



Cite this: *Phys. Chem. Chem. Phys.*, 2022, 24, 15937

# Molecular origins of the multi-donor strategy in inducing bathochromic shifts and enlarging Stokes shifts of fluorescent proteins†

Xia Wu,<sup>a</sup> Davin Tan,<sup>a</sup> Qinglong Qiao,<sup>b</sup> Wenting Yin,<sup>b</sup> Zhaochao Xu <sup>\*b</sup> and Xiaogang Liu <sup>\*a</sup>

Long-wavelength fluorescent proteins (LWFPs) and LWFP-based sensors are indispensable tools for bioimaging and biosensing applications. However, it remains challenging to develop LWFPs with outstanding brightness and/or sensitivities, largely due to the lack of simple and effective molecular design strategies. Herein, we rationalized the molecular origins of a multi-donor strategy that affords significant bathochromic shifts and large Stokes shifts with minimal structural changes in the resulting protein fluorophores. We analyzed three key factors that affect the spectral properties of these fluorophores, including the (1) substituent position, (2) electron-donating strength, and (3) number of electron-donating groups. We further demonstrated that this simple design strategy is generalizable to various fluorophore families. We expect that this work can provide rational guidelines for developing fluorescent proteins (and small-molecule fluorophores) with long emission wavelengths and large Stokes shifts.

Received 15th February 2022,  
Accepted 7th June 2022

DOI: 10.1039/d2cp00759b

[rsc.li/pccp](http://rsc.li/pccp)

## 1 Introduction

The availability of a platter of fluorescent proteins of different colours is vital for biological research.<sup>1–3</sup> Among various colours, long-wavelength fluorescent proteins (LWFPs) are particularly attractive due to their reduced phototoxicity, decreased autofluorescence background, and enhanced penetration depths in tissues.<sup>4–6</sup> However, the limited choices of residue moieties in natural amino acids and highly constrained inner cavities in proteins place a tight restriction on the length of  $\pi$ -conjugations. The development of LWFPs and associated probes thus becomes a significant challenge.

To address this challenge, extensive efforts have been directed to the rational design of fluorescent proteins, through a deep understanding of their formation processes and fluorescence mechanisms.<sup>7–9</sup> Yet, predicting the folding of proteins and in particular the reactions and rearrangement of the residues is a difficult task. To date, the performance of LWFPs is still outshined by a green fluorescent protein (GFP) in terms

of both the brightness and hue. Most LWFPs have undesirable photophysical properties, such as slow maturation rates and reduced photostability.<sup>10,11</sup> Moreover, many LWFP-based biosensors exhibit small dynamic ranges, undesired photoconversion, and mislocalization; the variety of these sensors is also highly limited, in comparison to their GFP analogues.<sup>12,13</sup> To circumvent these challenges, one may adopt a hybridization method by attaching synthetic dyes to proteins *via* various tags (such as SNAP-tag<sup>14</sup>/CLIP-tag,<sup>15</sup> and Halo-tag<sup>16</sup>), thus generating distinct colours of fluorescence.<sup>17–20</sup> Unfortunately, introducing extrinsic dyes in live systems faces many limitations in comparison to genetically encoded fluorescent proteins that are produced *in situ*. Notably, the cotranslational incorporation of unnatural amino acids (UAAs) has gained considerable progress in recent years.<sup>21,22</sup> UAAs greatly expanded the choice of residue structures, creating exciting opportunities for developing fluorescent proteins of novel structures and distinct properties. However, the chromophores in these proteins are still subjected to limited cavity sizes. Accordingly, the large expansion of  $\pi$ -conjugation, a popular strategy for producing long-wavelength small-molecule dyes, is inherently incompatible with LWFPs.

Instead of the expansion of  $\pi$ -conjugation, incorporating additional electron-donating groups at judiciously chosen positions of fluorophores affords greatly enlarged Stokes shifts ( $\Delta\lambda$ ) and leads to significant bathochromic shifts in the fluorescence spectra *via* minimal structural changes. We have first demonstrated that this multi-donor strategy is generalizable to a wide

<sup>a</sup> Fluorescence Research Group, Singapore University of Technology and Design, 8 Somapah Road, Singapore, 487372, Singapore. E-mail: [xiaogang\\_liu@sutd.edu.sg](mailto:xiaogang_liu@sutd.edu.sg)

<sup>b</sup> CAS Key Laboratory of Separation Science for Analytical Chemistry, Dalian Institute of Chemical Physics, Chinese Academy of Sciences, 457 Zhongshan Road, Dalian 116023, China. E-mail: [zcxu@dicp.ac.cn](mailto:zcxu@dicp.ac.cn)

† Electronic supplementary information (ESI) available. See DOI: <https://doi.org/10.1039/d2cp00759b>

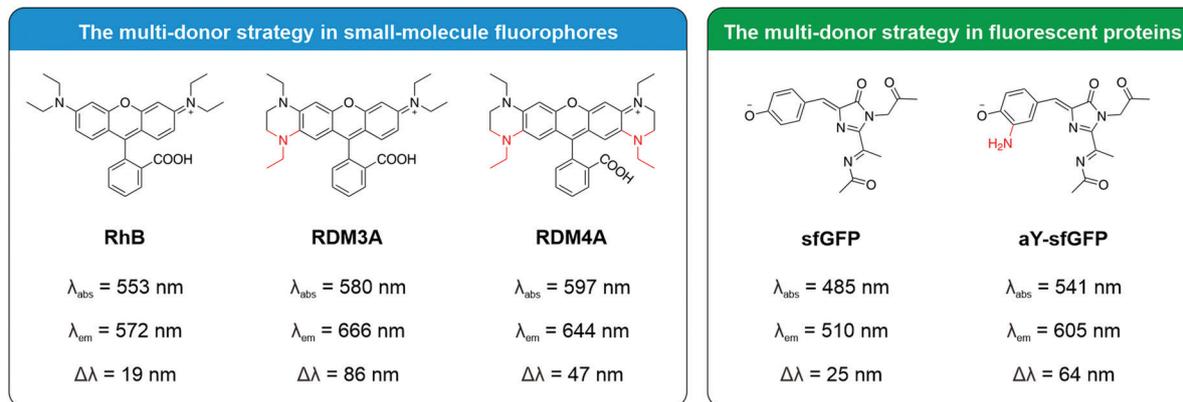


Fig. 1 Chemical structures and experimental peak UV-vis absorption wavelength ( $\lambda_{\text{abs}}$ ), peak emission wavelength ( $\lambda_{\text{em}}$ ), and Stokes shifts ( $\Delta\lambda$ ) of RhB, RDM3A and RDM4A (in ethanol) and sfGFP and aY-sfGFP.

range of small-molecule fluorophores, such as rhodamine, rhodol, fluorescein, naphthalimide, and naphthalene (Fig. 1, the left panel and Fig. S1, ESI<sup>†</sup>).<sup>23–27</sup> This multi-donor strategy has also been adopted by several research groups for developing small-molecule fluorophores and probes with long emission wavelengths and large Stokes shifts (Fig. S1, ESI<sup>†</sup>).<sup>28–31</sup> Notably, in a landmark study, Ai and co-workers have recently introduced a genetically encoded artificial amino acid, 3-aminotyrosine, to the fluorophore of GFP-like proteins and biosensors for “spontaneous and efficient green-to-red conversion” (Fig. 1, the right panel).<sup>32</sup> Their designs of fluorescent proteins employed one additional amino group and utilized the multi-donor design strategy. However, the mechanistic underpinnings of how these fluorescent proteins can exhibit significant red shifts and Stokes shifts remain unclear. Molecular design guidelines on enhancing these shifts are also lacking. Understanding these molecular origins and design rules, however, is critical for the rational deployment of Ai’s method to other families of fluorescent proteins, to realize enhanced bathochromic shifts and enlarged Stokes shifts with minimal structural changes in protein cavities.

Herein, we systematically studied the mechanism of creating LWFPs by introducing additional electron-donating groups to fluorophores (as did by the Ai group). By using density functional theory (DFT) and time-dependent density functional theory (TD-DFT) calculations, we analysed three important factors in this multi-donor design strategy: (1) the substitution positions, (2) the electron-donating strength of the additional donors, and (3) the number of donors. Moreover, we demonstrate that this molecular design strategy is generalizable to other fluorescent protein systems. We foresee that our results along with the rapid evolution of UAA techniques will afford a broad range of LWFPs with minimal structural changes whilst still retaining the desirable photophysical properties of their predecessors.

## 2 Computational methods

Geometry optimizations were carried out at the  $\omega$ B97XD/Def2svp level, unless stated otherwise.<sup>33</sup> Solvation effects (in water) were taken into account using the SMD model.<sup>34</sup> The excitation

and emission energies of all molecules were calculated using state-specific equilibrium solvation. Frequency calculations were performed to verify that the obtained structures are stable, without containing any imaginary vibration frequencies. All DFT and TD-DFT calculations were carried out using the Gaussian 16A suite of programs.<sup>35</sup> The molecular excitation properties were also investigated by hole–electron analysis using Multiwfn 3.6.<sup>36</sup>

## 3 Results and discussion

As we mentioned *vide supra*, Ai and co-workers introduced 3-aminotyrosine to the fluorophore of the green fluorescent protein-like sfGFP to obtain red fluorescent protein aY-sfGFP.<sup>32</sup> Essentially, the hydroxyl group in the sfGFP acts as the first electron-donating group, and the introduced amino group serves as the second donor, which led to a red-shift of the UV-vis absorption and emission peaks, and a large Stokes shift in the resulting aY-sfGFP. To elucidate the roles of this additional donor group, we performed DFT and TD-DFT calculations at the ground state ( $S_0$ ) and the first excited state ( $S_1$ ) of GFP fluorophore analogues in water.

### 3.1 The impact of linking an additional amino group on the NGFP

To this end,  $S_0$ – $S_1$  photoexcitation/deexcitation typically corresponds to electronic transitions between the highest occupied molecular orbital (HOMO) and the lowest unoccupied molecular orbital (LUMO). Accordingly, the electronic structures of the HOMO and LUMO, along with the HOMO–LUMO energy gap, can be used to predict the UV-vis absorption and emission wavelengths of fluorescent proteins. Indeed, the  $S_0$ – $S_1$  photoexcitation/de-excitation of all studied molecules in this work are mainly composed of the HOMO to LUMO transitions.

We note that the atomic contribution to the electron density at the 5-position in the NGFP is very high in the HOMO (0.12) but decreases considerably in the LUMO (0.02) (Fig. 2 and Fig. S2, ESI<sup>†</sup>). Therefore, introducing an electron-donating group at the 5-position can greatly destabilize the HOMO, but has little effect on the LUMO, thereby reducing the electronic

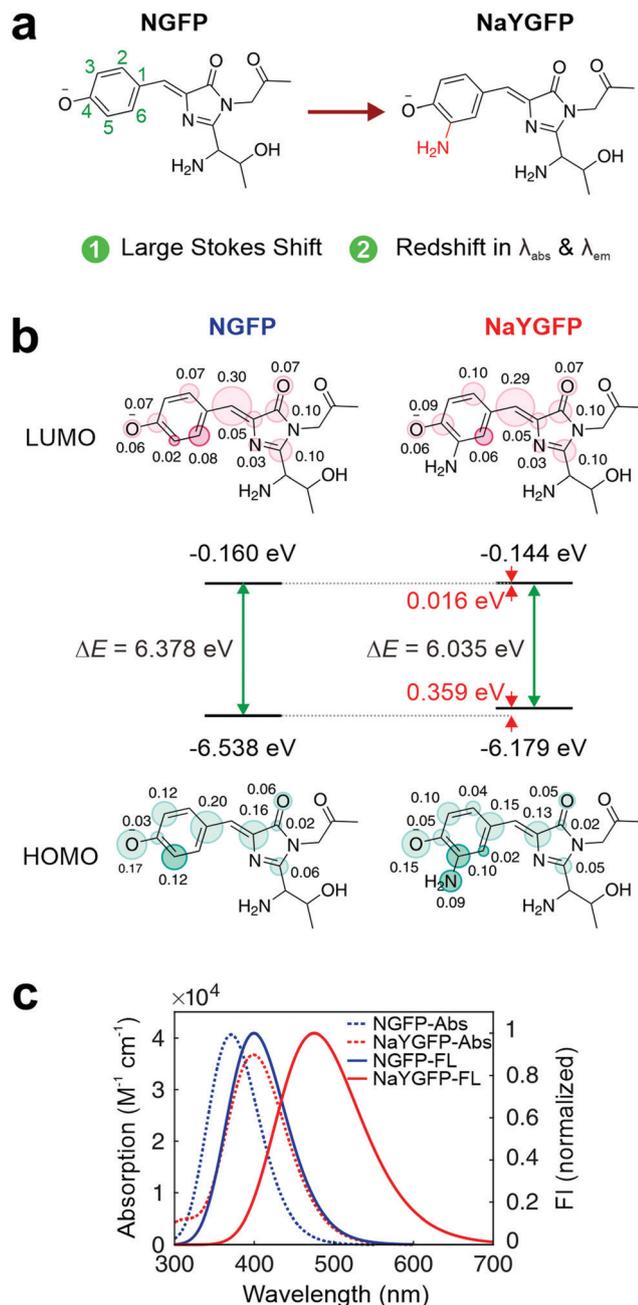


Fig. 2 (a) Illustration of the multi-donor strategy based on the NGFP. (b) Molecular frontier orbitals and electronic transitions and calculated (c) UV-vis absorption and normalized emission of the NGFP and NaYGFP in water.

gap of the NGFP. Indeed, our calculations show that the NaYGFP, which bears an additional amino donor at the 5-position, has a higher HOMO energy level (0.359 eV) and a smaller electronic gap ( $\Delta E_{\text{LUMO-HOMO}} = 6.035$  eV) compared with the NGFP. In parallel with this reduced electronic gap, the NaYGFP demonstrates considerable red shifts in both the UV-vis absorption and fluorescence spectra with respect to the NGFP.

Our calculations also reveal that the additional donor group in the NaYGFP has a significant impact on the Stokes shifts by

enhancing intramolecular charge transfer (ICT) upon photoexcitation.<sup>37,38</sup> Given that the  $S_1$  photoexcitation of the fluorophores (investigated in our study) is dominated by HOMO–LUMO transitions, the degree of ICT can be quantified *via* evaluating the spatial overlap between these two frontier molecular orbitals. Attaching an electron-donating group at the substituent position where a significant reduction in the electron density occurs upon the HOMO–LUMO transition could greatly enhance ICT.<sup>39</sup> In the NaYGFP, the HOMO–LUMO transition is accompanied by a considerable decrease of the electron density around the amino group at the 5-position, whereby the corresponding atomic contribution declines from 0.09 to almost 0. This large decrease indicates that the substantial ICT occurs in the NaYGFP upon photoexcitation.

To quantify the degree of ICT, we conducted electron-hole analysis using the Multiwfn program, by obtaining and comparing the  $t$  index (ESI† part 1).<sup>36,40,41</sup> The  $t$  index represents the degree of separation between the holes and electrons during charge transfer, whereby a larger  $t$  index corresponds to a higher degree of ICT. As shown in Table 1, our calculations revealed that the  $t$  index of the NaYGFP ( $-0.868$  Å) was larger than that of the NGFP ( $-1.933$  Å). Given the large extent of ICT and its associated geometry relaxation upon excitation, our calculations show that the NaYGFP exhibits a larger  $\Delta\lambda$  of 76 nm than the NGFP (30 nm; Table 1). This prediction is in good agreement with experimental observations.

To verify the accuracy of our computational results using  $\omega$ B97XD/Def2svp, we further calculated the NGFP and NaYGFP using different functionals and basis sets (Fig. S3, S4 and Tables S2, S3, ESI†). The results consistently showed that the additional donor group leads to red-shifts of  $\lambda_{\text{abs}}$  and  $\lambda_{\text{em}}$  as well as the increase of  $\Delta\lambda$ , in agreement with the results obtained using the  $\omega$ B97XD functional and the Def2svp basis set. Moreover, our results showed that the NaYGFP had longer  $\lambda_{\text{abs}}$  and  $\lambda_{\text{em}}$  as well as larger  $\Delta\lambda$  than the NGFP under various polarities (Fig. S5 and Table S4, ESI†).

These results reveal the molecular origins of the additional donor group in altering the electronic structures of the parent fluorophores, thus leading to significant red-shifts in both  $\lambda_{\text{abs}}$  and  $\lambda_{\text{em}}$ , and the enlargement of  $\Delta\lambda$ .

It is of note that many molecular design strategies could enhance the Stokes shifts, such as incorporating a rotary substituent,<sup>39</sup> introducing excited-state intramolecular proton transfer (ESIPT),<sup>42</sup> and enhancing the push-pull effects in organic fluorophores.<sup>43</sup> Stokes shifts are correlated with the degree of geometry relaxation between the Franck–Condon (FC) state and the adiabatic excited state. Such a geometry relaxation (or the change of nuclear coordinates) is essentially caused by intramolecular charge transfer. In fluorescent proteins, while the rotation of the chromophore is constrained and ESIPT is not applicable (to our systems), enhancing ICT (or increasing the push-pull effect) is an effective method to enlarge the geometrical relaxation and enhance the Stokes shift.

Inspired by the success in rationalizing the photophysical properties of fluorophores adopting the multi-donor strategy, we investigated various factors that correlate the donor groups

**Table 1** Calculated/experimental peak UV-vis absorption wavelength ( $\lambda_{\text{abs}}$ ), peak emission wavelength ( $\lambda_{\text{em}}$ ), Stokes shifts ( $\Delta\lambda$ ), oscillator strength ( $f$ ) of  $S_1$  and  $t$ -index of the GFP fluorophore analogues in water

Fluorophore analogues	$\lambda_{\text{abs}}$ (nm)	$\lambda_{\text{abs}}^a$ (nm)	$\lambda_{\text{em}}$ (nm)	$\lambda_{\text{em}}^a$ (nm)	$\Delta\lambda$ (nm)	$\Delta\lambda^a$ (nm)	$f$	$t$ index (Å)
NGFP	371	485	400	510	30	25	0.9810	-1.933
NaYGFP	398	541	474	605	76	64	0.6093	-0.868
NbYGFP	368		395		27		0.7859	-2.259
NMGFP	376		404		29		0.9888	-1.897
NMOGFP	375		406		31		0.9935	-1.934
N2aYGFP	422		462		41		0.8375	-1.391

<sup>a</sup> Experimental data are extracted from the literature.<sup>32</sup>

with the fluorescence properties of the resulting fluorescent proteins (Fig. 3).

### 3.2 The impact of the substituent position on the spectral properties of fluorescent proteins

First, we analysed the available positions in which the donors could be attached to the molecular scaffold of the proteins for achieving the largest red-shifts (Fig. 3(a)). In the phenol ring of the NGFP, four positions are available for substitution, including 2-, 3-, 5-, and 6-positions (Fig. 2(a)). We noted that, due to structural symmetry, the 2- and 6-positions possess similar electronic structures; so are 3- and 5-positions. Since we have analysed the 5-position substitutions in the NaYGFP, our subsequent discussion will focus on the 6-position substitution.

In contrast to the decrease of atomic contributions at the 5-position upon the HOMO-LUMO transition, the electron density at the 6-position exhibits an opposite trend in the NGFP (Fig. 3(b)). The atomic contribution is small in the HOMO ( $\sim 0$ ) but becomes higher in the LUMO (0.08). Therefore, attaching a donor group at the 6-position does not affect the HOMO much, but could largely destabilize the LUMO. Indeed, the resulting NbYGFP has a similar HOMO level to the NGFP, but a much higher LUMO than the NGFP. The  $\Delta E_{\text{LUMO-HOMO}}$  of 6.453 eV in the NbYGFP is larger than that of the NGFP (6.378 eV), causing a blue-shift in the predicted UV-vis absorption and emission peaks of the NbYGFP (in comparison to those of the NGFP and NaYGFP; Fig. 3(c)-(e) and Table 1).

The substituent position also has a significant impact on the degree of ICT and the associated Stokes shifts.<sup>37</sup> While the amino group at the 5-position greatly enhances the ICT of the NaYGFP, the atomic contributions of the additional amino group at the 6-position in the NbYGFP remained almost unchanged (from 0.02 in HOMO to 0.03 in LUMO), indicating a small extent of ICT.

As shown in Table 1, our calculations revealed that the  $t$  index of the NbYGFP ( $-2.259$  Å) is much smaller than that of the NaYGFP ( $-0.868$  Å). Given the small extent of ICT and its associated geometry relaxation upon excitation, our calculations show that the NbYGFP exhibits a smaller Stokes shift (27 nm) than the NaYGFP does (76 nm; Table 1).

These results demonstrate the importance of the position in which the donor group is located as it can directly affect the electronic structures and alter the UV-vis absorption and fluorescence properties of the proteins. To induce the maximum red-shifts, the positions to an electron-donating group

are only optimal when they possess large electron density in the HOMO and small electron density in the LUMO.

### 3.3 The impact of the electron-donating strength of the attached donors on the spectral properties of fluorescent proteins

We next explored how the electron-donating strength of donors can affect the photophysical properties of the fluorescent proteins. Based on our previous analysis, we attached these donors at the 5-position to achieve large red shifts. In these donor groups, the electron-donating strength can be described by the Hammett value, in which the donating strength increases in this order: methyl ( $-0.170$ ) < methoxy ( $-0.268$ ) < amino ( $-0.660$ ).<sup>44,45</sup>

Analysis of the frontier orbitals of the NMGFP, NMOGFP, and NaYGFP revealed a trend in the increasing HOMO energy levels as the strength of the donor increases with a decrease in  $\Delta E_{\text{LUMO-HOMO}}$  (Fig. 3(b)). Nevertheless, the difference between the HOMO of the NGFP, NMGFP, and NMOGFP is almost negligible, due to the weak electron-donating strength of the methyl and methoxy groups. As a result, the NMGFP exhibits only a  $\sim 5$  nm red shift in its UV-vis absorption and emission wavelengths compared to the NGFP, with a relatively small ICT (and a small  $t$  index; Table 1). The same trend was also observed for the NMOGFP. Hence, NGFP, NMGFP, and NMOGFP have similar spectral properties.

However, with a strong electron-donating group (*i.e.*, the amino group), the HOMO energy level of the NaYGFP increases by  $\sim 0.3$  eV compared to the rest of fluorophores in this series (Fig. 2). Moreover, during the HOMO-LUMO transition, the change of the atomic contribution around the amino group amounts to 0.09, which is 9 times greater than that of the NMOGFP, indicating a significant ICT effect in the NaYGFP. This change means that the NaYGFP exhibits a large Stokes shift with a great  $t$  index ( $-0.868$  Å). Conversely, NGFP, NMGFP, and NMOGFP have small Stokes shifts with similar  $t$  indexes (around  $-1.9$  Å). As a result, the NaYGFP displays a greater red-shift in  $\lambda_{\text{abs}}$  (398 nm) and especially in  $\lambda_{\text{em}}$  (474 nm) with a larger  $\Delta\lambda$  (76 nm) than those of NGFP, NMGFP, and NMOGFP (Table 1 and Fig. 3).

These calculations showed that it is possible to achieve larger Stokes shifts and red-shifted UV-vis absorption and emission spectra when attaching additional strong electron-donating groups onto the fluorophore scaffold at the selected position (*i.e.*, the 5-position for the NGFP). Indeed, the

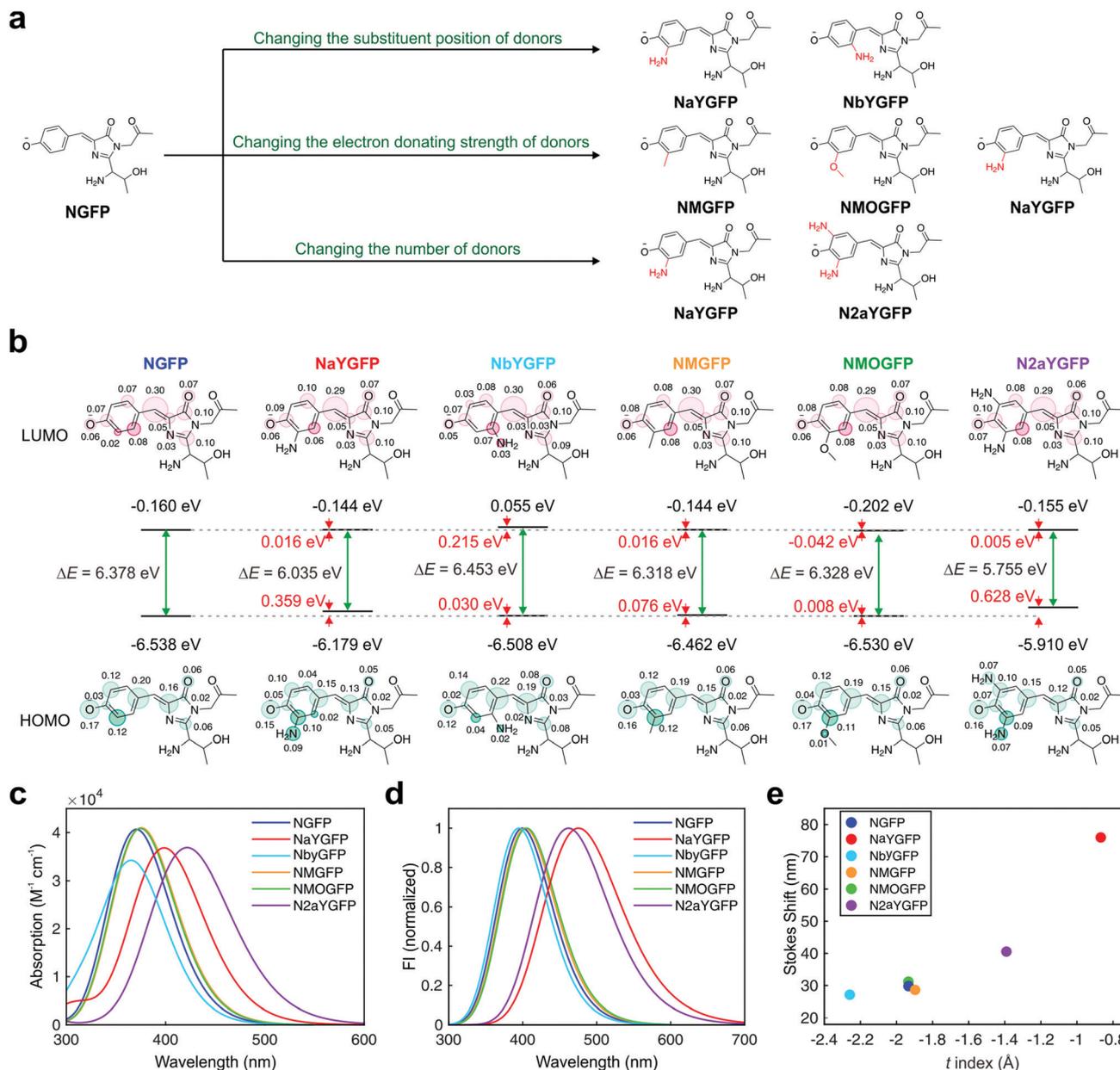


Fig. 3 (a) Chemical structures, (b) atomic contributions to the charge densities of the LUMO and HOMO, the energy levels of the HOMO and LUMO and the corresponding electronic gaps ( $\Delta E$ ), calculated (c) UV-vis spectra, (d) normalized emission spectra, and (f) Stokes shifts as a function of the  $t$  index in the NGFP analogues in water.

experimental data also show that increasing the electron-donating strength/decreasing the electron-withdrawing strength of the 5-position substituent leads to a red shift.<sup>46</sup>

### 3.4 The impact of the number of donors on the spectral properties of fluorescent proteins

Based on the results and discussion *vide supra*, we wondered if the incorporation of additional donor groups onto the main molecular scaffold can induce even greater red shifts. To this end, we designed and modelled a new compound N2aYGFP that contains two amino groups that are *ortho* to the phenol moiety (a total of three electron-donating groups; Fig. 3).

Our calculations revealed that the N2aYGFP has a higher HOMO energy level ( $-5.910$  eV) than NGFP ( $-6.538$  eV) and NaYGFP ( $-6.179$  eV) and had a smaller HOMO–LUMO gap ( $5.755$  eV) than NGFP ( $6.378$  eV) and NaYGFP ( $6.035$  eV). The reduced electronic gap corresponds to a red shift of  $\lambda_{\text{abs}}$  ( $51$  nm longer than that of NGFP and  $24$  nm longer than that of NaYGFP; Fig. 3).

It is interesting to note that an additional electron-donating group in the N2aYGFP compared to that in the NaYGFP does not induce further red shifts in  $\lambda_{\text{em}}$ . Our computational results revealed that the N2aYGFP had a blue-shifted  $\lambda_{\text{em}}$ , owing to a reduced Stokes shift when compared with the NaYGFP. To gain

further understanding, we analysed the charge distributions in the frontier molecular orbitals of both N2aYGFP and NaYGFP. The  $t$  index of the N2aYGFP is about  $-1.39 \text{ \AA}$ , vs.  $-0.87 \text{ \AA}$  for the NaYGFP. This difference means that the N2aYGFP experiences a smaller extent of ICT than the NaYGFP, upon photoexcitation. As a result, although it has an additional donor group, the N2aYGFP exhibits a smaller Stokes shift than the NaYGFP. These results suggest that equipping three donor groups (*i.e.*, in N2aYGFP) could lead to a more substantial red shift in  $\lambda_{\text{abs}}$ . However, having more donors does not always afford a longer  $\lambda_{\text{em}}$  than their double-donor analogues (*i.e.*, NaYGFP). The large degree of ICT and the associated Stokes shift play a crucial role in the long peak emission wavelength of the NaYGFP. Nevertheless, it is worth mentioning that the N2aYGFP still exhibits enhanced red shifts in both  $\lambda_{\text{abs}}$  and  $\lambda_{\text{em}}$ , as well as the enlarged  $\Delta\lambda$ , in comparison to the NGFP (a single-donor fluorophore).

Our conclusion is supported by the experimental data reported on small-molecule fluorophores (Fig. 1(a)). For example, Zhang *et al.* also synthesized RDM4A that contained four amino donor groups,<sup>47</sup> while Xu *et al.* developed a rhodamine derivative RDM3A bearing three amino donor groups (Fig. 1).<sup>24</sup> The peak

UV-vis absorption/emission wavelengths and the Stokes shift of RDM4A in ethanol are 597 nm/644 nm and 47 nm, respectively. In comparison, these values of RDM3A are 580 nm/666 nm and 86 nm, respectively. RDM4A possesses a longer  $\lambda_{\text{abs}}$  but a shorter  $\lambda_{\text{em}}$  than RDM3A. This difference in  $\lambda_{\text{em}}$  is caused by their different Stokes shifts. Note that the asymmetric molecular structure of RDM3A enhances ICT (in comparison to the symmetrical RDM4A), thus affording a larger Stokes shift than that of RDM4A. It is thus clear that the Stokes shift of the compound is independent of the total number of donor groups but is dependent on the extent of ICT.

It is worth highlighting that the large degree of ICT typically results in a decrease in the molar extinction coefficient. Enhancing the ICT effectively reduces the spatial overlap between the HOMO and LUMO, thus reducing the transition probability between these two molecular orbitals. This change in turn results in a decrease of the oscillator strength, which is related to the molar extinction coefficient. Indeed, it is worth highlighting that many "ICT" dyes (such as coumarins, naphthalimide, and NBD dyes) demonstrated much smaller molar extinction coefficients than "resonant" dyes (with a small degree of ICT, a.k.a. locally excited dyes).<sup>48</sup> We thus need to

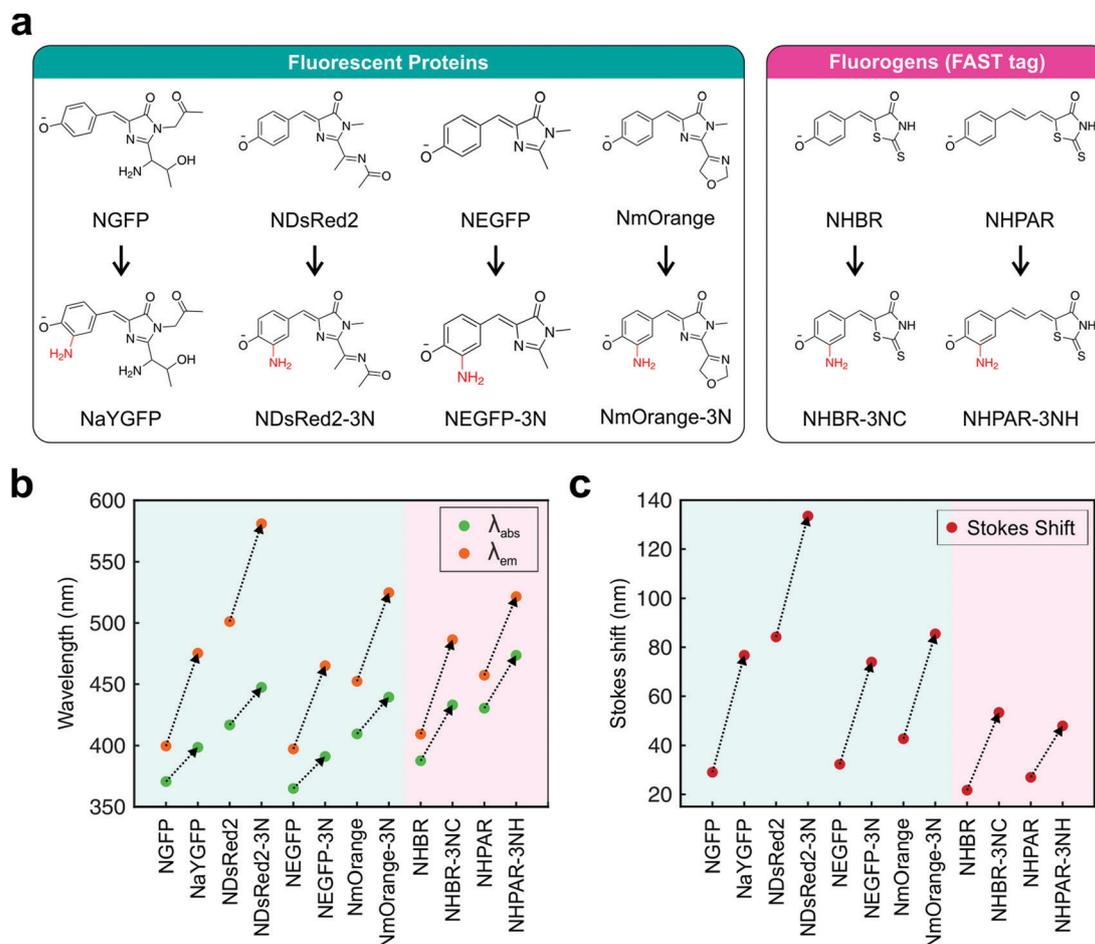


Fig. 4 (a) Chemical structures of various chromophores in fluorescent proteins, fluorogens, and their multi-donor derivatives. Calculated (b) peak UV-vis absorption wavelengths ( $\lambda_{\text{abs}}$ ), peak emission wavelengths ( $\lambda_{\text{em}}$ ), and (c) Stokes shifts of molecules in water.

carefully select the number of donors to suit the requirements of a given application.

This multi-donor design strategy can also be extended to other fluorescent protein families (Fig. 4), such as DsRed2, EGFP, and mOrange derivatives (Fig. S6–S8 and Tables S5–S7, ESI†). Additional calculations on these proteins were consistent with our previous results. That is, the multi-donor strategy can effectively shift their UV-vis absorption and emission bands into the long-wavelength region, with enlarged Stokes shifts. This further highlights the effectiveness and reliability of this strategy to expand the number of LWFPs, even into the near-infrared region.

Finally, the multi-donor design strategy can also be extended to “fluorogens” in tag proteins. Tag proteins could be divided into two types: (1) self-labelling tags that covalently bind an extrinsic dye upon a specific reaction, such as SNAP-tag<sup>14</sup>/CLIP-tag<sup>15</sup> and Halo-tag;<sup>16</sup> (2) fluorogenic activation proteins (FAPs) that non-covalently bind “fluorogens” at their reaction sites.<sup>49</sup> The reaction sites of FAPs have fixed cavity sizes and can only be activated by specific fluorophores.<sup>50,51</sup> This condition severely limits the type and number of available fluorophores that can be employed, and most used fluorophores can only emit in the short wavelength range. To overcome these shortcomings, it is important to modify fluorophores that bind to FAPs with minimal modification (to suit the cavity size) but maximal spectral shifts.

Our calculations show that the multi-donor strategy is also applicable to such fluorogens, with significant red shifts and enlarged Stokes shifts *via* minor structural modifications (Fig. 4 and Fig. S9, S10, Tables S8, S9, ESI†). To this end, it is worth noting that Gautier's group has designed a new fluorogenic dye by linking two methoxy groups to 4-hydroxybenzylidene-rhodanine (HBR), which they named HBR-3,5DOM. With the incorporation of multiple electron-donating groups, HBR-3,5DOM experiences a significant red shift (> 50 nm) in comparison to HBR and fluoresces red light, while still forming a tight complex with a fluorescence-activating and absorption-shifting tag (FAST).<sup>52</sup> Their results further corroborate the generalizability of the multi-donor strategy in LWFPs.

## 4 Conclusions

In conclusion, we investigated three important factors that contribute to the molecular origins of the multi-donor strategy in designing fluorescent proteins with desirable bathochromic shifts in the absorption and emission peaks and large Stokes shifts: (1) the substituent positions of the donor group in the main molecular scaffold of the protein fluorophores (or fluorogens) must have a large electron density in the HOMO and a small electron density in the LUMO; (2) at this selected position, attaching stronger electron-donating groups will enable larger red shifts and Stokes shifts; (3) the emission wavelengths and Stokes shifts of fluorescent proteins (fluorogens) are independent of the numbers of donors but are dependent on the extent of ICT. These results provide important guidelines for

engineering fluorescent proteins and protein-based fluorescent probes with significant red shifts *via* minimal structural changes. These guidelines are also applicable to small-molecule fluorophores. We thus believe that this work can provide a useful platform for dye chemists to expand the family of long-wavelength fluorophores.

## Conflicts of interest

There are no conflicts to declare.

## Acknowledgements

This work was supported by the Agency for Science, Technology and Research (A\*STAR) under its Advanced Manufacturing and Engineering Program (A2083c0051), the Ministry of Education, Singapore (MOE-MOET2EP10120-0007), the Singapore University of Technology and Design [SUTD-ZJU (VP)201905], and the Dalian Institute of Chemical Physics (No. DMTO201603 and TMSR201601). The authors are grateful for the computing service of SUTD-MIT IDC and the National Supercomputing Centre (Singapore).

## Notes and references

- 1 D. S. Bindels, L. Haarbosch, L. Van Weeren, M. Postma, K. E. Wiese, M. Mastop, S. Aumonier, G. Gotthard, A. Royant and M. A. Hink, *Nat. Methods*, 2017, **14**, 53–56.
- 2 S. J. Remington, *Protein Sci.*, 2011, **20**, 1509–1519.
- 3 H.-w. Ai, K. L. Hazelwood, M. W. Davidson and R. E. Campbell, *Nat. Methods*, 2008, **5**, 401–403.
- 4 E. C. Greenwald, S. Mehta and J. Zhang, *Chem. Rev.*, 2018, **118**, 11707–11794.
- 5 B. J. Bevis and B. S. Glick, *Nat. Biotechnol.*, 2002, **20**, 83–87.
- 6 J. V. Frangioni, *Curr. Opin. Chem. Biol.*, 2003, **7**, 626–634.
- 7 F. V. Subach, K. D. Piatkevich and V. V. Verkhusha, *Nat. Methods*, 2011, **8**, 1019–1026.
- 8 M. Zhang, H. Chang, Y. Zhang, J. Yu, L. Wu, W. Ji, J. Chen, B. Liu, J. Lu and Y. Liu, *Nat. Methods*, 2012, **9**, 727–729.
- 9 C. Fang, M. Drobizhev, H. L. Ng and P. Pantazis, *Front. Mol. Biosci.*, 2021, **8**, 701523.
- 10 I. Shemiakina, G. Ermakova, P. Cranfill, M. Baird, R. Evans, E. Souslova, D. Staroverov, A. Gorokhovatsky, E. Putintseva and T. Gorodnicheva, *Nat. Commun.*, 2012, **3**, 1–7.
- 11 R. E. Campbell, O. Tour, A. E. Palmer, P. A. Steinbach, G. S. Baird, D. A. Zacharias and R. Y. Tsien, *Proc. Natl. Acad. Sci. U. S. A.*, 2002, **99**, 7877–7882.
- 12 G. Ladds, A. Goddard and J. Davey, *Trends Biotechnol.*, 2005, **23**, 367–373.
- 13 D. L. Sheridan and T. E. Hughes, *BMC Biotechnol.*, 2004, **4**, 1–7.
- 14 A. Keppler, S. Gendreizig, T. Gronemeyer, H. Pick, H. Vogel and K. Johnsson, *Nat. Biotechnol.*, 2003, **21**, 86–89.
- 15 J. Ule, K. B. Jensen, M. Ruggiu, A. Mele, A. Ule and R. B. Darnell, *Science*, 2003, **302**, 1212–1215.

- 16 G. V. Los, L. P. Encell, M. G. McDougall, D. D. Hartzell, N. Karassina, C. Zimprich, M. G. Wood, R. Learish, R. F. Ohana and M. Urh, *ACS Chem. Biol.*, 2008, **3**, 373–382.
- 17 N. V. Povarova, S. O. Zaitseva, N. S. Baleeva, A. Y. Smirnov, I. N. Myasnyanko, M. B. Zagudaylova, N. G. Bozhanova, D. A. Gorbachev, K. K. Malyshevskaya and A. S. Gavrikov, *Chem. – Eur. J.*, 2019, **25**, 9592–9596.
- 18 K. S. Mineev, S. A. Goncharuk, M. V. Goncharuk, N. V. Povarova, A. I. Sokolov, N. S. Baleeva, A. Y. Smirnov, I. N. Myasnyanko, D. A. Ruchkin and S. Bukhdruker, *Chem. Sci.*, 2021, **12**, 6719–6725.
- 19 Y. Chekli, C. Peron-Cane, D. Dell’Arciprete, J.-F. Allemand, C. Li, J.-M. Ghigo, A. Gautier, A. Lebreton, N. Desprat and C. Beloin, *Sci. Rep.*, 2020, **10**, 1–13.
- 20 S. Leng, Q. Qiao, Y. Gao, L. Miao, W. Deng and Z. Xu, *Chin. Chem. Lett.*, 2017, **28**, 1911–1915.
- 21 J. Xie and P. G. Schultz, *Nat. Rev. Mol. Cell Biol.*, 2006, **7**, 775–782.
- 22 K. Lang and J. W. Chin, *Chem. Rev.*, 2014, **114**, 4764–4806.
- 23 X. Liu, PhD thesis, University of Cambridge, 2014.
- 24 Z. Xu, W. Yin and X. Liu, CN 104710816 B, 2017.
- 25 Z. Xu, S. Zheng and Y. Liu, CN 106866437 B, 2018.
- 26 Z. Xu, W. Yin and X. Liu, CN 104710815 B, 2017.
- 27 Z. Xu and W. Feng, CN 106867271 A, 2017.
- 28 T. B. Ren, W. Xu, W. Zhang, X. X. Zhang, Z. Y. Wang, Z. Xiang, L. Yuan and X. B. Zhang, *J. Am. Chem. Soc.*, 2018, **140**, 7716–7722.
- 29 W. Chen, S. Xu, J. J. Day, D. Wang and M. Xian, *Angew. Chem., Int. Ed.*, 2017, **129**, 16838–16842.
- 30 Y. Gan, G. Yin, T. Yu, Y. Zhang, H. Li and P. Yin, *Talanta*, 2020, **210**, 120612.
- 31 X. Ren, F. Zhang, H. Luo, L. Liao, X. Song and W. Chen, *Chem. Commun.*, 2020, **56**, 2159–2162.
- 32 S. Zhang and H.-w. Ai, *Nat. Chem. Biol.*, 2020, **16**, 1434–1439.
- 33 J.-D. Chai and M. Head-Gordon, *Phys. Chem. Chem. Phys.*, 2008, **10**, 6615–6620.
- 34 A. V. Marenich, C. J. Cramer and D. G. Truhlar, *J. Phys. Chem. B*, 2009, **113**, 6378–6396.
- 35 M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, G. A. Petersson, H. Nakatsuji, X. Li, M. Caricato, A. V. Marenich, J. Bloino, B. G. Janesko, R. Gomperts, B. Mennucci, H. P. Hratchian, J. V. Ortiz, A. F. Izmaylov, J. L. Sonnenberg, D. Williams, F. Ding, F. Lipparini, F. Egidi, J. Goings, B. Peng, A. Petrone, T. Henderson, D. Ranasinghe, V. G. Zakrzewski, J. Gao, N. Rega, G. Zheng, W. Liang, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, K. Throssell, J. A. Montgomery Jr., J. E. Peralta, F. Ogliaro, M. J. Bearpark, J. J. Heyd, E. N. Brothers, K. N. Kudin, V. N. Staroverov, T. A. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. P. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, J. M. Millam, M. Klene, C. Adamo, R. Cammi, J. W. Ochterski, R. L. Martin, K. Morokuma, O. Farkas, J. B. Foresman and D. J. Fox, *Gaussian 16 Rev. C.01*, Gaussian Inc., Wallingford, CT, 2016.
- 36 T. Lu and F. Chen, *J. Comput. Chem.*, 2012, **33**, 580–592.
- 37 X. Liu, J. M. Cole and Z. Xu, *J. Phys. Chem. C*, 2017, **121**, 13274–13279.
- 38 D. Huang, Y. Chen and J. Zhao, *Dyes Pigm.*, 2012, **95**, 732–742.
- 39 X. Liu, Z. Xu and J. M. Cole, *J. Phys. Chem. C*, 2013, **117**, 16584–16595.
- 40 Z. Liu, T. Lu and Q. Chen, *Carbon*, 2020, **165**, 461–467.
- 41 T. Le Bahers, C. Adamo and I. Ciofini, *J. Chem. Theory Comput.*, 2011, **7**, 2498–2506.
- 42 A. C. Sedgwick, L. Wu, H.-H. Han, S. D. Bull, X.-P. He, T. D. James, J. L. Sessler, B. Z. Tang, H. Tian and J. Yoon, *Chem. Soc. Rev.*, 2018, **47**, 8842–8880.
- 43 M. Ahn, M.-J. Kim, D. W. Cho and K.-R. Wee, *J. Org. Chem.*, 2020, **86**, 403–413.
- 44 C. Hansch, A. Leo and R. Taft, *Chem. Rev.*, 1991, **91**, 165–195.
- 45 D. H. McDaniel and H. C. Brown, *J. Org. Chem.*, 1958, **23**, 420–427.
- 46 M. G. Romei, C.-Y. Lin, I. I. Mathews and S. G. Boxer, *Science*, 2020, **367**, 76–79.
- 47 Z. Tian, B. Tian and J. Zhang, *Dyes Pigm.*, 2013, **99**, 1132–1136.
- 48 U. Resch-Genger, M. Grabolle, S. Cavaliere-Jaricot, R. Nitschke and T. Nann, *Nat. Methods*, 2008, **5**, 763–775.
- 49 C. Szent-Gyorgyi, B. F. Schmidt, Y. Creeger, G. W. Fisher, K. L. Zakel, S. Adler, J. A. Fitzpatrick, C. A. Woolford, Q. Yan and K. V. Vasilev, *Nat. Biotechnol.*, 2008, **26**, 235–240.
- 50 D. S. Waugh, *Trends Biotechnol.*, 2005, **23**, 316–320.
- 51 D. S. Waugh, *Protein Expression Purif.*, 2011, **80**, 283–293.
- 52 C. Li, M.-A. Plamont, H. L. Sladitschek, V. Rodrigues, I. Aujard, P. Neveu, T. Le Saux, L. Jullien and A. Gautier, *Chem. Sci.*, 2017, **8**, 5598–5605.