Chemical Approaches to the Development of Artificial Transcription Factors Based on Pyrrole - Imidazole Polyamides

AUTHOR(S):

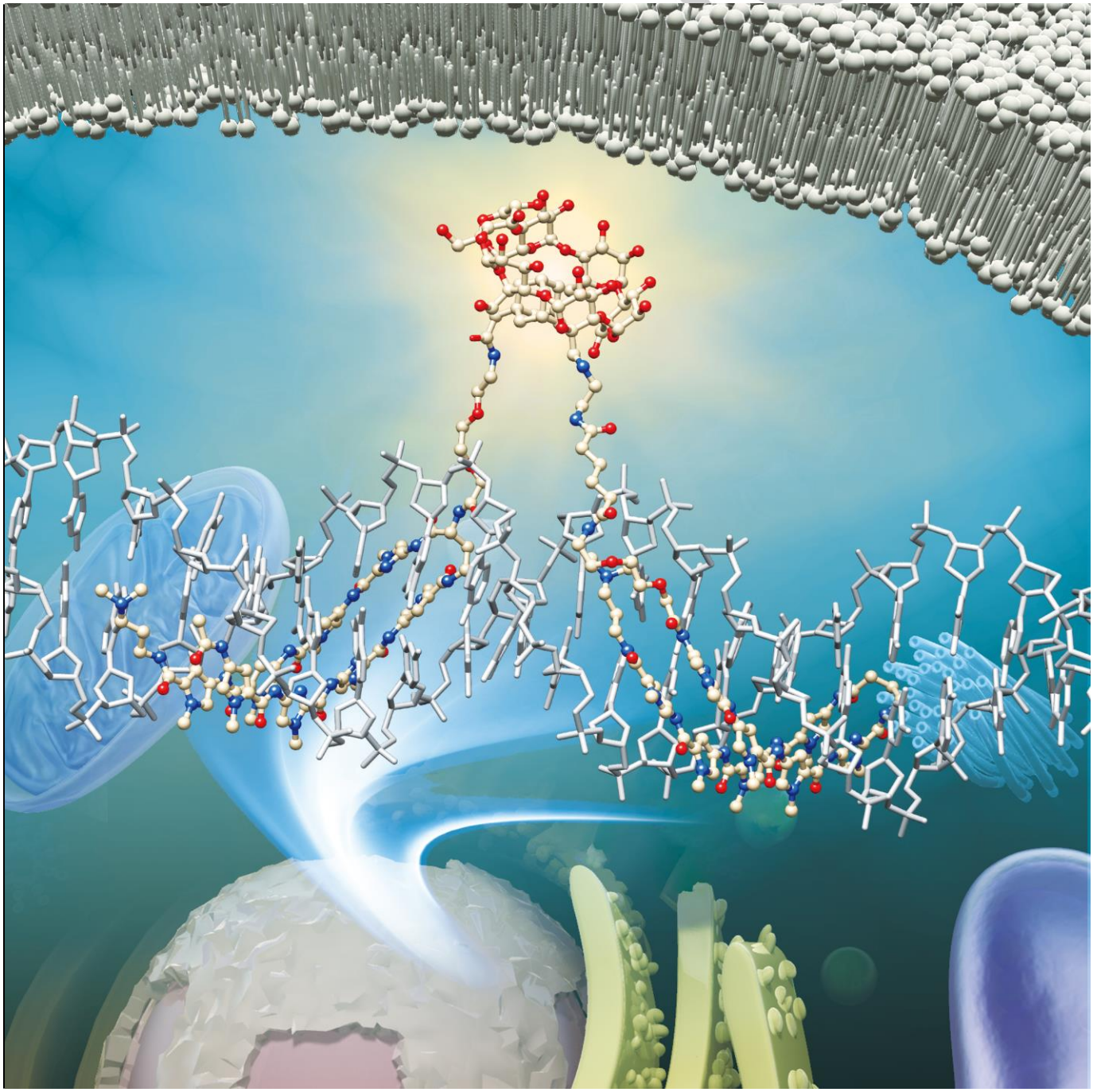
Hidaka, Takuya; Sugiyama, Hiroshi

WILEY-VCH

PERSONAL ACCOUNT

Chemical approaches to the development of artificial transcription factors based on pyrrole–imidazole polyamides

Takuya Hidaka, [a] and Hiroshi Sugiyama*[a, b]



Abstract: To maintain the functions of living organisms, cells have developed complex gene regulatory networks. Transcription factors have a central role in spatiotemporal control of gene expression and this has motivated us to develop artificial transcription factors that mimic their function. We found that three functions could be mimicked by applying our chemical approaches: i) efficient delivery into organelles that contain target DNA, ii) specific DNA binding to the target genomic region, and iii) regulation of gene expression by interaction with other transcription coregulators. We chose pyrrole–imidazole polyamides (PIPs), sequence-selective DNA binding molecules, as DNA binding domains, and have achieved each of the required functions by introducing other functional moieties. The developed artificial transcription factors have potential as chemical tools that can be used to artificially modulate gene expression to enable cell fate control and to correct abnormal gene regulation for therapeutic purposes.

1. Introduction

DNA encodes with four letters—A (adenine), T (thymine), G (guanine), and C (cytosine)—all the information required to build living organisms. The information contained in DNA is copied as mRNA, which is translated into a functional amino acid sequence. In addition to encoding materials, DNA also contains a description of how to build our body and make it functional by providing control layers that regulate gene expression in a spatiotemporal manner. Sequence-specific DNA binding and transcriptional modulation by transcription factors comprise the most canonical gene regulation strategy for cells to achieve cell-specific functions. For example, in pluripotent stem cells, expression of pluripotency-associated genes is induced by reprogramming factors such as OCT4 and SOX2, and SOX9 and MyoD act as master regulators to activate specific gene networks to promote chondrogenic and myogenic differentiation. These transcription factors modulate gene expression by recruiting transcription coregulators, which can also recruit other factors, change chromatin structure around the binding site or induce RNA species with transcriptional control activity such as micro RNA (miRNA) and long noncoding RNA (lncRNA). Precise control of gene networks is indispensable for maintaining cellular function, and abnormal gene regulation is related to many diseases, making it an important target for therapeutic purposes. Recent progress in high-throughput sequencing technology has been revealing the complex gene regulatory network, and increasingly detailed information about the function of DNA sequences is available. This knowledge is useful for artificial gene regulation, which has been partially achieved with molecular–biological approaches such as Cas9 (dCas9) and transcription activator-like effector (TALE), but there is a limited set of chemical tools that can utilize such information. This motivated us to develop compound-based artificial transcription factors using sequence-selective DNA binders: pyrrole–imidazole polyamides (PIPs). In this Personal Account, we would like to explain the functions required for artificial transcription factors and to introduce our chemical strategies to achieve those functions using PIPs.

2. Pyrrole–Imidazole polyamide (PIPs)

We have worked on the development of functional compounds that can be used to visualize and regulate cellular function based on DNA sequence information using pyrrole–imidazole polyamides (PIPs). PIPs are minor-groove DNA binders composed of *N*-methylpyrrole (Py) and *N*-methylimidazole (Im) (Figure 1).[1, 2] The advantage of PIPs is that we can program their binding sequence by modifying the combination of Py and Im rings. In this section, we describe rules of PIP design and applications of PIPs other than gene regulation: DNA probing and secondary structure control.

[a] T. Hidaka, H. Sugiyama
Department of Chemistry
Graduate School of Science, Kyoto University
Kitashirakawa-Oiwakecho, Sakyo-ku, Kyoto 606-8502, Japan
E-mail: hs@kuchem.kyoto-u.ac.jp
[b] H. Sugiyama
Institute for Integrated Cell-Material Sciences (WPI-iCeMS)
Kyoto University
Yoshida-Ushinomaecho, Sakyo-ku, Kyoto 606-8501, Japan

Takuya Hidaka joined Prof. Hiroshi Sugiyama group at Kyoto University in 2015 and started his doctoral study in 2018. He has been working on chemical biology researches and his research focuses on the development of functional sequence-specific DNA binders which can modulate DNA transcription and replication in living cells.

Hiroshi Sugiyama received his Ph.D. in 1984 with Teruo Matsuura at Kyoto University. After postdoctoral studies at the University of Virginia with Sidney M. Hecht, he returned to Kyoto University in 1986 as an Assistant Professor and became an Associate Professor in 1993. In 1996 he joined the Institute of Biomaterials and Bioengineering at Tokyo Medical and Dental University. He has been a Professor of Chemical Biology at Kyoto University since 2003. Among the honors he has received are the Nippon IBM Award and the Chemical Society of Japan Award for Creative Work.

2.1. Design of PIPs

DNA binding molecules have been studied for many years for their biological activity as antitumor and antibacterial drugs.[3, 4] Netropsin and distamycin A, which contain two or three Py rings, respectively, bind to the minor groove of AT-rich sequences. In 1985, Dickerson and colleagues solved the crystal structure of a 1:1 complex of netropsin and double-stranded DNA containing 5'-d(AATT)-3' sequence.[5] They showed that hydrogen bonds were formed between the amide of netropsin and the N-3 group of adenine and O-2 group of thymine. Based on the crystal structure, Dickerson and Lown proposed that selective binding to G/C base pairs can be achieved by substituting imidazole for pyrrole, because imidazole provides a nitrogen atom that can be an acceptor of a hydrogen bond with the N-2 amino group of guanines.[5, 6] After Wemmer and colleagues reported the homodimeric DNA binding of distamycin A and a four-ring Im-Py-Im-Py,[7, 8] Dervan and colleagues revealed the general rule of base-pair recognition by PIPs containing two antiparallel strands.[9] When we design PIPs targeting specific sequences, we are guided by the following principles:

1. To promote the antiparallel alignment of two PIP strands, PIPs that have two strands joined by linkers have been widely used. Structures such as cyclic-,[10] hairpin-,[11, 12] and H-pin-type PIPs have been proposed to date (Figure 1 and Figure 2a).[13, 14] Hairpin-type PIPs are the primary choice because of their ease of synthesis by solid-phase synthesis, [15, 16] but cyclic-type PIPs were reported to have better binding affinity and selectivity than hairpin-type PIPs. [17-20]
2. Antiparallel Py/Py pairs recognize A/T or T/A base pairs by forming two hydrogen bonds, whereas Py/Im pairs recognize G/C base pairs with three hydrogen bonds (Figure 2b). [21, 22] We usually use 4-4 or 6-6 ring-type hairpin PIPs, which have two strands containing four or six pyrrole/imidazole rings each. We can program their target sequences in four or six base pair lengths by rearranging the combination of pyrrole and imidazole rings.
3. Due to over-curvature, PIPs targeting DNA sequences longer than five base pairs do not fit into the shape of the DNA minor groove, reducing their affinity and selectivity. Replacing pyrroles with flexible β-alanine relaxes the curvature and allows efficient binding of long PIPs. [23] In the case of 6-6 ring-type PIPs, one β-alanine should be introduced instead of pyrrole at the central position in each strand (Figure 2c).
4. The turn unit (γ-aminobutyric acid) and β-alanine located at the C-terminal structure of PIPs are known to have A/T or T/A selectivity.[24, 25] Based on our experience, this effect is critical for turn units and the desired affinity is not obtained when turn units locate on G/C or C/G base pairs.

Generally, the binding affinity and sequence selectivity of PIPs are routinely evaluated by melting curve assay of the DNA–PIP complex, surface plasmon resonance (SPR) assay, and electrophoretic mobility shift assay (EMSA). High-throughput analyses based on microarrays and next-generation sequencing have also been reported to obtain more detailed information about sequence specificity in various sequence contexts and binding sites on the genome. [26-30]

2.2. Application of PIPs for DNA probing and DNA structure control

Sequence-selective DNA binding of PIPs enables us to deliver payload compounds to specific genomic sites. One application is the visualization of the telomere structures in nuclei. Telomeres locate at the terminal of chromosomes and have important roles in maintaining chromosome stability and replication. Human telomeres have a repetitive 5′-(TTAGGG)$_n$-3′ sequence and, to visualize this region, tandem PIPs that target 12, 18, and 24 bp sequences in telomeres conjugated with fluorescent compounds have been reported.[31-36] An advantage of PIP probes is that because PIPs bind to duplex DNA, denaturation steps such as heating, which are indispensable when nucleic acid probes are used, are not required for probing. This advantage enables us to perform live-cell imaging of telomeres, and we recently reported dynamics of telomeres in live cells using a PIP probe labeled with silicon rhodamine (SiR-TTet59B, Figure 3a) that targets a 24 bp sequence.[37]

It is also possible to control the secondary structure of DNA by giving local force to DNA. In 2018, we reported that G-quadruplex formation can be induced by pulling two DNA sequences flanking a G-quadruplex-forming sequence toward each other using PIP dimers (Figure 3b).[38] These studies indicate that PIP is a good platform to achieve artificial gene control based on the genomic sequence.

3. Development of artificial transcription factors using pyrrole–imidazole polyamides

Living organisms are maintained by a huge number of cells that have a wide range of functions. To achieve each specialized function, DNA binding proteins that are termed transcription factors, play an important role in regulating specific gene networks in a spatiotemporal manner. This suggests that mimicking the function of natural transcription factors with chemical compounds would enable us to artificially control gene expression, and this motivated us to work on the development of artificial transcription factors using PIPs. We found that there are three functions of transcription factors that should be mimicked by our chemical approaches: i) efficient delivery into organelles that contain the target DNA, ii) specific DNA binding to the target genomic region, and iii) regulation of gene expression by interacting with other transcription coregulators (Figure 4). In the following sections, we explain our approaches to achieve each function using PIPs.

3.1. Efficient delivery of PIPs into organelles

The efficient delivery of PIPs into organelles that contain DNA (i.e. nuclei and mitochondria) is important for the development of PIP-based artificial transcription factors. The cellular permeability of PIPs and their intracellular distribution have been extensively studied by flow cytometry and fluorescence imaging using

fluorescently labeled PIPs. Although some PIPs have been reported to accumulate in nuclei efficiently due to their affinity to DNA,[39] the permeability of PIP varies depending on a range of structural factors such as molecular size, pyrrole/imidazole content, the fluorescent group, and chemical modification at the turn and terminal structures.[40-44] This makes it difficult to predict their uptake efficiency a priori, so several approaches to achieve robust and efficient delivery of PIPs have been reported. For example, Dervan group reported that the introduction of isophthalic acid or fluorescein-5-isothiocyanate at the C-terminal and an aryl group at the turn moiety enhance cellular uptake of 4-4 ring hairpin-type PIPs.[45, 46]

In general, whereas short hairpin PIPs (e.g. 4-4 ring type) are favorable for efficient delivery, long PIPs (e.g. 6-6 ring type) offer advantages in binding affinity to DNA and sequence selectivity.[47] To improve this 'trade-off' relationship, we introduced a tri-arginine moiety into a 6-6 hairpin-type PIP targeting the SOX2 binding motif (Figure 5a). Flow cytometry analysis revealed that the tri-arginine vector improved the cellular uptake efficiency and the PIP and tri-arginine conjugates successfully accumulated into the nuclei.[48] The mechanism of enhanced cellular uptake by tri-arginine is not fully understood, but it is expected that increased local concentration of PIPs near the cell membrane resulting from electrostatic interaction between the tri-arginine moiety (positively charged) and cell membrane (negatively charged) and formation of hydrogen bonds between guanidino groups and sulfates of cellular components have important roles.[49, 50]

Although most genetic information is stored in the nuclear DNA, mitochondria also possess ca. 16.6 kbp circular DNA (mtDNA) encoding 37 genes, including 13 essential subunits of the oxidative phosphorylation system. A high level of reactive oxygen species in mitochondria generated from the latter system cause mutations in mtDNA, and some pathogenic mutations have been gaining attention as important therapeutic targets.[51]

To target mtDNA with PIPs, we developed "MITO-PIPs", which are conjugates of PIPs with mitochondria-penetrating peptides composed of cyclohexylalanine and arginine residues (Figure 5b).[52, 53] The mitochondrial membrane has a membrane potential that is three times higher than that of the cellular membrane, and positive charges of arginine residues generate the driving force for membrane penetration while hydrophobic cyclohexylalanines can reduce the energy barrier for permeation. The MITO-PIP targeting binding motif of mitochondrial transcription factor A (TFAM) was successfully delivered into mitochondria and repressed expression of a downstream gene (*MT-ND6*) by inhibiting DNA binding of TFAM.

As described above, it is possible to control the intracellular distribution of PIPs by using specific peptides, and both nuclear and mitochondrial DNA can be targeted by PIPs.

3.2. Site-specific binding of PIPs

After PIPs are delivered into organelles, they need to bind to specific sites in the gene control region. Natural transcription factors have a preference for DNA binding sequences (motifs) that are typically 4–10 bp in length. This sequence-specific binding can be easily achieved with PIPs by rearranging the combination of pyrrole and imidazole rings to program their target sequence. This enables us to achieve competitive inhibition of DNA binding of natural transcription factors to repress downstream gene expression. For example, a 6-6 ring hairpin-type PIP targeting the SOX2 binding sequence (5′- CTTTGTT) can repress the expression of SOX2-downstream genes, including SOX2 itself, and has been applied to mesoderm differentiation of induced pluripotent stem (iPS) cells and as an anticancer agent.[47, 54] In another study, a PIP that inhibits DNA binding of the REL/ELK1 to the minimal promoter region of the oncogenic *EVI1* gene showed anticancer activity, reducing migration and invasion activity of breast cancer cells.[55] Notably, such competitive inhibition indicates that PIPs (especially 6-6 ring PIPs) have DNA binding affinity comparable to that of natural transcription factors.

The recognition sequences of transcription factors are 4–10 bp long, which is short compared with the size of genomic DNA (ca. 3 billion bp) and there should be many possible binding sites on the genome. This is not favorable for spatial gene regulation and the natural system overcomes this problem by functioning as cooperative transcription factor pairs. In mammals, there are approximately 1000 transcription factors and it is expected that 55–70% of them work cooperatively by forming homo/heterodimers to ensure high binding affinity and extended recognition sequences.[56, 57]

To mimic this cooperative DNA recognition, we developed a "PIP-HoGu (host–guest)" system (Figure 6a).[58] In this system, one pair of PIPs is designed to locate next to each other and each PIP is conjugated with cyclodextrin and adamantane. The host–guest interaction between cyclodextrin and adamantane induces heterodimer formation of PIPs. Indeed, a reporter assay using luciferase gene, which is transcriptionally controlled under estrogen receptor α (ERα) homodimer, revealed that the reporter gene expression was repressed by this system more effectively compared with the repression efficiency obtained by individual PIPs.

We improved the PIP-HoGu system by changing the dimer formation domain from the cyclodextrin–adamantane pair to left-handed (LH) γPNA pairs, reported as "PIP-NaCo (nucleic acid-based cooperation system)" (Figure 6b).[59] Because dimer formation of PIP-NaCo is driven by duplex formation between complementary PNA strands, this system is expected to enable us to program PIP pairs that work cooperatively even in the presence of other PIPs. It is also advantageous to use LH γPNA because it forms an unnatural left-handed duplex and has no interaction with natural DNA and RNA, giving bioorthogonality to the dimer formation domain.

B-form duplex is the most dominant structure of DNA, but DNA can also take other secondary structures such as Z-form duplex and i-motif. G-quadruplex (G4) is one such non-B-form DNA structure, and extensive studies have revealed its biological roles in a wide range of cellular processes, including DNA replication, transcription, and epigenetics, which makes the G4 structure an important therapeutic target. One such therapeutic approach is the development of G4 ligands that stabilize G4 structures in the promoter region of oncogenes (e.g. c-Myc) and

repress their level of expression. Based on the reports that G4 structures recruit G4-specific protein factors for transcriptional control, they are attractive targets to give additional site-specificity to PIPs.

cIKP, which we reported in 2016, is a cyclic compound composed of lysine and *N*-methylimidazole with a flat aromatic structure.[60] cIKP preferentially binds to G4 over dsDNA and *in vitro* binding assays showed that cIKP has a higher binding affinity than the well-studied G4 ligand TMPyP4, which was enough to induce G4 formation in the dsDNA context. To recognize both G4 structure and an adjacent dsDNA sequence simultaneously, hybrid molecules of hairpin-type PIPs covalently linked with cIKP (Figure 6c) were examined. It was confirmed by circular dichroism (CD)/Förster resonance energy transfer (FRET) melting, SPR, and CD/NMR titration assays that the hybrid molecules synergistically recognize both G4 structure and dsDNA.[61] This approach indicates that structural information of DNA can also be utilized to regulate the binding site of PIPs, and that other non-B-form DNA structures can be targeted if specific ligands are available.

3.3. Gene expression control by PIPs

The PIPs localized at the targeted sites on the genome should cause transcriptional changes to act as artificial transcription factors. As described in Section 3.2, specific gene expression can be repressed by inhibiting the DNA binding of natural transcription factors that induce downstream gene transcription. While downregulation can be achieved by this simple approach, artificial gene activation is tricky because we need to utilize the natural transcription system in our chemical approaches.

In eukaryotic cells, genomic DNA is wrapped around histone octamers, forming nucleosomes and genome-wide chromatin structure. Nucleosomes restrict the access of transcription factors and act as an additional layer to regulate gene expression. Generally, transcription factors do not bind to highly packed nucleosomal DNA, and genes in this region are transcriptionally silenced. When nucleosomes are loosened and unfolded, released DNA is subjected to the binding of transcription factors initiating gene transcription. This switching between “OFF” and “ON” states is controlled by epigenetic modification of histone proteins. Lysine acetylation in the tail region is one of the histone modifications that is extensively studied. When lysin residues are acetylated by enzymes with histone acetyltransferase (HAT) activity, the positive charge of the amine group is neutralized and electrostatic interaction between DNA and histone is weakened. This loosens the nucleosome structure and gene expression is activated, so acetylation at K9 and K27 of histone H3 is recognized as a marker of active transcription. Conversely, histone deacetylation is mediated by histone deacetylase (HDAC) and downregulates gene expression (Figure 7a).

In our earlier study to develop artificial transcription factors that activate gene expression, we focused on suberoylanilide hydroxamic acid (SAHA), an HDAC inhibitor that is used as an anticancer drug (trade name Zolinza). Because SAHA can keep the genome in the transcriptionally ON state by inhibiting histone deacetylation, SAHA–PIP conjugates were expected to bind to a programmed sequence and activate genes around the binding site (Figure 7b).[62] To examine our hypothesis, we constructed a SAHA–PIP library containing 32 SAHA–PIPs targeting distinct sequences. Transcriptome analysis of human dermal fibroblasts treated with each SAHA–PIP revealed that transcription of different sets of genes was activated depending on the target sequence of SAHA–PIPs, supporting their sequence-specific gene activation (Figure 7c).[63]

In the following study, we used a HAT activator, *N*-(4-chloro-3-(trifluoromethyl) phenyl)-2-ethoxybenzamide (CTB), which was expected to have a similar effect to SAHA (Figure 7b). Interestingly, while different genes were activated by SAHA and CTB, SAHA–PIP and CTB–PIP conjugates targeting the same DNA sequence activated a similar set of genes (Figure 7d).[64] This result also suggests that site-specific gene activation can be achieved by combining the sequence-selectivity of PIPs and epigenetic modulation by functional compounds.

Although SAHA–PIPs and CTB–PIPs are promising artificial transcription factors, their induction of histone acetylation is an indirect outcome and acetylation would not occur if HATs were not available near the target region. To enhance histone acetylation more actively, a new type of artificial transcription factor, named Bi-PIP, was reported recently.[65] P300/CBP is a transcription coactivator with HAT activity and possesses several domains dedicated to interaction with other proteins. Among the interaction domains, bromodomain recognizes acetylated lysine of histone and, by working with the HAT domain, P300/CBP can acetylate other histones around the recognized one successively. Because Bi-PIP has a bromodomain inhibitor (Bi) that interacts with the bromodomain of P300/CBP, it can recruit the enzyme to a programmed sequence, promoting histone acetylation around the binding sites (Figure 8a). *In vitro* ChIP-qPCR showed sequence-selective histone acetylation by two different Bi-PIPs (Bi-PIP1 and Bi-PIP2), and the target sequence of each compound was enriched in differential ChIP-seq peaks of acetylated histone H3 between Bi-PIP1 and Bi-PIP2 (Figure 8b and c). Bi-PIP is expected to become one of the next generation of artificial transcription factors that control genes that cannot be targeted by SAHA–PIPs and CTB–PIPs.

4. Summary and Outlook

In this Personal Account, we introduced our approaches to convert PIPs into artificial transcription factors. The artificial gene regulation has great potential as a therapeutic approach, but careful optimization of PIP structure is required for *in vivo* application. Because PIPs are relatively large and pyrrole/imidazole rings are hydrophobic, structural factors such as charge, molecular size and conformation affect the solubility of PIPs in aqueous buffers significantly.[66] This optimization is also crucial for pharmacokinetics and toxicity *in vivo*.[67-69]

In the studies so far, we achieved each function required as transcription factors separately by using different chemical modifications. These approaches have been revealed to partially mimic the functions of natural transcription factors, but their selectivity was not comparable with that of the natural

transcription factors. To develop artificial transcription factors that can perform more precise gene regulation similar to that of natural transcription factors, these approaches should be used in combination to mimic the modularity of protein domains of natural transcription factors. For example, we recently reported a combinatorial system of PIP-HoGu and Bi-PIP that synergistically recruits P300 to the target site and enhances histone acetylation *in vitro*.[70] This system expands the length of the recognized DNA sequence and is expected to enable gene regulation in a more site-specific manner.

We believe that the PIP-based artificial transcription factors would be powerful tools for biological studies and drug candidates to correct abnormal gene regulation for therapeutic purposes.

Acknowledgements

This work was supported by AMED [JP18am0301005 (Basic Science and Platform Technology Program for Innovative Biological Medicine) and JP20am0101101 (Platform Project for Supporting Drug Discovery and Life Science Research (BINDS) to H.S.)] and JSPS [18J21755 (Grant-in-Aid for JSPS Fellows) to T.H. and 16H06356 (Grants-in-Aid for Scientific Research (S)) to H.S.].

Keywords: pyrrole–imidazole polyamide • artificial transcription factor • gene regulation • minor groove binder • epigenetics

[1] J. W. Trauger, E. E. Baird, P. B. Dervan, *Nature* **1996**, *382*, 559-561.
[2] D. M. Chenoweth, P. B. Dervan, *J. Am. Chem. Soc.* **2010**, *132*, 14521-14529.
[3] P. G. Baraldi, A. Bovero, F. Fruttarolo, D. Preti, M. A. Tabrizi, M. G. Pavani, R. Romagnoli, *Med. Res. Rev.* **2004**, *24*, 475-528.
[4] *DNA-targeting Molecules as Therapeutic Agents* (Ed.: M. J. Waring), The Royal Society of Chemistry, London, **2018**.
[5] M. L. Kopka, C. Yoon, D. Goodsell, P. Pjura, R. E. Dickerson, *J. Mol. Biol.* **1985**, *183*, 553-563.
[6] J. W. Lown, K. Krowicki, U. G. Bhat, A. Skorobogaty, B. Ward, J. C. Dabrowiak, *Biochemistry* **1986**, *25*, 7408-7416.
[7] J. G. Pelton, D. E. Wemmer, *Proc. Natl. Acad. Sci. USA* **1989**, 86, 5723-5727.
[8] B. H. Geierstanger, M. Mrksich, P. B. Dervan, D. E. Wemmer, *Science* **1994**, *266*, 646-650.
[9] S. White, E.E. Baird, P.B. Dervan, *Chem. Biol.* **1997**, *4*, 569-578.
[10] Junhyeong Cho, Michelle E. Parks, P. B. Dervan, *Proc. Natl. Acad. Sci. USA* **1995**, *92*, 10389-10392.
[11] M. Mrksich, M. E. Parks, P. B. Dervan, *J. Am. Chem. Soc.* **1994**, *116*, 7983-7988.
[12] J. M. Gottesfeld, L. Neely, J. W. Trauger, E. E. Baird, P. B. Dervan, *Nature* **1997**, *387*, 202-205.
[13] M. Mrksich, P. B. Dervan, *J Am Chem Soc* **1994**, *116*, 3663-3664.
[14] H. Mackay, T. Brown, P. B. Uthe, L. Westrate, A. Sielaff, J. Jones, J. P. Lajiness, J. Kluza, C. O'Hare, B. Nguyen, Z. Davis, C. Bruce, W. D. Wilson, J. A. Hartley, M. Lee, *Bioorg. Med. Chem.*. **2008**, *16*, 9145-9153.
[15] E. E. Baird, P. B. Dervan, *J. Am. Chem. Soc.* **1996**, *118*, 6141-6146.
[16] N. R. Wurtz, J. M. Turner, E. E. Baird, P. B. Dervan, *Org. Lett.* **2001**, *3*, 1201-1203.
[17] B. C. Li, D. C. Montgomery, J. W. Puckett, P. B. Dervan, *J. Org. Chem.* **2013**, *78*, 124-133.
[18] D. M. Herman, J. M. Turner, E. E. Baird, P. B. Dervan, *J. Am. Chem. Soc.* **1999**, *121*, 1121-1129.
[19] D. M. Chenoweth, D. A. Harki, P. B. Dervan, *J. Am. Chem. Soc.* **2009**, *131*, 7182-7188.
[20] Y. Hirose, K. Hashiya, T. Bando, H. Sugiyama, *Chem. Eur. J.* in press.
[21] 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.
[22] C. L. Kielkopf, E. E. Baird, P. B. Dervan, D. C. Rees, *Nat. Struct. Bio.* **1998**, *5*, 104-109.
[23] James M. Turner, Susanne E. Swalley, Eldon E. Baird, P. B. Dervan, *J. Am. Chem. Soc.* **1998**, *120*, 6219-6226.
[24] Michelle E. Parks, Eldon E. Baird, P. B. Dervan, *J. Am. Chem. Soc.* **1996**, *118*, 6147-6152.
[25] S. E. Swalley, E. E. Baird, P. B. Dervan, *J. Am. Chem. Soc.* **1999**, *121*, 1113-1120.
[26] J. W. Puckett, K. A. Muzikar, J. Tietjen, C. L. Warren, A. Z. Ansari, P. B. Dervan, *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 867-872.
[27] J. L. Meier, A. S. Yu, I. Korf, D. J. Segal, P. B. Dervan, *J. Am. Chem. Soc.* **2012**, *134*, 17814-17822.
[28] G. S. Erwin, D. Bhimsaria, A. Eguchi, A. Z. Ansari, *Angew. Chem. Int. Ed. Engl* **2014**, *53*, 10124-10128.
[29] G. S. Erwin, M. P. Grieshop, D. Bhimsaria, T. J. Do, J. A. Rodriguez-Martinez, C. Mehta, K. Khanna, S. A. Swanson, R. Stewart, J. A. Thomson, P. Ramanathan, A. Z. Ansari, *Proc. Natl. Acad. Sci. USA* **2016**, *113*, 7418-7427.
[30] A. Chandran, J. Syed, Y. Li, S. Sato, T. Bando, H. Sugiyama, *Chembiochem* **2016**, *17*, 1905-1910.
[31] K. Maeshima, S. Janssen, U. K.Laemmli, *EMBO J.* **2001**, *20*, 3218-3228.
[32] 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.
[33] Y. Kawamoto, T. Bando, F. Kamada, Y. Li, K. Hashiya, K. Maeshima, H. Sugiyama, *J. Am. Chem. Soc.* **2013**, *135*, 16468-16477.
[34] Y. Kawamoto, A. Sasaki, A. Chandran, K. Hashiya, S. Ide, T. Bando, K. Maeshima, H. Sugiyama, *J. Am. Chem. Soc.* **2016**, *138*, 14100-14107.
[35] Y. Kawamoto, A. Sasaki, K. Hashiya, S. Ide, T. Bando, K. Maeshima, H. Sugiyama, *Chem. Sci.* **2015**, *6*, 2307-2312.
[36] A. Sasaki, S. Ide, Y. Kawamoto, T. Bando, Y. Murata, M. Shimura, K. Yamada, A. Hirata, K. Nokihara, T. Hirata, H. Sugiyama, K. Maeshima, *Sci. Rep.* **2016**, *6*, 29261.
[37] Y. Tsubono, Y. Kawamoto, T. Hidaka, G. N. Pandian, K. Hashiya, T. Bando, H. Sugiyama, *J. Am. Chem. Soc.* **2020**, *142*, 17356-17363.
[38] S. Obata, S. Asamitsu, K. Hashiya, T. Bando, H. Sugiyama, *Biochemistry* **2018**, *57*, 498-502.
[39] Y. M. Lai, N. Fukuda, T. Ueno, H. Matsuda, S. Saito, K. Matsumoto, H. Ayame, T. Bando, H. Sugiyama, H. Mugishima, K. Serie, *J. Pharmacol. Exp. Ther.* **2005**, *315*, 571-575.
[40] J. M. Belitsky, S. J. Leslie, P. S. Arora, T. A. Beerman, P. B. Dervan, *Bioorg. Med. Chem.* **2002**, *10*, 3313-3318.
[41] T. P. Best, B. S. Edelson, N. G. Nickols, P. B. Dervan, *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 12063-12068.
[42] B. S. Edelson, T. P. Best, B. Olenyuk, N. G. Nickols, R. M. Doss, S. Foister, A. Heckel, P. B. Dervan, *Nucleic Acids Res.* **2004**, *32*, 2802-2818.
[43] C. F. Hsu, P. B. Dervan, *Bioorg. Med. Chem. Lett.* **2008**, *18*, 5851-5855.
[44] S. Nishijima, K. Shinohara, T. Bando, M. Minoshima, G. Kashiwazaki, H. Sugiyama, *Bioorg. Med. Chem.* **2010**, *18*, 978-983.
[45] J. L. Meier, D. C. Montgomery, P. B. Dervan, *Nucleic Acids Res.* **2012**, *40*, 2345-2356.
[46] N. G. Nickols, C. S. Jacobs, M. E. Farkas, P. B. Dervan, *Nucleic Acids Res.* **2007**, *35*, 363-370.
[47] J. Taniguchi, G. N. Pandian, T. Hidaka, K. Hashiya, T. Bando, K. K. Kim, H. Sugiyama, *Nucleic Acids Res.* **2017**, *45*, 9219-9228.

[48] T. Hidaka, Y. Tsubono, K. Hashiya, T. Bando, G. N. Pandian, H. Sugiyama, *Chem. Commun.* **2020**, *56*, 12371-12374.
[49] J. B. Rothbard, T. C. Jessop, R. S. Lewis, B. A. Murray, P. A. Wender, *J. Am. Chem. Soc.* **2004**, *126*, 9506-9507.
[50] I. Nakase, A. Tadokoro, N. Kawabata, T. Takeuchi, H. Katoh, K. Hiramoto, M. Negishi, M. Nomizu, Y. Sugiura, S. Futaki, *Biochemistry* **2007**, *46*, 492-501.
[51] G. S. Gorman, P. F. Chinnery, S. DiMauro, M. Hirano, Y. Koga, R. McFarland, A. Suomalainen, D. R. Thorburn, M. Zeviani, D. M. Turnbull, *Nat. Rev. Dis. Primers* **2016**, *2*, 16080.
[52] K. L. Horton, K. M. Stewart, S. B. Fonseca, Q. Guo, S. O. Kelley, *Chem. Biol.* **2008**, *15*, 375-382.
[53] T. Hidaka, G. N. Pandian, J. Taniguchi, T. Nobeyama, K. Hashiya, T. Bando, H. Sugiyama, *J. Am. Chem. Soc.* **2017**, *139*, 8444-8447.
[54] M. Malinee, A. Kumar, T. Hidaka, M. Horie, K. Hasegawa, G. N. Pandian, H. Sugiyama, *Bioorg. Med. Chem.* **2020**, *28*, 115248.
[55] J. Syed, G. N. Pandian, S. Sato, J. Taniguchi, A. Chandran, K. Hashiya, T. Bando, H. Sugiyama, *Chem. Biol.* **2014**, *21*, 1370-1380.
[56] G. Stampfel, T. Kazmar, O. Frank, S. Wienerroither, F. Reiter, A. Stark, *Nature* **2015**, *528*, 147-151.
[57] E. Morgunova, J. Taipale, *Curr. Opin. Struct. Biol.* **2017**, *47*, 1-8.
[58] Z. Yu, C. Guo, Y. Wei, K. Hashiya, T. Bando, H. Sugiyama, *J. Am. Chem. Soc.* **2018**, *140*, 2426-2429.
[59] Z. Yu, W. C. Hsieh, S. Asamitsu, K. Hashiya, T. Bando, D. H. Ly, H. Sugiyama, *Chem. Eur. J.* **2018**, *24*, 14183-14188.
[60] S. Asamitsu, Y. Li, T. Bando, H. Sugiyama, *Chembiochem* **2016**, *17*, 1317-1322.
[61] S. Asamitsu, S. Obata, A. T. Phan, K. Hashiya, T. Bando, H. Sugiyama, *Chem. Eur. J.* **2018**, *24*, 4428-4435.
[62] A. Ohtsuki, M. T. Kimura, M. Minoshima, T. Suzuki, M. Ikeda, T. Bando, H. Nagase, K. Shinohara, H. Sugiyama, *Tetrahedron Lett.* **2009**, *50*, 7288-7292.
[63] G. N. Pandian, J. Taniguchi, S. Junetha, S. Sato, L. Han, A. Saha, C. AnandhaKumar, T. Bando, H. Nagase, T. Vaijayanthi, R. D. Taylor, H. Sugiyama, *Sci. Rep.* **2014**, *4*, 3843.
[64] 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.
[65] J. Taniguchi, Y. Feng, G. N. Pandian, F. Hashiya, T. Hidaka, K. Hashiya, S. Park, T. Bando, S. Ito, H. Sugiyama, *J. Am. Chem. Soc.* **2018**, *140*, 7108-7115.
[66] A. E. Hargrove, J. A. Raskatov, J. L. Meier, D. C. Montgomery, P. B. Dervan, *J. Med. Chem.* **2012**, *55*, 5425-5432.
[67] J. A. Raskatov, J. W. Puckett, P. B. Dervan, *Bioorg. Med. Chem.* **2014**, *22*, 4371-4375.
[68] F. Yang, N.G. Nickols, B.C. Li, J.O. Szablowski, S.R. Hamilton, J.L. Meier, C.M. Wang, P.B. Dervan, *J. Med. Chem.* **2013**, *56*, 7449-7457.
[69] T.W. Synold, B.X. Xi, J. Wu, Y. Yen, B.C. Li, F. Yang, J.W. Phillips, N.G. Nickols, P.B. Dervan, *Cancer Chemother. Pharmacol.* **2012**, *70*, 617-625.
[70] Z. Yu, M. Ai, S. K. Samanta, F. Hashiya, J. Taniguchi, S. Asamitsu, S. Ikeda, K. Hashiya, T. Bando, G. N. Pandian, L. Isaacs, H. Sugiyama, *Chem. Commun.* **2020**, *56*, 2296-2299.

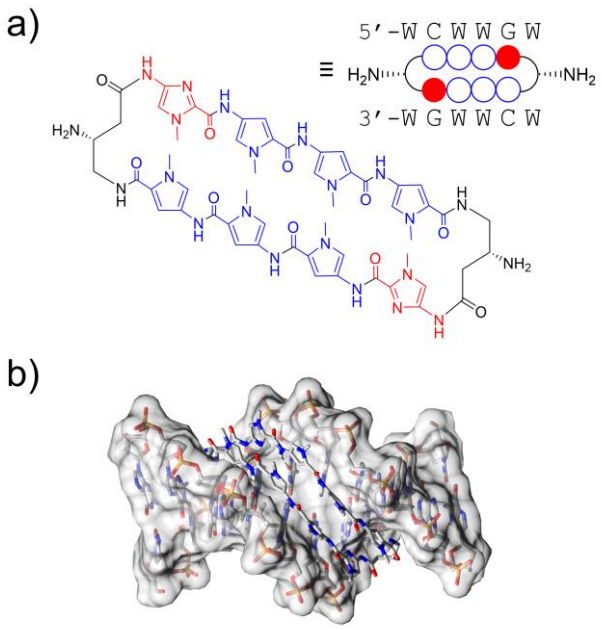


Figure 1. a) Chemical and schematic structure of a cyclic PIP and its recognition sequence. W represents adenine or thymine. b) Crystal structure of a complex of DNA and the cyclic PIP shown in Figure 1a (PDB: 3OMJ).[2]

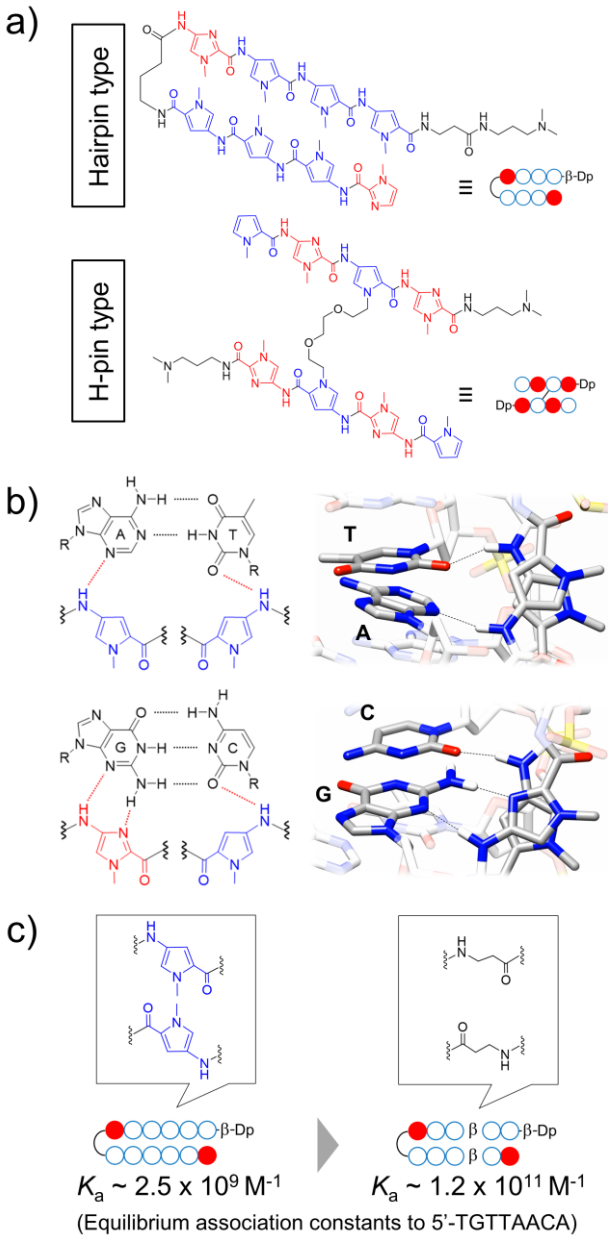


Figure 2. a) Chemical structure of hairpin and H-pin PIPs.[12, 14] Two PIP strands are joined by a linker moiety promoting an antiparallel alignment in the minor groove of duplex DNA. b) Base-pair recognition by PIPs. Pyrrole/pyrrole pairs recognize A/T or T/A base pairs whereas imidazole/pyrrole pairs recognize G/C base pairs.[21, 22] c) Introduction of flexible β-alanine instead of pyrroles into long PIPs improves their binding affinity to a target sequence.[23]

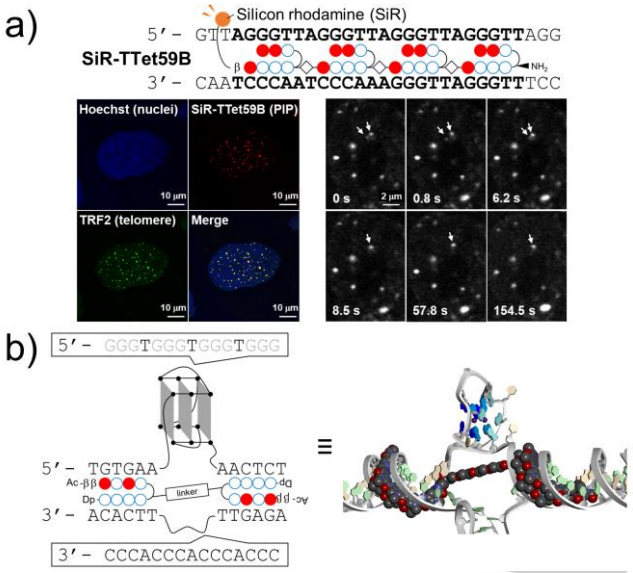


Figure 3. a) A tandem PIP targeting repetitive telomere sequence labeled with silicon rhodamine (SiR-TTet59B) can be used to visualize telomere structures. (bottom left) The signal from SiR-TTet59B overlapped with telomere structures labeled by immunostaining of TRF2. (bottom right) By introducing SiR-TTet59B using a peptide-based transfection reagent (Endo-Porter), dynamics of telomeres in living cells can be observed and telomere fusion was captured (indicated by white arrows).[37] (b) PIP dimers pulling two DNA sequences flanking a G-quadruplex-forming sequence toward each other enhance G-quadruplex formation.[38]

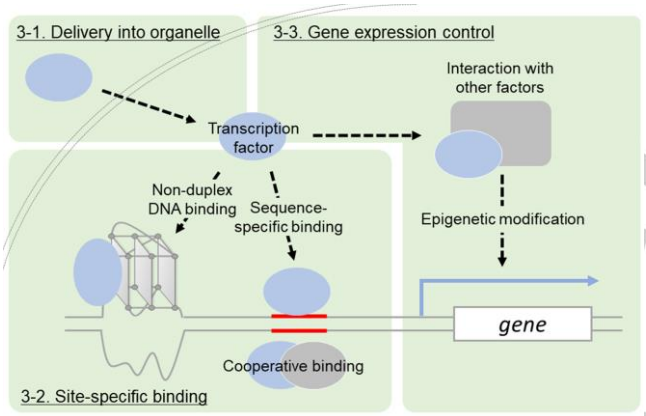


Figure 4. Three major functions of transcription factors that should be mimicked by our chemical approach using PIPs. Each function and our corresponding strategies are described in the sections noted in the figure.

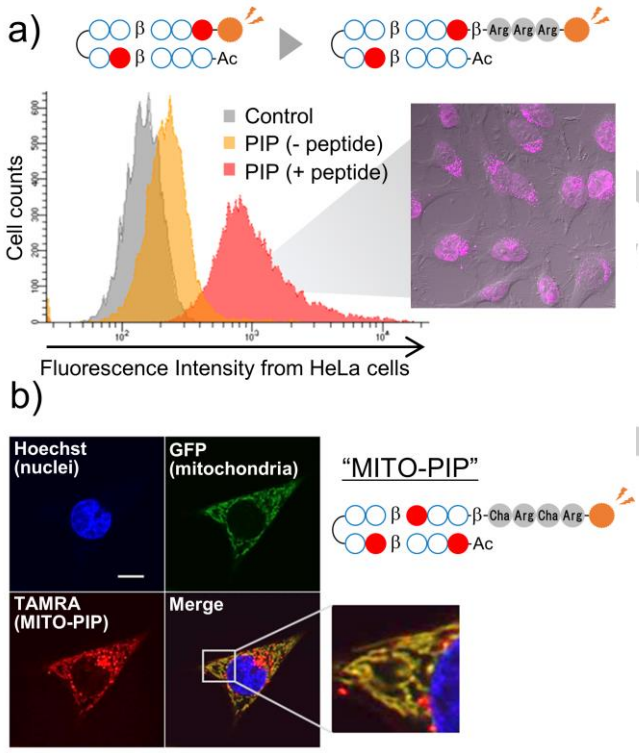


Figure 5. Organelle-specific delivery of PIPs using peptide vectors. A) Flow cytometry analysis of a TAMRA-labeled PIP conjugated with a tri-arginine vector revealed that the tri-arginine vector enhances cellular uptake of the PIP and efficient nuclear accumulation was confirmed from fluorescence images.[48] b) MITO-PIP conjugated with mitochondria-penetrating peptide composed of cyclohexylalanine and arginine residues is delivered into mitochondria. HeLa cells expressing mitochondria-specific mitochondria were treated with the TAMRA-labeled MITO-PIP. (Scale bar = 10 μm) [53]

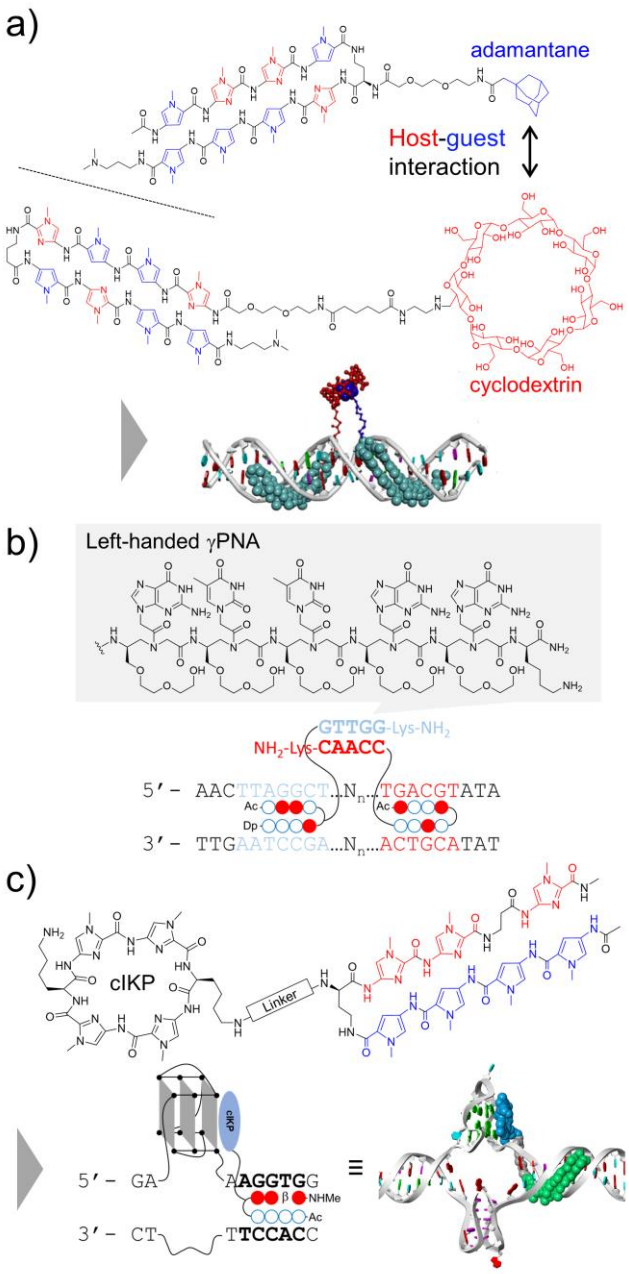


Figure 6. a) PIP-HoGu system mimicking cooperative DNA binding of natural transcription factor pairs.[58] b) PIP-NaCo has a bioorthogonal left-handed γPNA strand as a dimer formation domain.[59] c) PIP–cIKP conjugates can recognize both the G-quadruplex structure and the adjacent double-stranded DNA sequence simultaneously.[61]

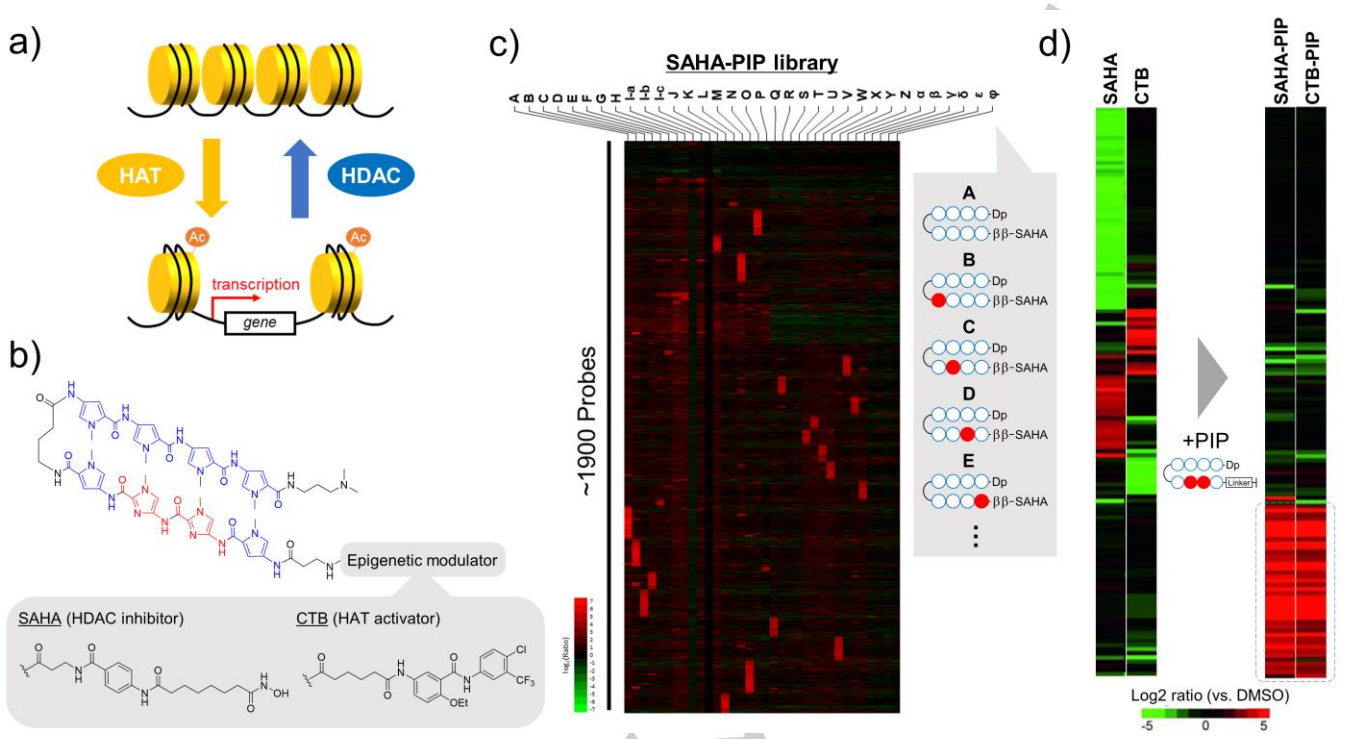


Figure 7. a) Schematic illustration of gene expression control by histone acetylation. b) Chemical structure of PIP conjugated with epigenetic modulators (SAHA and CTB) for transcription activation using histone acetylation machinery. c) Heat map of gene activation derived from transcriptome (microarray) analysis of human dermal fibroblast treated with a SAHA–PIP library containing 32 SAHA–PIPs targeting distinct sequences.[63] Each row corresponds to one probe of the microarray and red and green colors indicate that the expression of those genes is activated and repressed by SAHA–PIP, respectively. There is no significant overlap among the gene sets activated by each SAHA–PIP. d) A similar analysis to that described in (c) was performed with SAHA, CTB, SAHA–PIP, and CTB–PIP.[64] Whereas there was no overlap in activated genes between SAHA and CTB, many genes were commonly upregulated by SAHA–PIP and CTB–PIP.

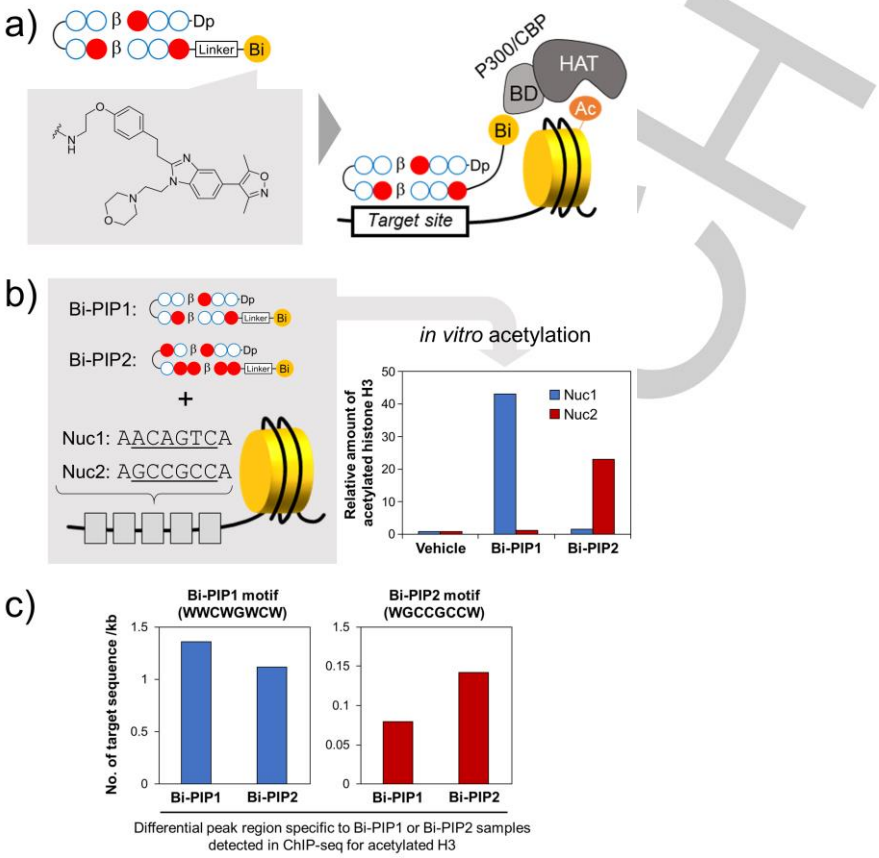


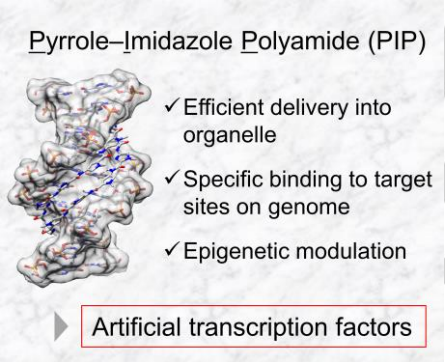
Figure 8. Compound design and activity of Bi-PIP.[65] a) Schematic illustration of gene activation mechanism by Bi-PIP. b) The *in vitro* acetylation assay using Bi-PIP1 and Bi-PIP2 and reconstituted nucleosomes containing the target sequence of each Bi-PIP. After *in vitro* acetylation reaction with P300, the relative amount of acetylated histone was quantified by ChIP-qPCR. c) Motif search analysis in differential peak region between Bi-PIP1 and Bi-PIP2 detected in ChIP-seq analysis for acetylated H3.

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

Layout 1:

PERSONAL ACCOUNT

Transcription factors have a central role in spatiotemporal control of gene expression, which has motivated us to develop artificial transcription factors by mimicking their function. We chose pyrrole–imidazole polyamides (PIPs), sequence-selective DNA binding molecules, as DNA binding domains and have achieved three functions required for artificial transcription factors: i) efficient delivery into organelles that contain target DNA, ii) specific DNA binding to a target genomic region, and iii) regulation of gene expression with other coregulators by epigenetic modulation.



*Takuya Hidaka, Hiroshi Sugiyama**

Page No. – Page No.

Chemical approaches to the development of artificial transcription factors based on pyrrole–imidazole polyamides

It would be our pleasure if our frontispiece image is also considered as a candidate for the journal Cover.