

Supporting Information for

Supramolecular Cages as Differential Sensors for Dicarboxylate Anions: Guest Length Sensing Using Principal Component Analysis of ESI-MS and ¹H-NMR Raw Data

Authors: Carlo Bravin,¹ Andrea Guidetti,¹ Giulia Licini¹ and Cristiano Zonta^{1*}

¹ Department of Chemical Sciences, University of Padova, via Marzolo 1, 35131 Padova

*Correspondence to: cristiano.zonta@unipd.it

Contents

Supramolecular Cages as Tool for Sensing of Guest Length Using Raw Data Analysis	1
1 General Methods	3
2 Synthesis and Characterization	4
3 Result and Discussions	5
3.1 General procedure for DCL formation with E-P-X linkers in the presence of the guests series	5
3.1.1 ESI-MS spectra.....	6
3.1.2 Selectivity profile of the $C_n@(R-R'-R'')$ cages	12
3.1.3 1H NMR spectra	13
3.2 General procedure for Data analysis with PCA	18
3.2.1 PCA of ESI-MS spectra	19
3.2.2 PCA of 1H -NMR spectra	21
3.3 Prediction of an unknown by projection in the PCA space.....	24
3.3.1 PCA of ESI-MS spectra with additional data.....	25
3.3.2 PCA of 1H -NMR spectra	26

1 General Methods

NMR spectra were recorded at 301 K on Bruker Avance-300. All the ^1H NMR spectra were referenced to residual isotopic impurity of DMSO- d_6 (2.50 ppm) or CD_3CN (1.98 ppm). The following abbreviations are used in reporting the multiplicity for NMR resonances; s=single, d=doublet, t= triplet, and m=multiplet. The NMR data were processed using MestReNova 10.0.2.

Low resolution electrospray ionization mass spectrometry LRMS (ESI-MS) experiments were carried out in positive mode with Agilent Technologies LC/MSD Trap SL AGILENT instrument (mobile phase Acetonitrile) with the following settings: Nebulizer = 20 psi, Dry gas = 5 L/min, Dry Temp = 325°C. The samples were infused from a syringe pump at the rate of 20 $\mu\text{L}/\text{min}$. MS peak intensity for each analysis was reported as monoisotopic mass and the data were processed with Data Explorer 4.2 and Openchrom 1.2.0 Community Edition. Principal Component Analysis were performed with MatLab R2017a.

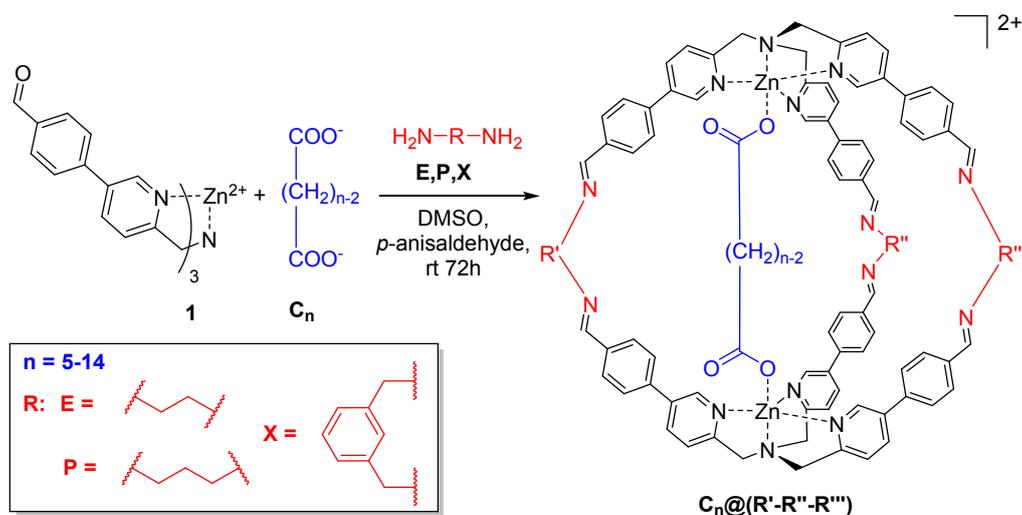
Chemicals were purchased from Sigma-Aldrich, TCI, or Acros and used without further purification.

2 Synthesis and Characterization

Compounds $C_n@E-E-E$, $C_n@P-P-P$, $C_n@X-X-X$ and complex **1** have been synthesized and fully characterized (1H NMR, 2D NMR, MS analysis) accordingly to: C. Bravin, E. Badetti, F.A. Scaramuzzo, G. Licini, C. Zonta. *J. Am. Chem. Soc.* **2017**; *139*, 6456-6460 and C. Bravin, E. Badetti, R. Puttreddy, F. Pan, K. Rissanen G. Licini, C. Zonta. *Chem. Eur. J.* **2018**; *24*, 2936.

3 Result and Discussions

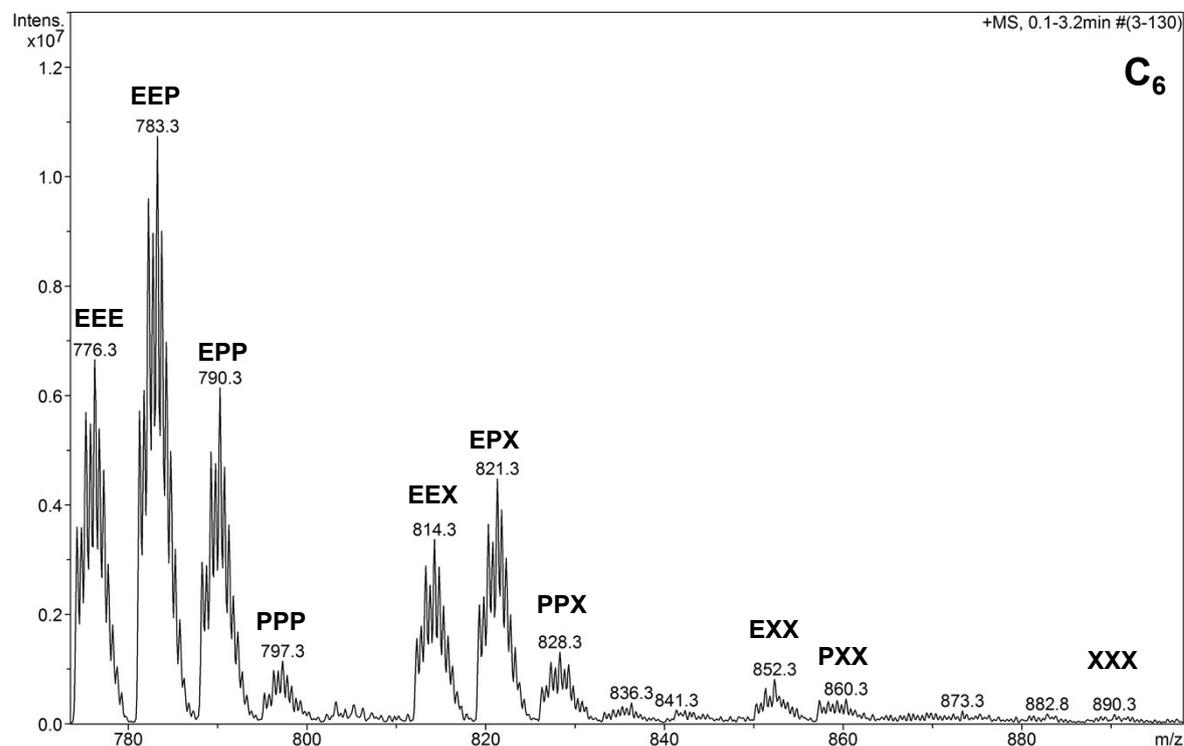
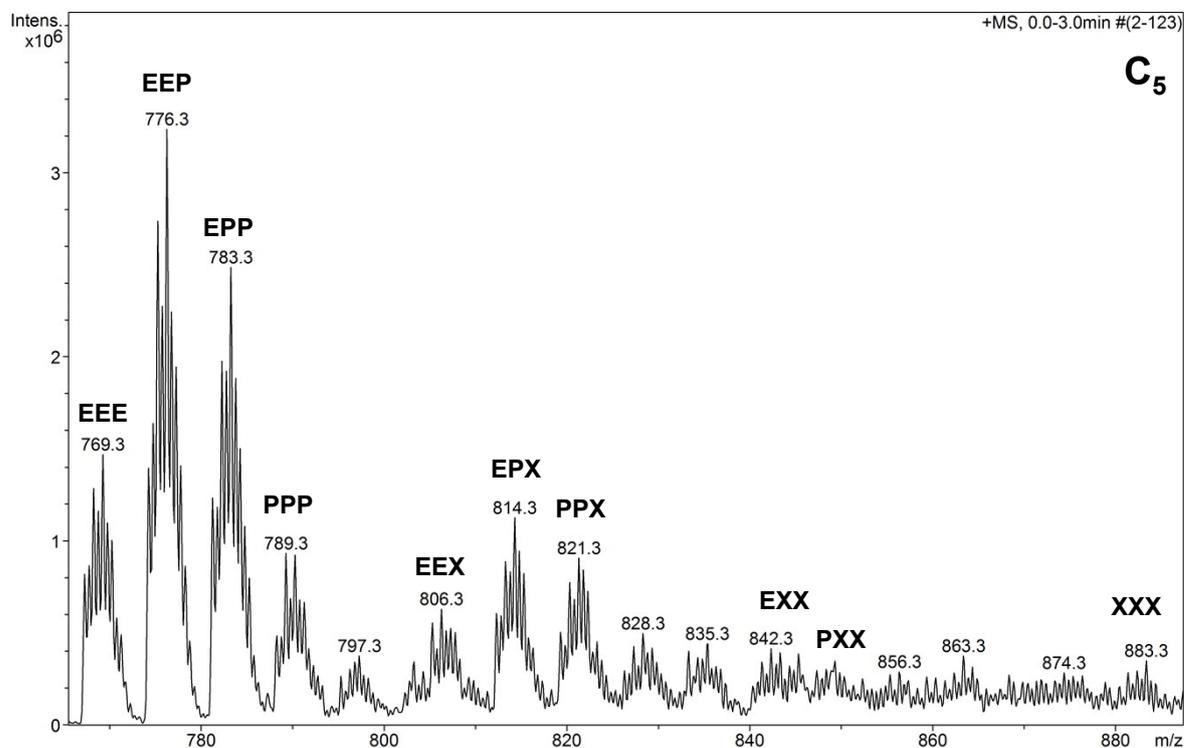
3.1 General procedure for DCL formation with E-P-X linkers in the presence of the guests series

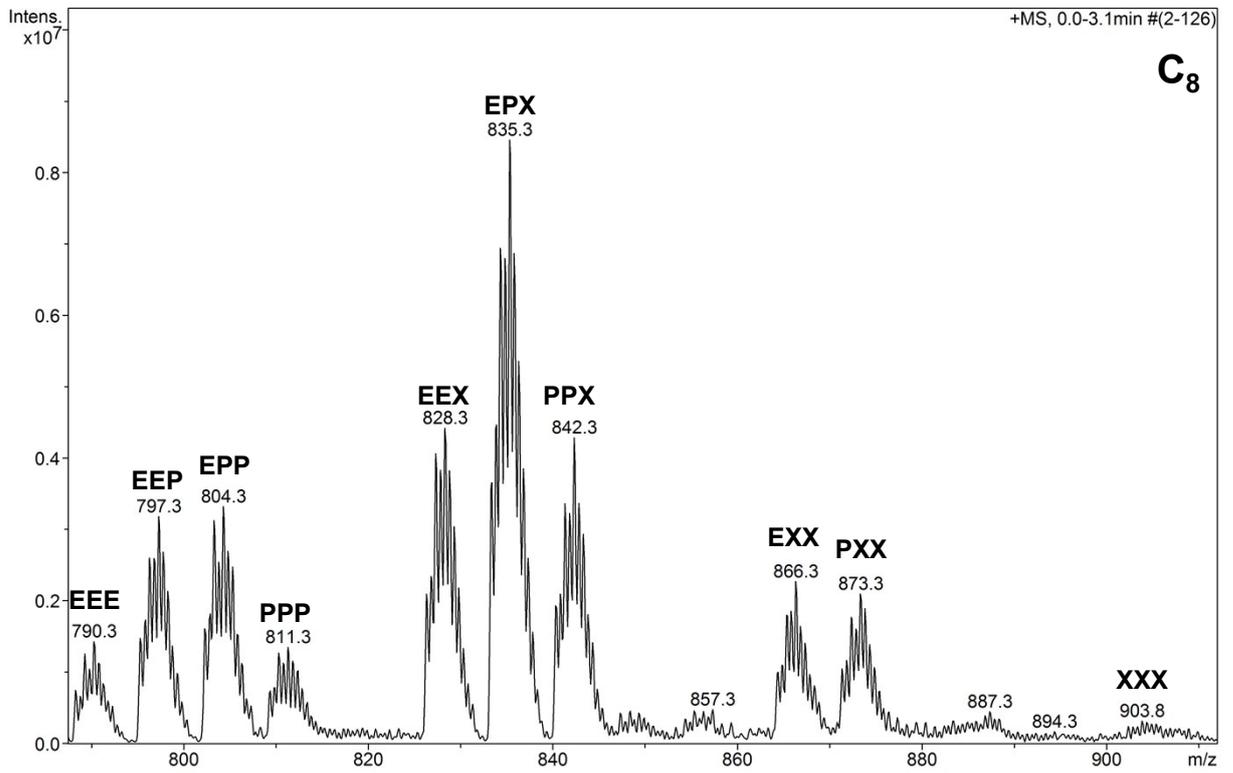
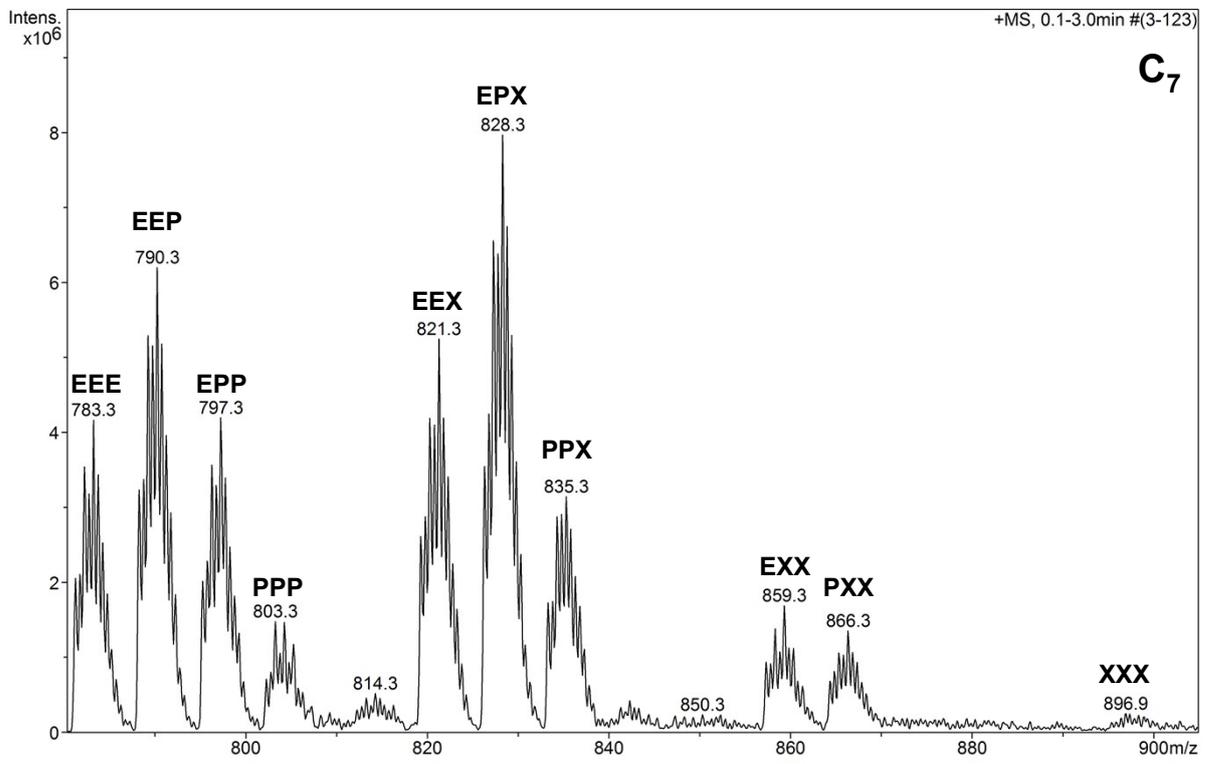


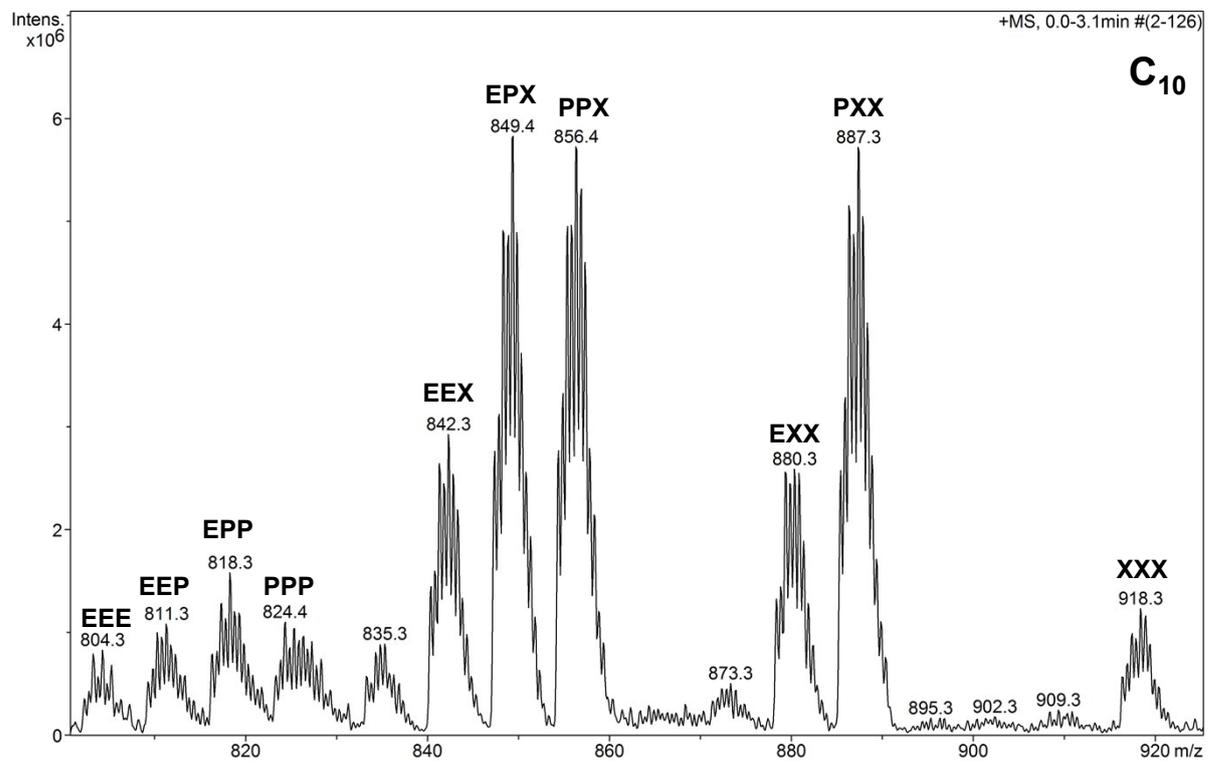
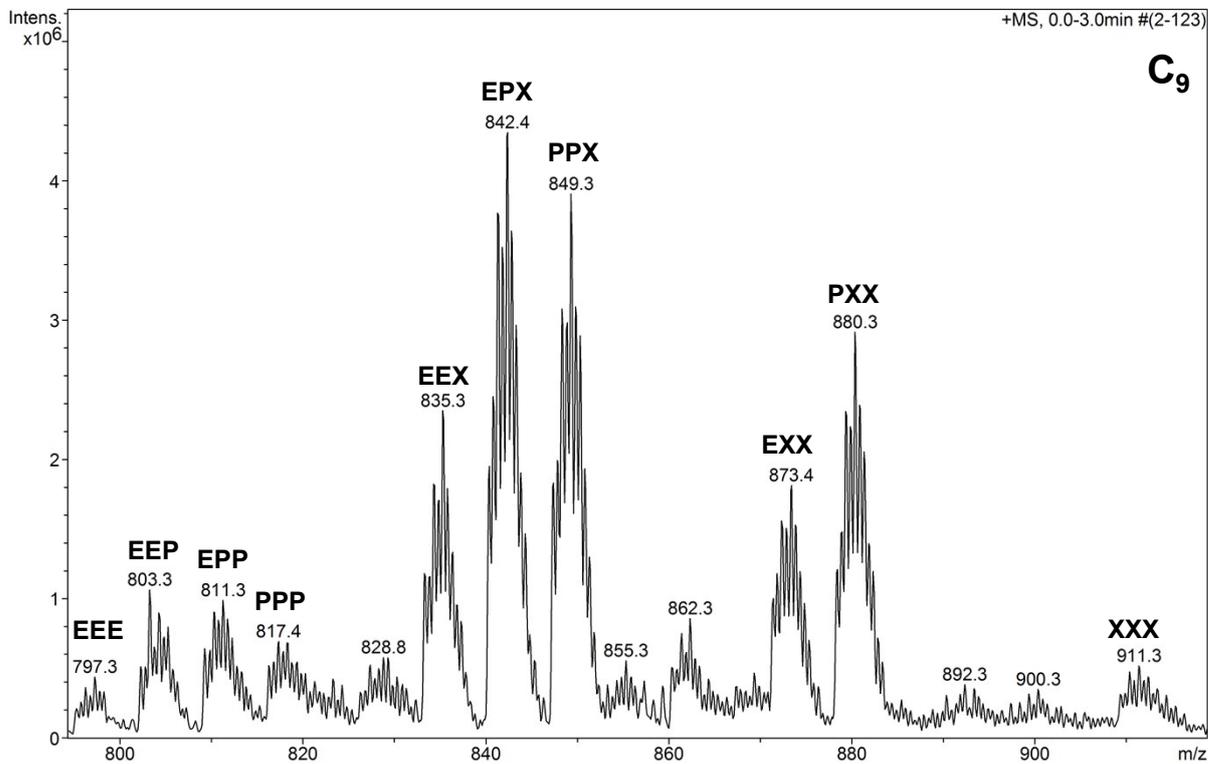
In a small flask were added 500 μl of a 0.002 M solution of complex **1** (0.001 mmol, 2 eq.), 25 μl of a 0.02 M solution of dicarboxylate **C_n** (0.0005 mmol, 1 eq.), 30 μl of a 0.2 M solution of *p*-anisaldehyde (0.006 mmol, 12 eq.) and 75 μl of a 0.02 M solution containing diamines **E**, **P** and **X** (0.0015 mmol, 3 eq. for each diamine) in DMSO. The mixture was left 3 days at room temperature and checked via ESI-MS (50 μl of the experiment mixture diluted in 3 ml of CH_3CN). The same experimental procedure was adopted for NMR experiment preparing the solution in $\text{DMSO-}d_6$.

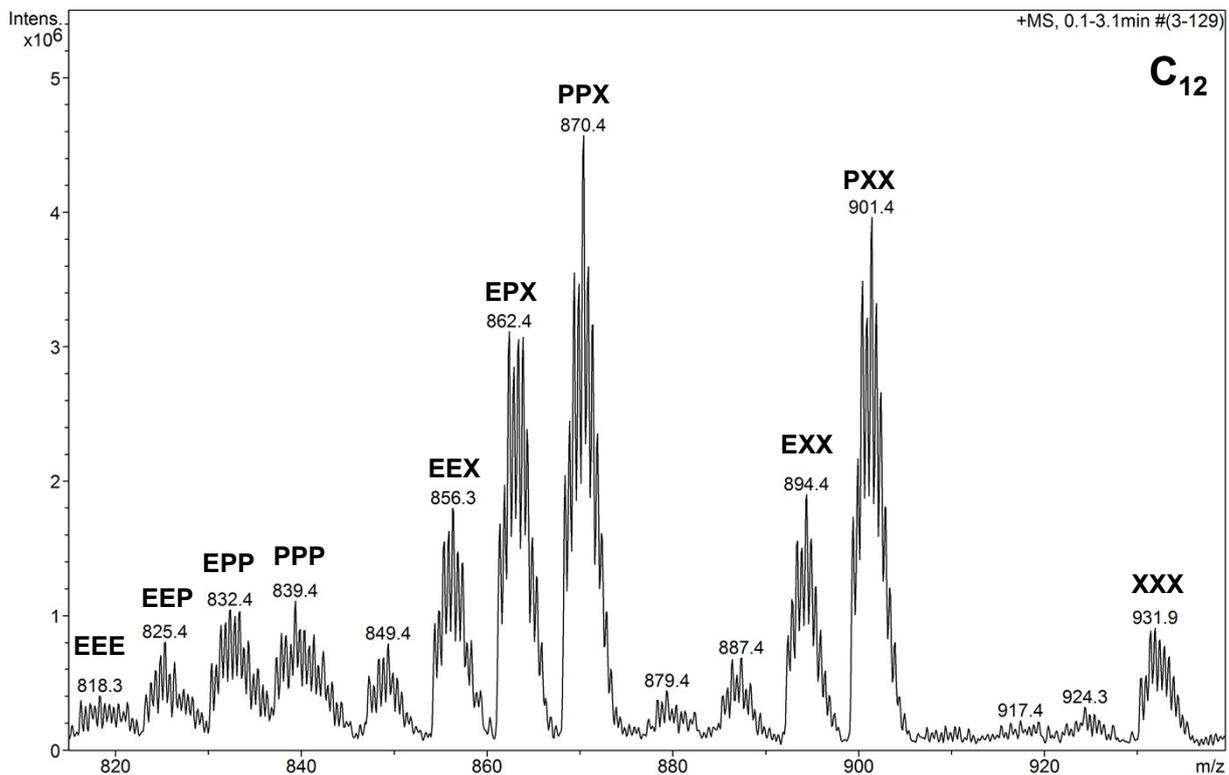
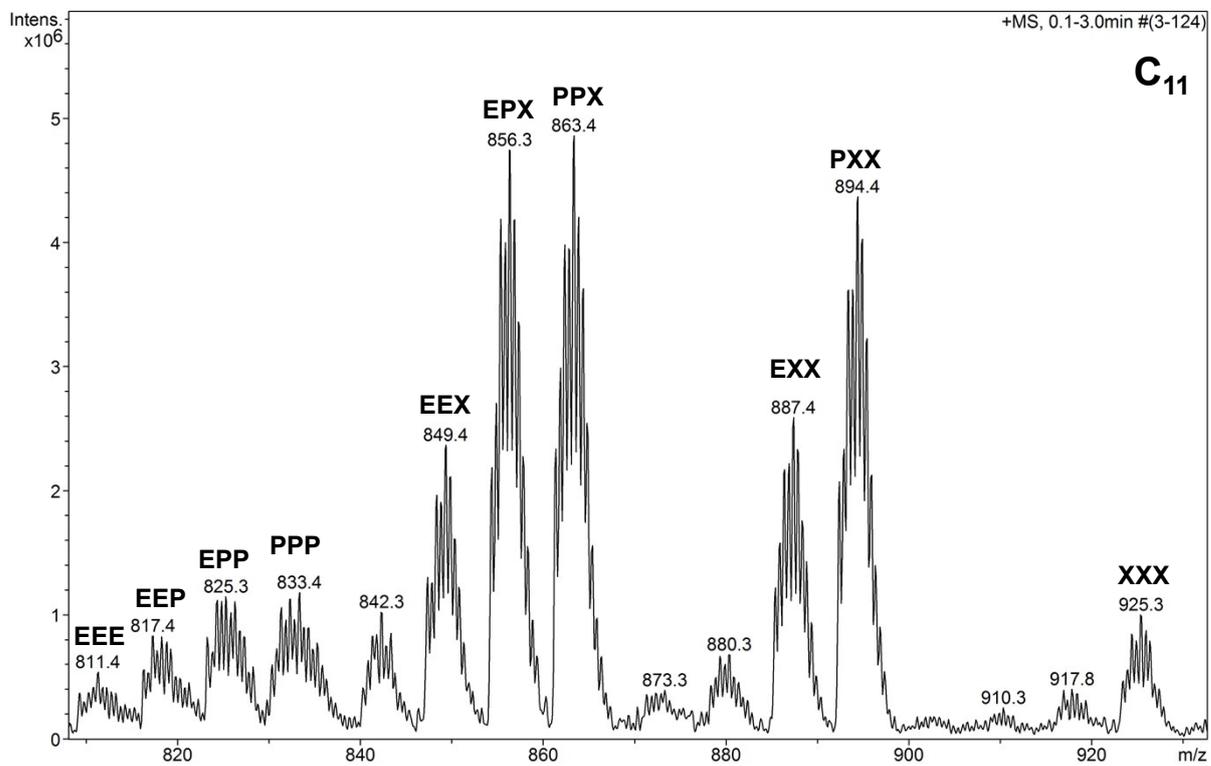
3.1.1 ESI-MS spectra

The spectral window containing all the di-charged peaks related to the cage species $C_n@1(R-R'-R'')$ for each guest of the C_n series. The symbols **E**, **P** and **X** represent the number of diamine linkers present in the structure (**E**=EDA, **P**=PDA, **X**=mXDA) while C_n represents the guest used in each experiment. In Figure S1 are reported the ESI MS spectra of the DCL formation while in Figure S2 are reported the corresponding selectivity profile for each experiment considering the relative intensity of the m/z peaks (Table S1). In Figure S3 are reported the corresponding 1H NMR spectra.









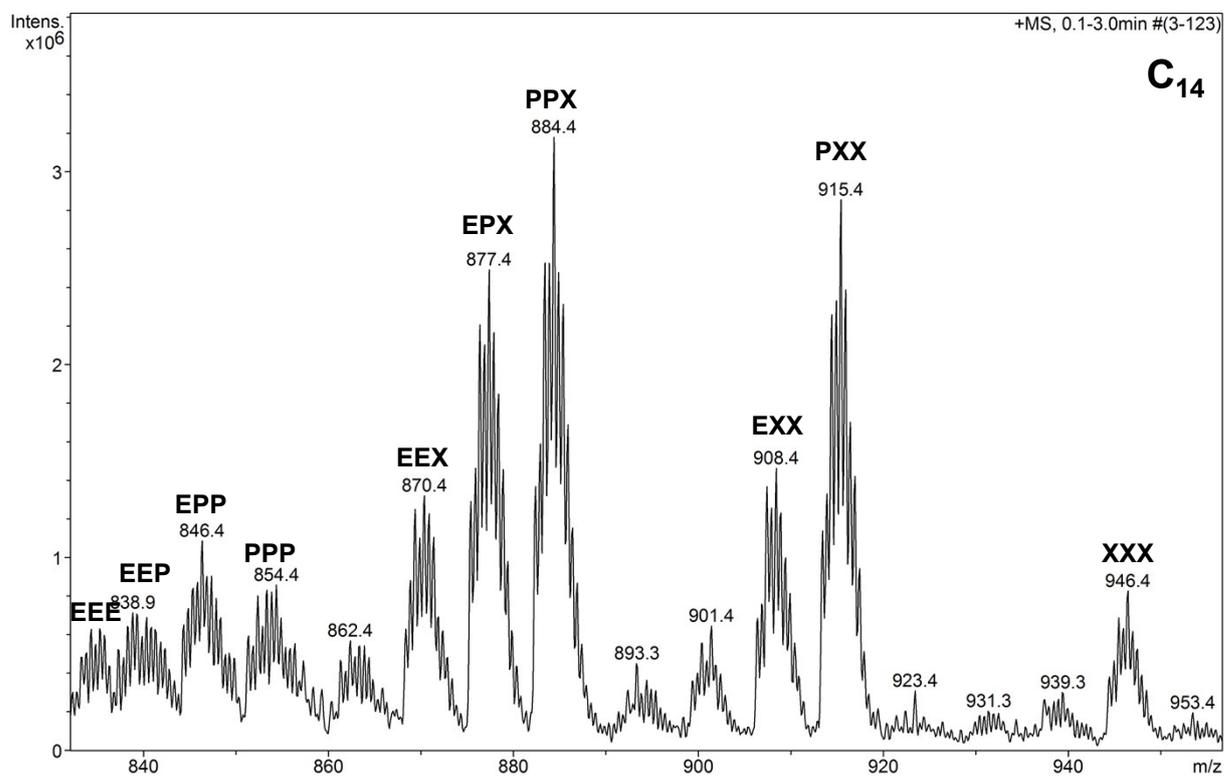
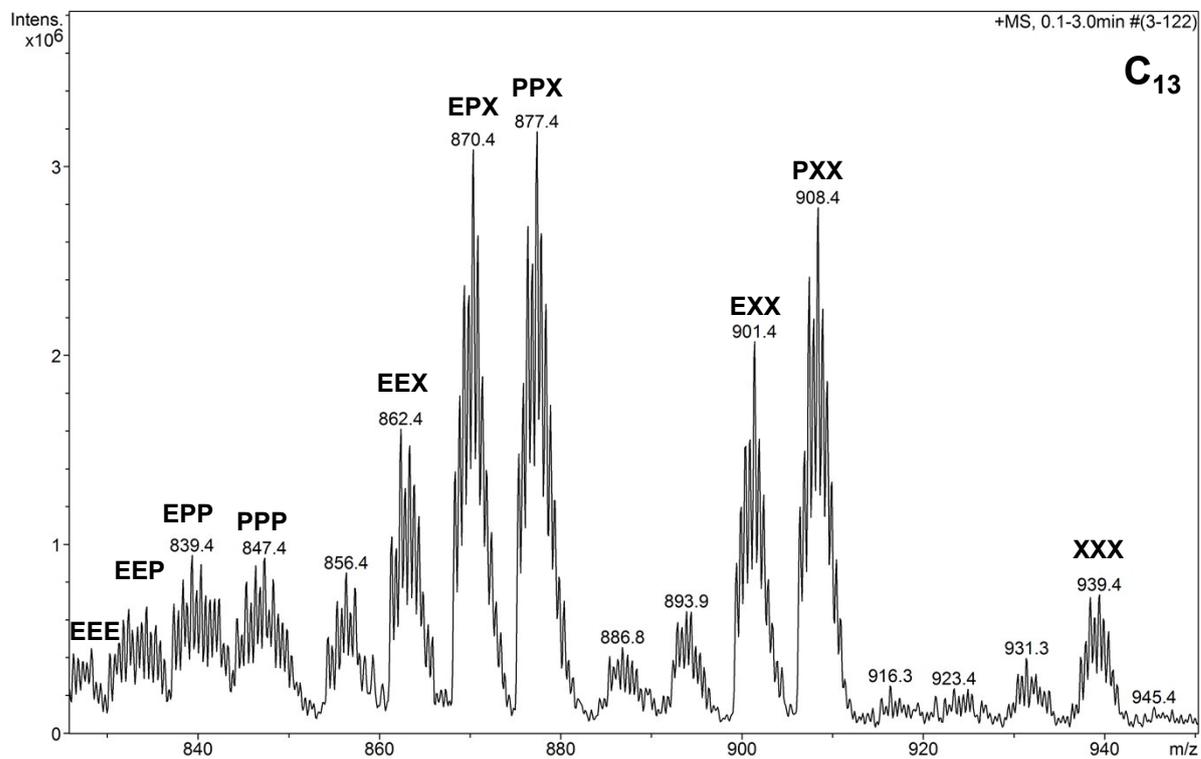


Figure S1 a) ESI-MS spectra of the E+P+X linkers experiments towards the guests series C_5 - C_{14} .

The values of relative intensity $I_{C_n}/\sum I_{C_n}$ for each characteristic peak in the scrambling experiment and their standard deviation are reported in Table S1. The error reported for the ESI-MS experiment are in agreement with our previous work (see ref. [26] in the article). The measures were repeated four times. The values obtained are showed in the histograms showed in Figure S2.

		$I_{C_n}/\sum I_{C_n}$				
Guest	Cage	1(E EE)	1(E EP)	1(E PP)	1(P PP)	1(E EX)
	C ₅		0.131±0.006	0.264±0.006	0.210±0.010	0.080±0.007
C ₆		0.187±0.008	0.311±0.011	0.170±0.008	0.035±0.002	0.091±0.010
C ₇		0.110±0.004	0.174±0.007	0.118±0.008	0.039±0.003	0.147±0.012
C ₈		0.046±0.004	0.098±0.008	0.108±0.014	0.052±0.002	0.143±0.001
C ₉		0.025±0.003	0.047±0.002	0.058±0.006	0.045±0.005	0.119±0.005
C ₁₀		0.027±0.001	0.041±0.002	0.052±0.005	0.039±0.003	0.104±0.003
C ₁₁		0.027±0.002	0.040±0.003	0.054±0.004	0.051±0.006	0.094±0.004
C ₁₂		0.025±0.002	0.043±0.005	0.061±0.004	0.053±0.004	0.095±0.003
C ₁₃		0.025±0.002	0.042±0.003	0.061±0.006	0.052±0.005	0.099±0.003
C ₁₄		0.025±0.002	0.052±0.003	0.077±0.002	0.060±0.002	0.084±0.010
Guest	Cage	1(E PX)	1(P PX)	1(E XX)	1(P XX)	1(X XX)
	C ₅		0.097±0.007	0.073±0.001	0.036±0.006	0.026±0.002
C ₆		0.124±0.007	0.040±0.008	0.024±0.001	0.014±0.001	0.005±0.001
C ₇		0.222±0.009	0.095±0.002	0.045±0.004	0.040±0.002	0.009±0.001
C ₈		0.244±0.017	0.144±0.006	0.075±0.004	0.079±0.006	0.011±0.001
C ₉		0.213±0.012	0.207±0.006	0.106±0.003	0.150±0.019	0.030±0.005
C ₁₀		0.203±0.005	0.202±0.011	0.095±0.007	0.193±0.006	0.044±0.001
C ₁₁		0.197±0.002	0.212±0.015	0.105±0.009	0.176±0.005	0.045±0.002
C ₁₂		0.166±0.003	0.215±0.012	0.100±0.003	0.191±0.003	0.051±0.005
C ₁₃		0.177±0.008	0.194±0.006	0.130±0.004	0.168±0.004	0.051±0.001
C ₁₄		0.158±0.008	0.202±0.002	0.105±0.009	0.181±0.011	0.057±0.003

Table S1 Relative intensity values of inclusion complexes $C_n@(\mathbf{R}-\mathbf{R}'-\mathbf{R}'')$ and relative standard deviation, calculated as the average of three peaks of each cluster for all the competition experiments.

3.1.2 Selectivity profile of the $C_n@(\text{R-R}'\text{-R}'')$ cages

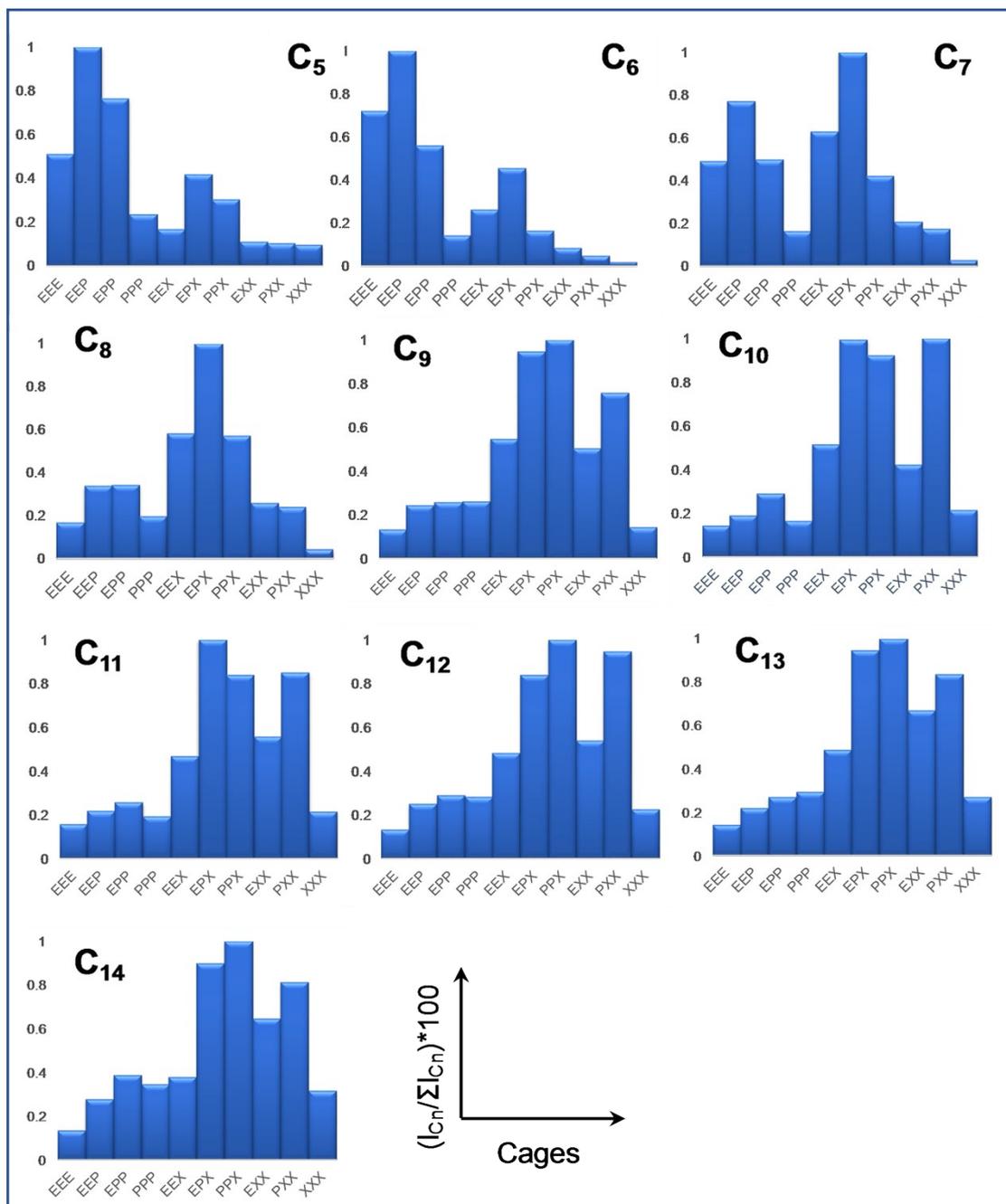
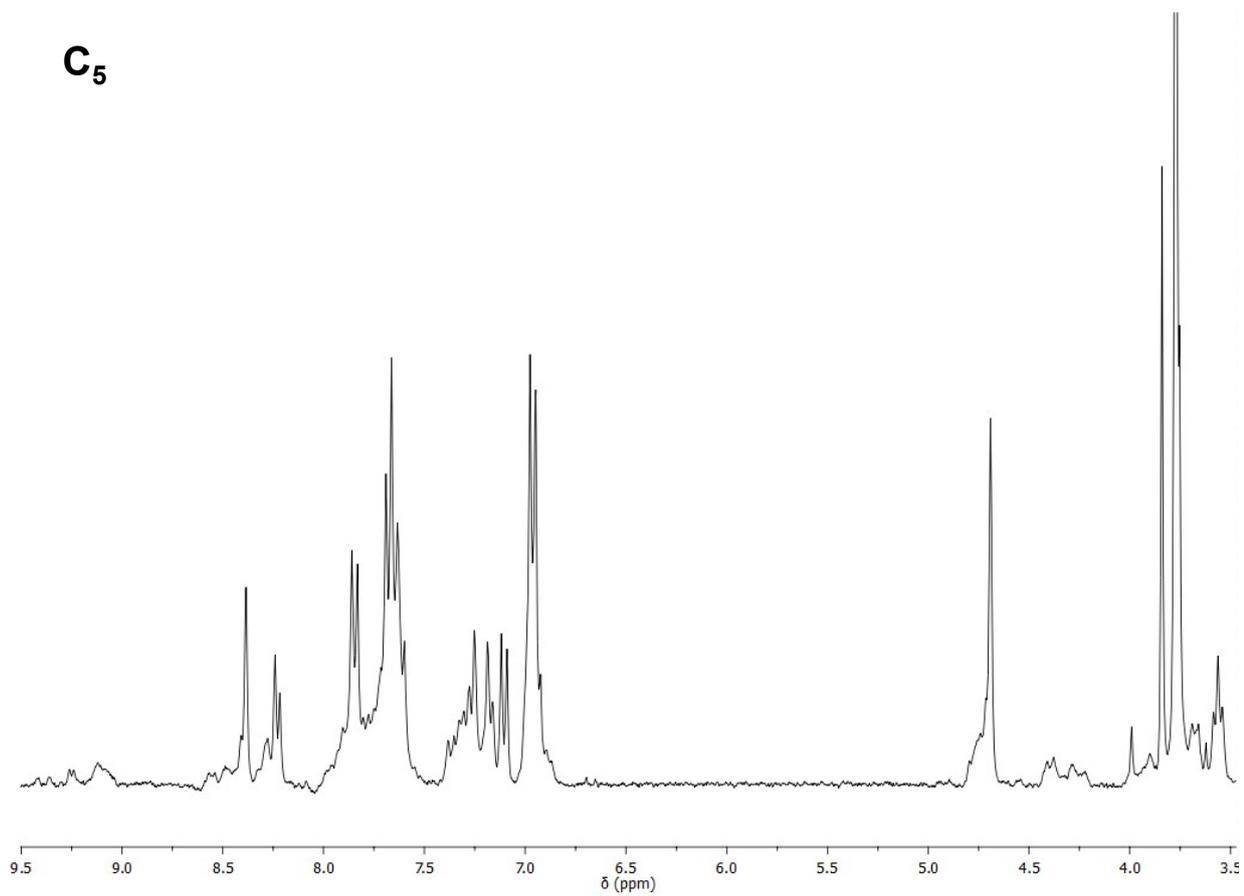


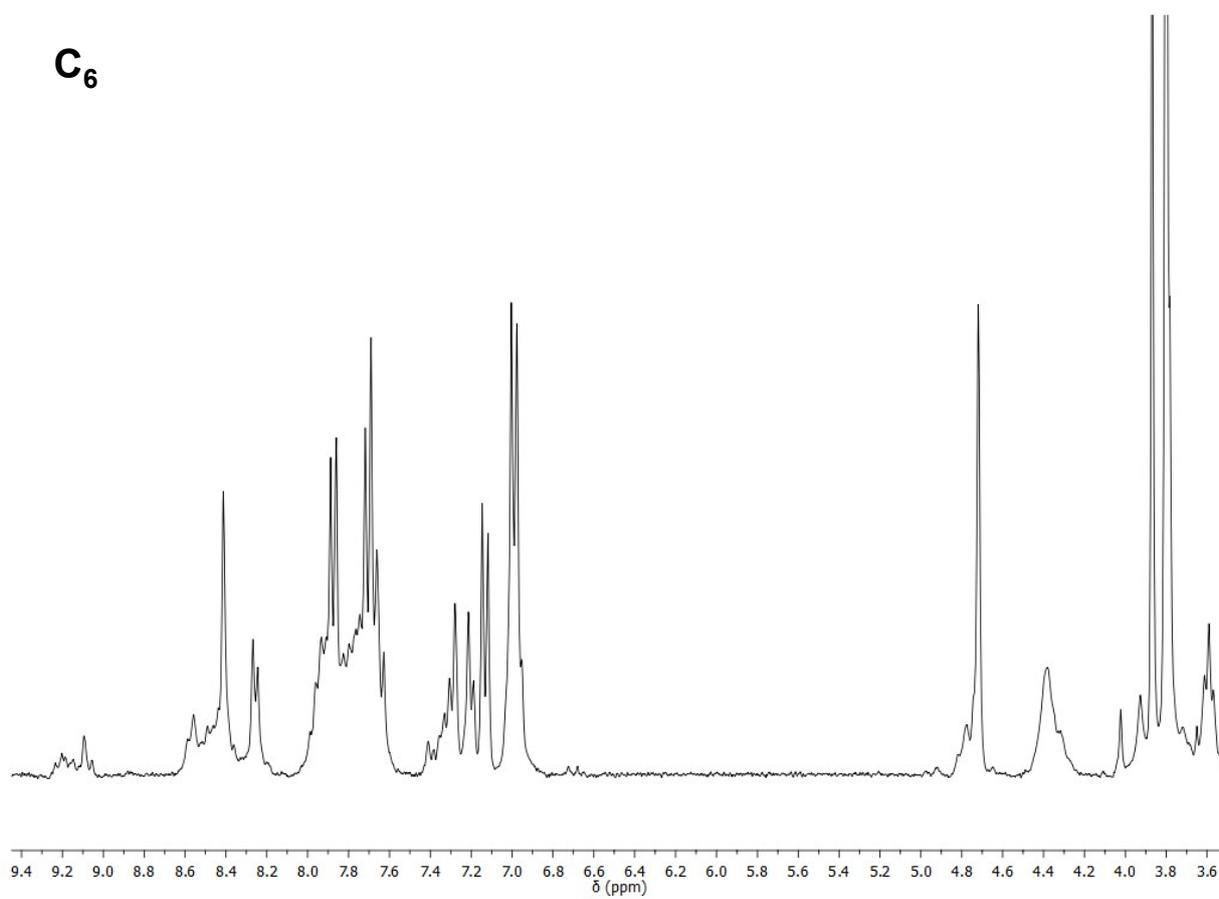
Figure S2 ESI-MS selectivity profiles for the E+P+X linkers experiments towards the guests series C₅-C₁₄.

3.1.3 ^1H NMR spectra

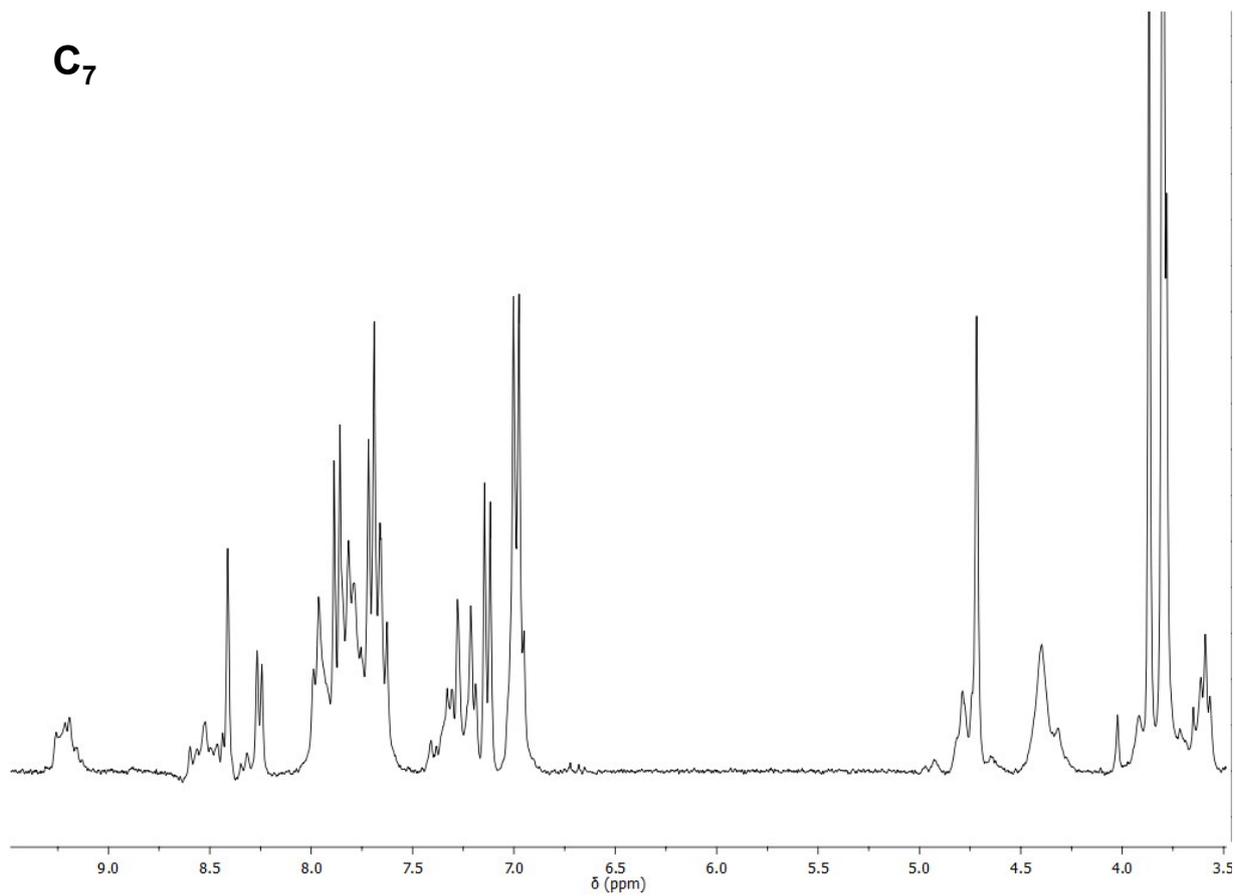
C₅



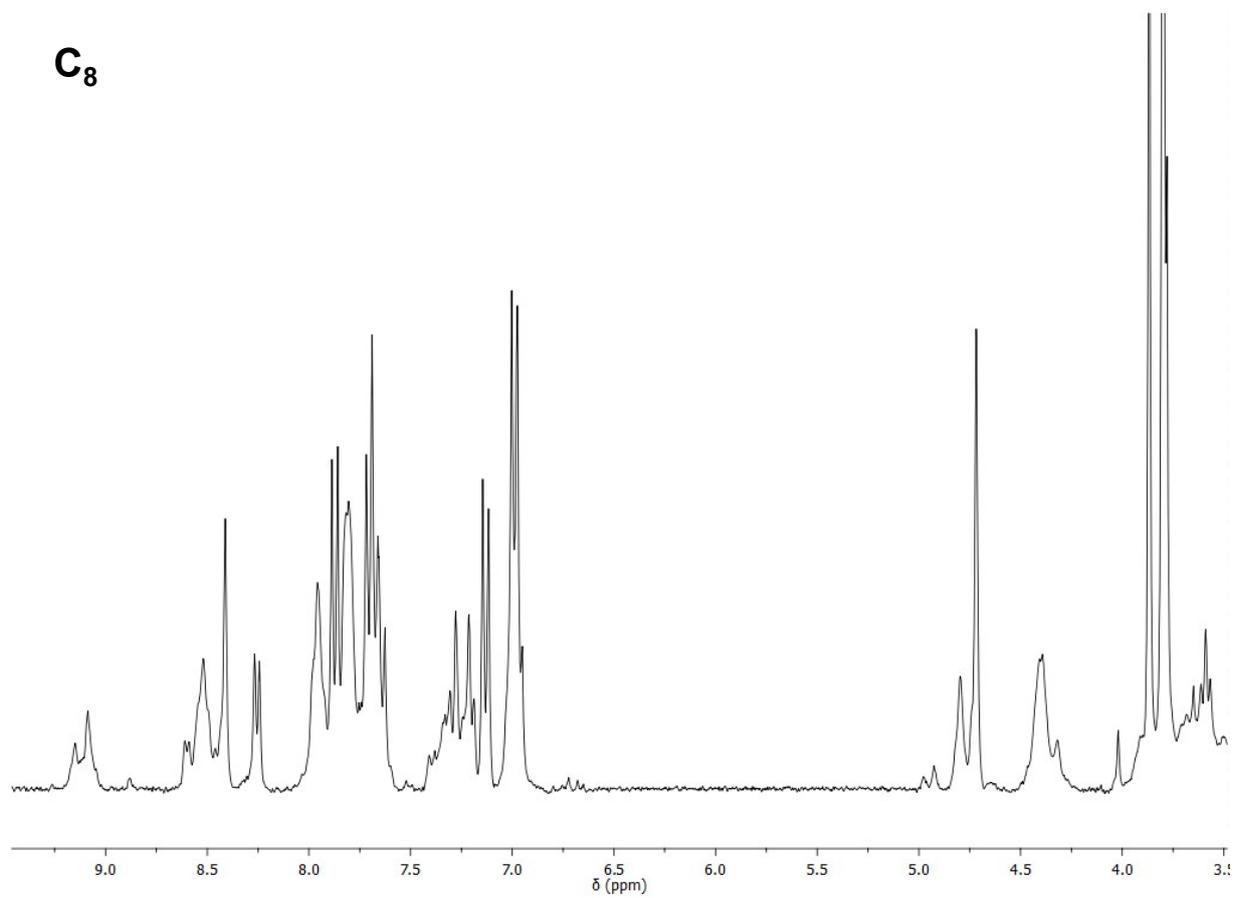
C₆



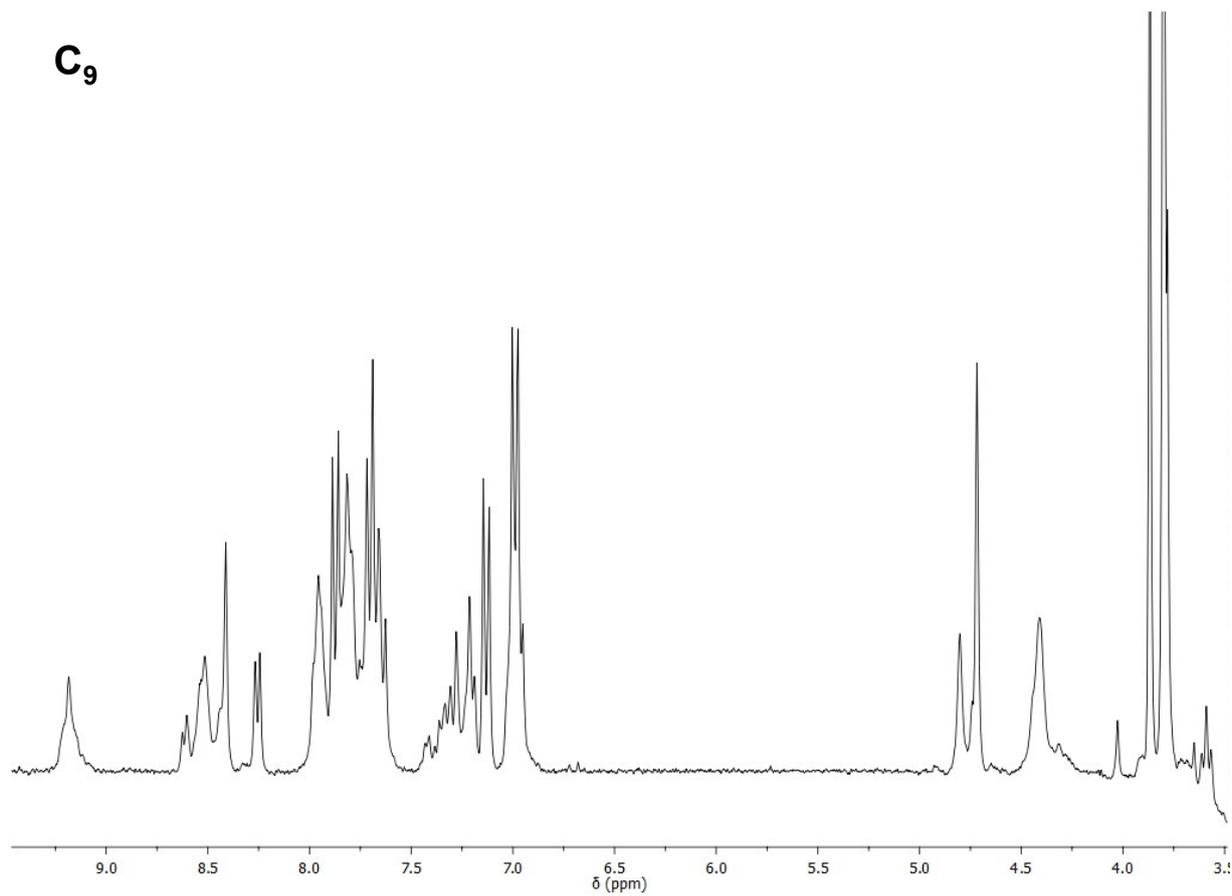
C₇



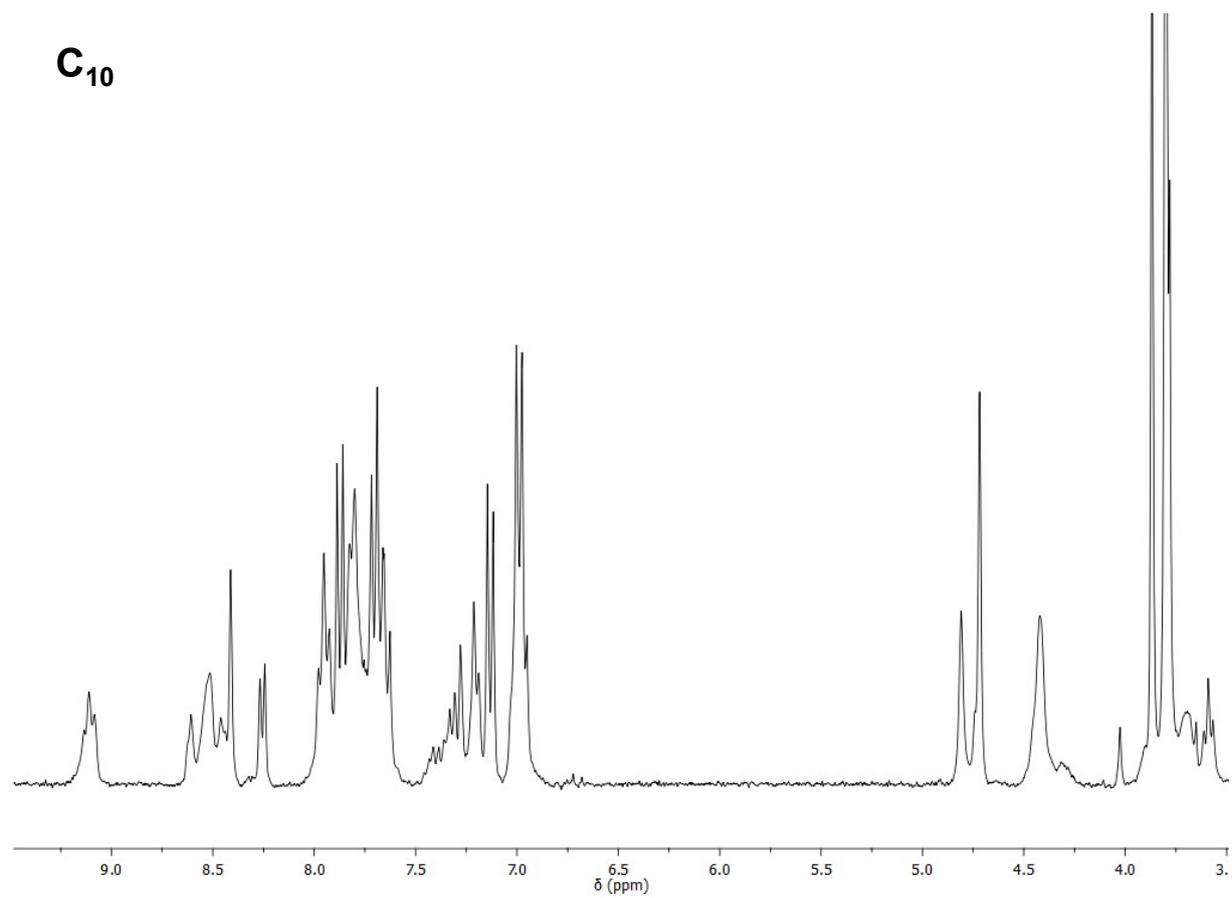
C₈



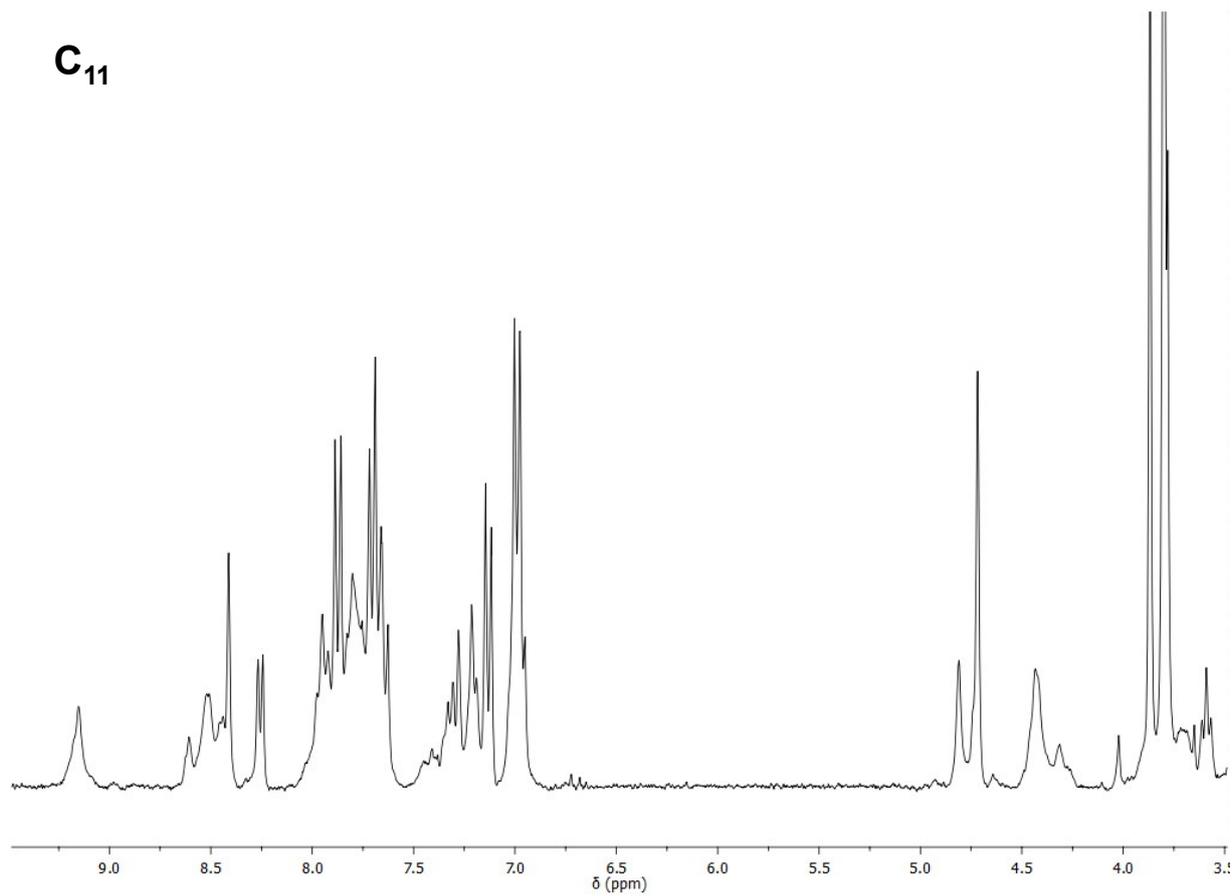
C₉



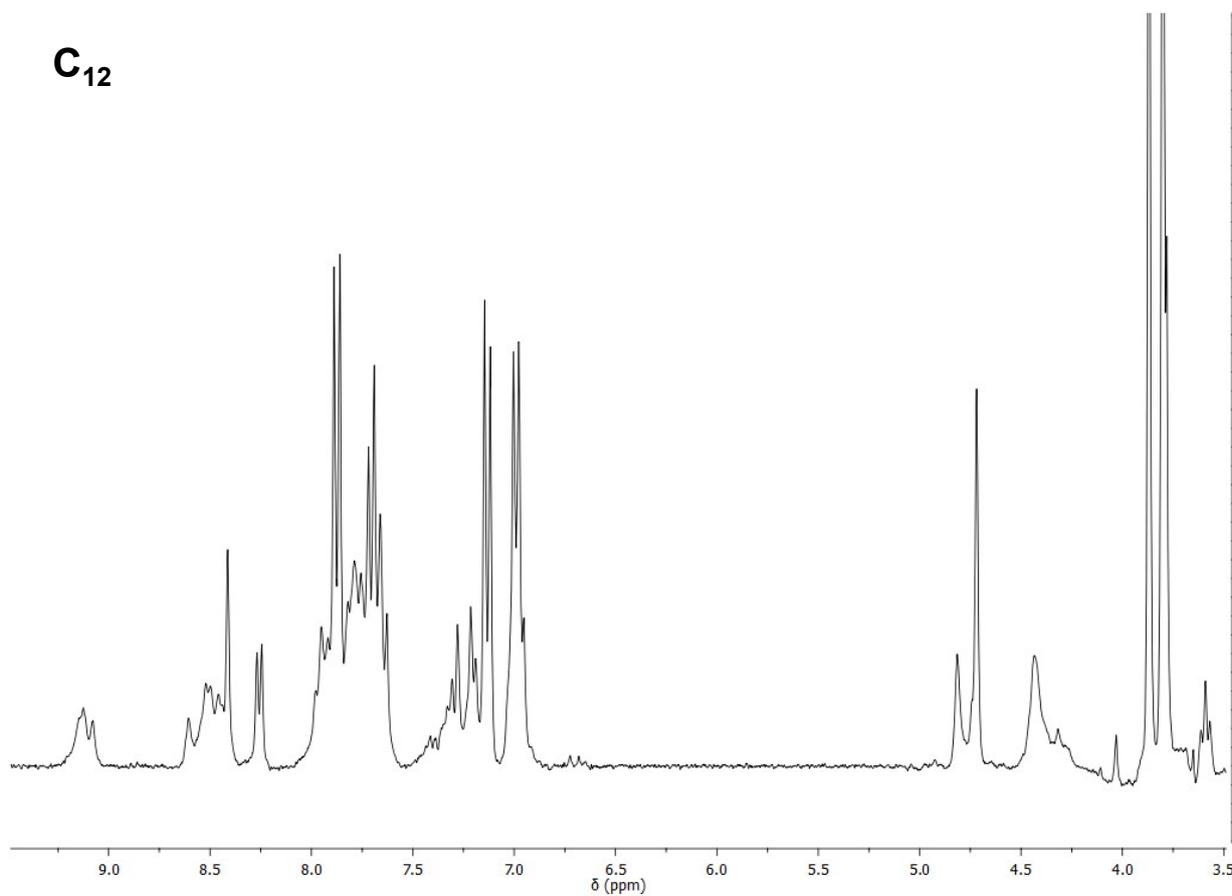
C₁₀



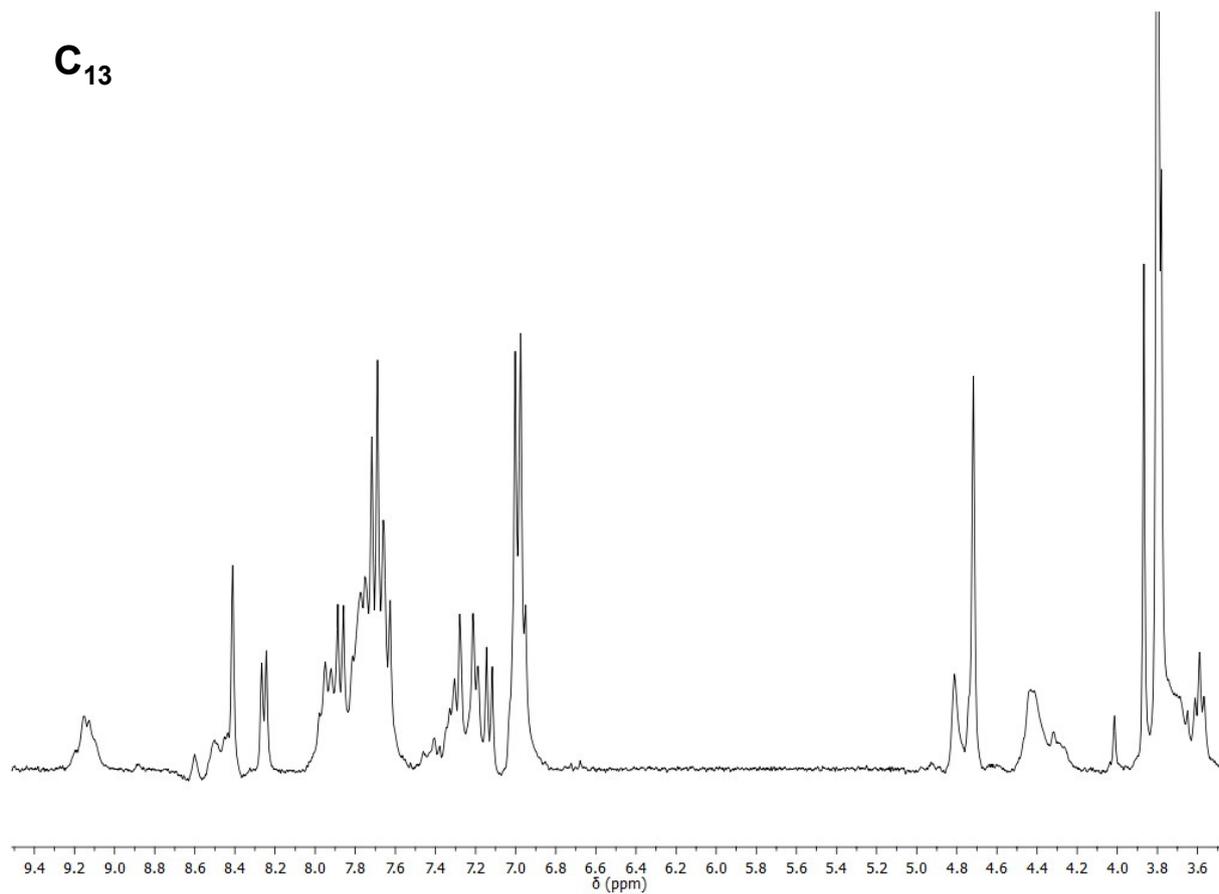
C₁₁



C₁₂



C₁₃



C₁₄

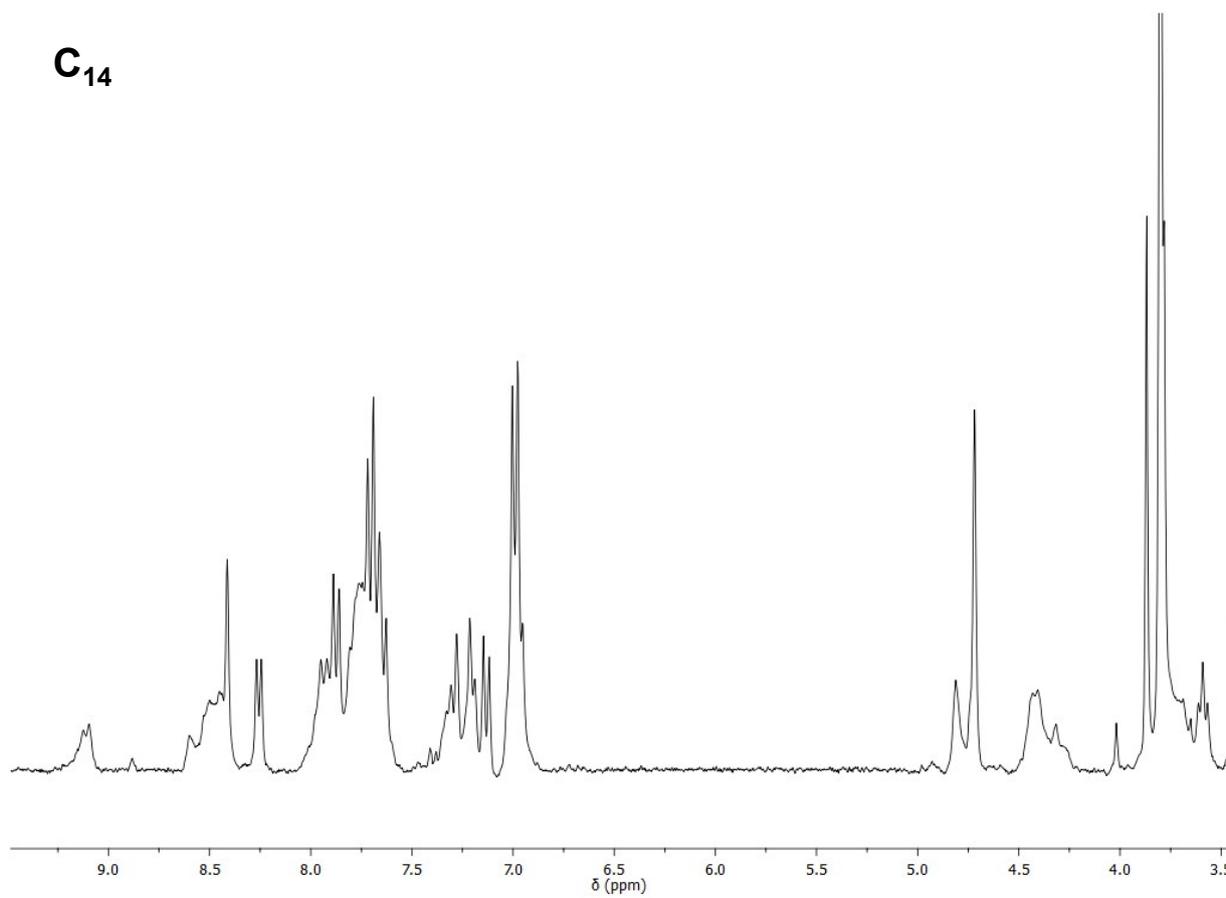


Figure S3 ¹H-NMR spectra of the **E+P+X** linkers experiments towards the guests series **C₅-C₁₄**

3.2 General procedure for Data analysis with PCA

The ESI-MS spectra were exported using the Data Analysis software of the ESI-MS and transformed into an excel sheet using OpenChrom. The peak intensities analysis was performed considering only the data corresponding to the m/z maximum of every di-charged peak clusters related to each cage, while the raw data analysis was performed choosing the spectral window containing all the di-charged peaks related to the $C_n@1(R-R'-R'')$ species.

The 1H NMR spectra were exported as ASCII using MestReNova taking into account the regions that presented the highest variance in the 3,8-9,5 ppm spectral window, which is diagnostic of the presence of the cages.

The experimental data collected by the two techniques were normalized internally, so that the highest point of intensity value for each experiment is considered at intensity equal to 1. In the case of ESI-MS data the mass of the guests C_n was previously removed. The data set was represented by a $n * m$ matrix X , where the n rows represent the experiments and the m columns represented the number of points acquired during the measurement.

The resulting set of data were run through the Principal Component Analysis (PCA) function of MatLabR2017a, using the following script:

```
[coeff,score,latent,~,explained] = pca(X)
```

The resulting PCA bidimensional plots were obtained from the score and coeff matrix from the pca function, considering the first and second principal components which account for at least 85% of the total variance of the data set.

3.2.1 PCA of ESI-MS spectra

The scatter plots of the two first principal components of the PCA of the data-set related to the DCL formed by **E-P-X** linkers. The peak intensity matrix is the matrix obtained from the normalized intensities of the ten peaks detected at ESI-MS. The corresponding Scores and Loadings plot of the first two components are showed in Figure S4 and Figure S6 respectively. The raw data matrix is generated from the normalized spectral data recorded by ESI-MS. The resulting scores plot is depicted in Figure S5.

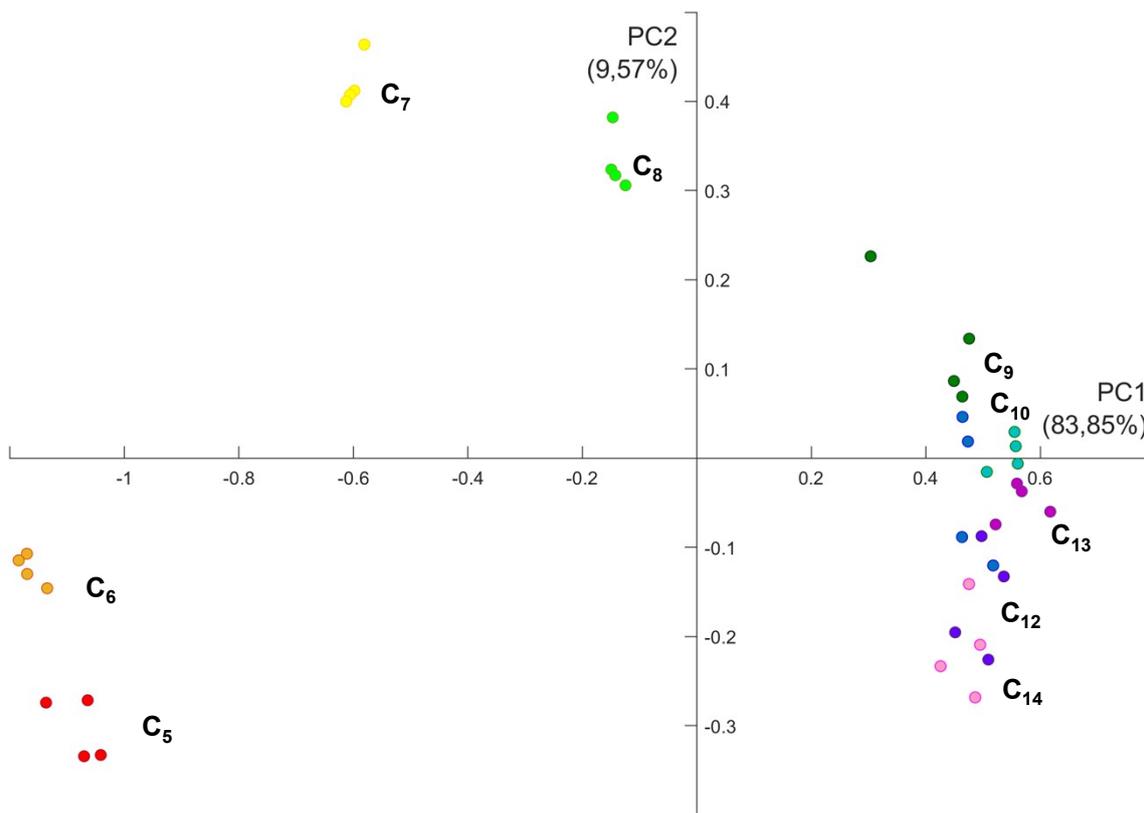


Figure S4 Scores plot of the first two principal components of the peak intensities matrix PCA

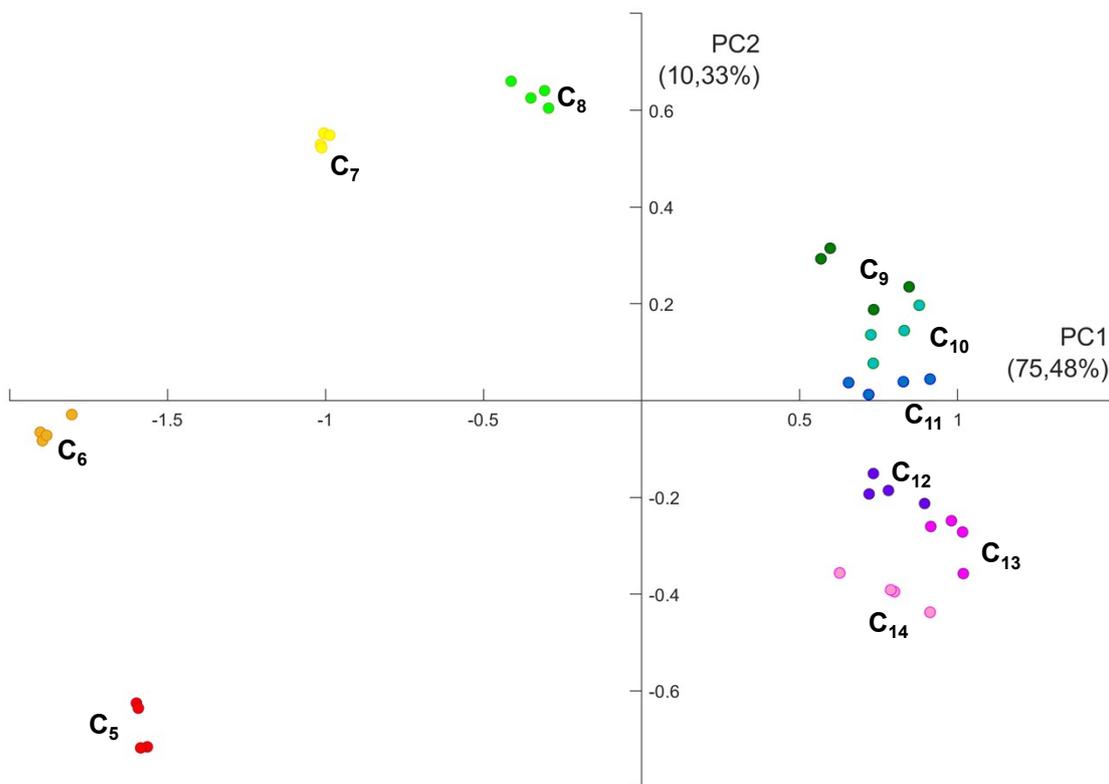


Figure S5 Scores plot of the first two principal components of the raw data matrix PCA

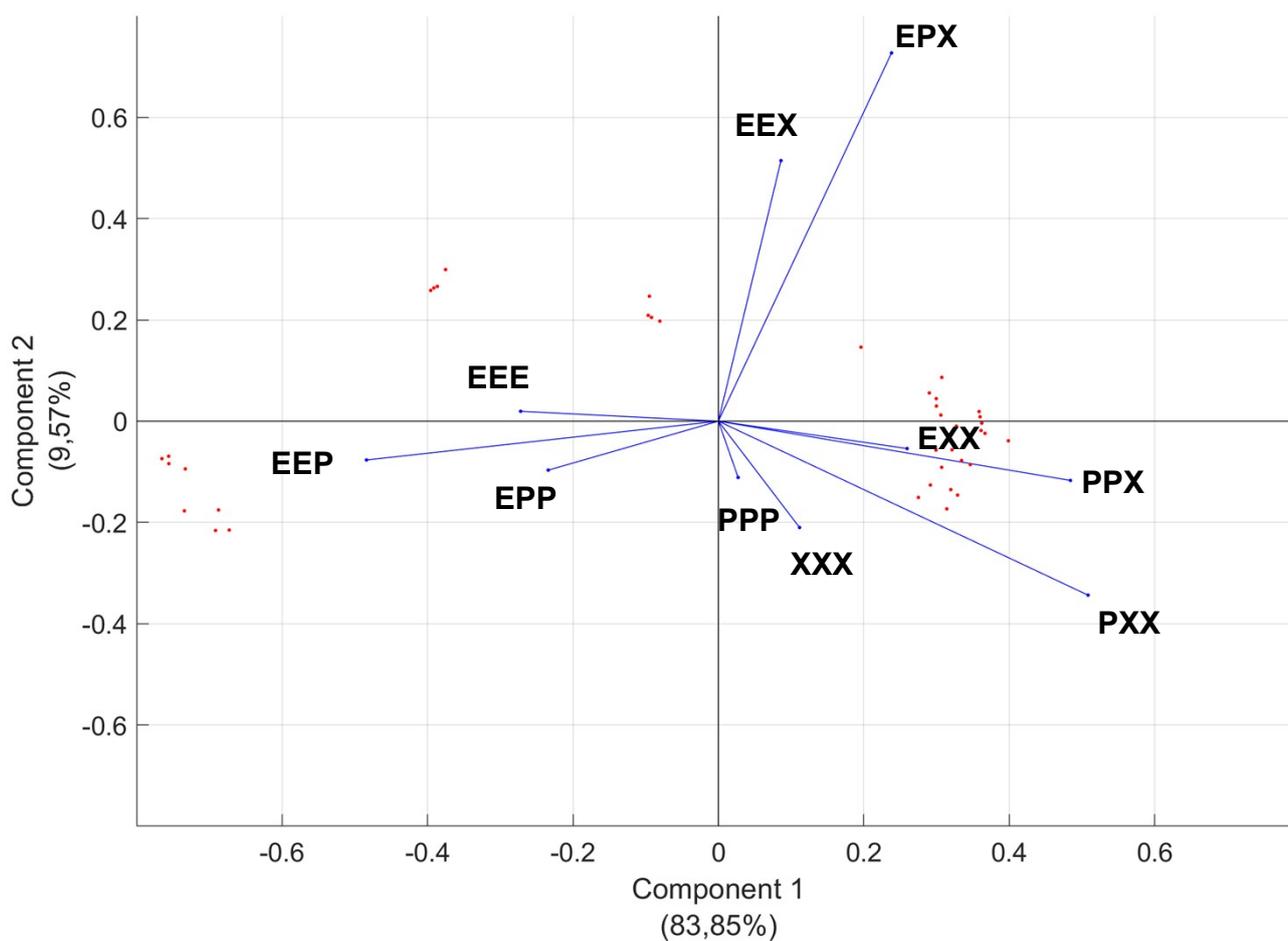


Figure S6 Loadings plot of the first two components of the peak intensities matrix PCA

3.2.2 PCA of $^1\text{H-NMR}$ spectra

The scatter plots of the two first principal components of the PCA of the data-set related to the experiment of **E+P+X** linkers. The raw data matrix is generated from the normalized spectral data. Each PCA considers a spectral window: 3,8-9,5 ppm (Figure S7); 7.7-7.9 ppm (Figure S8); 7.9-8.0 ppm (Figure S9); 8.3-8.6 ppm (Figure S10) and 8.9-9.3 ppm (Figure S11)

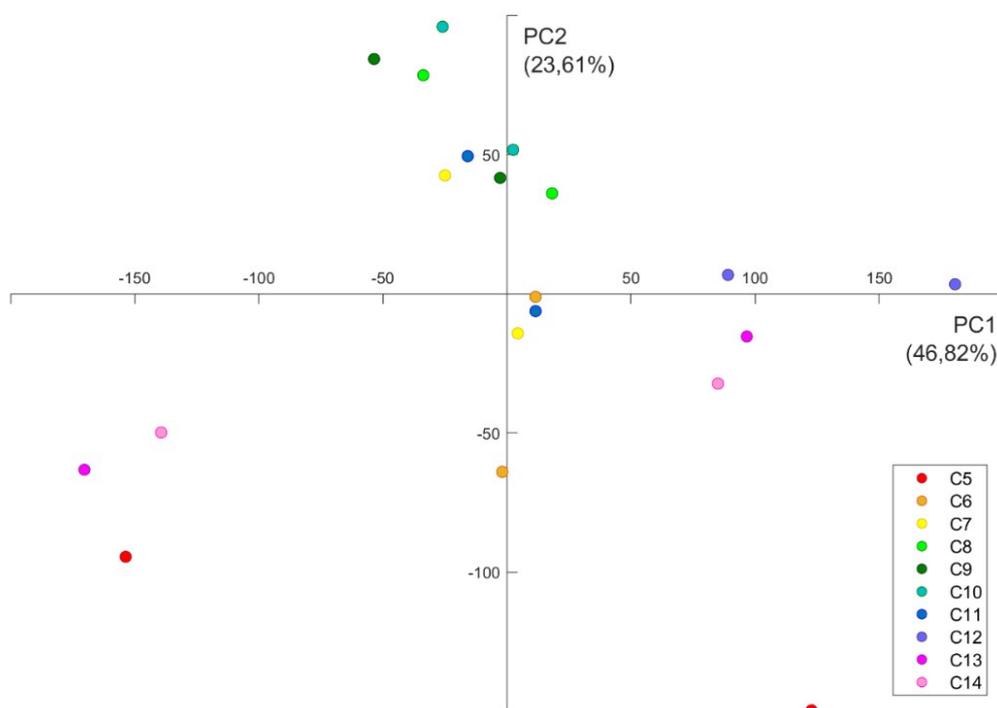


Figure S7 Scores plot of the first two principal components of the raw data matrix PCA for the spectral window 3.8-9.5 ppm

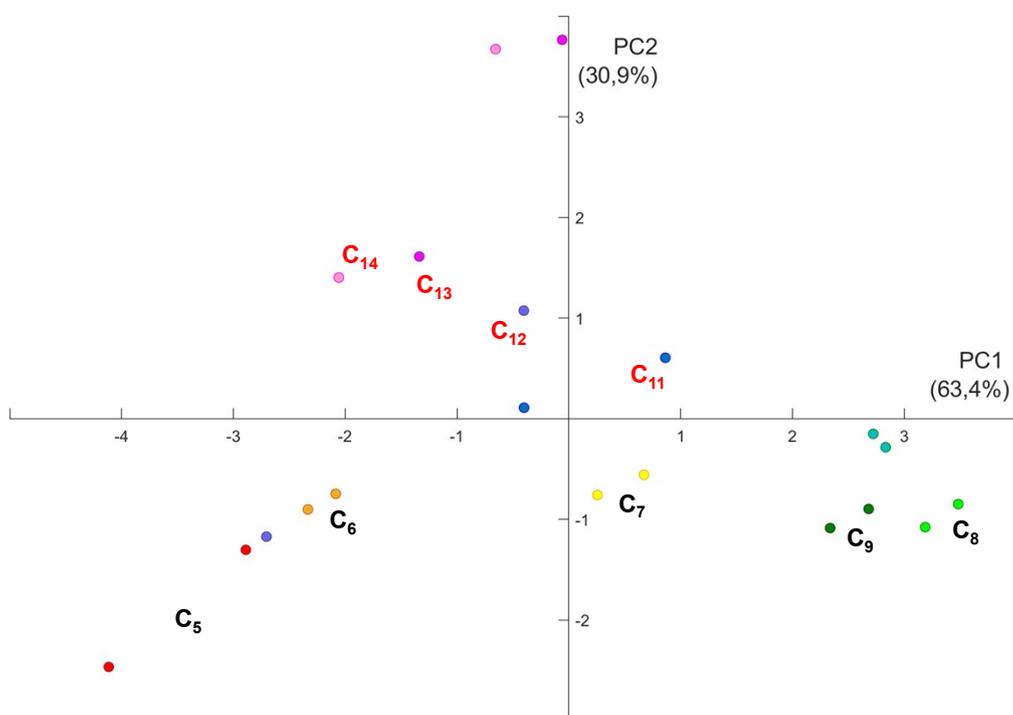


Figure S8 Scores plot of the first two principal components of the raw data matrix PCA for the spectral window 7.7-7.9 ppm. The points that are not correctly ordered are labelled in red.

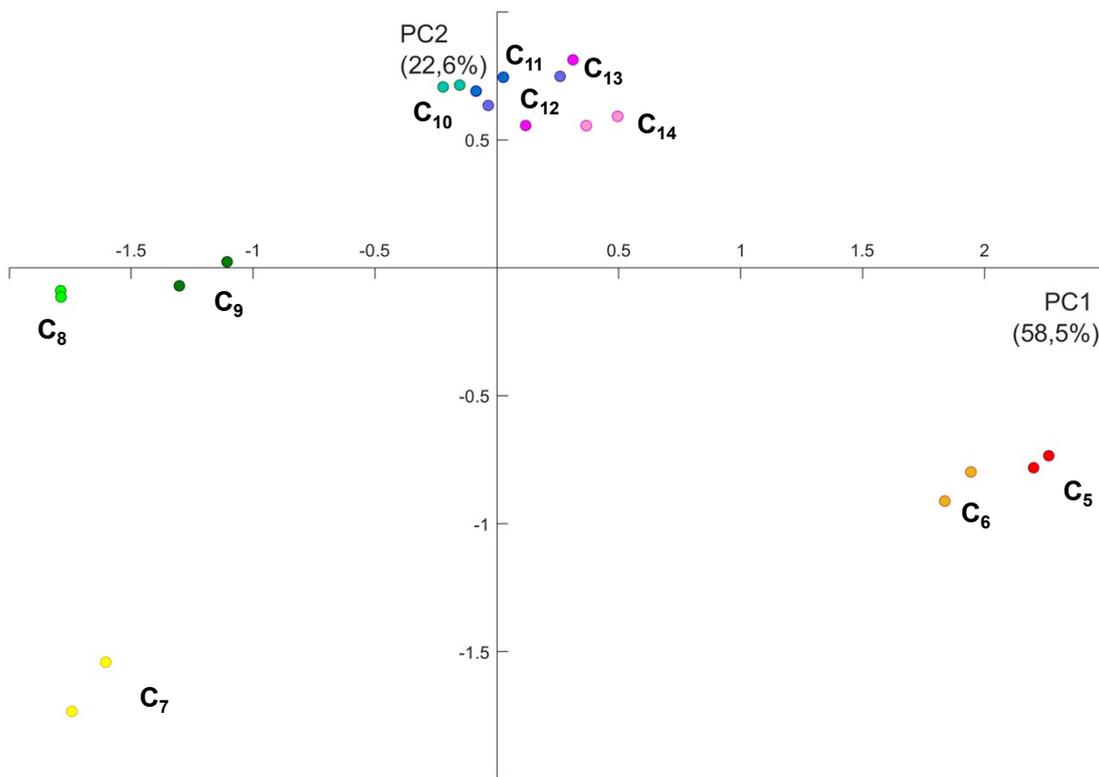


Figure S9 Scores plot of the first two principal components of the raw data matrix PCA for the spectral window 7.9-8.0 ppm.

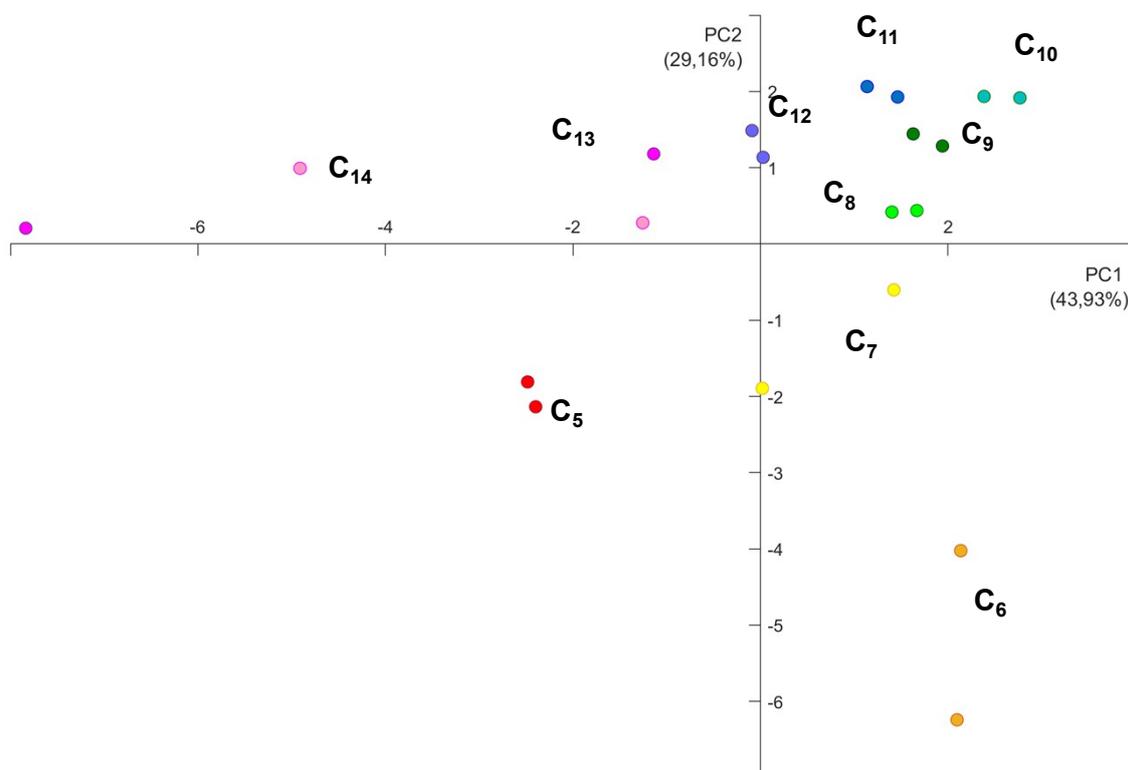


Figure S10 Scores plot of the first two principal components of the raw data matrix PCA for the spectral window 8.3-8.6 ppm.

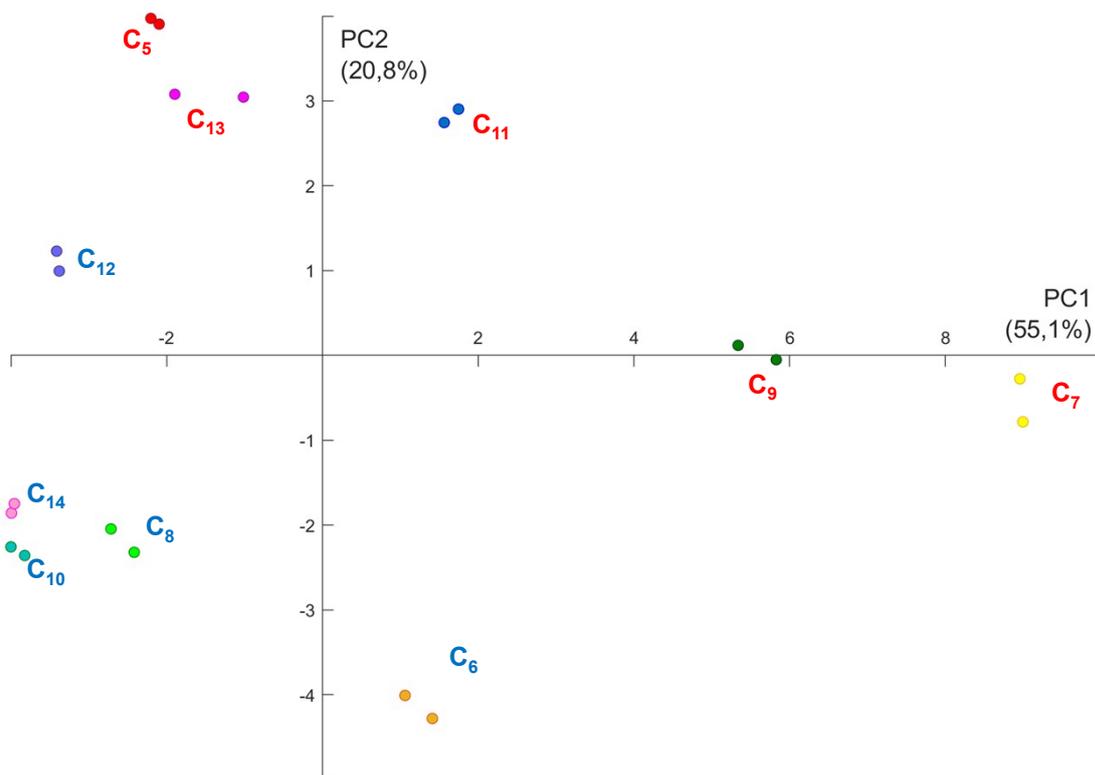


Figure S11 Scores plot of the first two principal components of the raw data matrix PCA for the spectral window 8.9-9.3 ppm. Points representing odd-length guests are labelled in red. Points representing even-length guests are labelled in blue.

3.3 Prediction of an unknown by projection in the PCA space

The new data were prepared as the previously described PCA data inputs. The original set of data were run through the Principal Component Analysis (PCA) function of MatLabR2018a. Afterwards, the unknown data were projected on the component space and compared to the original data using the following script:

```
[coeff,score] = pca(X);  
coeff = diag(std(X))\coeff;  
[X, mu, sigma] = zscore(X);  
sigma(sigma == 0) = 1;  
x = bsxfun(@minus, x, mu);  
x = bsxfun(@rdivide, x, sigma);  
y = (x\coeff');
```

With x being the new data and y the corresponding score value obtained by projection on the data covariance space. The resulting PCA bidimensional plots were obtained from the score and coeff matrix from the `pca` function, considering the first and second principal components which account for at least 85% of the total variance of the data set.

3.3.1 Projection of unknown in the PCA of ESI-MS spectra

The scatter plots of the two first principal components of the PCA of the data-set related to the DCL formed by **E-P-X** linkers. The peak intensity matrix is the matrix obtained from the normalized intensities of the ten peaks detected at ESI-MS. The corresponding Scores and Loadings plot of the first two components are showed in Figure S4 and Figure S6 respectively. The raw data matrix is generated from the normalized spectral data recorded by ESI-MS. The resulting scores plot is depicted in Figure S12.

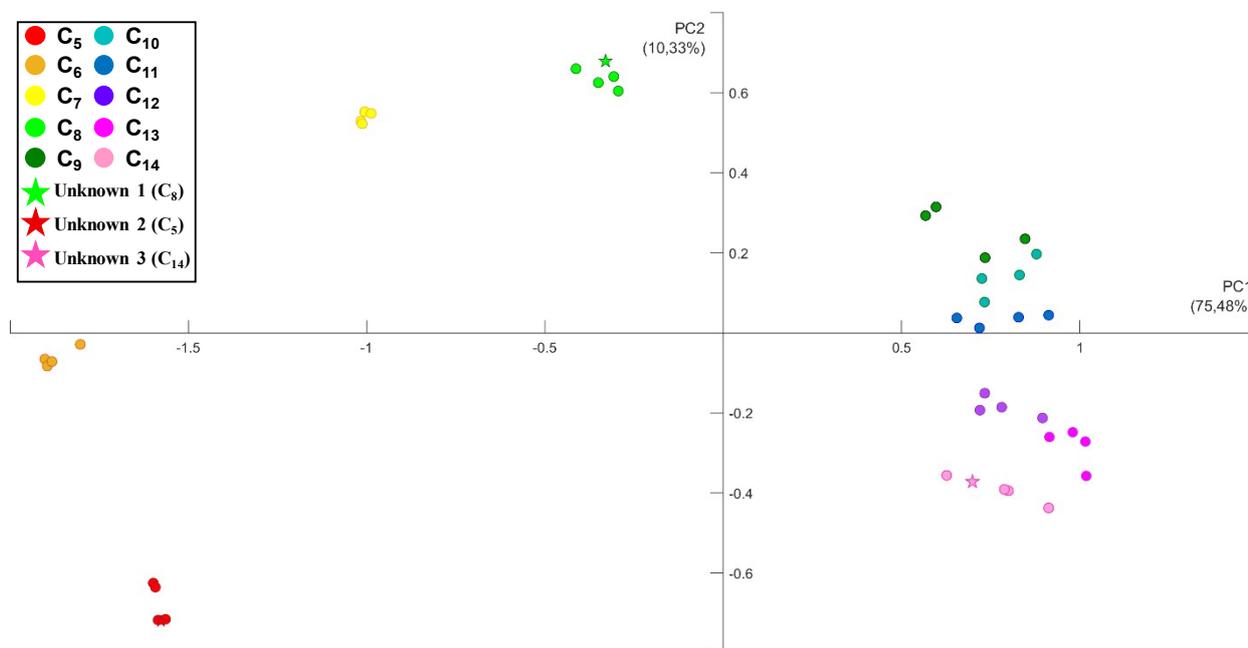


Figure S12 Scores plot of the first two principal components of the peak intensities matrix PCA with the projected additional data

3.3.2 Projection of unknown in the PCA of ¹H-NMR spectra

The scatter plots of the two first principal components of the PCA of the data-set related to the experiment of **E+P+X** linkers. The raw data matrix is generated from the normalized spectral data. Each PCA considers a spectral window: 7.9-8.0 ppm (Figure S13) and 8.9-9.3 ppm (Figure S14)

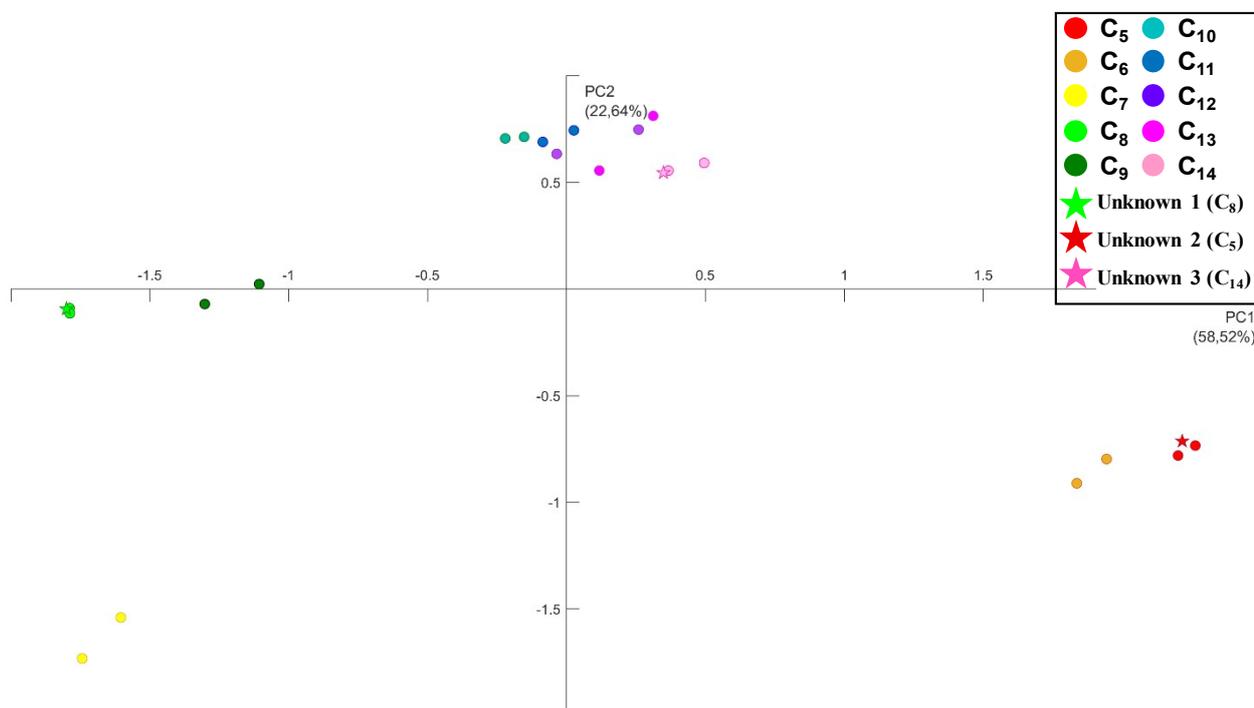


Figure S13 Scores plot of the first two principal components of the raw data matrix PCA for the spectral window 7.9-8.0, including the projections of the additional data points.

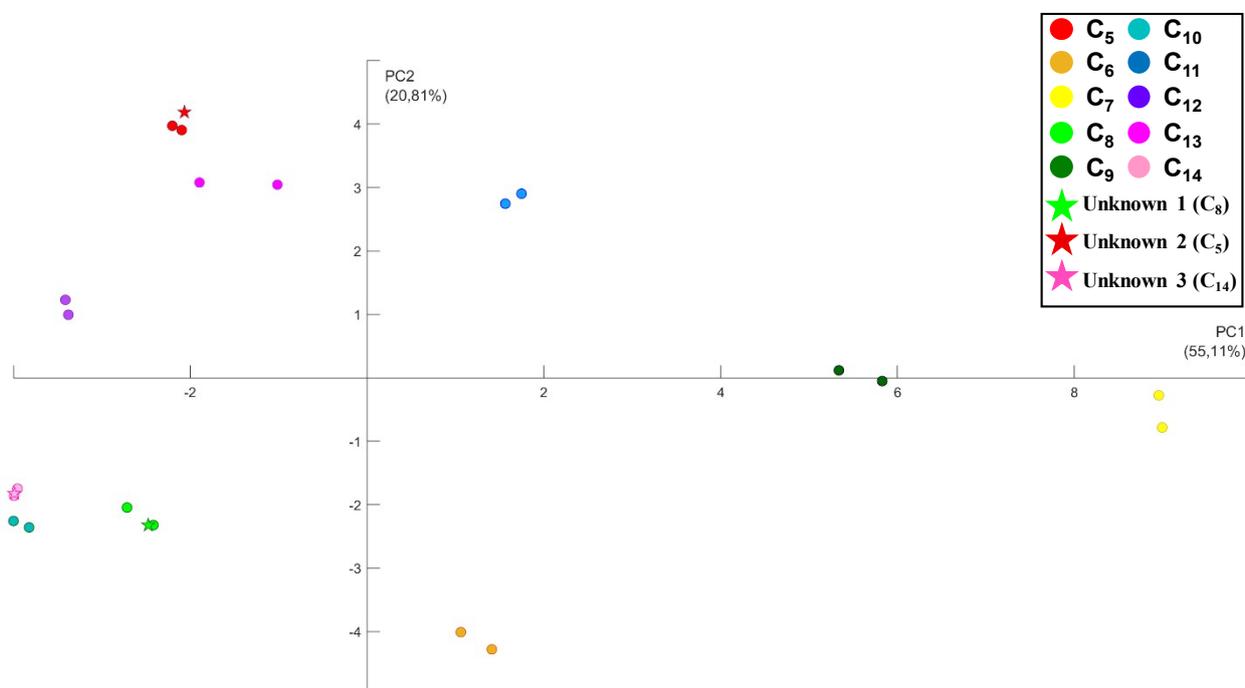


Figure S14 Scores plot of the first two principal components of the raw data matrix PCA for the spectral window 8.9-9.3 ppm, including the projection of the additional data points.