ChemComm

ROYAL SOCIETY OF CHEMISTRY

COMMUNICATION

Enantioselective Sensing of Insect Pheromones in Water

Briana L. Hickey,[a] Junyi Chen,[b] Yunfan Zou,[c] Adam D. Gill,[d] Wenwan Zhong,[a,b] Jocelyn G. Millar[a,c] and Richard J. Hooley[a,d]*

An arrayed combination of water-soluble deep cavitands and cationic dyes has been shown to optically sense insect pheromones at micromolar concentration in water. Machine learning approaches were used to optimize the most effective array components, which allows differentiation between small structural differences in target, including between different diastereomers, even though the pheromones have no innate chromophore. When combined with chiral additives, enantiodiscrimination is possible, dependent on the size and shape of the pheromone.

Water-soluble synthetic host molecules, when paired with fluorescent reporter molecules, are a powerful tool for sensing biorelevant targets.[1] Pairing multiple synthetic hosts and fluorescent guests enables array-based pattern recognition sensing,[2] which allows a wide scope of target sensing and can be exploited to distinguish between extremely small differences in structure.[3] We have previously shown that arrayed water-soluble deep cavitand hosts can be exploited as sensors for different biomolecules,[4] using multivariate analysis tools such as Principal Component Analysis (PCA).[5] Examples include post-translationally modified peptides,[4a] nucleotides, and small molecules such as drugs of abuse.[6] These sensors exploit multiple different recognition mechanisms,[4b] and a wide range of targets can be sensed by changing the host and/or dye structure. The recognition is most effective for "soft" cations such as choline, or highly hydrophobic guests[7] such as hydrocarbons or tetrahydrocannabinol. More water-soluble, neutral guests are challenging targets due to their weaker affinity for the cavitand. Furthermore, discriminating between stereoisomers (either diastereomers or enantiomers) via molecular recognition is a significant challenge. Chiral recognition is possible with asymmetric metal-ligand cage

complexes,[8] most commonly those formed from lanthanides,[8c] but they favor charged substrates, and while H-bonded organic capsules can show diastereoselectivity, they function best in organic solvents.[9] Simple optical detection of stereoisomers of neutral molecules via indicator displacement assays with macrocyclic hosts in water is underexplored.
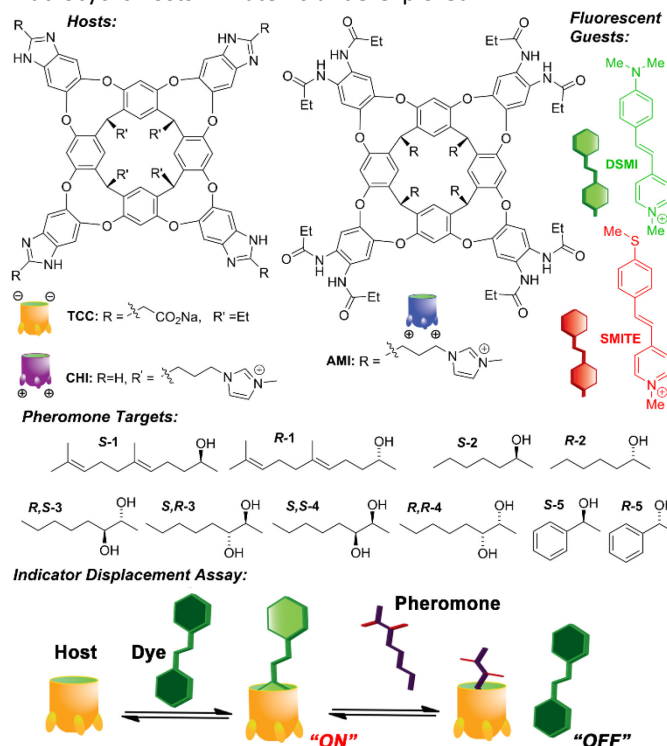

Figure 1. Components of the host:dye array tested, pheromone targets, and indicator displacement sensing mechanism.

The most notable examples of optically sensing and discrimination of stereoisomers have been demonstrated by Anslyn, who used metal-mediated dynamic combinatorial assembly to allow optical determination of enantiomeric excess[10] for small molecule alkanols, amines, and saccharides, among others. That innovative and impressive work focused on

[a1]Department of Chemistry; [b]Environmental Toxicology Graduate Program; [c]Department of Entomology; [d]Department of Biochemistry; University of California-Riverside, Riverside, CA 92521, U.S.A.* E-mail: richard.hooley@ucr.edu

recognizing the chiral center, and so did not effectively discriminate between differently sized alkanols. Nau combined decarboxylase enzymes with cucurbituril-based indicator displacement assays, which allow determination of enantiomeric excess (e.e.), but indirectly, as the cucurbituril does not bind the chiral species.[11] By employing host molecules that recognize the *alkyl components* of the target, size- and shape-based discrimination can theoretically be combined with diastereo- and enantioselective discrimination between molecules of similar structure. Here, we show that an arrayed combination of deep cavitand hosts, dyes, and chiral additives can optically sense and discriminate between insect pheromone targets in aqueous solution.

An extraordinary variety of small molecules are used by insects as pheromones for communication:[12] for example, fuscumol **1** (Fig. 1) is an aggregation pheromone used by a number of longhorn beetles (family Cerambycidae).[13] We chose a selection of suitable small molecule alkanols as targets, namely the homochiral isomers of **1-5** (Fig. 1). These single enantiomers were either purchased commercially or synthesized via known methods[14] and their enantiopurity was determined via GC (see Supplementary Information for details). They are suitably sized for binding inside deep cavitands, and if a dye is present that can competitively bind to the host, the pheromone target can displace this dye, turning it on, allowing optical detection. By combining multiple hosts and dyes with slight changes in structure in an arrayed format,[4] small changes in pheromone target can be sensed by multivariate analysis of the fluorescence signals.

The choice of host:guest system was complicated by the fact that alkanols can be weak guests for deep cavitands, therefore, we tested a series of hosts that have been previously used for selective sensing. All three hosts **TCC**, **CHI,** and **AMI** (Fig. 1) are water-soluble and can bind suitably sized guests in their cavities.[4,7] Two different dyes were chosen as reporters, the styrylpyridinium dyes **DSMI** and **SMITE** (Fig. 1). These guests bind in water-soluble deep cavitands and show enhanced fluorescence once bound. The initial test was a simple screening experiment to determine whether the alkanols could provide a fluorescence response when added to the host:guest systems. Single enantiomers (50 μM) of **1-5** (Fig. 2) were added to a solution of host (either **TCC**, **CHI**, or **AMI** at 20 μM) and guest (**DSMI** or **SMITE**, 3 μM, which has been shown to provide the most effective signal response when paired with these hosts[4a,b]) in 20 mM aqueous Tris buffer (pH 7.4). To further differentiate the response and provide more possible components to an array-based screen, we also added 50 μM heavy metal salt, either La(NO$_3$)$_3$, Ce(NO$_3$)$_3$, or UO$_2$(NO$_3$)$_2$ (as well as no metal), as these species have been shown to coordinate to water-soluble cavitands and modulate fluorescence response.[15]

As can be seen in Fig. 2 and ESI (for the full fluorescence responses with all the array components, see Figs S-12 – S-13), the pheromones indeed gave variable fluorescence responses when combined with the host:dye:metal combinations. The lowered fluorescence upon addition of pheromone indicates competitive fluorophore displacement from the hosts. Some of the responses were small, whereas others were quite large,

notably with the **AMI** host. However, when the data from the full 24-component array was subjected to Principal Component Analysis to determine discrimination, the results were underwhelming (Fig. 2c). While differently structured guests could be discriminated, the diastereo-discrimination was unsatisfactory, especially considering that a large 24-component array was used.
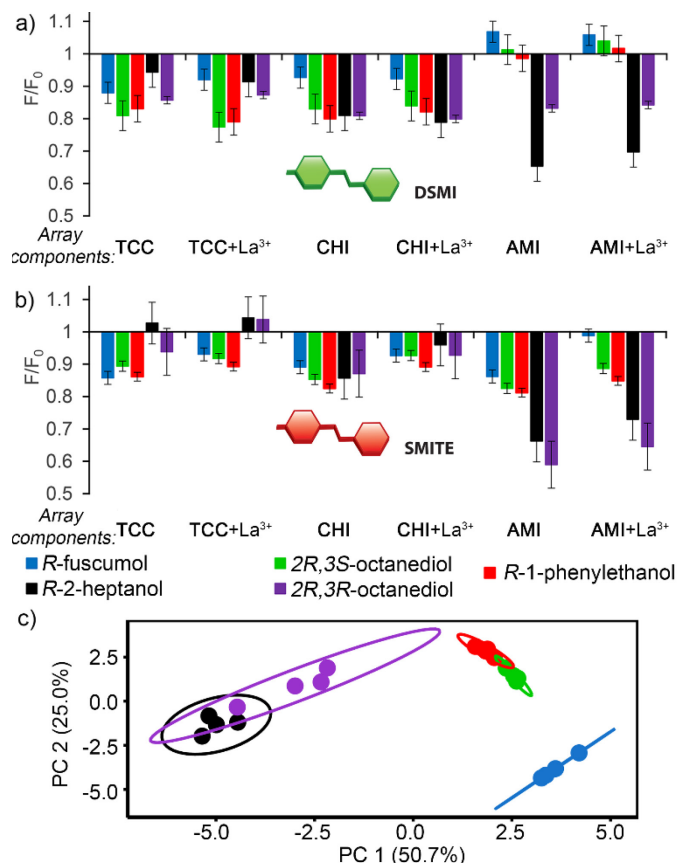

Figure 2. **Initial optical detection**. a) Relative fluorescence responses of **Host•dye•M⁺** complexes to pheromones **1-5** with a) **DSMI** dye; b) **SMITE** dye. c) PCA scores plot derived from the data in Figs. 2a, 2b and S12 – S13 using a 24-component array, ellipses indicate 95% confidence intervals. 20 mM Tris buffer, pH 7.4. [**Host**] = 20 μM, [**Dye**] = 3.0 μM, [**Metal**] = 50 μM, [**Pheromone**] = 50 μM.

The poor discrimination seen with the full array was unexpected, because simple visual inspection of the fluorescence responses in Figs 2a and 2b indicated that the different components did respond differently to the pheromones. Evidently, however, some combinations were ineffective and "damaged" the overall discrimination, so identifying the most important array components became necessary. Because manual evaluation of the data is time-consuming, error-prone, and uses a lot of material, we applied machine learning approaches[16] to select the best array elements from the original dataset to provide optimized sensing and discrimination. Machine learning is especially powerful for pattern recognition sensing, because it can detect hidden patterns in noisy or complex data sets. We have recently used this method to analyze DNA structures using arrayed sensors,[4b] and so we applied these techniques to the pheromone data to determine an optimized sensing array. The array data were

COMMUNICATION

treated with SVM-RFE (support vector machine recursive feature elimination), [17] using the sklearn library in Python 3.9. This is a supervised machine learning algorithm, in which a linear function hyperplane is used to separate data, and selects the optimal features by recursively removing non-important features. The procedure is illustrated schematically in Fig. 3a: first, StandardScaler was applied to standardize the fluorescence data, then the SVM-RFE algorithm was applied, and the individual sensor components were ranked. Six combinations were chosen for optimal discrimination, and then PCA was rerun using this optimized array. The six components were cross-validated and showed a cross-validation score > 0.99. The PCA scores plot of the results using this 6-component optimized array can be seen in Fig. 3b. Also, the data was processed using "sub-optimal arrays", i.e. the lowest ranked components, either the lowest 6 (Fig. 3c) or 12 (ESI, Fig. S-18).
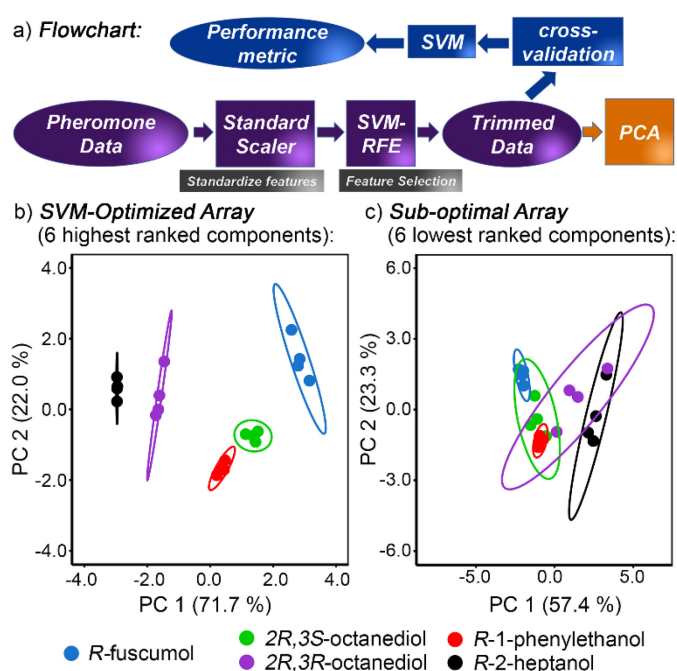

Figure 3. **Machine-Learning Optimized Sensing.** a) Flowchart illustrating the SVM-RFE process. b) PCA scores plots derived from the data in Figs. 2a, 2b and S12, S13 using b) an optimized array of the 6 most important components; c) PCA scores plot using a "sub-optimal" array of the 6 least important components selected by SVM-RFE; ellipses indicate 95% confidence intervals, 20 mM Tris buffer, pH 7.4. [**Host**] = 20 μM, [**Dye**] = 3.0 μM, [Metal] = 50 μM, [Pheromone] = 50 μM.

As can be seen in Fig. 3, the optimized array was far more successful than the full array in discriminating the pheromone stereoisomers, including the different diastereomers of 2,3-octanediol. Further analysis of the optimization rankings sheds some light on the results (for the full ranking table, see Fig. S-14). Most importantly, the cavitands are required for discrimination; using the dyes alone gave no discrimination (see Table S-1). The 6 most important elements were the combinations of **DSMI** with **AMI** and Ce^{3+}, **AMI** and La^{3+}, **AMI** alone, and **CHI** + Ce^{3+}, as well as **SMITE** with **CHI** alone or with **AMI** and Ce^{3+}. Overall, the differences between the combinations of **DSMI/SMITE**, **AMI/CHI** and metals were smallest, especially compared to those containing the anionic

TCC cavitand, which strongly populated the 6 "worst" combinations. This was surprising, as **TCC** has been our "workhorse" cavitand for biosensing in water,[4,6] and is the most affected by heavy metals.[15] A closer look at the guest binding properties of the cavitands sheds some light on this. The alkanol targets bind quite weakly in the hosts: while affinity is observed with **AMI** and **CHI**, NMR analysis shows that the guests display fast in/out exchange in the complexes and do not form long-lived Michaelis complexes (ESI, Figs S-8 – S-11). For example, fitting the NMR shifts of 2-heptanol upon titration into **AMI** to a 1:1 binding algorithm[18] yields K_a = 350 M^{-1}. This matches well with the **DSMI/SMITE** dye binding in those cavitands, as the affinity is lower than in **TCC**.[4b,6] The strong **DSMI/SMITE** binding affinity of **TCC** is a double-edged sword: while the host can bind the targets, its affinity for the dyes is strong, and the guests are incapable of sufficiently displacing either **DSMI** or **SMITE** from the host. The fluorescence plots in Fig. 2a illustrate this - the small changes in fluorescence are due to poor indicator displacement. Even though **AMI** and **CHI** are "poor" hosts for the targets, the affinity for dye is also poorer, so the combination of dye recognition and pheromone recognition is better matched, and so more effective discrimination occurs.
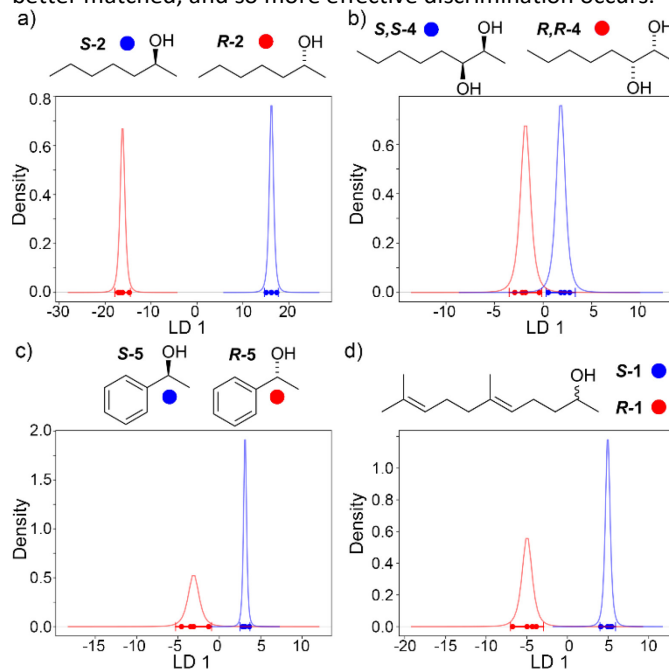

Figure 4. **Enantioselective Sensing.** 1D LDA plots showing discrimination between pheromone enantiomers. Array components: **AMI**, **CHI** hosts, **DSMI** dye and $La(NO_3)_3$, $Ce(NO_3)_3$, or $UO_2(NO_3)_2$ metal and L-sodium potassium tartrate as chiral additive; 20 mM Tris buffer, pH 7.4. [**Host**] = 20 μM, [**Dye**] = 3.0 μM, [Metal] = 50 μM, [Pheromone] = 50 μM, [Additive] = 50 μM. Red/blue dots = datapoints, curve = t-distribution probability density, vertical markers = 95% confidence intervals.

The optimized array was effective at sensing and discriminating the pheromone targets, but the most ambitious and important ability is to discriminate *enantiomers*. As the hosts are not homochiral, this requires an additional component that can allow enantiodiscrimination. The presence of heavy metals in the system introduces the possibility of adding chiral species to the array pool that can coordinate the metals and form larger chiral complexes, and the most

successful additive was L-sodium potassium tartrate (NaK-Tr, see Fig S-20 for data with other additives). This anionic species can coordinate to the added heavy metals, creating charged chiral complexes that can interact with the host:guest complexes, effecting enantiodiscrimination. As the additives change the nature of the sensing array, the array components were varied in this case. The ML data was used to remove cavitand **TCC** from the pool, and the 6-element sensing array consisted of **AMI**, **CHI** hosts, **DSMI** dye and La(NO$_3$)$_3$, Ce(NO$_3$)$_3$, or UO$_2$(NO$_3$)$_2$. Enantiodiscrimination requires direct comparison between two identified components, so PCA is not suitable, and we used supervised Linear Discrimination Analysis (LDA)[5b] as the discriminant method in this case. As only 2 classes are present the LDA reduces the data to a single dimension: the 1D LDA plots for sensing the two enantiomers of **R/S-1**, **R/S-2**, **RR/SS-4,** and **R/S-5** are shown in Fig. 4, and they contain LD 1 data points, t-distribution and 95% confidence intervals. All of the tested enantiomers could be discriminated at 95% confidence by the LDA method. The discrimination was most effective (Fig. 4a) for the 2-heptanol (cross-validation scores = 1.0), which positions its chiral center in the closest proximity to the cavitand upper rim when bound. While the t-distributions for the other guests tested were less well-separated (cross-validation scores >0.85), they could still be discriminated at 95% confidence on the scores plot (Fig 4a-c). The slightly weaker discrimination of these guests is most likely due to lower affinity and/or remote positioning of the stereocenter. Most importantly, the presence of cavitands was essential – when enantiodiscrimination was attempted with **R/S-2** or **RR/SS-4** using only dye/metal and NaK-Tr, significant overlap was observed in the distribution (Fig. S-22). The synergistic combination of host and chiral additive is essential for enantioselective discrimination of the pheromones.

In conclusion, we have shown that water-soluble deep cavitands are capable of stereoselective sensing of small molecule insect pheromone targets, when arrayed and combined with fluorescent dyes. The pheromones possess no optical detection motif, but can cause indicator displacement at micromolar concentrations. The presence of chiral additives allows enantiodiscrimination, and optimization of the sensing components can be achieved *via* machine learning techniques.

Conflicts of interest

There are no conflicts to declare.

Notes and references

1. (*a*) L. You, D. Zha and E. V. Anslyn, *Chem. Rev.* 2015, **115**, 7840; (*b*) R. Pinalli, A. Pedrini and E. Dalcanale, *Chem. Soc. Rev.* 2018, **47**, 7006; (*c*) R. N. Dsouza, A. Hennig and W. M. Nau, *Chem.-Eur. J.* 2012, **18**, 3444; (*d*) J. Beaver and M.L. Waters, *ACS Chem. Biol.* 2016, **11**, 643.
2. A. P. Umali and E. V. Anslyn, *Curr. Opin. Chem. Biol.* 2010, **14**, 685.
3. (*a*) B. C. Peacor, C. M. Ramsay and M. L. Waters, *Chem. Sci.*, 2017, **8**, 1422; (*b*) S. A. Minaker, K. D. Daze, M. C. F. Ma and F. Hof, *J. Am. Chem. Soc.*, 2012, **134**, 11674; (*c*) M. A. Beatty, A. J. Selinger, Y. Li and F. Hof, *J. Am. Chem. Soc.*, 2019, **141**, 16763.
4. (*a*) Y. Liu, L. Perez, M. Mettry, A. D. Gill, S. R. Byers, C. J. Easley, C. J. Bardeen, W. Zhong and R. J. Hooley, *Chem. Sci.,* 2017, **8**, 3960; (*b*) J. Chen, A. D. Gill, B. L. Hickey, Z. Gao, X. Cui, R. J. Hooley and W. Zhong. *J. Am. Chem. Soc.* 2021, **143**, 12791.
5. (*a*) S. Stewart, M. A. Ivy, and E. V. Anslyn, *Chem. Soc. Rev.* 2014, **43**, 70; (*b*) P. C. Jurs, G. A. Bakken and H. E. McClelland, *Chem. Rev.*, 2000, **100**, 2649.
6. (*a*) A. D. Gill, B. L. Hickey, W. Zhong and R. J. Hooley, *Chem. Commun.* 2020, **56**, 4352; (b) A. D. Gill, L. Perez, I. N. Q. Salinas, S. R. Byers, Y. Liu, B. L. Hickey, W. Zhong and R. J. Hooley, *Chem. – Eur. J.,* 2019, **25**, 1740.
7. R. J. Hooley, H. J. Van Anda and J. Rebek, Jr., *J. Am. Chem. Soc.*, 2007, **129**, 13464.
8. (*a*) D. Fiedler, D. H. Leung, R. G. Bergman and K. N. Raymond, *Acc. Chem. Res.* 2005, **38**, 351; (*b*) D. Fiedler, D. H. Leung, R. G. Bergman and K. N. Raymond, *J. Am. Chem. Soc.* 2004, **126**, 3674; (*c*) Q.-Y. Zhu, L.-P. Zhou, L.-X. Cai, X-Z. Li, J. Zhou and Q.-F. Sun *Chem. Commun.*, 2020, **56**, 2861.
9. J. M. Rivera, T. Martin and J. Rebek, Jr., *J. Am. Chem. Soc.* 2001, **123**, 5213.
10. (*a*) C.-Y. Lin, S. Lim and E. V. Anslyn. *J. Am. Chem. Soc.* 2016, **138**, 8045; (*b*) B. M. Chapin, P. Metola, S. L. Vankayala, H. L. Woodcock, T. J. Mooibroek, V. M. Lynch, J. D. Larkin and E. V. Anslyn, *J. Am. Chem. Soc.* 2017, **139**, 5568; (*c*) B. T. Herrera, S. L. Pilicer, E. V. Anslyn, L. A. Joyce and C. Wolf, *J. Am. Chem. Soc.* 2018, **140**, 10385.
11. D. M. Bailey, A. Hennig, V. D. Uzunova and W. M. Nau, *Chem.-Eur. J.,* 2008, **14**, 6069.
12. www.pherobase.org
13. (*a*) L. M. Hanks and J. G. Millar, *J. Chem. Ecol*. 2016, **42**, 631; (*b*) L. R. Meier, Y. Zou, J. A. Mongold-Diers, J. G. Millar, and L. M. Hanks, *J. Chem. Ecol.* 2020, **46**, 30.
14. (*a*) J. D. Wickham, J. G. Millar, L. M. Hanks, Y. Zou, J. C. H. Wong, R. D. Harrison and Y. Chen, *Environ. Entomol.* 2016, **45**, 223; (*b*) R. F. Mitchell, D. T. Hughes, C. W. Luetje, J. G. Millar, F. Soriano-Agatón, L. M. Hanks and H. M. Robertson, *Insect Biochem. Mol. Biol.* 2012, **42**, 499-505; (*c*) G. P. Hughes, Y. Zou, J. G. Millar and M. D. Ginzel, *Can. Entomol.* 2013, **145**, 327.
15. Y. Liu, M. Mettry, A. D. Gill, L. Perez, W. Zhong and R. J. Hooley, *Anal. Chem.* 2017, **89***,* 11113*.*
16. (*a*) S. D. Dreher and S. W. Krska, *Acc. Chem. Res.,* 2021, **54**, 1586; (*b*) Y. Shi, P. L. Prieto, T. Zepel, S. Grunert and J. E. Hein, *Acc. Chem. Res.,* 2021, **54**, 546; (*c*) J. Goecks, V. Jalili, L. M. Heiser and J. W. Gray, *Cell,* 2020, **181**, 92; (*d*) K. Jorner, A. Tomberg, C. Bauer, C. Sköld and P.-O. Norrby, *Nat. Rev. Chem.* 2021, **5**, 240.
17. (*a*) C. Cortes and V. Vapnik, *Machine Learning* 1995, **20**, 273; (*b*) O. Ivanciuc, *Rev. Comput. Chem.* 2007, **23**, 291.
18. Association constants calculated using BindFit software found at http://supramolecular.org.; D. B. Hibbert and P. Thordarson, *Chem. Commun.* 2016, **52**, 12792.