Structure, stability and optical absorption of a comprehensive li-brary of minimal eumelanin oligomer models

Jun Wang,[a,b] Lluís Blancafort[b,*]

[a]Jiangsu Key Laboratory for Chemistry of Low-Dimensional Materials, Jiangsu Engineering Laboratory for Environment Functional Materials, School of Chemistry and Chemical Engineering, Huaiyin Normal University, No.111 West Changjiang Road, Huaian 223300, Jiangsu Province, PR China.
[b]Institut de Química Computacional i Catàlisi and Department de Química, Universitat de Girona, Facultat de Ciències, C/M. A. Capmany 69, 17003 Girona, Spain.

ABSTRACT: Eumelanin is the black biopolymer responsible for photoprotection in living organisms. Lack of knowledge of the atomic-level structure limits understanding of its function and exploitation of its potential in material science. To overcome these limitations, we present a systematic density functional theory study of the stability and optical properties of a library of 830 dimers of 5,6-dihydroxyindole (DHI), which are minimal models of eumelanin oligomers. Our aim is to understand the principles that govern the formation of DHI oligomers, relate the optical properties of the dimers with their stability, and establish their possible role in the photophysics of the biopolymer. From the structural point of view, we find a preference for oxidized over reduced and cyclic over linear structures, which speaks in favor of polycyclic graphite-like structures for the larger oligomers. We present an electrocyclization mechanism leading to the cyclic structures. We also find that besides the widely considered quinone and quinone methide oxidation patterns where two heteroatoms per DHI fragment are oxidized, dimers with one or three oxidized sites per fragment and an interfragment double bond are also stable and may be present in eumelanin. As far as the optical properties are concerned, some oxidized dimers combine relative stability with absorption energies as low as 1.3 eV. Such fragments may be present as substructures in the naturally found oligomers and might have a relevant contribution to the absorption spectrum of the biopolymer. In addition to these insights into the struc-tural and optical properties of the oligomers, we introduce a new classification scheme and a representative set of 53 dimers combining thermodynamic stability with chemical diversity.

Introduction

Eumelanin is the black pigment found in skin, hair and eyes of living organisms, responsible for photoprotection against harmful light action.[1-3] It is one of the three main melanin variants together with pheomelanin and neuromelanin, and it is an insoluble and heterogeneous biopolymer. Eumelanin's photoprotective function is based on a broad absorption spectrum that enables it to absorb all UV and visible light of the environment.[4] It may also have a pathogenic role as generator of reactive oxygen species[5] and has raised interest in material science because its synthetic analogues, specially polydopamine, are versatile surface functionalization and coating materials combining attractive physical and chemical properties such as robustness, adhesivity, UV absorption, or antioxidant and free radical scavenger activity.[6-7]

Melanin is composed of oligomers made of two basic units, 5-6-dihydroxyindole (DHI) and 5,6-dihydroxyindole-2-carboxylic acid (DHICA). Its biological precursor is tyrosine. In the currently accepted biosynthetic path, tyrosine is transformed to DHI/DHICA in a sequence of enzymatic steps.[8] The route of eumelanin formation from DHI and DHICA is less well known, and it may involve oxidation by tyrosinase,[9] a redox reaction with dopaquinone or oxidation by O_2. In its turn, synthetic melanin analogues can be produced starting from tyrosine or dopa or directly from DHI or DHICA, using tyrosinase, peroxidase/hydrogen peroxide or oxygen as oxidants.[10]

Melanin oligomers contain 4 to 8 DHI/DHICA units.[11-12] DHI and DHICA are prone to oxidation, *ie* elimination of the pyrrolic or phenolic hydrogen atoms. Therefore, the DHI/DHICA fragments can also be present in quinone or quinone methide form, where two sites are oxidized on the same fragment (see Figure 1a). Further, assembly of two fragments with a single oxidized site results in structures with an interfragment double CC bond (Figure 1b). The fragments can assemble in linear arrangements (Figure 1b), but other conformations have been proposed such as 16-atom macrocycles[13] (Figure 1c) or polycyclic, graphite-like structures[14] (Figure 1d). According to the widely accepted hierarchical structural model,[6,15-16] the oligomers aggregate at different levels, yielding nanoparticles of > 100 nm size called melanosomes. However, the structural details are still largely unknown due to the heterogeneous and granular nature of eumelanin and its synthetic analogues, which prevents crystallographic characterization. Issues like the preferred connectivity between DHI/DHICA fragments, the preferred oxidation patterns, the presence of linear and cyclic structures or the relationship between optical properties and oligomer structure are largely unresolved. Clearly, lack of knowledge of the atomic-level structure limits our understanding of melanin function and our options of exploiting its potential in material science. To overcome these limitations, we use a bottom-up approach based on computational chemistry to understand the principles that govern the formation of DHI oligomers and the optical properties of the basic components.
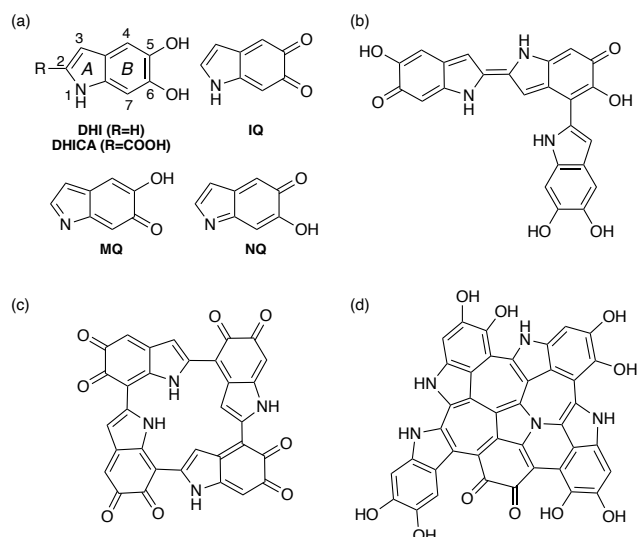
Figure 1. Chemical structures of eumelanin components and representative oligomers. (a) DHI/DHICA (with atom numbering and ring labeling) and oxidized indolequinone (IQ), quinone methide (MQ), and quinone imine (NQ) forms; (b) linear trimer with two singly oxidized fragments and an interfragment double bond; (c) macrocyclic tetramer; (d) polycyclic pentamer.

Computationally, the structural diversity of DHI oligomers has been recently studied in depth in two publications that have considered a total of more than 45000 structures (dimers up to tetramers) composed by the reduced DHI[17] and the oxidized indolequinone (IQ) and quinone methide (MQ) fragments.[18] Some connectivity and oxidation patterns were excluded *a priori* on the basis of stability arguments, and the large number of compounds remaining even after these constraints are applied gives an idea of the scale of the problem. These studies have yielded a set of representative oligomer models and provided important insights into the relationship between structure and stability, such as the preference for planar structures, the predominance of IQ fragments in the oxidized oligomers, and the favorable formation, from the energetic point of view, of macrocyclic tetramers. However, optical properties were not considered, and a large part of chemical space remained unexplored.

Turning to the optical properties, the main interest is understanding the origin of the absorption spectrum, which has a characteristic long tail that extends up to 800 nm.[4] Some computational studies have considered the radiationless decay mechanisms that may be responsible for the photoprotecting function,[19-20] but we center on absorption. Most studies of the absorption spectrum have been carried out on small compounds sets, and the results suggest that the low-energy absorption is due to the presence of oxidized structures. Thus, a study on ten structures (monomers up to pentamers), all in reduced form, gave absorption wave lengths lower than 490 nm,[21] whereas several studies on linear, macro- and polycyclic oligomers (see Figure 1) of up to 8 fragments with two to eight oxidized positions gave absorptions in the range 600 - 800 nm.[22-23] The effect of geometric disorder (*ie* aggregation) was shown to affect both the intensity and the shape of the absorptions. In spite of this significant progress, a systematic study of the optical properties of a larger library of compounds is still missing. Such a study is particularly important to gain insight into the relationship between oligomer structure, stability and absorption.

To address this issue, we have carried out a comprehensive study of the stability and optical absorption of a library of 830 DHI dimers, optimizing the structures and calculating the vertical absorption with density functional theory (DFT). The oligomers present in melanin are composed of 4 to 8 DHI fragments, and the dimers will only be present in small amounts, if any. However, our study of the dimers is fully justified because the oligomer properties will necessarily depend on those of the underlying substructures. In addition, the relatively small number of possible dimers compared to trimers or tetramers, where tens of thousands of structures are possible, makes it possible to consider all possible diversity. Therefore, we have considered all plausible connectivity and oxidation patterns (*ie* those with non-radical character), and we include cyclic structures that are possible constituents of graphitic-like polycycles.[14] We have determined the relationship between structure and absorption and have also considered the role of stability, which is crucial to understand the relevance of the different fragments for the photoprotecting function. We show that there is a thermodynamic preference for cyclic and oxidized structures, and the main optical absorption features of the biopolymer are already obtained at the level of dimers. Based on a new classification scheme that can be extended to larger oligomers we also propose a representative set of dimers that combines stability and diversity.

Computational Details. Structures are optimized with DFT, using the CAM-B3LYP functional and the 6-311G(d,p) basis set. Solvent effects (water) are included with the SMD continuum solvation model.[24] This approach was benchmarked against MP2/cc-pVTZ calculations for the reduced dimers (see section 1.1.1 of the Supporting Information, SI1.1.1). In addition to the 830 reported compounds, we considered six more isomers that were discarded, two because the structure is not stable (optimization led to a different dimer after tautomerization, see SI1.1.2), and four because the restricted DFT wave function had an instability indicative of radical character (see SI1.1.3). For every optimized structure, the vertical absorption energy and oscillator strength of the three lowest excited states is calculated with linear-response time-dependent DFT (TD-DFT), under the same conditions as the optimization. The suitability of TD-CAM-B3LYP was assessed comparing the vertical excitations of the reduced and oxidized monomers with MS-CASPT2 and EOM-EE-CCSD literature values.[25] The differences are approximately 0.2 eV, which lies in the expected precision range for TD-CAM-B3LYP (see SI1.1.4). All calculations were carried out with Gaussian 16.[26]

The reported energies include the thermal free energy correction (see SI1.2). In our discussion we use two energy scales: when we compare compounds of the same stochiometric formula (*ie* linear or cyclic compounds of the same oxidation state), the energies are provided relative to the lowest one of the set (Charts 1 and 2 and Figures 3, 4, and 6). However, to discuss the global stability (Figure 2) and compare structures with different stoichiometry, the hydrogen elimination reaction that relates compounds of different oxidation state is equalized with oxygen and water, mimicking some of the oxidation conditions used for synthetic melanin (see details in SI1.3).[10] Although this thermodynamic scheme is approximate, it allows for a fair comparison of compounds with different stoichiometry because the energy differences we discuss are > 10 kcal·mol^{-1}.

Library description and isomer classification. Our library contains all plausible linear dimers with CC interfragment connectivity, and cyclic dimers forming 5- or 6-membered

interfragment rings containing up to two nitrogen or oxygen atoms. As plausible we consider those dimers that have a total even number of oxidized sites. An odd number of oxidation sites results in a radical dimer which is probably too unstable, although radicals may become relevant for larger structures where the radical character can be stabilized by delocalization. For the linear dimers we consider *cis*/*trans* stereochemistry at the interfragment bond and include both stereoisomers in the set. The resulting set has a total of 538 linear and 292 cyclic dimers, and the complete library including structures and optical absorption data is available as Supporting Information.

To understand the relationship between structure, stability and optical properties it is convenient to have a structural classification where similar isomers are grouped in types. For this end we introduce a classification using three similarity criteria: oxidation state, connectivity, and resonance/bonding pattern:

(1) *Oxidation state.* We consider all isomers that can be obtained by eliminating up to three pairs of phenolic or pyrrolic hydrogen atoms. The compounds obtained by eliminating 2, 4 or 6 hydrogen atoms are classified as oxidation state 1, 2 and 3 (Ox1, Ox2 and Ox3).

(2) *Connectivity.* Linear isomers are classified according to the rings forming the interfragment bond as AA, AB or BB (see ring labeling in Figure 1a), whereas cyclic dimers are classified according to the structure of the interfragment ring as C6, C5 (interfragment ring formed by 6 or 5 carbon atoms) or CX (interfragment ring containing one or two nitrogen or oxygen atoms).

(3) *Resonance/bonding pattern.* Classifying the oxidation pattern with the monomer-based IQ/MQ/NQ types provides a large number of different types which is not convenient for analysis. Therefore we turn to a less detailed, more granular classification based on the number of oxidized sites on each fragment and the resonance structure describing the dimer, in particular the interfragment bond. When the dimer has an even or zero number of oxidized sites on each fragment (*ie* 2 and 0 oxidized sites in Ox1 and 2 and 2 in Ox2), its resonance structure has a single bond between the two fragments. This occurs with 50% of the dimers, which are classified as *s*. When the dimer has an odd number of oxidized sites on each fragment (*ie* 1 oxidized site on each fragment in Ox1, 1 and 3 sites in Ox2 and 3 sites on each fragment in Ox3), there are two more possibilities. In 20% of the dimers there is a neutral resonance structure with a double interfragment bond, and the dimer is classified as *d*. In the remaining 30% of the cases the resonance structure has zwitterionic character (no neutral one is possible), and the dimer is classified as *z*. The rules to classify the dimers as *s*, *d* or *z* can be derived from the fragments that compose the dimer and are explained in SI2.1. As we show below, the *s*/*d*/*z* classification is useful to analyze the optical absorption.
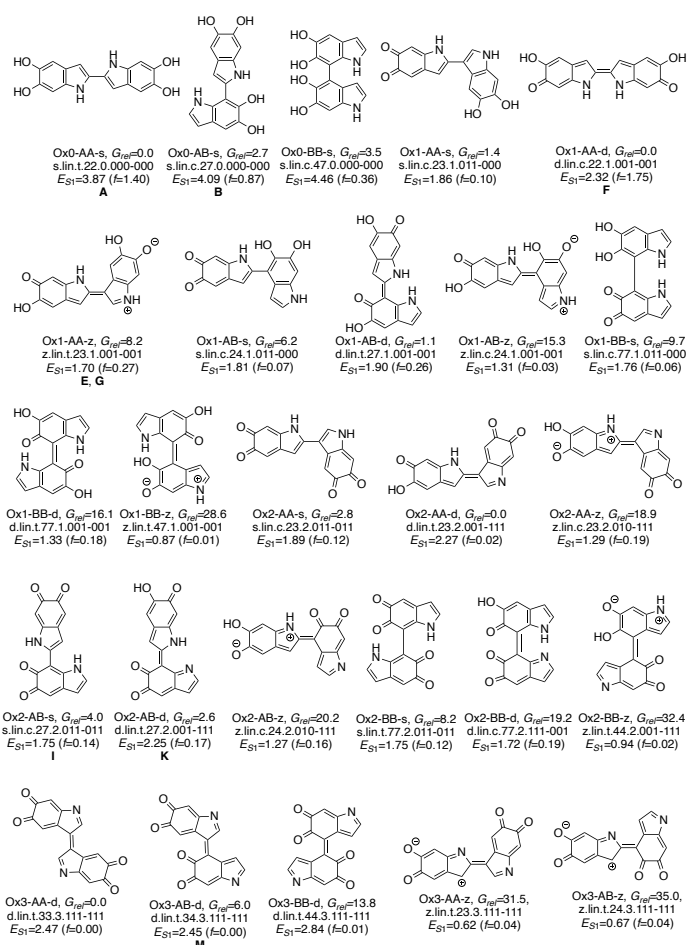
Chart 1. Most stable dimer for the 26 linear types. Labelling code: isomer type and free energy in kcal·mol^{-1} (relative to the most stable linear isomer of every oxidation state); full name; S$_1$ vertical absorption in eV and oscillator strength in brackets. Letters **A** - **F** refer to the signals of Figure 6.

Table 1. Number of dimers of each type

| | Linear isomers | | | | | | | | | |
	AA-s	AA-d	AA-z	AB-s	AB-d	AB-z	BB-s	BB-d	BB-z	Total
Ox0	6	-	-	8	-	-	6	-	-	20
Ox1	24	12	30	48	24	45[a]	24	20	19[a]	246
Ox2	42	6	18	72	20	28	42	16	8	252
Ox3	-	2	4	-	4	4	-	6	-	20
Total	72	20	52	128	48	80	72	42	30	538
	Cyclic isomers									
	C6-s	C6-d	C6-z	C5-s	C5-d	C5-z	CX-s	CX-d	CX-z	Total
Ox0	2	-	-	2	-	-	14	-	-	18
Ox1	6	5	7	12	7	11	46	17	57	168
Ox2	12	4	2	18	7	5	34	9	11	102
Ox3	-	2	-	-	2	-	-	-	-	4
Total	20	11	9	32	16	16	94	26	68	292

[a]Six other possible Ox1-AB-z and Ox1-BB-z isomers were discarded because of tautomerization or wave function instability (see SI1.1.2 and SI1.1.3).
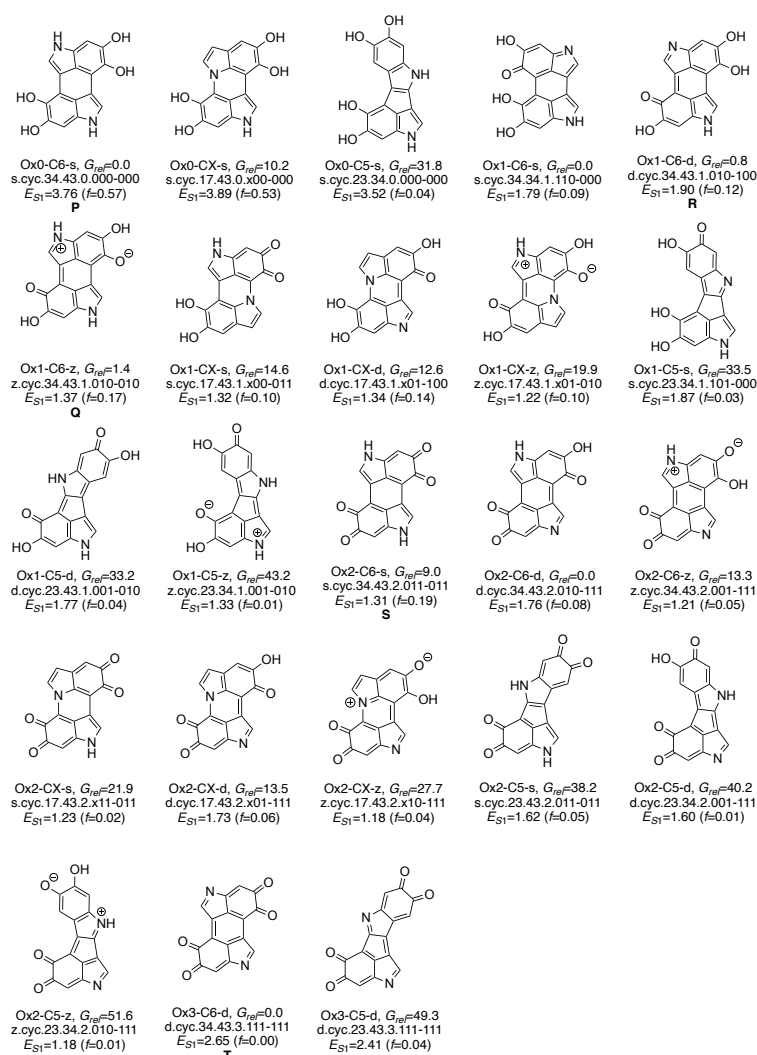
Ox0-C6-s, G_{ref}=0.0
s.cyc.34.43.0.000-000
E_{S1}=3.76 (f=0.57)
P

Ox0-CX-s, G_{ref}=10.2
s.cyc.17.43.0.x00-000
E_{S1}=3.89 (f=0.53)

Ox0-C5-s, G_{ref}=31.8
s.cyc.23.34.0.000-000
E_{S1}=3.52 (f=0.04)

Ox1-C6-s, G_{ref}=0.0
s.cyc.34.34.1.110-000
E_{S1}=1.79 (f=0.09)

Ox1-C6-d, G_{ref}=0.8
d.cyc.34.43.1.010-100
E_{S1}=1.90 (f=0.12)
R

Ox1-C6-z, G_{ref}=1.4
z.cyc.34.43.1.010-010
E_{S1}=1.37 (f=0.17)
Q

Ox1-CX-s, G_{ref}=14.6
s.cyc.17.43.1.x00-011
E_{S1}=1.32 (f=0.10)

Ox1-CX-d, G_{ref}=12.6
d.cyc.17.43.1.x01-100
E_{S1}=1.34 (f=0.14)

Ox1-CX-z, G_{ref}=19.9
z.cyc.17.43.1.x01-010
E_{S1}=1.22 (f=0.10)

Ox1-C5-s, G_{ref}=33.5
s.cyc.23.34.1.101-000
E_{S1}=1.87 (f=0.03)

Ox1-C5-d, G_{ref}=33.2
d.cyc.23.43.1.001-010
E_{S1}=1.77 (f=0.04)

Ox1-C5-z, G_{ref}=43.2
z.cyc.23.34.1.001-010
E_{S1}=1.33 (f=0.01)

Ox2-C6-s, G_{ref}=9.0
s.cyc.34.43.2.011-011
E_{S1}=1.31 (f=0.19)
S

Ox2-C6-d, G_{ref}=0.0
d.cyc.34.43.2.010-111
E_{S1}=1.76 (f=0.08)

Ox2-C6-z, G_{ref}=13.3
z.cyc.34.43.2.001-111
E_{S1}=1.21 (f=0.05)

Ox2-CX-s, G_{ref}=21.9
s.cyc.17.43.2.x11-011
E_{S1}=1.23 (f=0.02)

Ox2-CX-d, G_{ref}=13.5
d.cyc.17.43.2.x01-111
E_{S1}=1.73 (f=0.06)

Ox2-CX-z, G_{ref}=27.7
z.cyc.17.43.2.x10-111
E_{S1}=1.18 (f=0.04)

Ox2-C5-s, G_{ref}=38.2
s.cyc.23.43.2.011-011
E_{S1}=1.62 (f=0.05)

Ox2-C5-d, G_{ref}=40.2
d.cyc.23.34.2.001-111
E_{S1}=1.60 (f=0.01)

Ox2-C5-z, G_{ref}=51.6
z.cyc.23.34.2.010-111
E_{S1}=1.18 (f=0.01)

Ox3-C6-d, G_{ref}=0.0
d.cyc.34.43.3.111-111
E_{S1}=2.65 (f=0.00)
T

Ox3-C5-d, G_{ref}=49.3
d.cyc.23.43.3.111-111
E_{S1}=2.41 (f=0.04)

Chart 2. Most stable dimer for the 23 cyclic types. Labelling code: isomer type and free energy in kcal·mol^{-1} (relative to the most stable cyclic isomer of every oxidation state); full name; S$_1$ vertical absorption in eV and oscillator strength in brackets. Letters **J** - **N** refer to the signals of Figure 6.

Using the three criteria, we obtain 26 different linear isomer types and 23 cyclic ones. The number of isomers for each type is presented in Table 1, and the most stable isomers of each type are presented in Charts 1 and 2. These compounds form a representative set of the dimers that combines chemical diversity with stability. The nomenclature used to refer to individual isomers is explained in SI2.2.

Results and Discussion

Thermodynamics and stability. To assess thermodynamic preference for linear *vs* cyclic and oxidized *vs* reduced dimers, the relative energy distribution of the linear and cyclic dimers is plotted in Figure 2 (panels a and b, respectively). Oxidation states are coded by color. The number of Ox1 and Ox2 isomers in the library is much larger than that of Ox0 and Ox3 ones (see Table 1), which is reflected in the size of the histogram bars. Overall, oxidation is a thermodynamically favorable process for the dimers. Thus,

the bars shift progressively towards lower energies passing from fully reduced (Ox0, blue bars) to partially oxidized (Ox1 and Ox2, brown and red bars) and fully oxidized (Ox3, green bars) structures. The most stable linear and cyclic dimers are fully oxidized structures (d.lin.t.33.3.111-111 and d.cyc.34.43.3.111-111 in Charts 1 and 2). However, the preference for higher oxidation states depends on the connectivity (see Table SI2). In all linear dimers where one of the fragments is connected at C_2, the most stable isomer has oxidation state 2 because the Ox3 dimers with that connectivity all have a z resonance structure (eg the Ox3-AA-z compound in Chart 1), which is less stable than the neutral resonance structure found for Ox2 fragments. The preference for Ox2 over Ox3 in these fragments is 19 - 31 kcal·mol^{-1}. The situation is reversed for fragments connected at C_3 (except for 23 connectivity), where Ox3 is preferred by 12 - 21 kcal·mol^{-1}. Finally, for 44, 47 and 77 dimers there is no clear preference because the difference between the most stable Ox2 and Ox3 isomer is 1.5 - 7.5 kcal·mol^{-1}, which probably falls within the error margin of our calculated thermodynamics. As for the cyclic dimers, the highest oxidation state, Ox3, is preferred for the C6 isomers, and Ox2 for the C5 ones. For the CX isomers, the highest possible oxidation state is Ox2 because the heteroatom that is part of the interfragment ring cannot be oxidized, and this is also the preferred state.

The plot also shows that cyclization is thermodynamically favorable. The difference between the most stable cyclic and linear dimers with the same oxidation state is 50-70 kcal·mol^{-1} (bars of the same color in Fig. 2), whereas that between the most stable cyclic and linear dimers with the same stoichiometry (eg most stable linear Ox1 vs cyclic Ox0) is 20-50 kcal·mol^{-1}.
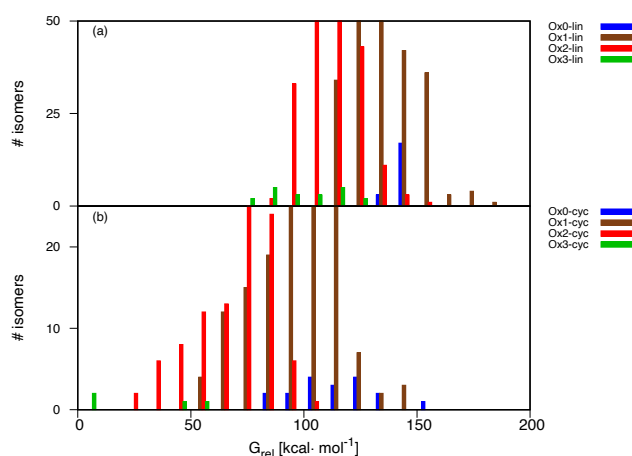

Figure 2. Free energy distribution of (a) linear and (b) cyclic dimers as a function of their oxidation state, relative to the most stable library dimer (y-axis cut at 50 and 25 isomers for clarity, see Figure SI3 for full plots).

As a possible route for formation of the cyclic structures we have investigated a two-step path that involves an electrocyclization, a variant of the hexatriene to cyclohexadiene rearrangement (see Scheme 1). This route is possible when the dimer has the *cis*-like triene arrangement highlighted in red. The cyclization step leads to an intermediate, followed by oxidation (H_2 elimination) to the cyclic isomers. This route was suggested by optimizations of some linear structures that lead spontaneously to the cyclized intermediates. It has been studied in detail for three representative examples,

one with an all-carbon hexatriene skeleton and two where a nitrogen or an oxygen atom replace one of the six carbon atoms. The reactions lead to an Ox1-C6 and two Ox1-CX products. The electrocyclization steps have low barriers of 12 - 22 kcal·mol^{-1} and are thermodynamically favorable by 13 - 50 kcal·mol^{-1}. The hydrogen elimination step, whose energy has been estimated in the same fashion as the oxidation steps, is also highly favorable due to aromaticity gain in the final products. These examples show that an alternative route to radical reaction is possible for the formation of the cyclic structures.[27]
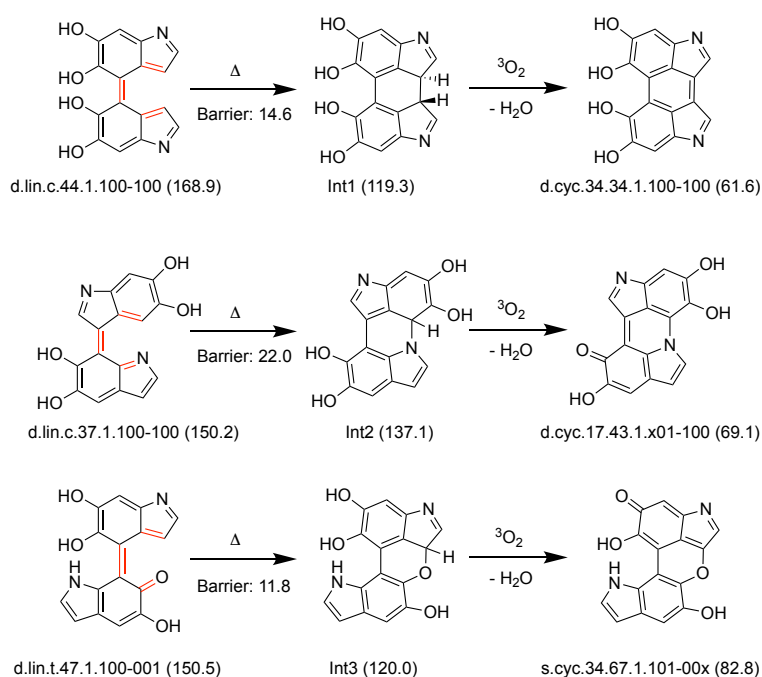
Scheme 1. Reaction scheme for formation of cyclic isomers via electrocyclization. Free energies in kcal·mol^{-1} relative to the most stable dimer (*ie* energy scale of Figure 2) provided in brackets.

To identify which are the thermodynamically most favorable types in each oxidation state, we present free energy distribution plots of the linear and cyclic isomer types in Figures 3 and 4. In the histogram bars, the isomers of each type are grouped according to their energy relative to the most stable isomer of every oxidation state, and we plot the most stable energy range of 0-20 kcal·mol^{-1}. Histograms for the complete sets covering ranges of up to 70 kcal·mol^{-1} are provided in the SI (Figures SI4, SI5).
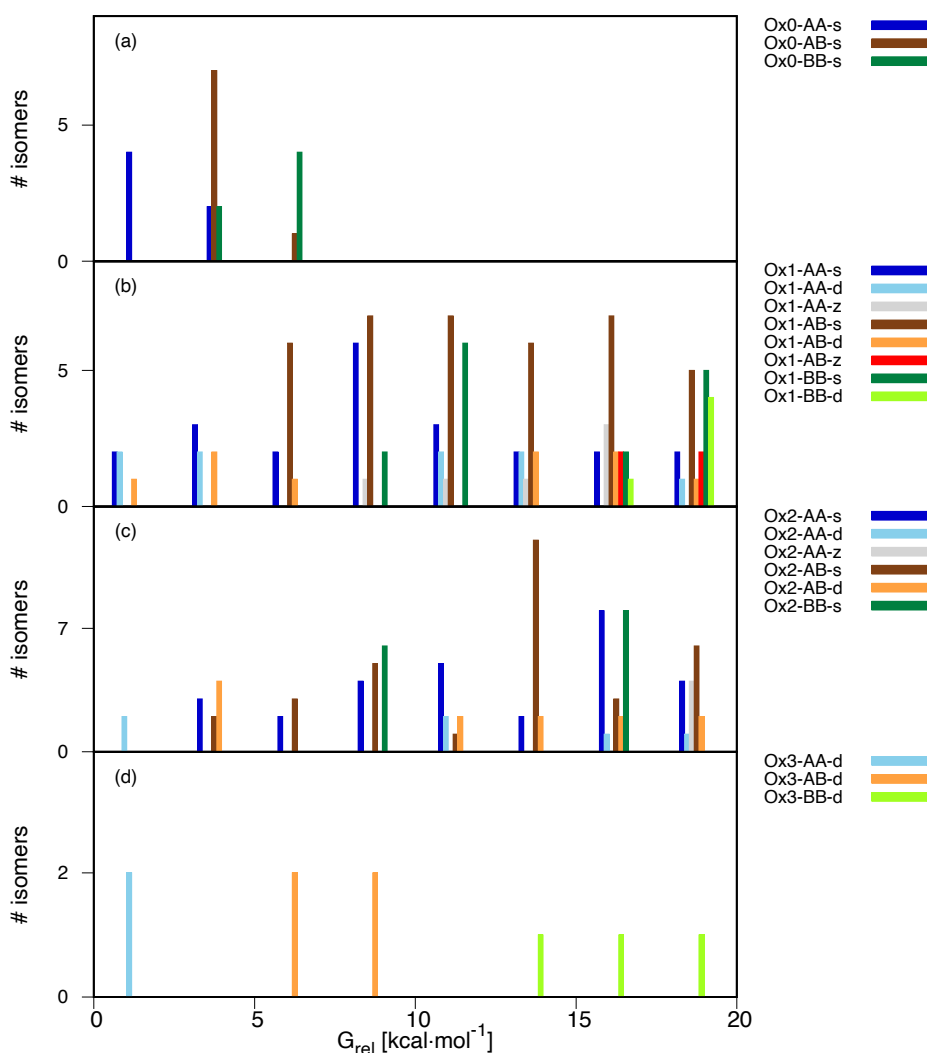
Figure 3. Free energy distribution of linear isomers as a function of their type (0 - 20 kcal·mol⁻¹ energy range, see Figure SI4 for the full range). Panels (a) - (d) display Ox0 - Ox3 dimers, and energies are relative to the most stable dimer of each group.

For the linear isomers (Figure 3), all Ox0 compounds fall into a narrow energy range of < 10 kcal·mol⁻¹. For the other oxidation states the energy spread is higher, up to 70 kcal·mol⁻¹ (see Figure SI4), and some trends regarding the stability are visible. First, for all oxidation states there is a preference for AA ring connectivity and, to a lesser extent, AB, over BB. In the three oxidation states the most stable isomer has AA connectivity, and the most stable BB structures lie 8-10 kcal·mol⁻¹ higher. This is related with the coplanarity of the DHI fragments. Coplanarity is favored in the AA case because the connection involves five-membered rings and there is less steric contact between the rings than in the BB case, where the connection involves two six-membered rings. More coplanarity results in an energetic stabilization thanks to an increase in electronic delocalization, and therefore the AA compounds are more stable. This preference for coplanar structures was also observed in previous studies of linear oligomer libraries.[17-18]

The preference for AA and AB connectivity is also consistent with experimental studies on dimers isolated from oxidative DHI oligomerization, where the preferred connectivity was 22, 24 and 27.[28] However there are some differences. First, according

to our studies 22 is the thermodynamically most stable connectivity for Ox0 and Ox1, but these dimers where only identified experimentally when the polymerization was carried out in the presence of transition metal ions. In addition, some of our most stable dimers have connectivity at C_3, which was not found experimentally. This suggests that other factors than the product stability, such as the thermodynamics of the radical precursors or the dimerization kinetics, are also important.
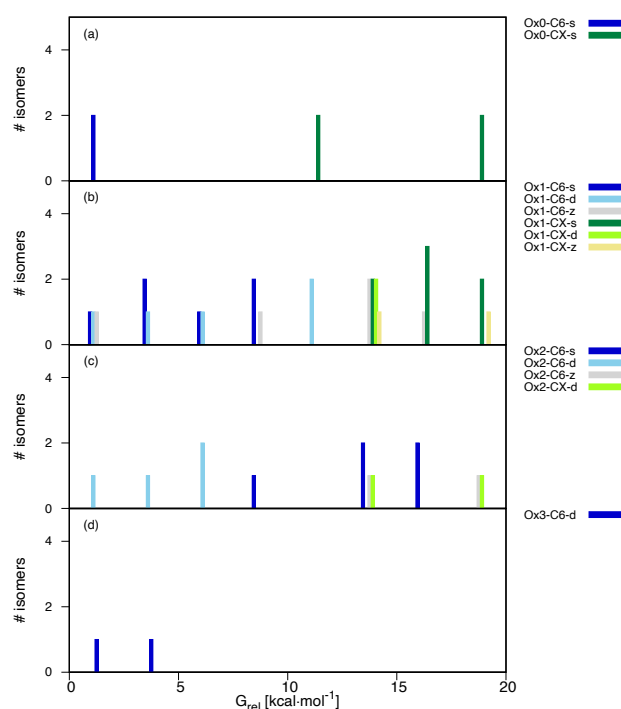

Figure 4. Free energy distribution of cyclic isomers as a function of their type (0 - 20 kcal·mol^{-1} energy range, see Figure SI5 for the full range). Panels (a) - (d) display Ox0 - Ox3 dimers, and energies are relative to the most stable dimer of each group.

Turning to the oxidation and bonding pattern, the *d* pattern is somewhat preferred over *s* for AA and AB compounds of Ox1 and Ox2 (*s* isomers: dark blue bars in Figure 3b,c; *d* isomers: light blue and orange bars; see also Chart 1, where the Ox1-AA-d dimer is more stable than the Ox1-AA-s one, and the same happens for Ox1-AB, Ox2-AA, Ox2-AB). The preference for an interfragment double bond may be due to a favorable electronic delocalization among the two fragments. occurs in the AA and AB compounds because the fragments are connected at the smaller rings, which allows them to be coplanar, as required by the double bond. In contrast to this, the Ox1-BB and Ox2-BB compounds prefer *s* connectivity (dark *vs* light green bars). Among the *s* compounds, those with IQ fragments are preferred, in line with previous studies.[18] Finally, the *z* compounds (grey and red bars in Figure 3b,c) are less stable than their *s* and *d* counterparts.

In the cyclic isomers (Figure 4), there is a clear preference for the C6 isomers (light and dark blue and grey bars) in all oxidation states, with the most stable CX and C5 isomers lying 10-15 and 30-35 kcal·mol^{-1} higher in energy, respectively (see Figure SI5 for the higher-energy range). In the CX isomers, the presence of a nitrogen atom in the ring is more favorable energetically than that of an oxygen. The most stable O-

containing CX isomer is about 10 kcal·mol^{-1} less stable that its N counterpart in Ox1 compounds, and 27 kcal·mol^{-1} in the Ox2 set.

Looking at the oxidation and bonding pattern, the *d* and *z* patterns, *ie* odd number of oxidized sites on every fragment, are also favorable for the cyclic compounds. For the Ox1-C6 compounds, the most stable group includes one isomer of each resonance/bonding (Figure 4b, see also Chart 2), and the same happens with the Ox1-C5 compounds. For Ox2 (Figure 4c) all Ox1-C6 isomers lie within 15 kcal·mol^{-1}, and the most favored connectivity is *d*. The fact that the *z* connectivity is comparatively more favored in the cyclic isomers may imply that the zwitterionic charges can be stabilized better than in the linear ones, presumably because delocalization is larger.

Optical properties. The relationship between structure and vertical S$_1$ absorption energy is illustrated in Figure 5. In this plot, every stick represents the absorption range of a group of compounds; the filled boxes represent the 50% of compounds in the group that absorb closest to the median absorption (second and third quartile), and the whiskers cover the full range of absorption for that group. The median is shown as a black line, and the label identifies the group and the number of compounds in it. The vertical dashed lines give the S$_1$ absorptions of the DHI, IQ, MQ and NQ monomers for reference, calculated at the same level of theory. Several trends can be recognized. For the fully reduced dimers, the linear ones (Ox0.lin.s, stick in the upper right part of the panel) absorb between 3.9 and 4.7 eV, which is in line with previous calculations giving 3.7 - 4.3 eV for the dimers.[21] Compared to the DHI excitation of 4.8 eV, the effect of dimerization is lowering the excitation energy by up to 0.9 eV. In turn, the absorption of the cyclic reduced dimers (Ox0.cyc.s) is 3.4 - 4.0 eV, implying that cyclization shifts the excitation energy down by another 0.5 eV.

In line with previous studies,[22-23] the oxidized dimers absorb at lower energies: the Ox1 dimers between 0.7 and 2.7 eV, and the Ox2 ones between 0.5 and 2.5 eV. The Ox3.lin.d and Ox3.cyc.d dimers absorb between 2.3 and 2.8 eV, substantially higher than the Ox1 and Ox2 compounds. Therefore, the trend is that the partially oxidized dimers absorb in the lower energy region, with small differences between oxidation states 1 and 2, whereas complete oxidation (with the exception of *z* resonance, see below) leads to an increase in the absorption energy. This is probably due to the fact that the fully oxidized structures with *d* resonance/bonding character recover aromaticity compared to the Ox1 and Ox2 ones, which increases the HOMO-LUMO gap and the excitation energy (see compound structures in Charts 1, 2). Turning to the relationship with bonding and resonance, if one compares the Ox1.lin.s, Ox1.lin.d and Ox1.lin.z sticks (upper half of the panel), it appears that the median absorption of the Ox1.lin.s and Ox1.lin.d compounds is similar, but the Ox1.lin.d compounds cover a broader range (0.7 - 2.7 eV) compared to Ox1.lin.s (1.4 - 2.2 eV). In turn, the Ox1.lin.z compounds absorb at lower energies, 0.7 - 1.7 eV. This trend is maintained for Ox2 and the cyclic compounds, and it suggests that compounds with *d* or *z* resonance/bond pattern may contribute to the lower energy absorption.
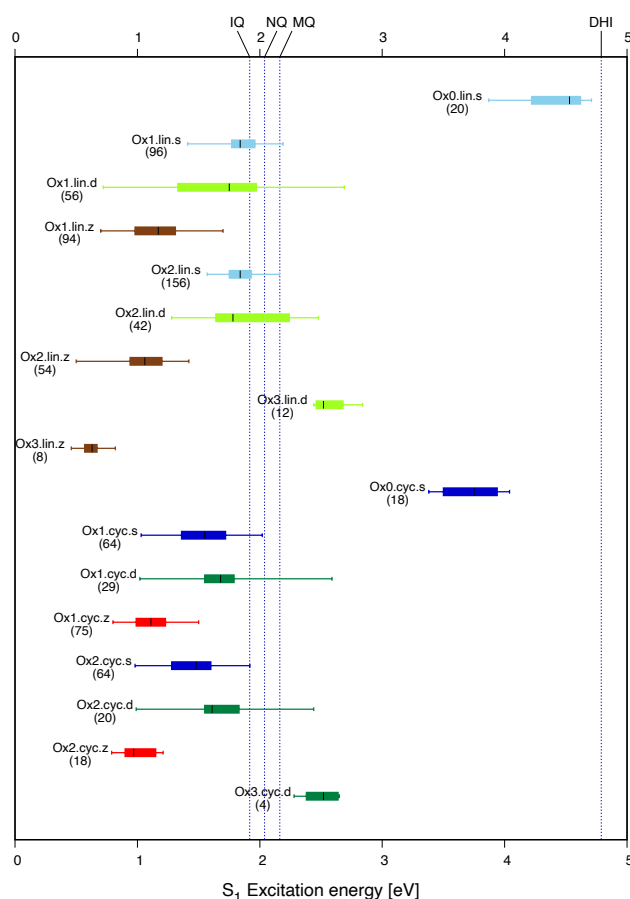
Figure 5. Distribution of S_1 excitation energy for the dimers grouped by oxidation state and resonance/bonding, separating linear from cyclic. The filled boxes cover the second and third quartile of each group, and the whiskers the full range from the lowest to the highest absorption. The label includes the number of compounds from each type. The vertical dashed lines are, for reference, the absorption of the DHI monomer and the oxidized IQ, NQ and MQ monomers.

To understand how the different dimers may contribute to the absorption spectrum, it is necessary to consider their relative stability. This is illustrated in Figure 6. The eight panels correspond to the linear and cyclic isomers of Ox0 - Ox3, and every signal on the plot arises from one or more compounds that absorb at the energy indicated on the x axis and have the relative free energy (G_{rel}) given on the y axis. The plot includes absorption of the three lowest states, S_1-S_3 (see SI5 for details). The absorption intensity is coded by the color bars on the right and is proportional to the calculated oscillator strength. The lower parts of the plots show how the most stable compounds of each group may contribute to the spectrum of a hypothetical compound mixture. The most relevant absorption features are highlighted with a white circle and labeled with the letters A-T.
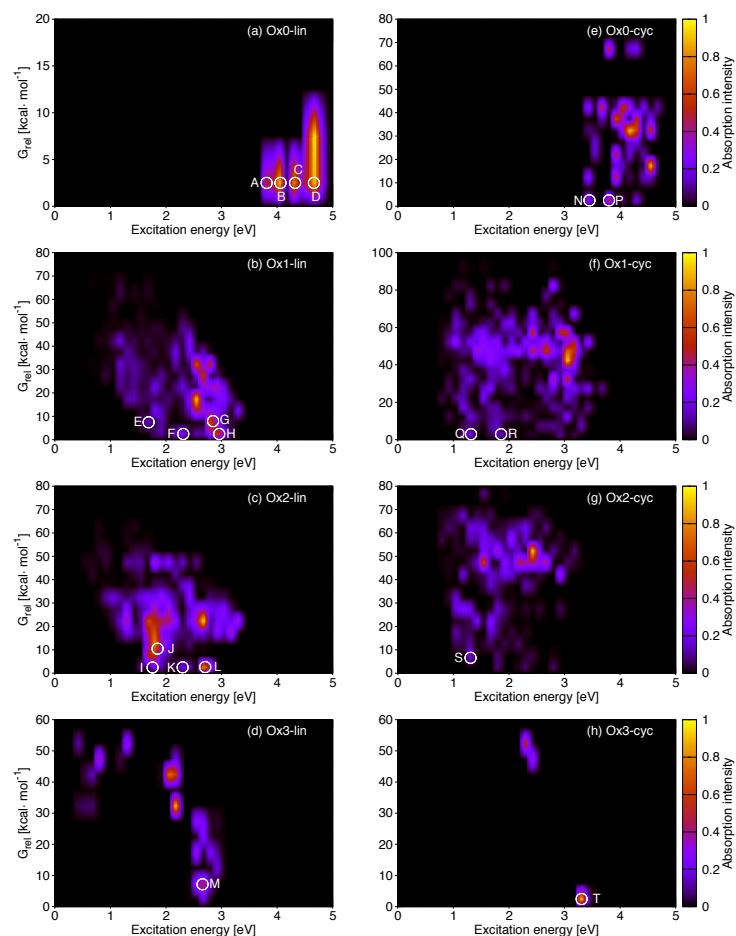
Figure 6. S_1-S_3 absorption properties (excitation energy, x axis, and oscillator strength, color bars) *vs* stability (energy relative to the most stable compound of every group) of (a)-(d) linear and (e)-(h) cyclic library compounds grouped by oxidation state. The letters A-T identify the most relevant features of stable compounds (low G_{rel}).

The upper part of Figure 6 (panels a and e) shows the absorption of the reduced (Ox0) linear and cyclic dimers. The absorption starts at 3.8 and 3.4 eV, respectively, which is consistent with the distribution of Figure 5. The linear dimers show higher absorption intensity than the cyclic ones, specially for the stable compounds (compare features A-D with N and P). Some of the cyclic Ox0 dimers have more intense absorption, but their G_{rel} is high, so they will probably be less relevant. Turning to the Ox1 and Ox2 compounds, the linear dimers with low G_{rel} (bottom of panels b and c, signals E-H and I-L) absorb between 1.5 and 3 eV. The dimers absorbing at lower energies in these groups (1 - 1.5 eV) are relatively unstable, with $G_{rel} \geq 15$ kcal·mol^{-1}. For the cyclic Ox1 and Ox2 dimers (panels f and g), the absorptions of the stable compounds (signals Q, R and S) are less intense than those found for their linear counterparts, but they appear at somewhat lower energy (1.3 - 1.4 eV, signals Q and S). As in the Ox0 case, the cyclic dimers with more intense absorption are relatively unstable. Finally, the fully oxidized Ox3 compounds absorb at significantly higher energy than the Ox1 and Ox2 analogues (panels d and h, signals M and T). This may partly be due to a gain in aromaticity, as discussed above, and it is reinforced by the fact that the lowest states have n,π* character, so that the signals correspond to higher-energy states (S_2 and S_3 for M and T,

respectively). Some Ox3 compounds absorb between 1 and 2 eV, but they are unstable (G_{rel} > 40 kcal·mol^{-1}) and probably not relevant.

The characterization of the compounds responsible for the highlighted signals in Figure 6 is provided in Table SI3. For each signal we present the most significant contributions (absorptions with higher oscillator strength), and the corresponding compounds are displayed in Charts 1, 2 and SI3. It is interesting to analyze the resonance/bonding pattern of these compounds to find how the different types may contribute to the spectrum. We center on the Ox1 and Ox2 groups because in the Ox0 case all dimers have *s* resonance/bonding, and in the Ox3 case all stable dimers are *d* type. In the Ox1-lin group, the two low-energy signals (E and F) arise from S_1 absorption of a *z* and a *d* compound, whereas the higher-energy ones (G and H) correspond to S_2 and S_3 absorptions from a larger group of compounds including *s*, *d* and *z* contributions. In contrast to this, the lowest-energy absorptions in the Ox2-lin group (I and J) arise from *s* compounds formed by two IQ fragments, whereas the higher-energy ones (K and L) come from *d* compounds (S_3 absorptions in the L case). Turning to the cyclic dimers, the Q and R signals come from a *z* compound and a group of *s* and *d* compounds, respectively, and signal S from an *s* compound formed by two IQ fragments.

To summarize, the absorption *vs* stability plots show that some of the thermodynamically stable Ox1 and Ox2 dimers have absorptions as low as 1.3 eV, with the absorptions of the cyclic compounds being weaker and lower in energy than the linear ones. Compounds with the three possible resonance/patterns, *s*, *d* and *z*, are responsible for the different signals. In particular, linear and cyclic dimers formed by two IQ fragments have an important contribution to the low energy absorptions (signals I, J, and S). We also find that some of the dimers have strong S_2 and S_3 absorptions below 3 eV. Population of these states may have implications for the excited-state dynamics and the non-radiative decay of the polymer.

Conclusions

Our work is the first exhaustive study of the stability and optical properties of a group of melanin components, the dimers, and their relationship with structure. With our classification in terms of oxidation state, connectivity and resonance/bonding, we have generated a representative set of the most stable 53 structures covering the chemical diversity of the dimers. Previous works have introduced representative sets of dimers up to the tetramers based solely on stability,[17-18] but these are found to be quite repetitive from the structural point of view. Our approach is complementary and compensates this by combining stability with diversity criteria.

From the structural point of view, our results show a preference for oxidized structures (Ox2 or Ox3, depending on the connectivity) which is consistent with the fact that DHI undergoes oxidative polymerization under mild conditions. The more reduced structures are less favorable, thermodynamically. Moreover cyclic structures, in particular those involving six-membered interfragment rings, are favored compared to linear ones. This speaks in favor of the polycyclic graphite-like structures,[14] which have been also favored by a recent Raman spectroscopic study of dopa melanin.[29] While our calculated C6 structures may be present as substructures in melanin oligomers, the C5 structures are probably not relevant because they are too high in energy. We have also

found an energetically accessible electrocyclic path for the formation of cyclic dimers which may complement the usually assumed radical oligomerization mechanism.

Turning to optical absorption, strongly red shifted absorptions down to about 1.3 eV ($l > 900$ nm) can already be seen for some dimers that are among the most stable ones. These fragments may have a relevant contribution, as substructures of larger oligomers, to the absorption of melanin samples. Our stability *vs* absorption analysis shows which dimers of the different oxidation states combine the necessary features of stability and absorption to contribute to the spectrum. Importantly, the compounds that contribute to the relevant low energy absorptions have *s*, *d* and *z* resonance/bonding pattern. This shows that theoretical melanin models should not be limited to indolequinone or quinone methide fragments with two oxidized heteroatoms per fragment. Instead, also structures with one or three oxidized heteroatoms per fragment and a CC double bond between the fragments can be thermodynamically stable and have relevant optical properties. In fact, even compounds whose resonance structure seems unfavorable, such as the ones we have termed as *z*, can be stable and relevant.

As an outlook, we plan to extend our approach to the DHICA dimers. In this case the preference for AA and AB connectivity found for the DHI dimers is likely to be reversed in favor of BB (*ie* connectivity at the C_4 and C_7 positions).[28] Since this is related with the preferred resonance/bonding character and the optical absorption, it will be interesting to see how the absorption of the DHICA dimers differs from the DHI ones, and also whether the preference for oxidized and cyclic structures is maintained. Our approach will also be extended to larger oligomers. One question here is whether the optical properties of the linear oligomers are dominated by those of the smaller fragments, or whether extended conjugation effects are dominant. It will also be interesting to determine whether cyclic and oxidized structures are also preferred for the larger oligomers, and in the case of the cyclic compounds we will include structures with seven-membered interfragment rings in our sets that cannot be formed at the dimer level. Given the complexity of the oligomerization scheme, it will probably also be necessary to go beyond thermodynamics and include kinetic aspects in the construction of representative theoretical melanin models.

Supporting Information

Supporting Information files include a document in PDF format containing computational details, details on the classification scheme and nomenclature, full histograms of energy distribution, a table with the most stable linear isomers of each connectivity, explanation of the construction of the absorption *vs* stability plots, list and structures of compounds responsible for the relevant signals, and complete Reference 26.

Acknowledgment

REFERENCES

1.   d'Ischia, M.; Wakamatsu, K.; Cicoira, F. *et al.* Melanins and melanogenesis: from pigment cells to human health and technological applications. *Pigment Cell Melanoma Res.* **2015,** *28*, 520-544.

2.   Meredith, P.; Sarna, T. The physical and chemical properties of eumelanin. *Pigment Cell Res.* **2006,** *19*, 572-594.

3.   d'Ischia, M.; Napolitano, A.; Ball, V.; Chen, C. T.; Buehler, M. J. Polydopamine and Eumelanin: From Structure-Property Relationships to a Unified Tailoring Strategy. *Acc. Chem. Res.* **2014,** *47*, 3541-3550.

4.   Nighswander-Rempel, S. P.; Riesz, J.; Gilmore, J.; Meredith, P. A quantum yield map for synthetic eumelanin. *J. Chem. Phys.* **2005,** *123*, 194901.

5.   Brash, D. E.; Goncalves, L. C. P.; Bechara, E. J. H. Chemiexcitation and Its Implications for Disease. *Trends Mol. Med.* **2018,** *24*, 527-541.

6.   d'Ischia, M.; Napolitano, A.; Pezzella, A.; Meredith, P.; Sarna, T. Chemical and Structural Diversity in Eumelanins: Unexplored Bio-Optoelectronic Materials. *Angew. Chem. Int. Ed.* **2009,** *48*, 3914-3921.

7.   Di Mauro, E.; Xu, R.; Soliveri, G.; Santato, C. Natural melanin pigments and their interfaces with metal ions and oxides: emerging concepts and technologies. *MRS Communications* **2017,** *7*, 141-151.

8.   Ito, S.; Wakamatsu, K. Chemistry of mixed melanogenesis - Pivotal roles of dopaquinone. *Photochem. Photobiol.* **2008,** *84*, 582-592.

9.   Roulier, B.; Peres, B.; Haudecoeur, R. Advances in the Design of Genuine Human Tyrosinase Inhibitors for Targeting Melanogenesis and Related Pigmentations. *J. Med. Chem.* **2020,** *63*, 13428-13443.

10.   d'Ischia, M.; Wakamatsu, K.; Napolitano, A. *et al.* Melanins and melanogenesis: methods, standards, protocols. *Pigment Cell Melanoma Res.* **2013,** *26*, 616-633.

11.   Okuda, H.; Yoshino, K.; Wakamatsu, K.; Ito, S.; Sota, T. Degree of polymerization of 5,6-dihydroxyindole-derived eumelanin from chemical degradation study. *Pigment Cell Melanoma Res.* **2014,** *27*.

12.   Pezzella, A.; Napolitano, A.; dIschia, M.; Prota, G.; Seraglia, R.; Traldi, P. Identification of partially degraded oligomers of 5,6-dihydroxyindole-2-carboxylic acid in Sepia melanin by matrix-assisted laser desorption/ionization mass spectrometry. *Rapid Commun. Mass Spectrom.* **1997,** *11*, 368-372.

13.   Kaxiras, E.; Tsolakidis, A.; Zonios, G.; Meng, S. Structural model of eumelanin. *Phys. Rev. Lett.* **2006,** *97*, 218102.

14.   Cheng, J.; Moss, S. C.; Eisner, M. X-ray characterization of melanins - II. *Pigment Cell Res.* **1994,** *7*, 263-273.

15.   Ju, K. Y.; Kang, J.; Chang, J. H.; Lee, J. K. Clue to Understanding the Janus Behavior of Eumelanin: Investigating the Relationship between Hierarchical Assembly Structure of Eumelanin and Its Photophysical Properties. *Biomacromolecules* **2016,** *17*, 2860-2872.

16.   Watt, A. A. R.; Bothma, J. P.; Meredith, P. The supramolecular structure of melanin. *Soft Matter* **2009,** *5*, 3754-3760.

17.   Chen, C. T.; Martin-Martinez, F. J.; Jung, G. S.; Buehler, M. J. Polydopamine and eumelanin molecular structures investigated with ab initio calculations. *Chem. Sci.* **2017,** *8*, 1631-1641.

18.   Chen, C. T.; Buehler, M. J. Polydopamine and eumelanin models in various oxidation states. *Phys. Chem. Chem. Phys.* **2018,** *20*, 28135-28143.

19.   Ghosh, D. Computational aspects towards understanding the photoprocesses in eumelanin. *WIREs Comput. Mol. Sci.* **2021,** *11*, e1505.

20.   Nogueira, J. J.; Corani, A.; El Nahhas, A.; Pezzella, A.; d'Ischia, M.; Gonzalez, L.; Sundstrom, V. Sequential Proton-Coupled Electron Transfer Mediates Excited-State Deactivation of a Eumelanin Building Block. *J. Phys. Chem. Lett.* **2017,** *8*, 1004-1008.

21.   Tuna, D.; Udvarhelyi, A.; Sobolewski, A. L.; Domcke, W.; Domratcheva, T. Onset of the Electronic Absorption Spectra of Isolated and pi-Stacked Oligomers of 5,6-Dihydroxyindole: An Ab Initio Study of the Building Blocks of Eumelanin. *J. Phys. Chem. B* **2016,** *120*, 3493-3502.

22.   Chen, C. T.; Chuang, C.; Cao, J. S.; Ball, V.; Ruch, D.; Buehler, M. J. Excitonic effects from geometric order and disorder explain broadband optical absorption in eumelanin. *Nat. Commun.* **2014,** *5*, 3859.

23.   Marchetti, B.; Karsili, T. N. V.; Ashfold, M. N. R.; Domcke, W. A 'bottom up', ab initio computational approach to understanding fundamental photophysical processes in nitrogen containing heterocycles, DNA bases and base pairs. *Phys. Chem. Chem. Phys.* **2016,** *18*, 20007-20027.

24.   Ghosh, S.; Verma, P.; Cramer, C. J.; Gagliardi, L.; Truhlar, D. G. Combining Wave Function Methods with Density Functional Theory for Excited States. *Chem. Rev.* **2018,** *118*, 7249-7292.

25.   Ghosh, P.; Ghosha, D. Elucidating the Photoprotection Mechanism of Eumelanin Monomers. *J. Phys. Chem. B* **2017,** *121*, 5988-5994.

26.   Frisch, M. J.; Trucks, G. W.; Schlegel, H. B. *et al. Gaussian 16 Rev. A.03*, Wallingford, CT, 2016.

27.   Pezzella, A.; Crescenzi, O.; Panzella, L.; Napolitano, A.; Land, E. J.; Barone, V.; d'Ischia, M. Free Radical Coupling of o-Semiquinones Uncovered. *J. Am. Chem. Soc.* **2013,** *135*, 12142-12149.

28.   D'Ischia, M.; Napolitano, A.; Pezzella, A.; Land, E. J.; Ramsden, C. A.; Riley, P. A. 5,6-dihydroxyrindoles and Indole-5,6-diones. Katritzky, A. R., Ed. Advances in Heterocyclic Chemistry, 2005; Vol. 89, pp 1-63.

29.   Grieco, C.; Kohl, F. R.; Hanes, A. T.; Kohler, B. Probing the heterogeneous structure of eumelanin using ultrafast vibrational fingerprinting. *Nat. Commun.* **2020,** *11*, 4569.

Structure, stability and optical absorption of a comprehensive library of minimal eumelanin oligomer models

Jun Wang,[a,b] Lluís Blancafort[b,*]

[a]Jiangsu Key Laboratory for Chemistry of Low-Dimensional Materials, Jiangsu Engineering Laboratory for Environment Functional Materials, School of Chemistry and Chemical Engineering, Huaiyin Normal University, No.111 West Changjiang Road, Huaian 223300, Jiangsu Province, PR China.

[b]Institut de Química Computacional i Catàlisi and Department de Química, Universitat de Girona, Facultat de Ciències, C/M. A. Capmany 69, 17003 Girona, Spain.

SUPPORTING INFORMATION

Contents

SI1. Computational details

SI1.1 Choice of electronic structure method

The calculations were carried out with density functional theory (DFT). We chose the CAM-B3LYP functional with the 6-311G(d,p) basis set, which gives good performance for the excited-state calculations.[1-2] In particular, this approach avoids spurious charge transfer stabilization effects[3] that can occur in the dimers. For consistency, restricted CAM-B3LYP was used both for structural optimization and time-dependent DFT (TD-DFT) vertical excitations. The use of a dispersion correction was discarded based on comparison with MP2 (see next section). Solvent effects were taken into account with the SMD continuum model[4] using water as solvent. All calculations were performed with Gaussian 16.[5]

SI1.1.1 Benchmark of ground-state approach

The validity of our approach to estimate the stability of the different compounds was tested for the set of 20 linear, fully reduced DHI dimers. As a benchmark method we used MP2 with Dunning's correlation-consistent polarized valence triple-z (cc-pVTZ) basis set and SMD solvation. We considered SMD-CAM-B3LYP/6-311G(d,p) with and without the atom-pair wise dispersion D3 correction, and compared the relative potential energies of the 22 compounds with the three approaches. The results are shown in Figure SI1, where we display the relative energies following the MP2 energies in decreasing order. In both charts the blue bars represent MP2 energies, and the red ones CAM-B3LYP without and with D3 correction[6] (Figures SI1a and b, respectively). The relative energies span a range of 10 kcal·mol^{-1} with MP2. The CAM-B3LYP energies are somewhat less dispersed, but the range of dispersion is closer to MP2 when the D3 correction is not included (6 *vs* 4 kcal·mol^{-1}). In addition, the order of the compounds is approximated better with CAM-B3LYP than by CAM-B3LYP-D3. In particular, CAM-B3LYP-D3 clearly underestimates the relative energies of the c.77 and c.47 isomers, which are the less stable ones at the MP2 level and the second and fourth more stable ones with CAM-B3LYP-D3. The root mean square of the difference between the relative energies is 3.1 kcal/mol for CAM-B3LYP and 5.0 kcal/mol for CAM-B3LYP-D3. Overall, the agreement with MP2 is better without the D3 correction. Therefore, the D3 scheme was discarded. The good agreement between CAM-B3LYP and MP2 is also seen comparing the energy distribution histograms as a function of the type (Figure SI2).

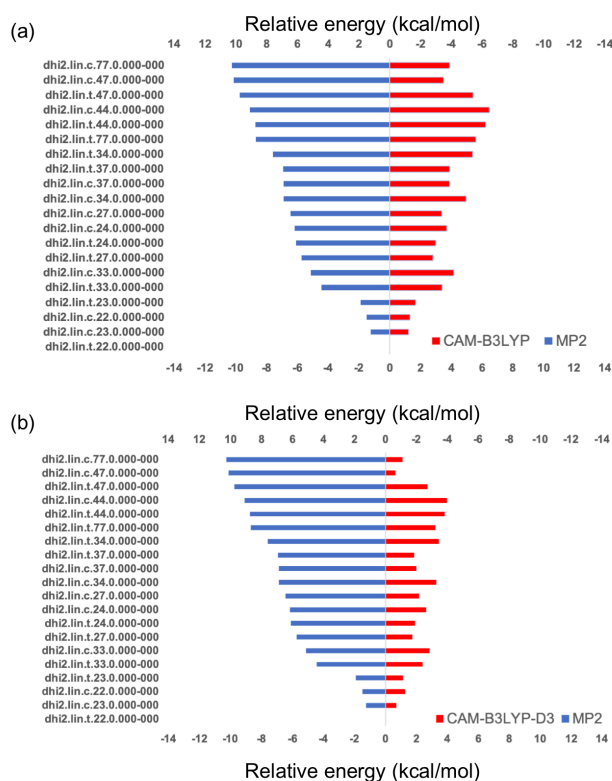**Figure SI1.** Relative potential energy of the Ox0-lin compounds calculated at the MP2/cc-pvtz level (blue bars) compared with CAM-B3LYP/6-311G** (red bars) without and with the D3 correction (panels (a) and (b), respectively).
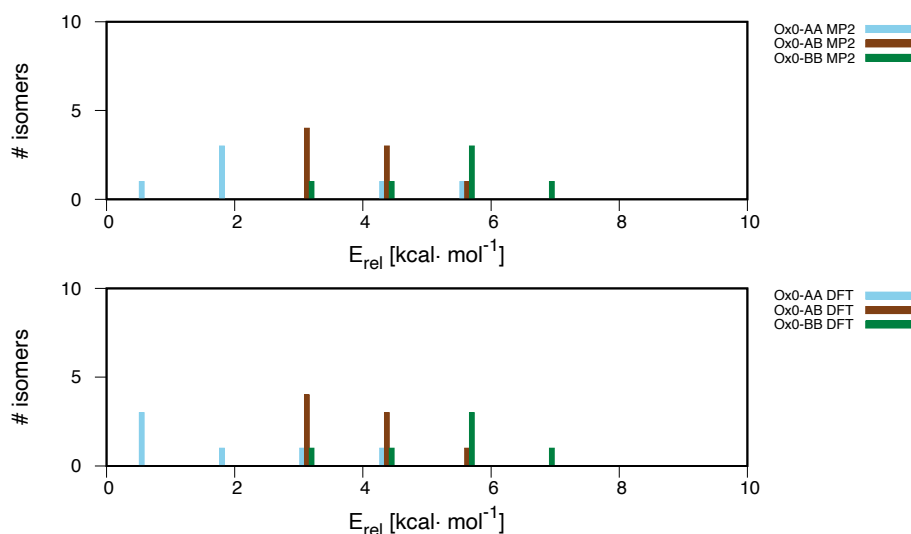

Figure SI2. Energy distribution of Ox0-lin isomers as a function of their type calculated with MP2/cc-pvtz and CAM-B3LYP/6-311G** (upper and lower panels, respectively).

SI1.1.2 Unstable tautomers

Two structures (z.lin.t.47.1.010-010 and z.lin.c.77.1.100-010) were not stable towards interfragment tautomerization. Upon optimization, they yielded structures z.lin.t.47.1.000-110 and z.lin.c.77.000-110, respectively. The unstable isomers were not included in the library.

SI1.1.3 Wave function stability of *z* structures

The structures with *z* resonance/bonding character may be prone to have a diradical character that makes them unsuitable for our theoretical approach. This possibility was examined reoptimizing all *z* structures with unrestricted CAM-B3LYP and comparing the results with the restricted approach. Four structures (z.lin.c.37.1.001-010, z.lin.c.37.1.010-001, z.lin.t.37.1.100-010 and z.lin.t.44.1.010-001) give a lower energy minimum with unrestricted CAM-B3LYP, which is indicative of instability of the restricted wave function and possible radical bicharacter. The structures were therefore discarded.

SI1.1.4 Benchmark of excited-state approach

The suitability of TD-CAM-B3LYP for the vertical excitations is assessed comparing the vertical excitations of the DHI, IQ and NQ monomers in the gas phase with high-level EOM-EE-CCSD and MS-CASPT2 calculations from Ref. [7] (Table SI1). TD-DFT excitations in solution are given in the fifth column. The differences of gas-phase TD-DFT with respect to EOM-EE-CCSD are 0.16 - 0.23 eV, and those with respect to MS-CASPT2 0-16 - 0.28 eV, *ie* overall there is a difference of approximately 0.2 eV which lies within the expected precision range of TD-CAM-B3LYP. Note that inclusion of solvation effects has a substantial effect on the excitation energies of IQ and NQ (decrease by approximately 0.2 eV) because of an increase in the dipole moment upon excitation.

Table SI1. Vertical S_1 excitation energy [eV] of the DHI monomer and the oxidized IQ and NQ forms calculated at different levels of theory.

	EOM-EE-CCSD/ 6-311++G(d,p)[a]	MS-CASPT2/ 6-311++G(d,p)[a]	TD-CAM-B3LYP/ 6-311G(d,p)[b]	TD-CAM-B3LYP/ 6-311G(d,p)[c]
DHI	4.55	4.58	4.78	4.79
IQ	2.33	2.43	2.15	1.92
NQ	2.37	2.05	2.21	2.04

[a]Gas phase, Ref. [7]. [b]Gas phase, this work. [c]Solution (SMD), this work.

SI1.2 Calculation of free energy

The free energy *G* of the different compounds is the free energy in solution at 298 K including translational, rotational, and vibrational contributions, and the electronic contribution for O_2. To save computational effort, the vibrational contribution of the dimers has been obtained from the SMD calculated frequencies for representative structures, namely the most stable cyclic and linear isomer of all oxidation states, and the same contribution has been used for all cyclic or linear isomers of the same oxidation state. Similarly, for all dimers the rotational entropy is approximated using the moments of inertia of the representative structures mentioned above. Following Eq. 6a of Ref. [8], for the Gibbs translational free energy correction $T \cdot S_{tr}$, the translational entropy is calculated using the free solvent volume of water with an intermediate value of 7 for

the C_{free} constant. For the dimers and water we assume standard conditions ($a_X = 1$), while for oxygen we assume an activity of 2.5e-4, which corresponds to the solubility of oxygen in water at 298 K (8 mg·L^{-1})[9]. The latter term is included as RT·ln(a_X) in $G(O_2)$.

SI1.3 Thermodynamic scheme to balance stoichiometry

To compare the energy of dimers with different stoichiometry, we consider that the oxidation proceeds with ground-state (triplet) oxygen, as in one of the common methods of dopa polymerization.[10] In this case, it is possible to compare the energy of isomers of different stoichiometry with the following equation, where the oxidation is balanced with O_2 and H_2O molecules (Equation S1):

$$C_{16}H_{12}N_2O_4 + m/2\ O_2 \longrightarrow C_{16}H_{12-2m}N_2O_4 + m\ H_2O \qquad [S1]$$

In Equation S1, $C_{16}H_{12}N_2O_4$ is the stochiometric formula of the fully reduced linear dimers (oxidation state 0). Using Equation S1, the energy of dimers with different composition can be compared calculating the following energy, which we term G^* (Equation S2):

$$G^*_{C_{16}H_{14-2m}N_2O_4} = G_{C_{16}H_{14-2m}N_2O_4} + mG_{H_2O} - (m/2)G_{O_2} \qquad [S2]$$

The linear dimers with oxidation state 0, 1, 2 and 3 correspond to m = 0, 1, 2 and 3. As for the cyclic dimers, they have 2 hydrogen atoms less than the linear counterparts. Therefore, cyclic dimers with oxidation state 0 - 3 have m = 1 - 4.

SI2. Classification scheme

SI2.1 Resonance/bonding classification

The resonance/bonding dimer classification is based on the fragment library shown in Charts SI1 and SI2. The monomer fragments (MF) are classified as MF-nH, where n is the number of oxidized sites, and labelled as b-$n.ijk$, where b is the bond connectivity at the bonding site, n the bonding site, and ijk codifies whether the N, O_5 or O_6 sites are reduced (digit=0) or oxidized (digit=1). In the cyclic structures where one of the heteroatoms is bonded, the corresponding digit is set to x. All linear and cyclic MF-0H and MF-2H fragments can be described by neutral resonance structures with single-bond connectivity at the connecting site(s), and are labeled s. In contrast, the MF-1H and MF-3H fragments can be described either by neutral resonance structures with double-bond connectivity at the connecting site(s), which are labeled d, or by zwitterionic resonance structures labeled z. The rules for the dimers are that two s fragments give an s dimer; two d fragments give a d dimer; and two z fragments or one d and one z fragment give a z dimer.

MF-0H

s-2.000 *s*-3.000 *s*-4.000 *s*-7.000

MF-1H

d-2.100 *d*-2.001 *d*-3.100 *d*-4.100 *d*-4.010 *d*-7.100

d-7.001 *z*-2.010 *z*-3.010 *z*-3.001 *z*-4.001 *z*-7.010

MF-2H

s-2.011 *s*-2.110 *s*-2.101 *s*-3.011 *s*-3.110 *s*-3.101

s-4.011 *s*-4.110 *s*-4.101 *s*-7.011 *s*-7.110 *s*-7.101

MF-3H

z-2.111 *d*-3.111 *d*-4.111 *d*-7.111

Chart SI1

6
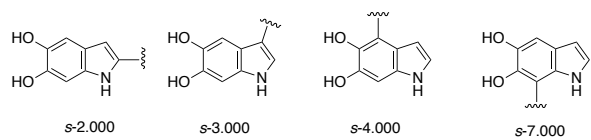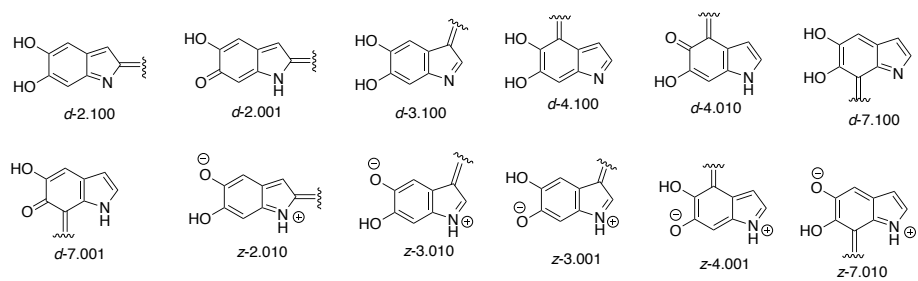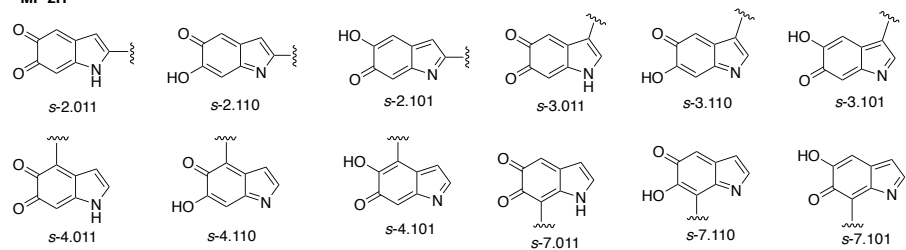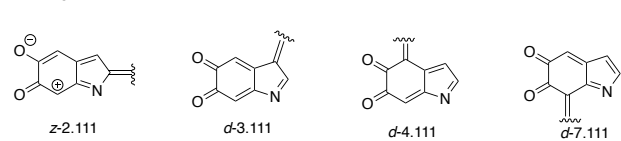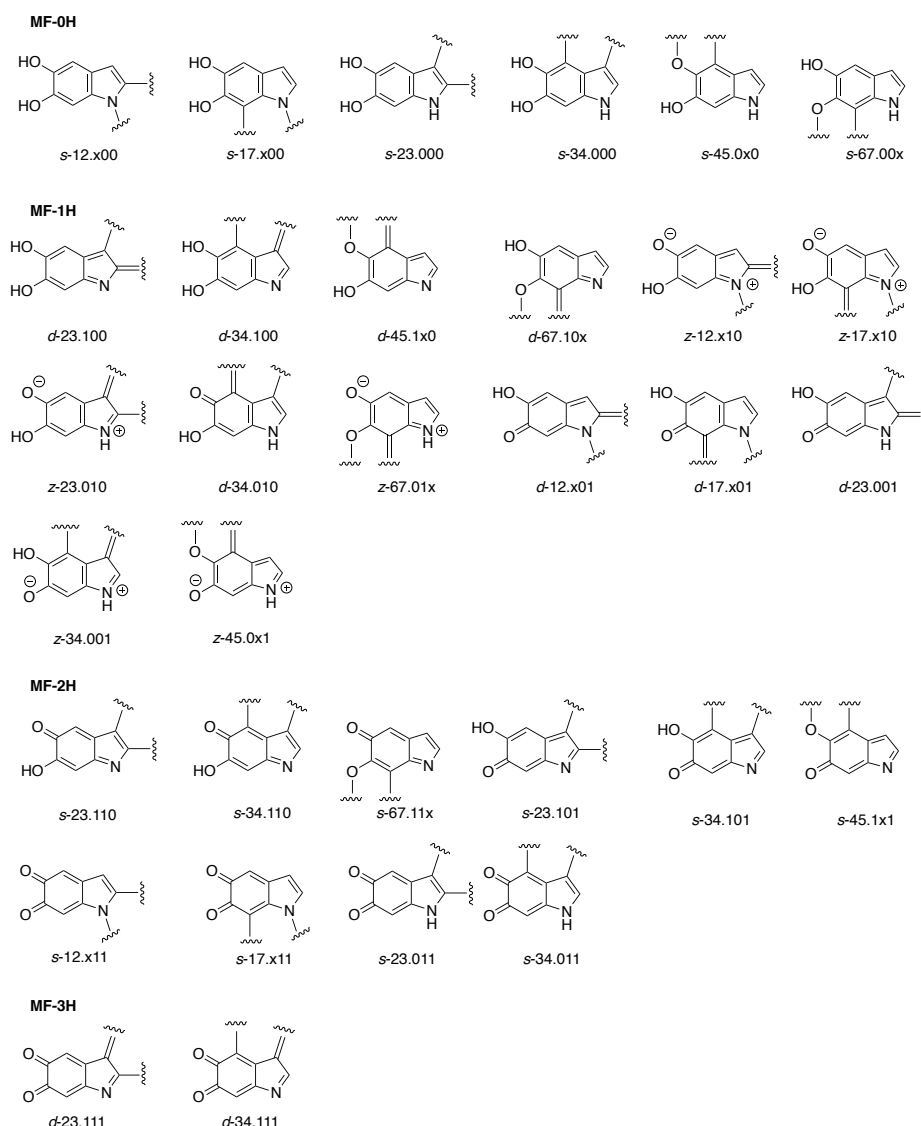
Chart SI2

SI2.2 Nomenclature

We introduce an unambiguous, specific nomenclature that describes the structural features in a compact way. For the linear structures, the name has the form *r.lin.u.mn.x.abc-def*, where *r* is the resonance/bonding key (*s, d* or *z*); *lin* stands for linear fragment; *u* describes the stereochemistry of the interfragment bond (*c* for cis and *t* for trans); *mn* gives the numbering of the carbons forming the interfragment bond; *x* gives the dimer oxidation state; and *abc* and *def* indicate whether N_1, O_5 and O_6 of the first and second fragment, respectively, are reduced (digit = 0) or oxidized (digit = 1). For the cyclic structures, the format is *r.cyc.mn.op.x.abc-def*, where *r*, *x* and *abc-def* have the same meaning that in the linear structures; *cyc* stands for cyclic fragment; and *mn* and *op* are the numbering of the interfragment connecting atoms of the first and second fragment (*m* connected with *o* and *n* with *p*). In the case where one of the N_1, O_5 and O_6 atoms is involved in the interfragment bond, the corresponding digit in the *abc-def* string is given the value *x*. Examples of the nomenclature are provided for the structures of Charts 1 and 2.
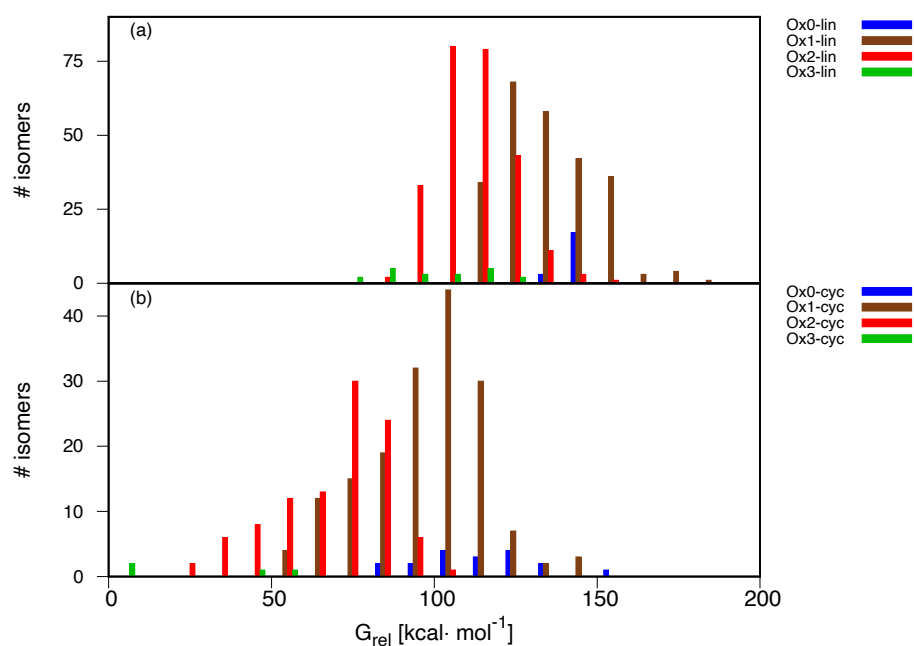
7

Full histograms of energy distribution

Figure SI3. Free energy distribution of (a) linear and (b) cyclic dimers as a function of their oxidation state, relative to the most stable library dimer (*cf* Figure 2).
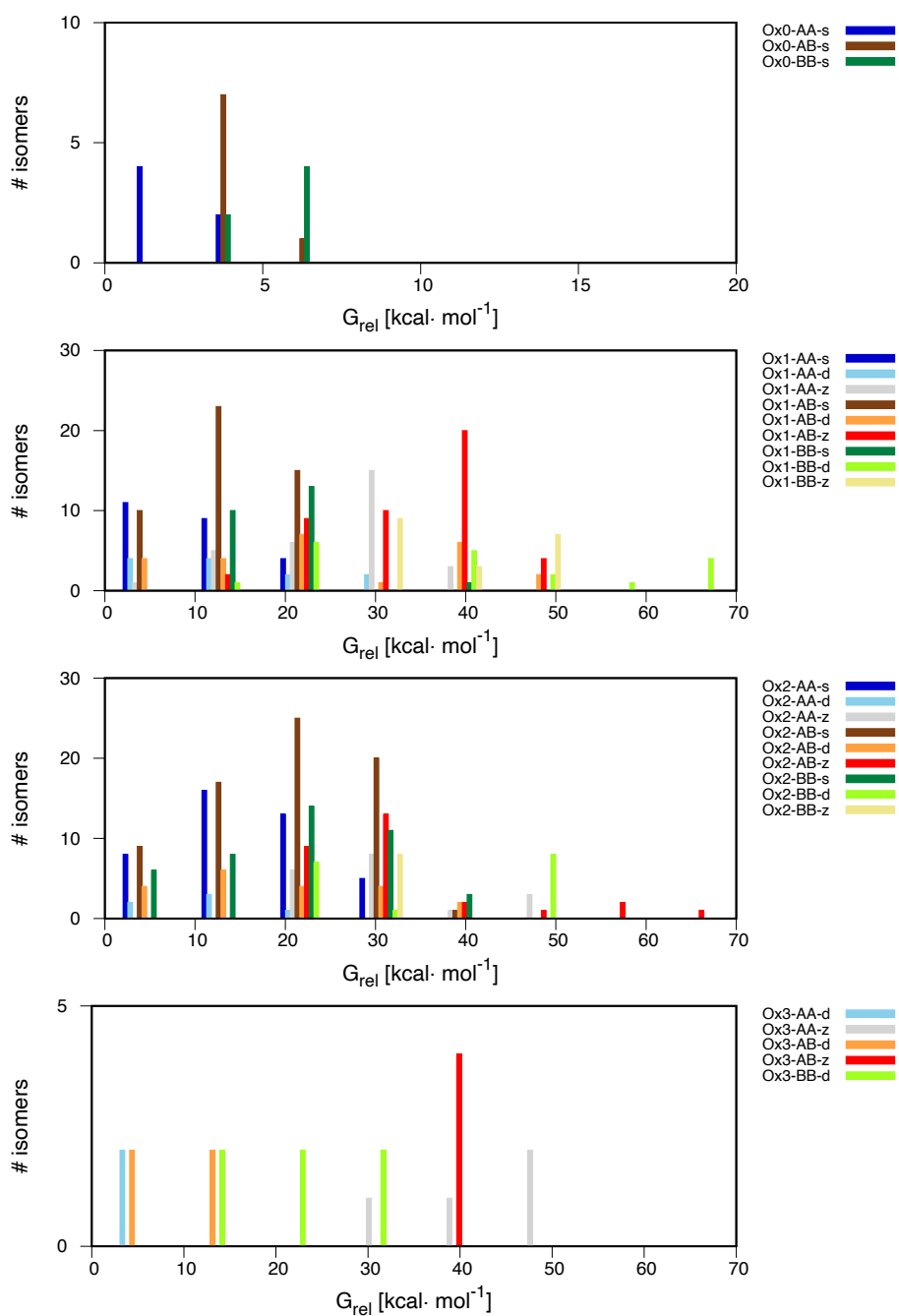
Figure SI4. Free energy distribution of linear isomers as a function of type. Compounds are grouped by oxidation state, and energies are relative to the most stable dimer of each group (*cf* Figure 3).

9
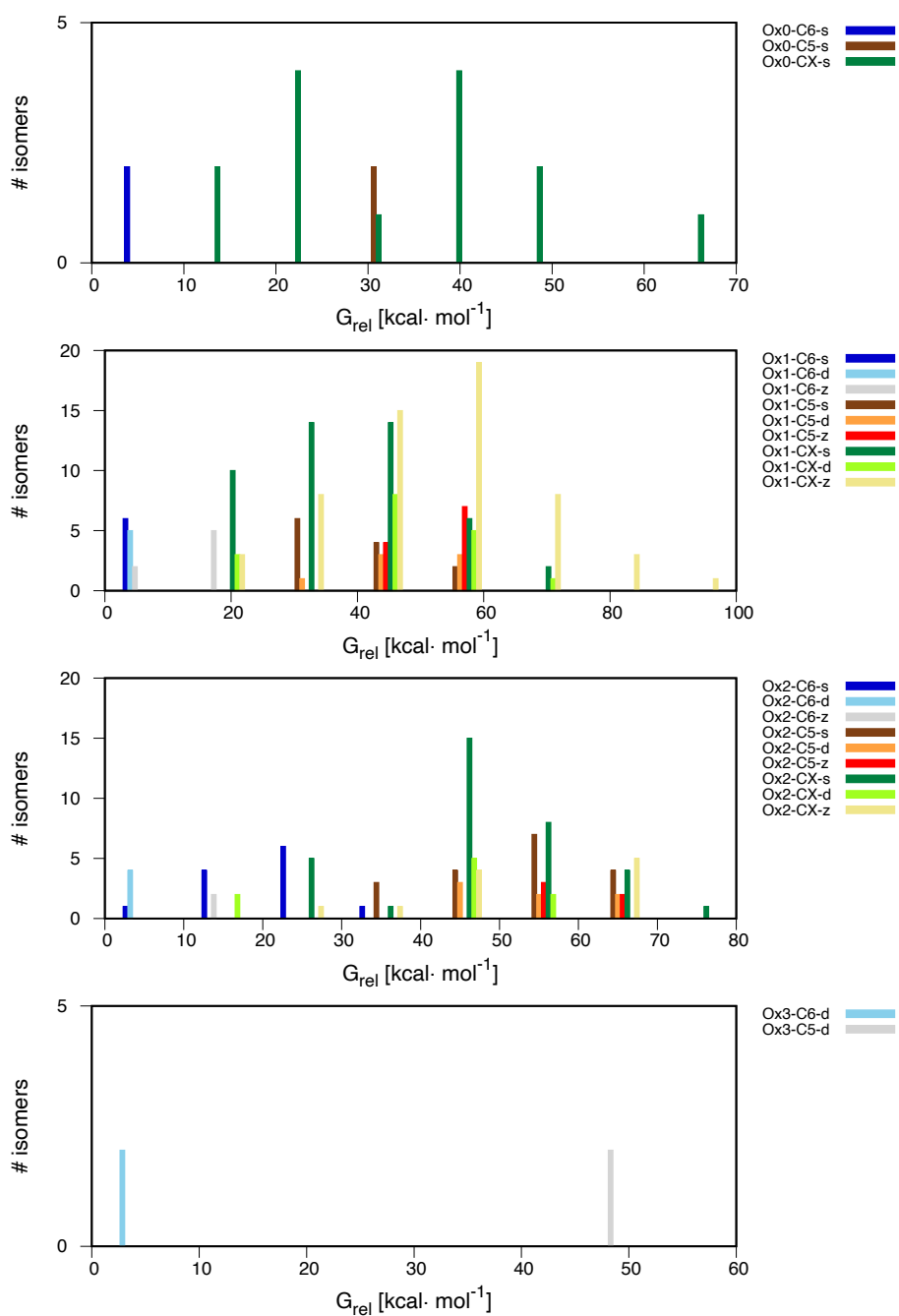
Figure SI5. Free energy distribution of cyclic isomers as a function of type. Compounds are grouped by oxidation state, and energies are relative to the most stable dimer of each group (*cf* Figure 4).

SI4. **Most stable linear isomers of each connectivity**

Table SI2. Most stable linear isomers of each connectivity.

Connectivity	Most stable isomer		ΔG [kcal·mol^{-1}][a]
	Name	Oxidation state	
22	s.lin.t.22.2.011-011	2	-30.7
23	d.lin.t.23.2.001-111	2	-18.5
24	d.lin.c.24.2.001-111	2	-19.2
27	d.lin.t.27.2.001-111	2	-26.0
33	d.lin.t.33.3.111-111	3	-20.6
34	d.lin.t.34.3.111-111	3	-14.8
37	d.lin.c.37.3.111-111	3	-12.0
44	d.lin.t.44.3.111-111	3	-7.5
47	d.lin.c.47.3.111-111	3	-1.5
77	s.lin.t.77.2.011-011	2	-5.2

[a]Energy relative to the lowest energy isomer with Ox3 (22, 23, 24, 27 and 77 connectivity) or Ox2 (33, 34, 37, 44, 47) oxidation state.

SI5. **Construction of absorption/stability plots**

To construct the absorption/stability plots (Figure 6), for every group of compounds a two-dimensional grid of bins is created with a spacing of 0.125 eV (x axis) and 5 kcal·mol^{-1} (y axis) (*eg* 41 x 5 for Figure 6a). The contents of every bin, z_{ij}, is the sum of the oscillator strengths assigned to that bin according to the following formula:

$$z_{ij} = \sum_{pq} f_{pq} \ \forall \left(G_p \in [G_j], E_{pq} \in [E_i] \right) \qquad [S3]$$

where the sum runs over p, the number of compounds in the group, and $q=1,3$, the number of vertical excitations calculated for every compound. The sum is performed when the energy of compound p, G_p, and the excitation energy E_{pq} fall into the bin relative energy and excitation ranges (G_j and E_i, respectively). After the sum is performed, the values of z are normalized to 1 dividing by the maximum of z_{ij}.

SI6. **Compounds contributing to the highlighted signals in the stability *vs* absorption plots (Figure 6)**

For every highlighted signal of Figure 6, the compounds that contribute most to the absorption (largest oscillator strength) are listed in Tables SI3 and SI4 (linear and cyclic dimers, respectively). The structures are displayed in Charts 1, 2 and SI3.

Table SI3. Most significant contributions to the highlighted absorption signals of the stability *vs* absorption plots, linear dimers (Figure 6a-d).

Group	Signal	G_{rel} [kcal·mol^{-1}]	E_{exc} [eV]	f	State	Compound name	Type	Chart
Ox0-lin	A	0.0	3.87	1.4002	S_1	s.lin.t.22.0.000-000	Ox0-AA-s	1
		1.2	3.89	1.3618	S_1	s.lin.c.22.0.000-000	Ox0-AA-s	SI3
	B	2.7	4.09	0.8725	S_1	s.lin.c.27.0.000-000	Ox0-AB-s	1
		2.9	4.00	0.8724	S_1	s.lin.c.24.0.000-000	Ox0-AB-s	SI3
		3.6	4.09	0.8481	S_1	s.lin.t.24.0.000-000	Ox0-AB-s	SI3
	C	1.1	4.26	0.8383	S_1	s.lin.c.23.0.000-000	Ox0-AA-s	SI3
		1.6	4.30	0.7923	S_1	s.lin.t.23.0.000-000	Ox0-AA-s	SI3
		3.3	4.28	0.7777	S_1	s.lin.t.27.0.000-000	Ox0-AB-s	SI3
	D	3.8	4.71	0.4052	S_2	s.lin.t.37.0.000-000	Ox0-AB-s	SI3
		3.3	4.75	0.3787	S_3	s.lin.t.33.0.000-000	Ox0-AA-s	SI3
		4.1	4.71	0.3159	S_1	s.lin.c.33.0.000-000	Ox0-AA-s	SI3
		3.8	4.65	0.2735	S_1	s.lin.c.37.0.000-000	Ox0-AB-s	SI3
		3.8	4.71	0.2504	S_2	s.lin.c.37.0.000-000	Ox0-AB-s	SI3
		1.1	4.76	0.2414	S_3	s.lin.c.23.0.000-000	Ox0-AA-s	SI3
Ox1-lin	E	8.2	1.70	0.2704	S_1	z.lin.t.23.1.001-001	Ox1-AA-z	1
	F	0.0	2.32	1.7509	S_1	d.lin.c.22.1.001-001	Ox1-AA-d	1
	G	6.0	2.83	1.2455	S_2	s.lin.c.22.1.101-000	Ox1-AA-s	SI3
		8.1	2.81	1.1386	S_2	s.lin.t.22.1.101-000	Ox1-AA-s	SI3
		8.2	2.77	0.8346	S_3	z.lin.t.23.1.001-001	Ox1-AA-z	SI3
		5.9	2.85	0.8222	S_3	d.lin.c.24.1.001-010	Ox1-AB-d	SI3
	H	4.4	2.88	1.1785	S_3	d.lin.c.23.1.001-100	Ox1-AA-d	SI3
		2.7	2.93	1.0876	S_3	s.lin.t.22.1.011-000	Ox1-AA-s	SI3
		4.4	2.95	1.0451	S_3	s.lin.c.22.1.011-000	Ox1-AA-s	SI3
		3.0	2.93	1.0347	S_3	d.lin.t.23.1.001-100	Ox1-AA-d	SI3
		3.2	2.80	0.8264	S_3	d.lin.c.24.1.001-010	Ox1-AB-d	SI3
Ox2-lin	I	4.6	1.87	0.1505	S_1	s.lin.t.22.2.011-011	Ox2-AA-s	SI3
		3.9	1.75	0.1401	S_1	s.lin.c.27.2.011-011	Ox2-AB-s	1
		4.0	1.82	0.1267	S_1	s.lin.t.23.2.011-011	Ox2-AA-s	SI3
	J	10.2	1.78	0.2333	S_1	s.lin.t.23.2.011-101	Ox2-AA-s	SI3
		5.9	1.64	0.2046	S_1	s.lin.t.24.2.011-011	Ox2-AB-s	SI3
		6.4	1.80	0.1776	S_1	s.lin.t.27.2.011-011	Ox2-AB-s	SI3
		9.9	1.73	0.1735	S_1	s.lin.t.24.2.101-011	Ox2-AB-s	SI3
	K	2.6	2.25	0.1723	S_1	d.lin.t.27.2.001-111	Ox2-AB-d	1
	L	2.7	2.73	0.9111	S_3	d.lin.c.24.2.001-111	Ox2-AB-d	SI3
		4.2	2.73	0.8692	S_3	d.lin.c.27.2.001-111	Ox2-AB-d	SI3
		4.7	2.78	0.9011	S_3	d.lin.t.24.2.001-111	Ox2-AB-d	SI3
Ox3-lin	M	9.2	2.65	0.2221	S_2	d.lin.t.37.3.111-111	Ox3-AB-d	SI3
		9.3	2.62	0.1924	S_2	d.lin.c.37.3.111-111	Ox3-AB-d	SI3

Table SI4. Most significant contributions to the highlighted absorption signals of the stability *vs* absorption plots, cyclic dimers (Figure 6a-d).

Group	Signal	G_{rel} [kcal·mol^{-1}]	E_{exc} [eV]	f	State	Compound name	Type	Chart
Ox0-cyc	N	1.8	3.38	0.3404	S$_1$	s.cyc.34.34.0.000-000	Ox0-C6-s	SI4
	P	0.0	3.76	0.5672	S$_1$	s.cyc.34.43.0.000-000	Ox0-C6-s	2
Ox1-cyc	Q	1.3	1.37	0.1701	S$_1$	z.cyc.34.43.1.010-010	Ox1-C6-z	2
		3.8	1.31	0.1064	S$_1$	s.cyc.34.43.1.000-011	Ox1-C6-s	SI4
		8.6	1.24	0.0889	S$_1$	z.cyc.34.43.1.001-010	Ox1-C6-z	SI4
		9.7	1.27	0.0874	S$_1$	s.cyc.34.43.1.000-110	Ox1-C6-s	SI4
		4.5	1.13	0.081	S$_1$	s.cyc.34.34.1.011-000	Ox1-C6-s	SI4
	R	0.7	1.90	0.1198	S$_1$	d.cyc.34.43.1.010-100	Ox1-C6-d	2
		2.8	1.76	0.0924	S$_2$	d.cyc.34.34.1.010-010	Ox1-C6-d	SI4
		0.0	1.79	0.0897	S$_1$	s.cyc.34.34.1.110-000	Ox1-C6-s	2
Ox2-cyc	S	9.1	1.31	0.1900	S$_1$	s.cyc.34.43.2.011-011	Ox2-C6-s	2
Ox3-cyc	T	0.0	3.30	0.0973	S$_3$	d.cyc.34.43.3.111-111	Ox3-C6-d	2

A
Ox0-AA-s, G_{rel}=1.2
s.lin.c.22.0.000-000
E_{S1}=3.89 (*f*=1.36)

B
Ox0-AB-s, G_{rel}=2.9
s.lin.c.24.0.000-000
E_{S1}=4.00 (*f*=0.87)

B
Ox0-AB-s, G_{rel}=3.6
s.lin.t.24.0.000-000
E_{S1}=4.09 (*f*=0.85)

C,D
Ox0-AA-s, G_{rel}=1.1
s.lin.c.23.0.000-000
E_{S1}=4.26 (*f*=0.84)
E_{S3}=4.76 (*f*=0.24)

C
Ox0-AA-s, G_{rel}=1.6
s.lin.t.23.0.000-000
E_{S1}=4.30 (*f*=0.79)

C
Ox0-AB-s, G_{rel}=3.3
s.lin.t.27.0.000-000
E_{S1}=4.28 (*f*=0.78)

D
Ox0-AB-s, G_{rel}=3.8
s.lin.t.37.0.000-000
E_{S1}=4.71 (*f*=0.41)

D
Ox0-AA-s, G_{rel}=3.3
s.lin.t.33.0.000-000
E_{S3}=4.75 (*f*=0.38)

D
Ox0-AA-s, G_{rel}=4.1
s.lin.c.33.0.000-000
E_{S1}=4.71 (*f*=0.32)

D
Ox0-AB-s, G_{rel}=3.8
s.lin.c.37.0.000-000
E_{S1}=4.65 (*f*=0.27)
E_{S2}=4.71 (*f*=0.25)

G
Ox1-AA-s, G_{rel}=6.0
s.lin.c.22.1.101-000
E_{S2}=2.83 (*f*=1.25)

G
Ox1-AA-s, G_{rel}=8.1
s.lin.t.22.1.101-000
E_{S2}=2.81 (*f*=1.14)

G
Ox1-AA-z, G_{rel}=8.2
z.lin.t.23.1.001-001
E_{S3}=2.77 (*f*=0.83)

G
Ox1-AB-d, G_{rel}=5.9
d.lin.t.24.1.001-010
E_{S3}=2.85 (*f*=0.82)

H
Ox1-AA-d, G_{rel}=4.4
d.lin.c.23.1.001-100
E_{S3}=2.88 (*f*=1.18)

H
Ox1-AA-s, G_{rel}=2.7
s.lin.t.22.1.011-000
E_{S3}=2.93 (*f*=1.09)

H
Ox1-AA-s, G_{rel}=4.4
s.lin.c.22.1.011-000
E_{S3}=2.95 (*f*=1.05)

H
Ox1-AA-d, G_{rel}=3.0
d.lin.t.23.1.001-100
E_{S3}=2.93 (*f*=1.03)

H
Ox1-AB-d, G_{rel}=3.2
d.lin.c.24.1.001-010
E_{S3}=2.80 (*f*=0.83)

I
Ox2-AA-s, G_{rel}=4.6
s.lin.t.22.2.011-011
E_{S1}=1.87 (*f*=0.15)

I
Ox2-AA-s, G_{rel}=4.0
s.lin.t.23.2.011-011
E_{S1}=1.82 (*f*=0.13)

J
Ox2-AA-s, G_{rel}=10.2
s.lin.t.23.2.011-101
E_{S1}=1.78 (*f*=0.23)

J
Ox2-AB-s, G_{rel}=5.9
s.lin.t.24.2.011-011
E_{S1}=1.64 (*f*=0.20)

J
Ox2-AB-s, G_{rel}=6.4
s.lin.t.27.2.011-011
E_{S1}=1.80 (*f*=0.18)

J
Ox2-AB-s, G_{rel}=9.9
s.lin.t.24.2.101-011
E_{S1}=1.73 (*f*=0.17)

L
Ox2-AB-d, G_{rel}=2.7
d.lin.c.24.2.001-111
E_{S3}=2.73 (*f*=0.91)

L
Ox2-AB-d, G_{rel}=4.2
d.lin.c.27.2.001-111
E_{S3}=2.73 (*f*=0.87)

L
Ox2-AB-d, G_{rel}=4.7
d.lin.t.24.2.001-111
E_{S3}=2.78 (*f*=0.90)

M
Ox3-AB-d, G_{rel}=9.2
d.lin.c.37.3.111-111
E_{S2}=2.65 (*f*=0.22)

M
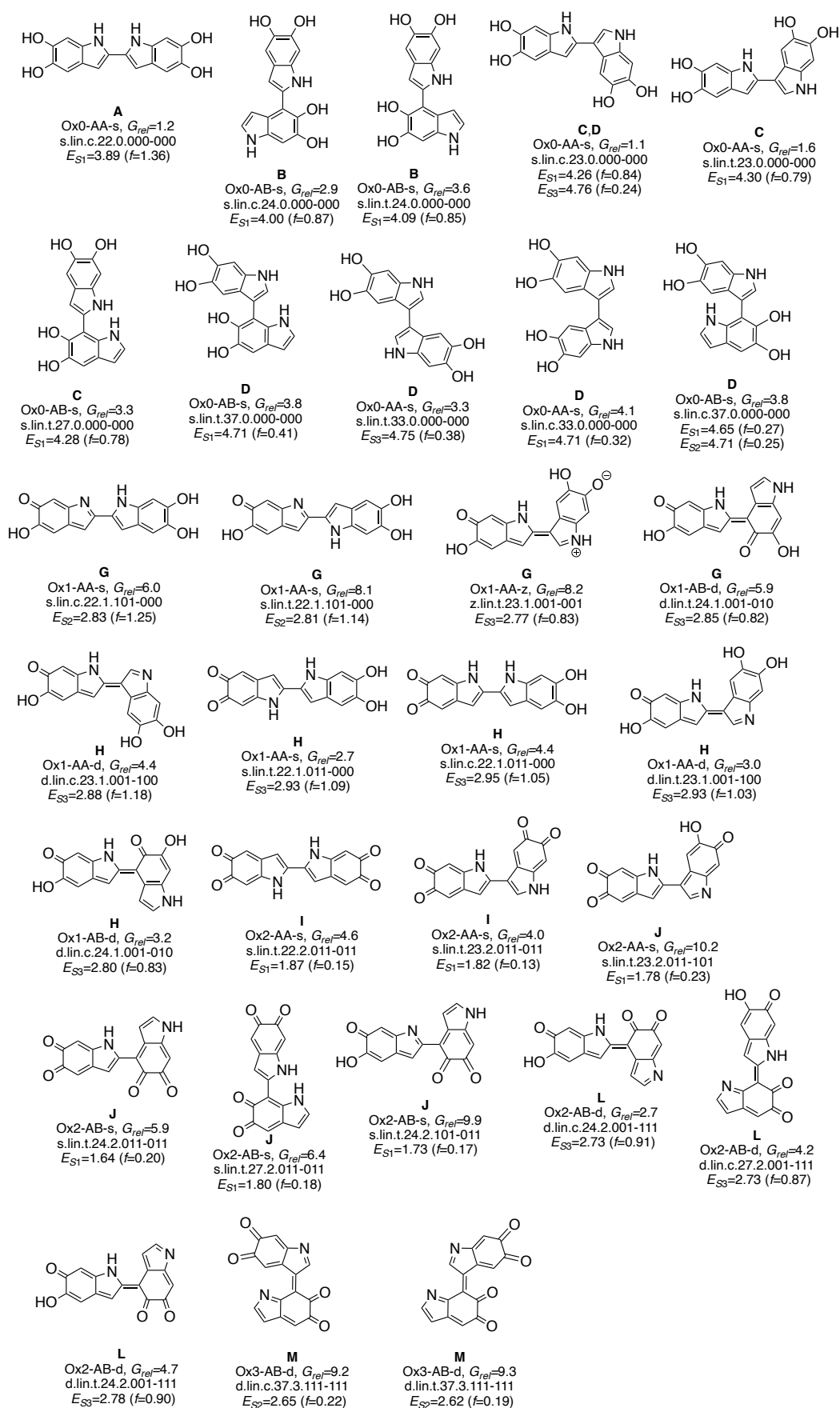Ox3-AB-d, G_{rel}=9.3
d.lin.t.37.3.111-111
E_{S2}=2.62 (*f*=0.19)

Chart SI3. Additional compounds highlighted in Figure 6 (linear dimers). Labelling follows the convention of Charts 1 and 2 in the main text.

14

N
Ox0-C6-s, G_{rel}=1.8
s.cyc.34.34.0.000-000
E_{S1}=3.38 (f=0.34)

Q
Ox1-C6-s, G_{rel}=3.8
s.cyc.34.43.1.000-011
E_{S1}=1.31 (f=0.11)

Q
Ox1-C6-z, G_{rel}=8.6
z.cyc.34.43.1.001-010
E_{S1}=1.24 (f=0.09)

Q
Ox1-C6-s, G_{rel}=9.7
s.cyc.34.43.1.000-110
E_{S1}=1.27 (f=0.08)

Q
Ox1-C6-s, G_{rel}=4.5
s.cyc.34.34.0.011-000
E_{S1}=1.13 (f=0.08)

R
Ox1-C6-d, G_{rel}=2.8
d.cyc.34.34.0.010-010
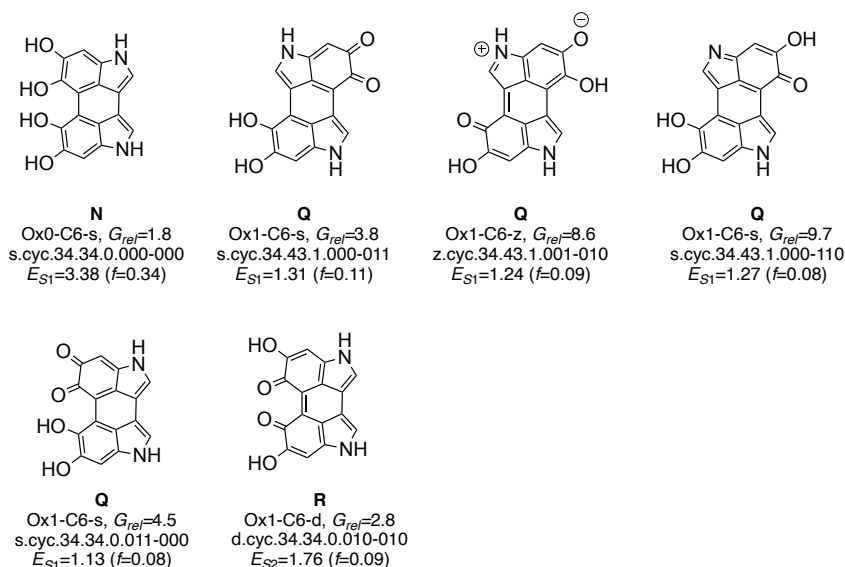E_{S2}=1.76 (f=0.09)

Chart SI4. Additional compounds highlighted in Figure 6 (cyclic dimers). Labelling follows the convention of Charts 1 and 2 in the main text.

SI7.   **Complete reference 26**

Gaussian 16, Revision A.03, Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Petersson, G. A.; Nakatsuji, H.; Li, X.; Caricato, M.; Marenich, A. V.; Bloino, J.; Janesko, B. G.; Gomperts, R.; Mennucci, B.; Hratchian, H. P.; Ortiz, J. V.; Izmaylov, A. F.; Sonnenberg, J. L.; Williams-Young, D.; Ding, F.; Lipparini, F.; Egidi, F.; Goings, J.; Peng, B.; Petrone, A.; Henderson, T.; Ranasinghe, D.; Zakrzewski, V. G.; Gao, J.; Rega, N.; Zheng, G.; Liang, W.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Throssell, K.; Montgomery, J. A., Jr.; Peralta, J. E.; Ogliaro, F.; Bearpark, M. J.; Heyd, J. J.; Brothers, E. N.; Kudin, K. N.; Staroverov, V. N.; Keith, T. A.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A. P.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Millam, J. M.; Klene, M.; Adamo, C.; Cammi, R.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Farkas, O.; Foresman, J. B.; Fox, D. J. Gaussian, Inc., Wallingford CT, 2016.

SI8.   **SI References**

1.   Laurent, A. D.; Jacquemin, D. TD-DFT benchmarks: A review. *Int. J. Quantum Chem.* **2013,** *113*, 2019-2039.

2.   Suellen, C.; Freitas, R. G.; Loos, P.-F.; Jacquemin, D. Cross-Comparisons between Experiment, TD-DFT, CC, and ADC for Transition Energies. *J. Chem. Theory Comput.* **2019,** *15*, 4581-4590.

3.   Dreuw, A.; Head-Gordon, M. Single-reference ab initio methods for the calculation of excited states of large molecules. *Chem. Rev.* **2005,** *105*, 4009-4037.

4.   Marenich, A. V.; Cramer, C. J.; Truhlar, D. G. Universal Solvation Model Based on Solute Electron Density and on a Continuum Model of the Solvent Defined by the Bulk Dielectric Constant and Atomic Surface Tensions. *J. Phys. Chem. B* **2009,** *113*, 6378-6396.

5.   Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.*, et al. Gaussian 09, Revision E.01*, Revision E.01; Gaussian, Inc.: Wallingford, CT, 2013.

6.   Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu. *J. Chem. Phys.* **2010,** *132*, 154104.

7.   Ghosh, D. Computational aspects towards understanding the photoprocesses in eumelanin. *WIREs Comput. Mol. Sci.* **2021,** *11*, e1505.

8.   Mammen, M.; Shakhnovich, E. I.; Deutch, J. M.; Whitesides, G. M. Estimating the entropic cost of self-assembly of multiparticle hydrogen-bonded aggregates based on the cyanuric acid center dot melamine lattice. *J. Org. Chem.* **1998,** *63*, 3821-3830.

9.   Kutty, M. N. Site Selection for Aquaculture. Chemical Features of Water. http://www.fao.org/3/AC183E/AC183E00.htm (accessed April 15th 2021).

10.   d'Ischia, M.; Wakamatsu, K.; Napolitano, A*., et al.* Melanins and melanogenesis: methods, standards, protocols. *Pigment Cell Melanoma Res.* **2013,** *26*, 616-633.