Published in final edited form as:

Biochemistry. 2019 June 04; 58(22): 2581–2583. doi:10.1021/acs.biochem.9b00274.

Eight-Letter DNA

Vivian T. Dien, Matthew Holcomb, Floyd E. Romesberg*

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

Abstract

Steve Benner and collaborators have recently reported an analysis of DNA containing eight nucleotide letters – the four natural letters dG, dC, dA, and dT, and four additional letters, called dP, dZ, dS, and dB. Their analysis demonstrates that the additional letters do not perturb the structure or stability of the base pairs formed between the natural letters, and remarkably, that the new base pairs, dP-dZ and dS-dB, behave virtually identically to the natural base pairs. This unprecedented result convincingly demonstrates that the thermodynamic and structural behavior previously thought to be the purview of only natural DNA is in fact not unique and can be imparted to suitably designed synthetic components. In addition, the first evidence that the eight letter DNA can be transcribed into RNA by a mutant RNA polymerase is presented, paving the way for the transfer of increased information from one biopolymer to another. Along with others working to develop unnatural DNA base pairs for both *in vitro* and *in vivo* applications, this work represents an important step toward the expansion of the genetic alphabet, a central goal of synthetic biology, and has profound implications for our understanding of the molecules and forces that can make life possible.

Graphical Abstract

The natural genetic alphabet encodes virtually all heritable information, but consists of only four nucleotide letters. Expansion of the natural alphabet would have profound conceptual and practical implications. In the late 1990's Steve Benner initiated efforts to discover additional nucleotide letters, and this work inspired others, including us and the Hirao lab, to seek six-letter DNA. Now, with his team at the Foundation for Applied Molecular Evolution (FfAME), Benner has reported their initial efforts toward eight-letter DNA.¹

The UBPs employed in eight-letter DNA are dP-dZ and dS-dB (Fig. 1). Both UBPs were designed to pair via complementary hydrogen bonding patterns that are orthogonal to those

*Corresponding Author floyd@scripps.edu.

Notes

F.E.R. has a financial interest (shares) in Synthorx, Inc., a company that has commercial interests in using UBPs to produce unnatural proteins.

that underlie the pairing of natural nucleotides. The dZ-dB UBP has been the focus of the Benner lab to create six-letter DNA since its initial report in 2006.² The dS-dB UBP resulted from the optimization of the disoG and disoC pair, originally suggested by Alex Rich, to overcome problems with tautomerization and nucleobase loss.³⁻⁶

To characterize the thermodynamics of their eight-letter DNA, the FfAME group turned to John SantaLucia Jr., President and CEO of DNA Software, Inc., who literally wrote the book on predicting the stability of natural DNA duplexes.⁷ To generalize the canonical nearest-neighbor analysis method, which treats the stability of the duplex DNA as the sum of stabilities of dimers, for example AC/TG, the DNA Software group determined the additional parameters required to predict the stability of the eight-letter DNA. Twenty-eight additional parameters were needed, twenty-two to describe dimer duplex stabilities for combinations of the natural and unnatural nucleotides, four to describe the stability of unnatural dimers, and two to describe the stability of terminal unnatural nucleotides. They assessed the quality of the parameterization by comparing the predicted stabilities with those of the duplex and found good agreement. Additionally, while the parameters were estimated from the melting of 8-mer duplexes, they also determined the melting temperature of a 16-mer duplex and found that it was well estimated by their method. Interestingly, while a dC-dG to dZ-dP and a dT-dA to dS-dB substitution each stabilize the duplex, the latter more so, the effects are sequence dependent. Overall, the results are remarkable — they reveal that the forces stabilizing the UBPs in duplex DNA are not identical to those that mediate the stability of their natural counterparts, possibly due to the strong packing mediated by the aryl-nitro group, but that their contribution was still predictable in different sequence contexts, just as with natural DNA. No other material has this property.

There is no structure more canonically associated with life than that of the DNA double helix. To characterize whether their eight-letter DNA duplexes adopt the same structures, the FfAME group turned to Millie Georgiadis at the Indiana University School of Medicine. Previously, the Georgiadis and Benner groups showed that DNA containing the dZ-dP UBP in various sequence contexts adopted natural, DNA-like structures.⁸ In Hoshika *et al.*,¹ they extend their collaborative work to three self-complementary 16-mer duplexes containing both the dP-dZ and dS-dB UBPs, with up to six in a row, and they again found that the unnatural duplexes adopted natural-like structures, with minor and major groove widths similar to GC-rich DNA, and with small deviations only in the opening angle of the dS-dB pair and buckle angle of dZ-dP pair, relative to a dG-dC pair at the same positions. Thus, as with six-letter DNA containing only the dP-dZ UBP, eight-letter DNA appears to adopt structures that are for all intents and purposes indistinguishable from the natural four-letter counterpart. Having demonstrated that eight-letter DNA adopts the same, canonical double helical structure of four-letter DNA, several previous studies must have loomed large in the back of the authors' minds. The authors had previously demonstrated that DNA containing the dP-dZ UBP could be transcribed into RNA, and moreover, in collaboration with Joseph Piccirilli's group at the University of Chicago, they showed that replacing a G-C pair within a riboswitch stem with the P-Z UBP did not significantly effect conformation nor ligand recognition.⁹ This suggested that the unnatural nucleotides may be transcribed into RNA and even incorporated into functional RNAs without ablating activity – setting the stage for

efforts to explore the conversion of the eight-letter DNA into eight-letter RNA, perhaps even functional eight-letter RNA.

Initial efforts to convert eight-letter DNA into eight-letter RNA failed, due to an apparent inability of natural RNA polymerases to insert rSTP opposite dB in a template. Thus, Benner turned to the research group of his former PhD student Andrew Ellington at the University of Texas at Austin. The Ellington group had been exploring the substrate repertoire of a collection of T7 RNA polymerase variants bearing mutations that confer increased thermal stability. As part of a previous collaboration with the Hirao group, this collection already produced a mutant that more efficiently transcribes DNA containing a predominantly hydrophobic UBP, which enabled the incorporation of functionalized analogs into RNA.¹⁰ The collection of thermostable polymerases again delivered for Hoshika *et al.*,¹ – providing a triple mutant, Y639F/H784A/P266L, referred to as FAL, that appeared to insert each of the new ribonucleoside triphosphates opposite its cognate nucleotide in a DNA template, including the insertion of rSTP opposite dB.

To demonstrate the FAL-mediated conversion of eight-letter DNA into eight-letter RNA, the authors transcribed a model template containing a single unnatural deoxynucleotide, or a template encoding the spinach aptamer with the dP-dZ and dS-dB UBPs positioned in a stem distal to the fluorophore binding site. While the fidelity of transcription is yet to be determined, the authors were able to demonstrate that transcription of the model template containing a single dB contained 1.2 ± 0.4 S ribonucleotides. They also observed no evidence that the unnatural ribotriphosphates were misincorporated opposite a natural nucleotide as the aptamer retained its bulk fluorescence, unlike aptamer produced with rZ incorporated proximal to the fluorophore binding site, which quenched fluorescence despite circular dichroism showing no evidence of structural perturbation.

In addition to its conversion into RNA, what makes natural DNA so special and uniquely able to mediate the heritable storage of information, is its replication by DNA polymerases. Six-letter DNA containing the dP-dZ UBP has been shown to be replicated by both *Taq* and Phusion DNA polymerases in a test tube, even when present in runs of up to four UBPs.¹¹ While the replication of eight-letter DNA remains to be demonstrated, the authors appear optimistic. This optimism is based on the assumption that the natural-like structure of eight-letter DNA will translate into efficient recognition by DNA polymerases. However, a previously reported study from the Georgiadis group revealed that stretches of six dP-dZ UBPs introduce significant duplex distortion, due to stacking of nitro-groups.⁸ While such structural perturbations may complicate replication, this would hardly limit the potential applications of the eight-letter DNA, or for that matter, any other expanded genetic alphabet.

In general, there are three major applications of expanded genetic alphabets being pursued (Figure 2). The first, is the use of the increased hybridization specificity inherent to DNA with more than four letters. For this application, runs of UBPs are not required, and impressively, the Benner group has already developed and commercialized DNA detection methods that take advantage of six-letter DNA for this purpose.^{12,13} Extension to eight-letter DNA promises even more success. The second application is to discover aptamers with increased functionality made available by the unnatural nucleotides. However, the ability to

incorporate consecutive or high-density unnatural nucleotides, via modification of a single nucleotide type, often 5-pyrimidines is already routine. Nonetheless, the real potential advantage of UBPs for aptamer discovery is the incorporation of a single or a few instances of an interesting functionality that better mimic the use of reactive functionality in proteins. The discovery of such aptamers has already been reported by both Hirao and Benner using six-letter DNA,^{14,15} and extension to eight-letter DNA will make the simultaneous use of increased functionality. The final goal is to use the UBPs as the basis of semi-synthetic organisms that store and retrieve increased information – meaning the creation of additional codon-anticodon pairs that can be used to expand the genetic code and expand the repertoire of amino acids that may be used to build proteins.¹⁶ Such unnatural codon-anticodon pairs need only a single UBP, and even runs of unnatural codons would never require adjacent unnatural nucleotides, much less six or more unnatural nucleotides.

This is just the beginning for eight-letter DNA. In addition to determining if its transcription by FAM or any other RNA polymerase proceeds with high fidelity, natural or mutant DNA polymerases must be identified that can mediate its replication. Nonetheless, the conceptual ramifications of what has already been shown should not be underestimated. Uncovering the mysteries behind the molecules of life is one of humanity's earliest pursuits. Are the molecules of life unique or are they just one possible solution? Conclusions that they were unique and in fact perfect were taken as evidence for their intelligent design, or alternately, conclusions that they were at least optimal were taken as evidence of the power of natural selection. But just how unique, perfect, or even optimal the molecules of life actually are has been impossible to gauge because we have had nothing to compare them to. Now for the first time, at least for the sequence specific stability of pairing and for the formation of a conserved duplex structure, we have a strong indication that the molecules that underlie information storage and retrieval are not unique, certainly not perfect, and perhaps only sufficiently optimal. Benner's pioneering work in the late 1980s inspired the rest of us to pursue the expansion of the genetic alphabet, and the conceptual implications of eight-letter DNA now adds to that inspiration.

Acknowledgments

Funding Sources

This work was supported by the National Institutes of Health (GM118178 to F.E.R.).

REFERENCES

- (1). Hoshika S, Leal NA, Kim MJ, Kim MS, Karalkar NB, Kim HJ, Bates AM, Watkins NE Jr., SantaLucia HA, Meyer AJ, DasGupta S, Piccirilli JA, Ellington AD, SantaLucia J Jr., Georgiadis MM, and Benner SA (2019) Hachimoji DNA and RNA: A genetic system with eight building blocks, *Science* 363, 884–887. [PubMed: 30792304]
- (2). Yang Z, Hutter D, Sheng P, Sismour AM, and Benner SA (2006) Artificially expanded genetic information system: a new base pair with an alternative hydrogen bonding pattern, *Nucleic Acids Res* 34, 6095–6101. [PubMed: 17074747]
- (3). Rich A (1962) On the problems of evolution and biochemical information transfer In *Horizons in Biochemistry* (Kasha MPB, Ed.), Academic Press, New York, NY.
- (4). Biondi E, and Benner SA (2018) Artificially expanded genetic information systems for new aptamer technologies, *Biomedicines* 6, 53.

- (5). Kim HJ, Leal NA, and Benner SA (2009) 2'-deoxy-1-methylpseudocytidine, a stable analog of 2'-deoxy-5-methylisocytidine, *Bioorg. Med. Chem* 17, 3728–3732. [PubMed: 19394831]
- (6). Martinot TA, and Benner SA (2004) Artificial genetic systems: exploiting the “aromaticity” formalism to improve the tautomeric ratio for isoguanosine derivatives, *J. Org. Chem* 69, 3972–3975. [PubMed: 15153036]
- (7). SantaLucia J Jr. (2000) The use of spectroscopic techniques in the study of DNA stability, In *Spectrophotometry and Spectrofluorimetry* (Gore MG, Ed.), pp 329–356, Oxford University Press, Oxford, U.K.
- (8). Georgiadis MM, Singh I, Kellett WF, Hoshika S, Benner SA, and Richards NG (2015) Structural basis for a six nucleotide genetic alphabet, *J. Am. Chem. Soc* 137, 6947–6955. [PubMed: 25961938]
- (9). Hernandez AR, Shao Y, Hoshika S, Yang Z, Shelke SA, Herrou J, Kim HJ, Kim MJ, Piccirilli JA, and Benner SA (2015) A crystal structure of a functional RNA molecule containing an artificial nucleobase pair, *Angew. Chem. Int. Ed* 54, 9853–9856.
- (10). Kimoto M, Meyer AJ, Hirao I, and Ellington AD (2017) Genetic alphabet expansion transcription generating functional RNA molecules containing a five-letter alphabet including modified unnatural and natural base nucleotides by thermostable T7 RNA polymerase variants, *Chem. Commun* 53, 12309–12312.
- (11). Yang Z, Chen F, Alvarado JB, and Benner SA (2011) Amplification, mutation, and sequencing of a six-letter synthetic genetic system, *J. Am. Chem. Soc* 133, 15105–15112. [PubMed: 21842904]
- (12). Glushakova LG, Bradley A, Bradley KM, Alto BW, Hoshika S, Hutter D, Sharma N, Yang Z, Kim MJ, and Benner SA (2015) High-throughput multiplexed xMAP Luminex array panel for detection of twenty two medically important mosquito-borne arboviruses based on innovations in synthetic biology, *J. Virol. Methods* 214, 60–74. [PubMed: 25680538]
- (13). Glushakova LG, Sharma N, Hoshika S, Bradley AC, Bradley KM, Yang Z, and Benner SA (2015) Detecting respiratory viral RNA using expanded genetic alphabets and self-avoiding DNA, *Anal. Biochem* 489, 62–72. [PubMed: 26299645]
- (14). Kimoto M, Yamashige R, Matsunaga K, Yokoyama S, and Hirao I (2013) Generation of high-affinity DNA aptamers using an expanded genetic alphabet, *Nat. Biotechnol* 31, 453–457. [PubMed: 23563318]
- (15). Zhang L, Yang Z, Le Trinh T, Teng IT, Wang S, Bradley KM, Hoshika S, Wu Q, Cansiz S, Rowold DJ, McLendon C, Kim MS, Wu Y, Cui C, Liu Y, Hou W, Stewart K, Wan S, Liu C, Benner SA, and Tan W (2016) Aptamers against cells overexpressing glypican 3 from expanded genetic systems combined with cell engineering and laboratory evolution, *Angew. Chem. Int. Ed* 55, 12372–12375.
- (16). Zhang Y, Ptacin JL, Fischer EC, Aerni HR, Caffaro CE, San Jose K, Feldman AW, Turner CR, and Romesberg FE (2017) A semi-synthetic organism that stores and retrieves increased genetic information, *Nature* 551, 644–647. [PubMed: 29189780]

Fig. 1.
The dZ-dP and dS-dB unnatural base pairs.

Fig. 2. Eight-letter DNA. Highlighted in green are the properties and applications demonstrated. Highlighted in orange are the properties and applications yet to be demonstrated.