

Sequence profiles

Morten Nielsen,
Department of Health Technology, DTU
and
Instituto de Investigaciones Biotecnológicas,
Universidad de San Martín, Argentina

Identification of essential residues in protein sequences

CENTER FOR BIOLOGICAL
SEQUENCE ANALYSIS CBS

Some one tells you that this protein contains one G that is essential for its function.

Which G would you select?

TKAVVLTFNTSVEICLVMQGTSIV---AAESHPLHLHGFnFPSNFNLVDGMERNTAGVP

What have we learned ...

- Sequence logo is a power tool to visualize (binding) motifs
 - Information content identifies essential residues for function and/or structural stability
- Weight matrices can be derived from very limited number of data using the techniques of
 - Sequence weighting
 - Pseudo counts

Sequence Profiles and Weight matrices

- Alignments based on conventional scoring matrices (BLOSUM62) scores all positions in a sequence in an equal manner
- Some positions are highly conserved, some are highly variable (more than what is described in the BLOSUM matrix)
- Sequence profile are ideal suited to describe such position specific variations

Sequence alignment

- Conventional sequence alignment uses a (BLOSUM) scoring matrix to identify amino acids matches in the two protein sequences

	A	R	N	D	C	Q	E	G	H	I	L	K	M	F	P	S	T	W	Y	V
A	4	-1	-2	-2	0	-1	-1	0	-2	-1	-1	-1	-1	-2	-1	1	0	-3	-2	0
R	-1	5	0	-2	-3	1	0	-2	0	-3	-2	2	-1	-3	-2	-1	-1	-3	-2	-3
N	-2	0	6	1	-3	0	0	0	1	-3	-3	0	-2	-3	-2	1	0	-4	-2	-3
D	-2	-2	1	6	-3	0	2	-1	-1	-3	-4	-1	-3	-3	-1	0	-1	-4	-3	-3
C	0	-3	-3	-3	9	-3	-4	-3	-3	-1	-1	-3	-1	-2	-3	-1	-1	-2	-2	-1
Q	-1	1	0	0	-3	5	2	-2	0	-3	-2	1	0	-3	-1	0	-1	-2	-1	-2
E	-1	0	0	2	-4	2	5	-2	0	-3	-3	1	-2	-3	-1	0	-1	-3	-2	-2
G	0	-2	0	-1	-3	-2	-2	6	-2	-4	-4	-2	-3	-3	-2	0	-2	-2	-3	-3
H	-2	0	1	-1	-3	0	0	-2	8	-3	-3	-1	-2	-1	-2	-1	-2	-2	2	-3
I	-1	-3	-3	-3	-1	-3	-3	-4	-3	4	2	-3	1	0	-3	-2	-1	-3	-1	3
L	-1	-2	-3	-4	-1	-2	-3	-4	-3	2	4	-2	2	0	-3	-2	-1	-2	-1	1
K	-1	2	0	-1	-3	1	1	-2	-1	-3	-2	5	-1	-3	-1	0	-1	-3	-2	-2
M	-1	-1	-2	-3	-1	0	-2	-3	-2	1	2	-1	5	0	-2	-1	-1	-1	-1	1
F	-2	-3	-3	-3	-2	-3	-3	-3	-1	0	0	-3	0	6	-4	-2	-2	1	3	-1
P	-1	-2	-2	-1	-3	-1	-1	-2	-2	-3	-3	-1	-2	-4	7	-1	-1	-4	-3	-2
S	1	-1	1	0	-1	0	0	0	-1	-2	-2	0	-1	-2	-1	4	1	-3	-2	-2
T	0	-1	0	-1	-1	-1	-2	-2	-1	-1	-1	-1	-2	-1	1	5	-2	-2	0	
W	-3	-3	-4	-4	-2	-2	-3	-2	-2	-3	-2	-3	-1	1	-4	-3	-2	11	2	-3
Y	-2	-2	-2	-3	-2	-1	-2	-3	2	-1	-1	-2	-1	3	-3	-2	-2	2	7	-1
V	0	-3	-3	-3	-1	-2	-2	-3	-3	3	1	-2	1	-1	-2	-2	0	-3	-1	4

Alignment scoring matrices

- Blosum62 score matrix. Fg=1. Ng=0?

	L	A	G	D	S	D
F						
I						
G						
D						
S						
L						

A	R	N	D	C	Q	E	G	H	I	L	K	M	F	P	S	T	W	Y	V	
A	4	-1	-2	-2	0	-1	-1	0	-2	-1	-1	-1	-2	-1	1	0	-3	-2	0	
R	-1	5	0	-2	-3	1	0	-2	0	-3	-2	2	-1	-3	-2	-1	-1	-3	-2	-3
N	-2	0	6	1	-3	0	0	0	1	-3	-3	0	-2	-3	-2	1	0	-4	-2	-3
D	-2	-2	1	6	-3	0	2	-1	-1	-3	-4	-1	-3	-3	-1	0	-1	-4	-3	-3
C	0	-3	-3	-3	9	-3	-4	-3	-3	-1	-1	-3	-1	-2	-3	-1	-1	-2	-2	-1
Q	-1	1	0	0	-3	5	2	-2	0	-3	-2	1	0	-3	-1	0	-1	-2	-1	-2
E	-1	0	0	2	-4	2	5	-2	0	-3	-3	1	-2	-3	-1	0	-1	-3	-2	-2
G	0	-2	0	-1	-3	-2	-2	6	-2	-4	-4	-2	-3	-3	-2	0	-2	-2	-3	-3
H	-2	0	1	-1	-3	0	0	-2	8	-3	-3	-1	-2	-1	-2	-1	-2	-2	2	-3
I	-1	-3	-3	-3	-1	-3	-3	-4	-3	4	2	-3	1	0	-3	-2	-1	-3	-1	3
L	-1	-2	-3	-4	-1	-2	-3	-4	-3	2	4	-2	2	0	-3	-2	-1	-2	-1	1
K	-1	2	0	-1	-3	1	1	-2	-1	-3	-2	5	-1	-3	-1	0	-1	-3	-2	-2
M	-1	-1	-2	-3	-1	0	-2	-3	-2	1	2	-1	5	0	-2	-1	-1	-1	-1	1
F	-2	-3	-3	-3	-2	-3	-3	-3	-1	0	0	-3	0	6	-4	-2	-2	1	3	-1
P	-1	-2	-2	-1	-3	-1	-1	-2	-2	-3	-3	-1	-2	-4	7	-1	-1	-4	-3	-2
S	1	-1	1	0	-1	0	0	-1	-2	-2	0	-1	-2	-1	4	1	-3	-2	-2	-2
T	0	-1	0	-1	-1	-1	-2	-2	-1	-1	-1	-1	-2	-1	1	5	-2	-2	0	
W	-3	-3	-4	-4	-2	-2	-3	-2	-2	-3	-2	-3	-1	1	-4	-3	-2	11	2	-3
Y	-2	-2	-2	-3	-2	-1	-2	-3	2	-1	-1	-2	-1	3	-3	-2	-2	2	7	-1
V	0	-3	-3	-3	-1	-2	-2	-3	-3	3	1	-2	1	-1	-2	-2	0	-3	-1	4

Alignment scoring matrices

- Blosum62 score matrix. Fg=1. Ng=0?

	L	A	G	D	S	D
F	0	-2	-3	-3	-2	-3
I	2 → -1	-4	-3	-2	-3	
G	-4	0	6	-1	0	-1
D	-4	-2	-1	6	0	6
S	-2	1	0	0	4	0
L	4	-1	-4	-4	-2	-4

A	R	N	D	C	Q	E	G	H	I	L	K	M	F	P	S	T	W	Y	V	
A	4	-1	-2	-2	0	-1	-1	0	-2	-1	-1	-1	-2	-1	1	0	-3	-2	0	
R	-1	5	0	-2	-3	1	0	-2	0	-3	-2	2	-1	-3	-2	-1	-1	-3	-2	-3
N	-2	0	6	1	-3	0	0	0	1	-3	-3	0	-2	-3	-2	1	0	-4	-2	-3
D	-2	-2	1	6	-3	0	2	-1	-1	-3	-4	-1	-3	-3	-1	0	-1	-4	-3	-3
C	0	-3	-3	-3	9	-3	-4	-3	-3	-1	-1	-3	-1	-2	-3	-1	-1	-2	-2	-1
Q	-1	1	0	0	-3	5	2	-2	0	-3	-2	1	0	-3	-1	0	-1	-2	-1	-2
E	-1	0	0	2	-4	2	5	-2	0	-3	-3	1	-2	-3	-1	0	-1	-3	-2	-2
G	0	-2	0	-1	-3	-2	-2	6	-2	-4	-4	-2	-3	-3	-2	0	-2	-2	-3	-3
H	-2	0	1	-1	-3	0	0	-2	8	-3	-3	-1	-2	-1	-2	-1	-2	-2	2	-3
I	-1	-3	-3	-3	-1	-3	-3	-4	-3	4	2	-3	1	0	-3	-2	-1	-3	-1	3
L	-1	-2	-3	-4	-1	-2	-3	-4	-3	2	4	-2	2	0	-3	-2	-1	-2	-1	1
K	-1	2	0	-1	-3	1	1	-2	-1	-3	-2	5	-1	-3	-1	0	-1	-3	-2	-2
M	-1	-1	-2	-3	-1	0	-2	-3	-2	1	2	-1	5	0	-2	-1	-1	-1	-1	1
F	-2	-3	-3	-3	-2	-3	-3	-3	-1	0	0	-3	0	6	-4	-2	-2	1	3	-1
P	-1	-2	-2	-1	-3	-1	-1	-2	-2	-3	-3	-1	-2	-4	7	-1	-1	-4	-3	-2
S	1	-1	1	0	-1	0	0	-1	-2	-2	0	-1	-2	-1	4	1	-3	-2	-2	0
T	0	-1	0	-1	-1	-1	-2	-2	-1	-1	-1	-1	-2	-1	1	5	-2	-2	0	
W	-3	-3	-4	-4	-2	-2	-3	-2	-2	-3	-2	-3	-1	1	-4	-3	-2	11	2	-3
Y	-2	-2	-2	-3	-2	-1	-2	-3	2	-1	-1	-2	-1	3	-3	-2	-2	2	7	-1
V	0	-3	-3	-3	-1	-2	-2	-3	-3	3	1	-2	1	-1	-2	-2	0	-3	-1	4

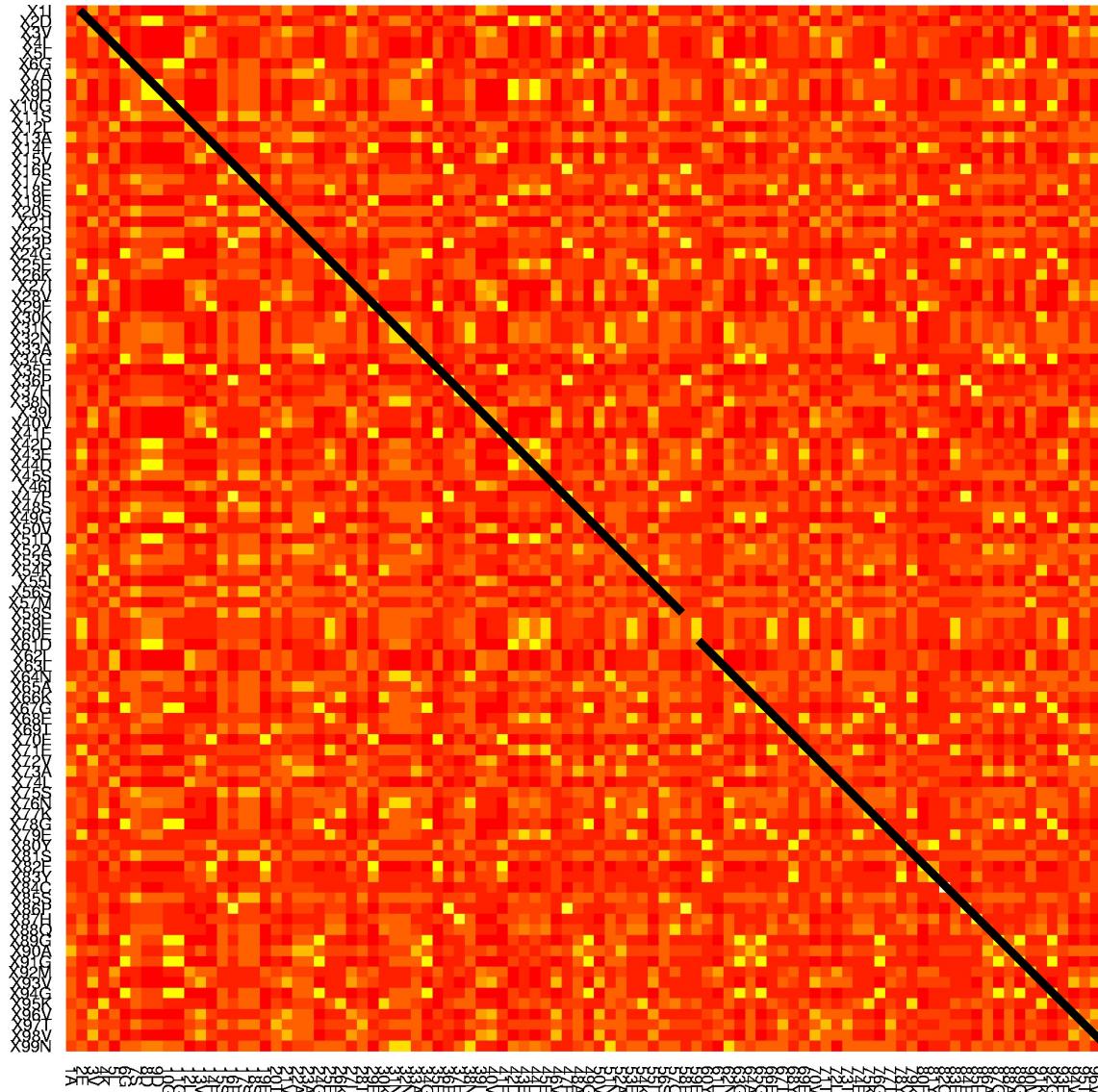
- Score = 2-1+6+6+4=17

LAGDS

I-GDS

When Blast works!

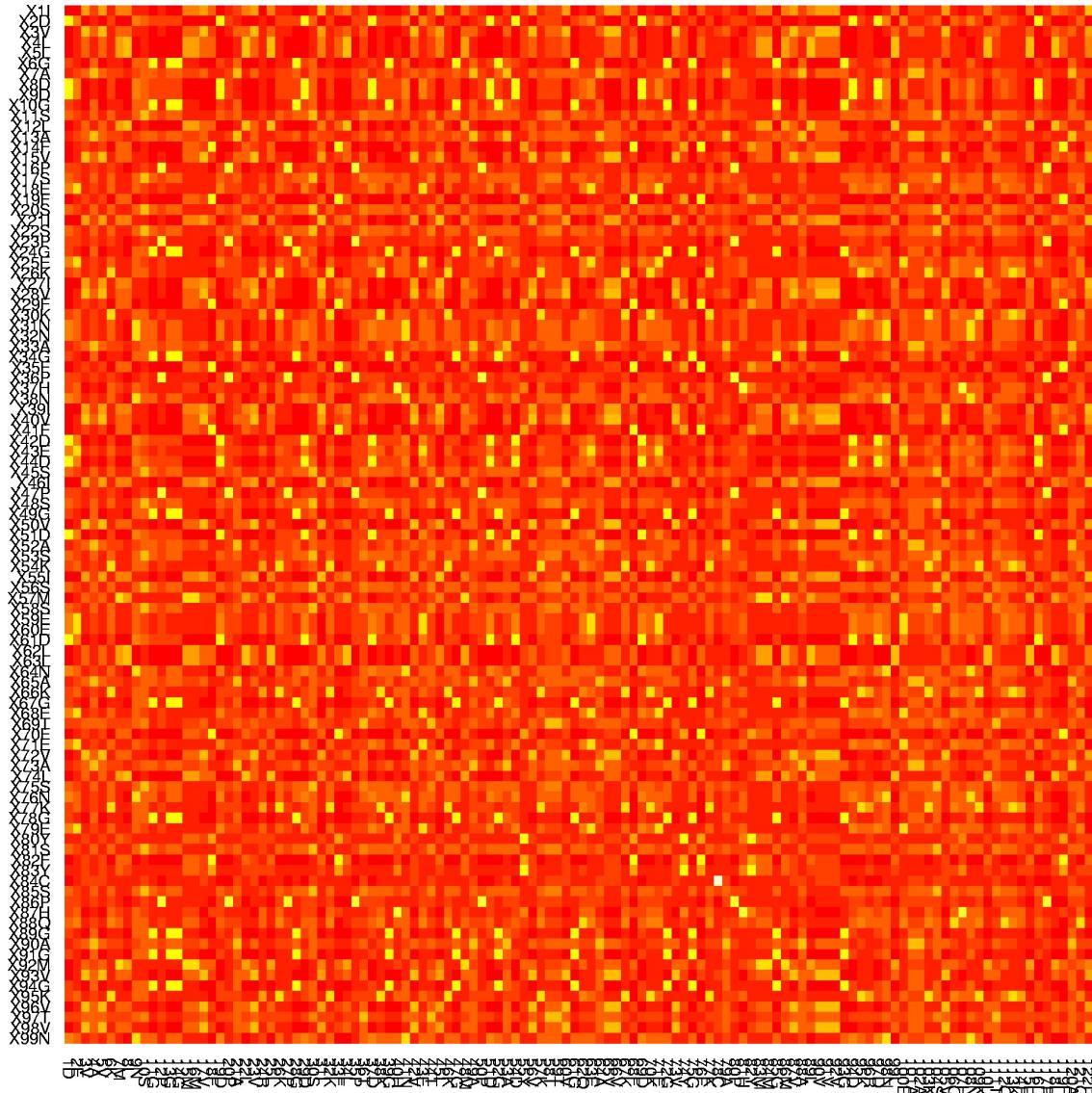
1PLC._



1PLB._

When Blast fails!

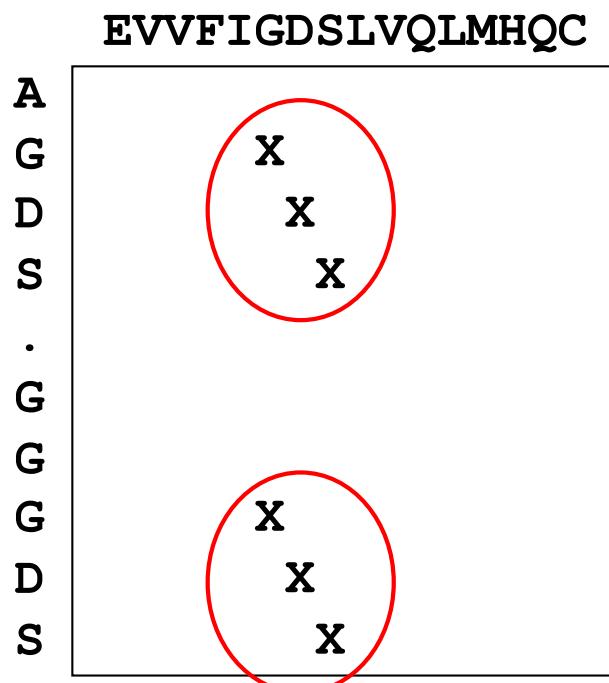
1PLC._



1PMY._

What goes wrong when Blast fails?

- Conventional sequence alignment uses a (Blosum) scoring matrix to identify amino acids matches in the two protein sequences
- This scoring matrix is identical at all positions in the protein sequence!



Alignment match scores

	A	R	N	D	C	Q	E	G	H	I	L	K	M	F	P	S	T	W	Y	V
A	4	-1	-2	-2	0	-1	-1	0	-2	-1	-1	-1	-1	-2	-1	1	0	-3	-2	0
R	-1	5	0	-2	-3	1	0	-2	0	-3	-2	2	-1	-3	-2	-1	-1	-3	-2	-3
N	-2	0	6	1	-3	0	0	0	1	-3	-3	0	-2	-3	-2	1	0	-4	-2	-3
D	-2	-2	1	6	-3	0	2	-1	-1	-3	-4	-1	-3	-3	-1	0	-1	-4	-3	-3
C	0	-3	-3	-3	9	-3	-4	-3	-3	-1	-1	-3	-1	-2	-3	-1	-1	-2	-2	-1
Q	-1	1	0	0	-3	5	2	-2	0	-3	-2	1	0	-3	-1	0	-1	-2	-1	-2
E	-1	0	0	2	-4	2	5	-2	0	-3	-3	1	-2	-3	-1	0	-1	-3	-2	-2
G	0	-2	0	-1	-3	-2	-2	6	-2	-4	-4	-2	-3	-3	-2	0	-2	-2	-3	-3
H	-2	0	1	-1	-3	0	0	-2	8	-3	-3	-1	-2	-1	-2	-1	-2	-2	2	-3
	I	-1	-3	-3	-3	-1	-3	-3	-4	-3	4	2	-3	1	0	-3	-2	-1	-3	-1
	L	-1	-2	-3	-4	-1	-2	-3	-4	-3	2	4	-2	2	0	-3	-2	-1	-2	-1
	K	-1	2	0	-1	-3	1	1	-2	-1	-3	-2	5	-1	-3	-1	0	-1	-3	-2
	M	-1	-1	-2	-3	-1	0	-2	-3	-2	1	2	-1	5	0	-2	-1	-1	-1	1
	F	-2	-3	-3	-3	-2	-3	-3	-3	-1	0	0	-3	0	6	-4	-2	-2	1	-1
	P	-1	-2	-2	-1	-3	-1	-1	-2	-2	-3	-3	-1	-2	-4	7	-1	-1	-4	-3
	S	1	-1	1	0	-1	0	0	-1	-2	-2	0	-1	-2	-1	4	1	-3	-2	-2
	T	0	-1	0	-1	-1	-1	-2	-2	-1	-1	-1	-1	-2	-1	1	5	-2	-2	0
	W	-3	-3	-4	-4	-2	-2	-3	-2	-2	-3	-2	-3	-1	1	-4	-3	-2	11	2
	Y	-2	-2	-2	-3	-2	-1	-2	-3	2	-1	-1	-2	-1	3	-3	-2	-2	2	7
	V	0	-3	-3	-3	-1	-2	-2	-3	-3	3	1	-2	1	-1	-2	-2	0	-3	-1
																				4

TVNGQ--FPGPRIAGVAREGDQVLVKVVNHVAENITIHWGVQLGTGWADGPAYVTQCPI

Sequence profiles

- In reality not all positions in a protein are equally likely to mutate
 - Some amino acids (active sites) are highly conserved, and the score for mismatch must be very high
 - Other amino acids can mutate almost for free, and the score for mismatch should be lower than the BLOSUM score
- Sequence profiles can capture these differences

Evolution has the answer - Sequence profiles

Conserved Non-conserved

Matching any thing
but G => large
negative score

Any thing can match

How to make sequence profiles

1. Align (BLAST) sequence against large sequence database (Swiss-Prot)
2. Select significant alignments and make sequence profile
3. Use profile to align against sequence database to find new significant hits
4. Repeat 2 and 3 (normally 3 times!)

Sequence logos. Visualization of sequence profiles

$$I = \sum_a p_a \log\left(\frac{p_a}{q_a}\right)$$

$$P_A = 6/10 = 0.6$$

$$P_G = 2/10 = 0.2$$

$$P_T = P_K = 1/10 = 0.1$$

$$P_C = P_D = \dots P_V = 0.0$$

$$q_A = 0.07$$

$$q_G = 0.07$$

$$q_T = 0.05$$

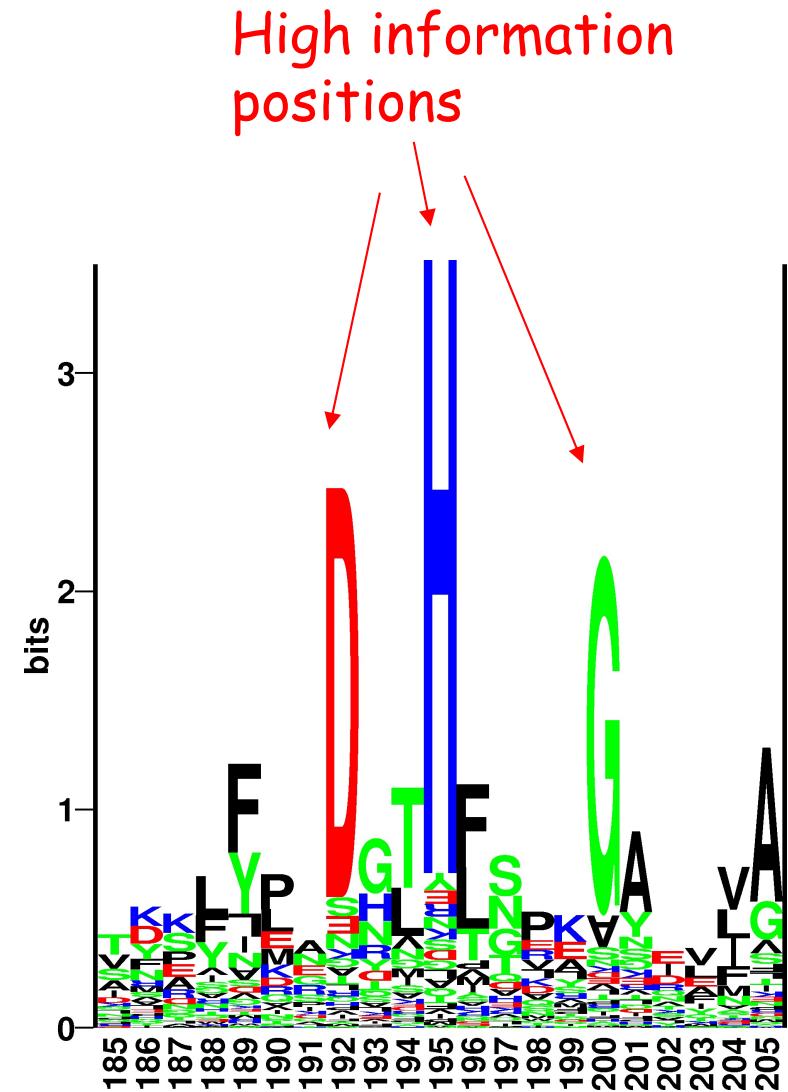
$$q_K = 0.06$$

ALAKAAAAAM
ALAKAAAAAN
ALAKAAAAAR
ALAKAAAAT
ALAKAAAAAV
GMNERPILT
GILGFVFTM
TLNAWVKVV
KLNEPVLLL
AVVPFIVSV

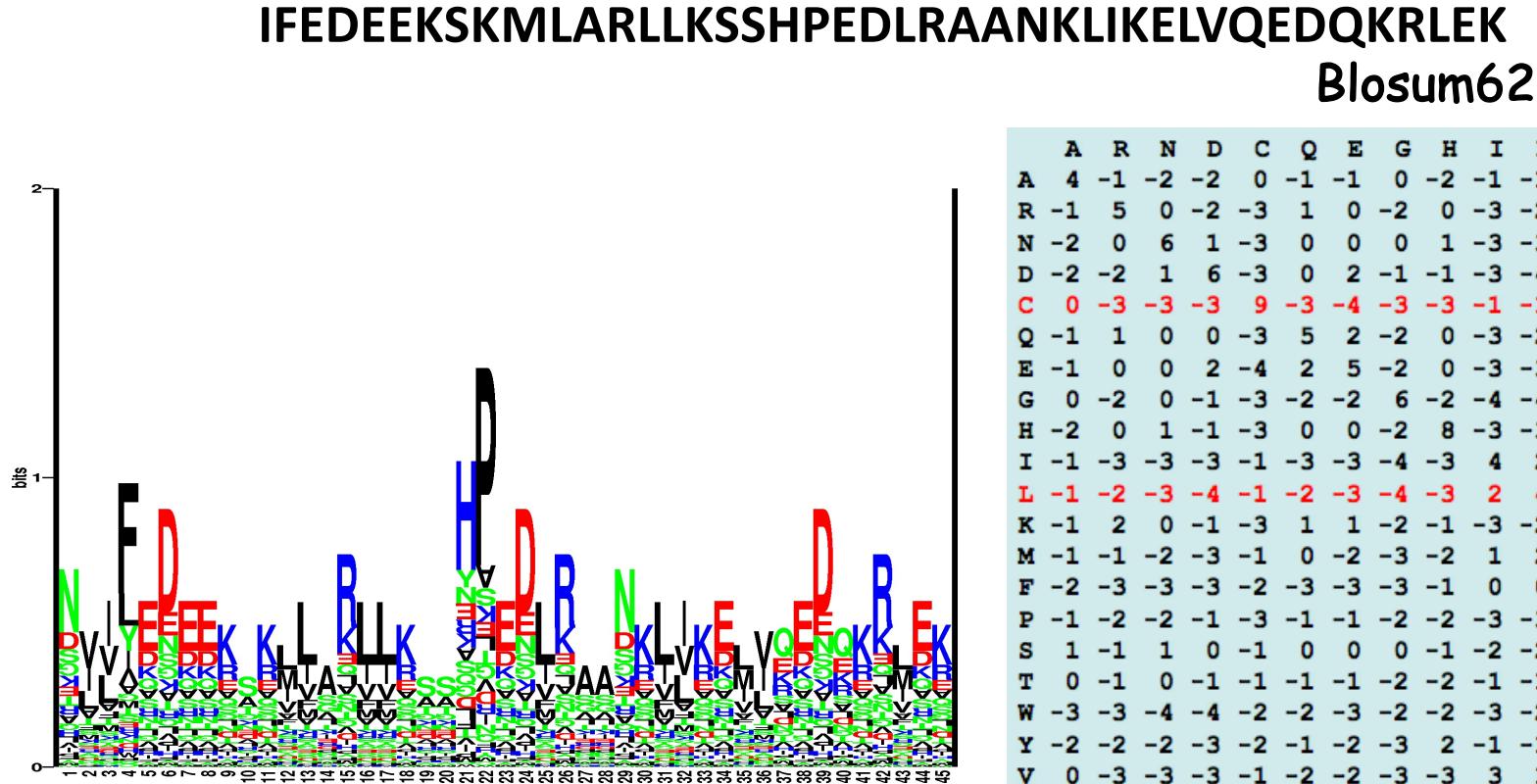
Sequence logos

$$I = \sum_a p_a \log\left(\frac{p_a}{q_a}\right)$$

- Height of a column equal to I
- Relative height of a letter is p
(letters are upside down if $q>p$)



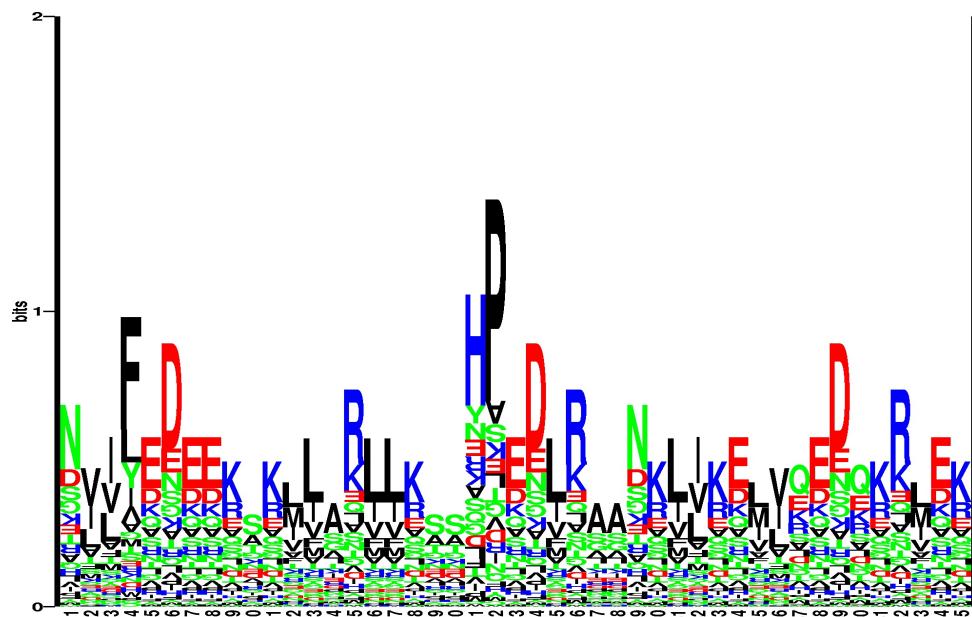
Sequence profiles (1J2J.B)



$$W_{ij} = \log(p_{ij}/q_j)$$

Sequence profiles (1J2J.B)

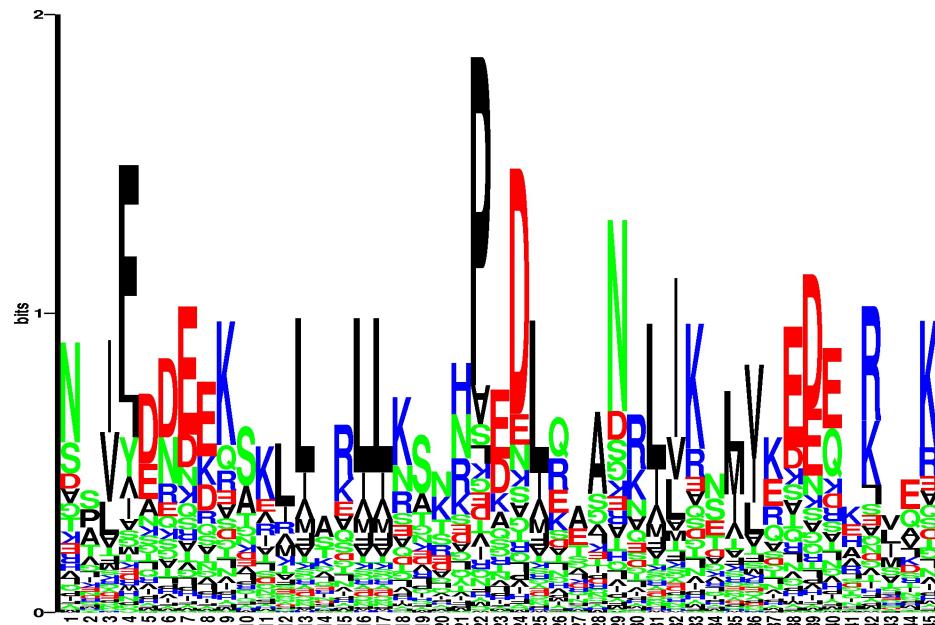
IFEDEEKSKMLARLLKSSHPEDLRAANKLIKELVQEDQKRLEK



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

Sequence profiles (1J2J.B)

IFEDEEKSKMLARLLKSHPEDLRAANKLIKLEVQEDQKRLEK



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

Blast2logo

Blast2logo 1.0 Server

[Instructions](#)

[Output format](#)

SUBMISSION

Paste a single sequence in [FASTA](#) format into the field below:

```
>Ex
VALAELYIPEVARRLGQQGWHEDECTFAEVТИGTRALQAILRDIATWSADEGGMRDGPALVLLPPG
EQHTLGMVAVAKLRRLGVSCLRMSTGPTELFGKRRFDAIMISLAHAEMLEVGRKLVKTLKD
MTGGRIPVAMGGALFLDGTEAASIPEADIVTNDIEALQ
```

Submit a file in [FASTA](#) format directly from your local disk:

Choose File no file selected

Upload a file in [BLAST PROFILE](#) format:

Choose File no file selected

Blast Database SP

Number of Blast iterations

Blast E-value cutoff

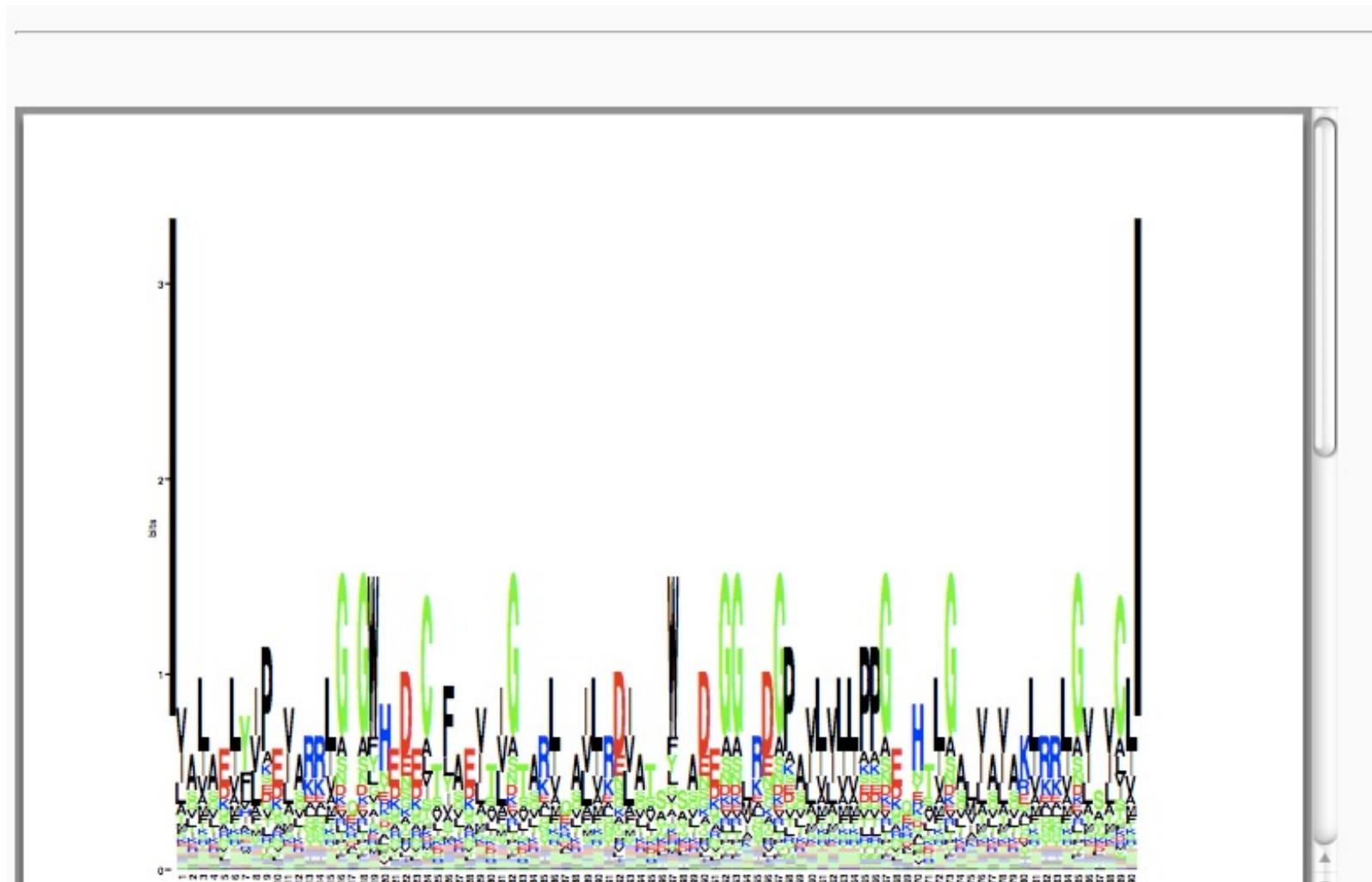
Stack Linesize

Plot Kullback-Leibler logo

File format for logo file PDF

Submit Clear fields

Blast2logo



Download logo file [Logo](#)

Link to Blastprofile output file [Blast.prof](#)

Blast2logo

CENTERFO
R BIOLOGI
CALSEQU
ENCEANA
LYSIS CBS

Last position-specific scoring matrix computed

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

A	R	N	D	C	Q	E	G	H	I	L	K	M	F	P	S	T	W	Y	V	
A	4	-1	-2	-2	0	-1	-1	0	-2	-1	-1	-1	-2	-1	1	0	-3	-2	0	
R	-1	5	0	-2	-3	1	0	-2	0	-3	-2	2	-1	-3	-2	-1	-1	-3	-2	-3
N	-2	0	6	1	-3	0	0	0	1	-3	-3	0	-2	-3	-2	1	0	-4	-2	-3
D	-2	-2	1	6	-3	0	2	-1	-1	-3	-4	-1	-3	-3	-1	0	-1	-4	-3	-3
C	0	-3	-3	-3	9	-3	-4	-3	-3	-1	-1	-3	-1	-2	-3	-1	-1	-2	-2	-1
Q	-1	1	0	0	-3	5	2	-2	0	-3	-2	1	0	-3	-1	0	-1	-2	-1	-2
E	-1	0	0	2	-4	2	5	-2	0	-3	-3	1	-2	-3	-1	0	-1	-3	-2	-2
G	0	-2	0	-1	-3	-2	-2	6	-2	-4	-4	-2	-3	-3	-2	0	-2	-2	-3	-3
H	-2	0	1	-1	-3	0	0	-2	8	-3	-3	-1	-2	-1	-2	-1	-2	-2	2	-3
I	-1	-3	-3	-3	-1	-3	-3	-4	-3	4	2	-3	1	0	-3	-2	-1	-3	-1	3
L	-1	-2	-3	-4	-1	-2	-3	-4	-3	2	4	-2	2	0	-3	-2	-1	-2	-1	1
K	-1	2	0	-1	-3	1	1	-2	-1	-3	-2	5	-1	-3	-1	0	-1	-3	-2	-2
M	-1	-1	-2	-3	-1	0	-2	-3	-2	1	2	-1	5	0	-2	-1	-1	-1	-1	1
F	-2	-3	-3	-3	-2	-3	-3	-3	-1	0	0	-3	0	6	-4	-2	-2	1	3	-1
P	-1	-2	-2	-1	-3	-1	-1	-2	-2	-3	-3	-1	-2	-4	7	-1	-1	-4	-3	-2
S	1	-1	1	0	-1	0	0	0	-1	-2	-2	0	-1	-2	-1	4	1	-3	-2	-2
T	0	-1	0	-1	-1	-1	-2	-2	-1	-1	-1	-1	-2	-1	1	5	-2	-2	0	
W	-3	-3	-4	-4	-2	-2	-3	-2	-2	-3	-2	-3	-1	1	-4	-3	-2	11	2	-3
Y	-2	-2	-2	-3	-2	-1	-2	-3	2	-1	-1	-2	-1	3	-3	-2	-2	2	7	-1
V	0	-3	-3	-3	-1	-2	-2	-3	-3	3	1	-2	1	-1	-2	-2	0	-3	-1	4

Blast2logo

Blast2logo 1.0 Server

[Instructions](#)

[Output format](#)

SUBMISSION

Paste a single sequence in [FASTA](#) format into the field below:

```
>Ex
VALAELYIPEVARRLGQQGWHEDECTFAEVТИCTARLQAILRDIATWSADEGGMRDGPALVLLPPG
EQHTLGAMAVAKLRLGVSVCLRMSTGPAELRELFKRRFDAIMISLAHAEMLLEVGRKLVKTLKD
MTGGRIPVAMGGALFLDGTEAASIPEADIVTN DIEAALQ
```

Submit a file in [FASTA](#) format directly from your local disk:

no file selected

Upload a file in [BLAST PROFILE](#) format:

no file selected

Blast Database

Number of Blast iterations

Blast E-value cutoff

Stack Linesize

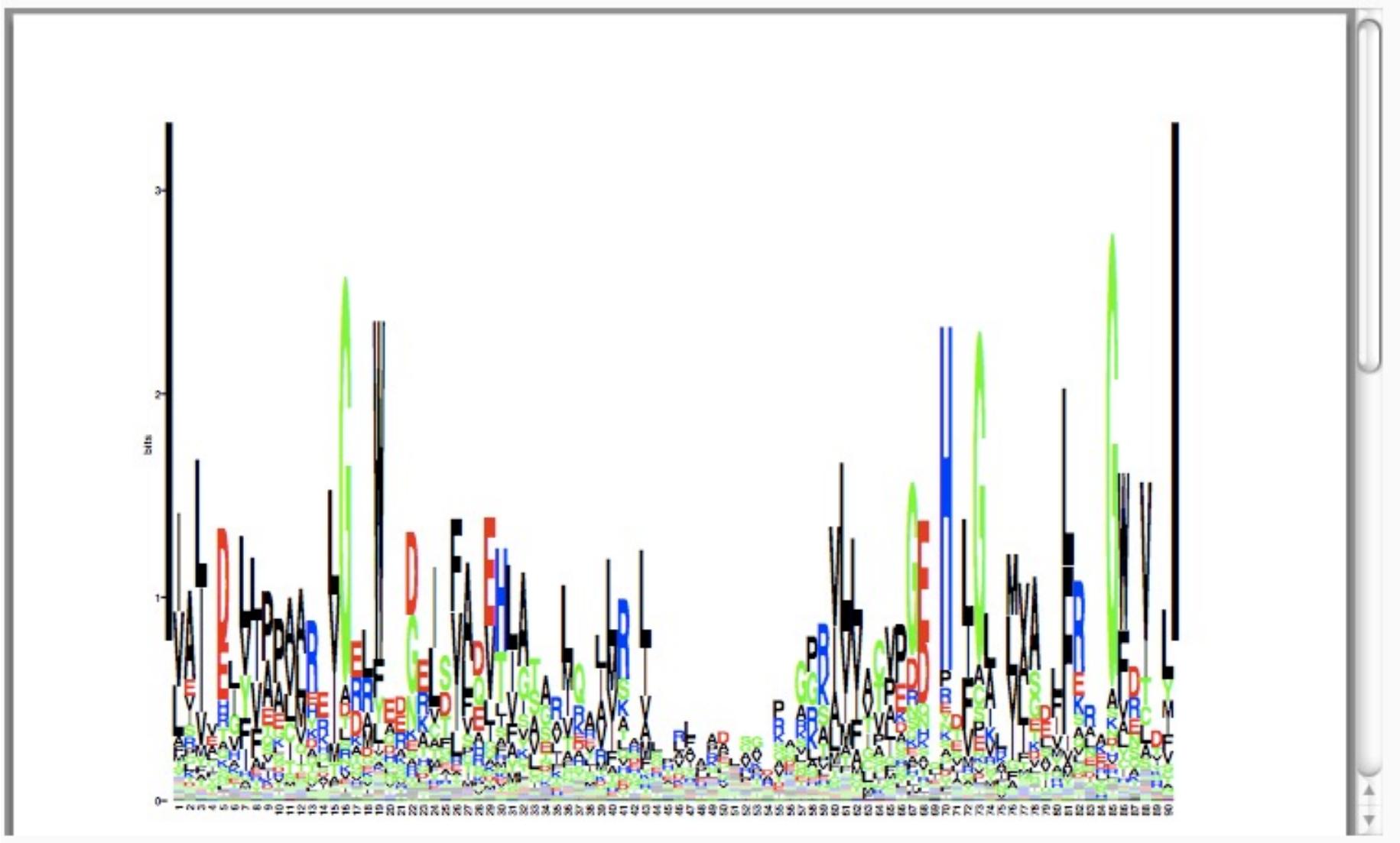
Plot Kullback-Leibler logo

File format for logo file

Restrictions:

At most 1 sequences per submission; each sequence not more than 20,000 amino acids.

Blast2logo

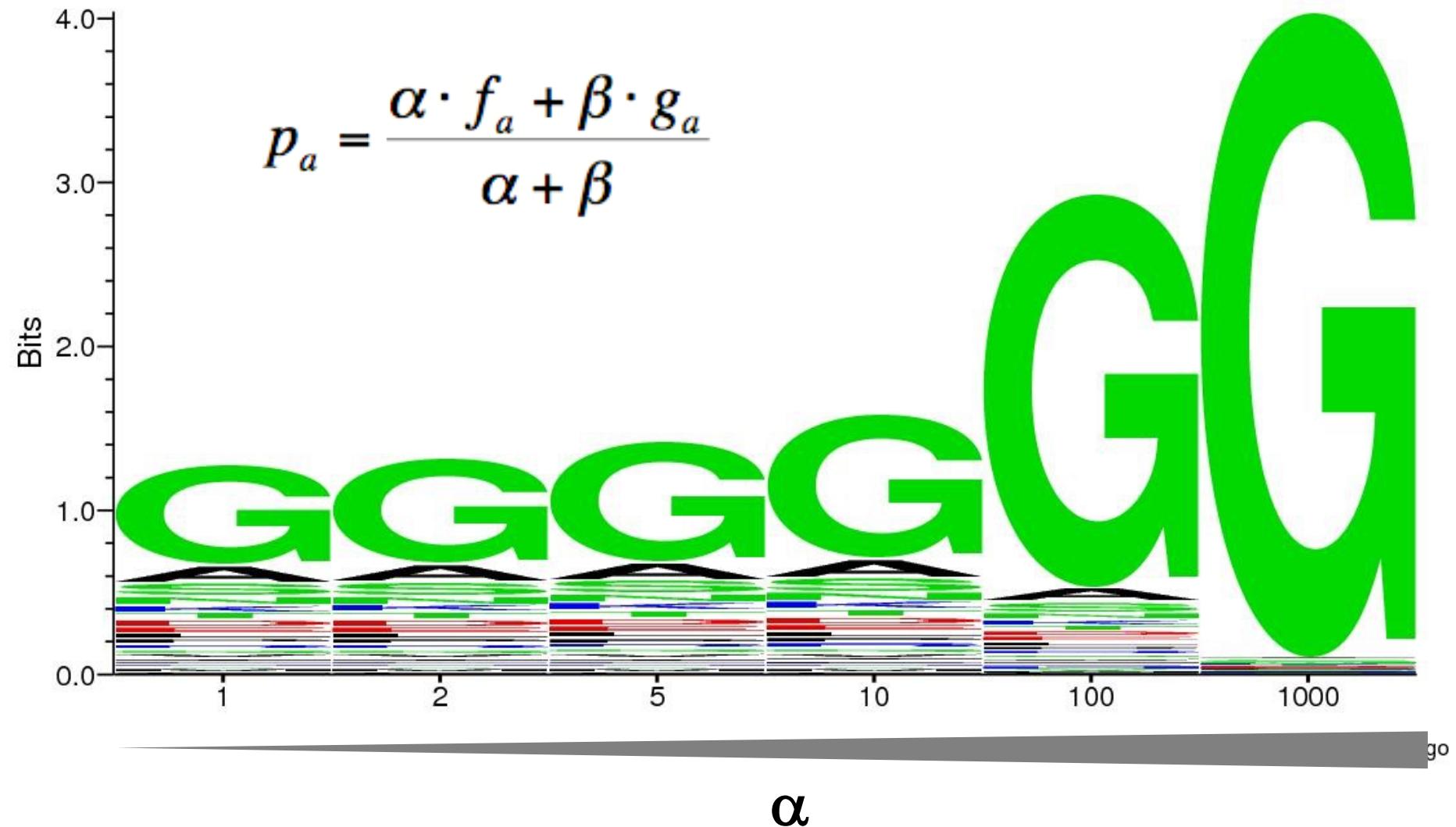


Blast2logo

Last position-specific scoring matrix computed,

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

Sequence profiles or Gaining confidence



Example.

>1K7C.A

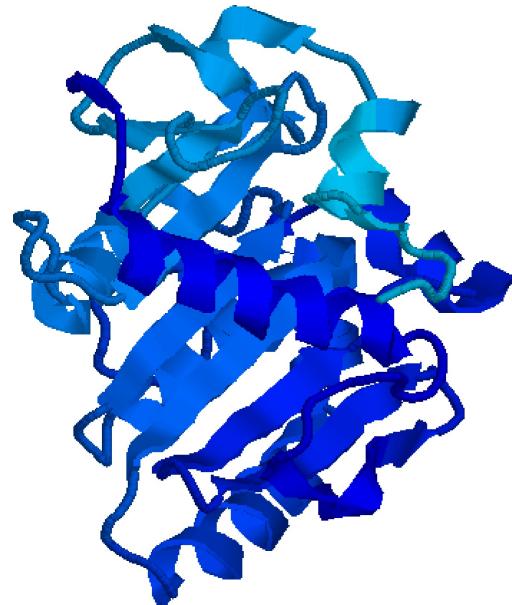
TTVYLAGDSTMAGNGGGSGTNGWGEYLASYLSATVVNDAVAGRSARSYTREGRFENIADV
VTAGDYVIVEFGHNDGGSLSTDNGRTDCSGTGAEVCYSVYDGVNETILTFPAYLENAAKL
FTAKGAKVILSSQTPNNPWETGTFVNSPTRFVEYAAEVAGVEYVDHWSYVDSIYETL
GNATVNSYFPIDHTHTSPAGAEVVAEFLKAVVCTGTSLKSVLTTSFEGTCL

- What is the function
 - Where is the active site?
-

What would you do?

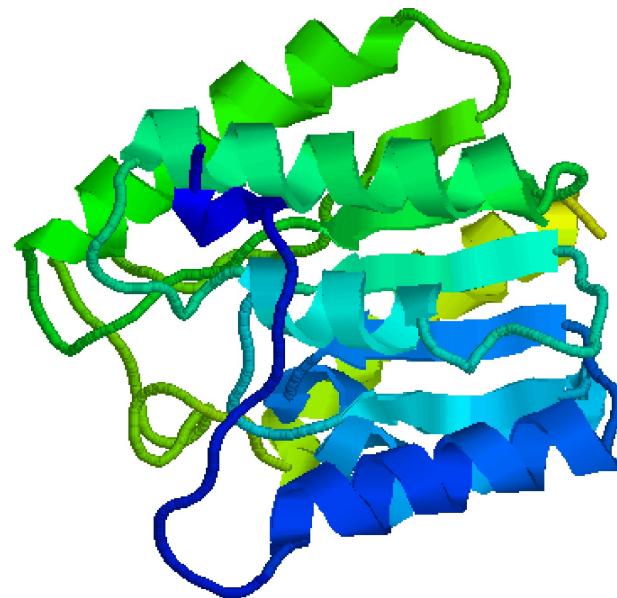
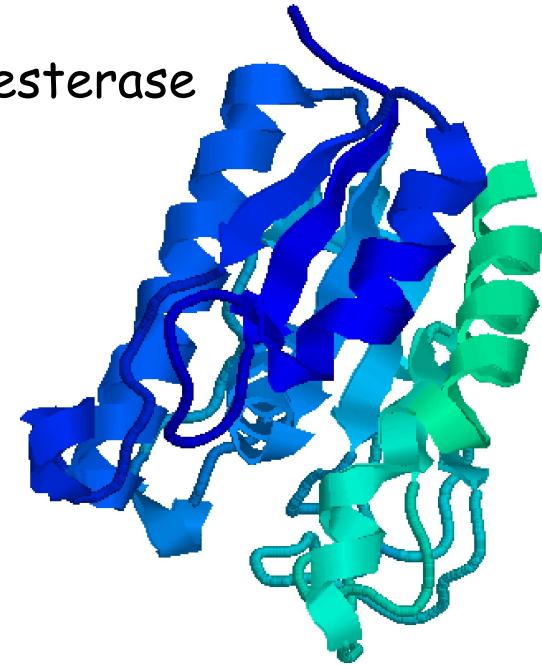
- Function
 - Run Blast against PDB
 - No significant hits
 - Run Blast against NR (Sequence database)
 - Function is Acetyl esterase?
- Where is the active site?

Example. Where is the active site?



1USW Hydrolase

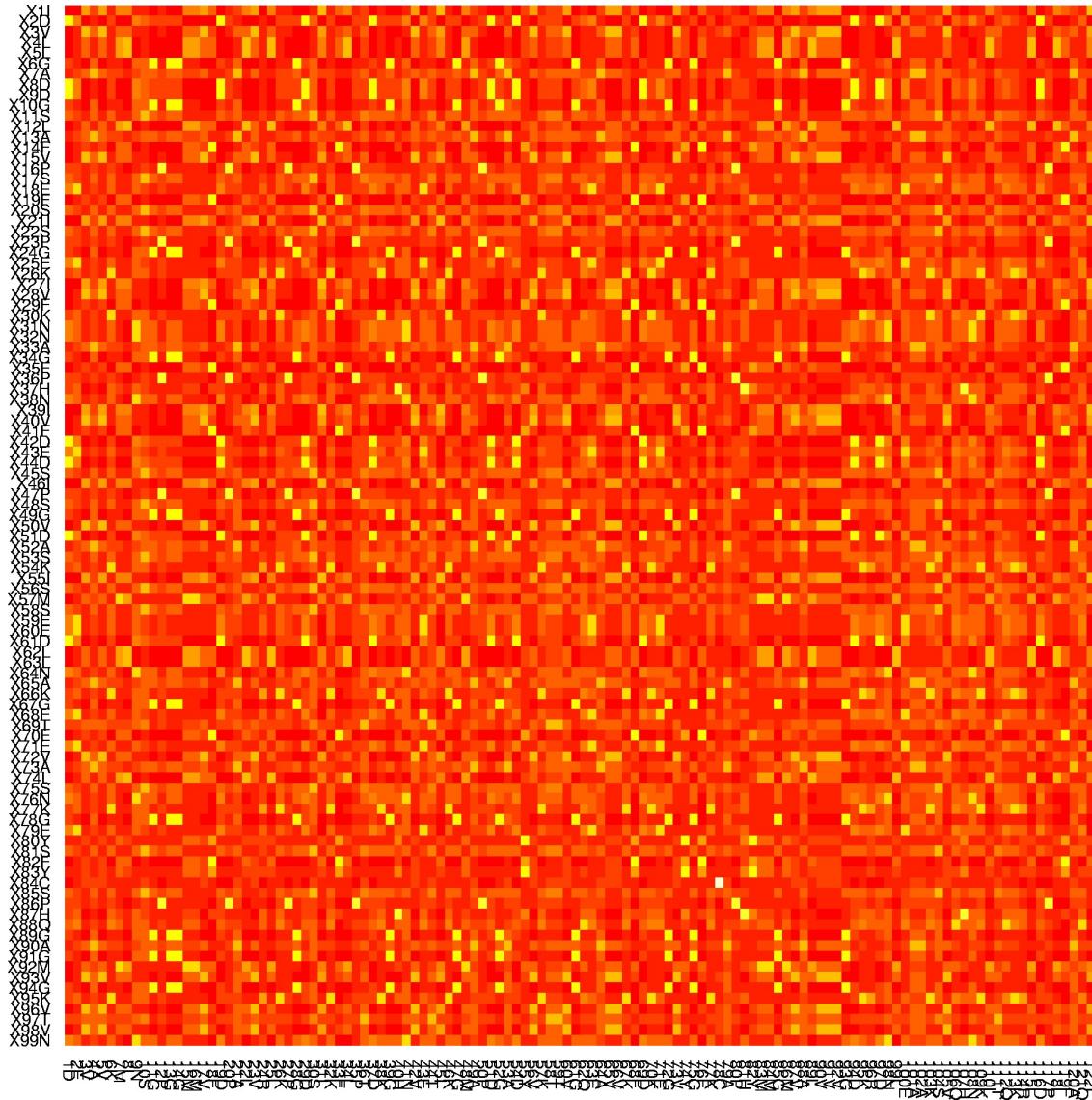
1G66 Acetylxyran esterase



1WAB Acetylhydrolase

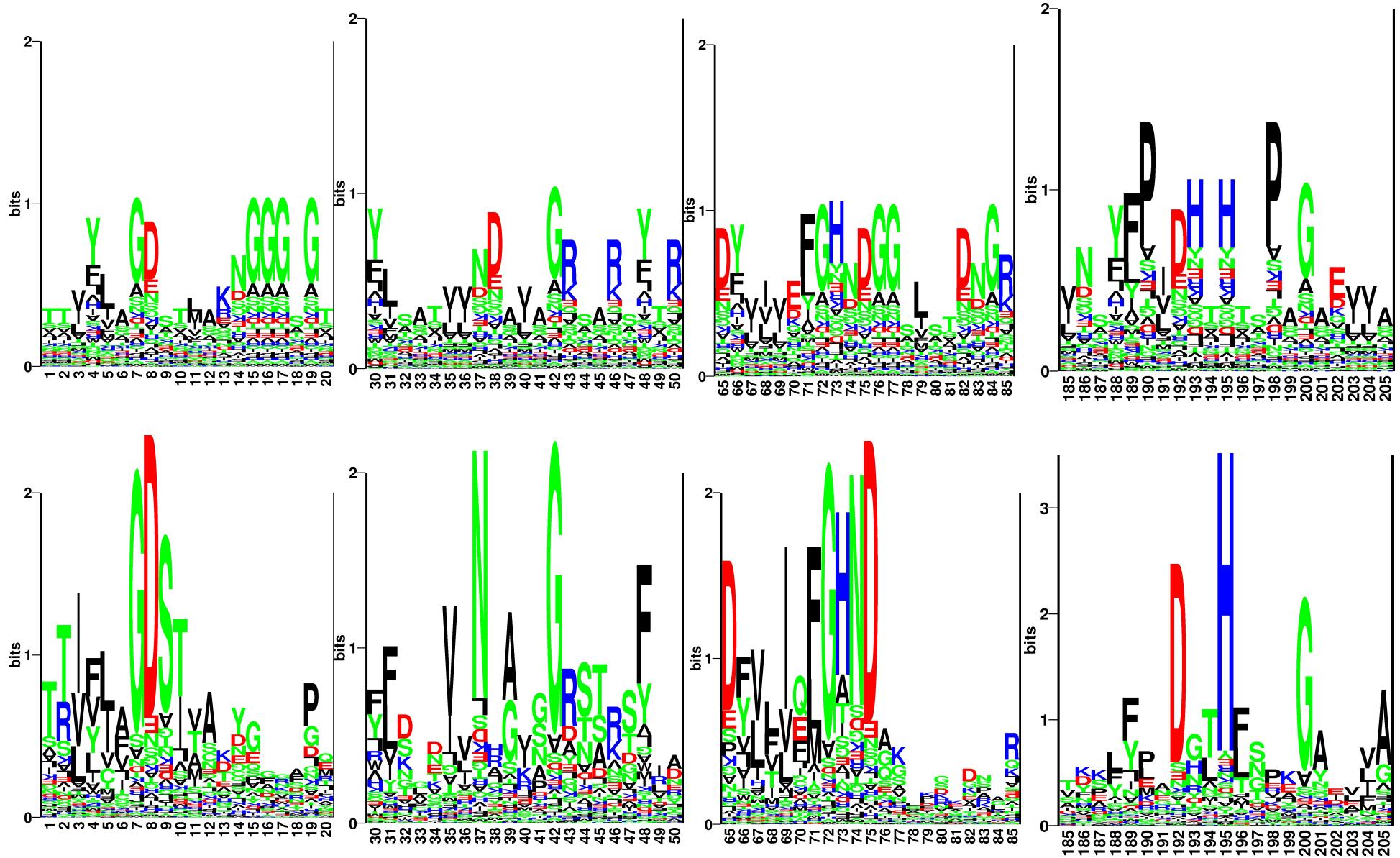
When Blast fails!

1K7A.A



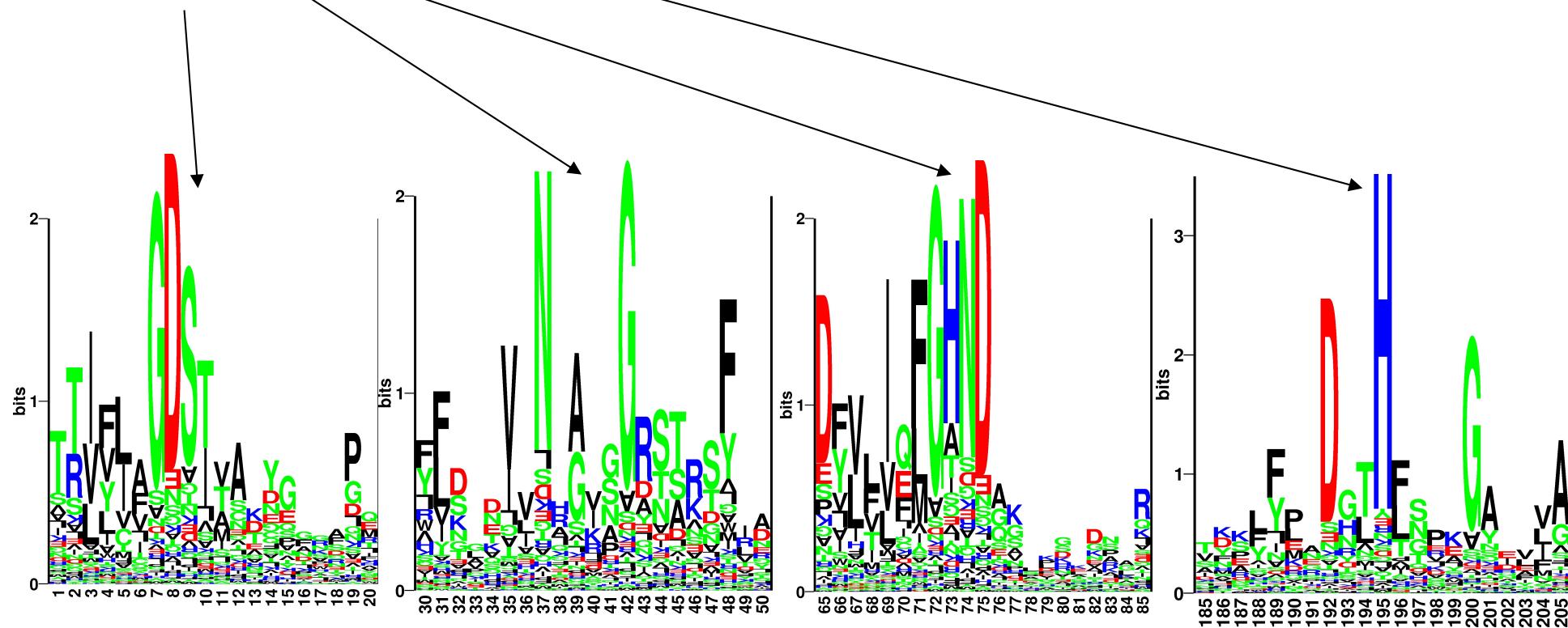
1WAB._

Example. (SGNH active site)

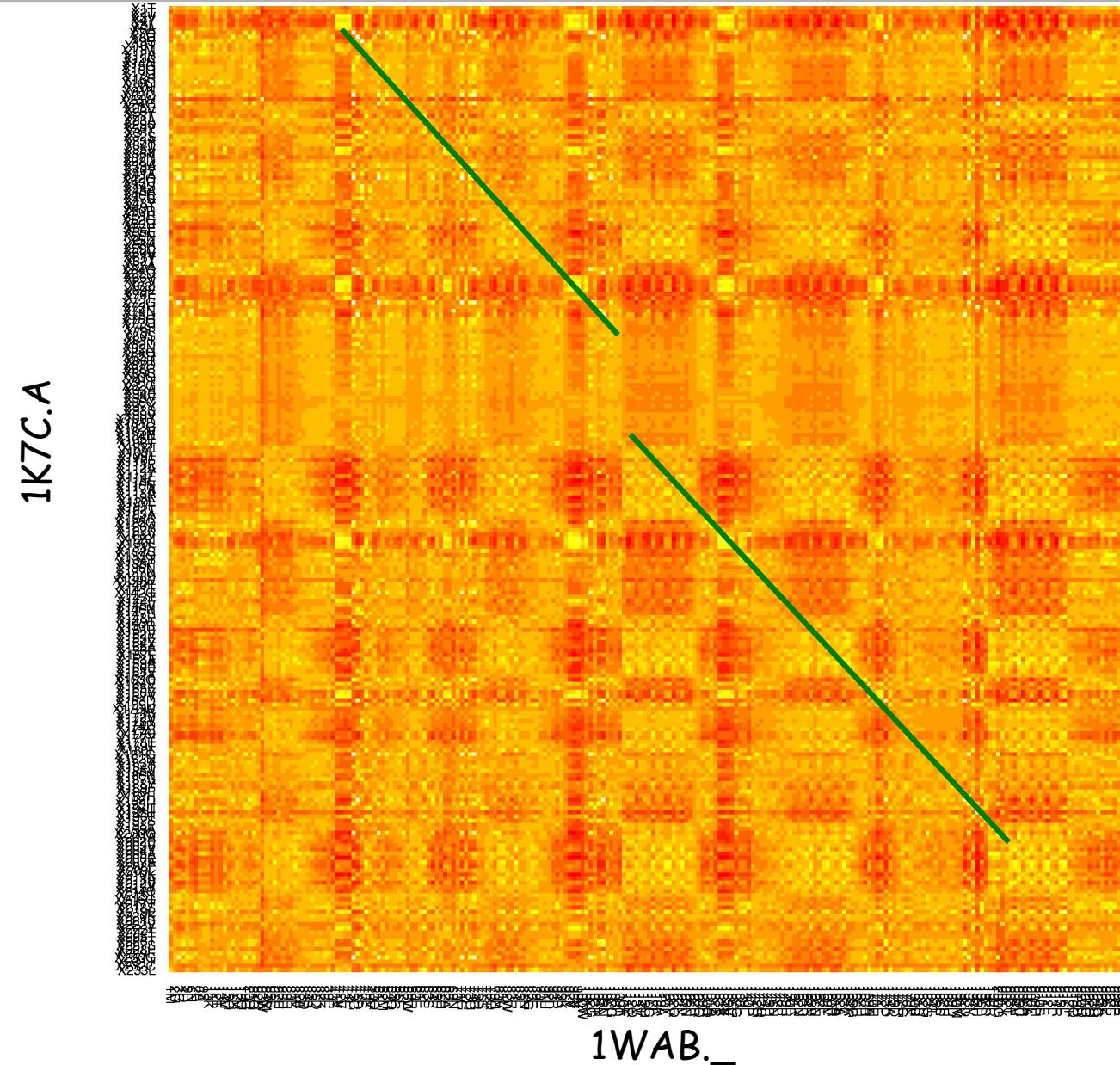


Example. Where is the active site?

- Sequence profiles might show you where to look!
- The active site could be around
 - S9, G42, N74, and H195



Profile-profile scoring matrix



Example. Where is the active site?

Align using sequence profiles

ALN 1K7C.A 1WAB._ RMSD = 5.29522. 14% ID

1K7C.A TVYLAGD**S**TMAKNGGGSGTNGW**G**EYLASYLSATVVNDAV**A**GRSARSYTREGRFENIADVVTAGDYVIVEFGH**N**DGGSLSTDN

1WAB._ EVVFIGD**S**LVQLMHQCE---IWRELFS---PLHALNFGIG**G**DSTQHVLW--RLEN~~GELEHIRPKIVVVWVG~~**TNNHG**-----

1K7C.A GRTDCSGTGAEV**C**YSVYDG**V**NETILT**F**PAYLENAAKLFTAK--GAKVILSSQT**P**NNPWE**T**GTFVNSPTRFVEYAEL-AAEVA

1WAB._ -----HTAEQVTGGIKAIVQLVNERQPQARVVVLGLLPRGQ-HPNPLREKNRRVNELVRAALAGHP

1K7C.A GVEYVDHSYVDSIYETLGNATVNSYFPIDHT**H**TSPAGAEVVAEAFLKAVVCTGTSL

1WAB._ RAHFLDADPG---FVHSDG--TISHHD**M**YDYL**H**LSRLGYTPVCRALHSLLLRL---L

Handout exercise

Using Psi-Blast Profiles

Sequence profiles take home message

- Blast will often fail to recognize sequence relationships for low homology sequence pairs
- Sequence profiles contain information on conserved/variable residues in a protein sequence
- Sequence profiles are calculated from (multiple) sequence alignments
- Iterative Blast enables homology recognition also for low sequence similarity
- Sequence profiles give information on residues essential for protein function and protein structure
- Can be used to predict impact of SNP's on protein function
 - This is often done using the Blosum matrix, but profiles are much more precise