

Chapter 2

Next Generation Sequencing Studies Guide the Design of Pyrrole-Imidazole Polyamides with Improved Binding Specificity by the Addition of β -Alanine

Abstract The identification of binding sites for small molecules in the genome space is important for various applications. Previously, we demonstrated rapid transcriptional activation by our small molecule SAHA-PIPs. However, it was not clear whether the strong biological effects exerted by SAHA-PIP were due to its binding specificity. Here, we used high-throughput sequencing (Bind-n-seq) to identify the binding specificity of SAHA-PIPs. Firstly sequence specificity bias was determined with SAHA-PIPs (**3** and **4**), which showed enhanced 6-bp sequence-specific binding compared with hairpin PIPs (**1** and **2**). This finding allowed us to investigate the role of β -alanine that links SAHA with PIP, which led to the design of $\beta\beta$ -PIPs (**5** and **6**). $\beta\beta$ -PIPs showed enhanced binding specificity. Overall, we demonstrated the importance of β -moieties for the binding specificity of PIPs, and the utilization of cost-effective high-throughput screening of these small molecules to the minor groove.

Keywords DNA recognition • High throughput screening • Bind-n-seq • Next generation sequencing • Synthetic biology • HDAC inhibitor

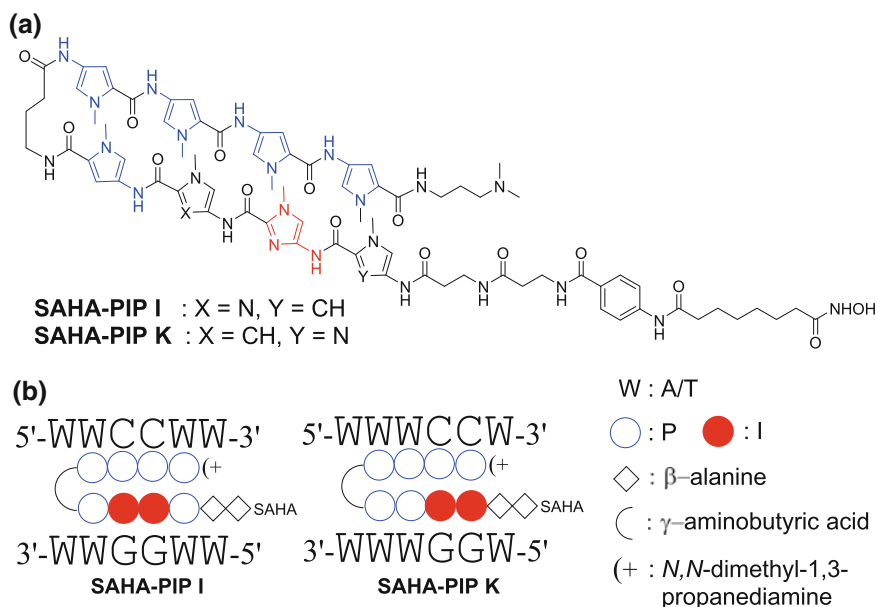
2.1 Introduction

Dysregulation of gene expression is a major cause of many diseases. Global gene expression is governed by transcriptional regulation, and genetic and epigenetic machinery plays major roles in this. Alteration of the epigenetic mechanism can provide targeted gene regulation and cellular reprogramming, and could enable the development of artificial genetic switches, governed by DNA-binding small molecules, which could bring back otherwise irredeemable dysfunctions connected with defective transcriptional/epigenetic machinery [1]. Hairpin *N*-methylpyrrole (P)–*N*-methylimidazole (I) polyamides (PIPs) are a class of small molecule that can bind in the minor groove of DNA [2] and can recognize each of the four Watson–Crick base pairs [3, 4]. When arranged side-by-side, I and P discriminate C·G from G·C base pairing [2], whereas the P and P pair can recognize both A·T and T·A [5].

Chromatin-modifying enzymes can alter transcriptional regulation randomly by epigenome modification [6]. Previously, we linked PIP to an epigenetically active histone deacetylase (HDAC) inhibitor (SAHA) to generate SAHA-PIP [7–10] and demonstrated the effect of SAHA-PIPs in mouse fibroblasts to induce differential activation of pluripotent stem cell-associated genes [8–10]. This study demonstrated the efficiency of SAHA-PIP in intervening in transcriptional activation and showed that the compound may activate the silent gene system in somatic cells. Our recent report on the microarray analysis of a SAHA-PIP library revealed that such compounds can trigger a unique set of gene clusters in human dermal fibroblast (HDF) cells [11]. However, whether the strong biological effects exhibited by SAHA-PIP were due to its binding specificity in a broad context was unclear.

Recently, the binding sites of hairpin PIPs in DNA at the genome level were reported [9]. In this approach, termed Bind-n-seq, PIPs were tagged with biotin and the bound DNA was analyzed by high-throughput sequencing methods [12]. Although the PIPs bound to the DNA with canonical pairing rules as predicted, some PIPs were identified with the reverse binding orientation [12, 13].

In this study, we identify the high-affinity binding sites of the previously reported germ cell gene activating SAHA-PIP K [14] and its structurally similar counterpart SAHA-PIP I [11] that activate an entirely different set of developmental genes (Scheme 2.1). By using the Bind-n-seq approach, we could identify the binding specificity of SAHA-PIP I (3) and SAHA-PIP K (4) in a broad context of



Scheme 2.1 **a** Chemical structures, and **b** DNA sequence-specific binding models of SAHA-PIP I and K

Table 2.1 Relative enrichment of 6-base pair “k”mers enriched by **3** in two separate binding and enrichment reactions (only top 30 sequences were shown)

Rank	Sequences	Enrichment ratio (BNS-1)	Enrichment ratio (BNS-1)
1	TACCAA	3.83	4.047
2	ATTACC	3.63	3.89
3	GGTACC	3.231	3.393
4	CTACCA	3.155	3.2
5	GTACCA	3.109	3.166
6	TGGTAA	3.006	3.085
7	AACCAA	2.959	3.169
8	ATGGTA	2.945	2.812
9	AATACC	2.853	3.083
10	ATACCA	2.846	2.918
11	GGTAGA	2.831	2.863
12	TTCCAA	2.795	2.931
13	CTTCCA	2.721	2.827
14	ACCAAT	2.651	2.779
15	GTTCCA	2.634	2.616
16	GGTACA	2.617	2.574
17	AGGTAG	2.559	2.675
18	ACCATC	2.55	2.331
19	CCAACC	2.439	2.503
20	GGAAGA	2.416	2.561
21	ATAACC	2.383	2.59
22	GGAACC	2.359	2.468
23	CCATCC	2.352	2.024
24	AACCAT	2.316	2.369
25	ACCAAG	2.313	2.46
26	GAACCA	2.308	2.38
27	ACTTCC	2.303	2.456
28	ACTACC	2.285	2.442
29	ACCAAC	2.279	2.547
30	AGGTAC	2.276	2.426

Its corresponding graph was shown in Fig. 2.3

oligo libraries. Both compounds showed high-affinity binding compared with PIP I (**1**) and PIP K (**2**) (PI polyamide without SAHA). This prompted us to redesign SAHA-PIP and replace the SAHA moiety with β -alanine (the β -moiety) in the non-core binding region of the 10-ring hairpin polyamide. The results showed that the β -moiety enforces binding of PIP in the closed form (Table 2.1).

Application of the epigenetically active and DNA-binding SAHA-PIPs I and K relies largely on high-affinity sites. SAHA-PIPs are expected to bind in the minor groove of DNA according to well-defined pairing rules (N-terminal to

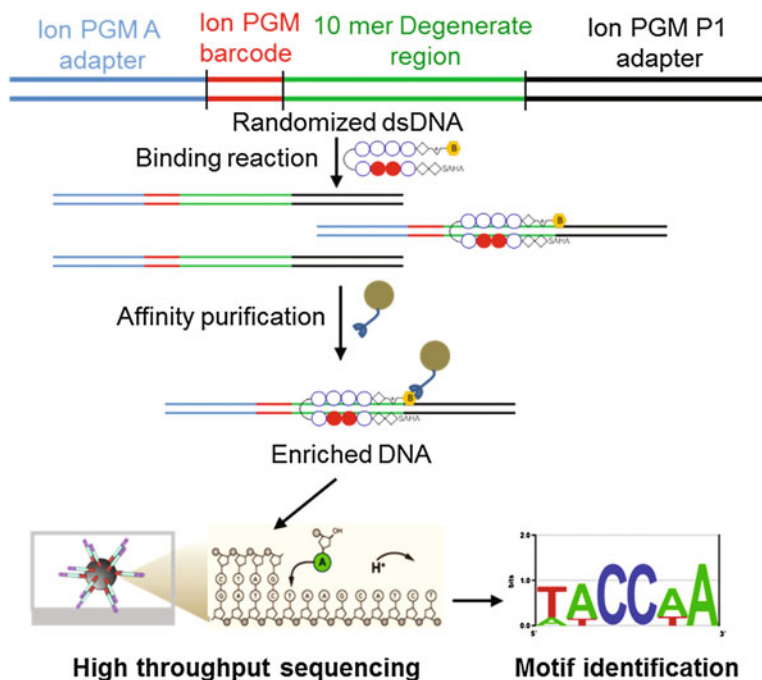


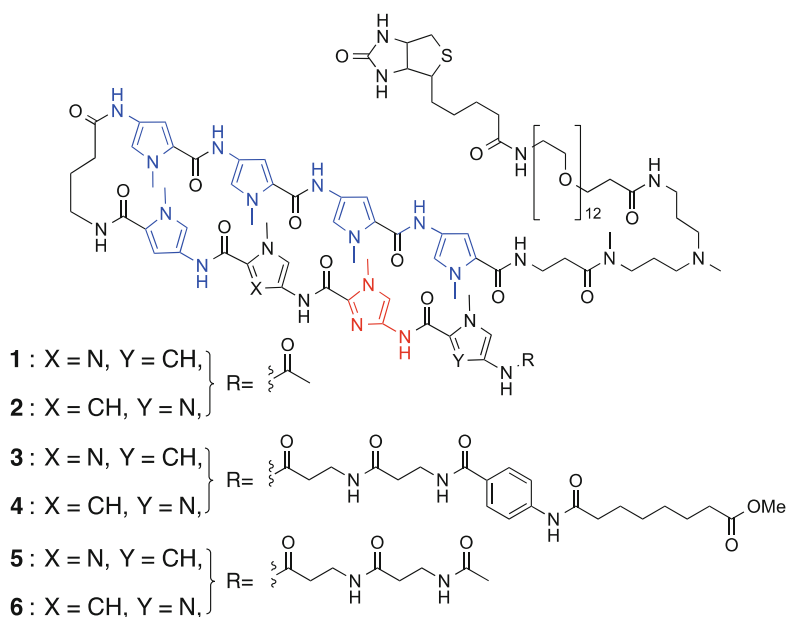
Fig. 2.1 High-throughput sequencing based Bind-n-seq experimental design and binding motif discovery

C-terminal:5'-3') [7]. Meier et al. studied the sequence selectivity and canonical pairing rules of PIP-DNA binding in a broad sequence context using affinity purification coupled with massively parallel sequencing in an Illumina sequencer [12]. To characterize the binding specificity of the PIP derivatives, we carried out high-throughput sequencing (Bind-n-seq) studies in a randomized DNA sequence space to define the recognition motifs of PIPs and SAHA-PIPs (Fig. 2.1) .

2.2 Results and Discussion

2.2.1 PIP and PIP Conjugates Synthesis and Bind-n-Seq Analysis

A small library of biotin-conjugated PIPs and SAHA-PIPs (**1–4**) (Scheme 2.2, **1–4**) was first synthesized (in the SAHA moiety of biotin-conjugated SAHA-PIPs (**3** and **4**), we used the methyl ester instead of hydroxamic acid for synthetic convenience. This change was not expected to affect the DNA-binding affinity of the SAHA-PIPs because SAHA does not bind DNA. Synthesis details are provided in the materials



Scheme 2.2 Chemical structures 1–6

and methods) and allowed to bind with 10-bp randomized DNA fragments equipped with an Ion PGM-compatible adapter sequence (for details of the design of the oligos and primer used to make duplex DNA, see the materials and methods). By using affinity purification, the enriched DNA sequences bound by the PIP derivatives were extracted and sequenced with an Ion PGM sequencer. After a quality check and adapter trimming of the sequenced reads, Bind-n-Seq analysis [15] revealed enrichment of “*k*-mer” ($k = 6$), and confirmed that the binding sites of **3** and **4** matched the canonical binding rule. Interestingly, **4** was found to bind in the forward orientation (N-terminal to C-terminal:5′-3′, Fig. 2.2d), but it was difficult to determine the binding orientation of **3** (Fig. 2.2c) because of its symmetrical binding site (5′-WWCCWW-3′/5′-WWGGWW-3′). The results obtained with **4** are significant because the PIPs have the ability to bind in both forward and reverse orientation and hence have a relatively large number of duplex DNA sites as potential targets. When the binding of **4** is principally in the forward orientation, then the number of potential target sites is reduced, which in turn suggests superior specificity over **2**. Motifs were identified by using the DREME [16] and enoLOGOS [17] algorithms, which are motif-binding programs developed to identify the short response elements typically bound by eukaryotic transcription factors. From the analysis described above, a graphical depiction of the strongest motif was generated for the highly enriched binding site of **1–4** (Fig. 2.2a); a table of enriched sequences is given in Tables 2.1 and 2.2. Replicate quantification validated the high-affinity binding; for example, the enrichment of DNA sequences

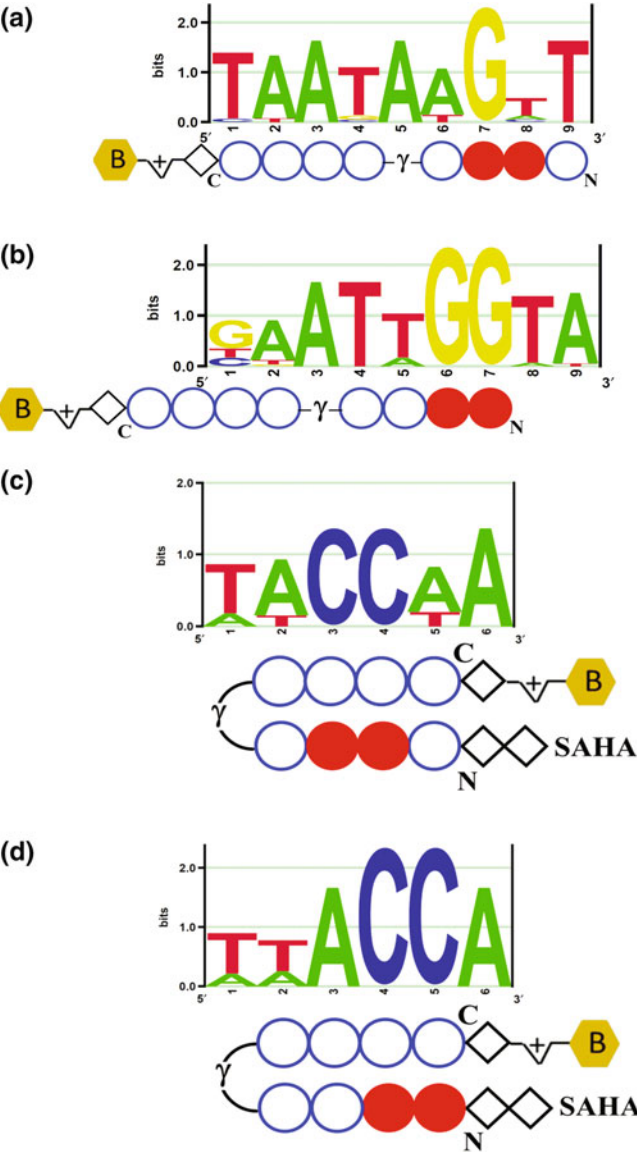


Fig. 2.2 Structure of Polyamides binding region and its primary binding motifs identified using Bind-n-seq through next generation sequencing: **a** **1** **b** **2** **c** **3** **d** **4**

bound by **3** was highly reproducible ($R^2 = 0.92$ for two separate binding and enrichment experiments; Fig. 2.3). The enriched sequence data and the similarity of enrichment between two different experiments are shown in the Table 2.1.

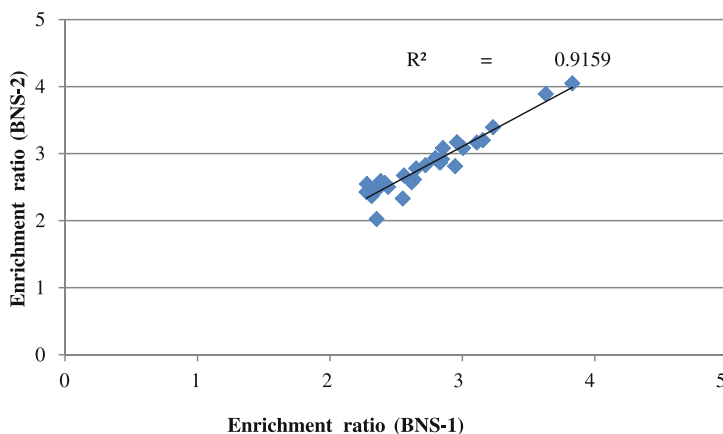


Fig. 2.3 Graphical representation of relative enrichment of two separate binding and enrichment reactions of **3** (sequence details are provided in Table 2.2)

2.2.2 PIP Redesign, $\beta\beta$ -PIP Synthesis and Bind-n-Seq Analysis

The results of the massively parallel sequencing analysis Bind-n-seq showed that **3** and **4** (Fig. 2.2c, d) had high-affinity binding toward the canonical binding sites. However, the results also showed that **1** and **2** bound with randomized DNA in an open conformation (Fig. 2.2a, b). In addition to the SAHA moiety, **3** and **4** contain two β -alanine moieties (used to conjugate the SAHA with PIP) in the non-core binding machinery; compounds **1** and **2** lacks these two β -moieties. The high-affinity binding of **3** and **4** prompted us to investigate the role of the β -alanine moiety in the construction of hairpin PIP. To this end, we designed PIP derivatives **5** and **6** (corresponding to SAHA-PIPs without SAHA) (Scheme 2.2, 5–6), having two β -alanine moieties in the non-core binding domain. Thus, $\beta\beta$ -PIP-biotin conjugates **5** and **6** were synthesized and subjected to binding reaction with randomized DNA with affinity purification followed by massively parallel sequencing.

Bind-n-seq data analyses were performed for sequence reads enriched by **5** and **6**. Assessment of the identified motifs (Fig. 2.4a, b, corresponding sequence details in Tables 2.3 and 2.4) revealed that the β -alanine moiety enforces high-affinity binding of the PIPs in the closed form. Compounds **5** and **6** bind with the WWCCWA and TWACCA/AWTGGT sites, respectively. Since we could not identify the binding orientation of **5**, we considered only the binding orientation of **6**. Interestingly, DNA motifs enriched after binding **6** revealed that this compound bound to the DNA with equal affinity of forward and reverse orientations. This investigation thus indicates that the β -alanine moiety induces binding of hairpin PIP in the closed form with DNA and does not affect the direction of binding.

2.2.3 Binding Affinity Conformation by SPR

To confirm the findings of Bind-n-seq analyses, we performed SPR analysis because previous reports have shown that the rates of association (k_a) and dissociation (k_d), and the dissociation constant (K_D) correlate well with the binding affinity [13]. For the SPR analysis, we synthesized PIPs **7**–**10** without biotin conjugates (Scheme 2.3). SPR analyses were performed with a Biacore X instrument with the match and mismatch DNA oligonucleotides from the human transcription factor sequences to identify whether PIP binding can selectively trigger human transcription factors (See materials and methods) [18]. SAHA-PIP K and I were shown to distinctively regulate the expression of silent developmental genes in human fibroblasts. Based on our previous reports, for **7** and **9** (similar binding sequence as SAHA-PIP I), we selected the match sequences in the promoter region of the gene encoding the developmental associated POU homeodomain [11] and for **8** and **10** (similar binding sequence as SAHA-PIP K), we selected the match (**8** and **10**) sequences in the promoter region of the gene encoding the germ cell associated *PIWIL 2* [14]. For checking the relative binding affinity, the match sequences of SAHA-PIP I (**7** and **9**) and K (**8** and **10**) were interchanged and used as the mismatch sequences. We compared the binding ability and efficiency of compounds **9** and **10** with those of **7** and **8** in the corresponding match and mismatch binding of DNA. The PIP derivatives were passed through a 5'-biotinylated hairpin DNA immobilized sensor chip by using a biotin-avidin system. The binding affinities of the PIP derivatives were measured by analyzing the data produced from the SPR method. Detailed assessment of the SPR dissociation constant (K_D) values showed that **9** and **10** had higher affinity toward the matching polyamide–DNA binding sites than **7** and **8**, with a K_D range of approximately 5–25-fold. These data correlate best with the binding sites identified by Bind-n-seq. The rates of association (k_a) and dissociation (k_d) and the dissociation constants (K_D) are shown in Table 2.5, and the sensorgrams are shown in the Supplementary Information (Figs. 2.5 and 2.6). Although many methods are under development, the targeting of dsDNA in a genomic sequence perspective by using small synthetic molecules remains difficult. The structural composition of polyamides and DNA sequence-dependent structural variations may decrease the binding specificity of PIP derivatives to the desired specific DNA-binding sites. This study delivers a low-cost sequencing method for designing and screening of sequence-specific DNA-binding molecules with the standardized experimental conditions. We have shown that the addition of a SAHA moiety to the N-terminal of PIPs does not affect the canonical binding rule of polyamides in a wide sequence context. Thus, the HDAC inhibitor SAHA can be used in a sequence-specific manner with the assistance of polyamides as we reported earlier [14].

Table 2.3 Relative enrichment of 6-base pair “k”mers enriched by **5** binding and enrichment reactions. (only top 30 sequences were shown)

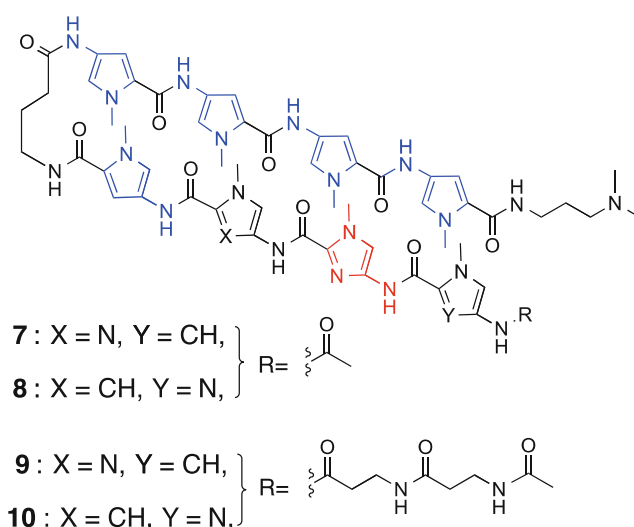
Rank	Sequence	Enrichment
1	TACCAA	3.635
2	ATTACC	3.585
3	CTACCA	2.993
4	TGGTAA	2.785
5	GTACCA	2.982
6	AACCAA	2.745
7	ATGGTA	2.754
8	TACCTA	2.325
9	ACTACC	2.541
10	GGTAGA	2.641
11	ATACCA	2.693
12	ACCAAT	2.539
13	CTTCCA	2.531
14	ACCAAC	2.197
15	GGTACA	2.393
16	TTCCAA	2.513
17	AGGTAG	2.568
18	GTTCCA	2.363
19	AATACC	2.722
20	CAACCA	2.085
21	GGAAGA	2.335
22	ACCAAG	2.215
23	AAGGTA	2.208
24	AGTACC	2.256
25	ATAACC	2.224
26	GAACCA	2.089
27	ACTTCC	2.232
28	ACCTAC	2.034
29	AGGTAC	2.234
30	ACCATC	2.445

2.2.4 Discussion

Despite the finding that polyamides bind according to the canonical binding rule, we have identified that **1** and **2** (without SAHA) can also preferentially bind to DNA in an open conformation instead of in the hairpin form. This phenomenon was not observed for **3** and **4** (PIPs with SAHA), which prompted us to look in more detail at the features leading to polyamide binding in the closed form. Our investigation focused on the two N-terminal β -alanine moieties in the non-core binding machinery. Several biologically active polyamides with β -residues in the core binding sites have been reported, and the importance of β -residue placement in the core binding region of polyamides has been examined [19]. However, none of the

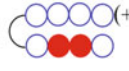
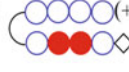
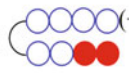
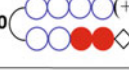
Table 2.4 Relative enrichment of 6-base pair “k”mers enriched by **6** binding and enrichment reactions

Rank	Sequence	Enrichment
1	AATGGT	3.147
2	TTACCA	3.142
3	ATGGTA	2.851
4	ATTACC	2.835
5	ATAACC	2.476
6	ATTGGT	2.395
7	TTGGTA	2.338
8	TAATGG	2.148
9	TATGGT	2.126
10	TAACCA	2.101

**Scheme 2.3** Chemical structures 7–10

studies investigated the effect of β -moieties in the N-terminal non-core binding machinery. We studied the Bind-n-seq offered high-affinity binding of polyamides with $\beta\beta$ -PIP derivatives **5** and **6** and established which sequences are enriched by $\beta\beta$ -PIPs with a high-throughput sequencing method. The identified motifs clearly indicate that the N-terminal β -residue can restore the forward polyamide-binding specificity. This conclusion was further supported by the results of SPR analysis. Although the $\beta\beta$ -PIP derivatives can restore the forward binding affinity, the motifs identified by **6** show both forward and reverse binding orientations, in which the core binding N-terminal of the polyamide is *N*-methylimidazole (I). This result suggests that the $\beta\beta$ -residues direct the polyamide to bind in the hairpin form. Previous studies have shown [12] that the replacement of a β -amino–GABA linker with an α -amino–GABA linker can restore the forward binding of polyamides. Our

Table 2.5 Binding affinities of polyamide 7–10

polyamide	5'-Biotin labeled- GCGC TTCTT CCCC _T 3'-CGCGG AAGGA GGGG _T		5'-Biotin labeled- CGTC TTCC AGCAG _T 3'-GCAGG AAAGT CGTC _T		specificity
	K_D (M) match	K_a (M ⁻¹ s ⁻¹) K_d (s ⁻¹)	K_D (M) 2-bp mismatch	K_a (M ⁻¹ s ⁻¹) K_d (s ⁻¹)	
7	 1.2×10^{-7}	1.4×10^5 1.6×10^{-2}	1.8×10^{-7}	1.0×10^5 1.8×10^{-2}	1.5
9	 2.8×10^{-8}	1.2×10^6 3.3×10^{-2}	1.4×10^{-7}	1.6×10^5 2.1×10^{-2}	5.0
8	 1.1×10^{-7}	1.4×10^5 1.6×10^{-2}	1.5×10^{-8}	2.5×10^5 3.7×10^{-3}	7.7
10	 6.5×10^{-8}	2.6×10^5 1.7×10^{-2}	2.5×10^{-9}	1.3×10^6 3.1×10^{-3}	26

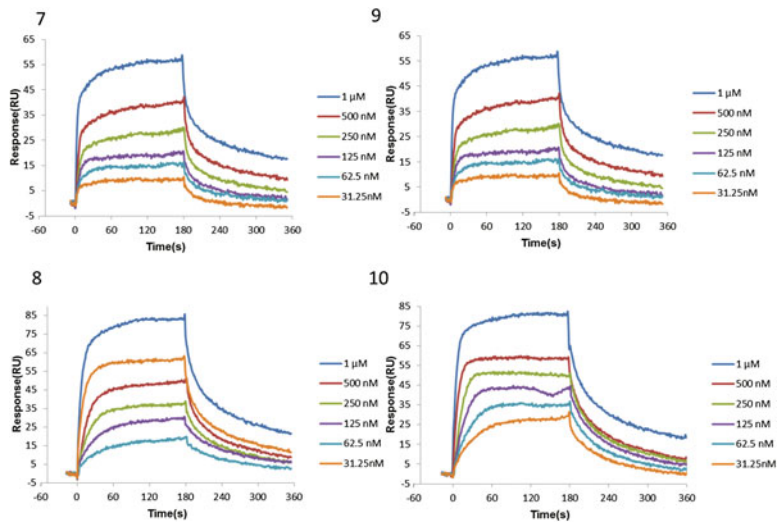


Fig. 2.5 SPR sensorgrams for the interactions of **7** and **9** with its match and **8** and **10** with its mismatch binding sequences in the promoter region of the gene encoding POU homeodomain are shown (Binding specificities were provided in Table 2.1)

results with **5** show that the N-terminal non-core β -residues can also restore the high forward binding affinity of polyamides, in which the core binding N-terminal of the polyamide is *N*-methylpyrrole (P). Further studies need to be carried out to confirm the exact number of β -residues required to restore the forward orientation of polyamide binding.

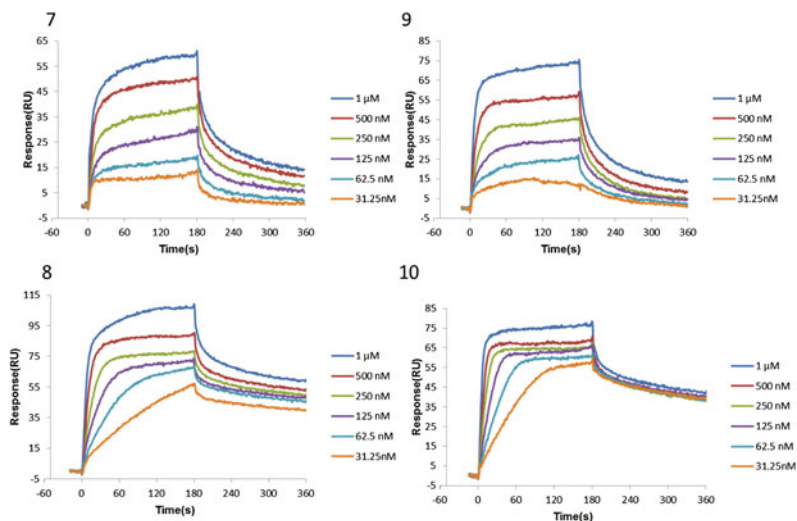


Fig. 2.6 SPR sensorgrams for the interactions of **7** and **9** with its mismatch and **8** and **10** with its match binding sequences in the promoter region of the gene encoding *PIWIL2* are shown (Binding specificities were provided in Table 2.1)

2.3 Materials and Methods

2.3.1 General Materials and Synthesis

Reagents and solvents were purchased from standard suppliers and used without further purification. The EZ-Link NHS-PEG₁₂-Biotin was purchased from Thermo Scientific (number 21312). The analytical HPLC was performed with a COSMOSIL 5C18-MS-II reversed phase column (4.6 × 150 mm, Nacalai) in 0.1% TFA in water with CH₃CN as eluent at a flow rate of 1.0 mL/min, and a linear gradient elution of 0–100% CH₃CN over 20 or 40 min with detection at 254 nm. The HPLC purification was performed with a COSMOSIL 5C18-MS-II reversed phase column (10 × 150 mm, Nacalai) in 0.1% TFA in water with CH₃CN as eluent. The final products were analyzed by ESI-TOF-MS (Bruker).

2.3.1.1 General Scheme for Synthesis of Biotin-Conjugated Polyamides

Ac-Py-Im-Im-Py-γ-Py-Py-Py-β-(+)-PEG₁₂-Biotin (**1**)

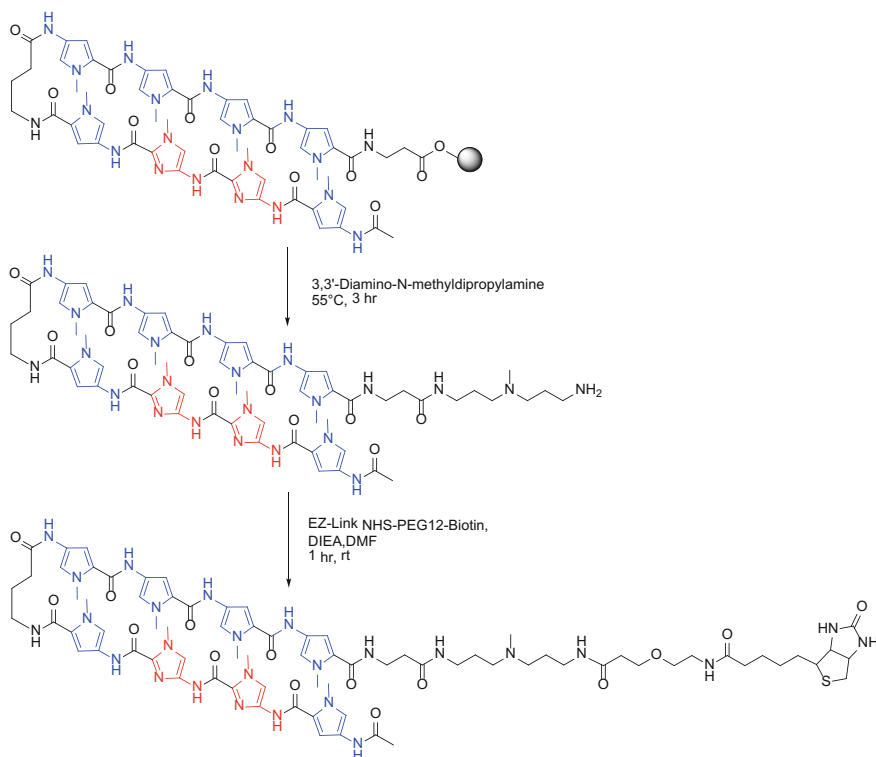
Py-Im polyamide supported by oxime resin was prepared in a stepwise reaction by a reported Fmoc solid-phase procedure. The product with oxime resin was cleaved

with 3,3'-Diamino-*N*-methyldipropylamine (500 μ L) at 55 $^{\circ}$ C for 3 h. After filtration and evaporation, the resulted oil was quenched by Et₂O. The obtained precipitation was washed with Et₂O for three times and dried in vacuum. The product was used for the next coupling steps without further purification.

A solution of polyamide (1.32 mg, 1 μ mol), EZ-Link NHS-PEG₁₂-Biotin (1 mg, 1.2 μ mol) and DIEA (2 μ L, 10 μ mol) in DMF (100 μ L) was stirred at room temperature for 1 h. After consumption of starting material was confirmed by HPLC, Et₂O was added to the mixture and the resultant was collected by centrifugation, and washed by Et₂O and CH₂Cl₂. The crude product was purified by reverse-phase HPLC. After lyophilization **1** was obtained (0.9 mg, yield 42%). Analytical HPLC: t_R = 15.5 min (0.1% TFA-CH₃CN, 0–100%, 40 min). ESI-TOF-MS m/z : calcd for C₉₉H₁₄₆N₂₆O₂₆S [M + 2H]²⁺ 2149.0622, found 2149.0460.

Ac-Im-Im-Py-Py- γ -Py-Py-Py-Py- β (+)-PEG₁₂-Biotin (**2**)

A similar synthetic procedure of **1** was used for the preparation of **2**. Analytical HPLC: t_R = 15.7 min (0.1% TFA-CH₃CN, 0–100%, 40 min). ESI-TOF-MS m/z : calcd for C₉₉H₁₄₆N₂₆O₂₆S [M + 2H]²⁺ 2149.0622, found 2149.1082.



SAHA- β - β -Py-Im-Im-Py- γ -Py-Py-Py-Py- β -(+)-PEG₁₂-Biotin (3)

A similar synthetic procedure of **1** was used for the preparation of **3**. Analytical HPLC: t_R = 17.2 min (0.1% TFA-CH₃CN, 0–100%, 40 min). ESI-TOF-MS m/z : calcd for C₁₁₉H₁₇₃N₂₉O₃₁S [M + 2H]²⁺ 2538.2573, found 2538.0851.

SAHA- β - β -Im-Im-Py-Py- γ -Py-Py-Py-Py- β -(+)-PEG₁₂-Biotin (4)

A similar synthetic procedure of **1** was used for the preparation of **4**. Analytical HPLC: t_R = 17.3 min (0.1% TFA-CH₃CN, 0–100%, 40 min). ESI-TOF-MS m/z : calcd for C₁₁₉H₁₇₃N₂₉O₃₁S [M + 2H]²⁺ 2538.2573, found 2538.2536.

Ac- β - β -Py-Im-Im-Py- γ -Py-Py-Py-Py- β -(+)-PEG₁₂-Biotin (5)

A similar synthetic procedure of **1** was used for the preparation of **5**. Analytical HPLC: t_R = 9.1 min (0.1% TFA-CH₃CN, 0–100%, 20 min). ESI-TOF-MS m/z : calcd for C₁₀₅H₁₅₆N₂₈O₂₈S [M + 2H]²⁺ 2291.1365, found 2290.8420.

Ac- β - β -Im-Im-Py-Py- γ -Py-Py-Py-Py- β -(+)-PEG₁₂-Biotin (6)

A similar synthetic procedure of **1** was used for the preparation of **6**. Analytical HPLC: t_R = 9.1 min (0.1% TFA-CH₃CN, 0–100%, 20 min). ESI-TOF-MS m/z : calcd for C₁₀₅H₁₅₆N₂₈O₂₈S [M + 2H]²⁺ 2291.1365, found 2290.8356.

2.3.2 Oligomer Sequences Used in Bind-n-Seq Method

Bind-n-Seq 92mer:

5'CCATCTCATCCCTGCGTGTCTCCGACTCAGXXXXXXXXXXXXNNNNNNNNNNNAAT
CACCGACTGCCCATAGAGAGAAAGCGGAGGCGTAGTGG 3'

- All barcoded Bind-n-Seq 92 mers were synthesized by Sigma Aldrich machine mixing, standard desalting purification.
- Barcode in yellow shadow (XXXXXXXXXXXX), 10-letter barcodes used as per Ion torrent sequencing technologies.

Primer 1:

5'-CCA CTA CGC CTC CGC TTT CCT CTC TA-3'

- Used in initial primer extension reaction.
- Synthesized by Sigma Aldrich, purification by standard desalting.

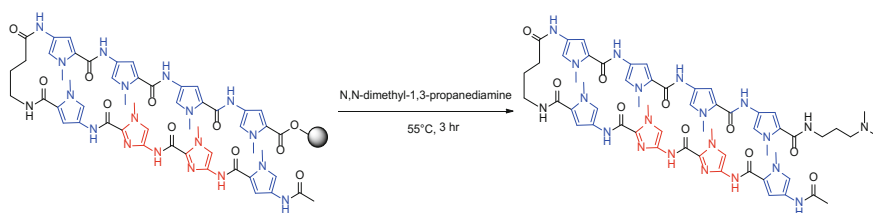
2.3.3 *Bind-n-Seq*

Previously reported Bind-n-Seq experiments [12, 13] and subsequent analysis to evaluate small molecule binding affinity towards specific DNA sequences in a broad context sequence pool were customized with the modifications suitable for Ion torrent PGM sequencer. The scheme involves the following three major steps,

- (1) Synthesis of biotinylated PIPs, and randomized oligonucleotides with high-throughput sequencing platform specific adapters (Ion torrent PGM). Adapter ligated oligonucleotides (3 μ M) were duplexed by primer extension with adapter specific primer 1 (9 μ M) in 25 μ L reaction contain 2 \times goTag PCR master mix with 2 mM Mg^{2+} . Reactions were performed at 95 $^{\circ}$ C (3 min), 60 $^{\circ}$ C (2 min), 70 $^{\circ}$ C (5 min) and then 4 $^{\circ}$ C using Bio-Rad thermocycler. Biotin conjugated PIPs (100 nM) were allowed to equilibrate with duplex random oligonucleotides for 14 h followed by Streptavidin M-280 Dynabeads (Beads were prepared based on the previous report [12, 13]) separation of the bound and unbound sequences using affinity purification.
- (2) Polyamides-enrichment recovered DNA was diluted 1:10 and amplified with sequencing library adapter specific primer for 15 cycles in order to obtain enough sequencing template. After purification enriched libraries were subjected to quality and quantity check with Agilent DNA High sensitivity BioAnalyzer kit (Agilent Technologies, USA). The qualified libraries were used for template preparation using Ion PGMTM template OT2 200 kit in Ion one touch2 system. The templates were then enriched using Ion one touch ES. The enriched libraries were sequenced as a single read sequencing with Ion PGM sequencer using Ion PGMTM sequencing 200 kit v2 and 318 chips (Life Technologies, USA) by following the manufacturer's instructions.
- (3) The sequenced reads (composed of A, C, T, or G) was then processed to obtain a valid constant region and unique random region) were retained and split into separate files through unique 10-nt ion Xpress-barcode using the Ion torrent suit 3.4.2 as mentioned before. To count the amount of PIP enriched unique DNA sequences, a sliding window of length k ($=6$) in MERMADE, a new pipeline for Bind-n-Seq analysis (<http://korflab.ucdavis.edu/Datasets/BindNSeq.27>) were used. Highly enriched motifs were confirmed with DREME primary motif analysis.

2.3.4 Surface Plasmon Resonance (SPR)

2.3.4.1 General Scheme for Synthesis of Polyamides Used for SPR Analysis



Ac-Py-Im-Im-Py- γ -Py-Py-Py-Dp (7)

Py-Im polyamide supported by oxime resin was prepared in a stepwise reaction by a reported Fmoc solid-phase procedure. The product with oxime resin was cleaved with *N,N*-dimethyl-1,3-propanediamine (500 μ L) at 55 °C for 3 h. After filtration and evaporation, the resulted oil was quenched by Et₂O. The obtained precipitation was washed with Et₂O for three times and dried in vacuo. The crude product was purified by reverse-phase HPLC. Analytical HPLC: t_R = 9.0 min (0.1% TFA-CH₃CN, 0–100%, 20 min). ESI-TOF-MS m/z : calcd for C₅₇H₆₉N₂₁O₁₀ [M + 2H]²⁺ 1209.5536, found 1209.3718.

Ac-Im-Im-Py-Py- γ -Py-Py-Py-Dp (8)

A similar synthetic procedure of **7** was used for the preparation of **8**. Analytical HPLC: t_R = 9.4 min (0.1% TFA-CH₃CN, 0–100%, 20 min). ESI-TOF-MS m/z : calcd for C₅₇H₆₉N₂₁O₁₀ [M + 2H]²⁺ 1209.5536, found 1209.4150.

Ac- β - β -Py-Im-Im-Py- γ -Py-Py-Py-Dp (9)

A similar synthetic procedure of **7** was used for the preparation of **9**. Analytical HPLC: t_R = 9.2 min (0.1% TFA-CH₃CN, 0–100%, 20 min). ESI-TOF-MS m/z : calcd for C₆₃H₇₉N₂₃O₁₂ [M + 2H]²⁺ 1351.6279, found 1351.5146.

Ac- β - β -Im-Im-Py-Py- γ -Py-Py-Py-Dp (10)

A similar synthetic procedure of **7** was used for the preparation of **10**. Analytical HPLC: t_R = 9.2 min (0.1% TFA-CH₃CN, 0–100%, 20 min). ESI-TOF-MS m/z : calcd for C₆₃H₇₉N₂₃O₁₂ [M + 2H]²⁺ 1351.6279, found 1351.5112.

2.3.4.2 SPR Analysis

SPR analyses were performed as described in previous studies [18] using BIACORE X instrument. In brief, we purchased biotinylated hairpin DNAs with match and mismatch PIP binding sites from JBioS (Tokyo, Japan) (**7** and **9** match/**8** and **10** mismatch: Sequence in the promoter region of the gene encoding POU homeodomain [8]-5'-Biotin- GCG CCT TCC TTC CCC TTT TGG GGA AGG AAG GCG C-3' and **7** and **9** mismatch/**8** and **10** match: Sequence in the promoter region of the gene encoding PIWIL2 [10]-5'-Biotin- CGT CCT TTC CAG CAG TTTTCT GCT GGA AAG GAC G-3'). The purchased DNAs were diluted to 100 nM. Then immobilization of hairpin biotinylated DNAs was performed on a streptavidin-coated sensor chip SA to obtain the desired immobilization level (approximately 1100–1200 RU rise). HBS-EP buffer (10 mM HEPES, pH 7.4, 150 mM NaCl, 3 mM EDTA, and 0.005% surfactant P20) with 0.1% DMSO at room temperature was used to carry out SPR assays. We prepared an order of **7–10** (without biotin) solutions with different concentrations in HBS-EP buffer with 0.1% DMSO and at the flow rate of 20 mL/min, the PIPs were passed on the immobilized oligonucleotide chips. The data processing was performed using BIAevaluation 4.1 program to measure the rates of association (k_a) and dissociation (k_d) and dissociation constant (K_D), with an appropriate fitting model.

References

1. Berdasco M, Esteller M (2010) Aberrant epigenetic landscape in cancer: how cellular identity goes awry. *Dev Cell* 19:698–711
2. Wade WS, Mrksich M, Dervan PB (1992) Design of peptides that bind in the minor groove of DNA at 5'-(A, T)G(A, T)C(A, T)-3' sequences by a dimeric side-by-side motif. *J Am Chem Soc* 114:8783–8794. doi:[10.1021/ja00049a006](https://doi.org/10.1021/ja00049a006)
3. Wemmer DE, Dervan PB (1997) Targeting the minor groove of DNA. *Curr Opin Struct Biol* 7:355–361. doi:[10.1016/S0959-440X\(97\)80051-6](https://doi.org/10.1016/S0959-440X(97)80051-6)
4. Dervan PB, Edelson BS (2003) Recognition of the DNA minor groove by pyrrole-imidazole polyamides. *Curr Opin Struct Biol* 13:284–299
5. Pelton JG, Wemmer DE (1989) Structural characterization of a 2:1 distamycin A.d (CGCAAATTGGC) complex by two-dimensional NMR. *Proc Natl Acad Sci U S A* 86:5723–5727. doi:[10.1073/pnas.86.15.5723](https://doi.org/10.1073/pnas.86.15.5723)
6. Kouzarides T (2007) Chromatin modifications and their function. *Cell* 128:693–705
7. Ohtsuki A, Kimura MT, Minoshima M et al (2009) Synthesis and properties of PI polyamide-SAHA conjugate. *Tetrahedron Lett* 50:7288–7292. doi:[10.1016/j.tetlet.2009.10.034](https://doi.org/10.1016/j.tetlet.2009.10.034)
8. Pandian GN, Shinohara KI, Ohtsuki A et al (2011) Synthetic small molecules for epigenetic activation of pluripotency genes in mouse embryonic fibroblasts. *ChemBioChem* 12:2822–2828. doi:[10.1002/cbic.201100597](https://doi.org/10.1002/cbic.201100597)
9. Pandian GN, Ohtsuki A, Bando T et al (2012) Development of programmable small DNA-binding molecules with epigenetic activity for induction of core pluripotency genes. *Bioorg Med Chem* 20:2656–2660. doi:[10.1016/j.bmc.2012.02.032](https://doi.org/10.1016/j.bmc.2012.02.032)

10. Pandian GN, Nakano Y, Sato S et al (2012) A synthetic small molecule for rapid induction of multiple pluripotency genes in mouse embryonic fibroblasts. *Sci Rep* 2:1–8. doi:[10.1038/srep00544](https://doi.org/10.1038/srep00544)
11. Pandian GN, Taniguchi J, Junetha S et al (2014) Distinct DNA-based epigenetic switches trigger transcriptional activation of silent genes in human dermal fibroblasts. *Sci Rep* 4:3843. doi:[10.1038/srep03843](https://doi.org/10.1038/srep03843)
12. Meier JL, Yu A, Korf I et al (2012) Guiding the design of synthetic DNA-binding molecules with massively parallel sequencing. *J Am Chem Soc* 134:17814–17822. doi:[10.1021/ja308888c](https://doi.org/10.1021/ja308888c)
13. Kang JS, Meier JL, Dervan PB (2014) Design of sequence-specific DNA binding molecules for DNA methyltransferase inhibition. *J Am Chem Soc* 136:3687–3694. doi:[10.1021/ja500211z](https://doi.org/10.1021/ja500211z)
14. Han L, Pandian GN, Junetha S et al (2013) A synthetic small molecule for targeted transcriptional activation of germ cell genes in a human somatic cell. *Angew Chem Int Ed* 52:13410–13413
15. Zykovich A, Korf I, Segal DJ (2009) Bind-n-Seq: high-throughput analysis of in vitro protein-DNA interactions using massively parallel sequencing. *Nucleic Acids Res* 37:e151. doi:[10.1093/nar/gkp802](https://doi.org/10.1093/nar/gkp802)
16. Bailey TL (2011) DREME: motif discovery in transcription factor ChIP-seq data. *Bioinformatics* 27:1653–1659. doi:[10.1093/bioinformatics/btr261](https://doi.org/10.1093/bioinformatics/btr261)
17. Workman CT, Yin Y, Corcoran DL et al (2005) enoLOGOS: a versatile web tool for energy normalized sequence logos. *Nucleic Acids Res*. doi:[10.1093/nar/gki439](https://doi.org/10.1093/nar/gki439)
18. Morinaga H, Bando T, Takagaki T et al (2011) Cysteine cyclic pyrrole-imidazole polyamide for sequence-specific recognition in the DNA minor groove. *J Am Chem Soc* 133:18924–18930. doi:[10.1021/ja207440p](https://doi.org/10.1021/ja207440p)
19. Minoshima M, Bando T, Sasaki S et al (2008) Pyrrole-imidazole hairpin polyamides with high affinity at 5'-CGCG-3' DNA sequence; influence of cytosine methylation on binding. *Nucleic Acids Res* 36:2889–2894. doi:[10.1093/nar/gkn116](https://doi.org/10.1093/nar/gkn116)

Advancing Development of Synthetic Gene Regulators
With the Power of High-Throughput Sequencing in
Chemical Biology

Chandran, A.

2018, XV, 114 p. 49 illus., 44 illus. in color., Hardcover

ISBN: 978-981-10-6546-0