J|A|C|S COMMUNICATIONS

Published on Web 03/19/2009

Transcription of an Expanded Genetic Alphabet

Young Jun Seo, Shigeo Matsuda, and Floyd E. Romesberg*

Department of Chemistry, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, California 92037

Received January 28, 2009; E-mail: floyd@scripps.edu

The genetic alphabet is constrained by the four natural nucleotides and the two base pairs that they form. An unnatural base pair that is selectively replicated, transcribed, and translated would dramatically expand the information potential of the genetic alphabet.[1−4] This would also increase the potential of the already ubiquitous methodologies based on DNA and RNA and their sequence-specific amplification. Moreover, in vivo expansion of the genetic alphabet would serve as the foundation of a semisynthetic organism with an expanded genetic code.

Toward this goal, we have focused on developing unnatural base pairs formed between predominantly hydrophobic unnatural nucleotides. These unnatural base pairs are stable and selectively replicated by DNA polymerases on the basis of complementary shape and packing interactions rather than complementary hydrogen bonding. Through screening of a library of nucleotide analogues followed by hit optimization, we identified the unnatural base pair formed between d**5SICS** and d**MMO2** (Figure 1),[2a] which is

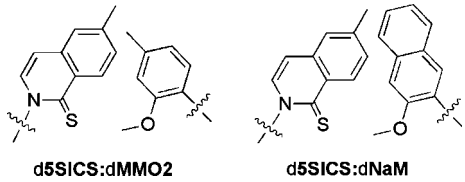

Figure 1. The d**5SICS**:d**MMO2** and d**5SICS**:d**NaM** base pairs. Sugar and phosphate backbone omitted for clarity.

relatively well recognized by a variety of different replicative DNA polymerases.[2b] Further optimization identified d**NaM**, which pairs with d**5SICS** to form an unnatural pair that is replicated with efficiencies and fidelities approaching those of a natural base pair.[2c]

Expansion of the genetic code requires unnatural base pairs that are not only replicable but also transcribed with good efficiency and selectivity in both strand contexts (i.e., d**X** must template **Y**TP insertion and d**Y** must template **X**TP insertion). Previous studies have examined the transcription of unnatural nucleotides bearing nucleobase analogues that pair on the basis of either orthogonal hydrogen bonding[1] or hydrogen bonding and hydrophobicity;[3a,b] however, in none of these cases was transcription of the unnatural base pair shown to be efficient and selective in both possible strand contexts. In addition, it remains unclear whether nucleobase shape and hydrophobicity alone are sufficient for transcription.

To characterize the transcription of the unnatural base pairs formed by d**5SICS** with d**MMO2** and d**NaM**, ribonucleotides and deoxynucleotides were synthesized and converted to the corresponding triphosphates or deoxyphosphoramidites, and the deoxyphorphoramidites were incorporated into DNA templates using automated DNA synthesis (see the Supporting Information). Transcription experiments were conducted with 100 nM DNA substrate, 1× Takara buffer [40 mM Tris-HCl (pH 8.0), 8 mM $MgCl_2$, 2 mM spermidine], DEPC-treated and nuclease-free sterilized water (Fisher), T7 RNA polymerase (50 units), 20 μM each natural NTP, α-^{32}P-ATP (2.5 μCi, MP Biomedicals), and either 5 μM **5SICS**TP, 10 μM **MMO2**TP, or 10 μM **NaM**TP. After incubation for 2 h at 37 °C, the reaction was quenched by the addition of 10 μL of gel loading solution (10 M urea, 0.05% bromophenol blue), and the reaction mixture was loaded onto a 20% polyacrylamide-7 M urea gel, subjected to electrophoresis, and analyzed by phosphorimaging (see the Supporting Information).

We first characterized the ability of d**5SICS** to template the transcription of RNA containing **MMO2** or **NaM** (Figure 2). In

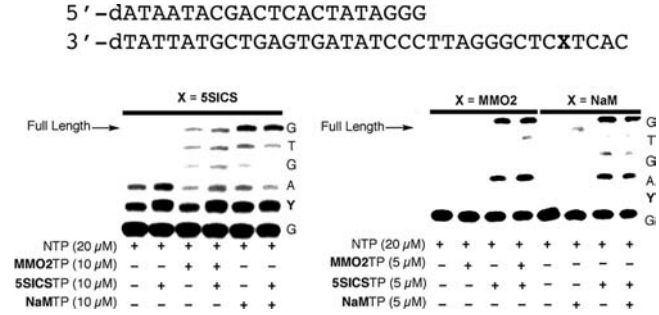

Figure 2. Full-length transcription of DNA containing **5SICS**, **MMO2**, or **NaM.** The template sequence is shown above. **X** and **Y** correspond to the indicated unnatural base in the template and transcript, respectively.

the absence of the unnatural triphosphates, no full length product was observed. Most of the truncated product corresponded to the termination of transcription immediately before d**5SICS** in the template, although a small amount corresponded to termination after mispairing or after single-nucleotide extension of a mispair. In contrast, in the presence of **MMO2**TP, a small amount of full-length transcript was observed, although a significant amount of truncated product remained after 2 h. In the presence **NaM**TP, significantly more full-length product was observed, revealing that d**5SICS** templates the incorporation of **NaM** into RNA more efficiently than it templates the incorporation of **MMO2**. This parallels the behavior observed with DNA polymerases,[2] suggesting that at least some aspects of unnatural base pair recognition are conserved between the two classes of enzymes. In the presence of either unnatural triphosphate, the major truncation products were those corresponding to termination immediately prior to and at the unnatural nucleotide in the template, which suggests that unnatural transcription is limited by both the rate at which the unnatural base pair is synthesized and the rate with which it is extended. This is again similar to what is observed with DNA polymerases.[2] Thus, we tentatively conclude that recognition of the unnatural base pairs with d**5SICS** in the template is similar for DNA and RNA polymerases. Importantly, the addition of **5SICS**TP did not alter the amount of transcript produced, suggesting that the self-pair does not inhibit transcription.

We next characterized transcription of the unnatural base pair in the opposite strand context by examining the ability of d**MMO2**

10.1021/ja9006996 CCC: $40.75

Table 1. Nucleotide Composition Analysis of T7 Transcription Products[a]

		normalized composition of nucleotides incorporated 5′ to A during transcription				
template dN	triphosphate	Ap	Gp	Cp	Up	**X**p
dA	none	1.01 ± 0.03 [1]	1.98 ± 0.03 [2]	n.d. [0]	0.99 ± 0.01 [1]	n.d [0]
d**5SICS**	**MMO2**TP	1.07 ± 0.03 [1]	1.96 ± 0.02 [2]	n.d. [0]	n.d. [0]	0.97 ± 0.03 [1]
d**5SICS**	**NaM**TP	0.98 ± 0.02 [1]	2.02 ± 0.03 [2]	n.d. [0]	n.d. [0]	0.99 ± 0.01 [1]
d**MMO2**	**5SICS**TP	1.07 ± 0.03 [1]	1.99 ± 0.05 [2]	n.d. [0]	n.d. [0]	0.93 ± 0.07 [1]
d**NaM**	**5SICS**TP	1.03 ± 0.04 [1]	2.02 ± 0.02 [2]	n.d. [0]	n.d. [0]	0.94 ± 0.03 [1]

[a] Values were determined by dividing the radioactivity observed for a particular radiolabeled monophosphate by the product of the total radioactivity and the expected number of nucleotides. Predicted values assuming 100% fidelity are shown in brackets. The error reported is the standard deviation of at least three independent determinations. n.d., not detected. See refs 3a and 3b and the Supporting Information for experimental details.

or d**NaM** to template the transcription of RNA containing **5SICS** (Figure 2). Again, in the absence of the unnatural triphosphates, virtually no full-length product was observed. With either unnatural nucleotide in the template, transcription occurred up to the unnatural nucleotide and then halted, yielding only truncated product. The addition of the cognate unnatural triphosphate, in this case **5SICS**TP, resulted in full-length transcript with both unnatural templates. The addition of either **MMO2**TP or **NaM**TP did not interfere with transcription, again indicating that transcription is not inhibited by self-pairing of the hydrophobic nucleobases. In contrast to transcription with d**5SICS** in the template, the observed pattern of truncated products with d**MMO2** or d**NaM** in the template suggests that the efficiency of continued transcription after synthesis of the unnatural base pair is higher than the efficiency of unnatural base pair synthesis. This contrasts with the behavior of DNA polymerases, for which synthesis of the unnatural pairs is more efficient than extension.[2]

We next used 2D TLC to confirm the high fidelity of unnatural base pair transcription (Table 1 and the Supporting Information)[3,5] Consistent with the observation that full-length transcription requires the presence of the unnatural triphosphate, we found that the fidelity of transcription, which was greater than 98% in all cases, was not significantly different than that for a natural base pair under identical conditions.

Having established that each nucleotide selectively templates the incorporation of its partner into RNA, we next examined transcription efficiency by measuring (at low percent conversion) the amount of full-length product formed as a function of time (see the Supporting Information). Relative to the rate at which a fully natural sequence is transcribed, the incorporation of a single **NaM** or **MMO2** opposite d**5SICS** reduces the rate of full-length transcription by a factor of only 16 or 41, respectively. The incorporation of a single **5SICS** opposite either d**MMO2** or d**NaM** reduces the rate of full-length transcription by a factor of 26 or 24, respectively. These relative rates of unnatural base pair transcription agree well with the qualitative gel data described above. The only ∼20-fold reduction in transcription efficiency of d**5SICS**:d**NaM** in both strand contexts is remarkable and again suggests that at least some of the determinants of substrate recognition are similar with DNA and RNA polymerases and that in contrast to previously characterized unnatural base pairs, these general determinants of recognition are possessed by d**5SICS**:d**NaM**.

Efforts to expand the genetic alphabet rely on the development of an unnatural base pair that is efficiently and selectively replicated and transcribed in both strand contexts. We have now demonstrated not only that d**5SICS**:d**NaM** and (to a somewhat lesser extent) d**5SICS**:d**MMO2** are efficiently and selectively replicated but also that d**5SICS**:d**NaM** is efficiently and selectively transcribed. This suggests that as in the case of replication, nucleobase shape and hydrophobicity are sufficient to underlie selective and efficient transcription of an unnatural base pair into RNA. Indeed, the d**5SICS**:d**NaM** unnatural base pair seems well-suited for use as part of an expanded genetic alphabet, with immediate in vitro applications,[6,7] as well as for the long-term goal of creating a semisynthetic organism with an expanded genetic code.[8,9]

Acknowledgment. We thank Prof. Ichiro Hirao for help with the 2D TLC experiments and the National Institutes of Health (GM060005) for funding.

Supporting Information Available: Details of compound synthesis and transcription analysis. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) (a) Piccirilli, J. A.; Krauch, T.; Moroney, S. E.; Benner, S. A. *Nature* **1990**, *343*, 33–37. (b) Yang, Z.; Sismour, A. M.; Sheng, P.; Puskar, N. L.; Benner, S. A. *Nucleic Acids Res.* **2007**, *35*, 4238–4249.
(2) (a) Leconte, A. M.; Hwang, G. T.; Matsuda, S.; Capek, P.; Hari, Y.; Romesberg, F. E. *J. Am. Chem. Soc.* **2008**, *130*, 2336–2343. (b) Hwang, G. T.; Romesberg, F. E. *J. Am. Chem. Soc.* **2008**, *130*, 14872–14882. (c) Seo, Y. J.; Hwang, G. T.; Ordoukhanian, P.; Romesberg, F. E. *J. Am. Chem. Soc.* **2009**, *131*, 3246–3252.
(3) (a) Mitsui, T.; Kimoto, M.; Harada, Y.; Yokoyama, S.; Hirao, I. *J. Am. Chem. Soc.* **2005**, *127*, 8652–8658. (b) Hirao, I.; Harada, Y.; Kimoto, M.; Mitsui, T.; Fujiwara, T.; Yokoyama, S. *J. Am. Chem. Soc.* **2004**, *126*, 13298–13305. (c) Hirao, I.; Mitsui, T.; Kimoto, M.; Yokoyama, S. *J. Am. Chem. Soc.* **2007**, *129*, 15549–15555.
(4) Krueger, A. T.; Lu, H.; Lee, A. H. F.; Kool, E. T. *Acc. Chem. Res.* **2007**, *40*, 141–150.
(5) Transcripts were synthesized in reactions containing α-^{32}P-ATP and then digested with RNAseI. This yielded mononucleotides that have a 3′ phosphate originating from the nucleotide that is immediately 3′ in the transcript. Mononucleotides were resolved by 2D TLC and quantified using phosphorimaging. Fidelity was measured as the ratio of measured radioactivity for each mononucleotide to the corresponding predicted value.
(6) (a) Seeman, N. C. *Trends Biochem. Sci.* **2005**, *30*, 119–235. (b) Seeman, N. C. *Mol. Biotechnol.* **2007**, *37*, 246–257.
(7) Stoltenburg, R.; Reinemann, C.; Strehlitz, B. *Biomol. Eng.* **2007**, *24*, 381–403.
(8) Wang, L.; Brock, A.; Herberich, B.; Schultz, P. G. *Science* **2001**, *292*, 498–500.
(9) Benner, S. A.; Sismour, A. M. *Nat. Rev. Genet.* **2005**, *6*, 533–543.

JA9006996