МОДЕЛИРОВАНИЕ ОБРАЗОВАНИЯ α-СПИРАЛЕЙ И β-ШПИЛЕК В ВОДОРАСТВОРИМЫХ БЕЛКАХ МЕТОДОМ КОДОВОЙ ФИЗИКИ

© *Б. В. Шестопалов*

Институт цитологии РАН, Санкт-Петербург;
электронный адрес: shest@mail.cytspb.rssi.ru

Один из возможных подходов для полного и окончательного решения проблемы определения трехмерной структуры белка по аминокислотной последовательности — моделирование образования трехмерной структуры белка. Для выполнения этой задачи автор предполагает использовать развиваемый им методом кодовой физики. В статье описано начало выполнения этого плана — моделирование образования α-спиралей и β-шпилек в водорастворимых белках. Результаты моделирования сопоставлены с данными эксперимента для 14 белков с числом аминокислот не более 50 и поэтому небольшим числом α-спиралей и β-нитей (чтобы учесть ограничения моделирования) и предсказаниями вторичной структуры по лучшим на сегодня методам предсказания вторичной структуры белка — PSIpred, PORTER и PROFsec. Вторичная структура белков, полученная после моделирования образования α-спиралей и β-шпилек методом кодовой физики, полностью согласуется с экспериментом, в то время как вторичная структура, предсказанная методами PSIpred, PORTER и PROFsec, содержит существенные отличия от данных эксперимента.

Ключевые слова: физика, кодирование, моделирование, трехмерная структура белка, α-спираль, β-структура.

При решении теоретической проблемы определения трехмерной структуры белка по аминокислотной последовательности наибольшее распространение получили сравнительные и подобные им эмпирические методы, основанные на прямом применении известных из эксперимента соответствий между аминокислотной последовательностью и трехмерной структурой, для всего белка или его фрагментов — от весьма коротких, в несколько аминокислот, до супервторичных сверток, при этом почти полностью вытеснив методы, основанные на использовании классических физических параметров (Moult, 2005). Эти эмпирические методы используют подобие аминокислотной последовательности данного белка, полной или ее фрагментов, аминокислотным последовательностям белков с известной трехмерной структурой и тем менее успешны, чем меньше это подобие как по длине сравниваемых участков, так и по доле идентичных или похожих аминокислот в них, в пределе уменьшаясь до неприемлемого уровня, когда правильное предсказание трехмерной структуры может быть получено только случайно. Очевидно, что эффективность таких методов связана со способностью используемых в них методов выравнивания аминокислотных последовательностей производить выравнивания, соответствующие выравниваниям трехмерных структур. Как только объем информации о трехмерных структурах белков станет достаточным для того, чтобы для любого белка с неизвестной трехмерной структурой можно будет найти трехмерные аналоги, выявляемые методами выравнивания аминокислотных последовательностей, эти методы станут не нужны.

Аналогично положение и в области решения частной задачи — определения вторичной структуры белка по аминокислотной последовательности. Наилучшие результаты, согласно данным сравнения методов сервера EVA (Eyrich et al., 2001; http://cubic.bioc.columbia.edu/eva/index.html), можно получить, используя подходы, основанные на самообучении предсказательной машины методом нейронных сетей на выборке белков с известной вторичной структурой, дополненным построением выравниваний аминокислотных последовательностей, подобных данной (Rost, Sander, 1993; McGuffin et al., 2000; Rost et al., 2004; Pollastri, McLysaght, 2005). Чем больше белков в выборке для обучения и чем больше последовательностей в группе выравнивания, тем выше точность предсказания. В настоящее время эти методы в среднем дают почти 80%-ную точность предсказания вторичной структуры, представляемой по модели трех состояний: спираль, β-структура и клубок. При этом точность предсказания может значительно снизиться, если не удается найти достаточно аминокислотных последовательностей, подобных данной.

Представленные выше наиболее успешные методы предсказания полной трехмерной и вторичной структуры белка не способны в принципе моделировать сам процесс образования трехмерной структуры белка. Решение именно этой задачи привело бы к полному и окончательному решению теоретической проблемы определения трехмерной структуры белка по аминокислотной последовательности и породило бы все возможные приложения теории как к фундаментальным, так и к прикладным исследованиям.

Для решения задачи моделирования образования трехмерной структуры белка автор предполагает развивать кодовые подходы, предложенные им ранее для предсказания вторичной структуры водорастворимых белков (Шестопалов, 1990; Shestopalov, 2003a, 2003b). Вторичная структура белка делится на 3 вида: α-спираль, β-нить и любой иной вид (клубок). α-Спирали могут образовывать суперспиральные структуры, а β-нити, как правило, входят в состав β-слоев, простейший вид которых — β-шпилька; α-спирали и β-нити могут образовывать разнообразные супервторичные структуры. По последним данным, α-спирали и β-нити доминируют во вторичной структуре белка — около 80 % аминокислот в среднем входят в их состав (Majumdar et al., 2005), поэтому естественно предположить, что точное предсказание этих элементов в результате успешного моделирования образования α-спиралей и β-структур обеспечит успешный старт моделирования всей трехмерной структуры белка.

В 1990 г. автор ввел кодируемые парами крайних аминокислот минимальные участки α-спирали, [i, i + 4], β-нити, [i, i + 2], клубка, [i, i + 1], или структуроны; их перекрывание дает участки любой длины (Шестопалов, 1990). В 2003 г. автор ввел кодовые числа кодонов (структуронов) и создал кодовую модель, на основе которой был создан метод предсказания вторичной структуры, суть которого состоит в поиске выборки неперекрывающихся кодонов различных структур с максимумом суммы кодовых чисел (Shestopalov, 2003a, 2003b). Метод предсказания вторичной структуры белка, созданный на основе кодовой модели (Шестопалов, 1990; Shestopalov, 2003a, 2003b), оказался лучшим при предсказании вторичной структуры трудных белков в 5-м всемирном эксперименте по критической оценке методов предсказания структуры белка **(**(Aloy et al., 2003) — http://www.russell.embl.de/casp5/SS/ss_results.txt**)**.

В настоящей статье описано дальнейшее развитие кодовой модели с целью моделирования процесса образования простейших элементов α-спирально-β-структурной архитектуры водорастворимых белков — α-спиралей и β-шпилек.

Метод

Для моделирования образования α-спиралей и β-шпилек вводятся новые объекты — пре-α-спирали и пре-β-нити. Кодовые числа получены по-новому: конформационое состояние пар аминокислот xy берется без разделения на x и y, добавлена нормировка для выравнивания встречаемостей пар xy (см. кодовые числа на странице автора (http://www.cytspb.rssi.ru/persons/shestopalov/code_table_999_fkkxy.pdf), детали будут опубликованы отдельно, поскольку кодовые числа имеют самостоятельную прикладную ценность).

Построение пре-структур (возможных α-спиралей и β-нитей)

Случай α-спирали: 1) поиск протеронов (от греч. proteros — первичный) — структуронов α-спирали, [i, i + 4], кодовые числа которых не меньше любого из кодовых чисел заключенных в них структуронов клубка ([i, i +1], [i + 1, i + 2], [i + 2, i + 3], [i +3, i +4]); N- и C-концевые структуроны не используются; 2) протерон расширяется, становясь ауксероном (от греч. auksano — расширяюсь), если при добавлении по одной аминокислоте с любой стороны (например, i –1) разность кодовых чисел кодонов добавляемых при этом структурона α-спирали, [i –1, i +3], и структурона клубка, [i –1, i], не меньше нуля; нерасширяемые протероны и ауксероны — тупиковые ауксероны; разность суммы кодовых чисел всех структуронов α-спирали и суммы кодовых чисел всех структуронов клубка ауксерона — кодовое число ауксерона; 3) перекрывающиеся тупиковые ауксероны объединяются в сюнерон (от греч. syn — вместе), если кодовое число объединения (вычисляется как для ауксерона) не меньше каждого из кодовых чисел объединяемых; испытываются все возможные объединения, начиная с тупиковых ауксеронов, затем с добавлением сюнеронов; все необъединяемые ауксероны и сюнероны — тупиковые сюнероны; примыкающие друг к другу сюнероны объединяются; 4) тупиковые сюнероны с неотрицательным кодовым числом — пре-α-спирали.

Случай β-шпильки: 1) построение пре-структур для β-шпилек сводится к построению пре-структур для β-нитей, пре-β-нитей, которое проводится с использованием структуронов β-нитей аналогично построению пре-α-спиралей; 2) пре-β-нити, не перекрываемые пре-α-спиралями, — про-пре-β-нити.

Алгоритм образования α-спиралей и β-шпилек

1. Все пре-α-спирали становятся α-спиралями; в дальнейшем из этих α-спиралей могут образовываться α-спиральные комплексы, но этот процесс здесь не моделируется.

2. Для образования β-шпильки необходима про-пре-β-нить; если она не одна, то начинаем с имеющей наибольшее кодовое число; β-шпилька образуется с добавлением ближайшего по цепи участка, если сумма кодовых чисел β-нитей, из которых составлена β-шпилька, минус сопутствующее уменьшение кодового числа α-спирали, если добавляемый участок — часть α-спирали, не отрицательная и как можно больше (можно удлинять и укорачивать про-пре-β-нить вплоть до трех аминокислот).

3. Если после п. 2 осталась всего одна пре-β-нить вне β-шпильки, то она присоединяется к β-шпильке с образованием новой β-шпильки или без этого: а) если она про-пре-β-нить, то при условии, что она может сблизиться с β-шпилькой в результате обычного изгибания белкового остова или образования суперспиральной структуры (это условие до решения задачи моделирования образования α-спиральных комплексов выполнимо однозначно только для белков с числом α-спиралей не более одной); б) если же она не про-пре-β-нить, то еще и при условии, что увеличение кодового числа в связи с ее добавлением больше, чем уменьшение кодового числа при сопутствующем разрушении α-спирали.

4. Если после п. 2 осталось более одной пре-β-нити вне β-шпильки и они не могут примкнуть к ее β-нитям по правилам п. 3, то рассматривается образование ими отдельной β-шпильки по правилам п. 2.

Как видно из описания метода моделирования образования α-спиралей и β-шпилек в водорастворимых белках, цель данного моделирования состоит не в том, чтобы любым способом правильно предсказать эти элементы вторичной структуры, а в том, чтобы правильно

описать основные стадии этого процесса; только такой подход позволит получить надежный метод предсказания α-спиралей и β-шпилек.

Использованный метод основан на статистическом анализе встречаемости физических объектов — пар аминокислот — в полученных физическими методами структурах белков — также физических объектах. При создании метода также использовано основное физическое понятие — взаимодействие, так как именно пары аминокислот, расположенные на расстояниях наибольшего взаимодействия боковых групп, использованы при создании метода. Кодовые числа, используемые при моделировании образования физических структур, получены при анализе физических структур, проведенном методами математической статистики. Кроме того, применение метода к аминокислотным последовательностям дает предсказания структур, которые поверяются результатами физического эксперимента. Как видим, с физики начинается, с физикой созидается и физикой заканчивается. Поэтому мы решили назвать описанный метод методом кодовой физики белка. Однако, поскольку предсказывается структура белков, важнейших для клеток и их сообществ молекул, этот метод через приложение полученных с его помощью результатов имеет важное значение и для цитологии.

Результаты и обсуждение

Согласно предложенной модели процесса образования α-спиралей и β-шпилек, образованию β-шпилек может предшествовать стадия образования одних только α-спиралей (п. 1 раздела «Алгоритм»). Это напоминает о гипотезе «избыточных» спиралей (Лим, 1975), позднее подтвержденной экспериментально для нескольких белков (см., например: Chikenji et al., 2004). Первоочередное образование α-спиралей представляется естественным, так как число элементарных актов при образовании β-шпильки больше, чем при образовании α-спирали, и поэтому требует больше времени: образование двух β-нитей и изгиба между ними против образования одного витка α-спирали. Даже если в данном месте возможная β-нить, пре-β-нить, по кодовому числу более выгодна, чем возможная α-спираль, пре-α-спираль, она не удержится, так как до образования полноценной, т. е. стабилизированной водородными связями β-структуры, например β-шпильки, она нестабильна, и в итоге здесь вначале образуется α-спираль. В пользу такой интерпретации говорят, например, данные о механизме формирования β-шпильки в белке FBP28 (Petrovich, 2006), согласно которым только после образования водородных связей между остовами нитей происходит стабилизация этой структуры.

Моделирование образования α-спиралей и β-шпилек методом кодовой физики испытано на малых белках с числом аминокислот не более 50 и поэтому небольшим числом α-спиралей и β-нитей (чтобы учесть ограничения моделирования). Результаты испытания сопоставлены с данными эксперимента и предсказаниями вторичной структуры по нашей предыдущей кодовой модели (Shestopalov, 2003a, 2003b) и по методам PSIpred (McGuffin et al., 2000; http:// bioinf.cs.ucl.ac.uk/psipred), PORTER (Pollastri, McLysaght, 2005; http://distill.ucd.ie/distill/home.html) и PROFsec (Rost, Sander, 1993; Rost et al., 2004; http://www.predictprotein.org) — см. рис. 1, 2.

Тестовая выборка белков

Тестовая выборка белковых структур состоит из 14 уникальных структур, не использованных для получения кодовых чисел. В нее включены все структуры белков из списка уникальных белковых структур (Rost, 1999), не использованные при получении кодовых чисел, и структура белка 1uaoa. Все структуры не имеют цистеина (теория не учитывает S—S-мостиков), модифицированных аминокислот и лигандов, соответствуют мономерному состоянию и получены методом ядерного магнитного резонанса. Идентификация α-спиралей и β-структур, по данным эксперимента, для этих белков представлена двумя принципиально различными методами: методом водородных связей Кабша и Сандера (Kabsch, Sander, 1983; Berman et al., 2000; Vriend, 1990) и геометрическим методом PALSSE (Majumdar et al., 2005).

На рис. 1 представлены белки, содержащие β-шпильки.

В белке 1uaoa нет пре-структур, поэтому в нем не может быть ни α-спиралей, ни β-структур. Интерпретация данных эксперимента противоречива: β-шпилька с β-нитями не короче 3 аминокислот есть только по данным метода PALSSE.

В белке 1e01a нет пре-α-спиралей, поэтому в нем не может быть α-спиралей; максимальное кодовое число имеет про-пре-β-нить 17—21, поэтому именно она инициирует образование β-шпильки (с кодовым числом 357) с ближайшим участником 25—27, укорачиваясь при этом до участка 19—21, а затем — β-шпильку с отдаленным участком 7—11, восстанавливая длину, что дает прибыль кодового числа 614. Присоединение далее второй про-пре-β-нити 3—5 затруднено малой величиной изгиба между участками 3—5 и 7—11. Следовательно, получается β-зигзаг из двух подряд β-шпилек с β-нитью 17—21 в центре, что соответствует обеим интерпретациям данных эксперимента.

В белке 1fsd есть единственная пре-α-спираль 10—26, следовательно, вначале имеем α-спираль 10—26, которая затем разрушается до α-спирали 15—26 при образовании β-шпильки 4—7/10—13 (кодовое число 507) с сопутствующим укорочением про-пре-β-нити 2—7. Интерпретация данных эксперимента противоречива. Результаты моделирования соответствуют интерпретации данных эксперимента методом PALSSE.

В белке 1e0na единственная пре-α-спираль 13—24 дает α-спираль 13—24, которая затем разрушается из-за выигрыша в кодовом числе при образовании β-шпильки 3—7/13—17 (кодовое число 1784) с участием единственной про-пре-β-нити 3—7 и последующем образовании β-шпильки 14—17/22—25 после присоединения участка 22—25. Следовательно, получается β-зигзаг из двух подряд β-шпилек с β-нитью 13—17 в центре, что соответствует обеим интерпретациям данных эксперимента.

В белке 1hnr вначале пре-α-спираль 27—36 порождает α-спираль 27—36. Затем происходит образование β-структуры. Хотя про-пре-β-нить 43—45 имеет максимальное кодовое число, она не порождает β-шпильки, так как сопутствующее этому разрушение α-спирали 27—36 невыгодно, а присоединяется к одной из β-нитей β-шпильки 5—10/16—21 (111), порожденной второй про-пре-β-нитью 5—10. Интерпретация данных эксперимента противоречива. Результаты моделирования соответствуют интерпретации данных эксперимента методом PALSSE.

На рис. 2 представлены белки, не содержащие β-структур.

Рис. 1. Моделирование образования α-спиралей и β-шпилек в водорастворимых белках методом кодовой физики в сопоставлении с экспериментальными данными и предсказаниями вторичной структуры другими способами.

Белки, содержащие β-шпильки. В строке 1: идентификатор структуры белка и число аминокислот в белке по Банку белковых структур (Berman et al., 2000); строка 2 — аминокислотная последовательность; строка pa — пре-α-спирали (если есть, то в следующей строке их кодовые числа); строка pb — пре-β-нити (если есть, то в следующей строке их кодовые числа); строка nc — α-спирали и β-нити по результатам моделирования (в следующей строке даны их кодовые числа, если α-спирали и β-нити отличаются от пре-α-спиралей и пре-β-нитей); строка hd — идентификация α-спиралей и β-нитей по данным эксперимента методом водородных связей Кабша и Сандера (Kabsch, Sander, 1983) по суммарным данным идентификаций (Vriend, 1990; Berman et al., 2000); строка gm — идентификация α-спиралей и β-нитей по данным эксперимента геометрическим методом (Majumdar et al., 2005) — α-, 3.10- и π-спирали не различаются; строка ps — предсказание α-спиралей и β-нитей методом нейронных сетей PSIpred (McGuffin et al., 2000); pr — предсказание α-спиралей и β-нитей методом нейронных сетей PORTER (Pollastri, McLysaght, 2005) — α-, 3.10- и π-спирали и β-нити и β-мостики не различаются; pf — предсказание α-спиралей и β-нитей методом нейронных сетей PROFsec (Rost, Sander, 1993; Rost et al., 2004; http://www.predictprotein.org), метод применим для белков не короче 18 аминокислот; строка oc — предсказание α-спиралей и β-нитей по статистической модели аминокислотного кода вторичной структуры белка (Shestopalov, 2003a, 2003b); α-спирали короче 5 и β-нити короче 3 аминокислот в данных эксперимента и PSIpred-предсказаниях перенесены в клубок, так как при моделировании они отнесены к клубку; H, h — α-спираль, E, e — β-нить, «–» — нет данных; при сопоставлении теории с экспериментальными данными достаточно согласия теории с данными хотя бы одной из строк (hd или gm); участки, в теории отличающиеся по структуре от данных эксперимента, даны строчными буквами; для 1e01a в строках ps и pr с-символами выделены участки, ошибочно предсказанные как клубок; символ m обозначает среднюю β-нить в β-слое; у 1hnr в строке gm β-нить 6—19 состоит из β-нитей 6—12 и 13—19.

В белке 1bzg пре-α-спираль 13—32 дает α-спираль 13—32. Единственная про-пре-β-нить 7—9 не дает β-шпильки, так как не выполняются правила п. 2 раздела «Алгоритм...». Результат моделирования соответствует обеим интерпретациям данных эксперимента.

В белке 1ifya три пре-α-спирали (7—16, 20—33 и 37—45) дают в этих местах α-спирали. β-Шпильки не образуются, так как единственная про-пре-β-нить 2—4 не может породить β-шпильку из-за невыполнения правила п. 2 раздела «Алгоритм...». Результат моделирования соответствует обеим интерпретациям данных эксперимента.

В белках 1wbr, 1jjsa, 1j5ba, 2vpu, 1bcv, 112ya и 1psm есть только пре-α-спирали поэтому образуются только α-спирали. Результат моделирования соответствует обеим интерпретациям данных эксперимента для всех белков, где они есть (с учетом того, что в методе PALSSE α-спирали и 3.10-спирали не различаются, и с учетом примечания к данным для белка 1vpu в подписи к рис. 2). Для белка 1bcv есть интерпретация данных эксперимента только по методу водородных связей. Результаты моделирования не согласуются с ней, однако известно (Pegna, 1996), что при добавлении трифторэтанола в этом белке образуется α-спираль, и именно там, где она образуется согласно моделированию.

При сопоставлении с данными эксперимента предсказаний вторичной структуры тестовых белков, полученных другими методами, выясняется, что прежний кодовый метод дает 13 существенных ошибок (определение существенной ошибки см. в подписи к рис. 1), метод PROFsec дает 2 существенные ошибки (отсутствуют две β-нити в белке 1e0na, причем вместо одной из них пред-

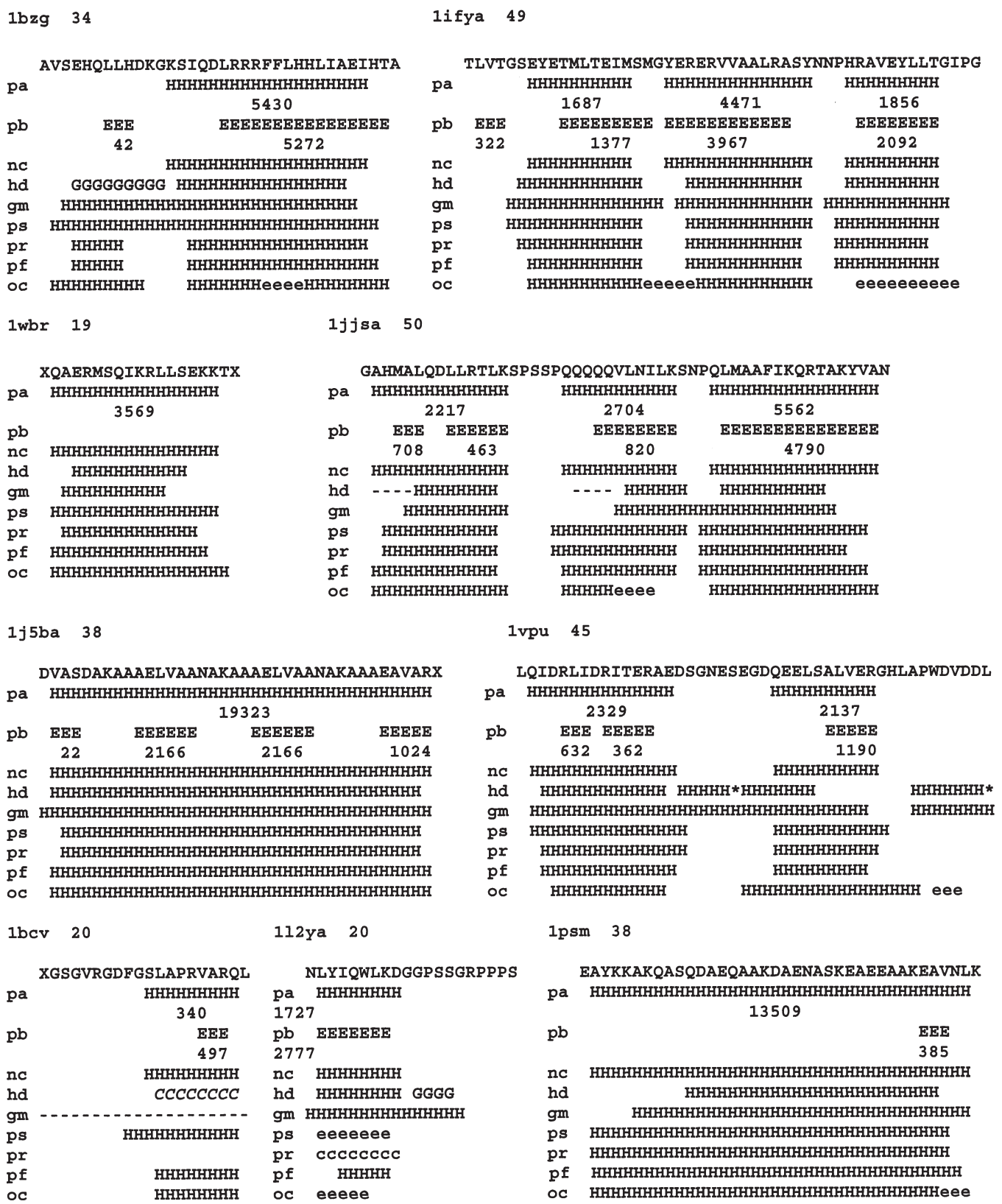

Рис. 2. Моделирование образования α-спиралей и β-шпилек в водорастворимых белках методом кодовой физики в сопоставлении с экспериментом и предсказаниями вторичной структуры другими способами.

Белки, не содержащие β-структур. См. основные пояснения в подписи к рис. 1; G — 3.10-спираль; в белке 1vpu α-спиралей 16—20* и 38—44* нет и в данных Банка белковых структур (Berman et al., 2000); для белка 1bcv *C*-символами обозначен участок клубка 12—19, переходящий в α-спираль при добавлении трифторэтанола (Pegna et al., 1996).

сказана α-спираль), метод PSIpred дает 3 существенные ошибки (отсутствует одна из β-нитей в белке 1e01a, в белке 1hnr вместо одной из β-нитей предсказана α-спираль, а в белке 112ya, напротив, вместо α-спирали предсказана β-нить), метод PORTER дает 4 существенные ошибки (отсутствует по одной β-нити в белках 1fsd и 1e0na, в белке 1hnr вместо β-нити предсказана α-спираль, а в белке 112ya α-спираль пропущена).

Из обсуждения результатов моделирования следует, что вторичная структура белков, полученная после моделирования образования α-спиралей и β-шпилек обновленным методом кодовой физики, полностью согласуется с экспериментом. Следовательно, можно сделать предварительный вывод о том, что решена задача моделирования образования α-спиралей и β-шпилек в малых водорастворимых белках (белках с небольшим числом α-спиралей и β-нитей) без S—S-связей, без модификаций аминокислот, без каких-либо лигандов, в мономерном состоянии. Для окончательного вывода необходимо расширить выборку тестовых белков. Также предварительно можно заключить, что результаты моделирования по методу кодовой физики можно использовать в качестве стартовых условий при моделировании образования полной трехмерной структуры белка.

Кроме того, очевидно, что наш метод предсказывает вторичную структуру исследованных тестовых малых белков существенно лучше, чем прежний кодовый метод (Shestopalov, 2003a, 2003b) и три самых лучших на сегодня метода предсказания вторичной структуры белка — PROFsec, PSIpred и PORTER. При этом наш метод существенно проще трех последних методов. Кроме того, наш метод в отличие от всех четырех сравниваемых с ним методов имеет физический смысл, моделируя процесс образования вторичной структуры, что можно проверить экспериментально. Следовательно, может оказаться, что для малых белков именно наш метод — лучший. Окончательный вывод можно будет сделать только после сравнения всех методов на расширенной тестовой выборке.

Полезное применение полученные результаты могут найти при поиске структурных подобий белков методом выравнивания различных объектов, полученных при моделировании образования α-спиралей и β-шпилек: последовательностей кодовых чисел, пре-структур, α-спиралей и β-шпилек.

При успешном моделировании образования α-спиралей и β-шпилек в малых белках были использованы новые объекты (кодоны, структуроны и пре-структуры), новые характеристики объектов (кодовые числа) и новые методы (получение кодовых чисел, построение и использование пре-структур). Такая совокупность новых объектов, характеристик и методов, позволившая успешно моделировать природное явление, может рассматриваться как основа или начало нового научного направления, которое мы назовем кодовой физикой. Предмет кодовой физики — описание систем, составленных из большого числа неодинаковых и не отделимых друг от друга элементов, т. е. систем, подобных белкам.

Описанный здесь нематричный код (кодирующее и кодируемое не отделимы друг от друга) из пар аминокислот и пре-структур с перекрыванием кодонов, возможно, существует в природе; поскольку код — это слова, а правила образования и использования слов — грамматика, представленное кодирование можно считать вкладом в языкознание белков.

ЭВМ-программы для получения кодовых чисел и построения пре-структур созданы совместно с Г. Р. Мавропуло-Столяренко и А. М. Лебедевым.

Благодарю акад. В. Т. Иванова за полезные советы и коллегу А. И. Петрова за полезные обсуждения.

Исследования выполнены при финансовой поддержке Российского фонда фундаментальных исследований (проект 04-04-48521).

Список литературы

Лим В. 1975. Структурные превращения белковой цепи при формировании нативной глобулы. Гипотеза «избыточных» спиралей. ДАН СССР. 222 (6) : 1467—1469.

Шестопалов Б. В. 1990. Предсказание вторичной структуры белка по методу дублетного кода. Молекуляр. биол. 24 (4) : 1117—1125. Engl. transl. — *Shestopalov B. V. 1990.* Prediction of protein secondary structure by doublet code method. Mol. Biol. (Moscow). 24 (4) : 900—907.

Aloy P., *Stark A.*, *Hadley C.*, *Russell R. B. 2003.* Predictions without templates: new folds, secondary structure, and contacts in CASP5. Proteins. 53 (Suppl. 6) : 436—456.

Berman H. M., *Westbrook J.*, *Feng Z.*, *Gilliland G.*, *Bhat T. N.*, *Weissig H.*, *Shindyalov I. N.*, *Bourne P. E. 2000.* The Protein Data Bank. Nucl. Acids Res. 28 : 235—242.

Chikenji G., *Fujitsuka Y.*, *Takada S. 2004.* Protein folding mechanisms and energy landscape of src SH3 domain studied by a structure prediction toolbox. Chem. Phys. 3007 : 157—162.

Eyrich V. A., *Marti-Renom M. A.*, *Przybylski D.*, *Madhusudhan M. S.*, *Fiser A.*, *Pazos F.*, *Valencia A.*, *Sali A.*, *Rost B. 2001.* EVA: continuous automatic evaluation of protein structure prediction servers. Bioinformatics. 17 : 1242—1243.

Kabsch W., *Sander C. 1983.* Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. Biopolymers. 22 : 2577—2637.

Majumdar I., *Krishna S. S.*, *Grishin N. V. 2005.* PALSSE: a program to delineate linear secondary structural elements from protein structures. BMC Bioinformatics. 6 : 202—225.

McGuffin L. J., *Bryson K.*, *Jones D. T. 2000.* The PSIPRED protein structure prediction server. Bioinformatics. 16 : 404—405.

Moult J. 2005. A decade of CASP: progress, bottlenecks and prognosis in protein structure prediction. Curr. Opin. Struct. Biol. 15 : 285—289.

Pegna M., *Molinari H.*, *Zetta L.*, *Melacini G.*, *Gibbons W. A.*, *Brown F.*, *Rowlands D.*, *Chan E.*, *Mascagni P. 1996.* The solution conformational features of two highly homologous antigenic peptides of foot-and-mouth disease virus serotype A, variant A and USA, correlate with their serological properties. J. Peptide Sci. 2 : 91—105.

Petrovich M., *Jonsson A. L.*, *Ferguson N.*, *Daggett V.*, *Fersht A. R. 2006.* Phi-analysis at the experimental limits: mechanism of beta-hairpin formation. J. Mol. Biol. 360 : 865—881.

Pollastri G., *McLysaght A. 2005.* Porter: a new, accurate server for protein secondary structure prediction. Bioinformatics. 21 : 1719—1720.

Rost B. 1999. Twilight zone of protein sequence alignments. Protein Engineering. 12 : 85—94.

Rost B., *Sander C. 1993.* Prediction of protein secondary structure at better than 70 % accuracy. J. Mol. Biol. 232 : 584—599.

Rost B., *Yachdav G.*, *Liu J. 2004.* The PredictProtein server. Nuc. Acids Res. 32 (WebServer issue): W321—326.

Shestopalov B. V. 2003a. Amino acid code of protein secondary structure. Цитология. 45 (7) : 702—706.

Shestopalov B. V. 2003b. Statistical model of amino acid code of protein secondary structure. Цитология. 45 (7) : 707—713.

Vriend G. 1990. WHAT IF: a molecular modeling and drug design program. J. Mol. Graph. 8 : 52—56.

Поступила 18 XII 2006

SIMULATION OF α-HELIX AND β-HAIRPIN FORMATION IN WATER-SOLUBLE PROTEINS BY THE CODE PHYSICS METHOD

B. V. Shestopalov

Institute of Cytology RAS, St. Petersburg;
e-mail: shest@mail.cytspb.rssi.ru

One of the possible ways for complete and final solution of the problem of determination of three-dimensional structure of proteins on amino acid sequence is simulation of protein three-dimensional structure formation. The use of the code physics method developed by the author has been suggested to fulfill this task. The simulation of α-helix and β-hairpin formation in water-soluble proteins as a start of realization of the plan is described here. The results of the simulation were compared with the experimental data for 14 proteins of no more than 50 amino acids and therefore with little number of α-helices and β-strands (to meet limits of simulation process) and with secondary structure predictions by the best to data methods of protein secondary structure prediction, PSIpred, PORTER and PROFsec. Secondary structure of the proteins, obtained as a result of the simulation of α-helix and β-hairpin formation using the code physics method, corresponded completely to experimental data while the secondary structure predicted by the PSIpred, PORTER and PROFsec methods differed from these data significantly.

K e y w o r d s: physics, encoding, simulation, protein three-dimensional structure, α-helix, β-structure.