Comparative Analysis of DNA-Binding Selectivity of Hairpin and Cyclic Pyrrole–Imidazole Polyamides Based on Next-Generation Sequencing

Gengo Kashiwazaki[+],[a] Anandhakumar Chandran[+],[a] Sefan Asamitsu,[a] Takashi Kawase,[b] Yusuke Kawamoto,[a] Yoshito Sawatani,[a] Kaori Hashiya,[a] Toshikazu Bando,*[a] and Hiroshi Sugiyama*[a, c]

Abstract: Many long pyrrole–imidazole polyamides (PIPs) have been synthesized to pursue higher specificity with the aim of realizing the great potential of such compounds in biological and clinical areas. Among several types of PIPs, we designed and synthesized hairpin and cyclic PIPs targeting identical sequences. Bind-n-Seq analysis revealed that they both bound to the predetermined sequences. However, adenines in the data analyzed by the previously reported Bind-n-Seq method appeared to be significantly higher in the motif ratio than thymines, even though the PIPs were not expected to distinguish A from T. We therefore examined the experimental protocol and analysis pipeline in detail and developed a new method named Bind-n-Seq-based motif identification with reference sequence (Bind-n-Seq-MR). High-throughput sequence analysis of the PIP-enriched DNA data by Bind-n-Seq-MR presented A and T comparably. Surface plasmon resonance assays were also performed to validate the new method of analysis.

Introduction

Synthetic chemical compounds that have the ability to bind DNA selectively have the potential to act as gene-modulating agents and to function as transcription factors. *N*-Methylpyrrole–*N*-methylimidazole polyamide (PIP) compounds constitute a class of synthetic molecules with such a function, and they have been shown to alter gene expression.[1] P/I and P/P pairs recognize C/G and A/T, respectively, by fitting into the minor groove of DNA in an antiparallel 2:1 binding mode.[2] Nevertheless, a robust understanding of the chemical background underlying the processes involved in PIP applications remains important.

For several decades, polyacrylamide gel electrophoresis (PAGE) has been one of the most useful methods with which to define PIP-binding sites using a technique called DNAse I footprinting,[3] or to define alkylated sites by PIP conjugates.[1j, 3f, 4] Melting temperature (T_m) measurements[5] and surface plasmon resonance (SPR) analyses[4e, 4f, 5c, 5d, 6] have also provided data on the interaction between DNA sequences and PIP compounds, and the results have aided investigations into affinity and specificity.

However, all the techniques mentioned above have an inherent limitation with respect to the number of sequences that can be tested at a time. Next-generation sequencing (NGS)/high-throughput sequencing is an elegant and powerful method that can be used to circumvent this issue because it can be used to analyze millions to several hundred millions of sequence reads in a single run.[7] In 2006, Ansari and coworkers reported a microarray platform that could be used to assess the sequence affinities of eight-ring hairpin PIPs and extradenticle-PIP complexes for every random sequence in a rapid, unbiased, and unsupervised manner.[8] Furthermore, quantification of the equilibrium association constant is possible by calibration with DNase I footprinting data.[9] A method of visualizing binding sites other than the use of a sequence motif was proposed, which was called a sequence specificity landscape to show binding sequences of natural DNA-binding proteins as well as engineered molecules, including hairpin PIPs.[3d] Later, biotinylated PIPs were applied to analyze their binding sequences and orientations.[3e] Recently, PIPs with biotin and psoralen (a photocrosslinker) were synthesized to perform crosslinking of small molecules for isolation of chromatin (COSMIC) in live human cells.[10]

A crucial characteristic of PIP is its ability to bind to a DNA sequence specifically. One can easily imagine that longer PIPs can recognize longer DNA sequences, resulting in better specificity. Hence, it is important to design and synthesize long PIPs without losing affinity toward DNA. To date, four types of PIPs have been proposed; namely, hairpin,[11] cyclic,[12] tandem hairpin,[13] and linear.[14]

Here, we examine and compare three types of PIPs by Bind-n-Seq utilizing the power of NGS. The hairpin form is widely synthesized because of its simplicity. However, the advantage of the cyclic type of PIP over hairpin is the exclusion of ambiguity of the binding mode: whereas hairpin structures can also bind in a linear mode, the cyclic form can only bind in a cyclic mode. There are two types of cyclic PIPs that differ in their linker structure: one linker is γ-aminobutyric acid, which was developed by Dervan and coworkers,[1c, 12a, 15] and the second is a sulfide bond formed by reaction between cysteine and a chloroacetyl group, which was developed by our group.[6d, 12c] The first cyclic PIP that consisted of a six-membered ring had an aminopropyl group at the 1-position of pyrrole;[15a] affinity and specificity were then further improved by constructing an eight-membered ring PIP with an amino group on the γ-turn.[12a] This

[a] Dr. G. Kashiwazaki, Dr. A. Chandran, S. Asamitsu, Y. Kawamoto, Y. Sawatani, K. Hashiya, Dr. T. Bando, Prof. Dr. H. Sugiyama
Department of Chemistry
Graduate School of Science, Kyoto University
Kitashirakawaoiwakecho, Sakyo, Kyoto, 606-8502, Japan
E-mail: bando@kuchem.kyoto-u.ac.jp
hs@kuchem.kyoto-u.ac.jp

[b] Dr. T. Kawase
Department of Systems Science
Graduate School of Informatics, Kyoto University
Yoshida-Honmachi 36-1, Sakyo, Kyoto, 606-8501, Japan

[c] Prof. Dr. H. Sugiyama
Institute for Integrated Cell-Material Sciences (iCeMS)
Kyoto University
Yoshida-Ushinomiyacho, Sakyo, Kyoto, 606-8501, Japan

[+] These authors contributed equally to this work.

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

type of cyclic PIP was also demonstrated to downregulate androgen receptor-activated prostate-specific antigen expression in cell culture after its induction with dihydrotestosterone.[1c] In addition, a library of cyclic PIPs with several protecting groups on the γ-turns was evaluated based on their DNA-binding affinity, cell permeability, and cytotoxicity.[15b]

As noted, PIP should be able to recognize a long DNA sequence. However, five contiguous pyrrole or imidazole rings constitute an upper limit to retention of affinity for the minor groove.[16] Two approaches have been proposed to address this issue. One is the insertion of β-alanine;[17] the second is the design of a tandem hairpin type that links two hairpin PIPs in a turn-to-tail mode.[13a] Although there are fewer reports of tandem hairpin PIPs than of hairpin forms because of the relative synthetic inaccessibility of the former type of PIP, they have exhibited some fascinating characteristics. For example, they can be used in the fluorescent visualization of telomeres[5c, 18] and for alkylating telomeric sequences.[4h] Our laboratory improved the synthetic route to tandem hairpins[18b] and the longest DNA recognition sequence of 18 bp was accomplished by using a tandem trimer PIP.[18c] The challenge of developing even longer recognition systems is ongoing.

All of the three types of PIPs were shown to bind their predetermined DNA sequences based on the pairing rules.[2] However, looking in more detail at the motif revealed that adenine groups tended to stand out, which seemed strange considering that P/P was not expected to discriminate between adenine and thymine. Indeed, this is the very fact that had driven Dervan and our group to find a TA-selector.[2a, 2c, 19] The currently used Bind-n-Seq analytical method to represent PIP-bound enriched DNA high-throughput sequence data as motif figures is basically designed to identify binding sites of molecules or proteins[20] that have a single dominant DNA sequence as its target. In this respect, there is a critical difference for PIP, which has degeneracy of recognizing adenine and thymine. Therefore, we developed the Bind-n-Seq-based motif identification with reference sequence (Bind-n-Seq-MR), which was optimized for PIP-binding analysis. As a result, A and T appear with comparable base ratio in motif analyses. Furthermore, we corroborated the results of Bind-n-Seq by conducting SPR-binding assays.

Results and Discussion

Design and Synthesis

The hairpin design has been widely adopted to evaluate PIP derivatives because of the convenient preparation of such molecules. Although tandem hairpin and cyclic types of PIP have also been developed so far, it was difficult to design a tandem hairpin to fulfill the conditions necessary for comparison: The molecule should contain the same number of rings (*N*-methylpyrrole or *N*-methylimidazole) and it should target a sequence that was identical to that of hairpin and cyclic PIP. Therefore, we synthesized only hairpin and cyclic PIP. The sequence of PIP was designed to be unsymmetrical so that it would be possible to distinguish the binding orientation later. The biotin conjugate hairpin **1** was prepared by cleavage from the resin with 3,3′-diamino-*N*-methyldipropylamine and subsequent coupling with NHS-PEG_{12}-Biotin. Cyclic PIP **2** was obtained by cleavage from the resin with TFA, leaving a carboxylic acid moiety, cyclization of the cysteine and chloroacetyl group, and coupling with 3,3′-diamino-*N*-methyldipropylamine and NHS-PEG_{12}-Biotin. The structures of biotin-PIP conjugates **1** and **2** for NGS analysis, and diamine-PIP **3** and **4**, for SPR assay, are shown in **Scheme**. Comparative data obtained by using the Bind-n-Seq technique can be beneficial for future PIP designs. Furthermore, we explored the specificities of PIPs **1** and **2** by employing a comprehensive randomized DNA library as substrates.

Bind-n-Seq

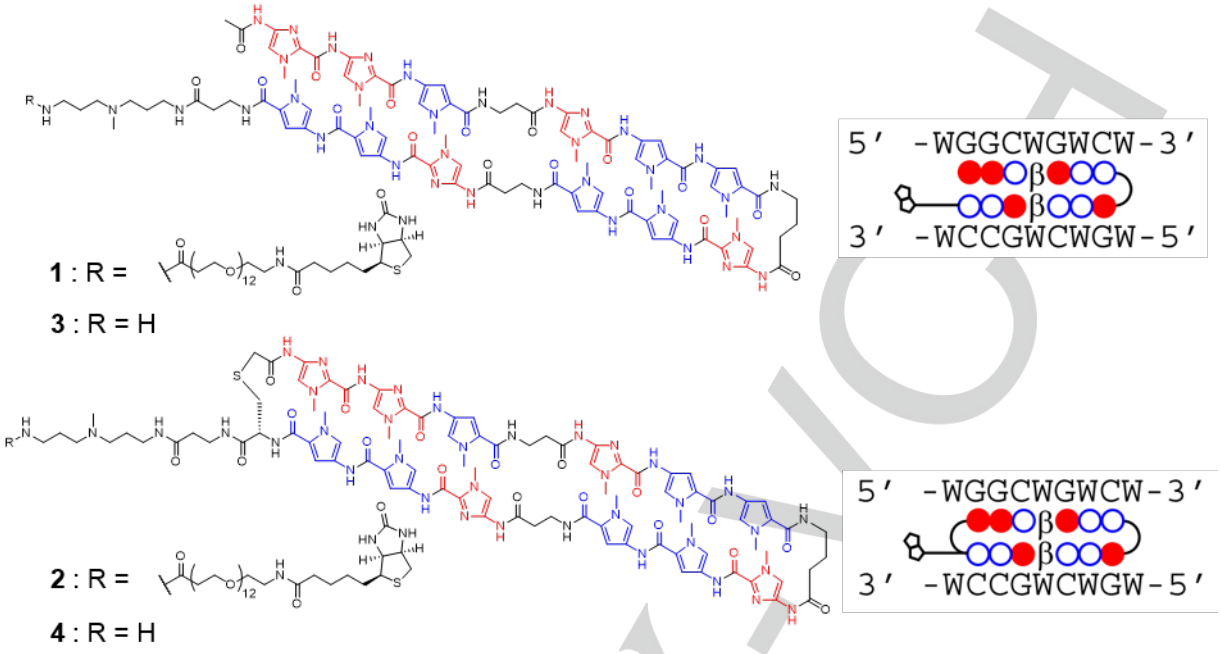
To compare the specificities of PIPs **1** and **2**, Bind-n-Seq[3e, 5b] experiments were performed with a library of oligomers containing randomized 21-mer, which were prepared according to a previous report.[6b] The motif results are presented in **Figures 1** and **2**. Nine bases for hairpin **1** and cyclic PIP **2** were extracted by using the MERMADE program.[20] The Bind-n-Seq mermade method of analysis gave the results presented in **Figures 1A** and **2A**. These motifs clearly show that both of the biotin conjugates bind to their target sites as expected based on the pairing rules. However, the apparent coincidence that the adenine groups shown in **Figures 1A** and **2A** were dominant in all the places where A and T were expected to appear equally raised questions regarding the validity of this analytical algorithm. We reasoned that the method produces motif results that rely mainly on the sequence ranked first while taking sequences with only one mismatch into consideration. This means that if the sequence ranked second has more than one mismatch from the first sequence, the second sequence would not be reflected in the motif result, which leads to overestimation of the sequence ranked first.

Thus, we developed Bind-n-Seq-MR to generate **Figures 1B** and **2B**. The procedure is as follows, accompanied by an explanation in the case of hairpin **1**.

(1) Decide the reference sequence to be read out, which is 5′-WGGCWGWCW-3′ in the case of hairpin **1**.

(2) Remove slipped sequences. This process is necessary to gain an equal sum of enrichment values among bases. For example, the sequence originally ranked third in **Table S1**, 5′-GACAGTCAC-3′, was omitted in Bind-n-Seq-MR analysis because the actual binding sequence should have been 5′-XGACAGTCA-3′, where X is an unknown base. Enrichment value is the ratio of the number of a certain sequenced DNA sequence derived from a sample PI polyamide versus that from an unspecific binding control by beads.

(3) Convert into the complementary sequence when the counterstrand (5′-WGWCWGCCW-3′) or its relevant sequence is ranked. The sequence originally ranked 30th in **Table S1**, 5′-AGTCAGGCA-3′, was converted into 5′-TGCCTGACT-3′, which is ranked 12th in **Figure 1C**. It should be noted that 5′-AG*T*CAG*G*CA-3′ is a two-base mismatch sequence, whereas 5′-TG*C*CTGACT-3′ is a one-base mismatch.



Scheme. Chemical structures of the three types of PIP-biotin conjugates: hairpin, cyclic, and tandem hairpin type, and their schematic representations with the target sequences. Biotinylated **1** and **2** were used in NGS; **3** and **4** were used for SPR analyses. W indicates A or T.

(4) Duplicate a sequence whose complementary sequence has the same number of mismatch bases and give half enrichment values to both the original and the complementary sequences. The sequence ranked second in **Table S1**, 5′-AG*A*CTGACA-3′, with an enrichment value of 69 has a one-base mismatch *A* shown in italic font. At the same time, its complementary sequence, 5′-TG*T*CAGTCT-3′, is also a one-base mismatch sequence. Although Bind-n-Seq cannot identify the orientation of PIP binding, we assumed that both orientations had equal possibilities and, accordingly, gave half enrichment values: the sequence ranked second, 5′-AGACTGACA-3′, and its complementary sequence, 5′-TGTCAGTCT-3′, in **Figure 1C** were given enrichment values of 35 (i.e., approximately half of 69).

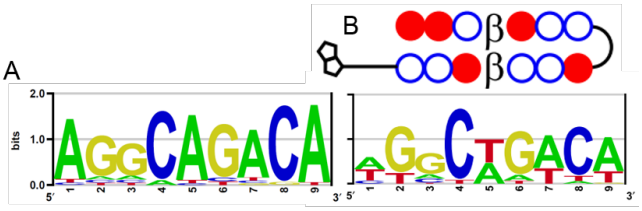
(5) Give each base an enrichment value of its sequence. In **Figure 1C**, every base in the sequence ranked first is given an enrichment value of 71.

(6) Sum the enrichment values allotted to each base. Thus, adenine at the 5′-end is given 550, which is the sum of enrichment values of ranks 1, 2, 4, 5, 7, 8, 9, 11, 16, 17, and 20.

(7) Produce motifs by enoLOGOS.[21] enoLOGOS converts the sum of enrichment values in **Figure 1C** into **Figure 1B**.

The application of the Bind-n-Seq mermade method and Bind-n-Seq-MR gave different results in the two cases (compare **Figures 1A** and **2A** with **Figures 1B** and **2B**). A particularly clear distinction is that the dominant adenines that were apparent after Bind-n-Seq mermade analysis became comparable frequencies of adenines and thymines after analysis of the raw data with Bind-n-Seq-MR. Those adenines are in the first, fifth, seventh, and ninth bases in **Figures 1A** and **2A**.

A



C

rank	sequence									enrichment
1	A	G	G	C	A	G	A	C	A	71
2	A	G	A	C	T	G	A	C	A	35
2	T	G	T	C	A	G	T	C	T	35
3	T	G	G	C	T	A	A	C	A	68
3	A	G	A	C	A	G	T	C	A	34
3	T	G	A	C	T	G	T	C	T	34
5	A	G	G	C	T	T	A	C	A	61
6	T	G	G	T	A	G	A	C	A	54
7	A	T	G	C	A	G	A	C	A	53
8	A	T	G	C	A	G	T	C	A	51
8	A	G	G	C	T	G	A	T	A	51
8	C	G	G	C	T	A	A	C	A	51
8	A	G	G	C	T	G	A	C	A	51
12	T	G	C	C	T	G	A	C	T	50
12	T	G	G	C	A	G	T	A	A	50
12	C	G	G	C	T	G	A	C	A	50
15	T	T	G	C	T	G	A	C	T	49
15	C	G	A	C	A	G	T	C	A	24.5
15	T	G	A	C	T	G	T	C	G	24.5
17	A	T	G	C	A	G	A	C	T	48
18	A	G	C	C	T	G	T	C	A	47
18	T	G	G	C	A	G	A	T	A	47
20	T	G	C	C	A	G	T	C	T	46

base	sum of enrichment values								
A	502	0	152	0	513	119	739	50	798
C	126	0	143	1030	0	0	0	936	0
G	0	883	755	0	0	904	0	0	25
T	457	201	35	54	571	61	346	98	262

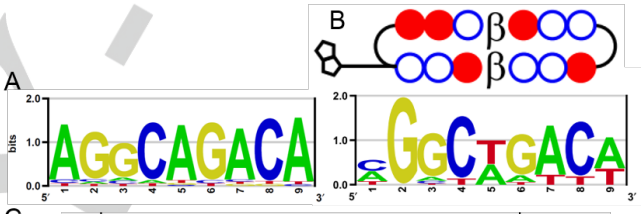
Figure 1. NGS results obtained with hairpin **1**. Motif analyses by using (A) the Bind-n-Seq mermade method and (B) the Bind-n-Seq-MR method. (C) Bind-n-Seq results obtained with **1**. The 9-mer sequences are shown in order of enrichment values.

The application of the Bind-n-Seq mermade method revealed that the two PIPs successfully bound to their target sequences as indicated in **Figures 1A** and **2A**. However, we were particularly intrigued by the observation that Bind-n-Seq mermade motif analyses showed that the adenines were dominant over thymines in all cases. This phenomenon probably occurred because Bind-n-Seq mermade analysis[20] is designed for DNA interactions with proteins and not with PIPs, which have P/P as a degenerate recognition pair for A/T and T/A, most likely with the same affinity. In an attempt to address these discrepancies, we developed Bind-n-Seq-MR as an alternative system of analysis. Whereas Bind-n-Seq mermade analysis reflects only the first rank and its one mismatch sequence, Bind-n-Seq-MR considers the top 20 ranks, irrespective of the number of mismatches that the ranked sequences contain. As a result, the dominant adenines were represented as comparable frequencies of adenines and thymines. Furthermore, thymines exceeded adenines in the fifth base position in **Figure 1B** and the fifth position in **Figure 2B**. The Bind-n-Seq-MR results of the top 1 to 5, 10, 20, 50, 100, and 1000 sequences for **1** and **2** are provided in **Figures S5** and **S6**. We selected the top 20 sequences as representative; although there is no clear criterion, when the number of sequences picked increased, the binding specificities became fuzzily represented: the difference of sizes between nucleobases in the motif figures became smaller and the binding characteristics would be difficult to extract.

Another significant difference between the two methods is that the adenine at the 5′-end was replaced by cytosine and adenine in **Figure 2A** and **2B**. This preference was unprecedented and could come from the sulfide-linked turn; however, further assessment of the generality of this hypothesis is required.

The plausible target sequence is 5′-WGGCWGWCW-3′, which represents 16 possible sequences because it contains four W positions (A or T). We therefore addressed the question: What were the ranks of those sequences? The results are shown in **Table S3**. Although the top sequence was the target for both PIPs, it was found that the least-enriched target sequence in the case of cyclic **2** was ranked 3,997 and that the least-enriched target sequence in the case of hairpin **1** was not even discovered. This result might suggest the inevitable off-target effects, although the levels of those effects should depend on the target sequence as well as on high-order structures, especially in cellular environments.

A



C

rank	sequence									enrichment
1	A	G	G	C	A	G	A	C	A	138
2	C	G	G	C	A	G	A	C	A	123
3	C	G	G	C	T	G	A	C	A	113
4	C	G	G	C	T	G	A	C	T	106
5	A	G	A	C	A	G	T	C	A	51
5	T	G	A	C	T	G	T	C	T	51
6	A	G	G	C	T	T	A	C	A	96
7	C	G	G	C	A	G	A	C	T	94
8	A	G	T	C	T	G	T	C	A	47
8	T	G	A	C	A	G	A	C	T	47
9	A	G	G	C	T	G	A	C	A	92
10	A	G	G	C	T	G	A	C	T	91
11	C	G	G	C	T	A	A	C	A	89
12	A	G	C	C	A	G	T	C	A	88
13	C	G	G	T	A	G	A	C	T	88
14	C	G	G	C	T	G	A	T	A	87
15	A	G	G	C	A	G	A	C	T	85
16	T	G	G	C	T	A	A	C	A	83
17	C	G	G	C	A	G	T	T	A	82
18	C	G	G	C	T	A	A	C	T	82
19	T	G	G	T	A	G	A	C	A	80
20	A	G	G	C	T	G	A	T	A	78

base	sum of enrichment values								
A	767	0	148	0	877	255	1576	0	1249
C	866	0	88	1725	0	0	0	1646	0
G	0	1894	1611	0	0	1543	0	0	0
T	261	0	47	169	1017	96	318	248	645

Figure 2. NGS results obtained with cyclic **2**. Motif analyses by using (A) the Bind-n-Seq mermade method and (B) the Bind-n-Seq-MR method. (C) Bind-n-seq results obtained with **2**. The 9-mer sequences are shown in order of enrichment values.

SPR Assay

Although the Bind-n-Seq results obtained using our new protocol seem reasonable because the P/P pair was expected to have no ability to differentiate A from T, the ambiguities highlighted in the

previous section prompted us to examine the affinities of hairpin **3** and cyclic **4** against ODN1 and ODN2. Hairpin **3** and cyclic **4** are precursors of biotin conjugates **1** and **2**, respectively. ODN1 (5′-Biotin-GCG**AGGCAGACA**CGCTTTTGCGTGTCTGCCTCGC) was designed to include the sequence ranked first (shown in bold). Coincidentally, the sequence ranked first from both hairpin **1** and cyclic **2** was the same, namely 5′-AGGCAGACA-3′. ODN2 (5′-Biotin-GCG**T**GGC**T**GT**C**T**T**CGCTTTTGCGAGACAGCCACGC) was composed of a sequence based on ODN1, but with every adenine position shown in bold in ODN1 being replaced by thymine, shown with underline in ODN2. The idea behind this design was to investigate whether P/P pairs in PIPs **3** and **4** have different affinities against adenine and thymine. The results presented in **Figures 1A** and **2A** suggest that both hairpin **3** and cyclic **4** would be expected to have much lower affinities toward ODN2 than ODN1. However, if **Figures 1B** and **2B** represent the binding characteristics more accurately, then the K_D values in the two cases would not be expected to differ significantly.

Sensorgrams of hairpin **3** and cyclic **4** against ODN1 immobilized on a streptavidin-functionalized SA sensor chip are shown in **Figure 3**. By fitting the sensorgrams in a predefined 1:1 binding model with mass transfer, association rate constants k_a, dissociation rate constants k_d, and equilibrium dissociation constants K_D were obtained. The k_a, k_d, and K_D values of hairpin **3** were 2.4 × 10^5 M^{-1} s^{-1}, 1.4 × 10^{-3} s^{-1}, and 5.9 nM, respectively, and those of cyclic **4** were 2.2 × 10^5 M^{-1} s^{-1}, 1.6 × 10^{-3} s^{-1}, and 7.0 nM, respectively.

In the case of ODN2, the k_a, k_d, and K_D values of hairpin **3** were 3.7 × 10^5 M^{-1} s^{-1}, 2.0 × 10^{-3} s^{-1}, and 5.4 nM, respectively, and those of cyclic **4** were 2.0 × 10^5 M^{-1} s^{-1}, 1.7 × 10^{-3} s^{-1}, and 8.6 nM, respectively. These kinetic and thermodynamic values are summarized in **Table 1**. The comparable K_D values for ODN1 and ODN2 with both hairpin **3** and cyclic **4** reinforced the validity of the Bind-n-Seq-MR analysis.

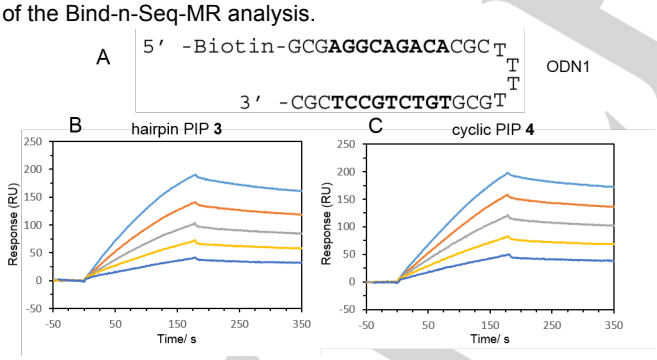


Figure 3. SPR-binding assays conducted to evaluate the binding properties of hairpin **3** vs. cyclic PIP **4**. (A) The sequence of ODN1, which has 5′-AGGCAGACA-3′ on the 5′ side. (B) SPR sensorgrams for interactions of the hairpin PIP. The concentrations were 100 nM (blue), 85 nM (orange), 70 nM (gray), 55 nM (yellow), and 40 nM (dark blue). (C) SPR sensorgrams for interactions of the cyclic PIP. The concentrations were 300 nM (blue), 250 nM (orange), 200 nM (gray), 150 nM (yellow), and 100 nM (dark blue).

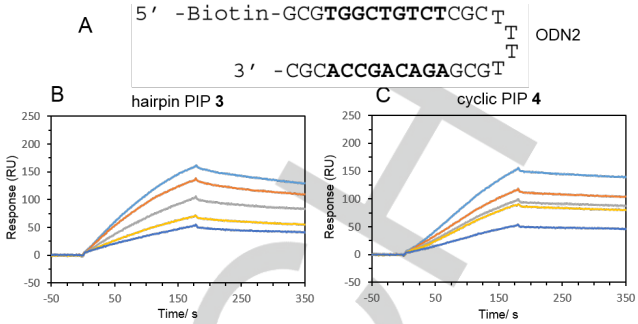


Figure 4. SPR-binding assays conducted to evaluate the binding properties of the hairpin **3** vs. cyclic PIP **4**. (A) The sequence of ODN2, which has 5′-TGGCTGTCT-3′ on the 5′ side. (B) SPR sensorgrams for interactions of the hairpin PIP. The concentrations were 100 nM (blue), 85 nM (orange), 70 nM (gray), 55 nM (yellow), and 40 nM (dark blue). (C) SPR sensorgrams for interactions of the cyclic PIP. The concentrations were 90 nM (blue), 80 nM (orange), 70 nM (gray), 60 nM (yellow), and 40 nM (dark blue).

Table 1. The values of the association rates (k_a) and dissociation rates (k_d) obtained from curve fittings of sensorgrams, and dissociation constants (K_D).

	ODN1			ODN2		
	k_a [M^{-1} s^{-1}]	k_d [s^{-1}]	K_D [nM]	k_a [M^{-1} s^{-1}]	k_d [s^{-1}]	K_D [nM]
hairpin **3**	2.4 x 10^5	1.4 x 10^{-3}	5.9	3.7 x 10^5	2.0 x 10^{-3}	5.4
cyclic **4**	2.2 x 10^5	1.6 x 10^{-3}	7.0	2.0 x 10^5	1.7 x 10^{-3}	8.6

Conclusions

For the purpose of comparison and characterization of hairpin, and cyclic PIPs, a Bind-n-Seq assay was executed with a library containing 21-mer randomized sequences. It was found that both PIP types exhibited reasonable specificities. However, it was necessary to reexamine the analytical mermade method because its approach has been designed to assess protein–DNA interactions, and its concept does not fit to the cases of PI polyamide. Binding characteristics is considered to follow the pairing rules, which entail P/P cannot distinguish A/T from T/A. In fact, the results of binding motifs produced by the mermade method with overwhelmingly large adenines over thymines, were revised by the Bind-n-Seq-MR analysis to solve that problem. Although the generality of the Bind-n-Seq-MR should be guaranteed by more examples of PI polyamide, the validity was confirmed by SPR assays and this program will be useful in the future for unbiased evaluation of DNA sequence-selectivity of PI polyamide binding.

Experimental Section

General

Reagents and solvents were purchased from standard suppliers and used without further purification. Flash column purifications were

performed by a CombiFlash Rf (Teledyne Isco, Inc.) with C18 RediSep Rf 4.3 Gram Flash Column. HPLC purifications were performed with a JASCO PU-2089 Plus pump and a JASCO UV-2075 Plus detector set at 254 nm. A Chemcobond 5-ODS-H column (4.6 mm I.D. ×150 mm) and COSMOSIL 5C_{18}-MS-II column (10 mm I.D. ×150 mm) were equipped and the mobile phase consisted of a gradient of acetonitrile in 0.1□% (v/v) TFA in water at a flow rate of 1.0 mL min^{-1} and 3.0 mL min^{-1} respectively with detection at 254 nm. ESI-TOF MS was performed on a Bio-TOF II (Bruker Daltonics) mass spectrometer by using positive ionization mode. Machine-assisted polyamide syntheses were performed on a PSSM-8 (Shimadzu) system with computer-assisted operation on a 40 μmol scale by using Fmoc chemistry. Bind-n-Seq analyses were conducted with Ion PGM™ System (ThermoFischer Scientific).

Fmoc Solid-Phase Peptide Synthesis of PI Polyamides.

The PI polyamides were prepared by Fmoc solid-phase machine-assisted synthesis procedure.[4e, 18c]

Synthesis of Hairpin PI Polyamide 3 and Its Biotin Conjugate 1 (AcIIP-β-IP-P-γ-IPP-β-IPP-β-diamine-PEG$_{12}$-biotin)

AcIIP-β-IPP-γ-IPP-β-IPP-β-Wang resin was treated with 400 μL of 3,3'-diamino-*N*-methyldipropylamine for 3 h at 45 °C, washed with dichloromethane, and the solvent was evaporated. The residue was dissolved in the minimum amount of dichloromethane and subsequently washed with diethyl ether to produce a 27 mg of white powder crude, which was used in next condensation reaction without further purification. But PIP **4** for SPR analysis was purified by flash column chromatography.

ESI-TOF-MS *m/z* calcd for $C_{89}H_{110}N_{36}O_{17}^{3+}$ [*M* + 3H]$^{3+}$ 652.6356, found 652.6329. To a solution of the crude PI polyamide (2.3 mg) in DMF (150 μL), DIEA (0.8 μL, 4.7 μmol) and NHS-PEG$_{12}$-Biotin (4.0 mg, 4.2 μmol, ThermoScientific) were added. The reaction mixture was incubated overnight at room temperature. Evaporation of the solvent gave a yellow oil, which was purified by HPLC to afford the hairpin PI Polyamide-biotin conjugate (0.8 mg, 0.29 μmol, 9.3%) as a light yellow powder. ESI-TOF-MS *m/z* calcd for $C_{126}H_{177}N_{39}O_{32}S^{3+}$ [*M* + 3H]$^{3+}$ 927.7787, found 927.7815.

Synthesis of Cyclic PI Polyamide 4 and Its Biotin Conjugate 2 (cyclo-(AcIIP-β-IPP-γ-IPP-β-IPP-(*R*)-Cys)-β-diamine-PEG$_{12}$-biotin)

AcIIP-β-IPP-γ-IPP-β-IPP-(*R*)-Cys-β-Wang resin was treated with a solution of trifluoroacetic acid 950 μL, triisopropylsilane 25 μL and water 25 μL for 30 min at room temperature. Precipitation by pouring the solution into diethyl ether obtained 22.3 mg of a brownish powder. To a solution of 21.2 mg of the crude in 2.0 mL DMF was added 40 μL of DIEA. Subsequently 5.0 mg of PyBOP ((Benzotriazol-1-yloxy)tripyrrolidinophosphonium hexafluorophosphate) and 10 μL of 3,3'-diamino-*N*-methyldipropylamine were added to that solution. After confirmation of the conversion, the product was precipitated by diethyl ether. 0.6 mg of cyclo-(AcIIP-β-IPP-γ-IPP-β-IPP-(*R*)-Cys)-β-diamine **5** was obtained from 10 mg of the crude polyamide after HPLC purification. ESI-TOF-MS *m/z* calcd for $C_{92}H_{116}N_{37}O_{18}S^{3+}$ [*M* + 3H]$^{3+}$ 686.3001, found 686.2964.

0.6 mg, 0.3 μmol of this amine was dissolved in 3 μL of DMF. To that solution were added NHS-PEG$_{12}$-Biotin DMF solution (0.3 mg, 0.3 μmol, 3 μL, 1eq.) and DIEA (0.3 μL, 2 μmol, 6eq.). Reaction proceeded quickly and the product was precipitated with diethyl ether and purified by HPLC (0.6 mg, 0.2 μmol, 67%). ESI-TOF-MS *m/z* calcd for $C_{129}H_{183}N_{40}O_{33}S_2^{3+}$ [*M* + 3H]$^{3+}$ 961.4432, found 961.4358.

Bind-n-Seq Analysis

Bind-n-Seq analysis were proceeded as previously reported.[6b] Briefly, oligonucleotides consisting of Ion torrent sequencing library adapter A1, Ion Express Barcode, 21-mer randomized sequence and another adapter P1 were purchased from Sigma-Aldrich and duplexed via primer extension; 500 nM of PIP-biotin conjugates were allowed to bind 1.5 μM of the library in 50 μL for 16 h at rt; enrichment by Streptavidin M-280 Dynabeads and elution from on the beads; PCR amplification, quantification by 2100 Bioanalyzer (Agilent Technologies) and emulsion PCR on Ion OneTouch 2 system (Thermo Fisher Scientific); sequencing by Ion Torrent PGM 318 chip (Thermo Fisher Scientific). The results were primarily analyzed by Bind-n-Seq analysis pipeline and the motifs were processed by enoLOGOS[21] (URL: http://www.benoslab.pitt.edu/cgi-bin/enologos/enologos.cgi).

Bind-n-Seq-MR Analysis.

The detailed information is described in "RESULTS" section. The Bind-n-Seq-MR program is available via H. S.

SPR Assay

SPR experiments were performed on a Biacore X instrument (GE Healthcare) according to previous reports.[4e, 6a] Biotinylated hairpin ODN1 (5'-Biotin-GCGAGGCAGACACGCTTTTGCGTGTCTGCCTCGC) and ODN2 (5'-Biotin-GCGTGGCTGTCTCGCTTTTGCGAGACAGCCACGC) were purchased from Japan Bio Services Co., LTD.（Saitama, Japan). The biotinylated DNA was immobilized to a streptavidin-functionalized SA sensor chip to obtain the desired immobilization level (approximately 1000 RU rise). SPR measurements were carried out using degassed and filtered HBS buffer (10 mM HPES pH 7.4, 150 mM NaCl, 3 mM EDTA, and 0.005% Surfactant P20) with 0.1% DMSO at 25 °C. A series of sample solutions with a wide range of concentrations were prepared in the buffer with 0.1% DMSO and injected at a flow rate of 20 μL/min. After each cycle, the samples remainig on the DNA was removed with 50 mM NaOH/1 M NaCl solution until the baseline of the sensorgrams was recovered. Optimized concentration range was then selected for the subsequent quantitive anlalysis. The resulting sensorgrams were fitted with 1:1 binding fitting model using BIAevaluation 4.1 program in order to obtain the values of kinetical parameters (k_a and k_d) and binding affinites (K_D).

Acknowledgements

This work was partially supported by JSPS KAKENHI (grant number 24225005 to H. S., 24310155 to T. B. and 15J00928 to Y. K.), Basic Science and Platform Technology Program for Innovative Biological Medicine from Japan Agency for Medical Research and Development, AMED, and JSPS-NSF International Collaborations in Chemistry (ICC) to H.S.

Conflict of Interest

The authors declare no conflicts.

Keywords: polyamide • NGS • Bind-n-Seq • small molecule •

[1] a) N. G. Nickols, P. B. Dervan, *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104*, 10418-10423; b) N. G. Nickols, C. S. Jacobs, M. E. Farkas, P. B. Dervan, *ACS Chem. Biol.* **2007**, *2*, 561-571; c) D. M. Chenoweth, D. A. Harki, J. W. Phillips, C. Dose, P. B. Dervan, *J. Am. Chem. Soc.* **2009**, *131*, 7182-7188; d) J. A. Raskatov, J. L. Meier, J. W. Puckett, F. Yang, P. Ramakrishnan, P. B. Dervan, *Proc. Natl. Acad. Sci.U. S. A.* **2012**, *109*, 1023-1028; e) J. A. Raskatov, N. G. Nickols, A. E. Hargrove, G. K. Marinov, B. Wold, P. B. Dervan, *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109*, 16041-16045; f) G. N. Pandian, Y. Nakano, S. Sato, H. Morinaga, T. Bando, H. Nagase, H. Sugiyama, *Sci. Rep.* **2012**, *2*, 544; g) N. G. Nickols, J. O. Szablowski, A. E. Hargrove, B. C. Li, J. A. Raskatov, P. B. Dervan, *Mol. Cancer Ther.* **2013**, *12*, 675-684; h) J. Syed, Ganesh N. Pandian, S. Sato, J. Taniguchi, A. Chandran, K. Hashiya, T. Bando, H. Sugiyama, *Chem. Biol.* **2014**, *21*, 1370-1380; i) L. Han, G. N. Pandian, A. Chandran, S. Sato, J. Taniguchi, G. Kashiwazaki, Y. Sawatani, K. Hashiya, T. Bando, Y. Xu, X. Qian, H. Sugiyama, *Angew. Chem. Int. Ed.* **2015**, *54*, 8700-8703; j) K. Hiraoka, T. Inoue, R. D. Taylor, T. Watanabe, N. Koshikawa, H. Yoda, K. Shinohara, A. Takatori, H. Sugimoto, Y. Maru, T. Denda, K. Fujiwara, A. Balmain, T. Ozaki, T. Bando, H. Sugiyama, H. Nagase, *Nat. Commun.* **2015**, *6*, 6706; k) J. Igarashi, N. Fukuda, T. Inoue, S. Nakai, K. Saito, K. Fujiwara, H. Matsuda, T. Ueno, Y. Matsumoto, T. Watanabe, H. Nagase, T. Bando, H. Sugiyama, T. Itoh, M. Soma, *PLoS One* **2015**, *10*, e0125295; l) J. Syed, A. Chandran, G. N. Pandian, J. Taniguchi, S. Sato, K. Hashiya, G. Kashiwazaki, T. Bando, H. Sugiyama, *ChemBioChem* **2015**, *16*, 1497-1501.

[2] a) C. L. Kielkopf, S. White, J. W. Szewczyk, J. M. Turner, E. E. Baird, P. B. Dervan, D. C. Rees, *Science* **1998**, *282*, 111-115; b) S. White, J. W. Szewczyk, J. M. Turner, E. E. Baird, P. B. Dervan, *Nature* **1998**, *391*, 468-471; c) P. B. Dervan, B. S. Edelson, *Curr. Opin. Struct. Biol.* **2003**, *13*, 284-299; d) C. F. Hsu, J. W. Phillips, J. W. Trauger, M. E. Farkas, J. M. Belitsky, A. Heckel, B. Z. Olenyuk, J. W. Puckett, C. C. C. Wang, P. B. Dervan, *Tetrahedron* **2007**, *63*, 6146-6151.

[3] a) J. W. Trauger, E. E. Baird, P. B. Dervan, *Nature* **1996**, *382*, 559-561; b) S. White, E. E. Baird, P. B. Dervan, *Chem. Biol.* **1997**, *4*, 569-578; c) J. W. Trauger, P. B. Dervan, *Methods in Enzymology,* **2001**, *340,* 450-466; d) C. D. Carlson, C. L. Warren, K. E. Hauschild, M. S. Ozers, N. Qadir, D. Bhimsaria, Y. Lee, F. Cerrina, A. Z. Ansari, *Proc. Natl. Acad. Sci. U. S. A.* **2010**, *107*, 4544-4549; e) J. L. Meier, A. S. Yu, I. Korf, D. J. Segal, P. B. Dervan, *J. Am. Chem. Soc.* **2012**, *134*, 17814-17822; f) N. R. Wurtz, P. B. Dervan, *Chem. Biol.* **2000**, *7*, 153-161.

[4] a) Y. D. Wang, J. Dziegielewski, N. R. Wurtz, B. Dziegielewska, P. B. Dervan, T. A. Beerman, *Nucleic Acids Res.* **2003**, *31*, 1208-1215; b) T. Bando, H. Sugiyama, *Acc. Chem. Res.* **2006**, *39*, 935-944; c) S. M. Tsai, M. E. Farkas, C. J. Chou, J. M. Gottesfeld, P. B. Dervan, *Nucleic Acids Res.* **2007**, *35*, 307-316; d) T. Yoshidome, M. Endo, G. Kashiwazaki, K. Hidaka, T. Bando, H. Sugiyama, *J. Am. Chem. Soc.* **2012**, *134*, 4654-4660; e) S. Asamitsu, Y. Kawamoto, F. Hashiya, K. Hashiya, M. Yamamoto, S. Kizaki, T. Bando, H. Sugiyama, *Bioorg. Med. Chem.* **2014**, *22*, 4646-4657; f) R. D. Taylor, S. Asamitsu, T. Takenaka, M. Yamamoto, K. Hashiya, Y. Kawamoto, T. Bando, H. Nagase, H. Sugiyama, *Chem. Eur. J.* **2014**, *20*, 1310-1317; g)R. D. Taylor, Y. Kawamoto, K. Hashiya, T. Bando, H. Sugiyama, *Chem. Asian J.* **2014**, *9*, 2527-2533; h) M. Yamamoto, T. Bando, Y. Kawamoto, R. D. Taylor, K. Hashiya, H. Sugiyama, *Bioconj. Chem.* **2014**, *25*, 552-559; i) R. D. Taylor, A. Chandran, G. Kashiwazaki, K. Hashiya, T. Bando, H. Nagase, H. Sugiyama, *Chem. Eur. J.* **2015**, *21*, 14996-15003; j) A. Y. Chang, P. B. Dervan, *J. Am. Chem. Soc.Com* **2000**, *122*, 4856-4864.

[5] a) D. S. Pilch, N. Poklar, C. A. Gelfand, S. M. Law, K. J. Breslauer, E. E. Baird, P. B. Dervan, *Proc. Natl. Acad. Sci.* **1996**, *93*, 8306-8311; b) J. S. Kang, J. L. Meier, P. B. Dervan, *J. Am. Chem. Soc.* **2014**, *136*, 3687-3694; c) A. Hirata, K. Nokihara, Y. Kawamoto, T. Bando, A. Sasaki, S. Ide, K. Maeshima, T. Kasama, H. Sugiyama, *J. Am. Chem. Soc.* **2014**, *136*, 11546-11554; d) C. Guo, Y. Kawamoto, S. Asamitsu, Y. Sawatani, K. Hashiya, T. Bando, H. Sugiyama, *Bioorg. Med. Chem.* **2015**, *23*, 855-860.

[6] a) E. R. Lacy, N. M. Le, C. A. Price, M. Lee, W. D. Wilson, *J. Am. Chem. Soc.* **2002**, *124*, 2153-2163; b) C. Anandhakumar, Y. Li, S. Kizaki, G. N. Pandian, K. Hashiya, T. Bando, H. Sugiyama, *ChemBioChem* **2014**, *15*, 2647-2651; c) Y.-W. Han, Y. Tsunaka, H. Yokota, T. Matsumoto, G. Kashiwazaki, H. Morinaga, K. Hashiya, T. Bando, H. Sugiyama, Y. Harada, *Biomater. Sci.* **2014**, *2*, 297-307; d) M. Yamamoto, T. Bando, H. Morinaga, Y. Kawamoto, K. Hashiya, H. Sugiyama, *Chem. Eur. J.* **2014**, *20*, 752-759.

[7] C. Anandhakumar, S. Kizaki, T. Bando, G. N. Pandian, H. Sugiyama, *ChemBioChem* **2015**, *16*, 20-38.

[8] C. L. Warren, N. C. Kratochvil, K. E. Hauschild, S. Foister, M. L. Brezinski, P. B. Dervan, G. N. Phillips, Jr., A. Z. Ansari, *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103*, 867-872.

[9] J. W. Puckett, K. A. Muzikar, J. Tietjen, C. L. Warren, A. Z. Ansari, P. B. Dervan, *J. Am. Chem. Soc.* **2007**, *129*, 12310-12319.

[10] G. S. Erwin, D. Bhimsaria, A. Eguchi, A. Z. Ansari, *Angew. Chem. Int. Ed.* **2014**, *53*, 10124-10128.

[11] a) M. Mrksich, M. E. Parks, P. B. Dervan, *J. Am. Chem. Soc.* **1994**, *116*, 7983-7988; b) R. P. L. de Clairac, B. H. Geierstanger, M. Mrksich, P. B. Dervan, D. E. Wemmer, *J. Am. Chem. Soc.* **1997**, *119*, 7909-7916.

[12] a) D. M. Herman, J. M. Turner, E. E. Baird, P. B. Dervan, *J. Am. Chem. Soc.* **1999**, *121*, 1121-1129; b) D. M. Chenoweth, P. B. Dervan, *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106*, 13175-13179; c) H. Morinaga, T. Bando, T. Takagaki, M. Yamamoto, K. Hashiya, H. Sugiyama, *J. Am. Chem. Soc.* **2011**, *133*, 18924-18930.

[13] a) D. M. Herman, E. E. Baird, P. B. Dervan, *Chem. Eur. J.* **1999**, *5*, 975-983; b) I. Kers, P. B. Dervan, *Bioorg. Med. Chem.* **2002**, *10*, 3339-3349.

[14] a) S. E. Swalley, E. E. Baird, P. B. Dervan, *Chem. Eur. J.* **1997**, *3*, 1600-1607; b) A. R. Urbach, P. B. Dervan, *Proc. Natl. Acad. Sci. U. S. A.* **2001**, *98*, 4343-4348; c) M. Minoshima, T. Bando, S. Sasaki, K.-i. Shinohara, T. Shimizu, J. Fujimoto, H. Sugiyama, *J. Am. Chem. Soc.* **2007**, *129*, 5384-5390; d) G. Kashiwazaki, T. Bando, K.-i. Shinohara, M. Minoshima, S. Nishijima, H. Sugiyama, *Bioorg. Med. Chem.* **2009**, *17*, 1393-1397; e) G. Kashiwazaki, T. Bando, K.-i. Shinohara, M. Minoshima, H. Kumamoto, S. Nishijima, H. Sugiyama, *Bioorg. Med. Chem.* **2010**, *18*, 2887-2893. f) R. E. Wang, R. K. Pandita, J. Cai, C. R. Hunt, J.-S. Taylor, *ChemBioChem* **2012**, *13*, 97-104.

[15] a) J. Cho, M. E. Parks, P. B. Dervan, *Proc. Natl. Acad. Sci. U. S. A.* **1995**, *92*, 10389-10392; b) B. C. Li, D. C. Montgomery, J. W. Puckett, P. B. Dervan, *J. Org. Chem.* **2013**, *78*, 124-133.

[16] J. M. Turner, E. E. Baird, P. B. Dervan, *J. Am. Chem. Soc.* **1997**, *119*, 7636-7644.

[17] J. M. Turner, S. E. Swalley, E. E. Baird, P. B. Dervan, *J. Am. Chem. Soc.* **1998**, *120*, 6219-6226.

[18] a) K. Maeshima, S. Janssen, U. K. Laemmli, *EMBO J.* **2001**, *20*, 3218-3228; b) Y. Kawamoto, T. Bando, F. Kamada, Y. Li, K. Hashiya, K. Maeshima, H. Sugiyama, *J. Am. Chem. Soc.* **2013**, *135*, 16468-16477; c) Y. Kawamoto, A. Sasaki, K. Hashiya, S. Ide, T. Bando, K. Maeshima, H. Sugiyama, *Chem. Sci.* **2015**, *6*, 2307-2312.

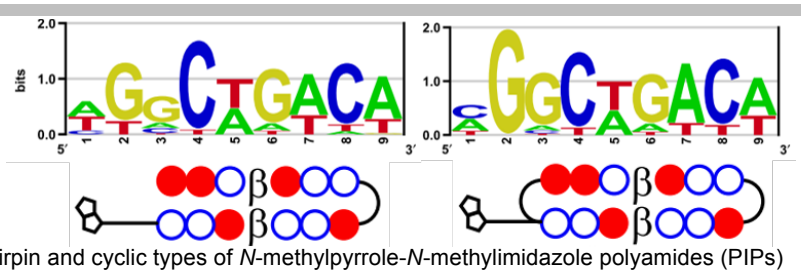
[19] a) C. L. Kielkopf, R. E. Bremer, S. White, J. W. Szewczyk, J. M. Turner, E. E. Baird, P. B. Dervan, D. C. Rees, *J. Mol. Biol.* **2000**, *295*, 557-567; b) C. A. Briehn, P. Weyermann, P. B. Dervan, *Chem. Eur. J.* **2003**, *9*, 2110-2122; c) W. Zhang, T. Bando, H. Sugiyama, *J. Am. Chem. Soc.* **2006**, *128*, 8766-8776; d) W. Zhang, M. Minoshima, H. Sugiyama, *J. Am. Chem. Soc.* **2006**, *128*, 14905-14912; e) W. Zhang, S.-K. Jiang, Y.-L. Wu, C.-X. Guo, H.-F. Zhang, H. Sugiyama, X.-L. Chen, *ChemBioChem* **2012**, *13*, 47-50; f) H.-F. Zhang, Y.-L. Wu, S.-K. Jiang, P. Wang, H. Sugiyama, X.-L. Chen, W. Zhang, Y.-J. Ji, C.-X. Guo, *ChemBioChem* **2012**, *13*, 1366-1374.

[20] A. Zykovich, I. Korf, D. J. Segal, *Nucleic Acids Res.* **2009**, *37*, e151.

[21] C. T. Workman, Y. Yin, D. L. Corcoran, T. Ideker, G. D. Stormo, P. V. Benos, *Nucleic Acids Res.* **2005**, *33*, W389-W392.

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

FULL PAPER



Hairpin and cyclic types of *N*-methylpyrrole-*N*-methylimidazole polyamides (PIPs) with biotin were synthesized, and their binding specificities against randomized DNA sequences were evaluated by using next-generation sequencing. Both PIP types were shown to recognize DNA sequences predetermined by the pairing rules. Along with these results, we propose a new analytical method for Bind-n-Seq that is particularly relevant for analysis of PIP binding.

G. Kashiwazaki, A. Chandran, S. Asamitsu, T. Kawase, Y. Kawamoto, Y. Sawatani, K. Hashiya, T. Bando, H. Sugiyama**

Page No. – Page No.

Comparative Analysis of DNA-Binding Selectivity of Hairpin and Cyclic Pyrrole-Imidazole Polyamides Based on Next-Generation Sequencing

Supporting Information

1. Analytical Data

Hairpin PI Polyamide 3 and Its Biotin Conjugate 1 (AcIIP-β-IP-P-γ-IPP-β-IPP-β-diamine-PEG_{12}-biotin)

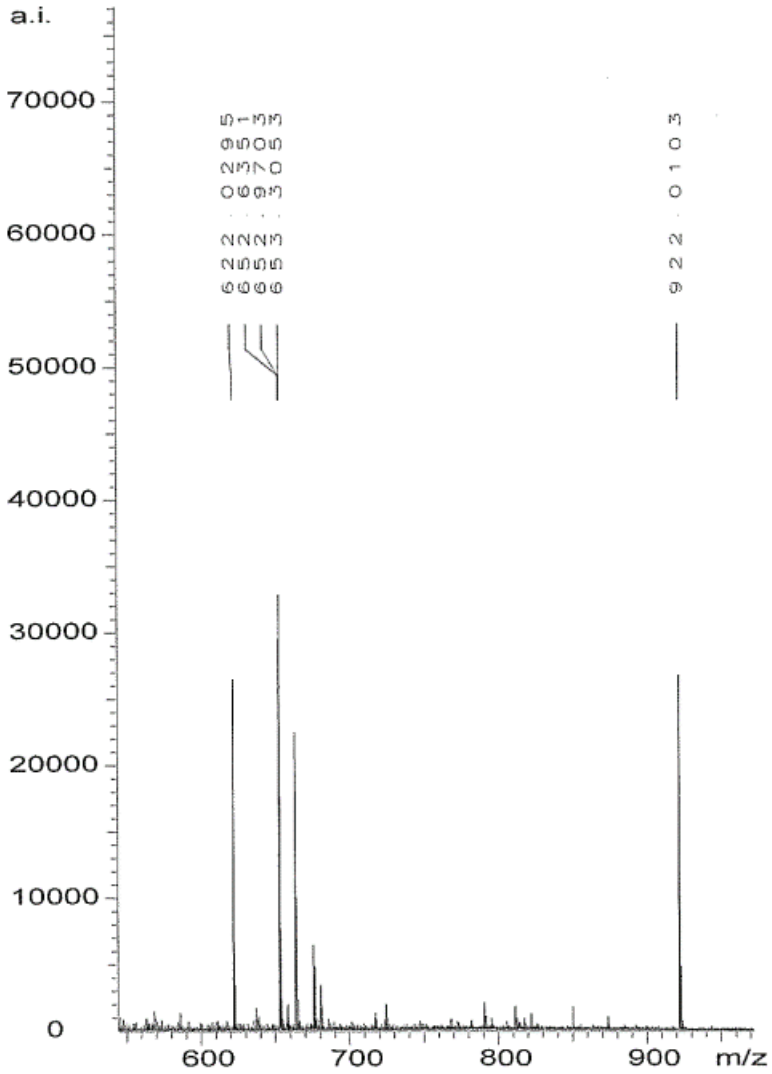


Figure S1. ESI-TOFMS spectrum of hairpin PI polyamide **3**. *m/z* calcd for $C_{89}H_{110}N_{36}O_{17}^{3+}$ $[M + 3H]^{3+}$ 652.6356, found 652.6329. The two peaks, 622.0295 and 922.0103, are the calibration peaks.

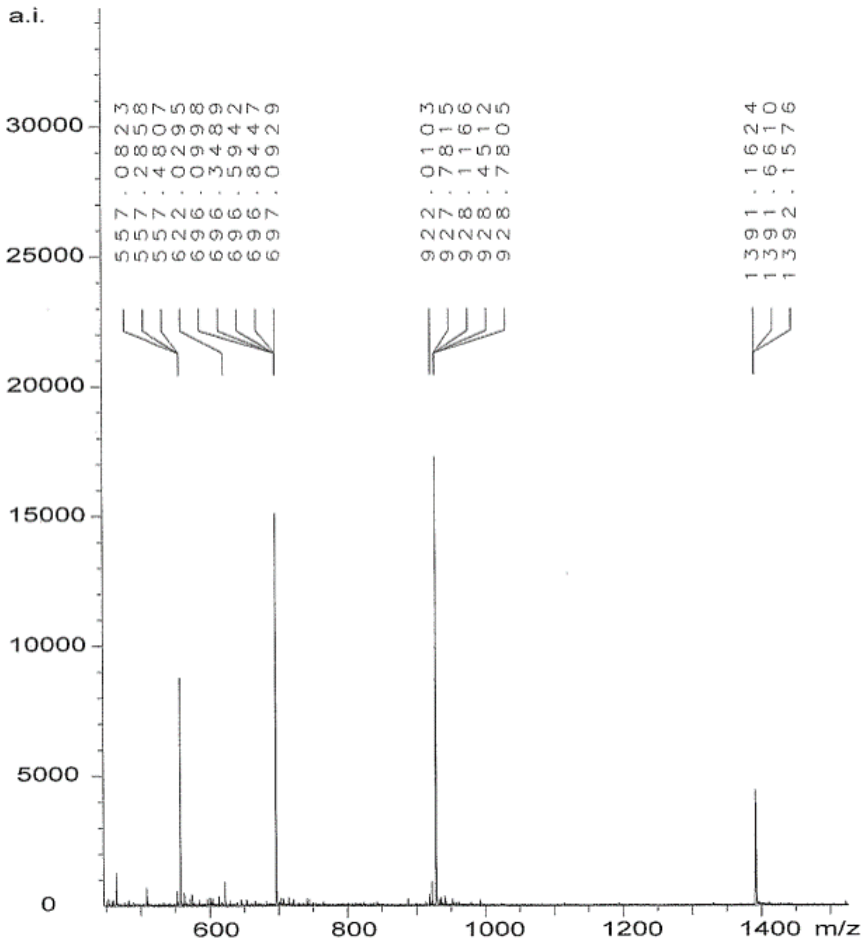


Figure S2. ESI-TOFMS spectrum of hairpin PI polyamide **1**. *m/z* calcd for $C_{126}H_{182}N_{39}O_{32}S^{4+}$ $[M + 5H]^{5+}$ 557.0701, found 557.0823. *m/z* calcd for $C_{126}H_{181}N_{39}O_{32}S^{4+}$ $[M + 4H]^{4+}$ 696.0858, found 696.0998. *m/z* calcd for $C_{126}H_{180}N_{39}O_{32}S^{3+}$ $[M + 3H]^{3+}$ 927.7787, found 927.7815. *m/z* calcd for $C_{126}H_{179}N_{39}O_{32}S^{2+}$ $[M + 2H]^{2+}$ 1391.1644, found 1391.1624. The two peaks, 622.0295 and 922.0103, are the calibration peaks.

Cyclic PI Polyamide 4 and Its Biotin Conjugate 2 (cyclo-(AcIIP-β-IPP-γ-IPP-β-IPP-(*R*)-Cys)-β-diamine-PEG$_{12}$-biotin)

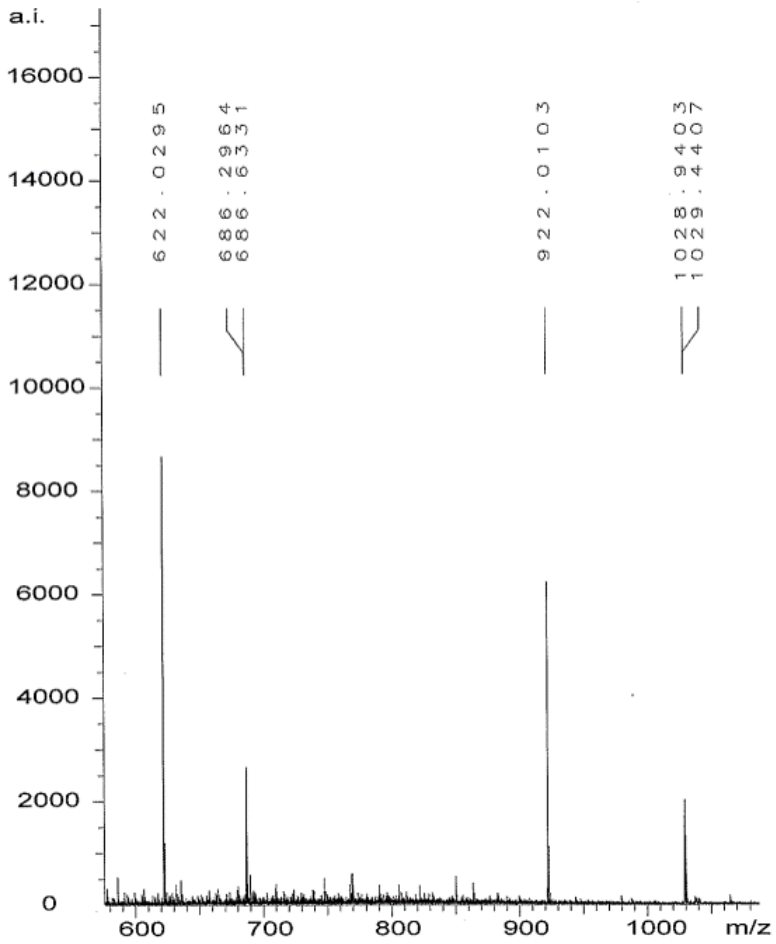


Figure S3. ESI-TOFMS spectrum of cyclic PI polyamide **4**. *m*/*z* calcd for $C_{92}H_{116}N_{37}O_{18}S^{3+}$ [*M* + 3H]$^{3+}$ 686.3001, found 686.2964. $C_{92}H_{115}N_{37}O_{18}S^{2+}$ [*M* + 2H]$^{2+}$ 1028.9465, found 1028.9403. The two peaks, 622.0295 and 922.0103, are the calibration peaks.

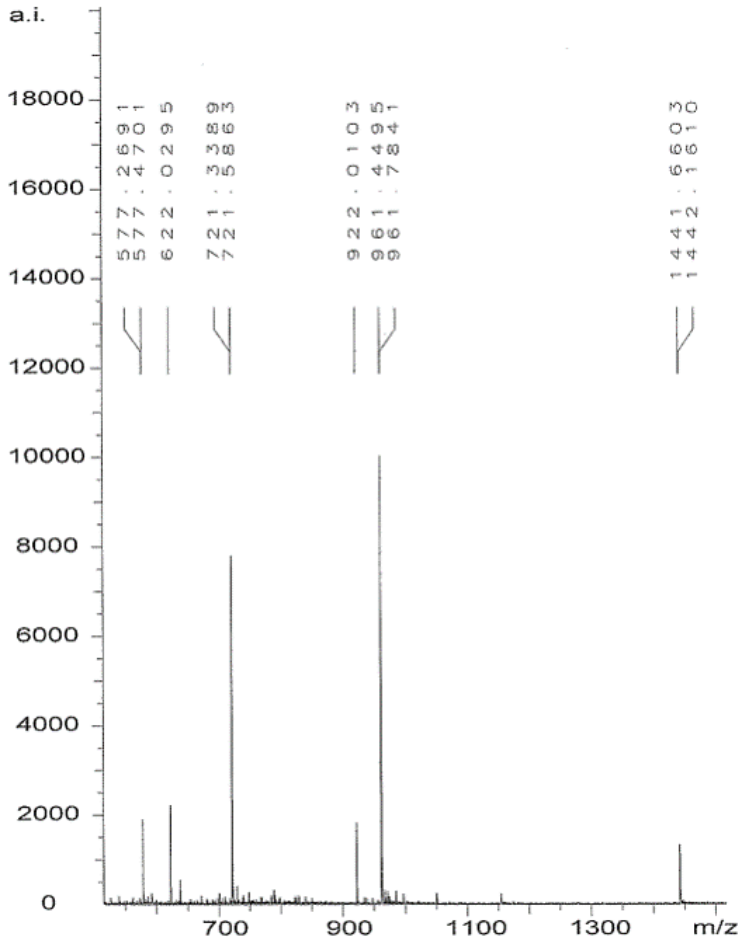


Figure S4. ESI-TOFMS spectrum of cyclic PI polyamide **2**. *m*/*z* calcd for $C_{129}H_{185}N_{40}O_{33}S_2^{4+}$ $[M + 5H]^{5+}$ 577.2688, found 577.2691. *m*/*z* calcd for $C_{129}H_{184}N_{40}O_{33}S_2^{4+}$ $[M + 4H]^{4+}$ 721.3342, found 721.3301. *m*/*z* calcd for $C_{129}H_{183}N_{40}O_{33}S_2^{3+}$ $[M + 3H]^{3+}$ 961.4432, found 961.4358. *m*/*z* calcd for $C_{129}H_{182}N_{40}O_{33}S_2^{2+}$ $[M+ 2H]^{2+}$ 1441.6612, found 1441.6603. The two peaks, 622.0295 and 922.0103, are the calibration peaks.

2. Bind-n-Seq

rank	kmer_selector output	enrichment	rank	kmer_selector output	enrichment
1	AGGCAGACA	71	29	CTGGCTAAC	50
2	AGACTGACA	69	30	AGTCAGGCA	50
3	GACAGTCAC	69	31	TGGCAGTAA	50
4	TGGCTAACA	68	32	CTGTCTTCC	50
5	AGACAGTCA	68	33	CGGCTGACA	50
6	CAGACAGTC	64	34	ATAGTGAGC	50
7	GGCTTACTA	63	35	AGTTGCTAT	49
8	AGGCTTACA	61	36	AGTCAGCAA	49
9	GGAAGACAA	60	37	CAGTCTACC	49
10	ATTGTCTGG	58	38	ATAGCAGAC	49
11	GTGTCAGCA	58	39	CGACAGTCA	49
12	ATTGGCAGA	58	40	CTGTCAGTC	49
13	ATTGTCAGC	57	41	ATATCAGCC	49
14	CAGTTAGCC	56	42	AGTGTCAGC	49
15	ATGTCTGGC	55	43	GATGGCTGA	49
16	TGGTAGACA	54	44	ATGTCAGTC	49
17	GCCAGACTA	54	45	GAGTGTGCC	48
18	GCAGACAGC	54	46	CAGTCAGCC	48
19	GCAGTCAAC	54	47	TAGACTGCA	48
20	ATGCAGACA	53	48	AGTCTGCAT	48
21	ATGTCTGCA	53	49	ACTGACAGC	47
22	GCAGACTTA	53	50	AGCCTGTCA	47
23	ATGCAGTCA	51	51	TATCTGCCA	47
24	AGGCTGATA	51	52	GCAGTCTTA	47
25	ATGGAAGAC	51	53	TGTCTGCAA	46
26	GGCTGTTAA	51	54	GCGGCAGAC	46
27	CGGCTAACA	51	55	AGCCAGGCT	46
28	AGGCTGACA	51			

Table S1. Bind-n-Seq results of hairpin **1**

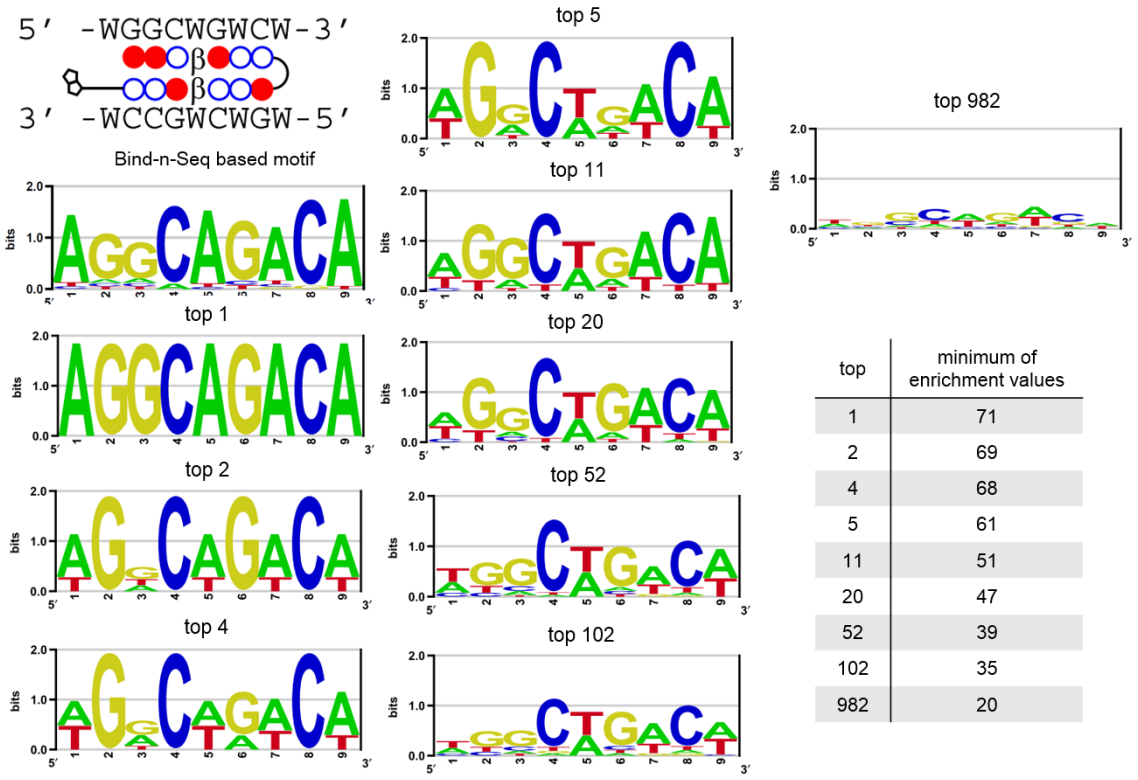
rank	kmer_selector output	enrichment	rank	kmer_selector output	enrichment
1	AGGCAGACA	138	23	ATTACTGCC	87
2	CGGCAGACA	123	24	CGGCTGATA	87
3	CGGCTGACA	113	25	GGCTGACAC	87
4	AGTCAGCCG	106	26	AGGCAGACT	85
5	AGACAGTCA	101	27	CCGGCAGAC	84
6	ATGTCTGCC	100	28	ATTGGCTGA	84
7	ACGGCTGAC	97	29	GGCTTACTA	84
8	AGGCTTACA	96	30	GGCAGACAA	84
9	ATGACAGCC	95	31	ATGTCAGTC	84
10	AGTCTGCCG	94	32	CTGACAGTC	83
11	AGTCTGTCA	93	33	TGGCTAACA	83
12	AGGCTGACA	92	34	CGGCAGTTA	82
13	AGGCTGACT	91	35	AGTTAGCCG	82
14	AAGACAGCC	90	36	GACAGTCAC	82
15	CGGCTAACA	89	37	CAGACAGCC	81
16	CAGACAGTC	89	38	GGCAGTAAC	80
17	ATGACTGTC	89	39	TGGTAGACA	80
18	AGCCAGTCA	88	40	CCGGCTTAC	79
19	AATACAGCC	88	41	CAGTCTGCC	79
20	AGTCTACCG	88	42	ATGTCAGCG	78
21	GGCTGACTA	88	43	GCGGCAGAC	78
22	ATGTCTGGC	87	44	AGGCTGATA	78

Table S2. Bind-n-Seq results of cyclic **2**

sequence	hairpin **1**	cyclic **2**
AGGCAGACA	1	1
AGGCAGACT	195	26
AGGCAGTCA	9255	3997
AGGCAGTCT	14030	1450
AGGCTGACA	28	12
AGGCTGACT	75	13
AGGCTGTCA	13847	926
AGGCTGTCT	–	1772
TGGCAGACA	4575	928
TGGCAGACT	3667	650
TGGCAGTCA	14318	1451
TGGCAGTCT	6317	2011
TGGCTGACA	3229	608
TGGCTGACT	9301	3439
TGGCTGTCA	4066	439
TGGCTGTCT	1421	133

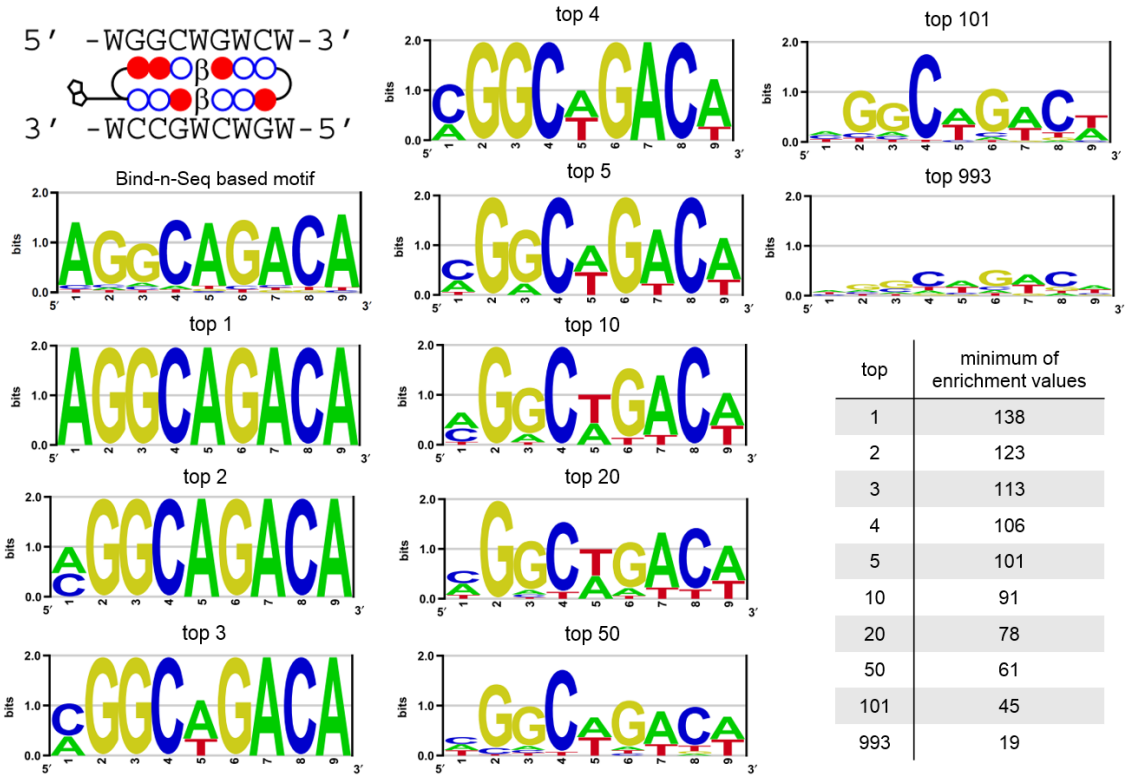
Table S3. All the predetermined sequences and its ranks in order of enrichment values for hairpin **1** and cyclic **2**.

3. Motif Analyses of NGS



top	minimum of enrichment values
1	71
2	69
4	68
5	61
11	51
20	47
52	39
102	35
982	20

Figure S5. NGS results of the hairpin **1**. Motif analyses by the Bind-n-Seq-based mermade and by Bind-n-Seq-MR with 5'-WGGCWGWCW-3' as the reference sequence. The table shows the minimum of enrichment values of sequences in the



top	minimum of enrichment values
1	138
2	123
3	113
4	106
5	101
10	91
20	78
50	61
101	45
993	19

Figure S6. NGS results of the cyclic **2**. Motif analyses by the Bind-n-Seq-based mermade and by Bind-n-Seq-MR with 5'-WGGCWGWCW-3' as the reference sequence. The table shows the minimum of enrichment values of sequences in the

4. SPR Fitting Curves

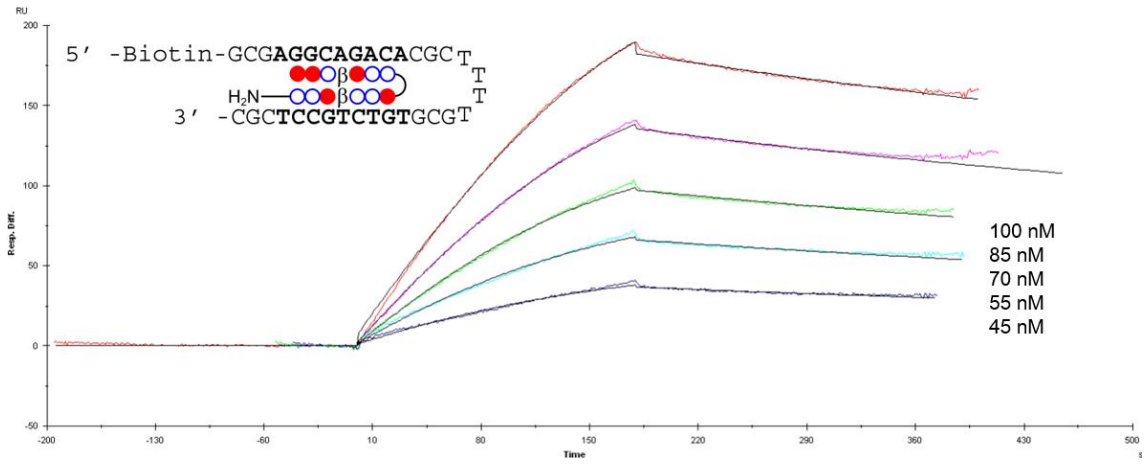


Figure S7. Fitting curves of SPR-binding assay for the combination of ODN1 and hairpin **4** in a fitting mode of 1:1 binding with mass transfer.

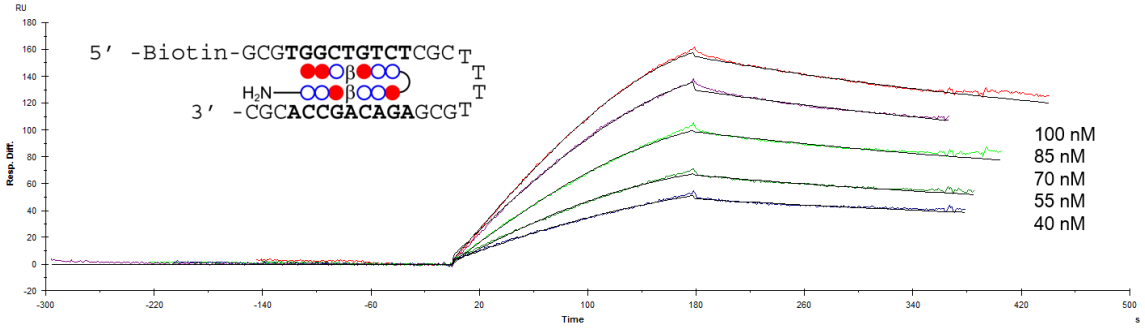


Figure S8. Fitting curves of SPR-binding assay for the combination of ODN2 and hairpin **4** in a fitting mode of 1:1 binding with mass transfer.

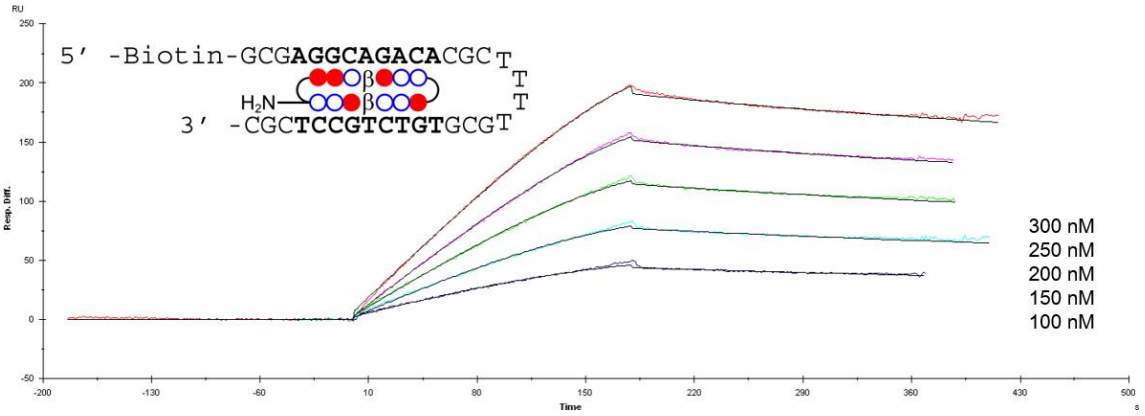


Figure S9. Fitting curves of SPR-binding assay for the combination of ODN1 and cyclic **5** in a fitting mode of 1:1 binding with mass transfer.

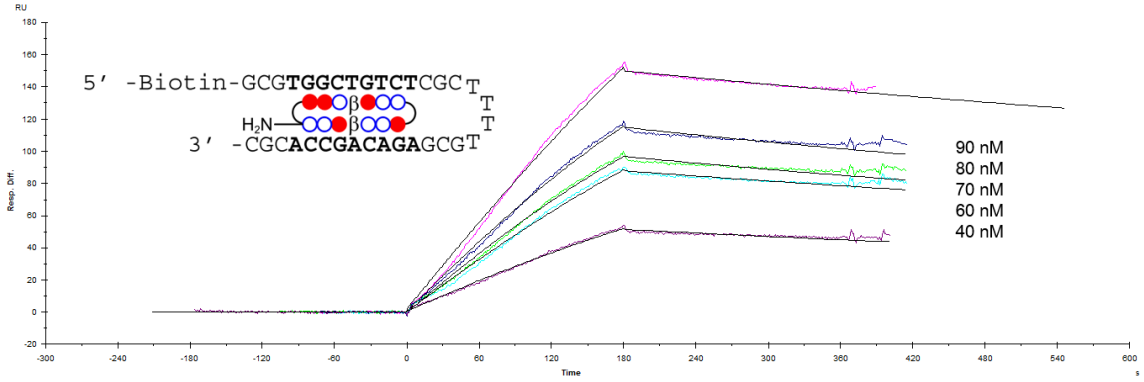


Figure S10. Fitting curves of SPR-binding assay for the combination of ODN2 and cyclic **5** in a fitting mode of 1:1 binding with mass transfer.