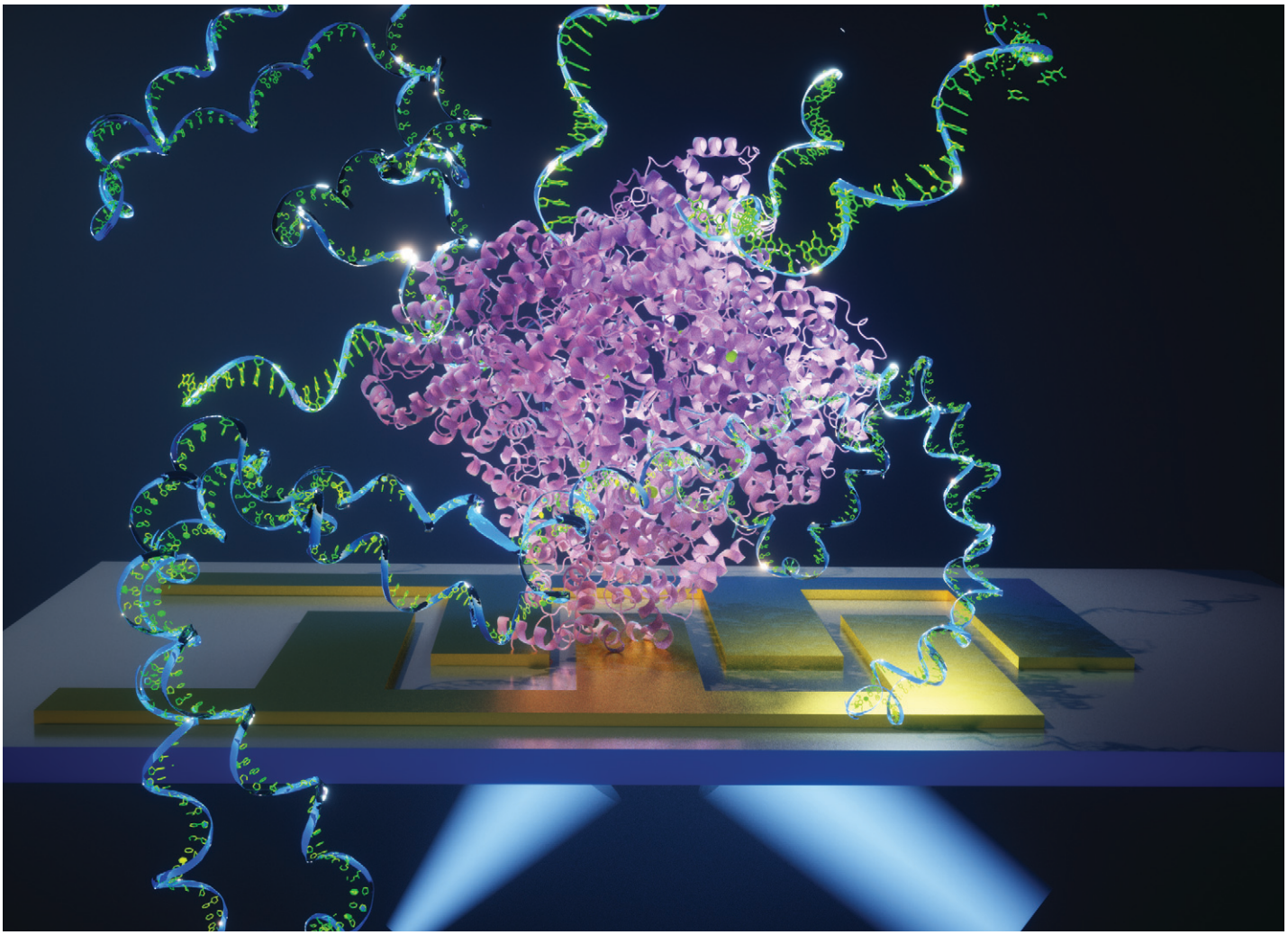


**Showcasing research from Professor Matej Butala's laboratory, Biotechnical Faculty, University of Ljubljana, Slovenia and the Marcin Nowotny's Laboratory of Protein Structure, International Institute of Molecular and Cell Biology in Warsaw, Poland.**

Data storage based on the absence of nucleotides using a bacteriophage abortive infection system reverse transcriptase

DNA molecules are a promising data storage medium for the future; however, effective de novo synthesis of DNA using an enzyme remains a challenge. We demonstrate that the polymerase AbiK from *Lactococcus lactis* facilitates such an approach. Using surface plasmon resonance and AbiK, DNA with segments of random length and a sequence consisting of only three of the four natural nucleotides can be synthesized. The information is encoded using the absence of one nucleotide in each segment. Our setup holds great potential for synthesizing DNA for data storage. *Artwork design by Ella Maru Studio.*

## As featured in:


See Matej Butala *et al.*, *Lab Chip*, 2025, **25**, 113.

ROYAL SOCIETY OF CHEMISTRY

rsc.li/loc

Registered charity number: 207890

Lab on a Chip

COMMUNICATION

View Article Online
View Journal | View Issue



Cite this: *Lab Chip*, 2025, **25**, 113

Received 11th September 2024,
Accepted 18th November 2024

DOI: 10.1039/d4lc00755g

rsc.li/loc



# Data storage based on the absence of nucleotides using a bacteriophage abortive infection system reverse transcriptase†

Gregor Bajc,[a] Anja Pavlin,[a] Małgorzata Figiel,[b] Weronika Zajko,[b] Marcin Nowotny[b] and Matej Butala *[a]

**DNA molecules are a promising data storage medium for the future; however, effective *de novo* synthesis of DNA using an enzyme that catalyzes the polymerization of natural nucleoside triphosphates in a user-defined manner, without the need for multiple injections of polymerase, remains a challenge. In the present study, we demonstrated that the bacteriophage abortive infection system reverse transcriptase AbiK from *Lactococcus lactis* facilitates such an approach. We employed surface plasmon resonance to monitor the polymerization of the DNA strand with a user-defined sequence of multiple segments through a sequential buffer exchange process. Using this method, we synthesized synthetic DNA with segments of random length and a sequence consisting of only three of the four natural nucleotides. The information is encoded using the absence of one nucleotide in each segment. We demonstrated that synthetic DNA can be stored on the chip, and when the DNA is released from the chip, the second strand can be synthesized and read by sequencing. Our setup facilitates a writing speed of one nucleotide in less than 1 s and holds enormous potential for synthesizing DNA for data storage.**

## Introduction

DNA is a sustainable data storage medium owing to its stability and remarkably high storage density.[1,2] Sequencing methods facilitate efficient access to the information stored in DNA; however, technological innovations for efficient DNA synthesis are still required.[3–5] A technique that has been used to synthesize DNA since the 1980s is phosphoramidite oligonucleotide synthesis.[6,7] Nowadays, this process is conducted in a microarray format, which increases productivity and allows the simultaneous synthesis of polynucleotide chains, each up to 200 nucleotides in length.[8] These chemically produced DNA fragments are assembled using Gibson assembly or PCR assembly into larger DNA fragments.[9,10] Phosphoramidite chemistry requires long synthesis cycles of stepwise addition of building blocks derived from 5′-protected-dimethoxytrityl nucleoside phosphoramidites and the use of expensive chemical reagents and organic solvents is not environment friendly. Enzymatic DNA synthesis emerged as a promising alternative that reduces the dependance on chemical synthesis, offering a potentially more cost-effective and environment-friendly approach to achieve DNA data storage.[3,4] In recent years, enzymatic DNA synthesis using terminal deoxynucleotidyl transferase (TdT) has been introduced as a template-independent approach.[2,3,11] Typically, the addition of selected nucleotides to a growing single-stranded (ssDNA) polymer is regulated using either nucleotides that are reversibly linked to TdT or engineered reversible terminator nucleotides and engineered TdT, thereby preventing subsequent coupling of the next nucleotide until further treatment.[2] To our knowledge, there are only two approaches that use natural nucleotides to synthesize a custom DNA sequence using TdT. One of the approaches utilizes the competition between TdT and the enzyme apyrase to facilitate data storage in ssDNA strands with short homopolymeric extensions.[12] The apyrase in the reaction mixture degrades the nucleoside triphosphates to limit the number of nucleotides with the same base per reaction step. In the other approach, enzymatic DNA polymerization by TdT is controlled by the uptake of the $Co^{2+}$ cofactor, which facilitates template-independent multiplexed DNA synthesis.[13] In both synthesis approaches, the demanding reaction conditions and the need to add the enzyme(s) at each cycle of the synthesis process increase the costs.

Here, we present an innovative approach based on the catalytic activity of AbiK, a reverse transcriptase of the phage abortive infection system of *Lactococcus lactis*. AbiK, a plasmid-encoded protein comprising 599 amino acids (molecular size, 71.4 kDa; isoelectric point, 7.98), is active in

[a] *Department of Biology, Biotechnical Faculty, University of Ljubljana, 1000 Ljubljana, Slovenia. E-mail: matej.butala@bf.uni-lj.si*
[b] *Laboratory of Protein Structure, International Institute of Molecular and Cell Biology in Warsaw, 02-109 Warsaw, Poland*

† Electronic supplementary information (ESI) available. See DOI: https://doi.org/10.1039/d4lc00755g

its homohexameric form.[14–16] It catalyzes the protein-primed synthesis of long, ssDNA strands with random sequences—independent of a DNA template—which confer resistance to phage infection.[17] The first nucleotide is covalently bound to the priming tyrosine in the amino-terminal region of the finger subdomain, and subsequent nucleotides are incorporated with equal frequency.[16,18] To our knowledge, previous studies have not demonstrated the ability of AbiK to synthesize oligonucleotides or polynucleotides in a controlled manner. In the present study, we used surface plasmon resonance (SPR) for sequential buffer exchange to control the DNA polymerization activity of AbiK in real time using natural nucleoside triphosphates. The results indicate that this system can be used to achieve the template-independent synthesis of ssDNA carrying a custom order of segments with sequences composed of selected three of the four nucleotides. This unique encoding method facilitates the information storage on the basis of the absence of one of the four nucleotides in each segment. We illustrate the technique by writing the word "DNA" and using a unique codec for the letters of the alphabet, which are written in nucleobases.

## Materials and methods

### Immobilization of AbiK onto a CM5 sensor chip

SPR measurements were performed at the Infrastructural Center for the Analysis of Molecular Interactions at the Department of Biology, University of Ljubljana, on a Biacore T200 system (GE Healthcare, Chicago, Illinois, USA) at 25 °C. AbiK (2.7–3.3 mg $mL^{-1}$) was isolated using a previously described protocol,[16] dissolved in sodium acetate buffer (pH 5.0) to a final concentration of 50 μg $mL^{-1}$, and then immobilized on a carboxymethylated dextran matrix (CM5) chip (Cytiva) by amine coupling. As shown in ESI† Fig. S1, the surface of the CM5 sensor chip was activated with a 12 min injection of a mixture of *N*-hydroxysuccinimide and 1-ethyl-3-(3-dimethylpropyl)-carbodiimide following manufacturer's instructions. Next, 50 μg $mL^{-1}$ AbiK diluted in 10 mM sodium acetate (pH 5.0) was applied on flow cell 2 or 4. Flow cells 1 and 3 were left unmodified and served as a control for non-specific binding of the test compounds. The sensorgrams shown in the results (shown in Fig. 1 and 2 and in ESI† Fig. S1) depict the responses obtained on flow cell 2 or 4, respectively, to the responses obtained on flow cell 1 or 3. The remaining active groups on the surface of both flow cells were deactivated with a 10 min injection of ethanolamine. For experiments in which the synthesized ssDNA was collected for sequencing, AbiK was immobilized on all four flow cells. The quantity of immobilized enzyme for each experiment was approximately 20 000 response units. The data were analyzed using the Biacore T200 evaluation software (GE Healthcare, Chicago, Illinois, USA).

### On-chip synthesis of ssDNA

For ssDNA synthesis, either single dNTPs (deoxyadenosine triphosphate [dATP], deoxycytidine triphosphate [dCTP], deoxythymidine triphosphate [dTTP], or deoxyguanosine triphosphate [dGTP]) or a mixture of dNTPs was injected over the immobilized AbiK on the chip. The dNTPs were diluted in a running buffer (20 mM Tris, pH 8.3; 140 mM NaCl; 2 mM $MgCl_2$; 0.005% P20) to a concentration of 100 μM and injected over AbiK at a flow rate of 10–40 μL $min^{-1}$ for the selected injection time (10–200 s). Between each injection of nucleotides, the system was washed with the running buffer using 15–30 s pulses. Polymerization reactions were performed at 25 °C. The nuclease Benzonase (Sigma, USA) was injected to degrade the synthesized DNA and regenerate the sensor chip surface with immobilized AbiK. Benzonase was injected over the chip at a concentration of 20 or 150 enzyme units (U) per mL at a flow rate of 10 μL $min^{-1}$ for up to 600 s as marked on the sensorgrams. To determine whether a DNA strand carrying ordered stretches of a selected nucleotide—dATP, dTTP, or dTGP—can be obtained with AbiK immobilized on the CM5 chip, 18-mer oligonucleotides carrying adenosine, thymidine, or cytosine bases (Microsynth, Austria; the poly-G sequence cannot be purchased) were injected over the chip at a concentration of 1 μM at a flow rate of 10 μL $min^{-1}$.

To store the information in synthetic DNA, we created a codec in which each letter of the alphabet is encoded by a unique combination of three DNA segments, with each segment containing only three of the four natural dNTPs (ESI† Table S1). Next, we prepared four 100 μM mixtures containing three dNTPs in the running buffer, in which either dATP, dCTP, dGTP, or dTTP was missing. Each mixture of nucleotides was injected at a flow rate of 40 μL $min^{-1}$ as a substrate for AbiK to generate an ssDNA with the desired sequence encoding the word "DNA". The word was written twice, and a homopolymeric DNA segment containing only guanine nucleobases followed the word, which facilitated the binding of an 18-mer poly-C oligonucleotide and subsequently the synthesis of the complementary strand using the Klenow fragment of *Escherichia coli* DNA polymerase I. At the beginning and end of the experiment, a mixture of all four nucleotides (100 μM each) was injected to generate short flanking sequences.

### Sequencing of the DNA fragments generated by AbiK

After DNA synthesis was complete, we removed the SPR chip containing the synthetic ssDNA covalently bound to AbiK from the SPR apparatus and treated the surface of the chip with proteinase K (0.4 mg $mL^{-1}$ in 50 μL EquiPhi29™ DNA polymerase buffer) at 55 °C for 1 h. After scoring the chip surface with a pipette tip, we transferred the mixture to a microcentrifuge tube and added double-distilled water to replace the evaporated water (to a volume of 50 μL). Proteinase K was then inactivated by incubation at 95 °C for 15 min, after which the mixture was cooled on ice for 5 min. To synthesize the second strand, 0.8 μM 18-mer poly-C oligonucleotides, 0.2 mM dNTPs, and the Klenow fragment reaction buffer at the working concentration were added. Then, the mixture was heated to 95 °C for 2 min and cooled on ice for 5 min. Next, the Klenow fragment lacking both the

3′ → 5′ and 5′ → 3′ exonuclease activity of DNA polymerase I (Thermo Fisher Scientific, Waltham, MA USA) was added at a concentration of 0.25 U $\mu L^{-1}$, and the mixture was incubated at 25 °C for 15 min and then at 37 °C for 1 h. The Klenow fragment was then thermally inactivated by incubation at 75 °C for 15 min. The mixtures were sequenced by Microsynth Seqlab (Göttingen, Germany) using Oxford Nanopore Technology (Full PlasmidSeq Service) after which we received the raw sequencing data.

## Results and discussion

### AbiK immobilized on the chip polymerizes DNA

The SPR technique is used to study the template-dependent polymerization of nucleotides;[19,20] therefore, we tested whether the SPR technique can be used for sequential buffer exchange to control DNA synthesis and follow it in real time. In our strategy, we immobilized the enzyme AbiK on the surface of a CM5 SPR chip using amine coupling. We chose AbiK as a catalyst for *de novo* DNA synthesis because: (i) it catalyzes protein-primed ssDNA synthesis, and no oligonucleotides are thus required to initiate the reaction; (ii) the growing ssDNA strand is covalently bound to AbiK; (iii) it efficiently elongates the polynucleotide chain independent of a template without favoring one natural nucleoside triphosphate over another.[16]

The SPR technology measures the change in refractive index at the surface of the SPR chip upon the binding of molecules, which is expressed in response units (RUs). We immobilized approximately 20 000 RU of AbiK, which corresponds to ~20 ng of protein per square millimeter of the SPR chip surface. This estimate is based on a previous finding that a 1000 RU corresponds to a density of 1 ng $mm^{-2}$ for globular proteins.[19] To test whether the immobilized enzyme retains DNA synthesis activity and whether it accepts each of the four natural nucleotides, we injected 100 μM dATP, dCTP, dTTP, or dGTP, respectively, with a wash step between each injection. The injection of each nucleoside triphosphate resulted in an increase in the measured RU (~80 to 250 RU per nucleotide), which was due to an increase in the local mass on the surface of the SPR chip, indicating that the nucleotides had been incorporated into the DNA chain (Fig. 1a). After the dNTPs were injected, the response remained stable, suggesting that the nucleoprotein complex detached only minimally from the chip. The RU signal decreased by 50% after the injection of benzonase (20 U $mL^{-1}$ for 600 s), indicating nuclease activity.

We used SPR to examine whether the AbiK enzyme generates homopolymeric segments of ssDNA. We injected 18-mer homopolymeric oligonucleotides containing cytosine, adenine, or thymine monomers (oligo-$C_{18}$, oligo-$A_{18}$, and oligo-$T_{18}$, respectively) both before and after the injection of the selected dNTP. We speculated that if AbiK synthesized ssDNA carrying a custom sequence of homopolymeric segments, the oligonucleotides would not anneal until a complementary strand was synthesized by AbiK. Indeed, the present findings suggest that oligo-$C_{18}$, oligo-$A_{18}$, or oligo-$T_{18}$

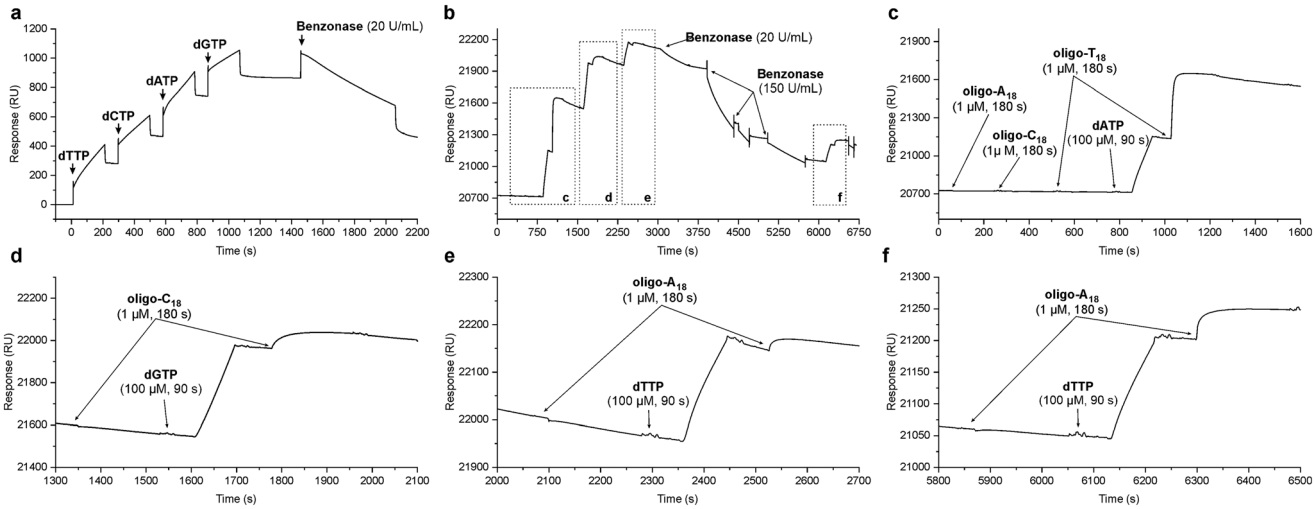
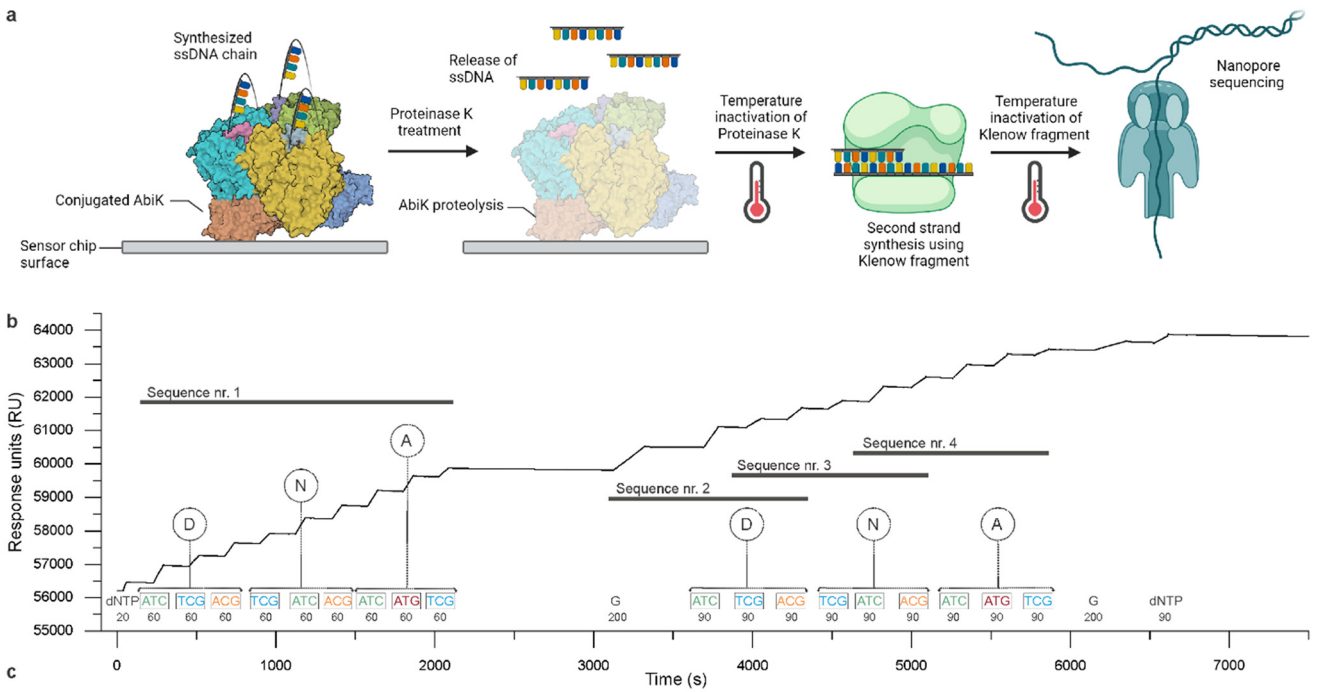



**Fig. 1** Template-independent single-stranded DNA synthesis using the bacteriophage abortive infection system reverse transcriptase AbiK immobilized on a CM5 sensor chip. (a) Sensorgram showing the polymerization of deoxyadenosine triphosphate (dATP), deoxycytidine triphosphate (dCTP), deoxythymidine triphosphate (dTTP), or deoxyguanosine triphosphate (dGTP) by the AbiK enzyme immobilized on the surface of the CM5 chip. Approximately 20 000 RU of AbiK was immobilized (set as 0 RU). Nucleotides were injected at indicated time points at a concentration of 100 μM for 200 s at a flow rate of 10 μL $min^{-1}$. The sensor chip surface with immobilized AbiK was regenerated by injecting the nuclease benzonase, which degraded the synthesized DNA. The graph depicts a 600 s injection of 20 enzyme units (U) per mL of benzonase. (b) Sensorgram of the injection of selected dNTPs over AbiK immobilized on the chip (approximately 21 000 RU). The dNTPs were injected at a concentration of 100 μM for 90 s at a flow rate of 10 μL $min^{-1}$. At the indicated time points, oligonucleotides oligo-$A_{18}$ (5′-AAAAAAAAAAAAAAAAAAAAA-3′), oligo-$C_{18}$ (5′-CCCCCCCCCCCCCCCCCCCC-3′), or oligo-$T_{18}$ (5′-TTTTTTTTTTTTTTTTTTTT-3′) were sequentially injected over the DNA synthesized by AbiK. The sensorgram in panel (b) depicts the complete cycle, and the marked regions of the sensorgram are shown enlarged in panels (c–f). All experiments were performed in a running buffer (containing 20 mM Tris [pH 8.3], 140 mM NaCl, 2 mM $MgCl_2$, and 0.005% P20) at a flow rate of 10 μL $min^{-1}$. All experiments were performed in triplicate, and the representative sensorgrams are shown.

hybridized to the ssDNA generated by AbiK only when the enzyme was prompted to generate a homopolymer with a sequence complementary to the injected oligonucleotide (Fig. 1b–f). This was signified by an increase in the RU after the injection of a selected oligonucleotide. The injection of benzonase (20–150 U $mL^{-1}$) eliminated most of the DNA synthesized by AbiK; however, AbiK remained active after benzonase treatment, as indicated by the fact that the binding of oligo-$A_{18}$ was observed only after dTTP was injected (Fig. 1b and f). Overall, these results suggest that the SPR system can be used to deliver selected deoxynucleotides to AbiK immobilized on the chip, which catalyzes the polymerization of homopolymeric nucleotide segments in a user-defined manner.

## Encoding the information in DNA using AbiK

Next, we tested whether the AbiK-SPR system facilitates the writing of digital information in DNA. However, we have not yet developed the method to a point where controlled writing by AbiK with single base resolution is possible. Moreover, encoding information in ssDNA carrying a custom order of different nucleotide homopolymers may lead to the formation of secondary DNA structures, which may impair the activity of the enzyme. Therefore, we developed a unique code for alphabet letters (ESI† Table S1). Then, we injected various mixtures of dNTPs—containing the selected three nucleotides in a particular order—over AbiK immobilized on the chip (Fig. 2b). The




c

**Sequence nr. 1:**

5'-AATATCAAAAAAATATCCAAACAAAATATACTCACACATATAAACCTGTGTGTTGCGTGGAAAGAGAAAGAAGAAAGAGCGCGCGCGCTGTGCGTC
CCTCGCTGTGTCACAAAATACCCAATACTCAAATAATAAACAAAATCAATAAAATAAAAAGGAAGAACGAAAGAGCGAAGAGAGAGAAAGAAAAGAGA
AAGAAGAATCTTTTACAACACCCAAAATTAATACCCATATATCAAACCATATATAAACACAACATAAGAGTGAGAGAGTTTGAAGGAGAAGGAAGGAA
GGATGGATAAGAATGGGAGAGAGAAATGAATAGGTAGAAATAATTGGGAAAAAAGAATAAAAAGAAGTTAGAATAAAGATGAAGTGAAAGTTGAGAAG
AGAGAGAAAGTGAAATGAGAGGGAGGGAGAATAATAAGAGAGTAAAAAGGAGCGCGGCTAG-3'

**Fig. 2** Encoding information in DNA using the bacteriophage abortive infection system reverse transcriptase AbiK. (a) Schematic representation of the workflow depicting the synthesis of DNA with the sequence encoding the word "DNA" twice, the release of single-stranded DNA (ssDNA) from AbiK by proteinase K treatment followed by the temperature inactivation of the protease, the synthesis of the second DNA strand by the Klenow fragment, temperature inactivation of the Klenow fragment, and the nanopore sequencing of the synthesized DNA (created with BioRender, **https://www.biorender.com/**). (b) Sensorgram depicting the polymerization of ssDNA encoding "DNA" and "DNA", followed by the injection of dNTPs or dGTP (G). Four different nucleotide mixtures containing three of the four nucleotides (dATP, dTTP, dGTP, or dCTP; mixtures: ACG, TCG, ATC, or ATG) were injected over immobilized AbiK (~20 000 RU) in a sequence encoding the word "DNA". A dilution of the single nucleotide dGTP was injected in the middle and at the end of the custom sequence. A mixture of all four nucleotides (dNTPs) was injected at the beginning and end of the experiment. All nucleotides or nucleotide mixtures were injected over AbiK at a concentration of 100 μM and a flow rate of 40 μL $min^{-1}$ for the number of seconds indicated under the marked nucleotide or nucleotide mixture in the sensorgram. The experiment was performed in a running buffer containing 20 mM Tris (pH 8.3), 140 mM NaCl, 2 mM $MgCl_2$, and 0.005% P20. The colored rectangles on the sensorgram represent selected sequences of the polymerized DNA. The detailed nucleotide sequences of DNA fragment 1 is shown in panel (c) and the sequences of other fragments in ESI† Table S2. The positions of the nucleotide sequences encoding the letters in fragments numbered 1–4 are deliniated next to the sensorgrams. (c) Nucleotide sequence of synthetic DNA fragment 1 that carries nucleotides encoding the word "DNA". The different sections of DNA are color-coded according to the nucleotide triplet: ATC in green, TCG in blue, ACG in orange, and ATG in red. The nucleotide A shown in black is likely a result of a sequencing error.

absence of one specific nucleotide in each injection mixture was used to convey the information.

To demonstrate that information can be stored in DNA synthesized by AbiK, we designed an assembly to synthesize ssDNA encoding the word "DNA" and read the stored information (Fig. 2a). Notably, AbiK was prompted to synthesize a sequence encoding the word "DNA" twice, flanked by a hompolymeric guanine nucleobase sequence, which enabled data retrieval. The sensorgram indicates that AbiK on the chip incorporated the nucleotides with each injection, suggesting that the information was encoded in the DNA (Fig. 2b). However, on the basis of the sensorgrams alone, we cannot determine whether the activity of the AbiK enzyme molecules immobilized on the chip was homogeneous, that is, whether all enzyme molecules came into contact with nucleotides and used the incoming nucleotides to extend the DNA strand. Therefore, we released the synthetic ssDNA from the SPR chip by AbiK proteolysis and generated the second strand using the large (Klenow) fragment of *E. coli* DNA polymerase I and an oligo-$C_{18}$ primer (Fig. 2a). Nanopore sequencing revealed that seven out of 89 sequences showed significant matches in the BLAST searches against a non-redundant database. These seven sequences were excluded from further analysis and the remaining sequences were considered as new fragments produced by AbiK. The DNA produced by AbiK was indeed heterogeneous, with regions of the DNA fragments carrying nucleotides for the corresponding letters of the alphabet ranging from a few hundred to 1012 bp (ESI† Table S2, ESI Data 1). Forty-two percent of the DNA fragments encoded at least one letter of the alphabet, *i.e.* they contained at least three different segments arranged in the correct order, with each segment consisting of only three of the four natural nucleotides. In the DNA fragments, multiple ~10–400-bp-long stretches of DNA consisting of only three of the four nucleotides were detected (Fig. 2c and ESI† Table S2). Moreover, DNA regions in which the segments signified the injection patterns (Fig. 2a) and represented specific letters (ESI† Table S2) or even the whole word "DNA" (Fig. 2c) were observed. Nevertheless, not all ssDNA strands formed by AbiK on the chip were the same, as signified by the different lengths of DNA reads, different lengths of segments containing three nucleotides, inconsistencies with the injection pattern and some "bleeding" of the nucleotides between various segments (ESI† Table S2, ESI Data 1).

The present study provides proof of principle that the reverse transcriptase AbiK of the bacteriophage abortive infection system facilitates custom DNA synthesis. We used the AbiK enzyme, which—unlike TdT—shows no preference for the incorporation of some nucleotides over others and catalyzes protein-primed ssDNA synthesis;[18,21] thus, only the dNTPs needed to be injected into the system, whereas the enzyme was immobilized on the solid phase. Furthermore, we used a unique DNA writing approach for information storage based on the absence of nucleotides, by injecting mixtures containing three of the four natural dNTPs. Using the SPR system, we directly monitored the activity of the AbiK enzyme. To our knowledge, this is the first report showing that DNA synthesized on the SPR chip can be released and sequenced. It is likely that a similar biophysical technique, such as biolayer interferometry (BLI), could be used analogously to SPR for the sequential buffer exchange and real-time monitoring of DNA polymerization by AbiK. The SPR method was preferred because it uses a continuous flow across AbiK immobilized on the sensor chip, whereas BLI relies on sequential dipping of the sensor tip in buffers containing mixtures of selected three nucleotides. We assumed that the dipping process could lead to nucleotide carryover from one buffer to another.

We hypothesize that other proteins belonging to the bacteriophage class 1 of the so-called unknown group and abortive infection (UG/Abi) reverse transcriptase family can also be used, but among those that were biochemically characterized, namely AbiK, AbiA, and Abi-P2, AbiK is the preferred enzyme because it incorporates all four natural dNTPs equally and synthesizes over 1000 nucleotide-long ssDNA products *in vitro*.[18,22] In contrast, AbiA synthesizes ssDNA mainly containing adenosine and cytosine bases with a length of only about 150 nucleotides, while Abi-P2 produces ssDNA of only a few tens of nucleotides.[16,18]

However, controlling the AbiK system such that ssDNA is elongated at one-base resolution remains a challenge. Furthermore, the immobilized AbiK enzyme molecules do not uniformly take up the incoming dNTPs; this "heterogeneity" effect needs to be addressed. In addition, scaling up the system with an alternative solid support will increase the efficiency and throughput, making the system more cost-effective and feasible for large-scale DNA synthesis applications. We believe that, with further development, AbiK and—most likely—other reverse transcriptase proteins from the abortive bacteriophage infection system could offer significant potential for DNA data storage and other nanobiotechnological applications.

## Data availability

All of the data presented in this study are available from the corresponding author upon reasonable request. The raw sequencing data are presented in ESI† Data 1.

## Conflicts of interest

The authors are listed as coinventors on a pending patent application related to this work filed at the European Patent Office in Luxemburg (application patent no. LU507011).

## Acknowledgements

This work was supported by the Slovenian Research and Innovation Agency [grant numbers J1-4394, P1-0207]. The authors thank Zdravko Podlesek and Jan Otoničar for their support with the initial expression of the AbiK protein.

## References

1 C. Bancroft, T. Bowler, B. Bloom and C. T. Clelland, *Science*, 2001, **293**, 1763–1765.
2 M. Yu, X. Tang, Z. Li, W. Wang, S. Wang, M. Li, Q. Yu, S. Xie, X. Zuo and C. Chen, *Chem. Soc. Rev.*, 2024, **53**, 4463–4489.
3 A. Hoose, R. Vellacott, M. Storch, P. S. Freemont and M. G. Ryadnov, *Nat. Rev. Chem.*, 2023, **7**, 144–161.
4 M. Pichon and M. Hollenstein, *Commun. Chem.*, 2024, **7**, 138.
5 R. Obexer, M. Nassir, E. R. Moody, P. S. Baran and S. L. Lovelock, *Science*, 2024, **380**, 1150–1154.
6 S. L. Beaucage and M. H. Caruthers, *Tetrahedron Lett.*, 1981, **22**, 1859–1862.
7 M. D. Matteucci and M. H. Caruthers, *J. Am. Chem. Soc.*, 1981, **103**, 3185–3191.
8 E. M. LeProust, B. J. Peck, K. Spirin, H. B. McCuen, B. Moore, E. Namsaraev and M. H. Caruthers, *Nucleic Acids Res.*, 2010, **38**, 2522–2540.
9 W. P. C. Stemmer, A. Crameri, K. D. Ha, T. M. Brennan and H. L. Heyneker, *Gene*, 1995, **164**, 49–53.
10 D. G. Gibson, L. Young, R. Y. Chuang, J. C. Venter, C. A. Hutchison and H. O. Smith, *Nat. Methods*, 2009, **6**, 343–345.
11 M. A. Jensen and R. W. Davis, *Biochemistry*, 2018, **57**, 1821–1832.
12 H. H. Lee, R. Kalhor, N. Goela, J. Bolot and G. M. Church, *Nat. Commun.*, 2019, **10**, 2383.
13 H. Lee, D. J. Wiegand, K. Griswold, S. Punthambaker, H. Chun, R. E. Kohman and G. M. Church, *Nat. Commun.*, 2020, **11**, 5264.
14 L. C. Fortier, J. D. Bouchard and S. Moineau, *J. Bacteriol.*, 2005, **187**, 3721–3730.
15 E. Emond, B. J. Holler, I. Boucher, P. A. Vandenbergh, E. R. Vedamuthu, J. K. Kondo and S. Moineau, *Appl. Environ. Microbiol.*, 1997, **63**, 1274–1283.
16 M. Figiel, M. Gapinska, M. Czarnocki-Cieciura, W. Zajko, M. Sroka, K. Skowronek and M. Nowotny, *Nucleic Acids Res.*, 2022, **50**, 10026–10040.
17 C. Wang, M. Villion, C. Semper, C. Coros, S. Moineau and S. Zimmerly, *Nucleic Acids Res.*, 2011, **39**, 7620–7629.
18 M. Gapińska, W. Zajko, K. Skowronek, M. Figiel, P. S. Krawczyk, A. A. Egorov, A. Dziembowski, M. J. O. Johansson and M. Nowotny, *Nucleic Acids Res.*, 2024, **52**, 4723–4738.
19 M. Buckle, R. M. Williams, M. Negroni and H. Buc, *Proc. Natl. Acad. Sci. U. S. A.*, 1996, **93**, 889–894.
20 S. J. Greive, S. E. Weitzel, J. P. Goodarzi, L. J. Main, Z. Pasman and P. H. Von Hippel, *Proc. Natl. Acad. Sci. U. S. A.*, 2008, **105**, 3315–3320.
21 A. A. Kuznetsova, T. E. Tyugashev, I. V. Alekseeva, N. A. Timofeyeva, O. S. Fedorova and N. A. Kuznetsov, *Life Sci. Alliance*, 2022, **5**, e202201428.
22 M. R. Mestre, L. A. Gao, S. A. Shah, A. López-Beltrán, A. González-Delgado, F. Martínez-Abarca, J. Iranzo, M. Redrejo-Rodríguez, F. Zhang and N. Toro, *Nucleic Acids Res.*, 2022, **50**, 6084–6101.