

УДК 541.64:543.544:547.963.32

О ПРИМЕНИМОСТИ КОНЦЕПЦИИ КРИТИЧЕСКОЙ ХРОМАТОГРАФИИ ДЛЯ ИССЛЕДОВАНИЯ ПОСЛЕДОВАТЕЛЬНОСТИ НУКЛЕОТИДОВ В ОДНОНИТЕВЫХ ДНК

© 2022 г. А. В. Горшков^{а,*}, И. Г. Роздина^а, В. В. Евреинов^а

^а Федеральный исследовательский центр химической физики им. Н.Н. Семенова Российской академии наук
119991, Москва, ул. Косыгина, 4, Россия

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

Поступила в редакцию 13.04.2021 г.

После доработки 15.09.2021 г.

Принята к публикации 25.10.2021 г.

Решеточная модель адсорбции макромолекул применена для описания разделения однонитевых ДНК в изократическом и градиентном режимах на обращенной фазе в смеси воды и ацетонитрила. Показано, что данная модель, дополненная корреляционными соотношениями зависимости эффективной энергии звена от состава растворителя, дает возможность с хорошей точностью описать закономерности хроматографического разделения как гомополимеров нуклеотидов, так и их гетерополимеров. Для гетерополимеров корреляция между расчетом и экспериментальными данными имеет меньшую точность, чем для гомополимеров, что связано с малым вкладом энтропийной составляющей адсорбционного взаимодействия при вариации последовательности. Обсуждаются варианты уточнения модели для учета эффекта соседних звеньев.

DOI: 10.31857/S2308112022010035

ВВЕДЕНИЕ

ДНК и РНК – высокомолекулярные гетерополимеры, играющие главную роль в передаче наследственных признаков и в управлении процессами биосинтеза белков. Они состоят из последовательности связанных в цепь нуклеотидов четырех типов, три из которых (А, Г и С) типичны для ДНК и РНК, а два других Т и U входят в состав только ДНК и РНК соответственно. Благодаря свойству комплементарности нуклеотиды попарно взаимодействуют как в одной цепи, так и между соседними цепями, образуя сложные пространственные структуры, определяемые последовательностью, иначе первичной структурой ДНК или РНК.

Для проявления текста последовательности ДНК, например, при адсорбционном взаимодействии с поверхностью в хроматографическом эксперименте, структура двойной спирали должна быть разрушена, переведена в однонитевое состояние. В таком состоянии адсорбционные свойства однонитевых ДНК и РНК оказываются весьма схожими. Олигонуклеотиды однонитевых ДНК с относительно небольшой степенью полимеризации $n = 10–60$ применяются в разнообразных биологических исследованиях. Для чтения текстов нуклеотидных последовательностей и отбора особых, важных для биологии олигонуклео-

тидов, используются методы жидкостной хроматографии с масс-спектрометрическим детектированием.

Вопросам хроматографического исследования последовательности олигонуклеотидов посвящено большое число работ. Основным итогом их явилось создание простой корреляционной теории, связывающей объем удерживания с составом цепи олигонуклеотида [1]. Такая теория основывается на аддитивном вкладе нуклеотида в объем удерживания, дополненном логарифмической поправкой на длину цепи n , и по этой причине связана в первую очередь со средним нуклеотидным составом цепи. Идеино данная теория весьма близка к различным вариантам аддитивных моделей, разработанных для предсказания объемов удерживания пептидов и полимеров, состоящих из аминокислот [2–5]. Как было показано нами в работе [6], эмпирические поправки, вводимые в аддитивные корреляционные теории для их уточнения, автоматически вытекают из физических моделей адсорбции, представляющих макромолекулы в виде жестких стержней или в виде цепи случайных блужданий, адсорбирующихся в ограниченных объемах.

Подобно любым макромолекулам, цепочечная структура онДНК приводит к фазовому характеру адсорбции, происходящей в узком диапазоне со-

става растворителя. Как и в любых цепочечных макромолекулах, объем удерживания одонитевых ДНК может определяться не только средним составом, но и характером чередования звеньев с различной энергией адсорбции. В частности, макромолекулы, имеющие один и тот же набор нуклеотидов, но различающиеся последовательностью их расположения в цепи, могут заметно отличаться по своим адсорбционным (и не только) свойствам.

Зависимость адсорбционных свойств от последовательности вытекает из физической сущности связывания звеньев в цепь и включения во взаимодействие с поверхностью энтропийной составляющей. В наиболее простом виде адсорбционное взаимодействие описывается моделью решеточной цепи [7], лежащей в основе модели BioLCCC [8]. В отличие от корреляционных теорий, в которых влияние вариации текста присутствует в виде поправок, в модели BioLCCC зависимость адсорбционных свойств от последовательности звеньев возникает как результат некоммутиативного произведения переходных матриц [7, 8].

Хотя для практического применения корреляционные теории в большинстве случаев достаточны, с идейной точки зрения представляет интерес возможность описания адсорбционных свойств одонитевых ДНК в рамках модели адсорбции решеточной цепи или жесткого стержня, что делается в настоящей работе. Такая модель дает глубокое понимание физики процесса адсорбции и его связи с общей теорией критических явлений. Взаимосвязь параметров, описывающих макромолекулу, адсорбент и растворитель, открывает широкие возможности для оптимизации разделения по тем или иным признакам строения цепи.

МОДЕЛЬ АДсорбЦИОННОГО ВЗАИМОДЕЙСТВИЯ

В отличие от пептидов макромолекулы одонитевых ДНК являются полиэлектролитами, в которых каждое звено может быть заряжено. Для жидкостной хроматографии, обычно проводимой в водных средах на обращенной фазе, используется растворитель сложного состава, содержащий, кроме воды и ацетонитрила, добавки ацетата триэтиламмония при pH 7.0, что позволяет нейтрализовать заряженность звена и включить его адсорбционное взаимодействие. В таком растворителе взаимодействие звена с поверхностью на молекулярном уровне имеет сложный характер и не сводится к чисто гидрофобному. Однако для рассматриваемых ниже моделей адсорбции решеточной цепи или стержня точный механизм взаимодействия в масштабе звена не важен. Важно лишь, чтобы это взаимодействие было локальным, т.е. действие адсорбционных

сил проявлялось на расстояниях порядка размера звена, что будет предполагаться ниже.

Значительным упрощением для построения модели адсорбционного взаимодействия онДНК является тот факт, что по крайней мере для трех оснований *T*, *C* и *A* возможен синтез гомополимеров, представляющих не что иное, как обычные макромолекулы олигомеров (здесь и далее используются стандартные буквенные обозначения нуклеотидов). Закономерности хроматографического разделения олигомеров с функциональными группами, в частности переход через критическую точку, смена режима разделения с эксклюзивного на адсорбционный, исчезновение молекулярно-массового распределения, хорошо известны [9]. Как и для полимеров, для гомополимеров одонитевых ДНК переход в адсорбированное состояние осуществляется в узком диапазоне состава растворителя и имеет черты фазового перехода. Это упрощает нахождение основного феноменологического параметра модели – эффективной энергии адсорбции звена ϵ . Подробное экспериментальное исследование разделения онДНК приведено в работе [1], данные которой будут сопоставлены с результатами расчета.

Эксперименты в работе [1] проводили на колонке Xterra MS C18 размером 30×4.6 мм (или 50×4.6 мм для разделения в градиенте), диаметр пор адсорбента 125 \AA . Для расчетов примем размер звена нуклеотида 5 \AA , тогда размер пор, выраженный в звеньях, $D = 25$. Отметим, что точный выбор размера звена не очень важен для применения модели, однако от него будет зависеть величина эффективной энергии.

Чтобы связать рассчитываемый в теории [6, 7] коэффициент распределения K_d (численно равный статистической сумме макромолекулы в поре) с экспериментально измеряемым коэффициентом емкости k примем, что объем жидкой фазы $V^0 = 0.9$ от общего объема колонки, а объем пор $V_p = 0.4 V^0$. Тогда коэффициент емкости можно записать как

$$k = \frac{V_p(K_d - 1)}{V^0}, \quad (1)$$

объем удерживания V_R макромолекул

$$V_R = V^0(1 + k),$$

а время удерживания t_R

$$t_R = V_R/u,$$

где u – объемная скорость подачи растворителя.

Эффективную энергию адсорбции звеньев ϵ всюду в дальнейшем будем выражать в единицах kT .

Первый вопрос, на который необходимо ответить, анализируя данные по разделению гомополимеров [1], какую модель макромолекулы, жест-

кого стержня [6] или решеточной цепи блужданий [7], взять за основу? Качественно эти модели в доступных экспериментально режимах разделения, например пептидов, близки и обе предсказывают существование критической точки с резким изменением адсорбционных свойств. Тем не менее, они имеют заметные различия в том, что касается зависимости адсорбционного взаимодействия от длины цепи n .

Принято считать, что цепи нуклеотидов достаточно жесткие. Кроме того, учитывая полиэлектролитный характер цепи, казалось бы, логично предположить, что ей ближе соответствует модель стержня. Однако анализ экспериментальных данных в изократическом режиме для гомополимеров типа $(dT)_n$ показал, что для таких гомополимеров зависимость k от степени полимеризации n точнее соответствует модели цепи случайных блужданий [7], а не модели жесткого стержня (рис. 1). Модель стержня [6] предсказывает более резкую зависимость, заметно отличающуюся от экспериментальной.

Возможным объяснением того, что решеточная цепь точнее соответствует эксперименту, может быть то, что состав растворителя в области, близкой к критической точке, нейтрализует заряженность звеньев цепи, и ее конформация становится близкой к клубку. Нельзя исключить, однако, что совпадение расчета в рамках модели BioLCCC и эксперимента обусловлено случайными причинами, которые рассматриваемая упрощенная модель не учитывает.

При каждом значении объемной доли V_b ацетонитрила в смеси молекулярно-массовые зависимости рис. 2 следуют модели решеточной цепи при одном и том же значении эффективной энергии адсорбции звена ϵ . Эффективные энергии звеньев, соответствующие экспериментальным точкам при разных V_b , приведены ниже.

ϵ	1.60	1.17	0.72	0.56	0.45	0.41	0.37	0.31	0.26
V_b	0.04	0.06	0.08	0.09	0.10	0.105	0.11	0.12	0.13
N_b	0.014	0.021	0.029	0.033	0.037	0.039	0.041	0.045	0.049

(N_b – мольная доля ацетонитрила).

Для точного совпадения экспериментальных и теоретических величин k в модель дополнительно была введена энергия адсорбции концевых звеньев $\epsilon'(T)$, отличающаяся на постоянную величину от энергии срединных звеньев цепи $\epsilon(T)$. Такое различие имеет определенное обоснование: химические структуры звена цепи T и концевого звена T' различаются. Разница энергий $\epsilon'(T) - \epsilon(T) \approx 1.0$, постоянная и не зависящая от состава растворителя, позволяет получить точное соответствие расчетных и экспериментальных величин (рис. 2).

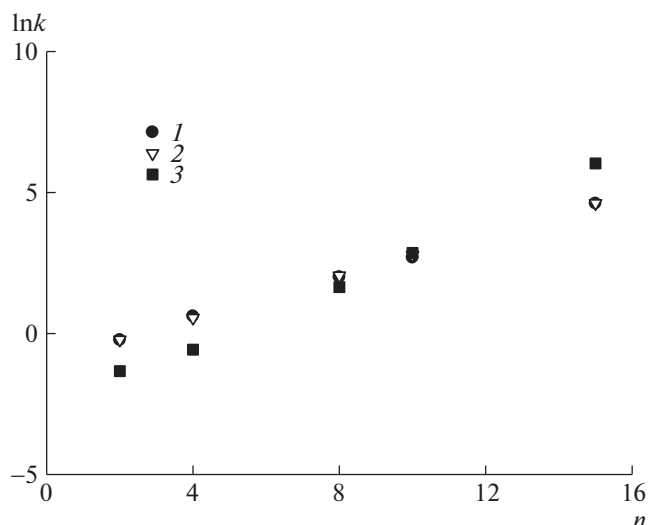


Рис. 1. Зависимость $\ln k$ от степени полимеризации n гомополимера $(dT)_n$ в изократических условиях при объемной доле ацетонитрила $V_b = 0.08$. 1 – эксперимент, 2 – расчет по модели клубка, 3 – расчет по модели стержня.

В целом решеточная цепь весьма хорошо описывает изократический эксперимент для гомополимера $T'(dT)_{n-2}T'$ (рис. 3). По этой причине далее в работе для односторонних ДНК разного состава будет применяться решеточная модель, аналогичная модели BioLCCC [8], ранее использованной авторами для описания разделения пептидов и белков.

Кратко содержание модели BioLCCC сводится к следующему. Поровое пространство адсорбента представляется в виде одномерной щели, размер которой D соответствует реальному размеру пор, выраженному в единицах размера звена. Макромолекула, адсорбирующаяся в такой щели, описывается цепью случайных блужданий длиной n на кубической решетке, размер ячеек которой совпадает с размером звена. Все возможные конфигурации цепи задаются произведением n переходных матриц $W(\epsilon_i)$, элементы которых зависят от энергии адсорбции ϵ_i соответствующего звена i последовательности, а размерность совпадает с размером поры D :

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

Коэффициент распределения K_d , численно совпадающий со статистической суммой макромолекулы, получается суммированием всех конфи-

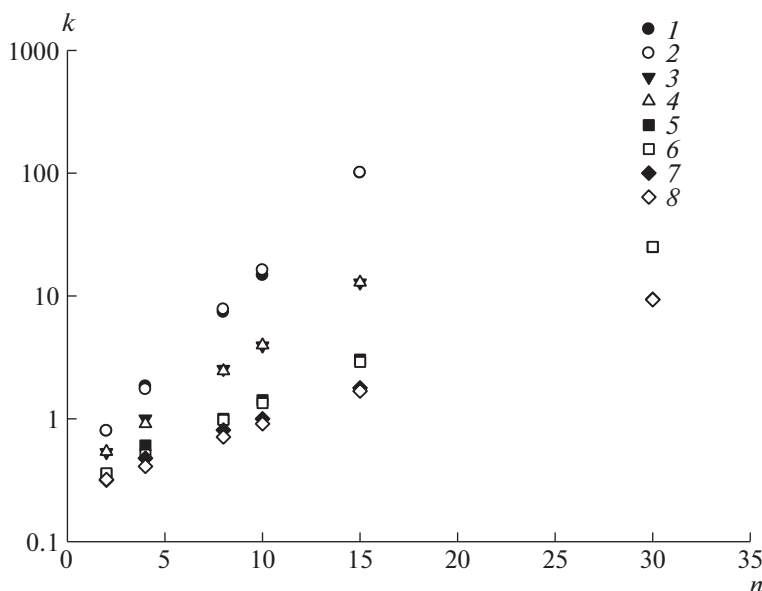


Рис. 2. Молекулярно-массовые зависимости коэффициента емкости k гомополимера $T'(dT)_{n-2}T'$ для изократического состава $V_b = 0.08$ (1, 2), 0.09 (3, 4), 0.1 (5, 6), 0.105 (7, 8). 1, 3, 5, 7 – эксперимент; 2, 4, 6, 8 – расчет.

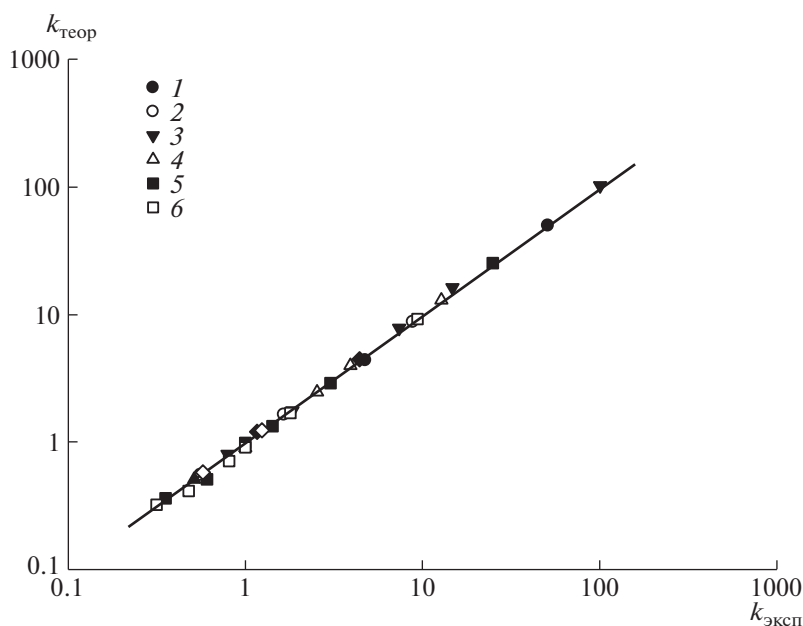


Рис. 3. Сравнение расчетных $k_{\text{теор}}$ и экспериментальных $k_{\text{эксп}}$ коэффициентов емкости в изократических режимах для гомополимера $T'(dT)_{n-2}T'$ с $n = 2$ (1), 4 (2), 8 (3), 10 (4), 15 (5) и 30 (6) в диапазоне составов растворителя с объемной долей ацетонитрила $V_b = 0.04-0.13$.

гураций цепи с помощью единичного вектора U , компоненты которого равны единице:

$$K_d = \frac{1}{D} U^T \times \left[\prod_i W(\epsilon_i) \right] \times U \quad (2)$$

Порядок произведения матриц соответствует последовательности разных нуклеотидов i в цепи. Произведение матриц некоммутативно, следова-

тельно, K_d зависит от последовательности нуклеотидов.

ОДНОНИТЕВЫЕ ДНК В ГРАДИЕНТЕ РАСТВОРИТЕЛЯ

Для описания зависимости V_R от градиентного профиля необходимо связать эффективную энергию адсорбции звена ϵ с составом растворителя –

объемной V_b или мольной долей N_b компонента b (ацетонитрила в данном случае). В модели BioLCCC используется приближение Ленгмюра, развитое для адсорбционной хроматографии в работах L.R. Snyder [10]. Хотя данное приближение для разделения на обращенной фазе плохо обосновано, как показывает опыт [7], оно оказывается вполне удовлетворительным для моделирования разделения пептидов и белков. Причина заключается в том, что для макромолекул даже не очень большой молекулярной массы существенное изменение адсорбционных свойств происходит в узком диапазоне состава растворителя вблизи критической точки адсорбции $\epsilon_c(N_c)$, и детальный вид зависимости энергии адсорбции звена $\epsilon(N_b)$ в этом узком диапазоне оказывается несущественным. Без потери общности для бинарного растворителя с компонентами a и b можно принять энергию адсорбции “слабого компонента” a (воды) на обращенной фазе $\epsilon_a = 0$, тогда зависимость $\epsilon_{ab}(N_b)$ смеси от концентрации “сильного компонента” b (ацетонитрила) записывается следующим образом [7]:

$$\epsilon_{ab} = \ln[N_b(e^{\epsilon_b} - 1) + 1],$$

где ϵ_b – энергия адсорбции компонента b . Для энергии звена $\epsilon(N_b)$ имеем

$$\begin{aligned} \epsilon(N_b) &= \epsilon^0 - \epsilon_{ab} \\ \epsilon(N_b) &\approx \epsilon_c + \ln \left[\frac{N_c}{N_b} \right] \end{aligned}$$

Здесь ϵ^0 – энергия звена в “чистом” компоненте a , ϵ_c – энергия звена, соответствующая критической точке (равная $\ln(6/5)$ для рассматриваемой модели), которой соответствует состав $N_c \approx 0.05$ для гомополимера $(dT)_n$. В последнем выражении учтено, что обычно энергия адсорбции сильного компонента (ацетонитрила) $\epsilon_b \gg 1$.

На рис. 4а показано соответствие экспериментально найденных эффективных энергий звеньев для растворителя разного состава (см. выше) и приведенного выше приближения Ленгмюра. Как видно, данное приближение неплохо соответствует экспериментальным точкам и могло бы быть взято для описания градиентного разделения.

Однако в настоящей работе мы не будем применять приближение Ленгмюра. Для общности мы покажем, что в модели BioLCCC в масштабе звена можно использовать любое корреляционное или эмпирическое выражение, описывающее зависимость его эффективной энергии от состава растворителя. Например, как следует из рис. 4б, для звеньев T имеет место практически линейная

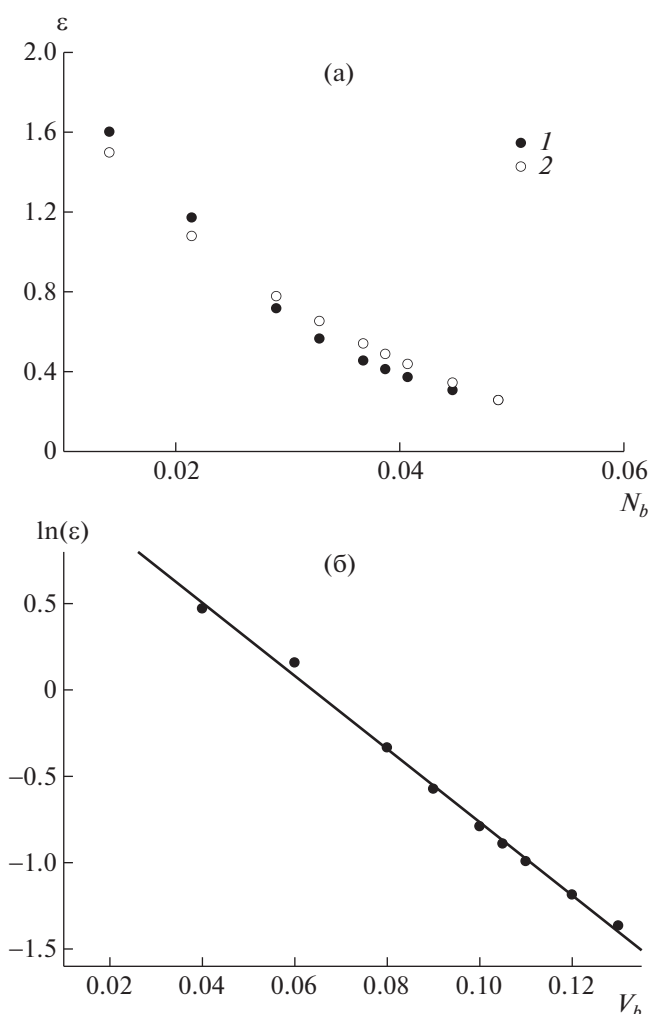


Рис. 4. Зависимость эффективной энергии ϵ звена от мольной доли ацетонитрила N_b в бинарной смеси для клубка (1) и приближения Ленгмюра (2) (а), а также зависимость $\ln(\epsilon)$ от объемной доли ацетонитрила V_b для гомополимера $(dT)_n$ (б).

зависимость $\ln \epsilon$ от объемной доли ацетонитрила V_b :

$$\ln(\epsilon) = \ln(\epsilon^0) - S \cdot V_b$$

Воспользуемся далее этой простой эмпирической зависимостью для описания градиентного разделения. Наблюдаемая линейная зависимость похожа на линейное приближение $\ln k$ от V_b [11]. Параметр $\ln(\epsilon^0)$ соответствует в данном случае энергии в некотором “стандартном” растворителе, например, в составе, реализующем критическую точку адсорбции, или в “чистом” компоненте a растворителя при $V_b = 0$.

В градиенте растворителя K_d и связанный с ним в соответствии с уравнением (1) коэффициент k зависят от состава растворителя $V_b(V)$, изменяющегося при изменении времени или объема V , прокачанного через колонку. Объем удержива-

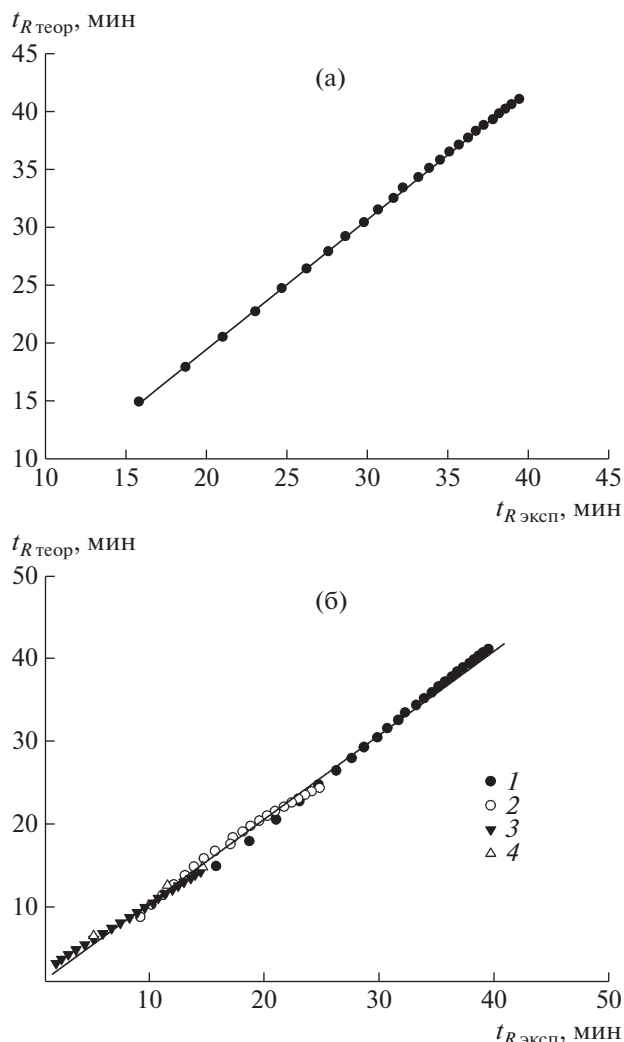


Рис. 5. Корреляция между расчетными $t_{R \text{ теор}}$ и экспериментальными $t_{R \text{ эксп}}$ временами удерживания при градиентном разделении для гомополимера $(dT)_n$ (а) и общая корреляция для гомополимеров $(dT)_n$ (1), $(dA)_n$ (2), $(dC)_n$ (3) и $(CCG)_n$ (4) при градиентном разделении [1] (б).

ния V_R (или время удерживания t_R) определяется известным уравнением

$$\int_0^{V_R - V^0} \frac{dV}{V^0 k(V)} = 1$$

Таблица 1. Эффективные энергии адсорбции звеньев ϵ^0 при $V_b = 0$ и ϵ при $V_b = 0.07$

Нуклеотид	$\ln(\epsilon^0)$	$\epsilon (V_b = 0.07)$
T	1.35	0.88
A	0.82	0.52
C	0.50	0.37
G	0.55	0.39

Примечание. Параметр $S = 21.185$ для всех типов звеньев, для концевых звеньев всех типов $\epsilon(X) = \epsilon(X) + 1.0$.

Результат расчета разделения гомополимера $(dT)_n$ для градиента с начальной точкой 5% и наклоном 0.15% ацетонитрила/мин [1] показан на рис. 5а. Как и следовало ожидать, из точного описания изократического режима следует весьма хорошее соответствие расчета и эксперимента для градиента.

Для других гомополимеров, для которых в работе [1] не приводятся результаты разделения в изократическом режиме, примем, что различие заключается только в разнице стандартных энергий адсорбции звеньев, в то время как наклон S для них одинаков. Из представленных в работе [1] экспериментальных данных по градиентному разделению гомополимеров можно найти энергии звеньев A и C . Для звена G энергию можно определить из данных по градиентному разделению макромолекул типа $(CCG)_n$. Соответствующие энергии нуклеотидов приведены в табл. 1.

С использованием найденных таким образом энергий адсорбции звеньев T, A, C и G были рассчитаны времена удерживания гетерополимеров однонитевых ДНК. Последовательность букв в структуре олигонуклеотида задает порядок произведения переходных матриц в выражении (2) с различной энергией звена, соответствующей букве. Результаты сопоставлены с экспериментальными данными в табл. 2 и на рис. 6. В работе [1] удерживание макромолекул олигонуклеотидов приведено к объемному содержанию ацетонитрила в растворителе, при котором макромолекула выходит из колонки. Эти величины были пересчитаны нами в t_R с помощью параметров градиентного профиля (начальной точки 5% и скорости изменения 0.15% ацетонитрила/мин).

РЕЗУЛЬТАТЫ И ИХ ОБСУЖДЕНИЕ

Как следует из приведенных выше результатов, модель решеточной цепи, дополненная корреляционной зависимостью $\epsilon(V_b)$, правильно схватывает основные черты разделения однонитевых ДНК. Необходимо, однако, отметить следующее.

Хотя общий коэффициент корреляции $R^2 = 0.99$ для гомополимеров близок к единице, на рис. 5б можно заметить, что для взятого отдельно гомополимера $(dA)_n$ корреляция между расчетом и экспериментом не столь хороша. Варьируя параметры $\ln(\epsilon^0)$ и S , вероятно, можно улучшить корреляцию и для $(dA)_n$. Как уже отмечалось, для модели совершенно неважно, каким образом описывать зависимость эффективной энергии от состава растворителя, в частности, можно использовать для этого и нелинейную аппроксимацию. В настоящей работе мы не стали прибегать к нелинейной аппроксимации, поскольку она вряд ли улучшит описание для гетероцепей одноните-

Таблица 2. Последовательности одонитевых ДНК и их экспериментальные и расчетные времена выхода t_R

<i>n</i>	Последовательность	%В* [1]	t_R эксп. мин	t_R теор. мин
16	<i>GTAGCAGCAGCCAGAC</i>	7.12	14.13	12.15
16	<i>GTCTGGCTGCTGCTAC</i>	7.66	17.73	17.71
16	<i>GACATGGGCTCCGCTC</i>	7.22	14.80	14.49
18	<i>GAGATCTAGGTGATGTGC</i>	7.84	18.93	20.48
18	<i>ACAGTATTGCTGGCGGAG</i>	7.92	19.47	18.32
19	<i>GAAGAATTGGTAGGAGTCC</i>	7.72	18.13	19.41
20	<i>ACTGGGAACGAGGGCGACGT</i>	7.77	18.47	16.79
20	<i>TCAGACGAGAAGGACCTGAC</i>	7.78	18.53	16.98
20	<i>TTCCGCCAGCAACAATGTGAC</i>	8.47	23.13	19.64
22	<i>GTCCTTTGAATGCAGAAATGGC</i>	8.57	23.80	22.84
22	<i>ATGGTTCCTGTGCAACTGTGAC</i>	8.75	25.00	23.36
23	<i>CAGCCCCTTCCCCTCTGCCGCCAG</i>	7.89	19.27	18.75
25	<i>ACCTCTGCACCCATCTCTCTCCTCA</i>	9.02	26.80	23.56
26	<i>GGGGTACCAGTATCGAAAGCACCTCC</i>	8.47	23.13	21.36
26	<i>GCGAATTCCAATACCGATATTTGGA</i>	9.26	28.40	25.14
26	<i>GCGAATTCTTATTTAGGATTGATAGC</i>	9.42	29.47	28.00
27	<i>TGGGAATTCAGTGCTTGGTCACTTCCG</i>	9.2	28.00	26.28
30	<i>GGGCTTCATGATGTCCCATAAATTTTGGC</i>	9.48	29.87	28.55
30	<i>AGCGTTGGCTACCCGTGATATTGCTGAAGA</i>	9.37	29.13	26.33
32	<i>CTTACCCAACAAGTGC GGCCGATCGAATTGAG</i>	9.22	28.13	25.12
32	<i>CTCAATTTCGATCGGGCCGCACTTGTGGGTAAG</i>	9.37	29.13	27.16
32	<i>GCCAGGTGCCAGGCTGGGTTGCCATTGATGGC</i>	8.82	25.47	24.88
32	<i>CCGCTCGAGGGTTCTCTGGATAACATCTCAGC</i>	9.19	27.93	26.14
33	<i>CTCCAACGGGAACCTGGCCCCCTGGGCAGGGGG</i>	8.75	25.00	21.18
33	<i>CCCCCTGCCAGGGGGCCAGGTTCCCGTTGGAG</i>	8.88	25.87	22.74
34	<i>CCTGCGCATCATTAAATCAGAATGGCATTACCCG</i>	9.71	31.40	27.86
34	<i>CGGTGAATGCCATTCTGATTAAATGATGCGCAGG</i>	9.53	30.20	27.98
37	<i>AATGGATCCGGAGGAGGGGACACCTTCATCCGTC</i>	9.56	30.40	25.76
39	<i>CCCAAGCTTATGGAAACGCGGCGGGTGGGCCAGGGC</i>	9.06	27.07	23.85
39	<i>CCATCCAGGGGGGAGCGGCCGCCCTCGACCTCAGAGCC</i>	8.88	25.87	22.44
39	<i>CATCCTCAGGTAGGCACTGAGCATCCCCCATAACGACC</i>	9.37	29.13	26.21
39	<i>CCTCAGGTAGGCACTGAGCATCCCCCATAAGCTCCACC</i>	9.29	28.60	25.59
41	<i>CATCCTCAGGTAGGCACTGAGCATCCCCCATAACTTCCTC</i>	9.34	28.93	28.11
48	<i>GGAGCCACTGGCCACGCCGCGCTTTGGATGCGCGGGAGGTGATTCCC</i>	9.56	30.40	27.57
54	<i>TATAAGCTTGAAGGGGAACGAAAGATGGCGGGGAAACACTGCTGTCCAGTTG</i>	10.08	33.87	29.43
56	<i>AGAGAATTCCACCATGGACTACAAGGACGACGATGACAAGTCTCAGAGCAACCGG</i>	9.87	32.47	28.88
57	<i>ATTTCTAGATTACTATTTGTGCATCGTCCTTATAGTCAAAGTCAAATTCGTCCAGGTC</i>	10.72	38.13	36.17
57	<i>CAATATTTTACATGAACTGGAGGTCCGTCAATGACAGTGTAGGCTGGAGCTGCTTCG</i>	10.17	34.47	31.97
60	<i>GACTTAGACTTAGACTTAGACTTAGACTTAGACTTAGACTTAGACTTAGACTTAGACTTA</i>	10.57	37.13	33.89

*Объемное содержание ацетонитрила в растворителе, при котором макромолекула выходит из колонки.

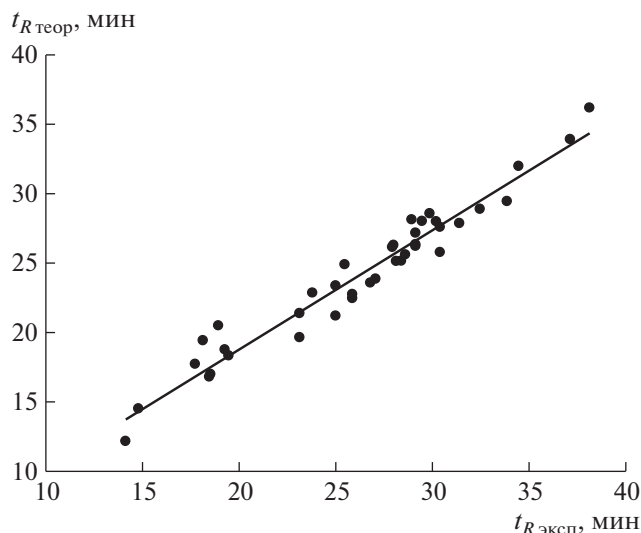


Рис. 6. Корреляция между расчетными $t_{R \text{ теор}}$ и экспериментальными $t_{R \text{ эксп}}$ временами удерживания для гетерополимеров онДНК со степенью полимеризации $n = 16\text{--}60$ при градиентном разделении [1].

вых ДНК, а модель при такой аппроксимации теряет простоту. Тем не менее, необходимо сказать, что исследование зависимости $\ln k$ гомополимера $(dA)_n$ от состава растворителя в изократическом режиме, подобное сделанному для $(dT)_n$, может прояснить особенности адсорбционного взаимодействия звеньев A с поверхностью и уточнить описание его градиентного разделения.

Также отметим, что принятая в работе корреляционная зависимость энергии от состава в отличие от приближения Ленгмюра нарушает аддитивность вкладов звена и растворителя, что неявно подразумевалось в модели BioLCCC и в ее приложениях для пептидов. Разница энергий взаимодействия нуклеотидов T, A, C, G не абсолютна, а зависит от состава растворителя.

Наконец, коэффициент корреляции $R^2 = 0.95$ между расчетом и экспериментом для гетерополимеров (рис. 6) оказывается заметно хуже, чем для гомополимеров. Другими словами, предсказание именно последовательности гетерополимера, а не только среднего состава, из измеренного времени удерживания t_R получается недостаточно точным. Рассмотрим данный вопрос более подробно.

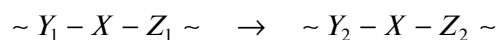
Зависимость объема или времени удерживания от последовательности имеет две составляющие — физическую и химическую. В рассматриваемой модели учитывается только физическая составляющая зависимости. Предполагается, что вариация последовательности, составленной из одного и того же набора “букв” — нуклеотидов, не

меняет их энергию, т.е. словами, при вариации последовательности средняя энергия звена сохраняется. В этом случае зависимость t_R от последовательности вблизи критической точки имеет в основном энтропийную природу (заметим, что в градиенте макромолекулы перемещаются в колонке в составе растворителя, близком к критическому). Чтобы вариация последовательности (например, перестановка звеньев) была значимой, согласно размерным оценкам работы [12], длина цепи n должна соответствовать условию

$$\Delta \epsilon \cdot \ln n > 1,$$

где $\Delta \epsilon$ — дисперсия энергии взаимодействия звеньев. Другими словами, для проявления последовательности при небольшой дисперсии цепь должна быть достаточно длинной. Оценка дисперсии взаимодействия звеньев гипотетической гетероцепи одонитивых ДНК с равным содержанием нуклеотидов A, T, G, C в составе растворителя, при котором она выходит из колонки (табл. 1), дает $\Delta \epsilon \approx 0.04$, так что для цепей с длиной $n = 40\text{--}60$ физическая составляющая, по-видимому, малозначима. Объем удерживания определяется длиной цепи и средним нуклеотидным составом, тем самым подтверждая аддитивный подход [1]. Макромолекулу гетерополимера при этом можно рассматривать как гомополимер с некоторой средней энергией взаимодействия звена с поверхностью. Конечно, возможно существование каких-то уникальных последовательностей, значительно отклоняющихся по времени удерживания от средней величины, но для большинства макромолекул одонитивых ДНК вариация последовательности приводит к отклонению времени удерживания порядка ширины пика.

Вместе с тем существует и химический фактор в зависимости адсорбционных свойств от последовательности нуклеотидов. При перестановке нуклеотид X , имеющий изначально соседями Y_1 и Z_1 , оказывается в локальном окружении других звеньев — соседей Y_2, Z_2 :



В соответствии с общими химическими законами, за счет эффектов соседа взаимодействие X с поверхностью может измениться. Тем самым, средняя энергия звена меняется, и вариация последовательности оказывается заметной уже за счет энергетической составляющей. Вероятно, для коротких цепей одонитивых ДНК химический фактор преобладает, что и вызывает расхождение расчета и эксперимента для гетерополимеров. Макромолекулы, составленные из одного и того же набора нуклеотидов, но различающиеся последовательностью, можно рассматривать как гомополимеры с различной средней энергией звена и, следовательно, с разной критической точкой и временем выхода.

В рамках рассматриваемой модели химический фактор можно учесть. Для этого необходимо синтезировать порядка 30 модельных одонитевых ДНК, имеющих в середине цепи структуру типа



в которой X , Y , Z последовательно замещаются на четыре типа возможных нуклеотидов. Исследование разделения таких макромолекул в *изократическом* режиме позволит определить поправки к энергии звеньев X , зависящие от соседей, и внести их в схему расчета.

Наконец, не стоит забывать, что одонитевая ДНК может замыкаться сама на себя за счет комплементарности составляющих ее нуклеотидов. Если замыкание происходит, то образующиеся вторичные структуры меняют адсорбционные свойства макромолекулы и, в какой-то степени, маскируют ее последовательность. Скорее всего в рассматриваемом сложном растворителе при температуре 60°C вероятность такого замыкания небольшая. В работе [13] показано, каким образом в модель BioLCCC можно включить различные вторичные структуры — складки, циклы, листы. В соответствии с расчетами для относительно коротких пептидов влияние таких структур малозаметно, что приближает их адсорбционные свойства к соответствующим принципу аддитивности.

ЗАКЛЮЧЕНИЕ

Подводя итог можно сказать, что модель адсорбции решеточной цепи, дополненная корреляционной зависимостью эффективной энергии звена от состава растворителя, правильно схватывает основные закономерности разделения олигонуклеотидов на обращенной фазе. Она может использоваться как для предсказания времен удерживания олигонуклеотидов с известной последовательностью для их отбора из сложной смеси, так и для оптимизации условий разделения макромолекул с различными последовательностями. В силу малой дисперсии энергий адсорбции разных нуклеотидов на обращенной фазе разделение в основном чувствительно к составу нуклеотидов в цепи, а не к тексту последовательности. По этой причине предсказательные возможности модели аналогичны аддитивной модели, предложенной в работе [1].

Модель оказывается весьма содержательной и единым образом описывает адсорбционные свойства широкого класса макромолекул, имеющих цепочечное строение — синтетических полимеров, пептидов, белков и одонитевых ДНК. Отметим также что данная модель перекидывает мост между макромолекулами и низкомолекулярными соединениями: она содержит в себе также и возможность описания хроматографии низ-

комолекулярных соединений, которым соответствует длина цепи $n = 1$. Это дает возможность рассматривать хроматографию единым образом, не выделяя отдельно хроматографию низкомолекулярных веществ и макромолекул. Наконец, модель имеет широкие возможности для более точного описания структуры макромолекулы и может быть основой для рассмотрения не только адсорбции на поверхности, но и других взаимодействий с участием макромолекул.

Современные методы компьютерного моделирования дают возможность детального описания структуры макромолекул. Однако надо иметь в виду, что для моделирования необходимо точное знание множества потенциалов взаимодействия внутри цепи, а также их зависимость от внешних условий. Определение потенциалов представляет собой сложную экспериментальную задачу. Детализация структуры макромолекулы требует столь же подробной детализации поверхности. Эти вопросы значительно усложняют как моделирование взаимодействия макромолекулы с поверхностью, так и интерпретацию результатов. Хотя такая программа исследования представляется перспективной, на наш взгляд, упрощенные модели, подобные рассмотренной в настоящей работе, правильно дополненные небольшим числом эмпирических параметров, не потеряли своей актуальности.

СПИСОК ЛИТЕРАТУРЫ

1. Gilar M., Fountain K.J., Budman Y., Neue U.D., Yardley K.R., Rainville P.D., Russel II R.J., Gebler J.C. // J. Chromatogr. A. 2002. V. 958. P. 167.
2. Guo D., Mant C.T., Taneja A.K., Parker J.M.R., Hodges R.S. // J. Chromatogr. 1986. V. 359. P. 499.
3. Krokhin O.V. // Anal. Chem. 2006. V. 78. I. 22. P. 7785.
4. Reimer J., Spicer V., Krokhin O.V. // J. Chromatogr. A. 2012. V. 1256. P. 160.
5. Gilar M., Hie H.W., Jaworski A. // Anal Chem. 2010. V. 82. P. 265.
6. Gorshkov A.B., 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.
7. Gorshkov A.V., Evreinov V.V., Tarasova I.A., Gorshkov M.V. // Polymer Science B. 2007. V. 49. № 3—4. P. 93.
8. 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.
9. Entelis S.G., Evreinov V.V., Gorshkov A.V. // Adv. Polym. Sci. 1986. V. 76. P. 129.
10. Snyder L.R. Principles of Adsorption Chromatography. New York: Marcel Dekker, 1968.
11. Snyder L.R., Dolan J.W., Gant J.R. // J. Chromatography B. 1979. V. 165. P. 3.
12. Обухов С.П. // Журн. эксперим. и теорет. физики. 1987. Т. 93. № 6. С. 1973.
13. Gorshkov A.V., Rozdina I.G., Pridatchenko M.L., Evreinov V.V. // Polymer Science A. 2020. V. 62. № 4. P. 437.