

СТРУКТУРНО-ФУНКЦИОНАЛЬНЫЙ АНАЛИЗ
БИОПОЛИМЕРОВ И ИХ КОМПЛЕКСОВ

УДК 577.150.2

L-МОДУЛИ В SH3-ПОДОБНЫХ ФОЛДАХ: СТРУКТУРА
И ОСОБЕННОСТИ ПОСЛЕДОВАТЕЛЬНОСТЕЙ© 2018 г. А. М. Каргатов^а, Е. В. Бражников^а, А. В. Ефимов^{а, *}^аИнститут белка Российской академии наук, Пушкино, Московская обл., 142290 Россия

*e-mail: efimov@protres.ru

Поступила в редакцию 17.04.2018 г.

Принята к печати 29.05.2018 г.

В результате стереохимического анализа негомологичных SH3-подобных фолдов выявлен новый повторяющийся L-образный модуль, структура которого может быть представлена как β -тяж–петля– β -тяж. β -Тяжи L-модуля расположены под углом $\sim 90^\circ$ в разных β -слоях, упакованных ортогонально, и вместе с петлей образуют в пространстве полвитка правой суперспирали. Чтобы изучить структурное сходство и различия L-модуля, создана база данных, включающая в себя 60 негомологичных SH3-подобных доменов. Определены частоты встречаемости L-модулей в зависимости от длины их петель. Показано, что L-модули с $\beta_m\alpha\alpha\beta_n$ - и $\beta_m\alpha\alpha\beta\alpha\beta_n$ -конформациями, где m и n – число остатков в первом и втором β -тяжах, встречаются наиболее часто (57 и 8% соответственно). С помощью компьютерных программ показано, что пространственные структуры L-модулей каждого типа хорошо совмещаются при наложении друг на друга. Проведено структурное выравнивание аминокислотных последовательностей L-модулей, что позволило определить ключевые позиции, в которых должны находиться гидрофобные, гидрофильные и пролиновые остатки.

Ключевые слова: α -спираль, β -тяж, структурный мотив, структурное сходство

DOI: 10.1134/S0026898418060095

ВВЕДЕНИЕ

SH3-домены и SH3-подобные домены широко распространены как в гомологичных, так и в негомологичных белках. В Банке белковых данных (Protein Data Bank, PDB) нами найдено свыше 800 PDB-файлов SH3-подобных доменов, из которых отобрано 60 негомологичных доменов согласно критериям программы для попарного выравнивания аминокислотных последовательностей [1]. В общем виде структура SH3-подобных доменов состоит из пяти–шести β -тяжей, образующих замкнутый β -цилиндр или β -бочонок, к которому могут быть пристроены несколько дополнительных α -спиралей и/или β -тяжей. Отличительная особенность β -бочонков SH3-подобных доменов (будем называть их в дальнейшем SH3-фолдами) состоит в том, что в них одновременно сосуществуют несколько структурных мотивов, которые широко распространены и в других белках. Так, первые три β -тяжа SH3-фолда образуют 3β -уголок [2]. Как правило, центральный β -тяж 3β -уголка сильно скручен и изогнут вплоть до 90° , и одна его половина расположена в одном, а другая – в другом β -слое β -бочонка, поэтому формально эти половины часто рассматривают как два β -тяжа (отсюда пять или шесть β -тяжей в SH3-фолдах). С другой стороны, все SH3-фолды содержат S-об-

разные β -листы, которые вместе с примыкающими к ним β -тяжами образуют правые суперспирали типа β -тяж–S-лист– β -тяж [3, 4]. Во многих SH3-фолдах первые два β -тяжа образуют сильно скрученные и изогнутые β -шпильки, которые имеют много общего со скрученными β -шпильками в других белках [5].

В настоящей работе обнаружен и охарактеризован еще один повторяющийся во всех SH3-фолдах структурный элемент, названный L-модулем, поскольку укладка его цепи напоминает латинскую букву L. Его структура может быть представлена как β -тяж–петля– β -тяж. β -Тяжи L-модуля расположены под углом $\sim 90^\circ$ в разных β -слоях SH3-фолда и вместе с петлей образуют полвитка правой суперспирали. В работе показано, что структуры негомологичных L-модулей одного типа хорошо совмещаются при наложении, а их аминокислотные последовательности имеют одинаковое расположение в цепи ключевых гидрофобных, гидрофильных и пролиновых остатков.

ЭКСПЕРИМЕНТАЛЬНАЯ ЧАСТЬ

Белки для исследования, содержащие SH3-фолды, отобраны из Банка белковых данных (PDB) с помощью разработанной в нашей группе

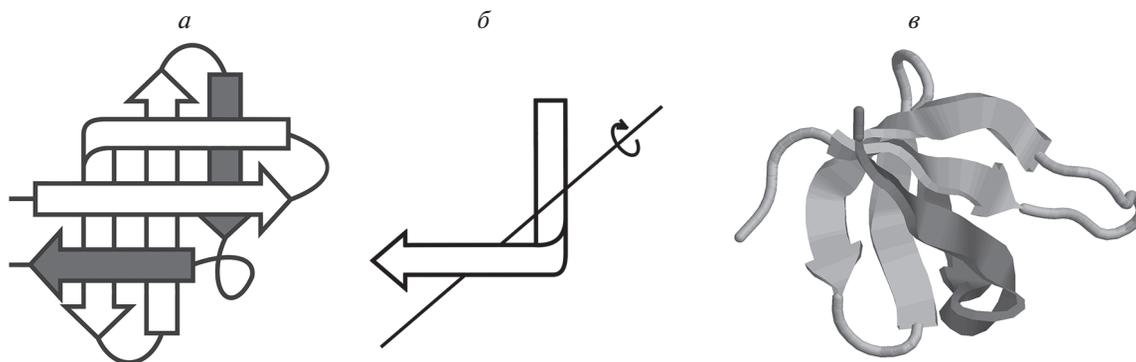


Рис. 1. *а* – Схематическое изображение общей укладки цепи в SH3-фолдах (L-модуль выделен серым цветом). *б* – Участок цепи L-модуля при переходе из одного β -слоя в другой образует полвитка правой суперспирали вокруг воображаемой оси. *в* – Изображение структуры SH3-фолда белка IQMC, полученное с помощью программы RasMol [7] (L-модуль выделен более темным цветом). Во всех случаях β -тяги показаны в виде стрелок, направленных от N- к C-концам.

структурной классификации PSBOST, которая доступна по адресу <http://strees.protres.ru/> [6], а также непосредственно с сервера PDB (<http://www.rcsb.org/pdb/>) с помощью ключевых слов. Отбирали для анализа негомологичные белки или домены. Проверку на гомологию проводили с помощью программы Blast 2 Sequences для попарного выравнивания аминокислотных последовательностей (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>) [1]. Негомологичными считали белки, для которых показатель E-value, учитывающий как полную идентичность участков последовательностей, так и замены на сходные аминокислоты при выравнивании, был больше $1e-04$.

В Банке белковых данных мы нашли свыше 800 PDB-файлов SH3-доменов и SH3-подобных доменов, из которых отобрали 60 негомологичных SH3-фолдов (табл. 1) в соответствии с критериями программы Blast 2 Sequences [1].

Основной объект исследования – L-модуль – определяли визуально с помощью программы молекулярной графики RasMol [7]. При вычислении углов ϕ и ψ использовали программу MolMol [8]. Конформации аминокислотных остатков обозначали набором символов в соответствии с номенклатурой, предложенной ранее [9]. Чтобы наложить пространственные структуры друг на друга, использовали программу DeepView/Swiss-PDBViewer [10]. Выравнивание аминокислотных последовательностей L-модулей, основанное на их структурном сходстве, проводили вручную.

РЕЗУЛЬТАТЫ ИССЛЕДОВАНИЯ

Основной объект исследования в настоящей работе – негомологичные SH3-подобные домены, а точнее их β -бочонки, которые названы здесь SH3-фолдами. Общая укладка полипептидной цепи в пространственной структуре SH3-фолдов представлена на рис. 1. Как отмечено выше, пер-

вые три (или четыре) β -тяги образуют 3β -уголок [2], в центральной части находится S-образный β -лист, который вместе с прилегающими по флангам β -тяжами образует виток правой суперспирали типа β -тяж–S-лист– β -тяж [3]. Как показал анализ, в C-концевой части всех SH3-фолдов найден структурный модуль типа β -тяж–петля– β -тяж, который назван здесь L-модулем (рис. 1*а, б*). β -Тяжи L-модуля расположены в разных β -слоях SH3-фолда под углом $\sim 90^\circ$ и вместе с петлей образуют в пространстве полвитка правой суперспирали (рис. 1*б*).

На рис. 2 представлена зависимость частоты встречаемости L-модулей от длины их петель. Как видно, значительное большинство L-модулей имеют петли из трех, пяти и шести остатков. Более детальный анализ показывает, что наиболее часто встречаются L-модули с $\beta_m\alpha\alpha\beta_n$ - и $\beta_m\alpha\alpha\beta\alpha\beta_n$ -конформациями, где m и n – число остатков в первом и втором β -тяжах соответственно. Принимая во внимание, что первый остаток на N-конце α -спирали, который образует первую водородную связь, имеет на самом деле β -конформацию [9] (аналогичная ситуация и в спиралях 3_{10}), можно сделать вывод, что значительное большинство L-модулей содержит виток α -спирали или спирали 3_{10} в петлях (см. рис. 3 и 4). Отметим, что этот первый остаток α -спирального витка – одновременно и последний остаток первого β -тяжа L-модуля. В петлях L-модулей с $\beta_m\alpha\alpha\beta\alpha\beta_n$ -конформациями дополнительно к виткам α -спирали есть стандартные β -полуобороты с $\beta\alpha\beta$ -конформациями (для их описания см. работу [9], а также рис. 4).

L-модули из различных негомологичных SH3-фолдов обладают очень схожими пространственными структурами и хорошо совмещаются при наложении. На рис. 3 представлено наложение друг на друга 13 немного укороченных с концов

Таблица 1. Список негомологичных SH3-фолдов, разрешение их структур и длины петель в L-модулях

№	PDB-код, цепь	Участок цепи	Разрешение, Å	Число остатков в петле
01	1f39 (A)	178–226	1.9	2
02	1gec (E)	129–215	2.1	16
03	1h5w (A)	61–123	2.1	11
04	1hyo (A)	18–117	1.3	3
05	1i1j (A)	21–90	1.39	3
06	1igq (A)	307–357	1.7	3
07	1jb0 (E)	5–68	2.5	3
08	1loj (A)	24–82	1.9	5
09	1m1g (A)	199–248	2.0	3
10	1m5q (A)	22–76	2.0	5
11	1ne8 (A)	6–95	2.1	3
12	1ou8 (A)	26–95	1.6	5
13	1oz2 (A)	242–292	1.55	5
14	1oz2 (A)	349–399	1.55	5
15	1oz2 (A)	453–503	1.55	5
16	1r77 (A)	284–362	1.75	11
17	1sf9 (A)	84–127	1.71	3
18	1tov (A)	150–219	1.77	3
19	1u1s (A)	21–65	1.6	5
20	1ugp (B)	143–226	1.63	3
21	1v76 (A)	51–99	2.0	3
22	1vq8 (Q)	37–94	2.2	3
23	1vqo (A)	103–157	2.2	2
24	1y71 (A)	9–81	1.95	3
25	1zuk (A)	7–67	1.9	3
26	1zuu (A)	3–57	0.97	3
27	2bjq (A)	32–81	1.75	8
28	2bjq (A)	118–166	1.75	6
29	2bjq (A)	216–263	1.75	6
30	2bjq (A)	295–343	1.75	6
31	2cb5 (A)	321–425	1.85	9
32	2evr (A)	14–79	1.6	6
33	2f5k (A)	15–65	2.2	3
34	2fpe (A)	3–58	1.75	3
35	2ox7 (A)	6–70	1.78	8
36	2pu9 (B)	5–73	1.65	3
37	2ra2 (A)	5–54	1.9	5
38	2rh2 (A)	28–77	0.96	3
39	2v1r (A)	14–76	2.1	3
40	2x4x (A)	1087–1165	1.85	3
41	2xc1 (A)	47–83	1.65	6
42	2zgw (A)	190–235	1.5	3
43	2zpl (A)	129–218	1.7	18
44	3bz7 (A)	33–80	2.0	3

Таблица 1. Окончание

№	PDB-код, цепь	Участок цепи	Разрешение, Å	Число остатков в петле
45	3fdr (A)	32–79	1.75	3
46	3g1j (A)	18–89	1.7	16
47	3g7z (A)	3–83	2.35	3
48	3h8z (A)	15–60	1.92	3
49	3h8z (A)	74–126	1.92	3
50	3pmt (A)	560–608	1.8	3
51	3qir (A)	31–115	2.45	3
52	4edj (A)	14–105	1.9	10
53	4edj (A)	112–196	1.9	10
54	4fw1 (A)	222–268	1.7	3
55	4kxt (A)	263–345	2.29	3
56	4rxj (A)	962–1019	2.1	3
57	4ytl (A)	538–583	1.6	3
58	4ytl (A)	588–632	1.6	3
59	5jqc (A)	256–329	2.15	3
60	5jqc (A)	335–411	2.15	3

L-модулей с $\beta_m\alpha\alpha\beta_n$ -конформацией. Репрезентативная выборка включала в себя следующие белки: 1M1G, 1TOV, 1Y71, 2PU9, 2RH2, 2X4X, 3H8Z (2 модуля), 3PMT, 4FW1, 4RXJ, 4YTL (2 модуля). На рис. 4 показано наложение немного укороченных с концов пяти L-модулей с $\beta_m\alpha\alpha\beta_n$ -конформацией, взятых из SH3-фолдов следующих белков: 1LOJ, 1M5Q, 1OU8, 1U1S, 2RA2. Отметим, что структуры L-модулей хорошо совмещаются друг с другом, несмотря на то, что они взяты из неродственных, негомологичных белков.

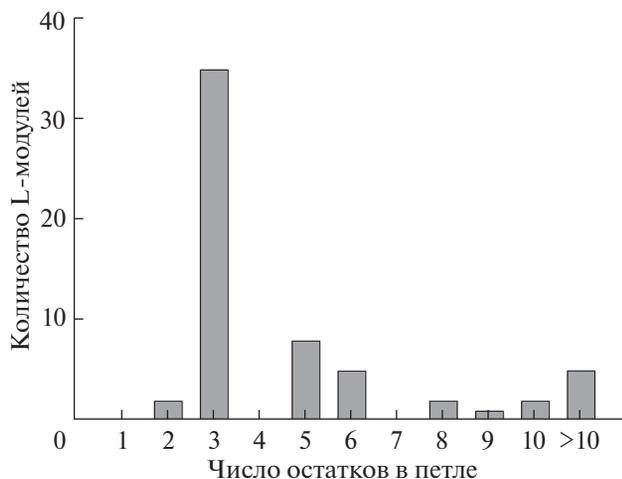


Рис. 2. Частоты встречаемости L-модулей в зависимости от длины петель.

Структурное сходство негомологичных L-модулей позволяет провести структурное выравнивание их аминокислотных последовательностей, сравнивая аминокислотные остатки, находящиеся в структурно-эквивалентных позициях. На рис. 5 представлено структурное выравнивание аминокислотных последовательностей 34 L-модулей с $\beta_m\alpha\alpha\beta_n$ -конформацией. Выравнивание проводили вручную. Участки последовательностей первых β -тяжей выравнивали по последнему остатку каждого β -тяжа так, чтобы внутренние (направленные в гидрофобное ядро) и внешние (направленные наружу) позиции попадали в одни и те же столбцы соответственно. Участки α -спиральных последовательностей выравнивали по первым остаткам α -спиралей, которые одновременно являются и последними остатками первых β -тяжей. Таким образом, в каждом столбце множественного выравнивания расположены структурно-эквивалентные остатки, конформация которых показана буквами α и β в верхней строке. Аналогичное выравнивание аминокислотных последовательностей L-модулей, имеющих $\beta_m\alpha\alpha\beta_n$ -конформации, представлено на рис. 6.

В соответствии с принципами структурной организации белков боковые цепи аминокислотных остатков, погруженные в гидрофобное ядро или локальное гидрофобное окружение молекулы, должны быть гидрофобными, а расположенные на поверхности — преимущественно гидрофильными. В идеальных случаях в двухслойных структурах β -тяжи должны иметь строгое чередование гидрофобных и гидрофильных остатков в цепи [11]. Однако это правило не всегда выполнено в

β -тяжах, расположенных на краях β -листов, поскольку их боковые цепи, направленные в сторону гидрофобного ядра, могут быть доступны молекулам воды и, следовательно, могут быть гидрофильными. Как видно на рис. 1а, β -тяжи, образующие L-модуль, расположены на краях β -слоев, поэтому многие боковые цепи этих β -тяжей, направленные в сторону гидрофобного ядра (их буквенные обозначения подчеркнуты на рис. 5 и 6), являются гидрофильными. Тем не менее, многие из них имеют длинные гидрофобные ножки (Lys, Arg, Glu, Gln). В L-модулях с $\beta_m\alpha\alpha\beta_n$ -конформациями наиболее строго правило выполнено в двух позициях – это предпоследняя позиция первого β -тяжа и первая позиция второго β -тяжа, которые в подавляющем большинстве случаев заняты гидрофобными остатками (они выделены серой заливкой на рис. 5). В L-модулях с $\beta_m\alpha\alpha\beta\alpha\beta_n$ -конформациями к таким относят три позиции – это предпоследняя позиция первого β -тяжа, β -позиция в петле и первая внутренняя позиция второго β -тяжа, которые заняты только гидрофобными остатками (выделены серой заливкой на рис. 6). Особый интерес представляют первые позиции α -спиралей, которые одновременно являются последними позициями первых β -тяжей в обоих типах L-модулей (они так же выделены серой заливкой на рис. 5 и 6). В этих случаях обнаружен так называемый “перпендикулярный вход” в α -спираль, который очень часто бывает на N-концах α -спиралей в белках и имеет $\beta\beta\alpha_m$ -конформацию, где m – число остатков в α -спирали [9, 12]. Статистический анализ показал, что в белках первые (β) позиции α -спиралей в большинстве случаев заняты гидрофильными (~80%, большинство из которых – Ser, Thr, Asp и Asn) и небольшими (~17%, это Gly, Ala, Pro) остатками. В целом, в первых (β) позициях α -спиралей L-модулей обнаружено аналогичное распределение остатков, однако встречается аномально большое количество пролинов. В L-модулях с $\beta_m\alpha\alpha\beta_n$ -конформацией 50% этих позиций занято пролинами. Кроме этого, пролины обнаружены во вторых и третьих позициях α -спиралей (рис. 5).

ОБСУЖДЕНИЕ РЕЗУЛЬТАТОВ

В глобулярных белках полипептидная цепь многократно сложена сама на себя и образует целый ряд часто повторяющихся блоков из двух и более соседних по цепи α -спиралей и/или β -тяжей, которые называются супервторичными структурами или структурными мотивами [13–16]. Особый интерес представляют структурные мотивы, имеющие компактные уникальные укладки цепей и определенную псевдохиральность [15, 16]. Каждый тип таких структурных мотивов можно охарактеризовать одинаковым количеством α -спиралей и/или β -тяжей и их одинаковым расположением

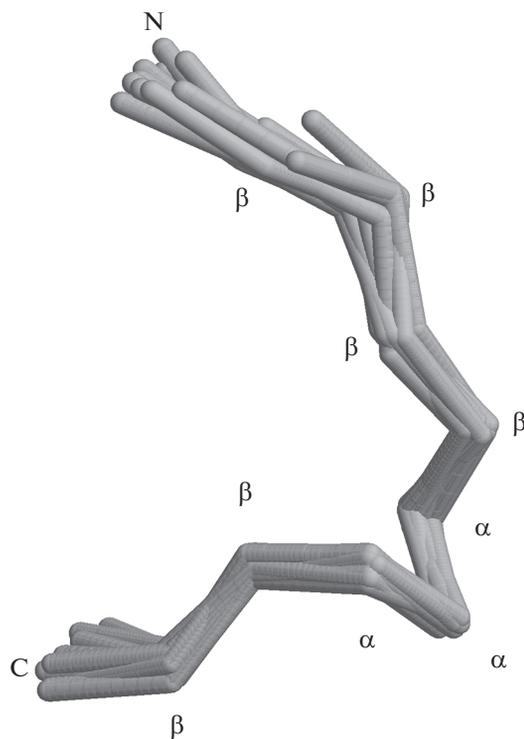


Рис. 3. Наложение друг на друга структур 13 L-модулей с $\beta_m\alpha\alpha\beta_n$ -конформацией с помощью программы DeepView/Swiss-PDBViewer [10]. Изображение получено с помощью программы RasMol [7]. Буквы α и β обозначают конформации соответствующих остатков, N и C – концы участков цепи.

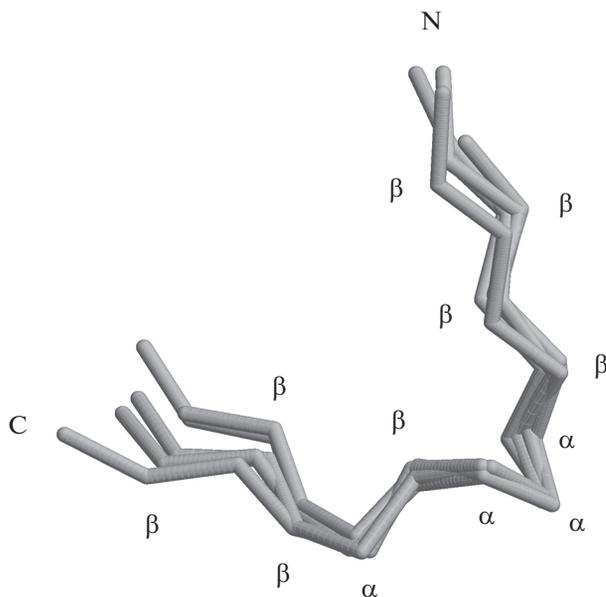


Рис. 4. Наложение друг на друга структур пяти L-модулей с $\beta_m\alpha\alpha\beta\alpha\beta_n$ -конформацией, полученное так же, как на рис. 3.

вдоль цепи и в пространстве. Считают, что такие структурные мотивы достаточно стабильны, способны свернуться в свои уникальные структуры

№	PDB-код	$\beta\beta\beta\beta\beta\beta$	$\alpha\alpha$	$\beta\beta\beta\beta$	Участок цепи
01	1HYO : A	<u>fts</u>	QAS	<u>atmh</u>	107–116
02	1I1J : A	<u>gyfp</u>	SSI	<u>vre</u>	77–86
03	1IGQ : A	<u>efean</u>	LAD	<u>vk\bar{l}</u>	342–352
04	1JB0 : E	<u>tnnfa</u>	LHE	<u>vqe</u>	57–67
05	1M1G : A	<u>rmtpveld</u>	FDQ	<u>vek</u>	234–247
06	1NE8 : A	<u>irtld</u>	KQR	<u>lt</u>	80–89
07	1SF9 : A	<u>ealp</u>	ISV	<u>lqe</u>	117–126
08	1TOV : A	<u>ggfvr</u>	PVD	<u>vk\bar{v}</u>	208–218
09	1UGP : B	<u>svyydcw</u>	EPY	<u>iel</u>	213–225
10	1V76 : A	<u>iwkvp</u>	KDV	<u>sifefe</u>	85–98
11	1VQ8 : Q	<u>kektiivt</u>	AAH	<u>lrr</u>	80–93
12	1Y71 : A	<u>qtnip</u>	EQM	<u>vk\bar{k}</u>	70–80
13	1ZUK : A	<u>viegifp</u>	KSF	<u>vav</u>	54–66
14	1ZUU : A	<u>etglvp</u>	TTY	<u>ir</u>	46–56
15	2F5K : A	<u>ewvp</u>	ESR	<u>vl\bar{k}</u>	55–64
16	2FPE : A	<u>rgvfp</u>	AYY	<u>aie</u>	47–57
17	2PU9 : B	<u>kahfr</u>	PDE	<u>v\bar{t}l</u>	62–72
18	2RH2 : A	<u>vqiyp</u>	VAA	<u>ler</u>	66–76
19	2V1R : A	<u>igyip</u>	YNY	<u>iei</u>	65–75
20	2X4X : A	<u>wgwl\bar{p}</u>	RTK	<u>lv\bar{p}</u>	1154–1164
21	3BZ7 : A	<u>drqv\bar{k}</u>	KDD	<u>an\bar{q}</u>	69–79
22	3FDR : A	<u>ngdcp</u>	LKD	<u>lra</u>	68–78
23	3G7Z : A	<u>asvp</u>	VSV	<u>ig</u>	69–77
24	3H8Z : A	<u>rqip</u>	FGD	<u>vr</u>	51–59
25	3H8Z : A	<u>eiv\bar{t}</u>	LER	<u>lr\bar{p}</u>	116–125
26	3PMT : A	<u>yeev\bar{l}</u>	LSN	<u>ik\bar{p}</u>	597–607
27	3QIR : A	<u>ill\bar{l}</u>	PEL	<u>sf</u>	106–114
28	4FW1 : A	<u>viwv\bar{p}</u>	SRK	<u>vk\bar{p}</u>	257–267
29	4KXT : A	<u>tyl\bar{p}</u>	LEV	<u>cn\bar{i}</u>	335–344
30	4RXJ : A	<u>dyywv\bar{h}</u>	QGR	<u>vfp</u>	1007–1018
31	4YTL : A	<u>lef\bar{p}</u>	IST	<u>lr\bar{k}</u>	573–582
32	4YTL : A	<u>evt\bar{i}</u>	ANN	<u>ls</u>	622–631
33	5JQC : A	<u>wm\bar{d}</u>	KRA	<u>fv\bar{y}</u>	320–328
34	5JQC : A	<u>wl\bar{d}</u>	ARA	<u>fh\bar{v}</u>	402–410

Рис. 5. Структурное выравнивание аминокислотных последовательностей L-модулей, имеющих $\beta_n\alpha\alpha\beta_n$ -конформации. Конформации остатков в столбцах показаны буквами α и β в верхней строке. α -Спиральные остатки показаны заглавными, β -структурные остатки – строчными буквами. Подчеркнуты β -остатки, направленные в сторону гидрофобного ядра. В каждой строке слева показан PDB-код белка, а справа – участок цепи, образующий L-модуль. См. также текст.

сами по себе и могут быть зародышами и/или готовыми структурными блоками при сворачивании белков [15, 17].

В предыдущих работах [18, 19] мы подробно рассмотрели различные варианты П-модулей, структура которых напоминает греческую (или русскую) букву П. По-видимому, сами по себе П-модули нестабильны, однако в комбинациях с

β -шпильками, S- и Z-образными β -листами и $\beta\alpha\beta$ -единицами они образуют компактные, замкнутые и стабильные структурные мотивы, которые широко распространены в белках. Как видно на рис. 1, L-модули – некомпактные структуры, их β -тяжи практически не взаимодействуют друг с другом, однако в пространстве они всегда расположены одинаково и вместе с петлей обра-

№	PDB-код	$\beta\beta\beta\beta\beta\beta\beta\beta$	$\alpha\alpha\beta\alpha$	$\beta\beta\beta\beta$	Участок цепи
01	1LOJ : A	<u>v</u> l <u>i</u> r	GDN <u>i</u> V	<u>y</u> i <u>s</u> r	69–81
02	1M5Q : A	r <u>v</u> f <u>i</u> m	YRY <u>i</u> V	<u>h</u> i <u>d</u>	62–74
03	1OU8 : A	<u>v</u> s <u>r</u> e <u>l</u> y <u>i</u> p	MGA <u>A</u> L	<u>a</u> i <u>y</u> a	78–94
04	1U1S : A	<u>q</u> m <u>v</u> y	KHA <u>i</u> S	<u>t</u> v <u>v</u> p	52–64
05	2RA2 : A	<u>k</u> q <u>q</u> i <u>n</u>	RTD <u>v</u> K	<u>e</u> m <u>v</u> a	40–53

Рис. 6. Структурное выравнивание аминокислотных последовательностей, кодирующих L-модули с $\beta_m\alpha\alpha\beta_n$ -конформациями. Обозначения такие же, как на рис. 5.

зуют полвитка правой суперспирали. Взаимодействие L-модуля с N-концевой частью SH3-фолда приводит к образованию замкнутой структуры типа β -бочонка, которая обладает повышенной кооперативностью и большей стабильностью (см., например, работу [20]). По-видимому, в этом и состоит основная структурная роль L-модулей.

Необходимо отметить, что впервые структуры, аналогичные L-модулям, обнаружены в пяти белках с ортогональной упаковкой β -слоев и описаны в работе [21]. Показано также, что β -тяги, которые переходят из одного β -слоя в другой, сильно скручены и изогнуты на $\sim 90^\circ$, образуя полвитка правой суперспирали, так же как в L-модуле. В $\beta\beta$ -уголках сразу два соседних β -тяжа образуют две правых суперспирали (т.е. два L-модуля) при переходе из одного β -слоя в другой [22]. Однако наиболее часто L-модули обнаружены в 3β -уголках и, следовательно, в содержащих их белках. Структура 3β -уголка представляет собой Z-образный β -лист, сложенный сам на себя таким образом, что центральный β -тяж при переходе из одного слоя в другой образует полвитка правой суперспирали, т.е. L-модуль [23]. В местах перехода центральные β -тяги могут приобретать конформацию β -сгиба (β -bend) [20] или образовывать небольшие стандартные структуры с $\beta\gamma\beta\beta$ -, $\beta\beta\alpha_L\beta$ -, $\beta\alpha\gamma\beta$ -, $\beta\gamma\epsilon\beta$ - или $\beta\epsilon\beta$ -конформациями [22–24]. При построении обновленного структурного древа мы нашли в Банке белковых данных (PDB) 720 белков, содержащих 3β -уголки, из них 224 – негомологичные белки (дерево доступно по адресу: http://strees.protres.ru/2_1.xml, см. также [6, 24]). В эту базу данных вошло большинство из 60 SH3-фолдов (не все, поскольку в 2014 г. список этих доменов был меньше [24]), потому что в каждом из этих фолдов первые с N-конца три (или четыре) β -тяжа образуют 3β -уголок и соответственно второй или второй и третий β -тяжи (в зависимости от конформации и количества остатков в переходе) образуют правый L-модуль. Таким образом, каждый SH3-фолд содержит два правых L-модуля, один из которых находится в N-концевой части и входит в состав 3β -уголка, а другой – обособленный – на C-конце. Как следует

из изложенного выше, настоящая работа посвящена, главным образом, этим обособленным L-модулям.

Сравнительный анализ показывает, что большинство L-модулей, входящих в состав 3β -углов SH3-фолдов, имеет более короткие переходы из одного слоя в другой, чем обособленные L-модули. Как правило, в состав первых входит от нуля до двух остатков в α -, α_L -, γ - или ϵ -конформациях (см. также работу [24]), а большинство вторых, как это видно на рис. 2, состоят из трех, пяти и более остатков. В образовании многих переходов L-модулей, входящих в 3β -уголки, ключевую роль играют остатки с α_L - или ϵ -конформацией, которые должны быть глицинами или остатками с гибкими боковыми цепями. В переходах (или петлях) большинства обособленных L-модулей остатков с α_L - и ϵ -конформациями нет, да и глицины встречаются относительно редко (см. рис. 5 и 6). Пролины, наоборот, обнаруживаются достаточно часто и, по-видимому, играют важную роль в образовании переходов в обособленных L-модулях.

Отметим в заключение, что обособленные L-модули есть и в белках других структурных классов и суперсемейств, отличных от SH3-фолдов. Например, L-модули с $\beta_m\alpha\alpha\beta_n$ -конформацией найдены в следующих белках из PDB: 1B9M, 1FR3, 1PL6, 1Q12, 2J7N, 2ZB3, 2ZB3, 5A4G и др. Несмотря на то, что эти L-модули найдены в различных структурных классах белков, они так же образуют правые суперспирали, хорошо совмещаются при наложении друг на друга и с L-модулями из SH3-фолдов и имеют схожее расположение в цепи и в структуре ключевых гидрофобных, гидрофильных и пролиновых остатков. Все это свидетельствует в пользу того, что в основе такого структурного сходства лежат общие физико-химические принципы, а не гомология.

Работа выполнена в рамках проекта гос. задания № 0115-2018-0008, а также при финансовой поддержке Российского фонда фундаментальных исследований (17-04-00242).

СПИСОК ЛИТЕРАТУРЫ

1. Tatusova T.A., Madden T.L. (1999) BLAST 2 Sequences, a new tool for comparing protein and nucleotide sequences. *FEBS Microbiol. Lett.* **177**, 187–188.
2. Efimov A.V. (1997) A structural tree for proteins containing 3 β -corners. *FEBS Lett.* **407**, 37–41.
3. Efimov A.V. (1998) A structural tree for proteins containing S-like β -sheets. *FEBS Lett.* **437**, 246–250.
4. Efimov A.V. (2010) Structural motifs are closed into cycles in proteins. *Biochem. Biophys. Res. Commun.* **399**, 412–415.
5. Бошкова Е.А., Бражников Е.В., Ефимов А.В. (2016) Взаимосвязь между структурой и аминокислотной последовательностью сильно скрученных и изогнутых β -шпилек в глобулярных белках. *Молекуляр. биология.* **50**, 880–886.
6. Gordeev A.B., Kargatov A.M., Efimov A.V. (2010) PCBOST: protein classification based on structural trees. *Biochem. Biophys. Res. Commun.* **397**, 470–471.
7. Sayle R.A., Milner-White E.J. (1995) RASMOL: Biomolecular graphics for all. *Trends Biochem. Sci.* **20**, 374–376.
8. Koradi R., Billeter M., Wüthrich K. (1996) MOLMOL: a program for display and analysis of macromolecular structures. *J. Mol. Graph.* **14**, 51–55.
9. Ефимов А.В. (1986) Стандартные конформации полипептидной цепи в нерегулярных участках белков. *Молекуляр. биология.* **20**, 250–260.
10. Guex N., Peitsch M.C. (1997) SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. *Electrophoresis.* **18**, 2714–2723.
11. Lim V.I. (1974) Structural principles of the globular organization of protein chains. A stereochemical theory of globular protein secondary structure. *J. Mol. Biol.* **88**, 857–872.
12. Ефимов А.В. (1984) Новая супервторичная структура белков: $\alpha\alpha$ -уголок. *Молекуляр. биология.* **18**, 1524–1537.
13. Rao S.T., Rossmann M.G. (1973) Comparison of supersecondary structures in proteins. *J. Mol. Biol.* **76**, 241–256.
14. Levitt M., Chothia C. (1976) Structural patterns in globular proteins. *Nature.* **261**, 552–558.
15. Efimov A.V. (1994) Favoured structural motifs in globular proteins. *Structure.* **2**, 999–1002.
16. Efimov A.V. (2018) Chirality and handedness of protein structures. *Biochemistry (Mosc.)*. **83** Suppl. 1, 103–110.
17. Efimov A.V. (1997) Structural trees for protein superfamilies. *Proteins.* **28**, 241–261.
18. Efimov A.V. (2017) Structural motifs in which β -strands are clipped together with the П-like module. *Proteins.* **85**, 1925–1930.
19. Каргатов А.М., Ефимов А.В. (2018) Уникальные комбинации $\beta\alpha\beta$ -единиц и П-образных модулей в белках и особенности их аминокислотных последовательностей. *Молекуляр. биология.* **1**, 43–50.
20. Кутышенко В.П., Гушина Л.В., Христофоров В.С., Прохоров Д.А., Тимченко М.А., Кудреватых Ю.А., Федюкина Д.В., Филимонов В.В. (2010) Структура и динамика химерного белка SH3-F2 по данным ЯМР. *Молекуляр. биология.* **44**, 1064–1074.
21. Chothia C., Janin J. (1982) Orthogonal packing of beta-pleated sheets in proteins. *Biochemistry.* **21**, 3955–3965.
22. Efimov A.V. (1991) Structure of coiled β - β -hairpins and β - β -corners. *FEBS Lett.* **284**, 288–292.
23. Efimov A.V. (1992) A novel super-secondary structure of β -proteins: a triple-strand corner. *FEBS Lett.* **298**, 261–265.
24. Ефимов А.В., Бошкова Е.А. (2014) Два механизма сворачивания белков: теоретический анализ. *Биорг. химия.* **40**, 665–672.

STRUCTURE AND FEATURES OF AMINO ACID SEQUENCES OF L-MODULES IN SH3-LIKE FOLDS

A. M. Kargatov¹, E. V. Brazhnikov¹, A. V. Efimov^{1, *}

¹*Institute of Protein Research, Russian Academy of Sciences, Pushchino, Moscow Region, 142290 Russia*

*e-mail: efimov@protres.ru

Here we present a detailed stereochemical analysis of non-homologous SH3-like folds and a novel commonly occurring module (L-module) with β -strand–loop– β -strand structure. The β -strands of the L-module are packed orthogonally in the top and bottom β -sheets of SH3-like folds. They bend through 90° when passing from one β -sheet to the other and together with the crossover loop form a half-turn of a right-handed superhelix. A database including 60 nonhomologous SH3-like domains has been compiled using the Protein Data Bank and analyzed for the L-modules occurrence. The most common L-modules have $\beta_m\alpha\alpha\alpha\beta_n$ - and $\beta_m\alpha\alpha\alpha\beta\alpha\beta_n$ -conformations, where m and n are numbers of β -residues in the first and second β -strands, these modules occur most in 57 and 8% SH2-like domain, respectively. When superimposed, these types of the L-modules exhibit much similarity between their 3D structures and a certain pattern of their key hydrophobic, hydrophilic and proline residues detected irrespective of whether they occur in homologous or nonhomologous proteins.

Keywords: α -helix, β -strand, structural motif, structural similarity