

SYNTHETIC BIOLOGY

Hachimoji DNA and RNA: A genetic system with eight building blocks

Shuichi Hoshika^{1,2*}, Nicole A. Leal^{1,2*}, Myong-Jung Kim^{1,2}, Myong-Sang Kim¹, Nilesh B. Karalkar^{1,2}, Hyo-Joong Kim¹, Alison M. Bates³, Norman E. Watkins Jr.⁴, Holly A. SantaLucia⁴, Adam J. Meyer⁵, Saurja DasGupta⁶, Joseph A. Piccirilli⁶, Andrew D. Ellington⁵, John SantaLucia Jr.⁴, Millie M. Georgiadis³, Steven A. Benner^{1,2†}

We report DNA- and RNA-like systems built from eight nucleotide “letters” (hence the name “hachimoji”) that form four orthogonal pairs. These synthetic systems meet the structural requirements needed to support Darwinian evolution, including a polyelectrolyte backbone, predictable thermodynamic stability, and stereoregular building blocks that fit a Schrödinger aperiodic crystal. Measured thermodynamic parameters predict the stability of hachimoji duplexes, allowing hachimoji DNA to increase the information density of natural terran DNA. Three crystal structures show that the synthetic building blocks do not perturb the aperiodic crystal seen in the DNA double helix. Hachimoji DNA was then transcribed to give hachimoji RNA in the form of a functioning fluorescent hachimoji aptamer. These results expand the scope of molecular structures that might support life, including life throughout the cosmos.

No behaviors are more central to biology than the storage, transmission, and evolution of genetic information. In modern terran biology, this is achieved by DNA double helices whose strands are joined by regularly sized nucleobase pairs with hydrogen bond complementarity (1). Schrödinger theorized that such regularity in size was necessary for the pairs to fit into an aperiodic crystal, which he proposed to be necessary for reliable molecular information storage and faithful information transfer (2). This feature is also essential for any biopolymer that might support Darwinian evolution, as it ensures that changes in the sequence of the informational building blocks do not damage the performance of the biopolymer, including its interactions with enzymes that replicate it.

Complementary interbase hydrogen bonding has been proposed to be dispensable in Darwinian molecules, provided that size complementarity is retained (3). Thus, hydrophobic nucleotide analogs have been incorporated into duplexes (4), aptamers (5), and living cells (6). These analogs increase the number of genetic letters from four to six. However, pairs lacking interbase hydrogen bonds evidently must be flanked by pairs joined by hydrogen bonds. Further, unless they are constrained by an enzyme active site, hydrophobic pairs can slip atop each other (7), shortening the rung in the DNA ladder, distorting the double

helix, and damaging the aperiodic crystal uniformity of the duplex.

When hydrogen bonding is used to give a third pair, behaviors characteristic of natural DNA are also reproduced (8). Thus, 6-letter DNA alphabets with interpair hydrogen bonds can be copied (9), polymerase chain reaction (PCR)–amplified and sequenced (10, 11), transcribed to 6-letter RNA and back to 6-letter DNA (12), and used to encode proteins with added amino acids (13). Six-letter alphabets with all pairs joined by interbase hydrogen bonds also support Darwinian selection, evolution, and adaptation (14), all hallmarks of the living state.

Here, we tested the limits of molecular information storage that combines Watson-Crick hydrogen bonding with Schrödinger’s requirement for crystal-like uniformity, building an alien genetic system from eight (“hachi”) building block “letters” (“moji”). This required the design of two sets of heterocycles that implement two additional hydrogen-bonding patterns that join two additional pairs (Fig. 1).

We first assessed the regularity and predictability of the thermodynamics of interaction between hachimoji DNA strands. With standard DNA, the energy of duplex formation is not accurately modeled by a single parameter for each base pair. Instead, to make usefully accurate predictions of duplex stability, predictive tools must account for sequence context (15). With standard DNA, this is done by obtaining nearest-neighbor thermodynamic parameters for all base pair dimers (BPDs) (15). Parameters are often added to account for the decrease in translational degrees of freedom when two strands become one duplex, and to specially treat the distinctively weak A:T pair at the ends of duplexes.

If context dependence is similar in hachimoji DNA, tools that make usefully accurate predictions should also require parameters for all BPDs

for an 8-letter alphabet. Of course, with eight building blocks instead of four, hachimoji DNA has many more BPDs to parameterize. After accounting for symmetry (e.g., AC/TG is equivalent to GT/CA), 40 parameters are required (28 more than the 12 required for standard DNA). These comprise 36 added BPDs plus four for pairs initiated with terminal G:C and terminal effects for A:T, S:B, and Z:P, where “S” is 3-methyl-6-amino-5-(1′-β-D-2′-deoxyribofuranosyl)-pyrimidin-2-one, “B” is 6-amino-9[(1′-β-D-2′-deoxyribofuranosyl)-4-hydroxy-5-(hydroxymethyl)-oxolan-2-yl]-1H-purin-2-one, “Z” is 6-amino-3-(1′-β-D-2′-deoxyribofuranosyl)-5-nitro-1H-pyridin-2-one, and “P” is 2-amino-8-(1′-β-D-2′-deoxyribofuranosyl)-imidazo-[1,2a]-1,3,5-triazin-[8H]-4-one.

To obtain these additional parameters, we designed 94 hachimoji duplexes (table S1) with standard nucleotides A, T, G, and C, purine analogs P and B, and pyrimidine analogs Z and S (Fig. 1). If the design is successful, these duplexes should be joined by P:Z and B:S pairs in addition to standard G:C and A:T pairs. The paired hachimoji DNA oligonucleotides were synthesized by solid-phase chemistry from synthetic phosphoramidites, assembled, and melted to collect thermodynamic data. These data were processed with Meltwin v.3.5 (16) to obtain a parameter set using both the T_m^{-1} versus $\ln(Ct)$ method and the Marquardt nonlinear curve fit method. The error-weighted average of the values from the two methods yielded the ΔG°_{37} and ΔH° values for the 94 duplexes (17).

This analysis allowed us to determine the 28 additional parameters for the 8-letter genetic system using singular value decomposition methods (tables S4, S7, and S10 and figs. S1, S3, and S5). Because this number of measurements over-determines the unknown parameters by a factor of 3.3, we could test the applicability of the BPD model using error propagation to derive standard deviations in the derived parameters (17). The parameters and standard deviations are in figs. S1, S3, and S5. A cross-validation approach gave the same result, as expected given the overdetermination.

The resulting parameters proved to usefully predict melting temperatures for hachimoji DNA. Plots of experimental versus predicted free energy changes and experimental versus predicted melting temperatures (Fig. 2) show that on average, T_m is predicted to within 2.1°C for the 94 GACTZPSB hachimoji duplexes, and ΔG°_{37} is predicted to within 0.39 kcal/mol (tables S3, S6, and S9). These errors are similar to those observed with nearest-neighbor parameters for standard DNA:DNA duplexes (15). Thus, GACTZPSB hachimoji DNA reproduces, in expanded form, the molecular recognition behavior of standard 4-letter DNA. It is an informational system.

We then asked whether hachimoji DNA might be mutable without damaging the Schrödinger aperiodic crystal required to support mutability and Darwinian evolution. High-resolution crystal structures were determined for three different hachimoji duplexes assembled from three self-complementary hachimoji duplexes (16-mers):

¹Firebird Biomolecular Sciences LLC, 13709 Progress Boulevard, No. 17, Alachua, FL 32615, USA. ²Foundation for Applied Molecular Evolution, 13709 Progress Boulevard, No. 7, Alachua, FL 32615, USA. ³Department of Biochemistry and Molecular Biology, Indiana University School of Medicine, Indianapolis, IN 46202, USA. ⁴DNA Software Inc., Ann Arbor, MI 48104, USA. ⁵Center for Systems and Synthetic Biology, University of Texas, Austin, TX 78703, USA. ⁶Department of Biochemistry and Molecular Biology and Department of Chemistry, University of Chicago, Chicago, IL 60637, USA. *These authors contributed equally to this work. †Corresponding author. Email: manuscripts@ffame.org

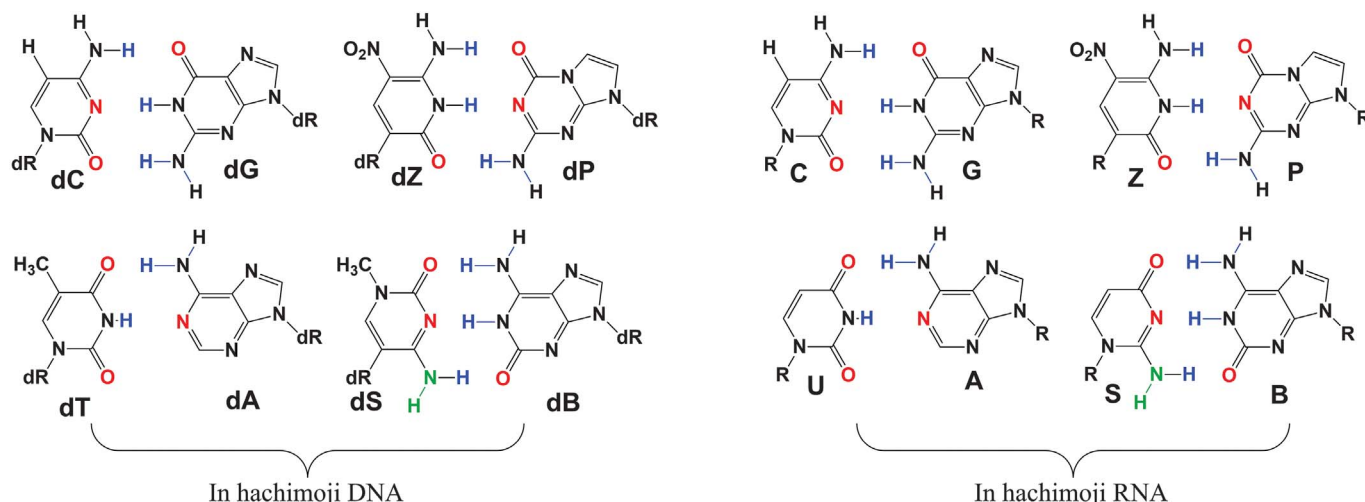


Fig. 1. The eight nucleotides of hachimoji DNA and hachimoji RNA are designed to form four size- and hydrogen bond-complementary pairs.

Hydrogen bond donor atoms involved in pairing are blue; hydrogen bond acceptor atoms are red. The left two pairs in each set are formed from the four standard nucleotides (note missing hydrogen-bonding group in the A:T pair, a peculiarity of standard terran DNA and RNA). The right two pairs in each set are formed from the four new nonstandard nucleotides. Notice the absence of electron density in the minor groove of S, which has a NH (green) moiety.

5'-CTTATPBTASZATAAG ("PB," 1.7 Å; PDB ID 6MIG), 5'-CTTAPCBTASGZTAAG ("PC," 1.6 Å; PDB ID 6MIH), and 5'-CTTATPPSBZZATAAG ("PP," 1.7 Å; PDB ID 6MIK). These duplexes were crystallized with Moloney murine leukemia virus reverse transcriptase to give a "host-guest" complex with two protein molecules (host) bound to each end of a 16-mer duplex (guest) (Fig. 3) (18). With interactions between the host and guest limited to the ends, the intervening 10 base pairs were free to adopt a sequence-dependent structure (Fig. 3A) (19).

The hachimoji DNA in all three structures adopted a B-form (Fig. 3, B to E) with 10.2 to 10.4 base pairs per turn, as analyzed by 3DNA (20). The major and minor groove widths for hachimoji DNA were similar to one another and to GC DNA (5'-CTTATGGGCCATAAG), but not AT DNA (5'-CTTATAAAATTTATAAG) (fig. S16). For nucleotide pair parameters, the S:B pairs at position 7 in both the PC and PB structures exhibited very similar propeller angles but slightly greater opening angles as compared to G:C in the same position. The P:Z pairs adjacent to natural pairs exhibited larger buckle angles relative to G:C pairs (Fig. 3, F to I).

Even with these differences, the structural parameters for the individual pairs and the dinucleotide steps of the hachimoji DNA fall well within the ranges observed for natural 4-letter DNA. Thus, it appears that hachimoji DNA meets the Schrödinger requirement for a Darwinian system, forming essentially the same "aperiodic crystal" regardless of the sequences. It is a mutable information storage system.

With the information storage and mutability properties shown for hachimoji DNA, we asked whether hachimoji information could also be transmitted, here to give hachimoji GACUZPSB RNA, where "S" is 2-amino-1-(1'-β-D-ribofuranosyl)-4(1H)-pyrimidinone, "B" is 6-amino-9[(1'-β-D-

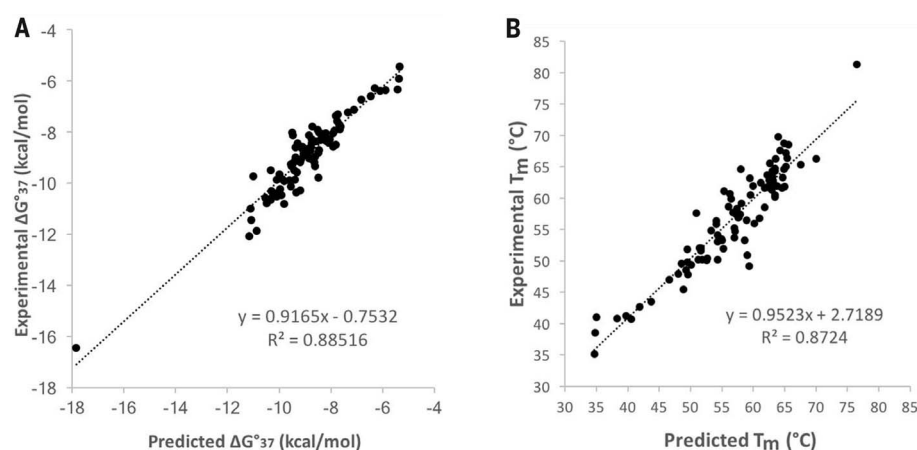


Fig. 2. Thermodynamics of hachimoji DNA duplexes. (A) Plot of experimental versus predicted free energy changes (ΔG°_{37}) for 94 SBZP-containing hachimoji DNA duplexes (tables S3, S6, and S9). (B) Plot of experimental versus predicted melting temperatures T_m of 94 SBZP-containing hachimoji DNA duplexes (tables S3, S6, and S9). The outlier is a sequence embedded in the PP guest (Fig. 3E).

ribofuranosyl)-4-hydroxy-5-(hydroxymethyl)-oxolan-2-yl]-1H-purin-2-one, "Z" is 6-amino-3-(1'-β-D-ribofuranosyl)-5-nitro-1H-pyridin-2-one, and "P" is 2-amino-8-(1'-β-D-ribofuranosyl)-imidazo-[1,2a]-1,3,5-triazin-[8H]-4-one. To develop RNA polymerases able to transcribe hachimoji DNA, we started with four model sequences that each contained a single nonstandard hachimoji component, B, P, S, or Z, each followed by a single cytidine. To analyze hachimoji RNA products, we labeled transcripts with [α - 32 P]cytidine 5'-triphosphate; digestion with ribonuclease T2 then generated the corresponding hachimoji 3'-phosphates (27). These were resolved in thin-layer chromatography (TLC) systems and compared with synthetic authentic nonstandard 3'-phosphates.

These experiments showed that native T7 RNA polymerase incorporates riboZTP opposite tem-

plate dP, riboPTP opposite template dZ, and riboBTP opposite template dS (figs. S11 and S12). However, incorporation of riboSTP opposite template dB was not seen with native RNA polymerase (22). This was attributed to an absence of electron density in the minor groove from the aminopyridone heterocycle on riboSTP (Fig. 1); polymerases are believed to recognize such density, which is presented by all other triphosphate substrates.

We therefore searched for T7 RNA polymerase variants able to transcribe a complete set of hachimoji nucleotides (table S11). One variant (Y639F H784A P266L, "FAL") was especially effective at incorporating riboSTP opposite template dB (figs. S13 and S14). FAL was originally developed as a thermostable polymerase to accept 2'-O-methyl triphosphates (23). High-performance

Fig. 3. Crystal structures of PB, PC, and PP hachimoji DNA.

(A) The host-guest complex with two N-terminal fragments from Moloney murine leukemia virus reverse transcriptase (in green and cyan) bound to a 16-mer PP hachimoji DNA; in the duplex sphere model, Z:P pairs are green and S:B pairs are magenta. The asymmetric unit includes one protein molecule and half of the 16-mer DNA, as indicated by the line. (B) Hachimoji DNA structures PB (green), PC (red), and PP (blue) are superimposed with GC DNA (gray). (C) Structure of hachimoji DNA with self-complementary duplex 5'-CTTATPBTASZATAAG ("PB"). (D) Structure of hachimoji DNA with self-complementary duplex 5'-CTTAPCBTASGZTAAG ("PC"). (E) Structure of hachimoji DNA with self-complementary duplex with six consecutive nonstandard 5'-CTTATPPSBZZATAAG (PP) components. DNA structures are shown as stick models with P:Z pairs (carbon atoms, green), B:S pairs (carbon atoms, magenta), and natural pairs (carbon atoms, gray). (F to I) Examples of largest differences in detailed structures. The Z:P pair from the PB structure (F) is more buckled than the corresponding G:C pair (G). The S:B pair from the PB structure (H) exhibits a propeller angle similar to that in the corresponding G:C pair (I).

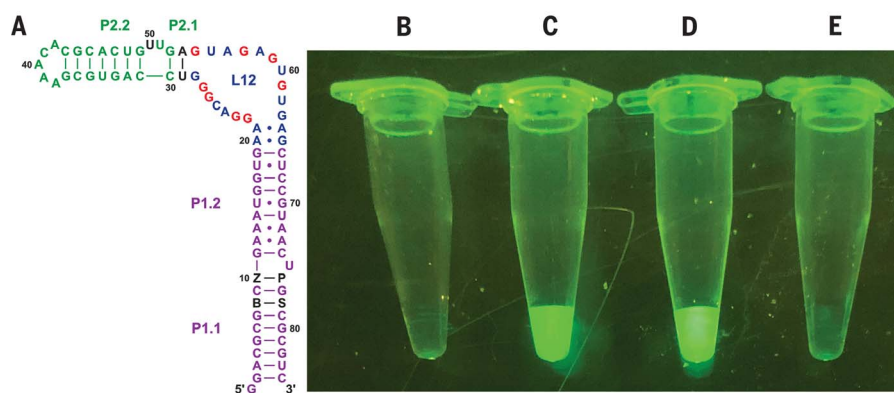
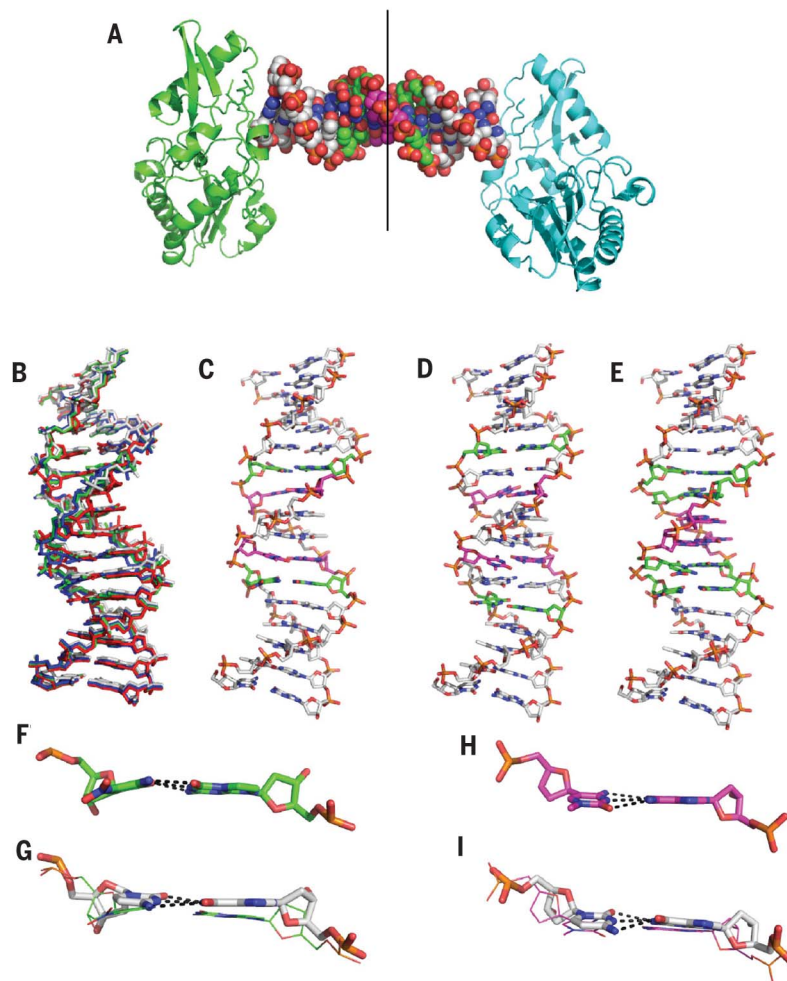


Fig. 4. Structure and fluorescent properties of hachimoji RNA molecules. (A) Schematic showing the full hachimoji spinach variant aptamer; additional nucleotide components of the hachimoji system are shown as black letters at positions 8, 10, 76, and 78 (B, Z, P, and S, respectively). The fluor binds in loop L12 (25). (B to E) Fluorescence of various species in equal amounts as determined by UV. Fluorescence was visualized under a blue light (470 nm) with an amber (580 nm) filter. (B) Control with fluor only, lacking RNA. (C) Hachimoji spinach with the sequence shown in (A). (D) Native spinach aptamer with fluor. (E) Fluor and spinach aptamer containing Z at position 50, replacing the A:U pair at positions 53:29 with G:C to restore the triple observed in the crystal structure. This places the quenching Z chromophore near the fluor; CD spectra suggest that this variant had the same fold as native spinach (fig. S8).

liquid chromatography (HPLC) analysis of its transcripts showed that 1.2 ± 0.4 riboSTP nucleotides were incorporated opposite a single template dB (fig. S9). FAL also incorporated the other non-standard components of the hachimoji system into transcripts (figs. S13 and S14).

We then designed a hachimoji variant of the spinach fluorescent RNA aptamer (24). In its standard form, spinach folds and binds 3,5-difluoro-4-hydroxybenzylidene imidazolinone, which fluoresces green when bound. An analogous hachimoji aptamer with nonstandard nucleotides placed strategically to avoid disrupting its fold (Fig. 4) was prepared by transcribing hachimoji DNA using the FAL variant of RNA polymerase. The aptamer's sequence was confirmed by label transfer experiments; incorporation of riboZ was further confirmed by HPLC and ultraviolet (UV) spectroscopy.

The hachimoji spinach fluoresced green (Fig. 4). As a control, a spinach variant was prepared with a Z incorporated at position 50, near enough to the bound fluor to quench its fluorescence. That variant did not fluoresce, even though circular dichroism (CD) experiments (fig. S8) suggested

that its overall fold was undisturbed. These results precluded the possibility that nonstandard hachimoji components are generally misincorporated throughout the aptamer, as these would disrupt the fold needed for the fluor to bind or place a quenching riboZ nucleotide near enough to loop L12 to eliminate its fluorescence.

This synthetic biology makes available a mutable genetic system built from eight different building blocks. With increased information density over standard DNA and predictable duplex stability across (evidently) all 8^n sequences of length n , hachimoji DNA has potential applications in bar-coding and combinatorial tagging, retrievable information storage, and self-assembling nanostructures. The structural differences among three different hachimoji duplexes are not larger than the differences between various standard DNA duplexes, making this system potentially able to support molecular evolution. Further, the ability to have structural regularity independent of sequence shows the importance of interbase hydrogen bonding in such mutable informational systems. Thus, in addition to its technical applications, this work expands the scope of the structures that we might encounter as we search for life in the cosmos.

REFERENCES AND NOTES

1. J. D. Watson, F. H. C. Crick, *Nature* **171**, 737–738 (1953).
2. E. Schrödinger, *Was ist Leben?* (Piper, 1943).
3. M. F. Goodman, *Nat. Biotechnol.* **17**, 640–641 (1999).
4. S. Moran, R. X.-F. Ren, S. Rumney IV, E. T. Kool, *J. Am. Chem. Soc.* **119**, 2056–2057 (1997).
5. M. Kimoto, R. Yamashige, K. Matsunaga, S. Yokoyama, I. Hirao, *Nat. Biotechnol.* **31**, 453–457 (2013).
6. Y. Zhang *et al.*, *Nature* **551**, 644–647 (2017).
7. K. Betz *et al.*, *Nat. Chem. Biol.* **8**, 612–614 (2012).
8. S. A. Benner *et al.*, *Cold Spring Harb. Perspect. Biol.* **8**, a023770 (2016).
9. Z. Yang, A. M. Sismour, P. Sheng, N. L. Puskar, S. A. Benner, *Nucleic Acids Res.* **35**, 4238–4249 (2007).
10. A. M. Sismour *et al.*, *Nucleic Acids Res.* **32**, 728–735 (2004).
11. Z. Yang, F. Chen, J. B. Alvarado, S. A. Benner, *J. Am. Chem. Soc.* **133**, 15105–15112 (2011).
12. N. A. Leal *et al.*, *ACS Synth. Biol.* **4**, 407–413 (2015).
13. J. D. Bain, C. Switzer, A. R. Chamberlin, S. A. Benner, *Nature* **356**, 537–539 (1992).
14. L. Zhang *et al.*, *J. Am. Chem. Soc.* **137**, 6734–6737 (2015).
15. J. SantaLucia Jr., *Proc. Natl. Acad. Sci. U.S.A.* **95**, 1460–1465 (1998).
16. J. A. McDowell, D. H. Turner, *Biochemistry* **35**, 14077–14089 (1996).
17. J. SantaLucia Jr., in *Spectrophotometry and Spectrofluorimetry: A Practical Approach*, M. G. Gore, Ed. (Oxford Univ. Press, 2000), pp. 329–356.
18. M. L. Coté, S. J. Yohannan, M. M. Georgiadis, *Acta Crystallogr. D* **56**, 1120–1131 (2000).
19. S. P. Montañó, M. L. Coté, M. J. Roth, M. M. Georgiadis, *Nucleic Acids Res.* **34**, 5353–5360 (2006).
20. X. J. Lu, W. K. Olson, *Nucleic Acids Res.* **31**, 5108–5121 (2003).
21. I. Hirao *et al.*, *Nat. Biotechnol.* **20**, 177–182 (2002).
22. C. Y. Switzer, S. E. Moroney, S. A. Benner, *Biochemistry* **32**, 10489–10496 (1993).
23. A. J. Meyer *et al.*, *Nucleic Acids Res.* **43**, 7480–7488 (2015).
24. J. S. Paige, K. Y. Wu, S. R. Jaffrey, *Science* **333**, 642–646 (2011).
25. H. Huang *et al.*, *Nat. Chem. Biol.* **10**, 686–691 (2014).

ACKNOWLEDGMENTS

We thank Accellero Biostructures Inc. for their rapid, high-quality x-ray diffraction data collection services. **Funding:** Supported by grants from NASA under award NNX15AF46G, the National Institute of General Medical Sciences (R41GM119494, R01GM128186, R01GM102489), and NSF under award CHE 1507816. This publication also was made possible through the support of a grant from the Templeton World Charity Foundation Inc. (0092/AB57). Use of the Stanford Synchrotron Radiation Lightsource (SLAC National Accelerator Laboratory) was supported by the U.S. Department of Energy (DOE) Office of Science, Office of Basic Energy Sciences under contract DE-AC02-76SF00515. The SSRL Structural Molecular Biology Program is supported by the DOE Office of Biological and Environmental Research and by the National Institute of General Medical Sciences (including P41GM103393). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the

authors and do not necessarily reflect the views of the NIH, NSF, DOE, NASA, or TWCF. **Author contributions:** S.H. synthesized and purified all of the hachimoji oligonucleotides and the hachimoji 3'-phosphates, and synthesized hachimoji RNA derivatives needed to support the label shift analyses. N.A.L. performed all the studies involving enzymatic reactions, including preparation of various RNA transcripts, and developed the hachimoji RNA analytical chemistry, including its 2D-TLC strategies. M.-J.K., M.-S.K., N.B.K., and H.-J.K. synthesized all of the hachimoji phosphoramidites and triphosphates. N.E.W., H.A.S., and J.S. designed the hachimoji DNA oligonucleotides, performed the melting temperature studies, and interpreted the melting temperature data. S.D. and J.A.P. performed the biophysical studies on the Z-variant of the spinach aptamer. A.M.B. and M.M.G. performed all of the crystallographic studies and analyzed the three crystal structures containing hachimoji DNA. A.J.M. and A.D.E. prepared variants of T7 RNA polymerase. M.M.G., J.S., and S.A.B. further directed the research and prepared the manuscript with the help of the other co-authors. **Competing interests:** S.A.B., N.A.L., S.H., and the Foundation for Applied Molecular Evolution have a financial interest in the intellectual property in the area of expanded genetic alphabets. S.A.B. owns Firebird Biomolecular Sciences, which makes various hachimoji reagents available for sale N.E.W., H.A.S., and J.S. are owners of DNA Software Inc., which owns intellectual property and software unrelated to this paper that makes thermodynamic predictions concerning nucleotide binding. A.D.E. has a financial interest in the T7 RNA polymerase variants used here. A.D.E. and A.J.M. are inventors on patent application US 15/127,617 held by the Board of Regents of the University of Texas System, which covers "T7 RNA polymerase variants with expanded substrate range and enhanced transcriptional yield." S.A.B. and N.A.L. filed patent application 16226963 "Enzymatic Processes for Synthesizing RNA Containing Certain Non-Standard Nucleotides" on 20 December 2018. **Data and materials availability:** Coordinates are deposited as PDB ID 6MIG, PDB ID 6MIH, and PDB ID 6MIK.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/363/6429/884/suppl/DC1
Materials and Methods
Figs. S1 to S16
Tables S1 to S13
References (26–39)

23 January 2018; resubmitted 25 September 2018
Accepted 31 January 2019
10.1126/science.aat0971

Hachimoji DNA and RNA: A genetic system with eight building blocks

Shuichi Hoshika, Nicole A. Leal, Myong-Jung Kim, Myong-Sang Kim, Nilesch B. Karalkar, Hyo-Joong Kim, Alison M. Bates, Norman E. Watkins Jr., Holly A. SantaLucia, Adam J. Meyer, Saurja DasGupta, Joseph A. Piccirilli, Andrew D. Ellington, John SantaLucia Jr., Millie M. Georgiadis and Steven A. Benner

Science **363** (6429), 884-887.
DOI: 10.1126/science.aat0971

Expanding the genetic code

DNA and RNA are naturally composed of four nucleotide bases that form hydrogen bonds in order to pair. Hoshika *et al.* added an additional four synthetic nucleotides to produce an eight-letter genetic code and generate so-called hachimoji DNA. Coupled with an engineered T7 RNA polymerase, this expanded DNA alphabet could be transcribed into RNA. Thus, new forms of DNA that add information density to genetic biopolymers can be generated that may be useful for future synthetic biological applications.

Science, this issue p. 884

ARTICLE TOOLS

<http://science.sciencemag.org/content/363/6429/884>

SUPPLEMENTARY MATERIALS

<http://science.sciencemag.org/content/suppl/2019/02/20/363.6429.884.DC1>

REFERENCES

This article cites 37 articles, 3 of which you can access for free
<http://science.sciencemag.org/content/363/6429/884#BIBL>

PERMISSIONS

<http://www.sciencemag.org/help/reprints-and-permissions>

Use of this article is subject to the [Terms of Service](#)