

ПОСТРОЕНИЕ ВОРОНОГО-ДЕЛОНЕ ДЛЯ МОДЕЛИРОВАНИЯ БЕЛОК-БЕЛКОВЫХ КОМПЛЕКСОВ

Анашкина А. А., Туманян В. Г.

(Россия, Москва)

Выявление аминокислот, играющих ключевую роль в образовании белок-белковых комплексов, необходимо для понимания природы этих комплексов. Мы решили задачу автоматического поиска контактирующих аминокислот в белок-белковых комплексах с помощью построения Вороного-Делоне. Был проведен расчет для 70 белок-белковых комплексов, и для оценки значимости каждого типа контакта была составлена матрица вероятности случайного его образования. Наибольшие отклонения от случайного распределения наблюдаются для контактов типа Leu-Val, Leu-Phe, Leu-Leu. Для белок-белковых комплексов одного типа (а именно антитело-антиген) наибольшие отклонения от случайного распределения наблюдаются для пар Leu-Phe, Phe-Ser.

Образование комплексов из двух или более белков является частью проблемы белок-белкового узнавания. В последнее время появилось несколько баз данных, содержащих информацию по взаимодействию между белками [1]. Причем это взаимодействие описывается на различных уровнях детализации и разрешения. Определение аминокислот, играющих ключевую роль в образовании белок-белковых комплексов, необходимо для понимания природы этих комплексов. Поэтому целью данного исследования было определение аминокислот, определяющих формирование комплексов. Для реализации этой цели мы воспользовались методом Вороного-Делоне для определения контактирующих аминокислот, сосредоточив свое внимание на анализе частоты встречаемости аминокислотных контактов.

Построение Вороного-Делоне и алгоритм программы. В основу созданной нами компьютерной программы, позволяющей однозначно определять контактирующие аминокислоты, а также площадь контакта между ними, было положено построение Вороного-Делоне. Вкратце суть построения сводится к сле-

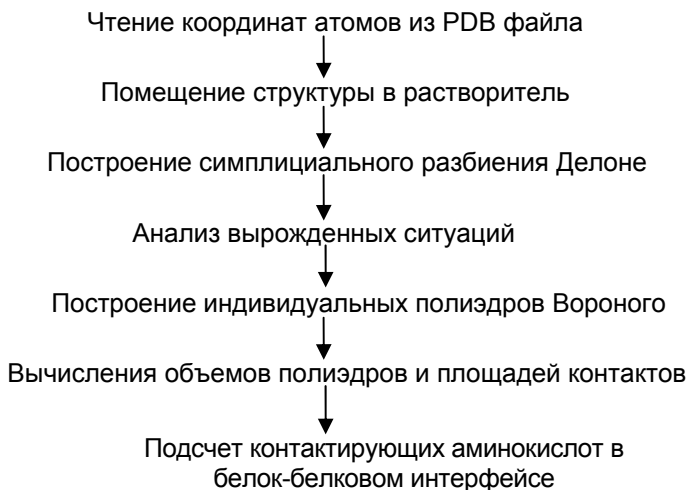
дующему. Для произвольной точки из системы точек $\{A\}$ можно указать область пространства, все точки которой ближе к данному центру, чем к любому другому центру системы [2]. Эта область называется областью Вороного. В трехмерном пространстве область Вороного любого центра i системы $\{A\}$ есть выпуклый многогранник, в двумерном — выпуклый многоугольник. Многогранники Вороного, построенные для каждого центра системы $\{A\}$, дают нам мозаику многогранников, которая называется *разбиением Вороного*.

Воспользуемся пустым шаром, который мы будем перемещать, изменяя его размер так, чтобы он мог касаться точек системы $\{A\}$, но всегда оставался пустым. Найденные четыре точки (i, j, k, l) определяют вершины тетраэдра, который характерен тем, что внутри его описанной сферы нет других точек системы $\{A\}$. Такой тетраэдр называется *симплексом Делоне*. Этот мысленный образ, предложенный Делоне, помогает увидеть проблему с новой стороны и яснее понять взаимосвязь и единство разбиений Вороного и Делоне.

Каждый центр системы $\{A\}$ определяет своих геометрических соседей двумя способами: посредством граней многогранника Вороного и посредством ребер симплексов Делоне. Те, в свою очередь, определяют своих соседей и т.д. Таким образом, применительно к координатам атомов, построение Вороного–Делоне однозначно определяет соседние атомы и строит для каждого атома индивидуальные ячейки (полиэдры) Вороного.

Как методом пустого шара Делоне, так и с помощью плоскостей Вороного мы выявляем одну и ту же систему точек. Каждый симплекс (многогранник мозаики Делоне) соответствует некоторой вершине мозаики Вороного, и наоборот, каждая вершина многогранника Вороного соотносится с определенным симплексом Делоне. Каждая грань любого симплекса Делоне однозначно связана с ребром мозаики Вороного. Каждое ребро мозаики Делоне соединяет пару геометрических соседей системы $\{A\}$, но та же самая пара определяет плоскость Вороного, на которой лежит соответствующая грань многогранника Вороного. Каждая вершина симплекса Делоне соотносится с некоторым своим многогранником Вороного. Вершина симплекса есть точка системы $\{A\}$, а каждая такая точка имеет свой многогранник Вороного. Эти построения называются дуальными, и они являются в топологическом смысле эквивалентными.

Мы воспользовались дуальностью разбиений Вороного и Делоне для написания программного алгоритма.



После нахождения всего ансамбля узлов Вороного и построения полиэдров для каждого атома вычислим площадь каждой грани и объем всего полиэдра. Для этого разобьем каждую грань, представляющую собой выпуклый многоугольник, на треугольники и найдем площадь грани $abcdef$ как сумму площадей треугольников abc , acd , ade и aef . Найдем объем полиэдра как сумму объемов пирамид с основаниями на гранях.

Объем каждой пирамиды V можно найти по формуле $V = \frac{1}{3}HS$, где S — площадь грани, а H — высота пирамиды, равная половине расстояния между атомами, между которыми проведена данная грань.

Основные соседи дают, как правило, самые большие грани многогранника. Неосновные — это обычно более далекие соседи, а их грани есть небольшие срезы вершин или ребер на многограннике Вороного.

Далее, суммируя объемы полиэдров отдельных атомов, мы вычислим объемы аминокислот, а, суммируя площади контактов между атомами, вычислим площади контактов между аминокислотами. Таким образом, мы написали программу, одно-

значно определяющую контактирующие аминокислоты, а также площадь контакта между ними.

Есть несколько принципиальных сложностей при работе с разбиением Вороного–Делоне. Первая состоит в том, что время работы почти всех алгоритмов с увеличением числа рассматриваемых атомов N растет как N^3 , а лучшие алгоритмы дают $N^2 \ln N$. Эта задача была решена следующим образом. Поскольку разбиение Вороного–Делоне является однозначным и не зависит от начальной точки разбиения, и для каждого атома существует некоторый эффективный радиус, дальше которого у него не обнаруживаются соседи, то вычисления можно локализовать, а именно выделить некоторое окружение для каждого атома и рассчитать его полиэдр, не рассматривая всего массива атомов. Понятно, что данный алгоритм будет сильно проигрывать при малом количестве атомов, но для больших N время работы программы линейно зависит от числа атомов, что позволяет существенно сократить время расчета. Вторая проблема заключается в создании граничных условий, позволяющих определить полиэдры Вороного не только для атомов внутри глобулы, но и для атомов на поверхности белка. Один из способов решения этой задачи — поместить белок в растворитель. Еще одна сложность — вырожденные ситуации, при которых существует несколько способов симплициального разбиения Делоне. Распространенный вариант решения данной проблемы — небольшое смещение, что снимает вырожденность ситуации. Однако эта задача была успешно решена другим способом. Поскольку при всех альтернативных разбиениях Делоне узлы Вороного определяются однозначно и не зависят от варианта разбиения Делоне, вся трудность заключается в том, чтобы вместо нескольких совпадающих центров описанной окружности для каждого симплекса Делоне для вырожденного случая мы видели один центр описанной окружности для всей вырожденной фигуры.

Таким образом, в нашей лаборатории разработана эффективная программа, производящая построение Вороного–Делоне и позволяющая получать информацию о соседях каждого атома, площади контактов с соседями, объемы полиэдров для каждого атома, площади контактов между аминокислотами и объемах аминокислот.

Возможности программы позволяют выписать результаты работы в виде следующих данных:

1. Индивидуальные полиэдры Вороного, построенные для каждого атома, с указанием соседей, площади взаимодействия с со-

седами, объема. Отдельный модуль программы позволяет получить графическое представление полиэдра.

2. Список взаимодействий между аминокислотами с указанием площадей контактов.
3. Список аминокислот с указанием площади поверхности и их объема.
4. Различные границы для графического представления — аминокислот, мест контактов, кластеров в виде соответствующих граней полиэдра.

Апробация программы. Для проверки компьютерной программы было проведено сравнение результатов расчета количества аминокислотных контактов и их площадей для комплекса рибонуклеазы Sa с барстаром (PDB код 1au7), полученных с помощью созданной нами программы, с соответствующими данными базы BIND [3] (www.bind.ca). При расчете предполагалось, что контакт между аминокислотами существует, если его площадь больше 3 \AA^2 , поскольку на гистограмме площадей контактов между аминокислотами в области 3 \AA^2 существует локальный минимум, разделяющий малые по площади контакта от остальных.

Как видно из Табл. 1, программа, основанная на методе Вороного–Делоне, выявила 32 аминокислотных контакта, из которых 28 указаны в базе BIND. Таким образом, кроме всех контактов, имеющих в базе BIND, мы видим еще 4 контакта, отсутствующих в этой базе. Три из 4 контактов, не указанных в базе BIND, характеризуются значительными площадями, что свидетельствует о высокой вероятности их реализации.

Определение контактирующих аминокислот. На следующем этапе исследования был проведен расчет количества контактов для 70 белок-белковых комплексов. Подборка этих комплексов была взята из работы [4] и включала 17 комплексов типа антиген-антитело, 8 ферментных комплексов, 11 G-белков, 5 больших протеазных комплексов, 19 комплексов типа протеазингибитор и еще 12 комплексов других типов. (PDB коды этих комплексов: 1a2k, 1acb, 1agr, 1ak4, 1ao7, 1atn, 1avw, 1brs, 1bth, 1cbw, 1cho, 1cse, 1dan, 1dfi, 1dhk, 1dkg, 1dvh, 1ebp, 1efn, 1efu, 1fbi, 1fc2, 1fin, 1fle, 1fss, 1gg2, 1gla, 1got, 1gua, 1hia, 1hwg, 1iai, 1igc, 1jhl, 1kb5, 1mct, 1mel, 1mkw, 1mlc, 1nca, 1nfd, 1nmb, 1nsn, 1osp, 1ppf, 1seb, 1stf, 1tbq, 1tco, 1tgs, 1toc, 1tx4, 1udi, 1vfb, 1ycs, 1ydr, 2btf, 2jel, 2kai, 2pcc, 2ptc, 2sic, 2sni, 2trc, 3sgb, 3tpi, 4cpa, 4htc). Результаты представлены в Табл. 2.

Раздел 8. Математические методы в биологии, экологии и химии
Part 8. Mathematical methods in biology, ecology and chemistry

Таблица 1. Контакты в комплексе рибонуклеазы Sa с барстаром (1au7), полученные методом Вороного–Делоне. Указаны контактирующие остатки, их номера, площадь контакта в Å² и наличие контакта в базе BIND

29	TYR	64	THR	11.68	Указан в BIND
29	TYR	66	GLY	18.14	Указан в BIND
29	TYR	67	THR	13.01	Указан в BIND
29	TYR	85	HIS	18.13	
30	TYR	85	HIS	8.45	Указан в BIND
31	GLY	85	HIS	13.73	Указан в BIND
33	ASN	41	GLU	11.49	Указан в BIND
33	ASN	85	HIS	7.30	Указан в BIND
33	ASN	86	TYR	9.68	Указан в BIND
33	ASN	87	ALA	5.77	Указан в BIND
34	LEU	40	ARG	22.72	Указан в BIND
34	LEU	41	GLU	17.86	Указан в BIND
35	ASP	37	PHE	11.09	Указан в BIND
35	ASP	39	ASN	4.72	
35	ASP	40	ARG	17.09	Указан в BIND
35	ASP	41	GLU	9.98	Указан в BIND
35	ASP	86	TYR	13.09	Указан в BIND
36	ALA	85	HIS	18.87	Указан в BIND
36	ALA	86	TYR	7.44	Указан в BIND
38	TRP	40	ARG	20.91	Указан в BIND
39	ASP	32	GLN	11.53	Указан в BIND
39	ASP	65	ARG	8.56	Указан в BIND
39	ASP	69	ARG	8.45	Указан в BIND
39	ASP	85	HIS	10.26	Указан в BIND
39	ASP	86	TYR	11.63	Указан в BIND
40	CYS	85	HIS	7.80	Указан в BIND
42	THR	32	GLN	10.11	Указан в BIND
43	GLY	32	GLN	5.52	Указан в BIND
44	TRP	64	THR	7.47	Указан в BIND
73	VAL	40	ARG	12.48	
76	GLU	38	GLN	11.89	
76	GLU	40	ARG	19.54	Указан в BIND

Таблица 2. Числа контактов между аминокислотами в полной выборке

	ALA	ARG	ASN	ASP	CYS	GLN	GLU	GLY	HIS	ILE	LEU	LYS	MET	PHE	PRO	SER	THR	TRP	TYR	VAL
ALA	36	39	27	30	26	32	46	54	24	35	79	28	30	59	37	46	28	23	47	48
ARG	39	68	81	110	48	44	147	120	41	45	103	55	30	64	67	84	63	95	94	51
ASN	27	81	52	41	8	36	69	56	49	28	66	33	19	52	43	45	83	33	79	46
ASP	30	110	41	22	16	35	29	38	40	32	63	101	16	34	39	68	33	22	76	25
CYS	26	48	8	16	58	27	17	38	28	45	45	30	17	17	15	43	21	25	34	32
GLN	32	44	36	35	27	72	52	75	16	40	81	49	13	50	31	54	58	23	96	49
GLU	46	147	69	29	17	52	70	51	39	50	62	141	25	70	34	56	60	37	85	50
GLY	54	120	56	38	38	75	51	118	58	39	102	56	42	63	56	63	42	59	84	61
HIS	24	41	49	40	28	16	39	58	20	29	41	22	21	24	25	42	53	16	65	25
ILE	35	45	28	32	45	40	50	39	29	42	97	27	23	70	31	52	33	39	45	61
LEU	79	103	66	63	45	81	62	102	41	97	206	58	55	193	73	103	105	77	110	217
LYS	28	55	33	101	30	49	141	56	22	27	58	22	22	38	36	81	67	30	60	30
MET	30	30	19	16	17	13	25	42	21	23	55	22	34	51	24	15	39	19	38	37
PHE	59	64	51	34	17	50	70	63	24	70	193	38	51	116	65	132	97	91	84	71
PRO	37	67	43	39	15	31	34	56	25	32	73	36	24	65	62	88	56	97	77	38
SER	46	83	45	68	43	54	56	63	42	52	102	81	15	132	88	78	85	53	124	70
THR	28	63	83	33	21	58	60	42	53	33	104	67	39	97	56	85	98	43	67	61
TRP	23	95	33	22	25	23	37	59	16	39	77	30	19	91	97	53	43	34	89	48
TYR	47	94	79	77	34	96	84	84	65	45	110	60	38	84	77	124	67	88	104	60
VAL	48	51	46	25	32	49	50	61	25	61	218	30	37	71	38	69	61	48	60	84

Таблица 3. Количества аминокислот, участвующих в формировании белок-белкового интерфейса (средняя строка) и полное число аминокислот данного типа в полной выборке (нижняя строка)

	ALA	ARG	ASN	ASP	CYS	GLN	GLU	GLY	HIS	ILE	LEU	LYS	MET	PHE	PRO	SER	THR	TRP	TYR	VAL
(средняя строка)	330	458	412	441	248	381	516	571	265	300	589	419	173	385	404	607	499	272	488	453
(нижняя строка)	2656	1768	1885	2176	1133	1597	2248	3128	815	1932	3157	2360	712	1470	1865	3387	2786	796	1594	2889

Из Табл. 2 видно, что возможно образование контактов любых типов, а полученная информация о количестве контактов в принципе позволяет определить приоритетность образования тех или иных пар. Расчет суммарных площадей аминокислотных остатков показал, что средние значения площадей контактов лежат в диапазоне от 7 до 16 Å².

Сопоставление числа аминокислот, участвующих в формировании белок-белкового интерфейса, с полным числом аминокислот в данной PDB выборке свидетельствует о вариабельности количества аминокислот разных типов, участвующих в связывании (Табл. 3).

Для оценки значимости того или иного вида контакта была составлена матрица вероятности случайного образования контакта для известного количества аминокислот, элементы которой (P_{ij}) вычислялись по следующей формуле:

$$P_{ij} = \frac{N_i N_j}{S(S-1)} \quad (i \neq j), \quad P_{ii} = \frac{N_i(N_i-1)}{S(S-1)} \quad (i = j),$$

где N_i — число аминокислот типа i , а $S = \sum_i N_i$.

Таким образом, была получена матрица вероятности случайного образования контактов. Сопоставление этой матрицы с матрицей количества аминокислотных контактов, нормированной на общее количество интерфейсных контактов, позволило получить отличия от случайного распределения (рис. 1).

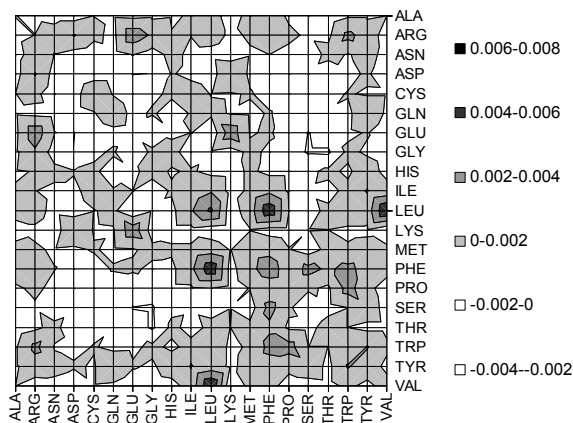


Рис. 1. Отклонения от вероятности случайного образования контакта

Для наглядности можно построить двумерную гистограмму, где на оси абсцисс — последовательность аминокислот, а по оси ординат — отклонение от случайного распределения.

Из рисунка 1 видно, что наибольшие отклонения от случайного распределения наблюдаются для контактов типа Leu-Val, Leu-Phe, Leu-Leu.

Насколько изменятся результаты, если рассмотреть белок-белковые комплексы одного типа? Рассмотрим отдельно результаты, полученные для 17 комплексов типа антиген-антитело. Поскольку выборка комплексов типа антиген-антитело значительно меньше, чем полная выборка из 70 комплексов, в ней есть «запреты»: пары ни разу не контактирующих аминокислот, например Ala-Met, Ile-Gln и много других (Табл.4). В Табл. 5 приведены количества аминокислот, участвующих в формировании белок-белкового интерфейса и полное число аминокислот данного типа в 17 комплексах типа антиген-антитело.

Составив для этих чисел матрицу вероятности случайного образования пары, получим отклонения от вероятности случайного образования контакта между двумя аминокислотами (рис. 2).

Хорошо видно, что наибольшие отклонения от случайного распределения наблюдаются для пар Leu-Phe, Phe-Ser.

Таблица 4. Числа контактов между аминокислотами в 17 комплексах типа антиген-антиген

	ALA	ARG	ASN	ASP	CYS	GLN	GLU	GLY	HIS	ILE	LEU	LYS	MET	PHE	PRO	SER	THR	TRP	TYR	VAL
ALA	8	7	4	3	1	4	5	9	4	9	7	8	0	2	1	1	3	1	1	3
ARG	7	1	1	2	2	2	1	2	5	1	1	1	1	1	1	2	2	3	4	5
ASN	4	1	2	1	3	3	7	1	3	4	1	9	6	2	1	2	3	1	4	1
ASP	3	2	1	8	2	6	1	1	1	7	1	3	3	6	6	2	1	1	4	2
CYS	1	2	3	2	1	0	2	3	0	0	0	0	0	0	1	3	1	0	0	2
GLN	4	2	3	6	0	3	3	1	2	0	3	2	3	2	6	3	1	8	7	5
GLU	5	1	7	1	2	3	8	9	2	1	1	3	2	2	1	1	6	8	4	4
GLY	9	2	1	1	3	1	9	4	6	6	1	6	7	3	1	3	1	1	4	8
HIS	4	5	3	1	0	2	2	6	2	1	3	2	5	2	2	1	1	1	1	0
ILE	9	1	4	7	0	0	1	6	1	0	1	3	0	4	1	6	4	1	1	4
LEU	7	1	1	1	0	3	1	1	3	1	3	1	6	1	3	4	3	1	6	6
LYS	8	1	9	3	0	2	3	6	2	3	1	6	3	1	4	2	2	1	3	2
MET	0	1	6	3	0	3	2	7	5	0	6	3	0	2	2	2	6	1	1	3
PHE	2	1	2	6	0	2	2	3	2	4	1	1	2	5	3	9	4	5	4	2
PRO	1	1	1	6	1	6	1	0	2	1	3	4	2	3	1	3	1	5	3	2
SER	1	2	2	2	3	3	1	3	1	6	4	2	2	9	3	4	3	3	7	3
THR	3	2	3	1	1	1	6	1	1	4	3	2	6	4	1	3	5	2	2	8
TRP	1	3	1	1	0	8	8	1	1	1	1	1	1	5	5	3	2	1	6	1
TYR	1	4	4	4	0	7	4	4	1	1	6	3	1	4	3	7	2	6	7	2
VAL	3	5	1	2	2	4	4	8	0	4	6	2	3	2	2	3	8	1	2	2

Таблица 5. Числа аминокислот, участвующих в формировании белок-белкового интерфейса (средняя строка) и полное число аминокислот данного типа в 17 комплексах типа антиген-антитело (нижняя строка)

	ALA	ARG	ASN	ASP	CYS	GLN	GLU	GLY	HIS	ILE	LEU	LYS	MET	PHE	PRO	SER	THR	TRP	TYR	VAL
	63	93	143	124	24	138	102	173	56	44	173	108	38	126	140	244	175	102	227	99
	717	474	575	580	326	530	515	948	190	460	811	667	162	377	580	1477	1096	298	575	793

Как видно из Табл. 6, для полной выборки из 70 комплексов и подвыборки из 17 комплексов типа антиген-антитело наблюдаются сходные закономерности, однако можно видеть отличия в абсолютном значении отклонения от случайного распределения для пар Leu-Phe, Phe-Trp, Trp-Pro в два раза, что позволяет предположить существование предпочтения определенных аминокислот в интерфейсах комплексов одного типа. Это утверждение, однако, требует дальнейшей проверки.

Для проверки гипотезы о случайном распределении частоты образования пар контактов мы воспользовались критерием χ^2 Пирсона. Величина χ^2 , рассчитанная для первой выборки, оказалась равной 0.2, а для подвыборки антиген-антитело 0.6, и, таким образом, можно утверждать, что статистика образования пар контактов в белок-белковых интерфейсах не является случайной.

Полученные результаты свидетельствуют, что наибольшее отклонение от случайного распределения наблюдается для контактов типа Leu-Val, Leu-Phe, Leu-Leu, то есть для крупных алифатических аминокислот, причем валин и фенилаланин имеют разветвление при C_β атоме. Для конкретного примера взаимодействий в группе антитело-антиген наибольшие отклонения от случайного распределения наблюдаются для пар Leu-Phe, Phe-Ser.

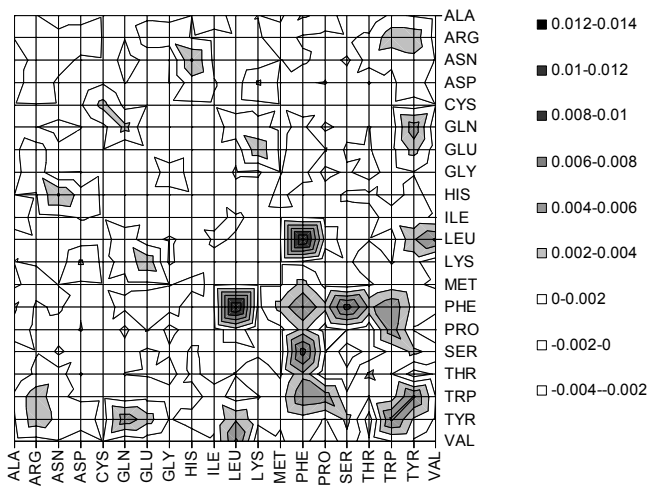


Рис. 2. Отклонения от вероятности случайного образования контакта для 17 комплексов типа антиген-антитело

Таблица 6. Сравнительная таблица максимальных отклонений от случайного распределения для двух выборок.

	Полная выборка из 70 комплексов		17 комплексов типа антиген-антитело
Leu-Val	0.006	Leu-Phe	0.012
Leu-Phe	0.006	Phe-Ser	0.009
Leu-Leu	0.004	Leu-Val	0.006
Glu-Lys	0.003	Phe-Phe	0.006
Arg-Glu	0.003	Phe-Trp	0.006
Trp-Pro	0.003	Trp-Tyr	0.006
Trp-Phe	0.003	Trp-Pro	0.005
Arg-Trp	0.003	Tyr-Gln	0.005
Arg-Asp	0.002	Tyr-Leu	0.005
Asp-Lys	0.002	Trp-Arg	0.004

Работа проводилась при поддержке гранта РФФИ 03-04-49017 и программы Молекулярная и клеточная биология.

Список литературы:

1. Xenarios I, Eisenberg D. Protein interaction databases // Curr Opin Biotechnol. 2001. V.12. №4. P.334–339.
2. Медведев Н.Н. Метод Вороного-Делоне в исследовании структуры некристаллических систем. – Новосибирск, СО РАН, 2000. 214с.
3. Bader G.D., Betel D., Hogue C.W. BIND: the Biomolecular Interaction Network Database // Nucleic Acids Res. 2003. V.31. №1. P.248–50.
4. Lo Conte L., Chothia C., Janin J. The atomic structure of protein-protein recognition sites // J Mol Biol. 1999. V.285 (Feb 5). №5. P.2177–2198.

**VORONOI –DELAUNE TESSELLATION FOR PROTEIN-
PROTEIN COMPLEXES MODELLING**

Anashkina A. A., Tumanyan V. G.

(Russia, Moscow)

Revelation of amino acids, playing a crucial role in protein-protein complex formation is essential for understanding its nature. We have solved problem of automatic retrieval information about contact amino acids in protein-protein complexes by Voronoi-Delaune tessellation. Computations were carried out for 70 protein-protein complexes. Probability matrix of chance contact formation was elucidated for valuation validity of different types of the contacts. The main deviation from chance distribution is observed for contacts types Leu-Val, Leu-Phe, Leu-Leu. For complexes of concrete type (antibody-antigen) the most deviation from chance distribution is observed for contacts types Leu-Phe, Phe-Ser.