

УДК: 519

Структура петлевых участков в β - α - и α - β - дугах abCd-единиц в глобулярных белках

Бражников Е.В., Каргатов А.М., Ефимов А.В.

Институт белка РАН, Пуцино, Московская область, ул. Институтская, д.4.

Аннотация. Проведен анализ конформаций около 600 петлевых участков (перетяжек) в β - α - и α - β -дугах, принадлежащих одному из структурных мотивов, часто встречающемуся в белках, abCd-единице. Всего в негомологичных белках отобрано 258 abCd-единиц с «обратным» ходом полипептидной цепи (236 PDB-файлов) и 69 abCd-единиц с «прямым» ходом (65 PDB-файлов). Исследованы четыре типа дуг: β - α - и α - β -дуги при «прямом» ходе цепи; β - α - и α - β -дуги при «обратном» ходе цепи. Для каждого типа дуг определены частоты встречаемости перетяжек разной длины и построены соответствующие гистограммы. Установлено, что наибольшую встречаемость имеют дуги длиной до трех аминокислотных остатков (57 %). β - α -Дуги из abCd-единиц с «прямым» ходом цепи чаще других состоят из двух аминокислотных остатков (44 %), которые в 86 % случаев имеют конформацию $\beta_m\alpha\beta_n$. В них нет глицинов и пролинов, а в позиции β часто встречается аспарагин. В этом типе дуг в перетяжках из одного остатка с конформациями $\beta_m\epsilon\alpha_n$ или $\beta_m\alpha_L\alpha_n$ чаще других встречаются глицины. В α - β -дугах с «прямым» ходом цепи чаще других (18 %) встречаются перетяжки из четырех аминокислотных остатков. Здесь нет преимущественной конформации перетяжек. В β - α -дугах с «обратным» ходом цепи чаще других (17 %) встречаются перетяжки из семи аминокислотных остатков, и большинство из них (88 %) имеют конформацию $\beta_m\alpha_L\beta\beta\alpha\beta\beta\alpha_n$. В α - β -дугах с «обратным» ходом цепи чаще других (32 %) встречаются перетяжки из одного аминокислотного остатка (все глицины) с конформацией дуг $\alpha_m\epsilon\beta_n$ или $\alpha_m\alpha_L\beta_n$. Проведенный в данной работе структурный анализ abCd-единицы содержит полезную информацию для предсказания пространственной структуры белков, а также для молекулярного моделирования белковых структур *de novo*.

Ключевые слова: структура белков, конформационный анализ, α - β -дуги, β - α -дуги, abCd-единица, глицин, пролин.

ВВЕДЕНИЕ

Понимание связи между аминокислотной последовательностью белка и его пространственной структурой имеет ключевое значение для развития методов предсказания этой структуры и моделирования каждого этапа ее сворачивания. Молекулы многих глобулярных белков содержат участки полипептидной цепи, имеющие сходную упаковку элементов вторичной структуры. Такие повторяющиеся участки принято называть структурными мотивами. Они могут выступать как в качестве зародышей, так и в качестве готовых блоков при сворачивании белков в уникальные структуры [1–5], что позволяет использовать их для создания системы структурной классификации белков [6].

В настоящей работе мы провели систематический анализ структуры перетяжек в abCd-единицах и показали, что несколько их типов, имеющих относительно

небольшую длину (до семи остатков), встречаются существенно чаще других. Также были установлены особенности аминокислотных последовательностей перетяжек, наиболее часто встречающихся в белках.

Структурный мотив, известный как $\alpha\beta\text{Cd}$ -единица [7], представляет собой компактную двухслойную замкнутую структуру, образованную непрерывным участком полипептидной цепи. Он состоит из трех антипараллельно упакованных β -тяжей и α -спирали, соединяющей крайние β -тяжи и упакованной в другом слое. Этот мотив может иметь как «прямой», так и «обратный» ход полипептидной цепи: при «прямом» ходе средний тяж β -листа оказывается первым в цепи, а при «обратном» – последним. Сама $\beta\alpha\beta$ -единица включает в себя две дуги: β - α -дугу и α - β -дугу (рис. 1).

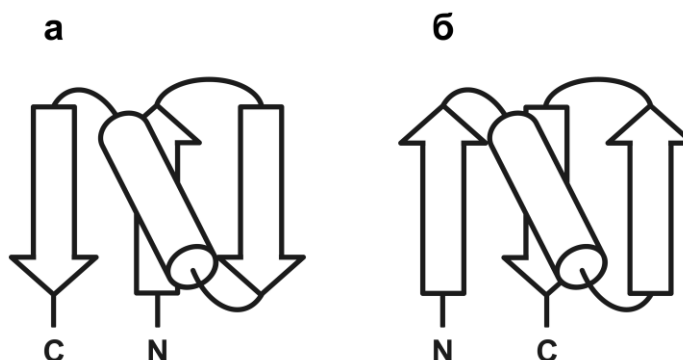


Рис. 1. Схематическое представление $\alpha\beta\text{Cd}$ -единицы с прямым (а) и обратным (б) ходом полипептидной цепи. Стрелками показаны β -тяжи, цилиндрами – α -спирали.

Ранее мы провели ряд исследований, направленных на поиск связи между аминокислотной последовательностью и пространственной структурой некоторых структурных мотивов. Была определена структура α -спиральных шпилек с короткими перетяжками (до четырех аминокислотных остатков): показано существование только пяти типов таких перетяжек, а также выявлено различие в распределении гидрофобных и гидрофильных остатков между правыми и левыми $\alpha\alpha$ -шпильками [8, 9]. При исследовании $\alpha\alpha$ -уголков было установлено, что гидрофобные остатки располагаются в них практически одинаково независимо от степени гомологии содержащих их белков [10]. В сильно скрученных и изогнутых β -шпильках обнаружено существенное различие во встречаемости некоторых аминокислотных остатков на вогнутой и выпуклой поверхностях шпильки. Показано, что внутренние позиции аномально часто содержат остатки глицина и аланина, особенно в местах наибольшей скрученности тяжей, и при этом никогда не содержат остатков пролина, тогда как во внешних позициях в области скручивания их встречается относительно много [11, 12]. Наконец, изучение 3β -уголков показало аномально высокую встречаемость во внутренних позициях остатков глицина, аланина, валина и полное отсутствие остатков пролина [4].

В нашей работе мы провели анализ более сложного структурного мотива – $\alpha\beta\text{Cd}$ -единицы - с использованием большой базы данных. Ранее проводились структурные исследования только $\beta\alpha\beta$ -единицы в работе [13] на примере 17 α/β -белков; β - α - и α - β -переходы (дуги) полипептидной цепи из одного слоя в другой изучались в работах [14–16], однако исследования проводились на малом количестве белков без привязки к определенным структурным мотивам по причине их недостаточного количества в банке белковых данных (Protein Data Bank, PDB) в то время.

ЭКСПЕРИМЕНТАЛЬНАЯ ЧАСТЬ

Для работы использовалась база данных по структурной классификации белков PCBOST, созданная нами ранее на основе банка белковых данных (PDB) и доступная по адресу <http://strees.protres.ru/>. Всего в этой базе данных имеется 926 abCd-единиц [6, 17], из которых 23 % имеют «прямой» ход полипептидной цепи, а 77 % – «обратный». Анализ проводили только на негомологичных белках, отобранных с помощью программы bl2seq (Blast 2 Sequences) методом попарного выравнивания (<http://blast.ncbi.nlm.nih.gov/Blast.cgi>) [18]. Выделение abCd-единиц из полной структуры белка и анализ их структуры осуществляли визуально при помощи программы молекулярной графики RasMol [19]. Углы внутреннего вращения ϕ и ψ вычисляли программой MOLMOL [20], а также с помощью нашей собственной программы, которая дополнительно позволяет определить все области конформаций аминокислотных остатков на карте Рамачандрана.

РЕЗУЛЬТАТЫ И ИХ ОБСУЖДЕНИЕ

Предварительный визуальный анализ пространственной структуры abCd-единиц в белках показал, что перетяжки в β - α - и α - β -дугах в этом структурном мотиве могут сильно отличаться по длине и конформации в зависимости от хода полипептидной цепи. Количественный анализ распределения перетяжек, принадлежащих β - α - и α - β -дугам, по длинам в негомологичных белках представлен в таблице 1 и на рисунке 2.

Таблица 1. Встречаемость перетяжек разной длины в β - α - и α - β -дугах с «прямым» и «обратным» ходом полипептидной цепи

| Прямой ход | β - α -дуги | | | | | | | | | | | | | |
|----------------------------|--------------------------|----|----|----|----|----|----|----|---|---|----|----|----|-----|
| Число остатков в перетяжке | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | >12 |
| Количество перетяжек | 10 | 3 | 28 | 7 | 4 | 2 | 3 | 0 | 1 | 2 | 1 | 0 | 1 | 1 |
| Всего перетяжек | 63 | | | | | | | | | | | | | |
| Прямой ход | α - β -дуги | | | | | | | | | | | | | |
| Число остатков в перетяжке | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | >12 |
| Количество перетяжек | 3 | 6 | 7 | 8 | 10 | 3 | 3 | 2 | 2 | 3 | 1 | 0 | 2 | 5 |
| Всего перетяжек | 55 | | | | | | | | | | | | | |
| Обратный ход | β - α -дуги | | | | | | | | | | | | | |
| Число остатков в перетяжке | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | >12 |
| Количество перетяжек | 29 | 14 | 32 | 28 | 15 | 17 | 10 | 39 | 9 | 6 | 9 | 6 | 3 | 8 |
| Всего перетяжек | 225 | | | | | | | | | | | | | |
| Обратный ход | α - β -дуги | | | | | | | | | | | | | |
| Число остатков в перетяжке | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | >12 |
| Количество перетяжек | 17 | 71 | 26 | 39 | 26 | 18 | 6 | 5 | 6 | 5 | 4 | 0 | 0 | 2 |
| Всего перетяжек | 225 | | | | | | | | | | | | | |

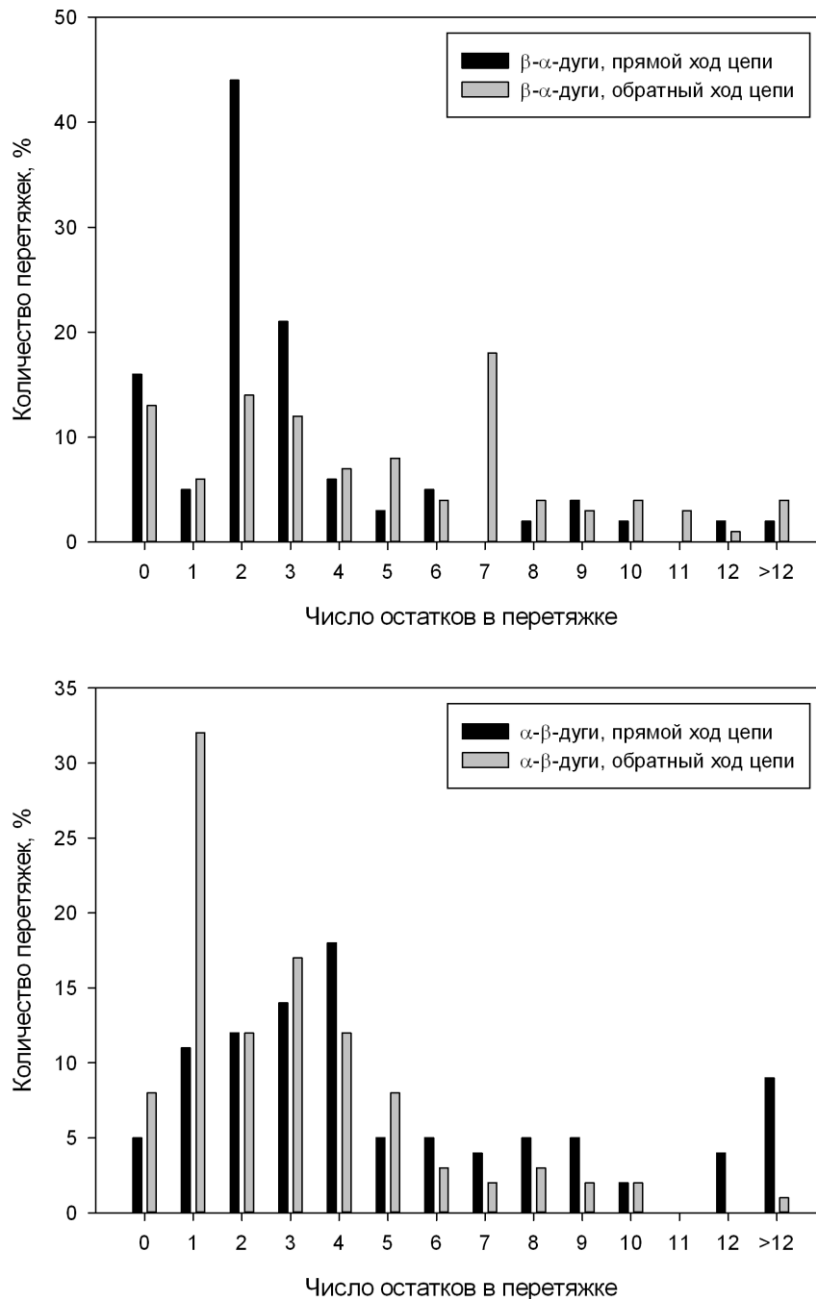


Рис. 2. Зависимость количества перетяжек от их длины в abCd-единицах с «прямым» и «обратным» ходом полипептидной цепи. Вверху – β - α -дуги, внизу – α - β -дуги. Темные линии – «прямой» ход, серые линии – «обратный» ход.

Установлено, что во всех типах дуг наибольшую распространённость имеют перетяжки длиной до трёх аминокислотных остатков (57 %). В β - α -дугах из abCd-единиц с «прямым» ходом цепи чаще других (44 %) встречаются перетяжки из двух остатков, а в β - α -дугах из abCd-единиц с «обратным» ходом цепи – перетяжки из семи остатков (17 %). В другом типе дуг - в α - β -дугах из abCd-единиц с «прямым» ходом цепи чаще других (18 %) встречаются перетяжки из четырех остатков, а в α - β -дугах из abCd-единиц с «обратным» ходом цепи – перетяжки из одного остатка (32 %). Таким образом, частота встречаемости перетяжек одинаковой длины во всех четырех случаях различна. Особенно заметно повышенное содержание перетяжек из двух остатков в β - α -дугах с «прямым» ходом цепи (44 %) и в α - β -дугах с одним остатком в перетяжке (32 %) в случае «обратного» хода полипептидной цепи.

Таблица 2. Основные конформации перетяжек в β - α - и α - β -дугах

| Количество остатков в перетяжке | Конформация перетяжки | |
|---------------------------------|---|--|
| | β - α -дуга | α - β -дуга |
| 0 | – | – |
| 1 | α_L, ε | α_L, ε^* |
| 2 | $\alpha\beta$ | $\beta\alpha, \alpha_L\alpha$ |
| 3 | $\alpha_L\beta\beta, \alpha\alpha\beta$ | $\beta\alpha\alpha, \alpha_L\beta\alpha$ |

*При «прямом» ходе цепи ε -конформация не встречается.

Примечания: конформации основных перетяжек при «прямом» и «обратном» ходе полипептидной цепи одинаковы, различается их количество. Дуги с перетяжками длиной два или три остатка с другими конформациями встречаются редко (по 1–2 раза).

Если перетяжка состоит только из одного остатка, то во всех случаях (как при «прямом», так и «обратном» ходе полипептидной цепи в β - α - и α - β -дугах) он имеет конформацию α_L или ε (табл. 2). Причем, большинство остатков (78 %) имеют конформацию α_L . Также наблюдается ограниченное число конформаций для перетяжек из двух и трех остатков. Перетяжки из двух остатков в β - α -дугах имеют, как правило, $\alpha\beta$ -конформацию, а в α - β -дугах – $\beta\alpha$ - или $\alpha_L\alpha$ -конформацию. В этом случае перетяжки с другими конформациями встречаются редко. В перетяжках из трех остатков в β - α -дугах преимущественно встречается $\alpha_L\beta\beta$ - и $\alpha\alpha\beta$ -конформации, а в α - β -дугах – $\beta\alpha\alpha$ - и $\alpha_L\beta\alpha$ -конформации. Встречаются также перетяжки с другими конформациями, но редко.

На рисунке 3 показано сопоставление аминокислотных последовательностей для $\alpha_m\alpha_L\beta_n$ -дуг с перетяжкой из одного аминокислотного остатка (здесь m и n – число остатков в α -спирали и β -тяже соответственно). Из 62 дуг с конформацией перетяжки α_L глицин присутствует в 44-х случаях, что составляет 71 %. В остальных перетяжках содержатся, в основном, остатки с гибкими боковыми цепями (лизин, глутаминовая кислота, аспарагин, аспарагиновая кислота). Отметим, что последний аминокислотный остаток α -спирали в этих дугах в большинстве случаев находится в γ -конформации.

Перетяжка с ε -конформацией (сопоставление аминокислотных последовательностей не показано) встречается реже – 15 дуг с «обратным» ходом полипептидной цепи (все – остатки глицина), а в дугах с «прямым» ходом ε -конформация не встречается.

В β - α -дугах перетяжек с одним остатком меньше (табл. 1), и они находятся также в ε - и α_L -конформациях. Дуги с «прямым» ходом встречаются три раза, а с «обратным» – четырнадцать. Глицин встречается в 12 перетяжках, остальные остатки – серин, аспарагиновая кислота, глутаминовая кислота, глутамин.

Перетяжки из двух остатков часто встречаются во всех четырех случаях, но их больше всего (44 %) в β - α -дугах с «прямым» ходом цепи. В α - β -дугах с «прямым» ходом цепи таких перетяжек только 13 %. Перетяжки в β - α -дугах имеют преимущественно $\alpha\beta$ -конформацию, а в α - β -дугах $\beta\alpha$ -конформацию. На рисунках 4 и 5 в качестве примера приведено сопоставление (выравнивание) аминокислотных последовательностей таких дуг при «прямом» и «обратном» ходе полипептидной цепи, соответственно.

| PDB код | αααααααααααα | α _L | ββββββββββββββ |
|---------|-----------------------------------|-------------------------|----------------------|
| | «Прямой ход» полипептидной цепи | | |
| 1. | 1B33 (N) | <u>YENWFREQQRIQKM</u> | GKIVKVEL |
| 2. | 1DT0 (A) | PLTI | DTPLLTCD |
| 3. | 1E8X (A) | <u>QWVRQCLKN</u> | EEIHLVLD |
| 4. | 1LDT (L) | SCIARCN | VSIKSE |
| 5. | 1PCE (A) | <u>ECKLCLARIEN</u> | QDIQIVKD |
| 6. | 1WI0 (A) | <u>AMLSYYYSTVMEQQVN</u> | QLIEPLQIFP |
| | «Обратный» ход полипептидной цепи | | |
| 1. | 1AUV (A) | <u>QHDFQDIASVVALT</u> | TYATTEP |
| 2. | 1AKO (A) | LEEVAKL | YNVIFYHQ |
| 3. | 1B24 (A) | <u>KALLEIVSRWLNNL</u> | VRNTIHL |
| 4. | 1CJW (A) | <u>DALVPEFYQRF</u> | FHPAGPC |
| 5. | 1DD5 (A) | <u>LSLIEKAINASDL</u> | LNPIN |
| 6. | 1DQ3 (A) | <u>IETFEIINDYLEKTF</u> | RKYSIQKD |
| 7. | 1DQ3 (A) | <u>IEDVTHYLNAL</u> | IKARIREK |
| 8. | 1EWQ (A) | <u>EDAERLARAL</u> | LVLTHKT |
| 9. | 1FJC (A) | <u>EDELKEVF</u> | DALEIRLVSQ |
| 10. | 1G6S (A) | <u>IRFADVLEKM</u> | ATICWG |
| 11. | 1G6S (A) | <u>YIDITLNLMKTF</u> | VEIENQ |
| 12. | 1GKZ (A) | <u>GLPTSRAYAEYL</u> | GSLQLQSLQIGTDVY |
| 13. | 1GMU (A) | <u>PFMLAKACYALGNR</u> | VPLQIMPGLRYH |
| 14. | 1GXT (A) | <u>FRPFVWQLAQQ</u> | LHGDCVNDGDGVEVRL |
| 15. | 1IQ4 (A) | <u>PKALDSAVEELTLIA</u> | QRPVVTR |
| 16. | 1JG5 (A) | <u>PELMQQL</u> | ASKRRVLGNNFYEYV |
| 17. | 1JJ2 (R) | <u>KGEVADAVEEQY</u> | VTV |
| 18. | 1J00 (A) | <u>RETKQLIINAIVRET</u> | AAQVQTIGHILVLY |
| 19. | 1JRM (A) | <u>KANREIIKEFSETF</u> | RDVEIVSGQKSRQKTI |
| 20. | 1JVA (A) | <u>TSLMERVTEYAEKL</u> | LCAEYKDRKEPQVAKT |
| 21. | 1JVA (A) | <u>TSVRDGLVSLARSL</u> | LVVSVNAE |
| 22. | 1K3E (A) | <u>ALEILNANLWFAENG</u> | PYLCYEAGAQSLLALR |
| 23. | 1KAF (A) | <u>EHHIQKFTDI</u> | MSCKIAKNGNVYLDIK |
| 24. | 1KN6 (A) | <u>PGGQEAASATAEEL</u> | YDLLGQIGSLENHYLFK |
| 25. | 1LFW (A) | <u>EAVKLAYESFLADK</u> | LDGSFEINDESADIVLI |
| 26. | 1NTF (A) | <u>GPACVKNFQSLTTSK</u> | YTKLKNITITETMGLTVYCL |
| 27. | 1O8B (A) | <u>RSAVARQLVKL</u> | GRPEYRQGVVTDN |
| 28. | 1PAV (A) | <u>KKDAPAWIQKS</u> | QELVGVFDRNGYIEIVMKK |
| 29. | 1QMI (A) | <u>HLTAVKAATEIC</u> | ATVEGAELGSQRLLFR |
| 30. | 1QMI (A) | <u>FIRRVLEPLLAKE</u> | IHQOTTL |
| 31. | 1R3N (A) | <u>KMIVAASEIAQRH</u> | GLFTCGIIDAK |
| 32. | 1R62 (A) | <u>GLSIARNLIDQH</u> | GKIEFTSWPGHTEFSVYL |
| 33. | 1SEI (A) | <u>KIKREIAEILKRE</u> | FIRDYEYIED |
| 34. | 1SPB (P) | <u>AAKKKDVISEK</u> | GKVQKQF |
| 35. | 1U8S (A) | <u>ICNEVRLVTOA</u> | CNIIDSRIAMF |
| 36. | 1U8S (A) | <u>LTEKFTQFFAQR</u> | IGMASLSAQTI |
| 37. | 1UAE (A) | <u>KDVDTSMKLLSQL</u> | AKVERNGSVHID |
| 38. | 1UAE (A) | <u>VDLHISGLEQL</u> | ATIKLEEGYVKAS |
| 39. | 1UAE (A) | <u>PEIVDTANFLITL</u> | AKISGQTDRIVIE |
| 40. | 1UAE (A) | <u>DAVLAKLRDA</u> | ADIEVGEDWISLD |
| 41. | 1UAE (A) | <u>MHVPELSRM</u> | AHAETESNTVICH |
| 42. | 1URR (A) | <u>FRKHTSHEAKRL</u> | VRGWCMNTRD |
| 43. | 1uv7 (A) | <u>LVSWIAYLQERQ</u> | VSVDAIDID |
| 44. | 1vi7 (A) | <u>LTGIEALLGQC</u> | GKIINSDYQAFVL |
| 45. | 1yre (A) | <u>LRAQGAIDKL</u> | AQREGVL |
| 46. | 1zo0 (A) | <u>DRAALLRTFSSFL</u> | FEIVRPGH |
| 47. | 1zvp (A) | <u>LTAAFATKLAEH</u> | ISANVI |
| 48. | 2a6o (A) | <u>MRLKEIIQEVAKEL</u> | VEIIEMQT |
| 49. | 2bun (A) | <u>LDIVETSQAHNARA</u> | LTGALFYS |
| 50. | 2cxa (A) | YQR | IFPWFSPGDPILWW |
| 51. | 2cy2 (A) | <u>KGRGFYEH</u> | GVLLGEREIEL |
| 52. | 2fgc (A) | <u>KQEIFQLVEIF</u> | GKIIDVSREGAI |
| 53. | 2fiu (A) | <u>YVSTAKPAFERF</u> | ANFLAR |
| 54. | 2fkl (A) | <u>HTVAKETCSEK</u> | TNL |
| 55. | 2i6c (A) | <u>AAGLLLYTQL</u> | YQPRAIAERHD |
| 56. | 2if1 (A) | <u>VKAFKKKF</u> | CNGTVIE |

Рис. 3. Структурное выравнивание аминокислотных последовательностей α-β-дуг с конформацией перетяжки α_L. Конформации остатков в столбцах показаны греческими буквами в верхней строке. Слева приведены PDB-коды соответствующих белков. Подчеркнуты остатки в перетяжках.

| № | PDB код | $\beta\beta\beta\beta\beta\beta\beta\beta\beta\beta\beta\beta\beta\beta$ | $\alpha\beta$ | $\alpha\alpha\alpha\alpha\alpha\alpha\alpha\alpha\alpha\alpha\alpha\alpha$ |
|----|----------|--|---------------|--|
| 1 | 1A9X (A) | IIRPSFTMGGSGGGIA | <u>YN</u> | REEFEEICARGLDL |
| 2 | 1AUV (A) | VVK_GHAHSG_GKVKV | <u>DN</u> | QHDFQDIASVVA |
| 3 | 1D4B (A) | LVLEEDGTAV | <u>DS</u> | EDFFQL |
| 4 | 1EHI (A) | VVKAANQGSSVGISRV | <u>TN</u> | AEEYTEALSDS |
| 5 | 1EN2 (A) | CSIWGWCG | <u>DS</u> | EPYC |
| 6 | 1GSO (A) | IKA_____KGVIVA | <u>MT</u> | LEEAEAAV |
| 7 | 1HPT (A) | CGTDGDTY | <u>PN</u> | ECVLCFENRK |
| 8 | 1HQZ (A) | IISPNAKKEYEPESTG | <u>SS</u> | FHDFL |
| 9 | 1LDT (L) | VCGSDGRTY | <u>AN</u> | SCIARCN |
| 10 | 1LR7 (A) | VCGLDGKTY | <u>RN</u> | ECALLKARCKE |
| 11 | 1M4J (A) | ISIENEQLVVGSCSPP | <u>SD</u> | SWEQDYDSFVLP |
| 12 | 1OEY (J) | NYRDAEGDLVRL | <u>LS</u> | DEIVAL_VRQA |
| 13 | 1PCE (A) | CGTDGVT | <u>ES</u> | ECKLCLARIEN |
| 14 | 1R0R (I) | CGSDNKTY | <u>GN</u> | KCNFCNAVVE |
| 15 | 1SPH (A) | VNLEYNGKTVNL | <u>KD</u> | IMGVMSL |
| 16 | 1TBR (R) | CGSDGETY | <u>SN</u> | PCTLNCAK |
| 16 | 1TBR (R) | CGSDDITY | <u>DN</u> | NCRLECAS |
| 17 | 1W96 (A) | ASEGGGKGIQV | <u>ER</u> | EEDFIALYHQAAN |
| 18 | 1WI0 (A) | EYEDGDGRITV | <u>RS</u> | DEEMKAMLSYYS |
| 19 | 1WMH (B) | GYTDAHGDLLPL | <u>TN</u> | DDSLHRA |
| 20 | 1YB3 (A) | LFIYG | <u>AE</u> | NFL |
| 21 | 2C60 (A) | NELSILL | <u>KN</u> | QDDLKDAIDILD |
| 22 | 2BBH (A) | YSIEEFREFKT | <u>TD</u> | VESVLPFR |
| 23 | 2SCU (B) | VKVV | <u>NS</u> | KEDIRAFENW |

Рис.4. Структурное выравнивание аминокислотных последовательностей β - α -дуг с конформацией перетяжки $\alpha\beta$ («прямой» ход). Конформации остатков в столбцах показаны греческими буквами в верхней строке. Слева приведены PDB-коды соответствующих белков. Подчеркнуты остатки в перетяжках.

| | PDB код | $\beta\beta\beta\beta\beta\beta\beta\beta\beta\beta\beta\beta$ | $\alpha\beta$ | $\alpha\alpha\alpha\alpha\alpha\alpha\alpha\alpha\alpha\alpha\alpha\alpha$ |
|----|----------|--|---------------|--|
| 1 | 1AUV (A) | HAHSG_GKVKV | <u>DN</u> | QHDFQDIASVVA |
| 2 | 1A79 (A) | VIV | <u>FD</u> | KNGISKLSA |
| 3 | 1B24 (A) | EYRVVITQ | <u>KS</u> | ENLIQHIAP |
| 4 | 1B24 (A) | IWN | <u>KN</u> | KALLEIVSRWLNN |
| 5 | 1B64 (A) | SILLVVKP | <u>WD</u> | DETMKLEECV |
| 6 | 1BHT (A) | TKKV | <u>NT</u> | ADQCANRCTR |
| 7 | 1CC8 (A) | HYQFNVM | <u>TC</u> | SGCSGAVNKVLT |
| 8 | 1DD5 (A) | LVIKP | <u>WD</u> | KSVLSLIEKAINA |
| 9 | 1DQ3 (A) | RIYD | <u>YR</u> | IETFEIINDYLEK |
| 10 | 1DQ3 (A) | IELGM | <u>VN</u> | KRLIEDVTHYLNA |
| 11 | 1FTR (A) | MARVLITA | <u>AS</u> | HKWAMIAVKEAT |
| 12 | 1FTR (A) | AGGNFYIMA | <u>ES</u> | QPAGLQAEAAVD |
| 13 | 1GMU (A) | SVVRC | <u>DD</u> | PFMLAKACYALGN |
| 14 | 1HKY (A) | EAKA | <u>AS</u> | SQACQELCEKD |
| 15 | 1I8N (A) | TKTEF | <u>TS</u> | VDECRKMCEE |
| 16 | 1J3E (A) | A | <u>AD</u> | EQTLLKN |
| 17 | 1JVA (A) | ATIKT | <u>IH</u> | TSVRDGLVSLARS |
| 18 | 1K8K (D) | TVVFSTVF | <u>KD</u> | DDDVVIGKVMQE |
| 19 | 1KR4 (A) | LVYSTF | <u>PN</u> | EEKALEIGRKLLE |
| 20 | 1N0U (A) | QVAVEV | <u>KN</u> | ANDLPKLVEGLKR |
| 21 | 1NSA (A) | VFRVNV | <u>ED</u> | ENDISELHELAST |
| 22 | 1PVZ (A) | KC | <u>AT</u> | DADCSRKCP |
| 23 | 1SIS (A) | PCFT | <u>TD</u> | PNMAKKCRDC |
| 24 | 1UTA (A) | CGSF | <u>RG</u> | AEQAETVRAQLAF |
| 25 | 1zpw (A) | LYAVAYDI | <u>PD</u> | DTRRVKLANLLKS |
| 26 | 2atz (A) | VFDYNG | <u>RN</u> | PERFYHKAQLLLR |

Рис. 5. Структурное выравнивание аминокислотных последовательностей β - α -дуг с конформацией перетяжки $\alpha\beta$ («обратный» ход). Конформации остатков в столбцах показаны греческими буквами в верхней строке. Слева приведены PDB-коды соответствующих белков. Подчеркнуты остатки в перетяжках.

Видно, что при прямом ходе имеется только одна перетяжка, в которой присутствует глицин. Первый остаток перетяжки (α -конформация) содержит маленькие остатки или остатки с гибкими боковыми цепями (четыре аспарагиновые кислоты, три серина, три треонина, два лизина, два аланина, две глутаминовые кислоты, один метионин, один пролин, один лейцин, один аспарагин), а во втором остатке (β -конформация) присутствуют одиннадцать аспарагинов, семь серинов, три аспарагиновые кислоты, один аргинин, одна глутаминовая кислота.

Аналогичное распределение аминокислотных остатков и в случае «обратного» хода цепи (26 перетяжек). Только в β -конформации наблюдается большее содержание глутаминовой кислоты (девять остатков). В случае прямого хода таких остатков только три. Мы не рассматриваем здесь перетяжки из двух остатков с другими конформациями, поскольку они встречаются редко.

| | PDB код | $\beta\beta\beta\beta\beta$ | $\alpha L\beta\beta\alpha\alpha\beta\beta$ | $\alpha\alpha\alpha\alpha\alpha\alpha\alpha\alpha\alpha\alpha$ |
|-----|----------|-----------------------------|--|--|
| 1. | 1B7F (A) | LIVN | YLPQDMT | DRELYALFRAI |
| 2. | 1B7F (A) | LYVT | NLPRTIT | DDQLDTIFGKY |
| 3. | 1FJC (A) | LLAK | NLSFNIT | EDELKEVF |
| 4. | 1H6K (X) | LYVG | NLSFYTT | EEQIYELFSK |
| 5. | 1HD1 (A) | MFIG | GLSWDTT | KKDLKDY |
| 6. | 1O00 (B) | ILFVT | SIHEEAQ | EDEIQEKFC |
| 7. | 1S79 (A) | YIK | GFPTDAT | LDDIKEWLE |
| 8. | 1SJR (A) | LRIIVE | NLFYPVT | LDVLHQIF |
| 9. | 1uw4 (A) | VVIR | RLPPTLT | KEQLQEH |
| 10. | 1we1 (A) | VYLK | GLPFEAE | NKHVIDF |
| 11. | 1wg1 (A) | LVK | NLPQDSN | CQEVHDLK |
| 12. | 1wg4 (A) | VLVS | GLPPSGS | WQDLKDHMRE |
| 13. | 1whw (A) | RLFVR | NLSYTSS | EEDLEKLFSA |
| 14. | 1why (A) | WVG | GLGPNTS | LAALAREFDR |
| 15. | 1wi8 (A) | TAFLG | NLPYDVT | EESIKEF |
| 16. | 1x4a (A) | IYVG | NLPDIR | TKDIEDVFK |
| 17. | 1x4e (A) | LYIR | GLQPGTT | DQDLVKL |
| 18. | 1x4g (A) | VYCG | GIASGLT | DQLMRQTFSPF |
| 19. | 1x4h (A) | FIR | NLSFDSE | EEALGEV |
| 20. | 1x5o (A) | LYIS | NLPLSMD | EQELENMLK |
| 21. | 1x5t (A) | IFIG | NLDPEID | EKLLYDTFSA |
| 22. | 2cpd (A) | LYVR | NLMLSTS | EEMIEKEFN |
| 23. | 2cpe (A) | IYVQ | GLNDSVT | LDDLADFFKQ |
| 24. | 2cpj (A) | LFVG | NLPDIT | EEEMRKL |
| 25. | 2cq0 (A) | TIRVT | NLSEDT | ETDLQEL |
| 26. | 2cq1 (A) | LHIR | KLPGEVT | ETEVIAGLPL |
| 27. | 2cq3 (A) | LHVS | NIPFRFR | DPDLRQMFQ |
| 28. | 2cq4 (A) | VFCM | QLAARIR | PRDLEDF |
| 29. | 2cqb (A) | YVG | GLAEEVD | DKVLHAAFI |
| 30. | 2cqd (A) | IFVG | GLPYHTT | DASLRKYFEG |
| 31. | 2cqh (A) | YIG | NLSPAVT | ADDLRQLFGD |
| 32. | 2dgx (A) | DVQVS | NIDYRLS | RKELQLLQEA |
| 33. | 2dis (A) | RLFIG | GIPKMKK | REEILEEIAKV |
| 34. | 2f9d (A) | LYIR | NLPYKIT | AEEMYDIF |
| 35. | 2if1 (A) | TTVQ | GIADDYD | KKKLVKAF |

Рис. 6. Структурное выравнивание аминокислотных последовательностей β - α -дуг с конформацией перетяжки $\alpha L\beta\beta\alpha\alpha\beta\beta$ («обратный» ход). Конформации остатков в столбцах показаны греческими буквами в верхней строке. Слева приведены PDB-коды соответствующих белков. Подчеркнуты остатки в перетяжках.

Перетяжки из трех остатков встречаются во всех типах дуг (не показаны). В β - α -дугах найдено семь перетяжек при «прямом» ходе цепи и 28 перетяжек при «обратном». Почти половина всех перетяжек имеет конформацию $\alpha L\beta\beta$. Чаще всего в первой позиции присутствуют глицины, встречаются также аспарагины, аспарагиновая кислота, лизины, но по одному - двум остаткам каждой аминокислоты. Есть несколько дуг с перетяжкой $\alpha\alpha\beta/\delta$. В α - β -дугах нет какой-либо преимущественной конформации перетяжек, а третий остаток перетяжки может быть в α -, αL - или ϵ -конформации.

Из дуг с более длинными перетяжками при «прямом» ходе чаще других встречаются α - β -дуги с перетяжкой из четырех остатков (18 %), а при «обратном» ходе - β - α -дуги с перетяжками из семи остатков (17 %). В перетяжках из четырех остатков нет преимущественной конформации перетяжки. Первый остаток может быть в α -, γ -, α_L -, ε - или β_P -конформации, а последний остаток чаще всего находится в β -конформации. Из 39 β - α -дуг с перетяжкой из семи остатков 35 дуг имеют первый остаток в конформации α_L (рис. 6). Преимущественно это остатки аспарагина (18 случаев) и глицина (12 случаев). Большинство таких дуг имеет конформацию $\beta_m\alpha_L\beta\beta\alpha\alpha\beta\beta\alpha_n$.

| PDB код | $\beta\beta\beta\beta\beta\beta\beta\beta\beta\beta\beta\beta$ | $\alpha\alpha\alpha\alpha\alpha\alpha\alpha\alpha\alpha\alpha\alpha$ |
|---------|--|--|
| | «Прямой» ход полипептидной цепи | |
| 1 | 1B33 (N) | YFTKLVP YENWFREQQRIQK |
| 2 | 1B6R (A) | QWRLR ANETEQLPAEC |
| 3 | 1DCQ (A) | ILTC IECSG IHREL |
| 4 | 1F0Z (A) | QIVP REQWA |
| 5 | 1G25 (A) | HTLC ESCVDLLEFVR |
| 6 | 1I35 (A) | RIKLITE EEFKQICFNS |
| 7 | 1ODH (A) | YLRPAIC DKARQKQQR |
| 8 | 1RM6 (B) | KRTVP VAQLFRESG |
| 9 | 1SEI (A) | VLT DKEARQKGTGGEII |
| 10 | 1T1H (A) | QTYE RSSIQKWLDA |
| | «Обратный» ход полипептидной цепи | |
| 1 | 1CC8 (A) | HYQFNVM TCSGCSGAVNKVLT |
| 2 | 1CJW (A) | AVLMCE DALVPFYQR |
| 3 | 1CKV (A) | VLVLKKS DEINTFIEEILLTD |
| 4 | 1EWQ (A) | YYLEVT RPYYER |
| 5 | 1G9Z (A) | QVTQK TQRRWFLDKLVDEI |
| 6 | 1GNK (A) | LVTVIK PFKLEDVREALSSI |
| 7 | 1I1G (A) | VTITGVDTK PEKLFVEAEKLE |
| 8 | 1IUJ (A) | VTMNRIPVR PEYAEQFEEAFRQR |
| 9 | 1KN6 (A) | AAEI PGGQEAASATAEEL |
| 10 | 1N0U (A) | VFLVEIQCP EQAVGGIYSVLNK |
| 11 | 1P8K (Z) | GIELS IKDVQLIYKIKKIL |
| 12 | 1R89 (A) | VFAVKFRKPDIV DDNLYPQLERASRK |
| 13 | 1RY9 (A) | VMLWANFDAPS DVKLQSSAYNILNL |
| 14 | 1S7H (A) | GFRFSLYPMT DDFISVIKSALKK |
| 15 | 1SEI (A) | LEVPAS KIKREIAEILKR |
| 16 | 1T8H (A) | VD DRVVDSL RPT |
| 17 | 1T95 (A) | KIP PEHTGRAISALYNF |
| 18 | 1uv7 (A) | VWIQPLP FSQLVSWIAYLQER |
| 19 | 1vi7 (A) | EYTLQCE YHQLTGIEALLGQ |
| 20 | 1vk8 (A) | TVSIKVVPAVE DGRLHEVIDRAIEK |
| 21 | 1whr (A) | KKFPQMT SYHRMLLHRVAAYF |
| 22 | 1xrs (B) | QLSFTLPLKNN ERSAEAAKQIAL |
| 23 | 2a1v (A) | LSLKVD PERGEELRQA |
| 24 | 2bop (A) | FALISGT ANQVKCYRFRVKKN |
| 25 | 2cy2 (A) | VWVL KENPKGRGFYEH |
| 26 | 2fb5 (A) | TLNAHLT APLLESI |
| 27 | 2ibo (A) | SIALQVLPVQ GIDRIAVIDQVIAYL |
| 28 | 2p8i (A) | WNAHVYFD ASSRDAAWTLREQIE |
| 29 | 2tbd (A) | FAIYTT KEKAALLYKKIMEKY |

Рис. 7. Структурное выравнивание аминокислотных последовательностей β - α -дуг с нулевой перетяжкой. Конформации остатков в столбцах показаны греческими буквами в верхней строке. Слева приведены PDB-коды соответствующих белков.

Нулевые перетяжки также встречаются во всех четырех типах дуг. На рисунке 7 представлена укладка таких дуг в abCd-единице с «прямым» и «обратным» ходом полипептидной цепи.

Последний остаток β -тяжа в β - α -дугах из abCd-единиц с «обратным» ходом цепи является преимущественно гидрофильным (76 %). Также здесь могут встречаться пролины (3 случая). Отсутствие массивных гидрофобных остатков объясняется их

свойством вызывать дегидратацию свободных NH-групп. Аналогичная ситуация наблюдается и в случае «прямого» хода цепи, где эта β -позиция также занята преимущественно гидрофильными остатками и пролинами. Заметим, что в перетяжках этого типа совсем не встречаются глицины, хотя в перетяжках из одного остатка они присутствуют почти всегда (рис. 3).

ЗАКЛЮЧЕНИЕ

Проведенный нами анализ структуры перетяжек в α - β - и β - α -дугах, содержащихся в abCd-единицах, показал, что перетяжки могут существенно отличаться по длине и конформации, а также иметь разный аминокислотный состав. Главный результат состоит в том, что большинство дуг в структуре abCd-единиц содержат короткие перетяжки (до трех аминокислотных остатков), и эти остатки имеют ограниченное число конформаций. Как правило, в перетяжках присутствуют глицины (за исключением перетяжек из двух аминокислотных остатков). В определенных ϵ - и α -позициях коротких перетяжек должны находиться глицины или остатки с гибкими боковыми цепями, а в первых позициях α -спиралей – небольшие или гидрофобные остатки. Как известно, α -спиральные и β -структурные участки также должны иметь определенное чередование гидрофобных и гидрофильных остатков. Все это определяет необходимые условия образования abCd-единиц. Полученные результаты имеют большое теоретическое значение и могут быть использованы как при моделировании белковых структур, содержащих abCd-единицы, так и для предсказания пространственной структуры белков.

Работа выполнена при финансовой поддержке Российского фонда фундаментальных исследований, грант РФФИ № 13-04-00150.

СПИСОК ЛИТЕРАТУРЫ

1. Phillips D.C. The hen egg-white lysozyme molecule. *Proc. Natl. Acad. Sci. USA*. 1967. V. 57. P. 483–495.
2. Птицын О.Б. Стадийный механизм самоорганизации белковых молекул. *Докл. АН СССР*. 1973. Т. 210. С. 1213–1215.
3. Wetlaufer D.B. Nucleation, rapid folding, and globular intrachain regions in proteins. *Proc. Natl. Acad. Sci. USA*. 1973. V. 70. P. 697–701.
4. Ефимов А.В., Бошкова Е.А. Два механизма сворачивания белков. Теоретический анализ. *Биоорг. хим.* 2014. Т. 40. С. 665–672.
5. Руднев В.П., Панкратов А.Н., Куликова Л.И., Дедус Ф.Ф., Тихонов Д.А., Ефимов А.В. Конформационный анализ структурных мотивов типа α - α -уголок в вычислительном эксперименте молекулярной динамики. *Математическая биология и биоинформатика*. 2014. Т. 9, № 2. С. 575–584.
6. Gordeev A.V., Kargatov A.M., Efimov A.V. PCBOST: Protein classification based on structural trees. *Biochemical and Biophysical Research Communications*. 2010. V. 397. P. 470–471.
7. Ефимов А.В. Супервторичная структура β -белков. *Молекулярная биология*. 1982. Т. 16. С. 799–806.
8. Efimov A.V. Structure of α - α -hairpins with short connections. *Protein Engineering*. 1991. V. 4. P. 245–250.
9. Бражников Е.В., Ефимов А.В. Структура α -спиральных шпилек с короткими перетяжками в глобулярных белках. *Молекулярная биология*. 2001. Т. 35. С. 100–108.
10. Ефимов А.В. Новая супервторичная структура белков: α - α -уголок. *Молекулярная биология*. 1984. Т. 18. С. 1524–1537.

11. Efimov A.V. Structure of coiled β - β -hairpins and β - β -corners. *FEBS Lett.* 1991. V. 284. P. 288–292.
12. Бошкова Е.А., Бражников Е.В., Ефимов А.В. Взаимосвязь между структурой и аминокислотной последовательностью сильно скрученных и изогнутых β -шпилек в глобулярных белках. *Молекулярная биология.* 2016. Т. 50. № 5.
13. Edwards M.S., Sternberg J.E., Thornton J.M. Structural and sequence patterns in the loops of $\beta\alpha\beta$ units. *Protein Engineering.* 1987. V. 1. P. 173–181.
14. Ефимов А.В. Стандартные структуры в белковых молекулах. I. α - β -Шпильки. *Молекулярная биология.* 1986. Т. 20 С. 329–339.
15. Ефимов А.В. Стандартные структуры в белковых молекулах. II. β - α -Шпильки. *Молекулярная биология.* 1986. Т. 20 С. 340–345.
16. Efimov A.V. Standard structures in proteins. *Progr. Biophys. Mol. Biol.* 1997. V. 60. P. 201–239.
17. Гордеев А.Б., Ефимов А.В. Новое структурное древо ($\alpha+\beta$)-белков, содержащих abCd-единицы. *Молекулярная биология.* 2009. Т. 43. С. 521–526.
18. Tatusova T.A., Madden T.L. Blast 2 sequences – a new tool for comparing protein and nucleotide sequences. *FEMS Microbiol. Lett.* 1999. V. 174 P. 247–250.
19. Sayle R., Milner-White J. RasMol: biomolecular graphics for all. *Trends Biochem. Sci.* 1995. V. 20. P. 374–376.
20. Koradi R., Billeter M., Wuthrich K. MOLMOL: a program for display and analysis of macromolecular structures. *J. Mol. Graph.* 1996. V. 14. P. 51–55.

Рукопись поступила в редакцию 06.06.2016.
Дата опубликования 18.07.2016.