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 Kim[1], Nilesh B. Karalkar[1,2], Hyo-Joong Kim[1], Alison M. Bates[3], Norman E. Watkins Jr.[4], Holly A. SantaLucia[4], Adam J. Meyer[5], Saurja DasGupta[6], Joseph A. Piccirilli[6], Andrew D. Ellington[5], John SantaLucia Jr.[4], Millie M. Georgiadis[3], 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 overdetermines 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):

[1]Firebird Biomolecular Sciences LLC, 13709 Progress Boulevard, No. 17, Alachua, FL 32615, USA. [2]Foundation for Applied Molecular Evolution, 13709 Progress Boulevard, No. 7, Alachua, FL 32615, USA. [3]Department of Biochemistry and Molecular Biology, Indiana University School of Medicine, Indianapolis, IN 46202, USA. [4]DNA Software Inc., Ann Arbor, MI 48104, USA. [5]Center for Systems and Synthetic Biology, University of Texas, Austin, TX 78703, USA. [6]Department 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