Unraveling the molecular mechanisms of color expression in anthocyanins†

Mariami Rusishvili, a,b,c Luca Grisanti, c,d,e Sara Laporte, c,d,e Marco Micciarelli, c,d,e Martinez, f Marta Rosa, g h Rebecca J. Robbins, c Tom Collins, d Alessandra Magistrato e and Stefano Baroni a, d

Anthocyanins are a broad family of natural dyes, increasingly finding application as substitutes for artificial colorants in the food industry. In spite of their importance and ubiquity, the molecular principles responsible for their extreme color variability are poorly known. We address these mechanisms by computer simulations and photoabsorption experiments of cyanidin-3-O-glucoside in water solution, as a proxy for more complex members of the family. Experimental results are presented in the range of pH 1–9, accompanied by a comprehensive systematic computational study across relevant charge states and tautomers. The computed spectra are in excellent agreement with the experiments, providing unprecedented insight into the complex behavior underlying color expression in these molecules. Besides confirming the importance of the molecule’s charge state, we also unveil the hitherto unrecognized role of internal distortions in the chromophore, which affect its degree of conjugation, modulating the optical gap and in turn the color. This entanglement of structural and electronic traits is also shared by other members of the anthocyanin family (e.g. pelargonidin and delphinidin) highlighting a common mechanism for color expression across this important family of natural dyes.

1 Introduction

Regulations on colorants are some of the oldest and more rigorous ones in the food industry, with the first list of prohibited additives dating back to 1906.1 Moreover, nowadays the food industry is experiencing additional mounting pressure from the public to substitute artificial additives with natural alternatives, perceived as healthier and safer. Obvious requirements that natural colorants have to fulfill for use in food applications include non-toxicity, abundance and cost-effectiveness, chemical stability, as well as versatility. While nature provides plenty of options for colors in the red-yellow-green gamut, choices are much more limited in the red-purple-blue range, particularly when it comes to the blue end of the palette.2 Anthocyanins are a broad family of dyes responsible for most of the colors in the red-purple-blue gamut found in flowers, fruits, vegetables, leaves, tubers and food products (wines, jams, and syrups) and they account for a large amount of color expression in nature.2–5

The extreme variety of hues, brightness, and nuances expressed by anthocyanins reflects their chemical variability. Substituents on the A, C and B rings (Fig. 1), including hydroxyl and glycosylic acids, will also modulate the color through both inter- and intra-molecular co-pigmentation. In addition, the optical properties of anthocyanins are highly sensitive to environmental factors such as the pH of the solution or the presence of metal ions.6 The red-colored flavylium cation exists at very acidic conditions (pH < 2) while in the pH range of 3–8, flavylium converts into neutral purple-colored forms, due to the deprotonation of one of its hydroxyls. Upon further pH increase, another hydroxyl will deprotonate, giving rise to the blue-colored anionic quinonoid forms.3,6

This variability and complexity of color expression makes the industrial utilization of anthocyanins a challenging endeavor, compared to synthetic dyes which are more stable and predictable. Thus a deep understanding of the mechanisms of color expression in these molecules is still needed, in spite of recent extensive experimental2,3,5,7,8 and theoretical9–16 efforts. Moreover, most of the available studies concern the cationic species—responsible for the red hues only—while only few studies have focused so far...
on the neutral and negative species, which are of striking interest for industrial applications.\textsuperscript{2,7}

The color expressed by an anthocyanin solution is determined by the optical properties of the solvated molecules, which can be measured experimentally through photoabsorption spectroscopy. Predicting the color using theory and computational methods, however, is not as straightforward.\textsuperscript{17,18} Efforts in this direction have comprised quantum chemical calculations performed at various levels of theory, ranging from semiempirical methods,\textsuperscript{13–15,19} to post Hartree–Fock\textsuperscript{12} and TDDFT,\textsuperscript{9,10,16} most of which, however, focused on cyanidin-3-glucoside (C3G, hereafter), chosen as a prototype for singly glycosylated anthocyanins, and which is highly abundant in nature. Fig. 1 shows the relevant molecular states (justified in Section 3.2) and chemical equilibria determining the optical properties of C3G throughout the 1–9 pH range: the flavylum cation (hereafter referred as A\(^{+}\)), two prototropic tautomeric states (protomers) of the neutral base, deprotonated at either position 4\(^{\prime}\) or 7 (hereafter referred to as A\(_{4}\)\(_{0}\) and A\(_{7}\)\(_{0}\)), and two protomers of the negative base (hereafter referred to as A\(_{4}\)\(_{7}\) and A\(_{5}\)\(_{5}\), deprotonated at positions 4\(^{\prime}\) and 7 or 4\(^{\prime}\) and 5, respectively). Experimentally, controlling the pH is relatively straightforward, however computationally there is no easy correspondence. For this reason, in this study we simulate each of the charge and protomeric states shown in Fig. 1 separately, and compare experimental spectra at a given pH with the computed spectra corresponding to the charge state which is expected to exist in majority at that pH. We have found that each molecular state can also adopt various conformations, whose effects on the optical properties are carefully accounted for in our simulations and discussed in this paper. We define and discuss the elements responsible for the coupling between electronic and structural degrees of freedom, the latter being mainly represented by the torsion in the chromophore between the two conjugated units. The agreement between our theoretical and experimental results is excellent, giving us confidence on the quality of the simulations and their ability to capture the fine details of microscopic mechanisms underlying color expression in this important family of natural dyes.

\section{Methods}
\subsection{Photoabsorption spectroscopy}
C3G solutions from pH 1–9 were prepared by adding 1 mL of cyanin stock solution to 20 mL Milli-Q deionized water, adjusting to respective pH using either HCl or NaOH, then completing

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{Fig1.png}
\caption{Scheme of the relevant molecular species of cyanidin-3-glucoside (C3G) in solution at pH 1–9. From left to right, the flavylum cation in red (A\(^{+}\)), in pink the neutral species deprotonated at either 4\(^{\prime}\) (A\(_{4}\)\(_{0}\)) or 7 (A\(_{7}\)\(_{0}\)), and in purple two negative species, deprotonated at 4\(^{\prime}\) and 5 (A\(_{5}\)\(_{5}\)), or at 4\(^{\prime}\) and 7 (A\(_{4}\)\(_{7}\)).}
\end{figure}
volume to a total of 24 mL in order to keep the concentration of cyanin constant.

The stock solution of cyanin was prepared by dissolving 11.60 mg cyanin chloride into 25 mL acidified water (0.01% HCl). Cyanin solutions were pH adjusted using various volumes of NaOH and HCl depending on desired final pH. The sodium hydroxide (2.0 M/0.2 M) (sodium hydroxide pellets 99.9% trace metal analysis from Sigma Aldrich – Cat. no. 306576) and HCl (1 M/0.1 M) stock solutions were prepared by dissolving NaOH pellets or diluting concentrated HCl (Hydrochloric Acid Trace-Metal Grade from Fisher Scientific – Cat. no. A508-P212), respectively into Milli-Q deionized water (EMD Millipore Milli-Q deionized water). The pH of the final cyanin-pH adjusted solutions were determined by using a Mettler-Toledo SevenEasy pH meter before measuring spectra.

To measure the spectra, an aliquot of pH adjusted cyanin solution was transferred to a 1 cm polystyrene UV visible cuvette (1 cm polystyrene UV visible cuvettes from Fisher Scientific – Cat. no. 14-955-125) on an Agilent 8453 UV visible spectrophotometer. UV visible absorbance data was obtained for wavelength range of 200–700 nm with 1 nm resolution and 0.5 second integration time using Agilent ChemStation software.

2.2 Computer simulations

We have developed a multi-scale computational protocol, enabling us to account for the many effects governing the optical properties of C3G in solution. Firstly, the conformational landscape of each molecular state is sampled using µs-scale enhanced-sampling molecular dynamics (MD). Then, each relevant conformer is simulated by means of ab initio MD (AIMD) including explicit solvent, thus accurately accounting for thermal fluctuations and hydrogen bonding effects. The optical spectra of selected statistically independent frames are then computed using time-dependent density functional theory (TDDFT), and the final spectrum is obtained as an average of these spectra. The color of the solution is finally estimated using the standard tri-stimulus model. While the details of the protocol will be presented and discussed in a future work, its main steps are summarized below.

MD. Classical MD simulations were run using GROMACS 4.5.5, upon adequate equilibration with all bonds to hydrogen atoms constrained using LINCS. General AMBER Force Field (GAFF) parameters were assigned using the antechamber module of AmberTools13, with RESP charges at the HF/6-31G* level, calculated on DFT-B3LYP optimized geometries, 50/200 Ry basis set. For each species of C3G studied, ten replicas were simulated using Hamiltonian replica-exchange MD (HREMD), as implemented in the PLUMED plugin, for a total of 10 µs (1 µs per replica). The molecules were fully solvated in TIP3P water, using cubic cells of 30 Å edge accommodating ~1000 water molecules, in periodic boundary conditions. The volume was kept fixed, and velocity rescaling was used to keep the average temperature constant (NVT ensemble). In each case only the trajectory of the first replica, at 300 K, was retained for analysis.

Conformational analysis. An advanced clustering algorithm was used to identify the most representative conformers, according to the three dihedrals illustrated in Fig. 2. These were selected as they describe distortions of the chromophore, which would in turn affect the color. The statistical weight of each conformer thus identified was evaluated as the normalized ratio between the residence time within each free-energy basin and the total simulation time. In general for each species four main clusters were identified, corresponding to four distinct orientations of θ. The relative orientations of the AC and B rings always deviate from planarity, due to steric hindrance of the substituents on the aromatic rings, while the varying position of the sugar results from competition between steric and solvent effects. The parameters used for the clustering algorithm are given in the ESL+.

AIMD. Explicit-solvent Car-Parrinello AIMD simulations were performed for each one of the four most populated conformers of the five C3G species investigated in this work, using the cp.x module of the Quantum ESPRESSO (QE) distribution, Ultrason pseudopotentials from the QE public repository were used with plane-wave basis sets (25/200 Ry) and the PBE exchange correlation-functional. The number of water molecules was reduced to ~200, so as to fit a simulation cell of linear dimensions ~20 Å. The NVT ensemble was used, with a Nosé–Hoover thermostat to maintain the temperature at 300 K.

TDDFT. For each relevant conformer, the absorption spectrum was computed using the methodology introduced in ref. 10. Molecular configurations of C3G were selected from the AIMD trajectories every 0.5 ps, ensuring statistical independence. Absorption spectra were computed for each such configuration using the TDDFT-based self-consistent continuum solvent (SCCS) approach, as implemented in Quantum ESPRESSO and using the B3LYP XC functional, which was found in ref. 10 and 42 to perform well for the optical properties of anthocyanins, using a 50/200 Ry basis set. The solvation model was enhanced by explicitly keeping solvent molecules which are found to be persistently H-bonded to the chromophore (typically, at positions 4’, 3’, 5 and 7; see Table S2, ESI†), thus better accounting for the effects of a protic solvent on the electronic structure of the solute. Finally, for each conformer the spectrum was evaluated as the average of the spectra computed for each selected frame.

Absorption spectra. The overall spectrum for the solution of each species was evaluated by averaging the TDDFT spectra computed for each conformer, weighted by the factors computed in the conformational analysis. The resulting color was finally estimated using the tri-stimulus model, as in ref. 10.

Additional quantum-chemical calculations were performed to corroborate some of our results. In particular, QM calculations included in the protocol are performed with QE (with SCCS for solvation), while extra analysis on the orbitals were run using Gaussian09 (with PCM for solvation). Further details are...
3 Results

3.1 Clustering analysis and molecular conformers

For each molecular state, our HREMD trajectory was analyzed in terms of the three dihedrals $\theta$, $\alpha_6$, and $\alpha_7$ (Fig. 2), in order to find representative geometries corresponding to conformers separated by free energy barrier of the order of a few $k_B T$, impossible to overcome on the timescale accessible to AIMD. Results from this analysis are shown in Fig. 3 for the $A_0^0$ molecular state, and the same clustering procedure was adopted for all the other studied species, giving qualitatively similar results. Conformers are detected as high density clusters in the dihedral space, here visible in the pair-wise distributions, depicted as 2D projections of a 3D density map. The clustering algorithm used in this work is based on density and distance cut-off criteria to group the three dimensional data points corresponding to molecular geometries from the MD trajectory. For each cluster a center is defined and the corresponding molecular geometry is selected as the most representative structure (Fig. 3, lower panel). Conformers were identified for each molecular state, as shown in detail in the ESI† for H-bond analysis.

Most previous studies have focused on the cationic form of anthocyanins, basically because full protonation lifts any ambiguity as to its tautomeric state and makes it the only relevant species at low pH, thus facilitating the comparison with experiments. In addition, in this work we investigate two out of the three protomers of both the neutral and negative species (excluding $A_0^0$ and $A_0^+$), selected according to their relative DFT energies (see Table 1). We thus expect C3G in solution to be predominantly a mix of the remaining five charge/protomeric states, whose relative abundance depends on pH.

Our measured and computed spectra are displayed and compared with each other in Fig. 4. The average spectra for each molecular species were computed by weighting the contributions from each conformer ($C_1$ to $C_4$) with their relative population (see Fig. S2, ESI†). Experimental spectra at different pH are grouped and compared with those computed for the charge species expected to be prevalent in solution at those pH values, considering the two successive pK values of 4 and 7 generally used as a reference for anthocyanins indicating that cyanin below pH 4 is expected to be positively charged, while above pH 7 it is expected to be negatively charged. At pHs between 4 and 7 the neutral species are expected to exist. Our predictions are in excellent agreement with experiment, as concerns the position and pH/charge dependence of the maximum absorption wavelength, $\lambda_{\text{max}}$. The latter is determined by the $S_0 \rightarrow S_1$ transition, which is essentially a single-configuration HOMO $\rightarrow$ LUMO transition.

reported in the ESI† Please see: Tables S6 and S7 (ESI†) for bond order calculations; Table S2 (ESI†) for H-bond analysis.

3.2 Optical spectra: effects of pH, charge, and protomeric state

![Fig. 3](image-url) Upper panel: 2D density maps of pairs of dihedral angles (in degrees) resulting from the HREMD trajectory for the $A_0^0$ state of C3G and used for the cluster analysis performed to select the dominant conformers. Lower panel: geometries of the four most representative conformers of the same $A_0^0$. Qualitatively similar results hold for the other molecular states investigated here (see Fig. S1, ESI† for analogous distributions for all species).
for the neutral and negative species at shorter wavelengths (λ ≈ 500 nm) and corresponding mostly to a HOMO – 1 → LUMO transition. This shoulder develops into a full-blown secondary peak in A45, which is remarkably visible in experimental spectra from intermediate to high acidity. This indicates that A45 is present in solution at room temperature in this pH range, contrary to the higher stability of A47 predicted by our static computations (see Table 1). Although this discrepancy could be due to the inadequacy of the energy functional or PCM model utilized here, yet unpublished results on the neutral species show that the relative stability of the A40 and A45 protomers is in fact reversed once entropic effects are taken into account, indicating that the difference between the minimum energies computed for different protomers in implicit solvent may not accurately reflect the true relative stability of each species, especially when the differences are small. A small entropic contribution to the free-energy difference between A47 and A45 could be enough to reverse the stability in this case, making A45 dominant, and explaining the prevalence of the two-peak feature in the experimental spectra. We notice that this secondary peak has an outstanding effect on the color expressed, as it subtracts blue components from the light diffused by the solution. We conclude that hindering deprotonation from site 5, e.g., by glycosylation, will likely enhance the blueness of the solution.

Table 1  Energy differences of different protomers of C3G, as predicted by energy minimization at the DFT level in implicit solvent (PCM44)

| Protomers       | Energy differences (kcal mol$^{-1}$) |
|-----------------|--------------------------------------|
| A40 – A45       | −15                                  |
| A40 – A47       | −26                                  |
| A45 – A47       | −5                                   |
| A47 – A45       | −24                                  |

Fig. 4  Experimental spectra at different pH values (top) and calculated for the species expected to be relevant for each pH interval.

(p → π*, accounting for ≈ 90% of the total oscillator strength). We also note the existence of a shoulder in the spectra predicted for the neutral and negative species at shorter wavelengths (λ ≈ 500 nm) and corresponding mostly to a HOMO – 1 → LUMO transition. This shoulder develops into a full-blown secondary peak in A45, which is remarkably visible in experimental spectra from intermediate to high acidity. This indicates that A45 is present in solution at room temperature in this pH range, contrary to the higher stability of A47 predicted by our static computations (see Table 1). Although this discrepancy could be due to the inadequacy of the energy functional or PCM model utilized here, yet unpublished results on the neutral species show that the relative stability of the A40 and A45 protomers is in fact reversed once entropic effects are taken into account, indicating that the difference between the minimum energies computed for different protomers in implicit solvent may not accurately reflect the true relative stability of each species, especially when the differences are small. A small entropic contribution to the free-energy difference between A47 and A45 could be enough to reverse the stability in this case, making A45 dominant, and explaining the prevalence of the two-peak feature in the experimental spectra. We notice that this secondary peak has an outstanding effect on the color expressed, as it subtracts blue components from the light diffused by the solution. We conclude that hindering deprotonation from site 5, e.g., by glycosylation, will likely enhance the blueness of the solution.

Fig. 5  HOMO – 1, HOMO and LUMO energy levels for all five molecular states, along with the frontier orbitals HOMO and LUMO. The HOMO–LUMO gap is here the spacing between the light brown lines while the HOMO – 1 level is drawn as a darker line. Energy level position and shapes of the orbitals remarkably differ in each molecular states.
These data show that deprotonation lifts the HOMO and LUMO orbital energies, in line with chemical intuition that suggest a decrease of the ionization potential, when increasing the negative charge of the molecule. To a lower extent a similar argument also holds for the LUMO, thus leading to a general decrease of the HOMO–LUMO (HL) gap from the positive to the negative states of C3G. Hence, the overall charge state of the molecule markedly affects the C3G optical properties, consistently with experimental evidence. Isosurfaces of the molecular orbitals are also qualitatively different from each other, and match an alternating distribution of bonding and antibonding characters, reflecting rearrangements of electronic distribution upon (de)protonation at different sites. Both the HOMO and the LUMO lie in the plane of the chromophore and extend over both conjugated moieties (AC and B rings); hence we also expect that distortions in the chromophore might decrease its overall conjugation, and result in a shift of the orbital

Fig. 6 Results for $A^+$, $A_4^0$, $A_{4^7}$, and $A_{4^5}$ species of C3G are reported from top to bottom. Columns from left to right: histograms showing the distribution of the $\theta$ dihedral angle for all clusters equilibrated in ab initio MD (AIMD), spectra of each of the conformers, and HOMO–LUMO gaps calculated for frames extracted from AIMD trajectories. All data referring to the $C_1$, $C_2$, $C_3$, and $C_4$ conformers are reported in cyan, blue, pink and orange, respectively.
energies and hence the color of the solution. This is discussed in more detail in the next section.

3.3 Conformational effects

The molecular distortion that mostly affects the optical properties of anthocyanins is the twist between the single (B) and double (AC) rings, described by the dihedral angle \( \theta \). In order to ascertain the effects of this distortion on the main absorption peak, we monitor the HL gap using the PBE functional along various AIMD trajectories. Computing these gaps is much less expensive than performing TDDFT excited-state calculations using the B3LYP hybrid functional, while the value of the gap correlates very accurately with the position of the \( S_0 \rightarrow S_1 \) transition, regardless of the functional used (as shown in the ESI† Section S1 and Fig. S0). The HL gaps were computed with QE in implicit solvent (SCCS) on 1000 geometries extracted from each AIMD trajectory, and were binned according to the \( \theta \) values of the corresponding structure with a bin-width of 5°. The average gap was computed for each bin with its statistical uncertainty (see Fig. 6, third column).

In Fig. 6 we show the distribution of the \( \theta \) dihedral for the five molecular species considered in this work (left panel), along with the average HL gaps recorded within each \( \theta \) bin for each conformer (right panel). The mid panel reports the average spectra recorded for each conformer. We notice that deviation from planarity induces a variation in the HL gaps, however this variation seems to depend on the molecular species. The nature of the frontier orbitals across the different molecular states is crucial to understand such trends. In order to get further insight, we utilize an indicator, \( B_{xy} \), of the bonding character along the C2-C1’ bond, connecting the AC and B rings, in the various frontier orbitals, defined as:

\[
B_{xy}(i) = \frac{|c_{x2} + c_{y1}|}{\sqrt{2(c_{x2}^2 + c_{y1}^2)}}
\]

(1)

where \( c_{xi} \) is the projection of the \( i \)-th molecular orbital \( \{i = \{\text{HOMO}, \text{LUMO}, \ldots \}\} \) over the 2pz atomic orbital of the \( x \)-th C atom \( \{x = \{1', 2\}\} \). The relative sign of \( c_{x2} \) vs. \( c_{y1} \) determines the resulting \( B_{xy} \): \( B_{xy} \sim 0 \) reflects a strong antibonding character, while \( B_{xy} \sim 1 \) corresponds to a fully bonding \( \pi \) orbital. The computed values of \( B_{xy} \) for optimized structures of each species are shown in Table 2. This simple quantity is able to explain the behavior of orbital energies upon distortion around the C2-C1’ bond: we expect the energy of a strongly bonding orbital to increase with distortion, whereas an antibonding orbital should be stabilized by distortion. The resulting effect on the gap depends on the relative variation of the HOMO and LUMO energies. From Table 2 we see that \( A_\pi^1 \), \( A_{\pi 7} \) and \( A_\pi^5 \) have a bonding HOMO \( (B_e = 0.98 \text{ to } 0.77) \) and an antibonding LUMO \( (B_e = 0.37 \text{ to } 0.45) \). Hence we expect to see a strong dependence of the gap on \( \theta \), and in particular that the gap should decrease upon distortion. This is indeed what we see in the right panel of Fig. 6. Conversely, \( A_\pi^4 \) which has a very antibonding HOMO \( (B_e = 0.08) \) and slightly bonding LUMO \( (B_e = 0.65) \), displays the opposite trend, as the gap increases with higher distortion. Finally, \( A_\pi^7 \) is an intermediate case in which the gap does not show a strong dependence on \( \theta \), although due to its slightly antibonding HOMO \( (B_e = 0.44) \) and bonding LUMO \( (B_e = 0.67) \) the trend is nevertheless reversed as in the \( A_\pi^5 \) case. Following this reasoning, we notice that the behavior of the two negative forms should be very similar to each other, as their \( B_{xy} \) values are very close—however the gap’s dependence on \( \theta \) is much more marked in the case of \( A_{\pi 7} \) than in that of \( A_{\pi 5} \).

This partial mismatch can be explained by the proximity of the HOMO – 1 and HOMO energy levels in this species (Fig. 5), which is related to the appearance of a second peak in the visible spectrum and partially undermines the single electron–hole pair picture of the optical transitions. We see that the bonding character depends on the protonation pattern more than the charge states—in fact both neutral protomers \( A_\pi^4 \) and \( A_\pi^7 \) are at the two extreme in bonding character for both HOMO and LUMO, and indeed the gaps clearly show opposite trends. The molecular orbitals reported in Fig. 5 qualitatively match \( B_{xy} \) data. Extra computational details about all these calculations are reported in Section S1 (ESI†).

While the above reasoning explains the dependence of the gap on \( \theta \), the behavior of the different conformers within a given species is a consequence of several competing effects, including electronic conjugation, which favors a planar geometry of the chromophore, and glycosylation at position 3, which exerts a torque between the sugar and the ortho (2’ or 6’) hydrogen on the phenyl B ring. Each molecular state displays a different degree of conjugation, as shown by the variation in bond orders in Table 2, so that the competition results in a different average value and distribution of \( \theta \). The details of such behavior are further discussed in the ESI† (Fig. S4), by explicitly defining other geometrical coordinates which can account for the fine details of the geometrical arrangement of the conjugated and glycosyl units. In addition to all the above, both the average and the spread in the distribution of \( \theta \) are crucially affected by the solvent. The \( C_5 \) and \( C_7 \) conformers are characterized by the presence of an H-bonding chain with two water molecules connecting the hydroxyl at position 3’ with the sugar hydroxyls. The presence of these rather persistent water wires locks in the sugar configuration, thus tuning and locking the \( \theta \) angle. This effect has been confirmed by comparing our explicit-solvent AIMD simulations with similar ones performed in vacuo, as shown in Fig. 7, where results for the \( C_2 \) and \( C_7 \) conformers are reported in blue and pink, respectively, as in Fig. 6. We notice that water has a restraining effect on \( z_6 \), whose values fluctuate more freely in vacuo, thus resulting in wider fluctuations for \( \theta \).
as well. Remarkably, the most likely values of $\alpha_6$ differ in water and in vacuo. This reveals that water molecules shape the conformational landscape of each cyanin species by creating an ordered solvation shell around the hydrophilic moiety of C3G.

3.4 Generalization to other anthocyanins

In an attempt at generalizing the relationship between internal distortions and optical properties we have also considered two other anthocyanins, pelargonin and delphinin, which differ from C3G in the number of hydroxyl groups on the B ring. Geometry optimizations of reduced model systems, with constrained values of $\theta$ (relaxed scan), reveal that, independently of the charge state and of the protomeric form (Fig. 8), the calculated HL gaps show that delphinidin assumes the bluest nuance, followed by cyanidin and pelargonidin, in line with experimental findings and theoretical investigations. Consistently with cyanin, the sensitivity of the HL gap to the $\theta$ distortion is confirmed. Indeed, for the same arguments that hold for C3G, a decrease of the gap along with an
increased internal distortion is again more pronounced in the neutral quinonoid form deprotonated at 4′ position, and damped when deprotonation occurs at positions 7 or 5. Trends on HL gaps quantitatively expressed above in term of $\beta_\theta$ would also explain results reported in Fig. 8. Behaviours are slightly different in $A^\alpha_{5/4}$ because of the already mentioned presence of a close-by second optical transition in the visible, with mixed contributions from (HOMO $\rightarrow$ LUMO) and (HOMO $\rightarrow$ 1 $\rightarrow$ LUMO).

4 Conclusions

In this work we provide the first systematic study of an anthocyanin, notably cyanidin-3-O-glucoside, over a wide (1–9) pH range. Our combination of computer simulations and photo-absorption experiments allow us to shed unprecedented light onto the molecular mechanisms of color expression in this family of natural dyes. Our results, in particular, highlight the prime and hitherto unrecognized role of internal molecular distortions and deprotonation pattern in determining the position of the frontier molecular states, which are in turn responsible for the color optical properties of the dye. The extent of internal distortions is markedly affected by the relative position of the sugar, which critically depends on the H-bond interactions with the water solvent. These findings could be key to guiding industrial efforts to identify natural non-toxic alternatives to some of the artificial colorants currently in use, according to their glycosylation pattern.

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

Enlightening discussions with Alessandro Laio in the first stages of this work are gratefully acknowledged. This work was supported in part by the European Union through the MaX Center of Excellence for supercomputing applications, grant no. 676598 and 824143, and by the HEAD program of the European Social Fund, project no. 2014/2020-25/15, managed by the Italian FVG regional authority, and by Mars Chocolate North America LLC, under the Mars-SISSA collaboration agreement.

Notes and references

1. S. Lehto, M. Buchweitz, A. Klimm, R. Strassburger, C. Bechtold and F. Ulbert, Food Addit. Contam., Part A, 2017, 34, 335–355.
2. K. Yoshida, M. Mori and T. Kondo, Nat. Prod. Rep., 2009, 26(7), 884–915.
3. P. Trouillas, J. C. Sancho-Garcia, V. De Freitas, J. Gierschner, M. Otyepka and O. Dangles, Chem. Rev., 2016, 116(9), 4937–4982.
4. A. Castañeda-Ovando, M. de Lourdes Pacheco-Hernández, M. E. Páez-Hernández, J. A. Rodríguez and C. A. Galán-Vidal, Food Chem., 2009, 113, 859–871.
5. G. T. Sigurdson and M. M. Giusti, J. Agric. Food Chem., 2014, 62(29), 6955–6965.
6. R. Brouillard, in Anthocyanins As Food Colors, ed. P. Markakis, Academic Press, 1982, pp. 1–40.
7. G. Sigurdson, R. Robbins, T. Collins and M. Giusti, Food Chem., 2017, 221, 1088–1095.
8. T. Goto and T. Kondo, Angew. Chem., Int. Ed., 1991, 30, 17–33.
9. X. Ge, I. Timrov, S. Binnie, A. Biancardi, A. Calzolari and S. Baroni, J. Phys. Chem. A, 2015, 119, 3816–3822.
10. I. Timrov, M. Micciarelli, M. Rosa, A. Calzolari and S. Baroni, J. Chem. Theory Comput., 2016, 12, 4423–4429.
11. I. Cacelli, A. Ferretti and G. Prampolini, Theor. Chem. Acc., 2016, 135, 156.
12. K. Sakata, N. Saito and T. Honda, Tetrahedron, 2006, 62, 3721–3731.
13. G. Rastelli, L. Costantino and A. Albasinii, THEOCHEM, 1993, 279, 157–166.
14. G. K. Pereira, P. M. Donate and S. E. Galembeck, THEOCHEM, 1997, 392, 169–179.
15. K. Torskangerpoll, K. J. Børve, O. M. Andersen and L. J. Sæthre, Spectrochim. Acta, Part A, 1999, 55, 761–771.
16. V. Barone, A. Ferretti and I. Pino, Phys. Chem. Chem. Phys., 2012, 14, 16130–16137.
17. D. Jacquemin, E. A. Perpête, I. Ciofini and C. Adamo, Acc. Chem. Res., 2009, 42, 326–334.
18. C. Adamo and D. Jacquemin, Chem. Soc. Rev., 2013, 42, 845–856.
19. G. K. Pereira and S. E. Galembeck, Spectrochim. Acta, Part A, 1998, 54, 339–348.
20. D. Loco, S. Jurinovich, L. Cupellini, M. F. S. J. Menger and B. Mennucci, Photochem. Photobiol. Sci., 2018, 17, 552–560.
21. E. R. Kjellgren, J. M. Haugaard Olsen and J. Kongsted, J. Chem. Theory Comput., 2018, 14, 4309–4319.
22. B. Hess, C. Kutzner, D. van der Spoel and E. Lindahl, J. Chem. Theory Comput., 2008, 4, 435–447.
23. D. Marx and J. Hutter, Ab initio molecular dynamics: basic theory and advanced methods, 2009.
24. E. Runge and E. K. U. Gross, Phys. Rev. Lett., 1984, 52, 997–1000.
The notation $X/\text{Ry}$ for wave functions and $Y/\text{Ry}$ for charge densities and potentials.