

In-vitro Molecular Computing Based on DNA Strands

An unconventional computing concept inspired by nature

PD Dr. Thomas Hinze

Brandenburg University of Technology Cottbus – Senftenberg
Institute of Computer Science and Information and Media Technology

`thomas.hinze@b-tu.de`



Brandenburgische
Technische Universität
Cottbus - Senftenberg

DNA as an Excellent Data Storage Medium

high storage
density up to

1 bit / nm³

L.M. Adleman. Molecular Computation of Solutions to Combinatorial Problems. *Science* 266:1021-1024, 1994

DNA as an Excellent Data Storage Medium

high storage
density up to

1 bit / nm³

2 bit per nucleotide or per
base pair in strand of
deoxyribonucleic acid (DNA)

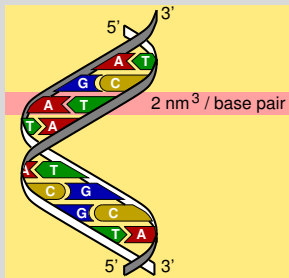


L.M. Adleman. Molecular Computation of Solutions to Combinatorial Problems. *Science* 266:1021-1024, 1994

DNA as an Excellent Data Storage Medium

high storage
density up to

1 bit / nm³



2 bit per nucleotide or per
base pair in strand of
deoxyribonucleic acid (DNA)

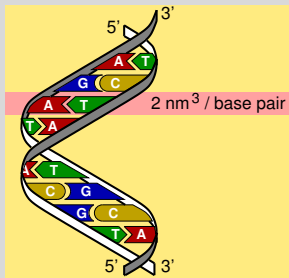


L.M. Adleman. Molecular Computation of Solutions to Combinatorial Problems. *Science* 266:1021-1024, 1994

DNA as an Excellent Data Storage Medium

high storage
density up to

1 bit / nm³



2 bit per nucleotide or per
base pair in strand of
deoxyribonucleic acid (DNA)



electronic
flash memory card
128 Gbit / 146 qmm
corresponds to

0.001 bit / nm³

DNA as an Excellent Data Storage Medium

**high storage
persistence up to
several thousand years**

W. Miller et al. Sequencing the nuclear genome of the extinct woolly mammoth. *Nature* 456:387-391, 2008

DNA as an Excellent Data Storage Medium

**high storage
persistence up to
several thousand years**

under optimal
environmental conditions



W. Miller et al. Sequencing the nuclear genome of the extinct woolly mammoth. *Nature* 456:387-391, 2008

DNA as an Excellent Data Storage Medium

**high storage
persistence up to
several thousand years**

under optimal
environmental conditions



www.zmescience.com

approx. 80% of genome reconstructed
from >20,000 years old mammoth

W. Miller et al. Sequencing the nuclear genome of the extinct woolly mammoth. *Nature* 456:387-391, 2008

DNA as an Excellent Data Storage Medium

**high storage
persistence up to
several thousand years**

under optimal
environmental conditions



www.zmescience.com

approx. 80% of genome reconstructed
from >20,000 years old mammoth



floppy disk:
5 ... 10 years



hard disk:
10 ... 15 years



DVD (expected):
30 ... 50 years

W. Miller et al. Sequencing the nuclear genome of the extinct woolly mammoth. *Nature* 456:387-391, 2008

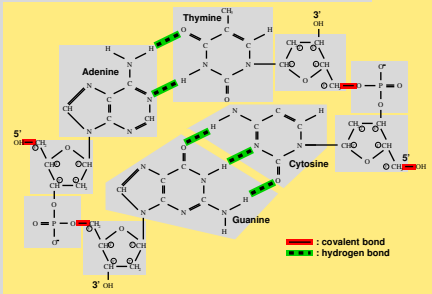
DNA as an Excellent Data Storage Medium

**highly efficient
chemical processing
by low energy consumption**

L. Kari. Arrival of biological mathematics. *The Mathematical Intelligencer* 19(2):9-22, 1997

DNA as an Excellent Data Storage Medium

**highly efficient
chemical processing
by low energy consumption**

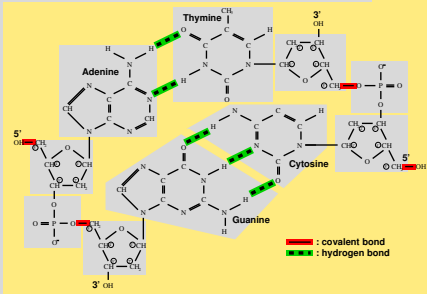


**up to 10^{18}
elementary operations
(break or form
internucleotide
chemical bond)
per Joule**

L. Kari. Arrival of biological mathematics. *The Mathematical Intelligencer* 19(2):9-22, 1997

DNA as an Excellent Data Storage Medium

**highly efficient
chemical processing
by low energy consumption**



**up to 10^{18}
elementary operations
(break or form
internucleotide
chemical bond)
per Joule**

**0 \triangleleft 1 bit set
or reset 10^{16}
electronically**

L. Kari. Arrival of biological mathematics. *The Mathematical Intelligencer* 19(2):9-22, 1997

DNA as an Excellent Data Storage Medium

**massive
data
parallelism**

L.M. Adleman. Molecular Computation of Solutions to Combinatorial Problems. *Science* 266:1021-1024, 1994

DNA as an Excellent Data Storage Medium

**massive
data
parallelism**

up to 10^{15}

DNA strands

per test tube (2ml)

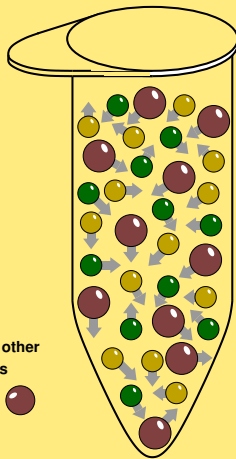
simultaneous

and autonomous

molecular interactions

without central control

**DNA and other
molecules**



L.M. Adleman. Molecular Computation of Solutions to Combinatorial Problems. *Science* 266:1021-1024, 1994

2.

Operations on DNA

Operations on DNA (Selection)

Gaining DNA strands

- **Synthesis** (oligos)
- **Isolation** (like plasmids or genomic DNA from organisms)



Operations on DNA (Selection)

Gaining DNA strands

- **Synthesis** (oligos)
- **Isolation** (like plasmids or genomic DNA from organisms)



Handling DNA solutions

- **Union** (merge)
- **Split** (aliquot)



Operations on DNA (Selection)

Gaining DNA strands

- **Synthesis** (oligos)
- **Isolation** (like plasmids or genomic DNA from organisms)



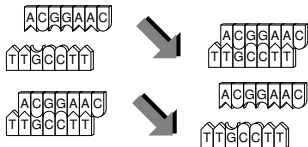
Handling DNA solutions

- **Union** (merge)
- **Split** (aliquot)



Forming and breaking hydrogen bonds

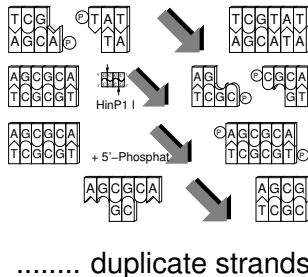
- **Annealing** (hybridisation)
- **Melting** (denaturation)



Operations on DNA (Selection)

Enzymatically catalysed reactions

- **Ligation** (concatenation).....
- **Digestion** (cleavage).....
- **Labelling** (strand end modification)
- **Polymerisation** (completion).....
- **PCR** (polymerase chain reaction)...



Operations on DNA (Selection)

Enzymatically catalysed reactions

- **Ligation** (concatenation).....
- **Digestion** (cleavage).....
- **Labelling** (strand end modification)
- **Polymerisation** (completion).....
- **PCR** (polymerase chain reaction)...



..... duplicate strands

Separation and analysis of DNA strands

- **Affinity purification** (sep. by biotin)...
- **Gel electrophoresis** (sep. by length)
- **Sequencing** (readout).....



sort and detect strands



3.

Algorithm for Solution to the NP-hard Knapsack Problem

Knapsack Problem

NP-hard decision problem, exponential need of resources

Definition

Given n natural numbers a_1, \dots, a_n and reference number $b \in \mathbb{N}$

Is there a subset $I \subseteq \{1, \dots, n\}$ with $\sum_{i \in I} a_i = b$?

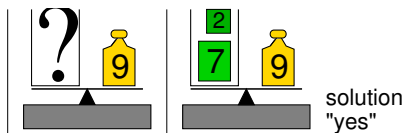
Explanation

a_1, \dots, a_n : **weights** of **objects** 1, ..., n .

Is there a possibility to pack a selection of these objects into the knapsack which exactly meets the reference weight b ?

Example

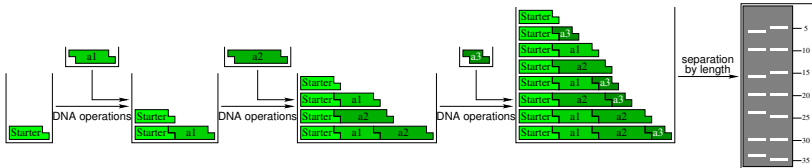
$a_1 = 5$	5	object 1
$a_2 = 7$	7	object 2
$a_3 = 2$	2	object 3
$b = 9$	2	object 3



Idea of Wetware Solution to Knapsack Problem

Brute force approach

- Encode a_1, \dots, a_n into DNA double strands by length ($c \cdot a_i$)
- Generate all solution candidates by a controlled **split-and-combine strategy**
- Separate final DNA pool by length
- Detect DNA at $\text{Starter length} + c \cdot b$ and answer yes

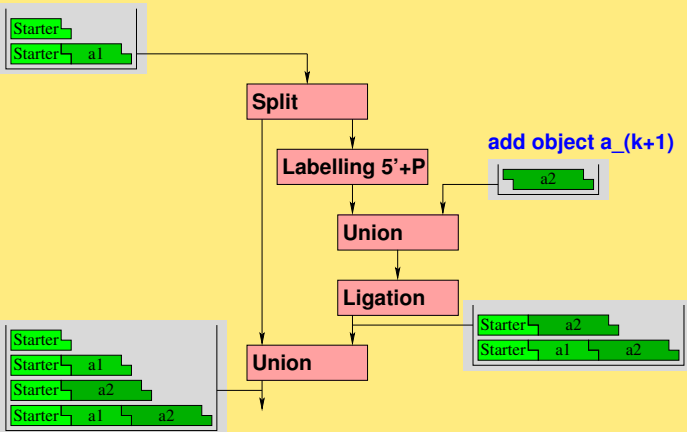


M. Sturm, T. Hinze. Verfahren zur Ausführung von mathematischen Operationen mittels eines DNA-Computers und DNA-Computer hierzu. Deutsches Patent DE 101 59 886 B4, IPC G06N 3/00, erteilt 2010

Split-and-combine Strategy

doubles number of combinations by addition of one object

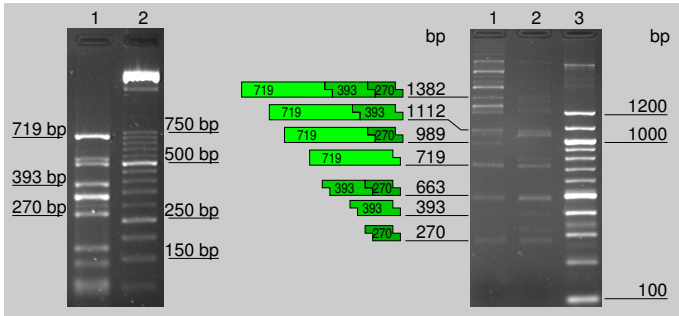
**DNA pool composed of combinations from k objects
start with $k = 0$ (starters only)**



DNA pool composed of combinations from $k+1$ objects

Simple Implementation of a Problem Instance

- $n = 3$ objects taken from plasmid (pQE30 cleaved with PvuII)
- $a_1 = 719$, $a_2 = 393$, $a_3 = 270$, $b = 1112$, $c = 1$
- Exponential need of resources moved from time to space
- Final sequencing of DNA band corresponding to b reveals “yes”
- Limited scalability of the algorithm due to **side effects** and amount of DNA



Side Effects and Perturbations

prevent DNA operations from running in an ideal manner

- Loss of DNA
- Incomplete reactions
- Non-specificities
- Malformed DNA (artefacts)
- DNA damage
- Contaminations or impurities of DNA solutions
- ... (many others)

Coping with side effects is a *hard challenge* to overcome in practical *in-vitro* DNA computation. Assuming an *error rate* of **5%** per operation and having a sequence of **10** operations, the *overall success rate* is merely $0.95^{10} \cdot 100 \approx 60\%$.

4.

Further Applications, Trends, and Perspectives

Milestones of DNA-based Computing

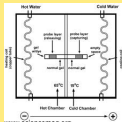
- Pioneering era after Adleman's experiment
- Refinement and improvement of techniques and encoding schemes complemented by much theoretical work



www.usc.edu

Adleman's
DNA-based solution to
Hamiltonian path problem
($n = 7$ nodes)

1994



www.sciencemag.org

DNA-based solution to
satisfiability problem
by brute force approach
($n = 20$ variables)

2002

Milestones of DNA-based Computing

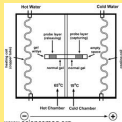
- Pioneering era after Adleman's experiment
- Refinement and improvement of techniques and encoding schemes complemented by much theoretical work
- *Addressable DNA-based memory* able to store data from files



www.usc.edu

Adleman's
DNA-based solution to
Hamiltonian path problem
($n = 7$ nodes)

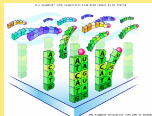
1994



www.sciencemag.org

DNA-based solution to
satisfiability problem
by brute force approach
($n = 20$ variables)

2002



www.newscientist.com

DNA chips for
massive parallel computing
by string matching
approx. 40,000 spots

2010

Milestones of DNA-based Computing

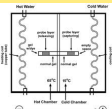
- Pioneering era after Adleman's experiment
- Refinement and improvement of techniques and encoding schemes complemented by much theoretical work
- *Addressable DNA-based memory* able to store data from files
- Computing by DNA self-assembly promising clue towards *freely programmable nanomachines*



www.usc.edu

Adleman's
DNA-based solution to
Hamiltonian path problem
($n = 7$ nodes)

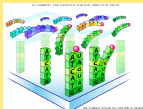
1994



www.sciencemag.org

DNA-based solution to
satisfiability problem
by brute force approach
($n = 20$ variables)

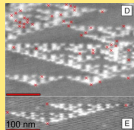
2002



www.newscientist.com

DNA chips for
massive parallel computing
by string matching
approx. 40,000 spots

2010



100 nm
PLOS Biology

algorithmic self-assembly
for cellular automaton and
universal DNA computer

2014

International Molecular Computing Community



≈ 500 researchers worldwide, conference series like CMC, DNA, UC, . . .

Take Home Message

Living organisms comprise almost perfect DNA-based computers. We are going to learn and to adapt the underlying principles for utilisation *in vitro*. There are first successes but there is still a lot of work to do.

Take Home Message

Living organisms comprise almost perfect DNA-based computers. We are going to learn and to adapt the underlying principles for utilisation *in vitro*. There are first successes but there is still a lot of work to do.

Further and more detailed information



- T. Hinze, M. Sturm. Rechnen mit DNA - Eine Einführung in Theorie und Praxis. De Gruyter, eBook, 2014
- T. Hinze. Computer der Natur. bookboon.com, eBook (for free), 2013