The effects of alkyl substitution on the aggregation of π-conjugated dyes: spectroscopic study and modeling

Brunella Bardi,[a] Alix Sournia-Saquet,[b,c] Alain Moreau,[b,c] Kathleen I. Moineau-Chane Ching,[b,c,*] and Francesca Terenziani[a,*]

[a] *Department of Chemistry, Life Sciences and Environmental Sustainability, University of Parma, Parco Area delle Scienze 17/a, 43124 Parma, Italy*

[b] *Laboratoire de Chimie de Coordination (LCC), CNRS, 205 route de Narbonne, FR-31077 Toulouse Cedex 4, France*

[c] *LCC-CNRS, Université de Toulouse, CNRS, 31077 Toulouse, France*

** Corresponding authors: kathleen.chane@lcc-toulouse.fr, francesca.terenziani@unipr.it*

SUPPORTING INFORMATION

Contents

1. Experimental details
   - *1.1 Synthetic procedures*
   - *1.2 Additional data*
   - *1.3 Deconvolution of the emission spectra in THF/water mixtures*
2. DFT and TDDFT calculations
   - *2.1. Computational details*
   - *2.2. DFT and TDDFT data*
3. Essential-state models
   - *3.1. Modelling of the solvated dye*
   - *3.2. Modelling of the aggregates*
4. References

1. Experimental details

All chemicals and reagents were used as received from commercial sources without purification and each reaction step of the syntheses was carried out under an argon atmosphere.

Mass spectrometry was performed by the Mass Spectrometry Core Facility of ICT (University Paul Sabatier, Toulouse). HRMS DCI/CH_4 experiments have been made on a GCT Premier spectrometer (Waters).

^{1}H and ^{13}C NMR spectra were acquired in deuterated solvents with Bruker Avance 300 and Avance 400 spectrometers and the chemical shifts are referenced to the corresponding solvent peaks.

Differential scanning calorimetry, DSC, was performed using a DSC 204 F1 Phoenixs system (NETZSCH). Samples (5-20 mg) were filled in 30 mL perforated aluminium pans and measurements were conducted under nitrogen atmosphere. Three different heating/cooling cycles were registered for each sample in temperature ranges between 0 and 250 °C with a 10 °C min^{-1} rate. With the exception of the first thermogram, all the cycles were reproducible. The values reported correspond to those determined from the position of the top of the peaks. The enthalpy in J/g is calculated from the area under each peak associated with the phase transition.

Spectroscopic characterization was performed on freshly prepared dilute solutions (absorbance < 0.1). Measurements were carried out in different solvents covering a wide polarity range. All solvents were HPLC or spectrophotometric grade and were used as received. UV-vis absorption spectra were collected with a Lambda35 and Lambda650 (Perkin Elmer) double-beam spectrophotometers. Fully corrected emission spectra were recorded with a FLS1000 (Edinburgh Instruments) fluorometer equipped with a 450 W Xenon lamp as the excitation source and a PMT-NIR detector in liquid nitrogen cooled housing. Fluorescence excitation anisotropy was measured in 2-methyltetrahydrofuran (HPLC grade, stored overnight on molecular sieves and filtered before use); the solutions were vitrified after rapid cooling in a OptistatDN (Oxford Instruments) liquid nitrogen cryostat adapted into the FLS1000 instrument.

Spectroscopic characterization in the solid state were performed both on the powders as received (pristine powders), and on the powders ground in a mortar. In the latter case, the specimens were obtained spreading the sample on a quartz plate, forming a thin film. Absorption measurements were performed in transmission, while emission was recorded in front-face configuration.

Voltametric measurements were carried out using a potentiostat Autolab PGSTAT100 controlled by GPES 4.09 software. Experiments were performed at room temperature in a homemade airtight three-electrode cell connected to a vacuum/argon line. The reference electrode consisted of a saturated calomel electrode (SCE) separated from the solution by a bridge compartment. The counter-electrode was a platinum wire of ca. 1 cm^2 apparent surface. The working electrode was a Pt microdisk (0.5 mm diameter) or a glassy carbon (1 mm diameter). The supporting electrolyte (nBu_4N)[PF6] (Fluka, 99% electrochemical grade) was used as received and simply degassed under argon. Dichloromethane was freshly distilled over CaH_2 prior to use. The solutions used during the electrochemical studies were typically 10^{-3} mol L^{-1} in compound and 0.1 mol L^{-1} in the supporting electrolyte. Before each measurement, the solutions were degassed by bubbling Ar and the working electrode was polished with a polishing machine (Presi P230).

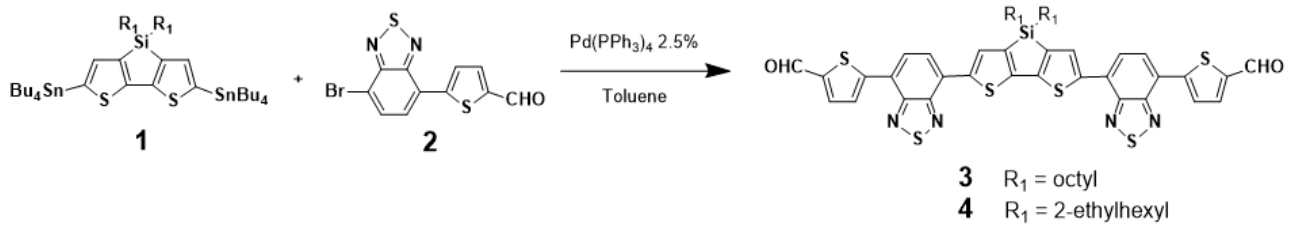
Aggregation studies were performed in mixtures of tetrahydrofuran (THF) and ultrapure milli-Q water of different composition. For each dye, 40 µL of a concentrated THF stock solution (2.5×10^{-4} mol L^{-1}) were added to 5.0 mL of THF/water mixture of given composition, achieving a nominal 2.0×10^{-6} mol L^{-1} concentration of the dye in the final mixture (absorbance < 0.1). The percentage (in volume) of water in the mixtures was varied from 0 (stock solution diluted in pure THF) to 100% (stock solution added to 5.0 mL of

pure water). Samples were stored in the dark and characterized within a working day. Suspensions (when formed) were stable for a few days under working conditions. Dynamic light scattering experiments were performed with a Malvern Zetasizer Nano ZSP instrument on freshly prepared samples.

Fluorescence quantum yields of solutions and THF/water mixtures were estimated using cresyl violet in ethanol (ϕ = 0.51, λ_{exc} = 550 nm) and 1,1',3,3,3',3'-hexa-methyl-indotricarbocyanine iodide (HITCI) in ethanol (ϕ = 0.283, λ_{exc} = 680 nm) as reference standards, respectively.

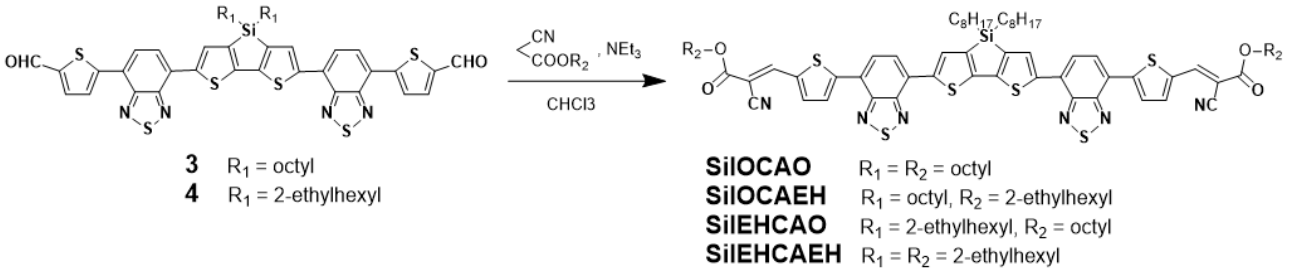
Two-photon absorption spectra were collected exploiting two-photon excited fluorescence (2PEF).[1,2] Solutions of the dyes in toluene (absorbance ≈ 0.1), contained in a standard 1 cm quartz cell, were excited by a fs-pulsed laser (Chameleon Discovery, Coherent) through a 25× water-dipping objective; emission was collected in epifluorescence mode and detected by a two-photon microscope (A1R MP+ Upright, Nikon) with a spectral GaAsP detector.

1.1. Synthetic procedures



Scheme S1. Synthetic route for building blocks **3** and **4**

A mixture of **1** (613 mg, 0.615 mmol), **2** (400 mg, 1.23 mmol), and $Pd(PPh_3)_4$ (35 mg, 0.03 mmol) in dry toluene (20 mL) was stirred and heated at 110 °C under argon for 12 h. After the reaction mixture was cooled to room temperature, the solvent was removed by rotary evaporation. Ethyl acetate was added to the resulting sticky mixture and then sonicated to generate the product, which was collected by filtration and rinsed 3 times with ethyl acetate, to afford the desired product as a black powder. After dry under vacuum, 400 mg of **3** (72%) and 325 of **4** (59%), respectively, are collected.; molecule **3** : ^{1}H NMR (300 MHz, $CDCl_3$) δ 10.00 (s, 2H), 8.26 (s, 2H), 8.24 (d, *J = 4.1 Hz*, 2H), 8.02 (d, *J = 7.7 Hz*, 2H), 7.94 (d, *J = 7.7 Hz*, 2H), 7.87 (d, *J = 4.1 Hz*, 2H), 1.55-1.45 (m, 4H), 1.44-1.35 (m, 4H), 1.33-1.19 (m, 16H), 1.15-1.02 (m, 4H), 0.92-0.78 (m, 6H); molecule **4** : ^{1}H NMR (400 MHz, $CDCl_3$) δ 10.00 (s, 2H), 8.29 (t, *J = 4.2 Hz*, 2H), 8.23 (d, *J = 4.0 Hz*, 2H), 8.02 (d, *J = 7.7 Hz*, 2H), 7.92 (d, *J = 7.7 Hz*, 2H), 7.87 (d, *J = 4.0 Hz*, 2H), 1.40-1.10 (m, 22H), 0.94-0.75 (m, 12H).



Scheme S2. Synthetic route for target compounds

A mixture of **3** (resp. **4**) (200 mg, 0.221 mmol), triethylamine (307 µL, 2.2 mmol), octylcyanoacetate (resp. ethylhexylcyanoacatate) (457 µL, 2.2 mmol) in chloroform (20 mL) was stirred for 24h at room temperature.

Then 20 mL of water were added, and the crude product was extracted twice with 30 mL of dichloromethane. After drying over magnesium sulphate and evaporation of the organic phases, the resulting black solid is dried under vacuum before recrystallisation in AcOEt-pentane.

SilOCAO, 205 mg (74%): ^{1}H NMR (10.8 mg/mL, 400 MHz, CD_2Cl_2) δ 8.16 (s, $2H_1$), 8.05 (s, $2H_6$), 7.95 (d, *J = 4.2 Hz*, $2H_3$), 7.70 (d, *J = 7.7 Hz*, $2H_5$), 7.69 (d, *J = 4.2 Hz*, $2H_2$), 7.54 (d, *J = 7.7 Hz*, $2H_4$), 4.25 (t, *J = 6.7 Hz*, $4H_{\text{-COO-CH2-}}$), 1.75 (m, $4H_{\text{-COO-C-CH2-}}$), 1.66-1.00 (m, $48H_{alkyl}$), 1.00-0.75 (m, $12H_{\text{-C-CH3}}$). ^{13}C NMR (101 MHz, CD_2Cl_2) δ 162.71, 151.95, 151.86, 151.71, 148.59, 145.67, 144.26, 140.98, 138.29, 136.09, 131.03, 127.77, 127.50, 127.05, 123.66, 122.65, 115.97, 98.20, 66.54, 33.40, 31.95, 31.81, 29.34, 29.32, 29.23, 29.20, 28.56, 25.83, 24.35, 22.71, 22.66, 13.89, 13.87, 11.83. UV-vis (CH_2Cl_2), λ_{max} (ε_{max})= 605 nm (84850 L.mol^{-1}.cm^{-1}). HRMS DCI/CH_4 (m/z, [M]+) Calcd for $C_{68}H_{80}N_6O_4S_6Si$ 1264.44, found 1264.4336.

SilOCAEH, 180 mg (65%): ^{1}H NMR (10.8 mg/mL, 400 MHz, CD_2Cl_2) δ 8.17 (s, 2 H_1), 8.05 (s, 2 H_6), 7.95 (d, *J = 4.2 Hz*, 2 H_3), 7.70 (d, *J = 7.7 Hz*, 2 H_5), 7.69 (d, *J = 4.2 Hz*, 2 H_2), 7.57 (d, *J = 7.7 Hz*, 2 H_4), 4.24-4.12 (m, $4H_{\text{-COO-CH2-}}$), 1.77-1.67 (m, $2H_{\text{-COO-C-CH-}}$), 1.65-1.00 (m, $44H_{alkyl}$), 1.00-0.70 (m, $18H_{\text{-C-CH3}}$). ^{13}C NMR (101 MHz, CD_2Cl_2) δ 162.78, 151.95, 151.84, 151.71, 148.56, 145.60, 144.28, 141.01, 138.28, 136.08, 131.06, 127.76, 127.54, 127.06, 123.66, 122.65, 115.92, 98.31, 68.74, 38.82, 33.41, 31.95, 30.31, 29.34, 29.32, 28.92, 24.35, 23.76, 22.99, 22.71, 13.89, 13.84, 11.84, 10.79. UV-vis (CH_2Cl_2), λ_{max} (ε_{max})= 603 nm (83540 L.mol^{-1}.cm^{-1}). HRMS DCI/CH_4 (m/z, [M]+) Calcd for $C_{68}H_{80}N_6O_4S_6Si$ 1264.44, found 1264.4426.

SilEHCAO, 270 mg (98%): ^{1}H NMR (11.8 mg/mL, 400 MHz, CD_2Cl_2) δ 8.14 (s, 2 H_1), 8.12 (m, 2 H_6), 7.94-7.89 (m, 2 H_3), 7.69-7.65 (m, 2 H_2), 7.61 (d, *J = 7.9 Hz*, 2 H_5), 7.47-7.42 (m, 2 H_4), 4.23 (t, *J = 6.8 Hz*, $4H_{\text{-COO-CH2-}}$), 1.79-1.69 (m, $4H_{\text{-COO-C-CH2-}}$), 1.69-1.08 (m, $42H_{alkyl}$), 1.00-0.82 (m, $18H_{\text{-C-CH3}}$). ^{13}C NMR (101 MHz, CD_2Cl_2) δ 162.68, 151.86, 151.64, 148.58, 145.61, 145.24, 141.00, 138.24, 136.05, 131.36, 127.74, 127.57, 126.91, 123.43, 122.54, 115.94, 98.17, 66.53, 36.09, 35.87, 31.82, 29.23, 29.21, 29.06, 29.00, 28.55, 25.82, 23.18, 22.67, 17.70, 14.11, 13.87, 10.72. UV-vis (CH_2Cl_2), λ_{max} (ε_{max})= 608 nm (83980 L.mol^{-1}.cm^{-1}). HRMS DCI/CH_4 (m/z, [M]+) Calcd for $C_{68}H_{80}N_6O_4S_6Si$ 1264.44, found 1264.4408.

SilEHCAEH, 215 mg (77%): ^{1}H NMR. (11.9 mg/mL, 400 MHz, CD_2Cl_2) δ 8.21 (s, 2 H_1), 8.20-8.17 (m, 2 H_6), 8.03-7.99 (m, 2 H_3), 7.74 (d, *J = 7.7 Hz*, 2 H_5), 7.76-7.71 (m, 2 H_2), 7.60-7.55 (m, 2 H_4), 4.24-4.12 (m, $4H_{\text{-COO-CH2-}}$), 1.78-1.68 (m, $2H_{\text{-COO-C-CH-}}$), 1.67-1.08 (m, $38H_{alkyl}$), 1.1-0.77 (m, 24 $H_{\text{-C-CH3}}$). ^{13}C NMR (101 MHz, CD_2Cl_2) δ 162.80, 152.02, 151.80, 151.55, 148.63, 145.65, 145.40, 141.04, 138.31, 136.13, 131.53, 127.89, 127.70, 127.11, 123.69, 122.76, 115.94, 98.40, 68.74, 38.82, 36.09, 35.84, 30.31, 29.04, 29.00, 28.92, 23.76, 23.14, 22.98, 17.70, 14.06, 13.83, 10.78, 10.70. UV-vis (CH_2Cl_2), λ_{max} (ε_{max})= 608 nm (83040 L.mol^{-1}.cm^{-1}). HRMS DCI/CH_4 (m/z, [M]+) Calcd for $C_{68}H_{80}N_6O_4S_6Si$ 1264.44, found 1264.4369.

Protons numbers refer to the labelled hydrogens shown on the pi-conjugated molecular fragment in Figure 1 of the main text.

Note that the ^{1}H NMR signals of protons situated in the vicinity of the branch point of the ethylhexyl chain appear as a multiplets instead of singulet or triplet due to the racemic character of ethylhexyl fragment: see for example comparison between the $4H_{\text{-COO-CH2-}}$ signals of **SilOCAO** and **SilOCAEH** and between the H6 signals of **SilOCAO** and **SilEHCAO**.

1.2. Additional data

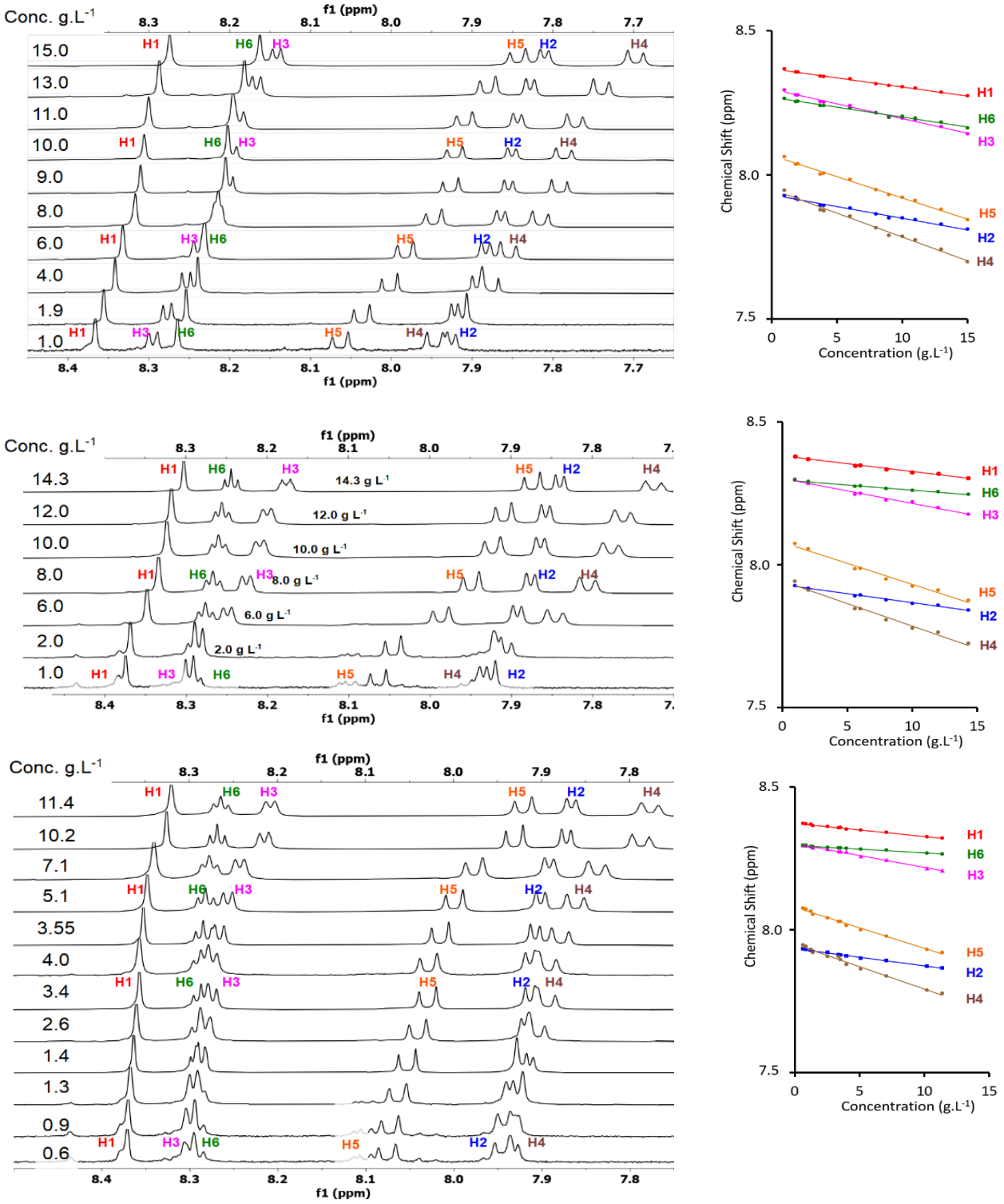


Figure S1. Left: aromatic part of ^{1}H-NMR spectra of **SilOCAEH**, **SilEHCAO** and **SilEHCAEH** at different concentrations in $CDCl_3$. Right: chemical shift versus concentration. Protons numbers refer to the labelled hydrogens shown on Figure 1.

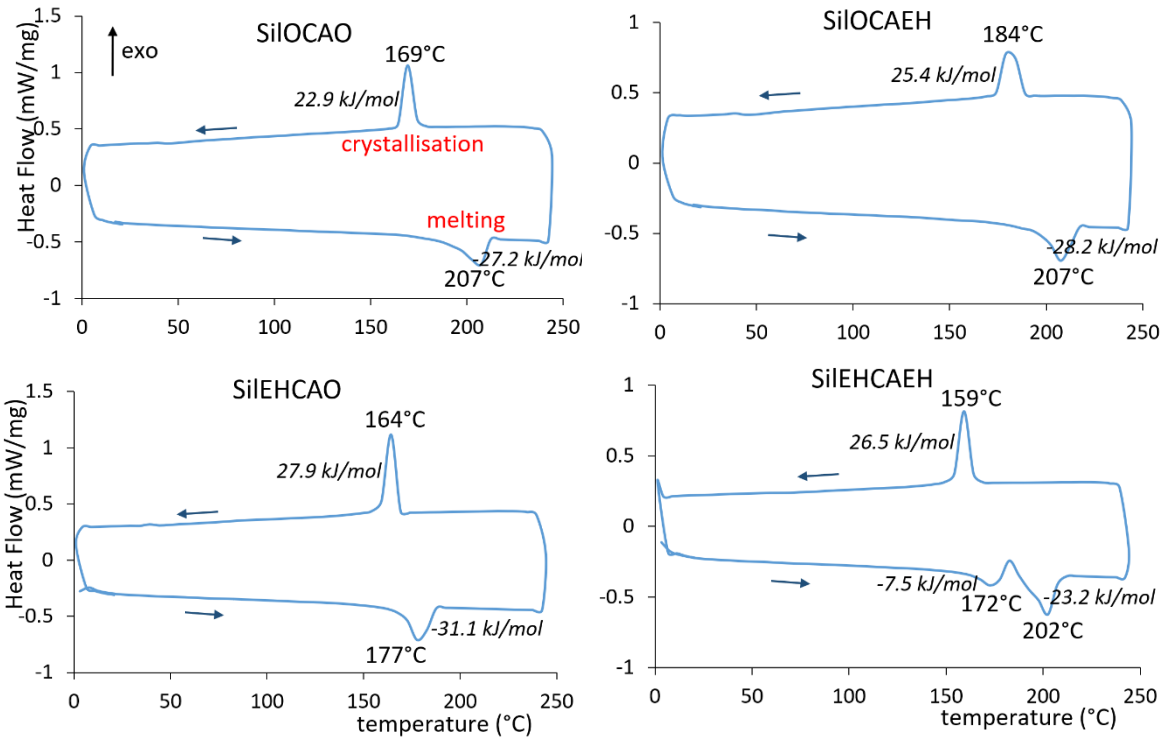


Figure S2. DSC thermograms of the four molecules showing the second heating and cooling cycle at 10°C/min under nitrogen flow. Enthalpies of the transitions are indicated in italic.

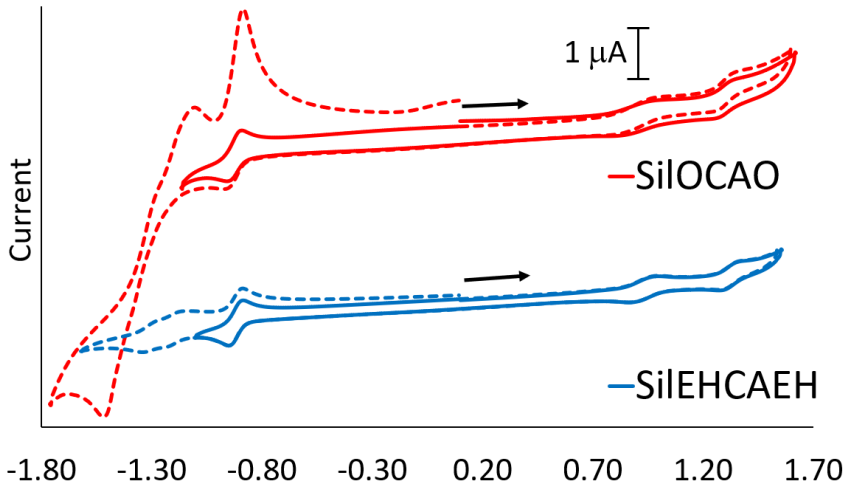


Figure S3. Cyclic voltammograms recorded in CH_2Cl_2 + tBu_4NPF_6 0.1 $molL^{-1}$ + SilOCAO 1.01 10^{-3} $molL^{-1}$ (red) and SilEHCAEH 9.65 10^{-4} $molL^{-1}$ (blue) on a platinum electrode (diameter 0.5 mm) at 0.2 V s^{-1}. Solid lines: potential range between -1.15 and +1.6 V/SCE; dotted lines: potential range between -1.75 (respect. -1.60) and +2.0 V/SCE for SilOCAO (respect. SilEHCAEH). The arrows indicate the direction of the potential scan.

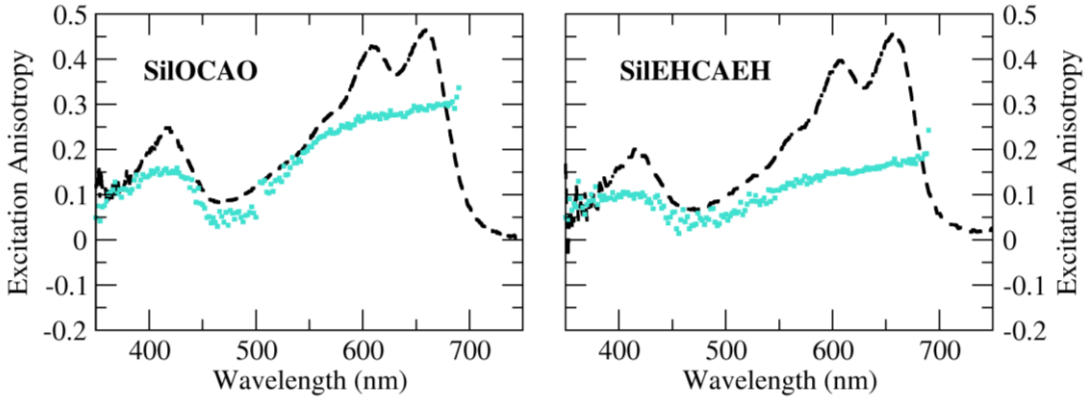


Figure S4. Fluorescence excitation anisotropy (squares) of **SilOCAO** and **SilEHCAEH** in vitrified 2-methyltetrahydrufuran at 77 K. Excitation spectra collected under the same experimental conditions (dashed lines) are reported as a guide to the eye.

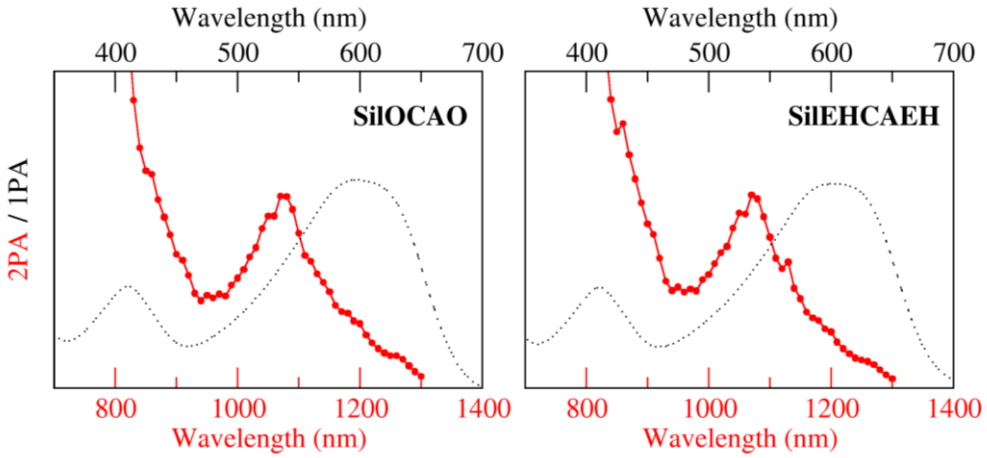


Figure S5. Normalized two-photon absorption (2PA) spectra (red dots) and linear absorption spectra (1PA, dotted lines) of **SilOCAO** and **SilEHCAEH** in toluene.

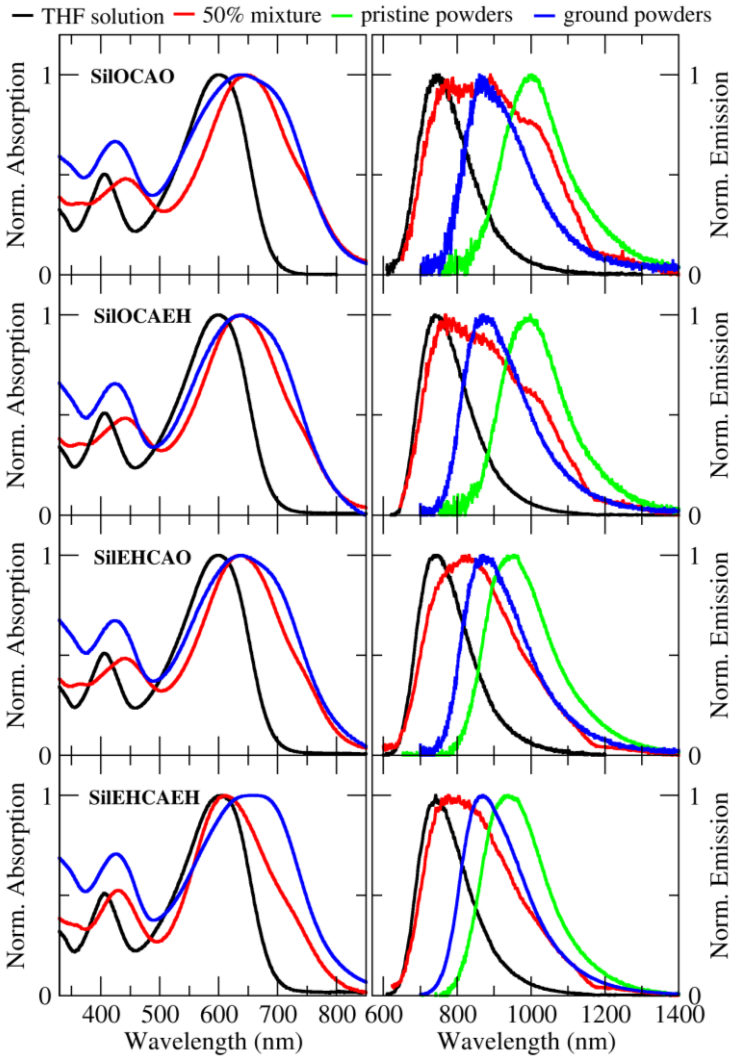


Figure S6. Comparison between the normalized absorption (left) and emission (right) spectra of the chromophores in different aggregation states: THF solution, 50:50 THF/water mixture (excitation on the band maximum), pristine powders and ground powders.

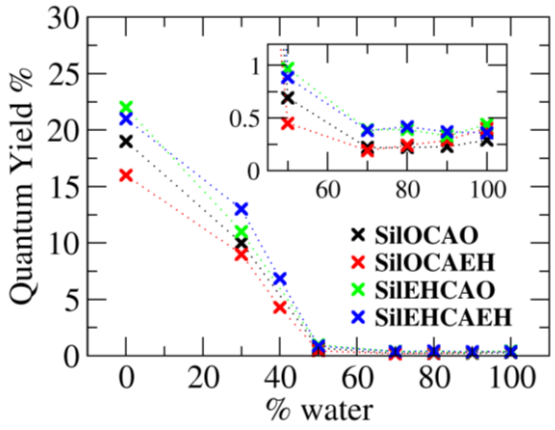


Figure S7. Fluorescence quantum yield of the dyes in THF/water mixtures of different composition (percentage of water) measured for excitation in proximity of the absorption maximum. The standard for fluorescence quantum yield measurements was a dilute solution of HITCI in ethanol (λ_{exc} = 680 nm, QY = 28.3%). The inset is a magnification of the region of the graph with low quantum yield.

Table S1. Mean diameter of the particles formed in the THF/water mixtures estimated from dynamic light scattering experiments on freshly prepared samples. The autocorrelation function was fitted with a multi-exponential distribution, of which only the main component is reported (values are the average of three measurements).

Compound	Mixture	Mean diameter
SilOCAO	70	90 ± 1 (87%)
	80	61 ± 3 (88%)
	90	44 ± 2 (95%)
	100	64.0 ± 0.6 (96%)
SilOCAEH	70	84 ± 1 (90%)
	80	58 ± 1 (91%)
	90	48.1 ± 0.9
	100	59.4 ± 0.6 (97%)
SilEHCAO	70	119 ± 5 (93%)
	80	76 ± 3 (95%)
	90	58 ± 3 (98%)
	100	87 ± 5 (97%)
SilEHCAEH	70	125.7 ± 0.5
	80	76 ± 4 (96%)
	90	61 ± 4 (98%)
	100	76 ± 2 (100%)

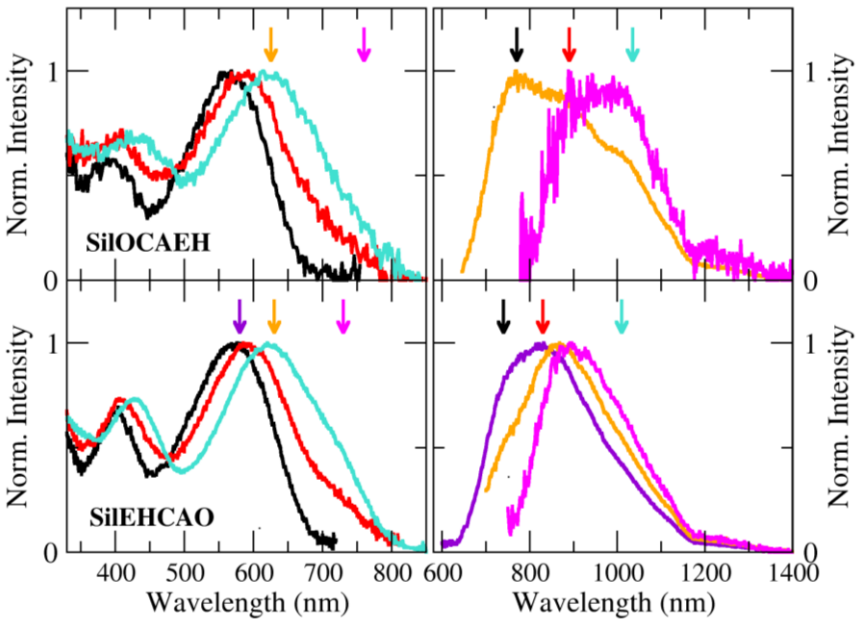


Figure S8. Excitation (left panels) and emission (right panels) spectra of **SilOCAEH** and **SilEHCAO** in a 50% THF/water mixture, collected for different emission/excitation wavelengths. The emission (excitation) wavelengths are indicated by the coloured arrows in the right (left) panels.

1.3. Deconvolution of the emission spectra in THF/water mixtures

Emission spectra $I(\tilde{\nu})$ of **SilOCAO** and **SilEHCAEH** in THF/water mixtures of different composition, collected for excitation at the absorption maximum, were fitted (on the wavenumber scale) with a sum of Gaussian functions of the form

$$I(\tilde{\nu}) = \sum_{i=1}^{n} G_i(\tilde{\nu})$$

with $$G_i(\tilde{\nu}) = a_i e^{-\frac{(\tilde{\nu}-b_i)^2}{2{c_i}^2}}$$

The fitting was performed with the least squares algorithm, allowing all parameters (a_i, b_i and c_i) to readjust freely. We fixed $n = 1 - 3$, depending on the mixture (Figure S9 and S10). The parameters obtained from the fit are provided in Table S2. In the lower panels of Figure S9 and S10, the Gaussian functions are labelled G0 ($n = 1$), G1 ($n = 2$) and G2 ($n = 3$).

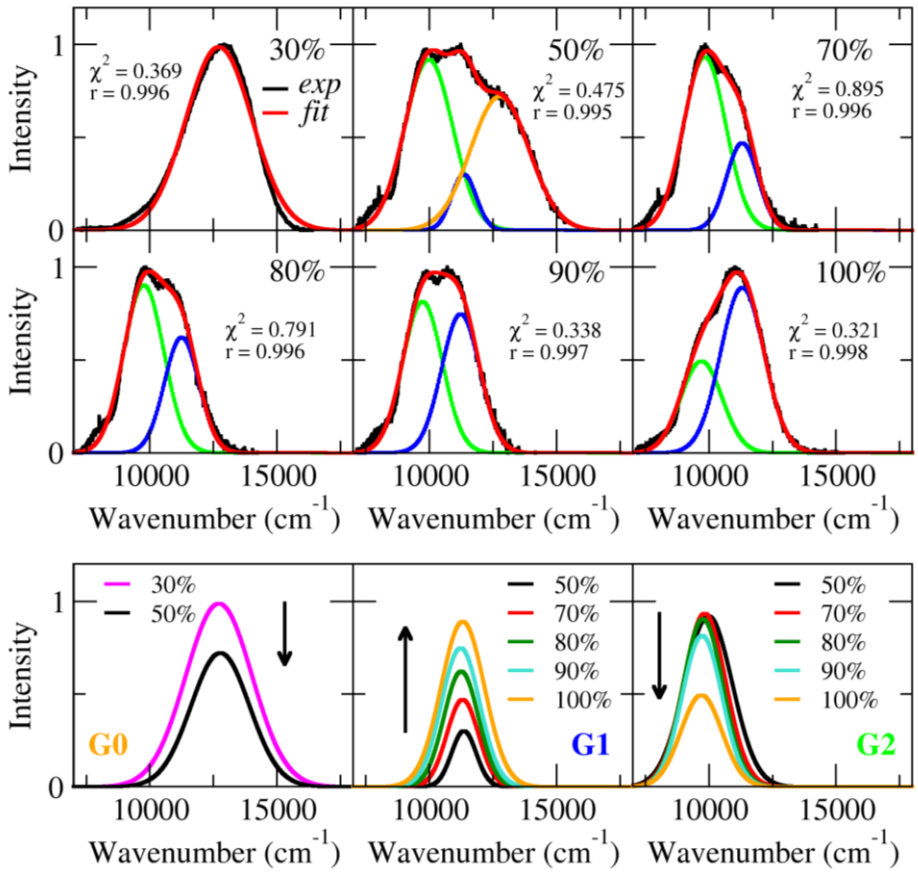


Figure S9. Deconvolution of **SilOCAO** emission spectra in THF/water mixtures of different composition (indicated as % of water).

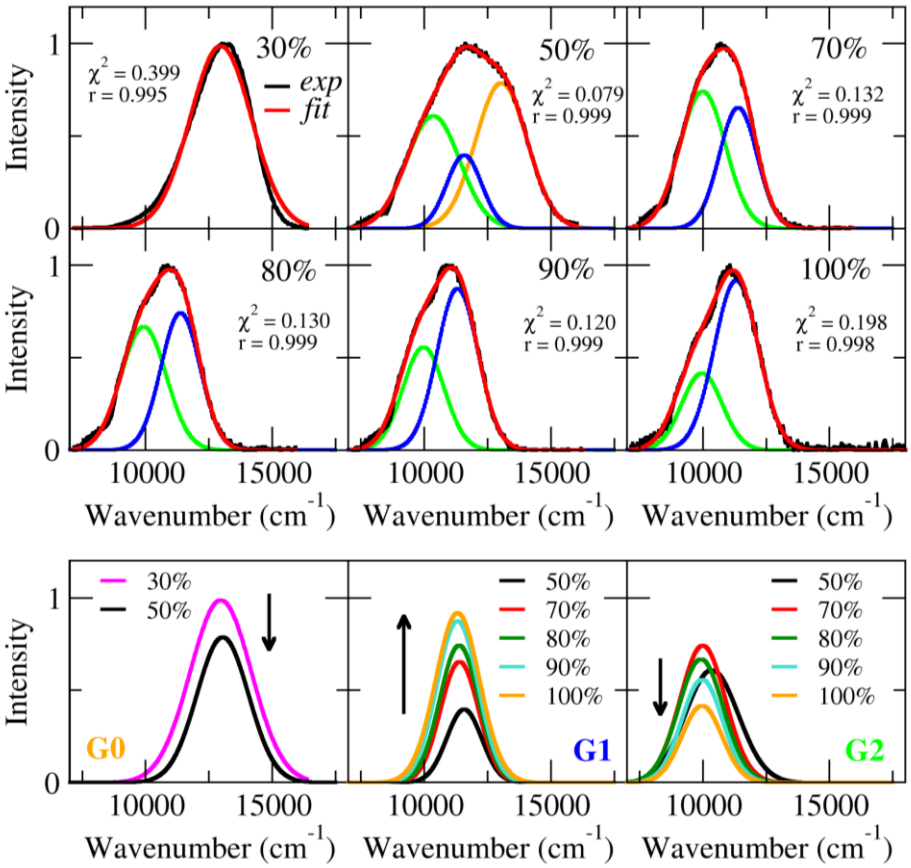


Figure S10. Deconvolution of **SilEHCAEH** emission spectra in THF/water mixtures of different composition (indicated as % of water).

Table S2. Parameters used for the fit of **SilOCAO** and **SilEHCAEH** emission spectra.

		SilOCAO			**SilEHCAEH**		
%water	*i*	*a*	*b*	*c*	*a*	*b*	*c*
30%	1	0.987	12717	1303	0.9868	12966	1209
50%	1	0.920	9974	944.0	0.7865	13043	1018
	2	0.301	11356	483.7	0.6088	10351	1047
	3	0.721	12770	1148	0.3958	11576	690.7
70%	1	0.934	9831	835.1	0.7413	9994	900.5
	2	0.470	11307	619.6	0.6541	11392	744.3
80%	1	0.904	9762	788.4	0.6680	9923	886.1
	2	0.622	11249	678.8	0.7426	11376	768.7
90%	1	0.815	9730	776.4	0.5575	9970	807.5
	2	0.748	11228	750.9	0.8737	11312	808.4
100%	1	0.494	9701	808.9	0.4162	9963	791.5
	2	0.890	11310	854.9	0.9182	11317	868.3

2. DFT and TDDFT calculations

2.1. Computational details

DFT and TDDFT calculations were performed with the Gaussian16 package.[3] The geometries of two selected conformations, ***a*** and ***b*** (Figure S11) were optimized in the gas phase using the hybrid exchange-correlation functional CAM-B3LYP[4] and the 6-31+G(d,p) basis set. To save computational resources, the long alkyl chains were replaced by methyl groups, as the nature of the aliphatic substituents does not affect the low energy photophysics of the chromophores. Stable minima were confirmed by frequency analysis. TDDFT calculations were performed on the optimized ground-state geometries at the same level of theory, asking for the lowest 15 electronic singlets states.

2.2. DFT and TDDFT data

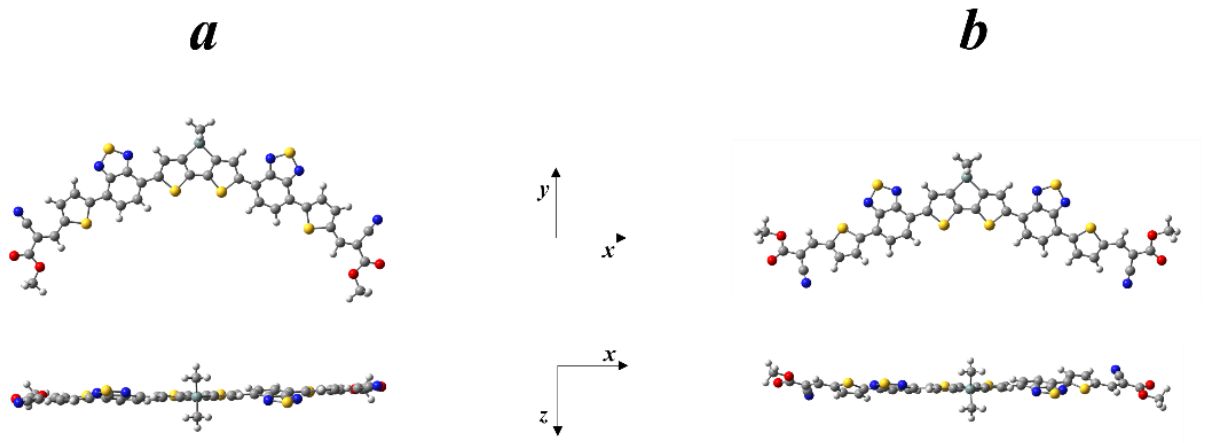


Figure S11. Front and side views of the optimized geometry of two conformers in the ground state (CAM-B3LYP/6-31+G(d,p) in gas phase). Conformer ***a*** is more stable by 0.37 kcal/mol.

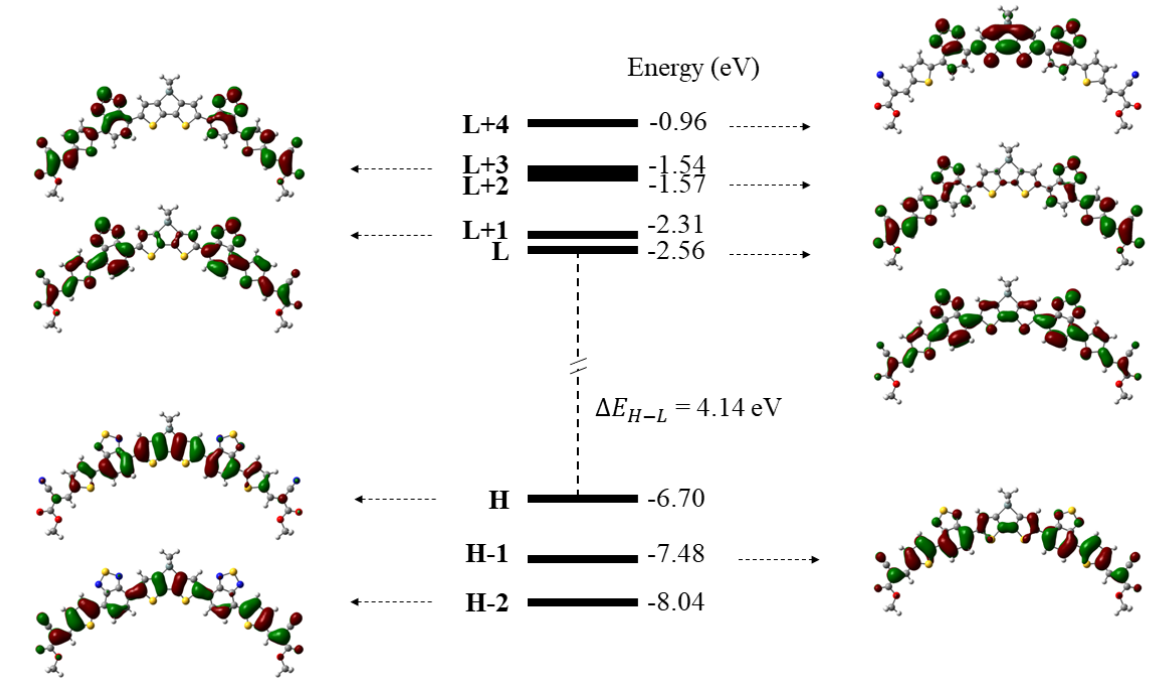


Figure S12. Frontier molecular orbitals (isovalue 0.02) of conformer ***a*** (CAM-B3LYP/6-31+G(d,p), gas phase).

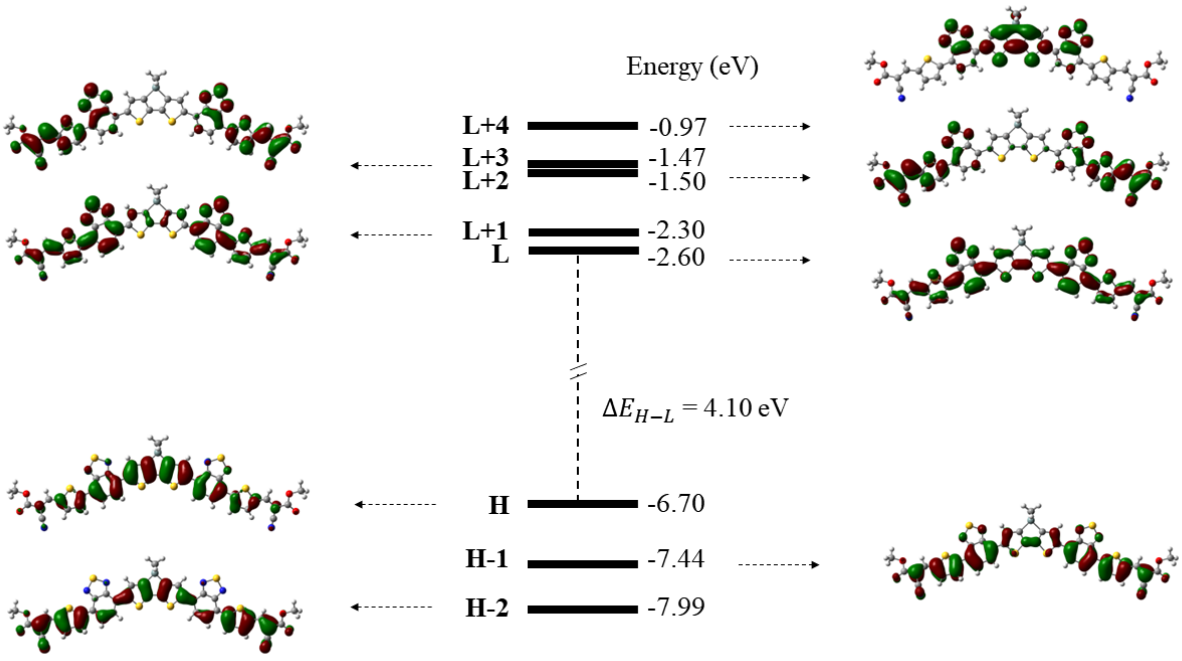


Figure S13. Frontier molecular orbitals (isovalue 0.02) of conformer ***b*** (CAM-B3LYP/6-31+G(d,p), gas phase).

Table S3. TDDFT data on two conformers in gas phase (CAM-B3LYP/6-31+G(d,p)): transition energies and wavelengths, oscillator strength (f), components of the transition dipole moment (μ_{ge}, see Cartesian axes in Figure S11), main orbital contributions.

Conformer	Transition	Energy (eV)	Wavelength (nm)	f	μ_{ge} (Debye)	Type (>20%)
a	$S_0 \rightarrow S_1$	2.32	535	2.141	-6.14 (x) 0.03 (z)	H → L (80%)
	$S_0 \rightarrow S_2$	2.72	455	0.306	-2.14 (y)	H-1 → L (22%) H → L+1 (68%)
	$S_0 \rightarrow S_3$	3.45	360	0.179	-1.45 (x) 0.11 (z)	H-2 → L (27%) H-1 → L+1 (31%)
	$S_0 \rightarrow S_4$	3.45	360	0.269	-1.78 (y)	H-2 → L+1 (21%) H-1 → L (40%)
	$S_0 \rightarrow S_5$	3.56	348	0.484	-2.35 (x) 0.04 (z)	H-1 → L+3 (22%) H → L+2 (30%) H → L+4 (35%)
	$S_0 \rightarrow S_6$	3.61	343	0.097	1.05 (y)	H-1 → L+2 (25%) H → L+3 (37%)
b	$S_0 \rightarrow S_1$	2.30	538	2.268	-6.34 (x) 0.04 (z)	H → L (81%)
	$S_0 \rightarrow S_2$	2.71	457	0.193	1.70 (y)	H-1 → L (22%) H → L+1 (69%)
	$S_0 \rightarrow S_3$	3.44	360	0.107	1.13 (y)	H-2 → L+1 (22%) H-1 → L (47%)
	$S_0 \rightarrow S_4$	3.45	359	0.123	-1.21 (x)	H-2 → L (28%) H-1 → L+1 (33%) H → L+4 (29%)
	$S_0 \rightarrow S_5$	3.58	346	0.841	-3.09 (x) -0.05 (z)	H-1 → L+3 (23%) H → L+2 (33%) H → L+4 (29%)
	$S_0 \rightarrow S_6$	3.64	340	0.068	0.87 (y)	H-1 → L+2 (28%) H → L+3 (43%)

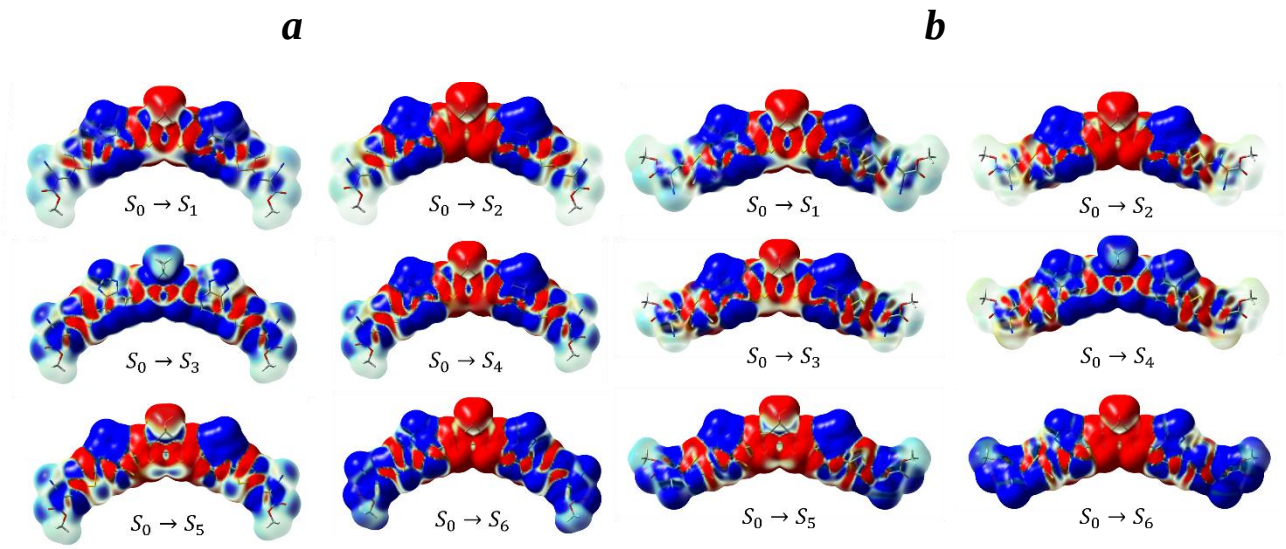


Figure S14. Density difference plots for the six lowest-energy electronic transitions of two conformers, ***a*** and ***b***. The difference between the electron density of the *n*-th excited state (S_n) and the electron density of the ground state is mapped on the ground state electron density (isovalue 0.0004). Red areas: negative sign (charge depletion); blue areas: positive sign (charge increase).

3. Essential-state models

3.1. Modelling of the solvated dye

The chromophores were modelled as V-shaped quadrupolar (A-D-A) dyes, with the molecular arms forming an angle α. To describe the non-interacting chromophore, we adopted the model described in Ref.[5], including electron-vibration coupling and polar solvation.

Shortly, we considered the main resonating structures of the dye, a neutral $|N\rangle$ (A-D-A) and two charge-separated $|Z_1\rangle$ and $|Z_2\rangle$ states (A^--D^+-A and A-D^+-A^-) as the electronic basis for the molecular Hamiltonian. We set 2η as the energy gap between the neutral state and the (degenerate) zwitterionic states, and $-\sqrt{2}t$ as the probability of electron transfer from the donor to either of the acceptor sites. To define the dipole moment operator on the chosen basis, all matrix elements are neglected except for the dipole moment of the zwitterionic states, μ_0, that is the leading term.

Exploiting symmetry, the two charge-separated states can be combined into a symmetric $|Z_+\rangle$ and an asymmetric $|Z_-\rangle$ state. The $|Z_+\rangle$ state mixes with the neutral state giving the ground state $|g\rangle$ and a higher energy excited-state $|e\rangle$, while $|Z_-\rangle$ stays unmixed and corresponds to the lowest-energy excited state $|c\rangle$. Thus, the model describes two electronic transitions, $|g\rangle \rightarrow |c\rangle$ and $|g\rangle \rightarrow |e\rangle$, both 1PA and 2PA allowed.

Electron-phonon coupling was included introducing two mutually independent effective vibrational coordinates with harmonic frequency ω_v and vibrational relaxation energy ε_v, describing nuclear relaxation along each molecular arm.

The coupled electron-vibration problem was solved non-adiabatically. The Hamiltonian was written on the basis given by the direct product of the three electronic basis states times the first 10 states of the harmonic oscillators associated to the vibrational coordinates, and diagonalized to get the numerically exact eigenstates. Absorption and emission spectra were obtained calculating the transition dipole moments between pairs of states and assigning to each transition a Gaussian band with half-width at half-maximum $\gamma = \sigma\sqrt{2\ln 2}$.

Specifically, the molar extinction coefficient $\varepsilon(\tilde{\nu})$ (units of M^{-1} cm^{-1}), was obtained from the following sum-over-states expression:

$$\varepsilon(\tilde{\nu}) = \frac{10\pi N_A \tilde{\nu}}{3\ln 10 \hbar c \varepsilon_0} \frac{1}{\sigma\sqrt{2\pi}} \sum_n \mu_{gn}^2 \exp\left[-\frac{1}{2}\left(\frac{\tilde{\nu}_{gn}-\tilde{\nu}}{\sigma}\right)\right] \quad \text{(S1)}$$

where $\tilde{\nu}$ is the wavenumber, N_A is the Avogadro number, c is the light speed, ε_0 is the vacuum dielectric constant, σ is the width of the Gaussian function assigned to each transition, μ_{gn} and $\tilde{\nu}_{gn}$ are the transition dipole moment and wavenumber of the transition from the ground state (g) to the excited state n and the sum runs over all the excited states.

The fluorescence spectrum $I(\tilde{\nu})$ was obtained as

$$I(\tilde{\nu}) \propto \frac{\tilde{\nu}^3}{\sigma\sqrt{2\pi}} \sum_n \mu_{fn}^2 exp\left[-\frac{1}{2}\left(\frac{\tilde{\nu}_{fn}-\tilde{\nu}}{\sigma}\right)^2\right] \quad \text{(S2)}$$

where f denotes the fluorescent state and the sum runs over all states having lower energy than f.

The effects of polar solvation were included in a reaction-field approach, considering the solute embedded in a continuum dielectric medium. In response to the electric field generated by the solute, the solvent molecules reorient forming a reaction-field F_{or} proportional to the solute dipole moment. The solvent polarity is described by a single parameter, ε_{or}, increasing with solvent polarity. To account for fluctuations of F_{or} around the equilibrium value (thermal disorder), responsible for inhomogeneous broadening effects, absorption and emission spectra (Eq. S1 and S2) were calculated over a grid of F_{or} values and then averaged according to the Boltzmann probability to get the final spectra.

An issue coming up when modelling the target dyes is the presence of many low-energy CT transitions (revealed by experimental data and confirmed by TDDFT calculations), while the three-state model describes only two electronic transitions. However, the linear properties (absorption and emission) in the visible window are dominated by the $S_0 \rightarrow S_1$, that we identified with the $|g\rangle \rightarrow |c\rangle$. The $S_0 \rightarrow S_2$ transition gives a minor contribution to 1PA and, since we are interested in the modelling of linear properties, it was disregarded. The $|g\rangle \rightarrow |e\rangle$ was assumed at higher energy, accounting in an averaged way for the many CT transitions in the UV that overlap giving the 400 nm absorption band. The set of model parameters that reproduces at best experimental properties (absorption and emission in different solvents) is provided in Table S4, while calculated spectra are provided in Figure S15. The agreement with experiment is good, especially in view of the limited number of parameters and the approximations introduced: not only bandshapes are well reproduced, but also the solvent effect is quantitatively accounted for, validating the three-state model for the chromophores at hand.

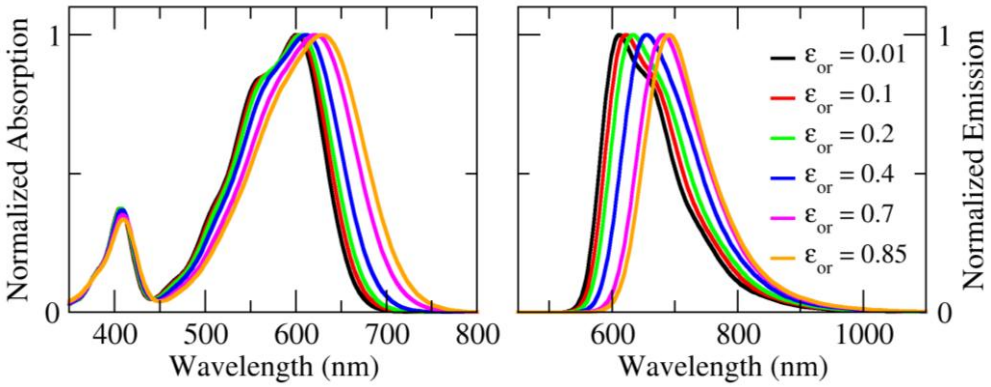


Figure S15. Absorption and emission spectra in solution calculated with the three-state model. Model parameters are listed in Table S4(i). The solvent relaxation energy ε_{or} mimicking different solvents is given in the legend (in eV): black: cyclohexane; red: toluene; green: chloroform; blue: acetone; magenta: dimethyl formamide; orange: dimethyl sulfoxide.

Table S4. Model parameters with (i) and without (ii) electron-vibration coupling.

	η (eV)	$\sqrt{2}t$ (eV)	ω_v (eV)	ε_v (eV)	γ (eV)	μ_0 (D)	α (deg)
i	0.81	0.95	0.19	0.68	0.10	26.0	125
ii	0.66	0.95	-	-	0.10	26.0	125

3.2. Modelling of the aggregates

Having validated the approach against experimental data of non-interacting chromophores, we can address the modelling of aggregates.[6,7] To reduce the computational cost (allowing to describe up to 6 interacting molecules), we moved to a simplified version of the model that neglects intramolecular vibrations, and calculations were performed in gas phase. To maintain the consistency with experiment, the model parameters were consequently adjusted to reproduce the experimental transition frequencies of the monomer (Table S4(ii) and Figure S16).

The Hamiltonian describing an aggregate of N chromophores is written on the (3^N-dimensional) basis obtained as the direct product of the three basis states of each chromophore ($|N\rangle$, $|Z_1\rangle$ and $|Z_2\rangle$), as the sum of the single-molecule Hamiltonians plus a term accounting for intermolecular interactions. Intermolecular interactions are introduced as Coulombic interactions between positive or negative integer point charges located on the donor and acceptor sites respectively and enter the Hamiltonian only when at least two molecules are in a zwitterionic state. Thus, the magnitude and sign of the intermolecular interactions depend on the geometry of the aggregate. In this work we considered two specific packings, called *up-up* and *up-down*, depicted in Figure 11a (main text). For aggregates made up of more than two molecules we assumed a ladder geometry, with the third molecule stacking on the second one at a distance z and same shift along the x and y coordinates and so on. Only nearest-neighbour interactions were accounted for. In the model, l is the length of the molecular arm and was kept fixed to $l = 8$ Å.

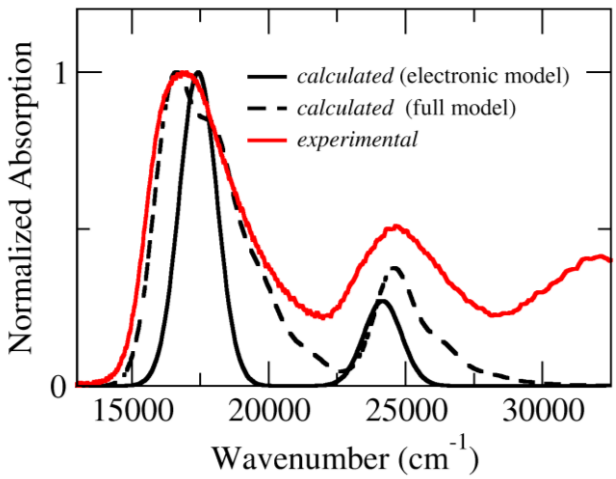


Figure S16. Comparison between the absorption spectrum of the monomer calculated in the gas phase neglecting electron-phonon coupling, after a renormalization of η (black line, parameters in Table S4(ii)), the absorption spectrum calculated in a low-polarity solvent (ε_{or} = 0.1 eV) explicitly accounting for vibrations (full model, dashed line), and the experimental absorption spectrum of **SilEHCAO** in cyclohexane (red line).

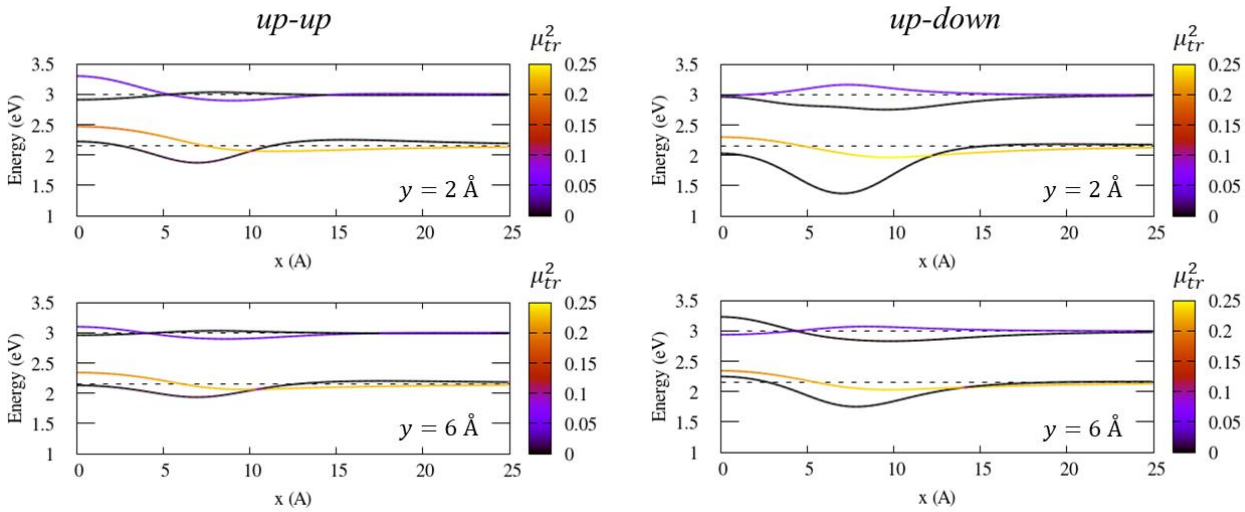


Figure S17. Energy of the lowest electronic states of a dimer (*up-up* and *up-down* packings) as a function of x for $y = 2$ or 6 Å and $z = 5$ Å. Ground state is taken as the zero of the energy. The colour palette is proportional to the squared transition dipole moment μ_{tr} from the ground state (i.e., black lines indicate a dark state, yellow lines fully allowed transitions). Dashed lines mark the transition energies of the monomer.

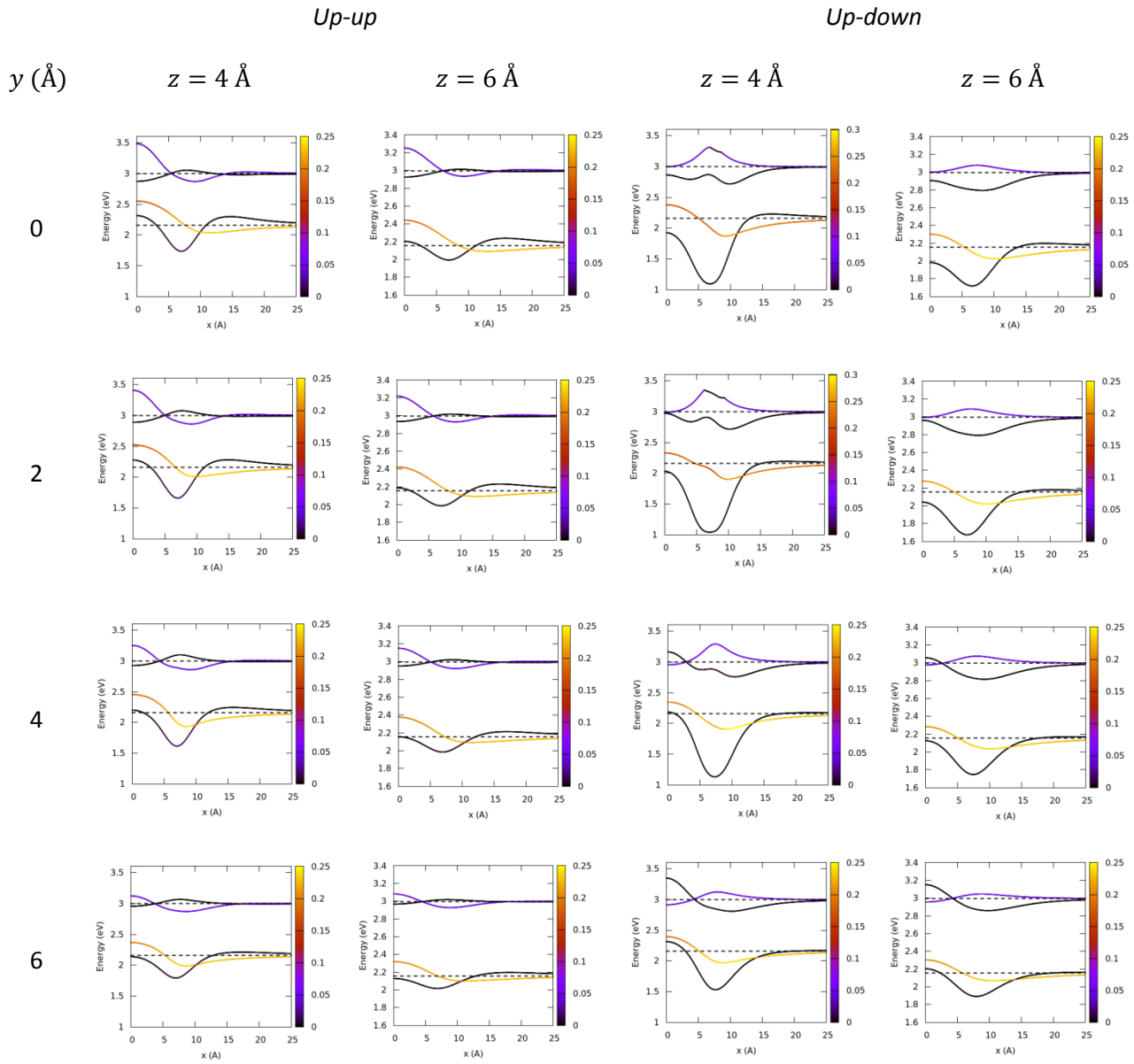


Figure S18. Energy of the lowest electronic states of a dimer (*up-up* and *up-down* packings) as a function of x for selected values of y and z. Ground state is taken as the zero of the energy. The colour palette is proportional to the squared transition dipole moment μ_{tr} from the ground state (i.e., black lines indicate a dark state, yellow lines fully allowed transitions). Dashed lines mark the transition energies of the monomer.

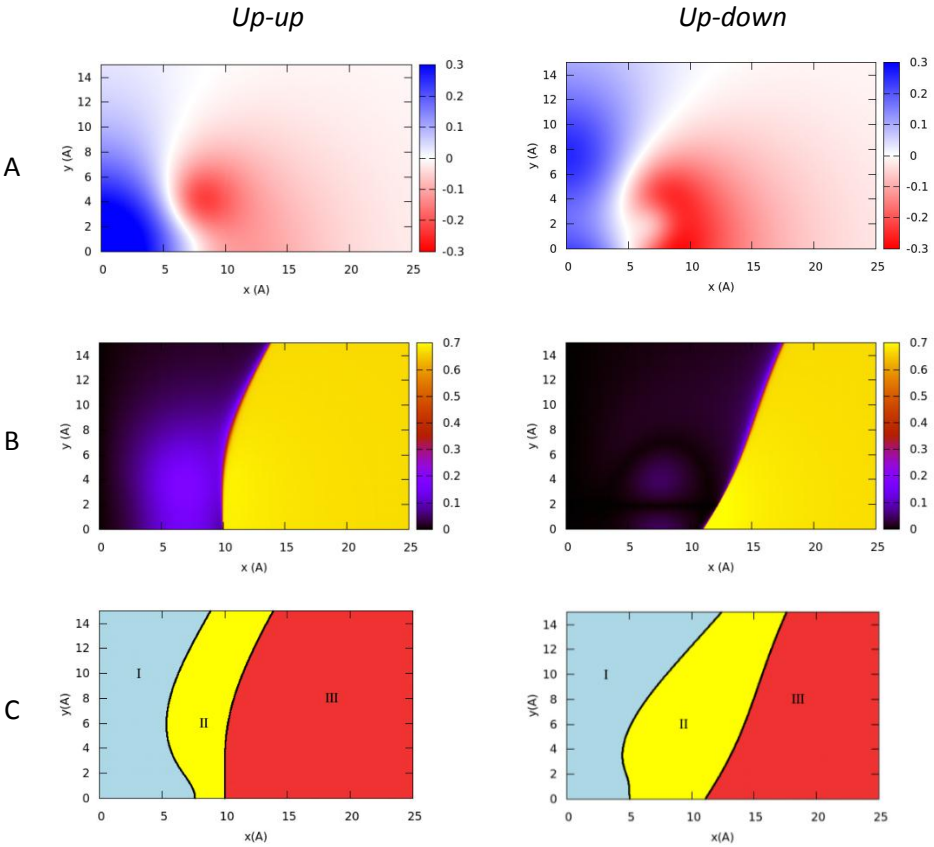


Figure S19. Dependence of the optical properties calculated for an *up-up* and *up-down* dimer on the aggregate geometry (variable x and y, $z = 4$ Å). (A) Absorption shift. The colour palette indicates (in eV) the difference between the absorption energy of the dimer and the absorption energy of the monomer (blue: blue shift, red: red shift). (B) Emission dipole moment (dimensionless). (C) Phase diagrams. Class I: blue-shifted absorption and fluorescence quenching (H-aggregates); class II: red-shifted absorption and fluorescence quenching (non-fluorescent J-aggregates or red-shifted H-aggregates); class III: red-shifted absorption and optically allowed emission (J-aggregates).

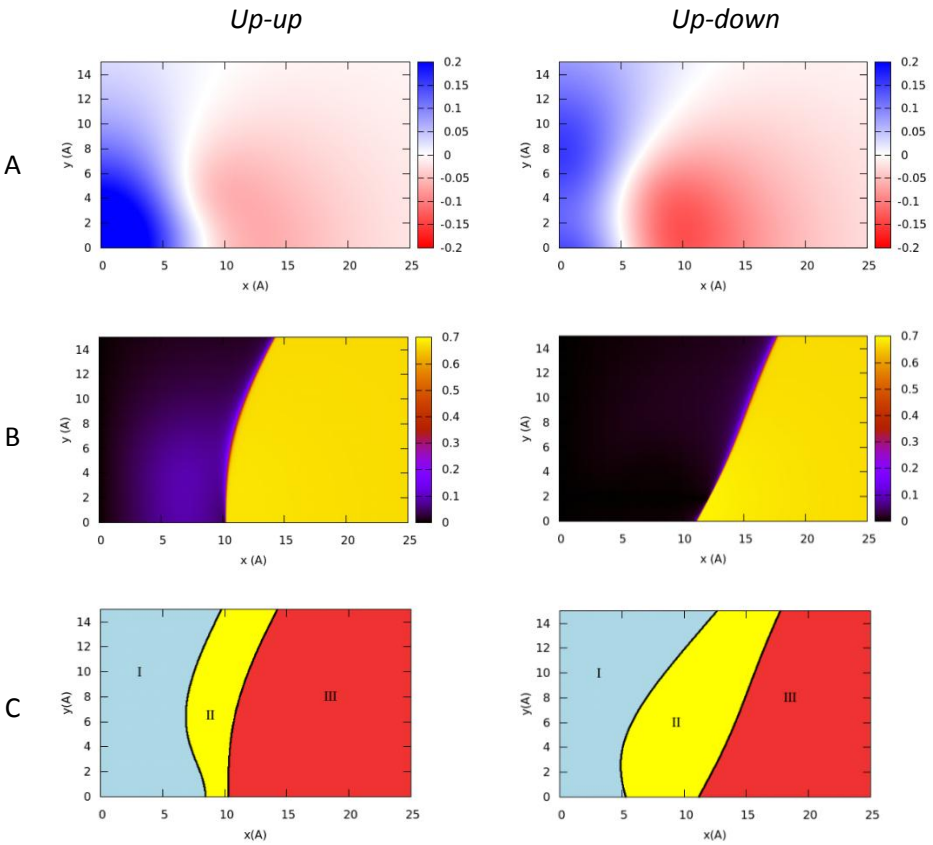


Figure S20. Dependence of the optical properties calculated for an *up-up* and *up-down* dimer on the aggregate geometry (variable x and y, $z = 6$ Å). (A) Absorption shift. The colour palette indicates (in eV) the difference between the absorption energy of the dimer and the absorption energy of the monomer (blue: blue shift, red: red shift). (B) Emission dipole moment (dimensionless). (C) Phase diagrams. Class I: blue-shifted absorption and fluorescence quenching (H-aggregates); class II: red-shifted absorption and fluorescence quenching (non-fluorescent J-aggregates or red-shifted H-aggregates); class III: red-shifted absorption and optically allowed emission (J-aggregates).

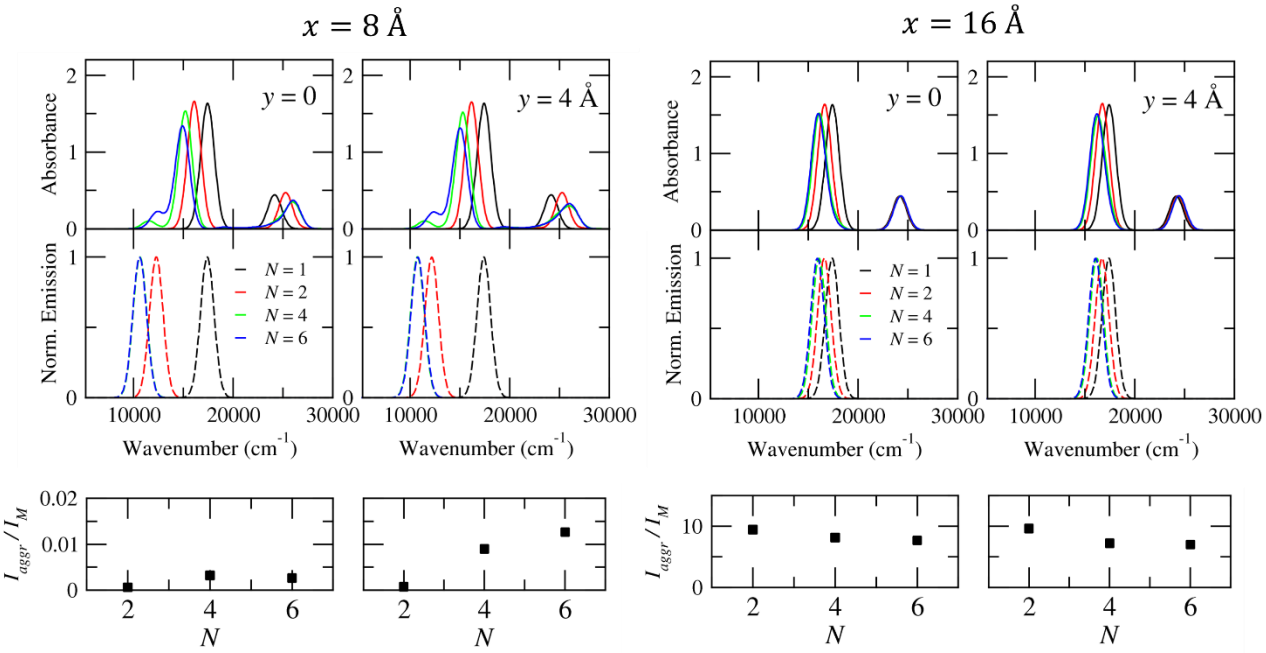


Figure S21. The same as in Figure 12 (main text) for the *up-down* packing.

4. References

1 C. Xu and W. W. Webb, *Journal of the Optical Society of America B*, 1996, **13**, 481.

2 M. A. Albota, C. Xu and W. W. Webb, *Appl Opt*, 1998, **37**, 7352.

3 M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, G. A. Petersson, H. Nakatsuji, X. Li, M. Caricato, A. V. Marenich, J. Bloino, B. G. Janesko, R. Gomperts, B. Mennucci, H. P. Hratchian, J. V. Ortiz, A. F. Izmaylov, J. L. Sonnenberg, D. Williams-Young, F. Ding, F. Lipparini, F. Egidi, J. Goings, B. Peng, A. Petrone, T. Henderson, D. Ranasinghe, V. G. Zakrzewski, J. Gao, N. Rega, G. Zheng, W. Liang, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, K. Throssell, J. E. Montgomery, J. A., Jr. Peralta, F. Ogliaro, M. J. Bearpark, J. J. Heyd, E. N. Brothers, K. N. Kudin, V. N. Staroverov, T. A. Keith, R. Kobayashi, K. Normand, J. Raghavachari, A. P. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, J. M. Millam, M. Klene, C. Adamo, R. Cammi, J. W. Ochterski, R. L. Martin, K. Morokuma, O. Farkas, J. B. Foresman and D. J. Fox, Gaussian 16, revision B.01, Gaussian Inc. Wallingford CT, 2016.

4 T. Yanai, D. P. Tew and N. C. Handy, *Chem. Phys. Lett.*, 2004, **393**, 51–57.

5 F. Terenziani, A. Painelli, C. Katan, M. Charlot and M. Blanchard-Desce, *J Am Chem Soc*, 2006, **128**, 15742–15755.

6 C. Sissa, F. Terenziani, A. Painelli, A. Abbotto, L. Bellotto, C. Marinzi, E. Garbin, C. Ferrante and R. Bozio, *Journal of Physical Chemistry B*, 2010, **114**, 882–893.

7 B. Bardi, C. Dall'Agnese, K. I. Moineau-Chane Ching, A. Painelli and F. Terenziani, *The Journal of Physical Chemistry C*, 2017, **121**, 17466–17478.