

ВВЕДЕНИЕ В БИОИНФОРМАТИКУ

Лекция №6

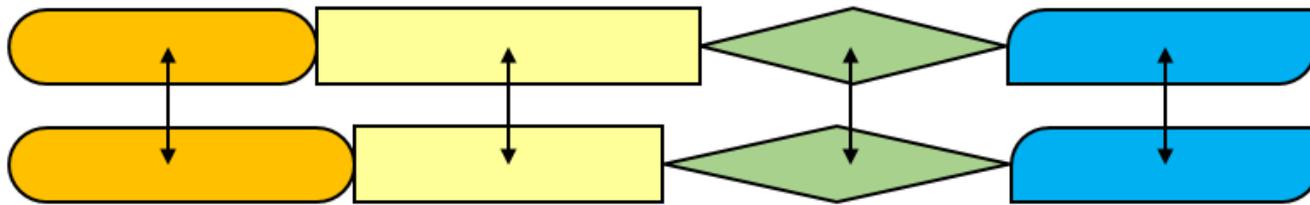
Значимость выравнивания. Экспресс-сравнение последовательностей (BLAST). Построение и визуализация профилей множественных выравниваний.

Новоселецкий Валерий Николаевич
к.ф.-м.н., доц. каф. биоинженерии
valery.novoseletsky@yandex.ru

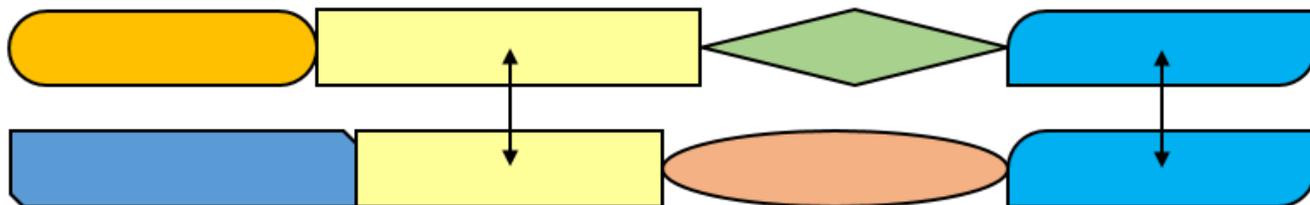
Сайт курса <http://intbio.org/bioinf2020-21>

Расчет выравнивания двух последовательностей

- Алгоритм Нидлмана-Вунша (1970) для глобального выравнивания
- Алгоритм Смита-Уотермана (1981) для локального выравнивания



Global Alignment



Local Alignment

Значимость выравнивания

Насколько значимо полученное выравнивание?

Имеет ли оно биологический смысл или образовалось случайно?

```
SEQUENCE 1 VLSAADKTNVKAAWSKVGGHAGEYGAEALERMFLGFPTTKTYFPHFDSLH 50
          |||.|||||||.|||.|||||||.|||||||.|||||||.
SEQUENCE 1 VLSPADKTNVKAAWGKVGHAHAGEYGAEALERMFLSFPTTKTYFPHFDSLH 50

SEQUENCE 51 GSAQVKAHGKKVADGLTLAVGHLDDLPGALSDSLNLHAHKLRVDPVNFKL 100
          |||||.|||||||.||.||. |:|:|.|||. |:|||||||.
SEQUENCE 51 GSAQVKGHGKKVADALTNAVAHVDDMPNALSALSDLHAHKLRVDPVNFKL 100

SEQUENCE 101 LSHCLLSTLAVHLPNDFTPAVHASLDKFLSSVSTVLTSKYR 141
          |||||.|||.|||. :|||||||. :|||||||.
SEQUENCE 101 LSHCLLVTLAAHLPAEFTPAVHASLDKFLASVSTVLTSKYR 141
```

```
# Matrix: EBLOSUM62
# Gap_penalty: 10.0
# Extend_penalty: 0.5
#
# Length: 141
# Identity: 123/141 (87.2%)
# Similarity: 128/141 (90.8%)
# Gaps: 0/141 ( 0.0%)
```

Значимость выравнивания

```
SEQUENCE 1 -VLSEGEWQLVLHVWAKVEADVAGHGQDILIRLFKSHPETLEKFDRFKHL 49
          .|:|.:.|..|..|..|:..|.....|.:.:.|... |.|..|
SEQUENCE 1 GALTESQAALVKSSWEEFNANIPKHTHRFFILVLEIAPAAK---DLFSFL 47
SEQUENCE 50 KTEAEM-KASEDLKKHGVTV-----LTALGAILKKKGHHEAELKP 88
          |.:|: :.:.|:..|...| |...|.:.: :|.|.|.
SEQUENCE 48 KGTSEVPQNNPELQAHAGKVFKLVEEAAIQLEVTGVVVT----DATLKN 92
SEQUENCE 89 LAQSHATKHKIPIKYLEFISEAIIHVLHSRHPGDFGADAQGAMNKALELF 138
          |...|.:.| .:..:..:..|:|:..:.. ..||.....:|.|...:
SEQUENCE 93 LGSVHVSK-GVADAHFPVVKEAILKTIKE---VVGAKWSEELNSAWTIA 137
SEQUENCE 139 RKDIAAKY-KELGYQG 153
          ..:|.:. |:....
SEQUENCE 138 YDELAIVIKKEMDDAA 153
```

```
# Matrix: EBLOSUM62
# Gap_penalty: 10.0
# Extend_penalty: 0.5
#
# Length: 166
# Identity: 35/166 (21.1%)
# Similarity: 62/166 (37.3%)
# Gaps: 26/166 (15.7%)
```

Значимость выравнивания



Кашалот



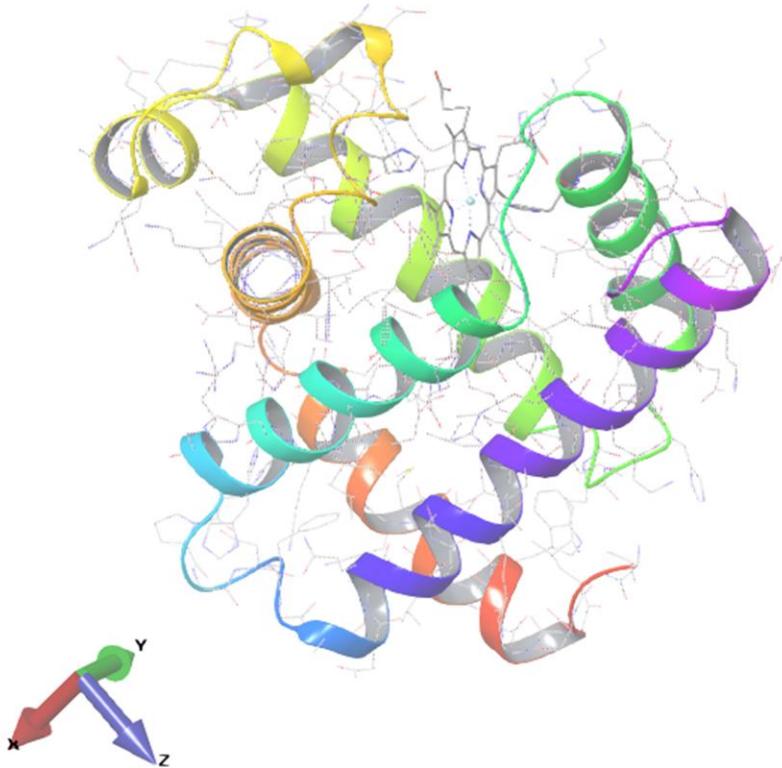
Люпин

Последний общий предок жил около 1 млрд лет назад!

Что общего между ними сейчас?

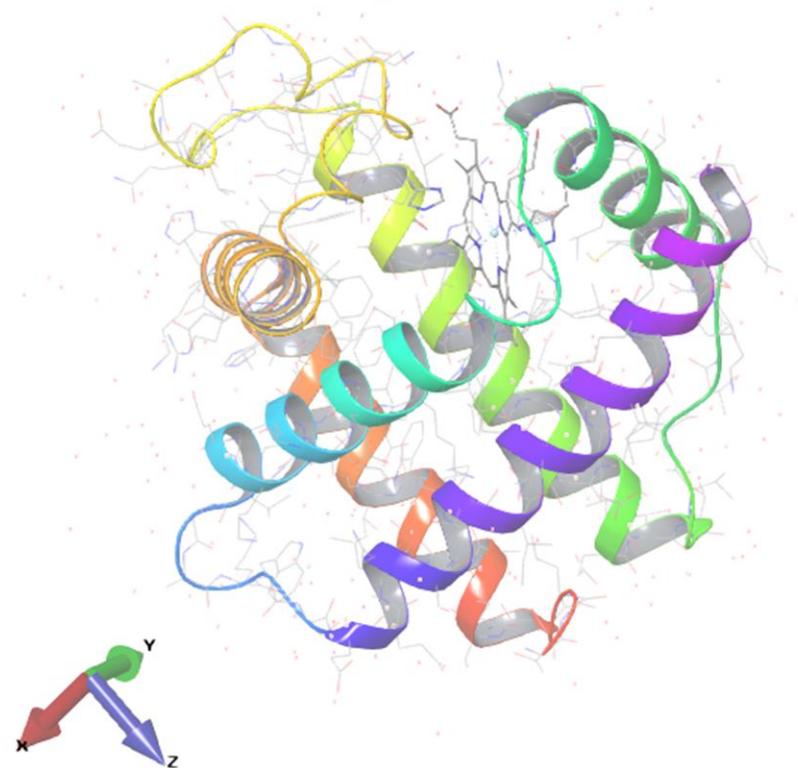
Значимость выравнивания

Title: 1MBN
PDB ID: 1MBN



Миоглобин кашалота (1mbn, 1969)

Title: 1GDJ
PDB ID: 1GDJ



Леггемоглобин люпина (1gdj, 1995) (ИК РАН)

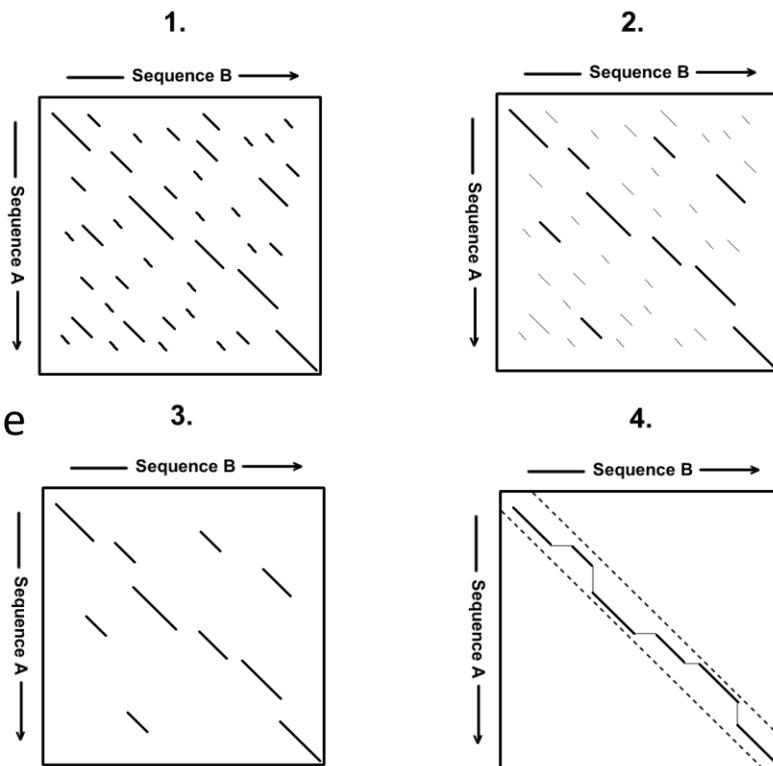
Экспресс-методы сравнения последовательностей.

FASTA

FASTA (Lipman and Pearson, 1985, Pearson and Lipman, 1988) предназначен для сравнения новых последовательностей с уже содержащимися в базах данных.

Алгоритм:

- Поиск идентичных k -буквенных слов ($k = 2$ для белков, 4-6 для нуклеотидов, 1 для коротких последовательностей) между последовательностями;
- Расширение точек идентичности до областей сходства с использованием матриц замен (**BLOSUM50** для белков, **единичная** для нуклеотидов);
- Переоценка выявленных областей сходства с помощью целевой матрицы замен, определение 10 областей с максимальной оценкой;
- Перебор возможных вариантов построения выравнивания с использованием этих областей и выбор оптимального с учётом оценок областей и штрафов за вставки;



FASTA и FASTQ форматы

Исходно – формат представления последовательностей для программы FASTA.

```
>sp|P69905|HBA_HUMAN Hemoglobin subunit alpha OS=Homo sapiens...
MVLSPADKTNVKAAWGKVGANHAGEYGAEALERMFLSFPTTKTYFPHFDLSHGSAQVKGHG
KKVADALTNAVAVHDDMPNALSALSDLHANHLR...
```

FASTQ - обобщение формата FASTA.

```
@SEQ_ID
GATTTGGGGTTCAAAGCAGTATCGATCAAATAGTAAATCCATTTGTTCAACTCACAGTTT
+
!''*(((((***+))%%%)++)(%%%) .1***-+*''))**55CCF>>>>>CCCCCCC65
```

Качество прочтения $Q = -10 \log_{10} e$, где e – вероятность **ошибочного** прочтения

94 градации в формате **ASCII**:

```
!"#$%&'()*+,-./0123456789:;<=>?@ABCDEFGHIJKLMNPO
QRSTUVWXYZ[\]^_`abcdefghijklmnopqrstuvwxyz{|}~
```

American Standard Code for Information Interchange (1967 - 2007)

ASCII Code Chart

	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
0	NUL	SOH	STX	ETX	EOT	ENQ	ACK	BEL	BS	HT	LF	VT	FF	CR	SO	SI
1	DLE	DC1	DC2	DC3	DC4	NAK	SYN	ETB	CAN	EM	SUB	ESC	FS	GS	RS	US
2		!	"	#	\$	%	&	'	()	*	+	,	-	.	/
3	0	1	2	3	4	5	6	7	8	9	:	;	<	=	>	?
4	@	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
5	P	Q	R	S	T	U	V	W	X	Y	Z	[\]	^	_
6	`	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o
7	p	q	r	s	t	u	v	w	x	y	z	{		}	~	DEL

128 символов ($256 = 2^8$ в расширенной версии), так что для кодирования необходим **1 байт на 1 символ** (команду или знак).

Кодировка ASCII заменена на Unicode (в настоящее время содержит 137 439 символов, охватывающих 146 современных и исторических шрифтов, математические символы, эмодзи и т.д.)

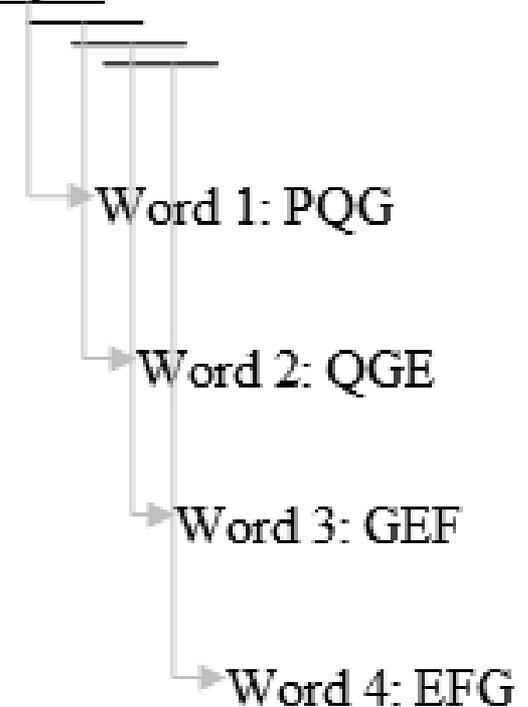
Экспресс-методы сравнения последовательностей. BLAST

BLAST - **B**asic **L**ocal **A**lignment **S**earch **T**ool (Altschul et al.,1990) предназначен для сравнения новых последовательностей с уже содержащимися в базах данных.

Алгоритм:

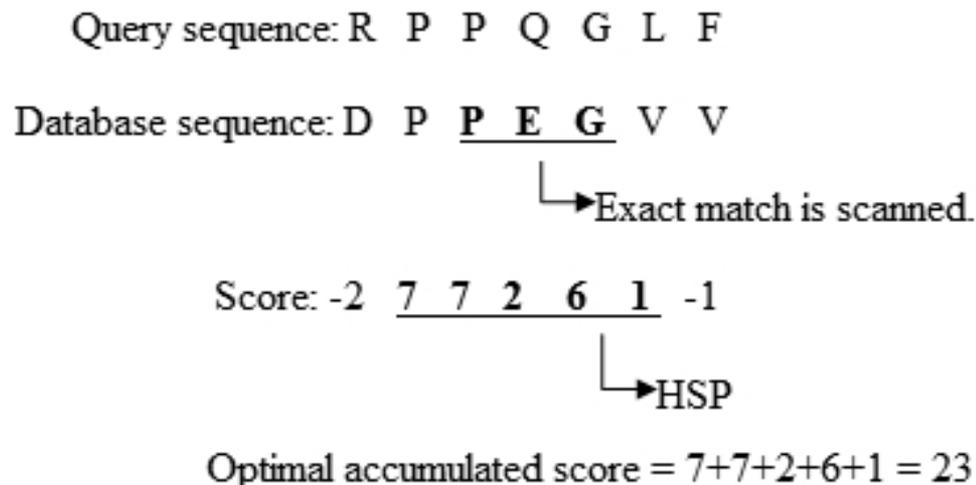
- Удаление малоинформативных участков последовательности (повторы и т.п.);
- Составление списка k -буквенных слов (K-tuple), присутствующих в последовательности запроса;
- Сопоставление этих слов со всеми возможными словами длины k и оценка сходства; отбор слов с оценкой, превышающей пороговую (например, для слова PQG сходными будут PNG, PEG и PDG, но не PQW)
- Сканирование последовательности из БД и поиск в ней слов с высокой оценкой, полученных на предыдущем шаге;

Query sequence: PQGEFG

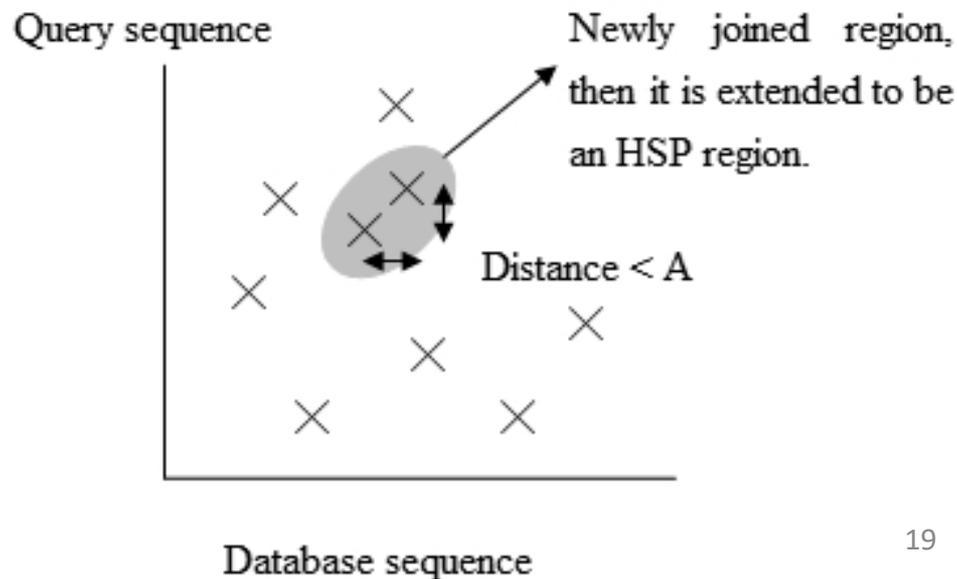


Экспресс-методы сравнения последовательностей. BLAST

- **Расширение локальных выравниваний** в обе стороны до тех пор, пока суммарная оценка выравнивания не начинает уменьшаться (построение сегментных пар (high-scoring segment pair, **HSP**));



- **Объединение сегментных пар**, лежащих на удалении меньше A;
- **Составление списка сегментных областей с высокой оценкой**;
- **Расчет статистической значимости этих оценок.**



Бросание кубика

Рассмотрим бросание кубика n раз.
Каково математическое ожидание
числа N_l повторов длины не менее l ?

$$E(N_l) = (n - l + 1)(1 - p)p^l \cong n(1 - p)p^l,$$

где p — вероятность выпадения грани



В случае наиболее длинного повтора $N = 1$. Тогда для
мат. ожидания длины L этого повтора справедливо

$$1 \cong n(1 - p)p^{E(L)} \Rightarrow E(L) = \log_{1/p} n(1 - p)$$

(Erdős, Rényi, 1970)

*«Математик — это автомат
по переработке кофе в теоремы»*



Альфред Реньи
(1921 – 1970)

Бросание кубика

Точечная матрица – двумерное обобщение бросания кубика

с заменой вероятности выпадения грани на вероятность совпадения букв.

Для последовательностей длиной n и m имеем

$$E(N) = (n - l + 1)(m - l + 1)(1 - p)p^l \cong mn(1 - p)p^l,$$

где p – вероятность совпадения

Более точное рассмотрение дает для мат. ожидания длины двумерного повтора выражение

$$E(L) = \log_{1/p}(mn) + \log_{1/p}(1 - p) + \gamma \log_{1/p} e - \frac{1}{2}$$

где «гамма» (постоянная Эйлера) $\approx 0,577215\dots$

$$\gamma = \lim_{n \rightarrow \infty} \left(\sum_{k=1}^n \frac{1}{k} - \ln n \right) = \lim_{n \rightarrow \infty} \left(1 + \frac{1}{2} + \frac{1}{3} + \dots + \frac{1}{n} - \ln n \right)$$

(1735 – Эйлер – 5 знаков, ... 1973 – Уотерман – 4879 знаков, ...)

Бросание кубика

Переходя к натуральному логарифму и вводя обозначения

$$\log_{1/p}(1-p) + \gamma \log_{1/p} e - \frac{1}{2} \equiv \log_{1/p} K, \quad \ln \frac{1}{p} \equiv \lambda$$

Получаем

$$E(L) = \frac{\ln(Kmn)}{\lambda}$$

Аналогичное справедливо ([Karlin, Altschul, 1990?](#)) и для максимальной оценки $S_{n,m}$ сегментной пары (**HSP**), образованной последовательностями длиной n и m :

$$S_{n,m} \propto \frac{\ln(Kmn)}{\lambda}$$

Нормализация (для исключения вклада величин n и m):

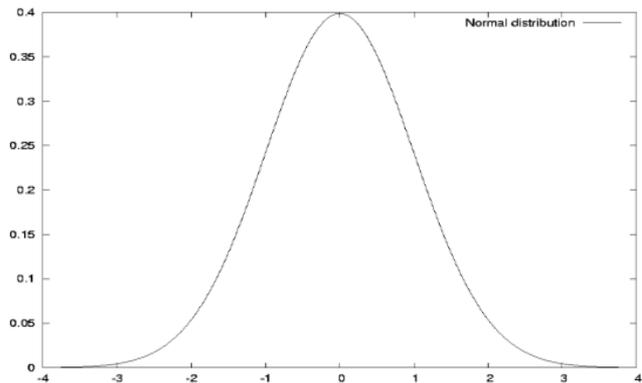
$$\tilde{S}_{n,m} = S_{n,m} - \frac{\ln(Knm)}{\lambda}$$

BLAST. Значимость выравнивания

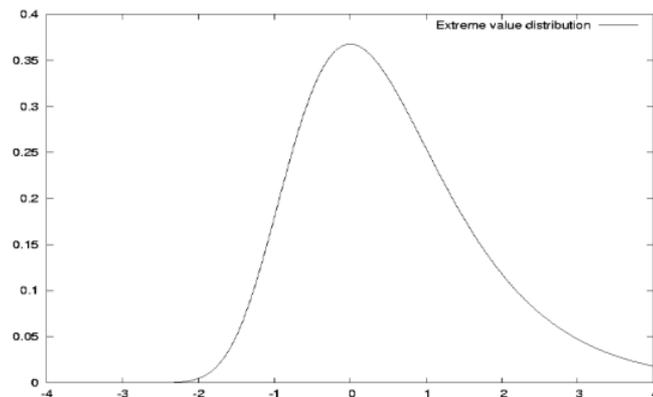
Распределение нормализованных максимальных оценок $\tilde{S}_{n,m}$ подчиняется распределению Гумбеля ([распределению экстремальных значений](#), [Gumbel, 1937](#); [Гнеденко, 1943](#)), для которого

$$P(\tilde{S}_{n,m} > S) \approx 1 - \exp(-K m n e^{-\lambda S}) \approx K m n e^{-\lambda S}$$

где S – случайная величина в интервале $(-\infty; +\infty)$ (но [последнее приближение действительно только для больших \$S\$](#) , когда $K m n * \exp(-\lambda S)$ мало!)



$$\varphi(x) = \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{x^2}{2}}$$



$$\varphi(x) = e^{-x} \cdot e^{-e^{-x}}$$

BLAST. Значимость выравнивания

$$P(\tilde{S}_{n,m} > S) \approx 1 - \exp(-K m n e^{-\lambda S}) \approx K m n e^{-\lambda S}$$

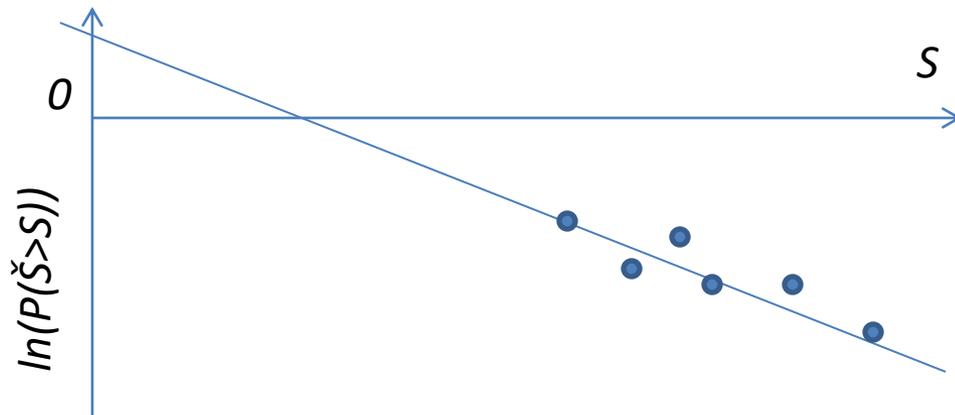
Но каковы в этом случае K и λ ?

В отличие от бросания кубика, вероятности совпадений букв заранее неизвестны.

Прологарифмируем выражение для вероятности:

$$\ln(P(\tilde{S} > S)) = \ln(K m n) - \lambda S$$

Таким образом, мы можем построить много выравниваний для одной и той же последовательности запроса длиной n и базы данных суммарной длиной m (причем почти все эти выравнивания будут заведомо неправильными) и для каждого выравнивания поставить точку в координатах $(S, \ln(P(\tilde{S} > S)))$:



Получившиеся точки можно экстраполировать прямой по методу наименьших квадратов, а наклон этой прямой и её пересечение с осью ординат позволят определить K и λ .

BLAST. Значимость выравнивания

Для BLOSUM62 были получены значения $K = 0.040$ и $\lambda = 0.254$ (Altshul, Gish, 1996)(?).

Подставляем их в формулу для расчета вероятности:

$$P(\tilde{S}_{n,m} > S) \approx 1 - \exp(-K m n e^{-\lambda S}) \approx K m n e^{-\lambda S}$$

$$P(\tilde{S}_{n,m} > S) \equiv E - value$$

$E < 0,02$ высокая вероятность гомологии

$0,02 < E < 1$ гомология не очевидна

$E > 1$ сходство случайно

Пример (миоглобин кашалота):

```
>SP:MYG_PHYCD P02185 Myoglobin OS=Physeter catodon OX=9755 GN=MB PE=1
SV=2 Length=154
```

```
Score = 104 bits (259),    Expect = 1e-29
```

```
Identities = 50/50 (100%), Positives = 50/50 (100%), Gaps = 0/50 (0%)
```

```
Query 1      LAQSHATKHKIPIKYLEFISEAIIHVLHSRHPGDFGADAQGAMNKALELF      50
             LAQSHATKHKIPIKYLEFISEAIIHVLHSRHPGDFGADAQGAMNKALELF
Sbjct 90      LAQSHATKHKIPIKYLEFISEAIIHVLHSRHPGDFGADAQGAMNKALELF      139
```

BLAST. Значимость выравнивания

Для BLOSUM62 были получены значения $K = 0.040$ и $\lambda = 0.254$ (Altshul, Gish, 1996)(?).

Подставляем их в формулу для расчета вероятности:

$$P(\tilde{S}_{n,m} > S) \approx 1 - \exp(-K m n e^{-\lambda S}) \approx K m n e^{-\lambda S}$$

$$P(\tilde{S}_{n,m} > S) \equiv E - value$$

$E < 0,02$ высокая вероятность гомологии

$0,02 < E < 1$ гомология не очевидна

$E > 1$ сходство случайно

Другой пример (калиотоксин):

```
>lcl|PDB:1TI5_A mol:protein length:46 plant defensin Length=46
```

```
Score = 25.8 bits (74), Expect = 1.9
```

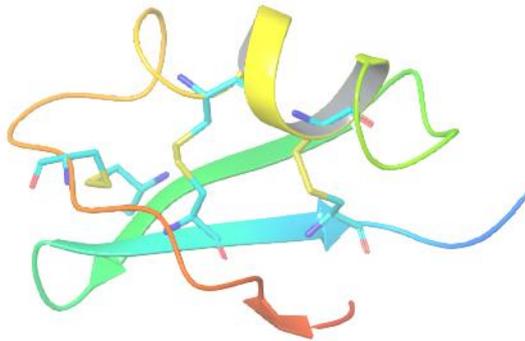
```
Identities = 12/31 (39%), Positives = 14/31 (45%), Gaps = 2/31 (6%)
```

```
Query 7 KCSGSPQCLKPCKDAGMRFGKC--MNRKCHC 35 калиотоксин
```

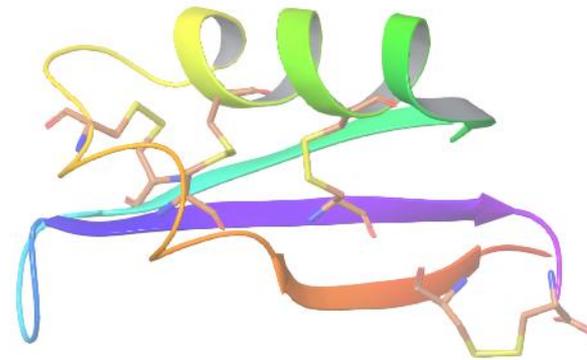
```
KC C CK+ G G C M R C+C
```

```
Sbjct 12 KCLIDTTCAHSCKNRGYIGGNCKGMTRTCCYC 42 дефензин
```

BLAST. Значимость выравнивания



калиотоксин



дефензин

```

3ODV_T  1  -GVEINV---KCSGSPQCLKPKDAGMRFGKCM-N-RKCHCTPK- 38
1TI5_A  1  RTCMIKKEGWGKCLIDTTCAHSCKNRGYIGGNCKGMTRTCyclVNC 46
          * :      ** . * : ** : *   * : *      * * : * :
    
```

>lcl|PDB:1TI5_A mol:protein length:46 plant defensin Length=46

Score = 25.8 bits (74), **Expect = 1.9**

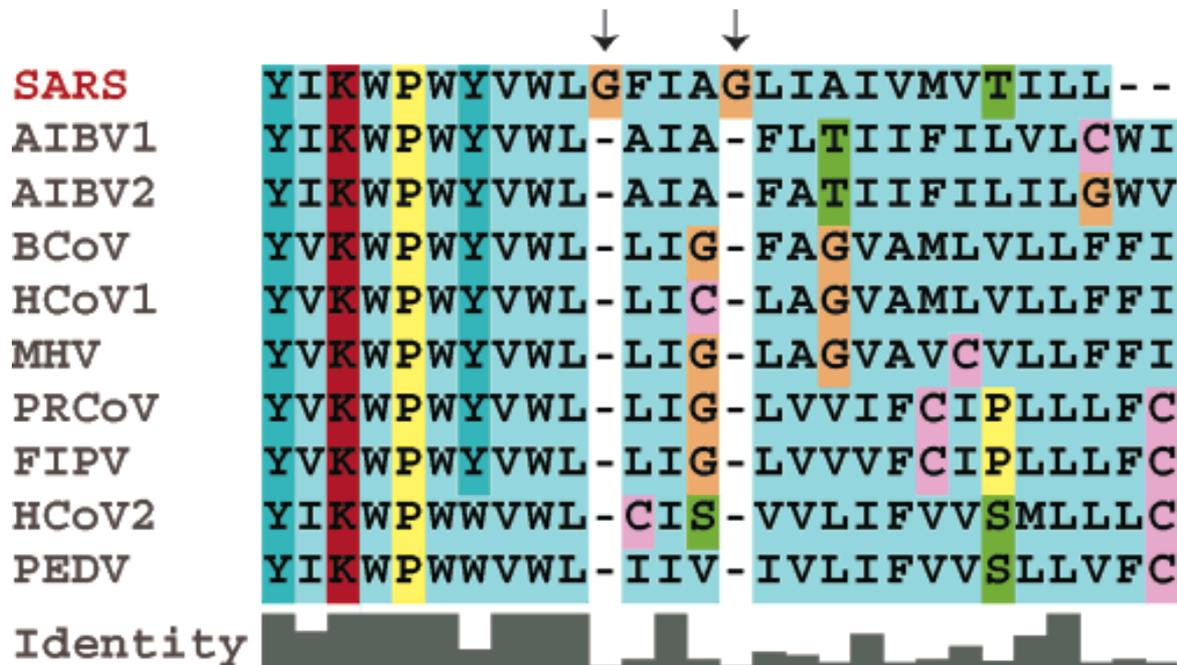
Identities = 12/31 (39%), Positives = 14/31 (45%), Gaps = 2/31 (6%)

```

Query  7  KCSGSPQCLKPKDAGMRFGKC--MNRKCHC 35  калиотоксин
          KC      C      CK+ G      G C      M R C+C
Sbjct 12  KCLIDTTCAHSCKNRGYIGGNCKGMTRTCYC 42  дефензин
    
```

Для коротких последовательностей сходство может быть НЕ случайным даже при $E > 1$!

Множественное выравнивание последовательностей



A Trimerizing GxxxG Motif Is Uniquely Inserted in the SARS Coronavirus Spike Protein Transmembrane Domain (2006)

Multiple sequence alignment of the putative spike protein TMD of 10 different coronaviruses. The SCoV spike protein TMD corresponds to residues Y1194-L1216. The coloring is according to the chemical nature of the amino acid. The arrows indicate the insertion of the glycine residues in the SCoV spike protein TMD.

Множественное выравнивание последовательностей

Цели:

- Построение филогенетических деревьев
- Выявление консервативных остатков и мотивов
- Построение профилей (визуализация)
- Итеративное выявление удаленной гомологии
- ...

Алгоритмы:

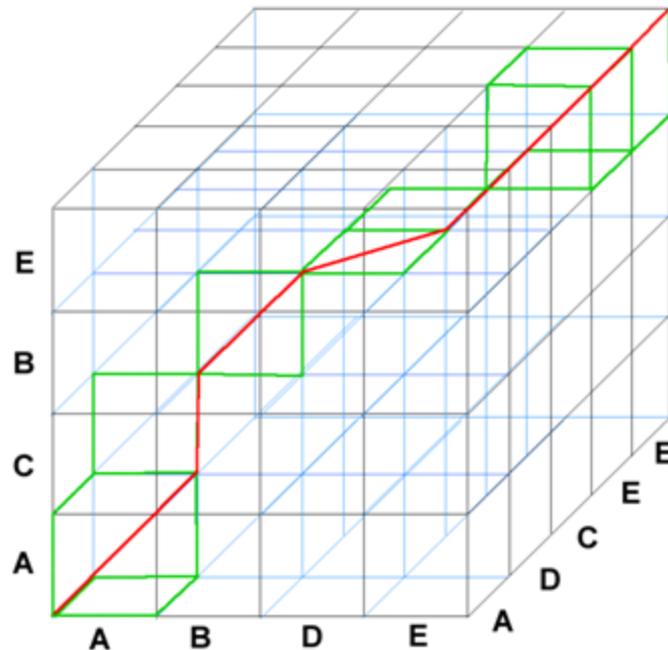
- **Динамическое программирование**
- Прогрессивное выравнивание
- Скрытые марковские модели
- Квантовые компьютеры?

Динамическое программирование

Прямой метод выполнения множественного выравнивания, обеспечивающий нахождение глобального оптимума.

Для выравнивания N последовательностей требуется построение N -мерной матрицы. Таким образом, пространство поиска растет экспоненциально с ростом N , а также зависит от длины последовательностей. Время поиска может быть оценено как $O(L^N)$.

A-BD-E-
ACB--E-
A--DCEE



Множественное выравнивание последовательностей

Цели:

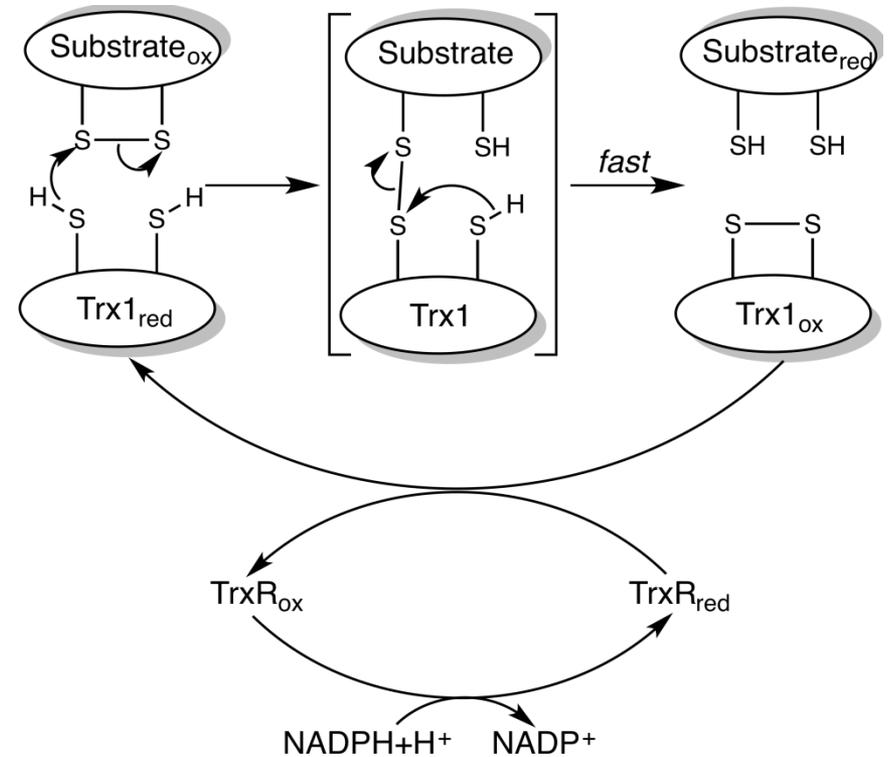
- Построение филогенетических деревьев
- **Выявление консервативных остатков и мотивов**
- **Построение профилей (визуализация)**
- Итеративное выявление удаленной гомологии
- ...

Алгоритмы:

- **Динамическое программирование – не годится**
- Прогрессивное выравнивание
- Скрытые марковские модели
- Квантовые компьютеры?

Построение и визуализация профилей

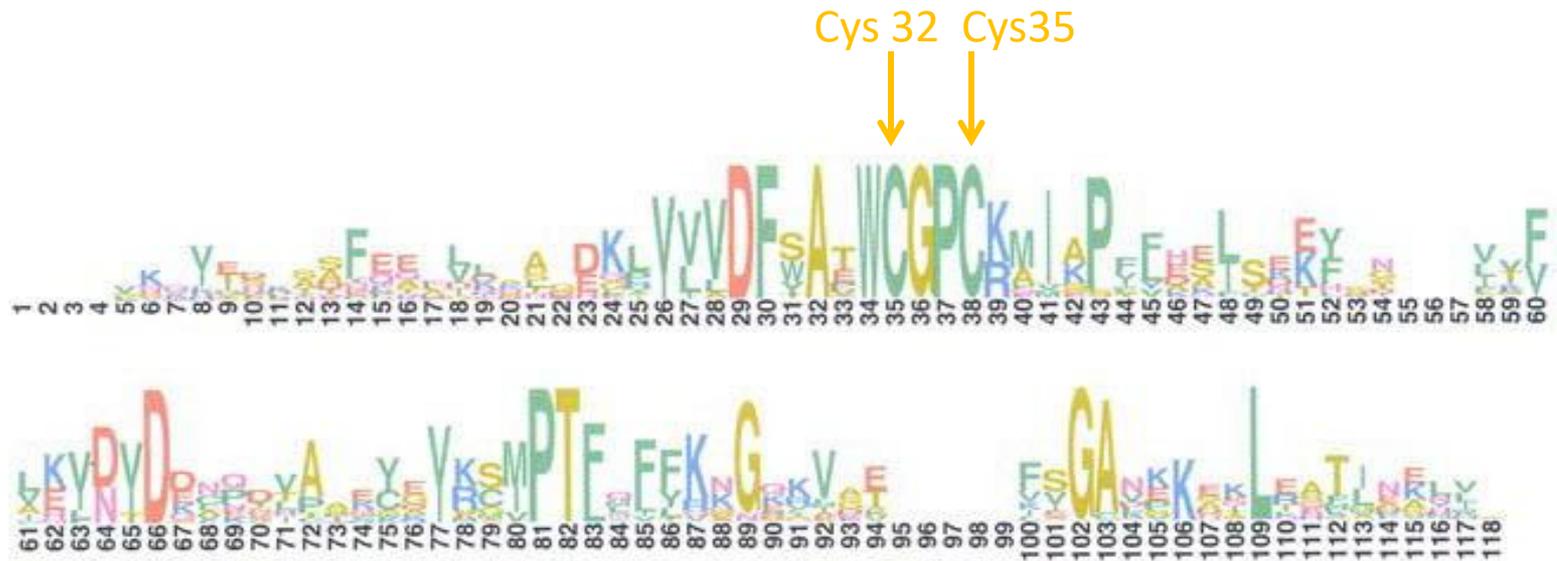
Тиоредоксины – семейство белков, отвечающих за восстановление дисульфидных связей в белках и встречающихся как в животном, так и в растительном мире.



Выравнивание структур тиоредоксина человека и мушки *Drosophila melanogaster*.

Построение и визуализация профилей

(b)



s – число символов в алфавите (4, 20, ...)

$p(a, i)$ – вероятность появления буквы a в позиции i

Sequence logos: a new way to display consensus sequences (Schneider & Stephens, 1990)

$$H_i = - \sum_{\text{по всему алфавиту}} p(a, i) \log_2 p(a, i)$$

Шенноновская энтропия i -той позиции

$$R_i = \log_2 s - H_i$$

Информационная значимость i -той позиции

$$h(a, i) = p(a, i) R_i$$

Высота символа в профиле

Построение и визуализация профилей

Sequence logos: a new way to display consensus sequences (Schneider & Stephens, 1990)

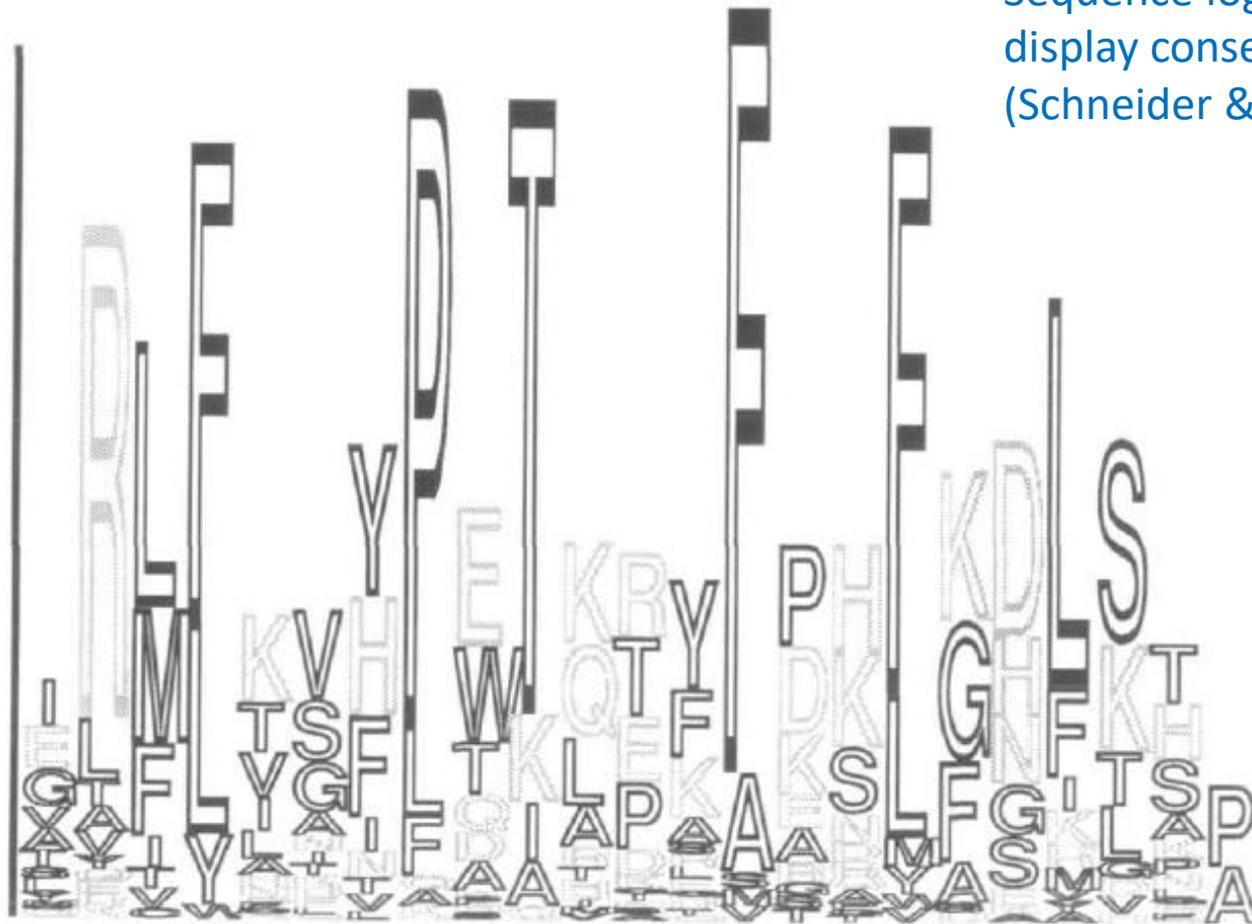


Figure 4. Logo for a portion of the globins. The alignment was performed by the program described by Hein (13) with the weight of an insertion-deletion of length k being $8 + 3k$. The 56 globin sequences supplied with the program were used. The logo is of the conserved packing and sliding contacts at the end of the B through the beginning of the D helices of the globins (21). The vertical bar is 3 bits high. Although outlined characters take longer to draw, they are easier to distinguish from one another when there are more than 4 characters. For example, an F on top of an L is distinguishable from a long E. As a substitute for color, external polar (N and Q) and charged (D, E, K, R and H) amino acids are lightly stippled while ambivalent (P, T, S, C, A, G, Y and W) and internal non-polar (F, L, I, M and V) amino acids are black (21).

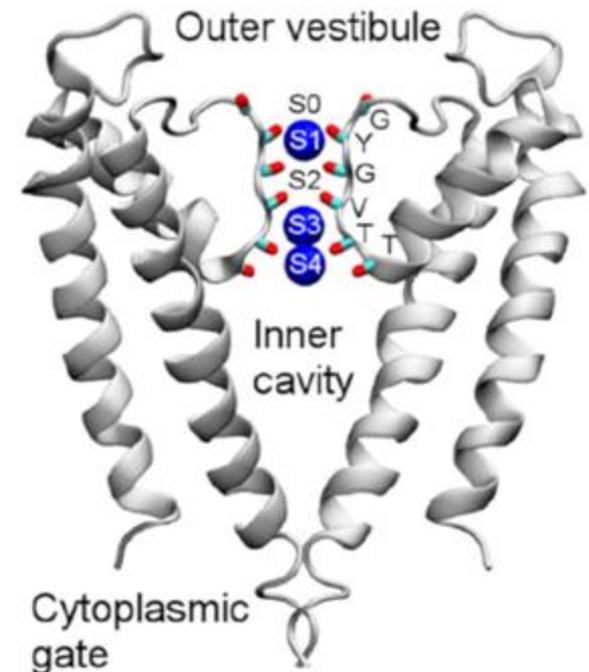
Консенсусная последовательность и мотивы

Консенсусная последовательность – последовательность, в каждой позиции которой находится символ, наиболее часто встречающийся в данной позиции выравнивания.

Мотив – набор символов, (предположительно) имеющий биологическое значение.

```
SLYFTMTSLTTIGFGNIAPTTDVEK  
ALYFTFSSTLSVGFGNVSPNTNSEK  
AFWWAVVTMTTVGYGDMTPVGVWVK  
AMWWAVVTLTTGYGDTIPQSFAGR  
: : : : : : : : : : * : * : : . * . :
```

[TS] [IVT]G[FY]G





MOTIF Search

Search Motif Library

Search Sequence Database

Generate Profile

KEGG2

[Help](#)

Compute

Clear

Enter query pattern or profile:

Pattern in PROSITE format

(Example) C-x-{C}-[DN]-x(2)-C-x(5)-C-C.

Each residue must be separated by - (**minus**).

x represents any amino acid.

[DE] means either D or E.

{FWY} means any amino acid except for F, W and Y.

A(2,3) means that A appears 2 to 3 times consecutively.

The pattern string must be terminated with . (**period**).

Result of Pattern Search

Number of found sequences: 1000 in SWISSPROT

Entry Name	Position	Description
12AH_CLOS4	18..22 Detail	[P21215] RecName: Full=12-alpha-hydroxysteroid dehydrogenase; EC=1.1.1.176; Flags: Fragment;
5HTR_BOMMO	333..337 Detail	[Q17239] RecName: Full=5-hydroxytryptamine receptor; » show all
AKA11_RAT	276..280 Detail	[Q62924] RecName: Full=A-kinase anchor protein 11; » show all
AKT1_ARATH	253..257 Detail	[Q38998] RecName: Full=Potassium channel AKT1;
AKT1_ORYSI	300..304 Detail	[P0C550] RecName: Full=Potassium channel AKT1; Short=OsAKT1;
AKT1_ORYSJ	300..304 Detail	[Q0JKV1] RecName: Full=Potassium channel AKT1; Short=OsAKT1;
AKT2_ARATH	277..281 Detail	[Q38898] RecName: Full=Potassium channel AKT2/3;



Database of protein domains, families and functional sites

 [SARS-CoV-2 relevant PROSITE motifs](#)

PROSITE consists of documentation entries describing protein domains, families and functional sites as well as associated patterns and profiles to identify them [[More...](#) / [References](#) / [Commercial users](#)].
PROSITE is complemented by [ProRule](#), a collection of rules based on profiles and patterns, which increases the discriminatory power of profiles and patterns by providing additional information about functionally and/or structurally critical amino acids [[More...](#)].

Release 2020_05 of 07-Oct-2020 contains 1865 documentation entries, 1311 patterns, 1288 profiles and 1321 ProRule.

Search

e.g. PDOC00022, PS50089, SH3, zinc finger

Search

Browse

- [by documentation entry](#)
- [by ProRule description](#)
- [by taxonomic scope](#)
- [by number of positive hits](#)

The MEME Suite

Motif-based sequence analysis tools

MEME Suite 5.1.0

► Motif Discovery

► Motif Enrichment

► Motif Scanning

► Motif Comparison

► Gene Regulation

► Manual

► Guides & Tutorials

► Sample Outputs

► File Format Reference

► Databases

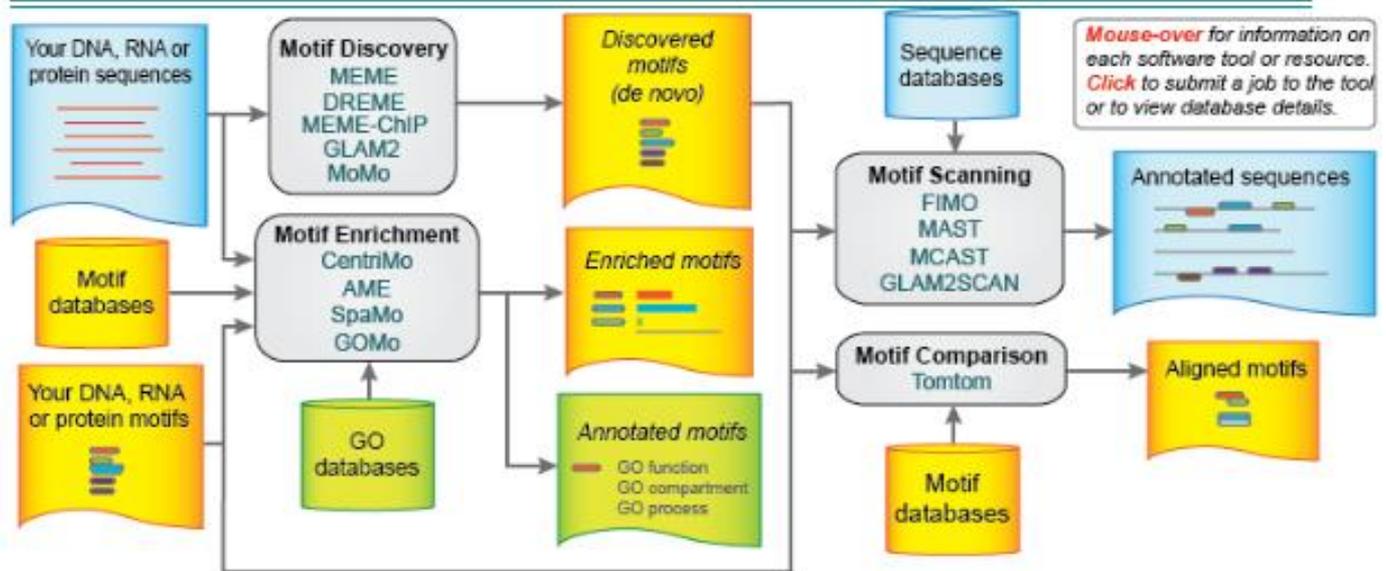
► Download & Install

► Help

► Alternate Servers

► Authors & Citing

► Recent Jobs



MEME
Multiple Em for Motif Elicitation

Improved

CentriMo
Local Motif Enrichment Analysis

FIMO
Find Individual Motif Occurrences

DREME
Discriminative Regular Expression Motif Elicitation

AME
Analysis of Motif Enrichment

Improved **MAST**
Motif Alignment & Search Tool

MEME-ChIP
Motif Analysis of Large Nucleotide Datasets

SpaMo
Spaced Motif Analysis Tool

MCAST
Motif Cluster Alignment and Search Tool

GLAM2
Gapped Local Alignment of Motifs

GOMo
Gene Ontology for Motifs

GLAM2Scan
Scanning with Gapped Motifs

MoMo
Modification Motifs

Improved **Tomtom**
Motif Comparison Tool

GT-Scan
Identifying Unique Genomic Targets

CisMapper
Predicting Regulatory Links

New

Множественное выравнивание последовательностей

Цели:

- Построение филогенетических деревьев
- **Выявление консервативных остатков и мотивов**
- **Построение профилей (визуализация)**
- **Итеративное выявление удаленной гомологии**
- ...

Алгоритмы:

- **Динамическое программирование – не годится**
- Прогрессивное выравнивание
- Скрытые марковские модели
- Квантовые компьютеры?

Position-Specific Iterative BLAST

Position-Specific Iterative BLAST (**PSI-BLAST**) (Altschul et al. 1997)

<http://www.ebi.ac.uk/Tools/sss/psiblast/>

Алгоритм:

- Просмотр БД и составление списка последовательностей, сходных с последовательностью запроса;
- Расчет множественного выравнивания и набора позиционно-специфичных матриц замен (PSSM) для него;
- Новый просмотр БД с использованием полученных PSSM вместо стандартных матриц;
- Повторение процедуры либо желаемое число раз, либо до достижения сходимости (состояния, когда новые последовательности не обнаруживаются).



Давид Липман

EN = 3

PSI-BLAST эффективен при поиске удалённых гомологов.

What about Impossible Burger?



<https://impossiblefoods.com/>

Множественное выравнивание последовательностей

Цели:

- Построение филогенетических деревьев
- **Выявление консервативных остатков и мотивов**
- **Построение профилей (визуализация)**
- **Итеративное выявление удаленной гомологии**
- ...

Алгоритмы:

- **Динамическое программирование – не годится**
- **Прогрессивное выравнивание**
- **Скрытые марковские модели**
- Квантовые компьютеры?

О математике для биологов

«В нашу современную жизнь вторгается математика с её особым стилем мышления, становящимся сейчас обязательным и для инженера, и для биолога.»



Гнеденко
Борис Владимирович
(1912 – 1995)
EN = 3

Поиск по MSC

Расстояние Сотрудничества

Современные Журналы

Современные Публикации

ISSN 2167-5163

MR Erdos Number = 3

Boris Vladimirovich Gnedenko	coauthored with	Andreï Nikolaevich Kolmogorov	MR0062975
Andreï Nikolaevich Kolmogorov	coauthored with	Alfréd Rényi	MR0113739
Alfréd Rényi	coauthored with	Paul Erdős ¹	MR0125031