Electronic Supplementary Material (ESI) for Chemical Science.

Emergent Supramolecular Assembly Properties of a Recognition-Encoded Oligoester

Supplementary Information

Filip T. Szczypiński, Luca Gabrielli and Christopher A. Hunter

Contents

1 General Methods

Commercial reagents were used as received without further purification. Dichloromethane and tetrahydrofurane were purified and dried using PureSolv MD 5 Solvent Purification System. Routine and characterisation NMR spectra were recorded on Bruker 400 MHz Avance III HD Smart Probe, 400 MHz Smart Probe, and 400 MHz Avance III HD Spectrometers at 298 K and using Wilmard 5 mm Thin Wall Precision NMR sample tubes. Characterisation spectra for **ADAD** were recorded by the NMR Service at the Department of Chemistry on 500 MHz DCH Cryoprobe and 400 MHz Neo Prodigy Spectrometers. ^{1}H-^{1}H DQF COSY, ^{1}H-^{13}C HMBC, ^{1}H-^{31}P HMBC, and ^{13}C-^{19}F HMBC of **ADAD** were recorded with 35.5% Poisson gap non-uniform sampling. ^{1}H-^{13}C HSQC, ^{13}C-^{19}F HSQC, and ^{13}C-^{19}F HMQC od **ADAD** were recorded with 25% Poisson gap non-uniform sampling. NMR dilutions were performed on Bruker 400 MHz Avance III HD Smart Probe Spectrometer by adding aliquotes of a concentrated solution into pure solvent. Upon each addition, the solution was manually shaken before acquiring the spectrum, which was sufficient time for equilibration to be reached. Variable temperature NMR experiments were performed on 500 MHz AVIII HD Smart Probe Spectrometer. DOSY experiments were performed on Bruker 400 MHz Avance III HD Smart Probe Spectrometer. The standard Bruker pulse program, ledbpgp2s, employing a stimulated echo and longitudinal eddy-current delay (LED) using bipolar gradient pulses for diffusion using 2 spoil gradients of 600 μs was utilized. Rectangular gradients were used with a total duration of 3 ms. Gradient recovery delays were 200 μs. Diffusion times were 160 μs. Individual rows of the quasi-2D diffusion databases were phased and baseline corrected. DOSY data analyses were performed using Bruker Dynamics Centre software, using area integrals of manually selected peaks. Chemical shifts for ^{1}H, ^{13}C, ^{19}F, and ^{31}P are reported in ppm on the δ scale; ^{1}H and ^{13}C were referenced to the residual solvent peak; ^{19}F and ^{31}P were unreferenced. Coupling constants (*J*) are reported in hertz (Hz). The following abbreviations are used to describe signal multiplicity for ^{1}H and ^{13}C NMR spectra: s: singlet, d: doublet, t: triplet, m: multiplet, br: broad. High resolution electrospray ionization mass spectrometry (HRMS-ESI) was performed on Waters LCT Premier TOF Spectrometer or by the Mass Spectrometry Service at the Department of Chemistry. Infrared (IR) spectra were recorded on Bruker Alpha FTIR Spectrometer with single reflection diamond Platinum ATR. The liquid chromatography mass spectrometry (LC-MS) analysis of samples was performed using Waters Acquity H-class UPLC coupled with a single quadrupole Waters SQD2. Acquity UPLC CSH C18 Column, 130Å, 1.7 μm, 2.1 mm X 50 mm was used as the UPLC column. The conditions of the UPLC method were as follows: solvent A: water +0.1% Formic acid; solvent B: acetonitrile +0.1% formic acid; gradient of 0-2 minutes 5% - 100%B + 1 minute

100%B with re-equilibration time of 2 minutes. Flow rate: 0.6 ml/min; column temperature of 40 °C; injection volume of 2 µL. The signal was monitored at 270 nm. Chromatographic separations were performed on Teledyne ISCO CombiFlash Rf+UV-Vis and CombiFlash Rf+Lumen, using prepacked cartridges of silica (25 µm or 50 µm PuriFlash Columns). Separated mixtures were solid loaded using Celite or silica gel 60 (Merck, 40–63 µm). The signal was monitored at 254 nm, 270 nm, and (if CombiFlash Rf+Lumen) using evaporative light scattering detector.

2 Synthesis

2.1 Synthesis of 3

A solution of previously reported compounds **1** (260 mg, 0.44 mmol) and **2** (215 mg, 0.44 mmol) in anhydrous dichloromethane (5 ml) with *N,N*-dimethylaminopyridine (11 mg, 0.1 mmol) and EDC·HCl (101 mg, 0.5 mmol) were stirred overnight under nitrogen atmosphere.[1] The mixture was poured into water (10 ml) and the aqueous layer was extracted with dichloromethane (3 × 5 ml). The combined organic extracts were washed with brine (20 ml), dried with anhydrous magnesium sulfate, filtered and concentrated under vacuum. The residue was purified using column chromatography (0-100% ethyl acetate in petroleum ether) to give compound **3** as a white wax (266 mg, 0.25 mmol, 57%).

^{1}H NMR (400 MHz, $CDCl_3$) δ 7.83 – 7.28 (m, 20H), 7.22 (d, J = 8.4 Hz, 2H), 6.71 (d, J = 8.4 Hz, 2H), 6.49 (d, J = 8.3 Hz, 2H), 5.18 (s, 2H), 4.38 (br t, J = 6.1 Hz, 2H), 4.08 (s, 2H), 4.03 (s, 2H), 3.81 (br t, J = 6.1 Hz, 2H), 3.70 (br t, J = 6.1 Hz, 2H), 3.54 (br t, J = 6.1 Hz, 2H), 2.35 (s, 3H), 2.05 – 1.56 (m, 6H), 1.09 – 0.98 (m, 15H), 0.84 (d, J = 6.8 Hz, 6H) ppm.

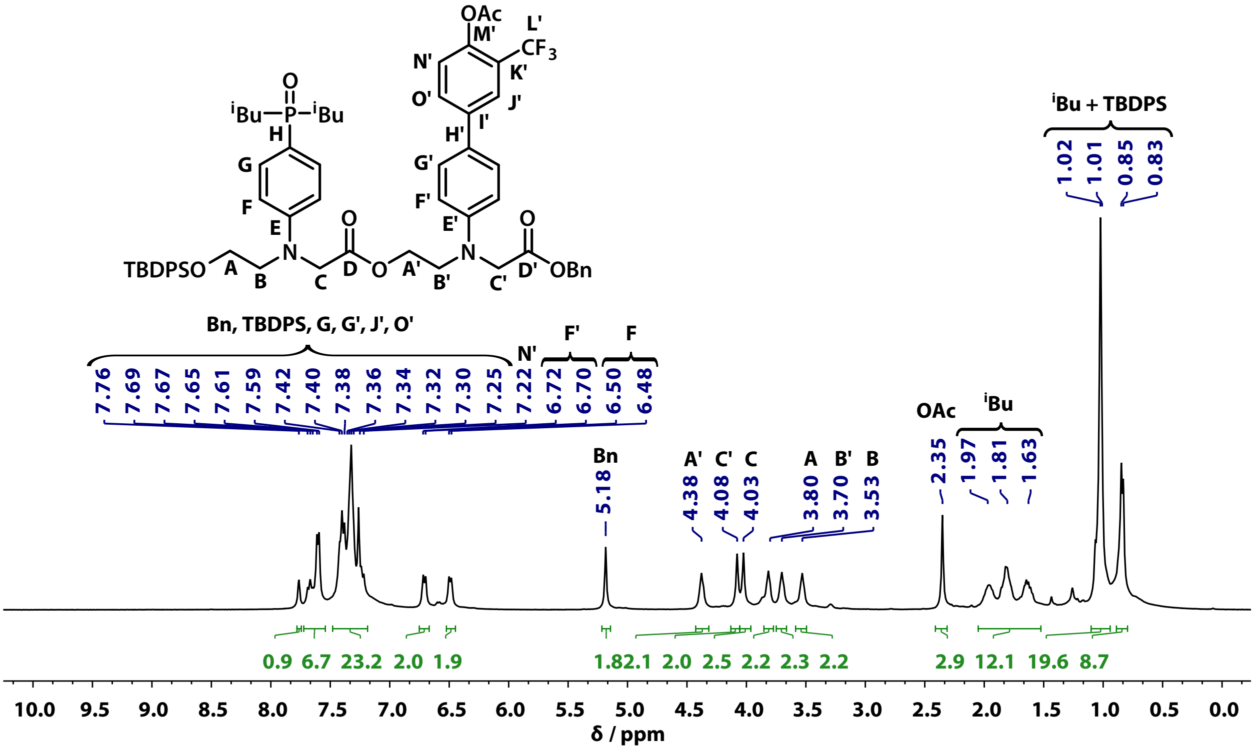
^{13}C NMR (101 MHz, $CDCl_3$) δ 170.1 (d, J = 13 Hz), 168.7, 149.5, 147.0, 138.7, 135.2, 135.0, 132.7, 131.5 (d, J = 10 Hz), 130.1, 129.4, 128.3, 128.2, 128.0, 128.0, 127.7, 127.4, 124.3, 124.1 (q, J = 5 Hz), 122.8 (q, J = 272 Hz), 119.3 (d, J = 100 Hz), 112.2, 111.0 (d, J = 12 Hz), 66.6, 61.9, 61.0, 53.0, 52.3 (d, J = 17 Hz), 49.9, 39.7 (d, J = 68 Hz), 26.4, 24.4 (d, J = 9 Hz), 24.2 (d, J = 8 Hz), 23.1 (d, J = 4.0 Hz), 20.4, 18.7 ppm.

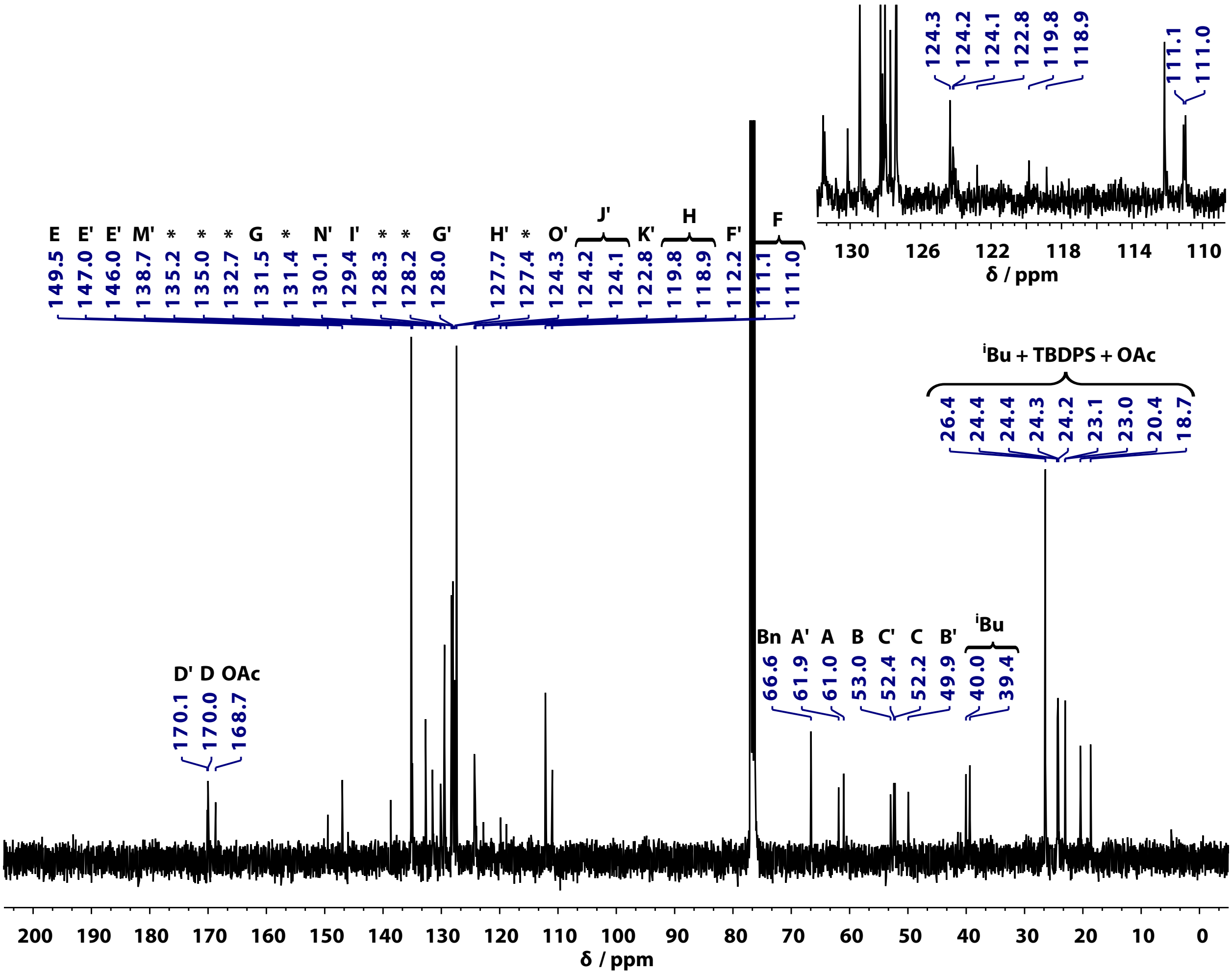
^{19}F NMR (376 MHz, $CDCl_3$) δ -61.9 ppm.

^{31}P NMR(162 MHz, $CDCl_3$) δ 41.2 ppm.

FT-IR (neat): 2955, 2925, 2894, 2862, 1772, 1745, 1598 cm^{-1}.

HR-MS (ESI): m/z = 1063.4680 $[M+H]^+$ (calcd. for $C_{60}H_{71}F_2N_2O_8PSi$: 1063.4669, Δ 1.0 ppm).


OAc
M'
L'
CF3
N'
K'
O'
J'
I'
H'
O
iBu—P—iBu
H
G
G'
F
F'
E
E'
TBDPSO
A
B
N
C
O
D
O
A'
B'
N
C'
O
D'
OBn
Bn, TBDPS, G, G', J', O'
7.76 7.69 7.67 7.65 7.61 7.59 7.42 7.40 7.38 7.36 7.34 7.32 7.30 7.25
N'
7.22
F'
6.72 6.70
F
6.50 6.48
Bn
5.18
A' C' C
4.38 4.08 4.03
A B' B
3.80 3.70 3.53
OAc
2.35
iBu
1.97 1.81 1.63
iBu + TBDPS
1.02 1.01 0.85 0.83
0.9 6.7 23.2 2.0 1.9
1.8 2.1 2.0 2.5 2.2 2.3 2.2
2.9 12.1 19.6 8.7
10.0 9.5 9.0 8.5 8.0 7.5 7.0 6.5 6.0 5.5 5.0 4.5 4.0 3.5 3.0 2.5 2.0 1.5 1.0 0.5 0.0
δ / ppm


124.3 124.2 124.1 122.8 119.8 118.9
111.1 111.0
130 126 122 118 114 110
δ / ppm
E E' E' M' * * * G * N' I' * * G' H' * O' J' K' H F' F
149.5 147.0 146.0 138.7 135.2 135.0 132.7 131.5 131.4 130.1 129.4 128.3 128.2 128.0 127.7 127.4 124.3 124.2 124.1 122.8 119.8 118.9 112.2 111.1 111.0
iBu + TBDPS + OAc
26.4 24.4 24.4 24.3 24.2 23.1 23.0 20.4 18.7
D' D OAc
170.1 170.0 168.7
Bn A' A B C' C B'
66.6 61.9 61.0 53.0 52.4 52.2 49.9
iBu
40.0 39.4
200 190 180 170 160 150 140 130 120 110 100 90 80 70 60 50 40 30 20 10 0
δ / ppm

2.2 Synthesis of 7

Nitrogen gas was bubbled for 15 min through a suspension of **3** (120 mg, 0.1 mmol) and palladium on carbon (11 mg, 10 wt% loading, 0.01 mmol) in absolute ethanol (10 ml). Hydrogen gas was then purged for 15 min through the suspension, which was subsequently left stirring under hydrogen atmosphere overnight. The reaction mixture was filtered through Celite, washed with ethanol (3×15 ml) and concentrated under vacuum to give intermediate **4** (**AD*-COOH**) as a white foam (87 mg, 0.09 mmol, 90%).

^{1}H NMR (400 MHz, $CDCl_3$) δ 7.74 (d, J = 2.1 Hz, 1H), 7.65 (dd, J = 8.4, 2.1 Hz, 1H), 7.63 – 7.29 (m, 14H), 7.19 (d, J = 8.4 Hz, 1H), 6.71 (d, J = 8.5 Hz, 2H), 6.51 (d, J = 8.2 Hz, 2H), 4.36 (t, J = 5.8 Hz, 2H), 4.05 (s, 2H), 3.89 (s, 2H), 3.82 (t, J = 5.8 Hz, 2H), 3.60 (t, J = 5.8 Hz, 2H), 3.55 (t, J = 5.8 Hz, 2H), 2.34 (s, 3H), 2.27 (br s, 1H), 2.04 – 1.63 (m, 6H), 1.02 (s, 9H), 0.99 (d, J = 6.6 Hz, 6H), 0.87 (d, J = 6.5 Hz, 6H) ppm.

HR-MS (ESI): m/z = 973.4228 $[M+H]^+$ (calcd. for $C_{53}H_{65}N_2O_8F_3PSi$: 973.4200, Δ 2.9 ppm).

In parallel, a solution of compound **3** (120 mg, 0.1 mmol) in anhydrous tetrahydrofuran (5 ml) with acetic acid (52 µl, 0.9 mmol) and *n*-tetrabutylammonium acetate (170 µl, 1 M in THF, 0.17 mmol) was stirred under nitrogen atmosphere overnight. Upon completion, the reaction mixture was diluted with water (10 ml) and extracted with ethyl acetate (3 × 10 ml). The combined organic extracts were washed with brine (10 ml), dried with anhydrous magnesium sulfate, filtered and concentrated under vacuum. The residue was filtered through a pad of silica (90% ethyl acetate in petroleum ether, then 5% methanol in dichloromethane) to give intermediate **5** (**HO-AD***) as a white foam (69 mg, 0.08 mmol, 76%).

^{1}H NMR (400 MHz, $CDCl_3$) δ 7.78 (d, J = 2.2 Hz, 1H), 7.70 (dd, J = 8.4, 2.3 Hz, 1H), 7.48 – 7.40 (m, 4H), 7.36 – 7.32 (m, 5H), 7.24 (d, J = 10.1 Hz, 3H), 6.75 (d, J = 8.7 Hz, 2H), 6.61 – 6.57 (m, 2H), 5.20 (s, 2H), 4.46 (t, J = 5.9 Hz, 2H), 4.15 (s, 2H), 4.09 (s, 2H), 3.78 (t, J = 5.7 Hz, 4H), 3.60 (m, 4H), 3.17 (s, 1H), 2.35 (s, 3H), 2.03 – 1.61 (m, 6H), 1.01 (d, J = 6.6 Hz, 6H), 0.84 (d, J = 6.4 Hz, 6H) ppm.

HR-MS (ESI): m/z = 825.3525 $[M+H]^+$ (calcd. for $C_{44}H_{53}F_3N_2O_8P$: 825.3492, Δ 4.0 ppm).

A solution of intermediates **4** (87 mg, 0.09 mmol) and **5** (69 mg, 0.08 mmol) in anhydrous dichloromethane (5 ml) with *N,N*-dimethylaminopyridine (11 mg, 0.1 mmol) and EDC·HCl (101 mg, 0.5 mmol) were stirred overnight under nitrogen atmosphere. The mixture was poured into water (10 ml) and the aqueous layer was extracted with dichloromethane (3 × 5 ml). The combined organic extracts were washed with brine (20 ml), dried with anhydrous magnesium sulfate, filtered and concentrated under vacuum. The residue was filtered through a pad of silica silica (60%, then 100% ethyl acetate in petroleum ether) to give compound **6** as a white wax (141 mg, 0.08 mmol, 99%).

^{1}H NMR (400 MHz, $CDCl_3$) δ 7.74 (m, 2H), 7.61 – 7.56 (m, 5H), 7.48 – 7.27 (m, 20H), 7.23 – 7.20 (m, 2H), 6.69 (d, J = 8.8 Hz, 2H), 6.67 – 6.61 (m, 4H), 6.49 – 6.45 (m, 2H), 5.16 (s, 2H), 4.40 – 4.27 (m, 6H), 4.07 (s, 2H), 4.03 (s, 2H), 3.98 (s, 2H), 3.94 (s, 2H), 3.80 (t, J = 5.9 Hz, 2H), 3.69 (t, J = 6.0 Hz, 2H), 3.62 (m, 4H), 3.53 (t, J = 5.7 Hz, 2H), 2.33 (s, 6H), 2.02 – 1.56 (m, 12H), 1.02 – 0.95 (m, 21H), 0.82 (d, J = 6.6 Hz, 12H) ppm.

HR-MS (ESI): m/z = 1779.7497 $[M+H]^+$ (calcd. for $C_{97}H_{115}O_{15}N_4F_6P_2Si$: 1779.7502, Δ -0.3 ppm).

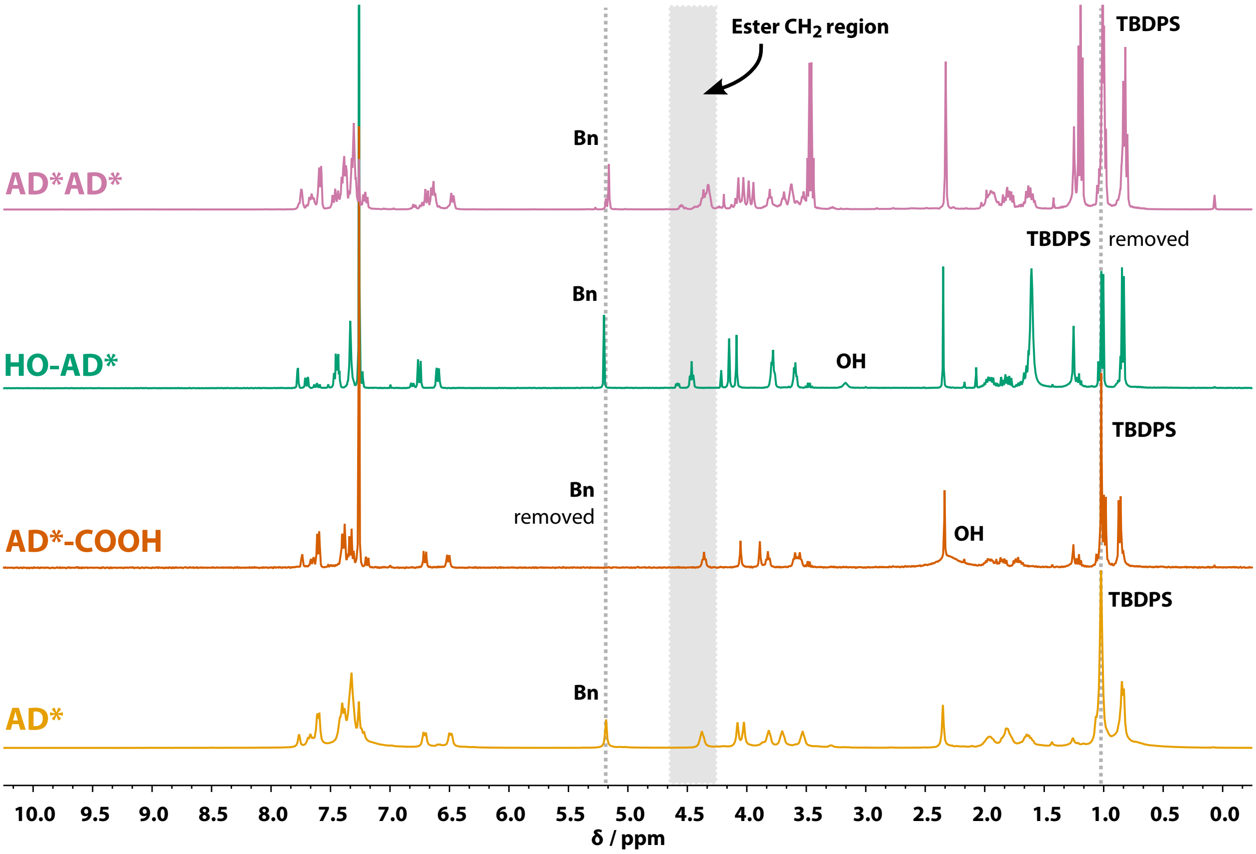


Fig. S1 Stack of ^{1}H NMR spectra ($CDCl_3$, 298 K) of intermediates **3** - **6**. Chemical shifts of the TBDPS and benzyl protecting groups, removed in the subsequent steps, are highlighted with dashed lines. Ester methylene region, where a new peak appears after EDC coupling, is highlighted in grey.

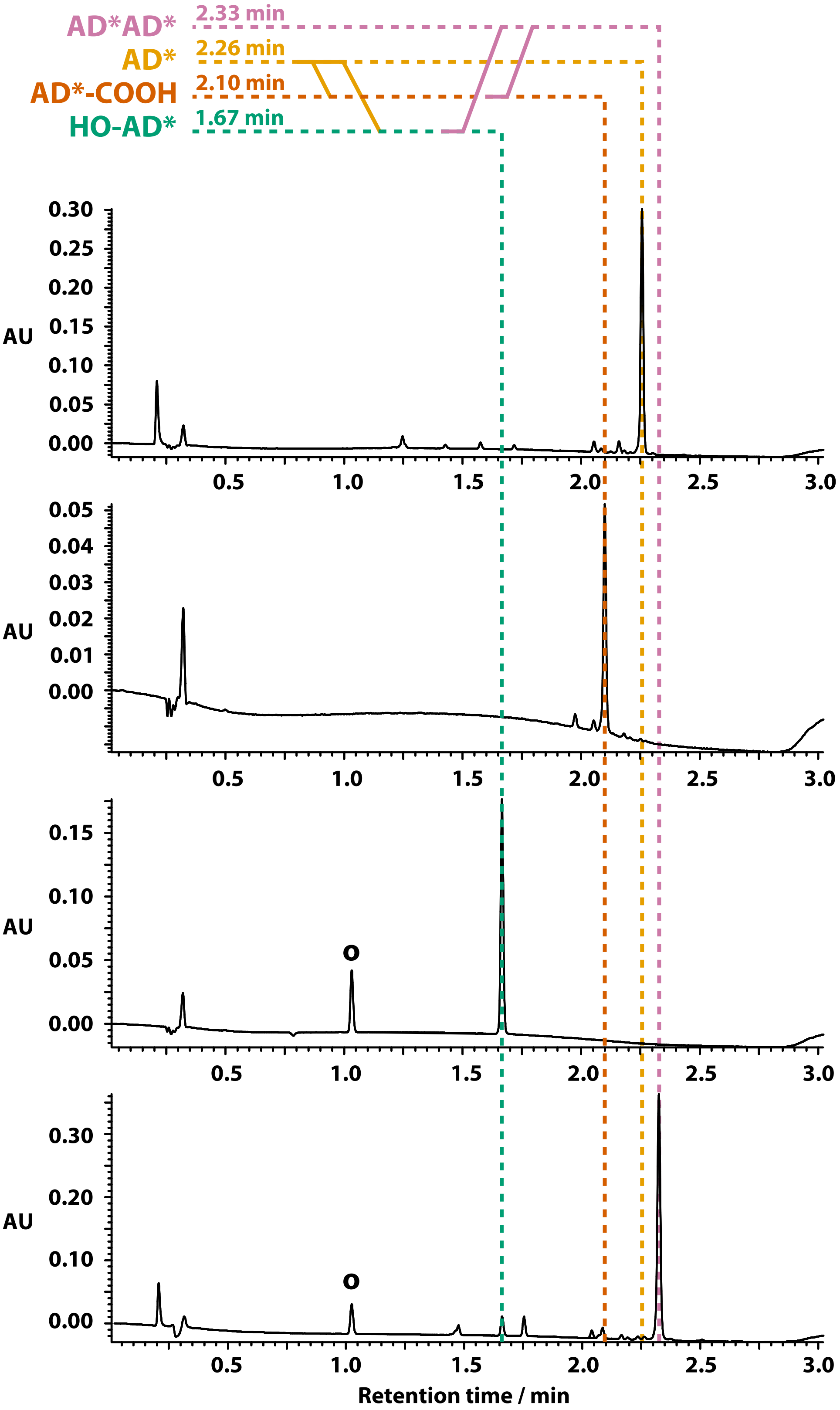


Fig. S2 LC-MS traces (monitored by UV at 270 nm) of the reactions converting **AD*** into **AD*AD***.

A portion of intermediate **6** (35 mg, 0.02 mmol) and ammonium acetate (38.5 mg, 0.5 mmol) were dissolved in a mixture of methanol (1 ml) and water (0.25 ml). The reaction mixture was stirred under nitrogen atmosphere overnight. After completion, the mixture was partitioned between water (20 ml) and ethyl acetate (10 ml). The layers were separated and the aqueous layer was subsequently washed with ethyl acetate (3 × 10 ml). The combined organic extracts were washed with brine (10 ml). The washed solution was dried with anhydrous magnesium sulfate, filtered and concentrated under vacuum. The residue was purified using column chromatography (0-20% methanol in ethyl acetate) to yield to yield compound **4.17** as a colourless oil (13 mg, 0.01 mmol, 38%).

^{1}H NMR (500 MHz, $CDCl_3$) δ 10.43 (s, 1H), 9.38 (s, 1H), 7.66 – 7.53 (m, 8H), 7.47 – 7.24 (m, 19H), 7.16 (d, J = 8.5 Hz, 1H), 7.10 (d, J = 8.5 Hz, 1H), 6.66 (d, J = 8.5 Hz, 2H), 6.59 – 6.52 (m, 4H), 6.41 (d, J = 6.8 Hz, 2H), 5.15 (s, 2H), 4.36 (t, J = 5.1 Hz, 2H), 4.29 (t, J = 5.6 Hz, 2H), 4.18 (t, J = 5.4 Hz, 2H), 4.02 (s, 2H), 3.98 (s, 2H), 3.90 (s, 2H), 3.80 (t, J = 5.5 Hz, 2H), 3.80 (s, 2H), 3.68 (t, J = 5.1 Hz, 2H), 3.56 (t, J = 5.6 Hz, 2H), 3.52 (t, J = 5.5 Hz, 2H), 3.46 (t, J = 5.4 Hz, 2H), 2.08 – 1.49 (m, 12H), 1.01 (s, 9H), 0.99 (d, J = 6.6 Hz, 6H), 0.97 (d, J = 6.6 Hz, 6H), 0.83 (d, J = 6.6 Hz, 6H), 0.80 (d, J = 6.6 Hz, 6H) ppm.

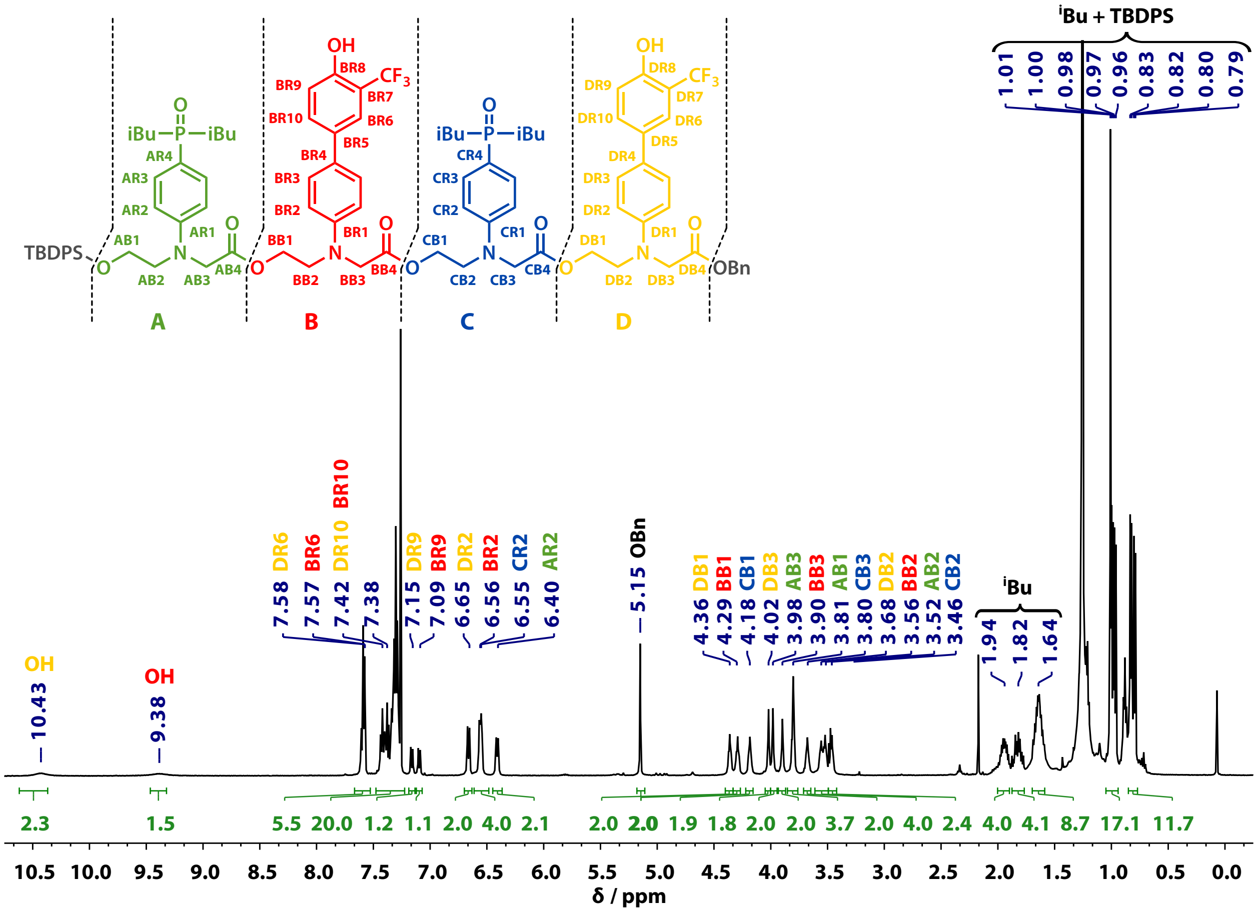
^{13}C NMR (126 MHz, $CDCl_3$) δ 170.7, 170.6, 170.3, 170.1, 155.1, 154.5, 150.1 (d, J = 2 Hz), 149.8 (d, J = 2 Hz), 146.7, 146.4, 135.5, 135.4, 133.0, 131.9 (d, J = 10 Hz), 131.8 (d, J = 10 Hz), 131.5, 131.0, 130.7, 130.4, 129.8, 129.6, 129.5, 128.6, 128.4, 128.3, 127.7, 127.5, 127.3, 124.2 (q, J = 273.1 Hz), 124.2 (q, J = 273.1 Hz), 124.1 (d, J = 5 Hz), 124.1 (d, J = 5 Hz), 119.5 (d, J = 101 Hz), 118.2 (d, J = 102 Hz), 118.0 (d, J = 5 Hz), 117.0 (q, J = 30 Hz), 116.7 (q, J = 30 Hz), 112.6, 112.4, 111.7 (d, J = 12 Hz), 111.5 (d, J = 12 Hz), 66.9, 62.3, 62.0, 61.3, 53.4, 52.5, 52.3, 52.1, 51.9, 50.1, 50.0, 49.8, 39.7 (d, J = 68 Hz), 39.7 (d, J = 68 Hz), 29.7, 24.7 (d, J = 9 Hz), 24.7 (d, J = 9 Hz), 24.5 (d, J = 8 Hz), 24.5 (d, J = 8 Hz), 23.4 (d, J = 4 Hz), 19.0 ppm.

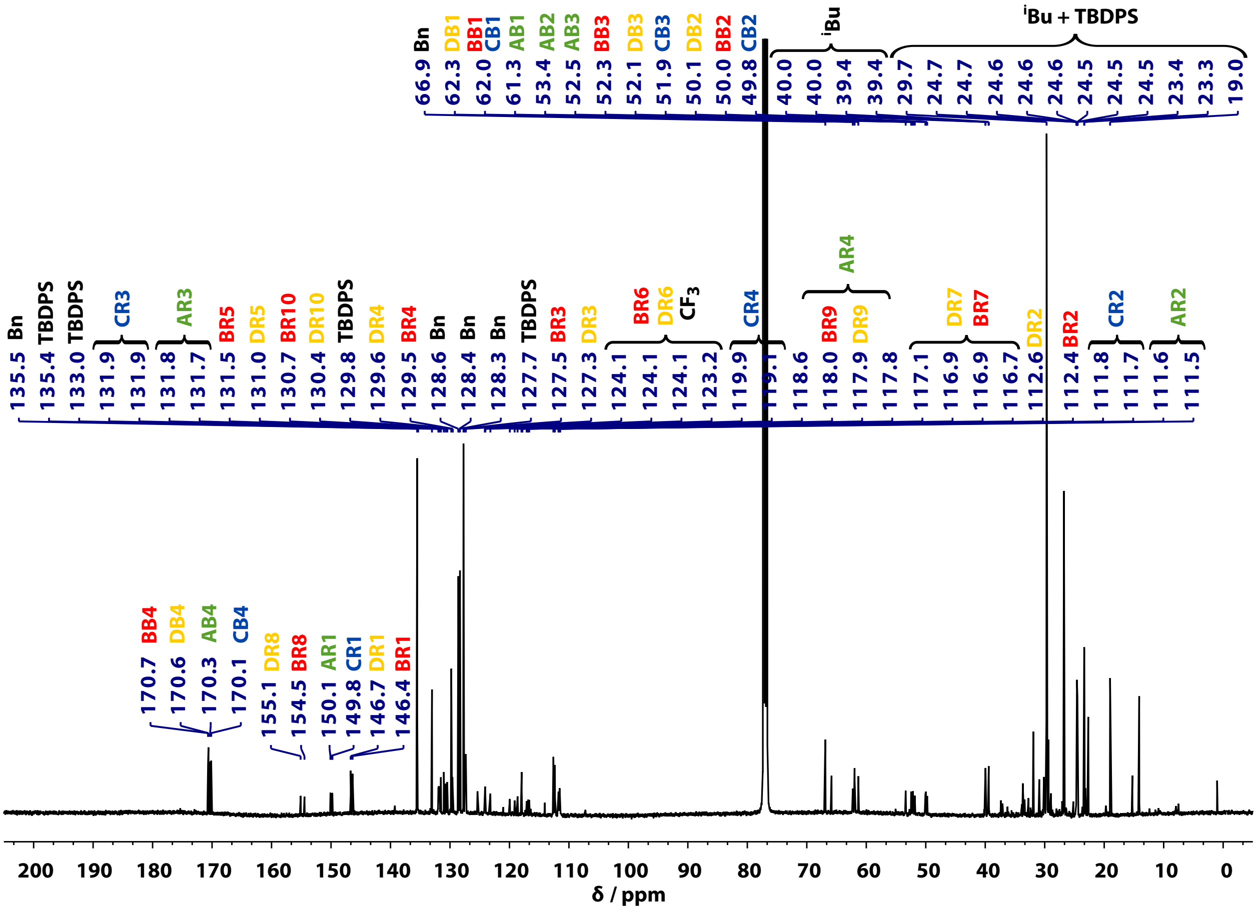
^{19}F NMR (376 MHz, $CDCl_3$) δ -61.7, -61.9 ppm.

^{31}P NMR(162 MHz, $CDCl_3$) δ 41.9, 41.1 ppm.

FT-IR (neat): 2955, 2918, 2850, 1740, 1612, 1598 cm^{-1}.

HR-MS (ESI): m/z = 1695.7255 $[M+H]^+$ (calcd. for $C_{93}H_{111}O_{13}N_4F_6P_2Si$: 1695.7291, Δ -2.1 ppm).


TBDPS
O
AB1
AB2
N
AB3
AB4
O
AR1
AR2
AR3
AR4
iBu—P—iBu
O
A
O
BB1
BB2
N
BB3
BB4
O
BR1
BR2
BR3
BR4
BR5
BR6
BR7
BR8
BR9
BR10
OH
CF3
B
O
CB1
CB2
N
CB3
CB4
O
CR1
CR2
CR3
CR4
iBu—P—iBu
O
C
O
DB1
DB2
N
DB3
DB4
O
OBn
DR1
DR2
DR3
DR4
DR5
DR6
DR7
DR8
DR9
DR10
OH
CF3
D
OH
10.43
OH
9.38
DR6 7.58
BR6 7.57
DR10 7.42
BR10 7.38
DR9 7.15
BR9 7.09
DR2 6.65
BR2 6.56
CR2 6.55
AR2 6.40
OBn 5.15
DB1 4.36
BB1 4.29
CB1 4.18
DB3 4.02
AB3 3.98
BB3 3.90
AB1 3.81
CB3 3.80
DB2 3.68
BB2 3.56
AB2 3.52
CB2 3.46
iBu
1.94
1.82
1.64
iBu + TBDPS
1.01
1.00
0.98
0.97
0.96
0.83
0.82
0.80
0.79
2.3
1.5
5.5
20.0
1.2
1.1
2.0
4.0
2.1
2.0
2.0
1.9
1.8
2.0
2.0
3.7
2.0
4.0
2.4
4.0
4.1
8.7
17.1
11.7
10.5
10.0
9.5
9.0
8.5
8.0
7.5
7.0
6.5
6.0
5.5
5.0
4.5
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
δ / ppm


Bn 66.9
DB1 62.3
BB1 62.0
CB1 61.3
AB1 53.4
AB2 52.5
AB3 52.3
BB3 52.1
DB3 51.9
CB3 50.1
DB2 50.0
BB2 49.8
CB2
iBu
40.0
40.0
39.4
39.4
iBu + TBDPS
29.7
24.7
24.7
24.6
24.6
24.6
24.5
24.5
24.5
23.4
23.3
19.0
Bn 135.5
TBDPS 135.4
TBDPS 133.0
CR3
131.9
131.9
AR3
131.8
131.7
BR5 131.5
DR5 131.0
BR10 130.7
DR10 130.4
TBDPS 129.8
DR4 129.6
BR4 129.5
Bn 128.6
Bn 128.4
Bn 128.3
TBDPS 127.7
BR3 127.5
DR3 127.3
BR6
DR6
CF3
124.1
124.1
124.1
123.2
CR4
119.9
119.1
AR4
118.6
BR9 118.0
DR9 117.9
117.8
DR7
BR7
117.1
116.9
116.9
116.7
DR2 112.6
BR2 112.4
CR2
111.8
111.7
AR2
111.6
111.5
BB4 170.7
DB4 170.6
AB4 170.3
CB4 170.1
DR8 155.1
BR8 154.5
AR1 150.1
CR1 149.8
DR1 146.7
BR1 146.4
200
190
180
170
160
150
140
130
120
110
100
90
80
70
60
50
40
30
20
10
0
δ / ppm

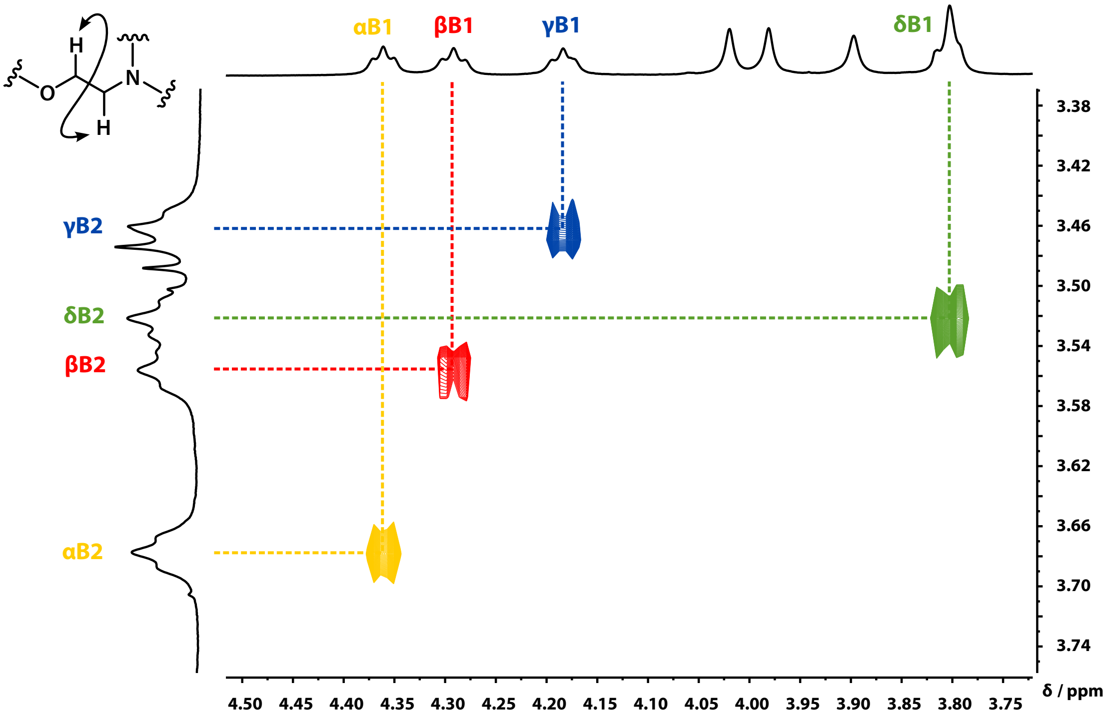


Fig. S3 Aliphatic region of the ^{1}H-^{1}H DQF COSY (500 MHz, $CDCl_3$, 298 K, *ca.* 5 mM) spectrum of the **ADAD** 4-mer. Cross-peaks corresponding to different residues are distinguished with colours and Greek letters. The substructure shows the highlighted correlations.

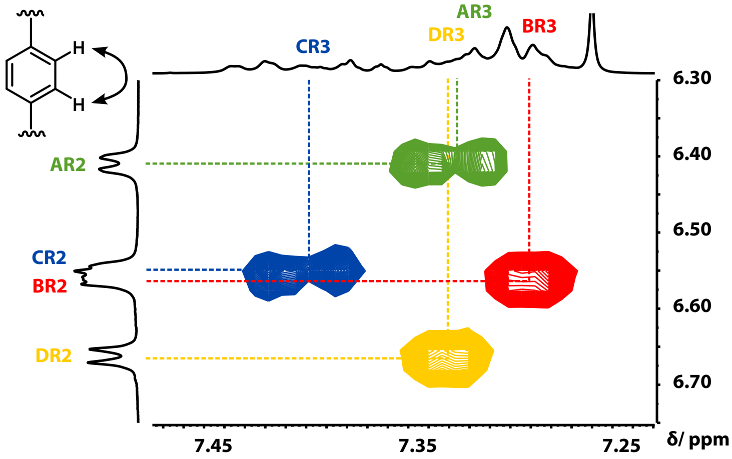


Fig. S4 Aromatic region of the ^{1}H-^{1}H DQF COSY (500 MHz, $CDCl_3$, 298 K, *ca.* 5 mM) spectrum of the **ADAD** 4-mer. The substructure shows the highlighted correlations between the signals on the aromatic signals.

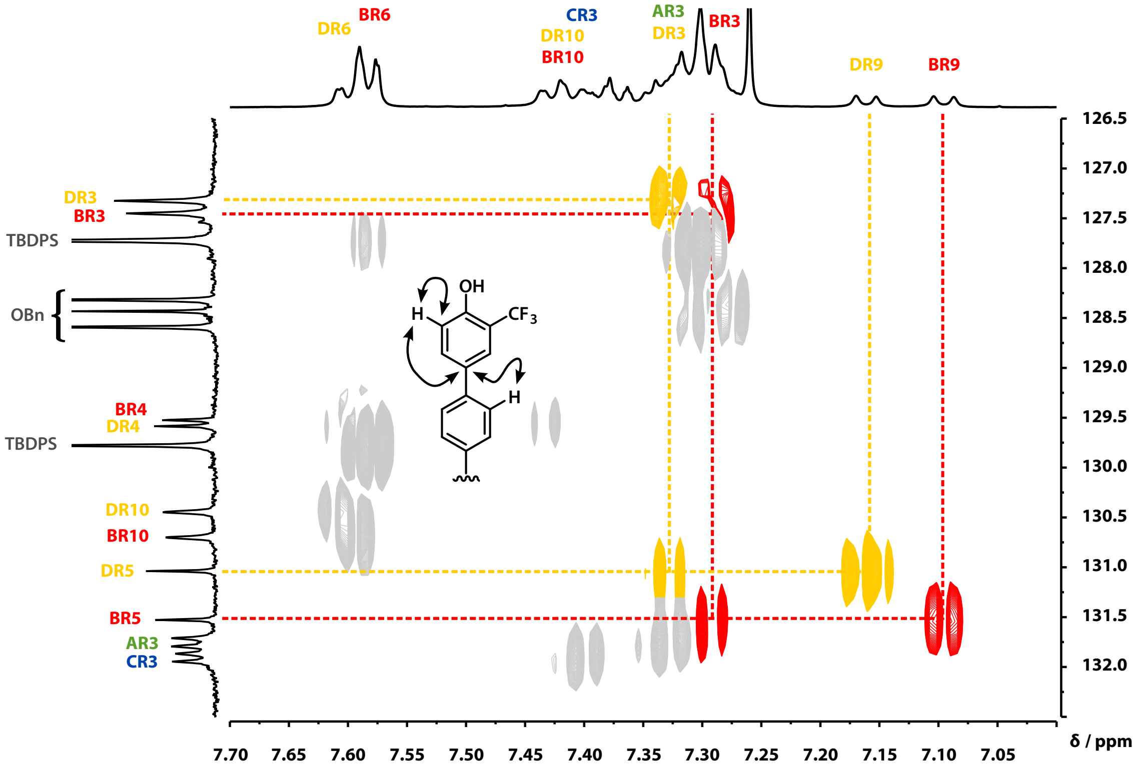


Fig. S5 Part of the ^{1}H-^{13}C HMBC (500 MHz, $CDCl_3$, 298 K, *ca.* 5 mM) spectrum of the **ADAD** 4-mer. The substructure shows the highlighted correlations between the two separate aromatic systems of the donor residues.

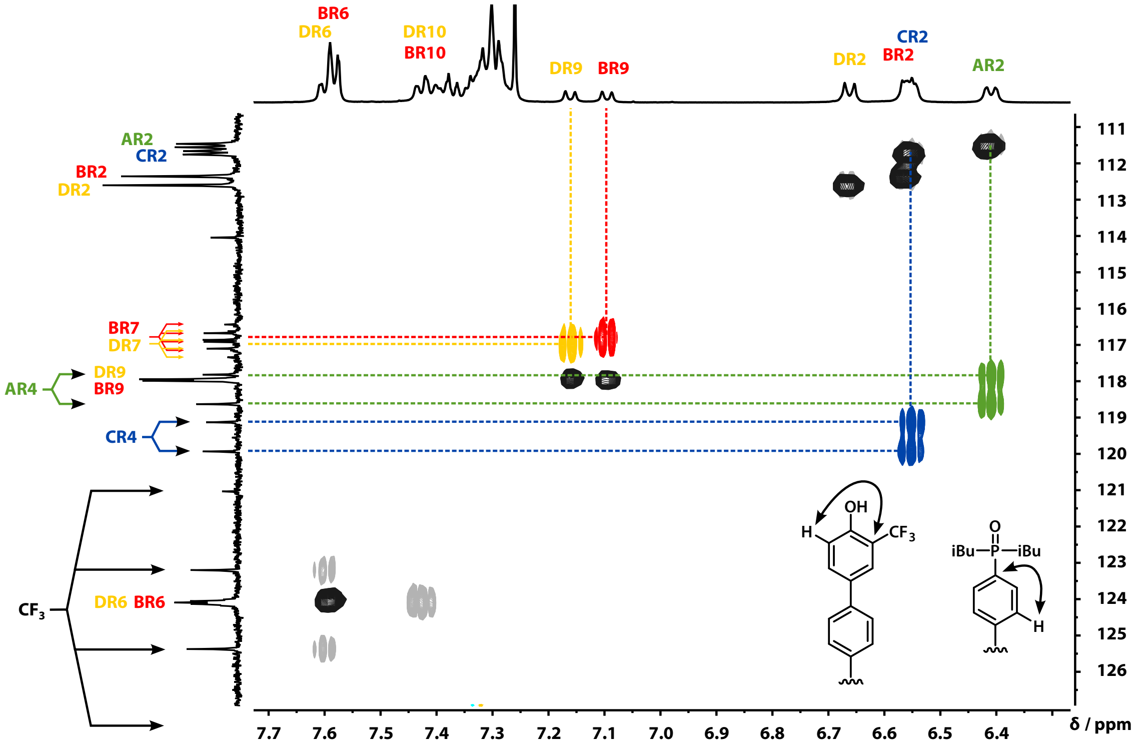


Fig. S6 Overlay of the aromatic regions of the ^{1}H-^{13}C HMBC (500 MHz, $CDCl_3$, 298 K, *ca.* 5 mM, light grey) and ^{1}H-^{13}C HSQC (500 MHz, $CDCl_3$, 298 K, *ca.* 5 mM, dark grey) spectra of the **ADAD** 4-mer. The substructure shows the highlighted correlations between the signals on the recognition units.

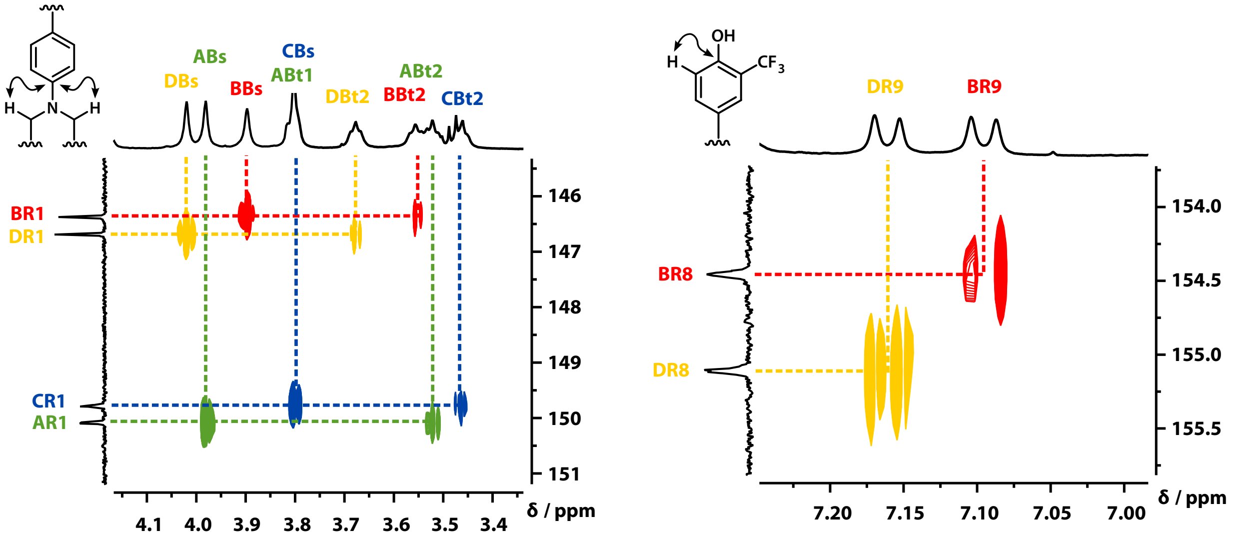


Fig. S7 Aromatic regions of the ^{1}H-^{13}C HMBC (500 MHz, CDCl$_3$, 298 K, *ca.* 5 mM) spectrum of the **ADAD** 4-mer. The substructure shows the highlighted correlations between the signals on the recognition units.

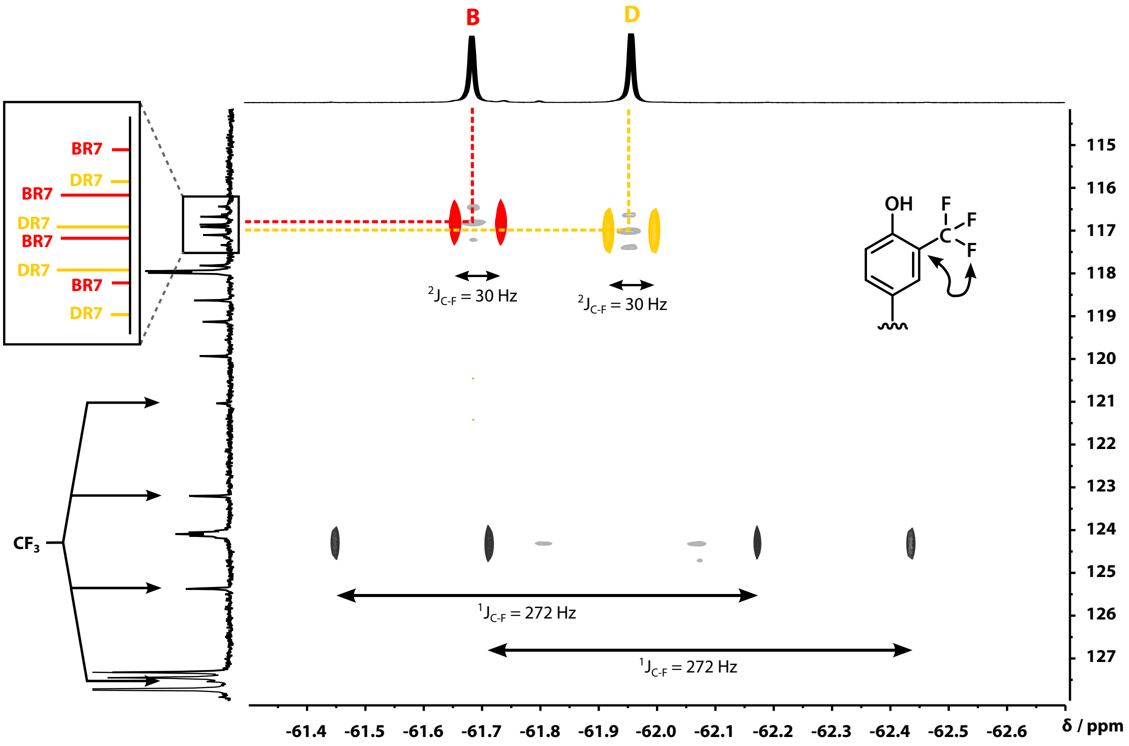


Fig. S8 Overlay of the ^{19}F-^{13}C HMBC (400 MHz, CDCl$_3$, 298 K, *ca.* 5 mM, dark grey) and ^{19}F-^{13}C HSQC (CDCl$_3$, 298 K, *ca.* 5 mM, light grey) spectra of the **ADAD** 4-mer. The substructure shows the highlighted correlation between the *ipso* carbon and the fluorine nucleus. One- and two-bond coupling constants are also shown.

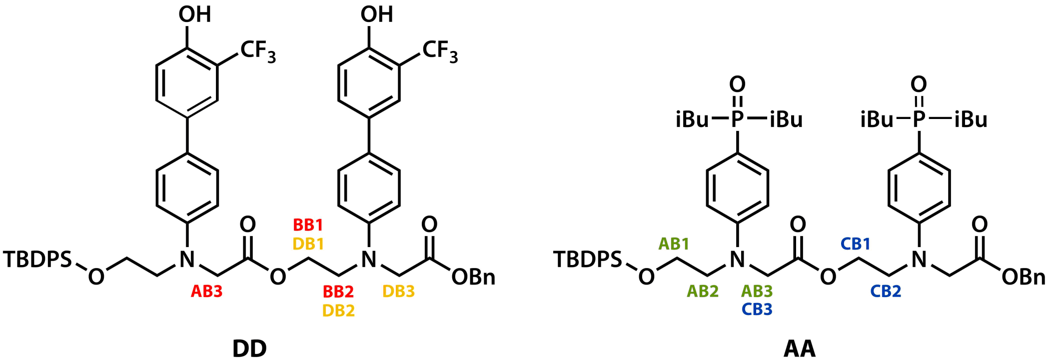


Fig. S9 Structures of **AA** and **DD** which cannot form hydrogen bonds and were used as the reference compounds for the backbone chemical shifts. Labelling corresponds to the labels of the equivalent **ADAD** residues. Unlabelled protons are not equivalent to any **ADAD** residues.

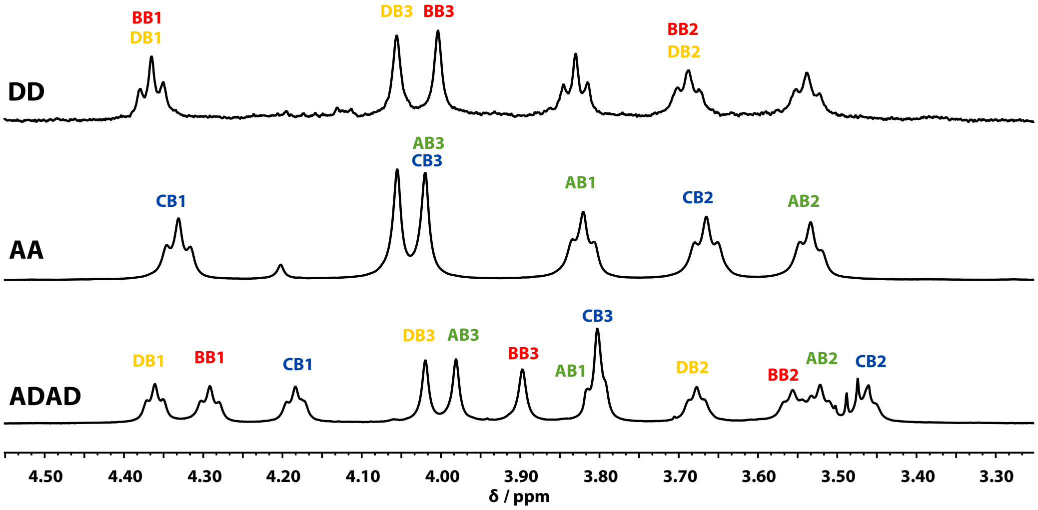


Fig. S10 Aliphatic region of the ^{1}H NMR ($CDCl_3$, 298 K) spectra of **DD** (400 MHz, *ca.* 5 mM), **AA** (400 MHz, *ca.* 1 mM), and **ADAD** (500 MHz, *ca.* 5 mM). Signals of the 2-mers are labelled using the corresponding **ADAD** labels. Unlabelled signals are not equivalent to any **ADAD** residues.

3 Melting experiments: Further Details

Two-State Melting Equilibrium

The resulting best fits of the two-state melting isochore to the variable temperature ^{19}F NMR data are shown in Fig. S11 (a). The ^{19}F NMR chemical shift of the outer **ADAD** donor residue in the fully denatured state (−57.7 ppm) is not consistent with the values from the dilution experiments, where the chemical shift for the monomeric **AD** and the inner residues of monomeric **ADAD** was around −59 ppm. The chemical shifts of all donor residues in fully assembled dimers of both **AD** and **ADAD** converged to the same value of −60.8 ppm, which is consistent with the dilution experiments. Therefore, we decided to fix the values of the complexation-induced limiting chemical shifts in **ADAD** to be equal to those of **AD**, but the resulting best-fit isochores (see Figure Fig. S11 (b)) did not model the experimental data well. These results suggest that the simple two-state melting equilibrium does not explain the behaviour of **ADAD**.

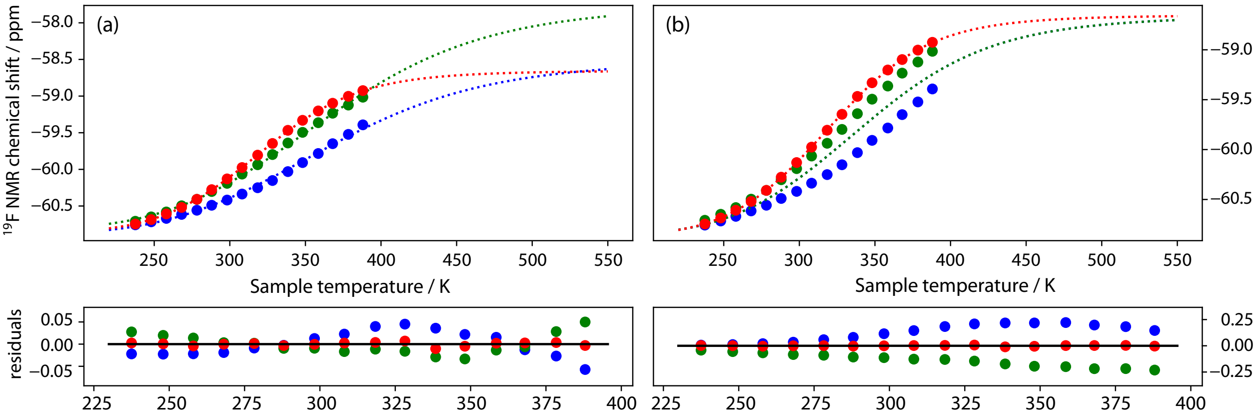


Fig. S11 Fitting of two-state equilibrium isochores to the ^{19}F NMR melting experiments of **AD** (red) and **ADAD** (outer ^{19}F green, inner ^{19}F blue) in TCE. (a) Four-parameters model with no constraints. (b) Monomer and dimer **ADAD** chemical shifts constrained to those of **AD**.

Three-State Melting Equilibrium

We assumed that the melting of the **ADAD** dimer proceeds through a highly populated intermediate **I**, giving a three-state equilibrium:

$$\mathbf{ADAD{\cdot}ADAD} \xrightleftharpoons{K_1(T,\Delta H_1,T_{\mathrm{m},1})} 2\,\mathbf{I} \tag{1}$$

$$\mathbf{I} \xrightleftharpoons{K_2(T,\Delta H_2,T_{\mathrm{m},2})} \mathbf{ADAD} \tag{2}$$

The chemical shifts of the fully bound and fully denatured strands were found to be consistent across all species when no constraints were imposed on the model (see Figure Fig. S12 (a)). In order to reduce the number of model parameters, we fixed the values of the limiting complexation-induced chemical shifts in **ADAD** to be equal to those of **AD**. The resulting best-fit isochores are shown in Figure Fig. S12 (b) and the thermodynamic fitting parameters are shown in Table S1. Melting temperatures for the two processes agree with the trends observed in the data. The enthalpy of melting for the intermolecular process ΔH_1 was found to be comparable with the value for the dissociation of **AD**, while the observed enthalpy of unfolding ΔH_2 was unexpectedly large for breaking of one hydrogen bond.

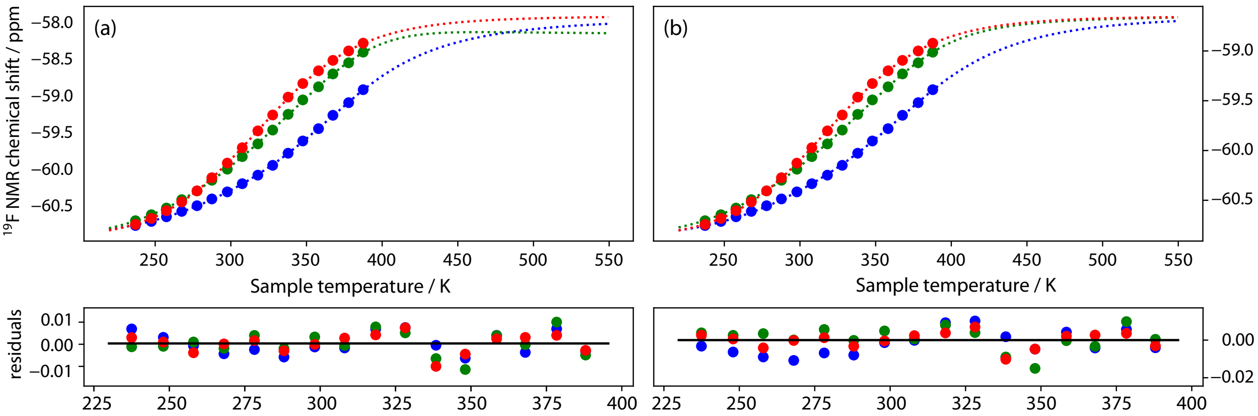


Fig. S12 Fitting of three-state equilibrium isochores to the ^{19}F NMR melting experiments of **AD** (red) and **ADAD** (outer ^{19}F green, inner ^{19}F blue) in TCE. (a) Seven-parameters model with no constraints. (b) Monomer and dimer **ADAD** chemical shifts constrained to be equal to those of **AD**.

Table S1 Thermodynamic parameters from NMR melting experiments on **ADAD** in TCE-d$_2$ at 7.4 mM.

Complex	δ_{kiss} / ppm	δ_{loop} / ppm	δ_{strand} / ppm	$T_{\text{m},1}$ / K	$T_{\text{m},2}$ / K	ΔH_1 / kJ mol^{-1}	ΔH_2 / kJ mol^{-1}
AD·AD	−60.9	—	−58.7	321	—	38.8	—
ADAD·ADAD (inner)	−60.9[§]	−58.9	−58.7[§]	324	404	31.4	37.1
ADAD·ADAD (outer)	−60.9[§]	−59.6	−58.7[§]				

[§] Parameter values fixed in the non-linear least-squares fit to be equal to the values for **AD**.

4 Dimerisation Isotherm Derivation and Implementation

Equilibrium constant $K_{\mathrm{M \cdot M}}$ the dimerisation of a self-complementary monomer **M** to form **M·M** is:

$$2\,\mathbf{M} \xrightleftharpoons{K_{\mathrm{M \cdot M}}} \mathbf{M \cdot M} \tag{3}$$

$$K_{\mathrm{M \cdot M}} = \frac{[\mathbf{M \cdot M}]}{[\mathbf{M}]^2} \tag{4}$$

where [**M**] and [**M·M**], are the equilibrium concentrations of the free monomer and the dimer. Assuming that no other equilibria occur in the mixture, the total concentration of the dimerising molecule $[\mathbf{M}]_0$is simply:

$$[\mathbf{M}]_0 = [\mathbf{M}] + 2[\mathbf{M \cdot M}] \tag{5}$$

Hence, the equilibrium constant $K_{\mathrm{M \cdot M}}$ can alternatively expressed as:

$$K_{\mathrm{M \cdot M}} = \frac{[\mathbf{M \cdot M}]}{\left([\mathbf{H}]_0 - 2[\mathbf{M \cdot M}]\right)^2} \tag{6}$$

which can be easily rearranged to give a quadratic in [**M·M**]:

$$[\mathbf{M \cdot M}]^2 - [\mathbf{M \cdot M}]\left([\mathbf{M}]_0 + \frac{1}{4K_{\mathrm{M \cdot M}}}\right) + \frac{1}{4}[\mathbf{M}]_0^2 = 0 \tag{7}$$

Equation 7 has only one physically meaningful root:

$$[\mathbf{M \cdot M}] = \frac{1}{2}\left([\mathbf{M}]_0 + \frac{1}{4K_{\mathrm{M \cdot M}}}\right) - \frac{1}{2}\sqrt{\left([\mathbf{M}]_0 + \frac{1}{4K_{\mathrm{M \cdot M}}}\right)^2 - [\mathbf{M}]_0^2} \tag{8}$$

which gives the molar fraction of the dimer $\chi_{\mathrm{M \cdot M}}$ as:

$$\chi_{\mathrm{M \cdot M}} = \frac{2[\mathbf{M \cdot M}]}{[\mathbf{M}]_0} = 1 + \frac{1}{4[\mathbf{M}]_0 K_{\mathrm{M \cdot M}}} - \sqrt{\left(1 + \frac{1}{4[\mathbf{M}]_0 K_{\mathrm{M \cdot M}}}\right)^2 - 1} \tag{9}$$

In the fast-exchange regime of NMR spectroscopy, the observed chemical shift δ_{obs} is a weighted average of the signals arising from all species present in solution:

$$\delta_{\mathrm{obs}} = \delta_{\mathrm{M}} \frac{[\mathbf{M}]}{[\mathbf{M}]_0} + \delta_{\mathrm{M \cdot M}} \frac{2[\mathbf{M \cdot M}]}{[\mathbf{M}]_0} = \delta_{\mathrm{M}}(1 - \chi_{\mathrm{M \cdot M}}) + \delta_{\mathrm{M \cdot M}}\chi_{\mathrm{M \cdot M}} \tag{10}$$

where δ_{M} and $\delta_{\mathrm{M\cdot M}}$ are the chemical shifts corresponding to the free **M** and the fully bound **M·M** complex, respectively. A model defined by Equations (9) and (10) can be fitted to the NMR data with $K_{\mathrm{M\cdot M}}$, $\delta_{\mathrm{M\cdot M}}$, and δ_{M} as parameters, using **lmfit** package in Python.[2]

Values in the text are quoted as arithmetic means and the errors were estimated as 95% confidence intervals based on at least two repetitions.

```
# FITTING DIMERISATION ISOTHERM TO NMR DILUTION DATA
# Non-linear regression done using lmfit
# https://lmfit.github.io/lmfit-py/model.html

from lmfit import Parameters, Model

# Define model parameters with initial values
# add with tuples: (NAME VALUE VARY MIN  MAX  EXPR  BRUTE_STEP)

params = Parameters()
params.add_many ((’K’, 100, True, None, None, None, None),
                (’d_bound’, -70, True, None, None, None),
                (’d_free’, -50, True, None, None, None, None))

# Molar fraction from the two-state dimerisation equilibrium

def alpha(c, K):
    return (1 + 1/(4*K*c) - np.sqrt((1 + 1/(4*K*c))*(1 + 1/(4*K*c)) - 1))

# MODEL FUNCTION: Observed chemical shift (weighted average)

def d_obs(c, K, d_bound, d_free):
    return d_bound * alpha(c, K) + d_free * (1 - alpha(c,K))

# Fitting parameters to the data
# xvalues : dataframe containing total concentration (M)
# yvalues : dataframe containing observed chemical shifts (ppm)
# List of minimisation methods: https://lmfit.github.io/lmfit-py/fitting.html
# Default minimisation method=’leastsq’ is Levenberg-Marquardt algorithm
# Here by default use robust Nelder-Mead method
#       and then estimate confidence interval from L-M covariance matrix
# results : contains best fit parameters

model = Model(d_obs)
results = model.fit(yvalues, params, c=xvalues, method=’nelder’)
results2 = model.fit(yvalues, params=results.params, c=xvalues, method=’leastsq’)

# Print best fit parameters and fit statistics

print(results.fit_report() + ’\n’ + results2.fit_report())
```

4.1 AD Dimerisation Isotherm in Chloroform - Repetition 1

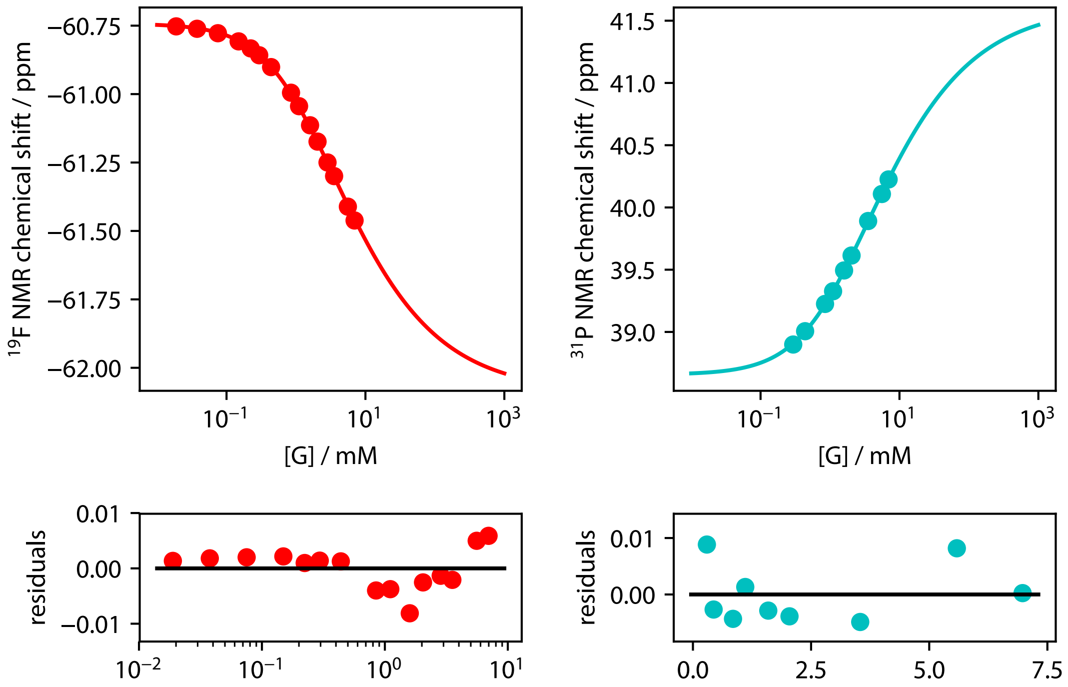


Fig. S13 Best fit dimerisation isotherm for **AD** chloroform-d at 298 K against the original data.

```
[[Fit Statistics]]
    # fitting method   = leastsq
    # function evals   = 7
    # data points      = 24
    # variables        = 5
    chi-square         = 4.0579e-04
    reduced chi-square = 2.1358e-05
    Akaike info crit   = -253.705261
    Bayesian info crit = -247.814992
[[Variables]]
    K:           170.751051 +/- 3.87320408 (2.27%) (init = 170.7511)
    d_freeP1:    38.6571557 +/- 0.00653719 (0.02%) (init = 38.65716)
    d_freeF1:   -60.7426747 +/- 0.00218359 (0.00%) (init = -60.74267)
    d_boundP1:   41.6223575 +/- 0.02004244 (0.05%) (init = 41.62236)
    d_boundF1:  -62.0914041 +/- 0.01230236 (0.02%) (init = -62.0914)
[[Correlations]] (unreported correlations are < 0.100)
    C(K, d_boundP1)          = -0.929
    C(K, d_boundF1)          =  0.894
    C(K, d_freeP1)           = -0.853
    C(d_boundP1, d_boundF1) = -0.830
    C(d_freeP1, d_boundF1)  = -0.762
    C(d_freeP1, d_boundP1)  =  0.641
    C(K, d_freeF1)           =  0.507
    C(d_freeF1, d_boundP1)  = -0.471
    C(d_freeP1, d_freeF1)   = -0.433
    C(d_freeF1, d_boundF1)  =  0.214
```

AD Dimerisation Isotherm in Chloroform - Repetition 2

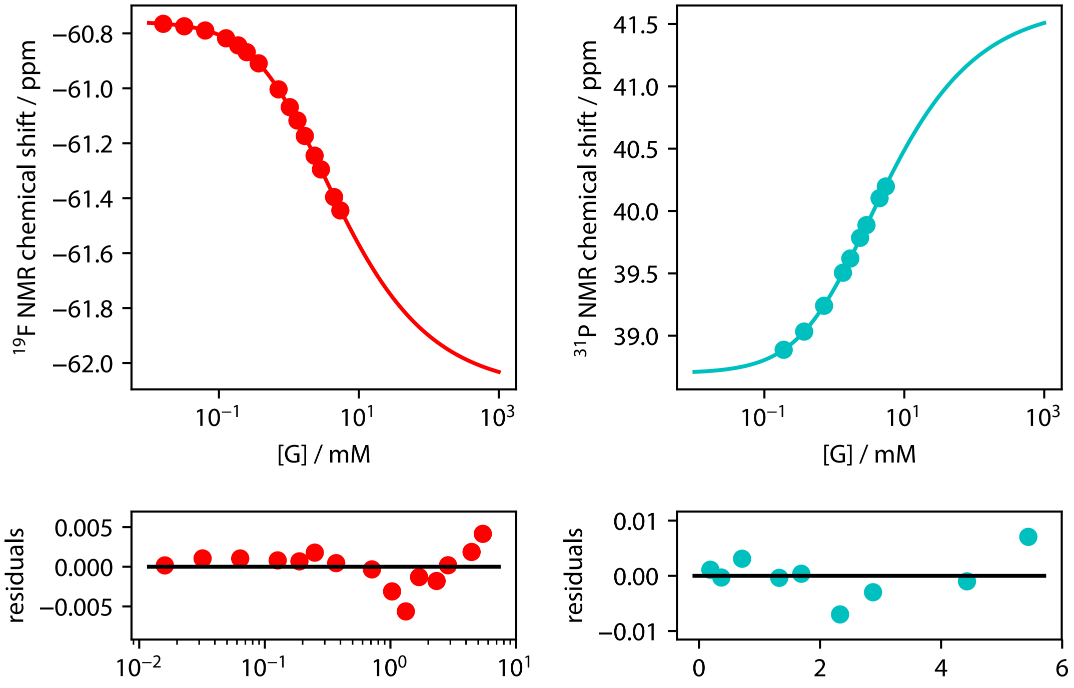


Fig. S14 Best fit dimerisation isotherm for **AD** chloroform-d at 298 K against the original data.

```
[[Fit Statistics]]
    # fitting method   = leastsq
    # function evals   = 7
    # data points      = 24
    # variables        = 5
    chi-square         = 1.9344e-04
    reduced chi-square = 1.0181e-05
    Akaike info crit   = -271.486881
    Bayesian info crit = -265.596612
[[Variables]]
    K:           194.406549 +/- 3.08573494 (1.59%) (init = 194.4065)
    d_freeP1:    38.6983112 +/- 0.00407552 (0.01%) (init = 38.69831)
    d_freeF1:   -60.7569883 +/- 0.00149070 (0.00%) (init = -60.75699)
    d_boundP1:   41.6546294 +/- 0.01531812 (0.04%) (init = 41.65463)
    d_boundF1:  -62.0986771 +/- 0.00891236 (0.01%) (init = -62.09868)
[[Correlations]] (unreported correlations are < 0.100)
    C(K, d_boundP1)          = -0.941
    C(K, d_boundF1)          =  0.895
    C(d_boundP1, d_boundF1) = -0.843
    C(K, d_freeP1)           = -0.805
    C(d_freeP1, d_boundF1)  = -0.721
    C(d_freeP1, d_boundP1)  =  0.599
    C(K, d_freeF1)           =  0.485
    C(d_freeF1, d_boundP1)  = -0.456
    C(d_freeP1, d_freeF1)   = -0.390
    C(d_freeF1, d_boundF1)  =  0.190
```

4.2 ADAD Dimerisation Isotherm in Chloroform - Repetition 1

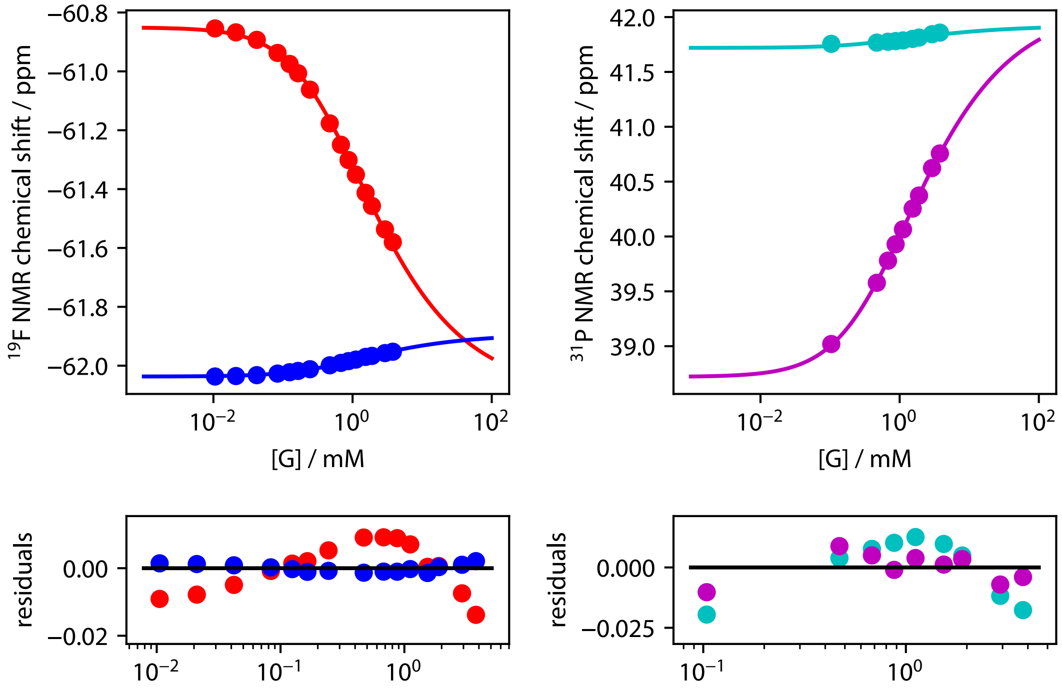


Fig. S15 Best fit dimerisation isotherm for **ADAD** chloroform-d at 298 K against the original data.

```
[[Fit Statistics]]
    # fitting method   = leastsq
    # function evals   = 11
    # data points      = 48
    # variables        = 9
    chi-square         = 0.00235235
    reduced chi-square = 6.0317e-05
    Akaike info crit   = -458.329984
    Bayesian info crit = -441.489175
[[Variables]]
    K:           488.092215 +/- 14.3160988 (2.93%) (init = 488.0922)
    d_freeP1:    41.7192611 +/- 0.00720932 (0.02%) (init = 41.71926)
    d_freeF1:   -60.8507328 +/- 0.00377699 (0.01%) (init = -60.85073)
    d_freeP2:    38.7203650 +/- 0.01242313 (0.03%) (init = 38.72036)
    d_freeF2:   -62.0368213 +/- 0.00335155 (0.01%) (init = -62.03682)
    d_boundP1:   41.9217981 +/- 0.01093760 (0.03%) (init = 41.9218)
    d_boundF1:  -62.0945051 +/- 0.01350667 (0.02%) (init = -62.09451)
    d_boundP2:   42.1200140 +/- 0.02564677 (0.06%) (init = 42.12001)
    d_boundF2:  -61.8924296 +/- 0.00788421 (0.01%) (init = -61.89243)
[[Correlations]] (unreported correlations are < 0.100)
    C(K, d_boundP2)          = -0.906
    C(K, d_boundF1)          =  0.817
    C(K, d_freeP2)           = -0.816
    C(d_freeP1, d_boundP1)   = -0.800
    C(d_boundF1, d_boundP2)  = -0.741
    C(d_freeP2, d_boundF1)   = -0.667
    C(d_freeF2, d_boundF2)   = -0.600
```

```
C(d_freeP2, d_boundP2)  =  0.539
C(K, d_freeF1)          =  0.464
C(d_freeF1, d_boundP2)  = -0.421
C(d_freeF1, d_freeP2)   = -0.379
C(K, d_boundF2)         = -0.163
C(d_boundP2, d_boundF2) =  0.147
C(d_boundF1, d_boundF2) = -0.133
C(d_freeP2, d_boundF2)  =  0.133
C(K, d_boundP1)         = -0.127
C(d_boundP1, d_boundP2) =  0.115
C(d_boundP1, d_boundF1) = -0.103
C(d_freeP2, d_boundP1)  =  0.103
```

ADAD Dimerisation Isotherm in Chloroform - Repetition 2

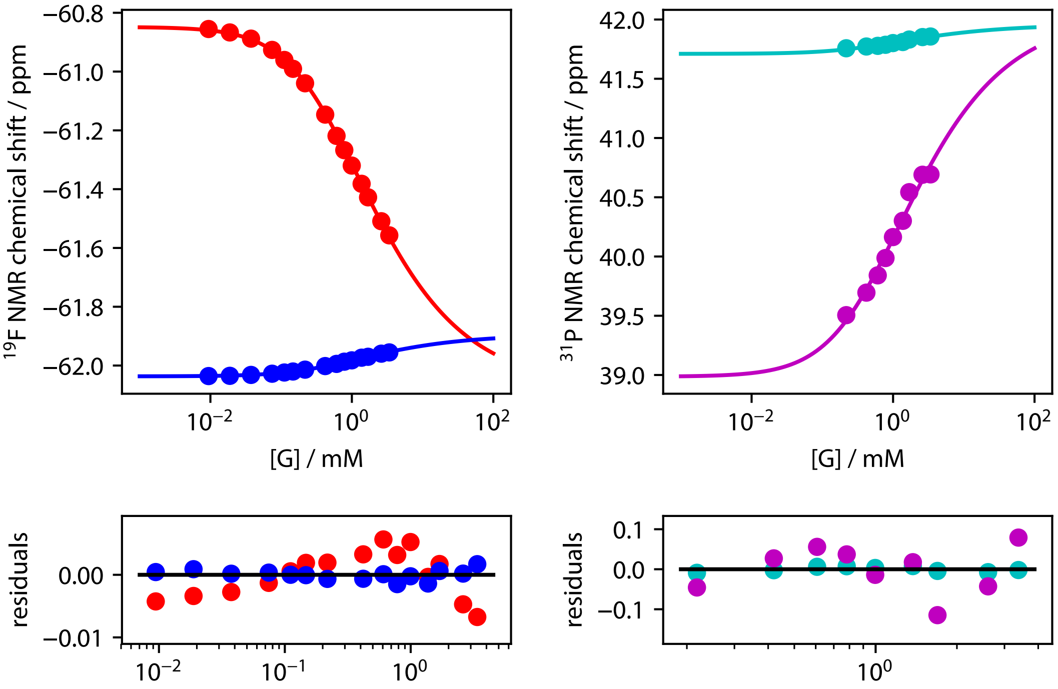


Fig. S16 Best fit dimerisation isotherm for **ADAD** chloroform-d at 298 K against the original data.

```
[[Fit Statistics]]
    # fitting method   = leastsq
    # function evals   = 31
    # data points      = 48
    # variables        = 9
    chi-square         = 0.02959961
    reduced chi-square = 7.5896e-04
    Akaike info crit   = -336.777371
    Bayesian info crit = -319.936562
[[Variables]]
    K:          495.830100 +/- 73.8830006 (14.90%) (init = 484.9741)
    d_freeP1:   41.7102884 +/- 0.02846814 (0.07%) (init = 41.71096)
    d_freeF1:  -60.8481837 +/- 0.01435614 (0.02%) (init = -60.84942)
    d_freeP2:   38.9861245 +/- 0.06450579 (0.17%) (init = 38.99468)
    d_freeF2:  -62.0366674 +/- 0.01177332 (0.02%) (init = -62.03652)
    d_boundP1:  41.9579335 +/- 0.04448221 (0.11%) (init = 41.95911)
    d_boundF1: -62.0758085 +/- 0.06439811 (0.10%) (init = -62.08438)
    d_boundP2:  42.0497488 +/- 0.10360027 (0.25%) (init = 42.06384)
    d_boundF2: -61.8941770 +/- 0.02943287 (0.05%) (init = -61.89318)
[[Correlations]] (unreported correlations are < 0.100)
    C(K, d_boundP2)          = -0.906
    C(K, d_freeP2)           = -0.900
    C(K, d_boundF1)          =  0.895
    C(d_boundF1, d_boundP2) = -0.811
    C(d_freeP2, d_boundF1)  = -0.806
    C(d_freeP1, d_boundP1)  = -0.803
    C(d_freeP2, d_boundP2)  =  0.658
```

```
C(d_freeF2, d_boundF2)  = -0.583
C(K, d_freeF1)          =  0.576
C(d_freeF1, d_boundP2)  = -0.522
C(d_freeF1, d_freeP2)   = -0.519
C(d_freeF1, d_boundF1)  =  0.290
C(K, d_boundF2)         = -0.227
C(d_boundP2, d_boundF2) =  0.206
C(d_freeP2, d_boundF2)  =  0.205
C(d_boundF1, d_boundF2) = -0.204
C(K, d_boundP1)         = -0.171
C(K, d_freeP1)          = -0.165
C(d_boundP1, d_boundP2) =  0.155
C(d_freeP2, d_boundP1)  =  0.154
C(d_boundP1, d_boundF1) = -0.153
C(d_freeP1, d_boundP2)  =  0.149
C(d_freeP1, d_freeP2)   =  0.148
C(d_freeP1, d_boundF1)  = -0.148
C(d_freeF1, d_boundF2)  = -0.131
```

4.3 AD Dimerisation Isotherm in TCE - Repetition 1

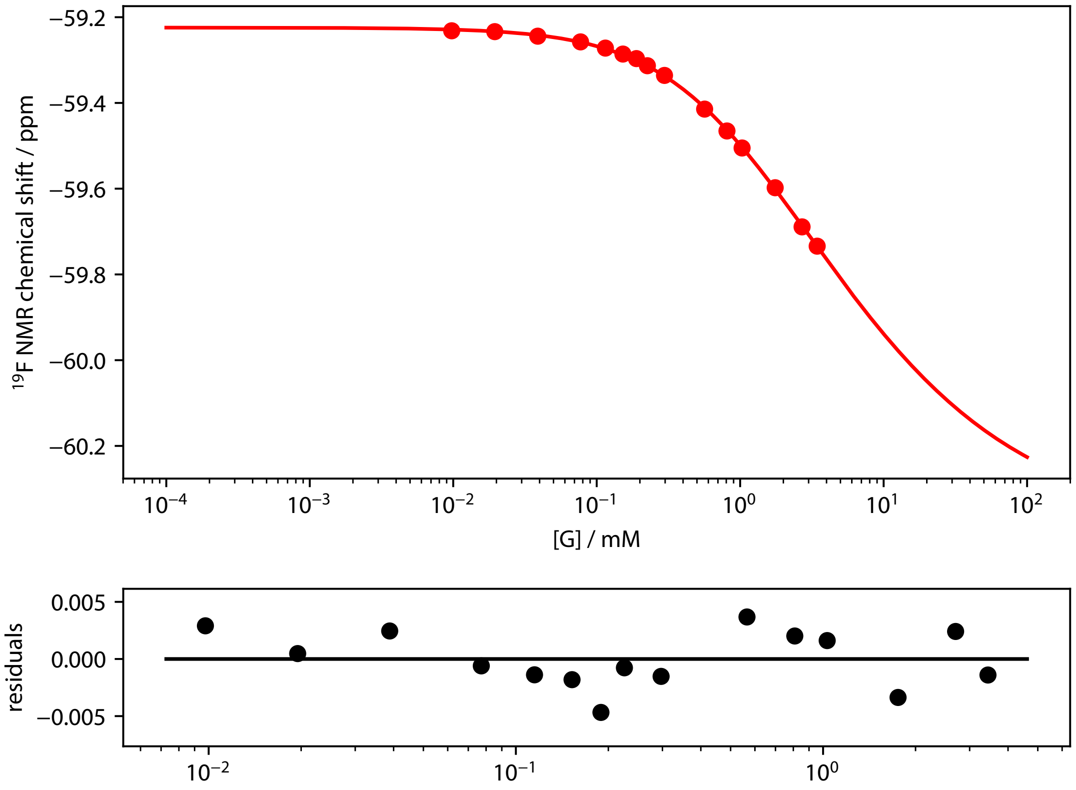


Fig. S17 Best fit dimerisation isotherm for **AD** TCE-d_2 at 298 K against the original data.

```
[[Model]]
    Model(d_obs)
[[Fit Statistics]]
    # fitting method   = leastsq
    # function evals   = 5
    # data points      = 15
    # variables        = 3
    chi-square         = 8.4285e-05
    reduced chi-square = 7.0237e-06
    Akaike info crit   = -175.340371
    Bayesian info crit = -173.216220
[[Variables]]
    K:         198.493145 +/- 6.40093251 (3.22%) (init = 198.4932)
    d_bound: -60.3982722 +/- 0.01701657 (0.03%) (init = -60.39827)
    d_free:  -59.2242828 +/- 0.00134973 (0.00%) (init = -59.22428)
[[Correlations]] (unreported correlations are < 0.100)
    C(K, d_bound)      =  0.968
    C(K, d_free)       =  0.688
    C(d_bound, d_free) =  0.558
```

AD Dimerisation Isotherm in TCE - Repetition 2

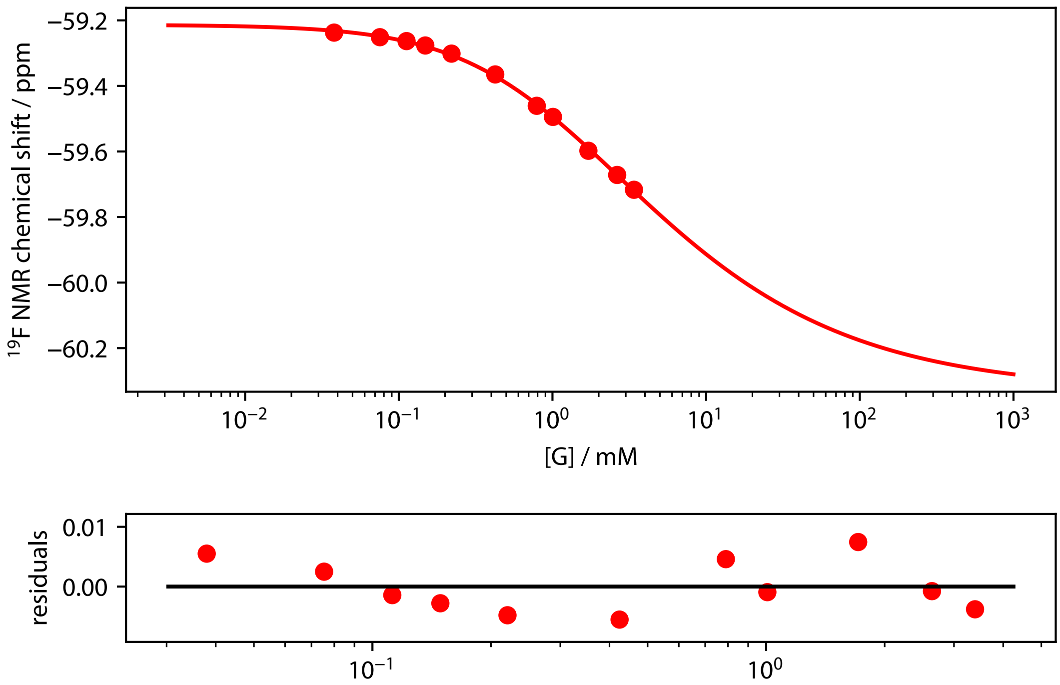


Fig. S18 Best fit dimerisation isotherm for **AD** TCE-d_2 at 298 K against the original data.

```
[[Fit Statistics]]
    # fitting method   = leastsq
    # function evals   = 5
    # data points      = 11
    # variables        = 3
    chi-square         = 1.9244e-04
    reduced chi-square = 2.4055e-05
    Akaike info crit   = -114.489807
    Bayesian info crit = -113.296122
[[Variables]]
    K:            224.029446 +/- 15.9937148 (7.14%) (init = 224.0292)
    d_freeF1:   -59.2135095 +/- 0.00355377 (0.01%) (init = -59.21351)
    d_boundF1: -60.3313485 +/- 0.03282292 (0.05%) (init = -60.33135)
[[Correlations]] (unreported correlations are < 0.100)
    C(K, d_boundF1)        =  0.969
    C(K, d_freeF1)         =  0.751
    C(d_freeF1, d_boundF1) =  0.622
```

4.4 ADAD Dimerisation Isotherm in TCE - Repetition 1

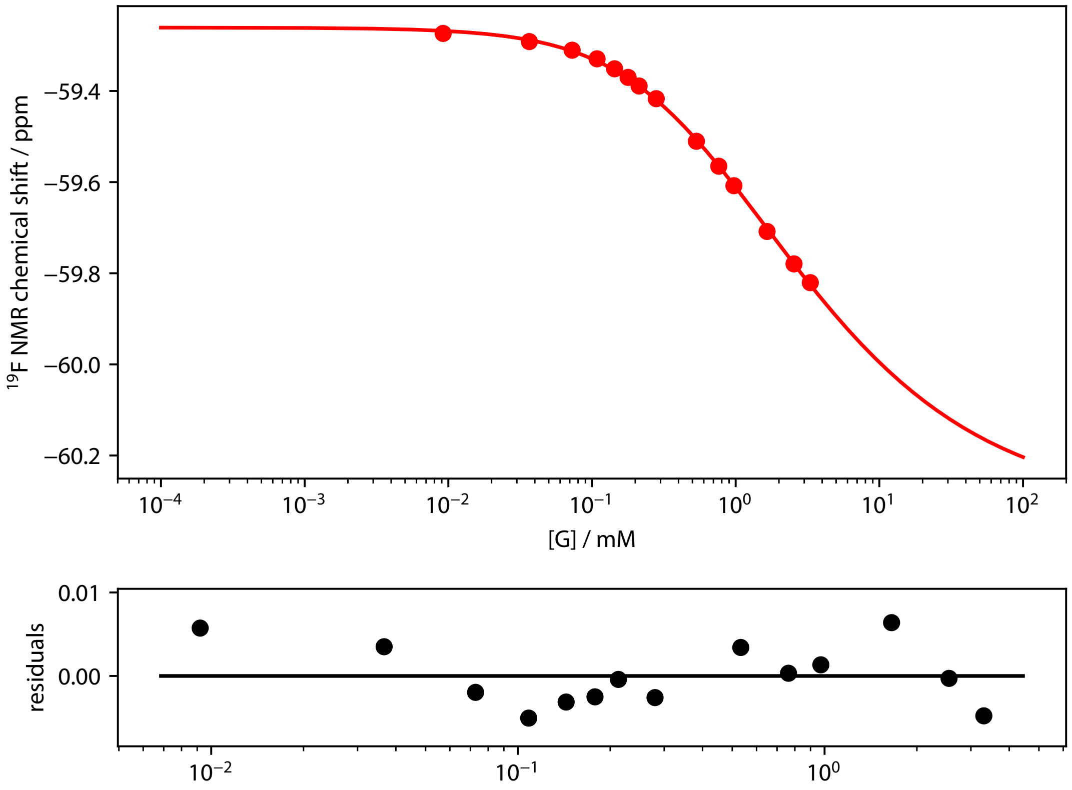


Fig. S19 Best fit dimerisation isotherm for **ADAD** TCE-d$_2$ at 298 K against the original data.

```
[[Model]]
    Model(d_obs)
[[Fit Statistics]]
    # fitting method   = leastsq
    # function evals   = 5
    # data points      = 14
    # variables        = 3
    chi-square         = 1.7440e-04
    reduced chi-square = 1.5855e-05
    Akaike info crit   = -152.104651
    Bayesian info crit = -150.187479
[[Variables]]
    K:         370.076582 +/- 15.6088765 (4.22%) (init = 370.0765)
    d_bound: -60.3197208 +/- 0.01598141 (0.03%) (init = -60.31972)
    d_free:  -59.2608510 +/- 0.00258315 (0.00%) (init = -59.26085)
[[Correlations]] (unreported correlations are < 0.100)
    C(K, d_bound)      =  0.948
    C(K, d_free)       =  0.760
    C(d_bound, d_free) =  0.592
```

ADAD Dimerisation Isotherm in TCE - Repetition 2

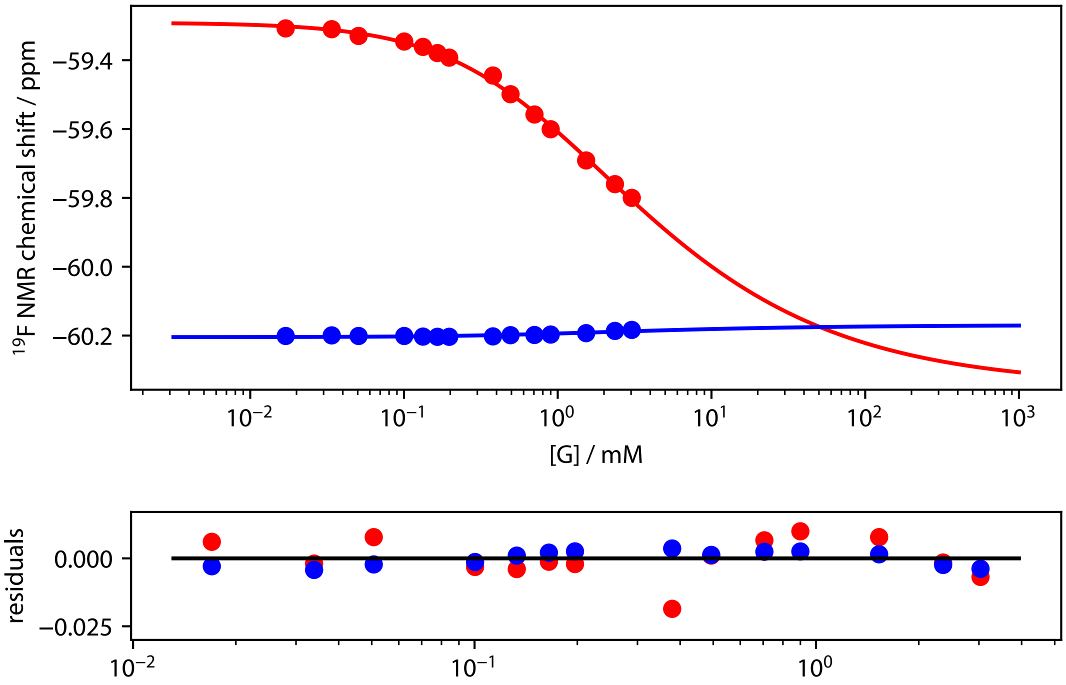


Fig. S20 Best fit dimerisation isotherm for **ADAD** TCE-d_2 at 298 K against the original data.

```
[[Fit Statistics]]
    # fitting method   = leastsq
    # function evals   = 38
    # data points      = 28
    # variables        = 5
    chi-square         = 8.2846e-04
    reduced chi-square = 3.6020e-05
    Akaike info crit   = -281.988120
    Bayesian info crit = -275.327098
[[Variables]]
    K:          306.798818 +/- 21.6241378 (7.05%) (init = 147.7316)
    d_freeF2:  -60.2044593 +/- 0.00247974 (0.00%) (init = -60.20388)
    d_freeF1:  -59.2908028 +/- 0.00363780 (0.01%) (init = -59.31508)
    d_boundF2: -60.1695750 +/- 0.00864069 (0.01%) (init = -60.15501)
    d_boundF1: -60.3484897 +/- 0.02968463 (0.05%) (init = -60.73865)
[[Correlations]] (unreported correlations are < 0.100)
    C(K, d_boundF1)         =  0.957
    C(K, d_freeF1)          =  0.732
    C(d_freeF2, d_boundF2)  = -0.620
    C(d_freeF1, d_boundF1)  =  0.577
    C(K, d_boundF2)         = -0.108
    C(d_boundF2, d_boundF1) = -0.104
```

5 Two-State Melting Isochore Derivation and Implementation

For a two-state melting equilibrium of a dimeric **M·M** into a self-complementary monomer **M**:

$$\mathbf{M{\cdot}M} \xrightleftharpoons{K_{\mathrm{M\cdot M}}(T, \varDelta H, T_{\mathrm{m}})} 2\mathbf{M} \tag{11}$$

the molar fractions for each species at a given temperature T are:

$$\chi_{\mathrm{M\cdot M}(T)} = \frac{2[\mathbf{M{\cdot}M}](T)}{[\mathbf{M}]_0} \tag{12}$$

$$\chi_{\mathrm{M}}(T) = \frac{[\mathbf{M}](T)}{[\mathbf{M}]_0} = 1 - \chi_{\mathrm{M\cdot M}}(T) \tag{13}$$

where $[\mathbf{M}]_0$ is the total concentration of **M** in all forms. The equilibrium constant for the melting reaction at temperature T can be written in terms of the molar fraction of the dimer as:

$$K(T) = \frac{[\mathbf{M}]^2(T)}{[\mathbf{M{\cdot}M}](T)} = \frac{2[\mathbf{M}]_0(1 - \chi_{\mathrm{M\cdot M}}(T))^2}{\chi_{\mathrm{M\cdot M}}(T)} \tag{14}$$

Differentiating the reaction isotherm equation ($\Delta G = -RT \ln K$) with respect to T^{-1}, under the assumption that ΔH and ΔS are independent of temperature, yields the van't Hoff equation:

$$\frac{\mathrm{d} \ln K(T)}{\mathrm{d}\,(1/T)} = -\frac{\Delta H}{R} \tag{15}$$

Integrating Equation (15) between a range of temperatures T_1 and T_2 gives:

$$\ln \frac{K(T_1)}{K(T_2)} = -\frac{\Delta H}{R} \left(\frac{1}{T_1} - \frac{1}{T_2} \right) \tag{16}$$

$$K(T_1) = K(T_2) \exp \left\{ -\frac{\Delta H}{R} \left(\frac{1}{T_1} - \frac{1}{T_2} \right) \right\} \tag{17}$$

At the melting temperature T_{m}, half of **M** exists as **M·M** and hence $\chi_{\mathrm{M\cdot M}}(T_{\mathrm{m}}) = \chi_{\mathrm{M}}(T_{\mathrm{m}}) = 0.5$:

$$K(T_{\mathrm{m}}) = \frac{2 \times (1 - 0.5)^2 \times [\mathbf{M}]_0}{0.5} = [\mathbf{M}]_0 \tag{18}$$

which provides the dimerisation melting isochore:

$$K(T) = \frac{1}{[\mathbf{M}]_0} \exp\left\{ -\frac{\Delta H}{R} \left(\frac{1}{T} - \frac{1}{T_\mathrm{m}} \right) \right\} \tag{19}$$

The observed chemical shift δ_obs is a weighted average for all species in fast exchange:

$$\delta_\mathrm{obs}(T) = \delta_\mathrm{M}\chi_\mathrm{M}(T) + \delta_\mathrm{M\cdot M}\chi_\mathrm{M\cdot M}(T) \tag{20}$$

which can be combined with Equation (13) to yield:

$$\delta_\mathrm{obs}(T) = \delta_\mathrm{M}(1 - \chi_\mathrm{M\cdot M}(T)) + \delta_\mathrm{M\cdot M}\chi_\mathrm{M\cdot M}(T) \tag{21}$$

Following standard dimerisation argument (see Subsection 4), $\chi_\mathrm{M\cdot M}(T)$ can be expressed as:[3]

$$\chi_\mathrm{M\cdot M} = \frac{2[\mathbf{M\cdot M}](T)}{[\mathbf{M}]_0} = 1 + \frac{K_\mathrm{M\cdot M}(T)}{4[\mathbf{M}]_0} - \sqrt{\left(1 + \frac{K_\mathrm{M\cdot M}(T)}{4[\mathbf{M}]_0}\right)^2 - 1} \tag{22}$$

A model defined by Equations (19), (21), and (22) can then fitted to the data obtained from a variable temperature ^{19}F experiment with ΔH, T_m, $\delta_\mathrm{M\cdot M}$, and δ_M as parameters, using **lmfit** package in Python.[2] Analysis was performed in Jupyter and the results were plotted using **matplotlib**.[4,5]

```
# FITTING TWO-STATE ISOCHORE TO VT NMR DATA
# Non-linear regression done using lmfit
# https://lmfit.github.io/lmfit-py/model.html

from lmfit import Parameters, Model
import numpy as np
import pandas as pd

# Define model parameters with initial values
# c : float containing total concentration (M)
# R : universal gas constant (J / mol K)

c = 0.00744
R = 8.314

params = Parameters()
params.add_many((’dH’, 38800, True, None, None, None, None),
                (’Tm’, 298, True, None, None, None, None),
                (’d_bound’, -60.860, True, None, None, None, None),
                (’d_free’, -58.653, True, None, None, None, None))

# Equilibrium constant from integrated van’t Hoff equation

def K(T, dH, Tm):
    return c * np.exp(-dH/R*(1/T-1/Tm))

# Molar fraction from two-state equilibrium constant

def alpha(T, dH, Tm):
    return (4*c + K(T, dH, Tm) - np.sqrt((4*c + K(T, dH, Tm))* \
                      (4*c + K(T, dH, Tm)) - 16*c*c))/(4*c)

# MODEL FUNCTION: Observed chemical shift (weighted average)

def d_obs(T, dH, Tm, d_bound, d_free):
    return alpha(T, dH, Tm)*d_bound + (1-alpha(T, dH, Tm))*d_free

# Fitting parameters to the data
# T : dataframe containing sample temperatures (K)
# d_F : dataframe containing observed chemical shifts (ppm)
# results : contains best fit parameters

model = Model(d_obs)
results = model.fit(d_F, params, T=T, method=’nelder’)
results2 = model.fit(d_F, results.params, T=T, method=’leastsq’)

print(results.fit_report() + ’\n’ + results2.fit_report())
```

5.1 AD Two-State Melting Isochore

```
[[Model]]
    Model(d_obs)
[[Fit Statistics]]
    # fitting method   = leastsq
    # function evals   = 6
    # data points      = 16
    # variables        = 4
    chi-square         = 2.6534e-04
    reduced chi-square = 2.2111e-05
    Akaike info crit   = -168.113666
    Bayesian info crit = -165.023311
[[Variables]]
    dH:       38863.2239 +/- 474.119538 (1.22%) (init = 38863.23)
    Tm:       321.141995 +/- 0.39350863 (0.12%) (init = 321.142)
    d_bound: -60.8600436 +/- 0.00690686 (0.01%) (init = -60.86004)
    d_free:  -58.6528140 +/- 0.01417487 (0.02%) (init = -58.65281)
[[Correlations]] (unreported correlations are < 0.100)
    C(dH, d_bound)     =  0.892
    C(dH, d_free)      = -0.890
    C(Tm, d_free)      =  0.816
    C(d_bound, d_free) = -0.692
    C(dH, Tm)          = -0.514
    C(Tm, d_bound)     = -0.186
```

5.2 ADAD Two-State Melting Isochore - No Constraints

```
[[Fit Statistics]]
    # fitting method   = leastsq
    # function evals   = 7
    # data points      = 32
    # variables        = 6
    chi-square         = 0.02004218
    reduced chi-square = 7.7085e-04
    Akaike info crit   = -224.020872
    Bayesian info crit = -215.226456
[[Variables]]
    dH:       -30948.6036 +/- 3012.73807 (9.73%) (init = -30948.6)
    Tm:        365.659008 +/- 16.0296604 (4.38%) (init = 365.659)
    d_bound1: -60.8846915 +/- 0.03514270 (0.06%) (init = -60.88469)
    d_free1:  -58.4917865 +/- 0.37601474 (0.64%) (init = -58.49179)
    d_bound2: -60.8178828 +/- 0.04410058 (0.07%) (init = -60.81788)
    d_free2:  -57.7287867 +/- 0.48485168 (0.84%) (init = -57.72879)
[[Correlations]] (unreported correlations are < 0.100)
    C(Tm, d_free2)        =  0.997
    C(Tm, d_free1)        =  0.996
    C(d_free1, d_free2)   =  0.995
    C(dH, d_free2)        =  0.949
    C(dH, d_free1)        =  0.948
    C(dH, Tm)             =  0.937
    C(dH, d_bound2)       = -0.918
    C(dH, d_bound1)       = -0.892
    C(d_bound1, d_bound2) =  0.887
    C(d_bound2, d_free2)  = -0.802
    C(d_free1, d_bound2)  = -0.789
    C(d_bound1, d_free1)  = -0.786
    C(d_bound1, d_free2)  = -0.768
    C(Tm, d_bound2)       = -0.767
    C(Tm, d_bound1)       = -0.745
```

5.3 ADAD Two-State Melting Isochore - Limiting Shifts Constrained

```
[[Fit Statistics]]
    # fitting method   = leastsq
    # function evals   = 3
    # data points      = 32
    # variables        = 2
    chi-square         = 0.71967752
    reduced chi-square = 0.02398925
    Akaike info crit   = -117.430015
    Bayesian info crit = -114.498543
[[Variables]]
    dH:       -33777.6504 +/- 2705.75829 (8.01%) (init = -33777.65)
    Tm:        342.835185 +/- 2.85285341 (0.83%) (init = 342.8352)
    d_bound1: -60.86 (fixed)
    d_free1:  -58.6528 (fixed)
    d_bound2: -60.86 (fixed)
    d_free2:  -58.6528 (fixed)
```

6 Three-State Melting Isochore Derivation and Implementation

Alternatively, melting might proceed through an intermediate **M***, giving a three-state equilibrium:

$$\mathbf{M{\cdot}M} \xrightleftharpoons{K_1(T, \Delta H_1, T_{\mathrm{m},1})} 2\,\mathbf{M^*} \tag{23}$$

$$\mathbf{M^*} \xrightleftharpoons{K_2(T, \Delta H_2, T_{\mathrm{m},2})} \mathbf{M} \tag{24}$$

where the molecular fractions of each species are:

$$\chi_{\mathrm{M{\cdot}M}}(T) = \frac{2[\mathbf{M{\cdot}M}](T)}{\mathbf{M}_0} \tag{25}$$

$$\chi_{\mathrm{M^*}}(T) = \frac{[\mathbf{M^*}](T)}{\mathbf{M}_0} \tag{26}$$

$$\chi_{\mathrm{M}}(T) = \frac{[\mathbf{M}](T)}{\mathbf{M}_0} \tag{27}$$

Equilibrium constants for the above equilibria can be expressed in terms of the molar fractions as:

$$K_1(T) = \frac{[\mathbf{M^*}]^2(T)}{[\mathbf{M{\cdot}M}](T)} = \frac{2\chi^2_{\mathrm{M^*}}(T)[\mathbf{M}]_0}{\chi_{\mathrm{M{\cdot}M}}(T)} \tag{28}$$

$$K_2(T) = \frac{[\mathbf{M}](T)}{[\mathbf{M^*}](T)} = \frac{\chi_{\mathrm{M}}(T)}{\chi_{\mathrm{M^*}}(T)} \tag{29}$$

which lets us express the molar fractions in terms of the common intermediate:

$$\chi_{\mathrm{M{\cdot}M}}(T) = \frac{2\chi^2_{\mathrm{M^*}}(T)[\mathbf{M}]_0}{K_1(T)} \tag{30}$$

$$\chi_{\mathrm{M}}(T) = K_2\chi_{\mathrm{M^*}}(T) \tag{31}$$

Noting that at every temperature $\sum_i \chi_i(T) = 1$, we obtain a quadratic equation in $\chi_{\mathrm{M^*}}(T)$:

$$\left(\frac{2[\mathbf{M}]_0}{K_1(T)}\right)\chi^2_{\mathrm{M^*}}(T) + (1 + K_2)\chi_{\mathrm{M^*}}(T) - 1 = 0 \tag{32}$$

which has only one physically meaningful root:

$$\chi_{\mathrm{M^*}}(T) = \frac{-(K_1(T) + K_1(T)K_2(T)) + \sqrt{(K_1(T) + K_1(T)K_2(T))^2 + 8[\mathbf{M}]_0K_1(T)}}{4[\mathbf{M}]_0} \tag{33}$$

Melting temperatures $T_{m,1}$ and $T_{m,1}$ can be defined as $K_1(T_{m,1}) = [\mathbf{M}]_0$ and $K_2(T_{m,2}) = 1$, respectively. Thus, Equation (15) gives the following isochores:

$$K_1(T) = [\mathbf{M}]_0 \exp\left\{-\frac{\Delta H_1}{R}\left(\frac{1}{T} - \frac{1}{T_{\mathrm{m},1}}\right)\right\} \tag{34}$$

$$K_2(T) = \exp\left\{-\frac{\Delta H_2}{R}\left(\frac{1}{T} - \frac{1}{T_{\mathrm{m},2}}\right)\right\} \tag{35}$$

$$\tag{36}$$

As before, the observed chemical shift δ_{obs} is a weighted average for all species in fast exchange:

$$\delta_{\mathrm{obs}} = \delta_{\mathrm{M}}\chi_{\mathrm{M}}(T) + \delta_{\mathrm{M}^*}\chi_{\mathrm{M}^*}(T) + \delta_{\mathrm{M\cdot M}}\chi_{\mathrm{M\cdot M}}(T) \tag{37}$$

which can be combined with Equations (30) and (31) to yield:

$$\delta_{\mathrm{obs}} = \delta_{\mathrm{M}}K_2\chi_{\mathrm{M}^*}(T) + \delta_{\mathrm{M}^*}\chi_{\mathrm{M}^*}(T) + \delta_{\mathrm{M\cdot M}}\frac{2\chi^2_{\mathrm{M}^*}(T)[\mathbf{M}]_0}{K_1(T)} \tag{38}$$

A model defined by Equations (33), (34), (35) and (38) is then fitted to the experimental data with ΔH_1, ΔH_2, $T_{\mathrm{m},1}$, $T_{\mathrm{m},2}$, $\delta_{\mathrm{M\cdot M}}$, δ_{M^*}, and δ_{M} as parameters, using **lmfit** package in Python.[2] Analysis was performed in Jupyter and the results were plotted using **matplotlib**.[4,5]

```python
# FITTING THREE-STATE ISOCHORE TO VT NMR DATA
# Non-linear regression done using lmfit
# https://lmfit.github.io/lmfit-py/model.html

from lmfit import Parameters, Model
import numpy as np

# Parameters definition with initial and fixed values
# add with tuples: (NAME VALUE VARY MIN  MAX  EXPR  BRUTE_STEP)

c = 0.00744   # Total concentration (M)
R = 8.314     # Universal gas constant (J / mol K)

fit_params = Parameters()
fit_params.add_many((’dH1’, 30000, True, None, None, None, None),
                (’Tg1’, 300, True, None, None, None, None),
                (’dH2’, 30000, True, None, None, None, None),
                (’Tg2’, 394, True, None, None, None, None),
                (’d_K1’, -60.8600, False, None, None, None, None),
                (’d_L1’, -60, True, None, None, None, None),
                (’d_U1’, -58.6528, False, None, None, None, None),
                (’d_K2’, -60.8600, False, None, None, None, None),
                (’d_L2’, -58, True, None, None, None, None),
                (’d_U2’, -58.6528, False, None, None, None, None))

```

```
# EQUILIBRIUM: K <=(K1)=> 2L // L <=(K2)=> U
# Molar fractions and equilibrium constants

def xL(T, dH1, Tg1, dH2, Tg2):
    return (K1(T,dH1,Tg1)*(-(1+(K2(T,dH2,Tg2))) + np.sqrt((1+(K2(T,dH2,Tg2)))* \
                  (1+(K2(T,dH2,Tg2))) + 8*c/K1(T,dH1,Tg1)) )/(4*c))

def xK(T, dH1, Tg1, dH2, Tg2):
    return 2*c*xL(T, dH1, Tg1, dH2, Tg2)* \
        xL(T, dH1, Tg1, dH2, Tg2) / (K1(T, dH1, Tg1))

def xU(T, dH1, Tg1, dH2, Tg2):
    return (K2(T, dH2, Tg2)) * xL(T, dH1, Tg1, dH2, Tg2)

def K1(T, dH1, Tg1):
    return c*np.exp(-dH1/R*(1/T-1/Tg1))

def K2(T, dH2, Tg2):
    return np.exp(-dH2/R*(1/T-1/Tg2))

# MODEL FUNCTION: Observed chemical shift (weighted average)

def d_obs(T, dH1, Tg1, dH2, Tg2, d_K, d_L, d_U):
    return xK(T, dH1, Tg1, dH2, Tg2)*d_K + xL(T, dH1, Tg1, dH2, Tg2)*d_L + \
                         xU(T, dH1, Tg1, dH2, Tg2)*d_U

# Objective function to be minimised
# Returns array of residuals of the model

  def fit_function(fit_params, T, F1, F2):
    dH1 = fit_params[’dH1’].value
    Tg1 = fit_params[’dH1’].value
    dH2 = fit_params[’dH1’].value
    Tg2 = fit_params[’dH1’].value
    d_K1 = fit_params[’d_K1’].value
    d_L1 = fit_params[’d_L1’].value
    d_U1 = fit_params[’d_U1’].value
    d_K2 = fit_params[’d_K2’].value
    d_L2 = fit_params[’d_L2’].value
    d_U2 = fit_params[’d_U2’].value

    modelF1 = d_obs(T, dH1, Tg1, dH2, Tg2, d_K1, d_L1, d_U1)
    modelF2 = d_obs(T, dH1, Tg1, dH2, Tg2, d_K2, d_L2, d_U2)

    residF1 = F1 - modelF1
    residF2 = F2 - modelF2

    return np.concatenate((residF1, residF2))

# Simultaneous fitting parameters to the two 19F signals
# T : dataframe containing sample temperatures (K)
# d_F1 : dataframe containing observed chemical shifts F1 (ppm)
# d_F2 : dataframe containing observed chemical shifts F2 (ppm)
# results : contains best fit parameters

results = minimize(fit_function, fit_params, args=(T, d_F1, d_F2), method=’leastsq’)

print(fit_report(results) + ’\n’ + fit_report(results2))
```

6.1 ADAD Three-State Melting Isochore - No Constraints

```
[[Fit Statistics]]
    # fitting method   = leastsq
    # function evals   = 12
    # data points      = 32
    # variables        = 10
    chi-square         = 7.5876e-04
    reduced chi-square = 3.4489e-05
    Akaike info crit   = -320.785961
    Bayesian info crit = -306.128602
[[Variables]]
    dH1:   29796.8620 +/- 4103.86205 (13.77%) (init = 29799.05)
    Tg1:   333.436383 +/- 34.4010560 (10.32%) (init = 333.4191)
    dH2:   33971.9826 +/- 9539.48370 (28.08%) (init = 33977.12)
    Tg2:   405.408462 +/- 16.0680797 (3.96%) (init = 405.4011)
    d_K1: -60.8544054 +/- 0.01954072 (0.03%) (init = -60.8544)
    d_L1: -59.5105048 +/- 0.54125404 (0.91%) (init = -59.51077)
    d_U1: -58.6751958 +/- 0.17105372 (0.29%) (init = -58.67521)
    d_K2: -60.8788143 +/- 0.03591207 (0.06%) (init = -60.8788)
    d_L2: -58.7014142 +/- 0.90702552 (1.55%) (init = -58.70186)
    d_U2: -58.8332475 +/- 0.53358457 (0.91%) (init = -58.833)
[[Correlations]] (unreported correlations are < 0.100)
    C(Tg1, d_L2)  =  1.000
    C(d_L1, d_L2) =  0.999
    C(Tg1, d_L1)  =  0.997
    C(d_L1, d_U2) = -0.992
    C(d_L2, d_U2) = -0.988
    C(Tg1, d_U2)  = -0.985
    C(dH1, d_K2)  =  0.972
    C(dH1, Tg1)   = -0.965
    C(dH1, d_L2)  = -0.960
    C(d_K1, d_K2) =  0.953
    C(dH1, d_L1)  = -0.949
    C(dH1, d_K1)  =  0.943
    C(dH1, d_U2)  =  0.935
    C(dH1, dH2)   =  0.929
    C(dH2, Tg2)   = -0.920
    C(Tg1, dH2)   = -0.912
    C(dH2, d_L2)  = -0.903
    C(Tg1, d_K2)  = -0.886
    C(Tg1, d_K1)  = -0.884
    C(d_K1, d_L2) = -0.883
    C(d_K1, d_L1) = -0.882
```

6.2 ADAD Three-State Melting Isochore - Limiting Shifts Constrained

```
[[Fit Statistics]]
    # fitting method   = leastsq
    # function evals   = 8
    # data points      = 32
    # variables        = 6
    chi-square         = 0.00124359
    reduced chi-square = 4.7830e-05
    Akaike info crit   = -312.975697
    Bayesian info crit = -304.181282
[[Variables]]
    dH1:   31403.5044 +/- 475.490681 (1.51%) (init = 31403.57)
    Tg1:   324.265341 +/- 1.87996359 (0.58%) (init = 324.2652)
    dH2:   37122.5690 +/- 2020.41044 (5.44%) (init = 37122.87)
    Tg2:   404.498504 +/- 1.59616141 (0.39%) (init = 404.4984)
    d_K1: -60.86 (fixed)
    d_L1: -59.6202713 +/- 0.02834152 (0.05%) (init = -59.62027)
    d_U1: -58.6528 (fixed)
    d_K2: -60.86 (fixed)
    d_L2: -58.9369289 +/- 0.03677951 (0.06%) (init = -58.93693)
    d_U2: -58.6528 (fixed)
[[Correlations]] (unreported correlations are < 0.100)
    C(d_L1, d_L2) =  0.944
    C(Tg1, d_L2)  =  0.942
    C(Tg2, d_L1)  =  0.838
    C(dH1, dH2)   =  0.834
    C(Tg2, d_L2)  =  0.823
    C(Tg1, d_L1)  =  0.810
    C(dH1, Tg1)   = -0.789
    C(Tg1, Tg2)   =  0.739
    C(Tg1, dH2)   = -0.681
    C(dH1, d_L2)  = -0.591
    C(dH2, d_L2)  = -0.430
    C(dH2, Tg2)   = -0.397
    C(dH1, d_L1)  = -0.357
    C(dH1, Tg2)   = -0.308
    C(dH2, d_L1)  = -0.178
```

7 Molecular Modelling

Molecular mechanics calculations were performed in Schrödinger Suite 2016-4 using MacroModel software.[6] Simplified structures were used, in which the end-capping protecting groups groups and the *iso*-butyl chains on the phosphine oxides were changed to methyl groups in order to reduce the computational cost. All structures were minimised first and the minimised structures were then used as the starting molecular structures for all MacroModel conformational searches. Two independent searches were performed, using MMFFs and OPLS3 as force fields with implicit solvation in chloroform, as implemented in the software.[7] The charges were defined by the force field library and no cut-off was used for non-covalent interaction. For the **ADAD** dimers, two hydrogen bonds between the inner residues were constrained and the outer residues were initially constrained as duplex or kissing stem-loops, with distance defined as (1.7 ± 0.5) Å and force constant of 100. No constraints were used for monomeric the **ADAD**. Mixed torsional/Large-Scale Low-Mode Sampling was used with Enhanced torsion sampling options, so as to include ester C–O bonds, and 100 steps per rotatable bond. Maximum of 10,000 iterations were performed per sample with redundant conformers eliminated using root mean square deviation (RMSD) of 2 Å. The minima converged on a Polak-Ribiere Conjugate Gradient (PRCG) with a threshold of 1.0.

The resulting lowest energy structures were used as the starting points for a further conformational search with only the two hydrogen bonds between the two inner residues in dimeric **ADAD** constrained. The hydrogen bond constraints for the outer residues were removed in the further conformational searches. The second search was only performed using OPLS3 force field. Above sampling parameters were changed to a maximum of 20,000 iterations and the structure redundancy criterion was reduced to 1 Å RMSD. Mixed torsional/Large-Scale Low-Mode Sampling was still used with Enhanced torsion sampling options and 100 steps per rotatable bond, and the minima converged on PRCG with a threshold of 1.0. The lowest energy conformation was further minimised with OPLS3 force field and PRCG with a threshold of 0.01.

References

(1) Szczypiński, F. T.; Hunter, C. A. *Chemical Science* **2019**, *10*, 2444–2451.

(2) Newville, M.; Stensitzki, T.; Allen, D. B.; Ingargiola, A. LMFIT: Non-Linear Least-Square Minimization and Curve-Fitting for Python., version 0.8.0, 2014.

(3) Thordarson, P. *Chem. Soc. Rev.* **2011**, *40*, 1305–1323.

(4) Perez, F.; Granger, B. E. *Comput. Sci. Eng.* **2007**, *9*, 21–29.

(5) Hunter, J. D. *Comput. Sci. Eng.* **2007**, *9*, 90–95.

(6) Schrödinger Release 2016-4: MacroModel., New York, NY, 2016.

(7) Harder, E. et al. *J. Chem. Theory Comput.* **2016**, *12*, 281–296.