Electronic Supplementary Material (ESI) for Physical Chemistry Chemical Physics.

Supporting information for: Unraveling the molecular mechanisms of color expression in anthocyanins

Mariami Rusishvili, Luca Grisanti, Sara Laporte, Marco Micciarelli, Marta Rosa, Rebecca Robbins, Tom Collins, Alessandra Magistrato, and Stefano Baroni*

E-mail: baroni@sissa.it

S1. Extra computational details

Geometry optimization and frontier orbitals calculations were performed at the DFT-B3LYP [S1] level of theory with the `Gaussian09` [S2] package, using the localized basis set 6-311g(d,p), and in implicit solvent (PCM) [S3] using default settings for water solvent. The relative energies of different tautomers of neutral and negative C3G, and for two other members of the anthocyanin family (pelargonidin and delphinidin) are shown in Table S1. Frontier orbital energies and isosurfaces were obtained for the relaxed geometries of C3G through single point calculations. The **Bonding characters** B_π referred to in the main text were also computed from these molecular orbitals.

HL gaps (HOMO-LUMO gaps) were calculated at PBE86 [S4] level in the presence of implicit solvent by extracting 1000 geometries from the trajectories, corresponding to the same frames used in calculating the absorption spectra. These were binned according to

their θ angle, with a bin width of 5 deg. The number of frames falling in each bin is of course θ-dependent and so is the statistical relevance, reflected by the error bars. In this way, HL gaps can be monitored during the AIMD dynamics and computed from many configurations. We justify monitoring these gaps by showing that for molecules with a single, intense, mostly H-L transition there is a linear correlation with the optical gap (vertical excitation energy $S_0 \rightarrow S_1$). Results are reported in Figure S0.

Relaxed scans of the θ angle were performed by fixing θ at different values and optimizing all other degrees of freedom, for all investigated charge states and tautomers of C3Gs as well as for pelargonidin and delphinidin. These were performed with QUANTUM ESPRESSO at the PBE86[S4] level of theory in the presence of implicit solvent.[S5] In these scans the glucosyl moiety was replaced by a methyl group to remove the steric effect of the sugar, while still accounting for its electronic properties. The standard deviation is reported to give an estimate of the error. The results are shown in the main text (Figure 8).

Clustering Analysis Clustering of the replica exchange MD trajectories was done according to the newly developed clustering method introduced in.[S6] After checking that the values of the parameters in a certain range produced qualitatively similar geometries for the cluster centers, with only little adjustment of the populations, a nominal parameter of $d_c = 0.02$ was retained for each analyzed trajectory. This ensured that no more than 10% of the configurations end up unassigned from clusters. Clustering was done according to the θ, α_6 and α_7 dihedral angles of pristine C3G (Figure 1 in the main text, and Figure S1). Histograms representing cluster populations are shown in Figure S2.

Spectrum for each snapshot was obtained convolving TDDFT poles with Lorentzian line-shape of FWHM=0.01 Ry (0.136 eV), according to broadening parameter in `turbo_davidson` package of the QUANTUM ESPRESSO suite.

Table S1: The relative energy (kcal/mol) of different protomers of different ANT secies investigated

Energy differences (kcal/mol)			
	Pelargonin	Cyanin (C3G)	Delphinin
$A^{\circ}_{4'}$- A°_{7}	0.7	-15	-3
$A^{\circ}_{4'}$- A°_{5}	0.8	-26	-3
$A^{-}_{4'7}$- $A^{-}_{4'5}$	0.6	-5	0.5
$A^{-}_{4'7}$- A^{-}_{57}	-4	-24	-8

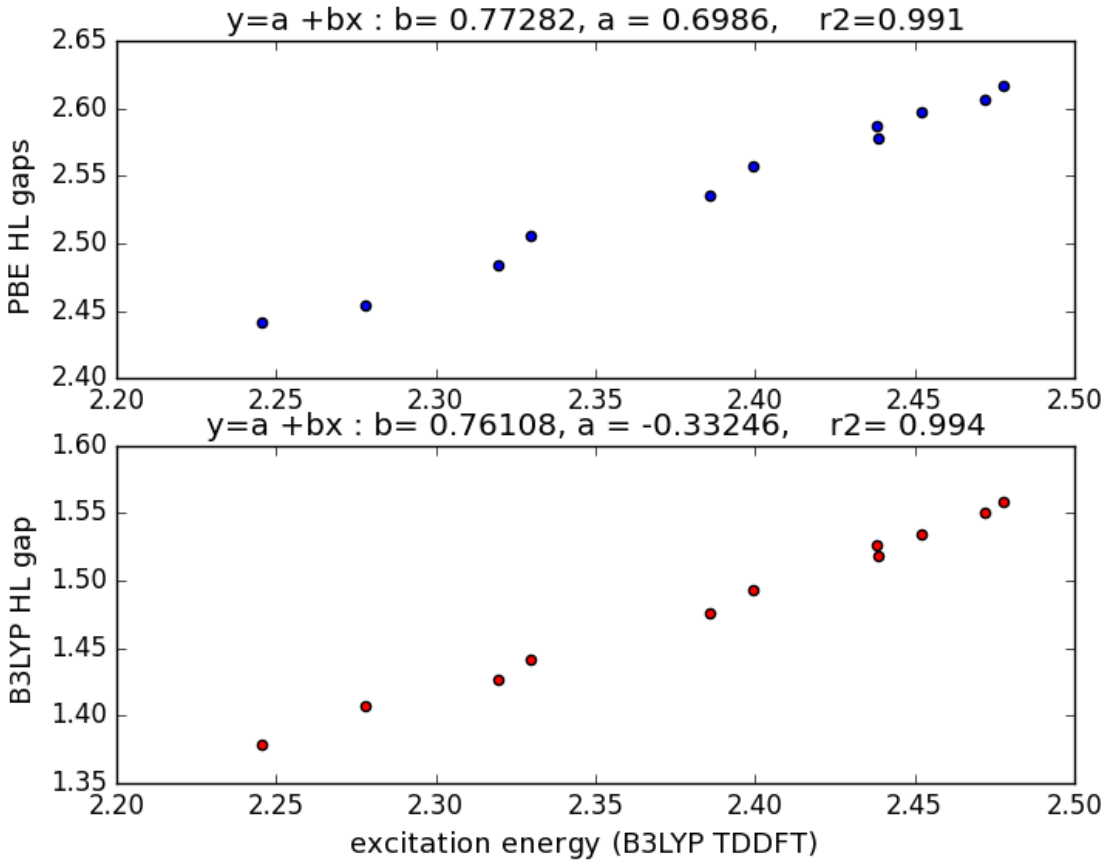

Figure S0: Correlation between HL gaps (calculated at both PBE (top y-axis) and B3LYP level (bottom, y-axis) and optical gap, i.e. excitation energy at TDDFT B3LYP level (x-axis) on structures from the relaxed scan at different values of θ. Linear regression parameters are also reported in the graph headers.

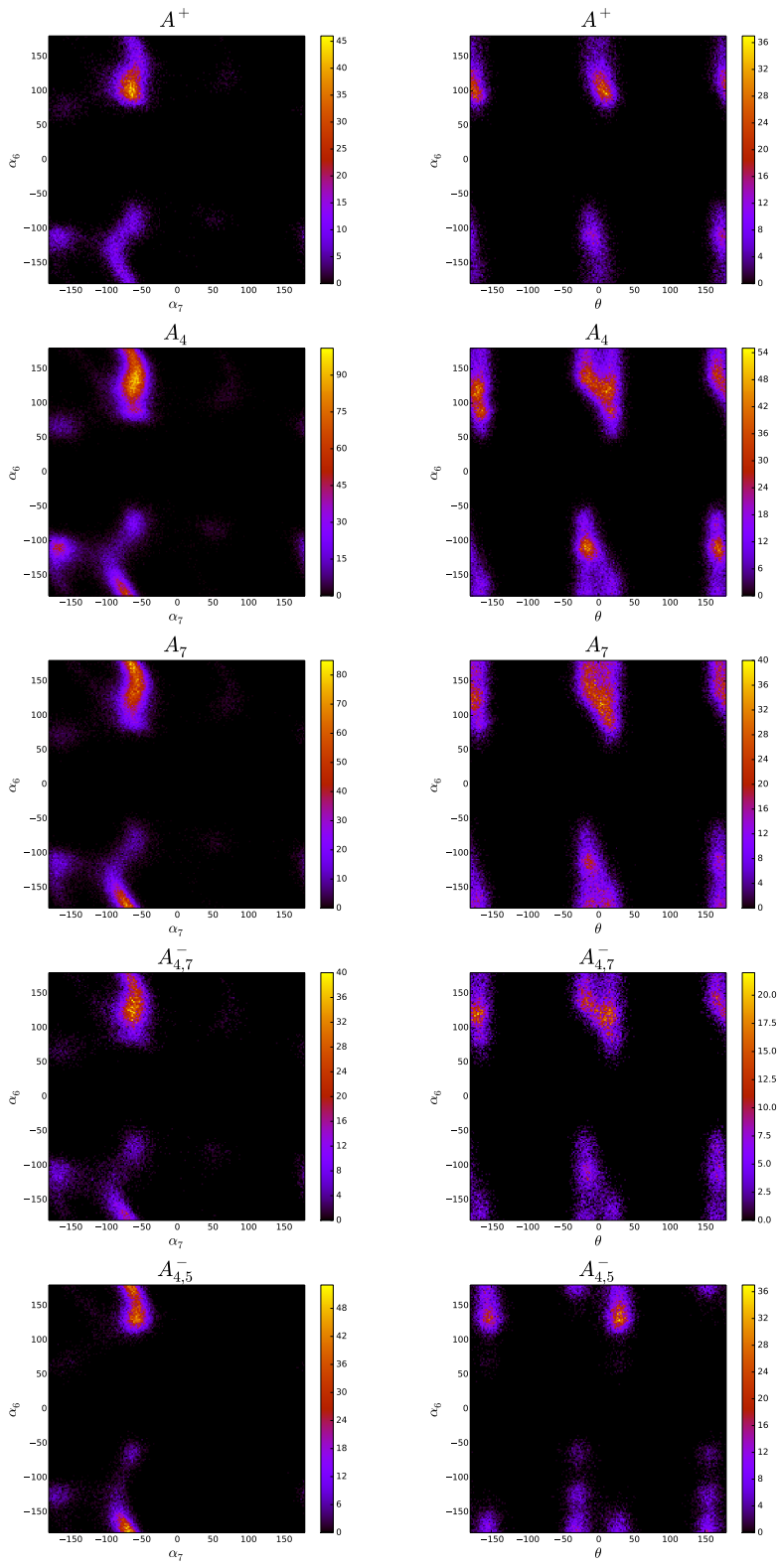

Figure S1: Heatmaps showing the distribution of α_7 vs α_6 and θ vs α_6 as obtained from the classical hamiltonian Replica Exchange molecular dynamics (HREMD) simulations. From top to bottom the molecular states: A^+, $A^\circ_{4'}$, A°_7, $A^-_{4'7}$ and $A^-_{4'5}$ are shown.

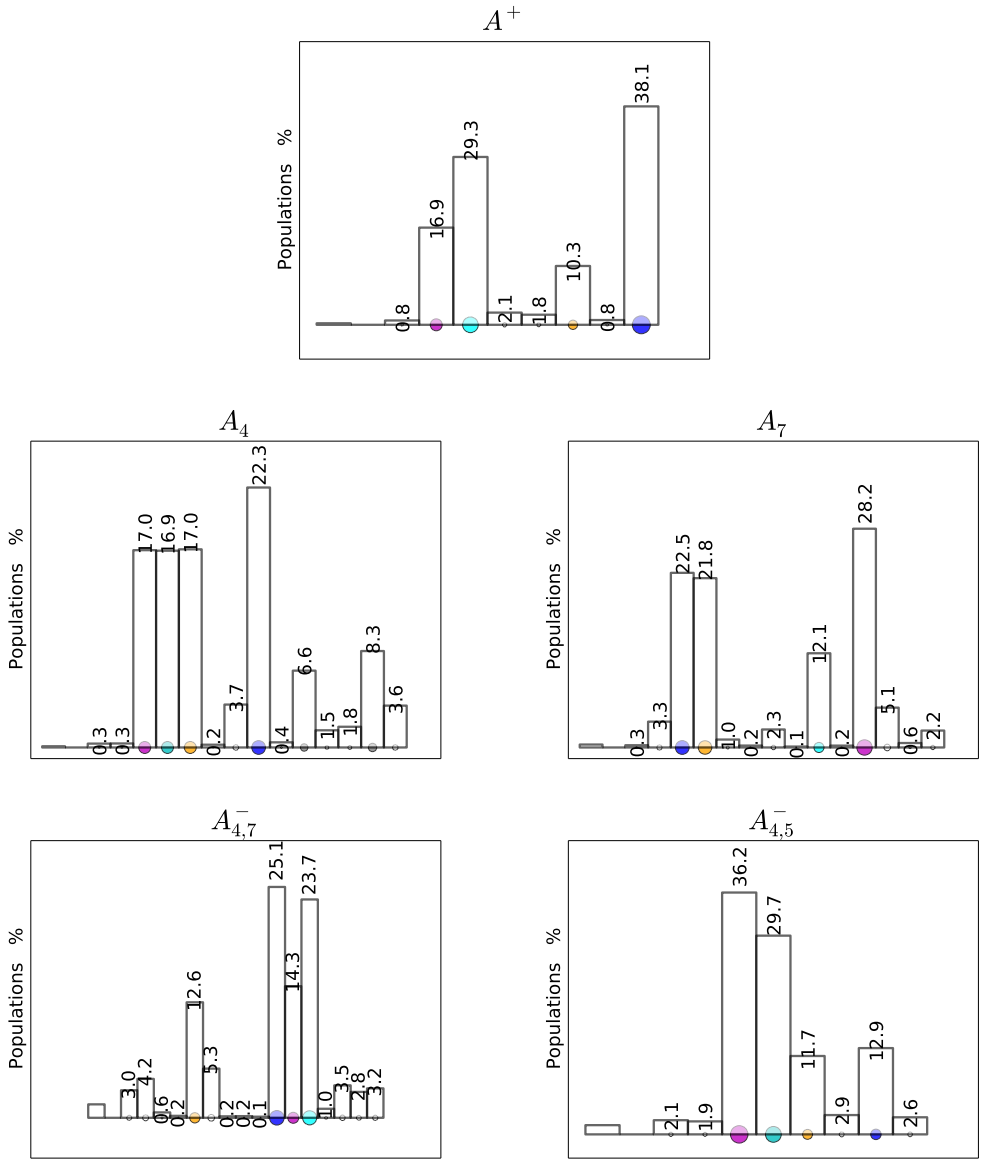

Figure S2: The relative populations (reported as %) of clusters as obtained by the analysis of HREMD trajectory of the A^+, $A^\circ_{4'}$, A°_7, $A^-_{4'7}$ and $A^-_{4'5}$ molecular states. The color code of the circles at the bottom of each bar corresponds to the main conformers: C_1, C_2, C_3 andC_4 as cyan, magenta, blue and orange respectively, and as it is in the main text. Grays are for negligible clusters (< 10%). Relative size of circles also corresponds to populations. Unlabeled separate bar on the left correspond to fraction of data unassigned to any clusters.

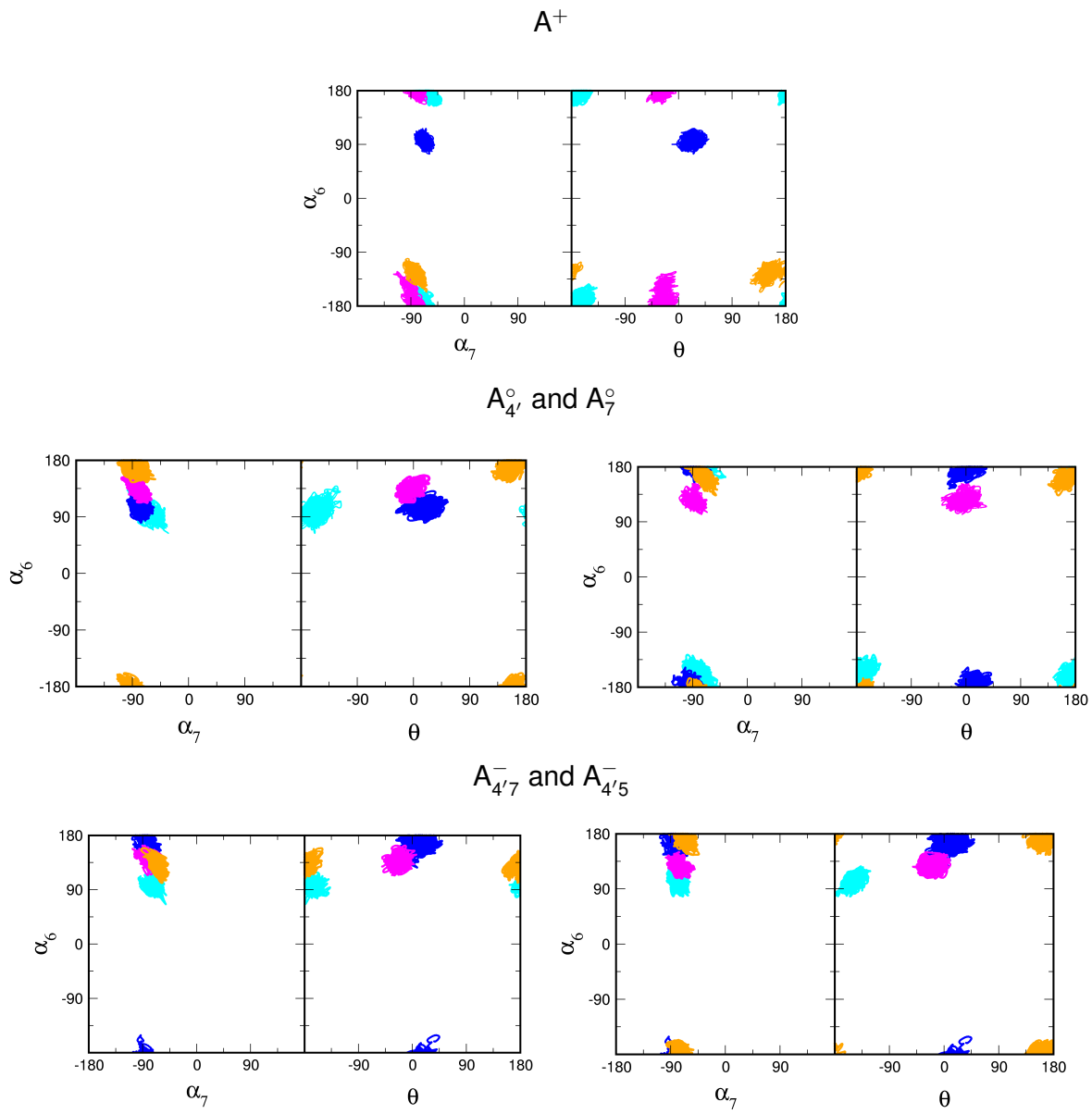

Figure S3: 2D plots of α_6 vs α_7, α_6 vs θ in AIMD calculations (color code corresponds to conformers: C_1 to C_4, cyan, magenta, blue and orange respectively). From top to bottom the molecular states: A^+, $A^\circ_{4'}$, A°_7, $A^-_{4'7}$ and $A^-_{4'5}$ are shown.

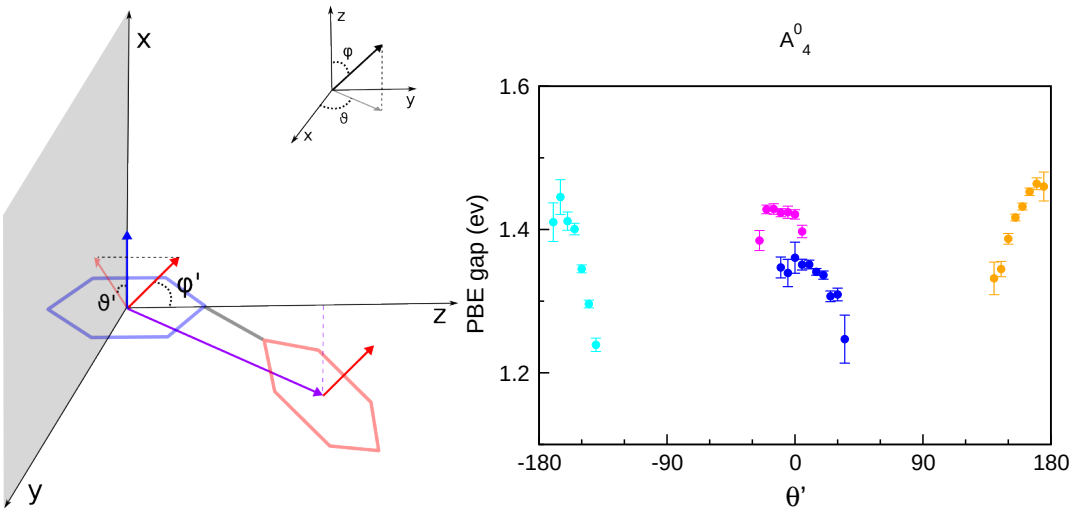

Figure S4: Refinement of geometrical descriptors. θ is defined as a dihedral angle formed by four atoms, and so largely describes torsion between ring planes. However, to properly capture out-of-plane bending, we refine our geometrical description by defining two orthogonal coordinates: θ' and ϕ'. These express the pure torsion between aromatic ring planes and the bending (out-of-plane) angle between plane of the B ring respect to the AC units. Left: diagram of the two contributions to chromophore distortion on rings C (blue) and B (red), with θ' representing a torsion between the two rings, and ϕ' being the out of plane bending contribution to the overall distortion. Right: Evolution of the gaps in the AIMD dynamics for $A^{\circ}_{4'}$binned respect to θ' (for a comparison see second row, right panel of Figure 6 of the main paper).

Table S2: The length of CP trajectories for each anthocyanin species investigated A^+, A_4, A_7, $A^-_{4,7}$ and $A^-_{4,5}$. Extra details about Car-Parinnello AIMD: plane-wave basis cutoffs of 25 and 200 Ry for wavefunctions and density, respectively. The masses of the hydrogen and oxygen atoms were scaled to be equal to the physical mass of carbon atoms (12 u),an effective electronic mass $\mu_e = 600 m_e$MD time step of 5 a.u ($\sim$0.12fs). The simulation box including explicit solvent was pre-equilibrated via classical MD at constant pressure/temperature. AIMD simulations were finally performed at constant volume and (room) temperature using the Nosè-Hoover thermostat[S7] for at least 15 ps per equilibrated trajectory.

Length of CP trajectories (ps)					
Conformer Name	A^+	$A^\circ_{4'}$	A°_7	$A^-_{4'7}$	$A^-_{4'5}$
C_1	**15**	**40**	**30**	**20**	**20**
C_2	**15**	**25**	**15**	**15**	**20**
C_3	**15**	**40**	**25**	**20**	**20**
C_4	**15**	**20**	**15**	**15**	**20**

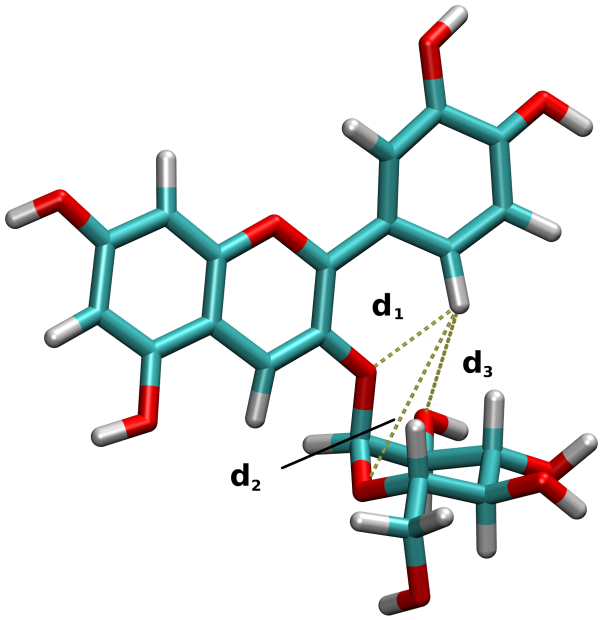

Figure S5: The distances which we examined to understand the asymmetries in the θ distribution among different protomers. (see the table S4 for detailed values of the bond lengths.

Table S3: Number of water molecules persistently H-bonding to each conformer of all anthocyanin species investigated as obtained from H-bonds analysis of the ab initio trajectories. Analysis of the H-bond persistence was done with the `ptraj` module of Ambertools 13. We remark that H-bond persistence may be overestimated due to the scaled massed used in the AIMD simulations.[S8]

Number of H-bonded waters					
Conformer Name	A^{+}	$A^{\circ}_{4'}$	A°_{7}	$A^{-}_{4'7}$	$A^{-}_{4'5}$
C_1	**7**	**6**	**7**	**7**	**7**
C_2	**7**	**7**	**6**	**8**	**7**
C_3	**8**	**6**	**6**	**7**	**6**
C_4	**7**	**6**	**6**	**8**	**8**

Table S4: Average distances and standard deviations (Angstrom) beween H1'/H6' and O1', O2' and O5' of the glycosyl moiety reported in d1, d2,d3, respectively for A^+, $A^{\circ}_{4'}$, A°_{7}, $A^{-}_{4'7}$and $A^{-}_{4'5}$. See figure S5.

	d (Å)	C_1	C_2	C_3	C_4
A^+	d_1	**2.2 ± 0.4**	**2.3 ± 0.4**	**2.3 ± 0.4**	**2.3 ± 0.4**
	d_2	**4.2 ± 0.4**	**3.4 ± 0.6**	**3.2 ± 0.6**	**4.5 ± 0.5**
	d_3	**2.9 ± 0.6**	**4.4 ± 0.6**	**4.5 ± 0.5**	**3.4 ± 0.8**
$A^{\circ}_{4'}$	d_1	**2.3 ± 0.4**	**2.1 ± 0.3**	**2.3 ± 0.3**	**2.3 ± 0.4**
	d_2	**3.2 ± 0.7**	**2.7 ± 0.5**	**2.7 ± 0.7**	**3.3 ± 0.7**
	d_3	**4.3 ± 0.8**	**4.2 ± 0.7**	**4.6 ± 0.4**	**4.3 ± 0.9**
A°_{7}	d_1	**2.2 ± 0.3**	**2.2 ± 0.3**	**2.2 ± 0.3**	**2.1 ± 0.3**
	d_2	**4.3 ± 0.3**	**3.0 ± 0.5**	**4.3 ± 0.5**	**3.4 ± 0.6**
	d_3	**2.9 ± 0.8**	**4.4 ± 0.6**	**3.3 ± 0.6**	**3.8 ± 0.8**
$A^{-}_{4'7}$	d_1	**2.2 ± 0.3**	**2.2 ± 0.3**	**2.3 ± 0.3**	**2.1 ± 0.3**
	d_2	**3.8 ± 0.7**	**2.9 ± 0.6**	**2.8 ± 0.5**	**3.4 ± 0.5**
	d_3	**3.9 ± 0.6**	**4.5 ± 0.5**	**4.4 ± 0.5**	**3.7 ± 0.5**
$A^{-}_{4'5}$	d_1	**2.4 ± 0.4**	**2.2 ± 0.3**	**2.3 ± 0.3**	**2.2 ± 0.4**
	d_2	**3.3 ± 0.7**	**2.5 ± 0.4**	**3.5 ± 0.7**	**3.8 ± 0.5**
	d_3	**4.8 ± 0.7**	**4.4 ± 0.6**	**4.0 ± 0.5**	**3.8 ± 0.7**

Table S5: Most relevant dihedrals (same used for clustering algorithm) for all the protomers: A^{+}, $A^{\circ}_{4'}$, A°_{7}, $A^{-}_{4'7}$and $A^{-}_{4'5}$. Means and standard deviations reported.

state		C_1	C_2	C_3	C_4
A^{+}	θ	$-164^{\circ}\pm25^{\circ}$	$-25^{\circ}\pm25^{\circ}$	$25^{\circ}\pm25^{\circ}$	$152^{\circ}\pm30^{\circ}$
	α_6	$-175^{\circ}\pm30^{\circ}$	$-178^{\circ}\pm25^{\circ}$	$95^{\circ}\pm20^{\circ}$	$-125^{\circ}\pm25^{\circ}$
	α_7	$-65^{\circ}\pm25^{\circ}$	$-85^{\circ}\pm20^{\circ}$	$-68^{\circ}\pm17^{\circ}$	$-80^{\circ}\pm20^{\circ}$
$A^{\circ}_{4'}$	θ	$-153^{\circ}\pm25^{\circ}$	$-10^{\circ}\pm25^{\circ}$	$18^{\circ}\pm30^{\circ}$	$157^{\circ}\pm25^{\circ}$
	α_6	$100^{\circ}\pm30^{\circ}$	$130^{\circ}\pm20^{\circ}$	$105^{\circ}\pm20^{\circ}$	$170^{\circ}\pm25^{\circ}$
	α_7	$-65^{\circ}\pm25^{\circ}$	$-80^{\circ}\pm20^{\circ}$	$-75^{\circ}\pm15^{\circ}$	$-85^{\circ}\pm25^{\circ}$
A°_{7}	θ	$-162^{\circ}\pm25^{\circ}$	$-6^{\circ}\pm30^{\circ}$	$7^{\circ}\pm25^{\circ}$	$170^{\circ}\pm30^{\circ}$
	α_6	$104^{\circ}\pm20^{\circ}$	$175^{\circ}\pm30^{\circ}$	$170^{\circ}\pm30^{\circ}$	$165^{\circ}\pm25^{\circ}$
	α_7	$-65^{\circ}\pm25^{\circ}$	$-85^{\circ}\pm30^{\circ}$	$-87^{\circ}\pm25^{\circ}$	$-70^{\circ}\pm20^{\circ}$
$A^{-}_{4'7}$	θ	$-167^{\circ}\pm25^{\circ}$	$-21^{\circ}\pm25^{\circ}$	$11^{\circ}\pm36^{\circ}$	$177^{\circ}\pm27^{\circ}$
	α_6	$97^{\circ}\pm20^{\circ}$	$-140^{\circ}\pm25^{\circ}$	$165^{\circ}\pm20^{\circ}$	$-130^{\circ}\pm25^{\circ}$
	α_7	$-75^{\circ}\pm20^{\circ}$	$-77^{\circ}\pm23^{\circ}$	$-80^{\circ}\pm22^{\circ}$	$-65^{\circ}\pm15^{\circ}$
$A^{-}_{4'5}$	θ	$-146^{\circ}\pm30^{\circ}$	$-25^{\circ}\pm25^{\circ}$	$27^{\circ}\pm25^{\circ}$	$166^{\circ}\pm26^{\circ}$
	α_6	$-103^{\circ}\pm22^{\circ}$	$125^{\circ}\pm17^{\circ}$	$107^{\circ}\pm20^{\circ}$	$165^{\circ}\pm32^{\circ}$
	α_7	$-77^{\circ}\pm15^{\circ}$	$-75^{\circ}\pm17^{\circ}$	$-53^{\circ}\pm20^{\circ}$	$-70^{\circ}\pm25^{\circ}$

Table S6: Bond order parameters of the C-C bond connecting the single and double rings for A^+, $A^{\circ}_{4'}$, A°_{7}, $A^{-}_{4'5}$and $A^{-}_{4'7}$of Cyanin, Pelargonin and Delphinin.

Bond Order					
	A^+	$A^{\circ}_{4'}$	A°_{7}	$A^{-}_{4'7}$	$A^{-}_{4'5}$
Cyanin	**1.17**	**1.29**	**1.15**	**1.24**	**1.23**
Pelargonin	**1.18**	**1.30**	**1.16**	**1.25**	**1.25**
Delphinin	**1.16**	**1.27**	**1.15**	**1.23**	**1.23**

Table S7: Bond order parameters of the C-C bond connecting the single and double rings for all most representative conformers of A^+, $A^{\circ}_{4'}$, A°_{7}, $A^{-}_{4'7}$ and $A^{-}_{4'5}$. *Bond order* calculations were performed using the Natural Bond Order (NBO) frame on single point `Gaussian09` calculations on selected snapshots along the AIMD dynamics (same used for spectrum).

Bond Order					
Conformer Name	A^+	$A^{\circ}_{4'}$	A°_{7}	$A^{-}_{4'7}$	$A^{-}_{4'5}$
C_1	**1.19**	**1.28**	**1.13**	**1.25**	**1.21**
C_2	**1.16**	**1.29**	**1.16**	**1.22**	**1.22**
C_3	**1.16**	**1.28**	**1.15**	**1.24**	**1.23**
C_4	**1.15**	**1.29**	**1.15**	**1.24**	**1.24**

References

(S1) Stephens, P. J.; Devlin, F. J.; Chabalowski, C. F.; Frisch, M. J. *The Journal of Physical Chemistry* **1994**, *98*, 11623–11627.

(S2) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, J. A., Jr.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, J. M.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, O.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. Gaussian-09 Revision E.01. Gaussian Inc. Wallingford CT 2009.

(S3) Tomasi, J.; Mennucci, B.; Cammi, R. *Chemical Reviews* **2005**, *105*, 2999–3094, PMID: 16092826.

(S4) Perdew, J. P.; Burke, K.; Ernzerhof, M. *Phys. Rev. Lett.* **1996**, *77*, 3865–3868.

(S5) Timrov, I.; Andreussi, O.; Biancardi, A.; Marzari, N.; Baroni, S. *J. Chem. Phys.* **2015**, *142*, 034111.

(S6) Rodriguez, A.; Laio, A. *Science* **2014**, *344*, 1492–1496.

(S7) WG., H. *Phys Rev A Gen Phys.* **1985**, *31*, 1695–1697.

(S8) Feenstra, K. A.; Hess, B.; Berendsen, H. J. C. *Journal of Computational Chemistry* **1999**, *20*, 786–798.