

# Sequence-Selective DNA Recognition with Peptide–Bisbenzamidine Conjugates

Mateo I. Sánchez, Olalla Vázquez, M. Eugenio Vázquez, José L. Mascareñas

## Peer reviewed version

This is the peer reviewed version of the following article: Sánchez, M. I., Vázquez, O., Vázquez, M. E. and Mascareñas, J. L. (2013), Sequence-Selective DNA Recognition with Peptide–Bisbenzamidine Conjugates. *Chem. Eur. J.*, 19: 9923-9929, which has been published in final form at <https://doi.org/10.1002/chem.201300519>. This article may be used for non-commercial purposes in accordance with Wiley Terms and Conditions for Use of Self-Archived Versions

## How to cite:

Sánchez, M. I., Vázquez, O., Vázquez, M. E. and Mascareñas, J. L. (2013), Sequence-Selective DNA Recognition with Peptide–Bisbenzamidine Conjugates. *Chem. Eur. J.*, 19: 9923-9929. doi:10.1002/chem.201300519

## Copyright information:

© 2013 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim. This article may be used for non-commercial purposes in accordance with Wiley Terms and Conditions for Use of Self-Archived Versions

## Sequence-Selective DNA Recognition with peptide-bisbenzamidine conjugates

Mateo I. Sánchez, Olalla Vázquez, M. Eugenio Vázquez,\* and José L. Mascareñas\*[a]

**Abstract:** Transcription factors (TFs) are specialized proteins that play a key role in the regulation of genetic expression. Their mechanism of action involves the interaction with specific DNA sequences, which usually takes place through specialized domains of the protein. However, achieving an efficient binding usually requires the presence of the full protein. This is the case for bZIP and zinc finger TF families, which cannot interact with their target sites when the DNA binding fragments are presented as isolated monomers. Herein we demonstrate that

the DNA binding of these monomeric peptides can be restored when conjugated to *aza*-bisbenzamidines, readily accessible molecules that interact with AT rich sites by insertion into their minor groove. Importantly, the fluorogenic properties of the *aza*-benzamidine unit allow to obtain details of the DNA interaction that are eluded in electrophoresis mobility shift assays (EMSA). The hybrids based on the GCN4 bZIP protein preferentially bind to composite sequences containing tandem bisbenzamidine-GCN4 binding sites (TCAT·AAATT). Fluorescence

reverse titrations show an interesting multiphasic profile consistent with the formation of competitive nonspecific complexes at low DNA/peptide ratios. On the other hand, the conjugate with the DNA binding domain of the zinc finger protein GAGA binds with high affinity ( $K_D \approx 12$  nM) and specificity to a composite AATTT·GAGA sequence containing both the bisbenzamidine and the TF consensus binding sites.

**Keywords:** molecular recognition • DNA binding • fluorescence • peptides • oligonucleotides

### Introduction

Gene expression is regulated by the action of specialized proteins called Transcription Factors (TFs) that bind to specific DNA regulatory sequences,<sup>1</sup> and hereafter promote the assembly of the multiprotein complex directly responsible for the initiation of transcription.<sup>2</sup> As a consequence of their fundamental role in the regulation of gene expression, it is not surprising that alterations in the activity of TFs are at the origin of many diseases, including cancer.<sup>3</sup> Therefore, the development of non-natural agents that can emulate or interfere with the dsDNA recognition of TFs remains a major goal in biological chemistry. These agents might have a great

impact in fundamental and applied biological research, and even lead to the development of gene-targeted therapies.<sup>4</sup> It is well established that although the DNA-binding of TF is mediated by relatively small peptide motifs, high-affinity DNA binding requires the full protein domain and, in many cases, the concerted action of multiple binding components.<sup>5</sup> Therefore, the preparation of small synthetic mimics of DNA-binding proteins is extremely challenging. Chemists have been able to engineer relatively small peptides capable of interacting with good affinity to specific dsDNA sequences.<sup>6</sup> Our own approach for the design of such minimized systems has relied on the conjugation of small peptide fragments derived from the DNA binding domains of transcription factors to tripyrrolic distamycin derivatives that bind to the minor groove of A/T-rich DNA sequences.<sup>7</sup> Although the strategy is effective, the synthesis of the tripyrrole moiety requires more than seven steps, and the handling of relatively unstable aminopyrrole synthetic intermediates. Therefore the use of other minor groove binding moieties that could be synthetically more accessible represents an important goal. In this context, we were attracted by propamidine and its derivatives, well-known A/T-rich DNA binders that display a much simpler structure than that of distamycin-related polyamides.<sup>8</sup> Herein we describe the synthesis of a hybrid between the basic region of GCN4 and propamidine, and demonstrate by electrophoresis mobility shift assays (EMSA) that this conjugate is capable of binding to designed DNA sites with good selectivity. More importantly, following our recent discovery of *aza*-

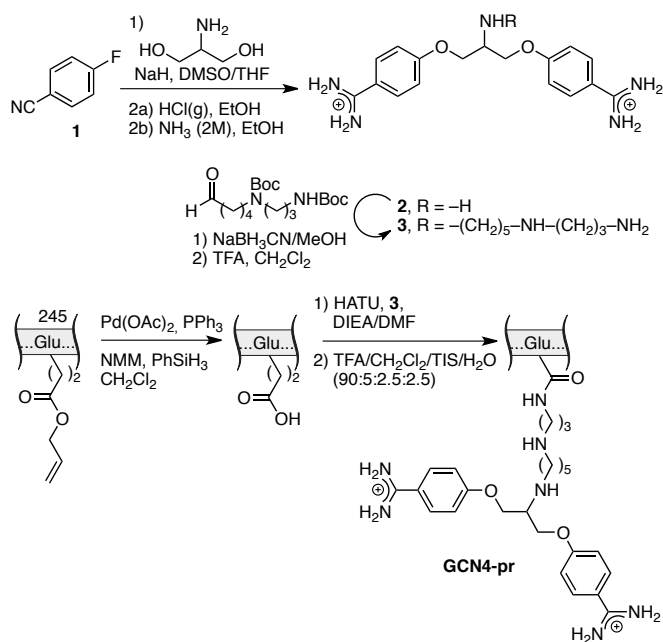
[a] Mateo I. Sánchez, Dr. Olalla Vázquez, Prof. M. Eugenio Vázquez\*, Prof. José L. Mascareñas\*  
Departamento de Química Orgánica and Centro Singular de Investigación en Química Biolóxica e Materiais Moleculares (CIQUS), and Unidad Asociada al CSIC  
Universidade de Santiago de Compostela  
15782 Santiago de Compostela, Spain  
Fax: (+34) 981 595 012  
E-mail: eugenio.vazquez@usc.es or joseluis.mascarenas@usc.es

Supporting information for this article is available on the WWW under <http://www.chemeurj.org/> or from the author.

bisbenzamidines as highly sensitive fluorogenic minor groove DNA binding agents,<sup>8a,9</sup> we also describe the synthesis of conjugates between one of these minor groove binders and fragments of the GCN4 and GAGA proteins, two representative members of the bZIP and zinc-finger families of transcription factors. In these conjugates, the *aza*-bisbenzamidine derivative not only promotes the interaction of the peptide with the DNA, but also acts as a sensing device that allows spectroscopic monitoring of the DNA binding process in solution.

## Results and Discussion

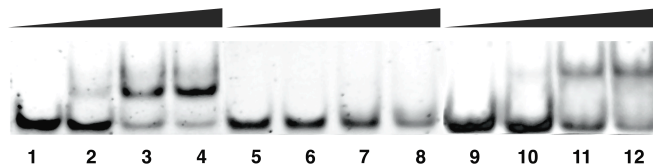
**Synthesis of the GCN4-propamidine hybrid (GCN4-pr).** The design of the GCN4-propamidine conjugate was based on the GCN4 peptide fragment comprising residues Asp226 to Gln248, which has been identified as the smallest peptide that retains specific DNA recognition properties as a dimer,<sup>10</sup> and was used in our previous studies with the tripyrrole conjugates.<sup>7c</sup> On the other hand, the propamidine derivative containing an appropriate linker for conjugation (**3**) was readily synthesized in three steps from commercially available *p*-fluorobenzonitrile (Scheme 1). The resulting propamidine-amino derivative was conjugated to the peptide fragment while it is still attached to the resin and selectively deprotected at glutamic acid 245, which replaces a native Arg residue in that position. After standard deprotection/cleavage and reverse-phase HPLC purification, the expected conjugate **GCN4-pr** was obtained in a good overall yield.



**Scheme 1.** *Top:* Synthesis of the propamidine derivative (**3**) for conjugation with the peptide. *Bottom:* Key steps in the synthesis of the conjugate between **3** and a minimal basic region of the GCN4 transcription factor (Asp226 to Gln252, see the supporting information; we only indicate the key Glu required for the coupling). The allyl removal and the coupling steps are done while the peptide is still attached to the solid support.

**DNA binding studies of GCN4-pr.** DNA binding experiments were carried out using standard non-denaturing EMSA protocols in polyacrylamide gels,<sup>11</sup> and using ds-oligonucleotides containing a composite binding site including the consensus recognition sequences of the GCN4 basic region and the propamidine binder, in

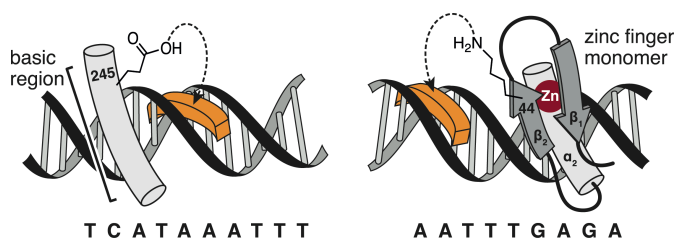
tandem (TCAT·AAATTT). As shown in Figure 1 (lanes 1-4), incubation of the target dsDNA oligo with the **GCN4-pr** hybrid leads to the formation of a relatively clean retention band, consistent with the formation of the desired complex in which the peptide derivative binds in major groove of its target sequence (TCAT), and the propamidine is inserted in the adjacent A/T-rich minor groove (AAATTT).



**Figure 1.** EMSA **GCN4-pr** DNA binding assays in Tris-HCl buffer. Lanes 1-4; target **API<sup>bs</sup>•AT** dsDNA (50 nM), lanes 2-4: 100, 300, 500 nM **GCN4-pr**; lanes 5-8: control **API<sup>bs</sup>•GC** dsDNA (50 nM), lanes 6-8: 100, 300, 500 nM **GCN4-pr**; lanes 9-12: control **GC•AT** DNA (50 nM), lanes 10-12: 100, 300, 500 nM **GCN4-pr**. dsDNA sequences (binding site in italics, only one strand is shown): **API<sup>bs</sup>•AT**: 5'-CGAACG *TCAT* AAATTT CCTC-3'; **API<sup>bs</sup>•GC**: 5'-ACGAACG *TCAT* GGCC GCCTC-3'; **GC•AT**: 5'-ACGAAG GCGGC *AAATT* CCTC-3'. Products were resolved by PAGE using a 10% non-denaturing polyacrylamide gel and 0.5X TBE buffer for 40 min at rt., and analyzed by staining with *SyBrGold* (Molecular Probes: 5  $\mu$ L in 50 mL of 1X TBE) for 10 min, and fluorescence visualization.

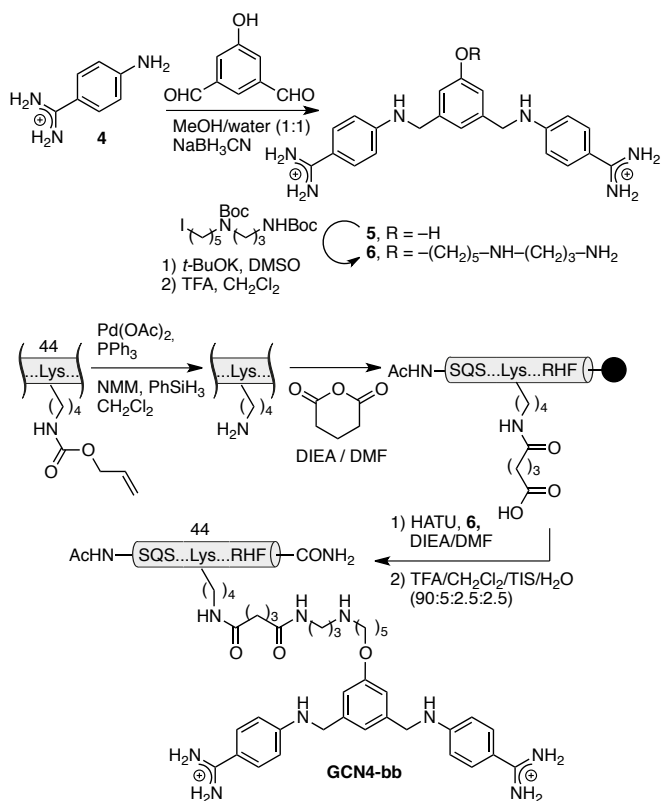
Consistent with that interaction model, circular dichroism experiments confirmed that the peptide chain folds into an  $\alpha$ -helix upon interaction with this DNA, as expected for a specific recognition (see the supporting information). A control oligonucleotide lacking the A/T-rich site fails to provide DNA-peptide complexes (Figure 1, lanes 5-8). On the other hand, incubation with oligonucleotides that do not exhibit a consensus peptide binding site (Figure 1, lanes 9-12), results in the formation of complexes of lower affinity that are more retained in the gel. These complexes most probably arise from non-specific DNA interactions involving the highly charged basic region, while the propamidine unit is inserted in the minor groove.<sup>7d,e</sup> Taken together, these results suggest that the propamidine is a viable alternative to the distamycin derivatives previously used as minor groove binders, although the performance in terms of affinity and selectivity seems poorer.

**Design and synthesis of bisbenzamidine conjugates of GCN4 and GAGA transcription factors.** At this stage of our research, we discovered that *aza*-bisbenzamidines, which feature a nitrogen instead of an oxygen atom at the *para*-position of the benzamidinium moiety, are weakly fluorescent in aqueous solvents but exhibit a great increase in the quantum yield of their fluorescence emission when bound to A/T-rich sites in the DNA.<sup>9</sup> We therefore envisioned the construction of a new type of hybrids featuring these chromophores as minor groove binding moieties, which might allow to gather information on the DNA interaction of the hybrids in solution by using spectroscopic methods. In particular, we decided to use as minor groove binding unit the phenol derivative **5**, which can be synthesized in a single step from commercial products, and contains a hydroxyl handle to engineer the connection to the peptide chains. With this compound at hand, we did not only approach the construction of the hybrid with the basic region of GCN4, but also a conjugate with a truncated module of the zinc finger protein GAGA (Ser28 to Phe58), a 30 residue peptide that by itself is incapable of showing significant DNA affinity.<sup>7a,12</sup>



**Figure 2.** *Left:* schematic representation of the simultaneous interaction of the GCN4 basic region peptide (gray cylinder) and the bisbenzamide (orange block), indicating the Glu245 residue selected for conjugation. *Right:* representation of a simultaneous interaction of the DNA binding domain of the GAGA transcription factor fragment in the major groove, and the bisbenzamide in the minor groove. In this case Lys44 is selected as an appropriate conjugation position. The models used to design the synthetic targets were based on the reported structures of the DNA binding complexes of the GCN4 and GAGA transcription factors,<sup>13</sup> and that of pentamidine,<sup>14</sup> as described in earlier reports of distamycin hybrids.<sup>7</sup>

The design of the GCN4-bisbenzamide hybrid (**GCN4-bb**) relies on the same Arg245 → Glu245 mutation of the GCN4 DNA binding domain fragment previously used for **GCN4-pr**. In this case, the conjugate was obtained by coupling the *aza*-bisbenzamide amine **6**, instead of the previously used propamidine-amine **3** (see the supporting information). Compound **6** was readily constructed by a reductive amination reaction between **4** and 5-hydroxyisophthalaldehyde, followed by alkylation of the phenolic oxygen in **5** with a Boc-protected derivative of *N*-(5-iodopentyl)propane-1,3-diamine, and a final TFA deprotection. Therefore the synthesis of bisbenzamide **6** is straightforward, and involves only two separate steps from commercial and inexpensive starting materials.

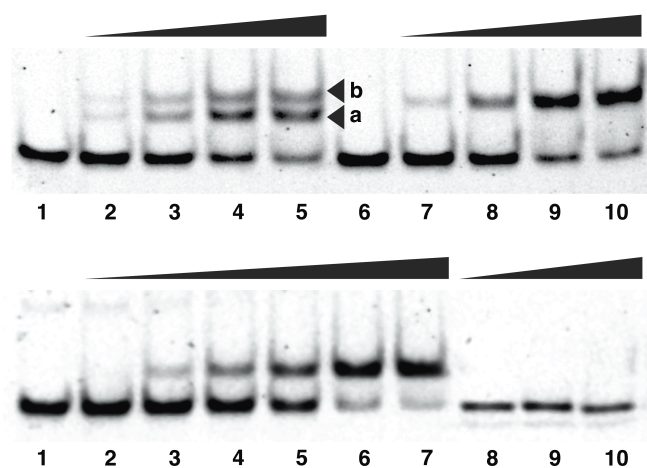


**Scheme 2.** *Top:* Synthesis of the bisbenzamide aromatic derivative **6** by reductive amination of 5-hydroxyisophthalaldehyde, introduction of the aminoalkyl side chain in the phenolic oxygen, and acidic deprotection. *Bottom:* Key steps in the synthesis of **GAGA-bb**. The black circle represents the solid resin used in the peptide synthesis.

The zinc finger conjugate was prepared from a peptide containing the truncated zinc finger unit of the GAGA transcription factor in which the Arg44 was replaced by Lys. This Lys44 was introduced with its side chain protected with an orthogonal alloc group that could be selectively removed in the solid phase (Scheme 2). The deprotected Lys side chain was then derivatized with glutaric anhydride, to increase the linker length and simultaneously install a carboxylic acid for attachment of the bisbenzamide. Once the carboxylic acid functionality was set in the peptide scaffold, the bisbenzamide **6** was coupled to the solid-phase linked peptide, using HATU as activating reagent. A standard cleavage and deprotection step involving treatment with a TFA cocktail, followed by reverse-phase HPLC purification gave the desired conjugate **GAGA-bb** in reasonable overall yield of  $\approx 19\%$ .

### DNA binding studies of GCN4-bb and GAGA-bb conjugates.

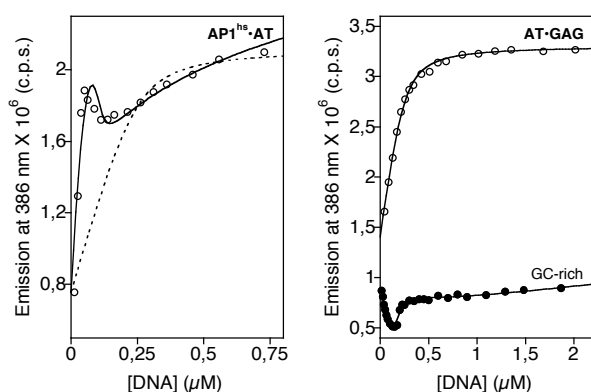
Having synthesized the desired peptide conjugates, we first assessed their DNA binding ability using standard EMSA experiments under non-denaturing conditions. Thus, incubation of the ds-oligonucleotide **API<sup>hs</sup>•AT**, containing the target composite binding site (TCAT•AAATT) at room temperature with increasing amounts of **GCN4-bb** led to the appearance of two new bands in addition to that corresponding to the free DNA (Figure 3 top, lanes 1-5). The faster, more intense band, is consistent with the expected compact hybrid-DNA complex (band a), while the slower migrating band could result from a complex in which the bisbenzamide unit is inserted into the minor groove of the A/T-rich region, while the extended peptide is making non-specific contacts with the DNA (Figure 3 top, band b). This assignment is supported by the observation of a similarly retarded band when **GCN4-bb** is incubated with an oligonucleotide featuring a mutated peptide-binding site (**API<sup>m</sup>•AT**) (Figure 3 top, lanes 6-10).<sup>7e</sup> As expected, incubation of **GCN4-bb** with a G/C-rich or the **API<sup>hs</sup>•GC** control oligonucleotide did not show any new bands (see the supporting information).



**Figure 3.** *Top:* EMSA assays of DNA recognition by **GCN4-bb**. Lanes 1-5: target **API<sup>hs</sup>•AT** dsDNA (50 nM). Lanes 2-5: 100, 300, 500, 800 nM **GCN4-bb**; lanes 6-10: control **API<sup>m</sup>•AT** dsDNA (50 nM). Lanes 7-10: 100, 300, 500, 800 nM **GCN4-bb**. *Bottom:* EMSA assays of DNA recognition by **GAGA-bb**. Lanes 1-7: target dsDNA **AT•GAG** (50 nM). Lanes 2-7: 100, 200, 300, 400, 500, 600 nM **GAGA-bb**. Lanes 8-10: control **AT<sup>m</sup>•GAG** dsDNA (50 nM); lanes 9-10: 300, 500 nM **GAGA-bb**. dsDNA sequences (binding site in italics): **API<sup>hs</sup>•AT**: 5'-CGAACG TCAT AAATTT CCTC-3'; **API<sup>m</sup>•AT**: 5'-CGAACG TGCT AAATTT CCTC-3'; **AT•GAG**: 5'-GACGG AATTT GAGAG CGTCG-3'; **AT<sup>m</sup>•GAG**: 5'-GACC GGGCC GAGAG TACGT-3'. EMSA assays of DNA recognition by **GCN4-bb** were done in the same conditions as for **GCN4-pr**; in the case of **GAGA-bb** the buffer used was slightly different (see the supporting information for details).

In the case of the conjugate **GAGA-bb**, we observed a single retarded band when treated with the **AT•GAG** oligonucleotide containing the composite binding site. This is consistent with the formation of the expected specific complex between the hybrid and the DNA (Figure 3 bottom, lanes 1-7). No retarded bands were observed when this same hybrid was incubated with a control dsDNA (**AT<sup>m</sup>•GAG**), lacking the minor groove recognition sequence (Figure 2 bottom, Lanes 8-10); likewise, incubation with a DNA containing the consensus minor groove binding sequence but a mutated peptide binding site failed to show any clear binding by EMSA (see the supporting information). The different behavior of the two conjugates (**GCN4-bb** and **GAGA-bb**) could be explained by the presence of a much higher number of basic residues in the case of the GCN4 hybrid, which promote electrostatic interactions with the negatively charged dsDNA, and hence the production of competitive less-specific DNA complexes.

**Steady-state fluorescence studies.** The presence of the fluorogenic minor groove *aza*-bisbenzamidine in the new conjugates allows monitoring of the DNA binding process by fluorescent spectroscopy, and hence obtaining complementary information that could not be derived from the EMSA studies. Thus, incubation of a 0.25  $\mu\text{M}$  solution of **GCN4-bb** with increasing concentrations of a double stranded oligonucleotide (**API<sup>hs</sup>•AT**) containing the composite consensus binding site (TCAT•AAATT) led to a notable and progressive increase in the fluorescence emission ( $\lambda_{\text{exc}}$  329 nm). Curiously, the resulting titration curve exhibited an atypical shape, consisting on a spike in the emission intensity at low DNA/hybrid ratios, followed by a gradual increase as in a conventional saturation curve, at higher DNA concentrations (Figure 4, left). This type of multiphasic profile has been previously observed in reverse titrations for the interaction of cationic proteins with DNA,<sup>15</sup> and could be explained in terms of the competitive formation of multiple non-specific complexes with the DNA oligomers at low DNA/ligand ratios. Upon increasing the DNA concentration, the 1:1 binding mode becomes predominant, and the curve follows a more familiar titration profile, consistent with the results observed in the gel shift experiments (note that the EMSA experiments were performed as forward titrations: constant DNA concentration and addition of the peptide).



**Figure 4.** Left: Fluorescence emission titration of a 0.25  $\mu\text{M}$  solution of **GCN4-bb** with a dsDNA containing the target sequence (**API<sup>hs</sup>•AT**) in Tris-HCl buffer 20 mM pH 7.5, 100 mM NaCl. Solid line represents a nonlinear fit to a mixed model considering a 1:1 complex and competitive species resulting from non-specific binding; dashed line represents the best fit to a simple 1:1 binding mode, discarding points 2 to 9 in the titration. Right: Titration of a 0.5  $\mu\text{M}$  solution of **GAGA-bb** with a DNA containing the composite recognition site (**AT•GAG**) in Tris-HCl buffer 20 mM pH 7.5, 100 mM NaCl. Solid line represents the nonlinear fit to a 1:1 binding mode for **AT•GAG**, and a mixed mode including competitive nonspecific binding for the GC-rich oligonucleotide.

In contrast with the relatively complex binding profile displayed by the **GCN4-bb** hybrid, fluorescence titration of a 0.50  $\mu\text{M}$  solution of the zinc-finger conjugate **GAGA-bb** with its target DNA (**AT•GAG**), containing the composite binding sequence (AATT•GAG), could be fitted to a simple 1:1 binding mode with an apparent  $K_D$  of  $\approx 12$  nM, (Figure 2, right).<sup>16</sup> Titration of this same hybrid with a G/C-rich double stranded DNA, resulted in a multiphasic profile with a very small increase in the emission intensity from the minor groove binder, in line with the expected low affinity for this oligonucleotide sequence. Control experiments with the bisbenzamidine **6**, which also contains a charged side chain with two amine groups, are consistent with this analysis, demonstrating low affinity for G/C-rich oligonucleotides and high affinity binding for A/T-rich sites (see the supporting information).

Taken together, these results demonstrate that bisbenzamidines are synthetically straightforward minor groove-binding handles for the construction of functional conjugates with TF peptide fragments, which, by themselves, are not capable of binding to the DNA. In addition to the thermodynamic stabilization of otherwise unstable complexes, these anchors display marked fluorogenicity that allowed us to observe molecular associations that are not evident in regular gel shift studies.

## Conclusion

In summary, we have demonstrated that conjugation of fragments of transcription factors to bisbenzamidines allows the selective DNA recognition of relatively long DNA sequences, containing composite sites of the original TF target sequence and A/T-rich sites targeted by the minor groove binder. Moreover, the fluorogenic nature of the minor groove binder allows to monitoring the DNA recognition process by fluorescence spectroscopy. Therefore, the combination of standard EMSA analysis and fluorescence titrations provides a more exact account of the processes taking place when these hybrids interact with the DNA. Given the ready accessibility of the minor groove binders, its optical sensing properties and the nanomolar affinities and selectivity exhibited by the conjugates, we expect a great future for this DNA binding strategy.

## Experimental Section

**General synthetic procedures:** All reagents were from commercial sources: DMF and TFA were purchased from *Scharlau*,  $\text{CH}_2\text{Cl}_2$  from *Panreac*,  $\text{CH}_3\text{CN}$  from *Merck*. The rest of reagents were from *Sigma-Aldrich*. When indicated, reactions were monitored by analytical RP-HPLC with an *Agilent 1100* series LC/MS with an *Eclipse XDBC18* ( $4.6 \times 150$  mm,  $5 \mu\text{m}$ ) analytical column. Standard conditions for analytical RP-HPLC consisted on an isocratic regime during the first 5 min, followed by a linear gradient from 5% to 75% of solvent B for 30 min at a flow rate of 1 mL/min (A: water with 0.1% TFA, B:  $\text{CH}_3\text{CN}$  with 0.1% TFA). Compounds were detected by UV absorption. Amidine derivatives **2**, **3** and **5** were purified on a Büchi *Sepracore* preparative system consisting on a pump manager C-615 with two pump modules C-605 for binary solvent gradients, a fraction collector C-660, and UV Photometer C-635. Purification was made using reverse phase linear gradients of MeOH/H<sub>2</sub>O 0.1% TFA in 30 min with a flow rate of 30 mL/min, using a pre-packed preparative cartridge ( $150 \times 40$  mm) with reverse phase RP18 silica gel (*Büchi* order #54863). The fractions containing the products were freeze-dried, and their identity confirmed by ESI-MS and NMR. Compounds were isolated as TFA salts. The peptide conjugates **GCN4-pr**, **GCN4-bb** and **GAGA-bb** were purified by analytical RP-HPLC, using an *Eclipse XDBC8* ( $4.6 \times 150$  mm,  $5 \mu\text{m}$ ) analytical column, following standard HPLC purification conditions. The sequences of the peptides are: Asp226 to Gln248 of GCN4: DPAALKRANTEAARRSRAE<sub>245</sub>KLQ; Ser28 to Phe58 of GAGA: SQSEQPATPCICYAVIK<sub>44</sub>QSRNLRRLHLELRHF. *N*-(3-((*tert*-butoxy)carbonyl)(5-oxopentyl)amino)propyl)carbamate **4**-[2-({3-(aminopropyl)amino}pentyl)amino]-3-(4-carbamimidoylphenoxy)propoxy] benzene carboximidamide (**3**). The amine **2<sup>nb</sup>** (90 mg, 0.13 mmol), and *tert*-butyl (3-((*tert*-butoxycarbonyl)amino)propyl)(5-oxopentyl)carbamate (39 mg, 0.10 mmol) were

dissolved in MeOH (1.25 mL) in a round bottom flask. NaBH<sub>3</sub>CN (8 mg, 0.12 mmol) was added over the solution, and the mixture was stirred for 5 h at rt. The solvent was removed under reduced pressure, and the residue was purified by preparative reverse-phase chromatography. The fractions containing the desired product were collected, concentrated and freeze-dried to obtain the desired conjugate. The isolated Boc-protected compound was dissolved in CH<sub>2</sub>Cl<sub>2</sub> (1 mL) and cooled to 0 °C. TFA (1 mL) was added dropwise, and the resulting solution was stirred at 0 °C for 1 h, and at rt for other 2 h. The solvent was removed under reduced pressure, and the residual TFA was removed by co-distillation with CH<sub>2</sub>Cl<sub>2</sub>. The freeze-dried white powder was identified as the desired product (**3**) (65.0 mg, 0.08 mmol, 65% overall yield for the two-step process). <sup>1</sup>H NMR (400 MHz, DMSO-*d*<sub>6</sub> δ): 1.38 (m, 2H), 1.62 (m, 2H), 1.71 (m, 2H), 1.91 (m, 2H), 2.88 (m, 4H), 2.99 (m, 2H), 3.16 (m, 2H), 4.08 (m, 1H), 4.54 (d, *J* = 4.5 Hz, 4H), 7.21 (d, *J* = 8.9 Hz, 4H), 7.87 (d, *J* = 8.8 Hz, 4H), 8.03 (s, 2H, NH), 8.89 (s, 1H, NH), 9.23 (s, 4H), 9.34 (s, 4H), 9.53 (s, 1H, NH). <sup>13</sup>C NMR (400 MHz, DMSO-*d*<sub>6</sub> δ): 23.5 (CH<sub>2</sub>), 24.2 (CH<sub>2</sub>), 25.4 (CH<sub>2</sub>), 25.6 (CH<sub>2</sub>), 36.6 (CH<sub>2</sub>), 44.3 (CH<sub>2</sub>), 46.2 (CH<sub>2</sub>), 47.0 (CH<sub>2</sub>), 55.7 (CH), 65.3 (CH<sub>2</sub>), 115.4 (CH), 121.0 (C), 139.6 (CH), 159.2 (q, C TFA), 162.2 (C), 165.2 (C). ESI<sup>+</sup>-MS: [M+H]<sup>+</sup> calcd. for C<sub>25</sub>H<sub>40</sub>N<sub>7</sub>O<sub>2</sub> = 470.3238 found 470.3236.

**Synthesis of 4-((3-((5-((3-aminopropyl)amino)pentyl)oxy)-5-((4-carbamimidoyl phenyl)amino)methyl)phenyl)methyl)amino)benzene-1-carboximidamide (**6**).** Potassium *tert*-butoxide (36.4 mg, 0.324 mmol, 4 equiv) was added to a solution of the bis-amino benzamide **5**<sup>8a</sup> (50 mg, 0.081 mmol) in dry DMSO (1.62 mL). After 30 min, *tert*-butyl-3-[[*tert*-butoxycarbonyl]amino]propyl(5-iodopentyl)carbamate (46.3 mg, 0.097 mol, 1.2 equiv) was added in portions. The reaction mixture was stirred under Ar at rt for 2 h. The crude reaction was directly purified by preparative reverse-phase chromatography (Büchi Sepacore) (gradient: 15% B, 5 min; 15% → 95% B, 30 min). The combined fractions were concentrated and freeze-dried.

The isolated Boc-protected amine was dissolved in CH<sub>2</sub>Cl<sub>2</sub> (2 mL) and cooled to 0 °C. TFA (2 mL) was added dropwise and the resulting solution was stirred at 0 °C for 1 h and at room temperature for other 2 h. The solvent was removed under reduced pressure, and the residual TFA was removed by co-distillation with CH<sub>2</sub>Cl<sub>2</sub>. The residue was purified by preparative reverse-phase chromatography (Büchi Sepacore) (gradient: 0% B, 5 min; 0% → 50% B, 30 min.). The freeze-dried solid was identified as the desired product (**6**) as a white powder (56.8 mg, 0.072 mmol, 89% overall yield for the two-step process). <sup>1</sup>H NMR (400 MHz, MeOD-*d*<sub>4</sub> δ): 1.32 (m, 2H), 1.55 (m, 2H), 1.77 (m, 2H), 2.07 (m, 2H), 3.05 (m, 2H), 3.11 (m, 2H), 3.94 (t, *J* = 6.1 Hz, 2H), 4.36 (s, 4H), 6.68 (d, *J* = 8.9 Hz, 4H), 6.79 (s, 2H), 6.94 (s, 1H), 7.57 (d, *J* = 8.9 Hz, 4H). <sup>13</sup>C NMR (400 MHz, MeOD-*d*<sub>4</sub> δ): 24.2 (CH<sub>2</sub>), 25.3 (CH<sub>2</sub>), 27.0 (CH<sub>2</sub>), 29.9 (CH<sub>2</sub>), 37.8 (CH<sub>2</sub>), 45.8 (CH<sub>2</sub>), 47.5 (CH<sub>2</sub>), 49.0 (CH<sub>2</sub>), 68.5 (CH<sub>2</sub>), 112.8 (CH), 113.1 (CH), 114.3 (C), 118.9 (CH), 130.6 (CH), 142.3 (C), 155.4 (C), 161.2 (C), 167.1 (C). ESI<sup>+</sup>-MS: [M+H]<sup>+</sup> calc. for C<sub>30</sub>H<sub>43</sub>N<sub>8</sub>O = 531.3554 found 531.3559.

**Peptide synthesis.** All peptide synthesis reagents and amino acid derivatives were from *GL Biochem* (Shanghai) and *Novabiochem*; amino acids were used as protected Fmoc amino acids with the standard side chain protecting scheme, except for the orthogonally protected Fmoc-Lys(alloc)-OH and Fmoc-Glu(alloc)-OH, which were purchased from *Bachem*. C-terminal amide peptides were synthesized on a 0.1 mmol scale using a Fmoc-PAL-PEG-PS resin from *Applied Biosystems*. All solvents were dry and synthesis grade. Peptides were synthesized using an automatic *PS3* peptide synthesizer from *Protein Technologies*. Amino acids were coupled in 4-fold excess using HBTU/HOBt as activating agent. Each amino acid was incubated with the HBTU/HOBt mixture for 30 seconds in DMF before being added onto the resin. Peptide bond-forming couplings were conducted for 30 min to 45 min.

**Deprotection of the temporal Fmoc protecting group.** Deprotection of the temporal Fmoc was performed by treating the resin with 20% piperidine in DMF solution for 10 min. ***N*-terminal acetylation.** The *N*-acetylation coupling was performed immediately after Fmoc deprotection by treating the resin with Ac<sub>2</sub>O/(DIEA/DMF 0.195 M) 2:1. ***Allyl and Alloc removal.*** The selective deprotection of the glutamic acid residues was carried out as following: approx. 3.5 μmol (25 mg of resin) of the peptide attached to the solid support were treated with PPh<sub>3</sub> (1.5 equiv), *N*-methyl morpholine (10 equiv) and phenylsilane (10 equiv) in CH<sub>2</sub>Cl<sub>2</sub> (400 μL) at rt for 15 min. Subsequently Pd(OAc)<sub>2</sub> (0.3 equiv) was added and the resulting mixture was stirred overnight. The selective deprotection of the lysine was carried out treating the peptide attached to the resin with Pd(PPh<sub>3</sub>)<sub>4</sub> (1 eq), morpholine (190 eq) in H<sub>2</sub>O/CH<sub>2</sub>Cl<sub>2</sub> 2% (400 μL). The resin was filtered and washed with DMF (2 × 1.5 mL × 2 min), diethyl dithiocarbamate (DEDTC, 25 mg in 5 mL of DMF, 2 × 1.5 mL × 5 min), DMF (2 × 1.5 mL × 2 min) and CH<sub>2</sub>Cl<sub>2</sub> (2 × 5 mL × 2 min). ***Cleavage/deprotection step.*** The resin containing the desired peptide is washed with CH<sub>2</sub>Cl<sub>2</sub> (2 × 1.5 mL × 2 min) and filtered; A mixture of 50 μL of CH<sub>2</sub>Cl<sub>2</sub>, 25 μL of H<sub>2</sub>O, 25 μL of triisopropylsilane (TIS), and 900 TFA μL for every 40 mg of resin is added onto the resin, and the mixture is shaken for 2.5 h. The resin is filtered and the liquid is added over 10 volumes of cold Et<sub>2</sub>O. After 5 min the suspension is centrifuged and the resulting white residue is washed twice with Et<sub>2</sub>O and dried under an Ar current.

**Synthesis of GCN4-pr.** The resin-bound selectively deprotected GCN4 peptide (25 mg, approx. 3.5 μmol) was shaken for 1 h in DMF and then filtered. A mixture of HATU (3 mg, 7.90 × 10<sup>-3</sup> μmol) in DMF (170 μL), and DIEA (28 μL, of a 0.5 M solution in DMF) was added to the resin, shaken for 5 min and the resin washed with DMF (2 × 1.5 mL × 2 min). Finally, a mixture of functionalized propamidine **3** (≈ 5 mg, ≈ 7 μmol) in DIEA (56 μL, of a 0.5 M solution in DMF) was added to the resin and shaken for 2 h. The resin was washed with DMF (2 × 1.5 mL × 2 min) and filtered. The peptide deprotection and resin cleavage step was carried out following standard conditions, and

the resulting crude mixture was purified by RP-HPLC obtaining the desired hybrid (3.0 mg, 1.74 μmol, overall yield considering the peptide synthesis ≈ 25%). Analytical Data of the purified products: MS: MALDI-TOF (M+H<sup>+</sup>): calcd. for C<sub>134</sub>H<sub>229</sub>N<sub>49</sub>O<sub>36</sub> = 3100.7, found 3101.2. Retention time = 22.0 min. Peptide GCN4-bb was synthesized following the same procedure but using the amine **6** for the coupling reaction (see the supporting information for experimental details).

**Synthesis of GAGA-bb.** The resin-bound GAGA peptide (28 mg, approx. 2.3 × μmol) was suspended in DMF (1 mL), shaken for 1 h and then filtered. A mixture of succinic anhydride (127 μL, 0.1 M in DMF, 18 μmol, 8 equiv), DIEA (36 μL, 0.5 M in DMF, 18 μmol, 8 equiv) and DMAP (28 μL, 0.2 M in DMF, 5.6 μmol, 2 equiv) was added, and the resulting resin suspension was shaken for 1 h at rt. The resin was then filtered and washed with DMF (3 × 0.6 mL). A mixture of HATU (172 μL, 52 mM in DMF, 9.1 μmol, 4 equiv) and DIEA (36 μL, 0.5 M in DMF, 18 μmol, 8 equiv) was added. The resulting mixture was shaken for 5 min and filtered. A solution of the bis-benzamide **6** (9.1 mg in 200 μL of DMF, 4 equiv.) and DIEA (36 μL, 0.5 M in DMF, 18 μmol, 8 equiv) was added. The mixture was shaken for 2.5 h after which the resin was filtered and washed with DMF (3 × 0.6 mL, 5 min) and Et<sub>2</sub>O (2 × 0.5 mL, 5 min). The cleavage step was performed following standard conditions, and the desired peptide conjugate purified by RP-HPLC. Analytical Data of the purified products: MS: MALDI-TOF (M+H<sup>+</sup>): calcd. for C<sub>195</sub>H<sub>304</sub>N<sub>60</sub>O<sub>49</sub>S<sub>2</sub> = 4337.0, found 4337.2. Retention time = 21.4 min.

**Fluorescence Spectroscopy.** Measurements were made with a *Jobin-Yvon Fluoromax-3*, (DataMax 2.20) coupled to a *Wavelength Electronics LFI-3751* temperature controller, using the following settings: increment: 1.0 nm; integration time: 0.2 s; excitation slit width: 3.0 nm; emission slit width: 6.0 nm; excitation wavelength 329 nm. The emission spectra were acquired from 345 to 500 nm at 20 °C. All titrations were made following the same procedure: to 1 mL of either compound (50 nM in the case of **6**, 250 nM in the case of GCN4-bb or 500 nM for GAGA-bb) in Tris-HCl buffer 20 mM pH 7.5, 100 mM NaCl, aliquots of ≈ 200 μM stock solutions of the respective dsDNAs were successively added, and the fluorescence spectrum was recorded after each addition. In the case of the titration of GAGA-bb before adding the first aliquot of dsDNA, 5 equiv of ZnCl<sub>2</sub> were added to make the complex.

**Electrophoretic mobility shift assays (EMSA)** were performed with a *BIO-RAS Mini Protean* gel system, powered by an electrophoresis power supplies *PowerPac Basic* model, maximum power 150 V, frequency 50.60 Hz at 140 V (constant V).

## Acknowledgements

We thank the financial support provided by the Spanish grants SAF2010-20822-C02, CTQ2009-14431/BQU, CTQ2012-31341, CSD2007-00006 *Consolider Ingenio 2010*, the *Xunta de Galicia* GRC2010/12, INCITE09 209 084PR, PGIDIT08CSA-047209PR. M.I.S. thanks the ministry for his PhD fellowship.

- 1 C. W. Garvie, C. Wolberger, *Mol. Cell* **2001**, *8*, 937–946.
- 2 a) D. S. Latchman, *Eukaryotic Transcription Factors*, Elsevier, London, 2004; b) N. M. Luscombe, S. E. Austin, H. M. Berman, J. M. Thornton, *Genome Biol.* **2000**, *1*, 1–10.
- 3 a) J. E. Darnell, *Nature Rev. Cancer* **2002**, *2*, 740–749; b) A. T. Look, *Science* **1997**, *278*, 1059–1064; c) P. Brennan, R. Donev, S. Hewamana, *Mol. Biosci.* **2008**, *4*, 909–919; d) P. K. Vogt, *Oncogene* **2001**, *20*, 2365–2377.
- 4 a) P. P. Pandolfi, *Oncogene* **2001**, *20*, 3116–3127; b) R. Pollock, M. Giel, K. Linher, T. Clackson, *Nature Biotechnol.* **2002**, *20*, 729–733; c) C. Denison, T. Kodadek, *Chem. Biol.* **1998**, *5*, R129–45; d) R. Yaghai, G. Cutting, *Mol Ther* **2002**, *5*, 685–694; e) A. K. Mapp, *Org. Biomol. Chem.* **2003**, *1*, 2217–2220.
- 5 a) H. C. Nelson, *Curr. Opin. Genet. Dev.* **1995**, *5*, 180–189; b) R. Moretti, A. Z. Ansari, *Biochimie* **2008**, *90*, 1015–1025; c) D. J. Segal, C. F. Barbas, *Curr. Opin. Chem. Biol.* **2000**, *4*, 34–39.
- 6 a) N. J. Zondlo, A. Schepartz, *J. Am. Chem. Soc.* **1999**, *121*, 6938–6939; b) R. L. Stafford, H.-D. Arndt, M. L. Brezinski, A. Z. Ansari, P. B. Dervan, *J. Am. Chem. Soc.* **2007**, *129*, 2660–2668; c) M. E. Vazquez, A. M. Caamaño, J. L. Mascareñas, *Chem. Soc. Rev.* **2003**, *32*, 338–349; d) E. Pazos, J. Mosquera, M. E. Vazquez, J. L. Mascareñas, *ChemBioChem* **2011**, *12*, 1958–1973.
- 7 a) O. Vazquez, M. E. Vazquez, J. B. Blanco-Canosa, L. Castedo, J. L. Mascareñas, *Angew. Chem. Int. Ed.* **2007**, *46*, 6886–6890; b) J. B. Blanco, V. I. Dodero, M. E. Vázquez, M. Mosquera, L. Castedo, J. L. Mascareñas, *Angew. Chem. Int. Ed.* **2006**, *45*, 8210–8214; c) J. B. Blanco, M. E. Vázquez, L. Castedo, J. L. Mascareñas, *ChemBioChem* **2005**, *6*, 2173–2176; d) J. B. Blanco, O. Vazquez, J. Martínez-Costas, L. Castedo, J. L. Mascareñas, *Chem. Eur. J.* **2005**, *11*, 4171–4178; e) J. B. Blanco, M. E. Vázquez, J. Martínez-Costas, L. Castedo, J. L. Mascareñas, *Chem. Biol.* **2003**, *10*, 713–722; f) M. E. Vázquez, A. M. Caamaño, J. Martínez-Costas, L. Castedo, J. L. Mascareñas, *Angew. Chem. Int. Ed.* **2001**, *40*, 4723–4725; g) M. E. Vazquez, A. M. Caamaño, L. Castedo, D. Gramberg, J. L. Mascareñas, *Tetrahedron Lett.* **1999**, *40*, 3625–3628; h) M. E.

- Vazquez, A. M. Caamaño, L. Castedo, J. L. Mascareñas, *Tetrahedron Lett.* **1999**, *40*, 3621–3624.
- 8 a) M. I. Sánchez, O. Vazquez, J. Martínez-Costas, M. E. Vazquez, J. L. Mascareñas, *Chem. Sci.* **2012**, *3*, 2383–2387; b) I. Haq, *Arch. Biochem. Biophys.* **2002**, *403*, 1–15; c) C. M. Nunn, S. Neidle, *J. Med. Chem.* **1995**, *38*, 2317–2325.
- 9 a) O. Vázquez, M. I. Sánchez, J. L. Mascareñas, M. E. Vázquez, *Chem. Commun.* **2010**, *46*, 5518–5520; b) O. Vázquez, M. I. Sánchez, J. Martínez-Costas, M. E. Vázquez, J. L. Mascareñas, *Org. Lett.* **2010**, *12*, 216–219; c) M. I. Sánchez, O. Vázquez, M. E. Vázquez, J. L. Mascareñas *Chem. Commun.* **2011**, *47*, 11107–11109
- 10 a) R. V. Talanian, C. J. McKnight, R. Rutkowski, P. S. Kim, *Biochemistry* **1992**, *31*, 6871–6875; b) R. V. Talanian, C. J. McKnight, P. S. Kim, *Science* **1990**, *249*, 769–771.
- 11 a) L. M. Hellman, M. G. Fried, *Nature Prot.* **2007**, *2*, 1849–1861; b) D. Lane, P. Prentki, M. Chandler, *Microbiol Rev* **1992**, *56*, 509–528.
- 12 Residue numbering taken from the corresponding pdb structures.
- 13 a) P. König, T. J. Richmond, *J. Mol. Biol.* **1993**, *233*, 139–154; b) T. E. Ellenberger, C. J. Brandl, K. Struhl, S. C. Harrison, *Cell* **1992**, *71*, 1223–1237; c) J. G. Omichinski, P. V. Pedone, G. Felsenfeld, A. M. Gronenborn, G. M. Clore, *Nature Struct. Mol. Biol.* **1997**, *4*, 122–132. pdb codes 1YSA and 1YUI for the GCN4-DNA and GAGA-DNA complexes, respectively
- 14 K. J. Edwards, T. C. Jenkins, S. Neidle, *Biochemistry* **1992**, *31*, 7104–7109. (pdb code: 1D64)
- 15 a) J. J. Hollenbeck, M. G. Oakley, *Biochemistry* **2000**, *39*, 6380–6389; b) Portela, F. Albericio, R. Eriřa, L. Castedo, J. L. Mascareñas, *ChemBioChem* **2007**, *8*, 1110–1114; c) A. V. Fedorova, I.-S. Chan, J. A. Shin, *BBA-Proteins* **2006**, *1764*, 1252–1259. This complex profile was analyzed with numerical methods as implemented in the *DynaFit* program. P. Kuzmic, *Anal. Biochem.* **1996**, *237*, 260–273.
- 16 Accurate determination of  $K_D$  values requires the titrations to be performed at concentrations below the  $K_D$ . However, the relatively low emission intensity of the benzimidazole fluorophore has not allowed us to perform the titrations at the required low nM concentration. Therefore the reported value should be taken with caution. a) M. R. Eftink, *Methods Enzymol.* **1997**, *278*, 221–257; b) V. J. LiCata, A. J. Wowor, *Methods Cell Biol.* **2008**, *84*, 243–262.

---

Received: ((will be filled in by the editorial staff))  
 Revised: ((will be filled in by the editorial staff))  
 Published online: ((will be filled in by the editorial staff))



## Entry for the Table of Contents

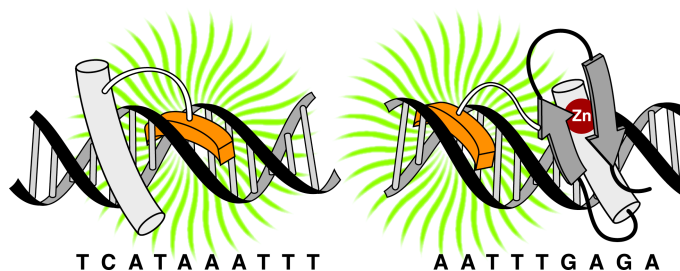
---

### Sequence-Selective DNA Recognition with peptide-bisbenzamidines

---

Mateo I. Sánchez, Olalla Vázquez, M. Eugenio Vázquez,\* and José L. Mascareñas\* ..... Page – Page

### Sequence-Selective DNA Recognition with peptide-bisbenzamidines



Conjugation of fragments of bZIP and zinc finger transcription factors to bisbenzamidines allows the selective DNA recognition of composite DNA sequences containing the original target sequence of the peptide fragment in tandem with A/T-rich sequences targeted by the minor groove binder unit. The fluorogenic *aza*-bisbenzamidine is exploited for the spectroscopic characterization of the DNA recognition process.

