

## ПЕРСПЕКТИВЫ СЕКВЕНИРОВАНИЯ БИМОЛЕКУЛ ПРИ ТРАНСЛОКАЦИИ ЧЕРЕЗ НАНОПОРЫ I. ОБЗОР МЕТОДОВ

© 2013 г. В. Л. Носик, Е. Б. Рудакова

Институт кристаллографии РАН, Москва

E-mail: v-nosik@yandex.ru

Поступила в редакцию 07.12.2012 г.

С накоплением данных о структуре макромолекул, полученных методами рентгеноструктурного анализа с использованием синхротронного излучения, ядерного магнитного резонанса и масс-спектрометрии все больше внимания уделяется изучению функциональных свойств биомолекул в нативных растворах, в частности их взаимодействию с мембранами. Это направление исследований тесно связано с развитием новых технологий секвенирования, т.е. определения последовательности нуклеотидов в биомолекуле ДНК. Одними из наиболее перспективных “физических” подходов к секвенированию являются методы, основанные на использовании наноканалов или нанопор, сквозь которые биомолекула в ионном растворе проходит под влиянием приложенного электрического поля. При этом нанопора обеспечивает пространственную локализацию молекул и позволяет осуществить детектирование сигнала (электрического, флуоресцентного и т.д.) от отдельного нуклеотида. В связи с появлением новых ярких импульсных источников рентгеновского излучения методы, основанные на исследовании флуоресцентного сигнала становятся все более привлекательными. Обсуждаются существующие методы моделирования движения биомолекул, определения их структуры и технологии секвенирования, а также перспективы дальнейшего развития исследований в этой области.

DOI: 10.7868/S0023476113060192

### ОГЛАВЛЕНИЕ

#### Введение

1. Структурные исследования белковых молекул и ДНК в кристаллической форме и в нативных растворах
2. Прямые техники секвенирования
3. Динамика ДНК
4. Нанопоры и транслокация полинуклеотидов
5. Синтетические поры
6. Методики секвенирования
7. Комбинированные методы с использованием рентгеновского и синхротронного излучений
8. Основные уравнения движения биополимера в нанопоре

#### Заключение

### ВВЕДЕНИЕ

Одним из перспективных направлений развития медицины являются геномные исследования, устанавливающие взаимосвязь предрасположенности к отдельным заболеваниям с набором генов. Однако помимо сложностей с установлением этой взаимосвязи на основе анализа статистических данных и теоретических моделей проблемой является и разработка новых технологий секвенирования<sup>1</sup> —

<sup>1</sup> Секвенирование биополимеров (белков и нуклеиновых кислот — ДНК и РНК) — это определение их первичной аминокислотной или нуклеотидной последовательности (от англ. *Sequence* — последовательность).

способа определения последовательности полинуклеотидов (ПН) в цепочке ДНК.

В настоящее время для определения структуры белковых молекул (БМ) в широком смысле слова используются методы рентгеновской дифракции, ядерно-магнитного резонанса (ЯМР), масс-спектрометрические и “химические” методы. Данные, полученные этими методами, обычно взаимодополняют друг друга, позволяя воспроизвести структуру с некоторой точностью, например с пространственным разрешением 1–2 Å. Наиболее популярные “химические методы”, восходящие к методикам Сангера, рассматриваются как технически сложные и дорогие, поэтому идет поиск альтернативных методик.

Один из наиболее многообещающих “физических” методов секвенирования основан на использовании прохождения (транслокации) развернутой цепочки биомолекулы в ионном растворе через искусственно созданные наноканалы или нанопоры (НП). При этом если первоначальные эксперименты ставили своей целью только определение длины отдельных фрагментов биомолекул по времени блокады ионного тока, то сейчас предлагаются более сложные схемы, измерение характеристик отдельных ПН. Например, в [1–3] было показано, что НП со специально смонтированными электродами позволяют различать электрические сигналы от разных оснований.

Нанопора обеспечивает не только пространственную локализацию отдельных оснований молекулы ДНК, но и движение оснований с почти постоянной скоростью: одно основание проходит НП в течение некоторого интервала времени (обычно от 100 фс до 5 мкс).

Настоящая работа – это обзор существующих методов секвенирования, в том числе с использованием НП, и их сравнение с традиционными структурными методами.

В следующей статье будут приведены данные по моделированию движения длинных биомолекул в НП с использованием гидродинамических подходов и уравнений Ланжевена и Смолуховского, а также результаты расчетов с учетом нелинейности уравнений движения и разных геометрий НП.

Третья работа будет посвящена вопросам идентификации отдельных оснований по спектрам флуоресценции и фотоэмиссии отдельных атомов (в основном атомов углерода и азота), находящихся в разных структурных положениях внутри оснований. Будут рассмотрены различные варианты фокусировки рентгеновского излучения на НП с использованием составных рефракционных линз, рентгеновских волноводов и изогнутых кристаллов (в геометрии Лауэ и Брэгга).

Использование для определения типа основания вторичного флуоресцентного и фотоэмиссионного сигнала современных спектрально чувствительных детекторов позволяет добиться временного разрешения на уровне 0.2 мкс. Поэтому на первый взгляд “тонкая” временная структура излучения рентгеновских лазеров на свободных электронах (РЛСЭ), 100 фс, для этих целей не нужна. Однако именно высокая яркость излучения РЛСЭ позволяет собрать достаточное количество флуоресцентных фотонов и определить тип основания за время его нахождения внутри поры.

Известно, что высокая яркость излучения РЛСЭ приводит к радиационным повреждениям рентгеновской оптики и разрушению самих исследуемых объектов. Главным образом это связано с разогревом вещества и тем, что при фотоионизации электроны достаточно быстро покидают исходные нейтральные атомы, оставляя их в виде положительно заряженных отталкивающихся ионов [4–6]. В случае транслокации через НП такие ионы могут заблокировать движение биомолекулы или даже (при достаточной их концентрации) разорвать пептидную связь. Поэтому конструирование НП для флуоресцентных спектральных измерений требует как разработки специальной схемы формирования падающего пучка, способов управления движением биомолекулы в НП, так и, возможно, систем компенсации зарядов, образующихся при фотоионизации.

Благодаря использованию синхротронных источников (СИ) и “фабрик” белковой кристаллизации собран большой объем информации о структуре белков в кристаллическом состоянии. Однако остаются вопросы о том, насколько меняются функциональные свойства белков с переходом из нативного (в растворах с разными физико-химическими характеристиками) состояния в кристаллическое. На данном этапе интерес к изучению физико-химических свойств биомолекул связан с развитием протеомики [7–9] и новых медицинских практик, когда используются методы диагностики и лечения, основанные на использовании генной информации и предрасположенности конкретного человека к тем или иным заболеваниям.

В перспективе речь идет о клиническом применении качественно новых способов лечения, использующих особенности функционирования белков и генов в живой клетке. В настоящее время накоплен достаточно большой объем эмпирических данных, активно ведутся работы по моделированию структуры и динамики как отдельных атомов при взаимодействии белков в растворе, так и многомасштабное моделирование отдельных фрагментов биомолекул.

Задача описания проникновения биомолекул через разного рода мембраны является одной из ключевых для понимания процессов их функционирования и специфичности [10–12].

Так, ученые из группы С. Деккера продемонстрировали транслокацию (“проскальзывание”) отдельной молекулы ДНК через НП в слое графена толщиной в один атом, в которой было проделано нанотверстие – НП – рекодер. Показано, что отдельные молекулы ДНК можно “протянуть” через такую графеновую НП без разрушения молекулы. Метод обнаружения отдельного основания базируется на том, что при приложении к НП электрического напряжения ионы в растворе начинают перемещаться через отверстие, генерируя электрический ток. Когда молекула ДНК входит в НП и частично блокирует поток ионов, ток становится меньше. При этом каждая отдельная перемещающаяся через пору молекула ДНК обнаруживается по специфическому падению электрического тока [13].

## 1. СТРУКТУРНЫЕ ИССЛЕДОВАНИЯ БЕЛКОВЫХ МОЛЕКУЛ И ДНК В КРИСТАЛЛИЧЕСКОЙ ФОРМЕ И НАТИВНЫХ РАСТВОРАХ

Традиционная парадигма, движущая геномику [14], основана на предположении, что химическая структура белка определяет его функции. Однако известно, что свойства БМ в клетке определяются не только ее химическим составом, но и

пространственным строением – вторичной и третичной структурами [15].

Секвенирование как процедура определения последовательности оснований в линейной цепочке ДНК – это только первый шаг к пониманию функциональных свойств макромолекул. Традиционная белковая кристаллография с использованием синхротронного излучения (СИ) позволяет судить о структуре основных субъединиц молекул, не давая полной информации о функциональных особенностях БМ (особенно в нативных растворах).

Поэтому перспективы изучения взаимодействия белковых молекул и ДНК в растворах с НП и наноструктурами намного шире. Это не просто еще один способ определения последовательности аминокислотных остатков БМ. В 1972 г. нобелевский лауреат К. Анфинсен показал, что трехмерное строение белка является следствием определенной последовательности аминокислотных остатков в его составе. Тем не менее до сих пор физико-химические причины, определяющие процесс сворачивания белка, полностью не описаны, и теоретически предсказать пространственную структуру белка в конкретном окружении по его аминокислотной последовательности с абсолютной точностью невозможно.

Микрофлюидика [16], или скорее наноплюидика, белковых растворов, к которой можно отнести и транслокацию денатурированных БМ через НП, и изучение их структуры в нативных растворах, открывает новые возможности для изучения связи структура–свойства.

Отличия микрофлюидики от “макрофлюидики” в основном сводятся к необходимости учета таких факторов, как поверхностное натяжение, диссипация энергии на границе раздела, квазиламинарные течения, сопротивление среды. Причем в силу принципа подобия числа Рейнольдса для таких течений могут быть очень малыми, т.е. отдельные слои жидкости могут не смешиваться друг с другом, а основным механизмом транспорта между ними становится диффузия [17]. В случае наноплюидики взаимодействие со средой более важно, чем для объемных течений. Например, резко увеличивается или уменьшается вязкость возле стенок нанокпилляров, изменяются термодинамические параметры жидкости, а также проявляется химическая активность на границе раздела твердой и жидкой фаз. Одна из причин такого поведения состоит в том, что характерные параметры структуры жидкости, такие как дебаевская длина, становятся сопоставимыми с размерами НП.

Изменение пространственной структуры, а следовательно, свойств и биологической активности нативного белка, называют денатурацией, которая может быть обратимой и необратимой. В первом случае нарушаются последовательно чет-

вертичная, третичная или вторичная структуры, но возможен обратный процесс восстановления структуры белка – ренатурация; во втором – происходит разрыв пептидных связей в составе первичной структуры (например, при температуре выше 45°).

Стандартная процедура определения структуры БМ состоит из нескольких последовательных этапов.

*Получение исходного материала*, содержащего требуемый белок. Для получения белка в больших количествах используют методы генной инженерии – в ДНК микроорганизма (например, кишечной палочки *Escherichia coli*) “встраивают” искусственно синтезированный ген, называемый вектором. В результате клетка начинает вырабатывать требуемый белок.

*Очистка материала* состоит из нескольких стадий – центрифугирование, гель-электрофорез, хроматографическая очистка.

*Определение структуры* очищенного белка может проводиться следующими методами:

– белковой кристаллографии или рентгеноструктурного анализа (РСА), когда определяется пространственное расположение отдельных атомов белка в кристалле с помощью дифракции рентгеновского излучения;

– спектроскопии ЯМР;

– масс-спектрометрии и др.

Методы ЯМР позволяют провести анализ параметров химических связей в специфических биомолекулах [18], так как одни и те же ядра атомов в различных окружениях в молекуле дают различные сигналы ЯМР. Отличие такого сигнала ЯМР от сигнала стандартного вещества позволяет определить так называемый химический сдвиг, который обусловлен химическим строением изучаемого вещества и экранированием внешнего магнитного поля электронами атомов. Однако при этом приходится использовать *a priori* информацию о связях стабильных оснований.

*Масс-спектрометрия*. В масс-спектрометрических методах по данным спектра определяется отношение массы к заряду,  $e/m$ , (качественные особенности) и количество заряженных частиц с таким отношением, образующихся при том или ином процессе ионизации образца (количественные характеристики). В случае белков [19, 20] приходится анализировать большое количество отдельных участков расщепленной БМ, или специфически модифицированного белка. Ключевым и критическим моментом здесь являются возможности химического разделения молекул белка и/или присоединения к ним разных молекул.

*Рентгеновские методы*. Основные проблемы белковой кристаллографии хорошо известны:

– вырастить совершенный кристалл белка достаточно сложно из-за большого размера молекул (масса атомов в одной молекуле белка может быть больше 100 кДа), сложностей с “доставкой” и правильной “укладкой” в кристаллическую структуру. Медленный процесс диффузии в этом случае предпочтительнее, чем “быстрая” конвекция;

– легкие атомы, составляющие основу любой БМ (водород в особенности), плохо различимы из-за того, что они слабо рассеивают и поглощают рентгеновские лучи. Несмотря на обширные планы по применению дифракции тепловых нейтронов как метода, комплементарного рентгеновским, эта методика не получила широкого распространения (было сделано несколько тестовых экспериментов с хорошо известными белками, например лизоцимом) из-за малого потока существующих нейтронных источников. Аналогичная ситуация в нейтронном эксперименте сложилась с идеей использования изотопов водорода, сечение неупругого рассеяния которых аномально велико. Методику предполагалось применить для визуализации положения молекул воды [21];

– количество дифракционных рефлексов при исследовании биомолекул исключительно велико, а их анализ затруднен без использования *a priori* информации о комплексах аминокислот или нуклеотидов. Наиболее удачные эксперименты связаны с определением положения тяжелых атомов металлов (например, железа в гемоглобине) и их окружения в реакционных центрах.

Существенной особенностью метода РСА является предварительная процедура *кристаллизации* белка, т.е. получение кристаллов с высокоупорядоченной периодической структурой. При облучении такого кристалла рентгеновскими лучами регистрируется дифракционная картина в виде набора отдельных рефлексов. Из-за того что возможности для регистрации слабых рефлексов увеличиваются как с ростом размера кристалла, так и с яркостью источника, широкое распространение получили эксперименты по определению структуры белков с помощью СИ. Сложности решения обратной задачи – определения 3D-распределения электронной плотности, хорошо известны. Прежде всего это неоднозначность точного решения обратной задачи рассеяния: определения распределения электронной плотности по набору интенсивностей отдельных рефлексов. Или так называемая “фазовая проблема” [22], когда необходимо определить амплитуду и фазу рассеяния, а в распоряжении экспериментатора есть только интенсивности отдельных рефлексов (модуль амплитуды).

Тем не менее разработаны надежные методики решения фазовой проблемы:

– *ab initio* определение фаз или прямые методы для малых молекул (<1000 неводородных атомов) с разрешением лучше, чем 1.4 Å,

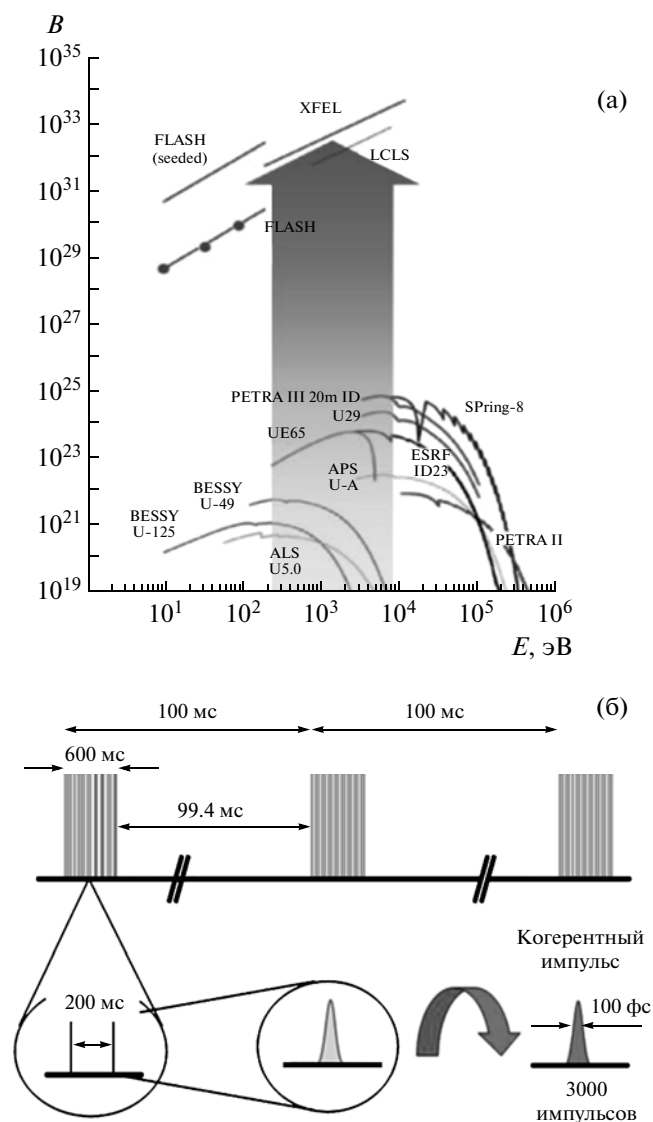
– молекулярного замещения,

– аномального рассеяния рентгеновских лучей (MAD или SAD phasing) вблизи краев поглощения, метод тяжелых атомов и др., анализ которых выходит за рамки настоящей работы.

Здесь можно упомянуть и новые фазочувствительные методы, такие как рентгеновская и фотоэлектронная голография [23], метод стоячих рентгеновских волн [24–26], позволяющие определить фазу рассеяния.

Отметим, что в [27] найден способ решения фазовой проблемы методом изоморфного замещения с помощью производных с тяжелыми атомами и продемонстрирована вся мощь метода рентгеновской дифракции на примере исследования структуры гемоглобина.

Сложной проблемой является подбор условий кристаллизации для выращивания биокристаллов высокого качества и большого размера. В настоящий момент кристаллизация таких сложных объектов, как БМ, не всегда дает 100%-ный результат, поэтому приходится перебирать массу вариантов условий кристаллизации (подбор растворов и осадителей и их концентраций), прежде чем удастся получить кристалл нужного качества и размера. Благодаря экспериментам в космосе установлено, что в условиях невесомости, когда явление конвекции можно исключить, удается выращивать более совершенные кристаллы [28]. Принципиально новые подходы к изучению свойств больших молекул и наночастиц появляются при использовании излучения РЛСЭ (X-ray Free Electron Laser, XFEL) [29, 30]. Сегодня действующие РЛСЭ – Linac Coherent Light Source (LCLS) в Стенфорде (США) и FLASH (Free-electron LASer in Hamburg) (ФРГ) начинают конкурировать с существующими источниками СИ: ESRF (European Synchrotron Radiation Facility) в Гренобле (Франция), Berliner Elektronenspeicherring-Gesellschaft für Synchrotronstrahlung (BESSY II – “Берлинское сообщество электронного накопительного кольца для синхротронного излучения”, ФРГ), SPring-8 (Super Photon ring – 8 GeV), с самой высокой в мире энергией – 8 ГэВ, в префектуре Хёго (Япония), APS (Advanced Photon Source) в Аргонской национальной лаборатории в Чикаго (США), Advanced Light Source (ALS) в Лоуренс Беркли, национальной лаборатории в Калифорнии (США), PETRA III – самым ярким источником СИ 3-го поколения в Гамбурге, и др. Благодаря яркости, в миллиарды раз превышающей яркость самых современных источников СИ (рис. 1а), и временной структуре импульса (рис. 1б) становится возможной регистрация нужного для определения структуры относительно сложных молекул числа рефлексов за один импульс. При этом отпадает необходимость в кристаллизации объектов, что особенно важно, на-



**Рис. 1.** Сравнение пиковой яркости  $B$  существующих источников СИ и XFEL (а); временная структура излучения XFEL характеризуется тремя уровнями: первый определяется скважностью импульсов – 30000 вспышек в секунду, внутри каждой вспышки можно выделить характерные временные структуры с периодом 200 нс – второй уровень, и отдельные “спайки” с длительностью 100 фс – третий уровень (б).

пример, при изучении мембранных белков, которые не растворяются в воде.

Однако при использовании излучения РЛСЭ возникает множество других проблем, которые до сих пор не решены. Например, из-за мгновенного выделения энергии под пучками РЛСЭ частицы становятся нестабильными. Необходимо научиться “доставлять” биомолекулы под пучок РЛСЭ (например, при помощи электронно-лучевой ловушки ионов, EBIT) [31, 32] и “сочленять” дифракционные картины, получающиеся при разной ориентации биомолекул [33]. Таким обра-

зом, одним из важнейших условий успешности эксперимента с РЛСЭ становится возможность зафиксировать пространственное положение молекулы, что, как показано ниже, становится вполне достижимым при использовании НП.

**Ядерно-магнитный резонанс.** Отметим, что интерпретация результатов ЯМР-исследований, основанных на поглощении электромагнитной энергии атомными ядрами с ненулевым спином, является достаточно сложной процедурой [34]. Раньше метод ЯМР использовался только при исследовании соединений с небольшой молекулярной массой, что было предопределено малой частотой первых магнитных ЯМР-спектрометров (~50 МГц). Эти приборы позволяли визуализировать отдельные специфические аминокислотные остатки. Первым белком, исследованным этим методом в 1957 г., была бычья панкреатическая рибонуклеаза [35].

Разрешение ЯМР-спектрометров существенно улучшилось с увеличением рабочей частоты спектрометра (220 МГц и выше), что позволило исследовать такие параметры, как химический сдвиг, площадь и ширину пика, и изучать не только структурное окружение, но и химические и колебательные свойства атомов в БМ, наблюдать структурные переходы и т.д.

Сегодня быстро развивается масс-спектрометрия сложных БМ [36]. Последние достижения в технике разделения белков с помощью двумерного гель-электрофореза и их последующий анализ с помощью масс-спектрометрии высокого разрешения создали предпосылки для дальнейшего развития структурной геномики, получившей название функциональной протеомики, которая исследует механизмы функционирования клетки в динамическом режиме [37].

Несмотря на активное применение ЯМР и масс-спектрометрических методик, основным методом, используемым сегодня для расшифровки структуры белков и других биомолекул, остается РСА.

С внедрением “фабрик по кристаллизации белка” [38] деятельность по выделению и наработке белка в количествах, достаточных для кристаллизации, приобретает “промышленный” характер. Процесс восстановления структуры белков по рентгendifракционным данным, полученным с использованием источников СИ, все более “стандартизируется” [39–41].

Деятельность по исследованию белков ведется в рамках масштабных программ (например, Protein Structure Initiative) и в основном ориентирована на определение структуры максимального числа белков, даже если функции соответствующих белков неизвестны [42]. Однако становится понятно, что изучение структуры ради процесса изучения (к тому же с низким разрешением) в на-

дежде, что кто-то позднее воспользуется этими данными, контрпродуктивно. Никто не будет [43] моделировать структуру белка, используя гомологию со структурами, хранящимися в таких хранилищах данных, как Protein Data Bank [44], если перед ним не стоит конкретная биологическая задача.

*Другие физические методы.* Между тем разработка и совершенствование хорошо известных и относительно простых методов изучения структуры макромолекул в растворах никогда не прекращались. К ним относятся прежде всего гидродинамические методы, основанные на анализе диффузии, седиментации и вязкости. Теоретические и экспериментальные основы этих подходов были заложены задолго до появления работ Крика и Уотсона [45].

В работах А. Стокса, Ф. Перрена и А. Эйнштейна показано, что сила трения в среде, действующая на молекулу, может быть определена из опытов по седиментации и диффузии. Это дает возможность вычислить “гидродинамические” размеры белков, которые зависят от конформации молекулы в растворе и, например, ионной “шубы”, возникающей вокруг молекулы в заряженных ионных растворах. Известно простое уравнение Марка—Куна—Хаувинка [46], связывающее характеристическую вязкость полимера в растворе и его молекулярную массу. Более того, при сочетании методов седиментации и диффузии можно определить молекулярные массы биомолекул.

С помощью метода седиментации получены принципиально новые результаты о структуре рибосом, изменении размеров молекулы гемоглобина при окислении (так называемом оксигенировании) и др. Измерение характеристической вязкости растворов биологических макромолекул позволило изучить эффекты ориентации макромолекул в гидродинамическом (эффект Максвелла) и электрическом поле (эффект Керра) [47]. При этом возникла относительно простая классификация биологических макромолекул (в частности, белков и вирусов) по трем классам. Первый класс составляют глобулярные белки и сферические вирусы, характеристическая вязкость которых не зависит от молекулярной массы и составляет  $4\text{--}6\text{ см}^3/\text{г}$  [48]. Второй класс – фибриллярные белки и палочкообразные вирусы, характеристическая вязкость которых быстро увеличивается с ростом молекулярной массы ( $100\text{--}1000\text{ см}^3/\text{г}$  в случае коллагенов) и степени “вытянутости” (асимметрии). К третьему классу можно отнести белки любой формы в “универсальном растворителе” (8 М мочевины или 6 М гуанидингидрохлорида с добавкой вещества, разрывающего дисульфидные связи), в котором молекула принимает конформацию развернутого беспорядочного клубка.

Помимо гидродинамических методов, начиная с XIX в., широко использовалась поляризованная флуоресценция, изучение которой было начато в работах Ф. Перрена и Г. Вебера. При флуоресценции между актами поглощения электромагнитного кванта квазиоптического диапазона и его испусканием всегда проходит конечное время ( $10^{-6}\text{--}10^{-11}\text{ с}$  в случае излучательного перехода с самого нижнего синглетного колебательного уровня в основное состояние), которое по порядку величины сравнимо со временем вращательной диффузии молекул глобулярных белков. Совпадение характерного времени позволяет определить коэффициенты вращательного трения БМ методом поляризованной флуоресценции как в статическом, так и в импульсном режимах.

Время жизни возбужденного состояния в случае аналогичных процессов при рентгеновской флуоресценции гораздо меньше ( $10^{-18}\text{ с}$  вдали от К-края поглощения и на несколько порядков меньше в условиях резонанса), поэтому можно считать, что испускание флуоресцентного рентгеновского кванта происходит за время, много меньшее времени прохождения биомолекулы через НП.

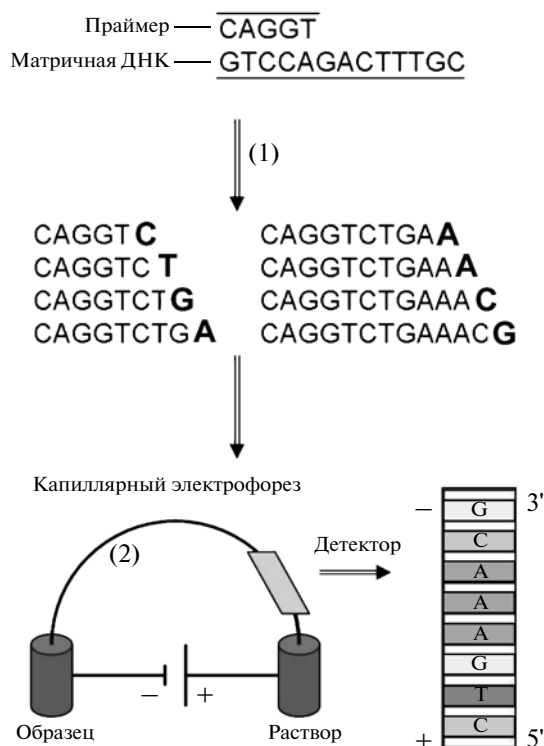
Принципиально новый шаг в развитии флуоресцентных методов был сделан Т. Ферстером, Л. Стриером и Р. Хогландом, которые показали, что, используя специальным образом подобранную пару молекул донор–акцептор, можно, исследуя особенности переноса энергии электронного возбуждения, определить расстояние между ними [49, 50]. Такая “*флуоресцентная линейка*” позволила определять расстояния в несколько десятков ангстрем между мечеными частями молекулы [51] и применяется сегодня для исследования самых разных биологических систем.

Использование высокоскоростной аппаратуры и импульсных лазеров позволило исследовать динамические процессы в окружении люминофора с высоким временным разрешением [52].

Регистрация спада интенсивности флуоресценции методом счета единичных фотонов с применением корреляционного анализа, использование лазеров на кристаллах сапфира с примесью титана, микроканальных пластинчатых фотоэлектронных фотоумножителей позволили достичь временного разрешения в несколько пикосекунд [53].

С развитием микрофлюидики основное направление исследования снова смещается в сторону гидродинамических методов.

В конце 70-х гг. прошлого века Ф. Сангер показал, что можно определить последовательность оснований молекулы ДНК, расчленив ее на отдельные части и используя химические методы разделения получившихся отрезков [54]. Сегодня эти исследования полностью автоматизированы,



**Рис. 2.** Матричная ДНК и праймер; 1 – ДНК-полимераза выстраивает последовательность, начиная с праймера и добавляя нуклеотиды к концу (3'). После встраивания дидезоксинуклеотида рост цепи прекращается. После денатурации цепочки ДНК различной длины разделяются с помощью капиллярного электрофореза.

и число установленных последовательностей растет каждый год. Фактически моментом рождения новой науки, *геномики*, стал 1995 г., когда был впервые расшифрован бактериальный геном *Mycoplasma genitalium* [55].

После десятилетий исследований [56] достигнут значительный прогресс в развитии методов секвенирования ДНК и определении структуры белков, позволивших на основе структурных данных расширить возможности предсказания наследственных заболеваний. Если стоимость первого расшифрованного генома человека [57–59] составляла десятки миллионов долларов США, а еще несколько лет назад – сотни тысяч, то уже через несколько лет она должна снизиться до 1 тысячи долларов. Сопоставление диагностируемых заболеваний и структуры генома пациента позволит связать последовательности (или точнее мутации) генов с наследственными заболеваниями, а также выяснить механизмы их инициации в зависимости от пола, возраста, региона проживания и расы. Геном каждого человека можно будет проверить на наличие известных генетических дефектов, что позволит предсказать предрасположенность к тем или иным заболеваниям и провести комплексное лечение. Персональная медици-

на как ожидается, станет гигантским прорывом в области диагностики и лечения в XXI в.

Тем не менее развитие прямых технологий определения нуклеотидов в ДНК “по Сангеру” сталкивается с многочисленными физическими и технологическими проблемами, что заставляет рассматривать принципиально новые подходы, которые при меньшей стоимости были бы более эффективными и точными [60, 61].

Одной из таких альтернатив является новая концепция, заключающаяся в использовании небольших, размерами порядка нанометра, пор для определения структурных и/или физических характеристик различных оснований ДНК. В отличие от существующих технологий, основанных на использовании химической техники дробления длинных биомолекул и детектировании отличий между получившимися цепочками ДНК при различных химических реакциях и физических процессах, в данном случае исследуется последовательность на всей доступной длине молекулы.

## 2. ПРЯМЫЕ ТЕХНИКИ СЕКВЕНИРОВАНИЯ

В [56, 59, 62] подробно описана технология “прямой” техники секвенирования. Современные методы секвенирования являются улучшенными вариантами метода Сангера, впервые примененного в 1977 г. [63].

В процессе прямого секвенирования можно выделить четыре этапа [56]:

- локализация молекулы ДНК, денатурация двухцепочечной молекулы;
- “разрезание” исследуемой цепочки на короткие участки длиной в несколько десятков нуклеотидов;
- непосредственное секвенирование каждого участка в отдельности;
- анализ полученных данных.

Третий этап схематически представлен на рис. 2. Методом химической элонгации получают цепочки ДНК с произвольно встроенными дидезоксинуклеотидами, которые участвуют в росте цепочки ДНК, однако после присоединения замыкающей метки рост цепи сразу останавливается. Впервые такой подход был предложен в [63]. Далее после электрофореза на пористой матрице происходит пространственное разделение цепочки ДНК на отрезки различной длины. Последнее основание каждой цепочки идентифицируется по лазерному возбуждению флуоресцентных меток, позволяющему распознать тип нуклеотида.

На четвертом этапе происходит заключительная обработка данных секвенирования, состоящая из “переборки” коротких последовательностей и восстановления полной последовательности оснований в исходной цепочке ДНК. Для успешной реконструкции методами “биоинфор-

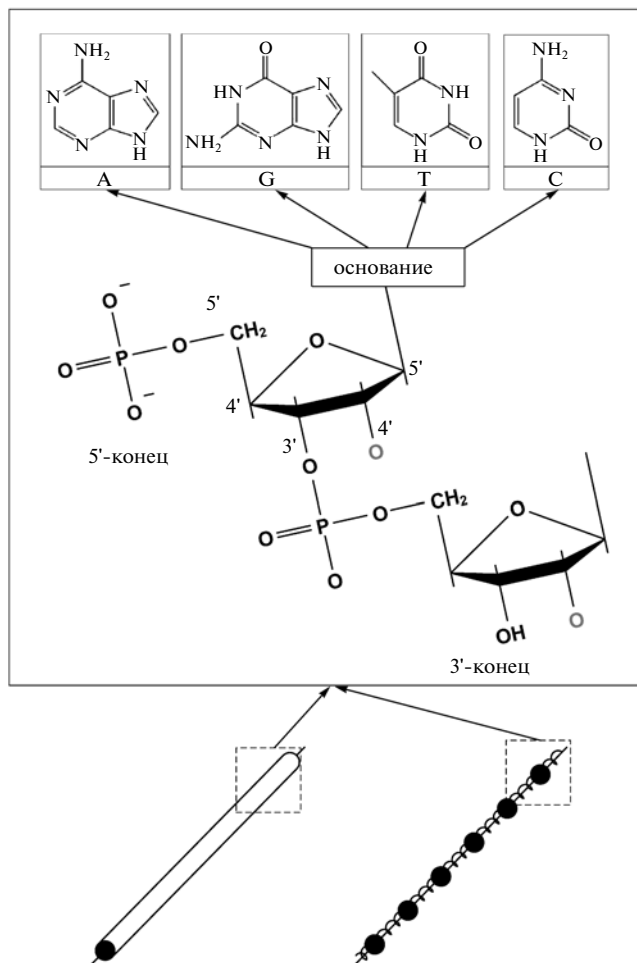
матики” необходимо, чтобы последовательности оснований в коротких цепочках перекрывались.

Прямые техники секвенирования требуют тщательной подготовки образцов и кропотливой обработки полученных данных. Основная трудность связана с тем, что существует верхний предел считываемой длины [56], при котором процесс секвенирования дает воспроизводимые результаты. Считываемая длина также ограничена чувствительностью выбранной методики электрофореза при разделении одноцепочечных ДНК с разными длинами. Очевидно, что процесс приговления образцов и обработка данных значительно упростятся, если станет возможным секвенировать более длинные цепи ДНК.

### 3. ДИНАМИКА ДНК

Молекула ДНК является одной из наиболее важных БМ и служит для хранения и передачи генетической информации. “Развернутая” молекула ДНК представляет собой сложную динамическую систему, состоящую из тысяч нуклеотидов с квазиодномерной структурой с несколькими степенями свободы (колебательными, вращательными и др.) и сложной симметрией. При этом задача описания ее движения на больших отрезках времени с использованием методов классической молекулярной динамики становится весьма сложной, а *ab initio* методами квантовой механики – практически не решаемой, даже при использовании современных компьютеров. Поэтому обычно используют те или иные приближения, описывающие динамику биомолекулы и ее взаимодействие с раствором и НП с некоторой точностью.

Известно, что молекула ДНК обладает квазирегулярной химической структурой. Регулярная часть (сахарофосфатный остов) образована строго чередующимися сахарами и фосфатными группами (рис. 3), объединенными в регулярно чередующиеся 3'-, 5'-фосфат-диэфирные звенья, а нерегулярная часть (боковые группы) формируется основаниями (рис. 3), соединенными с сахарами и образующими последовательность вдоль цепи. Эта последовательность определяет первичную структуру ДНК. Основания, входящие в состав нуклеотидов, разделяют на две группы: *пурины* (**Pu**) (*аденин* (*A*) и *гуанин* (*G*)) и *пиримидины* (**Py**) (*цитозин* (*C*) и *тимин* (*T*)). Известно, что некоторые физические свойства зависят от того, к какому семейству принадлежит основание. К примеру, *Pu* больше по размеру, следовательно, заранее можно предположить, что они будут сильнее взаимодействовать с поверхностями НП. Одним из наиболее важных свойств ПН является то, что они заряжаются в растворе. Для молекулы в среде раствора (без учета эффектов экранирования ионами раствора) можно считать, что каждый



**Рис. 3.** Две повторяющиеся структурные единицы сахарофосфатного остова ДНК, фосфатная группа отрицательно заряжена в растворе; *P* – атом углерода сахара связан с основанием, остальные атомы углерода помечены цифрами до 5'. Каждая цепочка ДНК может заканчиваться либо концом 3', либо 5'. Основания, образующие молекулу ДНК: *A* – аденин, *G* – гуанин, *T* – тимин, *C* – цитозин. Показаны простая структурная модель ДНК и ее дискретный аналог.

участок нуклеотида отрицательно заряжен. Благодаря этому становится возможным управление движением биомолекулы при помощи электрического поля (в частности электрофорез). При наличии мембраны с наноразмерным отверстием молекула ДНК, находящаяся в растворе по одну сторону от мембраны, пройдет через НП при наложении достаточно большого внешнего электрического поля.

Чтобы описать картину динамики молекулы ДНК, удобно классифицировать внутренние степени движения по значениям энергии (*E*), амплитуды (*A*) и характерного времени. В качестве примера в таблице представлена классификация, предложенная в [64].



Классификация экспериментальных и теоретических данных о внутренней подвижности ДНК [64]

Временная шкала	Пс	Нс	мкс	мс	с
Основные виды внутренних движений и их амплитуды ( $A$ )	Колебания отдельных атомов $A \cong 10^{-1} \text{ \AA}$	Колебания небольших групп атомов (сахара, фосфаты, основания); поперечные и вращательные движения элементов двойной цепочки с $A = 5-7 \text{ \AA}$	Изгибные движения; раскручивание и закручивание двойной спирали; раскрытие пар оснований	Движение молекулы ДНК как целого	Изомеризация, деление бактерий $A = 2-3 \text{ мкм}$
Энергия активации; возможные источники энергии	$E = 0.6 \text{ ккал/моль}$ ;	$E = 2-5 \text{ ккал/моль}$ ; столкновения с “горячими” молекулами раствора	$E = 5-20 \text{ ккал/моль}$ ; увеличение температуры, действие денатурирующих объектов	$E = 10-50 \text{ мкм/моль}$ ; взаимодействие с белками и другими лигандами	
Экспериментальные методы	ЯМР, инфракрасная спектроскопия, Рамановская спектроскопия, рентгеноструктурный анализ	ЯМР, ЭПР, Рамановская спектроскопия	ЯМР, ЭПР, водородный обмен	ЯМР, водородный обмен, гидродинамические методы	
Теоретические методы	Гармонический анализ, решетчатая модель, методы МД	Гармонический анализ, модель стержня, методы МД	Теория перехода спираль-клубок	Методы конформационной механики; топологические модели	

В первом приближении можно считать, что в основании молекулы ДНК находится длинная сахарофосфатная цепочка с точно повторяющимся рисунком атомов вдоль цепи атомов, которая обладает почти регулярной структурой (“скелет”). С этой точки зрения молекула ДНК имеет сходство с одномерными периодическими структурами, так называемыми квазиодномерными кристаллами.

Простейшая макроструктурная модель ДНК — однородный эластичный стержень с круговым сечением. Дискретный микроаналог стержневой модели состоит из цепочки связанных дисков (или шариков), каждый из которых имитирует фрагмент молекулы ДНК, содержащий одну пару оснований (рис. 3).

Исторически для описания макроскопических движений биополимера используется континуальная модель балки [65], а для описания возбуждений молекулы ДНК на микроуровне — ансамбль фононных колебаний.

В линейном приближении (справедливом для малых амплитуд смещений) нормальные фононные моды являются независимыми. Прохофски и соавторы [66–68] применили гармоническое приближение для описания возбуждений молекулы ДНК. Результаты вычислений, выполненных в линейном приближении, совпали с экспериментальными данными по скорости продольного звука [69], упругому рассеянию нейтронов [70–

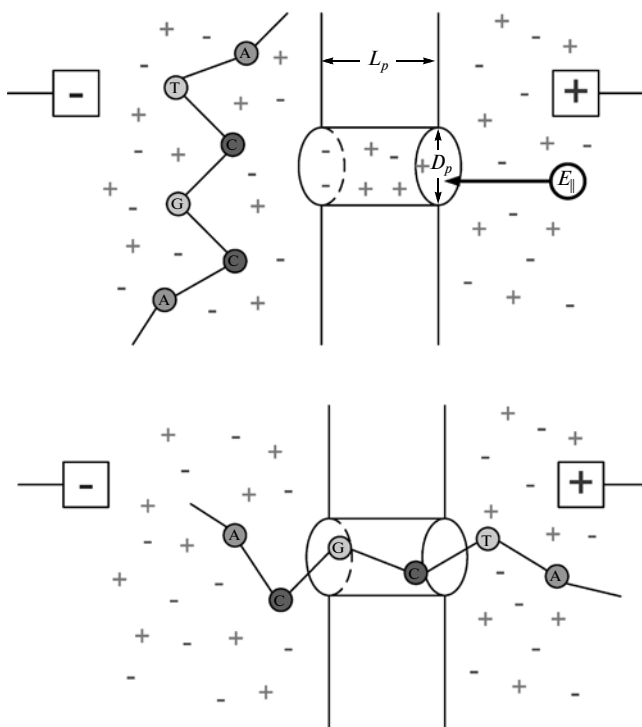
72], особенностям рамановских пиков [73, 74], низкочастотному спектру возбуждений ДНК [75–78] и данным экспериментов ЯМР [79–81].

#### 4. НАНОПОРЫ И ТРАНСЛОКАЦИЯ ПОЛИНУКЛЕОТИДОВ

Около десяти лет назад впервые удалось “протянуть” одноцепочечный ПН через биологическую НП [60], помещенную в ванну с раствором. Для этого к поре прикладывали напряжение, что помогало заряженной основе ПН “пролезть” через пору. Транслокация (перемещение) ПН регистрировалась по возникновению *блокирующего тока*. С тех пор НП активно используются для получения различной информации, характеризующей как сам процесс транслокации, так и динамику и взаимодействия одно- и двухцепочечных ПН в нанопоре [60, 61, 82, 83]. Дальнейший прогресс связывают с созданием синтетических НП с изменяемыми параметрами [84–87].

Для описания процесса транслокации ПН через пору обычно используют феноменологические подходы [88], объясняющие зависимость основных динамических характеристик от конечного набора различных параметров (длина полимера, размеры поры, прикладываемое напряжение и т.д.).

*Основная идея эксперимента.* В случае с длинными биомолекулами самым трудным было понять, почему ПН могут эффективно “проходить”



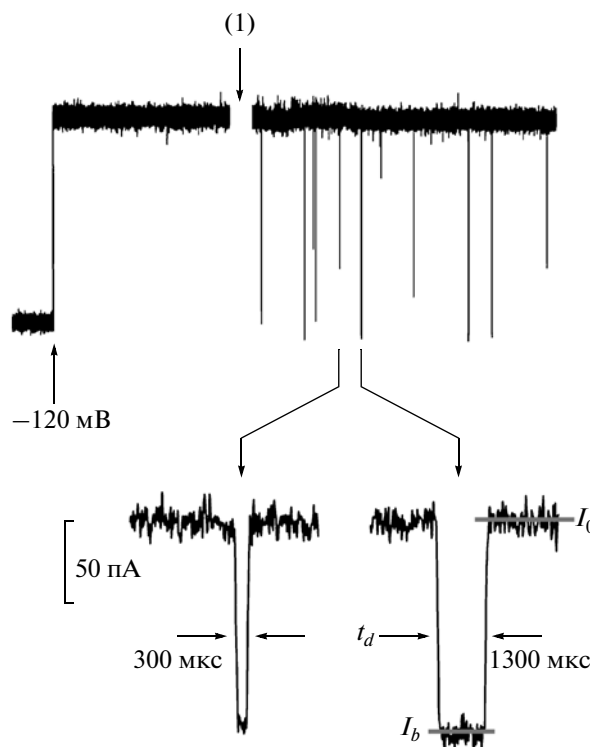
**Рис. 4.** Схема эксперимента по транслокации полинуклеотида через НП; а – разность потенциалов на концах мембраны создает электрическое поле  $E_{||}$ , сквозь пору течет ионный ток; б – поле “захватывает” в пору отрицательно заряженную основу ПН. Когда молекула ДНК локализована внутри поры, ионный ток уменьшается;  $L_p$ ,  $D_p$  – длина и средний диаметр поры, соответственно [60].

через пору. Схематично процесс изображен на рис. 4, где мембрана разделяет раствор в ячейке на две части, в каждую из которых помещены электроды. В отсутствие ПН электроды перемещают ионы раствора, создавая ионный ток  $I_0$  в открытом канале.

Благодаря отрицательно заряженным фосфатным группам ПН движется к катоду. В результате цепочка *захватывается* и проходит сквозь пору длиной  $L_p$  и средним диаметром  $D_p$ . В течение времени  $t_d$ , *периода транслокации*, цепочка частично блокирует ионы, не давая им проходить через пору (рис. 4).

Хотя нуклеотид заряжен, при его прохождении генерируется значительно меньший ионный ток через пору вследствие своей малой по сравнению с ионами скорости. Таким образом, о наличии ПН в поре сигнализирует резкое уменьшение блокадного тока  $I_b$  (рис. 5), что делает возможным регистрацию процесса транслокации, а также получение косвенной информации о длине ПН, его составе и динамических процессах, например об увеличении длины ПН из-за растяжения.

В настоящее время для проведения исследований используются два основных типа НП: биоло-



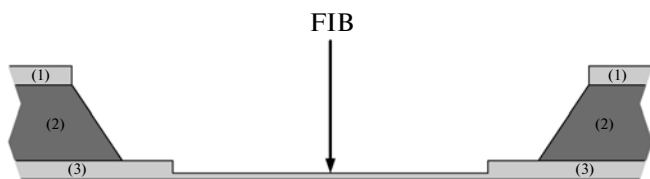
**Рис. 5.** В верхней части: сквозь пору течет ионный ток, напряжение 120 мВ. Когда ПН вносят в раствор (1), возникают блокадные события. В нижней части: примеры блокадных событий;  $I_0$  – среднее значение тока открытой поры,  $I_b$  – среднее значение тока блокады,  $t_d$  – среднее время транслокации [60].

гические и искусственные. Биологические поры (в основном поры  $\alpha$ -гемолизина, [89]) имеют поперечный размер в несколько нанометров и почти идеально подходят для транслокации цепей ПН. Однако работа с биологическими порами осложнена тем, что они “открыты” только при определенных условиях окружающей среды (температура, концентрация ионов), а изменить геометрические размеры такой поры практически невозможно.

### 5. СИНТЕТИЧЕСКИЕ ПОРЫ

Синтетические поры открывают дополнительные возможности для секвенирования ПН, делая возможным изменение размеров и свойств поры с учетом специфики ПН. Кроме того, возможности встраивания поверхностных сенсоров, таких как поперечные электроды, становится доступным более широкий диапазон изменения параметров эксперимента (концентрация солей, температура, напряжение, вязкость), которые напрямую влияют на скорость транслокации и захвата ПН [83, 90].

Процесс создания НП схематически представлен на рис. 6. В [91] показано, что, используя фо-



**Рис. 6.** Схема создания синтетической НП в тонкой мембране с помощью FIB; 1, 3 – слой SiO<sub>2</sub> толщиной 30 нм; 2 – слой Si толщиной 300 нм.

кусированный ионный пучок (**FIB**), можно “про-сверлить” отверстие в тонкой мембране диоксида кремния (толщина мембраны ~10 нм). Пластина кремния с двух сторон была покрыта слоем диоксида кремния толщиной 30 нм. Используя метод электронно-лучевой литографии и реактивного травления ионами, в слое оксида “открывают” площадку размером 1 × 1 мкм. Затем протравливают пластину кремния до нижнего слоя оксида, используя жидкостное травление с КОН, утончая мембрану диоксида кремния до толщины 30 нм. Используя FIB, можно добиться утоньшения мембраны и до 10 нм. Фокусированный ионный пучок (с размерами пятна порядка нескольких нанометров) используют для сверления пор в тонкой мембране диоксида кремния. Плотность тока в пучке ~10<sup>8</sup> А/м<sup>2</sup>, диаметр поры ~6 нм.

Полученную НП диаметром 6 нм можно “за-растить” до требуемого размера (вплоть до 1 нм), используя методику, описанную в [84]. На пору направляют пучок ионов Ag<sup>+</sup>, активирующий процесс диффузии, под влиянием которого пора начинает закрываться. Ток ионов Ag<sup>+</sup>, проходящих через пору, прямо пропорционален площади поры. Измеряя ток как функцию времени [84], можно контролировать сужение поры до необходимого размера.

Для проведения эксперимента НП помещают между двумя камерами, так что ионный раствор может протекать только через нее. При наложении напряжения отрицательно заряженная молекула ДНК под действием электрического поля проходит сквозь НП. Наличие молекулы внутри поры определяется по возникновению так называемого блокадного события. Время транслокации  $t_d$  и падение тока  $\Delta I_b$  зависят от свойств раствора (концентрация ионов, вязкость и температура), характеристик НП, прикладываемого напряжения и свойств молекулы. В процессе транслокации движущие силы (электростатическая и гидродинамическая) преодолевают силы вязкостного сопротивления и взаимодействуют со стенками самой НП. В случае ионного раствора для тока справедлив закон Ома:

$$\Delta I_b = \sigma V A_{DNA} / H, \quad (1)$$

где  $\sigma$  – проводимость раствора,  $V$  – приложенное вдоль поры напряжение,  $H$  – толщина поры,  $A_{DNA}$

– поперечное сечение проходящей сквозь пору молекулы. Время транслокации определяется выражением

$$t_d = K \eta L_{DNA} / \lambda V, \quad (2)$$

где  $\eta$  – коэффициент вязкости,  $\lambda$ ,  $L_{DNA}$  – линейная плотность зарядов и длина участка ДНК соответственно,  $K$  – коэффициент пропорциональности, расчет которого выходит за рамки этой простой модели. При этом  $\sigma$  зависит от  $\eta$  ( $\sigma \sim 1/\eta$ ), а  $\sigma$ ,  $\eta$  и  $\lambda$  зависят от температуры и концентрации ионов.

Блокадный ток обратно пропорционален вязкости (так как  $\sigma \sim 1/\eta$ ), а время транслокации линейно растет с вязкостью (2), что подтверждается экспериментально [82], и позволяет увеличить время транслокации посредством изменения вязкости раствора. Блокадный ток пропорционален приложенному напряжению, в то время как время транслокации обратно пропорционально. При характерном напряжении в 20 мВ ток составлял ~22 пА, но его падение при транслокации ДНК может быть легко измерено [82].

В [82] показано, что при снижении температуры от 22 до 4°С блокадный ток уменьшился с 210 до 140 пА, а время транслокации возросло от 165 до 280 мкс. Таким образом, увеличивая вязкость раствора в 2 раза (например, добавив глицерин), можно уменьшить скорость транслокации примерно в 5.5 раз. Комбинацией всех этих факторов в конкретном эксперименте можно добиться уменьшения скорости транслокации на порядок величины, что важно для регулирования времени задержки биомолекулы под пучком рентгеновского излучения в НП.

## 6. МЕТОДИКИ СЕКВЕНИРОВАНИЯ

“Электронные” методы детектирования типа полинуклеотида (схема на рис. 7) основаны на измерении либо ионного блокирующего тока в НП, [60, 92], либо поперечных характеристик молекул в НП, например поперечного тока между системами электродов при туннелировании через молекулу в НП [88, 93].

Эксперимент по исследованию ионного блокирующего тока описан в [94–100]. Отметим, что подобные эксперименты не позволяют идентифицировать одиночное основание, что необходимо для секвенирования. Существует отдельное направление исследований, использующее “измененные” поры [101, 102]. Когда в такую пору вводят БМ (например, циклодекстрин [103]) (рис. 7а), она по-разному взаимодействует с различными основаниями ДНК. Механизм действия такой системы секвенирования основан на химических реакциях, но не будем здесь останавливаться на нем подробно.

В методах, связанных с измерениями поперечного электронного сигнала, детектирование сигнала от отдельных оснований происходит с помощью вмонтированных в НП электродов, измеряющих поперечный ток через основания молекулы ДНК [104]. Такой метод в принципе позволяет определить электронную структуру оснований (так как туннельный ток зависит от особенностей электронных орбиталей проходящего нуклеотида), хотя получаемые результаты сильно зависят от конструкции электродов и системы усиления сигналов. Схема подобного опыта показана на рис. 7б. Подчеркнем, что расчеты “поперечного транспорта” обычно ведутся в рамках квантовой механики и сильно расходятся с результатами экспериментов по определению типа оснований [105–107].

Тем не менее теоретические расчеты и эксперименты для специфических нуклеотидов, находящихся между двумя электродами внутри НП, показали, что основания можно идентифицировать по значению поперечного тока [104]. Различия в значениях тока при прохождении конкретного основания зависят от расстояния между электродами и значительно меняются, если нуклеотиды могут свободно ориентироваться в пространстве. При характерном зазоре между электродами в 1.5 нм и при вертикальном положении нуклеотида относительно электродов токи для разных нуклеотидов различаются на порядок, вероятно, вследствие специфичности взаимодействия между основаниями и электродами [104].

Эксперименты показали, что поперечные токи довольно сильно отличаются для оснований *A* и *T*, а для оснований *G* и *C* – практически одинаковы, что затрудняет эффективное секвенирование [104]. При этом длина (вдоль НП) электрода должна быть меньше (или порядка) расстояния между основаниями в цепочке. Тогда для секвенирования одноцепочечной ДНК необходимо научиться делать очень тонкие мембраны толщиной всего около 7 Å.

Необходимость использования туннельного эффекта имеет обратную сторону: уже при зазоре между электродом и молекулой в 1.5 нм значение тока очень мало (~1 пА для основания *A*). Однако если сделать пору совсем узкой, то движение биомолекулы прекратится. Для повышения эффективности измерений необходимо использовать более узкие зазоры, наносить дополнительные слои и учитывать сложную гидродинамику течения через пору.

Для учета этих эффектов было проведено большое число экспериментов и попыток моделирования методами молекулярной динамики (МД) прохода одноцепочечной ДНК через пору [62, 108]. Контролировать поворот молекулы ДНК в пору достаточно сложно и ее ориентацию в пространстве можно считать случайной, что

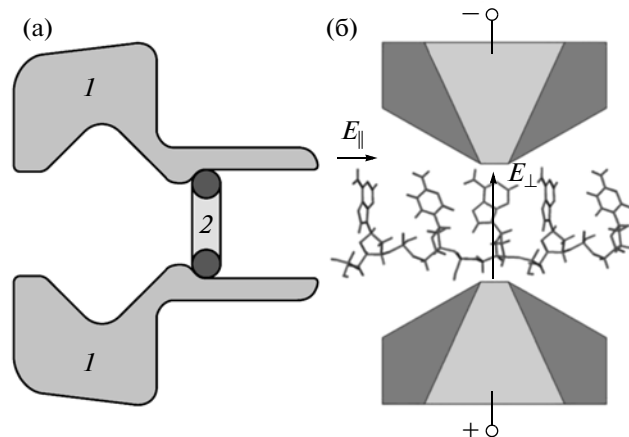


Рис. 7. Схема поры  $\alpha$ -гемолизина (1) с заряженной молекулой циклодекстрина (2) (показан продольный разрез) для расшифровки нуклеотидов посредством их ионной блокады (а). Схема двух электродов, монтированных в НП. Когда ДНК проходит через пору, электроды создают ток через нуклеотиды в поперечном направлении.  $E_{\parallel}$  – продольное поле, перемещающее ДНК через пору;  $E_{\perp}$  – поле, перпендикулярное поверхности электродов (б).

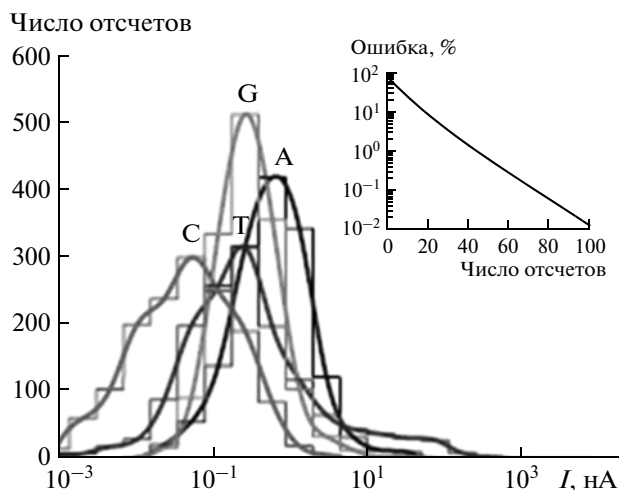
приводит к флуктуациям тока, по порядку величины, сравнимым с измеряемым значением.

Моделирование методами МД показывает, что относительно слабое поперечное поле может ориентировать нуклеотид (и всю цепочку) относительно системы электродов за меньшее время, чем потребуется нуклеотиду, чтобы пройти через соединение.

Если поперечное поле удовлетворяет условию  $E_{\parallel} \ll E_{\perp}$ , то ДНК будет медленно проходить через пору, причем поперечное поле успеет сориентировать нуклеотиды, пока они находятся в поле электродов. При таком контроле положения нуклеотидов можно составить “таблицу” секвенирования для данного типа поры, с помощью которой идентифицируются электрические характеристики для каждого нуклеотида (рис. 8). Помимо этого результаты зависят от выбора раствора, внешних условий, используемых методов детектирования и усиления сигнала и т.д.

При секвенировании ДНК протаскивать молекулу через пору нужно с такой скоростью, чтобы каждый нуклеотид оставался в области электрода в течение времени, достаточного для измерения тока и набора статистики. Оценки показывают, что на секвенирование ДНК одного человека по существующей технологии будет затрачиваться около 7 ч (без учета компьютерного анализа данных) [62].

Технологические трудности при реализации методики секвенирования с помощью НП с вмонтированными электродами обсуждались в [109]. В [110, 111] для контроля положения нуклеотидов было предложено использовать методы сканирующей туннельной микроскопии.



**Рис. 8.** Распределение токов для различных нуклеотидов (A, T, G, C) при зазоре между электродами в 1.25 нм и напряжении 1 В. Вставка дает число ошибок в идентификации оснований в зависимости от числа независимых отсчетов [108].

## 7. КОМБИНИРОВАННЫЕ МЕТОДЫ С ИСПОЛЬЗОВАНИЕМ РЕНТГЕНОВСКОГО И СИНХРОТРОННОГО ИЗЛУЧЕНИЙ

Одним из первых методов определения структуры протеинов и нуклеиновых кислот и их реактивной способности с использованием СИ стал

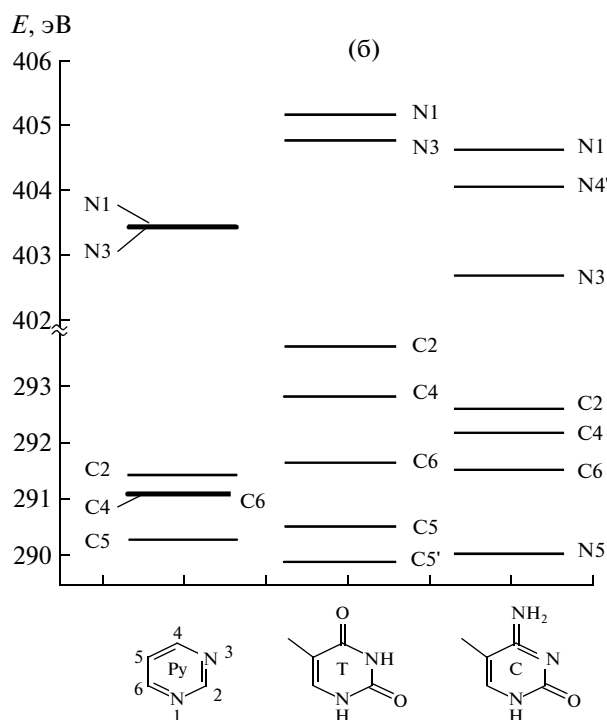
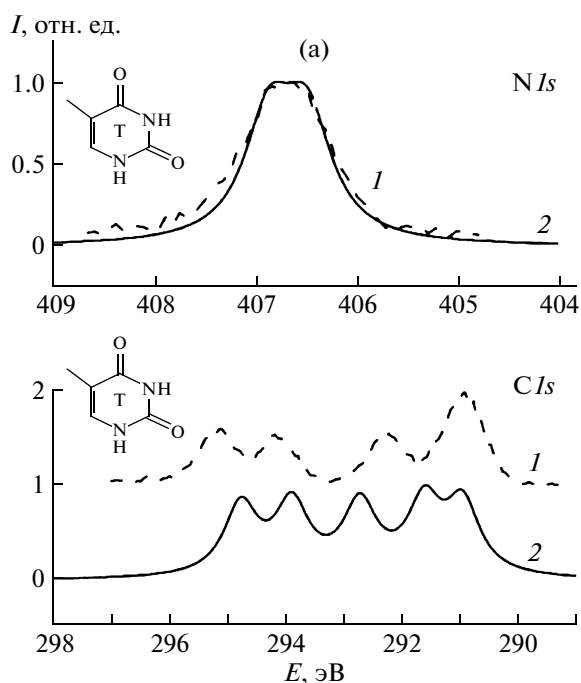
метод “сличения под пучками СИ” (Synchrotron Footprinting) [112].

Главное преимущество СИ – широкий спектральный диапазон, позволяющий провести выборочное возбуждение атомов определенного типа. Поэтому основные проблемы при проведении спектральных экспериментов связаны с использованием детекторов флуоресцентного сигнала и системами детектирования фотоэлектронов с высоким спектральным разрешением.

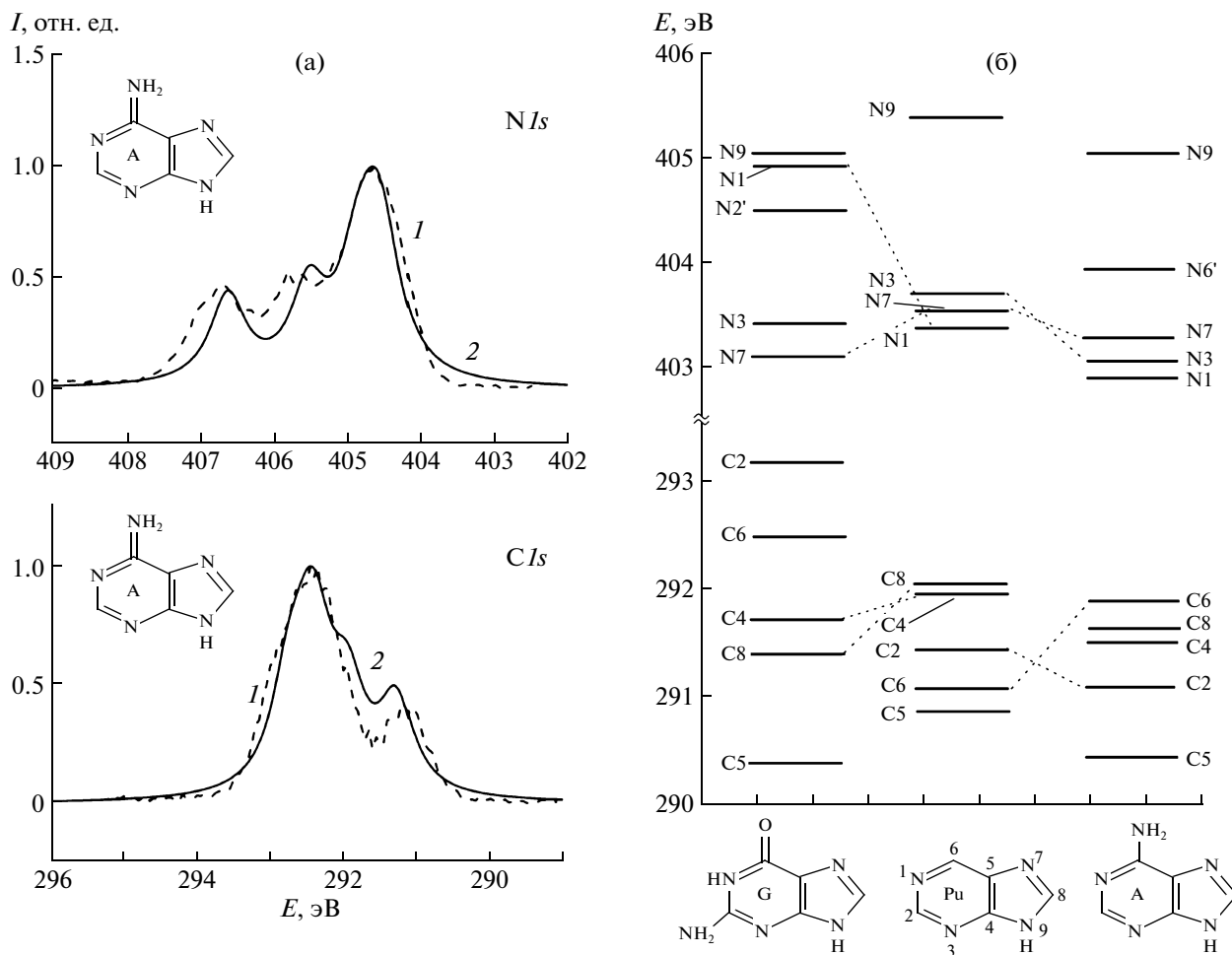
Анализ требуемого спектрального разрешения обычно основан на расчетах электронных состояний в различных основаниях молекулы ДНК. Для примера приведем потенциалы ионизации внутренних оболочек оснований ДНК, вычисленные с использованием модели LB94/TZ2P [113].

На рис. 9 представлена энергетическая диаграмма для атомов N1s и C1s Py-оснований, из которой видно, что изменения энергии атома в зависимости от его положения по отношению к незамещенному Py существуют. Пара атомов азота в незамещенном Py находится в одинаковом химическом окружении, поэтому атомы N(1) и N(3) полностью вырождены по энергии. Аналогичная ситуация складывается для атомов C(4) и C(6). В случае C и T подобная эквивалентность нарушается, и вырождение снимается.

На рис. 10 представлена энергетическая диаграмма для N1s и C1s пуриновых оснований.



**Рис. 9.** Сравнение экспериментальных данных, полученных на синхротроне (1) и теоретического расчета (2) флуоресцентного спектра тимина (T) (a); спектроскопическая диаграмма характерного флуоресцентного сигнала Py и его оснований (T, C) (б) [113].



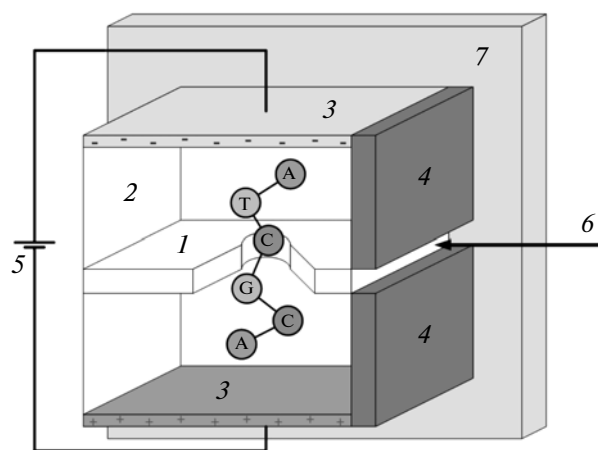
**Рис. 10.** Сравнение экспериментальных данных, полученных на синхротроне (1) и теоретического расчета (2) флуоресцентного спектра аденина (A) (a); спектроскопическая диаграмма характерного флуоресцентного сигнала Pu и его оснований (G, A) (б) [113].

Потенциалы ионизации отдельных атомов различных оснований различаются на величину  $\sim 1$  эВ при среднем значении потенциала ионизации 300 эВ, соответственно разрешение детектора должно быть лучше, чем

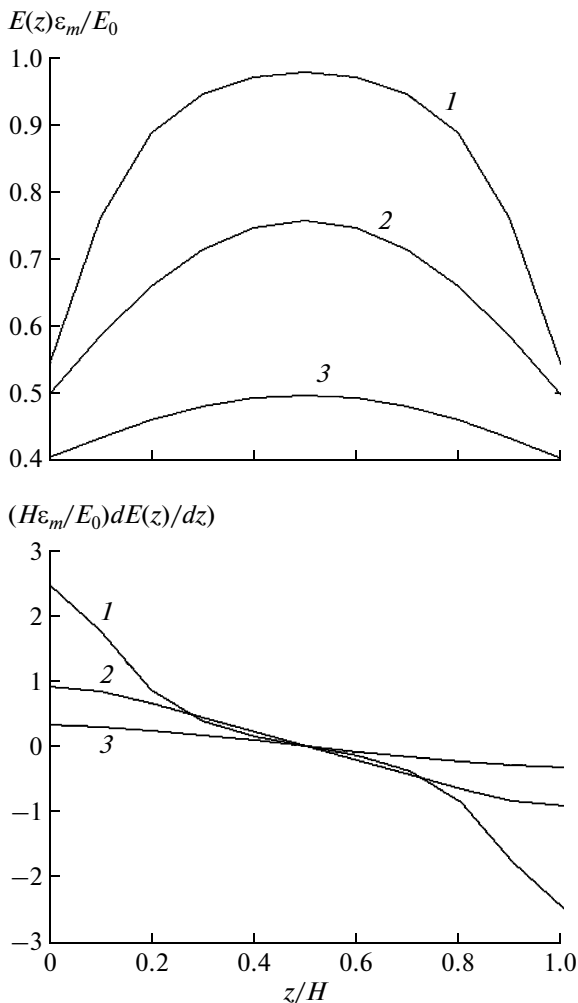
$$\Delta E/E < 10^{-3}.$$

Вопросы регистрации флуоресцентного и фотоэмиссионного сигналов от отдельных оснований будут рассмотрены в следующих работах. При этом мембрана, в которой расположена НП, может использоваться не только для пространственной локализации оснований (рис. 11), но и как рентгеновский волновод, позволяющий направить излучение на отдельное основание.

В последующих частях работы будет дано детальное описание структуры молекулы в ионном растворе в электрическом поле; динамики транслокации биомолекулы через НП; рентгеновских флуоресцентных и фотоэмиссионных методов идентификации оснований ДНК.



**Рис. 11.** Схема, иллюстрирующая возбуждение вторичного флуоресцентного излучения при транслокации молекулы ДНК: 1 – тонкая кремневая пластинка с наноразмерным отверстием, 2 – ионный раствор, 3 – металлические электроды, 4 – свинцовые пластины, 5 – источник тока, 6 – направленный пучок от импульсного источника, 7 – детектор.



**Рис. 12.** Напряженность электрического поля внутри поры по отношению к приложенному полю  $E_0$  для различных геометрий поры;  $\alpha = R/H$ ; 1 –  $\alpha = 0.2$ , 2 –  $\alpha = 0.5$ , 3 –  $\alpha = 1$  (а); градиент электрического поля в поре по отношению к приложенному полю  $E_0$  для различных геометрий поры,  $\alpha = R/H$  (б).

## 8. ОСНОВНЫЕ УРАВНЕНИЯ ДВИЖЕНИЯ БИОПОЛИМЕРА В НАНОПОРЕ

Как отмечалось выше, модель эластичного стержня описывает свойства полужесткого полимера [114–116]. Параметрами модели являются степень полимеризации цепочки – число элементов-мономеров  $N$ , полная длина  $L$ , коэффициент растяжения  $\kappa$ , коэффициент изгиба  $\epsilon$  и др. Конкретная конформация полимера описывается вектором  $\mathbf{r}(s)$ , отсчитываемым по длине дуги  $s$  от первого мономера; деформация изгиба задается изменением касательного вектора  $d^2\mathbf{r}/ds^2$  вдоль цепи, растяжение – изменением радиус-вектора  $d\mathbf{r}/ds$ . Необходимость введения деформации кручения цепи обсуждается в [117–121], однако здесь она не учитывается.

Движение полимера внутри поры описывается уравнением Ланжевена:

$$m \frac{\partial^2 \mathbf{r}(s, t)}{\partial t^2} + m\zeta \frac{\partial \mathbf{r}(s, t)}{\partial t} + \epsilon \frac{\partial^4 \mathbf{r}(s, t)}{\partial s^4} - \kappa \frac{\partial^2 \mathbf{r}(s, t)}{\partial s^2} - \mathbf{F} - \mathbf{A}(s, t) = 0. \quad (3)$$

$m$  – масса мономера. Второй член в уравнении – сила внутреннего трения. Следующие слагаемые – силы упругого растяжения и изгиба для непрерывной эластичной цепи. Внешняя сила дается выражением  $\mathbf{F} = \mathbf{F}_W + \mathbf{F}_E$ ; где  $\mathbf{F}_W$  – сила взаимодействия с порой и  $\mathbf{F}_E$  – электростатическая сила. Последнее слагаемое в (3) описывает случайные (стохастические) силы, возникающие из-за взаимодействия с ионами раствора.

Для учета взаимодействия мономера (участок полимера) с порой функция  $\mathbf{r}(s)$  может быть разложена по сферическим функциям  $Y_m(u)$  в направлении  $\frac{d\mathbf{r}}{ds} = \mathbf{u}(s, t)$  [122, 123]. Потенциал взаимодействия является отталкивающим на коротких расстояниях, в то время как для больших расстояний оказывается потенциалом притяжения. Полагают, что основной эффект, вносимый стенками поры, – это ориентация мономеров вдоль оси  $z$  поры [124, 125]. Ввиду симметрии задачи (как геометрии поры, так и цепи полимера) сила в направлении оси  $z$  является квадратичной функцией  $u_z^2$ :

$$F_{W,z} = \frac{w}{2} \left[ 3 \left( \frac{\partial z}{\partial s} \right)^2 - 1 \right]. \quad (4)$$

Среднее значение силы положительно,  $W \geq 0$ , параметр  $W$  является характеристикой конкретной поры.

Пусть мембрана с порой помещена в центр транслокационной камеры. Вдоль поры прикладывается постоянное напряжение, создающее внешнее поле  $E_0 = -\sigma/\epsilon_0$ , индуцирующее на поверхности диэлектрической мембраны заряды с поверхностной плотностью  $\pm\sigma$ . Учитывая, что над поверхностью поры этих зарядов нет, поле на входе в пору (рис. 12а) можно представить в виде

$$E(z) = -\frac{\sigma}{\epsilon_0 \epsilon_m} \left( \frac{z}{2\sqrt{z^2 + R^2}} + \frac{H-z}{2\sqrt{(H-z)^2 + R^2}} + \frac{\epsilon_m}{\epsilon_0} \right). \quad (5)$$

Сделаем несколько замечаний относительно области применения полученных формул.

Считается, что электростатическое взаимодействие между мономерами экранируется ионами раствора противоположного знака [126, 127].

В выражение для электростатической силы вдоль оси  $z$  необходимо включить член  $p_z dE(z)/dz$  ( $p_z$  – дипольный момент мономера), т.е. силу, действующую на диполь во внешнем поле (рис. 12б).

В [126, 127] для силы используют разложение в ряд по степеням  $z$  и ограничиваются линейными членами  $F_{E,z} = F_0 + F_1 z$ . Используя параболическое приближение для  $E(z)$ , для  $F_0$  и  $F_1$  получим

$$F_0 = eE(0) - \frac{4p_z}{3H} \left[ E\left(\frac{H}{2}\right) - E(0) \right],$$

$$F_1 H = -\frac{4}{3} \left[ E\left(\frac{H}{2}\right) - E(0) \right] \left( e - \frac{2p_z}{H} \right). \quad (6)$$

В общем случае для корректного решения уравнения (3) требуется учесть нелинейные и стохастические члены. В результате отдельные моды, описывающие колебания и вращения биомолекулы в НП, сильно взаимодействуют друг с другом, образуя новый спектр возбуждений системы. Все эти эффекты будут подробно рассмотрены в следующей работе.

В этих условиях выбор “правильной” геометрии фокусировки рентгеновского излучения в области НП является одной из ключевых задач. Различные варианты рентгенооптических систем будут рассмотрены в третьей работе вместе с методами детектирования флуоресцентных и фотоэмиссионных спектров.

### ЗАКЛЮЧЕНИЕ

Проведен анализ существующих методик секвенирования ДНК, РНК и других биомолекул, показаны достоинства и недостатки существующих методов.

Наиболее широко применяемым остается метод Сангера, основанный на химическом расщеплении молекул и электрофорезе (или анализе флуоресцентных меток), хотя для исследования структурной организации отдельных субъединиц широко используются комплементарные данные РСА, ЯМР и масс-спектрометрии.

Методики секвенирования, основанные на изучении структуры биомолекул в растворах, включая транслокацию через НП, пока не настолько популярны. Однако они позволяют эффективно изучать физико-химические свойства биополимеров в нативном состоянии и могут оказаться полезными при моделировании функциональных свойств биомолекул. Неразрушающие методики, основанные на определении различий в реакции отдельных оснований на внешние физические воздействия (изменение градиента тока, оптического сигнала, спектров флуоресцентного излучения, поглощенной дозы и др.), являются предпочтительными.

Для применения секвенирования с использованием НП и импульсных рентгеновских источников для исследования структуры биомолекул требуется выполнить большой объем работ по моделированию как взаимодействия молекулы ДНК

с раствором и НП, так и рентгенооптических систем, которым будут посвящены последующие работы.

### СПИСОК ЛИТЕРАТУРЫ

1. *Soni G., Singer A., Yu Z. et al. // Rev. Sci. Instrum.* 2010. V. 81. P. 014301.
2. *Chang S., Huang S., He J. et al. // Nano Lett.* 2010. V. 10. P. 1070.
3. *Ivanov A., Instuli E., McGilvery C. et al. // Nano Lett.* 2011. V. 11. P. 279.
4. *Hau-Riege S.P. // Appl. Phys. Lett.* 2007. V. 90. P. 173128.
5. *Juha L. // Proc. SPIE.* 2005. V. 5917. P. 91.
6. *Hau-Riege S.P., Chapman H.N., Krzywinski J. et al. // Phys. Rev. Lett.* 2007. V. 98. P. 145502.
7. *Anderson N.L., Anderson N.G. // Electrophoresis.* 1998. V. 19 (11). P. 1853.
8. *Blackstock W.P., Weir M.P. // Trends Biotechnol.* 1999. V.17 (3). P. 121.
9. *James P. // Quarterly Rev. Biophys.* 1997. V. 30 (4). P. 279.
10. [www.nature.com/subject/nextgenseq](http://www.nature.com/subject/nextgenseq)
11. *Stoddart D., Maglia G., Mikhailova E. et al. // Angew. Chem.* 2010. V. 49. P. 556.
12. *Tsutsui M., Rahong S., Iizumi Y. et al. // Nature Scientific Reports 1.* V. 46. P. 10.1038.
13. *Schneider G.F., Kowalczyk S.W., Calado V.E. et al. // Nano Lett.* 2010. V. 10 (8). P. 3163.
14. *Min Jou W., Haegeman G., Ysebaert M., Fiers W. // Nature.* 1972. V. 237 (5350). P. 82.
15. *Somerville R.A. // Trends in Biochemical Sciences.* 2002. V. 27. № 12. P. 606.
16. *Kirby B.J. Micro- and Nanoscale Fluid Mechanics: Transport in Microfluidic Devices.* Cambridge University Press. 2010.
17. *Tabelling P. Introduction to Microfluidics.* Oxford University Press. 2005.
18. *Дероум Э. Современные методы ЯМР для химических исследований.* М.: Мир, 1992. 401 с.
19. *Fenn J.B., Mann M., Meng C. K. et al. // Science.* 1989. V. 246 (4926). P. 64.
20. *Tanaka K., Waki H., Ido Y. et al. // Rapid Commun. Mass Spectrom.* 1988. V. 2 (20). P.151.
21. *Squires G.L. Introduction to the Theory of Thermal Neutron Scattering (2nd ed.).* Mineola, New York: Dover Publications Inc. 1996. 380 P.
22. *Kleywegt G.J. // Acta Cryst. D.* 2000. V. 56. P. 249.
23. *Jenkin J. // J. Electron Spectrosc. Relat. Phenom.* 1981. V. 23. P. 187.
24. *Ковальчук М.В., Кон В.Г. // УФН.* 1986. V. 149. P. 69.
25. *Молодкин В.Б., Шнак А.П., Ковальчук М.В. и др. // УФН.* 2011. Т. 181 (7). С. 681.
26. *Ковальчук М.В., Желудева С.И., Носик В.Л. // Природа.* 1997. Т. 2. С. 54.
27. *Rossmann M.G., Cullis A.F., Muirhead H. et al. // Nature.* 1960. V. 185. P. 416.



28. Смирнова Е.А., Кислицын Ю.А., Сосфенов Н.И. и др. // Кристаллография. 2009. Т. 54. № 5. С. 948.
29. Lorenz U., Kabachnik N.M., Weckert E., Vartanyants I.A. // Phys. Rev. E. 2012. V. 86. P. 051911.
30. www.xfel.eu
31. Lapierre A., Crespo Lopez-Urrutia J.R., Baumann T.M. et al. // Rev. Sci. Instrum. 2007. V. 78. P. 123105.
32. Lapierre A., Brodeur M., Brunner T. et al. // Nucl. Instrum. Methods Phys. Res. 2010. V. 624. P. 54.
33. Gulden J., Yefanov O. M., Mancuso A.P. et al. // Optics express. 2012. V. 20 (4). P. 4039.
34. Wutrich K. NMR in Biological Research: Peptides and Proteins. New York; Amsterdam: Elsevier, 1976.
35. Saunders M., Wishnia A., Kirkwood J.G. // J. Am. Chem. Soc. 1957. V. 79. P. 3289.
36. Kirpekar F., Douthwaite S., Roepstorff P. // RNA. 2000. V. 6. P. 296.
37. Senko M.W., McLafferty // Annu. Rev. Biophys. Biomol. Struct. 1994. V. 23. P. 763.
38. Ravin N.V., Galachyants Y.P., Mardanov A.V. et al. // Current Genetics. 2010. V. 56 (3). P. 215.
39. <http://www.embl.org/>
40. Laskin A.A., Skryabin K.G., Korotkov E.V. // J. Proteome Res. 2007. V. 6. P. 862.
41. Laskin A.A., Kudryashov N.A., Skryabin K.G., Korotkov E.V. // Comput. Biol. Chem. 2005. V. 29(3) P.229.
42. <http://www.nigms.nih.gov/Research/FeaturedPrograms/PSI/>
43. Moore P. // Structure. 2007. V. 15. P. 1350.
44. www.wwpdb.org
45. Watson J.D., Crick F.H.C. // Nature. 1953. V. 171. P. 737.
46. Houwink R. // J. Pract. Chem. 1940. V. 157. P. 15.
47. Melnichuk M., Wood L. T. // Phys. Rev. A. 2010. V. 82. P. 013821.
48. Cantor Ch. R., Shimmel P.R. Biophysical Chemistry. San Francisco: Freeman and Company, 1984. V. 2. 580 p.
49. Forster Th. // Ann. Phys. (Leipzig). 1948. V. 2. P. 55.
50. Stryer L., Haugland R.P. // Proc. Natl. Acad. Sci. USA. 1967. V. 58. P. 719.
51. Wu P., Brand L. // Anal. Biochem. 1994. V. 218. P. 1.
52. Millar D.P. // Curr. Opin. Struct. Biol. 1996. V. 6. P. 637.
53. Harada Y., Tokunaga M., Saito K., Yanigada T. // Nature. 1995. V. 374. P. 555.
54. Sanger F. // Science. 1981. V. 214. P. 1305.
55. Fraser C.M., Gocayne J.D., White O. et al. // Science. 1995. V. 270. P. 397.
56. Chan E.Y. // Mutat. Res. 2005. V. 573. P. 13.
57. Lander E.S. // Nature London. 2001. V. 409. P. 860921.
58. Venter J.C. // Science. 2001. V. 291. P. 1304.
59. Fredlake C.P., Hert D.G., Mardis E.R., Barron A.E. // Electrophoresis. 2006. V. 27. P. 3689.
60. Kasianowicz J.J., Brandin E., Branton D., Deamer D.W. // Proc. Nat. Acad. Sci. U.S.A. 1996. V. 93. P. 13770.
61. Kasianowicz J.J., Henrickson S.E., Weetall H.H., Robertson B. // Anal. Chem. 2001. V. 73. P. 2268.
62. Lagerqvist J., Zwolak M., Ventra M.Di. // Nano Lett. 2006. V. 6. P. 779.
63. Sanger F., Nicklen S., Coulson A.R. // Proc. Natl. Acad. Sci. U.S.A. 1977. V. 74. P. 5463.
64. Якушевич Л.В. // Молекулярная биология. 1989. Т. 23. С. 652.
65. Timoshenko S.P. // History of strength of materials, Dover Publications inc.: New York, 1983. 488 P.
66. Devi-Prasad K.V., Prohofsky E.W. // Biopolymers. 1984. V. 23. P. 1795.
67. Kim Y., Prohofsky E.W. // Phys. Rev. 1987. V. 36. P. 3449.
68. Prohofsky E.W. // Comments Mol. Cell. Biophys. 1983. V. 2. P. 65.
69. Hakim M.B., Lindsay S.M., Powell J. // Biopolymers. 1984. V. 23. P. 1185.
70. Grimm H., Stiller H., Maikrzak C. F. et al. // Phys. Rev. Lett. 1987. V. 59. P. 1780.
71. Prabhu V.V., Schroll W.K., Van Zandt L.L., Prohofsky E.W. // Phys. Rev. Lett. 1988. V. 60. P. 1588.
72. Schroll W.K., Prabhu V.V., Prohofsky E.W., Van Zandt L.L. // Biopolymers. 1989. V. 28. P. 1189.
73. Letellier A., Ghomi M., Taillandier E. // J. Biomol. Struct. Dyn. 1986. V. 3. P. 671.
74. Letellier A., Ghomi M., Taillandier E. // J. Biomol. Struct. Dyn. 1987. V. 4. P. 663.
75. Chou K.C. // Biochem. J. 1984. V. 221. P. 27.
76. Chou K.C. // Biophys. Chem. 1986. V. 25. P. 105.
77. Chou K.C., Chen N.Y., Forsen S. // Chemica Scripta. 1981. V. 18. P. 126.
78. Sobell H.M., Banerjee A., Lozansky E. D. et al. // Structure and Dynamics in Nucleic Acids and Proteins. New York: Adenine Press, 1983. P. 181.
79. Bolton P.H., James T.L. // J. Phys. Chem. 1979. V. 83. P. 3359.
80. Bolton P.H., James T.L. // J. Am. Chem. Soc. 1980. V. 102. P. 25.
81. Early T.H., Kearns D.R. // Proc. Natl. Acad. Sci. USA. 1979. V. 76. P. 4165.
82. Fologea D., Uplinger J., Thomas B. et al. // Nano Lett. 2005. V. 5. P. 1734.
83. Mathé J., Aksimentiev A., Nelson D. R. et al. // Proc. Natl. Acad. Sci. U.S.A. 2005. V. 102. P. 12377.
84. Li J., Gershow M., Stein D. et al. // Nat. Mater. 2003. V. 2. P. 611.
85. Li J., Stein D., McMullan C. et al. // Nature (London). 2001. V. 412. P. 166.
86. Storm A.J., Chen J.H., Ling X.S. et al. // Nat. Mater. 2003. V. 2. P. 537540.
87. Wanunu M., Meller A. // Nano Lett. 2007. V. 7. P. 1580.
88. Zwolak M., Ventra M. Di. // Rev. Mod. Phys. 2008. V. 80. P. 141.
89. Song L., Hobaugh M. R., Shustak C. et al. // Science. 1996. V. 274. P. 1859.
90. Henrickson S.E., Misakian M., Robertson B., Kasianowicz J.J. // Phys. Rev. Lett. 2000. V. 85. P. 3057.

91. *Chen G.S., Boothroyd C.B., Humphreys.* // *Philos. Mag.* 1998. V. 78. P. 491.
92. *Deamer D.W., Branton D.* // *Acc. Chem. Res.* 2002. V. 35. P. 817.
93. *Lee J.W., Meller A.* *Perspectives in Bioanalysis* / Ed. Mitchelson K. Amsterdam: Elsevier, 2007. 310 p.
94. *Fologea D., Gershow M., Ledden B. et al.* // *Nano Lett.* 2005. V. 5. P. 1905.
95. *Akeson M., Branton D., Kasianowicz J.J. et al.* // *Biophys. J.* 1999. V. 77. P. 3227.
96. *Ashkenasy N., Sanchez-Quesada J., Bayley H., Ghadiri M. R.* // *Angew. Chem., Int. Ed.* 2005. V. 44. P. 1401.
97. *DeGuzman V.S., Lee C.C., Deamer D.W., Vercoutere W.A.* // *Nucleic Acids Res.* 2006. V. 34. P. 6425.
98. *Winters-Hilt S., Vercoutere W., DeGuzman V. S. et al.* // *Biophys. J.* 2003. V. 84. P. 967.
99. *Vercoutere W., Winters-Hilt S., Olsen H., Deamer D. et al.* // *Nat. Biotechnol.* 2001. V. 19. P. 248.
100. *Vercoutere W.A., Winters-Hilt S., DeGuzman V.S. et al.* // *Nucleic Acids Res.* 2003. V. 31. P. 1311.
101. *Bayley H., Cremer P.S.* // *Nature (London).* 2001. V. 413. P. 226.
102. *Siwy Z., Trofin L., Kohli P. et al.* // *J. Am. Chem. Soc.* 2005. V. 127. P. 5000.
103. *Gu L.-Q., Braha O., Conlan S. et al.* // *Nature (London).* 1999. V. 398. P. 686.
104. *Zwolak M., Ventra M.Di.* // *Nano Lett.* 2005. V. 5. P. 421.
105. *Zwolak M., Ventra M.Di.* // *Encyclopedia of Nanoscience and Nanotechnology* / Ed. Singh-Nalwa H. New York: American Scientific Publishers, 2004. V. 2. P. 475.
106. *Endres R.G., Cox D.L., Singh R.R.P.* // *Rev. Mod. Phys.* 2004. V. 76. P. 195.
107. *Porath D., Cuniberti G., Felice R.D.* // *Top. Curr. Chem.* 2004. V. 237. P. 183.
108. *Lagerqvist J., Zwolak M., Ventra M.Di.* // *Biophys. J.* 2007. V. 93. P. 2384.
109. *Fischbein M.D., Drndic M.* // *Nano Lett.* 2007. V. 7. P. 1329.
110. *Xu M., Endres R. G., Arakawa Y.* *Nano Science and Technology* / Ed. Klitzing K.V. Heidelberg: Springer-Verlag, 2007. 408 p.
111. *Xu M., Endres R.G., Arakawa Y.* // *Small.* 2007. V. 3. P. 1539.
112. *Gupta S., Sullivan M., Toomey J. et al.* // *J. Synchrotron Rad.* 2007. V. 14. P. 233
113. *Wang F., Zhu Q., Ivanova E.* // *J. Synchrotron. Rad.* 2008. V. 15. P. 624.
114. *Bawendi M., Freed K.* // *J. Chem. Phys.* 1985. V. 83. P. 2491.
115. *Bosch A.* // *Macromol. Chem. Theory, Simulations.* 1993. V. 3. P. 851.
116. *Benetatos P.F.E.* // *Phys. Rev. E.* 2004. V. 70. P. 051806.
117. *Harris R., Hearst J.* // *J. Chem. Phys.* 1965. V. 44. P. 2595.
118. *Saito N., Takahasi K., Yunoki Y.* // *J. Phys. Soc. Jpn.* 1967. V. 22. P. 219.
119. *Soda K.* // *J. Phys. Soc. Jpn.* 1973. V. 35. P. 866.
120. *Aragon S., Pecora R.* // *Macromolecules.* 1985. V. 18. P. 1868.
121. *Maggs A.* // *Phys. Rev. Lett.* 2000. V. 85. P. 5472.
122. *Jerome B.* // *Rep. Prog. Phys.* 1991. V. 54. P. 393.
123. *Dzubiella J., Hansen J.* // *J. Chem. Phys.* 2005. V. 122. P. 234706.
124. *Durand E.* *Electrostatique I.* Paris: Masson, 1964. 302 p.
125. *Zhang F.* // *J. Chem. Phys.* 1999. V. 111. P. 9082.
126. *Barrett T.* // *Phys. Lett. A.* 1983. V. 94. P. 59
127. *Rayleigh J.* *The Theory of Sound.* V. 1 and 2, Dover Publications: New York, 1945. 640 P.