



Структура и функция биополимеров

УДК 577.112+371.24

С. З. Мальченко, Н. А. Чащин

О НЕКОТОРЫХ ЗАКОНОМЕРНОСТЯХ В ОРГАНИЗАЦИИ ПРОСТРАНСТВЕННОЙ СТРУКТУРЫ БЕЛКОВ

В работе проанализированы основные трудности в предсказании структуры белков по гомологии последовательностей и произведен предварительный анализ информационных сайтов белков. Обнаружено, что как для гомо-, так и для гетерогенных повторяющихся блоков аминокислот информационно значимым является определенное множество остатков. В работе приведен и обсуждается ряд принципов и закономерностей в организации пространственной структуры белков.

Введение. В настоящее время проблема прямого предсказания пространственной структуры белков по первичной последовательности далека от решения, поэтому предпринимаются попытки опосредованного ее разрешения — через предсказание вторичной структуры (ВС) белков. Показано, что, имея предсказание ВС белка с точностью 75—85 %, удастся составить общее представление о его пространственной структуре, а при увеличении точности — получить довольно близкую к реальной пространственную модель белка. При использовании современных предсказательных методов, вероятно, достигнут некий максимальный предел точности предсказания ВС для основной массы белков, равный ~70 % [1], перешагнуть который пока не удастся. В предыдущей работе [2] нами рассмотрены основные недостатки статистических методов предсказания и высказано предположение о вероятном пути прогресса в увеличении точности предсказания ВС белков. Этот путь заключается в разработке логических методов предсказания и анализе «информационных сайтов» белков. Как пример одного из логических методов анализа можно привести метод предсказания ВС по гомологии первичных последовательностей белков. Использование этого метода для высокогомологических белков позволяет достигнуть точности предсказания 95 %. В то же время существует ряд работ [3—5], в которых показано, что гомологичные участки белков могут принимать различную конформацию ВС. В данной работе предпринята попытка анализа сложившейся ситуации в методах предсказания по гомологии и описаны результаты поиска возможных информационных сайтов белков.

Материалы и методы. В работе использован банк данных, содержащий информацию о ВС 108 белков (20 000 аминокислот) с рентгеноструктурным разрешением менее 0,2 нм. Для выравнивания последовательностей и определения гомологичных районов в белках использовали программу MULTALIN [6]; в случае анализа банка данных ВС для поиска гомологичных участков и информационных фрагментов белков — оригинальную программу BankSearch. Исследуемые в работе последовательности белков получены из Брухевенского банка данных пространственных структур белков. Аминокислоты обозначали стандартными однобуквенными символами IUPAC-номенклатуры; конфор-

мации вторичной структуры α -спираль, β -складка и нерегулярная структура — соответственно Н(h), Е(e) и С(c).

Результаты и обсуждение. Первоначальной задачей был анализ встречаемости в ВС гомогенных повторяющихся последовательностей следующего типа: X_n , где X — одна из 20 природных аминокислот; n — степень повторяемости. Степень повторяемости, обнаруженная в используемой базе данных (БД), составила величины от 2 до 4, т. е. определены последовательности типа XX, XXX и XXXX (дипептиды, трипептиды и тетрапептиды), сведения о встречаемости которых во ВС представлены в табл. 1. Анализ данных таблицы выявляет ряд интересных фактов. Во-первых, при увеличении длины повторяющейся последовательности выделяются три алифатических остатка — G, A и V, обладающих высокой способностью к образованию нерегулярной конформации, α -спирали и β -складки соответственно. Во-вторых, имеются 13 остатков с полным набором свойств, необходимых для образования функциональных сайтов белков (гидрофобные и гидрофильные, отрицательно и положительно заряженные и т. д.) и с тенденцией находиться преимущественно в одной из трех конформаций ВС.

Далее, проведен анализ гетерогенных трипептидов типа YXY и YXY, где X, Y — два разных остатка из 20 природных аминокислот. Установление их встречаемости во ВС представляет определенный интерес, поскольку при объединении этих трипептидов образуется гетерогенная повторяющаяся последовательность типа (XY) $_n$. Для этой цели пользовались следующей стратегией — трипептид считали находящимся в какой-либо конформации, если все три остатка находились в этой конформации. При этом если трипептид с таким же аминокис-

Таблица 1

Данные о встречаемости последовательностей типа X_n во вторичной структуре белков

Пептид	Вторичная структура	Количество в БД	Информативность
Дипептид			
CC	ee, cc, ec	5 1 1 1	E*
MM	nh, cc, ee, ec	13 2 2 1	H*
HH	cc, hh	4 2	CH
WW	hh, hc, cc	5 1 1	HC
FF	hh, ee, hc, ec, cc	15 3 2 2 1	H*
PP	cc, hh, ee	31 1 1	C*
QQ	hh, ee, cc, ce, hc, ch, ec	14 7 3 2 1 1 1	CHE
Трипептид			
GGG	ccc, cce, cee	11 1 1	CE
AAA	hhh, ccc, hhc, eec	14 4 2	H*
III	eee, eec	1 1	EC
VVV	eee, cee, eec	6 2 1	EC
SSS	ccc, chh, eee, cee, eec, ecc, cce	9 3 2 1 1 1 1	C*
TTT	ccc, eec, eee, cee	2 2 1 1	CE
LLL	hhh, eee, cee, ecc	3 3 1 1	H*
PPP	ccc	1	C
RRR	ccc, eee	1 1	CE
NNN	ccc	1	C
DDD	ccc	4	C
EEE	hhh	2	H
KKK	hhh, ccc, cch	2 1 1	HC
YYY	ccc, cee, eec	1 1 1	CE
Тетрапептид			
GGGG	ccce	1	C
AAAA	hhhh	1	H
VVVV	eeee	2	E
SSSS	cccc, ceee, eeee	1 1 1	CE
IILL	eeee	1	E

Примечание. БД — база данных; Н(h), Е(e) и С(c) — α -спираль, β -складка и нерегулярная структура. * Структура имеет преимущественное значение, однако в этом же контексте могут встречаться и другие структуры.

лотным составом существовал и в другой конформации, исследуемый трипептид считался информационно незначимым. Данные об информационно значимых трипептидах представлены в табл. 2. При анализе данных обнаруживаются следующие информационные характеристики остатков: 1) остатки G, A и V являются самыми сильными (в смысле тенденции к образованию определенной конформации) в своих структурных группах триплетов; 2) выделяются информационно значимые остатки: для α -спирали — A, E, F и M, для β -складки — V и I, для нерегулярной структуры — G, P, D и N. Эти данные полностью совпадают с аналогичной информацией, полученной для гомогенных повторяющихся последовательностей. Таким образом, можно сделать вывод о том, что существует ряд остатков (см. табл. 1 и 2), несущих большую информационную нагрузку в детерминации ВС как для гомо-, так и для гетерогенных повторяющихся последовательностей остатков.

На сегодняшний день сильное возражение против методов предсказания по гомологии последовательностей содержится в работе [3], в которой показано, что идентичные пентапептиды могут находиться одновременно в разных конформациях ВС с вероятностью, превышающей случайное совпадение. Для объяснения результатов авторы предложили гипотезу высокой конформационной адаптивности коротких пептидов. В рамках этой гипотезы выдвигаются три предположения: 1) существуют идентичные пентапептиды, конформация которых полностью определяется окружающим аминокислотным контекстом; 2) существуют идентичные пентапептиды, конформация которых не зависит от контекста; 3) идентичность пентапептидов не отражает сходства ВС или эволюционного родства сравниваемых белков. В пользу своих выводов авторы приводят 25 примеров идентичных пентапептидов, 13 из которых подтверждают предположение 1 и 12 — предположение 2. Используя программу BankSearch, в настоящей работе сделаны поиск и анализ идентичных пентапептидов в БД с количеством аминокислот, в 2 раза превышающем таковое в работе [3]. Обнаружены 143 идентичных пентапептида из негомологичных белков, конформация которых с равной вероятностью была схожей и отличающейся. Также выделены 162 пентапептида в гомологичных белках, конформация которых полностью идентична. Исследование этих примеров показало, что как в случае несовпадения конформаций, так и в случае их совпадения определяющим фактором является наличие в ближайшем окружении (± 5 остатков) пентапептидов информационно значимых аминокислот (см. выше). «Конкурентная борьба» между информационно значимыми аминокислотами, детерминирующими определенную конформацию, в основном определяет локальную ВС участка белка.

Все методы предсказания ВС на основе гомологии последовательностей базируются на простом принципе: короткие пептиды с высокой степенью гомологии должны иметь схожую вторичную структуру. Однако до сих пор была сделана только одна попытка определить, насколько короткими должны быть пептиды и как высока должна быть степень гомологии [5]. Обобщая данные, полученные авторами этой работы, можно заключить следующее: для фрагментов белков длиной от 20 до 262 остатков недостаточна степень гомологии 25 % для того, чтобы они имели идентичную конформацию. Нами проведен анализ фрагментов белков длиной от 15 до 200 остатков со степенью гомологии выше 25 % для определения точных границ (в процентах) соответствия гомологии последовательностей и сходства их ВС. На основе полученных данных определены следующие границы: для фрагментов, гомологичных не более чем на 30 %, — несовпадающая ВС; с гомологией от 30 до 47 % — частично совпадающая ВС; с гомологией выше 47 % — практически полное совпадение ВС (рис. 1). Также были определены белки, названные нами самогомологичными белками. Они состоят из нескольких высокогомологичных блоков, имеющих идентичную ВС (см. рис. 1). При объединении таких блоков образуется симметричная структура белка (как в первичной последовательности, так

Таблица 2

Данные о встречаемости трипептидов типа ХУХ и УХУ, образующих регулярную последовательность типа (ХУ)_n

Структура (трипептид)	Количество в БД	1	2	Информативность
α-спираль				
AMA ACA AEA AWA AYA	2 1 7 2 4	16	5	+
LAL LQL	4 1	5	2	
ICI IDI IPI	1 1 1	3	3	
VMY VDV	1 1	2	2	
TMT	2	2	1	
MYM MSM MQM MRM	1 2 1 1	5	4	?
CNC	1	1	1	
DRD	1	1	1	
ELE EQE EHE EYE	3 3 1 1	8	4	+
QLQ QEQ QRQ QFQ	2 1 1 1	5	4	?
KQK KYK	1 1	2	1	
RTR RKR RHR RNR	1 2 1 2	6	4	?
HAH HTH	1 1	2	2	
WGW WRW	1 1	2	2	
FAF FLE FMF FKF	3 1 1 2	7	4	+
YDY	4	4	1	
PFP	1	1	1	
NFN	1	1	1	
β-складка				
LHL	1	1	1	
IAI IMI IWI	1 1 1 1	4	3	?
VIV VQV VVV	3 2 1	6	3	+
TDT TET	1 1	2	2	
MTM MKM	1 1	2	2	
CWC	1	1	1	
DHD	1	1	1	
KIK KWK KPK	2 1 1	4	3	?
RGR RSR	1 1	2	2	
HVII HEH	1 1	2	2	
FRE FHE FYF	1 1 1	3	3	?
YTY YCY YFY	1 1 1	3	3	?
NYN	1	1	1	
Нерегулярная				
GLG GIG GTG GMG GQG GKG	5 5 4 2 3 5	37	10	+
GRG GMG GFG GPG	5 2 3 3			
LCL	1	1	1	
SES SQS SRS SWS SPS	2 3 1 1 2	8	5	?
TET TWT TPT	1 1 4	6	3	
MGM MCM	1 1	2	2	
CYC	1	1	1	
DGD DSD DTD DCD DED	9 1 2 1 2	25	10	+
DQD DWD DYD DPD DND	3 1 1 3 2			
ESE ENE	3 1	4	2	
QSR QPQ	1 2	3	2	
KSK KHK KFK KNK	3 3 4 1	11	4	+
RIR RER RPR	1 1 1	3	3	
HGH HEH HKH	1 1 1	3	3	
WPW	1	1	1	
FDL FPL	4 1	5	2	
YVY YSY YEY	2 2 1	5	3	
PGP PAP PLP PIP PSP PTP	1 1 8 2 2 2	25	13	+
PMP PDP PQP PKP PKP PHP	1 2 2 2 1 1			
PWP	1			
PYP	1			
NGN NSN NCN NDN NEN NQN	1 3 1 1 3 1	11	7	+
NKN	1			

Примечание. В колонках 1 и 2 указаны общее количество трипептидов данного типа в БД и количество типов трипептидов в БД для данной аминокислоты соответственно. В колонке «Информативность» знаком «+» обозначается ярко выраженная тенденция к образованию определенной структуры; знаком «?» — возможное влияние остатков этого типа на детерминацию вторичной структуры. Для оценки информативности, кроме общего количества трипептидов в БД, важное значение имеет количество типов трипептидов.

влиять на образование или разрушение структуры, а служат, по-видимому, для тонкой корректировки расположения элементов ВС. В наблюдаемой нами БД делеции до пяти остатков приводили к сдвигу элементов ВС на столько позиций, на сколько произошла делеция.

5. Принцип понижения детерминированности структуры (рис. 3). Как показано в работе [5], степень гомологии ВС от 61 % и выше предполагает наличие схожих пространственных структур. В то же время

```

      X Y Z x y z
1: KSEEEIANGFRIFDKNIIDCFIDILFICEIIGIIE 93-126
   *****
2: GEEEEIAAAAADDNDNDDFDDDEEEIAAAAAGG
   *****
3: SEEEEEIAAAAADDNDNDDFDDDEEEIAAAAAGG
   *****
4: SNNNNNNNNNNNNSSSSSSSSNNNNNNNNNNSS

```

Рис. 2. Возможная эволюция Са-связывающего сайта. В колонках 1 и 4 представлены первичная последовательность Са-связывающего сайта и его вторичная структура соответственно [1]; в колонках 2 и 3 — возможные предковые последовательности Са-связывающего сайта, состоящие из повторяющихся блоков информационно значимых аминокислот. Символами X, Y, Z, x, y и z обозначены консервативные остатки в Са-связывающем сайте; символами «*» и «:» — соответственно идентичные и функционально схожие аминокислоты. Большие символы аминокислот отвечают информационно значимым аминокислотам, маленькие — неинформационным остаткам

Последовательность ДНК	Степень гомологии
1 GCGTTTCCGAGCAGCCGCGCGCGTTATTAGCGCG	25%
2 TTCCGCGCGTGTATAGCGCGCGCGCGCGCGTATT	
<i>Безклеточная последовательность</i>	
1 AFASSAGGJFSA	50%
2 FLASLAAGLGSF	
<i>Вторичная структура</i>	
1 NNNNINSSSSSS	91%
2 NNNNINSSSSSS	

Рис. 3. Понижение детерминированности структур от пространственной структуры к ДНК. Цифрами 1 и 2 обозначены две последовательности, имеющие схожую пространственную структуру. В колонке справа приведена степень гомологии двух последовательностей на различных уровнях организации структур. Символом «*» отмечена идентичность элементов вторичной структуры, аминокислот и нуклеотидов

нами показано, что для гомологии ВС необходимо совпадение как минимум 47 % аминокислотной последовательности. При этом, как показано на примере (рис. 3), достаточно совпадения 25 % нуклеотидной последовательности соответствующих генов.

6. Принцип избыточности пространственной структуры [7]. Ряд белков имеет 4-симметричную вторичную структуру, соответствующую четырем функциональным сайтам (лектин, Са-связывающий белок, гамма-макросталлин, рис. 1, A, B, D). Однако в исследуемой БД обнаружены белки с аналогичными им функциями, но имеющие 2-симметричную структуру, т. е. 4-симметричная структура является избыточной. Кроме этого, в БД содержатся восемь протеиназ длиной от 170 до 280 остатков, все они имеют один и тот же функциональный мотив длиной 120 остатков, достаточный для выполнения протеиназной функции. Таким образом, увеличение длины белка в отношении пространственной структуры является избыточным.

Summary. Paper present major difficulties during the prediction of protein structures on the base of sequences homology and the previous analysis of protein informative sites. It is determined that some set of residues is informatively important for homo- as for heterogenic repeated amino acids blocks. Several principles and regularities for the spatial protein structures organization are discussed.

СПИСОК ЛИТЕРАТУРЫ

- Garnier J., Levin J. M. The protein structure code: what is its present status? // CABIOS.— 1991.— 7, N 2.— P. 133—142.
- Мальченко С. З., Чащин Н. А. Предсказание вторичной структуры белков // Биополимеры и клетка.— 1992.— 8, № 5.— С. 21—30.

3. Kabsch W., Sander C. On the use of sequence homologies to predict protein structure: Identical pentapeptides can have completely different conformations // Proc. Nat. Acad. Sci. USA.— 1984. — 81, N 4.— P. 1075—1078.
4. Shaw G. S., Hodges S., Sykes B. D. Calcium-induced peptide association to form an intact protein domain: NMR structural evidence // Science.— 1990.— 249, N 4966.— P. 280—283.
5. Sternberg M. J. E., Islam S. A. Local protein sequence similarity does not imply a structural relationship // Prot. Eng.— 1990.— 4, N 2.— P. 125—131.
6. Corpet F. Multiple sequence alignment with hierarchical clustering // Nucl. Acids Res.— 1988.— 16, N 22.— P. 10881—10890.
7. Троцкий Г. В. Дефектные белки: постсинтетическая модификация.— Киев: Наук. думка, 1991.— 232 с.

Ин-т молекуляр. биологии и генетики АН Украины, Киев

Получено 31.08.92

УДК 591.175.4

Э. Б. Бабийчук, А. М. Филенко,
Д. Стемпковски, И. Конколь, В. М. Данилова

ВЛИЯНИЕ МОДИФИКАЦИИ РЕГУЛЯТОРНЫХ ЛЕГКИХ ЦЕПЕЙ МИОЗИНА НА АГРЕГАЦИОННЫЕ ХАРАКТЕРИСТИКИ ЕГО АКТИВНЫХ ФРАГМЕНТОВ

Проведено сравнительное исследование агрегационных характеристик (при 42°C) активных фрагментов миозина скелетных мышц кролика — химотриптического субфрагмента-1 (X-C1) без регуляторной легкой цепи (РЛЦ), Mg-напаинового субфрагмента-1 (Mg-C1), включающего РЛЦ, и тяжелого меромиозина (ТММ), содержащего дефосфорилированные или фосфорилированные РЛЦ. Показано, что помимо гидрофобных взаимодействий большое влияние на связывание РЛЦ с шарнирным сегментом головки миозина оказывают двухвалентные катионы. Полученные результаты позволяют предположить, что в присутствии двухвалентных катионов РЛЦ оказывает стабилизирующее воздействие на лабильные структуры головки миозина — «существенную» легкую цепь — и через нее на фрагмент с молекулярной массой 50 000. Действие фосфорилирования РЛЦ противоположно таковому двухвалентных катионов: оно приводит к некоторой дестабилизации структуры головки, что свидетельствует об ослаблении связывания фосфорилированной РЛЦ с головками и может служить косвенным указанием на «подстрочную» роль фосфорилирования РЛЦ при функционировании сократительной системы скелетных мышц.

Введение. Главную роль в системе сокращения мышц играет миозин. Его молекула состоит из нескольких структурных фрагментов, различающихся функционально. Эти фрагменты взаимодействуют между собой в процессе сократительного акта [1, 2]. Основная нагрузка при этом приходится на головки, представляющие собой глобулярные структуры, образованные N-концевыми участками двух тяжелых цепей миозина. На каждой головке размещены АТФазный и актинсвязывающий центры миозина. Помимо тяжелых цепей головки содержат также по две нековалентно связанные легкие цепи: «существенные» (СЛЦ), играющие важную роль в стабилизации структуры головок [3], и «регуляторные» (РЛЦ), которые в гладких мышцах и в мышцах беспозвоночных принимают непосредственное участие в регуляции сокращения. В таких сократительных системах регуляция осуществляется либо через Ca²⁺-зависимое фосфорилирование — дефосфорилирование РЛЦ, либо при непосредственном связывании Ca²⁺ с РЛЦ [4, 5]. РЛЦ миозина скелетных мышц также способны связывать Ca²⁺ и фосфорилироваться в процессе сокращения, однако непосредственного участия в регуляции запуска сокращения мышц они не принимают и, по-видимому, выполняют иные, неясные до сих пор функции.

Удобным объектом для изучения роли легких цепей миозина являются препараты изолированных головок миозина, известные под названием субфрагмент-1 (С1). Они несколько различаются в зависимости от способа получения [6—8]. Так, химотриптический С1 (X-C1) — ре-

© Э. Б. Бабийчук, А. М. Филенко, Д. Стемпковски, И. Конколь, В. М. Данилова, 1993