Using changes in speciation in a dynamic combinatorial library as a fingerprint to differentiate the methylation states of arginine

Journal:	*ChemComm*
Manuscript ID	CC-COM-01-2020-000415.R1
Article Type:	Communication

SCHOLARONE™
Manuscripts

COMMUNICATION

Using changes in speciation in a dynamic combinatorial library as a fingerprint to differentiate the methylation states of arginine[†]

Alexandria G. Mullins,[a] Lauren E. St. Louis,[a] and Marcey L. Waters*[a]

Herein we describe the development of a sensor array that utilizes the complex response of a dynamic combinatorial library (DCL) to discriminate all of the methylation states of Arg, previously unreported in a sensor array, as well as the methylation states of Lys. We find that the use of all species in the DCL, not just those that bind, allows for discrimination of analytes that are otherwise indistinguishable, demonstrating the value of utilizing a complex network of species for differential sensing.

Given their importance in epigenetic regulation of gene expression, there is great interest in sensing histone post-translational modifications (PTMs), including methylated lysine and arginine.[1–3] Recently, a number of synthetic hosts have been developed that bind methylated Lys[4–15] and Arg[7,16,17]. Several of these synthetic receptors have been used in *differential sensor arrays* to distinguish the methylation states of lysine,[18–21] but little work has focused on the sensing of methylated arginine,[19,20] despite the increasing implication of methylated Arg in a wide variety of diseases (Fig. 1).[22,23] Differential sensing,[24] sometimes referred to as "artificial nose" sensing, exploits several sensors to distinguish between a series of analytes via pattern recognition rather than relying on one selective "lock-and-key"[25] receptor. Typically, a solvatochromatic dye is found that binds to a synthetic receptor such that displacement of the dye by the analyte provides a signal, called an indicator displacement assay (IDA).[26] Combining several of these IDA sensors provides a unique pattern of signals, or fingerprint, for each analyte. While this approach is easier than achieving highly selective synthetic receptors, the need to identify and characterize a group of synthetic receptors with a range of binding properties and a corresponding solvatochromatic dye that functions in the right affinity range requires a significant commitment of time and resources (Fig. 2a).

We describe in this report a generalizable strategy for sensing the methylation states of Arg and Lys with two

significant advantages. First, this is the first sensor array capable of differentiating *all* of the methylation states of Arg, including the sensing of Rme1, which has not previously been accomplished. Second, this work provides a significantly streamlined workflow relative to traditional sensor arrays using changes in speciation in a dynamic combinatorial library (DCL)[27] directly as the fingerprint as described below (Fig. 2b and c).
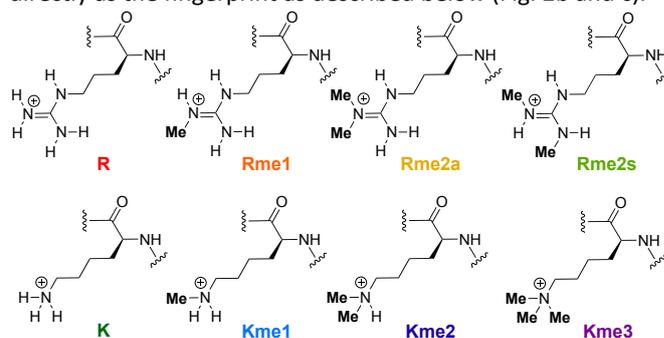

Fig. 1. Structures of the methylation states of Lys and Arg.

Dynamic combinatorial chemistry (DCC)[27] has proven to be a useful tool towards the development of differential sensor arrays, including development and isolation of synthetic receptors for traditional sensor arrays,[3] assessment of molecular similarity[28] and changes in the environment, such as pH and ionic strength.[29] The power of DCC to identify biologically-relevant analytes has also been demonstrated using metal-dye displacement[30] and covalent capture,[31] though these are not generalizable methods for sensing DCL analytes as most DCLs depend on noncovalent binding. We argue that a generalizable method for sensing DCL analytes has been accessible using the changes in speciation of DCLs as a fingerprint for identifying analytes, yet the power of this approach has been largely untapped (Fig. 2). Thus far, this strategy has only been reported for oligocarboxylates in DMSO.[32,33] The selective molecular recognition of analytes in water is a challenge[34] that continues to hinder the sensing of many biological analytes of interest. Therefore, a method that could be applied generally without the need to identify and isolate receptors could significantly decrease the burden of creating sensors for hydrophilic analytes in water.

a. *Department of Chemistry, CB 3290, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599*
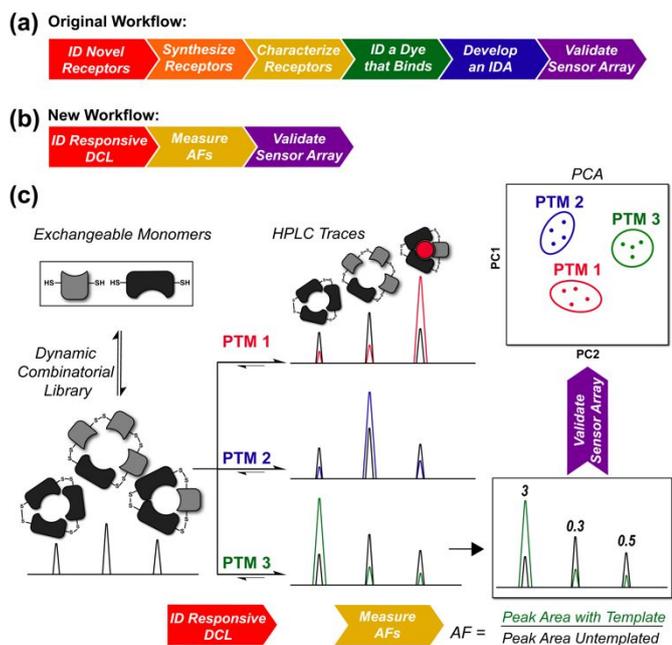
Fig. 2. The workflow for development of a traditional (a) and DCL-based sensor array(b and c): identification of a responsive DCL by HPLC, measurement of amplification factors (AFs) from the HPLC traces of DCLs, and generate a PCA score plot.

The methylation states of Arg have been a particularly challenging set of analytes to distinguish from each other and methylated Lys in water due to their similar size and hydrophilicity. To date, sensor arrays for methylated Lys have been developed using calixarenes, cavitands, and cyclophanes via the well-established IDA approach described above,[3–17,] but the bowl-shaped nature of these receptors provide poor binding to methylated Arg. To address this deficiency, we recently reported the isolation and characterization of a new high affinity, high selectivity synthetic receptor for asymmetric dimethylarginine (Rme2a), N_2G_2, which we identified from a dynamic combinatorial library (DCL).[16] DCLs are libraries of potential receptors under thermodynamic control, in which addition of an analyte perturbs the equilibrium to amplify receptor(s) that bind to it.[27] However, rather than pursuing a traditional sensor array as described above to achieve methylated Arg discrimination, we realized that the *DCL itself*, which was used to identity N_2G_2, *provided differential sensing directly*, as three unique receptors were amplified to various degrees in the presence of different methylation states of Arg and Lys (Fig. 2).

This approach provides rapid entry into differential sensor arrays by using the amplification factors (AFs) of the species in the library, as determined by HPLC integration, as the fingerprint using principal component analysis (PCA),[36] without requiring synthesis and isolation of the receptors or identification of a responsive dye. Moreover, we find that inclusion of data from nonbinding species can improve sensing, demonstrating the added value of information from the *entire network of species*, not just those that bind, to differentiate analytes. This approach can in theory be applied to any DCL for any analyte that has demonstrated a change in speciation.

Recently, we reported a DCL using monomers **E**, **G**, and **N** that exhibits unique amplification of four different receptors, depending on the guest: while N_2G_2 was amplified only in the presence of Rme2a, two isomers of ENG_2 were amplified in the presence of both Rme2a and Rme2s, while a fourth receptor, EG_3, was amplified in the presence of Kme3 (Figure 3).[16]

Exchangeable Monomers
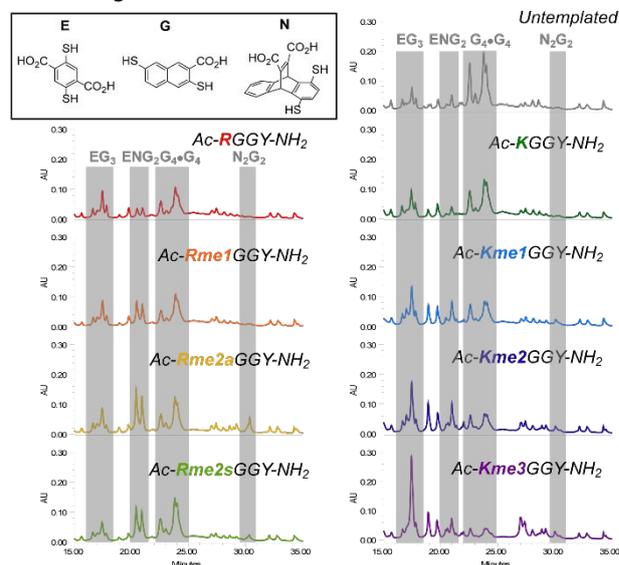
Combination of Amplification Factors Used as Pattern to Identify PTM

Fig. 3. The dynamic combinatorial libraries (DCLs) [16] of monomers **E**, **G**, and **N** used to generate an identifiable fingerprint for PTMs based on changes in the amplification factors of macrocycles EG_3, ENG_2, N_2G_2, and $G_4 \cdot G_4$ (highlighted in gray).

To determine whether these DCLs could be used directly for differential sensing, we first evaluated eight short XGGY peptides typically used for discovering new receptors, representing each of the possible methylation states of Lys and Arg (Table S1). All DCLs were equilibrated in 50 mM sodium borate buffer (pH 8.5) for 36 or more hours to allow full equilibration with 1.35 mM peptide guest, 0.34 mM **E**, 0.67 mM **G**, and 0.34 mM **N**. We then calculated the AFs for EG_3, ENG_2, and N_2G_2 with each peptide guest. The concentration of each receptor is determined by integrating the corresponding peak area, and amplification factors are calculated by dividing the concentration of each receptor in the templated library by the concentration in the untemplated library.[37] This data was analyzed by principal component analysis (PCA), an unsupervised multivariate statistical tool, to generate an easy-to-read representation of the data.[36] Each DCL was measured four times to provide four data points. We achieved excellent discrimination between each of the PTMs at 95% confidence (Fig. 4a), including Rme1 which has not previously been distinguished by a synthetic receptor or sensor array. Only the unmethylated Lys and Arg peptides exhibit some overlap of their 95% confidence ellipses. The vectors for each receptor
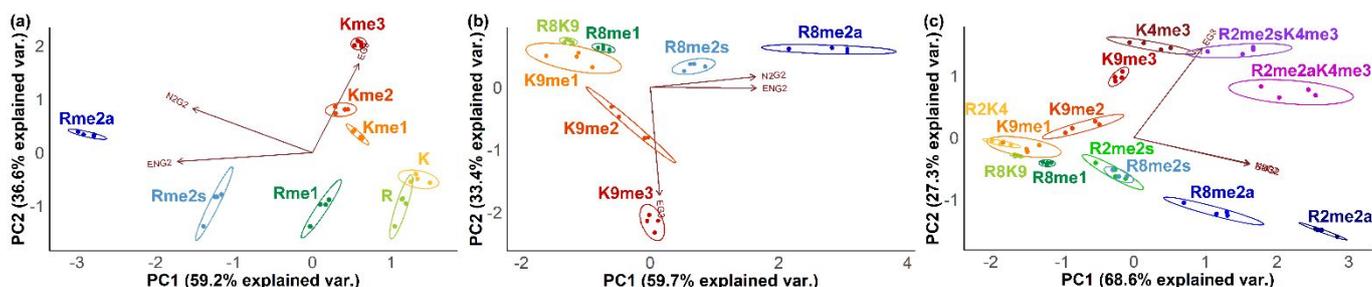
COMMUNICATION

Fig. 4. PCA score plots for different sets of analytes using AFs calculated from **EG$_3$**, **ENG$_2$**, and **N$_2$G$_2$** peaks with confidence ellipses drawn at 95%. Brown arrows represent the contribution of each receptor. (a) A PCA score plot for XGGY tetramer peptides. (b) A PCA score plot for histone peptides with methylation at R8 or K9 demonstrating methylation-state selectivity. (c) A PCA score plot for histone peptides with methylation at R2, R8, K4 and/or K9 demonstrating site selectivity.

that was monitored (**EG$_3$**, **ENG$_2$**, and **N$_2$G$_2$**) show the extent to which the amplification of each receptor contributed to distinguishing the PTMs. In agreement with the differences in amplification, the methylation states of lysine appear to be primarily discriminated by **EG$_3$**, and the methylation states of arginine by **ENG$_2$**. The unique selectivity of **N$_2$G$_2$** for Rme2a provides an additional factor for selectively, differentiating the mass-degenerate PTMs Rme2a and Rme2s. These results demonstrate the ability to achieve a selective sensor array using only three exchangeable monomers and their corresponding HPLC traces.

We then probed the ability of this technique to differentiate longer, more biologically relevant peptide sequences from the H3 histone protein (Table S1). We included peptides with Arg methylation at position 8 or Lys methylation at position 9 (Figure 4b). We also optimized an HPLC method from the original 60 minute method used for the XGGY peptides to a shorter 15 minute method (Fig. S11). Each of the methylation states of Lys and Arg in the R8K9 peptides are still differentiated at 95% confidence, including R8K9, R8me1, and K9me1. This is the first sensor array that has been shown to differentiate these three lower methylation states of Lys and Arg, and exhibits the power of the method, as none of the receptors is significantly amplified in the presence of these three peptides. Additionally, we demonstrated that separation was still observed between R2K4, R2me2a, and K4me3 at 10-fold lower peptide concentration (Fig. S15). Furthermore, we found that the sensor array can be used to quantify mixtures of analytes, such as different ratios of R2K4 and R2me2a, as would be needed for an enzyme assay[18,21,38] (Fig. S16).

We further analyzed whether the DCL could differentiate the same modification at different positions in the histone sequence, which is more challenging, by also evaluating peptides with Arg methylation at position 2 or Lys methylation at position 4 (Table S1 and Fig. 4c). The PCA plot successfully differentiated R2me2a and R8me2a as well as K4me3 and K9me3, demonstrating the ability to distinguish the position of the PTM without the need for LCMS despite the significant similarities in the sequences. We also investigated peptides with more than one PTM, and found that R2me2aK4me3 and R2me2sK4me3 can be differentiated from each other as well as from their corresponding singly modified peptides based on integration of the DCL (Fig. 4c). However, R2me2s and R8me2s

were indistinguishable using the sensor array based on **EG$_3$**, **ENG$_2$**, and **N$_2$G$_2$**.

Next we investigated whether there was more information in the system that could aid in differentiation of R2me2s and R8me2s. Monomer **G** is well known to form a **G$_4$•G$_4$** catenane in the absence of a template,[39] and this is observed in the untemplated library of monomers **E**, **G**, and **N**. We noted that this species decreases to different degrees in response to various analytes as monomer **G** is incorporated in different amplified species. We hypothesized that the decrease of **G$_4$•G$_4$** could provide an additional fingerprint for each analyte even though it does not bind them directly. Remarkably, while R2me2s and R8me2s are indistinguishable without inclusion of the **G$_4$•G$_4$** AFs, incorporating the AFs for the **G$_4$☐G$_4$** region significantly improved discrimination between R8me2s and R2me2s at 95% confidence (Fig. 5), with complete separation with 83% confidence ellipses (Fig. S17) by increasing the contribution of the PC2 component.[36] The ability to discriminate two analytes by inclusion of a response from a *non-binding species to increase differentiation* exemplifies the advantages of using a responsive network as a sensor array. This approach makes use of the complex information provided by the changes in speciation beyond the typical binding/no binding output.
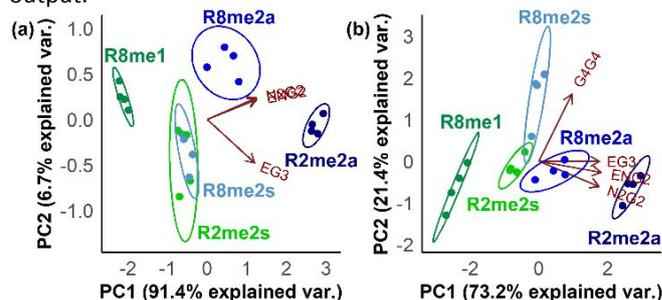

Fig. 5. PCA score plots demonstrating differentiation of histone peptides containing Rme2s without (a) and with (b) AFs calculated from the non-binding species **G$_4$☐G$_4$** in the DCL. Confidence ellipses drawn at 95%. Matrix generated using AFs calculated from **EG$_3$**, **ENG$_2$**, **N$_2$G$_2$**, as well as **G$_4$☐G$_4$** peaks in (b). Brown arrows represent the contribution of each receptor to the principal components.

In summary, we have demonstrated a generalizable method for achieving sensor arrays for methylated Arg and Lys by *directly* exploiting the differences in speciation as measured by HPLC resulting from differential amplification of a dynamic combinatorial library. Using a single library of only three monomers, we demonstrate the ability to differentiate 11

different combinations of PTMs, including position dependence and multiple PTMs in the same peptide. This is the first example of a sensor array capable of discriminating Rme1 from unmethylated or dimethylated Arg, a PTM linked to cancer and neurodegenerative disease.[22] This approach provides a significantly shorter workflow than for IDA-based sensor arrays for histone PTMs, without the need to design, isolate, and characterize selective receptors, and identify an appropriate solvatochromatic dye to achieve a sensor array. While it is not as high-throughput as a fluorescence or UV-vis based method, it is a general method that can be used *for many DCLs* in a rapid, straight forward manner, requiring only a unique pattern of amplification of the species in the library for different analytes. Furthermore, by using the DCL itself to discriminate unknowns, we make use of the rich chemical information encoded in the network of interconverting species, including species that do not bind to the analyte of interest, demonstrating the power of this Systems Chemistry approach.

Conflicts of interest

There are no conflicts to declare.

Notes and references

1   K. D. Daze and F. Hof, *Acc. Chem. Res.*, 2013, **46**, 937–945.
2   A. Shaurya, K. I. Dubicki and F. Hof, *Supramol. Chem.*, 2014, **26**, 538–590.
3   T. Gruber, *ChemBioChem*, 2018, **19**, 2324–2340.
4   C. S. Beshara, C. E. Jones, K. D. Daze, B. J. Lilgert and F. Hof, *ChemBioChem*, 2010, **11**, 63–66.
5   L. A. Ingerman, M. E. Cuellar and M. L. Waters, *Chem. Commun.*, 2010, **46**, 1839.
6   K. D. Daze, M. C. F. Ma, F. Pineux and F. Hof, *Org. Lett.*, 2012, **14**, 1512–1515.
7   M. A. Gamal-Eldin and D. H. Macartney, *Org. Biomol. Chem.*, 2013, **11**, 488–495.
8   N. K. Pinkin and M. L. Waters, *Org. Biomol. Chem.*, 2014, **12**, 7059–7067.
9   J. E. Beaver, B. C. Peacor, J. V. Bain, L. I. James and M. L. Waters, *Org. Biomol. Chem.*, 2015, **13**, 3220–3226.
10   H. Peacock, C. C. Thinnes, A. Kawamura and A. D. Hamilton, *Supramol. Chem.*, 2016, **28**, 575–581.
11   R. Pinalli, G. Brancatelli, A. Pedrini, D. Menozzi, D. Hernández, P. Ballester, S. Geremia and E. Dalcanale, *J. Am. Chem. Soc.*, 2016, **138**, 8569–8580.
12   J. E. Beaver and M. L. Waters, *ACS Chem. Biol.*, 2016, **11**, 643–653.
13   I. N. Gober and M. L. Waters, *Org. Biomol. Chem.*, 2017, **15**, 7789–7795.
14   T. Hanauer, R. J. Hopkinson, K. Patel, Y. Li, D. Correddu, A. Kawamura, V. Sarojini, I. K. H. Leung and T. Gruber, *Org. Biomol. Chem.*, 2017, **15**, 1100–1105.
15   F. Guagnini, P. M. Antonik, M. L. Rennie, P. O'Byrne, A. R. Khan, R. Pinalli, E. Dalcanale and P. B. Crowley, *Angew. Chemie Int. Ed.*, 2018, **57**, 7126–7130.
16   A. G. Mullins, N. K. Pinkin, J. A. Hardin and M. L. Waters, *Angew. Chemie Int. Ed.*, 2019, **58**, 5282–5285.
17   L. I. James, J. E. Beaver, N. W. Rice and M. L. Waters, *J. Am. Chem. Soc.*, 2013, **135**, 6450–6455.
18   M. Florea, S. Kudithipudi, A. Rei, M. J. González-Álvarez, A. Jeltsch and W. M. Nau, *Chem. - A Eur. J.*, 2012, **18**, 3521–3528.
19   S. A. Minaker, K. D. Daze, M. C. F. Ma and F. Hof, *J. Am. Chem. Soc.*, 2012, **134**, 11674–11680.
20   B. C. Peacor, C. M. Ramsay and M. L. Waters, *Chem. Sci.*, 2017, **8**, 1422–1428.
21   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.
22   R. S. Blanc and S. Richard, *Mol. Cell*, 2017, **65**, 8–24.
23   J. Jarrold and C. C. Davies, *Trends Mol. Med.*, 2019, **25**, 993–1009.
24   J. J. Lavigne and E. V. Anslyn, *Angew. Chemie Int. Ed.*, 2001, **40**, 3118–3130.
25   F. Cramer, *The Lock-and-Key Principle: The State of the Art 100 Years On*, John Wiley & Sons, Ltd, 1994.
26   L. You, D. Zha and E. V Anslyn, *Chem. Rev.*, 2015, **115**, 7840–7892.
27   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.
28   V. Saggiomo, Y. R. Hristova, R. F. Ludlow and S. Otto, *J. Syst. Chem.*, 2013, **4**, 2.
29   A. M. Valdivielso, F. Puig-Castellví, J. Atcher, J. Solà, R. Tauler and I. Alfonso, *Chem. - A Eur. J.*, 2017, **23**, 10789–10799.
30   A. Buryak and K. Severin, *Angew. Chemie Int. Ed.*, 2005, **44**, 7935–7938.
31   M. Lafuente, J. Solà and I. Alfonso, *Angew. Chemie Int. Ed.*, 2018, **57**, 8421–8424.
32   C. Bravin, A. Guidetti, G. Licini and C. Zonta, *Chem. Sci.*, 2019, **10**, 3523–3528.
33   F. Ulatowski and J. Jurczak, *Pure Appl. Chem.*, 2017, **89**, 801–807.
34   A. P. Davis, S. Kubik and A. Dalla Cort, *Org. Biomol. Chem.*, 2015, **13**, 2499–2500.
35   I. Alessandri, E. Biavardi, A. Gianoncelli, P. Bergese and E. Dalcanale, *ACS Appl. Mater. Interfaces*, 2016, **8**, 14944–14951.
36   S. Stewart, M. A. Ivy and E. V Anslyn, *Chem. Soc. Rev.*, 2014, **43**, 70–84.
37   P. T. Corbett, J. K. M. Sanders and S. Otto, *Chem. - A Eur. J.*, 2008, **14**, 2153–2166.
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   K. R. West, R. F. Ludlow, P. T. Corbett, P. Besenius, F. M. Mansfeld, P. A. G. Cormack, D. C. Sherrington, J. M. Goodman, M. C. A. Stuart and S. Otto, *J. Am. Chem. Soc.*, 2008, **130**, 10834–10835.