Secure data storage into DNA molecules compliant with biological constraints: Ensuring the confidentiality of data stored into DNA molecules
Chloé Berton, Gouenou Coatrieux, Dominique Lavenier

To cite this version:
Chloé Berton, Gouenou Coatrieux, Dominique Lavenier. Secure data storage into DNA molecules compliant with biological constraints: Ensuring the confidentiality of data stored into DNA molecules. RITS 2022 - Recherche en Imagerie et Technologies pour la Santé, May 2022, Brest, France. pp.1-1. hal-03817968

HAL Id: hal-03817968
https://hal.science/hal-03817968
Submitted on 17 Oct 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
Secure data storage into DNA molecules compliant with biological constraints
Ensuring the confidentiality of data stored into DNA molecules

Motivation - Considering the security of a promising storage medium

- **Context** – Actual storage technologies (flash memory, hard drives, magnetic tapes,…) are outpaced by the exponential rise of digital data production [1].
- **Advantages of DNA storage** [2] – Density of 10^21 bytes in one gram (10^6 times more compact than hard disks), durability for centuries, energy cost close to zero (molecules at room temperature with no maintenance).
- **Motivation** – Introducing security to ensure the confidentiality of the data stored into DNA molecules.

DNA data storage - A new storage medium

- **Principle** – [WRITING] Encode binary data into 4-base sequences following the DNA structure, transfer this data into synthetic DNA molecules. [READING] Amplify encoded sequences of interest and get several reads of them with a sequencing device. Reads are then processed and decoded back to binary data.
- **Constraints** – i) Biological DNA synthesis and sequencing are imperfect and introduce errors. ii) Devices have structural DNA requirements when generating 4-base sequences.
- **Vulnerabilities** – This chain is notably vulnerable to: theft or cloning of molecules; spying attacks on the sequencing or synthesis devices; DDoS attack by adding fake DNA sequence to confuse sequencing.

Dynamic encoding - a solution for confidentiality

- **Challenge** - Ensure confidentiality under biological constraints while approaching the ideal rate of 2 bits of information per base.
- **Solution** – A three step coding process that includes encryption, dynamic data encoding and error-correction code.

**Experimental results**

- **Simulation** of the biological processes using the simulator from [3].
- **Results** - no homopolymers longer than N, G-C content of 43-57%, data recovery without errors. For N=4, information rate of 1,875 bits per base.

**Conclusion and future work**

- **Confidentiality** in the entire storage chain that takes into account biological constraints.
- **Encoding solution** independent from encryption algorithm and error-correction code, adaptable to the size of unwanted patterns.
- Extend the approach to other synthesis and sequencing technologies.

Contact: chloe.berton@imt-atlantique.fr, gouenou.coatrieux@imt-atlantique.fr