

## СТРУКТУРНО-ФУНКЦИОНАЛЬНЫЙ АНАЛИЗ БИОПОЛИМЕРОВ И ИХ КОМПЛЕКСОВ

УДК 577.150.2

### НОВОЕ СТРУКТУРНОЕ ДРЕВО $\beta$ -БЕЛКОВ, СОДЕРЖАЩИХ abcd-ЕДИНИЦЫ

© 2008 г. А. Б. Гордеев, М. С. Кондратова, А. В. Ефимов\*

Институт белка Российской академии наук, Пушкино, Московская обл., 142290

Поступила в редакцию 16.07.2007 г.

Принята к печати 16.08.2007 г.

Создана база данных  $\beta$ -белков, содержащих abcd-единицы, которая включает в себя 528 белков из Банка белковых данных (PDB), в том числе 244 негомологичных (выборка из 1511 PDB-файлов). Построено обновленное структурное древо для белков этого класса, содержащее 153 возможные укладки полипептидной цепи. На основе построенного древа разработана современная структурная классификация  $\beta$ -белков, содержащих abcd-единицы. Вся информация о базе данных, а также структурное древо доступны в Интернете по адресу: <http://strees.protres.ru/>.

**Ключевые слова:** моделирование, классификация, сворачивание белков, структурный мотив, структурное сходство.

NOVEL STRUCTURAL TREE FOR  $\beta$ -PROTEINS CONTAINING abcd-UNITS, by A. B. Gordeev, M. S. Kondratova, A. V. Efimov\* (Institute of Protein Research, Russian Academy of Sciences, Pushchino, Moscow Region, 142290 Russia; \*e-mail: [efimov@protres.ru](mailto:efimov@protres.ru)). A database of 528  $\beta$ -proteins and  $\beta$ -domains containing abcd-units (among them 244 are nonhomologous) has been compiled from the Protein Data Bank (total 1511 PDB entries). A novel structural tree for this structural class of proteins that is composed of 153 possible polypeptide chain folds has been constructed. The structural classification of  $\beta$ -proteins containing abcd-unit based on the structural tree has been developed. Both the database and the structural tree are accessible at the web-site (<http://strees.protres.ru/>).

**Key words:** modeling, classification, protein folding, structural motif, structural similarity.

Структурное древо белков – это совокупность всех разрешенных промежуточных и конечных пространственных структур, которые могут быть получены из одной корневой (стартовой) структуры путем последовательного пристраивания к ней других элементов вторичной структуры в соответствии с набором правил, выведенных из известных принципов структурной организации белков. В качестве корневой структуры древа берется соответствующий структурный мотив, имеющий уникальную укладку цепи в пространстве. Возможные пути роста структур показываются линиями, которые в итоге объединяют все структуры в одно древо.

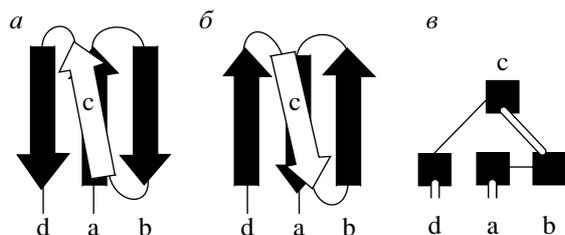
Первые варианты структурных деревьев построены около 10 лет назад [1–4]. За это время существенно выросло количество расшифрованных структур в Банке белковых данных (Protein Data Bank). Так, например, в состав структурного древа  $\beta$ -белков, содержащих abcd-единицы, варианта 1997 г. [2] входило около сорока известных белковых структур, а к настоящему времени мы собрали базу данных из 528 белков и доменов это-

го класса. Это предопределило необходимость построения обновленного структурного древа для этого класса белков. Построение и анализ обновленного древа позволяет исследовать новые пути роста структур, вести поиск новых способов укладки полипептидной цепи в пространстве, обнаруживать новые закономерности в белковых структурах и с учетом всего этого модернизировать правила построения структурных деревьев. Решению этих задач и посвящена настоящая работа.

#### ОБЪЕКТЫ И МЕТОДЫ ИССЛЕДОВАНИЯ

Базу данных  $\beta$ -белков, содержащих abcd-единицы, создавали с помощью системы структурной классификации SCOP (<http://scop.mrc-lmb.cam.ac.uk/scop/>). Отбор белков осуществляли вручную. Всего отобрано 528 белков и доменов, содержащих abcd-единицы, из них негомологичных – 244. Программу BLAST для попарного выравнивания (<http://www.ncbi.nih.gov/BLAST/>) использовали для выявления возможной гомологии. Анализ структуры белков проводили визуально с

\* Эл. почта: [efimov@protres.ru](mailto:efimov@protres.ru)



**Рис. 1.** Схематическое изображение abcd-единицы с прямым (а) и обратным (б) направлением цепи.  $\beta$ -Тяжи показаны в виде стрелок, направленных от N- к C-концам. в – Вид abcd-единицы с торца,  $\beta$ -тяжи изображены квадратиками; перетяжки, направленные к наблюдателю, показаны двойными линиями, а удаленные – одинарными линиями. Буквами а, б, с и d обозначены соответствующие  $\beta$ -тяжи.

использованием программы молекулярной графики RasMol [5].

Построение структурного дерева проводили в соответствии с ранее сформулированными правилами [1–4]:

1) в качестве корневой структуры дерева использована abcd-единица, которая, как известно [6], имеет уникальную пространственную укладку цепи. Простейший вариант abcd-единицы состоит из четырех следующих друг за другом вдоль цепи  $\beta$ -тяжей а, б, с и d, три из которых (а, б, и d) лежат в одном слое, а четвертый (с) – в другом слое. Тяжи б, с и d образуют правую суперспираль bcd. abcd-единица может быть в двух вариантах – с прямым и обратным ходом полипептидной цепи (рис. 1а и 1б). При построении дерева abcd-единица, так же, как и все другие структуры, представлена в упрощенном виде (рис. 1в) без указания направления цепи. При этом подразумевается, что разрешен как прямой, так и обратный ход цепи в каждом случае;

2) пристраивание других  $\beta$ -тяжей к растущим корневой и промежуточным структурам проводили последовательно, шаг за шагом; при этом каждая структура, полученная на предыдущем этапе, сохранялась в составе последующей и т.д. На каждом этапе  $\beta$ -тяж, расположенный ближе других к растущей структуре по цепи, пристраивался первым [2, 6];

3) в соответствии с принципом плотной упаковки все полученные структуры должны быть компактными;

4) пересечение перетяжек [7] и образование узлов [8] запрещено;

5) все структурные мотивы (т.е. не только корневые мотивы) должны иметь свойственную им хиральность и пространственную укладку цепи;

6) в двухслойных  $\beta$ -белках три подряд идущие по цепи  $\beta$ -тяжа не должны образовывать  $\beta$ - $\beta$ - $\beta$ -суперспираль, в которой первый и третий  $\beta$ -тяжи

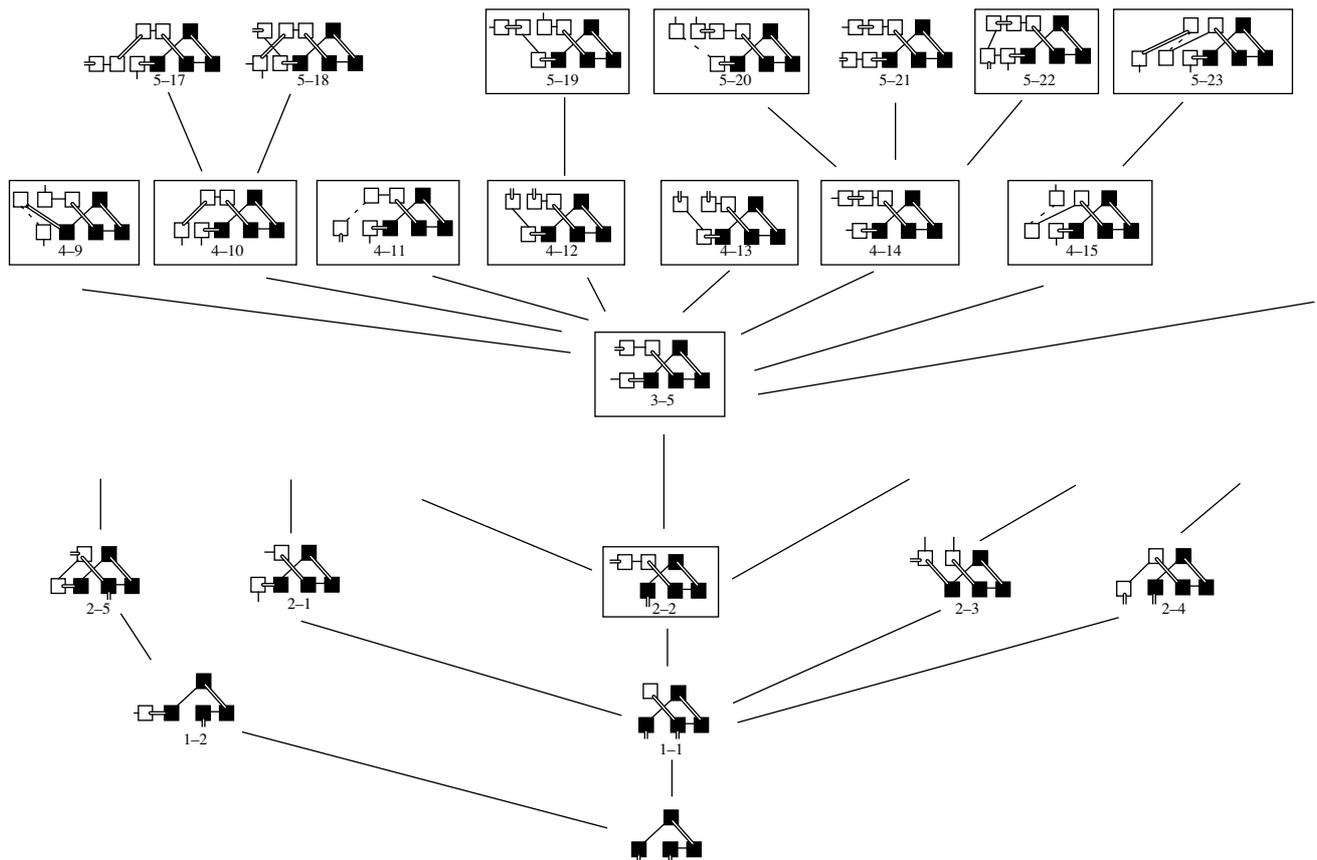
непосредственно взаимодействуют друг с другом с образованием параллельной  $\beta$ -структуры [2, 6]. Образование суперспирали тремя подряд идущими  $\beta$ -тяжами разрешено, если в  $\beta$ -слое между первым и третьим тяжами находится, по крайней мере, еще один тяж (например суперспираль bcd и а-тяж в abcd-единице). Как показано в данной работе, это правило хорошо срабатывает при пристраивании к abcd-единице ближних 3–4 тяжей и нарушается (приблизительно в 25–30% случаев) при завершении роста структур, когда пристраиваются пятый, шестой или седьмой  $\beta$ -тяжи.

## РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ

На рис. 2 представлен фрагмент построенного структурного дерева. Полное структурное дерево белков, содержащих abcd-единицы, размещено на сервере Института белка РАН по адресу (<http://strees.protres.ru/>). Как видно, в структурном древе можно выделить несколько уровней или рядов. В каждом уровне (ряду) находятся возможные укладки полипептидной цепи, состоящие из одинакового количества  $\beta$ -тяжей. Все укладки в древе пронумерованы, и каждая укладка цепи имеет свой номер, состоящий из двух чисел, разделенных точкой или тире. Первое число указывает количество добавленных к abcd-единице  $\beta$ -тяжей. Например, укладки первого уровня (ряда) содержат один добавленный  $\beta$ -тяж и обозначаются цифрой 1, второго ряда – цифрой 2 и т.д. Второе число обозначает порядковый номер укладки, если считать их слева направо в каждом ряду. Например, в четвертом ряду находятся всего 36 упаковок, и они имеют номера с 4–1 до 4–36.

С другой стороны, структурное дерево имеет несколько ветвей. В составе одной ветви укладка цепи, находящаяся на более высоком уровне, содержит в себе укладки, расположенные ниже. Укладки из разных ветвей содержат в себе одну и ту же укладку, находящуюся в месте разветвления. Чем выше в структурном древе находится точка разветвления, тем выше уровень структурного сходства между белками или доменами соответствующих ветвей.

Всего обновленное дерево содержит 153 укладки. До уровня пяти добавочных  $\beta$ -тяжей в древо включены все теоретически разрешенные укладки. На более высоких уровнях показаны только те пути роста структур, которые ведут к структурам известных белков. Одна из главных причин этого состоит в том, что белки и домены имеют ограниченный размер, и большая часть белков этого класса (~75%) состоит из abcd-единицы и 3–5 добавочных  $\beta$ -тяжей. Другая причина – в том, что количество возможных упаковок в верхних уровнях резко возрастает, и в целях экономии места на древе показана только часть из них. Тем не менее даже в таком виде обновленное древо со-



**Рис. 2.** Фрагмент структурного дерева  $\beta$ -белков, содержащих abcd-единицы. Все структуры ориентированы одинаковым образом, и показан их вид с торца, как на рис. 1в. Укладки цепи, найденные в белках, обведены в рамочки.

держит в несколько раз больше возможных укладок, чем первый вариант дерева (153 укладки по сравнению с 36 в первом варианте [2]), и почти на порядок больше известных белков и доменов (528 белков по сравнению с 40 в первом варианте). На обновленном дереве представлен также ряд новых путей роста промежуточных структур, которые приводят к новым типам укладок цепи. Интересно отметить, что из 18 укладок цепи в первом варианте дерева [2], которые не были в то время обнаружены в белках, 7 укладок сейчас найдены в известных структурах белков. Таким образом, правильно предсказаны, по крайней мере, 7 неизвестных в то время укладок цепи.

Анализ структурного дерева позволяет исследовать процесс “созревания” белковой глобулы или домена и, прежде всего, выяснить причины прекращения роста структур. Как отмечалось выше, белки и домены имеют ограниченный размер, и наблюдается больше всего белков и доменов с четырьмя добавочными  $\beta$ -тяжами (около 37%, если принимать во внимание только негомологичные белки). Населенность белками других уровней дерева быстро снижается, и нами найден всего один белок (из 244), имеющий 8 добавочных

$\beta$ -тяжей. Таким образом, “зрелый” белок или домен должен состоять из оптимального числа  $\beta$ -тяжей.

С другой стороны, нами обнаружено, что нарушения правил пристраивания элементов вторичной структуры к растущим структурам наблюдаются, как правило, на последних этапах роста структур. Например, правило 6 нарушается в 20 белках и доменах, относящихся к 7 различным укладкам, и во всех случаях это происходит при завершении роста структур. Аналогично, левые  $\beta\alpha\beta$ -единицы (как правило, они правые) в  $\alpha/\beta$ -белках также образуются на последних этапах роста этих белков или доменов [2]. По-видимому, такие “нарушения” являются одним из способов остановки роста структур. Однако понятно, что эта проблема требует дальнейшего изучения.

Одно из наиболее важных применений структурных деревьев – это их использование для структурной классификации белков. Все белки и домены, входящие в состав одного дерева, могут быть отнесены к одному структурному классу или суперсемейству. Белки и домены, принадлежащие ветвям структурного дерева, образуют подклассы. Как видно, такая классификация базируется только на сходстве пространственных струк-

тур и на общности смоделированных путей сворачивания. В ней не учитываются аминокислотные последовательности, а также информация о функциях и эволюционном родстве белков, что в той или иной степени учитывается в других известных классификациях [9–11].

На основе обновленного структурного древа нами создана иерархически организованная база данных всех  $\beta$ -белков, содержащих abcd-единицы, которая доступна в Интернете (<http://strees.protes.ru/>). С титульной страницы сайта можно перейти на страницы с фрагментами структурного древа или войти в систему поиска нужного белка по PDB-коду. Страницы логически связаны друг с другом с помощью контекстных переходов. Можно загрузить PDB-файлы всех содержащихся в базе данных белков (всего 1511 PDB-файлов для 528 белков и их мутантов) и посмотреть их с помощью любой программы молекулярной графики. Для облегчения работы с базой данных имеется страница с инструкцией.

В настоящее время ведется работа по созданию баз данных и построению обновленных структурных деревьев других структурных классов. Все структурные деревья и базы данных будут доступны в Интернете.

Работа выполнена при поддержке Российского фонда фундаментальных исследований (07-04-00659).

## СПИСОК ЛИТЕРАТУРЫ

1. Efimov A.V. 1996. A structural tree for  $\alpha$ -helical proteins containing  $\alpha$ - $\alpha$ -corners and its application to protein classification. *FEBS Lett.* **391**, 167–170.
2. Efimov A.V. 1997. Structural trees for protein superfamilies. *Proteins.* **28**, 241–260.
3. Efimov A.V. 1997. A structural tree for proteins containing  $3\beta$ -corners. *FEBS Lett.* **407**, 37–41.
4. Efimov A.V. 1998. A structural tree for proteins containing S-like  $\beta$ -sheets. *FEBS Lett.* **437**, 246–250.
5. Sayle R.A., Milner-White E.J. 1995. RASMOL – biomolecular graphics for all. *Trends Biochem. Sci.* **20**, 374–376.
6. Ефимов А.В. 1982. Супервторичная структура  $\beta$ -белков. *Молекуляр. биология.* **16**, 799–806.
7. Лим В.И., Мазанов А.Л., Ефимов А.В. 1978. Стереохимическая теория пространственной структуры глобулярных белков. I. Высокоспиральные промежуточные структуры. *Молекуляр. биология.* **12**, 206–213.
8. Richardson J.S. 1977.  $\beta$ -Sheet topology and relatedness of proteins. *Nature.* **268**, 495–500.
9. Murzin A.G., Brenner S.E., Hubbard T., Chothia C. 1995. SCOP: a structural classification of proteins database for the investigation of sequences and structures. *J. Mol. Biol.* **247**, 536–540.
10. Holm L., Sander C. 1999. Protein folds and families: sequence and structure alignments. *Nucl. Acids Res.* **27**, 244–247.
11. Orengo C.A., Michie A.D., Jones S., Jones D.T., Swindells M.B., Thornton J.M. 1997. CATH – a hierarchic classification of protein domain structures. *Structure.* **5**, 1093–1108.