EDGE ARTICLE

Check for updates

Supramolecular cages as differential sensors for dicarboxylate anions: guest length sensing using principal component analysis of ESI-MS and ^{1}H-NMR raw data†

Carlo Bravin, [ID] Andrea Guidetti, Giulia Licini [ID] and Cristiano Zonta [ID] *

Dynamic covalent libraries (DCLs) have been widely used in the development of differential sensors. Inspired by recent studies which use supramolecular recognition systems for sensing, we report the use of a tris(-pyridylmethyl)amine (TPMA)-based supramolecular cage as a differential sensor for dicarboxylate anions. In particular, a library of molecular cages constituted by linkers differing in size and flexibility was allowed to equilibrate toward a series of guests differing in molecular size. The differential system was able to discriminate a series of dicarboxylates depending on their chain length. This differentiation was evaluated through the application of the Principal Component Analysis (PCA) method using interpolated and raw data obtained from ESI-MS. Interestingly, while the ^{1}H NMR spectra of the differential system did not allow for the discrimination of the analytes by the naked eye, PCA of the raw data from the spectra revealed information on the chain length of the guest and also on the odd–even character of the carbon chain.

Introduction

Differential sensing has recently become an analytical method capable of replacing traditional techniques based on direct molecular recognition.[1–9] This approach takes inspiration from the olfactory sense of mammals which can discriminate between complex odorant mixtures without the necessity for highly specialized peripheral receptors.[10,11] In broad terms, differential sensing employs a collection of low-selectivity receptors which give a peculiar signal for the analyte present in a solution. However, the large number of low-selectivity sensors is the key to overcome the problem of complicated matrixes.[12–18] The discrimination made by the receptors is achieved by a characteristic "fingerprint" related to each system sensor analyte. This characteristic pattern is usually represented by an ensemble of parameters which are not easily described *via* simple calibration methods. In order to give an easy-to-read interpretation of the data collected, this sensing methodology was coupled to statistical analysis techniques, like discriminant analysis and principal component analysis (PCA), which are intensively used in many fields of academia and industry.[19–25]

Among the possible chemical approaches, the capability of dynamic covalent libraries (DCLs) to respond to external signals

Department of Chemical Sciences, University of Padova, Via Marzolo 1, 35131 Padova, Italy. E-mail: cristiano.zonta@unipd.it

has been widely used for the recognition and signalling of chemical stimuli.[26–28] In particular, it is possible to employ complex systems characterized by multiple equilibria which can be perturbed by the presence of an analyte toward a particular product distribution. In this context, we have recently developed a new class of supramolecular structures,[29–33] in particular molecular cages, which have been synthesized using imine dynamic covalent chemistry (DCC). These cages are obtained by self-assembly of modified tris(-pyridylmethyl)amine (TPMA) complexes and different diamine linkers.[34–36] Variation of diamine linkers has allowed us to create a library of cages which have shown different selectivities toward a series of dicarboxylates (Fig. 1). ESI-MS studies, combined with ^{1}H NMR, revealed that binding energies correlated the dimension of the cage with the length of the dicarboxylate (*viz.* the diamine linker length defines the distance between the two metal centres and, as a direct consequence, the preferred length of the guest included). For example, ethylenediamine linkers **E** in cage **C$_n$@E-E-E** direct the system toward the preferential inclusion of adipate C$_6$ as the best guest (Fig. 1a). On the other hand, the longer diamine linkers *m*-xylylenediamine **X** in cage **C$_n$@X-X-X** lead to preferential binding of sebacate C$_{10}$ (Fig. 1c).[34]

The peculiar capability of these cages to differentiate dicarboxylates by their length, combined with the dynamic nature of their formation, made this molecular system an ideal candidate for the development of a differential sensing array for dicarboxylate guests. Moreover, inspired by the systems reported by Anslyn[1,3,14] and Alfonso,[37–43] we checked if the differential
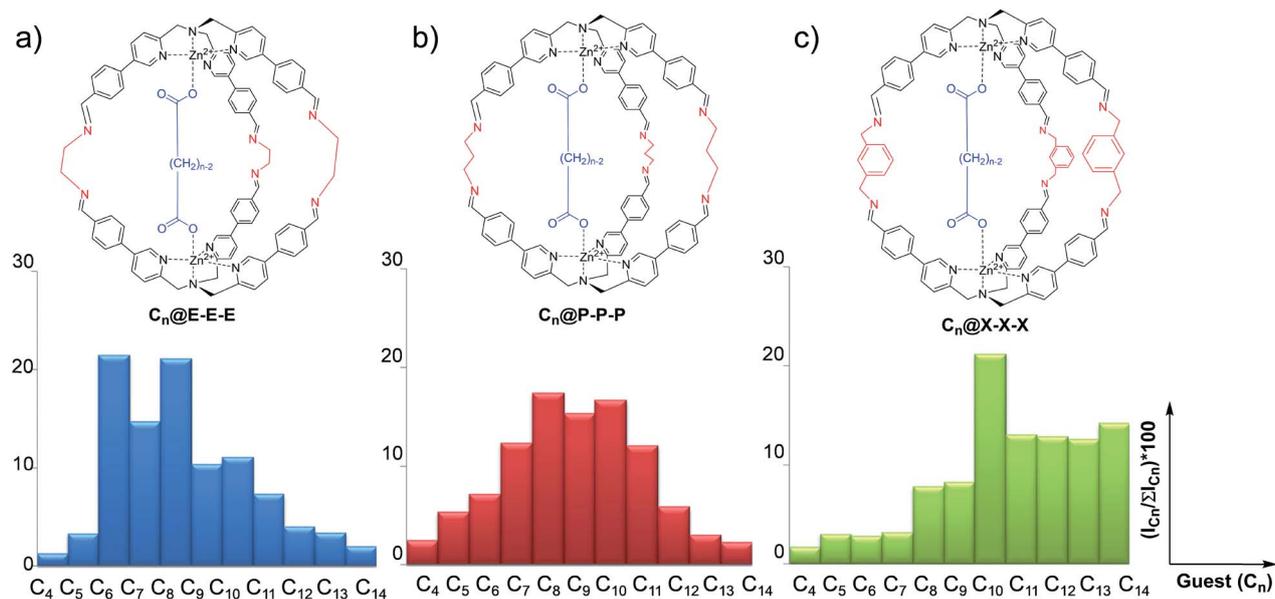
Fig. 1 Selectivity profiles for cages (a) C_n@E-E-E, (b) C_n@P-P-P and (c) C_n@X-X-X among the guest series ranging from C_4 to C_{14} ($I_{C_n}/\sum I_{C_n}$ represents the value of the relative intensity of the monoisotopic peak of each inclusion species among the guest series). The counteranion is perchlorate for the cage metals and triethylammonium for the carboxylate guests.

sensing technique worked with raw data extracted directly from ESI-MS or ^{1}H NMR analysis. This approach led to a system capable of discriminating dicarboxylate guests in the full range between C_5 and C_{14} and unexpected results from the use of ^{1}H NMR. The obtained results pave the way to another peculiar functional property, among the many other properties that supramolecular cages have shown so far.[44–53]

Results and discussion

Development of a differential sensing array using ESI-MS data

In order to develop a differential sensing system, we set up a series of experiments in which the DCL is allowed to form different cages incorporating different linkers (*viz.* diamines) in the presence of a single guest (*viz.* dicarboxylate). In other words, rather than focusing on the binding selectivity of a series

of guests towards a single cage, we investigated the formation of a cross-reactive array of multiple cages towards a single guest. In detail, the DCL consists of a solution containing one equivalent of dicarboxylate guests ranging from C_5 to C_{14}, two equivalents of complex **1** and a mixture of the selected diamines **E**, **P**, and **X**. Under the experimental conditions, 3 equivalents of each diamine were added. In addition, to compensate the amine excess, *p*-anisaldehyde was added to the reaction mixture (Fig. 2a).

In a typical experiment, the dynamic system explores all the possible combinations of binding between the dicarboxylate under study and all possible molecular cages, and the system equilibrates thermodynamically toward the more stable inclusion cages distribution.

72 hours after mixing, the reaction mixture, diluted to an appropriate concentration for the MS technique, was injected
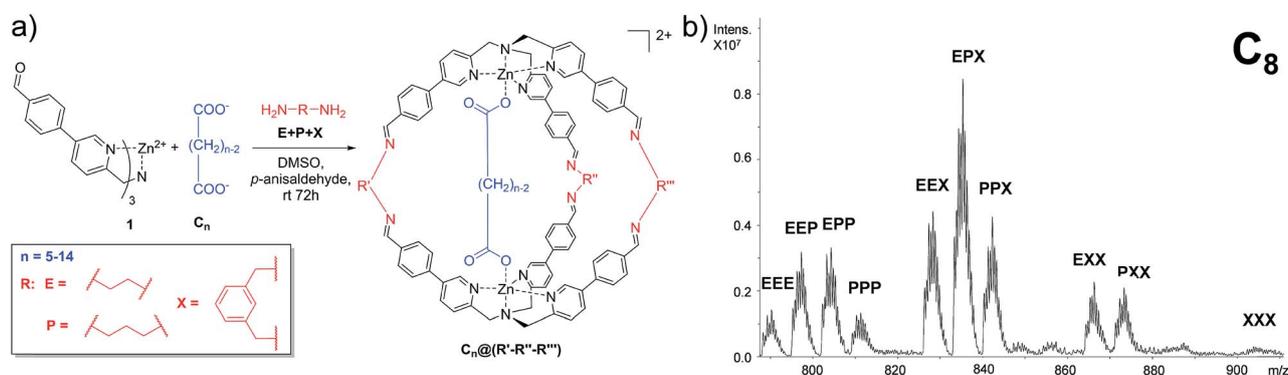
Fig. 2 (a) DCL using different linkers **E**, **P** and **X**. The ratio between the complex, guest, linkers and *p*-anisaldehyde is 2 : 1 : 9 : 12 ([**1**] = 1.6 mM, [C_n] = 0.8 mM, and [**E**] = [**P**] = [**X**] = 2.4 mM [*p*-anisaldehyde] = 9.5 mM). (b) ESI-MS spectrum of the mixture in the presence of C_8 as the guest after 72 hours.

into the ESI ion source. The typical MS trace displayed a series of m/z peaks corresponding to the different inclusion cages. For example, in Fig. 2b is reported the spectrum for the DCL experiment performed with C_8 as the guest in which the di-charged peaks related to all ten possible formed cages are present.

This experiment was extended to the series of guests ranging from C_5 to C_{14}, and the cage distribution is different for every single guest, as shown in Fig. 3a where the ESI-MS spectrum of each DCL experiment is reported (Fig. S1†).

At first glance, it can be reckoned that moving from C_5 towards longer guests results in an increase of the presence of cages containing the longest diamine in the series m-xylylene-diamine $\mathbf{X}$. A flattening toward a similar distribution can also be noted in the case of longer guests (C_{11}–C_{14}). In other words, the DCL system responds similarly to longer dicarboxylate.

In order to gather more information, we employed principal component analysis (PCA), an unsupervised technique used to reduce the dimensionality of data space. As the data source, we extrapolated the relative monoisotopic peak of the included cages for each dicarboxylate guest (Table S1 and Fig. S2 in the ESI†). This chemometric tool allowed us to generate a new set of variables, named principal components (PCs), to explain the variance of the system. Each DCL experiment was repeated four times for each guest to evaluate the strength of the analytical method (see Section S3.1 in the ESI†).

In the PCA of our DCL system, the different dicarboxylates showed effective separation, allowing for discrimination based on the chain length (Fig. 3b). The two main principal components PC1 and PC2, which account for almost 95% of the whole variance, perform an arch disposition which is in accordance with the length of the guests. Although PC2 accounts for 9.56% of the variance and it is not easily associated into a chemical property of the system, the disposition of the data in the score plot could be interpreted by taking into account the "horseshoe effect" which is typical for a unimodal distribution.[54,55] These
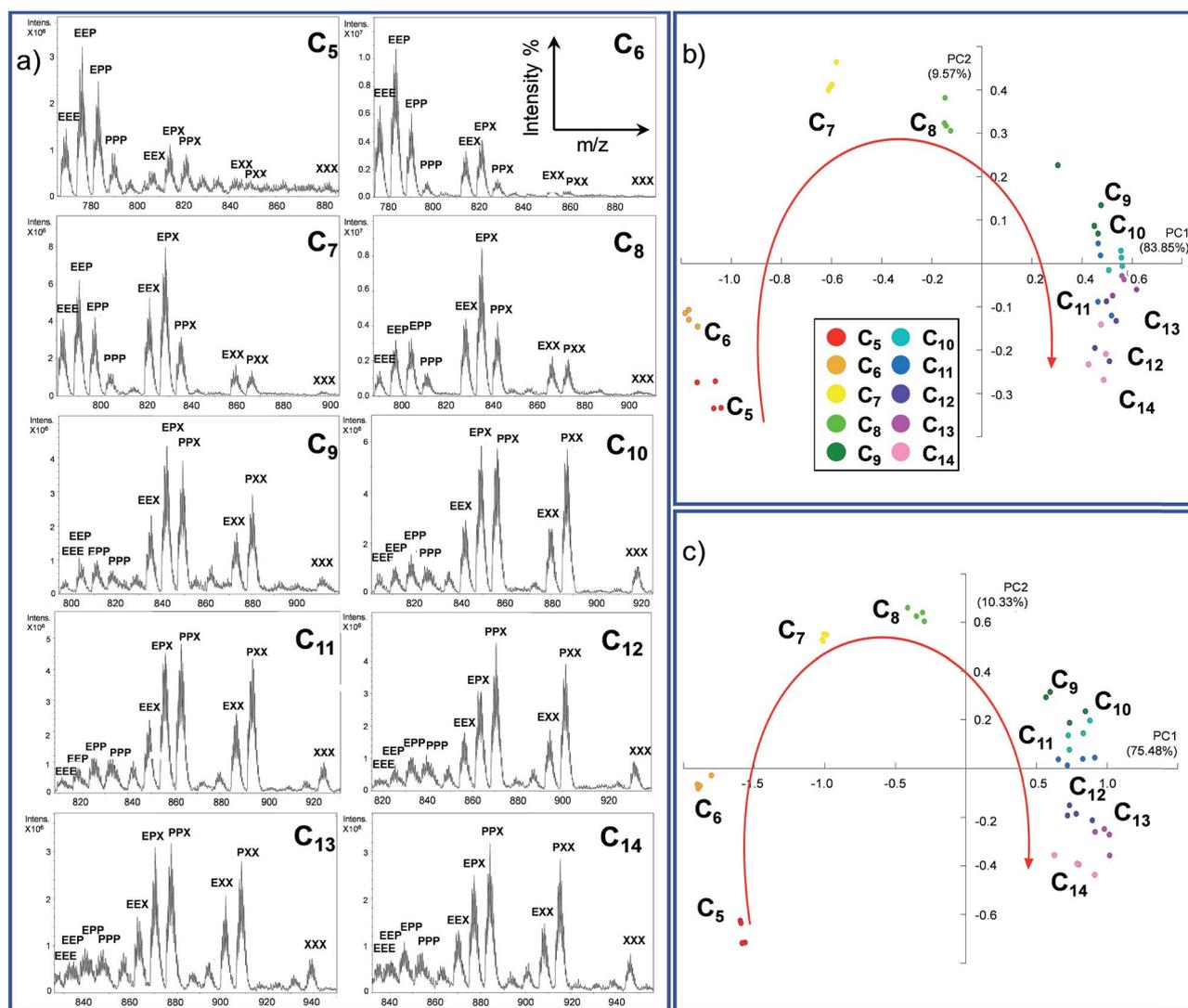
Fig. 3 (a) ESI-MS of the DCL ranging from C_5 to C_{14}. Ten isotope clusters corresponding to the ten possible cages are present (doubly charged peaks for each included system). (b) Principal component analysis scores of the spectral data for the E + P + X experiment taking into account the normalized intensities of the monoisotopic peaks in four different experiments and (c) PCA of the normalized raw data of the MS spectra.

results are also explained by the loading plot (Fig. S6 in the ESI†) which describes how positive values of PC1 correspond to the formation of cages containing longer linker **X**. In contrast, negative values indicate that mainly cages containing linker **E** are formed. However, the system has a strong tendency to promote the formation of mixed cages with the three different linkers C_n**@E-P-X**, due to the stoichiometric ratio of diamines in the DCL system.

The considerations made directly for the ESI-MS data reflect the results obtained with the PCA; while short chains are well distributed in the first part of the arch, a smaller distinction is observed for longer guests.

However, extrapolation of the monoisotopic peak intensities is already a reduction of the data information. For this reason, we decided to perform the PCA on the normalized raw ESI-MS data (Section S3.2, ESI†).[56] In this case, we introduce 126 points for each guest in the software four times. This reduces the data handling of the operator which could give rise to more information and thus better separation of the different guests. As expected, the results obtained by performing the PCA using the extrapolated monoisotopic peaks intensities (Fig. 3b) and the raw ESI-MS data are in agreement (Fig. 3c) in terms of the distribution of the different guests. However, the PCA obtained from the raw data analysis (i) increases the separation between the guests and (ii) allows for better clustering of the repeated analysis. In other words, the additional information present in the raw data allows for a better reproducibility of the measurement and it leads to a higher sensitivity of the method.

Differential sensing with ^{1}H NMR

PCA over ESI-MS data confirmed the discrimination capabilities of the cage sensor array. In addition, the PCA performed using raw data provides more information of the system allowing a better discrimination of the guest length.

However, the sensing capabilities were already noticeable from a "first glance" inspection of the ESI-MS trace. For this reason, we decided to investigate the response of our reactive array over ^{1}H NMR spectroscopy, a technique that in principle should display very similar spectra to the cages formed.

To explore this possibility, we tested our sensing system in DMSO-d_6 using the same experimental conditions described previously.

As expected, the resulting one dimensional ^{1}H NMR spectra recorded after 72 hours related to the ten different experiments are similar and the differences are difficult to interpret since each reaction mixture contains at least ten different cages (Fig. 4 and S3 in the ESI†).

Using the raw data from ^{1}H NMR, we built up a PCA taking into consideration the whole spectral region related to the cage signals (3.8–9.5 ppm). However, the whole spectra did not give any valuable information about the guest included (Fig. S7†). Therefore, we decided to focus our analysis on four different spectral windows which showed characteristic signal variations (Fig. 4).

The regions selected cover, respectively, the signals of the α-protons of the pyridine (8.9–9.3 ppm, green region in Fig. 4) and the imine protons of the cages (8.3–8.6 ppm, violet region) and two regions of the cage's aromatic protons related to the phenyl rings (7.9–8.0 ppm red region; 7.7–7.9 ppm blue region) (see ESI† Section S3.2).[57]

The PCA performed by analysing the data between 7.7 ppm and 7.9 ppm (Fig. 4d) shows that the system could discriminate between the length of different guests in the PC1/PC2 plane which accounts for 77% of the variance. In particular, only the intervals between C_5 and C_{10} are well differentiated. A particular discrimination arises from the PCA performed between 8.9 and 9.3 ppm where a clear distinction between the odd and even length alkyl chains of the guests is observed by plotting PC1 vs. PC2 (Fig. 4c). The observed result is in agreement with previously characterized cages which report a lower chemical shift of the α-protons of the pyridine in the presence of even length guests.[35] However, when the PCA was performed in the region between 7.9 ppm and 8.0 ppm, the length of the guests was discriminated and the PCA displays the above-mentioned "horseshoe effect" (Fig. 4a). A similar result was obtained considering the signals related to the imine protons of the cage (Fig. 4b).
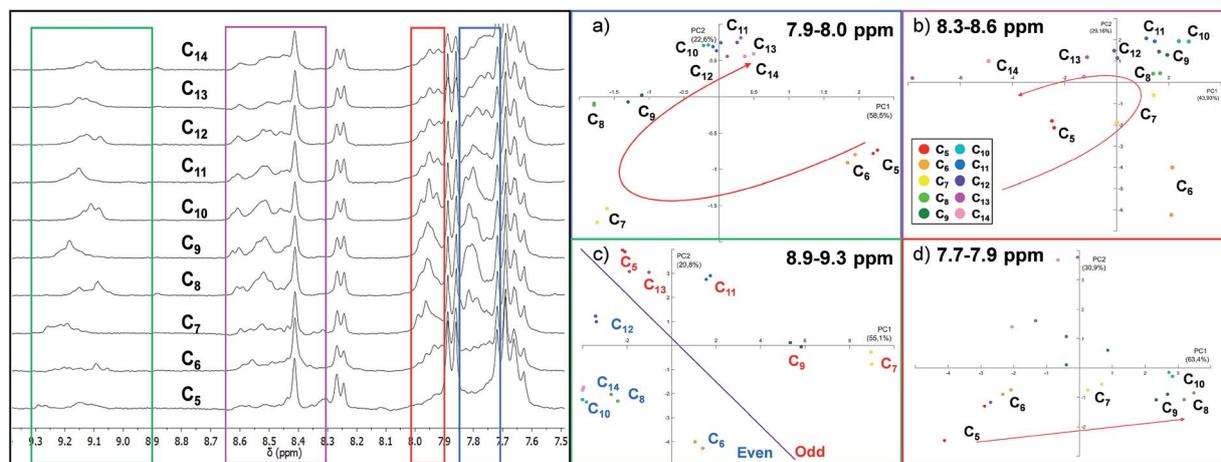
Fig. 4 On the left: ^{1}H NMR spectra (301 K, DMSO-d_6) of the DCL in the presence of the guests ranging from C_5 to C_{14}. On the right: PCA performed considering four regions of the ^{1}H NMR spectra (a) 7.9–8.0 ppm, (b) 8.3–8.6 ppm, (c) 8.9–9.3 ppm and (d) 7.7–7.9 ppm.

To summarize, from the analysis of ^{1}H NMR spectra which present indistinguishable differences, it is possible to gather information for the development of a sensor array. However, a careful evaluation of the spectral region should be taken. In particular, the large amount of numerical information corresponding to the whole spectra does not correspond to an increase in the discrimination capability of the sensor array, while this capability is achievable considering specific spectral regions.

Projection of an unknown sample in the PCA space

In order to validate the recognition system, three unknown samples were chosen to cover the whole carboxylate lengths (C_5, C_8 and C_{13}) and analyzed using MS and ^{1}H NMR spectra. The corresponding experimental data were projected on the component space and compared to the original data using a prediction script (see Section S3.3 in the ESI†). Interestingly, the unknown samples are close in the vectorial space to previous validation samples. MS is able to predict the length of the system, and NMR confirms both its capability to predict the length and the odd-even character of the carboxylates (Fig. S12–S14 in the ESI†).

Conclusions

In this work, a cross-reactive array of multiple cages for the differential sensing of guest length was developed. In order to achieve this objective, a series of experiments involving a mixture of three different linkers were performed and analyzed with ESI-MS and ^{1}H NMR. The data obtained from the resulting spectra were used to form matrix data-sets which were statistically analyzed through PCA. The resulting scores for the ESI-MS spectra show that the system was able to discriminate guests according to their length. In particular, the array was able to efficiently distinguish all the guests in the full range from C_5 to C_{14} using the monoisotopic peaks of the cages formed and the raw ESI-MS data as the input for the analysis. The PCA of ^{1}H NMR spectra was able to distinguish odd and even guests, therefore providing information on structural features related to the guests. In addition, the prediction of unknown guests within the PCA space was evaluated and the results allow us to extend the use of the developed methodology also for the evaluation of unknown samples. More importantly, the developed methodology which extends the chemometric analysis to two techniques less studied in combination with PCA highlights the advantages and precautions in the case of the use of raw data.

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

Notes and references

1 E. V. Anslyn, *J. Org. Chem.*, 2007, **72**, 687–699.
2 L. A. Joyce, M. S. Maynor, J. M. Dragna, G. M. da Cruz, V. M. Lynch, J. W. Canary and E. V. Anslyn, *J. Am. Chem. Soc.*, 2011, **133**, 13746–13752.
3 D. Leung, S. O. Kang and E. V. Anslyn, *Chem. Soc. Rev.*, 2012, **41**, 448–479.
4 L. You, D. Zha and E. V. Anslyn, *Chem. Rev.*, 2015, **115**, 7840–7892.
5 S. Stewart, M. A. Ivy and E. V. Anslyn, *Chem. Soc. Rev.*, 2014, **43**, 70–84.
6 M. M. Adams, L. A. Joyce and E. V. Anslyn, Uses of Differential Sensing and Arrays in Chemical Analysis, in *Supramolecular Chemistry*, ed. P. A. Gale and J. W. Steed, 2012, DOI: 10.1002/9780470661345.smc021.
7 N. A. Rakow and K. S. Suslick, *Nature*, 2000, **406**, 710.
8 J. R. Askim, M. Mahmoudi and K. S. Suslick, *Chem. Soc. Rev.*, 2013, **42**, 8649–8682.
9 P. Zardi, K. Wurst, G. Licini and C. Zonta, *J. Am. Chem. Soc.*, 2017, **139**, 15616–15619.
10 K. Persaud and G. Dodd, *Nature*, 1982, **299**, 352.
11 X. Grosmaitre, L. C. Santarelli, J. Tan, M. Luo and M. Ma, *Nat. Neurosci.*, 2007, **10**, 348.
12 D. Zamora-Olivares, T. S. Kaoud, K. N. Dalby and E. V. Anslyn, *J. Am. Chem. Soc.*, 2013, **135**, 14814–14820.
13 T. Zhang and E. V. Anslyn, *Org. Lett.*, 2006, **8**, 1649–1652.
14 H. H. Jo, X. Gao, L. You, E. V. Anslyn and M. J. Krische, *Chem. Sci.*, 2015, **6**, 6747–6753.
15 T. Zhang, N. Y. Edwards, M. Bonizzoni and E. V. Anslyn, *J. Am. Chem. Soc.*, 2009, **131**, 11976–11984.
16 S. L. Tobey and E. V. Anslyn, *Org. Lett.*, 2003, **5**, 2029–2031.
17 X. Li, D. Zamora-Olivares, K. L. Diehl, W. Tian and E. V. Anslyn, *Supramol. Chem.*, 2017, **29**, 308–314.
18 M. C. Janzen, J. B. Ponder, D. P. Bailey, C. K. Ingison and K. S. Suslick, *Anal. Chem.*, 2006, **78**, 3591–3600.
19 M. M. Adams and E. V. Anslyn, *J. Am. Chem. Soc.*, 2009, **131**, 17068–17069.
20 J. T. Brindle, H. Antti, E. Holmes, G. Tranter, J. K. Nicholson, H. W. L. Bethell, S. Clarke, P. M. Schofield, E. McKilligin, D. E. Mosedale and D. J. Grainger, *Nat. Med.*, 2002, **8**, 1439.
21 J. Lever, M. Krzywinski and N. Altman, *Nat. Methods*, 2017, **14**, 641.
22 R. Bro and A. K. Smilde, *Anal. Methods*, 2014, **6**, 2812–2831.
23 C. Zhang and K. S. Suslick, *J. Am. Chem. Soc.*, 2005, **127**, 11548–11549.
24 B. A. Suslick, L. Feng and K. S. Suslick, *Anal. Chem.*, 2010, **82**, 2067–2073.
25 J. J. Lavigne and E. V. Anslyn, *Angew. Chem., Int. Ed.*, 2001, **40**, 3118–3130.
26 L. You, S. R. Long, V. M. Lynch and E. V. Anslyn, *Chem.–Eur. J.*, 2011, **17**, 11017–11023.
27 S. J. Rowan, S. J. Cantrill, G. R. L. Cousins, J. K. M. Sanders and J. F. Stoddart, *Angew. Chem., Int. Ed.*, 2002, **41**, 898–952.

28 P. T. Corbett, J. Leclaire, L. Vial, K. R. West, J.-L. Wietor, J. K. M. Sanders and S. Otto, *Chem. Rev.*, 2006, **106**, 3652–3711.

29 F. A. Scaramuzzo, G. Licini and C. Zonta, *Chem.–Eur. J.*, 2013, **19**, 16809–16813.

30 F. A. Scaramuzzo, E. Badetti, G. Licini and C. Zonta, *Eur. J. Org. Chem.*, 2017, **2017**, 1438–1442.

31 E. Badetti, N. A. C. Dos Santos, F. A. Scaramuzzo, C. Bravin, K. Wurst, G. Licini and C. Zonta, *RSC Adv.*, 2018, **8**, 19494–19498.

32 E. Badetti, K. Wurst, G. Licini and C. Zonta, *Chem.–Eur. J.*, 2016, **22**, 6515–6518.

33 R. Berardozzi, E. Badetti, N. A. Carmo dos Santos, K. Wurst, G. Licini, G. Pescitelli, C. Zonta and L. Di Bari, *Chem. Commun.*, 2016, **52**, 8428–8431.

34 C. Bravin, C. Badetti, E. Putreddy, R. Pan, F. Rissanen, K. Licini and G. Zonta, *Chem.–Eur. J.*, 2018, **24**, 2936–2943.

35 C. Bravin, E. Badetti, F. A. Scaramuzzo, G. Licini and C. Zonta, *J. Am. Chem. Soc.*, 2017, **139**, 6456–6460.

36 C. Bravin, G. Licini, C. A. Hunter and C. Zonta, *Chem. Sci.*, 2019, **10**, 1466–1471.

37 E. Faggi, C. Vicent, S. V. Luis and I. Alfonso, *Org. Biomol. Chem.*, 2015, **13**, 11721–11731.

38 J. Atcher, A. Moure, J. Bujons and I. Alfonso, *Chem.–Eur. J.*, 2015, **21**, 6869–6878.

39 A. M. Valdivielso, F. Puig-Castellví, J. Atcher, J. Solà, R. Tauler and I. Alfonso, *Chem.–Eur. J.*, 2017, **23**, 10702.

40 E. Faggi, A. Moure, M. Bolte, C. Vicent, S. V. Luis and I. Alfonso, *J. Org. Chem.*, 2014, **79**, 4590–4601.

41 J. Atcher, J. Bujons and I. Alfonso, *Chem. Commun.*, 2017, **53**, 4274–4277.

42 M. Corredor, D. Carbajo, C. Domingo, Y. Pérez, J. Bujons, A. Messeguer and I. Alfonso, *Angew. Chem., Int. Ed.*, 2018, **57**, 11973–11977.

43 F. Puig-Castellví, Y. Pérez, B. Piña, R. Tauler and I. Alfonso, *Anal. Chem.*, 2018, **90**(21), 12422–12430.

44 F. J. Rizzuto, M. Kieffer and J. R. Nitschke, *Chem. Sci.*, 2018, **9**, 1925–1930.

45 A. J. McConnell, C. S. Wood, P. P. Neelakandan and J. R. Nitschke, *Chem. Rev.*, 2015, **115**, 7729–7793.

46 A. J. McConnell, C. J. E. Haynes, A. B. Grommet, C. M. Aitchison, J. Guilleme, S. Mikutis and J. R. Nitschke, *J. Am. Chem. Soc.*, 2018, **140**, 16952–16956.

47 F. J. Rizzuto, W. J. Ramsay and J. R. Nitschke, *J. Am. Chem. Soc.*, 2018, **140**, 11502–11509.

48 G. Cecot, M. Marmier, S. Geremia, R. De Zorzi, A. V. Vologzhanina, P. Pattison, E. Solari, F. Fadaei Tirani, R. Scopelliti and K. Severin, *J. Am. Chem. Soc.*, 2017, **139**, 8371–8381.

49 Z. Kostereli, R. Scopelliti and K. Severin, *Chem. Sci.*, 2014, **5**, 2456–2460.

50 W. Cullen, M. C. Misuraca, C. A. Hunter, N. H. Williams and M. D. Ward, *Nat. Chem.*, 2016, **8**, 231–236.

51 W. Cullen, S. Turega, C. A. Hunter and M. D. Ward, *Chem. Sci.*, 2015, **6**, 625–631.

52 G. Anderegg and F. Wenk, *Helv. Chim. Acta*, 1967, **50**, 2330–2332.

53 S. M. Jansze, G. Cecot and K. Severin, *Chem. Sci.*, 2018, **9**, 4253–4257.

54 J. Podani and I. Miklós, *Ecology*, 2002, **83**, 3331–3343.

55 M. W. Palmer, *Ecology*, 1993, **74**, 2215–2230.

56 The raw ESI† dataset contains 126 intensity *vs. m/z* corresponding to $C_n@(R'\text{-}R''\text{-}R''')$. In this set of data, the mass of the guest C_n is removed and the data are subsequently normalized to the highest intensity.

57 Signal assignment was possible due to the previously characterized cages reported in ref. 34.