Design of a D_{3h}-symmetry prismatic tris-(ferrocene-1,1′-diyl) molecular cage bearing boronate ester linkages

Electronic Supporting information (ESI)

Maurycy Krzyżanowski[a], Anna M. Nowicka[b], Krzysztof Kazimierczuk[c], Krzysztof Durka[a], Sergiusz Luliński[a], Artur Kasprzak[a]*

[a] Faculty of Chemistry, Warsaw University of Technology, Noakowskiego Str. 3, 00-664 Warsaw, Poland

* Corresponding author e-mail: akasprzak@ch.pw.edu.pl; artur.kasprzak@pw.edu.pl

[b] Faculty of Chemistry, University of Warsaw, Pasteura Str. 1, 02-093 Warsaw, Poland

[c] Centre of New Technologies, University of Warsaw, Banacha 2C, 02-097 Warsaw, Poland

Table on contents:

1. Optimization experiments

Table S1. Optimization experiments toward FcB-cage.

Entry	HHTP	Fc-[B(OH)$_2$]$_2$ (**1**)	Fc-[B(OMe)$_2$]$_2$ (**2**)	Solvent (mL)	time	temp.	Yield
1	0.06 mmol (1 eq)	0.09 mmol (1.5 eq)	-	MeOH (7 mL) 1,4-dioxane (7 mL)	24 h	rt	Recovery of starting material
2	0.06 mmol (1 eq)	0.09 mmol (1.5 eq)	-	MeOH (7 mL) 1,4-dioxane (7 mL)	24 h	40°C	Recovery of starting material
3	0.06 mmol (1 eq)	0.09 mmol (1.5 eq)	-	MeOH (3 mL) 1,4-dioxane (9 mL)	24h	rt	Recovery of starting material
4	0.06 mmol (1 eq)	0.09 mmol (1.5 eq)	-	MeOH (9 mL)	24h	rt	Recovery of starting material
5	0.06 mmol (1 eq)	0.09 mmol (1.5 eq)	-	MeOH (9 mL)	72h	rt	Recovery of starting material
6	0.06 mmol (1 eq)	0.09 mmol (1.5 eq)	-	EtOH (9 mL)	24h	rt	Recovery of starting material
7	0.06 mmol (1 eq)	0.09 mmol (1.5 eq)	-	1,4-dioxane (3 mL) Mesitylene (3 mL)	24h	rt	Recovery of starting material
8	0.06 mmol (1 eq)	0.09 mmol (1.5 eq)	-	1,4-dioxane (3 mL) Mesitylene (3 mL)	72h	rt	Recovery of starting material
9	0.06 mmol (1 eq)	0.09 mmol (1.5 eq)	-	1,4-dioxane (3 mL) Mesitylene (3 mL)	72h	85°C	Recovery of starting material
10	0.06 mmol (1 eq)	0.09 mmol (1.5 eq)	-	Acetone (5 mL) 1,4-dioxane (5 mL)	72 h	rt	Recovery of starting material
11	0.06 mmol (1 eq)	0.09 mmol (1.5 eq)	-	Acetone (2 mL) 1,4-dioxane (2 mL)	48h	rt	Recovery of starting material
12	0.06 mmol (1 eq)	0.09 mmol (1.5 eq)	-	Acetone (2 mL) 1,4-dioxane (2 mL)	48h	40°C	Recovery of starting material
13	0.06 mmol (1 eq)	0.09 mmol (1.5 eq)	-	DMF (6 mL)	48h	rt	40%
14	**0.06 mmol (1 eq)**	**-**	**0.09 mmol (1.5 eq)**	**DMF (6 mL)**	**48h**	**rt**	**68 %**
15	0.06 mmol (1 eq)	-	0.09 mmol (1.5 eq)	DMF (6 mL)	72h	rt	67 %
15	0.06 mmol (1 eq)	-	0.09 mmol (1.5 eq)	MeOH (6 mL)	48h	rt	Recovery of starting material

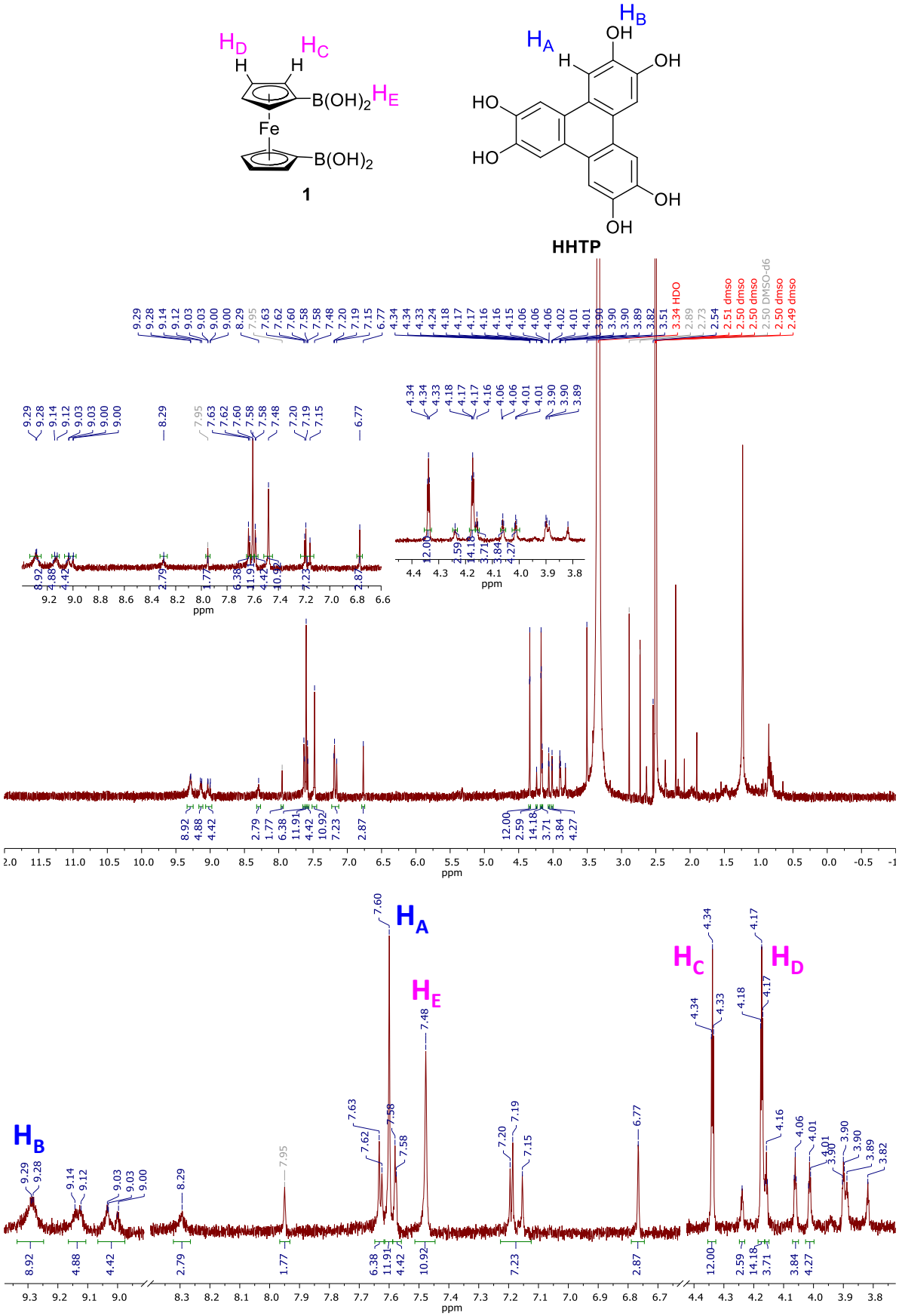


Figure S1. ^{1}H NMR (400 MHz, DMSO-d_6) spectrum of the product obtained in the reaction of **1** and HHTP in DMF (see discussion in the main text and data in Table S1, entry 14). The signals corresponding to the building blocks of hydrolyzed **FcB-cage** are marked with colors.

2. Spectral characterization

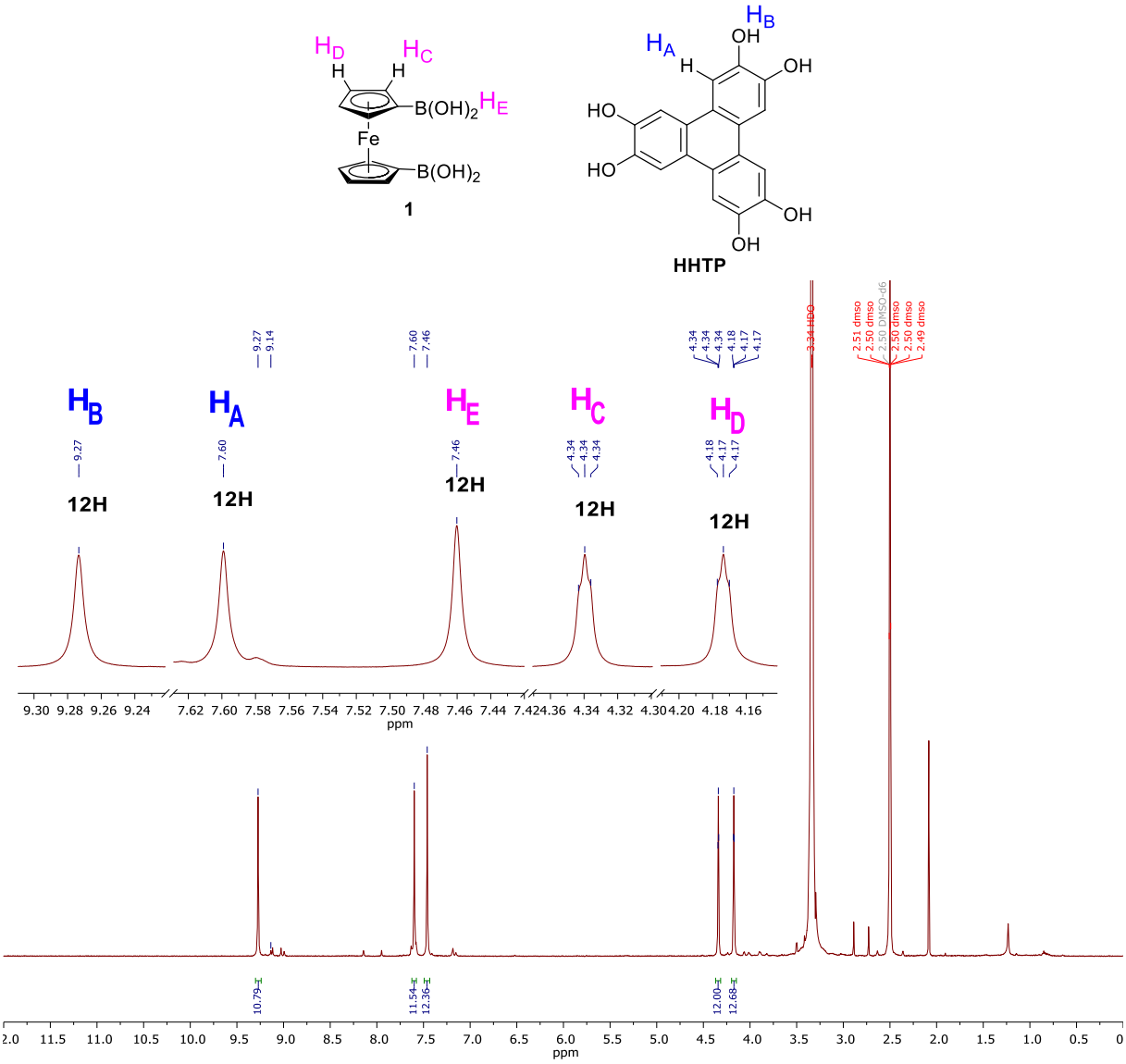


Figure S2. ^{1}H NMR (400 MHz, DMSO-d_6) spectrum of **FcB-cage** after the hydrolysis experiment.

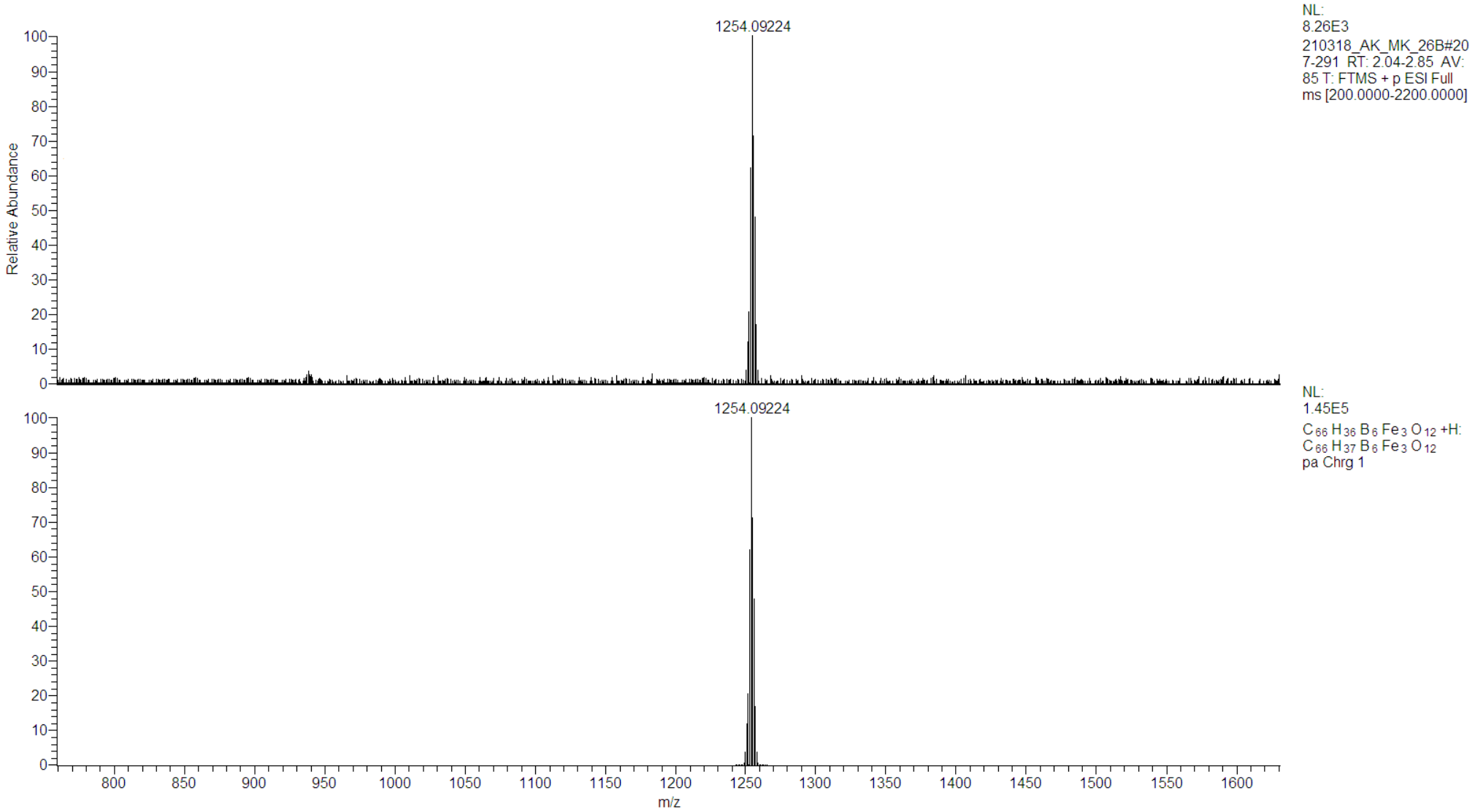


Figure S3. ESI-HRMS spectrum of **FcB-cage.**

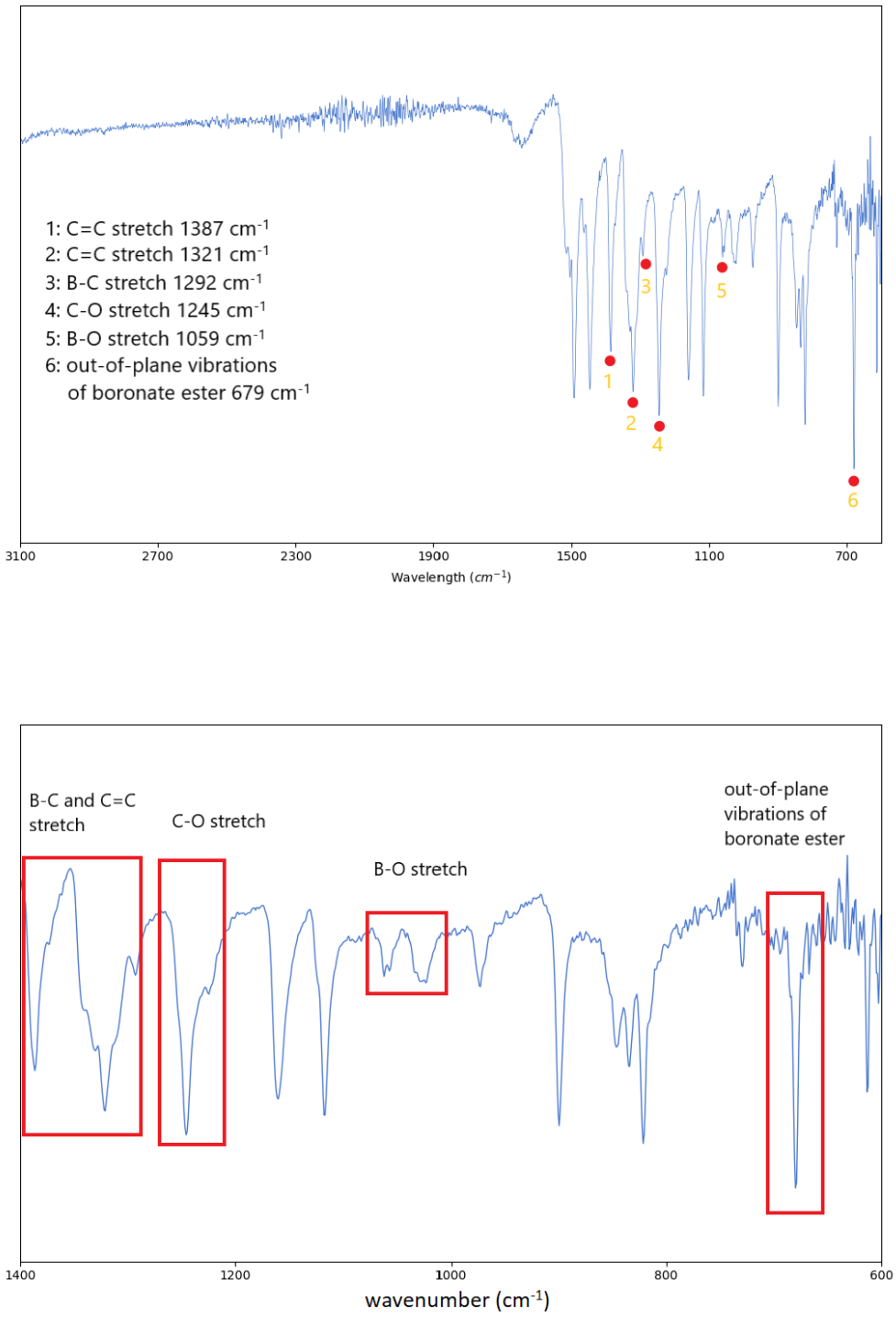


Figure S4. FT-IR spectrum of **FcB-cage** (top) and the inset presenting crucial vibrations (bottom).

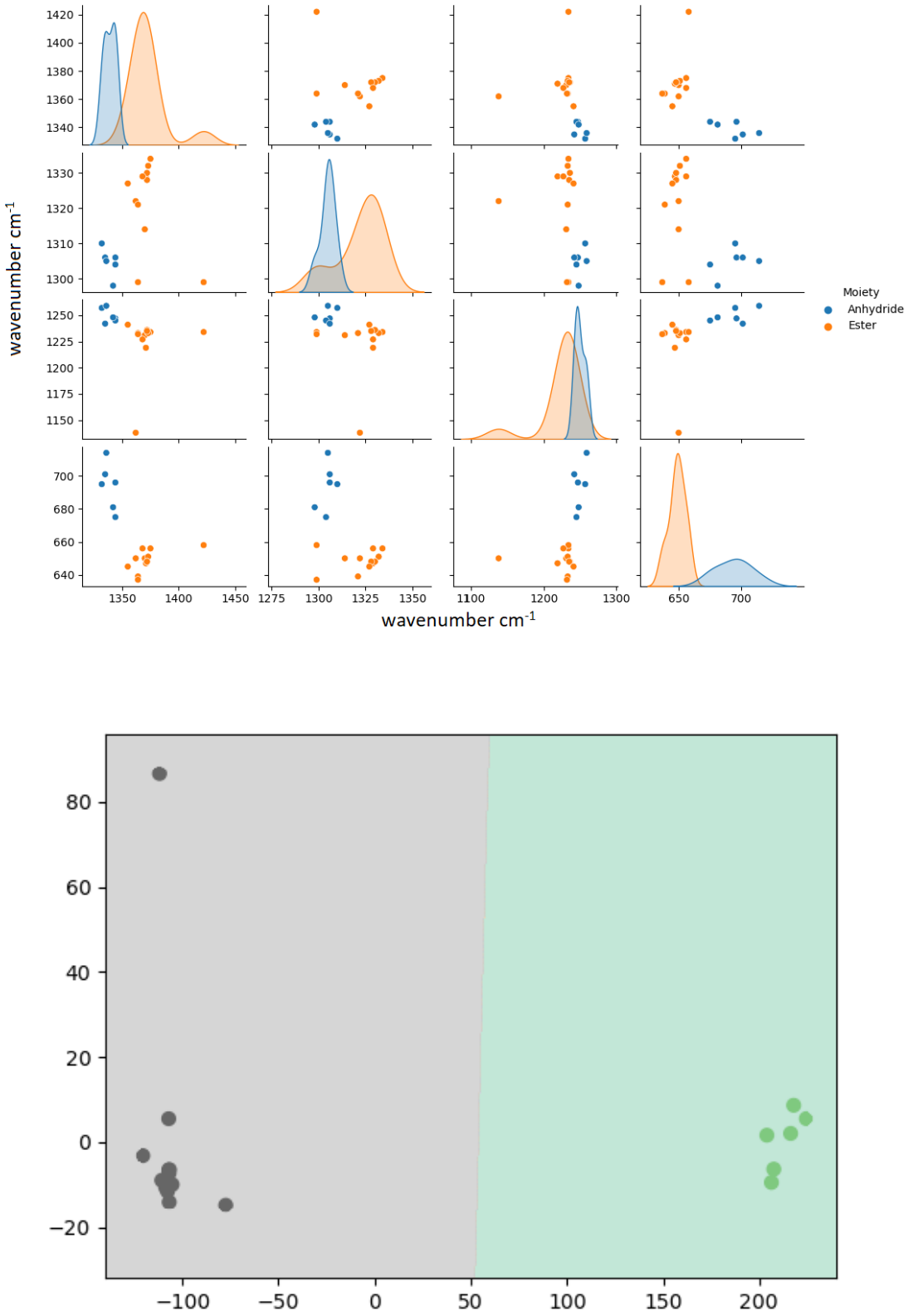


Figure S5. (top) Scatter matrix of the wavenumber values of anhydride and borate ester regarding the FT-IR analyses of **FcB-cage**, (bottom) Decision surface of rbf SVM classifier. In order to reduce dimensionality of the data to just two dimensions, so that it could be represented as a two-dimensional graph, principal component analysis (PCA) was applied to the data).

3. Additional discussion on a machine learning technique

The details on this technique can be found in the following references:

1. C. Cortes, V. Vapnik, Support-Vector Networks". *Machine Learning*, 1995, 20 (3): 273–297. doi:10.1007/BF00994018.
2. T. Hastie, R. Tibshirani, J H. Friedman. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. New York: Springer, 2001, ISBN: 978-1-0716-2122-6.
3. B. Schölkopf, A. Smola, Support Vector Machines and Kernel Algorithms. Encyclopaedia of Biostatistics, 2005, 5328-5335, ISBN: 978-0470849071.
4. F. Pedregosa, Scikit-learn: Machine Learning in Python*, Journal of Machine Learning*, 2011, 12, 2825-2830.
5. Support Vector Machines - scikit-learn 1.0.2 documentation

Machine learning can be categorized into supervised and unsupervised learning. In the case of our work, supervised learning was implemented - machine learning, where labelled examples are used. The libraries used were as follows: Scikit-learn - Python library for machine learning, Pandas - Python library for data manipulation and analysis, Matplotlib - Python figure plotting library. The algorithm predicts the output variable based on each input variable. Such a machine learning algorithm can either be a classification or regression algorithm. In the case of classification, the output variable is the discrete class value (the output = 1 means yes, the output = 0 means no). In regression, on the other hand, the output variable is continuous. For example, for a set of (x, y) data, a linear function was fit to the data through linear regression. That way, for each x value, a y value can be determined.

The output variable in our work was whether the product was anhydride or ester. It means that a classification algorithm needs to be used. The algorithm will therefore be called as a classifier. The input variables were the transmittance values of the infrared spectroscopy for a given wavelength value. In machine learning, the data is always divided into training and test sets. Training sets are used by the classifier to learn how to classify output values for given input values. A test set is used to test the classifier on previously unseen data.

The classifier used in the article was Support Vector Machine – SVM with Gaussian radial basis function as a kernel (Figure S6). SVM algorithm seeks, using the vectors, to find the best hyperplane that would separate/classify the data. Not all data can be easily separated by a hyperplane in two-dimensional space. However, if a data were aimed to be transformed into a three-dimensional space, their separation could be possible. Kernel is a method that enables classification of the data using a hyperplane in a higher-dimensional space.

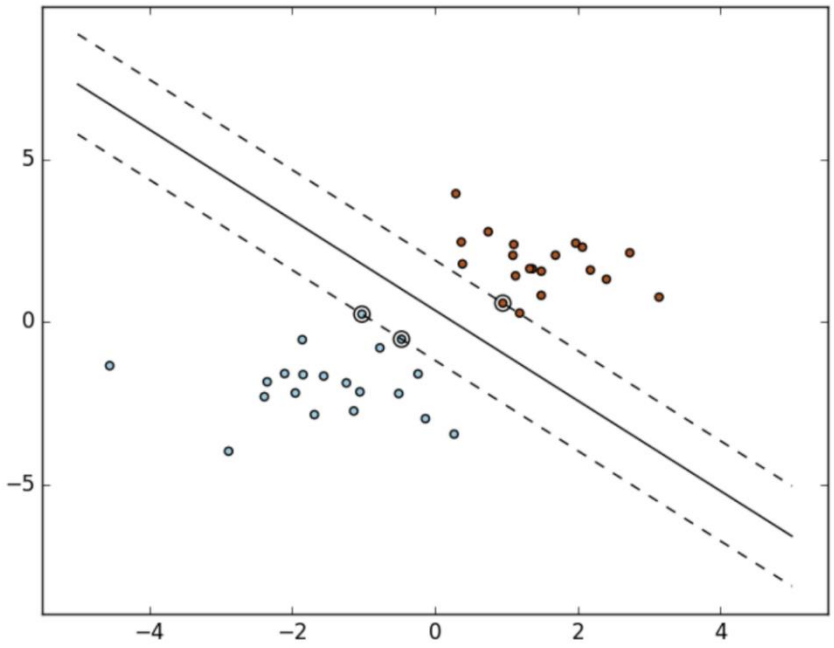


Figure S6. Hyperplane with margins that separates data (source: Support Vector Machines - scikit-learn 1.0.2 documentation).

However, directly transforming all the data into a three-dimensional space is not necessary. It is possible to find a linear decision boundary – hyperplane in the three-dimensional space while working only with data in the two-dimensional space. Such a possibility is called the kernel-trick. The kernel serves as the similarity function between two points in space because it computes the similarity between pairs of points in the high dimensional space. For the radial basis function kernel, the similarity between two points is an exponential Gaussian function:

$$k(\overrightarrow{x_i}, \overrightarrow{x_j}) = \exp\left(-\gamma \left\| \overrightarrow{x_i} - \overrightarrow{x_j} \right\|^2\right)$$

The gamma parameter, seen in the kernel function above, enables the control of the decision boundary should be. For small gamma values, the similarity boundaries of the set of points are big (Figure S7). It means that even points that are categorized otherwise can be incorporated into the boundaries set by the other points. However, large gamma values, although will not wrongfully incorporate as many points, may overfit the data.

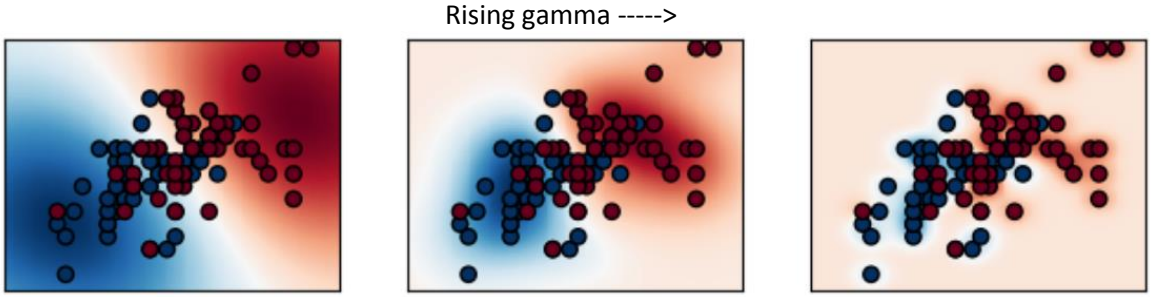


Figure S7. The influence of gamma parameter on decision boundary.

The other parameter is the regularisation parameter – C. For all kernels, this parameter is present because it is the parameter of the linear hyperplane itself. It tells how big the maximum margin criterion is. Big values of C mean that the maximum margin is small, and conversely, small values of C mean that the maximum margin is big (Figure S8). This parameter impacts just as much the decision

boundary as the gamma does. So as to not got through the work of finding the best values of the two parameters, a tool in scikit-learn called GridSearch was utilized, which finds those values.

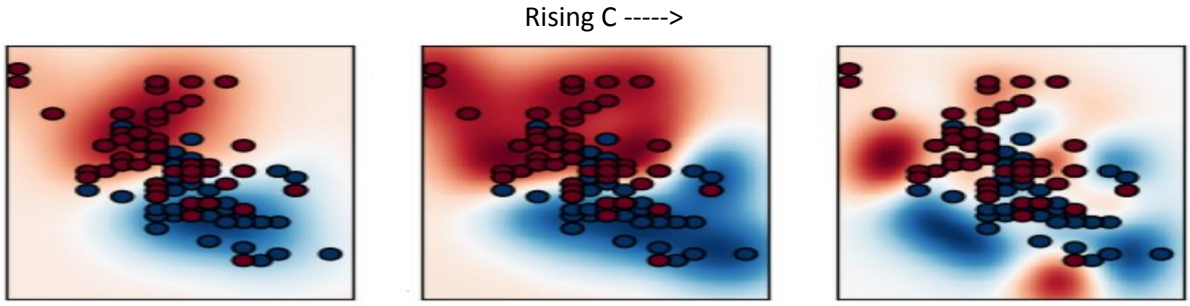


Figure S8. The influence of regularisation parameter on decision boundary.

4. Thermal properties

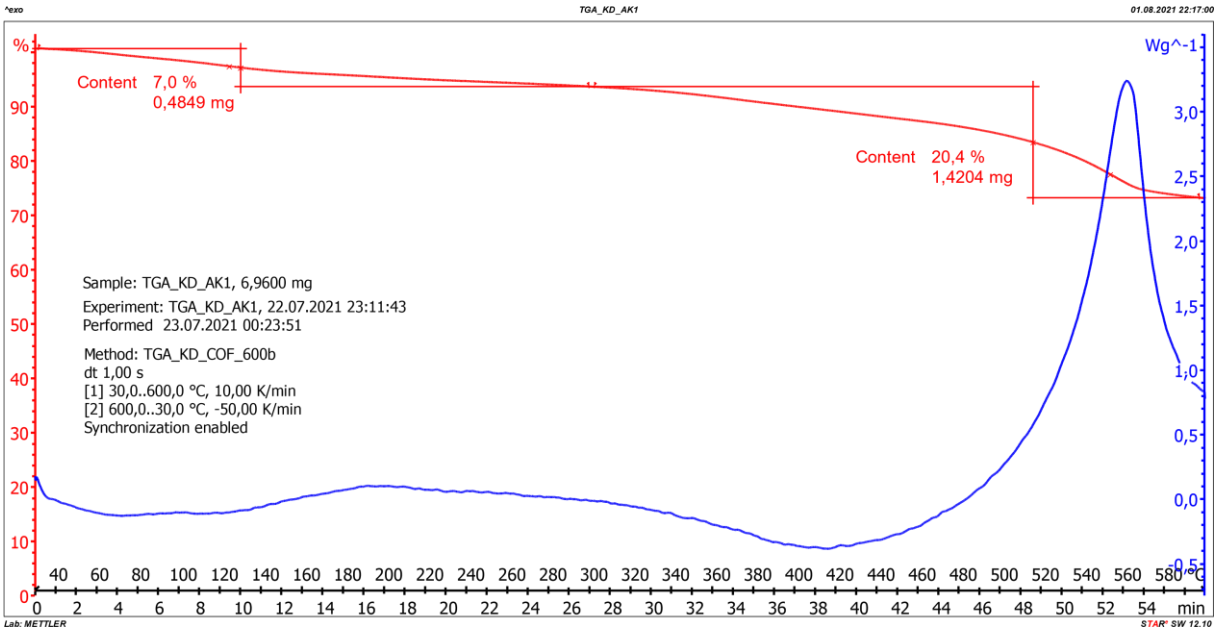


Figure S9. TGA plot of **FcB-cage** with the ramp rate of 10 K·min^{-1}.

5. SEM

Table S2. EDX data for **FcB-cage** together with the EDX spectrum at the bottom.

Element	Series	unn. C [wt.%]	norm. C [wt.%]	Atom. C [at.%]	Error (1 Sigma) [wt.%]
Boron	K-series	6.37	6.63	8.58	2.06
Carbon	K-series	59.20	61.59	71.77	7.78
Oxygen	K-series	17.90	18.62	16.29	2.98
Silicon	K-series	0.16	0.17	0.08	0.04
Iron	K-series	12.44	12.94	3.24	0.76
Sodium	K-series	0.06	0.06	0.04	0.04
	Total:	96.12	100.00	100.00	

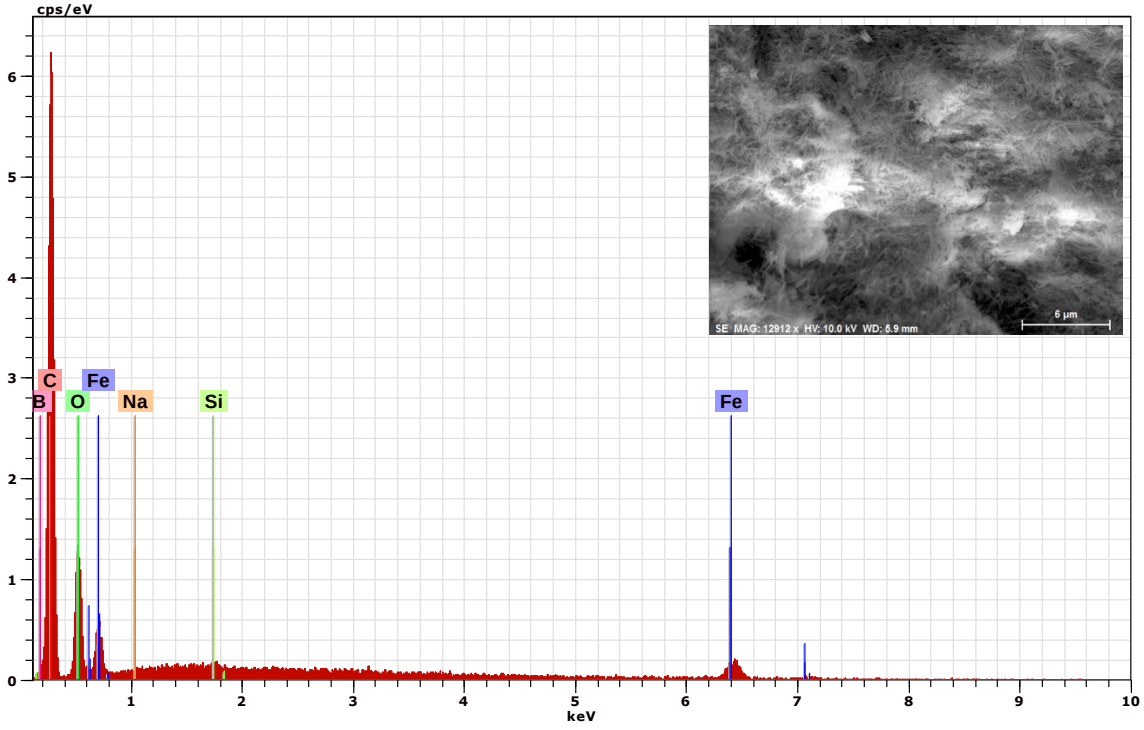


Figure S10. EDX profile of **FcB-cage**.

6. Theoretical calculations

Table S3. Computed total energy values and number of imaginary frequencies for the optimized structures (B97D/6-311++G(d,p)).

Compound	**G / a.u**	**N. im. Freq.**
2	-2159.899907	0
methanol	-115.671980	0
HHTP	-1143.860885	0
***syn*-2-HHTP**	-3072.423883	0
***anti*-2-HHTP**	-3072.424613	0
***syn*-2-HHTP$_2$**	-3984.967802	0
***anti*-2-HHTP$_2$**	-3984.945905	0
2$_2$-HHTP$_2$	-5682.199009	0
FcB-cage	-7379.440269	0

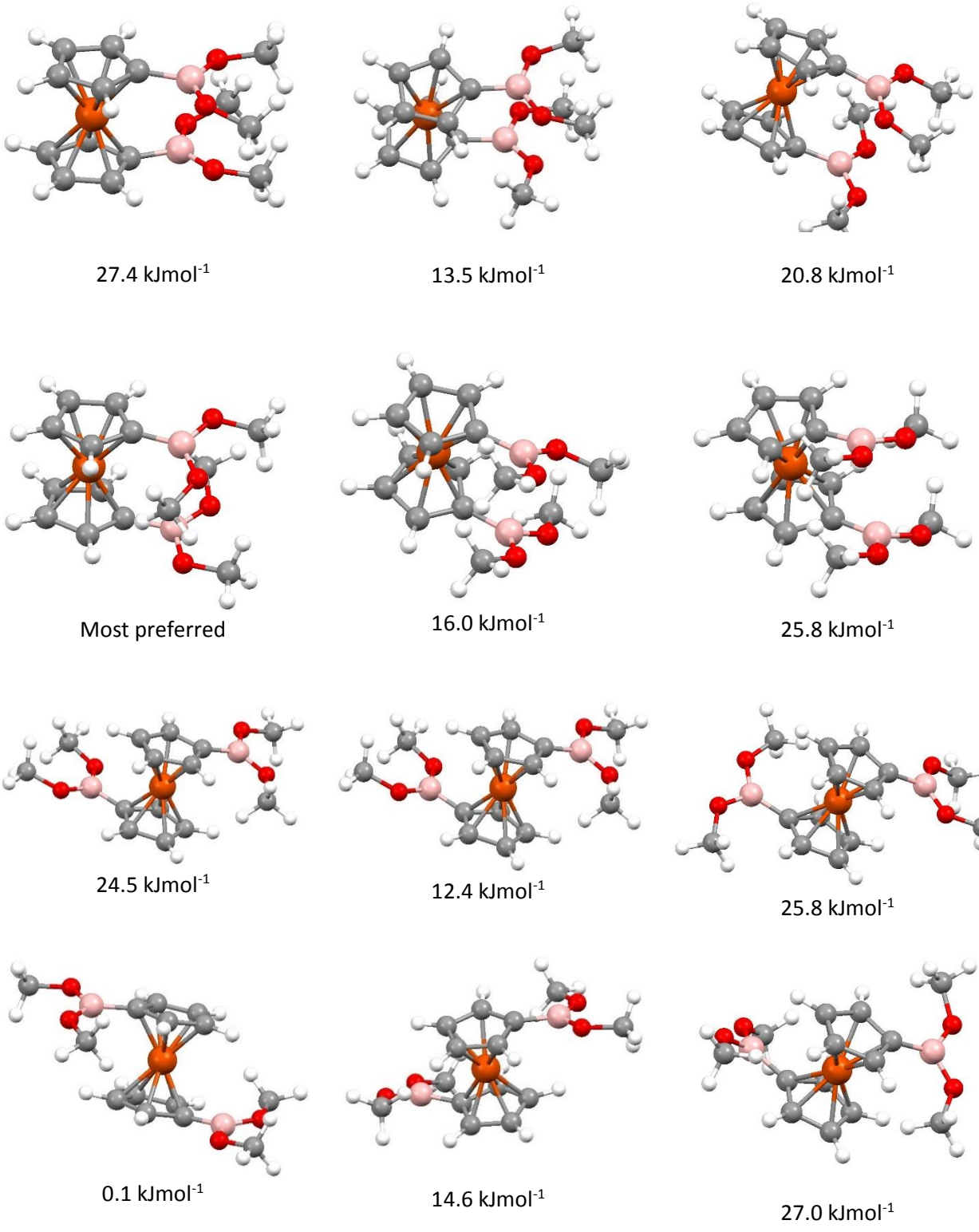
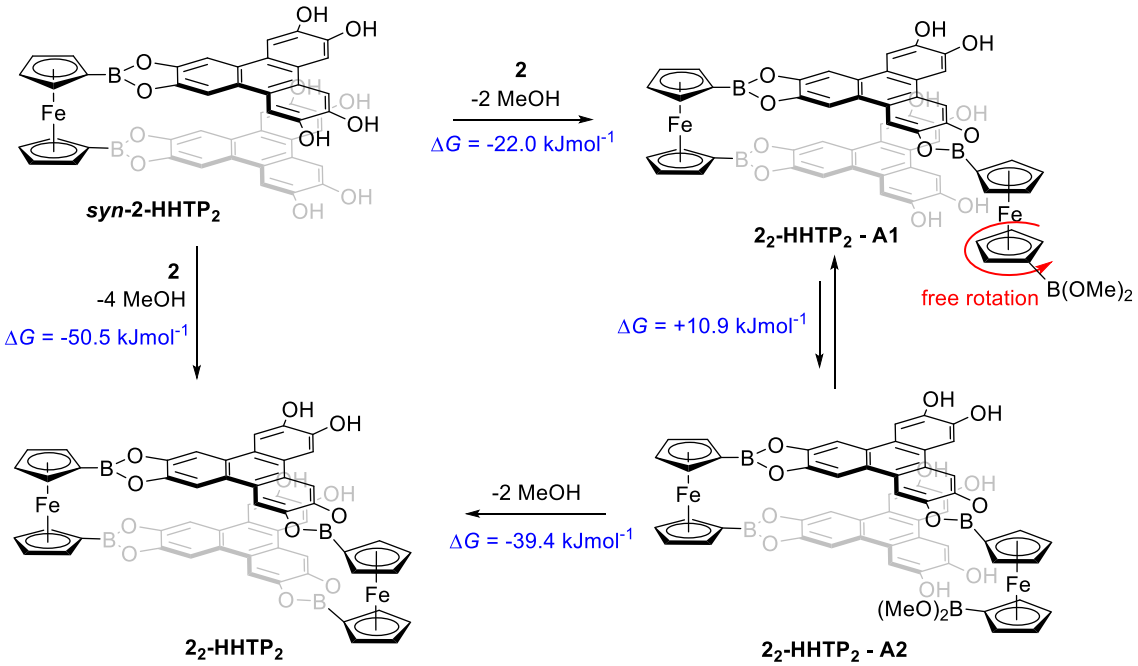
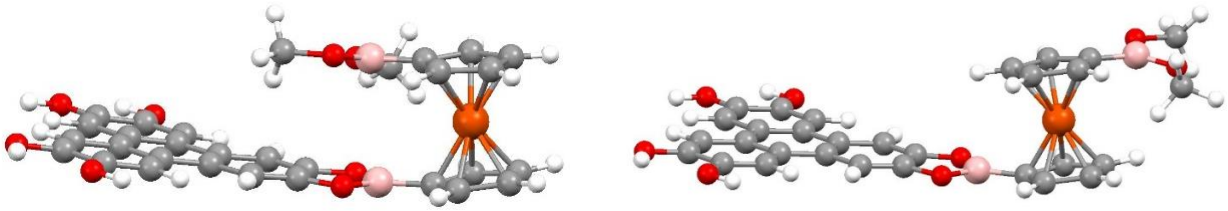


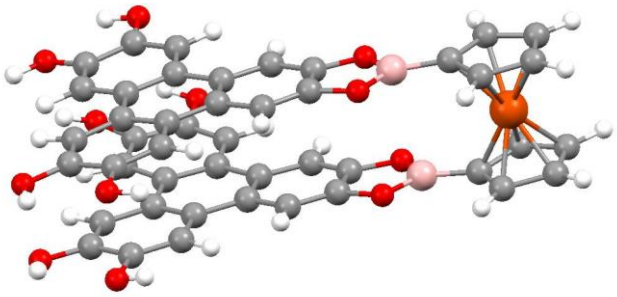
Figure S11. DFT-optimized structures of various conformers of **2**. The free enthalpies ($kJ \cdot mol^{-1}$) are given with respect to the most stable conformation.

Scheme S1. Stepwise formation of **2_2-HHTP$_2$** from the reaction of ***cis*-2-HHTP$_2$** with molecule **2**.



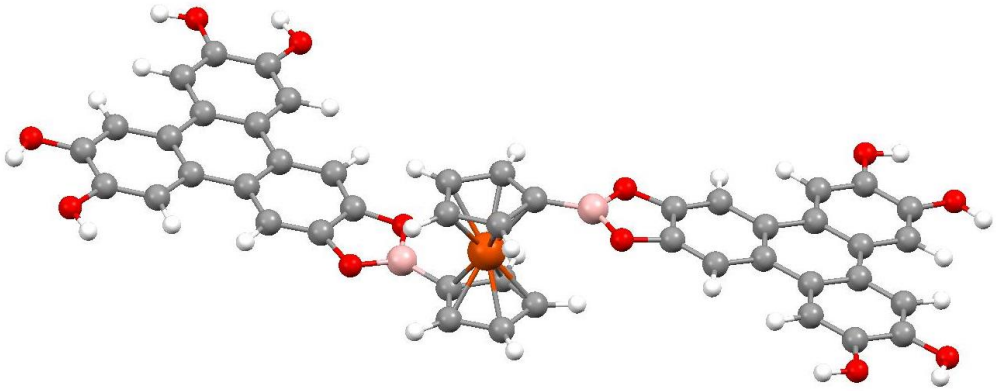


***syn*-2-HHTP** ***anti*-2-HHTP**

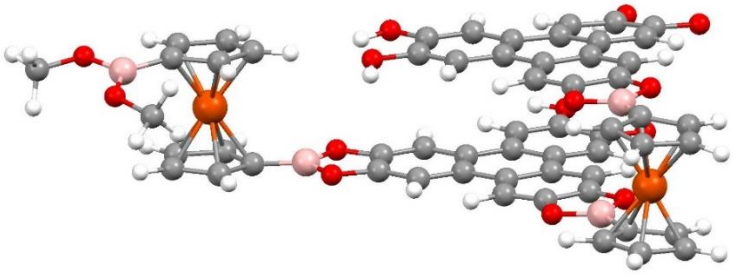


***Syn*-2-HHTP$_2$**

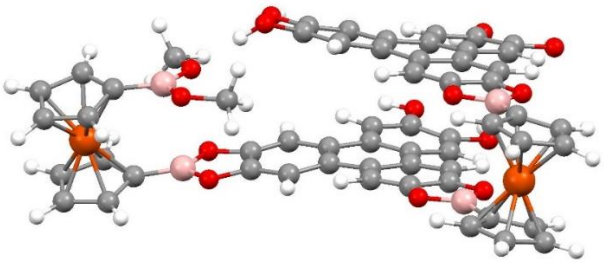
Figure S12. DFT-optimized structures considered on the course of the **Fc-cage** formation (part 1).



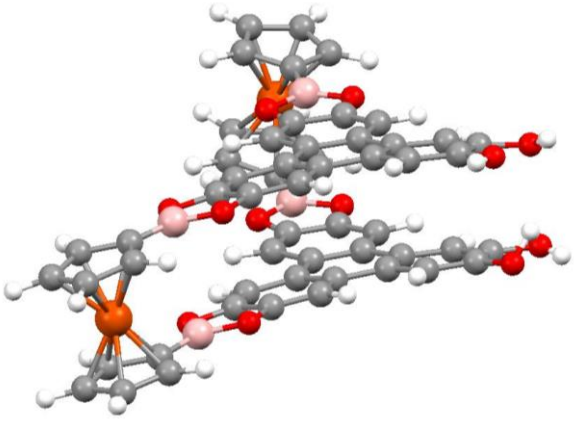
***anti*-2-HHTP$_2$**



2$_2$-HHTP$_2$ – A1



2$_2$-HHTP$_2$ – A2



2$_2$-HHTP$_2$

Figure S13. DFT-optimized structures considered on the course of the **Fc-cage** formation (part 2).

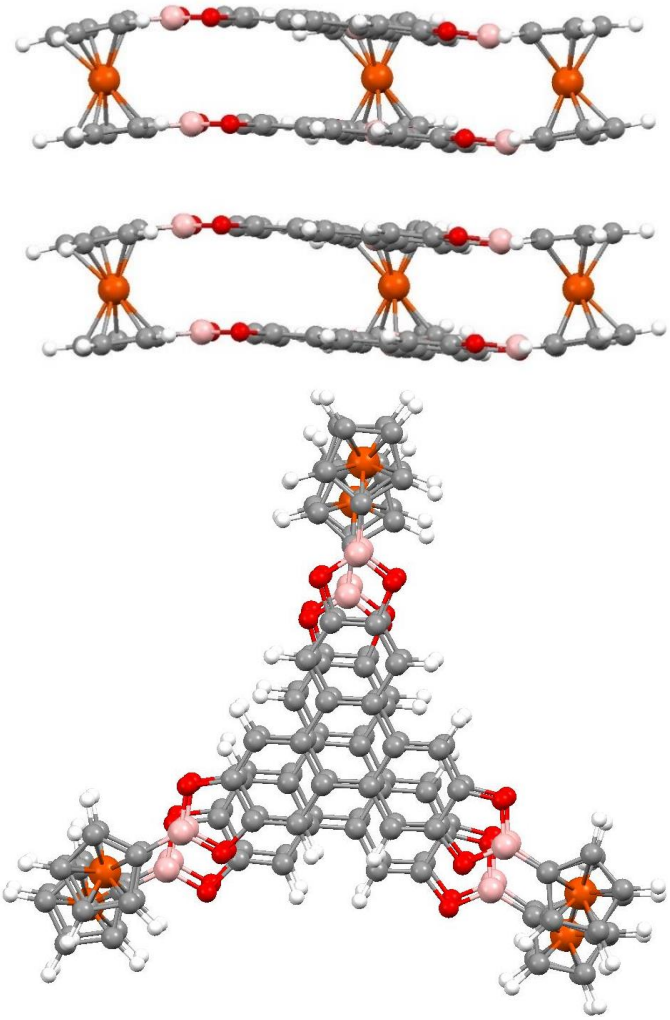


Figure S14. DFT-optimized **FcB-cage** dimer viewed from two perspectives.

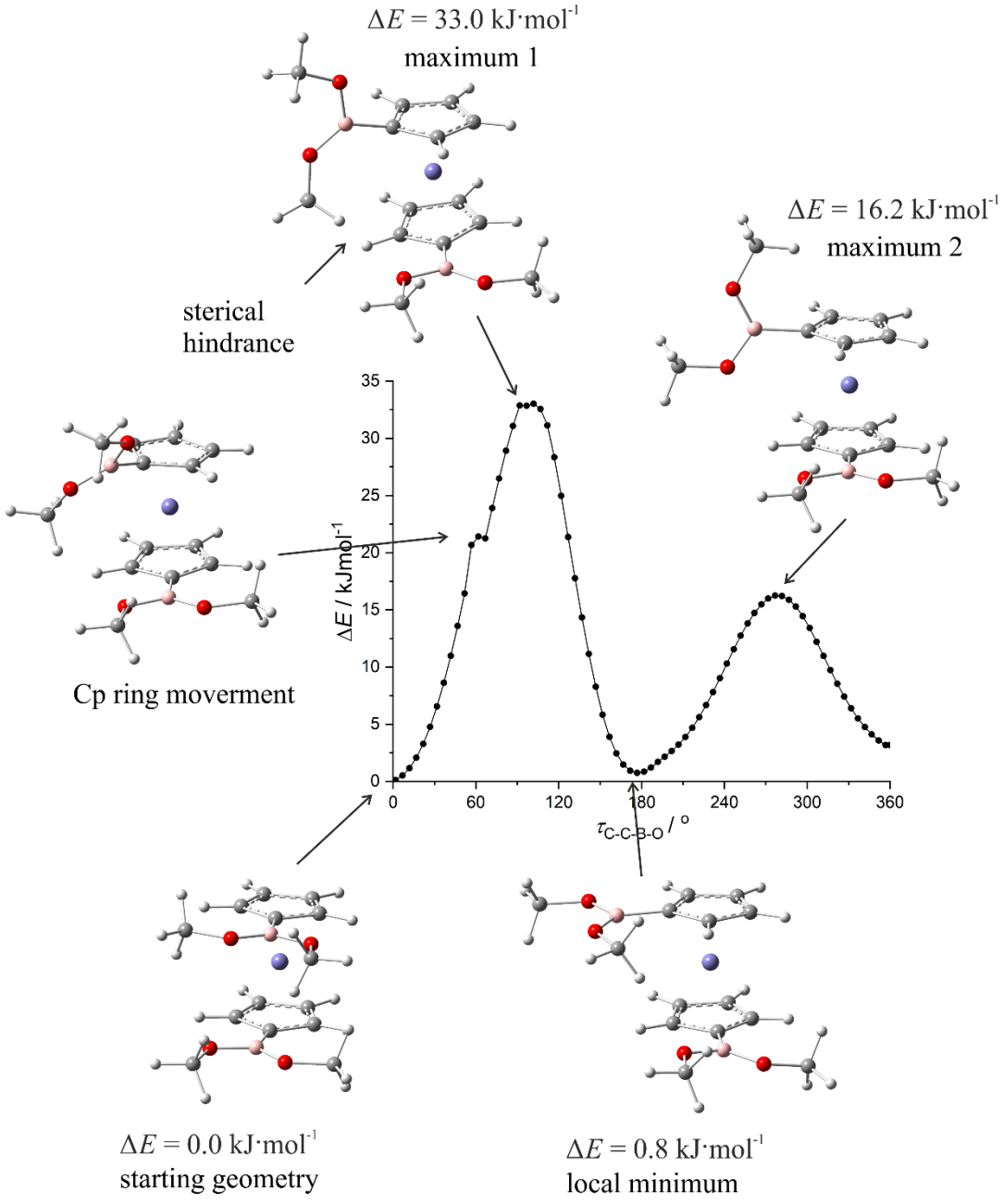


Figure S15. τ-Constrained optimization performed at the B97D/6-311++G(d,p) level of theory.

Comment: According to expectations, the most stable is the flat conformation. The $\tau_{C\text{-}C\text{-}B\text{-}O}$ constrained energy scan shows that there are two maxima corresponding to the perpendicular orientations of boronic ester groups with respect to Cp ring. The calculated rotation barrier is 33 kJmol-1, thus it is higher with respect to phenyl boronic acid (17 kJmol^{-1}, *Cryst. Growth Des.* **2012**, *12,* 3720−3734)., It originates from sterical hindrance of *anti*-oriented B-OMe group and Cp ring. The close proximity of Cp ring and *syn*-oriented B-OMe group does not produce such congestion and therefore the energy of second maximum of 16 kJmol^{-1} is comparable to the rotation barrier in phenylboronic acid.

7. PXRD experiments

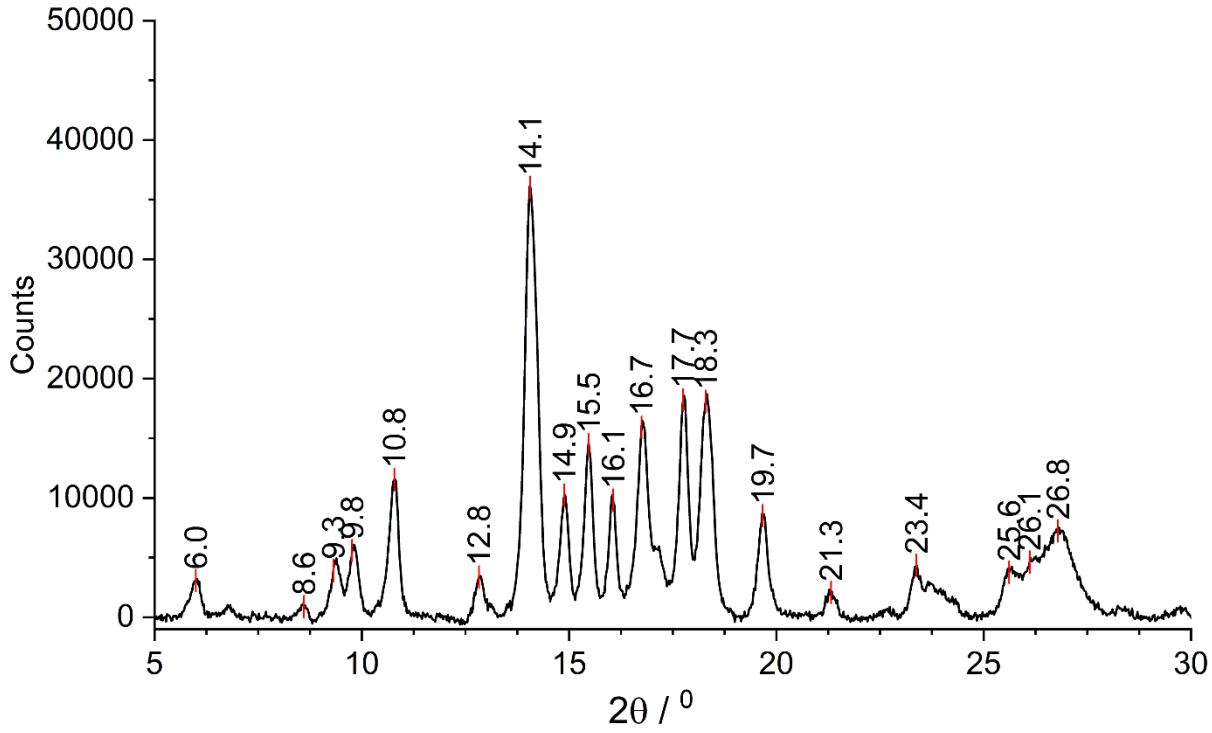


Figure S16. PXRD pattern of **FcB-cage** (background subtracted).

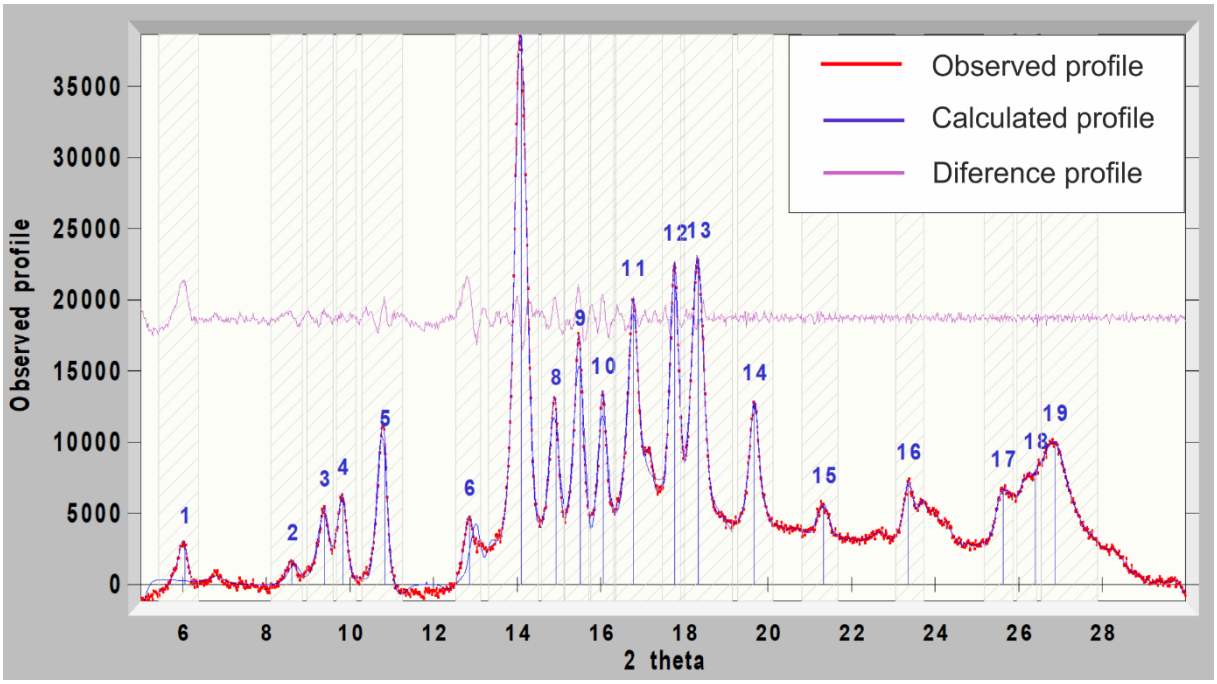


Figure S17. Results of the Pawley refinement for the monoclinic unit cell.

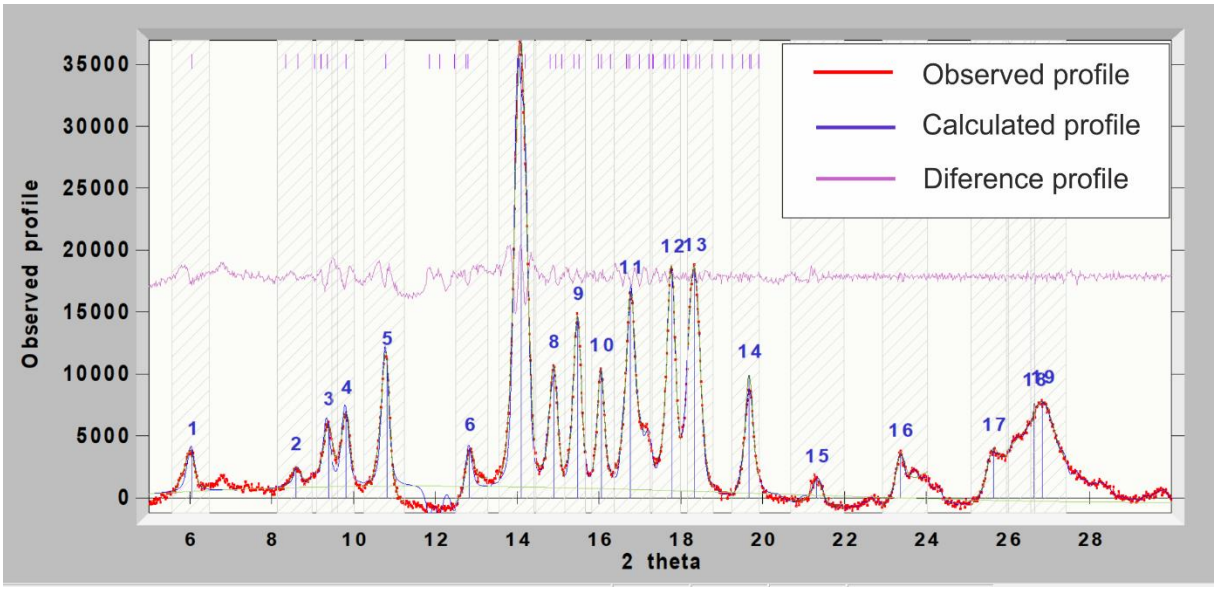


Figure S18. Results of the Pawley refinement for the triclinic unit cell.

Table S4. Summary of the Pawley refinement for proposed monoclinic and triclinic unit cells.

Crystal system	monoclinic	triclinic
Space group	*P*2	*P*1
a / Å	21.016	17.27
b / Å	18.078	12.56
c / Å	10.270	11.87
α / °	90	58.2
β / °	115.48	118.1
γ / °	90	117.3
Volume / Å^3	3522.4	1851.8
Radiation	CuK_α (λ = 1.54178)	CuK_α (λ = 1.54178)
Data/parameters	1164/154	1164/137
R_{wp}	9.7	13.0
R_{exp}	5.6	8.3
X^2	3.0	2.6

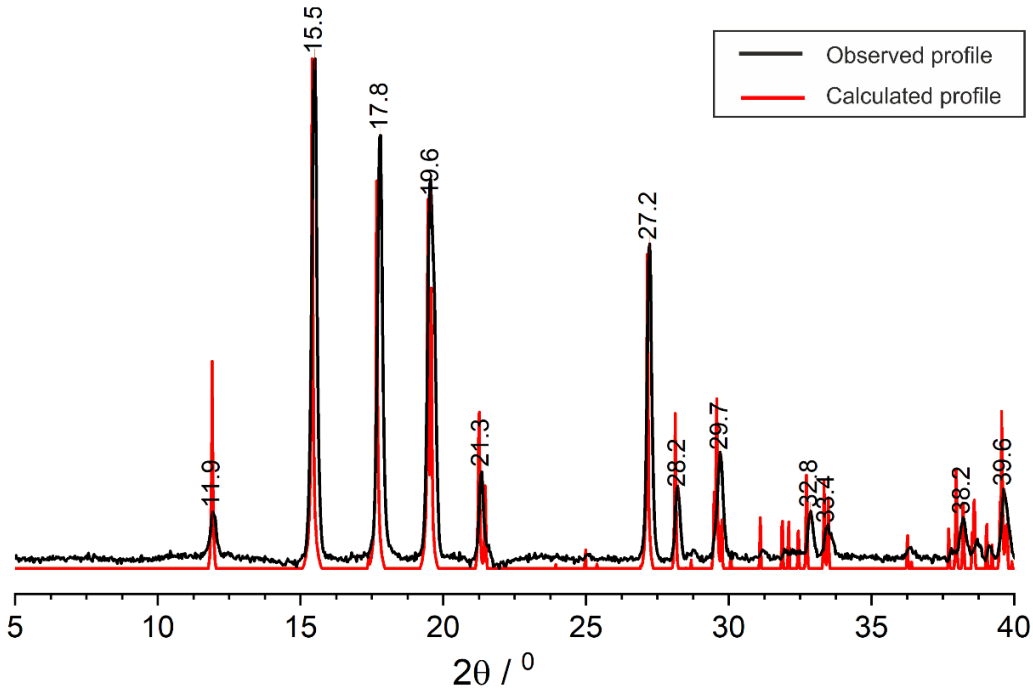


Figure S19. PXRD pattern of **1** together with its simulated powder pattern from X-ray crystal structure (M.Bolte, CSD Communication, 2011, REFCODE: OBEBUU, spacegroup: $P2_1/c$).

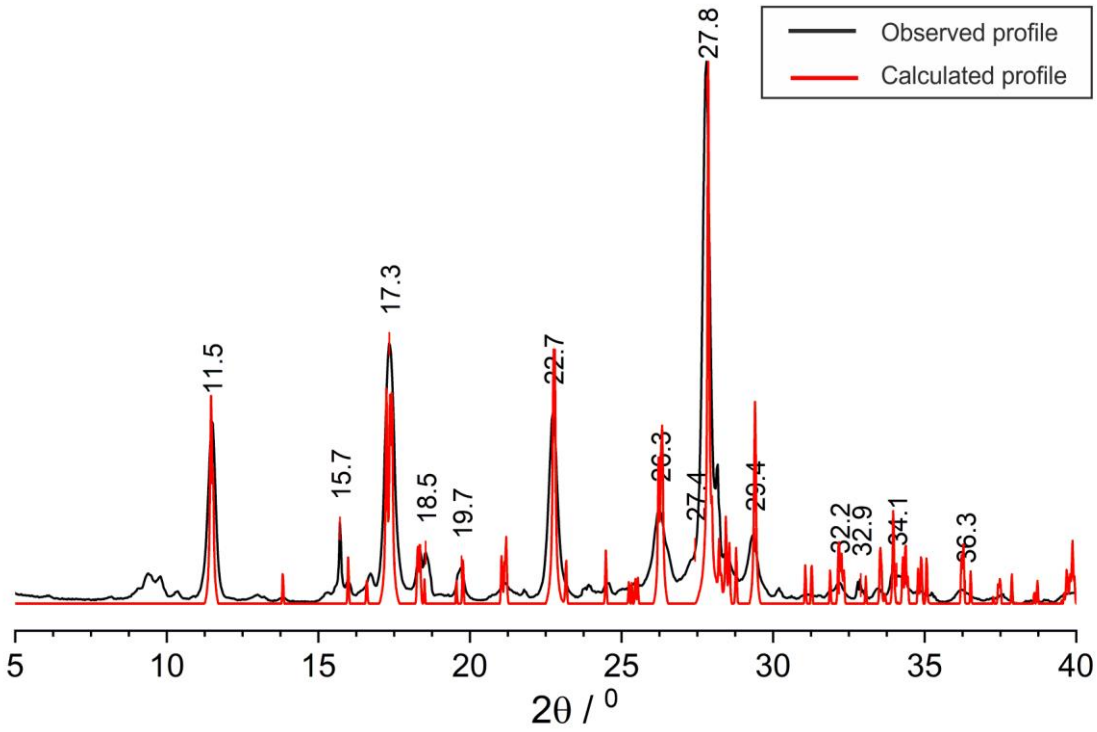


Figure S20. PXRD pattern of **HHTP** (commercial sample) together with its simulated powder pattern from X-ray crystal structure (T.L.Andresen, F.C.Krebs, N.Thorup, K.Bechgaard, Chem.Mater., 2000, 12, 2428, REFCODE: XEFSIK, spacegroup: $P2_1/c$).