A EUROPEAN JOURNAL OF CHEMICAL BIOLOGY

CHEMBIOCHEM

SYNTHETIC BIOLOGY & BIO-NANOTECHNOLOGY

Accepted Article

Title: DNA-binding properties of new fluorescent AzaHx-amides: Methoxy-pyridyl-aza-benzimidazole-pyrrole-imidazole/pyrrole

Authors: Beibei Liu, Luke Pett, Konstantinos Kiakos, Pravin C. Patil, Vijay Satam, John A. Hartley, Moses Lee, and W. David Wilson

To be cited as: *ChemBioChem* 10.1002/cbic.201800273

Link to VoR: http://dx.doi.org/10.1002/cbic.201800273

A Journal of

ChemPubSoc Europe

WILEY-VCH

www.chembiochem.org

DNA-binding properties of new fluorescent AzaHx-amides: Methoxy-pyridyl-aza-benzimidazole-pyrrole-imidazole/pyrrole

Beibei Liu,[a] Luke Pett,[b] Konstantinos Kiakos,[b] Pravin C. Patil,[c] Vijay Satam,[c] John A. Hartley,[b] Moses Lee,*[a,c,d] and W. David Wilson*[a]

Abbreviations

P	pyrrole
I	imidazole
f	formamide
Hx	2-(*p*-anisyl)-benzimidazole
AzaHx	2-*p*-(anisyl)-4-aza-benzimidazole
Pyr-AzaHx	pyridyl-aza-benzimidazole

Abstract: DNA minor groove binding polyamides have been extensively developed to control abnormal gene expression. Establishment of the novel, inherently fluorescent Hx-amides has provided an alternative path for studying DNA binding in cells by direct observation of cell localization. Because of the 2:1 antiparallel stacking homodimer binding mode of these molecules to DNA, modification of Hx-amides to AzaHx-amides has successfully extended the DNA recognition repertoire, from central CG (recognized by Hx-I) to central GC (recognized by AzaHx-P) recognition. To potentially target two consecutive GG bases, new modifications from AzaHx moiety to 3-Pyr-AzaHx and 2-Pyr-AzaHx moieties were developed. The newly designed molecules are also small-sized, fluorescent amides with Pyr-AzaHx connected to two conventional five-membered heterocycles. Complementary biophysical methods were carried out to investigate DNA binding properties of these molecules. The results showed that neither 3-Pyr-AzaHx nor 2-Pyr-AzaHx was able to mimic I-I to specifically target GG dinucleotides. Rather, 3-Pyr-AzaHx functions like AzaHx or f-I or P-I as a antiparallel stacked dimer. 3-Pyr-AzaHx-PI (**2**) binds 5'-ACGCGT'-3' with improved binding affinity and high sequence specificity when compared to its parent molecule AzaHx-PI (**1**). However, 2-Pyr-AzaHx is detrimental to DNA binding because of an unfavorable steric clash upon stacking in the minor groove.

Introduction

Polyamides are a set of well-known DNA-minor groove targeting small molecules that consist of *N*-methylpyrrole (P) and/or *N*-methylimidazole (I) heterocycles.[1] Modeled from naturally occurring DNA binding agents, Netropsin and Distamycin A,[2,3] these molecules can form 2:1 antiparallel side-by-side stacked dimers upon binding to the DNA minor groove.[1,4] These molecules interact with DNA by forming specific hydrogen bonds and van der Waals contacts with different bases. Such polyamides can be programmed to target any DNA sequence based on the recognition rules developed by Dervan and coworkers.[5,6] The recent development of polyamides focuses largely on the synthesis of large hairpin molecules or tandem polyamides with increased specificity for longer sequences[7,8] or modified polyamides conjugates that can serve as DNA probes and enable polyamide with various functions.[9,10] These polyamides exhibit the ability to manipulate gene expression, thus, advancing the role of polyamides in gene therapy.

However, challenges still exist for efficient cell uptake. This calls for the development of small-size, soluble, and stable polyamides that are more easily synthesized. To improve cell uptake, solubility, stability, and more importantly, the feasibility for nuclear localization testing, we previously developed a novel class of inherently fluorescent small-size diamino polyamides by introducing 2-(*p*-anisyl)-benzimidazole (Hx) moiety to function as two consecutive pyrroles, 'P-P' or formamido-pyrrole, 'f-P'.[11,12] These molecules are hybrids of Hx with two conventional 5-membered heterocycles (P or I). The new Hx-amides can form a 2:1 side-by-side, antiparallel, stacked homodimer in the DNA minor groove and retain strong DNA binding affinity and specificity. In addition, the intrinsically fluorescent properties confer these molecules with the ease of cell study. In a recent study, an extra positive charge was added to Hx-IP at either P or I. These molecules have demonstrated to not only have higher binding affinity but also improved cell uptake in NIH3T3 mouse fibroblast cells and A549 cancer cells.[13]

To expand the DNA recognition repertoire of Hx-amides and enhance sequence specificity, we recently reported the design and synthesis of a second generation *p*-anisyl-AzaHx-amides (AzaHx-amides).[14] Derived from Hoechst 33258, the AzaHx-amides are structurally similar to Hx-amides but contain 2-(*p*-anisyl)-4-azabenzimidazole (AzaHx) moiety at the *N*-terminus. The novel AzaHx moiety mimicked P-I or f-I in DNA recognition and consequently, enabled G/C base pair recognition of these molecules. Because of the side-by-side, antiparallel stacking manner of these molecules, introduction of the AzaHx moiety extended DNA site recognition from central CG (recognized by Hx-I) to central GC (recognized by AzaHx-P, Fig. 5), even CGCG recognition. Recognized selectively by AzaHx-PI (**1**), the 5'-ACGCGT-3' sequence is of great importance, because it resides in the control region of the human Dbf4 gene promoter: the Mlu1 Cell Cycle Box (MCB) sequence. Expression of this gene is closely related to the development of several cancers.[15-19]

[a] B. Liu, M. Lee, W. D. Wilson
Department of Chemistry, Georgia State University, Atlanta, GA 30303, USA
E-mail: mosesleenf@gmail.com; wdw@gsu.edu
[b] L. Pett, K. Kiakos, J. A. Hartley
Cancer Research UK Drug-DNA interactions Research Group, UCL Cancer Institute, London, WC1E 6BT, UK
[c] P. Patil, V. Satam, M. Lee
Department of Chemistry, Hope College, Holland, MI 49423, USA
[d] M. Lee
Current address: M. J. Murdock Charitable Trust, 703 Broadway Street, Suite 710, Vancouver, WA 98660, USA

Supporting information for this article is given via a link at the end of the document.

After successfully establishing that the fluorescent AzaHx moiety functions as f-I or P-I, we designed new molecules by introducing a pyridyl (Pyr) moiety in place of the *p*-anisyl moiety of *p*-anisyl-AzaHx-amides to further extend the recognition capability of AzaHx polyamides. The newly introduced pyridyl moiety can potentially function as a H-bond acceptor and target G as showed by previous literature.[20-23] Consequently, we hypothesized that the resulting Pyr-AzaHx could act as I-I and recognize two consecutive GG base alignments along a DNA sequence. It was thought that in a 2:1 antiparallel stacked manner, the new molecules would target a GGCC site.

To test our hypothesis, three different Pyr-AzaHx-amides, namely, 3-Pyr-AzaHx-PI (**2**), 3-Pyr-AzaHx-PP (**3**), and 2-Pyr-AzaHx-PP (**4**) (Fig.1) were designed and synthesized. Excitation of a 1 μM aqueous solution of compound **2** at 310 nm (λ_{max} in uv spectrum) produced a broad emission from 340 to 440 nm that maximized at about 380 nm. The position of the nitrogen in the pyridine ring was adjusted to identify the optimal position responsible for stronger binding. Their predicted cognate sequences are 5'-ACGCGT-3' and/or 5'-AGGCCT-3' for compound **2,** 5'-ATGCAT-3' and/or 5'-AGGCCT-3' for compound **3,** 5'-AGGCCT-3' and/or 5'-ATGCAT-3' for compound **4** and the binding schemes are presented in Fig.1. The predicted binding site maintains central GC base alignment but the immediate flanking bases on both ends are GC rich. In this study, an extensive series of biophysical methods including DNase I footprinting, thermal melting (Tm), surface plasmon resonance (SPR), circular dichroism (CD), and Spartan calculations was employed to investigate the DNA-binding properties of these novel molecules. The results have provided us with new guidance for development of next stage molecules.

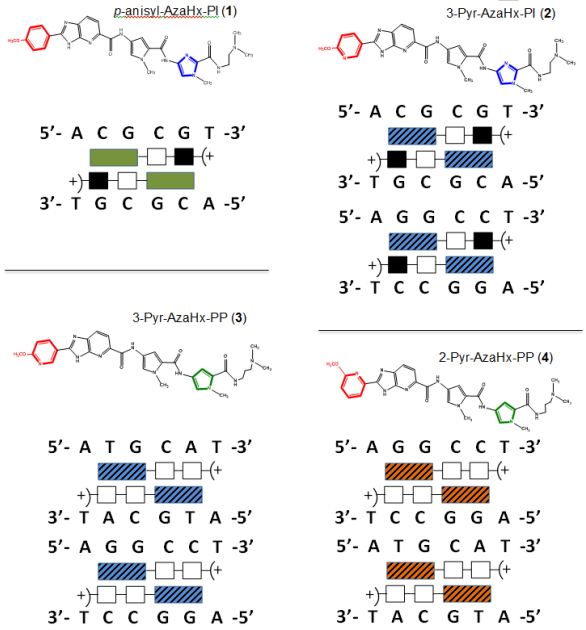


Figure 1. Structures and schematic presentation of predicted binding to DNA sites of three Pyr-AzaHx polyamides. Blue rectangles represent 3-Pyr-AzaHx moiety, empty squares represent pyrrole, filled squares represent imidazole, and orange rectangles represent 2-Pyr-AzaHx group. 2-Pyr-AzaHx-PP has two predicted recognition sites.

Results

Determination of relative DNA binding strength of Pyr-AzaHx polyamides using thermal melting (Tm)

To test the binding of these newly designed molecules to their predicted DNA sites, thermal melting experiments were carried out. ΔTm, the difference of melting temperature of the bound DNA complexes and melting temperature of DNA alone, reflects the extent to which DNA is stabilized by the molecules, and thus the relative binding strength. The ΔTm results of three molecules and their predicted cognate DNAs are listed in Table 1. It is clear that 3-Pyr-AzaHx-PI (**2**) binds to both 5'-ACGCGT-3' and 5'-ATGCAT-3' sequences quite strongly (15.9 °C and 15.5 °C, respectively), but does not have high affinity for 5'-AGGCCT-3' (8.8 °C). With 3-Pyr-AzaHx-PP (**3**), strong binding to 5'-ATGCAT-3' (14.5 °C) and weak binding to 5'-AGGCCT-3' (7.8 °C) and 5'-ACGCGT-3' (7.4 °C) were observed. These results suggest that the 3-Pyr-AzaHx moiety functions similarly to *p*-anisyl-AzaHx or f-I or PI and that the pyridine group does not make specific H-bond contacts with G. Thus, the observed target sequences of 3-Pyr-AzaHx-PI (**2**) and 3-Pyr-AzaHx-PP (**3**) are ACGCGT and ATGCAT, respectively.

We then modified the 3-Pyr-AzaHx by relocating the nitrogen to *meta*-position (Fig.1). The resulting 2-Pyr-AzaHx-PP (**4**) was tested against the three DNAs for its role in stabilization. Surprisingly, 2-Pyr-AzaHx-PP (**4**) showed hardly any affinity for all three DNAs.

Table 1. Thermal melting profile of three Pyr-AzaHx polyamides binding to their predicted cognate sequences and mutant sequences.

	CGCG	TGCA	GGCC
Tm of DNAs (°C)	68.8	61	67.2
Polyamides		ΔTm (°C)	
3-Pyr-AzaHx-PI (**2**)	15.9	15.5	8.8
3-Pyr-AzaHx-PP (**3**)	7.4	14.5	7.8
2-Pyr-AzaHx-PP (**4**)	1.2	0.8	0

Determination of DNA selectivity of Pyr-AzaHx polyamides by DNase I footprinting

DNase I footprinting was conducted to further investigate and compare the behavior of these molecules on a larger DNA scale. The binding affinity and selectivity of 3-Pyr-AzaHx-PI (**2**) was investigated using an engineered DNA fragment 1 that consists of the cognate sequence (based on Tm results): 5'-ACGCGT-3', as well as four mutant/non-cognate sites: 5'-ACCGGT-3', 5'-AAATTT-3', 5'-ACACGT-3', and 5'-AGCGCT-3' (Fig. 2A). The footprinting gel of 3-Pyr-AzaHx-PI (**2**) showed a clear footprint beginning at 0.5 μM for the cognate sequence (Fig. 2A), exhibiting a similar binding pattern to AzaHx-PI (**1**) (Fig. S5A).[14] Only a very weak and incomplete single mismatch protection site is shown at the sequence, 5'-ACACGT-3', indicating a weak tolerance of 3-Pyr-azaHx as well as azaHx (Fig. S5A) for a "A" over its preferred "G". No other significant protection sites were

evident at concentrations of up to 20 μM with this DNA sequence. The Tm and footprinting data together indicate that the 3-Pyr-AzaHx-PI (**2**) is very selective for the central GC dinucleotides recognition, with a selection factor of more than 40 fold according to the footprinting studies.

The DNA binding property of 3-Pyr-AzaHx-PP (**3**) was evaluated using engineered DNA fragment 2 (Fig. 2B). The molecule is predicted to target 5'-WWGCWW-3' (W = A or T). Two cognate sequences: 5'-ATGCAT-3' and 5'-AAGCAA-3', along with four non-cognate sites: 5'-ACGCGT-3', 5'-AAATTT-3', 5'-ACTGGT-3', and 5'-ATCGAT-3' were present in DNA fragment 2. A clear footprint was visible at both consensus sequences starting at 3 μM. However, evident footprint was also detected at the non-cognate site containing flipped central GC pair: 5'-ATCGAT-3' at 3 μM. The equally strong protection of this non-cognate site and the featured central base pair indicates that the molecule might bind to the sequence in the reversed orientation, that is, *N*-terminus of the molecule is aligned with 3' end of DNA. What also appeared on the gel was another footprint at 5'-ACTGGT-3' at 10 μM (Fig. 2B). The footprinting result suggests that 3-Pyr-AzaHx-PP (**3**) not only binds weakly

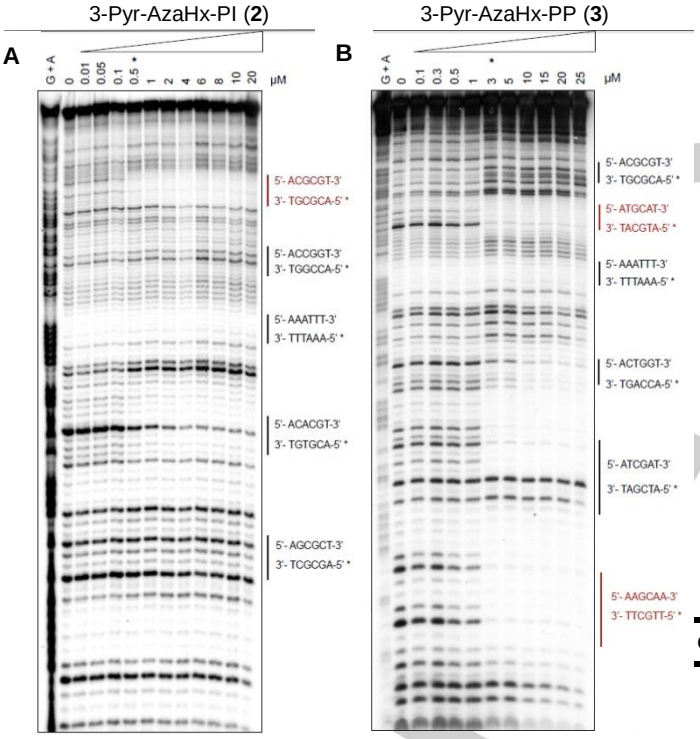


Figure 2. Autoradiograms of DNase I footprinting gels corresponding to 3-Pyr-AzaHx-PI, **2** (A) and 3-Pyr-AzaHx-PP, **3** (B). The binding of **2** and **3** were assessed using DNA fragment 1 and 2, respectively. The ^{32}P-radiolabelled DNA fragment was incubated for 1 h at 25°C in TN buffer (10 mM Tris-HCl, 10 mM NaCl, pH 7.0) containing the required polyamide concentration. DNase I digestion was initiated by the addition of DNase I solution (20mM NaCl, 2mM $MgCl_2$, 2mM $MnCl_2$, DNase I 0.02U, pH 8.0) and was terminated after 3 min by snap freezing the samples on dry ice. Following lyophilisation and resuspension in formamide loading dye (95% formamide, 20mM EDTA, 0.05% bromophenol blue, and 0.05% xylene cyanol), the cleavage products of the DNase I digestion reactions were resolved on 10% polyacrylamide gels. The concentrations (μM) used are shown at the top of each gel. 0; controlled cleavage reaction, containing no drug. The cognate sequences are labelled red and the non-cognate sites are also indicated. Asterisk marks the concentration (μM) at which footprints become evident. G+A represents a formic acid-piperidine marker specific for purines.

to its cognate 5'-ATGCAT-3' sequence compared to 3-Pyr-AzaHx-PI (**2**), but also has poor sequence selectivity, as evidenced by the off-target binding sites.

With the lack of binding according to the results from Tm studies, further investigation on 2-Pyr-AzaHx-PP (**4**) was conducted using a broad spectrum of DNA sequences to determine whether it resembles AzaHx-PP, recognizing 5'-WWGCWW-3', or IIPP, recognizing 5'-WGGCCW-3'. DNase I footprinting of 2-Pyr-AzaHx-PP (**4**) binding to DNA fragment 2 that contains two 5'-WWGCWW-3' sites (5'-ATGCAT-3' and 5'-AAGCAA-3') was carried out. No footprint was detected up to 25 μM (Fig. S2A). On another trial using the engineered fragment 3 containing 5'-WGGCCW-3' (5'-AGGCCT-3'), the results showed no binding up to 50 μM (Fig. S2B). These results agree with those collected from Tm studies, indicating that a change in the nitrogen position along with the attached methoxy unit at the 3-position of the pyridyl group at the *N*-terminus abrogated the binding of Pyr-AzaHx moiety, thus, making it incompetent structural element of design to target any DNA site.

Determination of binding kinetics and affinity of Pyr-AzaHx polyamides by SPR

SPR provides us with real-time information to quantitatively determine the interaction between molecules and DNAs. After confirming the corrected cognate DNA for each molecule and to assist evaluation of binding affinity and kinetics compared to previous work, SPR experiments were carried out with three predicted cognate DNA sequences that are immobilized on a sensorchip. The resulting SPR sensorgrams are shown in Fig. 3 and the kinetic and equilibrium binding constants are listed in Table 2. The strongest binding was observed between 3-Pyr-AzaHx-PI (**2**) and its cognate DNA, 5'-ACGCGT-3' (K_D = 0.75 ± 0.01 nM), followed by its interaction with 5'-ATGCAT-3' (K_D = 3.2 nM). The result shows that 3-Pyr-AzaHx-PI (**2**) is very specific to the central GC base alignment, but not very selective for the immediate flanking bases, especially among C, A or T. Strong interactions between 3-Pyr-AzaHx-PP (**3**) and its cognate DNA, 5'-ATGCAT-3' were also confirmed by SPR with K_D equals 1.1 nM. These results are in good agreement with both Tm and

Table 2. Kinetic rate constants and equilibrium constants derived from SPR

Compound	DNA	k_a (x10^4 M^{-1}s^{-1})	k_d(x10^{-3}s^{-1})	K_D (nM)
3-Pyr-AzaHx-PI	ACGCGT	16 ± 0.3	0.12 ± 0.001	0.75 ± 0.01
	ATGCAT	40 ± 0.7	1.30 ± 0.01	3.20 ± 0.2
	AGGCCT	1.70 ± 0.5	6.60 ± 0.03	397 ± 17
3-Pyr-AzaHx-PP	ACGCGT	3.30 ± 0.03	34 ± 0.04	1035 ± 22
	ATGCAT	35 ± 0.2	0.40 ± 0.001	1.10 ± 0.03
	AGGCCT	1.00 ± 0.01	40 ± 0.4	3996 ± 472
2-Pyr-AzaHx-PP	ACGCGT	ND[a]	ND[a]	1420 ± 700
	ATGCAT	2.30 ± 0.02	16 ± 0.06	698 ± 237
	AGGCCT	NB[b]	NB[b]	NB [b]

[a] The values are too fast to be determined. [b] No binding is detected.

footprinting data within error range, indicating that the molecules are well stacked and form strong hydrogen bonding with DNA base pairs. The strong interactions are also featured by very slow dissociation rate (~10^{-4}·s^{-1}, Table 2, Fig. 3). On the other hand, 3-Pyr-AzaHx-PP (**3**) was found to bind very weakly to 5'-ACGCGT-3' (K_D = 1035 ± 22nM), suggesting that a switch of the molecule structure from I to P is incompatible with G. Also consistent with the Tm result is the fact that 3-Pyr-AzaHx-PI (**2**) binds to the mismatched sequence 5'-AGGCCT-3', albiet with significantly reduced affinity compared to its cognate DNA. If 3-Pyr-AzaHx interacts with GG to some extent, then 3-Pyr-AzaHx-PP (**3**) should target 5'-AGGCCT-3' according to the binding scheme depicted in Fig. 1. To our surprise, 3-Pyr-AzaHx-PP (**3**) exhibited around 10 times weaker affinity to 5'-AGGCCT-3', indicating that the 3-Pyr-AzaHx moiety is not the determinant factor in this binding event. In an attempt to specifically recognize two consecutive GG base pairs, 3-Pyr-AzaHx was modified to 2-Pyr-AzaHx moiety, the binding of 2-Pyr-AzaHx-PP (**4**) was tested on SPR against the three DNA sequences. In agreement with the Tm data, 2-Pyr-AzaHx-PP (**4**) binds very weakly to 5'-ACGCGT-3' and 5'-ATGCAT-3' and showed no binding at all to 5'-AGGCCT-3'. Clearly, the relative position of the pyridine nitrogen along with the methoxy group is pivotal for binding. A close observation of the sensorgram shows that some fitting curves are diverged from experimental data. Especially, the experimental deviations are above the global fit. This indicates that non-specific binding begins to occur at higher concentrations.

Confirmation of binding of Pyr-AzaHx polyamides in DNA minor groove by CD

The binding mode of Pyr-AzaHx polyamides to their cognate DNA was evaluated by CD. Molecules that bind in the minor groove typically have positively induced band.[24-27] The Pyr-AzaHx molecules alone do not generate any peak. Upon titration of 3-Pyr-AzaHx-PI (**2**) and 3-Pyr-AzaHx-PP (**3**) to their cognate DNA, strong positive signals are induced at 350 nm for both molecules (Fig. 4).This is typical for polyamides and Hx-amides and is indicative of DNA minor groove binding.

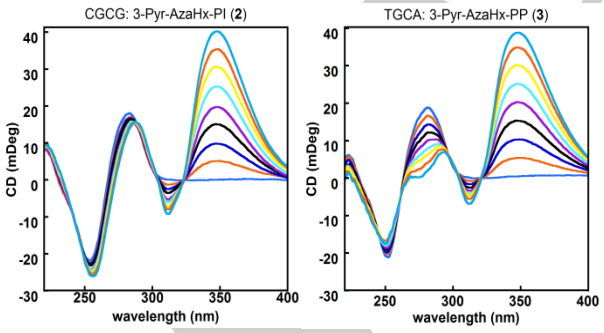


Figure 4. CD spectra depicting 3-Pyr-AzaHx molecules (**2**) and (**3**) binding to their cognate DNAs, respectively. The DNA sequences are listed in Supporting Information. The buffer consists of10 mM sodium phosphate, 1 mM EDTA, pH 6.2.

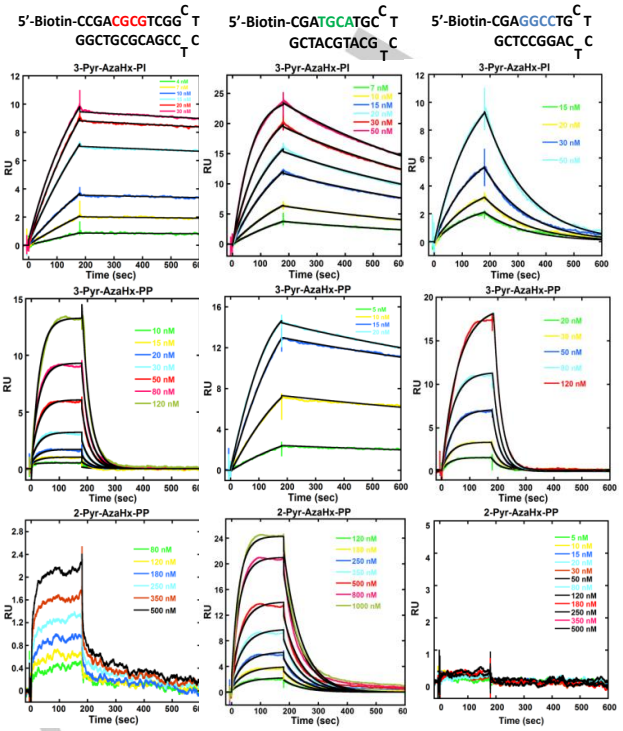


Figure 3. SPR sensorgrams of Pyr-AzaHx polyamides binding to different DNAs. The DNA sequences are listed at the top of the column and the compounds are specified on each sensorgram. The whole sequences of the DNAs are listed in the Supporting Information. The colored lines are experimental data and the black overlays are kinetic fitting. The sensorgram at the bottom left corner was fitted with steady state fitting and bottom right corner sensorgram shows no binding. The buffer used for SPR is 10 mM cacodylic acid, 100 mM NaCl, 1mM EDTA and 0.05% v/v surfactant Polysorbate 20 (P20), pH 6.2, 25°C.

Discussion

Previous development of hairpin polyamides involved incorporation of benzimidazole (Bi) or imidazolepyridine (Ip) into the polyamide core structure and showed that the resulting molecules displayed strong binding affinity and sequence specificity.[28-32] The need for more soluble, smaller polyamides to improve cell uptake and biological functions requires new strategies to design the molecules. Our previously designed small-size *p*-anisyl-Hx-amides mimic f-P or P-P, form antiparallel stacked dimer, and follow pyrrole-imidazole DNA recognition principles.[1,4] Biological studies of Hx-amides, specifically, Hx-IP targeting the inverted repeat of CCAAT Box 2 on the topoisomerase IIα promoter to inhibit NF-Y binding, have shown increased cell uptake and enhanced subsequent biological activities in NIH3T3 and A549 cells compared to high molar mass polyamides, such as hairpins and H-pins.[33-36] The improved cell permeability and DNA binding affinity from incorporating *N*-terminal *p*-anisyl-Hx moiety in small Hx-amides offers a promising way for designing small-size, soluble polyamides. However, the P-P like DNA binding property of *p*-anisyl-Hx is limited to A or T recognition. Especially, according to

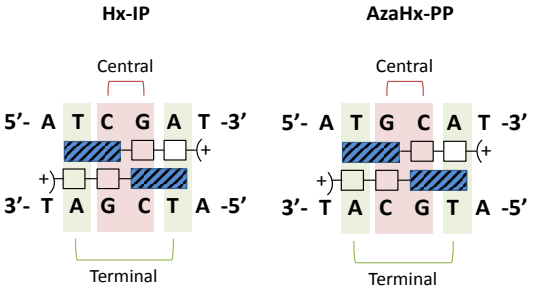


Figure 5. Central and terminal pairing rules illustrated by different binding of Hx-IP and AzaHx-PP. The central and terminal base pairs are indicated in pink and green boxes, respectively.

the central and terminal pairing rules (Fig. 5), the central two heterocycle pairs of a tri-amide or tetra-amide homodimer recognize the two central base pairs while the terminal formamido and imidazole/pyrrole (tri-amides) or the two "outer" heterocycles pairs (tetra-amides) specifically target the immediate flanking base pairs on either side of the central base pairs.[37,38] Therefore, by this recognition pattern, Hx-amides are restricted to central CG dinucleotide recognition and the subsequent target of biologically relevant sequences is constrained (Fig. 5).

The novel development of AzaHx-amides showed strong binding affinity and additional GC selectivity. The AzaHx moiety functions as an alternative to f-I or P-I capable of forming an antiparallel dimer. According to the central and terminal pairing rule, AzaHx-amides are able to confer central and outer GC dinucleotide recognition (Fig. 5), thus greatly extending the sequence it can target.

To further expand the recognition repertoire to GGCC, our new modification strategy is to replace the *p*-anisyl group in AzaHx-amdies with pyridyl groups to generate Pyr-AzaHx-amides. The central GC dinucleotide is unchanged, as is the AzaHx moiety. The extra nitrogen of *N*-terminal pyridyl moiety could provide an additional hydrogen bond acceptor for G recognition. To test this hypothesis, the DNA binding properties of the modified Pyr-AzaHx-amides are investigated by a complementary set of biophysical and biochemical methods.

Tm results show that 3-Pyr-AzaHx functions similarly to AzaHx or P-I or f-I

Tm studies were performed to initially screen the Pyr-azaHx complexes for binding DNA. The prediction was that Pyr-AzaHx would act similarly to two consecutive imidazole units, or I-I, to selectively recognize GG dinucleotides. But the low ΔTm value of 3-Pyr-AzaHx-PP (**3**) with 5'-AGGCCT-3' and high value with 5'-ATGCAT-3' (Table. 1) suggested that 3-Pyr-AzaHx could not form extra hydrogen bonds with two consecutive GG dinucleotides to stabilize the molecule-DNA complex. Rather, the 3-Pyr-azaHx moiety behaved more akin to AzaHx or P-I or f-I, and it anchored itself in the minor groove of TG dinucleotides with strong hydrophobic contacts. Evidence of the high ΔTm value of 3-Pyr-AzaHx-PI (**2**) with 5'-ACGCGT-3' and low value with 5'-AGGCCT-3' further confirmed the similar functionality of 3-Pyr-AzaHx to AzaHx. It is highly possible that the more polar nitrogen atom of the pyridyl moiety, compared to the opposite methinyl group, is rotated away from the floor of minor groove or is not positioned properly for hydrogen bonding. Instead, similar to pyrrole binding units, the hydrophobic C-H units in the pyridyl group were directed more favorably toward the hydrophobic floor in the minor groove. Thus, the preferred cognate binding sites for 3-Pyr-AzaHx-PI (**2**) and 3-Pyr-AzaHx-PP (**3**) were 5'-ACGCGT-3' and 5'-ATGCAT-3', respectively. It was also noticeable that 3-Pyr-AzaHx-PI (**2**) bound with high affinity to the mismatch site 5'-ATGCAT-3', indicating that when G was not available, imidazole could tolerate an A/T base pair quite well, thus compromising its selectivity. This result agreed with our previous work that showed imidazole on imidazole stacking in polyamides bound preferentially to T/G or G/T but it tolerated C/G and G/C base pairs.[26,39-41]

The extremely low ΔTm values of the modified 2-Pyr-AzaHx-PP (**4**) to any of the three sequences indicate that the position of the pyridine nitrogen along with the methoxy group can be quite important to DNA binding. In fact, to structurally visualize the relative position of the antiparallel stacked homodimer, Spartan´16 software was used for Molecular Mechanics calculation. Our models show that when two 2-Pyr-AzaHx-PP (**4**) molecules are anti-parallely stacked, the bulky 3-methoxy group attached to pyridine is sterically encumbered with the *N*-methyl group of the C-terminal pyrrole (Fig. S4). On the other hand, if the nitrogen of pyridine is attracted toward the minor groove floor, the 3-methoxy group appears to be too big and it also affects the stacking of two molecules. It is difficult for the minor groove to accommodate favorably. Therefore, either way the nitrogen is pointing, the 3-methoxy-2-pyridyl moiety would encounter steric clash upon binding to DNA as stacked dimer, resulting in thermodynamically unstable molecule-DNA complex.

Binding affinity and selectivity of 3-Pyr-AzaHx-amides

The sequence specificity of the new AzaHx-amides was investigated by DNase I footprinting. 3-Pyr-AzaHx-PI (**2**) showed a clear footprint toward 5'-ACGCGT-3' at 0.5 μM and more than 40-fold selectivity against the tested DNA fragment. This is very similar in terms of both affinity and selectivity, to its parent structure, AzaHx-PI (**1**), in which the AzaHx moiety contains *p*-anisyl instead of a pyridyl group. AzaHx-PI was previously reported to footprint at 5'-ACGCGT-3' using the same DNA fragment starting at 0.5 μM with at least 40-fold selectivity (Fig. S5A).[14] Since 3-Pyr-AzaHx is experimentally shown to function as AzaHx or P-I or f-I, two other AzaHx-amide analogs recognizing the same cognate DNA were also included into the comparison.[38,42-44] On the same DNA fragment, f-IPI[42,43], a N-formamido pyrrole/imidazole polyamide, generated a footprint at the cognate site, beginning at 0.05 μM. However, another clear footprint at 5'-AGCGCT-3' is produced at 1 μM, suggesting that the sequence selectivity of f-IPI is reduced even though it has relatively high binding affinity. On the other hand, PIPI[42], a non-formamido pyrrole/imidazole polyamide, footprinted starting at 1 μM but showed comparable sequence selectivity to 3-Pyr-AzaHx-PI (**2**).

With a slightly lower binding affinity than 3-Pyr-AzaHx-PI (**2**), 3-Pyr-AzaHx-PP (**3**) footprinted at 3 μM at its cognate site 5'-ATGCAT-3'. At the same concentration, another footprint appeared at 5'-ATCGAT-3'. Given the inverted central bases and the degenerate P/P recognition of A/T or T/A, it is likely that the molecule can bind in a reversed manner. Another footprint for 5'-ACTGGT-3' at 10 μM indicates that 3-Pyr-AzaHx-PP (**3**) is not very selective. When compared to its analogs that target the same cognate DNA, AzaHx-amide **3** showed higher binding affinity. AzaHx-PP, which also contains *p*-anisyl instead of a pyridyl group, generated a footprint at 10 μM at the cognate site, but one other footprint at 5'-ATCGAT-3' at 15 μM, indicating that AzaHx-PP could not differentiate the two sequences very well, either (Fig. S5B). For f-IPP, even though a clear footprint at the cognate sequence was evident at 3 μM, equally strong footprint also appeared at three non-cognate sequences (Fig. S5C).. These results suggest f-IPP has low selectivity. Taken together, 3-Pyr-AzaHx-PP (**3**) exhibited stronger binding affinity with slightly compromised selectivity compared to other strunctural similar compounds.

To more accurately compare the binding affinities, SPR was performed to obtain the binding kinetics and equilibrium binding affinities of the new molecules and three DNA sequences. The SPR results are in good agreement with Tm and footprinting data. Interestingly, the binding affinity of 3-Pyr-AzaHx-PI (**2**) (K_D = 0.75 nM) is much higher than its analogs: AzaHx-PI (**1**) (K_D = 400 nM),[14] f-IPI (K_D = 19 nM),[37,42]and PIPI (K_D = 7100 nM)[42] on binding to the cognate sequence, 5'-ACGCGT-3'. These results reveal that 3-Pyr-AzaHx-PI (**2**) has drastically improved affinity, at least 20 times over its previously reported analogs. In particular, 3-Pyr-AzaHx-PI (**2**) has much higher binding affinity than its closest analog and parent structure AzaHx-PI (**1**). When comparing 3-Pyr-AzaHx-PP (**3**) (K_D = 1.1 nM) to its analogs AzaHx-PP (K_D = 110 nM, data not published) and f-IPP (K_D = 120 nM),[37] 3-Pyr-AzaHx-PP (**3**) showed around 100 times higher binding affinity. These results together with the footprinting data suggest that the 3-Pyr-AzaHx moiety helps to improve binding affinity while maintains a reasonable level of selectivity.

Interestingly, in a separate study reported recently by us, adding an extra positive charge to Hx-amides showed that the resulting dicationic HxIP* and HxI*P have higher binding affinity but reduced selectivity toward their consensus site 5'-ATCGAT-3' (ΔTm = 32 °C and 30 °C, respectively) than their mono-charged analog, HxIP (ΔTm = 15 °C).[13] Therefore, we suspect that the high affinity of 3-Pyr-AzaHx-amides is due to protonation of the nitrogen on the pyridyl moiety, thus contributing to an extra positive charge for stronger attraction to the negatively charged DNA. In fact, previous literaturs have also shown that increasing positive charges can facilitate DNA binding.[45-47] To test this idea, the binding of 3-Pyr-AzaHx-PI (**2**) to 5'-ACGCGT-3' was performed by SPR at pH 9.0. The resulting affinity was lowered by about 16 times than that at pH 6.2 (Fig. 6), thus indicating that the increased binding affinity of 3-Pyr-AzaHx-PI (**2**) could be attributed to the protonation of the pyridine nitrogen. Altogether, these results suggest that modification of the *N*-terminal *p*-anisyl to *N*-terminal pyridiyl, or converting AzaHx to 3-Pyr-AzaHx not only retained comparable selectivity but also significantly improved binding affinity under biological conditions.

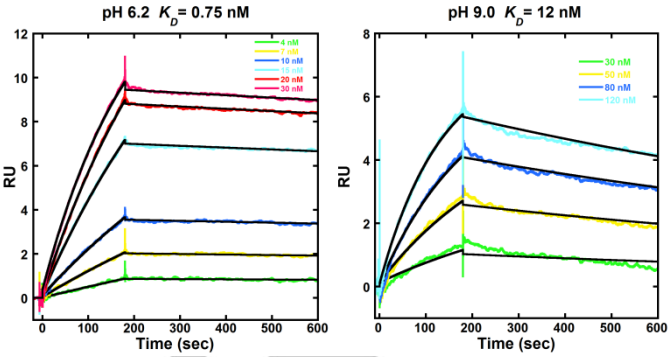


Figure 6. SPR sensorgrams and equilibrium binding constants of 3-Pyr-AzaHx-PI under different pH. The buffer used at pH 6.2 is consisted of 10 mM cacodylic acid, 100 mM NaCl, 1mM EDTA and 0.05% v/v surfactant Polysorbate 20 (P20), 25°C. The buffer at pH 9.0 contains 50 mM Tris-HCl, 100 mM NaCl, 1mM EDTA and 0.05% v/v P20, 25°C. The whole sequence of 5'-ACGCG-3' is listed in the Supporting Information.

Conclusions

In this study, we have designed and synthesized a novel class of small-sized polyamides, aiming to expand the DNA recognition repertoire to GGCC sequence. Our complementary biochemical and biophysical studies have shown that although the newly designed Pyr-AzaHx amides do not specifically recognize GG dinucleotides, the 3-Pyr-AzaHx amides do possess improved functionality compared to previous analogs. Especially, the increased binding affinity under biological conditions could be due to protonation of the pyridine nitrogen. The results have greatly extended our knowledge on molecular interactions and thus compound modification and optimization. Meanwhile, challenges on designing GG dinucleotides targeting, fluorescent compounds with good cell uptake are still needed.

Experimental Section

Experimental Details and supplementary figures are provided in the supporting information.

Acknowledgements

This work was supported by a grant from the National Science Foundation (CHE 0809162 to M.L and W.D.W), a program grant from Cancer Research UK (C2259/A16569 to J.A.H), and a Medical Research Council UK funded studentship (L.P). The biosensor- SPR work was supported by the National Institutes of Health for W.D.W(GM 111749).

Keywords: polyamides•AzaHx•3-Pyr-AzaHx•fluorescent•DNA recognition

[1] M. Mrksich, W. S. Wade, T. J. Dwyer, B. H. Geierstanger, D. E. Wemmer, P. B. Dervan, *Proc. Natl. Acad. Sci. U. S. A*., **1992**, *89*, 7586-7590.

[2] A. C. Finlay, F. A. Hochstein, B. A. Sobin, F. X. Murphy, *J. Am. Chem. Soc.,* **1951**, *73*, 341-343.

[3] F. Arcamone, S. Penco, P. Orezzi, V. Nicolella, A. Pirelli, *Nature,* **1964**, *203*, 1064-1065.

[4] S. W. Wade, M. Mrksich, P. B. Dervan, *J. Am. Chem. Soc.,* **1992**, *114*, 8783-8794.

[5] P. B. Dervan, *Bioorg. Med. Chem.,* **2001**, *9*, 2215-2235.

[6] P. B. Dervan, B. S. Edelson, *Curr. Opin. Struct. Biol.,* **2003**, *13*, 284-299.

[7] G. He, E. Vasilieva, G. D. Harris Jr., K. J. Koeller, J. K. Bashkin, C. M. Dupureur, *Biochimie,* **2014**, *102*, 83-91.

[8] Y. Kawamoto, A. Sasaki, A. Chandran, K. Hashiya, S. Ide, T. Bando, K. Maeshima, H. Sugiyama, *J. Am. Chem. Soc.,* **2016**, *138*, 14100-14107.

[9] Y. Kawamoto, T. Bando, H. Sugiyama, *Bioorg. Med. Chem.,* **2018**, *26*, 1393-1411**.**

[10] G. S. Erwin, M. P. Grieshop, A. Ali, J. Qi, M. Lawlor, D. Kumar, I. Ahmad, A. McNally, N. Teider, K. Worringer, R. Sivasankaran, D. N. Syed, A. Eguchi, M. Ashraf, J. Jeffery, M. Xu, P. M. C. Park, H. Mukhtar, A. K. Srivastava, M. Faruq, J. E. Bradner, A. Z. Ansari, *Science,* **2017**, *358*, 1617-1622.

[11] S. Chavda, Y. Liu, B. Babu, R. Davis, A. Sielaff, J. Ruprich, L. Westrate, C. Tronrud, A. Ferguson, A. Franks, S. Tzou, C. Adkins, T. Rice, H. Mackay, J. Kluza, S. A. Tahir, S. Lin, K. Kiakos, C. D. Bruce, W. D. Wilson, J. A. Harley, M. Lee, *Biochemistry,* **2011**, *50*, 3127-3136.

[12] V. Satam, P. Patil, B. Babu, M. Gregory, M. Bowerman, M. Savagian, M. Lee, S. Tzou, K. Olson, Y. Liu, J. Ramos, W. D. Wilson, J. P. Bingham, K. Kiakos, J. A. Hartley, M. Lee, *Bioorg. Med. Chem. Lett.,* **2013**, *23*, 1699-1702.

[13] L. Pett, K. Kiakos, V. Satam, P. Patil, S. Laughlin-Toth, M. Gregory, M. Bowerman, K. Olson, M. Savagian, M. Lee, W. D. Wilson, D. Hochhauser, J. A. Hartley, *Biochim. Biophys. Acta*., **2017**, *1860*, 617-629.

[14] V. Satam, B. Babu, P. Patil, K. A. Brien, K. Olson, M. Savagian, M. Lee, A. Mepham, L. B. Jobe, J. P. Bingham, L. Pett, S. Wang, M. Ferrara, C. D. Bruce, W. D. Wilson, M. Lee, J. A. Harley, K. Kiakos, *Bioorg. Med. Chem. Lett.,* **2015**, *25*, 3681-3685.

[15] X. Wu, H. Lee, *Oncogene,* **2002**, *21*, 7786-7796.

[16] J. W. Knockleby, J. Romero, K. A. Kylie, H. Lee, *Curr. Top. Biochem. Res*., **2010**, *12*, 43.

[17] D. H. Charych, M. Coyne, A. Yabannavar, J. Narberes, S. Chow, M. Wallroth, C. Shafer, A. O. Walter, *J. Cell. Biochem*., **2008**, *104*, 1075-1086.

[18] D. Bonte, C. Lindvall, H. Liu, K. Dykema, K. Furge, M. Weinreich, *Neoplasia*, **2008**,*10*, 920-931.

[19] Y. J. Sheu, B. Stillman, *Nature*, **2010**, *463*, 113-117.

[20] D. Renneberg, P. B. Dervan, *J. Am. Chem. Soc,* **2003**, *125*, 5707-5716.

[21] M. Munde, M. A. Ismail, R. Arafa, P. Peixoto, C. J. Collar, Y. Liu, L. Hu, M. H. David-Cordonnier, A. Lansiaux, C. Bailly, D. W. Boykin, W. D. Wilson, *J. Am. Chem. Soc.,* **2007**, *129*,13732-13743.

[22] M. I. Sánchez, O. Vázquez, J. Martínez-Costas, M. E. Vázquez, J. L. Mascarenas, *Chem. Sci.,* **2012**, *3*, 2383-2387.

[23] A. Paul, R. Nanjunda, A. Kumar, S. Laughlin, R. Nhili, S. Depauw, S. S. Deuser, Y. Chai, A. S. Chaudhary, M. H. David-Cordonnier, D. W. Boykin, W. D. Wilson, *Bioorg. Med. Chem.,* **2015**, *25*, 4927-4932.

[24] R. Lyng, A. Rodger, B. Norden*, Biopolymers*, **1992**, *32*, 1201-1214.

[25] X. L. Yang, C. Kaenzig, M. Lee, A. H. Wang, *Eur. J. Biochem*., **1999**, *263*, 646-655.

[26] E. R. Lacy, K. K. Cox, W. D. Wilson, M. Lee, *Nucleic. Acids. Res*., **2002**, *30*, 1834-1841.

[27] E. R. Lacy, N. M. Le, C. A. Price, M. Lee, W. D. Wilson, *J. Am. Chem. Soc*., **2002**, *124*, 2153-2163.

[28] D. Renneberg, P. B. Dervan, *J. Am. Chem. Soc*., **2003**, *125*, 5707-5716.

[29] C. A. Briehn, P. Weyermann, P. B. Dervan, *Chemistry*, **2003**, *9*, 2110-2122.

[30] M. A. Marques, R. M. Doss, S. Foister, P. B. Dervan, *J. Am. Chem. Soc*., **2004**, *126*, 10339-10349.

[31] R. M. Doss, M. A. Marques, S. Foister, D. M. Chenoweth, P. B. Dervan, *J. Am. Chem. Soc*., **2006**, *128*, 9074-9079.

[32] D. M. Chenoweth, A. Viger, P. B. Dervan, *J. Am. Chem. Soc*., **2007**, *129*, 2216-2217.

[33] K. Kiakos, L. Pett, V. Satam, P. Patil, D. Hochhauser, M. Lee, J. A. Hartley, *Chem. Biol*., **2015**, *22*, 862-875.

[34] A. Franks, C. Tronrud, K. Kiakos, J. Kluza, M. Munde, T. Brown, H. Mackay, W. D. Wilson, D. Hochhauser, J. A. Hartley, M. Lee, *Bioorg. Med. Chem.,* **2010**, *18*, 5553-5561.

[35] S. Nishijima, K. Shinohara, T. Bando, M. Minoshima, G. Kashiwazaki, H. Sugiyama, *Bioorg. Med. Chem.,* **2010**, *18*, 978-983.

[36] N. G. Nickols, C. S. Jacobs, M. E. Farkas, P. B. Dervan, *Nucleic Acids Res.*, **2007**, *35*, 363-370.

[37] K. L. Buchmueller, A. M. Staples, C. M. Howard, S. M. Horick, P. B. Uthe, N. M. Le, K. K. Cox, B. Nguyen, K. A. Pacheco, W. D. Wilson, M. Lee, *J. Am. Chem. Soc*., **2005**, *127*, 742-750.

[38] K. L. Buchmueller, A. M. Staples, P. B. Uthe, C. M. Howard, K. A. Pacheco, K. K. Cox, J. A. Henry, S. L. Bailey, S. M. Horick, B. Nguyen, W. D. Wilson, M. Lee, *Nucleic Acids Res.*, **2005**, *33*, 912-921.

[39] E. R. Lacy, B. Nguyen, M. Le, K. K. Cox, C. O'Hare, J. A. Hartley, M. Lee, W. D. Wilson, *Nucleic. Acids. Res*., **2004**, *32*, 2000-2007.

[40] V. C. Rucker, S. Foister, C. Melander, P. B. Dervan, *J. Am. Chem. Soc*., **2003**, *125*, 1195-1202.

[41] X. L. Yang, R. B. Hubbard, M. Lee, Z. F. Tao, H. Sugiyama, A. H. Wang, *Nucleic. Acids. Res*., **1999**, *27*, 4183-4190.

[42] T. Brown, H. Mackay, M. Turlington, A. Sutterfield, T. Smith, A. Sielaff, L. Westrate, C. Bruce, J. Kluza, C. O'Hare, B. Nguyen, W. D. Wilson, J. A. Hartley, M. Lee, *Bioorg.Med. Chem*., **2008**, *16*, 5266-5276.

[43] K. L. Buchmueller, S. L. Bailey, D. A. Mattews, Z. T. Taherbhai, J. K. Register, Z. S. Davis, C. D. Bruce, C. O'Hare, J. A. Hartley, M. Lee, *Biochemistry*, **2006**, *45*, 13551-13565.

[44] S. C. Lin, K. Kiakos, M. Lee, D. Hochhauser, J. A. Hartley, *Proc. 102nd Annual Meeting of the Am. Assoc. Cancer Res*. **2011**, Abs 665.

[45] B. Liu, S. Wang, K. Aston, K. J. Koeller, S. F. H. Kermani, C. H. Castaneda, M. J. Scuderi, R. Luo, J. K. Bashkin, W. D. Wilson, *Org. Biomol. Chem.,* **2017**, *15*, 9880-9888.

[46] J. L. Meier, D. C. Montgomery, P. B. Dervan, *Nucleic Acids Res*, **2012**, *40*, 2345-2356.

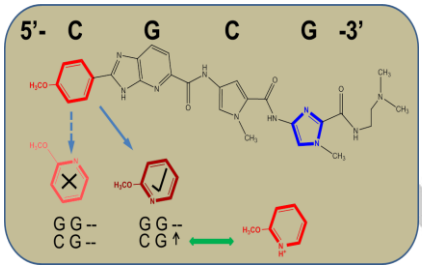
[47] P. B. Dervan, R. W. Burli, *Curr. Opin. Chem. Biol*, **1999**, *3*, 688-693.

Entry for the Table of Contents (Please choose one layout)

Layout 1:

FULL PAPER

Modification of AzaHx to pyrido-AzaHx does not enable the molecule to specifically recognize GG dinucleotides, instead, the 3-pyrido-AzaHx improved binding affinity, possibly due to protonation of the pyrido nitrogen.



*Author(s), Corresponding Author(s)**

Page No. – Page No.

Title

Layout 2:

FULL PAPER

((Insert TOC Graphic here; max. width: 11.5 cm; max. height: 2.5 cm))

*Author(s), Corresponding Author(s)**

Page No. – Page No.

Title

Text for Table of Contents