# Synthetic Biological Approaches for RNA Labelling and Imaging: Design Principles and Future Opportunities

Steven Pauff,[1] Jamie M. Withers,[1] Iain J. W. McKean,[1] Simon P. Mackay,[2] Glenn A. Burley[1*]

[1] *Department of Pure and Applied Chemistry, University of Strathclyde. Thomas Graham Building, 295 Cathedral Street, Glasgow, United Kingdom. G1 1XL. Email: glenn.burley@strath.ac.uk*

[2] *Strathclyde Institute of Pharmacy and Biomedical Science, University of Strathclyde. 161 Cathedral Street, Glasgow, United Kingdom. G4 0RE.*

## ABSTRACT

RNA is the most mercurial of all biomacromolecules. In contrast to DNA, where the predominant role is the storage of genetic information, the biological role of RNA varies; ranging from a template-based intermediary in gene expression to playing a direct role in catalysis. Their high turnover and metabolic lability makes the detection of specific sequences particularly challenging. This review describes the latest synthetic biological developments that enable the direct imaging of RNA both *in vitro* and in their native cellular environment.

This is a peer-reviewed, accepted author manuscript of the following research output: Pauff, S., Withers, J. M., McKean, I. J. W., Mackay, S. P., & Burley, G. A. (2017). Synthetic biological approaches for RNA labelling and imaging: design principles and future opportunities. *Current Opinion in Biotechnology*, 48, 153-158. DOI: 10.1016/j.copbio.2017.04.003

## INTRODUCTION

RNA is a highly dynamic biomacromolecule that exhibits a diverse range of biological functions. In contrast to DNA, where the predominant role is the storage of genetic information, the biological function of RNA varies; ranging from a template-based intermediary in gene expression to playing a direct role in catalysis.[1-4] As a reflection of these diverse roles, RNA is both structurally and spatiotemporally dynamic which, when combined with its varied levels of expression, and in many cases, rapid rates of turnover, poses significant challenges in reporting the synthesis, processing and trafficking of specific RNA molecules in real-time with suitable signal to noise.[5-7]

To address this need, a powerful palette of synthetic biology methods has emerged over the last 10 years which can interrogate RNA biology with high levels of sensitivity and with spatiotemporal control. Strategies such as the development of non-natural base-pairs can probe RNA dynamics down to base-pair level resolution.[8-13] In parallel, the development of aptamer technology has now reached an exciting stage where fluorogenic RNA motifs can detect transcription and even the presence of small molecule analytes.[12-13] Finally, RNA-binding proteins and gene-editing tools offer an auxiliary means of molecular recognition to detect RNA folds and motifs.[14-16] The aim of this review is to critique each of these fast moving areas and suggest potential opportunities that could further our understanding of the fundamental biology of this fascinating class of nucleic acids.

### Expanding Nature's Genetic Repertoire for Site-Specific RNA Labelling

The ability of nucleic acids to act as a template for the storage, replication and transfer of genetic information is borne out of the ability of A to pair with T/U and G pairing with C. As a consequence of the prevalence of all four of these nucleotides, labelling specific RNA molecules at defined internal positions poses significant challenges.[17] A nascent methodology is the development of unnatural nucleotides which pair with each other but not with naturally-occurring nucleotides.[18-19] The underlying pre-requisite of the success of this approach is the ability of these unnatural base-pairs (UBPs) to be replicated and transcribed with similar levels of fidelity to naturally-occurring base-pairs. In doing so, further derivatisation of UBPs opens up

opportunities to insert modifications into DNA and RNA molecules at defined sites using standard biochemical techniques such as PCR, transcription and reverse transcription.

The unnatural P-Z pair developed by the Benner group is one example of this approach. This UBP exhibits a unique hydrogen-bonding arrangement relative to naturally-occurring Watson-Crick base-pairs and has been explored in replication, transcription and reverse transcription.[20] However, mispairing of Z with G is prevalent, which poses difficulties in the wider utility of P-Z as an UBP platform for site-specific RNA labelling.[21] The Hirao and Romesberg groups have focused on the development of synthetic base-pairs which rely on complementarity of shape (Fig. 1a).[22-24] The hydrophobic Ds-Pa base-pair developed by the Hirao laboratory is replicable by PCR with fidelities approaching that of natural Watson-Crick base-pairs. Furthermore, this pairing regime can be used to site-specifically incorporate an alkyne functional group and fluorescent reporter molecules into RNA.[23] Post-synthetic labelling with either a fluorescent azide (*e.g.,* copper-catalyzed alkyne-azide cycloaddition, CuAAC) or cyclooctyne (*e.g.,* strain-promoted alkyne-azide cycloaddition, SPAAC) is then used to derivatise RNA at defined, internal locations.

The d5SICS-dNAM pair developed by the Romesberg group represents a far more structurally diverse example of shape complementarity UBP[24-25] where molecular recognition occurs *via* an intercalative mechanism (Fig. 1b).[25] Since the nature of shape complementarity is vastly different to the pairing exhibited by natural nucleotides, faithful replication of this UBP is observed *in vitro* and in *E. coli*.[26] This approach therefore opens up new synthetic biology applications, such as the development of semi-synthetic organisms with an expanded genetic alphabet.

The Romesberg group has applied these UBPs as a dual RNA-labelling strategy. Transcription of DNA incorporating the shape-complementary nucleotides dNAM and d5SICS in positons 704 and 750, using ribonucleoside triphosphates $5SICS^{CO}TP$ and $MMO2^{A}TP$, afforded the 243 nucleotide RNA fragment of the central 16S rRNA domain of *Thermus thermophiles.*[8] Post-synthetic functionalization of the $MMO2^{A}$ position with Cy3 NHS-ester and the $5SICS^{CO}$ position with Cy5 azide (CuAAC) produced an RNA fragment doubly-labelled in two defined locations, which was suitable to probe the dynamics of ribosome assembly using single molecule studies (Fig. 1c).

## Fluorogenic Aptamers as Reporters of RNA Synthesis, Dynamics and Localization

For over 23 years, Green Fluorescent Protein (GFP) has been an invaluable tool to report on the synthesis and localization of proteins both *in vitro* and *in vivo*.[27] The fluorescence characteristics of GFP arise from the formation of the 4-hydroxy-benzylidene-imidazolinone (HBI) fluorophore during protein synthesis. An equivalent GFP-like RNA platform has now been developed using SELEX.[28-30] Termed 'Spinach', this aptamer binds to the synthetic fluorophore, 3,5-difluoro-4-hydroxybenzylidene imidazolinone (DFHBI), with a $K_d$ ~ 500 nM and exhibits conditional green fluorescence (~ 200-fold increase; $\lambda_{max}$ 390 nm; emission ~ 475 nm) only when the Spinach-DFHBI complex is formed (Fig. 2a).

X-ray crystallographic studies have revealed that a structural hallmark of the complex is the presence of two G-quartets and a mixed tetrad, which induces the DFHBI chromophore to adopt a planar conformation. To accommodate DFHBI, Spinach folds into a hybrid structure with two coaxial duplex regions flanking the central G-quadruplex (Fig. 2b). When bound, DFHBI is sandwiched between a G-quartet and a U-A-U Hoogsten triplet, with hydrogen bonding from an unpaired guanine preventing lateral movement (Fig. 2c).[31-32] Extensive non-covalent interactions and $Mg^{2+}$ binding also assist in the binding of *cis*-DFHBI to Spinach.

This ground-breaking work has spurred the development of next-generation variants Spinach2,[33] Broccoli[34] and iSpinach.[35] The 49-nt Broccoli aptamer, for example, is significantly smaller than the first generation 98 nt Spinach aptamer and exhibits enhanced stability suitable for *in cellulo* imaging. To address thermal instability issues observed in earlier aptamer versions,[36] the Ryckelynck group developed the 68 nt iSpinach using *in vitro* compartmentalization to select aptamer populations based on fluorescence enhancement. The DFHBI-binding pocket appears to be conserved in all Spinach derivatives.

Spinach aptamers have now found extensive use as reporters in both prokaryotic and eukaryotic cells using Spinach fusion variants.[37] The Jäschke group developed a general transcription reporter platform using the first-generation Spinach construct.[38] The group prepared a DNA template bearing a T7 RNA polymerase

promoter upstream from an RNA sequence of interest (ROI), a hammerhead ribozyme sequence (HHR) and the Spinach reporter. The authors demonstrated the general applicability of the transcriptional reporter system using a range of RNA sequences and probed RNAP inhibition by the addition of Heparin, a known RNA polymerase inhibitor.

The Spinach aptamer has been further developed as a reporter platform for analyte detection in fusion aptamers where the binding of DFHBI is rendered conditional to analyte binding. This elegant sensing method has been used to detect small molecule metabolites (*e.g.,* c-di-GMP,[39-40] glycine,[41] ADP & SAM[42]) both *in vitro* and *in cellulo*. An exemplar of this approach is the development of conditional SAH sensors of methyltransferase activity (Fig. 2d).[42] The Hammond group developed a Spinach fusion where DFHBI fluorescence is conditional on SAH binding. This aptamer sensor was used to detect methylthioadenosine nucleosidase (MTAN) activity in *E. coli* and integrated into a high-throughput screening platform to explore inhibitors of this enzyme. This technique showed remarkable sensitivity for the target analyte at 10 μM, despite the presence of a number of other compounds, including structurally related compounds such as SAM and ATP at 100 μM and 3 mM respectively.

At present, the relatively weak dissociation constants of DFHBI-binding aptamers ($K_d$ > 400 nM[43]) are not suitable for single molecule applications where fluorophore binding needs to be in the low nanomolar range and exhibit slow off-rates. To address this limitation, the Mango thiazole orange (TO)-binding aptamer system was developed by SELEX. Mango binds to a biotinylated analogue TO-1 with a $K_D$ 3.2 ± 0.7 nM and induces 1100-fold enhancement in fluorescence emission.[43-44] The conditional fluorogenicity of Mango *in cellulo* was then shown by injecting biotin-TO and Mango RNA into the syncytial gonads of *C. elegans*.

**Protein-Derived Recognition of RNA Motifs**

RNA-binding GFP fusion proteins that have a high affinity for selected RNA secondary structures have been highly effective tools for RNA labelling. The coat protein of the MS2 phage (MCP) is the most extensively used to study RNA localization and dynamics in cells.[45] Several limitations of this approach do exist however. Firstly, long RNA tags (up to 1200 nt) are typically required in order to effect MCP-

RNA binding, which could impair RNA biology. Secondly, MCP recognition is not sequence specific. These limitations can partially be addressed by the use of the protein Pumilio, which directly recognises RNA sequences 8 nucleotides in length.[46] Sequence selectivity of Pumilio is programmable *via* the use of different amino acid substitutions and, when fused with GFP, Pumilio-GFP fusions have shown utility as imaging agents to track β-actin dynamics in mammalian cells.

One of the most biotechnologically important tools that has emerged over the past 5 years has been the gene-editing CRISPR-associated protein 9 nuclease (Cas9). Derived from *Streptococcus pyogenes*, sequence selectivity of the ribonucleoprotein Cas9 system is achieved by base-pairing to a 20 nt target DNA sequence and its associated single-guide RNA (sgRNA). The site of double-stranded DNA cleavage is directed to a site adjacent to protospacer adjacent motifs (PAMs).[47] Since the sequence selectivity is derived from the modularity of the sgRNA, Cas9 has found extensive use as a game-changing gene-editing and imaging tool for gene expression.[48]

The Yeo group has now extended the utility of the gene-editing Cas9 system as an imaging tool to detect specific single-stranded RNA sequences 39 nt in length.[15] In contrast to RNA targeting using Pumilio proteins, where each protein needs to be designed and validated for each RNA sequence, the recognition of arbitrary target RNA sequences using the Cas9 system (RCas9) is far simpler, requiring the introduction of the complementary sgRNA sequence and a mismatched synthetic PAMmer oligonucleotide. Fusing RCas9 with a fluorescent reporter protein (*e.g.,* GFP or mCherry) produced a platform to image and track RNA specific RNA molecules, such as *CCNA2* and *TFRC* which have low expression levels, into stress granules after the induction of cellular stress using sodium arsenite. Although the need to transfect the synthetic PAMmer is one limitation of this approach, the sensitivity of the fluorescence emission and the inherent programmability of the platform render the RCas9 system one of the most powerful emerging tools to track RNA synthesis and cellular distribution.

## Summary and Outlook

The multi-disciplinary nature of synthetic biology now offers a powerful palette of methods to detect and/or label one of the most challenging biomacromolecules found in living cells. Although each of the major labelling categories covered in this review do have their limitations, a blended approach where cross-fertilization of the distinct advantages of each of these approaches could offer opportunities to address some of the limitations of each method when used in isolation. For example, blending UBPs in the synthesis of PAMmer oligonucleotides for targeting RNA sequences using the RCas9 system could offer a new genetically-encoded approach to RNA imaging.[49] This and further developments to enhance the sensitivity of fluorescence-based approaches will undoubtedly further our understanding of RNA biology.

## References

Papers of particular interest, published within the period of review, have been highlighted as:

• of special interest •• of outstanding interest

1. Serganov A, Patel DJ: **Ribozymes, riboswitches and beyond: Regulation of gene expression without proteins.** *Nat Rev Genet* 2007, **8**:776-790.

2. He L, Hannon GJ: **Microrna: Small RNAs with a big role in gene regulation.** *Nat Rev Genet* 2004 **5**:522-531.

3. Fedor MJ, Williamson JR: **The catalytic diversity of RNAs.** *Nat Rev Mol Cell Biol* 2005, **6**:399-412.

4. Matlin AJ, Clark F, Smith CWJ: **Understanding alternative splicing: Towards a cellular code.** *Nat Rev Mol Cell Biol* 2005, **6**:386-398.

5. Rath AK, Rentmeister A: **Genetically encoded tools for RNA imaging in living cells.** *Curr Opin Biotechnol* 2015, **31**:42-49.

6. Pitchiaya S, Heinicke LA, Custer TC, Walter NG: **Single molecule fluorescence approaches shed light on intracellular RNAs.** *Chem Rev* 2014, **114**:3224-3265.

7. Hövelmann F, Gaspar I, Ephrussi A, Seitz O: **Brightness enhanced DNA fit-probes for wash-free RNA imaging in tissue.** *J Am Chem Soc* 2013, **135**:19025-19032.

•• 8. Lavergne T, Lamichhane R, Malyshey DA, Li ZT, Li LJ, Sperling E, Williamson JR, Millar DP, Romesberg FE: **Fret characterization of complex conformational changes in a large 165 ribosomal RNA fragment site-specifically labeled using unnatural base pairs.** *ACS Chem Biol* 2016, **11**:1347-1353.

This manuscript describes the first example of site-specific RNA labelling using derivatised UBPs.

9. Chen T, Hongdilokkul N, Liu ZX, Thirunavukarasu D, Romesberg FE: **The expanding world of DNA and RNA.** *Curr Opin Chem Biol* 2016, **34**:80-87.

10. Seo YJ, Malyshev DA, Lavergne T, Ordoukhanian P, Romesberg FE: **Site-specific labeling of DNA and RNA using an efficiently replicated and transcribed class of unnatural base pairs.** *J Am Chem Soc* 2011, **133**:19878-19888.

• 11. Someya T, Ando A, Kimoto M, Hirao I: **Site-specific labeling of RNA by combining genetic alphabet expansion transcription and copper-free click chemistry.** *Nucleic Acids Res* 2015, **43**:6665-6676.

This manuscript describes the site-specific RNA labelling at a single RNA site using an-alkyne derivatised UBP.

12. You MX, Jaffrey SR: **Structure and mechanism of RNA mimics of green fluorescent protein.** *Ann Rev Biophys.* 2015, **44**:187-206.

13. You MX, Litke JL, Jaffrey SR: **Imaging metabolite dynamics in living cells using a spinach-based riboswitch.** *Proc Natl Acad Sci U S A* 2015, **112**:E2756-E2765.

14. Biffi G, Di Antonio M, Tannahill D, Balasubramanian S: **Visualization and selective chemical targeting of RNA G-quadruplex structures in the cytoplasm of human cells.** *Nat Chem* 2014, **6**:75-80.

•• 15. Nelles DA, Fang MY, O'Connell MR, Xu JL, Markmiller SJ, Doudna JA, Yeo GW: **Programmable RNA tracking in live cells with CRISPR/Cas9.** *Cell* 2016, **165**:488-496.

This manuscript reports the first example of imaging RNA using an engineered CRISPR/Cas9 system. Key developments described in this manuscript include the development of modified oigonucleotides to render the system specific for target RNA sequence.

16. Liu Y, Holmstrom E, Zhang J, Yu P, Wang J, Dyba MA, De C, Ying J, Lockett S, Nesbitt DJ, Ferre-D/'Amare AR *et al*: **Synthesis and applications of RNAs with position-selective labelling and mosaic composition.** *Nature* 2015, **522**:368-372.

17. Malyshev DA, Romesberg FE: **The expanded genetic alphabet.** *Angew Chem-Int Edit* 2015, **54**:11930-11944.

18. Taylor AI, Arangundy-Franklin S, Holliger P: **Towards applications of synthetic genetic polymers in diagnosis and therapy.** *Curr Opin Chem Biol* 2014, **22**:79-84.

19. Zhang LQ, Yang ZY, Sefah K, Bradley KM, Hoshika S, Kim MJ, Kim HJ, Zhu GZ, Jimenez E, Cansiz S, Teng IT *et al*: **Evolution of functional six-nucleotide DNA.** *J Am Chem Soc* 2015, **137**:6734-6737.

20. Leal NA, Kim HJ, Hoshika S, Kim MJ, Carrigan MA, Benner SA: **Transcription, reverse transcription, and analysis of RNA containing artificial genetic components.** *ACS Synth Biol* 2015, **4**:407-413.

• 21. Reichenbach LF, Sobri AA, Zaccai NR, Agnew C, Burton N, Eperon LP, de Ornellas S, Eperon IC, Brady RL, Burley GA: **Structural basis of the mispairing of an artificially expanded genetic information system.** *Chem* 2016, **1**:946-958.

This manuscript reports the first example of mis-pairing of a synthetic base-pair (P-Z) with a naturally-occurring nucleotide. Mis-pairing is pH-dependent and occurs via hydrogen-bonding of the synthetic nucleotide Z with G.

22. Kimoto M, Mitsui T, Yamashige R, Sato A, Yokoyama S, Hirao I: **A new unnatural base pair system between fluorophore and quencher base analogues for nucleic acid-based imaging technology.** *J Am Chem Soc* 2010, **132**:15418-15426.

23. Ishizuka T, Kimoto M, Sato A, Hirao I: **Site-specific functionalization of RNA molecules by an unnatural base pair transcription system via click chemistry.** *Chem Commun* 2012, **48**:10835-10837.

24. Dhami K, Malyshev DA, Ordoukhanian P, Kubelka T, Hocek M, Romesberg FE: **Systematic exploration of a class of hydrophobic unnatural base pairs yields multiple new candidates for the expansion of the genetic alphabet.** *Nucleic Acids Res* 2014, **42**:10235-10244.

25. Betz K, Malyshev DA, Lavergne T, Welte W, Diederichs K, Romesberg FE, Marx A: **Structural insights into DNA replication without hydrogen bonds.** *J Am Chem Soc* 2013, **135**:18637-18643.

26. Malyshev DA, Dhami K, Lavergne T, Chen TJ, Dai N, Foster JM, Correa IR, Romesberg FE: **A semi-synthetic organism with an expanded genetic alphabet.** *Nature* 2014, **509**:385-388.

27. Giepmans BNG, Adams SR, Ellisman MH, Tsien RY: **Review - the fluorescent toolbox for assessing protein location and function.** *Science* 2006, **312**:217-224.

28. Song WJ, Strack RL, Jaffrey SR: **Imaging bacterial protein expression using genetically encoded RNA sensors.** *Nat Methods* 2013, **10**:873-875.

29. Paige JS, Nguyen-Duc T, Song WJ, Jaffrey SR: **Fluorescence imaging of cellular metabolites with RNA.** *Science* 2012, **335**:1194.

30. Paige JS, Wu KY, Jaffrey SR: **RNA mimics of green fluorescent protein.** *Science* 2011, **333**:642-646.

• 31. Huang H, Suslov NB, Li NS, Shelke SA, Evans ME, Koldobskaya Y, Rice PA, Piccirilli JA: **A G-quadruplex-containing RNA activates fluorescence in a GFP-like fluorophore.** *Nat Chem Biol* 2014 **10**:686-691.

• 32. Warner KD, Chen MC, Song WJ, Strack RL, Thorn A, Jaffrey SR, Ferre-D'Amare AR: **Structural basis for activity of highly efficient RNA mimics of green fluorescent protein.** *Nat Struct Mol Biol* 2014, **21**:658-663.

References 31 and 32 present the first crystal structures of the Spinach aptamer.

33. Strack RL, Disney MD, Jaffrey SR: **A superfolding Spinach2 reveals the dynamic nature of trinucleotide repeat-containing RNA.** *Nat Methods* 2013, **10**:1219-1224.

34. Filonov GS, Moon JD, Svensen N, Jaffrey SR: **Broccoli: Rapid selection of an RNA mimic of green fluorescent protein by fluorescence-based selection and directed evolution.** *J Am Chem Soc* 2014, **136**:16299-16308.

•• 35. Autour A, Westhof E, Ryckelynck M: **iSpinach: A fluorogenic RNA aptamer optimized for in vitro applications.** *Nucleic Acids Res* 2016, **44**:2491-2500.

This manuscript describes optimization of the fluorogenic properties of a next-generation Spinach aptamer.

36. Ketterer S, Fuchs D, Weber W, Meier M: **Systematic reconstruction of binding and stability landscapes of the fluorogenic aptamer spinach.** *Nucleic Acids Res* 2015, **43**:9564-9572.

•• 37. Guet D, Burns LT, Maji S, Boulanger J, Hersen P, Wente SR, Salamero J, Dargemont C: **Combining spinach-tagged RNA and gene localization to image gene expression in live yeast.** *Nat Commun* 2015, **19**: 8882.

This paper demonstrates the power of the Spinach aptamer system to conditionally image mRNA in *S. cerevisiae.* This work overcomes existing limitations associated with constant background fluoresence of MS2 approaches.

38. Höfer K, Langejürgen LV, Jäschke A: **Universal aptamer-based real-time monitoring of enzymatic RNA synthesis.** *J Am Chem Soc* 2013, **135**:13692-13694.

39. Nakayama S, Luo YL, Zhou J, Dayie TK, Sintim HO: **Nanomolar fluorescent detection of c-di-GMP using a modular aptamer strategy.** *Chem Commun* 2012, **48**:9059-9061.

40. Wang XC, Wilson SC, Hammond MC: **Next-generation RNA-based fluorescent biosensors enable anaerobic detection of cyclic di-GMP.** *Nucleic Acids Res* 2016, **44**:10.

41. Ketterer S, Gladis L, Kozica A, Meier M: **Engineering and characterization of fluorogenic glycine riboswitches.** *Nucleic Acids Res* 2016, **44**:5983-5992.

42. Su YC, Hickey SF, Keyser SGL, Hammond MC: **In vitro and in vivo enzyme activity screening via RNA-based fluorescent biosensors for *S*-adenosyl-L-homocysteine (SAH).** *J Am Chem Soc* 2016, **138**:7040-7047.

43. Dolgosheina EV, Jeng SCY, Panchapakesan SSS, Cojocaru R, Chen PSK, Wilson PD, Hawkins N, Wiggins PA, Unrau PJ: **RNA mango aptamer-fluorophore: A bright, high-affinity complex for RNA labeling and tracking.** *ACS Chem Biol* 2014, **9**:2412-2420.

44. Jeng SCY, Chan HHY, Booy EP, McKenna SA, Unrau PJ: **Fluorophore ligand binding and complex stabilization of the RNA mango and RNA spinach aptamers.** *RNA* 2016, **22**:1884-1892.

45. Bertrand E, Chartrand P, Schaefer M, Shenoy SM, Singer RH, Long RM: **Localization of ash1 mRNA particles in living yeast.** *Mol Cell* 1998, **2**:437-445.

46. Wang XQ, McLachlan J, Zamore PD, Hall TMT: **Modular recognition of RNA by a human pumilio-homology domain.** *Cell* 2002, **110**:501-512.

47. Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E: **A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity.** *Science* 2012, **337**:816-821.

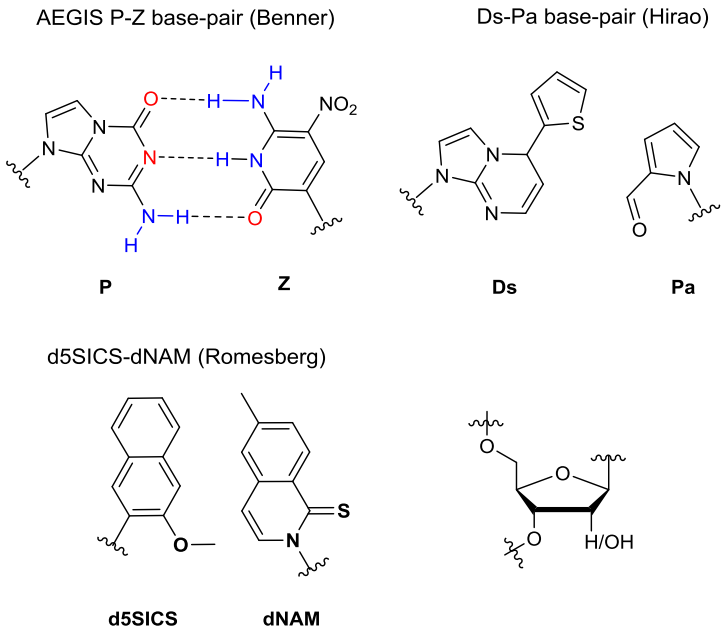
48. Jiang W, Bikard D, Cox D, Zhang F, Marraffini LA: **RNA-guided editing of bacterial genomes using CRISPR-Cas systems.** *Nat Biotech* 2013, **31**:233-239.

49. Zhang Y, Lamb BM, Feldman AW, Zhou AX, Lavergne T, Li L, Romesberg FE: **A semisynthetic organism engineered for the stable expansion of the genetic alphabet.** *Proc Natl Acad Sci USA* 2017, **114**:1317-1322.

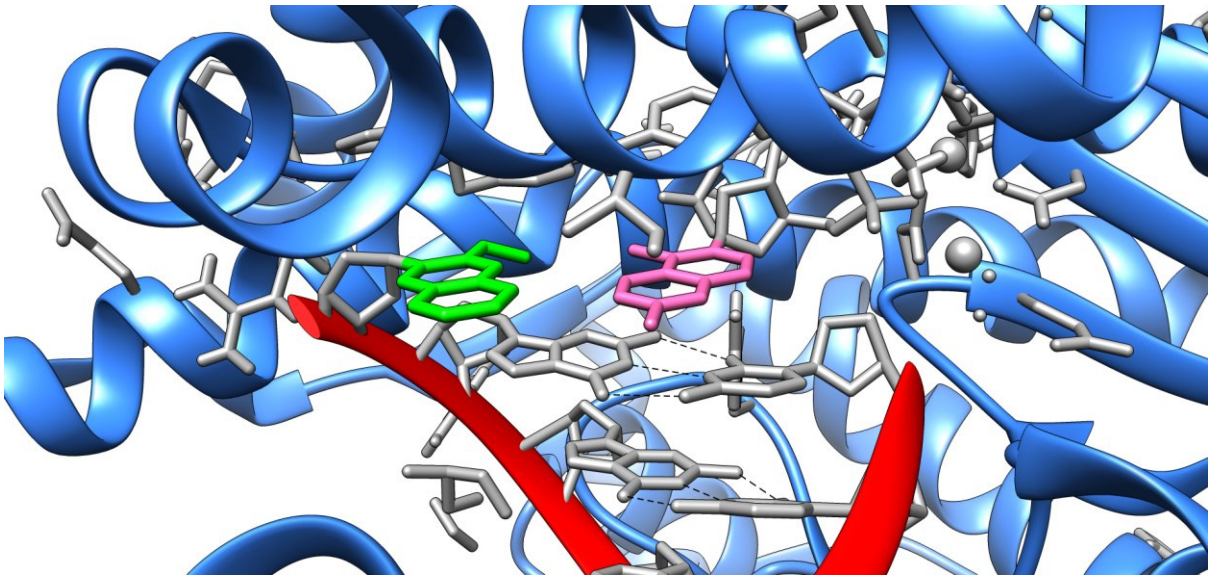
50. Betz K, Malyshev DA, Lavergne T, Welte W, Diederichs K, Dwyer TJ, Ordoukhanian P, Romesberg FE, Marx A: **KlenTaq polymerase replicates unnatural base pairs by inducing a Watson-Crick geometry.** *Nat Chem Biol* 2012, **8**:612-614.

## Figures

**A)**


AEGIS P-Z base-pair (Benner)
Ds-Pa base-pair (Hirao)
P
Z
Ds
Pa
d5SICS-dNAM (Romesberg)
d5SICS
dNAM
H/OH

**B)**



**C)**

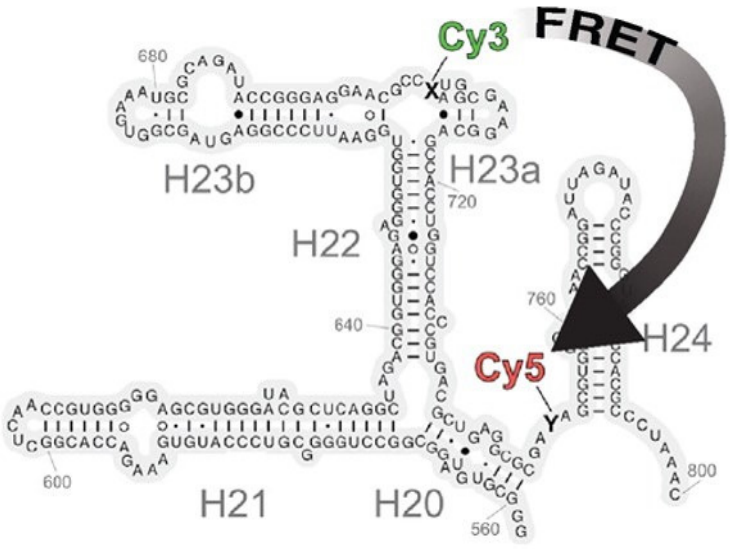
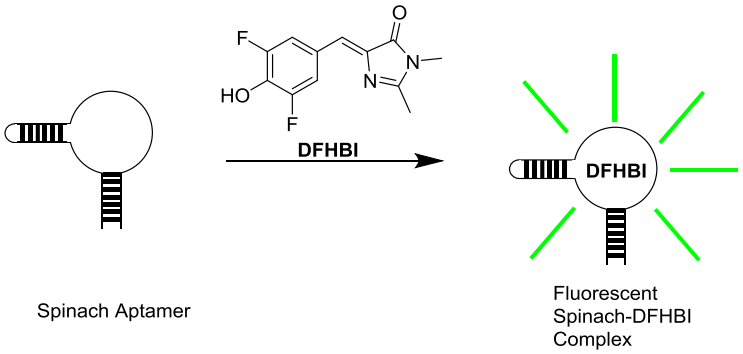
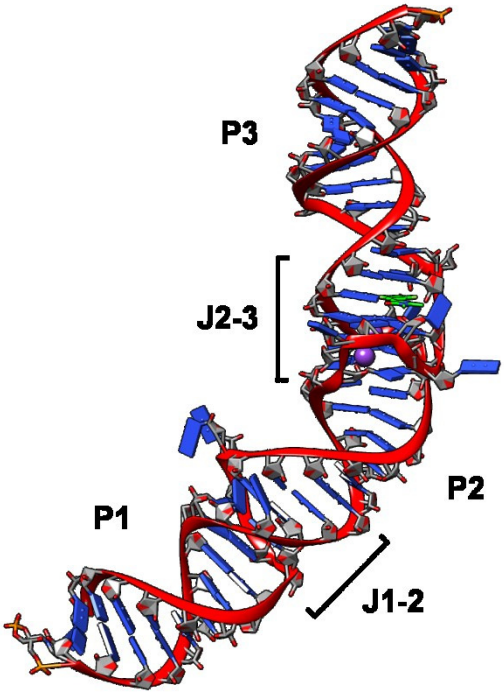



**Figure 1.** Current developments of unnatural base-pairs (UBPs). A) Structures of UBPs developed by Benner (P-Z), Hirao (Ds-Pa), and Romesberg (d5SICS-dNAM). B) Pairing of d5SICS (green) with dNAM (violet) within Klen Taq polymerase (PDB: 3SV3).[50] C) Dual labelling of 16S ribosomal RNA containing modified dNAM and d5SICS base-pairs. Adapted with permission from Lavergne T, Lamichhane R, Malyshey DA, Li ZT, Li LJ, Sperling E, Williamson JR, Millar DP, Romesberg FE: *ACS Chem Biol* 2016, **11**:1347-1353. [8] Copyright 2016 American Chemical Society.

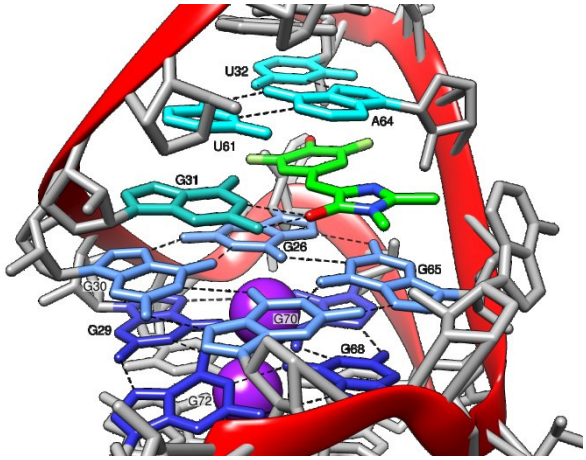
**A)**


F
HO
F
O
N
N
DFHBI
DFHBI
Spinach Aptamer
Fluorescent
Spinach-DFHBI
Complex

**B)**


P3
J2-3
P2
P1
J1-2

**C)**


U32
A64
U61
G31
G26
G65
G30
G70
G29
G68
G72

**D)**

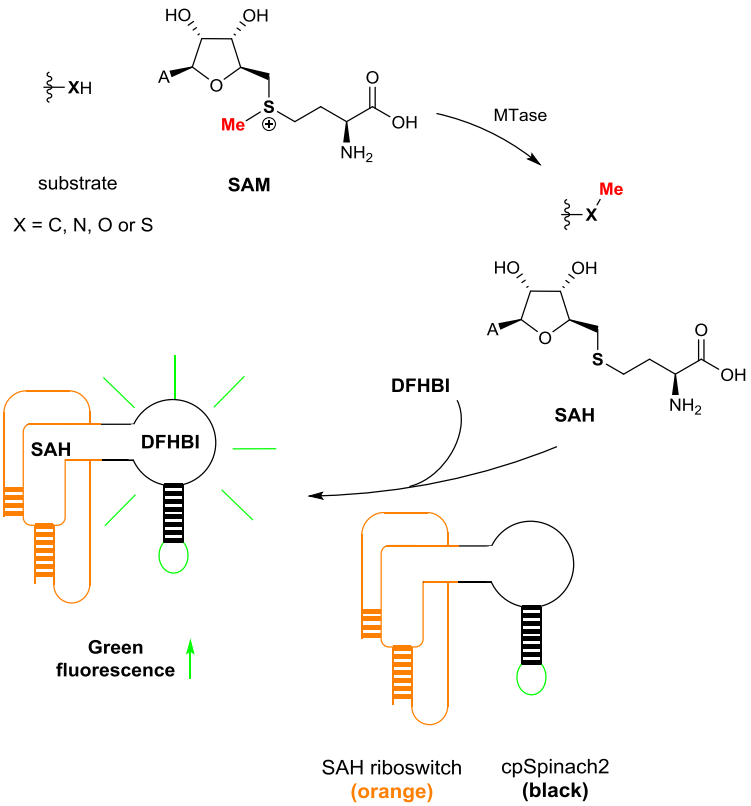



**Figure 2.** RNA labelling using the Spinach aptamer system. A) Fluorescence activation of DFHBI upon complexation with the Spinach aptamer. B) X-ray crystal structure of Spinach. Sections P1, P2, and P3 are base-pairing regions, whereas J1-2 and J2-3 represent junctions between regions. C) Structure of the DFHBI binding pocket of Spinach. DFHBI is highlighted in green, purple spheres are $K^+$ (PDB: 4TS2).[32] D) Schematic of a Spinach-based aptamer assay to assess C-methyltransferase activity. SAH riboswitch/cpSpinach2 fusion binds SAH. This induces a conformational change allowing for cpSpinach2 to bind DFHBI, which in turn switches on fluorescence of DFHBI. X = C, N, O or S in an appropriate substrate for a SAM dependent methyltransferase.