

УДК: 519

Конформационный анализ структурных мотивов типа α - α -уголок в вычислительном эксперименте молекулярной динамики

Руднев В.Р.¹, Панкратов А.Н.^{*1}, Куликова Л.И.¹, Дедус Ф.Ф.¹,
Тихонов Д.А.¹, Ефимов А.В.²

¹Институт математических проблем биологии РАН, 142290, Пущино, Московская область, ул. Институтская, д.4

²Институт белка РАН, 142290, Пущино, ул. Институтская, д.4

Аннотация. В настоящей работе рассматривается задача распознавания структурных мотивов белков типа α - α -уголков, решаемая с помощью аналитического описания основной цепи белковой глобулы и спектрального метода распознавания повторов. Описывается проведенная верификация найденных структур по конформационному шаблону, охарактеризованному предельными значениями углов на карте Рамачандрана. Методом молекулярной динамики показано, что α - α -уголки с короткой перетяжкой как автономные структуры устойчивы в водной среде.

Ключевые слова: распознавание образов, структурные мотивы белков, анализ конформаций белковых структур, карта Рамачандрана, молекулярная динамика.

ВВЕДЕНИЕ

Для предсказания пространственной структуры белков необходимо знать все возможные конформации полипептидной цепи и структуры, а также принципы укладки полипептидной цепи в пространственные структуры. Ранее были разработаны простые правила укладки белковой цепи [1], приведшие к построению структурных деревьев, которые описывают многообразие белковых структур. Это поставило задачу автоматического распознавания и предсказания различных структурных мотивов белков. В настоящее время объектом пристального внимания исследователей являются простые структурные мотивы, состоящие из двух элементов вторичной структуры, имеющих уникальные укладки полипептидной цепи в пространстве. Такой интерес к структурным мотивам обусловлен уникальностью структур и их возможностью быть зародышами в процессе сворачивания белков. Структурные мотивы могут быть также использованы в качестве стартовых структур для поиска возможных укладок полипептидной цепи при моделировании структуры белков.

В настоящее время известно несколько структурных мотивов, состоящих из двух элементов вторичной структуры, например, из двух α -спиралей, и имеющих уникальные укладки полипептидной цепи в пространстве (α - α -уголки, α - α -шпильки, L-образные и V-образные структуры и др.). Мотив α - α -шпильки - это супервторичная структура из двух соседних по цепи α -спиралей, которые связаны перетяжкой и упакованы антипараллельно. Шпилька может быть левой или правой в зависимости от того, как расположена вторая α -спираль относительно первой - справа или слева. Длина перетяжек между спиральями тоже может быть разной, при этом каждая стандартная α -

*pan@impb.ru

α -шпилька должна иметь строго определенное и свойственное только ей расположение в цепи гидрофобных, гидрофильных и глициновых остатков [2].

Супервторичные L- и V-образные структуры также образованы двумя спиралями. Особую роль в образовании L-образных структур играет пролин, который способствует излому между двумя α -спиралями. Как и шпильки, L-образные структуры могут быть правыми или левыми [3]. V-образные структуры также очень похожи на α - α -шпильки, у которых сильно удалены друг от друга несвязанные концы α -спиралей; они также похожи и на L-образные структуры. В V-образных структурах α -спирали имеют длину, как правило, не превышающую трех-четырёх витков.

Одним из часто встречаемых структурных мотивов в гомологичных и негомологичных белках являются α - α -уголки [4]. Эта супервторичная структура образована двумя соседними по полипептидной цепи α -спиралями, связанными между собой перетяжкой и упакованными в пространстве ортогонально. В белках α - α -уголки встречаются в форме левой суперспирали. Их последовательности имеют определенное расположение в цепи гидрофобных, гидрофильных и глициновых остатков.

Описанные выше мотивы образованы двумя α -спиралями, расположенными в полипептидной цепи друг за другом и соединенными между собой перетяжками, и представляют собой компактные пространственные структуры у которых есть гидрофобное ядро и полярная оболочка. Известно, что боковые цепи остатков, полностью утопленные в гидрофобном ядре, являются гидрофобными. Гидрофильные боковые цепи или другие полярные группы могут быть утоплены в гидрофобное ядро (или другое гидрофобное окружение), если они участвуют там в образовании водородных или солевых связей [1]. Отсюда следует, что каждая α -спираль должна иметь, по крайней мере, один гидрофобный остаток на виток; при этом гидрофобные остатки должны располагаться на одной стороне α -спирали и образовывать непрерывный гидрофобный кластер [4]. Также известно, что α -спирали в белках упакованы плотно. Наиболее плотная упаковка двух α -спиралей достигается при антипараллельной, перпендикулярной и так называемой скошенной ориентацией между спиралями, примерами такой упаковки являются рассмотренные нами супервторичные структуры: α - α -уголки, α - α -шпильки, L-образные и V-образные структуры.

Наиболее изученными среди α - α -уголков являются уголки с короткими перетяжками. Перетяжка относится к неупорядоченной части молекулы и, в принципе, каждый аминокислотный остаток в ней может иметь определенную конформацию из разрешенных областей на карте Рамачандрана. Поскольку конформационное изменение цепи молекулы значительное (близко к повороту), то конформации остатков короткой перетяжки оказываются зажатыми в более узком диапазоне. Эти закономерности были отмечены в работе [4] и требуют всестороннего исследования и статистического анализа.

Среди всех известных белков достаточно много небольших белков, состоящих, например, всего лишь из одного или двух известных структурных мотивов. А этот факт указывает на то, что такие структурные мотивы устойчивы сами по себе [4]. Устойчивость α - α -уголков была косвенно показана в 1993 году канадскими исследователями F.Tsai и J.Sherman в экспериментальной работе [5]. Кроме того, нами независимо была выдвинута и проверена гипотеза об автономной устойчивости структурных мотивов в вычислительных экспериментах молекулярной динамики (МД) [6]. С одной стороны, эта гипотеза позволяет надеяться на распознавание таких структур по аминокислотной последовательности. С другой стороны, проверка устойчивости *in silico* дает возможность производить автоматическую верификацию найденных структур.

В данной работе полученные траектории были изучены более подробно как с точки зрения феномена устойчивости, так и с точки зрения более точного определения и распознавания структуры.

МЕТОДЫ

1. Распознавание мотивов. Подготовка структур

В данной работе решается задача распознавания супервторичных структур в глобулярных белках с известными 3D структурами, определенными методами рентгеноструктурного анализа и ядерно-магнитного резонанса. Для этого были получены характеристические профили структур белков, представленных в формате PDB, а потом применен спектральный алгоритм поиска заданного паттерна [7] супервторичной структуры в исследуемых белках. Уникальные белковые профили получались путем аналитического описания координат C_{α} -атомов основной цепи белковой глобулы. Предложенный спектральный подход к решению задачи распознавания структурных мотивов описан в предыдущей работе авторов [6].

2. Молекулярная динамика

Моделирование молекулярной динамики (МД) молекулярного фрагмента α - α -уголка и последующий анализ результатов были выполнены с использованием пакета программ AMBER 11 [8]. Моделирование динамики проводилось с явным учетом водного окружения. Для этого при подготовке исходных данных в ячейке периодических граничных условий (ПГУ) был размещен комплекс, окруженный молекулами воды. Количество молекул воды для каждого молекулярного фрагмента было различным, но таким, что толщина водной прослойки от края молекулярного фрагмента до границы ячейки периодических граничных условий составляла 9 ангстрем. Силовые параметры молекул воды выбраны согласно модели TIP3P [8]. Моделирование осуществляли в силовом поле ff03 [10, 11] при температурах 300К. Вначале была произведена минимизация энергии системы при фиксированном положении координат атомов α - α -уголка. Это было сделано для того, чтобы упорядочить атомарные взаимодействия и возможно более равномерно распределить молекулы воды в ячейке ПГУ. Затем был выполнен небольшой разогрев системы при постоянном объеме, при этом положение координат атомов молекулы также было фиксированным. Эта часть расчетов необходима для того, чтобы правильно настроить молекулы воды, окружающие модельный α - α -уголок, относительно границ ячейки ПГУ. Отметим, что этот разогрев длился 10 пс. Далее, были выполнены две последовательные стадии минимизации энергии модели с частичным ослаблением сил, фиксирующих положения атомов молекулы α - α -уголка в пространстве. На следующем этапе вся система уже без ограничений была разогрета до температуры 300К при постоянном объеме. После разогрева, который длился 10 пс, было выполнено моделирование при постоянном давлении продолжительностью 15 пс. При этом ячейка периодических граничных условий подстраивалась таким образом, чтобы получить заданную плотность среды в системе. В результате выполнения перечисленных операций система приводилась к состоянию с желаемыми параметрами, из которого далее рассчитывалась МД траектория продолжительностью 11 нс при постоянном давлении 1 атм.

РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ

С помощью созданного авторами программно-аналитического комплекса «ProteinReviser», основанного на обобщенном спектрально-аналитическом методе [12] и разработанного для распознавания супервторичных структур в глобулярных белках с

решенными 3D структурами методами рентгеноструктурного анализа и ядерно-магнитного резонанса, была сформирована выборка α - α -уголков из базы данных PDB. Выборка представляет собой перечень белков с указанием координат атомов, образующих эти структуры. Из всех распознанных структур были отобраны для дальнейшего исследования 53 α - α -уголка (табл. 1). Все найденные с помощью программно-аналитического комплекса структуры были визуально проверены.

Таблица 1. Перечень отобранных белков с указанием координат атомов структурных мотивов типа α - α -уголок

№	Код PDB	Участок цепи, образующий α - α -уголок	№	Код PDB	Участок цепи, образующий α - α -уголок
1	3C07	41-60	28	1P48	106-137
2	2PIJ	15-36	29	1ELS	107-136
3	2OER	44-63	30	1F4L	55-87
4	2NYX	62-86	31	1ONE	107-137
5	1A76	200-221	32	3B97	106-137
6	2OEM	360-385	33	1B28	16-43
7	3BGW	234-256	34	2PSN	106-138
8	1QAY	14-38	35	1PFU	55-87
9	2CV2	252-272	36	3DBJ	114-145
10	1D1L	15-36	37	1C1D	314-345
11	1A77	200-221	38	1J6O	219-252
12	3F6W	26-46	39	1KCZ	127-159
13	1B9M	34-59	40	2PS6	242-278
14	1F36	74-94	41	3C4I	2-38
15	1B9N	34-59	42	1ETO	49-82
16	2DXI	252-272	43	2H1K	27-56
17	2G7U	30-55	44	2OPX	82-117
18	1E2X	33-58	45	2NX4	8-40
19	1LJ9	47-71	46	1BXG	305-347
20	2CSF	20-44	47	2AEK	242-278
21	1PDZ	107-136	48	2ANI	106-143
22	1PFV	55-87	49	1BW9	305-348
23	1PFY	55-87	50	2F93	70-117
24	2CV1	252-273	51	1Q3Q	96-143
25	1PFW	55-87	52	1A0D	266-286
26	1DU0	27-57	53	1E7P	606-638
27	1JDY	390-419			

Затем была выдвинута гипотеза об автономной устойчивости α - α -уголков, которая легла в основу исследования. Под автономной устойчивостью в данном случае понимается устойчивость пространственной структуры исследуемого структурного мотива отдельно от белковой молекулы, в которой данная структура была обнаружена.

Известно, что все структурные мотивы имеют свои, присущие только им конформационные шаблоны, основанные на распределении значений углов ϕ и ψ на карте Рамачандрана. Для описания конформации полипептидной цепи принято пользоваться определенной номенклатурой [4]: α , α_L , β области на карте Рамачандрана соответствуют правой α -спирали, левой спирали и β структуре соответственно.

Таким образом, конформация каждого остатка полипептидной цепи обозначается одним из символов номенклатуры. Конформацию супервторичной структуры типа α - α -уголок с короткой перетяжкой можно представить в виде шаблона:

$$\alpha_m \alpha_L \beta \alpha_n,$$

где n и m – число остатков α -спиралей.

Для проверки стабильности исследуемых уголков нами был проведен следующий вычислительный эксперимент. Согласно распределению предельных значений углов φ и ψ на карте Рамачандрана для всех возможных областей было проведено конформационное описание всех остатков изучаемых структур:

$$\alpha = \begin{cases} \varphi \in (-95, -40) \\ \psi \in (-60, -10) \end{cases},$$

$$\alpha_L = \begin{cases} \varphi \in (60, 110) \\ \psi \in (-20, 90) \end{cases},$$

$$\beta = \begin{cases} \varphi \in (-150, -60) \\ \psi \in (90, 180) \end{cases}.$$

Далее исследовали изменение конформации структуры уголка в каждом из 5000 кадров эксперимента МД (см. рис. 1 и рис. 2) путем регистрации частоты события попадания торсионных углов аминокислотного остатка в определенные области:

$$P(j, \omega) = \frac{1}{N} \sum_{i=1}^N I(i, j, \omega),$$

$$I(i, j, \omega) = \begin{cases} 1, & \text{если } (\varphi_{ij}, \psi_{ij}) \in \omega \\ 0, & \text{если } (\varphi_{ij}, \psi_{ij}) \notin \omega \end{cases},$$

где P – частота события, I – индикаторная функция, i – номер кадра, j – номер остатка, ω – определенная область на карте Рамачандрана, N – общее количество кадров.

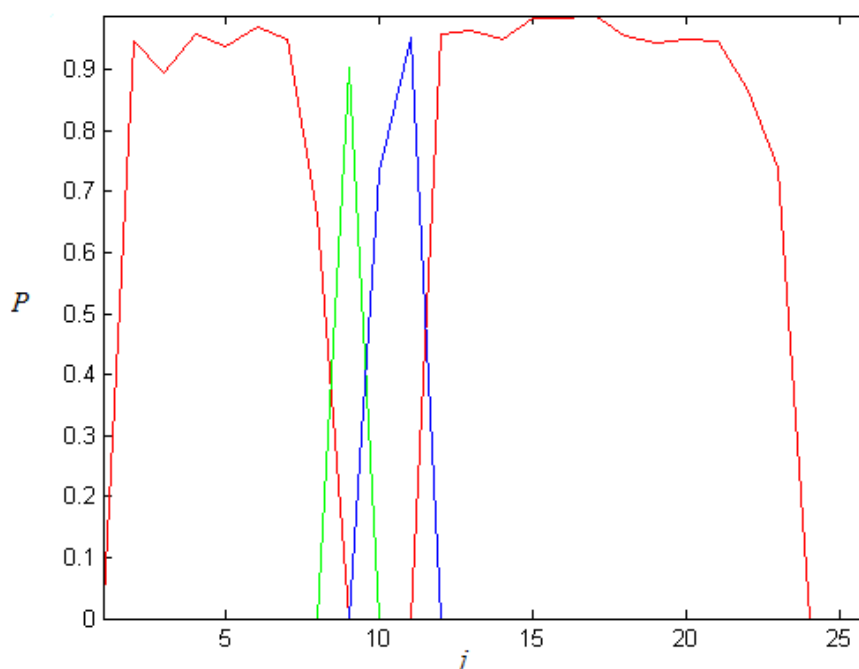


Рис. 1. Частота попадания углов аминокислотного остатка в области α (красным), α_L (зеленым), β (синим) для уголка с короткой перетяжкой белка 1B9M в эксперименте МД.

Результаты эксперимента были таковы: почти все мотивы показали свою автономную устойчивость в том смысле, что конформация структур сохранялась вдоль траектории молекулярной динамики (см. рис. 3 и рис. 4) с некоторыми особенностями, которые описаны ниже.

На рис. 1 представлен результат конформационного анализа вдоль траектории МД для примера α - α -уголка, который хорошо подходит под определение в виде шаблона. Другие структуры могут иметь меньшие частоты, но также значимые для того, чтобы идентифицировать эти структуры как удовлетворяющие шаблону.

Однако среди изученных структур есть и такие, которые не могут быть отнесены к заданному шаблону. Пример такой структуры представлен на рис. 2. Детальное рассмотрение этого мотива позволяет отнести его к уголкам с более длинной перетяжкой. Тем не менее, даже в этом случае вычисленные частоты конформаций остатков вдоль траектории МД позволяют утверждать, что эта структура устойчива.

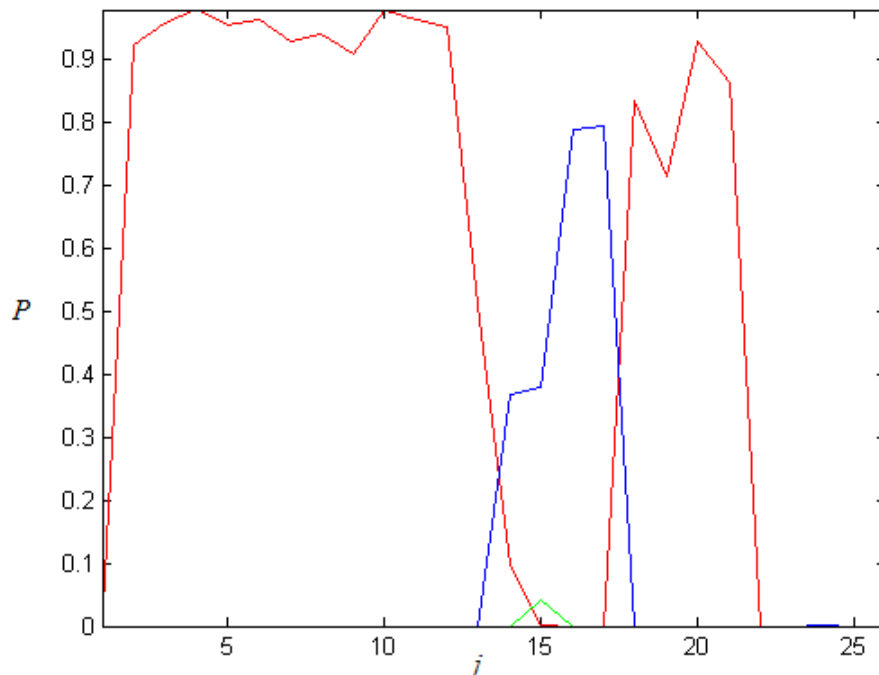


Рис. 2. Частота попадания углов аминокислотного остатка в области α (красным), α_L (зеленым), β (синим) для уголка с длинной перетяжкой белка 3BGW в эксперименте МД.

Группа структур, которые оказались менее устойчивы, и это показал вычислительный эксперимент МД, включила в себя структурные мотивы белков: 2NX4, 3BGW, 2PSN, 2OPX. Спирали в исследуемых мотивах данных белков в ходе эксперимента частично раскручиваются, но угол между спиральями сохраняется (см. рис. 4). Длина и конформация перетяжки α - α -уголков оказались менее стабильны. Таким образом, можно предположить, что конформация и длина перетяжки в подобных структурах вносят немаловажный вклад в устойчивость структуры. Хотя нужно отметить, что α - α -уголки с длинной перетяжкой являются менее изученными структурными мотивами.

На рис. 5 и рис. 6 представлены распределения торсионных углов для обоих классов структур соответственно. Видно, что структуры с короткой перетяжкой соответствуют шаблону и хорошо вписываются в рамки ограничений для углов. Структуры с длинной перетяжкой имеют другое распределение. Было выявлено различие в конформации α_L . Изменение распределения для структур с длинной перетяжкой показано на рис. 7, из которого видно, что произошло перераспределение значений углов в зону ϵ -конформации [1]:

$$\varepsilon = \left\{ \begin{array}{l} \varphi \in (60, 110) \\ \psi \in (-180, -150) \cup (150, 180) \end{array} \right\}.$$

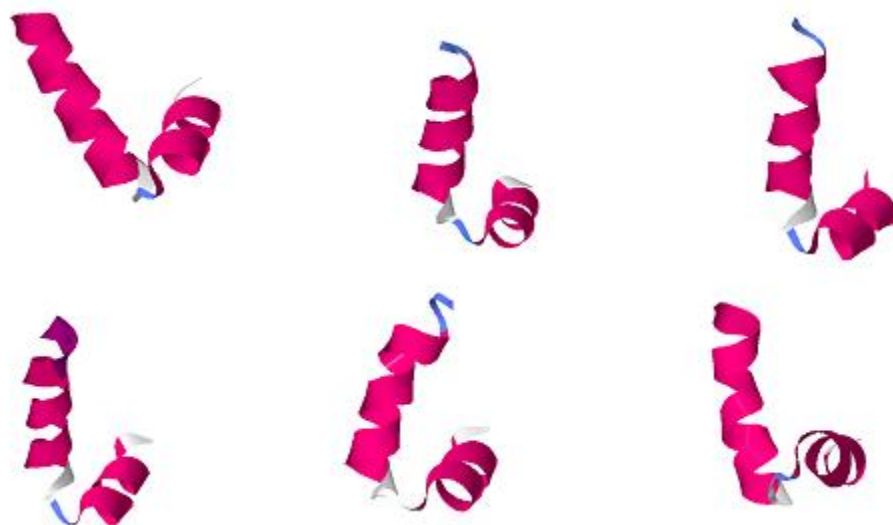


Рис. 3. Поведение α - α уголка с короткой перетяжкой в ходе численного эксперимента МД. Ленточная модель α - α уголка в белке 2G7U (см. табл. 1).

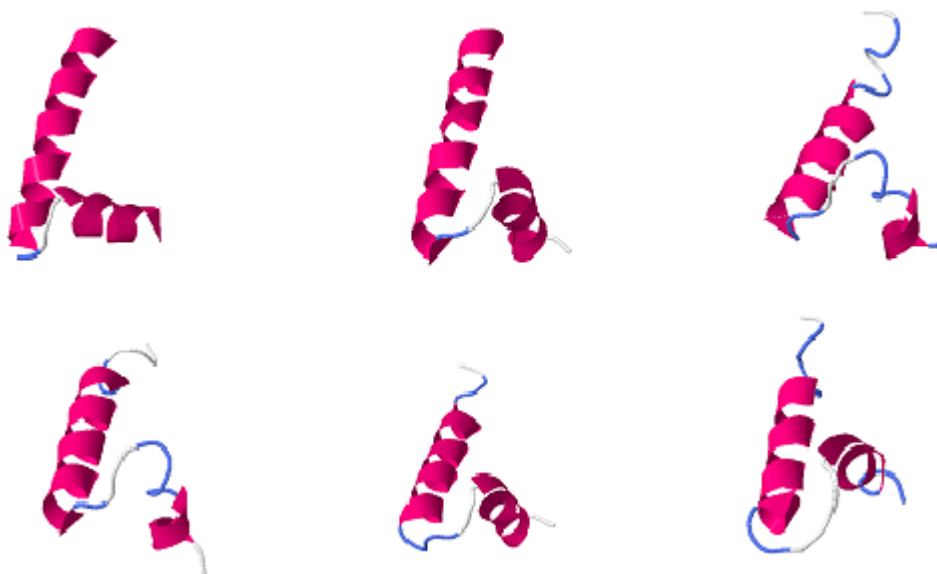


Рис. 4. Поведение α - α уголка с длинной перетяжкой в ходе численного эксперимента МД. Ленточная модель α - α уголка в белке 2PSN (см. табл. 1) в ходе численного эксперимента МД.

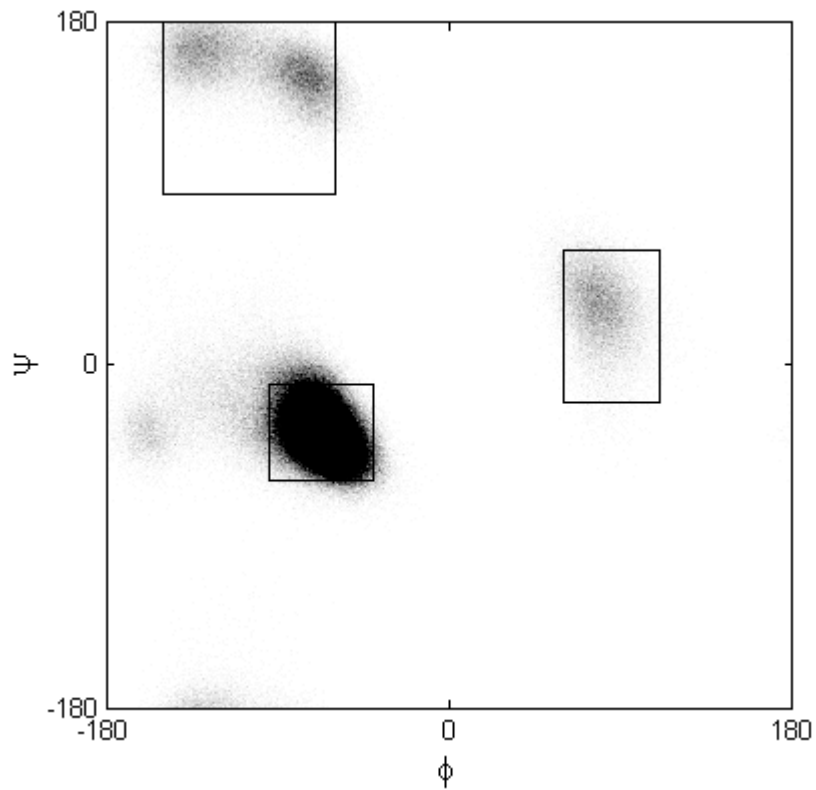


Рис. 5. Распределение торсионных углов всех остатков для структурных мотивов из белков 3C07, 1B9M, 1D1L, 2G7U, принадлежащих классу уголков с короткой перетяжкой, и границы областей для распознавания конформаций шаблона.

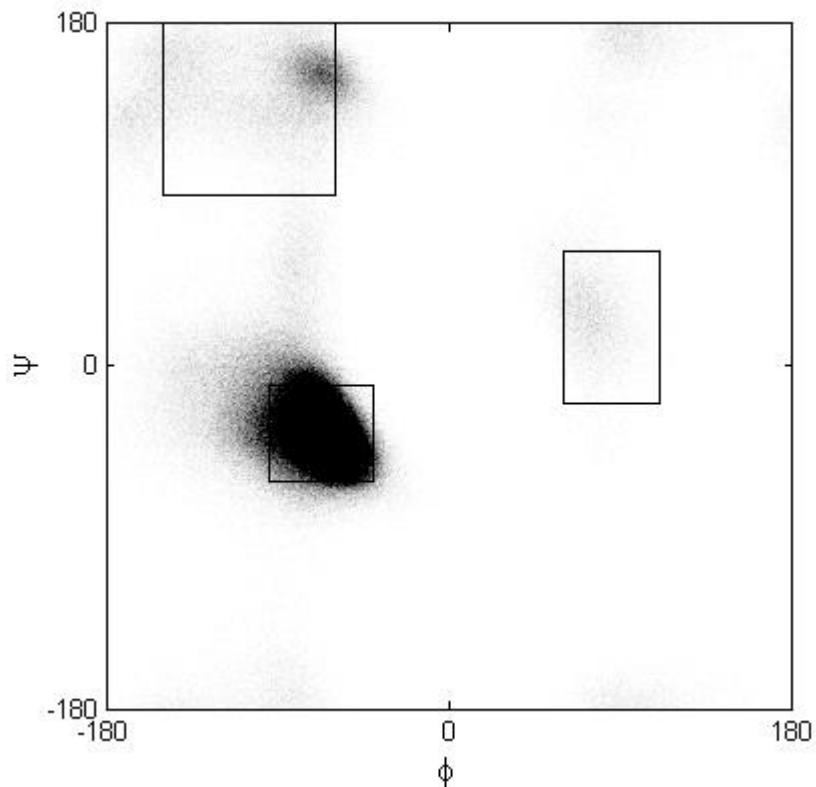


Рис. 6. Распределение торсионных углов всех остатков для структурных мотивов из белков 2NX4, 3BGW, 2PSN, 2OPX, принадлежащих классу уголков с длинной перетяжкой, и границы областей для распознавания конформаций шаблона.

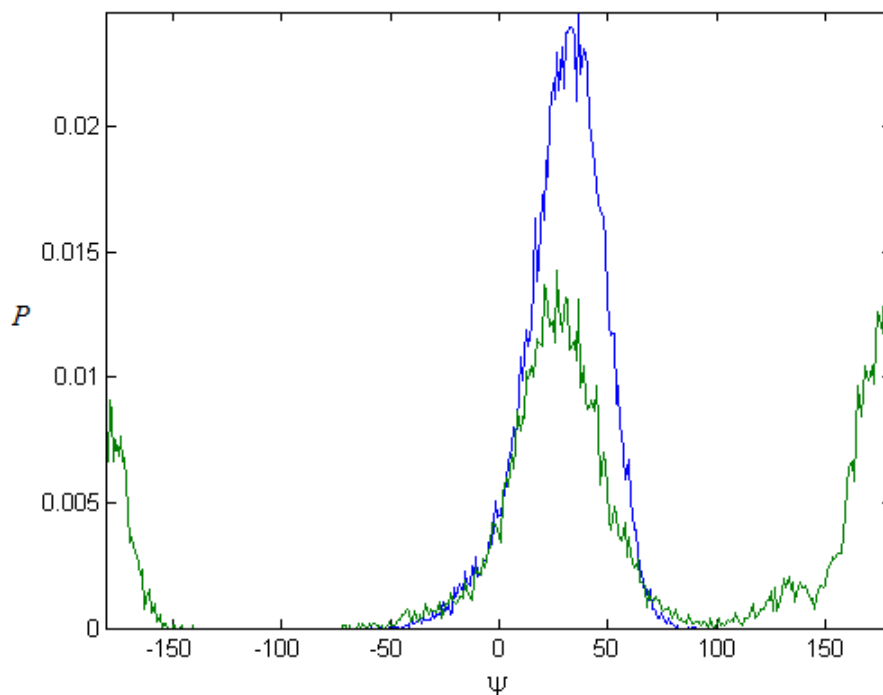


Рис. 7. Различие в распределении конформаций α - α -уголков с короткой (синяя кривая, группа белков из рис. 5) и длинной (зеленая кривая, группа белков из рис. 6) перетяжкой по углу ψ при значениях для угла ϕ , соответствующих α_L конформации.

ЗАКЛЮЧЕНИЕ

На первом этапе работы была сформирована выборка из 53 α - α -уголков в базе данных PDB. Для распознавания был предложен подход, основанный на континуальном описании цепи белковой молекулы и спектральном методе распознавания повторов.

На втором этапе проведен анализ автономной устойчивости α - α -уголков методом молекулярной динамики в водной среде. Эксперимент ставился на всей выборке из 53 α - α -уголков. Моделирование производилось средствами Amber 11. Исследование показало, что α - α -уголок является автономно устойчивой структурой.

На третьем этапе была проведена верификация найденных структур по конформационному шаблону, охарактеризованному предельными значениями углов на карте Рамачандрана. Конформационный анализ

- подтвердил устойчивость структур;
- позволил разделить класс структур с короткой перетяжкой и более общий случай;
- показал, что части длинных α -спиралей, выходящих за пределы гидрофобного ядра, менее устойчивы по сравнению с частями, входящими в гидрофобное ядро α - α -уголка, что проявлялось в частичном расплетании и самосборке спиралей при сохранении общей конформации уголка;
- показал появление статистически значимой части остатков в ϵ -конформации для структур с более длинной перетяжкой;
- позволил определить области возможных флуктуаций конформации структуры и статистически обосновать выбор границ областей на карте Рамачандрана в методе распознавания по конформационному шаблону.

Работа выполнена при поддержке грантов РФФИ, проекты № 13-04-00150, 14-07-31196, 14-07-00924, 13-01-00340.

СПИСОК ЛИТЕРАТУРЫ

1. Efimov A.V. Standard structures in proteins. *Prog. Biophys. Molec. Biol.* 1993. V. 60. P. 201–239.
2. Brazhnikov E.V., Efimov A.V. Structure of α - α -hairpins with short connections in globular proteins. *Molecular Biology*. 2001. V. 35. № 1. P. 89–97.
3. Ефимов А.В. L-Образная структура из двух α -спиралей с остатком пролина между ними. *Молекулярная биология*. 1992. Т. 26. С. 1370–1376.
4. Ефимов А.В. Новая супервторичная структура белков: α - α уголок. *Молекулярная биология*. 1984. Т. 18. С. 1524–1537.
5. Tsai F.C., Sherman J.C. Circular dichroism analysis of a synthetic peptide corresponding to the α - α -corner motif of hemoglobin. *Biochemical and biophysical research communications*. 1993. V. 196. № 1. P. 435–439.
6. Руднев В.Р., Панкратов А.Н., Куликова Л.И., Дедус Ф.Ф., Тихонов Д.А., Ефимов А.В. Распознавание и анализ устойчивости структурных мотивов типа α - α -уголок в глобулярных белках. *Математическая биология и биоинформатика*. 2013. Т.8. № 2. С.398–406. URL: http://www.matbio.org/2013/Rudnev_8_398.pdf (дата обращения: 17.12.2014).
7. Pankratov A.N., Gorchakov M.A., Dedus F.F., Dolotova N.S., Kulikova L.I., Makhortykh S.A., Nazipova N.N., Novikova D.A., Olshevets M.M., Pyatkov M.I., Rudnev V.R., Tetuev R.K., Filippov V.V. Spectral Analysis for identification and visualization of repeats in genetic sequences. *Pattern Recognition and Image Analysis*. 2009. V. 19. № 4. P. 687–692.
8. Case D.A., Cheatham T.E. 3rd, Darden T., Gohlke H., Luo R., Merz K.M. Jr, Onufriev A., Simmerling C., Wang B., Woods R.J. The Amber biomolecular simulation programs. *Journal of Computational Chemistry*. 2005. V. 26. P. 1668–1688.
9. Jorgensen W.L., Chandrasekhar J., Madura J.D., Impey R.W., Klein M.L. Comparison of simple potential functions for simulating liquid water. *Journal of Chemical Physics*. 1983. V. 79. P. 926–935.
10. Duan Y., Wu C., Chowdhury S., Lee M.C., Xiong G., Zhang W., Yang R., Cieplak P., Luo R., Lee T., Caldwell J., Wang J., Kollman P. A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations. *Journal of Computational Chemistry*. 2003. V. 24. P. 1999–2012.
11. Lee M.C., Duan Y. Distinguish protein decoys by using a scoring function based on a new Amber force field, short molecular dynamics simulations, and the generalized Born solvent model. *Proteins*. 2004. V. 55. P. 620–634.
12. Дедус Ф.Ф., Куликова Л.И., Панкратов А.Н., Тетуев Р.К. Классические ортогональные базисы в задачах аналитического описания и обработки информационных сигналов: учебное пособие. М.: Издательский отдел ВМиК МГУ, 2004. 147 с.

Материал поступил в редакцию 09.12.2014, опубликован 23.12.2014.