

# A Synthetic Biology Approach to the Expansion of the Genetic Alphabet: Molecular Design of Unnatural Base Pairs of DNA

Dr. S.K. Mishra

Principal, S. Sinha College, Aurangabad

---

## ABSTRACT

**Synthetic biology is a new research area, and its key objective is to create new biological systems by introducing non-natural (unnatural) components. Here, as one of the synthetic biology approaches, we describe studies of the expansion of the genetic alphabet of DNA by creating artificial base pairs (unnatural base pairs). Several unnatural base pairs that function as a third base pair in replication, transcription, and/or translation have recently been developed and are being utilized in a wide range of applications**

---

## INTRODUCTION

In 2010, Craig Venter's group reported the creation of a new *Mycoplasma* bacterium containing an artificially synthesized genome with 1.08 M bases.<sup>1</sup> This achievement was the culmination of the coordinated efforts of chemists and biologists through a 5-year study. The generation of the initial cells required great effort, but once the artificial cells were created, they proliferated in media like natural cells. Thus, the artificially designed cells can be reproduced with reasonable costs and used as a biological factory to synthesize useful proteins and other materials. This re-design of an existing biological system is an example of a synthetic biology **approach**.<sup>2</sup>

Another type of synthetic biology approach is the creation of a new biological system, constructed with newly designed biological components. In this approach, new, artificial components are developed to serve a certain purpose, and they function alongside the natural components in a biological system. The new components are created by repeated "proof of concept" experiments. A prototype component is designed based on a concept or an idea, and then it is improved according to the results from physical and biological experiments. Here, we describe this type of synthetic biology approach, through the creation of unnatural base pairs toward the expansion of the genetic alphabet of DNA.<sup>3-8</sup>

### 2. Development of unnatural base pairs

The genetic information of terrestrial life is encoded in DNA as a sequence comprising the four different bases, adenine (A), guanine (G), cytosine (C) and thymine (T), as alphabetical letters. In duplex DNA molecules, A selectively pairs with T, and G pairs with C. This base pairing rule is fundamental for the genetic information flow through replication, transcription, and translation. Thus, introducing an artificially created pair into DNA could increase the genetic alphabet and expand the genetic information, providing a new biotechnology capable of the site-specific incorporation of new components into nucleic acids and proteins (Figure 1).<sup>9</sup> In addition, recent physical and biological studies on replication and transcription, using artificial, unnatural bases, have revealed novel molecular interactions and biological reaction mechanisms, which had not been observed in conventional analyses using standard biomolecules with only the natural components. Furthermore, artificially increasing the genetic alphabet of a biological system is a formidable challenge for chemists.

The most important issue is that the unnatural base pair (X–Y) functions as a third base pair with highly exclusive selectivity; namely, X selectively pairs with Y, alongside the A–T and G–C pairs in the biological system (Figure 2). In replication, DNA polymerase binds to a partially double stranded DNA fragment between a primer and a template strand. Subsequently, a nucleoside 5'-triphosphate (dNTP, substrate) is imported into the polymerase–DNA complex. When the substrate base correctly pairs with its partner base in the template, then the oxygen atom of the 3'-hydroxy group of the primer attacks the  $\alpha$  position of the triphosphate in the substrate. This results in the formation of the phosphodiester bond between the primer and the imported substrate, and the release of pyrophosphoric acid (pyrophosphate) as a leaving group. After the incorporation of the correct substrate within the primer, the DNA polymerase slides along the template DNA strand, and the incorporation of the next substrate occurs. The selectivity of the natural base pairing in replication is extremely high. For example, in the case of the Klenow fragment of

*Escherichia coli* DNA polymerase I, an incorrect base substrate is incorporated once every 10,000 times,10 which means that the selectivity of the natural base pairing by the Klenow fragment is 99.99% per replication.

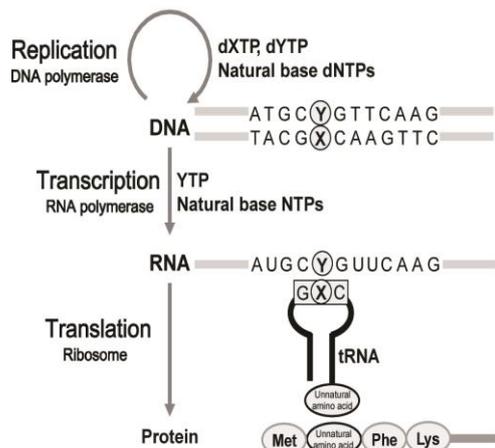


Figure 1. Expansion of the genetic alphabet by an unnatural base pair

### 3. Natural Base Pairs: A–T and G–C

To create an unnatural base pair, we need to understand why the natural A–T and G–C base pairs exhibit high selectivity

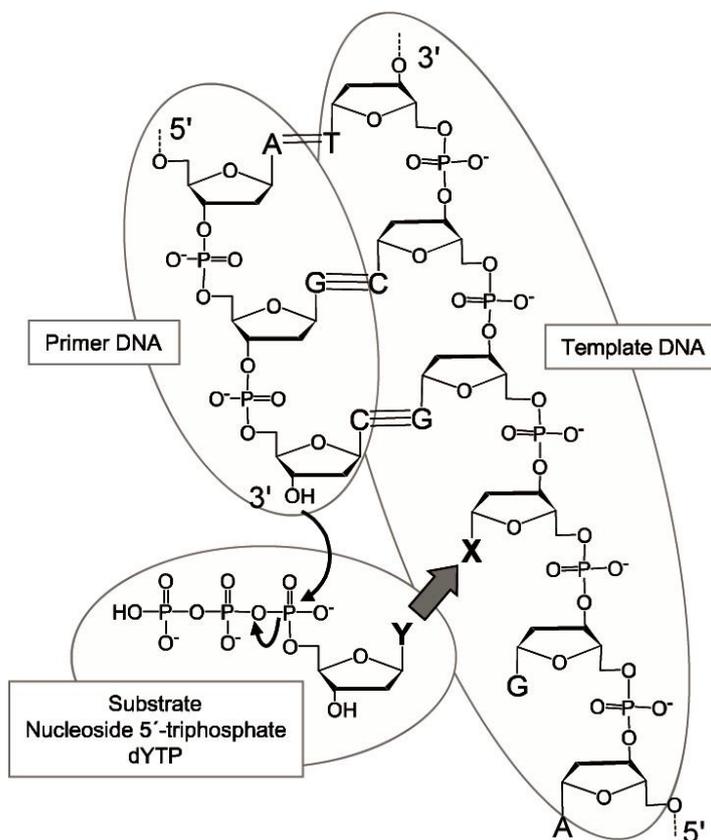
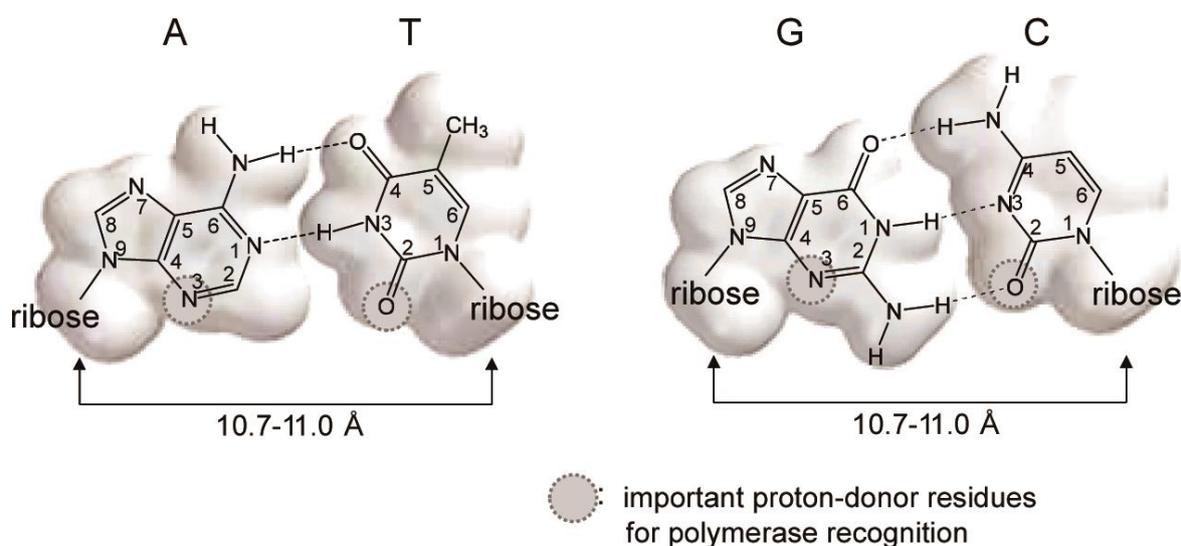


Figure 2. Replication mechanism involving an unnatural base pair

DNA polymerase binds to the partially double-stranded DNA between a primer and a template DNA. The substrate, dYTP, is imported into the protein-DNA complex. When the Y base correctly pairs with the X base in the template, the 3'-hydroxy group of the primer DNA attacks the  $\alpha$  position of the substrate phosphate, resulting in the formation of the phosphodiester bond and the release of the pyrophosphoric acid.



**Figure 3. The natural A-T and G-C base pairs**

Since the distances between the glycoside positions of the pairing bases are always around 11 Å, DNA forms several types of regular double-helices. The Kool group has shown the importance of the shape complementarity between pairing bases in replication.

In replication, in terms of their chemical, physical, and biological aspects. In the A–T and G–C pairings, purine bases (A or G) pair with pyrimidine bases (T or C), with two hydrogen bonds for A–T and three for G–C (Figure 3). Accordingly, the distances between the glycosidic bonds of each pairing base are always 10.7–11.0 Å, and the duplex DNA molecules adopt regular double-helical structures. Hydrogen bonds are formed between a proton donor residue and a proton acceptor atom or residue. Between the A–T and G–C pairs, the hydrogen bond patterns differ from each other, and thus A always pairs with T, and G always pairs with C in the regular helical structures. However, recent studies have revealed that hydrogen bond formation between pairing bases is not necessary for correct base pairing in replication.

In 1995, Eric Kool and his colleagues designed and synthesized a hydrophobic, unnatural Z–F pair, in which the shapes of Z and F mimic those of A and T, respectively (Figure 4).<sup>11-13</sup> For the Z base, the 1- and 3-nitrogens in A were replaced with carbons, and the 6-amino group was replaced with a methyl group. For the F base, the 3-imino group in T was replaced with a C–H group, and the 2- and 4-keto groups were replaced with fluorine atoms. In general, the proton acceptor ability of a fluorine atom is around ten times lower than that of a keto group. Kool's group chemically synthesized DNA templates containing Z or F at a specific position, as well as triphosphates of their nucleosides (dZTP and dFTP), as substrates for replication experiments *in vitro*. They showed that the Klenow fragment of *Escherichia coli* DNA polymerase I efficiently incorporated both dFTP and dTTP into DNA opposite Z or A in the templates with similar efficiencies, and also incorporated both dZTP and dATP opposite F or T. In contrast, the incorporation efficiencies of dCTP opposite Z and dGTP opposite F were very low. Thus, the hydrophobic, unnatural Z–F pair also works in replication and is compatible with the A–T pair, suggesting that the shape complementarity between pairing bases is more important than the hydrogen-bonding complementarity in replication.<sup>14</sup>

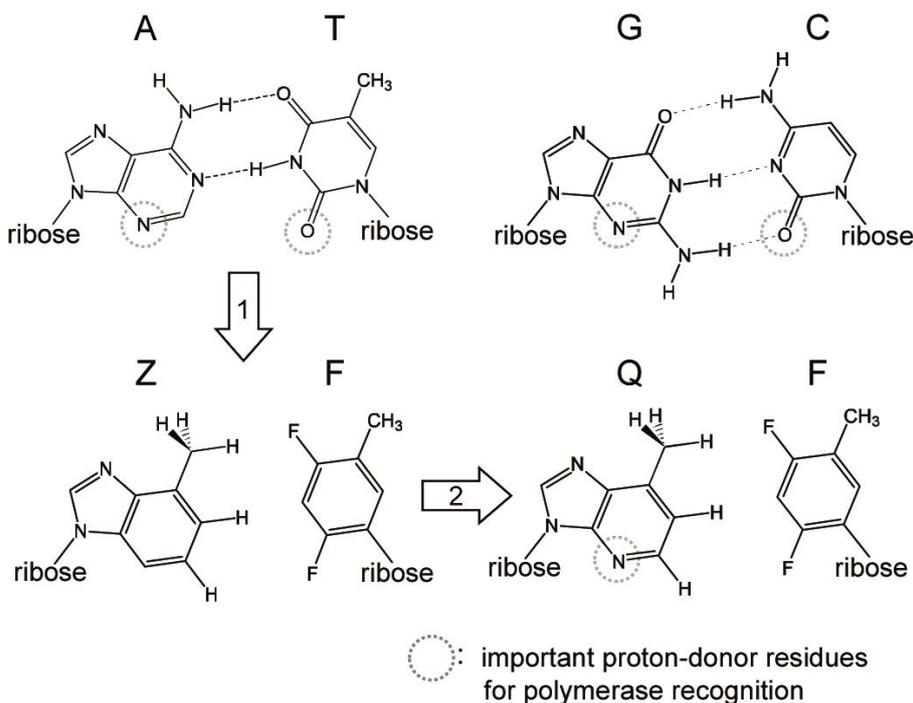
For recognition by polymerases, the 3-nitrogens of A and G, and the 2-keto groups of C and T are necessary as hydrogenbond acceptors oriented toward the minor groove, to interact with specific amino acid residues in polymerases (Figure 3).<sup>15</sup> For example, the substrates of pyrimidine analogs lacking the 2-keto groups are not incorporated into DNA in PCR (Polymerase Chain Reaction).<sup>16</sup> Kool's Z base also lacks the hydrogen acceptor atom corresponding to the 3-nitrogen of purines, and thus, they newly developed the Q base, which has a nitrogen at this position (Figure 4).<sup>17</sup> The Klenow fragment replicated the Q–F pair with higher efficiency than the Z–F pair. Other factors, such as the hydrophobicity of bases, the dipole moments of pairing bases, the stacking interactions among neighboring bases, and the CH– $\pi$  interactions<sup>18</sup>, are also important for the base pair selectivity in replication.

With these points in mind, various types of unnatural base pairs have been developed thus far. In the following sections, we introduce representative unnatural base pairs and their developmental processes.

#### 4. Hydrophilic unnatural base pairs by Benner's group

In the late 1980s, Steven Benner and his colleagues reported four types of unnatural base pairs bearing different hydrogen-bonding geometries, which were unlike those of the A–T and G–C pairs.<sup>19, 20</sup> A representative one is the base pair between 6-amino-2-ketopurine (isoG) and 2-amino-4- ketopyrimidine (isoC), which are the structural isomers of G and C, respectively (Figure 5).<sup>19</sup> Benner's group demonstrated that the isoG–isoC pair works complementarily in replication by the Klenow fragment. In addition, the isoG substrate (isoGTP) was incorporated into RNA opposite isoC in DNA templates by T7 RNA polymerase.<sup>21</sup> Furthermore, in 1992, the isoG–isoC pair was applied to an *in vitro* translation system, allowing the site-specific incorporation of a non-standard amino acid (3-iodotyrosine) into a peptide, using a short, chemically synthesized mRNA containing isoC and a 3-iodotyrosyl tRNA containing isoG.<sup>22</sup>

These pioneering studies by Benner's group attracted much attention to the genetic expansion system using unnatural base pairs. However, these unnatural base pairs still had some shortcomings. For example, the isoG base undergoes keto-enol tautomerization in solution, and the enol form of isoG pairs with T (Figure 5).<sup>21</sup>



**Figure 4. Kool's unnatural base pairs**

The Kool group synthesized an unnatural Z-F pair, mimicking the shapes of the natural A-T pair. The non-hydrogen-bonded Z-F pair functioned in replication with high selectivity, showing the importance of the shape complementarity, rather than hydrogen bonding, between pairing bases in DNA polymerase reactions.

Besides the isoG problem, the 2-amino group of isoC, corresponding to the 2-keto position of the natural bases, reduces the interaction with polymerases.<sup>21</sup> In addition, the nucleoside of isoC is unstable in aqueous solution, and 50% of the isoC nucleoside triphosphate is decomposed under neutral conditions in water at room temperature after four days.<sup>23</sup> Although the introduction of a methyl group to position 3 of the isoC base, corresponding to position 5 of the natural pyrimidine bases, improved the stability, the methyl-isoC ribonucleoside is still susceptible to the epimerization of its  $\beta$ -glycoside bond, from the  $\beta$ - to  $\alpha$  form. Recently, Benner's group addressed this problem by introducing a nitro group to position 3 of isoC,<sup>24</sup> as described below.

In 2005, Benner's group solved the problem of the isoG keto-enol tautomerization by using a modified T base, 2-thiothymine (TS), instead of T (Figure 5), for the isoG–isoC pair system.<sup>25</sup> The large size of the sulfur atom of TS prevents the pairing with the enol form of isoG, but it still pairs with a without steric repulsion. This strategy is similar to Kool's concept of shape fitting between pairing bases, as well as the idea of the unnatural base pair reported by Harry Rappaport in 1988,<sup>26</sup> who used a modified guanine base, 6-thioguanine, as a new base. The unnatural isoG–isoC pair functions together with the A–TS pair in PCR: the selectivity of the isoG–isoC pair reached 98% per replication, while the isoG–isoC selectivity without the help of TS was about 93% per replication.<sup>25</sup> However, 98% unnatural base pairing selectivity is not sufficient for replication, and the retention rate of the unnatural base pair in 20-cycle PCR amplified DNA fragments becomes about 67% ( $0.9820 = 0.67$ ). Thus, for the practical use of unnatural base pairs in replication, more than 99% selectivity per replication is required.

In 2007, Benner's group improved their unnatural base pair system and developed the P-Z pair, which functions in PCR amplification without the use of the A-TS pair (Figure 5).<sup>24</sup> Unlike isoG, the P base does not undergo the undesirable tautomerization in solution. The nitro group of the Z base prevents the epimerization of its nucleoside derivatives, due to the deprotonation of the imino group of the Z base. Although the selectivity of the P-Z pair in PCR was 97.5% at that time, the group recently reported improved selectivity to about 99.8% under optimized PCR conditions.<sup>27</sup>

### 5. Hydrophobic unnatural base pairs by Romesberg's group

Kool's pioneering experiments with the non-hydrogenbonded base pairs generated interest in hydrophobic base pairs as unnatural base pair candidates. Floyd Romesberg and his colleagues developed numerous hydrophobic unnatural base pairs and examined their abilities in replication systems. In 1999, they reported a hydrophobic, isoquinoline derivative, PICS (Figure 6), which pairs self-complementarily in doublestranded DNA fragments with high thermal stability.<sup>28</sup> In addition, the PICS substrate was enzymatically incorporated into DNA opposite PICS in templates by the Klenow fragment. Unfortunately, after the incorporation of the PICS substrate opposite PICS, the primer extension paused. This is because the shape of the PICS-PICS pair is too large for accommodation within the regular DNA double helix, in which the PICS bases stack on each other, and does not allow polymerase recognition. Therefore, they exhaustively designed and synthesized other hydrophobic, unnatural base pairs and examined their abilities in replication.<sup>29-35</sup>

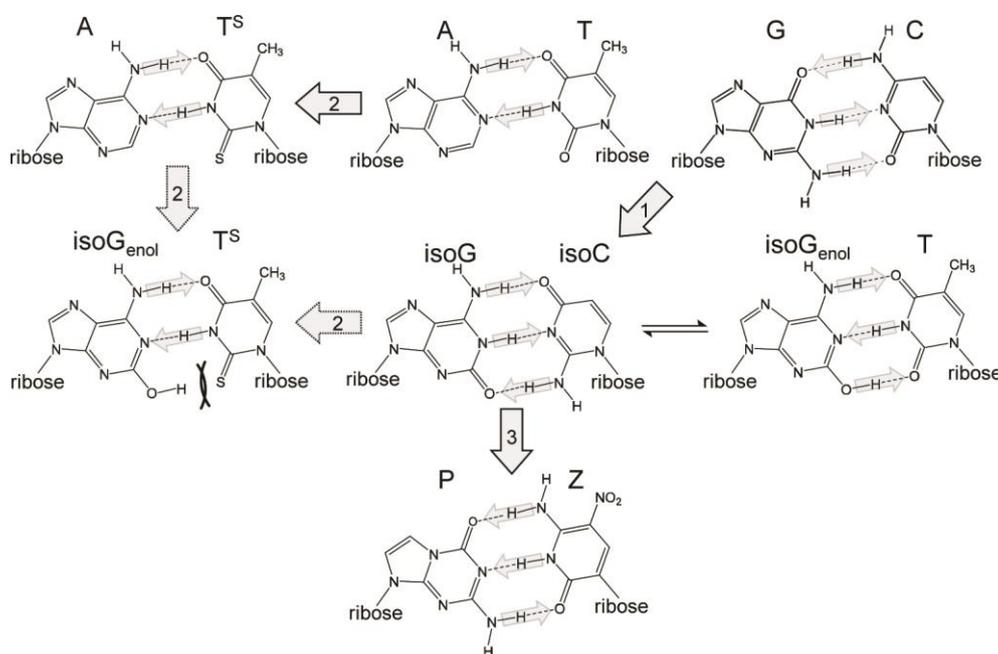


Figure 5. Benner's unnatural base pairs

In 2009, they succeeded in developing the hydrophobic, unnatural base pairs 5SCIS-MMO<sup>2</sup> and 5SCIS-NaM (Figure 6), which function as a third base pair in PCR

The Benner group developed unnatural base pairs, such as isoG-isoC and Z-P, with different hydrogen bond donor-acceptor patterns from those of the A-T and G-C pairs. The P-Z pair functions in PCR amplification. Although the selectivity of the isoG isoC pair in replication is not high because of the keto-enol tautomerization, the isoG isoC pair can be used in PCR by using A-TS in place of the A-T pair. amplification.<sup>36, 37</sup> As compared to the previous PICS-PICS pair, the shape complementarity of these base pairs was improved. The best fidelity of the 5SCIS-NaM pair in PCR reached 99.8%, although it depended on the sequence contexts around the unnatural base pair. These hydrophobic base pairs also function in transcription by T7 RNA polymerase.<sup>38</sup> The group also engineered DNA polymerases by an evolutionary engineering technique, and employed mutated polymerases to solve the polymerase pausing problem when using the PICS-PICS pair in replication.<sup>39</sup>

### 6. A series of unnatural base pairs by Hirao's group

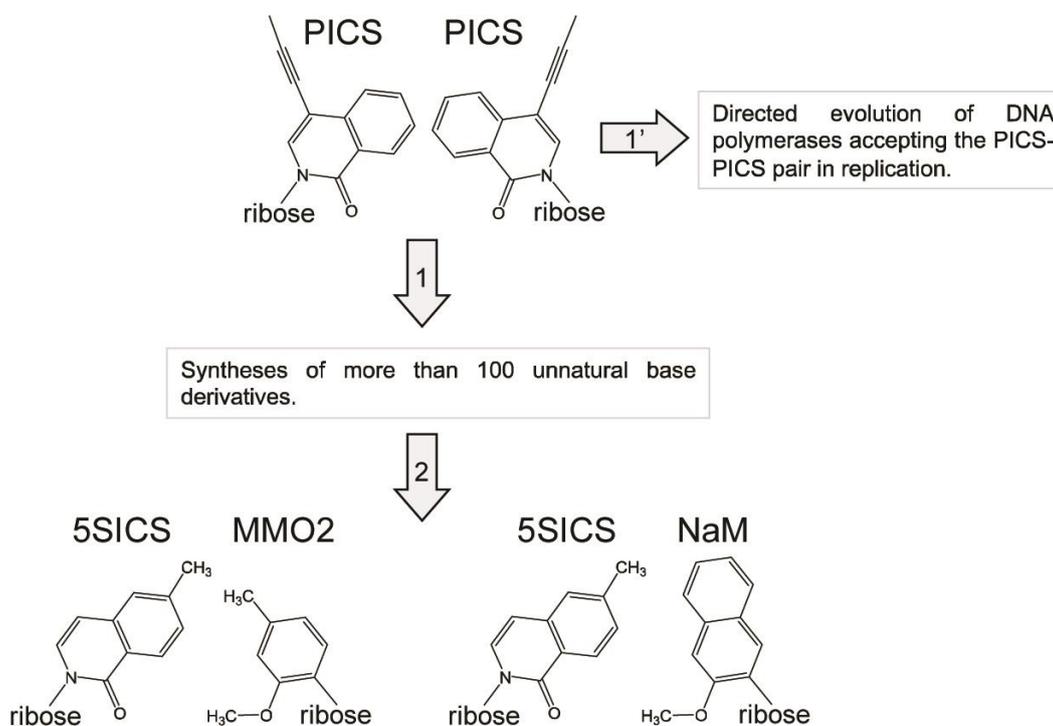
Our group has also been studying unnatural base pairs since 1997. After struggling through lots of 'proof of concept' experiments, we developed our first unnatural base pair, between 2-amino-6-dimethylaminopurine (x) and 2-oxypyridine (y), in 2001 (Figure 7).<sup>40</sup> The x-y was designed by combining two concepts: Benner's hydrogen-bonding patterns, which differ from those of the natural base pairs, and a steric hindrance effect. The x-y pair has two hydrogen bonds, and the 2-aminopurine moiety of x may pair with T through a similar hydrogen bonding interaction as that with y. Thus, we added a bulky dimethylamino group to position 6 of x, to exclude the x-T pairing by steric hindrance between the 6-dimethylamino group and the 4-keto group of T. In contrast to T, y has a hydrogen atom, instead of a

keto group. The x–y pair functions in transcription with high selectivity, and T7 RNA polymerase incorporated the triphosphate substrate of y (yTP) into RNA, opposite x in DNA templates, with more than 95% selectivity.

Furthermore, we improved the shape complementarity of the unnatural base pair by replacing the dimethylamino group of x with a thienyl group. Thus, we developed 2-amino-6-thienylpurine (s) as the pairing partner of y (Figure 7).<sup>41</sup> The planarity of the thienyl group excludes the mispairing with T more efficiently than the dimethylamino group of x. In addition, the planarity increased the stacking ability of the s base with neighboring bases in DNA strands. The selectivity and efficiency of the s–y pair in transcription were greatly improved, as compared to those of the x–y pair, and not only yTP but also several modified yTPs, such as biotin- and fluorophorelinked y bases, were incorporated into RNA, opposite s in DNA templates, by T7 RNA polymerase.<sup>42-44</sup> In 2002, by combining the specific transcription involving the s–y pair and an *in vitro* *E. coli* translation system, we succeeded in the site-specific incorporation of a non-standard amino acid, 3-chlorotyrosine, into a protein.<sup>41</sup> Thus, the s–y pair functions as a third base pair in *in vitro* transcription and translation systems.

We sought to improve our unnatural base pairs further. One major obstacle was the insufficient selectivity of the s–y pair for replication. After 10 cycles of PCR with a DNA fragment containing the s–y pair, around 40% of the unnatural base pair was replaced with the natural base pairs. The selectivity of the s–y pair is ~95% per replication (40% of replacement after 10-cycle PCR:  $1 - 0.95^{10} = \sim 0.4$ ), and thus, more than 99% selectivity is required for unnatural base pairs in replication, in which ~90% ( $0.99^{10} = \sim 0.90$ ) of an unnatural base pair could be retained in its amplified DNA after 10 cycles of PCR. Therefore, we strictly refined the shape complementarity between the pairing bases and designed a five-membered ring nucleobase analog, imidazoline-2-one (z),<sup>45</sup> instead of y with the six-membered ring, as the pairing partner of s (Figure 7). In addition, we removed the atoms and residues involved in hydrogen-bonding interactions between the pairing bases from the s–z pair,<sup>46</sup> and developed a hydrophobic, unnatural base pair between 7-(2-thienyl)imidazo[4,5-*b*]-pyridine (Ds) and pyrrole-2-carbaldehyde (Pa) (Figure 7, see Appendix for chemical syntheses).<sup>47</sup> We added an aldehyde group to the Pa base, as a proton acceptor for interactions with polymerases.

The problem with the Ds – Pa pair is that self-complementary Ds–Ds pairing also competitively occurred in replication. As also found with Romesberg's PICS–PICS pair, the extension paused after the Ds incorporation opposite Ds. Fortunately, we found that a modified triphosphate substrate, the  $\alpha$ -amidotriphosphate, of Ds was selectively incorporated opposite Pa, but showed greatly decreased incorporation opposite Ds.<sup>47</sup> After the modified substrate is incorporated, the DNA has the native phosphodiester linkage, because the amidopyrophosphate moiety of the substrate is



**Figure 6. Romesberg's unnatural base pairs**

The Romesberg group developed PICS, the first replicable self-pair. Then, they screened an unnatural hydrophobic base library, and discovered the highly selective 5SICS-*MMO2* and 5SICS-*NaM* pairs. They also studied polymerase mutations for enhanced incorporation of the PICS-PICS pair.

removed as a leaving group during polymerization. For PCR amplification, we also used the  $\gamma$ -amidotriphosphate of A, besides Ds, to reduce the misincorporation of dATP opposite Pa. In 2006, we successfully performed highly selective PCR amplification involving the Ds–Pa pair using the  $\gamma$ -amidotriphosphates of A and Ds, in which the selectivity of the Ds–Pa pairing reached more than 99% per replication. As a serendipitous mistake, we accidentally synthesized the  $\gamma$ -amidotriphosphates by treating the triphosphate synthesis intermediates, the cyclic 5'-triphosphates, with concentrated ammonia. We further improved the Ds–Pa pair, and replaced the aldehyde group of Pa with a nitro group, thus designing 2-nitropyrrole (Pn) (Figure 7).<sup>48</sup> The nitro group of Pn can reduce the misincorporation of A opposite Pa, by the electrostatic repulsion between the oxygen of the nitro group and the 1-nitrogen of A. In addition, we added a propynyl group to position 4 of the Pn base, and developed 2-nitro-4-propynylpyrrole (Px) as a pairing partner of Ds (Figure 7).<sup>49</sup> The propynyl group increases the hydrophobicity, which strengthens the interaction with polymerases. Thus, the Ds–Px pair functions in PCR without the aid of the  $\gamma$ -amidotriphosphates. DNA fragments containing the Ds–Px pair were amplified 108-fold by 40 cycles of PCR, and more than 97% of the Ds–Px pair was retained in the amplified DNA. The selectivity of the Ds–Px pair reached more than 99.9% per replication, which is the highest selectivity among the unnatural base pairs reported thus far.<sup>50</sup> Diagnostic and therapeutic applications using the Ds–Px pair are in progress. In addition to these unnatural base pairs mentioned here, we also reported unique base pairs with fluorophore or quencher abilities and their applications.

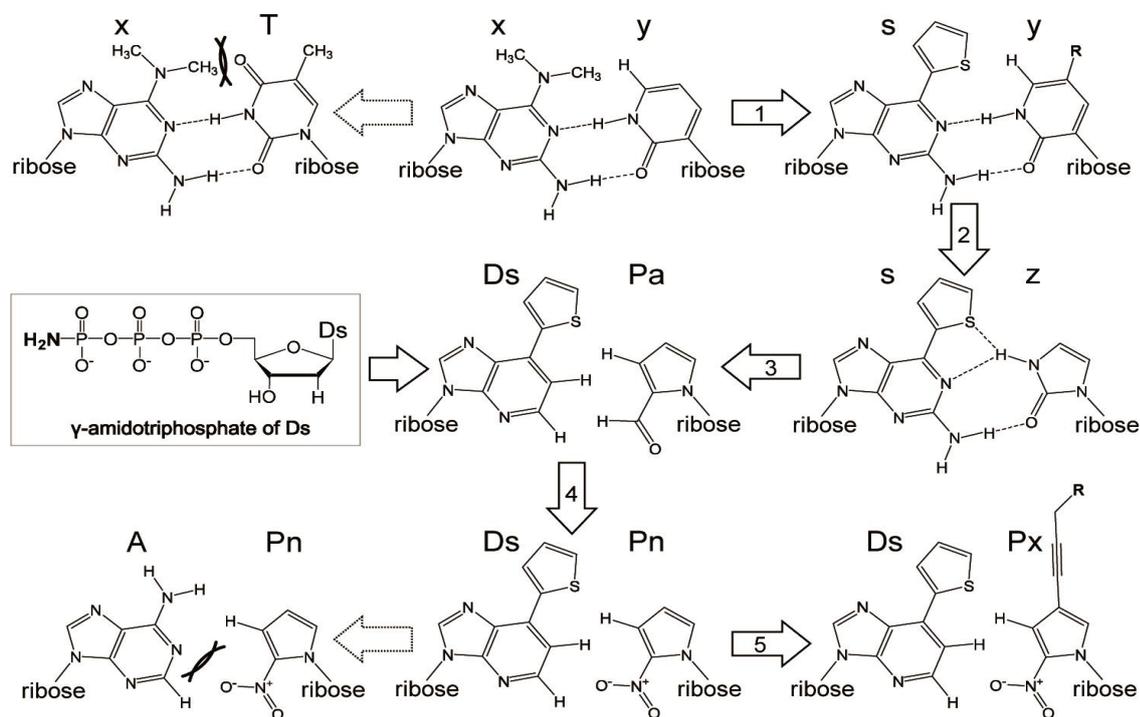
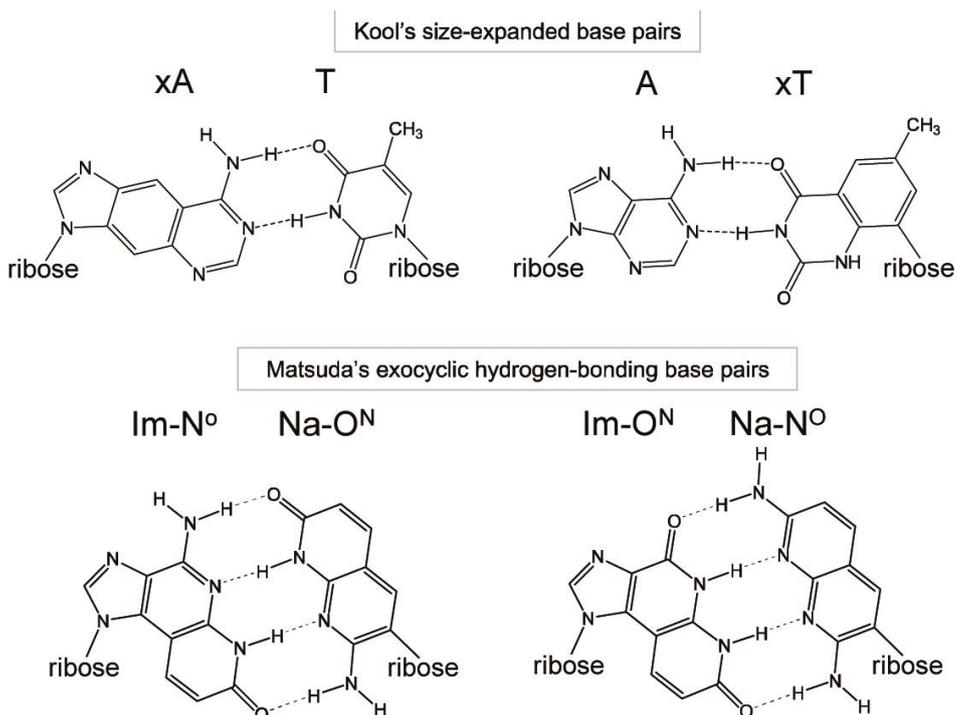


Figure 7. Hirao's unnatural base pairs

I develop several unnatural base pairs by combining the designed concepts of hydrogen-bonding patterns, shape complementarity with steric hindrance and electrostatic repulsion, and hydrophobicity. The Ds-Px pair exhibits high fidelity and efficiency in PCR amplification.

## CONCLUSION

The synthetic biology of unnatural base pairs has rapidly advanced over the past 20 years, and various types of unnatural base pairs were developed (Figure 8).<sup>4, 51-54</sup> These unnatural base pairs have been applied to site-specific fluorescent labeling,<sup>42, 55</sup> immobilization,<sup>43, 47</sup> specific detection,<sup>56-60</sup> and structural analyses<sup>61-63</sup> of DNA and RNA molecules. Although the applications of unnatural base pairs to translation are still being developed,<sup>22, 41</sup> it is only a matter of time until a protein synthesis technology is developed for the site-specific incorporation of non-standard amino acids into proteins. The previous unnatural base pair studies were limited to *in vitro* experiments. However, we now have several types of unnatural base pairs with the potential for use in *in vivo* experiments. In the future, expanded genetic alphabets using unnatural base pairs will be applied to cells, such as Venter's artificial cell, to efficiently produce artificial proteins and to trace target gene expression in the cell. We look forward to further advancements in this area, toward next generation biotechnologies.

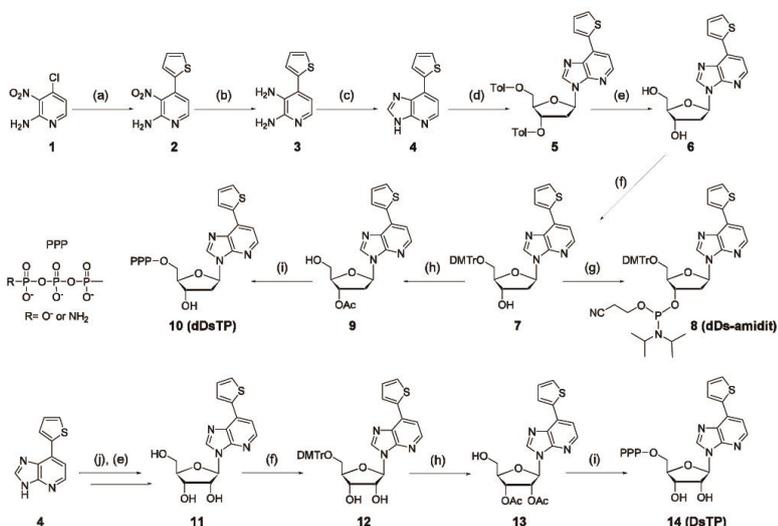


**Figure 8.** Other unnatural base pairs

Synthesis of nucleoside derivatives of Ds Reagents and abbreviations: (a) dichlorobis(triphenylphosphine)palladium, 2-(tributylstannyl)thiophene, DMF; (b) palladium on carbon, sodium borohydride, ethanol, ethylacetate; (c) formic acid; (d) NaH, 2-deoxy-3,5-di-*O*-*p*-toluoyl- $\alpha$ -D-erythro-pentofuranosyl chloride, CH<sub>3</sub>CN; (e) NH<sub>3</sub>, methanol; (f) 4,4'-dimethoxytrityl chloride, pyridine; (g) 2-cyanoethyl tetraisopropylphosphordiamidite, tetrazole, CH<sub>3</sub>CN; (h) acetic anhydride, pyridine, then dichloroacetic acid, dichloromethane; (i) 2-chloro-4H-1,3,2-benzodioxaphosphorin-4-one, dioxane, pyridine, tributylamine, bis(tributylammonium)pyrophosphate, DMF, then I<sub>2</sub>/pyridine, water, NH<sub>4</sub>OH (for triphosphate), I<sub>2</sub>/pyridine, NH<sub>4</sub>OH (for  $\gamma$ -amidotriphosphate); (j) tetra-*O*-acetyl- $\beta$ -D-ribofuranose, chloroacetic acid. Tol: toluoyl, DMT: 4,4'-dimethoxytrityl, Ac: acetyl.

Reagents and abbreviations: (a) NaH, 2-deoxy-3,5-di-*O*-*p*-toluoyl- $\alpha$ -D-erythro-pentofuranosyl chloride, CH<sub>3</sub>CN; (b) NH<sub>3</sub>, methanol; (c) 4,4'-dimethoxytrityl chloride, pyridine; (d) 2-cyanoethyl *N,N*-diisopropylaminochlorophosphoramidite, diisopropylethylamine, THF; (e) 1,8-bis(dimethylamino)naphthalene, POC13, trimethyl phosphate, then tributylamine, bis(tributylammonium)pyrophosphate, DMF; (f) NaH, CH<sub>3</sub>CN, then 2,3,5-tri-*O*-benzyl-D-ribofuranosyl chloride; (g) BBr<sub>3</sub>, dichloromethane. Tol: toluoyl, DMT: 4,4'-dimethoxytrityl, Ac: acetyl.

### Appendix 1



## ACKNOWLEDGMENTS

I would like to extend my heartiest thanks to Prof. U.S. Yadav Retd. HOD, Deptt. of Chemistry, J.P.U. Chapra, Prof. Rabindra Singh, Deptt. of Chemistry, J.P.U. Chapra, Prof. Brij Bihari Sharma, Retd. HOD, Deptt. of Chemistry, M.U. Bodhgaya, Prof. R.P.S. Chauhan, Retd. HOD, Deptt. of Chemistry, M.U. Bodhgaya, Prof. Udai Arbind, Deptt. of Chemistry, J.P.U. Chapra, for providing academic support and valuable suggestions for doing so and impetus force for moving ahead.

## REFERENCES

- [1] D. G. Gibson, J. I. Glass, C. Lartigue, V. N. Noskov, R. Y. Chuang, M. A. Algire, G. A. Benders, M. G. Montague, L. Ma, M. M. Moodie, C. Merryman, S. Vashee, R. Krishnakumar, N. Assad-Garcia, C. Andrews-Pfannkoch, E. A. Denisova, L. Young, Z. Q. Qi, T. H. Segall-Shapiro, C. H. Calvey, P. P. Parmar, C. A. Hutchison, 3rd, H. O. Smith, J. C. Venter, *Science* **2010**, *329*, 52–56.
- [2] D. Sprinzak, M. B. Elowitz, *Nature* **2005**, *438*, 443–448.
- [3] D. E. Bergstrom, *Curr. Protoc. Nucleic Acid Chem.* **2009**, Chapter 1, Unit 1 4.
- [4] A. T. Krueger, E. T. Kool, *Chem. Biol.* **2009**, *16*, 242–248.
- [5] I. Hirao, *Curr. Opin. Chem. Biol.* **2006**, *10*, 622–627.
- [6] A. A. Henry, F. E. Romesberg, *Curr. Opin. Chem. Biol.* **2003**, *7*, 727–733.
- [7] S. A. Benner, A. M. Sismour, *Nat. Rev. Genet.* **2005**, *6*, 533–543.
- [8] I. Hirao, M. Kimoto, R. Yamashige, *Acc. Chem. Res.* in press.
- [9] R. F. Service, *Science* **2000**, *289*, 232–235.
- [10] T. A. Kunkel, K. Bebenek, *Annu. Rev. Biochem.* **2000**, *69*, 497–529.
- [11] S. Moran, R. X. Ren, E. T. Kool, *Proc. Natl. Acad. Sci. USA* **1997**, *94*, 10506–10511.
- [12] J. C. Morales, E. T. Kool, *Nat. Struct. Biol.* **1998**, *5*, 950–954.
- [13] K. M. Guckian, T. R. Krugh, E. T. Kool, *Nat. Struct. Biol.* **1998**, *5*, 954–959.
- [14] T. J. Matray, E. T. Kool, *Nature* **1999**, *399*, 704–708.
- [15] E. T. Kool, *Annu. Rev. Biochem.* **2002**, *71*, 191–219.
- [16] M. J. Guo, S. Hildbrand, C. J. Leumann, L. W. McLaughlin, M. J. Waring, *Nucleic Acids Res.* **1998**, *26*, 1863–1869.
- [17] J. C. Morales, E. T. Kool, *J. Am. Chem. Soc.* **1999**, *121*, 2323–2324.
- [18] O. Takahashi, Y. Kohno, M. Nishio, *Chem. Rev.* **2010**, *110*, 6049–6076.
- [19] C. Switzer, S. E. Moroney, S. A. Benner, *J. Am. Chem. Soc.* **1989**, *111*, 8322–8323.
- [20] J. A. Piccirilli, T. Krauch, S. E. Moroney, S. A. Benner, *Nature* **1990**, *343*, 33–37.
- [21] C. Y. Switzer, S. E. Moroney, S. A. Benner, *Biochemistry* **1993**, *32*, 10489–10496.
- [22] J. D. Bain, C. Switzer, A. R. Chamberlin, S. A. Benner, *Nature* **1992**, *356*, 537–539.
- [23] J. J. Coegel, S. A. Benner, *Helv. Chim. Acta* **1996**, *79*, 1863–1880.
- [24] Z. Yang, A. M. Sismour, P. Sheng, N. L. Puskar, S. A. Benner, *Nucleic Acids Res.* **2007**, *35*, 4238–4249.
- [25] A. M. Sismour, S. A. Benner, *Nucleic Acids Res.* **2005**, *33*, 5640–5646.
- [26] H. P. Rappaport, *Nucleic Acids Res.* **1988**, *16*, 7253–7267.
- [27] Z. Yang, F. Chen, J. B. Alvarado, S. A. Benner, *J. Am. Chem. Soc.* **2011**, *133*, 15105–15112.
- [28] D. L. McMinn, A. K. Ogawa, Y. Wu, J. Liu, P. G. Schultz, F. E. Romesberg, *J. Am. Chem. Soc.* **1999**, *121*, 11585–11586.
- [29] E. L. Tae, Y. Wu, G. Xia, P. G. Schultz, F. E. Romesberg, *J. Am. Chem. Soc.* **2001**, *123*, 7439–7440.
- [30] A. A. Henry, C. Yu, F. E. Romesberg, *J. Am. Chem. Soc.* **2003**, *125*, 9638–9646.
- [31] A. A. Henry, A. G. Olsen, S. Matsuda, C. Yu, B. H. Geierstanger, F. E. Romesberg, *J. Am. Chem. Soc.* **2004**, *126*, 6923–6931.
- [32] A. M. Leconte, S. Matsuda, G. T. Hwang and F. E. Romesberg, *Angew. Chem. Intl. Ed. Engl.* **2006**, *45*, 4326–4329.
- [33] S. Matsuda, J. D. Fillo, A. A. Henry, P. Rai, S. J. Wilkens, T. J. Dwyer, B. H. Geierstanger, D. E. Wemmer, P. G. Schultz, G. Spraggon, F. E. Romesberg, *J. Am. Chem. Soc.* **2007**, *129*, 10466–10473.
- [34] A. M. Leconte, G. T. Hwang, S. Matsuda, P. Capek, Y. Hari, F. E. Romesberg, *J. Am. Chem. Soc.* **2008**, *130*, 2336–2343.
- [35] Y. J. Seo, G. T. Hwang, P. Ordoukhanian, F. E. Romesberg, *J. Am. Chem. Soc.* **2009**, *131*, 3246–3252.
- [36] D. A. Malyshev, Y. J. Seo, P. Ordoukhanian, F. E. Romesberg, *J. Am. Chem. Soc.* **2009**, *131*, 14620–14621.
- [37] D. A. Malyshev, D. A. Pfaff, S. I. Ippoliti, G. T. Hwang, T. J. Dwyer, F. E. Romesberg, *Chemistry* **2010**, *16*, 12650–12659.
- [38] Y. J. Seo, S. Matsuda, F. E. Romesberg, *J. Am. Chem. Soc.* **2009**, *131*, 5046–5047.
- [39] A. M. Leconte, L. Chen, F. E. Romesberg, *J. Am. Chem. Soc.* **2005**, *127*, 12470–12471.
- [40] T. Ohtsuki, M. Kimoto, M. Ishikawa, T. Mitsui, I. Hirao, S. Yokoyama, *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 4922–4925.
- [41] I. Hirao, T. Ohtsuki, T. Fujiwara, T. Mitsui, T. Yokogawa, T. Okuni, H. Nakayama, K. Takio, T. Yabuki, T. Kigawa, K. Kodama, K. Nishikawa, S. Yokoyama, *Nat. Biotechnol.* **2002**, *20*, 177–182.

- [42] R. Kawai, M. Kimoto, S. Ikeda, T. Mitsui, M. Endo, S. Yokoyama, I. Hirao, *J. Am. Chem. Soc.* **2005**, *127*, 17286–17295.
- [43] K. Moriyama, M. Kimoto, T. Mitsui, S. Yokoyama, I. Hirao, *Nucleic Acids Res.* **2005**, *33*, e129.
- [44] I. Hirao, *Biotechniques* **2006**, *40*, 711–715.
- [45] I. Hirao, Y. Harada, M. Kimoto, T. Mitsui, T. Fujiwara, S. Yokoyama, *J. Am. Chem. Soc.* **2004**, *126*, 13298–13305.
- [46] T. Mitsui, A. Kitamura, M. Kimoto, T. To, A. Sato, I. Hirao, S. Yokoyama, *J. Am. Chem. Soc.* **2003**, *125*, 5298–5307.
- [47] I. Hirao, M. Kimoto, T. Mitsui, T. Fujiwara, R. Kawai, A. Sato, Y. Harada, S. Yokoyama, *Nat. Methods* **2006**, *3*, 729–735.
- [48] I. Hirao, T. Mitsui, M. Kimoto, S. Yokoyama, *J. Am. Chem. Soc.* **2007**, *129*, 15549–15555.
- [49] M. Kimoto, R. Kawai, T. Mitsui, S. Yokoyama, I. Hirao, *Nucleic Acids Res.* **2009**, *37*, e14.
- [50] R. Yamashige, M. Kimoto, Y. Takezawa, A. Sato, T. Mitsui, S. Yokoyama, I. Hirao, *Nucleic Acids Res.* in press.
- [51] J. Gao, H. Liu and E. T. Kool, *J. Am. Chem. Soc.* **2004**, *126*, 11826–11831.
- [52] N. Minakawa, S. Ogata, M. Takahashi, A. Matsuda, *J. Am. Chem. Soc.* **2009**, *131*, 1644–1645.
- [53] S. Hikishima, N. Minakawa, K. Kuramoto, Y. Fujisawa, M. Ogawa, A. Matsuda, *Angew. Chem. Int. Ed. Engl.* **2005**, *44*, 596–598.
- [54] C. Kaul, M. Muller, M. Wagner, S. Schneider, T. Carell, *Nat. Chem.* **2011**, *3*, 794–800.
- [55] M. Kimoto, T. Mitsui, S. Yokoyama, I. Hirao, *J. Am. Chem. Soc.* **2010**, *132*, 4988–4989.
- [56] C. B. Sherrill, D. J. Marshall, M. J. Moser, C. A. Larsen, L. Daude-Snow, S. Jurczyk, G. Shapiro, J. R. Prudent, *J. Am. Chem. Soc.* **2004**, *126*, 4550–4556.
- [57] S. C. Johnson, D. J. Marshall, G. Harms, C. M. Miller, C. B. Sherrill, E. L. Beaty, S. A. Lederer, E. B. Roesch, G. Madsen, G. L. Hoffman, R. H. Laessig, G. J. Kopish, M. W. Baker, S. A. Benner, P. M. Farrell, J. R. Prudent, *Clin. Chem.* **2004**, *50*, 2019–2027.
- [58] P. Sheng, Z. Yang, Y. Kim, Y. Wu, W. Tan, S. A. Benner, *Chem. Commun.* **2008**, 5128–5130.
- [59] S. Hoshika, F. Chen, N. A. Leal, S. A. Benner, *Angew. Chem. Int. Ed. Engl.* **2010**, *49*, 5554–5557.
- [60] R. Yamashige, M. Kimoto, T. Mitsui, S. Yokoyama, I. Hirao, *Org. Biomol. Chem.* **2011**, *9*, 7504–7509.
- [61] M. Kimoto, M. Endo, T. Mitsui, T. Okuni, I. Hirao, S. Yokoyama, *Chem. Biol.* **2004**, *11*, 47–55.
- [62] M. Kimoto, T. Mitsui, Y. Harada, A. Sato, S. Yokoyama, I. Hirao, *Nucleic Acids Res.* **2007**, *35*, 5360–5369.
- [63] Y. Hikida, M. Kimoto, S. Yokoyama, I. Hirao, *Nat. Protoc.* **2010**, *5*, 1312–1323.
- [64]