

## О РАЗДЕЛЕНИИ ПЕПТИДОВ С ИНВЕРТИРОВАННЫМИ ПОСЛЕДОВАТЕЛЬНОСТЯМИ

© 2024 г. А. В. Горшков<sup>1, \*</sup>, И. Г. Роздина<sup>1</sup>, М. Л. Придатченко<sup>2</sup>

<sup>1</sup>Федеральный исследовательский центр химической физики им. Н. Н. Семенова Российской академии наук  
119991 Москва, ул. Косыгина, 4

<sup>2</sup>Институт энергетических проблем химической физики им. В.Л. Тальрозе Федеральный исследовательский центр  
химической физики им. Н.Н. Семенова Российской академии наук 119334 Москва, Ленинский пр., 38

\*e-mail: alex.gorshkov@gmail.com

Поступила в редакцию 16.04.2024 г.

После доработки 13.08.2024 г.

Принята к публикации 16.09.2024 г.

Экспериментально и теоретически исследовано разделение пептидов с инвертированными от N-конца к С-концу последовательностями. Хотя такие последовательности состоят из одинаковых аминокислотных остатков, их адсорбционные свойства различаются. Теоретический расчет в рамках модели разделения пептидов *BioLCCC* показал, что разница в адсорбционных свойствах таких последовательностей обусловлена различием в пространственной корреляции связанных в цепь аминокислотных остатков и концевых групп при их взаимодействии с поверхностью.

DOI: 10.31857/S2308112024030089, EDN: LWKPPM

### ВВЕДЕНИЕ

Важным свойством биомакромолекул (пептидов, белков, РНК и ДНК) является зависимость их свойств от порядка чередования мономеров разного химического строения в цепи, называемого иначе первичной структурой или текстом последовательности. От текста последовательности зависит как возможность сворачивания биомакромолекулы в те или иные плотные, глобулярные структуры, определяющие биологические функции белков, так и их взаимодействие друг с другом с образованием сложных трехмерных комплексов. Иногда даже перестановка или модификация нескольких звеньев в длинном тексте последовательности цепи приводит к невозможности правильного сворачивания макромолекулы и, следовательно, к потере ее биологической активности. По этой причине установление первичной структуры биомакромолекулы, ее модификации остается одной из задач современной науки, решаемой разнообразными химическими и физическими методами.

Одним из проявлений зависимости свойств биомакромолекул от текста последовательности является их адсорбционное взаимодействие с поверхностью. Такое свойство пептидов и белков используется в хроматографии для их

разделения и поиска изменений в последовательности, вызванных теми или иными точечными мутациями. Разделение пептидов, получаемых методами ферментативного гидролиза сложной смеси белков, выделенных из организма, дополненное методами масс-спектрометрического детектирования, играет важную роль в протеомике для идентификации белков и их модификаций. Дополнение данных масс-спектрометрии временем удерживания пептида дает возможность повысить достоверность идентификации белков, содержащих данный пептид, и заметно сократить время анализа. При таком подходе можно ограничиться точным измерением масс пептидов и исключить их фрагментацию [1]. Для этого необходимо иметь в арсенале исследователя модель разделения, дающую возможность рассчитать адсорбционное взаимодействие пептида с известной последовательностью и тем самым предсказать его время выхода или объем удерживания.

Зависимость адсорбционного взаимодействия пептида от текста последовательности в настоящее время можно предсказать на основе нескольких моделей. Аддитивная модель, типичная в практике хроматографии, предполагает, что время удерживания определяется суммой коэффициентов удерживания, которые

различаются для разных остатков [2–5]. Она имеет понятные физические обоснования для низкомолекулярных соединений или коротких пептидов, которые взаимодействуют с поверхностью целиком, всеми входящими в их состав аминокислотными остатками. В простой аддитивной модели отсутствует зависимость адсорбционных свойств пептида от перестановки букв в тексте последовательности, тем не менее она отражает его аминокислотный состав. Учитывая то, что пептиды, получаемые в результате ферментативного гидролиза белков, как правило, отличаются и по длине, и по аминокислотному составу, эта модель может с успехом применяться как начальная точка для предсказания времени выхода таких пептидов в дополнение к данным по массе пептида и массам его фрагментов.

Развитие аддитивной модели привело к пониманию того, что даже относительно короткие пептиды, в которых чередуются гидрофобные и гидрофильные звенья, могут при взаимодействии с поверхностью образовывать хвосты. В частности, заряженный положительно N-конец пептида, отталкивающийся от гидрофобной поверхности, с необходимостью удаляет от неё близкие к N-концу аминокислотные остатки, тем самым уменьшая их эффективную энергию взаимодействия с поверхностью. Если в аддитивную модель внести дополнительные связанные с этим явлением поправки, то ее предсказания становятся заметно точнее [6–8]. Далее можно внести поправки на длину цепи пептида, на спиральные участки или на какие-то другие особенности строения цепи. Однако такой подход порождает огромное количество дополнительных эмпирических коэффициентов, которые, хотя и описывают уже какие-то особенности вариации текста последовательности, имеют отдаленный физический смысл и затрудняют перенос модели, например, на другие системы разделения. В целом же модифицированная аддитивная модель представляет собой попытку описания взаимодействия пептида с поверхностью со стороны низкомолекулярного соединения.

Другим вариантом является использование нейронной сети для предсказания времени выхода разных последовательностей пептидов. Правильно сконструированная и обученная на широком тестовом наборе последовательностей нейронная сеть дает наиболее точные предсказания времени удерживания пептидов [9–12]. Нет сомнения, что при дальнейшем «обучении» такой сети в нее будут включены практически все пептиды белков, существенные для их идентификации: количество таких пептидов, конечно, гигантское, но все же ограниченное. Однако физическая причина,

по которой вариации последовательностей имеют разные адсорбционные взаимодействия, в нейронной сети не очень прозрачна.

Наконец, вариант модели разделения, развиваемый авторами, ориентирован на подход к пептидам со стороны макромолекул. Пептиды, хотя и относительно короткие, имеют цепочечную, макромолекулярную структуру. Если их длина порядка 20 звеньев, то и в адсорбционном взаимодействии проявляются свойства макромолекул. Для последних характерен резкий переход в адсорбированное состояние в узком диапазоне состава растворителя и температуры. Адсорбционные свойства макромолекул и их связь с хроматографическим разделением хорошо описываются моделью цепи случайных блужданий в щелеобразной поре. Эта модель в разных вариантах – гауссовой цепи или решеточной цепи, предсказывает существование критической точки адсорбции. Для гомополимеров в критической точке исчезает молекулярно-массовое распределение как следствие масштабной инвариантности, в том числе и для относительно коротких цепочек олигомеров, что открывает широкие возможности для исследования других типов неоднородности цепи [13–15]. Естественное обобщение для гетерополимеров с разным характером взаимодействия звеньев с поверхностью реализовано в модели *BioLCCC* [16–21]. В ней зависимость адсорбционного взаимодействия от текста последовательности автоматически возникает как следствие энтропийного вклада петель и хвостов, существующих в конфигурации адсорбированной цепи вблизи критической точки. Перестановка звеньев с различной энергией адсорбции меняет размер петель и их энтропийный вклад в адсорбционное взаимодействие, следовательно, зависимость от последовательности имеет в основном энтропийную природу. Поскольку изменение адсорбционных свойств макромолекулы происходит в узком диапазоне внешних параметров, в модель можно включить зависимость энергии адсорбции звена от состава растворителя с помощью корреляционной теории, что дает возможность описать разделение в градиентах, на разных адсорбентах и в разных системах разделения. В настоящей работе рассматривается простая вариация текста последовательности – его переворот от N- к C-концу (будем называть такие последовательности инвертированными) и выясняется физическая причина, по которой такой переворот приводит к изменению адсорбционных свойств макромолекулы.

## МАТЕРИАЛЫ И УСЛОВИЯ ЭКСПЕРИМЕНТА

Последовательности изученных пептидов приведены в табл. 1. Как можно заметить, в образцах

Таблица 1. Исследованные в работе пептиды с инвертированными последовательностями

Пептид, №	Последовательность	Пептид, №	Инвертированная последовательность
1	<i>ALQQQQQQQQQK</i>	1'	<i>QQQQQQQQQLAK</i>
2	<i>NPQNSSQSADGLR</i>	2'	<i>LGDASQSSNQPNR</i>
3	<i>ILELDQFKGQQGQKR</i>	3'	<i>RKQGQQGKFQDLELI</i>
4	<i>KQQNQELQEQLR</i>	4'	<i>RLQEQLEQNQQK</i>
5	<i>FLITHNPTNATLNK</i>	5'	<i>KNLTANTPNHTILF</i>

1 и 2 последовательность инвертирована без аминокислотного остатка на С-конце, тогда как в образцах 3, 4 и 5 последовательность инвертирована полностью. Этих пептиды исследовали в разное время на нанохроматографе “Dionex UltiMate 3000”, соединенном с масс-спектрометром “Amazon SL” в разных системах.

**Система I:** колонка Acclime perMap 100 Å, размер колонки 150 × 0.075 мм, фаза C<sub>18</sub>, размер частиц адсорбента 3 мкм, T = 40 °C, растворитель А – вода + 0.1% муравьиной кислоты (МК), растворитель В – ацетонитрил (АЦН) 80% + вода 20% + 0.1% МК, рН 2.0, скорость подачи растворителя 300 нл/мин, объем вводимой пробы 1 мкл.

**Система II:** колонка YMC 300 Å, размер колонки 150 × 0.075 мм, фаза C<sub>18</sub>, размер частиц адсорбента 5 мкм, T = 45 °C, растворитель А – вода + 0.1% МК + 0.01% трифторуксусной кислоты (ТФУ), растворитель В – АЦН + 0.1% МК + 0.01% ТФУ, рН 2.0, скорость подачи растворителя 300 нл/мин, объем вводимой пробы 1 мкл.

**Система III:** колонка Peaky 120 Å, размер колонки 5 см × 0.075 мм, фаза C<sub>18</sub>, размер частиц адсорбента 1.9 мкм, T = 65 °C, растворитель А – вода + 0.1% МК, растворитель В 80% АЦН + 20% вода + 0.1% МК, рН 2.0, скорость подачи растворителя 1500 нл/мин, объем вводимой пробы 1 мкл.

**Система IV:** колонка Acclime perMap 100 Å, размер колонки 15 см × 0.075 мм, размер частиц адсорбента 3 мкм, T = 40 °C, растворитель А – вода + 100 мМоль бикарбоната аммония (БА), растворитель В – 80% АЦН + 20% вода + 100 мМоль БА, рН 8.0, скорость подачи растворителя 300 нл/мин, объем вводимой пробы 1 мкл.

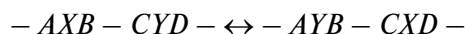
Градиенты, использованные для разделения пептидов в разных системах, приведены на рис. 1.

Эксперименты проводили в разное время, и для подключения колонок к нанонасосу и масс-спектрометру использовали капилляры разной длины и разного диаметра. Эти экстраколочные объемы для некоторых систем были больше или порядка объема наноклонок, что приводило к задержке градиента до колонки и задержке транспортировки пептида от колонки

к узлу ввода в масс-спектрометр. Данные задержки учитывались нами при теоретических расчетах приблизительно путем добавления к расчетному времени удерживания времени задержки. Детектирование проводили по уникальным массам пептидов, а идентификация прямых и инвертированных последовательностей – по массам фрагментов. Показанные на рис.1 градиентные профили было бы правильнее отображать не во временной шкале, а в шкале объема растворителя, прокачиваемого через колонку. При такой нормировке градиент в системе III представляется очень пологим, что оправдано для короткой колонки, учитывая резкий переход пептида из адсорбированного состояния в раствор. Однако мы оставили традиционное представление градиентов во временной шкале.

### Модель хроматографического разделения

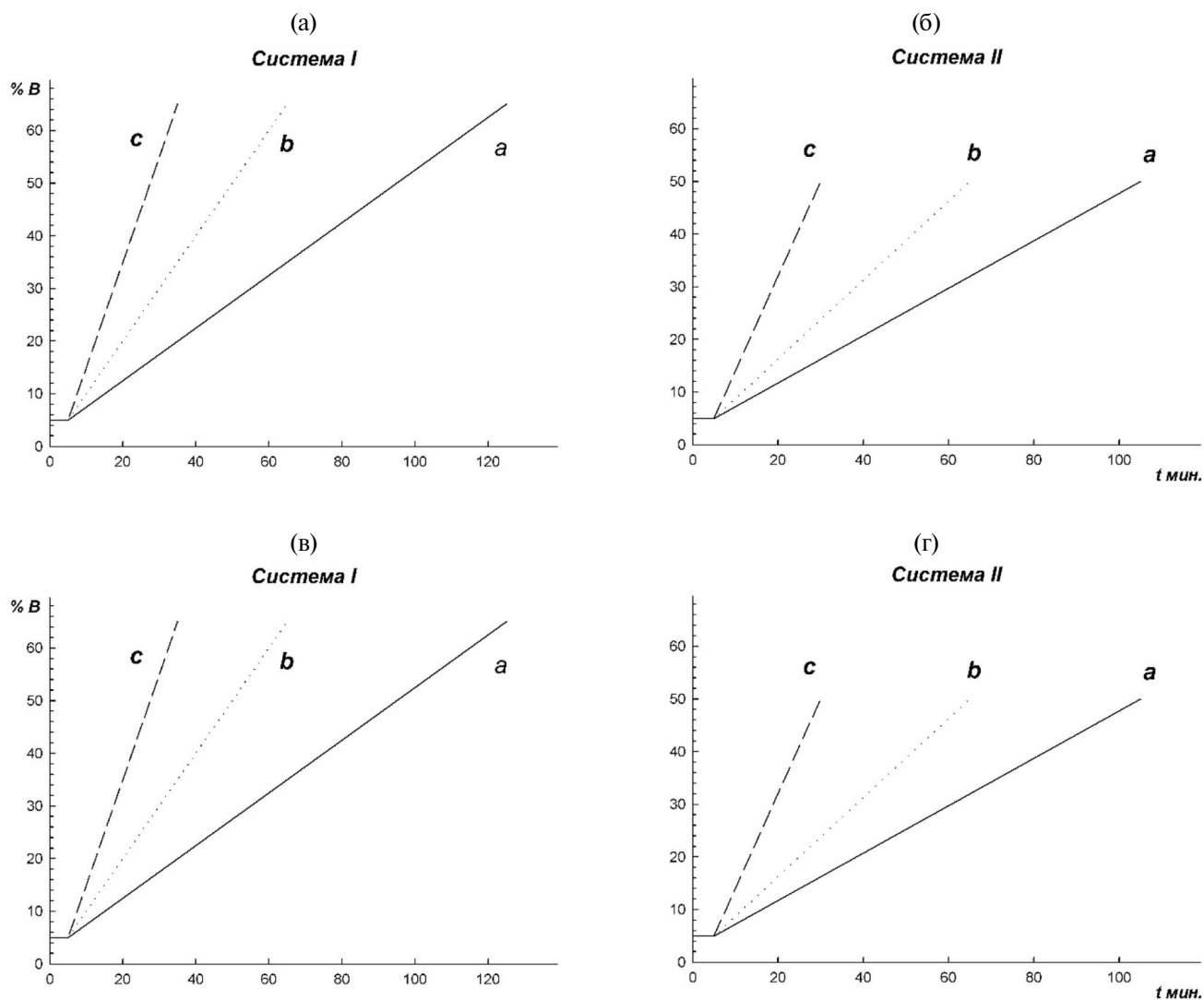
Прежде чем перейти к обсуждению модели разделения, сделаем несколько замечаний. Элементарную вариацию текста последовательности аминокислотных остатков в цепи пептида дает перестановка двух букв – X и Y:



Если перестановка происходит в цепи гомополимера, что обычно имеет место для олигомеров, содержащих небольшое число реакционноспособных групп либо на концах, либо в самой цепи, то переставляемые звенья в разных местах находятся в окружении одинаковых мономеров



и их энергия взаимодействия с поверхностью за счет соседей не меняется. Строго говоря, это не совсем так, поскольку локальная плотность мономеров M зависит от места переставляемой группы в цепи и различна в середине и на конце цепи. Если учесть объемные взаимодействия между звеньями, которые имеют отталкивательный



**Рис. 1.** Профили градиентных программ для систем I–IV. Система I – (время  $t$ , мин : B, %): (0 : 5, 5 : 5, 125 : 65) **a**, (0 : 5, 5 : 5, 65 : 65) **b**, (0 : 5, 5 : 5, 35 : 65) **c**; система II – (0 : 5, 5 : 5, 105 : 50) **a**, (0 : 5, 5 : 5, 65 : 50) **b**, (0 : 5, 5 : 5, 30 : 50) **c**; система III – (0 : 5, 5 : 5, 65 : 65) **a**; система IV – (0 : 5, 5 : 5, 65 : 65) **a**, (0 : 5, 5 : 5, 35 : 65) **b**.

характер в хорошем растворителе, то это может сказаться на адсорбционном взаимодействии переставленных звеньев. Однако точность хроматографического эксперимента вряд ли достаточна для обнаружения такого различия.

В гетерополимере для переставленных X и Y соседями справа и слева оказываются разные звенья. В результате из-за химического эффекта соседа энергия адсорбции звеньев X и Y в новых местах может оказаться различной. Соответственно после такой перестановки средняя энергия адсорбции звеньев цепи изменяется, и перестановка может быть учтена в приближении *среднего гомополимера* такой же длины с некоторой *средней энергией адсорбции звена*. Разделение пептидов в этом случае можно трактовать как разделение смеси гомополимеров с разной энергией адсорбции звеньев. Такая составляющая

влияния последовательности, которую мы называем химической, не универсальна и зависит от конкретной реализации текста последовательности.

Вместе с тем существует другая, универсальная составляющая влияния последовательности, которую мы называем физической: даже если энергия адсорбции переставляемых групп в цепи не меняется в зависимости от соседей и никак не проявляется в приближении среднего гомополимера, взаимодействие всей гетерополимерной цепи с поверхностью при перестановке оказывается различным. Для цепи случайных блужданий, лежащей в основе модели *BioLCCC*, математическим выражением этой зависимости является некоммутативность произведения переходных матриц, описывающего статистическую сумму гетерополимера вблизи поверхности.

Именно такая физическая составляющая влияния перестановки на адсорбционные свойства будет рассматриваться ниже. Какой фактор — химический или физический — доминирует, в общем случае зависит от последовательности конкретного пептида.

Для понимания дальнейшего кратко изложим содержание модели *BioLCCC* [16, 17]. Макромолекула представляется в виде цепи случайных блужданий на кубической решетке, задаваемой переходной матрицей  $W(\varepsilon_i)$ , зависящей от энергии адсорбции соответствующего аминокислотного остатка:

$$W(\varepsilon_i) = \begin{bmatrix} \frac{2}{3}e^{\varepsilon_i} & \frac{1}{6}e^{\varepsilon_i} & 0 & \dots & 0 & 0 \\ \frac{1}{6} & \frac{2}{3} & \frac{1}{6} & \dots & 0 & 0 \\ 0 & \frac{1}{6} & \frac{2}{3} & \dots & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & \frac{2}{3} & \frac{1}{6} \\ 0 & 0 & 0 & \dots & \frac{1}{6}e^{\varepsilon_i} & \frac{2}{3}e^{\varepsilon_i} \end{bmatrix}$$

Размер ячеек решетки равен размеру звена макромолекулы  $a$ , все другие величины размерности длины, например размер пор  $D$ , выражаются в единицах размера звена  $D/a$  и соответствует размерности переходной матрицы. Величины размерности энергии далее выражаются в единицах  $kT$ , при этом положительная энергия  $\varepsilon_i$  соответствует притяжению, а отрицательная — отталкиванию звена от поверхности. Статистическая сумма цепи в щелеобразной поре, моделирующей адсорбент, записывается в виде

$$Z = U^T \times W(0) \times \left[ \prod_i W(\varepsilon_i) \right] \times U, \quad (1)$$

где  $U$  — единичный вектор начала цепи, матрица  $W(0)$  с нулевой энергией соответствует концу цепи, вектор  $U^T$  суммирует все возможные конфигурации цепи в поре. Произведение берется по всем звеньям цепи пептида, включая  $N$ - ( $i = 1$ ) и  $C$ - ( $i = n$ ) концы. В такой модели предполагается, что с поверхностью взаимодействуют стыки звеньев, играющие роль аминокислотных остатков. Для дальнейшего заметим, что вектор  $U'$ ,

$$U' = U^T \times W(0) = [5/6 \ 1 \ 1 \ \dots \ 5/6],$$

является левым собственным вектором переходной матрицы в критической точке адсорбции.

Статсумма, отнесенная к единице объема, представляет собой коэффициент распределения макромолекул между подвижной и неподвижной фазами  $K_d = Z/D$ , задающий объем удерживания пептида  $V_R$  через объем пор  $V_p$  и объем межчастичного пространства в колонке  $V_o$ :

$$V_R = V_o + V_p \cdot K_d.$$

В свою очередь энергия звеньев  $\varepsilon_i$  зависит от состава бинарного растворителя; в приближении Лэнгмюра эта зависимость имеет вид

$$\varepsilon_i = \varepsilon_i^o - \varepsilon_{ab}(N_b) = \varepsilon_i^o - \ln(N_b e^{\varepsilon_b} + 1 - N_b), \quad (2)$$

где  $N_b$  — мольная доля растворителя  $b$  (ацетонитрила в нашем случае). Мольная доля естественным образом связана с изменяющейся в соответствии с градиентом объемной долей  $V_b$  компонента  $b$  через плотность и молекулярную массу. Величину  $\varepsilon_i^o$  можно понимать как энергию адсорбции аминокислотного остатка в чистом растворителе  $a$  (в нашем случае в воде). Данные энергии, а также энергия адсорбции ацетонитрила  $\varepsilon_b = 2.49$  найдены нами ранее по модельным пептидам [16] и приведены в табл. 2.

Приближение Ленгмюра для учета влияния растворителя на адсорбционные свойства звеньев не является оптимальным для гидрофобных взаимодействий звеньев на обращенной фазе типа  $S_{18}$ , которая обычно используется для разделения пептидов. Однако для длинных, наиболее информативных в плане идентификации белков, последовательностей  $n \sim 20$  переход из адсорбированного состояния в раствор осуществляется в узком диапазоне составов растворителя. Зависимость энергии адсорбции от состава растворителя в узком диапазоне, по сути, линейна, так что применение любого корреляционного подхода, в том числе и приближения Ленгмюра, оправдано. Зависимость энергии звеньев от состава  $V_b$  дает возможность описать разделение макромолекул в градиенте любого профиля, вводя при этом в модель единственный дополнительный параметр — энергию адсорбции компонента  $b$ . Заметим, что по форме зависимость энергии адсорбции звена от состава растворителя совпадает с корреляционной теорией Снайдера [22], развитой для низкомолекулярных соединений. Энергии адсорбции в табл. 2 представляют собой не что иное, как изменение свободной энергии в системе макромолекула—поверхность—растворитель при замене контакта растворитель—поверхность на контакт звено цепи—поверхность.

Введем понятие *среднего гомополимера* и определим среднюю энергию его звена  $\bar{\varepsilon}$  следующим образом:

$$e^{\bar{\varepsilon}} = \frac{\sum e^{\varepsilon_i}}{n}, \quad (3)$$

( $n$  – длина последовательности). Такое среднее соответствует приближению макромолекулы системой разорванных звеньев, локализованных в размере клубка  $R$ .

Используя средний гомополимер, представим, следуя работе [23], переходную матрицу в виде

$$W(\varepsilon_i) = \begin{bmatrix} \frac{2}{3}(e^{\varepsilon_i} + e^{\bar{\varepsilon}} - e^{\bar{\varepsilon}}) & \dots & 0 \\ \frac{1}{6} & \dots & 0 \\ \dots & \dots & \dots \\ 0 & \dots & \frac{2}{3}(e^{\varepsilon_i} + e^{\bar{\varepsilon}} - e^{\bar{\varepsilon}}) \end{bmatrix} =$$

$$= W(\bar{\varepsilon}) + (e^{\varepsilon_i - \bar{\varepsilon}} - 1) \cdot \Delta W(\bar{\varepsilon}) = W(\bar{\varepsilon}) + \beta_i \cdot \Delta W(\bar{\varepsilon}),$$

где  $\Delta W$  – поправочная матрица, в которой все элементы равны нулю, за исключением первой и последней строк, описывающих попадание звеньев на поверхность.

$$\Delta W(\bar{\varepsilon}) = \begin{bmatrix} \frac{2}{3}e^{\bar{\varepsilon}} & \frac{1}{6}e^{\bar{\varepsilon}} & \dots & 0 & 0 \\ 0 & 0 & \dots & 0 & 0 \\ 0 & 0 & \dots & 0 & 0 \\ \dots & \dots & \dots & 0 & \dots \\ 0 & 0 & \dots & \frac{1}{6}e^{\bar{\varepsilon}} & \frac{2}{3}e^{\bar{\varepsilon}} \end{bmatrix}$$

Параметр  $\beta_i = (e^{\varepsilon_i - \bar{\varepsilon}} - 1)$  учитывает отличие энергии адсорбции  $i$ -го звена от средней величины.

Раскрывая произведение (1), записанное через матрицы среднего гомополимера  $W = W(\bar{\varepsilon})$  и  $\Delta W$ , для статистической суммы получим выражение

$$Z = Z(\bar{\varepsilon}) + \sum_i \beta_i U' W^{i-1} \Delta W W^{n-i} U +$$

$$+ \sum_{i < k} \beta_i \beta_k U' W^{i-1} \Delta W W^{n-k-1} \Delta W W^{n-k} U + \dots$$

**Таблица 2.** Эффективные энергии аминокислотных остатков в модели *BioLCCC*, определенные экспериментально по последовательностям *G(XX)LLLKK* [2] с разными концевыми группами Ac-, NH<sub>2</sub>-, -OH, -amide

Аминокислота	$\varepsilon^o$	Аминокислота	$\varepsilon^o$
К	0.266	С	1.296
Н	0.386	У	1.686
Р	0.516	В	1.751
Н	0.614	М	1.822
Г	0.656	И	2.156
С	0.698	Л	2.298
Q	0.746	F	2.319
D	0.781	W	2.436
T	0.893	NH <sub>2</sub>	-1.69
E	0.984	OH	-0.03
A	1.143	Ac	0
P	1.143	amide	0

Представляя  $\Delta W$  в виде

$$\Delta W = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \\ 0 \end{bmatrix} \times \begin{bmatrix} \frac{2}{3}e^{\bar{\varepsilon}} & \frac{1}{6}e^{\bar{\varepsilon}} & \dots & 0 & 0 \end{bmatrix} +$$

$$+ \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix} \times \begin{bmatrix} 0 & 0 & \dots & \frac{1}{6}e^{\bar{\varepsilon}} & \frac{2}{3}e^{\bar{\varepsilon}} \end{bmatrix},$$

легко заметить, что статсумма гетерополимера записывается как статсумма среднего гомополимера  $Z(\bar{\varepsilon})$ , «исправленная» вкладом точечных одно-, двух- ...  $n$ -частичных взаимодействий разных звеньев с поверхностью. При этом статсуммы хвостов  $T$  и петель  $L$ , соединяющих звенья  $n$ -частичной конфигурации взаимодействия цепи с поверхностью, выражаются через таковые для среднего гомополимера. Статсуммы левого  $T'$  и правого  $T$  хвостов, начинающихся на поверхности и заканчивающихся внутри поры, имеют вид

$$T_i' = [5/6 \ 1 \ 1 \ \dots \ 5/6] \times W^{i-1} \times \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \\ 0 \end{bmatrix}$$

и

$$T_{n-i} = e^{\bar{\varepsilon}} \times [2/3 \ 1/6 \ 0 \ \dots \ 0] \times W^{n-i} \times \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \\ 1 \end{bmatrix},$$

соответственно. Что касается петель, то для них возможен вариант, когда они связывают звенья  $i$  и  $k$ , взаимодействующие с одной поверхностью, петли  $L$ , или звенья, взаимодействующие с противоположными поверхностями, петли  $L'$ . Статсуммы таких петель записываются как

$$L_{k-i} = e^{\bar{\varepsilon}} \times [2/3 \ 1/6 \ 0 \ \dots \ 0] \times W^{k-i-1} \times \begin{bmatrix} 1 \\ 0 \\ 0 \\ \dots \\ 0 \end{bmatrix}$$

и

$$L'_{k-i} = e^{\bar{\varepsilon}} \times [2/3 \ 1/6 \ 0 \ \dots \ 0] \times W^{k-i-1} \times \begin{bmatrix} 0 \\ 0 \\ 0 \\ \dots \\ 1 \end{bmatrix}.$$

Реальная практика протеомных исследований такова, что пептиды с длиной  $n \sim 30$  звеньев имеют размер  $R$ , меньший размера пор  $D \sim 100 \dots 300 \text{ \AA}$ , так что петлями  $L'$ , связывающими звенья на противоположных поверхностях, можно пренебречь. Пептиды, исследованные в настоящей работе, соответствуют указанному условию. В результате получим выражение, учитывающее независимое взаимодействие цепи одновременно только с одной из двух поверхностей щелеобразной поры:

$$Z = Z(\bar{\varepsilon}) + 2 \times \sum_i T_i' \beta_i T_{n-i} + 2 \times \sum_{i < k} \beta_i \beta_k T_i' L_{k-i} T_{n-k} + \dots$$

Заметим, что если мы имеем дело с набором пептидов, различающихся по аминокислотному

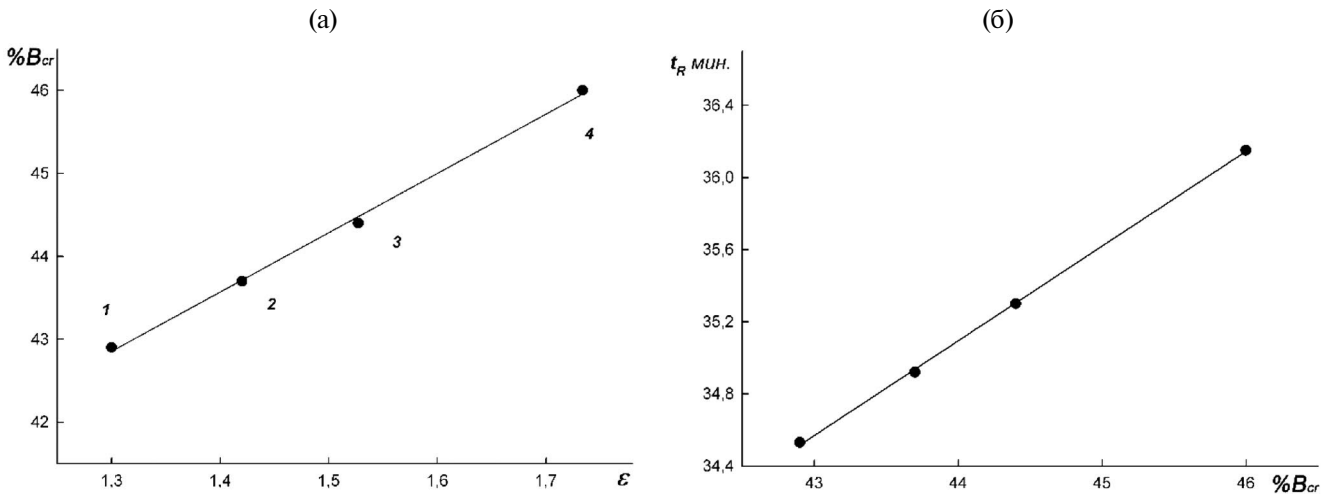
составу и, следовательно, по средней энергии звена, то приближение среднего гомополимера может оказаться вполне достаточным для предсказания их порядка выхода.

Выбор средней энергии звена в разложении статсуммы гетерополимера может быть произвольным. Однако средняя энергия звена в приближении системы разорванных звеньев (3) точнее коррелирует с критической точкой гетерополимера, чем, например, средняя арифметическая энергия звеньев.

Чтобы показать это, рассмотрим регулярный гетерополимер типа  $[AB]_n$ , который на больших масштабах при достаточной степени полимеризации  $n$  представляет собой не что иное, как гомополимер со средней энергией адсорбции звена. Для него, очевидно, существует критическая точка, определяемая нами как состав растворителя, в котором исчезает зависимость адсорбции от размера макромолекулы (или длины последовательности при  $n \gg 1$ ). Критическая энергия звена такого регулярного гетерополимера не является средним арифметическим энергий звеньев А и В. Одна и та же среднеарифметическая энергия звена, например  $\varepsilon = 1.3$ , может быть реализована различным заданием энергий звеньев А и В:  $\varepsilon(A, B) = (1.3, 1.3), (1.8, 0.8), (2.0, 0.6), (2.3, 0.3)$ . При этом рассчитанный по формулам (1)–(2) состав растворителя, соответствующий критической точке и исчезновению зависимости статсуммы от степени полимеризации, для разных реализаций такого гетерополимера будет различным.

Определенная по формуле (3) средняя экспоненциальная энергия звена, хотя также не соответствует критическим точкам регулярного гетерополимера, тем не менее, хорошо коррелирует в диапазоне энергий, типичных для аминокислотных остатков (табл. 2), с расчетным критическим составом растворителя для разных реализаций энергии звеньев А и В регулярного гетерополимера (рис. 2а).

Наконец, отметим, что в стандартном градиенте рассчитанный по формулам (1)–(2) объем удерживания последовательностей  $[AB]_n$  определяется их критическими точками (рис. 2б). Следовательно, при вариации последовательности можно рассмотреть различия в адсорбционном взаимодействии и соответствующее изменение порядка выхода пептидов в приближении среднего гомополимера, находящегося в критической точке. Аналогичный вывод справедлив и для другого предельного случая – моделирования пептида жестким стержнем [20]. Для полноты картины заметим, что любое отличие последовательности  $[AB]_n$  от регулярной, например, случайное распределение звеньев А и В по цепи, приводит к увеличению адсорбционного взаимодействия.



**Рис. 2.** Корреляция между расчетным критическим составом растворителя  $B_{cr}$  и средней энергии звена  $\epsilon$  по уравнению (3) для регулярного гетерополимера с разной энергией звеньев (а); корреляция между расчетным временем выхода  $t_R$  и составом  $B_{cr}$ , соответствующим критической точке бинарного гетерополимера в стандартном градиенте ( $t, \text{ мин} : B, \%$ ): (0:0, 100:100) на колонке  $4 \times 250 \text{ мм}$  (б). Пояснения в тексте.

Включение влияния растворителя в энергию звена среднего гомополимера и выбор состава так, чтобы энергия звена была равна критической для цепи блужданий на кубической решетке

$$\bar{\epsilon} = \epsilon_c = \ln \frac{6}{5},$$

заметно упрощает выражение для  $Z$ , и  $K_d$  гетерополимера приобретает вид

$$K_d \approx 1 + \frac{2}{D} \frac{5}{6} \cdot \sum_i \beta_i + \frac{2}{D} \frac{5}{6} \cdot \sum_{i < k} \beta_i \beta_k L_{k-i} + \frac{2}{D} \frac{5}{6} \cdot \sum_{i < k < m} \beta_i \beta_k \beta_m L_{k-i} L_{m-k} \dots \quad (4)$$

Статсуммы хвостов  $T'$  и  $T$  в критической точке равны  $5/6$  и единице соответственно, независимо от их длины. Статсумма петли, связывающей звенья  $k$  и  $i$  на поверхности, определяется как

$$L_{k-i} = \begin{bmatrix} 1 \\ 0 \\ 0 \\ \dots \\ 0 \end{bmatrix} \times W(\epsilon_c)^{k-i-1} \times \begin{bmatrix} 1 \\ 0 \\ 0 \\ \dots \\ 0 \end{bmatrix} \quad (5)$$

Выбор средней энергии в виде (3) приводит к тому, что линейный по  $b$  член в выражении (4) тождественно равен нулю, и различие в объемах удерживания при вариации последовательности обусловлено различием в двухчастичных (будем

называть их далее для краткости бинарными) взаимодействиях звеньев с поверхностью. Выражение (4) можно рассматривать как разложение статсуммы макромолекулы вблизи критической точки адсорбции по параметру  $\beta = (e^{\epsilon - \bar{\epsilon}} - 1)$ . Параметр  $b$  не зависит от состава растворителя и полностью определяется последовательностью. Он может быть не мал для конкретной реализации последовательности, так что необходимо учитывать все члены ряда (это автоматически учитывается произведением переходных матриц). Заметим, что взаимодействия с поверхностью более высоких порядков для цепи блужданий выражаются через бинарные. Ниже для оценок при выяснении физической причины в различии адсорбционных свойств инвертированных последовательностей мы ограничимся бинарными взаимодействиями.

Разложение статсуммы вблизи критической точки не ограничивает общность рассмотрения: как следует из рис. 2б в градиенте объем удерживания гетерополимера однозначно определяется критической точкой. Можно сказать, что в градиенте в длинной колонке гомополимеры в смеси распределяется по составам растворителя, близким к их критическим точкам, после чего они движутся с одинаковой скоростью, близкой к скорости движения градиента. Таким образом, рассмотрение взаимодействия в критической точке для определенного выше среднего гомополимера справедливо, по крайней мере качественно, и для окрестности критической точки, т.е. для состава растворителя, в котором макромолекула выходит из колонки.

Статсумма (5) петли, связывающей звенья  $i$  и  $k$  на поверхности, важная для дальнейшего рассмотрения, в критической точке при увеличении



длины цепи быстро становится близкой к таковой для гауссовой цепи, в чем можно убедиться прямым вычислением:

$$L_{k-i} \approx \frac{C}{\sqrt{|k-i|}}$$

где  $C$  – константа порядка единицы, а  $|k-i|$  – расстояние по цепи между звеньями. Конфигурация гауссовой цепи в критической точке соответствует конфигурации цепи в неограниченном объеме, и статсумма петли, начинающейся и заканчивающейся на поверхности, равна [24]

$$L_{k-i} = \sqrt{\frac{6}{\pi|k-i|}}$$

Для оценки влияния вариации последовательности можно использовать гауссово приближение и тем самым получить аналитические выражения для разницы времен удерживания.

Возвращаясь к элементарной вариации – перестановке групп  $X$  и  $Y$ , для разницы в объемах удерживания  $\Delta V_R$  в приближении бинарных взаимодействий получаем выражение

$$\Delta V_R \sim \Delta Z \sim (\beta(X) - \beta(Y)) \cdot \sum_i \left[ \beta(i) \left( \frac{1}{\sqrt{|i-i_X|}} - \frac{1}{\sqrt{|i-i_Y|}} \right) \right] \quad (6)$$

В формуле (6) опущены численные коэффициенты. Ниже нас будет в основном интересовать знак различия при перевороте последовательности, так как абсолютное изменение объема удерживания при такой вариации, как правило, незначительно. Изменение объема удерживания пропорционально изменению времени удерживания, будем считать эти переменные эквивалентными.

Структуру уравнения (6) проще всего понять, если воспользоваться диаграммным представлением бинарных взаимодействий (рис. 3). Большинство диаграмм в разнице объемов удерживания (6) взаимно сокращаются, остаются лишь те, которые связывают переставляемые группы  $X$  и  $Y$  со звеном  $i$  в цепи; различие между ними связано с разной длиной петель. Если энергии переставляемых групп отличаются мало, то множитель перед суммой в выражении (6) мал и  $\Delta V_R$  незначительно. Также различие мало, если профиль энергии вдоль цепи  $b(i)$  имеет определенную симметрию относительно переставляемых групп. Например, если переставляются разные концевые группы, а профиль энергии симметричен относительно центра цепи, то различие в объемах удерживания близко к нулю.

## РЕЗУЛЬТАТЫ И ИХ ОБСУЖДЕНИЕ

С учетом сказанного выше рассмотрим разделение инвертированных последовательностей. Времена удерживания пептидов из табл. 1 на разных системах приведены в табл. 3. Заметим, что в инвертированных последовательностях соседи звеньев не меняются, так что упомянутый выше химический фактор влияния, в отличие от перестановок, минимален. На исследованных системах I–III, в которых pH 2.0, наблюдается разделение инвертированных последовательностей, причем знак различия времен удерживания  $\Delta t_R$  для всех систем одинаков.

Рассмотрим далее вопрос о согласованности времен удерживания одних и тех же пептидов в разных градиентах (рис. 4). Такое сравнение важно: в протеомных исследованиях научные коллаборации используют разные протоколы разделения пептидов, и для приведения в базы данных времен удерживания к единой временной шкале часто используется линейное преобразование. Кроме того, в разных градиентах, даже на одной и той же системе, порядок выхода некоторых пептидов может меняться. На рис. 4а приведено сравнение экспериментального времени удерживания  $t_R$  исследованных пептидов в градиентах  $b$ ,  $c$  и  $t_R$  тех же пептидов в градиенте  $a$  на системе I.

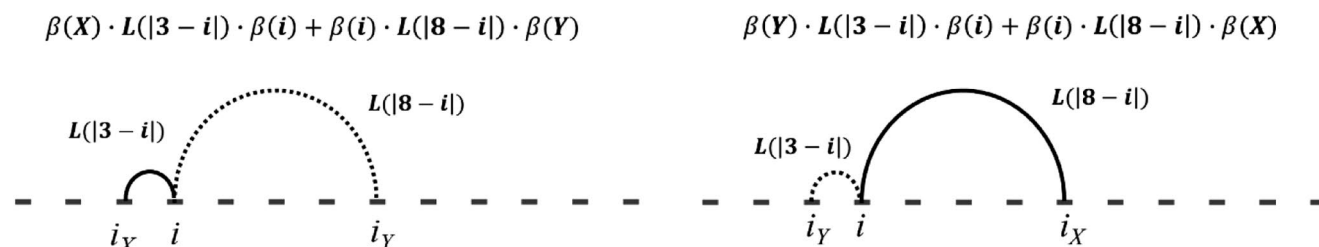


Рис. 3. Диаграммное представление несократимых бинарных взаимодействий для исходной последовательности и последовательности с переставленными звеньями  $X$  и  $Y$ .  $L$  – петли, связывающие переставляемые звенья  $X$  ( $i = 3$ ) и  $Y$  ( $i = 8$ ) со звеном  $i$ . Для получения уравнения (5) необходимо просуммировать разность по всем звеньям цепи  $i$ .

Как видно, на этой системе наблюдается линейное соответствие времен удерживания в разных градиентах, и для определения коэффициентов линейного преобразования достаточно небольшого набора пептидов с известными последовательностями. Линейная зависимость имеет место и для градиентов системы IV (рис. 4б).

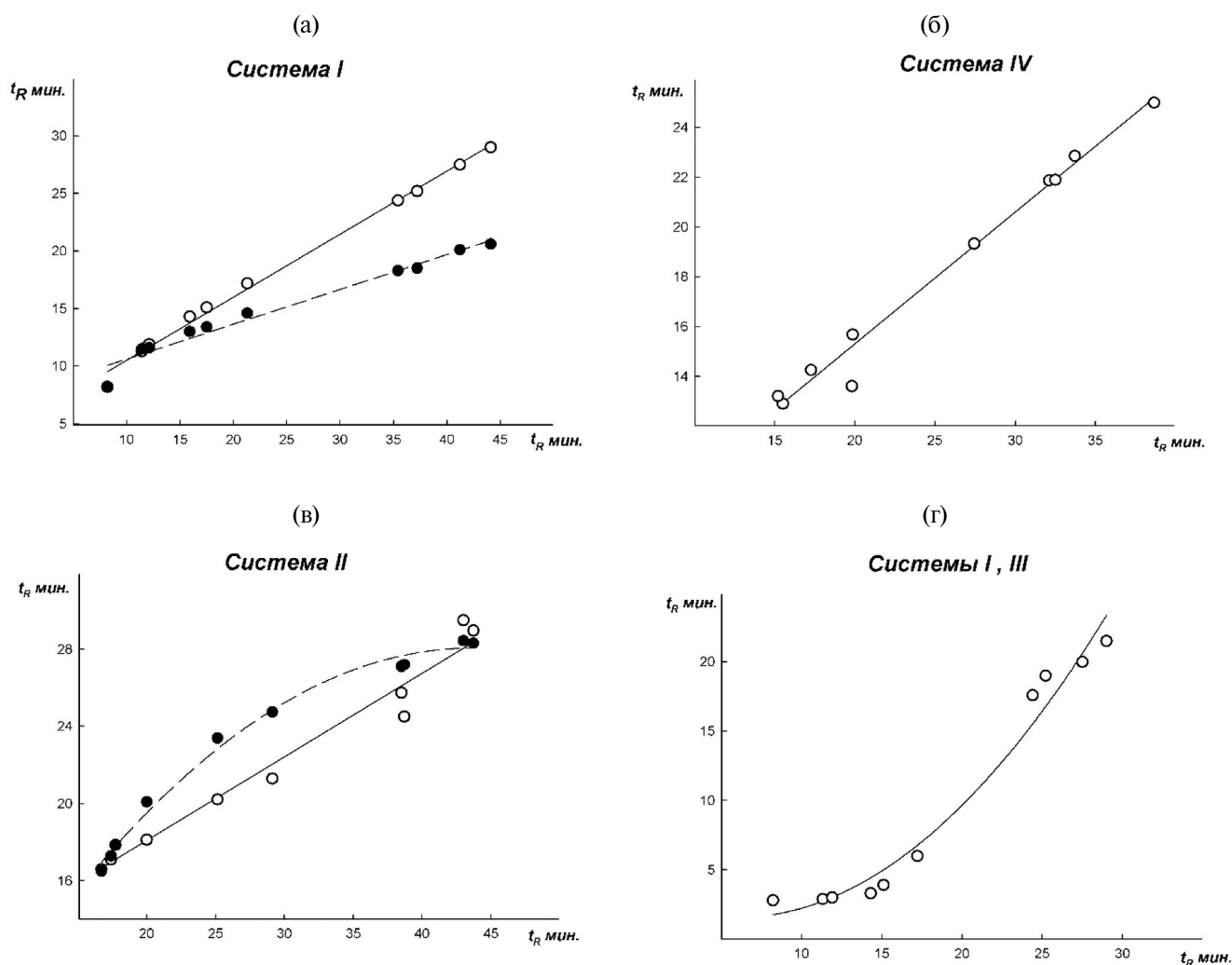
Однако аналогичное сравнение  $t_R$  пептидов в градиентах **b**, **c** и их  $t_R$  в градиенте **a**, для системы II демонстрирует нелинейность на одной и той же колонке (рис. 4в). Нелинейность наблюдается и при сравнении  $t_R$  для одинаковых во временной шкале градиента **b** в системе I и градиента **a** в системе III (рис. 4г). Таким образом, даже для рассматриваемой нами небольшой выборки пептидов связь времен удерживания в разных градиентах на одной колонке может быть нелинейной. Такая нелинейность, на наш взгляд, затрудняет выравнивание времен удерживания в единой шкале, поскольку для ее корректного учета необходим гораздо больший набор пептидов с известными последовательностями. По этой причине при выравнивании времен удерживания, полученных в разных протоколах разделения, необходимо убедиться в их линейности. Теоретический расчет времен удерживания в модели *BioLCCC* по уравнениям (1)–(2) предсказывает такую нелинейную зависимость для разных градиентов системы II, и для проверки линейности в разных протоколах разделения при планировании эксперимента можно воспользоваться предварительным теоретическим расчетом.

Сравним далее экспериментальные времена удерживания с рассчитанными по модели *BioLCCC* (рис. 5). Как видно, несмотря на простоту модели, результаты расчета хорошо коррелируют с экспериментом и, что на наш взгляд важнее, правильно предсказывают знак различия времен удерживания инвертированных последовательностей (табл. 3).

Чтобы разобраться в физической причине различий в объемах удерживания инвертированных последовательностей, рассмотрим более детально профиль энергии взаимодействия (рис. 6). Следует отметить, что при рассмотрении взаимодействия биомакромолекул с поверхностью необходимо проводить анализ текста последовательности не в пространстве букв, а в пространстве энергий взаимодействия звеньев. При этом может оказаться, что в пространстве букв никакой периодичности не наблюдается, но профиль энергии взаимодействия демонстрирует определенную периодичность или заметную асимметрию от N- к C-концу. На рис. 6 величина  $\Delta\epsilon_i$  представляет собой разницу между энергией взаимодействия звена  $\epsilon_i$  и средней энергией  $\bar{\epsilon}$  вдоль последовательности от N- к C-концу (отрицательная

Таблица 3. Экспериментальные времена удерживания инвертированных пептидов в разных системах разделения и разных pH в сравнении с расчетными

Пептид, №	$t_R$ эксп / $t_R$ теор, мин						$t_R$ расчет, мин					
	Система I (pH 2.0)			Система II (pH 2.0)			Система III (pH 2.0)	Система IV (pH 8.0)		Skyline SSRCalc [25]		
	a	b	c	a	b	c		a	b			
1	8.20/11.57	8.20/10.65	8.20/9.61	16.71/16.80	16.60/16.10	16.49/16.80	2.80/0.70	19.80	13.60	4.09	47.08	12.90
1'	11.40/14.16	11.30/12.50	11.50/10.80	17.73/17.98	17.85/16.84	17.84/17.97	2.90/0.90	32.10	21.87	5.54	55.29	15.49
2	15.90/16.57	14.30/13.98	13.00/11.67	20.00/19.32	18.12/17.52	20.08/19.11	3.30/1.20	15.50	12.90	9.73	44.01	15.38
2'	12.10/12.91	11.90/11.52	11.60/10.12	17.40/17.40	17.10/16.40	17.28/17.40	3.00/0.80	15.20	13.20	6.39	83.61	12.82
3	35.40/42.56	24.40/28.18	18.30/19.28	38.70/39.59	24.51/28.99	27.20/26.81	17.60/11.50	33.70	22.85	14.60	162.74	34.03
3'	41.20/54.13	27.50/34.77	20.10/23.04	43.72/48.16	28.96/34.57	28.31/29.64	20.00/15.60	27.43	19.33	10.60	179.63	34.48
4	21.30/26.24	17.20/19.89	14.60/15.13	29.13/25.34	21.29/21.10	24.74/22.43	6.00/2.70	17.25	14.25	22.99	71.67	21.22
4'	17.50/22.48	15.10/17.60	13.40/13.78	25.14/22.93	20.21/19.61	23.39/21.27	3.90/2.00	19.85	15.67	26.66	67.88	17.85
5	37.20/48.39	25.20//30.03	18.50/20.16	38.49/43.12	25.74/31.05	27.11/27.71	19.00/13.70	38.65	25.00	25.50	143.07	31.58
5'	44.10/55.50	29.00/35.26	20.60/23.20	42.99/49.94	29.50/35.48	28.44/29.93	21.50/17.00	32.50	21.90	29.96	162.18	35.49



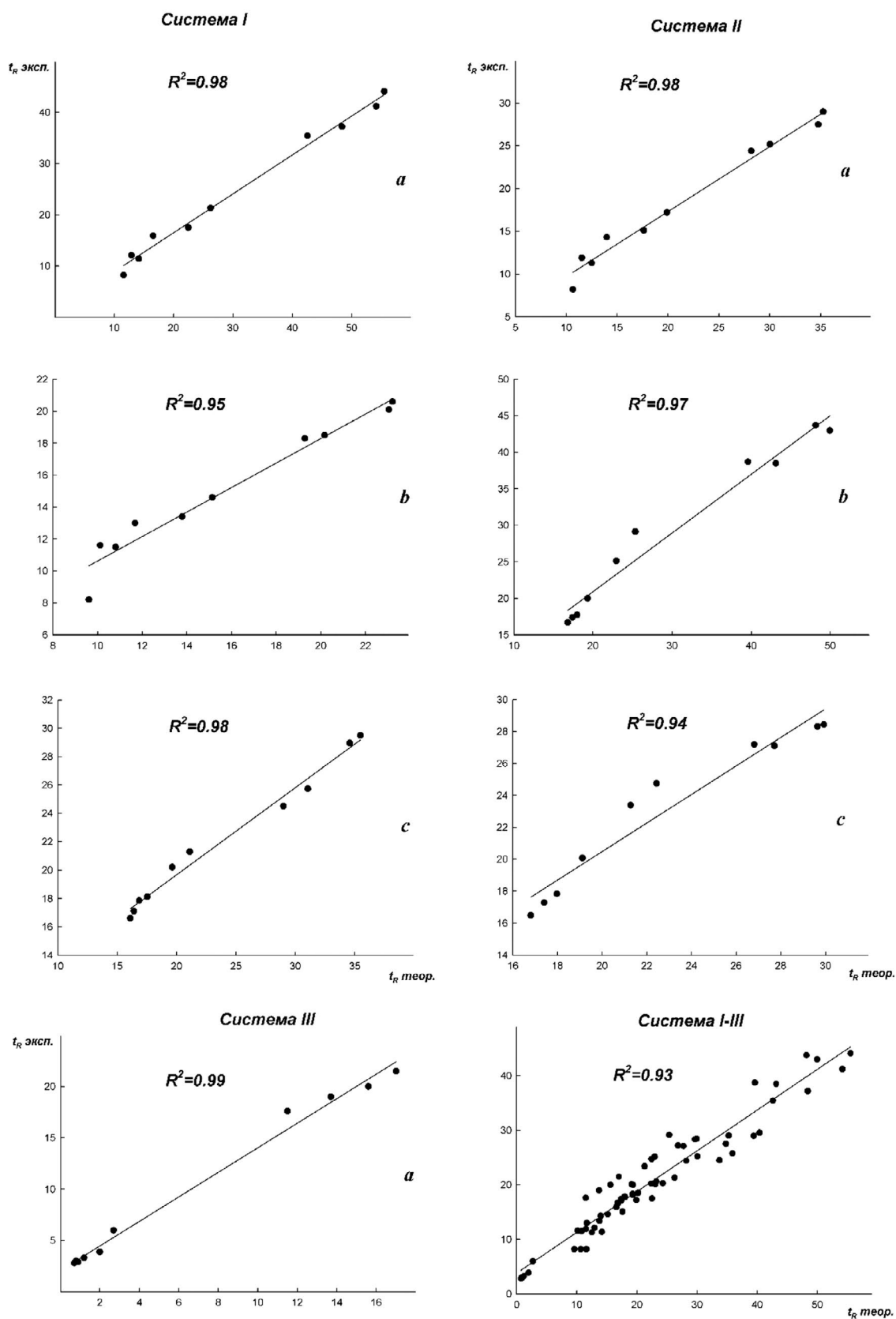
**Рис. 4.** Сравнение экспериментальных времен удерживания  $t_R$  пептидов в разных градиентах на разных системах: а – система I, градиент *b* (светлые точки) и градиент *c* (темные), абсцисса, относительно градиента *a*, ордината; б – система IV, градиент *b* относительно градиента *a*; в – система II, градиент *b* (светлые точки) и градиент *c* (темные) относительно градиента *a*; г – градиент *a* система III относительно градиента *b* система I.

величина соответствует отталкиванию звена от поверхности, положительная – притяжению).

Вид профиля энергии на рис. 6 позволяет качественно понять знак различия во взаимодействии при инверсии последовательности. Действительно, в условиях разделения при pH 2.0 N-конец оказывается положительно заряженным,  $\text{NH}_3^+$ . Заряженность звена вне зависимости от знака означает его сильное отталкивание от гидрофобной поверхности в водном растворе. Причина этого в том, что «изображение» заряда в границе вода–гидрофобная поверхность оказывается того же знака (см. подробнее в работе [25] для аналогичного вопроса о взаимодействии заряженных аминокислот с гидрофобным ядром глобулярных белков). Такое отталкивание приводит к тому, что N-конец оказывается удаленным от поверхности, как и близкие связанные с ним звенья. Это

уменьшает вероятность взаимодействия близких к N-концу звеньев с поверхностью. Следовательно, если адсорбирующиеся (гидрофобные) звенья располагаются в цепи вблизи N-конца, то при перевороте последовательности они окажутся вблизи гораздо слабее отталкивающегося C-конца, вероятность их взаимодействия с поверхностью возрастет и адсорбционное взаимодействие всей цепи увеличится. Знак различия во временах удерживания инвертированных последовательностей и профили энергии на рис. 6 полностью соответствуют этим качественным рассуждениям.

Рассмотрим данный вопрос количественно, для чего найдем во втором порядке разложения разницу в объемах удерживания инвертированных последовательностей в соответствии с выражением (6). Инверсию последовательности можно интерпретировать как перестановку конечных



**Рис. 5.** Сравнение расчетных  $t_{R\text{теор}}$  и экспериментальных  $t_{R\text{эксп}}$  времен удерживания (табл. 3) для разных градиентов для системы I (a, b, c), II (a, b, c) и III (a), а также общая корреляция для систем I – III, pH 2.0.

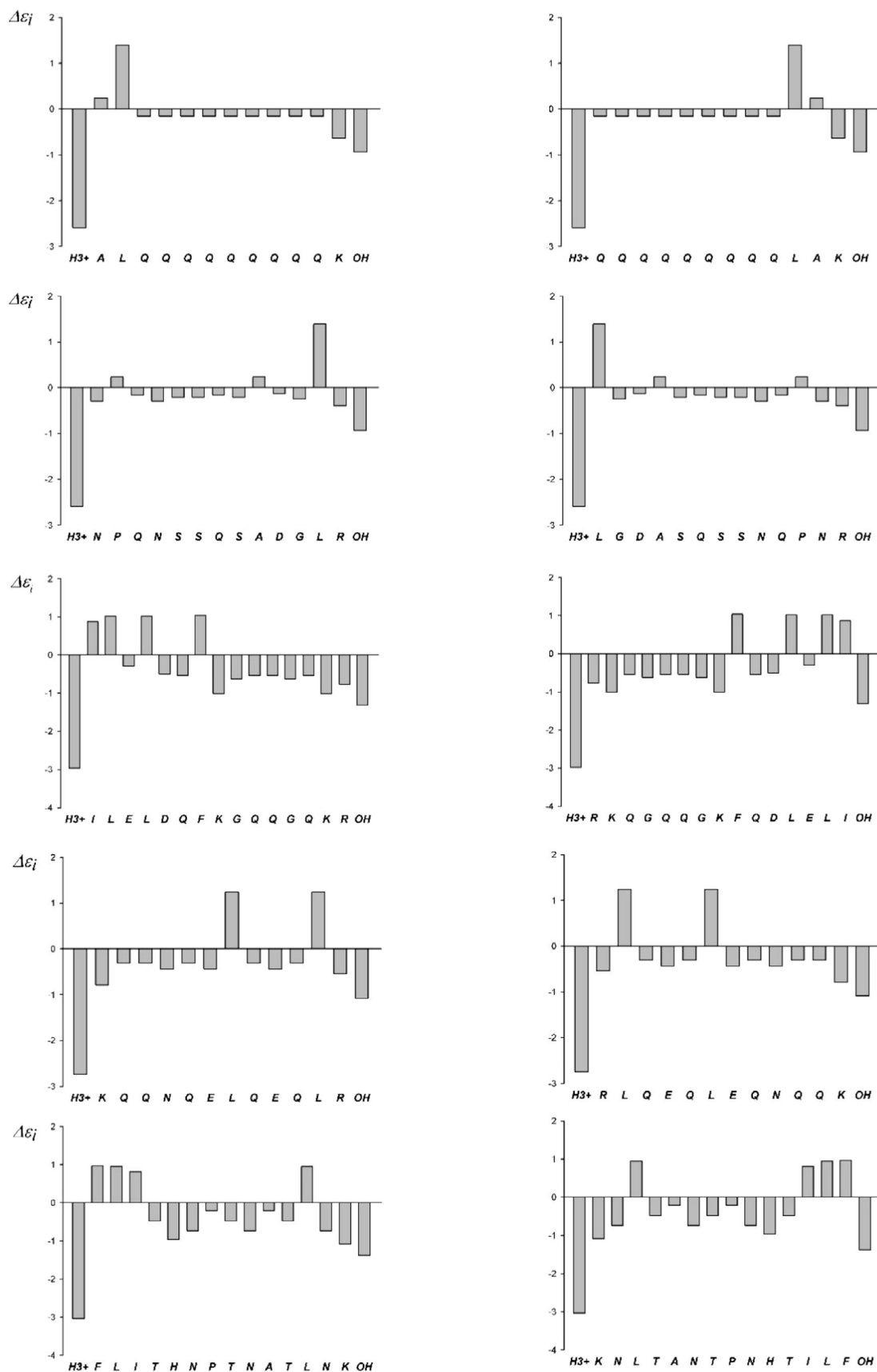
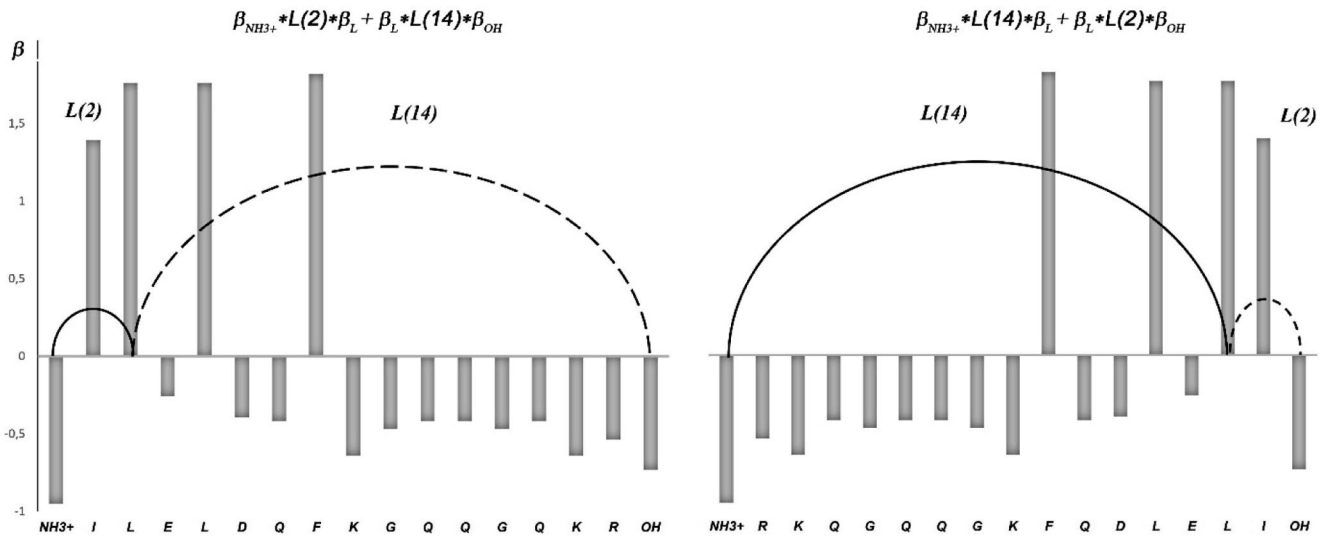


Рис. 6. Профиль энергии взаимодействия звеньев от N- к С-концу для последовательностей в табл. 1.



**Рис. 7.** Диаграммное представление бинарных взаимодействий при инвертировании последовательности 3 из табл. 1. Для примера показаны несократимые диаграммы, связывающие аминокислотный остаток *L* с концевыми группами; для получения разницы необходимо просуммировать по всем остальным остаткам между концами цепи.

групп. Несократимые диаграммы, которые связывают концевые группы с остатком *i* в цепи, приведены на рис. 7 для одной из последовательностей табл.1. Суммирование по всем звеньям между N-и C-концами приводит к выражению

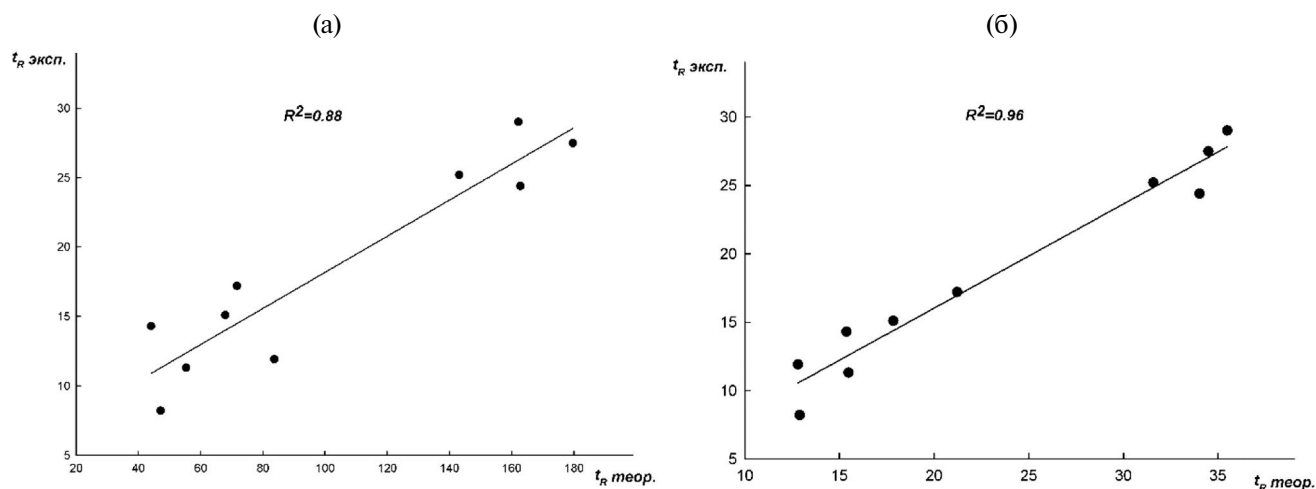
$$\Delta t_R \sim \Delta V_R \sim (\beta(NH_3) - \beta(OH)) \cdot \sum_i \left[ \beta(i) \cdot \left( \frac{1}{\sqrt{|i-1|}} - \frac{1}{\sqrt{|i-n|}} \right) \right]. \quad (7)$$

Как видно, различие объемов удерживания при инвертировании последовательности связано с различием в длине петель, связывающих звено *i* с концевыми группами.

Расчет  $DV_R$  по уравнению (7) в приближении бинарных взаимодействий дает правильный знак изменения адсорбционного взаимодействия для пептидов 3, 4 и 5 и их порядка выхода, соответствующего табл. 3. Для последовательностей 1 и 2 после инвертирования последовательности необходимо переместить звенья *K* и *R* к C-концу, например, для последовательности *I'* путем серии перестановок  $2K \ll 11Q, 11K \ll 12L, 12K \ll 13A$ . Учесть влияние таких перестановок на изменение объема удерживания можно с помощью диаграмм, аналогичных рис. 3, и уравнения (6), в котором задействованы петли, связывающие звенья цепи с переставляемыми. В результате для пептидов 1 и 2 в приближении бинарных взаимодействий также получается правильный знак изменения объема удерживания  $DV_R$ .

Рассмотрим качественно разделение инвертированных последовательностей в системе *IV* при pH 8.0. В этих условиях N-конец нейтрализуется, и его отталкивание от гидрофобной поверхности резко уменьшается, а C-конец заряжается отрицательно и начинает отталкиваться от поверхности, уменьшая вероятность взаимодействия с поверхностью связанных с ним звеньев. Таким образом, можно предположить, что при pH 8.0 знак разности времен удерживания при инверсии последовательности должен измениться. Теперь, если гидрофобные адсорбирующиеся звенья находятся вблизи C-конца, то вероятность их взаимодействия с поверхностью должна уменьшиться. В целом данная тенденция в изменении знака различия при pH 8.0 наблюдается для пептидов 3, 4, 5 (табл. 3). Однако для пептидов 1, 2 такая перемена знака  $DV_R$  не наблюдается. Для выяснения этого вопроса необходимо переопределить энергии взаимодействия аминокислотных остатков при pH 8.0 по стандартным последовательностям, так как изменение pH влияет на адсорбционное взаимодействие не только концевых групп. Мы предполагаем, что при pH 8.0 степень заряженности C-конца не столь высокая, чтобы заметно изменить характер взаимодействия связанных с ним звеньев и, следовательно, порядок выхода пептидов.

Расчет влияния инверсии последовательности в рамках модифицированной аддитивной модели для систем с pH 2.0, выполненный нами в программе Skyline Proteomics [26], в которую встроен калькулятор времен удерживания пептидов SSR-Calcul [6, 7], также дает правильный знак различия объемов удерживания, за исключением последовательностей 3 и 4 (табл. 3). Поправки, вводимые



**Рис. 8.** Корреляция между экспериментальными  $t_{R\text{эксп}}$  и рассчитанными в программе DeepLC  $t_{R\text{теор}}$  временами удерживания инвертированных пептидов: а – тренировочный набор из 1366 пептидов в неизвестном градиенте, б – тот же набор в градиенте **b** системы I с теоретическими по модели *BioLCCC* временами удерживания.

в аддитивную модель в программе SSRCalc, основаны на влиянии заряженного N-конца на адсорбционное взаимодействие ближайших к этому концу аминокислотных остатков, так что данные поправки должны давать правильную картину. По сути, это альтернативный способ учета связанности звеньев в цепь вблизи заряженной группы, аналогичный качественному рассмотрению влияния инверсии на рис. 6. Расхождения с экспериментом для последовательностей 3 и 4 могут быть связаны с тем, что мы использовали некоммерческий вариант программы, возможно, не оптимальный для исследованных в работе градиентов. В отличие от работы [6, 7] модель *BioLCCC* дает правильный знак различия в удерживании инвертированных последовательностей для всех градиентов, не вводя никаких дополнительных поправочных коэффициентов и оставаясь в рамках минимального возможного для описания разделения пептидов числа (20) энергий адсорбции аминокислотных остатков.

Сравним также экспериментальные результаты с расчетом по модели Deep LC [12], используемой для предсказания времен удерживания вариант обучаемой нейронной сети. В версии 2.2.26. модели в качестве обучающей выборки первоначально был взят набор из 1366 известных последовательностей пептидов с соответствующими временами удерживания в некотором неизвестном градиенте, DeepLC(1) (табл. 3). На рис. 8а показана корреляция между предсказанными с таким набором и экспериментальными временами удерживания для системы I **b**. Как видно, получающаяся при таком тренировочном наборе корреляция не слишком хорошая и знак различия объемов удерживания для некоторых последовательностей не соответствует эксперименту. Если в качестве обучающей выборки взять

те же 1366 последовательностей, но использовать для них времена удерживания, рассчитанные по модели *BioLCCC* для системы I **b**, то корреляция заметно улучшается (DeepLC(2), табл. 3, рис. 8б). Знак различия в объемах удерживания инвертированных последовательностей при этом становится соответствующим эксперименту.

Аналогичные результаты сравнения с DeepLC получаются и для других исследованных нами систем разделения. Тем самым можно сделать вывод о том, что в качестве обучающей выборки в программе DeepLC можно использовать расчётные величины времен удерживания. Это может оказаться полезным в том случае, когда по каким-то причинам отсутствует набор известных последовательностей и их времен удерживания в градиенте, в котором проводится разделение неизвестного набора пептидов.

## ЗАКЛЮЧЕНИЕ

Подводя итог, можно сказать, что модель *BioLCCC*, основанная на модели адсорбции цепи случайных блужданий, дополненная феноменологическими параметрами энергий адсорбции аминокислотных остатков, дает правильное понимание физической причины изменения адсорбционных свойств пептидов при инверсии последовательности. Как показано, причина заключается в различии статстумм петель, связывающих концевые группы с другими звеньями в цепи. Влияние концевых групп на адсорбционное взаимодействие какого-то аминокислотного остатка проявляется через вероятность столкновения этого остатка с поверхностью. Отталкивающаяся от гидрофобной поверхности заряженная концевая группа существенно уменьшает вероятность стол-

кновения близких к ней аминокислотных остатков с поверхностью, в то время как для удаленных по цепи звеньев её влияние быстро затухает.

В исследованных пептидах определяющим для изменения адсорбционных свойств при инверсии последовательности является, по-видимому, физический фактор. Инверсию последовательности можно трактовать как перестановку концевых групп, имеющих разную энергию взаимодействия с поверхностью. Если в гомополимере такая перестановка не приводит к изменению адсорбционных свойств всей цепи, то гетерогенность цепи пептида как бы усиливает воздействие перестановки на адсорбционные свойства. В исследованных пептидах переставлялись группы с заметно различающимися энергиями адсорбции, что приводило к заметному коэффициенту перед суммой в уравнениях (6) и (7). Однако в других случаях, когда переставляются звенья, не столь различные по своим адсорбционным свойствам, или цепь имеет определенную симметрию вдоль последовательности в пространстве энергий взаимодействия звеньев, эффект перестановки концевых групп или инверсии последовательности может быть не столь заметен и не соответствовать рассматриваемой модели. Для уточнения описания необходимо будет вводить коррекцию энергии взаимодействия звена в зависимости от соседей по цепи. Хотя идейно такую коррекцию в модель *BioLCCC* ввести не сложно, для экспериментального определения коэффициентов коррекции необходимо провести большую работу по синтезу соответствующих моделей.

Работа выполнена при финансовой поддержке Российского научного фонда (проект 20-14-00229-П).

#### СПИСОК ЛИТЕРАТУРЫ

1. *Ivanov M.V., Bubis Ju.A., Gorshkov V., Tarasova I.A., Levitsky L.I., Lobas A.A., Solovyeva E.M., Pridatchenko M.L., Kjeldsen F., Gorshkov M.V.* // *Anal. Chem.* 2020. V. 92. № 6. P. 4326.
2. *Guo D., Mant C.T., Taneja A.K., Parker J.M.R., Hodges R.S.* // *J. Chromatogr.* 1986. V. 359. P. 499.
3. *Mant C.T., Lorne Burke T.W., Black J.A., Hodges R.S.* // *J. Chromatogr.* 1988. V. 458. P. 193.
4. *Gilar M., Hie H.W., Jaworski A.* // *Anal. Chem.* 2010. V. 82. P. 265.
5. *Azarova I.N., Baram G.I., Gol'dberg E.L.* // *Russ. J. Bioorg. Chem.* 2006. V. 32. № 1. P. 50.
6. *Krokhin O.V., Craig R., Spicer V., Ens W., Standing K.G., Beavis R.C., Wilkins J.A.* // *Mol. Cell Proteom.* 2004. V. 3. № 9. P. 908.
7. *Krokhin O.V.* // *Anal. Chem.* 2006. V. 78. № 22. P. 7785.
8. *Reimer J., Spicer V., Krokhin O.V.* // *J. Chromatogr. A.* 2012. V. 1256. P.160.
9. *Palmblad M., Ramström M., Markides K. E., Håkansson P., Bergquist J.* // *Anal. Chem.* 2002. V. 74. P. 5826.
10. *Guan, S., Moran, M. F., Ma, B.* // *Mol. Cell. Proteom.* 2019. V. 18. P. 2099.
11. *Gessulat S., Schmidt T., Zolg D. P., Samaras P., Schnatbaum K., Zerweck J., Knaute T., Rechenberger J., Delanghe B., Huhmer A., Reimer U., Ehrlich H.-Ch., Aiche S., Kuster B., Wilhelm M.* // *Nat. Methods* 2019. V. 16. P. 509.
12. *Bouwmeester R., Gabriels R., Hulstaert N., Martens L., Degroeve S.* // *Nat. Methods* 2021. V. 18. P. 1363.
13. *Entelis S.G., Evreinov V.V., Gorshkov A.V.* // *Adv. Polym. Sci.* 1986. V. 76. P. 129.
14. *Entelis S.G., Evreinov V.V., Kuzaev A.I.* // *Reactive Oligomers.* Utrecht: VSP, 1989.
15. *Pasch H., Trathnigg B.* // *HPLC of Polymers.* Berlin: Springer, 1999.
16. *Gorshkov A.V., Tarasova I.A., Evreinov V.V., Savitski M.M., Nielsen M.L., Zubarev R.A., Gorshkov M.V.* // *Anal. Chem.* 2006. V. 78. № 22. P. 7770.
17. *Gorshkov A. V., Evreinov V. V., Tarasova I. A., Gorshkov M. V.* // *Polymer Science B.* 2007. V. 49. № 3—4. P. 93.
18. *Tarasova I. A., Gorshkov A. V., Evreinov V. V., Adams K., Zubarev R. A., Gorshkov M. V.* // *Polymer Science A.* 2008. V. 50. № 3. P. 309
19. *Tarasova I.A., Goloborodko A.A., Perlova T.Y., Pridatchenko M.L., Gorshkov A.V., Evreinov V.V., Ivanov A.R., Gorskov M.V.* // *Anal. Chem.* 2015. V. 87. P. 6562.
20. *Gorshkov A. V., Goloborod'ko A. A., Pridatchenko M. L., Tarasova I. A., Rozdina I. G., Evreinov V. V., Gorshkov M. V.* // *Polymer Science A.* 2017. V. 59. № 3. P. 354.
21. *Tarasova I.A., Goloborodko A.A., Perlova T.Y., Pridatchenko M.L., Gorshkov A.V., Evreinov V.V., Ivanov A.R., Gorshkov M.V.* // *Anal. Chem.* 2015 V. 87. P. 6562.
22. *Snyder L.R.* // *Principles of Adsorption Chromatography.* New York: M. Dekker, 1968.
23. *Обухов С. П.* // *Журн. эксперим. и теорет. физики.* 1987. Т 93. № 6. С. 1973.
24. *Eisenriegler E., Kremer K., Binder K.* // *J. Chem. Phys.* 1982. V. 77. P. 6296.
25. *Финкельштейн А. В., Птицын О. Б.* // *Физика белка.* М.: Книжный дом Университет, 2012.
26. *MacLean B., Tomazela D. M., Shulman N., Chambers M., Finney G. L., Frewen B., Kern R., Tabb D. L., Liebler D. C., MacCoss M. J.* // *Bioinformatics.* 2010. V. 26. I. 7. P. 966.