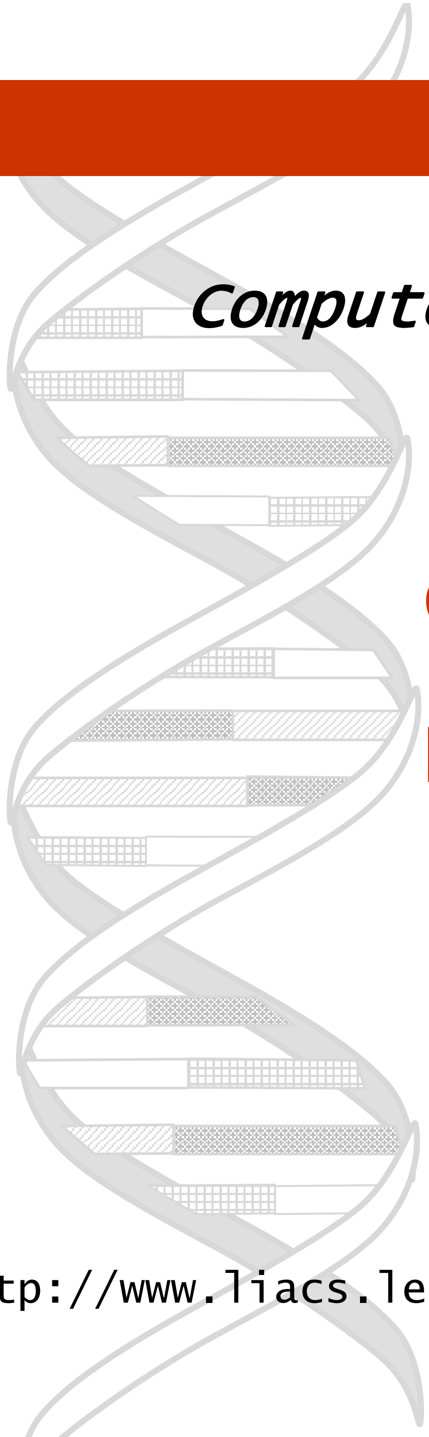
PRE-class



Computer Science: Working Smarter!

Computational Biology

Hendrik Jan Hoogeboom

jan 2021

http://www.liacs.leidenuniv.nl/~hoogeboomhj/praatjes/lapp-top/ »

fundamenteel

fi1 discrete wiskunde

fi2 formele talen

fi3 bereken-baarheid

theory van concurrency

programmeerlijn

Programmeer-methoden

Algoritmiek

Datastructuren

Databases

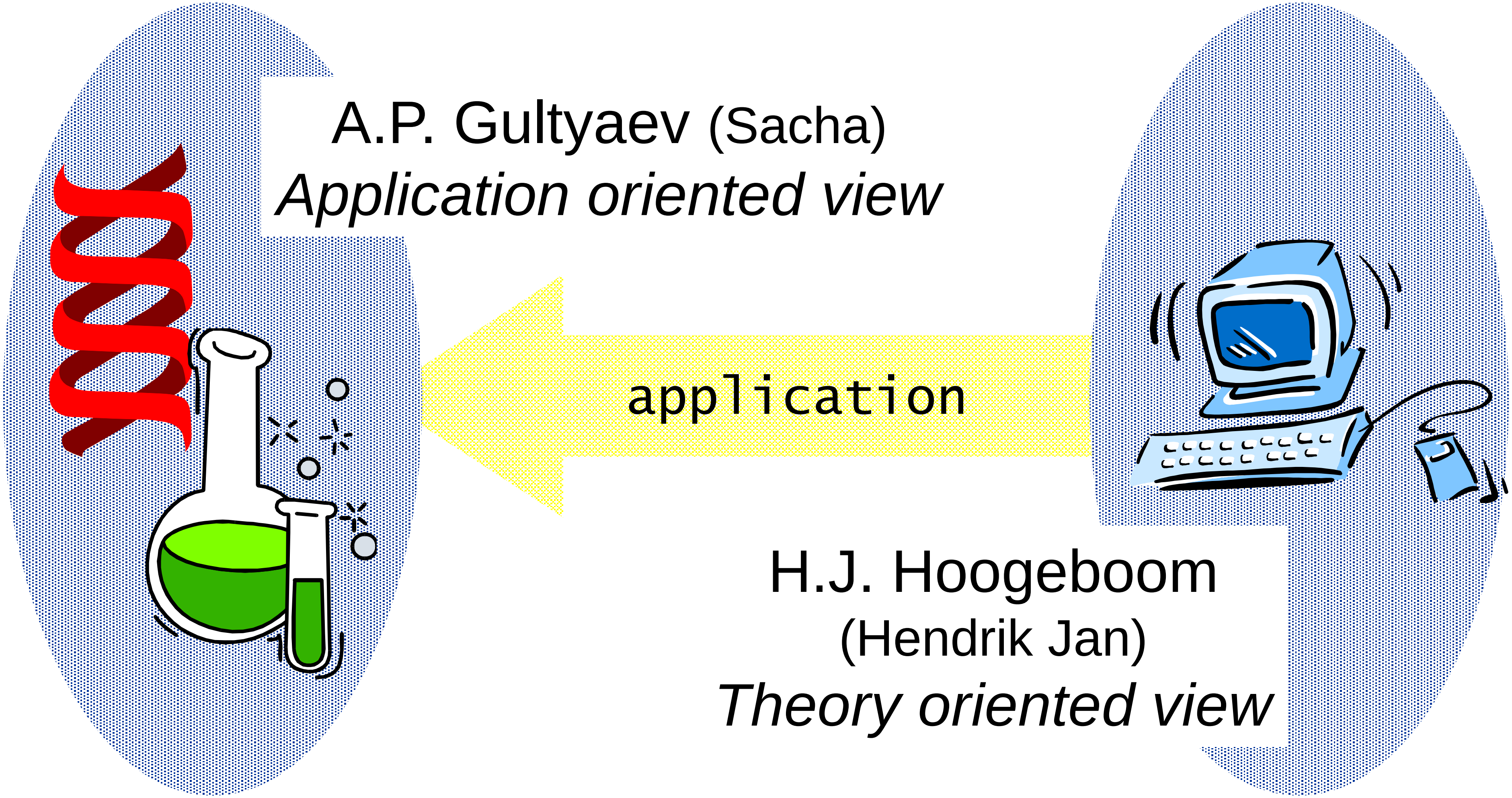
Software engineering

service

challenges
project android

COMPMOLBIOL

Computational Biology



A.P. Gultyaev (Sacha)

Application oriented view

application

H.J. Hoogeboom

(Hendrik Jan)

Theory oriented view

www.liacs.nl/home/hoogeboo/mcb/

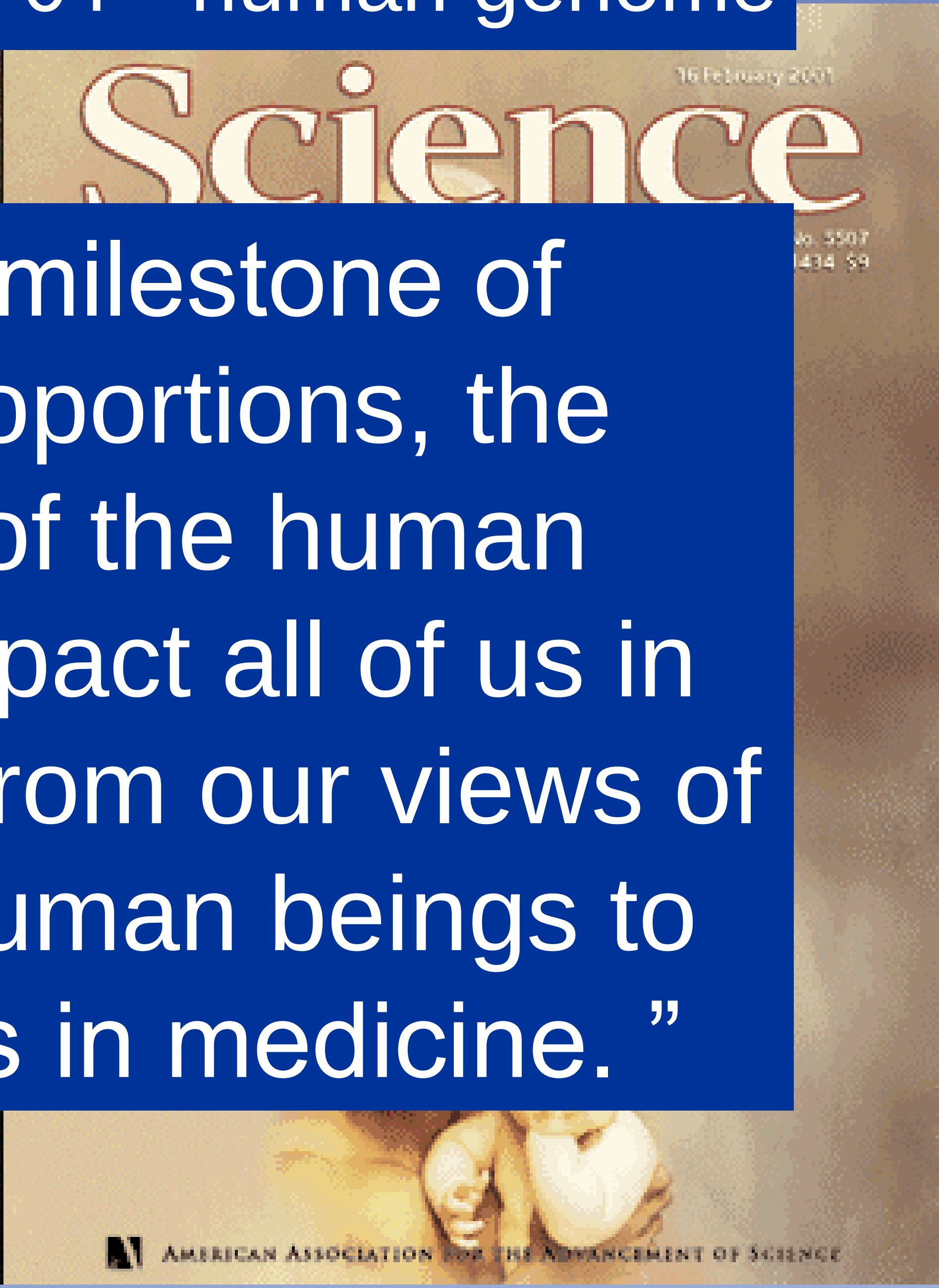
sequence alignment

TCAGACGATTG
TCGGAGCTG

are sequences similar?

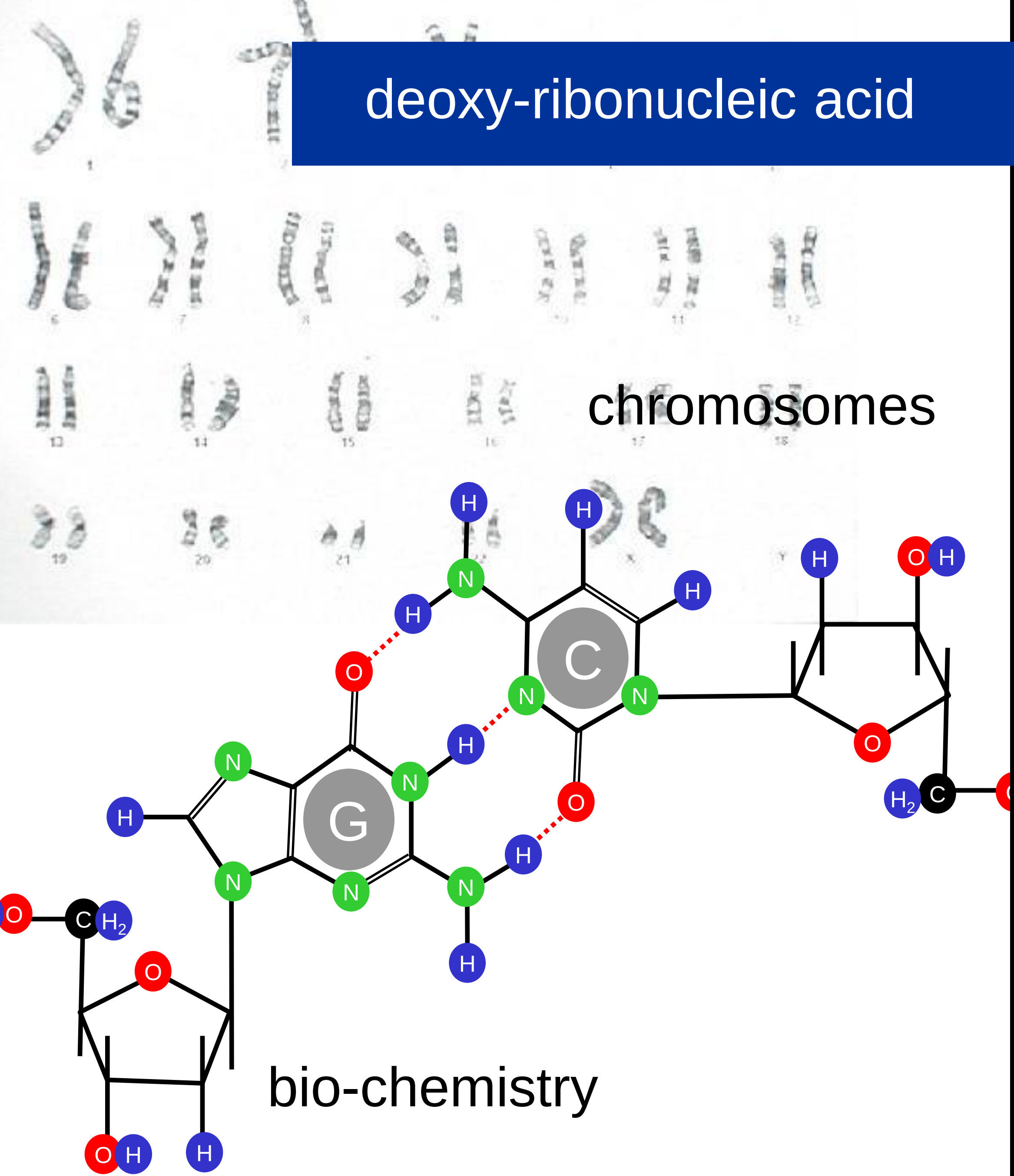
what does that mean …
how do we compute it …

feb’01 - human genome



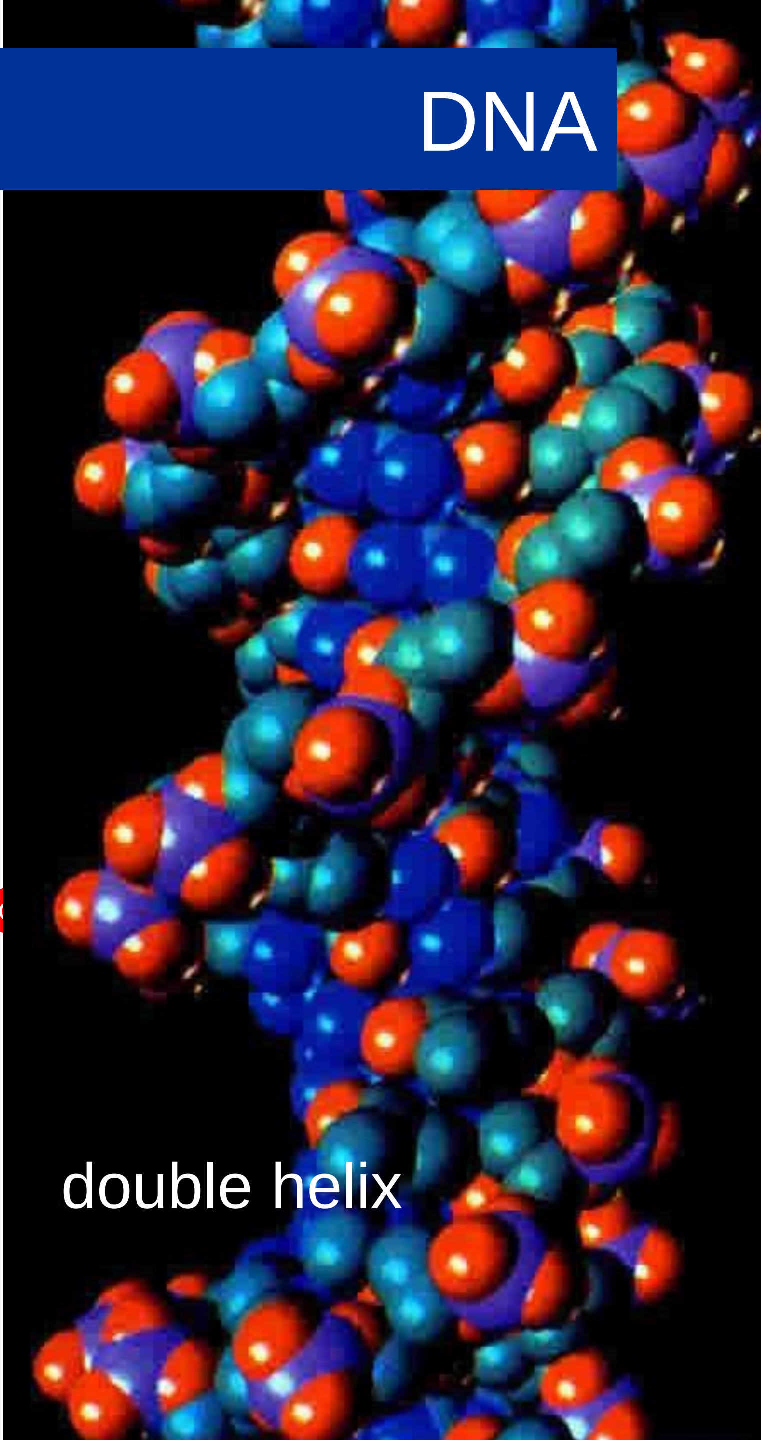
“A scientific milestone of enormous proportions, the sequencing of the human genome will impact all of us in diverse ways – from our views of ourselves as human beings to new paradigms in medicine. ”

deoxy-ribonucleic acid DNA



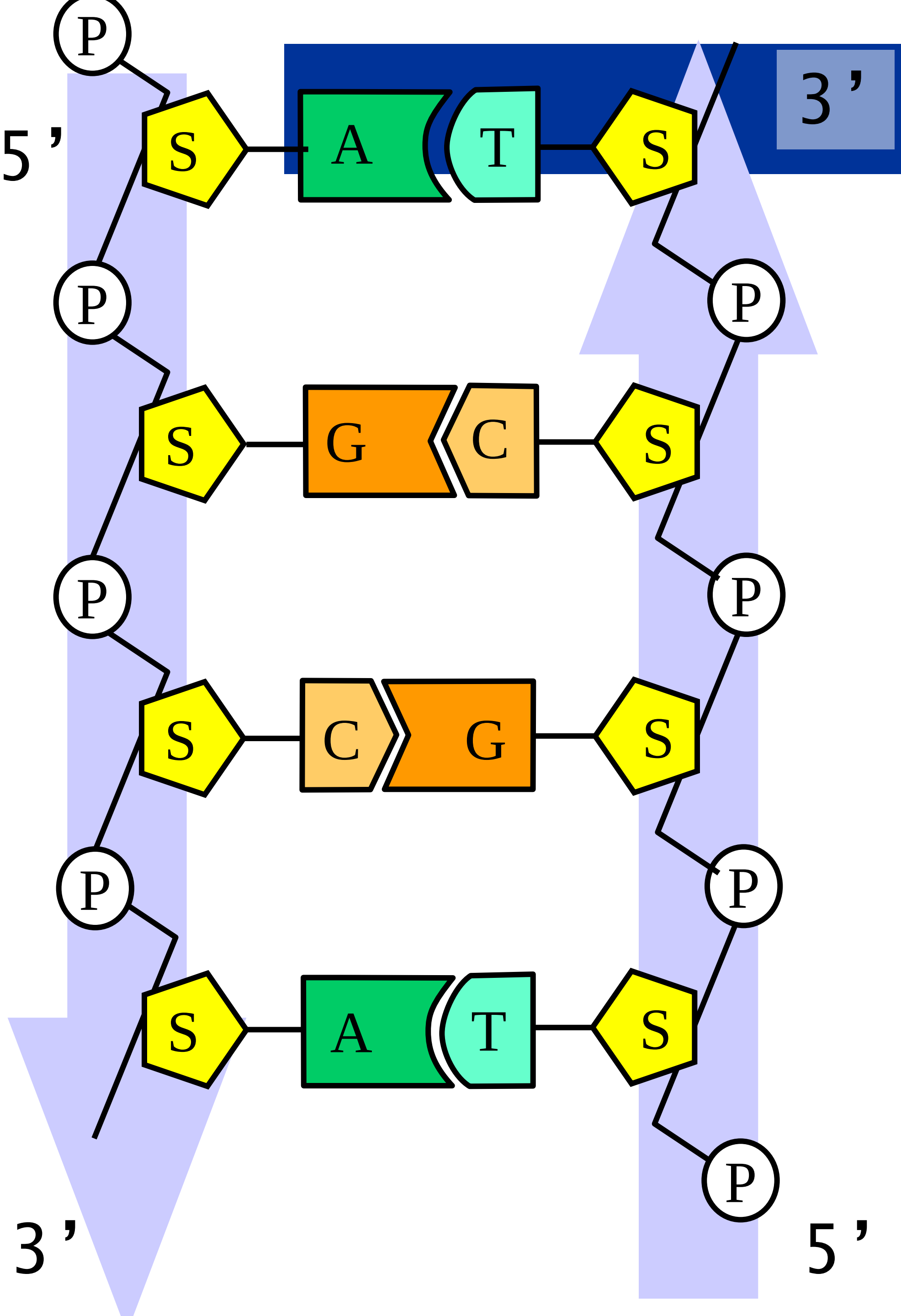
chromosomes

bio-chemistry



double helix

DNA

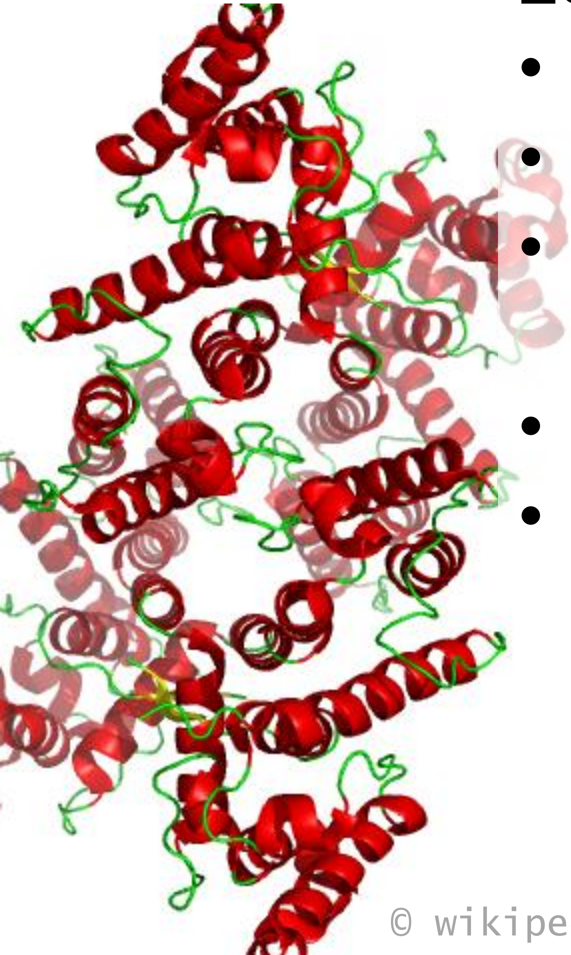


sugar
phosphor
nucleotides

basepairs
Watson & Crick

A=T
adenine - thymine
C≡G
guanine - cytosine

statistics human genome



© wikipedia

23 pairs of chromosomes

- 3.1 10^9 nucleotide bases
- 20,000 - 30,000 genes
- avarage 3000 bases / gene

  dystrophin 2.4 million
- protein variants 1 million (splicing)
- 99.9% exactly the same in humans

statistics human genome

20,000 - 30,000 - 40,000 genes?

The haploid human genome contains ca. 23,000 protein-coding genes, far fewer than had been expected before its sequencing. In fact, only about 1.5% of the genome codes for proteins, while the rest consists of non-coding RNA genes, regulatory sequences, introns, and noncoding DNA (once known as "junk DNA").

wikipedia http://en.wikipedia.org/wiki/Human_genome feb 2011

There are an estimated 20,000-25,000 human protein-coding genes. The estimate of the number of human genes has been repeatedly revised down from initial predictions of 100,000 or more as genome sequence quality and gene finding methods have improved, and could continue to drop further. Protein-coding sequences account for only a very small fraction of the genome (approximately 1.5%), and the rest is associated with non-coding RNA molecules, regulatory DNA sequences, LINEs, SINEs, introns, and sequences for which as yet no function has been determined.

mar 2016

themes

problem ⇒ model (eg. graph)

- known algorithms
- characterization

unprecise data

complexity

⇒ heuristics

what *is* the right answer?

two alphabets

DNA

bases

4 symbols

a c t g

RNA

a c u g

eiwitten

proteins

amino acids

20 symbols

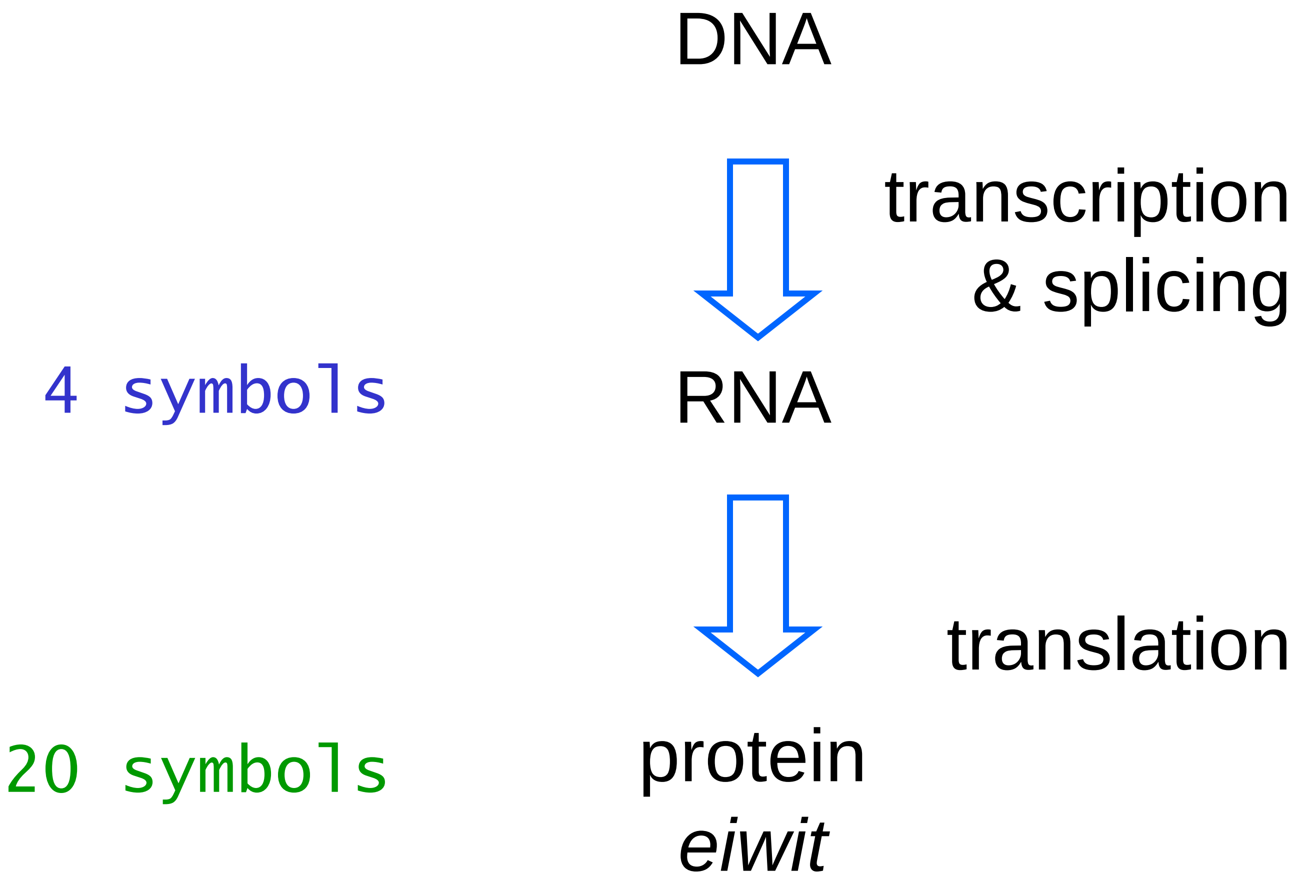
A R D N C

E Q G H I

L K M F P

S T W Y V

central dogma



DNA

transcription
& splicing

4 symbols

RNA

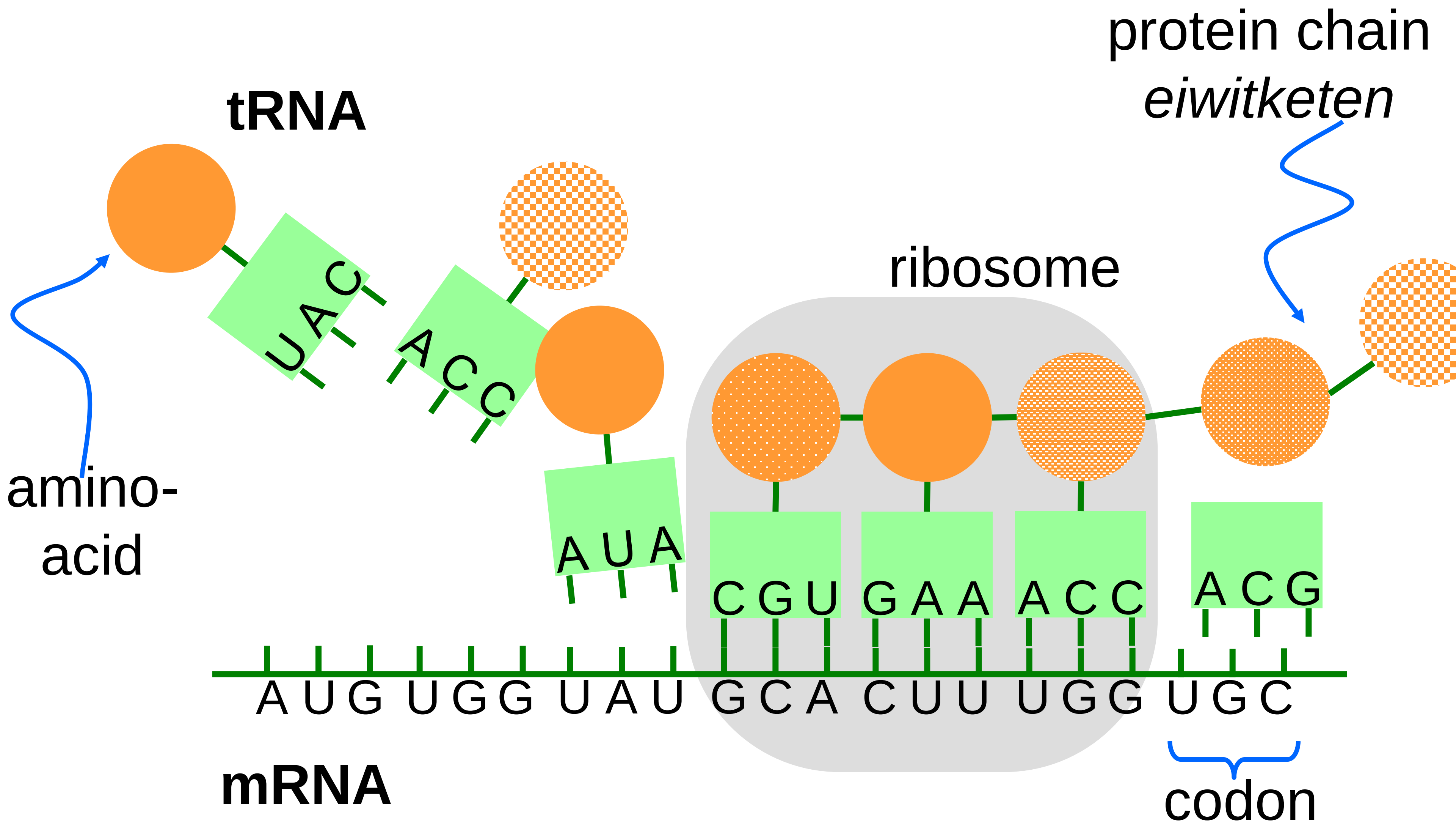
translation

20 symbols

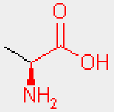
protein
eiwit

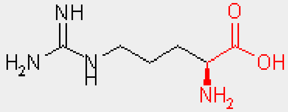
‘gene expression’

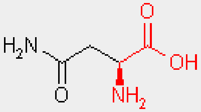
translation

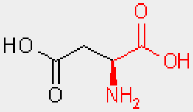


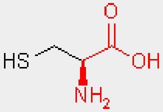
20 amino acids

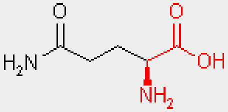

Ala

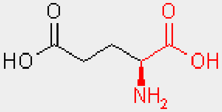

Arg

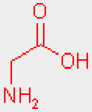

Asn

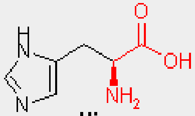

Asp

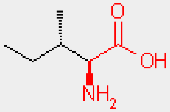

Cys

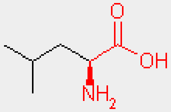

Gln

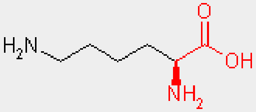

Glu

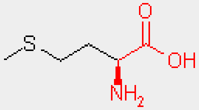

Gly

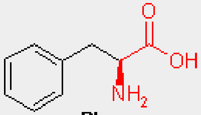

His

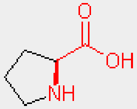

Ile

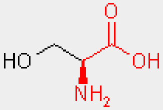

Leu

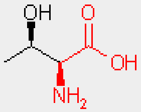

Lys

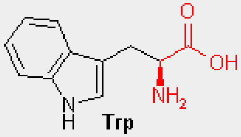

Met

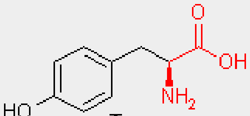

Phe

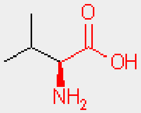

Pro

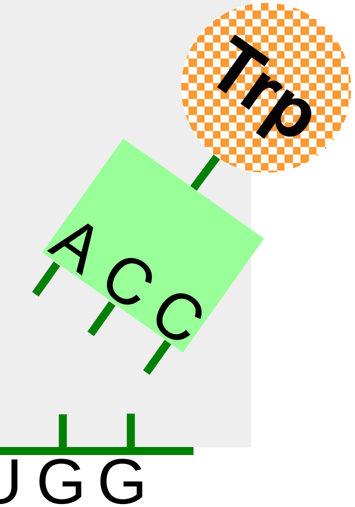

Ser


Thr


Trp


Tyr


Val



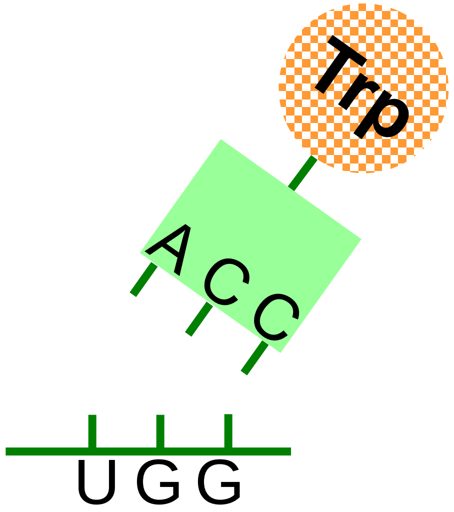
genetic code

	U	C	A	G	
U	Phe	Ser	Tyr	Cys	U
	Phe	Ser	Tyr	Cys	C
	Leu	Ser	Stop	Stop	A
	Leu	Ser	Stop	Trp	G
C	Leu	Pro	His	Arg	U
	Leu	Pro	His	Arg	C
	Leu	Pro	Gln	Arg	A
	Leu	Pro	Gln	Arg	G
A	Ile	Thr	Asn	Ser	U
	Ile	Thr	Asn	Ser	C
	Ile	Thr	Lys	Arg	A
	Met	Thr	Lys	Arg	G
G	Val	Ala	Asp	Gly	U
	Val	Ala	Asp	Gly	C
	Val	Ala	Glu	Gly	A
	Val	Ala	Glu	Gly	G

UGG

⇓

Trp



questions

- Lookup
  - Is the gene known for my protein (or vice versa)?
  - **On which chromosome is the gene located?**
  - What sequence patterns are present in my protein?
  - Are the mutations known which cause this disease?
  - **To what class or family does my protein belong? What is known?**
- Compare
  - Are there sequences in the database resembling my protein?
  - How can I optimally align the members of this protein family?
  - Are these two sequences similar?
- Predict
  - Can I predict the active site residues of this enzyme?
  - Why are these patients ill?
  - **Can I make a 3D model for my protein?**
  - Can I predict a (better) drug for this target?
  - How can I improve the thermostability? (protein engineering)
  - How can I predict the genes located on this genome?

BLAST

The following DNA sequence fragment, containing some mutation, was isolated from a patient:

tttgctccccgcgcgctgtttttctcagtgactttcagcgggcggaaaag

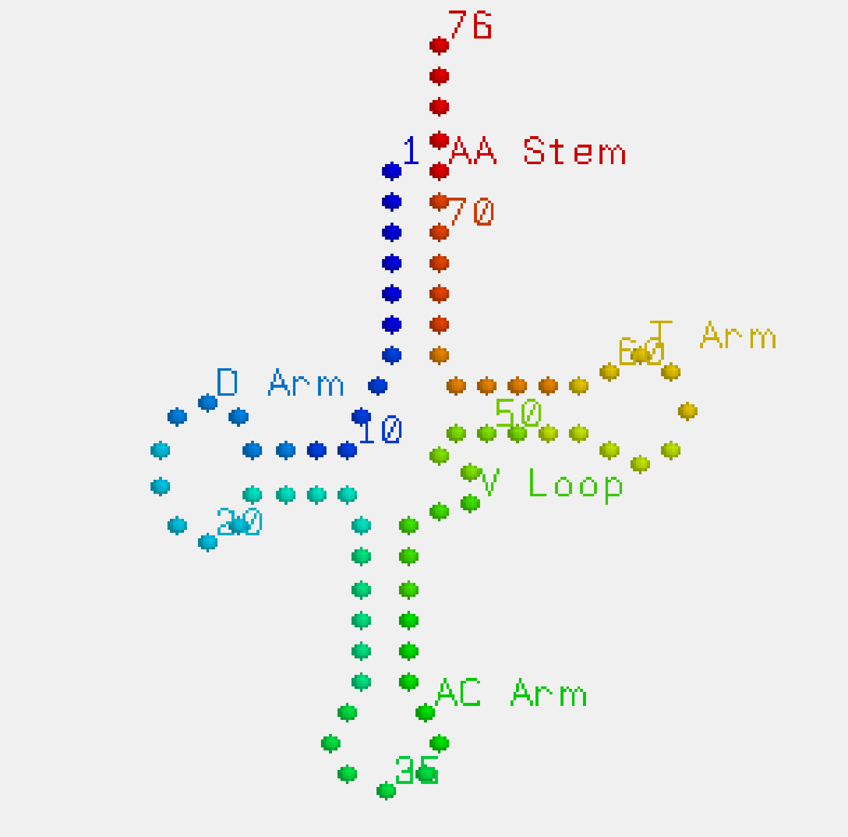
(a) In what gene the mutation is located? On which chromosome? How many nucleotides are changed?

(b) Could you indicate a possible disease determined by this mutation?

http://www.ncbi.nlm.nih.gov/

nucleotide blast

2D & 3D Structures of Yeast Phenylalanyl-Transfer RNA

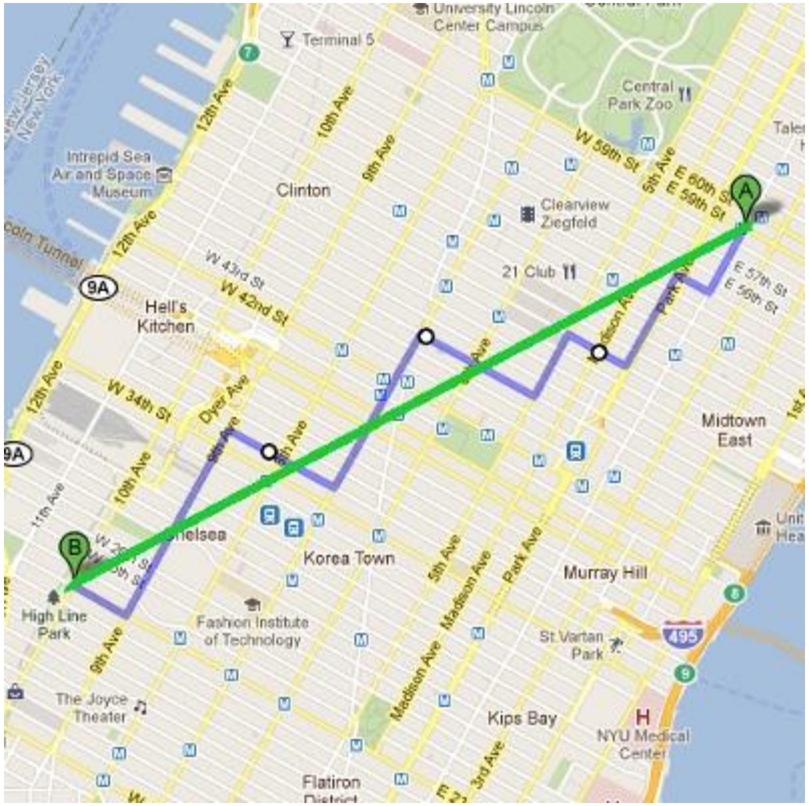


2D Structure

3D Structure

ALIGNMENT

Manhattan

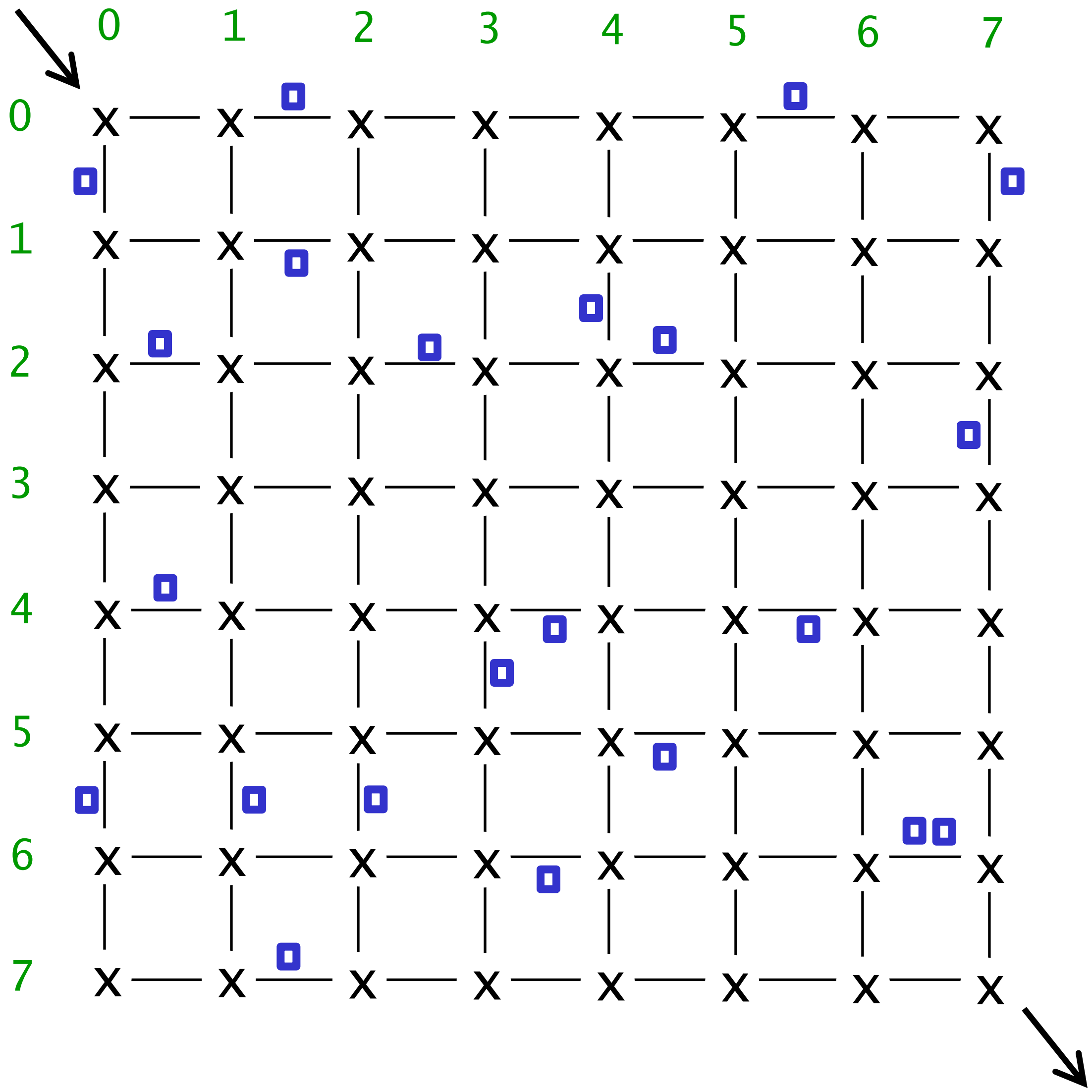


MOOC Pevzner

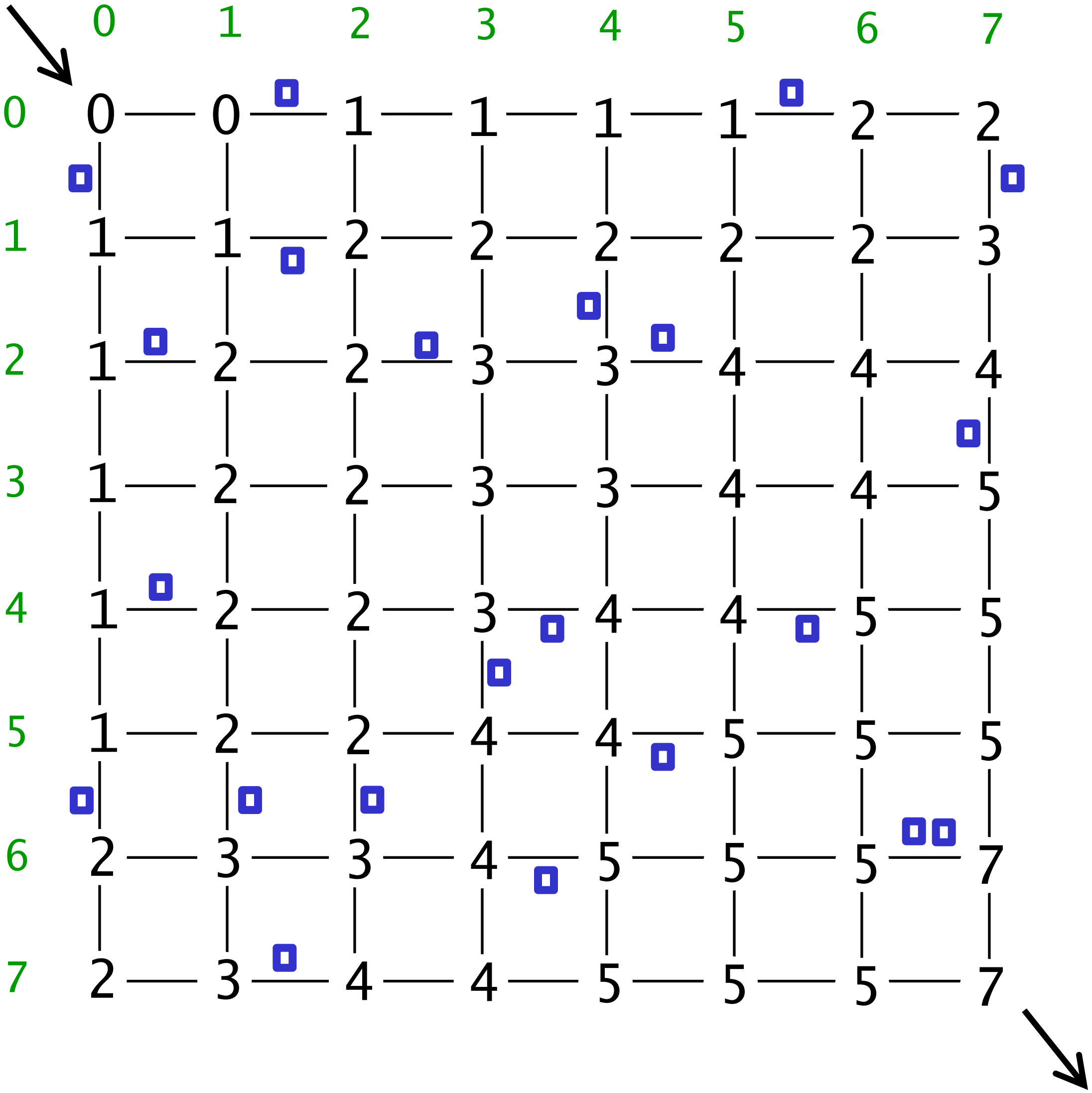
https://www.youtube.com/watch?v=wmIMeyWDZI0

- maximalizeren
- hoeveel mogelijkheden?

Manhattan

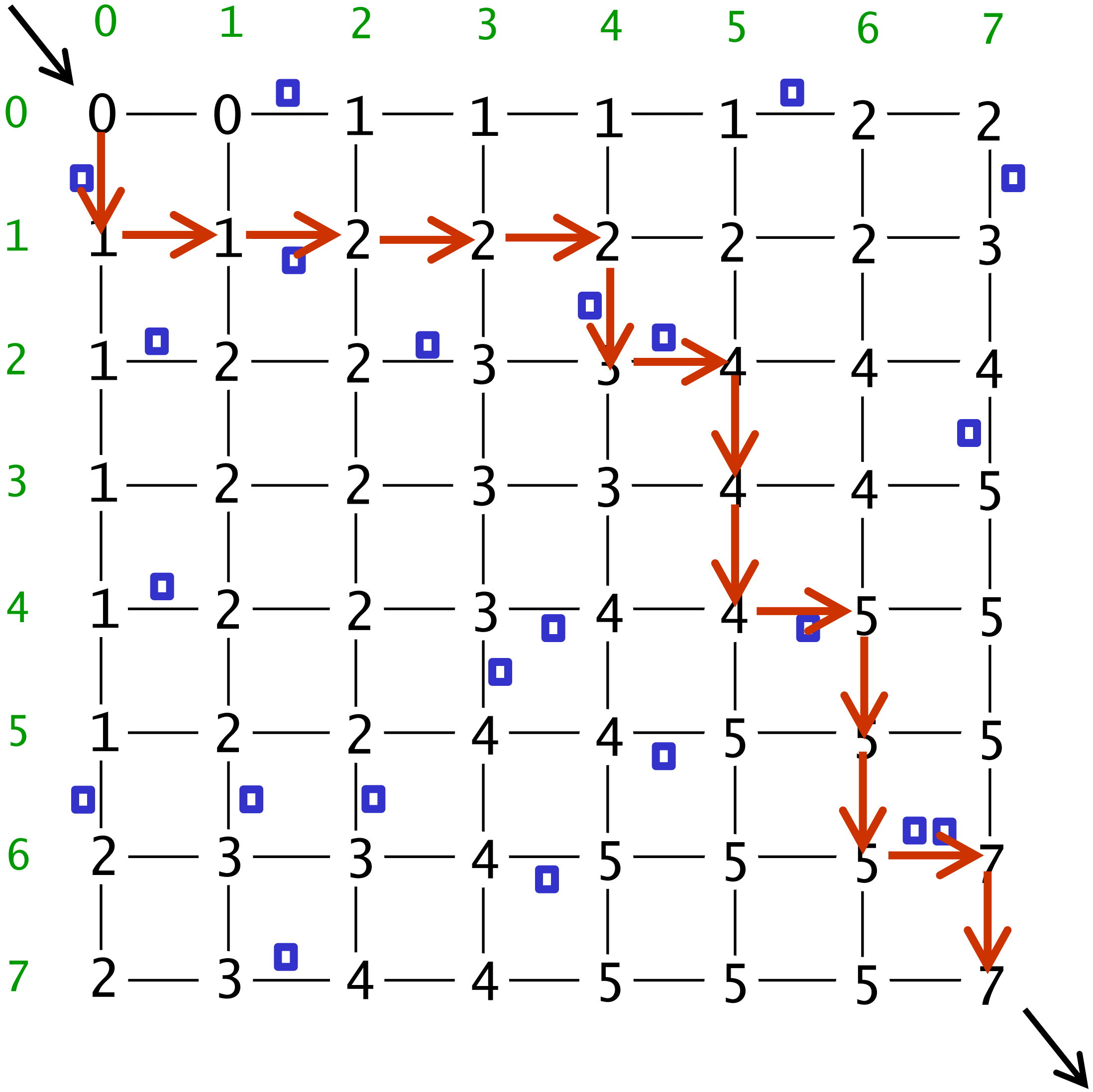


Manhattan



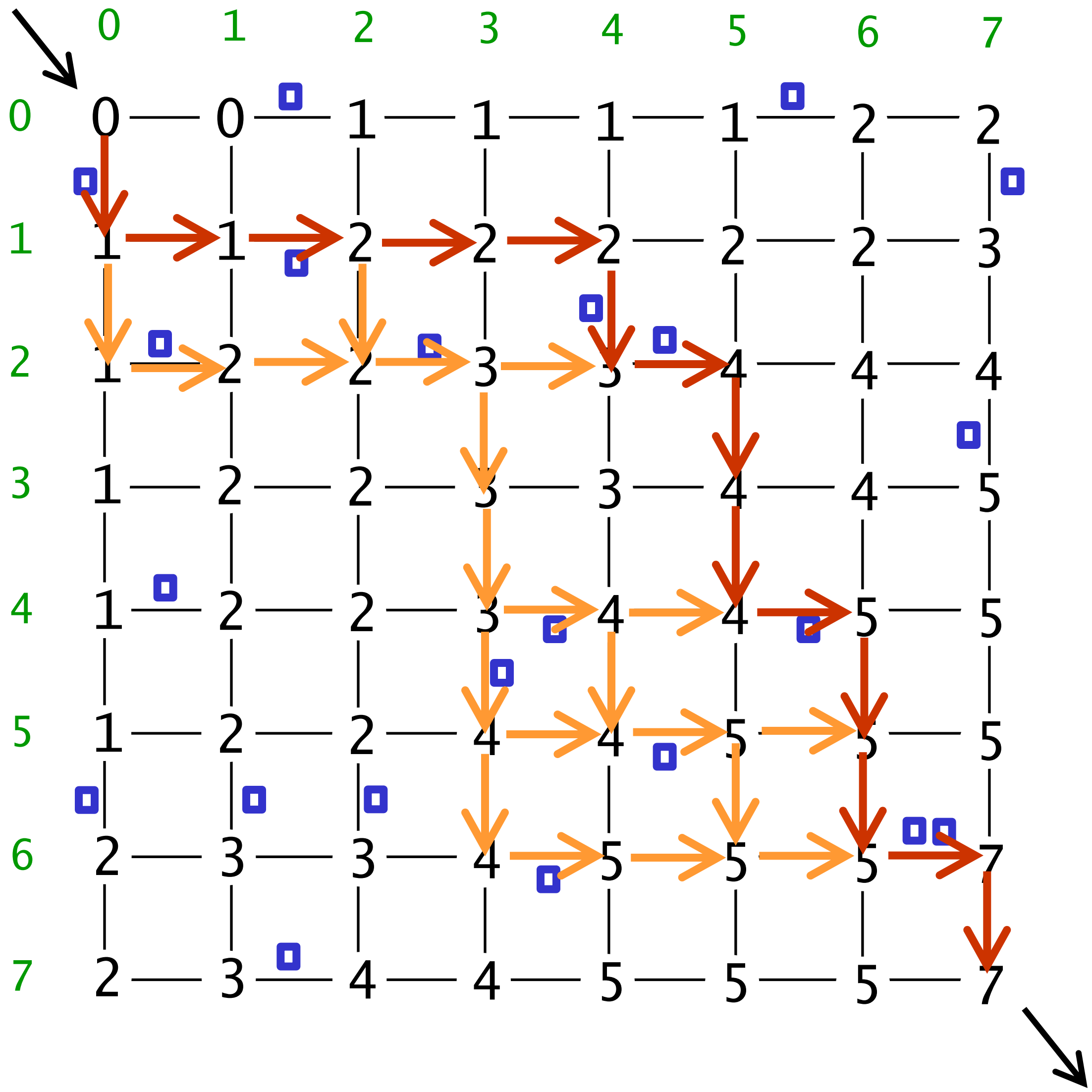
	0	1	2	3	4	5	6	7
0	0	0	1	1	1	1	2	2
1	1	1	2	2	2	2	2	3
2	1	2	2	3	3	4	4	4
3	1	2	2	3	3	4	4	5
4	1	2	2	3	4	4	5	5
5	1	2	2	4	4	5	5	5
6	2	3	3	4	5	5	5	7
7	2	3	4	4	5	5	5	7

Manhattan



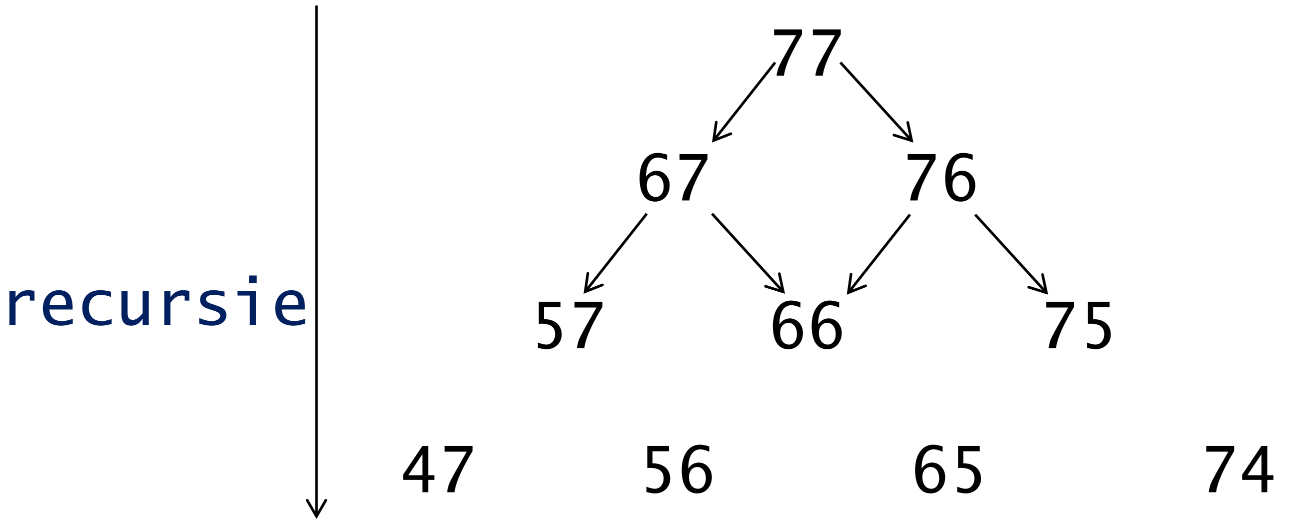
	0	1	2	3	4	5	6	7
0	0	0	1	1	1	1	2	2
1	1	1	2	2	2	2	2	3
2	1	2	2	3	3	4	4	4
3	1	2	2	3	3	4	4	5
4	1	2	2	3	4	4	5	5
5	1	2	2	4	4	5	5	5
6	2	3	3	4	5	5	5	7
7	2	3	4	4	5	5	5	7

Manhattan



	0	1	2	3	4	5	6	7
0	0	0	1	1	1	1	2	2
1	1	1	2	2	2	2	2	3
2	1	2	2	3	3	4	4	4
3	1	2	2	3	3	4	4	5
4	1	2	2	3	4	4	5	5
5	1	2	2	4	4	5	5	5
6	2	3	3	4	5	5	5	7
7	2	3	4	4	5	5	5	7

Manhattan



77

67 76

recursie

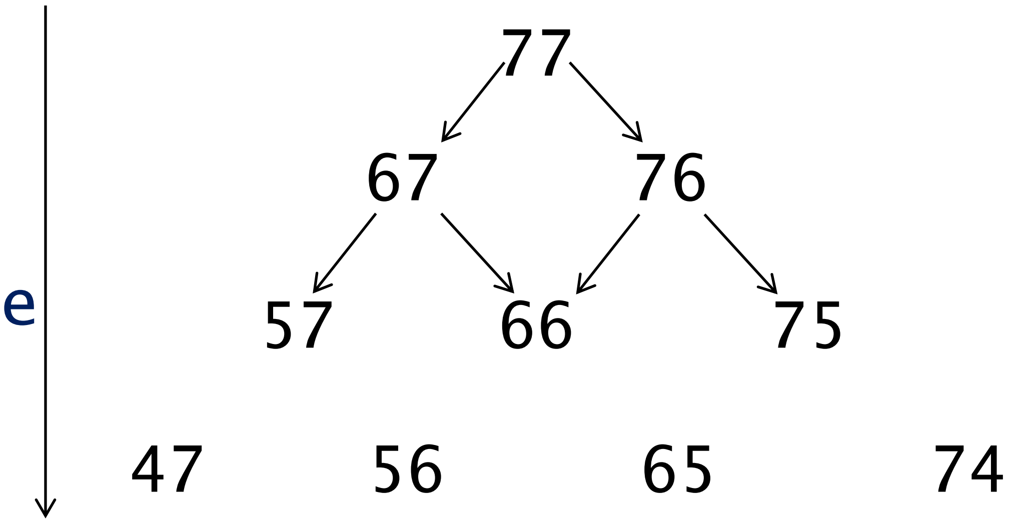
57 66 75

47 56 65 74

2^n

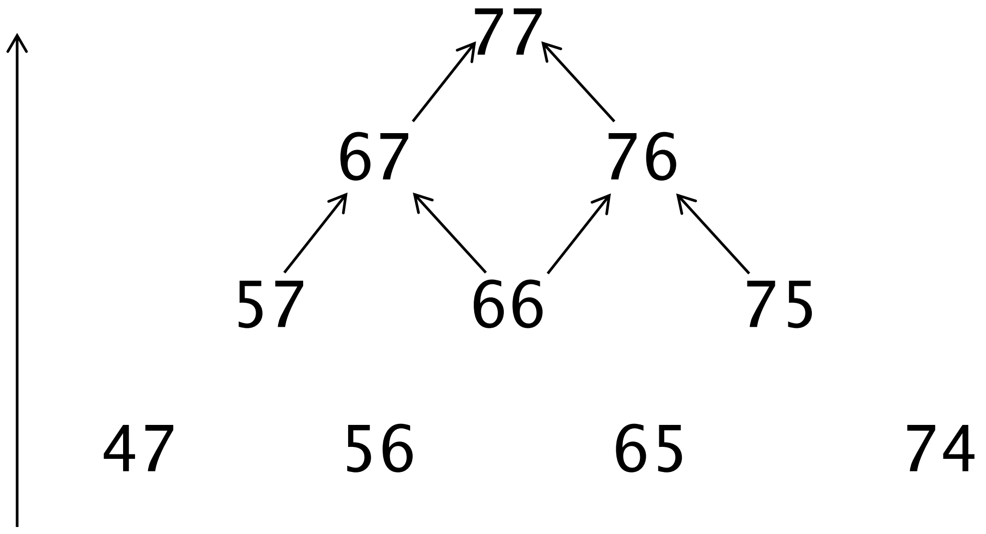
Manhattan

recursie



2^n

dynamisch programmeren



n^2

sequence alignment

are sequences similar?

what does that mean …
how do we compute it …

sequence alignment

TCAGACGATTG
TCGGAGCTG

```
TCAG-ACG-ATTG
TC-GGA-GC-T-G
```

insertion
deletion
substitution

```
TCAGACGATTG
TCGGAGC--TG
```

match
mismatch
gap

```
TCAG-ACGATTG
TC-GGA-GCT-G
```

sequence alignment

```
TCAGACGATTG
  TCGGAGCTG
```

```
TCAG-ACG-ATTG
TC-GGA-GC-T-G     14-6=8
```

```
TCAGACGATTG
TCGGAGC--TG       12-3-2=7
```

```
TCAG-ACGATTG
TC-GGA-GCT-G      14-1-4=9
```

match	+2
mismatch	-1
gap	-1

kan dit beter?

sequence alignment

```
TCAGACGATTG
 TCGGAGCTG
```

```
TCAG-ACG-ATTG
TC-GGA-GC-T-G
```

```
TCAGACGATTG
TCGGAGC--TG
```

```
TCAG-ACGATTG
TC-GGA-GCT-G
```

```
TCAGACGATTG
TCGGA-GCT-G
```

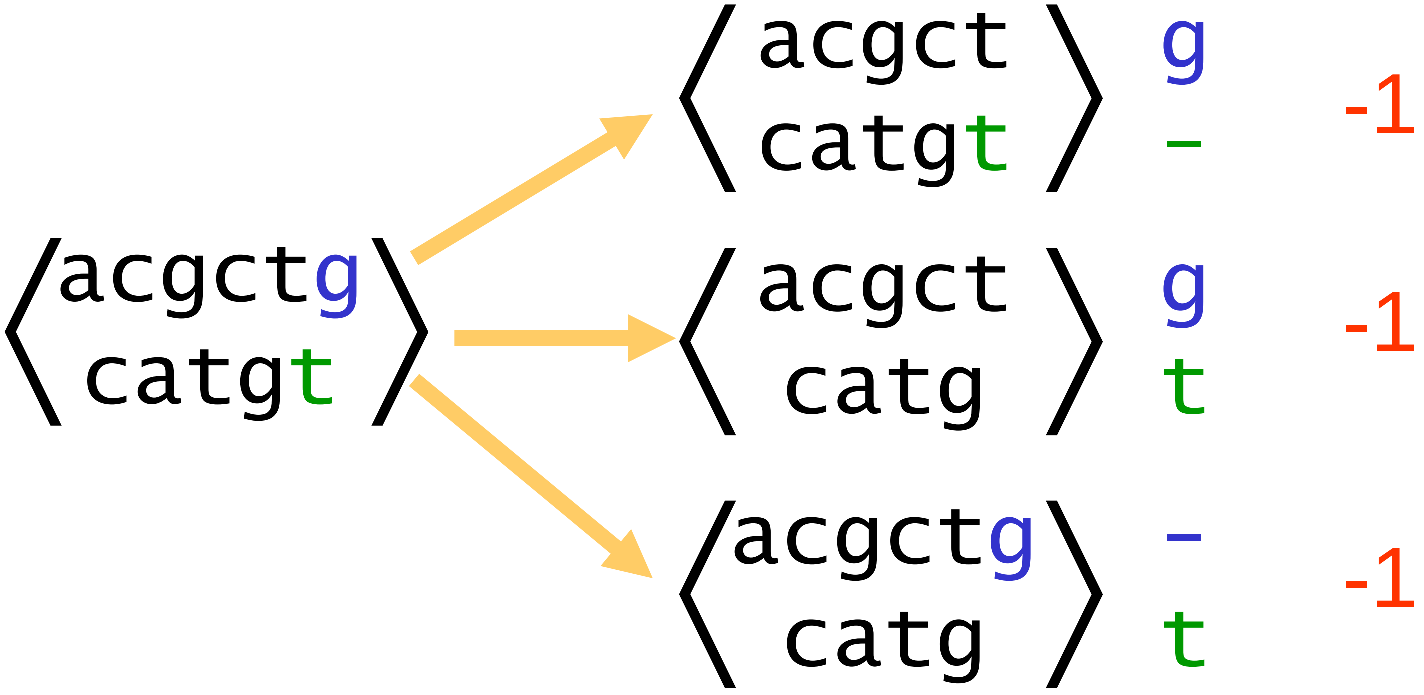
14-2-2=10

basic algorithm

alignment

- recursive principle
- dynamic programming

alignment: recursion



$$\left\langle \begin{matrix}\texttt{acgctg}\\ \texttt{catgt}\end{matrix} \right\rangle$$

$$\left\langle \begin{matrix}\texttt{acgct}\\ \texttt{catgt}\end{matrix} \right\rangle \begin{matrix}\texttt{g}\\ \texttt{-}\end{matrix} \qquad -1$$

$$\left\langle \begin{matrix}\texttt{acgct}\\ \texttt{catg}\end{matrix} \right\rangle \begin{matrix}\texttt{g}\\ \texttt{t}\end{matrix} \qquad -1$$

$$\left\langle \begin{matrix}\texttt{acgctg}\\ \texttt{catg}\end{matrix} \right\rangle \begin{matrix}\texttt{-}\\ \texttt{t}\end{matrix} \qquad -1$$

$\sigma(-,x) = -1$

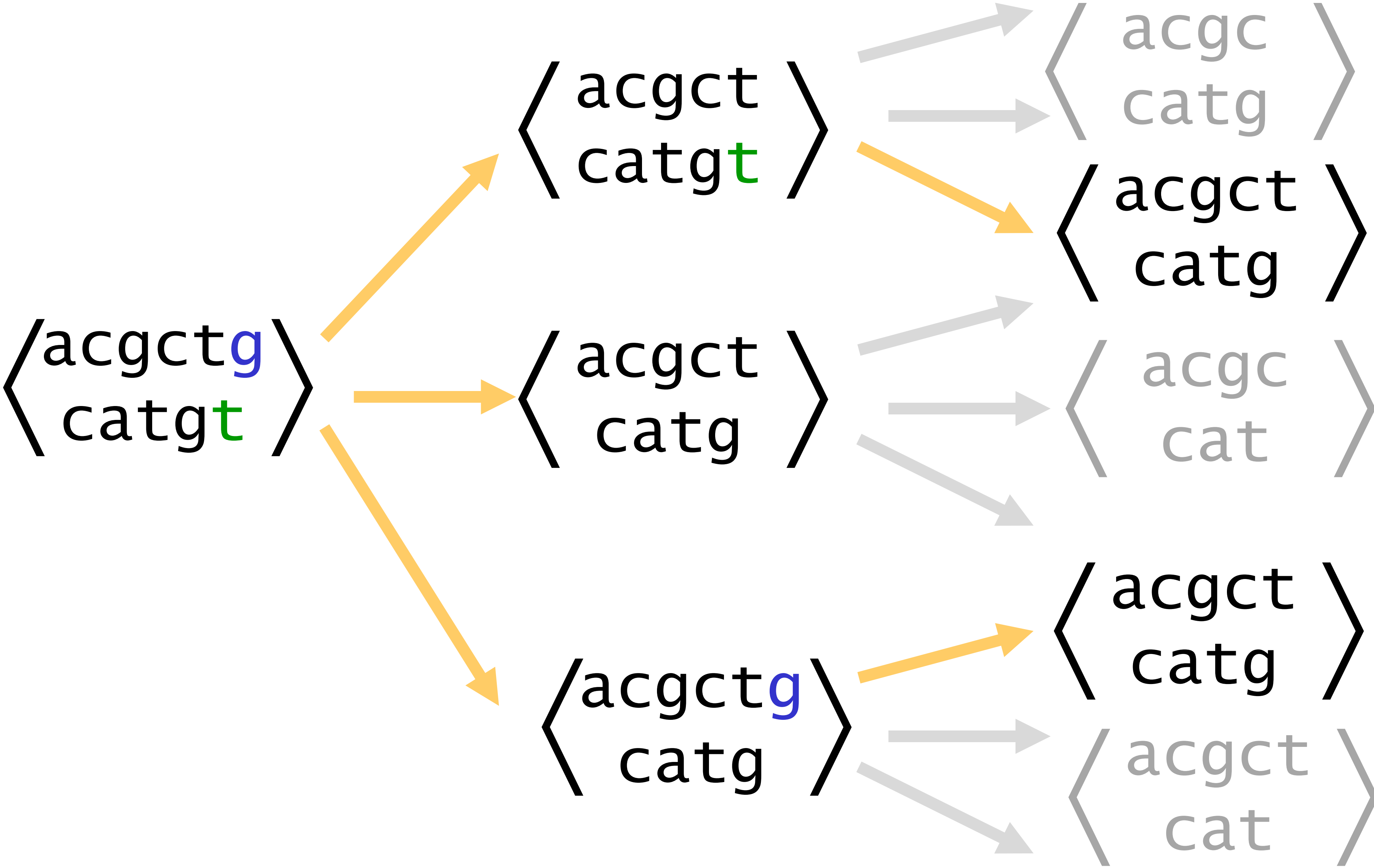
$\sigma(x,-) = -1$

$\sigma(x,y) = -1$

$\sigma(x,x) = \ 2$

‘reward and punishment’

alignment: recursion



⟨acgctg
catgt⟩

⟨acgct
catgt⟩

⟨acgct
catg⟩

⟨acgctg
catg⟩

⟨acgc
catg⟩

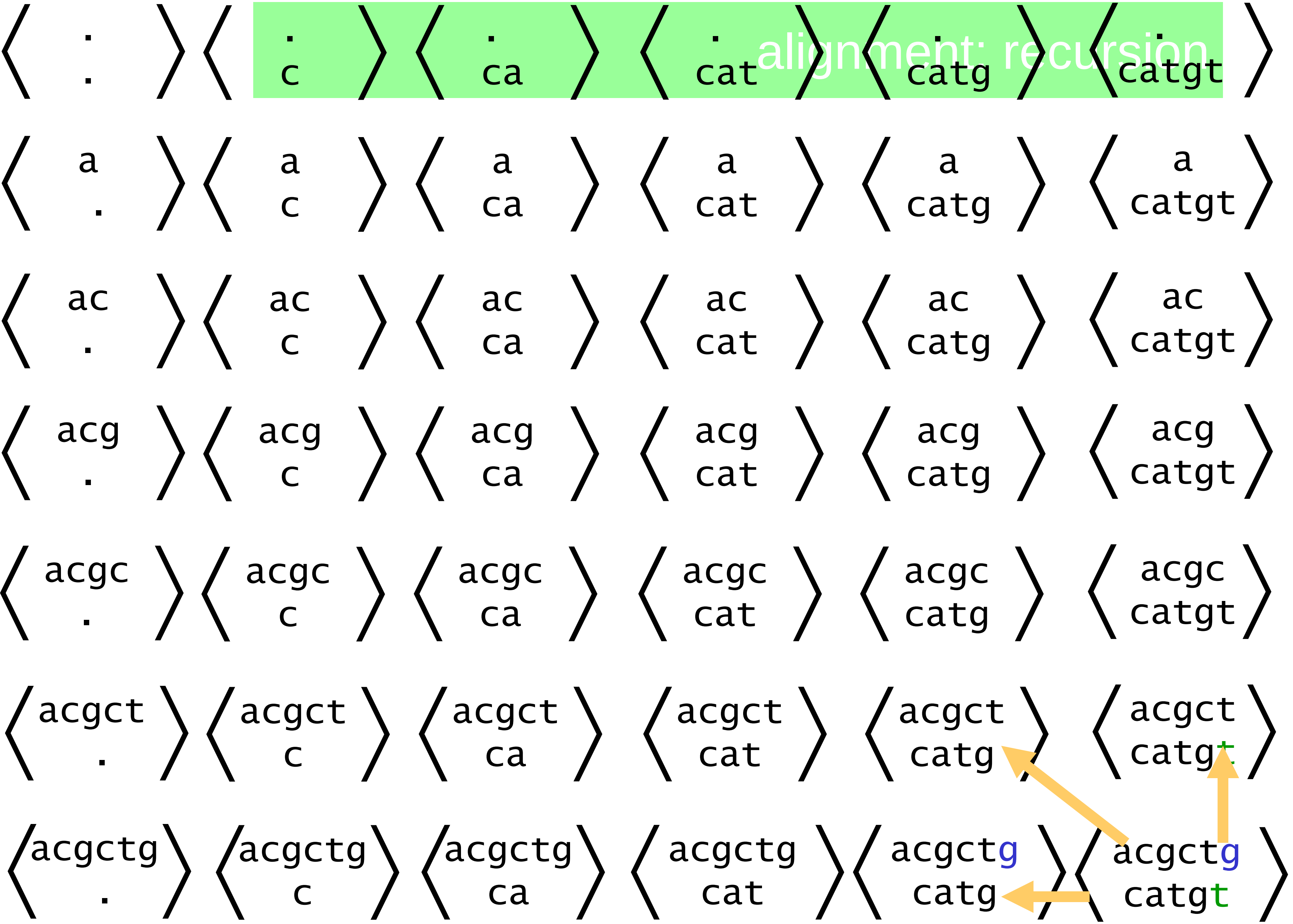
⟨acgct
catg⟩

⟨acgc
cat⟩

⟨acgct
catg⟩

⟨acgct
cat⟩

alignment: recursion



. .	. c	. ca	. cat	. catg	. catgt
a .	a c	a ca	a cat	a catg	a catgt
ac .	ac c	ac ca	ac cat	ac catg	ac catgt
acg .	acg c	acg ca	acg cat	acg catg	acg catgt
acgc .	acgc c	acgc ca	acgc cat	acgc catg	acgc catgt
acgct .	acgct c	acgct ca	acgct cat	acgct catg	acgct catgt
acgctg .	acgctg c	acgctg ca	acgctg cat	acgctg catg	acgctg catgt

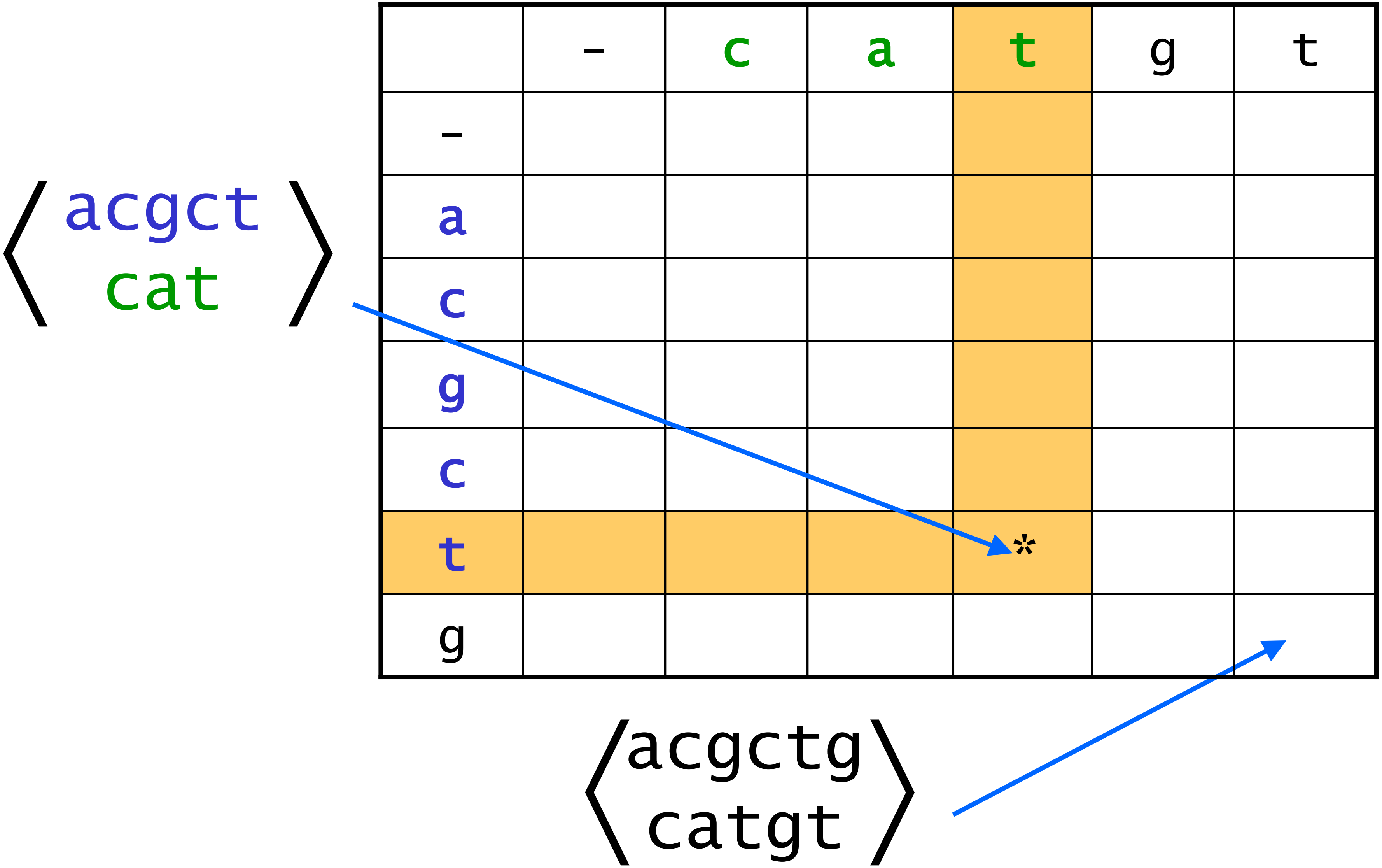
dynamic programming

$$\left\langle \begin{array}{c} \texttt{acgct} \\ \texttt{cat} \end{array} \right\rangle$$

	-	c	a	t	g	t
-	0	-1	-2	-3	-4	-5
a	-1	-1	1	0	-1	-2
c	-2	1	0	0	-1	-2
g	-3	0	0	-1	2	1
c	-4	-1	-1	-1	1	1
t	-5	-2	-2	1	0	3
g	-6	-3	-3	0	3	**2**

$$\left\langle \begin{array}{c} \texttt{acgctg} \\ \texttt{catgt} \end{array} \right\rangle$$

dynamic programming



	-	c	a	t	g	t
-						
a						
c						
g						
c						
t				*		
g						

⟨ acgct
cat ⟩

⟨ acgctg
catgt ⟩

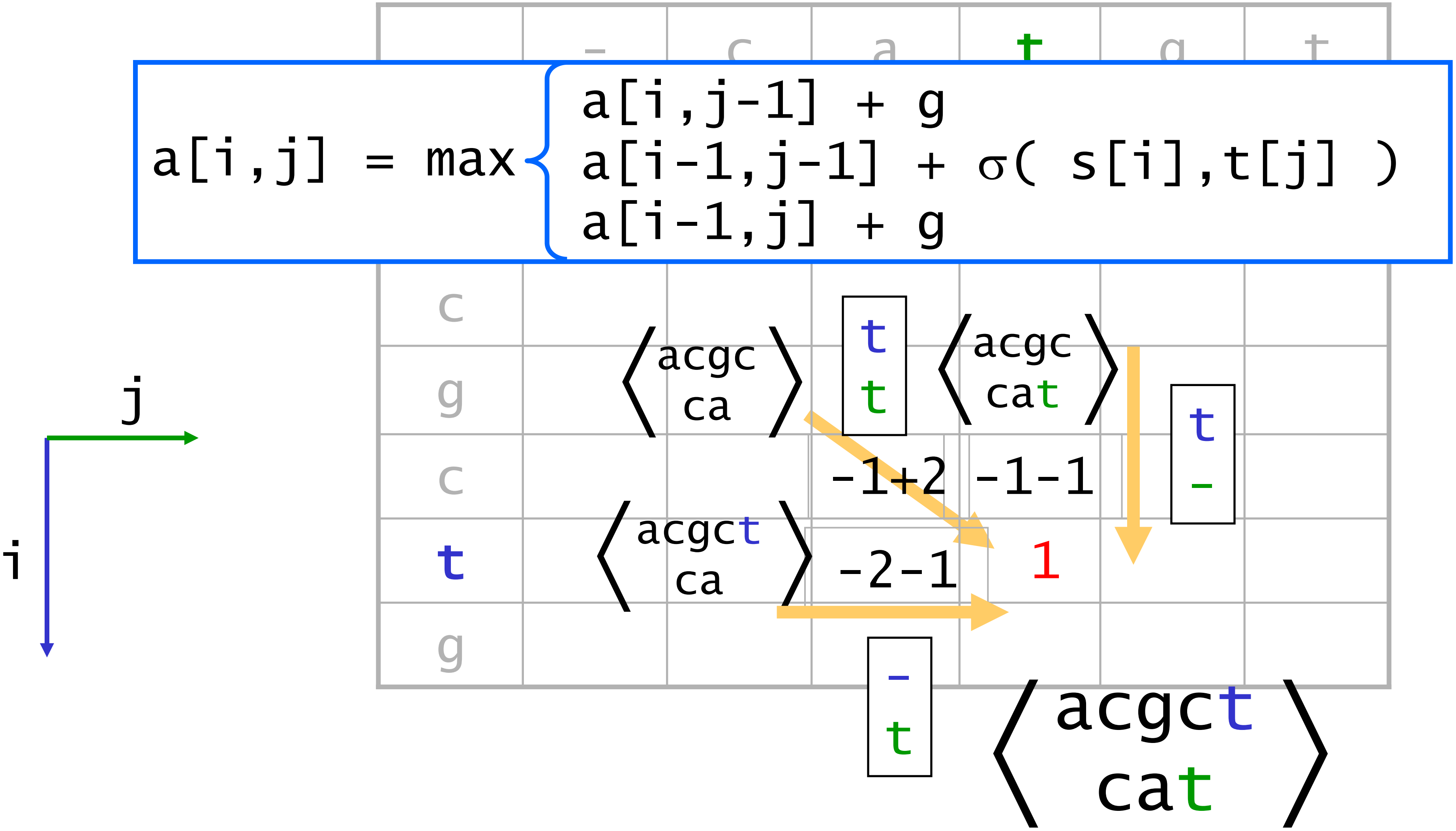
initialization

$$\left\langle \begin{array}{c} \texttt{---} \\ \texttt{cat} \end{array} \right\rangle$$

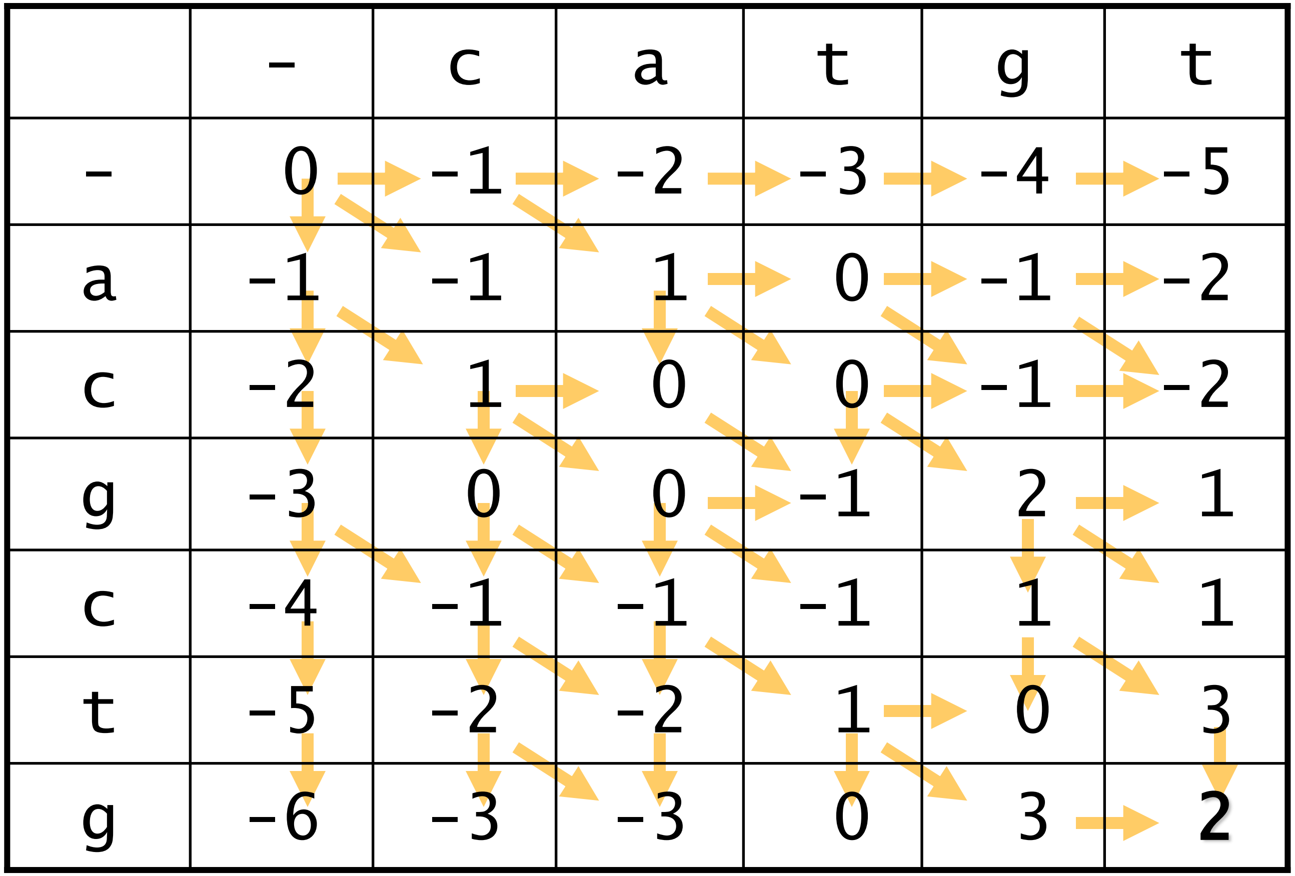
	-	c	a	t	g	t
-	0	-1	-2	-3	-4	-5
a	-1					
c	-2					
g	-3					
c	-4					
t	-5					
g	-6					

dynamic programming

$$a[i,j] = \max \begin{cases} a[i,j-1] + g \\ a[i-1,j-1] + \sigma(\ s[i],t[j]\ ) \\ a[i-1,j] + g \end{cases}$$



alignment



	-	c	a	t	g	t
-	0	-1	-2	-3	-4	-5
a	-1	-1	1	0	-1	-2
c	-2	1	0	0	-1	-2
g	-3	0	0	-1	2	1
c	-4	-1	-1	-1	1	1
t	-5	-2	-2	1	0	3
g	-6	-3	-3	0	3	2

$$\left\langle \begin{array}{c} \texttt{acgctg} \\ \texttt{catgt} \end{array} \right\rangle$$

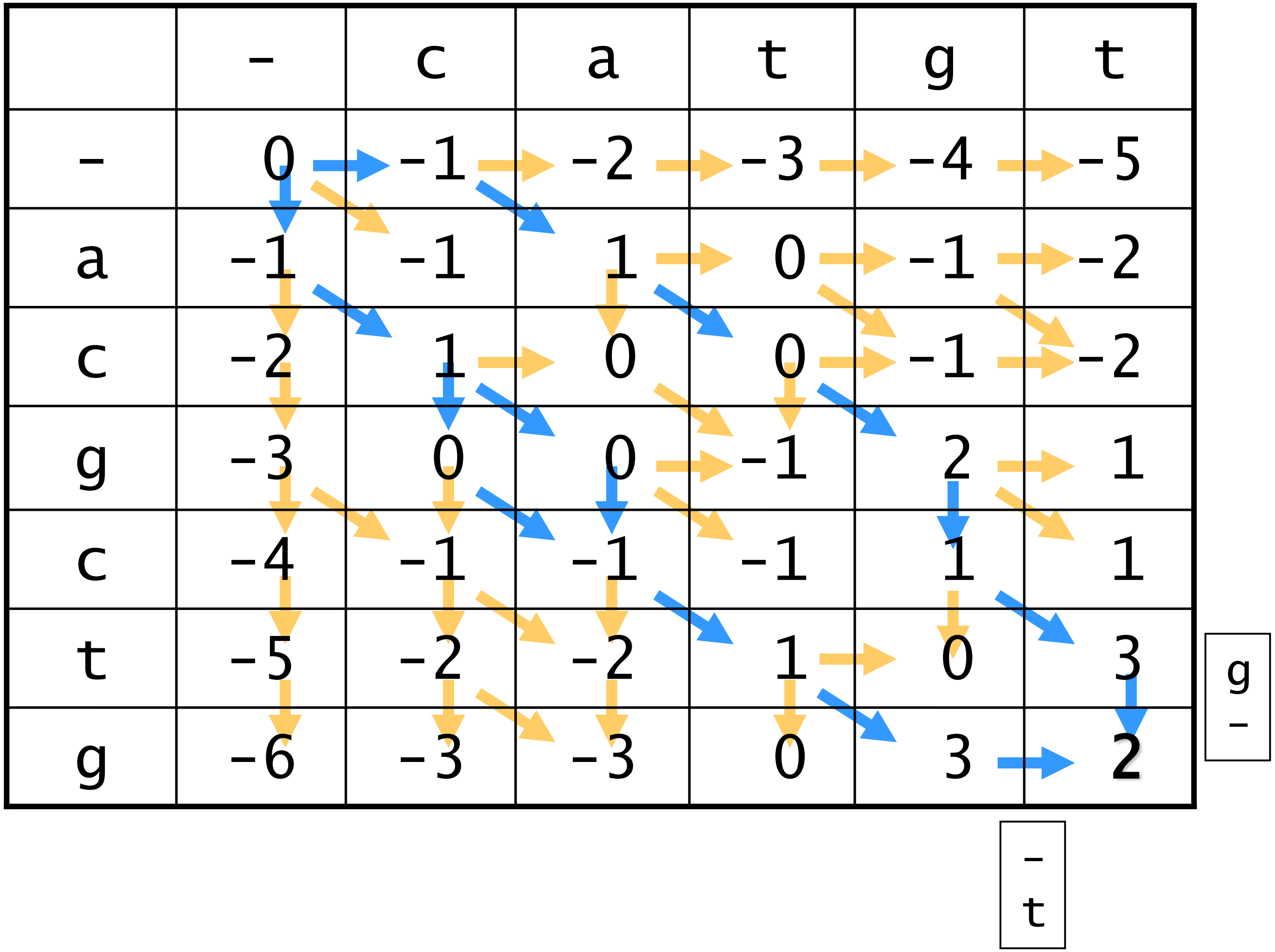
traceback

-acgctg
catg-t-

acgctg-
-ca-tgt

acgctg-
-c-atgt

⟨acgctg
catgt⟩



	-	c	a	t	g	t
-	0	-1	-2	-3	-4	-5
a	-1	-1	1	0	-1	-2
c	-2	1	0	0	-1	-2
g	-3	0	0	-1	2	1
c	-4	-1	-1	-1	1	1
t	-5	-2	-2	1	0	3
g	-6	-3	-3	0	3	2

Eddy S.R. (2004a)

```
 /* Recursion: the heart of the DP algorithm.*/
 /* Initialization. */

 S[0][0] = 0;
 for (i = 1; i <= M; i++)  S[i][0] = i * INDEL;
 for (j = 1; j <= N; j++)  S[0][j] = j * INDEL;

 /* Dynamic programming, global alignment (Needleman/Wunsch) recursion. */

for (i = 1; i <= M; i++)
   for (j = 1; j <= N; j++)
     {                                  /* case #1: i,j are aligned */
       if (x[i] == y[j]) S[i][j] = S[i-1][j-1] + MATCH;
       else              S[i][j] = S[i-1][j-1] + MISMATCH;

       sc = S[i-1][j] + INDEL;           /* case #2: i aligned to -  */
       if (sc > S[i][j]) S[i][j] = sc;

       sc = S[i][j-1] + INDEL;           /* case #3: j aligned to -  */
       if (sc > S[i][j]) S[i][j] = sc;
     }

 /* The result (optimal alignment score) is now in S[M][N]. */
```

program Eddy

```
[hoogeboo@tin 22:01 ~/colleges/mcb/programs] > ./align
Sequence X: TTCATA
Sequence Y: TGCTCGTA
Scoring system: 5 for match; -2 for mismatch; -6 for gap

Dynamic programming matrix:
                T    G    C    T    C    G    T    A
           0   -6  -12  -18  -24  -30  -36  -42  -48
     T    -6    5   -1   -7  -13  -19  -25  -31  -37
     T   -12   -1    3   -3   -2   -8  -14  -20  -26
     C   -18   -7   -3    8    2    3   -3   -9  -15
     A   -24  -13   -9    2    6    0    1   -5   -4
     T   -30  -19  -15   -4    7    4   -2    6    0
     A   -36  -25  -21  -10    1    5    2    0   11

Alignment:
X: T--TCATA
Y: TGCTCGTA
Optimum alignment score: 11
```

scoring and parameter choice

```
A       A
   vs.
A       C

A       A-
   vs.
C       -C

AT      -AT
   vs.
TA      TA-
```

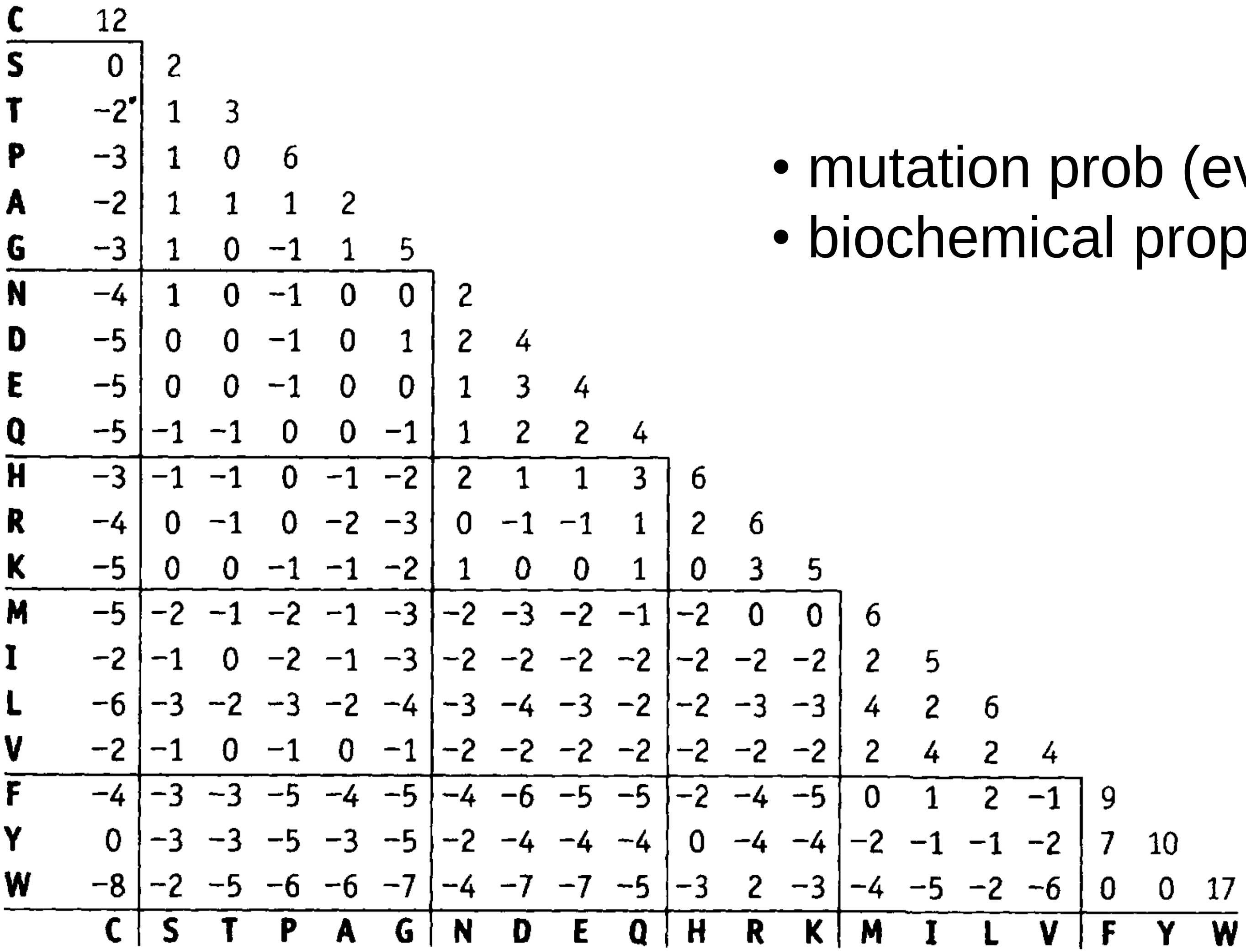
match	M	+2	+2
mismatch	m	-1	-2
gap	g	-1	-1

	A	C	G	T
A	91	-114	-31	-123
C		100	-125	-31
G			100	-114
T				91

gap 400 + 30k

ask your favourite molecular biologist !

PAM250 Matrix



C	12																			
S	0	2																		
T	-2	1	3																	
P	-3	1	0	6																
A	-2	1	1	1	2															
G	-3	1	0	-1	1	5														
N	-4	1	0	-1	0	0	2													
D	-5	0	0	-1	0	1	2	4												
E	-5	0	0	-1	0	0	1	3	4											
Q	-5	-1	-1	0	0	-1	1	2	2	4										
H	-3	-1	-1	0	-1	-2	2	1	1	3	6									
R	-4	0	-1	0	-2	-3	0	-1	-1	1	2	6								
K	-5	0	0	-1	-1	-2	1	0	0	1	0	3	5							
M	-5	-2	-1	-2	-1	-3	-2	-3	-2	-1	-2	0	0	6						
I	-2	-1	0	-2	-1	-3	-2	-2	-2	-2	-2	-2	-2	2	5					
L	-6	-3	-2	-3	-2	-4	-3	-4	-3	-2	-2	-3	-3	4	2	6				
V	-2	-1	0	-1	0	-1	-2	-2	-2	-2	-2	-2	-2	2	4	2	4			
F	-4	-3	-3	-5	-4	-5	-4	-6	-5	-5	-2	-4	-5	0	1	2	-1	9		
Y	0	-3	-3	-5	-3	-5	-2	-4	-4	-4	0	-4	-4	-2	-1	-1	-2	7	10	
W	-8	-2	-5	-6	-6	-7	-4	-7	-7	-5	-3	2	-3	-4	-5	-2	-6	0	0	17
	C	S	T	P	A	G	N	D	E	Q	H	R	K	M	I	L	V	F	Y	W

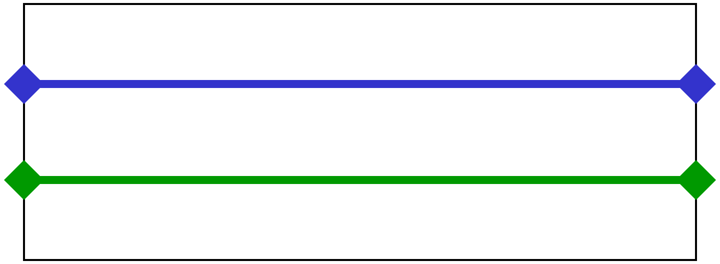
- mutation prob (evolution)
- biochemical properties

alignment

```
CAGCACTTGGATTCTCGG
CAGC-----G-T----GG
```

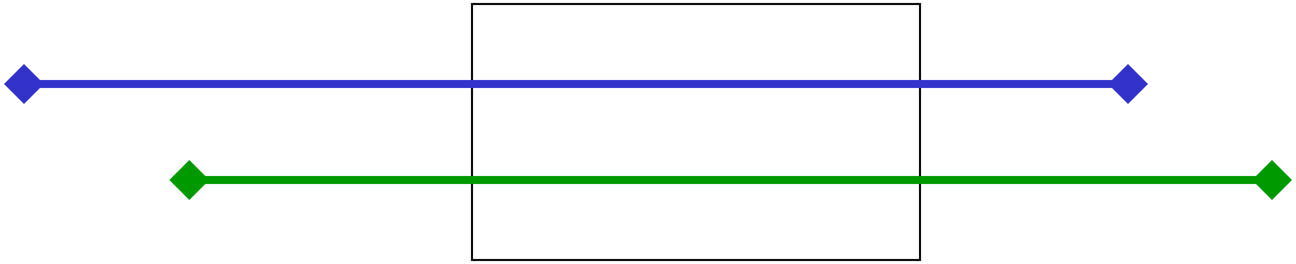
```
CAGCA-CTTGGATTCTCGG
---CAGCGTGG--------
```

variants of alignment



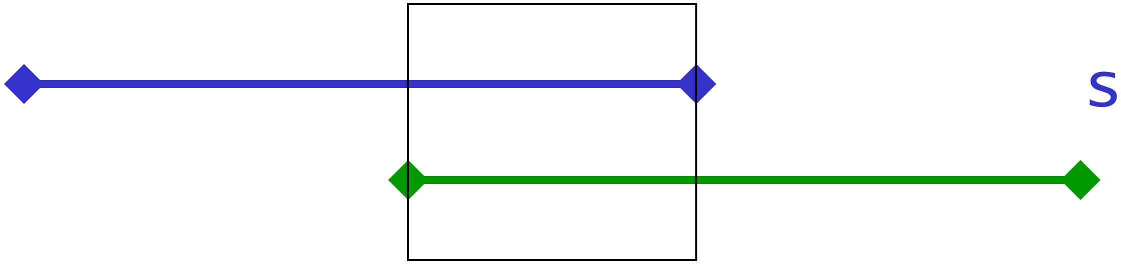
> global

Needleman & Wunsch

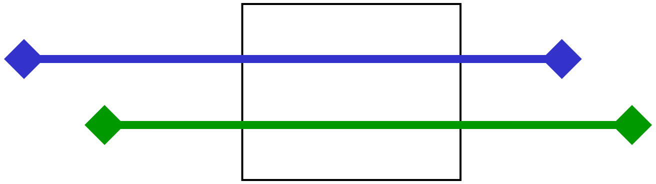


local

Smith & Waterman

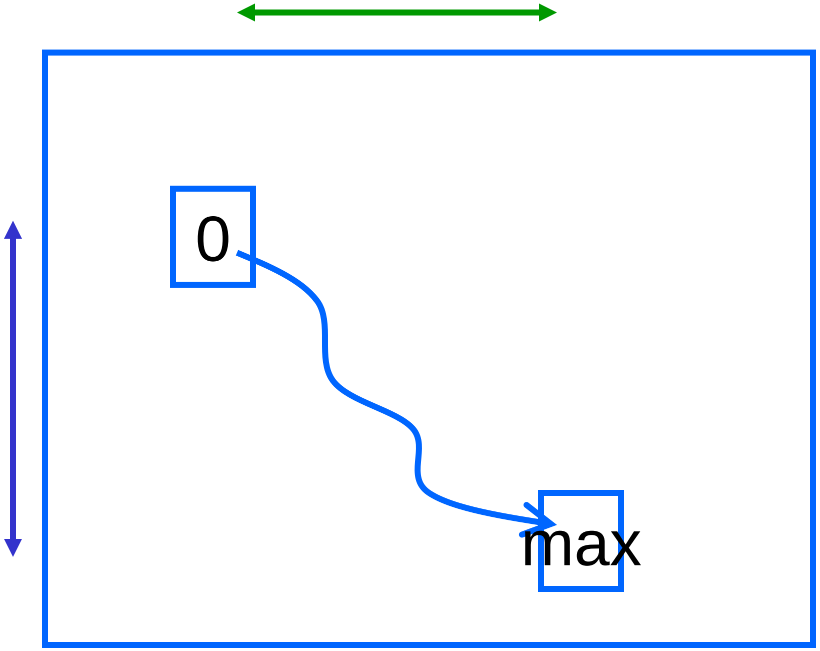


semi-global
end



local alignment

$$a[i,j] = \max \begin{cases} a[i,j-1] + g \\ a[i,j] + \sigma(s[i],t[j]) \\ a[i-1,j] + g \\ 0 \end{cases}$$



solution (*traceback*)
from max to 0

problem solved !?

much too slow

- long strings
- huge databases

heuristics

- **FASTA** along diagonal
- **BLAST** minimal close match

multiple alignment

(several strings)

NP complete … exponential

Basic Local Alignment Search Tools

query

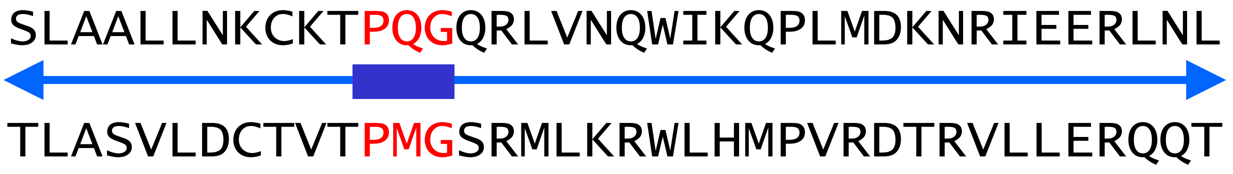
query word (W=3)

GSVEDTTGSQSLAALLNKCKTPQGQRLVNQWIKQPLMDKNRIEERLNL

neighbourhood words

PQG	18
PEG	15
PRG	14
PMG	13
PQA	12
PQN	12

threshold



SLAALLNKCKTPQGQRLVNQWIKQPLMDKNRIEERLNL

TLASVLDCTVTPMGSRMLKRWLHMPVRDTRVLLERQQT

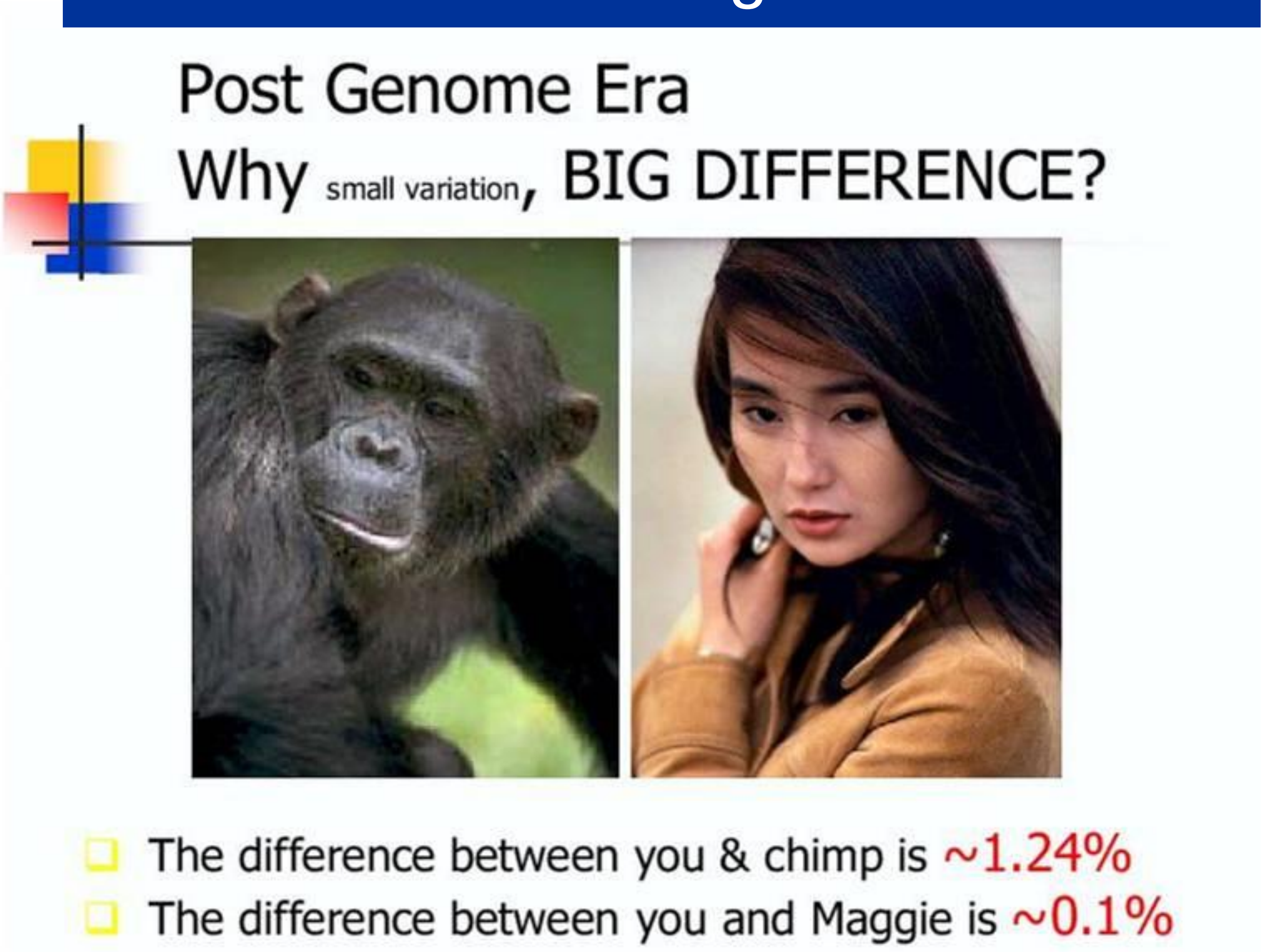
subject

high-scoring segment pair (HSP)

parameters:
W word size
T threshold
S score
E expected

om thuis over na te denken …

genetic variation



Post Genome Era

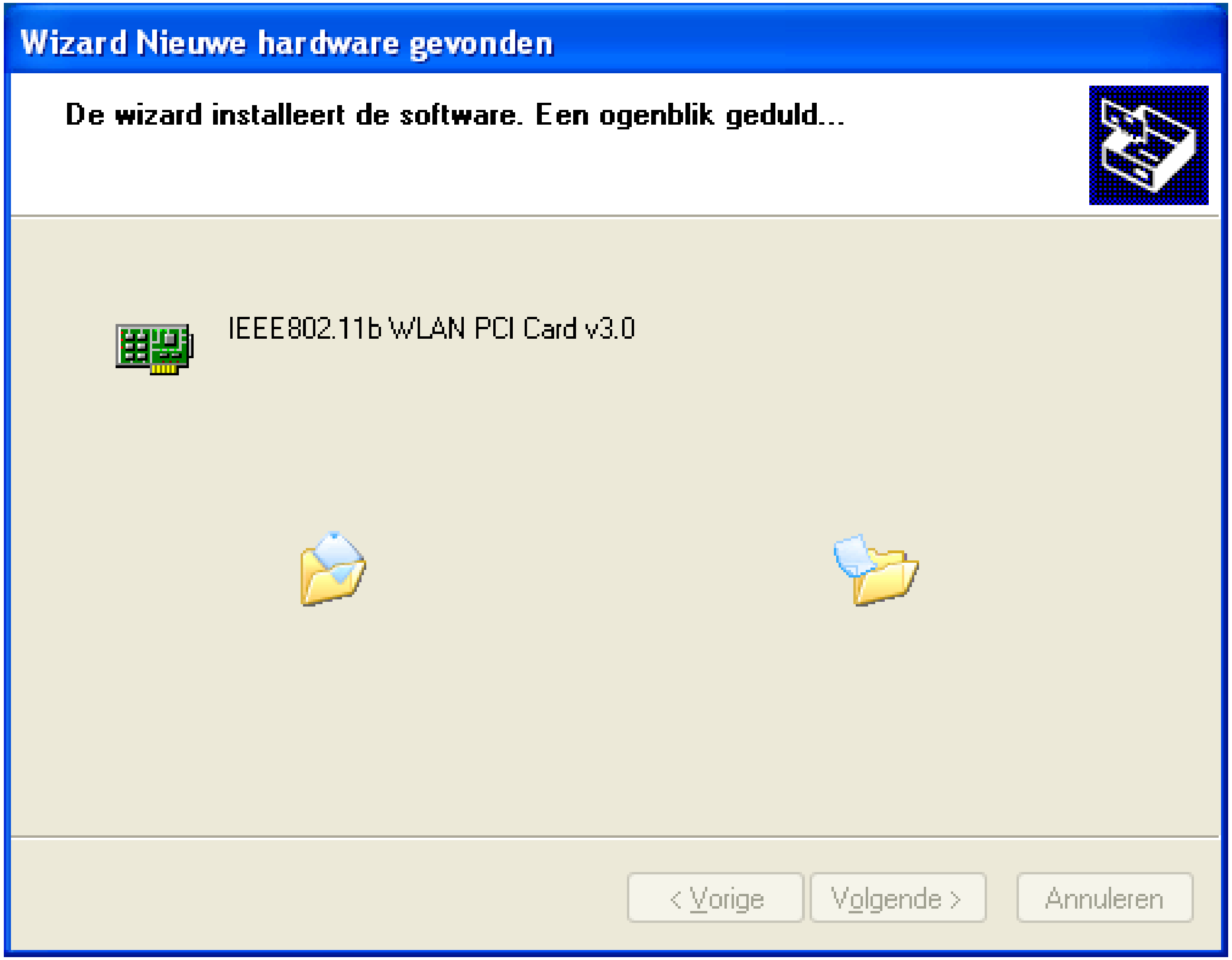
Why small variation, BIG DIFFERENCE?

- The difference between you & chimp is ~1.24%
- The difference between you and Maggie is ~0.1%

thanks google

BAVARIANBEER

nog even geduld …



Wizard Nieuwe hardware gevonden

De wizard installeert de software. Een ogenblik geduld...

IEEE802.11b WLAN PCI Card v3.0

som

```
1 + 2 * 3 - 4
```

BAPC 2006

BENELUX
ALGORITHM
PROGRAMMING
CONTEST
2006 LEIDEN

A. Geodes
B. Bowling
C. High spies
D. Digital Friends
E. The Bavarian Beer Party
F. Evacuation
G. Decompression
H. Rummikub
I. Make it Manhattan

website BAPC

programmeerwedstrijd voor teams

E. The Bavarian Beer Party

Description

The professors of the *Bayerische Mathematiker Verein* have their annual party in the local *Biergarten*. They are sitting at a round table each with his own pint of beer. As a ceremony each professor raises his pint and toasts one of the other guests in such a way that no arms cross.

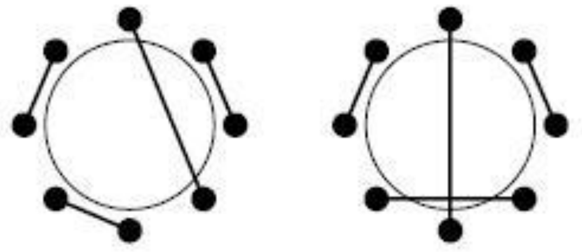


Figure 2: Toasting across a table with eight persons: no arms crossing(left), arms crossing(right)

We know that the professors like to toast with someone that is drinking the same brand of beer, and we like to maximize the number of pairs of professors toasting with the same brand, again without crossing arms. Write an algorithm to do this, keeping in mind that every professor should take part in the toasting.

E. The Bavarian Beer Party

Input

The first line of the input contains a single number: the number of test cases to follow. Each test case has the following format:

One line with an even number p, satisfying 2 ≤ p ≤ 1000: the number of participants,

One line with p integers (separated by single spaces) indicating the beer brands for the consecutive professors (in clockwise order, starting at an arbitrary position). Each value is between 1 and 100 (boundaries included).

Output

For every test case in the input, the output should contain a single number on a single line: the maximum number of non-intersecting toasts of the same beer brand for this test case.

Sample Input

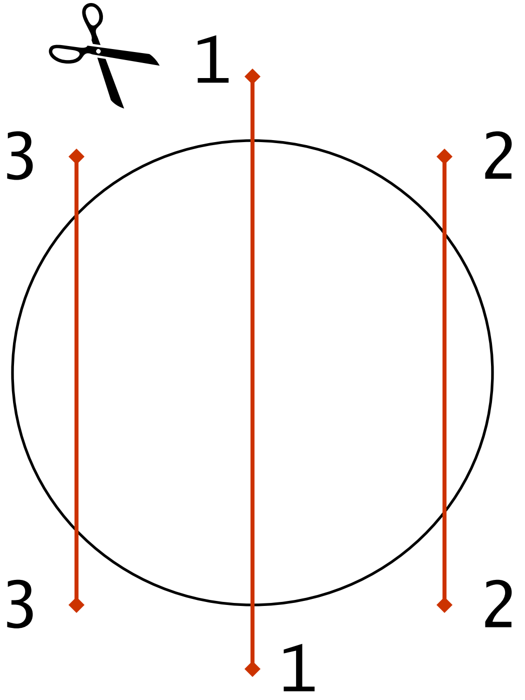
```
2
6
1 2 2 1 3 3
22
1 7 1 2 4 2 4 9 1 1 9 4 5 9 4 5 6 9 2 1 2 9
```

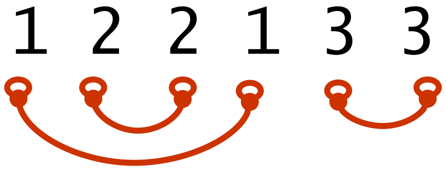
Sample Output

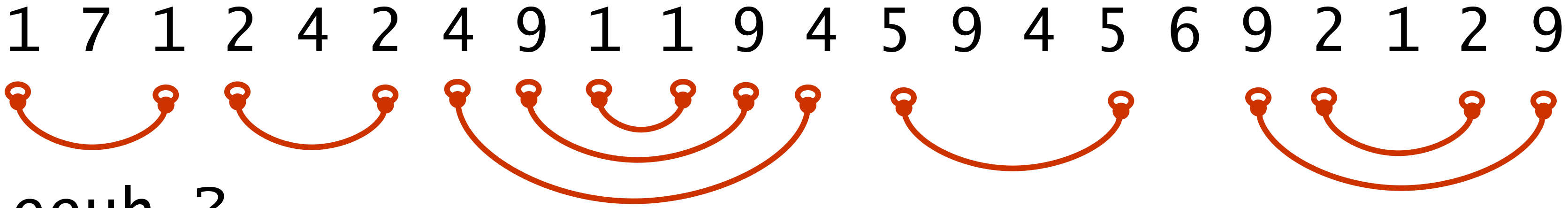
```
3
6
```

bijvoorbeeld

```
Sample Input
2
6
1 2 2 1 3 3
22
1 7 1 2 4 2 4 9 1 1 9 4 5 9 4 5 6 9
2 1 2 9
Sample Output
3
6
```

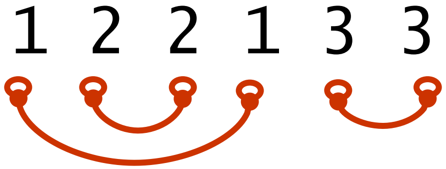


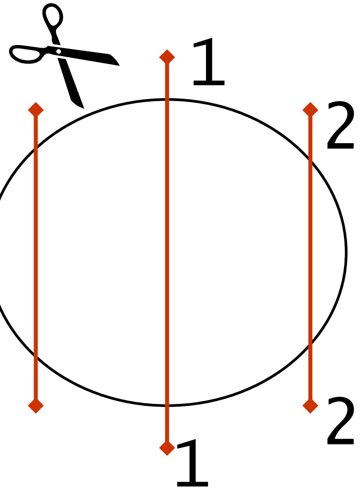


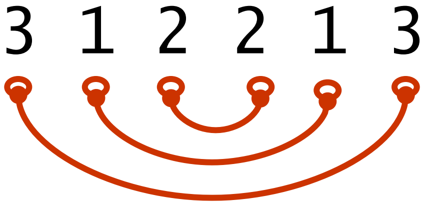


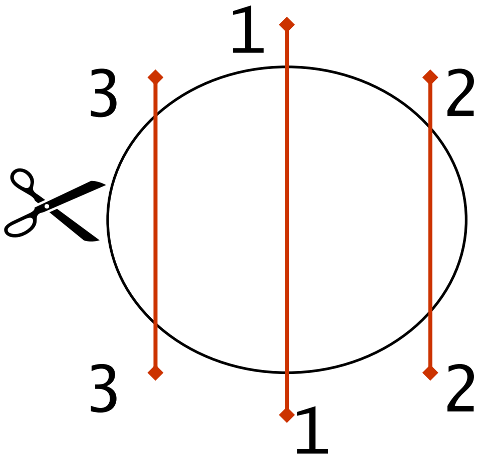
eeuh ?

open knippen

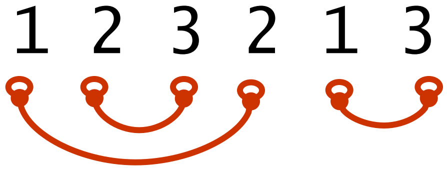


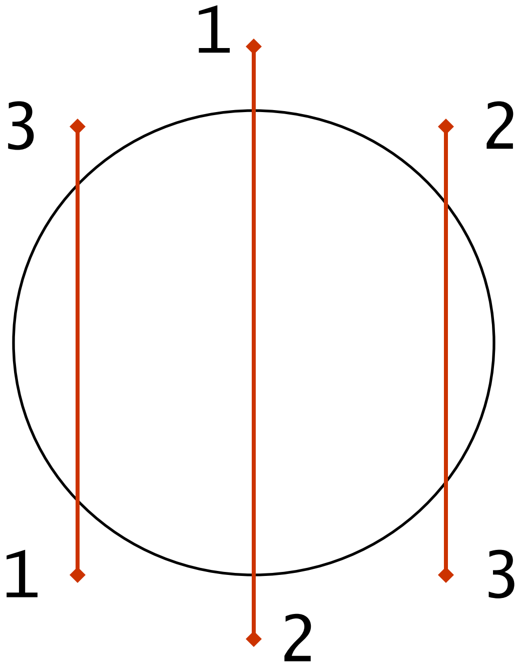






alle mogelijkheden?

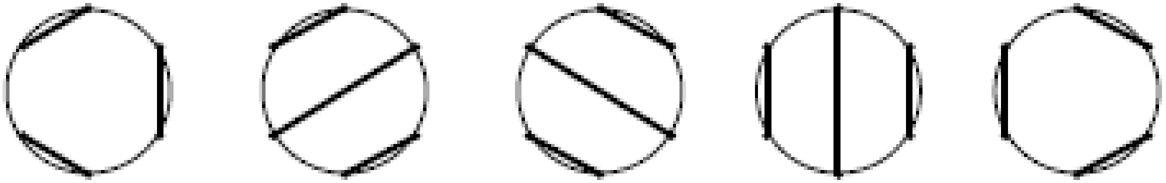




alle mogelijkheden?

handshaking problem :- Catalan numbers

For $n = 3$, there are 5 ways:

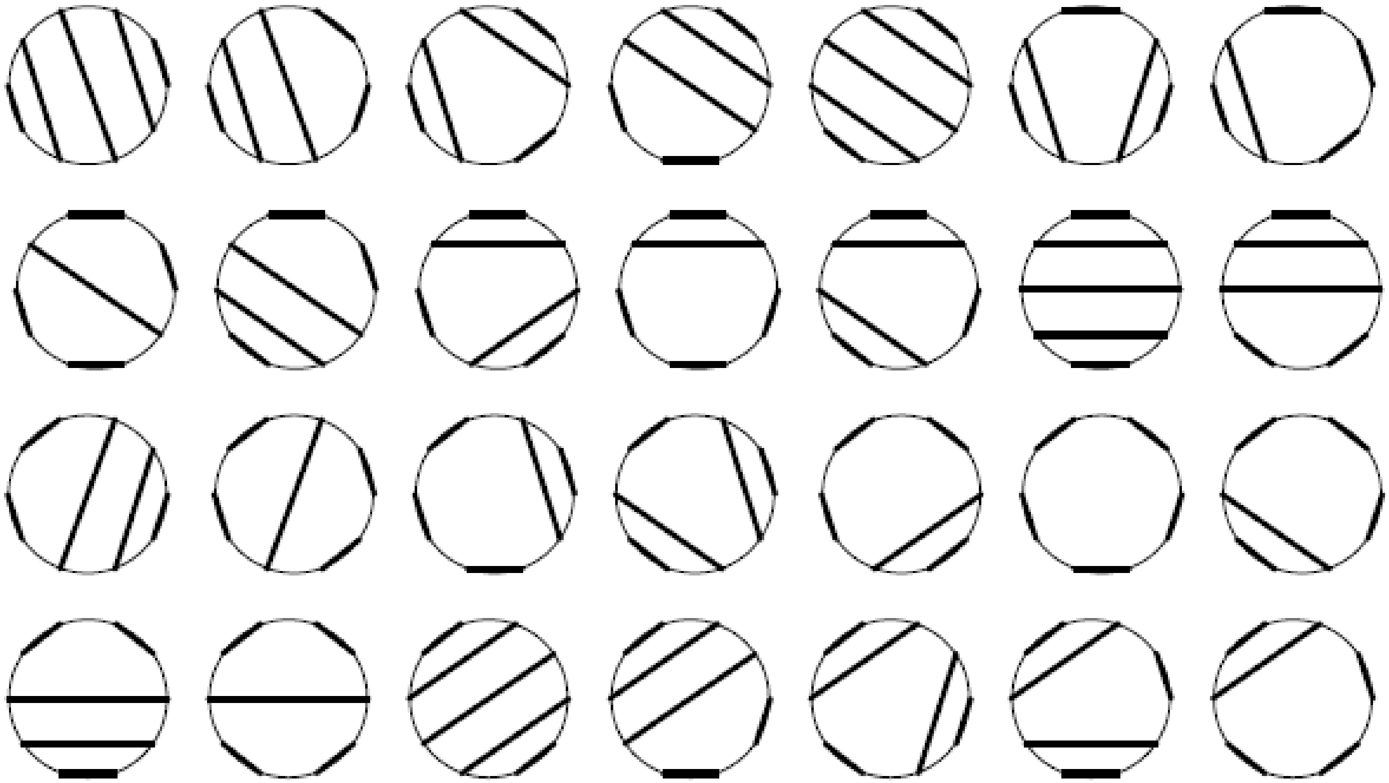


opgave: teken bijbehorende *smiling faces*

alle mogelijkheden?

handshaking problem :- Catalan numbers

For $n = 5$, there are 42 ways:



http://www.maths.usyd.edu.au/u/kooc/catalan.html »
klik op ‘smiling faces’ en ‘handshaking problem’

hoeveel mogelijkheden?

‘Sloane’ On-Line Encyclopedia of Integer Sequences

http://oeis.org/search?q=1+2+5+14+42

A000108 Catalan numbers: C(n) = binomial(2n,n)/(n+1) = (2n)!/(n!(n+1)!). Also called Segner numbers.

- The solution to Schroeder's first problem. A very large number of combinatorial interpretations are known - see references, esp. Stanley, Enumerative Combinatorics.
- Number of ways to insert n pairs of parentheses in a word of n+1 letters. E.g. for n=3 there are 5 ways: **((ab)(cd))**, **(((ab)c)d)**, **((a(bc))d)**, **(a((bc)d))**, **(a(b(cd)))**.
- Shifts one place left when convolved with itself.
- **Ways of joining 2n points on a circle to form n nonintersecting chords**.
- Arises in Schubert calculus - see Sottile reference.
- Inverse Euler transform of sequence is A022553.
- With interpolated zeros, the inverse binomial transform of the Motzkin numbers A001006.

[… etc …]

Eugène Charles Catalan

Eugène Charles Catalan (Brugge, 30 mei 1814 – Luik, 14 februari 1894) was een Belgisch wiskundige

wiki:catalan

hoeveel?

1	1
2	2
3	5
4	14
5	42
6	132
7	429
8	1430
9	4862
10	16796
11	58786
12	208012
13	742900
14	2674440
15	9694845
16	35357670

Sloane: Catalan numbers

http://www.research.att.com/~njas/sequences/A000108 »

$$\frac{1}{n+1}\binom{2n}{n} = \frac{(2n)!}{(n+1)!n!}$$

(precies)

$$\frac{4^n}{(n+1)\sqrt{\pi n}}$$

(ongeveer)

a(n) ~ 4^n / (sqrt(Pi*n) * (n+1))

hoeveel?

0	1	2	3	4	5	6
1	1	2	5	14	42	132

	# paar	aantal
( ) x x x x x x x x	(0)4	1 x14
(x x) x x x x x x	(1)3	1 x 5
(x x x x) x x x x	(2)2	2 x 2
(x x x x x x) x x	(3)1	5 x 1
(x x x x x x x x)	(4)0	14 x 1 +
		42

“Shifts one place left when convolved with itself”

hoeveel → algoritme

```
bav(1,10) = maximum van …

( ) x x x x x x x x                 bav(3,10) +match(1,2)
( x x ) x x x x x x   bav(2,3) +bav(5,10) +match(1,4)
( x x x x ) x x x x   bav(2,5) +bav(7,10) +match(1,6)
( x x x x x x ) x x   bav(2,7) +bav(9,10) +match(1,8)
( x x x x x x x x )   bav(2,9)            +match(1,10)
```

dynamisch programmeren

hoeveel → algoritme

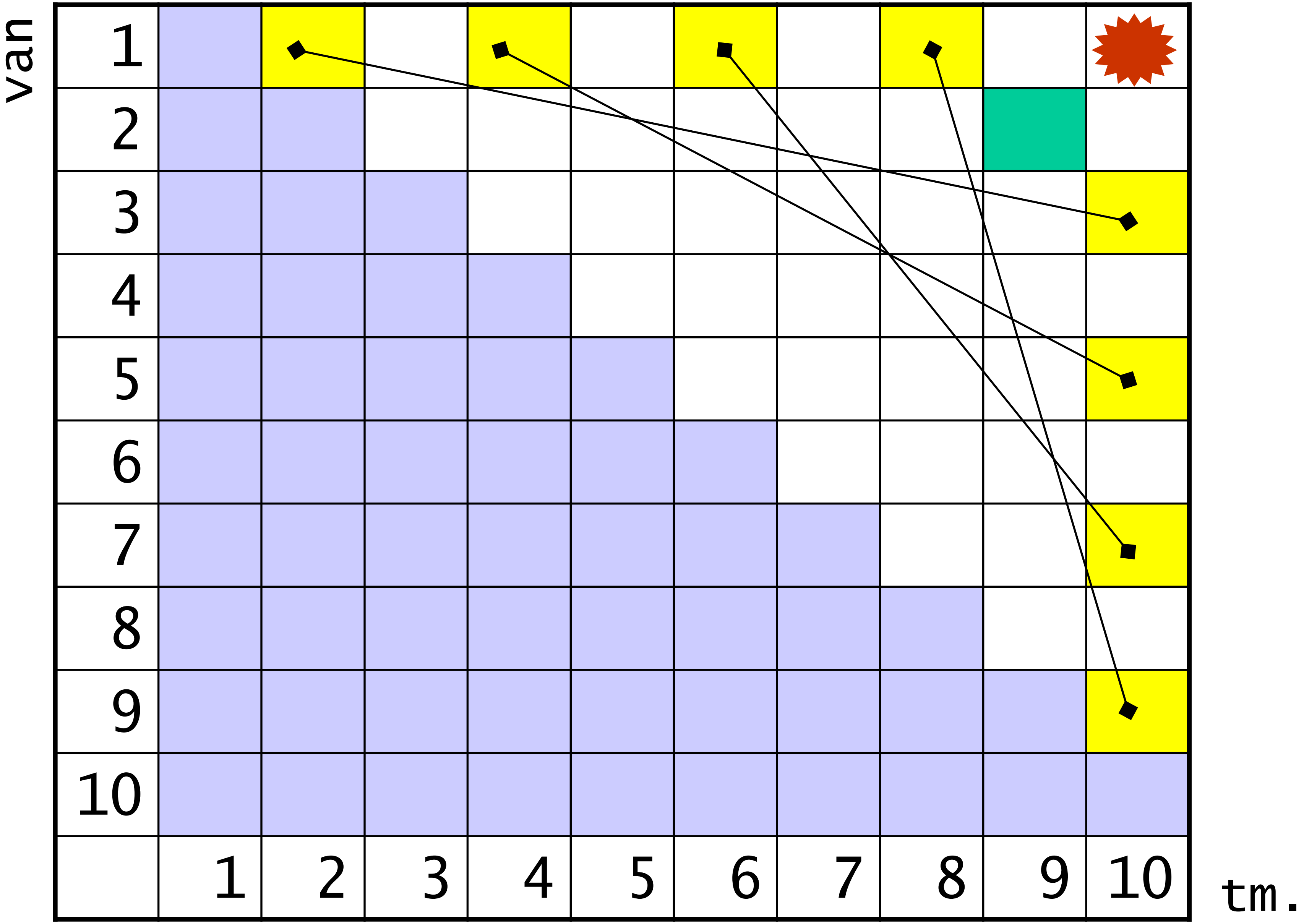
[verwarrend]
eigenlijk hetzelfde
algo maar match op
andere plek berekend

bav(1 maximum van …

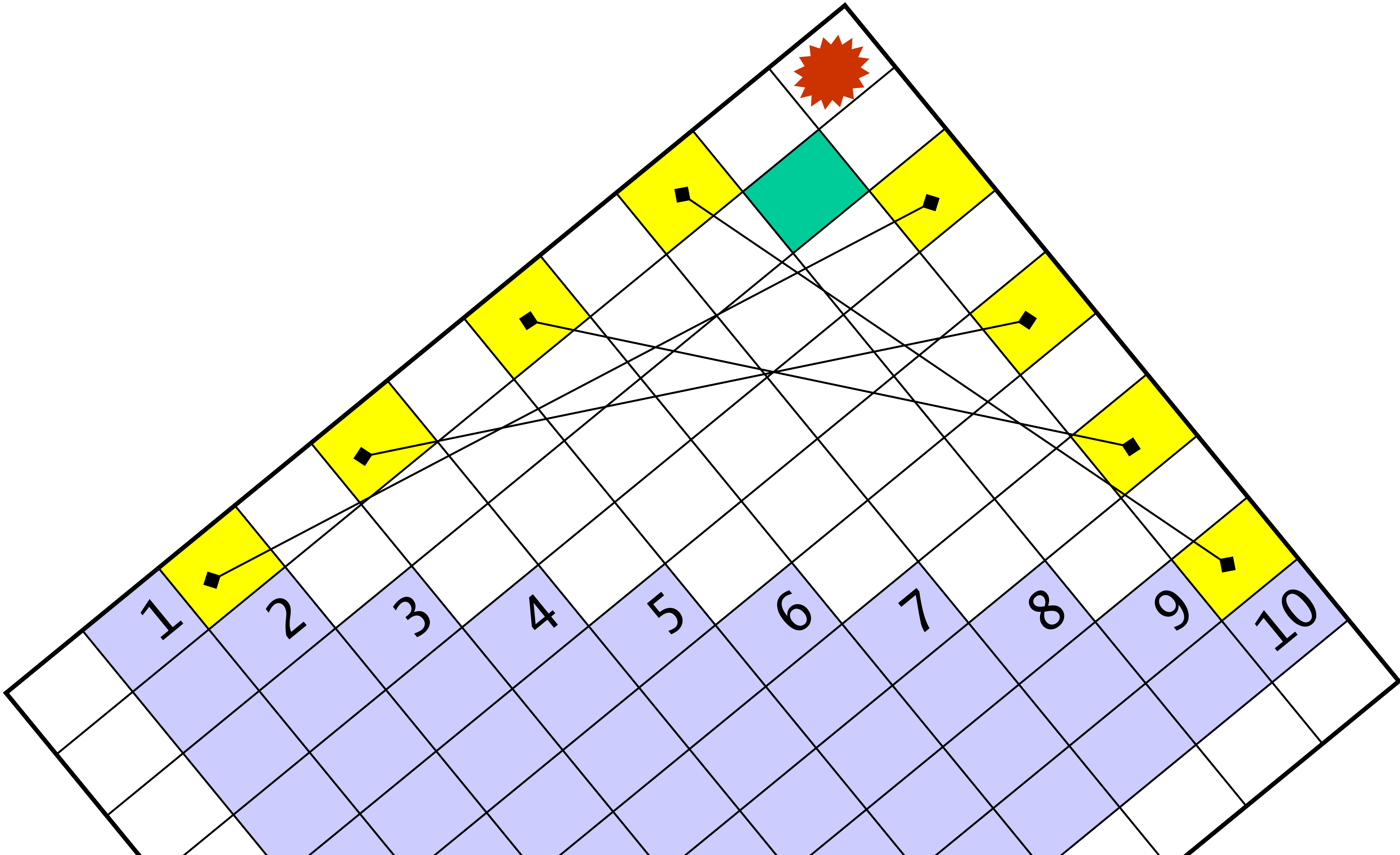
```
x x|x x x x x x x x    bav(1,2)+bav(3,10)

x x x x|x x x x x x    bav(1,4)+bav(5,10)

x x x x x x|x x x x    bav(1,6)+bav(7,10)

x x x x x x x x|x x    bav(1,8)+bav(9,10)

( x x x x x x x x )    bav(2,9)+match(1,10)
```

dynamisch programmeren

bav(1,10) = maximum van …

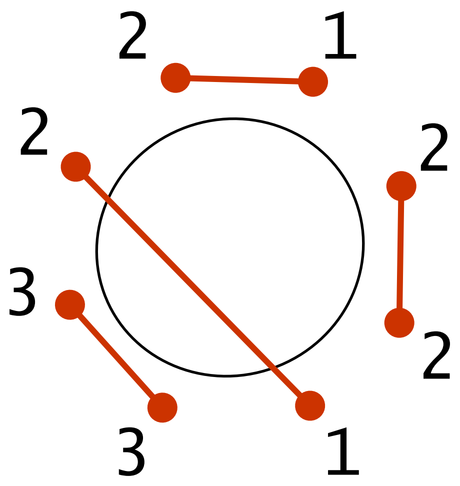


bav(1,10) = maximum van …

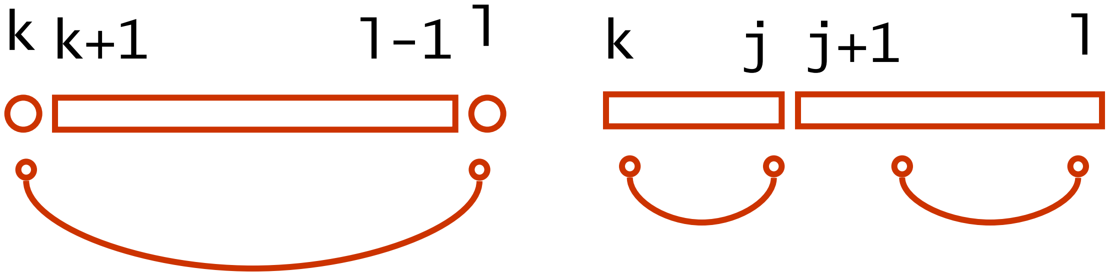


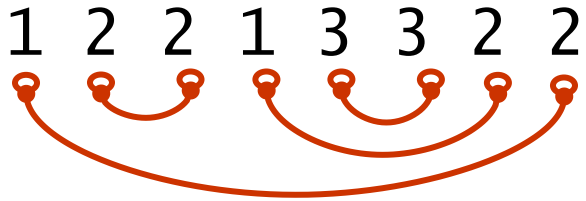
Bavarian Beer Party

optimale score interval [k,l]



bav[k,l]





bav[k,l] = max

- bav[k+1,l-1] + match(k,l)
- (∀j) bav[k,j]+bav[j+1,l]

dynamic programming

even length segments

Bavarian Beer Party

```
int proost ( int totaal )
{ int i, eerst, laatst, aantal, best, som;

     // initialisatie: paren
  for (i=1; i<totaal; i++)
  { bav[i][i+1] = ( brand[i] == brand[i+1] ); }

  if (totaal == 2) { return bav[1][2] ; }

  for (aantal=4; aantal<=totaal; aantal+=2)
  {    // dynamische loop: 'aantal' toastende professoren
       // aantal == laatst - eerst + 1
    for (eerst=1; eerst<totaal-aantal; eerst++)
    {
      laatst = eerst + aantal -1;
      best = (brand[eerst] == brand[laatst]);
      best += bav[eerst+1][laatst-1];
      for (i=eerst+1; i<=laatst-2; i=i+2)
      {   // alleen even matches!
        som = bav[eerst][i] + bav[i+1][laatst];
        if (som > best)  best = som;
      }
      bav[eerst][laatst] = best;
    }
  }// loop
  return bav[1][totaal] ;
}// proost
```

complexiteit

$$\frac{4^n}{(n+1)\sqrt{\pi n}}$$

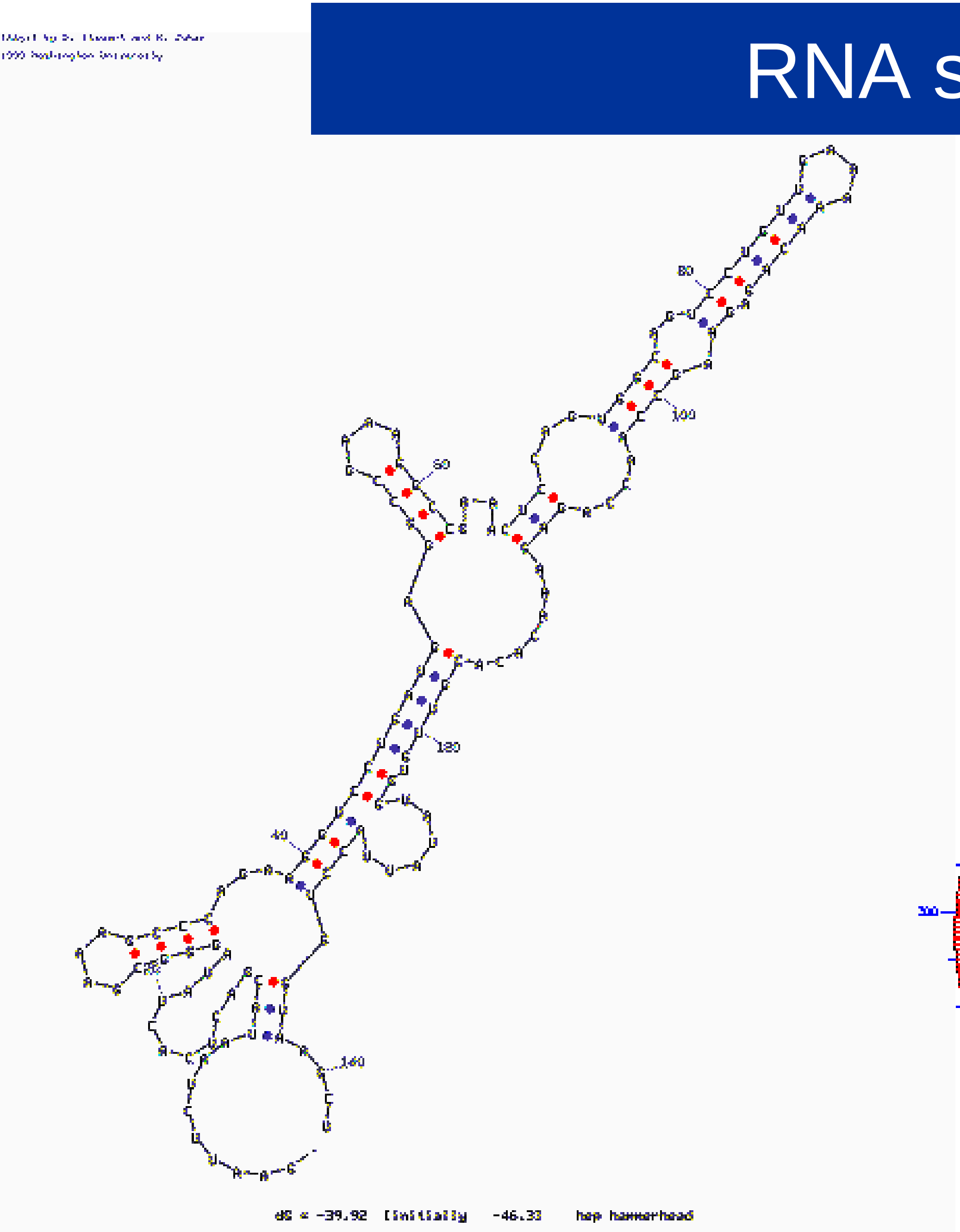
Hoeveel werk is dit dan nu?

4^{100} = 1606938044258990275541962092341162602522202993782792835301376

100^3 = 1000000

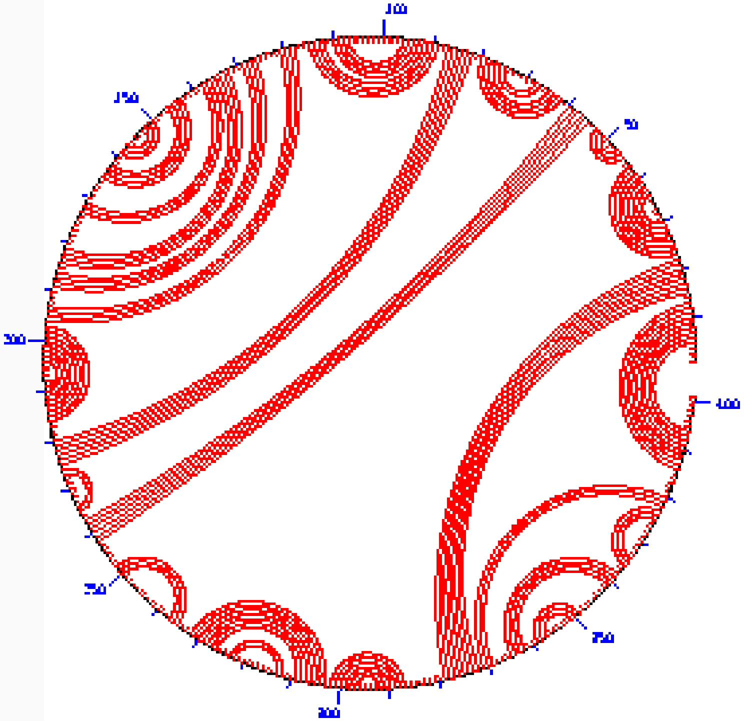
lekker belangrijk …

RNA secondary structure


mfold by D. Stewart and M. Zuker
© 2000 Washington University
80
60
100
180
40
140
dG = -39.92 [initially -46.33] hep hammerhead

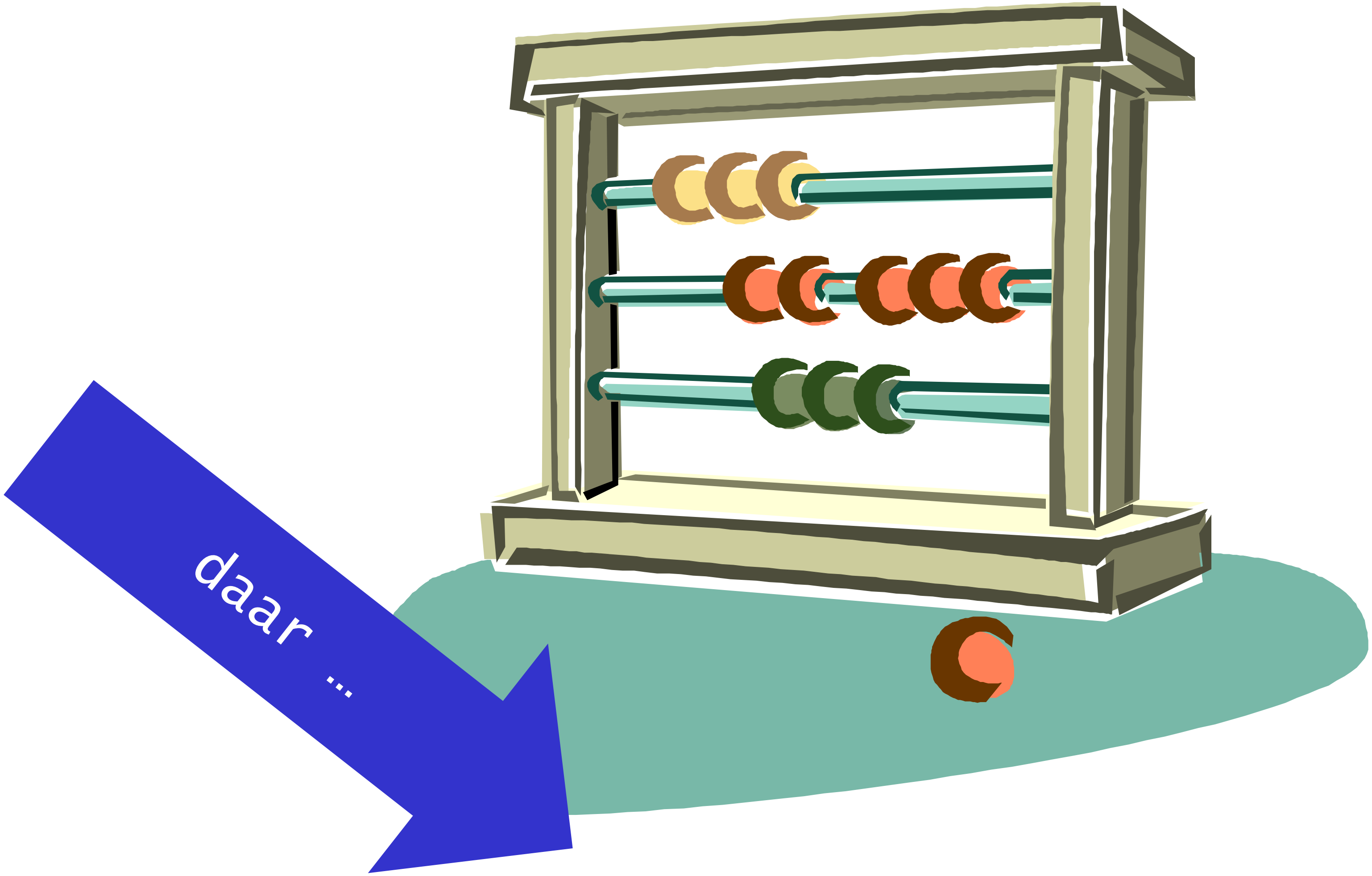
A Adenine
C Cytosine
G Guanine
U Uracil

A=U G≡C



PRACTICUM

practicum



daar ...

http://www.liacs.nl/home/hoogeboo/praatjes/lapp-top/ »

verse computers …

ULCN account gebruiken:

Windows
Linux

(wat je zelf wil)

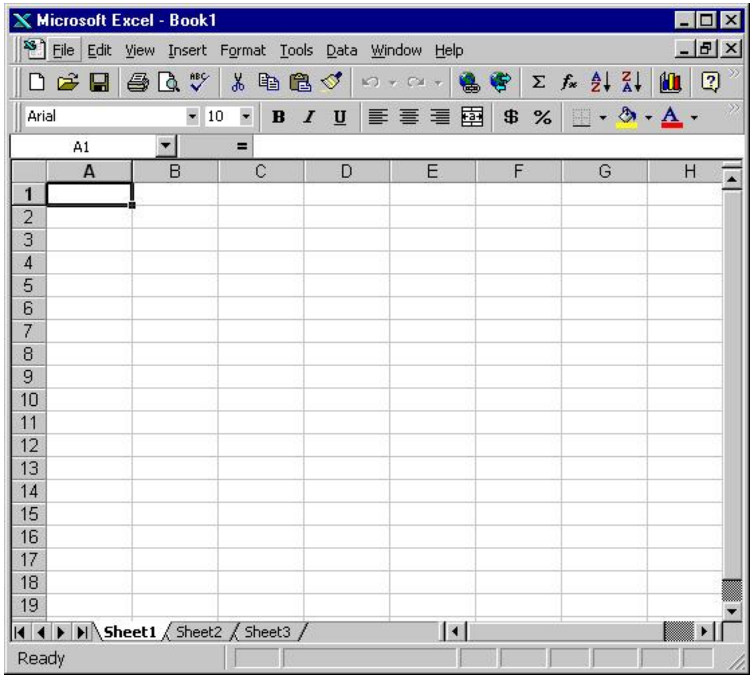
‘rekenblad’ Excel

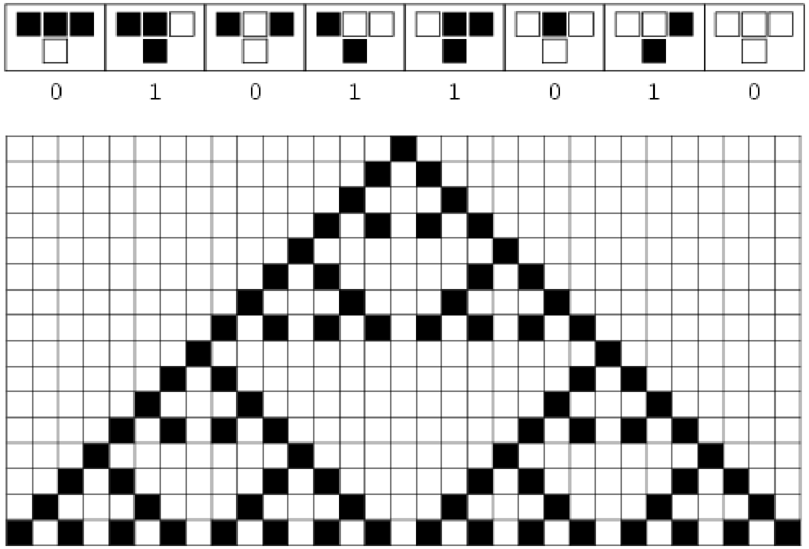
laboratorium

excel

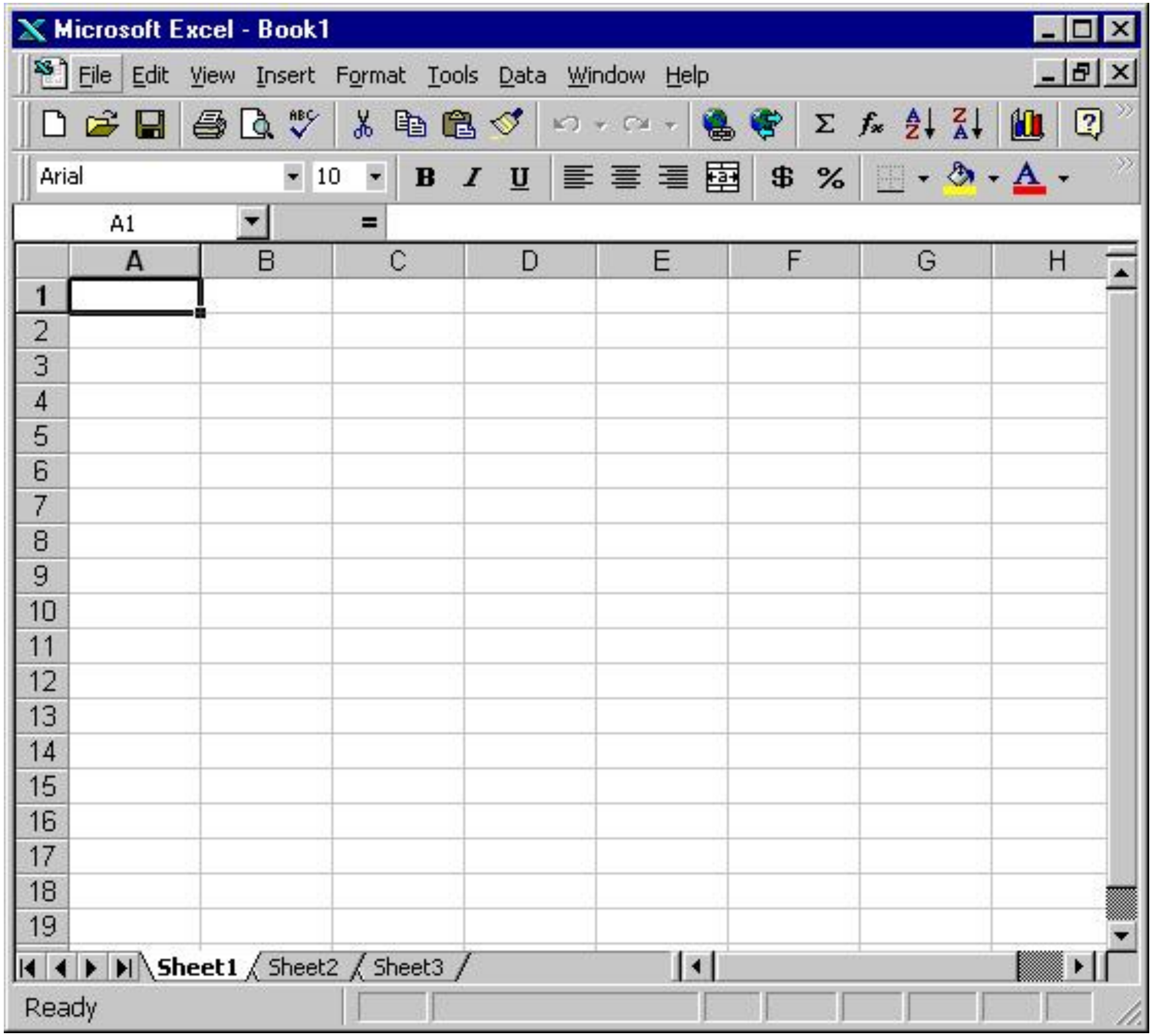
of C++ natuurlijk

- driehoek van Pascal
- cellulaire automaat
- alignment





excel



gegevens
adressen
formules

excel

gegevens

getallen, strings,
datum, geld

adressen

A1	relatief
$A$1	absoluut

formules

```
=A2+B1
IF(A2=B1,1,2)
SUM(B9:B12)
```

adressen

A1	relatief
$A$1	absoluut

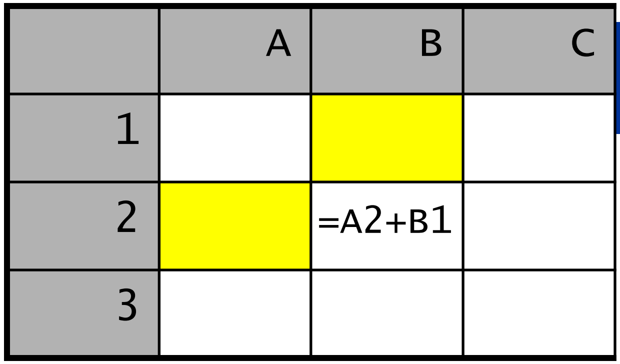
	A	B	C	D	E	F
1		🟨				
2	🟨	=A2+B1			🟨	
3				🟨	=D3+E2	

adressen

A1	relatief
$A$1	absoluut

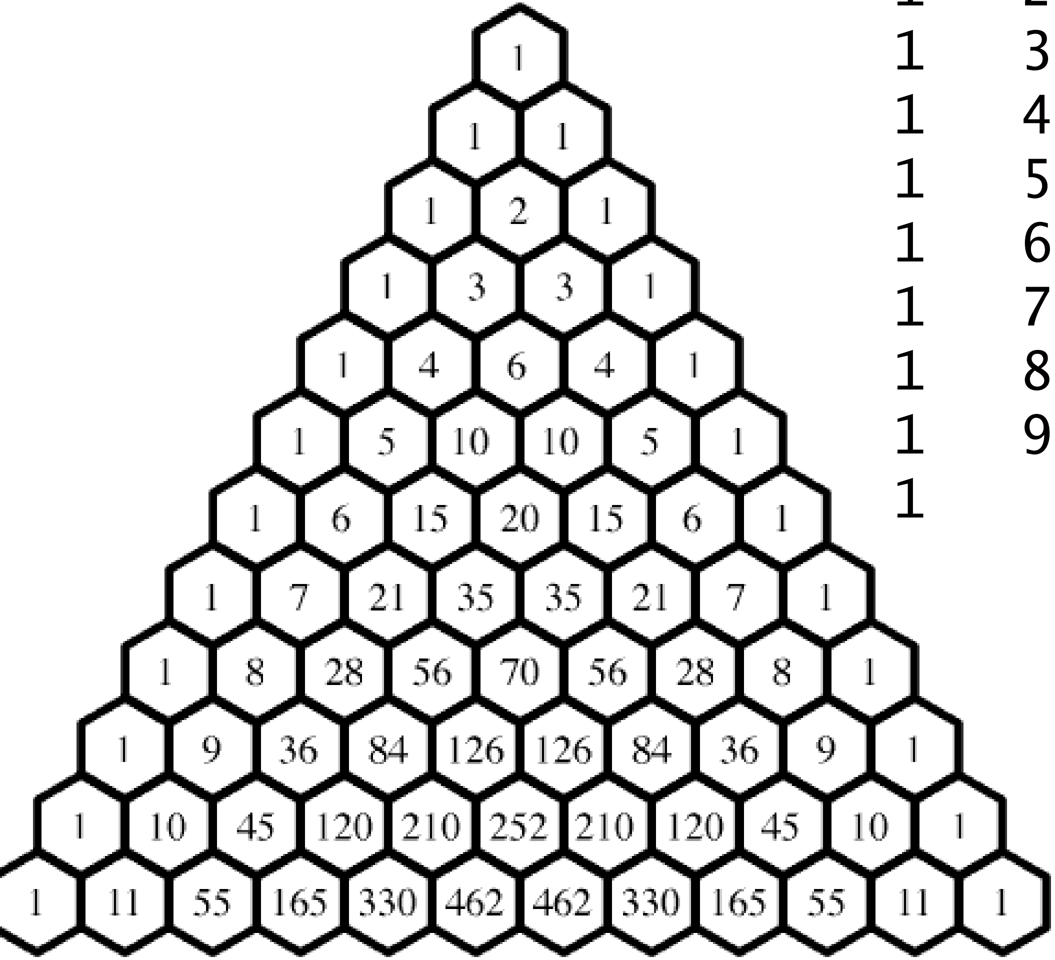
	A	B	C	D	E	F
1		🟨			🟩	
2	🟨	=$A2+B$1				
3	🟩				=$A3+E$1	

driehoek van Pascal



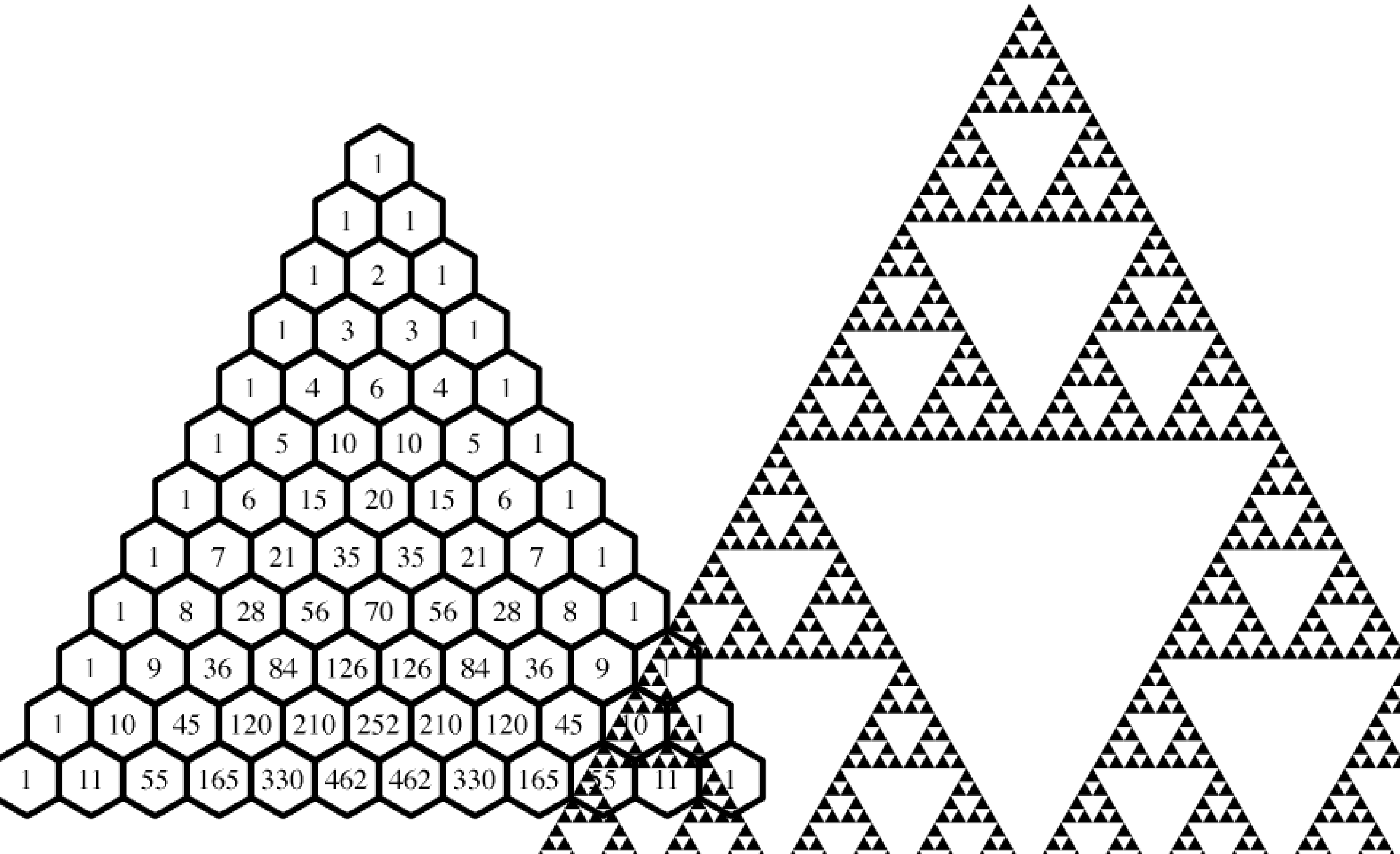
	A	B	C
1			
2		=A2+B1	
3			

1	1	1	1	1	1	1	1
1	2	3	4	5	6	7	8
1	3	6	10	15	21	28	36
1	4	10	20	35	56	84	
1	5	15	35	70	126		
1	6	21	56	126			
1	7	28	84				
1	8	36					
1	9						
1							



1
1 1
1 2 1
1 3 3 1
1 4 6 4 1
1 5 10 10 5 1
1 6 15 20 15 6 1
1 7 21 35 35 21 7 1
1 8 28 56 70 56 28 8 1
1 9 36 84 126 126 84 36 9 1
1 10 45 120 210 252 210 120 45 10 1
1 11 55 165 330 462 462 330 165 55 11 1

welke getallen gekleurd ?



1

1 1

1 2 1

1 3 3 1

1 4 6 4 1

1 5 10 10 5 1

1 6 15 20 15 6 1

1 7 21 35 35 21 7 1

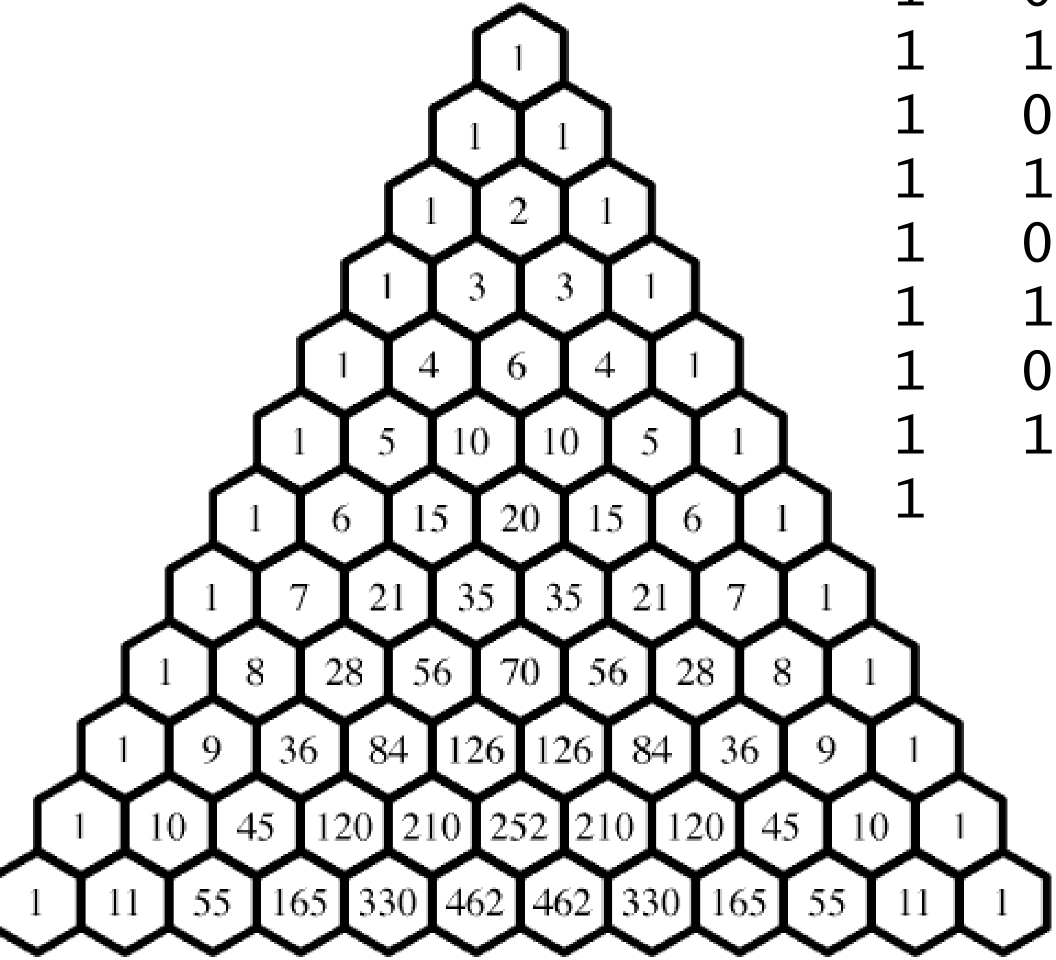
1 8 28 56 70 56 28 8 1

1 9 36 84 126 126 84 36 9 1

1 10 45 120 210 252 210 120 45 10 1

1 11 55 165 330 462 462 330 165 55 11 1

welke getallen gekleurd ?



1
1 1
1 2 1
1 3 3 1
1 4 6 4 1
1 5 10 10 5 1
1 6 15 20 15 6 1
1 7 21 35 35 21 7 1
1 8 28 56 70 56 28 8 1
1 9 36 84 126 126 84 36 9 1
1 10 45 120 210 252 210 120 45 10 1
1 11 55 165 330 462 462 330 165 55 11 1

```
1   1   1   1   1   1   1   1
1   0   1   0   1   0   1   0
1   1   0   0   1   1   0   0
1   0   0   0   1   0   0
1   1   1   1   0   0
1   0   1   0   0
1   1   0   0
1   0   0
1   1
1
```

exclusive or

even / oneven

x	y	x+y-2xy
0	0	0
0	1	1
1	0	1
1	1	0

∆ van Pascal

waarom eigenlijk niet gewoon ‘rechtop’?
(dat lukt met ‘gaten’ tussen de waardes)

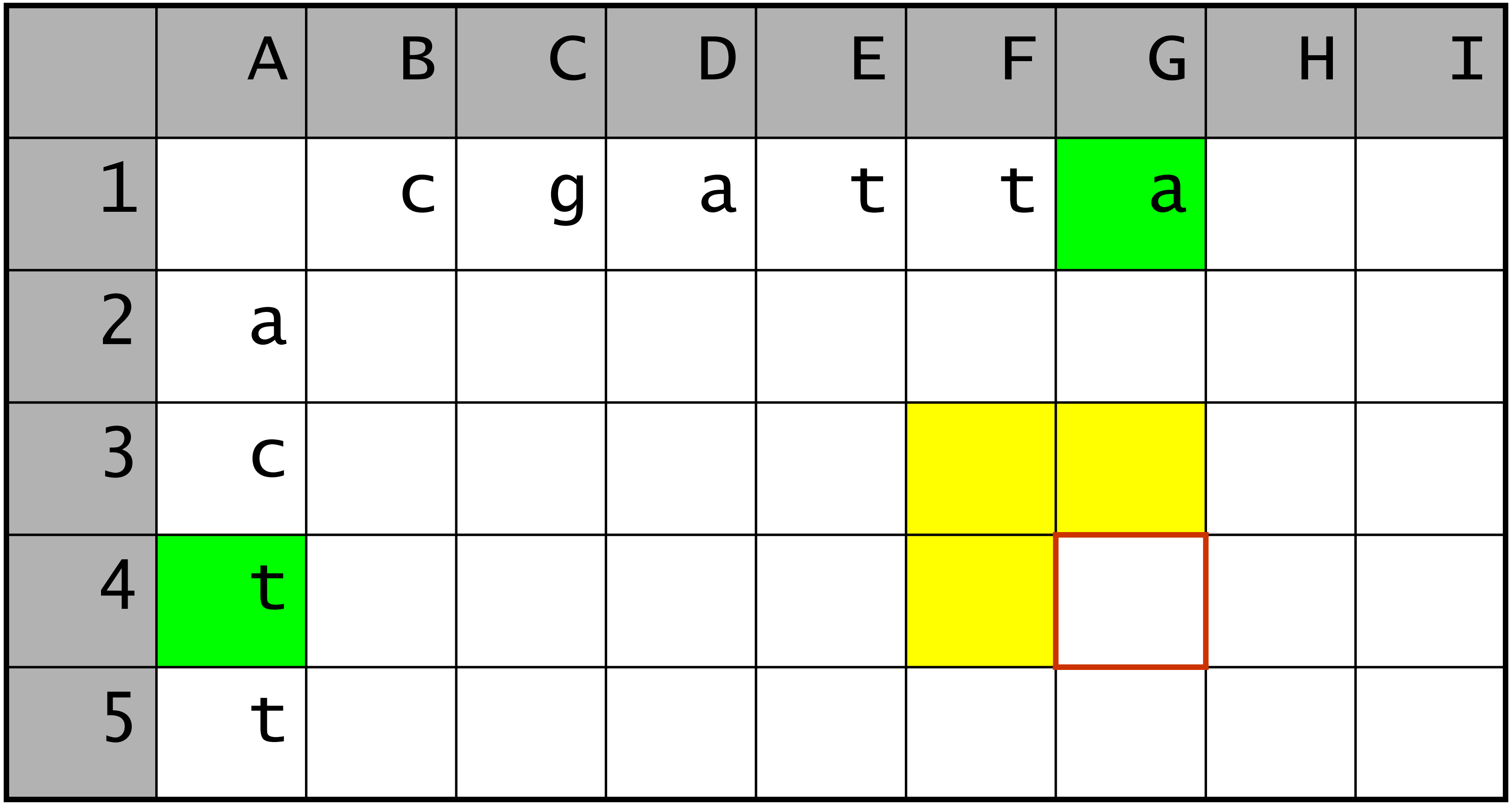
0	0	0	0	0	0	1	0	0	0	0	0	0
0	0	0	0	0	1	0	1	0	0	0	0	0
0	0	0	0	1	0	2	0	1	0	0	0	0
0	0	0	1	0	3	0	3	0	1	0	0	0
0	0	1	0	4	0	6	0	4	0	1	0	0
0	1	0	5	0	10	0	10	0	5	0	1	0
1	0	6	0	15	0	20	0	15	0	6	0	1

zie cellulaire automaat
‘rule 90’
(extra excel oefening)

A1		C1
	=A1+C1	

alignment

excel - alignment

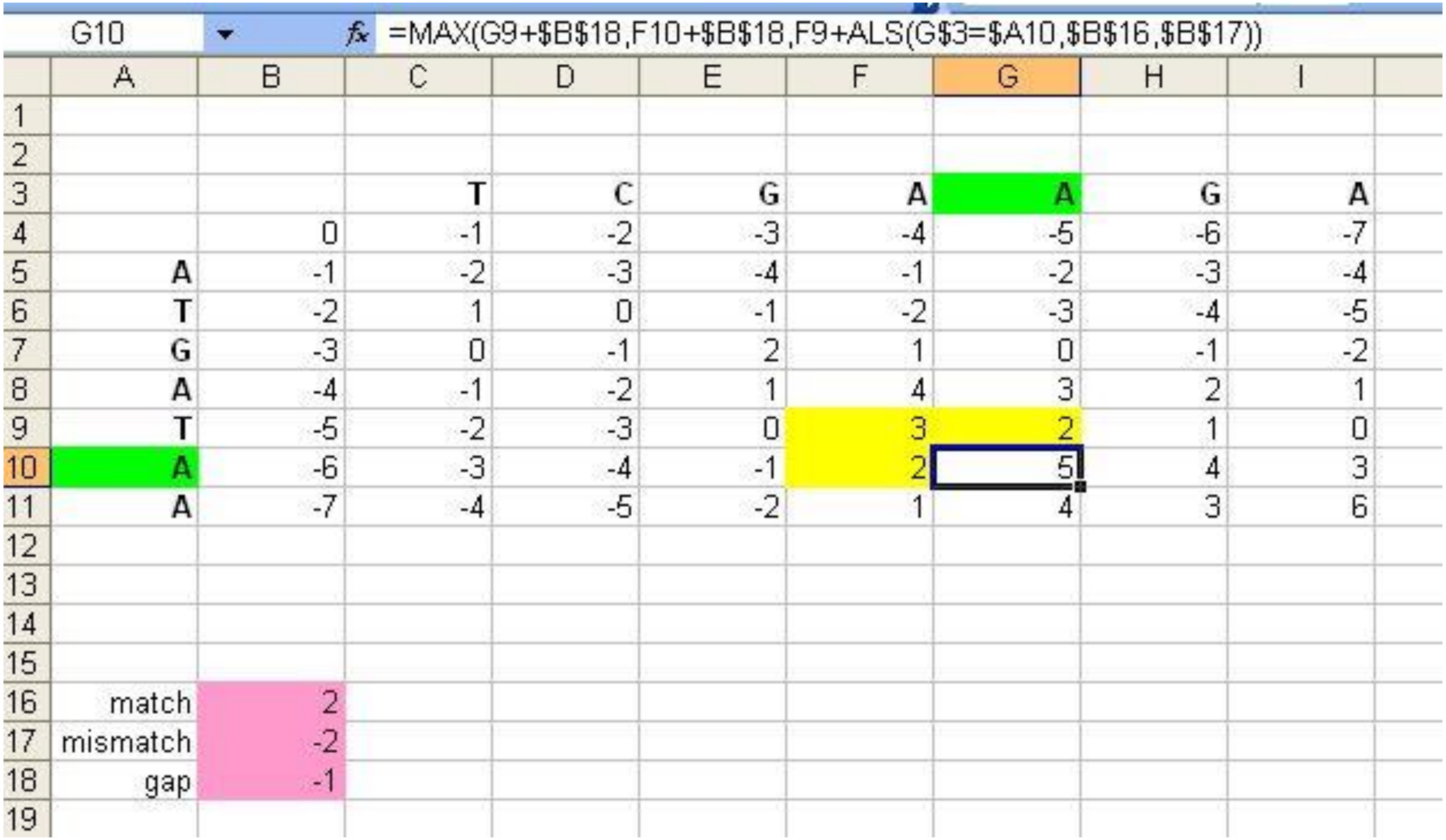


	A	B	C	D	E	F	G	H	I
1		c	g	a	t	t	a		
2	a								
3	c								
4	t								
5	t								

🟨 MAX *drie buren* ± gap / (mis)match

🟩 IF/ALS *letters* → match/mismatch

excel - alignment

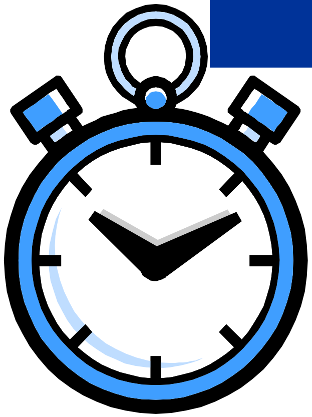


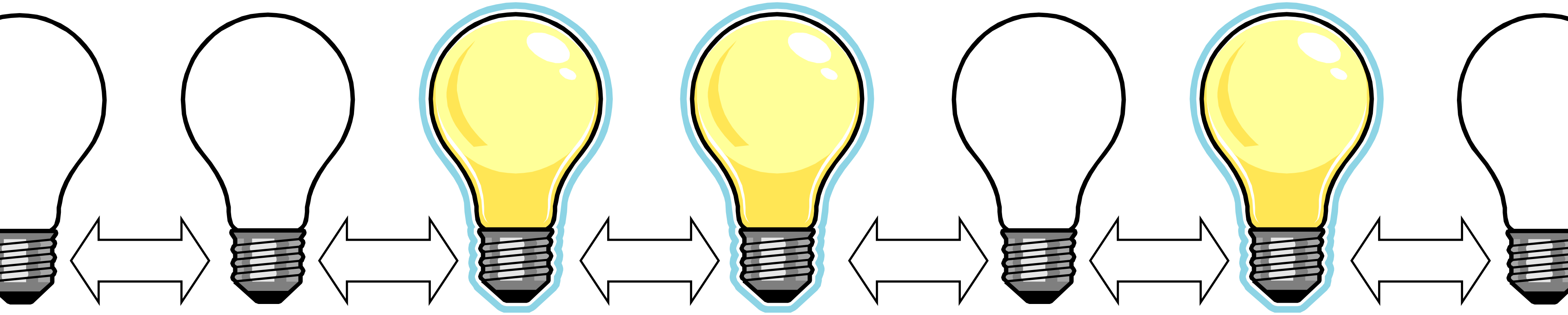
G10 =MAX(G9+$B$18,F10+$B$18,F9+ALS(G$3=$A10,$B$16,$B$17))

	A	B	C	D	E	F	G	H	I
1									
2									
3			T	C	G	A	A	G	A
4		0	-1	-2	-3	-4	-5	-6	-7
5	A	-1	-2	-3	-4	-1	-2	-3	-4
6	T	-2	1	0	-1	-2	-3	-4	-5
7	G	-3	0	-1	2	1	0	-1	-2
8	A	-4	-1	-2	1	4	3	2	1
9	T	-5	-2	-3	0	3	2	1	0
10	A	-6	-3	-4	-1	2	5	4	3
11	A	-7	-4	-5	-2	1	4	3	6
12									
13									
14									
15									
16	match	2							
17	mismatch	-2							
18	gap	-1							
19									

cellulaire automaat

cellulaire automaat

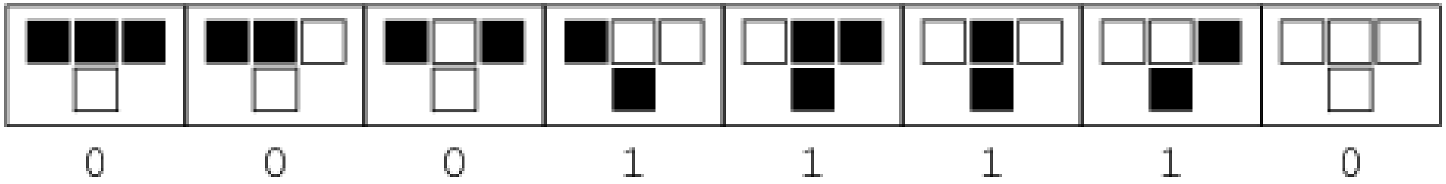


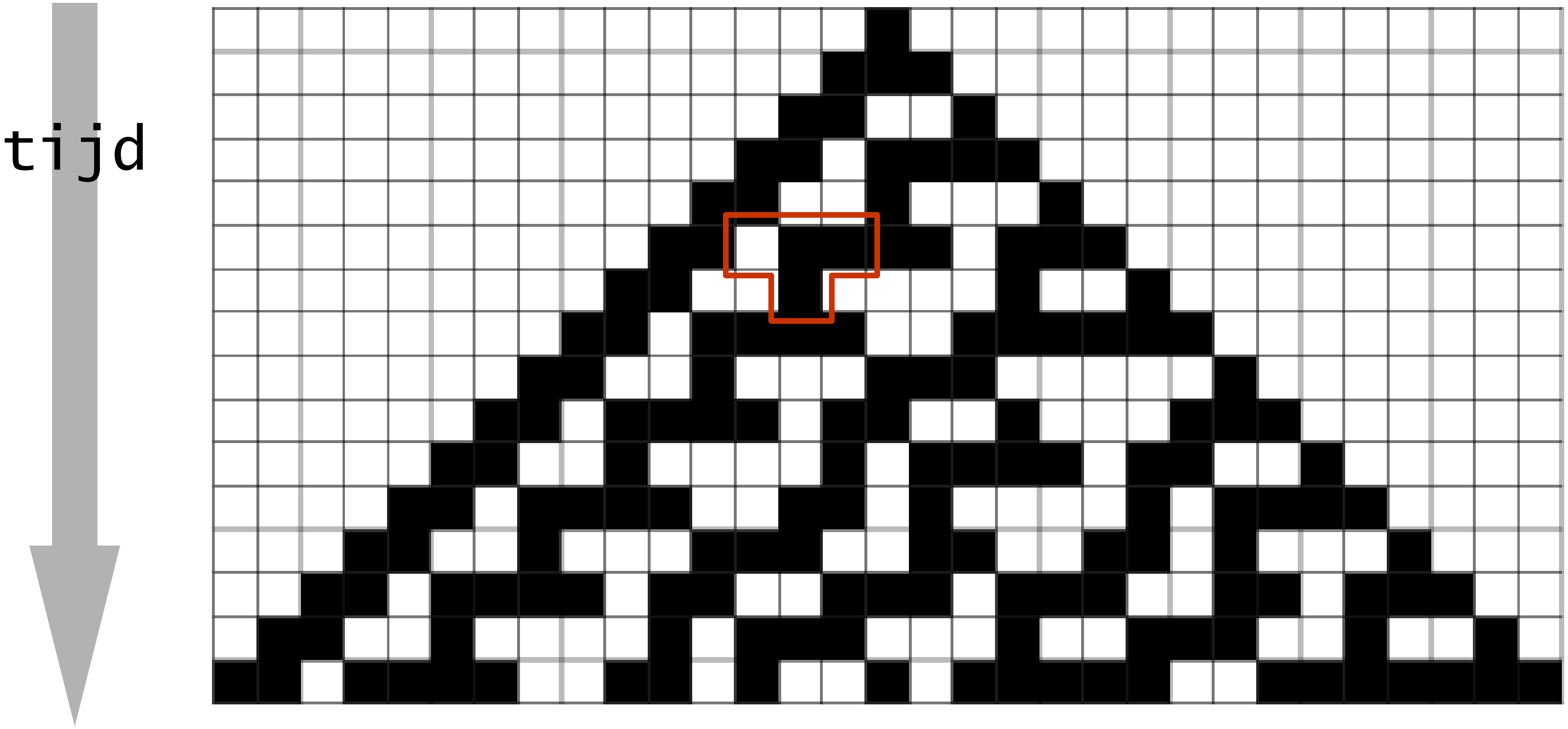


http://mathworld.wolfram.com/CellularAutomaton.html

http://nl.wikipedia.org/wiki/Elementaire_cellulaire_automaat »

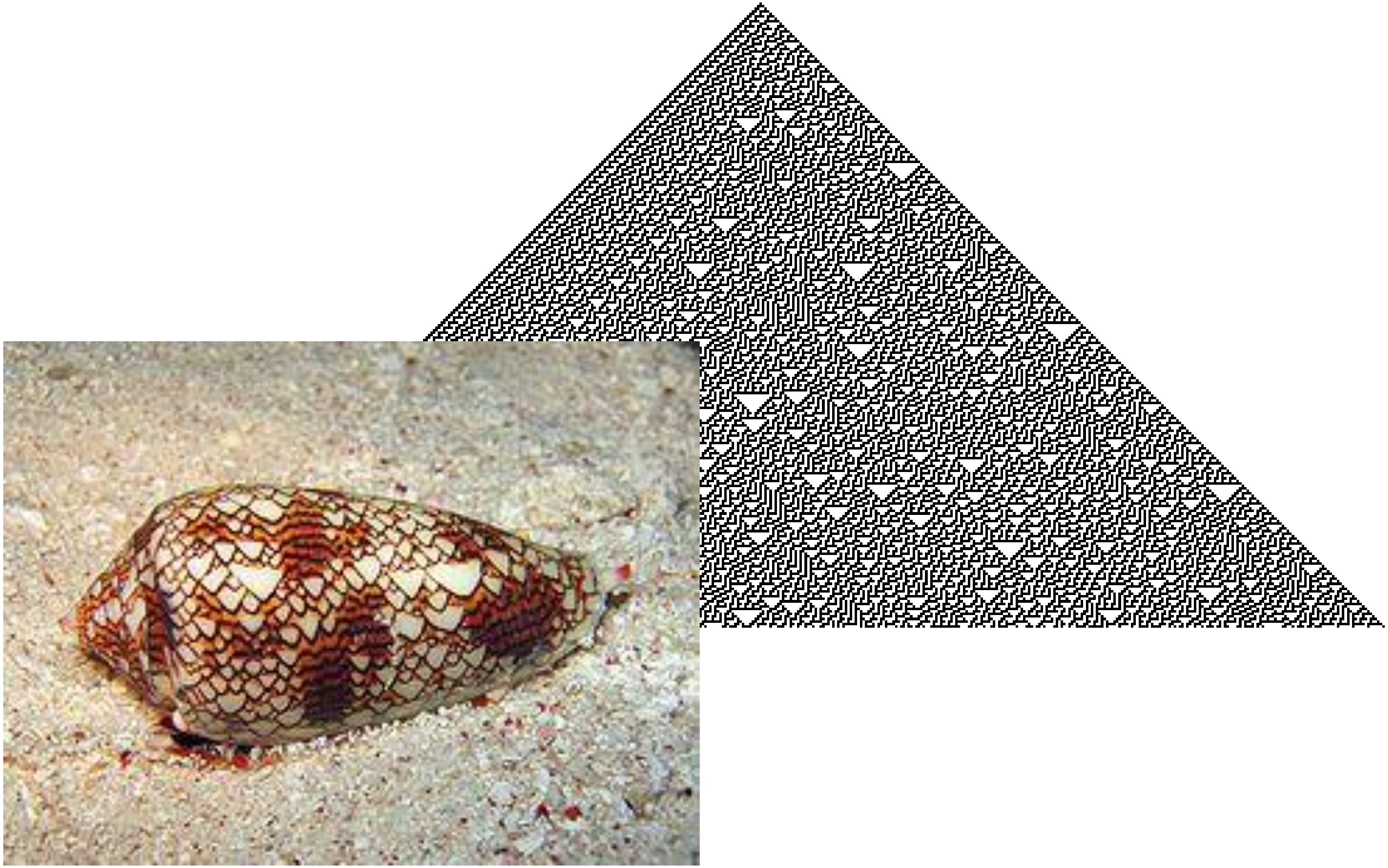
rule 30





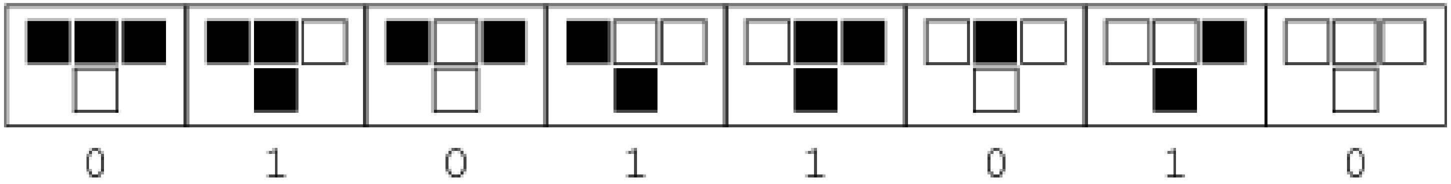
http://mathworld.wolfram.com/Rule30.html

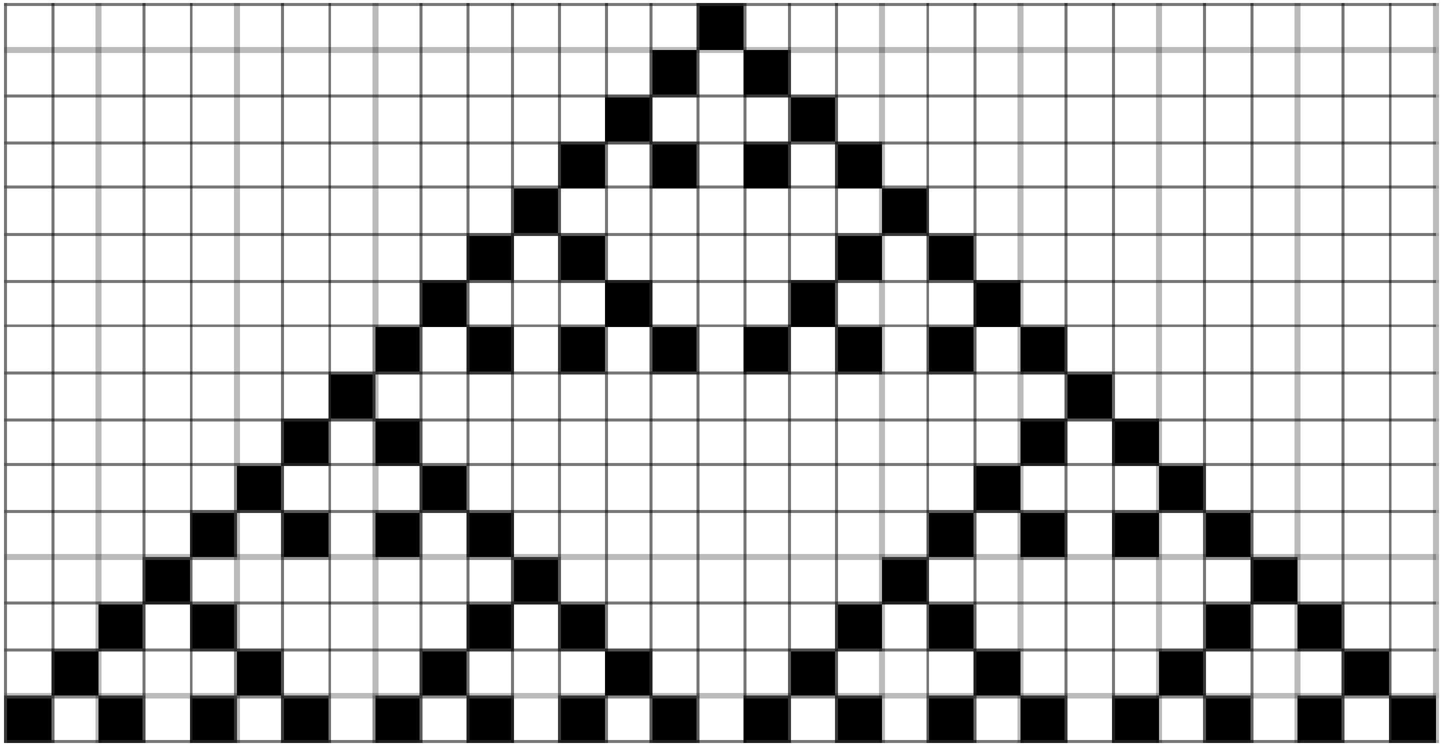
Conus textile



http://en.wikipedia.org/wiki/Conus_textile

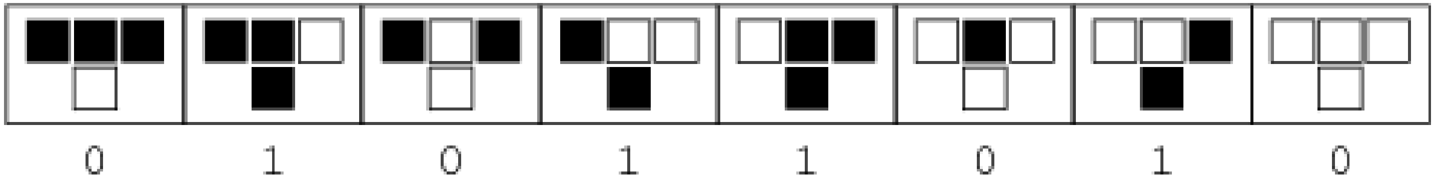
rule 90





http://mathworld.wolfram.com/Rule90.html

rule 90



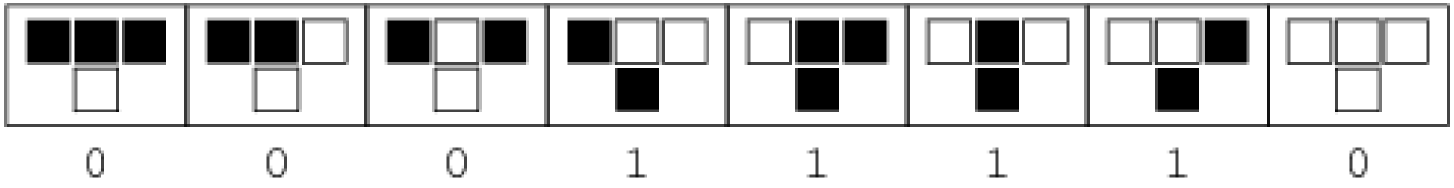
xyz	‘90’
000	0
001	1
010	0
011	1
100	1
101	0
110	1
111	0

A1 x	B1 y	C1 z
	?	

‘90’ als functie van xyz

formule ?

rule 30



xyz	‘30’
000	0
001	1
010	1
011	1
100	1
101	0
110	0
111	0

$$(1-x)\cdot(y+z-y\cdot z)$$

$$+$$

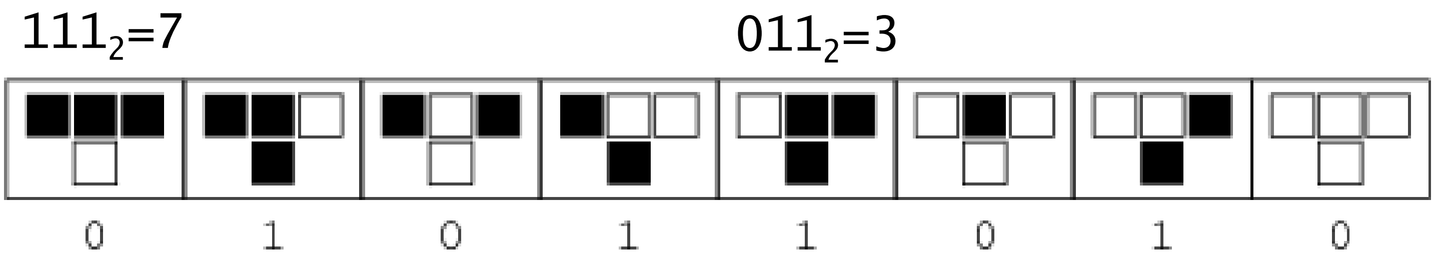
$$x\cdot(1-y)\cdot(1-z)$$

in logische terminologie:

$\neg x\wedge(y\vee z) \;\vee\; x\wedge\neg(y\vee z)$ ‘xor’

$x\oplus(y\vee z)$

lookup ipv. formule



	P	Q	R
	000	0	0
	001	1	1
	010	2	0
	011	3	1
	100	4	1
	101	5	0
	110	6	1
	111	7	0

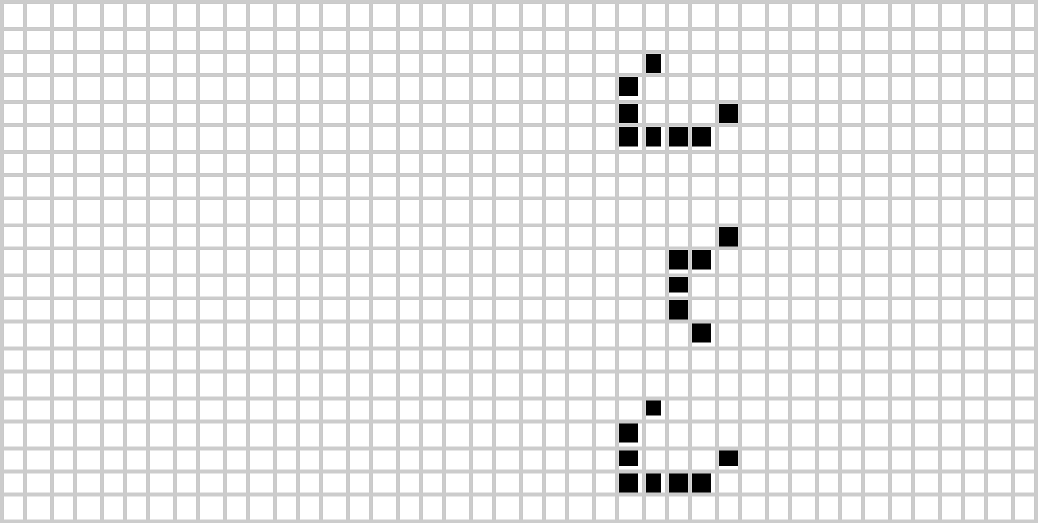
invoer uitvoer

=Lookup(b, Q1:Q8, R1:R8)

b = binaire waarde buren:

A1	B1	C1
= 4*A1+2*B1+1*C1		

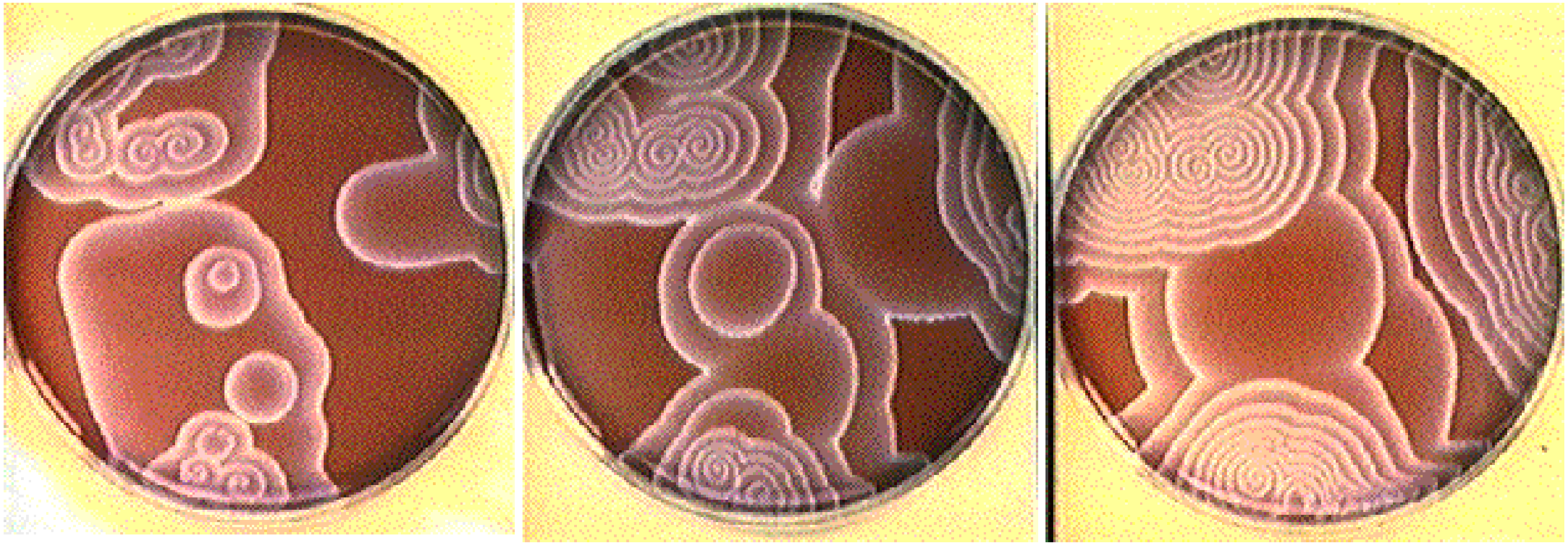
2-dim cellulaire automaat



game of life ‘puffer train’

http://www.collidoscope.com/modernca/

2-dim cellulaire automaat



http://www.ux.uis.no/~ruoff/BZ_Phenomenology.html