Bioengineered 5:3, 214–215; July/August 2014; © 2014 Landes Bioscience

Genetics just got SEXY

Sequences encoding XY

Roy D Sleator

Department of Biological Sciences; Cork Institute of Technology; Cork, Ireland

A semi-synthetic organism with an extended genetic alphabet heralds a new era in synthetic biology.

In the May 15th 2014 issue of Nature,[1] Floyd E Romesberg and colleagues unveiled the first semi-synthetic organism with an expanded genetic alphabet; heralding a new era in synthetic biology.[2]

Since the dawn of life on earth, all genetic information has been coded by just 5 natural bases (Adenine [A], Guanine (G], Cytosine [C], and Thymine [T], which is replaced with Uracil [U] in RNA).[3] However, it wasn't until the 1960s, a decade after Watson and Crick revealed that the structure of DNA[4] is mediated by the specific pairing of A to T and G to C, that scientists began to consider the possibilities of incorporating synthetic analogs of the natural bases as additional functional pairs.[5] Almost 30 years later, Steven Benner, a pioneer of synthetic biology, described the incorporation of non-standard amino acids into polypeptides by ribosome-based translation.[6] Benner's team described the expansion of the genetic code through the creation of a 65th codon-anticodon pair from unnatural base pairs (UBPs) exhibiting non-standard hydrogen-bonding patterns. This non-canonical codon-anticodon pair supported translation in vitro to yield peptides containing a non-standard amino acid. Schweitzer and Kool[7] extended these findings, showing that hydrogen bonds are not necessarily required for the formation of a stable duplex DNA-like structure, suggesting that hydrogen bonds are more important for specificity of pairing than for affinity in DNA. The ability to substitute hydrogen bonding with hydrophobic or indeed steric interactions enabled the design of increasingly more versatile nucleotide analogs.[8-10] An added advantage of the differing chemistries (i.e., hydrophobic interactions vs. hydrogen bonding) is that it prevents the natural and synthetic nucleotides from mispairing with one another. This repartitioning could potentially be used to design hybrid systems: semi-synthetic cells operating two separate genetic codes. The native code would run normal cellular processes, while the parallel synthetic code would allow the cell to act as a micro-factory producing novel proteins.

In 2012, Romesberg and colleagues,[8] having synthesized and tested more than 300 artificial nucleotides, developed a class of UBPs, exemplified by d5SICS-dNaM (abbreviated as X and Y), formed between nucleotides bearing hydrophobic nucleobases. In vitro testing, involving PCR and PCR-based applications, found d5SICS-dNaM to be functionally equivalent to natural base pairs. However, while replication of UBPs in vitro is one thing, in vivo replication is a completely different proposition. Until now, the major stumbling block in moving from in vitro synthesis to in vivo replication was in getting the synthetic nucleotide analogs inside the bacterial cell in the first place. While DNA integrated nucleotides contain a single phosphate group, they require two additional phosphates before being incorporated into replicating DNA. Removal of the extra phosphates is achieved via a dephosphorylation reaction, providing the energy to power the replication process. Passive diffusion of free nucleosides into the cell, followed by conversion to the corresponding triphosphate via the nucleoside pathway, proved inefficient.[11] Given that in vivo analog assembly from precursors was unlikely to work, at least in the short-term, Romesberg's group circumvented this step by engineering the host cell to accumulate the pre-formed nucleoside triphosphates. This simple fix elegantly incorporates a failsafe bio-containment measure; because the *E. coli* host cannot synthesize the analogs, they need to be added exogenously. In the absence of unnatural triphosphates in the growth medium, the UBP is eventually lost. Furthermore, this loss was found to be the result of replication-mediated mispairing as opposed to the activity of DNA repair pathways.[1]

The researchers exploited the fact that certain intracellular bacteria and algal plastids fail to synthesize their own nucleotides, instead accumulating them from the external environment via nucleotide triphosphate transporters (NTTs).[12,13] Of 8 NTTs tested, the *Phaeodactylum tricornutum* NTT (*Pt*NTT2)[14] proved most effective when heterologously expressed against an *Escherichia coli* host, resulting in 90 μM d5SICSTP and 30 μM dNaMTP in the cell cytoplasm 30 min after addition to the growth media at a concentration of 0.25 mM. In addition to the *Pt*NTT2 encoding plasmid, named pACS (for accessory plasmid), the *E. coli* host also harbors pINF (the information plasmid). Based on the pUC19 backbone, the in vitro synthesized pINF harbors a dTPT3-dNaM pair in place of dA-dT at position 505, 362 bp downstream of the ColE1 origin of replication and within the TK-1 Okazaki processing site,[15] where DNA replication is mediated by DNA Pol I (previously shown to efficiently replicate DNA containing d5SICS-dNaM in vitro[16]). The use of dTPT3,[17] an analog of d5SICS, in the in vitro synthesized pINF provided an efficient means of confirming in vivo replication; replacement of dTPT3 with d5SICS, supplied exogenously in

Correspondence to: Roy D Sleator; Email: roy.sleator@cit.ie
Published Online: 05/21/2014
http://dx.doi.org/10.4161/bioe.29306

the medium, only occurs if the plasmid is replicated in vivo. Indeed, Romesberg's team found that DNA containing the UBP is replicated in vivo with at least 99.4% fidelity[1]; corresponding to an error rate of ~10^{-3}, equivalent to that seen in natural DNA.[18]

Having proved that UBPs can be readily incorporated and replicated in DNA, the next step is to demonstrate that they can be transcribed into RNA in vivo. Such modified RNAs might well lead to improved functional RNA elements such as riboswitches, ribozymes, and ribonucleoproteins[5]—an entire suite of new synthetic biology tools. However, perhaps the most exciting aspect of this research is the ability to artificially extend our genetic alphabet. While the natural genetic code builds proteins with just 20 amino acids building blocks,[19] Romesberg's expanded code incorporates up to 172 amino acids, circumventing the need to recode the translational functions of existing codons as proffered by Lajoie et al.[20]

The obvious commercial potential of the research is not lost on Romesberg who is co-founder of Synthorx (http://www.synthorx.com), a San Diego based biotech company, the launch of which coincided with the publication of the Malyshev et al., paper.[1] A stated aim of Synthorx is to exploit the potential of UBPs to improve the discovery and development of new drugs, diagnostics, and vaccines as well as creating innovative products, including research reagents, aptamers, and nanomaterials.

A mere 53 years after Nirenberg and Matthaei's pioneering work to crack the genetic code,[21] we stand poised to write a new one. In Romesberg's own words; the increased information offered by this newly expanded code will allow us "to write more interesting words, bigger words, more complicated words, more nuanced words, better stories."

Disclosure of Potential Conflicts of Interest

No potential conflict of interest was disclosed.

Acknowledgments

R.D.S. is coordinator of the EU FP7 project ClouDx-i.

References

1. Malyshev DA, Dhami K, Lavergne T, Chen T, Dai N, Foster JM, Corrêa IR Jr., Romesberg FE. A semi-synthetic organism with an expanded genetic alphabet. Nature 2014; 509:385-8; PMID:24805238; http://dx.doi.org/10.1038/nature13314
2. Sleator RD. The synthetic biology future. Bioengineered 2014; 5:5; PMID:24561910; http://dx.doi.org/10.4161/bioe.28317
3. Di Giulio M. The origin of the genetic code: matter of metabolism or physicochemical determinism? J Mol Evol 2013; 77:131-3; PMID:24162920; http://dx.doi.org/10.1007/s00239-013-9593-9
4. Watson JD, Crick FH. Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. Nature 1953; 171:737-8; PMID:13054692; http://dx.doi.org/10.1038/171737a0
5. Thyer R, Ellefson J. Synthetic biology: New letters for life's alphabet. Nature 2014; 509:291-2; PMID:24805244; http://dx.doi.org/10.1038/nature13335
6. Bain JD, Switzer C, Chamberlin AR, Benner SA. Ribosome-mediated incorporation of a non-standard amino acid into a peptide through expansion of the genetic code. Nature 1992; 356:537-9; PMID:1560827; http://dx.doi.org/10.1038/356537a0
7. Schweitzer BA, Kool ET. Hydrophobic, Non-Hydrogen-Bonding Bases and Base Pairs in DNA. J Am Chem Soc 1995; 117:1863-72; PMID:20882111; http://dx.doi.org/10.1021/ja00112a001
8. Malyshev DA, Dhami K, Quach HT, Lavergne T, Ordoukhanian P, Torkamani A, Romesberg FE. Efficient and sequence-independent replication of DNA containing a third base pair establishes a functional six-letter genetic alphabet. Proc Natl Acad Sci U S A 2012; 109:12005-10; PMID:22773812; http://dx.doi.org/10.1073/pnas.1205176109
9. Yamashige R, Kimoto M, Takezawa Y, Sato A, Mitsui T, Yokoyama S, Hirao I. Highly specific unnatural base pair systems as a third base pair for PCR amplification. Nucleic Acids Res 2012; 40:2793-806; PMID:22121213; http://dx.doi.org/10.1093/nar/gkr1068
10. Yang Z, Chen F, Alvarado JB, Benner SA. Amplification, mutation, and sequencing of a six-letter synthetic genetic system. J Am Chem Soc 2011; 133:15105-12; PMID:21842904; http://dx.doi.org/10.1021/ja204910n
11. Wu Y, Fa M, Tae EL, Schultz PG, Romesberg FE. Enzymatic phosphorylation of unnatural nucleosides. J Am Chem Soc 2002; 124:14626-30; PMID:12465973; http://dx.doi.org/10.1021/ja028050m
12. Amiri H, Karlberg O, Andersson SG. Deep origin of plastid/parasite ATP/ADP translocases. J Mol Evol 2003; 56:137-50; PMID:12574860; http://dx.doi.org/10.1007/s00239-002-2387-0
13. Horn M, Wagner M. Bacterial endosymbionts of free-living amoebae. J Eukaryot Microbiol 2004; 51:509-14; PMID:15537084; http://dx.doi.org/10.1111/j.1550-7408.2004.tb00278.x
14. Ast M, Gruber A, Schmitz-Esser S, Neuhaus HE, Kroth PG, Horn M, Haferkamp I. Diatom plastids depend on nucleotide import from the cytosol. Proc Natl Acad Sci U S A 2009; 106:3621-6; PMID:19221027; http://dx.doi.org/10.1073/pnas.0808862106
15. Allen JM, Simcha DM, Ericson NG, Alexander DL, Marquette JT, Van Biber BP, Troll CJ, Karchin R, Bielas JH, Loeb LA, et al. Roles of DNA polymerase I in leading and lagging-strand replication defined by a high-resolution mutation footprint of ColE1 plasmid replication. Nucleic Acids Res 2011; 39:7020-33; PMID:21622658; http://dx.doi.org/10.1093/nar/gkr157
16. Seo YJ, Matsuda S, Romesberg FE. Transcription of an expanded genetic alphabet. J Am Chem Soc 2009; 131:5046-7; PMID:19351201; http://dx.doi.org/10.1021/ja9006996
17. Li L, Degardin M, Lavergne T, Malyshev DA, Dhami K, Ordoukhanian P, Romesberg FE. Natural-like replication of an unnatural base pair for the expansion of the genetic alphabet and biotechnology applications. J Am Chem Soc 2014; 136:826-9; PMID:24152106; http://dx.doi.org/10.1021/ja408814g
18. Goodman MF. Error-prone repair DNA polymerases in prokaryotes and eukaryotes. Annu Rev Biochem 2002; 71:17-50; PMID:12045089; http://dx.doi.org/10.1146/annurev.biochem.71.083101.124707
19. Sleator RD. Proteins: form and function. Bioeng Bugs 2012; 3:80-5; PMID:22095055; http://dx.doi.org/10.4161/bbug.18303
20. Lajoie MJ, Rovner AJ, Goodman DB, Aerni HR, Haimovich AD, Kuznetsov G, Mercer JA, Wang HH, Carr PA, Mosberg JA, et al. Genomically recoded organisms expand biological functions. Science 2013; 342:357-60; PMID:24136966; http://dx.doi.org/10.1126/science.1241459
21. Nirenberg MW, Matthaei JH. The dependence of cell-free protein synthesis in E. coli upon naturally occurring or synthetic polyribonucleotides. Proc Natl Acad Sci U S A 1961; 47:1588-602; PMID:14479932; http://dx.doi.org/10.1073/pnas.47.10.1588