

## Elucidation of the chemical composition of avian melanin

*Suet Yi Liu,<sup>1</sup> Matthew Shawkey,<sup>2</sup> Dilworth Parkinson,<sup>3</sup> Tyler Patrick Troy,<sup>1</sup> Musahid Ahmed<sup>1,\*</sup>*

<sup>1</sup> Chemical Science Division, Lawrence Berkeley National Laboratory, Berkeley, CA, USA

<sup>2</sup> Department of Biology, University of Akron, Akron, OH, USA

<sup>3</sup> Advanced Light Source, Lawrence Berkeley National Laboratory, Berkeley, CA, USA

\* mahmed@lbl.gov

### Supporting Information:

Figure S1 is a schematic of the experimental apparatus.

Description of experimental procedure and statistical analysis.

Figure S2 and S3 shows the laser power and synchrotron energy dependence on the mass spectra for three representative samples for black, brown and grey melanin.

Figure S4 shows the mass spectra of extracted melanin from wild turkey with synchrotron-LDPI at 10.5 eV photon energy.

Figure S5 shows structures for identified masses.

Figure S6 shows the chemical structures for eumelanin and pheomelanin and their fragmentation patterns as identified in the mass spectra

Figure S7 shows the mass spectra of melanin samples from 26 different sets recorded at 10.5 eV synchrotron photon energy

Discussion on the peak probability contrast (PPC) analysis

Figures S8 and S9 shows the results of peak probability contrast analysis on the mass spectra shown in figure S7.

Methodology for the simulation of the absorption spectra for distinct melanin monomer units

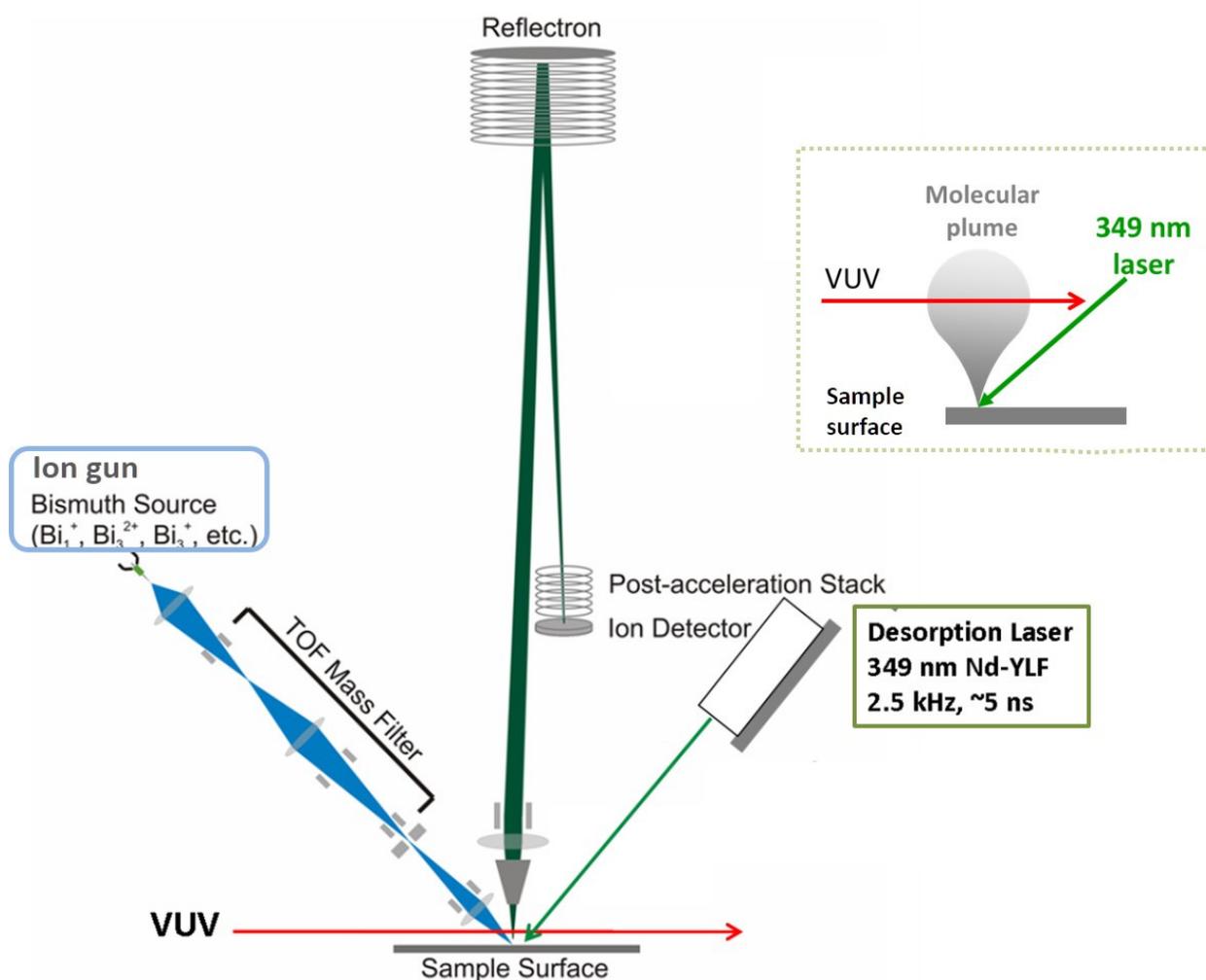
Figure S10 shows the calculated absorption spectrum of selected melanin monomer units.

Table S1 indicates sample abbreviations and classification.

Table S2 indicates the common peak lists observed in the statistical training PPC analysis.

## Methods:

Experimental:



**Figure S1:** Schematic of the experimental apparatus used to measure mass spectra.

The experiments were performed on a modified commercial reflectron-type time-of-flight secondary-ion mass spectrometer (TOF.SIMS V; IonTOF, Germany) coupled to a synchrotron VUV light port and equipped for laser desorption (Figure S1).<sup>[1]</sup> A 349 nm Nd:YLF desorption laser emitting 8.5 ns pulses is focused to a spot diameter of ~30  $\mu\text{m}$  and irradiates the sample

surface at an angle of 45 degrees. Depending on the sample type and analysis desired, various laser peak power densities ranging from  $\sim 0.7 \text{ MW cm}^{-2}$  to  $17.8 \text{ MW cm}^{-2}$  were used. To optimize the signal to noise (S/N) ratio, each data set presented here is the sum of mass spectra collected for  $\approx 16384$  laser shots on the sample surface. To avoid pyrolysis from laser heating and sample damage, mass spectra are typically collected while the sample is linearly scanned by rastering the sample stage at a fixed speed of 2 mm/sec over a 20 mm distance. A neutral molecular plume desorbed by the laser pulse starts to spread perpendicular to the sample surface, and is intersected by the synchrotron VUV beam, which is approximately 50  $\mu\text{m}$  to 150  $\mu\text{m}$  above the sample surface. Samples are held at ground potential, 1.5 mm away from the analyzer extraction cone of the mass spectrometer. The molecules, after being ionized by the VUV light, continue unaltered in their velocity till application of an extraction electrical field pulse. A 3  $\mu\text{s}$ -long -2kV extraction pulse is applied 2  $\mu\text{s}$  after the desorption laser shot. This delay is used to accumulate more ions in the interaction region and eventually obtain a mass spectrum with a better signal-to-noise ratio. The spectral width of the ionizing VUV light is  $\sim 0.2 \text{ eV}$  and when necessary, flux-limiting slits are employed to reduce the photon flux, avoiding detector saturation.

For extant sampling, we first isolate intact melanosomes following the enzymatic extraction method of Liu et al.<sup>[2]</sup> This method has been shown to retain the integrity of the chemical composition of melanin, compared to other harsher extraction methods traditionally employed in melanin extraction. Samples analyzed by VUV-LDPI are prepared on silicon substrates (Wafer World, Inc.; P/N 1183) by directly depositing the melanin sample (about 5 mg) onto the wafer and subsequently dissolving and dispersing the applied compound with high-purity (99.9%) methanol. Samples are allowed to air-dry.

## Statistical Analysis:

From our MS data, a randomized training set consisting of 17 spectra (8 spectra for black color, 5 spectra for brown color and 4 spectra for grey color), and a testing set composed of the remaining 9 spectra was constructed. Table 1 lists all the sample names, training and testing sets. Of the 26 samples, seven were extracted from feathers with some degree of iridescence (see table 1). Iridescence is produced by light scattering from organized arrays of melanosomes.<sup>[3]</sup> No common peaks could be identified in their mass spectra, suggesting that the iridescent effect dominates color composition and hence mask the chemical contribution from melanin. Moreover, this pattern suggests that, despite the similarity in the shapes of melanosomes from iridescent feathers, their chemistry is highly variable. We therefore do not consider them as a separate group in the analyses, and classify them according to their primary (non-iridescent) color.

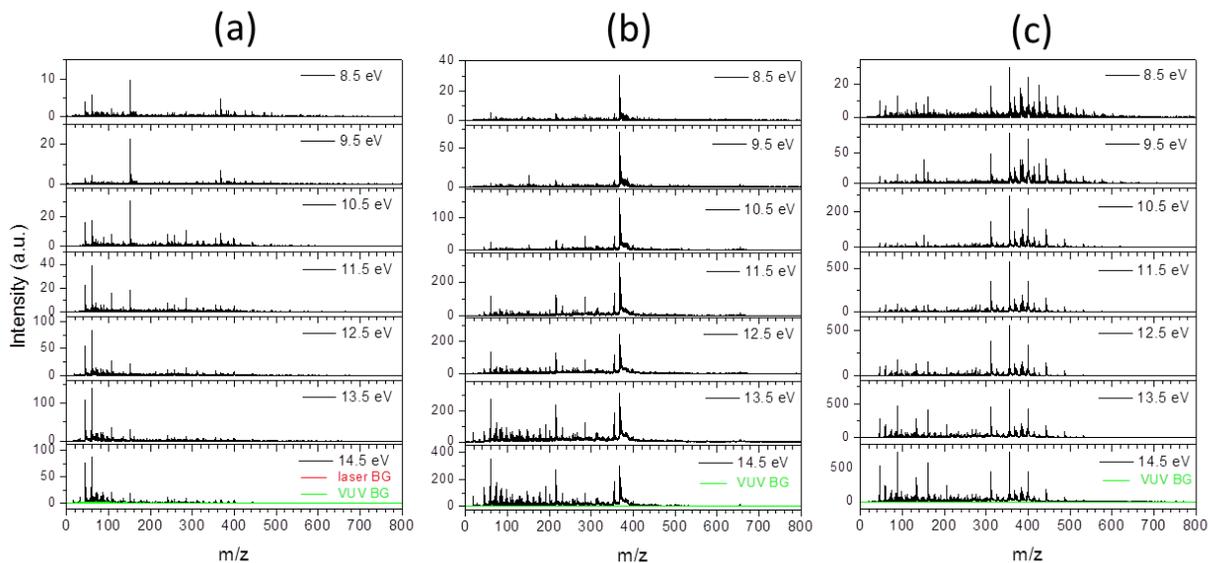
The raw mass spectral data files obtained from the mass spectrometer was imported into Origin 9.0 (OriginLab, Northampton, MS, USA) for data manipulation. The major preprocessing steps prior to classification included intensity normalization and peak picking using the Origin program. Since all 26 samples were measured at various times over two years, the signal to noise ratio of each spectrum differed considerably. Note that the  $m/z$  range was reduced from  $m/z$  1 to 700 to  $m/z$  200 to 700 since most of the characteristic peaks appear above  $m/z$  200. To avoid inference from the noise during the classification procedure, the signal mass peaks in each spectrum were picked using the “find peak” function in Origin software (see supplementary Fig. S5), and the resulting peak lists were used in PPC analysis. Almost all signal peaks in each mass spectrum were used in statistical analysis (Figure S5).

A number of methods are available for mass spectral classification. We selected Peak Probability Contrasts (PPC) because while it has performance comparable to other leading methods, it yields results that directly support conclusions about the underlying chemistry. It does this by providing a ranked list of mass peaks in the order in which they are useful for distinguishing the different classes of samples. (Supplementary figure 6 & 7) PPC is robust to variations in  $m/z$  locations for different samples, which can be due to biochemical or experimental reasons. And because peak detection is carried out before PPC, automated peak identification routines can be used, while still providing an easy to implement opportunity to manually identify and remove spurious peaks that are due to sample preparation or experimental system problems, before beginning the classification analysis.

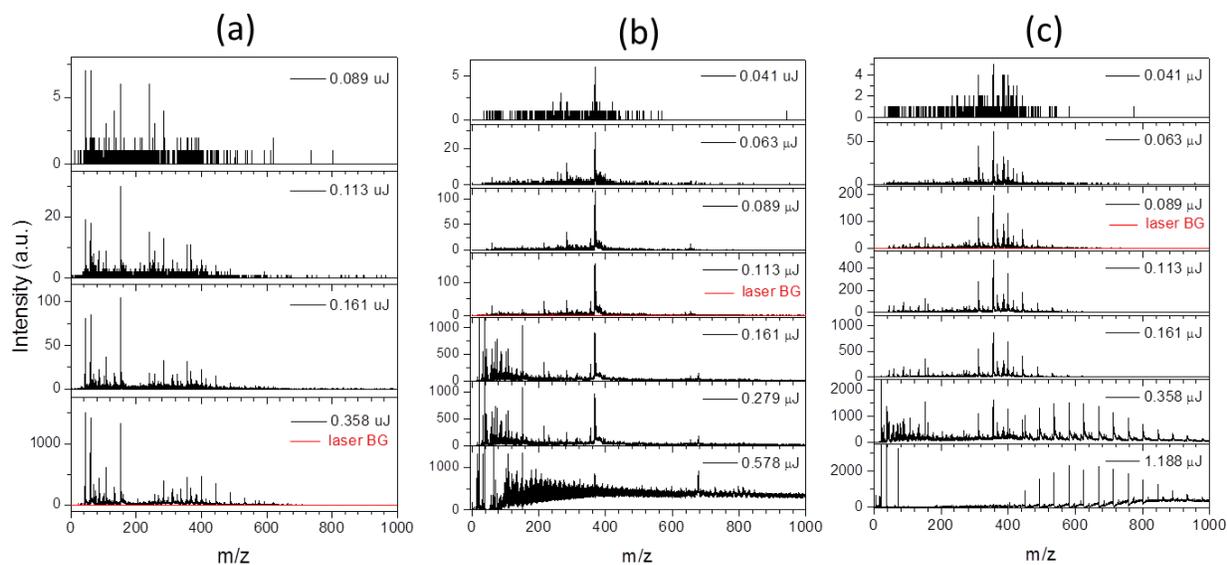
PPC was performed on the training set using the R package (A software environment for statistical computing and graphics) in the R programming language, yielding 33 common peaks in the training set (see supplementary Table S1). Some mass peaks in the training set are common to all three colors but the proportions differ (see supplementary Table S1).

## Determination of optimum laser power and synchrotron energy for photoionization

After the threshold laser power was determined, (see supplementary figure S2 and S3) mass spectra for all melanin samples were collected at different photon energies from 8.5 eV to 14.5 eV in 1 eV intervals. Signals at 10.5 eV provide the most reasonable signal-to-noise ratio and minimal fragmentation.



**Figure S2:** Mass spectra of (a) CHIP (b) JNCO and (c) CHIX at various synchrotron photon energies ranging between 8.5 and 14.5 eV. The bottom panel shows the synchrotron only background in green at 14.5 eV. The laser power intensities are (a) 0.111  $\mu\text{J}$  (b) 0.113  $\mu\text{J}$  (c) 0.089  $\mu\text{J}$  on a diameter of 30  $\mu\text{m}$  spot size.



**Figure S3:** Mass spectra of (A) CHIP (B) JNCO and (C) CHIX at various laser desorption energies. The bottom left panel shows the laser background only in red. The laser power intensities are marked in the figure. The laser spot size is 30  $\mu\text{m}$  diameter.

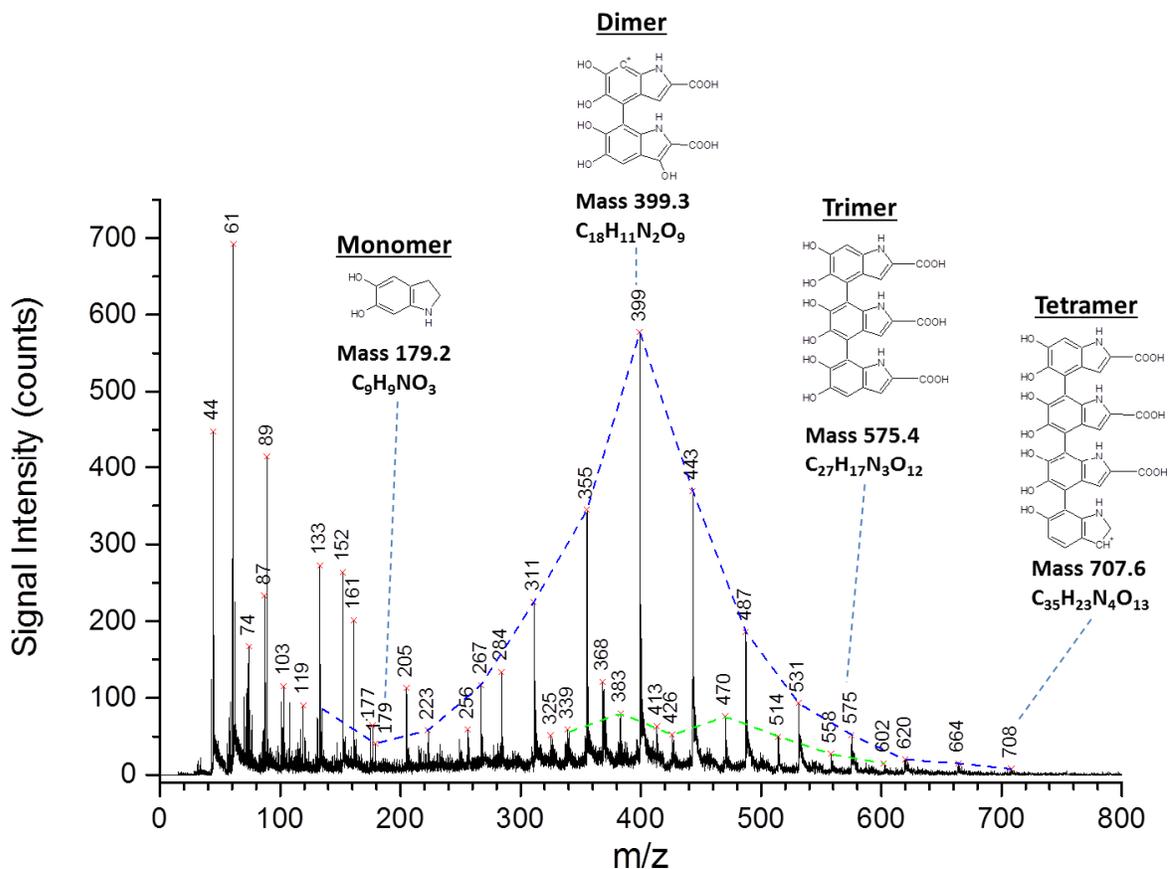


Figure S4. Mass spectra of extracted melanin from wild turkey with synchrotron-LDPI at 10.5 eV. Blue and green dash lines indicated two sets of CO<sub>2</sub> loss sequence, and these two sets have a mass different of 17 Da which corresponds to the loss of OH group.

Figure S4 shows mass spectra of extracted melanin from wild turkey (*Melleagris gallopavo*) with synchrotron-LDPI at 10.5 eV. The synchrotron-LDPI spectra of melanin exhibit clear and discrete peaks with a mass interval of 44 Da (loss of CO<sub>2</sub> group) in the mass region between mass-to-charge ratios (m/z) of 170–710. Two sets of CO<sub>2</sub> loss sequences are indicated as blue and green dashed lines in figure 1, and these two sets have a mass difference of 17 Da which correspond to the loss of a hydroxyl (OH) group. Both sets are expected for a macromolecule composed of various redox forms of 5,6-dihydroxyindole (DHI) and 5,6-dihydroxyindole-2-

carboxylic acid (DHICA), which are the two main eumelanin precursors (see supplementary Fig. S5).<sup>[4]</sup> The tetramer observed in the spectrum could be due to small oligomeric units formed by breaking the  $\pi$ - $\pi$  bonding between the indole layers, supporting the assumption of 4-6 monomers composed of one oligomeric unit.<sup>[5]</sup> This spectral pattern (sequential loss of CO<sub>2</sub> fragment) could be a signature of eumelanin since similar spectral patterns were also observed in black feathers considered in this study.

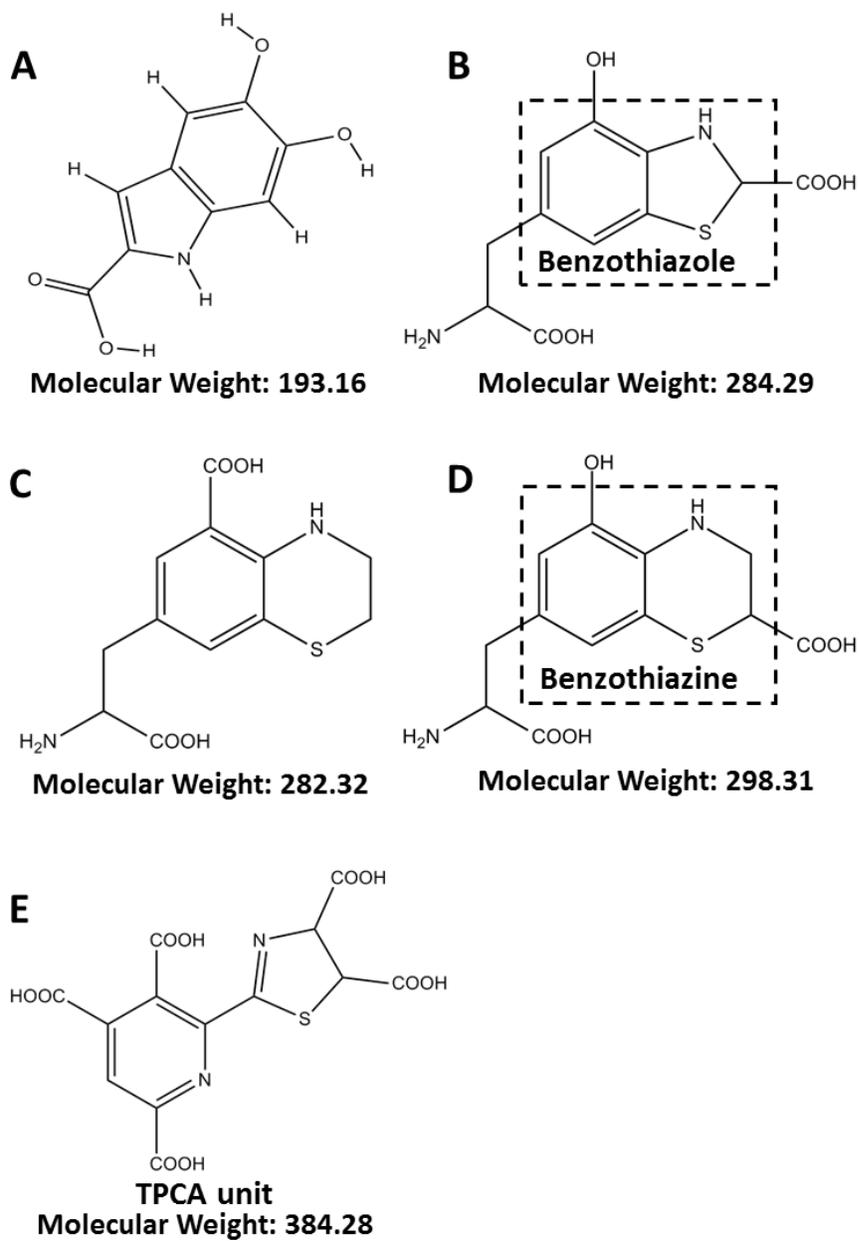
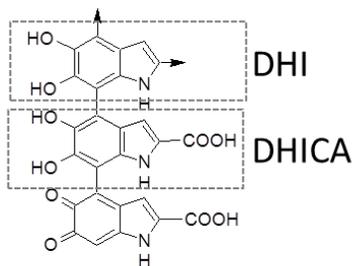
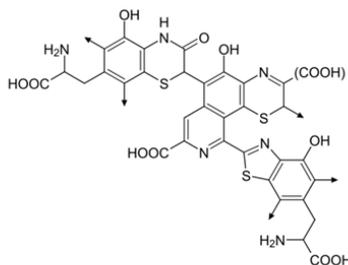


Figure S5. Structures proposed for indicated masses.

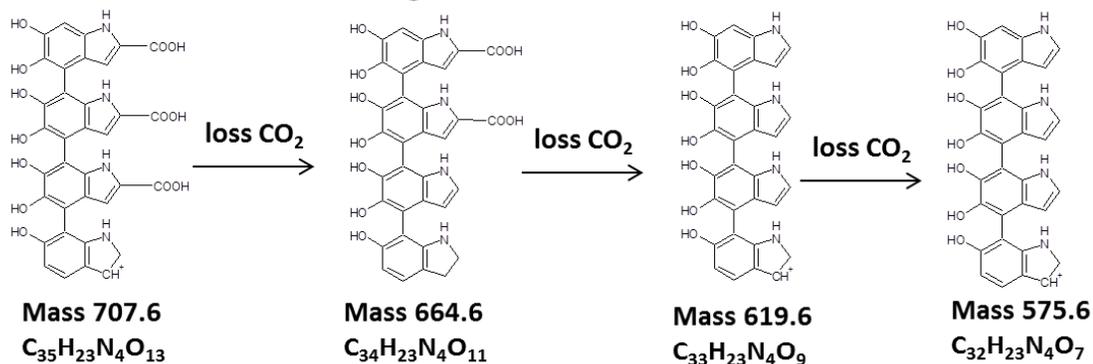
**(a) Eumelanin**



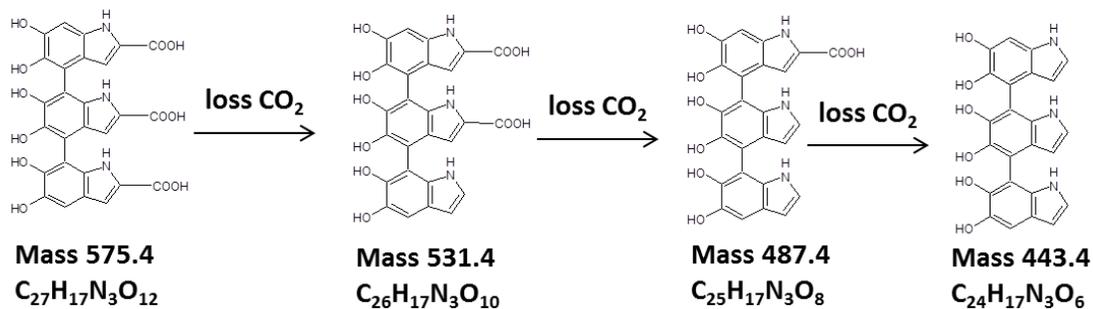
**Pheomelanin**



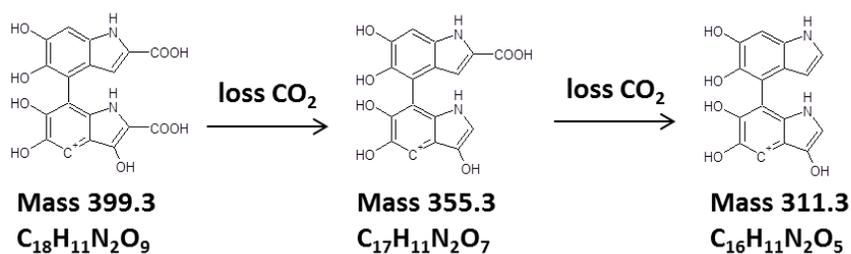
**(b) DHICA tetramer and its fragments**



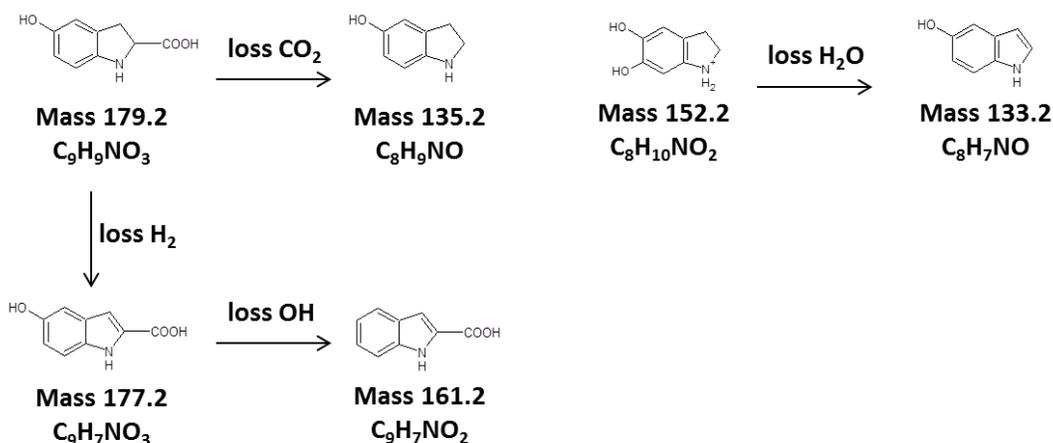
**DHICA trimer and its fragments**



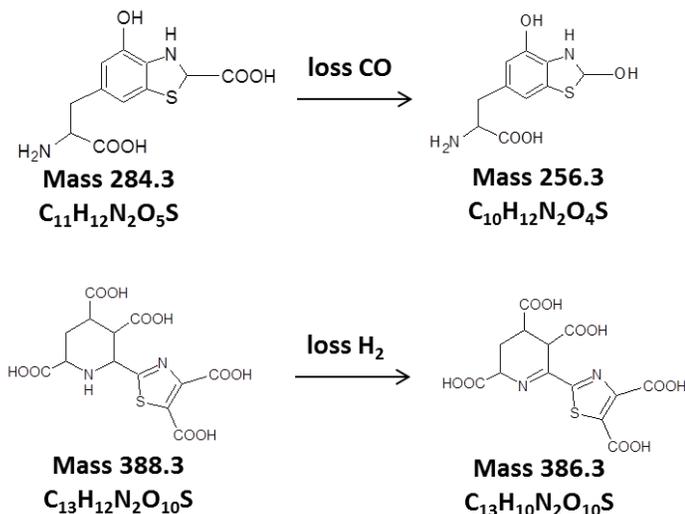
**DHICA dimer and its fragments**



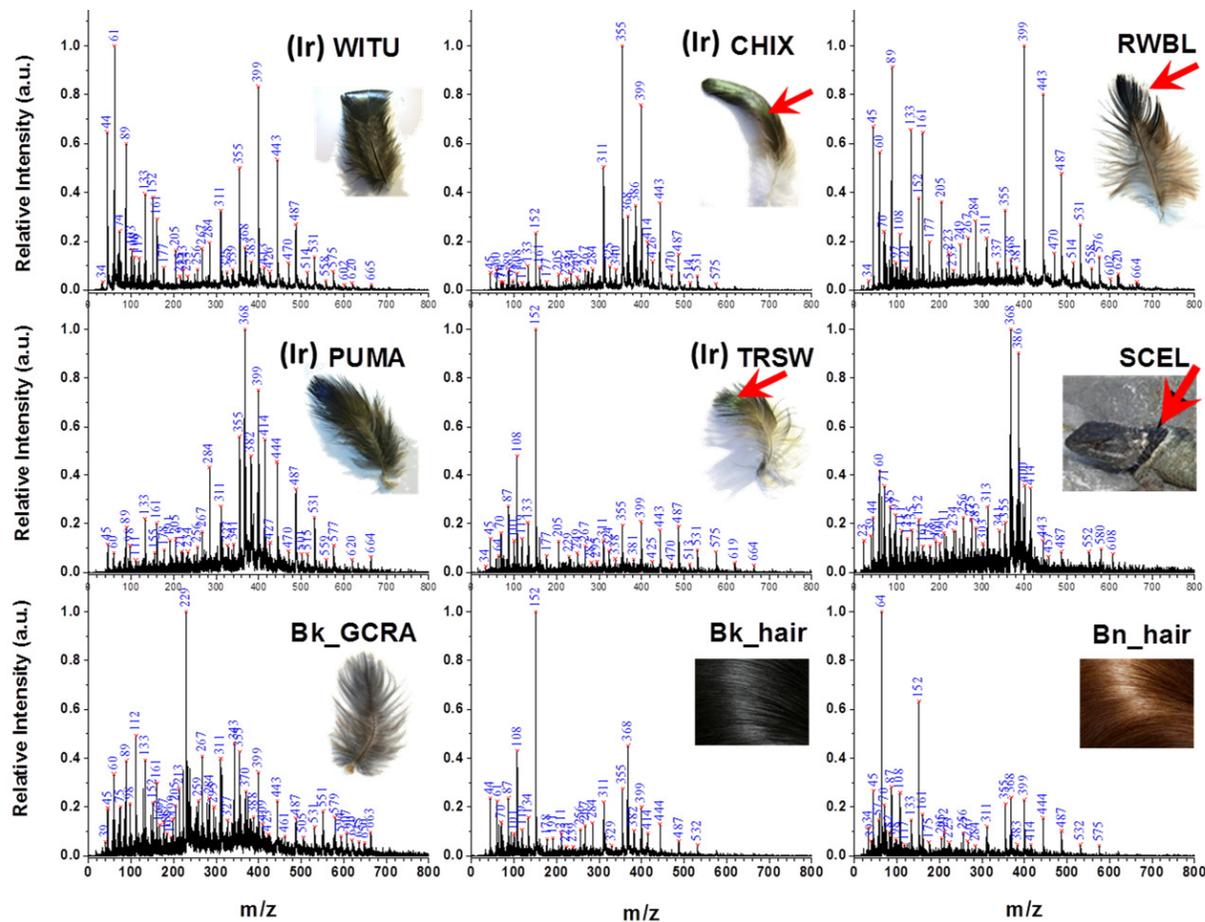
### Monomer and its fragments



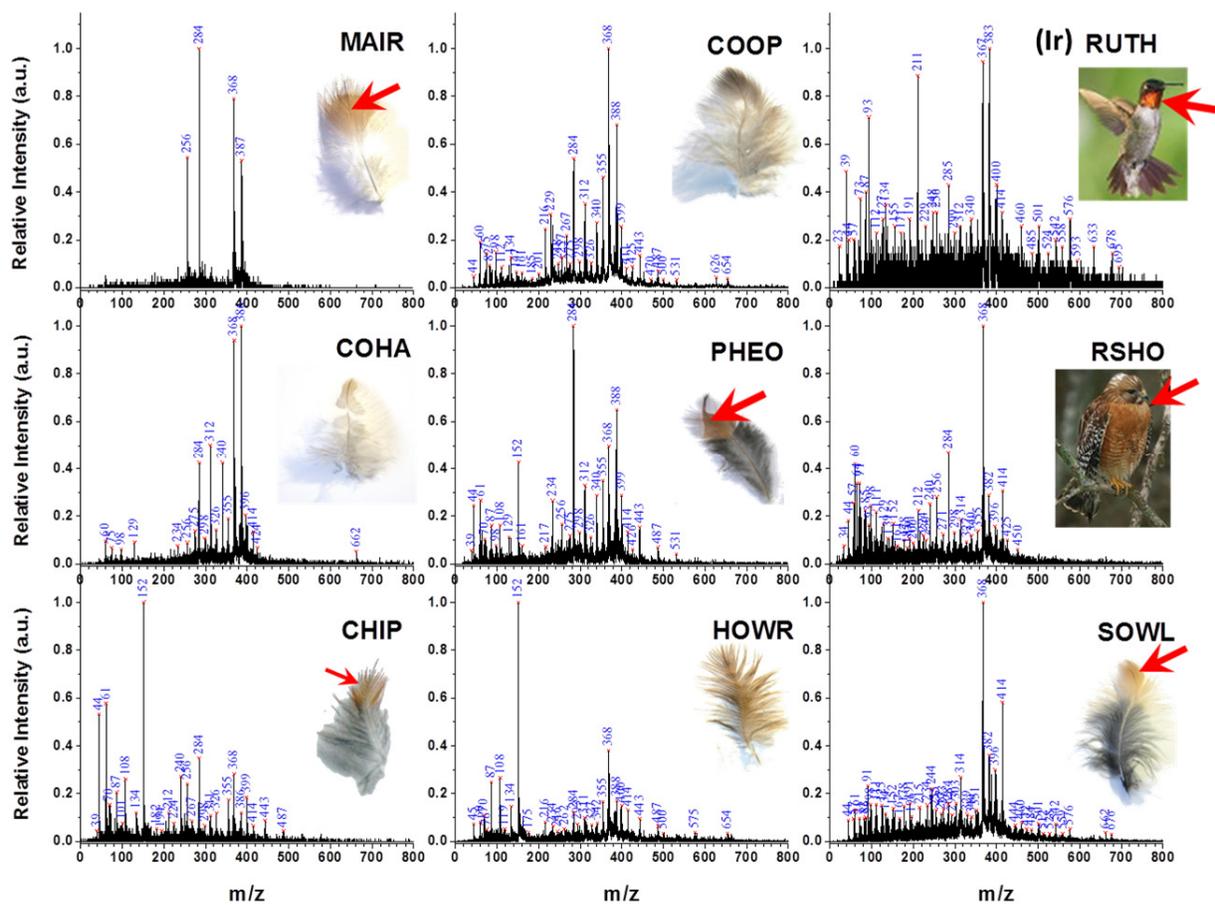
### (c) Degradation product of pheomelanin



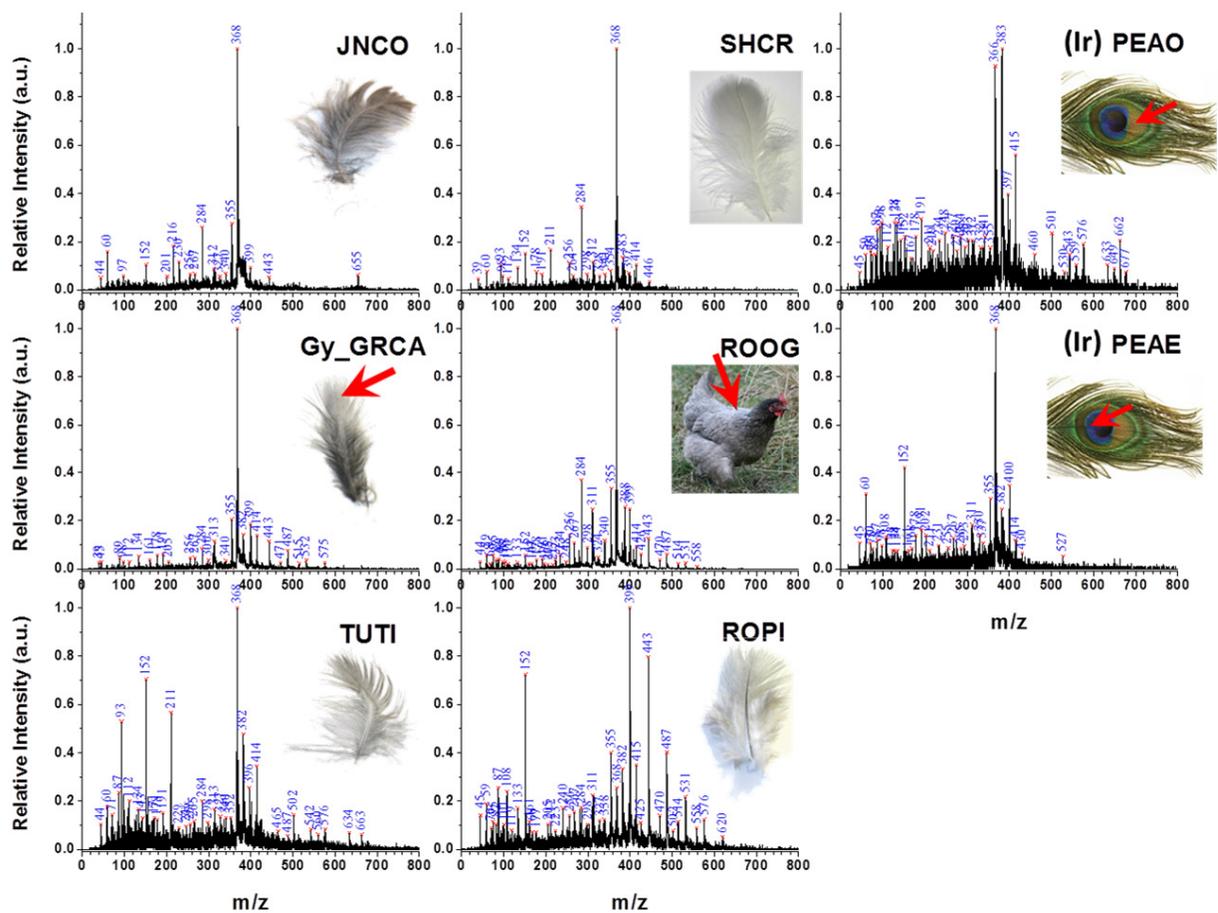
**Figure S6:** (a) General proposed structure of eumelanin and pheomelanin: Eumelanin polymers have been thought to comprise numerous cross-linked 5,6-dihydroxyindole (DHI) and 5,6-dihydroxyindole-2-carboxylic acid (DHICA) polymers, and the oligomer structure of pheomelanin have been thought to incorporate benzothiazine and benzothiazole units that are produced, instead of DHI and DHICA. The lists of the possible peak assignments based on those found in (b) black color feather (wild turkey), (c) brown color feather (mallard).



**Figure S7: A – black**, Mass spectra of 9 samples at 10.5 eV photon energy. The intensity of all mass spectra was normalized. Ir in ( ) indicated iridescent feather.



**Figure S7: B- Brown.** Mass spectra of 9 samples at 10.5 eV photon energy. The intensity of all mass spectra was normalized. Ir in () indicated iridescent feather.

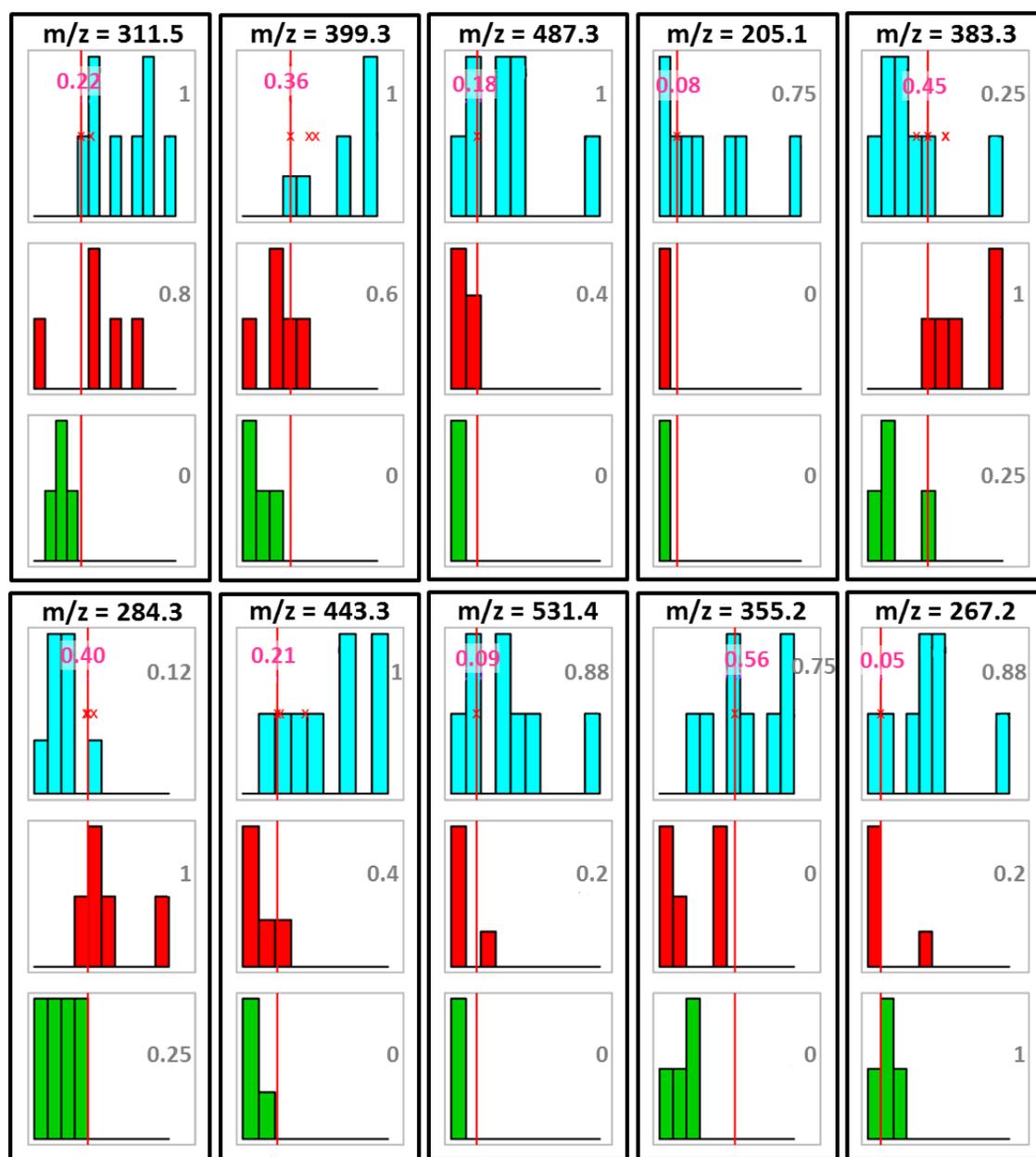


**Figure S7: C- Grey and Peacock.** Mass spectra of 8 samples at 10.5 eV photon energy. The intensity of all mass spectra was normalized. Ir in ( ) indicated iridescent feather.

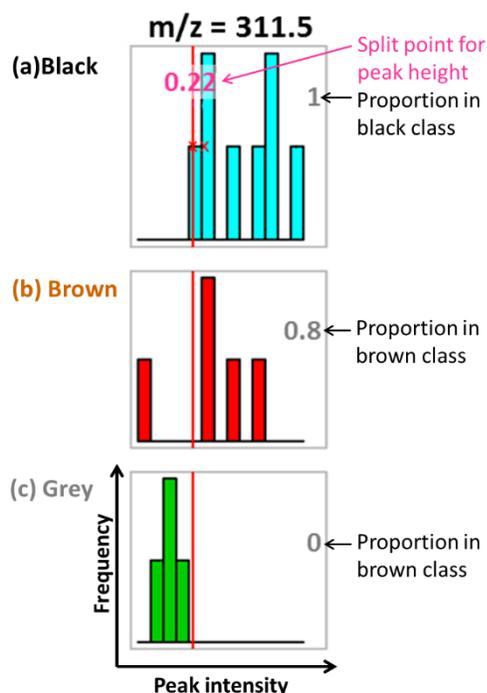
### **Description of the PPC process.**

The procedure for PPC is described in detail in reference 6, but we summarize the procedure here. We performed the following steps:

- a. Find peaks in the mass spectra (this is independent of PPC). PPC takes as input the list of peak heights and positions for each sample.
- b. For the training data set, estimate the common set of peaks via hierarchical clustering. The size of the  $m/z$  range that can be counted as a single cluster is an adjustable parameter. We used a range plus or minus 3% around each  $m/z$  value. For each cluster, a “center”  $m/z$  value is determined.
- c. From the list of common peaks, for each individual data set determine which peaks are present (in our case, if a data set’s peak was within 3% of a common, the peak was counted).
- d. For each peak from the common peaks list, find the peak height or area cutoff that maximally discriminates between the different groups in the training set. It is possible to rank the peaks in order of discriminating power. Some of the peaks will be 100% useful (all of the spectra in the training set for one group fall above the cutoff, all the spectra in the other groups fall below the cutoff), while other peaks may be more ambiguous. The list of peaks that are most useful for classification helps draw conclusions about the chemistry underlying the differences.
- e. With common peaks and cutoff points determined, new spectra can be analyzed and their classification can be determined. The “answer” for which group the new spectrum falls in is done by a nearest shrunken centroid method.



**Figure S8:** Results of PPC method on the extracted melanin samples. Each panel shows a histogram of peak heights in the training set at one mass peak ( $m/z$  value in black type in top), for black feather (blue), brown feather (red) and grey feather (green). The peaks are ordered from strongest to weakest, as measured by the difference in proportions (grey type in middle), starting in the top left corner and moving down the left column. Only the top 10 peaks are shown out of a total of 33 peak sites. The split point for peak height was shown as pink type.



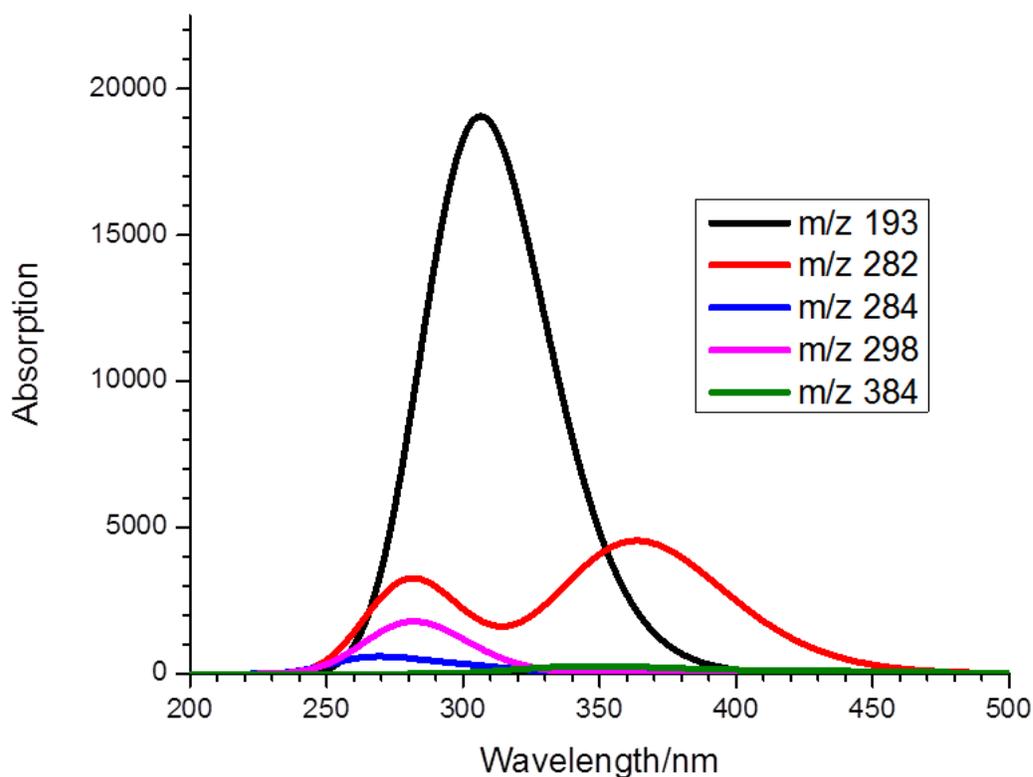
**Figure S9:** Exploded view of the top left panel of figure S7 (the histogram for mass 311 Da).

The vertical red line shows the estimated optimal height split point and the value of split point is shown as pink type. The proportions of samples in each class having peaks higher than the split point is indicated as grey text. The horizontal axis is the peak intensity, and the vertical axis is frequency. For example, in our training set, there are 8 samples in black color class, 5 samples in brown color class and 4 samples in grey color class. (a) shows 6 bars, which means 6 different peak heights were found in 8 black samples, with two of the bars representing two samples with same peak height (to within the size of the histogram bin), and the peak heights of  $m/z$  311.5 in all 8 black samples are higher than the split point 0.22, therefore the proportion of  $m/z$  311.5 in the black class is 1. (b) shows that for the 5 brown samples, only 4 samples have peak heights higher than split point, therefore the proportion in brown class is 0.8. (c) shows no peak height of  $m/z$  311 are higher than split point in all 4 grey samples, therefore the proportion in grey class is 0. Peaks that are important for the “signature” of a particular group have the characteristic that

most of the samples from a given group are either above or below the split point. In this case, 100% of black samples were above the split point, 100% of gray samples were below the split point, and 80% of brown samples were above the split point, meaning this m/z value is useful for distinguishing the appropriate group to which a sample belongs to.

## Computational Methods

The monomers depicted in figure 4 were treated computationally in order to simulate their absorption spectra. In these calculations, density functional theory (DFT) as implemented by the Gaussian 09 package<sup>7</sup> was used to generate optimized geometries of the structures A-E obtained at the B3-LYP/6-31+G(d) level. A frequency analysis at the same level of theory yielded real frequencies thus validating the relaxed geometries. Those geometries were then treated using time-dependent density functional theory TD-DFT in order to determine their first root excited-state energies from which the absorption spectra were simulated as depicted in figure S8.



**Figure S10:** Calculated absorption spectra for structures shown in figure S5.

**Table S1: Samples from which melanin was extracted and analyzed with synchrotron-LDPI**

Sample no.	Abbreviation	Full name	Training/Testing set
1	Bk_hair	Black hair	Training (Black)
2	RWBB	Red winged blackbird	Training (Black)
3	SCEL	Sceloporus lizard	Training (Black)
4	CHIX	Chicken (Ir)	Training (Black)
5	PUMA	Purple martin (Ir)	Training (Black)
6	TRSW	Tree swallow (Ir)	Training (Black)
7	WITU	Wild turkey (Ir)	Training (Black)
8	Bn_hair	Blonde hair (eumelanin)	Training (Black)
9	RUTH	Ruby throated hummingbird (Ir)	Training (Brown)
10	COHA	Cooper's hawk	Training (Brown)
11	COOP	Cooper's hawk	Training (Brown)
12	HOWR	House Wren	Training (Brown)
13	MAIR	Mallard	Training (Brown)
14	Gy_GCRB	Grey catbird (body)	Training (Grey)
15	JNCO	Junco	Training (Grey)
16	SHCR	Sandhill crane	Training (Grey)
17	TUTI	Tufted titmouse	Training (Grey)
18	Bk_GRCA	Grey catbird (head)	Testing (Black)
19	CHIP	Chipping sparrow	Testing (Brown)
20	PHEO	Light brown chicken	Testing (Brown)
21	SOWL	Screech owl	Testing (Brown)
22	RSHO	Red-shouldered hawk	Testing (Brown)
23	ROPI	Rock Pigeon	Testing (Grey)
24	ROOG	Rooster grey	Testing (Grey)
25	PEAE	Peacock-inner bluish part (Ir)	Testing (Iridescent)
26	PEAO	Peacock-outer brownish part (Ir)	Testing (Iridescent)

\*Ir in () indicates iridescent feather. The training and testing set for PPC analysis are also shown.

**Table S2:** The common peaks list observed in training set (total 17 spectra) using PPC analysis.

# of peak	Peak position	Number of spectra	Proportion in Black training set	Proportion in Brown training set	Proportion in Grey training set
1	205.1	7	0.75	0	0
2	212.4	7	0.5	0.2	0.5
3	223.2	5	0.625	0	0
4	233.3	8	0.625	0.4	0.25
5	239.3	1	0.125	0	0
6	248.2	7	0.25	0.6	0
7	256.3	13	0.375	0.6	0
8	267.2	12	0.875	0.2	1
9	275.3	3	0.125	0.4	0
10	284.3	17	0.125	1	0.25
11	294.2	6	0.125	0.6	0.5
12	298.5	7	0.125	0.6	0.75
13	311.5	16	1	0.8	0
14	326.4	11	0.625	0.6	0.25
15	340.3	14	0.25	0.8	0
16	355.2	15	0.75	0	0
17	368.4	16	0.5	0.6	1
18	383.3	16	0.25	1	0.25
19	399.3	16	1	0.6	0
20	414.3	13	0.625	0.4	0.25
21	425.8	6	0.5	0	0
22	443.3	13	1	0.4	0
23	470.3	8	0.5	0	0
24	487.3	13	1	0.4	0
25	507.4	10	0.625	0.6	0.25
26	531.4	10	0.875	0.2	0
27	542.4	2	0	0.2	0.25
28	558.4	6	0.5	0.2	0.25
29	575.4	11	0.75	0.4	0
30	602.5	4	0.375	0.2	0
31	623.1	5	0.5	0.2	0
32	663.5	9	0.25	0.6	0.5
33	686.3	1	0	0.2	0

### Supplementary references:

- [1] (a) L. K. Takahashi, J. Zhou, K. R. Wilson, S. R. Leone, M. Ahmed, *J. Phys. Chem. A* **2009**, *113*, 4035-4044; (b) O. Kostko, L. K. Takahashi, M. Ahmed, *Chem. Asian. J.* **2011**, *6*, 3066-3076.
- [2] Y. Liu, V. R. Kempf, J. B. Nofsinger, E. E. Weinert, M. Rudnicki, K. Wakamatsu, S. Ito, J. D. Simon, *Pigment Cell Res.* **2003**, *16*, 355-365.
- [3] M. D. Shawkey, A. M. Estes, L. Siefferman, G. E. Hill, *Biol. J. Linn. Soc.* **2005**, *84*, 259-271.
- [4] S. Ito, K. Wakamatsu, M. d'ischia, A. Napolitano, A. Pezzella, in *Melanins and Melanosomes*, Wiley-VCH Verlag GmbH & Co. KGaA, **2011**, pp. 167-185.
- [5] P. Meredith, T. Sarna, *Pigment Cell Res.* **2006**, *19*, 572-594.
- [6] R. Tibshirani, T. Hastie, B. Narasimhan, S. Soltys, G. Shi, A. Koong, Q.T. Le. *Bioinformatics* **20**, 3034-3044 (2004).
- [7] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. A. Montgomery Jr., J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J. M. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, O. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski and D. J. Fox, *Gaussian 09, Revision A. 1*, Gaussian, Inc., Wallingford, CT, 2009.