Development and Mechanistic Studies of an Optimized Receptor for Trimethyllysine using Iterative Redesign by Dynamic Combinatorial Chemistry

Journal:	*Organic & Biomolecular Chemistry*
Manuscript ID:	OB-ART-06-2014-001249.R1
Article Type:	Paper
Date Submitted by the Author:	21-Jul-2014

Complete List of Authors:	Pinkin, Nicholas; The University of North Carolina at Chapel Hill, Department of Chemistry Waters, Marcey; The University of North Carolina at Chapel Hill, Department of Chemistry

Organic and Biomolecular Chemistry RSCPublishing

ARTICLE

Development and Mechanistic Studies of an Optimized Receptor for Trimethyllysine using Iterative Redesign by Dynamic Combinatorial Chemistry

Cite this: DOI: 10.1039/x0xx00000x

Received ooth January 2012,
Accepted ooth January 2012

DOI: 10.1039/x0xx00000x

www.rsc.org/

Nicholas K. Pinkin and Marcey L. Waters[a]

A new small molecule receptor, **A_2N**, has been identified that binds specifically to trimethyllysine (Kme_3) with sub-micromolar affinity. This receptor was discovered through the iterative redesign of a monomer known to incorporate through dynamic combinatorial chemistry (DCC) into a previously reported receptor for Kme_3, **A_2B**. In place of monomer **B**, the newly designed monomer **N** introduces an additional cation-π interaction into the binding pocket, resulting in more favorable binding to Kme_3 by 1.3 kcal/mol, amounting to a 10-fold improvement in affinity and a 5-fold improvement in selectivity over Kme_2. This receptor exhibits the tightest affinity and greatest selectivity for KMe_3-containing peptides reported to date. Comparative studies of **A_2B** and **A_2N** provide mechanistic insight into the driving force for both the higher affinity and higher selectivity of **A_2N**, indicating that the binding of KMe_3 to **A_2N** is both enthalpically and entropically more favorable. This work demonstrates the ability of iterative redesign coupled with DCC to develop novel selective receptors with the necessary affinity and selectivity required for biological applications.

Introduction

Tools that enable the identification and characterization of post-translational modifications (PTMs) in histone proteins are essential for advancing the field of epigenetics. On the histone tail, the site and extent of lysine (Lys) methylation is linked to activation and repression of gene expression.[1,2] Moreover, dysregulation of Lys methylation has been associated with a number of types of cancers.[3–7] Methylation can occur up to three times, giving mono-, di- or trimethyllysine (Kme, Kme_2, or Kme_3, Figure 1). With increasing methylation, the size and hydrophobicity of the terminal ammonium group increases while the hydrogen-bonding capacity, and thus the cost of desolvation, decreases. While these changes are subtle, reader proteins are capable of site-specifically recognizing Kme, Kme_2, or Kme_3 and they facilitate downstream events in response to specific recognition events.[8]

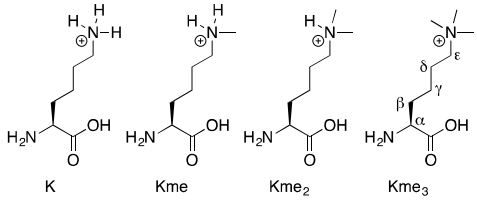


Figure 1. Methylation states of Lys that are found in proteins.

Antibodies are powerful and commonly employed tools for characterizing PTMs. Due to their high specificity and well understood structure, they enable individual PTMs to be tagged and enriched from complex samples, which is key to methods such as ELISA, western blotting, and microarray technology.[9–12] Unfortunately, the high specificity is both a strength and limitation; because site-specificity stems from interactions with residues neighboring a specific PTM, it is challenging to use antibodies to discover new sites of Lys methylation. Furthermore, studies of the histone code are often hampered by cross reactivity and interference with neighboring PTMs. [13–17]

Synthetic small molecule receptors are emerging as promising new tools for studying PTMs.[18–21] In the past decade, a variety of interesting receptors have been reported that bind specifically to Kme_3 through combinations of non-covalent interactions. Several of these receptors bind tighter and more selectively to Kme_3 than native reader proteins, despite being thousands of daltons smaller. [22–27]

We previously reported one such receptor, **A_2B** (Figure 2), which was found to bind to Kme_3 peptides in aqueous solution with comparable affinity and selectivity to the HP1 chromodomain, a reader protein known to bind histone 3 (H3) $K9me_{2/3}$.[22,28] **A_2B** was identified using dynamic combinatorial

chemistry (DCC), a competitive selection method that simplifies the discovery of synthetic receptors by relying on thermodynamic equilibria to select for hosts that interact most favorably with the guest of interest.[29] Because the monomers self-assemble into macrocyclic hosts, we proposed that we could improve the binding and selectivity of **A_2B** for Kme_3 by simply redesigning the constituent monomers. By focusing on monomer redesign, the chemistry required to make changes to a macrocycle's binding pocket is simplified to the modification of a small molecule. This reduces the challenges inherent in a *de novo* approach to macrocycle modification, namely protecting group optimization and targeted modification of single functional groups when multiple identical groups exist. DCC allows new monomers to be rapidly screened for their propensity to incorporate into selective hosts, as the composition of dynamic combinatorial libraries (DCLs) with varying guests is often indicative of host selectivity.

Using iterative monomer redesign to optimize the affinity and selectivity of the lead receptor, **A_2B**, herein we describe the development and characterization of a new receptor, **A_2N**, with 300 nM affinity for Kme_3-containing peptides. Moreover, A_2N exhibits ~10-fold tighter binding to Kme_3, 5-fold greater selectivity over Kme_2, and > 4-fold greater selectivity over unmethylated Lys relative to **A_2B**. The degree of affinity and selectivity of **A_2N** makes it a promising candidate to move forward with applications for sensing Kme_3. Moreover, analysis of the enthalpy and entropy of binding to each of the methylation states of Lys to these two receptors provides mechanistic insight into the factors providing affinity and selectivity.

Results and Discussion.

System Design.

A_2B binds preferentially to Kme_3 over the lower methylation states of Lys via cation-π interactions in a binding pocket made up of five aromatic rings.[22] However, the selectivity over Kme_2 is a modest 2-fold. Computational modelling of **A_2B** suggested that the binding cavity is shallow, which may be responsible for the low selectivity for Kme_3 over Kme_2 (Figure 2). We envisioned that a new monomer, **N**, if incorporated in place of monomer **B** into a similar receptor, **A_2N**, would provide a deeper pocket and additional CH(δ+)-π interactions with Kme_3. Furthermore, we anticipated that Lys guests would require greater desolvation to bind into the deeper binding pocket of **A_2N**, which we expected would improve selectivity for Kme_3.

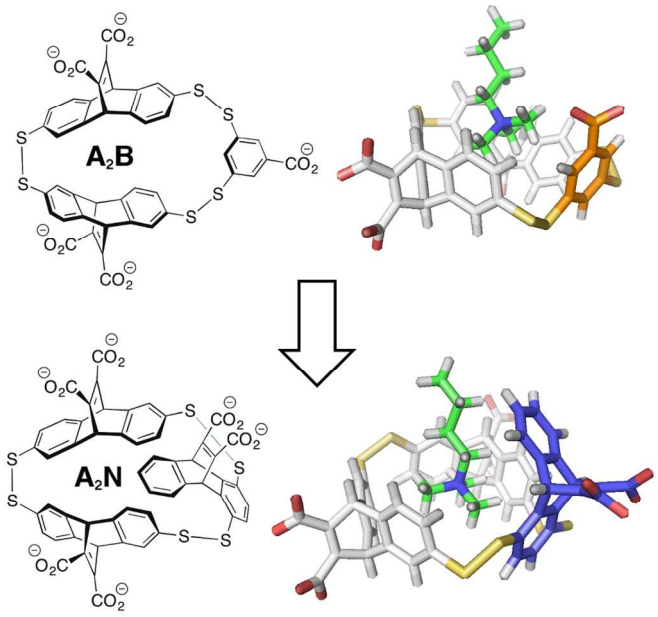
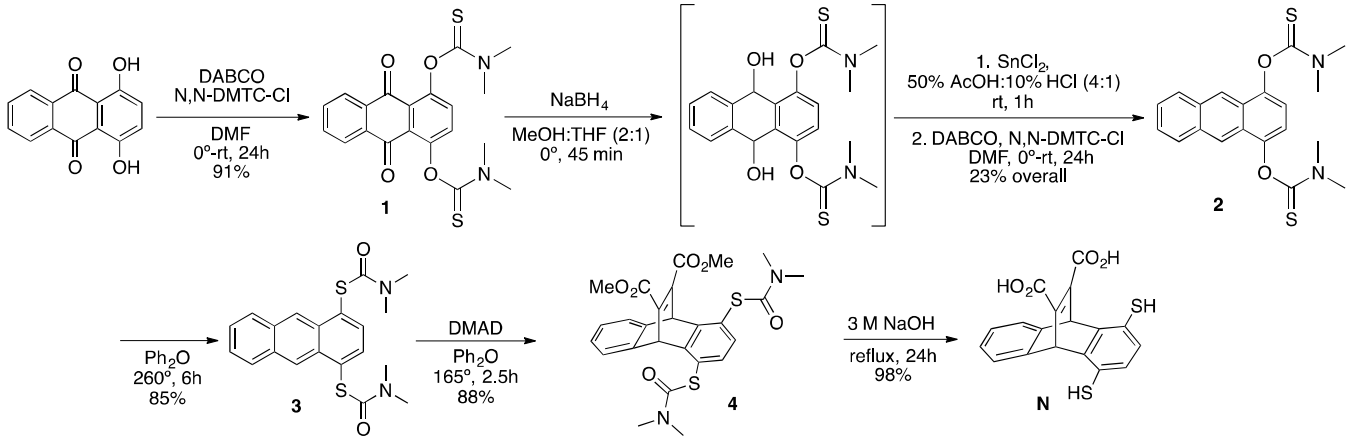


Figure 2. Monomer **B** of **A_2B** (top) was redesigned into monomer **N** (bottom) to deepen the binding pocket of **A_2B** and provide an extra cation-π interaction.

Synthesis.

Monomer **N** was synthesized using an approach similar to that reported by Otto and Sanders for synthesizing the isomeric monomer **A** (Scheme 1).[30] Initial efforts toward the dithiocarbamate anthracene **2** relied on previous reports of the synthesis and modification of 1,4-anthracenediol, but were unsuccessful due to rapid degradation of all intermediates.[31] Instead, we found that the thiocarbamate group can act as a protecting group for the reduction of the anthraquinone to the anthracene, allowing the protected anthracene **2** to be reached in acceptable yield over three steps. Compound **1** is first reduced to the intermediate diol using $NaBH_4$, then a reductive elimination using $SnCl_2$ in aqueous acid and a subsequent reprotection of the hydroxyl groups yields anthracene **2** (the intermediate diol rapidly degrades in the presence of air and light and is not isolated). The O-thiocarbamate anthracene **2** is subjected to a Newman-Kwart rearrangement to yield the S-thiocarbamate anthracene **3**, which subsequently undergoes a Diels Alder cycloaddition with dimethyl acetylene dicarboxylate (DMAD) to afford **4**. A final base-promoted hydrolysis gives monomer **N** cleanly and in high yield.



Scheme 1. Synthesis of monomer **N**.

Library Screening.

DCC was used to rapidly screen for novel receptors for Kme_3. Disulfide exchange was used as the reversible reaction because it occurs in aqueous solution at close to neutral pH and is stable toward most biological functional groups.[32] Dynamic combinatorial libraries (DCLs) were set up with 2.5 mM each of monomers A and N and guest concentration equal to the total combined monomer concentration (ie. 5 mM for a 2 monomer library) in 50 mM borate buffer, pH 8.5. Simple peptides with the sequence Ac-Kme_XGGL-NH_2 (X=0-3) were used as guests to limit non-specific interactions that could interfere with Lys recognition. Leu was incorporated to decrease the polarity of the peptides, which simplified their purification by reversed phase HPLC. For each combination of monomers, five DCLs were set up in parallel: four with one of the Lys guests and one untemplated library that lacked a guest. DCLs were monitored by LC-MS after three days, twelve days and three weeks. A species that was amplified in one library more than any other library was pursued as a potential selective receptor for the guest causing the amplification.

In DCLs containing only monomer **A**, $\mathbf{A_3}$ was amplified with increasing methylation on Lys (Figure S16). In DCLs containing only monomer **N**, no change in the library composition was observed in the presence of any of the guests (Figure S17). Instead, the monomer assembles into various forms of the homocyclic tetramer $\mathbf{N_4}$. When both monomers are combined in equal concentrations, there is significant amplification of three species in the presence of Kme_3 (Figure 3a).

Because **N** is an isomer of **A**, it was impossible to identify by mass the exact identity of the three amplified species, all of which were trimers. Nonetheless, a comparison to the DCLs of the individual monomers (analyzed by LC-MS using the same method) suggested that the new species must be heterotrimers of **A** and **N**, since their retention times were different from $\mathbf{A_3}$, and $\mathbf{N_3}$ was not amplified in the library of **N** (Figure S18). Once isolated, treatment of the receptors with TCEP resulted in reduction to a 2:1 mixture of monomers **A** and **N**, establishing that the three species are all isomers of $\mathbf{A_2N}$ (Figure S30).

Comparing the amplification of $\mathbf{A_2N}$ and $\mathbf{A_2B}$ (by peak area) in similar DCLs, $\mathbf{A_2N}$ is amplified 30-fold in the presence of Kme_3 over an untemplated library, while $\mathbf{A_2B}$ is only amplified about 10-fold. (Figure 3b) In the presence of the lower methylation states, similar amplification is observed for both receptors. This observation suggests that $\mathbf{A_2N}$ is a more selective receptor for Kme_3 than $\mathbf{A_2B}$.

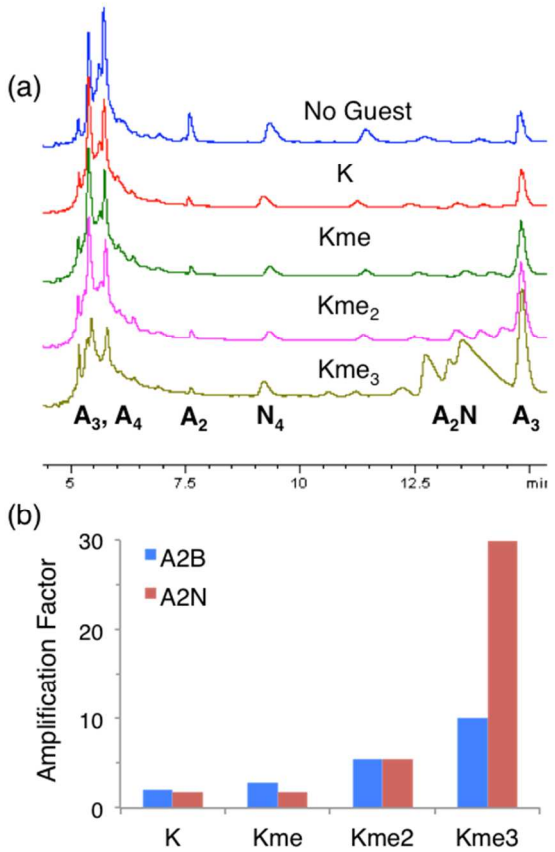


Figure 3. (a) DCLs of monomers **A** and **N** with various guests (2.5 mM A, 2.5 mM N, 5 mM Ac-Kme_XGGL-NH_2, 50 mM borate buffer, pH 8.5) (b) Calculated amplification of $\mathbf{A_2B}$ and $\mathbf{A_2N}$ by peak area in the presence of different guests compared to the untemplated DCL.

Structural Characterization.

The isomeric $\mathbf{A_2N}$ macrocycles were made on a preparative scale using DCC under similar conditions but with acetylcholine chloride (AcCh) as a template. AcCh was used instead of Kme_3 because it was found to similarly amplify $\mathbf{A_2N}$ when present in excess in a DCL, yet is commercially available. After five days, the isomers of $\mathbf{A_2N}$ were isolated by RP-HPLC. Under optimized conditions two isomers nearly co-elute (in 22% yield), but the third isomer is better resolved and is easily isolated in 23% yield (Figure S19).

Because monomer **A** is used in libraries as a racemic mixture, we expected that the three $\mathbf{A_2N}$ species must be two distinct *meso* isomers and a pair of *enantiomers*. Initial experiments revealed that at room temperature in methanol-*d4* or D_2O, the proton resonances of all three isomers of $\mathbf{A_2N}$ were significantly broadened, indicating that all isomers of $\mathbf{A_2N}$ are dynamic and that rotation is on the NMR timescale. The para-substitution of the thiols on **N** likely enables the monomer to rotate about an axis created by the C-S bonds, making the receptor quite flexible. While cooler temperatures only increased the broadening, mild heating sharpened the resonances significantly. In methanol-*d4*, less heating was required to sharpen peaks compared to in D_2O; therefore, all structural characterization of $\mathbf{A_2N}$ alone was performed in methanol-*d4* (Figure S22-S29).

A simple comparison of the ^{1}H NMR spectra allowed the two *meso* isomers to be assigned as the second and third species that elute during purification, as their ^{1}H spectra contained fewer peaks than that of the first species. Because the first *meso* isomer, *meso*$_1$-$\mathbf{A_2N}$, co-elutes with *rac*-$\mathbf{A_2N}$, a pure sample could not be obtained. However, the resonances of *meso*$_1$-$\mathbf{A_2N}$ were distinguishable in the mixed spectrum, which enabled further 2D NMR characterization. While *rac*-$\mathbf{A_2N}$ could be isolated with careful purification, there was significant peak overlap in the ^{1}H spectrum and further structural characterization was not pursued.

Proton assignments were made using the TOCSY and COSY spectra of the two *meso*- isomers. The ROESY spectrum of *meso*$_2$-$\mathbf{A_2N}$ revealed NOEs between protons 2, 3, & 4 on monomer **N** and protons 10, 11, & 12 on monomer **A**, confirming the orientation of **N** in *meso*$_2$-$\mathbf{A_2N}$ as shown in Figure 4. In contrast, no inter-monomer NOEs were observed in the ROESY spectrum of *meso*$_1$-$\mathbf{A_2N}$. This suggests that *meso*$_1$-$\mathbf{A_2N}$ contains a more open binding pocket than *meso*$_2$-$\mathbf{A_2N}$, which may help to explain the subsequent observation that *meso*$_2$-$\mathbf{A_2N}$ binds tighter and more selectively than *meso*$_1$-$\mathbf{A_2N}$ to Kme_3 (*vide infra*).

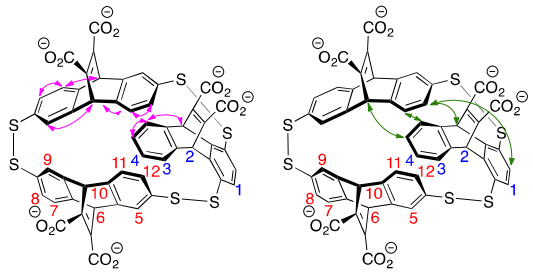


Figure 4. Intra-monomer (left) and Inter-monomer (right) NOEs observed for *meso*$_2$-$\mathbf{A_2N}$. Numbering and NOEs are identical for each half of the σ symmetric receptor.

To determine the mode of binding to Kme_3, an NMR analysis of the dipeptide Ac-Kme_3G-NH_2 in the presence of excess *meso*$_2$-$\mathbf{A_2N}$ was performed in D_2O under saturating conditions. Significant upfield shifting ranging from 0.6 to 3.5 ppm was observed for the β, γ, δ, ε, and methyl protons of Kme_3 (Table 1 and Figure S32), indicating close proximity of these positions to the face of the aromatic rings of the receptor. This is the largest degree of upfield shifting observed for any Kme_3 receptor reported to date. Compared to *rac*-$\mathbf{A_2B}$, *meso*$_2$-$\mathbf{A_2N}$ shifts the protons of Kme_3 within its binding pocket ~1 ppm further upfield. For both receptors, the ε protons exhibit the greatest degree of upfield shifting and the extent of upfield shifting of the other protons within the binding pocket decreases with increasing distance from the ε protons. In contrast, there is no significant upfield shifting of any other protons in the peptide, suggesting that the receptor interacts primarily with the sidechain of Kme_3.

Table 1. Change in Chemical Shift (Δδ) observed for Ac-Kme_3-Gly-NH_2 upon binding to an excess of *rac*-$\mathbf{A_2B}$ or *meso*$_2$-$\mathbf{A_2N}$.

Peptide Protons	*rac*-A_2B Δδ (ppm)	*meso*$_2$-A_2N Δδ (ppm)
$\mathbf{Nme_3}$	-1.59	-2.46
ε	-2.59	-3.45
δ	-2.11	-3.25
γ	-1.15	-2.09
β	-0.58	-0.60
α	-0.05	+0.08
Gly	+0.18	+0.25
Ac	+0.20	+0.33

Comparing the upfield shifting of the methylene protons between both receptors, the greatest difference in shift is observed for the γ and δ methylenes, which *meso*$_2$-$\mathbf{A_2N}$ shifts 0.94 and 1.14 ppm further upfield than does *rac*-$\mathbf{A_2B}$, respectively. Comparitively, the ε and Nme_3 protons are both shifted 0.86 ppm further upfield by *meso*$_2$-$\mathbf{A_2N}$. We expected that incorporation of monomer **N** into $\mathbf{A_2N}$ would result in a deeper binding pocket that is capable of participating in additional cation-π and CH(δ^+)-π interactions with the methyl groups and methylenes in the sidechain. This is evidenced by the ~1 ppm further upfield shifting of all β, γ, δ, ε, and methyl protons of Kme_3 bound to *meso*$_2$-$\mathbf{A_2N}$ compared to *rac*-$\mathbf{A_2B}$.

Binding Studies.

The previously reported K_d for *rac*-**A_2B** binding to Kme_3 in the context of the histone 3 (H3) peptide, FAM-QTAR-K9me_3-STG-NH_2 (where FAM is carboxyfluorescein), was determined using fluorescence anisotropy (FA) to be 25 μM.[22] To gain more mechanistic insight into the driving force for binding, we turned to isothermal titration calorimetry (ITC) to characterize binding of **A_2N** to Kme_3. To make direct comparisons between *rac*-**A_2B** and **A_2N**, we repeated measurements of the binding of *rac*-**A_2B** to the peptide H3 K9me_X (Table 2), which corresponds to residues 5-12 of histone 3, using ITC. The binding to this sequence was studied for comparison to previous data for *rac*-**A_2B**.[22] A WGGG sequence was added to the N-terminus of all peptides to enable concentration to be determined by UV. To verify that the receptors do not interact with the Trp tag, binding was measured to an H3 peptide whose basic Arg8 and Lys9 residues were mutated to Gly in order to eliminate any cation-π or charge-charge interactions with the receptors (Table 2 and Table 3, entries 10 and 18). These control experiments verified that the Trp tag does not bind to *rac*-**A_2B** or **A_2N**.

Table 2. Peptides used for ITC titrations.

H3 K9me$_X$ (X=0-3)	Ac-WGGG-QTARKme$_X$STG-NH_2
H3 R8G-K9me$_X$ (X=0-3)	Ac-WGGG-QTAGKme$_X$STG-NH_2
H3 K9G	Ac-WGGG-QTARGSTG-NH_2
H3 R8G-K9G	Ac-WGGG-QTAGGSTG-NH_2
H3 K36me$_X$ (X=0,3)	Ac-WGGG-TGGVKme$_X$KPH-NH_2

Interestingly, ITC measurements of the binding of *rac*-**A_2B** to H3 K9me_X gave affinities to all methylation states that are ~10-fold tighter than previously reported by FA (Table 3, entries 11-14), with a K_d of 2.6 μM for H3 K9me_3, although the selectivity for different methylation states is similar. We attribute this difference in affinity to a systematic error in the determination of receptor concentration due to incomplete desalting in the FA experiments that influenced the reported K_d. Since analysis of ITC data is not dependent on an absolute concentration of host for determination of K_a, ΔH, and ΔS, rather the relative ratio of host concentration to K_a and the absolute guest concentration,[33,34] we believe the ITC data is a more accurate measure of the binding affinity.

COMPARISON OF H3 K9ME$_3$ BINDING TO A_2B VERSUS A_2N. The binding of all three isomers of **A_2N** to H3 K9me_3 was studied (See SI, Table S2), but after determining that *meso$_2$*-**A_2N** binds the tightest and most selectively to an H3 K9me_3 peptide (Table 2), further studies focused on this isomer. *Meso$_2$*-**A_2N** was found to bind H3 K9me_3 with 300 nM affinity, as compared to the 2.6 μM affinity of *rac*-**A_2B** (Table 3, entries 1 and 11). This amounts to a 1.3 kcal/mol difference in affinity arising from the introduction of an additional aromatic ring. This value is consistent with previous measurements in cyclophanes and β-hairpins, which showed that the cation-π interaction with quaternary ammonium ions can contribute ~0.5-1.1 kcal/mol per aromatic ring to the binding of cationic guests.[35–38]

Inspection of ΔH and ΔS indicate that the difference in affinity is due to small improvements in *both* the enthalpy and entropy of binding for *meso$_2$*-**A_2N** relative to *rac*-**A_2B**. This goes against the typical trend of enthalpy-entropy compensation.[39,40] The more favorable enthalpy of *meso$_2$*-**A_2N** binding is most easily explained by greater van der Waals and cation-π interactions with the Kme_3 sidechain. The more favorable entropy observed for *meso$_2$*-**A_2N** may be due to a greater contribution of the classical hydrophobic effect[41] due to the larger surface area of the receptor cavity, as well as a larger number of favorable binding orientations.

SELECTIVITIES OF A_2N AND A_2B FOR DIFFERENT METHYLATION STATES OF LYS. *Meso$_2$*-**A_2N** exhibits markedly improved selectivity for Kme_3 over all other methylation states of Lys relative to *rac*-**A_2B.** *Meso$_2$*-**A_2N** binds to H3 K9me_3 with 14-, 130-, and 35-fold selectivity over H3 K9me_2 H3 K9me, and H3 K9, respectively (Table 3, entries 1-4). In contrast, *rac*-**A_2B** was found to bind the same H3 K9me_3 peptide with only 2.4-, 5.4-, and 8.3-fold selectivity over H3 K9me_2 H3 K9me, and H3 K9, respectively (Table 3, entries 11-14). Thus, the deeper aromatic pocket in **A_2N** results in a significant improvement in selectivity.

Comparison of ΔH and ΔS for binding of the H3 Kme_{1-3} peptides provides some insight into the observed selectivity. The driving force for *meso$_2$*-**A_2N** binding methylated Lys in the H3 K9 series is a favorable enthalpic term that is fairly constant for the three guests, Kme_{1-3}. The selectivity for Kme_3 arises primarily from a decrease in the entropic penalty of binding with increasing methylation on Lys. A similar trend has been seen with a beta-hairpin system that investigated the role of Lys methylation on cation-π interactions.[38,42] There are several factors that may contribute to this entropic effect. The peptide-receptor complex may have a larger number of favorable binding conformations for Kme_3 than for Kme_2 and Kme. Additionally, greater methylation would be expected to result in a larger contribution of the classical hydrophobic effect to binding. Lastly, it may reflect different degrees of ordered water molecules within the pocket upon binding different methylation states, since Kme and Kme_2 can form hydrogen bonds, unlike Kme_3.[43]

Binding of *meso$_2$*-**A_2N** to H3 K9 does not follow the same trend. *Meso$_2$*-**A_2N** exhibits a tighter affinity for H3 K9 than H3 K9me (Table 3, entries 3 and 4). The binding of H3 K9 is considerably less exothermic than binding to the methylated residues, thus its tighter affinity over H3 K9me can be attributed to more favorable entropy of binding (compare entries 3 and 4). This suggests a change in mechanism of binding, such as H3 K9 binding to the exterior of the receptor via electrostatic interactions between the carboxylates and both R8 and K9. The favorable entropy of binding is consistent with both the fact that there are multiple possible orientations for binding and that electrostatic interactions with both ammonium and guanidinium groups have been shown to be entropically favorable in other systems.[44,45] The role of R8 is explored further below.

Binding of *rac*-**A_2B** to methylated Lys in the H3 K9me$_X$ (X = 0-3) peptides is also driven by a release of heat that overcomes the entropic cost of binding (Table 3, entries 11-14). However, in contrast to *meso$_2$*-**A_2N**, the selectivity for H3 K9me$_3$ is not purely entropy driven; instead it arises from a combination of enthalpic and entropic effects. With increasing methylation up to Kme$_2$, the binding to *rac*-**A_2B** becomes more exothermic, but more entropically disfavored, thus displaying typical enthalphy-entropy compensation.[46,47]

Table 3. Thermodynamic data obtained for the binding of *rac*-**A_2B** and *meso$_2$*-**A_2N** to the peptides shown in Chart 1 as measured by ITC.

Entry	Receptor	Peptide	Charge	K_d [b] (uM)	Selectivity factor [c]	ΔG [b] (kcal/mol)	ΔH [b] (kcal/mol)	$T\Delta S$ [b] (kcal/mol)
1	A_2N	H3 K9me$_3$	+2	0.30 ± 0.04	-	-8.91 ± 0.07	-12.0 ± 0.5	-3.1 ± 0.5
2	A_2N	H3 K9me$_2$	+2	4.1 ± 0.5	14	-7.36 ± 0.04	-12.5 ± 0.4	-5.1 ± 0.4
3	A_2N	H3 K9me	+2	40 ± 4	130	-6.01 ± 0.06	-12.0 ± 0.5	-6.0 ± 0.5
4	A_2N	H3 K9	+2	10.5 ± 0.9	35	-6.80 ± 0.05	-7.3 ± 0.3	-0.5 ± 0.3
5	A_2N	H3 R8G-K9me$_3$	+1	1.3 ± 0.2	-	-8.05 ± 0.08	-13.4 ± 0.5	-5.3 ± 0.6
6	A_2N	H3 R8G-K9me$_2$	+1	35 ± 1	28	-6.1 ± 0.4	-	-
7	A_2N	H3 R8G-K9me	+1	~150 [d,e]	120 [d,e]	~ -5.2	-	-
8	A_2N	H3 R8G-K9	+1	~360 [d,e]	280 [d,e]	~ -4.7	-	-
9	A_2N	H3 R8-K9G	+1	~300 [d,e]	-	~ -4.8	-	-
10	A_2N	H3 R8G-K9G	0	NB [e]	-	-	-	-
11	A_2B	H3 K9me$_3$	+2	2.6 ± 0.1	-	-7.63 ± 0.03	-11.26 ± 0.05	-3.61 ± 0.05
12	A_2B	H3 K9me$_2$	+2	6.3 ± 0.3	2.4	-7.10 ± 0.07	-11.65 ± 0.09	-4.5 ± 0.1
13	A_2B	H3 K9me	+2	13.9 ± 0.1	5.4	-6.64 ± 0.01	-9.65 ± 0.06	-3.00 ± 0.07
14	A_2B	H3 K9	+2	22 ± 1	8.3	-6.38 ± 0.02	-9.2 ± 0.2	-2.9 ± 0.3
15	A_2B	H3 R8G-K9me$_3$	+1	17.1 ± 0.1	-	-6.52 ± 0.01	-12.37 ± 0.01	-5.84 ± 0.02
16	A_2B	H3 R8G-K9	+1	~140 [d,e]	8.2 [d,e]	~ -5.3	-	-
17	A_2B	H3 R8-K9G	+1	~150 [d,e]	-	~ -5.2	-	-
18	A_2B	H3 R8G-K9G	0	NB [e]	-	-	-	-
19	A_2N	H3 K36me$_3$	+2	0.3 ± 0.1	-	-8.9 ± 0.2	-	-
20	A_2N	H3 K36	+2	~ 70 [d,e]	200 [d,e]	~ -5.7	-	-

[a] Conditions: 26 °C in 10 mM borate buffer, pH 8.5. [b] Errors are from averages. [c] Selectivity is calculated as the factor-fold difference in affinity for Kme$_3$ over the designated methylation state in that row. [d] These values are approximate because the c-value for these experiments was <1. [e] For these experiments, the N-value was fixed at 1 for one-site fitting.

Investigation of electrostatic contributions in the H3 K9 peptide. Intrigued by the peculiar tighter binding of *meso$_2$*-**A_2N** to H3 K9 over H3 K9me, we mutated the neighboring Arg8 to Gly to see what impact the nearby charge has on binding to Kme$_x$ (Table 3, entries 5-8). Upon mutation, we observed ~4-fold weaker binding to H3 R8G-K9me$_3$ as compared to the unmutated H3 K9me$_3$ peptide, amounting to a loss of about 0.9 kcal/mol. As the affinities decreased in the series, we were unable to achieve c-values greater than the accepted minimum of 1, thus the K_d values reported in these situations are approximate and a thermodynamic analysis is not made.[33,48] Nonetheless, the selectivity for H3 K9me$_3$ over H3 K9me$_2$ and H3 K9me is relatively unaffected by the R8G mutation (compare entries 1-3 to entries 5-7). In contrast, mutation of R8 has an immense effect on binding to the unmethylated K9, with a decrease in binding affinity of more than 30-fold (> 2 kcal/mol). This results in a much improved selectivity for Kme$_3$ over K of >250-fold in this mutant series. Comparing H3 K9 to H3 R8G-K9 (entries 4 and 8), the difference in binding affinity amounts to at least 2 kcal/mol, compared to about 1 kcal/mol for H3 K9me$_3$ versus H3 R8G-K9me$_3$ (entries 1 and 5). Thus, R8 contributes more to binding of the unmethylated Lys than to any of the H3 Kme$_{1-3}$ peptides. It is important to note, however, that mutation of K9, giving H3 R8-K9G (entry 9) results in similar weak binding observed for H3 R8G-K9 (entry 8), indicating that Arg is not a significant binder on its own. Furthermore, methylation of Arg8 to any of the three methylated states found on histone tails (Rme, sRme$_2$, and aRme$_2$) leads to weaker binding compared to unmethylated Arg, presumably due to weakening of the unique interaction of *meso$_2$*-**A_2N** with unmethylated Arg8 and Lys9 (Table S1).

Taken together, these results suggest that the presence of R8 results in a different binding mechanism of H3 K9 to *meso$_2$*-**A_2N** that is much less entropically costly than those with methylated K9. Because we observe similar selectivities for Kme$_3$ over Kme$_2$ and Kme despite the R8G mutation, the 250-fold selectivity over K in the R8G series of peptides more accurately represents the selectivity of *meso$_2$*-**A_2N** in the absence of other neighboring interactions.

The role of R8 was also investigated in the binding of *rac*-**A_2B** to K9me$_3$ and K9. In this case, mutation of R8 has about the same effect on binding to K9me$_3$ or K9: loss of R8 results in a 1.1 kcal/mol decrease in binding, regardless of the methylation state of Lys (compare entries 11 and 15 to entries 14 and 16).

Comparison of the enthalpy and entropy of binding of H3 K9me$_3$ and H3 R8G-K9me$_3$ with **A_2N** or **A_2B** provides additional insights into the role of Arg in the presence of methylated Lys. With both receptors, mutation of Arg8 to Gly results in a *more favorable enthalpy* of binding by 1.1-1.4 kcal/mol and a *less favorable entropy* of binding by 2.2 kcal/mol (compare entries 1 to 5 and 11 to 15). Thus, the contribution of Arg to binding is entropic, not enthalpic. This may suggest additional contributions, such as Arg stacking with the aromatic rings on the exterior of the receptor, which may release water molecules and strengthen the electrostatic

interaction with the carboxylates on each monomer (Figure 5). Evidence for this mode of binding comes from several model systems.[44,45,49] Further investigation into the mechanism of this interaction is underway.

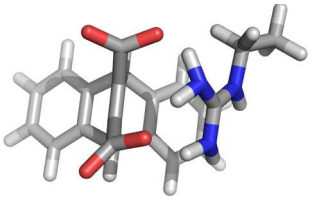
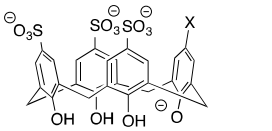


Figure 5. Computational model of the interaction of a guanidinium group with both the carboxylates and aromatic ring of the exterior of an **A** or **N** monomer.

Comparison of H3 K9me$_3$ and H3 K36me$_3$. To better understand the impact of the surrounding sequence on the recognition of Kme$_3$ over unmodified Lys by **A$_2$N**, we measured the binding of *meso$_2$*-**A$_2$N** to trimethylated and unmethylated H3 K36 peptides, which have the same net charge but contain a neighboring Lys in place of Arg (see Table 2 and Table 3, entries 19-22). The binding affinity of *meso$_2$*-**A$_2$N** to H3 K36me$_3$ was identical to that of H3 K9me$_3$, validating that, with the exception of basic residues, the surrounding sequence does not have a significant impact on affinity. Interestingly, however, the selectivity for H3 K36me$_3$ over H3 K36 is very similar to that observed for the mutated H3 R8G-K9 peptide series, which has a +1 charge (compare entries 19 and 20 to entries 5 and 8). This supports the fact that **A$_2$N** recognizes Kme$_3$ with approximately 250-fold selectivity over K and suggests that a neighboring Arg can interact with the receptor in a unique manner.

Comparison to Other Synthetic Receptors for Kme$_3$. Several synthetic receptors that bind Kme$_3$ either as a single amino acid[23,26] or within the context of a histone tail peptide[18,21] have been reported to date. Because the zwitterionic nature of the amino acid influences binding in ways that are not relevant to recognition of PTMs in proteins, only comparison to receptors that bind Kme$_3$ in the context of peptides is made here (Table 4). It is clear that all receptors reported to date are influenced by the net charge of the peptide, such that significantly tighter binding can be achieved with more basic peptides (compare Table 4, entries 3 and 4, for example). A careful analysis of the effect of these nonspecific electrostatic interactions on selectivity over the unmodified peptide has not been fully investigated for any systems. Nonetheless, comparing binding to peptides of the same net charge, *meso$_2$*-**A$_2$N** demonstrates the tightest binding affinity and highest selectivity over the unmethylated state reported to date (Table 4, entries 1, 2, and 3). Interestingly, the extra aromatic ring in CX$_4$ArCO$_2^-$ relative to CX$_4$ does not provide any additional affinity (Table 4, entries 4 and 5), unlike the additional aromatic ring in **A$_2$N** relative to **A$_2$B** (Table 4, entries 1 and 2). The rigid nature of the rings in **N** as well as the methine linkers between the rings (versus the sulfonamide linker in CX$_4$ArCO$_2^-$ and CX$_4$ArBr) may be important in providing additional binding affinity.



CX$_4$: X = SO_3^-
CX$_4$ArCO$_2^-$: X = NH-SO$_2$-C$_6$H$_4$-CO_2^-
CX$_4$ArBr: X = NH$_2$-SO$_2$-C$_6$H$_4$-Br

Figure 6. Structure of other reported hosts for KMe$_3$ in the context of peptides.

Conclusions.

In summary, we have used iterative design coupled with DCC to optimize a receptor for recognition of Kme$_3$, resulting in a 300 nM binder for H3 K9me$_3$ with 10-fold improvement in binding affinity and a 5-fold improvement in selectivity over Kme$_2$. Further, *meso$_2$*-**A$_2$N** is the tightest and most selective receptor for Kme$_3$ in the context of a peptide reported to date. NMR data indicate that the Kme$_3$ sidechain binds inside the aromatic pocket, while the peptide backbone is not involved in binding. The improved selectivity over the original receptor arises from both more favorable enthalpy and entropy of binding, while the improved selectivity over the lower methylation states of Lys arise from more favorable entropy. This work demonstrates the utility of DCC coupled with iterative design for generating new receptors with affinity and selectivity necessary for biological applications and provides new insights into the driving force for achieving both affinity and selectivity for this class of modified amino acids in aqueous solution. Molecular recognition in water is an ongoing challenge in supramolecular chemistry, but this work demonstrates the strength of iterative redesign coupled with DCC for meeting this challenge.[50]

Table 4. Comparison of binding affinities and selectivities of synthetic receptors for Kme$_3$ peptides.

Entry	Host	Histone 3 Peptide	Peptide Charge	K_d (μM)	Selectivity (Kme$_3$/K)	Reference
1	*rac*-**A$_2$B**	Ac-WGGG-QTARKme$_3$STG-NH$_2$	+2	2.6[a]	8	This work
2	*meso*2-**A$_2$N**	Ac-WGGG-QTARKme$_3$STG-NH$_2$	+2	0.3[a]	35	This work
3	CX$_4$	Ac-TARKme$_3$STGY-NH$_2$	+2	7.2[b]	14	Ref 18
4	CX$_4$	H-ARTKQTARKme$_3$STGY-NH$_2$	+5	0.17[c]	NR	Ref 19
5	CX$_4$ArCO$_2^-$	H-ARTKQTARKme$_3$STGY-NH$_2$	+5	0.19[c]	NR	Ref 19
6	CX$_4$ArBr	H-ARTKQTARKme$_3$STGY-NH$_2$	+5	4.8[c]	NR	Ref 19

[a] 10 mM borate buffer, pH 8.5, 26 °C; [b] 40 mM phosphate buffer, pH 7.4, 30 °C; [c] 10 mM phosphate buffer, pH 7.4, 25 °C; [d] NR = not reported

Organic and Biomolecular Chemistry RSCPublishing

ARTICLE

Experimental.

A detailed synthetic procedure for the preparation of monomer **N** can be found in the SI. All dynamic combinatorial libraries were prepared by dissolving monomers at 2.5 mM and peptide guests at 2.5-5 mM (equal to total monomer concentration) in 50 mM sodium borate buffer (pH 8.5). DCLs were monitored at various time points on an Agilent Rapid Resolution LC-MS system equipped with an online degasser, binary pump, autosampler, heated column compartment, and diode array detector. All separations were performed using gradients between water (A) and acetonitrile (B) containing 5 mM NH_4OAc at pH 5 on a Zorbax Extend C18 (2.1 X 50 mm, 1.8 μm) column. Using 1 μL injections, libraries were monitored at various time points with the following gradient: 3-27 %B from 0-2 min, 27-29.2 %B from 2-12 min, then 100 %B for 5 min. The MS was performed using a single quad mass spectrometer. Mass spectra (ESI-) were acquired in ultrascan mode by using a drying temperature of 350 °C, a nebulizer pressure of 45 psi, a drying gas flow of 10 L/min, and a capillary voltage of 3000 V. Data analysis was performed using the software Agilent ChemStation.

A_2N was prepared on a preparative scale using acetylcholine chloride (AcCh) as a guest. The preparative DCLs were prepared at 2 mM **A** and **N** and 10 mM AcCh in 50 mM sodium borate buffer (pH 8.5) and were allowed to equilibrate for five days before purification on an Atlantis PrepT3 5 μm 10x100mm C18 column in NH_4OAc buffered solvents. Extended lyopholization removed the majority of NH_4OAc salts for an accurate determination of the extinction coefficient.

All 1D and 2D NMR experiments were performed using a Bruker 400 MHz or Bruker 600 MHz instrument, as noted. Data analysis was performed using Topspin 3.1 software. VT 1D NMRs were collected on a Bruker 500 MHz instrument using cooled or heated nitrogen gas to control the temperature of the sample. NMR binding experiments were performed on a Bruker 600 MHz instrument at 25 °C in 10 mM borate buffererd D_2O (pH 8.67). Concentration determination of Ac-Kme_3G-NH_2 was performed using DSS as an internal standard. A 600 μM peptide stock solution was used to dissolve lyophilized *meso*$_2$-**A_2N** or *rac*-**A_2B**. 1D spectra were collected with 128 scans. Proton assignments were made using TOCSY analysis.

All ITC titrations were performed using a MicroCal Auto-iTC200 at 26 °C. Data analysis was performed using the built in Origin 7 software using a one site binding model. Unless otherwise noted, titrations were performed in triplicate. A 10 mM pH 8.5 sodium borate buffer was used for all experiments. All concentrations were determined using a NanoDrop2000 with a xenon flash lamp, 2048 element linear silicon CCD array detector, and 1 mm path length. Peptides were desalted using Thermo Scientific polyacrylamide 1800 MWCO desalting columns. ~0.5-3 mM solutions of peptide were titrated into ~20-200 μM solutions of A_2N (all isomers were studied, see SI for data on the *rac*- and *meso*$_1$- isomers) or *rac*-A_2B using 2 μL injections every 3 minutes. Heats of dilution of peptides were subtracted prior to analysis in Origin.

Acknowledgements

We gratefully acknowledge funding from the W. M. Keck foundation for this work. This material is based in part upon work supported by the National Science Foundation under Grant No. CHE-1306977 and also the National Science Foundation Graduate Research Fellowship to N.K.P. under grant no. DGE-1144081.

Notes and References

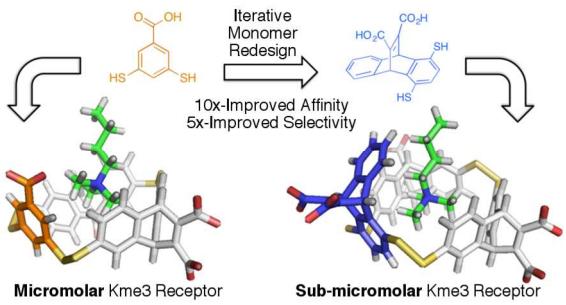
[a] Department of Chemistry, CB 3290, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599. E-mail: mlwaters@unc.edu

Electronic Supplementary Information (ESI) available: Synthesis and characterization of monomer N; NMR characterization of receptors; ITC binding data. See DOI: 10.1039/b000000x/

1. C. Martin and Y. Zhang, *Nat. Rev. Mol. Cell Biol.*, 2005, **6**, 838–849.
2. T. Kouzarides, *Cell*, 2007, **128**, 693–705.
3. S. B. Hake, A. Xiao, and C. D. Allis, *Br. J. Cancer*, 2004, **90**, 761–769.
4. G. G. Wang, C. D. Allis, and P. Chi, *Trends Mol. Med.*, 2007, **13**, 363–372.
5. H.-C. Tsai and S. B. Baylin, *Cell Res.*, 2011, **21**, 502–517.
6. M. Rodríguez-Paredes and M. Esteller, *Nat. Med.*, 2011, **17**, 330–339.
7. W. Timp and A. Feinberg, *Nat. Rev. Cancer*, 2013, **13**, 497–510.
8. S. D. Taverna, H. Li, A. J. Ruthenburg, C. D. Allis, and D. J. Patel, *Nat. Struct. Mol. Biol.*, 2007, **14**, 1025–1039.
9. D. E. Schones and K. Zhao, *Nat. Rev. Genet.*, 2008, **9**, 179–191.
10. Y. Zhao and O. N. Jensen, *Proteomics*, 2009, **9**, 4632–4641.
11. S. B. Rothbart, K. Krajewski, B. D. Strahl, and S. M. Fuchs, *Methods Enzymol.*, 2012, 107–135.
12. Z. Su, M. D. Boersma, J.-H. Lee, S. S. Oliver, S. Liu, B. A. Garcia, and J. M. Denu, *Epigenetics Chromatin*, 2014, **7**, 7.
13. S. Thorpe, *Nature*, 2007, **447**, 741–744.
14. I. Bock, A. Dhayalan, S. Kudithipudi, O. Brandt, P. Rathert, and A. Jeltsch, *Epigenetics*, 2011, **6**, 256–263.

15. T. A. Egelhofer, A. Minoda, S. Klugman, K. Lee, P. Kolasinska-Zwierz, A. A. Alekseyenko, M.-S. Cheung, D. S. Day, S. Gadel, A. A. Gorchakov, T. Gu, P. V Kharchenko, S. Kuan, I. Latorre, D. Linder-Basso, Y. Luu, Q. Ngo, M. Perry, A. Rechtsteiner, N. C. Riddle, Y. B. Schwartz, G. A. Shanower, A. Vielle, J. Ahringer, S. C. R. Elgin, M. I. Kuroda, V. Pirrotta, B. Ren, S. Strome, P. J. Park, G. H. Karpen, R. D. Hawkins, and J. D. Lieb, *Nat. Struct. Mol. Biol.*, 2011, **18**, 91–93.
16. S. M. Fuchs, K. Krajewski, R. W. Baker, V. L. Miller, and D. Brian, *Curr. Biol.*, 2012, **21**, 53–58.
17. S. Nishikori, T. Hattori, S. M. Fuchs, N. Yasui, J. Wojcik, A. Koide, B. D. Strahl, and S. Koide, *J. Mol. Biol.*, 2012, **424**, 391–399.
18. 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.
19. M. Florea, S. Kudithipudi, A. Rei, M. J. Gonzalez-Alvarez, A. Jeltsch, and W. M. Nau, *Chem. A Eur. J.*, 2012, 3521–3528.
20. S. A. Minaker, K. D. Daze, M. C. F. Ma, and F. Hof, *J. Am. Chem. Soc.*, 2012, 11674–11680.
21. H. F. Allen, K. D. Daze, T. Shimbo, A. Lai, C. A. Musselman, J. K. Sims, P. A. Wade, F. Hof, and T. G. Kutateladze, *Biochem. J.*, 2014, **459**, 505–512.
22. L. A. Ingerman, M. E. Cuellar, and M. L. Waters, *Chem. Commun.*, 2010, **46**, 1839–1841.
23. K. D. Daze, M. C. F. Ma, F. Pineux, and F. Hof, *Org. Lett.*, 2012, **14**, 1512–1515.
24. A. L. Whiting and F. Hof, *Org. Biomol. Chem.*, 2012, **10**, 6885–6892.
25. K. D. Daze and F. Hof, *Acc. Chem. Res.*, 2013, **46**, 937–945.
26. M. A. Gamal-Eldin and D. H. Macartney, *Org. Biomol. Chem.*, 2013, **11**, 488–495.
27. L. I. James, J. E. Beaver, N. W. Rice, and M. L. Waters, *J. Am. Chem. Soc.*, 2013, **135**, 6450–6455.
28. S. A. Jacobs and S. Khorasanizadeh, *Science*, 2002, **295**, 2080–2083.
29. 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.
30. P. T. Corbett, J. K. M. Sanders, and S. Otto, *Chem. A Eur. J.*, 2008, **14**, 2153–2166.
31. D. H. Hua, M. Tamura, X. Huang, H. A. Stephany, B. A. Helfrich, E. M. Perchellet, B. J. Sperfslage, J.-P. Perchellet, S. Jiang, D. E. Kyle, and P. K. Chiang, *J. Org. Chem.*, 2002, **67**, 2907–2912.
32. S. Otto, R. L. E. Furlan, and J. K. M. Sanders, *Science*, 2002, **297**, 590–593.
33. T. Wiseman, S. Williston, J. F. Brandts, and L. N. Lin, *Anal. Biochem.*, 1989, **179**, 131–137.
34. J. Tellinghuisen, *J. Phys. Chem. B*, 2005, **109**, 20027–20035.
35. H. Schneider, T. Matter, and S. Simova, *J. Chem. Soc. Chem. Commun.*, 1989, 580–581.
36. 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.
37. C. D. Tatko and M. L. Waters, *Protein Sci.*, 2003, 2443–2452.
38. R. M. Hughes and M. L. Waters, *J. Am. Chem. Soc.*, 2005, **127**, 6518–6519.
39. S. L. Hauser, E. W. Johanson, H. P. Green, and P. J. Smith, *Org. Lett.*, 2000, **2**, 3575–8.
40. M. V Rekharsky, T. Mori, C. Yang, Y. H. Ko, N. Selvapalam, H. Kim, D. Sobransingh, A. E. Kaifer, S. Liu, L. Isaacs, W. Chen, S. Moghaddam, M. K. Gilson, K. Kim, and Y. Inoue, *Proc. Natl. Acad. Sci. U. S. A.*, 2007, **104**, 20737–20742.
41. C. Tanford, *Science*, 1978, **200**, 1012–1018.
42. R. M. Hughes, M. L. Benshoff, and M. L. Waters, *Chem. A Eur. J.*, 2007, **13**, 5753–5764.
43. F. Biedermann, V. D. Uzunova, O. A. Scherman, W. M. Nau, and A. De Simone, *J. Am. Chem. Soc.*, 2012, **134**, 15318–15323.
44. S. E. Thompson and D. B. Smithrud, *J. Am. Chem. Soc.*, 2002, **124**, 442–449.
45. S. L. Tobey and E. V Anslyn, *J. Am. Chem. Soc.*, 2003, **125**, 14807–14815.
46. E. Grunwald and C. Steel, *J. Am. Chem. Soc.*, 1995, 5687–5692.
47. M. V. Rekharsky and Y. Inoue, *Chem. Rev.*, 1998, **98**, 1875–1918.
48. W. Turnbull and A. Daranas, *J. Am. Chem. Soc.*, 2003, 14859–14866.
49. X. Wang, J. Post, D. K. Hore, and F. Hof, *J. Org. Chem.*, 2014, **79**, 34–40.
50. G. V Oshovsky, D. N. Reinhoudt, and W. Verboom, *Angew. Chem. Int. Ed. Engl.*, 2007, **46**, 2366–2693.

TOC Graphic:



Iterative monomer redesign leads to a Kme_3-peptide receptor with 10-fold tighter affinity and 5-fold improved selectivity over Kme_2 than the original receptor. Thermodynamic analysis provides insight into this improvement.