New Trends in DNA-Based Data Storage

BOOK OF ABSTRACTS

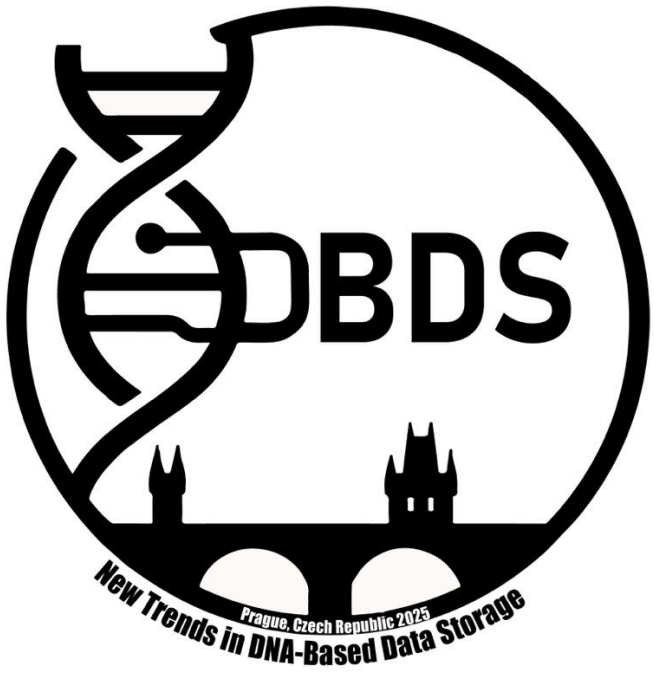
3–6 June, 2025
Prague, Czech Republic

New Trends in DNA-Based Data Storage

3–6 June 2025

Profesní dům, Malostranské nám. 2/25

Prague, Czech Republic



Organized by:

Local Organizing Committee:

- Jaroslav Kočišek
- Katarína Jakubová (Conference Secretary)
- Leo Sala
- David Ruiz
- Tereza Zápotocká
- Karen Cardos

J. Heyrovský Institute of Physical Chemistry, Czech Academy of Sciences

International Scientific Committee:

- Adrian Keller (University of Paderborn, Germany)
- Wooli Bae (University of Surrey, UK)
- Steven de Feyter (KU Leuven, Belgium)
- Thomas Heinis (Imperial College, UK)

SPONSORS

tilibit nanosystems

DATANA by BioSistemika

BIOCOMPUTE

NEO

Co-funded by the European Union

Schedule

Start	End		
Monday, June 2, 2025			
12:20	13:45		Start of registration (Kaiserstein Palace)
13:45	18:15		DigNA meeting for portfolio members
19:00			Welcome reception (Kaiserstein Palace)
Tuesday, June 3, 2025			
8:45	9:00		**Opening**
9:00	9:45	Eitan Yaakobi	**Olgica Milenkovic**: Challenges in DNA-based data storage
9:45	10:00		**Ramy Khabbaz**: Novel efficient codes for correcting random edit errors in DNA Storage
10:00	10:30		**Ian Hoffecker**: Writing digital data to DNA without de novo synthesis
10:30	10:50		**Coffee Break**
10:50	11:20	Wooli Bae	**Petr Šulc**: Modeling-based design and information storage in self-assembled DNA nanotechnology systems
11:20	11:50		**Natália Neto Pereira Cerize**: Bridging bytes and bases: CODEC strategies for parallel oligonucleotide synthesis in DNA data storage
11:50	12:20		**Friedrich Simmel**: Electromechanical switching of bistable DNA nanostructures for at least 200,000 times
12:20	14:00		**Lunch**
14:00	14:45	Fulvio Bernardini	**Robert Grass:** DNA stability and preservation
14:45	15:00		**Isabel Obieta Vilallonga:** DigNA portfolio introduction
15:00	15:15		**Fabian Schroeder:** Encoding schemes for DNA data storage via oligonucleotide assembly
15:15	15:30		**Gangamallaiah Velpula:** Controlling the orientation of chiral double-L DNA origami on mica
15:30	15:45		**Pierre-Yves Burgi:** DNAMIC Project: DNA microfactory for autonomous archiving
15:45	16:15		**Coffee Break**
16:15	16:30	Thomas Heinis	**Marthe Colotte:** The preservation of DNA integrity in the context of DNA data storage
16:30	16:45		**Cecilia Wetzl:** Integrating DNA-based memory in water-resistant electrospun polymer fibers for non-destructive data retrieval
16:45	17:00		**Joel Spratt:** DropSCRIBE: using droplet microfluidics to store digital data
17:00	17:30		**Damien Woods:** Thermodynamically favoured DNA computing and storage
17:30	18:00		**Zohar Yakhini:** Composite alphabets for DNA based data applications
18:00			**Poster Session I & Refreshments**

Start	End		
Wenesday, June 4, 2025			
9:00	9:45	Raja Apuswamy	**Laura Na Liu:** Engineering complexity: the promise of DNA superstructures
9:45	10:00		**Omer Sabary:** Scalable and robust DNA-based storage via coding theory and deep learning
10:00	10:30		**Manish K. Gupta:** A bird's-eye view on DNA storage simulators and computing tools
10:30	10:50		**Coffee Break**
10:50	11:20	Ilko Bald	**Veikko Linko:** Stability of nucleic acid origami
11:20	11:50		**Tim Liedl:** Bridging the scales with DNA origami
11:50	12:20		**Yixin Zhang:** Non-linear manipulation of DNA codes
12:20	14:00		**Lunch/Conference Photo**
14:00	14:45		**Mark Bathe** (Virtual)**:** Scalable search of massively pooled nucleic acid samples: from cold data archives to global genomics
14:45	19:00		**Free Afternoon / City Tours**
19:00			**Conference Dinner at the Strahov Monastery Brewery**

CONFERENCE DINNER VENUE

Strahov Monastery Brewery

Strahovské nádvoří 301, 118 00 Praha 1-Hradčany

Tucked inside the historic Strahov Monastery, founded in 1142, this brewery continues a brewing tradition more than 600 years old. It combines time-honored craftsmanship with modern techniques to create over 25 distinctive beers. Their signature brew, St. Norbert's beer, pays tribute to the founder of the Premonstratensian Order, which has called the monastery home for centuries. With its vaulted stone halls and peaceful courtyard, the setting is as memorable as the beer itself. A visit here offers not just great flavor, but a true taste of Czech brewing heritage.

Directions from the conference venue to the Strahov Brewery:

www.klasterni-pivovar.cz

Start	End		
Thursday, June 5, 2025			
9:00	9:45	Natália N. P. Cerize	**David Soloveichik:** Bridging DNA storage and computation
9:45	10:00		**Asad Usmani:** Fast RDF-analytics from large-scale DNA-based archives
10:00	10:30		**Amelie Heuer-Jungemann:** Silicification for robust and reversible protection of DNA nanostructures
10:30	10:50		**Coffee Break**
10:50	11:20	Ivan Barišić	**Jie Song:** Xeno nucleic acids for robust long-stranded data storage
11:20	11:50		**Raja Appuswamy:** Motif-based approaches for scaling read/write cost in DNA storage
11:50	12:20		**Albert Keung:** Thermodynamic and kinetic challenges of DNA-based data
12:20	14:00		**Lunch**
14:00	14:45	David Soloveichik	**Marc Antonini:** Towards efficient compression for DNA-based data storage
14:45	15:00		**Maria Abu-Sini:** Coding for the DNA decay problem along with edit errors
15:00	15:15		**Abeer Eshra:** Programmable and renewable thermodynamically favoured DNA computing and storage
15:15	15:30		**Kaya Wernhart:** DNAByte: a modular framework for the simulation and analysis of DNA-based data storage.
15:30	15:45		**Dganit Hanania:** On the capacity of DNA labeling
15:45	16:15		**Coffee Break**
16:15	16:30	Robert Peharz	**Mayan Rivlin Gilboa:** Sequence determinants of CRISPR-Cas12a off-target reaction
16:30	16:45		**Robin T. Vermathen:** Molecular encryption for DNA data storage
16:45	17:00		**Yiwei Zhang:** Secondary structure avoidance codes for DNA storage and computing
17:00	17:30		**Vladimíra Petráková:** DNA-mediated assembly of plasmonic and fluorescent nanoparticles for nanophotonics
17:30	18:00		**Eitan Yaakobi:** Minimizing the Sequencing Coverage in DNA-based Data Storage Systems
18:00			**Poster Session II & Refreshments**

Friday, June 6, 2025			
9:00	9:45	Amelie Heuer-Jungemann	**Chueh Loo Poh:** DNA sequence-based data storage
9:45	10:00		**Jokin Yeregui Elosua:** Programmable vesicles embedded in electrospun nanofibers for the triggered retrieval of encoded DNA digital data
10:00	10:30		**Matteo Castronovo:** Towards enzymatically modifying DNA within compact memory storage devices
10:30	10:50		**Coffee Break**
10:50	11:20	Adrian Keller	**David Turek:** Creating an associative memory through combinatorial assembly
11:20	11:50		**Tuvi Etzion:** Representing codes on DNA using enzymatic labeling
11:50	12:20		**Ivan Barišić:** CATANA, an online modelling environment for proteins and nucleic acid nanostructures
12:20	13:45		**Conference closing & announcement of student prizes for the best hot topic and poster presentations**

List of Contributions

I. Talks

● Plenary ● Invited ● Hot Topic

II. Posters

ABSTRACTS

I. TALKS

CHALLENGES IN DNA-BASED DATA STORAGE

Milenkovic O.

Department of Electrical and Computer Engineering, University of Illinois, Urbana-Champaign, 311 Coordinated Science Lab
milenkov@illinois.edu

We provide an overview of the first known example of topological native DNA based data storage systems, including DNA Punchcards [1], 2DDNA [2] and DNA Tails [3]. DNA Punchcards use Crispr and other nicking enzymes to store binary information in the form of nicks/no nicks at predefined sites of the sugar phosphate backbone of native bacterial DNA. 2DDNA extends this paradigm by using nicks to store metadata on synthetic DNA, with the added benefit of ease of erasing and rewriting. In this context, we also describe the first machine learning approach that mitigates the use of error-correction coding to restore corrupted images. We then proceed to describe DNA Tails, and in particular, a DNA Flash Memory based on tails, where at each nicking site, tails of different lengths are grown enzymatically to increase the encoding alphabet. In addition to experimental results for all three storage models, we present a cohort of theoretical findings pertaining to specialized error-control coding techniques.

References

[1] Tabatabaei, S. Kasra, Boya Wang, Nagendra Bala Murali Athreya, Behnam Enghiad, Alvaro Gonzalo Hernandez, Christopher J. Fields, Jean-Pierre Leburton, David Soloveichik, Huimin Zhao, and Olgica Milenkovic. "DNA punch cards for storing data on native DNA sequences via enzymatic nicking." *Nature communications* 11, no. 1 (2020): 1742.

[2] Pan, Chao, S. Kasra Tabatabaei, SM Hossein Tabatabaei Yazdi, Alvaro G. Hernandez, Charles M. Schroeder, and Olgica Milenkovic. "Rewritable two-dimensional DNA-based data storage with machine learning reconstruction." *Nature communications* 13, no. 1 (2022): 2984.

[3] Sima, Jin, Chao Pan, S. Kasra Tabatabaei, Alvaro G. Hernandez, Charles M. Schroeder, and Olgica Milenkovic. "DNA Tails for Molecular Flash Memory." *arXiv preprint arXiv:2505.03629* (2025).

NOVEL EFFICIENT CODES FOR CORRECTING RANDOM EDIT ERRORS IN DNA STORAGE

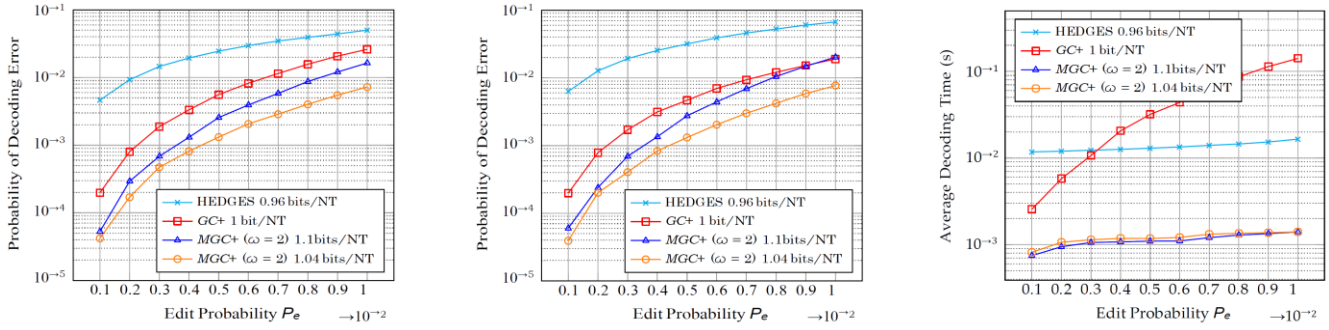
Khabbaz R., Antonini M., Hanna S. K.

Côte d'Azur University, CNRS, I3S, Sophia Antipolis, France,
ramy.khabbaz@etu.univ-cotedazur.fr, marc.antonini@cnrs.fr, serge.kas-hanna@cnrs.fr

DNA is a promising data storage medium due to its high density and durability, but ensuring data reliability is challenging due to *edit* errors arising from synthesis, sequencing, and storage, including *deletions* and *insertions* (*indels*), as well as substitutions. Current synthesis limitations require storing data in short DNA strands (oligos), typically a few hundred nucleotides long, necessitating short blocklength edit-correcting codes. Most prior work targets specific errors (e.g., deletions only) rather than handling all three types simultaneously, often relying on asymptotic results requiring impractically long codes. Instead, existing techniques for DNA storage predominantly rely on sequencing redundancy (coverage) to enhance reliability. Notable exceptions like HEDGES [1] and GC+ [2] directly correct edit errors in short sequences, offering practical solutions.

GC+ codes [2] correct indels by mapping them from the DNA (quaternary) field to erasures and substitutions in a higher field, which are then handled via Reed-Solomon (RS) codes. Its decoding relies on a *guess-and-check* mechanism that generates and validates offset hypotheses using the RS code. We enhance GC+ by inserting periodic markers (period ρ) and designing a novel Maximum A Posteriori (MAP)-based decoding algorithm that estimates the most likely offset pattern. Rather than *exhaustive* guessing, the decoder generates a small set of *informed* guesses based on the MAP estimate. While the inclusion of periodic markers introduces additional redundancy, the new algorithm reduces the redundancy required by the RS code, resulting in an overall improvement in code rate. The resulting Marker-GC+ (MGC+) code outperforms HEDGES and GC+ across all three key metrics: information density, decoding latency, and reliability (lower probability of decoding error).

In the simulations below, we compare the decoding error and decoding time under random i.i.d. edit errors with probability P_e for messages of length 176 bits encoded into DNA. Fig. (a) assumes equal probabilities for deletions, insertions, and substitutions ($P_d = P_\mathrm{i} = P_s = P_e/3$), while Fig. (b) uses unequal ratios inspired by recent experimental data [3]. Fig. (c) compares decoding times for the error ratios in (b). The results show that MGC+ significantly improves reliability and decoding latency while achieving higher information densities.



References:

[1] Press, William H., et al. "HEDGES error-correcting code for DNA storage corrects indels and allows sequenceconstraints." Proceedings of the National Academy of Sciences 117.31 (2020): 18489-18496.

[2] Kas Hanna, Serge "Short systematic codes for correcting random edit errors in dna storage." 2024 IEEE International Symposium on Information Theory (ISIT). IEEE, 2024.

[3] Gimpel, Andreas L., et al. "A digital twin for DNA data storage based on comprehensive quantification of errors and biases." Nature Communications 14.1 (2023): 6026.

WRITING DIGITAL DATA TO DNA WITHOUT DE NOVO SYNTHESIS

Hoffecker I.

Dept. of Gene Technology, KTH Royal Institute of Technology, Science for Life Laboratory, Solna, Sweden
Ian.hoffecker@scilifelab.se

TBA

MODELING-BASED DESIGN AND INFORMATION STORAGE IN SELF-ASSEMBLED DNA NANOTECHNOLOGY SYSTEMS

Šulc P.

[1]*The Biodesign Institute, Arizona State University, Tempe, AZ 85287, USA*
[2]*Department of Biosciences, Technical University of Munich, ZNN Garching, 85748, Germany*
psulc@asu.edu

We will present a mechanism to design self-assembly of arbitrary shape from individual building blocks made from DNA origami. Our computational pipeline can find the smallest number of particles that can reliably assemble into target design. We show applications of our framework to some highly coveted structures, such as fractal Menger cube design or diamond lattices. We then show how our framework can be extended to switchable nanostructures with internal "state" that can be switched between different conformations. Our design and modeling pipeline is based on the oxDNA ecosystem tools which we use to inform the experimental realization.

BRIDGING BYTES AND BASES: INTEGRATED CODEC AND SYNTHESIS STRATEGIES FOR DNA-BASED DATA STORAGE

da Costa Martins A. G.[1], Wisinewski H. R.[1], Dalibera N.C.[1];
Verona B. M.[1]; Cerize, N. N. P.[1*]

**corresponding author: NNP Cerize - ncerize@ipt.br*

The Prometheus Project, a collaboration between the Institute for Technological Research (IPT) and Lenovo, pursues DNA-based cold-data storage by integrating a complete molecular stack that spans binary-to-nucleotide encoding/decoding (CODEC) and DNA synthesis hardware. We developed Pantheon, a modular CODEC that converts digital files into four-letter nucleotide strings while adding indexing, addressability, Reed–Solomon error correction, and a Boot-Sector layout recommended by the Storage Networking Industry Association (SNIA). Adaptive decoding recovers data across heterogeneous sequencing platforms and compensates for synthesis-specific error spectra.

Concurrently, we engineered Brazil's first DNA synthesizer: a microfluidic, valve-controlled platform that supports both phosphoramidite "base-by-base" chemistry and template-independent enzymatic extension. Miniaturised reaction chambers reduce reagent volumes, shorten cycle times, and enable electronically gated nucleotide delivery, providing first-level spatial parallelisation. Integrated temperature regulation maintains enzymatic activity for long, mixed-base runs. Facility-level design considerations align the instrument with data-center requirements for footprint, energy consumption, and automation.

We validated our encoding and decoding system by system by a 2 MB dataset, synthesising 15 oligonucleotide batches, and recovering the files after high-throughput sequencing (>60 000× mean coverage per oligo). Measured deletion rates remained below 0.3 %, enabling loss-free decoding. On hardware, benchmarking against commercial DNA writers used on biotech showed a fourfold reduction in cost per nucleotide and a 2.5-fold increase in throughput. Outcomes include seven patent applications, two peer-reviewed articles, and public demonstrations at national and international venues.

The Prometheus Project established an integrated ecosystem in Brazil for DNA data storage innovation, covering hardware, software, and biochemical process development. It positions the country as one of the few with capabilities across the full technology stack required for molecular information systems.

ELECTROMECHANICAL SWITCHING OF BISTABLE DNA NANOSTRUCTURES FOR AT LEAST 200,000 TIMES

Simmel F. C.

Department of Bioscience, TUM School of Natural Sciences, Technical University Munich, Am Coulombwall 4a, Munich, D-85748, Germany
simmel@tum.de

Bistable switches are fundamental components for the realization of memory elements and can also serve as building blocks for switching networks capable of performing simple information processing tasks. In this talk, we present a mechanically bistable nanoscale switch constructed from DNA origami, inspired by macroscopic snap-through mechanisms. We demonstrate that this DNA-based switch can be reliably toggled between two distinct mechanical states using an externally applied electric field. Readout of the switch state can be achieved either via fluorescence measurements or through light scattering from gold nanoparticles attached to the structures. The latter method is not affected by photobleaching, making it particularly suitable for long-term monitoring of device performance, including the study of device failure and fatigue phenomena. In our experiments, we show that, under our current reaction conditions, individual switches can undergo more than 100,000 successful switching cycles.

DNA STABILITY AND PRESERVATION

Grass R.

D-CHAB, ETH Zurich, Vladimir-Prelog-Weg 1, 8093 Zurich ,Switzerland
rograss@ethz.ch

The scientific information on the stability of DNA is Janus-like, where it is considered highly stable in the context of paleontology,[1] but is regarded as sensitive for longer storage horizons in biochemistry.[2] Possible reasons behind this contradiction are the high redundancy of DNA in ancient samples, and the inorganic matrix (bones, teeth) in which natural samples are preserved. During this talk I will be discussing aspects of DNA stability in the context of DNA data storage applications. Very similar to ancient DNA, we can profit from both DNA redundancy and protection of the molecule during storage in appropriate physical matrixes.

Analogy with ancient samples and profiting from an improved understanding of the DNA decay promoting reactions[3] has aided the design of DNA protection measures. This has resulted in DNA encapsulation matrixes and containers, which protect the DNA molecules from reaction with water and oxygen, the main decay promoting reactants.

An additional highly important parameter in the scope of data storage is the length of the individual DNA strands.[4] Somewhat counterintuitively, short DNA strands are more stable than long DNA strands, as the susceptibility of DNA breakage scales linearly with the strand length. To increase the usability of long DNA strands, DNA repair reactions have been investigated and can repair hydrolysis related DNA damage.[5]

The combination of synthetic DNA and stable storage in inorganic matrixes enables several applications, including DNA barcoding,[6] DNA of things [7] and in-product cryptography [8], which I will shortly introduce during the presentation.

References:

[1] M. E. Allentoft, *Proc. Biol. Sci. 279*, 4724-33, 2012.

[2] N.V. Ivanova et al., *Mol. Ecol. Resources 13*, 890-898, 2013.

[3] Bonnet et al. *Nucl. Acid. Res. 38*, 1531-1546 2010.

[4] K. Matange et al., *Nat. Commun. 12*, 1358, 2021.

[5] Meiser et al. *Commun. Biol. 5*, 1117 2022.

[6] Paunescu et al. *Angew. Chem. Int. Ed. 52*, 4269-4727, 2013.

[7] Koch et al. *Nat. Biotechnol. 38*, 39-43, 2020.

[8] Luescher et al. *ACS Nano 18*, 30774, 2024.

ENCODING SCHEMES FOR DNA DATA STORAGE VIA OLIGONUCLEOTIDE ASSEMBLY

Schroeder F.[1], Wernhart K.[1], Barisic I.[2]

[1]Eko Refugium, Crno Vrelo 2, Slunj Croatia
[2]Austrian Institute of Technology, Giefinggasse 4, 1210 Vienna, Austria
fabian.schroeder@ekorefugium.com

DNA is the information storage medium of nature. It is highly dense and extremely stable and, thus, has the potential to solve mankind's increasingly pressing cold data storage problems. While several proof-of-concepts have been demonstrated, the prohibitive cost of synthesis remains a major barrier to large-scale adoption. While it is unclear whether the de-novo synthesis of DNA will ever reach an effective cost level, the assembly of shorter oligonucleotides offers a cost-effective alternative.

In this presentation, we will introduce a general framework for assembly-based DNA data encoding, which consists of two basic components: a predefined library of oligonucleotides and an encoding function that maps bit sequences onto sequences of these library elements. We will explore some examples of this approach and analyse critical design parameters, including (i) data density, (ii) library size, and (iii) oligonucleotide length.

Furthermore, we will discuss the inherent error correcting properties of such encodings, arising from their constrained sequence space. Finally, we present preliminary results from both in silico simulations and in vitro experiments aimed at optimizing oligo libraries for reliable and efficient data encoding.

References:

[1] K. A. S. Immink and K. Cai, "Properties and Constructions of Constrained Codes for DNA-Based Data Storage," in *IEEE Access*, vol. 8, pp. 49523-49531, 2020, doi: 10.1109/ACCESS.2020.2980036.

CONTROLLING THE ORIENTATION OF CHIRAL DOUBLE-L DNA ORIGAMI ON MICA

Velpula G.[1], Tomm E.[2], Shen B.[3], Mali K.S.[1], Keller A.[2], De Feyter S.[1]

[1]*Department of Chemistry, Division of Molecular Imaging and Photonics, KU Leuven, B-3001 Leuven,Belgium*
[2]*Paderborn University, Technical and Macromolecular Chemistry, Warburger Str.100, Paderborn 33098,Germany*
[3]*Department of Medical Biochemistry and Biophysics, Karolinska Institutet, 17177 Stockholm, Sweden*
gm.velpula@kuleuven.be; adrian.keller@ uni-paderborn.de; steven.defyter@kuleuven.be

The ability to precisely control DNA origami orientation holds immense potential for a wide range ofapplications.[1] This includes the development of advanced metamaterials, highly sensitive chiral sensing platforms, high-density data storage devices, and sophisticated drug delivery systems. Any method to achieve DNA origami orientation control is therefore attractive for both fundamental research and technological innovation.

This presentation explains the simple yet powerful approach to control the orientation of DNA origaminanostructures upon deposition on mica, an intrinsically negatively charged substrate. By varying theMg^{2+} concentration of the buffer solution, we demonstrate the ability to control the orientation of a chiral 2D DNA origami shape on the mica surface (Figure 1). A chiral double-L (**CDL**) DNA origamistructure was used that can adopt either an **S** or **Z** orientation upon adsorption. **CDL** adsorption on mica was probed by atomic force microscopy (AFM), both for dried samples as well as at the liquid-solid interface. Distributions of **S** and **Z** orientations are shown to depend dramatically on the Mg^{2+} concentration, ranging from randomly oriented **CDL**s to exclusive **S**. The results are explained by considering Mg^{2+} induced conformational transitions in the 3D shape of the 2D **CDL DNA** origami.

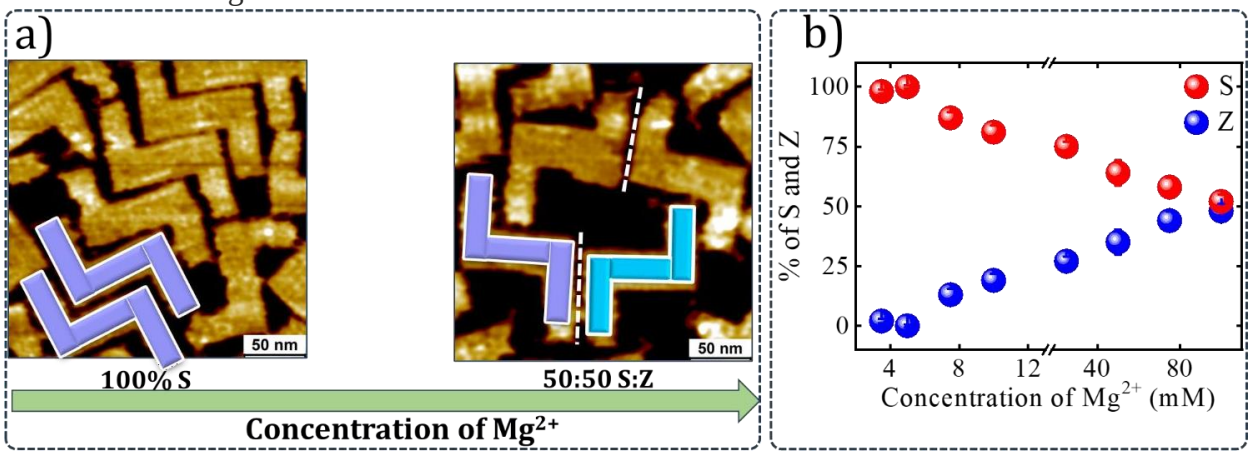


Figure1. (a) Schematic representations and high-resolution AFM images (scale bar is 50 nm) illustratethe mirror-image **S** and **Z** orientations of **CDL DNA** origami adsorbed on mica as a function of Mg^{2+}concentration. (b) % of **S** and **Z** orientations of the **CDL DNA** origami as a function of Mg^{2+} concentration at TAE buffer/mica interface.

References:

[1] Shen *et al*. Plasmonic nanostructures through DNA-assisted lithography. Sci. Adv. 2018, 4 (2), eaap8978.

DNAMIC PROJECT: DNA MICROFACTORY FOR AUTONOMOUS ARCHIVING

Burgi P.-Y.

IT Department, University of Geneva, 1205 Geneva, Switzerland
pierre-yves.burgi@unige.ch

The DNAMIC project[1] aims to develop, among others, an autonomous DNA microfactory for secure, long-term data storage, leveraging DNA as a high-density medium. Our microfactory concept (Figure 1) will combine autonomous, independent modules for DNA processing from encoding to decoding, including synthesis, amplification, library preparation and sequencing. The information to be archived is first formatted using the OLOS.swiss long-term preservation system, powered by the DLCM technology [1]. This platform represents one of the core elements of the proposed DNA archiving system and conforms to the ISO 14721 reference model (OAIS), which covers data ingestion, archiving, based on Archival Information Packages (AIP), and access. In addition, thanks to metadata and self-documentation, the standard enables content archived in DNA to continue to be interpreted in the future, even if the original software or systems are not available in the long term. By maintaining an agnostic view of file formats, OLOS makes it possible to preserve any type of file while securing their integrity by means of checksums. The AIPs, transmitted to the microfactory supervisor, which distributes the tasks to the appropriate modular units, are first processed by a CODEC to generate oligonucleotide sequences for all data and query strands. To retrieve an AIP, the synthesizer fabricates the query primers, and the samples are sent to the amplification module. The data is then divided into two sets. One is sent back to be stored for later sequencing, and the other is forwarded to the sequencing stage. A bidirectional sequencing algorithm allows execution to be stopped as soon as all data has been decoded, minimizing time and reagent consumption. Once decoded into AIP format, the data is returned to the OLOS archive system for retrieval. Thanks to its modular approach, such microfactory-based automation can, in the longer term, act as a catalyst for other applications serving a wider community.

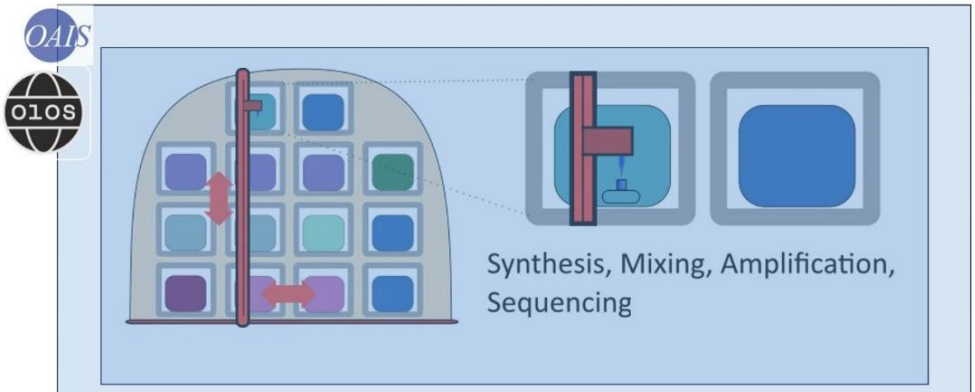


Figure 1: The DNA microfactory concept

References:

[1] Burgi, P. Y. et al. OAIS-compliant digital archiving of research and patrimonial data in DNA. *iPRES - 18th International Conference on Digital Preservation,* 2022

[1] DNAMIC (https://dnamic.org) has received funding from the European Union's Horizon 2020 Research and Innovation Programme under Grant Agreement no. 101115389. Swiss participants in this project are supported by the Swiss State Secretariat for Education, Research and Innovation (SERI) under contract numbers 23.00300 and 23.00487.

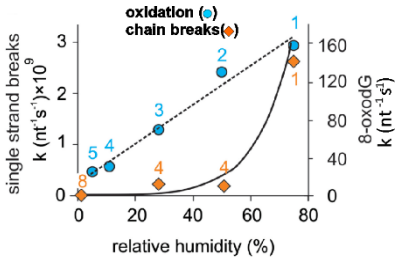
THE PRESERVATION OF DNA INTEGRITY IN THE CONTEXT OF DNA DATA STORAGE

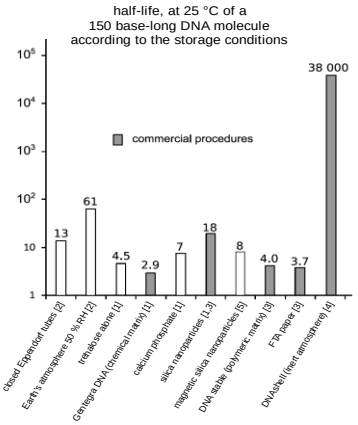
Colotte M.[1], Bonnet J.[1,2]

[1]*Imagene company, Gradignan, France*
[2] *BRIC, U 312 INSERM, Université de Bordeaux, France*
colotte@imagene.eu

The preservation of DNA integrity is a critical challenge in various fields, including genomics, biobanking, forensic science, biotechnology, and, more recently, data storage. DNA degradation occurs through well-documented mechanisms such as hydrolysis or oxidation, leading to base loss or modification, chain fragmentation, and ultimately, loss of information. From these degradation pathways it appears that protecting DNA from oxygen, ozone and especially water is essential for developing effective storage solutions.

A variety of DNA preservation methods have been designed to mitigate these effects. The conventional cold storage techniques, remain widely used but require continuous energy input or maintenance and offer limited scalability, constraints that become particularly significant in the context of DNA data storage, where the sheer volume of information demands more sustainable and autonomous solutions. Alternative strategies have emerged, such as drying on solid surfaces, inclusion in protective matrices, chemical encapsulation and especially physical encapsulation in an anoxic and anhydrous environment which enables storage at room temperature, offering a reliable and energy-independent solution.





The evaluation of storage systems is based on key performance indicators, including DNA recovery efficiency, sequence fidelity, volume and resistance to environmental factors. A systematic review of the various proposed procedures highlights the advantages and limitations of each method, providing valuable insights into their relevance for different applications, particularly in DNA data storage. This state-of-the-art analysis offers a comprehensive overview of the most effective approaches and supports informed decision-making in selecting the optimal DNA storage strategy.

References:

[1] Organick et al., An Empirical Comparison of Preservation Methods for Synthetic DNA Data Storage. Small Methods, 2021. 5(5): p. 2001094.
[2] Bonnet et al., Chain and conformation stability of solid-state DNA: implications for room temperature storage. Nucleic Acids Res, 2010. 38(5): p. 1531-46.
[3] Grass et al., Robust Chemical Preservation of Digital Information on DNA in Silica with Error-Correcting Codes. Angew Chem Int Ed Engl, 2015. 54(8): p. 2552-2555.
[4] Coudy et al., Long term conservation of DNA at ambient temperature. Implications for DNA data storage. PLoS One, 2021. 16(11).
[5] Chen et al., Combining Data Longevity with High Storage Capacity—Layer-by-Layer DNA Encapsulated in Magnetic Nanoparticles. Advanced Functional Materials, 2019. 29(28): p. 1901672.

INTEGRATING DNA-BASED MEMORY IN WATER-RESISTANT ELECTROSPUN POLYMER FIBERS FOR NON-DESTRUCTIVE DATA RETRIEVAL

Wetzl C.[1], Soukarie D.[1], Elosua J. Y.[1,2], Bittner A.[1,3], Santiago I.[1]
[1]*CIC nanoGUNE BRTA, Donostia-San Sebastián, 20018, Spain*
[2]*University of the Basque Country UPV/EHU, Donostia-San Sebastián, 20018, Spain*
[3]*IKERBASQUE Basque Foundation for Science, 48009 Bilbao, Spain*
email to: c.wetzl@nanogune.eu

DNA is a digital memory storage medium with advantageous properties, including high information density, centuries-long and energy-efficient retention of data.[1] Thanks to the advances in data-to-DNA encoding schemes together with the development of next-generation synthesis and sequencing methods, DNA-based solutions demonstrated to be a reliable option for the future of massive digital data storage. In this context, the employment of DNA dry preservation systems has emerged as a viable method, since it does not require refrigeration, which enhances its cost-effectiveness and reduces the environmental impact.[2]

To ensure the stability of dry DNA for long-term archival storage, embedding it in a robust matrix is essential. Here, we present a versatile way to store digital information in synthetic DNA embedded in electrospun polymer fibers. We developed an end-to-end workflow from data-to-DNA message encoding to DNA-to-data decoding passing through preservation into polymeric fibers and DNA controlled release (**Figure 1**). We built a library of electrospinnable polymers compatible with DNA encapsulation, with improved stability to environmental conditions and a facile retrieval strategy achieved by using hydrophobic and hydrophilic cross-linked polymers. We demonstrated the on-demand retrieval of short and long messages encoded in a single oligonucleotide and a pool of oligonucleotides, respectively. Its water resistance makes our system a true non-destructive readout memory: repetitive access to messages stored into fibers is achieved without damaging the integrity of the fibers or DNA. We envisage our simple and robust manufacturing approach will contribute to the development of scalable and accessible DNA data storage solutions.

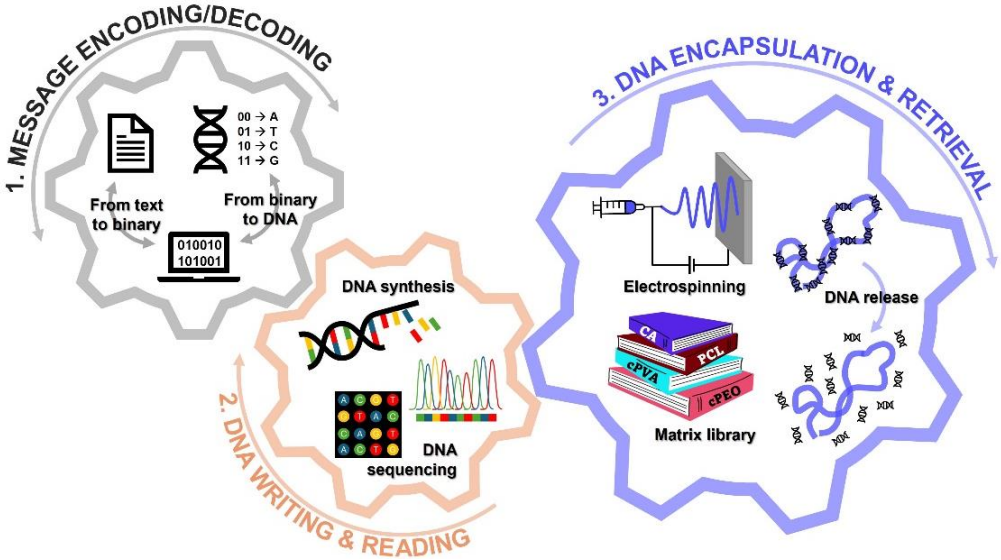


Figure 1. Schematic overview of our experimental end-to-end workflow

References:

[1] Doricchi, A. et al., *ACS Nano,* 16 (11), 17552–17571, 2022
[2] Matange, K. et al., *Nat Commun*, 12 (1), 1358, 2021

DROPSCRIBE: USING DROPLET MICROFLUIDICS TO STORE DIGITAL DATA

Spratt J.[1], Lang S.[1], Fördős F.[2], Thobaben R.[3], Jönsson H.[4], Hoffecker I. T.[1]

[1] Dept. of Gene Technology, KTH Royal Institute of Technology, Science for Life Laboratory, Solna, Sweden
[2] Dept. of Medical Biochemistry and Biophysics, Karolinska Institutet, Stockholm, Sweden
[3] Dept. of Intelligent Systems, KTH Royal Institute of Technology, Stockholm, Sweden
[4] Dept. of Protein Science, KTH Royal Institute of Technology, Science for Life Laboratory, Solna, Sweden
joel.spratt@scilifelab.se

THERMODYNAMICALLY-FAVOURED DNA COMPUTING AND STORAGE

Stérin T., Eshra A., Adio J., Evans C., Woods D.

Hamilton Institute, Department of Computer Science, Maynooth University, Co. Kildare, Ireland.
damien.woods@mu.ie

Imagine if our molecular-scale DNA data storage devices came with their own molecular computer to read, write and compute on data stored in DNA!

We present a new form of DNA computing amenable to dynamic DNA data storage, called the Scaffolded DNA Computer (SDC) [1]. The SDC is grounded in mathematical, physical and computer science principles so that the output is thermodynamically-favoured, meaning that it is encoded in the ground state(s) of the system and any errors are unfavourable. We demonstrate ten DNA programs including MULTIPLICATION-by-3, DIVISION-by-2, 8-bit PARITY-detection, and ADDITION of up to 25-bit numbers. SDC algorithms have simple experimental protocols, can be reused dozens of times and small instances run in under a minute. Unlike previous forms of DNA computing, no error-correction nor precise step-by-step kinetic control are required, and its thermodynamic and computational properties mean it is both scalable and programmable.

References:

[1] T. Stérin*, Eshra*, A., J. Adio, C.G. Evans, and D. Woods, A Thermodynamically Favoured Molecular Computer: Robust, Fast, Renewable, Scalable. Under review. *Joint lead authors.

COMPOSITE ALPHABETS FOR DNA BASED DATA APPLICATIONS

Yakhini, Z.

[1]*Faculty of Computer Science, Technion, Haifa, Israel*
[2]*School of Computer Science, Reichman University, Hertzliya, Israel*

SHAPE MATTERS

Na Liu L.

Second Physics Institute, University of Stuttgart, Germany
Max Planck Institute for Solid State Research, Stuttgart, Germany
na.liu@pi2.uni-stuttgart.de

The shape of biological matter is central to cell function at different length scales and determines how cellular components recognize, interact, and respond to one another. However, their shapes are often transient and hard to reprogram. In this work, we construct a synthetic cell model comprised of signal-responsive DNA nanorafts, biogenic pores, and giant unilamellar vesicles (GUVs). We demonstrate that reshaping of DNA rafts at the nanoscale can be coupled to reshaping of GUVs at the microscale. The nanorafts collectively undergo reversible transitions between isotropic and short-range local order on the lipid membrane, programmably remodeling the GUV shape. Assisted by the biogenic pores, during GUV shape recovery the locally-ordered DNA rafts perforate the membrane, forming sealable synthetic channels for large cargo transport. Our work describes a versatile platform for interfacing reconfigurable DNA nanostructures with synthetic cells, expanding the potential of DNA nanotechnology in synthetic biology.

SCALABLE AND ROBUST DNA-BASED STORAGE VIA CODING THEORY AND DEEP LEARNING

Bar-Lev D.[1], Orr I.[2,3], Sabary O.[2], Etzion T.[2], Yaakobi E.[2]

[1]*Center of Memory and Recording Research, University of California, San Diego, La Jolla, CA 92093, USA*

[2]*Computer Science Faculty, Technion - Israel Institute of Technology, Haifa, Israel, 3200003*

[3]*UVeye Ltd., Tel Aviv, Israel*

omersabary@cs.technion.ac.il

The global data sphere is expanding exponentially, projected to reach 400 zettabytes by 2028, while current storage technologies cannot scale at the required pace [1]. DNA-based storage has emerged as a promising solution to address this gap, allowing digital information to be archived in DNA molecules. This method offers significant advantages over existing storage solutions, including exceptional information density, enhanced durability, and negligible power consumption for data integrity maintenance [2]. However, its practical implementation faces challenges, such as synthesis and sequencing errors, computational complexity, and the tradeoff between scalability and accuracy in information retrieval pipelines.

To overcome these challenges, we propose DNAformer [3], a modular and holistic pipeline integrating Deep Neural Networks (DNNs) trained on simulated data, Tensor-Product [4] (TP)-based Error-Correcting Codes (ECC), and a safety margin mechanism. This unified approach addresses the inherent redundancy and errors in DNA synthesis and sequencing, ensuring efficient and accurate information retrieval. We validated our method on 3.1MB of data using two sequencing technologies, demonstrating a 3200× speed improvement, a 40% increase in accuracy, and a code rate of 1.6 bits per base under high-noise conditions.

The DNA storage pipeline involves encoding binary data into DNA sequences, which are synthesized as strands (oligos), each with multiple copies that are stored unordered in a storage container. During reading, noisy reads of the strands are obtained using sequencing. To retrieve the data from the reads, our solution utilizes three main steps. 1) Clustering step in which the noisy reads are partitioned according to their corresponding strand. The clustering is performed based on predesigned indices, which allow an efficient process. 2) Reconstruction step in which an estimation of the original sequence of each cluster is predicted using a DNN that combines convolution and transformer layers, which also outputs confidence scores. An additional algorithm, Conditional Probability Logic, reprocesses clusters with low confidence, boosting reconstruction accuracy. 3) Decoding step that applies a TP-based ECC to correct any remaining errors to restore the data completely.

By addressing critical computational bottlenecks and ensuring scalability, DNAformer overcomes the limitations of current retrieval processes and paves the way for the future of data storage.

References:

[1] Taylor, P. Amount of data created, consumed, and stored 2010–2023, with forecasts to 2028. *Statista* (2024); available at https://www.statista.com/statistics/871513/worldwide-data-created/.

[2] Meiser, L. et al. Synthetic DNA applications in information technology. *Nat. Commun.* **13**, Article 1234 (2022).

[3] Bar-Lev, D. et al. Deep DNA storage: Scalable and robust DNA storage via coding theory and deep learning. *arXiv preprint arXiv:2109.00031* (2024).

[4] Wolf, J. On codes derivable from the tensor product of check matrices. *IEEE Trans. Inf. Theory* **11**, 281–284 (1965).

A BIRD'S-EYE VIEW ON DNA STORAGE SIMULATORS AND COMPUTING TOOLS

Gupta M. K.

Kaushalya the Skill University, India
mankg@guptalab.org

DNA computing is a fascinating field that emerged from Adleman's seminal work in 1994 [1]. It has since evolved into several branches, including DNA self-assembly, DNA nanostructures, chemical computing, DNA origami, and DNA data storage, among others. In this talk, we will highlight key tools used in the field and review various software platforms that simulate DNA storage across different domains [2].

References:

[1] Adleman, L. M. (1994). Molecular computation of solutions to combinatorial problems. *Science*, *266*(5187), 1021–1024. https://doi.org/10.1126/science.7973651

[2] Sanket Doshi, Mihir Gohel and Manish K. Gupta, A Bird-Eye View on DNA Storage Simulators, 2024, arXiv, 2404.04877, cs.IT, https://arxiv.org/abs/2404.04877

STABILITY OF NUCLEIC ACID ORIGAMI

Linko V.

Institute of Technology, University of Tartu, Nooruse 1, 50411 Tartu, Estonia
Department of Bioproducts and Biosystems, Aalto University School of Chemical Engineering, Kemistintie 1, 02150 Espoo, Finland
veikko.pentti.linko@ut.ee

The structural integrity of nucleic acid nanostructures might get compromised under various application-specific conditions [1], such as in organic solvents [2]. This is also a major challenge in physiological media, where the sufficient pharmacokinetic availability of the structures needs to be ensured [3]. Here we focus on the stability of DNA origami nanostructures (DONs) in the biologically relevant context, *i.e*., under low-magnesium conditions [4] and in nuclease-rich environments [5]. Furthermore, we have shown that these features can be tuned by internal design [6], and that in general, the DON stability is superstructure-dependent [4,5,7].

The stability of DONs can also be improved by intercalators and versatile, protective molecular coatings. We have combined DONs with doxorubicin [7-9], phthalocyanines [10], and lipids [11,12], as well as with "camouflaging" albumins and antigen-targeting antibody fragments [13] to modulate the stability and enhance the functionality of DONs. Recently, we have coated DONs and hybrid mRNA-DNA origami structures with virus capsid proteins for directing capsid polymorphism, improving the stability and delivery of nucleic acids, and assembling nanoreactors through compartmentalization of enzymes [14-16].

References:

[1] Ramakrishnan, S. et al., *Computational and Structural Biotechnology Journal 16*, 342–349, 2018
[2] Enlund, E. et al., *Nanoscale 16*, 13407–13415, 2024
[3] Linko, V. & Keller, A., *Small 19*, 2301935, 2023
[4] Kielar, C. et al., *Angewandte Chemie International Edition 57,* 9470–9474, 2018
[5] Ramakrishnan, S. et al., *ChemBioChem 20*, 2818–2823, 2019
[6] Xin, Y. et al., *Small 18*, 2107393, 2022
[7] Ijäs, H. et al., *Nucleic Acids Research 49*, 3048–3062, 2021
[8] Lisitsyna, E.S. et al., *bioRxiv*, 10.1101/2024.06.20.599777, 2024
[9] Klose, A. et al., *Nanoscale 16*, 17585–17598, 2024
[10] Shaukat, A. et al., *Chemical Communications 56*, 7341–7344, 2020
[11] Julin, S. et al., *Angewandte Chemie International Edition 60*, 827–833, 2021
[12] Julin, S. et al., *Chemical Communications 59*, 14701–14704, 2023
[13] Seitz, I. et al., *ACS Applied Materials & Interfaces 14*, 38515–38524, 2022
[14] Seitz, I. et al., *Nature Nanotechnology 18*, 1205–1212, 2023
[15] Seitz, I. et al., *Advanced Materials 37*, 2417632, 2025
[16] Seitz, I. et al., *submitted* 2025

DNA-ASSEMBLY FOR PHOTONICS, PLASMONICS AND BIOSENSING

Liedl T.

Fakultät für Physik and Center for NanoScience, Ludwig-Maximilians-UniversitätMünchen;
tim.liedl@physik.lmu.de

Over the last decades, DNA self-assembly in general and DNA origami in particular have matured at a breathtaking pace and DNA architectures are today routinely usedfor the constructing of functional two- and three-dimensional nanomachines andmaterials [1,2]. Our group has contributed complex and nanometer-precise assemblies of biomolecules, organic fluorophores and inorganic nanoparticles [3].

We now demonstrated diamond-type DNA origami lattices exhibiting structural color. For this we first grew single crystals with overall dimensions of 10 - 20 micrometers. This step was followed by sol-gel type silicification [4] and subsequent atomic layer deposition (ALD) of several nanometers of TiO_2 to ensure a high enough refractive index contrast to open a photonic band gap, which could then be observed in the reflection spectrum of the crystals [5]. We are currently exploring further properties of this unique type of material.

Another interest of our group is to employ gold and silver nanoparticles for plasmonic sensing by making use of circular dichroism, absorbance and scattering properties of such versatile and nanoscopic reporters [5].

In this talk I will report on our ongoing efforts to build functional plasmonic devices on the one hand and materials that are designed on the molecular level while reaching macroscopic dimensions on the other.

References:

[1] P. W. K. Rothemund, Nature 440, 297–302 (2006)
[2] N. C. Seeman, Annu. Rev. Biochem. 79, 12.1 (2010)
[3] R. Schreiber et al. Nature Nanotechnology 9, 74-78 (2014)
[4] L. Nguyen et al. Angew. Chem. Int. Ed., 58, 912-916 (2019)
[5] G. Posnjak et al. Science 384, 781-785 (2024)
[6] F. Schuknecht et al. Nature Communications 14, 7192 (2023)

NON-LINEAR MANIPULATION OF DNA CODES

Zhang Y.

B CUBE, TU Dresden, Germany
yixin.zhang1@tu-dresden.de

DNA-encoded chemical library (DEL) has become one of the leading techniques in drug screening [1, 2]. While tens to hundreds of trillions of compounds have been synthesized in the past decade, they present not only one of the largest collections of chemical structures, but also the largest collection of man-made DNA codes. The DEL technology assumes a linear correlation through the synthesis and screening processes. To investigate and explore the non-linearity can help us to develop new method in DEL screening, and in fields beyond that, e.g. in the field of diversity analysis, information encryption, and data procession [3,4,5].

References:

[1] Mannocci, L. et al., *PNAS* 105, 17670 – 5, 2008

[2] Reddavide, F. et al., *Angew. Chem. Int. Ed.*, 54, 7924 –7928, 2015

[3] Andrade, H. et al., *ChemBioChem* 21, 1144 –9, 2020

[4] Cui, M. et al., *ChemBioChem* 21, 2503 –11, 2020

[5] Cui, M. et al., *Nucleic Acids Research*, 50, 15, 8974–8985, 2022

SCALABLE SEARCH OF MASSIVELY POOLED NUCLEIC ACID SAMPLES: FROM COLD DATA ARCHIVES TO GLOBAL GENOMICS

Bathe M.

Department of Biological Engineering, MIT, 77 Mass. Ave., Cambridge MA 02139 USA
mark.bathe@mit.edu

Biotechnology is powering a global industrial revolution largely enabled by innovation waves in nucleic acid synthesis, sequencing, and editing. Aside from its classical application to therapeutics and clinical diagnostics, alternative applications of synthetic DNA include generic information storage of digital datasets [1], as metadata labels for storing and retrieving biomolecular samples [2, 3], and as programmable materials interfaced with silicon semiconductors [4]. In digital data storage, DNA offers long-term storage spanning centuries to millennia at ultra-high densities up to zettabytes per cubic centimeter, with zero energy consumption for ongoing data maintenance [1]. Random file access can be performed efficiently from such pooled databases using classical Boolean operations, with DNA sequencing to read out file contents [1]. This application is ideally suited to cold data such as archived media or compliance records, requiring infrequent and limited file access. Alternatively, such a file storage and retrieval system can be used for native biological samples such as clinical genomic DNA and RNA from humans, pathogens, and other organisms, in applications to global biobanking, genomics, ecological conservation, and pathogen surveillance to detect or thwart epidemics or pandemics [2, 3]. Finally, interfacing these DNA datasets with silicon chips offers alternative read-outs from sequencing, empowering next-generation formats for nucleic acid storage and memory [4]. These and emerging innovations in DNA data will help empower advances in 21st century medicine, computing, sensing, and materials, as discussed in this presentation.

References:

[1] Banal, J.L. et al., *Nature Materials 20*, 1272–1280, 2021

[2] Banal, J. *ACS Applied Materials & Interfaces 13*, 49729–49736, 2021

[3] Berleant, J.D., et al., *medRxiv 2024.04.12.24305660*; doi: https://doi.org/10.1101/2024.04.12.24305660

[4] Luo, X., et al., *bioRxiv 2025.01.23.634416;* doi: https://doi.org/10.1101/2025.01.23.634416

BRIDGING DNA STORAGE AND COMPUTATION

Soloveichik D.

University of Texas at Austin, USA
david.soloveichik@utexas.edu

DNA is an incredibly dense storage medium for digital data. However, computing on the stored information is expensive and slow, requiring an inefficient change of domain from the chemical to the electronic and back. The field of DNA computation, pioneered over three decades ago and since significantly evolved into the discipline of molecular programming, provides a framework for performing computational operations directly within the DNA domain itself. I will overview ways that digital logic and analog dynamics can be programmed into DNA chemical systems using systematic bottom-up design, and I will explore the critical interplay between computational parallelism and autonomy in DNA-based systems. I will conclude with insights on DNA's natural affinity for "similarity match" operations based on hybridization free energy. By understanding what DNA naturally excels at computationally, we can design systems that leverage these inherent strengths rather than forcing incompatible paradigms. While DNA storage and computation have largely developed as separate domains, there are compelling potential convergence points where computational principles could enhance storage systems.

FAST RDF-ANALYTICS FROM LARGE-SCALE DNA-BASED ARCHIVES

Usmani A., Wiese L.

Department of Computer Science, Goethe University
Frankfurt,60325, Robert-Mayer-Str. 10, Frankfurt am Main
usmani@mathematik.uni-frankfurt.de

DNA, or Deoxyribo Nucleic Acid, storage has emerged for inexpensive long-term massive data archives, leading to future data analytics using knowledge graphs. Objectively, partial data retrieval over unwanted whole data from a DNA storage system avoids unnecessary sequencing costs and latency. Hence, suitable DNA-based graph data storage is highly desirable for effective query processing retrieval. Regarding this, an early effort was made to obtain as few DNA strands as possible while executing SPARQL queries over RDF graph data storage in DNA. Despite partial data retrieval being studied earlier, the measurement of the corresponding sequencing runs —a key feature specifying the query latency —has still not been taken into account. In this paper, we address both of the issues making query processing more effective over RDF graph data. The idea is to minimize as many sequencing runs as possible using a block-based technique. There, each block of strands is associated with a standalone primer address instead of one-to-one individual strand mapping. Favorably, we achieved a reduction in sequencing runs of at least seven times over the previous work. As a result, query execution costs and time are further optimized for DNA-based RDF graphs.

Query-oriented partial data retrieval could minimize sequencing costs and time by reducing sequence runs instead of whole data retrieval in large data archives. Partial data retrieval can be accomplished using RDF graph data storage and SPARQL query processing. The study of partial data retrieval versus sequencing runs is therefore of the utmost importance when minimizing sequencing costs when only a portion of the data is needed using query processing over RDF graphs. Previously, we proposed a DNA storage model [1] that allows us to query an RDF graph over 200MB while extracting less than 1% of its mapping data. However, it cannot scale due to the need for a large number of primers, besides the slow speed of obtaining results in a DNA-based system. This redirects us to find a more practical solution to these problems. Thus, in our newly proposed architecture, we intend to fetch a complete block composed of many DNA fragments instead of a single DNA strand. Multiple related DNA strands are connected to the same primer address, creating one block or cluster of DNA strands. Since there are several DNA strands in each block, thedesired output has a better chance of being present completely in the same block and having enough information to fetch the appropriate next block if it is not found. The retrieval of each block requiresa single sequencing run; therefore, the results of a query should be based on a minimum of blocks. The approach increases the data retrieval rate, however, based on how many data strands are incorporated within a single block.

References:

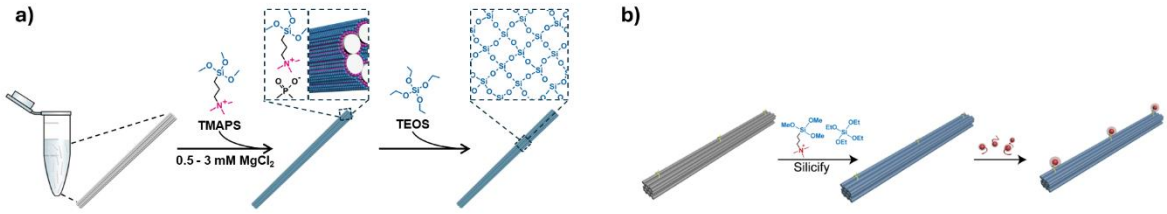
[1] Usmani, Asad, and Lena Wiese. "DNA-based storage of RDF graph data: a futuristic approach to data analytics." *IEEE access* 11 (2023): 129931-129944.

SILICIFICATION FOR ROBUST AND REVERSIBLE PROTECTION OF DNA NANOSTRUCTURES

Heuer-Jungemann A.[1,2], Baptist, A.[1,2]

[1]*Max Planck Institute of Biochemistry, Am Klopferspitz 18, 82152 Martinsried and Center for Nanoscience, LMU Munich, Germany*
[2]*TU Dortmund University, Otto-Hahn-Str. 6, 44227 Dortmund, Germany*
Amelie.heuer-jungemann@tu-dortmund.de

DNA origami has found a plethora of potential applications ranging from biomedicine to biocatalysis and data storage. However, one big bottleneck for most real-world applications is DNA origami's inherent instability to many environmental factors such as heat, low salt conditions, extreme pH or the presence of nuclease enzymes. To increase the stability of structures, different methods have been proposed, including e.g. internal cross-linking[1], coating with polymers[2, 3] or biominerals like calcium carbonate, calcium phosphate[4] or silica[5-7]. Of these, silicification (see **Figure 1a**) has emerged as an excellent tool to create highly stable DNA origami nanostructures with retained shape/size, as we all site-specific addressability (**Figure 1b**) [8]. I will highlight recent insights gained into the silicification reaction and show how the use of modified pre-cursors can result in additional properties of the silica coating, making the process even more versatile for a broad range of applications.



References:

[1] Gerling T. et al., *Science Advances*, eaau1157, 2018.
[2] Ponnuswamy N. et al., *Nature Communications*, 15654, 2017.
[3] Agarwal N. P. et al., *Angewandte Chemie*, 5552–5556, 2017.
[4] Wu S. et al., *ACS Nano*, 1555–1565, 2021.
[5] Jing X. et al., *Nature Protocols*, 2416–2436, 2019.
[6] Nguyen L. et al., *Angewandte Chemie International Edition*, 912–916, 2019.
[7] Nguyen M.-K. et al., *Chemistry of Materials*, 6657–6665, 2020.
[8] Wassermann L. M. et al., *Advanced Materials*, 2212024, 2023.

XENO NUCLEIC ACIDS FOR ROBUST LONG-STRANDED DATA STORAGE

Song J.

Hangzhou Institute of Medicine, Chinese Academy of Sciences, Hangzhou 310022, China
songjie@him.cas.cn

Over the past decade, the exponential growth of global data volumes has driven significant advancements in information storage and processing technologies. DNA, with its remarkable information density, programmability, and energy efficiency, has emerged as a promising candidate for next-generation storage material. However, its application in complex environmental conditions remains limited due to challenges such as susceptibility to cell lysis, extreme pH and nucleophilic solutions. To address these limitations, the incorporation of non-canonical nucleic acids, which offer enhanced biochemical stability without compromising storage density, presents a promising avenue for the development of robust storage technologies.

Non-canonical nucleic acids represent a diverse class of molecules that have been naturally or chemically modified in their sugar moiety, phosphate backbone, nucleobases, or mirror-image isomerism. In this talk, we will summarize the fundamental properties of non-canonical nucleic acids, including chemical and biological stability, hybridization affinity, and thermal stability. Furthermore, we introduce the general workflow for storing information in nucleic acid molecules and discuss recent advances in the synthesis and sequencing of non-canonical nucleic acids. We also highlight the key challenges in their application as storage materials and outline potential research directions to achieve more stable, efficient, and accessible storage solutions.

MOTIF-BASED APPROACHES FOR SCALING READ/WRITE COST IN DNA STORAGE

Appuswamy R.

Data Science Department,
EURECOM,
450 Route des Chappes, 06410 Biot, France
raja.appuswamy@eurecom.fr

The surge in demand for cost-effective long-term archival media, coupled with density limitations of contemporary magnetic media, has resulted in synthetic DNA emerging as a promising new alternative. Despite its benefits, storing data in DNA poses several challenges as the technology used for writing data on DNA is very expensive, and technology used for reading data is error prone. Thus, it is important to design pipelines that can efficiently use redundancy to mask errors without amplifying read/write cost. In this talk, we will present the benefits of using motifs, which are short sequences of nucleotides, as fundamental building blocks for encoding/decoding digital data to/from DNA instead of the traditional approach of using nucleotides.

We first present CMOSS [1], a reliable, motif-based Columnar Molecular Storage System that focuses on reading the read cost of DNA data storage. CMOSS differs from state-of-the-art approaches on three fronts. First, it uses a motif-based, columnar layout in contrast to nucleotide-based row layout for designing oligos. Second, thanks to use of motifs, it performs integrated consensus calling and decoding to effectively handle reliability bias in DNA storage. Third, it provides a flexible, block-based data organization for random access over DNA storage in contrast to object-based organization used by state-of-the-art. Using results from large-scale wet-lab experiments, we demonstrate the benefit of CMOSS' motif-based design in reducing read cost.

Next, we focus on reducing the benefit of using motifs for reducing the write cost of DNA storage. In particular, we will present the *composite motifs* [2]—a framework that uses a mixture of prefabricated motifs sequences assembled as building blocks of synthesis to reduce write cost by scaling logical density. We will provide an overview of synthesis and sequencing techniques, consensus calling methods, and encoding/decoding algorithms customized to the composite motifs framework. Using these tools, we will present the results from wet-lab experiments that show how the use of motifs as building blocks can provide an order of magnitude reduction in write cost as well.

References:

[1] Marinelli, E., Yan, Y., Tattini, L., Magnone, V., Barbry, P., and Appuswamy, R.. CMOSS: A Reliable, Motif-based Columnar Molecular Storage System. In Proceedings of the 17th ACM International Systems and Storage Conference (2024), https://doi.org/10.1145/3688351.3689162

[2] Yan, Y., Pinnamaneni, N., Chalapati, S. et al. Scaling logical density of DNA storage with enzymatically-ligated composite motifs. Sci Rep 13, 15978 (2023). https://doi.org/10.1038/s41598-023-43172-0

THERMODYNAMIC AND KINETIC CHALLENGES OF DNA-BASED DATA STORAGE AND COMPUTATION

Keung A. J.

Department of Chemical & Biomolecular Engineering, North Carolina State University, Raleigh, NC, USA
ajkeung@ncsu.edu

Molecular information storage in DNA presents several theoretical advantages if they can be practically realized. DNA can theoretically store orders of magnitude more data per volume and compute for orders of magnitude less energy than current state of the art systems. Here we will present our analysis of how these advantages also present fundamental challenges unique to molecular information systems, and how they must be solved to become practical technologies. We then describe our work addressing these challenges through molecular biology, biochemistry, simulations, and materials science.

In particular, we describe how thermodynamics presents key challenges of undesired non-specific interactions becoming dominant in highly dense systems. We show how we solve this challenge through hierarchical file address systems and through leveraging non-specific interactions to program useful functions like Previewing a file. We continue by describing an end-to-end DNA computer that exploits a unique high surface area material for immobilizing DNA yet maintaining its accessibility while also improving its stability. We demonstrate a 10^4 TB cm^{-3} storage density, half-lives over 2 million years at -18C, and we compute simple Sudoku and chess problems in a reuseable manner. We end by describing three unpublished studies. First we describe a system for enzymatic synthesis of modified DNAs. Second, we describe a new fundamental phenomena of DNA molecules in solution that we exploit for controllably obfuscating information. And we end by describing a new machine learned model of DNA-DNA interactions that outperforms state of the art models by nearly 30% in accuracy.

TOWARD EFFICIENT COMPRESSION FOR DNA-BASED DATA STORAGE

Antonini M.

I3S laboratory, University Côte d'Azur and CNRS, France
Marc.antonini@cnrs.fr

The rapid and exponential growth of digital data - 90% of which has been generated in the last two years - poses a significant challenge for long-term storage due to limited resources, energy consumption, and the short lifespan of conventional storage media. Moreover, about 70% of this data is "cold," rarely accessed and needing preservation for 10 years or more, further highlighting the need for durable and scalable storage solutions.

Recent advances identify DNA as a highly promising medium for next-generation data storage, offering an extraordinary theoretical capacity of up to 215 petabytes per gram and the potential for data stability over centuries through synthetic DNA encapsulated in specialized microcapsules. Retrieval is enabled by advanced sequencing technologies.

This presentation will review the current state of the art in DNA-based data storage, with a particular focus on the efficient compression and encoding of digital data into the quaternary code of DNA's four nucleotides -Adenine (A), Thymine (T), Cytosine (C), and Guanine (G). We will introduce JPEG DNA, an emerging standard specifically designed for image compression and coding tailored to DNA's unique biochemical constraints.

Furthermore, inspired by classical image compression principles, our work at the I3S laboratory introduces a novel approach that jointly designs compression and DNA encoding within a closed-loop framework. Unlike traditional pipelines that apply standard image compression followed by DNA encoding - resulting in suboptimal, open-loop workflows - our method integrates source and channel coding in a way that directly accounts for the biochemical constraints of DNA synthesis and sequencing. This joint optimization improves compression efficiency without compromising image quality.

Our work specifically addresses the unique challenges of DNA-based data storage by co-optimizing encoding strategies and molecular constraints. This enables more practical, cost-effective, and scalable solutions for digital image archiving in synthetic DNA.

CODING FOR THE DNA DECAY PROBLEM ALONG WITH EDIT ERRORS

Abu-Sini M., Heckel R.

Department of Computer Engineering, Technical University of Munich, DE-80333, Munich, Germany
maria.abu-sini@tum.de

According to Gimpel et al. [2], the DNA decay problem remains among the main unresolved challenges in the DNA storage systems. In these systems, a set of information sequences is stored by synthesizing for each sequence (multiple) noisy DNA strands that suffer from insertions, deletions, and substitutions. Storing these strands for a long period of time results in their decay, i.e., their breakage to shorter fragments. Thus, to enable storing data for long periods of time in a DNA storage system, one needs a scheme that reconstructs a set of sequences from their noisy fragments. Such a scheme was proposed by Song et al. [4], but without rigorous analysis and guarantees on the number of breakages and edit errors it can tackle.

In our study, we aim to provide such a rigorous scheme. Thus, as a first step we focus on a very simple setting, in which a single long information sequence is stored using a single DNA strand, and should be later recovered from a set of fragments received from at most t breakings and t_e edit errors. For this setting we propose a reconstruction scheme by combining tools from [1] and [5]. Specifically, to ensure successfull reconstruction of a length-m information sequence, our scheme maps the sequence to an encoded one of length $m + \Theta\left((t + t_e)\log m \log\frac{m}{t+t_e}\right)$ and recovers it with run-time complexity $O(t!\, m^4)$, which can be further optimized to get a polynomial (of n) run-time complextity if it is known that all the errors in the strand are substitutions solely.

Our main contribution is suggesting a technique to correct both breakages and errors, though for a very simple setting solely. We recall here that rigorous schemes were proposed in [3], [5], and [6], yet similarly just for simple settings. As a future extension of this work, one may try to apply it to the real decay setting or incorporate it with the more applicable scheme proposed in [6].

References:

[1] K. Cheng et al., "Deterministic document exchange protocols and almost optimal binary codes for edit errors," *Journal of the ACM*, vol. 69, no. 6, pp. 1--39, 2022.

[2] A. L. Gimpel et al., "Challenges for error-correction coding in DNA data storage: Photolithographic synthesis and DNA decay," *Digital Discovery*, vol. 3, no. 12, pp. 2497--2508, 2024.

[3] S. Nassirpour et al., "DNA merge-sort: A family of nested Varshamov-Tenengolts reassembly codes for out-of-order media," *IEEE Transactions on Communications*, vol. 72, no. 3, pp. 1303--1317, Mar. 2023.

[4] L. Song et al., "Robust data storage in DNA by de Bruijn graph-based de novo strand assembly," *Nature Communications*, vol. 13, no. 1, pp. 5361, 2022.

[5] C. Wang et al., "Break-resilient codes for forensic 3D fingerprinting," *IEEE International Symposium on Information Theory (ISIT)*, pp. 3148--3153, Athens, Greece, Jul. 2024.

[6] Y. Yehezkeally et al., "Generalized unique reconstruction from substrings," *IEEE Transactions on Information Theory*, vol. 59, no. 9, pp. 5648--5659, 2023.

PROGRAMMABLE AND RENEWABLE THERMODYNAMICALLY FAVOURED DNACOMPUTING AND STORAGE

Eshra A.[1], Stérin T.[2], Woods D.[1]

[1]*Hamilton Institute, Department of Computer Science, Maynooth University, Co. Kildare, Ireland.*
[2] *prgm.dev, 9 rue des colonnes, Paris, France*
abeer.eshra@mu.ie

Most DNA computing systems are single-use. After a computation is completed, the DNA strands in the test tube are discarded, requiring new material for each run. This increases cost, waste, and experimental effort. In molecular computing systems, kinetic proofreading and other error-correcting mechanisms prevent error states, but at an energy cost or complexity scale-up. Unlike thesesystems, thermodynamically favoured computers reach correct outputs naturally by moving to equilibrium. However, implementing this principle in molecular computing has remained a challengefor decades [1].

Our thermodynamically favoured Scaffolded DNA Computer (SDC) executes computations using a simple thermal anneal [2]. We successfully demonstrated computations like arithmetic and cellular automata in under a minute encoding data on a 12-bit or 24-bit memory, and on 75 bits over a few hours, all without needing explicit error correction. Unlike traditional DNA computing approaches that require new material for each run, the SDC is *intrinsically renewable*. Instead of beingdiscarded, the same mix can be reset and reused multiple times with just two DNA strands, one to reset the program and another to introduce the new input. We validated the renewal process on multiple programs, showing steady performance over repeated cycles.

We contend that these principles also apply to DNA storage platforms. Many existing DNA Storage platforms rely on destructive sequencing [3]. However, by leveraging thermodynamically favoured states, DNA storage can become computable, rewritable and sustainable. Our approach integrates renewability into molecular computation, reducing material waste, minimizing energy consumption, and paving the way for long-term, low-energy alternatives to traditional computing andstorage technologies.

References:

[1] Bennett, C.H., *Physical Review Letters 53* (12). 1202-1202, 1984
[2]Eshra*, A., T. Stérin*, J. Adio, and D. Woods, *A Thermodynamically Favoured MolecularComputer: Robust, Fast, Renewable*, Under review.
[3] Doricchi, A. et al., *ACS Nano 16* (11). 17552-17571, 2022

DNABYTE: A MODULAR FRAMEWORK FOR THE SIMULATION AND ANALYSIS OF DNA-BASED DATA STORAGE

Wernhart K. [1], Schroeder F. [1], Barišić I.[1,2]

[1]Eko Refugium, Crno Vrelo 2, 47240 Slunj, Croatia
[2]Austrian Institute of Technology, Giefinggasse 4, 1210 Vienna, Austriakaya.wernhart@ekorefugium.com

DNA-based data storage has emerged as the most promising solution for the cold storage of digital files. Various DNA codes have been proposed for encoding data into DNA strands [1]. However, the design of a DNA data storage pipeline is a complex process, which involves multiple interdependent stages, ranging from encoding algorithms to DNA synthesis, storage, and sequencing and finally error correction and decoding. In-silico simulation of the end-to-end process can offer detailed analytics and reduce the need for costly and time-consuming experiments [2].

DNAByte (dnabyte.ekorefugium.com) is a modular simulation and analysis framework that covers the entire DNA data storage process. It supports encoding and decoding of files as well as simulating errors that occur in the synthesis, storage, and sequencing of DNA. The modular architecture allows the user to run simulations with any combination of encoding, decoding, and error correction strategies. Furthermore, it is easily extendable as every module can be substituted by custom-made functions, which can be incorporated into the framework. This flexibility is essential as new DNA codes continue to be developed. The framework currently supports several module categories, including (i) inner error correction (e.g., Reed-Solomon), (ii) outer error correction (e.g., LT codes), (iii) encoding/decoding algorithms, and (iv) error channels (e.g., sequencing errors, storage conditions, etc.) [3]. To analyze the possibility of storing large amounts of data, DNAbyte can also create multiple split DNA strands containing fragments of the data that can be stored together as well as it including a ligation simulation.

As a web-based application, DNAByte offers an easily accessible and comprehensive suite of analysis tools for tracking and quantifying errors and error correction performance. It provides detailed statistics on encoding performance, error correction efficiency, and storage conditions and thereby accelerates the development of robust and scalable DNA data storage pipelines.

References:

[1] Yan, Y., Pinnamaneni, N., Chalapati, S. *et al.* Scaling logical density of DNA storage with enzymatically-ligated composite motifs. *Sci Rep* **13**, 15978 (2023). https://doi.org/10.1038/s41598-023-43172-0

[2] Yuan, L., Xie, Z., Wang, Y. *et al.* DeSP: a systematic DNA storage error simulation pipeline. *BMC Bioinformatics* **23**, 185 (2022). https://doi.org/10.1186/s12859-022-04723-w

[3] Schwarz, M., Welzel, M., Kabdullayeva, T., Becker, A., Freisleben, B. and Heider, D., 2020.MESA: automated assessment of synthetic DNA fragments and simulation of DNA synthesis, storage, sequencing and PCR errors. Bioinformatics, 36(11), pp.3322-3326.

ON THE CAPACITY OF DNA LABELING

Hanania D.[1], Bar-Lev D.[1], Nogin Y.[2], Shechtman Y.[2,3,4], Yaakobi E.[1]

[1]*Department of Computer Science, Technion—Israel Institute of Technology, Haifa 3200003, Israel*
[2]*Russel Berrie Nanotechnology Institute, Technion, Haifa 320003, Israel*
[3]*Department of Biomedical Engineering, Technion, Haifa 320003, Israel*
[4]*Lorry I. Lokey Center for Life Sciences and Engineering, Technion, Haifa 320003, Israel*
dganit@cs.technion.ac.il

DNA labeling is a powerful tool in molecular biology and biotechnology that allows for the visualization, detection, and study of DNA at the molecular level. Under this paradigm, a DNA molecule is being labeled by specific k patterns and is then imaged. Then, the resulted image is modeled as a $(k+1)$-ary sequence in which any non-zero symbol indicates on the appearance of the corresponding label in the DNA molecule. For example, let $\boldsymbol{\alpha_1} = AGT$, $\boldsymbol{\alpha_2} = CG$ be two labels. For the DNA sequence $\boldsymbol{x} = AGTAGTCCCGCGATCGAGTCC$, it holds that the received output labeling sequence is $\boldsymbol{z} = 1001000020200020010000$.

The first goal of this work is to study the labeling capacity, defined as the maximum information rate achievable through the labeling process. For the case of using a single label, we demonstrate that the labeling capacity depends on the label's length and some additional properties, such as periodicity. We compute the labeling capacity for almost any label and organize all labels of length $\ell \leq 5$ according to their labeling capacity. We extend the definition of labeling capacity to multiple labels and present several results for these cases.

The second goal of our work is to study the minimum number of labels of the same length required to achieve the maximum labeling capacity of 2 for DNA sequences or $\log(q)$ for an arbitrary alphabet of size q. For DNA sequences, for labels of length 1, three different labels are necessary and sufficient to decode every DNA sequence from its labeling and to have capacity 2. For labels of length 2, it is clear that achieving labeling capacity of 2 may be obtained using any 15 different labels. Our main result here claims that 10 is the optimal number of labels to reach the maximum capacity. We also extend these results for arbitrary alphabet size q. The solution to this problem requires the study of path unique subgraphs of the de Bruijn graph with the largest number of edges. In [1], [2], we provide the optimal minimal number of labels of length one or two, over any alphabet of size q and in [3] we provide upper and lower bounds on this value for longer labels.

References:

[1] D. Hanania, D. Bar-Lev, Y. Nogin, Y. Shechtman, and E. Yaakobi, "On the capacity of DNA labeling," in 2023 IEEE International Symposium on Information Theory (ISIT). IEEE, 2023, pp. 567–572.

[2] D. Hanania, D. Bar-Lev, Y. Nogin, Y. Shechtman, and E. Yaakobi, "On the capacity of DNA labeling," arXiv preprint arXiv:2305.07992, 2024.

[3] C. Hofmeister, A. Gruica, D. Hanania, R. Bitar, and E. Yaakobi, "Achieving DNA labeling capacity with minimum labels through extremal de bruijn subgraphs," in 2024 IEEE International Symposium on Information Theory (ISIT), 2024, pp. 452–457.

SEQUENCE DETERMINANTS OF CRISPR-CAS12A OFF-TARGETREACTION

Gilboa M. R.[1], Assa G.[1], Das A.[3], Yakhini Z.[1,2], Somoza M.[3]

[1]Faculty of Computer Science, Technion, Haifa, Israel
[2]School of Computer Science, Reichman University, Hertzliya, Israel
[3]Leibniz Institute for Food Systems Biology, Technical University of Munich, Munich, Germany
mayan.rivlin@campus.technion.ac.il

Understanding off-target editing activity is crucial for improving the specificity and appli- cability of CRISPR technology [1]. Recent work suggests several CRISPR-based protocols for authentication and information storage applications. Characterizing cleavage profiles is impor- tant for these and other protocols. In this study, we present a comprehensive large-scale analysis of off-target cleavage patterns in Cas12a mediated single-stranded DNA (ssDNA) editing. Utiliz- ing micro-arrays synthesized by photolithography [2], we systematically map cleavage efficiencyacross a dataset of of over 57,000 unique sequences, spanning all possible single, double, and triplemodification variants within 24-nucleotide target sites.

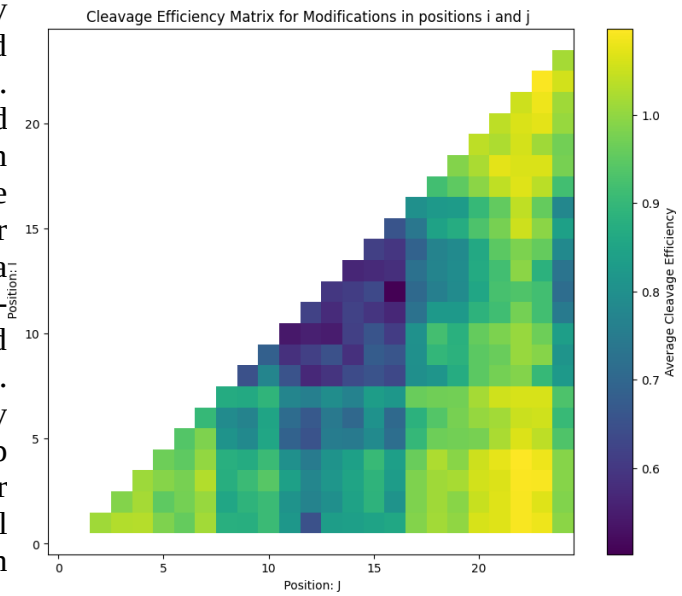


Figure 1: Cleavage efficiency heatmap for double nucleotide modifications across different positions (i, j) in a 24 bp target sequence. Darker regions indicate lower cleavage efficiency, highlighting po- sitions critical for reaction determination.

Our results reveal a position-dependent effect, identifying the 9-17 bp region as a critical determinant of cleavage efficiency, suggesting a potential alternative seed recognition mechanism specific to Cas12a ssDNA cleavage. Heatmaps (see Figure 1) and statistical enrichment analyses performed using mHG [3][4], further demonstrate that modifications located centrally within the target sequence significantly impact cleavage activity. These findings provide essential insights for enhancing CRISPR precision and its use in CRISPR-Cas based storage protocols.

Future work will focus on refining CRISPR-based authentication frameworks, characterizing Cas9 cleavage mechanism, and optimizing guide RNAs to enhance specificity.

References

[1] Guy Assa et al. Quantifying allele-specific crispr editing activitywith crispector2.0. *Nucleic Acids Research*, 2024.

[2] Kathrin Hölz et al. Multi-level patterning nucleic acid photolithography. *Nature Communications*, 2019.

[3] Eran Eden et al. Discovering motifsin ranked lists of dna sequences. *PLoS Computational Biology*, 2007.

[4] Fabian Wagner. Thexl-mhg test forenrichment. *arXiv*, 2015.

MOLECULAR ENCRYPTION FOR DNA DATA STORAGE

Vermathen R.T.[1,2], Bögels B.W.A.[1,2,3,4], Yurchenko A.V.[1,2,3], Takahashi C.N.[5], Markvoort A.J.[1,3], de Greef T.F.A. [1,2,3,4,6,7].

[1]*Institute for Complex Molecular Systems (ICMS), Eindhoven University of Technology, Eindhoven, The Netherlands*
[2]*Laboratory of Chemical Biology, Department of Biomedical Engineering, Eindhoven University of Technology, Eindhoven, The Netherlands*
[3]*Computational Biology Group, Department of Biomedical Engineering, Eindhoven University of Technology, Eindhoven, The Netherlands*
[4]*Department of Bioscience, School of Natural Sciences, Technical University Munich, Garching, Germany*
[5]*Paul G. Allen School of Computer Science and Engineering, University of Washington, Seattle, Washington, USA*
[6]*Institute for Molecules and Materials, Radboud University, Nijmegen, The Netherlands*
[7]*Center for Living Technologies, Eindhoven-Wageningen-Utrecht Alliance, Utrecht, The Netherlands*
t.f.a.d.greef@tue.nl

Inspired by its natural function as an information carrying molecule and motivated by its promising properties such as high stability [1] and high information density [2], DNA has emerged as an alternative medium for the long-term storage of digital information. As with any data storage medium, protection of the encoded information is desired for DNA data storage. Previous work has focussed on hiding DNA in objects [3], drowning out the signal of a message with a high background signal [4], or permanently deleting all data upon heating [5]. However, these methods either greatly decrease the information density of DNA or lead to loss of the stored data. Hence, we present a method of physical encryption in the form of a molecular locker. The locker consists of a DNA strand that is complementary to part of the data encoding strands of the file and contains a 3'-inverted T base. Due to sequence overlap, the locker competes with primers during an attempted amplification of the DNA file, greatly reducing the presence of the locked strands in the amplification product. This artificially creates drop-out when attempting to read the file using next-generation sequencing, preventing the data from being decoded. The locker can be removed by adding a password strand that is complementary to the locker. This allows for the repeated locking and unlocking of a DNA file without the loss of data.

References:

[1] Organick et al. 2021, An Empirical Comparison of Preservation Methods for Synthetic DNA Data Storage. Small Methods. 5.

[2] Organick et al. 2020. Probing the physical limits of reliable DNA data retrieval. Nature Communications. 11.

[3] Koch et al. 2020. A DNA-of-things storage architecture to create materials with embedded memory. Nature Biotechnology. 38:39-43.

[4] Clelland et al. 1999. Hiding messages in DNA microdots. Nature. 399:533-534.

[5] Kim et al. 2020. Metastable hybridization-based DNA information storage to allow rapid and permanent erasure. Nature Communications. 11.

SECONDARY STRUCTURE AVOIDANCE CODES FOR DNA STORAGE AND COMPUTING

Zhang Y.[1], Wang C.[1], Chu H.[1], Ge G.[2]

[1]*School of Cyber Science and Technology, Shandong University, Qingdao 266237, Shandong, China*
[1]*School of Mathematical Science, Capital Normal University, Beijing 100048, China*
ywzhang@sdu.edu.cn

In DNA storage, the secondary structure refers to the tendency of a single stranded DNA sequence to fold back upon itself, which is usually caused by the existence of two non-overlapping reverse complement substrings. We study the constrained system which avoids the appearance of secondary structure of a given length \(m\), by analyzing the capacity of the system and presenting some explicit constructions using TC-dominant sequences. In particular, our constructions have optimal rates 1.1679 bits/nt and 1.5515 bits/nt when \(m=2\) and \(m=3\), respectively. This work has been published in [1].

References:

[1] Chen Wang, Hui Chu, Gennian Ge, and Yiwei Zhang, The Capacity of Secondary Structure Avoidane Codes for DNA Sequences, *IEEE Transactions on Molecular, Biological, and Multi-Scale Communications,* vol. 10, no. 2, pp. 359-367, 2024.

DNA-MEDIATED ASSEMBLY OF PLASMONIC AND FLUORESCENT NANOPARTICLES FOR NANOPHOTONICS

Hansen N.[1], Petráková V.[1]

[1]*J. Heyrovský Institute of Physical Chemistry Czech Academy of Sciences, Dolejškova 3, 182 23 Prague, Czechia*
vladimira.petrakova@jh-inst.cas.cz

Various photonic phenomena rely on precise positioning of photon sources in close configurations with nanometer precision. DNA nanotechnology offers a solution for this, taking advantage of the programmable nature of DNA spatial configuration. Moreover, it offers an opportunity to assemble additional light-modulating elements, like plasmonic nanoparticles, opening new exciting avenues to improve our fundamental understanding of plasmonic interactions.

Here, we present a highly adaptable and easy-to-use platform that harnesses DNA nanotechnology to precisely position nanodiamonds containing fluorescent nitrogen-vacancy (NV) centers and plasmonic nanoparticles to investigate the modulations in the position of light emitted by an NV center using superresolution microscopy. NV centers are particularly well-suited for this investigation due to their broad absorption and emission spectra, large Stokes shift, and resistance to photobleaching.

To demonstrate the capabilities of our platform, we conducted several case studies in which we varied the distance between nanodiamond and gold nanoparticles, the number of assembled nanoparticles, and the controlled orientation of assembly of asymmetric nanostructures such as gold nanorods. By doing so, we were able to engineer the position of the plasmon resonance that suits preferred applications and needs. We evaluate the plasmonic enhancement mechanism of coupled NV centers and the shifts in their projected position based on the assembly design.

Additionally, by demonstrating controlled positioning of diamond nanoparticles containing NV centers, we open routes to use DNA-mediated assembly to improve magnetometry and other applications in the field of quantum optics where NV centers in diamond play a prominent role.

References:

[1] Hansen, N. et al., *arXiv,* http://arxiv.org/abs/2504.01749*,* 2025

MINIMIZING THE SEQUENCING COVERAGE IN DNA-BASED STORAGE SYSTEMS

Yaakobi, E.

Department of Computer Science, Technion, Israel

Although the expenses associated with DNA sequencing have been rapidly decreasing, the current cost of sequencing information stands at roughly $120/GB, which is dramatically more expensive than reading from existing archival storage solutions today. In this work, we aim to reduce not only the cost but also the latency of DNA storage by initiating the study of the **DNA coverage depth problem**, which aims to reduce the required number of reads to retrieve information from the storage system. Under this framework, our main goal is to understand the effect of error-correcting codes and retrieval algorithms on the required sequencing coverage depth. We establish that the expected number of reads that are required for information retrieval is minimized when the channel follows a uniform distribution. We also derive upper and lower bounds on the probability distribution of this number of required reads and provide a comprehensive upper and lower bound on its expected value. We further prove that for a noiseless channel and uniform distribution, MDS codes are optimal in terms of minimizing the expected number of reads. Additionally, we study the DNA coverage depth problem under the random-access setup, in which the user aims to retrieve just a specific information unit from the entire DNA storage system. We prove that the expected retrieval time is at least k for [n,k] MDS codes as well as for other families of codes. Furthermore, we present explicit code constructions that achieve expected retrieval times below k and evaluate their performance through analytical methods and simulations. Lastly, we provide lower bounds on the maximum expected retrieval time. Our findings offer valuable insights for reducing the cost and latency of DNA storage.

DNA SEQUENCE-BASED DATA STORAGE

Poh C. L.[1,2,3]

[1]*Department of Biomedical Engineering, National University of Singapore, Singapore*
[2]*Synthetic Biology for Clinical and Innovation (SynCTI), National University of Singapore, Singapore*
[3]*National Centre for Engineering Biology, Singapore*
biepcl@nus.edu.sg

As data is being generated at an ever-increasing rate, traditional storage methods face challenges related to maintenance costs, density limitations, and energy consumption. DNA data storage presents an attractive alternative solution for storing digital information as it offers several advantages [1]. DNA is extremely dense with very high information capacity, and has been shown to be very long-lasting. Due to it being the basis of life itself, DNA as a material will be eternally relevant to study and manipulation. DNA is particularly attractive as an alternative for long term shortage of data. With the advancement in DNA data storage over the years [2,3], researchers have successfully stored multiple forms of media in DNA sequence, including a Netflix television series and time capsules. In this talk, I will present current DNA sequence-based data storage workflows and related enabling technologies, explore new emerging solutions and applications, and discuss the challenges involved in making DNA data storage efficient and practical. I will also present on our recent effort in developing method that utilizes optogenetic circuits to capture 2-dimensional light patterns using multiple wavelengths of light directly into DNA in living cells [4], bypassing de novo DNA synthesis. This work demonstrates a direct capture of images into DNA, like a 'living digital camera', opening new possibilities for greater integration of biological systems with digital technologies for DNA data storage.

References:

[1] GM Church, Y Gao, & S Kosuri. Next-generation digital information storage in DNA. Science 337, 1628–1628 (2012).

[2] A Doricchi et al. Emerging Approaches to DNA Data Storage: Challenges and Prospects. ACS Nano 2022, 16, 11, 17552–17571

[3] CK Lim, S Nirantar, WS Yew, CL Poh. Novel modalities in DNA data storage. Trends in Biotechnology, (2021)

[4] CK Lim, JW Yeoh, AA Kunartama, et al. A biological camera that captures and stores images directly into DNA. Nat Commun 14, 3921 (2023).

PROGRAMMABLE VESICLES EMBEDDED IN ELECTROSPUN NANOFIBERS FOR THETRIGGERED RETRIEVAL OF ENCODED DNA DIGITAL DATA

Elosua J. Y.[1,2], Wetzl C.[1], Soukarie D.[1], Bittner A.[1,3], Santiago I.[1]

[1]CIC nanoGUNE BRTA, Donostia-San Sebastián, 20018, Spain
[2]University of the Basque Country, UPV/EHU, Donostia-San Sebastián, 20018, Spain[3]IKERBASQUE, Basque Foundation for Science, 48009 Bilbao, Spain
j.yeregui@nanogune.eu

DNA has emerged as a promising medium for digital data storage. To ensure the durability and preservation of the stored data, many strategies have been explored to protect DNA from oxidative. Additionally, accessing specific DNA sequences on demand co-located in the same space is a desirable random-access property of memory systems. This random access mostly relies on PCR selection **[1]**, which is susceptible to error introduction and data corruption. It also makes the whole process more energy- and time-consuming, as well as expensive. Moreover, with this method, the non-selected information may be lost during the process **[1]**.
Here, we combine the electrospun fiber technology **[2]** with a new strategy for random access based on selective release of the target DNA from the matrix. To this aim, several messages encoded in different sets of DNA sequences are encapsulated inside various stimuli-responsive SUVs (small unilamellar vesicles) before embedding them into the same electrospun water-resistant nanofiber mesh. We selectively release the target sequences using orthogonal physical triggers (e.g. temperature,pH; **Figure 1**). The orthogonality of the used triggers enables the preservation of the SUVs containingthe non-selected DNA, maintained in the fiber mesh for future retrieval purposes.
We show initial steps towards a molecular form of random access in water resistant fibers that can find use in DNA-based memory systems.

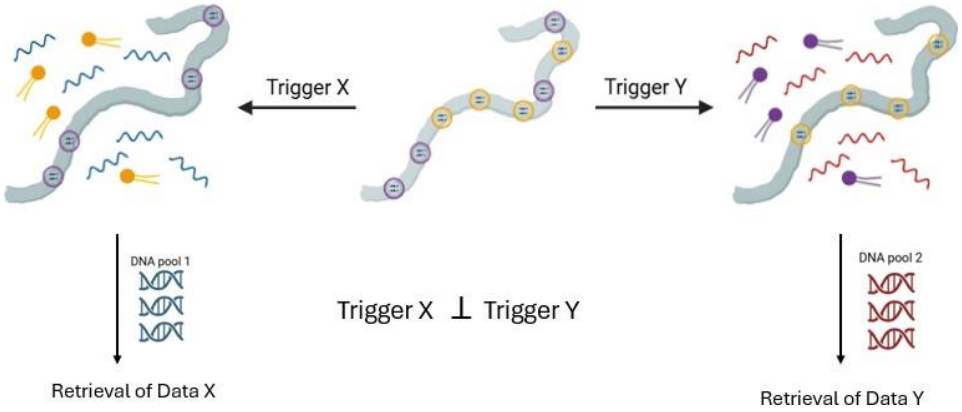


Figure 1. Scheme of the triggered retrieval of the desired data set [3].

References:

[1] Ceze, L., Nivala, J. & Strauss, K. Molecular digital data storage using DNA. *Nat Rev Genet* **20**, 456–466 (2019).
[2] Soukarie, D., Nocete, L., Bittner, A. M., & Santiago, I. DNA data storage in electrospun and melt-electrowrittencomposite nucleic acid-polymer fibers. *Materials Today Bio*, **24**, 100900 (2023).
[3] Created with BioRender.com

TOWARDS ENZYMATICALLY MODIFYING DNA WITHIN COMPACT MEMORY STORAGE DEVICES

Castronovo M.[1,2], Adedeji Olulana A.F.[1,§], Stopar A.[3], Suma A.[4,5], Aboagye-Mensah D.[1,†], Kailas L.[2,6], Thomson N.H.[6,7], Nicholson A.W.[3], Carnevale V.[3,5]

[1]*School of Food Science and Nutrition, University of Leeds, Leeds, UK*
[2]*Bragg Centre for Materials Research, University of Leeds, Leeds, UK*
[§]*current address: School of Mathematical and Physical Sciences, University of Sheffield, Sheffield, UK*
[3]*Department of Biology, Temple University, Philadelphia, PA, USA*
[4]*Department of Physics, University of Bari, Bari, Italy*
[5]*Institute of Computational Molecular Science, Temple University, USA*
[†]*Current address: Department of Structural and Molecular Biology, University College London, London, UK*
[6]*School of Physics and Astronomy, University of Leeds, Leeds, UK*
[7]*School of Dentistry, University of Leeds, Leeds, UK*
M.Castronovo@leeds.ac.uk

Current DNA engineering approaches are conceptualized around the classic vision that the function (and manipulation) of DNA is mainly determined by its one-dimensional (1D) sequence, such that DNA sequence recognition is the primary determinant of functional biological reactions and processes, including enzymatic DNA modification. My team has uncovered fundamentally new mechanisms, extending beyond the 1D-DNA dogma, for the control of enzymatic processing of DNA nanostructures. We have characterized the effect of nanoscale molecular confinement on the action of restriction endonucleases (REases) [1, 2] and DNA ligase in 2D DNA nanostructures, revealing that the interaction of these enzymes with DNA nanostructures can be quantitatively and precisely regulated by specific structural determinants that are not available to 1D DNA forms. For instance, REase cleavage of a 2D DNA nanostructure triangle exhibits a novel, digital ”on/off” behaviour, in that for each recognition site in the triangle, enzymatic cleavage is either efficient or fully inhibited [2]. By investigating REase action in surface-bound, nanoscale DNA “brush” assemblages, we have found that off-target DNA cutting can be regulated on/off by only varying DNA density [3]. Furthermore, T4 DNA ligase action in 2D DNA nanostructures can be suppressed (“off”) or strongly promoted (“on”) only by varying nanostructure shape. In the latter case, T4 DNA ligation efficiency does not depend on the number of ligation sites available on the nanostructure surface [4]. In this talk, I will present these studies in the perspective of applying DNA origami nanostructures and enzymatic DNA processing to add information in novel memory storage devices.

References:

[1] Stopar, A. et al. *Nucleic Acids Research* 46 (2), 995-1006, 2018

[2] Suma, A. et al. *Nucleic Acids Research* 48 (9), 4672-4680, 2020

[3] Adedeji Olulana, A.F. et al. *Molecules* 27 (16), 5262, 2022

[4] Aboagye Mensah et al. *manuscript in preparation*

CREATING AN ASSOCIATIVE MEMORY THROUGH COMBINATORIAL ASSEMBLY

Turek D.

CATALOG, USA

DNA storage presents the opportunity for efficient scalable computing by avoiding the data movement bottlenecks associated with conventional computing architectures. This presentation will describe a DNA data encoding strategy based on combinatorial assembly and show how this is used to create an associative memory of extreme scale to support a variety of computing applications. Through examples, the presentation will show how computing and storage can be combined in a single location where computing is brought to the data, avoiding the penalties associated with the Von Neumann Bottleneck

REPRESENTING CODES ON DNA USING ENZYMATIC LABELING

Bar-Lev D.[1], Etzion T.[2], Yaakobi E.[2], Yakhini Z.[3]

[1]*Center for emory and Recording Research, UCSD, USA*
[2]*Department of Computer Science, Technion, Israel*
[3]*School of Computer Science, Reichman University, Israel*
etzion@cs.technion.ac.il

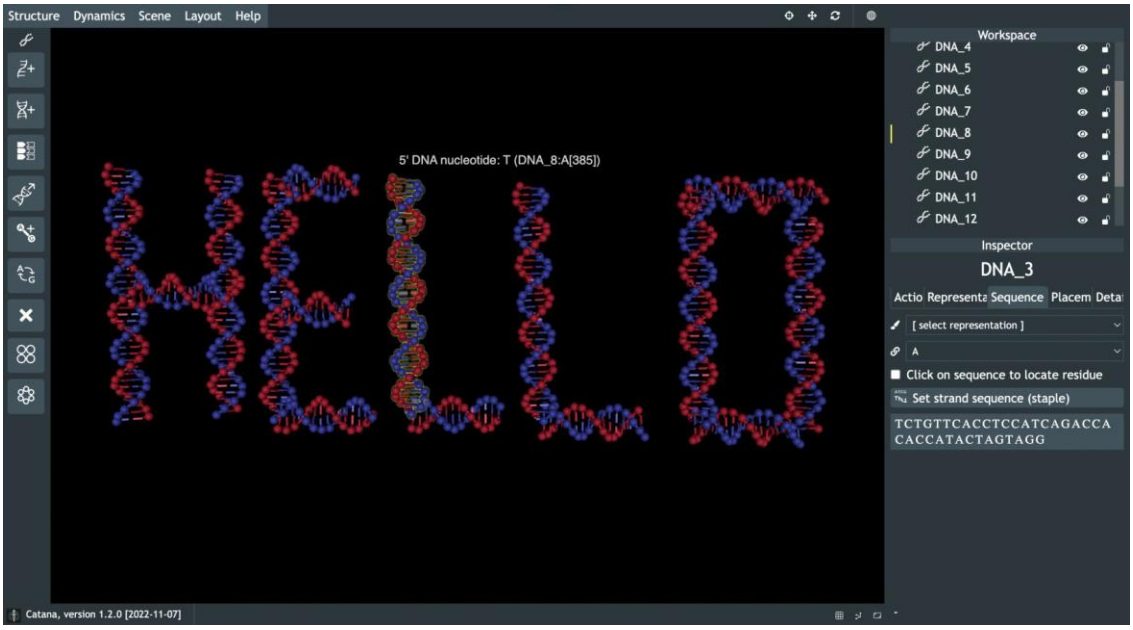
Enzymatic DNA labeling is a powerful tool with applications in biochemistry, molecular biology, biotechnology, medical science, and genomic research. This work contributes to the evolving field of DNA-based data storage by presenting a formal framework for modeling DNA labeling in strings, specifically tailored for data storage purposes. Our approach involves a known DNA molecule as a label template, employing patterns induced by a set of designed labels to represent information. One hypothetical implementation can use CRISPR-Cas9 and gRNA reagents for labeling. Various aspects of the general labeling channel, including fixed-length labels, are explored, and upper bounds on the maximum size of the corresponding codes are given. The study includes the development of an efficient encoder-decoder pair that is proven to be optimal in terms of maximum code size under specific conditions.

CATANA, AN ONLINE MODELLING ENVIRONMENT FOR PROTEINS AND NUCLEIC ACID NANOSTRUCTURES

Barišić I.[1,2], Schroeder F.[2], Kutak D.[2]

[1]*Molecular Diagnostics, AIT Austrian Institute of Technology, Vienna, Austria*
[2]*Eko Refugium, Slunj, Croatia*
Ivan.barisicr@ait.ac.at

In the last decade, significant advances have been made towards the rational design of proteins, DNA, and other organic nanostructures. The emerging possibility to precisely engineer molecular structures resulted in a wide range of new applications in fields such as biotechnology or medicine. The complexity and size of the artificial molecular systems as well as the number of interactions are greatly increasing and are manifesting the need for computational design support. In addition, a new generation of AI-based structure prediction tools provides researchers with completely new possibilities to generate recombinant proteins and functionalized DNA nanostructures. In this talk, we present Catana, a web-based modelling environment suited for proteins and DNA nanostructures.[1] User-friendly features were developed to create and modify recombinant fusion proteins, predict protein structures based on the amino acid sequence, and manipulate DNA origami structures. Moreover, Catana was jointly developed with the novel Unified Nanotechnology Format (UNF).[2] Therefore, it employs a state-of-the-art coarse-grained data model, that is compatible with other established and upcoming applications. A particular focus was put on an effortless data export to allow even inexperienced users to perform in silico evaluations of their designs by means of molecular dynamics simulations. After the publication of the manuscript, we continued the development and released a plugin system that allows external developers to add new features and functions to our online environment. Catana is freely available at http://catana.ait.ac.at/.



References:

1. Kuťák, D. et al. CATANA: an online modelling environment for proteins and nucleic acid nanostructures. Nucleic acids research; 10.1093/nar/gkac350 (2022).

2. Kuťák, D., Poppleton, E., Miao, H., Šulc, P. & Barišić, I. Unified Nanotechnology Format: One Way to Store Them All. Molecules (Basel, Switzerland) 27; 10.3390/molecules27010063 (2021).

ABSTRACTS

II. POSTERS

ID-BASED SELECTIVE TRANSCRIPTION FOR RANDOM ACCESS IN DNA-BASED DIGITAL DATA STORAGE

Lindberg A., Fördös F., Högberg B.

Department of Medical Biochemistry and Biophysics, Karolinska Institutet, Stockholm, Sweden
alexander.lindberg@ki.se

Random access of information is a key aspect for the successful deployment of DNA-based digital data storage. As the amount of data stored increases, the ability to selectively retrieve the specific data of interest without having to read out the entire pool of information becomes increasingly important. Currently, most systems rely on either PCR-based selection or physical separation of the data strands for random access [1]. PCR-based approaches use unique orthogonal primer pair combinations to selectively amplify the desired data strands from the pool. Physical separation, on the other hand, employs probes to fish out the data strands of interest from the pool, or relies on partitioning of the data and storage in separate compartments. However, these approaches carry several limitations. These include the controlled temperature cycling and time requirements of PCR reactions as well as biases and errors introduced during amplification, or the loss of data from the pool following retrieval for read out, requiring constant renewal. Recently, a system was developed using transcription-based read out of DNA strands adsorbed on specific dendricolloidal materials [2]. By using different promoter and RNA polymerase pairs, random access could be obtained, although at a limited scale. In this project, we aimed to combine an *in vitro* transcription-based read out approach with primer-based selectivity for improved random access, circumventing several of the limitations of previous approaches. Using single-stranded DNA information strands carrying a unique primer binding site acting as ID, the targets of interest can be selectively double-stranded and then read out through T7 RNA polymerase-based transcription. By investigating different polymerases, reaction conditions and template/primer combinations, we have developed an optimized system for the concurrent and selective double-stranding and transcription of DNA data strands, providing a new approach for random access of information.

References:

[1] Zhou, Y. et al. *ACS Applied Materials & Interfaces*, *16*(33), 43102–43113, 2024

[2] Lin, K. N. et al. *Nature Nanotechnology*, *19*(11), 1654–1664, 2024

MONITORING OF CHEMICAL REACTIONS BY SINGLE-MOLECULE SERS USING DNA ORIGAMI-BASED PLASMONIC NANOANTENNAS

Bald I., Kanehira Y., Beresowski C., Kogikoski Jr. S.

Institute of Chemistry, Hybrid Nanostructures, University of Potsdam, 14476 Potsdam, Germany
ilko.bald@uni-potsdam.de

DNA origami nanostructures are ideally suited to arrange both plasmonic nanoparticles as well as receptors for analyte molecules with nanometer precision. In this way they can provide optimized substrates for surface-enhanced Raman scattering (SERS), where the strongest signal enhancement is localized in nanometric hot spots and where the DNA origami can be used to precisely position the molecules of interest. In recent years we have demonstrated the few- and single-molecule SERS detection in different nanoparticle arrangements.[1-5]

The monitoring of chemical reactions on a single-molecule level is particularly interesting, because relevant intermediates can be identified and elementary reaction steps can be revealed. As an example for a rather complex molecular system we studied the enzymatic activity of horseradish peroxidase (HRP) on the single-molecule level by SERS and identified single reaction steps within a catalytic cycle in-situ.[6] However, a major complication in the interpretation of single-molecule SERS data is a potential interaction of the excited surface-plasmon resonance and the molecules detected by SERS. Consequently, we have studied in detail, how plasmon-induced chemical reactions take place in small organic molecules. First results on single-molecule SERS measurements of plasmon-induced chemical reactions will be shown.

References:

[1] S. Kogikoski Jr et al., *Chem. Commun. 59*, 4726, 2023

[2] K. Tapio et al., *ACS Nano 15*, 7065, 2021

[3] A. Dutta et al., *Nanoscale 14*, 16467, 2022

[4] A. Mostafa et al., *Nano Lett. 24*, 6916, 2024

[5] Y. Kanehira et al., *ACS Nano 17*, 21227, 2023

[6] Y. Kanehira et al., *ACS Nano 18*, 20191, 2024

COMPLEX DNA SYNTHESIS SEQUENCES

Moav B.[1], Yaakobi E.[1], Gabrys R.[2]

[1]*Computer Science, Technion, Haifa, Israel* [2]*CalIT2, UCSD, San Diego, California, USA*

boazmoav@cs.technion.ac.il

DNA-based data storage offers unprecedented information density,but widespread adoption is hindered by high synthesis costs. A keyfactor in these costs is the number of synthesis cycles required. This research presents a theoretical framework to minimize synthesis cycles.

Our "complex synthesis sequence" model bridges two predominant DNA synthesis approaches: (1) enzymatic synthesis, in which the synthesis machine permits the addition of any one nucleotide to individual strands per cycle. (2) photolithographic synthesis, in which the synthesis machine enables selective nucleotide addition across multiple strands simultaneously.

Our key contributions follow next. First, the introduction of the hybrid synthesis model that allows selection from a fixed number of nucleotides within each synthesiscycle. Moreover, we extend the mathematical concepts of subsequences and supersequences for complex synthesis sequences. And lastly, extending Lenz et al.'s [1] information rate definition and results, using an analog of the deletion ball, to analyze the novel framework.

Specifically, we define a model where a synthesis machine can access w distinct nucleotides from a total set of q nucleotides, with each strand receiving at most one nucleotide per cycle. For example, if $q = 4$, i.e., the DNA alphabet, and $w = 2$, the complex symbols are $\{A, C\}$, $\{A, G\}$, $\{A, T\}$, $\{C, G\}$, $\{C, T\}$ and $\{G, T\}$, and ifthe synthesis machine has $\{A, C\}$ in its first synthesis cycle (as in Figure 1), it can either append A to a subset of the synthesized strands, append C, or do nothing.

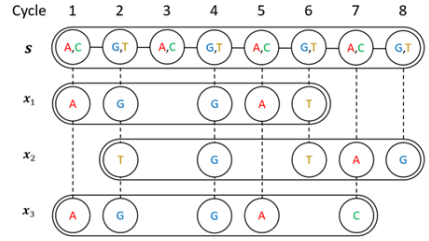


Figure 1. Cycles of complex synthesis, where $\boldsymbol{x}_1 = AGGAT$, $\boldsymbol{x}_2 = TGTAG$ and $\boldsymbol{x}_3 = AGGAC$ are synthesized using a periodic complex synthesis sequence of length 8.

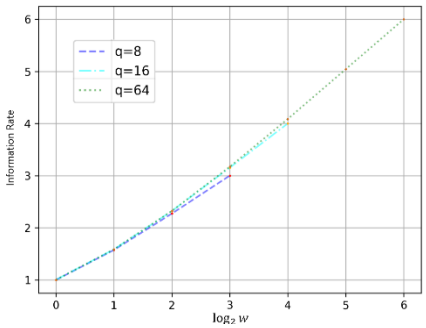


Figure 2. Maximum information rate for values of q and w.

We identify the complex sequences that maximize the informationrates under our synthesis model, which are the periodic complex sequences, and find specific values of the maximal information rate for different values of q and w. Moreover, we develop a dynamic programing algorithm to determine optimal synthesis sequences for known strands, showing that the solution is analogous to finding a shortest common supersequence (SCS). Notably, our results align with Lenz et al. [1] when $w = 1$.

To conclude, we demonstrate the results using the DNA case study. Consider $q = 4$, for the DNA alphabet. Then, for $w = 2$, the periodic complex sequence $(\{A, C\}, \{G, T\}, \{A, C\}, \{G, T\}, \dots)$ maximizes the information rate as a complex synthesis sequence. Furthermore, the value of the maximal information rate is 1.45, compared to 0.947 when $w = 1$. More results are in Figure 2.

References:

[1] A. Lenz, Y. Liu, C. Rashtchian, P. H. Siegel, A. Wachter-Zeh, and E. Yaakobi, "Coding for efficient DNA synthesis," in ISIT2020. IEEE, Jun. 2020.

REPLICABLE DATA STORAGE IN DNA ORIGAMI STRUCTURES

Fördös F., Lindberg A., Kloosterman A., Shen B., Högberg B.

Department of Medical Biochemistry and Biophysics, Karolinska Institutet, Stockholm, Sweden
bjorn.hogberg@ki.se

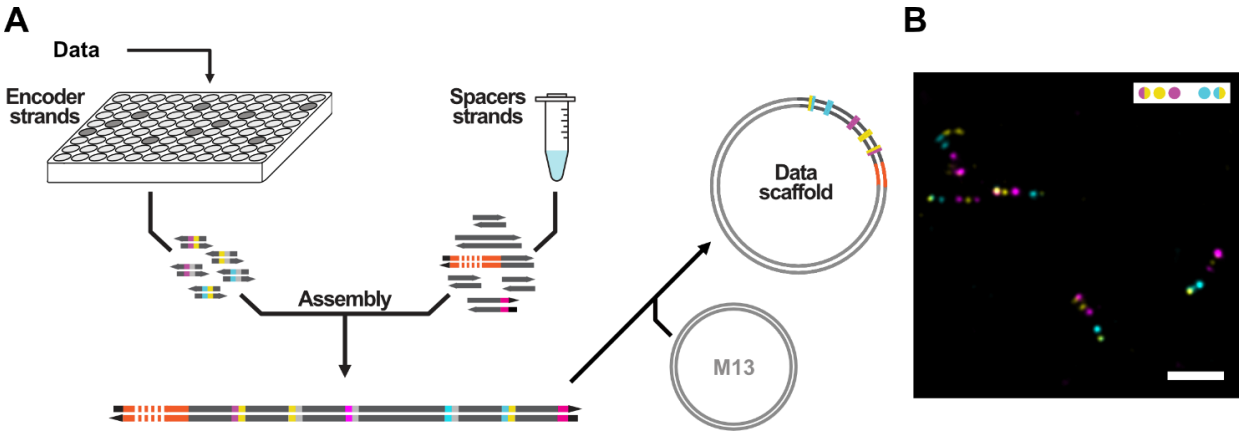


Fig. 1 Data storage in replicable DNA origami structures *A) Schematic of the workflow for constructing data encoding scaffold molecules. B) DNA PAINT super resolution image of data encoding scaffold molecules folded into a DNA origami structure (scale bar: 200nm)*

The looming data storage crisis caused by the rate of data production predicted to be outpacing the creation of new storage capacity in the coming decade led to the search for alternative storage media [1]. DNA being one of the leading candidates, due to its high information density and long-term stability, still has fundamental limitations that prevent it from being a commercially viable alternativeon larger scales, one of it being the prohibiting high costs of de-novo DNA synthesis [1]. In order to overcome this technology bottleneck strategies that rely on the usage of sets of DNA molecules produced in bulk are being developed [1]. DNA origami provides a valuable tool for these approaches,as it can create nanoparticles in which information can be encoded and stored using bulk synthetizedoligonucleotides that can be read with non-sequencing-based technologies such as AFM or microscopy [2]. One limitation of these techniques is that the stability of the information stored in thenanostructures is inherently coupled to the stability of the structures and thus to DNA hybridization, as the information is not encoded into molecules used to construct the structures but rather the assembly itself, making it sensitive to temperature and changing salt conditions and preventing the copying of the stored information using DNA replication [3]. Here we present a method with which we overcome these limitations by encoding information into libraries of DNA origami structures by introducing data-coding inserts constructed with combinatorial assembly. The produced structures canbe folded with a single set of staples and can be read out using DNA PAINT [4]. Additionally, the approach allows the replication of the library enzymatically or through clonal expansion. This approach thus provides a more stable and economical way to store data in DNA origami structures.

References:

[1] Doricchi, A. et al., *ACS Nano 16,* 11, 17552-17571, 2022
[2] Dickinson, G.D. et al., *Nat Commun 12*, 2371, 2021
[3] Ramakrishnan, S. et al., *CSBJ 16*, 342-249, 2018
[4] Schnitzbauer, J. et al., *Nat Protoc 12*, 1198–1228, 2017

ENHANCED CLUSTERING METHODS FOR DECAYED SYNTHETIC DNA IN DATASTORAGE APPLICATIONS

Assa G.[1], Grass R.[2], Yakhini Z.[1,3]

[1]*The Henry & Marilyn Taub Faculty of Computer Science, Technion, Israel*
[2] *Institute for Chemical and Bioengineering, Department of Chemistry and Applied Biosciences, ETH,Zurich, Switzerland*
[3]*School of Computer Science, Reichman University, Hertzliya, Israel*
guy.assa@campus.technion.ac.il

Clustering synthetic DNA presents significant challenges for data storage applications. Synthetic DNA breaks occur spontaneously and lack inherent structure, making reconstruction difficult.

Additionally, storage conditions accelerate DNA decay, introducing unexpected breaks that further complicate clustering [1]. Existing methods address these challenges with varying trade- offs. Methods like k-median and k-means clustering is computationally greedy, with runtime scaling with input size [2], while other approaches face similar efficiency constraints [3]. Deep learning approaches [4] sacrifice accuracy for improved runtime, potentially compromising data integrity—the primary concern in storage applications. Our method, which leverages user- defined primers and inherent DNA features, demonstrates the potential for accurate recovery of designed sequences from decayed DNA. Our approach uses primer-base reconstruction, adding optimization to the approach described in [5]. Our preliminary analysis of data from Meiser *et al.*

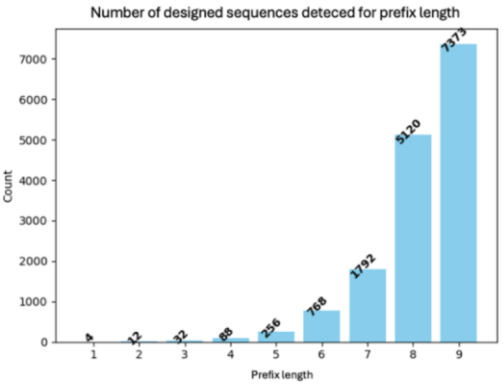


Figure 3

[1] reveals that the first 9 bases following the forward primer potentially allow detection of all7373 original designed sequences (Figure 1). Combining the reverse primer information and using optimization and advanced statistical techniques suggests a promising enhanced data retrieval from degraded DNA.

References

[1] Meiser, L.C., Gimpel, A.L., Deshpande, T. et al. Information decay and enzymaticinformation recovery for DNA data storage. Commun Biol 5, 1117 (2022). https://doi.org/10.1038/s42003-022-04062-9

[2] S. Guha, Y. Li, and Q. Zhang. Distributed Partial Clustering.arXiv preprint arXiv:1703.01539, 2017.

[3] M.-F. Balcan, Y. Liang, and P. Gupta. Robust Hierarchical Clustering. Journal of Machine Learning Research, 15(1):3831–3871, 2014.

[4] B. Betancourt, G. Zanella, J. W. Miller, H. Wallach, A. Zaidi,and B. Steorts. Flexible Models for Microclustering with Application to Entity Resolution. In NIPS, 2016.

[5] Antkowiak, P.L., Lietard, J., Darestani, M.Z. et al. Low cost DNA data storage using photolithographic synthesis and advanced information reconstruction and error correction. Nat Commun 11, 5345 (2020). https://doi.org/10.1038/s41467-020-19148-3

This work was funded by the European Union (DiDAX, 101115134). Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Council Executive Agency. Neither the European Union nor the granting authority can be held responsible for them.

COVERING ALL BASES: THE NEXT INNING IN DNA SEQUENCING EFFICIENCY

Abraham H.[1], Gabrys R.[2], Yaakobi E.[1]

[1] *The Henry & Marilyn Taub Faculty of Computer Science, Technion, Israel*
[2] *Calit2, University of California, San Diego*
hadasabraham@campus.technion.ac.il

The rapid growth of digital data, projected to reach 180 zettabytes by 2025, is causing a data storage crisis, with demand surpassing supply [1]. Existing storage technologies face challenges meeting big data demands. In response, DNA emerges as a promising medium due to its density and durability.

In this study we extend recent research addressing the *coverage depth problem* [2] by generalizing it to a more practical scenario. Specifically, we consider a container storing m files, each composed of k information strands. These strands are encoded into mn strands using some coding scheme, and the objective is to recover a files out of the total m. Our focus is on investigating the required coverage depth, considering factors such as the DNA storage channel and the error-correcting code. Additionally, we aim to explore the optimal pairing of an error-correcting code with a given DNA storage system to minimize coverage depth. This investigation is conducted within the framework of random access settings, where the user seeks to retrieve only a fraction of the stored information. In this context, we conduct both theoretical and experimental analyses to examine the expectation and probability distribution of the number of samples needed to fully recover the specified a files.

The study sheds light on the structural attributes of various coding schemes that impact random access expectations and probability distributions.

References

[1] J. Rydning, "Worldwide IDC global datasphere forecast, 2022–2026: Enterprise organizations driving most of the data growth," tech. rep., Technical Report, 2022.

[2] D. Bar-Lev, O. Sabary, R. Gabrys, and E. Yaakobi, "Cover your bases: How to minimize the sequencing coverage in DNA storage systems," IEEE Transactions on Information Theory, 2024.

SAVING INFORMATION IN SHAPE: PROGRAMMABLE DNA ORIGAMI

Haggenmueller S.[1], Matthies M.[1], Sulc P.[1,2,3]

[1]*School of Natural Sciences, Department of Bioscience, Technical University Munich, 85748 Garching,Germany*
[2]*Schoool of Molecular Sciences and Center for Molecular Design and Biomimetics, The BiodesignInstitute, Arizona State University, Tempe, AZ 85281, USA*
[3]*Center for Biological Physics, Arizona State University, Tempe, AZ 85287, USA*
psulc@asu.edu

Nanoscale structures that can change shape on command have exciting applications in medicine [1], biosensing [2], and smart materials [3]. DNA origami technology has enabled the creation of such dynamic nanostructures with precise control over their shape and functionality [4]. In this project, we designed a DNA origami composed of triangular units that can fold into two different 3D shapes depending on external signals. The project was divided into two parts: (1) reconfiguration of the 3D shape through strand displacement reactions, and (2) reconfiguration driven by temperature changes. In the first part, strand displacement was used, allowing for precise and programmable control over shape transformation. This approach has potential applications in molecular cargo delivery, where the structure could release or encapsulate a payload based on biochemical triggers, such as miRNAs. In the second part, the DNA origami net is designed to be able to reconfigure based on change in ambient temperature. Such thermally responsive nanomaterials could be used in biosensing, environmental monitoring, or even as molecular memory devices. Temperature changes could be "recorded" in the structure's shape, allowing it to retain a history of thermal events for environmental data tracking.

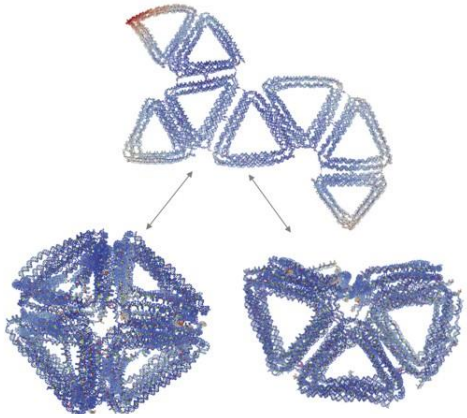


Figure 1) oxDNA simulations of net (top), and 3Dshapes octahedron (left) and boat (right)

To optimize the design, we performed oxDNA simulations. These simulations provided valuable insights into stability, reaction kinetics, and potential failure modes before testing them in the lab. We present preliminary experimental data of the system using TEM and fluorophore-quencher assays.

This project highlights the potential of integrating computational modeling with experimental testing to design smart, responsive nanomaterials. The ability to control shape at the nanoscale with both chemical and physical stimuli opens up new possibilities for smart materials,adaptable nanorobotics, responsive biomedical devices, and molecular data storage systems.

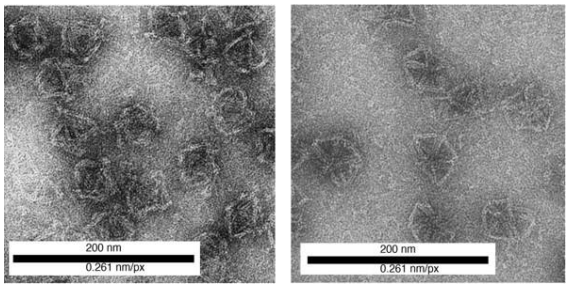


Figure 2) TEM images of DNA origami folding into octahedrons (left) and boat (right)

References:

[1] Cabane, E. et al., *Biointerphases 7*, 1-4, 2012

[2] Ramesh, M. et al., *Biosensors 13*, 1 40, 2022

[3] Tanjeem, N. et al., *Advanced materials 34*, 3,2022

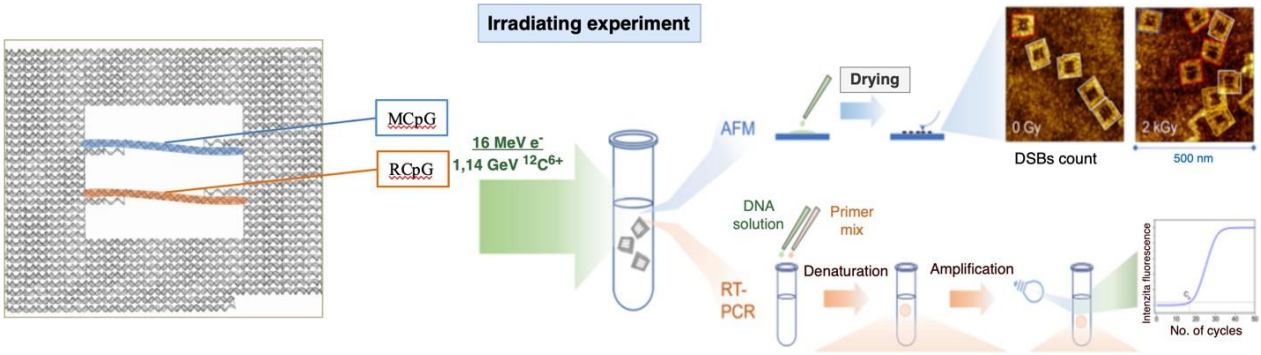
[4] Rothemund, P., *Nature 440*, 297–302, 2006

IMPACT OF CYTOSINE METHYLATION ON RADIATION DAMAGE TO DNA

Zápotocká T.[1,2], Kočišek J.[1], Sala L.[1]

[1]*J. Heyrovský Instituteof Physical Chemistry, Dolejškova 3, 18200, Prague CZ*
[2]*Faculty of Biomedical Engineering, Czech Technical University in Prague, 27201, Kladno, CZ*
tzapotocka@gmail.com

Understanding how cytosine methylation affects DNA damage induced by ionizing radiationis essential for advancing cancer radiotherapy, as DNA methylation may influence cellular responsesto treatment. [1,2] In this study, DNA origami nanoframes were employed as a platform to investigateradiation-induced damage to methylated DNA. The nanoframes, based on a design by Endo et al. [3],supported two parallel DNA sequences: one containing methylated CpG sites (MCpG) and one unmodified control (RCpG). These structures were irradiated using both low-linear energy transfer (LET) radiation (electrons) and high-LET radiation (carbon ions). Double-strand breaks (DBSs) wereassessed via atomic force microscopy (AFM), while overal DNA damage was quantified using qPCR. [4] The results revealed differences in DSB frequency between methylated and control sequences depending on radiation type. Notably, qPCR analysis consistently indicated reduced damagein methylated DNA, suggesting a protective role of cytosine methylation against radiation-induced lesions.



[3] [4]

Acknowledgment: EISMEA project no. 101115317 (NEO: Next Generation Molecular Data Storage)

References:

[1] HUANG, Rui-Xue and ZHOU, Ping-Kun 2020. DNA damage response signaling pathways andtargets for radiotherapy sensitization in cancer. *Signal Transduction and Targeted Therapy*, 5, 60.

[2] MIOUSSE, Isabelle R., et al. 2017 Effects of ionizing radiation on DNA methylation: from experimental biology to clinical applications. *International Journal of Radiation Biology*, 93, 5, 457.

[3] ENDO, Masayuki, et al. 2010. Regulation of DNA Methylation Using Different Tensions of Double Strands Constructed in a Defined DNA Nanostructure. *J. Am. Chem. Soc.*, *132*, 1592–1597.

[4] SALA, Leo, et al. 2022 Different Mechanisms of DNA Radiosensitization by 8-Bromoadenosineand 2′-Deoxy-2′-fluorocytidine Observed on DNA Origami Nanoframe Supports. Online. *The Journal of Physical Chemistry Letters*. 2022, 13, 17, 3922.

OPTIMALIZATION OF CISPLATIN-CROSS-LINKED DNA ORIGAMI NANOSTRUCTURES FOR DRUG DELIVERY APPLICATIONS

Jakubová K.[1,2], Perečko T.[3], Kočišek J.[1], Sala L.[1]

[1]*J. Heyrovký Institute of Physical Chemistry, Dolejškova 3, 18200, Prague, CZ*
[2]*Faculty of Biomedical Engineering, Czech Technical University in Prague, 27201, Kladno, CZ*
[3]*Department of Cell Biology and Radiobiology, Institute of Biophysics of the Czech Academy of Sciences, Kralovopolska 135, 61265, Brno, CZ*
katarina.jakubova@jh-inst.cas.cz

Currently, cancer persists as one of the greatest challenges in medicine, necessitating the development of new and more effective approaches for diagnosis and therapy. Nanomedicine, utilizing the unique physical properties of nanomaterials, emerges as a promising strategy for combating this heterogeneous disease. Thanks to these properties, drug delivery systems can be designed to be safer and more effective.

DNA nanostructures, due to their high biocompatibility, programmability, control over size, shape, and chemical surface composition represent an ideal tool for precision medicine, which has become a new paradigm in cancer treatment. DNA nanostructures possess key characteristics of an efficient and adaptable delivery system for various proteins, nucleic acids, or small molecules [1].

This project focuses on the development and comprehensive characterization of DNA nanostructures, specifically DNA nanoblocks [2], loaded with cisplatin [3] — a first-line drug widely used in cancer treatment, whose therapeutic efficacy is often limited by tumor cell resistance**.** Beyond its cytotoxic activity, cisplatin also enhances the structural stability of the nanoblocks and allows for high drug-loading capacity. To optimize the design and assess the biomedical potential of these nanostructures, we employed a combination of analytical techniques, including atomic force microscopy (AFM), spinning disk confocal microscopy (SDCM), and *in vitro* experiments on model cancer cell lines (FaDu and MCF7). The results demonstrate the potential of DNA nanoblocks as effective and modular platforms for targeted cancer therapy.

Acknowledgment: Czech Science Foundation supported the work via grant no. 24-11503S

References:

[1] Weiden, Jorieke and Bastings, Maartje M.C. 2021. DNA origami nanostructures for controlled therapeutic drug delivery. *Current Opinion in Colloid & Interface Science*. 52: 1 – 14
[2] Zhong, Yi-Fang et al. 2020. DNA Nanostructures as Pt(IV) Prodrug Delivery Systems to Combat Chemoresistance. *Smalll*. 16: 1 – 6
[3] Sala, Leo et al. 2022. Cisplatin-Cross-Linked DNA Origami Nanostructures for Drug Delivery Applications ACS Appl. Nano Mater.

PRIMER DESIGN FOR DNA STORAGE RANDOM ACCESS

Mateos J.[1,3], Lavenier D.[2], Dimopoulou M.[3], Genot A.[4], Antonini M.[1]

[1]I3S laboratory, Côte d'Azur University CNRS, UMR 7271, Sophia Antipolis, France
[2]University Rennes, Inria, CNRS, IRISA, Campus de Beaulieu, Rennes, France
[3]PEARCODE, Sophia-Antipolis, France
[4]LIMMS (IRL2820)/CNRS-IIS, University of Tokyo, Tokyo, Japan
Email: mateos@i3s.unice.fr, dominique.lavenier@irisa.fr, melpomeni@pearcode.io, genot@iis.u-tokyo.ac.jp,am@i3s.unice.fr

DNA is a promising alternative to traditional storage media due to the molecule's high density and long term stability. However, this novel medium presents challenges, particularly regarding the addressing to allow retrieving specific data from pooled DNA sequences, a process known as random access. This is achieved by designating an addressing zone on each DNA sequence known as a primer, a short DNA segment that acts as a file identifier for the stored information. Establishing a random access is crucial to optimize the efficiency and flexibility of data retrieval. The efficiency of the random access highly depends on the quality of the primer sequences which is strongly conditioned by structural constraints of DNA sequences. In this article, we propose a methodology for generating high-stringency primers that meet specific biochemical constraints, avoiding sequences that can form undesired shapes or loops that hinder DNA amplification and data retrieval. The tool uses a computational approach to predict the binding affinity and specificity of primers based on thermodynamic calculations. Users can adjust parameters to align with their specific wet lab protocols, providing accurate simulations that enhance comprehension, optimization, and efficiency in data retrieval during biochemical processes.

References:

[1] Lee Organick, Siena Dumas Ang, Yuan-Jyue Chen, Randolph Lopez, Sergey Yekhanin, Konstantin Makarychev, Miklos Z Racz, Govinda Kamath, Parikshit Gopalan, Bichlien Nguyen, etal., "Random access in large-scale dna data storage," Nature biotechnology, vol. 36, no. 3, pp. 242–248, 2018.

[2] Jian Ye, George Coulouris, Irena Zaretskaya, Ioana Cutcutache, Steve Rozen, and Thomas LMadden, "Primer-blast: a tool to design target-specific primers for polymerase chain reaction," BMC bioinformatics, vol. 13, pp. 1–11, 2012.

[3] Michela Testolina, Vlad Hosu, Mohsen Jenadeleh, Davi Lazzarotto, Dietmar Saupe, and TouradjEbrahimi, "Jpeg aic-3 dataset: Towards defining the high quality to nearly visually lossless quality range," pp.55–60, 2023.

[4] Richard Owczarzy, Bernardo G Moreira, Yong You, Mark A Behlke, et Joseph A Walder. Predicting stability of dna duplexes in solutions containing magnesium and monovalent cations.Biochemistry, 47(19):5336–5353, 2008

MOLECULAR DAMAGE IN DNA ORIGAMI INDUCED BY IONIZING RADIATION

Cardos K.[1,2], Sala L.[1], Kočišek J.[1]

[1] *Dept. of Dynamics of Molecules and Clusters, J. Hevrovský, Institute of Physical Chemistry of the CAS v. v. i., The Czech Academy of Science, Dolejškova 2155/3, 182 00 Prague 8, Czechia*
[2] *Department of Condensed Matter Physics, Faculty of Mathematics and Physics, Charles University, Ke Karlovu 5, 121 16 Prague 2, Czechia*
karen.cardos@jh-inst.cas.cz

DNA Origami techniques have been proven to be versatile tools for various applications in different fields, including its recent exploration as a storage medium. Its exceptional information density, stability, longevity, and sustainability make it a promising alternative to conventional storage systems [1]. Additionally, DNA Origami enables the precise nanoscale positioning of organic and inorganic materials, facilitating the encoding, storage, and retrieval of digital data.

This research continues our systematic exploration of the stability of DNA origami nanostructures to evaluate the radiation resistance of DNA-based memory systems to ensure extremely long-term data preservation [2]. A pre-designed rectangular DNA structure was used to arrange streptavidin in a specific pattern. We examined its interactions with high-energy projectiles and analyzed their surface modifications that impact structural integrity and information encoding. The findings will serve as proof-of-concept for a scalable, efficient, and durable DNA-based data storage technology, contributing to the development of next-generation digital storage solutions.

Acknowledgment: EISMEA project no. 101115317 (NEO: Next Generation Molecular Data Storage)

References:

[1] Doricchi, A.; Platnich, C.; Gimpel, A.; *et. al.*, *ACS Nano 16(11)*, 17552-17571, 2022.

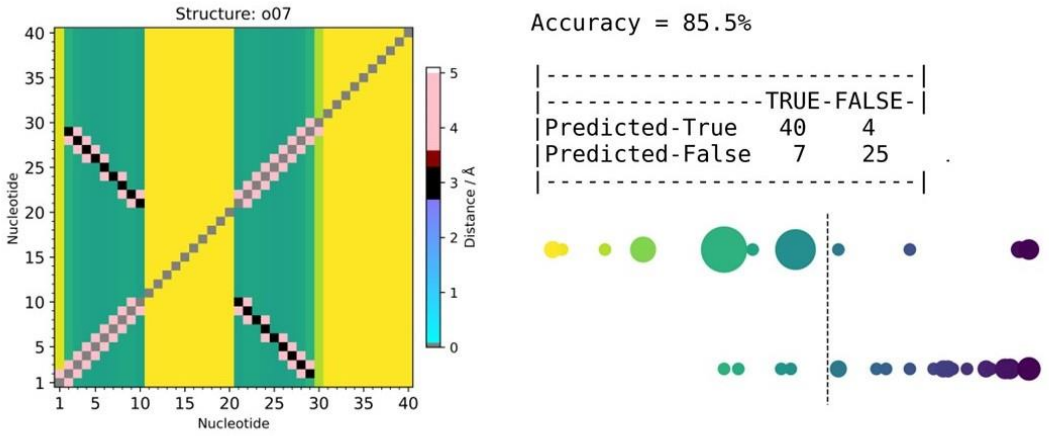
[2] Ameixa, J.; Sala, L.; Kocišek, J.; *et. al., ChemPhysChem 26(1),* 1-17, 2025.

OPTIMIZING OLIGONUCLEOTIDE LIBRARY FOR DNA DATA STORAGE WITH MOLECULAR DYNAMICS

Kožić M.[1], Wernhart K.[2], Schroeder F.[2], Barišić I.[3], Branimir Bertoša B.[1]

[1] *University of Zagreb, Faculty of Science, Department of Chemistry, Horvatovac 102a, 10000 Zagreb, Croatia*
[2] *Eko Refugium, Crno Vrelo 2, 47240 Slunj, Croatia*
3 Austrian Institute of Technology, Giefinggasse 4, 1210 Vienna, Austria
mkozic@chem.pmf.hr

Optimized through two billion years of evolution, DNA is an attractive target for modern data storage solutions. For companies to be able to cold-store massive amounts of data for e.g. disaster recovery, archiving, etc, DNA data storage system would have to be optimized (among other things) in terms of cost, efficiency, and stability. For that reason, a predetermined set of oligonucleotides with desired properties, an "oligolibrary", would be defined. Each oligonucleotide would map onto a packet of information (e.g. one byte), and oligonucleotides would have specific complements so that they assemble into a long strand of DNA that then encodes desired information. However, besides the desired interactions, oligonucleotides could bind in all other undesired ways. To optimize the selection of oligonucleotides it is necessary to estimate the affinity of intended (I) and unintended (UI) matches. Molecular dynamics (MD) is employed for the goal of assessing I and UI bindings between oligonucleotides from an oligolibrary. A structure is derived from sequences using AlphaFold3[1], and molecular dynamics using Gromacs[2] are run to research dynamical properties of pairs of oligonucleotides. A modified contact map depicting stable base pairings during simulations is used to predict if an oligonucleotide pair belongs to intended or unintended binding group. The modified map, called "Base Pairing Plot", is also discussed as a useful tool for batch analysis of DNA molecular dynamics simulations.



References:

[1] Abramson, J. *et al., Nature* 630, 493 – 500, 2024.

[2] Bekker, H. *et al., Physics computing* 92, 252 – 256,1993.

TOWARD HIGH-DENSITY STREPTAVIDIN PATTERNS ON DNA ORIGAMI NANOSTRUCTURES

Rabbe L.[1], Tomm E.[1], Grundmeier G.[1], Keller A.[1]

[1] *Paderborn University, Technical and Macromolecular Chemistry, Warburger Str. 100, 33098 Paderborn, Germany*
lukrabbe@mail.uni-paderborn.de

The functionalization of DNA origami has led to a wide range of applications including biochemistry [1], drug discovery [2], and drug delivery [3]. The precise and efficent decoration of DNA origami with high-density protein arrays, a topic of high interest, remains a challenge. In this study, we investigated the binding of streptavidin (SAv) to biotinylated staples on a rectangular DNA origami for different geometric arrangments having different distances between binding sites and different spacer lengths.. The design of the functionalized DNA origami with the different binding site arrangements is shown in **Figure 1** (left), with example AFM images (right), verifying SAv binding to neighboring sites. Our results reveal that maximum binding yields require a distance between binding sites of 12 nm, whereas a shorter distance of 6 nm results in notably decreased binding due to bidentate SAv-biotin binding and steric hindrance between neighboring proteins. However, even under such optimized conditions, binding yields rarely exceed 80% and decrease further when the array size is increased to twenty binding sites.With these results, this works contributes insights into how to design a twist corrected rectangular DNA origami in a way to achieve a possible high binding yield of SAv on a DNA origami surface.

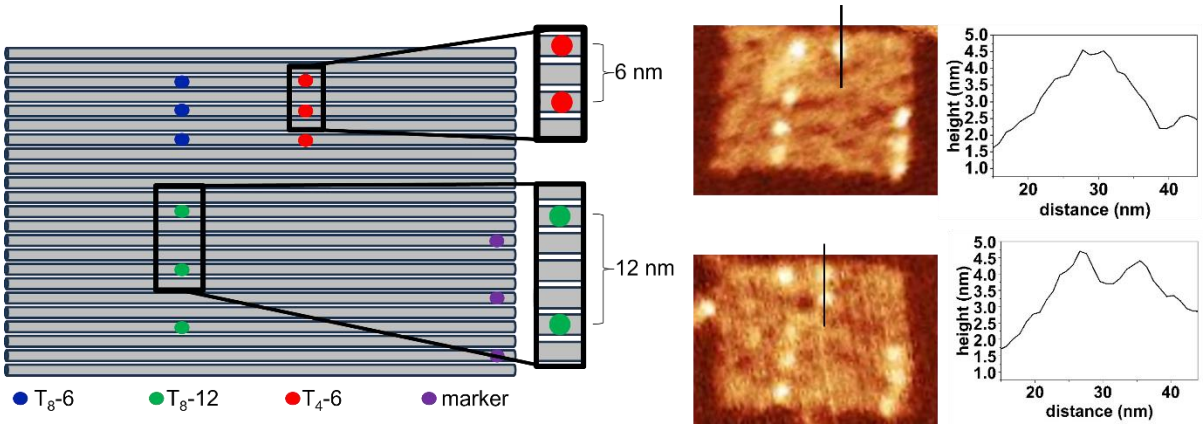


Figure 1. Arrangement of different SAv binding sites on a rectangular DNA origami with corresponding AFM images and height profiles.

References:

[1] J. Rackwitz, et al., *Angew. Chem. Int. Ed.* **2016**, 55, 10248.

[2] C. Kielar, et al., *Angew. Chem. Int. Ed.* **2018**, *57*, 14873.

[3] I. Mela, et al., *Angew. Chem. Int. Ed.* **2020**, *59*, 12698.

HIGH-YIELD ASSEMBLY OF PLASMON-COUPLED COLOR CENTER SYSTEMS VIA DNA ORIGAMI FOR TUNABLE LIGHT EMISSION

Hansen N.[1,3], Čopák J.[2,4], Kindermann M.[2], Roesel D.[1], Scollo F.[1], Bald I.[1,6], Cígler P.[2], Petráková V.[1,5]

[1]*J. Heyrovský Institute of Physical Chemistry, Czech Academy of Sciences, Dolejškova 3, 182 23 Prague, Czechia*
[2]*Institute of Chemistry and Biochemistry, Czech Academy of Sciences, Flemingovo naměstí 542/2, 160 00 Prague, Czechia*
[3]*Department of Physical Chemistry, Faculty of Chemical Engineering, University of Chemistry and Technology, Technická 5, 166 28 Prague 6, Czechia*
[4]*Department of Physical and Macromolecular Chemistry, Faculty of Science, Charles University, Hlavova 2030/8, 128 40 Prague 2, Czechia*
[5]*Faculty of Biomedical Engineering, Czech Technical University, Naměstí Sítná 3105, 27201 Kladno 2, Czechia*
[6]*Institut für Chemie, Universität Potsdam, Karl-Liebknecht-Straße 24-25, 14476 Potsdam, Germany*

vladimira.petrakova@jh-inst.cas.cz, petr.cigler@uochb.cas.cz

Spatial arrangement of optically active elements is a crucial factor in natural and engineered photonic systems. Color centers in nanodiamond offer a diverse portfolio of advantages in quantum sensing and information processing. However, precise and reproducible positioning, as well as coupling, into complex optical nanoarchitectures is challenging. DNA origami offers an elegant solution with the construction of fully addressable nanostructures, as previously demonstrated with precise positioning of various kinds of nanoparticles. Here, we present a robust method for covalent functionalization of fluorescent nanodiamonds, enabling the high-yield assembly of nanodiamonds and gold nanoparticles on DNA origami nanostructures. We reveal a distant dependent modulation of the color center emission with varying distance to the plasmonic nanoparticles via correlative Atomic Force and Fluorescence Lifetime Microscopy. Our findings further indicate selective plasmon-driven effects regarding radiative and non-radiative processes. We overcome key limitations in current nanodiamond assembly strategies and provide insight into the modulation of nitrogen vacancy color centers via plasmonic coupling that will advance to quantum photonic and sensing applications.

OPTIMIZING THE DECODING PROBABILITY AND COVERAGE RATIO OF COMPOSITE DNA

Cohen T., Yaakobi E.

Faculty of Computer Science, Technion-Israel Institute of Technology, Israel
Tomer.cohen@campus.technion.ac.il

The main challenge of making DNA storage systems competitive relative to existing storage technologies is the synthesis cost. The simplest and straightforward approach to reduce the cost is to increase the volume of data coded into a given length of oligos, while the information capacity can be measured by bits/symbol or bits/synthesis-cycle. The naive approach, working over A, C, G, T has a theoretical limit of $\log_2 4 = 2$ bits/symbol, while using error-correction codes can significantly decrease this limit. For example, the information rate in [3] and [4] is at most $\log_2 3 \approx 1.58$ since they imposed every two consecutive symbols to be distinct. On the other hand, if additional encoding characters are introduced, it is possible to increase the capacity and thus reduce the cost.

Composite DNA symbols were first introduced in [1], [2] to leverage the significant information redundancies built into the synthesis and sequencing technologies. A composite symbol is a representation of a position in a sequence that does not store just a single nucleotide, but a mixture of the four nucleotides. That is, a composite symbol can be abstracted as a quartet of probabilities $\{p_A, p_C, p_G, p_T\}$, such that p_b describes the fraction of the nucleotide $b \in \{A, C, G, T\}$ present in the mixture, where $0 \leq p_b \leq 1$ and $p_A + p_C + p_G + p_T = 1$.

This work studies two important aspects of the DNA composite model. The first problem is concerned with the sequencing process of DNA in which strands are read randomly. Since every read provides one base of the composite symbol, it is necessary to read many copies of the strands inorder to decode the composite, even in the noiseless case. In the second problem, we investigate how to carefully choose the mixtures of the probabilities. Namely, assume one wishes to design a DNA storage system with m composite symbols. Then, we study how to choose them so the decoding success probability of the Maximum Likelihood Decoder will be maximized.

References:

[1] L. Anavy, I. Vaknin, O. Atar, R. Amit, and Z. Yakhini, “Data storage in DNA with fewer synthesis cycles using composite DNA letters," *Nature Biotechnology*, vol. 37, no. 10, pp. 1229–1236, 2019.

[2] Y. Choi, T. Ryu, A. Lee, H. Choi, H. Lee, J. Park, S.H., Song, S. Kim, H. Kim, W. Park, and S. Kwon, “High information capacity DNA-based data storage with augmented encoding characters using degenerate bases," *Scientific Reports*, vol.9, 2019.

[3] N. Goldman et al., “Towards practical, high-capacity, low-maintenance information storage insynthesized DNA," *Nature*, vol. 494, no. 7435, pp. 77–80, 2013.

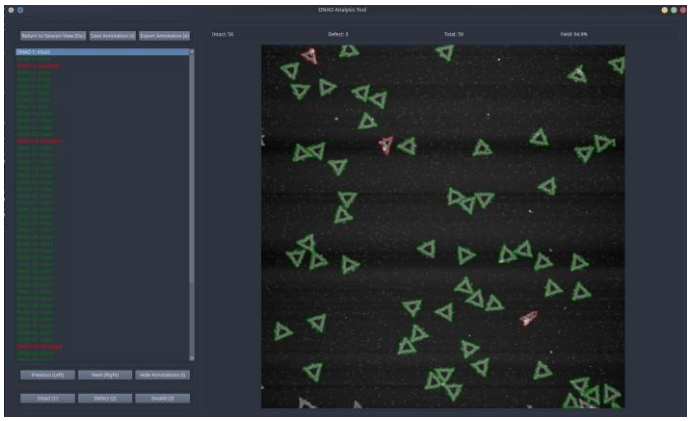
[4] H. Lee, R. Kalhor, N. Goela, J. Bolot, and G.M. Church, “Terminator-free template-independent enzymatic DNA synthesis for digital information storage," *Nature Communications*, vol. 10, no. 1, Jun., 2019.

ZERO-SHOT IMAGE SEGMENTATION FOR DNA ORIGAMI IN AFM IMAGES

Plantosar P.[1], Peharz R.[1]

[1]*Institute of Machine Learning and Neural Computation, Graz University of Techonolgy, Inffeldgasse 16b/I, 8010 Graz, Austria*
plantosar@tugraz.at

Research involving DNA origami [1] frequently relies on the manual annotation and classification of Atomic Force Microscopy (AFM) images to quantify the success of experiments. In recent years, efforts have been made to minimize the tedium of these tasks using automation. While modern computer vision methods such as YOLO [2] have shown promising results (e.g. [3]), they rely on training data manually annotated or at least verified by human experts. This limits how applicable these models are, especially if the experimental setup changes frequently. More recently, so-called foundation models have gained significant attention, not only within computer vision but across the broader machine learning community. What makes these models so powerful is their generalizability across a wide range of tasks, even with limited task-specific training. In this work, we focus on the Segment Anything Model (SAM) [4] a state of the art foundation model for image segmentation. We evaluate its effectiveness as a core component in a pipeline for detecting and classifying DNA origami structures in AFM images. This pipeline is then integrated into a custom software tool that aims to speed up the entire annotation process. Our overall goal is to assess whether the generalization capabilities of such models can reduce the reliance on extensive task-specific annotation in the context of nanoscale imaging. A screenshot of the tool is shown below.



References

[1] Rothemund, P. Folding DNA to create nanoscale shapes and patterns. *Nature 440*, 297–302 (2006).
[2] Joseph Redmon, Santosh Divvala, Ross Girshick, Ali Farhadi, You Only Look Once: Unified, Real-Time Object Detection, Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 779-788.
[3] Chiriboga, M., Green, C.M., Hastman, D.A. *et al.,* Rapid DNA origami nanostructure detection and classification using the YOLOv5 deep convolutional neural network., *Sci Rep 12*, 3871 (2022).
[4] Alexander Kirillov, Eric Mintun, Nikhila Ravi, Hanzi Mao, Chloe Rolland, Laura Gustafson, Tete Xiao, Spencer Whitehead, Alexander C. Berg, Wan-Yen Lo, Piotr Dollar, Ross Girshick, Segment Anything, International Conference on Computer Vision (ICCV), 2023, pp. 4015-4026.

HIERARCHICAL ENCODING OF JPEG2000-COMPRESSED IMAGES FOR DNA DATA STORAGE

Pic X.[1], Appuswamy R.[1]

[1]*Data Science Department, EURECOM, 450 Rte des Chappes, 06410 Biot, France*
xavier.pic@eurecom.fr

The increase in storage demand and low lifespan of conventional data storage media has transformed long-term archival and preservation into key bottlenecks for the data storage industry. Thus, researchers are now investigating innovative data storage media techniques. DNA molecules, with their high density, long lifespan and low energy needs, are promising candidates for alternative long-term data archival systems. However, current DNA data storage technologies are facing challenges with respect to cost (reading and writing on DNA is expensive) and reliability (reading and writing data is error prone). Thus, data compression and error correction are crucial to scale DNA storage and make it technologically and economically viable. Additionally, the DNA molecules encoding different files are very often stored in the same place, called an oligo pool. For this reason, without random access solutions, it is relatively impractical to decode a specific file from the pool, because all the oligos from all the files need to first be sequenced, which greatly deteriorates the read cost.

This paper introduces a solution (**Fig. 1**) to efficiently encode and store images into DNA molecules, that aims at reducing the read cost necessary to retrieve a resolution-reduced version of an image. This image storage system is based on the Progressive Decoding Functionality of the JPEG2000 codec and can be adapted for any other codec that enables a progressive decoding function. Each resolution layer is encoded into a set of oligos using the Raptor code [1] provided in the JPEG DNA VM software, with primers specific to the resolution layer attached to them. Depending on the desired resolution to be read, the set of oligos to be sequenced and decoded is adjusted accordingly. These oligos will be selected, augmented and sequenced through a PCR process run with the layer specific primers. The ReadUntil functionality of the Nanopore Sequencer can replace the PCR runs. It provides a system to reject at sequencing time the oligos that do not match a given template. This template can be dynamically modified during sequencing, allowing for better automation of the whole layer access process.

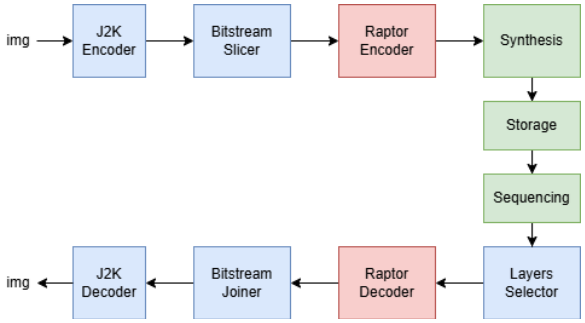


Figure 1: General Workflow of the submitted DNA-Adapted JPEG 2000 Progressive Encoder/Decoder

References:

[1] Lazzarotto, D. et al., "Technical description of the EPFL submission to the JPEG DNA CfP", *jpeg.org*, 2023

A STOCHASTIC PERFORMANCE ANALYSIS FRAMEWORK FOR DNA BASED AUTHENTICATION PROTOCOLS

Shafir R.[1], Limor O.[1], Yakhini Z.[1,2]

[1]*Faculty of Computer Science, Technion – Israel Institute of Technology, Haifa, Israel*[2]*School of Computer Science, Reichman University, Herzeliya, Israel*
RoyShafir8@gmail.com, Orielimor@gmail.com

Synthetic DNA is attracting increasing interest for its potential in many fields [1], in particular, a growing attention is being directed towards product tagging, driven by the need for enhanced security, precise traceability, and reliable authenticity verification [2,3]. However, the field of DNA-based authentication protocols is still in its early stages. Consequently, in this work, we aim to provide a framework that enables the analysis and comparison of these protocols' performance. We provide precise definitions, describe results, and present an analysis of example protocols, with a particular focus on a protocol inspired by the work of Volf et al. [2], as well as various modifications of it. Additionally, we analyze other DNA-based protocols, including one based on Chemical Unclonable Functions (CUFs) [3], and introduce an alternative set of protocols utilizing CRISPR-edited guide-RNAs. In [2], we assume that a manufacturer sends a cell-line to a client as a product. For authentication, the manufacturer first mutates the cell-line at N specific locations known only to them, and provides the client with a subset of these locations via primers. The client can then verify that the mutations appear as expected. Our proposed metric examines the relationship between the number ofproducts an adversary purchases and their probability of counterfeiting a product that will produce identical results when authenticated by the manufacturer. This probability is given by

$$\pi(t) = \sum_{b=0}^{n} P(\Gamma_{t-1} = \gamma) \sum_{\ell=0}^{\min(k_t,\gamma)} \frac{\binom{\gamma}{\ell}\binom{n-\gamma}{k-\ell}}{\binom{n}{k}} \left(\frac{1}{2}\right)^{k_t-\ell}$$

where n is the total number of locations and k_t isthe size of the subset sent at time t. $P(\Gamma_{t-1} = \gamma)$, calculated using dynamic programming, is the probability that the adversary knows γ modification locations after $t-1$ purchases. The second term represents the probability that the adversary guesses the outcome of a subset of k_t locations given prior knowledge of γ locations, following a hypergeometric distribution, and a coin toss for the remaining locations.

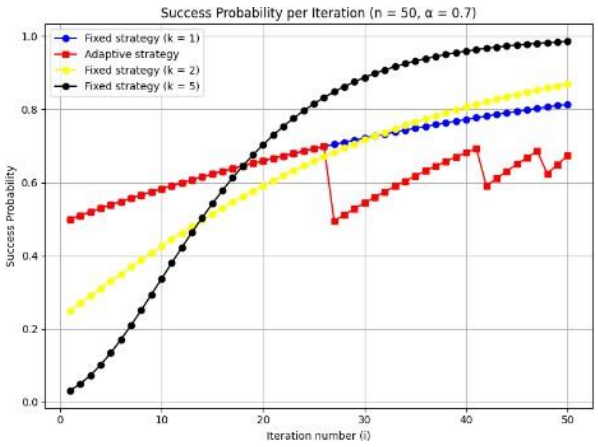


Figure 1. A comparison between the performance of strategies using different sizes of subsets sent to the client. We define α as the threshold accuracy of the adversary, at which the manufacturer considers the key to be compromised. This analysis led to the development of the α-optimal scheme for different values of k – the adaptive strategy (in red), which is greedily choosing the lowest k possible so that the success probability is lower than α.

References:

[1] Goldman, N. *et al.* "Towards practical, high-capacity, low-maintenance information storage in synthesized DNA". *Nature* **494**, 77–80 (2013).

[2] Volf, V. *et al.* "Cryptography in the DNA of living cells enabled by multi-site base editing." bioRxiv (2023): 2023-11.

[3] Luescher, A. *et al.* "DNA-Based Chemical Unclonable Functions for Cryptographic Anticounterfeit Tagging of Pharmaceuticals". ACS nano 18, no. 44 (2024): 30774-30785.

This work was funded by the European Union (DiDAX, 101115134). Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Council Executive Agency. Neither the European Union nor the granting authority can be held responsible for them.

EVALUATION OF STRUCTURAL VULNERABILITIES IN DNA ORIGAMI DATA STORAGE NANOPLATFORMS USING IONIZING RADIATION

Sala L.[1], Olšanský V.[2], Chvátil D.[2], Chevalier F.[3], Vizcaino V.[3], Méry A.[3], Kočišek J.[1]

[1]*Department of Dynamics of Molecules and Clusters, J. Heyrovský Institute of Physical Chemistry of the CAS, Dolejškova 3, 182 23, Prague, Czechia*
[2]*Nuclear Physics Institute of the CAS, Řež 130, 250 68, Řež, Czechia*
[3]*Normandie Univ, ENSICAEN, UNICAEN, CEA, CNRS, CIMAP, Boulevard Henri Becquerel, BP 5133, 140 70, Caen cedex 5, France*
leo.sala@jh-inst.cas.cz

Efforts to develop next-generation long-term archival media are driven by increasing data storage demands and the limitations of conventional disk- and tape-based systems [1]. DNA-based platforms such as DNA origami offer promising solutions due to their potential for high storage density via nanoscale bit encoding [1]. Understanding their long-term stability and damage mechanisms is crucial for optimizing these systems. We previously exposed DNA origami to ionizing radiation [2,3], which could model aging and damage pathways. Using atomic force microscopy (AFM) and gel electrophoresis, we observed remarkable stability of DNA origami in aqueous environments under gamma [2] and high-energy electron [4] irradiation, as well as in vacuum under ion beam exposure [3]. Recently, using the same methodology, we tested the radiation response in solution of three 2D DNA origami shapes designed for bit coding. High energy electrons (16 MeV) and ^{12}C ions (1.14 GeV) were used to model low and high linear energy transfer (LET) radiation, respectively, and to encompass both total and single-event effects. AFM analyses showed that damage from radiation exposure correlates with the total length of exposed edges—highlighting the role of both direct ionizing effects and secondary radical species from radiolysis. We are now refining this protocol to apply it to bit-coded structures, aiming to identify vulnerable regions and inform error correction strategies to aid future data recovery from anticipated losses.

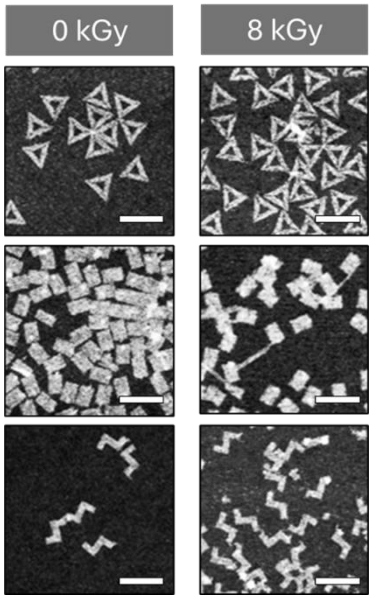


Figure 4. *AFM images of control and 16 MeV electron-irradiated 2D DNA origami (8 kGy total dose) in solution highlighting typical damage features (scale bar: 200 nm).*

Acknowledgment: EISMEA project no. 101115317 (NEO: Next Generation Molecular Data Storage)

References:

[1] Dickinson G.D. et al., *Nature Communications* 12, 2371, 2021
[2] Sala L. et al., *Nanoscale* 13, 11197-11203, 2021
[3] Sala L. et al., *Beilstein Journal of Nanotechnology* 15, 207-214, 2024
[4] Sala L. et al., *Journal of Physical Chemistry Letters* 13, 3922-3928, 2022

NOVEL STRAND DISPLACEMENT MECHANISMS FOR DATA MANIPULATION IN DNANANOSTRUCTURE-BASED DATA STORAGE DEVICES

Chan S.M.[1], Malyshava A.[2], Bae W.[2], Ouldridge T.[1]

[1] *Centre for Synthetic Biology, Department of Bioengineering, Imperial College London, London, UnitedKingdom.*
[2] *Soft Matter Group, Department of Physics, Faculty of Engineering and Physical Sciences, University ofSurrey, Guildford, United Kingdom*
sing-ming.chan@imperial.ac.uk

As global data storage demands approach 394 zettabytes by 2028, conventional storage technologies face critical challenges of scaling, obsolescence, and environmental sustainability. DNA, with its exceptional durability and information density, offers a promising alternative for sustainable data storage. However, traditional approaches rely heavily on costly DNA synthesis for writing operations, requiring unique sequences for each data bit. Our work addresses this challenge by developing two novel strand displacement-based mechanisms for writing and updating information on DNA nanostructure-based storage devices: 4-way/6-way strand displacement and associative handhold mediated strand displacement (AHMSD) [1].

Our approach encodes data using molecular 'bumps' on DNA nanostructures, in a manner analogous to how information is stored on compact discs where the presence of a bump represents a '1' and its absence a '0'. These mechanisms achieve high orthogonality through different architectures, enabling parallel updating of information. The key innovation in both mechanisms is their combinatorial architecture by using separate recognition domains for structures and positions, we dramatically reduce the number of unique DNA strands needed for information updating, eliminating the need for repeated DNA synthesis during writing operations.

Our preliminary results demonstrate successful data writing and deletion, with distinct kinetic profiles for each mechanism. The 4-way strand displacement achieves ~60% completion within 3 minutes in solution. AHMSD, using a 1-nt toehold and a 10-nt handhold, reaches approximately 50% yield within 2 hours. For deletion, 6-way strand displacement achieves 50% completion in ~15 minutes on DNA duplexes in solution, while AHMSD takes around 2 hours using a 2-nt toehold and 10-nt handhold. We also demonstrated this position-specific writing through AFM imaging, where a writing complex delivers a biotin-labelled overhang visualized by streptavidin binding as a ~4 nm bump at a targeted location. To evaluate the uniformity and orthogonality of these processes, we designed 9 orthogonal toehold pairs (81 combinations) for the 4-way strand displacement mechanism, and 10 distinct handhold sequences for AHMSD. However, both mechanisms showed variable kinetics and occasional leakage depending on sequence design.

To conclude, our research establishes fundamental design principles for DNA-based data storage through novel strand displacement mechanisms and lays crucial groundwork for scalable data manipulation in a localized environment of DNA nanostructure devices.

References:

[1] Cabello-Garcia, J., Bae, W., Stan, G.-B. V. & Ouldridge, T. E. *ACS Nano* **15**, 3272–3283 (2021).

COST-EFFICIENT FOLDING OF FUNCTIONALIZED DNA ORIGAMI NANOSTRUCTURES VIA STAPLE RECYCLING

Tomm E., Grundmeier G., Keller A.

Paderborn University, Technical and Macromolecular Chemistry, Warburger Str. 100, Paderborn 33098, Germany
Adrian.keller@uni-paderborn.de

DNA origami nanostructures (DONs) offer a highly programmable platform for applications ranging from biosensing to drug delivery – and increasingly, for molecular data storage [1]. However, high production costs, especially for modified staples, limit their broader use. Standard assembly protocols require a large staple excess [2], much of which is discarded after folding.

Here, we present a cost-efficient and straightforward strategy to recover and recycle unmodified, biotinylated and ATTO488-labeled staples. A rectangular DON served as the model system and molecular weight cut-off filtration enables effective separation of folded structures from excess staples. The recovered staples retained folding efficiency, structural integrity and functionality over at least five folding cycles as verified by agarose gel electrophoresis and atomic force microscopy.

Recycling reduces staple costs by 33% after five cycles, with savings up to 41% projected for extended reuse and high-density modifications. This approach improves resource efficiency and is particularly relevant for DNA-based data storage applications requiring extensive staple functionalization.

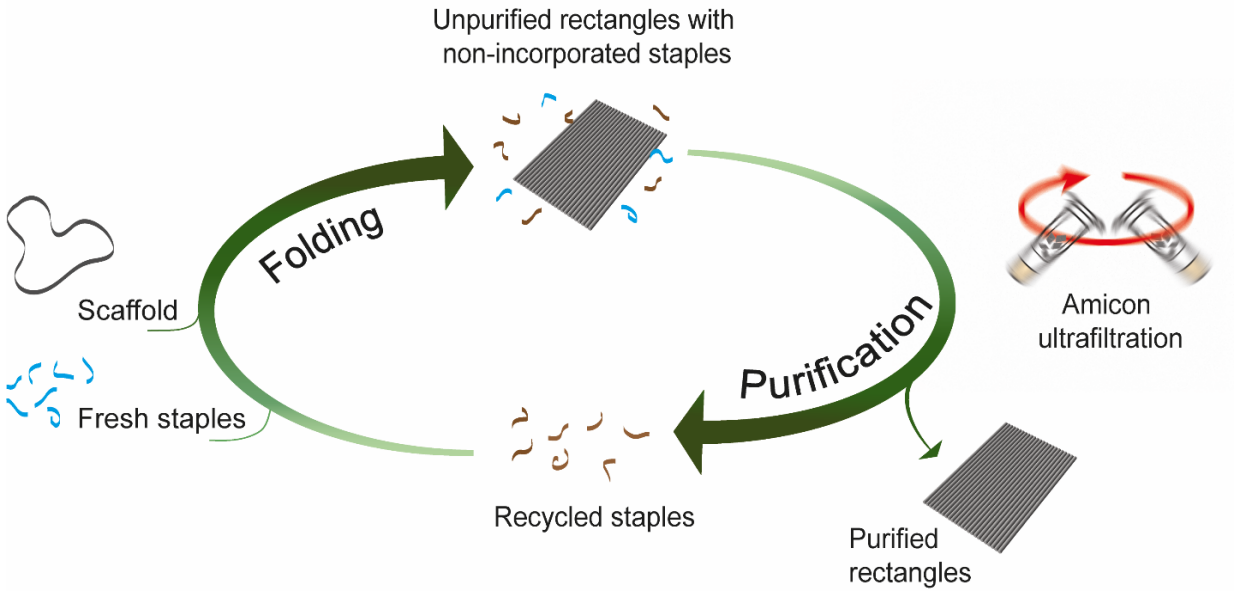


Figure 5 Schematic representation of the DON recycling cycle. Rectangular structures are folded using a mixture of fresh and recycling staple strands. Amicon filtration separates well-folded rectangles from non-incorporated staples which are collected and reused.

References:

[1] Doricchi, A., et al., *ACS Nano 16*, 17552-17571, 2022

[2] Rothemund, P. W. K., *Nature 440*, 297-302, 2006

CHROMOPHORES ASSEMBLIES AS NUCLEOTIDE-FREE-LETTERS TOWARDS ORIGAMI BASED DATA STORAGE

Malynovskyy V.L.

J Heyrovsky Institute of Physical Chemistry, CAS, Dolejškova 2155/3, 182 00 Prague, CZ
volodymyr.malynovskyy@jh-inst.cas.cz

The DNA based data storage systems, DBDS, are grounded on primary order of natural nucleotides (sequence-based approach), or on programable nanostructures of DNA-origami (structure-based approach).

Potential and limitations of DBDS approaches are essentially dependent on progress in synthesise (writing) of corresponding oligonucleotide sequences and/or their nanostructures, and on robust access to encoded information (reading). The field of DNA based data storage is predominantly designed on utilisation of natural letters, nucleobases (A, T, C, G) or their modifications (e.g. with periphery-appended fluorophores). [1; 2]

Nucleotide-free organic dyes, e.g. as cyanines, pyrenes, perylenediamines, others can be incorporated into sequences of natural oligonucleotides via classical phosphoramidite chemistry approach. [3]

Herein, incorporation of *nucleotide-free-letters* (NFL) into oligonucleotides, and its potential in data storage applications will be discussed. Figure 1.

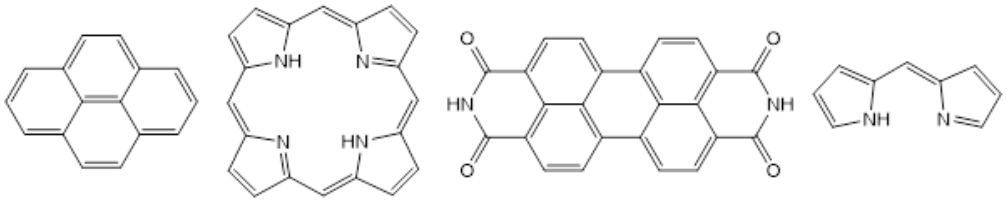


Figure 1. Cores of representative organic chromophores applicable as NFL in DBSB

Selected examples of chromophoric-NFL which do not disturb structure (shape, twist, recognition and hybridisation abilities, etc) within the natural oligonucleotides part, or can adapt and mimic DNA-like shape, induce and promote chirality, and extend an optical window up to 750 nm, will be presented.

It will be demonstrated that information might well be encoded in error-free mode into optical signals of chromophoric-NFL (absorbance, fluorescence, CD, CPL, excitonic interaction, etc.) with high fidelity.

Acknowledgement: Ministry of Education, Youth, and Sports of the Czech Republic via project CZ.02.01.01/00/22_008/0004649 QUEENTECH co-funded by the European Union

References:

[1] Doricchi, A. et all., ASC Nano, 16, 17552-17571, 2022

[2] Yang, S. et all., Nature Reviews Chemistry 8, 179-194, 2024

[3] Malinovskii V.L (Malynovskyy, V.L.) et all., Chem. Soc. Rev., 39 (2), 410–422, 2010

EVALUATING THE LONG-TERM STABILITY OF DNA ORIGAMI NANOSTRUCTURES

Li X.[1], Grundmeier G.[1], Keller A.[1]

[1]*Technical and Macromolecular Chemistry, Paderborn University, Warburger Str. 100, 33098 Paderborn, Germany*
xinyangl@mail.uni-paderborn.de

DNA origami nanostructures [1] are assembled by folding a long single-stranded DNA scaffold with the help of numerous short staple strands, enabling precise control over molecular architecture at the nanoscale. Their structural versatility and addressability have led to widespread applications in areas such as molecular data storage [2], biosensing [3], and drug delivery [4]. However, many of these applications expose DNA origami nanostructures to environmental or processing conditions that can compromise their structural integrity. Ensuring the stability of DNA origami nanostructures is crucial for maintaining their functional performance.

To investigate the factors influencing DNA origami stability and to predict their structural lifetime, we subjected two-dimensional DNA origami structures to mildly acidic conditions (pH 5.0, 5.2, and 5.4) and elevated temperatures (50.7 °C, 55.7 °C, and 60.7 °C). Structural integrity over time was monitored using atomic force microscopy (AFM) and agarose gel electrophoresis. We observed that lower pH values and higher temperatures significantly accelerated degradation. The degradation behavior is currently being fitted to a first-order kinetic model. Although the degradation process is complex, it can be effectively described by an apparent rate constant. An Arrhenius analysis of these rate constants allows estimation of an apparent activation energy, providing a practical method for predicting the lifetime of DNA origami structures under specific conditions.

References:

[1] Rothemund, P.W.K., *Nature* 440, 297–302, 2006.

[2] Dickinson, G.D. et al., *Nat. Commun.* 12, 2371, 2021.

[3] Daems, D. et al., *ACS Appl. Mater. Interfaces* 10, 23539–23547, 2018.

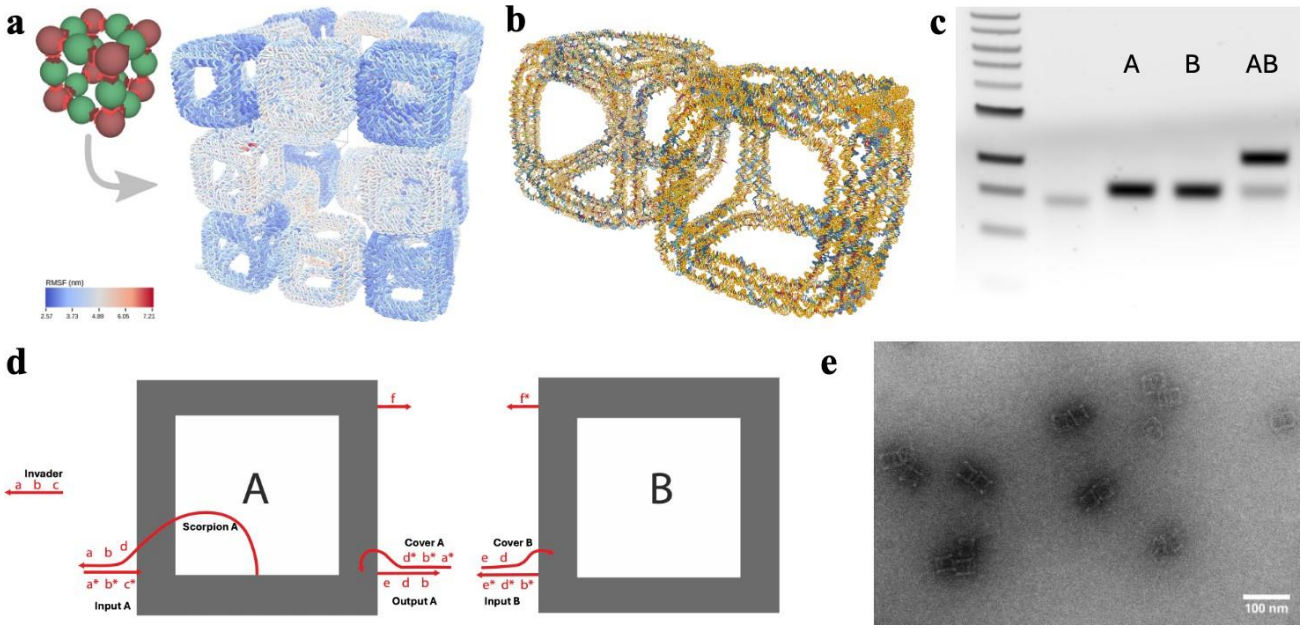
[4] Li, J. et al., *Adv. Mater.* 25, 4386–4396, 2013.

SIGNAL PASSING INSIDE A WIREFRAME DNA ORIGAMI CUBE

Hanrieder F.[2], Matthies M.[2], Šulc P.[1,2]

[1]*The Biodesign Institute, Arizona State University, Tempe, AZ 85287, USA*
[2]*Department of Biosciences, Technical University of Munich, ZNN Garching, 85748, Germany*
psulc@asu.edu

Constructing larger and more complex structures from DNA origami [1] requires the controlled assembly of multiple units. We implement these dynamic units via a modular DNA origami wireframe cube.
We experimentally demonstrate the simplest form of interaction - dimerization - where complementary single-stranded overhangs [2] on distinct cube species (A and B) drive specific binding. Furthermore, we developed a conditional dimerization mechanism in which binding is triggered only upon the addition of an input strand. This input initiates a strand displacement cascade [3] that removes the protective cover from one cube and subsequently enables interaction with another cube. All these interactions are optimized in silico (with the use of oxDNA simulations [4]) before experimental verification.
We hope this strategy can ultimately be used as a platform for programmable, signal-responsive 3D DNA assemblies capable of storing and propagating information.



a) Polycube assembly [2] b) oxView representation of the dimerized cubes c) Gel electrophoresis showing the difference between monomers (A, B) and dimers (AB) d) Proposed full signaling mechanism e) TEM image of a dimer sample

References:

[1] Rothemund, P. W. K., *Nature*, 440(7082), 297–302, 2006
[2] Bohlin, J. et al., *ACS Nano*, 17(6), 5387–5398, 2023
[3] Simmel, F. C. et al., *Chemical Reviews*, 119(10), 6326–6369, 2019
[4]Poppleton, E. et al., *The Journal of Open Source Software*, 8(81), 4693, 2023

HYBRID DNA ORIGAMI NANOSTRUCTURES

Ruiz Arce D.D.[1], Jonáš A.[2], Sala L.[1], Zemánek P.[2], Kočišek J.[1]

[1]*J. Heyrovský Institute of Physical Chemistry of the CAS, Dolejškova 2155/3, 18200 Prague, Czech Republic*
[2]*Institute of Scientific Instruments of the CAS, Královopolská 147, 61200 Brno, Czech Republic*
david.ruiz@jh-inst.cas.cz

Hybrid DNA origami complexes provide a versatile platform to fabricate photonic systems [1]. In this work, we present a strategy to functionalize DNA Origami structures via site-specific modifications, allowing object attachment and fluorescent labeling. Synthetic DNA strands were designed and chemically modified to incorporate fluorophores at defined positions, allowing precise visualization and tracking. Additionally, we introduced linker groups at specific sites to support covalent or non-covalent binding interactions. The successful integration of these modifications was verified using a combination of fluorescence imaging and structural characterization techniques. These hybrid structures establish a foundation for the development of tailored optomechanical probes, advancing micromanipulation and optical trapping applications.

Acknowledgment:

This work was supported by MEYS, OP JAK project CZ.02.01.01/00/22_008/0004558 "AMULET" and 24-11503S Czech Science Foundation.

References:

[1] Kuzyk, A.; Jungmann, R.; Acuna, G. P.; Liu, N. DNA Origami Route for Nanophotonics. ACS Photonics 2018, 5, 1151–1163.

PROBING FLUOROPHORE – PLASMON INTERACTIONS WITH FLUORESCENT GOLD NANOCLUSTERS

Umar K.[1,2], Zawada Z.[1], Hansen N.[1,3], Petráková V.[1]

[1]*J. Heyrovsky Institute of Physical Chemistry, 3, Dolejškova 2155, 182 00 Prague, Czech Republic.*
[2]*Czech Technical University in Prague, Náměstí sítná 3105, 272 01 Kladno, Czech Republic.*
[3]*University of Chemistry and Technology, Technická 5, 160 00 Prague, Czech Republic.*
vladimira.petrakova@jh-inst.cas.cz

Gold nanoclusters (AuNC) have attracted significant attention due to their unique photophysical properties, making them ideal as labels in fluorescent microscopy and for other nanophotonic applications. Here, we aim to synthesize highly fluorescent AuNC, enhance its photophysical properties through surface modification, and enable its conjugation to DNA. Using these DNA-conjugated AuNCs as fluorescent probes, we seek to explore the mechanisms underlying plasmon–fluorophore interactions.

Current studies on fluorescence enhancement and positional shifts induced by plasmonic nanoparticle/rod interactions have largely focused on using organic dyes, which exhibit small Stokes shifts, making it challenging to independently target their excitation and emission spectra since the plasmon resonance of plasmonic nanoparticle/rod are very broad. Unlike organic dyes, AuNC exhibit both large Stokes shift (> 180 nm) and exceptional photostability, making it better suited for studying plasmon – fluorophore interactions.

We developed a modular chemical architecture that enhances fluorescence and enables DNA labelling of AuNC, which is essential for employing DNA origami technology in the defined spatial configuration of AuNC and gold nanoparticles. Here we present two distinct synthetic strategies for producing highly fluorescent, DNA-functionalizable AuNC. The first strategy employs carboxylate-decorated AuNC synthesized in aqueous solution. The second approach utilizes amino-decorated AuNC prepared in dipolar aprotic solvents.

Both systems successfully incorporate azide groups for subsequent DNA conjugation via copper-free click chemistry, yielding DNA-decorated AuNC suitable for use in DNA origami nanotechnology. Combining them with DNA nanotechnology, these DNA-functionalized AuNCs can be precisely arranged on DNA origami nanostructures with site-selective addressability in the presence of plasmonic nanostructures. This capability facilitates distance-dependent studies of plasmonic coupling effects, enabling deeper insights into nanoscale photonic interactions

CUSTOM OPTOMECHANICAL PROBES BASED ON DNA ORIGAMI

Jonáš A.[1], Ruiz Arce D. D.[2], Sala L.[2], Benešová M.[1], Pilát Z.[1], Ježek J.[1], Zemánek P.[1], Kočišek J.[2]

[1]*Institute of Scientific Instruments of the CAS, Královopolská 147, 61200 Brno, Czech Republic*
[2]*J. Heyrovský Institute of Physical Chemistry of the CAS, Dolejškova 2155/3, 18200 Prague, Czech Republic*
sasa@isibrno.cz

Optical micromanipulations enable confinement and directed transport of micron-sized particles suspended in liquids or gases, with applications in research fields ranging from physics through chemistry to biology [1]. Typically, optically manipulated particles have spherical or nearly spherical shapes. However, there is an increasing interest in optical manipulation of objects with custom-defined shapes and optical properties (polarizability, chirality) that determine object's behavior upon illumination with the trapping light and enable novel types of optomechanical experiments [2]. Self-assembly mediated by carefully designed DNA origami scaffolds offers an attractive path towards fabrication of trapping objects with the desired characteristics from suitable building blocks (colloidal particles, quantum dots, fluorophores) [3].

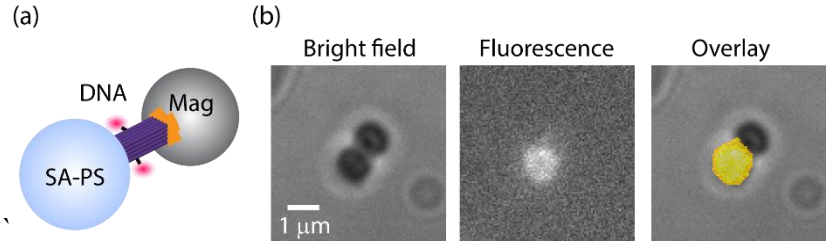


Fig. 1: (a) Hybrid optomechanical probe formed by a 1 μm streptavidin-coated polystyrene particle (SA-PS) and 0.9 μm paramagnetic particle (Mag) connected by 24HB DNA origami (DNA) labeled with Cy3 fluorescent dye. (b) Optically trapped hybrid probe in different imaging modes.

We report on the pilot optical trapping experiments with custom made optomechanical probes self-assembled from streptavidin-coated polystyrene particles and paramagnetic particles connected by fluorescently labeled DNA origami linkers [see Fig. 1(a)]. We manipulated the synthesized hybrid probes using a single-beam optical trap and recorded both bright-field and fluorescence images of the trapped probes [see Fig. 1(b)]. Our experiments showed that the probes are mechanically stable and emit specific fluorescence even after prolonged exposure with the trapping light and their orientation in the trapping beam can be controlled by adjusting the manipulation conditions. Furthermore, the presence of a paramagnetic particle also renders the probes sensitive to external magnetic fields.

The reported asymmetric probes that can be manipulated via optical, magnetic, and hydrodynamic forces represent a unique experimental tool for the investigation of coupled translational-rotational dynamics with non-linear and nonconservative character, and study of self-arranged dynamic metamaterials formed by optical binding of multiple structures illuminated by structured light fields.

References:

[1] Zemánek, P. et al., *Advances in Optics and Photonics 11*, 577–678, 2019

[2] Gieseler, J. and Millen, J., *Entropy 20*, 326, 2018

[3] Kuzyk, A. et al., *ACS Photonics 5*, 1151–1163, 2018

ELECTRON ACCELERATOR MICROTRON MT25 AS SUITABLE DEVICE FOR DNA ORIGAMI IRRADIATION

Olšanský V.[1,2], Chvátil D.[2], Kočišek J.[1], Sala L.[1]

[1]*Department of Dynamics of Molecules and Clusters, J. Heyrovský Institute of Physical Chemistry, Dolejškova 3, 182 23 Prague 8, Czech Republic*
[2]*Department of Accelerators, Nuclear Physics Institute of the CAS, Husinec – Řež 130, 250 68 Řež*
vaclav.olsansky@jh-inst.cas.cz

Microtron MT25 is an electron accelerator of the Nuclear Physics Institute of the Academy of Sciences of the Czech Republic, located in a detached laboratory in Prague. It is a cyclic electron accelerator with Kapitza's resonator (accelerating cavity) [1]. The principle of the Microtron accelerator is based on the passage of electrons through the accelerating cavity, where the electrons are accelerated by RF electric field with constant amplitude and frequency. Accelerated electrons move in circular trajectories in a homogenous magnetic field. The accelerating cavity is located at the common tangent point of these trajectories. The principle of the acceleration process is shown in Fig. 1.

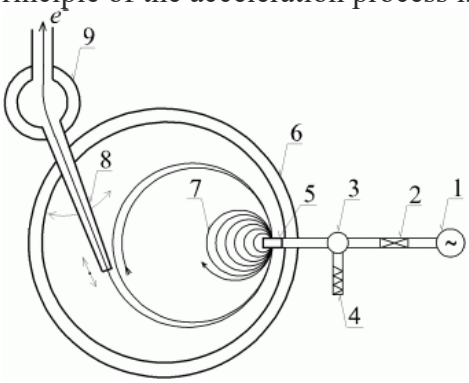


Fig. 1 Schematic layout of the Microtron MT 25 [2]: 1 – magnetron, 2 – phase shifter, 3 – circulator, 4 – water load, 5 – accelarating cavity, 6 – main magnet (vacuum chamber), 7 – electron trajectories, 8 – adjustable beam extractor, 9 – first deflector

The accelerated electron beam is monoenergetic and the output energy of the beam can be adjusted from 6 to 25 MeV with an energy dispersion of around tens of keV. The primary electron beam can be used to irradiate samples or can be converted into bremsstrahlung (photon beam) using a tungsten target. If a target made of a material with a high effective cross-section for the (γ, n) reaction (e.g. lead) is placed behind the W-target, a neutron field is emitted around the target.

The use of the electron beam is suitable for irradiation of DNA origami [3]. These experiments were performed using the electron beam with an energy of 16.5 MeV and absorbed doses in the samples were from 50 Gy to 8 kGy.

Acknowledgment: EISMEA project no. 101115317 (NEO: Next Generation Molecular Data Storage)

References:

[1] Kapitza, S.P., Melekhin, V.N.., *London: Harwood Academic Publishers*, 204, 1978

[2] Krist, P., Bíla, J., *Nuclear Physics Methods an Accelerators in Biology and Medicine 1204*, 183-185, 2009

[3] Ameixa, J., Sala, L. et al., *Chemphyschem 26*, 1, 2025

CONVERTING 2D RECTANGULAR SHEET DNA ORIGAMI TO 3D NANOTUBE DNA ORIGAMI NANO STRUCTURES

Ozra, M., Stevens A.

Department of Chemistry, University of Saskatchewan, Saskatoon, Canada
Amy.stevens@usask.ca

DNA origami has emerged as a powerful and programmable technique for constructing DNA nanostructures [1] with precise control over shape, size, and functionality. Since its introduction by Rothemund in 2006 [2]. DNA origami has facilitated the assembly of a wide array of two-dimensional (2D) and three-dimensional (3D) architectures. This method relies on the folding of a long single-stranded DNA (ssDNA) scaffold into a predefined structure, guided by numerous short ssDNA staple strands that induce controlled crossover interactions. In this study, we report the transformation of rectangular DNA origami sheet into bent sheet (nanotube) structures by manual modification of staple sequences. We applied two methodologies to optimize the yield of the product: 1-simultaneous folding of scaffold (M13Mp18) with main and modified staples in one pot and 2- sequential approach where the rectangular sheet was first formed, followed by the addition of modified staples. We validated the successful folding strategies and evaluated the structural integrity using a combination of Atomic Force Microscopy (AFM), Transmission Electron Microscopy (TEM), and Synchrotron Small-Angle X-ray Scattering (SAXS). SAXS offers a solution-phase, ensemble-averaged characterization that enables the differentiation between 2D rectangular sheets and their 3D nanotube counterparts, providing critical insights into their structural transformations. Our results demonstrate that SAXS is a robust and reliable technique for validating DNA nanotube formation and assessing folding efficiency, complementing AFM and TEM analysis. This study provides a robust framework for the design and construction of complex DNA nanostructures, broadening the scope of DNA origami applications.

References:

[1] N. C. Seeman, *Annual Review of Biophysics and Biomolecular Structure*, vol. 27, pp. 225–248, 1998.

[2] P. W. Rothemund, *Nature*, vol. 440, no.7082, pp. 297–302, 2006.

EXPLORING MOLECULAR PATTERN RECOGNITION VIA MULTIPLEX QPCR

Kotsch F., Hebel A., Bögels B., Simmel F. C.

Department of Bioscience, TUM School of Natural Sciences, Technical University Munich, Am Coulombwall 4a, Munich, D-85748, Germany
franziska.kotsch@tum.de, alexander.hebel@tum.de

Molecular systems for information processing have been explored in various forms, including DNA-based logic circuits [1], strand-displacement cascades [2], and dynamic reaction networks [3]. Recent advancements have demonstrated the potential of integrating machine learning with molecular techniques. For instance, convolutional neural networks (CNNs) have been employed to classify qPCR amplification curves by transforming them into images, enabling the differentiation of various targets [4]. Separately, methods have been developed to encode image-derived feature vectors into synthetic DNA sequences, facilitating similarity searches within DNA data storage systems [5]. Building upon these developments, we propose a distinct approach that encodes digital input features into synthetic DNA sequences and uses multiplex quantitative PCR (qPCR) for biochemical readout. By coupling programmable hybridization with fluorescence detection and machine learning, our system enables molecular pattern recognition and supports flexible, high-dimensional classification within a standard laboratory framework. Pattern recognition is a central task in many domains of science and engineering, and its implementation in molecular systems offers exciting new opportunities for bio-integrated diagnostics, smart therapeutics, and unconventional computing. In the proposed system, photographic images are first compressed into feature vectors using a convolutional neural network. Each feature vector is then encoded into a single short DNA "feature strand." These strands are designed to hybridize with specific "node" strands in a synthetic DNA pool. Upon hybridization, linear-cycle PCR is initiated, amplifying the corresponding node strand. The resulting product is detected at read-out nodes, each equipped for fluorescence-based detection via multiplex qPCR. Instead of mapping each class to a single fluorescence channel, the system allows each amplified node strand to interact with multiple read-out nodes. This multiplexed interaction expands the effective readout space, enabling high-dimensional representation of input features. A machine-learning model interprets the resulting fluorescence patterns for image classification. To expand the system's computational capacity, we aim to implement a network architecture in which amplified products can act as primers for downstream nodes. This would enable indirect signal propagation and the construction of distributed amplification pathways. Short-term memory may be introduced by periodically spiking the reservoir with selected feature strands, which preferentially amplify associated nodes. Such conditioning would allow the system to encode information about previously presented inputs and their temporal order, enabling time-series prediction via memory-like dynamics.

References

[1] Qian, L., & Winfree, E., Scaling up digital circuit computation with DNA strand displacement cascades. *Science*, 332(6034), 1196–120, 2011
[2] Zhang, D. Y., & Seelig, G. Dynamic DNA nanotechnology using strand-displacement reactions. *Nature Chemistry*, 3(2), 103–113, 2011
[3] Chen, Y.-J. et al. Programmable chemical controllers made from DNA. *Nature Nanotechnology*, 8(10), 755–762, 2013
[4] Romana Cavallo, F., et al., A Point-of-Care Device for Fully Automated, Fast and Sensitive Protein Quantification via qPCR. ACS Central Science, 8(7), 931–942, 2022
[5] Bee, Callista, et al., Molecular-level similarity search brings computing to DNA data storage, Nature communications 12.1: 4764, 2021

DNA AS A REWRITABLE STORAGE MEDIUM

Rajesh A.

BioCompute, Delaware (United States) & Bengaluru (India)
chief@biocomputeinc.com

As the field of DNA data storage advances toward commercialization, the high cost and slow turnaround time of DNA synthesis remain critical bottlenecks[1]. Moreover, for DNA storage systems to become comparable to silicon-based memory, they must enable multiple read–write cycles rather than operate as read-once archives.

We focus on a new frontier: rewritable DNA data storage systems integrated with end-to-end hardware solutions. We explore how specially engineered enzymes, whose activity can be modulated by physical stimuli such as light [2], temperature, or electric fields [3], enable controlled, reversible modifications to DNA strands, in order to encode data. These modifications are specific base conversions, epigenetic modifications or structural modifications in DNA that correspond to a specific data encoding and error correction schema.

In addition to rewritability, the DNA has to be sequenced in a manner that enables us to retrieve the sample in an intact manner. Current sequencing devices tend to consume or cleave the DNA sequence in the process of sequencing. This is not conducive to a data storage application. Hence, we are developing solid state nanopores with improved signal to noise ratio to retrieve data.

We also extend this to the design of microfluidic and nanofluidic systems to automate encoding, editing, and retrieval workflows, thus reducing human intervention and operational cost; enabling usto put together a plug-and-play DNA data storage device.

References:

[1] Lin, K.N., Volkel, K., Tuck, J.M. *et al.* Dynamic and scalable DNA-based information storage. *Nat Commun* 11, 2981 (2020). https://doi.org/10.1038/s41467-020-16797-2

[2] Lim, C.K., Yeoh, J.W., Kunartama, A.A. *et al.* A biological camera that captures and stores images directly into DNA. *Nat Commun* 14, 3921 (2023). https://doi.org/10.1038/s41467-023-38876-w

[3] Wang, F., Liu, X., & Willner, I. (2015). DNA switches: From principles to applications. *Angewandte Chemie International Edition*, 54(4), 1098–1129. https://doi.org/10.1002/anie.201404652

DATANA

Enabling Real-World DNA Data Storage

The scientific community already recognizes DNA's potential as a high-density, long-term storage medium. However, until now, the high cost and complexity of DNA synthesis have prevented this potential from being realized in large-scale, real-world applications. The DATANA project is a patented innovation aimed at changing that.

DATANA is a European project focused on building the infrastructure necessary to transform DNA data storage from a fascinating possibility into a usable and scalable solution. The team behind DATANA is developing a novel DNA writer—a device that combines software and microfluidics to efficiently and cost-effectively encode digital data into DNA.

The project's core innovation lies in a proprietary DNA synthesis method implemented in a high-throughput, ultra-low volume liquid handling system. This not only reduces costs but also opens the door for smaller, more agile devices, pushing DNA storage out of centralized facilities and into the hands of commercial users.

DATANA isn't reinventing the field; it's focusing on making it usable, scalable, and accessible. From system integration and error-correcting code to software interfaces and real-world testing, the project is building a full-stack DNA data writing platform that lowers the barrier to entry and accelerates adoption.

By bridging the gap between laboratory demonstrations and deployable technology, the project marks a pivotal shift toward real-world DNA data storage. More than just a technological step forward, DATANA is an enabler of progress, paving the way for the widespread adoption of DNA as a sustainable solution to the world's growing data storage demands.

tilibit®
nanosystems

as global leaders in **DNA nanotechnology**, we deliver precision-engineered nanostructures and high-quality materials that turn ambitious visions into reality. **Partner with our experts** to elevate your innovation and break new scientific ground.

single-stranded scaffold DNA

M13mp18 based

- 4844 bp
- 7249 bp
- 7560 bp + 2 orthogonals (CS3-L + CS4)
- 8064 bp
- 8634 bp
- 48892 bp - coming soon

2D & 3D DNA origami

computer-aided custom design

assembly, purification & precise placement of:

- modifications, e.g., dyes, chemical linkers
- peptides, antibodies, proteins
- nanoparticles (e.g., gold)
- addition of dynamic features

***meet us for your custom discounts: order scaffold and get discounted oligos!**