ACS Synthetic Biology

CADS: CRISPR/Cas12a-Assisted DNA Steganography for Securing the Storage and Transfer of DNA-encoded Information

Shiyuan Li, Jia-Kun Liu, Guoping Zhao, and Jin Wang

ACS Synth. Biol., **Just Accepted Manuscript** • DOI: 10.1021/acssynbio.8b00074 • Publication Date (Web): 29 Mar 2018

CADS: CRISPR/Cas12a-Assisted DNA Steganography for Securing the Storage and Transfer of DNA-encoded Information

Shi-Yuan Li[1,2], Jia-Kun Liu[1,2] , Guo-Ping Zhao[1] and Jin Wang[1,*]

[1]Key Laboratory of Synthetic Biology, Institute of Plant Physiology and Ecology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai 200032, China

[2]Shanghai Tolo Biotechnology Company Limited, Shanghai 200233, China

*To whom correspondence should be addressed. Tel.: +86-21-54971125; Fax: +86-21-54971126

Email: wangj01@hotmail.com

ABSTRACT

Because DNA has the merit of high capacity and complexity, DNA steganography, which conceals DNA-encoded messages, is very promising in information storage. Classical DNA steganography method hides DNA with "secret message" in a mount of junk DNA, and the message can be extracted by polymerase chain reaction (PCR) using specific primers (key), followed by DNA sequencing and sequence decoding. As leakage of the primer information may result in message insecurity, new methods are needed to better secure the DNA information. Here, we develop pre-key by either mixing specific primers (real key) with non-specific primers (fake key) or linking real key with 3′-end redundant sequences. Then, the single-stranded DNA (ssDNA) *trans* cleavage activity of CRISPR/Cas12a is employed to cut fake key or remove 3′-end redundant sequences, generating real key for further information extraction. Therefore, with the Cas12a-assisted DNA steganography method (CADS), both storage and transfer of DNA-encoding data can be better protected.

KEYWORDS: CRISPR, Cas12a, *trans* cleavage, DNA steganography, DNA storage

Because of the high capacity, complexity and long retention time of DNA, more attention has been paid to DNA steganography, which conceals DNA-encoded messages, in fields of both biotechnological industry and national security[1]. Nowadays, with the rapid growth of DNA reading and writing technologies[2-4], use of artificial synthetic DNA for data storage and communication has become a reality[5]. Clelland *et. al* first used the biosteganography strategy to hide a piece of meaningful DNA fragment in genomic DNA and the decoding process was performed by polymerase chain reaction (PCR) with a pair of primers (key) and subsequent DNA sequencing[6]. Therefore, similar to digital information, the "key" is also critical to the security of both storage and transfer of DNA-encoded data.

In a recent study, we showed that CRISPR associated protein Cas12a had *trans*-cleavage activity that cut non-targeted single-stranded DNAs (ssDNAs) but protecting targeted ssDNAs within the ternary complex of Cas12a/crRNA/targeted ssDNA[7]. Besides, Chen *et al.* also found the *trans*-cleavage activity of the ternary complex almost at the same time[8]. Here, using both the *trans*-cleavage activity and the protective effect against non-targeted and targeted ssDNAs,

respectively, we developed a DNA steganography method namely Cas12a-assisted DNA steganography (CADS) to better secure the DNA-encoded information.

As illustrated in the abstract figure, Alice wants to send secret messages to Bob, and the messages are hidden in DNA sequences. Alice first secretly discusses with Bob about the DNA encryption and DNA steganography rules and gives him the key (primers). Then Alice synthesizes the DNA sequences containing the secret messages and mixes it with junk DNAs. When Bob gets the mixed DNA, he uses the key to extract the DNA sequences and reads the message by following the DNA decoding rule. However, once Eve has the chance to steal the materials, the secret DNA sequences can be easily extracted and the messages will leak when the data are stored following a common rule, which is illustrated in Situation 1 in the abstract figure.

To solve the problem, CADS is developed to add an additional safeguard for the secret messages, and the messages will still be safe even if the mixed DNA is stolen and the key leaks (Situation 2 in in the abstract figure). In CADS, the key consists of two components (pre-key and the Cas12a/crRNA complex), which is similar to a combination key in digital cryptography that increases the data security. To prepare the pre-key, real primers are mixed with fake primers, and the pre-key has to be treated with the Cas12a/crRNA complex to generate the mature key for further information extraction. In detail, the Cas12a/crRNA complex targets the real primers and forms the ternary complex of Cas12a/crRNA/real primers, which protects the targeted real primers but meanwhile has the *trans*-cleavage activity against collateral ssDNAs and therefore cleaves the fake primers[7].

We first tested the efficiency of fake primer removal by mixing the real primer of 428-T1(R) (for primer sequences, please *ref to* **Table S1**) and the fake primer of -376(R) to make a pre-key. The pre-key was treated by Cas12a/crRNA-T1, where the targeted 428-T1(R) was protected but the non-targeted -376(R) was *trans*-cleaved by the Cas12a complex, leaving the real primer of 428-T1(R) for further PCR amplification. The treated key was then paired with primer -577F, and a single target band appeared (lane 4 in **Figure 1a**) when plasmid pUC18-T1 was employed as the PCR template (**Figure S1**). Otherwise, if the pre-key was not treated with the Cas12a complex to remove the fake primer, two PCR bands could be observed, resulting in the failure in extracting the target DNA sequences (lane 3 in **Figure 1a**).

We then increased the complexity of the CADS pre-key system, using specific primers mixed

with more fake primers (**Figure 1b** and **1c**) or the target DNA mixed with more junk DNA (*e.g.*, *E. coli* MG1655 genome) (**Figure 1d** and **1e**), and the target product of a single band was produced when the pre-key was correctly treated with the Cas12a complex; otherwise, two or three products would appear (**Figure 1b-1e**). Moreover, to prevent the decryption of crRNA or primer sequences, fake crRNAs (**Figure 1d**) and random N25 primers (**Figure 1e**) were added, respectively, which increased the CADS security but had no influence on PCR amplification.

Alternatively, the pre-key can be comprised of a pair of long oligonucleotides with the fake and real primers linked together, where the fake primer is on the 3′ terminus while the real primer locates at the 5′ terminus (**Figure 2a**). When the pre-key is directly used for PCR amplification, fake products are generated, while real products could only be obtained when the pre-key is treated by the Cas12 complex. This idea was successfully tested with the employment of plasmid pUC18-T1 as the template and primers of 1-461-F and T1-1260-R as the pre-key, and a 422-bp real PCR product was obtained after the Cas12a treatment (**Figure S2**). To further prove its practicability in information extraction, we synthesized both fake and true messages (**Supplementary Sequence**) according to the encryption rule described in a previous study[9], and mixed the messages with junk DNA (genomic DNA of both *E. coli* MG1655 and human 293T cell). Following the procedure in **Figure 2**, which included Cas12a digestion (**Figure 2a**), PCR amplification (**Figure 2b**), TA cloning, and Sanger DNA sequencing, the true message was successfully extracted (**Figure 2c**).

In this study, we employed the ssDNA *trans*-cleavage activity of the ternary complex Cas12a/crRNA/target DNA to treat pre-key, and developed the CADS method for securing the storage and transfer of DNA-encoded information. As it is difficult to get the full sequence of a small amount of short RNAs using present methods, the information of crRNA (*e.g.* around 40 nts in length) can still be safte even if the Cas12a/crRNA complex is stolen. Besides, fake crRNAs can be added in CADS to increase the difficulty to decrypt the true crRNAs, provding an additional protective layer for the security of the DNA-encoded information. Furthermore, the work described here may provide a base for further development of new concepts for keeping data safe, and improved CADS designs may include more fake primers, junk DNA messages and crRNAs in the reaction system, or be combined with the DNA encryption algorithms.

METHODS

Plasmid Construction

To construct plasmid pUC18-T1, PCR amplification was performed with primers of pUC18-T1-F and pUC18-T1-R (**Table S1**), using Tolo 2× HIFI DNA polymerase and employing plasmid pUC18 as the template. Then, the PCR amplicon was purified and self-ligated at the presence of T4 DNA ligase (Tolo Biotech.), T4 PNK (Tolo Biotech.) and DpnI (NEB). Both fake data and true data sequences were synthesized by Tolo Biotech., and were then inserted into the BamHI and XhoI sites of plasmid pET28a, obtaining pET28a-fake and pET28a-true, respectively.

Transcription of crRNAs and purification of Cas12a protein

Transcription of crRNAs: As previously described[10], the transcription templates were prepared through annealing of the synthesised oligonucleotides with T7-crRNA-F, and crRNAs were synthesised using the T7 High Yield Transcription Kit (Thermo Fisher Scientific). RNA was purified using the RNA Clean & Concentrator^{TM}-5 (Zymo Research) and quantified with NanoDrop 2000C (Thermo Fisher Scientific).

Purification of Cas12a protein: Cas12a in this study is FnCas12a from *Francisella tularensis* subsp. novicida U112, the protein expression and purification were same as previously study[10].

The CADS method

Cas12a reaction: Primers (0.5 μM each), crRNA(s) (1 μM each) and Cas12a (0.5 μM) were added and incubated with NEB buffer 3 in a 20-μl Cas12a reaction system. Normally, the reactions were performed for 1 h at 37 °C. Reactions were stopped by heating at 98 °C for 10 min.

PCR reaction: PCR was performed with Easytaq (Transgen) (**Figure S1**) or high-fidelity KOD FX (ToYoBo) (**Figure 2** and **Figure S2**). The reaction was performed in a 20-μl system with 0.25 μM primers (or 10-μl Cas12a-treated system) and 20-ng plasmid DNA (pUC18-T1) or 5-ng plasmid DNA (pUC18-T1) mixed with 2-μg genomic DNA (*E. coli* MG1655). In Figure 2B, the template for PCR amplification was the mixed DNA of pET28a-fake data (5 ng), pET28a-true data (5 ng) and genomic DNA (1-μg genomic DNA of *E. coli* MG1655 and 1-μg genomic DNA of human 293T cell).

TA cloning: The PCR products were first cloned using the TA cloning method before they

were applied for Sanger sequencing analysis (**Figure 2**). After PCR amplification using KOD FX, Easytaq was added and the reaction was maintained at 72 °C for 10 min to add a 3′-terminal "A", which was then purified and ligated to the T-vector of pMD19T (Takara), using T4 DNA Ligase. Subsequent Sanger sequencing was performed using the universal primer M13F-47.

FUNDING

This work was supported by grants from the Strategic Priority Research Program of the Chinese Academy of Sciences (Grant No. XDB19040200), the Youth Innovation Promotion Association CAS (No. 2017322) and the National Natural Science Foundation of China (31421061, 31430004 and 31300031).

ACKNOWLEDGEMENTS

We also thank Jiaoyun Ma for helpful discussion.

SUPPORTING INFORMATION

Supplementary tables, sequences and figures as described in the text.

AUTHOR CONTRIBUTION

S-Y. L., G-P. Z. and J. W. designed the experiments. S-Y. L. conducted most of the experiments. J-K. L. purified the Cas12a protein. S-Y. L. and J. W. wrote the paper. All authors read and approved the final version of the manuscript. J. W. supervised the whole project.

CONFLICT OF INTEREST

The authors declare no competing financial interest.

REFERENCES

1. Marwan, S.; Shawish, A.; Nagaty, K., DNA-based cryptographic methods for data hiding in DNA media. *Biosystems* **2016,** *150*, 110-118.
2. Carlson, R., The changing economics of DNA synthesis. *Nature biotechnology* **2009,** *27* (12),

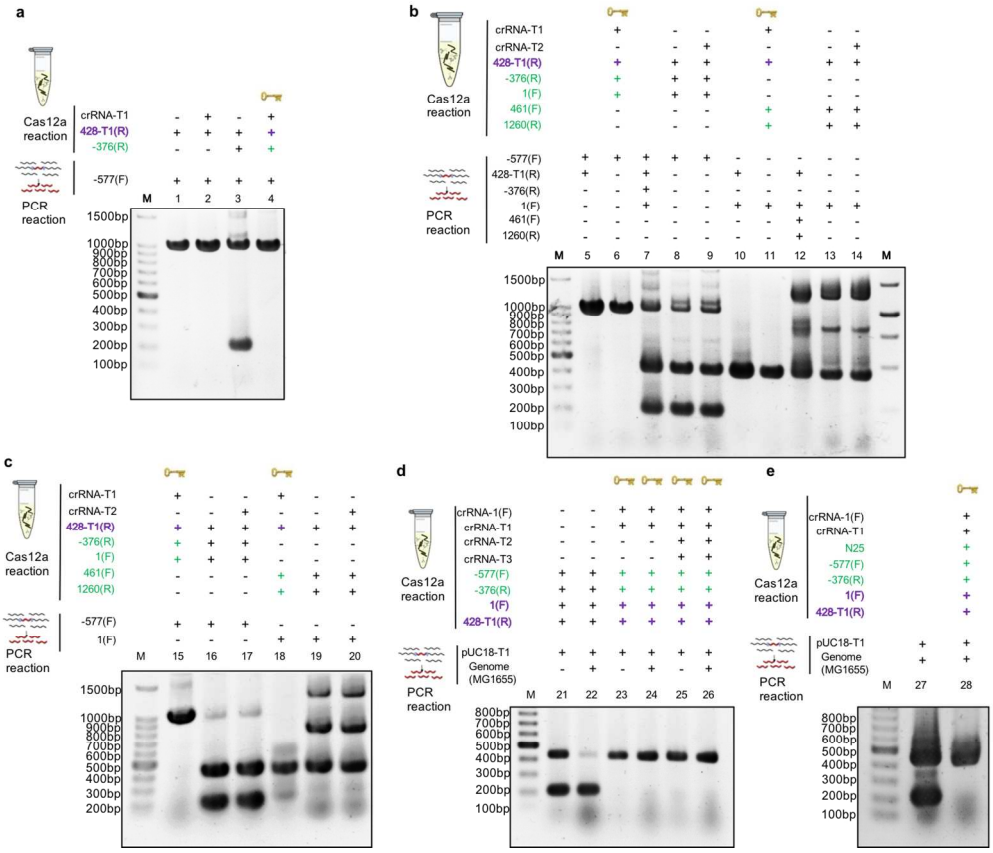
1091.
3. Medini, D.; Serruto, D.; Parkhill, J.; Relman, D. A.; Donati, C.; Moxon, R.; Falkow, S.; Rappuoli, R., Microbiology in the post-genomic era. *Nature Reviews Microbiology* **2008,** *6* (6), 419-430.
4. Carr, P. A.; Church, G. M., Genome engineering. *Nature biotechnology* **2009,** *27* (12), 1151-1162.
5. Castillo, M., From hard drives to flash drives to DNA drives. *AJNR Am J Neuroradiol* **2014,** *35* (1), 1-2.
6. Clelland, C. T.; Risca, V.; Bancroft, C., Hiding messages in DNA microdots. *Nature* **1999,** *399* (6736), 533-534.
7. Li, S.; Cheng, Q. X.; Liu, J. K.; Nie, X. Q.; Zhao, G. P.; Wang, J., CRISPR-Cas12a has both *cis*- and *trans*-cleavage activities on single-stranded DNA. *Cell Research* **2018**, DOI: 10.1038/s41422-018-0022-x.
8. Chen, J. S.; Ma, E.; Harrington, L. B.; Da Costa, M.; Tian, X.; Palefsky, J. M.; Doudna, J. A., CRISPR-Cas12a target binding unleashes indiscriminate single-stranded DNase activity. *Science* **2018**, DOI: 10.1126/science.aar6245.
9. Zakeri, B.; Carr, P. A.; Lu, T. K., Multiplexed Sequence Encoding: A Framework for DNA Communication. *PLoS One* **2016,** *11* (4), e0152774.
10. Li, S. Y.; Zhao, G. P.; Wang, J., C-Brick: A New Standard for Assembly of Biological Parts Using Cpf1. *ACS Synth Biol* **2016**, 5(12), 1383-1388

FIGURE LEGENDS

Figure 1. The CADS method for extracting true products by PCR amplification. (a) Employing the template of pUC18-T1, a 1-kb true PCR product was obtained with a pair of true primers of 428-T1(R) and -577(F) (lanes 1 and 2). While at the presence of the interference primer of -376(R), an additional product of about 200 bps in length appeared (lane 3). Then, with the treatment of Cas12a/crRNA complex, the non-targeted interference primer of -376(R) was *trans*-cleaved, and the fake band disappeared, leaving only the true band (lane 4). (b) Test of CADS with more fake primers added. pUC18-T1 was used as the template for PCR amplification. True bands were indicated in lanes 5 and 10. If pre-key was treated with the Cas12a/crRNA complex (*e.g.* without the Cas12a treatment in lanes 7 and 12, or treated with no crRNA in lanes 8 and 13, or treated with wrong crRNA (crRNA-T2) in lanes 9 and 14), fake primers of -376(R)/1(F) and 461(F)/1260(R) could not be removed, resulting in the appearance of three PCR bands with the true band concealed. Only after pre-key was treated by Cas12a/crRNA-T1, fake primers were successfully *trans*-cleaved, then generating the true band only (lanes 6 and 11). (c) Test of CADS with more complicated junk genomic DNA. Reactions in lanes 15-20 were the same as those in lanes 6, 8, 9,

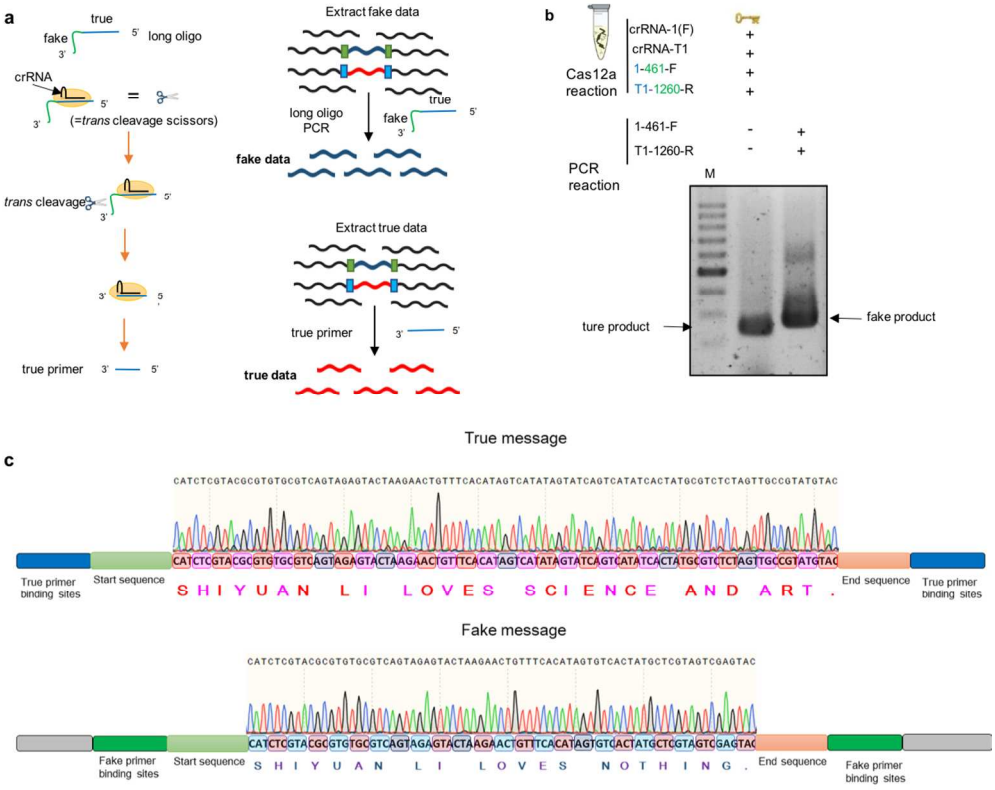
11, 13 and 14, respectively, except that template pUC18-T1 was mixed with the genomic DNA of *E. coli* MG1655. (d) Increasing the CADS complexity and security by addition of more fake crRNAs. Two pairs of primers of -577(F)/-376(R) and 1(F)/428-T1(R) were added in the PCR reaction (lanes 21 and 22), producing two PCR bands. When Cas12a together with crRNA-1(F) and crRNA-T1, which targeted the pair of primers of 1(F)/428-T1(R), respectively, were used to treat the primer pairs, the non-target primers were *trans*-cleaved and then the true product of about 400 bps in length could be obtained (lanes 23 to 26). To avoid the decryption of the crRNA information, misleading fake crRNAs (*i.e.* crRNA-T2 and crRNA-T3) were added, which had no influence on the PCR amplification (lanes 25 and 26). (e) Increasing the CADS complexity by addition of more junk primers (N25). In lane 27, pre-key primers (N25/-577(F)/-376(R)/1(F)/428-T1(R)) were used to perform PCR amplification. After Cas12a digestion, pre-key primers were *trans*-cleaved and 1(F)/428-T1(R) remained, and single true band appeared after PCR amplification (lane 28). As illustrated, partial primers were first added into the Cas12a reaction, while the rest were added prior to the PCR amplification, when the Cas12a had already been heat inactivated. Besides, golden keys were indicated above each reaction that employed the CADS treatment. Green colored primers stood for non-target ssDNAs (*i.e.* fake primers), which could be *trans*-cleaved by the Cas12a/crRNA/target ssDNA complex, while purple colored primers stood for target ssDNAs (*i.e.* true primers as the key to extract true data) that could be protected from *trans*-cleavage through binding of Cas12a/crRNA to form a complex. The experiments were repeated at least twice.

Figure 2. The CADS method for extraction of secret messages. (a) Illustration of the CADS method. The long oligos (pre-key) contained two parts, *i.e.* the 3′-end fake primer and the 5′-end true primer. Only when the pre-key was treated by the Cas12a/crRNA complex, true products could be obtained. (b) Extraction of the true products by CADS. Without Cas12a treatment, a fake band was produced. This experiment was repeated at least three times. (c) Decoding secret messages after TA cloning and subsequent Sanger DNA sequencing.

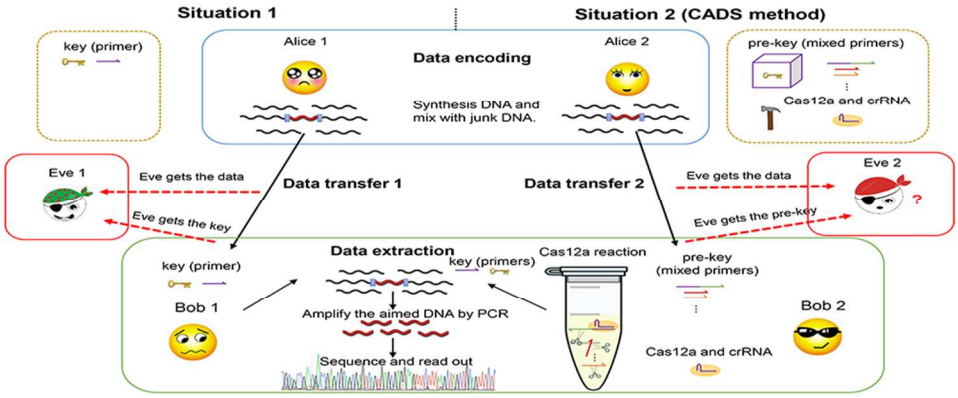


The CADS method for extracting true products by PCR amplification.

265x225mm (150 x 150 DPI)



The CADS method for extraction of secret messages.



For Table of Contents Use Only

88x34mm (300 x 300 DPI)