

БЕЛОК-БЕЛКОВОЕ УЗНАВАНИЕ: ПРОВЕРКА НЕКОТОРЫХ ГИПОТЕЗ О СТРУКТУРЕ КОНТАКТНЫХ ОБЛАСТЕЙ В КОМПЛЕКСАХ БЕЛКОВ НА МАТЕРИАЛЕ БРУКХЕЙВЕНСКОГО БАНКА ДАННЫХ

Л.Н.ДРОЗДОВ-ТИХОМИРОВ*, Д.М.ЛИНДЕ**, В.В.ПОРОЙКОВ**,
М.А.МОКУЛЬСКИЙ*.

*Институт молекулярной генетики РАН, пл. Курчатова, Москва, 123128.

**Институт биомедицинской химии РАМН, Погодинская 10, Москва, 119832.

Эл. почта: dlinde@ru.

Проверены 4 гипотезы о структуре контактов в комплексах белков на 122 комплексах негомологичных белков из Брукхейвенского банка данных. Ни одна из гипотез не получила подтверждения.

Ключевые слова: комплекс белок-белок, белок-белковое узнавание, контактирующие аминокислоты, комплементарные аминокислоты (Брукхейвенский банк данных).

ВВЕДЕНИЕ. Гипотеза о комплементарности аминокислот в комплексах пептидов, предложенная Меклером в 1969 году [1-2], предполагает образование стабильного комплекса между пептидами, прочитанными по комплементарным цепям ДНК. Было предложено два способа построения гипотетических матриц комплементарности аминокислотных остатков на основе генетического кода [3,4]. Каждому триплету генетического кода можно приписать комплементарный ему триплет и инвертированный комплементарный триплет (прочтение в направлении 5'-3'). Таким образом, каждому аминокислотному остатку ставят в соответствие один или несколько "комплементарных" аминокислотных остатков, кодируемых комплементарным (гипотеза 1) триплетом или инвертированным (гипотеза 2) комплементарным триплетом ДНК.

Особенностью обеих этих матриц комплементарности является то, что при переходе от любой аминокислоты к "комплементарной" ей, значения гидрофобности/гидрофильности остатков изменяются на противоположные. Так как второй нуклеотид в триплете не меняется при инверсии, все триплеты с тиминном во втором положении, будут комплементарны триплетам с остатком аденина во втором положении и наоборот. Так как триплеты с тиминном во втором положении кодируют гидрофобные аминокислотные остатки (L, F, I, M, V), а триплеты с аденином - гидрофильные (Y, H, Q, N, K, D, E), то при прочтении комплементарного триплета в любом направлении получающаяся "комплементарная" аминокислота будет иметь гидрофобность/гидрофильность противоположную той, которую имеет исходная аминокислота.

В литературе опубликованы данные об энергиях взаимодействия между некоторыми "комплементарными" пептидами и белками (построенными по смысловой и антисмысловой цепям ДНК). В частности, такие данные получены для АКТИГ ($K_d=3 \cdot 10^{-9}M$) [5,6], S-пептида рибонуклеазы ($8 \cdot 10^{-7}M$) [3], инсулина ($K_d=3,1 \cdot 10^{-9}M$) [7], соматостатина [8], и др. [1]. Эти

данные можно считать подтверждением гипотезы Меклера о существовании комплементарного взаимодействия аминокислот в комплексах пептидов.

В 1995 г. Blalock и Smith предложили строить комплементарные пептиды не по комплементарным, а по инвертированным триплетам [9]. Тогда второй нуклеотид в триплете остается не измененным. В построенной на такой основе матрице комплементарности (3-я гипотеза) гидрофобным аминокислотным остаткам будут комплементарны гидрофобные, а гидрофильным - гидрофильные.

Генетический код является вырожденным. В частности, замена последнего нуклеотида в триплете часто не приводит к изменению кодируемого аминокислотного остатка [10]. Так как генетический код мог быть вначале двухбуквенным, мы предположили, что комплементарными могут быть аминокислотные остатки, построенные по инвертированным двухбуквенным кодонам (4-я гипотеза). Согласно этой гипотезе комплементарный аминокислотный остаток, не обязательно имеет ту же гидрофобность/гидрофильность, что и исходная аминокислота.

Целью проводимых нами исследований является проверка гипотез белок-белкового узнавания путем сопоставление структур областей контактов в белок-белковых комплексах известной структуры (Брукхейвенский банк), со структурами этих областей, прогнозируемыми в соответствии с гипотезами.

Ранее [11] мы провели статистическую проверку двух гипотез о комплементарности (1- и 2-ой гипотез Меклера) и показали, что ни одна из них не обладает статистической достоверностью на суммарном материале по аминокислотным контактам, обнаруженных нами в выборке белковых комплексов из Брукхейвенского банка данных.

Полученный результат означал бы, что ни одна из гипотез Меклера не соответствует действительности, если бы то, что мы называем аминокислотным контактом в белок-белковом комплексе, полностью соответствовало тому, что под этим понимал Меклер.

Согласно его представлениям область контакта двух белков можно рассматривать как составленную из некоторого числа пар связанных аминокислот, одна из которых в каждой паре принадлежит одному белку комплекса, а другая другому. Используемый нами способ определения "контактных пар аминокислот" (по расстоянию между наиболее сближенными атомами) довольно часто дает для аминокислоты одного белка комплекса несколько "контактирующих" с ней аминокислот другого белка этого комплекса. В предыдущей работе все определенные таким методом контакты включались в рассмотрение и, фактически, считались равноценными, хотя, очевидно, что вклад во взаимодействие одних может весьма существенно отличаться от вклада других.

Если принять во внимание вышесказанное, то придется признать результат предыдущей работы [11] недостаточным для окончательного вывода о несостоятельности гипотез Меклера.

В настоящей работе для проверки гипотез мы решили применить другой подход, основанный на исследовании структуры области контакта каждого белок-белкового комплекса в отдельности.

МЕТОДИКА. Мы отобрали 122 белковых комплексов и белков, состоящих из нескольких субъединиц, из категории негомологичных белков (EMBL, Heidelberg) [12] Брукхейвенского банка данных (PDB, выпуск январь 1995) [13]. Используемая выборка содержит аминокислотные последовательности, имеющие менее 25% гомологии.

Для проведения работы было введено понятие контакта между аминокислотами в белковом комплексе. Контактom считали пару остатков, принадлежащих разным субъединицам или белкам, для которых хотя бы один атом одного из остатков расположен на расстоянии не более 2,8 Å по отношению хотя бы к одному атому другого остатка (радиус

молекулы воды $R=1,4 \text{ \AA}$). При таком определении контакта с одним аминокислотным остатком, находящимся на контактной поверхности одного из белков комплекса, "контактировало" от одного до шести аминокислотных остатков, находящихся на контактной поверхности другого белка комплекса. Каждый такой контакт включался в список контактных пар комплекса. Среднее число контактных пар на одну аминокислоту на комплекс колеблется около 3.

Для каждого (n - го) комплекса из выборки была построена матрица контактов между аминокислотными остатками этого комплекса $-|K|^n$ ($n=1,2,\dots,122$), размерностью 20×20 . Элемент K_{ij}^n матрицы был определен равным общему числу встречающихся в n -ом комплексе контактов между i -ой и j -ой аминокислотами (каждой из 20 аминокислот присвоен условный номер l ($l=1,2,\dots,20$)). Половина суммы всех элементов n -ой матрицы контактов $Mn = 1/2 * \sum_{i=1}^{20} \sum_{j=1}^{20} K_{ij}^n$, очевидно, равна числу всех контактов в n -ом комплексе

. Каждая гипотеза в свою очередь была представлена матрицей комплементарности $|H|^m$ ($m=1,2,3,4$), размерностью 20×20 . Элемент H_{ij}^m матрицы комплементарности равен 1, если согласно m -ой гипотезе j -я аминокислота комплементарна i -ой, и равен 0, если не комплементарна.

Первые строки матриц $|K|^n$ и $|H|^m$ можно рассматривать как вектора \vec{K}_i^n и \vec{H}_i^m . Нетрудно видеть, что половина суммы скалярных произведений $(\vec{K}_i^n \times \vec{H}_i^m)$ равна числу контактов в n -ом комплексе, являющихся комплементарными согласно m -ой гипотезе.

$$F_{nm} = 1/2 \times \sum_{i=1}^{20} (\vec{K}_i^n * \vec{H}_i^m)$$

Знание величин Mn и F_{nm} для комплекса позволяет судить о соответствии структуры межбелкового контакта в этом комплексе m -ой гипотезе. При этом мы должны принять во внимание, что согласно современным представлениям число связей, обеспечивающих существование белок-белкового комплекса не может быть меньше 2 [14]. Очевидно, что зная величины F_{nm}/Mn и F_{nm} можно сделать одно из двух заключений:

1) если $F_{nm} \geq 2$ и $F_{nm}/Mn \geq 0,333$, то структура межбелкового контакта в n -ом комплексе не противоречит m -ой гипотезе.

2) если $F_{nm} < 2$ или $F_{nm}/Mn < 0,333$, то структура межбелкового контакта n -го комплекса противоречит m -ой гипотезе

РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ. Для каждого из 122 белковых комплексов были рассчитаны величины Mn , равные общему числу пар аминокислот, находящихся в области контакта данного комплекса ($n=1,2,\dots,122$).

Для каждой из четырех рассматриваемых гипотез, представленных 4-мя матрицами комплементарности $|H|^m$ ($m=1,2,3,4$) была определена последовательность величин F_{nm} ($n=1,2,\dots,122$), каждая из которых равна числу комплементарных по данной гипотезе пар аминокислот, находящихся в области межбелкового контакта n -го комплекса.

В таблице 1 для каждой из четырех гипотез представлены данные о числе пар комплементарных (по m -ой гипотезе) аминокислот среди контактных пар n -го комплекса - F_{nm} , и их доле от общего числа контактных пар в этом комплексе - F_{nm}/Mn . В таблице 1 внизу суммировано число комплексов, существование которых противоречит каждой из гипотез.

Как видно из таблицы, структуры межбелковых контактов всех 122 проанализированных белковых комплексов противоречат первой и третьей гипотезам. Вторую гипотезу опровергают все комплексы, кроме единственного 2dgr_A, структуру контакта которого можно рассматривать как непротиворечивую второй гипотезе. В случае четвертой гипотезы также найден всего один комплекс 2zta_A из 122 обследованных, структура контакта которого не противоречит гипотезе. Практически все обследованные комплексы имеют структуру контактных областей не соответствующую ни одной из гипотез. Таким образом, для каждой гипотезы найдено не по одному опровергающему примеру, что уже было бы достаточно для их опровержения, а значительно больше.

Таблица. Количество реальных и гипотетических контактных АА-пар в межбелковых контактах белок-белковых комплексов из Бруксейвенской базы данных.

Код белка в БХ - базе данных	Общее число контактов	Гипотеза 1		Гипотеза 2		Гипотеза 3		Гипотеза 4	
		Mi	Fi1	Fi1/Mi	Fi2	Fi2/Mi	Fi3	Fi3/Mi	Fi4
1alk_a	432	37	0.1	64	0.1	40	0.1	94	0.2
1aoz_a	59	5	0.1	4	0.1	1	0.0	9	0.2
1apm_e	96	5	0.1	19	0.2	6	0.1	14	0.1
1atn_a	93	5	0.1	13	0.1	6	0.1	25	0.3
1aya_a	60	6	0.1	7	0.1	4	0.1	7	0.1
1azc_a	48	6	0.1	6	0.1	14	0.3	4	0.1
1bab_b	51	2	0.0	8	0.2	3	0.1	5	0.1
1bar_a	100	11	0.1	15	0.2	10	0.1	24	0.2
1bbp_a	112	1	0.0	8	0.1	8	0.1	15	0.1
1bcf_a	73	15	0.2	8	0.1	14	0.2	23	0.3
1bge_b	13	1	0.1	1	0.1	2	0.2	3	0.2
1bmd_a	166	7	0.0	19	0.1	23	0.1	15	0.1
1bsa_a	8	0	0.0	2	0.3	1	0.1	0	0.0
1cau_a	141	13	0.1	19	0.1	12	0.1	33	0.2
1cdt_A	47	0	0.0	4	0.1	2	0.0	14	0.3
1chm_A	322	14	0.0	26	0.1	32	0.1	61	0.2
1cmc_A	169	4	0.0	20	0.1	16	0.1	49	0.3
1col_a	56	0	0.0	5	0.1	10	0.2	15	0.3
1cpc_A	282	12	0.0	41	0.1	15	0.1	64	0.2
1cpc_B	165	7	0.0	19	0.1	10	0.1	34	0.2
1csk_A	77	4	0.1	4	0.1	8	0.1	19	0.2
1cta_a	65	3	0.0	8	0.1	6	0.1	25	0.4
1d66_A	51	5	0.1	11	0.2	12	0.2	18	0.4
1dfn_A	45	5	0.1	5	0.1	4	0.1	11	0.2
1dsb_A	62	7	0.1	8	0.1	7	0.1	7	0.1
1epa_b	46	2	0.0	4	0.1	8	0.2	11	0.2
1fba_A	265	18	0.1	40	0.2	23	0.1	41	0.2
1fc2_c	57	3	0.1	4	0.1	4	0.1	10	0.2
1fcd_C	164	14	0.1	26	0.2	9	0.1	27	0.2
1frp_A	268	24	0.1	46	0.2	32	0.1	52	0.2
1gd1_O	421	25	0.1	36	0.1	56	0.1	102	0.2
1gdh_A	330	28	0.1	60	0.2	24	0.1	49	0.1
1ghs_A	21	3	0.1	6	0.3	2	0.1	4	0.2
1glc_G	51	0	0.0	5	0.1	5	0.1	1	0.0
1gmf_A	27	1	0.0	8	0.3	0	0.0	4	0.1
1qp1_A	66	6	0.1	12	0.2	5	0.1	9	0.2

1hcn_A	214	20	0.1	41	0.2	18	0.1	34	0.2
1hgl_A	481	24	0.0	61	0.1	30	0.1	93	0.2
1hle_A	263	14	0.1	22	0.1	29	0.1	67	0.3
1hmc_A	17	1	0.1	2	0.1	2	0.1	2	0.1
1hne_e	15	0	0.0	2	0.1	0	0.0	3	0.2
1hrt_j	270	16	0.1	42	0.2	19	0.1	35	0.1
1hum_a	122	8	0.1	20	0.2	11	0.1	27	0.2
1isc_A	73	3	0.0	4	0.1	14	0.2	16	0.2
1isu_A	21	0	0.0	1	0.0	0	0.0	2	0.1
1kan_A	36	6	0.2	10	0.3	8	0.2	2	0.1
1len_A	518	28	0.1	70	0.1	37	0.1	93	0.2
1lga_A	30	0	0.0	1	0.0	0	0.0	4	0.1
1lld_A	106	8	0.1	16	0.2	16	0.2	20	0.2
1lob_D	448	31	0.1	70	0.2	43	0.1	94	0.2
1lts_A	144	8	0.1	17	0.1	14	0.1	18	0.1
1lts_C	116	8	0.1	24	0.2	8	0.1	15	0.1
1lts_D	269	16	0.1	29	0.1	34	0.1	53	0.2
1min_A	596	20	0.0	53	0.1	40	0.1	83	0.1
1min_B	367	20	0.1	45	0.1	41	0.1	47	0.1
1myp_C	49	6	0.1	12	0.2	4	0.1	11	0.2
1nba_A	549	39	0.1	80	0.1	59	0.1	82	0.1
1nip_A	159	9	0.1	16	0.1	7	0.0	29	0.2
1olb_A	51	1	0.0	3	0.1	3	0.1	4	0.1
1pdg_A	116	2	0.0	7	0.1	4	0.0	18	0.2
1pox_A	129	14	0.1	16	0.1	8	0.1	22	0.2
1prc_C	443	36	0.1	57	0.1	40	0.1	68	0.2
1prc_L	434	26	0.1	57	0.1	48	0.1	75	0.2
1psp_A	48	2	0.0	3	0.1	3	0.1	11	0.2
1pxt_B	215	17	0.1	44	0.2	22	0.1	42	0.2
1pya_B	167	11	0.1	21	0.1	7	0.0	17	0.1
1rib_A	266	21	0.1	55	0.2	26	0.1	58	0.2
1sac_A	147	13	0.1	19	0.1	14	0.1	16	0.1
1scm_A	304	19	0.1	41	0.1	30	0.1	62	0.2
1slt_A	53	5	0.1	11	0.2	8	0.2	15	0.3
1srg_A	183	13	0.1	58	0.3	17	0.1	38	0.2
1sry_A	250	19	0.1	27	0.1	20	0.1	60	0.2
1tab_I	91	9	0.1	11	0.1	4	0.0	6	0.1
1tgs_I	103	8	0.1	14	0.1	8	0.1	19	0.2
1tnr_A	61	5	0.1	11	0.2	7	0.1	13	0.2
1tpk_A	33	4	0.1	9	0.3	5	0.2	6	0.2
1tpl_A	191	11	0.1	20	0.1	10	0.1	25	0.1
1trk_A	470	28	0.1	68	0.1	34	0.1	87	0.2
1trz_B	143	6	0.0	17	0.1	19	0.1	28	0.2
1tss_A	125	13	0.1	18	0.1	12	0.1	28	0.2
1vaa_A	197	8	0.0	22	0.1	16	0.1	28	0.1
1whit_A	447	22	0.0	69	0.2	40	0.1	68	0.2
1wsy_A	122	10	0.1	16	0.1	10	0.1	16	0.1
1ypt_B	52	4	0.1	17	0.3	9	0.2	4	0.1
2atc_B	61	2	0.0	3	0.0	4	0.1	10	0.2
2bbk_H	315	22	0.1	43	0.1	21	0.1	46	0.1
2bbv_C	402	25	0.1	60	0.1	55	0.1	88	0.2
2cht_C	227	16	0.1	27	0.1	22	0.1	37	0.2
2drp_A	5	1	0.2	2	0.4	0	0.0	0	0.0
2er7_E	56	1	0.0	5	0.1	1	0.0	8	0.1
2gst_A	107	6	0.1	7	0.1	13	0.1	27	0.3

2hbm_A	165	14	0.1	26	0.2	9	0.1	38	0.2
2hip_A	16	1	0.1	3	0.2	1	0.1	2	0.1
2hpd_A	36	1	0.0	2	0.1	1	0.0	8	0.2
1hpe_A	169	17	0.1	19	0.1	32	0.2	46	0.3
2lgs_A	17	0	0.0	0	0.0	1	0.1	3	0.2
2mad_L	61	7	0.1	13	0.2	2	0.0	12	0.2
2mta_C	42	5	0.1	6	0.1	4	0.1	6	0.1
2ohx_A	172	11	0.1	28	0.2	13	0.1	35	0.2
2pcd_A	392	25	0.1	46	0.1	25	0.1	58	0.1
2pfk_D	90	2	0.0	5	0.1	11	0.1	25	0.3
2pmg_A	59	4	0.1	6	0.1	3	0.1	11	0.2
2rsp_B	149	15	0.1	17	0.1	27	0.2	24	0.2
2scp_A	107	9	0.1	10	0.1	13	0.1	23	0.2
2tbv_A	152	12	0.1	9	0.1	8	0.1	23	0.2
2tpr_A	314	17	0.1	29	0.1	35	0.1	50	0.2
2zta_A	63	9	0.1	13	0.2	12	0.2	25	0.4
3aah_A	308	21	0.1	31	0.1	15	0.0	46	0.1
3gap_B	149	11	0.1	20	0.1	22	0.1	25	0.2
3hhr_C	50	3	0.1	10	0.2	7	0.1	9	0.2
3mdd_A	147	5	0.0	17	0.1	10	0.1	15	0.1
3mon_A	124	10	0.1	23	0.2	9	0.1	23	0.2
3sic_I	103	7	0.1	11	0.1	11	0.1	9	0.1
4cpa_I	66	2	0.0	9	0.1	2	0.0	8	0.1
4sgb_I	92	7	0.1	16	0.2	7	0.1	14	0.2
4xia_A	165	9	0.1	26	0.2	11	0.1	33	0.2
5rub_A	337	15	0.0	41	0.1	36	0.1	96	0.3
8atc_A	76	0	0.0	4	0.1	3	0.0	11	0.1
8cat_A	464	44	0.1	57	0.1	56	0.1	72	0.2
8tln_E	28	3	0.1	8	0.3	2	0.1	4	0.1
9wga_A	32	1	0.0	5	0.2	0	0.0	7	0.2
Число обнаруженных противоречий		122		121		122		121	

ВЫВОДЫ. Проведенное исследование показывает, что структура белок-белковых контактов практически всех 122-х обследованных белок-белковых комплексов, описанных в Брукхейвенском банке данных, не соответствует ни одной из рассмотренных четырех гипотез о комплементарных парах аминокислот в межбелковых контактах. Таким образом, все четыре гипотезы можно считать несостоятельными.

Работа выполнена при частичной финансовой поддержке РФФИ (грант № 96-04-495545)

ЛИТЕРАТУРА.

1. Tropsha A., Kizer J.S., Chaiken I.M. (1992). J. of Molecular Recognition. 5. 43-54.
2. Меклер Л.Б. (1970), Биофизика, 14, 581-584
3. Blalock J.E., Bost K.L. (1986). Biochem.J. 234. 679-683.
4. Bost K.L., Smith E.M., Blalock J.E. (1985). Proc.Natl.Acad.Sci. USA. 82. 1372-1375.
5. Shai Y., Flashner M., Chaiken I.M. (1987). Biochemistry. 26, 669-675.
6. Knutson V.P. (1988). J. Biol. Chem. 263. 14145-14151.
7. Campbell-Thompson M. (1993). Amer. J. Med. Sci. 305. 365-373.
8. Fassina G., Zamai G., Brigham-Burke M., Chaiken I.M. (1989). Biochemistry. 28. 8811-8818.
9. Blalock J.E. (1995) Nat Med Sep 1:9 876-87
10. Волькенштейн М.В. (1975). Молекулярная биофизика. М.: Наука, 591
11. Линде Д.М., Амелин К.А., Поройков В.В., Арчаков А.И. (1997), Мол. биология, 31, 430-434.
12. Hobohm U., Sander C. (1994). Protein Science. 3, 522.
13. Bernstein F.C. (1977). J.Mol.Biol. 112, 535-542.
14. Stites W.E. (1997), Chem. Rev., 97, 1233-1250

PROTEIN-PROTEIN RECOGNITION: CONTROL OF SOME HYPOTHESES OF AMINOACIDE'S COMPLEMENTARITY EXISTENCE USING THE BROOKHAVEN DATA BANK DATA

DROZDOV-TIKHOMIROV L.N.*, LINDE D.M.* *, POROIKOV V.V.* *, MOKULSKI M.A*.

*Institute of Molecular Genetics RAS, Kurchatova sq., Moscow, 123128.

**Institute of Biomedical Chemistry RAMS, Pogodinskaya 10, Moscow, 119832.

Four amino acid's complementarity hypotheses have been checked using the data on the structure of 122 protein complexes taken from protein Brookhaven Data Bank. No one of hypotheses was conformed at the analysis of the contact region structures of the protein complexes examined.

Key Words: protein-protein complex, protein-protein recognition, contacting amino acids, complementary amino acids. Brookhaven Data Bank.