MEME
Multiple Em for Motif Elicitation

Программа **MEME**

Множественное локальное выравнивание

http://meme.nbcr.net/meme4_3_0/cgi-bin/meme.cgi

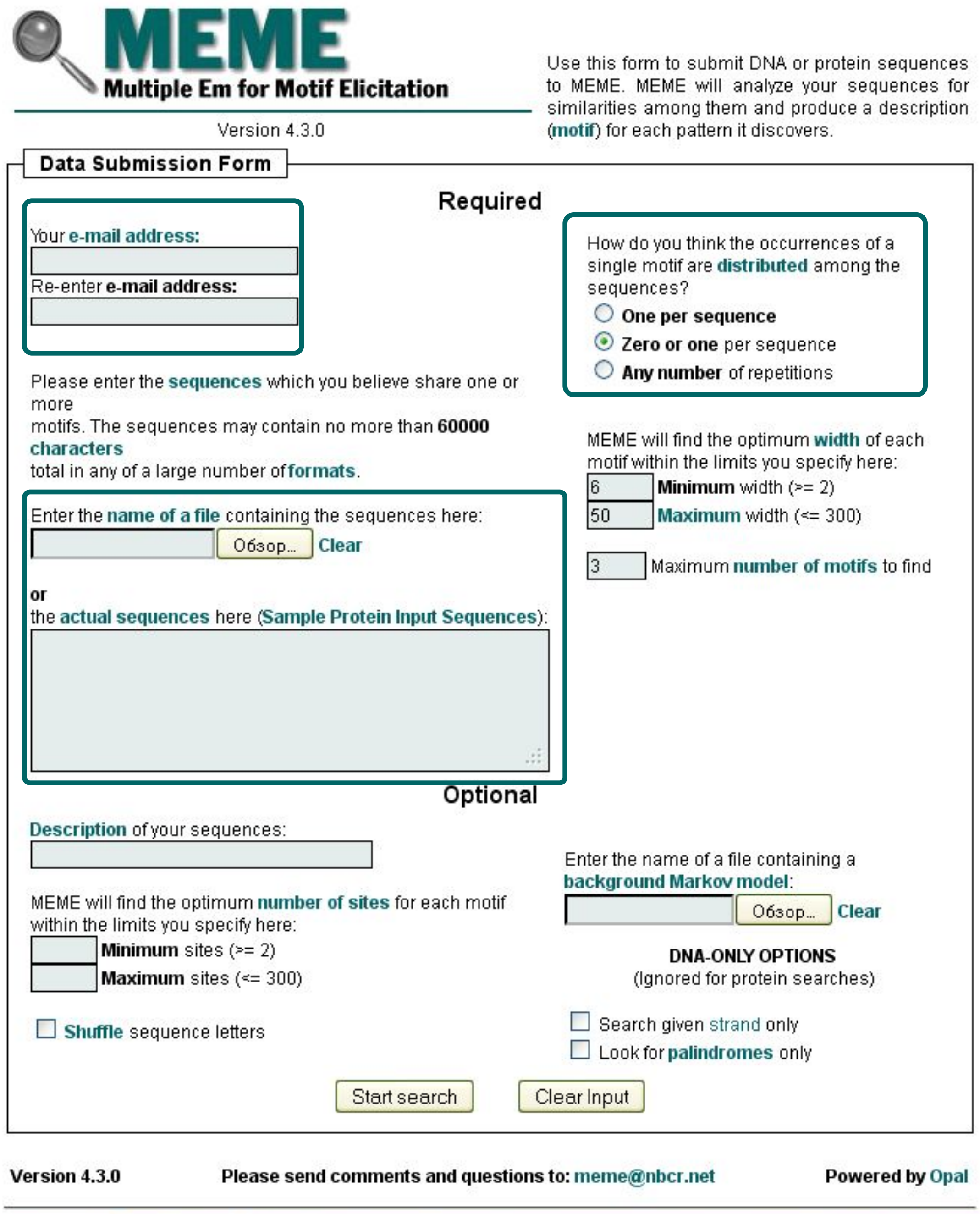
MEME
Multiple Em for Motif Elicitation

- Локальное выравнивание довольно часто имеет больший биологический смысл, чем глобальное

- MEME – наиболее популярная программа множественного локального выравнивания

- MEME находит **блоки**, т.е.участки локальных выравниваний <u>без</u> гэпов

Онлайн форма **MEME**



Version 4.3.0

Use this form to submit DNA or protein sequences to MEME. MEME will analyze your sequences for similarities among them and produce a description (**motif**) for each pattern it discovers.

Data Submission Form

Required

Your **e-mail address:**

Re-enter **e-mail address:**

Please enter the **sequences** which you believe share one or more motifs. The sequences may contain no more than **60000 characters** total in any of a large number of **formats**.

Enter the **name of a file** containing the sequences here:

or
the **actual sequences** here (**Sample Protein Input Sequences**):

How do you think the occurrences of a single motif are **distributed** among the sequences?

○ **One per sequence**
● **Zero or one** per sequence
○ **Any number** of repetitions

MEME will find the optimum **width** of each motif within the limits you specify here:

6 **Minimum** width (>= 2)
50 **Maximum** width (<= 300)

3 Maximum **number of motifs** to find

Optional

Description of your sequences:

MEME will find the optimum **number of sites** for each motif within the limits you specify here:

Minimum sites (>= 2)
Maximum sites (<= 300)

☐ **Shuffle** sequence letters

Enter the name of a file containing a **background Markov model**:

DNA-ONLY OPTIONS
(Ignored for protein searches)

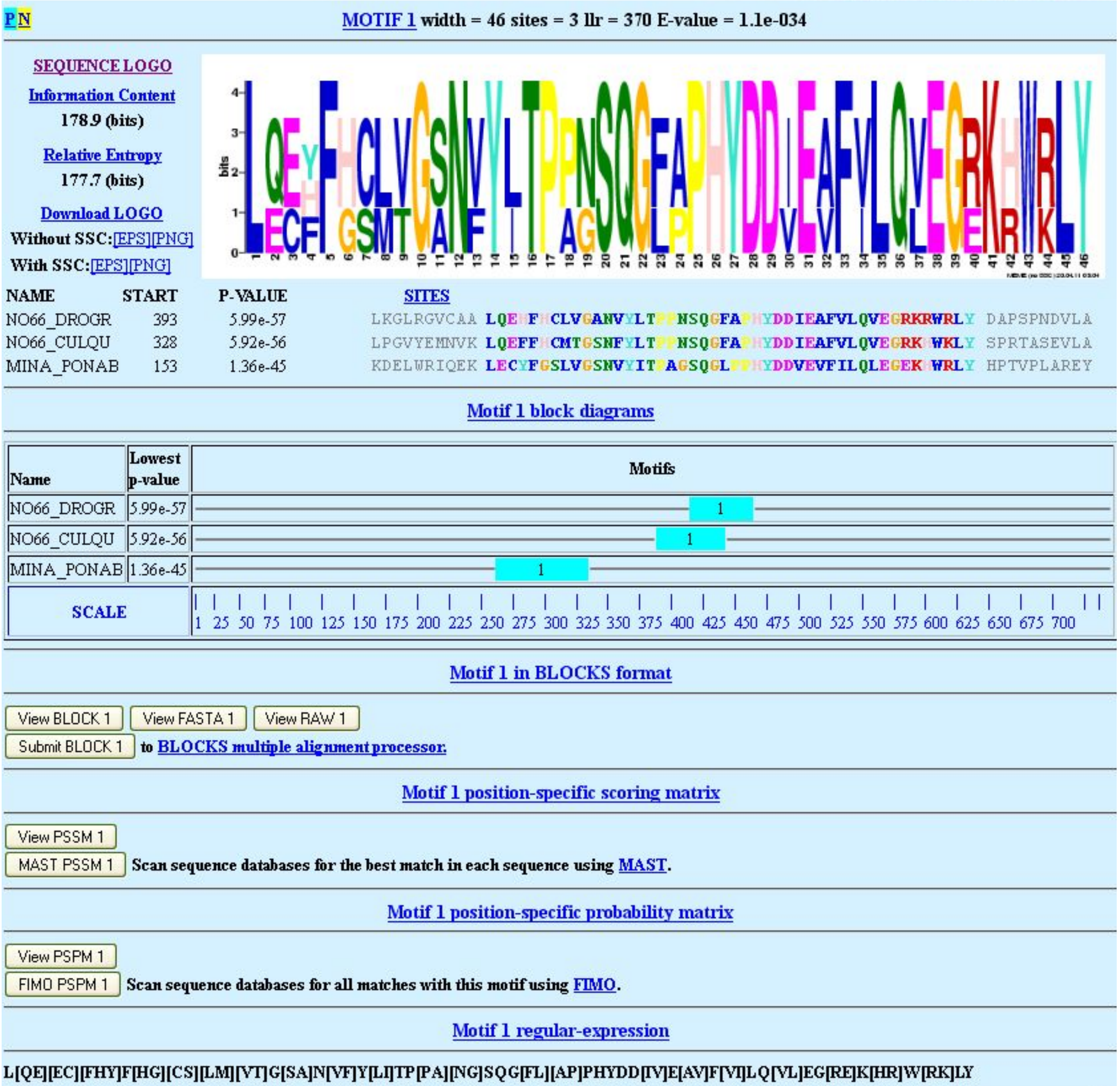
☐ Search given strand only
☐ Look for **palindromes** only

Version 4.3.0 **Please send comments and questions to:** meme@nbcr.net **Powered by** Opal

Следует заполнить:

1. Email
2. Загрузить файл с последовательностями в FASTA-формате (или paste в текстовое поле)
3. Указать, сколько раз найденный блок может встречаться в заданном выравнивании
    - Один
    - От нуля до одного
    - Любое число раз

Основные результаты MEME



MOTIF 1 width = 46 sites = 3 llr = 370 E-value = 1.1e-034

SEQUENCE LOGO

Information Content
178.9 (bits)

Relative Entropy
177.7 (bits)

Download LOGO
Without SSC:[EPS][PNG]
With SSC:[EPS][PNG]

NAME	START	P-VALUE	SITES		
NO66_DROGR	393	5.99e-57	LKGLRGVCAA	LQEHFHCLVGANVYLTPPNSQGFAPHYDDIEAFVLQVEGRKRWRLY	DAPSPNDVLA
NO66_CULQU	328	5.92e-56	LPGVYEMNVK	LQEFFHCMTGSNFYLTPPNSQGFAPHYDDIEAFVLQVEGRKHWKLY	SPRTASEVLA
MINA_PONAB	153	1.36e-45	KDELWRIQEK	LECYFGSLVGSNVYITPAGSQGLPPHYDDVEVFILQLEGEKHWRLY	HPTVPLAREY

Motif 1 block diagrams

Name	Lowest p-value	Motifs
NO66_DROGR	5.99e-57	1
NO66_CULQU	5.92e-56	1
MINA_PONAB	1.36e-45	1
SCALE		1 25 50 75 100 125 150 175 200 225 250 275 300 325 350 375 400 425 450 475 500 525 550 575 600 625 650 675 700

Motif 1 in BLOCKS format

to BLOCKS multiple alignment processor.

Motif 1 position-specific scoring matrix

Scan sequence databases for the best match in each sequence using MAST.

Motif 1 position-specific probability matrix

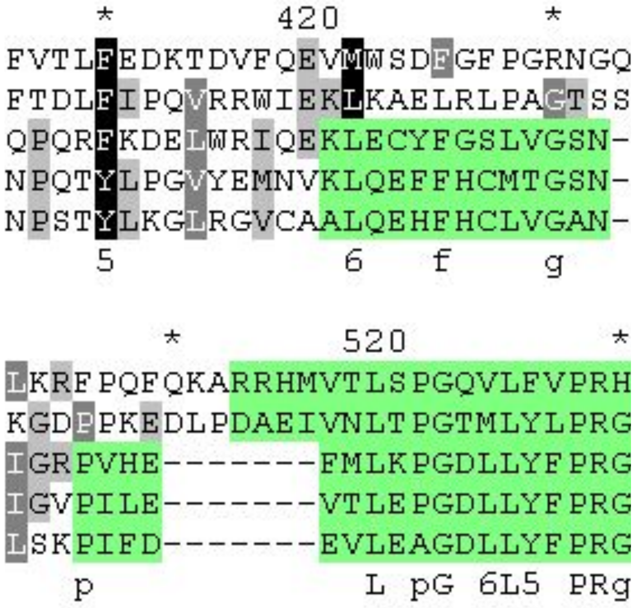
Scan sequence databases for all matches with this motif using FIMO.

Motif 1 regular-expression

L[QE][EC][FHY]F[HG][CS][LM][VT]G[SA]N[VF]Y[LI]TP[PA][NG]SQG[FL][AP]PHYDD[IV]E[AV]F[VI]LQ[VL]EG[RE]K[HR]W[RK]LY

Для каждого **блока**
- SEQUENCE LOGO
- log likelihood ratio(LLR)
- E-value
- p-value
- PSSM (position-specific scoring matrix
- PSPM (position-specific probability matrix)

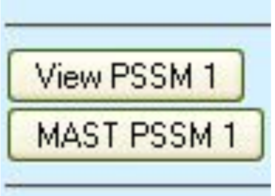
MEME vs Muscle



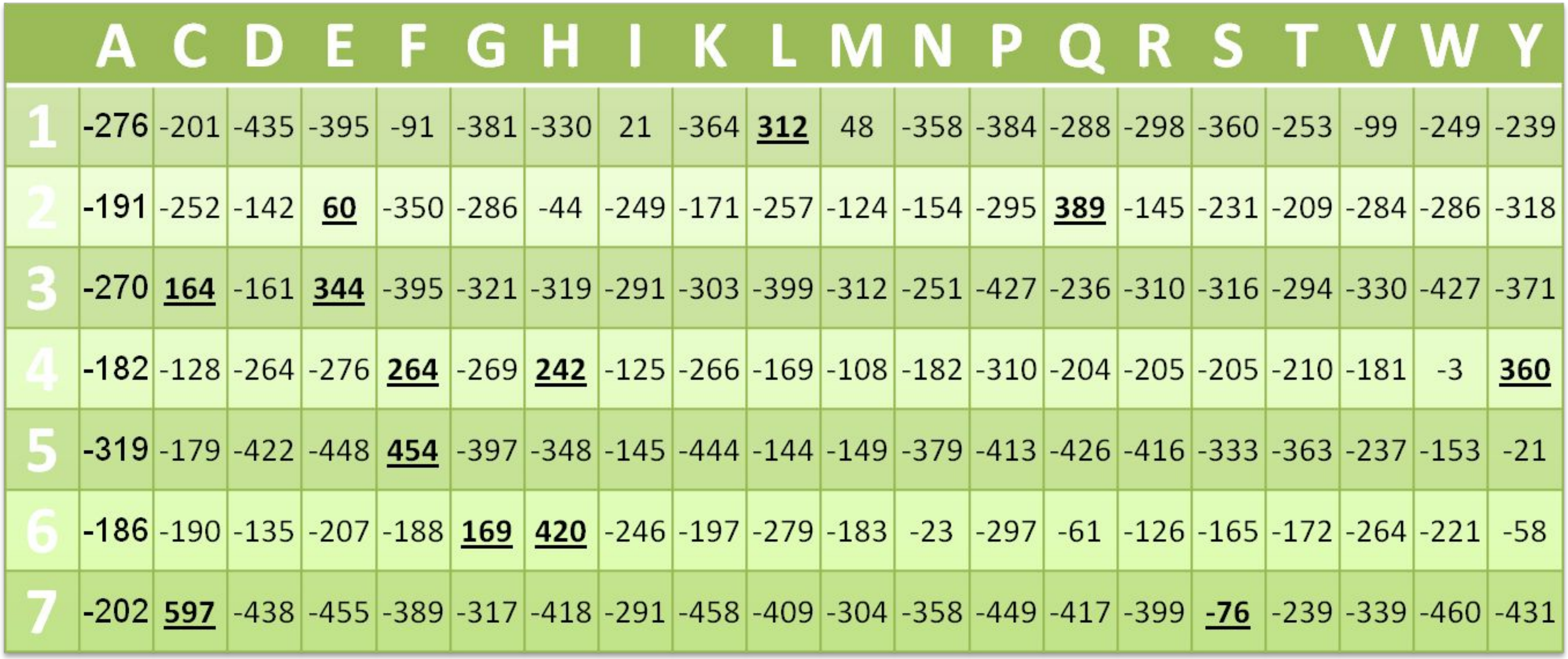
```
    *          420          *
FVTLFEDKTDVFQEVMWSDFGFPGRNGQ
FTDLFIPQVRRWIEKLKAELRLPAGTSS
QPQRFKDELWRIQEKLECYFGSLVGSN-
NPQTYLPGVYEMNVKLQEFFHCMTGSN-
NPSTYLKGLRGVCAALQEHFHCLVGAN-
    5          6   f    g

        *       520         *
LKRFPQFQKARRHMVTLSPGQVLFVPRH
KGDPPKEDLPDAEIVNLTPGTMLYLPRG
IGRPVHE-------FMLKPGDLLYFPRG
IGVPILE-------VTLEPGDLLYFPRG
LSKPIFD-------EVLEAGDLLYFPRG
  p             L pG 6L5 PRg
```

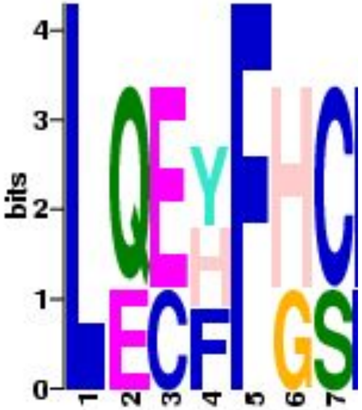
- Блоки MEME не обязательно во всех последовательностях
- Блоки MEME могут не совпадать с выравниванием Muscle



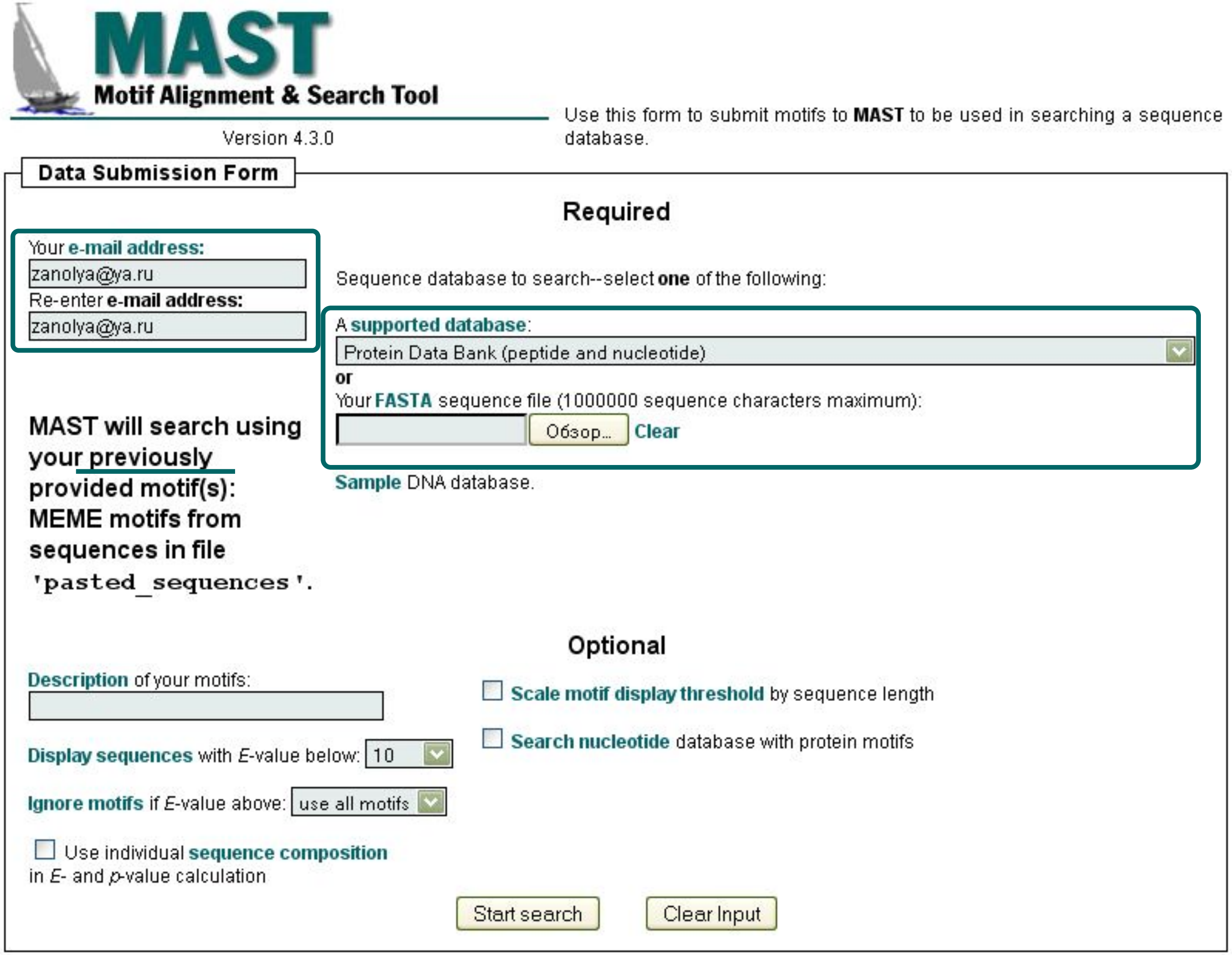
PSSM
(position-specific scoring matrix)



	A	C	D	E	F	G	H	I	K	L	M	N	P	Q	R	S	T	V	W	Y
1	-276	-201	-435	-395	-91	-381	-330	21	-364	**312**	48	-358	-384	-288	-298	-360	-253	-99	-249	-239
2	-191	-252	-142	**60**	-350	-286	-44	-249	-171	-257	-124	-154	-295	**389**	-145	-231	-209	-284	-286	-318
3	-270	**164**	-161	**344**	-395	-321	-319	-291	-303	-399	-312	-251	-427	-236	-310	-316	-294	-330	-427	-371
4	-182	-128	-264	-276	**264**	-269	**242**	-125	-266	-169	-108	-182	-310	-204	-205	-205	-210	-181	-3	**360**
5	-319	-179	-422	-448	**454**	-397	-348	-145	-444	-144	-149	-379	-413	-426	-416	-333	-363	-237	-153	-21
6	-186	-190	-135	-207	-188	**169**	**420**	-246	-197	-279	-183	-23	-297	-61	-126	-165	-172	-264	-221	-58
7	-202	**597**	-438	-455	-389	-317	-418	-291	-458	-409	-304	-358	-449	-417	-399	**-76**	-239	-339	-460	-431



LQEHFHC
LQEFFHC
LECYFGS
1234567


MAST
Motif Alignment & Search Tool
Version 4.3.0

Use this form to submit motifs to **MAST** to be used in searching a sequence database.

Data Submission Form

Required

Your **e-mail address:** zanolya@ya.ru
Re-enter **e-mail address:** zanolya@ya.ru

Sequence database to search--select **one** of the following:

A **supported database**: Protein Data Bank (peptide and nucleotide)
or
Your **FASTA** sequence file (1000000 sequence characters maximum):

Sample DNA database.

MAST will search using your previously provided motif(s): MEME motifs from sequences in file `'pasted_sequences'`.

Optional

Description of your motifs:

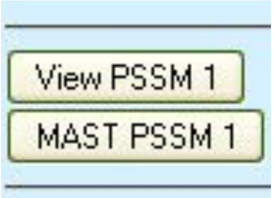
Display sequences with *E*-value below: 10

Ignore motifs if *E*-value above: use all motifs

☐ Use individual **sequence composition** in *E*- and *p*-value calculation

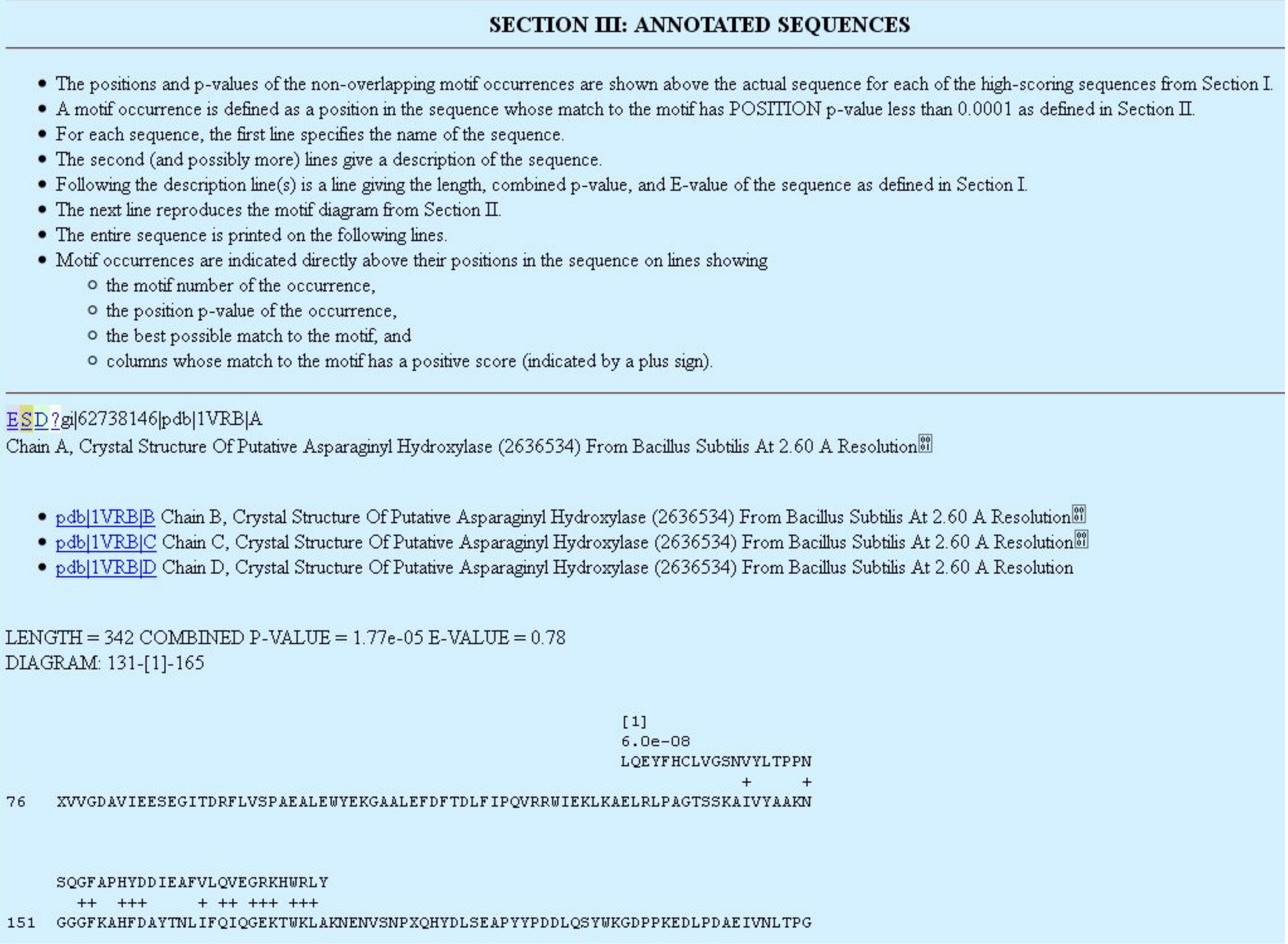
☐ **Scale motif display threshold** by sequence length

☐ **Search nucleotide** database with protein motifs



- Motif file name: **MEME motifs from sequences in file pasted_sequences.**
- Number of motifs: **1**
- Total motif columns: **46**
- Motif alphabet: **PROTEIN**
- Database to search: **Protein Data Bank (peptide and nucleotide) (pdb.aa)**
- Displaying sequences with ***E*-value < 10**
- Motif display threshold: **motif *p*-value < 0.0001**

Основные результаты MAST



SECTION III: ANNOTATED SEQUENCES

- The positions and p-values of the non-overlapping motif occurrences are shown above the actual sequence for each of the high-scoring sequences from Section I.
- A motif occurrence is defined as a position in the sequence whose match to the motif has POSITION p-value less than 0.0001 as defined in Section II.
- For each sequence, the first line specifies the name of the sequence.
- The second (and possibly more) lines give a description of the sequence.
- Following the description line(s) is a line giving the length, combined p-value, and E-value of the sequence as defined in Section I.
- The next line reproduces the motif diagram from Section II.
- The entire sequence is printed on the following lines.
- Motif occurrences are indicated directly above their positions in the sequence on lines showing
  - the motif number of the occurrence,
  - the position p-value of the occurrence,
  - the best possible match to the motif, and
  - columns whose match to the motif has a positive score (indicated by a plus sign).

ESD?gi|62738146|pdb|1VRB|A
Chain A, Crystal Structure Of Putative Asparaginyl Hydroxylase (2636534) From Bacillus Subtilis At 2.60 A Resolution

- pdb|1VRB|B Chain B, Crystal Structure Of Putative Asparaginyl Hydroxylase (2636534) From Bacillus Subtilis At 2.60 A Resolution
- pdb|1VRB|C Chain C, Crystal Structure Of Putative Asparaginyl Hydroxylase (2636534) From Bacillus Subtilis At 2.60 A Resolution
- pdb|1VRB|D Chain D, Crystal Structure Of Putative Asparaginyl Hydroxylase (2636534) From Bacillus Subtilis At 2.60 A Resolution

LENGTH = 342 COMBINED P-VALUE = 1.77e-05 E-VALUE = 0.78
DIAGRAM: 131-[1]-165

```
                                                             [1]
                                                             6.0e-08
                                                             LQEYFHCLVGSNVYLTPPN
                                                                         +     +
76   XVVGDAVIEESEGITDRFLVSPAEALEWYEKGAALEFDFTDLFIPQVRRWIEKLKAELRLPAGTSSKAIVYAAKN


     SQGFAPHYDDIEAFVLQVEGRKHWRLY
       ++  +++     + ++ +++ +++
151  GGGFKAHFDAYTNLIFQIQGEKTWKLAKNENVSNPXQHYDLSEAPYYPDDLQSYWKGDPPKEDLPDAEIVNLTPG
```