

Comparing pairs of sequences

Objectives:

To align two sequences in a biologically logical fashion.

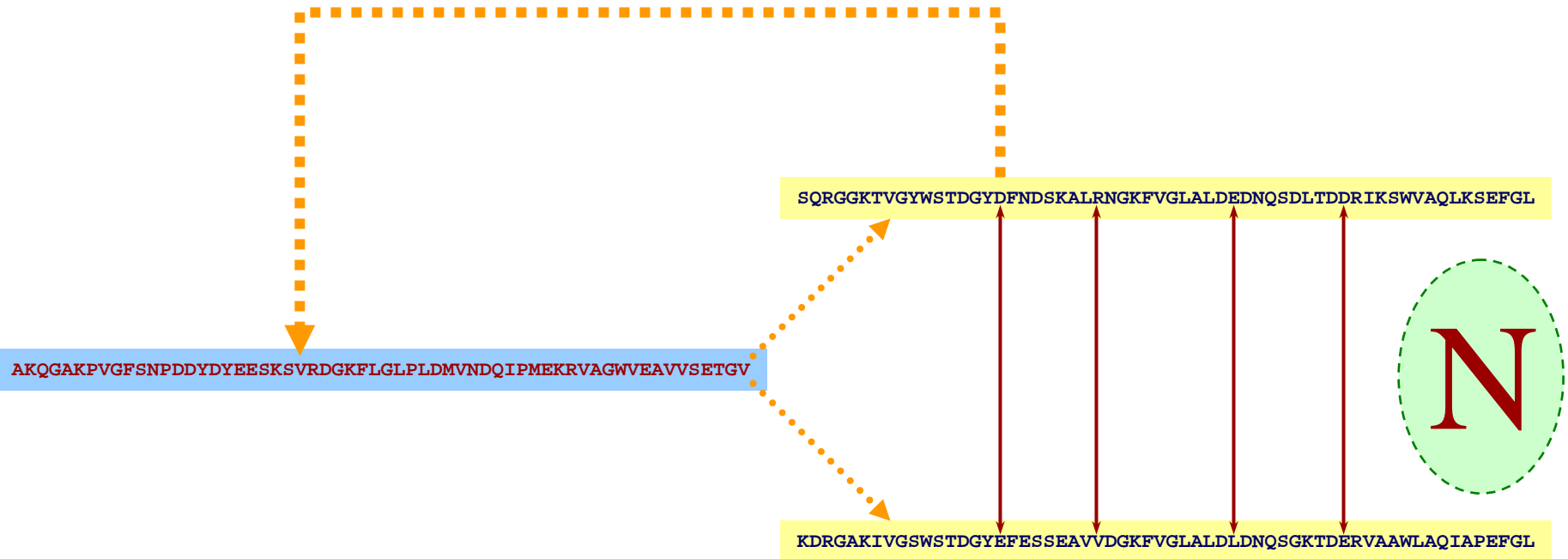
To generate a score representing the “quality” of the alignment.

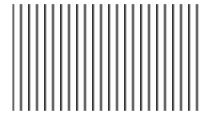
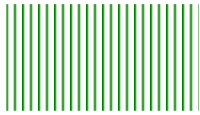
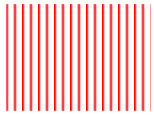
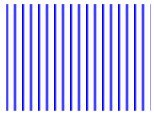
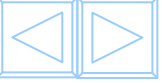
{ Sadly, the computed score has little absolute meaning }

Assumption:

The sequences being compared are homologous.

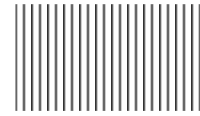
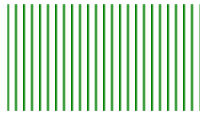
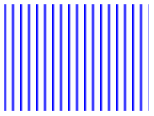
{ so differences between the sequences are exclusively due to evolutionary processes }

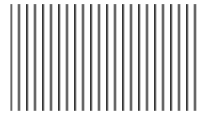
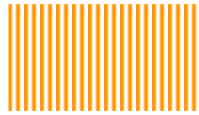
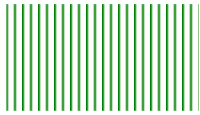
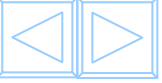




ACBEERGYALEDILAGERAFGSTOUTFAWATERM

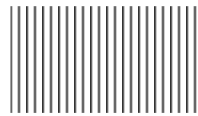
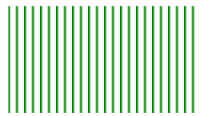
ABEERNALEDLAGERDFWGALSTOUTWRARWATERA

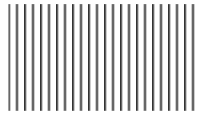
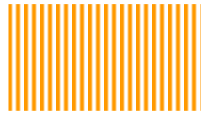
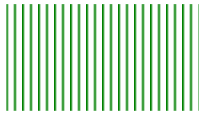
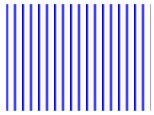
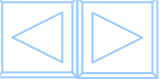




ACBEERGYALEDILAGERAFGSTOUTFAWATERM

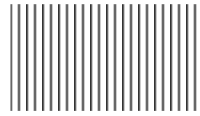
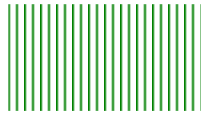
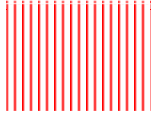
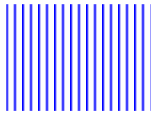
ABEERNALEDLAGERDFWGALSTOUTWRARWATERA

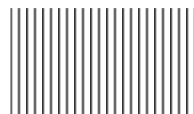
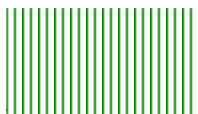
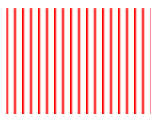
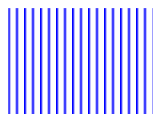
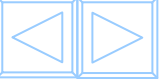




ACBEERGYALEDILAGERAFGSTOUTFAWATERM

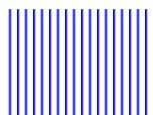
ABEERNALEDLAGERDFWGALSTOUTWRARWATERA

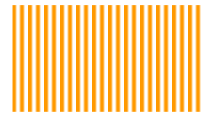
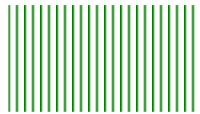
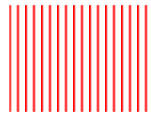
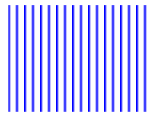
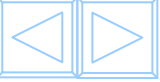




ACBEERGYALEDILAGERAFGSTOUTFAWATERM

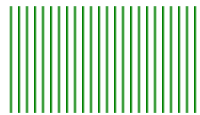
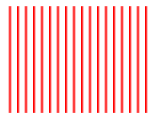
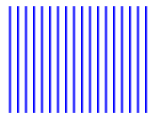
ABEERNALEDLAGERDFWGALSTOUTWRARWATERA





ACBEERGYALEDILAGERAFGSTOUTFAWATERM

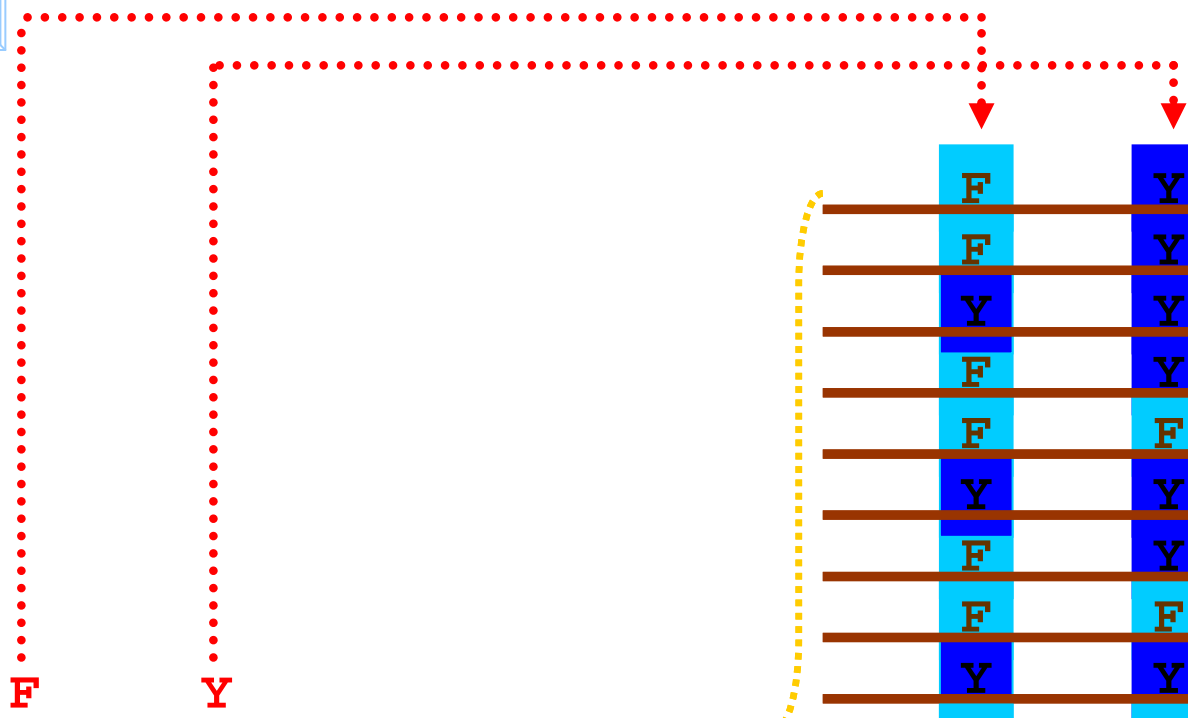
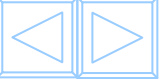
ABEERNALEDLAGERDFWGALSTOUTWRARWATERA



Original protein

**The PAM matrices
are computed from
aligned families
of proteins.**

Aligned current proteins



F	Y
F	Y
Y	Y
F	Y
F	F
Y	Y
F	Y
F	F
Y	Y
F	F
Y	Y
F	F
F	Y
Y	F
F	F
F	Y
Y	Y
F	Y
F	F

Original protein

The High $F \leftrightarrow Y$ score implies:

Where F is conserved

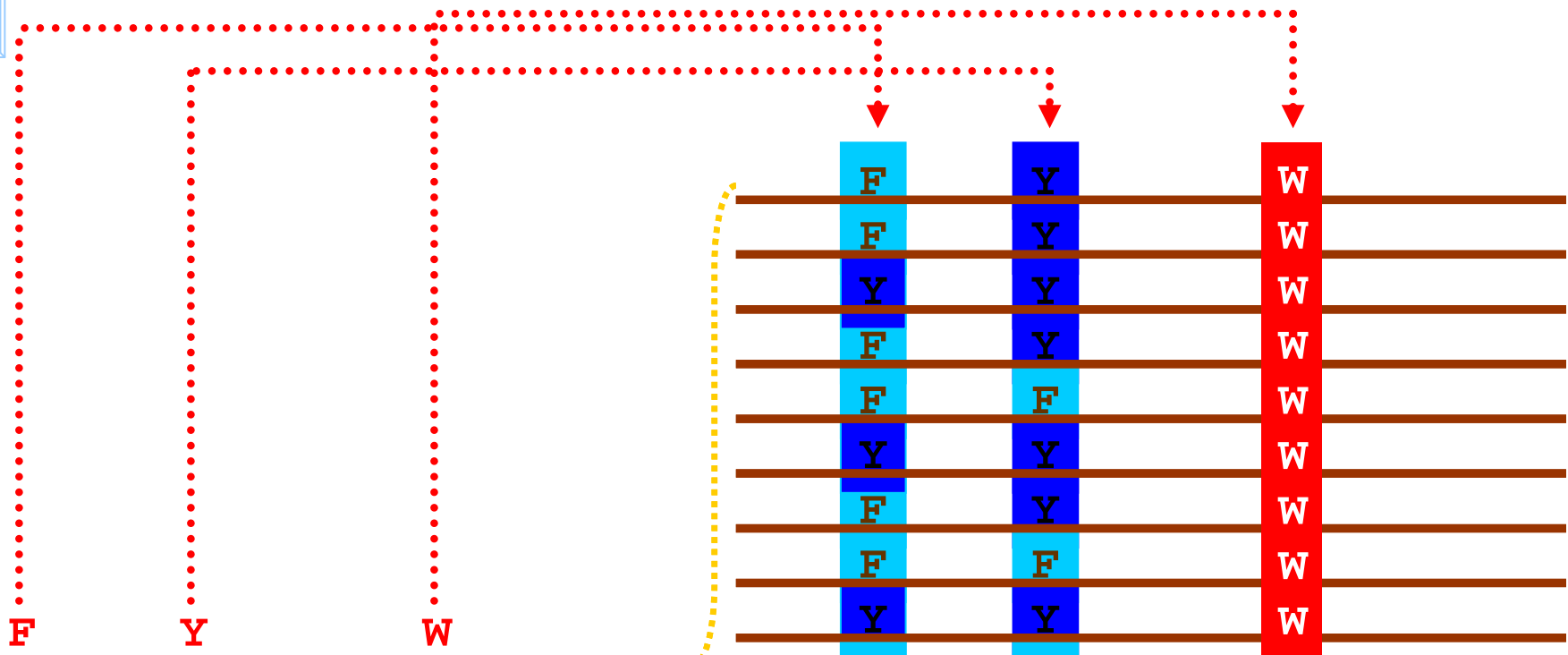
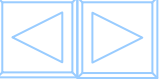
F -> Y substitutions are common

+

Where Y is conserved

Y -> F substitutions are common

Aligned current proteins



F Y W

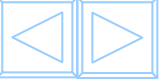
Original protein

The High $W \longleftrightarrow W$ score implies:

Where W is conserved

Any substitutions are uncommon

Aligned current proteins



Typical Amino Acid composition

{according to Argos and McCaldon}

Alanine is very common

AMINO ACIDS	%
A	8.3
C	1.7
D	5.3
E	6.2
F	3.9
G	7.2
H	2.2
I	5.2
K	5.7
L	9.0
M	2.4
N	4.4
P	5.1
Q	4.0
R	5.7
S	6.9
T	5.8
V	6.6
W	1.3
Y	3.2

Tryptophan is relatively rare

PAM - Point Accepted Mutation

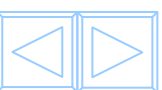
PAM - is a measure of evolutionary distance

An evolutionary distance of 1 PAM indicates the probability of a residue mutating during a distance in which 1 point mutation was accepted per 100 residues.

Relationship between Observed difference and PAM value.

Observed %
Difference Evolutionary
Distance (PAMs)

1	1
10	11
20	23
30	38
40	56
50	80
60	112
70	159
80	246



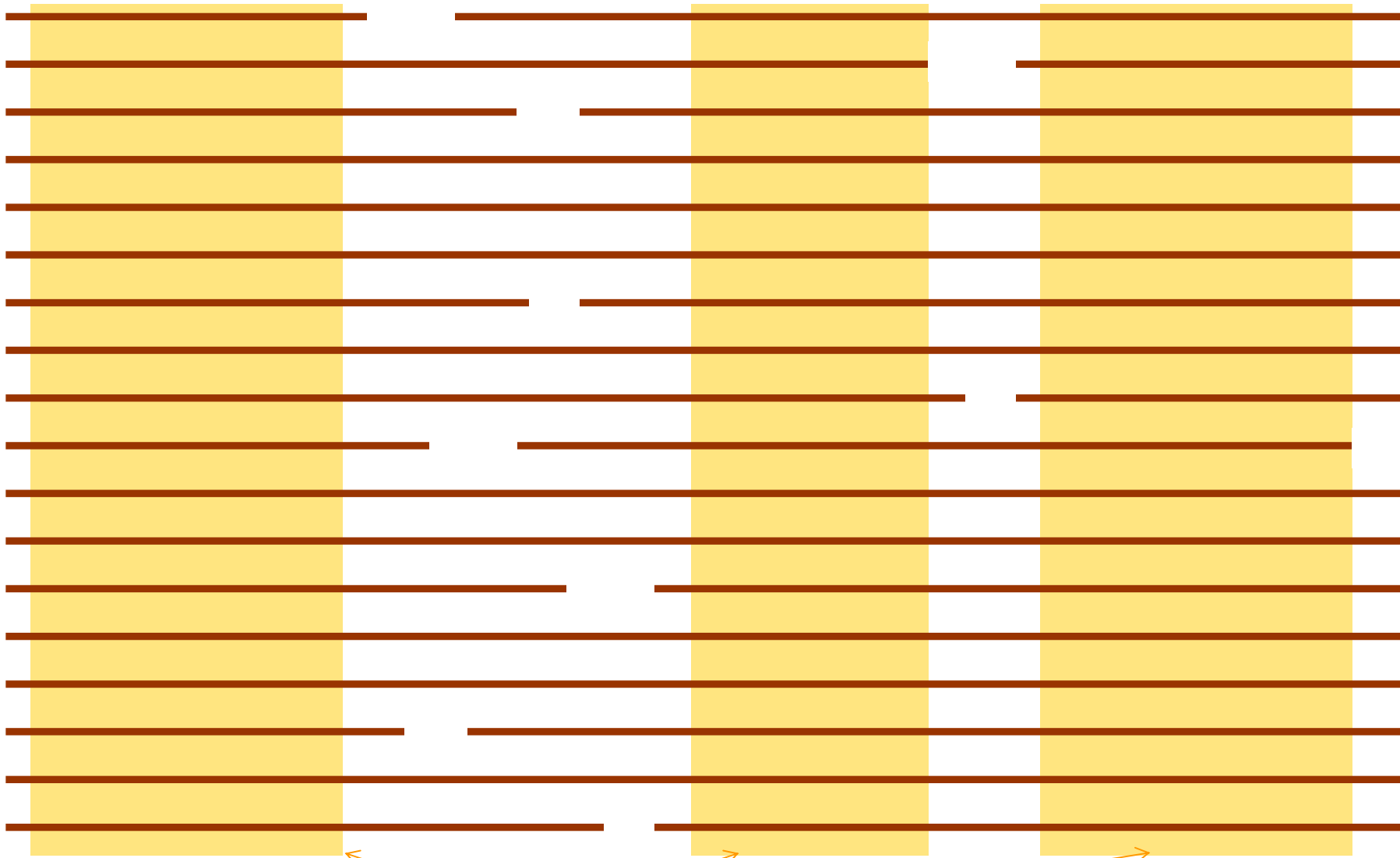
The BLOSUM scoring matrices

Original protein

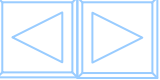
The BLOSUM matrices are also computed from aligned families of proteins.

Aligned current proteins

The BLOSUM scoring matrices



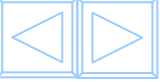
Regions of high conservation, stored in the BLOCKS database



Insertions/Deletions => GAPS

ACBEERGYALEDILAGERAFGSTOUTFAWATERM

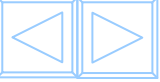
ABEERNALEDLAGERDFWGALSTOUTWRARWATERA



Insertions/Deletions => GAPS

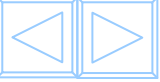
ACBEERGYALEDILAGERAFGSTOUTFAWATERM

ABEERNALEDLAGERDFWGALSTOUTWRARWATERA



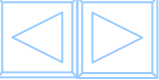
Insertions/Deletions => GAPS

ACBEERGYALEDILAGERAFGSTOUTFAWATERM
ABEERN-ALEDLAGERDFWGALSTOUTWRARWATERA



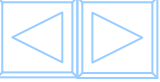
Insertions/Deletions => GAPS

ACBEERGYALEDILAGERAFGSTOUTFAWATERM
ABEERN-ALED-LAGERDFWGALSTOUTWRARWATERA



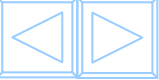
Insertions/Deletions => GAPS

ACBEERGYALEDILAGERAFG-STOUTFAWATERM
ABEERN-ALED-LAGERDFWGALSTOUTWRARWATERA



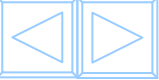
Insertions/Deletions => GAPS

ACBEERGYALEDILAGERAFG--STOUTFAWATERM
ABEERN-ALED-LAGERDFWGALSTOUTWRARWATERA



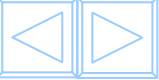
Insertions/Deletions => GAPS

ACBEERGYALEDILAGERAFG---STOUTFAWATERM
ABEERN-ALED-LAGERDFWGALSTOUTWRARWATERA



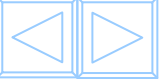
Insertions/Deletions => GAPS

ACBEERGYALEDILAGERAFG---STOUTFA-WATERM
ABEERN-ALED-LAGERDFWGALSTOUTWRARWATERA



Insertions/Deletions => GAPS

ACBEERGYALEDILAGERAFG---STOUTFA--WATERM
ABEERN-ALED-LAGERDFWGALSTOUTWRARWATERA

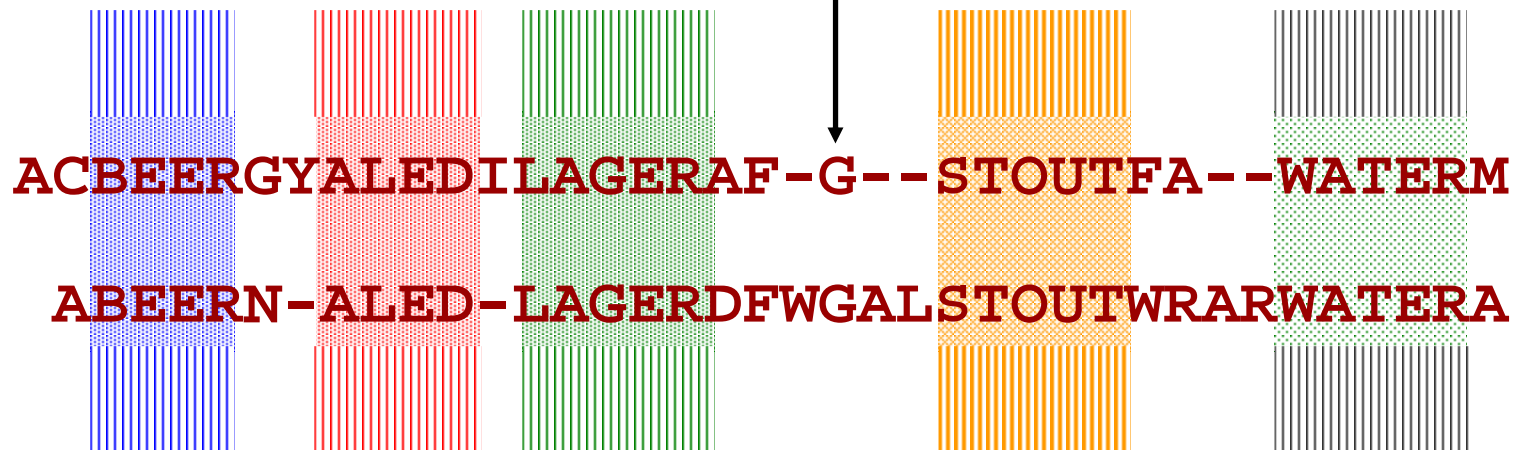


Insertions/Deletions => GAPS

?????? - but adds 12 to the score

G<->W = -7

G<->G = +5

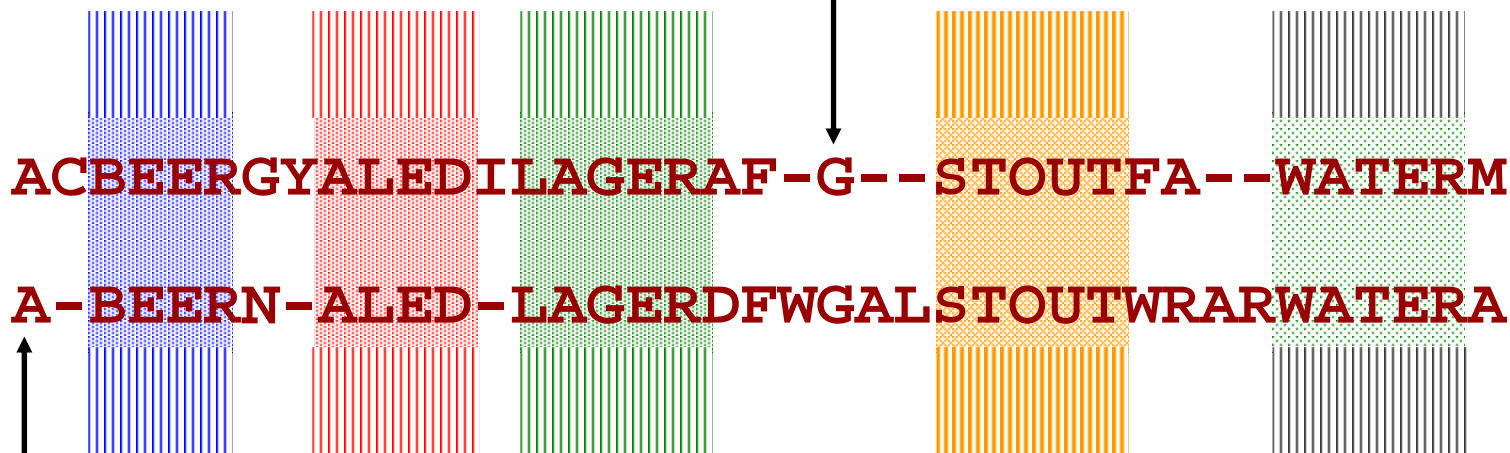




The need for penalising GAPS.

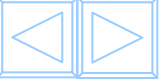
?????? - but adds 12 to the score

G<->W = -7
G<->G = +5



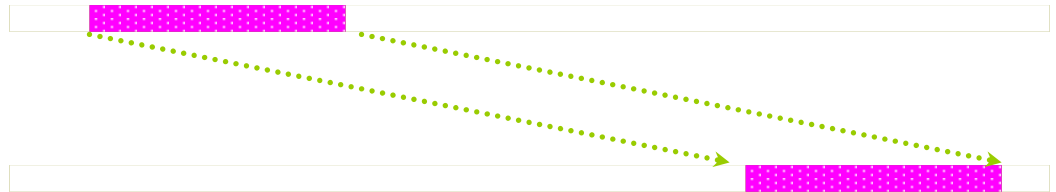
??????? - but adds 4 to the score

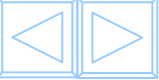
A<->C = -2
A<->A = +2



LOCAL and GLOBAL implementations of pairwise alignment

LOCAL Alignment





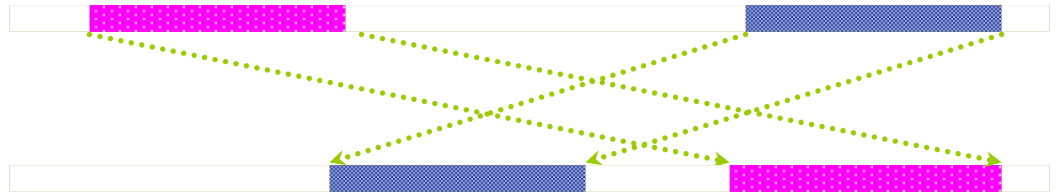
LOCAL and GLOBAL implementations of pairwise alignment

LOCAL Alignment

GCG : **bestfit**

Staden : **spin**

Emboss : **water/matcher**



GLOBAL Alignment

GCG : **gap**

Staden : **spin**

Emboss : **needle/stretcher**

