

# Strong and Specific Recognition of CAG/CTG Repeat DNA (5'-dWGCWGCW-3') by a Cyclic Pyrrole-Imidazole Polyamide

Yuki Hirose,<sup>[a]</sup> Tomo Ohno,<sup>[a]</sup> Sefan Asamitsu,<sup>[a, b]</sup> Kaori Hashiya,<sup>[a]</sup> Toshikazu Bando,<sup>\*[a]</sup> and Hiroshi Sugiyama<sup>\*[a, c]</sup>

Abnormally expanded CAG/CTG repeat DNA sequences lead to a variety of neurological diseases, such as Huntington's disease. Here, we synthesized a cyclic pyrrole-imidazole polyamide (cPIP), which can bind to the minor groove of the CAG/CTG DNA sequence. The double-stranded DNA melting temperature

( $T_m$ ) and surface plasmon resonance assays revealed the high binding affinity of the cPIP. In addition, next-generation sequencing showed that the cPIP had high specificity for its target DNA sequence.

Although DNA repeat sequences are normally present in the human genome, abnormally elongated repeats can lead to a variety of diseases.<sup>[1,2]</sup> The abnormal elongation of CAG/CTG repeat sequences causes Huntington's disease, spinocerebellar ataxia, and myotonic dystrophy. Compounds that bind to the CAG/CTG repeat sequences have been studied to develop therapeutic agents for these neurological disorders.<sup>[3-5]</sup> Notably, Pearson et al. recently achieved a reduction in the number of repeats *in vivo* using a compound that binds to the hairpin structure formed by the CAG repeats.<sup>[6]</sup> These studies targeted the r(CUG) repeats<sup>[3,4]</sup> or d(CAG/CAG) hairpin structures in the CAG repeat regions,<sup>[5,6]</sup> whereas our group previously developed pyrrole-imidazole polyamides (PIPs) that bind to the d(CAG/CTG) sequences in a sequence-specific manner.<sup>[7,8]</sup>

PIPs are one of the well-studied DNA-binding compounds, which were developed by Dervan et al.<sup>[9,10]</sup> They bind to the minor groove of B-DNA, recognizing Watson-Crick base pairs by antiparallel pairings of their *N*-methylpyrrole (Py) and *N*-meth-

ylimidazole (Im) moieties: a Py/Im pair recognizes a C/G pair and a Py/Py pair recognizes an A/T or T/A pair. A  $\gamma$ -aminobutyric acid turn ( $\gamma$ -turn) connects two arrangements of Py and Im to create hairpin PIPs (hPIPs), which have been frequently used in many studies. For effective binding of long (more than 10 heteroaromatic rings) hPIPs, a  $\beta$ -alanine moiety is used as a substitute for a Py to ease the curvature of the hPIPs: a Py/ $\beta$  and  $\beta$ / $\beta$  pair recognizes an A/T or T/A pair.<sup>[11]</sup> Addition of a second  $\gamma$ -turn to close the hairpin structure results in cyclic PIPs (cPIPs). In previous studies, cPIPs with eight rings were reported to have higher DNA-binding affinity than the corresponding hPIPs.<sup>[12,13]</sup> In terms of the specificity of base recognition, an 8-ring cPIP distinguished a single base mismatch better than an hPIP.<sup>[12]</sup>

To create a ligand that binds to the CAG/CTG repeat with high affinity and specificity, we designed and synthesized cPIP 1, which targets the sequence 5'-WGCWGCW-3'. We also synthesized hPIP 2 with the same recognition sequence for comparison (Figure 1). The (*R*)- $\alpha$ -substituted  $\gamma$ -turn was used to enhance the binding affinity and orientation specificity of each PIP.<sup>[14,15]</sup> The synthetic routes of PIPs 1 and 2 are shown in the Experimental Section.

First, we evaluated the binding affinity of PIPs 1 and 2 for their target sequence using the double-stranded DNA melting temperature ( $T_m$ ) assay with a sequence 5'-CGAGCAGCAG-3'/3'-GCTCGTCTGTC-5' (bases in bold font are recognized by the PIPs). The measured  $T_m$  and  $\Delta T_m$  ( $\Delta T_m = T_m[\text{DNA} + \text{PIP}] - T_m[\text{DNA}]$ ) values are presented in Table 1 and the representative melting curves are in Figure S5. As expected, cPIP 1 showed a higher binding affinity than hPIP 2 ( $\Delta \Delta T_m = 12.4^\circ\text{C}$ ). To further investigate the binding property of these PIPs, we performed  $T_m$  assays using 5'-(CAG)<sub>10</sub>-3' (5'-CAGCAGCAGCAGCAGCAGCAGCAGCAGCAG-3') and 5'-(CTG)<sub>10</sub>-3' (5'-CTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG-3') sequences and obtained values that are shown in Table 1 and Figure S6. Each of these sequences forms a self-complementary hairpin structure, which contains A/A or T/T mismatch pairs. Expanded CAG/CTG repeat sequences are known to partially form such structures.<sup>[16]</sup> Interestingly,

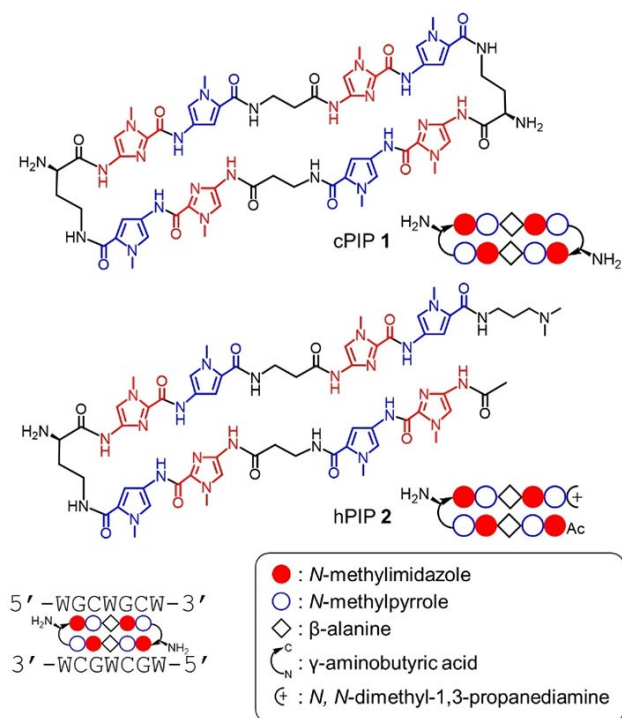
[a] Y. Hirose, T. Ohno, Dr. S. Asamitsu, K. Hashiya, Dr. T. Bando, Prof. Dr. H. Sugiyama  
Department of Chemistry,  
Graduate School of Science,  
Kyoto University  
Kitashirakawa-oiwakecho  
Sakyo-ku, Kyoto, 606-8502 (Japan)  
E-mail: bando@kuchem.kyoto-u.ac.jp  
hs@kuchem.kyoto-u.ac.jp

[b] Dr. S. Asamitsu  
Present address: Department of Genomic Neurology,  
Institute of Molecular Embryology and Genetics,  
Kumamoto University, Honjo  
Chuo-ku, Kumamoto, 860-0811 (Japan)

[c] Prof. Dr. H. Sugiyama  
Institute for Integrated Cell-Material Science (iCeMS),  
Kyoto University, Yoshida-ushinomiya-cho  
Sakyo-ku, countryPart/> Kyoto, 606-8501 (Japan)

Supporting information for this article is available on the WWW under <https://doi.org/10.1002/cbic.202100533>

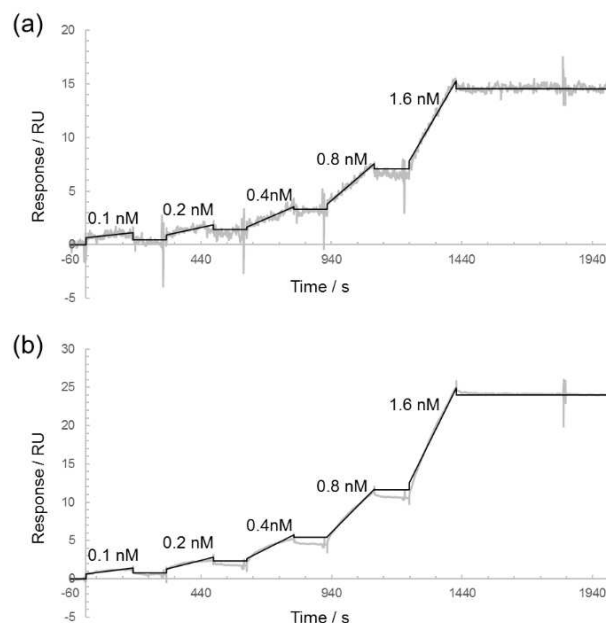
© 2021 The Authors. ChemBioChem published by Wiley-VCH GmbH. This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.



**Figure 1.** Chemical structures and ball-and-stick notation of cPIP 1 and hPIP 2.

although there are three A/A or T/T mismatches per binding site, both of these PIPs bound to 5'-(CAG)<sub>10</sub>-3' and 5'-(CTG)<sub>10</sub>-3' sequences strongly. As with the CAG/CTG sequence, cPIP 1 also showed a higher binding affinity than hPIP 2 in this case ( $\Delta T_m = 9.4^\circ\text{C}$  and  $6.9^\circ\text{C}$ , respectively). These findings indicated that cPIP 1 could bind to CAG/CTG repeat sequences even when the sequences formed unnatural structures. Moreover, findings of the  $T_m$  assays using 5'-r(CAG)<sub>10</sub>-3' and 5'-r(CUG)<sub>10</sub>-3' clearly indicated that these PIPs bound with high specificity to DNA rather than RNA (Figure S7).

In addition, we performed the surface plasmon resonance (SPR) assay to obtain the kinetic constants for the interaction between these PIPs and their target DNA. A 5'-biotin-labeled hairpin DNA (5'-biotin-CGCGAGCAG-CACGCGTTTTTCGCTGCTGCTCGCG-3') was immobilized on a streptavidin-coated sensor chip, and various concentrations of the polyamide solution were added (Figure S8). The sensorgrams shown in Figure 2 were obtained using the single-cycle program on the Biacore T200 system. The rates of association



**Figure 2.** SPR sensorgrams using (a) cPIP 1 and (b) hPIP 2. Gray and black curves represent the experimental data and fitting curves, respectively.

**Table 2.** Binding affinities of compounds 1 and 2.

PIP	5'-biotin-CGCGAGCAGCA CGCG T <sub>T</sub> 3'-GCGCTCGTCTGTCGTCGCG T <sub>T</sub>		
	$k_a/M^{-1}s^{-1}$	$k_d/s^{-1}$	$K_D/M$
cPIP 1	$4.1 \times 10^4$	$5.2 \times 10^{-7}$	$1.3 \times 10^{-11}$
hPIP 2	$2.2 \times 10^6$	$2.7 \times 10^{-6}$	$1.2 \times 10^{-12}$

All values are determined by fitting with a 1:1 binding model.

( $k_a$ ) and dissociation ( $k_d$ ) and the dissociation constant ( $K_D$ ) are listed in Table 2 and other detailed values in Figure S9. Notably, the U-values, which indicate the reliability of the data, were large in each case. This might be due to the linearity of the association and dissociation regions, the large  $R_{max}$  and the  $k_d$  values below the specification range of the system (Figure S9). Surprisingly, contrary to the findings of our  $T_m$  assay, cPIP 1 showed a larger  $K_D$  value than hPIP 2. We do not have a certain explanation for this inconsistency, but it might be due to differences in the measurement methods and PIP concentrations. Interestingly, these two PIPs showed different kinetic properties: cPIP 1 exhibited lower  $k_a$  and  $k_d$  values than hPIP 2.

**Table 1.**  $T_m$  and  $\Delta T_m$  values of compounds 1 and 2.

PIP	Match		5'-(CAG) <sub>10</sub> -3'		5'-(CTG) <sub>10</sub> -3'	
	$T_m/^\circ\text{C}$	$\Delta T_m/^\circ\text{C}$	$T_m/^\circ\text{C}$	$\Delta T_m/^\circ\text{C}$	$T_m/^\circ\text{C}$	$\Delta T_m/^\circ\text{C}$
-	41.9 ( $\pm 0.5$ )	-	49.0 ( $\pm 0.7$ )	-	50.6 ( $\pm 0.3$ )	-
cPIP 1	93.2 ( $\pm 0.2$ )	51.2 ( $\pm 0.6$ )	92.7 ( $\pm 0.5$ )	43.7 ( $\pm 0.8$ )	92.0 ( $\pm 0.6$ )	41.4 ( $\pm 0.7$ )
hPIP 2	80.7 ( $\pm 0.4$ )	38.8 ( $\pm 0.7$ )	83.2 ( $\pm 0.6$ )	34.3 ( $\pm 0.9$ )	85.1 ( $\pm 0.4$ )	34.5 ( $\pm 0.5$ )

Therefore, although hPIP 2 binds to the target DNA faster than cPIP 1, it is more difficult to dissociate the latter than the former after binding. We expect that such differences in the kinetic properties will affect the function of these PIPs *in cellulo* and *in vivo*. Moreover, experiments using disease models are currently underway.

The findings of the  $T_m$  and SPR assays showed high binding affinity of cPIP 1 to its target DNA sequence. However, we had not considered its sequence specificity. To investigate the binding specificity of PIPs for a broad range of DNA sequences, the Bind-n-Seq method, which is a powerful tool that utilizes large-scale parallel sequencing, can be used.<sup>[17–20]</sup> We prepared a biotinylated cPIP (cPIP 3 in Figure 3a) and performed the Bind-n-Seq method using random DNA sequences as described previously.<sup>[21]</sup>

The enriched sequences are shown in Figure 3b in the order of enrichment values, and the binding motif result is presented in Figure 3c. The binding motif obtained using the Bind-n-Seq method matched almost perfectly with the target sequence of cPIP 3 (5'-WGCWGCW-3'). This finding suggests that the original PIP 1 recognizes its target sequence with high specificity.

In summary, we synthesized cPIP 1 and hPIP 2, which target the CAG/CTG repeat sequences, and examined their DNA-binding properties *in vitro*. The findings of the  $T_m$  assay indicated that cPIP 1 has a higher binding affinity than hPIP 2. The SPR assays revealed that these PIPs possess different kinetic properties on binding to their target DNA sequence. In addition, the findings of the Bind-n-Seq method indicated that PIP 1 had high sequence specificity to the target DNA sequence. Taken together, these findings suggest that cPIP 1 can be used as a ligand to distinguish the CAG/CTG repeat sequences. Further experiments using cellular and animal disease models should

be performed to develop the use of cPIP 1 as a therapeutic drug and molecular probe.

## Experimental Section

**General:** Reagents and solvents were purchased from standard suppliers and used without further purification. HPLC analysis was performed on a Jasco Engineering PU-2089 plus series system using a COSMOSIL 150×4.6 mm 5 C<sub>18</sub>-MS-II Packed Column (Nacalai Tesque, Inc.) in 0.1% TFA in water with acetonitrile as the eluent at a flow rate of 1.0 mL/min and a linear gradient elution of 0–100% acetonitrile in 40 min with detection at 254 nm. Collected fractions were analyzed by MALDI-TOF-MS microflex-KS II (Bruker). HPLC purification was carried out by Jasco engineering PU-2080 or PU-2089 plus series using a COSMOSIL 150×10 mm 5 C<sub>18</sub>-MS-II Packed Column (Nacalai Tesque, Inc.) in 0.1% TFA in water with acetonitrile as the eluent at a flow rate of 3.8 mL/min and a linear gradient elution of +1%/min acetonitrile for 20 min with detection at 254 nm. Master DMSO solutions of PIP were prepared based on the following formula.

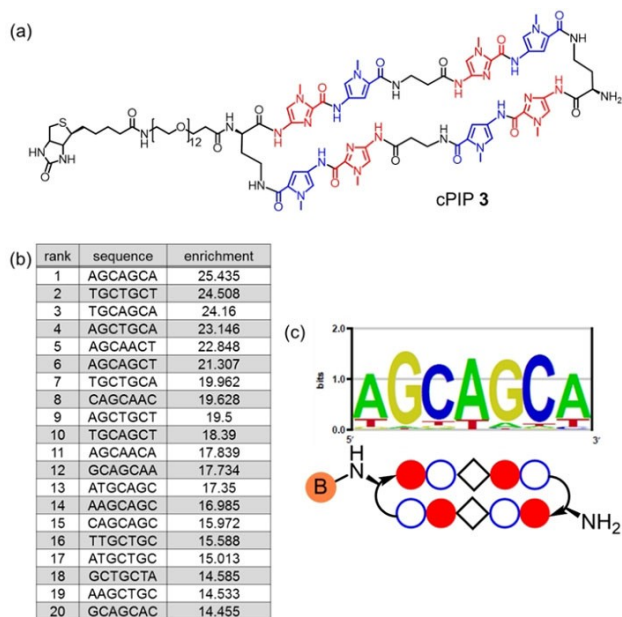
$$\varepsilon = 9900 \times (\text{sum number of Py and Im})$$

$$\text{Abs} = \varepsilon c l$$

$\varepsilon$ , Abs,  $c$ , and  $l$  are molar extinction coefficients of PIPs in DMSO solution at around 302 nm, absorbance at 302 nm measured by Nanodrop 1000 spectrophotometer (Thermo Fisher Scientific Inc.), molar concentration, and the path length, respectively. <sup>1</sup>H-NMR spectra were measured on JEOL JNM ECA-600 spectrometer (600 MHz for <sup>1</sup>H), with chemical shifts reported in parts per million relative to residual solvent and coupling constants in hertz. The following abbreviations were applied to spin multiplicity: s (singlet), d (doublet), t (triplet), and m (multiplet).

**Fmoc solid-phase synthesis of PIPs:** The solid-phase synthesis of each PIP was performed on a PSSM-8 system (Shimadzu) as described previously.<sup>[8]</sup> The building blocks used in this study are FmocHN-Py-CO<sub>2</sub>H, FmocHN-Im-CO<sub>2</sub>H, FmocHN-Py- $\beta$ -CO<sub>2</sub>H, FmocHN-Boc-(*R*)- $\alpha$ -aminobutyric acid. Each of them was introduced sequentially to FmocHN-Py-trityl resin (for 1) or FmocHN-Py-oxime resin (for 2). The N-terminus of hairpin PIP (2) was capped by acetyl group using 20% Ac<sub>2</sub>O in DMF.

**cyclo-(ImPy $\beta$ ImPy-(*R*)<sup>( $\alpha$ -NH<sub>2</sub>, $\gamma$ -ImPy $\beta$ ImPy-(*R*)<sup>( $\alpha$ -NH<sub>2</sub>, $\gamma$ -)</sup> (1):</sup>** Using 86 mg (49 mg + 37 mg) FmocHN-Py-trityl resin (0.335 mmol/g + 0.137 mmol/g) and proper building blocks, H<sub>2</sub>N-(*R*)<sup>( $\alpha$ -NH<sub>2</sub>, $\gamma$ -ImPy $\beta$ ImPy-(*R*)<sup>( $\alpha$ -NH<sub>2</sub>, $\gamma$ -)</sup> and proper building blocks, H<sub>2</sub>N-(*R*)<sup>( $\alpha$ -NH<sub>2</sub>, $\gamma$ -ImPy $\beta$ ImPy-(*R*)<sup>( $\alpha$ -NH<sub>2</sub>, $\gamma$ -)</sup> was synthesized by the Fmoc solid-phase synthesis. After H<sub>2</sub>N-(*R*)<sup>( $\alpha$ -NH<sub>2</sub>, $\gamma$ -ImPy $\beta$ ImPy-(*R*)<sup>( $\alpha$ -NH<sub>2</sub>, $\gamma$ -)</sup> resin was cleaved with 30% hexafluoroisopropanol (HFIP) in dichloromethane (DCM) for 3 hours at room temperature, resin was removed by filtration and the filtrate was dropped into Et<sub>2</sub>O to obtain 26.1 mg brown powder of H<sub>2</sub>N-(*R*)<sup>( $\alpha$ -NH<sub>2</sub>, $\gamma$ -ImPy $\beta$ ImPy-(*R*)<sup>( $\alpha$ -NH<sub>2</sub>, $\gamma$ -)</sup> (Analytical HPLC:  $t_R$  = 18.3 min. MALDI-TOF MS:  $m/z$  calcd for C<sub>68</sub>H<sub>89</sub>N<sub>26</sub>O<sub>17</sub><sup>+</sup> [M+H]<sup>+</sup> 1541.68, found; 1541.75). This crude sample was dissolved in DMF 16.9 mL (1 mM) for the next intramolecular cyclization step. After the addition of pentafluorophenyl diphenylphosphinate (FDPP, 3 equiv.) and *N,N*-diisopropylethylamine (DIEA, 6 equiv.), the mixture was stirred for 24 hours at room temperature. Then the solvent was evaporated and dried *in vacuo*. The vacuum-dried brown oil was dissolved in the minimum volume of MeOH/DCM 1:1 mixture dropped into Et<sub>2</sub>O. Then Et<sub>2</sub>O was removed *in vacuo* to obtain 23.9 mg brown powder of cyclo-(ImPy $\beta$ ImPy-(*R*)<sup>( $\alpha$ -NH<sub>2</sub>, $\gamma$ -ImPy $\beta$ ImPy-(*R*)<sup>( $\alpha$ -NH<sub>2</sub>, $\gamma$ -)</sup>. This crude sample was dissolved in trifluoroacetic acid (TFA)/DCM</sup></sup></sup></sup></sup>



**Figure 3.** (a) The chemical structure of compound 3. (b) Seven-bp sequences in the order of enrichment values. (c) Motif analysis.

2:5 mixture 700  $\mu\text{L}$  and stirred for 30 minutes at room temperature to deprotect two Boc groups. After the reaction, the reaction mixture was dropped into  $\text{Et}_2\text{O}$  and powdered. The powder was dried *in vacuo* and 23.5 mg brown powder of cyclo-(ImPy $\beta$ ImPy-( $R$ ) $^{\alpha-\text{NH}_2}$ - $\gamma$ -ImPy $\beta$ ImPy-( $R$ ) $^{\alpha-\text{NH}_2}$ - $\gamma$ ) (1) was obtained. A part of this crude sample (7.9 mg) was dissolved in DMF and purified by HPLC. 1.2 mg of purified sample was obtained as off-white powder (0.91  $\mu\text{mol}$ , 13% yield for 13 steps). Analytical HPLC:  $t_{\text{R}} = 13.5$  min. MALDI-TOF MS:  $m/z$  calcd for  $\text{C}_{58}\text{H}_{71}\text{N}_{26}\text{O}_{12}^+ [\text{M} + \text{H}]^+$  1323.57, found; 1323.69.  $^1\text{H-NMR}$  (600 MHz,  $\text{DMSO-d}_6$ ):  $\delta$  11.02 (s, 2H), 10.25 (s, 2H), 9.98 (s, 2H), 9.97 (s, 2H), 8.31 (d,  $J = 4.2$  Hz, 6H), 8.20 (t,  $J = 5.4$  Hz, 2H), 8.04 (t,  $J = 6.0$  Hz, 2H), 7.50 (s, 2H), 7.46 (s, 2H), 7.30 (d,  $J = 1.8$  Hz, 2H), 7.24 (d,  $J = 1.8$  Hz, 2H), 7.00 (d,  $J = 1.8$  Hz, 2H), 6.82 (d,  $J = 1.8$  Hz, 2H), 4.05–4.00 (m, 2H), 3.95 (s, 6H), 3.94 (s, 6H), 3.81 (s, 12H), 3.28–3.23 (m; partially overlapped with  $\text{H}_2\text{O}$ , 8H), 2.60–2.57 (m; partially overlapped with DMSO, 4H), 2.00–1.96 (m, 4H).

**AcImPy $\beta$ ImPy-( $R$ ) $^{\alpha-\text{NH}_2}$ - $\gamma$ -ImPy $\beta$ ImPyDp (2):** Using 85 mg FmocHN-Py-oxime resin (0.328 mmol/g) and proper building blocks, AcImPy $\beta$ ImPy-( $R$ ) $^{\alpha-\text{NH}_2}$ - $\gamma$ -ImPy $\beta$ ImPy-oxime resin was synthesized by the Fmoc solid-phase synthesis. After AcImPy $\beta$ ImPy-( $R$ ) $^{\alpha-\text{NH}_2}$ - $\gamma$ -ImPy $\beta$ ImPy-oxime resin was cleaved with *N,N*-dimethyl-1,3-propanediamine (Dp) for 3 hours at 55  $^\circ\text{C}$ , resin was removed by filtration and the filtrate was dropped into  $\text{Et}_2\text{O}$  to obtain 37.0 mg brown powder of AcImPy $\beta$ ImPy-( $R$ ) $^{\alpha-\text{NH}_2}$ - $\gamma$ -ImPy $\beta$ ImPyDp (Analytical HPLC:  $t_{\text{R}} = 18.2$  min. MALDI-TOF MS:  $m/z$  calcd for  $\text{C}_{66}\text{H}_{87}\text{N}_{26}\text{O}_{14}^+ [\text{M} + \text{H}]^+$  1467.68, found; 1467.66). This crude sample was dissolved in TFA/DCM 2:5 mixture 700  $\mu\text{L}$  and stirred for 30 minutes at room temperature to deprotect a Boc group. After the reaction, the reaction mixture was dropped into  $\text{Et}_2\text{O}$  and powdered. The powder was dried *in vacuo* and 46.4 mg brown powder of AcImPy $\beta$ ImPy-( $R$ ) $^{\alpha-\text{NH}_2}$ - $\gamma$ -ImPy $\beta$ ImPyDp (2) was obtained. A part of this crude sample (10.7 mg) was dissolved in DMF and purified by HPLC. 4.4 mg of purified sample was obtained as off-white powder (3.2  $\mu\text{mol}$ , 50% yield for 12 steps). Analytical HPLC:  $t_{\text{R}} = 14.4$  min. MALDI-TOF MS:  $m/z$  calcd for  $\text{C}_{61}\text{H}_{87}\text{N}_{26}\text{O}_{12}^+ [\text{M} + \text{H}]^+$  1367.63, found; 1367.78.

**Cyclo-(ImPy $\beta$ ImPy-( $R$ ) $^{\alpha-\text{NH}}-\text{PEG}_{12}$ -Biotin- $\gamma$ -ImPy $\beta$ ImPy-( $R$ ) $^{\alpha-\text{NH}_2}$ - $\gamma$ ) (3):** The crude powder of compound (4.0 mg) 1 was mixed with DIEA (6 equiv.) and EZ-Link<sup>TM</sup> NHS-PEG<sub>12</sub>-Biotin (Thermo Fisher Scientific, 0.7 equiv.) in 50  $\mu\text{L}$  of DMF. The mixture was stirred for 25 min at r.t. Evaporation of the solvent yielded a yellow oil, which was purified by HPLC to afford compound 3 as a light yellow powder (0.4 mg, 0.30  $\mu\text{mol}$ , 5% yield for 14 steps). HPLC and MALDI-TOF mass spectroscopy were carried out to identify the synthetic compound. Analytical HPLC:  $t_{\text{R}} = 16.0$  min. MALDI-TOF MS:  $m/z$  calcd for  $\text{C}_{95}\text{H}_{138}\text{N}_{29}\text{O}_{27}\text{S}^+ [\text{M} + \text{H}]^+$  2149.0, found; 2149.13.

**$T_{\text{m}}$  analysis:** Four DNA oligomers (5'-CGAGCAGCACG-3'/3'-GCTCGTCTGTC-5', 5'-(CAG)<sub>10</sub>-3' and 5'-(CTG)<sub>10</sub>-3') used in this analysis were purchased from Sigma and two RNA oligomers (5'-r(CAG)<sub>10</sub>-3' and 5'-r(CTG)<sub>10</sub>-3') were purchased from Thermo Fisher. The analytical buffer for  $T_{\text{m}}$  analysis was an aqueous solution of 2.5 mM sodium chloride and 10 mM Tris-HCl at pH 7.5 containing 0.375% v/v DMSO. The concentration of dsDNA was 2.5  $\mu\text{M}$ . The concentration of polyamides was 3.75  $\mu\text{M}$  (1.5 equiv.). Before the analysis, the samples were annealed from 95  $^\circ\text{C}$  to 20  $^\circ\text{C}$  at a rate of 1.0  $^\circ\text{C}/\text{min}$ . Absorbance at 260 nm was recorded from 20  $^\circ\text{C}$  to 95  $^\circ\text{C}$  at a rate of 1.0  $^\circ\text{C}/\text{min}$  using a spectrophotometer V-750 (JASCO) with a thermocontrolled PAC-743R cell changer (JASCO) and a thermal circulator CTU-100 (JASCO). The  $T_{\text{m}}$  values shown in Table 1 and Figures S5–S7 are the average of all data ( $n \geq 2$ ). The representative denaturing graphs of each compound are shown in Figures S5–S7.

**SPR assays:** SPR assays were performed on a Biacore T200 instrument (GE Healthcare) following the manufacturer's instruction. A

biotinylated DNA oligomer (5'-biotin-CGCGAGCCAG-CACGCGTTTCGCGTCTGCTCGCG-3') was purchased from Sigma and immobilized to the streptavidin-functionalized SA sensor chip (GE Healthcare) to obtain the desired immobilized level (approximately 630 RU rise). SPR assays were carried out using HBS-EP buffer (GE Healthcare, 10 mM HEPES pH 7.4, 150 mM NaCl, 3 mM EDTA, and 0.005% Surfactant P20) with 0.1% DMSO at 25  $^\circ\text{C}$ . A series of sample solutions with various concentrations were prepared in HBS-EP buffer with 0.1% DMSO. The contact time, dissociation time and flow rate were set for 180 s, 600 s, and 100  $\mu\text{L}/\text{min}$ , respectively. To measure the rates of association ( $k_{\text{a}}$ ) and dissociation ( $k_{\text{d}}$ ) and dissociation constant ( $K_{\text{d}}$ ), data processing was performed by using the Biacore T200 Evaluation Software version 1.0. The sensorgrams were fitted by using a 1:1 binding model. All sensorgrams and all values are shown in Figures 2 and S9 and Table 2.

**Bind-n-Seq analysis:** Bind-n-Seq analysis was conducted with an Ion PI System (Thermo Fisher Scientific). Bind-n-Seq and subsequent analysis to evaluate small-molecule binding affinity towards specific DNA sequences in a broad context sequence pool were modified for the Ion Proton sequencer (ThermoFisher Scientific). The scheme involves three major steps.

- 1) Synthesis of biotinylated PIPs and randomized oligonucleotides with high-throughput sequencing platform-specific adapters (Ion torrent PI; 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). Oligonucleotides (3  $\mu\text{M}$ ) were duplexed by primer extension with adapter-specific primer 1 (9  $\mu\text{M}$ ) in 25 mL reactions containing GoTaq Green (Promega) PCR master mix (2 $\times$ ) with  $\text{Mg}^{2+}$  (2 mM). Reactions were performed at 95  $^\circ\text{C}$  (2 min), 63  $^\circ\text{C}$  (1 min), 72  $^\circ\text{C}$  (4 min), and then 4  $^\circ\text{C}$  in a thermocycler (BioRad). Biotin-conjugated PIPs (100 nM) were allowed to equilibrate with duplex random oligonucleotides for 20 h followed by the addition of Dynabeads M-280 Streptavidin (Thermo Fisher Scientific; beads prepared based on the previous report), then separation of bound and unbound sequences by affinity purification.
- 2) Polyamide-enrichment recovered DNA was diluted (1:10) and amplified with sequencing library adapter-specific primer for 15 cycles to obtain sufficient sequencing template. After purification, enriched libraries were subjected to quality and quantity checking with a DNA High sensitivity BioAnalyzer kit (Agilent Technologies). The qualified libraries were used for template preparation by using an Ion PI Hi-Q OT2 200 kit in an Ion OneTouch 2 system (Thermo Fisher Scientific). The templates were then enriched by using an Ion OneTouch ES enrichment system. The enriched libraries were sequenced by single-read sequencing (Ion Proton sequencer with Ion PI Hi-Q sequencing 200 kit and Ion PI Chip v3; Thermo Fisher Scientific) by following the manufacturer's instructions.
- 3) The sequenced reads (A, C, T, and G) were then processed to obtain a valid constant region and unique random region and retained and split into separate files through a unique 10-nt ion Xpress Barcode. To count the number of PIP enriched unique DNA sequences, a sliding window (length, 7) in MERMADE, and a new pipeline for Bindn-Seq analysis (<http://korflab.ucdavis.edu/Datasets/BindNSeq>) were used. The motifs were processed by enoLOGOS (<http://www.benoslab.pitt.edu/cgi-bin/enologos/enologos.cgi>).

## Acknowledgements

We thank Dr. Kei Okatsu and Dr. Syuya Fukai (Department of Chemistry, Graduate School of Science, Kyoto University) for helping SPR assay. This research was supported by AMED [Grant No. JP21am0101101, Platform Project for Supporting Drug Discovery and Life Science Research (BINDS) to H.S.] and JSPS KAKENHI [Grant No. JP16H06356, No. 21H04705, and 20H05936] and NIH award RO1CA236350 to H.S.; 21J22276, Grant-in-Aid for JSPS Research Fellows to Y.H.].

## Conflict of Interest

The authors declare no conflict of interest.

**Keywords:** CAG/CTG repeats · DNA recognition · pyrrole-imidazole polyamides · triplet repeats

- [1] S. M. Mirkin, *Nature* **2007**, *447*, 932–940.
- [2] K. Usdin, E. Grabczyk, *Cell. Mol. Life Sci.* **2000**, *57*, 914–931.
- [3] M. T. Wheeler, K. Sobczak, J. D. Lueck, R. J. Osborne, X. Lin, R. T. Dirksen, C. A. Thornton, *Science* **2009**, *325*, 336–339.
- [4] L. Guan, M. D. Disney, *Angew. Chem. Int. Ed.* **2013**, *52*, 10010–10013; *Angew. Chem.* **2013**, *125*, 10194–10197.
- [5] K. Nakatani, S. Hagihara, Y. Goto, A. Kobori, M. Hagihara, G. Hayashi, M. Kyo, M. Nomura, M. Mishima, C. Kojima, *Nat. Chem. Biol.* **2005**, *1*, 39–43.
- [6] M. Nakamori, G. B. Panigrahi, S. Lanni, T. Gall-Duncan, H. Hayakawa, H. Tanaka, J. Luo, T. Otabe, J. Li, A. Sakata, M. C. Caron, N. Joshi, T. Prasolava, K. Chiang, J. Y. Masson, M. S. Wold, X. Wang, M. Lee, J. Huddlestone, K. M. Munson, S. Davidson, M. Layeghifard, L. M. Edward, R.

- Gallon, M. Santibanez-Koref, A. Murata, M. P. Takahashi, E. E. Eichler, A. Shlien, K. Nakatani, H. Mochizuki, C. E. Pearson, *Nat. Genet.* **2020**, *52*, 146–159.
- [7] J. Fujimoto, T. Bando, M. Minoshima, S. Uchida, M. Iwasaki, K. Shinohara, H. Sugiyama, *Bioorg. Med. Chem.* **2008**, *16*, 5899–5907.
- [8] S. Asamitsu, Y. Kawamoto, F. Hashiya, K. Hashiya, M. Yamamoto, S. Kizaki, T. Bando, H. Sugiyama, *Bioorg. Med. Chem.* **2014**, *22*, 4646–4657.
- [9] P. B. Dervan, B. S. Edelson, *Curr. Opin. Struct. Biol.* **2003**, *13*, 284–299.
- [10] P. B. Dervan, R. W. Bürli, *Curr. Opin. Chem. Biol.* **1999**, *3*, 688–693.
- [11] J. M. Turner, S. E. Swalley, E. E. Baird, P. B. Dervan, *J. Am. Chem. Soc.* **1998**, *120*, 6219–6226.
- [12] D. M. Herman, J. M. Turner, E. E. Baird, P. B. Dervan, *J. Am. Chem. Soc.* **1999**, *121*, 1121–1129.
- [13] D. M. Chenoweth, D. A. Harki, J. W. Phillips, C. Dose, P. B. Dervan, *J. Am. Chem. Soc.* **2009**, *131*, 7182–7188.
- [14] D. M. Herman, E. E. Baird, P. B. Dervan, *J. Am. Chem. Soc.* **1998**, *120*, 1382–1391.
- [15] Y. Hirose, S. Asamitsu, T. Bando, H. Sugiyama, *J. Am. Chem. Soc.* **2019**, *141*, 13165–13170.
- [16] A. M. Gacy, G. Goellner, N. Juranić, S. Macura, C. T. McMurray, *Cell* **1995**, *81*, 533–540.
- [17] J. L. Meier, A. S. Yu, I. Korf, D. J. Segal, P. B. Dervan, *J. Am. Chem. Soc.* **2012**, *134*, 17814–17822.
- [18] J. S. Kang, J. L. Meier, P. B. Dervan, *J. Am. Chem. Soc.* **2014**, *136*, 3687–3694.
- [19] C. Anandhakumar, S. Kizaki, T. Bando, G. N. Pandian, H. Sugiyama, *ChemBioChem* **2015**, *16*, 20–38.
- [20] G. Kashiwazaki, C. Anandhakumar, S. Asamitsu, T. Kawase, Y. Kawamoto, Y. Sawatani, K. Hashiya, T. Bando, H. Sugiyama, *ChemBioChem* **2016**, *17*, 1752–1758.
- [21] C. Anandhakumar, Y. Li, S. Kizaki, G. N. Pandian, K. Hashiya, T. Bando, H. Sugiyama, *ChemBioChem* **2014**, *15*, 2647–2651.

Manuscript received: October 5, 2021

Revised manuscript received: October 26, 2021

Version of record online: November 18, 2021