ChemComm

Biosensing and Biorecognition with Water-Soluble Macrocyclic Hosts

Journal:	*ChemComm*
Manuscript ID	CC-FEA-04-2025-002339.R1
Article Type:	Feature Article

ChemComm

FEATURE ARTICLE

Biosensing and Biorecognition with Water-Soluble Macrocyclic Hosts

Arman C. Garcia[1], Ria Lian[1], Wenwan Zhong[2] and Richard J. Hooley*[1].

Received 00th January 20xx,
Accepted 00th January 20xx

DOI: 10.1039/x0xx00000x

This Feature Article review discusses our efforts and those of others to use water-soluble macrocyclic receptors in the recognition, sensing and discrimination of a wide-ranging scope of biological targets. By using multiple hosts and fluorescent guests in an arrayed manner, a variety of biomolecules can be selectively detected and discriminated *via* simple optical methods. Targets include post-translationally modified peptides (with modifications as varied as lysine methylations, serine phosphorylation and peptide epimers), enzyme activity (including methyltransferases, demethylases, kinase and phosphatases), as well as anions that vary in size from iodide to non-canonically folded DNA structures. Using pattern recognition techniques allows excellent selectivity, but the key to broad-ranging target scope is to employ multiple orthogonal recognition mechanisms. By varying the recognition mechanism, large target scope and strong selectivity are not inversely correlated: simple host:guest complexes can be a powerful, tunable sensing platform that can be applied to detect multiple different biorelevant species.

Introduction

Water-soluble synthetic receptors have a variety of applications, from molecular recognition to catalysis, drug delivery and sensing, among many others.[1] These applications take advantage of a defined internal cavity in the receptor, which has been the driving force for the creation of these "protein-mimicking" species,[2] as the cavities are inspired by those in natural bio-receptors. There are a number of methods to introduce a defined cavity to a concave molecule, including hydrogen-bond or metal-mediated self-assembly, which have all been covered in myriad reviews.[1,3] Here, we will limit our discussion to covalent macrocyclic hosts, especially those that have been used for molecular recognition and sensing of biologically important analytes.

There are a variety of macrocyclic scaffolds that can be applied for biorecognition. Some involve cyclization of polar subunits (e.g. cyclodextrins,[4] cucurbit[n]urils[5] or bambusurils[6]), but most exploit aromatic species, such as calixarenes,[7] pillararenes,[8] cyclophanes[9] or deep cavitands.[10] These hosts can bind targets of suitable size, shape and charge, exploiting CH-π, π-π or cation-π interactions,[11] desolvation and hydrophobic driving forces,[12] and other weak forces such as hydrogen or halogen bonds[13] for target affinity and selectivity.

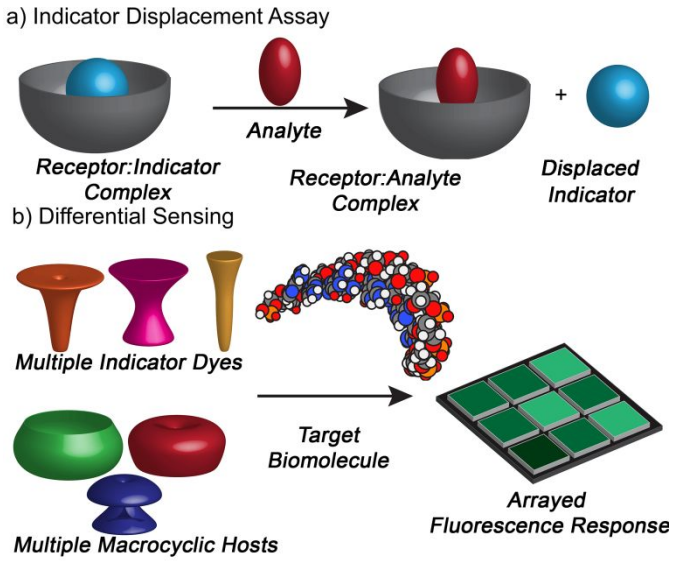

Figure 1. General techniques in biosensing: a) Illustration of the Indicator Displacement Assay concept; b) Differential Sensing approach.

While there are many methods to create a defined cavity, conferring water-solubility on the system is another challenge.[14] Some macrocycles such as cyclodextrins or cucurbit[n]urils (CB[n]) are natively water-soluble, whereas

[1]*Department of Chemistry, University of California–Riverside, Riverside, CA 92521, U.S.A. *E-mail: richard.hooley@ucr.edu*
[2]*State Key Laboratory of Precision and Intelligent Chemistry; Department of Chemistry, School of Chemistry and Materials Science; University of Science and Technology of China, Hefei, Anhui, 230026, P.R. China.*

other hosts such as calixarenes, pillar[n]arenes, cyclophanes or deep cavitands require the introduction of pendant solubilizing groups. These groups are often charged, or can be based on neutral oligo-ethyleneglycol motifs, etc. They are usually "silent", in that their only function is to confer solubility on the system – the recognition process is purely driven by the cavity. However, this is not always the case, as these pendant functions can either alter target binding affinity, act as the recognition element themselves, or provide a secondary binding element to enhance affinity. Examples of these properties can be seen in Gibb's use of NMe_3^+ groups to bind anions at cavitand lower rim crowns,[15] or the enhanced binding affinity of bis-functionalized ammonium species in CB[7] *via* hydrogen bonding.[16] As we will discuss here, pendant groups have a wide array of effects on guest binding and selectivity that can be exploited for selective biorecognition.

While target recognition is essential, the scope of biosensing is vastly enhanced when the recognition event is coupled to a detection motif, which allows analysis of the recognition *via* optical methods. Optical sensors offer high sensitivity, minimal sample preparation, and low noise, making them ideal for studying biological systems. Some hosts are natively fluorescent, and some can be covalently attached to fluorophores, but the most desirable detection method is the indicator displacement assay (IDA, Figure 1a).[17] IDAs offer a simpler approach by utilizing the displacement of an indicator molecule, often a fluorophore, from its host, generating a detectable optical signal. Designing an effective fluorescent sensor for biological applications requires careful consideration of several factors, especially since the target analyte is often present at low concentrations alongside competing analytes. It is essential to account for physiological pH, high salt concentrations, and the need for a strong optical signal, particularly in bioimaging studies.[18] The field of IDAs is now broad,[19] and covers a wide array of targets. The initial examples focused on detecting acetylcholine[20,21] using resorcinarenes, and cationic hydrogen-bonding receptors were shown to sense anions such as citrate[22] or fluoride[23] using carboxyfluorescein or nitrophenolate as indicators, respectively. This concept has since been expanded to allow detection of a wide range of targets, with myriad macrocyclic hosts in a variety of solvents.[17,19]

Substrate selectivity can be enhanced by applying the hosts as differential sensors.[24] By using multiple hosts and fluorophores in an array, a fluorescence fingerprint can be generated that, after statistical data processing,[25] provides a simple output that allows differentiation and classification of highly similar substrates (illustrated in Figure 1b). Different statistical analysis techniques are possible, including Principal Component Analysis (PCA), Linear Discriminant Analysis (LDA) Canonical Discriminant Analysis (CDA) and t-SNE, among others.[26] The differential sensing strategy allows synthetic receptor-based sensors to approach levels of selectivity in biosensing that are usually only associated with antibodies or other platforms that exploit biomacromolecules, with none of the synthetic complexity or expense associated with biological sensors.[27]

Our focus in this review is to discuss examples of complex bio-recognition and differential biosensing with small molecule receptors, and illustrate our contributions to the field. For the interests of space, we will not discuss the many beautiful examples of bio-recognition using cucurbit[n]urils[5] or pillar[n]arenes,[8] as they are extensive and have been exhaustively reviewed already.[1-3] We will highlight some choice examples of complex biorecognition, organized by target type, and then discuss how our work fits into the broader field.

Biosensing Applications of Defined Macrocyclic Cavities

As there are numerous reviews focusing on small molecule biosensing and drug delivery/intracellular detection, we will focus on a smaller subset of work that attempts to mimic the scope and selectivity of natural bioreceptors. Antibodies are capable of site- and state-selective differentiation of identical functional groups in complex biomolecules. This selectivity, although challenging, can be approached with synthetic macrocycles in certain cases.

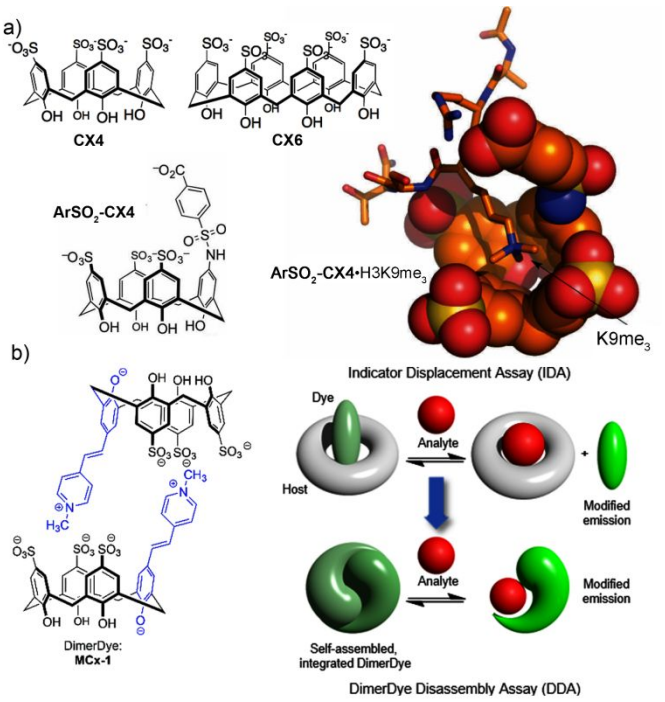

Figure 2. Calixarene-based PTM detection. a) Sulfonated calixarene receptors and a structure of **$ArSO_2$-CX4** binding H3K9me_3.[35] b) DimerDye structure and recognition mechanism. Adapted with permission from ref 38, copyright 2017, American Chemical Society.

Peptide Post-translational Modifications

Peptide and protein post-translational modifications (PTM) are epigenetic markers that allow expansion of the genetic code beyond the natural peptides coded from DNA.[28] These

modifications are myriad, including phosphorylation, methylation, acetylation, glycosylation, and ubiquitination.[29] Some of these modifications are very small, such as the addition of methyl groups to lysine or arginine sidechains – the challenge in their detection is generating a macrocycle that can selectively recognize the group in the complex environment of an oligopeptide or protein.

Fortunately, lysine methylations[30] are ideally suited to recognition by synthetic macrocycles. Due to their ease of synthesis, most synthetic macrocycles (except cucurbit[n]urils and cyclodextrins) are composed of multiple electron-rich aromatic rings, and this arrangement mirrors that seen in the acetylcholinesterase active site, where an "aromatic box" is formed with tyrosine and tryptophan residues,[31] allowing selective recognition of R-NMe_3^+ groups *via* cation-π interactions.[32] As such, acetylcholine recognition is quite simple with synthetic receptors, and other lysine and arginine methylations can also be targeted. The challenge lies in site- and state-selectivity of recognition. Common receptors such as p-tetrasulfonatocalix[4]arene (**CX4**) or p-hexasulfonato-calix[6]arene (**CX6**) can provide a core recognition scaffold for these targets.[33] By varying the structure of the scaffold, selectivity can be tuned to favor trimethylated lysines. Hof demonstrated state-selective recognition of methylated lysine over lysine with a strapped variant of **CX4** that covalently links the lower rim phenolates with ethylene glycol units. The flexibility of the host favors binding of methylated ammonium groups, with the collapse of the host necessary for binding.[34] By varying the pendant group in the calixarene, selectivity for Kme_3 peptides such as H3K27me_3 could be enhanced, allowing disruption of protein and protein reader interactions.[35] The maximal selectivity for Kme_3 peptides was seen with the sulfonated variant **$ArSO_2$-CX4** (Figure 2a). Binding of methylated lysine groups was also observed in proteins, including Kme_2-lysozymes.[36]

This recognition selectivity was extended to optical detection by indicator displacement, and by formation of differential sensing arrays. Hof applied IDA to study histone modifications using sulfonated calixarenes (**CX4**, **CX6** and a modified variant containing an aryl bromide) paired with pyridinium dyes (lucigenin (**LCG**) or the styrylpyridinium dye (**PSP**).[37] The simple **CX6**/**PSP** pairing could easily distinguish peptides with lysine methylations after LDA by methylation state (i.e. Kme vs Kme_2 vs Kme_3), and a larger array consisting of three calixarenes and LCG was capable of discriminating identical methylation states at different sites of histone peptides, such as H3K36me_3 and H3K9me_3. This array could also identify different types of lysine modifications, including differentiating between H3K9ac, H3K9, H3K9me_3, and H3K9me_3S10ph, as well as symmetric and asymmetric dimethylation on H4R3 peptides.

The original work required a separate dye for IDA, but by tailoring the calixarene synthetic route, Hof and Beatty created the self-reporting "DimerDye" scaffold.[38] The DimerDye system exploits hydrophobic and cation-π interactions to form a dimer that is self-quenched – binding an analyte breaks the dimer, turning on the sensor. The DimerDye system has the advantage of being highly salt-tolerant, and the simple synthetic route allows creation of a wide array of structure, ideal for differential sensing. The merocyanine calixarene **MCx-1** (Figure 2b) is effective at monitoring both methylation of H3K4 to H3K4me_3 and demethylation of H3K9me_3, enabling the detection of multiple enzymatic processes. Expanding the DimerDye suite to incorporate other fluorophores resulted in the DimerDye II system that was effective in discriminating between key drugs of abuse such as nicotine, MDMA, and cocaine in water and saliva.[39]

Calixarenes are cone-shaped, so are well-suited for detecting tetrahedral targets such as Kme_3. In contrast, planar arginine residues are more suitable substrates for toroidal macrocycles that can allow cation-π interactions with the upper and lower faces. Cyclophanes, originally created by Dougherty,[40] are the paradigmatic example of synthetic hosts exploiting cation-π interactions for target binding. The Waters group have vastly expanded this initial concept by using dynamic combinatorial libraries (DCL) of cyclophane hosts.[41] By mixing various oligomercaptan components (such as, among others, E, G and N, Figure 3), reversible disulfide bond formation can create a library of synthetic hosts that can be selected for a single specific target *via* templation, shifting the equilibrium toward the formation of a host cyclophane that binds the guest. The initial macrocycles, selected by using a trimethylated lysine dipeptide as a template, showed selectivity for trimethyllysine, similar to Dougherty's examples of acetylcholine-selective cyclophanes. Affinities were K_d = 25–30 μM, comparable to that of the native HP1 chromodomain. These synthetic receptors demonstrated over two-fold selectivity for KMe_3 over KMe_2 and over four-fold selectivity for KMe_3 over unmethylated lysine, attributed to stronger cation-π interactions with the methylated ammonium group.

By expanding the dynamic library, enhanced receptors could be formed that favored binding of unsymmetrical dimethylarginine, a target that is challenging for bowl-and cone-shaped macrocycles.[42] Low micromolar affinities of $aRme_2$-containing peptides were seen, and up to 8-fold selectivity for $aRme_2$ over $sRme_2$, depending on sequence, was observed. This initial work was again enhanced by optimization of the monomers used for library selection in the DCL process (Figure 3a,b).[43] By using different monomers—E, G, and N— in their DCL, distinct response patterns were observed *via* HPLC analysis. For example, N_2G_2 was amplified by Rme_{2a}, EG_3 by Kme_3, and ENG_2 by both Rme_{2a} and Rme_{2s}. The N_2G_2 macrocycle showed affinity K_d = 1.2 μM for H3R8me_{2a}, and 10-fold selectivity over other methylated arginines. The approach was expanded by developing differential sensor arrays using multiple synthetic receptors to distinguish analytes *via* pattern recognition.[44]

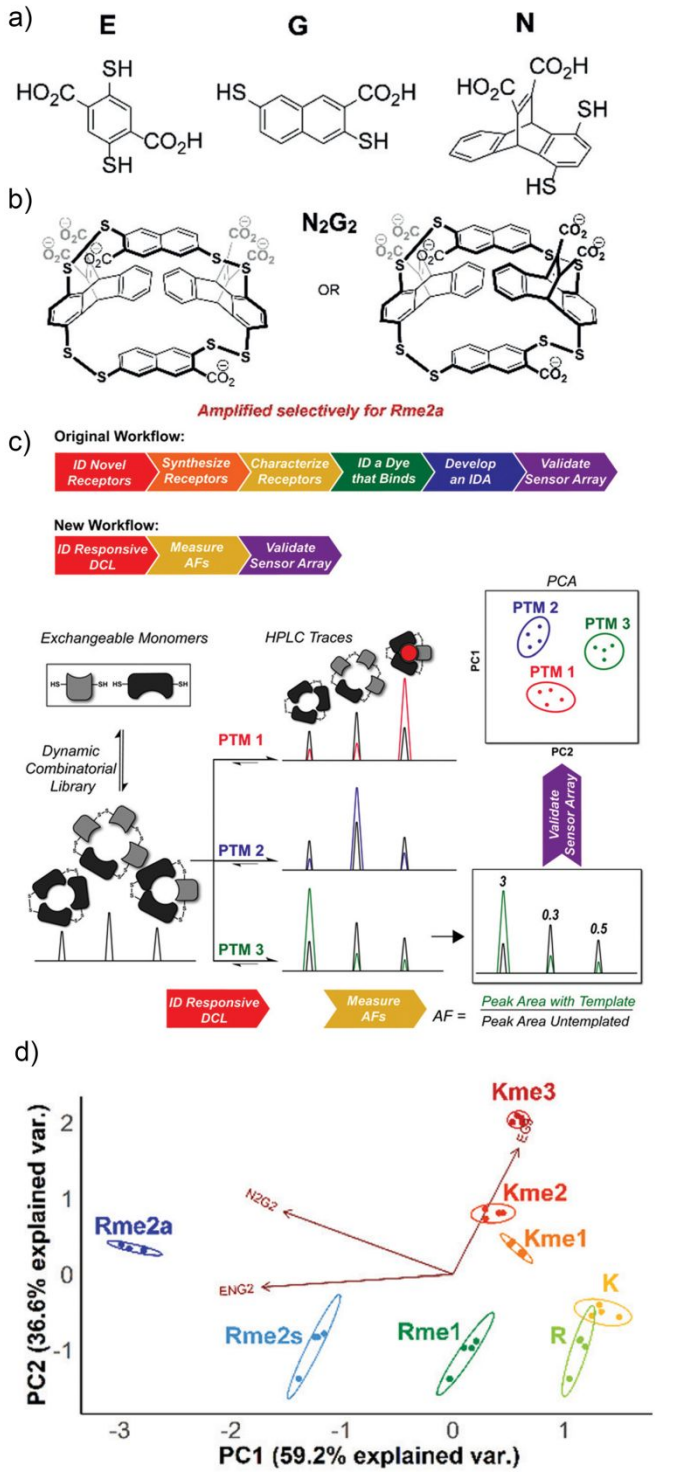

Figure 3. **Cyclophane Dynamic Combinatorial Libraries.** a) Selected bis-mercaptan elements used for DCL cyclophane formation. b) Examples of selected cyclophanes that favorably bind modified amino acids. c) Dynamic Combinatorial Library selection workflow. d) PCA scores plot for XGGY tetramer peptides using AFs calculated from HPLC peaks with confidence ellipses drawn at 95%.[44]

A system composed of eight short XGGY peptides equilibrated with three monomers was analyzed. Using the amplification factors (AFs) of the species in the library as determined by HPLC integration as the fingerprint for PCA (Figure 3d), as opposed to an indicator displacement process, enabled excellent discrimination of all tested PTMs, including Rme, which had not been previously distinguished by synthetic receptors. Further extension of this work to the Histone H3 tail sequence enabled a broad fingerprinting of the histone code for multiple different methylation sites and states.[45]

As the isolation step of the DCL concept was somewhat laborious, the Waters group more recently established the "imprint-and-report" method, a two-step process.[46] In the first step, a guest is added to the DCL, forming a distinct set of molecular interactions that imprint a memory of the analyte. In the second step, a second guest is introduced, competing with the first for binding to the templated species, producing a unique signal. The DCL does not re-equilibrate after the first step, ensuring that the imprint of the original guest remains intact. This method was successfully applied to histone peptides bearing methylated lysine and arginine residues. Building blocks E, G, and N, paired with fluorescent receptors (**LCG**, acridine orange, and thioflavin T), were used to probe seven Ac-YGGQTARme$_x$Kme$_y$STG-NH$_2$ peptides. The arrayed fluorescence data, analyzed *via* PCA, allowed full discrimination of seven methylation PTMs in the sequence, and the imprint-and-report method vastly improved the workflow and speed of the analysis, making it a more widely applicable sensor than the original process, with no loss of efficiency. The method has subsequently been applied to detect trimethylamine N-oxide in complex mixtures,[47] and detect perfluoroalkylsubstances at low concentrations in lake water.[48]

Supramolecular Tandem Enzyme Assays

The logical extension of recognizing and sensing biomolecules is to couple the recognition with enzymatic reactions, allowing the *in* situ monitoring of enzyme function.[49] This obviously has numerous additional challenges – the receptor:reporter system must be highly selective in discriminating reactant and product, tolerant to the enzyme itself, as well as being unaffected by all the cofactors and biomedia necessary for effective function. When using synthetic macrocycles as recognition units, the process has been termed a supramolecular tandem assay (STA).[50] The receptor can either recognize the reactant or the product of the enzymatic reaction; either way, the indicator dye acts as a competitor, and as the reaction proceeds, a change in fluorescence is triggered. The direction of the fluorescence response, whether an increase (switch-ON) or decrease (switch-OFF), depends on the specific properties of the dye and host used. Numerous groups have applied STAs to their selective modification detection systems, notably Nau, who implemented this approach in monitoring histone lysine methyltransferases activity using a **CX4-LCG** reporter pair.[51] The Dim-5 enzyme from *Neurospora crassa*, which trimethylates histone peptide H3K9, was applied to two oligopeptides from the histone H3 tail. As unmodified lysine residues bind weakly to **CX4**, **LCG** will outcompete the reactant and bind in the **CX4** cavity which causes quenching. Upon enzymatic trimethylation, the Kme$_3$ product binds more strongly to **CX4** and displaces **LCG**, leading to fluorescence enhancement. The selectivity is such

that there is minimal interference from the enzyme itself, or its cofactors, and this simple optical process allows real-time monitoring of the Dim-5 activity. This concept has been used widely by the Nau group, with examples focusing on enzymatic redox processes,[52] proteases[53] and many others.[54] Guo has used cationic calixarenes as hosts, and monitored the formation of pyridoxal phosphate[55] and pepsin activity.[56]

Waters et. al took a different approach in studying histone deacylase processes, developing a method that enables site-selective covalent labelling.[57] The strategy involved noncovalent recognition of Kme_3 peptides by a trisulfonated calix[4]arene conjugated to a fluorogenic electrophile (O-aryl nitro-benzoxadiazole, **ONBD**). Binding to Kme_3 brings **ONBD** near a neighboring unmethylated lysine, triggering proximity-driven covalent labelling. This allows application to histone deacylase activity, as the removal of the acetyl group in Kac forms a free lysine, which can be covalently labelled with **ONBD**. Thus, as the process requires a Kme_3 group in proximity to the requisite lysine for labelling, site-selectivity in deacylase monitoring is possible.

Intracellular Detection

Target detection in water, buffer or serum has its challenges due to external interferents, but intracellular detection with cavity-based receptors is far more complex. If the binding affinity and selectivity is suitably strong, however, intracellular detection is feasible. An early example came from the Nau group, who utilized the **CX4**/**LCG** pair to study analytes in live cells.[58] **CX4** has a sufficiently high affinity for acetylcholine (K_d = 10 μM) and protamine (K_d = 0.8 nM) to allow IDAs to occur in cytoplasm – the targets can displace **LCG** from **CX4**, leading to enhanced fluorescence. However, simple calixarenes such as **CX4** are somewhat limited in their selectivity and performance, so are not widely used. Guo has shown substantial applications of amphiphilic calixarenes in intracellular detection, drug transport and imaging.[59] The amphiphilic calixarenes are biostable and bind cationic peptides, allowing membrane transport.[60,61,62] This allows a variety of medicinal applications to be performed with these calixarenes, including phototheranostics,[63] inhibition of amyloid fibrillation,[64] cell line discrimination[65] and the detection of cancer biomarkers.[66] This work is far too substantial to discuss in detail here, and readers are directed to some reviews on the topic.[67]

Arrayed Deep Cavitands as Biosensors

Our contributions to the field center around the use of flexible, water-soluble deep cavitands for molecular recognition and biosensing purposes. Examples of relevant cavitand structures are shown in Figure 4 - as pioneered to Cram[68] and Rebek,[69] the cavities are generally deeper than those of calixarenes, and the flexible "walls" are held together by self-complementary hydrogen bonds. Benzimidazole cavitands (such as **TCC**, **CHI**, **NHC**, Figure 4) exploit hydrogen bonds with water for folding, so predominantly exist in the folded "vase" conformation in aqueous solvents. In contrast, the amide seam in octamide cavitands such as **AMI**, **AMP** is interrupted by water, so they exist in the open "kite" conformation unless a suitable guest is added.

Cavity-Based Sensing

Our initial applications focused on cavity-based recognition and indicator displacement assays.[70] The benzimidazole cavitands were initially created as acetylcholine hosts,[71] and are highly selective for trimethylammonium ($R\text{-}NMe_3^+$) salts. This suggested that the hosts would be able to bind lysine methylations, notably Kme_3. The cavity strongly binds the NMe_3^+ group, and its open-ended nature allows for indicator displacement sensing allowing for the detection of Kme_3 PTMs at various positions on longer peptide strands.

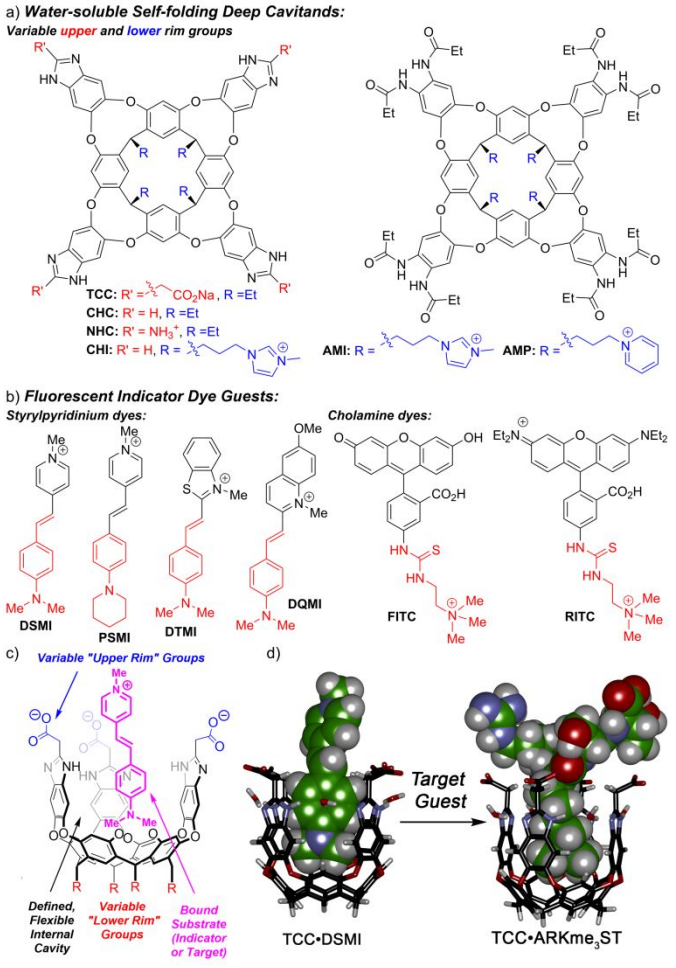

Figure 4. Structures of receptor-reporter pairs used for biosensing. a) Deep cavitand hosts. b) Indicator dyes. c) Illustration of the host:dye complexation event. d) Optimized structures of cavitand **TCC** binding **DSMI** dye or the modified pentapeptide strand $ARKme_3ST$.

When paired with an $R\text{-}NMe_3^+$-containing cholamine dye such as **FITC** (Figure 4), the **TCC•FITC** host:guest complex self-aggregates, leading to fluorescence quenching, which is disrupted upon indicator displacement by a suitable target. The **TCC•FITC** sensor could detect trimethylated peptides based on

the Histone H3 sequence, for example H3K9Me$_3$(1-21). The cavitand:dye sensor can easily differentiate between unmethylated, monomethylated, and trimethylated histone H3K9 peptides at low μM [peptide].

The easily tunable nature of the **TCC** scaffold allows this concept to be extended to differential sensing. Combining three cavitands (**TCC, CHC, NHC**) with a rhodamine-based cholamine dye variant (**RITC**) in a 6-component array (Figure 5a) enables selective discrimination of H3Kme$_3$ modifications at different positions on the peptide backbone - while the core cavity recognizes the R-NMe$_3^+$ group (i.e. state-selectivity), the variable upper rim functions interact differently with adjacent sidechains on the peptide, allowing site-selectivity in discrimination.[72] The differences in affinity between the hosts and targets are small, but these small differences can be teased out by differential sensing.

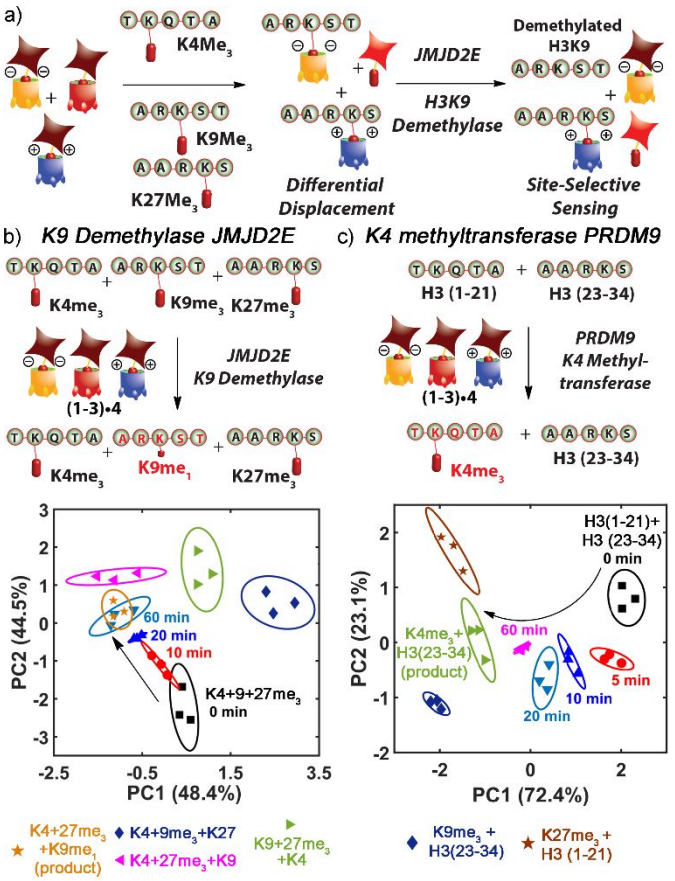

Figure 5. a) Illustration of the supramolecular tandem assay concept. b) PCA scores plot of the selective demethylation of K9me$_3$ in a mixture of K4me$_3$, K9me$_3$ and K27me$_3$ peptides by JMJD2E with the **TCC/CHC/NHC/RITC** array; c) PCA scores plot of the selective methylation of K4me$_3$ in a mixture of H3 (1-21) and H3 (23-24) peptides by PRDM9 with the **TCC/CHC/NHC/RITC** array. Ellipses drawn at 95% confidence. Adapted with permission from ref 73, copyright 2017, American Chemical Society.

10 different peptides were applied to the array, including those with Kme$_3$ modifications at different sites of the H3 strand (K4, K9, K27, K36, K79), peptides of different length (21 residues vs 11) and peptides with multiple modifications (K4me$_3$K9AcS10p and K79me$_3$T80p). Full discrimination was observed after PCA with the array at 95% confidence – the use of multiple cavitands at varying pH enables selective charge matching or mismatching interactions with the non-Kme$_3$ groups at nearby residues, enabling site and state-selectivity. This sensing can also be extended to mixtures of Kme$_3$ modified peptides.[73] Three different peptide strands (H3K4me$_3$, H3K9me$_3$ and H3K27me$_3$) were combined and analyzed by the array. When one of the strands was replaced by its unmethylated counterpart in the mixture, full discrimination was possible – the sensing was so sensitive that sequential changes of Kme$_3$/K ratio could be detected, which led to the use of the array as a site-selective demethylase sensor.

The selectivity for Kme$_3$ groups is so strong that the cavitand:dye arrays can function in the presence of chromatin writer and eraser enzymes (and their cofactors).[73] The first example was shown with the **TCC•FITC** pair, which successfully monitored the activity of histone demethylase JMJD2E, an enzyme that catalyzes the demethylation of H3K9Me$_3$. The H3K9me$_3$ peptide outcompetes the **FITC** dye from the host, but as the peptide was demethylated by the enzyme, the dye is bound by the cavitand, leading to fluorescence quenching. The host:dye pair is fully tolerant to all the necessary cofactors in the enzyme reactions, including 2-oxoglutarate, Fe^{2+} salts, and ascorbate, as well as the enzyme itself.

The success of the **TCC•FITC** system in demethylase monitoring was encouraging, but not unexpected, as Nau had already shown that **CX4** could monitor that process.[51] The benefit of cavitands, however, is in *site-selective* discrimination. Whereas most supramolecular tandem assays focus on changes in small molecules, lysine demethylase and methyltransferase can act on specific sites on the peptide. As the **TCC/CHC/NHC•RITC** array was capable of site-selective Kme$_3$ sensing with purified peptides and in peptide mixtures, we next asked the question whether it could also monitor enzyme activity with site-selectivity. JMJD2E is a K9-selective demethylase, so the **TCC/CHC/NHC/RITC** array was applied to sense JMJD2E demethylase activity in a mixture of H3K4me$_3$, H3K9me$_3$ and H3K27me$_3$ peptides (Figure 5b). Only the H3K9me$_3$ peptide was demethylated in the reaction, and the final product (as validated by MALDI-MS) was the monomethylated H3K9me peptide. The **TCC/CHC/NHC/RITC** array was able to monitor the selective removal of the methyl groups in real time – the peaks in the PCA scores plot move towards the location of H3K9me/H3K4me$_3$/H3K27me$_3$ peptide product mixture over time, nicely illustrating the site-selectivity. Interestingly, the array was also able to identify the state-selectivity of the reaction, as the final product mixture was fully separated from the "complete" demethylation product mixture that was not observed with this particular batch of JMJD2E. The array was tolerant to all the reaction components, including cofactors, physiological buffer conditions and the enzyme itself, and still maintained full fidelity in site- and state-selective reaction monitoring. It was also capable of monitoring the reverse reaction – when the methyltransferase PRDM9 was used to

selectively methylate a mixture of 2 peptide strands H3(1-21) and H3(23-24), trimethylation occurs selectively at K9. Exposing the reaction mixture over time to the **TCC/CHC/NHC/RITC** array allowed selective monitoring of the site-selective methylation process (Figure 5c).

Intracellular Detection

The charged deep cavitands are also able to recognize suitable targets in cellular environments. These systems are not amenable to differential sensing, but the affinity and selectivity of **TCC** for R-NMe_3^+ salts is strong enough that molecular recognition of these species is possible, even in the presence of the myriad constituents of a cell.[74] Obviously, this presents a challenge in detection, so the initial tests were performed with fluorescent guests to simplify analysis. **TCC** is highly lipophilic, and it is strongly bound in artificial membranes such as supported lipid bilayers[75] or micelles/vesicles.[76] When added to HeLa cells in the presence of the **FITC** dye, selective transport of the dye occurred into the cytoplasm of the cells. The mechanism of transport was shown to be endocytosis – the **FITC** dye is poorly transported across the cell membrane in the absence of host, but when incubated with **TCC**, molecular recognition occurs at the external membrane, allowing endocytosis of the **TCC•FITC** dye complex. This did not occur in giant unilamellar artificial vesicles, and the endocytosis could be slowed in the presence of growth inhibitors. The transport could occur in other cell types, such as astrocytes, but the process was limited by the fact that the **TCC** host released the dye after transport, and the dye was solely localized in cellular endosomes. Still, this illustrates the power of the host system – it is capable of selective R-NMe_3^+ salt recognition in highly complex environments.

While the **FITC** dye was limited in use, as it was unable to escape endosomes upon endocytosis, the styrylpyridinium dyes **DSMI** and **DTMI** are more interesting.[77] The dyes show strong affinity for nucleotide-rich cellular regions, especially the nucleolus and mitochondria. When these dyes were complexed with cavitands, the location of the dyes within the cell could be controlled. Specifically, pre-mixing the dye with **TCC** before adding it to the cells prevented the dye from being washed out, unlike in the absence of cavitand. **TCC** also shifted the dye from nucleotide-rich areas to the outer membrane, demonstrating its ability to sequester and control the dye's cellular distribution (Figure 6). Intracellular indicator displacement assays were also possible. Strongly binding R-NMe_3^+ salts such as choline and butyrylcholine displaced the dye from the cavitand, resulting in fluorescence changes ("turn-off" or "turn-on", depending on the cavitand used), while weakly bound substrates like betaine had minimal effect. This demonstrated that cavitands can selectively bind and release dyes based on guest binding strength, which controls the dye localization within the cell. The cavitand-dye systems also did not cause any substantial cell death, making them suitable for more advanced biological applications.

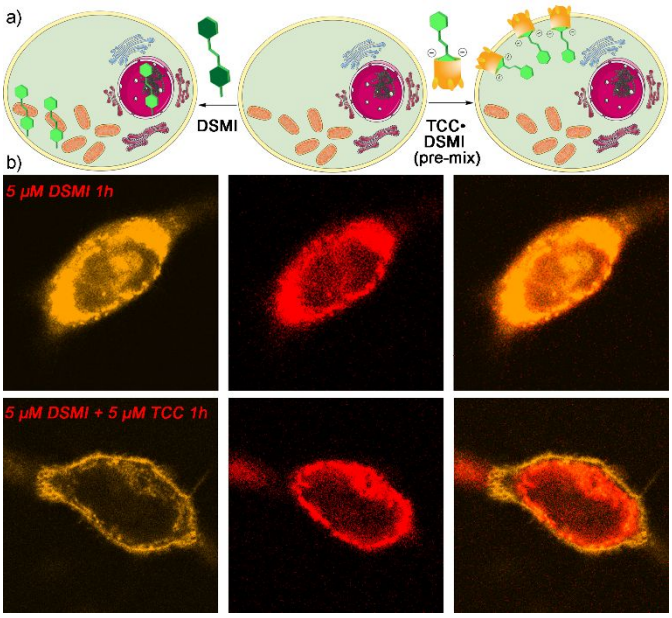

Figure 6. a) Illustration of the different cellular locations shown by **DSMI** and the **TCC•DSMI** premixed complex. b) Confocal fluorescence images of live HeLa cells treated by 5 μM **DSMI** for 1 h (top), and 5 μM **DSMI•TCC** (**pre-mix**) for 1 h (bottom). The images of 2 channels including both **DSMI**, and MitoView650, as well as their Merged images, are listed from left to right. Adapted with permission from ref 77, copyright 2023, American Chemical Society.

Indirect Sensing of Peptide Modifications

Using a defined cavity for target recognition is extremely powerful, but it can be limited in scope. Specific targets are perfectly suited for certain receptors - deep cavitands[72] and functionalized calixarenes[34] are highly selective for Kme_3, phosphonate cavitands are selective for NH_3^+,[3c] and slightly flattened toroidal cyclophanes show excellent selectivity for Rme_2^+.[43] The challenge is that creating a cavity-based receptor for other targets requires a full reimagining of the structure. This is possible, and some powerful hosts have been created for other species such as phosphate,[78] among many others, but this is time-consuming. What if a single scaffold could recognize and/or optically multiple different types of targets *via* orthogonal mechanisms? This is the advantage of the deep cavitand system - by changing the nature of the flexible structure, including the folding motif and the appended charged groups, as well as the nature of the bound indicator dye, the central scaffold can be repurposed to detect species that *do not bind in the central cavity.*

The initial example of this was discovered somewhat accidentally - the original work using **TCC** to sense Kme_3 modifications[70] showed evidence that the anionic cavitand could bind with non-modified cationic Histone H3 peptides at higher peptide concentration. The **TCC•RITC** system was not appreciably affected by this interaction, so strong selectivity for Kme_3 peptides was possible over unmodified ones. However, simply changing the dye to the styrylpyridinium **DSMI** enabled alternate detection selectivity. Whereas the affinity of **TCC** for the **RITC** dye (K_d = 1.5 μM) was very strong, and only strongly

binding Kme_3 targets could displace it, the affinity for **DSMI** is slightly weaker (K_d = 23.1 μM), and this allows indicator displacement upon a non-cavity-based interaction between anionic **TCC** and cationic histone H3 peptides. As the interaction is based on the cavitand's charge and lipophilicity, altering the charge of the target peptide changes the affinity for the host – the **TCC•DSMI** host:guest system was capable of detecting phosphorylation modifications on H3 peptides (Figure 7a), even in the presence of other modifications.[79]

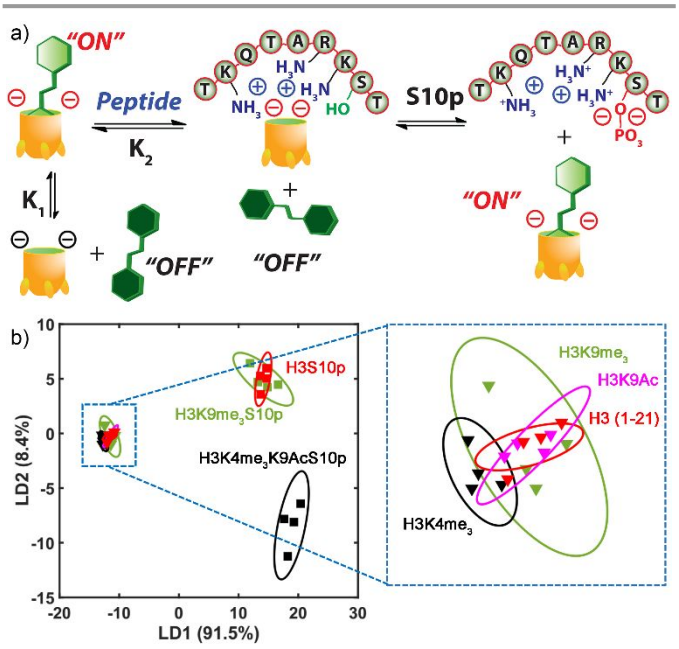

Figure 7. a) Illustration of the indirect phosphopeptide sensing mechanism with the **TCC** cavitand. b) LDA scores plot showing selective phosphopeptide sensing with a 4-component array with **TCC**, **DSMI** and metal salts as components, derived from individual sensor component responses to different PTM types. Adapted with permission from ref 79, copyright 2018, American Chemical Society.

Importantly, as the detection mechanism is indirect, and does not involve binding the added phosphate, the sensing was functional in phosphate buffer. By adding heavy metal ions that can chelate to the **TCC** cavitand[80] to the system, an array could be formed that allowed discrimination between phosphorylated and non-phosphorylated peptides, including 21-residue H3S10p and $H3K4me_3K9AcS10p$ (Figure 7). The sensor also detected phosphorylation on tyrosine (e.g. CREBtide-Yp) and threonine (e.g. MBP-Tp), and the array is unaffected by small-molecule phosphates (e.g., ATP, cAMP), ensuring the lack of interference by these cofactors. Because of this, the **TCC:DSMI** system could be used to monitor real-time phosphorylation of H3 peptides by protein kinase A (PKA) or Aurora B Kinase, as well as dephosphorylation by alkaline phosphatase. This recognition system was also able to detect other small charge based variations in peptide structure, such as the conversion of arginine sidechains to citrullines – converting the cationic guanidinium units in R sidechains to neutral ureas reduces affinity of the **TCC** cavitand for the peptide, enabling dye binding.[81] This concept was subsequently used by Nau and Hennig to monitor tyrosine phosphorylation.[82]

Cationic octapeptides bind 10-100 times more strongly to the anionic **CX4** than their phosphotyrosine variants. The receptors also sense phosphorylation by observing a reduction in substrate affinity, so is less sensitive to interference from common kinase cofactors like ATP or cAMP, and can be applied to monitor the dephosphorylation of phosphotyrosine by alkaline and acid phosphatases.

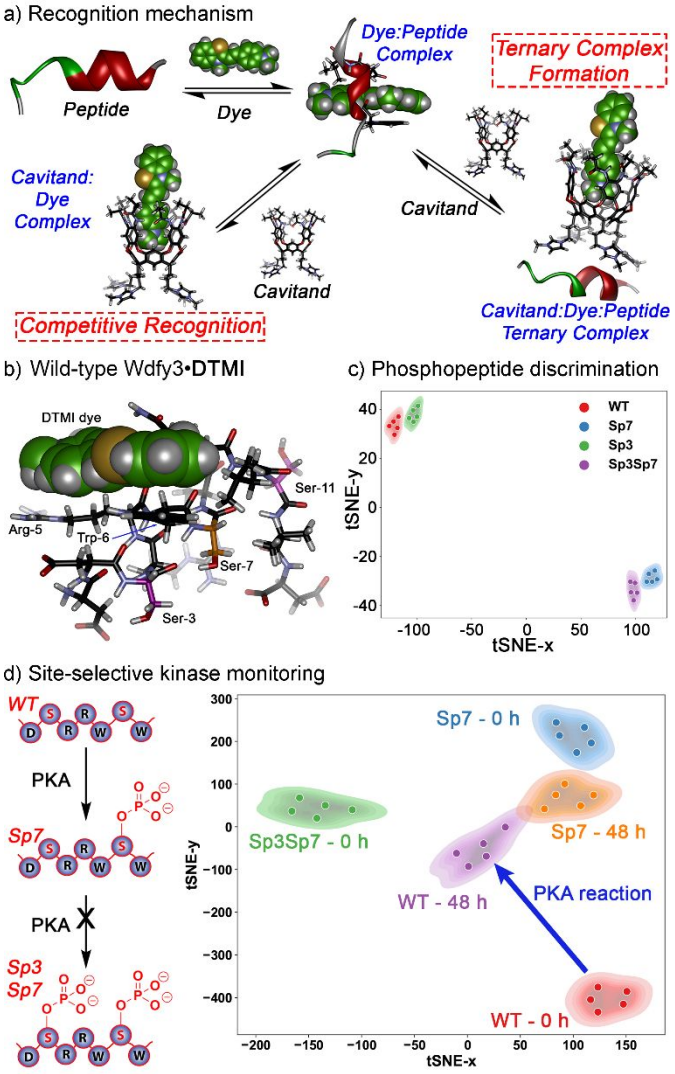

Figure 8. a) Illustration of the "inverted" phosphopeptide recognition process, showing various complexes formed by the host, dye and peptide combinations. b) Optimized structure of **DTMI** and wild-type Wdfy3 peptide determined by molecular dynamics simulation. c) t-SNE plot of the fluorescence responses from an optimized 3-element array, discriminating WT vs Sp7 vs Sp3 vs Sp3Sp7 Wdfy3. d) t-SNE plot from F/F_0 data of the PKA reaction with various Wdfy3 peptides using the optimal 3-element array. Adapted with permission from ref 83, copyright 2025, American Chemical Society.

The **TCC•DSMI** sensing array was effective in detecting phosphorylations, but the target scope was quite limited, requiring strongly cationic peptides such as H3(1-21) for **TCC** affinity. In addition, the targets chosen were not amenable to creating site-selective phosphorylation and kinase sensors. A more desirable target is the Wdfy3 peptide sequence, which is a substrate for PKA, and is selectively phosphorylated by the

enzyme at Ser-7. This sequence contains two other serine residues (Ser-3 and Ser-11), so is an excellent target to illustrate site-selective sensing. It is, however, anionic, and has no affinity for **TCC** at all. To solve this problem, we focused on an alternate detection mechanism - instead of using the cavitand as the "host" for the biological target and varying the nature of the indicator dye, can the dye be used as the recognition motif and the cavitand(s) added as "mediator" of these interactions?

Fortunately, the styrylpyridinium dyes are able to recognize anionic peptides such as Wdfy3, so long as they contain an electron-rich residue such as Trp.[83] Figure 8 shows the recognition system, which is quite complex, but centers around the affinity of dyes such as **DTMI** for the peptide. Cation-π and π-π stacking interactions drive the affinity, and cause an enhanced fluorescence response, which is altered when the dyes are added to phosphorylated peptides: in Wdfy3 Sp3, the phosphate group was oriented away from the dye, showing similar fluorescence to the unphosphorylated peptide. In Sp7, the phosphate group disrupted interactions, reducing fluorescence. The addition of cavitands changed the recognition in two ways – either the host competitively removed the dye from the peptide (e.g. for anionic **TCC**), or formed heteroternary cavitand•dye•peptide complexes (for cationic **AMI**).

This allowed a cavitand-dye array to be used for site- and state-selective phosphorylation sensing. This challenging selectivity required an optimized sensing array, which was achieved by applying machine learning techniques. Techniques such as SVM-RFE (support vector machine recursive feature elimination)[84] are powerful pattern recognition tools, as they can detect hidden patterns in complex data sets, as well as allowing prediction of unknown groups *via* data set training. By determining the sensor elements that provided the largest contribution to the discrimination, a minimal array can be determined. In this case SVM-RFECV suggested that three elements were sufficient: **DTMI•AMI**, **DQMI•AMI**, and **DSMI•TCC**. Analysis of the fluorescence responses by t-SNE plot[85] showed full differentiation of all 4 Wdfy3 variants, including the unmodified peptide, Sp3, Sp7 and the diphosphorylated Sp3Sp7 (Figure 8c). The array was also tolerant to kinase enzymes and their cofactors, as was the **TCC•DSMI** array described above.[79] As this array showed site-selectivity in phosphopeptide sensing, we applied it for site-selective kinase activity monitoring. PKA phosphorylates the Wdfy3 strand at Ser-7 only, so the optimized 3 component array (Figure 8d, consisting of 3 combinations of **TCC**, **AMI**, **DTMI** and **PTMI**, optimized by SVM-RFE) was applied to monitor the selective incorporation of phosphate at Ser-7. The fluorescence responses obtained as aliquots of the reaction mixture were added to the optimized array were subjected to t-SNE analysis, and this clearly showed the responses moving towards overlap with the Wdfy3Sp7 product. The sensing was somewhat challenging, as the PKA enzyme did not completely phosphorylate the peptide (yield as determined by MALDI-MS was ~ 70%), so perfect overlap was not obtained, but the results illustrate the sensor selectivity.

The "inverted" dye-centric recognition can also be applied to non-phosphorylation modifications.[86] Peptide isomerization occurs in long-lived proteins such as αB–crystallin, and proteins strongly associated with aging and brain functions, like Aβ and Tau.[87] Identification of peptide isomers is challenging, as they show no mass changes and cannot be easily differentiated by MS without peptide modification or complex ionization processes.[88] Optical detection is even more difficult, as there are no changes in functional group upon isomerization, merely the larger folded structure of the protein. This is an enticing target for the inverted detection mechanism, though, as all that is required is for the peptides to bind the dye. The αB–crystallin 57-69 strand contains a W59 residue and is overall anionic (pI = 4.09), so is well-suited to bind the cationic dyes in a similar manner to the Wdfy3 sequence described above. Indeed, **DTMI** shows a low micromolar affinity for the αB 57-69 sequence in Tris buffer, although it should be noted that the affinity profile is complex and multiple different stoichiometries and conformations of the **DTMI•**αB 57-69 occur.

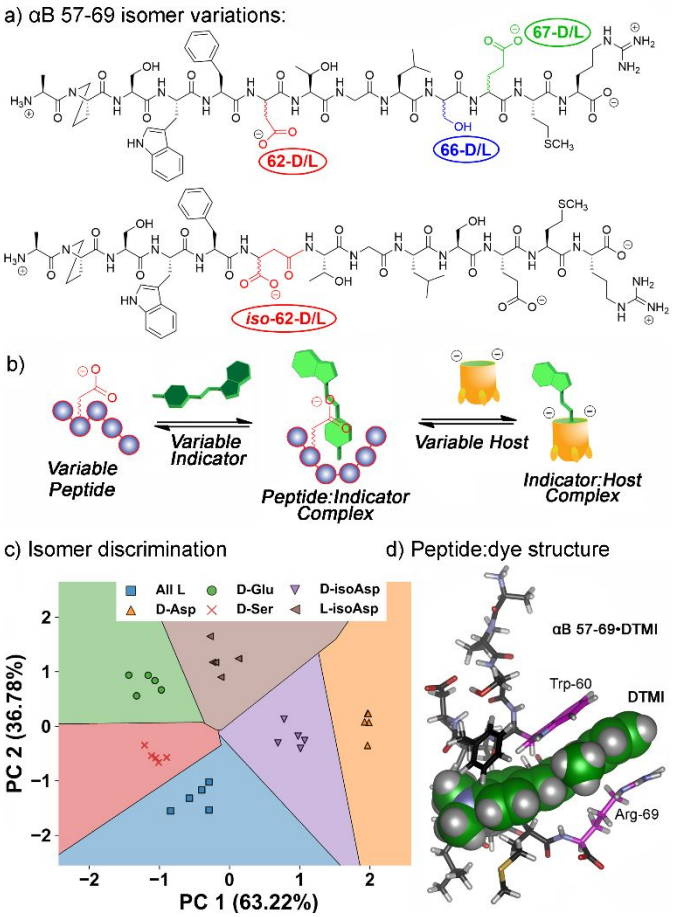

Figure 9. a) Isomeric variations in αB 57-69 tested. b) Illustration of the peptide isomer recognition and sensing mechanism. c) Decision region boundary plot for isomer discrimination using the PCA of the 2 most optimal elements selected by SVM-RFECV: **TCC•4-DSMI** and **TCC•DTMI**. d) The optimized structure of **DTMI•**[αB 57-69] modeled by molecular dynamics simulations.[86]

The most favorable conformation was determined by molecular dynamics calculations and is shown in Figure 9d, again

illustrating the dye:tryptophan affinity. When the **TCC** cavitand was added to the dye:peptide mix, a competitive equilibrium was established (Figure 9b), which allowed a sensing array to be constructed, consisting of **TCC** and the **DSMI/DTMI** dyes. This array allowed discrimination of a series of peptide isomers that varied in the stereochemistry of a single residue in the αB 57-69 strand (Figure 9a). These included D/L isomers of Asp62, Ser66 and Glu67, as well as D/L iso-Asp62 residues (isomers commonly seen in aged eye tissue). After array optimization by SVM-RFE as before, it was established that a simple 2-component array of **TCC•DSMI** and **TCC•DTMI** was able to discriminate all 6 isomers fully (Figure 9c). This illustrates the power of the detection system – simply combining one cavitand and two dyes allows discrimination of identical peptide sequences that vary only in the stereochemistry at a single atom in a 13-residue peptide strand!

Indirect Sensing of Anions in Aqueous Solutions

The formation of ternary complexes between anionic peptides, cationic dyes and the cationic **AMI**-type cavitands introduces many more possibilities for detection. The cavity can be used for dye recognition, and the lower rim functional groups can be exploited for target recognition. If this lower rim recognition event causes an allosteric change in the cavity conformation, the dye affinity will change, and sensing is possible. This can be seen with the **AMI** cavitand and small anions such as I-.[89]

The lower rim imidazoles in **AMI** are not simply solubilizing groups, they can coordinate anions *via* CH-anion interactions. The flexible $(CH_2)_3$-imidazolium groups can chelate anions at the lower rim of the cavitand in pure water, as well as in buffer. The affinity is highest for I^- (K_d = 0.2 mM), but Br^- can also be bound (K_d = 2.6 mM). The more important facet of the recognition is that the lower rim binding causes an allosteric change in the cavity of **AMI** – when the anion is bound at the base, the walls of the cavitand are forces into the open kite conformation (Figure 10a), which expunges any bound dye, lowering observed fluorescence and enabling optical sensing of iodide in aqueous solution. The greatest sensitivity was seen for the **AMI•PSMI** host:dye pairing, which had a limit of detection for I^- of 21 µM in 10X PBS buffer. This illustrates the selectivity of the sensing, as 10X PBS buffer contains 1.7 M Cl^- ions! When other cavitands such as **CHI** were used, the emission changes were far smaller, indicating that the combination of cation recognition elements and a flexible, allosterically regulated cavity is necessary for effective sensing.

A more interesting and challenging anionic target is DNA. DNA G-quadruplex (G4) structures are non-canonical folds that exhibit subtle structural variations critical to their biological functions.[90] Discriminating between structural variations in DNA can be challenging, as most analysis techniques use sequencing, rather than structural analysis, and scXRD or NMR structural determination is time-consuming and requires large amounts of material.[91] Again, the inverted, dye-centric recognition mechanism is extremely useful. The styrylpyridinium dyes in Figure 4 are mitochondrial DNA stains, so show affinity for anionic oligonucleotides in cells, as described above.[77] This can also be leveraged to detect small structural modifications in DNA. Specifically, our system can distinguish between different G4 structures that share the same folding topology but vary in the number of G-quartet stacks or in the sequence at the 3' and 5' ends. Additionally, the system can classify these structures into the three primary topological types: parallel, antiparallel, and hybrid.

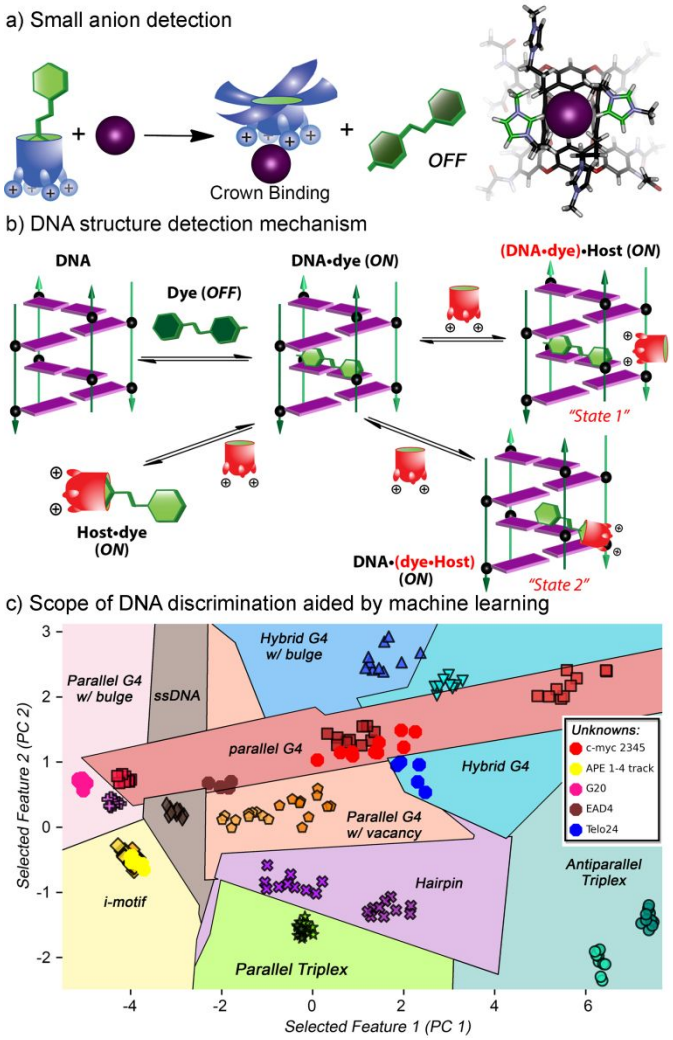

Figure 10. a) Illustration of the small anion sensing mechanism shown by the **AMI** cavitand, including the DFT-optimized structure of **AMI•I**.[89] b) Schematic illustration of the multiple interaction equilibria between DNA, host, and dye leading to differential sensing of DNA folding motifs, using an i-motif as the example structure. c) Decision Region Boundary plot using PC 1 and PC 2 obtained from subjecting the 16-element array data acquired from the 18-DNA pool by PCA-SVM-RFE. Five unknowns were projected to the regions representing the predicted folding structures. Adapted with permission from ref 93, copyright 2021, American Chemical Society.

A host:dye array involving 5 cavitands and two dyes was initially employed to sense the structural differences between G4 structures.[92] The **PSMI** and **DSMI** dyes bind folded DNA, and show increased fluorescence upon binding. When the dyes are combined with the cavitands, multiple different equilibria are established, which enables differential sensing. Single-stranded

DNA (A20, G20, T20, and C20), parallel G4 (c-kit1, c-myc 2345) hybrid G4 (bcl-2 2345) and an antiparallel G4 (TBA) could be easily discriminated with the initial array. The detection was applied to a wider scope of targets, notably 23 different G4 structures. PCA analysis demonstrated discrimination between subsets of the DNA strands, separating the DNAs by G4 length (2-stack, 3-stack or 4-stack G4s), as well as by topology (parallel vs antiparallel). Using Canonical Discriminant Analysis (CDA) enabled classification of the different DNAS by grouping, demonstrating a 98% precision in classification between the antiparallel, hybrid, and parallel G4 topologies.

The sensing mechanism is somewhat complex, but uses the same "indirect" sensing concept as described above. The dyes bind the folded DNA strands (as illustrated in Figure 10b), which causes an increase in emission. The added cavitands (notably **TCC**) can either act as simple competitive receptors for the dye (similar to the mechanism seen with αB crystallin peptides[86]), or they can form heteroternary **host•dye•DNA** complexes. These heteroternary complexes form with cationic cavitands such as **AMI**, reminiscent of the Wdfy3 phosphate sensing mechanism.[83] The synergistic combination of multiple equilibria in a complex sensing mixture enables the excellent differential selectivity of the sensor(s).

This sensor system was also capable of discriminating a far larger suite of targets.[93] In addition to the various G4 topologies, DNA G4s with bulges or vacancies, triplexes, i-motifs and hairpins were analyzed by the array. As the data suite was very large, we employed machine learning techniques as described above. The SVM-RFE algorithm was trained with data from a known 18-DNA data set (180 samples in total), and a classification model was obtained that permits the use of the fluorescence responses from an "unknown" DNA to predict its folding motif. Five DNA strands were chosen as "unknown" targets to test the predictive abilities of the algorithm, and as can be seen in Figure 10c, each of the "unknown" i-motif, parallel G4 and hybrid G4 repeats were successfully projected into the "correct" folding region. The only "failure" was single-stranded G20, which does not form a single conformation in solution, so was added as a control to validate the model. This method also allowed discrimination of "damaged" DNA G4 structures containing oxidation or methylation modifications on individual guanine bases in the strand.[94] The modifications can make subtle changes to the G4 folding that are undetectable by circular dichroism, but induce differential fluorescence responses in the array. The array is functional in the complex matrix of diluted serum, illustrating the selectivity of the sensor for small differences in DNA structure that can be as small as changing a single CH for a C=O group in a single base, even in the presence of interferents in real-world biomedia.

Future work

The host:dye array sensing systems described here are both tunable and powerful – multiple orthogonal recognition mechanisms can be applied to favor selective recognition of biorelevant targets with vastly different structures. Even using just a handful of hosts and guests, numerous sensor element combinations are possible. One of the advantages of using self-folding deep cavitands for this process is their propensity for aggregation. Aqueous receptors that are prone to aggregation are often avoided due to their challenges in characterization, but the formation of calixarene-based "receptor micelles" has been shown by Guo to enable a far wider scope of recognition and transport.[67] In our system, reversible aggregation mechanisms provide a second layer to recognition. The cavitands do not rely solely on cavity-based recognition, but can aggregate with themselves and/or large biomolecules, and when combined with indicator dyes, allow multiple detection mechanisms. This allows application of the system to multiple targets, not just cavity-binding R-NMe_3^+ salts. Using arrayed synthetic receptors has many advantages over other technical detection methods such as tandem LC-MS/MS. The detection is cheap and technically simple, is functional on a wide variety of unmodified substrates, and doesn't require complex separations. When compared to antibodies, synthetic receptors are far simpler to validate, enabling detailed analysis of the recognition mechanism.

This leads into some of the complexities of this method, which can be described as both an advantage and disadvantage. The complex mechanisms give rise to excellent target selectivity, but can be challenging to interpret – for example, how to choose the correct sensor elements for a particular target, and how to determine the dominant recognition mechanism in each case are often not obvious. In the age of Big Data, more advanced machine learning techniques and AI seem ideally suited to this task. Applying a "target-agnostic" strategy for analyte detection, whereby a large pool of hosts and dyes are applied in a high-throughput manner and the results analyzed by machine learning, could identify selectivity for complex targets not yet tested. These are powerful techniques that are seeing wide application in "omics" fields, but have not yet gained wide popularity in supramolecular host:guest chemistry, likely due to the relatively small number of privileged macrocyclic hosts that are effective as biosensors. As this pool increases in size, more complex, AI-driven analysis will become essential to manage large datasets. This will also be extremely important for analyzing targets (including mixtures of targets) in more complex media, especially biofluids, serum and cells where interferents are still highly problematic. This is the key limitation of the method at the moment – certain specific examples are functional in serum or living cells, but the high background fluorescence in serum is a challenge for optical detection methods. Salt and buffer are extremely well tolerated, but additives that cause quenching, such as heavy metals, complicate the detection. These limitations suggest some specific areas of future investigation – how can we apply these receptors for biomedical applications more broadly, and how can they target other, important biological species that are inaccessible with common synthetic receptors? Can these mechanisms be used to inhibit or enhance natural enzymatic

processes? As more synthetic receptor types are created, the further the field can expand.

Biographies

Arman C. Garcia received his MS degree from Cal. State Los Angeles with Professor Matthias Selke in 2019. He then continued his academic studies at University of Oregon and joined the laboratory of Professor Michael D. Pluth. Upon completing his Ph.D. in 2024 he began his postdoctoral studies at University of California, Riverside, with Professor Richard J. Hooley. His research focuses on aqueous supramolecular chemistry and molecular recognition.

Ria Lian received her B.S. in Chemistry with a biochemistry emphasis from California State Polytechnic University, Pomona in 2022. She began pursuing her Ph.D. in Analytical Chemistry at the University of California, Riverside in 2023. Under the guidance of Professor Richard J. Hooley, her research centers around applying a range of analytical methods to study molecular recognition. She is particularly interested in host-guest systems and is working to better understand the mechanisms that drive selective binding at the molecular level.

Wenwan Zhong is Professor of Chemistry at University of Science and Technology of China (USTC), specializing in separation and sensing. She received her B.S. degree in Applied Chemistry from USTC and her Ph.D. in Analytical Chemistry at Iowa State University under the direction of Prof. Edward S. Yeung. After finishing her postdoctoral training at the Los Alamos National Laboratory, she established her independent research group at University of California-Riverside in 2006, and was promoted to full Professor in 2016. She joined USTC in January 2024. Prof. Zhong's research mainly focuses on: 1) Development of liquid biopsy platforms to enhance early disease diagnosis; 2) Design of supramolecular host-guest sensor arrays to differentiate biomolecules; and 3) Advancement of separation technologies to assist with -omics study. Her work utilizes synthetic receptors, functional nucleic acids, and nanomaterials for target recognition and signal amplification and also employs open-channel separation and optical spectroscopy to assist with marker discovery.

Richard J. Hooley obtained his undergraduate degrees in Natural Sciences from Emmanuel College, Cambridge University, and his Ph.D. in the laboratory of Prof. Martin F. Semmelhack at Princeton. Following a postdoctoral position in the lab of Prof. Julius Rebek Jr. at The Scripps Research Institute in La Jolla, CA he began his independent career in the Department of Chemistry at the University of California – Riverside in 2008 and was promoted to Full Professor in 2018. The Hooley group works on various aspects of organic, inorganic and supramolecular chemistry, with a focus on the synthesis and applications of biomimetic self-assembled cages and hosts, and their functions in supramolecular catalysis, molecular recognition and biosensing.

Acknowledgements

The authors would like to thank the National Science Foundation (CHE-2305089 to R.J.H.) for funding, and the University of Science and Technology of China (to W.Z.) for support.

Conflicts of interest

There are no conflicts to declare.

Data availability

No primary research, software, or code has been included and no new data was collected, generated, or analysed as part of this feature article.

Notes and references

1 (*a*) R. Pinalli, A. Pedrini and E. Dalcanale, *Chem. Soc. Rev.*, 2018, **47**, 7006-7026. (*b*) S. van Dun, C. Ottmann, L.-G. Milroy and L. Brunsveld, *J. Am. Chem. Soc.*, 2017, **139**, 13960-13968. (*c*) J. Krämer, R. Kang, L. M. Grimm, L. De Cola, P. Picchetti and F. Biedermann, *Chem. Rev.*, 2022, **122**, 3459-3636.

2 (*a*) J.-H. Tian, Z. Zhang, Y.-C. Pan, Y. Wang and D.-S. Guo, *Responsive Materials.* 2025, **3**, e20240036. (*b*) Y.-L. Lu, X.-D. Zhang, Y.-H. Qin, J.-Q. Song, Y.-H. Huang, C.-H. Liu, J.-J. Chen, H.-S. Xu, M. Pan and C.-Y. Su, *Chem*, 2023, **9**, 2144-2160. (*c*) G. T. Williams, C. J. E. Haynes, M. Fares, C. Caltagirone, J. R. Hiscock and P. A. Gale, *Chem. Soc. Rev.*, 2021, **50**, 2737-2763 (*d*) M. J. Webber and R. Langer, *Chem. Soc. Rev.*, 2017, 46, 6600-6620.

3 (*a*) D.-S. Guo and Y. Liu, *Acc. Chem. Res.*, 2014, **47**, 1925-1934. (*b*) W. C. Geng, Z.-T. Jiang, S.-L. Chen and D.-S. Guo, *Chem. Sci.*, 2024, **15**, 7811-7823. (*c*) R. Pinalli and E. Dalcanale, *Acc. Chem. Res.*, 2013, **46**, 399-411. (*d*) L. Escobar and P. Ballester, *Chem. Rev.,* 2021, **121**, 2445−2514. (*e*) F. Hof, S. L. Craig, C. Nuckolls and J. Rebek, Jr. *Angew. Chem. Int. Ed.,* 2002, **41**, 1488−1508. (*f*) C. T. McTernan, J. A. Davies and J. R. Nitschke, *Chem. Rev.,* 2022, **122**, 10393−10437.

4 T. L. Mako, J. M. Racicot and M. Levine, *Chem. Rev.*, 2019, **119**, 322−477.
5 S. J. Barrow, S. Kasera, M. J. Rowland, J. del Barrio and O. A. Scherman, *Chem. Rev.*, 2015, **115**, 12320–12406.
6 T. Lizal and V. Sindelar, *Isr. J. Chem.*, 2018, **58**, 326–333.
7 K. D. Daze and F. Hof, *Acc. Chem. Res.*, 2013, **46**, 937–945.
8 T. Ogoshi, T. Kakuta and T. Yamagishi, *Angew. Chem. Int. Ed.*, 2019, **58**, 2197−2206.
9 J. E. Beaver and M. L. Waters, *ACS Chem. Biol.*, 2016, **11**, 643−653.
10 W. Zhong and R. J. Hooley, *Acc. Chem. Res.*, 2022, **55**, 1035–1046.
11 J. C. Ma and D. A. Dougherty, *Chem. Rev.*, 1997, **97**, 1303-1324.
12 J. H. Jordan and B. C. Gibb, *Chem. Soc. Rev.*, 2015, **44**, 547−585.
13 A. M. S. Riel, R. K. Rowe, E. N. Ho, A.-C. C. Carlsson, A. K. Rappé, O. B. Berryman and P. S. Ho, *Acc. Chem. Res.*, 2019, **52**, 2870–2880.
14 S. M. Biros, J. Rebek, Jr., *Chem. Soc. Rev.*, 2007, **36**, 93–104.
15 (*a*) J. H. Jordan, H. S. Ashbaugh, J. T. Mague and B. C. Gibb, *J. Am. Chem. Soc.*, 2021, **143**, 18605-18616. (*b*) H. R. Aziz, W. Yao, J. H. Jordan and B. C. Gibb, *Supramol. Chem.*, 2021, **33**, 266-271.
16 L. Cao, M. Šekutor, P. Y. Zavalij, K. Mlinarić-Majerski, R. Glaser and L. Isaacs, *Angew. Chem. Int. Ed.*, 2014, **53**, 988-993.
17 A. C. Sedgwick, J. T. Brewster II, T. Wu, X. Feng, S. D. Bull, X. Qian, J. L. Sessler, T. D. James, E. V. Anslyn, and X. Sun, *Chem. Soc. Rev.*, 2021, **50**, 9−38.
18 J. Chan, S. C. Dodani and C. J. Chang, *Nat. Chem.* 2012, **4**, 973–984.
19 B. T. Nguyen and E. V. Anslyn, *Coord. Chem. Rev.*, 2006, **250**, 3118–3127.
20 M. Inouye, K. Hashimoto and K. Isagawa, *J. Am. Chem. Soc.* 1994, **116**, 5517–5518.
21 K. N. Koh, K. Araki, A. Ikeda, H. Otsuka and S. J. Shinkai, *J. Am. Chem. Soc.* 1996, **118**, 755–758.
22 S. C. McCleskey, P. N. Floriano, S. L. Wiskur, E. V. Anslyn and J. T. McDevitt, *Tetrahedron*, 2003, **59**, 10089–10092.
23 P. A. Gale, L. J. Twyman, C. I. Handlin and J. L. Sessler, *Chem. Commun.* 1999, **35**, 1851–1852.
24 L. You, D. Zha and E. V. Anslyn, *Chem. Rev.*, 2015, **115**, 7840−7892.
25 P. C. Jurs, G. A. Bakken and H. E. McClelland, *Chem. Rev.*, 2000, **100**, 2649−2678.
26 S. Stewart, M. A. Ivy and E. V. Anslyn, *Chem. Soc. Rev.*, 2014, **43**, 70−84.
27 J.-H. Tian, X.-Y. Hu, Z.-Y. Hu, H.-W. Tian, J.-J. Li, Y.-C. Pan, H.-B. Li and D.-S. Guo, *Nat. Commun.* 2022, **13**, 4293.
28 A. C. Conibear, *Nat. Rev. Chem.*, 2020, **4**, 674−695.
29 K. W. Barber and J. Rinehart, *Nat. Chem. Biol.*, 2018, **14**, 188−192.
30 M. Luo, *Chem. Rev.*, 2018, **118**, 6656-6705.
31 S. D. Taverna, H. Li, A. J. Ruthenburg, C. D. Allis and D. J. Patel, *Nat. Struct. Mol. Biol.*, 2007, **14**, 1025−1040.
32 W. K. Paik, D. C. Paik and S. Kim, S., *Trends Biochem. Sci.,* 2007, **32**, 146−152.
33 C. S. Beshara, C. E. Jones, K. D. Daze, B. J. Lilgert and F. Hof, *ChemBioChem*, 2010, **11**, 63–66.
34 K. D. Daze, T. Pinter, C. S. Beshara, A. Ibraheem, S. A. Minaker, M. C. F. Ma, R. J. M. Courtemanche, R. E. Campbell, and F. Hof, *Chem. Sci.*, 2012, **3**, 2695–2699.
35 S. Tabet, S. F. Douglas, K. D. Daze, G. A. E. Garnett, K. J. H. Allen, E. M. M. Abrioux, T. T. H. Quon, J. E. Wulff and F. Hof, *Bioorg. Med. Chem.* 2013, **21**, 7004–7010.
36 R. E. McGovern, B. D. Snarr, J. A. Lyons, J. McFarlane, A. L. Whiting, I. Paci, F. Hof and P. B. Crowley, *Chem. Sci.*, 2015, **6**, 442–449.
37 S. A. Minaker, K. D. Daze, M. C. F. Ma and F. Hof, *J. Am. Chem. Soc.*, 2012, **134**, 11674–11680.
38 M. A. Beatty, J. Borges-González, N. J. Sinclair, A. T. Pye and F. Hof, *J. Am. Chem. Soc.*, 2018, **140**, 3500–3504.
39 M. A. Beatty, A. J. Selinger, Y. Li and F. Hof, *J. Am. Chem. Soc.*, 2019, **141**, 16763−16771.
40 P. C. Kearney, L. S. Mizoue, R. A. Kumpf, J. E. Forman, A. McCurdy and D. A. Dougherty, *J. Am. Chem. Soc.*, 1993, **115**, 9907–9919.
41 L. A. Ingerman, M. E. Cuellar and M. L. Waters, *Chem. Commun.*, 2010, **46**, 1839–1841.
42 L. I. James, J. E. Beaver, N. W. Rice and M. L. Waters, *J. Am. Chem. Soc.,* 2013, **135**, 6450–6455.
43 A. G. Mullins, N. K. Pinkin, J. A. Hardin and M. L. Waters, *Angew. Chem. Int. Ed.* 2019, **58**, 5282–5285.
44 A. G. Mullins, L. E. St. Louis and M. L. Waters, *Chem. Commun.*, 2020, **56**, 3947–3950.
45 B. C. Peacor, C. M. Ramsay and M. L. Waters, *Chem. Sci.*, 2017, **8**, 1422–1428.
46 E. E. Harrison, B. A. Carpenter, L. E. St. Louis, A. G. Mullins and M. L. Waters, *J. Am. Chem. Soc.*, 2021, **143**, 14845–14854.
47 E. E. Harrison and M. L. Waters. *Angew. Chem. Int. Ed.* 2022, **61**, e202205193.
48 E. E. Harrison and M. L. Waters, *Chem. Sci.* 2023, **14**, 928–936.
49 W. M. Nau, G. Ghale, A. Hennig, H. Bakirci and D. M. Bailey, *J. Am. Chem. Soc.*, 2009, **131**, 11558−11570.
50 R. N. Dsouza, A. Hennig and W. M. Nau, *Chem. Eur. J.*, 2012, **18**, 3444–3459.
51 M. Florea, S. Kudithipudi, A. Rei, M. J. González-Álvarez, A. Jeltsch and W. M. Nau, *Chem. Eur. J.*, 2012, **18**, 3521–3528.
52 F. Biedermann, D. Hathazi and W. M. Nau *Chem. Commun.,* 2015, **51**, 4977–4980.
53 G. Ghale, V. Ramalingam, A. R. Urbach and W. M. Nau, *J. Am. Chem. Soc.* 2011, **133**, 7528–7535.
54 G. Ghale and W. M. Nau *Acc. Chem. Res.* 2014, **47**, 2150−2159.
55 Y.-X. Yue, Y. Kong, F. Yang, Z. Zheng, X.-Y. Hu and D.-S. Guo, *ChemistryOpen* 2019, **8**, 1437−1440.
56 Z. Zheng, W.-C. Geng, H.-B. Li and D.-S. Guo, *Supramol. Chem.* 2021, **33**, 80−87.
57 I. N. Gober and M. L. Waters, *J. Am. Chem. Soc.*, 2016, **138**, 9452−9459.
58 A. Norouzy, Z. Azizi and W. M. Nau, *Angew. Chem. Int. Ed.*, 2015, **54**, 792–795.
59 Y.-C. Pan, X.-Y. Hu and D.-S. Guo, *Angew. Chem. Int. Ed.*, 2021, **60**, 2768−2794.
60 Y.-C. Pan, A. Barba-Bon, H.-W. Tian, F. Ding, A. Hennig, W. M. Nau and D.-S. Guo, *Angew. Chem. Int. Ed.* 2021, **60**, 1875–1882.
61 D.-Y. Zhang, Z. Zheng, H. Zhao, H.-Y. Wang, F. Ding, H.-B. Li, Y.-C. Pan and D.-S. Guo, *Chem. Commun.* 2021, **57**, 12627−12630.
62 A. Barba-Bon, Y.-C. Pan, F. Biedermann, D.-S. Guo, W. M. Nau and A. Hennig, *J. Am. Chem. Soc.* 2019, **141**, 20137−20145.
63 J. Gao, J. Li, W.-C. Geng, F.-Y. Chen, X. Duan, Z. Zheng, D. Ding and D.-S. Guo, *J. Am. Chem. Soc.* 2018, **140**, 4945−4953.
64 Z. Xu, S. Jia, W. Wang, Z. Yuan, B. J. Ravoo and D.-S. Guo, *Nat. Chem.* 2019, **11**, 86−93.
65 X.-Y. Hu, Z.-Y. Hu, J.-H. Tian, L. Shi, F. Ding, H.-B. Li and D.-S. Guo, *Chem. Commun.*, 2022, **58**, 13198–13201.
66 Z. Zheng, W.-C. Geng, J. Gao, Y.-Y. Wang, H. Sun and D.-S. Guo *Chem. Sci.*, 2018, **9**, 2087–2091.
67 (*a*) Y.-C. Pan, J.-H. Tian and D.-S. Guo *Acc. Chem. Res.* 2023, **56**, 3626−3639. (*b*) W.-C. Geng, J. L. Sessler and D.-S. Guo *Chem. Soc. Rev.*, 2020, **49**, 2303–2315.
68 J. R. Moran, J. L. Ericson, E. Dalcanale, J. A. Bryant, C. B. Knobler and D. J. Cram, *J. Am. Chem. Soc.*, 1991, **113**, 5707–5714.

69 D. M. Rudkevich, G. Hilmersson, J. Rebek Jr., *J. Am. Chem. Soc.* 1998, **120***,* 12216–12225.
70 Y. Liu, L. Perez, M. Mettry, C. J. Easley, R. J. Hooley and W. Zhong, *J. Am. Chem. Soc.*, 2016, **138**, 10746–10749.
71 F. Hof, L. Trembleau, E. C. Ullrich and J. Rebek Jr., *Angew. Chem. Int. Ed.*, 2003, **42**, 3150–3153.
72 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**, 3960–3970.
73 Y. Liu, L. Perez, A. D. Gill, M. Mettry, L. Li, Y. Wang, R. J. Hooley and W. Zhong, *J. Am. Chem. Soc.*, 2017, **139**, 10964–10967.
74 Y.-J. Ghang, M. P. Schramm, F. Zhang, R. A. Acey, C. N. David, E. H. Wilson., Y. Wang, Q. Cheng and R. J. Hooley, *J. Am. Chem. Soc.*, 2013, **135**, 7090–7093.
75 L. Perez, Y.-J. Ghang, P. B. Williams, Y. Wang, Q. Cheng and R. J. Hooley *Langmuir*, 2015, **31**, 11152–11157.
76 Y.-J. Ghang, J. L. Lloyd, M.P. Moehlig, J. K. Arguelles, M. Mettry, X. Zhang, R. R. Julian, Q. Cheng and R. J. Hooley, *Langmuir*, 2014, **30***,* 10161–10166.
77 J. Chen, E. Z. Tabaie, B. L. Hickey, Z. Gao, A. A. P. Raz., Z. Li, E. H. Wilson, R. J. Hooley and W. Zhong, *ACS Sens.*, 2023, **8**, 3195–3204.
78 (*a*) N. A. Esipenko, P. Koutnik, T. Minami, L. Mosca, V. M. Lynch, G. V. Zyryanov and P. Anzenbacher, *Chem. Sci.*, 2013, **4**, 3617-3623. (*b*) T. Zhang, N. Y. Edwards, M. Bonizzoni and E. V. Anslyn, *J. Am. Chem. Soc.,* 2009, **131**, 11976−11984.
79 Y. Liu, J. Lee, L. Perez, A. D. Gill, R. J. Hooley and W. Zhong, *J. Am. Chem. Soc.*, 2018, **140**, 13869–13877.
80 Y. Liu, M. Mettry, A. D. Gill, L. Perez, W. Zhong and R. J. Hooley, *Anal. Chem.*, 2017, **89**, 11113–11121.
81 A. D. Gill, B. L. Hickey, S. Wang, M. Xue, W. Zhong and R. J. Hooley, *Chem. Commun.*, 2019, **55**, 13259–13262.
82 Y. Liu, S. Peng, L. Angelova, W. M. Nau and A. Hennig, *ChemistryOpen,* 2019, **8**, 1350–1354.
83 J. Chen, P. Fashianifard, R. Lian, L. J. Gibson-Elias, J. L. Moreno Jr., C.-E. A. Chang, W. Zhong and R. J. Hooley, *J. Am. Chem. Soc.*, 2025, **147**, 841–850.
84 (*a*) C. Cortes and V. Vapnik, *Machine Learning,* 1995, **20**, 273–297. (*b*) O. Ivanciuc, *Rev. Comput. Chem.*, 2007, **23**, 291–400.
85 L. van der Maaten and G. Hinton, *J. Mach. Learn. Res.*, 2008, **9**, 2579–2605.
86 J. Chen, P. Fasihianifard, A. A. P. Raz, B. L. Hickey, J. L. Moreno Jr., C.-E. A. Chang, R. J. Hooley and W. Zhong, *Chem. Sci.*, 2024, **15**, 1885–1893.
87 T. R. Lambeth, D. L. Riggs, L. E. Talbert, J. Tang, E. Coburn, A. S. Kang, J. Noll, C. Augello, B. D. Ford and R. R. Julian, *ACS Cent. Sci.*, 2019, **5**, 1387–1395.
88 Y. A. Lyon, G. M. Sabbah and R. R. Julian, *J. Proteome Res.*, 2017, **16**, 1797–1805.
89 B. L. Hickey, A. A. P. Raz, J. Chen, J. L. Moreno Jr., J. D. Hartman, W. Zhong and R. J. Hooley, *Chem. Commun.*, 2023, **59***,* 7819–7822.
90 C. K. Kwok and C. J. Merrick, *Trends Biotechnol.*, 2017, **35**, 997–1013.
91 F. Zhang, J. Nangreave, Y. Liu. and H. Yan, *J. Am. Chem. Soc.*, 2014, **136**, 11198–11211.
92 J. Chen, B. L. Hickey, L. Wang, J. Lee, A. D. Gill, A. Favero, R. Pinalli, E. Dalcanale, R. J. Hooley and W. Zhong, *Nat. Chem*, 2021, **13**, 488–495.
93 J. Chen, A. D. Gill, B. L. Hickey, Z. Gao, X. Cui, R. J. Hooley and W. Zhong, *J. Am. Chem. Soc.*, 2021, **143**, 12791-12799.
94 J. Chen, B. L. Hickey, A. A. P. Raz, Z. Gao, R. J. Hooley and W. Zhong, *ACS Sens.*, 2022, **7**, 2164–2169.

Data availability

No primary research, software, or code has been included and no new data was collected, generated, or analysed as part of this feature article.