Задание 6

Трещева Мария, Б06-207



Домены SH2 связываются с фосфорилированными остатками тирозина на других белках, изменяя функцию или активность SH2-содержащего белка. Домен SH2 можно считать модульным доменом белок-белкового взаимодействия, обеспечивающим передачу сигналов.

* Взяли протеомы Danio rerio и Carcharodon carcharias (далее – акула).

Сначала получили файлы sh2\_danio\_rerio.tblout и sh2\_shark.tblout. Наименьшие e-value были у нескольких идущих подряд изоформ tyrosine-protein phosphatase non-receptor type. Сначала я решила, что выравнивать одни и те же изоформы не очень интересно и поэтому взяла:

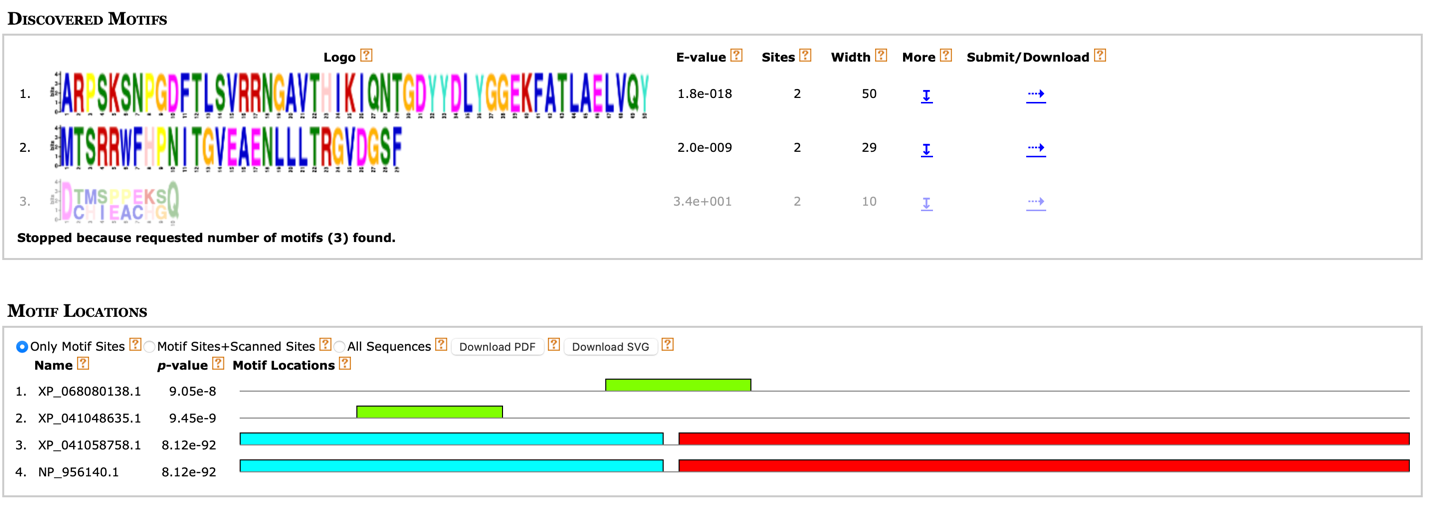
1. **изоформу tyrosine-protein phosphatase non-receptor type 11** для Danio rerio и акулы (у них абсолютный min E-value) и
2. а. **ras GTPase-activating protein 1 isoform X1** для *Danio rerio*

б. **1-phosphatidylinositol 4,5-bisphosphate phosphodiesterase gamma-2** для *акулы*

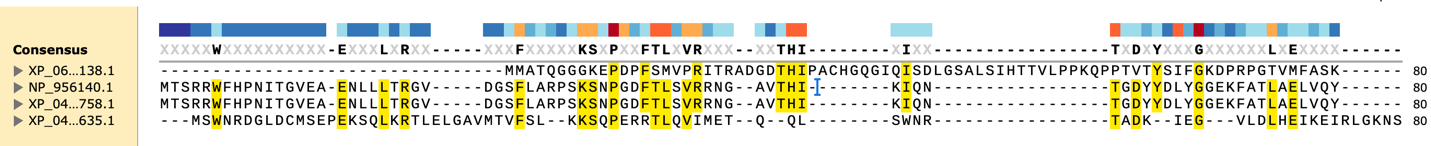
Это были следующие белки с минимальным E-value, не являющиеся изоформами первого белка.

* Множественное выравнивание.

Meme смог выровнять между собой tyrosine-protein phosphatase. Между ГТФазой и фосфодиэстеразой он нашел совпадение участка – зеленым цветом, но в окне ‘Discovered motifs’ видно, что совпадений там мало(. Сначала было задано минимальное количество мотивов=3



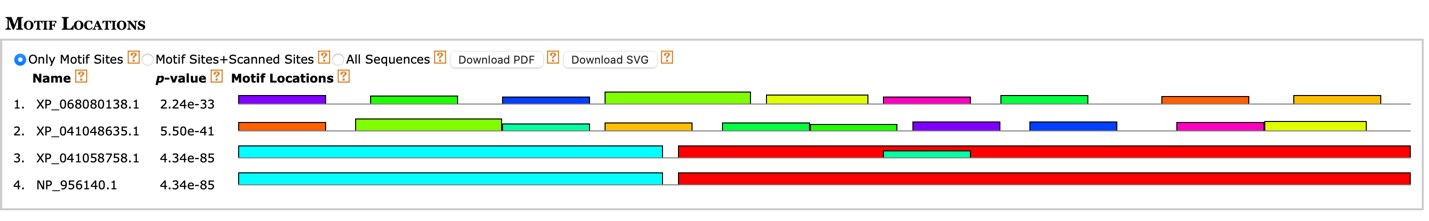
Чтобы увидеть, как много совпадений есть у данных последовательностей, я выровняла с помощью Clustal Omega. Видно, что непосредственно совпадающих аминокислот мало у первого белка (это ras GTPase-activating protein 1 isoform X1 [Danio rerio]). Видимо, совпадают каталитические аминокислоты, а скэффолд белка имеет не очень много прямых пересечений.



Затем я задала ему минимальное количество мотивов=6, но он все равно находил очень сильные совпадения между последовательностями фосфатазы. Тем не менее, появились новые найденные участки совпадения



Минимальное количество мотивов=14 тоже не дало ожидаемых результатов по выравниванию с фосфатазами:



Потом я решила взять и посмотреть, что будет при выравнивании четырех изоформ:

>NP\_956140.1 tyrosine-protein phosphatase non-receptor type 11 [Danio rerio]

MTSRRWFHPNITGVEAENLLLTRGVDGSFLARPSKSNPGDFTLSVRRNGAVTHIKIQNTGDYYDLYGGEKFATLAELVQY

>XP\_005162214.1 tyrosine-protein phosphatase non-receptor type 11b isoform X1 [Danio rerio]

MVRWFHPNITGIEAEHLLLTRGVHGSFLARPSKSNPGDFTLSVRRNDEVTHIKIQNSGDYYDLYGGEKFATLAELVQYYT

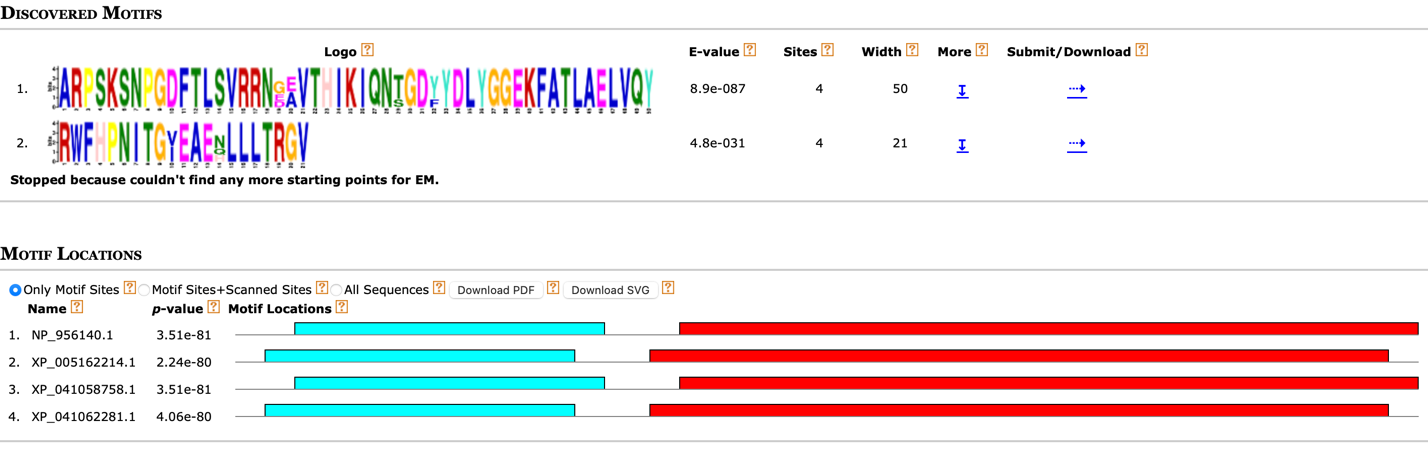
>XP\_041058758.1 tyrosine-protein phosphatase non-receptor type 11 [Carcharodon carcharias]

MTSRRWFHPNITGVEAENLLLTRGVDGSFLARPSKSNPGDFTLSVRRNGAVTHIKIQNTGDYYDLYGGEKFATLAELVQY

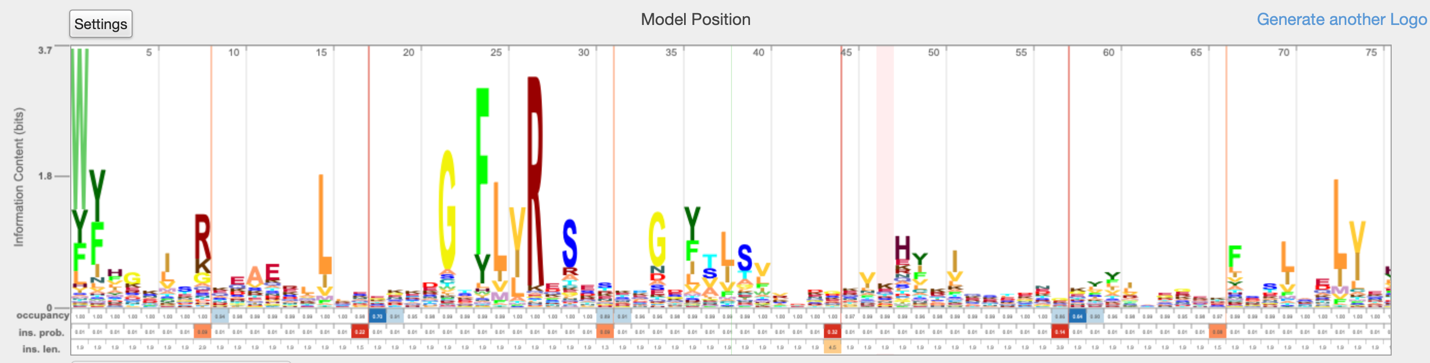
>XP\_041062281.1 tyrosine-protein phosphatase non-receptor type 11b isoform X3 [Carcharodon carcharias]

MVRWFHPNITGIEAEQLLLTRGVHGNFLARPSKSNPGDFTLSVRRNEEVTHIKIQNTGDFYDLYGGEKFATLAELVQYYT

Было ожидаемо, что изоформы почти одинаковые, поэтому у них найдены одинаковые участки, которые включают в себя не только SH2-домен, но и остальную гомологичную часть. Видно, что большая часть аминокислот совпадает, там где есть несовпадения – несколько букв по вертикали:



Сравним этот профиль с визуализированным hmm-профилем файла PF00017.hmm (с помощью сайта skylign.org), по которому мы делали поиск домена в протеомах:



Между профилем выравнивания и профилем SH2-домена больше всего совпадают консервативные **W, F,** I, G, E, A, E, **L**-аминокислоты в начале. Я думаю, что выравнивание по красному домену связано с тем, что выбранные белки высокогомологичны, а выравнивание по второму – именно с доменом SH2.