# Selective Targeting of the KRAS Codon 12 Mutation Sequence by Pyrrole–Imidazole Polyamide *seco*-CBI Conjugates

Rhys D. Taylor[a], Anandhakumar Chandran[a], Gengo Kashiwazaki[a], Kaori Hashiya[a], Toshikazu Bando*[a], Hiroki Nagase[b], Hiroshi Sugiyama*[a, c, d]

**Abstract:** Mutation of KRAS is a key step in many cancers. Mutations occur most frequently at codon 12, but targeting KRAS is notoriously difficult. We recently demonstrated selective reduction in tumors volume harboring the KRAS codon 12 mutation in a mouse model using an alkylating hairpin *N*-methylpyrrole-*N*-methylimidazole polyamide *seco*-CBI conjugate (conjugate **4**) designed to target the KRAS codon 12 mutation sequence. Here, we compared the alkylating activity of **4** against three other conjugates also designed to target the KRAS codon 12 mutation sequence. Conjugate **4** displayed greater affinity for the G12D mutation sequence than for the G12V sequence. A computer-minimized model suggested that conjugate **4** could bind more efficiently to the G12D match sequence than to a one-base pair mismatch sequence. Conjugate **4** was modified for next-generation sequencing. Bind-n-Seq analysis supported the evidence showing that conjugate **4** could target the G12D mutation sequence with exceptionally high affinity and the G12V mutation sequence with much higher affinity than that for the wild-type sequence.

## Introduction

The RAS family is a known family of oncogenes, such as HRAS, NRAS, and KRAS. This family is associated with poor prognosis[1,2] and is the most common oncogene family, with 1 in 3 cancers bearing an RAS mutation.[3,4] Around 90% of pancreatic cancers and 50% of colorectal cancers contain a KRAS mutation.[5,6] Conventional chemotherapeutic drugs such as DNA-alkylating agents, cisplatin, the anthracycline doxorubicin, and other antitumor antibiotics such as bleomycin and mitomycin-C are used routinely to treat cancer.[7] However, these drugs lack selectivity and can attack both healthy cells, notably labile cells such as bone marrow and reproductive cells, and cancerous cells.

Although there has recently been some limited success using small-molecule inhibitors designed to target KRAS,[8–10] no successful clinical inhibitors of KRAS have been developed, which has led many to think of KRAS as "undruggable."[6,11] Point mutations most frequently occur at codons 12, 13, and 61.[12–15] Mutations at codons 12 and 13 cause steric hindrance of the KRAS protein through a conformational change that inhibits GAP-induced GTP (active) hydrolysis to GDP (inactive) and causes the KRAS protein to be permanently activated.[16] Therefore, the mutation sequences at codons 12 and 13 are therapeutically attractive targets.[17]

*N*-methylpyrrole-*N*-methylimidazole polyamides are an important class of small molecules that can bind in a sequence-specific manner to the DNA minor groove.[18–20] Antiparallel pairing of imidazole (Im) opposite pyrrole (Py) recognizes G–C base pairs (bps), and a Py–Py pair recognizes A–T or T–A bps. $\gamma$-Aminobutyric acid ($\gamma$) is used as the turn moiety to form a hairpin motif that recognizes A–T or T–A bps. For hairpin Py–Im polyamides with a chain of five or more continuous Py–Im moieties, $\beta$-alanine ($\beta$) is required to ease the steric hindrance and to allow efficient binding to the minor groove, because Py–Im polyamides are overcurved compared with the DNA minor groove.[21-24] The inclusion of $\beta$ can replace Py, so that $\beta$–$\beta$, $\beta$–Py, and Py–$\beta$ pairings recognize A–T or T–A bps in alkylating Py–Im polyamides.[25] It is important to consider the positioning of $\beta$ because it can influence both the binding orientation[26] and the binding affinity[27,28] of the Py–Im polyamide in the minor groove.

The binding of transcription factors can be inhibited to selectively silence genes using Py–Im polyamides.[29] However, many genes share the same transcription factors, and targeting unique sequences in the coding region is preferable. RNA polymerase removes Py–Im polyamides during transcription. Conjugating an alkylating moiety to a Py–Im polyamide can form a covalent adduct with N3 of adenine at a predetermined sequence. Alkylating Py–Im polyamides can selectively target the coding region that inhibits transcription[30,31] and can selectively silence genes.[32] For efficient alkylation, we conjugate Py–Im polyamides with the alkylating moiety, 1,2,9,9a-tetrahydrocyclopropa[1,2-*c*]benz[1,2-*e*]indol-4-one (CBI)[33–35] via an indole linker.[36] Recently, we demonstrated selective targeting of the KRAS codon 13 mutation sequence using *seco*-CBI–Py–Im polyamides.[37] Conjugate **4** has also recently been shown to selectively target tumors harboring the KRAS codon 12 mutation in a mouse model.[38] Here, we report on the selective targeting of the KRAS codon 12 mutation sequence using *seco*-CBI–Py–Im polyamides. Conjugate **4** and three other conjugates were designed with the same core recognition motif to target the KRAS codon 12 mutation sequence, but with the $\beta$ pairing at various positions (Figure 1). The alkylating activity of these compounds was evaluated using high-resolution denaturing gel electrophoresis. Based on these results, conjugate **4** was selected for further evaluation. Using Bind-n-Seq with 10 bp randomized oligonucleotide sequences and computer-minimized structures, we confirmed its high affinity for the codon 12 mutation sequence over other sequences, demonstrating a potential route for gene sequence-based cancer therapy.

[a] Dr. R. D. Taylor, A. Chandran, Dr. G. Kashiwazaki, 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] Prof. Dr. H. Nagase
Chiba Cancer Center Research Institute, Chuo-Ku, Chiba, 260-8717 (Japan)
[c] Prof. Dr. H. Sugiyama
Institute for Integrated Cell-Materials Sciences (iCeMS) Kyoto University, Yoshida-ushinomiyacho, Sakyo-Ku, Kyoto, 606-8501 (Japan)
[d] Prof. Dr. H. Sugiyama
CREST (Japan) Science and Technology Corporation (JST), Sanbancho, Chiyoda-Ku, Tokyo, 102-0075 (Japan)

## Results and Discussion

**Molecular Design and Synthesis of Alkylating Py–Im Polyamide Conjugates Targeting the KRAS Codon 12 Mutation.**

In human and rodent tumors, the most frequent point mutations in KRAS are at codon 12 and involve the changes GGT→GAT (G12D) and GGT→GTT (G12V),[39] which change the amino acid code from glycine to aspartic acid and to valine, respectively. Four alkylating hairpin Py–Im polyamides were designed with the same core moiety to recognize a 9 bp sequence containing the KRAS codon 12 mutation site and to alkylate A in the template strand (Figure 1). To strongly block RNA polymerase, the DNA alkylation must target the template strand.[30] Binding occurs in the N–C direction with regard to the 5′–3′ direction of the DNA duplex[40] and recognizes the sequences 5′-TGWTGGCGT-3′ of the coding strand and 5′-ACGCCAWCA-3′ for the template strand, with W denoting A/T of the G12D and G12V sequence.

The chemical structures of conjugates **1–4** differ only in the position of the $\beta$ pair used in place of two Py moieties, which results in different combinations of Py–Py, $\beta$–Im, $\beta$–Py, and $\beta$–$\beta$. This difference causes minor conformational changes in the structures of **1–4**, which affect their binding affinity for the DNA minor groove.[41,42] The Py–Im polyamide carboxylic acids were synthesized by Fmoc solid-phase synthesis on an oxime resin for **1**, **2,** and **4**, or on $\beta$-Wang resin for **3**, followed by cleavage using NaOH or TFA, respectively. The Py–Im polyamide carboxylic acids were converted into **1–4** by coupling with $NH_2$-indole-*seco*-CBI via an activated ester using PyBOP.[43–45]

Following HPLC purification, conjugates **1–4** were verified by reversed-phase HPLC and ESI–TOF mass spectrometry. The sequence-specific DNA-alkylating activity of conjugates **1–4** against the KRAS codon 12 mutation sequence was evaluated using high-resolution denaturing gel electrophoresis.

**Evaluation of Specific DNA-Alkylating Activities against KRAS Codon 12 Sequences.**

The alkylating activity against the KRAS codon 12 mutation was evaluated using 5′-Texas Red-labeled 208 bp DNA fragments and an automated DNA sequencer. A 16 bp oligonucleotide was ligated into the pGEM-T Easy vector. The top T7 strand contained the KRAS G12D sequence 5′-ACGCCATCA-3′, and the bottom SP6 strand contained the KRAS G12V sequence 5′-ACGCCAACA-3′.

DNA alkylation was conducted at 25°C for 18 h and was then quenched with calf thymus DNA. DNA alkylation was visualized by thermal cleavage of the 5′-Texas Red-labeled DNA strands at the alkylated sites,[46] which showed cleavage bands quantitatively on the polyacrylamide gel. Sanger sequencing was used to identify the alkylation sites.[47–50] Figure 2 shows the alkylation by conjugates **1–4** of the T7 DNA strand containing the KRAS G12D mutation, and Figure 3 shows the alkylation by conjugates **1–4** of the SP6 DNA strand containing the KRAS G12V mutation.

Conjugate **1** displayed very weak alkylation of the G12D mutation sequence at 500 nM and 1 $\mu$M but failed to show any alkylating activity over the entire concentration range for the G12V mutation sequence. This can probably be attributed to the

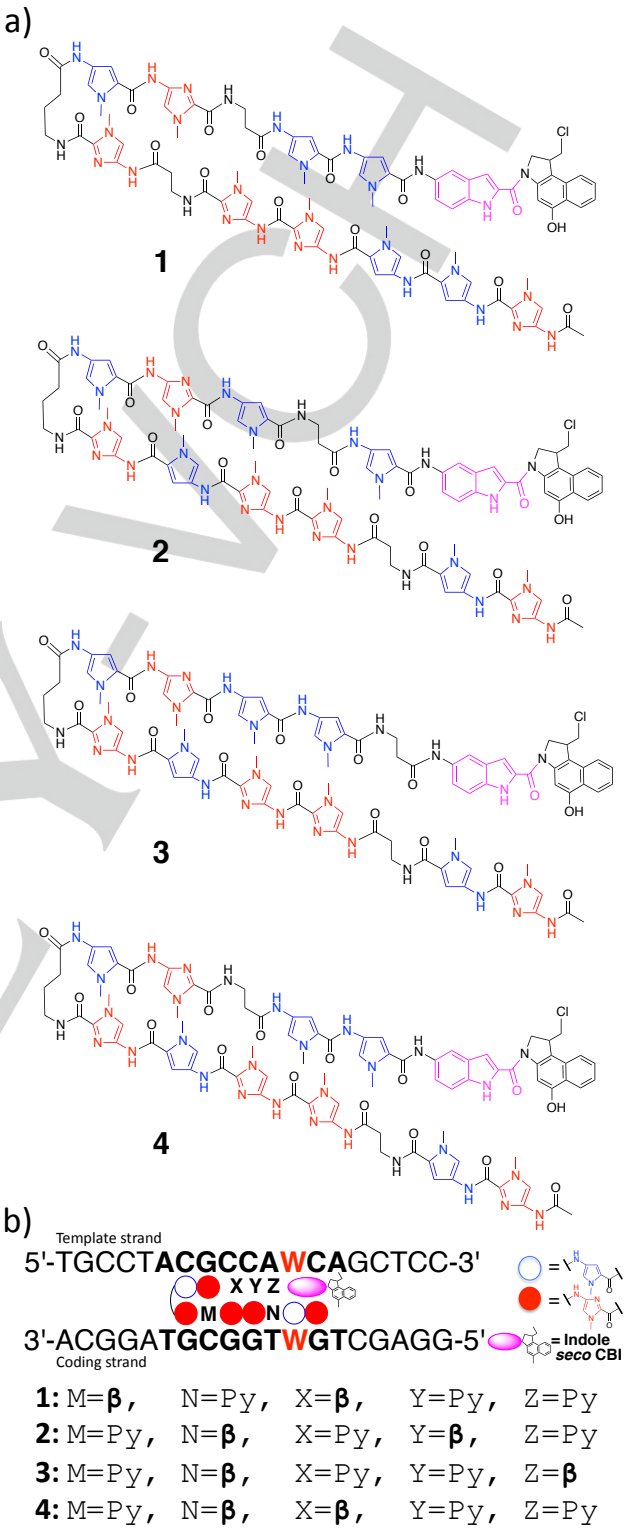



**Figure 1.** a) Chemical structures of Py–Im polyamide *seco*-CBI conjugates **1–4**. b) Schematic representation of DNA alkylation of the KRAS G12D mutation sequence. X, Y, Z, M, and N represent the possible substitution of Py for $\beta$, indicated to the right. W represents A/T. 5′-GAT-3′ on the coding strand represents the G12D mutation, and 5′-GTT-3′ on the coding strand represents the G12V mutation.

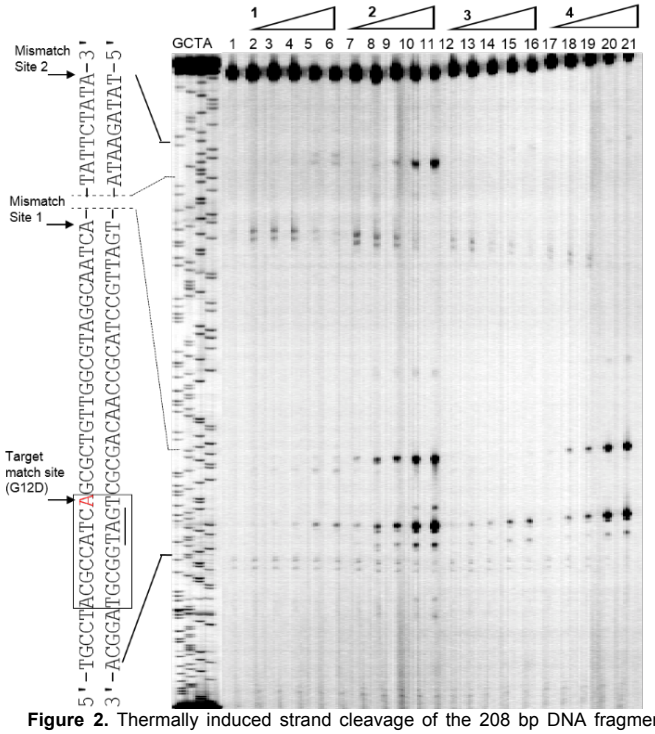



**Figure 2.** Thermally induced strand cleavage of the 208 bp DNA fragments containing the G12D mutation sequence were labeled with 5′-Texas Red (10 nM). Polyamides **1**–**4** were incubated for 18 h at 25°C at the following Py–Im polyamide concentrations: lane 1=DNA control; lanes 2–6=10, 50, 100, 500, and 1000 nM of polyamide **1**; lanes 7–11=10, 50, 100, 500, and 1000 nM of polyamide **2**; lanes 12–16=10, 50, 100, 500, and 1000 nM of polyamide **3**; and lanes 17–21= 10, 50, 100, 500, and 1000 nM of polyamide **4.**

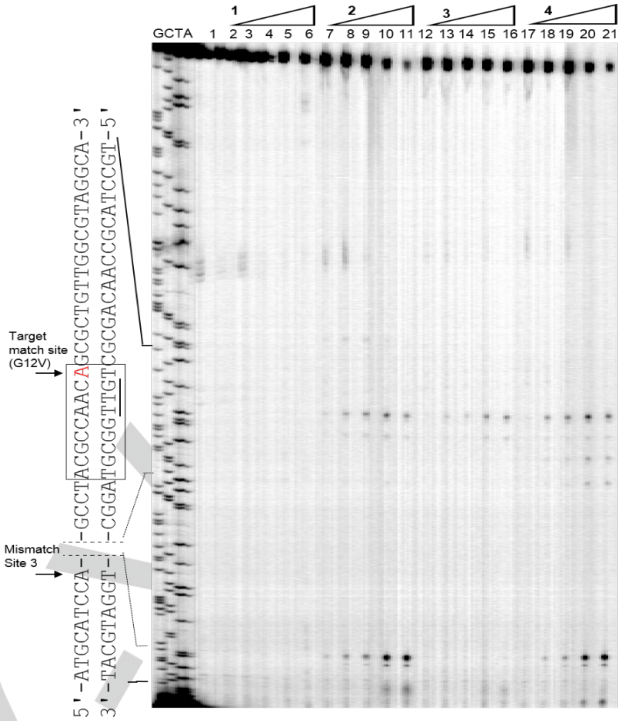



**Figure 3.** Thermally induced strand cleavage of 208 bp DNA fragments containing the G12V mutation sequence were labeled with 5′-Texas Red (10 nM). Polyamides **1**–**4** were incubated for 18 h at 25°C at the following Py–Im polyamide concentrations: lane 1=DNA control; lanes 2–6=10, 50, 100, 500, and 1000 nM of polyamide **1**; lanes 7–11=10, 50, 100, 500, and 1000 nM of polyamide **2**; lanes 12–16=10, 50, 100, 500, and 1000 nM of polyamide **3**; and lanes 17–21= 10, 50, 100, 500, and 1000 nM of polyamide **4**.

placement of $\beta$ at position M (Figure 1). The amino terminus contains five continuous heterocyclic Py–Im; this conformation would be over curved compared with the minor groove and would therefore inhibit the Py–Im polyamide from binding effectively and thus inhibit alkylation. This confirms previous reports suggesting that $\beta$ should be included after a maximum of four consecutive Py–Im moieties.[21–24]

Conjugate **2** alkylated the G12D mutation site with high reactivity (Figure 2). Alkylation was visible from 50 nM and potent alkylation at 500 nM and 1 $\mu$M. However lower reactivity was observed at the G12V site (Figure 3). Mismatch alkylation was also observed at three separate sites: mismatch sites 1, 2, and 3. Conjugate **2** displayed high reactivity toward mismatch site 1 located on the T7 strand, with alkylation comparable to the G12D match site. Alkylation was also observed at 500 nM and 1 $\mu$M at mismatch site 2, which is also located on the T7 strand (Figure 2). Visible alkylation was also evident at mismatch site 3 on the SP6 strand at 500 nM and 1 $\mu$M (Figure 3).

Conjugate **3** displayed lower reactivity toward both the G12D and G12V target sites and only minor alkylation at 500 nM and 1 $\mu$M. The lower reactivity of conjugate **3** can probably be attributed to $\beta$ at position Z (Figure 1), which is next to the indole linker. The consecutive degenerate binding of $\beta$ and the indole linker probably inhibits the efficient binding of the alkylating moiety to the minor groove.

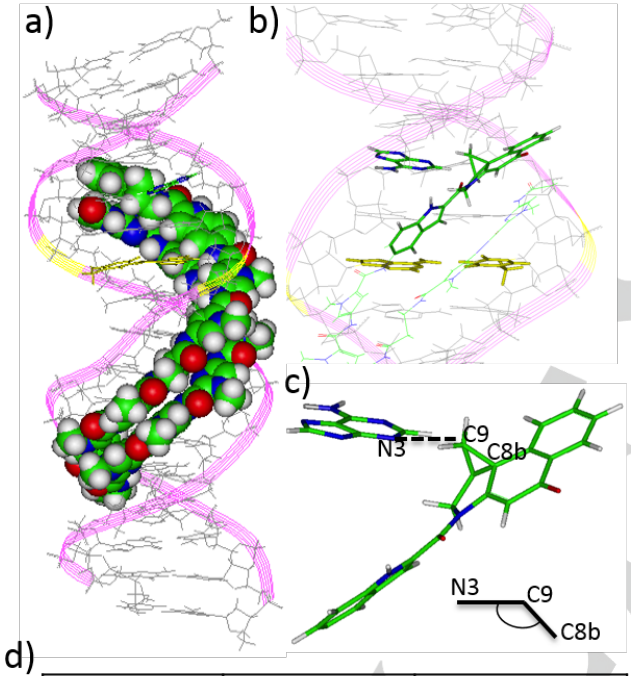
Conjugate **4** targeted the G12D with high reactivity, and alkylation was observed from 50 nM and potent alkylation at 500 nM and 1 $\mu$M. Lower reactivity was again observed for the G12V site over the same concentration range. Mismatch alkylation was also observed at two separate sites. At mismatch site 1, minor alkylation was seen at 50 nM and 100 nM, and higher reactivity was seen at 500 nM and 1 $\mu$M. Mismatch alkylation at mismatch site 3 was also observed at 500 nM and 1 $\mu$M but, importantly, no mismatch alkylation was observed at mismatch site 2.

Interestingly, conjugates **2** and **4** had much higher reactivity for the G12D mutation sequence, which is the most frequently occurring KRAS mutation.[51] Conversely, the reactivity was lower for the G12V mutation sequence. The reason for this is not clear from the Py–Im polyamide binding rules,[18–20] and further investigation is warranted.

Based on the results of PAGE, substitution of Py with $\beta$ at position M was unsuccessful because the aromatic chain was too long; therefore, substitution at position N is preferred. Additionally, substitution at position X gives higher reactivity and selectivity compared with positions Y and Z. The two consecutive Py moieties probably stabilize the indole-*seco*-CBI in the minor groove to allow for efficient alkylation. Therefore, the binding specificity of conjugate **4** was investigated further. These data support our recently published data demonstrating the ability of conjugate **4** to suppress tumor growth selectively in mice with tumors harboring the KRAS codon 12 mutation.[38]

**Molecular Modeling Studies of Conjugate 4.**

To gain further insight into the alkylating reactivity of conjugate **4**, we performed molecular modeling studies of polyamide **4** [5′-dGCCTACGCCAACAGCTC-3′/5′-dGAGCTGTTGGCGTAGGC-3′ (G12D match sequence)] and **4** [5′-dGCCTACGCCACCAGCTC-3′/5′-dGAGCTGGTGGCGTAGGC-3′ (1 bp mismatch sequence)] (Figure 4). The energy-minimized structure for the G12D sequence indicated that conjugate **4** could bind tightly to the minor groove of the DNA helix. The model displayed a distance of 3.03 Å between the C9 of the cyclopropane subunit of CBI and the nucleophilic N3 of adenine. This is within an appropriate distance to allow efficient alkylation to occur. The distance between the same units for the 1 bp mismatch sequence was 3.12Å, which is consistent with the evidence of efficient sequence-specific alkylation of the G12D sequence.




| Conjugate **4** | Distance (N3-C9) | Angle (N3-C9-C8b) |
|---|---|---|
| Match site | 3.03 Å | 150° |
| 1 bp mismatch | 3.12 Å | 149° |

**Figure 4.** a) Energy-minimized structure of the complex between **5** [5′-dGCCTACGCCAACAGCTC-3′/5′-dGAGCTGTTGGCGTAGGC-3′] and the KRAS G12D sequence. The G12D mutation is highlighted in yellow. The water layer, H atoms, and $Na^+$ cations are omitted for clarity. b) Magnified view of the alkylation site. The indole-CBI and adenine moieties are highlighted. c) The distance between C9 of the cyclopropane subunit of CBI and N3 of adenine, and the angle between the C8b and C9 of the cyclopropane and the N3 of adenine. d) Table comparing distance and angle of the cyclopropane unit and adenine between the match sequence and the 1 bp mismatch sequence.

The angle of the N3-C9-C8b for the G12D sequence was 150°, thus the cyclopropane ring was positioned in an efficient orientation for alkylation to occur.[36] The G12D-minimized structure shows that the $\beta$–$\beta$ pairing at positions X and N of **4** works well to allow the Py–Im polyamide to curve efficiently around the DNA helix, while remaining tight inside the minor groove. The flexibility of $\beta$ at position N allows the terminal Py–Im to hold the indole-*seco*-CBI close to the nucleophilic N3 of the adenine. The distance and angle of the alkylating moiety of conjugate **4** are consistent with the evidence showing good sequence-specific DNA alkylating reactivity. These results also show that careful placement of the $\beta$ pairing should be considered when designing alkylating Py–Im polyamides to recognize longer DNA sequences to allow for efficient binding to the DNA helix.

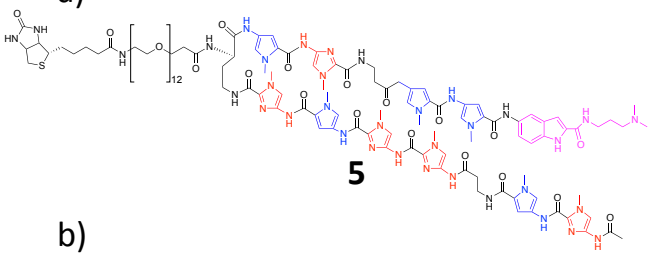
**High-Throughput Sequencing of Conjugate 5.**

Bind-n-Seq enables high-throughput sequencing of Py–Im polyamide binding sites[52-54] and allowed us to investigate the binding sequences of conjugate **4** in much greater detail than with PAGE. To investigate the binding specificity of Py–Im polyamide **4**, conjugate **5** was synthesized. Conjugate **5** contains the same Py–Im polyamide core as conjugate **4** with the addition of $PEG_{12}$-biotin on the $\gamma$ turn to facilitate the affinity pull down; and the alkylating CBI moiety was substituted with dimethylaminopropylamine (Dp) to ensure that alkylation did not interfere with the binding. This reduced the recognition sequence by **1** to 8 bp (Figure 5b). An 8 bp DNA sequence contains 32,896 combinations of possible bp sequences.[55] Following incubation of conjugate **5** with 10 bp randomized DNA sequences with Ion PGM-compatible adapter sequences, pull down with streptavidin beads and affinity purification retained the Py–Im polyamide-bound DNA sequences.[56] Following amplification by PCR of the retained DNA sequences, high-throughput sequencing was performed using the Ion PGM sequencer and was confirmed using the DREME primary motif analysis algorithm.[57]

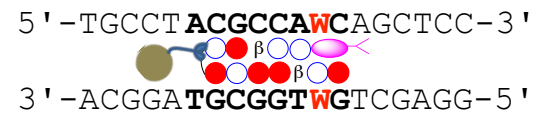
According to the bind-n-seq analysis parameter used in previous reports, [52-54, 56] the highest enriched sequence is used for the binding site motif construction (Figure 5c). Since our top three sequences share very close sequence similarity by default the Bind-n-seq motif construction program constructed the preliminary motif (Figure 5c) from the top three sequences.

In the Bind-n-Seq analysis shown in Figure 5d, the top-three binding sequences represent match sequences for conjugate **5** with a high fold enrichment. Figure 5b shows the hit-binding sequence of conjugate **5**, 5′-ACGCCATC-3′, which corresponds to the G12D mutation sequence that bound with very high specificity, with a significant 89-fold enrichment. Third was the G12V sequence, which had a considerable 69-fold enrichment. Importantly, the second sequence from the Bind-n-Seq analysis 5′ACGCCTTC-3′, with 76-fold enrichment, is also a match sequence for conjugate **5** according to the established binding rules.[18–20] The fold enrichment declined drastically for the mismatch binding sites, suggesting unfavorable binding; for example, by the ninth place, the fold enrichment was less than half that of the G12D sequence. Importantly, the corresponding WT KRAS sequence 5′-ACGCCACC-3′ was located at position 354, with a nonsignificant fourfold enrichment. These results clearly show that the Py–Im polyamide motifs of 4 and 5 targeted the KRAS codon 12 mutation sequences, especially the G12D mutation sequence, with high affinity.

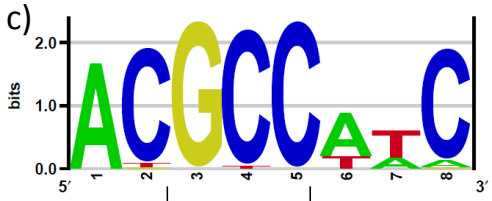
a)




b)




c)




d)

| Rank | Sequence | Fold-enrichment |
|---|---|---|
| 1 | ACGCCATC | 89.097 |
| 2 | ACGCCTTC | 76.887 |
| 3 | ACGCCAAC | 69.049 |
| 4 | ATGAAGGC | 64.357 |
| 5 | ATGATGGC | 59.01 |
| 6 | AACGCCAT | 53.967 |
| 7 | CGCCATCA | 53.508 |
| 8 | GCCTTCAA | 44.989 |
| 9 | CGCCAACA | 44.346 |
| 10 | AACGCCAA | 43.933 |
| **354** | **ACGCCACC** | **4.104** |

**Figure 5.** a) Structure of Py–Im polyamide **5**. **5** contained the same core moiety as **4**, but the CBI unit was replaced with Dp and $PEG_{20}$-biotin was conjugated to the $\gamma$ turn. b) Schematic representation of the target KRAS-binding sequence. c) Primary binding motif identified using Bind-n-Seq and next-generation sequencing. This sequence corresponded to the G12D mutation sequence. d) Table of the top 10 binding sequences for **5** identified by Bind-n-Seq. Note the G12V sequence ranked third and the KRAS WT sequence ranked at 354.

The mismatch site 1 observed on the gel (Figure 2), AGGCAATCA was also partially observed in the Bind-n-Seq experiments as a 7 base site AGGCAAT with 4.2 fold enrichment. Also mismatch site 3, ATGCATCCA was partially observed for the reverse 7 base sites, TACGAAG and TACGATG with 2.6 and 2.4 fold enrichment respectively. These low enrichments of the mismatch sites could be attributed to a broad context of sequences used in Bind-n-Seq. In other words, alkylation of the mismatch sites observed in PAGE was overestimated due to a restricted availability of sequences, and the results from Bind-n-Seq reflected more actual binding characteristics of Py-Im polyamide.

Even without the *seco*-CBI moiety, the G12D sequence was the preferred target for the Py–Im polyamide followed by the G12V mutation sequence, which is in agreement with the PAGE results. Intriguingly, mismatch sites 1 (5′-AGGCAATC-3′) or 2 (5′-TATTCTAT-3′) could not be detected in the Bind-n-Seq analysis, suggesting that the CBI moiety might play a role in the mismatch alkylation observed in PAGE.

## Conclusions

Mutant KRAS is a therapeutically important target, and there are at present no drugs that successfully target mutant KRAS. Conventional drugs aim at inhibiting the KRAS protein or downstream targets. We decided to take a different approach by targeting the KRAS codon 12 mutation directly. We evaluated the sequence-specific alkylating activity of four Py–Im polyamide conjugates in targeting the KRAS codon 12 G12D and G12V mutation sequences. Two of the alkylating Py–Im polyamides displayed high reactivity for the target site; conjugate **4** displayed the highest selectivity and was selected for further evaluation. The computer-minimized structure of the G12D mutation sequence suggested that the structural motif of **4** was appropriate for efficient sequence-specific alkylation of the target sequence. High-throughput sequencing (Bind-n-Seq) revealed the binding specificity of conjugate **5** against 10 bp randomized sequences. Conjugate **5** displayed significant affinity for the KRAS codon 12 mutation sequences with 89- and 69-fold enrichment for the G12D and G12V sequences, respectively, but only fourfold enrichment for the WT sequence. The 9 base pair recognition of Py-Im polyamide **4** allows the recognition of approximately 9,121 sites in the human genome (GRCh38/hg38) containing the KRAS match sequence for preferential alkylation.[38] Our results strongly suggest that *seco*-CBI Py–Im polyamides are potential agents for the treatment of cancer based on individual genetic mutations. This approach may help minimize side effects and overcome the problem of drug resistance. Specifically targeting the KRAS codon 12 mutation sequence may be a viable therapeutic route for targeting tumors harboring mutant KRAS.

## Experimental Section

**General Methods.** Reagents and solvents were purchased from standard suppliers and used without further purification. High-performance liquid chromatography (HPLC) analysis was performed with a JASCO PU-2080 Plus HPLC pump and using a JASCO 807-IT HPLC UV/Vis detector and a Chemcobond 5-ODS-H reversed-phase column (4.6 x150 mm) in 0.1% TFA in water with acetonitrile as eluent at a flow rate of 1.0 mL $min^{-1}$ with detection at 254 nm. Electrospray ionization time-of-flight mass spectrometry (ESI-TOF- MS) was produced on a Bio-TOF II (Bruker Daltonics) mass spectrometer using a positive ionization mode. Machine-assisted polyamide syntheses were performed on a

PSSM-8 (Shimadzu) with computer-assisted operation system at a 20 μmol scale by using Fmoc chemistry. The oligonucleotides, 5'-Texas Red labelled primers, and cold primers were purchased from Sigma–Aldrich. Ex Taq DNA polymerase was purchased from Takara. Thermo Sequence core sequencing kit and loading dye (formamide with fuchsin red) were purchased from GE Healthcare. Polymerase Chain Reaction (PCR) was performed on an iCycler (BIO-RAD). Long Ranger gel solution (50%) was purchased from FMC bioproducts. Polyacrylamide gel electrophoresis was performed on a HITACHI 5500-S DNA sequencer, and data were analysed by FLAGLYS version 2 software (HITACHI). $NH_2$ –indole-*seco*-CBI was prepared by previous methods.[33–36] Oxime resin (200– 400 mesh) was purchased from Peptides International and Novabiochem. Randomized oligonucleotides were modified using the Ion torrent PGM sequencer (Life Technologies).

**Solid-Phase Synthesis of Py-Im Polyamides.** Py-Im polyamides supported by oxime resin (for the synthesis of the conjugates **1**, **2**, **4** and **5**) or *β*-Wang resin (for conjugate **3**) were prepared in a stepwise reaction by reported fluorenylmethyloxycarbonyl (Fmoc) solid-phase synthesis using Fmoc-Py-$CO_2H$, Fmoc-Im-$CO_2H$, Fmoc-PyIm-$CO_2H$, Fmoc-*β*-$CO_2H$, Fmoc-GABA-OH (for conjugates **1**-**4**), Boc-Dab (Fmoc) –OH (for conjugate **5**), O- (1H-6-chlorobenzotriazole-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate (HCTU), *N, N*-diisopropylethylamine (DIEA), 1-methyl-2-pyrrolidone (NMP).[18] Cleavage from the oxime resin was carried out in basic conditions using 2M NaOH and 1, 4-dioxane (1:1) at 55°C for 3 h, and *β*-Wang resin using TFA, water and triisopropylsilane (95:2.5:2.5) at room temperature for 30 min. For **1**, **2**, **4** and **5**, the resulting carboxylic acids were acidified with $CH_3COOH$ to pH 6, precipitated with $Et_2O$, washed with milliQ water and lyophilized. The target carboxylic acid **3** was collected after precipitation with $Et_2O$. The corresponding crudes were used without further purification in the coupling reaction with $NH_2$-indole-*seco*-CBI.

**Synthesis of Conjugate 1 by the Coupling with $NH_2$-Indole *seco*-CBI.** A solution of Py-Im polyamide carboxylic acid **1** (1.0 mg, 0.66 μmol), DMF (100 μL), DIEA (0.5 μL, 2.86 μmol), and benzotriazol-1-yl-oxytripyrrolidinophosphonium hexafluorophosphate (PyBOP, 1.0 mg, 1.95 μmol) were stirred for 30 min at RT. The formation of the activated 1-hydroxybenzotriazole ester was checked by analytical HPLC (tR = 13.8 min, 0.1% TFA/MeCN, linear gradient 0–100%, 0–20 min); then, $NH_2$ -indole-*seco*-CBI (0.6 mg, 1.58 μmol) was added and the mixture was stirred overnight at room temperature under a $N_2$ atmosphere. $Et_2O$ was added and the supernatant was removed. The crude was washed with $Et_2O$ and $CH_2Cl_2$, followed by HPLC purification (0.1% TFA/ MeCN, linear gradient 20–65%, 0–30 min). After collecting the main peak fraction and lyophilization, conjugate **1** was obtained as a yellow solid (1.0 mg, 0.54 μmol, 80% yield). MS (ESI-TOF): *m/z* calcd for $C_{89}H_{94}ClN_{31}O_{16}^{2+}$ $[M+2H]^{2+}$ 943.8586; found 943.8555.

Following the same synthetic procedure conjugates **2-4** were synthesized:

Conjugate **2** was obtained as a yellow powder. MS (ESI-TOF): *m/z* calcd for $C_{89}H_{94}ClN_{31}O_{16}^{2+}$ $[M+2H]^{2+}$ 943.8586; found 943.8482; HPLC: tR = 13.8 min (0.1% TFA/MeCN, linear gradient 0–100%, 0–20 min).

Conjugate **3** was obtained as a yellow powder. MS (ESI-TOF): *m/z* calcd for $C_{89}H_{94}ClN_{31}O_{16}^{2+}$ $[M+2H]^{2+}$ 943.8586; found 943.8552; HPLC: tR = 13.7 min (0.1% TFA/MeCN, linear gradient 0–100%, 0–20 min).

Conjugate **4** was obtained as a yellow powder. MS (ESI-TOF): *m/z* calcd for $C_{89}H_{94}ClN_{31}O_{16}^{2+}$, $[M+2H]^{2+}$ 943.8586; found 943.8572; HPLC: tR = 13.9 min (0.1% TFA/MeCN, linear gradient 0–100%, 0–20 min)

**Synthesis of Conjugate 5**

$NO_2$-indole-3-dimethylaminopropylamine (DMAPA) (6 mg, 20 μmol) was reduced under a $H_2$ atmosphere with Pd/C (5 mg, 46 μmol) in DMF (400 μL) for 2 h. The Pd/C was removed by filtration and the remaining solution was immediately used in the next step. Simultaneously, DIEA (0.5 μL, 2.86 μmol) and PyBOP (1.0 mg, 1.95 μmol) were added to a solution of Py-Im polyamide carboxylic acid **5** (1.0 mg, 0.38 μmol) in DMF (50 μL). The reaction mixture was stirred for 30 min at room temperature. After the conversion from the carboxylic acid into the activated 1-hydroxybenzotriazole ester, the $NH_2$-indole-DMAPA solution was added and stirred overnight at room temperature under $N_2$ atmosphere. $Et_2O$ was added and the supernatant was removed. The crude was washed with $Et_2O$ and $CH_2Cl_2$ and dried. Boc deprotection was carried out in TFA (300 μL) and DCM (300 μL) for 1 h at room temperature. The solvent was removed, and the remaining solid was filtered, washed with milliQ (500 μL) and subjected to HPLC purification (0.1% TFA/MeCN 30–75% linear gradient, 0–30 min). The peak fraction containing the product was collected, and lyophilized. The purified compound was dissolved in DMF (100 μL) and DIEA (0.5 μL, 2.86 μmol) and NHS-$PEG_{12}$-Biotin was added and stirred for 1 h at room temperature. $Et_2O$ was added and the supernatant was removed. Conjugate **5** was washed with $CH_2Cl_2$ and $ET_2O$ to give a light brown powder (0.5 mg, 0.19 μmol, 50%). MS (ESI-TOF): *m/z* calcd for $C_{118}H_{164}N_{36}O_{30}S^{2+}$ $[M+2H]^{2+}$ 1298.6062; found: 1298.5363; HPLC: tR = 10.9 min (0.1% TFA/ MeCN, linear gradient 0–100%, 0–20 min).

**Cloning of Plasmid DNA**

Oligonucleotides were purchased from Sigma–Aldrich and annealed at a concentration of 10 μM, followed by ligation into the pGEM-T Easy vector (Promega). *Escherichia coli* DH5α competent cells (TOYOBO) were transformed and cultured overnight at 37 °C on an LB plate with 32 μg of X-gal (20 mg $mL^{-1}$), 25 μL/IPTG 100 mM, and 100 μg $mL^{-1}$ ampicillin. White colonies were identified by colony PCR in 20 μL of the reaction mixtures containing 250 nM of each primer (SP6 primer, 5'-TATTTAGGTGACACTATAG-3'; T7 primer, 5'-TAATACGACTCACTATAGGG-3'), 200 μM dNTPs (Sigma Aldrich), two units of Taq polymerase, and 1xThermoPol reaction buffer (New England Bio Labs). Amplification of the DNA fragments was carried out incubating at 95 °C for 5 min, followed by 35 cycles of 95 °C for 35 s, 50 °C for 35 s, 72 °C for 30 s, with a final extension step of 72 °C for 7 min. The appropriate colony was selected and transferred to 5 mL of LB medium with 100 mg $mL^{-1}$ ampicillin and cultured overnight at 37 °C. The plasmids were extracted using a GenElute Plasmid Miniprep Kit (Sigma Aldrich) and identified by PCR (program and reaction mixtures were the same as above). The 5'-Texas Red-modified 208 bp DNA fragment was prepared by PCR with 5'-Texas Red-modified SP6 and T7 primers from 1 ng $L^{-1}$ of the 31 bp fragments 5'-GCCTACGCCAACAGCGCTGATGGCGTAGGCA-3'/5'-GCCTACGCCATCAGCGCTGTTGGCGTAGGCA-3' inserted into the pGEM-T Easy vector (program and reaction mixtures were the same as above). Fragments were purified by GenElute PCR Clean-up Kit (Sigma Aldrich), and their concentrations were determined by UV absorption.

**High-Resolution Gel Electrophoresis**

The 5'-Texas Red labelled 208 bp DNA fragment was prepared by PCR using the plasmid with inserts described above and the 5'-Texas Red labelled SP6 and T7 primers. The sequences on SP6 side was 5'-TATTTAGGTGACACTATAGAATACTCAAGCTATGCATCCAACGCGTT
GGGAGCTCTCCCATATGGTCGACCTGCAGGCGGCCGCGAATTCAC
TAGTGATTGCCTACGCCAACAGCGCTGATGGCGTAGGCAATCGAAT
TCCCGCGGCCGCCATGGCGGCCGGGAGCATGCGACGTCGGGCCC

AATTCGCCCTATAGTGAGTCGTATTA-3'. Obtained 5'-Texas Red labelled DNA fragments were purified using PCR-Clean Up Kit (Sigma) and their concentrations were determined by UV absorption. A reaction mixture (10 μL) containing each of alkylating conjugates, a 5'-Texas Red labelled DNA fragment (10 nM, duplex concentration), and 50 mM sodium phosphate buffer (pH 7.0) was incubated for 18 h at 25 °C. The samples were quenched by the addition of 1 mM calf thymus DNA (1 μL) and then heated at 95°C for 5 min to cleave the DNA strands at the specific alkylation sites. After the removal of solvents under reduced pressure, loading dye (7 μL) was added. The samples were heated at 95 °C for 25 min and immediately cooled on ice. Samples (1.2 μL) were subjected to electrophoresis on 6% denaturing polyacrylamide gel using Hitachi 5500-S DNA Sequencer.

**Molecular Modelling Studies on the DNA Complex of Conjugate 4**

Minimizations were performed with the Discover (MSI, San Diego, CA) program using CVFF force-field parameters. The starting structure was constructed using Insight II program and the builder module of the program using standard bond lengths and angles. Where the three upper and lower sides of Watson–Crick base pairs were fixed, conjugate **4** was inserted into the 5'- dGCCTACGCCAACAGCTC-3'/5'-dGAGCTGTTGGCGTAGGC-3' duplex (match sequence) and 5'-dGCCTACGCCACCAGCTC-3'/5'-dGAGCTGGTGGCGTAGGC-3' duplex (1 bp mismatch sequence). Thirty-four $Na^+$ cations were placed at the bifurcating position of the O-P-O angle at a distance of 2.21 Å from the phosphorus atom. The resulting complex was soaked in a 15 Å layer of water. The layer of water was minimized without constraints to the stage where the rms was less than 0.001 kcal $mol^{-1}$Å using the steepest and then the conjugate algorithm; successively the whole complex was also minimized in the same way.

**Bind-n-Seq:**

Bind-n-Seq[52-54] and subsequent analysis to evaluate the binding affinity of conjugate **5** towards 10 bp randomized sequences were modified the Ion torrent PGM sequencer (Life Technologies). Randomized oligonucleotides with high-throughput sequencing platform-specific adapters were synthesized according to the literature procedures.[56] All barcoded Bind-n-Seq 92 mers were synthesized by Sigma Aldrich machine mixing, standard desalting purification, 5'-CCATCTCATCCCTGCGTGTCTCCGACTCAGXXXXXXXXXXNNNNNNNNNNAATCACCGACTGCCCATAGAGAGGAAAGCGGAGGCGTAGTGG-3`. The barcode represented by XXXXXXXXXX, are 10-letter barcodes used as per Ion torrent sequencing technologies, and N represents the randomized sequence. Adapter ligated oligonucleotides (3 μM) were duplexed by primer extension with the adapter-specific primer 5`-CCACTACGCCTCCGCTTTCCTCTCTA-3` (9 μM) in 25 μL reactions containing GoTaq Green (Promega) PCR master mix (2x) with $Mg^{2+}$ (2 mM). The primer extension reactions were performed at 95 °C (3 min), 60 °C (2 min), 70 °C (5 min), and then 4 °C in a thermocycler (Bio- Rad). The binding reaction was performed using duplex DNA and conjugate **5**-bound DNA that was enriched followed by Next Generation Sequencing. To count the number of conjugate **5** enriched unique DNA sequences, MERMADE with the k-mer sliding window (k = 8 bp) and a new pipeline for Bind- n-Seq analysis (http://korflab.ucdavis.edu/Datasets/BindNSeq) were used. Highly enriched motifs were confirmed with DREME primary motif analysis.[57]

## Acknowledgements

This work was supported by JSPS KAKENHI (Grant Number 24225005), "Basic Science and Platform Technology Program for Innovative Biological Medicine" and "JSPS-NSF International Collaborations in Chemistry (ICC)" to HS, and by JSPS KAKENHI (Grant Number 24310155) to TB.

**Keywords:** KRAS • DNA alkylation • pyrrole-imidazole polyamides • PAGE • Bind-n-Seq

[1] W. Pao, T. Y. Wang, G. J. Riely, V. A. Miller, Q. Pan, M. Ladanyi, M.F. Zakowski, R. T. Heelan, M. G. Kris, H. E. Varmus, *PLoS Med.* **2005**, *2*, e17.
[2] S. D. Richman, M. T. Seymour, P. Chambers, F. Elliott, C. L. Daly, A. M. Meade, G. Taylor, J. H. Barrett, P. Quirke, *J. Clin. Oncol.* **2009**, *27*, 5931–5937.
[3] A. A. Adjei, *JNCI J. Natl. Cancer Inst.* **2001**, *93*, 1062–1074.
[4] A. D. Cox, C. J. Der, *Small GTPases* **2010**, *1*, 2–27.
[5] S. Eser, A. Schnieke, G. Schneider, D. Saur, *Br. J. Cancer* **2014**, *111*, 817–822.
[6] A. D. Cox, S. W. Fesik, A. C. Kimmelman, J. Luo, C. J. Der, *Nat. Rev. Drug Discov.* **2014**, *13*, 828–851.
[7] S. R. Rajski, R. M. Williams, *Chem. Rev*. **1998**, *98*, 2723–2796.
[8] G. Zimmermann, B. Papke, S. Ismail, N. Vartak, A. Chandra, M. Hoffmann, S. A. Hahn, G. Triola, A. Wittinghofer, P. I. H. Bastiaens, H. Waldmann, H. *Nature* **2013**, *497*, 638–42.
[9] J. M. Ostrem, U. Peters, M. L. Sos, J. A. Wells, K. M. Shokat, *Nature* **2013**, *503*, 548–51.
[10] Y. Wang, C. E. Kaiser, B. Frett, H.-Y. Li, *J. Med. Chem*. **2013**, *56*, 5219−5230.
[11] N. M. Baker, C. J. Der, *Nature* **2013**, *497*, 577–578.
[12] T. Deramaudt, A. K. Rustgi, *Biochim. Biophys. Acta* **2005**, *1756*, 97-101.
[13] O. Kranenburg, *Biochim. Biophys. Acta* **2005**, *1756*, 81–2.
[14] V. T. Smit, A. J. Boot, A. M. Smits, G. J. Fleuren, C. J. Cornelisse, J. L. Bos, *Nucleic Acids Res*. **1988**, *16*, 7773–7782.
[15] J. L. Bos, *Cancer Res*. **1989**, *49*, 4682–4689.
[16] Y. Pylayeva-Gupta, E. Grabocka, D. Bar-Sagi, *Nat. Rev. Cancer* **2011**, *11*, 761–774.
[17] B. B. Friday, A. A. Adjei, *Biochim. Biophys. Acta* **2005**, *1756*, 127–144.
[18] P. B. Dervan, *Bioorg. Med. Chem*. **2001**, *9*, 2215–2235.
[19] P. B. Dervan, B. S. Edelson, *Curr. Opin. Struct. Biol.* **2003**, *13*, 284–299.
[20] P. B. Dervan, R. M. Doss, M. A. Marques, *Curr. Med. Chem. Anti-Cancer Agents* **2005**, *5*, 373–387.
[21] M. Parks, E. Baird, P. B. Dervan, *J. Am. Chem. Soc.* **1996**, *7863*, 6147–6152.
[22] P. B. Dervan, A. R. Urbach, *Essays in Contemporary Chemistry, Verlag Helvetica Chimica Acta*, Zürich, **2001**.
[23] C. C. C. Wang, U. Ellervik, P. B. Dervan, *Bioorg. Med. Chem*. **2001**, *9*, 653–657.
[24] C. R. Woods, T. Ishii, B. Wu, K. W. Bair, D. L. Boger, *J. Am. Chem. Soc*. **2002**, *124*, 2148–2152.
[25] T. Bando, M. Minoshima, G. Kashiwazaki, K. Shinohara, S. Sasaki, J. Fujimoto, A. Ohtsuki, M. Murakami, S. Nakazono, H. Sugiyama, *Bioorg. Med. Chem*. **2008**, *16*, 2286–2291.
[26] V. C. Rucker, C. Melander, P. B. Dervan, *Helv. Chim. Acta* **2003**, *86*, 1839–1851.
[27] S. Wang, R. Nanjunda, K. Aston, J. K. Bashkin, W. D. Wilson, *Biochemistry* **2012**, *51*, 9796–9806.
[28] J. K. Bashkin, K. Aston, J. P. Ramos, K. J. Koeller, R. Nanjunda, G. He, C. M. Dupureur, W. D. Wilson, *Biochimie* **2013**, *95*, 271–279.
[29] J. Syed, G. N. Pandian, S. Sato, J. Taniguchi, A. Chandran, K. Hashiya, T. Bando, H. Sugiyama, *Chem. Biol*. **2014**, *21*, 1370-1380.

[30] T. Oyoshi, W. Kawakami, A. Narita, T. Bando, H. Sugiyama, *J. Am. Chem. Soc.* **2003**, *125*, 4752–4754.

[31] K. Shinohara, S. Sasaki, M. Minoshima, T. Bando, H. Sugiyama, *Nucleic Acids Res.* **2006**, *34*, 1189–1195.

[32] K. Shinohara, A. Narita, T. Oyoshi, T. Bando, H. Teraoka, H. Sugiyama, *J. Am. Chem. Soc*. **2004**, *126*, 5113–5118.

[33] D. L. Boger, T. Ishizaki, P. A. Kitos, O. Suntornwat, *J. Org. Chem*. **1990**, *55*, 5823–5832.

[34] D. L. Boger, W. Yun, B. R. Teegarden, *J. Org. Chem*. **1992**, *57*, 2873–2876.

[35] D. L. Boger, J. A. McKie, *J. Org. Chem.* **1995**, *60*, 1271–1275.

[36] T. Bando, S. Sasaki, M. Minoshima, C. Dohno, K. Shinohara, A. Narita, H. Sugiyama, *Bioconjug. Chem*. **2006**, *17*, 715–720.

[37] 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.

[38] K. Hiraoka, T. Inoue, R. D. Taylor, T. Watanabe, N. Koshikawa, K. Shinahara, H. Yoda, K. Sugimoto, Y. Maru, T. Denda, K. Fujiwara, A. Balmain, T. Ozaki, T. Bando, H. Sugiyama, H. Nagase, *Nat. Commun.* **2015**, *6*, 6706.

[39] P. B. McKinzie, R. R. Delongchamp, T. Chen, B. L. Parsons, *Mutagenesis* **2006**, *21*, 391–397.

[40] S. White, E. E. Baird, P. B. Dervan, *J. Am. Chem. Soc*. **1997**, *119*, 8756–8765.

[41] Y.-W. Han, G. Kashiwazaki, H. Morinaga, T. Matsumoto, K. Hashiya, T. Bando, Y. Harada, H. Sugiyama, *Bioorg. Med. Chem*. **2013**, *21*, 5436–5441.

[42] S. Wang, K. Aston, K. J. Koeller, G. D. Harris, N. P. Rath, J. K. Bashkin, W. D. Wilson, *Org. Biomol. Chem*. **2014**, *12*, 7523–7536.

[43] E. E. Baird, P. B. Dervan, *J. Am. Chem. Soc*. **1996**, *118*, 6141–6146.

[44] N. R. Wurtz, J. M. Turner, E. E. Baird, P. B. Dervan, *Org. Lett*. **2001**, *3*, 1201–1203.

[45] J. M. Belitsky, D. H. Nguyen, N. R. Wurtz, P. B. Dervan, *Bioorg. Med. Chem*. **2002**, *10*, 2767–2774.

[46] T. Bando, A. Narita, I. Saito, H. Sugiyama, *Chem. Eur. J.* **2002**, *8*, 4781–4790.

[47] F. Sanger, A. R. Coulson, *J. Mol. Biol*. **1975**, *94*, 441–448.

[48] F. Sanger, S. Nicklen, A. R. Coulson, *Proc. Natl. Acad. Sci.* **1977**, *74*, 5463–5467.

[49] L. M. Smith, S. Fung, M. W. Hunkapiller, T. J. Hunkapiller, L. E. Hood, *Nucleic Acids Res.* **1985**, *13*, 2399–2412.

[50] L. M. Smith, J. Z. Sanders, R. J. Kaiser, P. Hughes, C. Dodd, C. R. Connell, C. Heiner, S. B. Kent, L. E. Hood, *Nature* **1986**, *321*, 674–679.

[51] J. Neumann, E. Zeindl-Eberhart, T. Kirchner, A. Jung, *Pathol. Res. Pract.* **2009**, *205*, 858–862.

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

[53] J. L. Meier, A. S. Yu, I. Korf, D. J. Segal, P. B. Dervan, *J. Am. Chem. Soc*. **2012**, *134*, 17814–17822.

[54] J. J. S. Kang, J. J. L. Meier, P. B. Dervan, *J. Am. Chem. Soc.* **2014**, *136*, 3687–3694.

[55] P. B. Dervan, *Science* **1986**, *232*, 464–471.

[56] A. Chandran, Y. Li, S. Kizaki, G. N. Pandian, K. Hashiya, T. Bando, H. Sugiyama, ChemBioChem **2014**, *8501*, 1–6.

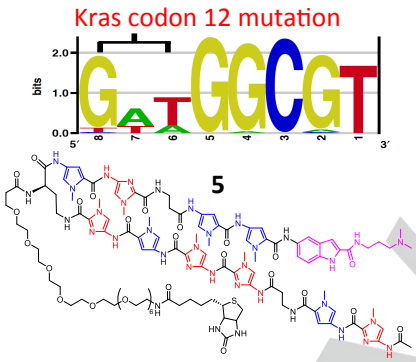
[57] T. L. Bailey, *Bioinformatics* **2011**, *27*, 1653–1659.

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

Layout 1:

# FULL PAPER

*N-methylpyrrole-N-methylimidazole (Py-Im) polyamide seco-*CBI Conjugates *target the Kras codon 12 mutation with high affinity and specificity. Following optimization of the β-alanine pairings of four Py-Im polyamide conjugates for sequence specific alkylation, conjugate 5 was synthesized for comprehensive evaluation of Py-Im polyamides sequence specificity for the Kras codon 12 mutation sequence using Bind-n-Seq.*




*Rhys D. Taylor, Anandhakumar Chandran, Gengo Kashiwazaki, Kaori Hashiya, Toshikazu Bando*, Hiroki Nagase, Hiroshi Sugiyama**

***Page No. – Page No.***

**Selective Targeting of the KRAS Codon 12 Mutation Sequence by Pyrrole–Imidazole Polyamide seco-CBI Conjugates**

Supporting Information

# Selective Targeting of the KRAS Codon 12 Mutation Sequence by Pyrrole-Imidazole Polyamide *seco*-CBI Conjugates

Rhys D. Taylor[a], Anandhakumar Chandran[a], Gengo Kashiwazaki[a], Kaori Hashiya[a], Toshikazu Bando*[a], Hiroki Nagase[b], Hiroshi Sugiyama*[a, c, d]

List of top 50 Bind-n-Seq sequences for Py-Im polyamide **5** with their fold enrichment

| **Rank** | **Sequence** | **Fold-enrichment** |
|---|---|---|
| 1 | ACGCCATC | 89.097 |
| 2 | ACGCCTTC | 76.887 |
| 3 | ACGCCAAC | 69.049 |
| 4 | ATGAAGGC | 64.357 |
| 5 | ATGATGGC | 59.01 |
| 6 | AACGCCAT | 53.967 |
| 7 | CGCCATCA | 53.508 |
| 8 | GCCTTCAA | 44.989 |
| 9 | CGCCAACA | 44.346 |
| 10 | AACGCCAA | 43.933 |
| 11 | ATGGCGTG | 43.67 |
| 12 | CGCCTTCA | 43.474 |
| 13 | AACGCCTT | 42.673 |
| 14 | GCCATCAA | 41.35 |
| 15 | GATGGCGA | 37.998 |
| 16 | AGATGGCG | 36.909 |
| 17 | GCCTTCAC | 36.587 |
| 18 | ACGATGGC | 36.426 |
| 19 | GCCATCAC | 34.206 |

| | | |
|---|---|---|
| 20 | ACGAAGGC | 34.005 |
| 21 | ATGGCGTA | 33.954 |
| 22 | ATGGCGTC | 33.872 |
| 23 | AAGATGGC | 33.394 |
| 24 | CTGAAGGC | 31.669 |
| 25 | ATGTTGGC | 31.45 |
| 26 | CAGATGGC | 31.216 |
| 27 | GCCATCTA | 31.148 |
| 28 | CTGATGGC | 30.388 |
| 29 | CACGCCAA | 28.893 |
| 30 | GCCAACAA | 28.527 |
| 31 | GCCAACAC | 28.487 |
| 32 | AAGGCGTA | 28.268 |
| 33 | TACGCCAA | 28.267 |
| 34 | CGATGGCG | 27.318 |
| 35 | AATGGCGT | 27.023 |
| 36 | AAGGCGTG | 26.651 |
| 37 | GCCATCGA | 25.266 |
| 38 | GACGCCAA | 24.57 |
| 39 | ACGCCATA | 23.506 |
| 40 | GAAGGCGA | 23.445 |
| 41 | GAACGCCA | 22.484 |
| 42 | GCCTTCGA | 21.442 |

| | | |
|---|---|---|
| 43 | AGTTGGCG | 21.399 |
| 44 | GAGATGGC | 21.194 |
| 45 | ATCGCCAT | 20.564 |
| 46 | CTGTTGGC | 20.173 |
| 47 | CCGATGGC | 19.686 |
| 48 | CAGAAGGC | 19.578 |
| 49 | TAACGCCA | 18.75 |
| 50 | GCCTTCTA | 18.617 |