

Практикум 8

Задание 1.

В первом задании я работала со структурой транскрипционного фактора FadR из *Escherichia coli*. Аффинность данного белка к молекуле ДНК контролируется путем связывания с белком длинноцепочечных ацетил-CoA молекул. Таким образом осуществляется контроль метаболизма жирных кислот в этой бактерии.

Сперва я попробовала предсказать наличие доменов методом SplitValue. С помощью этой функции можно предсказать остатки, которые находятся на границе двух доменов. Эта функция реализована в алгоритме DOMAK, который основывается на определении структурного домена, согласно которому остатки одного домена имеют много взаимодействий между собой и имеют мало взаимодействий с остатками других доменов.

Я реализовала эту функцию и получила значения SplitValue для всех остатков.

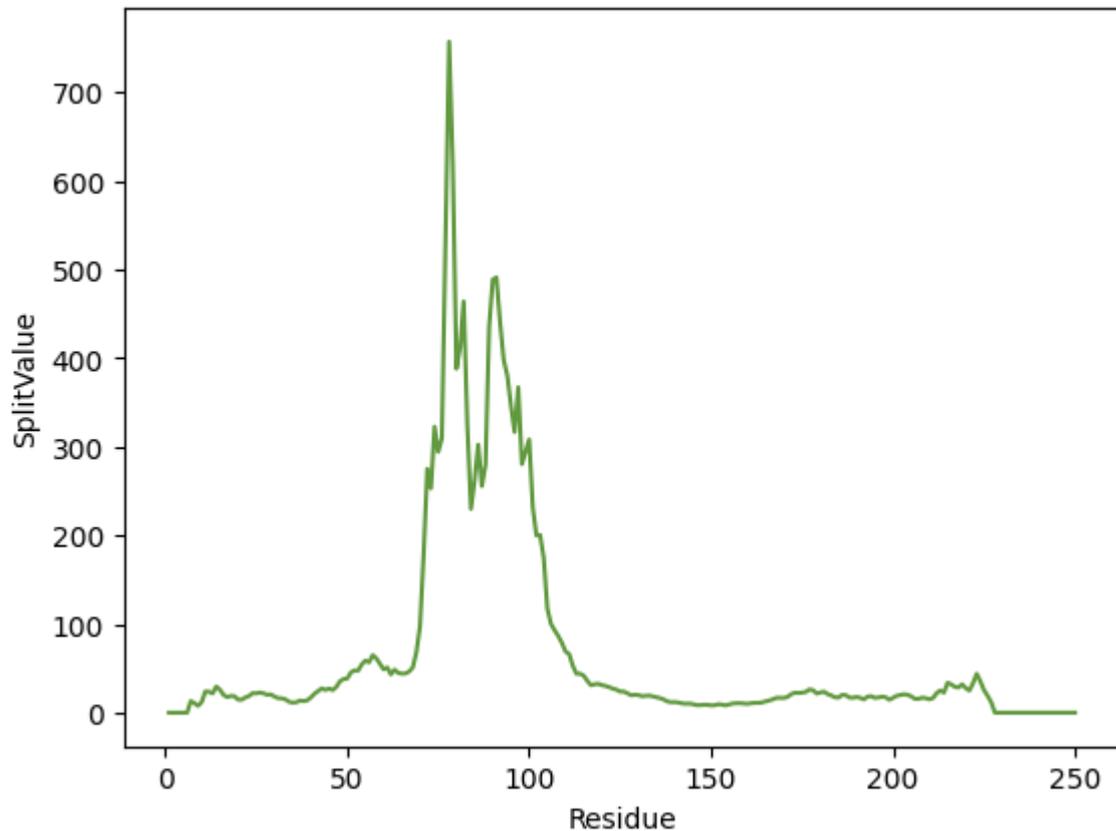


Рис. 1. SplitValue в зависимости от номера остатка. Два пика соответствуют остаткам 77 и 90.

Исходя из полученных результатов, я провела окраску по доменам для данной структуры. Результаты представлены на картинке ниже.

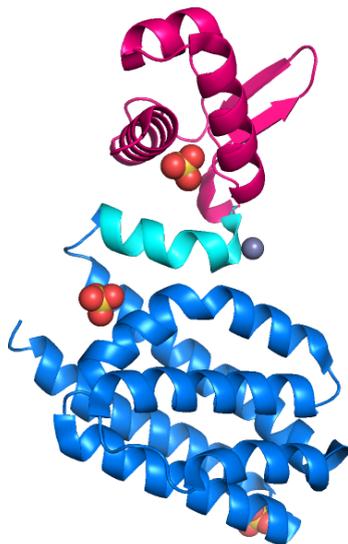


Рис. 2. Доменная организация структуры 1HW1, согласно SplitValue. Розовым отмечены остатки 1-77, голубым – 77-90, синим – 90-236.

Задание 2

Далее я рассмотрела выдачу баз данных CATH и SCOP.

По базе данных CATH оба домена преимущественно альфа. На рисунке ниже представлено разделение на домены согласно базе данных CATH.

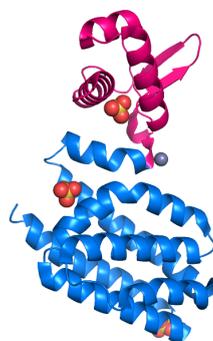
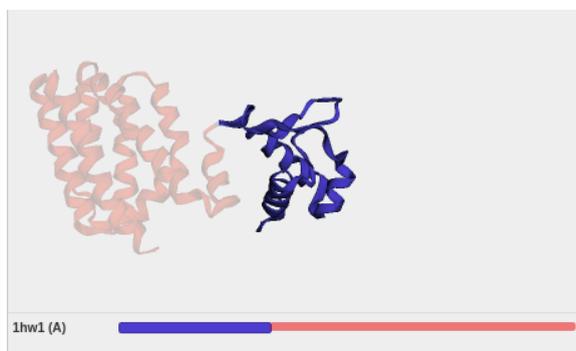


Рис. 3. Слева рисунок выдачи базы данных CATH. Справа – изображение в Pymol. Границы доменов – розовый – 5-79, синий – 80-230.

Далее я рассмотрела выдачу базы данных SCOP. Границы доменов по этой базе данных немного отличаются: для малого домена – 5-73, для большого – 92-230. Т. е. согласно базе данных SCOP, как и согласно алгоритму ДОМАК, выделяется 3 домена.

UniProtP0A8V6
This domain

MVIKAQSPAG FAEFYIIESI WNNRFPPGTI LPAERELSEL IGVTRTTLRE VLQRLARDGW LTIQH GKPTK VNNFWETSGL NILETLARLD
HESVPLIDN LLSVRTNIST IFIRTAFRQH PDKAQEVLAT ANEVADHADA FAELDYNIFR GLAFASGNPI YGLILNGMKG LYTRIGRHYF
ANPEARSLAL GFYHKLSALC SEGANDQVYE TVRRYGHESG EIWHRMQKNL PGDLAIQGR

Рис. 4. Выдача базы данных SCOP для малого домена. Границы домена– 5-73.

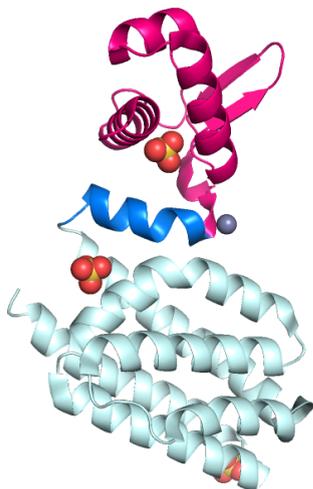


Рис. 5. Границы доменов согласно базе данных SCOP. Границы доменов: розовый– 5-73, синий– 73-92, голубой– 92-230.

Таким образом есть вероятность, что база данных CATN ошибается в расположении границ доменов.

Задание 3.

Далее я посмотрела выдачу базы данных InterPro.



Рис. 6. Выдача базы данных InterPro.

База данных нашла в последовательности 2 домена: GntR – бактериальный регуляторный белок семейства GntR и FadR_C – C-концевой домен, похожий на

регуляторный белок метаболизма жирных кислот FadR. Первый домен обнаружен с 5 по 71 аминокислотные остатки, второй – с 72 по 234.

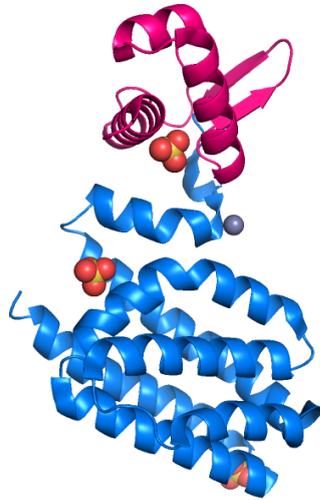


Рис. 7. Доменная организация согласно базе данных InterPro. Розовые – 5-71 остатки, синий – 72-234 остатки.

Выводы:

Все программы немного по-разному определяют границы доменов. Вероятнее всего в структуре присутствуют два домена, функции которых согласуются с доменными функциями в InterPro.