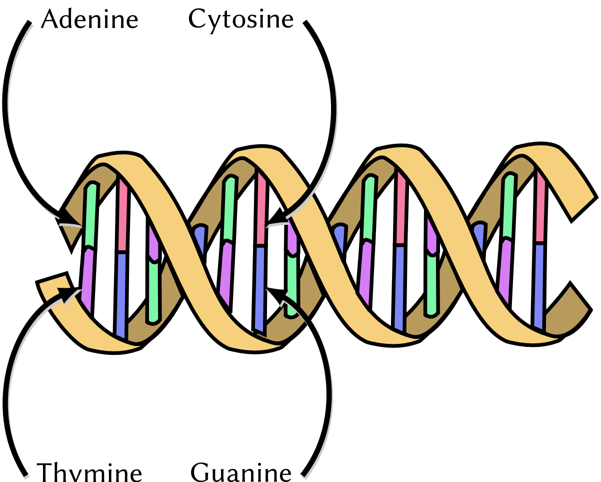
Toehold DNA Languages are Regular

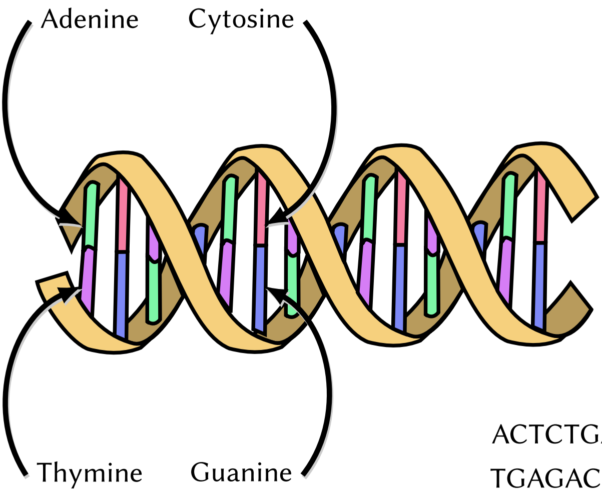
Sebastian Brandt *Nicolas Mattia* ***Jochen Seidel*** *Roger Wattenhofer*

ETH Zurich – Distributed Computing Group – www.disco.ethz.ch

DNA



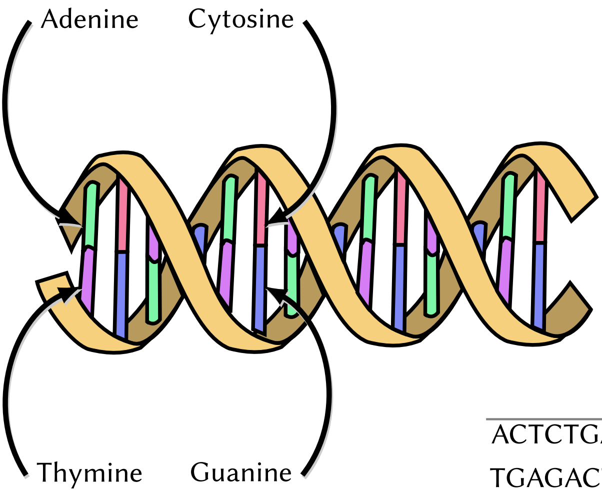
DNA



ACTCTGAGACTCTG

TGAGACTCTGAGAC

DNA



ACTCTGAGACTCTG
TGAGACTCTGAGAC

Domains & Complements

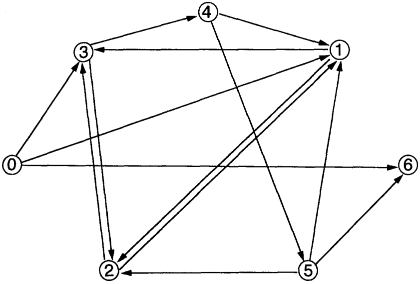
$$\overrightarrow{\mathbf{a}} \equiv \overrightarrow{\mathrm{ATCG}} \qquad \overrightarrow{\bar{\mathbf{a}}} \equiv \overrightarrow{\overline{\mathrm{ATCG}}}$$

$$\overrightarrow{\mathbf{b}} \equiv \overrightarrow{\mathrm{ATT}} \qquad \overrightarrow{\bar{\mathbf{b}}} \equiv \overrightarrow{\overline{\mathrm{ATT}}}$$

$$\overrightarrow{\mathbf{c}} \equiv \overrightarrow{\mathrm{CTC}} \qquad \overrightarrow{\bar{\mathbf{c}}} \equiv \overrightarrow{\overline{\mathrm{CTC}}}$$

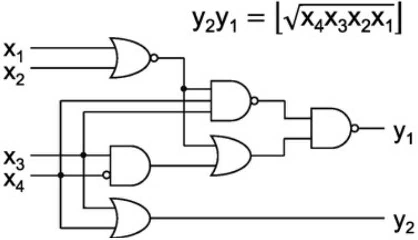
$$\begin{matrix} \overrightarrow{\mathbf{abc}} \\ \underleftarrow{\bar{\mathbf{a}}\bar{\mathbf{b}}\bar{\mathbf{c}}} \end{matrix} \equiv \begin{matrix} \overrightarrow{\mathrm{ATCGATTCTC}} \\ \underleftarrow{\mathrm{TAGCTAAGAG}} \end{matrix}$$

Hamiltonian Path



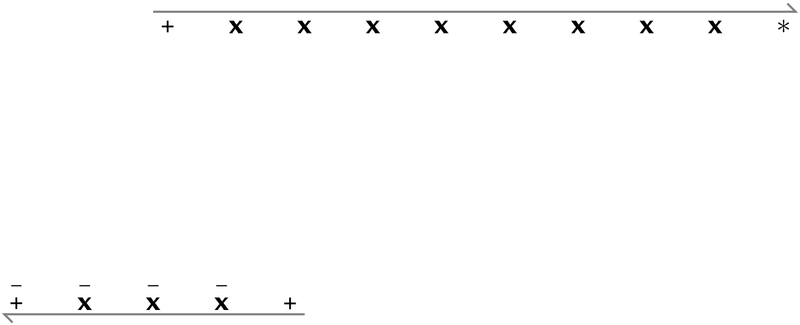
Adleman 1994

Square Root

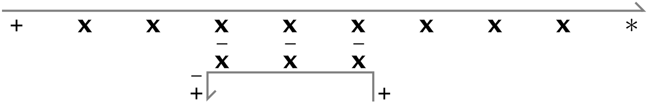


Qian, Winfree 2011

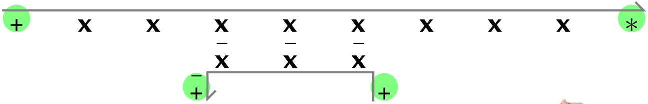
Strand Binding



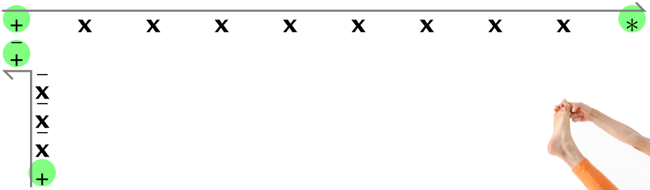
Strand Binding



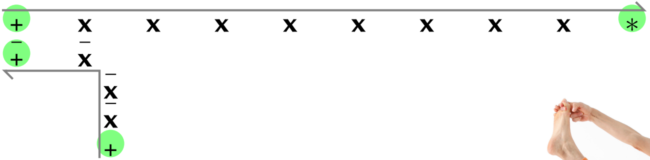
Strand Binding, Toeholds



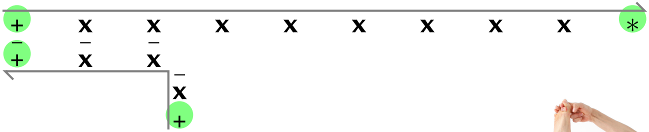
Strand Binding, Toeholds



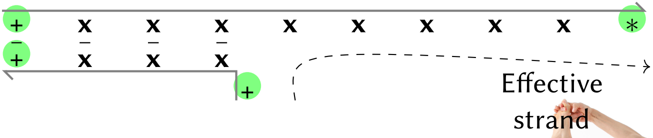
Strand Binding, Toeholds



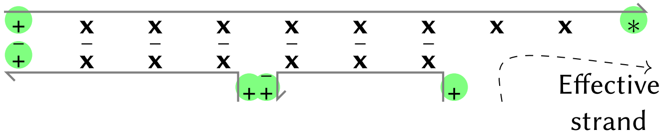
Strand Binding, Toeholds



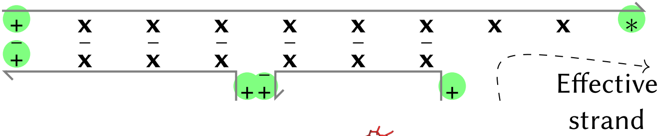
Strand Binding, Toeholds, Effective Strands

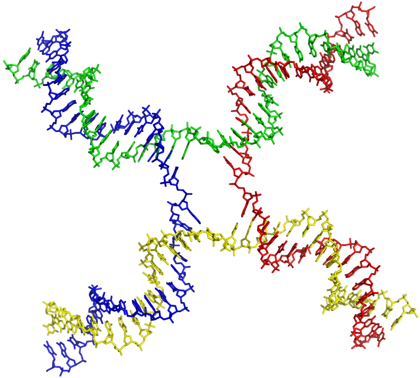


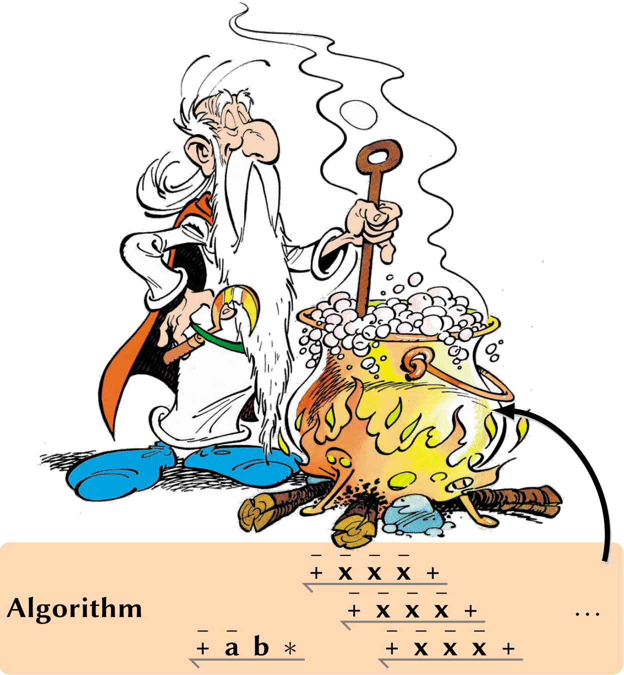
Strand Binding, Toeholds, Effective Strands



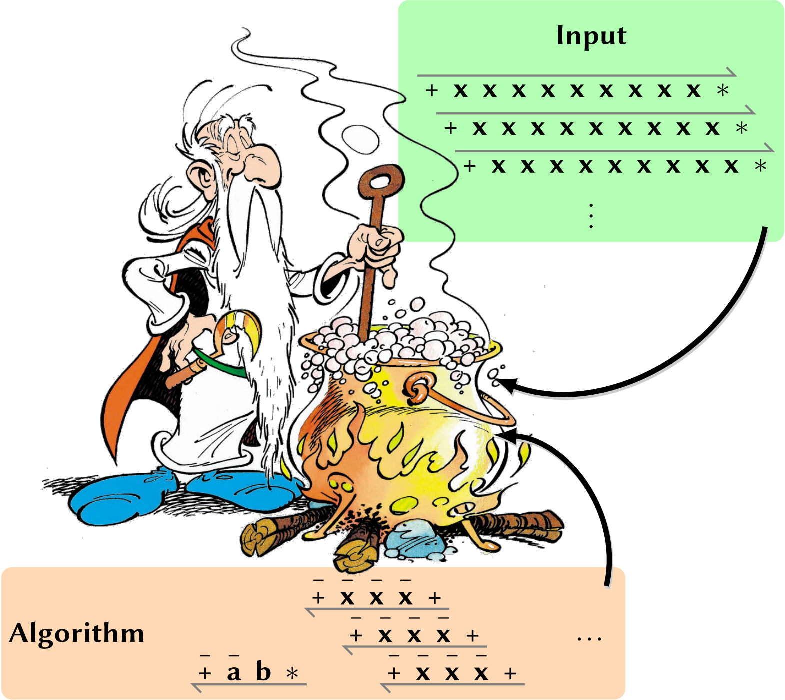
Strand Binding, Toeholds, Effective Strands

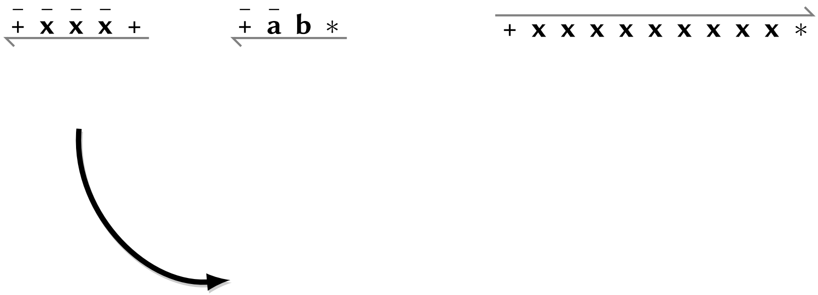




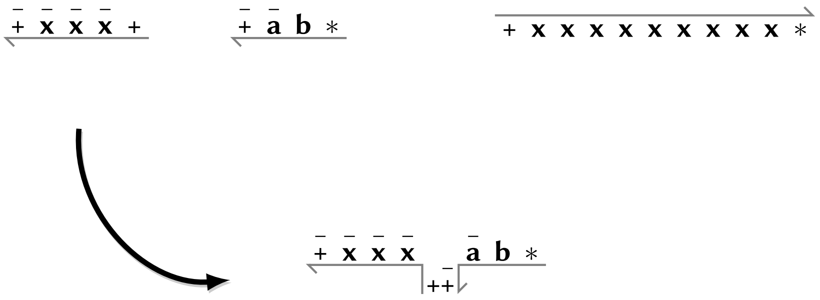


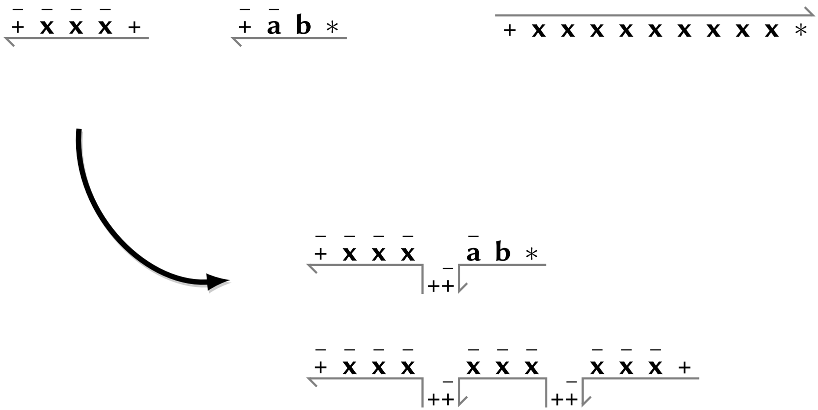
Algorithm
+ x x x +
+ x x x +
+ x x x +
+ a b ∗
...


Input
+ x x x x x x x x x *
+ x x x x x x x x x *
+ x x x x x x x x x *
⋮
Algorithm
+ x x x +
+ x x x +
+ a b *
+ x x x +
...

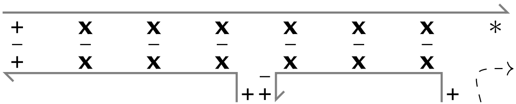


+ x x x +
+ a b *
+ x x x x x x x x x *


+ x x x +
+ a b *
+ x x x x x x x x x *
+ x x x
++
a b *

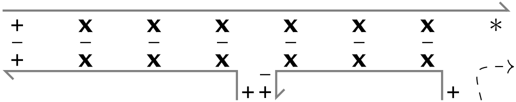


+ x x x +
+ a b *
+ x x x x x x x x x *
+ x x x
++
a b *
+ x x x
++
x x x
++
x x x +

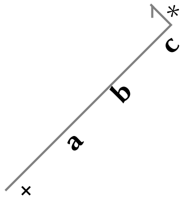


+ x x x x x x *
+ x x x x x x
+ + +

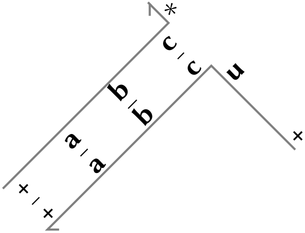

Yes!



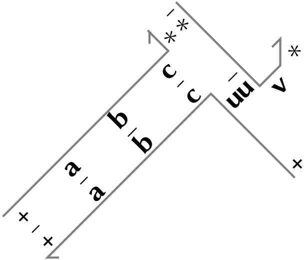
Substitution: +abc$*$ $\rightarrow$ +z$*$



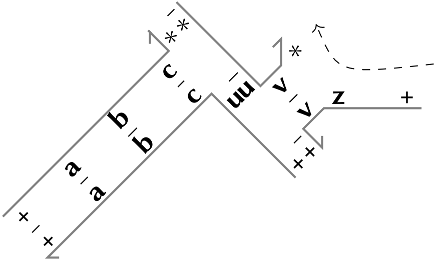
Substitution: +abc$*$ $\rightarrow$ +z$*$



Substitution: +abc$*$ $\rightarrow$ +z$*$

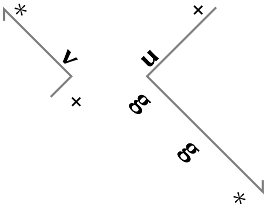


Substitution: +abc$*$ $\rightarrow$ +z$*$



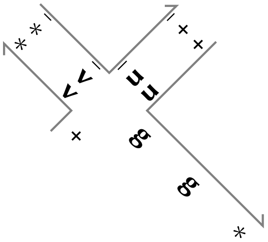
Aggregation: $+\mathbf{abc}* \wedge +\mathbf{def}* \rightarrow +\mathbf{gg}*$

1. $+\mathbf{abc}* \rightarrow +\mathbf{ugg}*$
2. $+\mathbf{def}* \rightarrow +\mathbf{v}*$
3. $\overline{*\mathbf{vu}+}$



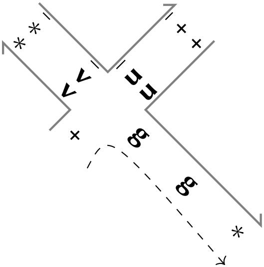
Aggregation: $+\mathbf{abc}* \wedge +\mathbf{def}* \rightarrow +\mathbf{gg}*$

1. $+\mathbf{abc}* \rightarrow +\mathbf{ugg}*$
2. $+\mathbf{def}* \rightarrow +\mathbf{v}*$
3. $\overline{*}\overline{\mathbf{v}}\overline{\mathbf{u}}\overline{+}$



Aggregation: $+\mathbf{abc}* \wedge +\mathbf{def}* \rightarrow +\mathbf{gg}*$

1. $+\mathbf{abc}* \rightarrow +\mathbf{ugg}*$
2. $+\mathbf{def}* \rightarrow +\mathbf{v}*$
3. $\overline{*\mathbf{vu}+}$



Primality

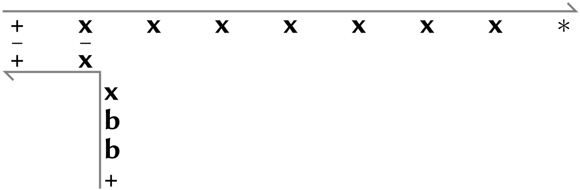
For $2 \leq j \leq \sqrt{n}$ and $k \leq j - 1$:

1. $\overrightarrow{+\mathbf{b}^j\mathbf{x}\bar{\mathbf{x}}\bar{+}}$
2. $\overrightarrow{+\mathbf{b}^j\bar{\mathbf{x}}^j\bar{\mathbf{b}}^j\bar{+}}$
3. $\overrightarrow{+\mathbf{b}^j\mathbf{x}^k*} \rightarrow \overrightarrow{+\mathbf{b}^j\mathbf{f}*}$
4. $\overrightarrow{+\mathbf{b}^2\mathbf{f}*} \wedge \overrightarrow{+\mathbf{b}^3\mathbf{f}*} \wedge \cdots \wedge \overrightarrow{+\mathbf{b}^{\sqrt{n}}\mathbf{f}*} \rightarrow \overrightarrow{+*}$

Primality

For $2 \leq j \leq \sqrt{n}$ and $k \leq j - 1$:

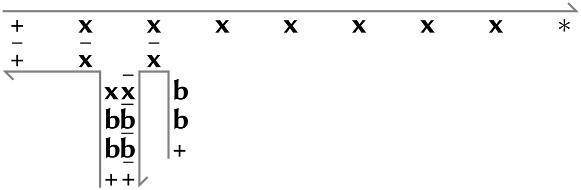
1. $\overrightarrow{+\mathbf{b}^{j}\mathbf{x}\bar{\mathbf{x}}\bar{+}}$
2. $\overrightarrow{+\mathbf{b}^{j}\bar{\mathbf{x}}^{j}\bar{\mathbf{b}}^{j}\bar{+}}$
3. $\overrightarrow{+\mathbf{b}^{j}\mathbf{x}^{k}*} \rightarrow \overrightarrow{+\mathbf{b}^{j}\mathbf{f}*}$
4. $\overrightarrow{+\mathbf{b}^{2}\mathbf{f}*} \wedge \overrightarrow{+\mathbf{b}^{3}\mathbf{f}*} \wedge \cdots \wedge \overrightarrow{+\mathbf{b}^{\sqrt{n}}\mathbf{f}*} \rightarrow \overrightarrow{+*}$



Primality

For $2 \leq j \leq \sqrt{n}$ and $k \leq j - 1$:

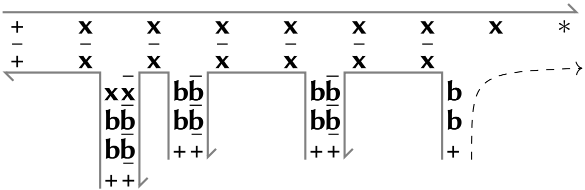
1. $\overrightarrow{+\mathbf{b}^j\mathbf{x}\bar{\mathbf{x}}\bar{+}}$
2. $\overrightarrow{+\mathbf{b}^j\bar{\mathbf{x}}^j\bar{\mathbf{b}}^j\bar{+}}$
3. $\overrightarrow{+\mathbf{b}^j\mathbf{x}^k*} \to \overrightarrow{+\mathbf{b}^j\mathbf{f}*}$
4. $\overrightarrow{+\mathbf{b}^2\mathbf{f}*} \wedge \overrightarrow{+\mathbf{b}^3\mathbf{f}*} \wedge \cdots \wedge \overrightarrow{+\mathbf{b}^{\sqrt{n}}\mathbf{f}*} \to \overrightarrow{+*}$



Primality

For $2 \leq j \leq \sqrt{n}$ and $k \leq j - 1$:

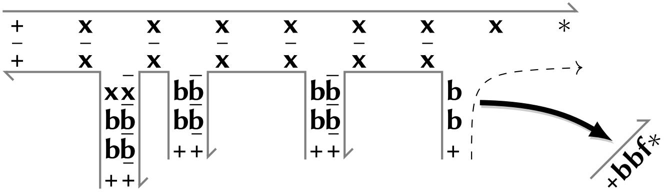
1. $\overrightarrow{+\mathbf{b}^j\mathbf{x}\bar{\mathbf{x}}\bar{+}}$
2. $\overrightarrow{+\mathbf{b}^j\bar{\mathbf{x}}^j\bar{\mathbf{b}}^j\bar{+}}$
3. $\overrightarrow{+\mathbf{b}^j\mathbf{x}^k*} \rightarrow \overrightarrow{+\mathbf{b}^j\mathbf{f}*}$
4. $\overrightarrow{+\mathbf{b}^2\mathbf{f}*} \wedge \overrightarrow{+\mathbf{b}^3\mathbf{f}*} \wedge \cdots \wedge \overrightarrow{+\mathbf{b}^{\sqrt{n}}\mathbf{f}*} \rightarrow \overrightarrow{+*}$



Primality

For $2 \leq j \leq \sqrt{n}$ and $k \leq j - 1$:

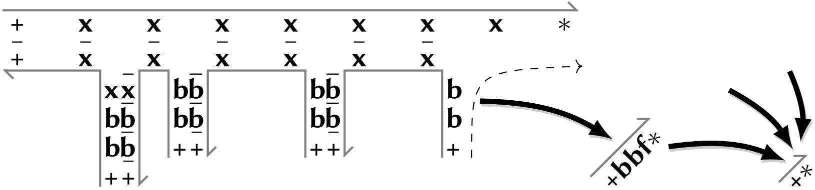
1. $\overrightarrow{+\mathbf{b}^j\mathbf{x}\bar{\mathbf{x}}\bar{+}}$
2. $\overrightarrow{+\mathbf{b}^j\bar{\mathbf{x}}^j\bar{\mathbf{b}}^j\bar{+}}$
3. $\overrightarrow{+\mathbf{b}^j\mathbf{x}^k*} \rightarrow \overrightarrow{+\mathbf{b}^j\mathbf{f}*}$
4. $\overrightarrow{+\mathbf{b}^2\mathbf{f}*} \wedge \overrightarrow{+\mathbf{b}^3\mathbf{f}*} \wedge \cdots \wedge \overrightarrow{+\mathbf{b}^{\sqrt{n}}\mathbf{f}*} \rightarrow \overrightarrow{+*}$



Primality

For $2 \leq j \leq \sqrt{n}$ and $k \leq j - 1$:

1. $\overrightarrow{+\mathbf{b}^j\mathbf{x}\bar{\mathbf{x}}\bar{+}}$
2. $\overrightarrow{+\mathbf{b}^j\bar{\mathbf{x}}^j\bar{\mathbf{b}}^j\bar{+}}$
3. $\overrightarrow{+\mathbf{b}^j\mathbf{x}^k*} \rightarrow \overrightarrow{+\mathbf{b}^j\mathbf{f}*}$
4. $\overrightarrow{+\mathbf{b}^2\mathbf{f}*} \wedge \overrightarrow{+\mathbf{b}^3\mathbf{f}*} \wedge \cdots \wedge \overrightarrow{+\mathbf{b}^{\sqrt{n}}\mathbf{f}*} \rightarrow \overrightarrow{+*}$



Primality

For $2 \leq j \leq \sqrt{n}$ and $k \leq j - 1$:

1. $\overrightarrow{+\mathbf{b}^j\mathbf{x}\bar{\mathbf{x}}\bar{+}}$
2. $\overrightarrow{+\mathbf{b}^j\bar{\mathbf{x}}^j\bar{\mathbf{b}}^j\bar{+}}$
3. $\overrightarrow{+\mathbf{b}^j\mathbf{x}^k*} \rightarrow \overrightarrow{+\mathbf{b}^j\mathbf{f}*}$
4. $\overrightarrow{+\mathbf{b}^2\mathbf{f}*} \wedge \overrightarrow{+\mathbf{b}^3\mathbf{f}*} \wedge \cdots \wedge \overrightarrow{+\mathbf{b}^{\sqrt{n}}\mathbf{f}*} \rightarrow \overrightarrow{+*}$

- $O(n)$ strands
- Binary Encoding: $O(\sqrt{n} \cdot \log n)$ strands
- Good: $O(\sqrt{n} \cdot \log n) \subsetneq O(\frac{n}{\log n})$

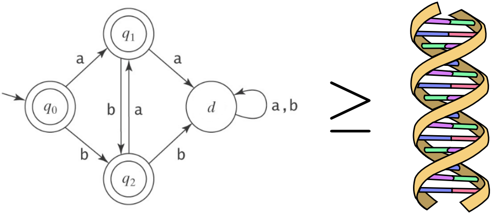
Primality

For $2 \leq j \leq \sqrt{n}$ and $k \leq j-1$:

1. $\overrightarrow{+\mathbf{b}^j\mathbf{x}\bar{\mathbf{x}}\bar{+}}$
2. $\overrightarrow{+\mathbf{b}^j\bar{\mathbf{x}}^j\bar{\mathbf{b}}^j\bar{+}}$
3. $\overrightarrow{+\mathbf{b}^j\mathbf{x}^k*} \rightarrow \overrightarrow{+\mathbf{b}^j\mathbf{f}*}$
4. $\overrightarrow{+\mathbf{b}^2\mathbf{f}*} \wedge \overrightarrow{+\mathbf{b}^3\mathbf{f}*} \wedge \cdots \wedge \overrightarrow{+\mathbf{b}^{\sqrt{n}}\mathbf{f}*} \rightarrow \overrightarrow{+*}$

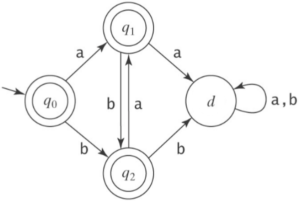
- $O(n)$ strands
- Binary Encoding: $O(\sqrt{n} \cdot \log n)$ strands
- Good: $O(\sqrt{n} \cdot \log n) \subsetneq O(\frac{n}{\log n})$

Constant Size Algorithms?

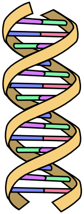


Theorem

Let $\mathcal{A}$ be a DNA algorithm.
If $\mathcal{A}$, interpreted as a set of strands, has finite size,
then $\mathcal{A}$ decides a regular language.



$\geq$



Theorem

Let $\mathcal{A}$ be a DNA algorithm.
If $\mathcal{A}$, interpreted as a set of strands, is a regular language,
then $\mathcal{A}$ decides a regular language.

Basic idea: Define language $\mathcal{I}$ such that

1. $\mathcal{I}$ is regular,
2. every $\iota \in \mathcal{I}$ is accepted by $\mathcal{A}$, and
3. every ι accepted by $\mathcal{A}$ is in $\mathcal{I}$.

The Language $\mathcal{I}$ ...

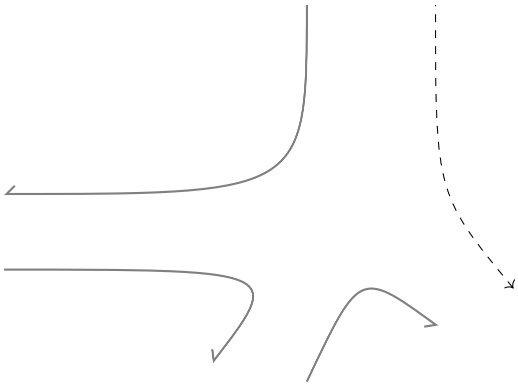
$$\mathcal{I} := \bigcup_{w:\delta(w,*)\in F} \mathcal{X}(\delta(s,+), w, Q \times Q \setminus \{(\delta(s,+), w)\}),$$

$$\mathcal{X}(v,w,J) := \bigcup_{q\in I(v,w,J)} \left(\bigcap_{\substack{(q_i,q_{i+1}):\\ q_i\neq NULL\neq q_{i+1}}} C(q_i, q_{i+1}, J) \ \cap \bigcap_{i:q_i=NULL} D(q_{i-1}, q_{i+1}, J) \right)$$

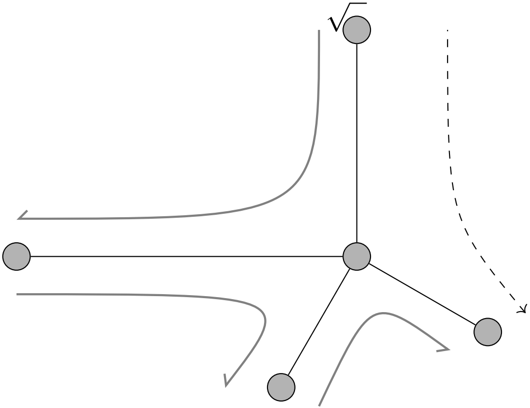
$$C(x,y,J) := \begin{cases} +\Sigma^* *, & \text{if } \exists (z_1,z_2) \in H_{x,y,J} \text{ s.t. } z_2 = s \text{ and } z_1 \in F \\ \bigcup_{(z_1,z_2)\in H_{x,y,J}} \mathcal{X}(z_1, z_2, J \setminus \{(x,y),(z_1,z_2)\}), & \text{otherwise.} \end{cases}$$

$$D(x,y,J) := \bigcup_{K\subseteq J\setminus\{(x,y)\}} \left(\overline{\mathcal{L}(B_{K,x,y})} \cap \bigcap_{(z_1,z_2)\in K} \mathcal{X}(z_1, z_2, J \setminus \{(x,y),(z_1,z_2)\}) \right)$$

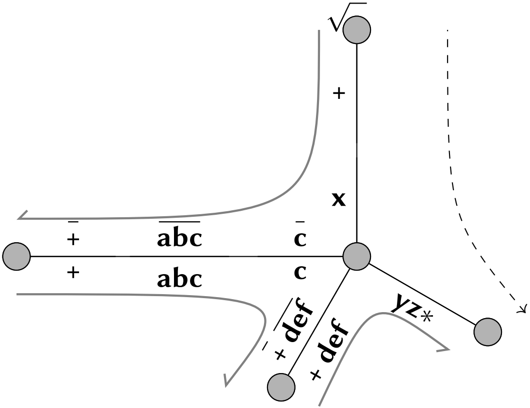
Assemblies



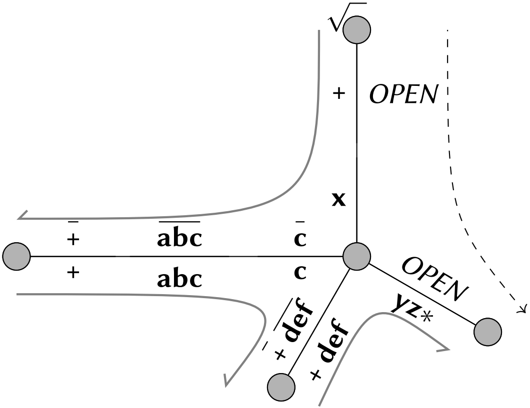
Assemblies



Assemblies

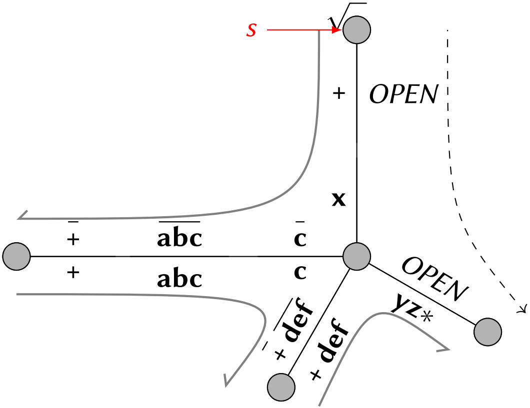


Assemblies



+
OPEN
x
$\overline{+}$
$\overline{\mathbf{abc}}$
$\bar{\mathbf{c}}$
+
abc
c
OPEN
yz*
$\overline{+}\ \overline{\mathbf{def}}$
+ def

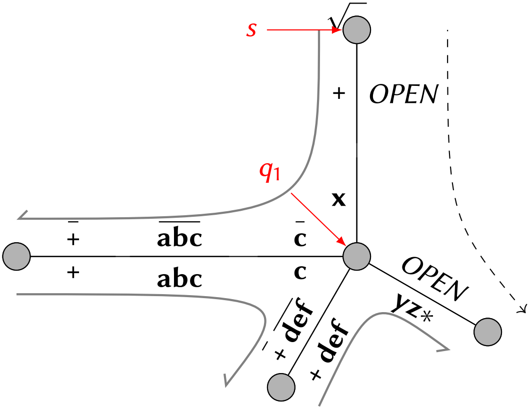
Assemblies & Automata



DNA Algorithm ≈ Finite Automaton

$(Q, \Sigma, \delta, s, F)$

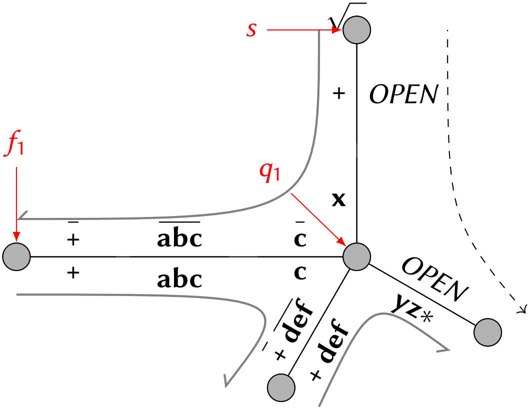
Assemblies & Automata



DNA Algorithm ≈ Finite Automaton

$(Q, \Sigma, \delta, s, F)$

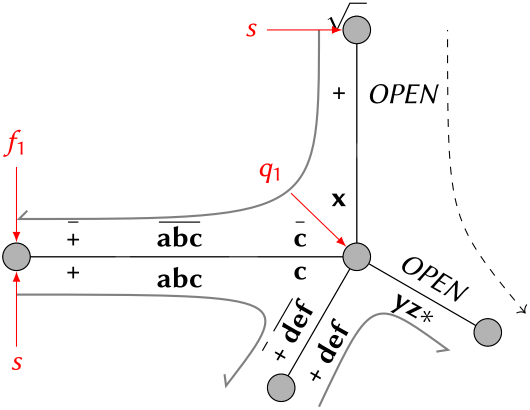
Assemblies & Automata



DNA Algorithm ≈ Finite Automaton

$(Q, \Sigma, \delta, s, F)$

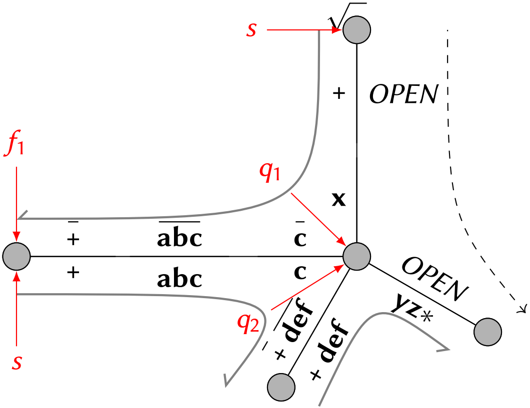
Assemblies & Automata



DNA Algorithm ≈ Finite Automaton

$(Q, \Sigma, \delta, s, F)$

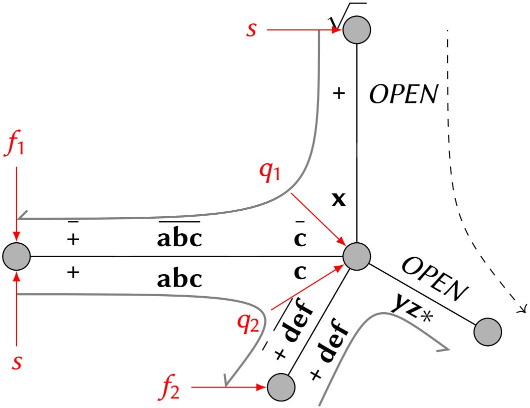
Assemblies & Automata



DNA Algorithm ≈ Finite Automaton

$(Q, \Sigma, \delta, s, F)$

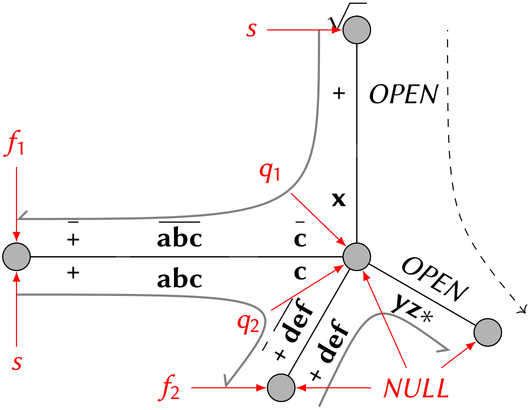
Assemblies & Automata



DNA Algorithm ≈ Finite Automaton

$(Q, \Sigma, \delta, s, F)$

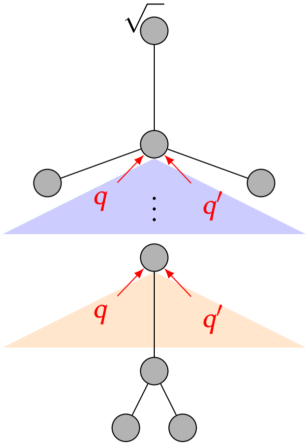
Assemblies & Automata



DNA Algorithm ≈ Finite Automaton

$(Q, \Sigma, \delta, s, F)$

Minimal Assemblies



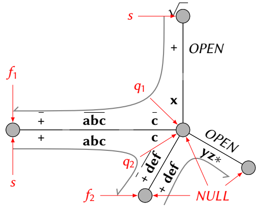
The Laborious Part . . .

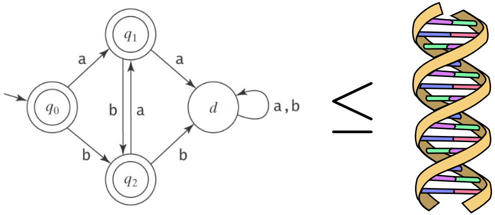
$$\mathcal{I} := \bigcup_{w:\delta(w,*)\in F} \mathcal{X}(\delta(s,+), w, Q\times Q \setminus \{(\delta(s,+), w)\})$$

$$\mathcal{X}(v,w,J) := \bigcup_{q\in I(v,w,J)} \left(\bigcap_{\substack{(q_i,q_{i+1}):\\ q_i\neq NULL\neq q_{i+1}}} C(q_i, q_{i+1}, J) \;\cap \bigcap_{i:q_i=NULL} D(q_{i-1}, q_{i+1}, J) \right)$$

$$C(x,y,J) := \begin{cases} +\Sigma^* * , & \text{if } \exists (z_1,z_2)\in H_{x,y,J} \text{ s.t. } z_2 = s \text{ and } z_1 \in F \\ \bigcup_{(z_1,z_2)\in H_{x,y,J}} \mathcal{X}(z_1, z_2, J\setminus\{(x,y),(z_1,z_2)\}) , & \text{otherwise.} \end{cases}$$

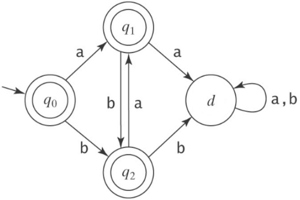
$$D(x,y,J) := \bigcup_{K\subseteq J\setminus\{(x,y)\}} \left(\overline{\mathcal{L}(B_{K,x,y})} \cap \bigcap_{(z_1,z_2)\in K} \mathcal{X}(z_1,z_2,J\setminus\{(x,y),(z_1,z_2)\}) \right)$$



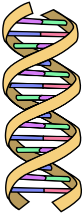


Theorem

Let $\mathcal{L}$ be a regular language.
There is a constant size DNA algorithm that decides $+\mathcal{L}*$.



$\leq$



Theorem

Let $\mathcal{L}$ be a regular language.
There is a constant size DNA algorithm that decides $+\mathcal{L}*$.

- Regular Language $\rightarrow$ FA $(Q, \Gamma, \delta, s, F)$
- $\overrightarrow{+\bar{s}+}$ for the starting state s
- $\overrightarrow{+q'\bar{x}\bar{q}+}$ for all possible transitions $\delta(q, x) = q'$
- $\overrightarrow{\bar{*}\bar{f}*}$ for all $f \in F$.

Summary