

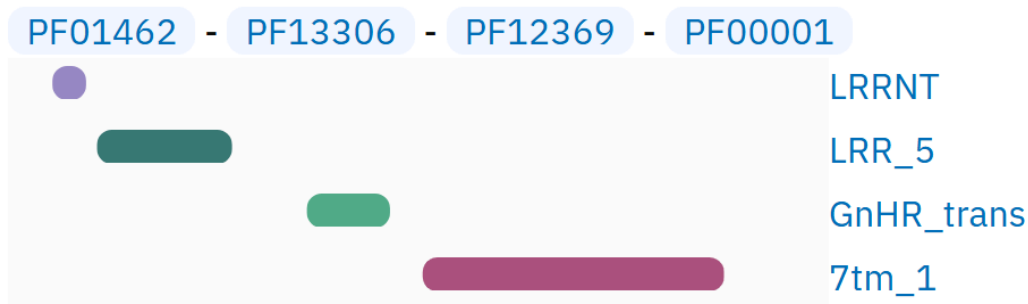
Семейство Pfam и его подсемейство

Для практикума выбрали семейство [PF00001](#) (7 transmembrane receptor (rhodopsin family)). Это семейство, помимо прочих G-белок-сопряжённых рецепторов (GPCR), включает представителей семейства опсинов, которые считаются типичными членами родопсиновой суперсемьи. Они разделяют несколько общих мотивов, главным из которых являются семь трансмембранных спиралей, характерных для GPCR родопсиновой суперсемьи. Все опсины связывают хромофор, например 11-цис-ретиналь. Функции большинства опсинов, за исключением фотоизомераз, разделены на два этапа: поглощение света и активация G-белка. Фотоизомеразы, напротив, не сопряжены с G-белками; предполагается, что они генерируют и поставляют хромофор, который используется зрительными опсинами.

- **Name** — 7 transmembrane receptor
- **AC** — PF00001
- **ID** — 7tm_1
- **Клан** — GPCR_A
- **Количество белков в seed** — 63
- **Количество белков в full** — 289903

Выберем подсемейство со следующей доменной архитектурой:

There are 66 proteins with this architecture (represented by [L5KTB5](#)):



Доменная архитектура подсемейства белков, выделенного в семействе на основе общей доменной архитектуры

Поиск сигналов и мотивов в представителях семейства

Последовательности 66 белков, входящих в данное подсемейство, были скачаны и выровнены при помощи программы MAFFT с параметрами по умолчанию.

В полученном выравнивании был выделен домен GnHR_trans — Gonadotropin hormone receptor transmembrane region.

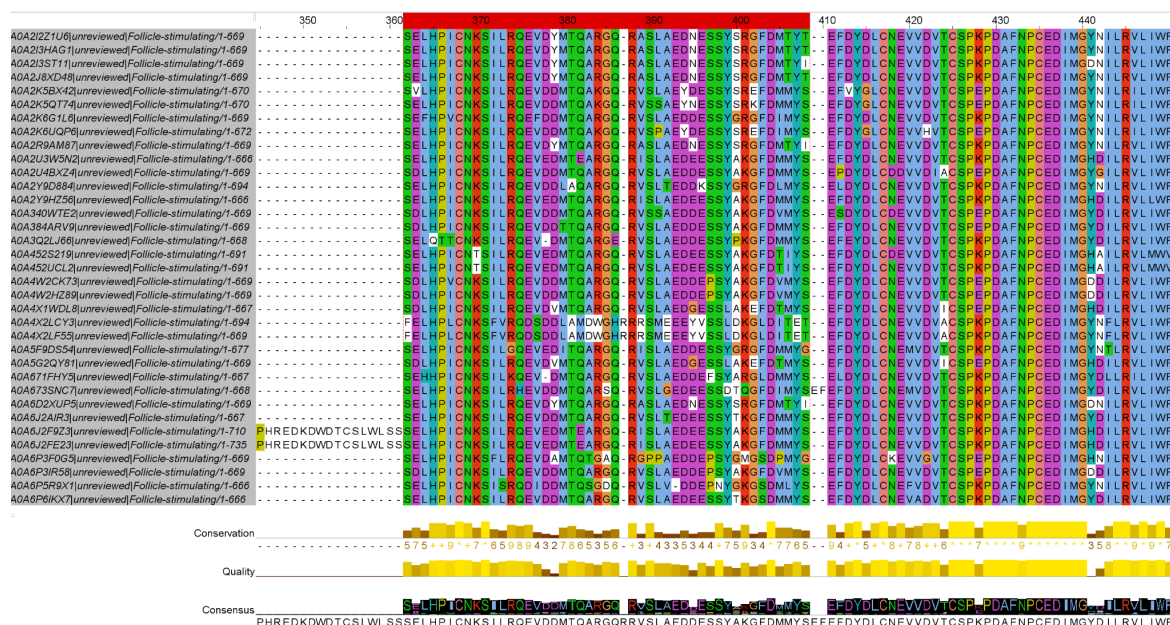


Рис. 2 Выравнивание белков рассматриваемого подсемейства, в котором выделен изучаемый домен(позиции 363-408)

Построение профиля HMM для подсемейства и определение порога для его выделения

С использованием программы hmmbuild по полученному выравниванию был построен HMM-профиль домена GnHR_trans.

hmmbuild --amino hmm.out GnHR_trans.fasta

Затем мы скачали все аннотированные белки из семейства 7 transmembrane receptor (1863 белка) и в них поискали последовательности, которые лучше всего соответствуют предоставленному профилю:

hmmsearch -o res_2.out hmm.out protein-matching-PF00001.fasta

В результате поиска профилем было обнаружено всего-то 15 последовательностей, соответствующих исследуемому подсемейству Follicle-stimulating hormone рецепторов. Значения score для найденных белков варьировались от 115.9 до 24.9. Наиболее достоверные совпадения

имели E-value порядка 10^{-35} – 10^{-20} , что свидетельствует о высокой специфичности профиля для данного подсемейства.

Минимальный score для уверенно идентифицируемых представителей подсемейства составил около 68–70. Однако использование такого порога приводило к появлению большого числа ложноположительных результатов, поскольку в выборку начинали попадать более удалённые представители семейства GPCR.

Было замечено, что большинство белков исследуемого подсемейства имеют score выше 90. Поэтому порог score ≥ 90 был выбран в качестве оптимального для выделения подсемейства. Такой порог позволяет сохранить высокую чувствительность профиля и одновременно уменьшить количество ложноположительных совпадений. Несмотря на то, что данный порог не является абсолютно строгим, среди протестированных значений он обеспечивает наилучший баланс между чувствительностью и специфичностью поиска.

	True	False
Positive	10	0
Negative	1863	5