# Introduction to Bioinformatics

**JTMS-19**

Marc-Thorsten Hütt mhuett@constructor.university
Felix Jonas fjonas@constructor.university

session Wed, October 9, 2024

# Key idea of probability models (repeated), multiple sequence alignment, final exam: first preview

## What is this session about?

One key idea of probability models is revisited. An algorithm for multiple sequence alignment is introduced. A first preview version of the final exam is discussed.

## How can you revise the material after the session?

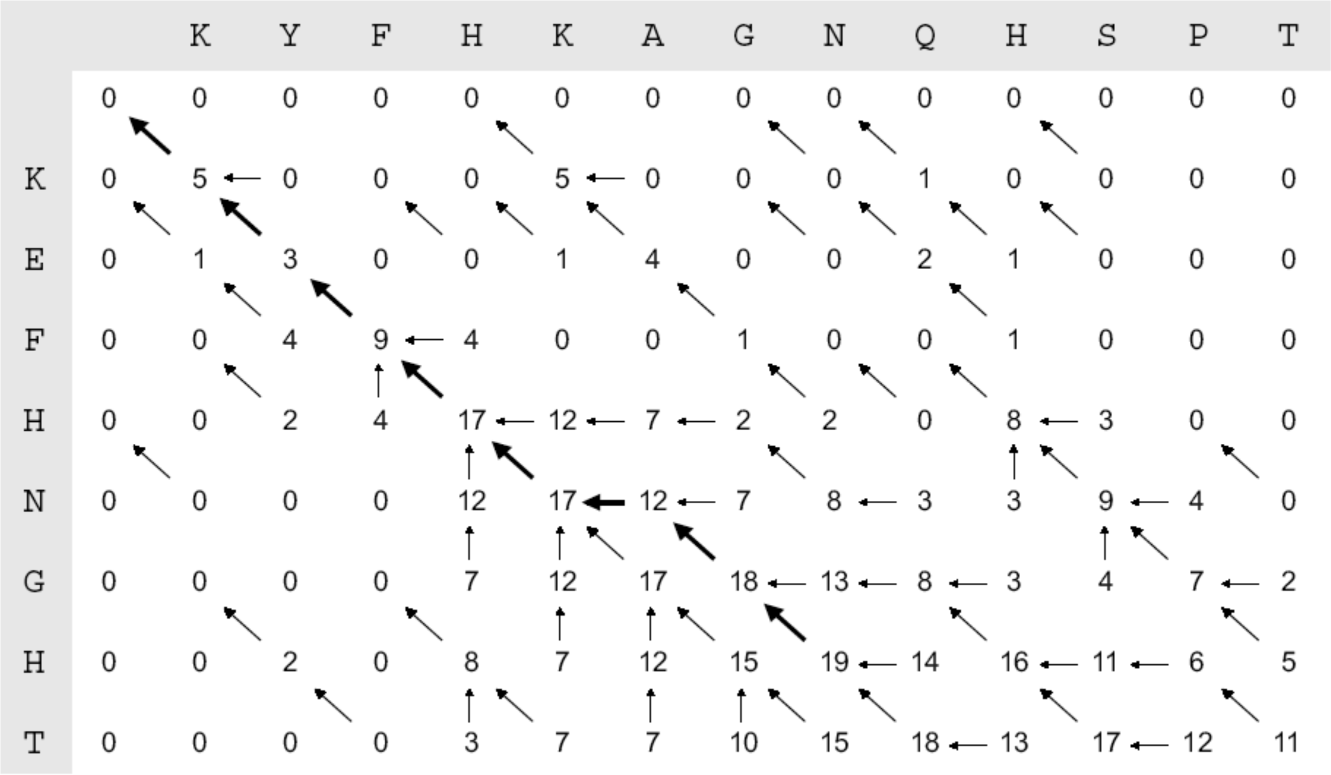
Read Durbin et al. chapters 6.1-6.4

*alternative reading*: Hütt/Dehnert chapter 3.2.2

## ... from the previous lecture

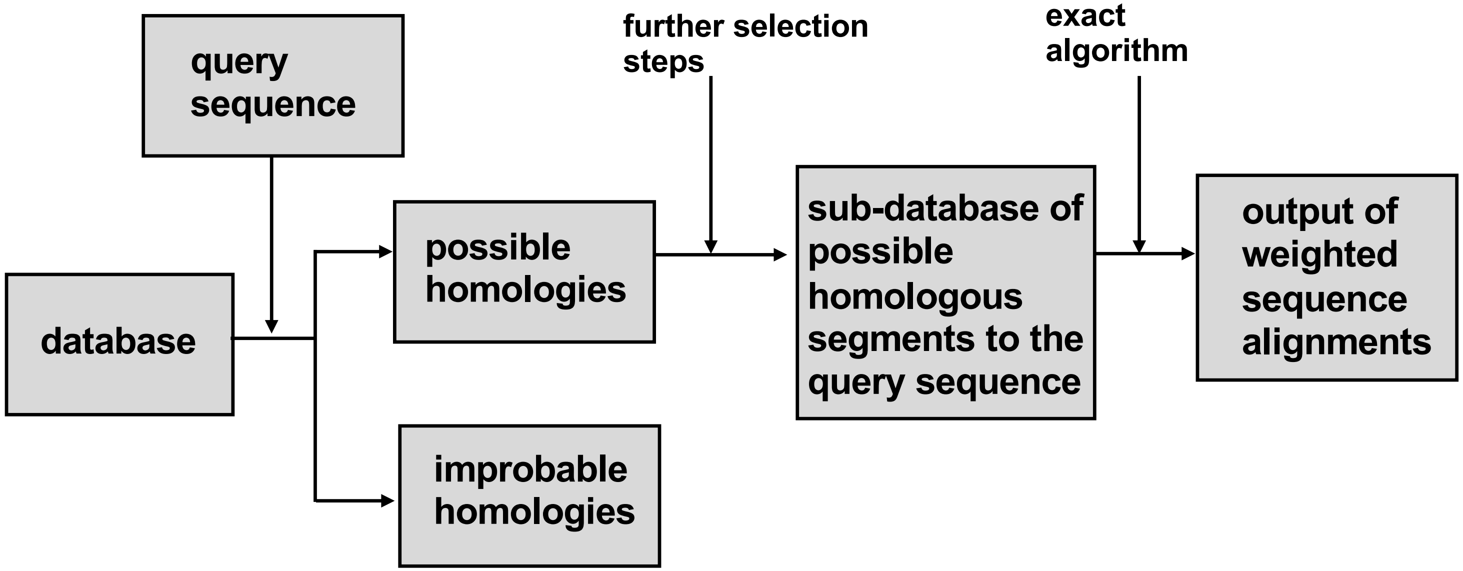
$$F_{i,j} = \max \begin{cases} 0 \\ F_{i-1,j-1} + s(x_i, y_j) \\ F_{i-1,j} - d \\ F_{i,j-1} - d\ . \end{cases}$$

```
KEFHN-GH
|  ||   |
KYFHKAGN
```



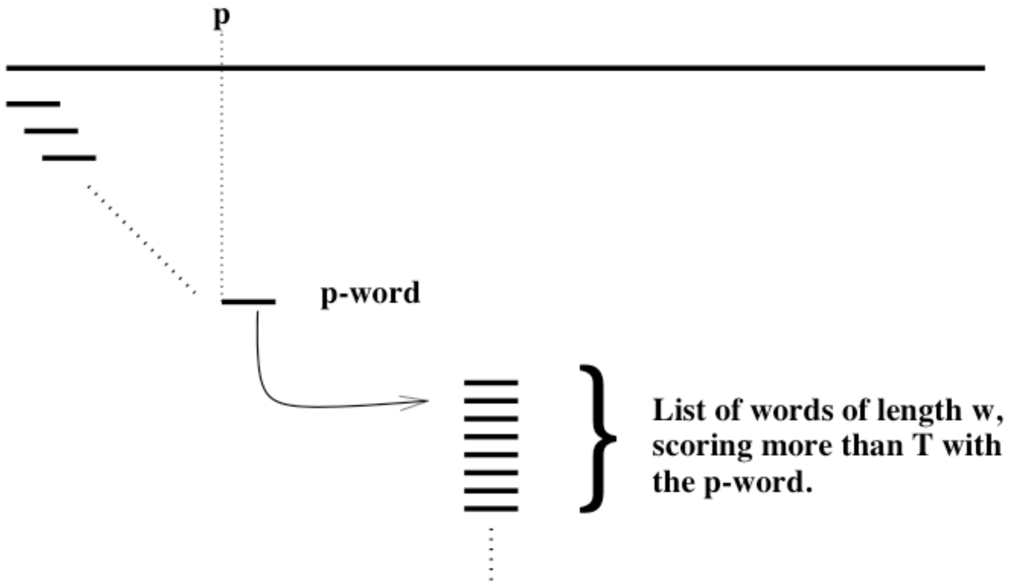
|   |   | K | Y | F | H | K | A | G | N | Q | H | S | P | T |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
|   | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| K | 0 | 5 | 0 | 0 | 0 | 5 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 |
| E | 0 | 1 | 3 | 0 | 0 | 1 | 4 | 0 | 0 | 2 | 1 | 0 | 0 | 0 |
| F | 0 | 0 | 4 | 9 | 4 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 0 |
| H | 0 | 0 | 2 | 4 | 17 | 12 | 7 | 2 | 2 | 0 | 8 | 3 | 0 | 0 |
| N | 0 | 0 | 0 | 0 | 12 | 17 | 12 | 7 | 8 | 3 | 3 | 9 | 4 | 0 |
| G | 0 | 0 | 0 | 0 | 7 | 12 | 17 | 18 | 13 | 8 | 3 | 4 | 7 | 2 |
| H | 0 | 0 | 2 | 0 | 8 | 7 | 12 | 15 | 19 | 14 | 16 | 11 | 6 | 5 |
| T | 0 | 0 | 0 | 0 | 3 | 7 | 7 | 10 | 15 | 18 | 13 | 17 | 12 | 11 |

# Looking inside BLAST


query sequence
database
possible homologies
improbable homologies
further selection steps
sub-database of possible homologous segments to the query sequence
exact algorithm
output of weighted sequence alignments

## Looking inside BLAST

A: For each position p of the query, find the list or words of length w scoring more than T when paired with the word starting at p:



List of words of length w, scoring more than T with the p-word.

## Looking inside BLAST

**query sequence: QLNFSAGW**

**(1) parameters**

**word length w = 2**
**score threshold T = 8**

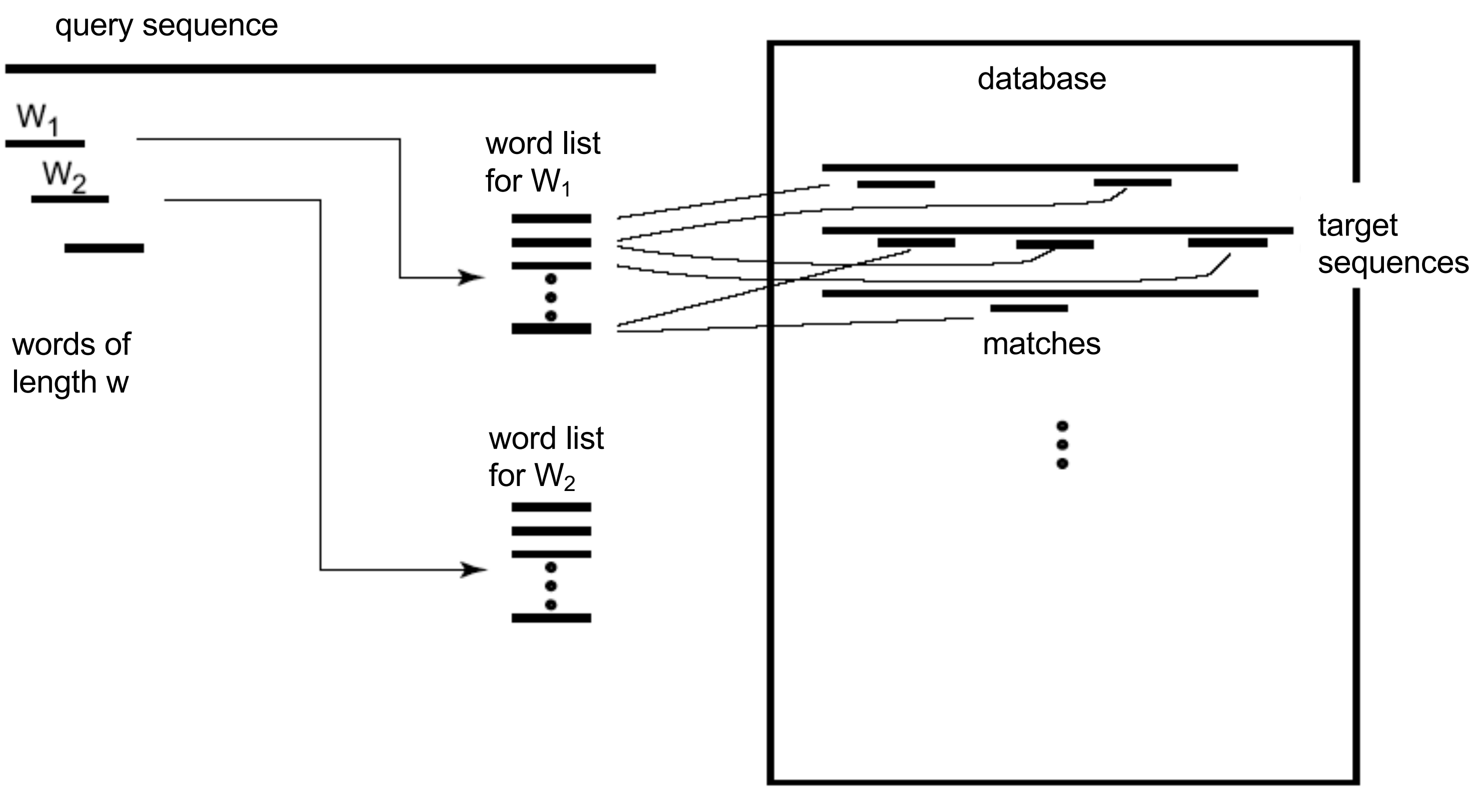
**(2) determine all words of length w in the query sequence:**

**QL LN NF FS SA AG GW**

**(3) for each word determine a word list with an alignment score larger than (or equal to) the threshold T:**

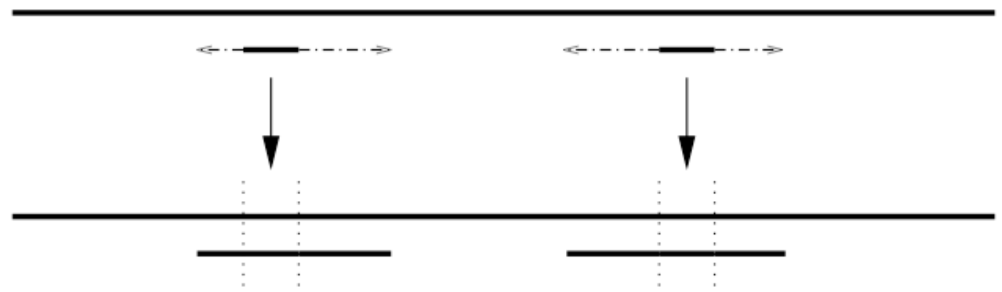
| | |
|---|---|
| **QL:** | **QL=11, QM=9, HL=8, ZL=9** |
| **LN:** | **LN=9, LB=8** |
| **NF:** | **NF=12, AF=8, NY=8, DF=10, ...** |
| **...** | |

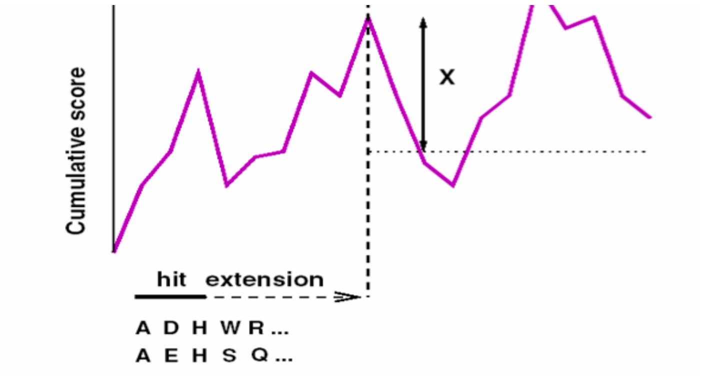
## Looking inside BLAST



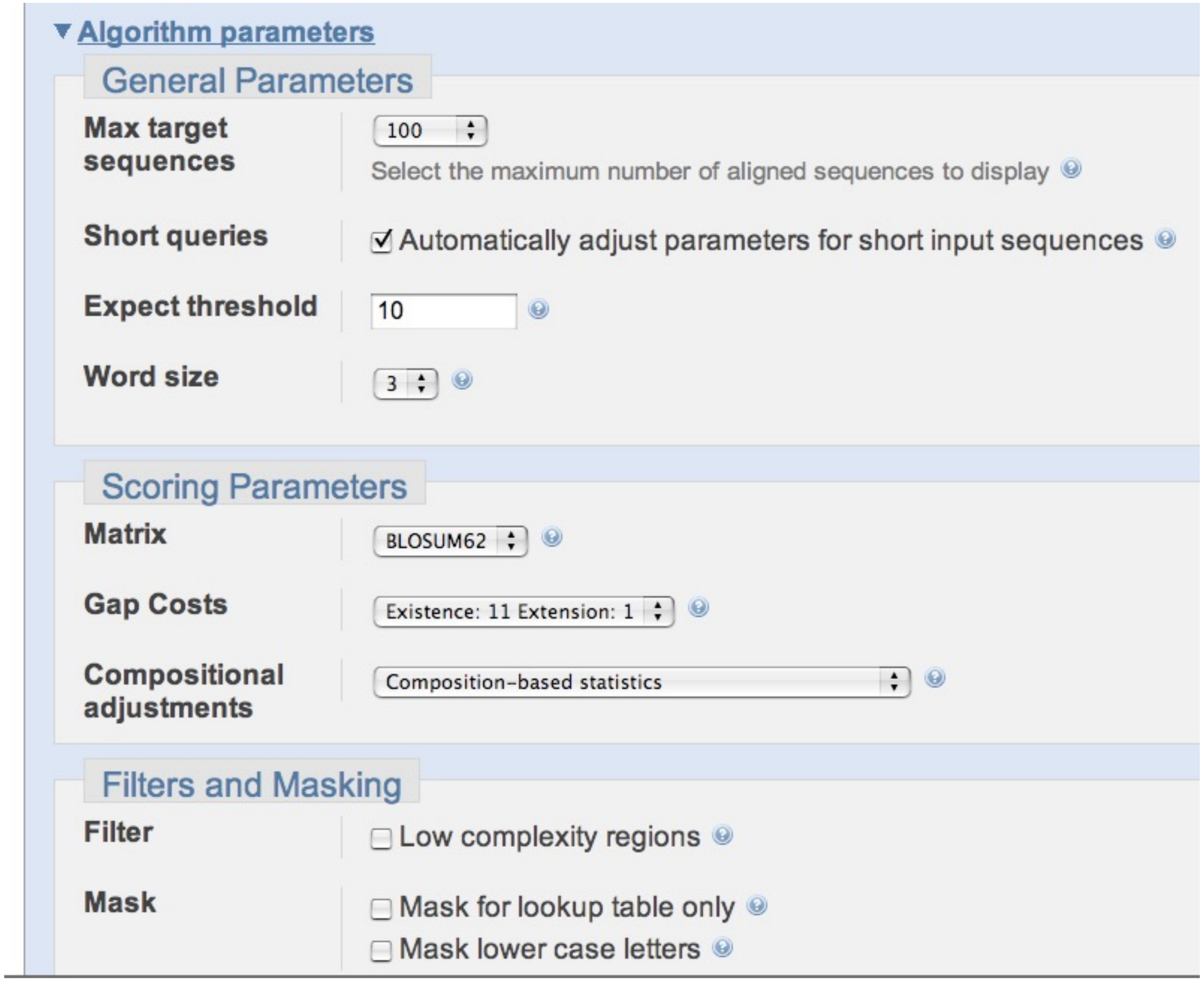
## Looking inside BLAST

C: For each word match («hit»), extend ungapped alignment in both directions. Stop when S decreases by more than X from the highest value reached by S.





## Looking inside BLAST



▼ Algorithm parameters

General Parameters

Max target sequences: 100
Select the maximum number of aligned sequences to display

Short queries: ☑ Automatically adjust parameters for short input sequences

Expect threshold: 10

Word size: 3

Scoring Parameters

Matrix: BLOSUM62

Gap Costs: Existence: 11 Extension: 1

Compositional adjustments: Composition-based statistics

Filters and Masking

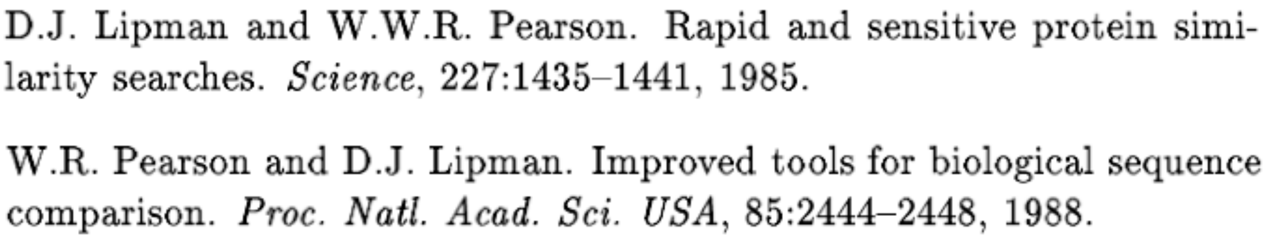
Filter: ☐ Low complexity regions

Mask: ☐ Mask for lookup table only
☐ Mask lower case letters

## Looking inside BLAST

### heuristic methods of sequence alignment

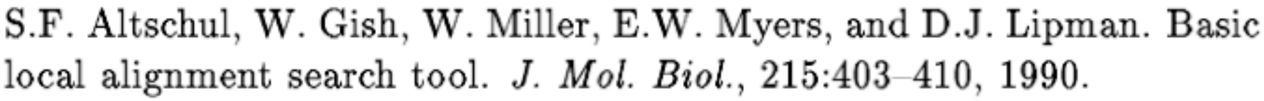
**FastA = fast Alignment**



D.J. Lipman and W.W.R. Pearson. Rapid and sensitive protein similarity searches. *Science*, 227:1435–1441, 1985.

W.R. Pearson and D.J. Lipman. Improved tools for biological sequence comparison. *Proc. Natl. Acad. Sci. USA*, 85:2444–2448, 1988.

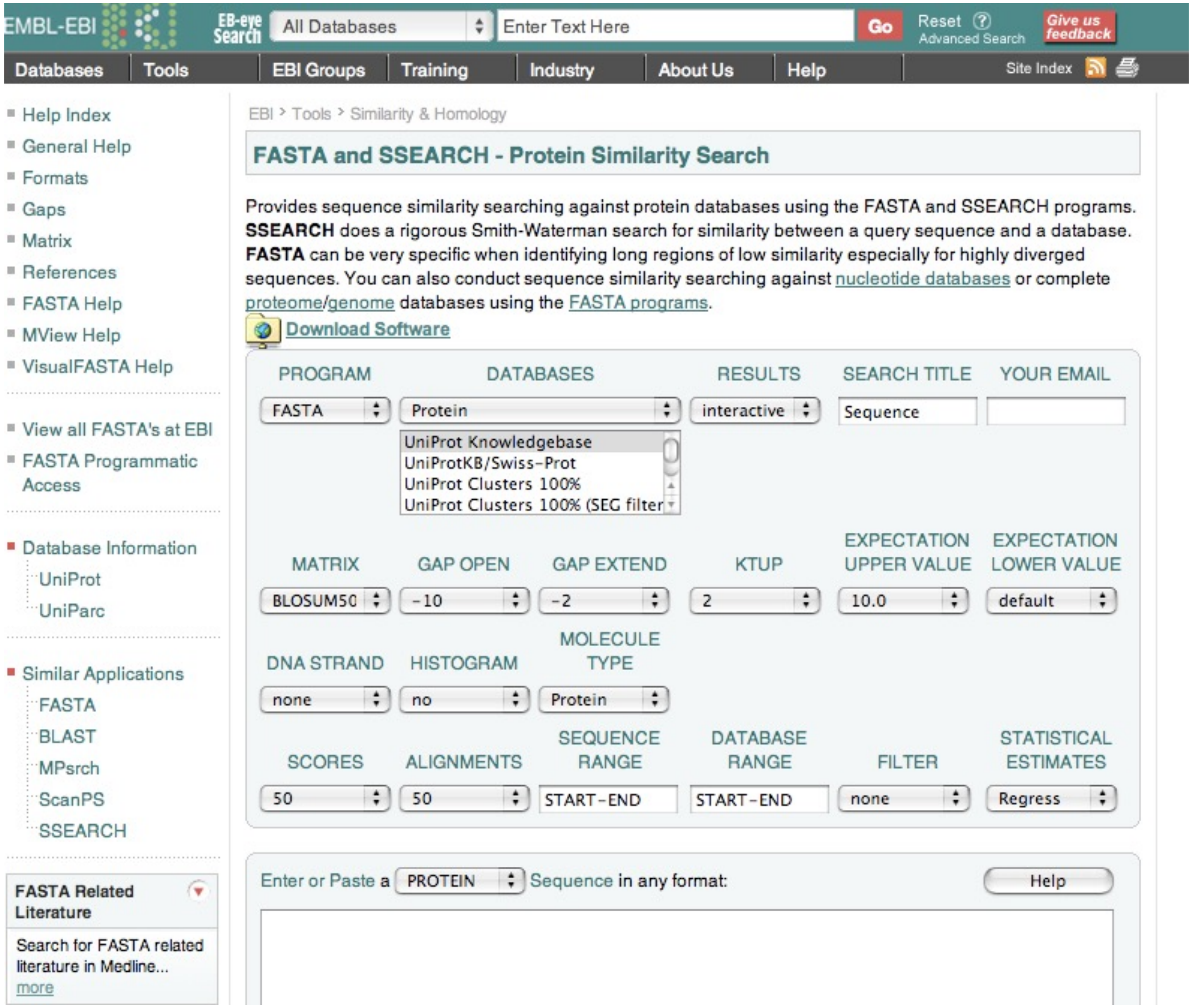
**BLAST = Basic Local Alignment Search Tool**



S.F. Altschul, W. Gish, W. Miller, E.W. Myers, and D.J. Lipman. Basic local alignment search tool. *J. Mol. Biol.*, 215:403–410, 1990.

[a good introduction to these methods:
Frédérique Galisson, The fasta and blast programs, 2002]

## Looking inside FastA



FASTA Related Literature

Search for FASTA related literature in Medline... more

### FASTA and SSEARCH - Protein Similarity Search

Provides sequence similarity searching against protein databases using the FASTA and SSEARCH programs. **SSEARCH** does a rigorous Smith-Waterman search for similarity between a query sequence and a database. **FASTA** can be very specific when identifying long regions of low similarity especially for highly diverged sequences. You can also conduct sequence similarity searching against nucleotide databases or complete proteome/genome databases using the FASTA programs.

| PROGRAM | DATABASES | RESULTS | SEARCH TITLE | YOUR EMAIL |
|---|---|---|---|---|
| FASTA | Protein<br>UniProt Knowledgebase<br>UniProtKB/Swiss-Prot<br>UniProt Clusters 100%<br>UniProt Clusters 100% (SEG filter | interactive | Sequence | |

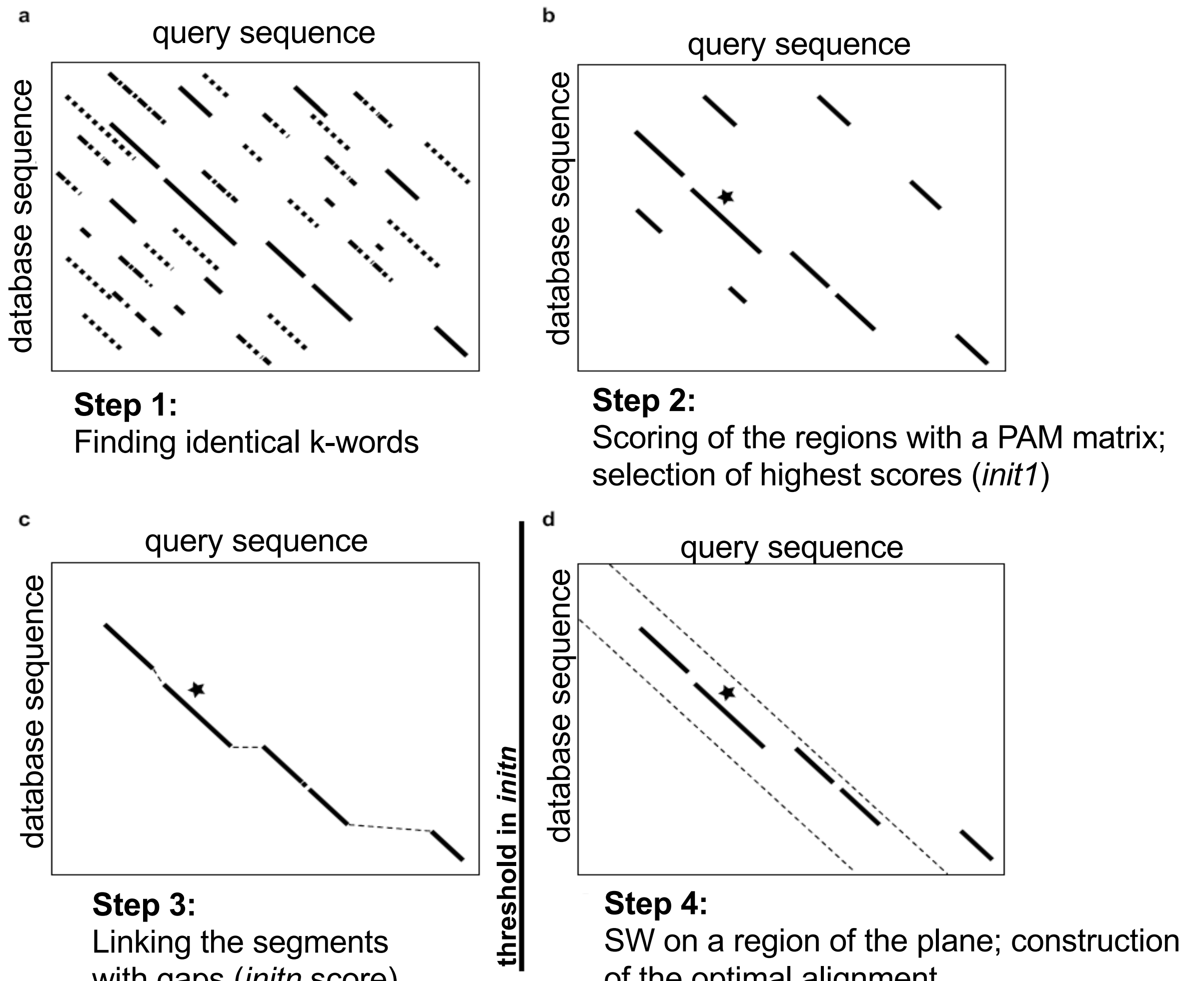
| MATRIX | GAP OPEN | GAP EXTEND | KTUP | EXPECTATION UPPER VALUE | EXPECTATION LOWER VALUE |
|---|---|---|---|---|---|
| BLOSUM50 | -10 | -2 | 2 | 10.0 | default |

| DNA STRAND | HISTOGRAM | MOLECULE TYPE |
|---|---|---|
| none | no | Protein |

| SCORES | ALIGNMENTS | SEQUENCE RANGE | DATABASE RANGE | FILTER | STATISTICAL ESTIMATES |
|---|---|---|---|---|---|
| 50 | 50 | START-END | START-END | none | Regress |

Enter or Paste a PROTEIN Sequence in any format:

## Looking inside FastA



**Step 1:**
Finding identical k-words

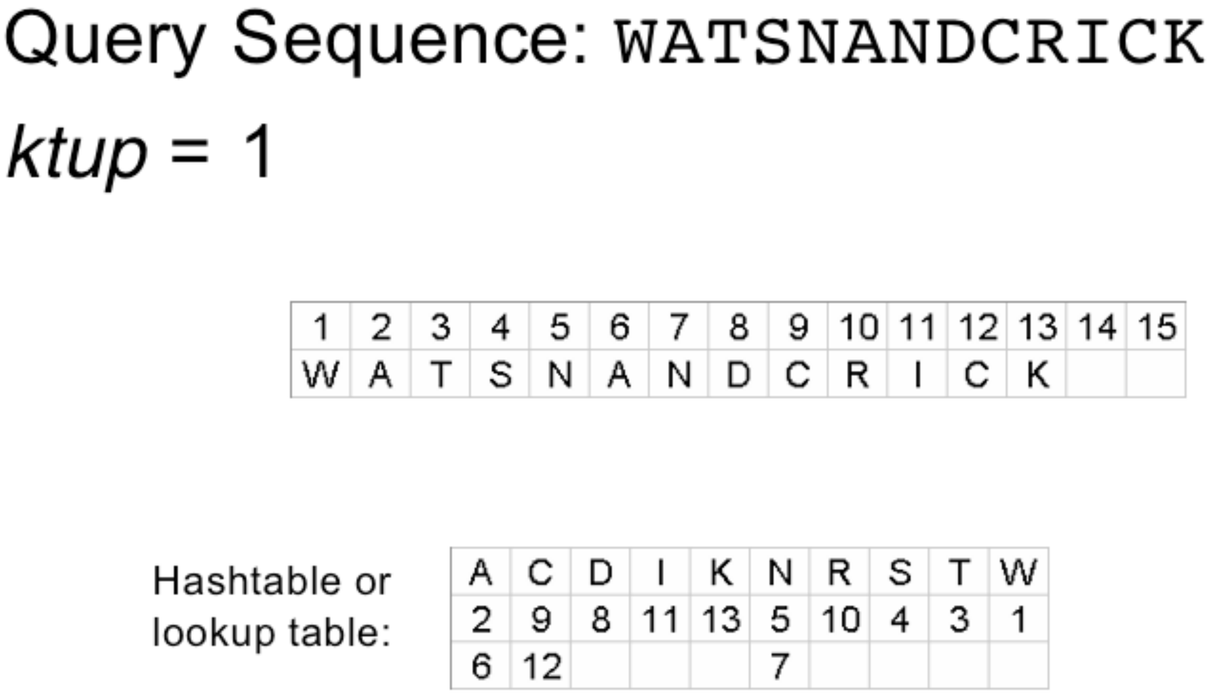
**Step 2:**
Scoring of the regions with a PAM matrix; selection of highest scores (*init1*)

**Step 3:**
Linking the segments with gaps (*initn* score)

**Step 4:**
SW on a region of the plane; construction of the optimal alignment

## Looking inside FastA

remark on step 1



Query Sequence: WATSNANDCRICK

*ktup* = 1

| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| W | A | T | S | N | A | N | D | C | R | I | C | K | | |

Hashtable or lookup table:

| A | C | D | I | K | N | R | S | T | W |
|---|---|---|---|---|---|---|---|---|---|
| 2 | 9 | 8 | 11 | 13 | 5 | 10 | 4 | 3 | 1 |
| 6 | 12 | | | | 7 | | | | |

## Looking inside FastA

remark on step 1

Target Sequence: BASEBALLANDCRICKET

Query Hashtable:

| A | C | D | I | K | N | R | S | T | W |
|---|---|---|---|---|---|---|---|---|---|
| 2 | 9 | 8 | 11 | 13 | 5 | 10 | 4 | 3 | 1 |
| 6 | 12 | | | | 7 | | | | |

Target Hashtable:

| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| B | A | S | E | B | A | L | L | A | N | D | C | R | I | C | K | E | T |
| | 0 | 1 | | | -4 | | | -7 | -5 | -3 | -3 | -3 | -3 | -6 | -3 | | -15 |
| | 4 | | | | 0 | | | -3 | -3 | | 0 | | | -3 | | | |

Offset Table:

| -15 | -14 | -13 | -12 | -11 | -10 | -9 | -8 | -7 | -6 | -5 | -4 | -3 | -2 | -1 | 0 | 1 | 2 | 3 | 4 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | | | | | | | | 1 | 1 | 1 | 1 | 6 | | | 3 | 1 | | | 1 |

| | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 |
| | | | W | A | T | S | N | A | N | D | C | R | I | C | K | | |
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 |
| B | A | S | E | B | A | L | L | A | N | D | C | R | I | C | K | E | T |

**... summary/repetition of probability models**

## ▶ Concepts from statistics
(probabilities, transition probabilities, Markov models)

### General idea:

Probability models can be seen as 'generators' of signal (e.g., DNA sequences) that can then be compared to real-life signals.

Probability models contain parameters.

Each 'run' of a probability model will give you (generally speaking) a different output. Observing a large number of such outputs may allow you to extract the underlying parameters.

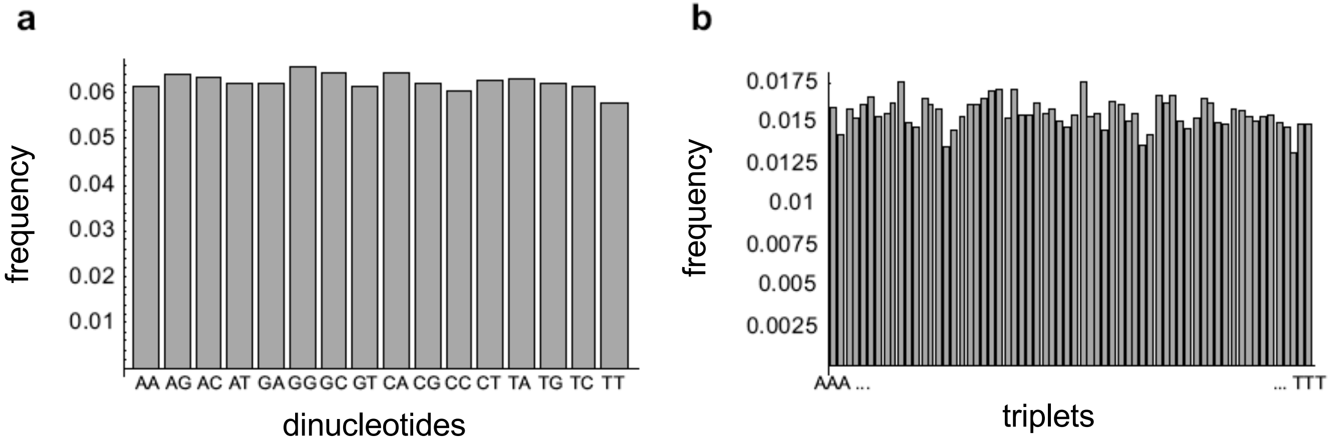
In the case of Markov models, these parameters are transition probabilities from one state (e.g., symbol in a sequence) to the next.

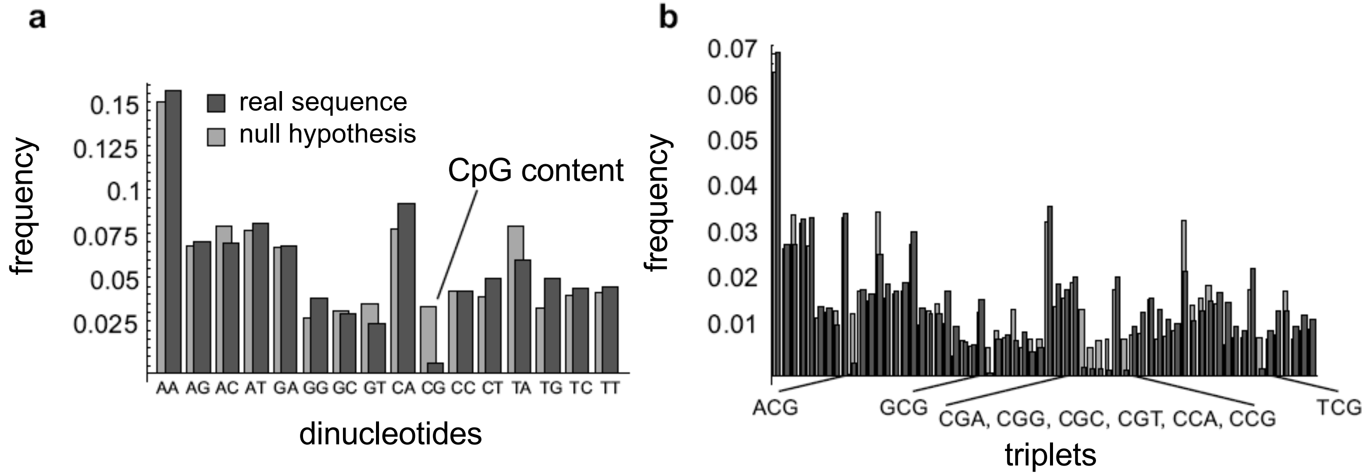
### Training:

Parameter estimation from data.

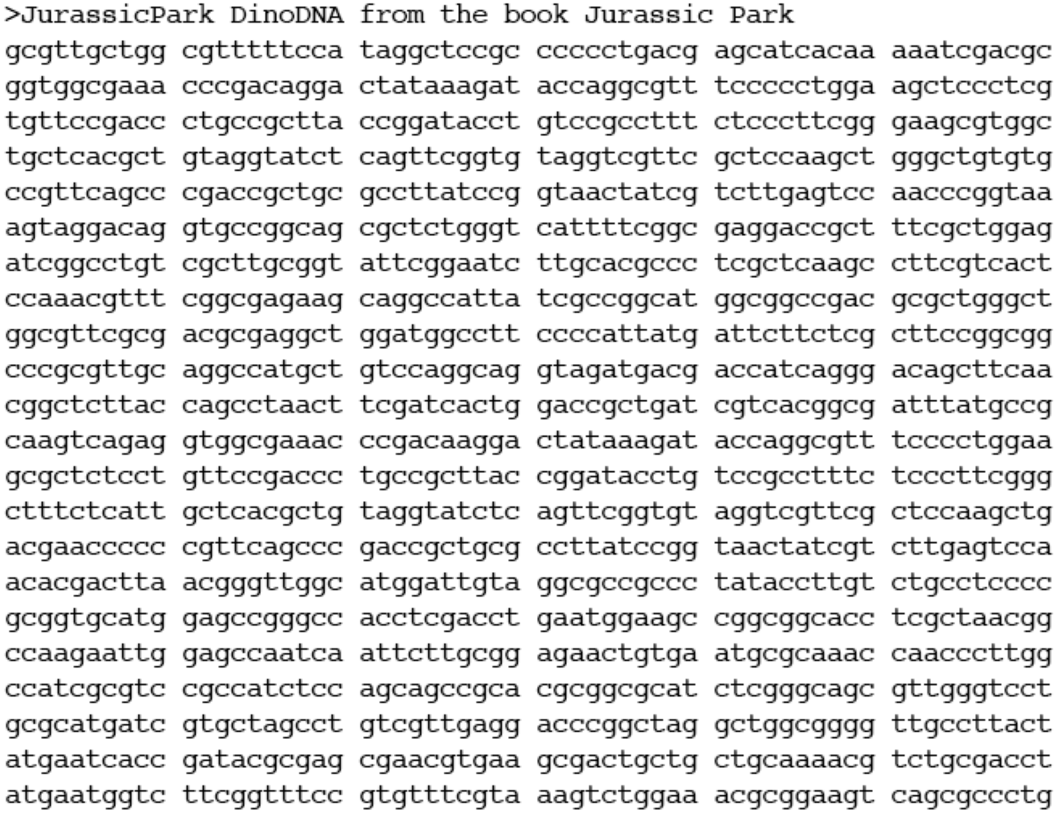
### Scoring:

What is the probability that a given model can produce a given sequence?


a
b
frequency
0.06
0.05
0.04
0.03
0.02
0.01
AA AG AC AT GA GG GC GT CA CG CC CT TA TG TC TT
dinucleotides
frequency
0.0175
0.015
0.0125
0.01
0.0075
0.005
0.0025
AAA ...
... TTT
triplets


a
b
frequency
0.15
0.125
0.1
0.075
0.05
0.025
real sequence
null hypothesis
CpG content
AA AG AC AT GA GG GC GT CA CG CC CT TA TG TC TT
dinucleotides
frequency
0.07
0.06
0.05
0.04
0.03
0.02
0.01
ACG
GCG
CGA, CGG, CGC, CGT, CCA, CCG
TCG
triplets

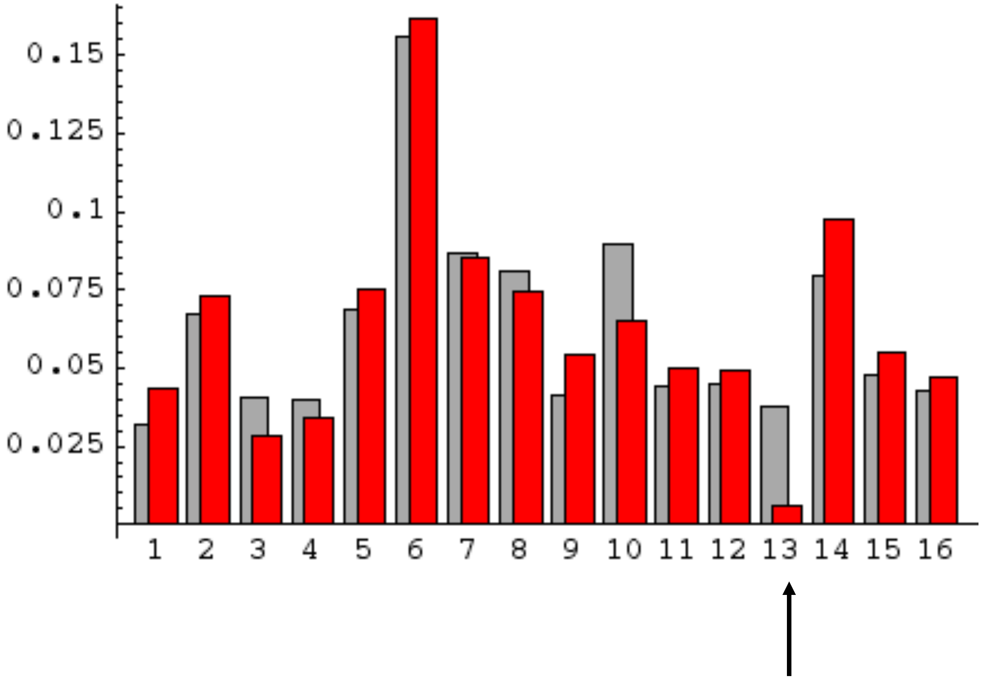
## DNA sequence from *Jurassic Park*



```
>JurassicPark DinoDNA from the book Jurassic Park
gcgttgctgg cgtttttcca taggctccgc ccccctgacg agcatcacaa aaatcgacgc
ggtggcgaaa cccgacagga ctataaagat accaggcgtt tccccctgga agctccctcg
tgttccgacc ctgccgctta ccggatacct gtccgccttt ctcccttcgg gaagcgtggc
tgctcacgct gtaggtatct cagttcggtg taggtcgttc gctccaagct gggctgtgtg
ccgttcagcc cgaccgctgc gccttatccg gtaactatcg tcttgagtcc aacccggtaa
agtaggacag gtgccggcag cgctctgggt cattttcggc gaggaccgct ttcgctggag
atcggcctgt cgcttgcggt attcggaatc ttgcacgccc tcgctcaagc cttcgtcact
ccaaacgttt cggcgagaag caggccatta tcgccggcat ggcggccgac gcgctgggct
ggcgttcgcg acgcgaggct ggatggcctt ccccattatg attcttctcg cttccggcgg
cccgcgttgc aggccatgct gtccaggcag gtagatgacg accatcaggg acagcttcaa
cggctcttac cagcctaact tcgatcactg gaccgctgat cgtcacggcg atttatgccg
caagtcagag gtggcgaaac ccgacaagga ctataaagat accaggcgtt tcccctggaa
gcgctctcct gttccgaccc tgccgcttac cggatacctg tccgcctttc tcccttcggg
ctttctcatt gctcacgctg taggtatctc agttcggtgt aggtcgttcg ctccaagctg
acgaaccccc cgttcagccc gaccgctgcg ccttatccgg taactatcgt cttgagtcca
acacgactta acgggttggc atggattgta ggcgccgccc tataccttgt ctgcctcccc
gcggtgcatg gagccgggcc acctcgacct gaatggaagc cggcggcacc tcgctaacgg
ccaagaattg gagccaatca attcttgcgg agaactgtga atgcgcaaac caacccttgg
ccatcgcgtc cgccatctcc agcagccgca cgcggcgcat ctcgggcagc gttgggtcct
gcgcatgatc gtgctagcct gtcgttgagg acccggctag gctggcgggg ttgccttact
atgaatcacc gatacgcgag cgaacgtgaa gcgactgctg ctgcaaaacg tctgcgacct
atgaatggtc ttcggtttcc gtgtttcgta aagtctggaa acgcggaagt cagcgccctg
```

**Boguski, M.S.**
**A Molecular Biologist Visits Jurassic Park. (1992) BioTechniques 12(5):668-669).**

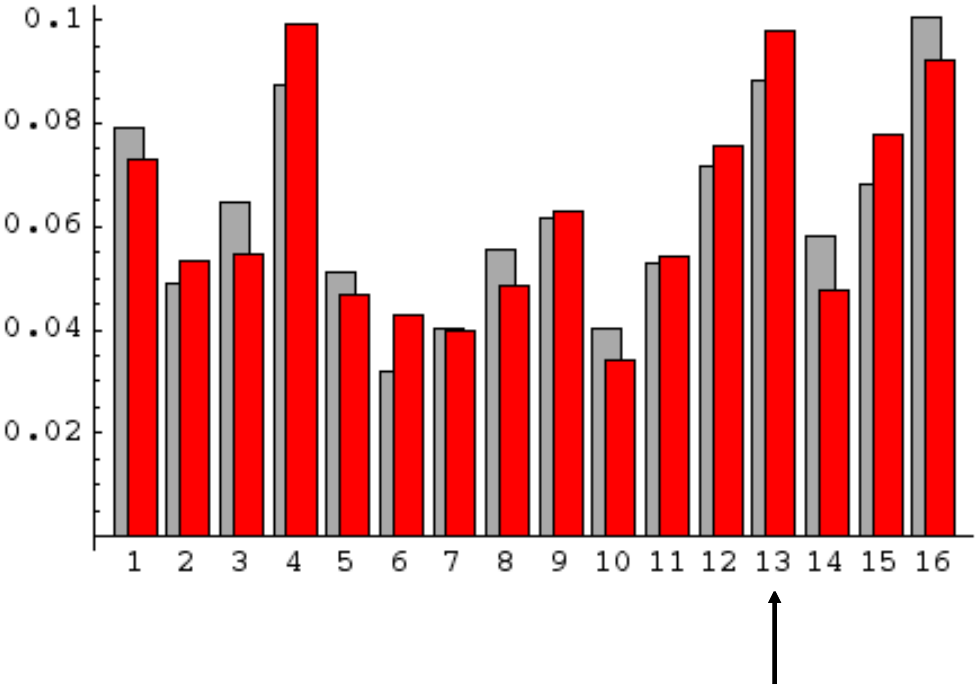
## elementary analysis of pair probabilities



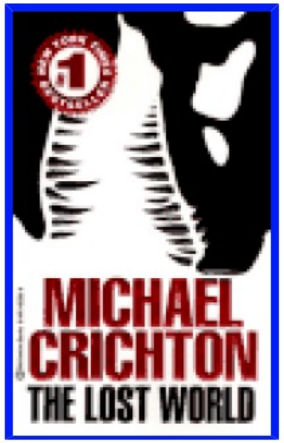
**reduction of the CpG content**
**(lower probability for the dinucleotide "CG")**

## elementary analysis of pair probabilities

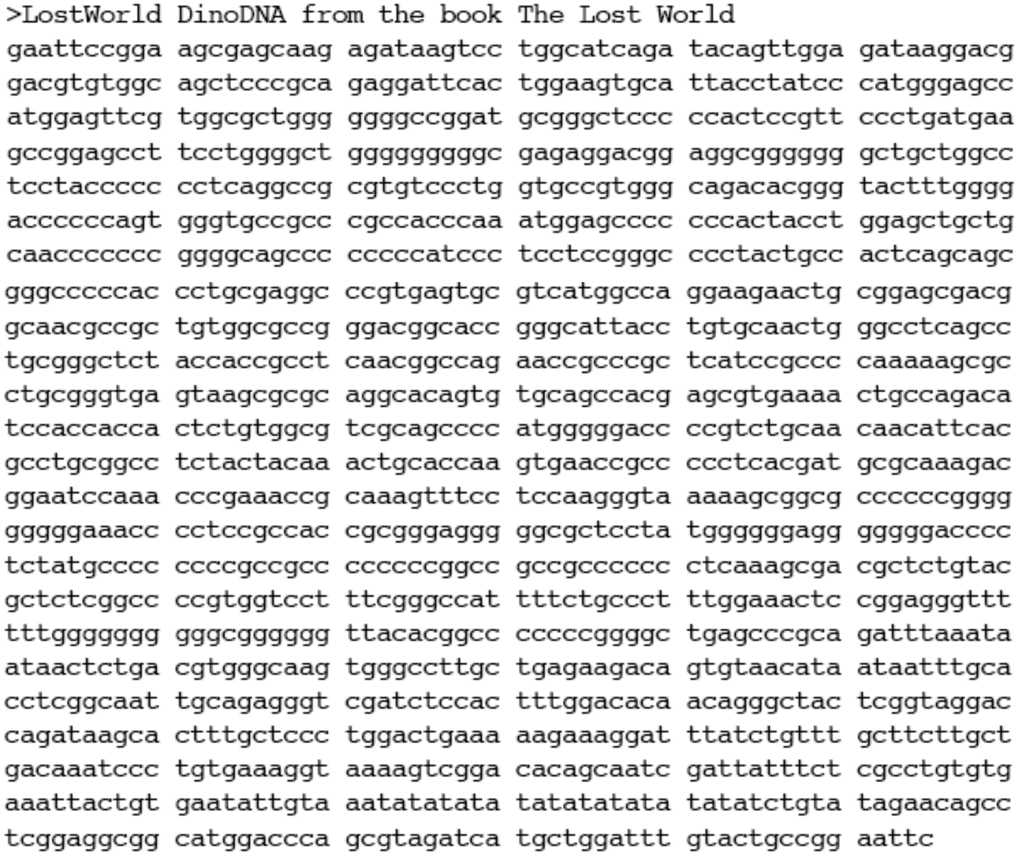
"Jurassic Park"-Sequenz



**no reduction of the CpG content**

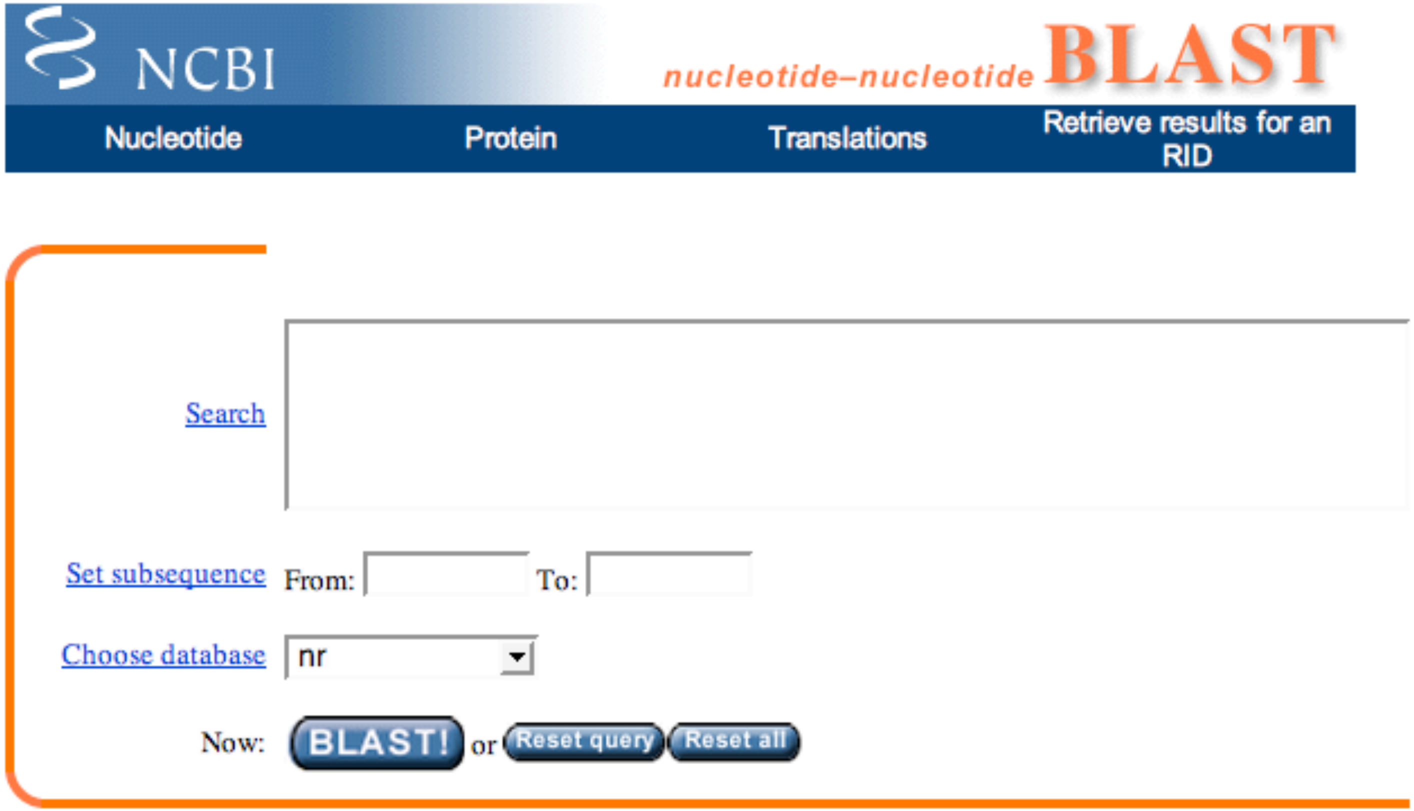


## follow-up: "The Lost World"



```
>LostWorld DinoDNA from the book The Lost World
gaattccgga agcgagcaag agataagtcc tggcatcaga tacagttgga gataaggacg
gacgtgtggc agctcccgca gaggattcac tggaagtgca ttacctatcc catgggagcc
atggagttcg tggcgctggg ggggccggat gcgggctccc ccactccgtt ccctgatgaa
gccggagcct tcctggggct gggggggggc gagaggacgg aggcgggggg gctgctggcc
tcctaccccc cctcaggccg cgtgtccctg gtgccgtggg cagacacggg tactttgggg
acccccccagt gggtgccgcc cgccacccaa atggagcccc cccactacct ggagctgctg
caaccccccc ggggcagccc cccccatccc tcctccgggc ccctactgcc actcagcagc
gggcccccac cctgcgaggc ccgtgagtgc gtcatggcca ggaagaactg cggagcgacg
gcaacgccgc tgtggcgccg ggacggcacc gggcattacc tgtgcaactg ggcctcagcc
tgcgggctct accaccgcct caacggccag aaccgcccgc tcatccgccc caaaaagcgc
ctgcgggtga gtaagcgcgc aggcacagtg tgcagccacg agcgtgaaaa ctgccagaca
tccaccacca ctctgtggcg tcgcagcccc atgggggacc ccgtctgcaa caacattcac
gcctgcggcc tctactacaa actgcaccaa gtgaaccgcc ccctcacgat gcgcaaagac
ggaatccaaa cccgaaaccg caaagtttcc tccaagggta aaaagcggcg cccccgggg
gggggaaacc cctccgccac cgcgggaggg ggcgctccta tgggggagg gggggacccc
tctatgcccc ccccgccgcc ccccccggcc gccgcccccc ctcaaagcga cgctctgtac
gctctcggcc ccgtggtcct ttcgggccat tttctgccct ttggaaactc cggagggttt
tttgggggg gggcgggggg ttacacggcc cccccggggc tgagcccgca gatttaaata
ataactctga cgtgggcaag tgggccttgc tgagaagaca gtgtaacata ataatttgca
cctcggcaat tgcagagggt cgatctccac tttggacaca acaggctac tcggtaggac
cagataagca ctttgctccc tggactgaaa aagaaaggat ttatctgttt gcttcttgct
gacaaatccc tgtgaaaggt aaaagtcgga cacagcaatc gattatttct cgcctgtgtg
aaattactgt gaatattgta aatatatata tatatatata tatatctgta tagaacagcc
tcggaggcgg catggaccca gcgtagatca tgctggattt gtactgccgg aattc
```

# follow-up: "The Lost World"



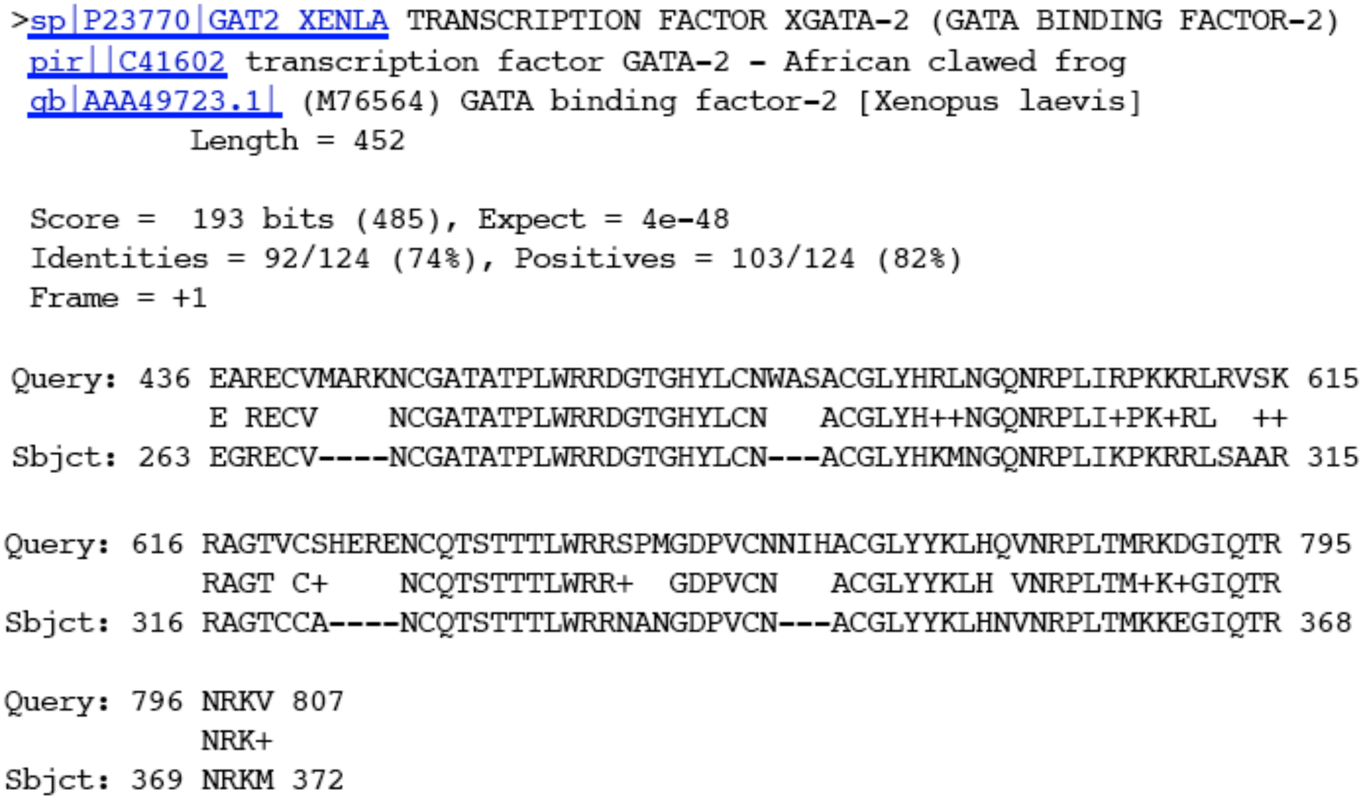
Search

Set subsequence From: To:

Choose database nr

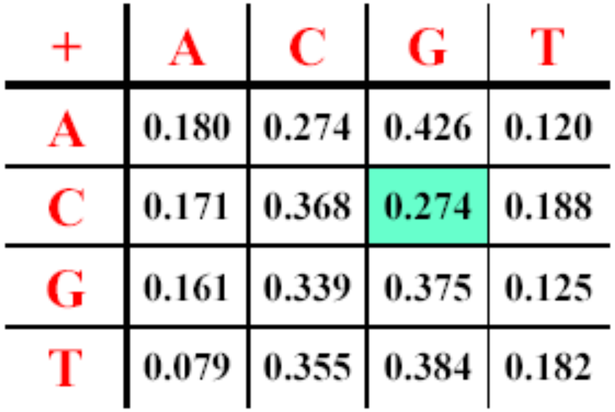
Now: BLAST! or Reset query Reset all

## follow-up: "The Lost World"

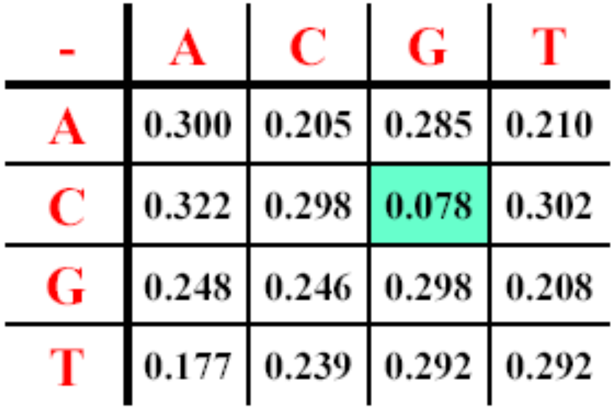


```
>sp|P23770|GAT2_XENLA TRANSCRIPTION FACTOR XGATA-2 (GATA BINDING FACTOR-2)
 pir||C41602 transcription factor GATA-2 - African clawed frog
 gb|AAA49723.1| (M76564) GATA binding factor-2 [Xenopus laevis]
           Length = 452

 Score =  193 bits (485), Expect = 4e-48
 Identities = 92/124 (74%), Positives = 103/124 (82%)
 Frame = +1

Query: 436 EARECVMARKNCGATATPLWRRDGTGHYLCNWASACGLYHRLNGQNRPLIRPKKRLRVSK 615
            E RECV    NCGATATPLWRRDGTGHYLCN   ACGLYH++NGQNRPLI+PK+RL  ++
Sbjct: 263 EGRECV----NCGATATPLWRRDGTGHYLCN---ACGLYHKMNGQNRPLIKPKRRLSAAR 315

Query: 616 RAGTVCSHERENCQTSTTTLWRRSPMGDPVCNNIHACGLYYKLHQVNRPLTMRKDGIQTR 795
            RAGT C+    NCQTSTTTLWRR+  GDPVCN   ACGLYYKLH VNRPLTM+K+GIQTR
Sbjct: 316 RAGTCCA----NCQTSTTTLWRRNANGDPVCN---ACGLYYKLHNVNRPLTMKKEGIQTR 368

Query: 796 NRKV 807
            NRK+
Sbjct: 369 NRKM 372
```

# Markov chains as a tool for studying CpG islands



| + | A | C | G | T |
|---|---|---|---|---|
| A | 0.180 | 0.274 | 0.426 | 0.120 |
| C | 0.171 | 0.368 | 0.274 | 0.188 |
| G | 0.161 | 0.339 | 0.375 | 0.125 |
| T | 0.079 | 0.355 | 0.384 | 0.182 |



| - | A | C | G | T |
|---|---|---|---|---|
| A | 0.300 | 0.205 | 0.285 | 0.210 |
| C | 0.322 | 0.298 | 0.078 | 0.302 |
| G | 0.248 | 0.246 | 0.298 | 0.208 |
| T | 0.177 | 0.239 | 0.292 | 0.292 |

$$S(x) = \log\left(\frac{P(x\,|\,\text{model }+)}{P(x\,|\,\text{model }-)}\right) = \log\left(\frac{P(B)\prod_{i=1}^{L} a^{+}_{x_{i-1}x_i}}{P(B)\prod_{i=1}^{L} a^{-}_{x_{i-1}x_i}}\right) = \sum_{i=1}^{L}\log\left(\frac{a^{+}_{x_{i-1}x_i}}{a^{-}_{x_{i-1}x_i}}\right) = \sum_{i=1}^{L}\beta_{x_{i-1}x_i}$$

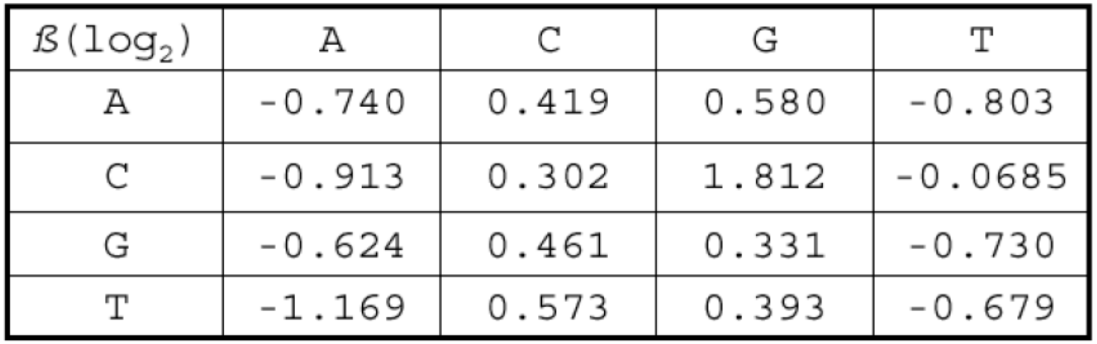
a number for each sequence x

→ histogram of score values S(x) for many sequences x

a number for each dinucleotide

→ table of "log-likelihoods"

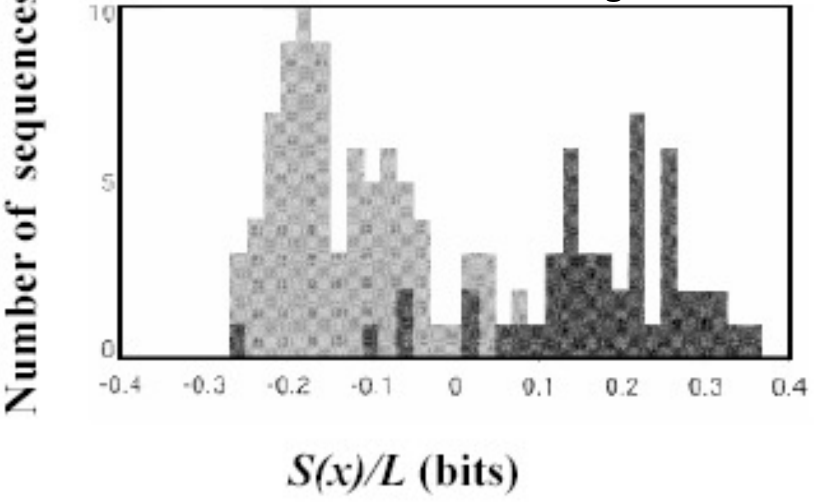
## Markov chains as a tool for studying CpG islands

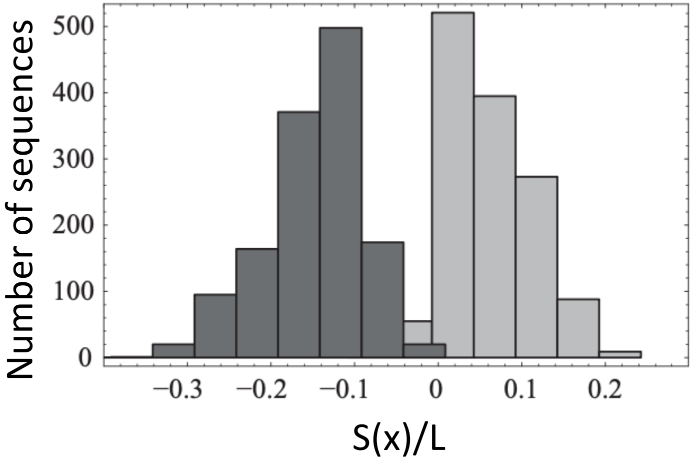


| ß(log$_2$) | A | C | G | T |
|---|---|---|---|---|
| A | -0.740 | 0.419 | 0.580 | -0.803 |
| C | -0.913 | 0.302 | 1.812 | -0.0685 |
| G | -0.624 | 0.461 | 0.331 | -0.730 |
| T | -1.169 | 0.573 | 0.393 | -0.679 |

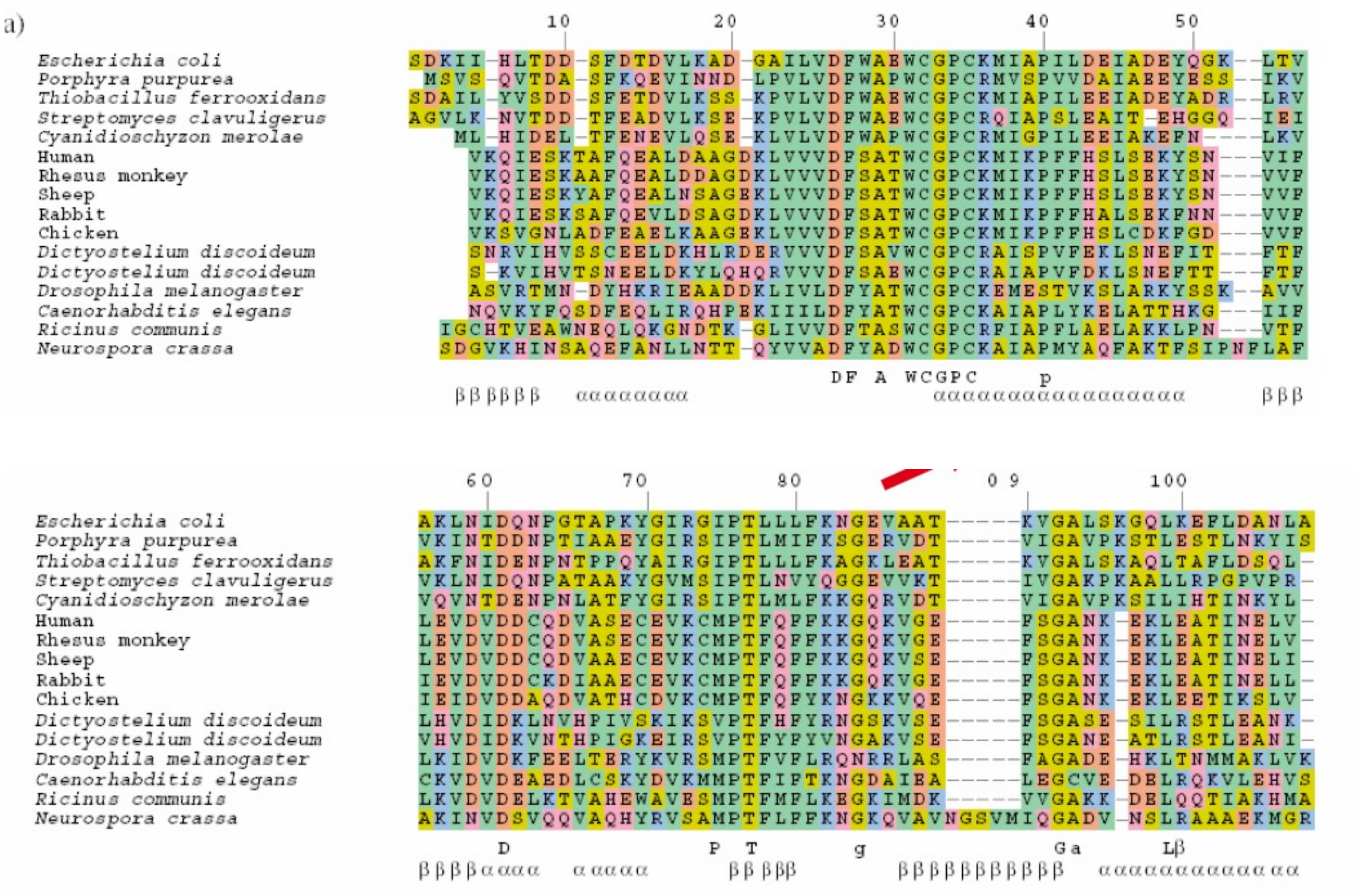
**table of "log-likelihoods"**

**histogram of scores**



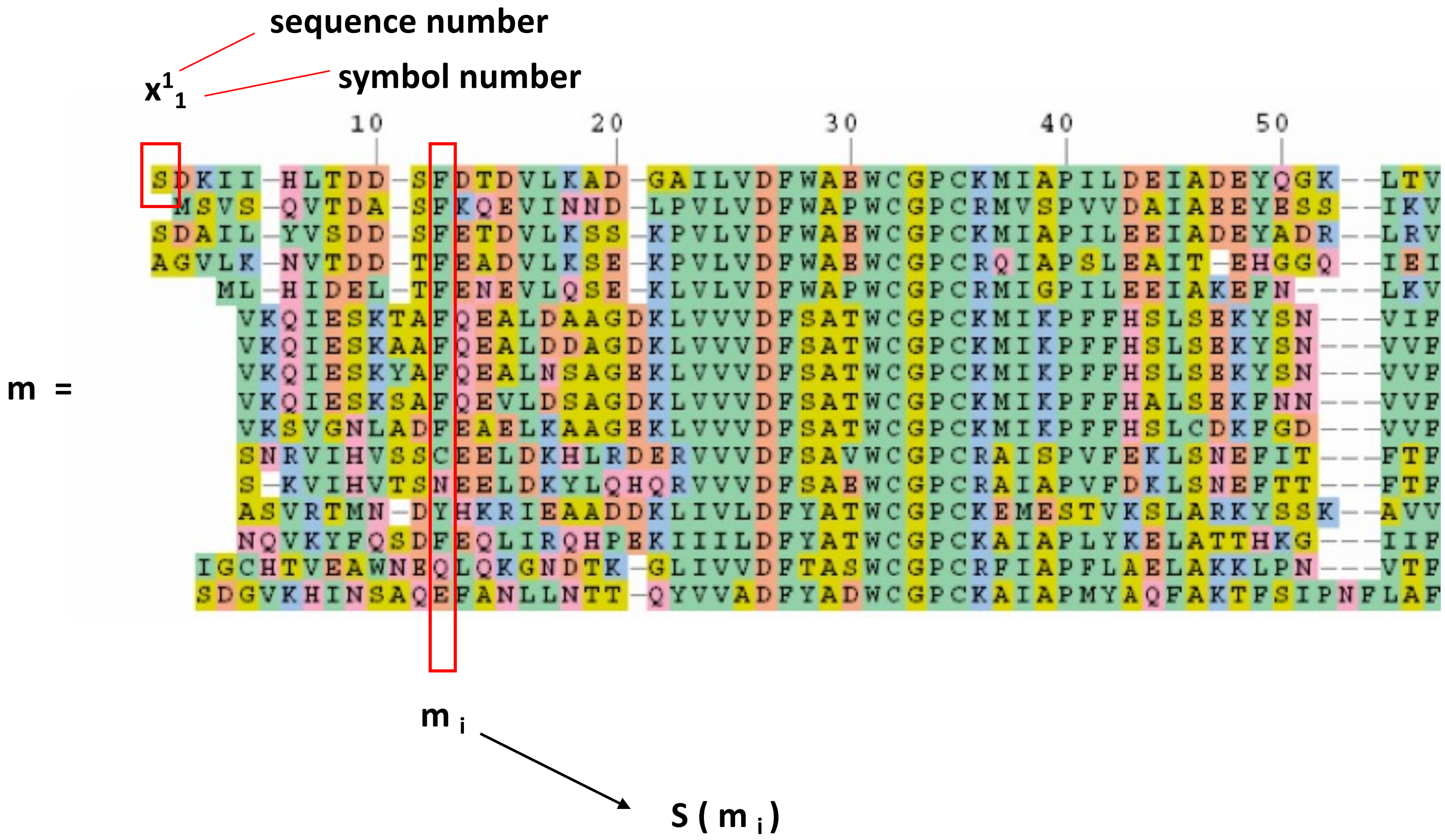


## Multiple sequence alignment



a)

```
                                    10         20          30        40        50
Escherichia coli           SDKII-HLTDD-SFDTDVLKAD-GAILVDFWAEWCGPCKMIAPILDEIADEYQGK--LTV
Porphyra purpurea           MSVS-QVTDA-SFKQEVINND-LPVLVDFWAPWCGPCRMVSPVVDAIAEEYESS--IKV
Thiobacillus ferrooxidans  SDAIL-YVSDD-SFETDVLKSS-KPVLVDFWAEWCGPCKMIAPILEEIADEYADR--LRV
Streptomyces clavuligerus  AGVLK-NVTDD-TFEADVLKSE-KPVLVDFWAEWCGPCRQIAPSLEAIT-EHGGQ--IEI
Cyanidioschyzon merolae       ML-HIDEL-TFENEVLQSE-KLVLVDFWAPWCGPCRMIGPILEEIAKEFN----LKV
Human                          VKQIESKTAFQEALDAAGDKLVVVDFSATWCGPCKMIKPFFHSLSEKYSN---VIF
Rhesus monkey                  VKQIESKAAFQEALDDAGDKLVVVDFSATWCGPCKMIKPFFHSLSEKYSN---VVF
Sheep                          VKQIESKYAFQEALNSAGEKLVVVDFSATWCGPCKMIKPFFHSLSEKYSN---VVF
Rabbit                         VKQIESKSAFQEVLDSAGDKLVVVDFSATWCGPCKMIKPFFHALSEKFNN---VVF
Chicken                        VKSVGNLADFEAELKAAGEKLVVVDFSATWCGPCKMIKPFFHSLCDKFGD---VVF
Dictyostelium discoideum       SNRVIHVSSCEELDKHLRDERVVVDFSAVWCGPCRAISPVFEKLSNEFIT---FTF
Dictyostelium discoideum       S-KVIHVTSNEELDKYLQHQRVVVDFSAEWCGPCRAIAPVFDKLSNEFTT---FTF
Drosophila melanogaster        ASVRTMN-DYHKRIEAADDKLIVLDFYATWCGPCKEMESTVKSLARKYSSK--AVV
Caenorhabditis elegans         NQVKYFQSDFEQLIRQHPEKIIILDFYATWCGPCKAIAPLYKELATTHKG---IIF
Ricinus communis             IGCHTVEAWNEQLQKGNDTK-GLIVVDFTASWCGPCRFIAPFLAELAKKLPN---VTF
Neurospora crassa            SDGVKHINSAQEFANLLNTT-QYVVADFYADWCGPCKAIAPMYAQFAKTFSIPNFLAF

                                                       DF A WCGPC   p
                              ββββββ  αααααααα                αααααααααααααααααα     βββ

                              60         70        80            0 9       100
Escherichia coli           AKLNIDQNPGTAPKYGIRGIPTLLLFKNGEVAAT-----KVGALSKGQLKEFLDANLA
Porphyra purpurea          VKINTDDNPTIAAEYGIRSIPTLMIFKSGERVDT-----VIGAVPKSTLESTLNKYIS
Thiobacillus ferrooxidans  AKFNIDENPNTPPQYAIRGIPTLLLFKAGKLEAT-----KVGALSKAQLTAFLDSQL-
Streptomyces clavuligerus  VKLNIDQNPATAAKYGVMSIPTLNVYQGGEVVKT-----IVGAKPKAALLRPGPVPR-
Cyanidioschyzon merolae    VQVNTDENPNLATFYGIRSIPTLMLFKKGQRVDT-----VIGAVPKSILIHTINKYL-
Human                      LEVDVDDCQDVASECEVKCMPTFQFFKKGQKVGE-----FSGANK-EKLEATINELV-
Rhesus monkey              LEVDVDDCQDVASECEVKCMPTFQFFKKGQKVGE-----FSGANK-EKLEATINELV-
Sheep                      LEVDVDDCQDVAAECEVKCMPTFQFFKKGQKVSE-----FSGANK-EKLEATINELI-
Rabbit                     IEVDVDDCKDIAAECEVKCMPTFQFFKKGQKVGE-----FSGANK-EKLEATINELL-
Chicken                    IEIDVDDAQDVATHCDVKCMPTFQFYKNGKKVQE-----FSGANK-EKLEETIKSLV-
Dictyostelium discoideum   LHVDIDKLNVHPIVSKIKSVPTFHFYRNGSKVSE-----FSGASE-SILRSTLEANK-
Dictyostelium discoideum   VHVDIDKVNTHPIGKEIRSVPTFYFYVNGAKVSE-----FSGANE-ATLRSTLEANI-
Drosophila melanogaster    LKIDVDKFEELTERYKVRSMPTFVFLRQNRRLAS-----FAGADE-HKLTNMMAKLVK
Caenorhabditis elegans     CKVDVDEAEDLCSKYDVKMMPTFIFTKNGDAIEA-----LEGCVE-DELRQKVLEHVS
Ricinus communis           LKVDVDELKTVAHEWAVESMPTFMFLKEGKIMDK-----VVGAKK-DELQQTIAKHMA
Neurospora crassa          AKINVDSVQQVAQHYRVSAMPTFLPFKNGKQVAVNGSVMIQGADV-NSLRAAAEKMGR

                                D              P T      g            Ga     Lβ
                           ββββαααα  ααααα     ββ βββ        βββββββββββ  αααααααααααααα
```



sequence number

symbol number

$x^1_1$

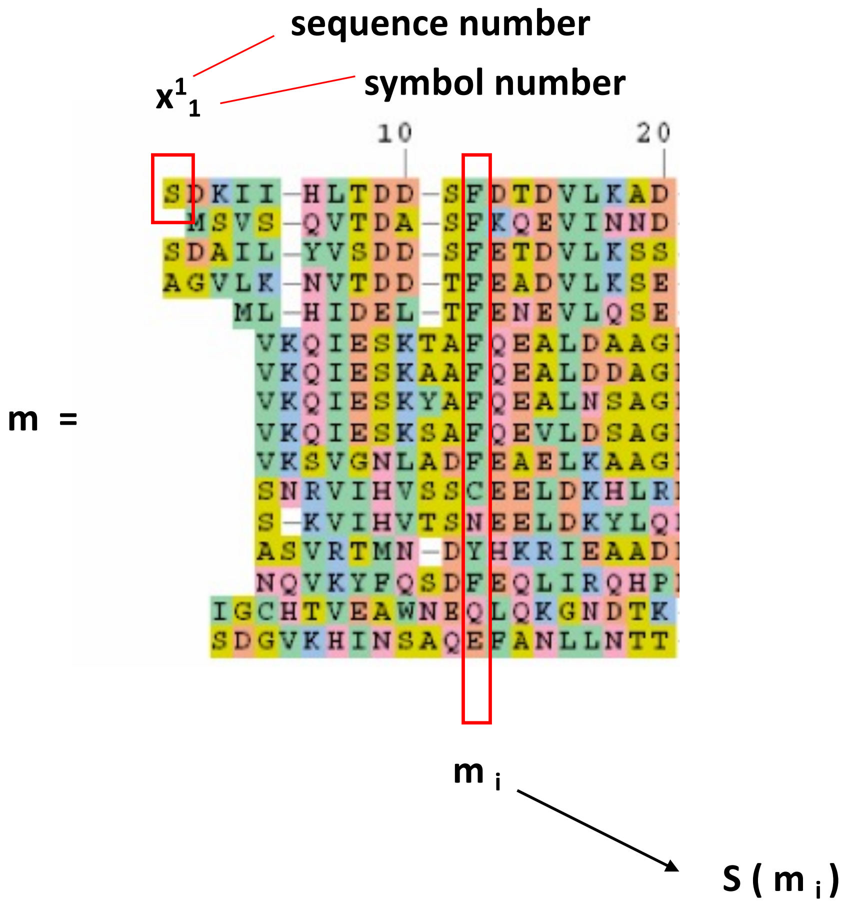
m =

```
         10        20        30        40        50
SDKII-HLTDD-SFDTDVLKAD-GAILVDFWAEWCGPCKMIAPILDEIADEYQGK--LTV
 MSVS-QVTDA-SFKQEVINND-LPVLVDFWAPWCGPCRMVSPVVDAIAEEYESS--IKV
SDAIL-YVSDD-SFETDVLKSS-KPVLVDFWAEWCGPCKMIAPILEEIADEYADR--LRV
AGVLK-NVTDD-TFEADVLKSE-KPVLVDFWAEWCGPCRQIAPSLEAIT-EHGGQ--IEI
   ML-HIDEL-TFENEVLQSE-KLVLVDFWAPWCGPCRMIGPILEEIAKEFN----LKV
    VKQIESKTAFQEALDAAGDKLVVVDFSATWCGPCKMIKPFFHSLSEKYSN---VIF
    VKQIESKAAFQEALDDAGDKLVVVDFSATWCGPCKMIKPFFHSLSEKYSN---VVF
    VKQIESKYAFQEALNSAGEKLVVVDFSATWCGPCKMIKPFFHSLSEKYSN---VVF
    VKQIESKSAFQEVLDSAGDKLVVVDFSATWCGPCKMIKPFFHALSEKFNN---VVF
    VKSVGNLADFEAELKAAGEKLVVVDFSATWCGPCKMIKPFFHSLCDKFGD---VVF
    SNRVIHVSSCEELDKHLRDERVVVDFSAVWCGPCRAISPVFEKLSNEFIT---FTF
    S-KVIHVTSNEELDKYLQHQRVVVDFSAEWCGPCRAIAPVFDKLSNEFTT---FTF
    ASVRTMN-DYHKRIEAADDKLIVLDFYATWCGPCKEMESTVKSLARKYSSK--AVV
    NQVKYFQSDFEQLIRQHPEKIIILDFYATWCGPCKAIAPLYKELATTHKG---IIF
  IGCHTVEAWNEQLQKGNDTK-GLIVVDFTASWCGPCRFIAPFLAELAKKLPN---VTF
  SDGVKHINSAQEFANLLNTT-QYVVADFYADWCGPCKAIAPMYAQFAKTFSIPNFLAF
```

$m_i$

$S(m_i)$

Sum of pairs (SP) score



```
SDKII-HLTDD-SFDTDVLKAD
 MSVS-QVTDA-SFKQEVINND
SDAIL-YVSDD-SFETDVLKSS
AGVLK-NVTDD-TFEADVLKSE
   ML-HIDEL-TFENEVLQSE
    VKQIESKTAFQEALDAAGI
    VKQIESKAAFQEALDDAGI
    VKQIESKYAFQEALNSAGI
    VKQIESKSAFQEVLDSAGI
    VKSVGNLADFEAELKAAGI
    SNRVIHVSSCEELDKHLRI
    S-KVIHVTSNEELDKYLQI
    ASVRTMN-DYHKRIEAADI
    NQVKYFQSDFEQLIRQHPI
  IGCHTVEAWNEQLQKGNDTK
  SDGVKHINSAQEFANLLNTT
```

$$S(m_i) = \sum_{j<k} s(m_i^j, m_i^k)$$

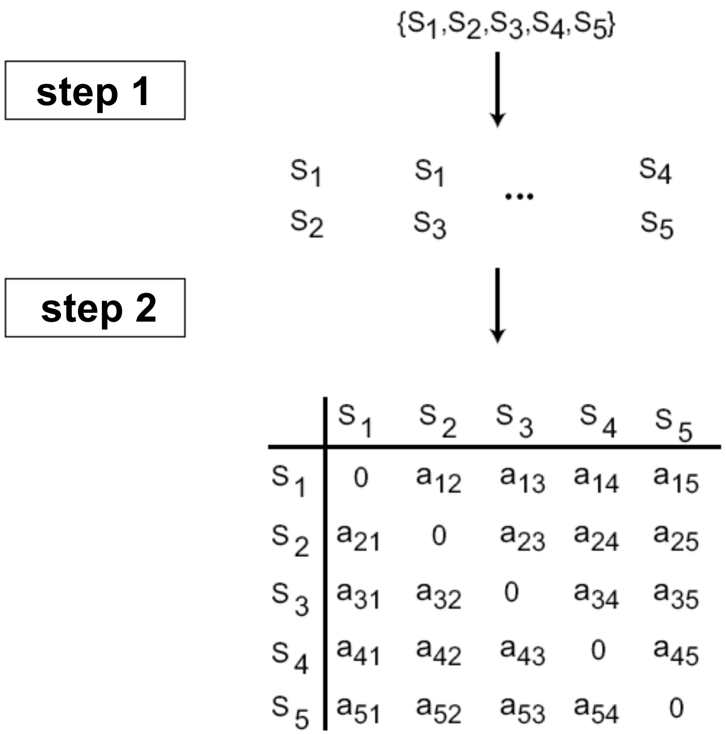
$$s(a, -) = s(-, a) = -d, \quad s(-, -) = 0$$

$$\alpha_{i_1,i_2,\ldots,i_N} = \max \begin{cases} G_0 \\ G_1 \\ G_2 \\ \vdots \\ G_{N-1} \end{cases},$$

$$G_0 = \alpha_{i_1-1,i_2-1,i_3-1,\ldots,i_N-1} + S\left(x^1_{i_1}, x^2_{i_2}, x^3_{i_3}, \ldots, x^N_{i_N}\right)$$

$$G_1 = \begin{cases} \alpha_{i_1,i_2-1,i_3-1,\ldots,i_N-1} + S\left(-, x^2_{i_2}, x^3_{i_3}, \ldots, x^N_{i_N}\right) \\ \alpha_{i_1-1,i_2,i_3-1,\ldots,i_N-1} + S\left(x^1_{i_1}, -, x^3_{i_3}, \ldots, x^N_{i_N}\right) \\ \quad\vdots \\ \alpha_{i_1-1,i_2-1,i_3-1,\ldots,i_N} + S\left(x^1_{i_1}, x^2_{i_2}, x^3_{i_3}, \ldots, -\right) \end{cases}$$

$$G_2 = \begin{cases} \alpha_{i_1,i_2,i_3-1,\ldots,i_N-1} + S\left(-, -, x^3_{i_3}, \ldots, x^N_{i_N}\right) \\ \alpha_{i_1-1,i_2,i_3,\ldots,i_N-1} + S\left(x^1_{i_1}, -, -, \ldots, x^N_{i_N}\right) \\ \quad\vdots \end{cases}$$



$\{S_1, S_2, S_3, S_4, S_5\}$

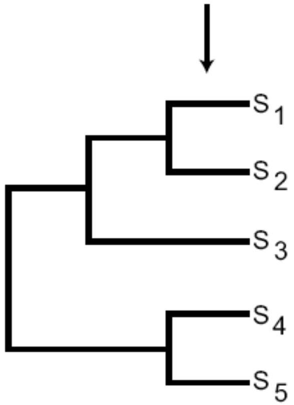
**step 1**

$S_1$ $S_2$ $S_1$ $S_3$ ... $S_4$ $S_5$

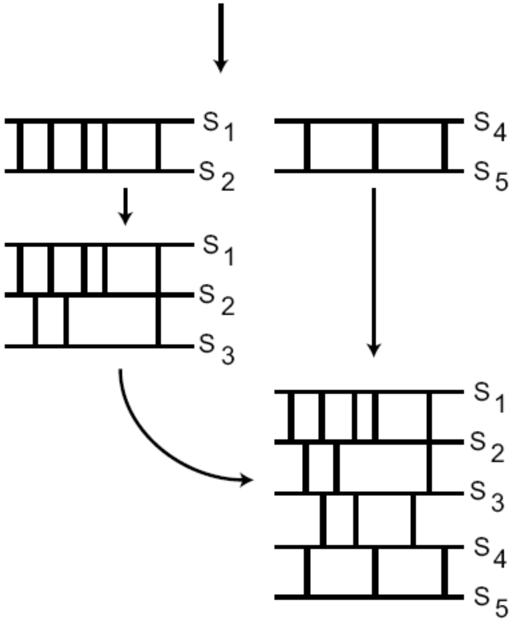
**step 2**

$$
\begin{array}{c|ccccc}
 & S_1 & S_2 & S_3 & S_4 & S_5 \\
\hline
S_1 & 0 & a_{12} & a_{13} & a_{14} & a_{15} \\
S_2 & a_{21} & 0 & a_{23} & a_{24} & a_{25} \\
S_3 & a_{31} & a_{32} & 0 & a_{34} & a_{35} \\
S_4 & a_{41} & a_{42} & a_{43} & 0 & a_{45} \\
S_5 & a_{51} & a_{52} & a_{53} & a_{54} & 0
\end{array}
$$

step 3



step 4



**a**

```
KEFHN-G--H--T
| ||  |  |  |
KYFHKAGNQHSPT
```

S=11

**b**

```
KEFH---NGHT
| ||   ||||
KYFHKAGNGHT
```

S=27

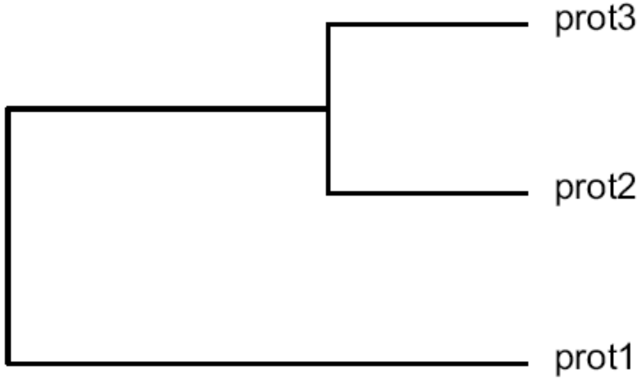
**c**

```
KYFHKAGNQHSPT
|||||||| |  |
KYFHKAGNGH--T
```

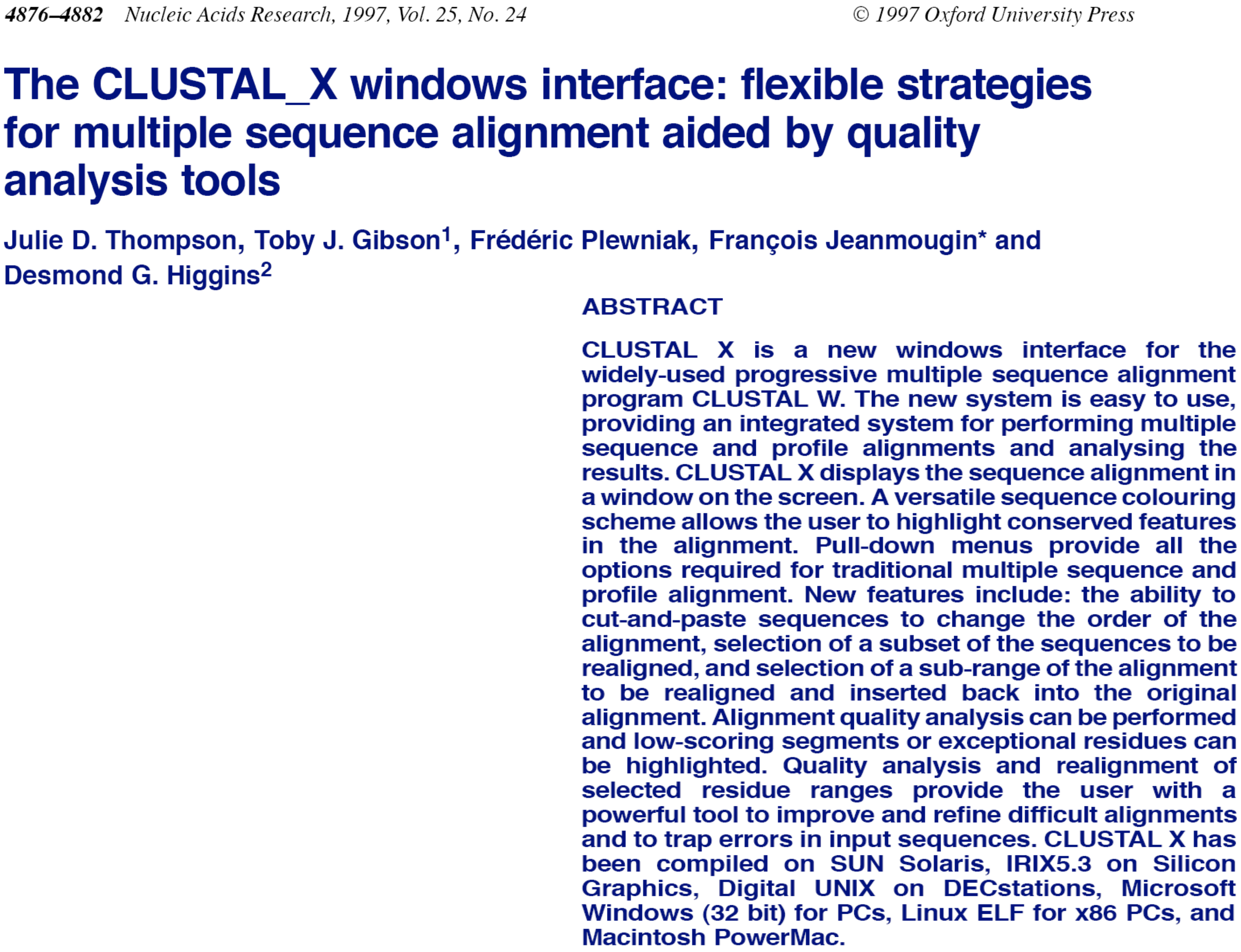
S=48

$$d = -\log \frac{S - S_{rand}}{S_{max} - S_{rand}}$$

| | prot1 | prot2 | prot3 |
|---|---|---|---|
| prot1 | 0 | 1.30429 | 0.75107 |
| prot2 | 1.30429 | 0 | 0.393446 |
| prot3 | 0.75107 | 0.393446 | 0 |



```
prot2   KYFHKAGNQHSPT
prot3   KYFHKAGNGH--T
prot1   KEFH---NGH--T
```



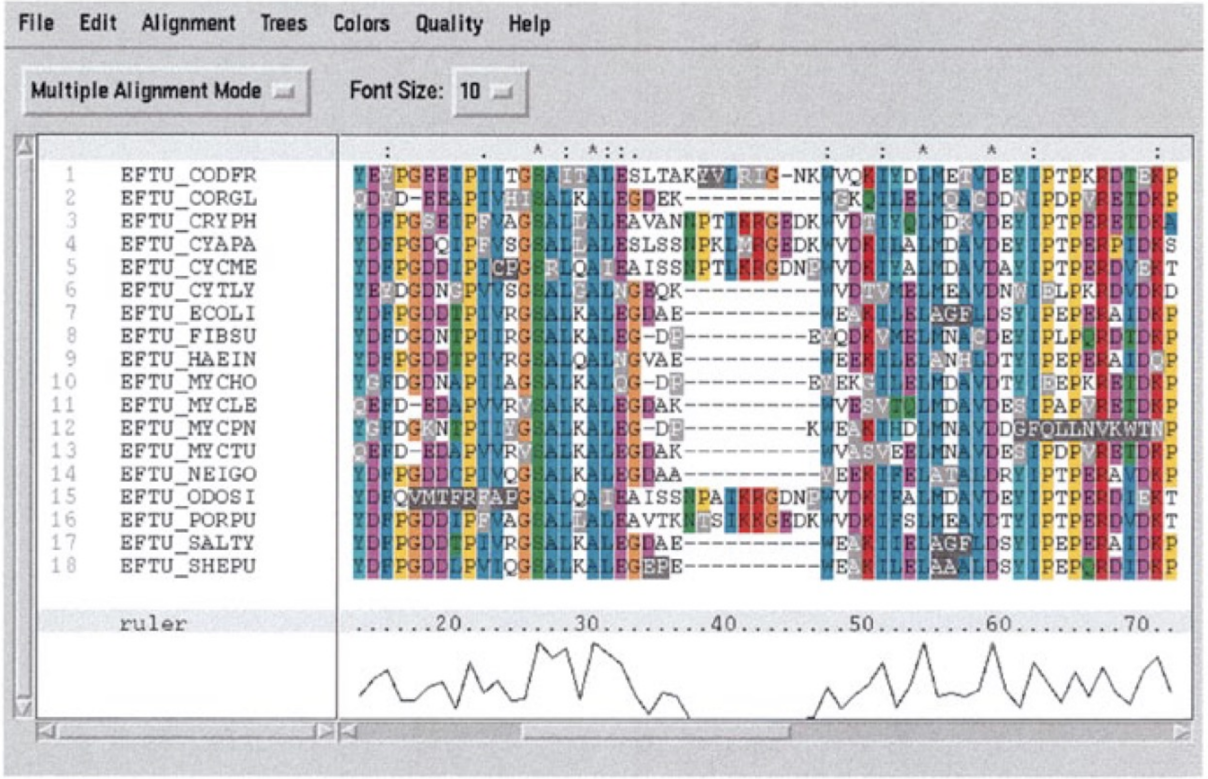
**4876–4882** *Nucleic Acids Research, 1997, Vol. 25, No. 24* *© 1997 Oxford University Press*

# The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools

Julie D. Thompson, Toby J. Gibson[1], Frédéric Plewniak, François Jeanmougin* and Desmond G. Higgins[2]

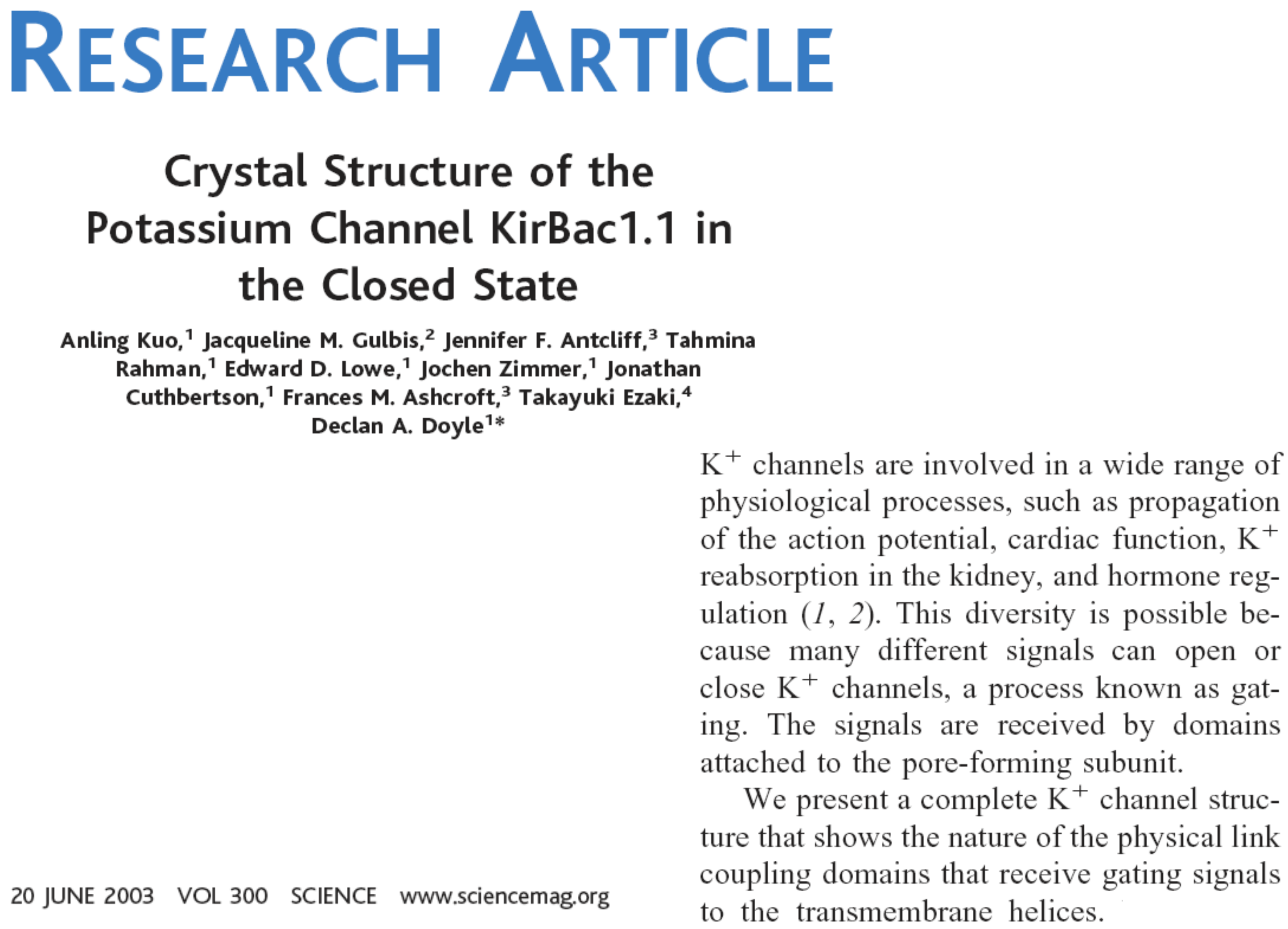
## ABSTRACT

CLUSTAL X is a new windows interface for the widely-used progressive multiple sequence alignment program CLUSTAL W. The new system is easy to use, providing an integrated system for performing multiple sequence and profile alignments and analysing the results. CLUSTAL X displays the sequence alignment in a window on the screen. A versatile sequence colouring scheme allows the user to highlight conserved features in the alignment. Pull-down menus provide all the options required for traditional multiple sequence and profile alignment. New features include: the ability to cut-and-paste sequences to change the order of the alignment, selection of a subset of the sequences to be realigned, and selection of a sub-range of the alignment to be realigned and inserted back into the original alignment. Alignment quality analysis can be performed and low-scoring segments or exceptional residues can be highlighted. Quality analysis and realignment of selected residue ranges provide the user with a powerful tool to improve and refine difficult alignments and to trap errors in input sequences. CLUSTAL X has been compiled on SUN Solaris, IRIX5.3 on Silicon Graphics, Digital UNIX on DECstations, Microsoft Windows (32 bit) for PCs, Linux ELF for x86 PCs, and Macintosh PowerMac.



```
                         :       .    * : *::.                 :   :  *   *  :          :
  1   EFTU_CODFR   YEYPGEEIPIITGSAITALESLTAKYVLRIG-NKWVQKIYDLMETVDEYIPTPKRDTEKP
  2   EFTU_CORGL   QDYD-EEAPIVHISALKALEGDEK----------WSKQILELMQACDDNIPDPVRETDKP
  3   EFTU_CRYPH   YDFPGSEIPFVAGSALLALEAVANNPTIKRGEDKWVDTIYQLMDKVDEYIPTPERETDKA
  4   EFTU_CYAPA   YDFPGDQIPFVSGSALLALESLSSNPKLMRGEDKWVDKILALMDAVDEYIPTPERPIDKS
  5   EFTU_CYCME   YDFPGDDIPICPGSRLQAIEAISSNPTLKRGDNPWVDKIYALMDAVDAYIPTPERDVEKT
  6   EFTU_CYTLY   YEYDGDNGPVVSGSALGALNGEQK----------WVDTVMELMEAVDNWIELPKRDVDKD
  7   EFTU_ECOLI   YDFPGDDTPIVRGSALKALEGDAE----------WEAKILELAGFLDSYIPEPERAIDKP
  8   EFTU_FIBSU   YDFDGDNTPIIRGSALKALEG-DP---------EYQDKVMELMNACDEYIPLPQRDTDKP
  9   EFTU_HAEIN   YDFPGDDTPIVRGSALQALNGVAE----------WEEKILELANHLDTYIPEPERAIDQP
 10   EFTU_MYCHO   YGFDGDNAPIIAGSALKALQG-DP---------EYEKGILELMDAVDTYIEEPKRETDKP
 11   EFTU_MYCLE   QEFD-EDAPVVRVSALKALEGDAK----------WVESVTQLMDAVDESIPAPVRETDKP
 12   EFTU_MYCPN   YGFDGKNTPIIYGSALKALEG-DP---------KWEAKIHDLMNAVDDGFQLLNVKWTNP
 13   EFTU_MYCTU   QEFD-EDAPVVRVSALKALEGDAK----------WVASVEELMNAVDESIPDPVRETDKP
 14   EFTU_NEIGO   YDFPGDDCPIVQGSALKALEGDAA----------YEEKIFELATALDRYIPTPERAVDKP
 15   EFTU_ODOSI   YDFQVMTFRFAPGSALQAIEAISSNPAIKRGDNPWVDKIFALMDAVDEYIPTPERDIEKT
 16   EFTU_PORPU   YDFPGDDIPFVAGSALLALEAVTKNTSIKKGEDKWVDKIFSLMEAVDTYIPTPERDVDKT
 17   EFTU_SALTY   YDFPGDDTPIVRGSALKALEGDAE----------WEAKIIELAGFLDSYIPEPERAIDKP
 18   EFTU_SHEPU   YDFPGDDLPVIQGSALKALEGEPE----------WEAKILELAAALDSYIPEPQRDIDKP

      ruler        ......20........30........40........50........60........70..
```

**Figure 1.** The CLUSTAL X window in multiple alignment mode. An alignment of some EFTU proteins is displayed. Low-scoring segments are highlighted using a white character on a black background. Exceptional residues are shown as a white character on a grey background. The quality analysis reveals two anomalously low scoring regions, ruler positions 16–25 in EFTU_ODOSI and 61–71 in EFTU_MYCPN. These were found to be caused by frameshift errors. Two more sequences (EFTU_RICPR and EFTU_SPIPL), not shown here, have 4-residue sequencing errors in this region which CLUSTAL X will also highlight.



# RESEARCH ARTICLE

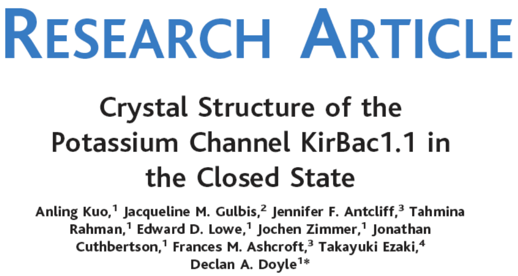
## Crystal Structure of the Potassium Channel KirBac1.1 in the Closed State

Anling Kuo,[1] Jacqueline M. Gulbis,[2] Jennifer F. Antcliff,[3] Tahmina Rahman,[1] Edward D. Lowe,[1] Jochen Zimmer,[1] Jonathan Cuthbertson,[1] Frances M. Ashcroft,[3] Takayuki Ezaki,[4] Declan A. Doyle[1]*

20 JUNE 2003 VOL 300 SCIENCE www.sciencemag.org

$K^+$ channels are involved in a wide range of physiological processes, such as propagation of the action potential, cardiac function, $K^+$ reabsorption in the kidney, and hormone regulation (*1*, *2*). This diversity is possible because many different signals can open or close $K^+$ channels, a process known as gating. The signals are received by domains attached to the pore-forming subunit.

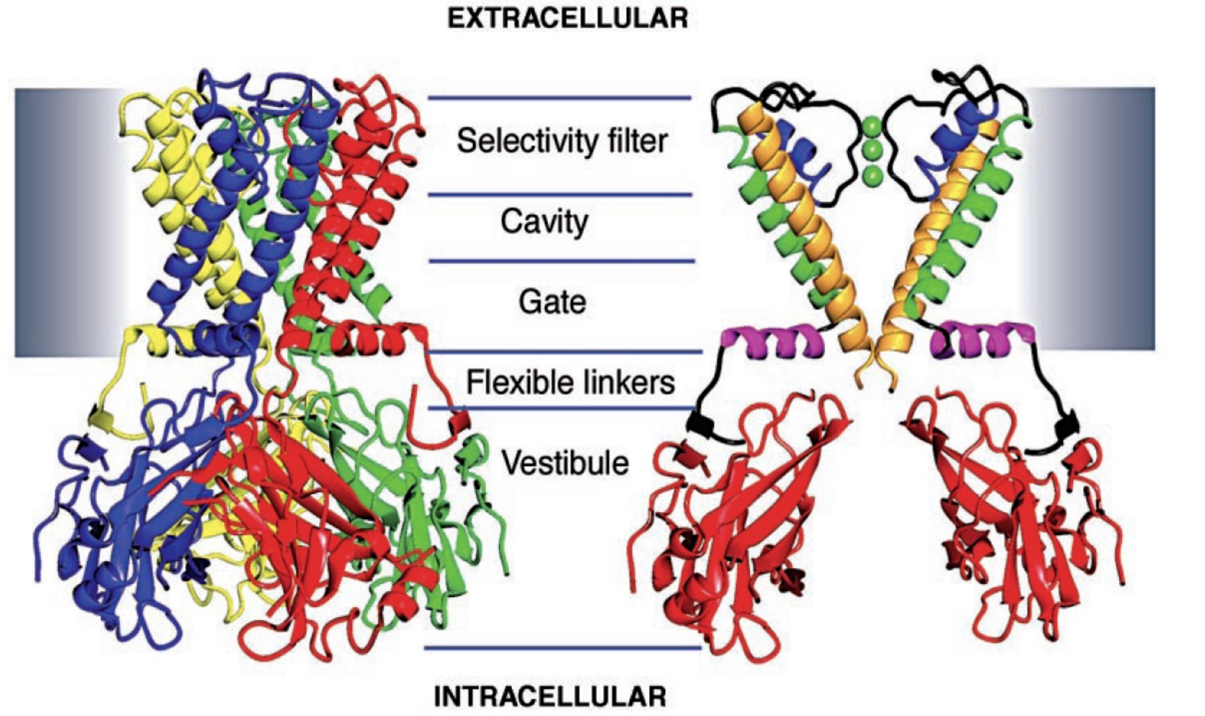
We present a complete $K^+$ channel structure that shows the nature of the physical link coupling domains that receive gating signals to the transmembrane helices.

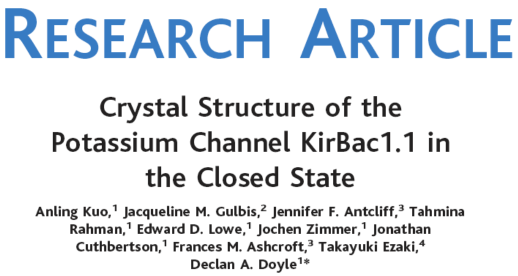


# Research Article

## Crystal Structure of the Potassium Channel KirBac1.1 in the Closed State

Anling Kuo,[1] Jacqueline M. Gulbis,[2] Jennifer F. Antcliff,[3] Tahmina Rahman,[1] Edward D. Lowe,[1] Jochen Zimmer,[1] Jonathan Cuthbertson,[1] Frances M. Ashcroft,[3] Takayuki Ezaki,[4] Declan A. Doyle[1*]

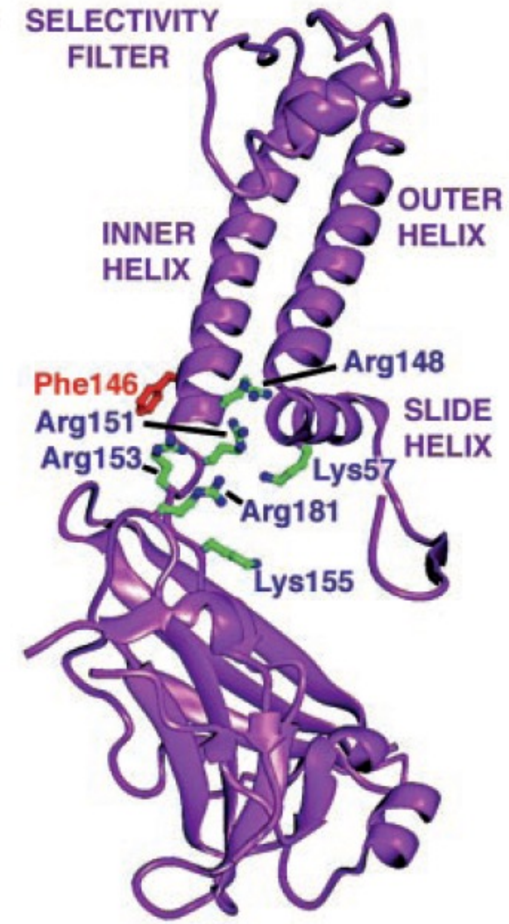




# Research Article

## Crystal Structure of the Potassium Channel KirBac1.1 in the Closed State

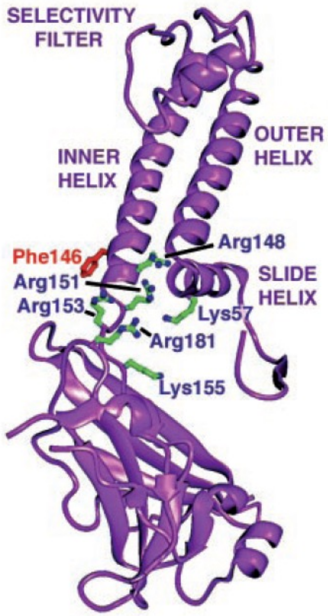
Anling Kuo,[1] Jacqueline M. Gulbis,[2] Jennifer F. Antcliff,[3] Tahmina Rahman,[1] Edward D. Lowe,[1] Jochen Zimmer,[1] Jonathan Cuthbertson,[1] Frances M. Ashcroft,[3] Takayuki Ezaki,[4] Declan A. Doyle[1*]

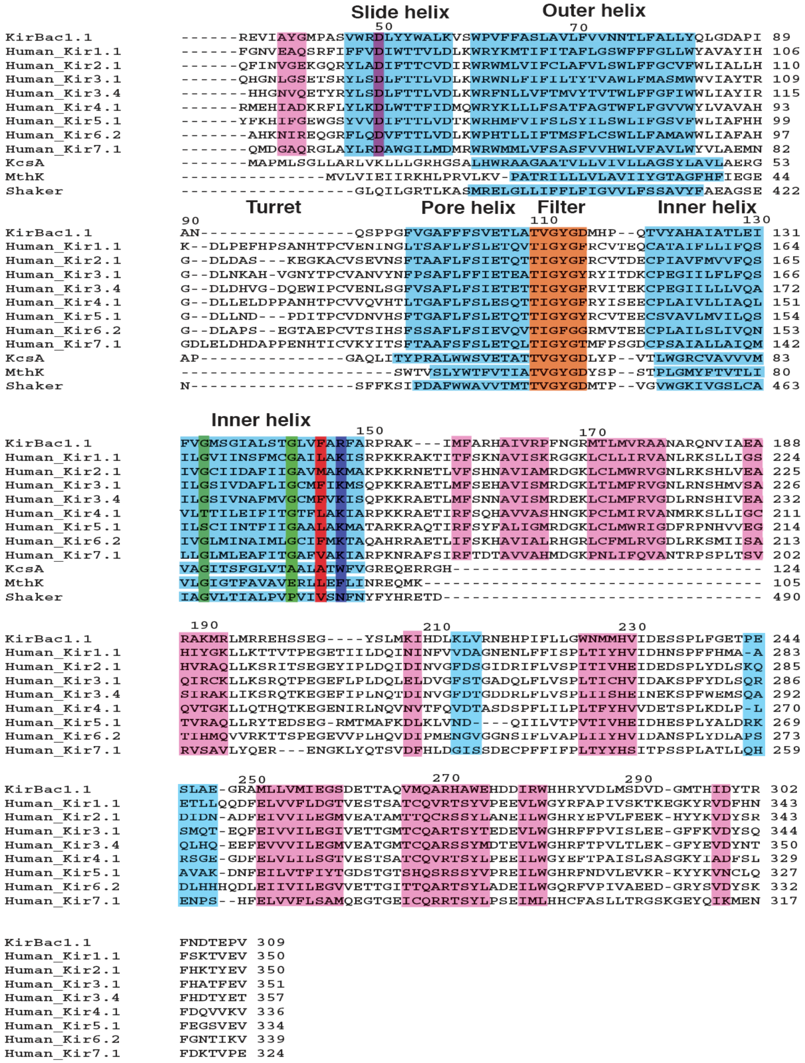


# Research Article

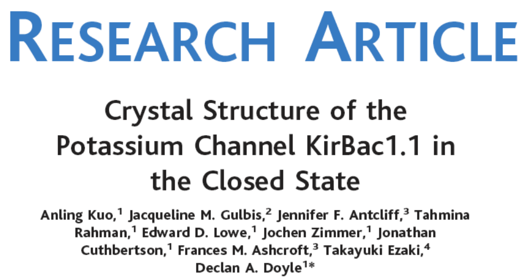
## Crystal Structure of the Potassium Channel KirBac1.1 in the Closed State

Anling Kuo,[1] Jacqueline M. Gulbis,[2] Jennifer F. Antcliff,[3] Tahmina Rahman,[1] Edward D. Lowe,[1] Jochen Zimmer,[1] Jonathan Cuthbertson,[1] Frances M. Ashcroft,[3] Takayuki Ezaki,[4] Declan A. Doyle[1]*





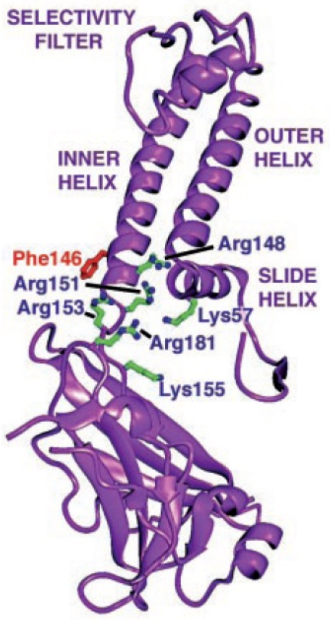
```
                                       Slide helix        Outer helix
                                         50                 70
KirBac1.1        ------REVIAYGMPASVWRDLYYWALKVSWPVFFASLAVLFVVNNTLFALLYQLGDAPI 89
Human_Kir1.1     ------FGNVEAQSRFIFFVDIWTTVLDLKWRYKMTIFITAFLGSWFFFGLLWYAVAYIH 106
Human_Kir2.1     ------QFINVGEKGQRYLADIFTTCVDIRWRWMLVIFCLAFVLSWLFFGCVFWLIALLH 110
Human_Kir3.1     ------QHGNLGSETSRYLSDLFTTLVDLKWRWNLFIFILTYTVAWLFMASMWWVIAYTR 109
Human_Kir3.4     -------HHGNVQETYRYLSDLFTTLVDLKWRFNLLVFTMVYTVTWLFFGFIWWLIAYIR 115
Human_Kir4.1     ------RMEHIADKRFLYLKDLWTTFIDMQWRYKLLLFSATFAGTWFLFGVVWYLVAVAH 93
Human_Kir5.1     ------YFKHIFGEWGSYVVDIFTTLVDTKWRHMFVIFSLSYILSWLIFGSVFWLIAFHH 99
Human_Kir6.2     -------AHKNIREQGRFLQDVFTTLVDLKWPHTLLIFTMSFLCSWLLFAMAWWLIAFAH 97
Human_Kir7.1     -------QMDGAQRGLAYLRDAWGILMDMRWRWMMLVFSASFVVHWLVFAVLWYVLAEMN 82
KcsA             -------MAPMLSGLLARLVKLLLGRHGSALHWRAAGAATVLLVIVLLAGSYLAVLAERG 53
MthK             ---------------MVLVIEIIRKHLPRVLKV-PATRILLLVLAVIIYGTAGFHFIEGE 44
Shaker           ------------------GLQILGRTLKASMRELGLLIFFLFIGVVLFSSAVYFAEAGSE 422

                        Turret             Pore helix  Filter      Inner helix
                 90                                   110                 130
KirBac1.1        AN-----------------QSPPGFVGAFFFSVETLATVGYGDMHP--QTVYAHAIATLEI 131
Human_Kir1.1     K--DLPEFHPSANHTPCVENINGLTSAFLFSLETQVTIGYGFRCVTEQCATAIFLLIFQS 164
Human_Kir2.1     G--DLDAS---KEGKACVSEVNSFTAAFLFSIETQTTIGYGFRCVTDECPIAVFMVVFQS 165
Human_Kir3.1     G--DLNKAH-VGNYTPCVANVYNFPSAFLFFIETEATIGYGYRYITDKCPEGIILFLFQS 166
Human_Kir3.4     G--DLDHVG-DQEWIPCVENLSGFVSAFLFSIETETTIGYGFRVITEKCPEGIILLLVQA 172
Human_Kir4.1     G--DLLELDPPANHTPCVVQVHTLTGAFLFSLESQTTIGYGFRYISEECPLAIVLLIAQL 151
Human_Kir5.1     G--DLLND---PDITPCVDNVHSFTGAFLFSLETQTTIGYGYRCVTEECSVAVLMVILQS 154
Human_Kir6.2     G--DLAPS--EGTAEPCVTSIHSFSSAFLFSIEVQVTIGFGGRMVTEECPLAILSLIVQN 153
Human_Kir7.1     GDLELDHDAPPENHTICVKYITSFTAAFSFSLETQLTIGYGTMFPSGDCPSAIALLAIQM 142
KcsA             AP----------------GAQLITYPRALWWSVETATTVGYGDLYP--VTLWGRCVAVVVM 83
MthK             ------------------------SWTVSLYWTFVTIATVGYGDYSP--STPLGMYFTVTLI 80
Shaker           N-----------------SFFKSIPDAFWWAVVTMTTVGYGDMTP--VGVWGKIVGSLCA 463

                     Inner helix
                                150                 170
KirBac1.1        FVGMSGIALSTGLVFARFARPRAK---IMFARHAIVRPFNGRMTLMVRAANARQNVIAEA 188
Human_Kir1.1     ILGVIINSFMCGAILAKISRPKKRAKTITFSKNAVISKRGGKLCLLIRVANLRKSLLIGS 224
Human_Kir2.1     IVGCIIDAFIIGAVMAKMAKPKKRNETLVFSHNAVIAMRDGKLCLMWRVGNLRKSHLVEA 225
Human_Kir3.1     ILGSIVDAFLIGCMFIKMSQPKKRAETLMFSEHAVISMRDGKLTLMFRVGNLRNSHMVSA 226
Human_Kir3.4     ILGSIVNAFMVGCMFVKISQPKKRAETLMFSNNAVISMRDEKLCLMFRVGDLRNSHIVEA 232
Human_Kir4.1     VLTTILEIFITGTFLAKIARPKKRAETIRFSQHAVVASHNGKPCLMIRVANMRKSLLIGC 211
Human_Kir5.1     ILSCIINTFIIGAALAKMATARKRAQTIRFSYFALIGMRDGKLCLMWRIGDFRPNHVVEG 214
Human_Kir6.2     IVGLMINAIMLGCIFMKTAQAHRRAETLIFSKHAVIALRHGRLCFMLRVGDLRKSMIISA 213
Human_Kir7.1     LLGLMLEAFITGAFVAKIARPKNRAFSIRFTDTAVVAHMDGKPNLIFQVANTRPSPLTSV 202
KcsA             VAGITSFGLVTAALATWFVGREQERRGH-------------------------------- 124
MthK             VLGIGTFAVAVERLLEFLINREQMK----------------------------------- 105
Shaker           IAGVLTIALPVPVIVSNFNYFYHRETD--------------------------------- 490

                  190                210                230
KirBac1.1        RAKMRLMRREHSSEG----YSLMKIHDLKLVRNEHPIFLLGWNMMHVIDESSPLFGETPE 244
Human_Kir1.1     HIYGKLLKTTVTPEGETIILDQINIFVVDAGNENLFFISPLTIYHVIDHNSPFFHMA-A 283
Human_Kir2.1     HVRAQLLKSRITSEGEYIPLDQIDINVGFDSGIDRIFLVSPITIVHEIDEDSPLYDLSKQ 285
Human_Kir3.1     QIRCKLLKSRQTPEGEFLPLDQLELDVGFSTGADQLFLVSPLTICHVIDAKSPFYDLSQR 286
Human_Kir3.4     SIRAKLIKSRQTKEGEFIPLNQTDINVGFDTGDDRLFLVSPLIISHEINEKSPFWEMSQA 292
Human_Kir4.1     QVTGKLLQTHQTKEGENIRLNQVNVTFQVDTASDSPFLILPLTFYHVVDETSPLKDLP-L 270
Human_Kir5.1     TVRAQLLRYTEDSEG-RMTMAFKDLKLVND----QIILVTPVTIVHEIDHESPLYALDRK 269
Human_Kir6.2     TIHMQVVRKTTSPEGEVVPLHQVDIPMENGVGGNSIFLVAPLIIYHVIDANSPLYDLAPS 273
Human_Kir7.1     RVSAVLYQER---ENGKLYQTSVDFHLDGISSDECPFFIFPLTYYHSITPSSPLATLLQH 259

                      250                270                290
KirBac1.1        SLAE-GRAMLLVMIEGSDETTAQVMQARHAWEHDDIRWHHRYVDLMSDVD-GMTHIDYTR 302
Human_Kir1.1     ETLLQQDFELVVFLDGTVESTSATCQVRTSYVPEEVLWGYRFAPIVSKTKEGKYRVDFHN 343
Human_Kir2.1     DIDN-ADFEIVVILEGMVEATAMTTQCRSSYLANEILWGHRYEPVLFEEK-HYYKVDYSR 343
Human_Kir3.1     SMQT-EQFEIVVILEGIVETTGMTCQARTSYTEDEVLWGHRFFPVISLEE-GFFKVDYSQ 344
Human_Kir3.4     QLHQ-EEFEVVVILEGMVEATGMTCQARSSYMDTEVLWGHRFTPVLTLEK-GFYEVDYNT 350
Human_Kir4.1     RSGE-GDFELVLILSGTVESTSATCQVRTSYLPEEILWGYEFTPAISLSASGKYIADFSL 329
Human_Kir5.1     AVAK-DNFEILVTFIYTGDSTGTSHQSRSSYVPREILWGHRFNDVLEVKR-KYYKVNCLQ 327
Human_Kir6.2     DLHHHQDLEIIVILEGVVETTGITTQARTSYLADEILWGQRFVPIVAEED-GRYSVDYSK 332
Human_Kir7.1     ENPS--HFELVVFLSAMQEGTGEICQRRTSYLPSEIMLHHCFASLLTRGSKGEYQIKMEN 317

KirBac1.1        FNDTEPV 309
Human_Kir1.1     FSKTVEV 350
Human_Kir2.1     FHKTYEV 350
Human_Kir3.1     FHATFEV 351
Human_Kir3.4     FHDTYET 357
Human_Kir4.1     FDQVVKV 336
Human_Kir5.1     FEGSVEV 334
Human_Kir6.2     FGNTIKV 339
Human_Kir7.1     FDKTVPE 324
```

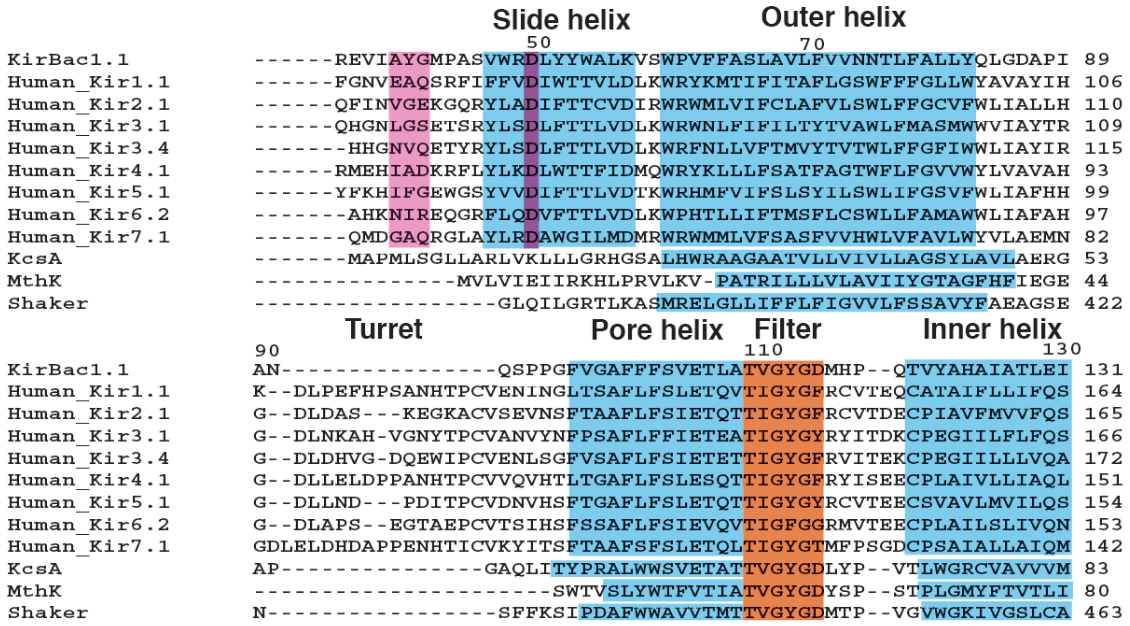


# Research Article

## Crystal Structure of the Potassium Channel KirBac1.1 in the Closed State

Anling Kuo,[1] Jacqueline M. Gulbis,[2] Jennifer F. Antcliff,[3] Tahmina Rahman,[1] Edward D. Lowe,[1] Jochen Zimmer,[1] Jonathan Cuthbertson,[1] Frances M. Ashcroft,[3] Takayuki Ezaki,[4] Declan A. Doyle[1]*





```
                                    Slide helix        Outer helix
                                    50                  70
KirBac1.1     ------REVIAYGMPASVWRDLYYWALKVSWPVFFASLAVLFVVNNTLFALLYQLGDAPI 89
Human_Kir1.1  ------FGNVEAQSRFIFFVDIWTTVLDLKWRYKMTIFITAFLGSWFFFGLLWYAVAYIH 106
Human_Kir2.1  ------QFINVGEKGQRYLADIFTTCVDIRWRWMLVIFCLAFVLSWLFFGCVFWLIALLH 110
Human_Kir3.1  ------QHGNLGSETSRYLSDLFTTLVDLKWRWNLFIFILTYTVAWLFMASMWWVIAYTR 109
Human_Kir3.4  -------HHGNVQETYRYLSDLFTTLVDLKWRFNLLVFTMVYTVTWLFFGFIWWLIAYIR 115
Human_Kir4.1  ------RMEHIADKRFLYLKDLWTTFIDMQWRYKLLLFSATFAGTWFLFGVVWYLVAVAH 93
Human_Kir5.1  ------YFKHIFGEWGSYVVDIFTTLVDTKWRHMFVIFSLSYILSWLIFGSVFWLIAFHH 99
Human_Kir6.2  -------AHKNIREQGRFLQDVFTTLVDLKWPHTLLIFTMSFLCSWLLFAMAWWLIAFAH 97
Human_Kir7.1  -------QMDGAQRGLAYLRDAWGILMDMRWRWMMLVFSASFVVHWLVFAVLWYVLAEMN 82
KcsA          -------MAPMLSGLLARLVKLLLGRHGSALHWRAAGAATVLLVIVLLAGSYLAVLAERG 53
MthK          ---------------MVLVIEIIRKHLPRVLKV-PATRILLLVLAVIIYGTAGFHFIEGE 44
Shaker        ------------------GLQILGRTLKASMRELGLLIFFLFIGVVLFSSAVYFAEAGSE 422

                    Turret                    Pore helix   Filter     Inner helix
              90                                  110                   130
KirBac1.1     AN----------------QSPPGFVGAFFFSVETLATVGYGDMHP--QTVYAHAIATLEI 131
Human_Kir1.1  K--DLPEFHPSANHTPCVENINGLTSAFLFSLETQVTIGYGFRCVTEQCATAIFLLIFQS 164
Human_Kir2.1  G--DLDAS---KEGKACVSEVNSFTAAFLFSIETQTTIGYGFRCVTDECPIAVFMVVFQS 165
Human_Kir3.1  G--DLNKAH-VGNYTPCVANVYNFPSAFLFFIETEATIGYGYRYITDKCPEGIILFLFQS 166
Human_Kir3.4  G--DLDHVG-DQEWIPCVENLSGFVSAFLFSIETETTIGYGFRVITEKCPEGIILLLVQA 172
Human_Kir4.1  G--DLLELDPPANHTPCVVQVHTLTGAFLFSLESQTTIGYGFRYISEECPLAIVLLIAQL 151
Human_Kir5.1  G--DLLND---PDITPCVDNVHSFTGAFLFSLETQTTIGYGYRCVTEECSVAVLMVILQS 154
Human_Kir6.2  G--DLAPS--EGTAEPCVTSIHSFSSAFLFSIEVQVTIGFGGRMVTEECPLAILSLIVQN 153
Human_Kir7.1  GDLELDHDAPPENHTICVKYITSFTAAFSFSLETQLTIGYGTMFPSGDCPSAIALLAIQM 142
KcsA          AP---------------GAQLITYPRALWWSVETATTVGYGDLYP--VTLWGRCVAVVVM 83
MthK          ----------------------SWTVSLYWTFVTIATVGYGDYSP--STPLGMYFTVTLI 80
Shaker        N-----------------SFFKSIPDAFWWAVVTMTTVGYGDMTP--VGVWGKIVGSLCA 463
```