

## Техническое задание по проекту “Ugene” (Унипро)

**Цель.** Разработать инструмент в составе программы “Ugene”, позволяющий определить аминокислотные остатки полипептидных цепей компонентов (доменов) макромолекулярного комплекса, непосредственно участвующие в его образовании, исходя из кристаллографической структуры комплекса, записанной в файле формата pdb.

**Метод определения участвующих в образовании комплекса аминокислотных остатков.** Пусть имеется комплекс из  $i$  белковых доменов, не связанных ковалентно (или комплекс белок-ДНК), а полипептидная цепь каждого домена представлена  $n_i$  аминокислотными остатками или  $N_i$  атомами (включая боковые радикалы). Пусть  $\vec{E}_i = (E_{m_1} \dots E_{m_i})$ ,  $m = n$  или  $N$  — энергия сольватации доступных растворителю аминокислотных остатков (или атомов) полипептидной цепи  $i$ -того домена в составе комплекса (или комплекса ДНК-белок), а  $\vec{e}_i = (e_{m_1} \dots e_{m_i})$  — энергия сольватации  $i$ -того домена вне комплекса, но в той же конформации, как и в комплексе. Пусть  $\vec{\Delta}_i = (\Delta_{m_1} \dots \Delta_{m_i}) = \vec{E}_i - \vec{e}_i$ , тогда  $\Delta_{m_j} \neq 0$  ( $j = 1..i$ ) будет соответствовать аминокислотным остаткам (или атомам)  $i$ -того домена недоступным молекулам растворителя в составе комплекса, то есть лежащим в область взаимодействия доменов. Энергии  $\vec{E}_i$  могут быть определены из pdb-файлов, при помощи метода, предложенного в [1]. Энергии  $\vec{e}_i$  определяются аналогично  $\vec{E}_i$ , но используется pdb-файл комплекса, из которого удалена информация о всех доменах, исключая  $i$ -тый. Расчет энергий сольватации с использованием [1] реализован в программе GetArea (<http://curie.utmb.edu/getarea.html>), однако после расчета приходится вручную сравнивать энергии сольватации для каждого атома/аминокислоты, а затем в одной из структурных программ вручную выделять цветом определенные аминокислотные остатки.

### Задачи.

1. Включить в состав 3D-viewer инструмент, позволяющий «графически» редактировать pdb-файл (выделять и удалять из структуры элементы, а затем сохранить новый pdb-файл без удаленных элементов, для дальнейшего определения  $\vec{e}_i$ ).
2. Разработать инструмент, позволяющий рассчитать энергии  $\vec{E}_i$  и  $\vec{e}_i$ , а затем автоматически ищущий разницу между этими энергиями  $\vec{\Delta}_i$  и выделяющий в структуре комплекса аминокислотные остатки, для

которых  $\Delta_{m_j} \neq 0$  (с генерацией таблицы, в которой приводится соответствие этих  $\Delta_{m_j}$  аминокислотам или атомам).

### **Примечания.**

1. Не все определенные этим методом аминокислотные остатки будут участвовать в связывании, но все будут лежать в областях поверхностей белковых доменов, в результате взаимодействия которых происходит образование комплекса. Для определения конкретных аминокислот, участвующих в связывании нужен дополнительный анализ полученных данных.
2. При решении задачи 2 входными данными будут pdb-файлы комплекса и  $i$ -того домена (последний будет получен при решении задачи 1). Не стоит автоматизировать «разложение» комплекса на домены (к примеру, при помощи поиска не связанных ковалентно полипептидных цепей в комплексе), т.к. при кристаллизации комплекса возможно использование «неестественных» ковалентных сшивок (например, при получении комплексов ферментов репарации с неповрежденной ДНК).

### **Пример выполнения поставленной задачи при помощи программ PyMol (De Lano Scientific), GetArea и Microsoft Exel.**

Задача: определить аминокислотные остатки, лежащие в области связывания урацил-ДНК- $N$ -гликозилазы человека (UNG) с ДНК исходя из pdb-файла комплекса UNG с ДНК (№ в базе данных Protein Data Bank 1EMH).

#### Выполнение.

Загрузить в PyMol pdb-файл 1EMH, выделить цепи ДНК и удалить их (...→actions→remove atoms), сохранить как новый pdb-файл<sup>1</sup> (1EMHwitoutDNA.pdb).

Загрузить поочередно исходный и полученный файлы в программу GetArea (онлайн-версия: <http://curie.utmb.edu/getarea.html>). Полученные энергии сольватации скопировать в одну таблицу Microsoft Exel и для каждой аминокислоты взять разницу в энергиях сольватации (табл. 1). В PyMol в структуре 1EMH выделить цветом те аминокислотные остатки, для которых разность в энергиях сольватации отлична от нуля (рис. 1).

### **Ссылки**

---

<sup>1</sup> Также можно удалить информацию из pdb-файла «вручную» редактированием 1EMH в блокноте

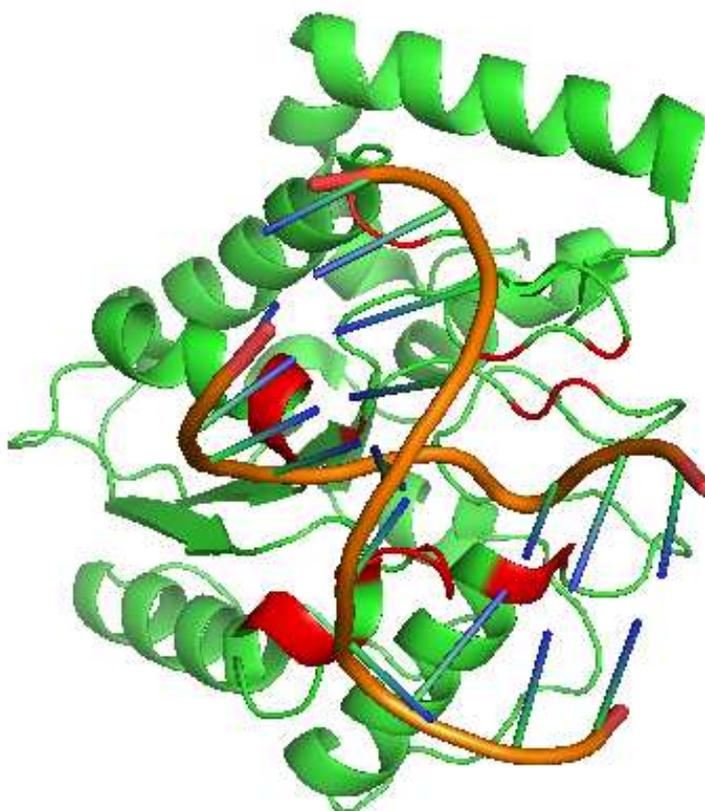
1. Fraczkiewicz, R. and Braun, W. (1998) "Exact and Efficient Analytical Calculation of the Accessible Surface Areas and Their Gradients for Macromolecules" J. Comp. Chem., 19, 319-333<sup>2</sup>

---

<sup>2</sup>Есть по адресу: [http://onlinelibrary.wiley.com/doi/10.1002/\(SICI\)1096-987X\(199802\)19:3%3C319::AID-JCC6%3E3.0.CO;2-W/abstract](http://onlinelibrary.wiley.com/doi/10.1002/(SICI)1096-987X(199802)19:3%3C319::AID-JCC6%3E3.0.CO;2-W/abstract)

DNA complex(1EMH.pdb)								w/o DNA (1EMHwitoutDNA.pdb)								Difference						
Residue	Res№	Total	Apolar	Backbone	Sidechain	Ratio(%)	In/Out	Residue	Res№	Total	Apolar	Backbone	Sidechain	Ratio(%)	In/Out	Residue	Res№	Total	Apolar	Backbone	Sidechain	Ratio(%)
VAL	140	0,0	0,0	0,0	0,0	0,0	i	VAL	140	0,0	0,0	0,0	0,0	0,0	i	VAL	140	0,0	0,0	0,0	0,0	0,0
ILE	141	0,6	0,6	0,0	0,6	0,4	i	ILE	141	0,6	0,6	0,0	0,6	0,4	i	ILE	141	0,0	0,0	0,0	0,0	0,0
LEU	142	0,0	0,0	0,0	0,0	0,0	i	LEU	142	0,0	0,0	0,0	0,0	0,0	i	LEU	142	0,0	0,0	0,0	0,0	0,0
GLY	143	0,0	0,0	0,0	0,0	0,0	i	GLY	143	0,0	0,0	0,0	0,0	0,0	i	GLY	143	0,0	0,0	0,0	0,0	0,0
GLN	144	9,2	2,9	0,5	8,8	6,1	i	GLN	144	40,2	3,1	0,5	39,7	27,6		GLN	144	30,9	-0,3	0,0	30,9	21,5
ASP	145	20,2	0,0	14,2	6,0	5,3	i	ASP	145	20,2	0,0	14,2	6,0	5,3	i	ASP	145	0,0	0,0	0,0	0,0	0,0
PRO	146	11,7	9,6	6,9	4,8	4,6	i	PRO	146	11,7	9,6	6,9	4,8	4,6	i	PRO	146	0,0	0,0	0,0	0,0	0,0
TYR	147	28,9	28,3	4,9	24,1	12,5	i	TYR	147	32,5	31,8	4,9	27,6	14,3	i	TYR	147	-3,5	-3,5	0,0	-3,5	-1,8
HIS	148	20,6	17,7	0,5	20,2	13,0	i	HIS	148	69,2	48,9	0,5	68,8	44,5		HIS	148	48,6	31,2	0,0	48,6	31,5
GLY	149	13,0	13,0	13,0	0,0	14,9	i	GLY	149	13,0	13,0	13,0	0,0	14,9	i	GLY	149	0,0	0,0	0,0	0,0	0,0

**Табл. 1.** Фрагмент таблицы Microsoft Exel с данными по энергиям сольватации (рассчитанным при помощи GetArea) для аминокислотных остатков комплекса 1EMH (UNG-ДНК), 1EMHwitoutDNA и разница этих энергий для разных аминокислот. Зеленым выделена разница, отличная от нуля.



**Рис. 1.** Аминокислотные остатки полипептидной цепи UNG, образующие ДНК-связывающую область фермента (выделено красным). На рисунке приведена структура комплекса UNG-ДНК 1EMH. ДНК с основаниями обозначены темно-оранжевыми «жгутами» и зелено-синими отрезками соответственно. Молекула UNG обозначена зеленым.