

УДК 541.64:543.544

О ПРИМЕНИМОСТИ КОНЦЕПЦИИ КРИТИЧЕСКОЙ ХРОМАТОГРАФИИ К ЗАДАЧАМ ПРОТЕОМИКИ. ЭКСПЕРИМЕНТАЛЬНОЕ ИССЛЕДОВАНИЕ ЗАВИСИМОСТИ ВРЕМЕНИ УДЕРЖИВАНИЯ ПЕПТИДОВ ОТ ПОСЛЕДОВАТЕЛЬНОСТИ АМИНОКИСЛОТНЫХ ОСТАТКОВ В ЦЕПИ¹

© 2008 г. И. А. Тарасова*, А. В. Горшков**, В. В. Евреинов**, К. Адамс***,
Р. А. Зубарев***, М. В. Горшков*

*Институт энергетических проблем химической физики
Российской академии наук
119334 Москва, Ленинский пр., 38, корп. 2

**Институт химической физики им. Н.Н. Семенова
Российской академии наук
119991 Москва, ул. Косягина, 4

***Uppsala University, Biological and Medical Center, Laboratory
for Biological and Medical Mass Spectrometry
Box 583, S-75 123 Uppsala, Sweden

Поступила в редакцию 18.01.2007 г.
Принята в печать 27.08.2007 г.

Представлены экспериментальные данные по разделению синтетических и природных пептидов, рассмотренные в рамках предложенной авторами модели разделения, которая учитывает связанность аминокислотных остатков в цепь и коллективный характер их взаимодействия с поверхностью. Показано, что модель правильно предсказывает разделение пептидов одинакового аминокислотного состава, имеющих различия в чередовании остатков в цепи. Такие различия могут быть обусловлены перестановками аминокислотных остатков, наличием в цепи концевых групп, изомеров аминокислот или зеркальных последовательностей. Модель разделения применена для предсказания времен удерживания пептидов, полученных путем ферментативного гидролиза трипсином белков *E. coli* и белка альбумина BSA. Показано, что модель в целом правильно отражает совокупность экспериментальных данных по разделению таких пептидов и впервые позволяет связать последовательность в цепи с объемом удерживания.

ВВЕДЕНИЕ

Ранее авторами [1, 2] была сформулирована модель разделения биомакромолекул, основанная на представлении об адсорбции макромолекулы как адсорбции цепи случайных блужданий. Обработка в рамках этой модели существующих в литературе [3–5] экспериментальных данных по хроматографии синтетических и природных пеп-

тидов показала, что она непротиворечиво описывает наблюдаемые закономерности. В отличие от альтернативных моделей разделения [6, 7], широко используемых для предсказания времени удерживания пептидов, предложенная модель дает возможность связать объем или время удерживания не только с аминокислотным составом, но и с последовательностью аминокислотных остатков в цепи пептида – его “текстом”. В настоящей работе рассмотрено соответствие модели экспериментальным данным по разделению синтетических и природных пептидов, полученных на разных неподвижных фазах типа $-C_{18}$ и на разных экспериментальных установках.

¹ Работа выполнена при финансовой поддержке Отделения химии и наук о материалах РАН (программа “Создание эффективных методов химического анализа и исследования структуры веществ и материалов”), а также Российского фонда фундаментальных исследований (код проекта 06-04-49632) и фонда INTAS (гранты 04-83-2643 и Genomics 05-1000004-7759).

E-mail: evreinov@polymer.chph.ras.ru (Евреинов Виктор Викторович).

МОДЕЛЬ ХРОМАТОГРАФИЧЕСКОГО РАЗДЕЛЕНИЯ ПЕПТИДОВ BioLCCC

Система уравнений, описывающая хроматографическое разделение пептидов в условиях градиентного элюирования в бинарном растворителе с компонентами А и В и профилем $N_B(V)$ (V – объем прокачанного через колонку растворителя) подробно рассмотрена в работе [1]. Эти уравнения представляются в виде

$$\begin{aligned}\epsilon_{AB}(V) &= \epsilon_A + \ln(N_B(V)e^{(\epsilon_B - \epsilon_A)} - 1 + N_B(V)) \\ \epsilon_i(V) &= \alpha(X_i^0 - \epsilon_{AB}(V)) \\ K_d(V) &= \frac{1}{D} U^T \prod_i W(\epsilon_i(V)) P_0 \quad (1) \\ \int_0^{V_p - V_0} \frac{dV}{V_p K_d(V)} &= 1\end{aligned}$$

Кратко уравнения (1) могут быть интерпретированы следующим образом. Градиентный профиль (мольная доля N_B компонента В) определяет в каждой точке объема растворителя “элюирующую силу” смеси $\epsilon_{AB}(V)$, зависящую от силы отдельных компонентов ϵ_A и ϵ_B . В свою очередь “текст” последовательности аминокислот пепти-

да задает адсорбционный энергетический профиль макромолекулы $\{X_i^0\}$, где X_i^0 – стандартная энергия адсорбции i -го аминокислотного остатка. В каждой точке рассчитывается $\epsilon_i(V)$ – энергия адсорбции i -го аминокислотного остатка цепи с учетом десорбции компонентов растворителя. Масштабный фактор α учитывает небольшие различия в энергии адсорбции остатков на разных неподвижных фазах типа $-C_{18}$. По “тексту” последовательности аминокислотных остатков пептида строится цепь случайных блужданий с переходными матрицами $W(\epsilon_i(V))$ и вычисляется статистическая сумма, численно совпадающая с коэффициентом распределения $K_d(V)$ в поре размером D . Наконец, рассчитывается объем или время удерживания интегрированием функции от коэффициента распределения. Константы колонок V_0 и V_p соответствуют объему межчастичного пространства и объему пор.

Стандартные энергии адсорбции аминокислотных остатков $\{X_i^0\}$ – основные феноменологические параметры модели, приведены ниже. Они определены экспериментально для бинарного растворителя вода–ацетонитрил с добавками трифтормукусной кислоты на обращенной фазе типа $-C_{18}$ [1, 2].

K	H	R	N	G	S	Q	D	T	E
0.266	0.386	0.516	0.614	0.656	0.698	0.746	0.781	0.876	0.984
A	P	C	Y	V	M	I	L	F	W
1.143	1.143	1.296	1.686	1.751	1.822	2.156	2.298	2.319	2.436

Феноменологические параметры, описывающие энергию адсорбции компонентов растворителя, для воды и ацетонитрила приняты соответственно $\epsilon_A = 0$ и $\epsilon_B = 0.45$.

Система уравнений (1) подробно рассмотрена в работе [1]. Она осуществляет перенос идей метода критической хроматографии макромолекул (LCCC) на синтетические и природные биомакромолекулы. Модель разделения биомакромолекул, описываемая системой (1), будет называться в дальнейшем моделью BioLCCC (моделью критической хроматографии в приложении к биомакромолекулам).

Прежде чем перейти к рассмотрению приложения модели BioLCCC для объяснения экспериментальных закономерностей разделения пептидов, сделаем несколько общих замечаний относительно пределов ее применимости.

Основной целью хроматографии макромолекул является в общем случае их разделение по характеру чередования разных мономеров в цепи. Для макромолекул пептидов разделение должно осуществляться как по набору аминокислотных остатков (которые можно условно назвать “буквами” с различными энергиями взаимодействия с поверхностью), так и по их последовательности

(“тексту”, составленному из этих букв). Если возможность разделения макромолекул пептидов по набору их аминокислотных остатков очевидна, то разделение по “текстам” макромолекул, содержащих перестановки аминокислотных остатков – не вполне очевидный факт. Это особенно заметно, если подходить к проблеме “со стороны” низкомолекулярных соединений. Обычно такой подход сводится к вариантам аддитивного приближения, когда каждый остаток дает определенный “вклад” в статистическую сумму, в коэффициент распределения или во время удерживания [6, 7]. Ясно, что при таком подходе ни о какой связи объема удерживания с “текстом” последовательности цепи речи не идет.

Напротив, подход со стороны высокомолекулярных соединений органически содержит в себе зависимость удерживания от последовательности звеньев в цепи, что непосредственно следует из некоммутативности произведения переходных матриц W , описывающих блуждание звеньев по решетке в уравнении (1). Возможность разделения макромолекул гетерополимера по “текстам” связана с тем, что взаимодействие разных звеньев с поверхностью скоррелировано из-за сильных ковалентных связей, соединяющих мономеры в цепь. Для выбранного звена его взаимодействие с поверхностью зависит от удаленных по цепи соседей, и их перестановки могут заметно повлиять на это взаимодействие [8].

Существует два предельных случая корреляции звеньев в цепи гетерополимера. Гибкая макромолекула моделируется цепью случайных блужданий, в которой каждый остаток представляется звеном кубической решетки. Жесткая макромолекула моделируется стержнем, длина которого много больше ее диаметра, и который может изгибаться в результате тепловых флюктуаций. В модели BioLCCC макромолекула моделируется гибкой цепью случайных блужданий. Помимо математической простоты существуют еще две причины выбора такого приближения в качестве исходного.

Во-первых, опыт исследования олигомеров показывает, что даже в предельном случае очень коротких цепей модель случайных блужданий (или эквивалентная ей гауссова модель) дает хорошее количественное описание разделения макромолекул самой разной природы: от жестких

макромолекул эпоксидных смол, поликарбонатов и сложных полиэфиров до гибких цепей политетиленоксида или полипропиленоксида.

Во-вторых, корреляции в цепи случайных блужданий дают максимальную зависимость объема удерживания от “текста” последовательности, что позволяет оценить предельные возможности хроматографии для разделения пептидов с разными последовательностями.

Модель адсорбции флюктуирующего гетерогенного стержня также может быть использована в BioLCCC. Возможно, что такая модель наиболее адекватна макромолекулам пептидов. Однако ее математические аспекты, а главное – возможность соотнесения феноменологических параметров гибкости цепи экспериментально наблюдаемым величинам, представляются более сложными. Эти вопросы нами пока подробно не рассматривались, однако качественный вывод для модели жесткого гетерогенного стержня достаточно очевиден: перестановки остатков в середине цепи, по-видимому, мало существенны, а важно лишь то, какие остатки располагаются ближе к концам. Вообще говоря, цепь, моделируемая жестким стержнем, должна быть ближе к широко распространенным аддитивным моделям: основной вклад дают состояния, в которых макромолекула располагается на поверхности целиком, при этом каждый остаток взаимодействует с поверхностью.

Гибкая цепь также может располагаться на поверхности целиком. Это возможно в пределе очень коротких цепей или очень сильной энергии взаимодействия звеньев с поверхностью, близкой к энергии химической связи. По этой причине для гибких коротких цепей аддитивное приближение дает вполне разумную связь между составом цепи и объемом удерживания. Однако по мере роста длины цепи реализация состояния полностью адсорбированной цепи быстро становится затруднительной, так как требуется слишком большая энергия взаимодействия. В общем случае равновесное состояние адсорбированной цепи представляет собой чередующиеся участки взаимодействующих с поверхностью адсорбированных звеньев и связывающих их десорбированных петель и хвостов. Тем самым условия аддитивности нарушаются и появляется зависимость адсорбционных свойств от последовательности составляю-

щих цепь звеньев. Равновесное состояние макромолекулы реализуется в результате тонкого баланса энергии притяжения звеньев и потери энтропии цепи, связанной с ограничениями в их расположении вблизи поверхности по сравнению с неограниченным пространством.

Высказанные выше качественные представления справедливы и вблизи критической точки перехода макромолекулы из адсорбированного в десорбированное состояние: очевидно, что такому переходу предшествует образование петель и хвостов и, как следствие, нарушение принципа аддитивности. Адсорбционно-десорбционный переход всегда имеет место при градиентном элюировании макромолекул. В начальный момент макромолекулы адсорбированы на поверхности, затем, по мере увеличения элюирующей силы растворителя, они десорбируются и перемещаются вдоль колонки. Тем самым вблизи критической точки адсорбции осуществляются условия, в которых должна проявляться последовательность аминокислотных остатков в цепи.

Следует отметить, что в десорбированном состоянии при движении вдоль колонки макромолекулы разделяются по размерам по эксклюзионному механизму. В адсорбционном режиме макромолекулы также медленно мигрируют вдоль колонки и разделяются по разным признакам их первичной структуры. Модель BioLCCC учитывает все эти особенности реального разделения. Вместе с тем переход пептидов в адсорбированное состояние (в зависимости от конкретной первичной последовательности), по-видимому, осуществляется достаточно резко в узкой области состава растворителя. Поэтому основной вклад в разделение вносит область состава растворителя вблизи критической точки адсорбции. Предельно упрощая, можно сказать, что пептиды распределяются в градиенте состава раствора вблизи своих критических точек, зависящих от их конкретной первичной последовательности.

Помимо учета линейной памяти (связанности звеньев в цепь) модель BioLCCC основана на линейных соотношениях свободной энергии. Это позволяет записать в корреляционном виде зависимость эффективной энергии взаимодействия аминокислотного остатка от состава раствора и типа адсорбента в уравнении (1). Такой подход опирается на простейшие представления об

адсорбции компонентов растворителя, восходящие к изотерме Ленгмюра. В рамках линейного приближения можно “расцепить” адсорбционные свойства собственно аминокислотного остатка, подвижной и неподвижной фазы и тем самым предсказать изменение характера разделения при изменении формы градиентного профиля.

Изложенные выше качественные представления являются первым приближением к реальной картине, однако они дают зависимость объема удерживания от последовательности аминокислотных остатков в цепи.

Следует сделать еще одно замечание. В модели, описываемой уравнением (1), неявным образом предполагается короткодействующий характер взаимодействия аминокислотных остатков с поверхностью. Это может быть неверно для остатков, содержащих ионогенные группы $-NH_2$. Если количество кислоты в растворителе соответствует условию $pH 2.0$, то такие группы ионизируются, и их взаимодействие с поверхностью становится нелокальным. Пептиды, получающиеся в результате ферментативного гидролиза белков трипсином (для простоты будем называть их триптическими пептидами), имеют по крайней мере одну такую заряженную группу $-NH_3^+$ на N-конце. Если же в цепи есть несколько аминокислотных остатков с такими группами, то необходимо принять во внимание и возможность их электростатического взаимодействия друг с другом. Оценка влияния заряженных групп на картину разделения в рамках модели BioLCCC будет предметом дальнейших экспериментальных исследований.

Изменение формы энергетического профиля цепи при перестановке остатков или замене одного из них меняет объем удерживания. Модель BioLCCC позволяет для любой заданной последовательности оценить изменение объема удерживания, а следовательно, и возможность обнаружения перестановки или замены методом хроматографии. В общем случае перестановки могут приводить и к близким, неразличимым объемам удерживания. Поэтому решение обратной задачи (однозначного “чтения текста” неизвестного пептида из точно измеренного объема удерживания) в рамках хроматографии вряд ли возможно в общем случае. Необходимо привлечение “ортого-

нальных" хроматографии методов масс-спектрометрии.

Отметим, что масс-спектрометрия сама по себе также не дает ответа на вопрос о последовательности пептидов. Более того, точное измерение массы пептида отвечает лишь на вопрос о его аминокислотном составе. В пределах погрешности измерения, типичной для современных масс-спектрометров (10–100 м.д.), число возможных вариантов аминокислотных составов, соответствующих измеренной массе, может достигать сотен и тысяч (соответственно, возможных вариантов "текстов" последовательностей может быть на несколько порядков больше) [9].

Задача "чтения текста" пептида решается методом тандемной масс-спектрометрии, с использованием фрагментации макромолекул по пептидным связям и измерения масс фрагментов. Но даже и в этом случае однозначное установление последовательности затруднительно в силу ряда причин.

Во-первых, эффективность существующих методов фрагментации, как правило, невысока, не все пептидные связи удается разрушить и тем самым получить полный спектр фрагментации.

Во-вторых, ряд аминокислот имеет близкие либо совпадающие массы. Например, в паре глютамин (Q) и лизин (K) отличие в массах составляет всего 0.036, а массы лейцина (L) и изолейцина (I) просто одинаковы. Это делает затруднительным или невозможным их однозначную идентификацию. Даже для таких сложных методов измерения, каким является масс-спектрометрия ионного циклотронного резонанса [10], точность измерения масс в спектрах фрагментов относительно невысока.

Наконец, даже если не рассматривать задачу "чтения текста" неизвестного пептида, а решать типичную задачу протеомики – идентификацию пептида (а следовательно, и белка, которому он принадлежит), используя поиск по базам данных масс и последовательностей известных белков, то также можно столкнуться со значительными сложностями. Существующие поисковые системы, такие как Mascot или SEQUEST, могут содержать ошибочные данные о последовательности (по некоторым оценкам до 30%) или давать несколько возможных "текстов".

Несмотря на то, что решение обратной задачи с использованием BioLCCC также дает множество возможных "текстов" последовательностей, это множество не совпадает с множеством "текстов", получаемых из масс-спектрометрических данных. Следовательно, объединение метода BioLCCC и тандемной масс-спектрометрии в едином исследовательском комплексе значительно уменьшает число возможных "текстов"-кандидатов по сравнению с тем, что дают эти методы по отдельности.

Перед тем, как перейти к обсуждению экспериментальных данных, необходимо отметить некоторые особенности применяемых в настоящее время в исследованиях пептидов хроматографических систем. В отличие от стандартных аналитических хроматографических систем, использующих большие колонки и малые внеколоночные объемы, разделение пептидов обычно осуществляется в наносистемах. Последние имеют очень небольшой объем колонки, который много меньше, чем, например, объем вводимой пробы, что приводит к необходимости использования специфических способов ввода образца в колонку. Обычно исследуемый образец в объеме 5 мкл вводится в колонку объемом 300 нл в течение 30 мин, при этом неявно предполагается, что в течение времени ввода все пептиды сильно адсорбируются в начале колонки и не мигрируют вдоль нее. Помимо того, что для гидрофильных пептидов такое предположение не вполне правильно, при указанном способе ввода в начальный момент в колонке создаются "плохие" хроматографические условия, соответствующие высокой концентрации пептидов на поверхности и их взаимодействию друг с другом. Последнее обстоятельство никак не учитывается в модели BioLCCC. Таким образом, при интерпретации экспериментальных данных по разделению пептидов в наносистемах необходимо принимать во внимание и эти факторы.

С учетом сказанного выше рассмотрим совокупность экспериментальных данных по разделению синтетических пептидов с разными последовательностями, а также сложных смесей пептидов, получаемых при ферментативном гидролизе трипсином белков *E. coli* и BSA.

УСЛОВИЯ ЭКСПЕРИМЕНТА

Эксперименты по разделению и идентификации индивидуальных пептидов и их смесей проводили на двух хроматографических системах: макроскопической (стандартной) аналитической системе ВЭЖХ HP1100 ("Hewlett Packard/Agilent", США) и наносистеме ВЭЖХ Agilent 1100. Стандартная система ВЭЖХ HP1100 имела в составе УФ-детектор на основе диодной матрицы, позволяющий определять объем удерживания любых пептидов на нескольких длинах волн. Наносистема ВЭЖХ Agilent 1100 была объединена с гибридным масс-спектрометром LTQ-FT ("Thermo Electron", США). Масс-спектрометр был оборудован наноэлектроспреем (Proxeon Biosystems, Дания), позволяющим получать макромолекулярные ионы, а также использовать несколько методов их фрагментации, включая диссоциацию, активированную соударениями (ДАС) с молекулами буферного газа в ионной ловушке, и рекомбинантную диссоциацию при электронном захвате (ДЭЗ). Эти методы использовали для последующей идентификации последовательностей с помощью поисковой системы Mascot, доступной через Интернет (Matrix Science, США, www.matrix-science.com).

Разделение синтетических пептидов, содержащих изомерные остатки изолейцина (I) и лейцина (L), H-RAAAAIAAAAR-OH и H-RAAAA-LAAAAR-OH, осуществляли на стандартной системе HP1100 на колонках 4.0 × 150 мм, заполненных фазой Диасфер 110-C₁₈ (БиоХимМак СТ, Москва) с размером частиц 5 мкм и размером пор 110 Å. Компонент А мобильной фазы представлял собой воду с добавкой 0.1 об. % трифторуксусной кислоты (ТФУК), компонент В – ацетонитрил, также содержащий 0.1 об. % ТФУК. Разделение пептидов осуществляли при линейном по времени градиенте состава растворителя от начальной точки 2 об. % В до конечной точки 30 об. % В в течение 30 мин. Скорость подачи растворителя 0.5 мл/мин, объем вводимой пробы 20 мкл, концентрация образцов в пробе порядка 0.1 мг/мл.

Для разделения зеркальных последовательностей H-GALYIYLGGLDTADAEG-amide, H-GEADATDLGDGLYIYLAG-amide и таких же последовательностей с OH-группами на C-конце использовали стандартную хроматографическую

систему HP1100, аналитическую колонку Vydac 218TP54 размером 4.6 × 250 мм, заполненную фазой –C₁₈ с размером частиц 5 мкм и размером пор 300 Å. Объем вводимой пробы 20 мкл, концентрация образцов в пробе 0.1 мг/мл. Компонент А мобильной фазы представлял собой воду с добавкой 0.1 об. % ТФУК, компонент В – смесь 90.0 об. % ацетонитрила, 9.9 об. % воды и 0.1 об. % ТФУК, что соответствует pH 2.0. Пептиды с одинаковыми концевыми группами, содержащиеся в смеси в соотношении 2 : 1, были попарно проанализированы в нескольких хроматографических экспериментах при одних и тех же условиях разделения: линейный по времени градиент от начальной точки 2 об. % компонента В до 60 об. % В в течение 60 мин со скоростью 0.5 мл/мин.

Модельные пептиды получали путем автоматизированного твердофазного синтеза на синтезаторе "Intavis ResPep Synthesizer" ("Intavis Bioanalytical Instruments AG", Германия).

Разделение пептидов, образующихся в результате ферментативного гидролиза трипсином белков, выделенных из бактерии E. Coli, а также гидролизата белка альбумина (BSA) осуществляли на наносистеме Agilent 1100 с использованием в качестве детектора масс-спектрометра LTQ FT [11]. В наносистеме Agilent 1100 в качестве колонки использовали кварцевый капилляр размером 0.075 × 150 мм ("Proxeon Biosystems", Дания), заполненный фазой ReproSil Pur C18 AQ с размером частиц 3 мкм и размером пор 120 Å ("Dr. Maisch GmbH", Германия). Хроматограммы получали по регистрации полного тока ионов после ионизации электрораспылением. Объем вводимой пробы 8 мкл, количество вещества для анализа – порядка 500 фемтомолей. Компонент А мобильной фазы представлял собой воду, содержащую 0.5 об. % уксусной кислоты (pH 3.0). Компонент В – смесь 89.5 об. % ацетонитрила, 10 об. % воды и 0.5 об. % уксусной кислоты. В экспериментах использовали следующие условия хроматографического разделения: изократическое элюирование со скоростью 500 нл/мин в течение 20 мин смесью, содержащей 4 об. % компонента В, за которым следует линейный градиент от 4 до 45 об. % В в течение 90 мин со скоростью 200 нл/мин.

Идентификацию триптических пептидов белков бактерии E. coli, взятых из 20 фракций электрофореза, проводили на основе анализа спек-

Таблица 1. Последовательности триптических пептидов белков *E. coli* в образцах I и II

Последовательность	t_R эксп., МИН	t_R^* теор., МИН	Последовательность	t_R эксп., МИН	t_R^* теор., МИН
Образец I			Образец II		
TGVETR	31.910	31.875/30.144	REDVHR	27.520	30.875/29.132
FRDEGR	33.280	31.250/32.053	NDVSDSEK	30.850	32.125/29.571
GSGGGSSGGSIGGR	34.220	38.750/27.473	AREELRK	32.050	33.375/33.418
ENAANIK	34.510	34.375/40.959	SFSHQAGASSK	34.130	36.625/33.418
VIEPVKR	38.040	41.375/51.900	RVDQLK	34.940	33.750/36.714
VIEPVKR	38.040	41.375/51.900	GPAAIQK	35.240	35.750/37.484
VQLSDGR	39.770	35.250/44.903	AHVTAETPK	35.300	36.750/33.308
AGVEVDDR	39.870	38.625/45.666	PAVTAHPEQK	35.580	38.750/35.066
VLSGPQAQPAGDK	42.280	51.125/44.775	GSFEGR	37.040	32.375/33.418
GVNANHIIR	42.880	47.625/43.503	KSVAGFK	37.140	36.250/33.967
AQFLQK	44.050	39.875/59.534	NDEALNSQR	37.310	37.250/31.549
VHVTDYTNASR	45.170	45.125/37.014	QGIESATQK	37.880	36.125/30.670
AQYEDIAQK	45.230	44.250/36.887	AFAHYR	39.670	35.875/34.846
LSSPATLNSR	46.880	47.500/42.104	EVTQLR	40.440	36.125/35.176
SLLEGEKGSSGGGR	47.050	50.500/42.231	FSSSSGYGGGSSR	41.420	38.875/36.824
EHILLGR	47.160	53.125/47.320	FLEQQNK	41.600	36.875/40.121
EAFDTGVR	47.410	42.625/40.450	YSDEVVR	44.210	38.375/40.560
VAVDSEVTVR	49.840	55.125/49.737	STMQUELNSR	44.540	41.375/39.571
SLVNLGGSK	52.160	47.625/42.104	LIAVNR	44.880	42.750/46.385
GVQSILQR	53.690	49.125/41.722	AEAESLYQSK	46.320	46.375/42.648
DHFAILK	53.930	54.500/43.630	SQILDEAK	46.530	44.250/40.341
TNLIGAVAR	55.090	58.250/49.737	FSAASQPAAPVTK	48.260	54.500/51.330
LDVLDER	55.520	47.875/49.483	TLLEGEESR	48.540	47.125/52.868
GPLTTPVGGGIR	55.740	59.875/51.773	TLNDAVEVK	49.070	48.000/53.088
VVASPGQQQQFEIQASK	56.170	67.875/59.661	EDLLASGR	49.580	45.625/49.901
FLEQQNQVLQTK	56.490	64.375/66.913	GELFGAK	49.970	42.500/47.374
NLALNIESR	58.520	59.750/60.424	VQAQIQQGDEIR	50.080	51.625/48.692
SYYALAESVK	59.100	62.750/60.424	LANELS DAAENK	50.210	51.375/46.714
VSLVYQQMNEPPGNR	59.860	69.875/58.134	IVGDGIAIKPTGNK	52.210	59.125/55.066
GAFVSQVLPNSSAAK	60.960	73.250/58.007	VATVSLPR	52.550	48.375/47.264
VVEEAPAPGITPELR	61.800	77.375/68.058	YAIVANDVR	52.930	54.750/52.868
IVDLLTER	63.270	65.125/80.399	ATNLLYTR	53.480	53.625/49.791
GYTVSIFNR	64.770	59.750/73.147	EIADGLASAER	53.600	53.750/47.813
LINDAYDSEYFATK	66.130	75.375/63.732	ALEESNYELEGK	54.160	57.250/49.242
GNFDLEGLER	66.260	62.750/61.060	AEAQVIIEQANK	54.900	61.875/50.670
YDIPVVMDSAR	67.860	70.875/58.261	VQALEEEANNDLENK	55.980	58.625/53.088
DWTIEQITR	70.690	60.125/53.045	SEFITVAR	56.220	55.375/58.253
SNTGLVIDPYFSGTK	71.570	81.250/60.933	IEISELNR	56.710	49.875/58.582
DVLEAMQADSGIR	72.110	69.375/55.844	SVEEILGK	56.810	49.500/54.736
AAPSYEELSNSQELLETGK	74.480	84.250/77.473	FGSELLAK	57.150	53.375/59.901
AVEIGSFLLGR	78.470	84.500/101.65	FLEQQNQVLQTK	57.280	60.875/53.967
EEFGGELIDGGPWLK	81.780	87.500/85.361	LIYTASDLK	57.530	59.000/52.648
YLLVASAIGSLYK	82.340	93.000/97.956	YEELQITAGR	57.630	54.750/47.593
SLDLDSIIAEVK	85.570	81.750/82.562	DYQELMNTK	58.060	46.875/39.571
IATDPFVGNLTFRR	85.620	94.000/89.941	AAAQLQQGLADTSDENLK	58.800	65.750/60.231
DVIADAFLQQILLR	100.43	108.25/99.101	ALLENTELSAR	59.290	63.625/66.275

Таблица 1. Окончание

Последовательность	t_R эксп, мин	t_R^* теор, мин	Последовательность	t_R эксп, мин	t_R^* теор, мин
DGYADGWAQAGTAR	59.430	55.625/56.165	IDVLDSIDPADTGVK	69.330	69.625/74.846
SPMVGTFYR	59.530	58.125/64.297	AVQLGGVALGTTQVINSK	69.910	75.500/77.484
SLEEIIR	60.030	53.500/68.143	ELLVGFGTQVR	70.400	72.625/72.319
SGGGFSSGSAGIINYQR	61.300	64.125/52.538	LGEHNIDVLEGNEQFINAAK	70.740	79.125/77.593
QVIGQVAADLR	61.550	63.250/75.176	DLSDVTLGQFAGK	74.000	65.500/75.505
YPEYNFYR	62.050	52.625/77.703	SILSELVR	74.180	63.125/76.824
EAGLSDLSLK	62.310	57.625/70.121	EQIIFPEIDYDK	75.720	80.125/84.407
SQTVHFQGNPVTWANSIPQAGSK	62.670	70.500/65.176	LLDNAAADLAAISGQK	76.520	73.250/82.978
LQLVGVGYR	62.960	60.875/86.165	TEIEDVLILEPR	78.930	86.375/70.121
QVAILAVAGAEK	63.970	71.875/86.165	NPLYEEIADVTIR	80.900	78.125/66.714
VSVSIFGR	64.020	52.875/80.670	VGLFQDTSAF	81.880	66.125/67.703
GVSADQISIVSYGK	64.420	65.750/72.758	VDLLNQEIEFLK	84.090	85.750/72.978
LNDLEDALQQAK	66.420	59.375/68.692	ALLNSMVIGVTEGFTK	87.390	85.500/80.231
VILAGEVTPVTVR	66.730	77.125/64.736	QGISLSYLEQLFSR	95.070	86.750/73.967
DEVFALSNIQK	69.010	67.125/83.198			

* Концевые группы последовательностей $\text{H}_2\text{N}-$ (~) –COOH. Здесь и в табл. 2 времена t_R теор – предсказываемые в рамках модели BioLCCC (числитель) и по аддитивной модели [4] (знаменатель).

тров фрагментации с использованием поисковой системы Mascot [11]. Для настоящей работы из 20 фракций выбрали два образца, обозначенных в табл. 1 как образцы I и II. Основываясь на результатах поиска в системе Mascot, в этой таблице для сравнения предсказываемых теорией и экспериментально наблюдаемых времен удерживания оставили только те последовательности, достоверность идентификации которых превышала 95%.

Все хроматографические системы были стандартизованы с помощью пептидного стандарта S₁–S₅ (“Pierce Inc.”, USA), который представляет собой смесь синтетических пептидов, четыре из которых имеют общую структуру Ac-RGxxGLGLGK-amide, где xx соответствует GG (S₂), AG (S₃), VG (S₄) и VV (S₅). Стандарт S₁ имеет последовательность остатков, идентичную S₃, однако отличается от последнего наличием свободной α -аминогруппы. Такая стандартизация позволяет определить активность поверхности адсорбента (масштабный фактор α в уравнении (1)) для разных колонок.

Для разделения пептидов использовали неподвижные фазы типа –C₁₈ на основе силикагеля. Применение фаз разных производителей было

сделано осознанно, чтобы показать общность предсказательных свойств модели BioLCCC.

РЕЗУЛЬТАТЫ И ИХ ОБСУЖДЕНИЕ

Начнем обсуждение с экспериментальных результатов разделения стандартов S₁–S₅. На рис. 1 приведена хроматограмма для этих стандартов, полученная на колонке ReproSil Pur C₁₈ AQ. На рис. 2а представлена корреляция между экспериментальными временами удерживания и предсказанными в рамках модели BioLCCC. Как видно, наблюдается хорошее соответствие между предсказанными и реальными временами удерживания. Для примера на рис. 2б приведена аналогичная корреляция между экспериментальными и расчетными временами удерживания, полученными в рамках аддитивной модели [4]. Даже для достаточно коротких цепей предсказания модели BioLCCC оказываются сопоставимыми по точности с предсказаниями аддитивной модели.

Вместе с тем стоит отметить, что стандарт S₁ явно выпадает из линейной зависимости на рис. 2. Отличие этого стандарта от S₃ заключается в том, что он имеет на конце ионизированную группу H₃N⁺. Очевидно, что указанная группа сильно гидрофильна и отталкивается от поверхности

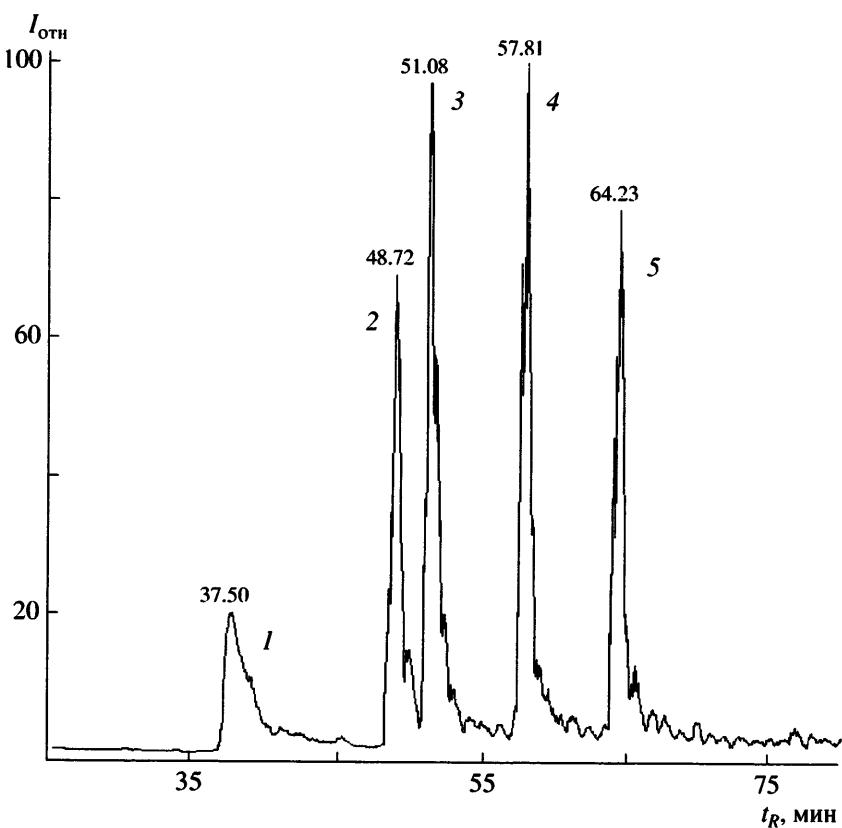


Рис. 1. Хроматограмма пептидного стандарта S₁–S₅ (1–5), полученная на наносистеме Agilent 1100. Условия разделения: колонка 0.075 × 150 мм, фаза ReproSil Pur C18 AQ, размер частиц 3 мкм, размер пор 120 Å, объем вводимой пробы 8 мкл, изократическое элюирование 20 мин смесью, содержащей 4 об. % компонента В (89.5 об. % ацетонитрила, 10 об. % воды и 0.5 об. % уксусной кислоты), линейный градиент 4 – 45 об. % В в течение 90 мин.

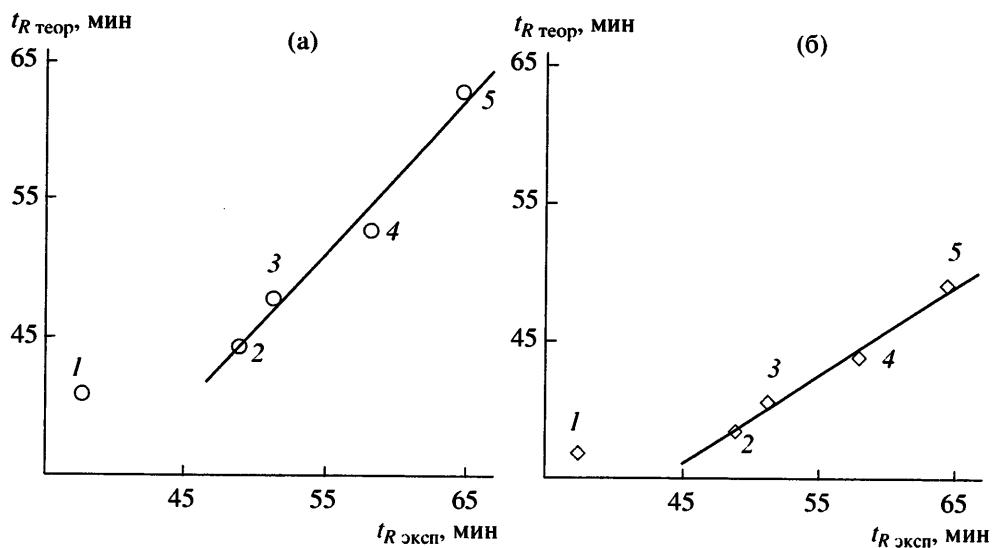


Рис. 2. Корреляционная зависимость между предсказанными в рамках модели BioLCCC (а) и аддитивной модели [4] (б) и экспериментально наблюдаемыми временами удерживания последовательностей стандарта S₁–S₅. Коэффициент корреляции $R^2 = 0.974$ (а) и 0.983 (б).

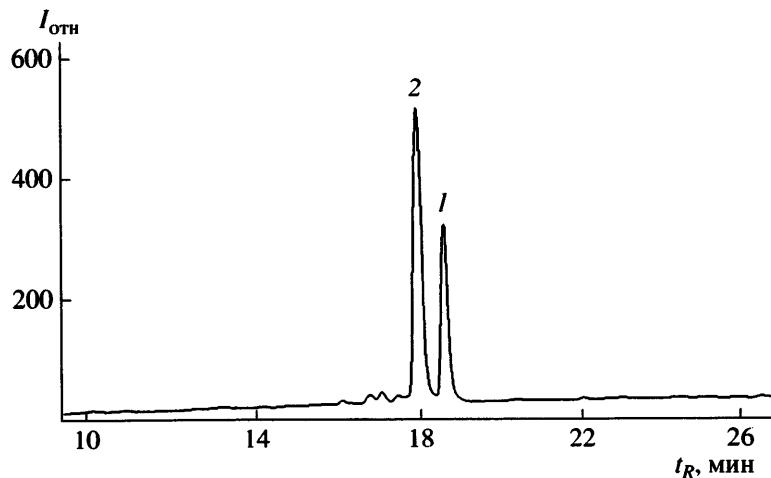


Рис. 3. Разделение синтетических пептидов H-RAAAAIAAAAR-OH (1) и H-RAAAALAAAAR-OH (2), последовательности которых содержат изомеры лейцина. Расчетные в рамках модели BioLCCC и экспериментально наблюдаемые времена удерживания соответственно 18.50 и 18.54 мин (1), 17.90 и 17.88 мин (2). Условия разделения: хроматографическая система HP1100, колонка 4.0 × 150 мм, фаза Диасфер 110-C₁₈, размер частиц 5 мкм, размер пор 110 Å, объем пробы 20 мкл, концентрация образца 0.1 мг/мл, линейный градиент от начальной точки 2 об. % В (ацетонитрил, содержащий 0.1 об. % ТФУК) до 30 об. % В в течение 30 мин.

(другими словами, отталкивание обусловлено тем, что “изображение” заряда на границе раздела полярной фазы (воды) и неполярной –C₁₈ имеет тот же знак). Поэтому объем удерживания стандарта S₁ заметно меньше объема удерживания S₃, имеющего тот же “текст”, но незаряженную концевую группу на N-конце. Модель BioLCCC учитывает такое отталкивание (энергия взаимодействия группы H₃N⁺ – принята равной –1.69 [1]). Однако, если предположить бесконечно сильное отталкивание концевого звена, это приводит при вычислении статсуммы к исключению лишь небольшого числа конфигураций цепи, а именно тех, которые начинаются на поверхности. По-видимому, этого недостаточно для наблюданного уменьшения объема удерживания стандарта S₁. Необходимо учитывать нелокальность взаимодействия заряженных концевых групп, что достаточно просто сделать в переходных матрицах уравнения (1). Отметим, что такой учет нелокальности идеально близок к поправкам для энергий адсорбции аминокислотных остатков, “привязанных” к N-концу пептида в аддитивной модели [7]. Тем не менее правильный учет взаимодействия заряженных концевых групп NH₃⁺ – (а также их взаимного влияния друг на дру-

га в цепи) в рамках модели BioLCCC пока остается открытым.

Рассмотрим далее разделение макромолекул, содержащих изомеры L и I. Интерес к хроматографическому разделению и идентификации таких макромолекул состоит в том, что указанные последовательности не идентифицируются (или, по крайней мере, очень сложно идентифицируются) любыми методами масс-спектрометрии, в том числе и tandemной масс-спектрометрии по спектрам фрагментации. Вместе с тем хроматография позволяет легко разделить последовательности, что видно из хроматограммы, представленной на рис. 3. Согласие предсказанного времени выхода пептидов с экспериментальным оказывается очень хорошим. Такой пример подтверждает сформулированный выше тезис о том, что хроматографические сведения о последовательности, получаемых в рамках модели BioLCCC, дополняют данные масс-спектрометрии: для масс-спектрометрии пептидная пара из рассмотренного примера неразличима, в то время как для хроматографии – это два разных “текста”.

Интересно отметить возможность разделения таких последовательностей не только по типу изомеров лейцина, но также и по их месту в цепи. Например, для последовательностей типа H-RAAAALAAAAR-OH и H-RALAAAAAAAR-OH

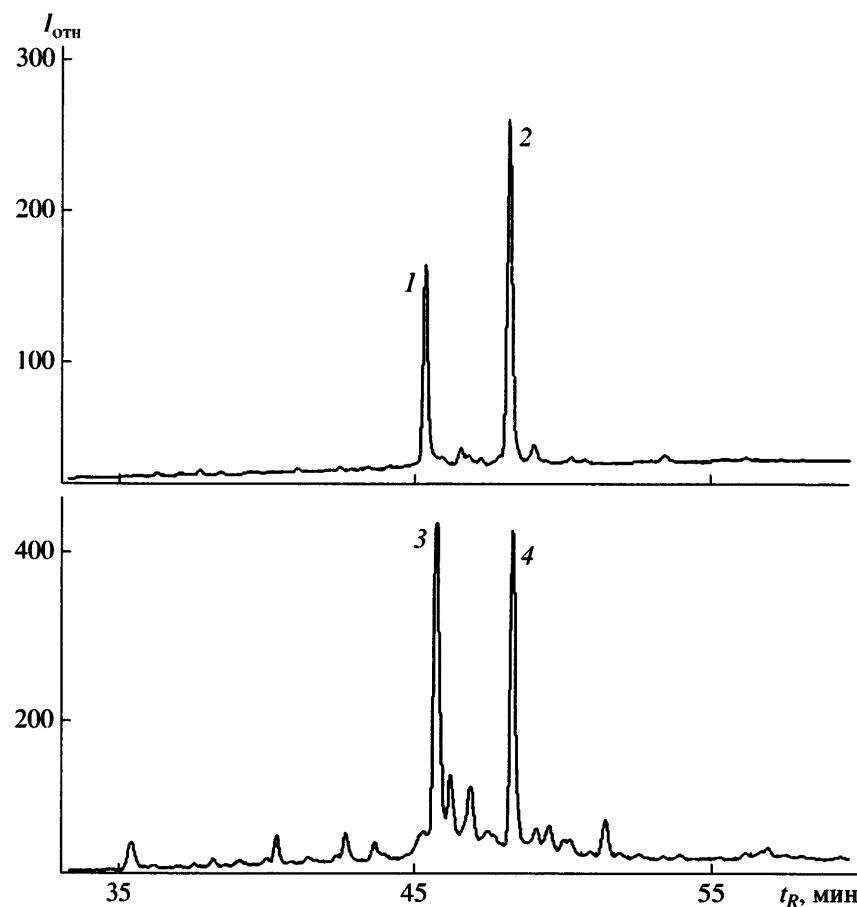


Рис. 4. Разделение “зеркальных” пептидов типа H-GALYIYLGDGLDTADAEG-amide (1), H-GEADATDLGDGLIYILAG-amide (2) и H-GALYIYLGDGLDTADAEG-OH (3), H-GEADATDLGDGYIYILAG-OH (4). Расчетные в рамках модели BioLCCC и экспериментально наблюдаемые времена удерживания соответственно 42.04 и 45.34 мин (1), 43.92 и 48.22 мин (2), 41.99 и 45.70 мин (3), 43.83 и 48.31 мин (4). Условия разделения: хроматографическая система HP1100, колонка Vydac 218TP54 4.6 × 250 мм, фаза -C₁₈, размер частиц 5 мкм, размер пор 300 Å, объем пробы 20 мкл, концентрация образца 0.1 мг/мл, линейный градиент от 2 об. % В (90.0 об. % ацетонитрила, 9.9 об. % воды и 0.1 об. % ТФУК) до 60 об. % В в течение 60 мин.

расчетные объемы удерживания различны для условий разделения, приведенных на рис. 3. Экспериментальная проверка этого факта будет служить прямым доказательством возможности использования модели BioLCCC для определения места (номера остатка) модифицированных и(или) изомерных форм аминокислот в цепи, что является одной из наиболее трудных задач протеомики. Исследование закономерностей разделения модифицированных пептидов и биомакромолекул будет предметом дальнейших экспериментальных исследований.

Следующий пример разделения пептидов идеально гораздо более сложен. Он относится к разделению перевернутых, “зеркальных” последовательностей. В первой части работы [1] была

предсказана возможность разделения таких последовательностей, имеющих асимметричный энергетический профиль. Последовательности имеют, очевидно, одинаковую массу, так что их “чтение” с помощью точного измерения массы макромолекулярного иона невозможно и требует использования методов тандемной масс-спектрометрии (однако даже в этом случае возможны ситуации, когда спектры фрагментации не позволяют различить пары зеркальных изомеров).

Для доказательства возможности разделения “зеркальных” пептидов были синтезированы последовательности типа H-GALYIYLGDGLDTADAEG-amide и H-GEADATDLGDGYIYILAG-amide, отличающиеся переворотом внутренней части. Также были синтезированы те же последо-

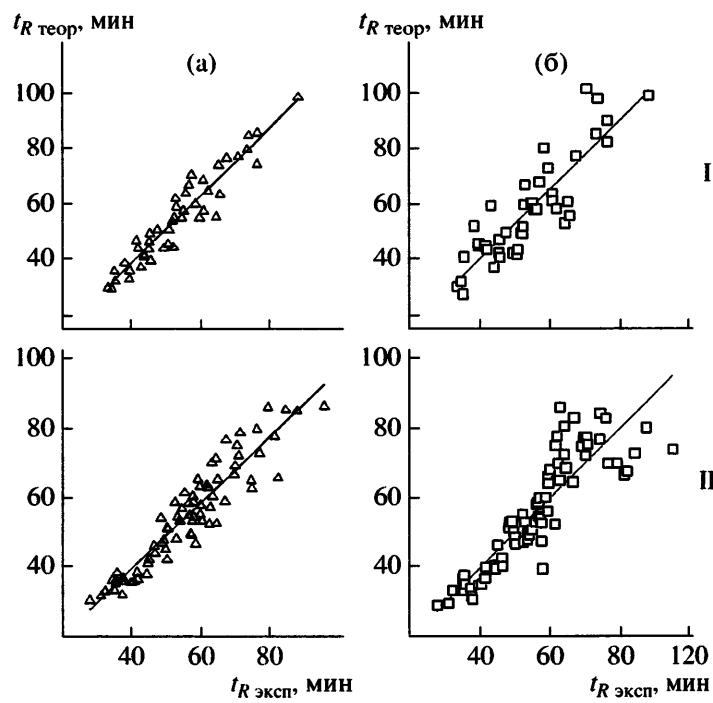


Рис. 5. Сравнение рассчитанных по модели BioLCCC (а) и аддитивной модели [4] (б), а также экспериментально наблюдаемых времен удерживания (мин) последовательностей пептидов белков *E. coli*, идентифицированных в образцах I и II (табл. 1). Коэффициент корреляции и дисперсия времен удерживания (мин) соответственно: $R^2 = 0.906$ и $\sigma = 0.76$ (Ia), 0.780 и 1.32 (Ib), 0.879 и 0.69 (IIa), 0.762 и 0.93 (IIb).

вательности, отличающиеся типом концевой группы на С-конце: H-GALYIYLGDGLDTADAEG-OH и H-GEADATDLGDIYIYLAG-OH. “Зеркальные” последовательности можно представить как перестановку концевых аминокислотных остатков. Как показано на рис. 4, эти последовательности действительно разделяются в полном соответствии с предсказаниями теории BioLCCC. Естественно, что разделение таких структур не следует из аддитивной модели или ее аналогов. Здесь практически в чистом виде проявляется эффект связанности остатков в цепь и коллективный характер их взаимодействия с поверхностью. Учет влияния перестановки аминокислотных остатков в цепи на статистическую сумму гетерополимера вблизи поверхности несложно сделать в рамках подхода, развитого в работе [8]. Отметим существенную роль гетерогенности цепи: для макромолекул гомополимера перестановка разных концевых групп вырождена в смысле взаимодействия с поверхностью.

Наконец, рассмотрим возможность использования модели BioLCCC для анализа хроматографических данных, полученных в исследованиях

смесей триптических пептидов белков бактерии *E. coli*. Дайджесты белков *E. coli* изучены достаточно хорошо, встречающиеся в них последовательности приведены в табл. 1. Эти последовательности были идентифицированы программой Mascot с достоверностью, превышающей 95%. Вместе с тем следует сказать, что далеко не для всех из них были получены полные спектры фрагментации высокого качества, так что достоверность их первичной структуры не абсолютная.

Как следует из рис. 5а, и для реальных систем наблюдается хорошая корреляция (коэффициент корреляции $R^2 \sim 0.9$) между предсказанными и реальными временами удерживания со стандартной ошибкой предсказания менее 1 мин. Тем самым информация о возможном “тексте” последовательности, получаемая из времени удерживания, может быть использована для повышения достоверности идентификации пептидов, а значит, и белков протеома организма. Как следует из рис. 5б, корреляция между предсказанным и наблюдаемым временем удерживания в модели BioLCCC заметно лучше, чем в аддитивной модели.

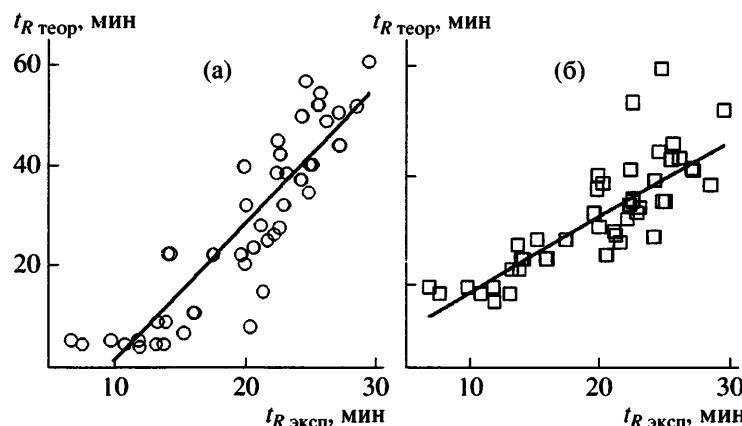


Рис. 6. Сравнение расчетных в рамках модели BioLCCC (а) и аддитивной модели [4] (б) и экспериментально наблюдаемых времен удерживания последовательностей пептидов белка BSA (табл. 2). $R^2 = 0.79$ (а) и 0.66 (б).

Такой же анализ можно применить и к последовательностям триптических пептидов белка альбумина BSA. Для них корреляция между предсказанными и экспериментальными временами удерживания приведена на рис. 6. В целом, однако, степень корреляции в данном случае заметно хуже, чем для белков E. coli. Причина этого пока не вполне ясна. В отличие от белков E. coli для альбумина мы не ограничивались только последовательностями, достоверность которых более 95%. На рис. 6 и в табл. 2 представлены все последовательности, идентифицированные Mascot по массам именно как последовательности, присутствующие в альбумине. На наш взгляд, это лишний раз доказывает, что идентификация последовательностей пептидов только по массам (даже пептидов, полученных из белка с известной первичной структурой) обладает невысокой достоверностью.

ЗАКЛЮЧЕНИЕ

На основании приведенных экспериментальных исследований можно сделать следующий вывод: модель BioLCCC правильно схватывает основные черты хроматографического разделения пептидов и позволяет предсказать возможность разделения последовательностей, имеющих весьма тонкие различия в их первичной структуре. Масштаб различий в стандартных энергиях адсорбции аминокислотных остатков на обращенной фазе типа $-C_{18}$ в паре вода–ацетонитрил с добавками кислоты (рН 2.0), по-видимому, соответствует действительности. Тем самым модель

BioLCCC может быть использована для оценки возможностей разделения тех или иных последовательностей пептидов, определения “дефектности” этих последовательностей, их модификаций в реальных условиях хроматографического эксперимента.

Следует сказать, что эта модель требует уточнения в тех случаях, когда в цепи имеются заряженные аминокислотные остатки.

Наконец, “разрезание” длинных цепей белков путем ферментативного гидролиза трипсином, возможно, и не является необходимым с точки зрения их хроматографического разделения. Восприимчивость взаимодействия макромолекулы с поверхностью к слабым изменениям ее “текста” (например, к перестановкам двух остатков, замене или модификации одной аминокислоты) может оказаться заметно сильнее именно для длинных цепей, что следует из теоретического анализа адсорбции гетерополимеров [8]. Для хроматографического исследования первичной последовательности белков и применения модели BioLCCC для поиска оптимальных условий разделения необходимо перевести белки из глобулярного в клубкообразное состояние (денатурировать). При нахождении белков в глобулярном состоянии во взаимодействии с поверхностью будут участвовать лишь аминокислотные остатки, находящиеся снаружи глобулы. В данном случае разделение скорее отражает вторичную или третичную структуру белка, нежели его последовательность. Особый интерес представляет воз-

Таблица 2. Последовательности триптических пептидов альбумина BSA, использовавшиеся при сопоставлении предсказанных и экспериментальных времен удерживания

Последовательность	Score*	t_R эксп., мин	t_R теор., мин	Последовательность	Score*	t_R эксп., мин	t_R теор., мин
SEIAHR	38	6.7	5.55/19.90	GACLLPK	34	22.18	26.40/32.30
TPVSEK	20	7.58	4.70/18.70	EYEATLEECCAK	80	22.36	38.65/34.90
SEIAHR	38	9.72	5.55/19.90	EYEATLEECCAK	62	22.39	38.65/34.90
TPVSEK	20	10.75	4.70/18.70	CCAADDKEACFAVEGPK	87	22.43	45.05/41.40
SEIAHR	45	11.77	5.55/19.90	AEFVEVTK	58	22.58	27.75/53.60
NYQEAK	22	11.88	4.15/17.20	NECFLSHKDSPDLPK	29	22.61	42.35/35.50
TPVSEK	35	13.11	4.70/18.70	ECCHGDLLEADDR	51	22.64	42.30/36.00
ATEEQLK	57	13.22	9.15/23.10	EACFAVEGPK	67	22.9	32.35/33.50
AFDEK	22	13.71	4.75/27.60	EACFAVEGPK	57	22.93	32.35/33.50
ATEEQLK	46	13.85	9.15/23.10	YNGVFQECCQAEDK	57	23.14	38.50/34.40
TCVADESHAGCEK	48	14.01	22.60/24.90	YNGVFQECCQAEDK	33	23.17	38.50/34.40
TCVADESHAGCEK	67	14.15	22.60/24.90	DDPHACYSTVFDK	44	24.21	37.25/29.00
IETMR	22	15.21	6.90/28.50	DDPHACYSTVFDK	14	24.24	37.25/29.00
QNCDQFEK	41	15.9	10.95/25.00	KVPQVSTPTLVEVSR	71	24.28	49.80/39.40
QNCDQFEK	54	16.01	10.95/25.00	KVPQVSTPTLVEVSR	27	24.31	49.80/39.40
SHCIAEVEK	58	17.43	22.45/28.50	RHPEYAVSVLLR	88	24.59	56.70/44.60
SHCIAEVEK	64	17.47	22.45/28.50	YLYEIAR	55	24.82	34.80/59.70
LCVLHEK	41	19.6	22.40/33.40	LKPDPNTLCDEFK	56	24.87	40.30/35.50
LCVLHEK	40	19.63	22.40/33.40	LKPDPNTLCDEFK	8	25.1	40.30/35.50
LKECCDKPLLEK	87	19.84	39.85/37.80	KQTALVELLK	83	25.5	52.10/43.20
LKECCDKPLLEK	89	19.88	39.85/37.80	KQTALVELLK	32	25.56	52.10/43.20
NECFLSHK	44	19.94	20.60/40.30	RPCFSALTPDETYVPK	69	25.65	54.40/46.10
ECCDKPLLEK	56	20.04	32.20/30.80	VPQVSTPTLVEVSR	56	26.14	48.85/43.50
KFWGK	29	20.31	8.20/38.90	SLHTLFGDELCK	21	27.06	50.50/41.60
CCTESLVNR	63	20.56	23.90/25.70	LVNELTEFAK	57	27.12	44.15/41.20
CCTESLVNR	63	20.6	23.90/25.70	LVNELTEFAK	21	27.16	44.15/41.20
YICDNQDTISSK	83	21.15	28.20/30.00	QTALVELLK	54	28.46	51.80/38.60
DLGEEHFK	42	21.31	15.15/29.30	LGEYGFQNALIVR	20	29.42	60.65/52.20
FKDLGEEHFK	82	21.66	25.25/28.10				

* Параметр Score, определяемый в программе Mascot, характеризует достоверность идентификации последовательности (чем больше, тем выше достоверность).

можность разделения модифицированных белков, для которых можно поставить в том числе и задачу разделения таких белков по месту модификации. Этот вопрос требует дальнейшего экспериментально исследования. Модель BioLCCC позволяет решать задачу в общем виде и дать оценку возможностям хроматографии на обращенной фазе для “чтения текстов” таких белков.

Что касается других вариантов хроматографического разделения белков (ионнообменной или аффинной), то модель BioLCCC не специфична к конкретному типу взаимодействий. Для ее применимости важно, чтобы взаимодействие было короткодействующим. В других вариантах хроматографии необходимо переопределить стандартные энергии адсорбции остатков с помощью модель-

ных пептидов, как было сделано для хроматографии на обращенной фазе.

СПИСОК ЛИТЕРАТУРЫ

1. Горшков А.В., Тарасова И.А., Евреинов В.В., Горшков М.В. // Высокомолек. соед. Б. 2007. Т. 49. № 4. С. 732.
2. Gorshkov A.V., Tarasova I.A., Evreinov V.V., Savitski M.M., Nielsen M.L., Zubarev R.A., Gorshkov M.V. // Analyt. Chem. 2006. V. 78. P. 7770.
3. Guo D., Mant C.T., Taneja A.K., Parker J.M.R., Hodges R. // J. Chromatogr. 1986. V. 359. P. 499.
4. Guo D., Mant C.T., Taneja A.K., Hodges R. // J. Chromatogr. 1986. V. 359. P. 519.
5. Mant C.T., Burke L.T.W., Black J.A., Hodges R. // J. Chromatogr. 1988. V. 458. P. 193.
6. Mant C.T., Burke L.T.W., Zhou N.E., Parker R.J.M., Hodges R. // J. Chromatogr. 1990. V. 485. P. 365.
7. Krokhin O.V., Craig R., V. Spicer V., Ens W., Standing K.G., Beavis R.C., Wilkins J.A. // Mol. Cell. Proteomics. 2004. V. 3. № 9. P. 908.
8. Обухов С.П. // Журн. эксперим. и теорет. физики. 1987. Т. 93. № 6. С. 1973.
9. Spengler B. // J. Am. Soc. Mass Spectrom. 2004. V. 15. P. 703.
10. Marshall A.G., Verdun F.R. Fourier Transforms in NMR, Optical, and Mass Spectrometry: A User's Handbook. Amsterdam: Elsevier, 1990.
11. Nielsen M.L., Savitski M.M., Zubarev R.A. // Mol. Cell. Proteomics. 2005. V. 4. № 6. P. 835.

Applicability of the Critical Chromatography Concept to Proteomics Problems: Experimental Study of the Dependence of Peptide Retention Time on the Sequence of Amino Acids in the Chain

I. A. Tarasova^a, A. V. Gorshkov^b, V. V. Evreinov^b, K. Adams^c,
R. A. Zubarev^c, and M. V. Gorshkov^a

^a Institute of Energy Problems of Chemical Physics, Russian Academy of Sciences,
Leninskii pr. 38/2, Moscow, 119334 Russia

^b Semenov Institute of Chemical Physics, Russian Academy of Sciences,
ul. Kosygina 4, Moscow, 119991 Russia

^c Laboratory for Biological and Medical Mass Spectrometry, Biological and Medical Center,
Uppsala University, Box 583, S-75 123 Uppsala, Sweden
e-mail: evreinov@polymer.chph.ras.ru)@

Abstract—Experimental data on the separation of synthetic and natural peptides are presented as treated in terms of the separation model proposed by the authors that allows for the chain connectivity of amino acid residues and the cooperative character of their interaction with the surface. It was shown that the model accurately predicts the separation of peptides with identical amino acid contents and different sequences of units in the chain. The differences in the sequence may be permutation of amino acid residues and the presence of terminal groups, amino acid isomers, or mirror sequences in the chain. The separation model was used to predict the retention times of peptides prepared via the enzymatic hydrolysis of *E. coli* proteins and bovine serum albumin with trypsin. It was shown that the model in general accurately explains the array of experimental data on the separation of such peptides, thus being the first successful attempt to relate the chain sequence to the retention volume.