

3.3 Database searching for similar sequences

3.3.1 Importance of database searches for similar sequences

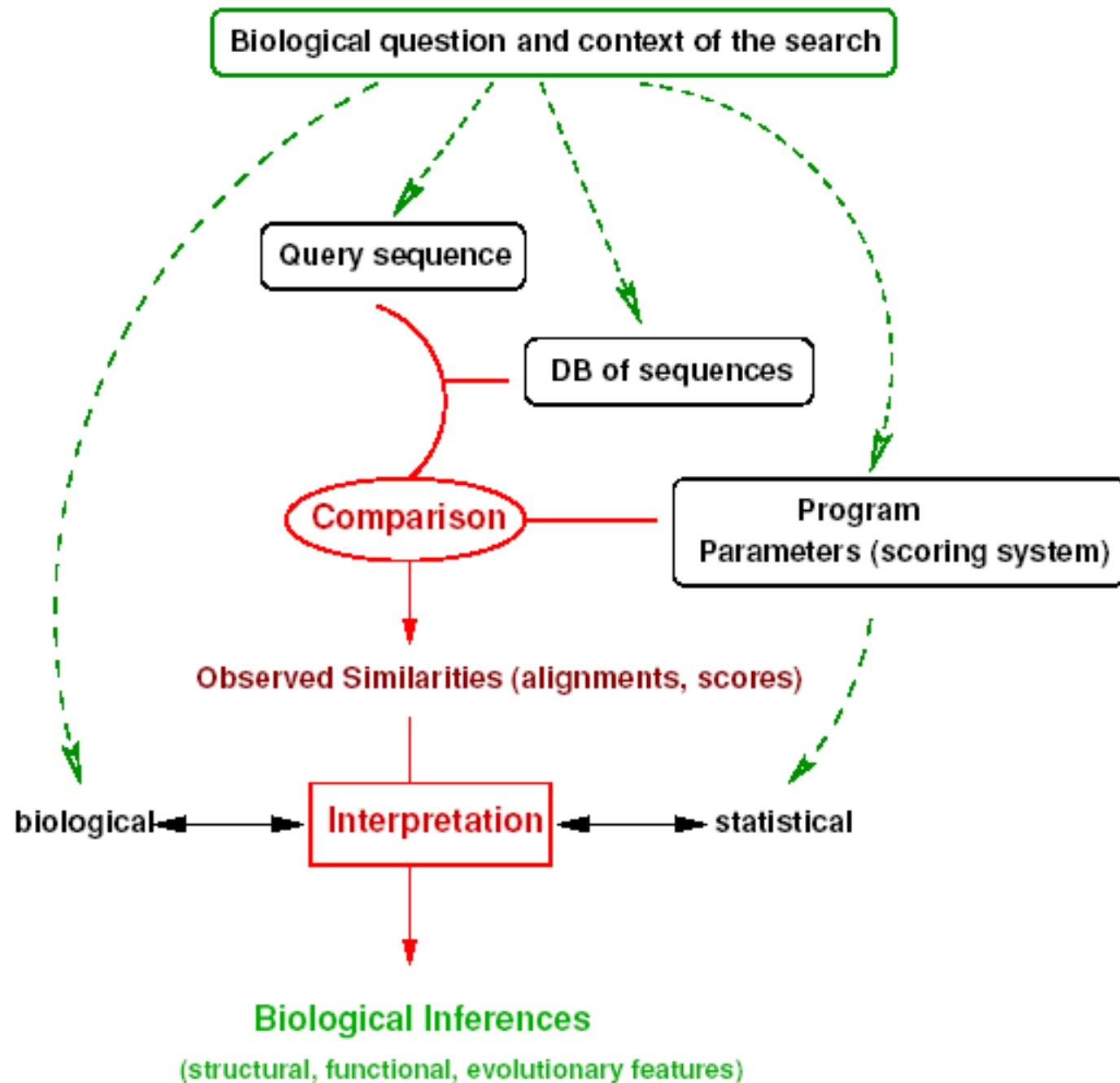
“The massive quantities of data generated by genomic research ... We are swimming in a rapidly rising sea of data...how do we keep from drowning?”

Roos DS. Bioinformatics----Trying to swin in a sea of data. Science, 2001,291(5507):1260-1261

Doolittle RF, et al. Simian sarcoma virus onc gene, v-sis, is derived from the gene (or genes) encoding a platelet-derived growth factor. *Science*. 1983 Jul 15;221(4607):275-7

simian sarcoma virus (SSV) oncogene, v-sis

- **platelet-derived growth factor (PDG)**
- **31AA (26 similar) / 39AA (35 similar)**
- **It appeared likely that the retrovirus had acquired the gene from the host cell as some kind of genetic exchange event and then had produced a mutant form of the protein that could compromise the function of the normal protein when the virus infected another animal**



Uses of Similarity Searches

- Find orthologous proteins sequences in different organisms (ortho = same function)
- Find paralogous sequences in same organism (para=originate from gene duplication)
- Align genomic sequences and expressed gene sequences (cDNAs, ESTs) to locate gene and ORFs in gene.

Difficulties in Similarity Searching and Solutions

- Slowness of dynamic programming alignment algorithm - **use a word search, heuristic method (nearly always works but is not a guaranteed algorithm).**
- Low complexity regions in sequences give false high scores - **filter out these regions in the query sequence**
- How good is the match? - **Produce an optimal local alignment**
- How significant is the alignment score? - **Provide an expect score (E value) for finding a match with an unrelated sequence in searching a sequence database of the same size**

3.3.2 Word or *k*-tuple methods

- Used by the FASTA and BLAST programs
- They align two sequences very quickly, by first searching for identical short stretches of sequences (called words or *k*-tuples) and by then joining these words into an alignment by the dynamic programming method.
- They are heuristic methods, reliable in a statistical sense, and usually provide a reliable alignment.
- Heuristic methods: an empirical method of computer programming in which rules of thumb are used to find solutions and feedback is used to improve performance.

Heuristic method for sequence alignment is used to try to find a reasonable alignment. An example of a heuristic alignment method is starting with a search for similar sequence patterns and then building a longer alignment based on these patterns.

Rather than comparing individual residues in the two sequences, FASTA and BLAST search for matching sequence patterns or words (Wilbur and Lipman, 1983). The programs then attempts to build a local alignment based on these word matches.

Wilbur WJ and Lipman DJ. 1983. Rapid similarity searches of nucleic acid and protein data banks. PNAS, 80:726-730

Methods of Similarity Searching

- FASTA (fast alignment) and BLAST (basic local alignment search tool) algorithms
- PSI-BLAST (position-specific-iterated) and PHI-BLAST(pattern-hit-iterated)
- MAST and MetaMEME on MEME server: search with PSSMs or HMMs representing domains

FASTA & BLAST story

1985 : FASTP (D. Lipman and W. Pearson)

Global gapped alignments

1988 : FASTA (W. Pearson and D. Lipman)

Local gapped alignments

1990 : BLAST1

(S. Altschul, W. Gish, W. Miller, E. Myers, and D. Lipman)

Local ungapped alignments

Gapped BLASTs :

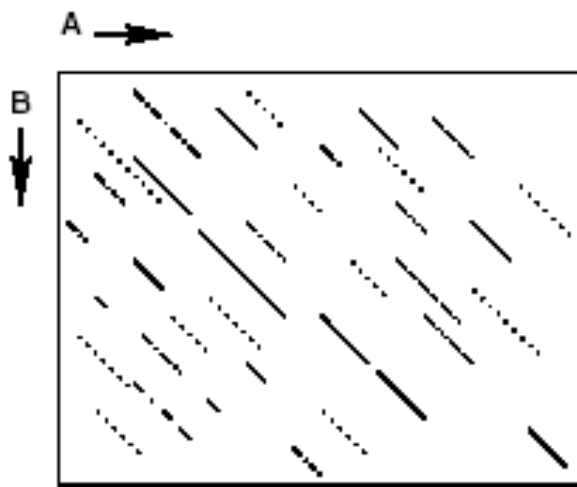
1996: WU-BLAST2 (W. Gish)

1997: NCBI-BLAST2 (and PSI-BLAST)

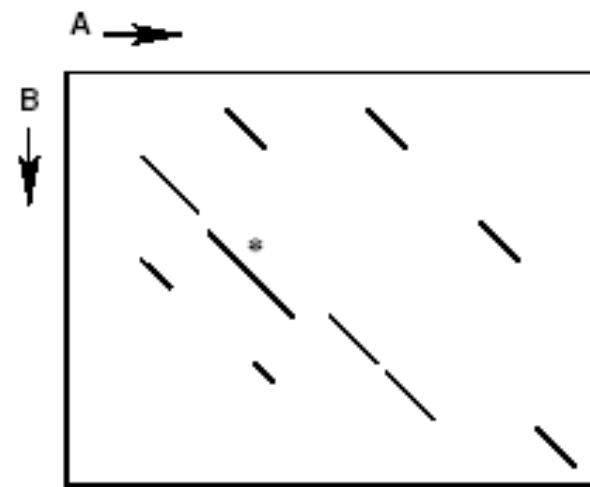
**(S. Altschul, T. Madden, A. Schaffer, J. Zhang, Z. Zhang,
W. Miller and D. Lipman)**

FASTA Algorithm

k -tuple 1-2
for amino
acid and 4-6
for
nucleotide

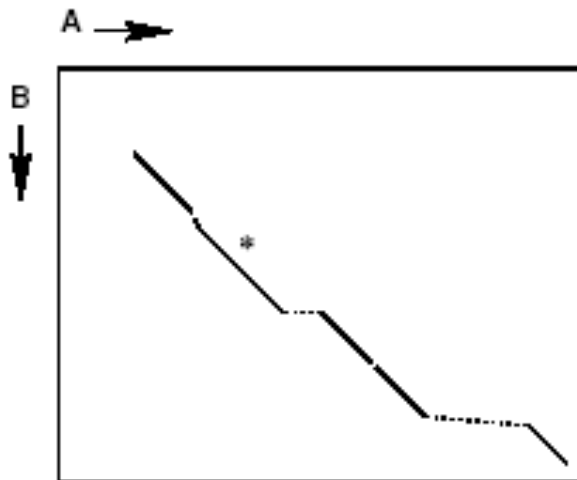


Identify all k -tuple matches



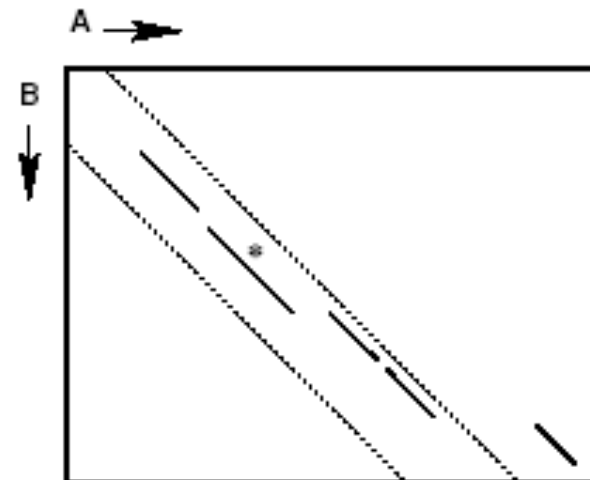
score the 10 best scoring regions
using a scoring matrix

→ Init1 score



Apply joining procedure

→ Initn score



Apply limited DP

→ Opt score

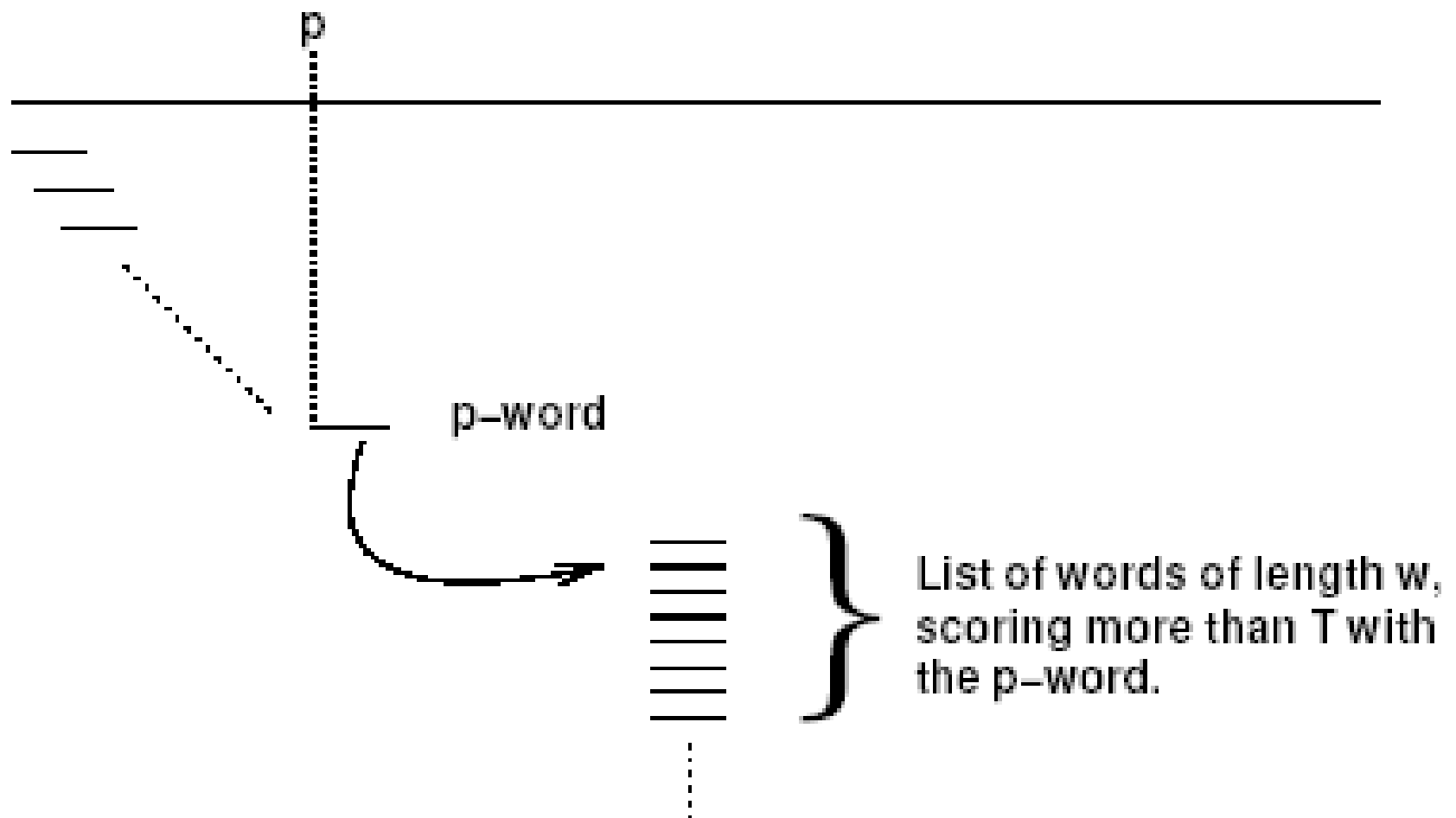
BLAST

- BLAST(Basic Local Alignment Search Tool) (Altschul et al 1990), a heuristic method designed to detect alignments that optimize a measure of local similarity known as the maximal segment pair (MSP) score.
- Word length: 11 (DNA sequences) and 3 (protein)
- Word index
- High-scoring segment pairs (HSP)

BLAST1 algorithm

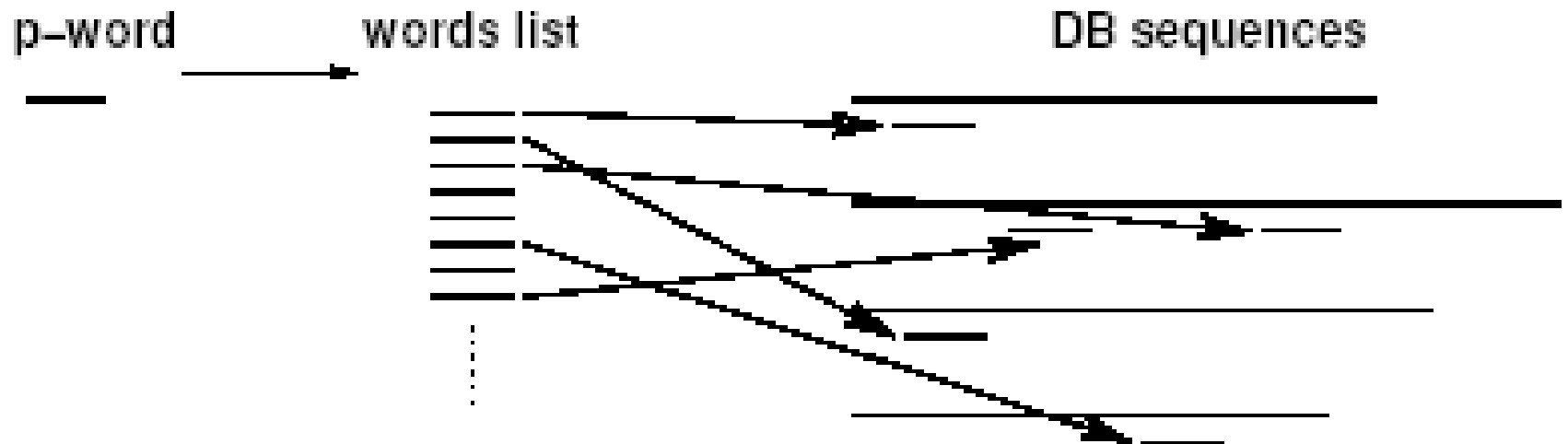
First step:

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



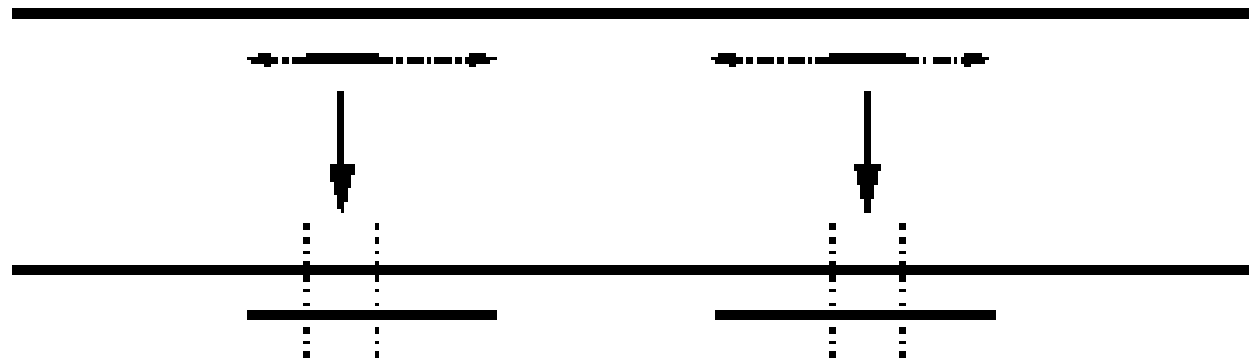
Second step:

For each words list, identify all exact matches with DB sequences:



Third step:

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 .



HSP = High Scoring Segment Pair

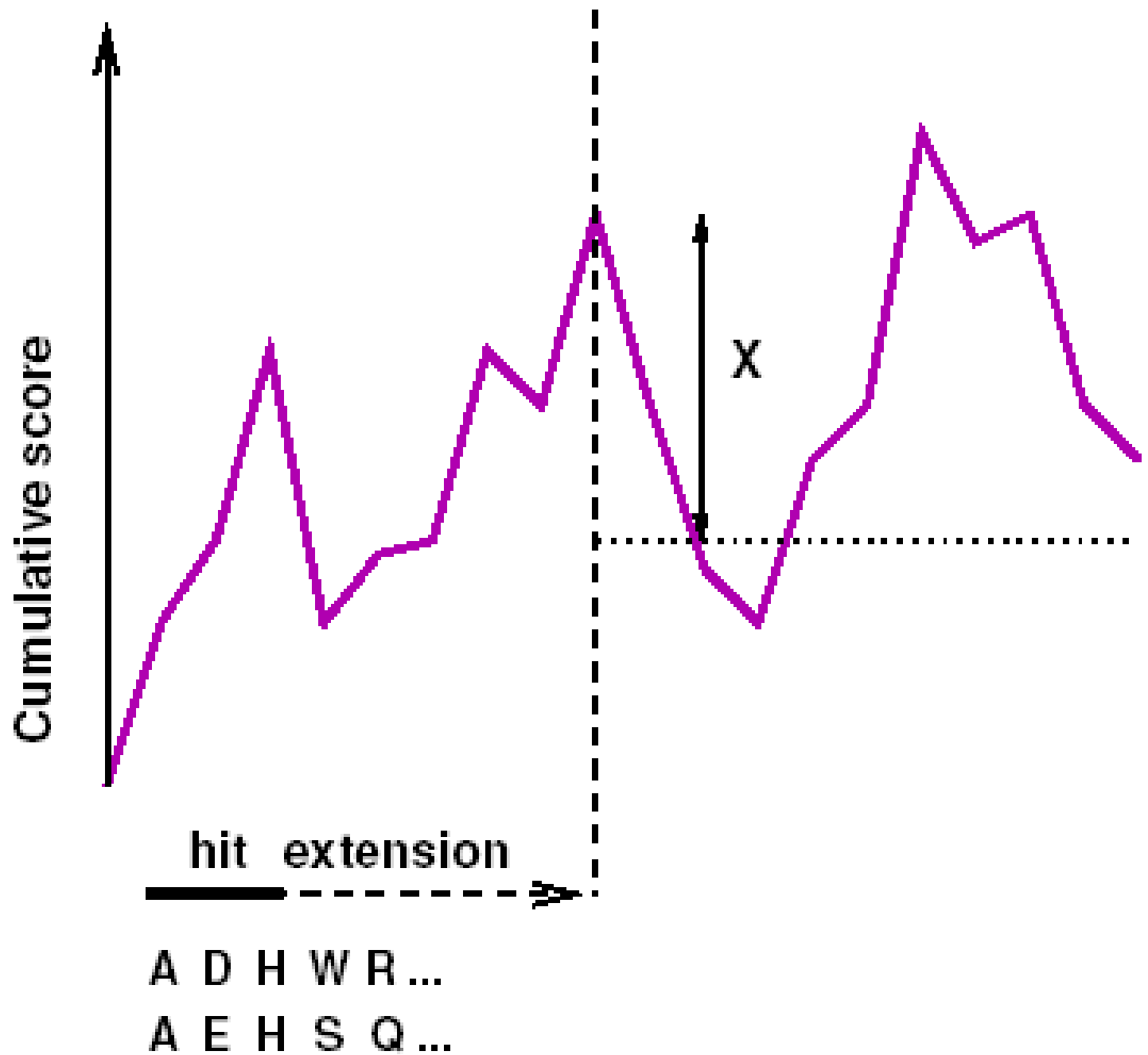
MSP = Maximal Segment Pair

Reports all HSPs having score S above a threshold, or equivalently, having E -value below a threshold.

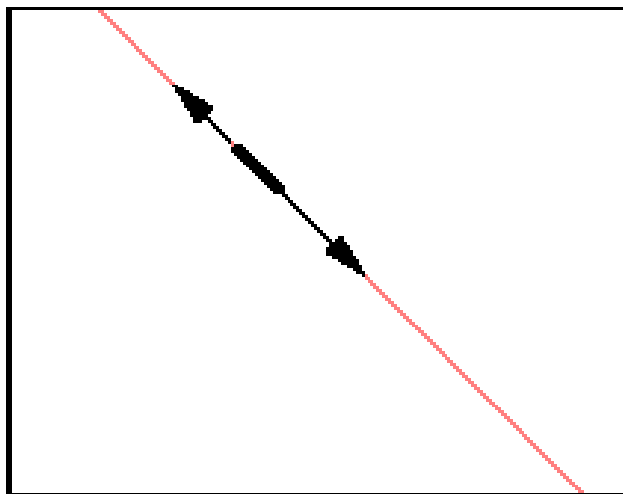
E -value = the number of HSPs having score S (or higher) expected to occur only by chance.

Apply sum-statistics to evaluate the significance of a combination of HSPs involving the same DB sequence.

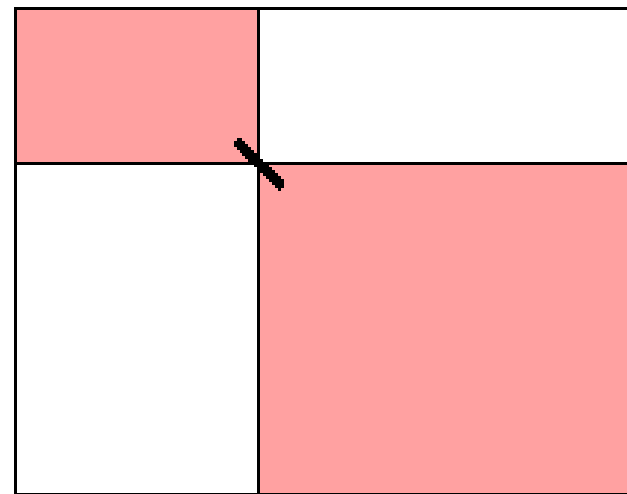
Ungapped extension of hits



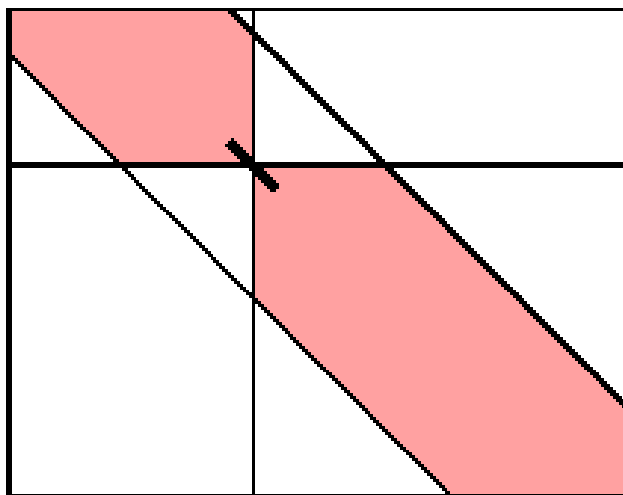
Ungapped and gapped extensions



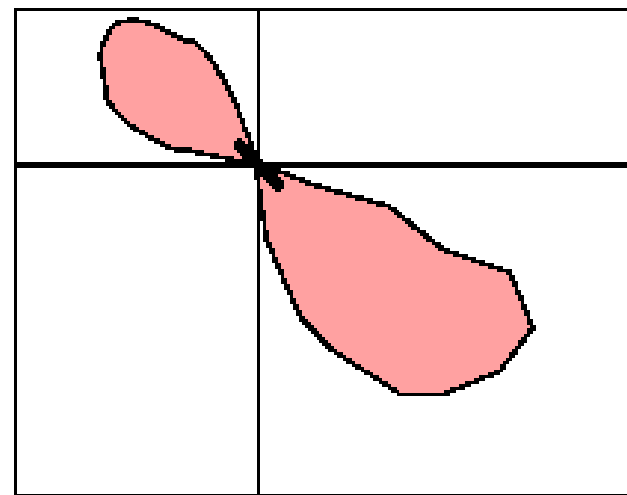
Ungapped extension



Gapped extension by full DP



Gapped extension by «banded DP»



Gapped extension by «score-limited DP»

BLAST*all*

PSI-BLAST
PHI-BLAST
rpsBLAST
.....

megaBLAST
.....

Query Sequence

Amino acid Sequence

DNA Sequence

BLASTp

tBLASTn

BLASTn

BLASTx

tBLASTx

Protein Database

Nucleotide Database

Translated

Translated

Web-based BLAST Input

NCBI Blast - Microsoft Internet Explorer

文件(F) 编辑(E) 查看(V) 收藏(A) 工具(T) 帮助(H)

后退 前进 停止 刷新 主页 搜索 收藏 历史 邮件 打印 编辑

地址(O) http://www.ncbi.nlm.nih.gov/blast/Blast.cgi?CMD=Web&LAYOUT=TwoWindows&AUTO_FORMAT=Semi&ALIGNMENTS=50&ALIGNMENT_VIEW=Pairwise&CLIENT=web&DATA 转到

Customize Search Bookmarks My Yahoo! Yahoo! Games Yahoo! Mail Shopping News

NCBI nucleotide-nucleotide **BLAST**

Nucleotide Protein Translations Retrieve results for an RID

[Search](#)

[Set subsequence](#) From: To:

[Choose database](#) nr

Now: **BLAST!** or **Reset query** **Reset all**

Options for advanced blasting

[Limit by entrez query](#) or select from: (none)

[Choose filter](#) ☒ Low complexity ☐ Human repeats ☐ Mask for lookup table only ☐ Mask lower

[Expect](#) 10

batchseq[1] - 记事本

文件(F) 编辑(E) 格式(O) 帮助(H)

```
gb|AY137580.1| Homo sapiens cell division cycle protein 25A splice variant 2 (CDC25A) mRNA, complete cds; alternatively spliced
ATGGAACCTGGGCGGAGCCCGCCGACCGCCGCGCGCTGCTCTTCGCCTGCAGCCCCCTCCCGGCTCGC
AGCCCGTCGTGAAGGCGCTATTTGGCGCTTCAGCCGCGCGGGGCACTGTCGCTGTCCACCAACCTGACCGT
CACTATGACCAGCTGCAGGCTCGGGCACTGATTATGAGCAACCACTGGAGGTGAGAACCAACAGTAAT
CTGAGAGAAATGGGCTCCTCCGAGTCAACAGATTCAAGTTTCTGTCTAGATTCTCCTGGGCCATTGGACA
GTAAGAGAAACCTTGAAATCCTATGAGAGAAATACATTCCCTACCTCAGAGCTGTTGGGATGATGCTCC
AGCTCTGAGAGAGGAGCCATTCTGATTCTCTTGACCATGACATCTTTCAGCTCATCGCCAGATGAGAAC
AAGCAAAATCTTTCCTCAATGAAGAGATAGCAGTGAACCAAGGAATTCATTCCTCTTTTACACCCCT
AGTCACCTGTGACAGCCACTTTGTCTGATGAGGATGATGGCTTCGTGGACCTCTCGATGGAGAGATCT
GAAGAAATGAGGAGAGACCCCTCGTGATGGCAAGCCTCGGACAGCTCCTCTCGTATGAGAACTACA
AACCTTGACAGCCGATGCAAGCTGTTGACTCCCTTCCCTGTGAGCTCCAGCACTCGGTCACTGTTGA
AGAGACCAAGAGATCTCAAGAGGAGTCTCCACCTGGAAGTACAAGAGAGAGAGAGAGCTGCTGGGGC
CAGCCCCAAGAGTCAACTAATCCAGAGAGGCCCATGAGACTCTTCATCAGCTTTATCCCTGGCATCT
TCCCCAAGAGAACCAATTGAGAACATTTGGACAATGACCAAGGGACCTTATAGGAGACTTCTCAAGG
GTTATCTCTTTATACAGATTGCTGGGAACATCAGGATTAAATATACATCTCCAGAAATTTGGCATC
TGTTTTGAATGGCAAGTTTGCAACCTCATTAAAGAGTTTGTATCATCGACTGTCGATACCCATATGAA
TACAGGGAGAGCCCATCAAGGCTCAAGTGAACCTTGACATGCAAGAGAGAGGTTGAGAGACTTCTATTGA
AGAGAGCCATTGTACCTACTGATGGCAAGCGTGTCTATTGTTGTTTCACTGCGAGTTTCTTCTGAGAG
AGTCCCCGATGTCCGGTATGTGAGAGAGAGAGATCGCCTGGGTAATGAATACCCAACTCCACTAC
CTGAGCTGTATGCTGAGGGGGGATACAGGAGTTCTTTATGAATGCCAGTCTTACTGTGAGCCCCC
CTAGCTACCGCCCATGCACCACGAGGACTTTAAGAGAGACTGAAGAGTTCCGACCAAGAGCGCGAC
CTGGGACGGGAGAGAGCAAGAGGAGATGTACAGTCTGTCGAAGAGCTCTGA
```

Web-based BLAST Output

i. Overview of Query

BLASTN 2.2.4 [Aug-26-2002]

Reference:

Altschul, Stephen F., Thomas L. Madden, Alejandro A. Schäffer, Jinghui Zhang, Zheng Zhang, Webb Miller, and David J. Lipman (1997), "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs", Nucleic Acids Res. 25:3389-3402.

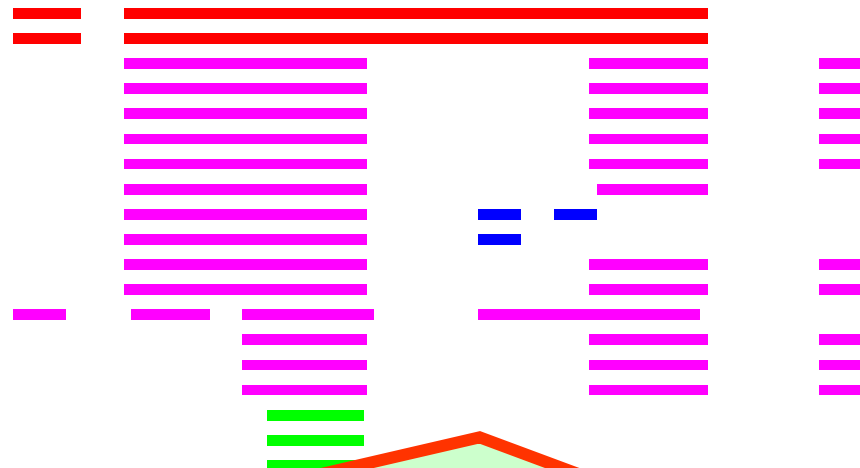
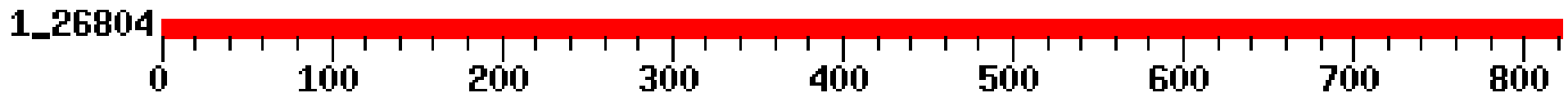
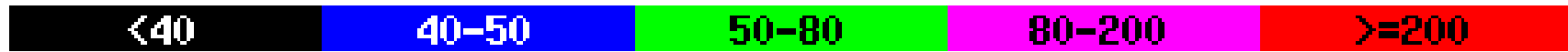
RID: 1031981000-015047-25897

Query= cluster8190_1
(1447 letters)

Database: All GenBank+EMBL+DDBJ+PDB sequences (but no EST, STS, GSS, or phase 0, 1 or 2 HTGS sequences)
1,367,736 sequences: 6,442,224,816 total letters

Web-based BLAST Output

Color Key for Alignment Scores



ii. Graphical overview of database matches

Web-based BLAST Output

iii. Descriptions of each hit (alignment)

Sequences producing significant alignments.			Score (bits)	E Value
gi 21314556 gb AF513016.1 	Sus scrofa myosin regulatory lig...	654	0.0	
gi 21261723 emb AJ487671.1 SSC487671	Sus scrofa mRNA for my...	654	0.0	
gi 17226389 gb AF440218.1 AF440218	Canis familiaris ventric...	521	e-145	
gi 22061915 ref XM_027060.4 	Homo sapiens myosin, light pol...	420	e-114	
gi 21411328 gb BC031006.1 	Homo sapiens, myosin, light poly...	420	e-114	
gi 21410232 gb BC031008.1 	Homo sapiens, myosin, light poly...	420	e-114	
gi 16198354 gb BC015821.1 BC015821	Homo sapiens, myosin, li...	420	e-114	
gi 516580 gb S69022.1 S69022	Homo sapiens myosin light chai...	420	e-114	
gi 4557774 ref NM_000432.1 	Homo sapiens myosin, light poly...	404	e-109	
gi 34845 emb X66141.1 HSMYLC2	H.sapiens mRNA for cardiac ve...	404	e-109	
gi 2460246 gb AF020768.1 AF020768	Homo sapiens cardiac vent...	400	e-108	
gi 1220300 gb M22815.1 HUMCMCLC	Human (clone PWHCLC2-8) card...	385	e-103	
gi 34686 emb X14332.1 HSMLC2	Human mRNA for ventricular myo...	371	2e-99	
gi 56682 emb X07314.1 RNMLC2R	Rat heart myosin light chain ...	264	3e-67	
gi 20841890 ref XM_132324.1 	Mus musculus myosin light chai...	262	1e-66	
gi 6754777 ref NM_010861.1 	Mus musculus myosin light chain...	262	1e-66	
gi 12832295 dbj AK002367.1 	Mus musculus adult male kidney ...	262	1e-66	
gi 199984 gb W91602.1 MUSMYLTC	M. musculus myosin light cha...	262	1e-66	

Web-based BLAST Output

```
☐ >gi|21314556|gb|AF513016.1 Sus scrofa myosin regulatory light chain ventricular isoform  
(MLC-2V) mRNA, complete cds  
Length = 672
```

```
Score = 654 bits (330), Expect = 0.0  
Identities = 359/370 (97%)  
Strand = Plus / Plus
```

```
Query: 95 tccaccatgtcacctaagaaagccaagaagagagcagatggagccaaagnnncgtgttc 154  
|||||  
Sbjct: 1 tccaccatgtcacctaagaaagccaagaagagagcagatggagccaattccaacgtgttc 60
```

```
Query: 155 tccatgtttgaacagaccagattcaggaatttaaggaggccttcaccatcatggaccag 214  
|||||  
Sbjct: 61 tccatgtttgaacagaccagattcaggaatttaaggaggccttcaccatcatggaccag 120
```

```
Query: 215 aacagggatggcttcatagacaagaatgatctgaggacacctttgctgctcttgggcgt 274  
|||||  
Sbjct: 121 aacagggatggcttcatagacaagaatgatctgaggacacctttgctgctcttgggcgt 180
```

```
Query: 275 gtgaacgtgaaaaatgaggaaattgatgaaatgatcaaggaagctccaggtccattaac 334  
|||||  
Sbjct: 181 gtgaacgtgaaaaatgaggaaattgatgaaatgatcaaggaagctccaggtccattaac 240
```

```
Query: 335 tttactgtgttcctaacgatgtttggggagaaacttaagggggcagaccctg 394  
|||||  
Sbjct: 241 tttactgtgttcctaacgatgtttggggagaaacttaagggggcagaccctg 300
```

```
Query: 395 gcccttgacgcgttccaagtgtttgaccctgaaggcaaagggtgctcagggtgattat 454  
|||||  
Sbjct: 301 atccttaacgcgttccaagtgtttgaccctgaaggcaaagggtgctcagggtgattat 360
```

iv. Pairwise
alignments

Web-based BLAST Output

Database: All GenBank+EMBL+DDBJ+PDB sequences (but no EST, STS, GSS, or phase 0, 1 or 2 HTGS sequences)

Posted date: Sep 4, 2002 12:27 AM

Number of letters in database: 2,147,257,520

Number of sequences in database: 1,367,736

Lambda	K	H
1.37	0.711	1.31

Gapped

Lambda	K	H
1.37	0.711	1.31

Matrix: blastn matrix:1 -3

Gap Penalties: Existence: 5, Extension: 2

Number of Hits to DB: 3,320,183

Number of Sequences: 1367736

Number of extensions: 3320183

Number of successful extensions: 18698

Number of sequences better than 10.0: 207

length of query: 1447

length of database: 6,442,224,816

effective HSP length: 22

effective length of query: 1425

effective length of database: 6,412,134,624

effective search space: 9137291839200

effective search space used: 9137291839200

T: 0

A: 30

X1: 6 (11.9 bits)

X2: 15 (29.7 bits)

S1: 12 (24.3 bits)

S2: 20 (40.1 bits)

v. Statistical details
of the search

Local BLAST

On Windows PC:

- (1) Download from NCBI and install
- (2) Configuration: using the ncbi.ini file

[NCBI]

Data="C:\path\data\“

- (3) Create or format your databases

Difference between FASTA and BLAST

- Whereas FASTA searches for all possible words of the same length, BLAST confines the search to the words that are the most significant.
- For the BLAST algorithm, the word length is fixed at 3 (formerly 4) for proteins and 11 for nucleic acids (3 if the sequences are translated in all six reading frames). This length is the minimum needed to achieve a word score that is high enough to be significant but not so long as to miss short but significant patterns.
- FASTA theoretically provides a more sensitive search of DNA sequence databases because a shorter word length may be used.

Beyond sequences: database searching using scoring matrices, profiles and PSSMs

Previously used a single sequence to search a sequence database to locate similar sequences

Now interested in searching a database with a representation of a local multiple sequence alignment – a PSSM (no gaps), profile (gaps) or HMM (gaps)

Can also search a sequence for matches to a database of models in the form of PSSMs, profiles or HMMs (e.g. the PFAM database)

- PSSM: position-specific scoring matrix; HMM: Hidden Markov Models

How do PSSMs etc. find additional matches than searching with single sequences?

If search with a set of single query sequences, then matches must align with query with Blosum62 substitutions at each position

If using a PSSM, profile of HMM then can also match combinations of amino acids

Blast search

Query ..GYD..

Database ..GYD..

sequences ..GWD..

matched ..GWE..

(Blosum62
substitutions)

PSSM or HMM search

Query ..GYD..

Database ..GYD..

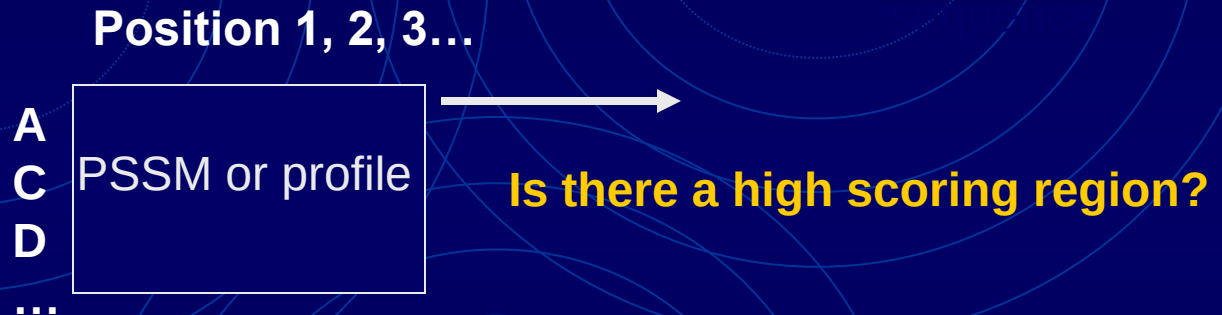
sequences ..SYE..

matched ..SWE..

(PSSM or HMM
substitutions)

Searching a database with a PSSM or HMM

- Recall that the PSSM is a matrix of scores with the columns representing positions in a local multiple sequence alignment and the rows the 20 amino acids

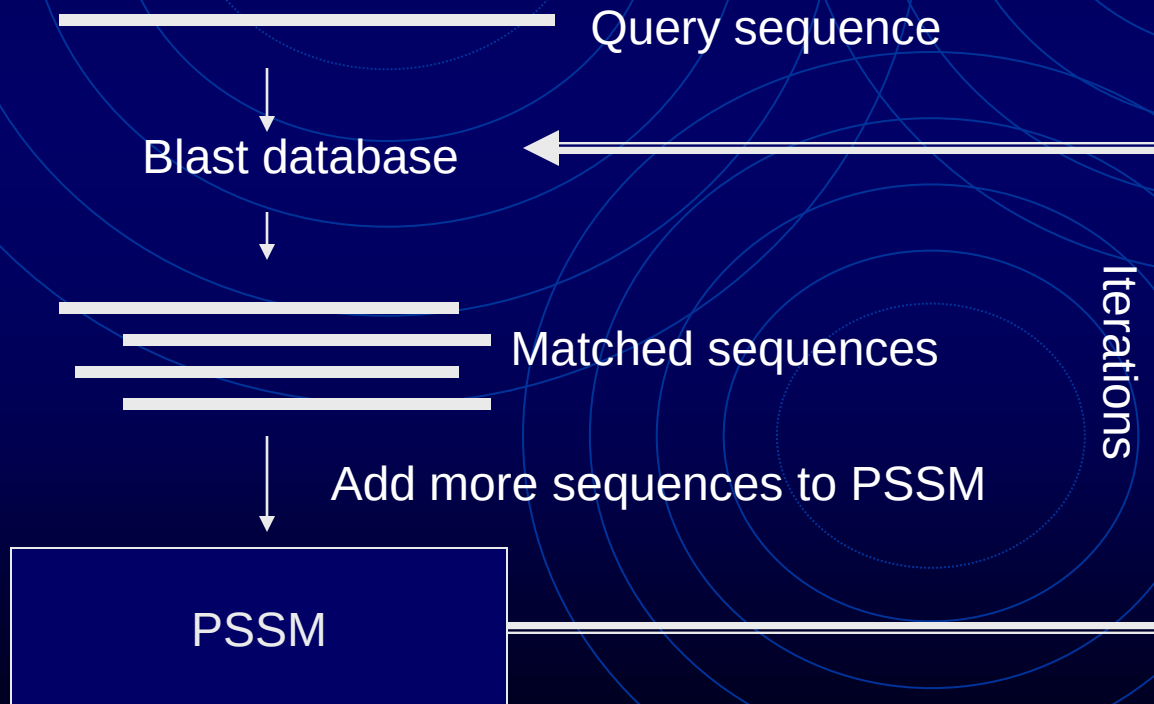


Similarly, an HMM is a probabilistic model of a multiple sequence alignment.



Other tools: PSI-BLAST

- **PSI-Blast is position-specific-iterated-blast**
Generates a PSSM from the matched sequences in a BLAST search – it uses the entire sequence



Caution!
Can switch
domains

PHI-BLAST: regular expressions

Regular expressions provide a wide range of pattern finding options e.g. [LIVMF]-G-E-x (5,11)-[IV] implemented in the prosite catalog and used in phi-blast

Prosite entry for ATP/GTP binding site motif A (P loop) is [AG] - x(4)-G-K-[ST]

In phi-blast, a regular expression is provided – database is searched for sequences with pattern, and then psi-blast used to expand search

3.3.4 Assessing the significance of sequence alignments

Global alignments:

the distribution is not known

Local alignments without gaps:

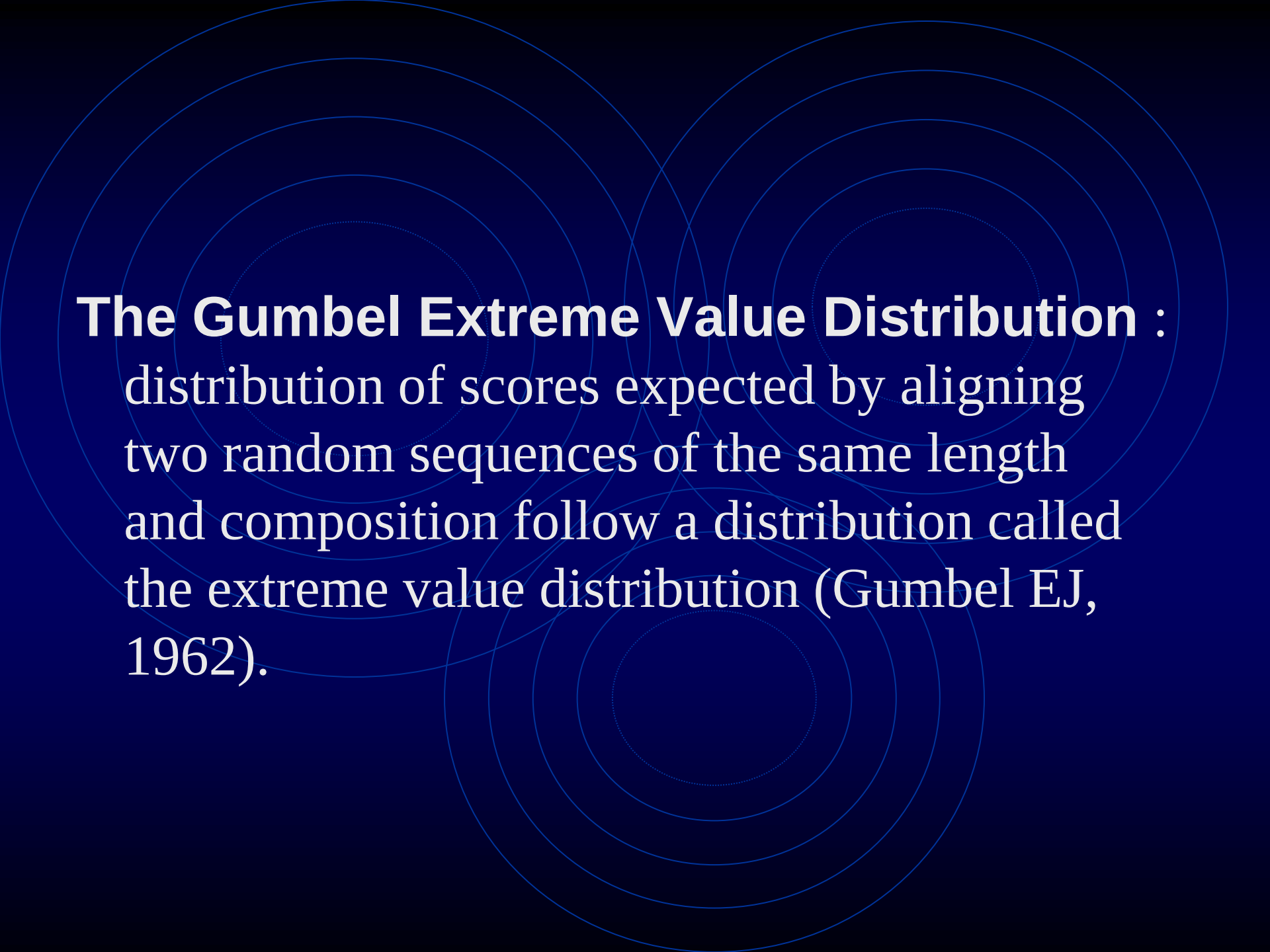
theoretical work: Karlin–Altschul statistics

→ Extreme-value distribution

Local alignments with gaps:

empirical studies

→ Extreme-value distribution

The background of the slide features a dark blue field with several sets of concentric circles in a lighter blue color. These circles are centered at various points, creating a pattern of overlapping rings that fills the entire background.

The Gumbel Extreme Value Distribution :
distribution of scores expected by aligning
two random sequences of the same length
and composition follow a distribution called
the extreme value distribution (Gumbel EJ,
1962).

A. Score distribution

```

z'  opt      E()
< 20 176      0:==
22  1  0:=
24  2  0:=
26  2  2:*
28  21 17:*
30 109 102:*
32 355 393:==*
34 1053 1067:==*
36 2214 2191:=====
38 4004 3620:=====
40 5285 5050:=====
42 6350 6173:=====
44 6716 6809:=====
46 6847 6935:=====
48 6661 6640:=====
50 5777 6059:=====
52 5015 5327:=====
54 4344 4550:=====
56 3772 3801:=====
58 3025 3120:=====
60 2475 2528:=====
62 1968 2026:=====
64 1607 1612:=====
66 1362 1274:=====
68 983 1002:=====
70 823 785:=====
72 597 614:=====
74 584 478:=====
76 456 372:=====
78 306 289:=====
80 238 225:=====
82 230 172:=====
84 141 136:=====
86 131 105:=====

```

92	55	49:*	=====*
94	41	38:*	=====
96	37	29:*	=====
98	20	23:*	=====*
100	20	17:*	=====
102	16	14:*	=====*
104	7	10:*	=====*
106	9	8:*	=====
108	7	6:*	=====
110	3	5:*	=====*
112	6	4:*	=====
114	2	3:*	=====*
116	4	2:*	=====

Gaussian and extreme value density functions

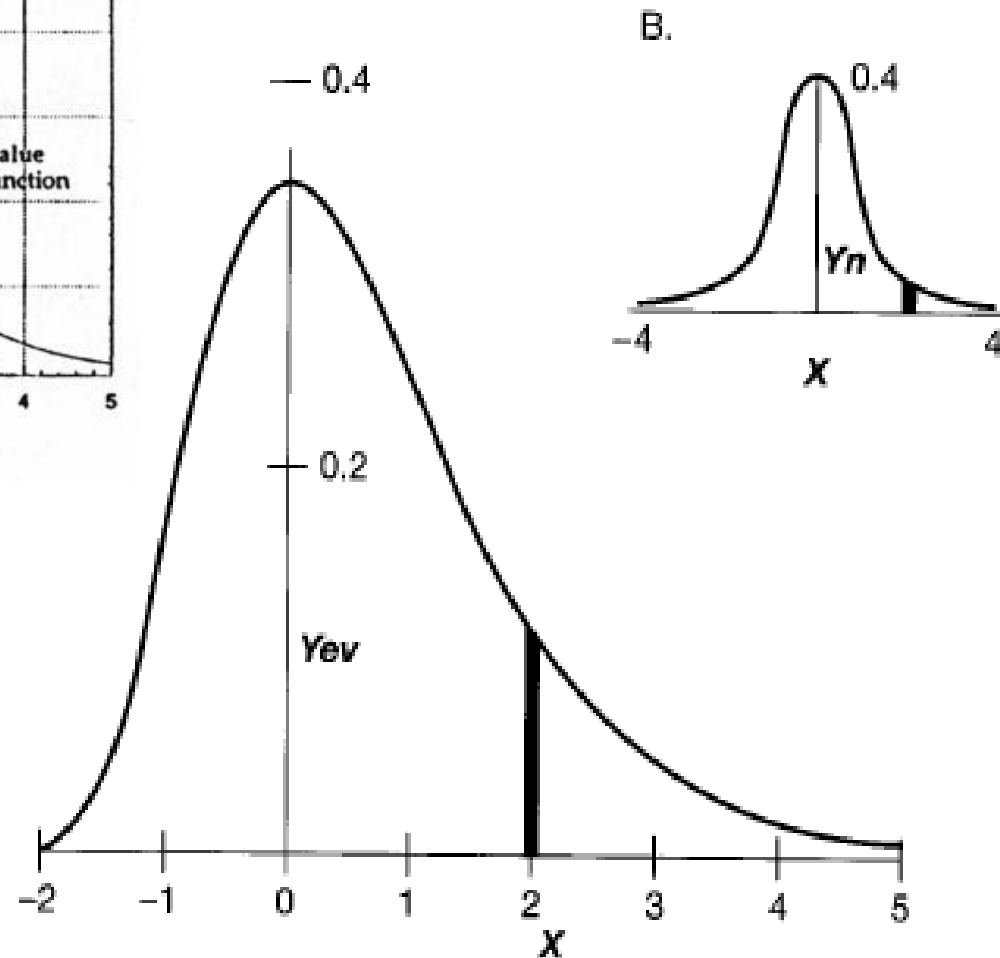
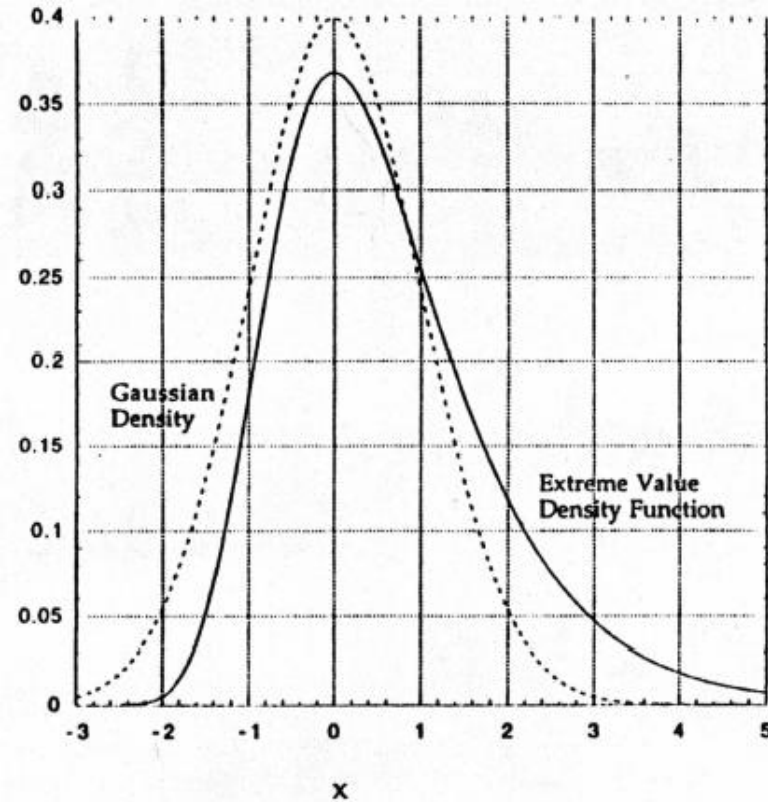


Figure 3.17. Probability values for the extreme value distribution (A) and the normal distribution (B). The area under each curve is 1.

The extreme value distribution

$$Y_{ev} = \exp[-x - e^{-x}]$$

The normal distribution

$$Y_n = 1 / \sqrt{(2\pi)} \exp[-x^2 / 2]$$

The probability that score S will be less than value x , $P(S < x)$, is obtained by calculating the area under curve from $-\infty$ to x


$$P(S < x) = \exp[-e^{-x}]$$

$$P(S \geq x) = 1 - \exp[-e^{-x}]$$

$$= 1 - \exp[-Kmne^{-\lambda x}]$$

P-value: probability that there is at least one random MSP having score S or greater

E-value: the number of HSPs having score S (or higher) expected to occur only by chance

E-value is given by the formula:

$$E = Kmne^{-\lambda x}$$



$$P(S' \geq x) = 1 - \exp -E$$

$$P(S' \geq x) \approx E$$

$$P(S \geq x) = 1 - \exp[-e^{-x}] \rightarrow e^{-x}$$

The Gumbel extreme value distribution

$$P(s \geq x) = 1 - \exp[-e^{\lambda(x-\mu)}]$$

$$\lambda = \pi(\sigma\sqrt{6})$$

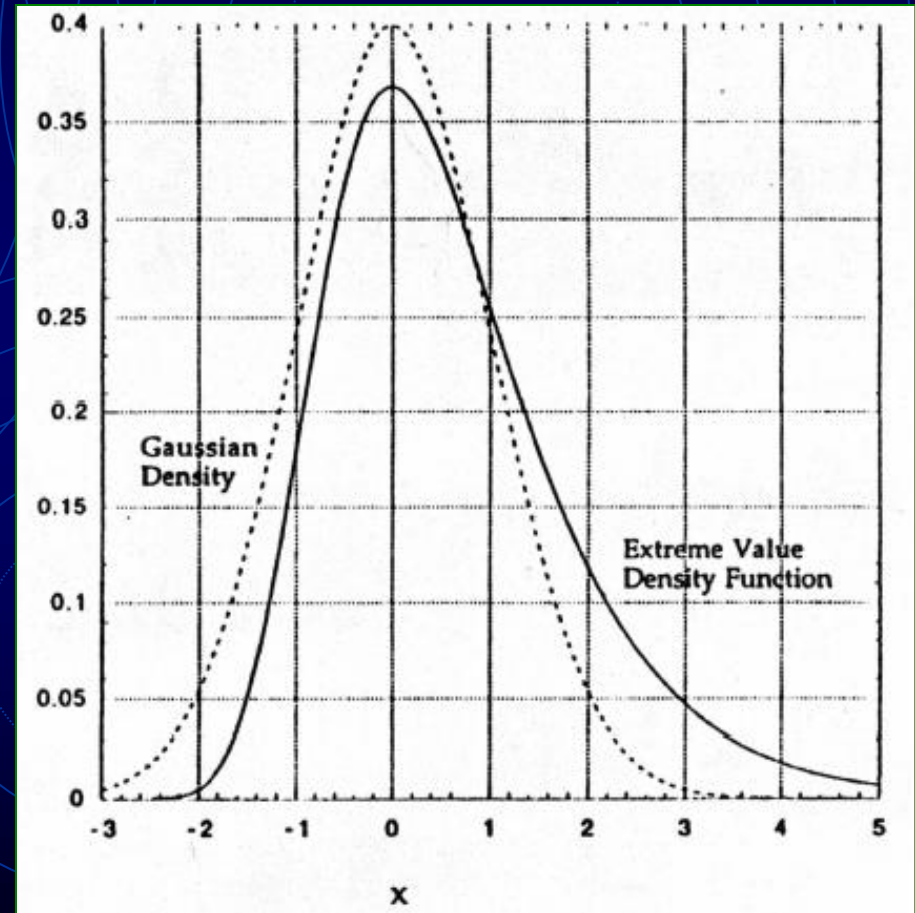
$$\mu = \underline{x} - \gamma / \pi = \underline{x} - 0.450\sigma$$

$$= (\ln Kmn) / \lambda$$

$$P(s \geq x) = 1 - e^{Kmn e^{-\lambda s}}$$

$$\approx Kmn e^{-\lambda s}$$

$$P(S \geq x) = 1 - \exp[-e^{-x}] \rightarrow e^{-x}$$



$$P(S \geq x) = 1 - \exp[-e^{-x}] \rightarrow e^{-x}$$


$$P(S \geq x) \approx e^{-x}$$

$$P(S' \geq x) \approx Kmne^{-\lambda x}$$

$$P(S' \geq x) \approx E$$

By normalizing a raw score using
the formula

$$S' = \lambda S - \ln Kmn$$

Need of empirical simulations of the random distribution in order to calculate its parameters

- **Blast2 :**

artificial random sequences

- **Fasta:**

uses results from the search: real unrelated sequences

BLASTP 2.0.5 [May-5-1998]

Query= human XP-F repair gene (905 letters)

Database: Non-redundant SwissProt sequences 74,596 sequences; 26,848,718 total letters

B.

Color Key for Alignment Scores

<40

40-50

80-200

>=200

QUERY

0 125 250 375 500 625 750 875

Distribution of 11 BLAST Hits on the Query Sequence

Sequences producing significant alignments:

Score E
(bits) Value

sp Q92889 XPF_HUMAN	DNA-REPAIR PROTEIN COMPLEMENTING XP-F CELL ...	1659	0.0
sp P36617 RA16_SCHPO	DNA REPAIR PROTEIN RAD16	485	e-136
sp P06777 RAD1_YEAST	DNA REPAIR PROTEIN RAD1	231	4e-60
sp P40562 YIS2_YEAST	PUTATIVE ATP-DEPENDENT RNA HELICASE YIR002C	37	0.17
sp Q10202 YAXB_SCHPO	PUTATIVE ATP-DEPENDENT RNA HELICASE C13F4.11C	36	0.38

Database: Non-redundant SwissProt sequences
Number of letters in database: 26,848,718
Number of sequences in database: 74,596
Lambda K H
0.320 0.136 0.394
Gapped
Lambda K H
0.270 0.0470 0.230
Matrix: BLOSUM62
Gap Penalties: Existence: 11, Extension: 1
Number of Hits to DB: 42777291
Number of Sequences: 74596
Number of extensions: 1706128
Number of successful extensions: 4638
Number of sequences better than 10.0: 12
Number of HSP's better than 10.0 without gapping: 4
Number of HSP's successfully gapped in prelim test: 8
Number of HSP's that attempted gapping in prelim test: 4616
Number of HSP's gapped (non-prelim): 16
length of query: 905
length of database: 26848718
effective HSP length: 55
effective length of query: 850
effective length of database: 22745938
effective search space: 19334047300
effective search space used: 19334047300
T: 11
A: 40
X1: 16 (7.4 bits)
X2: 38 (14.8 bits)
X3: 64 (24.9 bits)
S1: 41 (21.8 bits)
S2: 68 (30.9 bits)

FASTA (3.14 April, 1998) function (optimized, BL50 matrix) ktup: 2
join: 39, opt: 27, gap-pen: -12/ -2, width: 16 reg.-scaled

The best scores are:			initn	initl	opt	z-sc	E()
XPF_HUMAN	11/97	(905)	5893	5893	5893	6529.7	0
RA16_SCHPO	11/97	(892)	1569	519	749	827.2	2.1e-39
RAD1_YEAST	11/97	(1100)	975	362	619	681.7	2.7e-31
YIS2_YEAST	11/95	(993)	37	37	161	174.6	0.0047
YAXB_SCHPO	10/96	(578)	91	91	133	147.1	0.16