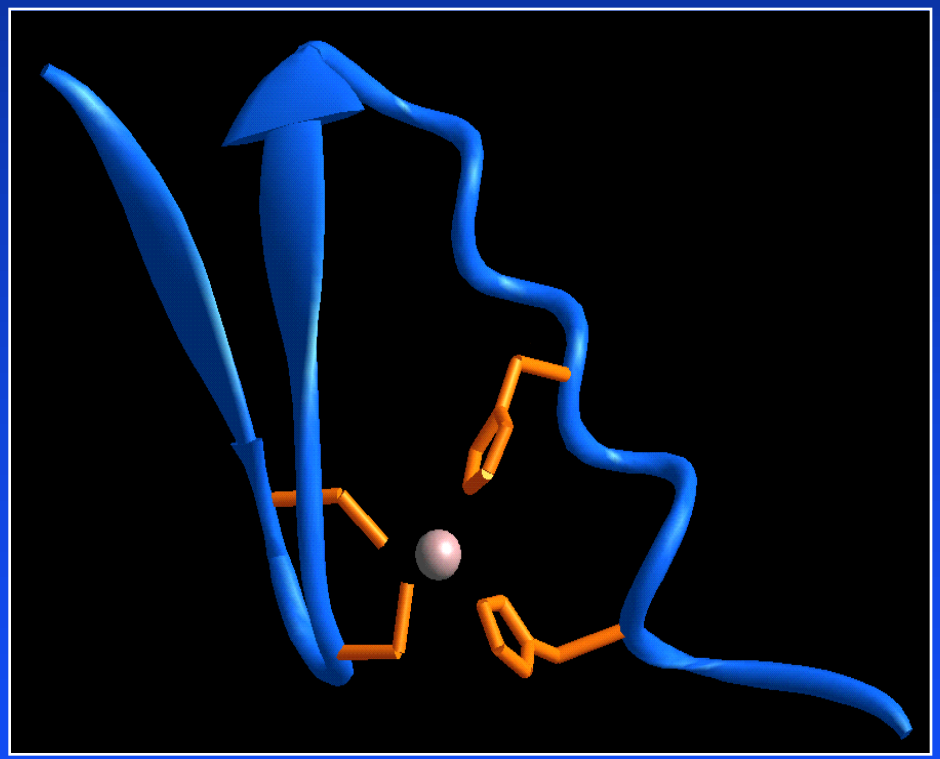
Computational Molecular Biology
Discrete Protein Sequence Motifs

Biochemistry 218/Medical Information Sciences 231

November 8, 1999



Thomas D. Wu, Craig G. Nevill-Manning
Jane Su, Jimmy Huang, Steven P. Bennett,
Douglas L. Brutlag

PROSITE Patterns

(http://www.expasy.ch/prosite/)

Active site of trypsin-like serine proteases

G D S G G

Zinc Finger (C_2H_2 type)

C x{2,4} C x{12} H x {3,5} H

Homeobox Domain Signature

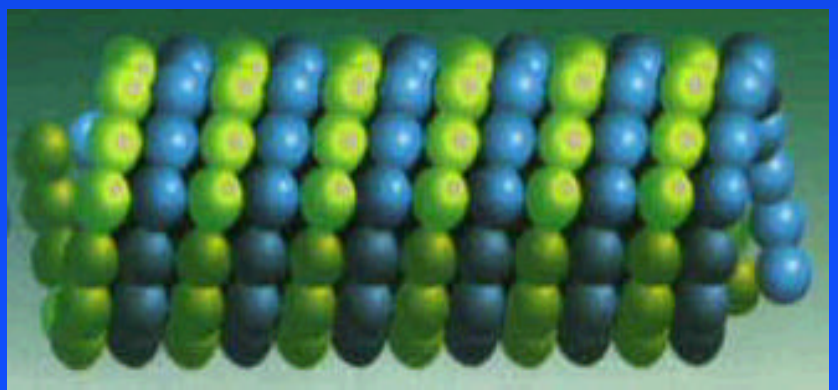
[LIVMF] x{5} [LIVM] x{4} [IV] [RKQ] xW x{8} [RK]

*e*MOTIF Advantages
(http://motif.stanford.edu/emotif/)

- Discrete motifs that represent specific functions
- Highly specific motifs for searching entire proteomes
- Maintain sensitivity with multiple motifs
- Generate motifs automatically from alignments
- Resistant to errors in sequence or alignment or protein classification
- Robust with respect to protein subclasses
- Generates structural motifs & potential drug targets
- Biological generalization from known examples

Generating Motifs from Aligned Protein Sequences

```
TEAESNMNDPVAEYQQY
TDARQDLYELEVDYANL
TEARENIAVLERDFEEV
TEAESNMNDLVSEYQQY
TEVRANMNDLVAEYQQY
SEAESNMNDLVSEYQQY
TEAREDLAALEKDYEEV
TEAREDLAALERDYIEV
SEAREDLAALEKDYEEV
AEAREDLAALEKDYIEV
SEAREDLAALEKDYEEV
SEAREDLAALERDYEEV
```



Generating Motifs from Aligned Protein Sequences

```
TEAESNMNDPVAEYQQY
TDARQDLYELEVDYANL
TEARENIAVLERDFEEV
TEAESNMNDLVSEYQQY
TEVRANMNDLVAEYQQY
SEAESNMNDLVSEYQQY
TEAREDLAALEKDYEEV
TEAREDLAALERDYIEV
SEAREDLAALEKDYEEV
AEAREDLAALEKDYIEV
SEAREDLAALEKDYEEV
SEAREDLAALERDYEEV
```

```
TEARENIAVLERDFEEV
SDVESDNNDPVAEYIQL
    A LYE  V  ANY
    Q   A  S  Q
           K
```

Generating Motifs from Aligned Protein Sequences

```
TEAESNMNDPVAEYQQY
TDARQDLYELEVDYANL
TEARENIAVLERDFEEV
TEAESNMNDLVSEYQQY
TEVRANMNDLVAEYQQY
SEAESNMNDLVSEYQQY
TEAREDLAALEKDYEEV
TEAREDLAALERDYIEV
SEAREDLAALEKDYEEV
AEAREDLAALEKDYIEV
SEAREDLAALEKDYEEV
SEAREDLAALERDYEEV
```

```
TEAREDLAALERDYEEV
S          K  I
```

Generating Motifs from Aligned Protein Sequences

```
TEAESNMNDPVAEYQQY
TDARQDLYELEVDYANL
TEARENIAVLERDFEEV
TEAESNMNDLVSEYQQY
TEVRANMNDLVAEYQQY
SEAESNMNDLVSEYQQY
TEAREDLAALEKDYEEV
TEAREDLAALERDYIEV
SEAREDLAALEKDYEEV
AEAREDLAALEKDYIEV
SEAREDLAALEKDYEEV
SEAREDLAALERDYEEV
```

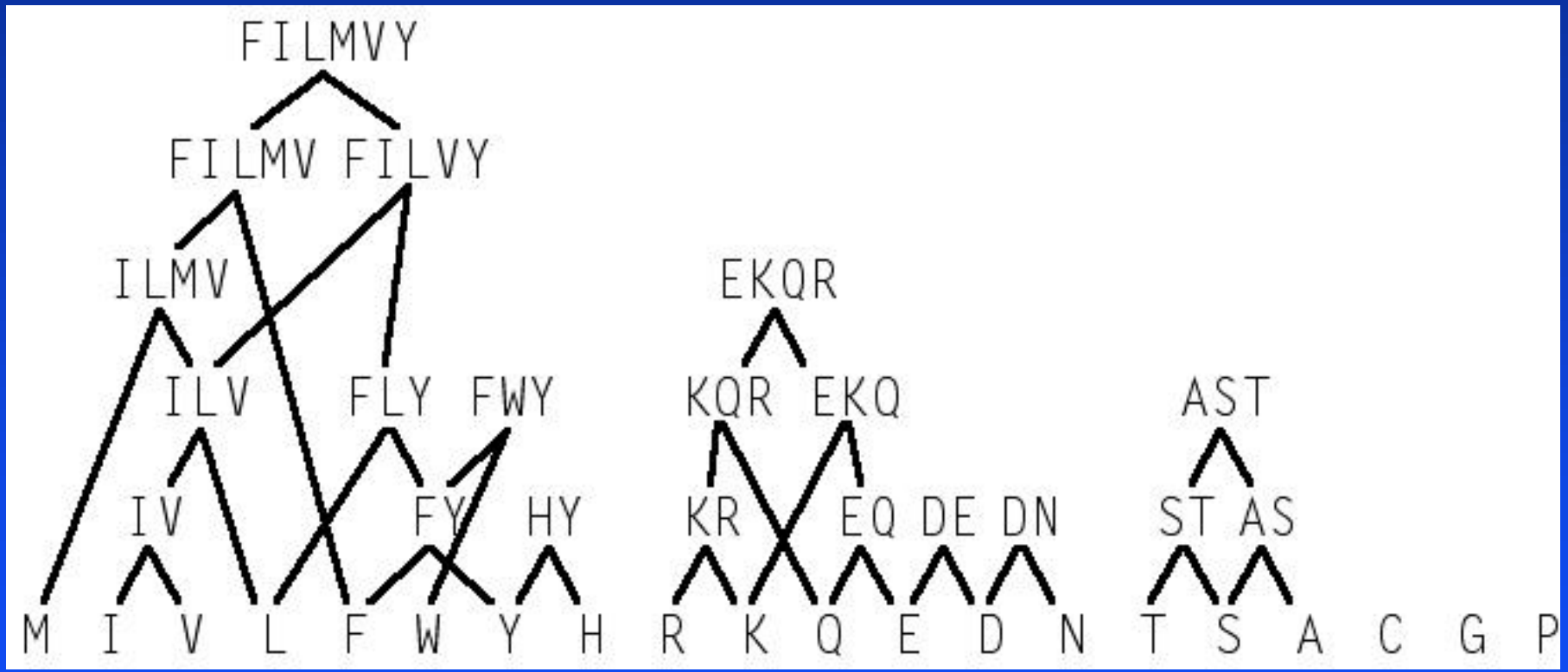
```
TEARENIAVLERDFEEV
SDVESDNNDPVAEYIQL
    A LYE  V  ANY
    Q   A  S  Q
           K
```

Amino Acid Substitution Groups Based on Physical Properties

Only allow groups of amino acids sharing some chemical or physical property

Group	Property
AG	Tiny
ST	Hydroxyl
PAGST	Small
QN	Glutamine/Glutamate
QNED	Acidic/Polar
KR	Strongly Basic
VLI	Small hydrophobic
VLIM	Small hydrophobic
FYW	Aromatic
KRH	Basic
DE	Acidic

Allowable Amino Acid Substitution Groups



Evaluating *e*MOTIFs

For each motif, calculate:

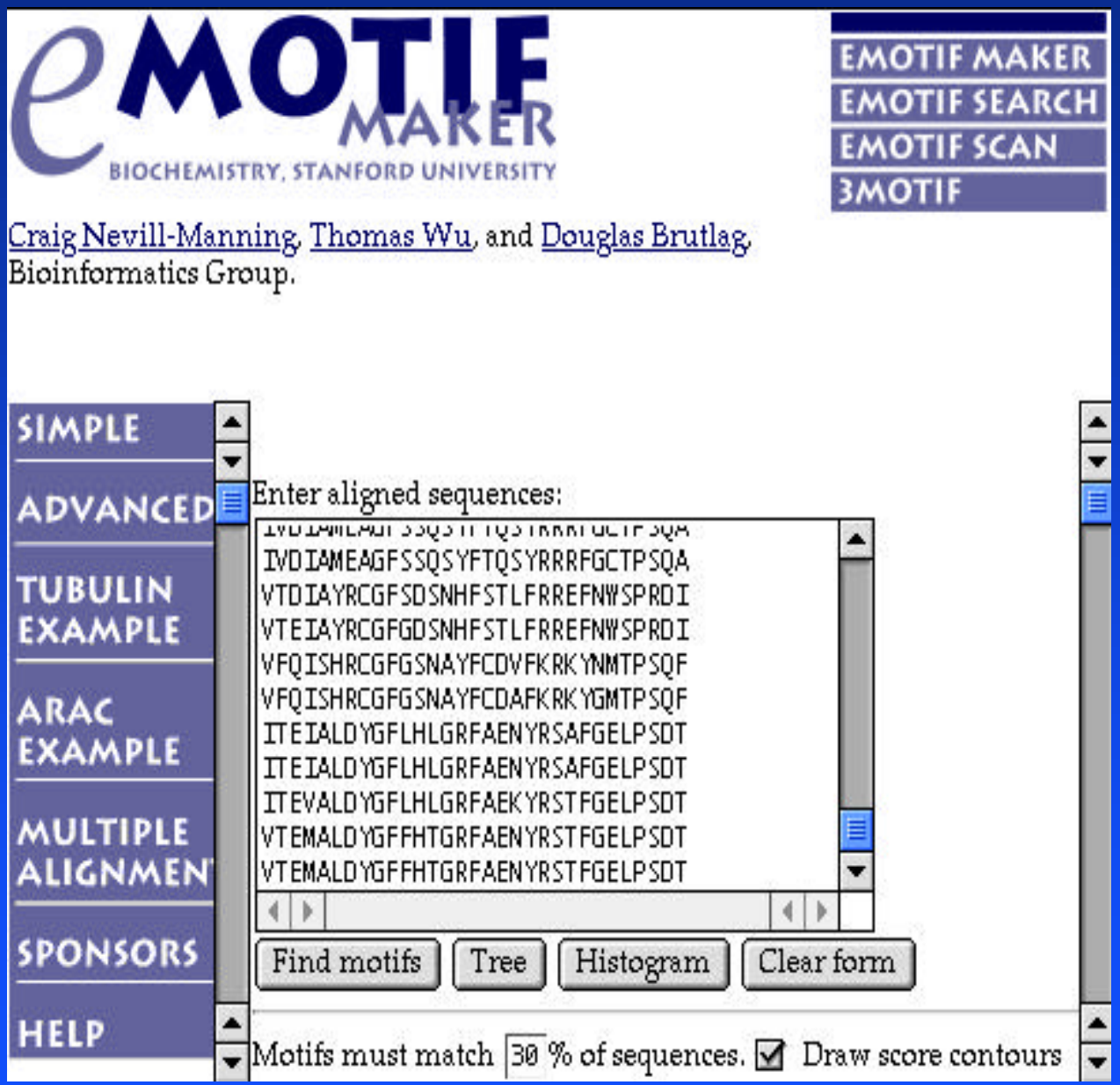
- Sensitivity (fraction of training set covered):

$$\frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}}$$

- Specificity (expected frequency of false predictions) :

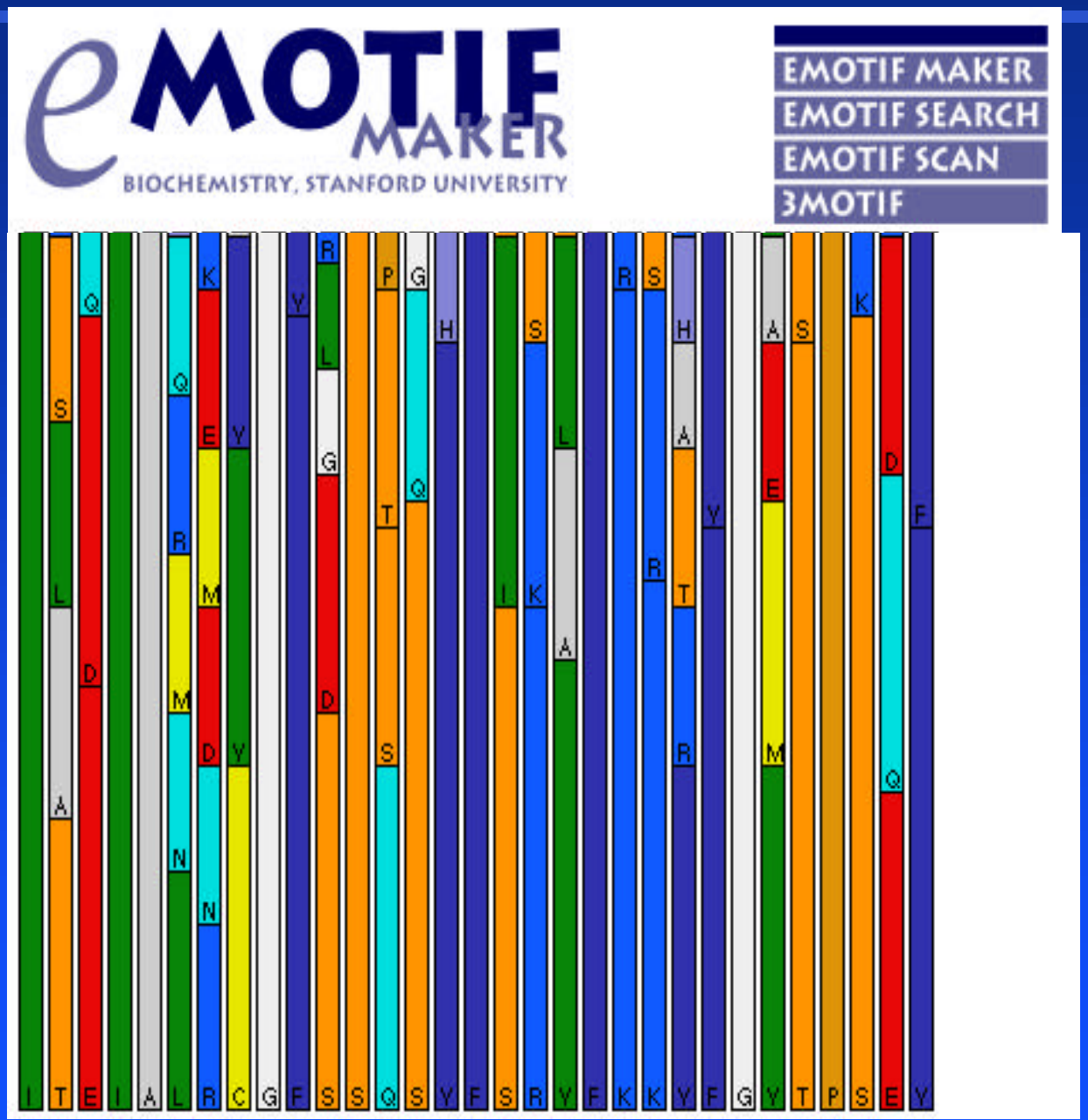
$$\prod_{\text{Positions}} \left[\sum_{\text{AA}} \text{Probability(Amino Acids in Group)} \right]$$

Finding *e*MOTIFs
(http://motif.stanford.edu/emotif/)



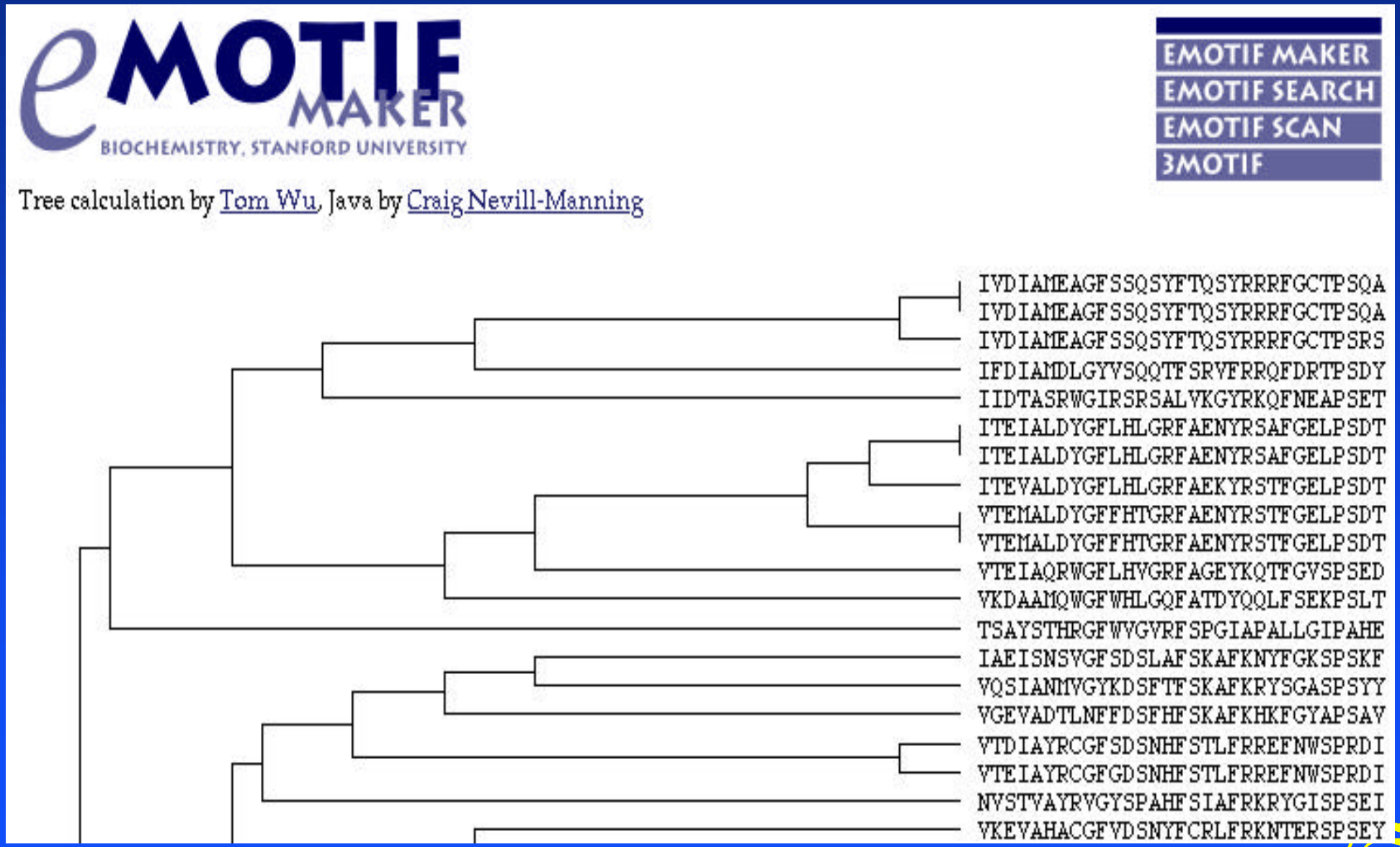
Histogram of Amino Acid Frequencies

(http://motif.stanford.edu/emotif/)



eMOTIF Dendrogram

(http://motif.stanford.edu/emotif/)



The score represents the number of bits saved if the sequences were transmitted with respect to the motif. For practical purposes, though, just the ranking is significant.

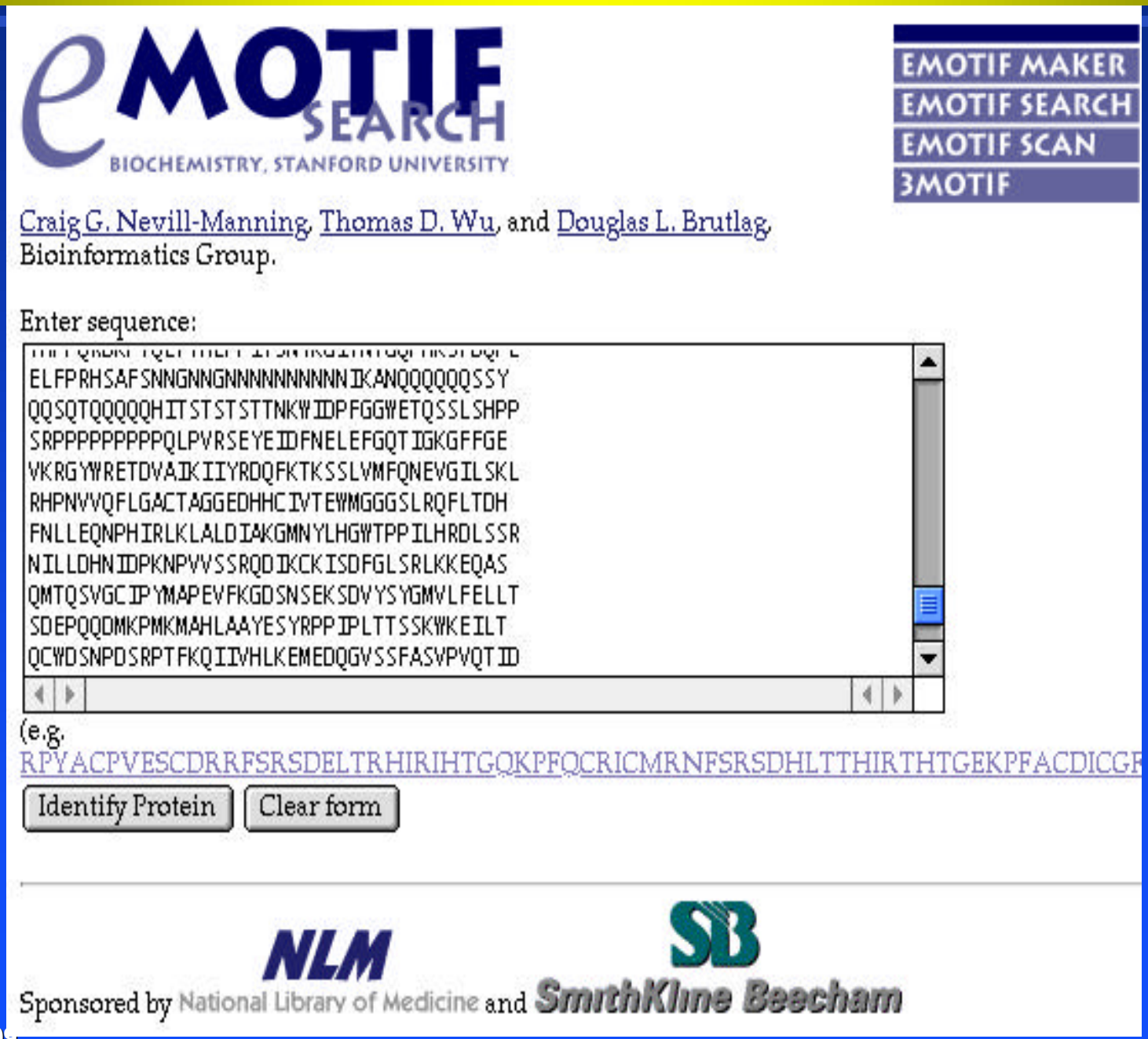
score	matches	expected	motif
971	32	10^{-2}	[ilv]..[iv]....g[filvy].....f...f......[ast]p..[fwy]
961	31	10^{-2}	[ilv]..[iv]....g[filvy].....f...f......[st]p..[fwy]
937	30	10^{-2}	[ilv]..[iv]....g[filvy].....f...f......[st]p..[fy]
923	29	10^{-2}	[iv]..[iv]....g[filvy].....f...f......[st]p..[fwy]
905	28	10^{-2}	[ilv]..[iv]....g[fy].....f...f......[ast]p..[fwy]
903	27	10^{-3}	[ilv]..[iv]....g[filvy]....[hy]f...f......[st]p..[filvy]
915	26	10^{-3}	[ilv]..[iv]....g[filvy]....[hy]f...f......[st]p..[fwy]
899	25	10^{-3}	[ilv]..[iv]....g[filvy]....yf...f......[st]p..[fwy]
869	24	10^{-3}	[ilv]..[iv]....g[filvy]....yf...f......[st]p..[fy]
846	23	10^{-4}	[iv]..[iv]....g[filvy]....yf...f......[st]p..[fwy]
818	22	10^{-4}	[ilv]..[iv]....g[fy]....[hy]f...f......[st]p..[fwy]
811	21	10^{-4}	[ilv]..[iv]....g[fy]....[hy]f...f[kr].....[st]p..[filvy]
773	20	10^{-4}	[iv]..[iv][ast]...g[fy].....f...[fy][kr]..[fy]..[st]p...
772	19	10^{-5}	[ilv]..[iv][ast]...g[filvy].s..[hy]f...[fy]...[fy]..[st]p...
766	18	10^{-5}	[ilv]..[iv][ast]...g[filvy].s..yf...[fy]...[fy]..tp...
746	17	10^{-6}	[ilv]..[iv][as]...g[filvy].s.[as][hy]f...[fy]...[fy]..[st]p...

Total Number of *e*MOTIFs in The *e*MOTIF-Search Database

Specificity

Database	# Alignments	10^{-10}	10^{-9}	10^{-8}	10^{-7}	10^{-6}	Total
BLOCKS	3,363	5,415	5,049	4,796	4,511	4,207	23,978
PRINTS	3,504	5,893	5,505	5,180	4,796	4,398	25,772
Total	7,867	11,308	10,554	9,976	9,307	8,602	49,747

Identifying Protein Function with *e*MOTIF Search (http://motif.stanford.edu/emotif-search/)



Identifying Protein Function with *e*MOTIF Search

eMOTIF SEARCH
BIOCHEMISTRY, STANFORD UNIVERSITY

EMOTIF MAKER
EMOTIF SEARCH
EMOTIF SCAN
3MOTIF

At a stringency of at least one in 10^{10} (no false positives expected) no matches.

At a stringency of at least one in 10^{9} (no false positives expected) no matches.

At a stringency of at least one in 10^{8} (no false positives expected)

Name	Description	Motif	Specificity
TYRKINASE	TYROSINE KINASE CATALYTIC DOMAIN SIGNATURE	[ilmv]..cw.....rp.f	$10^{-8.2}$
	positions 1537-1552	...RPPIPLTTSSKWKEILTQC▼DSNPDSRPTFKQIIVHLKEMEDQGV...	

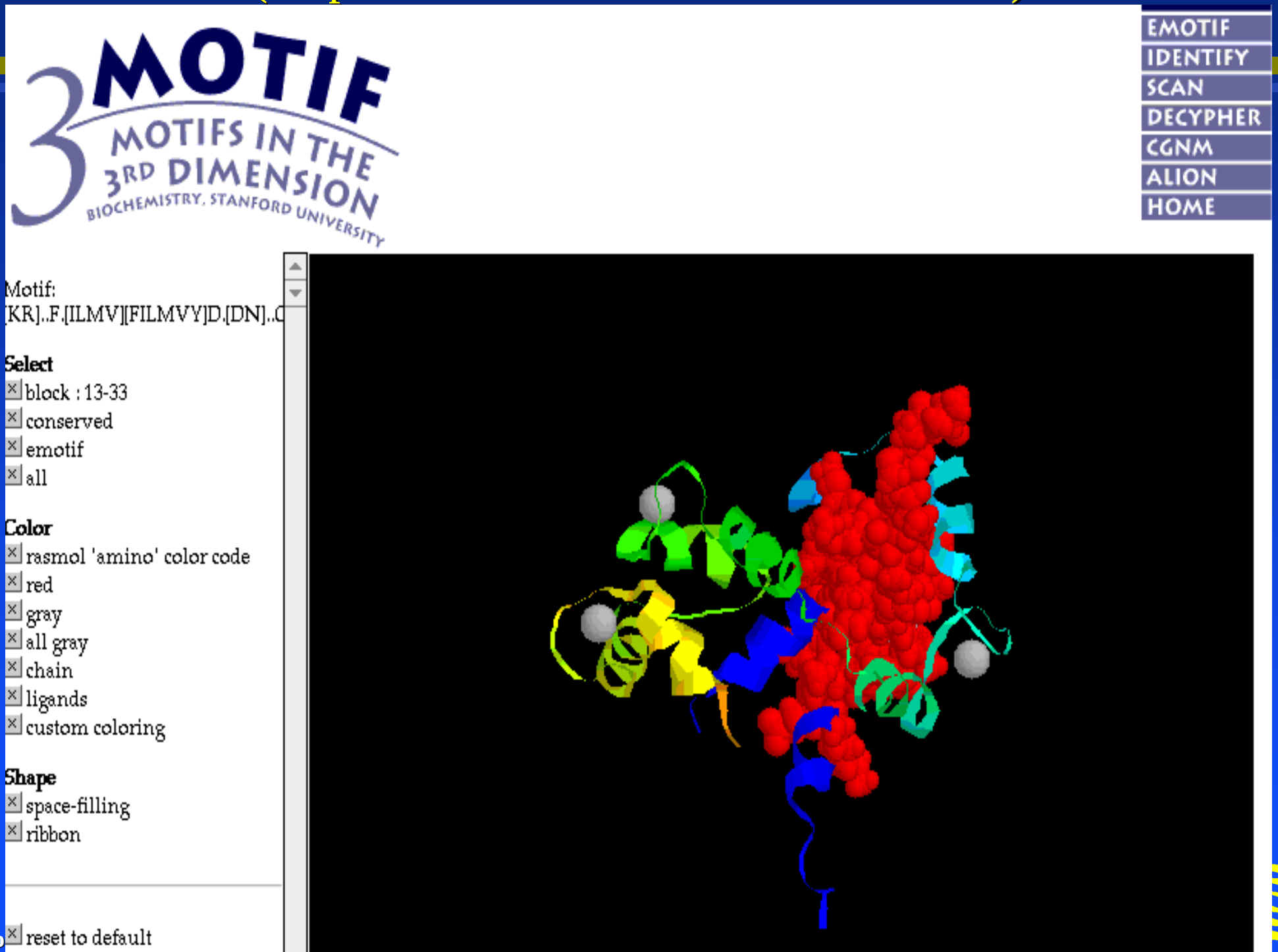
At a stringency of at least one in 10^{7} (no false positives expected)

Name	Description	Motif	Specificity
PROTEIN_KINASE_ATP	Protein kinases ATP-binding region proteins.	[hy]rd[ilv]...n.[filmv][ilmv]	10^{-7}
	positions 1414-1425	...AKGMNYLHGWTPPILHRDLSSRNILLDHNIDPKNPVVSSRQ...	3D
		wi...ggw	$10^{-7.3}$
	positions 1245-1253	...QQQHITSTSTSTTNK▼IDPFGG▼ETQSSLSHPPSRPPP...	

At a stringency of at least one in 10^{6} (expect one false positive)

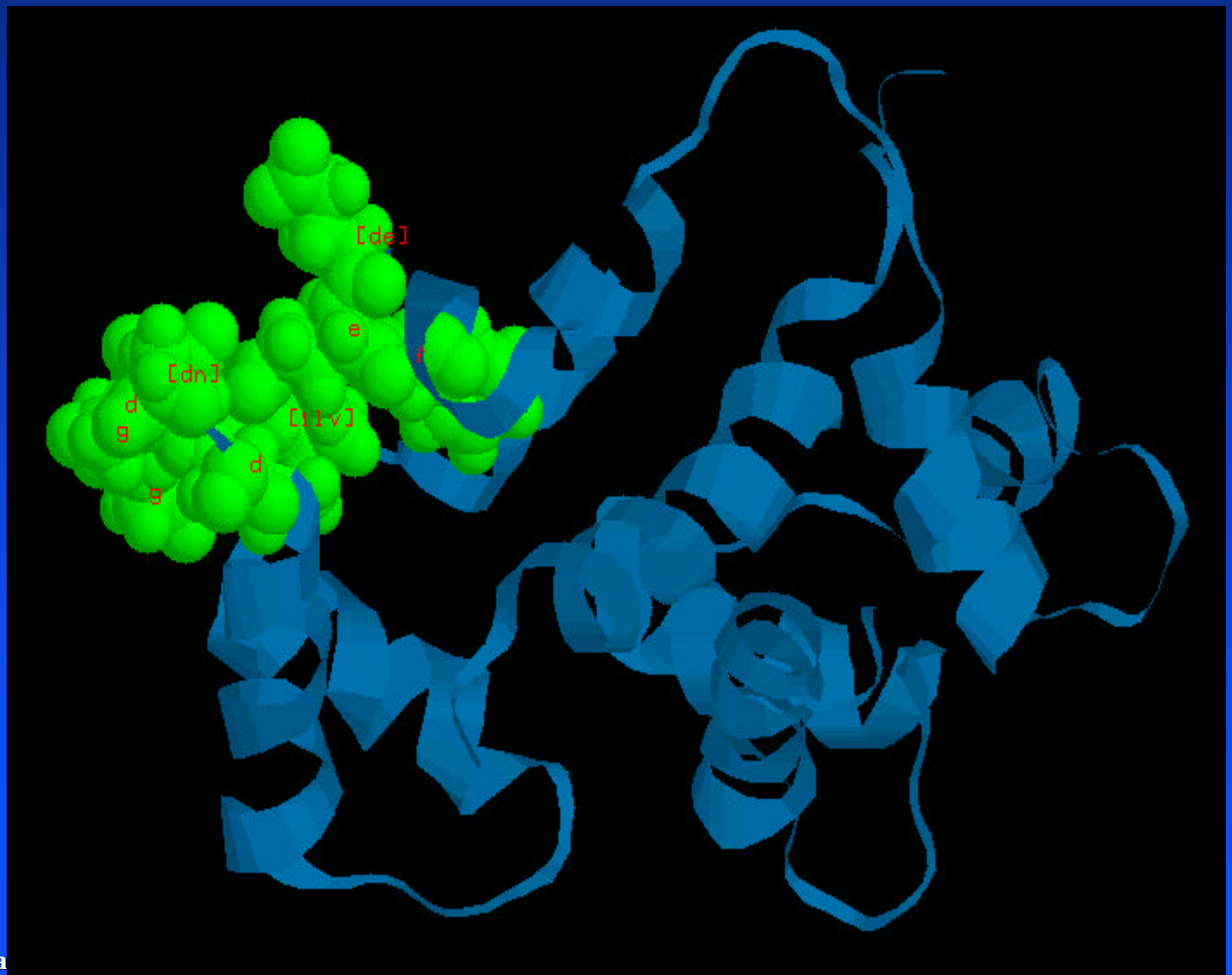
Name	Description	Motif	Specificity
PROTEIN_KINASE_ATP	Protein kinases ATP-binding region proteins.	[filmvy].[hy].d[filmv]...n.[filmv][filmvy]	10^{-6}
	positions 1412-1425	...DIAKGMNYLHGWTPPILHRDLSSRNILLDHNIDPKNPVVSSRQ...	3D

Mapping Sequence Motifs to Structural Motifs
(http://motif.stanford.edu/3motif/)



3MOTIF Labeling

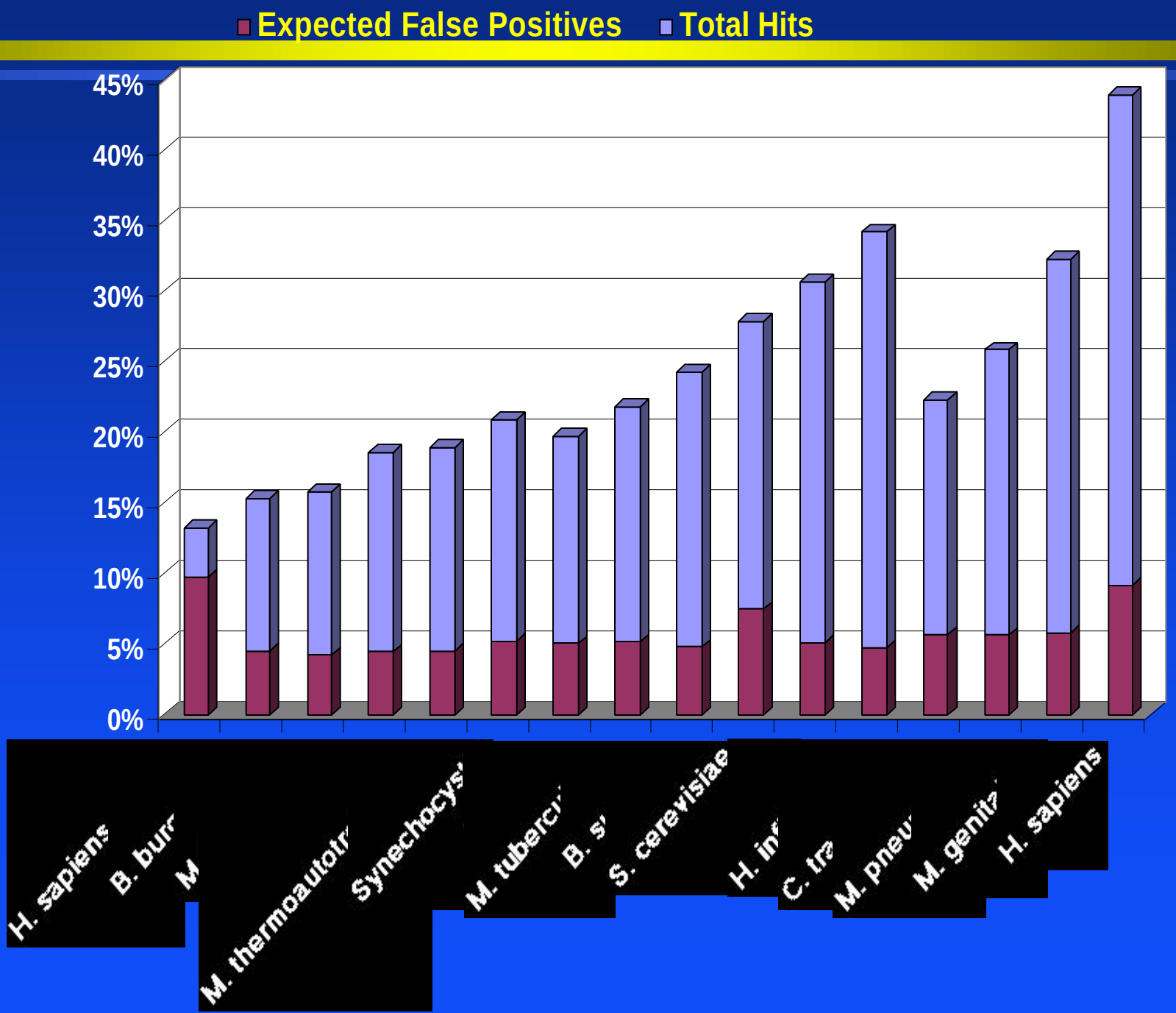
(http://motif.stanford.edu/3motif/)



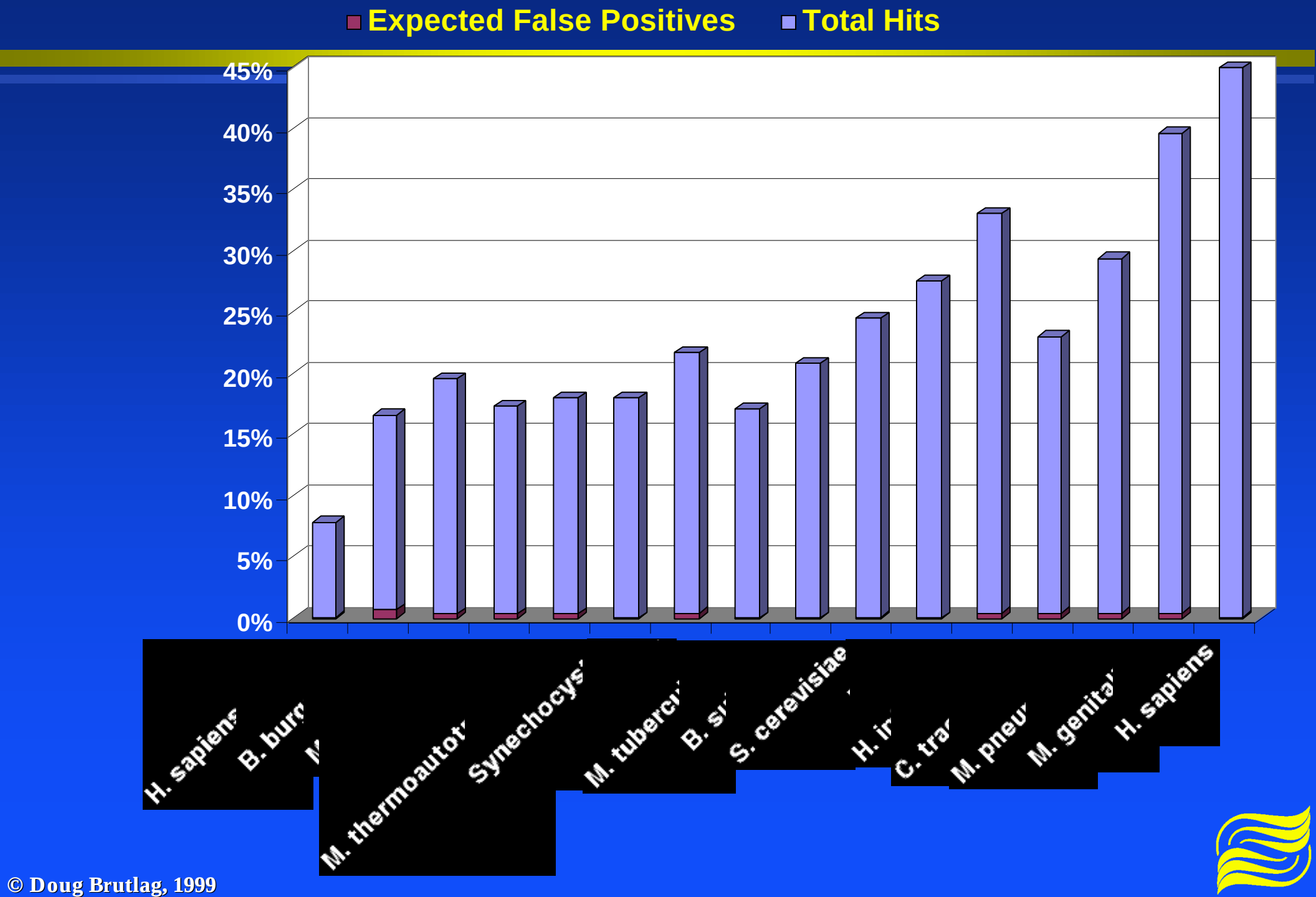
ORFs & ESTs in Proteomes
(http://motif.stanford.edu/genomes/)

Organism	Residues/ Proteome	Total ORFs
H. sapiens (ESTs)	3,500,344	5,733
B. burgdorferi	359,448	1,258
M. jannaschii	450,513	1,680
A. fulgidus	675,579	2,409
M. thermoautotrophicum	527,376	1,869
Synechocystis	1,036,375	3,169
H. pylori	498,404	1,565
M. tuberculosis	953,064	2,936
B. subtilius	1,239,329	4,100
S. cerevisiae	2,906,567	6,218
E. coli	1,366,029	4,286
H. influenzae	511,253	1,697
C. trachomatis	323,614	899
M. pneumoniae	238,328	677
M. genitalium	170,576	467
H. sapiens	3,301,896	5,733
Total	18,058,695	44,696

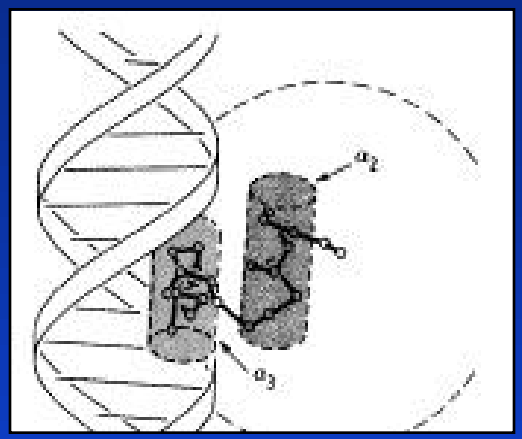
Fraction of ORFs with Prosite Hits



Fraction of ORFs with *e*MOTIF Hits



eMATRIX: Position-Specific Scoring Matrices



Structural or functional motif

⇓

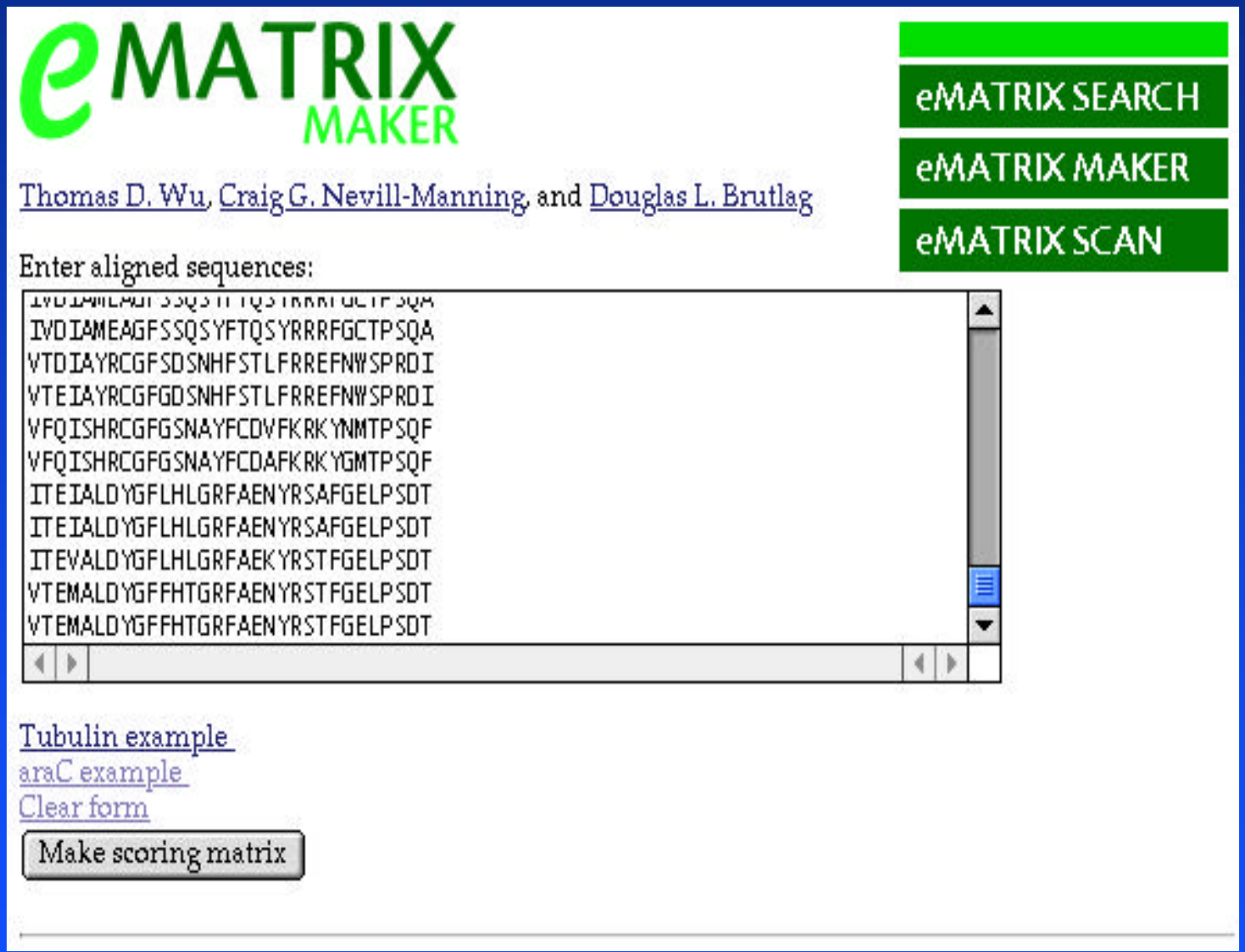
Examples of motif

HSGEQLAETLGMSRAAINKHIQ
VTLYDVAEYAGVSYQTVSRVVN
AMIKDVALKAKVSTATVSRALM
ATIKDVAKRAGVSTTTVSHVIN
ITIYDLAELSGVSASAVSAILN
LHLKDAAALLGVSEMTIRRDLN
TAYAELAKQFGVSPGTIHVRVE
GSLTEAAHLLGTSQPTVSRELA
MSQRELKNELGAGIATITRGSN
ITRQEIGQIVGCSRETVGRILK
FDIASVAQHVCLSPSRLSHLFR
LRIDEVARHVCLSPSRLAHLFR
MTRGDIGNYLGLTVETISRLLG
VTLEALADQVGMSPFHLHRLFK
.

⇒

	Position																					
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
A	2	1	3	13	10	12	67	4	13	9	1	2	4	3	6	15	4	4	4	11	0	10
R	7	5	8	9	4	0	1	16	7	0	1	0	1	16	6	6	0	11	28	3	0	16
N	0	8	0	1	0	0	0	2	1	1	10	0	7	1	3	1	0	4	8	0	1	11
D	0	1	0	1	13	0	0	12	1	0	4	0	1	2	0	0	0	0	1	1	0	3
C	0	0	1	0	0	0	0	0	0	2	2	1	0	0	0	0	0	0	0	1	0	0
Q	1	1	21	8	10	0	0	7	6	0	0	2	1	17	7	7	0	2	12	5	2	4
E	2	0	0	9	21	0	0	15	7	3	3	0	1	6	11	0	0	2	0	1	13	6
G	9	7	1	4	0	0	8	0	0	0	46	0	6	0	7	1	0	3	1	1	0	4
H	4	3	1	1	2	0	0	2	2	0	5	0	3	3	0	2	0	2	4	5	0	2
I	10	0	11	1	2	10	0	4	9	3	0	16	0	2	0	1	26	1	0	8	16	0
L	16	1	17	0	1	31	0	3	11	24	0	14	0	2	0	1	21	1	1	12	20	0
K	3	4	5	10	11	1	1	13	10	0	5	2	1	4	1	1	0	1	8	4	5	14
M	7	1	1	0	0	0	0	0	5	7	1	8	0	0	2	0	2	0	0	2	0	1
F	4	0	3	0	0	4	0	0	0	10	0	0	0	0	1	0	0	1	1	1	11	0
P	0	6	0	1	0	0	0	0	0	0	0	0	1	12	7	0	0	0	0	0	0	3
S	1	17	0	8	3	1	3	0	2	2	2	0	37	1	24	5	0	29	3	0	1	3
T	5	22	3	11	1	5	0	2	2	2	0	5	16	4	2	38	0	4	1	0	4	3
W	2	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0	2	10	0	0
Y	1	0	4	2	0	1	0	0	2	4	0	1	1	2	0	2	0	15	5	7	0	0
V	6	3	1	1	2	15	0	0	2	12	0	28	0	5	3	0	27	0	1	8	7	0

eMATRIX Maker
(http://motif.stanford.edu/ematrix-maker/)



eMATRIX SEARCH

To paste this matrix into eMATRIX-SCAN, click here.

```
ID
AC
DE
MA
  A   B   C   D   E   F   G   H   I   K   L   M   N   P   Q   R   S   T   V   W   X   Y   Z   *   -
-51 -36 -46 -60 -63 -46 -63 -59  38 -59  -8 -40  -4 -64 -61 -60 -58  12  21 -61 -36 -54 -62   0   0
  6  -9 -17 -14 -23  -2   0  -4   2   8   6 -15  -2 -28 -10 -19   7  15   2 -28  -4 -22 -15   0   0
  4   8 -33  24  22 -37 -32 -26 -36 -25 -38  -9 -12 -37  19  -5   8  -4 -20 -40 -13  -7  20   0   0
  3 -59 -43 -59 -60  -3 -58 -53  37 -56 -10  -4 -58 -61 -58 -57 -54   1  16 -53 -29   9 -59   0   0
 35 -57 -48 -58 -58 -59  -1 -56 -53 -55  -7 -53 -56 -60 -57 -58  19 -16  -9 -64 -31 -60 -57   0   0
  5  -1  -1  -7  -2  -7 -21  17  -5 -15   5  20   7 -27   9   4  -7   2  -4 -23  -3   3   4   0   0
 -7   9   5   8   3 -17 -19  13  -9   5  -6  14  10 -24  13  10  -3  -4 -11 -23  -3   1   8   0   0
  4 -21  37 -33 -32 -18 -31   1   3 -26  -3 -20  -6 -37  -6   8 -17 -14  14  27 -10  21 -18   0   0
-60 -42 -67 -68 -73 -15  42 -68 -69 -68 -74 -70 -10 -77  -2 -17 -66 -70   7 -74 -46 -74 -34   0   0
 -4 -59 -48 -60 -60  42 -10 -40  10 -57 -41 -43 -57 -64 -58 -56 -56 -54 -43 -36 -36  38 -59   0   0
 -6   7 -23  13  -6   1  -2  -3 -23 -12  -7  12  -1  -6  -4   5  14 -14 -11  30  -3   2  -5   0   0
-53   8 -58  21 -57 -67 -59  26 -64 -58 -68 -64 -10 -66 -14 -61  36 -54  -3 -72 -40 -63 -33   0   0
 -1 -18 -24 -26  -2 -31  -2 -24 -11 -22  -3 -23  -8  17  22  -8   9  13  -1 -35  -6 -31  11   0   0
  0 -21 -37 -40 -40  18   7 -37 -39 -38  -9 -38   2 -11  19 -40  26 -36  -2 -44 -14  -7  -7   0   0
 -4 -33 -45 -48 -45 -29 -49  38 -46 -40 -25 -43 -14 -55  -5  17 -11   6 -45 -36 -24  42 -23   0   0
-84 -91 -78 -92 -93  52 -89 -77 -68 -89   6 -68 -90 -96 -92 -87 -89 -85 -73 -67 -70 -56 -92   0   0
  7 -42  23 -44 -45 -41 -42  16  20 -43 -40 -39 -40 -49 -44 -45  27  11 -12 -48 -22   7 -45   0   0
-28  -5  -2 -10  -2 -35 -18 -25  -7  16 -33 -30   1   2  -1  28   4   7  -8 -39  -8 -11  -1   0   0
 10  -5 -18  -7   3 -22   3 -24   1 -10   7 -15  -2 -31 -23  -5  -9   4  17 -31  -4 -26 -12   0   0
-69 -76 -64 -77 -77  49 -74 -58  11 -74 -54 -56 -74 -81 -76 -72 -74 -70 -57 -51 -57  28 -77   0   0
 -5 -27 -55 -50 -46 -60 -22 -46 -58  37 -58 -52   2 -58  -4  27 -20 -21 -28 -62 -28 -56 -23   0   0
  2  -3 -33  -8   0 -38 -32   3 -38  27 -38 -33   5   2  10  17   0  -8 -35 -40 -11  -7   6   0   0
  2 -18   8 -24  -2  -2 -26  13  -4   3  -4 -18 -12 -31   4  12  -9   8  -4 -20  -5  26   1   0   0
-40 -46 -36 -48 -48  36 -47 -32 -11 -45   5   8 -44 -52 -46 -44 -21  17 -10 -29 -23  32 -47   0   0
-20   4 -53  -3 -10 -59  38 -51 -59 -17  -9 -55  12 -61  -5 -54  -7 -54 -58 -60 -26 -61  -7   0   0
 -7 -33  18 -33   7   4  -3 -28   6 -19  -6  31 -34 -40   0 -10 -33 -11  18  10  -8   4   3   0   0
 -1 -47 -43 -49 -50 -52 -49 -50   0 -11  -8 -46 -45  -2 -50 -49  23  34 -43 -58 -24 -53 -50   0   0
-14 -88 -88 -87 -87 -97 -89 -88 -91 -84 -94 -91 -90  51 -88 -15 -86 -87 -90 -99 -73 -95 -88   0   0
  5 -45 -45 -47 -46 -54 -17   8 -53  18 -21 -49 -43 -53  -3  10  31  -9 -50 -58 -22 -52 -22   0   0
  2   3 -27  13  18 -27  -8  20 -28   7  -6 -24 -11 -31  19  -2  -4 -24 -27 -31  -7   2  19   0   0
-18 -29 -36 -18  -3  27 -45 -28   1 -41  -7 -32 -42 -50 -42 -42 -24  10 -12  30 -18  39 -24   0   0
//
```

eMATRIX Search

eMATRIX SEARCH

eMATRIX SEARCH
eMATRIX MAKER
eMATRIX SCAN

Thomas D. Wu, Craig G. Nevill-Manning, and Douglas L. Brutlag

Desired significance threshold: 10e -8

Threshold on information: 60.0

Enter sequence:

HRDLSSRNILLDHNIDPKNPVVSSRQDIKCKISDFGLSRLKKEQASQMTQSVGCIPYMAPEVFKGDSNSE
KSDVYSYGMVLFELLTSDEPQQDMKPMKMAHLAAYESYRPPIPLTTSSKWKEILTQCWDSNPDSRPTFKQ
IIVHLKEMEDQGVSSFASVPVQTIDTGVYA

Fill in example
Clear form
Search for function

eMATRIX is based on minimal-risk scoring matrices, optimized for speed and accuracy. To cite this work, use:

Thomas D. Wu, Craig G. Nevill-Manning, and Douglas L. Brutlag, "Minimal-risk scoring matrices for sequence analysis", Journal of Computational Biology, 1999, in press.

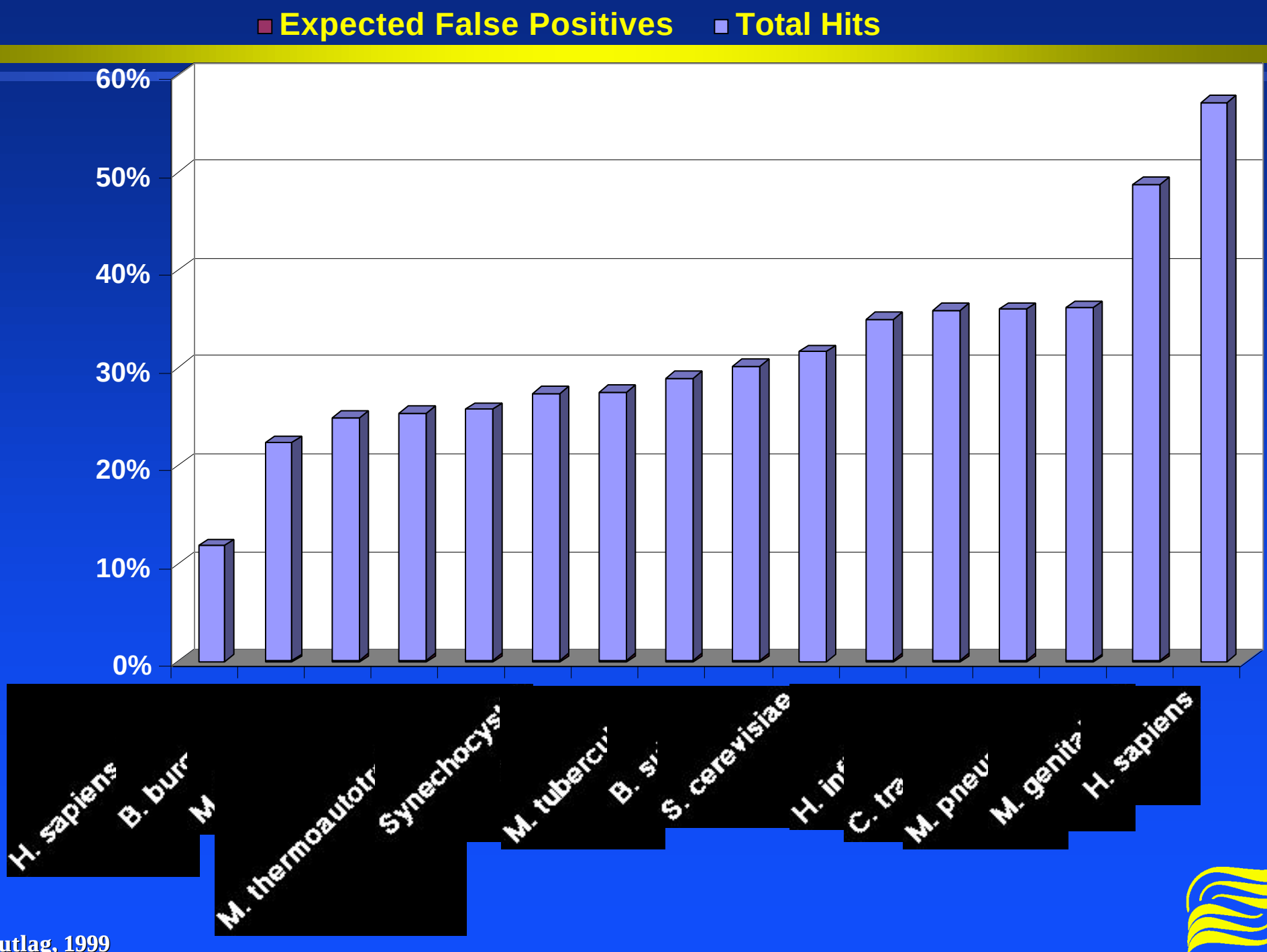
eMATRIX Search Results

(http://motif.stanford.edu/ematrix-search/)

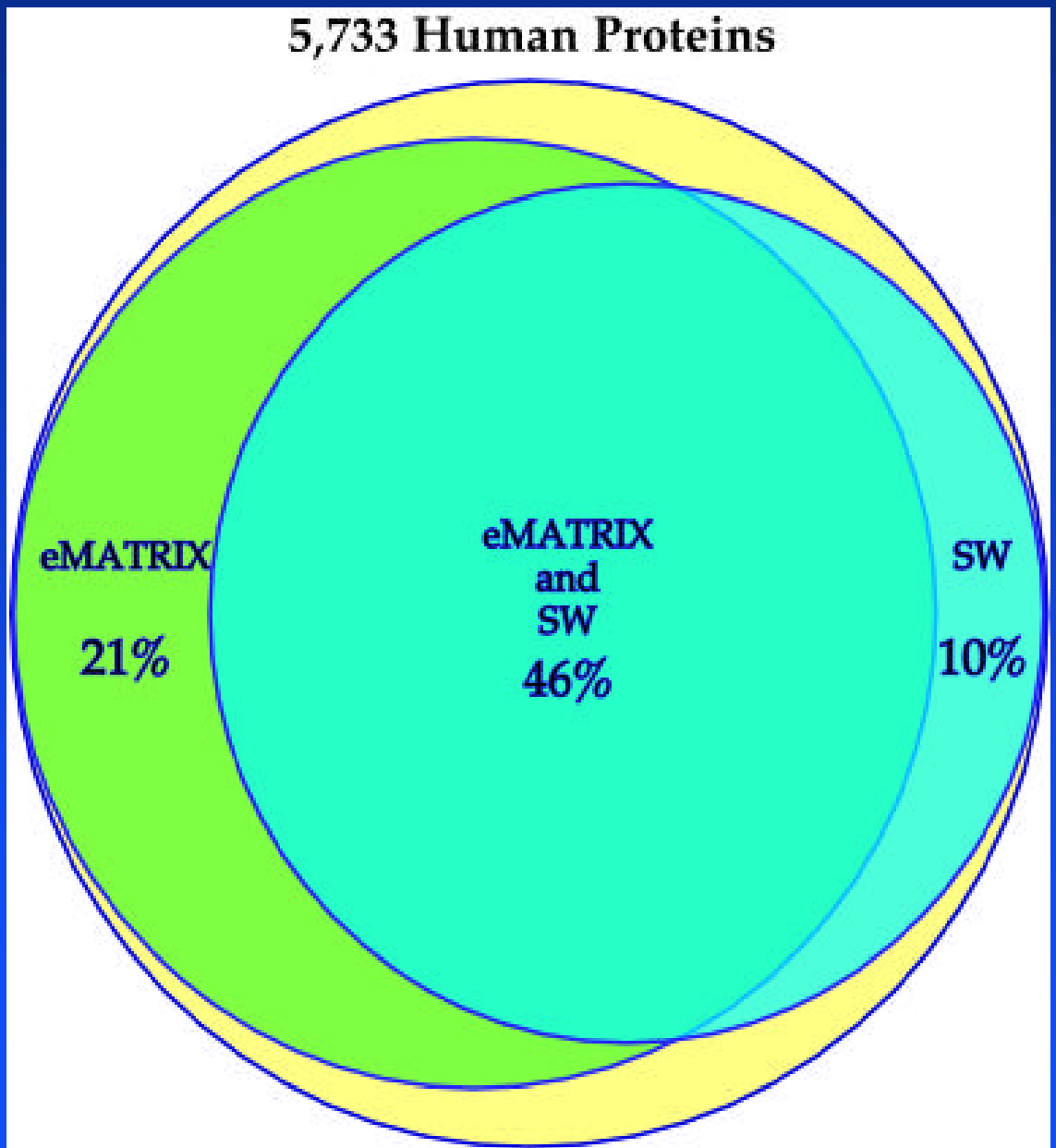
eMATRIX SEARCH

Rank	Prob.	Profile and Matching Segment
1.	5.765e-12	PR00109D TYROSINE KINASE CATALYTIC DOMAIN SIGNATURE 72 SDVYSYGMVLFELLTSDEPQQDM 95
2.	3.876e-11	BL00239E Receptor tyrosine kinase class II proteins. 43 EQASQMTQSVGCIPYMAPEVFKGDSNSEKSDVYSYGMVLFELLTSDEPQQ 93
3.	7.253e-11	BL00240G Receptor tyrosine kinase class III proteins. 89 EPQQDMKPMKMAHLAAYESYRPPIPLTTSSKWKEILTQCWDSNPDSRPTFKQI 142
4.	5.596e-10	PR00109E TYROSINE KINASE CATALYTIC DOMAIN SIGNATURE 116 TSSKWKEILTQCWDSNPDSRPTF 139
5.	4.757e-09	BL00790Q Receptor tyrosine kinase class V proteins. 108 YRPPIPLTTSSKWKEILTQCWDSNPDSRPTFKQIIVHLKEMEDQGVSSF 157
6.	6.824e-09	BL00107B Protein kinases ATP-binding region proteins. 71 KSDVYSYGMVLFELLT 87
7.	7.247e-09	BL00239F Receptor tyrosine kinase class II proteins. 97 MKMAHLAAYESYRPPIPLTTSSKWKEILTQCWDSNPDSRPTFKQI 142
8.	9.224e-09	BL00240F Receptor tyrosine kinase class III proteins. 42 KEQASQMTQSVGCIPYMAPEVFKGDSNSEKSDVYSYGMVLFELLTSDE 90

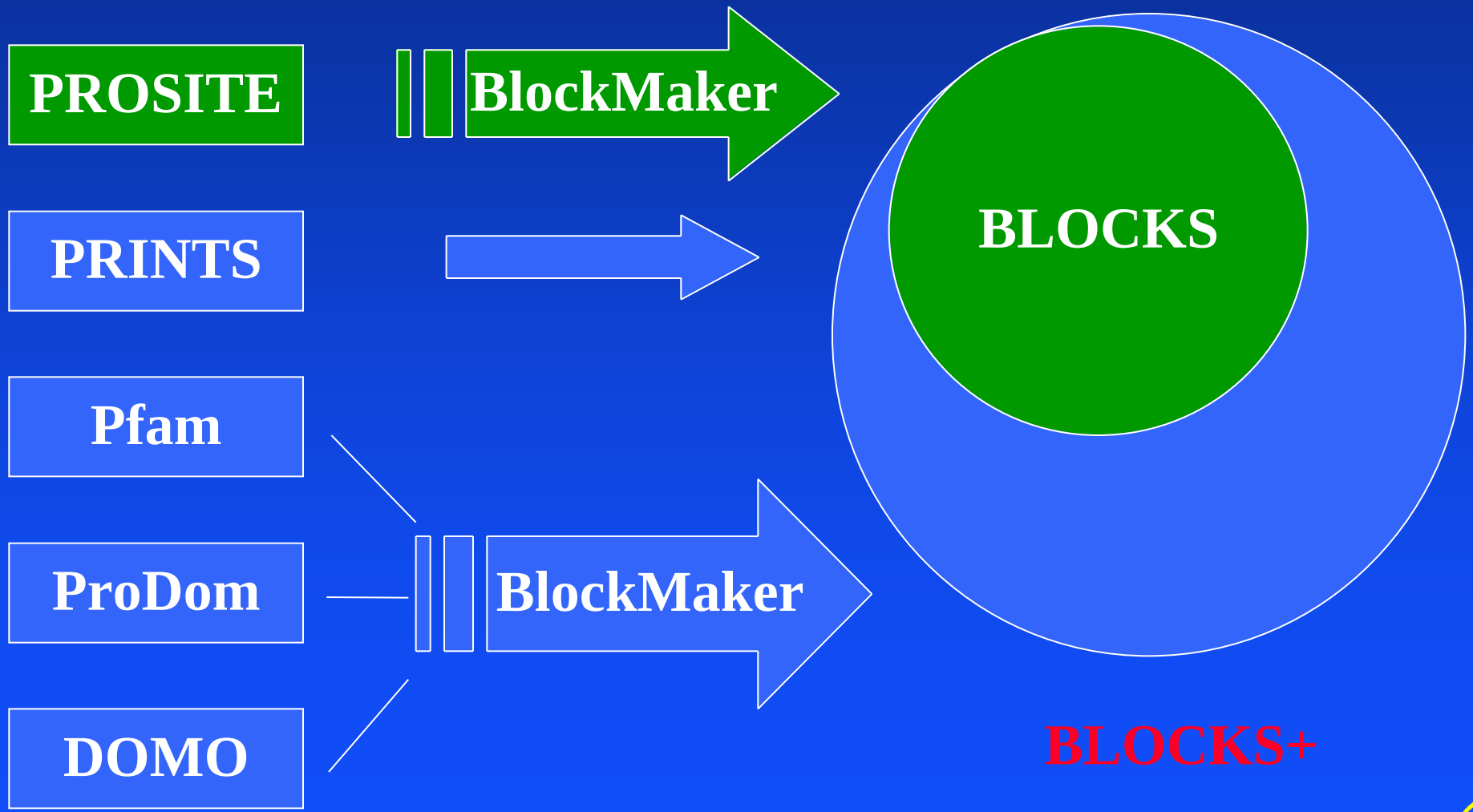
Fraction of ORFs with *e*MATRIX Hits



eMATRIX & Smith Waterman Identify 77% of Human Proteins



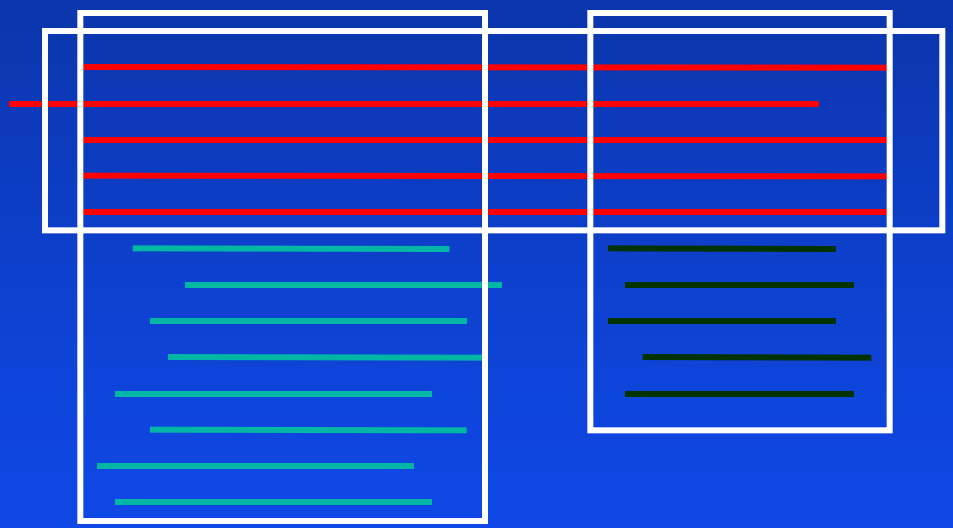
BLOCKS+ Is Based On Several Protein Family Databases



PSI-BLAST In Brief

1) Compare the query sequence to database
2) Construct profile from significant similarities
3) Compare the profile to database
4) Repeat step 2 and 3 until convergence

PSI-BLAST Results Contain Multiple Similar Regions



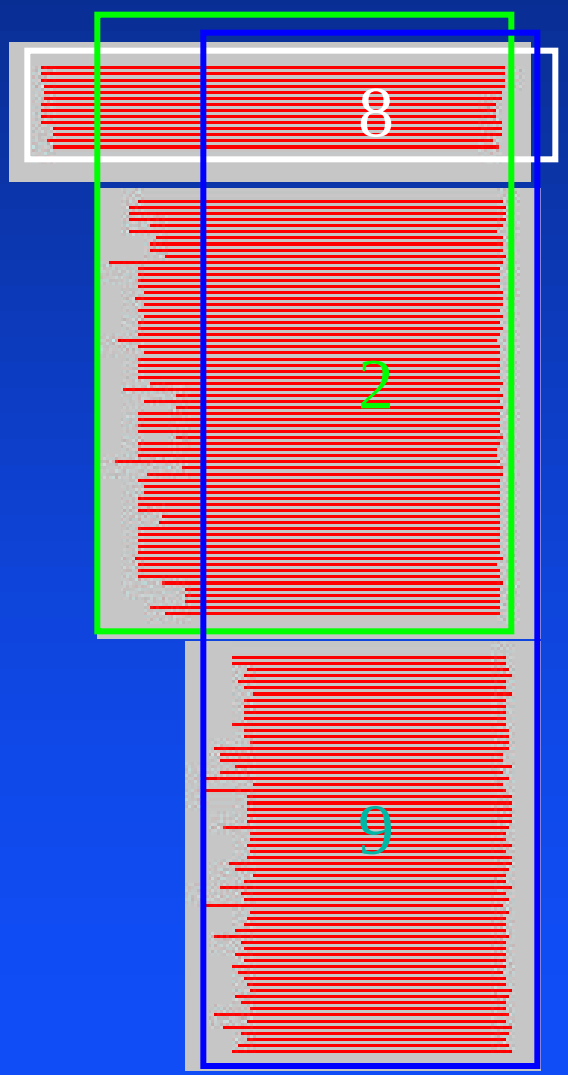
Major Steps:

1) Clustering

2) Alignment

3) Trimming

Clusters Are Organized Into Groups



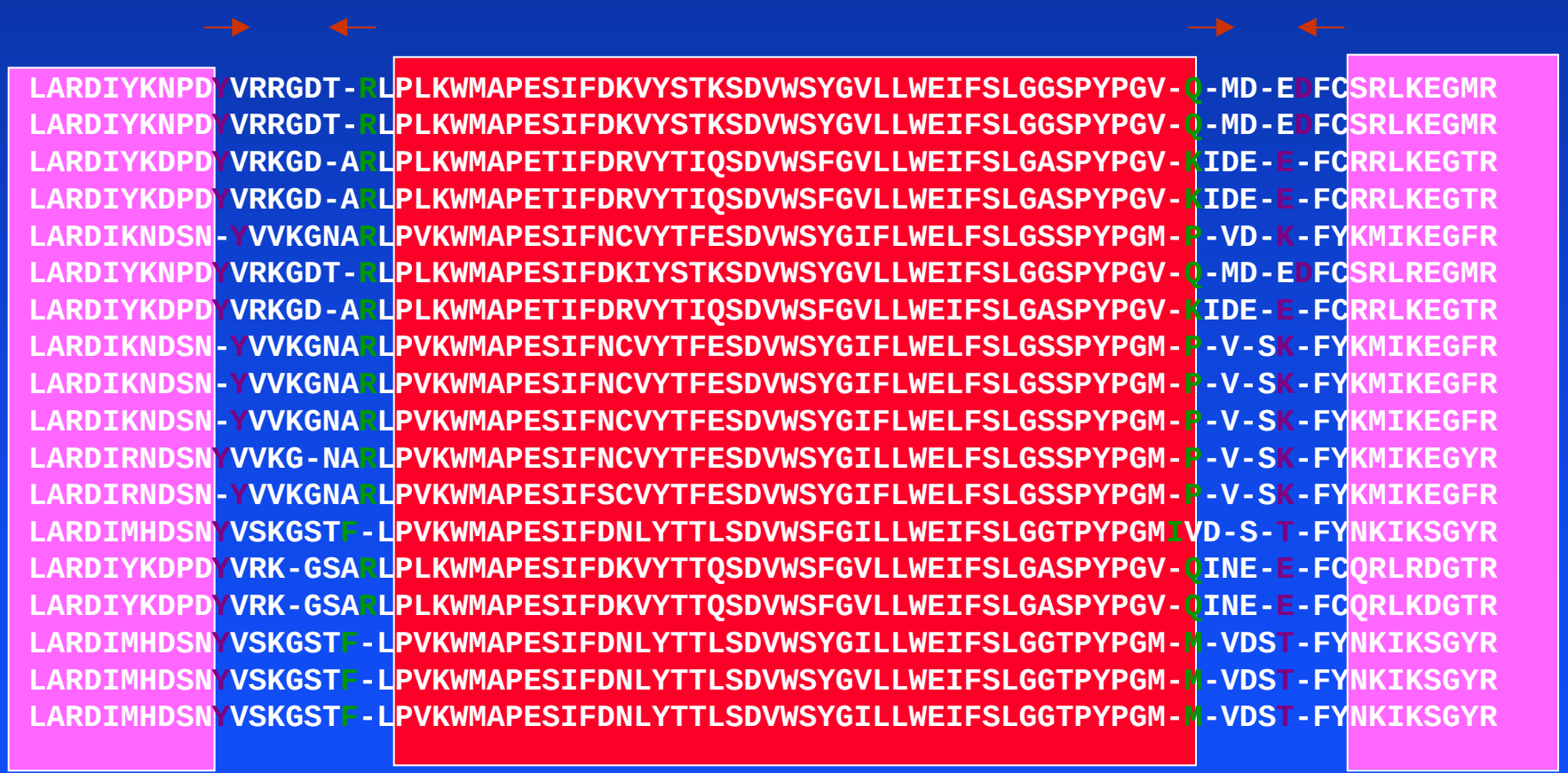
Gaps Are Propagated To Make Alignment

```
QUERY:     MQQL-DNPYIVRMIGICEAE-SWM
SUBJECT1:  MGQF-DHPNIIRLEGVVTKSRPVM

QUERY:     MQQL-DNPYIVRMIGICEAE-SWM
SUBJECT2:  MKQL-QHPRLVRLYAVVTQE-PIY

QUERY:     MQQL-DNPYIVRMIGICEAE-SWM
SUBJECT3:  MKMIGKHKNIINLLGACTQDGPLY
```

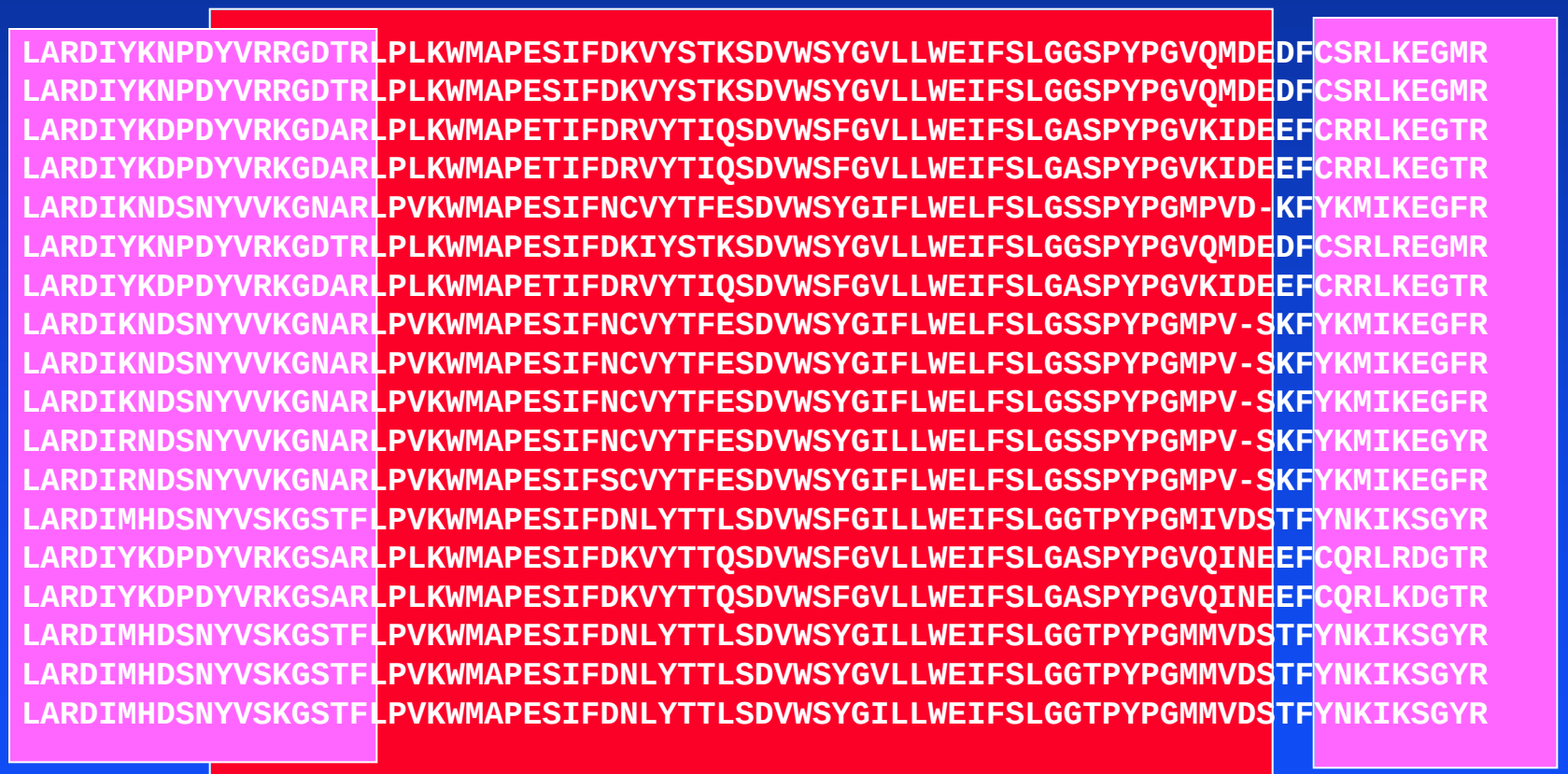
Gapped Alignments Are Trimmed to Form UnGapped Blocks

Edges Are Extended If Pass Test

$$D(i) = \log_2 20 + \sum_{k=1}^{20} p_k(i) \log_2 p_k(i) > 2(bits)$$



Overlapping Blocks Are Merged



Overlapping Blocks Are Merged

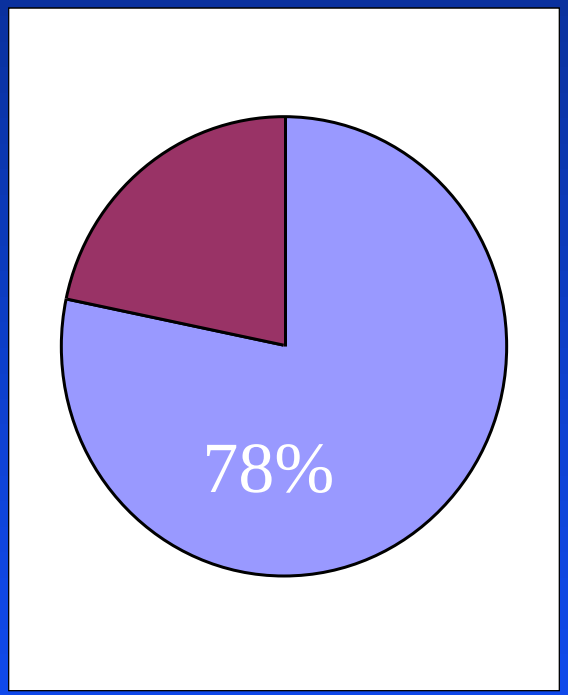
eBLOCKs Summary

- SWISS-PROT
  - 79,449 Sequences
- Filtered Target Set
  - 57,266 Sequences
- PSI-BLAST Searches
  - 17,415
- Final Number Of Seed Sequences
  - 9,503
- Final Number Of Groups
  - 16,658
- Final Number Of Blocks
  - 74,396

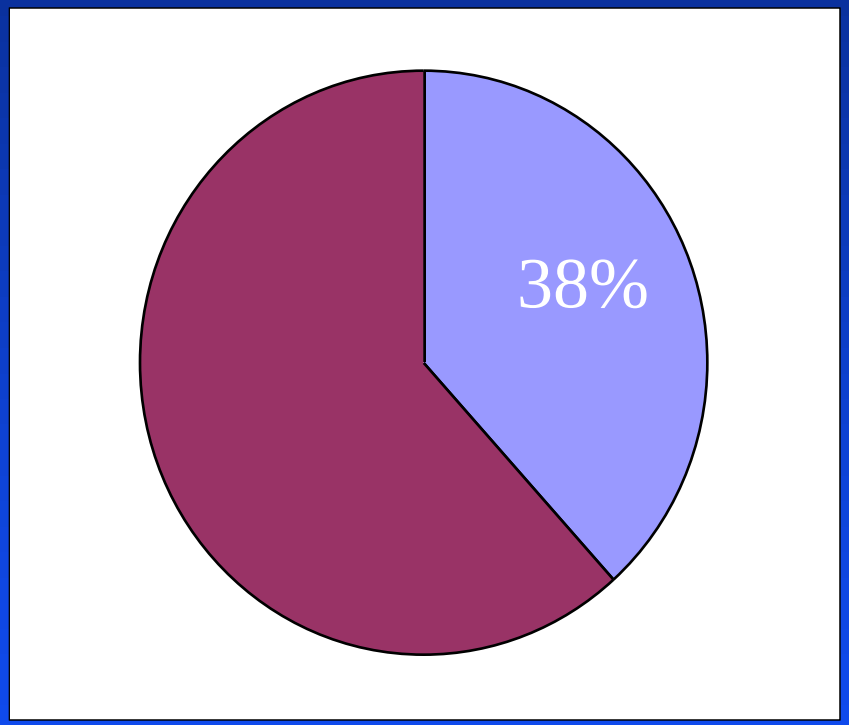
Comparing With BLOCKS+

- LAMA
  - Compares PSSMs in Pairs Of Blocks
- Blocks Contents
  - At Least One Common Sequence

eBLOCKs Is More Comprehensive



BLOCKS+
(9,498)



eBLOCKs
(74,396)

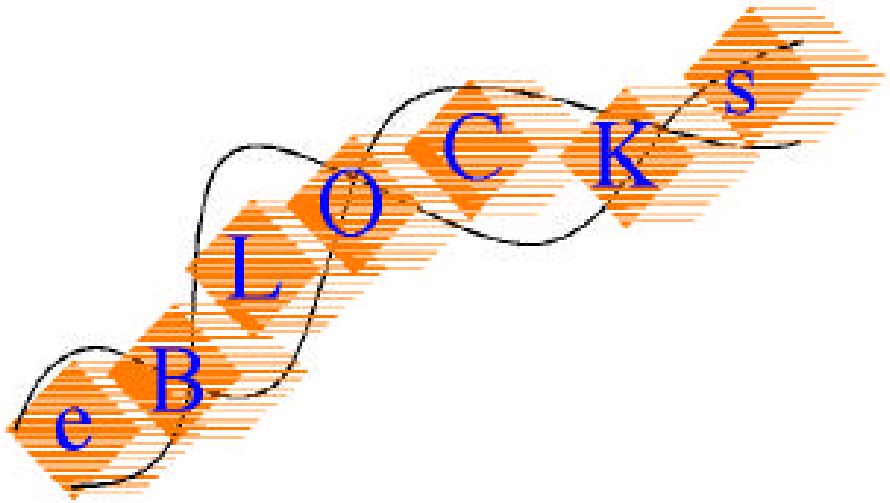
Web Access to eBLOCKS

(http://eblocks.stanford.edu)

Netscape: eBLOCKs Home

Back Forward Reload Home Search Guide Images Print Security Stop

Netsite: http://eblocks/



Enumeration of Blocks From PSI-Blast Results

About eBLOCKs

Search By Accession

Search By Keyword

Search A Sequence

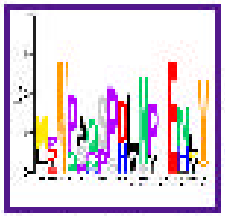
Qiaojuan Su and Douglas Brutlag
Brutlag Bioinformatics Group
Stanford University

An Entry From eBLOCKs

Netscape: eBLOCK P29358G1B1

eBLOCK P29358G1B1

- Sequence Logo



A PostScript File Will Be Generated. The Left Picture (GIF) Is Illustration Only.
You Could Use Ghostscript To View PostScript Files.
[About Sequence Logo]

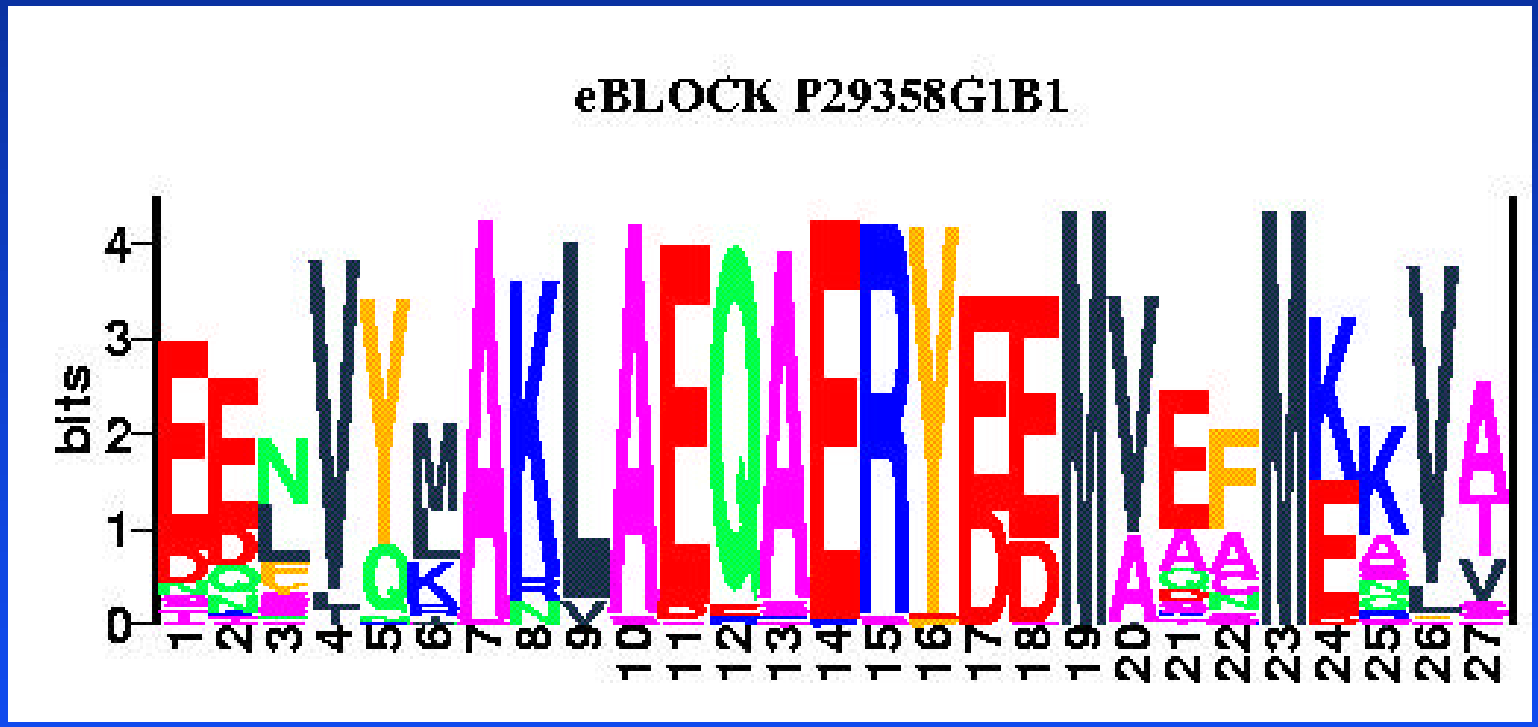
- Links To Blocks+ And Source Documents

  BLOCKS+ BL00796A: PROSITE PS00796

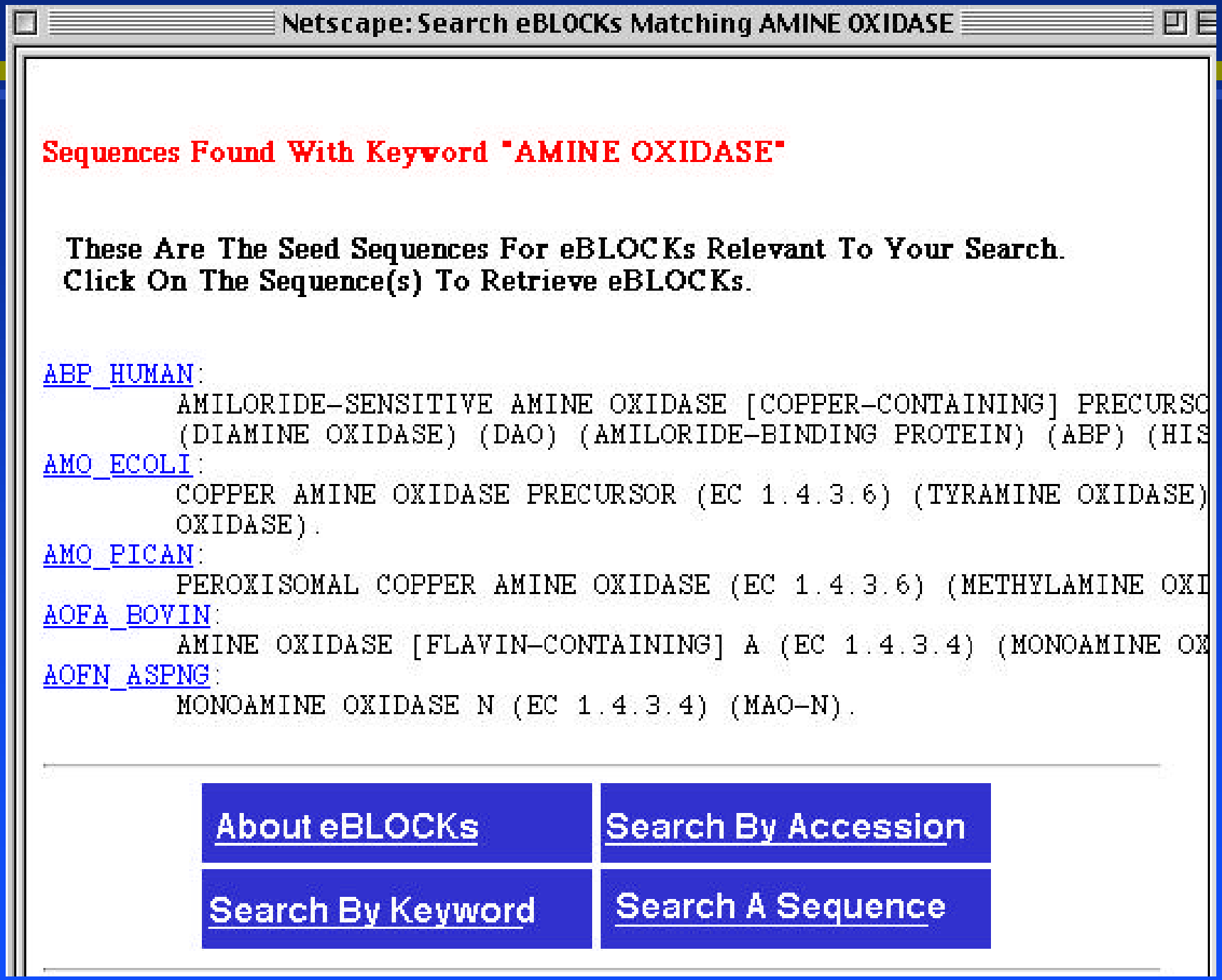
- **All eBLOCKs From The Same Seed Sequence** (143B_BOVIN)

```
ID   143B_BOVIN;
AC   P29358G1B1
DE   14-3-3 PROTEIN BETA/ALPHA (PROTEIN KINASE C INHIBITOR PROTEIN-1)
DE   (KCIP-1).
BL   width=27 seqs=72

143B_TOBAC (O49995) (    4) EENVYMAKLAEQAERYEEMVSFMEKVS  18
1435_SOLTU (P93784) (    6) EENVYMAKLAEQAERYEEMVEFMEKVV  13
1433_PEA   (P46266) (    9) EENVYMAKLAEQAERYEEMVEFMEKVS  15
1433_MESCR (P93259) (    8) EENVYMAKLAEQAERYEEMVEFMEKVA  13
143C_TOBAC (P93343) (    9) EENVYMAKLAEQAERYEEMVEFMEKVS  15
143O_ARATH (Q01525) (    6) EELVYMAKLAEQAERYEEMVEFMEKVS  15
```

An Entry From eBLOCKs



A Sample Keyword Search



Netscape: Sequence Search Result

eBLOCKs search results for: Sample Query Sequence

	Specificity	eBLOCK	Motif
1.	8.570e-22	P29358G1B2	l..e[de]r.l[ilv]s..ykn.[
		43--76 VEFMEKVSANADSEE LTVEERNLLSVAYKNVIGARRASW	
	14-3-3 PROTEIN BETA/ALPHA (PROTEIN KINASE C INHIBITOR PR		
	(KCIP-1).		
2.	1.621e-18	P29358G1B8	[ilv]...[eq].....a..[i
		185--226 PPTHPIRLGLALNFS VFYYEILNSPDRACNL	
	14-3-3 PROTEIN BETA/ALPHA (PROTEIN KINASE C INHIBITOR PR		
	(KCIP-1).		
3.	1.660e-15	P29358G1B5	[de]...f..k[ilmv][ekq]gd
		122--142 LKLLDTRLIPSASSG DSKVFYLKMKGDYHRY	
	14-3-3 PROTEIN BETA/ALPHA (PROTEIN KINASE C INHIBITOR PR		
	(KCIP-1).		
4.	2.399e-13	P29358G1B9	imql[fly].dn[fly]t.w
		226--238 ELDTLGEESYKDSTL IMQLLRDNLTLW TSD	
	14-3-3 PROTEIN BETA/ALPHA (PROTEIN KINASE C INHIBITOR PR		
	(KCIP-1).		
5.	6.784e-13	P29358G1B1	[ilv]..[as].[ilv][as].[e
		12--32 ....MAAAHTPREEN VYMAKLAEQAERYEEMVEFM EKV	
	14-3-3 PROTEIN BETA/ALPHA (PROTEIN KINASE C INHIBITOR PR		
	(KCIP-1).		