno, R. Ando, F. De Schryver, A. Miyawaki, J. Hofkens, *J. Am. Chem. Soc.* **2007**, *129*, 4870–4871.
- [7] M. Andresen, A. Stiel, J. Fölling, D. Wenzel, A. Schönle, A. Egner, C. Eggeling, S. Hell, S. Jakobs, *Nat. Biotechnol.* **2008**, *26*, 1035–1040.
- [8] H. Mizuno, T. Mal, M. Walchli, A. Kikuchi, T. Fukano, R. Ando, J. Jeyakanthan, J. Taka, Y. Shiro, M. Ikura, et al., *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 9227–9232.
- [9] V. Adam, K. Nienhaus, D. Bourgeois, G. Nienhaus, *Biochemistry-US* **2009**, *48*, 4905–4915.
- [10] T. Brakemann, G. Weber, M. Andresen, G. Groenhof, A. C. Stiel, S. Trowitzsch, C. Eggeling, H. Grubmüller, S. W. Hell, M. C. Wahl, S. Jakobs, *J Biol Chem* **2010**, *285*, 14603–9.
- [11] H. Chang, M. Zhang, W. Ji, J. Chen, Y. Zhang, B. Liu, J. Lu, J. Zhang, P. Xu, T. Xu, *Proc. Natl. Acad. Sci. USA* **2012**, *109*, 4455–4460.
- [12] D. Smyrnova, B. Moeyaert, S. Michielsens, J. Hofkens, P. Dedecker, A. Ceulemans, *J Phys Chem B* **2015**, *119*, 12007–16.
- [13] R. Ando, C. Flors, H. Mizuno, J. Hofkens, A. Miyawaki, *Biophys. J.* **2007**, *92*, L97–L99.
- [14] V. Adam, M. Lelimosin, S. Boehme, G. Desfonds, K. Nienhaus, M. Field, J. Wiedenmann, S. McSweeney, G. Nienhaus, D. Bourgeois, *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 18343–18348.
- [15] F. Pennacchietti, E. O. Serebrovskaya, A. R. Faro, I. I. Shemyakina, N. G. Bozhanova, A. A. Kotlobay, N. G. Gurskaya, A. Bodén, J. Dreier, D. M. Chudakov, K. A. Lukyanov, V. V. Verkhusha, A. S. Mishin, I. Testa, *Nat Methods* **2018**, *15*, 601–604.
- [16] M. Andresen, A. C. Stiel, S. Trowitzsch, G. Weber, C. Eggeling, M. C. Wahl, S. W. Hell, S. Jakobs, *Proceedings of the National Academy of Sciences* **2007**, *104*, 13005–13009.

- [17] M. Kaucikas, M. Tros, J. J. van Thor, *J Phys Chem B* **2015**, *119*, 2350–62.
- [18] S. P. Laptinok, A. A. Gil, C. R. Hall, A. Lukacs, J. N. Iuliano, G. A. Jones, G. M. Greetham, P. Donaldson, A. Miyawaki, P. J. Tonge, S. R. Meech, *Nature Chemistry* **2018**, *10*, 845–852.
- [19] S. Duwé, E. De Zitter, V. Gielen, B. Moeyaert, W. Vandenberg, T. Grotjohann, K. Clays, S. Jakobs, L. Van Meervelt, P. Dedecker, *ACS Nano* **2015**, *9*, 9528–9541.
- [20] E. De Zitter, D. Thédié, V. Mönkemöller, S. Hugelier, J. Beaudouin, V. Adam, M. Byrdin, L. Van Meervelt, P. Dedecker, D. Bourgeois, *Nat Methods* **2019**, *16*, 707–710.
- [21] E. De Zitter, J. Ridard, D. Thédié, V. Adam, B. Lévy, M. Byrdin, G. Gotthard, L. Van Meervelt, P. Dedecker, I. Demachy, D. Bourgeois, *J Am Chem Soc* **2020**, *142*, 10978–10988.
- [22] D. A. Zacharias, J. D. Violin, A. C. Newton, R. Y. Tsien, *Science* **2002**, *296*, 913–6.
- [23] R. Ando, H. Mizuno, A. Miyawaki, *Sci. Signal.* **2004**, *306*, 1370–1373.
- [24] T. Grotjohann, I. Testa, M. Leutenegger, H. Bock, N. T. Urban, F. Lavoie-Cardinal, K. I. Willig, C. Eggeling, S. Jakobs, S. W. Hell, *Nature* **2011**, *478*, 204–208.
- [25] T. Roebroek, S. Duwé, W. Vandenberg, P. Dedecker, *Int J Mol Sci* **2017**, *18*.
- [26] D. Morozov, G. Groenhof, *Angew Chem Int Ed Engl* **2016**, *55*, 576–578.
- [27] R. Bro, A. K. Smilde, *Analytical Methods* **2014**, *6*, 2812–2831.
- [28] A. Shinobu, N. Agmon, *J Phys Chem B* **2015**, *119*, 3464–78.
- [29] N.-E. Christou, I. Ayala, K. Giandoreggio-Barranco, M. Byrdin, V. Adam, D. Bourgeois, B. Brutscher, *Biophys J* **2019**, *117*, 2087–2100.
- [30] P. W. Fenimore, H. Frauenfelder, B. H. McMahon, F. G. Parak, *Proc Natl Acad Sci U S A* **2002**, *99*, 16047–51.
- [31] H. Frauenfelder, G. Chen, J. Berendzen, P. W. Fenimore, H. Jansson, B. H. McMahon, I. R. Stroe, J. Swenson, R. D. Young, *Proceedings of the National Academy of Sciences* **2009**, *106*, 5129–5134.
- [32] M.-C. Bellissent-Funel, A. Hassanali, M. Havenith, R. Henchman, P. Pohl, F. Sterpone, D. van der Spoel, Y. Xu, A. E. Garcia, *Chem Rev* **2016**, *116*, 7673–97.
- [33] B. Storti, E. Margheritis, G. Abbandonato, G. Domenichini, J. Dreier, I. Testa, G. Garau, R. Nifosì, R. Bizzarri, *ACS Chem Biol* **2018**, *13*, 2082–2093.
- [34] N. Coquelle, M. Sliwa, J. Woodhouse, G. Schirò, V. Adam, A. Aquila, T. R. M. Barends, S. Boutet, M. Byrdin, S. Carbajo, E. De la Mora, R. B. Doak, M. Feliks, F. Fieschi, L. Foucar, V. Guillon, M. Hilpert, M. S. Hunter, S. Jakobs, J. E. Koglin, G. Kovacsova, T. J. Lane, B. Lévy, M. Liang, K. Nass, J. Ridard, J. S. Robinson, C. M. Roome, C. Ruckebusch, M. Seaberg, M. Thepaut, M. Cammarata, I. Demachy, M. Field, R. L. Shoeman, D. Bourgeois, J.-P. Colletier, I. Schlichting, M. Weik, *Nat Chem* **2018**, *10*, 31–37.
- [35] J. Chang, M. G. Romei, S. G. Boxer, *J Am Chem Soc* **2019**, *141*, 15504–15508.
- [36] M. G. Romei, C.-Y. Lin, I. I. Mathews, S. G. Boxer, *Science* **2020**, *367*, 76–79.
- [37] P. Amat, R. Nifosì, *J Chem Theory Comput* **2013**, *9*, 497–508.
- [38] B. L. Grigorenko, A. V. Nemukhin, I. V. Polyakov, D. I. Morozov, A. I. Krylov, *J Am Chem Soc* **2013**, *135*, 11541–9.
- [39] M. Nadal-Ferret, R. Gelabert, M. Moreno, J. M. Lluch, *J Chem Theory Comput* **2013**, *9*, 1731–42.
- [40] K. E. Merabti, S. Azizi, J. Ridard, B. Lévy, I. Demachy, *Chemical Physics* **2017**, *493*, 157–165.
- [41] R. Nifosì, B. Mennucci, C. Filippi, *Phys Chem Chem Phys* **2019**, *21*, 18988–18998.
- [42] M. E. Martin, F. Negri, M. Olivucci, *J Am Chem Soc* **2004**, *126*, 5452–64.
- [43] J.-y. Hasegawa, T. Ise, K. J. Fujimoto, A. Kikuchi, E. Fukumura, A. Miyawaki, Y. Shiro, *J Phys Chem B* **2010**, *114*, 2971–9.
- [44] M. Drobizhev, P. R. Callis, R. Nifosì, G. Wicks, C. Stoltzfus, L. Barnett, T. E. Hughes, P. Sullivan, A. Rebane, *Sci Rep* **2015**, *5*, 13223.
- [45] R. M. Wachter, M. A. Elsliger, K. Kallio, G. T. Hanson, S. J. Remington, *Structure* **1998**, *6*, 1267–77.
- [46] M. Z. Lin, M. R. McKeown, H.-L. Ng, T. A. Aguilera, N. C. Shaner, R. E. Campbell, S. R. Adams, L. A. Gross, W. Ma, T. Alber, R. Y. Tsien, *Chem Biol* **2009**, *16*, 1169–79.

- [47] J. W. Park, Y. M. Rhee, *Phys Chem Chem Phys* **2016**, *18*, 3944–55.
- [48] C. Filippi, F. Buda, L. Guidoni, A. Sinicropi, *J Chem Theory Comput* **2012**, *8*, 112–24.
- [49] D. Smyrnova, K. Zinovjev, I. Tuñón, A. Ceulemans, *J Phys Chem B* **2016**, *120*, 12820–12825.



## TOC

De Zitter *et al.* study how the anisotropic chromophore environment of a fluorescent protein determines its properties using a large structural and spectroscopic dataset of closely-related variants.