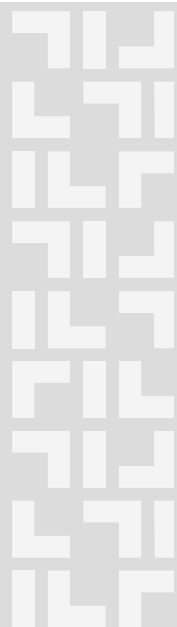




Univerza v Ljubljani

Fakulteta  
za računalništvo  
in informatiko



## 4. DNK procesiranje (DNA computing)

Vsebina 4.poglavja predavanj (4.UNI/RS)

Avtor: Izr.prof.dr. Miha Mraz

Šol.letu: 2010/2011



## 4.1. Uvod

- DNK – deoksiribonukleinska kislina (angl. *Deoxyribonucleic acid* – DNA)
- Namesto klasičnih polprevodniških tehnologij (silicijeva osnova) za gradnjo rač.sistemov uporabljamo DNK
- Temelji na biokemiji in molekularni biologiji (angl. *molecular computing*)
- DNK: v njej je “shranjena” genska zasnova živih organizmov (angl. *blue print*)

## 4.2. Koncept pomnjenja v DNK

- Klasično računalniško pomnjenje: niz binarnih podatkov kodiran z sekvenco 0 in 1
- DNK pomnjenje: DNK je “kodirana” – sestavljena iz poljubno dolge sekvence štirih baz A,C,T in G (baze ali nukleotidi adenin, gvanin, citozin in timin) [1], ki tvorijo dve vijačnici (BluePrint)
- Enote baz nanizane na pribl. 0,35 nm  $\Rightarrow$  gostota “podatkov” je velikostnega reda 18Mbits/inch
- Če vrednotimo omenjeno gostoto na površini dobimo velikostni red gostote  $10^6$  Gbits/sq.inch (današnje HD tehnologije: pribl.10Gbits/sq.inch)
- Prednost: IZREDNO VISOKA GOSTOTA ZAPISA





- Primer zapisa v DNK: S=ATTACGTCG
- Komplementarni pari A-T in C-G
- Pojem komplementa S: S'=TAATGCAGC
- S in S' tvorita dvojno vijačnico
- Ker imamo venomer shranjen podatek (S) in njegov komplement (S') imamo na razpolago enostaven mehanizem za DETEKCIJO NAPAK
- Vzroki pojavitve napak v DNA:
  - Napake pri sestavi DNA (opravljajo jo encimi) – verjetnost napak  $p < 10^{-9}$
  - Prisotnost UV sevanja, termične energije,...



- Osnovni manipulatorji nad DNK so encimi:
  - Sestavljanje DNK
  - Rezanje DNK
  - Kopiranje DNK
  - Popravljanje DNK
- Vse našteje operacije smatramo kot osnovne operacije procesiranja (ekvivalent dvojiškimi logičnimi operacijam)
- “Procesiranje” vršijo encimi
- Paralelizem: nad DNK deluje več encimov nad različnimi segmenti vijačnic
- Osnovna ideja DNK procesiranja:
  - DNK dvojna vijačnica: pomnjenje
  - Delo encimov nad vijačnico: procesiranje (bližanje trendu samosestavljanja (angl. *self-assembly*))



- *gen*: del DNK vijačnice – enota dedovanja
- *genom*: skupek vseh genov organizma (celoten dedni zapis)
- človeški genom: 20.000 – 25.000 genov (2009)
- dolžina celotne človeške vijačnice:  $3 \cdot 10^9$  BP (*baznih parov*) (2009)

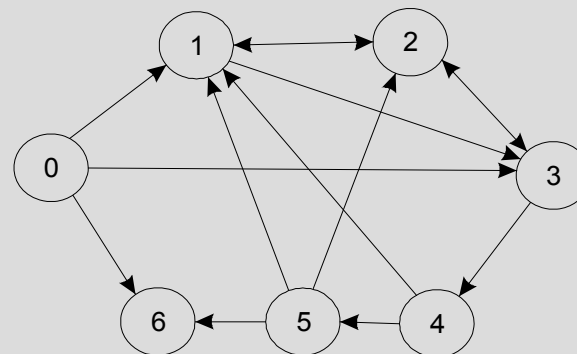
## 4.3 Adlemanov eksperiment kot zgled DNK procesiranja

- 1994: potrditev ideje (angl. *proof of the concept*), da je mogoče z visokim paralelizmom procesiranja nad DNK reševati NP-polne probleme
- Primer reševanega problema:
  - Hamiltonove poti po usmerjenem grafu
  - Imamo  $n$  vozlišč, med katerimi vodi poljubno število eno ali dvosmernih povezav;
  - Problem: ali je ob danem izhodiščnem in ciljnem vozlišču možno prepotovati vsa vozlišča natanko enkrat;
- Ni nujno, da rešitev obstaja, če pa obstaja, kompleksnost njenega iskanja strmo raste z višanjem števila mest





- Konkretni problem ob robnih pogojih, ki ga rešuje Adleman, predvideva da se pot začne v 0 in konča v 6; rešitev predstavlja sekvenca obiskov 0, 1, 2, 3, 4, 5, 6;





- Ideja za reševanje problema: prostorska potratnost naj bi nadomestila procesno;
- Psevdo algoritem:
  - Vhod: usmerjeni graf  $G$  z  $n$  vozlišči in izvornim ter ponornim vozliščem
  - 1.korak: generiraj potencialne poti v grafu  $G$  naključno in v velikih količinah
  - 2.korak: zavrži vse poti, ki nimajo ustreznega ponora in izvora
  - 3.korak: zavrži vse poti, ki ne obišejo natanko  $n$  poti



- 4.korak: ponavljaj za vsako od  $n$  vozlišč: preglej vse generirane poti in posamezno pot zavrži, če ne vsebuje trenutno opazovanega vozlišča;
- Izhod: če je v množici poti še kakšna “ustrezna” odgovori pozitivno, drugače negativno;
- Hipoteza: obstoj platforme, ki za posamezni korak algoritma potrebuje le en “procesni korak”
- Koraki 1, 2, 3: enkratni koraki (konstantna odvisnost)
- Korak 4: linearno odvisen od števila  $n$
- Omenjeno hipotezo omogočata nedeterminizem in visok paralelizem procesiranja v mediju DNK



- Adlemanov pristop k kodiranju:
  - posamezno od 7 vozlišč zakodiramo z 20 baznimi pari (A,C,T,G)
  - kodiranje naključno (z dolžino zagotovljena dovoljšnja “razdalja” kodov)
- zgledi kodiranja vozlišč (niz 10+10):
  - s2=TATCGGATCG GTATATCCGA
  - s3=GCTATTCGAG CTTAAAGCTA
  - s4=GGCTAGGTAC CAGCATGCTT



- Cilj iskanja: možne poti po grafu
- Kodiranje poti:
  - 20 kodni niz = [izvor[11..20], ponor[1-10]]
  - Koda poti: neg(stik(izvor[11-20], ponor[1-10]))
  - Dolžina niza kode prehoda kot dela poti: 20 nukleotidnih baz = 20 znakov
  - Niz prehoda negiramo na osnovi Watson-Crickovega komplementa, ki deluje po izrazu na desni:

$$\bar{T} = A, \bar{A} = T, \bar{C} = G, \bar{G} = C$$



- Tehnološke izvedbe psevd algoritma:
  - Sinteza ob prisotnosti encimov za 1.fazo
  - Posledica: “negacije” – prehodi na 2.vijačnici
  - 2.faza: PCR (angl. *Polymerase Chain Reaction*) – izločanje verig z neustreznim izvorom in ponorom
  - 3.faza: GE (angl. *Gel Electrophoresis*) – izločanje neustrezno dolgih verig



- Slabosti eksperimenta:
  - velika poraba materije (200 vozlišč –  $3 \cdot 10^{25}$  kg DNK materije – teža planeta???)
  - stohastičnost procesa (veliko število iteracij ???)

## 4.4. Sistemska biologija

<http://lrs.fri.uni-lj.si/sl/teaching/ont/default.asp?lc=sl>

*Miha Moškon, et.al: Model realizacije funkcionalosti  
RS pomnilne celice v biološkem sistemu.  
Elektrotehniški vestnik. (Osnove postavitve  
biološkega vezja)*



## 4.5. Repozitorij BioBricks gradnikov

- Odprtokodno skladišče
- Sistemizirana DNK zaporedja
- Hierarhična delitev

### Osnovni elementi



- Regulator
- Mesto vezave ribosoma
- Zaporedje kodirajočega proteina
- Terminator
- Konjugator

### Naprave



- Reporter
- Inverter
- Signalizator
- Generator proteina
- Sestavljena naprava
- Meritvena naprava

### Sistemi



- Projekti

- **Okolje** - organizem, medij, gostitelj



### ***Priprava na modeliranje***

Določitev cilja biološkega vezja

Izbira ustreznega gostiteljskega organizma

Izbira ustreznih biokock

### ***Modeliranje***

4) Priprava matematičnega modela

5) Raziskava dinamike vezja *in silico*

### ***Implementacija, testiranje, razhroščevanje***

6) Določitev ustrezne DNK implementacije vezja

7) Sestava večih verzij, v kolikor obstaja kritičen parameter

8) Testiranje vezja



### Vhodne naprave

- Vhodni signali
- PoPS konvertor

### Izhodne naprave

- Izhodni signali

### Biološka logična vrata

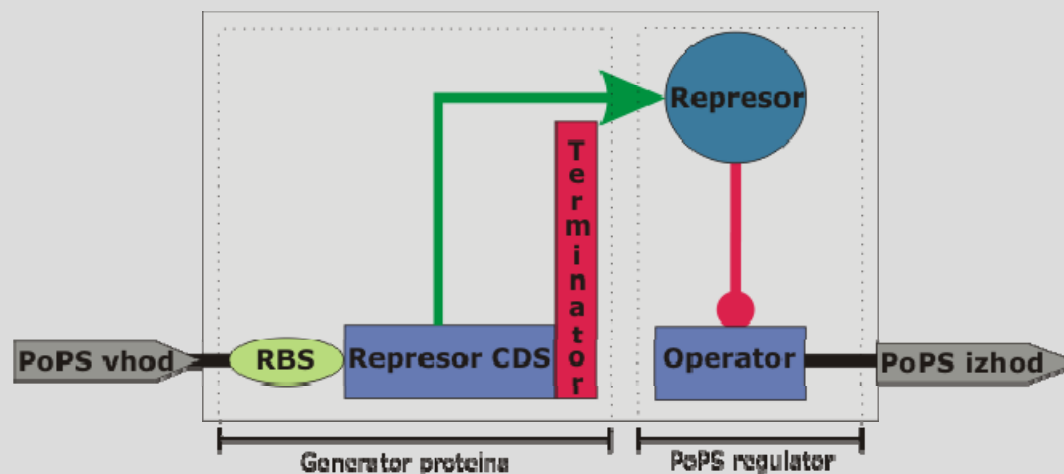
- Generator proteina
- PoPS regulator

### PoIn nabor

$\{\neg, \wedge\}, \{\neg, \vee\}, \{\uparrow\}, \{\downarrow\}$

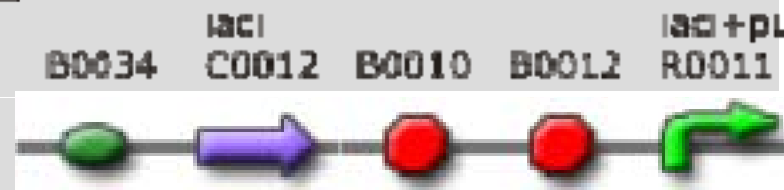


## Bio NOT vrata



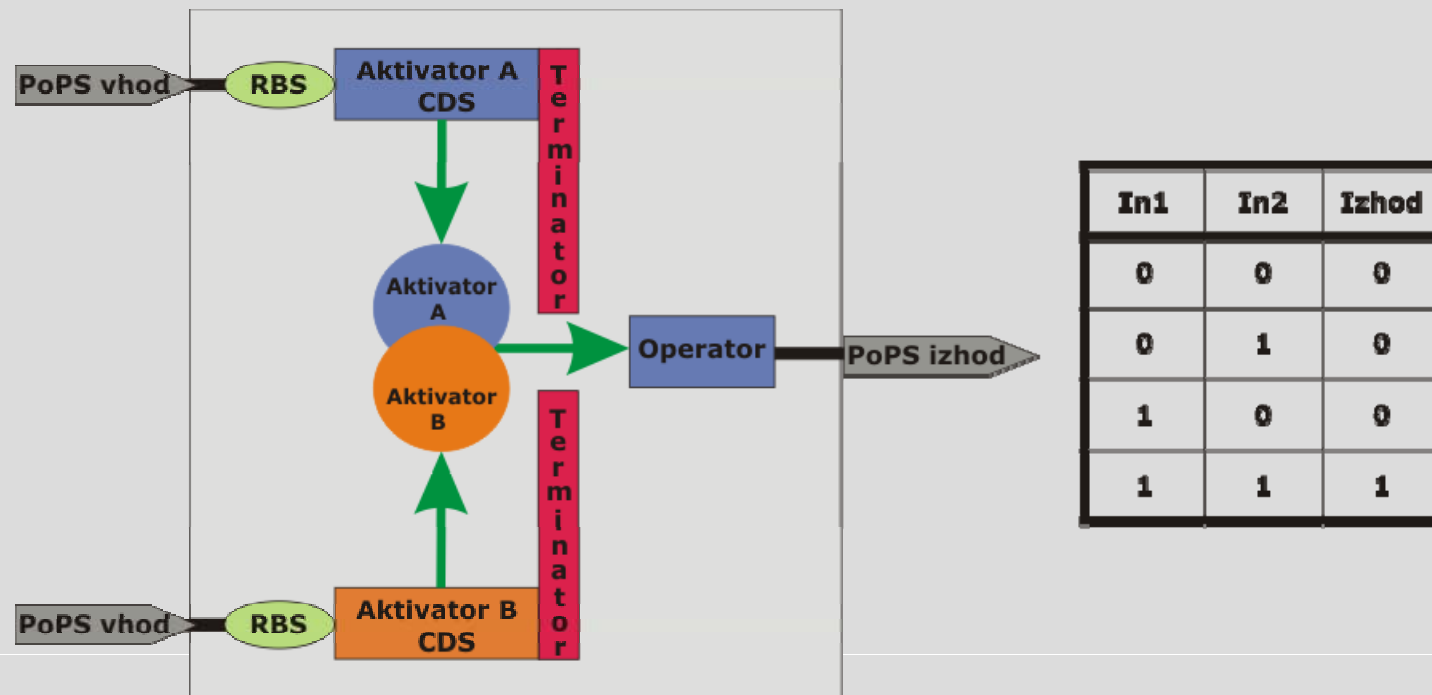
Vhod	R2	Izhod
0	0	1
1	1	0

Biokocka *BBa\_Q04121*:





## Bio AND vrata





- Razpoložljiva orodja:
  - [SynBioSS](#)
  - [Athena](#)
  - [GenoCAD](#)
  - [Bio-Spice](#)
  - [Cell Designer](#)
  - [GeNetDes](#)

## 4.6. Ostali primeri DNK procesiranja

- Računalnika MAYA I, MAYA II: računalnik za igranje igre TIC-TAC-TOE (igra križcev in krogcev)
- 2002: Weizmanov institut (Izrael)
  - Prvi univerzalni DNA računalnik (iz encimov in DNA molekul)
  - Zmogljivost:  $330 \cdot 10^{12}$  operacij/sec. ( $10^5$  x hitreje od danes običajnih komercialnih izvedb osebnih računalnikov)

## 4.7. Povzetek DNA koncepta procesiranja

- Paralelizem -> doseganje velikih hitrosti
- Izredno visoka miniaturizacija pomnjenja
- **Možnost implementacije procesiranja v fluidnih okoljih** (npr. krvi, itd.)
- Energetska nepotravnost