Computational screening of bioinspired mixed ionic-electronic conductors

Tristan Stephens-Jones[a] and Micaela Matta[a,*]

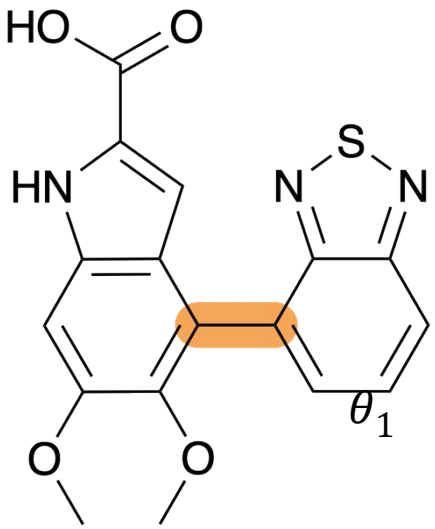
[a.] King's College Department of Chemistry, Strand, WC2R 2LS

micaela.matta@kcl.ac.uk

Table of Contents

S1 – Planarity Calculation

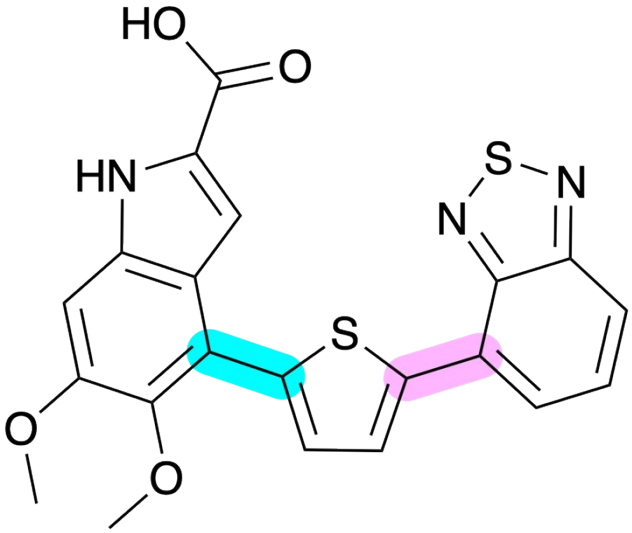
a)



$$P = cos^2(\theta)$$

θ_2

b)



$$P = \frac{cos^2(\theta_1) + cos^2(\theta_2)}{2}$$

Figure S1. Donor-Acceptor inter-ring planarity definition for **a)** systems featuring single bond linkers and **b)** double, triple, imine and thiophene linkers.

S2 – SAscore

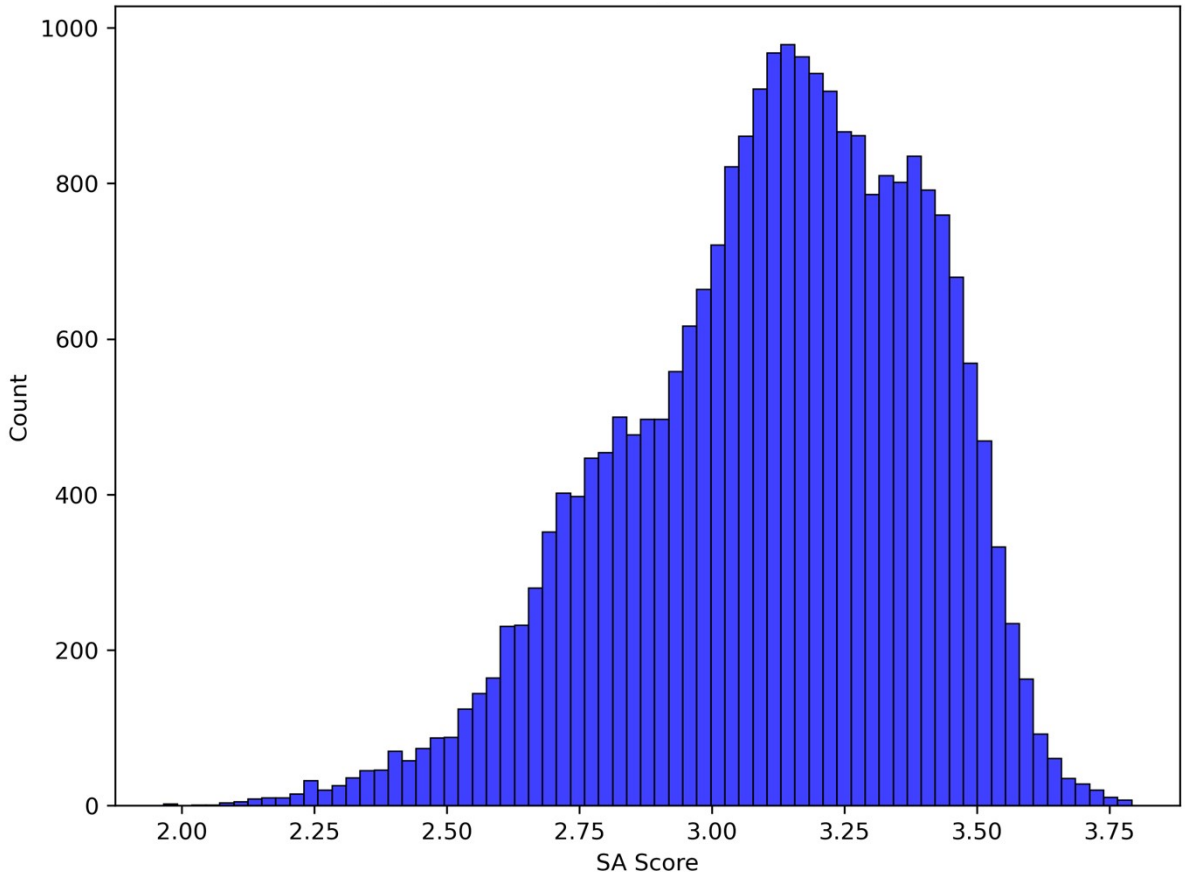


Figure S2. Distribution of the synthetic accessibility of the 24,990 strong system dataset. The score ranges from 1-10, where 1 is 'most' synthetically accessible and 10 being the least synthetically accessible. The SA score distribution of this eumelanin inspired dataset shows good predicted synthetically accessibility.

S3 – Planarity and Energy Gap with Different Functional Groups

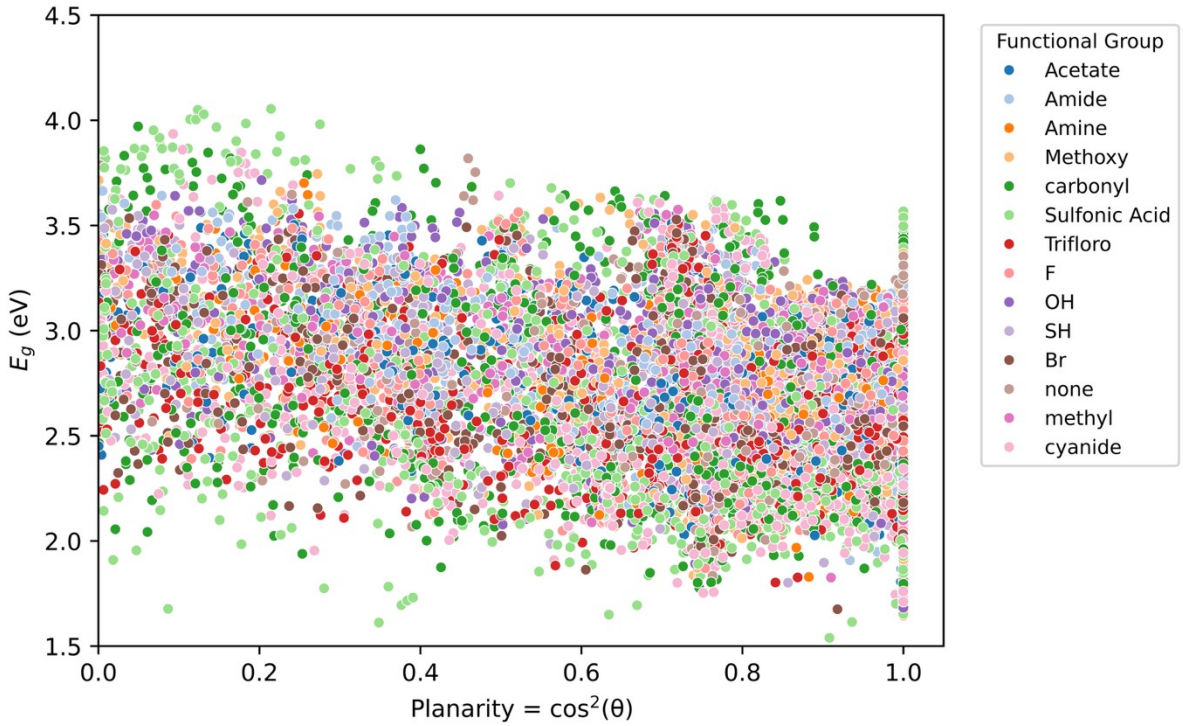


Figure S3. Effect of different functional groups on the planarity score and E_g. Of the 14,363 D-L-A systems surveyed at this stage, 3,353 (23.3%) of them fit within the filter regime of planarity >0.67 and energy gap < 2.5 eV.

S4 – Functional Groups Effect on Top Performing Planarity and Energy Gap Systems

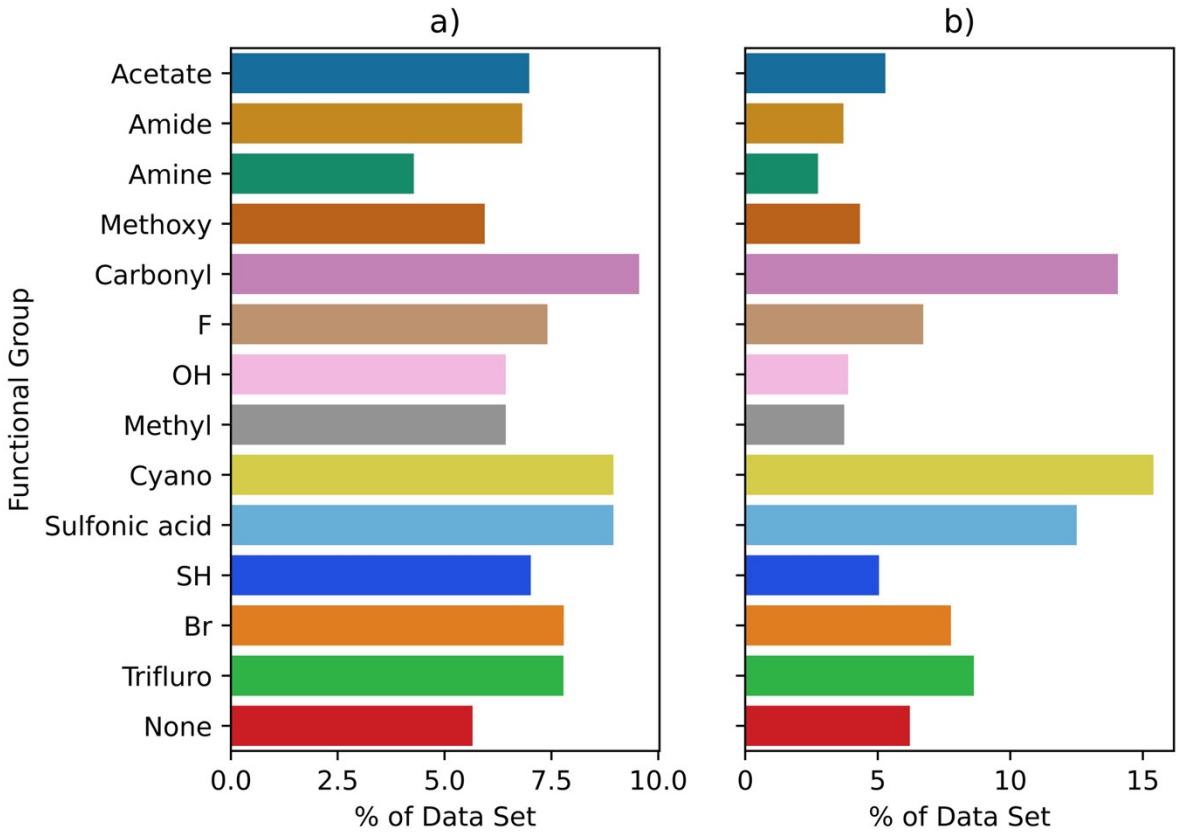


Figure S4. a) 14,363 molecules that underwent planarity and energy gap calculations. **b)** is 3,353 molecules who have energy gap <2.5 eV and planarity >0.67

S5 – Linkers Effect on Top-Performing Planarity and Energy Gap Systems

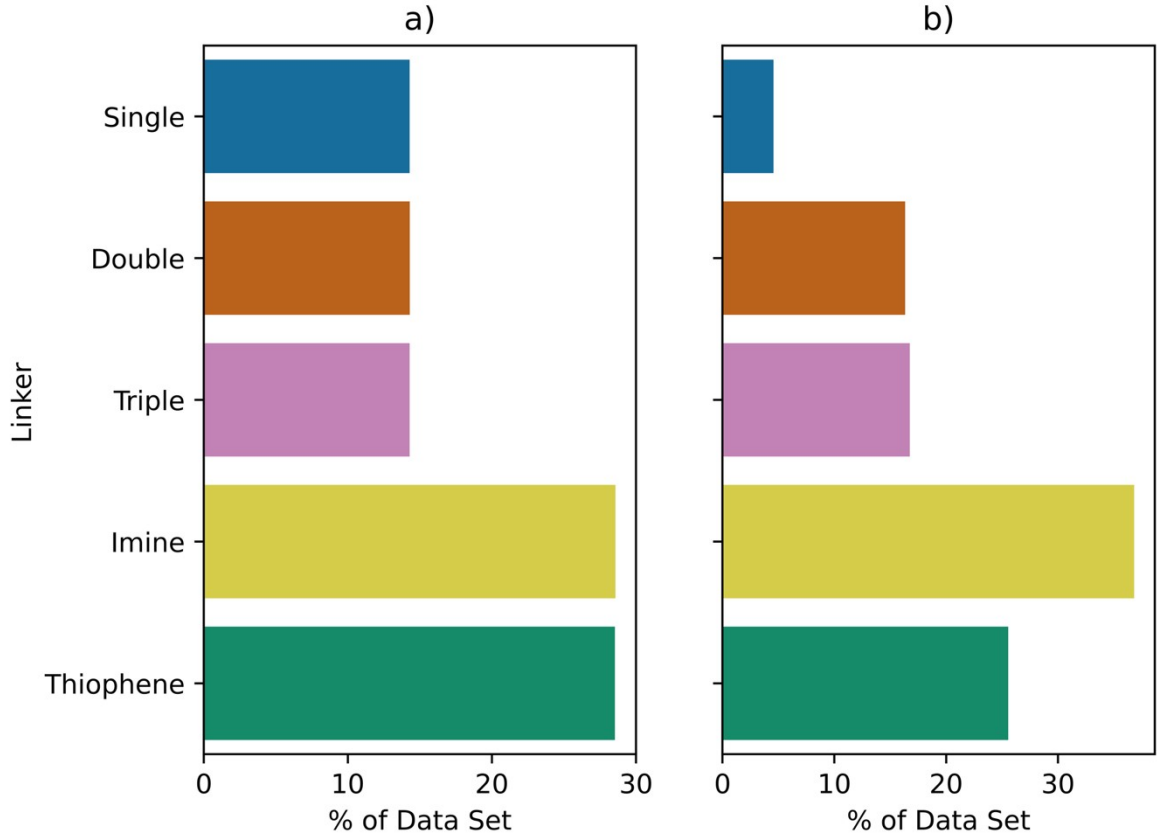


Figure S5. a) 14,363 molecules that underwent planarity and energy gap calculations. **b)** is 3,353 molecules who have energy gap <2.5 eV and planarity >0.67

S6 - Eumelanin Bonding Position Effect on Top-Performing Planarity and Energy Gap Systems

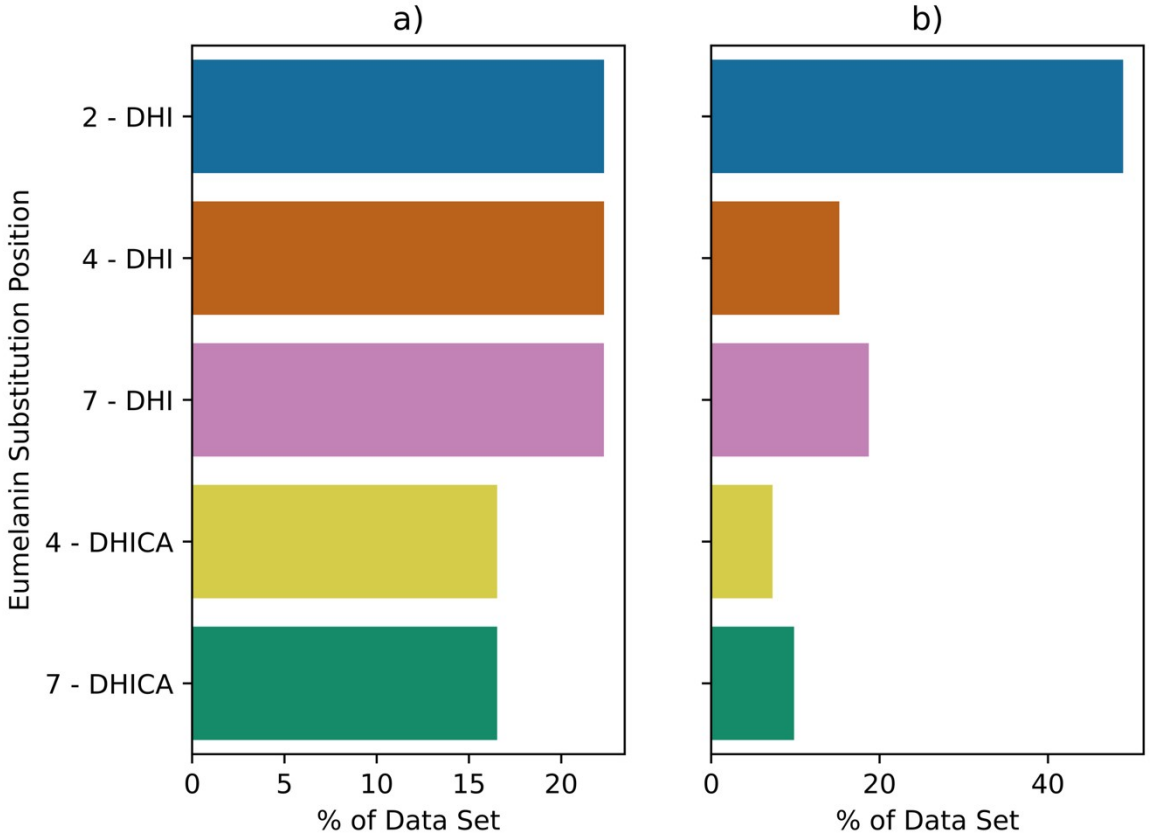


Figure S6. The bonding position of the linker and fragment to the DHI and DHICA eumelanin is depicted for **a)** all the energy gap and planarity systems and **b)** the top performing systems. Position 2 is shown to be most favourable.

S7 - Eumelanin Functionalisation Effect on Top-Performing Planarity and Energy Gap Systems

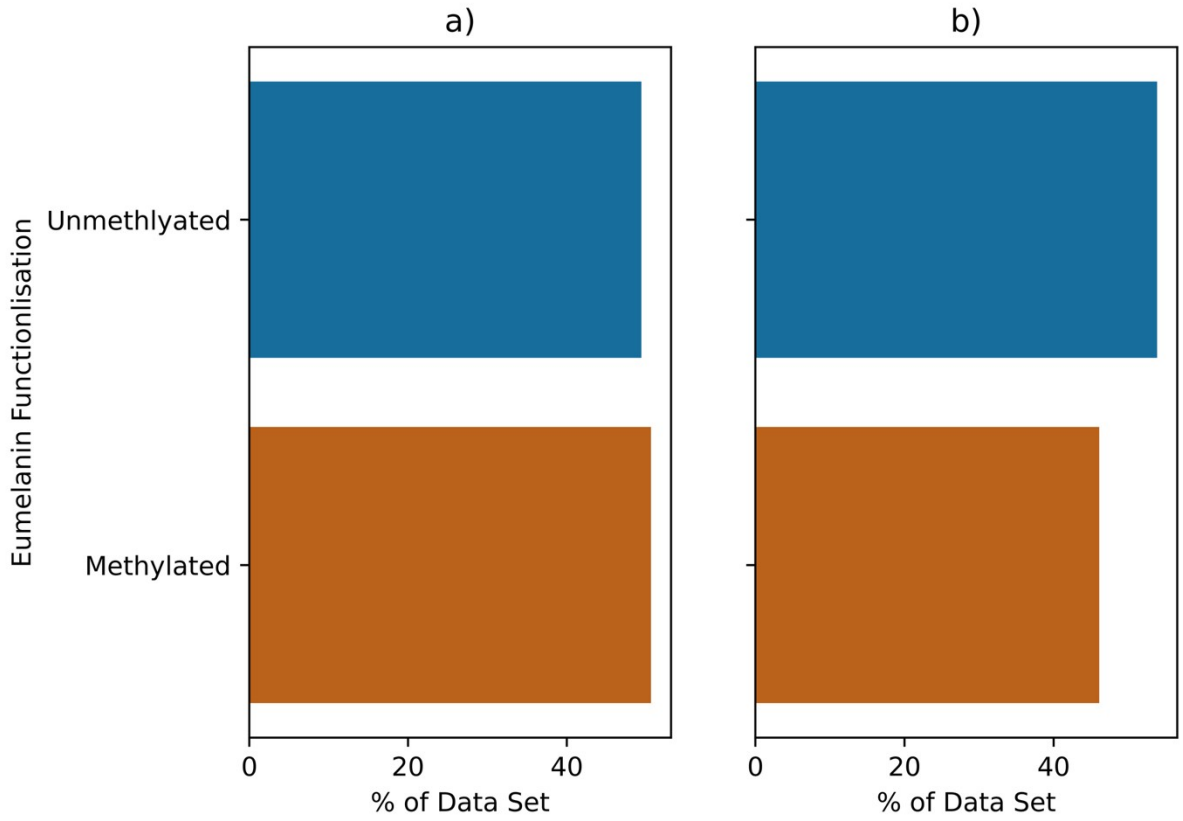


Figure S7. a) all the energy gap and planarity systems and **b)** the top performing systems. The nature of the nitrogen substitution within the DHI and DHICA fragment has no significant effect on the energy gap and planarity of the molecule.

S8 – Linkers Effect on Planarity and Energy Gap

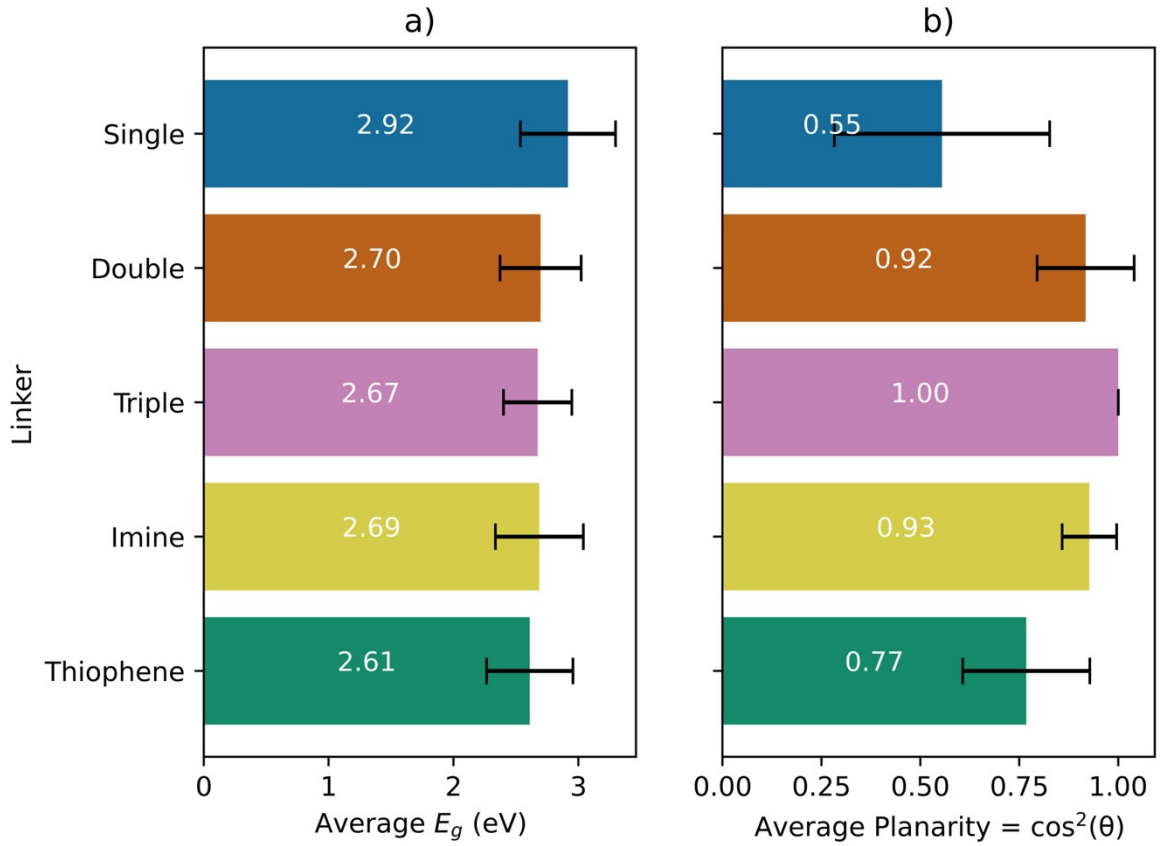


Figure S8. The effect that different linkers have on the; **a)** average energy gap, where thiophene has the lowest average **b)** average planarity where the triple bond linker has the highest planarity.

S9 –Effect of Functional Group on Planarity and Energy Gap

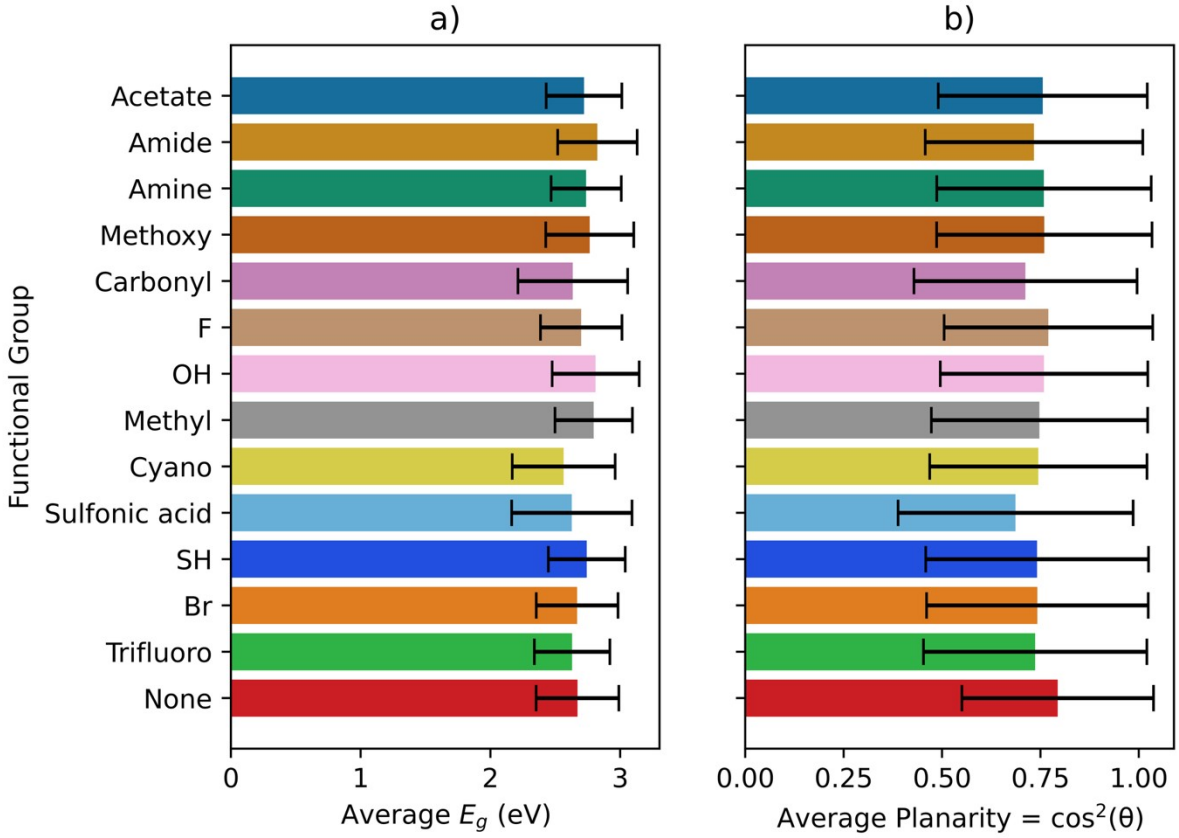


Figure S9. Effect of functional group substitution on average values of **a)** energy gap E_g and **b)** planarity (14,363 D-L-A systems considered).

S10 – Linkers Effect on Reorganisation Energy

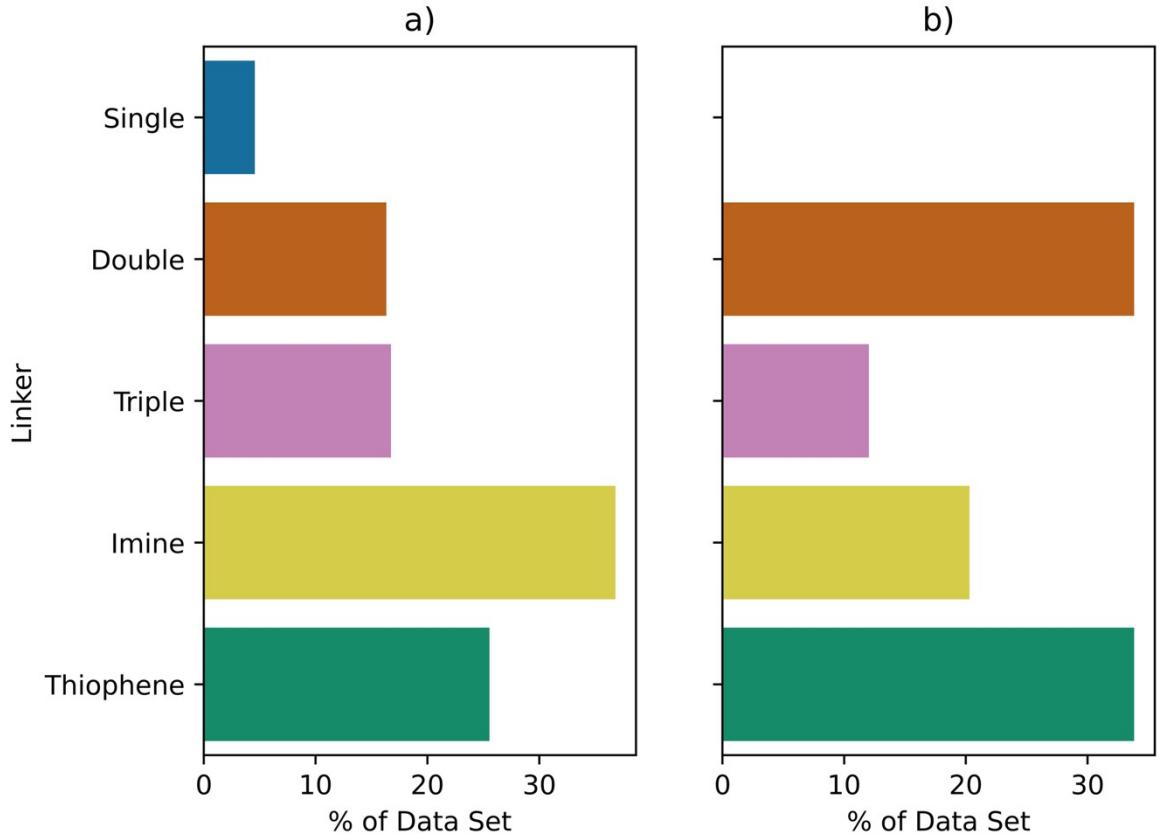


Figure S10. Linker distribution across **a)** the 3,350 D-L-A systems selected for reorganisation energy calculation and **b)** the 133 D-L-A systems with λ^- below the target threshold of 250 meV. Double bond and thiophene linkers are overrepresented in low λ^-structures.

S11 – Functional Group Effect on Reorganisation Energy

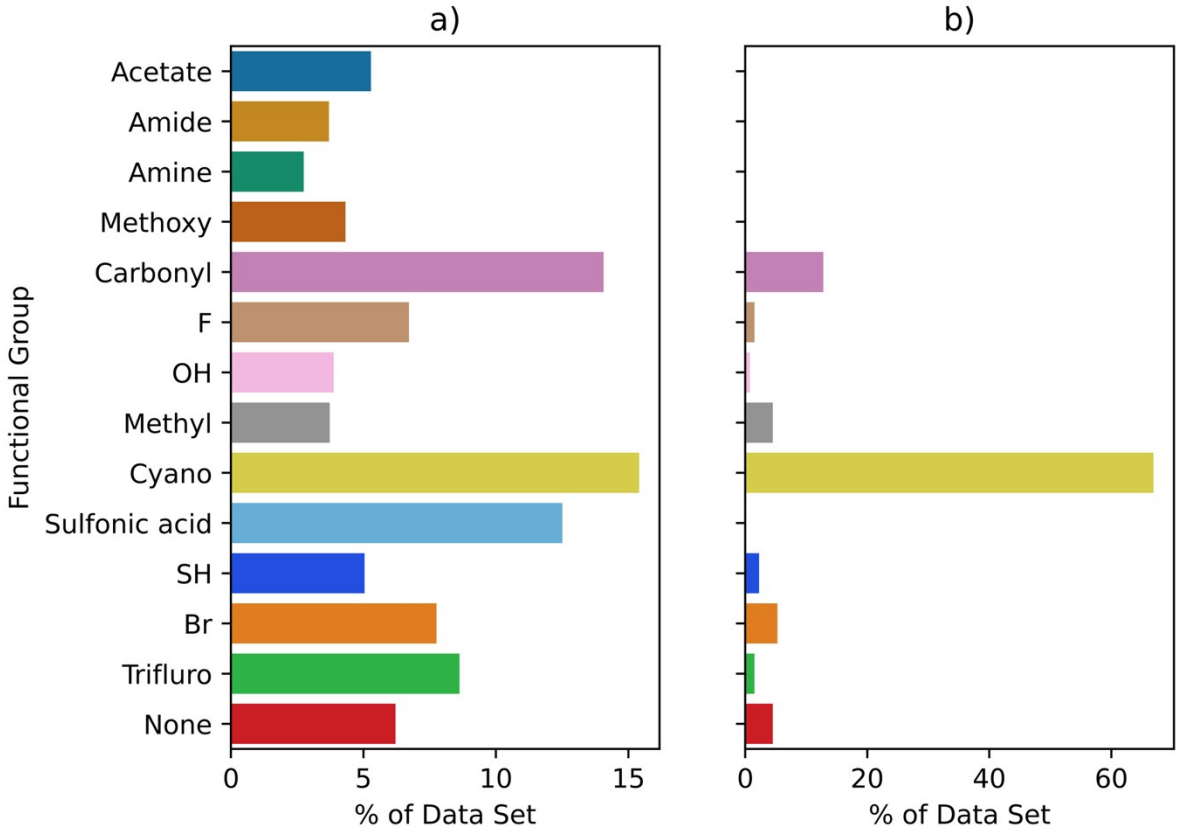


Figure S11. Functional group distributions for **a)** the 3,350 D-L-A systems that undergo reorganisation energy calculation; **b)** the 133 dyads having $\lambda^{+/-} \leq 250$ meV. The cyano functional group is overrepresented in eumelanin- inspired D-L-A systems with low λ^{-}.

S12 – Eumelanin Bonding Position Effect on Reorganisation Energy

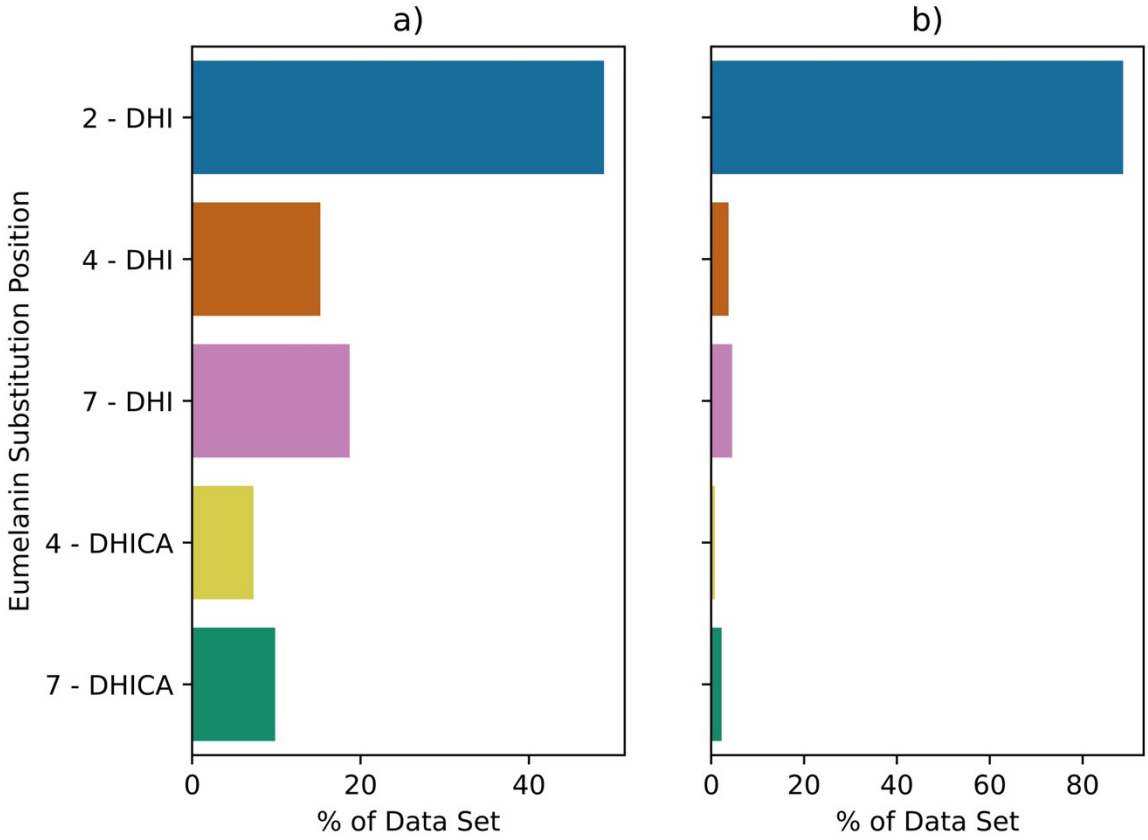


Figure S12. The bonding position of the linker and fragment to the DHI and DHICA eumelanin is depicted for **a)** all the reorganisation energy systems and **b)** the top performing systems. Position 2 is shown to be most favourable.

S13 – Eumelanin Functionalisation Effect on Reorganisation Energy

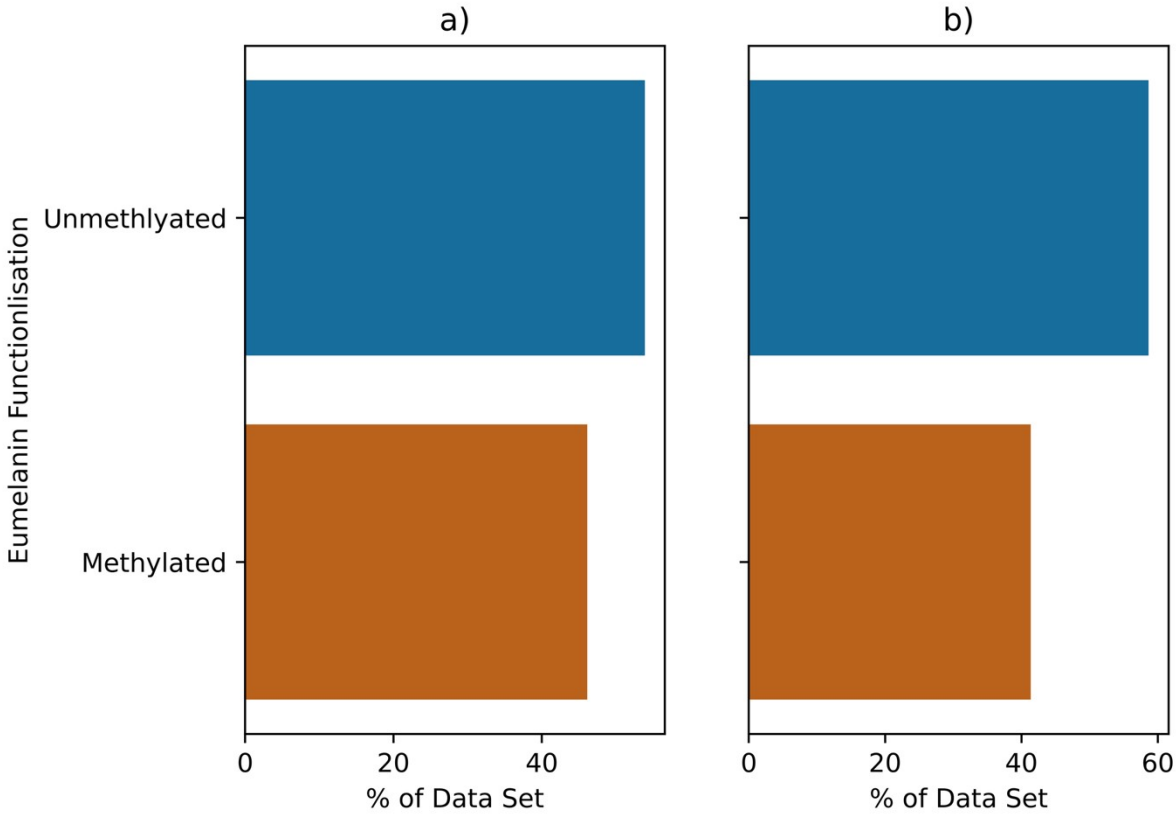


Figure S13. The nature of the nitrogen substation within the DHI and DHICA fragment has no significant bearing on the λ of the systems.

S14 – Cost Function Explanation

Two cost function equations are utilised within the GA. Equation 3 is the cost function carried out at the stage for picking the 25% of the loop to go forward for λ calculations. Equation 4 in the cost function carried out to rank the elite systems for mutation.

$$C_{elite} = C_P + C_{Eg} + 0.25 \cdot C_{SAscore} \quad \text{(Equation 3)}$$

$$C_{mut} = C_{elite} + 2 \cdot C_{\lambda^-} + 2 \cdot C_{\lambda^+} \quad \text{(Equation 4)}$$

Here we take equation 3 and show an example of how it may look in action, all these numbers are fabricated as this is just an example.

Table S14a. This table shows fabricated results of 5 systems.

System	Eg /eV	P	SAscore
1	3.2	0.9	3.2
2	2.5	0.77	2.4
3	1.7	0.85	2.7
4	2.1	0.7	3.1
5	3.1	0.97	2.9

Table S14b. The cost functions are applied to each of the systems and their final score.

System	C_{Eg}	C_P	$C_{SAscore}$	C_{tot}
1	5	2	5	8.25
2	3	4	1	7.25
3	1	3	2	4.5
4	2	5	4	8
5	4	1	3	5.75

An example cost function calculation for **system 1**:

$C_{elite} = 2 + 5 + 0.25 \cdot 5 = 8.25$

Table S14c. The final ranking for this set of 5 systems would be as follows.

Place	System
1st (best)	3
2nd	5
3rd	2
4th	4
5th (worst)	1

S15 – Overlap Matrix for GA Starting Position

Table S15. Percentage overlap matrix of the starting systems of all GA runs.

	R_1	R_2	R_3	R_4	R_5
R_1		0.025	0.0	0.01	0.02
R_2	0.025		0.01	0.01	0.005
R_3	0.0	0.01		0.02	0.005
R_4	0.01	0.01	0.02		0.015
R_5	0.02	0.005	0.005	0.015	

S16 – Loop Similarity Threshold

Each GA is halted when the similarity, S, is greater or equal to 0.95 for 10 consecutive loops. The average number of top performers discovered up to the start of the loop plateaus to 3.8 top performing per run. During the plateau on average only 0.7 new top performers are discovered per loop showing the GA is beginning to slow down its discovery and finally we applied a safety test all the way up to 45 loops which showed and even greater drop off in discovery rate of 0.31 top performers per loop.

Table S16. The drop off in discovery rate after the GA halt point of 10 consecutive runs of S>0.95.

	Loop Similarity % Plateau Start	Total Number of Top Performers Discovered at Start Plateau	Top Performers per Loop Before Plateau	Loop Similarity % Plateau End	Total Number of Top Performers Discovered at During Plateau	Top Performers per Loop During Plateau	Number of Top Performers at run 45	Total Top Performers Discovered per Loop After Plateau
R_1	23	89	3.9	33	8	0.8	2	0.17
R_2	24	91	3.8	34	3	0.3	7	0.64
R_3	24	85	3.5	34	6	0.6	3	0.27
R_4	23	88	3.8	33	10	1.0	3	0.25
R_5	22	84	3.8	32	6	0.6	3	0.23
R_{avg}	23.2	87	3.8	33.2	7	0.7	3.6	0.31

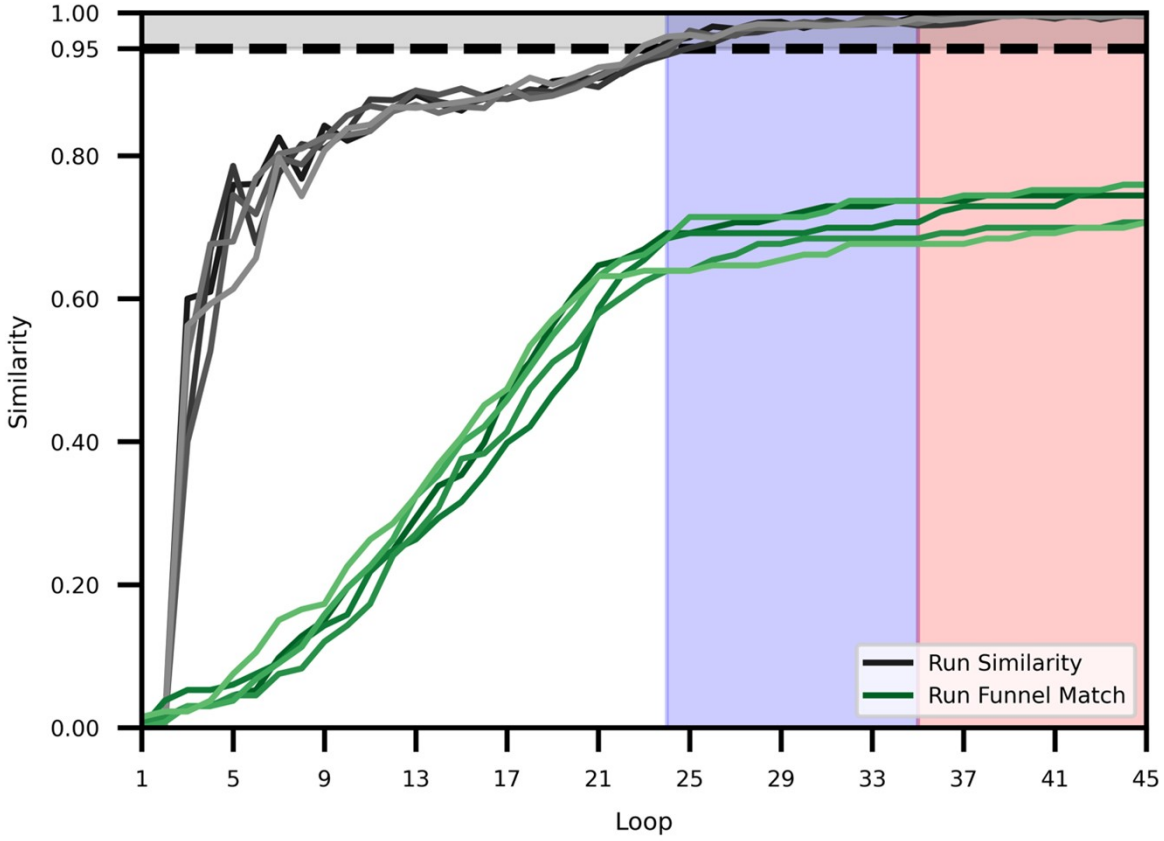


Figure S16. Similarity as a function of the number of loops for 5 independent GA runs. The grey lines indicate the loop *S* for the 5 GA runs, the green lines indicate the *S* match of the systems discovered with a *P*>0.67, E_g<2.5eV and $\lambda^{-/+}$<250meV compared to the total found in the funnel. The black dotted line at 0.95 shows the *S* threshold. The blue box highlights the 10 consecutive loops where *S*>0.95; the red box shows additional loops beyond the S threshold yielding only marginal gains and is the safety test that was carried out to prove the GA has truly halted discovery.

S17 – SAscore Weighting

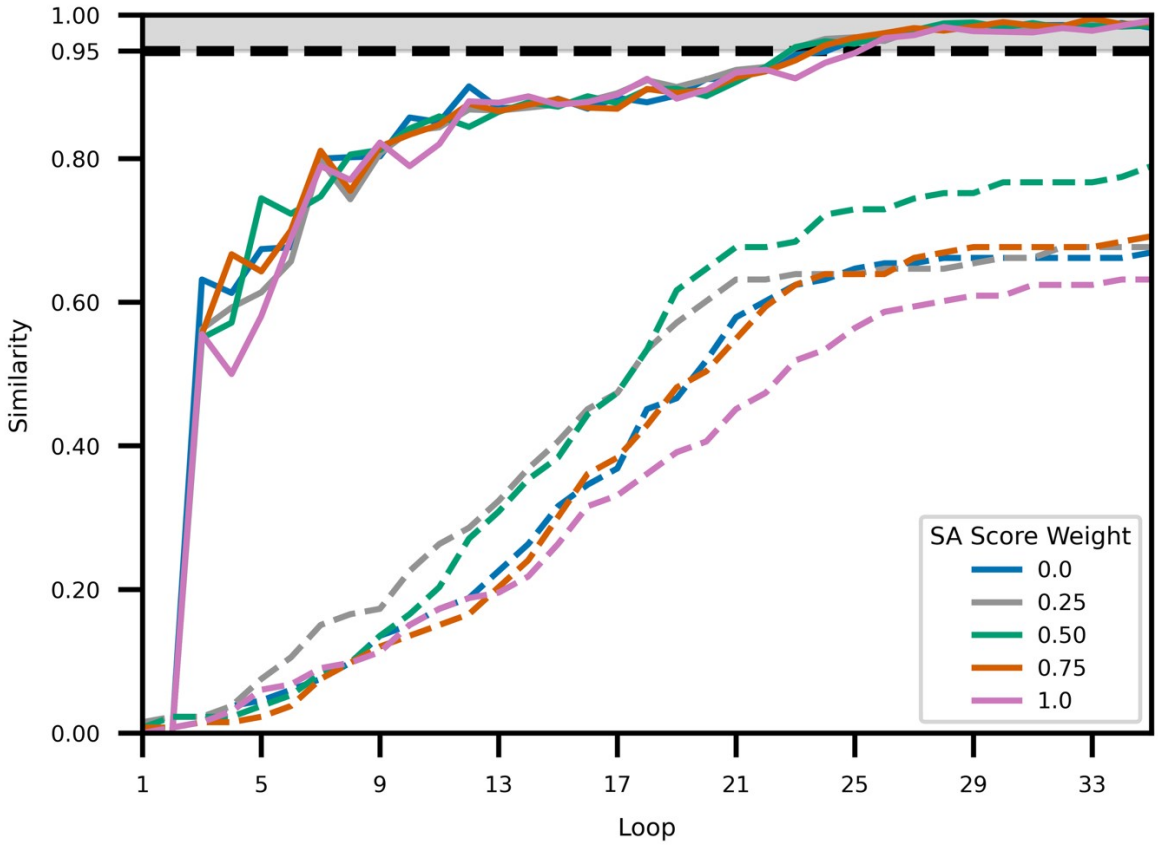


Figure S17. Shows the effect of changing the weight of SA score has on the GA. The solid lines indicate the loop S for the 5 GA runs with different weights. Dashed lines indicate the S match of the systems discovered in GA with a *P*>0.67, E_g<2.5eV and $\lambda^{-/+}$<250meV compared to the total found in the funnel.

S18 – Overlap Matrix of Top performing Systems

Table S18. Percentage overlap matrix of the top-performing systems of all GA runs and the funnel.

	R_1	R_2	R_3	R_4	R_5
R_1		0.907	0.907	0.907	0.876
R_2	0.936		0.904	0.926	0.904
R_3	0.967	0.934		0.923	0.934
R_4	0.898	0.888	0.857		0.867
R_5	0.955	0.955	0.955	0.955	
Funnel	0.729	0.707	0.684	0.737	0.669

S19 – Efficiency of GA vs Funnel

Table S19. Comparing the efficiency of the GA to the funnel approach and how the GA was able to discover top-performing molecules with a reduced number of calculation

	Loop Similarity % Plateau Start	Molecules Studied	% Chemical Space Explored	P, Eg and SA Calculations	λ Molecules Calculated	Total λ<350 meV	Top Performers	P, Eg and SA Efficiency %	<350 meV Efficiency %	Top Performers λ Efficiency %	Total Efficiency %
R_1	23	7161	29	5008	1108	787	89	22.1	71.0	8.0	1.24
R_2	24	6956	28	4765	1037	731	91	21.8	70.5	8.8	1.31
R_3	24	7286	29	5084	1087	760	85	21.4	69.9	7.8	1.17
R_4	23	6995	28	4896	1068	767	88	21.8	71.8	8.2	1.26
R_5	22	6665	27	4628	1006	723	84	21.7	71.9	8.3	1.26
R_{avg}	23.2	7013	28	4876	1061	754	87	21.8	71.0	8.2	1.25
Funnel	-	24990	100	14363	3350	1724	133	23.3	51.5	4.0	0.53

S20 – Unique Molecules only in Funnel

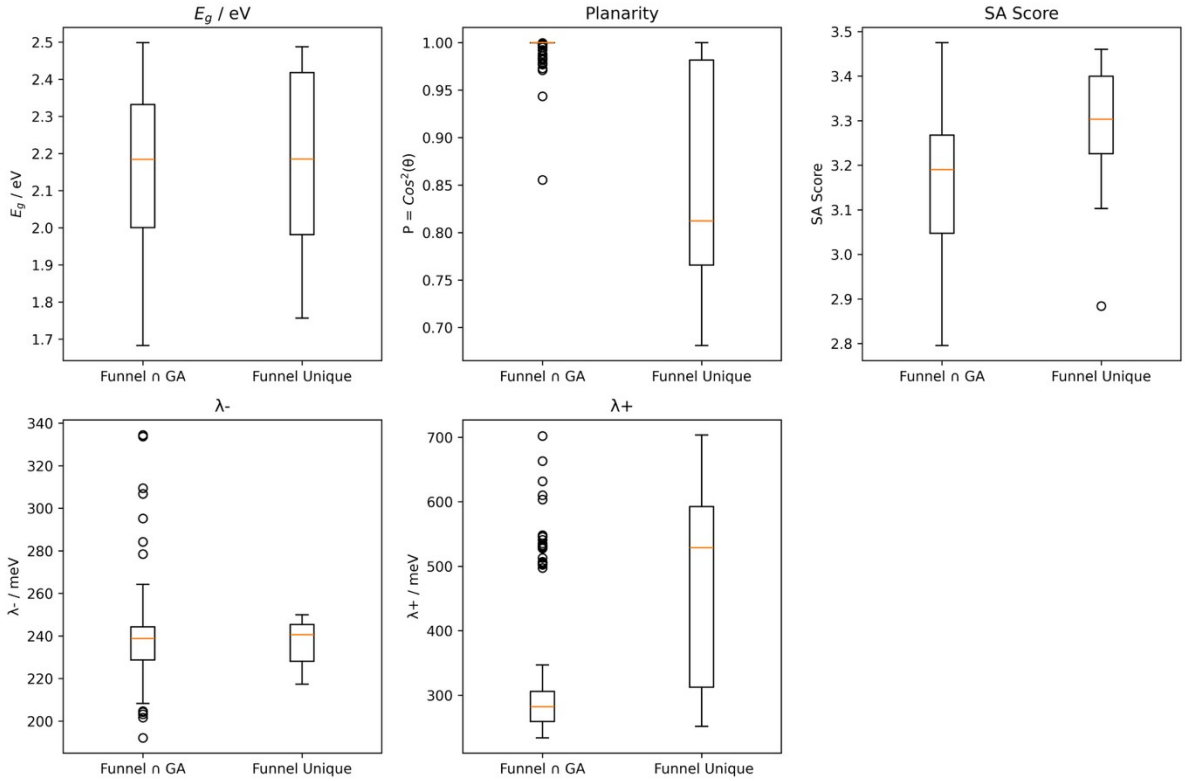


Figure S20. Average, standard deviation and outliers across key parameters between i) the D-L-A systems discovered by both the funnel and the GA, and ii) the 20 unique systems only selected by the funnel.

S21 – Down-selected Molecules

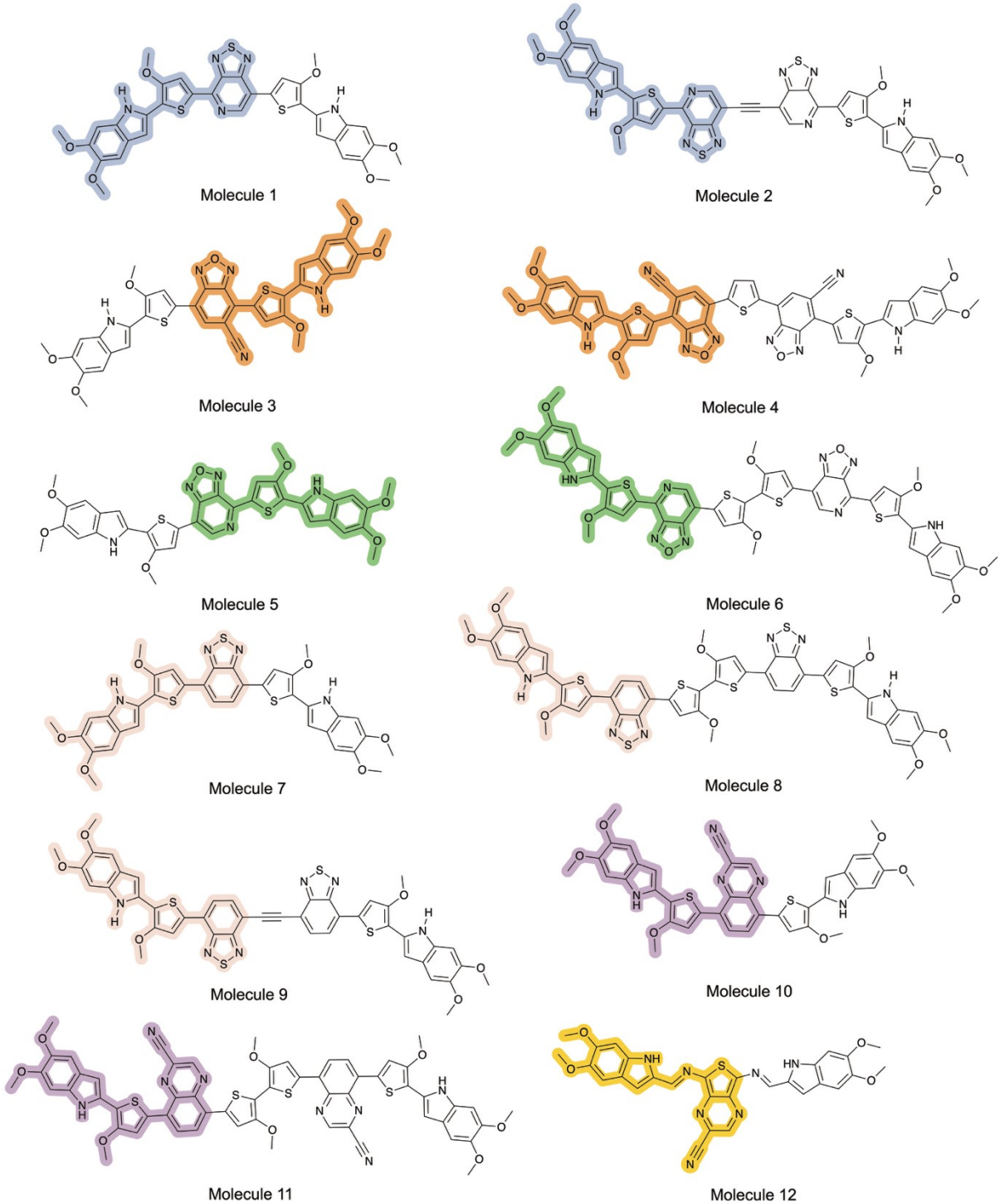
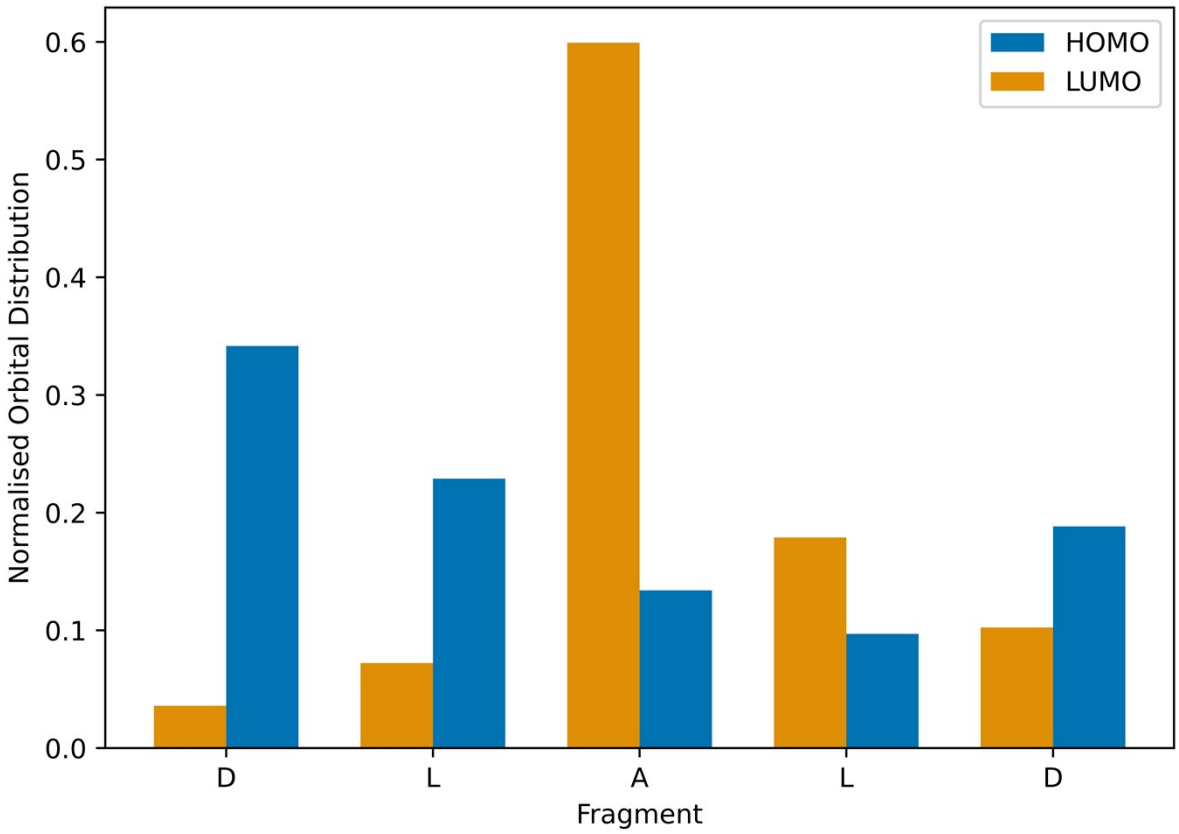


Figure S21. 12 molecules developed from 7 D-L-A systems discovered (highlighted).

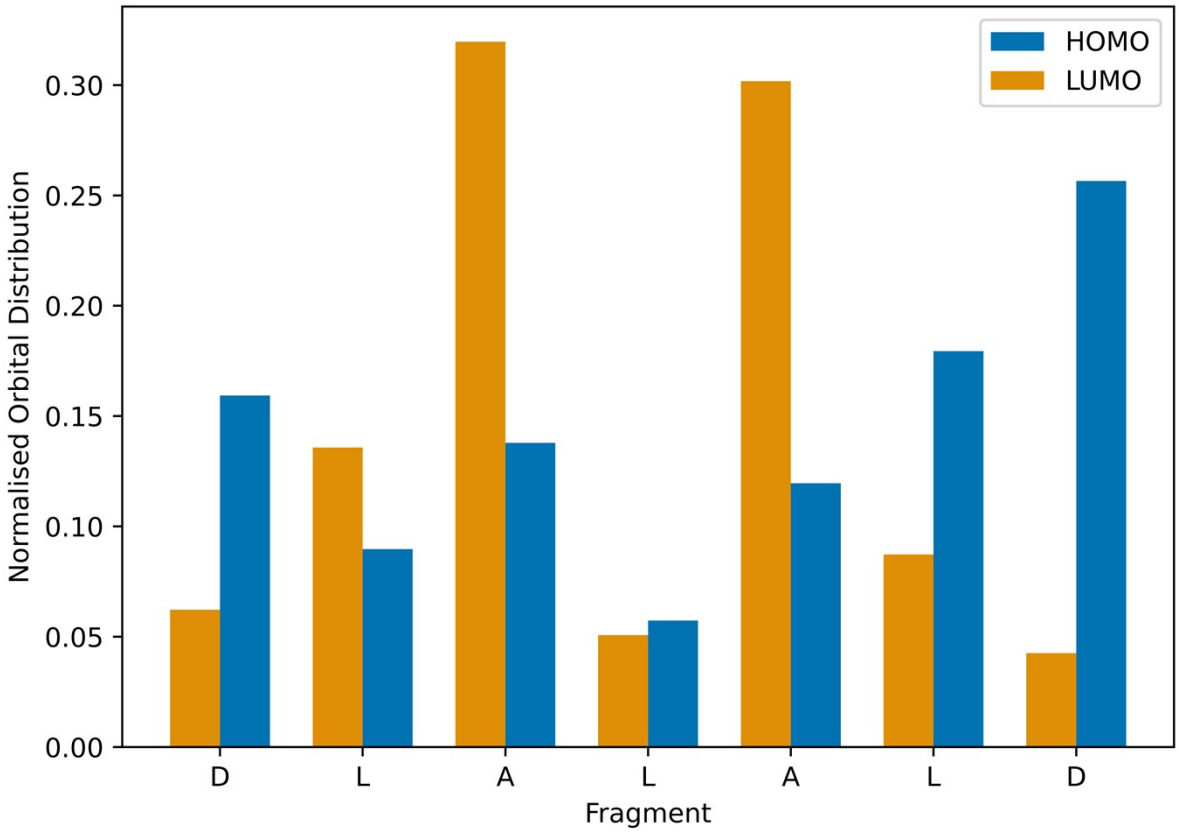
Table S22. The 7 fragments that are used to design the 12 molecules.

Molecule Fragment	E_g (eV)	λ^+ (meV)	λ^- (meV)
1, 2	1.89	288	230
3, 4	2.09	702	228
5, 6	1.98	286	245
7, 8, 9	2.03	299	238
10, 11	1.92	513	248
12	2.22	290	241

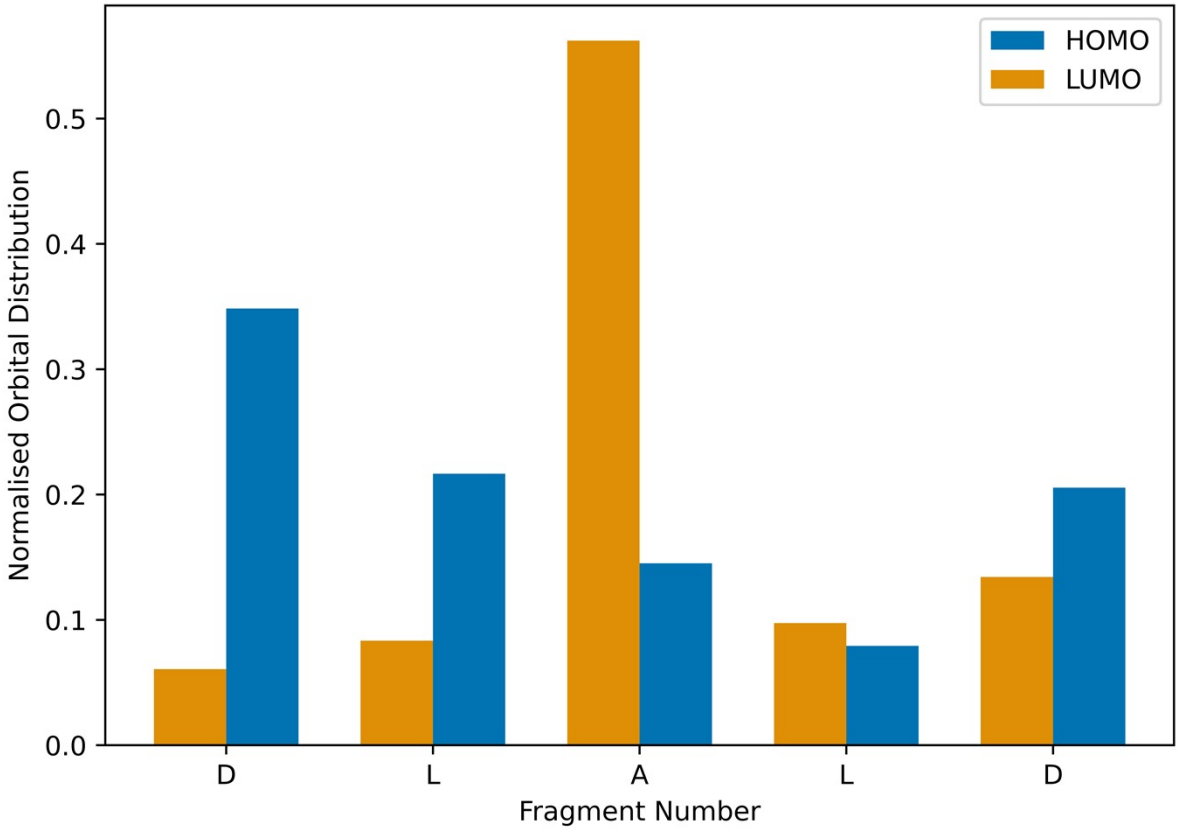
S23 – Orbital Distribution Molecule 1



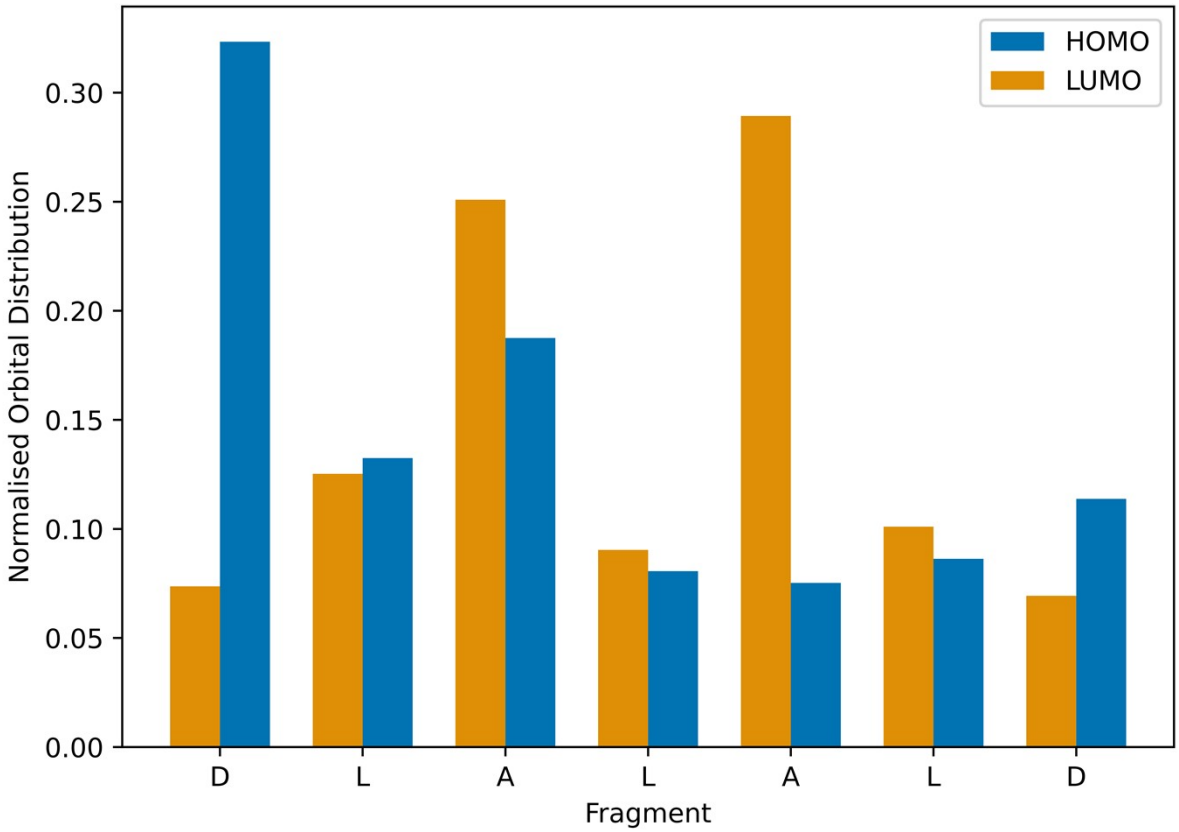
S24 – Orbital Distribution Molecule 2



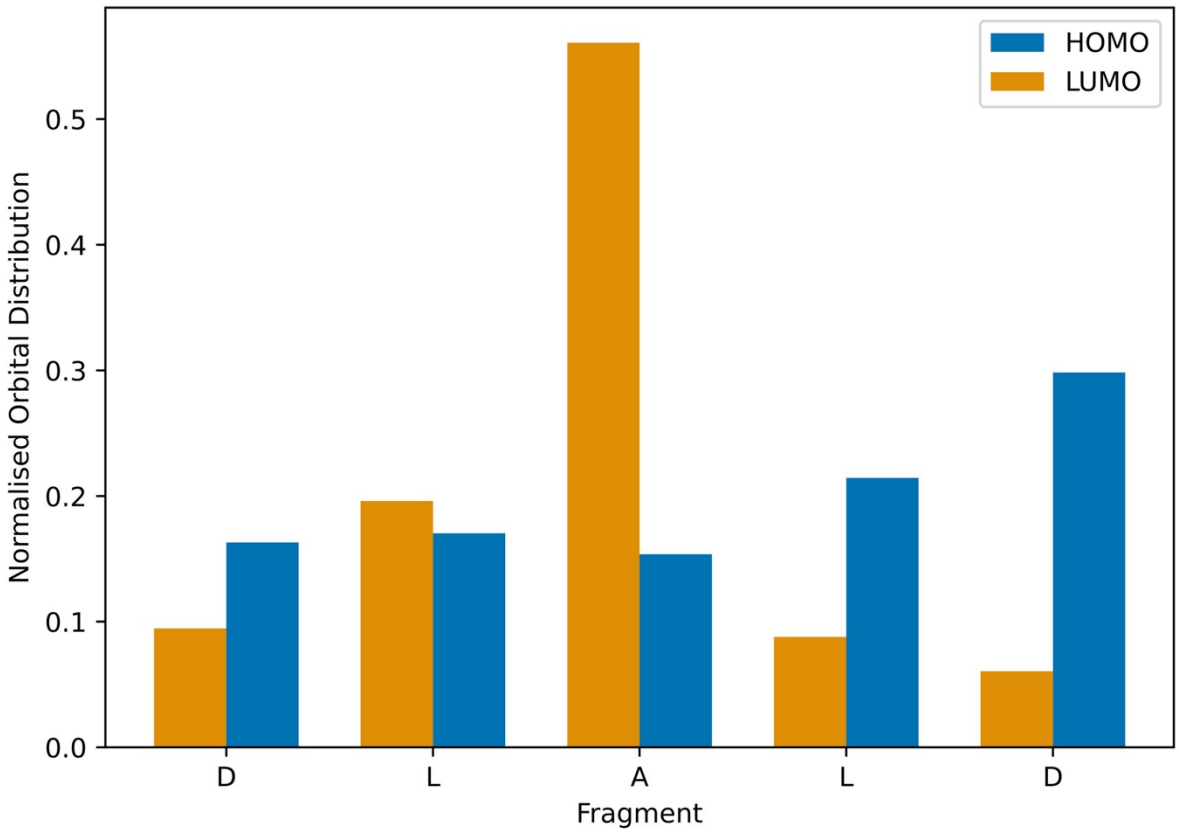
S25 – Orbital Distribution Molecule 3



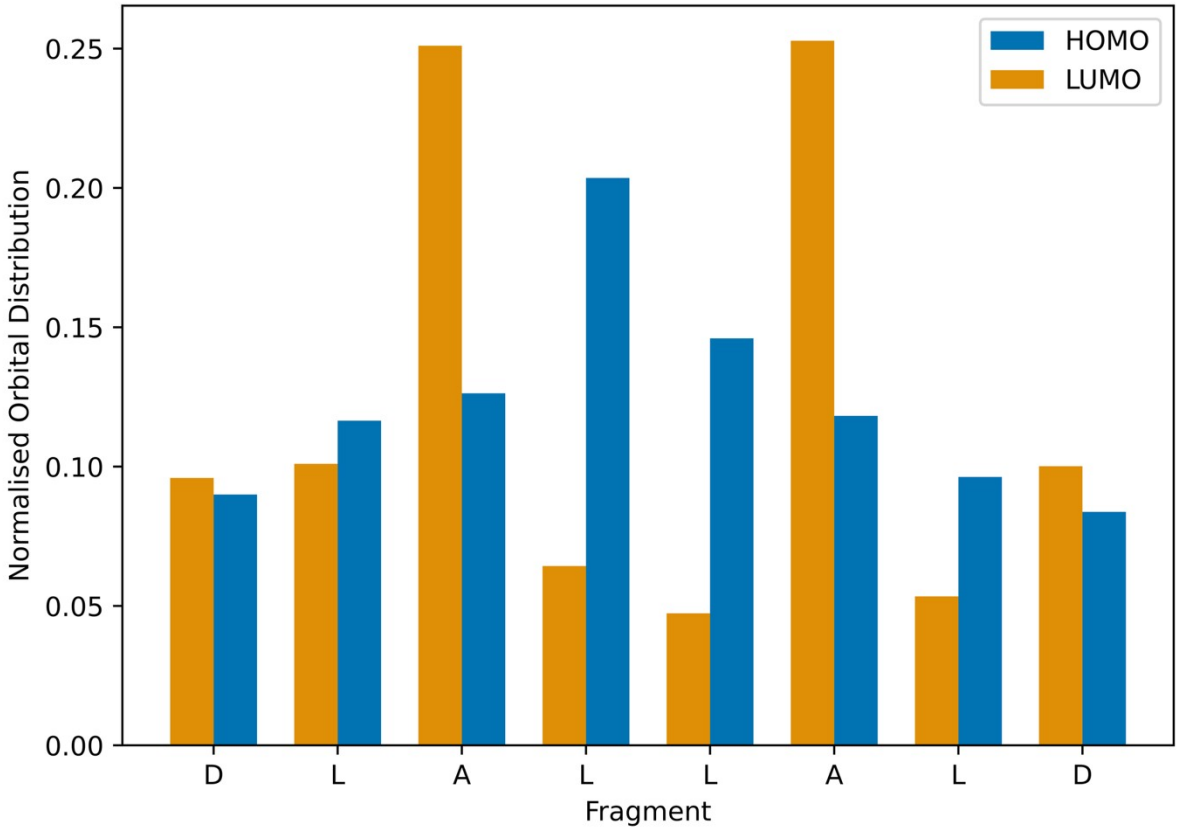
S26 – Orbital Distribution Molecule 4



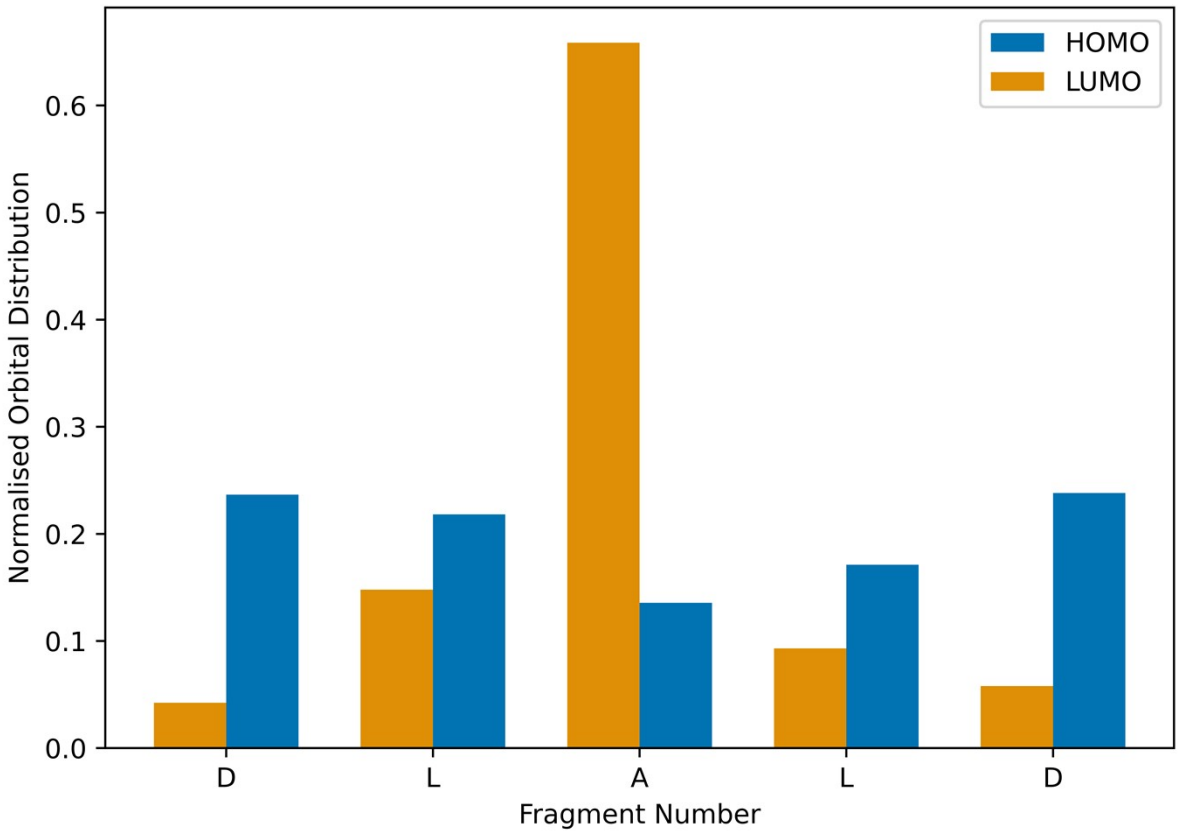
S27 – Orbital Distribution Molecule 5



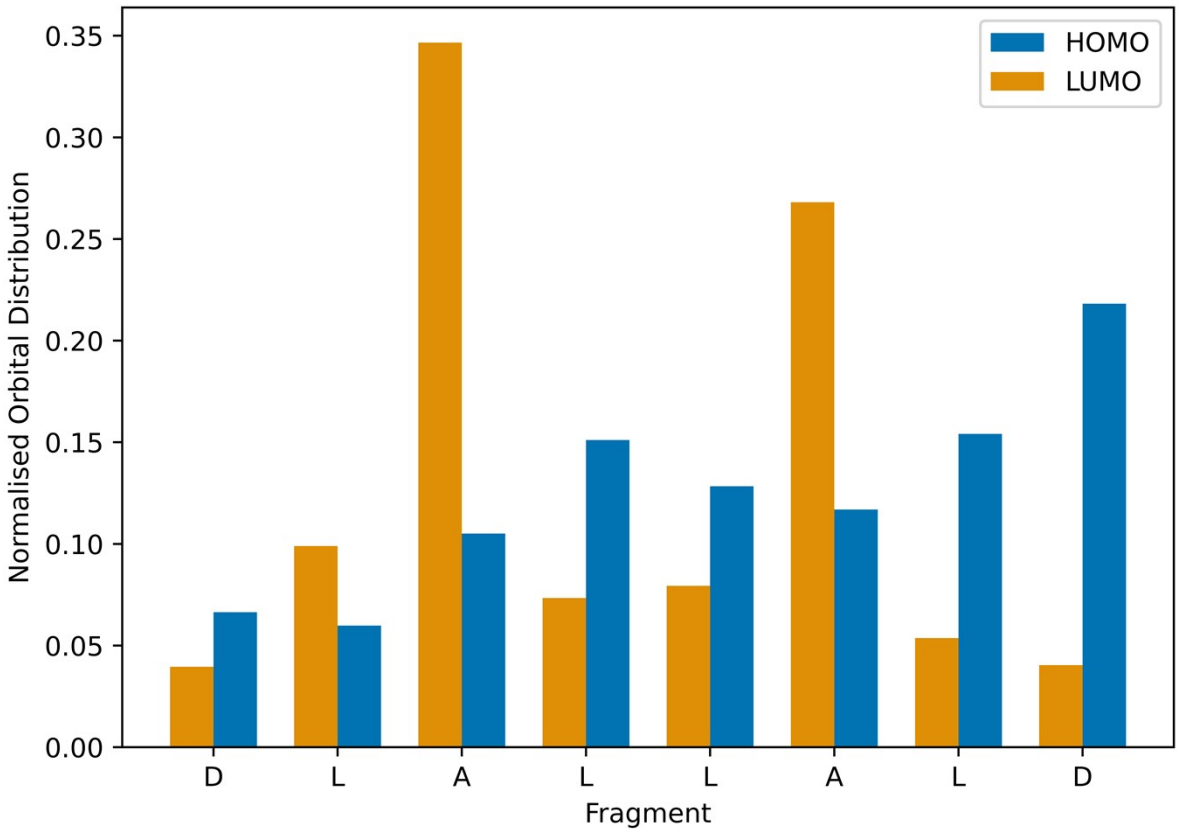
S28 – Orbital Distribution Molecule 6



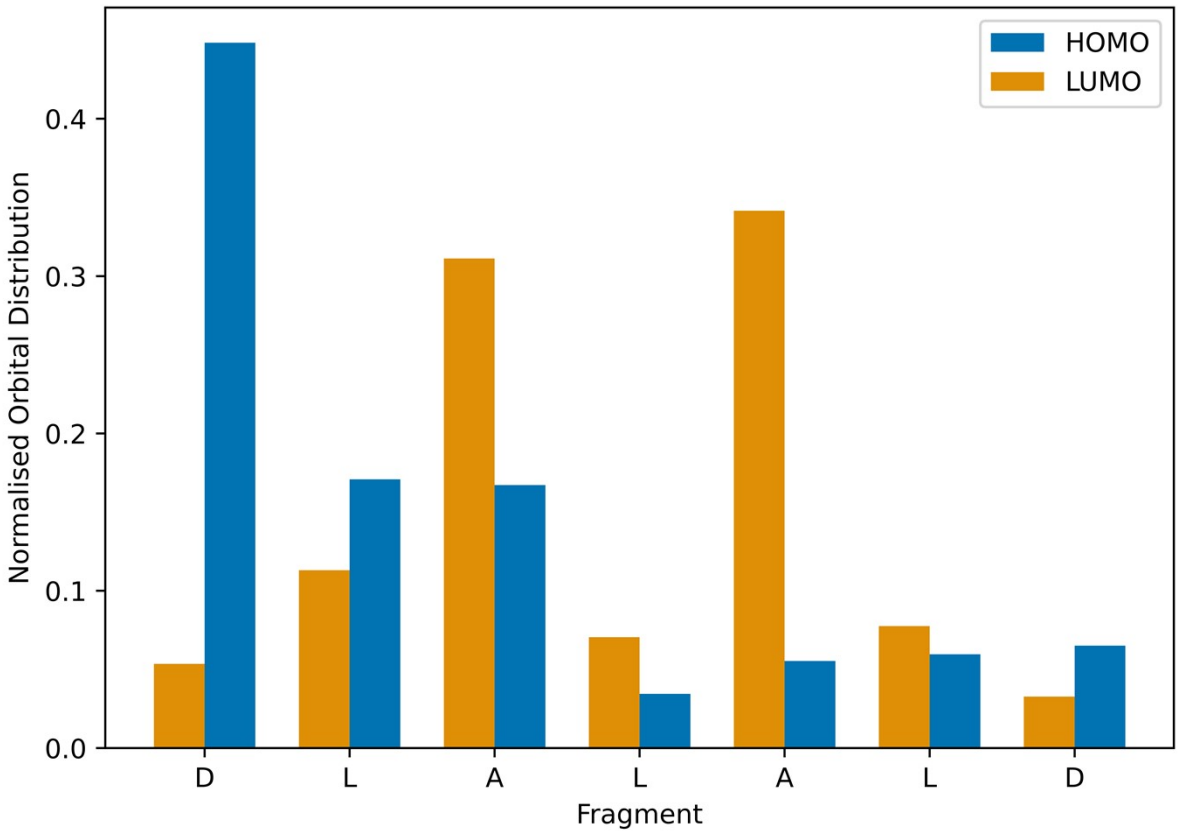
S29 – Orbital Distribution Molecule 7



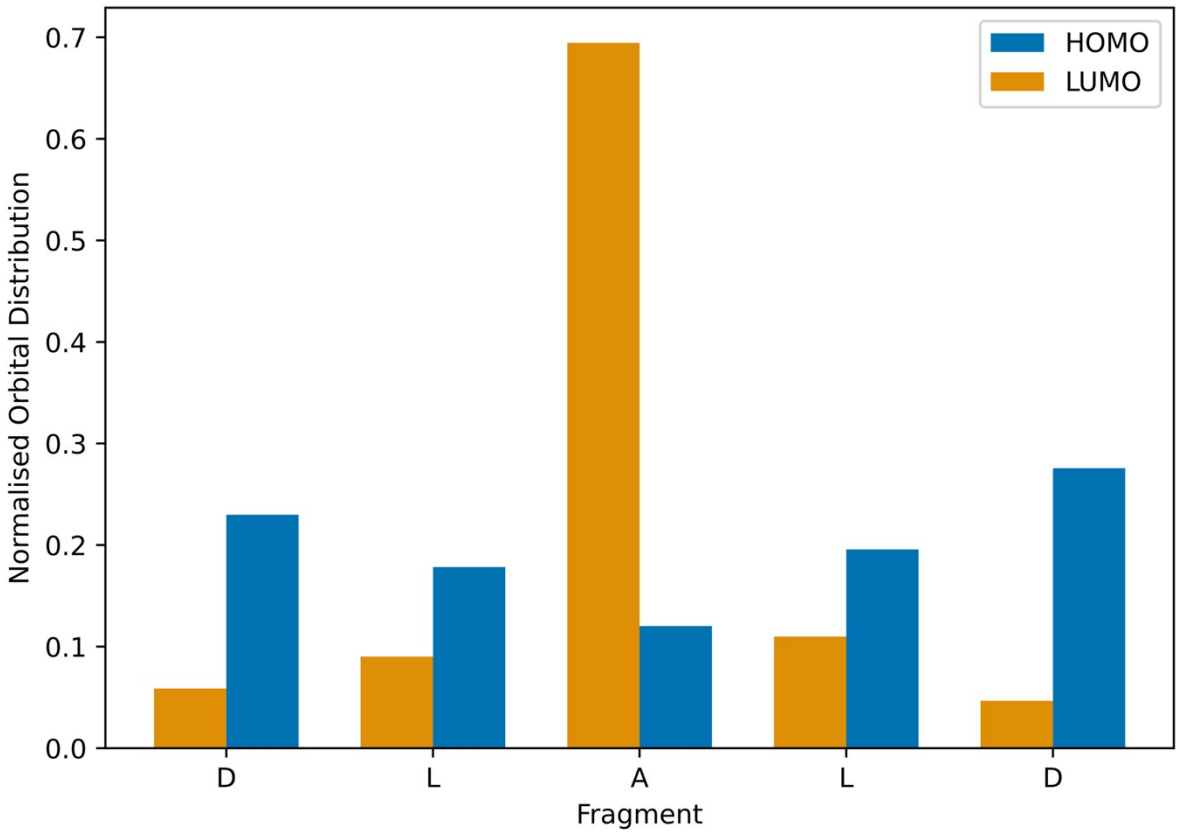
S30 – Orbital Distribution Molecule 8



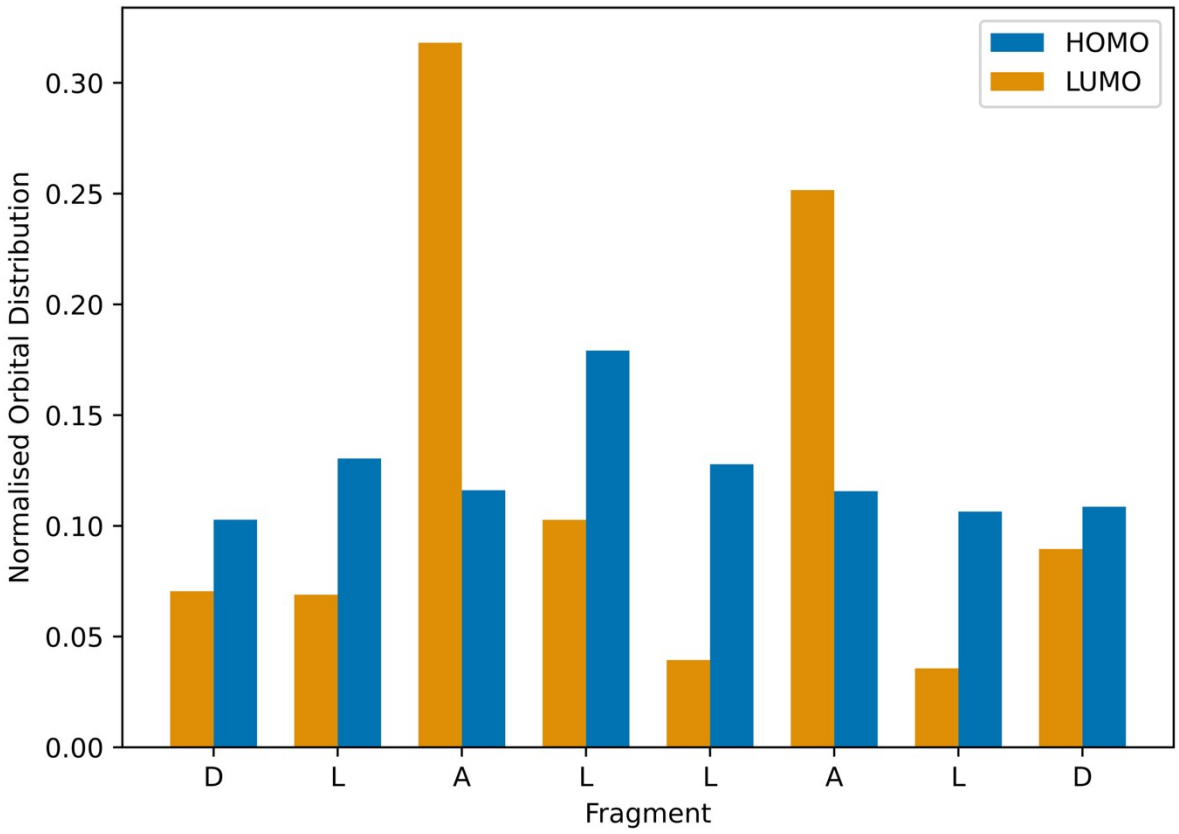
S31 – Orbital Distribution Molecule 9



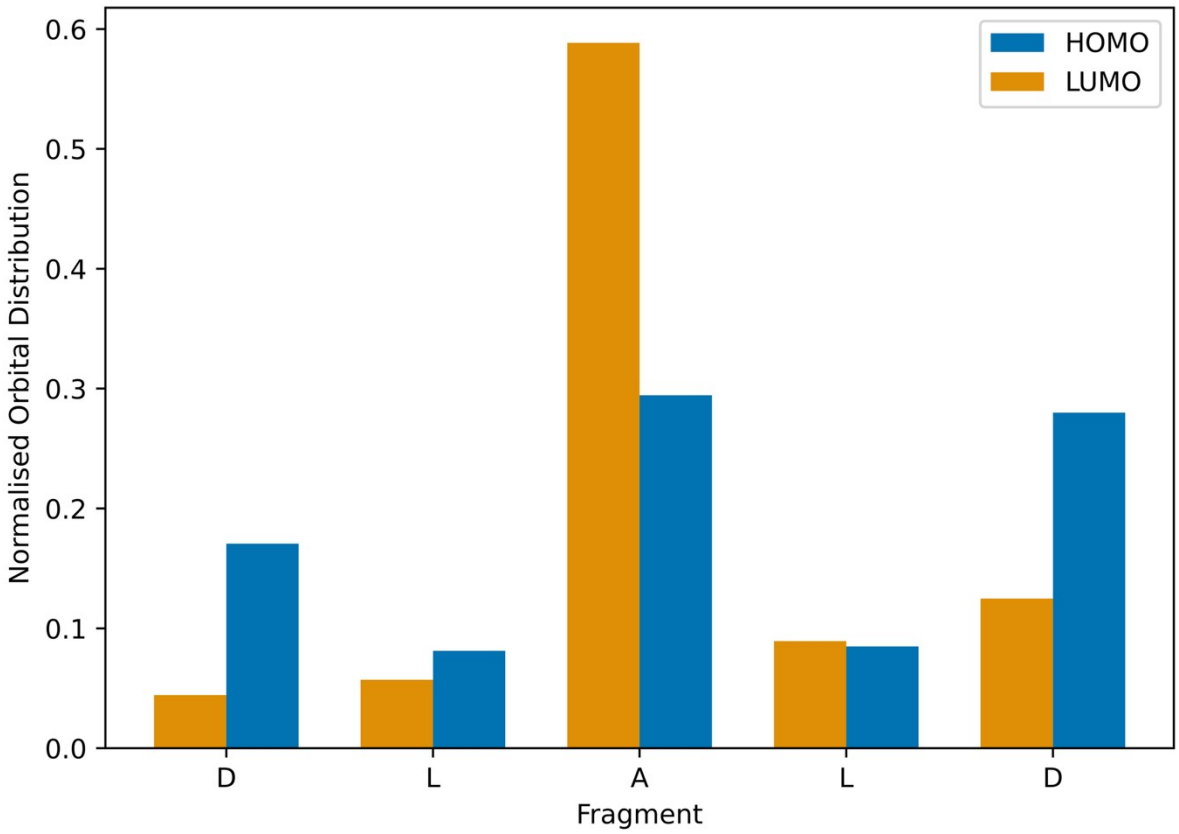
S32 – Orbital Distribution Molecule 10



S33 – Orbital Distribution Molecule 11



S34 – Orbital Distribution Molecule 12



S35 – RDKit Planarity vs Gaussian 16 Geometry Optimised Planarity

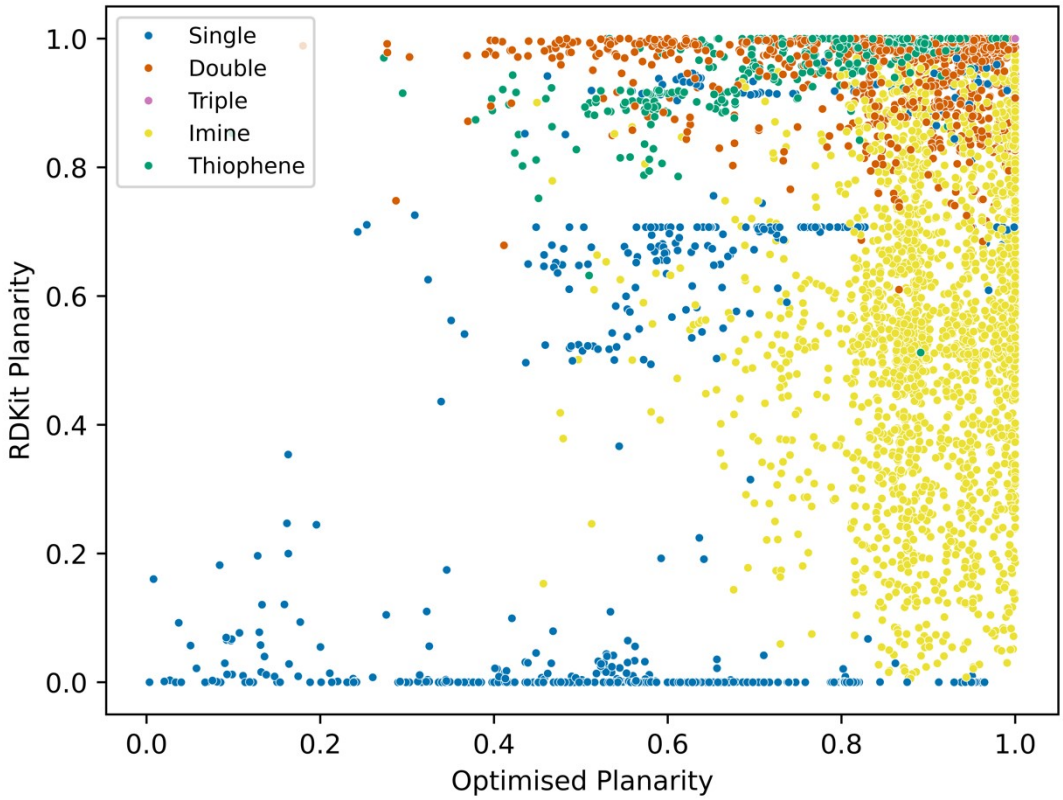


Figure S35. Comparing the planarity of the RDKit predicted geometry before any optimisation to the planarity once the molecule has undergone geometry optimisation at B3LYP/6-31G*. The graph shows the importance of the geometry optimisation step.

S36 – RDKit OptimizeMoleculeConfs Planarity vs Gaussian 16 Geometry Optimised Planarity

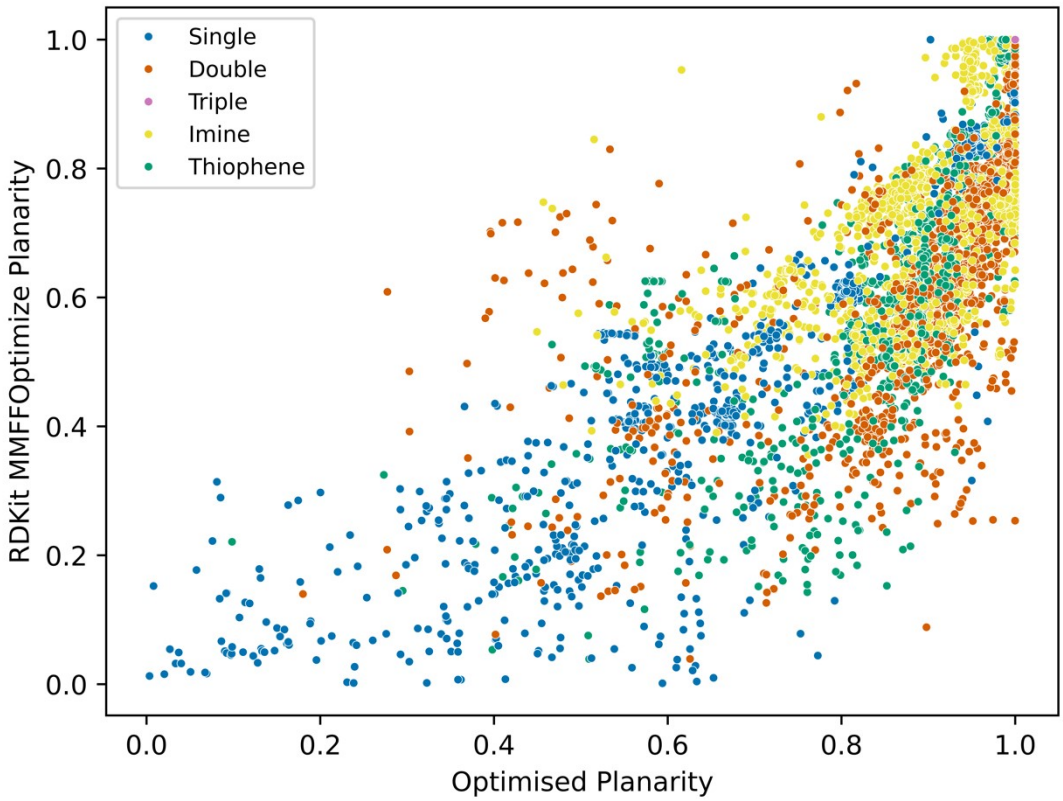


Figure S36. Comparing the planarity of the RDKit predicted geometry when it has undergone conformer searching using AllChem.MMFFOptimizeMoleculeConfs function to the planarity once the molecule has undergone geometry optimisation at B3LYP/6-31G*. The number of conformers studied was 100 per molecule using MMFF94 default.

S37 – RDKit OptimizeMoleculeConfs Planarity vs RDKit Unoptimised Planarity

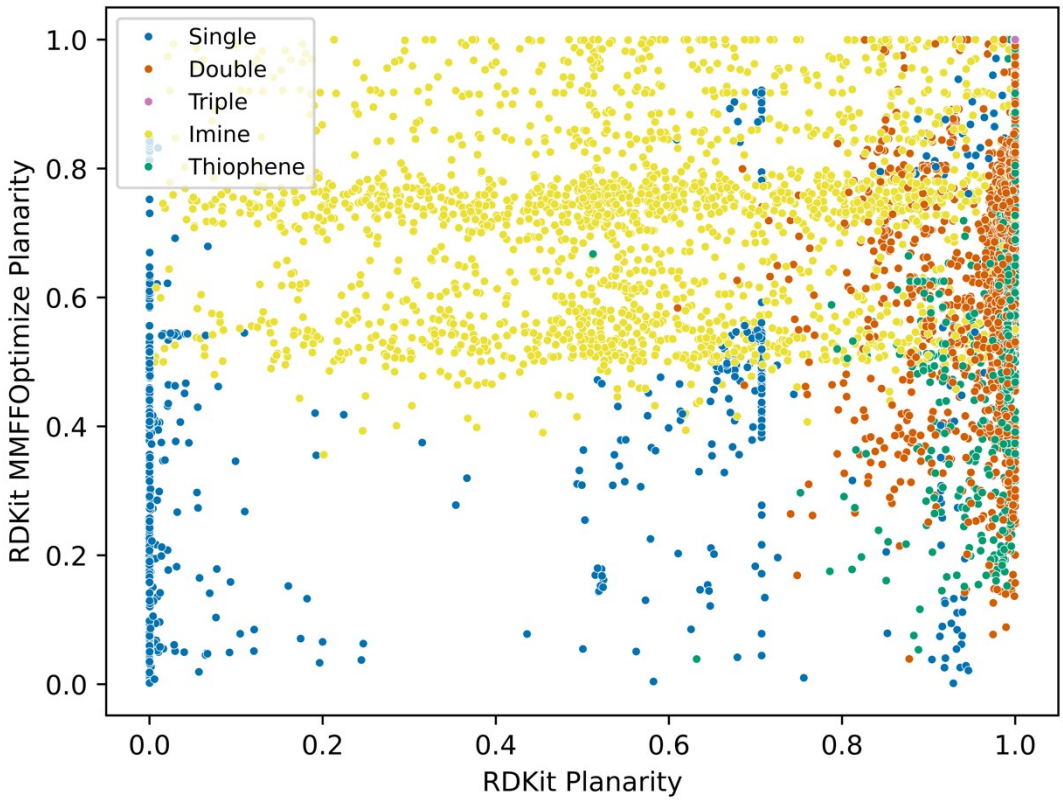


Figure S37. Comparing the planarity of the RDKit predicted geometry when it has undergone conformer searching using AllChem.MMFFOptimizeMoleculeConfs function to the geometry from the initial RDKit guess from SMILES strings.

S38 – Pre-Structure Prediction vs Gaussian 16 Optimisation

Table S38. Comparing the average difference between planarity of structures.

Comparison	Average Difference	Standard Deviation
RDKit to RDKit MMFFOptimize	0.209	0.210
RDKit to G16 Optimised	0.161	0.222
RDKit MMFFOptimize to G16 Optimised	0.154	0.144