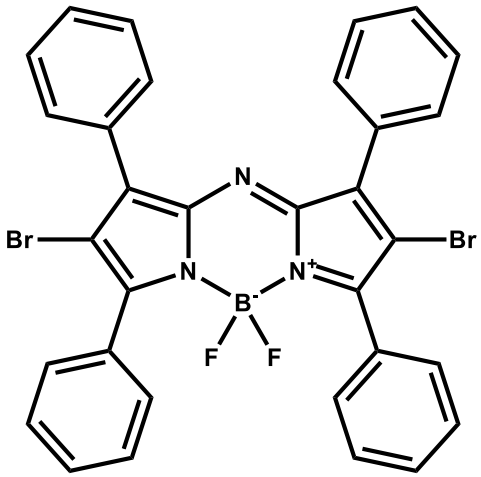
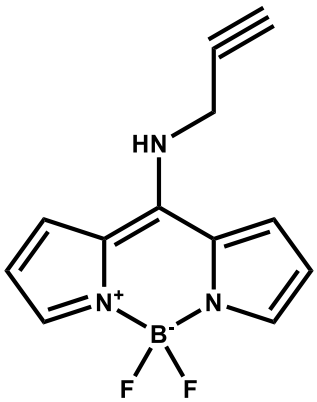
Electronic Supplementary Material (ESI) for Physical Chemistry Chemical Physics.

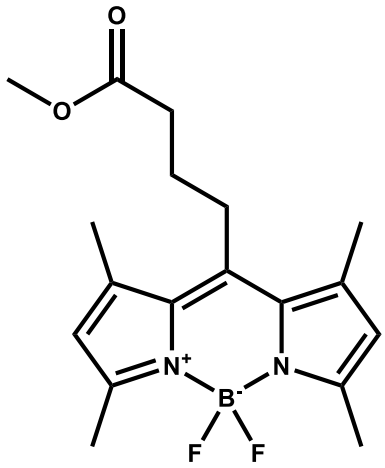
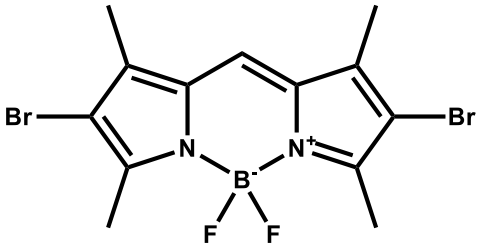
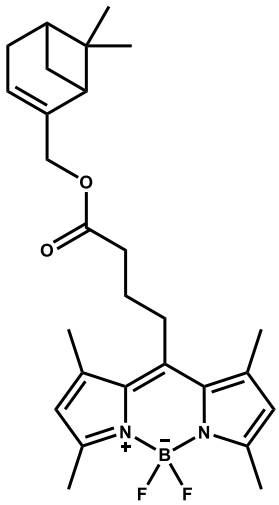
Electronic Supplementary Information

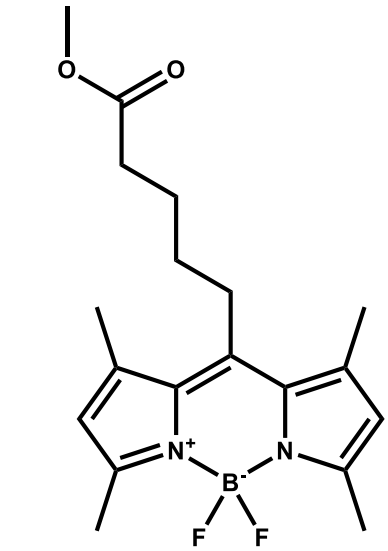
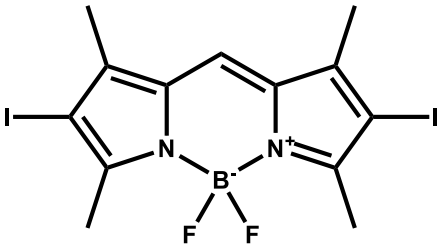
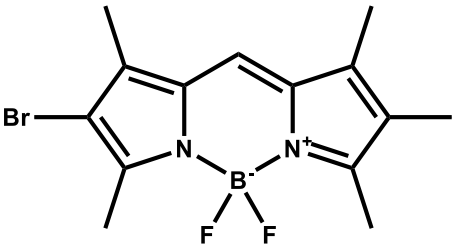
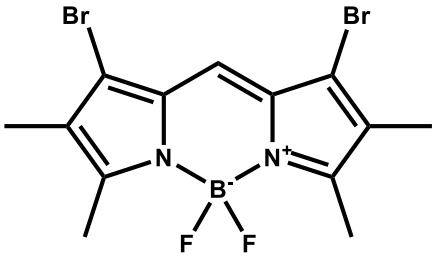
Accurate prediction of ^{11}B NMR chemical shift of BODIPYs via machine learning

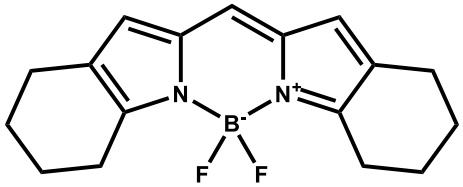
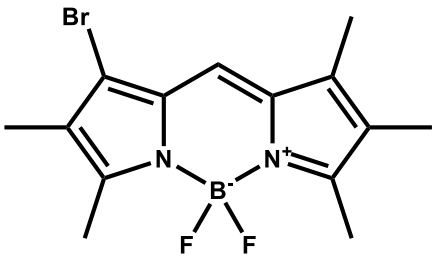
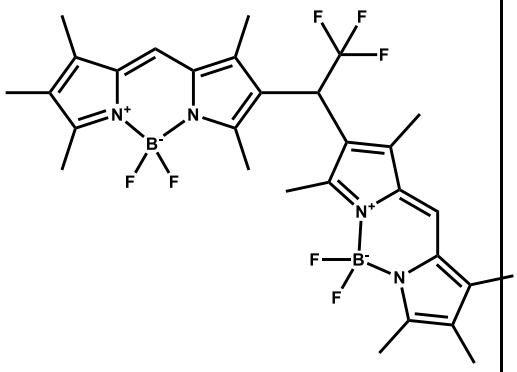
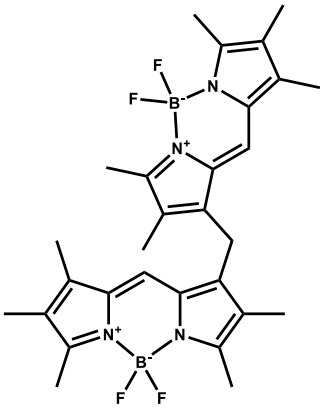
Alexander A. Ksenofontov, Yaroslav I. Isaev, Michail M. Lukanov, Dmitry M. Makarov, Varvara A. Eventova, Ilya A. Khodov, Mechail B. Berezin

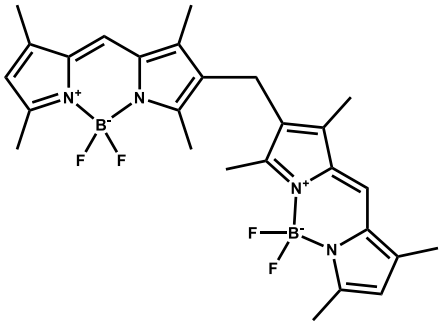
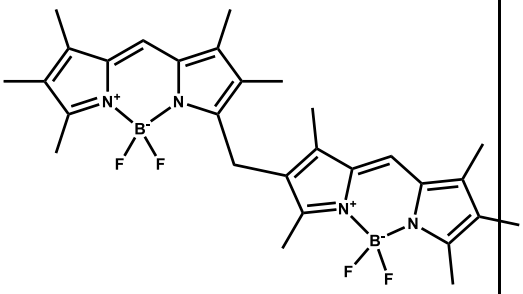
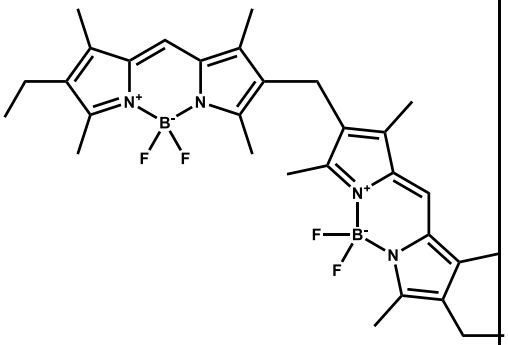
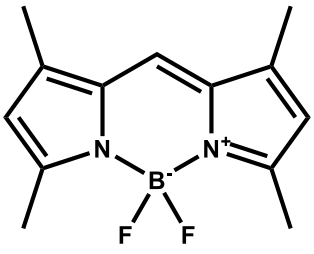
Table S1. DFT/GIAO calculations of ^{11}B chemical shift for compounds in the TEST set.

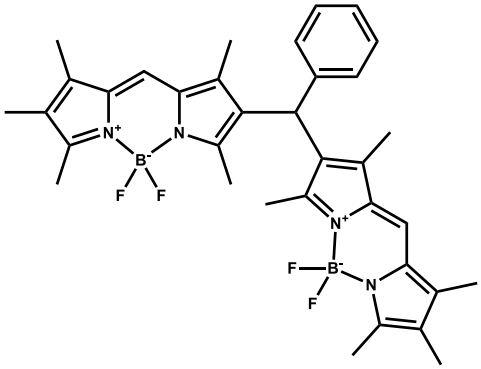
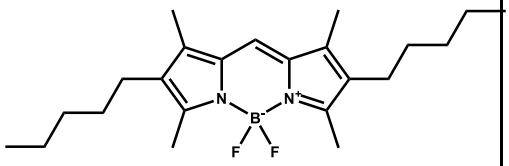
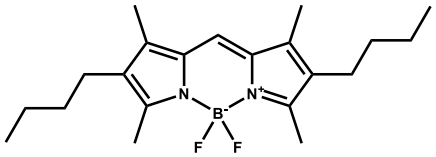
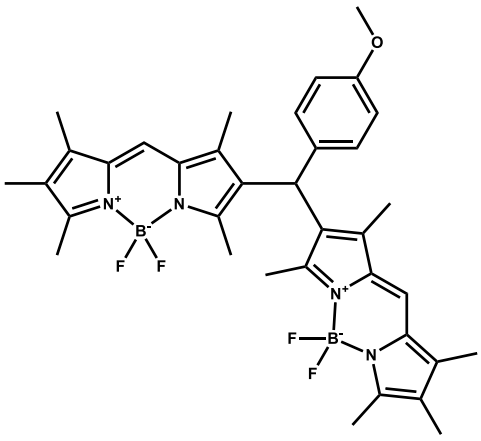
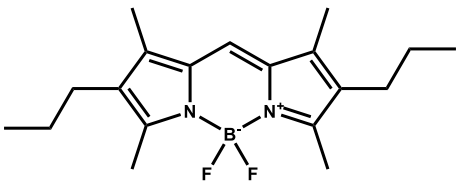
№	Compound	aug-cc-pVDZ		aug-cc-pVTZ	
		σ, ppm	**Δδ,ppm**	**σ, ppm**	**Δδ,ppm**
0	$BF_3 \cdot EtO_2$	110.754	0.000	104.070	0.000
1		108.166	2.588	103.352	0.718
2		111.152	-0.397	105.951	-1.881

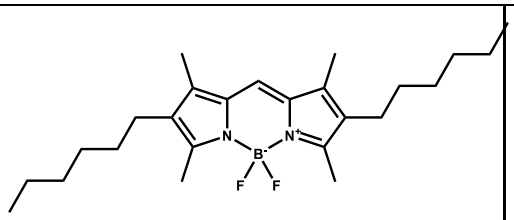
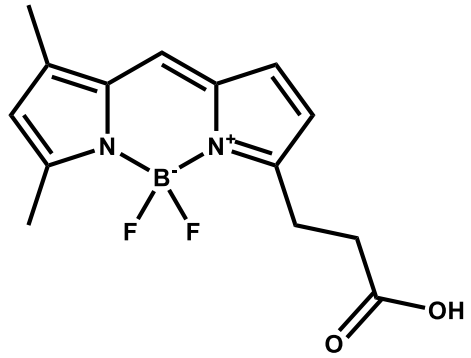
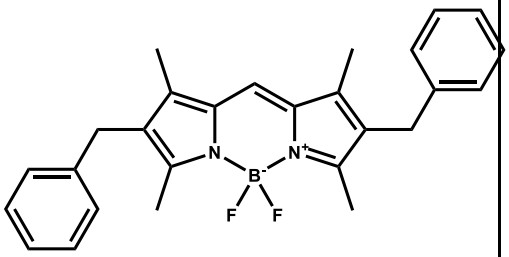
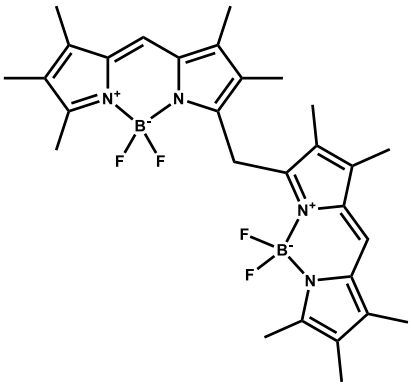
3		110.860	-0.106	105.463	-1.393
4		110.558	0.196	105.743	-1.673
5		111.111	-0.357	105.530	-1.460

6		110.897	-0.143	105.530	-1.460
7		110.8429	-0.089	105.776	-1.706
8		110.503	0.251	105.604	-1.534
9		110.575	0.179	105.492	-1.422

10		110.550	0.204	105.653	-1.583
11		110.540	0.214	105.556	-1.486
12		110.283	0.472	105.479	-1.409
13		110.644	0.110	105.308	-1.238

14		110.494	0.260	105.542	-1.472
15		110.126	0.628	105.193	-1.123
16		110.343	0.411	105.552	-1.482
17		110.437	0.317	105.372	-1.302

18		110.369	0.385	105.547	-1.477
19		110.424	0.330	105.384	-1.314
20		110.446	0.308	105.393	-1.323
21		110.855	-0.101	105.458	-1.388
22		110.469	0.285	105.411	-1.341

23		110.431	0.323	105.312	-1.242
24		110.358	0.396	105.411	-1.341
25		110.351	0.403	105.311	-1.241
26		110.578	0.177	105.783	-1.713

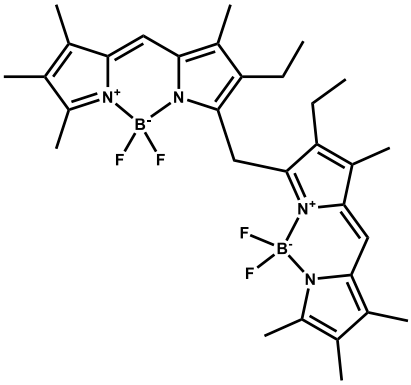
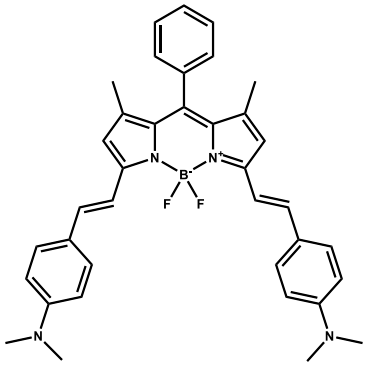
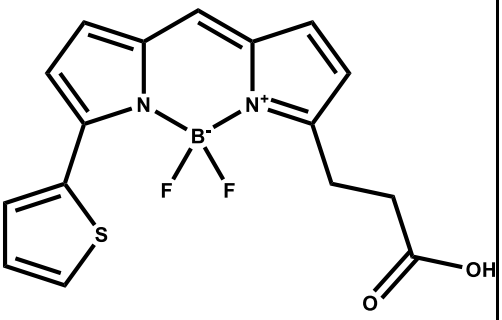
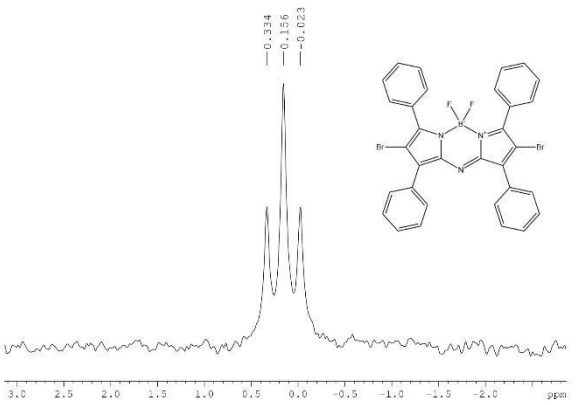
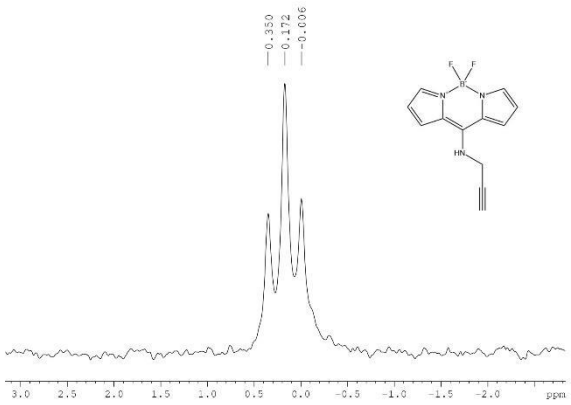
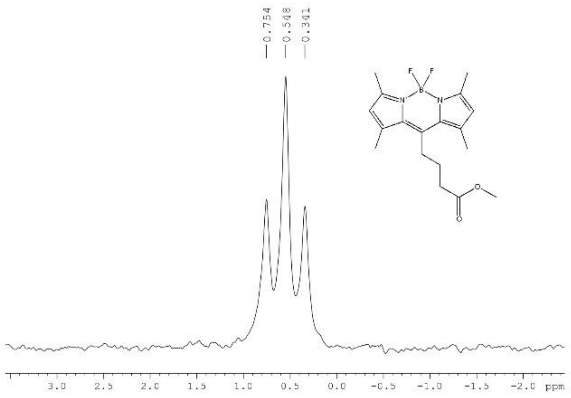
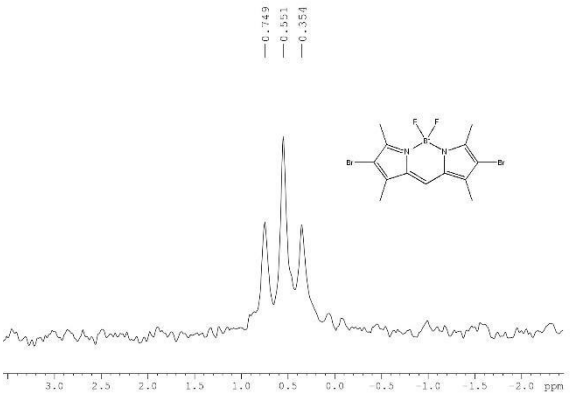
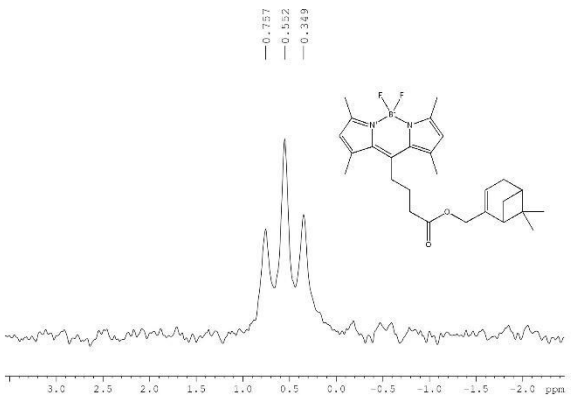
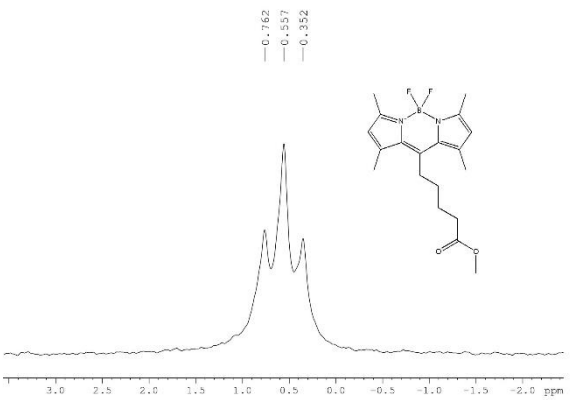
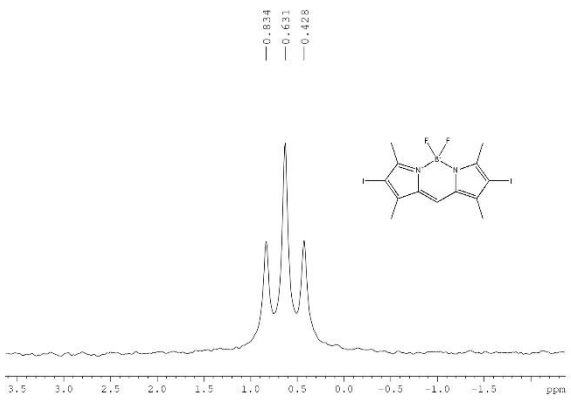
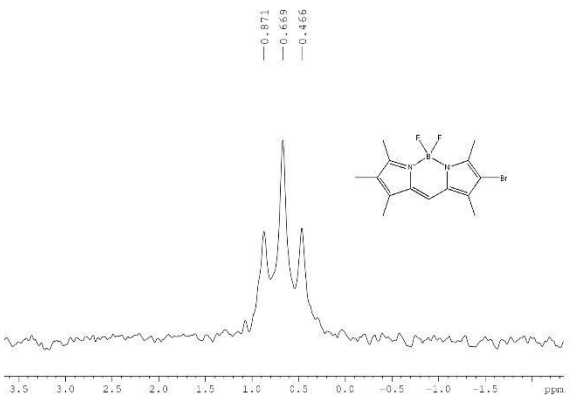
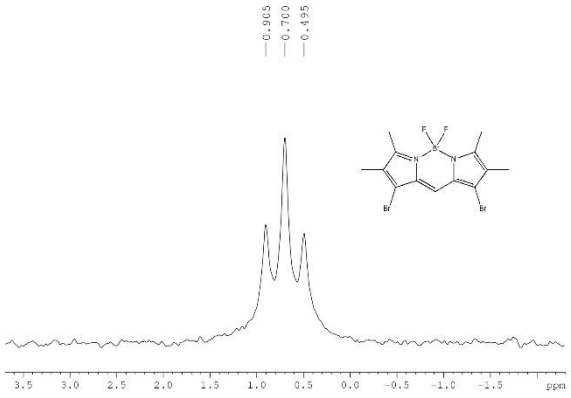
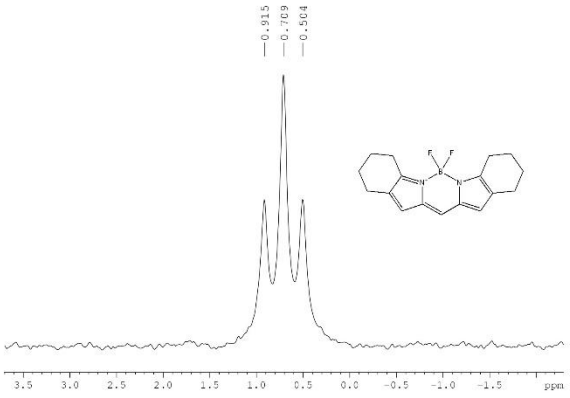
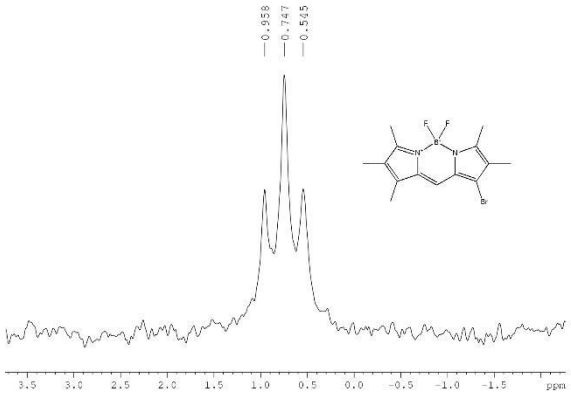
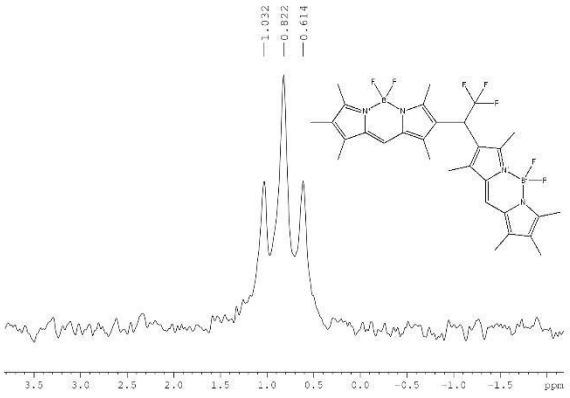
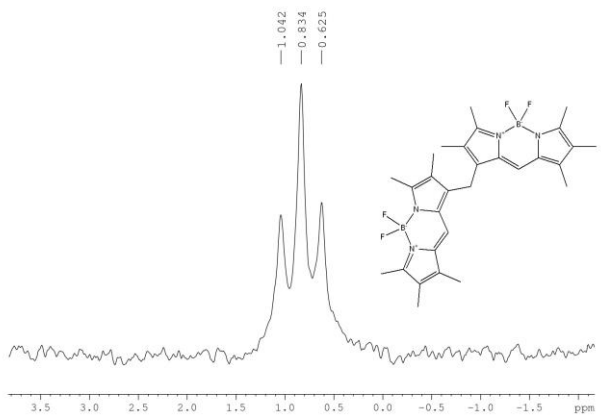
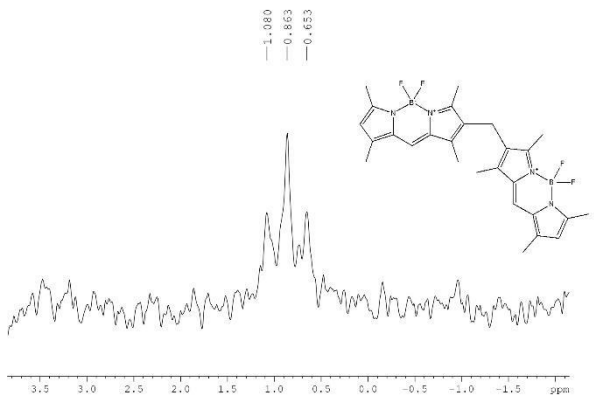
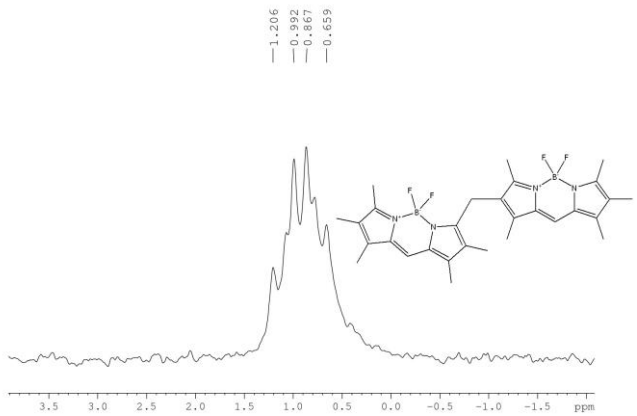
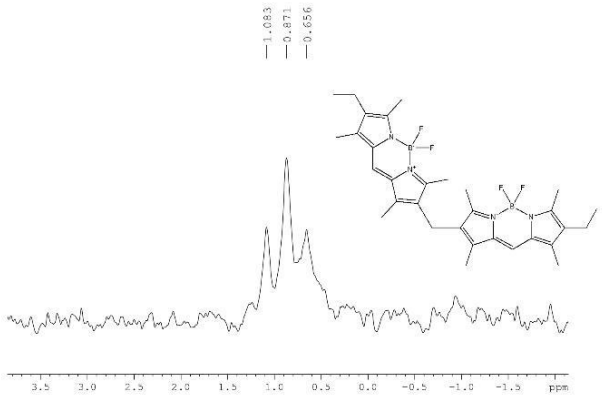
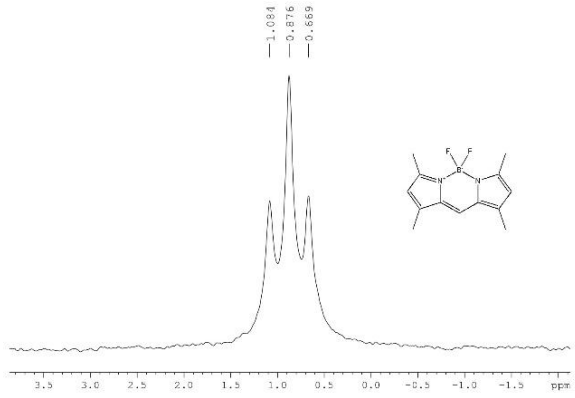
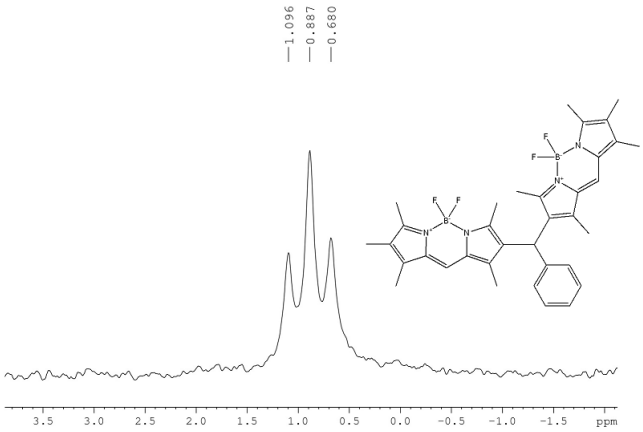
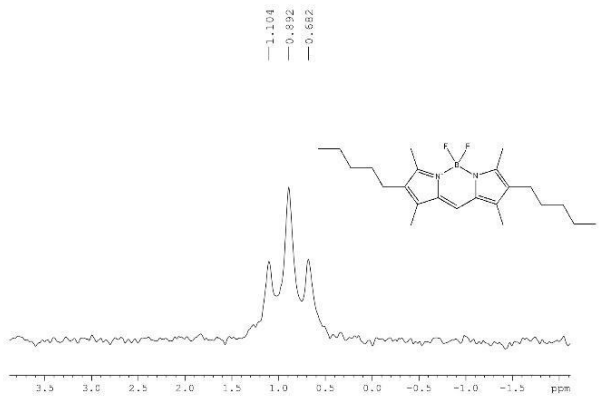
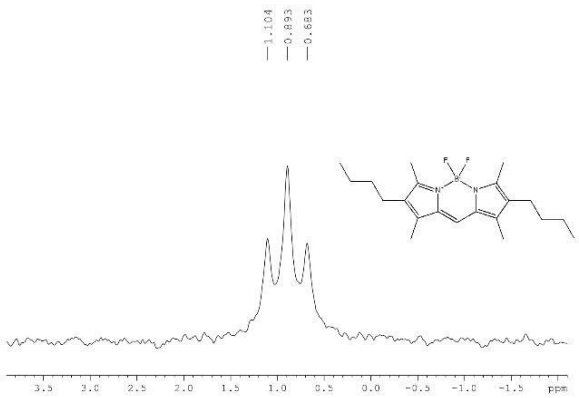
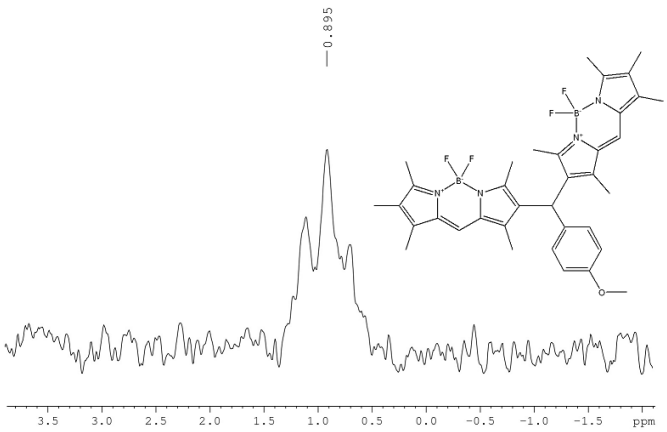
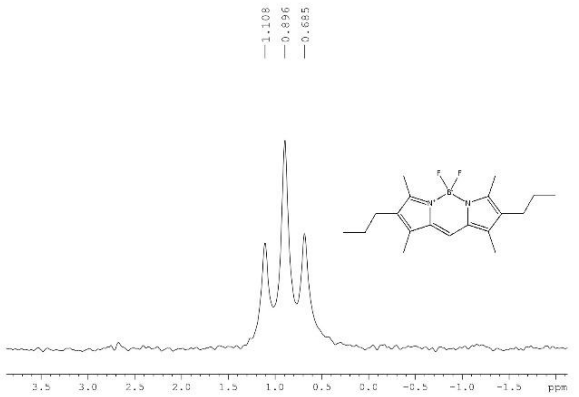
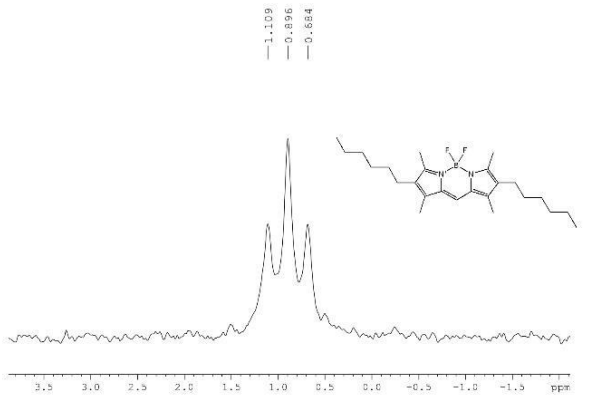
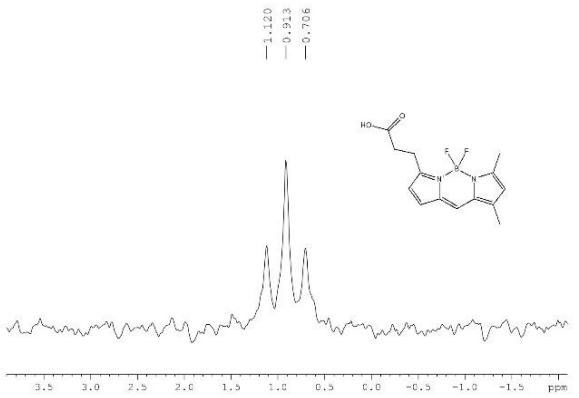
27		109.989	0.765	105.404	-1.334
28		109.467	1.287	105.219	-1.149
29		109.836	0.919	104.714	-0.644

Table S2. ^{11}B NMR spectra of BODIPYs in $CDCl_3$.

1	2
	
^{11}B NMR: δ (ppm) 0.1557 (t, $^{1}J_{BF}$ = 28.64 Hz, -BF_2-)	^{11}B NMR: δ (ppm) 0.1724 (t, $^{1}J_{BF}$ = 28.59 Hz, -BF_2-)
3	**4**
	
^{11}B NMR δ (ppm) 0.5481 (t, $^{1}J_{BF}$ = 33.14 Hz, -BF_2-)	^{11}B NMR: δ (ppm) 0.5515 (t, $^{1}J_{BF}$ = 31.73 Hz, -BF_2-)
5	**6**
	
^{11}B NMR: δ (ppm) 0.5517 (t, $^{1}J_{BF}$ = 32.74 Hz, -BF_2-)	^{11}B NMR: δ (ppm) 0.5568 (t, $^{1}J_{BF}$ = 32.93 Hz, -BF_2-)

7	8
	
^{11}B NMR: δ (ppm) 0.6309 (t, $^{1}J_{BF}$ = 32.55 Hz, -BF_2-)	^{11}B NMR: δ (ppm) 0.6692 (t, $^{1}J_{BF}$ = 32.55 Hz, -BF_2-)
9	**10**
	
^{11}B NMR: δ (ppm) 0.6999 (t, $^{1}J_{BF}$ = 32.90 Hz, -BF_2-)	^{11}B NMR: δ (ppm) 0.7094 (t, $^{1}J_{BF}$ = 32.94 Hz, -BF_2-)
11	**12**
	
^{11}B NMR: δ (ppm) 0.7475 (t, $^{1}J_{BF}$ = 33.16 Hz, -BF_2-)	^{11}B NMR: δ (ppm) 0.8225 (t, $^{1}J_{BF}$ = 33.53 Hz, -BF_2-)

13	14
 ^{11}B NMR: δ (ppm) 0.8345 (t, $^{1}J_{BF}$ = 33.47 Hz, -BF_2-)	 ^{11}B NMR: δ (ppm) 0.8635 (t, $^{1}J_{BF}$ = 34.27 Hz, -BF_2-)
15	**16**
 ^{11}B NMR: δ (ppm) 0.8665 (m, -BF2-)	 ^{11}B NMR: δ (ppm) 0.8707 (t, $^{1}J_{BF}$ = 34.28 Hz, -BF_2-)
17	**18**
 ^{11}B NMR: δ (ppm) 0.8764 (t, $^{1}J_{BF}$ = 33.25 Hz, -BF_2-)	 ^{11}B NMR: δ (ppm) 0.8872 (t, $^{1}J_{BF}$ = 33.39 Hz, -BF_2-)

19	20
	
^{11}B NMR: δ (ppm) 0.8920 (t, $^{1}J_{BF}$ = 33.80 Hz, -BF_2-)	^{11}B NMR: δ (ppm) 0.8929 (t, $^{1}J_{BF}$ = 33.84 Hz, -BF_2-)
21	**22**
	
^{11}B NMR: δ (ppm) 0.8949 (t, $^{1}J_{BF}$ = 32.69 Hz, -BF_2-)	^{11}B NMR: δ (ppm) 0.8960 (t, $^{1}J_{BF}$ = 33.89 Hz, -BF_2-)
23	**24**
	
^{11}B NMR: δ (ppm) 0.8962 (t, $^{1}J_{BF}$ = 34.12 Hz, -BF_2-)	^{11}B NMR: δ (ppm) 0.9132 (t, $^{1}J_{BF}$ = 33.25 Hz, -BF_2-)

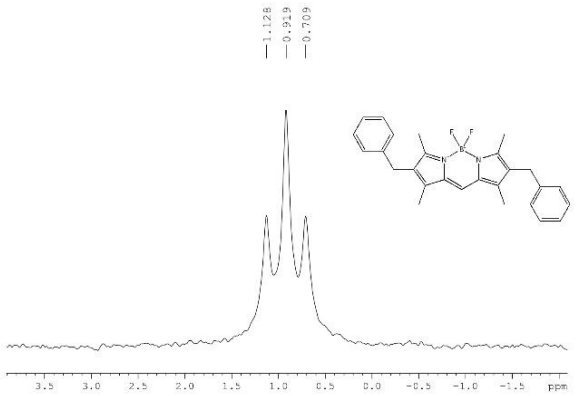
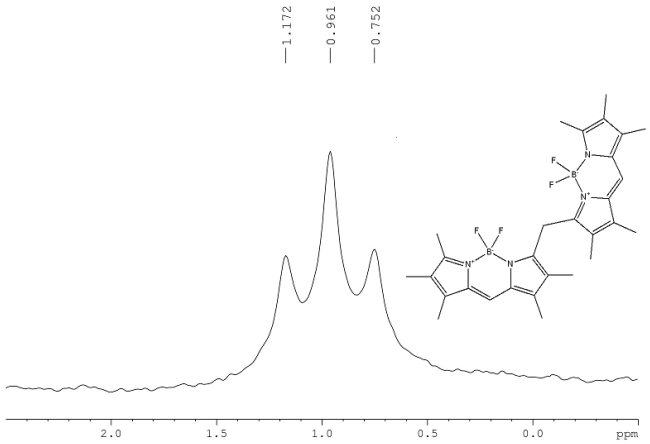
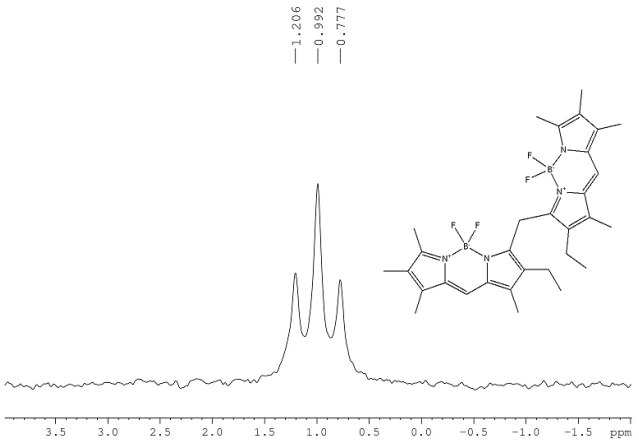
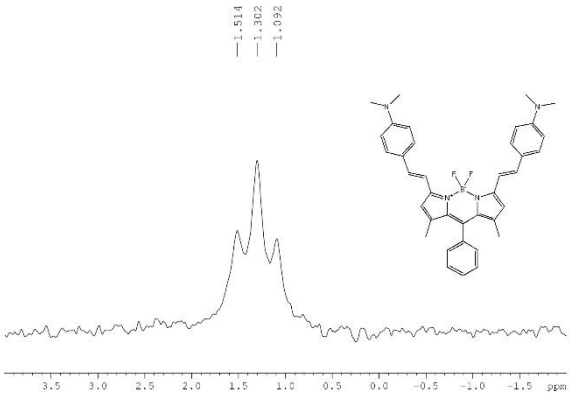
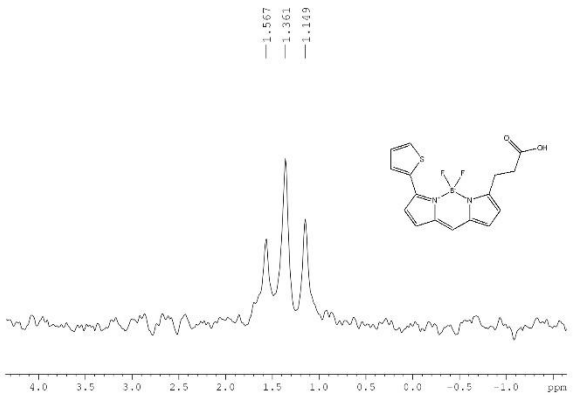
25	26
 ^{11}B NMR: δ (ppm) 0.9188 (t, $^{1}J_{BF}$ = 33.60 Hz, -BF_2-)	 ^{11}B NMR: δ (ppm) 0.9613 (t, $^{1}J_{BF}$ = 33.71 Hz, -BF_2-)
27	**28**
 ^{11}B NMR: δ (ppm) 0.9924 (t, $^{1}J_{BF}$ = 34.39 Hz, -BF_2-)	 ^{11}B NMR: δ (ppm) 1.3020 (t, $^{1}J_{BF}$ = 33.85 Hz, -BF_2-)
29	
 ^{11}B NMR: δ (ppm) 1.3608 (t, $^{1}J_{BF}$ = 33.53 Hz, -BF_2-)	

Table S3. Statistical coefficients calculated by different models for predicting the ^{11}B NMR chemical shift.

Method/Descriptors	R^2	RMSE, ppm	MAE, ppm	R^2	RMSE, ppm	MAE, ppm
	FINAL training set			TEST set		
RFR/alvaDesc	0.67 ± 0.03	0.52 ± 0.03	0.32 ± 0.01	0.40 ± 0.20	0.23 ± 0.04	0.12 ± 0.02
RFR/Avalon	0.71 ± 0.03	0.47 ± 0.02	0.27 ± 0.01	0.30 ± 0.20	0.21 ± 0.05	0.14 ± 0.03
RFR/ISIDA	0.80 ± 0.03	0.40 ± 0.02	0.25 ± 0.01	0.50 ± 0.20	0.17 ± 0.05	0.12 ± 0.02
RFR/OEstat	0.67 ± 0.03	0.50 ± 0.03	0.30 ± 0.01	0.31 ± 0.07	0.26 ± 0.03	0.22 ± 0.02
RFR/SOAP	0.54 ± 0.03	0.58 ± 0.03	0.34 ± 0.01	0.26 ± 0.08	0.28 ± 0.03	0.22 ± 0.02
XGBoost/alvaDesc	0.66 ± 0.03	0.51 ± 0.03	0.31 ± 0.01	0.30 ± 0.10	0.22 ± 0.03	0.17 ± 0.02
XGBoost/Avalon	0.71 ± 0.03	0.47 ± 0.03	0.27 ± 0.01	0.20 ± 0.10	0.27 ± 0.05	0.17 ± 0.04
XGBoost/ISIDA	0.79 ± 0.02	0.40 ± 0.02	0.24 ± 0.01	0.44 ± 0.09	0.26 ± 0.06	0.18 ± 0.04
XGBoost/OEstat	0.68 ± 0.04	0.49 ± 0.03	0.28 ± 0.01	0.20 ± 0.20	0.28 ± 0.04	0.21 ± 0.03
DNN/alvaDesc	0.52 ± 0.04	0.63 ± 0.03	0.42 ± 0.01	0.06 ± 0.10	0.70 ± 0.10	0.52 ± 0.10
DNN/Avalon	0.71 ± 0.03	0.47 ± 0.03	0.27 ± 0.01	0.20 ± 0.10	0.27 ± 0.05	0.17 ± 0.04
DNN/ISIDA	0.61 ± 0.03	0.56 ± 0.02	0.36 ± 0.01	0.10 ± 0.10	0.38 ± 0.06	0.28 ± 0.05
DNN/OEstat	0.62 ± 0.04	0.54 ± 0.03	0.32 ± 0.01	0.10 ± 0.10	0.32 ± 0.05	0.24 ± 0.04
Transformer CNF	0.67 ± 0.03	0.51 ± 0.03	0.31 ± 0.01	0.60 ± 0.10	0.19 ± 0.02	0.16 ± 0.02
Transformer CNN	0.71 ± 0.03	0.47 ± 0.03	0.29 ± 0.01	0.30 ± 0.20	0.28 ± 0.06	0.19 ± 0.04
SheNet	0.43 ± 0.04	0.66 ± 0.03	0.45 ± 0.01	0.16 ± 0.09	0.38 ± 0.06	0.28 ± 0.05