nature biotechnology

ARTICLES

https://doi.org/10.1038/s41587-021-00969-6

Check for updates

Bioorthogonal information storage in L-DNA with a high-fidelity mirror-image *Pfu* DNA polymerase

Chuyao Fan[1,2], Qiang Deng[1,2] and Ting F. Zhu[1]✉

Natural DNA is exquisitely evolved to store genetic information. The chirally inverted L-DNA, possessing the same informational capacity but resistant to biodegradation, may serve as a robust, bioorthogonal information repository. Here we chemically synthesize a 90-kDa high-fidelity mirror-image *Pfu* DNA polymerase that enables accurate assembly of a kilobase-sized mirror-image gene. We use the polymerase to encode in L-DNA an 1860 paragraph by Louis Pasteur that first proposed a mirror-image world of biology. We realize chiral steganography by embedding a chimeric D-DNA/L-DNA key molecule in a D-DNA storage library, which conveys a false or secret message depending on the chirality of reading. Furthermore, we show that a trace amount of an L-DNA barcode preserved in water from a local pond remains amplifiable and sequenceable for 1 year, whereas a D-DNA barcode under the same conditions could not be amplified after 1 day. These next-generation mirror-image molecular tools may transform the development of advanced mirror-image biology systems and pave the way for the realization of the mirror-image central dogma and exploration of their applications.

The concept of a mirror-image world of biology was first proposed more than 160 years ago by Pasteur soon after his discovery of molecular chirality[1], and yet to date, a mirror-image form of life has not been discovered in nature or synthesized in the laboratory. Such mirror-image biology systems would require chirally inverted versions of the enzymes and substrates involved in the central dogma of molecular biology[2–4]. On the nucleic acid front, traditional column-based phosphoramidate chemistry has enabled efficient mirror-image (L-) oligonucleotide synthesis of up to ~150 nucleotides (nt) for DNA and ~70 nt for RNA[5,6]. On the protein front, the conjunction of solid-phase peptide synthesis (SPPS)[7] and native chemical ligation (NCL)[8–10] has yielded effective means for the total chemical synthesis of various mirror-image (D-) proteins[2,11–17]. Despite the recent development of new peptide synthesis methods, such as automated fast-flow peptide synthesis capable of producing peptide chains of up to 164 amino acids (aa)[18], chemical protein synthesis has remained limited to relatively small proteins; the synthesis of proteins larger than ~400 aa is much harder to achieve, mainly owing to the limited synthesis and ligation efficiencies of peptide segments. The lack of methods to synthesize long mirror-image nucleic acid and large mirror-image protein molecules has prohibitively constrained the development of advanced mirror-image biology systems.

We have demonstrated that one way to overcome the bottleneck of synthesizing long L-nucleic acid molecules is through enzymatic polymerization by mirror-image polymerases. We initially developed a mirror-image genetic replication and transcription system based on the mirror-image version of the 174-aa African swine fever virus polymerase X (ASFV pol X)[2]. This was followed by a more efficient and thermostable 352-aa *Sulfolobus solfataricus* P2 DNA polymerase IV (Dpo4) developed by us and others[15–17]. Despite the high error rates owing to the erroneous nature of the polymerases[15,19], as a proof of concept, these earlier studies demonstrated error-prone genetic replication and transcription[2], error-prone mirror-image polymerase chain reaction (MI-PCR) with assembly and amplification of short genes[15–17], and error-prone mirror-image gene transcription and reverse transcription[19]. In particular, using a mutant version of mirror-image Dpo4 (D-Dpo4-5m-Y12S), we enzymatically transcribed a full-length 120-nt mirror-image 5S ribosomal RNA (rRNA), which was too long to be chemically synthesized[19]. These early generation mirror-image polymerases represented a reluctant compromise between polymerase size and efficiency[2,15–17], as small polymerases, such as ASFV pol X and Dpo4 (with error rates on the order of 10^{-2} and 10^{-4}, respectively)[20–22], have intrinsically poor enzymatic activity and fidelity, making them unsuitable for the faithful assembly, amplification and transcription of long mirror-image genes[2,15,16,19], and for realization of practical mirror-image DNA information systems.

Harnessing mirror-image versions of the best and typically larger enzymatic tools that nature offers is key to the development of advanced mirror-image biology systems, but has remained challenging. Here we set out to chemically synthesize the mirror-image version of the 775-aa, hyperthermostable, high-fidelity *Pyrococcus furiosus* (*Pfu*) DNA polymerase, which possesses the highest fidelity among known natural thermostable DNA polymerases[23] and is one of the most widely used PCR enzymes in laboratories. Its total chemical synthesis faced substantial obstacles as the molecular mass of *Pfu* DNA polymerase, at 90 kDa, is about twice that of the 40-kDa Dpo4 (the previously reported largest chemically synthesized non-homopolymeric protein)[15–17] and the ~50-kDa hexa-ubiquitin– and tetra-ubiquitin–α-globin proteins (with side chain homopolymer ubiquitin segments included)[24,25]. Therefore, we introduced new strategies to facilitate the chemical synthesis of large mirror-image proteins, including split-protein designs and systematic isoleucine substitution, leading to the successful synthesis of the 90-kDa high-fidelity mirror-image *Pfu* DNA polymerase, which enabled accurate assembly of a kilobase-sized mirror-image gene and use of L-DNA for bioorthogonal information storage.

[1]School of Life Sciences, Tsinghua-Peking Center for Life Sciences, Beijing Frontier Research Center for Biological Structure, Beijing Advanced Innovation Center for Structural Biology, Center for Synthetic and Systems Biology, Ministry of Education Key Laboratory of Bioorganic Phosphorus Chemistry and Chemical Biology, Ministry of Education Key Laboratory of Bioinformatics, Tsinghua University, Beijing, China. [2]These authors contributed equally: Chuyao Fan, Qiang Deng. ✉e-mail: tzhu@tsinghua.edu.cn

Results

Design and synthesis of the 90-kDa high-fidelity mirror-image *Pfu* DNA polymerase. We reasoned that using split-protein designs could simplify the problem of chemically synthesizing large proteins into the preparation of two or more smaller protein fragments that can co-fold in vitro into a functionally intact enzyme. In addition, this strategy would allow the synthesis, purification, ligation and desulfurization of each split-protein fragment to be performed in parallel, reducing the overall time and costs needed for chemically synthesizing large proteins, as well as for corrections when synthesis failure on certain fragment(s) occurs. Many enzymes have natural or engineered split versions, including the *Pfu* DNA polymerase: a previously discovered split site between K467 and M468 in the coiled coil motif of its finger domain divides the polymerase into two fragments (a 467-aa Pfu-N fragment and a 308-aa Pfu-C fragment; Fig. 1a) without substantially affecting its PCR activity and fidelity[26]. Nevertheless, the synthesis of the Pfu-N fragment with 467 aa (54 kDa) alone, much larger than Dpo4 with 352 aa (40 kDa), still poses considerable challenges. One of the hurdles is that NCL of synthetic peptides prepared by SPPS requires an amino-terminal cysteine residue at the ligation site, and yet the wild-type (WT) *Pfu* DNA polymerase has only four cysteine residues (C429 and C443 in the Pfu-N fragment, and C507 and C510 in the Pfu-C fragment). Although we took advantage of a previously reported metal-free radical-based desulfurization approach to convert unprotected cysteine to alanine residue after NCL[27] so that another eight ligation sites with alanine residues (A40, A163, A223 and A408 in the Pfu-N fragment, and A501, A596, A652 and A715 in the Pfu-C fragment) could be also used, some of the peptide segments were still too long to be prepared by SPPS directly. Therefore, we designed a mutant version of the *Pfu* DNA polymerase with five point mutations (encoding E102A, E276A, K317G and V367L in the Pfu-N fragment, and I540A in the Pfu-C fragment) based on multiple sequence alignment (MSA) to introduce additional ligation sites (Methods and Supplementary Fig. 1), without substantially affecting the PCR activity of the polymerase (split Pfu-5m; Supplementary Fig. 2).

Another challenge is the synthesis and ligation of hydrophobic peptide segments under aqueous conditions. Current methods to overcome this problem mainly focus on introducing various chemical modifications to the peptide, such as an *N*-(2-hydroxy-4-methoxybenzyl) (Hmb) moiety[28,29], removable solubilizing tags[30,31], pseudoprolines[32,33] and depsipeptide (*O*-acyl isopeptide)[34,35], although their practical use is often constrained by the laborious procedures involved and the requirement of special amino acid derivatives. Here we addressed this problem by an alternative approach through introducing mutations. For example, we found the Pfu-C-4 segment difficult to synthesize by standard 9-fluorenylmethyloxycarbonyl (Fmoc)-based SPPS (Supplementary Fig. 3a), with poor solubility in aqueous acetonitrile and 6 M Gn·HCl solutions for downstream purification and NCL. We reasoned that isoleucine is one of the most bulky and hydrophobic proteinogenic amino acids[36,37], and thus substituting the isoleucine(s) in a hydrophobic peptide with other less bulky or hydrophobic amino acids (such as valine, leucine, alanine and so on) may substantially alter the physicochemical properties of the peptide segment. We introduced a systematic isoleucine substitution approach based on MSA and structural data to substitute all seven isoleucine residues in this segment (I598V, I605T, I611V, I619A, I631L, I643V and I648T; Methods and Supplementary Fig. 1) without substantially affecting the PCR activity of the polymerase (Supplementary Fig. 2). Indeed, with these seven point mutations, the synthesis of this peptide segment was readily achieved (Supplementary Fig. 3b), and it also became soluble in aqueous acetonitrile and 6 M Gn·HCl solutions for downstream purification and NCL, allowing us to bypass the need to resort to other chemical modifications for its synthesis. Although other bulky and hydrophobic amino acids such as phenylalanine and tyrosine can also be substituted using similar methods, in our current work, we chose to substitute isoleucine as it is typically more abundant in natural proteins[37] (in fact, no phenylalanine or tyrosine exists in Pfu-C-4), and systematic isoleucine substitution also drastically reduces the costs for synthesizing the D-polymerase (Methods). Therefore, we substituted a large number (41 out of 71, or 58%) of isoleucine residues in the *Pfu* DNA polymerase to other amino acids (such as valine, leucine and alanine and so on) without substantially affecting the PCR activity of the polymerase (split Pfu-5m-30I; Supplementary Fig. 2).

We used these strategies to chemically synthesize both the natural and mirror-image versions of the *Pfu* DNA polymerase. The Pfu-N fragment was divided into 9 peptide segments ranging from 40 to 62 aa in length (Extended Data Fig. 1), and the Pfu-C fragment was divided into 6 segments ranging from 33 to 63 aa (Extended Data Fig. 2). The peptide segments were prepared by Fmoc-based SPPS, purified by reversed-phase high-performance liquid chromatography (RP-HPLC) and assembled by hydrazide-based NCL with a convergent assembly strategy[38,39], followed by metal-free

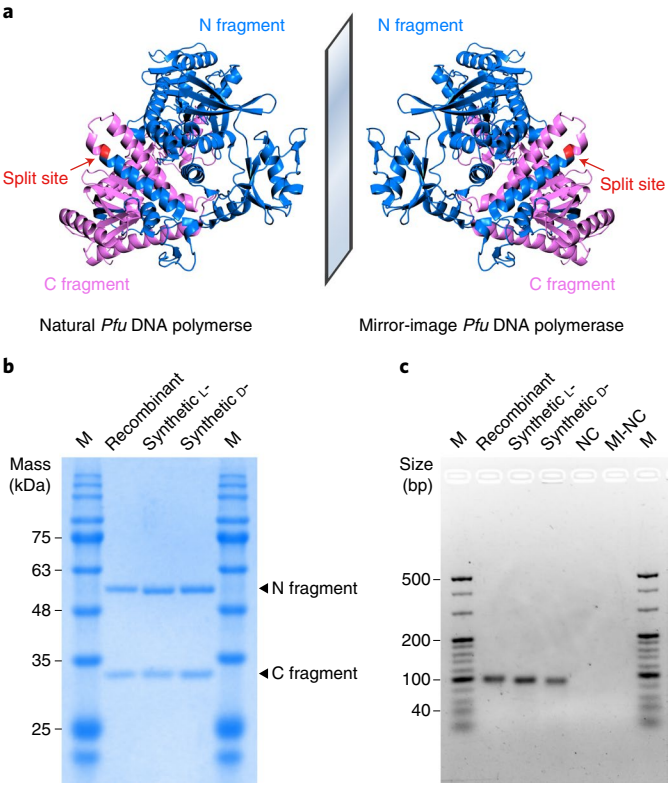


Fig. 1 | Synthetic natural and mirror-image *Pfu* DNA polymerases. a, The structure of the WT *Pfu* DNA polymerase (PDB: 3A2F) in its natural and mirror-image forms, showing the N fragment in blue and the C fragment in magenta. The split-site residues (K467 and M468) are highlighted in red. **b**, The recombinant split, mutant *Pfu* DNA polymerase expressed and purified from *E. coli* (Recombinant) and synthetic natural and mirror-image *Pfu* DNA polymerases of the same sequence (Synthetic L- and Synthetic D-, respectively), analyzed by 12% SDS–PAGE, stained with Coomassie brilliant blue. M, protein marker. The experiment was performed twice with similar results. **c**, PCR amplification of a 100-bp DNA sequence by recombinant split, mutant *Pfu* DNA polymerase (Recombinant), synthetic natural and mirror-image *Pfu* DNA polymerases of the same sequence (Synthetic L- and Synthetic D-, respectively), and negative controls without natural or mirror-image polymerase (NC and MI-NC, respectively), analyzed by 3% sieving agarose gel electrophoresis and stained with ExRed. M, DNA marker. The experiment was performed twice with similar results.

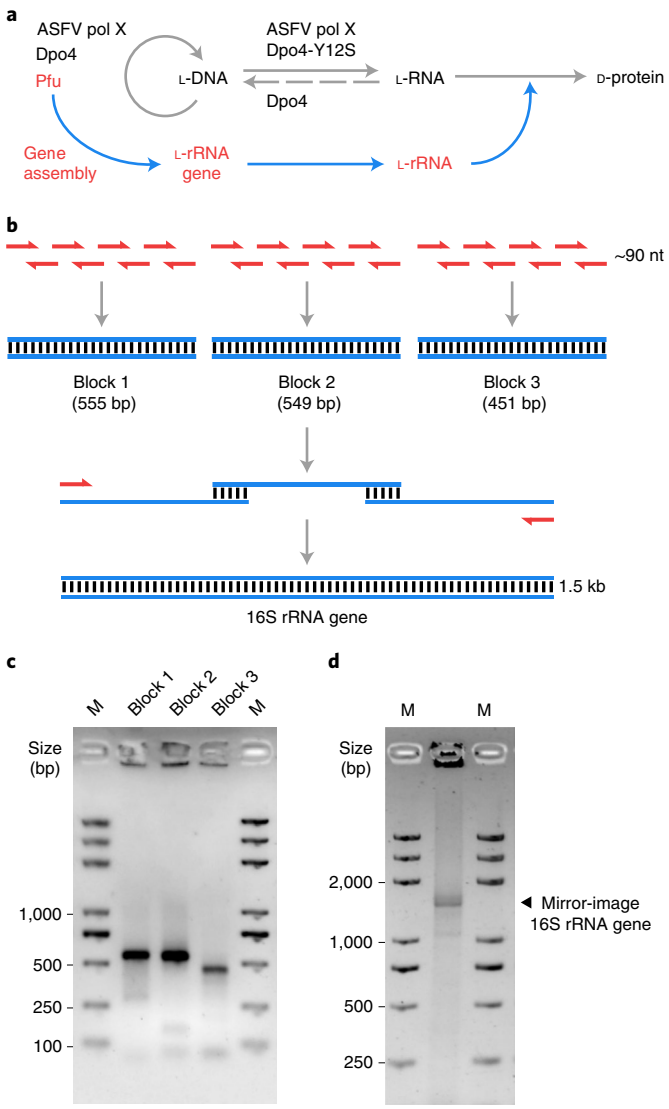


Fig. 2 | Mirror-image gene assembly by the mirror-image *Pfu* DNA polymerase. a, A mirror-image version of the central dogma of molecular biology (gray arrows), including mirror-image genetic replication (ASFV pol X, Dpo4 and Pfu), transcription (ASFV pol X and Dpo4-Y12S), reverse transcription (Dpo4) and translation, and the role of the high-fidelity mirror-image *Pfu* DNA polymerase and mirror-image gene assembly in realizing it (blue arrows). **b**, A schematic overview of the two-step assembly of a 1.5-kb mirror-image 16S rRNA gene. **c**, L-DNA blocks of 450–600 bp assembled by mirror-image assembly PCR from short, synthetic L-DNA oligonucleotides of ~90 nt, analyzed by 1.5% agarose gel electrophoresis and stained with ExRed. **d**, A full-length 1.5-kb mirror-image 16S rRNA gene obtained by mirror-image assembly PCR from the three L-DNA blocks, analyzed by 1.5% agarose gel electrophoresis and stained with ExRed. M, DNA marker. The experiment was performed once.

radical-based desulfurization[27] (Extended Data Figs. 1 and 2, and Supplementary Figs. 15–60). After several synthesis attempts, we obtained a total of 4.3 mg L-Pfu-N fragment with an observed molecular mass of 54,830.0 Da (calculated molecular mass of 54,829.9 Da; Supplementary Fig. 29) and 2.2 mg L-Pfu-C fragment with an observed molecular mass of 35,563.2 Da (calculated molecular mass of 35,563.0 Da; Supplementary Fig. 37) for the L-polymerase; a total of 16.5 mg D-Pfu-N fragment with an observed molecular mass of 54,829.5 Da (calculated molecular mass of 54,829.9 Da; Supplementary Fig. 52) and 11.9 mg D-Pfu-C fragment with an observed molecular mass of 35,561.9 Da (calculated molecular mass of 35,563.0 Da; Supplementary Fig. 60) for the D-polymerase. Both the synthetic L- and D-polymerases were folded by dialysis, followed by heat precipitation at 85 °C, which further improved the purity of the correctly folded protein (Supplementary Fig. 4). Next, we analyzed the folded polymerases by sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS–PAGE; Fig. 1b), and tested their PCR activities on short, 100-base-pair (bp) synthetic D- or L-DNA templates, and measured comparable amplification efficiencies between the recombinant, and synthetic L- and D-polymerases (Fig. 1c and Supplementary Fig. 5). We also quantified the fidelity of the synthetic L-polymerase on a 1.2-kilobase (kb) DNA template, and Sanger sequencing of the PCR products measured an error rate of 3.6×10^{-6} (Supplementary Table 1), consistent with that of the WT *Pfu* DNA polymerase reported in previous studies[23].

Assembly of a kilobase-sized mirror-image gene. We carried out the assembly of a full-length 1.5-kb mirror-image 16S rRNA gene (Fig. 2a). As many of the enzymatic tools and molecular cloning techniques are currently unavailable in the mirror-image system, we developed an approach for the accurate assembly of kilobase-sized mirror-image genes with the limited, currently available mirror-image molecular tools. We began by testing the gene assembly using synthetic L- *Pfu* DNA polymerase on D-DNA with a modified two-step assembly procedure[40]: DNA blocks of 450–600 bp were first assembled from short, synthetic oligonucleotides of ~90 nt (Supplementary Table 2), followed by a second step to assemble the DNA blocks into a full-length 1.5-kb 16S rRNA gene (Fig. 2b). However, in our initial attempt, Sanger sequencing of the assembled full-length sequences indicated that only ~40% of them were correct (Supplementary Table 1), with most of the errors being nucleotide deletions, likely arising from the minus 1- and 2-nt products from oligonucleotide synthesis. Thus, we modified the oligonucleotide purification approach using denaturing PAGE with single-nucleotide resolution to substantially improve the quality of the synthetic oligonucleotides by removing the majority of the plus/minus 1- and 2-nt sequences, after which most of the deletion errors were eliminated, and ~90% of the final assembled sequences were correct (the rest contained only single-nucleotide substitutions; Supplementary Table 1). Therefore, using the same oligonucleotide purification approach and mirror-image assembly PCR, we performed the assembly of a full-length 1.5-kb mirror-image 16S rRNA gene (Fig. 2c,d), which will become a template for the future enzymatic transcription into mirror-image 16S rRNA, a linchpin in building a functional mirror-image ribosome towards realizing the mirror-image central dogma[2–4] (Fig. 2a).

Mirror-image DNA information storage. We explored the application of the high-fidelity mirror-image *Pfu* DNA polymerase in storing information. We selected a paragraph from the 1860 publication by Pasteur in which the concept of a mirror-image world of biology was first proposed[1] (Fig. 3a), encoded the information into DNA sequences (Extended Data Fig. 3 and Supplementary Table 3) and archived them into 11 double-stranded L-DNA segments of 220 bp in length (Supplementary Table 4), each assembled from 4 short, synthetic L-DNA oligonucleotides of 70–90 nt by the mirror-image *Pfu* DNA polymerase (Fig. 3b,c). The L-DNA storage library containing all 11 segments (L-S1 to L-S11) was biostable in that each of the segments was amplifiable even after being treated by natural DNase I (Fig. 3d,e). The reading of L-DNA can be realized through L-DNA chemical sequencing[41], or sequencing-by-synthesis using mirror-image polymerases such as Dpo4 and the *Pfu* DNA polymerase, by the phosphorothioate approach with L-deoxynucleoside α-thiotriphosphates (L-dNTPαSs) and cleavage by 2-iodoethanol[42,43], or by the chain-termination approach with

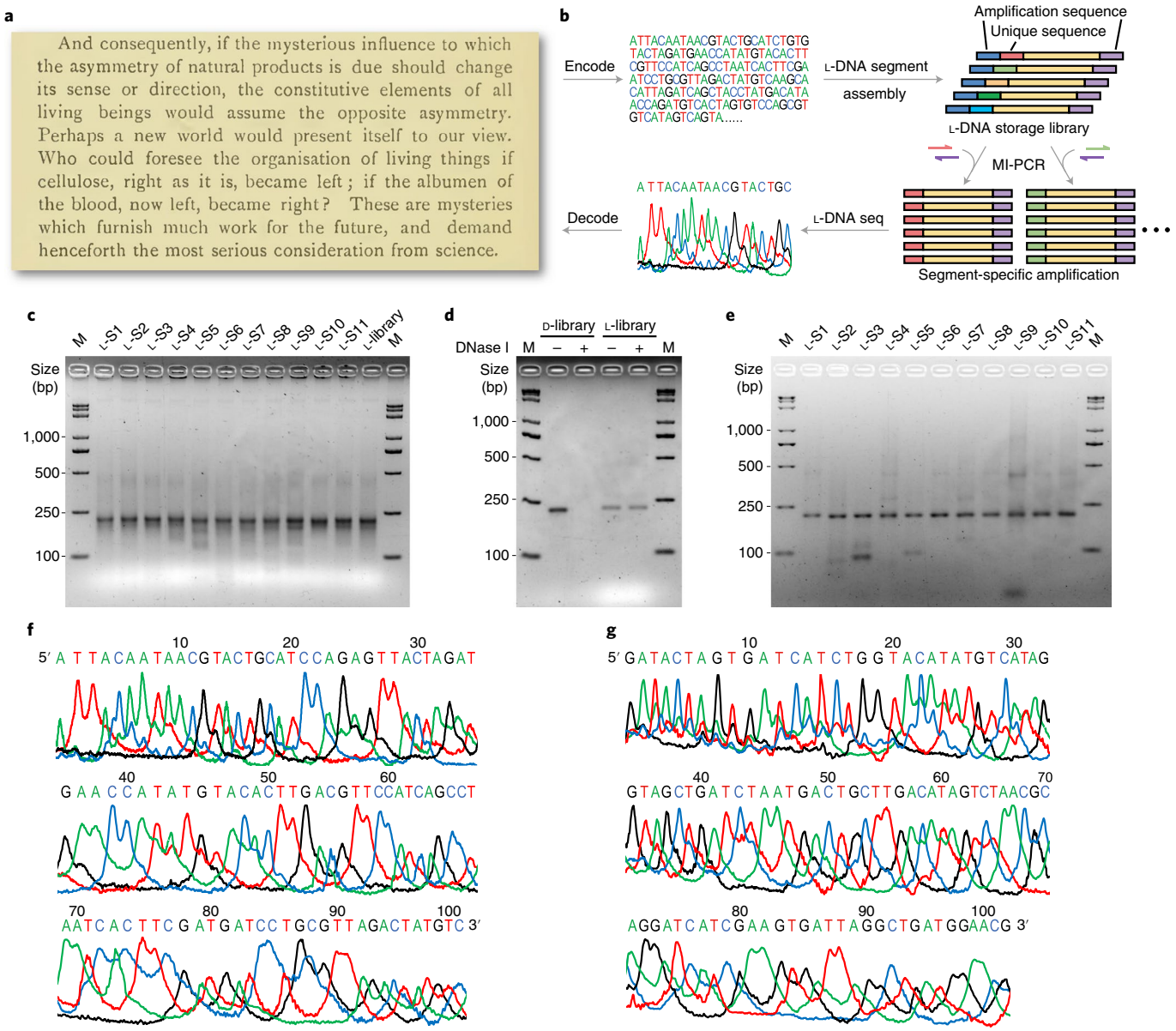


Fig. 3 | Mirror-image DNA information storage. a, The selected paragraph from Pasteur's 1860 publication in which the concept of a mirror-image world of biology was first proposed[1]. **b**, A schematic overview of mirror-image DNA information storage. L-DNA-seq, L-DNA sequencing. **c**, Information-storing double-stranded L-DNA segments of 220 bp (L-S1 to L-S11), each assembled by the mirror-image *Pfu* DNA polymerase using mirror-image assembly PCR from 4 short, synthetic L-DNA oligonucleotides of 70–90 nt, and the L-DNA storage library containing all 11 segments (L-library), analyzed by 2.5% agarose gel electrophoresis and stained with ExRed. The experiment was performed twice with similar results. **d**, Amplified D- and L-DNA storage libraries were treated by natural DNase I, analyzed by 2.5% agarose gel electrophoresis and stained with ExRed. The experiment was performed twice with similar results. **e**, Information-storing L-DNA segments of 203 bp, each amplified by D-Dpo4-5m from the DNase I-treated L-DNA storage library with segment-specific sequencing primers, analyzed by 2.5% agarose gel electrophoresis and stained with ExRed. M, DNA marker. The experiment was performed twice with similar results. **f**, Sequencing chromatogram of the information-storing L-DNA segment S1 by D-Dpo4-5m with L-dNTPαSs and 5′-FAM-labeled forward sequencing primer (with the corresponding sequencing gels shown in Supplementary Fig. 7a). The experiment was performed twice with similar results. **g**, Sequencing chromatogram of the information-storing L-DNA segment S1 by D-Dpo4-5m with L-dNTPαSs and 5′-Cy5-labeled reverse sequencing primer (with the corresponding sequencing gels shown in Supplementary Fig. 7b). The experiment was performed twice with similar results.

L-dideoxynucleoside triphosphates[44]. Here, for the convenience of synthesizing mirror-image Dpo4 and L-dNTPαSs, we chose to apply D-Dpo4-5m (a mutant version of Dpo4 to facilitate its chemical synthesis)[15,16] and L-dNTPαSs for L-DNA sequencing-by-synthesis (Methods). We also applied a bi-directional sequencing approach using 5′-labeled primers each with different dyes (FAM or Cy5), which improved the maximum read length in a single reaction to ~180 bp by denaturing PAGE (Supplementary Fig. 7). The information-storing L-DNA segments of 203 bp were each amplified from the DNase I-treated L-DNA storage library by D-Dpo4-5m with segment-specific sequencing primers (Fig. 3e), and the L-DNA segment S1 was sequenced by D-Dpo4-5m to retrieve the encoded digital text without detectable error (Fig. 3f,g and Supplementary Figs. 6a and 7), highlighting the abilities of the mirror-image DNA information system to faithfully write and read the information contents stored in L-DNA.

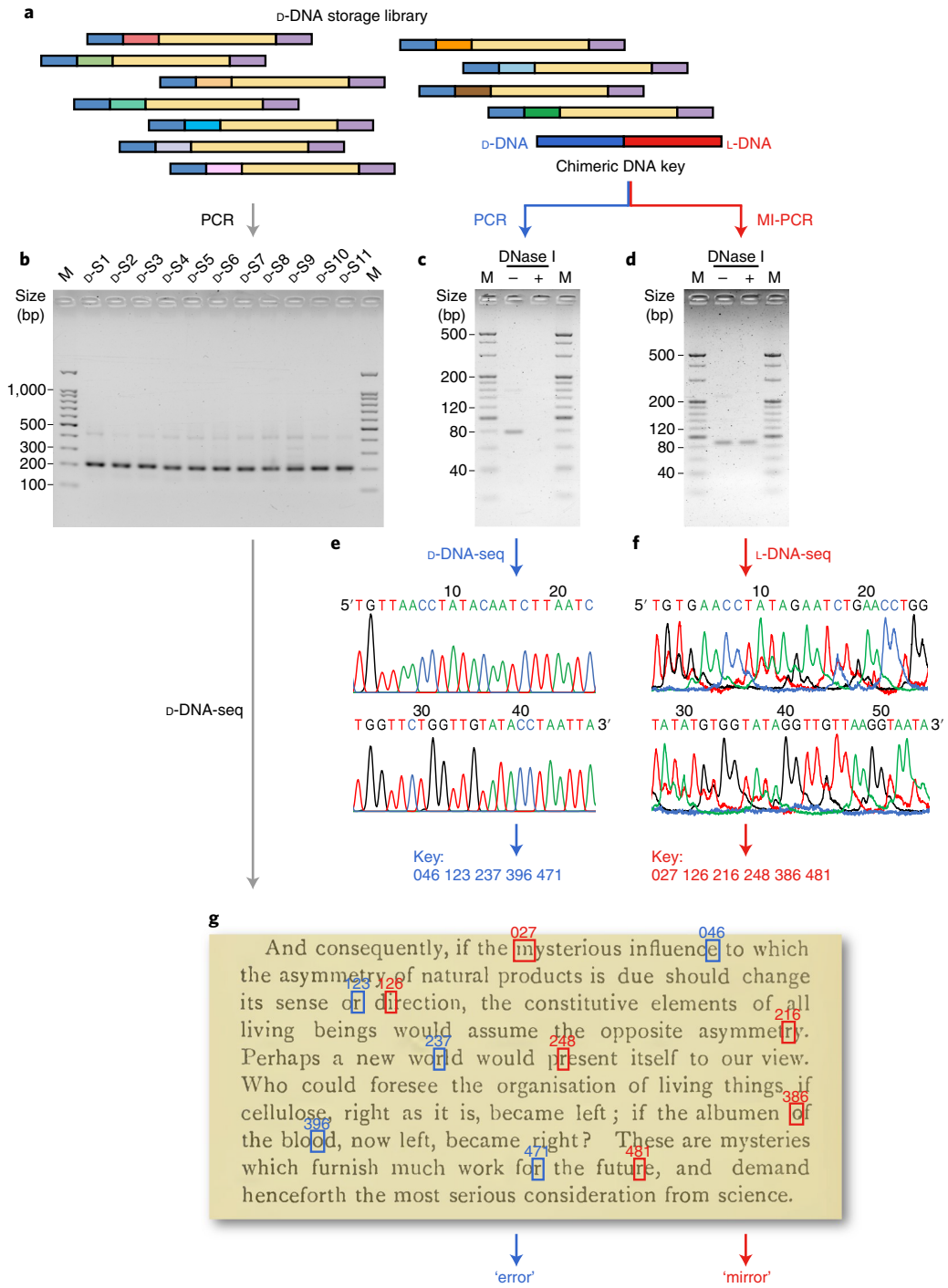


Fig. 4 | Chiral steganography. a, A D-DNA storage library encodes the selected paragraph from Pasteur's 1860 publication[1] as a 'cover text', embedded with a chimeric double-stranded D-DNA/L-DNA key molecule, which conveys either a false message 'error' (blue) or a secret message 'mirror' (red) depending on the chirality of reading. **b**, The information-storing D-DNA segments of 203 bp (D-S1 to D-S11) were each amplified from the D-DNA storage library with segment-specific sequencing primers for subsequent Sanger sequencing, analyzed by 2% agarose gel electrophoresis and stained with ExRed. The experiment was performed twice with similar results. **c**, The D-DNA part of the chimeric DNA key of 79 bp was PCR amplified from the storage library with D-DNA primers for subsequent Sanger sequencing, analyzed by 4% sieving agarose electrophoresis and stained with ExRed. The experiment was performed twice with similar results. **d**, The L-DNA part of the chimeric DNA key of 88 bp was MI-PCR amplified from the storage library by D-Dpo4-5m with L-DNA primers for subsequent L-DNA phosphorothioate sequencing, analyzed by 3.5% sieving agarose electrophoresis and stained with ExRed. M, DNA marker. The experiment was performed twice with similar results. **e**, Sanger sequencing chromatogram of the D-DNA part of the chimeric DNA key with reverse sequencing primer. The experiment was performed twice with similar results. **f**, Sequencing chromatogram of the L-DNA part of the chimeric DNA key by D-Dpo4-5m with L-dNTPαSs and 5′-Cy5-labeled reverse sequencing primer (with the corresponding sequencing gels shown in Supplementary Fig. 9). The experiment was performed once. **g**, The decoded D-DNA key reveals the false message 'error', whereas the decoded L-DNA key reveals the secret message 'mirror'.

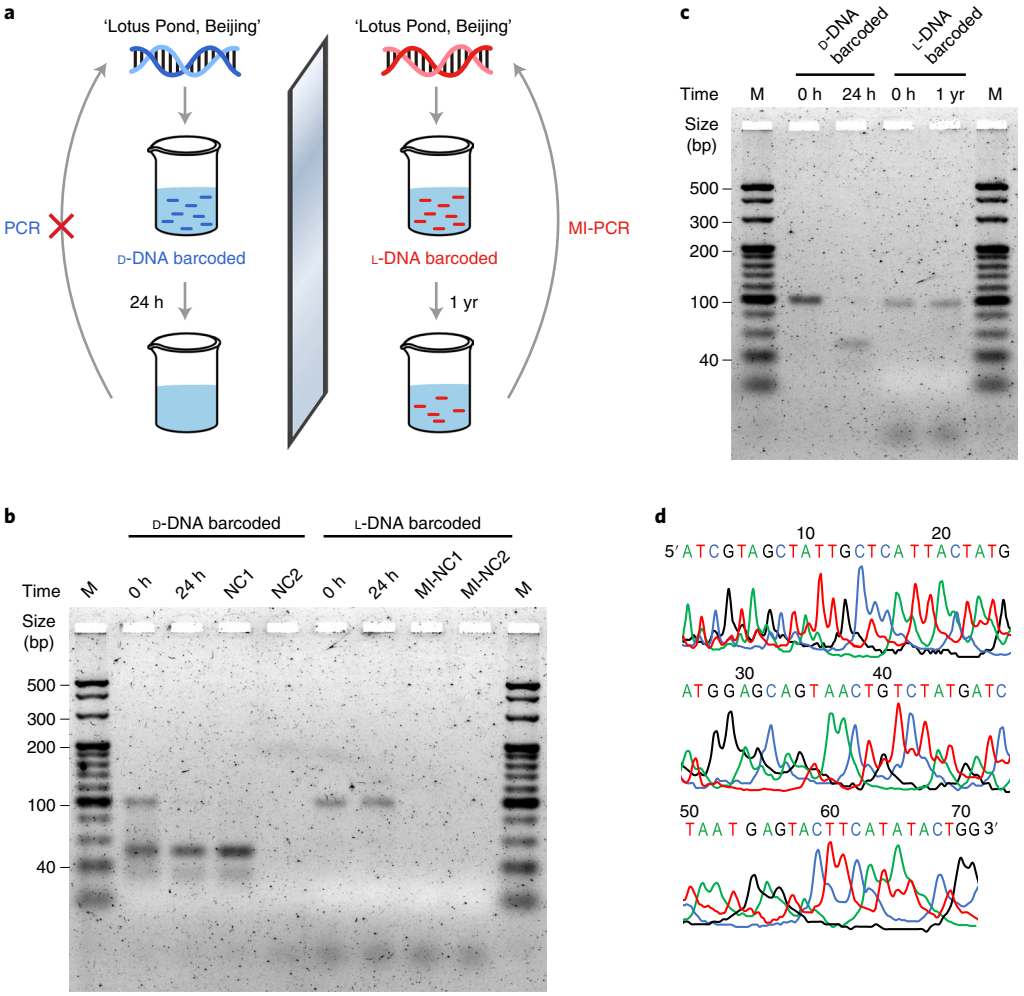


Fig. 5 | Mirror-image DNA barcoding of environmental water samples. a, A schematic overview of L-DNA barcoding of environmental water samples. **b**, PCR amplification of D-DNA barcode by L-Dpo4-5m in 2-ml pond water samples after 24 h, and MI-PCR amplification of L-DNA barcode by D-Dpo4-5m in 2-ml pond water samples after 24 h, analyzed by 3% sieving agarose gel electrophoresis and stained with ExRed. NC1 and MI-NC1 are negative controls without D- or L-DNA barcode, respectively. NC2 and MI-NC2 are negative controls without natural or mirror-image polymerase, respectively. The experiment was performed three times with similar results. **c**, PCR amplification of D-DNA barcode by L-Dpo4-5m in 40-ml pond water samples after 24 h, and MI-PCR amplification of L-DNA barcode by D-Dpo4-5m in 40-ml pond water samples after 1 year, analyzed by 3% sieving agarose gel electrophoresis and stained with ExRed. The experiment was performed twice with similar results. M, DNA marker. **d**, Sequencing chromatogram of the amplified L-DNA barcode after 1 year by D-Dpo4-5m with L-dNTPαSs and 5′-Cy5-labeled reverse sequencing primer (with the corresponding sequencing gels shown in Supplementary Fig. 12). The experiment was performed three times with similar results.

Chiral steganography. We applied the mirror-image DNA information system to secure communication in a chiral steganography experiment, where a D-DNA storage library encodes the selected paragraph from Pasteur's 1860 publication[1] as a 'cover text' (Fig. 4a), and an L-DNA key helps to decrypt the 'stego text' (secret message). A total of 550 characters in the paragraph were serially numbered from 001 to 550. Embedding an L-DNA key molecule in the D-DNA storage library would enable the conveyance of a secret message; yet, to make the secret message even more disguised, we designed a chimeric D-DNA/L-DNA key molecule (prepared by D-DNA/L-DNA oligonucleotide synthesis and enzymatic ligation of the double-stranded DNA; Supplementary Fig. 8 and Supplementary Table 5), which conveys either a false message 'error' or a secret message 'mirror' depending on the chirality of reading (Fig. 4). All of the 11 information-storing D-DNA segments were each amplified from the D-DNA storage library and sequenced by Sanger sequencing to retrieve the 'cover text' (Fig. 4b and Supplementary Table 6). We show that using natural PCR, one can amplify and sequence only the D-DNA part of the chimeric DNA key embedded in the storage library, revealing the false message 'error' (Fig. 4c,e,g), whereas using mirror-image PCR, one can amplify and sequence the L-DNA part of the chimeric DNA key, revealing the secret message 'mirror' (Fig. 4d,f,g and Supplementary Figs. 6b and 9).

Mirror-image DNA barcoding of environmental water samples. To demonstrate that information stored in L-DNA evades biodegradation and contamination from natural environments, we collected unpurified environmental water samples from a local pond and added a trace amount of 100-bp, double-stranded D- or L-DNA barcode encoding the location information of sample collection ('Lotus Pond, Beijing'; Fig. 5a and Supplementary Table 4) to the pond water samples (to a final concentration of 50 μg l^{-1}, or 770 pM, stored at 4 °C). The information-storing L-DNA barcode remained amplifiable for up to 1 year (Fig. 5c) and potentially beyond, with the stored information contents reliably retrieved by L-DNA phosphorothioate sequencing (Fig. 5d and Supplementary Figs. 6c and

12). In comparison, the D-DNA barcode of the same sequence and concentration could no longer be amplified after merely 1 day under the same conditions (Fig. 5b,c and Supplementary Fig. 10). Furthermore, we show that spiking the information-storing L-DNA barcode into the microbial (D-) DNA extracted from the pond water samples was also bioorthogonal in that it was specifically amplifiable by the mirror-image *Pfu* DNA polymerase with L-DNA primers (Supplementary Fig. 13a), and did not affect the (D-DNA) metagenomic sequencing results. No contaminating barcode sequences were found in more than 11 million Illumina high-throughput sequencing reads, with the top 10 genera of microbial organisms present in the pond water samples identified (Supplementary Fig. 13b), whereas in the control experiment using D-DNA barcode, more than 20 thousand contaminating barcode sequences were present (Supplementary Table 7).

Discussion

In this work, we developed an effective mirror-image DNA information system by chemically synthesizing a 90-kDa high-fidelity mirror-image *Pfu* DNA polymerase. We used it to accurately assemble a full-length 1.5-kb mirror-image 16S rRNA gene without relying on mirror-image DNA ligase[45,46]. The average size of natural proteins is ~270–470 aa (~30–50 kDa, corresponding to coding gene sequences of ~0.9–1.5 kb)[47], and >90% of the proteins in *Escherichia coli* and >70% of the proteins in other organisms are <600 aa (~65 kDa, corresponding to coding gene sequences of ~1.8 kb)[48]. Thus, the abilities to synthesize mirror-image versions of enzymatic proteins as large as the *Pfu* DNA polymerase at 90 kDa, and to assemble kilobase-sized mirror-image genes in turn, will become a key enabling technology and an important stepping stone towards synthesizing a mirror-image form of life. Surpassing in efficiency and fidelity the first-generation error-prone mirror-image ASFV pol X at 20 kDa and the second-generation error-prone mirror-image Dpo4 at 40 kDa, the development of this third-generation high-fidelity mirror-image *Pfu* DNA polymerase at 90 kDa should open opportunities for realizing advanced mirror-image biology systems and expanding the mirror-image molecular toolbox for applications in biotechnology and medicine.

The next crucial step in establishing the mirror-image central dogma is to realize mirror-image translation through building a functional mirror-image ribosome[2–4]. Although we have recently overcome the limitations of L-RNA chemical synthesis (typically less than ~70 nt)[6] by transcribing a synthetic L-DNA template into full-length 120-nt mirror-image 5S rRNA[19], more efficient enzymatic tools capable of transcribing longer mirror-image genes into longer L-RNAs are required for obtaining the 1.5-kb 16S and 2.9-kb 23S rRNAs, as well as messenger RNAs (mRNAs) for translation. One possibility is to transform DNA polymerases into DNA-dependent RNA polymerases as previously demonstrated[49]. Indeed, we have succeeded in reengineering the split, mutant *Pfu* DNA polymerase (with an additional six point mutations, encoding V93Q, D141A, E143A, Y410G, A486L and E665K) into an efficient DNA-dependent RNA polymerase (Supplementary Fig. 14). However, the preparation and purification of long single-stranded L-DNA templates pose another challenge and should be addressed first. Alternatively, synthesizing the mirror-image version of the T7 RNA polymerase, which uses double-stranded L-DNA templates, should enable the enzymatic transcription of the long mirror-image rRNAs and mRNAs required for mirror-image translation. Moreover, the synthesis of a mirror-image DNase may also help digest the L-DNA templates for L-RNA purification after transcription[19], and eliminate the information-storing L-DNA molecules after use as a biocontainment strategy.

The increasingly rapid pace at which data are being generated worldwide has created a growing need for reliable, high-density storage media to preserve the massive amount of information. Storage in DNA, nature's molecule of choice for encoding vast genomic instructions in tightly packed chromosomes, has emerged as a potential solution[50–52]. With the same informational capacity as natural DNA, mirror-image DNA is uniquely suited for the task of bioorthogonal information storage thanks to its abilities to evade biodegradation and contamination, a concept we proposed earlier with the development of mirror-image DNA chemical sequencing[41]. Mirror-image DNA information storage and chiral steganography can be also applied in conjunction with DNA cryptography to provide an extra layer of security through encryption. Future efforts to reengineer the mirror-image polymerases (for example, by synthesizing mutant or truncated versions without 3′–5′ exonuclease activity) for L-DNA Sanger sequencing, to synthesize specialized mirror-image polymerases or helicases for L-DNA nanopore sequencing, or to develop automated, high-throughput L-DNA sequencing techniques may lead to new applications such as large-scale, dynamic L-DNA information storage and direct selection of L-nucleic acid aptamer drugs. The efficient assembly, amplification and sequencing of information-storing L-DNA may present opportunities for applications in data storage, DNA computing, L-DNA-encoded library screening, environmental barcoding, food and drug labeling, medical implant monitoring, anti-counterfeiting tagging and secure communication. The accurate assembly of mirror-image genes and even entire genomes in the future could also make the mirror-image DNA information system suitable for producing mirror-image genome backup copies of natural organisms for genome banking purposes.

Undoubtedly, the information-storing L-DNAs are still susceptible to physical stress and chemical degradation[53], and future studies to better understand the environmental processing and degradation of mirror-image molecules under various temperature, humidity and pH conditions, as well as their potential interactions with natural biology systems should be carried out. Incorporation into other DNA-protecting storage architectures may help further improve their stability in practical settings[54,55]. Conversely, given their unique ability to evade biodegradation, L-DNAs may become an excellent model system for studying DNA stability under physical stress and chemical degradation. Currently, applications of the mirror-image DNA information system are limited by high costs and slow writing and reading speeds, as well as the relative inconvenience of L-DNA sequencing and lack of error-correction mechanisms. Nonetheless, because mirror-image molecules behave in a reciprocal manner to their natural twins[2,11–17,56], our vast knowledge about the biochemistry, biophysics and design of natural proteins and nucleic acids can be immediately applied to the mirror-image system, making it uniquely robust and practical among the numerous synthetic non-canonical molecular systems that have been proposed.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41587-021-00969-6.

Received: 23 November 2020; Accepted: 31 May 2021;
Published online: 29 July 2021

References

1. Pasteur, L. *Researches on the Molecular Asymmetry of Natural Organic Products (Société Chimique de Paris, 1860)* Reprint No. 14 (Alembic Club, 1905).
2. Wang, Z., Xu, W., Liu, L. & Zhu, T. F. A synthetic molecular system capable of mirror-image genetic replication and transcription. *Nat. Chem.* **8**, 698–704 (2016).
3. Peplow, M. Mirror-image enzyme copies looking-glass DNA. *Nature* **533**, 303–304 (2016).

4. Peplow, M. A conversation with Ting Zhu. *ACS Cent. Sci.* **4**, 783–784 (2018).
5. Beaucage, S. L. & Caruthers, M. H. Deoxynucleoside phosphoramidites - a new class of key intermediates for deoxypolynucleotide synthesis. *Tetrahedron Lett.* **22**, 1859–1862 (1981).
6. Liu, Y. et al. Synthesis and applications of RNAs with position-selective labelling and mosaic composition. *Nature* **522**, 368–372 (2015).
7. Merrifield, R. B. Solid phase peptide synthesis. 1. Synthesis of a tetrapeptide. *J. Am. Chem. Soc.* **85**, 2149–2154 (1963).
8. Dawson, P., Muir, T., Clark-Lewis, I. & Kent, S. Synthesis of proteins by native chemical ligation. *Science* **266**, 776–779 (1994).
9. Yan, L. Z. & Dawson, P. E. Synthesis of peptides and proteins without cysteine residues by native chemical ligation combined with desulfurization. *J. Am. Chem. Soc.* **123**, 526–533 (2001).
10. Fang, G.-M. et al. Protein chemical synthesis by ligation of peptide hydrazides. *Angew. Chem. Int. Ed. Engl.* **50**, 7645–7649 (2011).
11. Milton, R., Milton, S. & Kent, S. Total chemical synthesis of a D-enzyme: the enantiomers of HIV-1 protease show reciprocal chiral substrate specificity. *Science* **256**, 1445–1448 (1992).
12. Zawadzke, L. E. & Berg, J. M. A racemic protein. *J. Am. Chem. Soc.* **114**, 4002–4003 (1992).
13. Weinstock, M. T., Jacobsen, M. T. & Kay, M. S. Synthesis and folding of a mirror-image enzyme reveals ambidextrous chaperone activity. *Proc. Natl Acad. Sci. USA* **111**, 11679–11684 (2014).
14. Vinogradov, A. A., Evans, E. D. & Pentelute, B. L. Total synthesis and biochemical characterization of mirror image barnase. *Chem. Sci.* **6**, 2997–3002 (2015).
15. Xu, W. et al. Total chemical synthesis of a thermostable enzyme capable of polymerase chain reaction. *Cell Discov.* **3**, 17008 (2017).
16. Jiang, W. et al. Mirror-image polymerase chain reaction. *Cell Discov.* **3**, 17037 (2017).
17. Pech, A. et al. A thermostable D-polymerase for mirror-image PCR. *Nucleic Acids Res.* **45**, 3997–4005 (2017).
18. Hartrampf, N. et al. Synthesis of proteins by automated flow chemistry. *Science* **368**, 980–987 (2020).
19. Wang, M. et al. Mirror-image gene transcription and reverse transcription. *Chem* **5**, 848–857 (2019).
20. Lamarche, B. J., Kumar, S. & Tsai, M. D. ASFV DNA polymerase X is extremely error-prone under diverse assay conditions and within multiple DNA sequence contexts. *Biochemistry* **45**, 14826–14833 (2006).
21. Ling, H., Boudsocq, F., Woodgate, R. & Yang, W. Crystal structure of a Y-family DNA polymerase in action: a mechanism for error-prone and lesion-bypass replication. *Cell* **107**, 91–102 (2001).
22. Boudsocq, F., Iwai, S., Hanaoka, F. & Woodgate, R. *Sulfolobus solfataricus* P2 DNA polymerase IV (Dpo4): an archaeal DinB-like DNA polymerase with lesion-bypass properties akin to eukaryotic polη. *Nucleic Acids Res.* **29**, 4607–4616 (2001).
23. Cline, J., Braman, J. C. & Hogrefe, H. H. PCR fidelity of *Pfu* DNA polymerase and other thermostable DNA polymerases. *Nucleic Acids Res.* **24**, 3546–3551 (1996).
24. Tang, S. et al. Practical chemical synthesis of atypical ubiquitin chains by using an isopeptide-linked Ub isomer. *Angew. Chem. Int. Ed. Engl.* **56**, 13333–13337 (2017).
25. Sun, H. & Brik, A. The journey for the total chemical synthesis of a 53 kDa protein. *Acc. Chem. Res.* **52**, 3361–3371 (2019).
26. Hansen, C. J., Wu, L., Fox, J. D., Arezi, B. & Hogrefe, H. H. Engineered split in *Pfu* DNA polymerase fingers domain improves incorporation of nucleotide ɣ-phosphate derivative. *Nucleic Acids Res.* **39**, 1801–1810 (2011).
27. Wan, Q. & Danishefsky, S. J. Free-radical-based, specific desulfurization of cysteine: a powerful advance in the synthesis of polypeptides and glycopolypeptides. *Angew. Chem. Int. Ed. Engl.* **46**, 9248–9252 (2007).
28. Hyde, C., Johnson, T., Owen, D., Quibell, M. & Sheppard, R. Some 'difficult sequences' made easy. *Int. J. Pept. Protein Res.* **43**, 431–440 (1994).
29. Johnson, T., Quibell, M. & Sheppard, R. C. *N,O*-bisFmoc derivatives of *N*-(2-hydroxy-4-methoxybenzyl)-amino acids: useful intermediates in peptide synthesis. *J. Pept. Sci.* **1**, 11–25 (1995).
30. Zheng, J. S. et al. Robust chemical synthesis of membrane proteins through a general method of removable backbone modification. *J. Am. Chem. Soc.* **138**, 3553–3561 (2016).
31. Jacobsen, M. T. et al. A helping hand to overcome solubility challenges in chemical protein synthesis. *J. Am. Chem. Soc.* **138**, 11775–11782 (2016).
32. Wöhr, T. et al. Pseudo-prolines as a solubilizing, structure-disrupting protection technique in peptide synthesis. *J. Am. Chem. Soc.* **118**, 9218–9227 (1996).
33. Pascal Dumy, M. K., Ryan, D. E., Rohwedder, B., Wöhr, T. & Mutter, M. Pseudo-prolines as a molecular hinge: reversible induction of cis amide bonds into peptide backbones. *J. Am. Chem. Soc.* **119**, 918–925 (1997).
34. Sohma, Y. et al. 'O-Acyl isopeptide method' for the efficient synthesis of difficult sequence-containing peptides: use of 'O-acyl isodipeptide unit'. *Tetrahedron Lett.* **47**, 3013–3017 (2006).
35. Coin, I. The depsipeptide method for solid-phase synthesis of difficult peptides. *J. Pept. Sci.* **16**, 223–230 (2010).
36. Kyte, J. & Doolittle, R. F. A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.* **157**, 105–132 (1982).
37. Ellington, A. & Cherry, J. M. Characteristics of amino acids. *Curr. Protoc. Mol. Biol.* **33**, A.1C.1–A.1C.12 (2001).
38. Fang, G. M., Wang, J. X. & Liu, L. Convergent chemical synthesis of proteins by ligation of peptide hydrazides. *Angew. Chem. Int. Ed. Engl.* **51**, 10347–10350 (2012).
39. Zheng, J. S., Tang, S., Qi, Y. K., Wang, Z. P. & Liu, L. Chemical synthesis of proteins using peptide hydrazides as thioester surrogates. *Nat. Protoc.* **8**, 2483–2495 (2013).
40. Xiong, A. S. et al. A simple, rapid, high-fidelity and cost-effective PCR-based two-step DNA synthesis method for long gene sequences. *Nucleic Acids Res.* **32**, e98 (2004).
41. Liu, X. & Zhu, T. F. Sequencing mirror-image DNA chemically. *Cell Chem. Biol.* **25**, 1151–1156 (2018).
42. Nakamaye, K. L., Gish, G., Eckstein, F. & Vosberg, H.-P. Direct sequencing of polymerase chain reaction amplified DNA fragments through the incorporation of deoxynucleoside α-thiotriphosphates. *Nucleic Acids Res.* **16**, 9947–9959 (1988).
43. Gish, G. & Eckstein, F. DNA and RNA sequence determination based on phosphorothioate chemistry. *Science* **240**, 1520–1522 (1988).
44. Sanger, F., Nicklen, S. & Coulson, A. R. DNA sequencing with chain-terminating inhibitors. *Proc. Natl Acad. Sci. USA* **74**, 5463–5467 (1977).
45. Zhang, B. et al. Ligation of soluble but unreactive peptide segments in the chemical synthesis of *Haemophilus influenzae* DNA ligase. *Angew. Chem. Int. Ed. Engl.* **58**, 12231–12237 (2019).
46. Weidmann, J., Schnolzer, M., Dawson, P. E. & Hoheisel, J. D. Copying life: synthesis of an enzymatically active mirror-image DNA-ligase made of D-amino acids. *Cell Chem. Biol.* **26**, 645–651 (2019).
47. Tiessen, A., Perez-Rodriguez, P. & Delaye-Arredondo, L. J. Mathematical modeling and comparison of protein size distribution in different plant, animal, fungal and microbial species reveals a negative correlation between protein size and protein number, thus providing insight into the evolution of proteomes. *BMC Res. Notes* **5**, 85 (2012).
48. Zhang, B. C. et al. Chemical synthesis of proteins containing 300 amino acids. *Chem. Res. Chin. Univ.* **36**, 733–747 (2020).
49. Cozens, C., Pinheiro, V. B., Vaisman, A., Woodgate, R. & Holliger, P. A short adaptive path from DNA to RNA polymerases. *Proc. Natl Acad. Sci. USA* **109**, 8067–8072 (2012).
50. Church, G. M., Gao, Y. & Kosuri, S. Next-generation digital information storage in DNA. *Science* **337**, 1628 (2012).
51. Goldman, N. et al. Towards practical, high-capacity, low-maintenance information storage in synthesized DNA. *Nature* **494**, 77–80 (2013).
52. Ceze, L., Nivala, J. & Strauss, K. Molecular digital data storage using DNA. *Nat. Rev. Genet.* **20**, 456–466 (2019).
53. Matange, K., Tuck, J. M. & Keung, A. J. DNA stability: a central design consideration for DNA data storage systems. *Nat. Commun.* **12**, 1358 (2021).
54. Paunescu, D., Fuhrer, R. & Grass, R. N. Protection and deprotection of DNA–high-temperature stability of nucleic acid barcodes for polymer labeling. *Angew. Chem. Int. Ed. Engl.* **52**, 4269–4272 (2013).
55. Koch, J. et al. A DNA-of-things storage architecture to create materials with embedded memory. *Nat. Biotechnol.* **38**, 39–43 (2020).
56. Wade, D. et al. All-D amino acid-containing channel-forming antibiotic peptides. *Proc. Natl Acad. Sci. USA* **87**, 4761–4765 (1990).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature America, Inc. 2021

Methods

Materials. L-DNA oligonucleotides were synthesized on the H-8 oligonucleotide synthesizer (K&A Laborgeraete) with L-deoxynucleoside phosphoramidites (ChemGenes). L-deoxynucleoside triphosphates (L-dNTPs) and L-dNTPαSs were synthesized from L-deoxynucleosides (ChemGenes)[57]. Primers for recombinant protein expression were ordered from Genewiz. Primers for 16S rRNA gene assembly were purified by denaturing PAGE with single-nucleotide resolution. Other DNA oligonucleotides were purified by oligonucleotide purification cartridges (Ruibiotech). The PAGE DNA Purification Kit was purchased from Tiandz. ExRed was purchased from Beijing Zoman Biotech. Tris base, NP-40, Tween 20, KCl, guanidine hydrochloride (Gn·HCl) and β-mercaptoethanol (β-ME) were purchased from Amresco. Imidazole and EDTA were purchased from Solarbio Life Sciences. 2-Indoethanol was purchased from Aladdin Bio-Chem Technology. 2-Chlorotrityl chloride resin (loading = 0.6 mmol g^{-1}) was purchased from Tianjin Nankai Hecheng Science & Technology. Wang Chemmatrix resin was purchased from CSBio. Fmoc-D-amino acids, Fmoc-L-amino acids and *O*-(6-chlorobenzotriazol-1-yl)-*N*,*N*,*N*′,*N*′-tetramethyluronium hexafluorophosphate (HCTU) were purchased from GL Biochem. *N*,*N*-Diisopropylethylamine (DIEA), trifluoroacetic acid, *N*,*N*-dimethylformamide (DMF), thioanisole, triisopropylsilane, 1,2-ethanedithiol, palladium chloride ($PdCl_2$), sodium 2-mercaptoethanesulfonate and 2,2′-azobis[2-(2-imidazolin-2-yl) propane] dihydrochloride (VA-044) were purchased from J&K Scientific. 4-Mercaptophenylacetic acid (MPAA) was purchased from Alfa Aesar. Piperidine, $Na_2HPO_4{\cdot}12H_2O$, $NaH_2PO_4{\cdot}2H_2O$, sodium nitrite ($NaNO_2$) and acetic anhydride were purchased from Sinopharm Chemical Reagent Co. NaCl, NaOH and hydrochloric acid were purchased from Sinopharm Chemical Reagent Co. Dichloromethane was purchased from Shanghai Titan Scientific Co. Tris(2-carboxyethyl)phosphine hydrochloride (TCEP·HCl), Fmoc-$NHNH_2$, ethyl cyanoglyoxylate-2-oxime (Oxyma), *N*,*N*′-diisopropylcarbodiimide (DIC) and DL-1,4-dithiothreitol (DTT) were purchased from Adamas Reagent Co. Glutathione (reduced form) was purchased from Acros Organics. Anhydrous ether was purchased from Beijing Tongguang Fine Chemicals Co. Acetonitrile (HPLC grade) was purchased from J. T. Baker.

Fmoc-based SPPS. All peptides were synthesized by Fmoc-based SPPS on the Liberty Blue automated microwave peptide synthesizer (CEM) and Prelude X automated peptide synthesizer (Protein Technologies). Peptides with a carboxy-terminal carboxylate such as Pfu-N-9 and Pfu-C-6 were synthesized on Wang Chemmatrix resin (CSBio) preloaded with the first C-terminal residue. All of the other peptides were synthesized on Fmoc-hydrazine 2-chlorotrityl chloride resin to prepare peptide hydrazides[58]. For each peptide acid, the first residue was manually attached to the Wang Chemmatrix resin by a double coupling method. In the first coupling reaction, amino acid was coupled for 1 h at 30 °C using 4 equivalents (equiv.) of amino acid, 3.8 equiv. HCTU and 8 equiv. DIEA, and the resin was washed with DMF and dichloromethane. Without deprotection, the second coupling reaction was carried out overnight at 25 °C with 4 equiv. amino acid, 4 equiv. Oxyma and 4 equiv. DIC. All resins were swelled in DMF for 5–10 min before use. The Fmoc groups of both resins and the assembled amino acids were removed by treatment with 20% piperidine and 0.1 M Oxyma in DMF at 85 °C. Coupling of amino acids except Fmoc-Cys(Trt)-OH and Fmoc-His(Trt)-OH was carried out at 85 °C using 4 equiv. amino acid, 4 equiv. Oxyma and 8 equiv. DIC. The coupling reactions for Fmoc-Cys(Trt)-OH and Fmoc-His(Trt)-OH were carried out at 50 °C for 10 min to avoid side reactions at high temperature. Trifluoroacetyl thiazolidine-4-carboxylic acid-OH (ref. [59]) was coupled using Oxyma/DIC activation at room temperature overnight. After the completion of peptide chain assembly, peptides were cleaved from resin using H_2O/thioanisole/triisopropylsilane/1,2-ethanedithiol/trifluoroacetic acid (0.5/0.5/0.5/0.25/8.25) (vol/vol). The cleavage reaction took 2.5 h under agitation at 27 °C. Most of the trifluoroacetic acid in the mixture was removed by N_2 blowing, and cold ether was added to precipitate the crude peptide. After centrifugation, the supernatant was discarded and the precipitates were washed twice with ether. The crude peptides were dissolved in CH_3CN/H_2O, analyzed by RP-HPLC and electrospray ionization mass spectrometry (ESI-MS), and purified by semi-preparative HPLC.

NCL. The C-terminal peptide hydrazide segment was dissolved in acidified ligation buffer (aqueous solution of 6 M Gn·HCl and 0.1 M NaH_2PO_4, pH 3.0). The mixture was cooled in an ice–salt bath (−10 °C), and 10 equiv. $NaNO_2$ in acidified ligation buffer (pH 3.0) was added. The activation reaction system was kept in an ice–salt bath under stirring for 25 min, after which 40 equiv. MPAA in ligation buffer and 1 equiv. N-terminal cysteine peptide were added, and the pH of the solution was adjusted to 6.5 at room temperature. After overnight reaction, 150 mM TCEP in ligation buffer (pH adjusted to 7.0) was added to dilute the system twice and the reaction system was kept at room temperature for 30 min with stirring. Finally, the ligation product was analyzed by HPLC and ESI-MS, and purified by semi-preparative HPLC. Notably, during the ligation of the Pfu-C-1 and Pfu-C-2 segments, we discovered that the ligation was very inefficient due to the insoluble Pfu-C-2 segment, and thus we increased the initial concentration of Gn·HCl to 8 M (final Gn·HCl concentration at ~7 M), which substantially improved the solubility and ligation efficiency of the two peptide segments.

Desulfurization. Cysteine-containing peptide (3 mg ml^{-1}) was dissolved in desulfurization buffer (0.1 M aqueous phosphate buffer containing 6 M Gn·HCl, 200 mM TCEP, 40 mM reduced L-glutathione and 20 mM VA-044, pH 6.8). The mixture was stirred at 37 °C overnight, and the desulfurization product was analyzed by HPLC and ESI-MS, and purified by semi-preparative HPLC.

Acm deprotection. The Acm group was removed by the Pd-assisted deprotection strategy[60]. Acm-protected peptide was dissolved in Acm deprotection buffer (aqueous solution of 6 M Gn·HCl, 0.1 M phosphate and 40 mM TCEP, pH 7.0) to a final concentration of 1 mM, after which 20 equiv. $PdCl_2$ was added. The reaction mixture was incubated with agitation at 25 °C overnight. DTT was added to 50 mM final concentration to quench the reaction. The reaction mixture was stirred for 1 h and purified by semi-preparative HPLC.

RP-HPLC and ESI-MS. All RP-HPLC analyses and purification procedures were carried out on Shimadzu Prominence HPLC systems (Shimadzu) with SPD-20A ultraviolet–visible detectors and LC-20AT solvent delivery units. An Ultimate XB-C4 column (5 μm, 4.6 × 250 mm; Welch Materials) was used for analysis at a flow rate of 1 ml min^{-1} to monitor the ligation reactions and analyze the purity of the peptide products. Ultimate XB-C4 and C18 columns (5 μm, 21.2 × 250 mm or 5 μm, 10 × 250 mm; Welch Materials) were used to separate the crude peptides and ligation products, respectively, at a flow rate of 4–8 ml min^{-1}. The purified products were characterized by ESI-MS on a Shimadzu LC/MS-2020 system (Shimadzu).

Design of the mutant *Pfu* DNA polymerase. To design a mutant *Pfu* DNA polymerase to facilitate its chemical synthesis and reduce the synthesis costs for the mirror-image version, MSA was performed among *Pfu*, KOD1, *Tgo*, 9°N-7 and *Tok* polymerases, revealing the highly conserved and the less conserved residues (Supplementary Fig. 1). To avoid affecting the PCR activity and fidelity of the polymerase, the highly conserved residues revealed by MSA were largely left unchanged, while some of the less conserved ones were substituted to introduce additional NCL sites and less bulky or hydrophobic amino acids. In addition, while the mirror-image versions of all proteinogenic amino acids are commercially available and most have similar prices to their natural counterparts, D-isoleucine is ~300-fold more expensive than L-isoleucine and the rest of the D-amino acids (in part due to the existence of two chiral centers that makes its synthesis and purification difficult), often accounting for 80–90% of the D-amino acid costs when synthesizing mirror-image proteins (depending on the abundance of isoleucine, typically at ~5% in natural proteins[37]). Thus, we substituted the isoleucine residues (although other bulky and hydrophobic amino acids such as phenylalanine and tyrosine can also be substituted using similar methods) to other less bulky or hydrophobic amino acids[36,37] to facilitate the chemical synthesis and reduce the synthesis costs for the mirror-image version. For some of the isoleucine substitutions (I171A, I191V, I256V, I268L, I401V, I598V, I619V, I643V, I734V and I745V) in the conserved regions revealed by MSA, structural data (PDB: 3A2F) were taken into consideration to substitute isoleucine with less bulky or hydrophobic amino acids that can fit into the structure and would have comparable interactions with the surrounding residues (atomic distance <3.5 Å)[61]. This approach resulted in reducing by approximately half the D-amino acid costs for synthesizing the mirror-image *Pfu* DNA polymerase, which may benefit its large-scale synthesis and practical applications in the future.

Protein expression and purification. The gene of *Pfu* DNA polymerase was cloned into the pET-28c plasmid, and mutants were constructed by the pEASY-Uni Seamless Cloning and Assembly Kit (TransGen Biotech). Proteins fused to an N-terminal His_6 tag were expressed using *E. coli* strain BL21 (DE3) in lysogeny broth (LB) medium. The induced cells were collected and resuspended in lysis buffer (40 mM Tris-HCl, 300 mM NaCl, 10 mM imidazole, 10 mM β-ME and 10 mg ml^{-1} lysozyme, pH 8.0). Cell lysate was heated at 85 °C for 15 min, and the thermolabile proteins were subsequently removed by centrifugation at 20,000*g* for 40 min at 4 °C. The supernatant was incubated in Ni-NTA Superflow resin (Senhui Microsphere Technology) for 1 h at 4 °C. The resin was washed by a buffer containing 40 mM Tris-HCl (pH 8.0), 300 mM NaCl, 40 mM imidazole and 10 mM β-ME, which was then eluted by a buffer containing 40 mM Tris-HCl (pH 8.0), 300 mM NaCl, 250 mM imidazole and 10 mM β-ME. The purified and concentrated *Pfu* DNA polymerase and mutants were dialyzed against a storage buffer containing 100 mM Tris-HCl (pH 8.0), 50% glycerol, 0.2 mM EDTA, 0.2% NP-40 nonionic detergent, 0.2% Tween 20 and 2 mM DTT. L-Dpo4-5m was expressed and purified from *E. coli*[15,16]. D-Dpo4-5m was synthesized and folded according to our previously reported methods, except that automated peptide synthesizers were used, and norleucine was replaced by methionine[15,16].

Folding of synthetic *Pfu* DNA polymerases in vitro. Lyophilized Pfu-N fragment or Pfu-C fragment was dissolved in 4 M or 5 M Gn·HCl containing 10 mM β-ME, respectively. Protein folding in vitro was performed by mixing equal concentrations of the two fragments (0.5 μM), followed by dialyzing against a buffer containing 40 mM Tris-HCl (pH 7.5), 1 mM EDTA, 100 mM KCl and 10% glycerol, overnight at 4 °C. The folded *Pfu* DNA polymerase was heated to 85 °C for 15 min to precipitate thermolabile peptides, which were subsequently removed

by centrifugation at 20,000*g* for 40 min at 4 °C. The supernatant was concentrated and dialyzed against a storage buffer of 100 mM Tris-HCl (pH 8.0), 50% glycerol, 0.2 mM EDTA, 0.2% NP-40 nonionic detergent, 0.2% Tween 20 and 2 mM DTT.

PCR activity and fidelity of synthetic *Pfu* DNA polymerases. The natural and mirror-image PCR reactions were performed in a 50-µl reaction system containing 1× Pfu buffer (Solarbio Life Sciences), with 200 µM (each) dNTPs, 0.2 µM (each) primers, template and polymerase. To quantify the PCR activity of *Pfu* DNA polymerase and its mutants, the polymerases were adjusted to the same concentration with WT *Pfu* DNA polymerase by 12% SDS–PAGE. The PCR program settings were 94 °C for 3 min (initial denaturation); 94 °C for 30 s, 50–65 °C (melting temperature (T_m) dependent) for 30 s, and 72 °C for 1–7 min (depending on the amplicon length), for 10–35 cycles; 72 °C for 10 min (final extension). To quantify the amplification efficiency of synthetic *Pfu* DNA polymerase, a 100-bp DNA sequence was used as the template. The amplification products of the first nine cycles were analyzed with the ImageJ software (https://imagej.nih.gov/ij). To examine the fidelity of synthetic *Pfu* DNA polymerase, products of natural PCR (on a 1.2-kb D-DNA sequence from the pUC19 plasmid) after cycle 45 were purified by the V-elute Gel Mini Purification Kit (Beijing Zoman Biotechnology) and cloned by the Zero background ZT4 Simple-Blunt Fast Clone Kit (Beijing Zoman Biotechnology) for Sanger sequencing, and the polymerase fidelity was calculated with the duplication number set to 12, according to previously described methods[62].

DNA-templated RNA polymerization. RNA polymerization was performed in 1× Thermopol buffer (New England Biolabs), 3 mM $MgSO_4$, 0.625 mM (each) NTPs, 0.5 µM 5′-FAM-labeled DNA primer (21 nt), 1 µM single-stranded DNA template (41 nt) and polymerase. Before the addition of polymerase, the reaction system was heated to 94 °C for 30 s and slowly cooled to 4 °C for annealing. The primer extension reaction took place at 65 °C for 10 min. The reaction was stopped by the addition of loading buffer containing 98% formamide, 0.25 mM EDTA and 0.0125% SDS, and the products were analyzed by 20% denaturing PAGE in 8 M urea.

Assembly of 16S rRNA gene. Synthetic oligonucleotides of ~90 nt in length at concentrations of 0.005–0.02 µM each (inner) or 0.2 µM each (outer) were assembled into a full-length gene in two steps by synthetic L- *Pfu* DNA polymerase. In the first step, the assembly PCR program settings were 94 °C for 3 min (initial denaturation); 94 °C for 30 s, 60 °C for 30 s, and 72 °C for 3 min, for 35 cycles; 72 °C for 10 min (final extension). In the second step, the previously assembled DNA blocks of 450–600 bp in length were purified by 1.5% agarose gel before being subject to assembly PCR. The assembly PCR program settings were 94 °C for 3 min (initial denaturation); 94 °C for 30 s, 60 °C for 30 s, and 72 °C for 7 min, for 35 cycles; 72 °C for 10 min (final extension). The assembled product was further amplified with PCR program settings: 94 °C for 3 min (initial denaturation); 94 °C for 30 s, 60 °C for 30 s, and 72 °C for 7 min, for 35 cycles; 72 °C for 10 min (final extension). The final D-DNA products of natural assembly PCR were purified by the V-elute Gel Mini Purification Kit (Beijing Zoman Biotechnology), and cloned by the Zero Background ZT4 Simple-Blunt Fast Clone Kit (Beijing Zoman Biotechnology) for Sanger sequencing. The assembly of a full-length 1.5-kb mirror-image 16S rRNA gene was performed using the same oligonucleotide purification approach and mirror-image assembly PCR by the mirror-image *Pfu* DNA polymerase.

L-DNA phosphorothioate sequencing. The L-DNA phosphorothioate sequencing approach used in this work was modified from methods for D-DNA phosphorothioate sequencing reported in the literature[42,43]. The L-DNA to be sequenced was specifically amplified with 5′-FAM-labeled (forward) and/or 5′-Cy5-labeled (reverse) sequencing primers by D-Dpo4-5m in four separate PCR reactions, within which one of the L-dNTPs was replaced by the corresponding L-dNTPαS. The PCR program settings were 86 °C for 3 min (initial denaturation); 86 °C for 30 s, 54 °C (T_m dependent) for 1 min, and 65 °C for 1–2.5 min (depending on the amplicon length), for 45 cycles; 65 °C for 5 min (final extension). The PCR products (mixed 1:20 w/w with unlabeled carrier double-stranded D-DNA of the same length and sequence) were purified by 8% or 10% denaturing PAGE and dissolved in water to a concentration of ~200 ng µl^{-1}. For each sequencing reaction, 2.5 µl of labeled L-DNA was mixed with 2.5 µl of a denaturation buffer (98% formamide and 0.25 mM EDTA) containing 2% (v/v) 2-iodoethanol, followed by heating to 95 °C for 3 min, before immediate transfer to ice. The samples were loaded on slabs of 0.4 mm × 340 mm × 300 mm, separated by 10% polyacrylamide gel in 8 M urea. The gel was pre-run at 50 W (constant power) for 2 h before being heated to 40–50 °C. After loading, the gel was run at 50 W (constant power) for 1.5 h and paused for fluorescent scanning, following which the gel continued running and was scanned every other hour until a total running time of up to 3.5 h (depending on the L-DNA length). The polyacrylamide gel was scanned by a Typhoon Trio$^+$ system operated under Cy2 and/or Cy5 modes. Gel quantification and chromatogram analysis were performed by the ImageJ software (https://imagej.nih.gov/ij).

Mirror-image DNA information storage. The selected paragraph from Pasteur's 1860 publication[1] containing 550 characters was encoded into a DNA sequence with 1,650 nucleotides according to Supplementary Table 3, and archived in 11 double-stranded L-DNA segments of 220 bp in length, each assembled from 4 short, synthetic L-DNA oligonucleotides of 70–90 nt by the mirror-image *Pfu* DNA polymerase. The assembly PCR program settings were 94 °C for 3 min (initial denaturation); 94 °C for 30 s, 55 °C for 30 s, and 72 °C for 2 min (depending on the amplicon length), for 35 cycles; 72 °C for 10 min (final extension). All 11 L-DNA segments in the L-DNA storage library were amplified by the mirror-image *Pfu* DNA polymerase with L-M13-F and L-M13-R primers and treated with natural DNase I (Turbo DNase, Invitrogen) to demonstrate their resistance to natural nuclease digestion. The information-storing L-DNA segments were each amplified from the DNase I-treated L-DNA storage library, and the S1 segment was specifically amplified with 5′-FAM-labeled (forward) and 5′-Cy5-labeled (reverse) sequencing primers by D-Dpo4-5m in four separate PCR reactions, within which one of the L-dNTPs was replaced by the corresponding L-dNTPαS for L-DNA phosphorothioate sequencing (Supplementary Figs. 6a and 7).

Chiral steganography. The chimeric D-DNA/L-DNA oligonucleotides were synthesized with D- and L-deoxynucleoside phosphoramidites using the methods described above. The oligonucleotides D-F1, D-R1, D/L-F2 and D/L-R2 (Supplementary Table 5) were heated to 95 °C for 3 min and slowly cooled to 4 °C for annealing, and the annealed double-stranded DNAs were ligated by the T3 DNA ligase (New England Biolabs; the T4 DNA ligase was also capable of the ligation, as shown in Supplementary Fig. 8b) at 25 °C for 1.5 h. The D-DNA storage library encoding the selected paragraph from Pasteur's 1860 publication[1] that served as a 'cover text' was prepared by the TransStart FastPfu Fly polymerase (TransGen Biotech) using similar methods as for the L-DNA storage library. The chimeric double-stranded D-DNA/L-DNA key purified by agarose gel electrophoresis was added to the D-DNA storage library at the same molar concentration as each D-DNA segment. The 11 information-storing D-DNA segments and the D-DNA part of the chimeric DNA key were each amplified with segment-specific primers from the storage library and cloned by the Zero Background ZT4 Simple-Blunt Fast Clone Kit (Beijing Zoman Biotechnology) for Sanger sequencing (Supplementary Table 6). The L-DNA part of the chimeric DNA key was amplified with L-M13F and L-M13R primers by D-Dpo4-5m from the storage library, with the following PCR program settings: 86 °C for 3 min (initial denaturation); 86 °C for 30 s, 55 °C for 1 min, and 65 °C for 1 min, for 45 cycles; 65 °C for 5 min (final extension). The amplified L-DNA sequence was further amplified with 5′-Cy5-labeled reverse sequencing primer by D-Dpo4-5m and sequenced by the phosphorothioate approach (Supplementary Figs. 6b and 9).

Mirror-image DNA barcoding of environmental water samples. Unpurified water samples were collected from the Lotus Pond at Tsinghua University (40° 0′ 27″ N, 116° 19′ 34″ E) on 8 December 2019. Synthetic D- or L-DNA oligonucleotides were heated to 95 °C for 5 min and slowly cooled to 4 °C for annealing, and the annealed double-stranded D- or L-DNA barcode was added to the fresh pond water samples to a final concentration of 50 µg l^{-1}, or 770 pM, after which the samples were stored at 4 °C without avoiding light. The amplification and sequencing-by-synthesis of L-DNA barcode can be performed by both the mirror-image Dpo4 and *Pfu* DNA polymerase, although for the convenience of synthesizing mirror-image Dpo4, we primarily used D-Dpo4-5m for amplifying and sequencing the short L-DNA barcode. To amplify the D- or L-DNA barcode, 2 ml of the pond water samples was filtered by an Acrodisc Syringe Filter with 0.2 µm Supor Membrane (Pall), and resuspended in diethyl pyrocarbonate (DEPC)-treated water by an Amicon Utra centrifugal filter unit (0.5 ml, 10,000 molecular weight cutoff), before being amplified by L- or D-Dpo4-5m. The PCR program settings were 86 °C for 3 min (initial denaturation); 86 °C for 30 s, 55 °C for 10 s, and 65 °C for 10 s, for 30 cycles; 65 °C for 5 min (final extension). The MI-PCR program settings were 86 °C for 3 min (initial denaturation); 86 °C for 30 s, 55 °C for 30 s, and 65 °C for 30 s, for 30 cycles; 65 °C for 5 min (final extension). To amplify the L-DNA barcode after long-term (1 year) storage, 40 ml of the pond water samples was filtered and concentrated in DEPC-treated water for amplification by D-Dpo4-5m. The MI-PCR program settings were 86 °C for 3 min (initial denaturation); 86 °C for 30 s, 55 °C for 1 min, and 65 °C for 1 min, for 30 cycles; 65 °C for 5 min (final extension). The amplified L-DNA barcode was further amplified with 5′-Cy5-labeled reverse sequencing primer by D-Dpo4-5m in four separate PCR reactions, within which one of the L-dNTPs was replaced by the corresponding L-dNTPαS for L-DNA sequencing. The amplification and sequencing of L-DNA barcode in pond water samples after 8 months were performed before the initial submission of this paper (Supplementary Fig. 11), and the amplification and sequencing results after 1 year were obtained during revision (Fig. 5 and Supplementary Figs. 6c and 12). For microbial (D-) DNA extraction, the pond water samples were filtered with a MicroFunnel Filter Funnel with 0.2 µm Supor Membrane (Pall), and the microbial DNA was extracted by the DNeasy PowerSoil Kit (Qiagen). The D- or L-DNA barcode was spiked into the microbial DNA extracted from the pond water samples, and amplified by synthetic natural and mirror-image *Pfu* DNA polymerases. An amount of ~2 µg extracted DNA from each sample was used for metagenomic sequencing library preparation. Sequencing libraries were generated and index codes were added, and sequenced by the Illumina NovaSeq 6000 sequencing platform (Illumina) with paired-end reads of 150 bp. The raw reads (available at the National Center for Biotechnology Information under the BioProject

number PRJNA707266) were pre-processed by Fastp[63], and the clean reads were analyzed by FastQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/). Metagenomic analysis of microbial organisms present in the pond water samples was performed by MetaPhlAn 2.7.7 (http://huttenhower.sph.harvard.edu/metaphlan2 (ref. [64]); Supplementary Fig. 13b and Supplementary Table 7).

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

The data that support the findings of this study are available within the paper and its Supplementary Information. Metagenomic sequencing data are available at the National Center for Biotechnology Information under the BioProject number PRJNA707266.

References

57. Caton-Williams, J., Hoxhaj, R., Fiaz, B. & Huang, Z. Use of a novel 5′-regioselective phosphitylating reagent for one-pot synthesis of nucleoside 5′-triphosphates from unprotected nucleosides. *Curr. Protoc. Nucleic Acid Chem.* **52**, 1.30.1–1.30.21 (2013).
58. Huang, Y.-C. et al. Facile synthesis of C-terminal peptide hydrazide and thioester of NY-ESO-1 (A39-A68) from an Fmoc-hydrazine 2-chlorotrityl chloride resin. *Tetrahedron* **70**, 2951–2955 (2014).
59. Huang, Y. C. et al. Synthesis of L- and D-ubiquitin by one-pot ligation and metal-free desulfurization. *Chemistry* **22**, 7623–7628 (2016).
60. Maity, S. K., Jbara, M., Laps, S. & Brik, A. Efficient palladium-assisted one-pot deprotection of (acetamidomethyl)cysteine following native chemical ligation and/or desulfurization to expedite chemical protein synthesis. *Angew. Chem. Int. Ed. Engl.* **55**, 8108–8112 (2016).
61. Burley, S. K. & Petsko, G. A. Weakly polar interactions in proteins. *Adv. Protein Chem.* **39**, 125–189 (1988).
62. Lundberg, K. S. et al. High-fidelity amplification using a thermostable DNA polymerase isolated from *Pyrococcus furiosus*. *Gene* **108**, 1–6 (1991).
63. Chen, S., Zhou, Y., Chen, Y. & Gu, J. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* **34**, i884–i890 (2018).
64. Segata, N. et al. Metagenomic microbial community profiling using unique clade-specific marker genes. *Nat. Methods* **9**, 811–814 (2012).

Acknowledgements

We thank J. Chen, M. Chen, W. Jiang, J. J. Ling, G. Wang, Y. Xu and R. Zhao for assistance with the experiments, and W. Jiang, M. J. McFall-Ngai, Y. Shi, J. W. Szostak, H. W. Wang, E. Winfree and N. Yan for comments on the manuscript. T.F.Z. was supported by funding from the National Natural Science Foundation of China (21925702, 32050178 and 21750005), the Tsinghua-Peking Center for Life Sciences, the Tencent Foundation, the Beijing Advanced Innovation Center for Structural Biology and the Beijing Frontier Research Center for Biological Structure.

Author contributions

C.F. performed the chemical synthesis. Q.D. performed the biochemistry experiments. All authors analyzed and discussed the results. T.F.Z. designed and supervised the study, and wrote the paper.

Competing interests

The authors have filed a provisional patent application related to this work.

Additional information

Extended data is available for this paper at https://doi.org/10.1038/s41587-021-00969-6.

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41587-021-00969-6.

Correspondence and requests for materials should be addressed to T.F.Z.

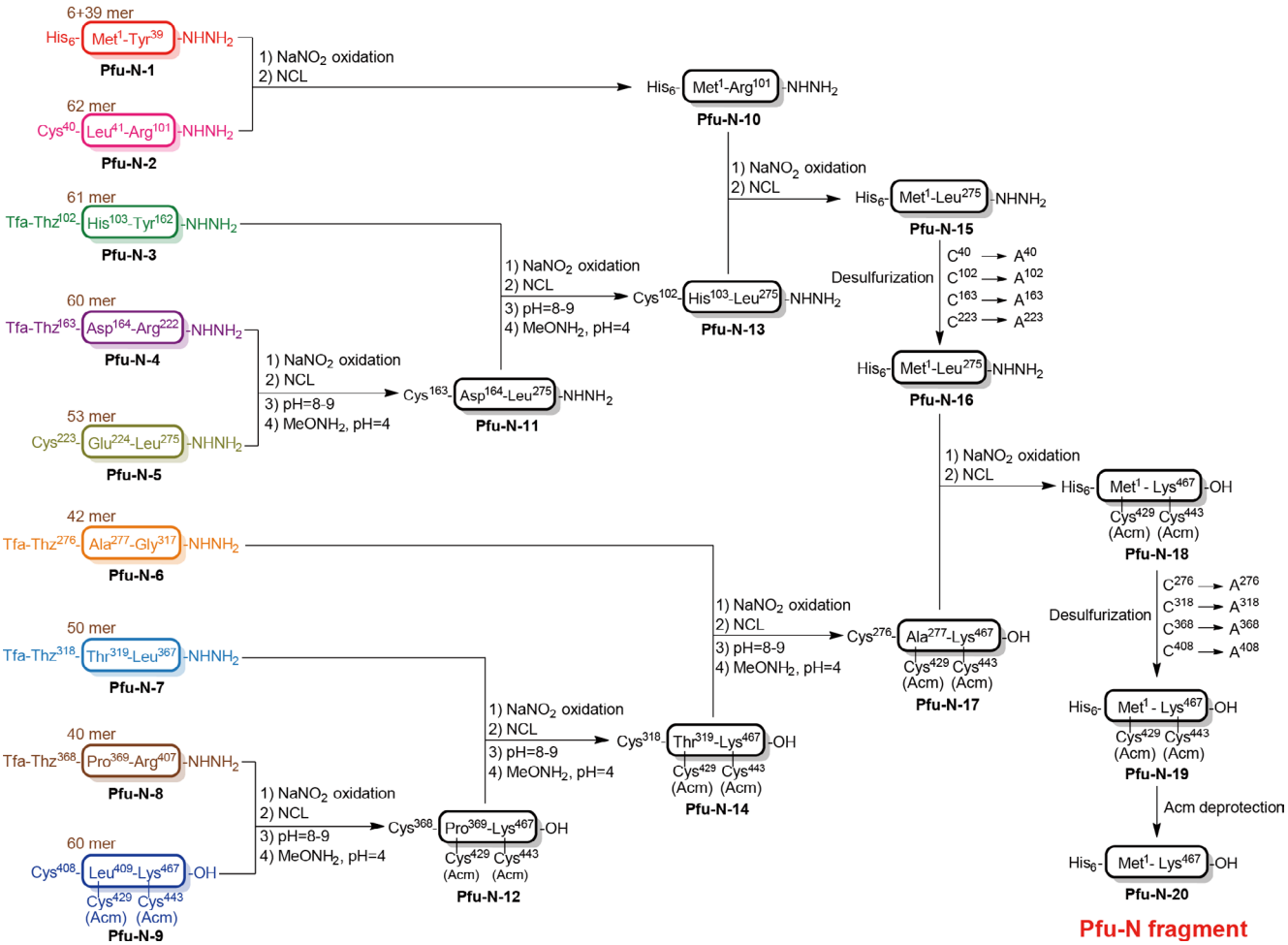
Peer review information *Nature Biotechnology* thanks the anonymous reviewers for their contribution to the peer review of this work.

Reprints and permissions information is available at www.nature.com/reprints.

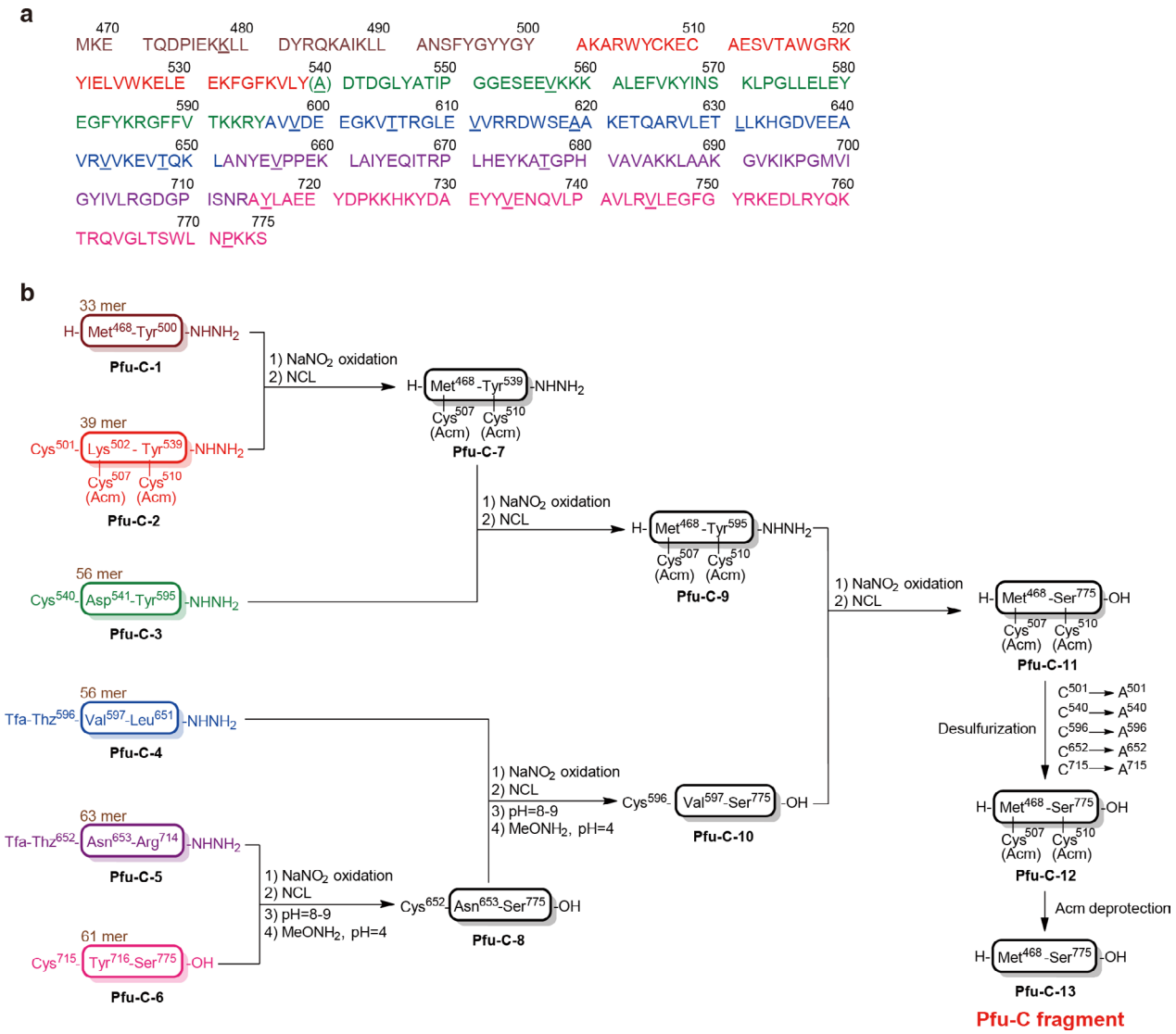
a

```
                10                  20                  30                  40                  50                  60
(His6)MILDVDYITE  EGKPVIRLFK  KENGKFKIEH  DRTFRPYEYA  LLRDDSKIEE  VKKITGERHG
                70                  80                  90                 100                 110                 120
KVVRVVDVEK  VEKKFLGKPV  TVWKLYLEHP  QDVPTIREKV  R(A)HPAVVDIF  EYDIPFAKRY
               130                 140                 150                 160                 170                 180
LIDKGLVPME  GEEELKMLAF  DIETLYHEGE  EFGKGPILMI  SYADENEAKV  ATWKNVDLPY
               190                 200                 210                 220                 230                 240
VEVVSSEREM  VKRFLRVVRE  KDPDVLVTYN  GDSFDFPYLA  KRAEKLGVKL  TLGRDGSEPK
               250                 260                 270                 280                 290                 300
MQRMGDMTAV  EVKGRVHFDL  YHVATRTLNL  PTYTL(A)AVYE  AVFGKPKEKV  YADEIAKAWE
               310                 320                 330                 340                 350                 360
SGENLERVAK  YSMEDA(G)ATY  ELGKEFLPME  AQLSRLVGQP  LWDVSRSSTG  NLVEWFLLRK
               370                 380                 390                 400                 410                 420
AYERNE(L)APN  KPSEEEYQRR  LRESYTGGFV  KEPEKGLWEN  VVYLDFRALY  PSIIITHNVS
               430                 440                 450                 460                 467
PDTLNLEGCK  NYDVAPQVGH  KFCKDFPGFI  PSLLGHLLEE  RQKIKTK
```

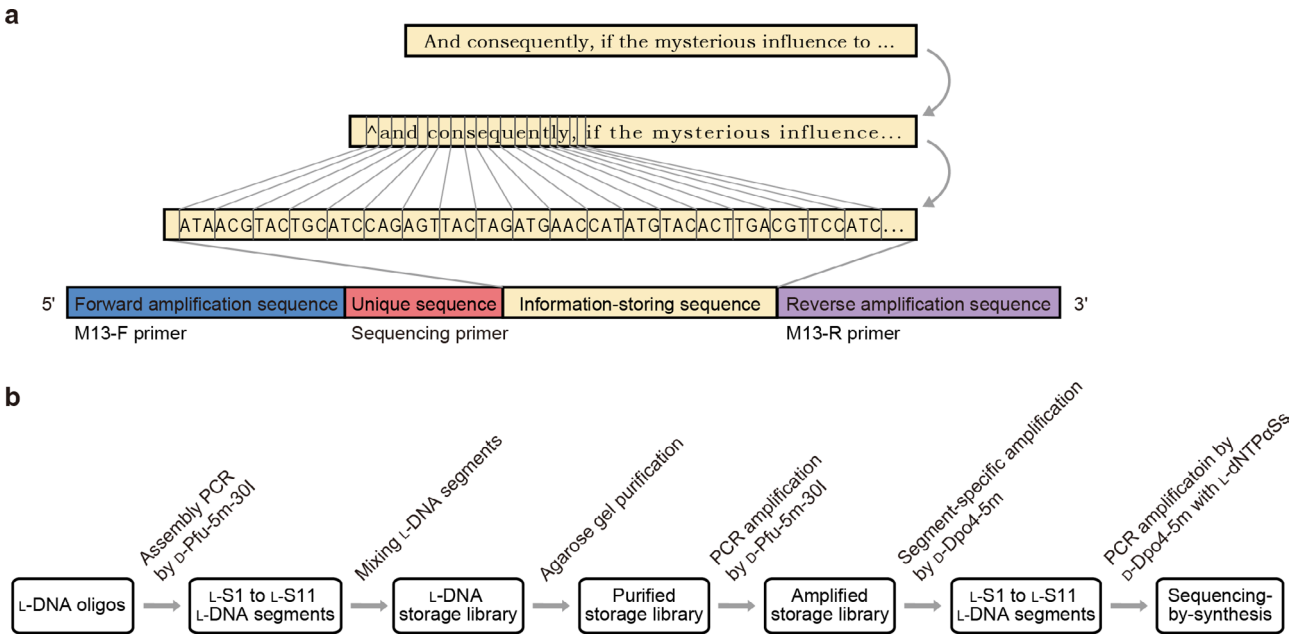
b



Extended Data Fig. 1 | Design of the mutant Pfu-N fragment. a, Mutant Pfu-N fragment amino acid sequence with an N-terminal His_6 tag and 4 point mutations (E102A, E276A, K317G, V367L, in parentheses) to introduce additional NCL sites. In addition, 25 isoleucine residues (underlined) were substituted to facilitate the chemical synthesis and reduce the synthesis costs for the mirror-image version. Colors of the amino acid sequences correspond to the peptide segment colors used in panel b. **b**, Synthetic route for synthesizing the mutant Pfu-N fragment.



Extended Data Fig. 2 | Design of the mutant Pfu-C fragment. a, Mutant Pfu-C fragment amino acid sequence with 1 point mutation (I540A, in parentheses) to introduce an additional NCL site. In addition, 16 isoleucine residues (underlined) were substituted to facilitate the chemical synthesis and reduce the synthesis costs for the mirror-image version. Colors of the amino acid sequences correspond to the peptide segment colors used in panel b. **b**, Synthetic route for synthesizing the mutant Pfu-C fragment.



Extended Data Fig. 3 | Mirror-image DNA information storage. a, Design of information-storing L-DNA segments. Caret, uppercase. **b**, Experimental procedures for mirror-image DNA information storage.

nature research

Corresponding author(s): Ting Zhu

Last updated by author(s): May 16, 2021

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed
☐	☒ The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☒	☐ The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐ A description of all covariates tested
☒	☐ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☒	☐ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☒	☐ For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐ Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	All RP-HPLC and ESI-MS analyses were performed on Shimadzu Prominence HPLC and LC/MS-2020, respectively. Agarose gels were scanned by Bio-Rad ChemiDoc XRS+. Polyacrylamide gels were scanned by Typhoon Trio+. Metagenomic sequencing data were obtained using Illumina NovaSeq 6000.
Data analysis	Fastp (version 0.20.1), FastQC (version 0.11.8), ImageJ (version 1.53a), MetaPhlAn2 (version 2.7.7)

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

The data that support the findings of this study are available within the paper and the Supplementary Information. Metagenomic sequencing data are available at the NCBI under BioProject number PRJNA707266.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	We arbitrarily chose to encode a paragraph containing 550 characters in 11 L-DNA segments of 220 bp in length, which is efficient for mirror-image assembly PCR and phosphorothioate sequencing.
Data exclusions	The metagenomic sequencing reads were processed for the removal of low-quality and contaminating sequences.
Replication	The L- and D-polymerases were synthesized consistently. PCR amplifications analyzed by agarose gel electrophoresis produced similar results in multiple replicates. The phosphorothioate sequencing of L-DNAs was performed multiple times with the encoded information reliably retrieved.
Randomization	Randomization is not applicable to this study.
Blinding	Blinding is not applicable to this study.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Human research participants
☒	☐ Clinical data
☒	☐ Dual use research of concern

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging