

# ПОИСК МОТИВОВ

программа MEME

# Мотив (в данном случае) – частичное множественное выравнивание без разрывов

VPFIVNDDVSL  
VPFIVNDDVAL  
VPFIVNDDVEL  
IPFIVNDDVDL  
VPFIVNDDVEL  
IPFIINDDVEL  
IPFIINDDVDL  
VPFIINDDVAL  
VPFIVNDDVEL  
VPFIIDDDIDL  
VSFIVNDDVEL

**VPFIVNDDVEL**

Консенсус

[VI]-[PS]-F-I-[VI]-[DN]-D-D-[VI]-[ASDE]-L

Паттерн

# Консенсус

Сайты связывания PurR *E. coli*

<i>сврА</i>	ССТАСГСАААСГТТТТСТТТТТ
<i>purM</i>	GTCTCGCAAACGTTTGCTTTCC
<i>purT</i>	CACACGCAAACGTTTTCGTTTA
<i>purL</i>	TCCACGCAAACGGTTTCGTCAG
<i>purE</i>	GCCACGCAACCGTTTTCTTGC
<i>purC</i>	GATACGCAAACGTGTGCGTCTG
<i>purB</i>	CCGACGCAATCGGTTACCTTGA
<i>purH</i>	GTTGCGCAAACGTTTTCGTTAC
<i>purA</i> <sub>1</sub>	TTGAGGAAAACGATTGGCTGAA
<i>purA</i> <sub>2</sub>	TTTAAGCAAACGGTGATTTTGA
<i>guaB</i>	TAGATGCAATCGGTTACGCTCT
<i>purR</i> <sub>1</sub>	TAAAGGCAAACGTTTACCTTGC
<i>purR</i> <sub>2</sub>	AACGAGCAAACGTTTCCACTAC
consensus	ACGCAAACGTTTTCGT

# Матрица частот (frequency matrix)

Сайты связывания PurR *E. coli*

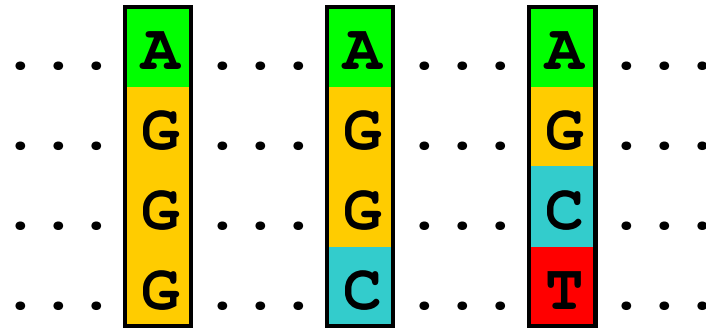
	<b>a</b>	<b>m</b>	<b>G</b>	<b>A</b>	<b>A</b>	<b>A</b>	<b>a</b>	<b>C</b>	<b>G</b>	<b>k</b>	<b>T</b>	<b>T</b>	<b>w</b>	<b>C</b>	<b>w</b>	<b>T</b>
<i>b</i>																
<b>A</b>	10	2	0	1	13	13	10	0	0	1	0	0	4	0	1	0
<b>C</b>	0	8	0	12	0	0	1	13	0	0	0	0	1	11	4	2
<b>G</b>	2	2	13	0	0	0	0	0	13	4	1	1	3	1	5	0
<b>T</b>	1	1	0	0	0	0	2	0	0	8	12	12	5	1	3	11

## Профиль(PSSM)

$$W(b, j) = \ln [N(b, j)+0,5] - 0,25 \sum_i \ln [N(i, j)+0,5]$$

<b>A</b>	1,6	-0,2	-0,8	0,0	2,5	2,5	1,6	-0,8	-0,8	-0,4	-1,1	-1,1	0,3	-1,3	-0,8	-1,2
<b>C</b>	-1,4	1,0	-0,8	2,1	-0,8	-0,8	-0,3	2,5	-0,8	-1,5	-1,1	-1,1	-0,8	1,8	0,3	0,4
<b>G</b>	0,2	-0,2	2,5	-1,1	-0,8	-0,8	-1,4	-0,8	2,5	0,7	0,0	0,0	0,0	-0,2	0,5	-1,2
<b>T</b>	-0,3	-0,7	-0,8	-1,1	-0,8	-0,8	0,2	-0,8	-0,8	1,3	2,1	2,1	0,5	-0,2	0,0	1,9

# Сравнение матрицы частот и PSSM



Частоты

<b>A</b>	<b>1</b>	<b>1</b>	<b>1</b>
<b>C</b>	0	1	1
<b>G</b>	3	2	1
<b>T</b>	0	0	1

Беза (scores)

<b>A</b>	<b>0,34</b>	<b>0,15</b>	<b>0,00</b>
<b>C</b>	<b>-0,76</b>	<b>0,15</b>	<b>0,00</b>
<b>G</b>	<b>1,18</b>	<b>0,66</b>	<b>0,00</b>
<b>T</b>	<b>-0,76</b>	<b>-0,95</b>	<b>0,00</b>

# Информационное содержание

$$I_j = \sum_b f(b, j) \log_2 [f(b, j) / p(b)]$$

$$I = \sum_j I_j$$

где  $b$  пробегает буквы алфавита (например, {A, T, G, C}),  
 $j$  пробегает позиции выравнивания,

$f(b, j) = N(b, j)/N$ , т.е. доля буквы  $b$  в позиции  $j$  ( $N$  – число последовательностей)

$p(b)$  – «априорная» доля буквы (в выборке или в базе данных;  
в случае нуклеотидных последовательностей обычно принимают  $p(b) = 1/4$ )

Числа  $f(b, j)$  образуют т.н. «матрицу вероятностей» (probability matrix)

# Диаграмма Лого (Logo)

Сайты связывания PurR *E. coli*

$$I_j = \sum_b f(b, j) \log_2[f(b, j) / p(b)]$$

Все  $p(b) = 1/4 \Rightarrow$  максимальное значение  $I_j$  равно 2



Высота буквы  $b$  в позиции  $j$  рисуется равной слагаемому  $f(b, j) \log_2[f(b, j)/p(b)]$

# MEME

<http://meme-suite.org/tools/meme>

В составе EMBASSY

(интегрируется в EMBOSS, но не mEMBOSS ☹ )

## **Нахождение мотивов в данных последовательностях**

- `meme` (выдача в html-формате)
- `memetext` (та же информация в виде текста)

## **Поиск по мотиву в банке последовательностей**

- `emast`