A Journal of the Gesellschaft Deutscher Chemiker

Angewandte Chemie

GDCh

International Edition

www.angewandte.org

Accepted Article

Title: The Network of Replication, Transcription, and Reverse Transcription of a Synthetic Genetic Cassette

Authors: Hui Yang, Elena Ereemeeva, Mikhail Abramov, and Piet Herdewijn

This manuscript has been accepted after peer review and appears as an Accepted Article online prior to editing, proofing, and formal publication of the final Version of Record (VoR). This work is currently citable by using the Digital Object Identifier (DOI) given below. The VoR will be published online in Early View as soon as possible and may be different to this Accepted Article as a result of editing. Readers should obtain the VoR from the journal website shown below when it is published to ensure accuracy of information. The authors are responsible for the content of this Accepted Article.

To be cited as: *Angew. Chem. Int. Ed.* 10.1002/anie.202011887

Link to VoR: <https://doi.org/10.1002/anie.202011887>

RESEARCH ARTICLE

The Network of Replication, Transcription, and Reverse Transcription of a Synthetic Genetic Cassette

Hui Yang^{[a]†}, Elena Ereemeeva^{[a]†}, Mikhail Abramov^[a], and Piet Herdewijn^{*[a]}

[a] Mrs. H. Yang, Dr. E. Ereemeeva, Dr. M. Abramov, Prof. P. Herdewijn
Medicinal Chemistry
Rega Institute for Medical Research
KU Leuven
Herestraat 49, Box – 1041, B-3000 Leuven, Belgium
E-mail: piet.herdewijn@kuleuven.be
† Authors contributed equally

Supporting information for this article is given via a link at the end of the document.

Abstract: Synthetic nucleic acids, with four non-canonical nucleobases, can function as genetic materials. A comprehensive analysis of PCR amplification, transcription, reverse transcription, and cloning was done to screen for alternative genetic monomers. A small library of six modified nucleobases was selected: the modified 2'-deoxyribonucleoside (dZTPs) and ribonucleoside (rZTPs) triphosphates of 7-deaza-adenine, 5-chlorouracil, 7-deaza-guanine or inosine together with 5-fluorocytosine or 5-bromocytosine. The fragments composed of one to four modified nucleotides (denoted as DZA) have been successfully recognised and transcribed to natural or modified RNA (denoted as RZA) by T7 RNA polymerase. The fully modified RZA fragment could be reverse transcribed and then amplified in the presence of various dZTPs. Noticeably, modified fragments could function as genetic templates *in vivo* by encoding the 678 base pair gene of a fluorescent protein in bacteria. These results demonstrate the existence of a fully simulated genetic circuit that uses synthetic materials.

Introduction

Replication, transcription, and reverse transcription are the three central biological processes with which natural nucleic acids duplicate themselves or interconvert into each other.^[1] These processes are widely used in biotechnology and molecular biology applications such as the polymerase chain reaction (PCR), *in vitro* RNA synthesis, molecular cloning, sequencing, and the analysis of gene expression.^[2] They are the foundations of the advanced gene and genome manipulation technologies extensively used in modern bioengineering.^[3] The fidelity of these processes is based on the Watson-Crick base-pairing system. Along with natural monomers, non-canonical nucleotides are abundantly distributed in cellular RNA and DNA.^[4] They play a pivotal role in the biological functions of nucleic acids, raising the question of whether other synthetic nucleobases can be successfully involved in genetic information inheritance.

Previously, we have investigated whether the fully base modified nucleic acids (DZA, Figure 1) could replicate *in vitro* (by PCR amplification) and serve as a genetic template *in vivo*.^[5] Long (0.5 and 1.5 kb) DZA fragments were obtained enzymatically using two different genetic cassettes composed of four synthetic nucleobases. These modified DNAs were protected from cleavage by restriction endonucleases and could encode

antibiotic resistance in bacteria.^[5–7] We have now investigated the potential to involve fully base modified RNA (denoted as RZA) in the network of replication, transcription, and reverse transcription. In this way, we established an alternative genetic flow based on DZA and RZA, with the potential to cross-communicate with the natural genetic circuit composed of DNA and RNA (Figure 1b). Despite the several attempts to partially reproduce information flow using synthetic nucleic acids,^[8] the direct information exchange from fully substituted DNA to cognate RNA and back to fully modified DNA, however, has never been established.

Figure 1. Chemical structures of modified nucleotides and general experimental scheme. **a.** The chemical structures of non-canonical nucleobases (T1, A1, A2, C1-C3, G1, G2) as well as their 2'-deoxyribonucleoside (dZTP) and ribonucleoside (rZTP) 5'-triphosphates. **b.** General scheme of *in vitro* and *in vivo* studies. Base-modified DZA templates were PCR synthesised, transcribed into unmodified RNA or base-modified RZA sequences, and reverse transcribed to DNA or DZA *in vitro*. In bacteria, DZA templates could serve as genetic templates encoding red fluorescent protein mTangerine. Unnatural genetic information flow established in the present research is shown in red.

Herein, the alternative genetic flow of DZA to RZA and back to DZA comprised of the modified bases 5-chlorouracil, 7-deazaadenine, 2,6-diaminopurine, 5-fluorocytosine, 5-bromocytosine, 5-methylcytosine, 7-deazaguanine, and inosine [5-ClU (T1), 7-deazaA (A1), DAP (A2), 5-FC (C1), 5-BrC (C2), 5-MeC (C3), 7-deazaG (G1), and I (G2), respectively, Figure 1a] was analysed *in vitro* and the resulting genetic templates were tested to encode red fluorescent protein mTangerine in bacteria.^[9]

RESEARCH ARTICLE

Given the growing interest in the nucleobase-modified genetic material in therapy and diagnostics (e.g., mRNA therapy, RNA interference, aptamers, DNA/RNAzymes),^[10–12] RZA and DZA are of particular importance. These synthetic polymers possess exceptional biocompatibility with the natural biosystems while enhancing the chemical repertoire and restriction enzyme resistance of natural nucleic acids.

Results and Discussion

PCR synthesis of modified DZA fragments (DNA→DZA)

Different combinations of non-canonical 2'-deoxynucleoside triphosphates (dZTPs, Figure 1 and S1) were tested as potential substrates for Taq DNA polymerase in a PCR amplification assay of an 87-mer DNA template.^[13] The 104-bp PCR products were obtained in all cases (Figure 2a,b), even when all four dNTPs were substituted for non-canonical counterparts. In Figure 2b, it is shown that a single nucleotide substitution did not affect PCR amplification. The product yield was similar or even higher in the case of pyrimidine modifications (i.e., T1, C1, and C2) than with the four natural dNTPs. The most profound effect on PCR amplification was observed when dCTP was substituted for dC2TP analogue in dNTPs mixture resulting in 133% yield. In contrast, the guanosine-modified triphosphates decreased the efficiency of PCR amplification to 40% and 16%, for G1 and G2, respectively. Surprisingly, the PCR amplification efficiencies with modified dGTPs were improved by combination with dC2TP (95% for G1+C2 and 32% for G2+C2). This positive effect on dsDZA synthesis remained even after adding the two other modified triphosphates, dA1TP and dT1TP, to the reaction mixture. The PCR amplification with fully modified dZTPs sets yielded 77% and 27% for G1+C2+A1+T1 and G2+C2+A1+T1, respectively. This data confirm previous observations, i.e., that the use of certain combinations of modified nucleoside triphosphates could synergistically improve the PCR amplification efficiency compared to single modification.^[6,7]

In an effort to improve PCR synthesis of G2-modified DZA fragments, we examined the dZTP combinations composed of 5-MeC and DAP (C3 and A2) triphosphates together with dG2TP (Figure 1). Inosine (G2) is a DNA duplex destabilising nucleotide.^[14] In contrast, 5-MeC and DAP, can improve the DNA hybridisation strength, and, therefore, potentially increase the PCR amplification yield.^[15,16] However, all dZTP cassettes containing A2 and C3 had reduced PCR efficiencies compared to A1 and C2 combinations, though C3 was preferable to C1 in any combination with G2 (Figure S2). The simultaneous incorporation of several types of duplex-stabilizing nucleotides (i.e., A2 together with C2 or C3) might hamper the PCR amplification due to over-stabilisation of the amplicon and its inability to reanneal during PCR.^[16]

Consistent with our previous results,^[5,6] the most successful dZTP set to produce the G2-modified dsDZA fragments was G2+C2+A1+T1, whereas the triphosphate set composed of G1+C1+A1+T1 resulted in the highest yield (96%) of fully modified dsDZA product, compared to the natural product.

In vitro transcription of modified DZA templates with natural rNTPs (DZA→RNA)

The resulting base-modified dsDZA templates were used in an *in vitro* transcription assay with T7 RNA polymerase (T7 RNAP) and natural rNTPs as reagents. As shown in Figure 2c, the dsDZA fragments containing single pyrimidine substitution (i.e., 5-CIU, 5-FC, or 5-BrC) were efficiently recognised by T7 RNAP as templates providing 83%, 82%, and 90% of relative transcription yield, respectively. On the other hand, the presence of purine-modified nucleotides, A1 and G2, in the templates reduced transcription efficiency to 53% and 20%, respectively. When unnatural nucleotides replaced all four letters in the DNA template, i.e., G2+C2+A1+T1, the yield of transcripts formation dramatically decreased to 1%. The transcription efficiency of the fully modified templates was slightly improved to 5% by using the increased amounts of T7 RNAP (from 60 to 100 U, Figure 2c) or by increasing Mg²⁺ and rNTPs concentrations (from 8 to 10 mM for Mg²⁺ and from 2 to 3 mM for rNTPs, Figure S3a). In general, the presence of a modified G base (7-deazaG and I, Figure S3b) in the DZA template reduced the transcription efficiency more significantly than all other modifications. The G2-substituted dsDZA fragments were somewhat preferable to G1-substituted templates by T7 RNAP, especially in combination with C2 (G2+C2 fragments gives 40% relative yield vs 10% of G2-containing template and 7% for G1-containing template). It was shown that T7 RNAP has a strong sequence preference for guanosines at positions +1 and +2 of the promoter, and three consecutive modified G in the DZA template can certainly hamper RNA synthesis.^[17] Nevertheless, the fully modified RZA transcript composed of G2+C2+A1+T1 can be synthesised in optimised reaction conditions with the relative yield of 11% (Figure S3b).

To identify how modified bases in dsDZA templates affect the accuracy of *in vitro* transcription, sequencing analysis of the products after reverse transcription and PCR amplification (RT-PCR) was performed. The mutation frequency in the dsDNAs after transcription and RT-PCR was compared to those of the initial dsDZA templates. The sequencing data shown in Figure S4a,b and Table S2A,B, demonstrate that fully modified dsDZA fragment (G2+C2+A1+T1) was correctly amplified by Taq DNA polymerase, despite the fact that the incorporation of A1 (single or in combination with T1 and C1) during PCR amplification increased the error rate. These results are in line with our previous report,^[5] where A1 was somewhat the most mutagenic nucleotide in PCR among the other non-canonical nucleotides. On the other hand, the final dsDNA fragments obtained after *in vitro* transcription, reverse transcription, and PCR amplification exhibited higher mutation frequency than the initial dsDZAs (Figure S4b and Table S2B). The sequencing results demonstrate that the mutation frequency of RT-PCR products is $1.2\text{--}4.8 \cdot 10^{-3}$ with a relatively low error rate for a single modification (all 0.12%) increasing with multiple non-canonical nucleotides in dsDZA templates. This increase in the mutation rate can be due to cumulative infidelity of enzymatic transformations, including RNA synthesis, reverse transcription, and PCR by T7 RNAP, M-MuLV, and Taq polymerase, respectively. The T7 RNAP fidelity can vary ranging $10^{-6}\text{--}10^{-4}$, and the error rate of M-MuLV reverse transcriptase is reported as $3\text{--}8.4 \cdot 10^{-5}$.^[18] Additionally, the RNA synthesis using modified dsDZA templates was performed in 'forcing' conditions with increased concentration of T7 RNAP, rNTPs, and Mg²⁺ that further supports mispairing.

RESEARCH ARTICLE

Figure 2. Enzymatic transformations of base-modified fragments. **a.** General experimental scheme of *in vitro* studies (the detailed scheme is shown in Supporting Information Figure S1). **b.** The relative PCR amplification efficiencies to natural product formation for each composition of dZTPs. **c.** The relative efficiencies of *in vitro* transcription of modified dsDZA templates. The reactions were performed with 60 or 100 U of T7 RNAP at 37°C for 1 h. **d.** The relative efficiencies of *in vitro* transcription of natural DNA template with different compositions of modified rZTPs. Reactions were performed with 40 U T7 RNAP at 37°C for 2 h. **e.** Optimisation of RNA synthesis of G1+C2+A1+T1 transcript followed by 3'-Cy5 labelling for visualisation of the product. **f.** The relative efficiency of *in vitro* transcription of modified DZA templates with the corresponding modified rZTPs followed by 3'-Cy5 labelling of transcripts for visualisation. The modified samples were loaded to the gel in a 5-fold excess compare to PC. **g.** The relative efficiencies of reverse transcription reaction of fully modified RZA transcript (G1+C2+A1+T1) with various dZTPs. The PC indicates the positive control, enzymatic reactions with all-natural dNTPs/rNTPs and/or templates is taken as 100%. NC is referred to negative control, reaction without the corresponding rNTPs/dNTPs. The gels were scanned either upon Cy3 (**b** and **g**) or Cy5 channels (**e** and **f**); alternatively, they were first stained by SYBR Gold DNA dye and then scanned upon Cy2 channel (**c** and **d**).

Despite the accumulation of some errors in the final RT-PCR products, their mutation frequency was comparable to the one of all-natural dsDNA template (PC, positive control, 0.12%, Figure S4b).

In vitro transcription of natural DNA template with modified rZTPs (DNA→RZA)

Next, the *in vitro* transcription of a natural 104-bp dsDNA template with modified ribonucleoside triphosphates (rZTPs) using T7 RNA polymerase was investigated. From Figure 2d, it is evident that

A1, T1, C1, and C2 ribonucleoside triphosphates were alternative substrates to the natural rNTPs for T7 RNAP. The transcription yield with 5-Br-rCTP (C2) was 28% higher than with rCTP. Evidently, both brominated analogues, 5-bromocytidine and 5-bromo-2'-deoxycytidine triphosphates, are preferable substrates over natural ones for the corresponding polymerases, T7 RNAP and Taq (Figure 2b,d). This opens an avenue for the highly efficient production of bromo-functionalised DNA and RNA polymers that can be used in studies such as UV-mediated cross-linking assays, protein-nucleic acid interaction studies, X-ray

RESEARCH ARTICLE

diffraction experiments, as well as aptamer sequence optimisation.^[19]

With an increasing number of modifications introduced in rZTP set, the efficiency of RZA synthesis decreased. As such, the A1+T1 combination of triphosphates provided 58% relative transcription yield, while A1+T1+C1 yielded 26% of the full-length product. With the introduction of modified guanosine triphosphates in rZTP, the efficiency of detectable *in vitro* transcription markedly decreased, resulting in 10% relative yield with single G1 triphosphate in the mixture and 1% or less for all other G-modified groups.

The decline in the efficiency of *in vitro* transcription might be due to multiple factors. First, 7-deazaguanosine-containing nucleic acids inhibit the binding ability of fluorescent DNA dyes (e.g., SYBR Gold, ethidium bromide).^[6,20] Therefore, the resulting RZA/DZA fragments can have reduced band intensities in the gel upon staining by the SYBR Gold solution compared to unmodified fragments. From Figure S5a, it is evident that G-modified DNA samples quenched the SYBR Gold fluorescence, whereas all fragments were detectable upon Cy3 scanning. To overcome this problem, the post-transcriptional 3'-termini labelling of RNA/RZA transcripts by using T4 RNA ligase and the fluorescent compound pCp-Cy5-3' was carried out.^[21] To guarantee the specific Cy5-end-labelling of RNA, we evaluated various templates, including 104-nt antisense and sense DNA strands, 104-bp Cy3-dsDNA, and 87-nt RNA transcript in the ligation reaction. As shown in Figure S5b, under the same reaction conditions, only the RNA sample was successfully conjugated with Cy5-fluorophore dye, which confirms the specific and efficient labelling of the transcripts.

With this post-transcriptional labelling technique, the transcription of G-modified RZA products was re-evaluated. After Cy5-conjugation, all transcripts can be detected, albeit with significantly reduced yields compared to natural RNA (Figure S6a). The yield of G1-modified transcripts was higher (16% for G1, 9% for G1+C1, and 18% for G1+C2) compared to G2-modified ones (less than 2% for all cases).

The second factor that significantly affects transcription efficiency is the structure of initiating and elongating rGTP substrates. It has been shown that the rGTP plays a critical role in the transcription initiation process and T7 RNAP•DNA complex orientation.^[22–24] The purine N7 position is in particular importance for specific interactions with Arg632 of the enzyme.^[25,26] Not surprisingly, rGTP analogues are difficult substrates for T7 RNAP, explaining the limited number of publications using rGTP analogue in the RNA synthesis.^[27]

To improve the efficiency of *in vitro* transcription with modified rGTP, we employed guanosine monophosphate (rGMP) as an initiating substrate that can successfully replace rGTP during the transcription initiation.^[22] As shown in Figure S6b, the yield of fully modified transcripts was enhanced with the use of rGMP as an initiator (i.e., for G2+C2+A1+T1 from 0 to 11% and for G1+C2+A1+T1 from 5 to 16%). Interestingly, the 7-deaza-rGTP (rG1TP) was always preferable by T7 RNAP as a substrate compared with rITP (rG2TP) in any rZTP configuration (Figure 2d and S6). Previously, the K_m values for the initiating reaction by rGTP, rG1TP, and rG2TP were determined as 0.6, 0.75, and 1.4 mM, respectively,^[22,25] demonstrating the preference of rG1TP over rG2TP by T7 RNA polymerase. Inorganic pyrophosphatase (PPase) was also evaluated in order to enhance the transcription reactions;^[28] however, it did not affect the total RNA yield (Figure S6b).

To further improve the production of fully modified RNA composed of G1+C2+A1+T1 nucleotides, we used increasing concentrations of Mg^{2+} -ions (to a total of 18 mM), or T7 RNAP (to a total of 100 U), or by adding 10% dimethyl sulfoxide (DMSO) to the reaction mixture.^[29] Figure 2e shows that apart from rGMP, additional Mg^{2+} could also significantly improve the transcription yield by a factor of 20 compared with the control (reaction without additives). In addition, 10% DMSO enhanced the transcription efficiency 7-fold, whereas increasing T7 RNAP concentration did not affect the RZA synthesis.

To verify the base composition of the resulting RNA transcripts as well as DZA templates, the LC-MS/MS analysis was performed. The data shown in Figure S7 demonstrate that the modified nucleotides were successfully incorporated during RNA synthesis by T7 RNAP. The resulting fully modified transcript composed of G1+C2+A1+T1 contained no natural nucleosides. The sequence composition and transcription accuracy were further evaluated by sequencing (Figure S4c). The sequencing analysis of RZA after RT-PCR demonstrated that rZTP were accepted by T7 RNAP with a high degree of accuracy. However, the introduction of modified purines generated some mutations in the final fragments (mutation frequency in $\sim 0.2\text{--}0.6 \cdot 10^{-3}$ range, Table S2C).

***In vitro* transcription of modified DZA templates with modified rZTPs (DZA→RZA)**

In order to examine whether the flow of genetic information can be transferred from DZA to RZA, the *in vitro* transcription of modified dsDZA templates was carried out using the corresponding modified rZTPs. In the absence of modified guanosines in the dsDZA template and the rZTP set, the RZA synthesis was successfully achieved, even when three nucleotides were substituted both in the template and the rNTP set (Figure S8a). The relative efficiency of transcription with the A1+T1+C1 group was 13%, and 30% with the A1+T1+C2 group. In general, C2 substitution of cytidine in the template and rNTP set was preferred over C1 substitution by T7 RNAP (82% vs 71% for C2- and C1-modified groups, respectively, Figure S8a). As expected from previous results, the introduction of modified guanosines in dsDZA and rZTP dramatically reduced the yields of the final transcripts compared to natural RNA products (Figure S8b). As shown in Figure 2f, in the presence of rGMP in the reaction mixture, the fully modified transcripts comprising G2+C2+A1+T1 or G1+C2+A1+T1 were obtained with a relative yield of 14% or 3%, respectively. The sequencing data shown in Figure S4d and Table S2D confirmed correct, fully modified, product formation for both G2+C2+A1+T1 and G1+C2+A1+T1 groups.

In the case of DZA→RZA flow, T7 RNA polymerase should accept modifications in both the rGTP substrates and the promoter sequence, which are equally critical for the success of RNA synthesis.^[25] For these reasons, although to a limited extent, the production of fully modified RZA transcripts using dsDZA is noteworthy. Given the growing importance of nucleic acid therapeutics and the solid-phase synthesis of long modified fragments (in particular, RNA and mRNA) can be technically sophisticated and expensive, our findings take a step forward in the production of functional DNA/RNA material for biomedical research.

RESEARCH ARTICLE

Reverse transcription of modified RZAs with modified dZTPs (RZA→DZA)

To study whether RZA can be directly converted back to DZA, the fully modified RZA composed of G1+C2+A1+T1 was reverse transcribed using either Avian Myeloblastosis Virus (AMV) or Moloney Murine Leukemia Virus (M-MuLV) enzymes with a single or fully modified dZTPs and Cy3-labelled reverse primer (Figure 2g and Figure S9). The results demonstrate that the successful conversion of RZA transcripts to ssDZA fragments can be accomplished with M-MuLV reverse transcriptase. The efficiency of reverse transcription of all modified groups was similar to the unmodified control, even when all-four substitutions were introduced in the dZTP set (relative yield is 84%).

The analysis of sequencing results of the reverse transcription products (after PCR amplification with Taq pol and various dZTP sets) revealed that both enzymes, T7 RNAP and Taq pol, well-tolerated most of the single modifications in RZA templates and the corresponding rZTP/dZTP substrates (Figure S4e and Table S2E). As an exception, G substitution for G2 resulted in sequence disruption, including deletions and frameshifts in all analysed fragments. In general, G2 was somewhat more mutagenic over G1 in almost all studied enzymatic transformation: PCR amplification, *in vitro* transcription, and reverse transcription (Table S2). It is well known that inosine (G2) is a mutagenic nucleotide analogue used in error-prone PCR assay.^[30] Together with other mutagenic nucleotides, such as A1 and C2, G2 might accumulate a fatal amount of errors. This might be beneficial for developing new random mutagenesis approaches. In contrast, information encoded in a fully modified G1+C2+A1+T1 genetic polymer could overcome several enzymatic transformations with sufficient accuracy (Table S2E).

***In vivo* assay with modified genetic material**

Complete information circuit from DZA to RZA and back to DZA was demonstrated *in vitro*. The next question was whether the genetic information encoded in the DZA template could be functional *in vivo*. Previously, the accurate information transfer from the DZA template was illustrated using an antibiotic-resistant gene, R67 dihydrofolate reductase, that is a short gene of only 237-bp.^[5,6] Now, a fluorescent protein mTangerine was chosen as a model system for expression in *E.coli*.^[9] To this end, we first performed PCR synthesis of DZA fragments that embrace the 678-bp mTangerine gene. As such, the ribosome binding site (RBS) and start codon were also modified. The silent mutations were introduced into each DZA strand through the mutated primers in order to identify the final chimeric plasmids from the initial vector (pDZA_mTangerine and pmTangerine, respectively, Figure 3a and S10). The final 777-bp DZA genetic fragments were ligated to the ampicillin-resistant pJET1.2 vector followed by treatment with two different restriction enzymes, EcoRV and EcoRI, that hydrolyse the parental pmTangerine plasmid but do not act on the pDZA_mTangerine vector. The resulting chimeric vectors with various DZA genes were transformed into chemically competent bacterial cells (Figure 3a). The cultures were grown on media with ampicillin, and after incubation, the viable red and white colonies from different groups were counted and analysed by sequencing. Figure 3b-d shows that the red fluorescent colonies could be successfully detected from all modified samples. Similar to the *in vitro* studies, the pyrimidine-substituted genes (T1 and C1) were relatively more successfully transliterated than

other groups of modified templates. The A1-modified mTangerine genes had one of the lowest efficiencies of converting information. Presumably, due to high mutagenicity of the A1 modification as it was revealed from sequencing data shown in Figure 3e (details see in Table S3). All genetic samples containing 7-deazaA substitution possessed the highest number of mutations compared to other groups. Despite the incorporation of some mutations, most likely during PCR synthesis, highly modified genetic templates were successfully encoded for the functional protein. The bacterial replication machinery was able to recognise the DZA fragments providing fluorescent red colonies even when the entire gene, together with RBS and start codon, were substituted for the non-canonical analogues (G1+C1+A1+T1 gene resulted in 32% of relative colony formation, Figure 3d). The formation of white colonies was associated rather with cloning artifacts, such as incorrect gene orientation, immature or no PCR product insertion, than critical missense mutations that render the mTangerine protein unfunctional (Table S4).

Conclusion

The present life employs two informative polymers to encode for hereditary information, DNA and RNA, comprising of repeating units of four monomeric nucleotides, G, C, A, T or U, that are linked together by a universal motive, the phosphodiester backbone. To evaluate whether other chemistries could be involved in information flux, we performed the screening of alternative non-canonical nucleotides as building blocks for genetic biopolymers. The resulting polymers composed of the purine bases 7-deaza-adenine, 7-deaza-guanine or inosine together with the pyrimidine bases 5-chlorouracil, 5-fluorocytosine, or 5-bromocytosine (deoxy)ribonucleotides were denoted as DZA and RZA. Our study revealed that information encoded in entirely morphed informational polymers could be successfully propagated by natural polymerases. DZA can be replicated by Taq DNA polymerase, and RZA can be synthesised by T7 RNA polymerase, that, in their turn, can be reverse transcribed by the M-MuLV enzyme using all four modified dZTPs as substrates. Notably, bacterial replication machinery recognises the artificial DZA gene providing the functional fluorescent protein, even when all-four letters in the genetic fragment are substituted.

Chemical modification of nucleic acid diminishes a number of inherent drawbacks of natural biopolymers, including biological instability and limited chemical and structural diversity, making synthetic nucleic acids key targets of biomedical research. In particular, base-modified nucleic acids have already demonstrated their applicability in numerous studies. It was shown that the introduction of naturally occurring non-canonical nucleotides (e.g., pseudouridine, 2-thiouridine, and 5-MeC), reduced the immunogenicity of therapeutic mRNA while enhancing their efficacy and stability.^[10] It was also demonstrated that chemically modified genes (containing glucosyl-hydroxymethylcytosine) are invisible for an attack by the CRISPR-Cas9 system that might be useful in genome engineering.^[31] Additionally, *in vitro* selection technology of affinity aptamers and catalytic nucleic acids significantly benefited from chemical alteration of initial libraries.^[12] Multiple chemical substitutions of aptamers and nucleic acid enzymes markedly enhance their potency, i.e., catalytic and binding efficiency.

RESEARCH ARTICLE

Figure 3. Analysis of DZA genetic material *in vivo*. **a.** General scheme of *in vivo* studies. PCR amplification of pmTangerine plasmid (encoding a red fluorescent protein, mTangerine) was performed to produce various 777-bp DZA genetic fragments. The resulting fragments were assembled in ampicillin-resistant vector (pJET1.2), transformed in *E. coli*, and grown in LB with ampicillin. To eliminate the parental pmTangerine plasmid, the EcoRV and EcoRI cleavage of ligation products was performed. The viable red and white colonies were counted, and plasmids were analysed by sequencing to identify unique signatures (i.e., silent mutations introduced in the synthetic DZA fragments) in each DNA strand. **b.** Expression of mTangerine in *E. coli* transformed with the chimeric pDZA_mTangerine plasmid. The examples of LB agar plates containing ampicillin (200 mg L⁻¹) are shown. The plates were scanned under Cy3 channel. **c.** The number of red colonies normalised to 50 μ L of culture. **d.** The normalised ratio of red colonies to the total number of viable colonies (red and white) on the plate. **e.** Sequencing analysis of the 777-bp fragments. The mutation frequency was found in the 777-bp unmodified and modified dsDNA/dsDZA fragments. Mutation frequency [mutation/base in %] calculated as [the total number of mutations / (number of colonies $\times$ the length of fragments)] $\times$ 100%. The data are collected from four independent experiments. The composition of the modified dsDZA fragments is shown on the X-axis. PC refers to the positive control, natural dsDNA fragment. Sequencing data are presented in Supporting Information Table S3.

Nevertheless, despite the remarkable advances that nucleic acid alteration can offer, there are no reports yet involving fully modified systems as a genetic replacement of natural polymers in biological processes.

The chemical modifications of DNA and RNA presented here do not disturb the base-pairing structure and replication/transcription functions of natural polymers, whereas broadening their chemical space. Halogen atoms, like F, Cl, and Br, were excluded from the evolution of natural biopolymers (nucleic acids, proteins, and polysaccharides) though, they are valuable atoms in medicinal chemistry design.^[32] Halogenation engenders changes in the interaction of biopolymers with their molecular targets, e.g., improving ligand-binding affinities and substrate selectivity. Halogen-substituted pyrimidines are potent drugs.^[33] They are excellent substrates for cellular polymerases and can be incorporated into the bacterial or eukaryotic genome.^[34] Furthermore, the successful selection and efficacy of 5-fluorouracil comprising aptamer as anticancer prodrug has been demonstrated.^[35] The modified purines used in the study are also established biomedical tools. Derivatives and isomers of 7-

deazaA (e.g., Tubercidin and Remdesivir) hold high antiviral potency,^[36] while inosine and 7-deazaG are widely used in molecular biology approaches, for example, to improve PCR amplification and sequencing of GC-rich templates (7-deaza-dGTP)^[37] or to enhance the error rate in PCR-based random mutagenesis (dITP).^[30] Taken together, these observations provide compelling evidence that DZA and RZA should be considered as a potent tool in biomedical and genomic research.

In summary, this study confirms that DZA genetic templates could be successfully replicated *in vitro* and serve as genetic templates *in vivo*.^[5,6] In addition to the successful transliteration (DZA $\leftrightarrow$ DNA $\rightarrow$ RNA), this study involved fully modified RNA in the context of information transfer (Figure 1). Both *in vitro* transcription (DNA/DZA $\rightarrow$ RNA/RZA) and reverse transcription (RNA/RZA $\rightarrow$ DNA/DZA) have been demonstrated. Alternative genetic polymers, RZA and DZA, were fully biocompatible with natural polymerases. Thus, the synthetic 'DZA $\rightarrow$ RZA $\rightarrow$ DZA' information flow opens a new avenue to re-engineer the chemistry of life using natural or evolved polymerases. To complete the central dogma steps, the RZA translation yet needs to be verified.

RESEARCH ARTICLE

As we advance, the beneficial characteristics of DZA and RZA biomolecules can make them a novel tool not only for *in vitro* or *in vivo* expression of therapeutic proteins and vaccination but also for gene manipulations (e.g., using CRISPR-Cas9 system) and oligonucleotide drug design.

Acknowledgements

This work was supported by the KU Leuven Research Fund (OT/14/128). We wish to acknowledge the Fonds Wetenschappelijk Onderzoek (FWO, Flanders Research Foundation) [G.078014N to P.H., 12Q8619N to E.E.]. H.Y. thanks the China Scholarship Council (CSC) for funding (Grant 201707040069). The authors are indebted to Prof. Jef Rozenski for LC-MS/MS analysis and Chantal Biernaux for editorial help.

Keywords: artificial gene • non-canonical nucleotides • nucleic acids • synthetic biology • transcription

- [1] F. Crick, *Nature* **1970**, 227, 561–563.
- [2] a) R. Saiki, S. Scharf, F. Faloona, K. Mullis, G. Horn, H. Erlich, N. Arnheim, *Science* **1985**, 230, 1350–1354; 2 b) B. Alberts, A. Johnson, J. Lewis, M. Raff, K. Roberts, P. Walter, *Mol. Biol. Cell* **2002**, 1–16.
- [3] A. J. Simon, S. D'Oelsnitz, A. D. Ellington, *Nat. Biotechnol.* **2019**, 37, 730–743.
- [4] a) E. Eremeeva, P. Herdewijn, *Curr. Opin. Biotechnol.* **2019**, 57, 25–33; b) H. Grosjean, in *DNA RNA Modif. Enzym. Struct. Mech. Funct. Evol.* (Ed.: Henri Grosjean), Landes Bioscience, **2009**, pp. 1–18.
- [5] E. Eremeeva, M. Abramov, L. Margamuljana, J. Rozenski, V. Pezo, P. Marlière, P. Herdewijn, *Angew. Chemie Int. Ed.* **2016**, 55, 7515–7519.
- [6] E. Eremeeva, M. Abramov, L. Margamuljana, P. Herdewijn, *Chem. - A Eur. J.* **2017**, 23, 9560–9576.
- [7] E. Eremeeva, M. Abramov, P. Marlière, P. Herdewijn, *Org. Biomol. Chem.* **2017**, 15, 168–176.
- [8] a) H. Maruyama, K. Furukawa, H. Kamiya, N. Minakawa, A. Matsuda, *Chem. Commun.* **2015**, 51, 7887–7890; b) V. Pezo, F. W. Liu, M. Abramov, M. Froeyen, P. Herdewijn, P. Marlière, *Angew. Chemie Int. Ed.* **2013**, 52, 8139–8143; c) C. Liu, C. Cozens, F. Jaziri, J. Rozenski, A. Maréchal, S. Dumbre, V. Pezo, P. Marlière, V. B. Pinheiro, E. Groaz, et al., *J. Am. Chem. Soc.* **2018**, 140, 6690–6699; d) M. Luo, E. Groaz, M. Froeyen, V. Pezo, F. Jaziri, P. Leonczak, G. Schepers, J. Rozenski, P. Marlière, P. Herdewijn, *J. Am. Chem. Soc.* **2019**, 141, 10844–10851; e) H. Nguyen, M. Abramov, E. Eremeeva, P. Herdewijn, *ChemBiochem* **2020**, 21, 272–278; f) N. Minakawa, M. Sanji, Y. Kato, A. Matsuda, *Bioorg. Med. Chem.* **2008**, 16, 9450–9456; g) M. Flamme, L. K. McKenzie, I. Sarac, M. Hollenstein, *Methods* **2019**, 161, 64–82; h) V. Raindlová, M. Janoušková, M. Slavičková, P. Perlíková, S. Boháčová, N. Milisavljevič, H. Šanderová, M. Benda, I. Barvík, L. Krásný, et al., *Nucleic Acids Res.* **2016**, 44, 3000–3012.
- [9] N. C. Shaner, R. E. Campbell, P. A. Steinbach, B. N. G. Giepmans, A. E. Palmer, R. Y. Tsien, *Nat. Biotechnol.* **2004**, 22, 1567–1572.
- [10] a) K. Karikó, H. Muramatsu, F. A. Welsh, J. Ludwig, H. Kato, S. Akira, D. Weissman, *Mol. Ther.* **2008**, 16, 1833–1840; b) M. S. D. Kormann, G. Hasenpusch, M. K. Aneja, G. Nica, A. W. Flemmer, S. Herber-Jonat, M. Huppmann, L. E. Mays, M. Illenyi, A. Schams, et al., *Nat. Biotechnol.* **2011**, 29, 154–159; c) K. Bahl, J. J. Senn, O. Yuzhakov, A. Bulychyev, L. A. Brito, K. J. Hassett, M. E. Laska, M. Smith, Ö. Almarsson, J. Thompson, et al., *Mol. Ther.* **2017**, 25, 1316–1327.
- [11] H. Peacock, A. Kannan, P. A. Beal, C. J. Burrows, *J. Org. Chem.* **2011**, 76, 7295–7300.
- [12] a) B. N. Gawande, J. C. Rohloff, J. D. Carter, I. von Carlowitz, C. Zhang, D. J. Schneider, N. Janjic, *Proc. Natl. Acad. Sci.* **2017**, 114, 2898–2903; b) L. Ma, J. Liu, *iScience* **2020**, 23, 100815; c) E. Eremeeva, P. Herdewijn, in *Enzym. Chem. Synth. Nucleic Acid Deriv.* (Ed.: J. Fernández-Lucas), Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim, Germany, **2018**, pp. 159–194; d) M. Hollenstein, *Curr. Opin. Chem. Biol.* **2019**, 52, 93–101.
- [13] T. Kojima, K. Furukawa, H. Maruyama, N. Inoue, N. Tarashima, A. Matsuda, N. Minakawa, *ACS Synth. Biol.* **2013**, 2, 529–536.
- [14] Y. Kawase, S. Iwai, H. Inoue, K. Miura, E. Ohtsuka, *Nucleic Acids Res.* **1986**, 14, 7727–7736.
- [15] a) C. Bailly, *Nucleic Acids Res.* **1998**, 26, 4309–4314; b) V. Boudou, L. Kerremans, B. De Bouvere, E. Lescriner, G. Schepers, R. Busson, A. Van Aerschot, P. Herdewijn, *Nucleic Acids Res.* **1999**, 27, 1450–1456.
- [16] I. V. Kutyavin, *Biochemistry* **2008**, 47, 13666–13673.
- [17] J. F. Milligan, D. R. Groebe, G. W. Witherell, O. C. Uhlenbeck, *Nucleic Acids Res.* **1987**, 15, 8783–8798.
- [18] a) V. Potapov, X. Fu, N. Dai, I. R. Corrêa, N. A. Tanner, J. L. Ong, *Nucleic Acids Res.* **2018**, 46, 5753–5763; b) J. Huang, L. G. Briebe, R. Sousa, *Biochemistry* **2000**, 39, 11571–11580.
- [19] a) L. Celewicz, *J. Photochem. Photobiol. B Biol.* **1995**, 30, 129–137; b) M. Zdrowowicz, P. Wityk, B. Michalska, J. Rak, *Org. Biomol. Chem.* **2016**, 14, 9312–9321; c) S. Adhya, P. Sarkar, D. Valenzuela, U. Maitra, *Proc. Natl. Acad. Sci. U. S. A.* **1979**, 76, 1613–1617; d) C. Baugh, D. Grate, C. Wilson, *J. Mol. Biol.* **2000**, 301, 117–128.
- [20] a) F. Seela, A. Röling, *Nucleic Acids Res.* **1992**, 20, 55–61; b) Z. Dudová, J. Špaček, L. Havran, H. Pivoňková, M. Fojta, *Monatshefte für Chemie* **2016**, 147, 13–20.
- [21] H. Wang, R. A. Ach, B. O. Curry, *RNA* **2007**, 13, 151–159.
- [22] C. T. Martin, J. E. Coleman, *Biochemistry* **1989**, 28, 2760–2762.
- [23] Y. Jia, S. S. Patel, *J. Biol. Chem.* **1997**, 272, 30147–30153.
- [24] W. P. Kennedy, J. R. Momand, Y. W. Yin, *J. Mol. Biol.* **2007**, 370, 256–268.
- [25] I. Kuzmine, P. A. Gottlieb, C. T. Martin, *J. Biol. Chem.* **2003**, 278, 2819–2823.
- [26] M. L. Gleghorn, E. K. Davydova, R. Basu, L. B. Rothman-Denes, K. S. Murakami, *Proc. Natl. Acad. Sci. U. S. A.* **2011**, 108, 3566–3571.
- [27] N. Milisavljevič, P. Perlíková, R. Pohl, M. Hocek, *Org. Biomol. Chem.* **2018**, 16, 5800–5807.
- [28] P. R. Cunningham, J. Ofengand, *Biotechniques* **1990**, 9, 713–714.
- [29] Z. Chen, Y. Zhang, *Biochem. Biophys. Res. Commun.* **2005**, 333, 664–670.
- [30] J. H. Spee, W. M. de Vos, O. P. Kuipers, *Nucleic Acids Res.* **1993**, 21, 777–778.
- [31] A. L. Bryson, Y. Hwang, S. Sherrill-Mix, G. D. Wu, J. D. Lewis, L. Black, T. A. Clark, F. D. Bushman, *MBio* **2015**, 6, e00648.
- [32] a) M. Hernandez, S. M. Cavalcanti, D. R. Moreira, W. de Azevedo Junior, A. C. Leite, *Curr. Drug Targets* **2010**, 11, 303–314; b) L. Mendez, G. Henriquez, S. Sirimulla, M. Narayan, *Molecules* **2017**, 22, 1397.
- [33] a) N. Tsesmetzis, C. B. J. Paulin, S. G. Rudd, N. Herold, *Cancers (Basel)* **2018**, 10, 240; b) K. L. Seley-Radtke, M. K. Yates, *Antiviral Res.* **2018**, 154, 66–86.
- [34] a) S. M. Morris, *Mutat. Res. Genet. Toxicol.* **1993**, 297, 39–51; b) P. Marlière, J. Patrouix, V. Döring, P. Herdewijn, S. Tricot, S. Cruveiller, M. Bouzon, R. Mutzel, *Angew. Chemie Int. Ed.* **2011**, 50, 7109–7114.
- [35] S. Kruspe, U. Hahn, *Angew. Chemie Int. Ed.* **2014**, 53, 10541–10544.
- [36] a) G. Acs, E. Reich, M. Mori, *Proc. Natl. Acad. Sci. U. S. A.* **1964**, 52, 493–501; b) M. K. Lo, R. Jordan, A. Arvey, J. Sudhamsu, P. Shrivastava-Ranjan, A. L. Hotard, M. Flint, L. K. McMullan, D. Siegel, M. O. Clarke, et al., *Sci. Rep.* **2017**, 7, 43395; c) S. Vittori, D. Dal Ben, C. Lambertucci, G. Marucci, R. Volpini, G. Cristalli, *Curr. Med. Chem.* **2006**, 13, 3529–3552.
- [37] N. Guido, E. Starostina, D. Leake, I. Saaem, *PLoS One* **2016**, 11, e0156478.

RESEARCH ARTICLE

Entry for the Table of Contents

Alternative genetic information flow encoded in four non-canonical nucleotides can be carried out by natural biological systems. Natural polymerases can synthesize the entirely morphed DNA (DZA) using DNA template that, in turn, can serve as a template to produce a fully modified RNA (RZA), and be reverse transcribed back to DZA. In bacteria, various DZA genes can successfully encode for a functional fluorescent protein.