

Аннотация

Вирусы представляют собой наноразмерные белковые оболочки (капсиды), содержащие генетический материал и обладающие специфическими термодинамическими и механическими свойствами. Эти свойства играют большую роль в важнейших этапах процесса инфицирования, таких как взросление капсиды, выпускание генома и связывание с рецептором.

Для исследования вирусов используются современные методы электронной микроскопии (ЭМ), благодаря которым, помимо структуры белковой оболочки вирусов, удалось определить, что геном в капсиде упакован слоями. Однако для более детальной картины и исследования свойств вирусов появляется необходимость в использовании вычислительных и/или теоретических моделей.

Основной цель данной работы – создание «гибридной» модели вируса (капсида с ДНК) и исследование его свойств. В качестве вируса был выбран Гепатит В, поскольку он обладает оптимальным для изучения *in silico* размером как капсиды, так и ДНК. В ходе работы была создана модель ДНК этого вируса, на основе MRG-CG, проведена упаковка ДНК в капсиду и исследована структура и поведение ДНК при упаковке и в конечной конформации. Существующая модель капсиды была соединена с ДНК, и проведены вычислительные эксперименты по наноиндентированию капсиды. Проведено сравнение результатов с наноиндентированием пустой капсиды, а также с существующими экспериментами *in vitro*.

Содержание

1 Введение	5
2 Вирус Гепатита В (HBV)	10
2.1 Вирион и капсида	10
2.2 Генетический материал	11
3 Исследование ДНК при упаковке в сферическую оболочку	12
3.1 Крупнозернистая модель ренормализационных молекулярных групп для ДНК	14
3.1.1 Термостат Нозе-Хувера	17
3.1.2 Термостат Ланжевена	19
3.2 Упаковка ДНК в капсиду вируса	21
3.2.1 Создание начальной структуры и топологии модели ДНК	21
3.2.2 Подготовка системы для упаковки ДНК	23
3.2.3 Сжатие ДНК в сферическую оболочку	25
3.3 Анализ системы, а также структуры ДНК в процессе сжатия	27
3.3.1 Давление внутри капсиды при разных условиях	27
3.3.2 Определение центральной линии ДНК	30
3.3.3 Расчет персистентной длины	32
3.3.4 Вычисление радиального распределения ДНК и ионов внутри капсиды	35
3.3.5 Определение количества слоев ДНК и расстояния между ними .	39
4 Вычислительные методы для описания механических свойств вирусов	42
4.1 Модели для описания вируса Гепатита: пустого и с генетическим материалом	44
4.1.1 Модель самоорганизующегося полимера для описания капсиды вируса	44
4.1.2 Модель капсиды и вирусной частицы Гепатита В с генетическим материалом	46
4.2 Наноиндентирование <i>in silico</i>	50
4.2.1 Наноиндентирование пустой капсиды	53
4.2.2 Наноиндентирование вируса Гепатита В	55

1 Введение

Вирусы занимают огромную часть живой среды, как по количеству, так и по общей массе. Все живые существа ежедневно сталкиваются с миллиардами вирусных частиц, в том числе и люди. При таком постоянном воздействии наш организм научился защищать себя от большинства инфекций, однако некоторые вирусные инфекции (полиомиелит, корь, ВИЧ и др.) бывают опасны для жизни и могут привести к летальному исходу.

Несмотря на способность к инфицированию (как человека, так и других живых организмов) вирусы могут быть полезны. Например, вирусы могут защитить от невирусных заболеваний: диабет типа 1 можно предотвратить путем инфицирования лимфотропными вирусами [1]. Кроме того ослабленные вирусные частицы используются в медицине для вакцинации, помогая усилить иммунитет к болезни, что впоследствии приводит к предотвращению заражения или ослаблению его отрицательных последствий. Также с развитием технологий и более глубоким пониманием свойств и поведения вирусных систем стало возможным использовать вирусные частицы для точной доставки лекарственных препаратов в инфицированные (например, раковые) клетки [2]. Прежде всего это связано с тем, что определенный вирус может размножаться только в клетках определенного типа, обеспечивая селективность доставки нуклеиновой кислоты в определенные клетки и ткани. Все эти примеры иллюстрируют необходимость подробного изучения вирусов, поскольку с пониманием их свойств становится возможным использование вирусов во благо.

Ключевым звеном любого вируса является генетический материал (одно- или двухцепочечная РНК или ДНК), содержащий в себе информацию для выполнения всех вирусных функций, включая вирусную репликацию. Нуклеиновая кислота окружена белковой оболочкой (капсидой) нанометрового размера, предназначенная для защиты и доставки генома. Длина генома определяет количество белков, которые могут быть в ней закодированы. Более того нуклеиновая кислота ограничена размерами капсиды. Поэтому большинство капсид содержат одну или несколько белковых последовательностей, расположенных с высокой степенью симметрии. Выделяют два основных класса вирусов: палочковидные (напр. Вирус табачной мозаики, TMV) и сферические (напр. Гепатит B, HBV). Капсидные белки палочковидных вирусов расположены с винтовой симметрией вокруг нуклеиновой кислоты, а капсиды большинства сферических вирусов расположены с икосаэдрической симметрией.

Поскольку вирусы имеют регулярную структуру, их можно исследовать методом электронной микроскопии (ЭМ) (Рисунок 2(а)). Этот метод основан на окрашивании вирусов (или участков инфицированных клеток) электронно-плотным материалом. Однако даже если структура хорошо сохраняется и может быть достигнута высокая степень контраста, разрешения ЭМ ($50\text{--}75 \text{ \AA}$) недостаточно для получения атомарной структуры вируса. Так что зачастую используют более современный метод криоэлектронной микроскопии (крио-ЭМ). В этом случае исследуемый образец замораживается и исследуется при низких температурах в некристаллическом, стекловидном состоянии, что позволяет сохранить нативную структуру. В сочетании с современными математическими методами анализа изображений отдельных частиц и трехмерной реконструкции крио-ЭМ может увеличить разрешение до атомного уровня (Рисунок 2(с-е)). Этот подход может быть дополнен криоэлектронной томографией, в которой регистрируются двумерные изображения, когда исследуемый образец наклоняется под разными углами к электронному пучку и затем объединяется в трехмерную карту плотности (Рисунок 2(б)). Метод крио-ЭМ часто используется в комбинации с рентгеноструктурным анализом, который позволяет определить атомарную структуру отдельных белков оболочки вируса — капсомеров (белковых субъединиц, организованных по двум типам симметрии — кубическому или спиральному). К полученным атомарным структурам затем применяются геометрические преобразования таким образом, чтобы они наиболее точно совпадали с результатами электронной микроскопии, что позволяет получить атомарную структуру целой оболочки. Разрешенные этими способами структуры вирусных частиц можно найти в базе данных PDB (Рисунок 1): как целые капсиды, так и отдельные капсомеры (rcsb.org) [3]. На данный момент это базе собраны более 152800 биологических структур, разрешенных разными способами. Из них всего 7719 вирусов, и это далеко не все существующие вирусы.

Еще одним важным процессом, связанным с вирусами, является упаковка генома — один из этапов самосборки вирусов. Существует две основные стратегии. Многие вирусы собирают свои капсиды вокруг конденсированного генома. Другие вирусы сначала формируют белковую оболочку, а затем упаковывают свой геном в капсиду. В случае последних упаковку совершает мотор, работающий за счет энергии, поставляемой гидролизом АТФ. Конечная организация генома варьируется независимо от способа упаковки. У большинства икосаэдрических вирусов нет явного порядка в упакованном геноме, тогда как у некоторых вирусов геном приобретает частичную икосаэдрическую симметрию в своей вторичной структуре. Многие вирусы, особенно ДНК-бактериофаги,

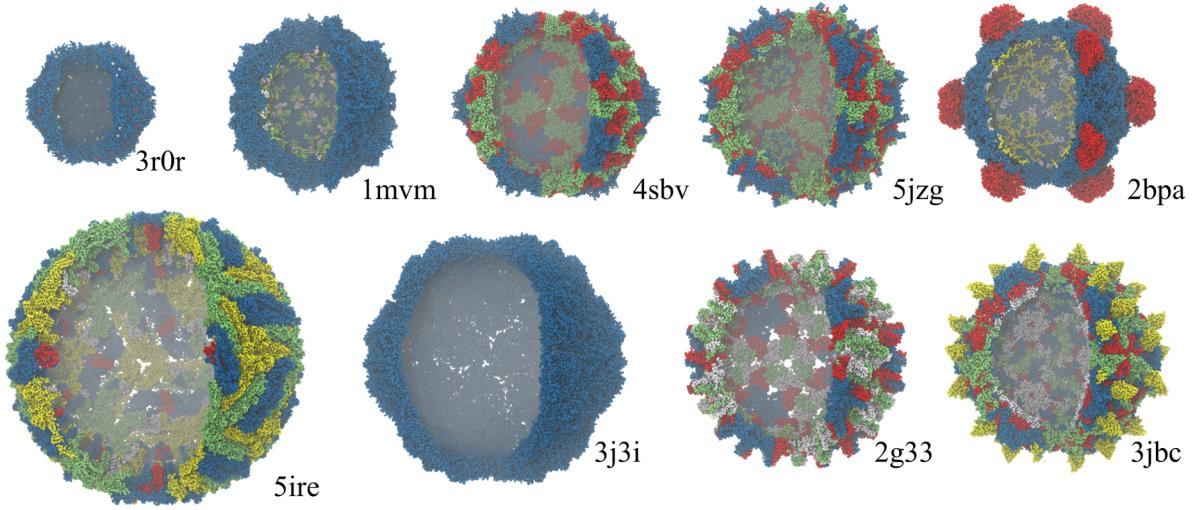


Рис. 1: Примеры разрешенных трехмерных структуры некоторых вирусов, которые можно найти в базе данных PDB.

демонстрируют систематические слои упакованной ДНК (Рисунок 2(с-е)). Другие, в основном нитевидные, вирусы упаковывают свой геном как спирали, окруженные белками. Однако современные методы ЭМ не могут разрешить внутреннее строение вирусов в полном виде. Так что открытый вопрос, каким образом нуклеиновая кислота упаковывается внутрь капсиды, и как она там располагается [4–6].

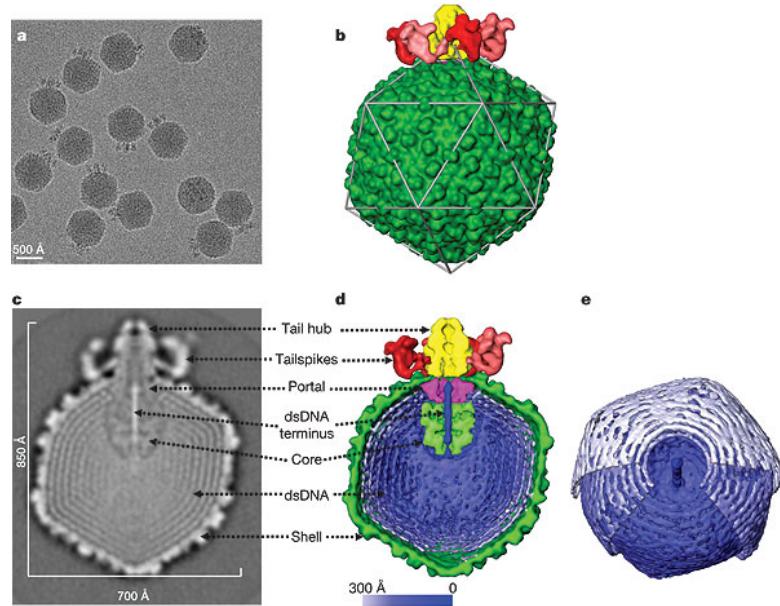


Рис. 2: Изображение частиц бактериофага $\epsilon 15$, адсорбированные на поверхности, полученное при помощи ЭМ (а). Срез и 3D структура этого же фага, полученная с использованием крио-ЭМ с разрешением 20 Å (б-д): цветом выделены структурные компоненты бактериофага. 3D изображение генетического материала фага (е). Изображение взято из [7].

Что касается исследования механических свойств вирусов, то для этого все чаще

используются такие экспериментальные технологии как атомно-силовая микроскопия (АСМ) [8–10]. Принцип работы этого метода основан на отслеживании силового взаимодействия между поверхностью образца и кантileвером, располагающимся на конце упругой панели. Сила, действующая на кантileвер со стороны поверхности исследуемого образца (например, капсиды), приводит к изгибу панели. Определяя величину изгиба, можно сделать вывод о силе взаимодействия капсиды с зондом, а перемещая зонд по вертикальной оси — о жесткости вируса. Таким методом можно как визуализировать поверхность образца вируса, так и определять его характеристики, такие как жесткость, эластичность и др. (Рисунок 3). Однако результаты подобных экспериментов сложно интерпретировать без использования вычислительного и/или теоретического моделирования, способного описывать полный спектр наблюдаемых механических явлений. В связи с чем, были созданы модели, позволяющие повторить работу АСМ, но уже *in silico* [11–15].

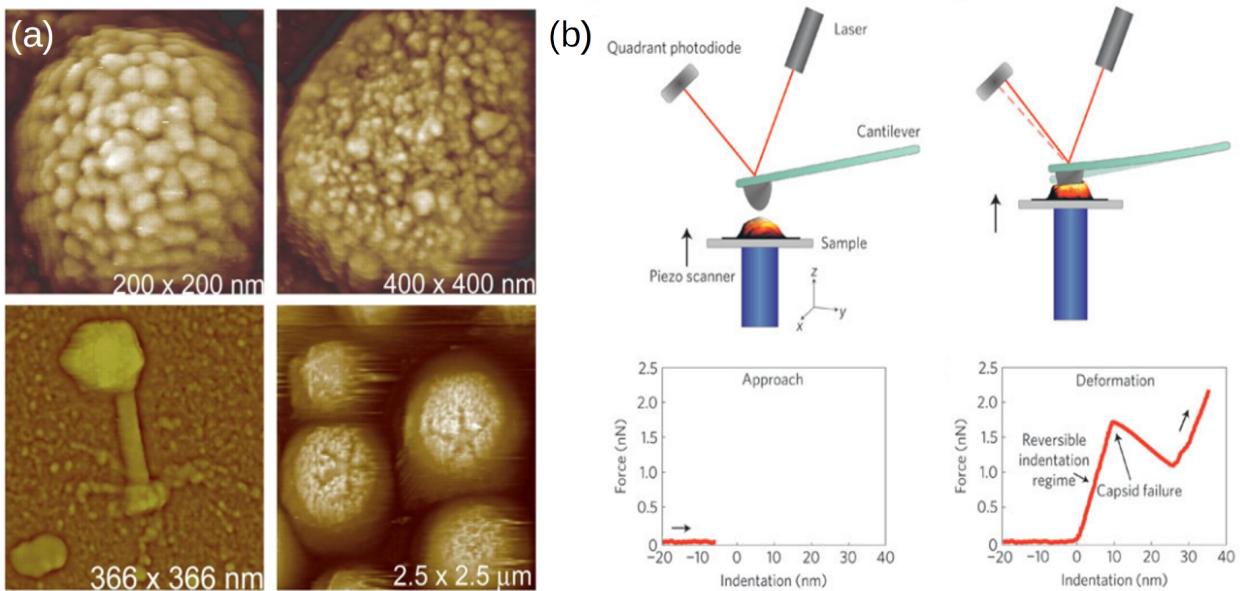


Рис. 3: (а) АСМ изображения, заимствованные из статьи [16]: вирион вируса лейкоза; поверхность вириона вируса коровьей оспы; бактериофаг, заражающий цианобактерии из морской среды; мимивирус, крупнейший вирус, известный в природе. (б) Схема АСМ наноиндентирования, взятая из статьи [8]: 1 — наконечник АСМ еще не коснулся поверхности вируса, и, следовательно, приложенная сила равна нулю; 2 — наконечник АСМ индентирует вирус и заметен изгиб кантileвера.

Существующие вычислительные эксперименты были проведены на пустых оболочках вирусов, что не дает полной информации о свойствах вирусов в реальной жизни. В связи с чем в этой работе была создана модель вируса, состоящая не только из оболочки, а также с генетическим материалом внутри. В работе использовалось круп-

нозернистое моделирование по нескольким причинам. Во-первых, огромный размер исследуемых биомолекул не дает возможности использовать стандартные методы полноатомного моделирования, а крупнозернистые модели позволяют обойти это ограничение используя меньшее число частиц на один аминокислотный остаток (для капсиды) и нуклеотид (для ДНК) и используя неявно заданный растворитель. Во-вторых, крупнозернистый подход позволяет ускорить обход фазового пространство, что ещё больше увеличивает доступные фактически временные интервалы моделирования. Успешная и точная крупнозернистая модель обычно опирается на достаточно сложный Гамильтониан [17–22]. Поэтому ее реализация требует дополнительных потенциалов взаимодействия частиц.

Для создания таких моделей в нашей лаборатории была разработана программная платформа ASAP-MD. С ее помощью были созданы модели ДНК, капсиды, которые впоследствии были объединены в единую "гибридную" модель вируса. На основе этих моделей был изучен процесс упаковки ДНК в капсиду вируса, проведен анализ ее структуры в упакованном виде, также были исследованы механические свойства вирусов — капсид без и с генетическим материалом внутри — которые были получены приnanoиндентировании *in silico*. Более того благодаря существующим исследованиям вирусов *in vitro* оказалось возможным сравнить наши модели с реальными вирусными частицами.

Результаты этой работы были представлены на постерных сессиях XXIV-ой и XXV-ой международных конференций "Математика. Компьютер. Образование." (www.mce.su), на одной из которых (XXIV) был оценен как лучший.

2 Вирус Гепатита В (HBV)

В работе исследуется вирусная частица Гепатита В. Она имеет небольшой размер капсиды (внешний радиус 42 нм), а также достаточно короткую ДНК (3182 базовые пары). Такие размеры являются оптимальными для проведения вычислительных экспериментов без особых вычислительных и временных затрат. Также существует хорошо разрешенная полноатомная структура капсиды HBV, что позволяет создать модель, хорошо описывающую свойства оболочки этого вируса. Более того, для Гепатита В проведено много экспериментов, в том числе и наноиндентирование, что позволяет нам сверять результаты, полученные вычислительными методами, с реальными.

Вирус гепатита В (HBV) является патогеном млекопитающих, птиц, а также человека. Точное количество зараженных HBV людей неизвестно, по примерным оценкам это около 400 миллионов человек. Большинство зараженных живут в Азии, многие в Африке. Существует высокий процент людей, инфицированных на крайнем севере Северной Америки и Гренландии, но из-за низкой численности населения этих регионов число инфицированных относительно невелико. Вирус присутствует в крови и сперме инфицированных людей, таким образом передача инфекции возможна либо половым путем, либо через повреждения в тканях/коже, а также от матери к ребенку. Ежегодно регистрируется более 50 миллионов новых случаев заражения гепатитом В, в основном среди детей, которые заразились от своих матерей. Считается, что более 8 миллионов инфекций в год являются результатом повторного использования шприцев и игл для инъекций, главным образом в развивающихся странах. Многие инфекции HBV вызывают легкое недомогание или протекают бессимптомно, особенно у детей. Однако именно у них инфекция HBV, скорее всего, вызовет хроническую болезнь (около 90–95% инфицированных новорожденных живут с хроническим HBV по сравнению с 1–10% инфицированных в зрелом возрасте). Люди, имеющие хроническую болезнь, могут оставаться здоровыми в течение долгого времени, но у некоторых может развиться тяжелый гепатит, который может привести к циррозу и, в конечном итоге, к раку печени. Эти заболевания, вызванные инфекцией HBV, вызывают около миллиона смертей ежегодно.

2.1 Вирион и капсида

Вирусная частица (вирион) HBV является одной из самых маленьких: она представляет собой квази-сферу диаметром 42 нм. Состоит вирион из внешней липидной

оболочки и капсиды, состоящей из белков и содержащей вирусную ДНК и ДНК- полимеразу. Капсида имеет икосаэдрическую симметрию, помимо этого в ней есть отверстия и короткие шипы, выступающие из ее поверхности (Рисунок 4). Оболочка вириона содержит три вида белка, обозначенные как маленький (S), средний (M) и большой (L). Белки M и L являются более длинными версиями белка S, наиболее распространенного из трех. Поверхностные области белков оболочки составляют поверхностный антиген гепатита В (HBsAg). Область закрепления вируса находится рядом с N-концом белка L, но только около 50% молекул L имеют N-конец снаружи вириона; N-концы оставшихся L-молекул на внутренней стороне связаны с капсидой.

2.2 Генетический материал

HBV имеет маленький геном — всего 3182 базовые пары (bp) — который используется с большой экономией для кодирования вирусных белков и для контроля экспрессии генов вируса. Помимо этого репликация этого вируса включает обратную транскрипцию, поэтому они сильно отличаются от других ДНК-вирусов, реплицирующих свою ДНК непосредственно в ДНК. Геном HBV состоит из двух нитей ДНК, одна из которых является неполной; следовательно, ДНК частично одноцепочечная и частично двухцепочечная. Короткая часть ее даже является трехцепочечной в результате комплементарной последовательности в точке 5', и это приводит к тому, что ДНК имеет круговую конформацию.

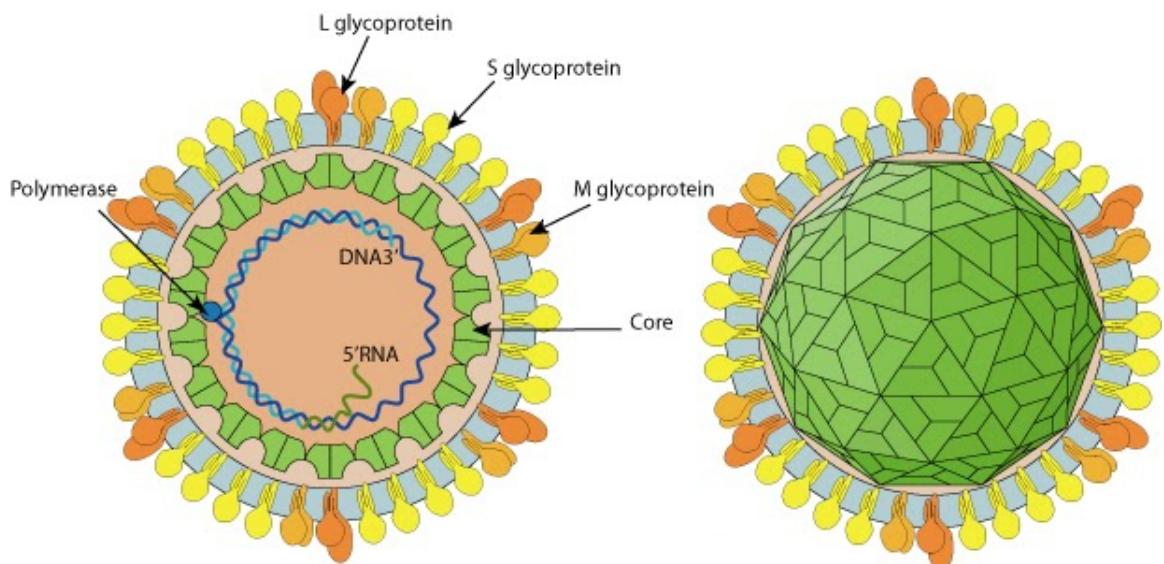


Рис. 4: Оболочка вируса Гепатита В имеет диаметр 42 нм и обладает икосаэдрической симметрией. Изображение взято из [23].

3 Исследование ДНК при упаковке в сферическую оболочку

Нуклеиновая кислота является носителем наследственной информации, которая участвует в синтезе и регуляризации огромного количества молекул в клетке. Эти процессы основаны на уникальных механических, химических и термодинамических свойствах генома, которые позволяют эффективно считывать с него информацию, в то время как цепь может быть сильно конденсирована: человеческий геном (примерно метр длиной) помещается в ядро клетки длиной несколько микрометров. Такое сжатие, достигаемое в физиологических условиях, удивительно, поскольку ДНК является сильно заряженным и жестким биополимером. Однако несмотря на жесткость генома *in vivo* часто встречаются случаи, когда цепочки нуклеиновой кислоты укладываются в вирусные капсиды <100 нм, конденсируются на хроматине диаметром от 10 до 100 нм, что предполагает различные (сильно изогнутые и скрученные) геометрические конфигурации. Понимание механического и термодинамического поведения генома представляет огромный интерес не только в биологии, но и в материаловедении. Поскольку благодаря программируемой природе последовательностей молекулы нуклеиновой кислоты стали привлекательными материалами для конструирования различных нанообъектов.

В последнее десятилетие наше понимание физики генома значительно улучшилось благодаря различным методам, позволяющим механически манипулировать отдельными молекулами, а также методам классического молекулярно-динамического моделирования [24]. Например, для более глубокого понимания биофизики нуклеиновых кислот *in vitro* активно используются полноатомные модели [25–27]. Однако такой подход, хоть и является химически более точным, не всегда реалистичен, если говорить, например, о масштабах в процессе упаковки генома в капсиду вируса. Существуют крупнозернистые модели генома, которые позволяют объяснить поведение, наблюдаемого в экспериментах на микроскопическом уровне, уменьшая при этом вычислительные затраты [27, 28]. Помимо этого такие модели могут использоваться не только для объяснения и дополнения экспериментов, но и для исследования и визуализации пространственных и временных масштабов, недоступных для экспериментальных методов. В частности, одна из текущих проблем для моделей крупнозернистых моделей генома состоит в том, чтобы достичь достаточной надежности и позволить точно описывать поведение целых нуклеосом и хроматина, структуры которых еще не определены экспериментальным путем.

Задача создания точных и переносимых крупнозернистых моделей генома сталкивается со многими проблемами, такими как необходимость учитывать сильные электростатические взаимодействия, связь цепей генома с окружающими ионами, влияние последовательности на локальную кривизну молекул и структурную жесткость. Чтобы решить это, было предложено много моделей, с различными уровнями структурного разрешения и с разными предположениями [18, 27, 28]. Создать универсальную и достаточно общую крупнозернистую модель, которая хорошо описывает и структурные, и механические, и термодинамические свойства нуклеиновой кислоты почти невозможно. Однако в последние годы было разработано несколько крупнозернистых моделей, которые хорошо описывают различными структурные и термодинамические свойства одноцепочечных и двухцепочечных геномов в различных условиях.

3.1 Крупнозернистая модель ренормализационных молекулярных групп для ДНК

Поскольку геном вируса Гепатита В большей частью состоит из двухцепочечной ДНК, было решено в данной работе для описания генома использовать крупнозернистую модель ренормализационных молекулярных групп (англ. molecular renormalization group coarse-graining, MRG-CG) [29, 30].

MRG-CG основан на идее определения парных и многочастичных потенциалов взаимодействия между частицами путем сопоставления статистической суммы микроскопических и крупнозернистых моделей. В нынешней форме метод обеспечивает систематический способ получения гамильтониана для ДНК и, в принципе, для произвольного полимера в присутствии подвижных ионов. В этом подходе каждый нуклеотид представляется в виде взаимодействующего центра, и в сравнении с полноатомными моделями, этот метод продемонстрировали высокий уровень точности как для крупномасштабных, так и более мелких механических и электростатических свойств ДНК. Примечательным аспектом MRG-CG является его связь с равновесной статистической физикой и критическими явлениями, а именно с теорией ренормализационных групп (RG), которая предоставляет возможность для систематических улучшений силового поля. В принципе, эти улучшения могут быть достигнуты с желаемой степенью точности в зависимости от характера проблемы и других методов крупнозернистого моделирования.

Силовое поле, используемое в этой модели, состоит из взаимодействий внутри цепи, межионных и между ДНК и ионами, которые точно описывают связь ДНК с ионной средой:

$$U_{MRG-CG} = U_{bond} + U_{ang} + U_{fan} + U_{el}, \quad (1)$$

здесь первые два слагаемых учитывают флуктуации угла сцепления и угла изгиба (взаимодействия внутри цепи), а последние два слагаемых относятся к взаимодействиям между цепями (представленные пунктирными линиями на Рисунке 5) и электростатическим взаимодействиям соответственно.

Функциональная форма индивидуальных вкладов в потенциалы $U_{bond,fan}$ и U_{ang} основана на полиномах, дабы учесть асимметричную форму структурных колебаний ДНК:

$$U_{bond,fan} = \sum_{\alpha=2}^4 K_{\alpha} (l - l_0)^{\alpha}, \quad U_{ang} = \sum_{\alpha=2}^4 K_{\alpha} (\theta - \theta_0)^{\alpha}, \quad (2)$$

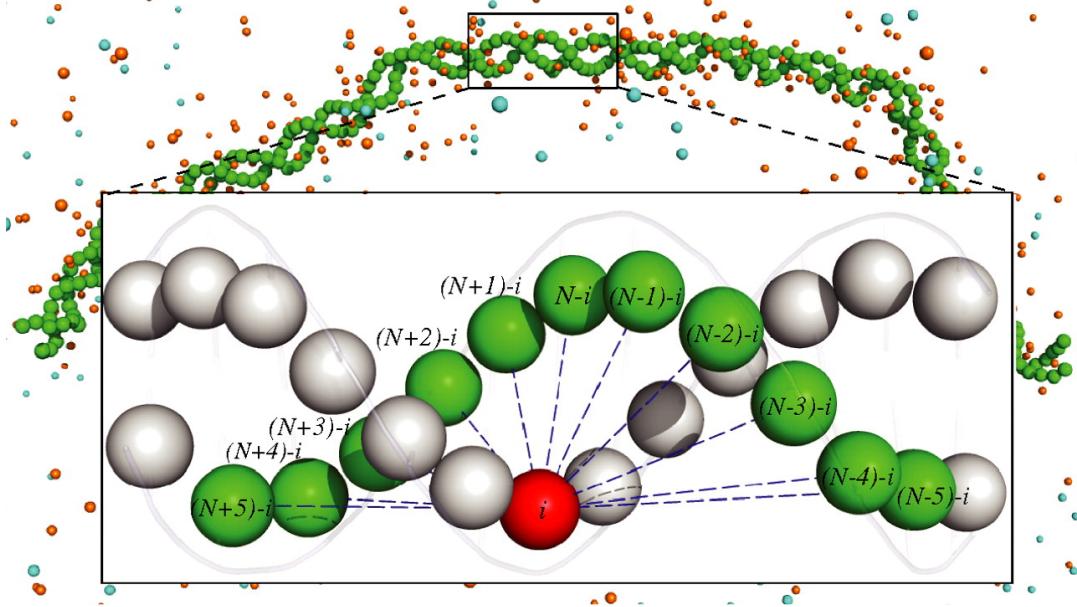


Рис. 5: Крупнозернистая модель для двухцепочечной ДНК в присутствии явных ионов. Каждая пара ДНК-ДНК представлена двумя частицами, каждая из которых помещена в геометрический центр соответствующего нуклеотида. Синие пунктирные линии указывают на межцепочечное взаимодействие. Обозначения частиц показывают, что частица с индексом i одной цепи взаимодействует с частицей $[(N\pm 0.5)-i]$ из другой цепи. В результате имеется 11 межцепочечных взаимодействий для каждой частицы. Изображение заимствовано из статьи [30].

Равновесная длина связи l_0 и равновесный угол между связями θ_0 , а также набор коэффициентов $K_\alpha^{(0)}$, были получены из полноатомного моделирования при помощи планомерной параметризации, которая использовала линейность гамильтониана по параметрам, (Таблица 1).

Однако это только первый шаг, за которым следует оптимизация всех параметров взаимодействия MRG-CG с использованием итерационной схемы. Ниже приведены потенциалы меж-ионного взаимодействия ($U_{ion-ion}$) и потенциалы между частицами ДНК и ионами ($U_{DNA-ion}$), вместе составляющие U_{el} :

$$U_{ion-ion} = \sum_{i>j} \left[\frac{A}{r_{ji}^{12}} + \sum_{k=1}^5 B^{(k)} e^{-C^{(k)}(r_{ij}-R^{(k)})^2} + \frac{q_i q_j}{4\pi\epsilon_0\epsilon r_{ij}} \right], \quad (3)$$

$$U_{DNA-ion} = \sum_{i>j} \left[\frac{A}{r_{ji}^6} + \sum_{k=1}^3 B^{(k)} e^{-C^{(k)}(r_{ij}-R^{(k)})^2} + \frac{q_i q_j}{4\pi\epsilon_0\epsilon r_{ij}} \right],$$

где A , $B^{(k)}$, $C^{(k)}$ и $R^{(k)}$ – параметры модели, которые были взяты из статьи [30] (Таблица 2).

В крупнозернистой модели (Рисунок 5) каждый нуклеотид отображается в один шарик, что приводит к более чем 30-кратному уменьшению степеней свободы цепи

Таблица 1: Параметры для гамильтониана MRG-CG, определяемого полиномами в Уравнениях (2). Параметры K_1 и K_2 и K_3 для всех типов взаимодействий, кроме потенциала угла изгиба, приведены в единицах [ккал/моль · Å⁻²], [ккал/моль · Å⁻³] и [ккал/моль · Å⁻⁴] соответственно. Аналогичные параметры для потенциала угла изгиба измеряются в единицах [ккал/моль · рад⁻²], [ккал/моль · рад⁻³] и [ккал/моль · рад⁻⁴], соответственно. l_0 - это равновесная длина связи (в Å), а θ_0 - равновесный угол изгиба (в градусах). Обозначения «+1..5» и «-1..5» соответствуют частицам [(N±0..5)-i] как в обозначениях на Рисунке 5.

		K_1	K_2	K_3	l_0, θ_0
Внутри цепи	Связь	2.625	-0.226	0.0148	4.96
	Угол	9.22	4.16	1.078	156
Между цепями	+5	0.3626	-0.077	0.005	16.9
	+4	0.1386	-0.0568	0.005	12.5
	+3	0.1378	-0.0527	0.005	10.2
	+2	0.0955	-0.0459	0.00502	9.2
	+1	0.115	-0.041	0.0058	9.9
	0	2.92	0.41	0.072	11.3
	-1	0.040	-0.01	0.0008	12.3
	-2	0.123	-0.04	0.0037	13.45
	-3	0.085	-0.0444	0.005	14.7
	-4	1.324e-06	-0.0122	0.00185	16.35
	-5	0.0467	0.0021	0.000146	17.1

ДНК, что позволяет проводить эксперименты в большем масштабе. Еще одной примечательной особенностью является учет явного присутствия ионов в системе. Это устраняет многие артефакты, присутствующие в большинстве крупнозернистых моделей ДНК, где корреляции среди ионов вблизи сильно заряженных поверхностей ДНК в значительной степени игнорируются. Также следует заметить, что эта модель обладает большим преимуществом, поскольку она правильно описывает персистентную длину (жесткость) ДНК при различной концентрации ионов (Раздел 3.3.3) [30].

Использование молекулярной динамики приводит к микроканоническому ансамблю (NVE) в системе (постоянны количество частиц, объем и полная энергии). Но многие величины, которые нам интересны, могут получены только из канонического ансамбля (NVT), где температура, количество частиц и объем фиксируются в заданных значениях. Температура в отличие от количества частиц и объема — интенсивный параметр. Аналогом температуры является кинетическая энергия, которая связана с ней уравнением:

$$T = \frac{2}{3} \frac{E}{Nk_B} = \frac{1}{3Nk_B} \sum_{i=1}^N \frac{p_i^2}{m} \quad (4)$$

Таблица 2: Оптимизированные параметры, определяющие взаимодействия между ионами и ДНК, Уравнение (3). Параметры A и $B_1 \dots B_5$ приведены в единицах [ккал/моль · Å⁶] и [ккал/моль], соответственно. Гауссовские отклонения $C_1 \dots C_5$ имеют размерность [Å⁻²], в то время как позиции гауссовых функций $R_1 \dots R_5$ измеряются в Å.

	Na ⁺ -ДНК	Cl ⁻ -ДНК	Na ⁺ -Na ⁺	Na ⁺ -Cl ⁻	Cl ⁻ -Cl ⁻
A	$1.82 \cdot 10^4$	$1.48 \cdot 10^3$	$8.4 \cdot 10^5$	$1.14 \cdot 10^5$	$2.7 \cdot 10^7$
B_1	0.11	-0.06	-0.013	-1.06	-0.59
B_2	0.173	-0.19	0.61	2.61	0.33
B_3	0.035	-0.06	-0.2	-0.29	-0.053
B_4	—	—	0.086	0.276	0.078
B_5	—	—	-0.023	-0.072	-0.022
C_1	7.8	9.0	5	8.4	5.8
C_2	1.2	0.5	2.46	3.5	3.12
C_3	0.31	0.11	4.38	3.52	3.5
C_4	—	—	3.73	5.66	4
C_5	—	—	7.4	5	9.9
R_1	6.5	5.1	3.8	2.75	4.9
R_2	8.6	7.5	4.75	3.52	6.1
R_3	11.5	11.0	6.2	5.1	7.5
R_4	—	—	7.2	6	8.5
R_5	—	—	8.7	7.3	9.7

Для перехода к каноническому ансамблю в системе необходимо ввести термостат, чтобы поддерживать среднюю температуру в желаемых пределах. Существует множество методов, позволяющих реалистично добавлять и удалять энергию с границы системы, приближая ее к каноническому ансамблю. Популярными методами контроля температуры являются термостаты Андерсона, Беренсена, Нозе-Хувера и Ланжевена (стохастический).

Для этой работы был выбран интегральный термостат для поддержания необходимой средней температуры, более того было проведено сравнение двух термостатов для упаковки генома в вирусную капсиду, а именно Нозе-Хувера [31, 32] и Ланжевена.

3.1.1 Термостат Нозе-Хувера

Интегральный термостат Нозе-Хувера вводит дополнительные степени свободы в гамильтониан системы, для которого можно вывести уравнение движения. Эти уравнения для дополнительных степеней свободы интегрируются вместе с «обычными» уравнениями для пространственных координат и импульсов (то есть с уравнениями Ньютона).

на).

Идея метода, предложенного Нозе [31], заключалась в том, чтобы уменьшить влияние внешней системы, выступающей в качестве теплового резервуара, до дополнительной степени свободы. Этот тепловой резервуар контролирует температуру данной системы, таким образом температура колеблется вокруг заданного значения. Фактически, тепловое взаимодействие между тепловым резервуаром и системой приводит к обмену кинетической энергией между ними. Нозе ввел два набора переменных: $\{p_i, q_i\}$ — реальные момент и координаты i -го атома и $\{\pi_i \rho_i\}$ — виртуальные. Виртуальные переменные являются производными от так называемого преобразования Сундмана [33]:

$$s = \frac{d\tau}{dt}, \quad (5)$$

где τ — виртуальное время, t — реальное время, а s — коэффициент масштабирования, который так же рассматривается как динамическая переменная. Преобразование виртуальных переменных $\{\pi_i \rho_i\}$ в реальные $\{p_i, q_i\}$ выполняется в соответствии с:

$$\begin{aligned} p_i &= \pi_i, \\ q_i &= \rho_i. \end{aligned} \quad (6)$$

Введение эффективной массы M_s также связывает импульс с дополнительной степенью свободы π_s . Полученный гамильтониан, выраженный через виртуальные координаты, можно записать в виде:

$$H^* = \sum_{i=1}^N \frac{\pi_i^2}{2ms} + U(\rho_1, \rho_2, \dots, \rho_N) + \frac{\pi_s^2}{2M_s} + gk_B T \ln s, \quad (7)$$

где $g = 3N + 1$ — число степеней свободы расширенной системы ($+1$ — новая степень свободы).

Из гамильтониана можно получить новые уравнения движения:

$$\begin{aligned} \frac{d\rho_i}{d\tau} &= \frac{\partial H^*}{\partial \pi_i} = \frac{\pi_i}{ms^2}, \\ \frac{d\pi_i}{d\tau} &= -\frac{\partial H^*}{\partial \rho_i} = -\frac{\partial U}{\partial \rho_i}, \\ \frac{ds}{d\tau} &= \frac{\partial H^*}{\partial \pi_s} = \frac{\pi_s}{M_s}, \\ \frac{d\pi_s}{d\tau} &= -\frac{\partial H^*}{\partial s} = \frac{gkT}{s} + \sum_{i=1}^N \frac{\pi_i^2}{2ms}, \end{aligned} \quad (8)$$

Если преобразовать эти уравнения в реальные переменные $\{p_i, q_i\}$ и ввести новую переменную $\zeta = s \frac{ds}{dt} = s \frac{ds}{d\tau} \frac{d\tau}{dt} = s \frac{\partial H^*}{\partial \pi_s} \frac{d\tau}{dt} = s^2 \frac{\pi_s}{M_s}$, то получим:

$$\begin{aligned}\frac{d\mathbf{q}_i}{dt} &= \frac{\mathbf{p}_i}{m_i}, \\ \frac{d\mathbf{p}_i}{dt} &= -\frac{\partial U}{\partial \mathbf{q}_i} - \zeta \mathbf{p}_i, \\ \frac{\partial \ln s}{dt} &= \zeta, \\ \frac{d\zeta}{dt} &= -\frac{1}{M_s} \left(\sum_{i=1}^N \frac{p_i^2}{2m_i} - gk_B T \right), p_i \equiv |\mathbf{p}_i|\end{aligned}\tag{9}$$

Эти уравнения описывают термостат Нозе-Хувера. Параметр M_s является параметром тепловой инерции, который определяет скорость передачи тепла. Значение этого параметра необходимо устанавливать осторожно, поскольку, если оно выбрано слишком маленьким, фазовое пространство системы не будет каноническим, и если оно выбрано слишком большим, управление температурой будет неэффективным. С помощью термостата Нозе-Хувера мы можем добиться среднего значения температуры температуры, равное заданному значению.

3.1.2 Термостат Ланжевена

Когда мы рассматриваем движение крупных частиц в жидкости, нужно учитывать уравнение Ланжевена:

$$\ddot{x} = \nabla \phi - \gamma \dot{x} + \sigma \xi\tag{10}$$

или

$$\begin{aligned}\frac{d\mathbf{q}_i}{dt} &= \frac{\mathbf{p}_i}{m}, \\ \frac{d\mathbf{p}_i}{dt} &= -\frac{\partial \phi(\mathbf{q})}{\partial \mathbf{q}_i} - \gamma \mathbf{p}_i + \sigma \xi_i\end{aligned}\tag{11}$$

Здесь маленькие частицы добавляют силу трения в момент $(-\gamma \mathbf{p}_i)$, поскольку крупные частицы выталкивают более мелкие с пути. Помимо этого маленькие (тепловые) частицы движутся с кинетической энергией и случайно ударяются о крупные. Для того, чтобы восстановить распределение канонического ансамбля, σ, γ связаны друг с

другом флюктуационно-диссипативным соотношением:

$$\sigma^2 = 2\gamma m k_B T \quad (12)$$

Уравнение Ланжевена можно использовать в молекулярной динамике, предполагая, что моделируемые атомы окружены гораздо более меньшими вымыщленными частицами. Во многих системах особый интерес представляет поведение растворенного вещества, а поведение растворителя — нет (например, белки, ДНК, наночастицы в растворе). В этих случаях растворитель влияет на динамику растворенного вещества (наночастиц) посредством случайных столкновений и путем воздействия силы трения на движение наночастиц в растворителе. И уравнение движения Ланжевена позволяет объединить эти два эффекта.

В каждый момент времени термостат Ланжевена изменяет уравнение движения так, чтобы изменение импульса было:

$$\Delta \mathbf{p}_i = \left(\frac{\partial \phi(\mathbf{q})}{\partial \mathbf{q}_i} - \gamma \mathbf{p}_i + \delta p \right) \Delta t, \quad (13)$$

где γ — коэффициент трения, а δp — случайная сила, имеющая Гауссово распределение с вероятностью:

$$\rho(\delta p) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{|\delta p|^2}{2\sigma^2}\right) \quad (14)$$

и стандартным отклонением, описанными в Уравнении 12

Случайная флюктуирующая сила моделирует тепловые удары от мелких частиц. Использование силы трения и случайной силы вместе позволяют добиться правильного канонического ансамбля.

Основное преимущество термостата Ланжевена состоит в том, что, благодаря диссипативному члену, мы можем выбрать относительно большой шаг по времени Δt (в 2-3 раза больший, чем в молекулярной динамике), поскольку трение стабилизирует уравнения движения. Кроме того, поскольку термостат Ланжевена заменяет самые быстрые частотные движения в реальной системе стохастическими терминами, шаг по времени теперь выбирается такой, чтобы разрешить более медленные степени свободы, и, таким образом, Δt в несколько сотен раз больше, чем в исходной системе.

3.2 Упаковка ДНК в капсиду вируса

Как уже говорилось в Разделе 1 существует два основных способа упаковки генетического материала при самосборке вируса: (1) сборка белковой капсиды вокруг генома; (2) формирование белковой оболочки, и последующая упаковка ДНК при помощи мотора. Однако для наших целей был выбран упрощенный способ упаковки ДНК — при помощи сферы, проницаемой для ионов, радиус которой уменьшается на каждом шаге по времени. Весь вычислительный эксперимент состоит из четырех этапов: создание модели, минимизация, уравновешивание системы и сжатия генома.

3.2.1 Создание начальной структуры и топологии модели ДНК

Для создания структуры ДНК за основу был взят исходный код для создания MRG-CG модели, написанный на языке программирования Python. Данный скрипт позволяет описать координаты и взаимодействия всех частиц системы, в соответствие с Уравнениями (1-3), и сохранить их для дальнейших вычислений в форматах: xyz (описание координат всех частиц системы: ДНК и ионы) и top (набор таких параметров, как углы, длины связи, заряды, характерных для структуры). Каждая частица ДНК в создаваемой системе имеет единичный отрицательный заряд, так что для нулевого суммарного заряда систему необходимо нейтрализовать ионами Na^+ , а также дополнить молекулами NaCl для воспроизведения необходимой концентрации. Положение ионов задается случайно, но с учетом координат частиц ДНК, дабы избежать их наложения. Помимо явных ионов в системе используется неявный растворитель, поскольку мы исключаем множество атомов и включаем их неявно, используя стохастические термины, тем самым уменьшая количество вычислений за один шаг.

Исходная модель создавалась для двухцепочечной прямой ДНК, состоящей из 200 bp. Однако для этой работы длина ДНК должна быть в несколько раз больше (в работе были исследованы ДНК длинами от 100 до 10000 bp), и если создавать прямую нить, то необходимо увеличивать размеры расчетной области ($\times 3$). В связи с этим начальная конформация ДНК была создана в виде спиралевидной сферы (Рисунок 8(а)), уравнение которой представлено следующей параметрической формулой:

$$\begin{cases} x = r\sqrt{1 - t^2} \cos(a\pi t), \\ y = r\sqrt{1 - t^2} \sin(a\pi t), \\ z = rt, \end{cases} \quad (15)$$

Здесь радиус сферы (r) в начальный момент времени был определен из реальной длины генома (произведение расстояния между нуклеотидами и длины генома).

Для различных исследований модели были проведены различные вычислительные эксперименты для разных длин ДНК и при разном количестве ионов в системе. В Таблице 3 приведены все исследуемые системы.

Таблица 3: Конфигурации ДНК и параметров расчетной системы, исследованные в этой работе (1 kbp = 1000 базовых пар).

Длина ДНК, kbp	Концентрация NaCl, мМ	Размер расчетной области, нм ³
4	0.1	200x200x200
	2	200x200x200
	5	250x250x250
	12	190x190x190
	50	120x120x120
0.5	0.5	
	1	
	2	
	6	50
	8	
	10	
0.1	0.1	200x200x200
	2	200x200x200
	5	250x250x250
	12	190x190x190
	50	120x120x120
0.2	0.1	200x200x200
	2	200x200x200
	5	250x250x250
	12	190x190x190
	50	120x120x120
0.5	0.1	200x200x200
	2	200x200x200
	5	250x250x250
	12	190x190x190
	50	120x120x120

3.2.2 Подготовка системы для упаковки ДНК

Начальный вариант системы создавался из соображений наименьшего объема, не учитывая растяжение цепи на «экваторе» и сжатие на «полюсах». Так что перед непосредственной упаковкой ДНК в сферическую оболочку энергию систему необходимо минимизировать, чтобы избежать нестабильности. Физически минимизация энергии соответствует мгновенному замораживанию системы: статическая структура, в которой ни один атом не чувствует суммарную силу, соответствует температуре 0 К. Потенциальная энергия, рассчитанная путем суммирования энергии всех взаимодействий (Уравнение 1), является численной величиной для текущей конформации. Минимизация энергии обычно выполняется градиентной оптимизацией: здесь мы используем метод наискорейшего спуска (steepest descent). Таким образом новые координаты частиц рассчитываются таким образом, чтобы влияние суммарной силы на каждую из частиц уменьшалось. В минимизированной структуре силы, действующие на каждую частицу небольшие, и, следовательно, такая структура будет отличной отправной точкой для дальнейшего моделирования.

Как уже было сказано, в качестве оптимизационного алгоритма был использован градиентный спуск. Вместо того, чтобы производить вычисления многочисленных вторых производных, в данном методе используются приближения: здесь вторая производная предполагается постоянной. Следовательно, уравнение для обновления координат выглядит следующим образом:

$$\mathbf{r}_{i+1} = \mathbf{r}_i - \gamma_i U'(\mathbf{r}_i), \quad (16)$$

где γ задает скорость спуска (может быть постоянной, убывающей в процессе спуска или гарантирующей наискорейший спуск); U — гамильтониан системы (Уравнение 1); \mathbf{r} — координаты в трехмерном пространстве. Первые производные все же нужно считать в каждой точке, но не вторые, за счет чего метод сходится намного быстрее, чем, например, метод Ньютона–Рафсона. Однако из-за приближений градиентный спуск не так эффективен, и поэтому для нахождения минимума обычно требуется большое количество итераций (в нашем случае минимизация осуществлялась в течение 10 нс с шагом по времени 10 фс). Направление, в котором минимизируется энергия, противоположно направлению, в котором градиент максимальен в i -ой точке. Как только достигается минимум в первом направлении, выполняется второй этап минимизации, который начинается с этой точки и двигается в самом «кругом» оставшемся направлении (Рису-

нок 6). Этот процесс продолжается до тех пор, пока не будет достигнут минимум во всех направлениях с необходимой точностью (другими словами пока не будет найден локальный или глобальный минимум на поверхности потенциальной энергии). Структура, полученная после минимизации энергии показана на Рисунке 8(b).

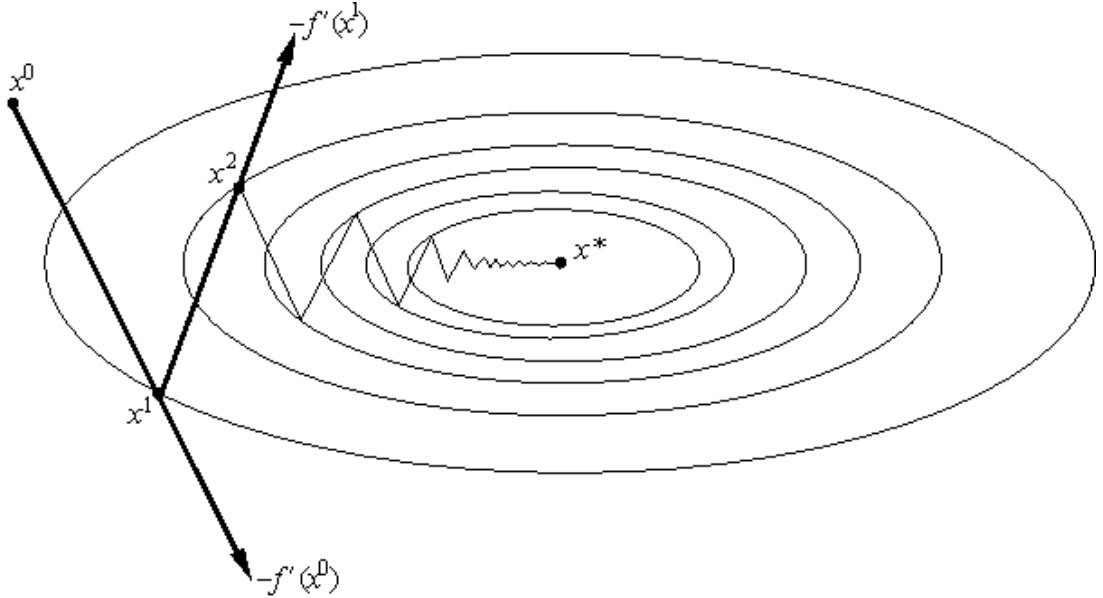


Рис. 6: Геометрическая интерпретация метода градиентного спуска для двумерного случая. $f(x)$ — минимизируемая функция.

Помимо минимизации было проведено так называемое уравновешивание системы. На этом этапе подготовки системы для интегрирования уравнений движения был использован метод с перешагиванием (более известен как Leap Frog). Этот метод имеет такое интересное название поскольку положение частиц \mathbf{r} и их скорость \mathbf{v} «перепрыгивают» через друг друга (Рисунок 7). Leap Frog использует \mathbf{r} — координаты в момент времени t , \mathbf{v} — скорости в момент времени $t - \frac{1}{2}\Delta t$ и обновляет их, используя силы $\mathbf{F}(t)$ в соответствие с Уравнениями 17. Алгоритм с перешагиванием имеет третий порядок по \mathbf{r} и обратим во времени.

$$\begin{aligned} \mathbf{v}(t + \frac{1}{2}\Delta t) &= \mathbf{v}(t - \frac{1}{2}\Delta t) + \frac{\Delta t}{m} \mathbf{F}(t), \\ \mathbf{r}(t + \Delta t) &= \mathbf{r}(t) + \Delta t \mathbf{v}(t + \frac{1}{2}\Delta t). \end{aligned} \quad (17)$$

Для данного этапа подготовки системы для поддержания средней температуры в системе были использованы интеграторы, описанные в Разделах 3.1.1 и 3.1.2. При таких интеграторах система уравновешивалась при постоянной средней температуре 300 К. Шаг по времени был 5 фс, что оказалось допустимым для избежания неустойчивости

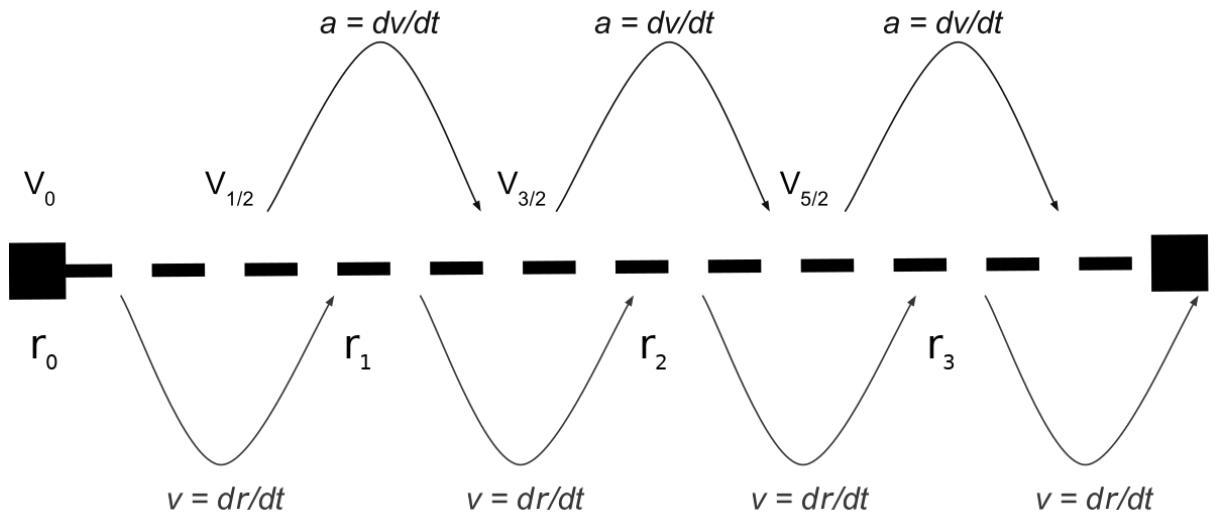


Рис. 7: Визуальное представление алгоритма с перешагиванием (Leap Frog).

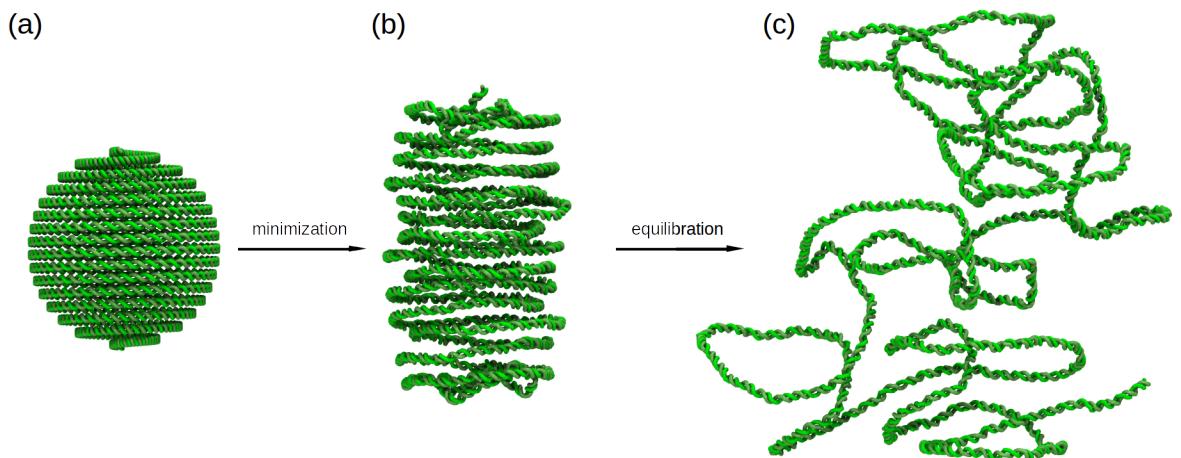


Рис. 8: Иллюстрация процессов минимизации и уравновешивания ДНК вируса Гепатита В (частично двухцепочечная, частично одноцепочечная), которая на рисунке обозначена зеленым цветом. На (а) изображена начальная конформация ДНК — сферическая спираль; (б) — структура, полученная после первого этапа — минимизации энергии; на (с) представлена уже уравновешенная система, готовая к упаковке в сферическую оболочку.

при вычислениях. Уравновешивание происходило течение 5 нс, в результате чего была получена конформация ДНК, изображенная на Рисунке 8(с).

3.2.3 Сжатие ДНК в сферическую оболочку

Для создания гибридной модели вируса необходимо сжать полученную структуру ДНК до размеров чуть меньших внутреннего радиуса капсида (для Гепатита В это 11 нм). Для сжатия используется потенциал Pushing Sphere, который представляет

собой сферу, радиус которой уменьшается с каждый шагом:

$$\Delta R = (R_0 - R_f) \frac{\Delta t}{N}, \quad (18)$$

здесь R_0 – начальный радиус сферы; а R_f – величина, до которой радиус будет уменьшаться; Δt – значение шага по времени; N – общее количество шагов. Все эти параметры задаются в конфигурационном файле.

Стоит заметить, что эта сфера проницаема для ионов, таким образом сжимается только ДНК, и давление на сферу оказывается только частицами ДНК, что позволяет ионам Na^+ свободно конденсироваться на молекуле ДНК. Давление внутри сферы рассчитывалось через силу действия атомов ДНК на сферу, которая, в свою очередь, определялась отталкивающим потенциалом Леннарда-Джонса:

$$\begin{aligned} p &= F/(4\pi R^2), \\ \mathbf{F} &= -\nabla U, \\ U &= \sum_i \varepsilon (\sigma/r_i)^2, \end{aligned} \quad (19)$$

где r_i – расстояние от частицы i до сферы, значения ε , σ также задаются в конфигурационном файле.

Для вычислений, как и в случае уравновешивания, используются интегральные термостаты (Разделы 3.1.1 и 3.1.2), поддерживающие температуру 300К.

В процессе упаковки ионы натрия конденсируются на ДНК (это хорошо видно на увеличенной части Рисунка 9(а)), нейтрализуя ее сильный отрицательный заряд, тем самым уменьшая отталкивание между цепями генома, что позволяет упаковать достаточно длинные цепи генома (4000 базовых пар имеют длину 1360 нм) в радиус порядка 10 нм.

3.3 Анализ системы, а также структуры ДНК в процессе сжатия

Одной из основных целей данной работы было отслеживание характеристик системы во время упаковки ДНК, а также анализ полученной структуры. В процессе были получены такие параметры, как давление со стороны ДНК на сферическую оболочку, разные энергии системы, а также непосредственно положение ДНК в каждый момент времени упаковки. Благодаря этим параметрам удалось определить персистентную длину, плотность и структуру упаковки нуклеиновой кислоты внутри сферы, а также внутреннее давление капсиды в зависимости от ее длины и от концентрации ионов натрия и хлора в системе. Эти данные позволяют проверить правильность используемой модели путем сравнения с экспериментальными данными, а также дополнить их.

3.3.1 Давление внутри капсиды при разных условиях

При упаковке ДНК были использованы термостаты Нозе-Хувера и Ленжеvена (подробнее о каждом написано в Разделах 3.1.1 и 3.1.2). Из полученных графиков давления (Рисунок 9(а)) видно, что, в принципе, разницы в терминах давления в этих двух интеграторах нет, однако в случае Нозе-Хувера обход фазового пространства происходит быстрее, так что в дальнейших расчетах использовался именно он. Более того, был проведен подбор шага по времени для сжатия ДНК, в результате чего для Нозе-Хувера был выбран шаг в 1 фс.

Когда термостат был выбран, в первую очередь для проверки корректности его параметров, а также чтобы убедиться в независимости сжатой структуры от начальной конформации нуклеиновой кислоты, были проведены вычислительные эксперименты по уравновешиванию и упаковке двухцепочечной ДНК после минимизации энергии системы. Расчеты были проведены для пяти разных траекторий, а именно при разных значениях случайного числа (сида), которое используется для инициализации генератора псевдослучайных чисел. Все остальные параметры были одинаковы: эксперименты проходили при одной концентрации ионов в системе, равной 50 мМ; шаг по времени был выбран 1 фс и эксперимент проводился в течение 100 нс. Сравнение полученных значений давления в зависимости от радиуса сжатия сферы представлено на Рисунке 9(б). Как видно из графика, все траектории ведут себя одинаково, вне зависимости от инициализации, что говорит о том, что можно проводить дальнейшие исследования

с разными заданными значениями сида и не бояться, что он повлияет на вычисления.

Помимо этого была исследована упаковка ДНК разной длины и при разных ионных силах в системе (Таблица 3 — первые две строки). На основе этих расчетов были получены графики давления (Рисунок 9(с-d)), из которых видно, что количество ионов в системе слабо влияет на давление внутри капсиды. А вот длина ДНК имеет очень большое влияние, что вполне объяснимо: при упаковке большего генома в один и тот же объем расстояние между цепями ДНК будет уменьшаться, и поскольку сконденсированные на ней ионы натрия уже не вмещаются в это расстояние, сила отталкивания ДНК от самой себя увеличивается, создавая большее давление на капсиду изнутри.

Также стоит заметить, что давление, оказываемое со стороны ДНК, сжатой до внутреннего радиуса капсиды Гепатита В (12.5 нм), на сферу почти не зависит от концентрации соли в системе и составляет порядка 100 атмосфер. В то время как реальное внутреннее давление вируса составляет несколько десятков атмосфер [34, 35]. Такая разница объясняется тем, что реальный геном HBV имеет отрезок одноцепочечной ДНК (Раздел 2.2). Однако, если заменить часть (а именно отрезок из 671 базовых пар) двухцепочечной ДНК на одноцепочечную, давление получается равным примерно 40 атмосфер (Рисунок 9(е)), что уже гораздо лучше соотносится с экспериментальными результатами.

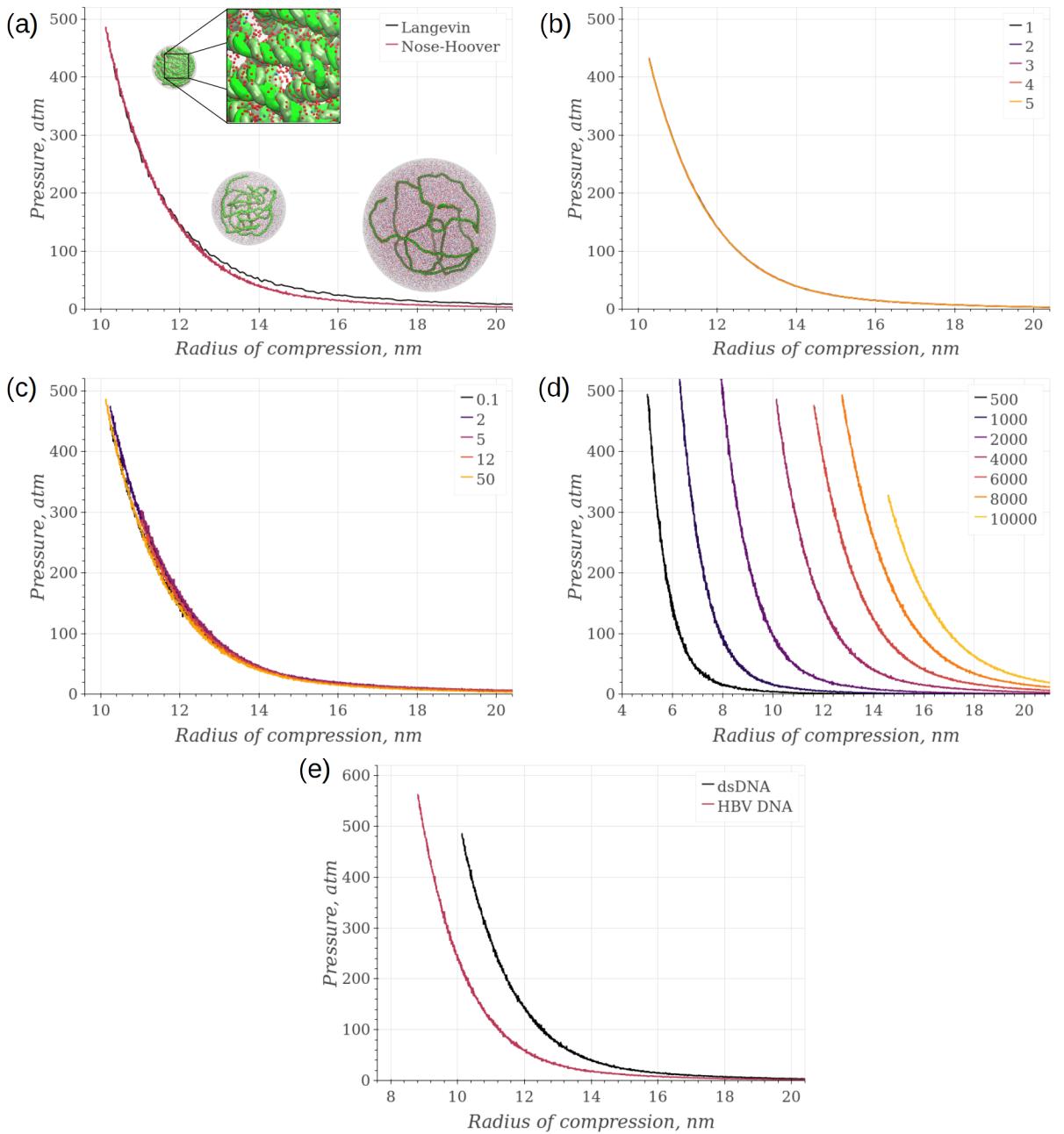


Рис. 9: Графики зависимости давления от степени сжатия ДНК. (а) Сравнение двух термостатов: Ланжевена и Нозе-Хувера (Разделы 3.1.1 и 3.1.2) с визуализацией структуры ДНК и сжимающей ее сферы. На увеличенном снимке сильно сжатой ДНК видно, как ионы натрия конденсируются внутри сферы. Поведение давления внутри капсиды не зависит от начальной конфигурации системы (б). Более того, внутреннее давление капсиды слабо зависит от концентрации соли NaCl в системе (с), однако от нее зависит до какой степени ДНК сожмется (здесь использовалась двухцепочечная ДНК длиной 4000 базовых пар). А вот от длины ДНК зависит не только радиус сферы (д), в которую ДНК можно упаковать, но и давление внутри (при этом ионная сила была одна и та же: 60 мМ). В природе внутреннее давление вируса — несколько десятков атмосфер, при упаковке двухцепочечной ДНК длиной в 3182 базовую пару в капсиду Гепатита В (внутренний радиус 12,5 нм), давление получается порядка 100 атмосфер (е). Однако реальный геном Гепатита В содержит часть одноцепочечной, и при упаковке такого генома внутреннее давление вируса аналогично реальному (≈ 40 атмосфер).

3.3.2 Определение центральной линии ДНК

Поскольку в данной работе мы имеем дело с двухцепочечной и «смешанной» ДНК, для того, чтобы рассчитать такие характеристики как перsistентная длина, расположение ДНК внутри капсиды и других, необходимо найти ее центральную линию (линию, которая проходит через центр спирали нуклеиновой кислоты). Из-за наличия в структуре ДНК так называемых бороздок просто взять центр отрезка, соединяющего базовую пару, недостаточно.

ДНК не имеет гладких сторон, как обычный цилиндр: на поверхности молекулы ДНК между фосфатными группами образуются бороздки. Между сахарофосфатными каркасами спирали вокруг внешней поверхности ДНК расположены две бороздки разного размера, известные как основная и дополнительная бороздки (Рисунок 10). Некоторые белки связываются с ДНК, чтобы изменить ее структуру, регулировать транскрипцию или репликацию, и им легче взаимодействовать с основаниями (внутренними частями молекулы ДНК) на стороне главной бороздки. Основная бороздка приблизительно на 50% шире, чем второстепенная. Такая конформация ДНК является наиболее термодинамически стабильной, она максимизирует прочность водородных связей между основаниями и сводит к минимуму стерическое и электростатическое отталкивание. Побочная бороздка — это всего лишь трехмерная форма двухцепочечной ДНК, которая отражает молекулы, из которых она состоит. Большие фосфатные группы можно рассматривать как широкие трубы с обеих сторон, в то время как основания между нитями являются плоскими и, таким образом, образуют бороздку, независимо от того, намотана ли она в спираль или вытянута.

Рассмотрим i -ую базовую пару: центры масс ее нуклеотидов расположены в точках p_1 и p_2 соответственно (Рисунок 11). Единичный вектор, проведенный между этими центрами масс (n_1) и центр отрезка между ними (K_1) задают плоскость Π_1 . Для следующая пары мы проделываем аналогичные действия и получаем плоскость Π_2 ⁷. Тогда плоскость (Π_3), которой принадлежит искомая центральная точка (C_1) включает в себя точку K_1 и перпендикулярна вектору, образуемому векторным произведением нормальных векторов плоскостей Π_1 и Π_2 (n_1 и n_2). Тогда искомая точка C_1 находится как пересечение этой плоскости с вектором пересечения плоскостей Π_1 и Π_2 .

Следующую базовую пару мы выбираем не по порядку, поскольку в таком случае эти векторы, образуемые нуклеотидами, будут почти параллельны, и угол между ними будет очень мал. Так что при решении системы уравнений, полученной из всех

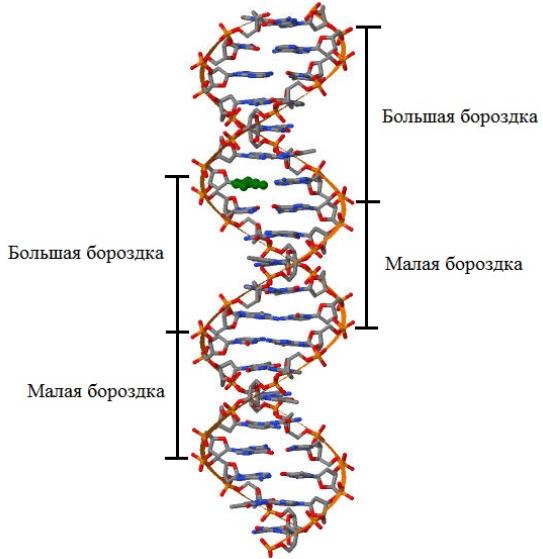


Рис. 10: Нити ДНК расположены ближе друг к другу на одной стороне спирали, чем на другой, образуя бороздки. Бороздки закручиваются вокруг молекулы с противоположных сторон. Изображение заимствовано из [36].

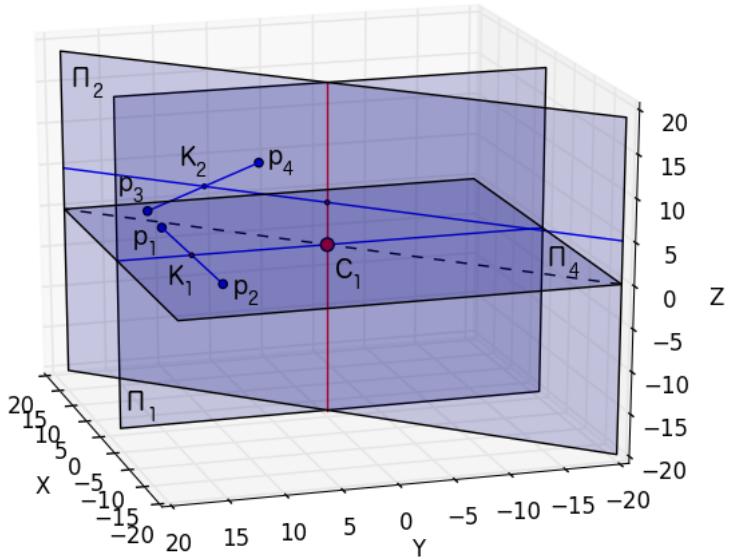


Рис. 11: Схематическая визуализация процесса нахождения точки C_1 , центральной для базовой пары $\{p_1, p_2\}$. Изображение заимствовано из статьи [30].

этих уравнений плоскостей, будем получать близкие к нулю значения, что приведет к вычислительным проблемам. Таким образом из ± 3 соседних пар выбирается такая, что косинусное расстояние между ней и i -ой парой максимально:

$$1 - \cos(\widehat{n_1 n_2}) \rightarrow \max_{n_2}.$$

Этот процесс повторяется для каждой базовой пары, и таким образом находится центральная линия для двухцепочечной ДНК (Рисунок 12).

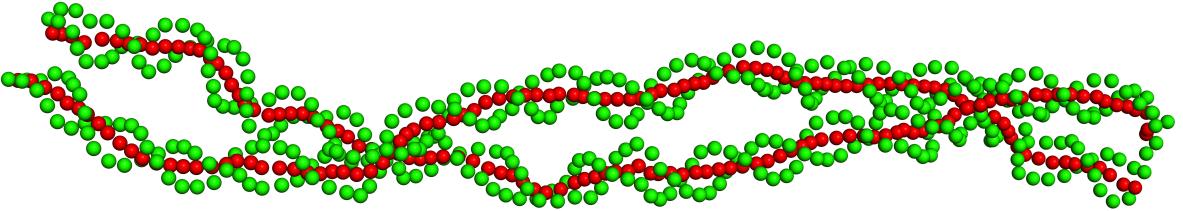


Рис. 12: Визуализация полученной центральной линии для двухцепочечной ДНК. Зеленым обозначены частицы, представляющие собой нуклеотиды, красным — центральная линия. Изображение заимствовано из статьи [30].

3.3.3 Расчет персистентной длины

Персистентная длина является основным свойством, определяющим жесткость полимера (а в нашем случае ДНК). Полимерная цепь, длина которой меньше персистентной, ведет себя как гибкий упругий стержень. Если же длина цепи больше персистентной, то ее свойства могут описываться только статистически. Другими словами персистентная длина (L_p) — это длина, на которой не чувствуются корреляции в направлении касательных полимерной цепи.

Для заряженных полимеров персистентная длина за счет электростатического экранирования зависит от концентрации соли в среде. В этой работе мы воспользуемся этим свойством и сравним персистентную длину, вычисленную для используемой модели двухцепочечной ДНК, с ее экспериментально полученными значениями [37, 38].

Для расчетов персистентной длины использовалась двухцепочечная ДНК длиной в 100 базовых пар при разных концентрациях ионов натрия и хлора (Таблица 3). Для этих параметров модели оценка персистентной длины происходит в несколько этапов:

1. Мы находим центральную линию ДНК (подробно этот процесс описан в Разделе 3.3.2).
2. Далее строим унитарные векторы \mathbf{u}_i для 90 пар нуклеотидов в середине цепи (а именно с 5 по 95), взяв вектор, соединяющий точку центральной линии i -й пары с аналогичной для $i + 1$ пары:

$$\mathbf{u}_i = \frac{\mathbf{r}_{c,i+1} - \mathbf{r}_{c,i}}{|\mathbf{r}_{c,i+1} - \mathbf{r}_{c,i}|} \quad (20)$$

3. Вычисляем косинус угла между двумя векторами \mathbf{u}_i и \mathbf{u}_j разделенными разным количеством базовых пар: $N_a = 2, 3, \dots, 84$.
4. Для каждого N_a мы вычисляем персистентную длину L_p , используя следующее выражение:
$$L_p = -\frac{N_a a}{\log(\langle \cos \alpha \rangle)}, \quad (21)$$
где $a = 0.356$ нм — длина базовой пары, а $\langle \cos \alpha \rangle$ — средний косинус, взятый для каждой траектории в каждый сохраненный момент времени и для каждого унитарного вектора \mathbf{u}_i , разделенного N_a базовыми парами. Вычисленная таким образом персистентная длина (Уравнение 22) получается зависимой от N_a (Рисунок 13). Для точности вычислений было использовано несколько траекторий, которые после были усреднены (жирные кривые на графике).
5. Усредненные кривые далее аппроксимируются функцией

$$L_p^{current}(N_a) = L_p(1 - \exp(-kx)), \quad (22)$$

дабы получить L_p . Здесь мы предполагаем, что $L_p(N_a)$ при больших значениях N_a выйдет на плато.

На Рисунке 14 точками показана персистентная длина для разных концентраций соли в растворе, полученная экспериментально [37, 38] (черные квадратики) и вычисленные по описанному выше способу для используемой модели MRG-CG (красные кружки). Как мы видим из полученных графиков значения персистентной длины для модели получились ниже, чем экспериментальные значения. Одной из причин этого является значение длины базовой пары, $a = 0,356$ нм. Оценки это значение из модели: $a = 0,387 \pm 0,047$ нм, что немного больше реального значения.

Однако стоит отметить, что качественно зависимость персистентной длины от концентрации ионов для используемой модели соответствует экспериментальным значениям. Более того при физиологических значениях концентрации ионов NaCl, которые представляют наибольший интерес, полученные результаты близки к экспериментальным.

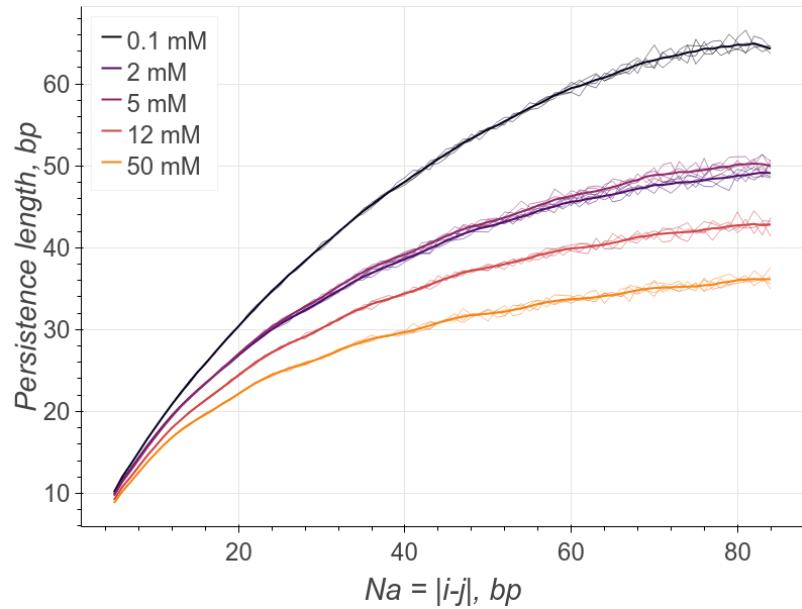


Рис. 13: Результаты моделирования с использованием термостата Нозе-Хувера. На графике представлена зависимость персистентной длины L_p от расстояния разделения N_a для различных концентраций $NaCl$, рассчитанная с использованием метода, описанного в Разделе 3.3.3. Тонкие линии представляют необработанные данные для нескольких траекторий, жирные линии — их усреднение.

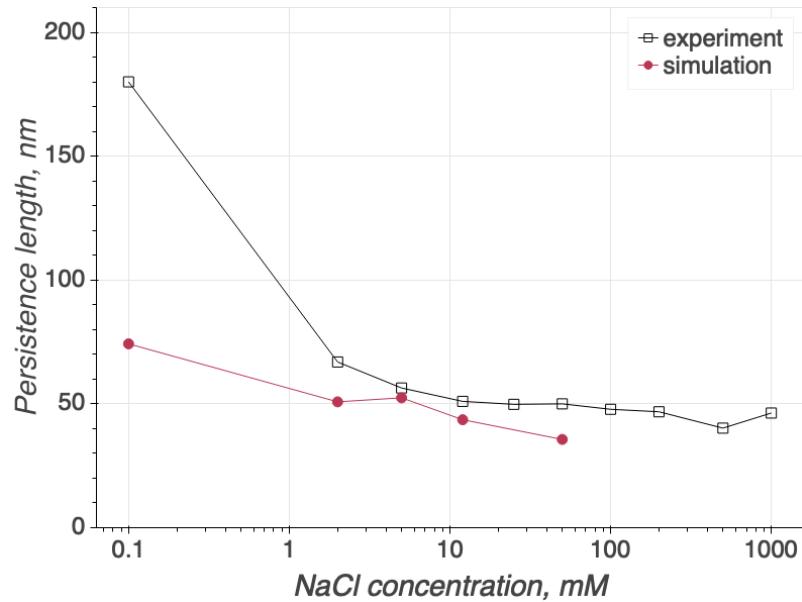


Рис. 14: Зависимость персистентной длины ДНК от концентрации $NaCl$. Экспериментальные данные отображены на чёрной кривой. Красная линия визуализирует результаты моделирования на GPU с использованием термостата Нозе-Хувера.

3.3.4 Вычисление радиального распределения ДНК и ионов внутри капсиды

Как было сказано в Разделе 1 структура ДНК внутри капсиды до конца не разрешена и ее точная конформация на данный момент неизвестна. Однако благодаря полученной возможности рассматривать ДНК как единую цепь (а не скрученные две цепи), можно определить как молекула ДНК ведет себя на каждом шаге упаковки, а также ее конечную структуру. Для вычисления всех интересующих нас параметров системы (таких как количество слоев ДНК, расстояние между слоями) в первую очередь необходимо найти радиальное распределение частиц.

Функция радиального распределения $g(r)$ описывает изменение плотности частиц в системе относительно любого выбранной частицы. Эта функция есть частный случай корреляционной функции $g^{(n)}(r_1, \dots, r_n)$. В общем случае выражение для $g^{(n)}(r_1, \dots, r_n)$ записывается следующим образом:

$$g^{(n)}(r_1, \dots, r_n) = \frac{V^n N!}{N^n (N-n)!} \frac{\int \cdots \int \exp\left(-\frac{U_N}{k_B T}\right) dr_{n+1} \dots dr_N}{Z_N}, \quad (23)$$

где U_N — полная потенциальная энергия системы — функция от всех координат частиц, N — количество частиц в системе, V и T — объем и температура соответственно, k_B — постоянная Больцмана, а Z_N называется классическим конфигурационным интегралом (Уравнение 24). Этот интеграл является классическим аналогом канонической статистической суммы (термин «классическая» подразумевает, что энергетические уровни предполагаются непрерывными, что верно для большинства систем).

$$Z_N = \int \cdots \int \exp\left(-\frac{U_N}{k_B T}\right) dr_1 \dots dr_N \quad (24)$$

$g^{(n)}(r_1, \dots, r_n)$ описывает корреляцию между частицами, возникающую за счет межатомное взаимодействие U_N . Если частицы абсолютно независимы друг от друга (другими словами $U_N = 0$), то $g^{(n)}(r_1, \dots, r_n) = \rho^n$, где ρ — объемная плотность системы. Если система однородна $\lim_{r \rightarrow \infty} g(r) = 1$. Помимо этого применяется условие нормировки:

$$4\pi\rho \int_0^\infty g(r) r^2 dr = N - 1 \approx N. \quad (25)$$

$g(r)$ также можно рассматривать как нормированную функцию распределения плотности:

$$g(r) = \frac{\rho(r)}{\rho}, \quad (26)$$

где $\rho(r)$ — локальная плотность, а ρ — объемная плотность.

В случае молекулярной динамики мы будем рассматривать не непрерывную, а дискретную функцию радиального распределения. Функция радиального распределения частиц определялась на каждом шаге по времени (поскольку с каждым шагом уменьшается радиус сферы и сжимается ДНК, меняя распределение и частиц ДНК, и ионов). Помимо этого плотность можно представить как $\rho = N/V$. Таким образом функция радиального распределения относительно начала координат (поскольку это есть центр сжимающей сферы) высчитывалась следующим образом:

$$g(r) = \frac{dN/N}{dV/V}, \quad (27)$$

где dN — количество частиц в кольце объема dV (на расстоянии $(r + dr)$ от начала координат). То есть мы находим радиус-вектор каждой частицы, и считаем количество радиус-векторов попавших в полуинтервал $[r, r + dr]$, тогда количество всех частиц — это количество всех векторов в расчетной области. Полный объем расчетной области: $V = 4/3\pi R^3$; расчетная область у нас кубическая, однако мы область наших интересов — сфера, поэтому ионами в углах расчетной области мы пренебрегаем. В свою очередь объем слоя толщиной $r + dr$ вычисляется следующим образом:

$$dV = 4/3\pi(r + dr)^3 - 4/3\pi r^3 = 4\pi r^2 dr, \quad (28)$$

здесь мы пренебрегли квадратными и кубическими членами, содержащими dr .

Таким образом мы оцениваем функцию радиального распределения путем дискретизации радиуса в интервалы:

$$g(r) = \frac{4/3\pi R^3}{4\pi r^2 dr} \frac{\sum_i [r \leq r_i < r + dr]}{\sum_i [r_i]}, \quad (29)$$

где r_i — радиус вектор i -ой частицы, R — радиус расчетной области, а r — текущий радиус (от 0 до R с шагом dr).

Стоит отметить, что полный объем системы и общее количество частиц — константы, так что учитывая, что $g(r)$ мы нормируем, ими можно пренебречь. Более того,

полученная функция распределения — дискретна, что не совсем удобно при анализе количества слоев ДНК (которое равно количеству максимумов функции радиального распределения). Поэтому можно применить ядерную оценку плотности как альтернативу гистограмме для сглаживания $g(r)$:

$$\hat{g}(r) = 2 \frac{1}{Nh} \sum_{i=1}^N K\left(\frac{r - r_i}{h}\right), \quad (30)$$

где K — ядерная функция (гладкая симметричная функция); $h > 0$ — полоса сглаживания, которая контролирует величину сглаживания. В этой работе в качестве ядерной функции была выбрана функция Гаусса, а величина полосы сглаживания $h = 0.1$ нм.

Такие вычисления были проведены для каждого кадра, отдельно для частиц ДНК, натрия и хлора. Функция радиального распределения позволяет проследить как меняется положение ДНК и ионов в зависимости от степени сжатия сферой (было смонтировано видео для визуализации процесса упаковки с отображением функции радиального распределения). На Рисунке 15 показаны функции радиального распределения для разных радиусов сжатия, где каждому графику приведены соответствующие структуры ДНК (с указанием радиуса сжимающей сферы и давления внутри капсиды).

Благодаря этим графикам было обнаружено, что ДНК действительно образует слои при упаковке, причем с уменьшением радиуса сжатия слои становятся более выраженным и расстояние между ними уменьшается. Также хорошо заметно на графиках (Рисунок 15), как ионы натрия конденсируются на молекуле ДНК, при этом полностью выталкивая ионы хлора из капсиды. Помимо этого была определена конечная конформация упакованной ДНК — на Рисунке 15 видно, что ДНК завернута в клубок (дальнейший анализ слоев и структуры ДНК — в Разделе 3.3.5).

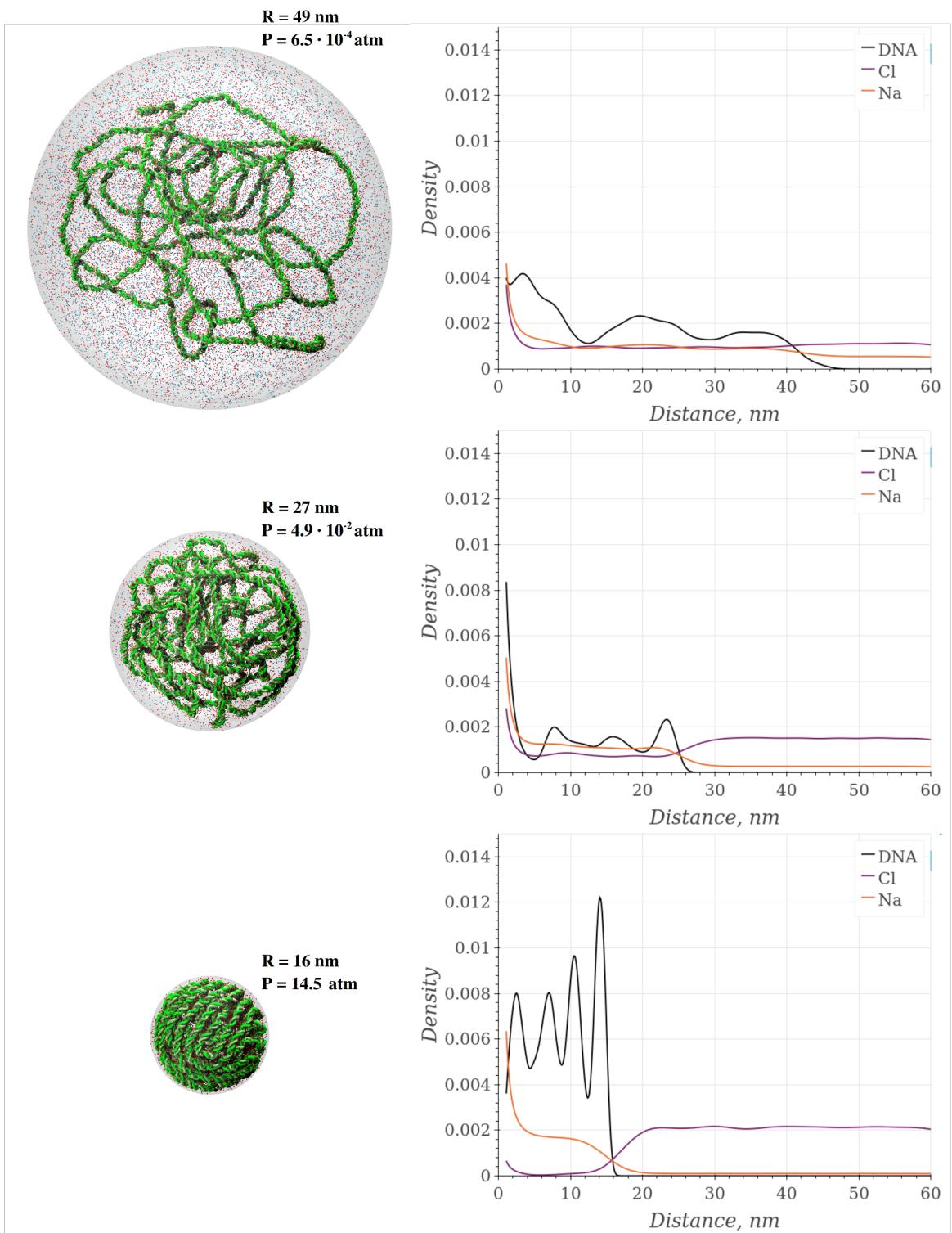


Рис. 15: Функция распределения частиц ДНК, Na и Cl во время процесса упаковки ДНК. Графикам справа соответствуют изображения слева. С уменьшением радиуса сжимающей сферы слои ДНК становятся более явными: это видно как на графике справа (черная кривая соответствует ДНК), так и по структуре ДНК слева (при максимальном сжатии ДНК приняла форму клубочка). Поскольку цепочки ДНК сильно отрицательно заряжены и при сжатии они оказываются близко друг к другу, увеличивается сила отталкивания между ними. Для уменьшения этой силы положительные ионы натрия конденсируются на ДНК (оранжевая кривая на графиках справа), выталкивая отрицательно заряженные ионы хлора (фиолетовая кривая на графиках).

3.3.5 Определение количества слоев ДНК и расстояния между ними

Как видно на Рисунке 15 максимумы функции радиального распределения указывают на большую плотность. Другими словами, максимумы функции $g(r)$, если говорить о молекуле ДНК, — это слои. Следовательно посчитав эти максимумы, и определив положение, можно найти количество слоев и расстояние между ними. Что и было сделано в рамках этой работы.

Для нахождения максимумов использовалось исследование функции радиального распределения через производные, а именно поиск точек, где производная меняет знак с + на -. Значения функции радиального распределения дискретны (с шагом в 0.1 нм), так что для нахождения производных в точках использовался метод численного дифференцирования. Для аппроксимации производной $\frac{df}{dx}$ в точке x_i разложение в ряд Тейлора даст:

$$\frac{df}{dx}(x_i) \approx \frac{1}{\Delta x} \left[\sum_{k=m}^n c_k f(x_{i+k}) + \sum_{l=n-m+1}^{\infty} D_l \frac{f^{(l)}(x_i) \Delta x^l}{l!} \right], \quad (31)$$

где $f^{(l)}$ — производная функции f l -го порядка.

Первая часть Уравнения 31 — это схема конечных разностей, использующаяся для аппроксимации частичной разности, а вторая часть — ошибка усечения. Порядок точности в Уравнении 31 определяется разностью $|n - m|$. Коэффициенты c_k и D_l — известны. Для нашей задачи нам нужно найти первую производную с достаточно хорошей точностью, которую можно добиться методом центральной конечной разности 6 и 8 порядка точности. Было выбрано несколько вариантов, дабы исключить ошибки. Коэффициенты c_l даны в Таблице 4.

Таблица 4: Коэффициенты для метода центральной конечной разности для первой производной в точке x_i

Точность	x_{i-4}	x_{i-3}	x_{i-2}	x_{i-1}	x_i	x_{i+1}	x_{i+2}	x_{i+3}	x_{i+4}
6		$-\frac{1}{60}$	$\frac{3}{20}$	$-\frac{3}{4}$	0	$\frac{3}{4}$	$-\frac{3}{20}$	$\frac{1}{60}$	
8	$\frac{1}{280}$	$-\frac{4}{105}$	$\frac{1}{5}$	$-\frac{4}{5}$	0	$\frac{4}{5}$	$-\frac{1}{5}$	$\frac{4}{105}$	$-\frac{1}{280}$

Итак, начиная с момента, когда функция начинает возрастать, мы движемся вдоль функции радиального распределения до границы расчетной области, считая, сколько раз ее производная меняет знак, а так же записывая ее положение (рассто-

жение от начала координат). Аналогичные вычисления проводились для каждого кадра траектории.

С использованием такого подхода были проанализированы траектории для разных длин и разных концентраций. Результаты вычислений показаны в виде нормированной гистограммы на Рисунке 17(а-б). Из графиков видно, что как и в случае с давлением, количество слоев слабо зависит от концентрации ионов в системе (всего 4-5 слоев образуется при упаковке 4000 базовых пар ДНК). При этом для разных длин ДНК количество слоев различается: от 2 до 6-7 слоев. Изменение количества слоев для одной и той же траектории во время упаковки вполне логично: при увеличении давления ДНК хочет занять наиболее выгодную с точки зрения энергии форму, и если в ней присутствуют какие-то большие изгибы, она перестраивается, чтобы таких изгибов не было, что позволяет сжимать ДНК до малых размеров.

Как уже упоминалось, были зафиксированы положения максимумов функции радиального распределения относительно начала координат. Эта информация была использована для послойной визуализации ДНК (Рисунок 16). На этом рисунке показаны срезы слоев ДНК, визуализированные при помощи программы VMD (Visual Molecular Dynamics). Здесь хорошо видно, что ДНК свернулась в клубок, при этом большая часть нитей ДНК расположена параллельно, поскольку в таком положении отталкивание между нитями минимальное. Всего в этой ДНК выделено 4 слоя: внешние три похожи по структуре — параллельные равномерные нити, иногда загибаются; внутренний слой отличается от всех остальных — он состоит из 4 нитей ДНК расположенных друг относительно друга крестом.

Помимо визуализации положение максимумов позволило найти расстояние между нитями ДНК в процессе ее упаковки, поскольку нам всего лишь осталось найти расстояние между этими максимумами, что и было сделано (Рисунок 17(с-д)). Такие вычисления также были проведены для различных систем (Таблица 3).

По графикам на Рисунке 17(с) можно увидеть, при каком сжатии образуется дополнительный слой (разрыв второго рода на прямой) для маленьких концентраций соли NaCl, а для больших концентраций такого разрыва не происходит. Этот же эффект видим на Рисунке 17(д) — все прямые могут быть аппроксимированы непрерывной функцией. Более того, эти графики позволяют определить расстояние между ДНК в сжатом виде: эта величина для ДНК длиной в 4000 базовых пар при сжатии в сферу с радиусом, равным внутреннему радиусу вируса Гепатита В (12.5 нм) равна 2-3 нм. Этот результат хорошо согласуется с реальными значениями расстояния между слоями

ДНК для Гепатита В, которое равное 2.6-2.9 нм [39].

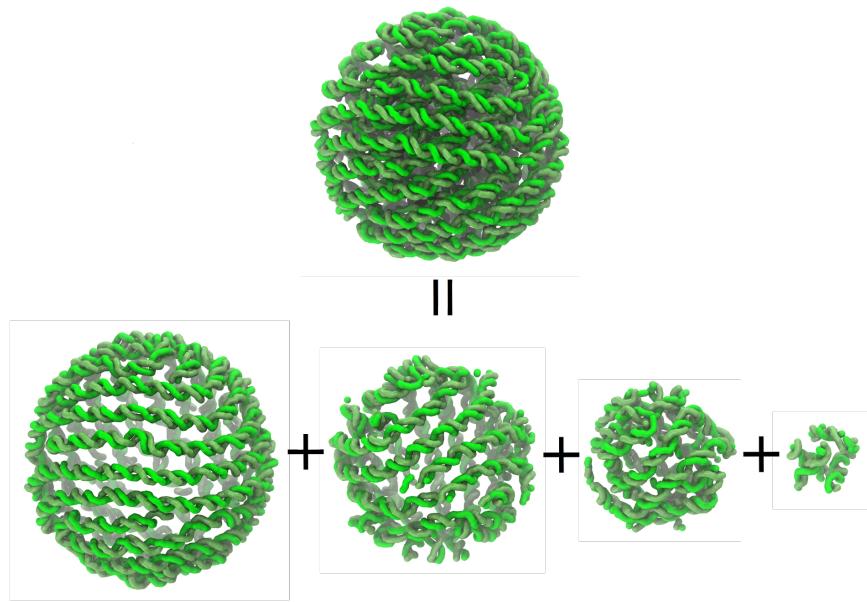


Рис. 16: На основании положения максимумов функции радиального распределения ДНК были разрешены и визуализированы слои сжатой ДНК. Каждый слой кроме внутреннего имеет структуру клубочка: большинство нитей ДНК расположены параллельно друг другу, однако некоторые сильно загнуты. Внутренний слой представляет из себя перекрестье четырех цепей ДНК.

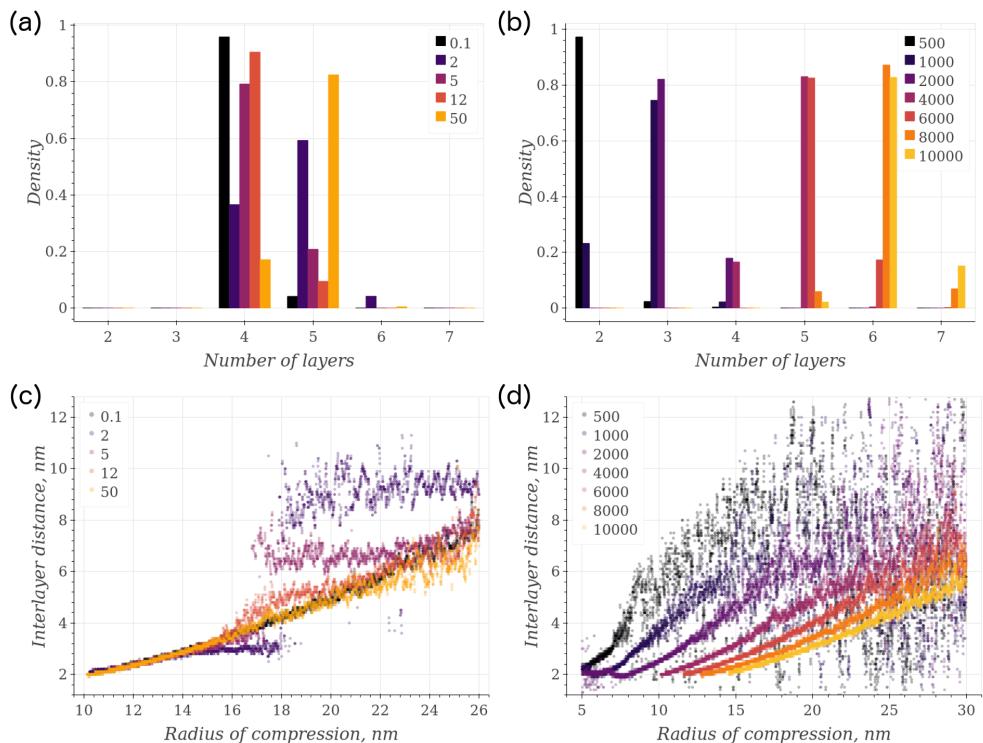


Рис. 17: Распределение количества слоев и расстояния между слоями во время упаковки ДНК длиной 4000 базовых пар при разных концентрациях ионов: (а) и (с) соответственно, а также для системы с ионной силой 50 мМ, но для ДНК разного размера: (б) и (д).

4 Вычислительные методы для описания механических свойств вирусов

Вирусные капсиды должны отвечать противоречивым целям: они должны быть достаточно стабильными, чтобы защитить свой геномный материал, и при этом быть достаточно нестабильными, чтобы они высвобождали свой геном в процессе инфицирования клеток хозяина. В вирусных оболочках не содержится какого-либо метаболического механизма, следовательно взаимосвязь между структурой, механикой и функцией в этих системах можно исследовать, используя равновесную термодинамику и неравновесную статистическую механику.

Одним из основных способов, с помощью которого были экспериментально описаны вирусы и вирусоподобные структуры, является измерение силы во время их наноиндентирования при помощи атомно-силового микроскопа (ACM) [4]. Схема установки для измерения динамической силы на основе ACM для биологической частицы (вируса) представлена на Рисунке 22. Хотя этот метод и обладает относительно высоким разрешением, ему не хватает контроля по ряду причин: нет точного знания молекулярного положения, при котором происходит наноиндентирование; ему не хватает точного представления о том, как объект (вирусная частица) деформируется кончиком ACM; и как результат, нельзя точно соотнести изменения энергии с процессом наноиндентирования. Но основная проблема заключается в том, что эти параметры можно определить из динамической силы, полученной при помощи ACM *in vitro* [4, 40], поскольку такое сжатие деформирует структуру вируса и разрывает нековалентные межбелковые связи. На практике, поскольку вирусная капсида состоит из нескольких копий капсомеров молекулярная интерпретация экспериментальных результатов на уровне одинарных связей и в масштабе нанометровых расстояний практически невозможна.

В этом разделе мы представляем подход к многомасштабному моделированию, которое называется «наноиндентирование *in silico*». Этот метод позволяет численно исследовать биомеханику вирусных частиц. Помимо исследования пустой капсиды Гепатита В, было также проведено наноиндентирование вирусной частицы с генетическим материалом внутри. В предыдущих разделах была получена модель ДНК, упакованная в сферическую оболочку радиуса соответствующего внутреннему радиусу оболочки Гепатита В. Однако для моделирования целого вируса также необходимо создать модель для капсиды. Здесь так же как и для моделирования ДНК для уменьшения количества степеней свободы было решено использовать крупнозернистую модель. Для

сохранения структурного разрешения был использован C_{α} -крупнозернистый подход, в котором каждая аминокислота была представлена своим атомом C_{α} . Этот метод хорошо показал себя в экспериментах по наноиндентированию капсид вируса крапчатости фасоли (Cowpea chlorotic mottle virus, CCMV) [9].

4.1 Модели для описания вируса Гепатита: пустого и с генетическим материалом

4.1.1 Модель самоорганизующегося полимера для описания капсиды вируса

Модель самоорганизующегося полимера (СОП) позволяет описывать большие белки и белковые комплексы, включая капсиды вирусов [12, 41, 42]. Эта модель может быть использована для описания белковых субъединиц, из которых состоит капсida вируса. Модель СОП основана на топологии, где каждый аминокислотный остаток описывается одним центром взаимодействия — атомом C_α (Рисунок 18), а белковая цепь — набором ковалентных связей $C_\alpha - C_\alpha$ с расстоянием связи $a = 3.8 \text{ \AA}$ (длина пептидной связи).

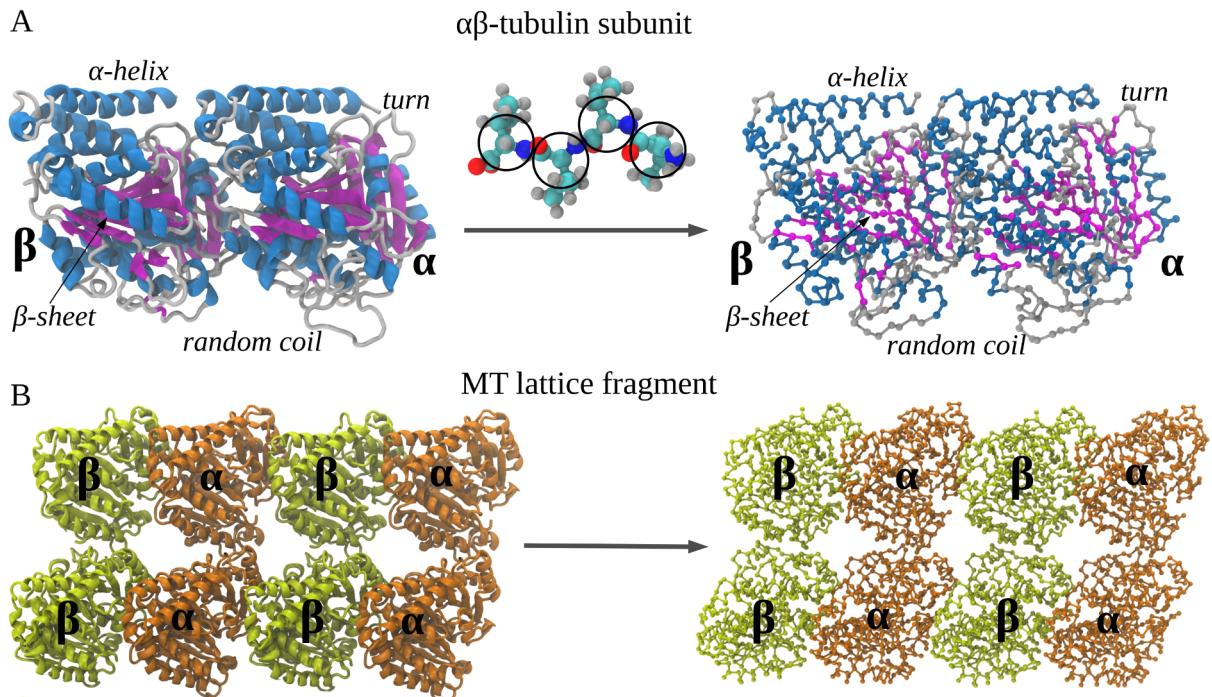


Рис. 18: Пример перехода от полноатомной к крупнозернистой модели самоорганизованного полимера (СОП) полипептидной цепи. (а) иллюстрирует этот процесс для атомной структуры димера $\alpha\beta$ -тубулина — структурной единицы цилиндра микротрубочек. Аминокислотные остатки заменены единичными центрами взаимодействия (сферическими частицами) с координатами C_α -атомов (обозначены черными кружками). Четыре представительных круга показаны в качестве примера процесса крупнозернистой структуры. Следовательно, белковый каркас заменяется набором ковалентных связей $C_\alpha - C_\alpha$ с расстоянием 3.8 \AA . На (б) изображены результаты крупнозернения небольшого фрагмента цилиндра микротрубочек, полученные аналогичным способом. Иллюстрация заимствована из sop-gpu.readthedocs.io.

Потенциальная энергия взаимодействий U_{SOP} определяется в терминах коорди-

нат C_α атомов и записывается в следующем виде [42, 43]:

$$U_{SOP} = U_{FENE} + U_{NB}^{ATT} + U_{NB}^{REP}, \quad (32)$$

U_{FENE} описывает ковалентные связи.

$$U_{FENE} = - \sum_{i,j}^{bonds} \frac{kR_0}{2} \log \left(1 - \frac{(r_{ij} - r_{ij}^0)^2}{R_0^2} \right). \quad (33)$$

$k = 14$ Н/м – есть жесткость, $R_0 = 2.0$ Å – допустимое изменение длины ковалентной связи. Расстояние между частицами C_α - C_α i и j – есть r_{ij} , а r_{ij}^0 – его значение в исходной (нативной) структуре. Суммирование производится по всем ковалентно связанным парам i и j.

Потенциал U_{NB}^{ATT} учитывает нековалентные взаимодействия, стабилизирующие нативную структуру белка, и описывается потенциалом Леннарда-Джонса :

$$U_{NB}^{ATT} = \sum_{i,j}^{native} \varepsilon_h [(r_{ij}^0/r_{ij})^{12} - 2(r_{ij}^0/r_{ij})^6] \quad (34)$$

r_{ij} – это расстояние между двумя атомами i и j. Параметры ε_h и r_{ij}^0 определяют силу и равновесное расстояние нативных контактов. Суммирование идет по всем атомам, образующим нативные контакты. Если два атома i и j разделены более чем двумя ковалентными связями и $r_{ij} < R_C$, то они образуют нативную связь. Параметры ε_h и r_{ij}^0 определяют силу и равновесное расстояние нативных контактов и должны подбираться с использованием полноатомного моделирования того же белка.

Ненативные (несвязанные) взаимодействия рассматриваются как отталкивающие:

$$U_{NB}^{REP} = \sum_{i,j}^{rep} \varepsilon_r (\sigma_r/r_{ij})^6 \quad (35)$$

Значения параметров $\varepsilon_l = 1$ ккал/моль и $\sigma_l = 3.8$ Å. Они определяют силу и диапазон отталкивания. Благодаря этому U_{NB}^{REP} исключает самопересечения полипептидной цепи.

Динамика системы получается путем численного решения уравнений Ланжеvena движения для каждого положения частицы r_i без учета инерции:

$$\xi \frac{dr_i}{dt} = -\frac{\partial U_i(r_i)}{\partial r_i} + G_i(t) \quad (36)$$

$U_i(r_i)$ здесь – полная потенциальная энергия, учитывающая все взаимодействия между частицами в молекуле – U_{SOP} из Уравнения (32), $G_i(t)$ – случайная сила, а ξ – коэффициент трения [44, 45].

4.1.2 Модель капсида и вирусной частицы Гепатита В с генетическим материалом

Для описания капсида Гепатита В из представленных в базе данных PDB (Protein Data Bank) была выбрана структура с кодом 1QGT. Эта структура была получена методом дифракции рентгеновского излучения с разрешением 3.3 Å [46] и сохранена в базе в виде последовательностей четырех цепей (Рисунок 19). Однако капсида была разрешена не полностью, в связи с чем недостающие элементы последовательности белковой структуры были подобраны с использованием программы MODELLER при помощи UCSF Chimera [47].

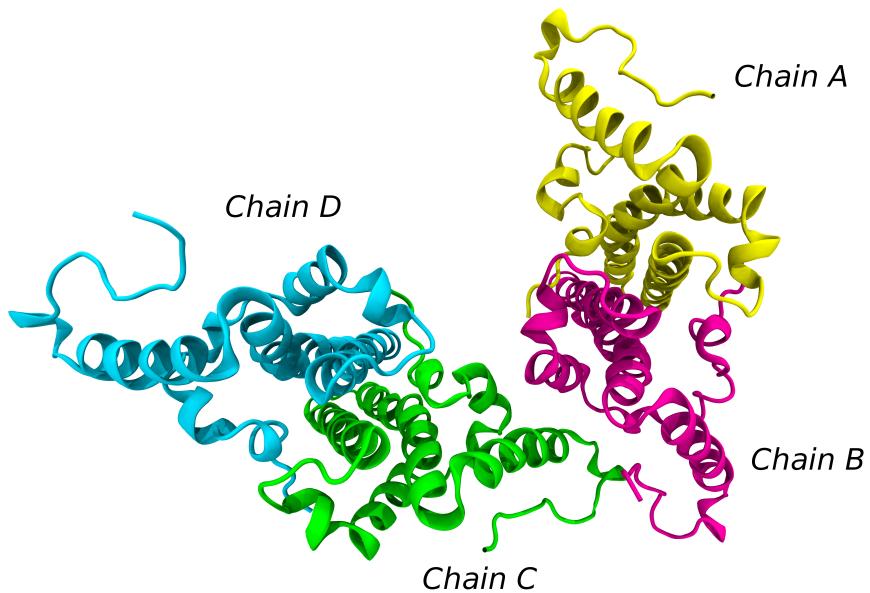


Рис. 19: Капсомер структуры 1QGT вируса Гепатита В. Разные участки цепи обозначены разными цветами.

Затем, капсомеры были сдвинуты при помощи геометрических преобразований (параллельный перенос, поворот) так, чтобы капсомеры компоновались в соответствии с общей структурой вируса, полученного при помощи электронной микроскопии, которая также была взята из базы данных PDB. Для вычисления матриц поворота систем использовался алгоритма Кабша (англ. Kabsch algorithm) [48]. Таким образом была получена полноатомная модель вируса гепатита В.

Поскольку полноатомное моделирование требует большой вычислительной мощности и длительных вычислений, была создана крупнозернистая модель (Раздел 4.1.1), при помощи которой можно существенно сократить число степеней свободы молекулы. Таким образом группы атомов (в этой работе аминокислоты) заменяются одним центром, взаимодействия между которыми задаются при помощью потенциальной энергии (Уравнение 32).

Таким образом мы имеем модель MRG-CG для двухцепочечной ДНК (Раздел 3), а также модель капсиды Гепатита В. Соединив модели для двухцепочечной ДНК длиной 3182 базовых пары (что равно длине генома HBV) и капсиды, была получена гибридная модель вируса Гепатита В.

Поскольку ДНК находится внутри капсиды и не проникает сквозь нее, для описания взаимодействия ДНК с оболочкой вируса использовался отталкивающий потенциал Леннарда-Джонса, описанный в Уравнении (35). Однако помимо ДНК в системе присутствуют ионы, которые также оказывают влияние на капсиду. Взаимодействие белка с ионами описывается так же как и взаимодействие ДНК с ионами, а именно Уравнением (3). Радиус взаимодействия определяется как среднее арифметическое радиусов нуклеотида и аминокислоты (4.4 \AA). Заряды считаются размещенными на C_α атомах заряженных аминокислот.

Для подтверждения корректности гибридной модели, в первую очередь были проведены вычислительные эксперименты без приложения сил, при температуре 0 К в окружении ионов NaCl. Однако такой геном оказывал слишком сильное давление на капсиду и в итоге разрушал ее изнутри (Рисунок 20).

Это связано с тем, что мы использовали модель для двухцепочечной ДНК, в то время как реальный геном вируса Гепатита В имеет часть одноцепочечной. В связи с этим было решено убрать 671 частицу из одной нити ДНК, но с теми же параметрами что и для двухцепочечной ДНК. Помимо этого геном Гепатита В замкнутый, что также было изменено в существующей модели ДНК.

После проведения всех этапов подготовки и упаковки уже для смешанной ДНК (все этапы описаны в Разделе 3), была получена модель Гепатита В, не разрушающаяся при отсутствии внешних сил и хорошо описывающая структуру этого вируса.

После благополучного тестирования модели, была проведена подготовка к наноиндентированию ее *in silico*. В первую очередь была минимизирована полная энергия системы: сначала с фиксированной капсидой — для минимизации энергии ДНК внутри капсиды, а после и целого вируса (минимизация энергии проводилась с использованием

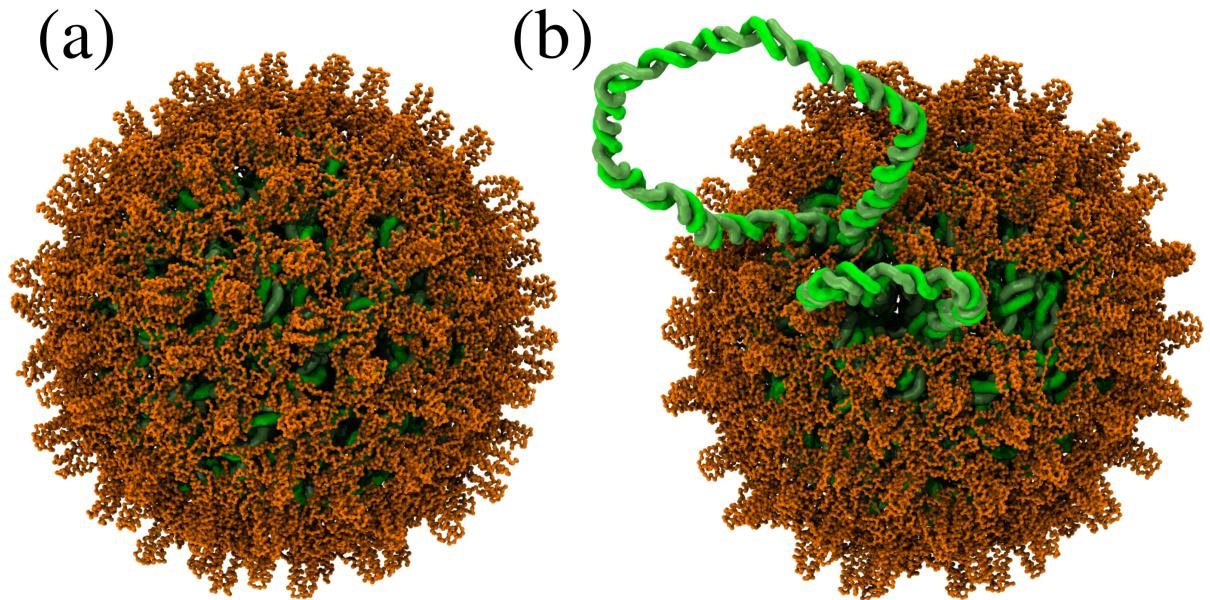


Рис. 20: Крупнозернистая модель вируса Гепатита В (белковая капсида, содержащая двухцепочечную ДНК длиной в 3182 базовые пары) (а). Под действием высокого давления на стенки капсиды геном “вырывается” наружу, разрушая капсиду (б).

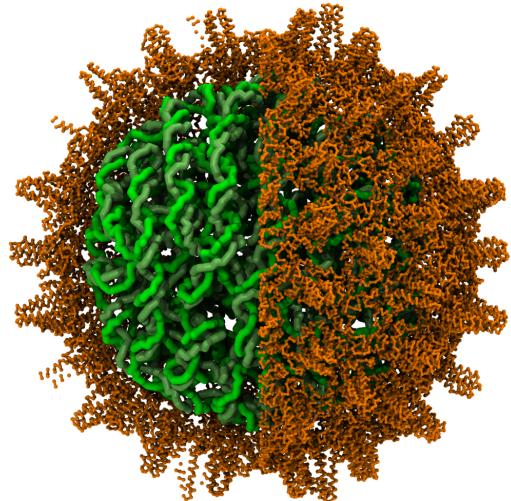


Рис. 21: После обнаружения проблемы с полностью двухцепочечной ДНК, было решено использовать более реалистичную модель ДНК: замкнутую и с отрезком одноцепочечной. Конечная гибридная модель вируса не разрушалась. Такая модель вируса была подготовлена для наноиндентирования: была минимизирована энергия, а также проведен процесс уравновешивания системы. На рисунке показана полученная структура вируса, в том числе и структура ДНК (ионы в системе присутствуют, но не отражены на иллюстрации).

алгоритма градиентного спуска, подробно описанная в Разделе 3.2.2). Далее аналогично подготовке ДНК для упаковки было проведено уравновешивание системы с использованием термостата Нозе-Хувера при помощи алгоритма перешагивания (Раздел 3.2.2):

но здесь в первой итерации уравновешивания капсида не фиксировалась, а во второй — фиксировалась, поскольку для наноиндентирования необходимо, чтобы вирус располагался строго под кончиком кантилевера. Результат подобной подготовки показан на Рисунке 21.

4.2 Наноиндентирование *in silico*

Измерение динамической силы позволяет исследовать биомеханику различных биологических систем. Эксперименты по деформации на основе АСМ дают кривые зависимости силы от деформации (F_X), которые содержат информацию о жесткости, эластичность, силах, необходимых для обратимой деформации и необратимого разрушения конструкций, и т.д. [8]. Несколько вирусных оболочек были исследованы с использованием подхода, основанного на АСМ: вирусы человека - вирус иммунодефицита человека, вирус гепатита В, аденовирус и вирус простого герпеса [49–52]; а также другие вирусы, инфицирующие эукариотические клетки. Деформация частицы при помощи АСМ может быть выполнена в нескольких режимах: (1) прямого вдавливания, когда основание кантилевера опускается с постоянной скоростью (v_f) по направлению к частице, постепенно увеличивая приложенную силу; (2) обратного принудительного втягивания, когда направление движения основания меняется на противоположное и сила уменьшается до нуля. Экспериментальные результаты часто позволяют определить сложный механизм деформации, но отсутствие структурной информации не дает детально интерпретировать наноразмеры и моделировать формы спектральных линий [4, 40].

Такие ограничения в экспериментальном подходе требуют разработки альтернативных стратегий исследования. Поэтому часто помимо экспериментального наноиндентирования частиц применяется методология «наноиндентирования *in silico*», которая позволяет понять механизмы деформации и разрушения оболочки вируса на структурном уровне (Рисунок 22) [9]. В этом подходе механическую загрузку и разгрузку биологической частицы осуществляют *in silico* (то есть вычислительными методами). Важно, что измерения наноиндентирования выполняются в экспериментальных условиях приложения динамической силы $f(t) = r_f t$ (Рисунок 22), т.е. при моделировании используются экспериментальные скорости нагружения силой (r_f). Структурные переходы могут быть определены при помощи анализа положения аминокислотных остатков, а биомеханические характеристики и термодинамические величины — посредством анализа энергии.

Наноиндентирование *in silico* обеспечивает детальное моделирование всего процесса деформации, релаксации и разрушения частиц, следовательно метод может использован для интерпретации экспериментальных результатов *in vitro*. Кроме того наноиндентирование *in silico* позволяет полностью контролировать протокол моделиро-

вания: амплитуду и направление приложенной силы, точное расположение контакта между концом кантileвера и биологической частицей. Аналогично экспериментам *in vitro*, мы можем изменить направление приложенной силы, величину амплитуды и др. Такой контроль над системой на протяжении всего процесса наноиндентирования может быть использован для изучения наноиндентирования при различных положениях частицы (будет ли кончик кантileвера давить на стык капсомеров или поверхность одного капсомера), а также для определения связи изменения характеристик системы со структурными изменениями частицы. Кроме того, благодаря режимам прямого и обратного хода кантileвера, мы можем моделировать повторяющиеся циклы вдавливания и релаксации, чтобы понять процесс «усталости» и структурного разрушения исследуемой биологической частицы.

При наноиндентировании *in silico* основание кантileвера представлено сферой, соединенной гармонической пружиной со сферическим кончиком радиуса R_{tip} , имитирующим кончик кантileвера (индентора) (Рисунок 22). Наконечник взаимодействует с биологическими частицами через отталкивающий потенциал Леннарда-Джонса:

$$U_{tip} = \sum_i^N \varepsilon_{tip} \left(\frac{\sigma_{tip}}{|r_i - r_{tip}| - R_{tip}} \right), \quad (37)$$

тем самым создавая углубление на внешней поверхности частицы. В Уравнении 37 r_i и r_{tip} — координаты i -й частицы и центра индентора соответственно; $\varepsilon_{tip} = 4.18$ кДж/моль и $\sigma_{tip} = 1.0$ Å являются параметрами взаимодействия, и суммирование выполняется по всем частицам под индентором.

Для кончика кантileвера численно решается следующее уравнение движения Ланжеvena:

$$\eta \frac{dr_{tip}}{dt} = -\frac{\partial U_{tip}(r_{tip})}{\partial r_{tip}} + \kappa \left((r_{tip}^0 - v_f t) - r_{tip} \right), \quad (38)$$

где r_{tip}^0 — начальное положение центра индентора; v_f — базовая скорость кантileвера; η — коэффициент трения ($7.0 \cdot 10^6$ нН с/м)* κ — жесткость пружины кантileвера.

Движущееся с постоянной скоростью (v_f) основание кантileвера действует через кончик кантileвера: зависящая от времени сила $\mathbf{f}(t) = f(t)\mathbf{n}$ в направлении \mathbf{n} перпендикулярна верхней поверхности вируса. В результате величина силы $f(t) = r_f t$, действующая на частицу, линейно возрастает за время t со скоростью нагружения силой $r_f = \kappa v_f$. При моделировании прямого хода основание кантileвера и индентор

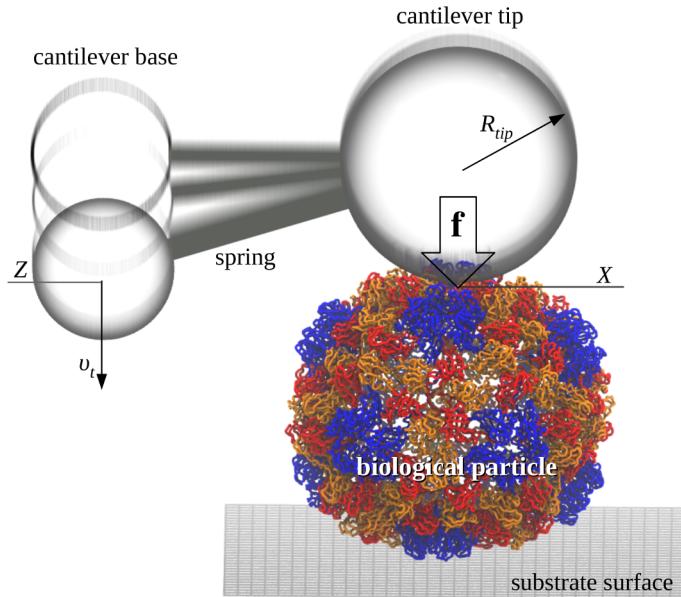


Рис. 22: Схема установки, используемой при наноиндентировании *in vitro* или *in silico*. Биологическая частица располагается на подложке. В режиме прямого хода основание кантилевера движется с постоянной скоростью v_f в направлении, перпендикулярном верхней поверхности частицы. Это создает силу, которая передается на индентор (сфера радиуса R_{tip}) по гармоническому закону. Величина силы (f) линейно возрастает со временем: $f(t) = r_f t$, где $r_f = k v_f$ — скорость нагружения. Отклик частицы измеряется путем оценки силы вдавливания F в зависимости от смещения основания кантилевера Z или как функции глубины вдавливания X . Изображение взято из работ [9, 11].

движутся в направлении вируса; при моделировании «втягивания силой» направление движения основания и кончика кантилевера консоли меняется на обратное, что приводит к постепенному уменьшению силы сжатия до нуля. Мы контролируем смещение основания кантилевера Z и положение кончика кантилевера X , которое определяет глубину вдавливания. Сила сопротивления F вирусной частицы, которая соответствует экспериментально измеренной силе вдавливания, может быть легко рассчитана с использованием выходной энергии. Кроме того, чтобы предотвратить скатывание капсида и имитировать адсорбцию частицы на подложке, мы всегда можем зафиксировать нижнюю часть частицы, фиксируя несколько атомов C_α , контактирующих с поверхностью подложки. Благодаря ускорению вычислений в ~ 200 раз, достигнутому на современных GPU, можно использовать экспериментально значимые значения скорости кантилевера ($v_f = 0.1\text{-}1.0 \text{ мкм/с}$), что является существенным преимуществом этого метода.

4.2.1 Наноиндентирование пустой капсида

Были проведены измерения динамической силы *in silico* для оболочки вируса Гепатита В. Мы использовали модель СОП для пустой оболочки HBV, полученной из полноатомной структуры капсида (код PDB: 1QGT) после всех этапов подготовки (Раздел 4.1.1). В качестве индентора была использована сфера радиуса $R_{tip} = 20$ нм. В этой работе были проведены вычислительные эксперименты по наноиндентированию только для прямого хода кантилевера: величина силы $f(t) = r_f t$ увеличивается со временем при скорости силового нагружения $r_f = \kappa v_f$, где $v_f = 0.001$ нм/пс $\kappa = 100$ кДж/моль·нм². Для проверки точности модели и параметров для поведения кантилевера были выполнены несколько вычислительных экспериментов. Эксперименты проводились при температуре 300 К, для фиксации капсида строго под кончиком кантилевера были точечно зафиксированы нижние части капсида. Полученные в результате зависимости силы от деформации для прямого вдавливания для оболочки вируса Гепатита В представлены на Рисунке 24).

На графиках видно, что кривая сила-деформация для оболочки Гепатита В характеризуются сильным увеличением силы деформации при небольшой глубине деформации. Также можно заметить что эта кривая практически линейна с примерно постоянным наклоном, dF/dZ , который количественно определяет механическую податливость частицы вируса к внешней силе сжатия. Этот режим сохраняется до тех пор, пока не будет достигнуто критическое значение деформации $Z^* = 8$ нм, и в этот момент сила деформации достигает критического (максимального) значения силы $F^* \approx 0.3$ нН. Эти величины определяют пределы механической устойчивости вирусной частицы. После перехода через эту точку в течение следующих 5 нм деформации частица находится в состоянии полуустойчивости: на графике видно кривая вышла на плато. А после этого этапа частица входит в динамический режим, в котором она подвергается самопроизвольному переходу с изменением формы из нативного (недеформированного) состояния в глобальный коллапс (полностью деформированный) состояние.

Подобный эксперимент был проведен для нескольких капсид Гепатита В *in vitro* [53]. Результаты этой статьи представлены на Рисунке 23. Как мы видим при сжатии капсида на 20 нм *in vitro* значение динамической силы составляет порядка 0.6 нН, в то время как для вычислительной модели (Рисунок 24) это значение равно 0.5 нН. Это говорит о том, что модель капсида корректна для проведения такого эксперимента. Однако как мы видим в результатах эксперимента *in vitro* нет плато, которое присутствует

приnanoиндентировании *in silico*.

Визуализация nanoиндентирования *in silico* также представлена на Рисунке 24. Все изображения соответствуют моменту сдвига основания кантелеverа на значение, указанное на рисунке.

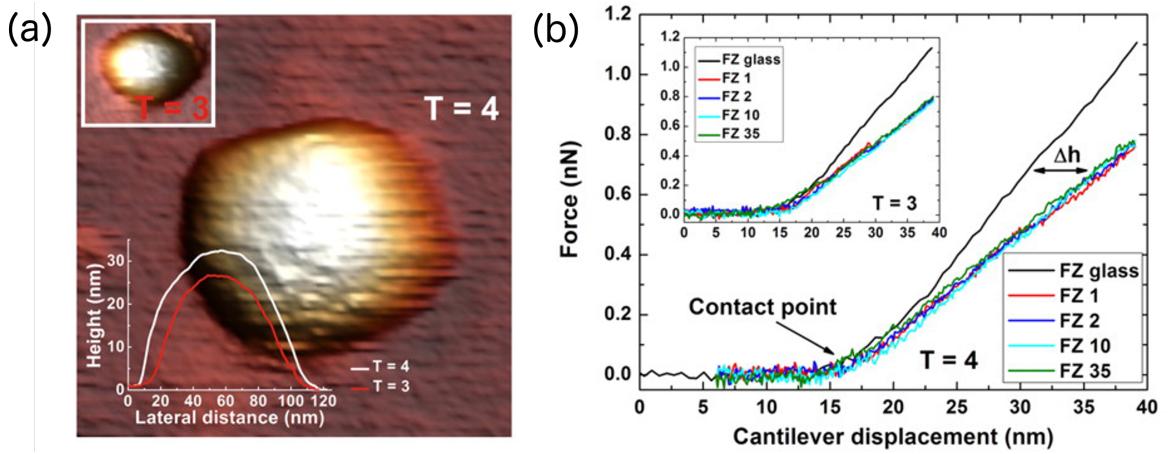


Рис. 23: Данные ACM для капсид вируса HBV, полученные экспериментально, взятые из статьи [53]. (а) ACM-изображения капсид T4 и T3 (верхняя левая вставка). Соответствующие профили высоты (нижняя левая вставка) показывают различие капсиды T3 и T4 (разница в диаметре составляет 15%). Боковые видимые диаметры размыты эффектом свертывания наконечника ACM. (б) Пример кривых зависимостей силы от расстояния (*FZ*) для частиц T4 и T3 (вставка). Эталонная кривая стекло *FZ* показана вместе с 1-й, 2-й, 10-й и 35-й кривыми *FZ*.

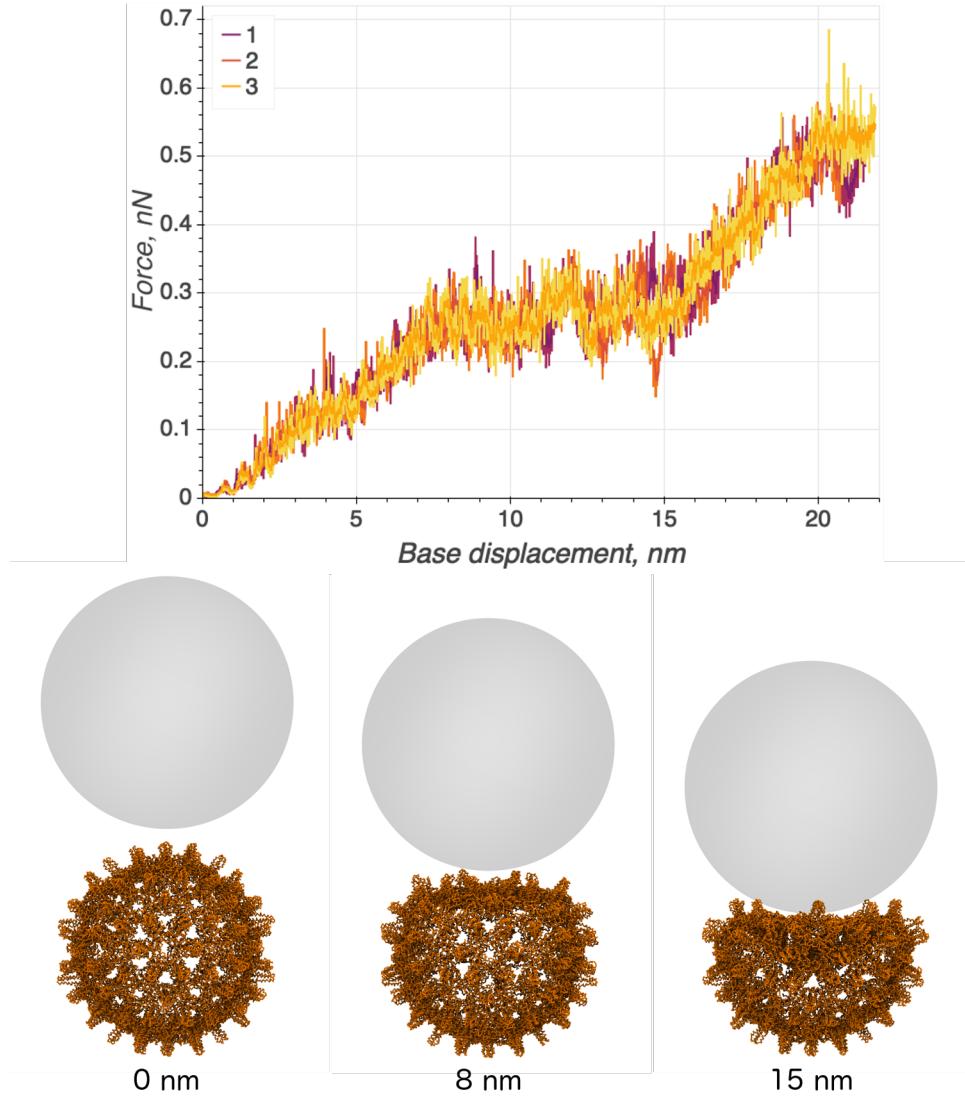


Рис. 24: Графики зависимости динамических сил при разных траекториях для пустой капсида вместе с визуализацией процесса при разной глубине вдавливания. Для каждой траектории были зафиксированы: сила, действующая со стороны кончика кантилевера на исследуемую частицу (светлая кривая), и сила сопротивления со стороны частицы (более темная кривая).

4.2.2 Наноиндентирование вируса Гепатита В

Эксперименты по наноиндентированию *in silico* были проведены для многих вирусных капсид [9, 11]. Однако в природе особый интерес представляют вирусы с генетическим материалом внутри, а не только их оболочка. В связи с чем в этой работе была создана гибридная модель вируса (белковая оболочка с геномом внутри). Благодаря этой модели стало возможно измерить динамическую силу при помощи наноиндентирования *in silico* и исследовать вирусную частицу уже с нуклеиновой кислотой внутри.

Использовалась та же установка что и в Разделе 4.2.1, с такими же параметрами (скорость движения основания кантилевера, жесткость пружины между ней и индентатором и пр.) и аналогичным способом фиксации вируса строго под кончиком кантилеве-

ра. Единственным различием этих двух экспериментов является исследуемая частица: в этом Разделе мы работаем с полной моделью вируса Гепатита В. Благодаря такой модели мы можем оценить вклад ДНК в устойчивость вируса к внешним силам.

Результаты проведенных вычислений для разных положений вируса показаны на Рисунке 25. Очевиден огромный вклад ДНК в жесткость капсиды: с первых шагов кантилевера угол наклона кривой силы (F) больше чем для пустой капсиды, хотя сама по себе кривая на этом этапе так же может быть аппроксимирована прямой, но только с большим тангенсом угла наклона. Исходя из этого можно сделать вывод, что жесткость вируса составляет не только его капсида, но и его генетический материал. Этот режим держится до тех пор, пока сила деформации не достигнет критического значения $F^* = 0.85$ нН.

При достижении этого значения силы деформации (при перемещении основания кантилевера на 13 нм) капсида не выдерживает давления сверху и изнутри (поскольку молекула ДНК тоже взаимодействует с белками оболочки отталкивающим потенциалом), и происходит разрыв капсиды. Из-за этого разрыва ДНК высвобождается из капсиды (Рисунок 25). Этот динамический переход также отражается резким уменьшением силы вдавливания от максимального значения до меньших значений. После достижения минимума сила деформации снова начинает увеличиваться, поскольку кончик кантилевера начинает вдавливать нижнюю часть структуры раздавленного вируса к прикрепленным (гармонически) его же частям.

Помимо этого видно, что для пустой капсиды положение вируса не существенно для профиля силы, в то время как для полной графики силы деформации выглядят по разному для разной траектории. Это объясняется в первую очередь положением ДНК: поскольку ДНК внутри может принимать любые положения, возможна ситуация, когда конформация ДНК на момент взаимодействия индентора с вирусом была не оптимальная, и, следовательно разрушение вируса произошло быстрее (траектории 2 и 3 на Рисунке 25).

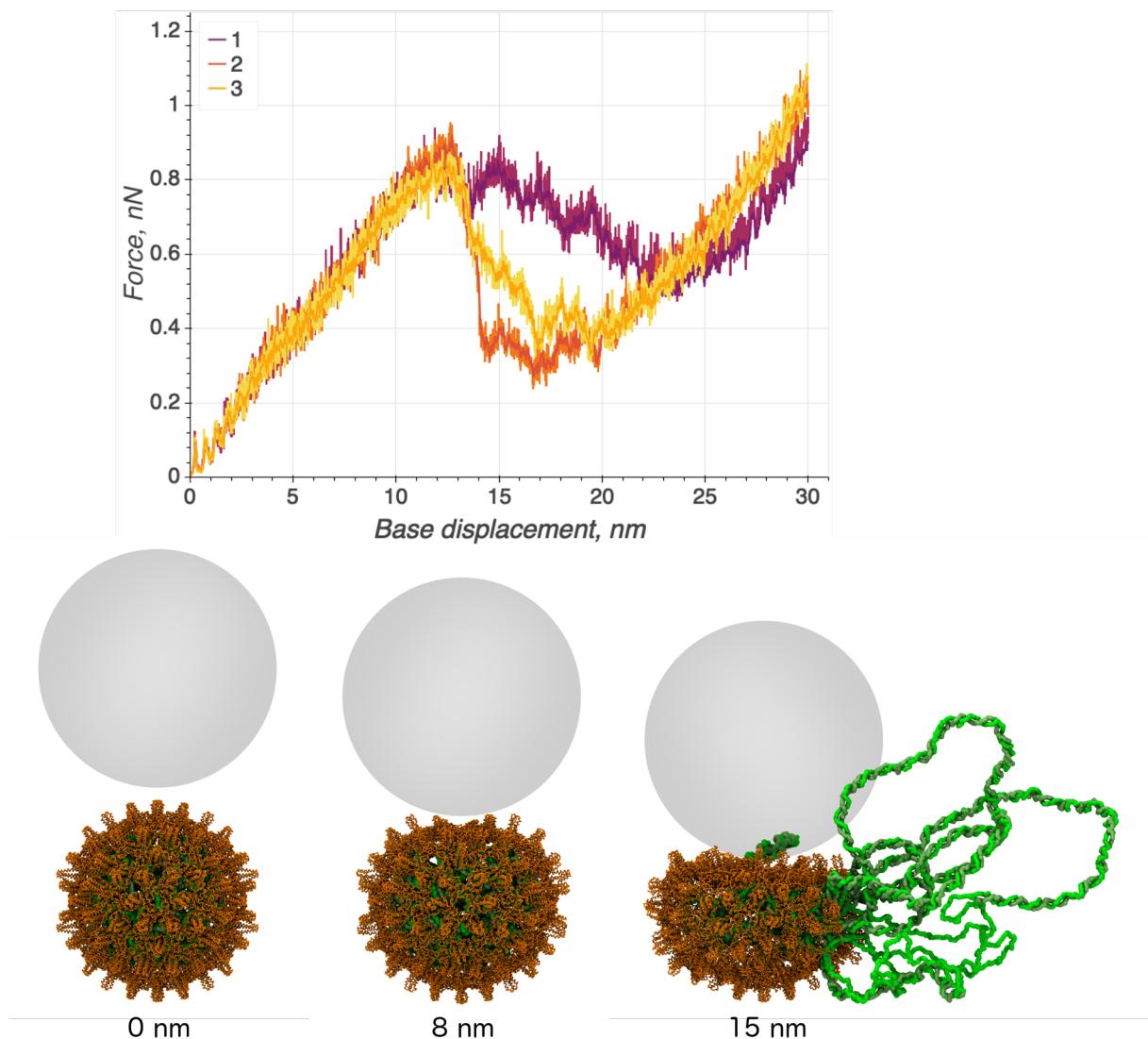


Рис. 25: Графики зависимости динамических сил при разных траекториях для целого вируса Гепатита В, а также визуализация процесса на разных этапах. Более светлая кривая отображает силу, действующую со стороны кончика кантилевера на исследуемую частицу, а темная — силу сопротивления со стороны вирусной частицы.

5 Заключение

Эта работа состоит из двух основных частей: (1) упаковка генетического материала в сферическую оболочку и исследование свойств системы во время упаковки, а также упакованной структуры ДНК; (2) объединение модели для ДНК с моделью, описывающей оболочку вируса, и проведение вычислительного эксперимента по наноиндентированию пустой оболочки и целого вируса.

Для описания свойств генома была использована модель двухцепочечной ДНК MRG-CG (Раздел 3.1), а для капсиды — СОП метод (Раздел 4.1.1). В системе помимо генетического материала и капсиды присутствуют явные ионы, однако растворитель используется неявный. Для поддержания постоянной средней температуры в расчетной области были протестированы два термостата: Ланжевен и Нозе-Хувер, в результате сравнения был выбран последний, поскольку он позволяет быстрее обойти фазовое пространство.

На этапе упаковки ДНК в сферическую оболочку было обнаружено, что свойства системы почти не зависят от ионной силы раствора, но сильно зависят от длины упаковываемого генома. Помимо этого была посчитана персистентная длина ДНК, которая слабо отличается от экспериментально, особенно при больших концентрациях NaCl. Было обнаружено, что ДНК в сжатом виде принимает структуру клубочка с количеством слоев от 2 до 6 (в зависимости от количества базовых пар ДНК), расстояние между которыми соответствует таковому в реальной жизни (2.6–2.8 нм).

При объединении моделей ДНК и капсиды мы столкнулись с тем, что полностью двухцепочечная ДНК разрушает капсиду вируса Гепатита В изнутри. В первую очередь это связано с тем, что реальный геном Гепатита В имеет одноцепочечную часть и замкнут, также это могло возникнуть из плохой параметризации модели, описывающей капсид. Чтобы проверить эти гипотезы, была создана модель максимально соответствующая реальному геному Гепатита В. И после объединения в вирусную частицу капсиды с этой моделью, таких проблем уже не было. Таким образом было решено использовать именно такой вариант модели для дальнейшего исследования.

Модель пустой капсиды, а также объединенная модель были исследована методом наноиндентирования *in silico* (Раздел 4.2). В первую очередь было проведено сравнение сил деформации полученных в процессе наноиндентирования пустой капсиды с экспериментальными результатами для вируса HBV [53]. Сравнение показало, что значения силы деформации для нашей модели очень близки к реальным (0.5 нН

для модели против 0.6 нН *in vitro*). Следовательно, и модель, и метод ее исследования согласовываются с экспериментальными данными, а значит, их можно использовать и для исследований целого вируса. После проведения индентирования на пустой капсиде, этим способом при таких же параметрах была исследована целая вирусная частица, состоящая из капсиды и генетического материала внутри. По результатам эксперимента, можно сделать вывод, что ДНК вносит вклад в жесткость вируса, поскольку сила деформации полной капсиды заметно больше, чем для пустой (Рисунок 25). Более того была найдена критическая точка деформации вируса Гепатита В, при котором происходит полное разрушение капсиды и высвобождение ДНК из нее: $F^* = 0.9$ нН при смещении основания кантилевера на 13 нм.

Список литературы

- [1] M. Oldstone, “Prevention of type i diabetes in nonobese diabetic mice by virus infection,” *Science (New York, N.Y.)*, vol. 239, pp. 500–2, 02 1988.
- [2] R. Eager and J. Nemunaitis, “Clinical development directions in oncolytic viral therapy,” *Cancer gene therapy*, vol. 18, pp. 305–17, 03 2011.
- [3] H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, and P. E. Bourne, “The Protein Data Bank,” *Nucleic Acids Research*, vol. 28, pp. 235–242, 01 2000.
- [4] W. Roos, I. Ivanovska, A. Evilevitch, and G. Wuite, “Viral capsids: mechanical characteristics, genome packaging and delivery mechanisms,” *Cell. Mol. Life Sci.*, vol. 64, no. 12, pp. 1484–1497, 2007.
- [5] S. Sun, V. Rao, and M. Rossmann, “Genome packaging in viruses,” *Curr. Opin. Struct. Biol.*, vol. 20, no. 1, pp. 114–120, 2010.
- [6] J. Speir and J. Johnson, “Nucleic acid packaging in viruses,” *Curr. Opin. Struct. Biol.*, vol. 22, no. 1, pp. 65–71, 2012.
- [7] W. Jiang, J. Jakana, P. Weigele, J. King, and W. Chiu, “Structure of epsilon15 bacteriophage reveals genome organization and dna packaging/injection apparatus,” *Nature*, vol. 439, no. 7076, pp. 612–616, 2006.
- [8] W. Roos, R. Bruinsma, and G. Wuite, “Physical virology,” *Nat. Phys.*, vol. 6, no. 10, pp. 733–743, 2010.
- [9] O. Kononova, J. Snijder, M. Brasch, J. Cornelissen, R. Dima, K. Marx, G. Wuite, W. Roos, and V. Barsegov, “Structural transitions and energy landscape for cowpea chlorotic mottle virus capsid mechanics from nanomanipulation *in vitro* and *in silico*,” *Biophys. J.*, vol. 105, no. 8, pp. 1893–1903, 2013.
- [10] W. Roos, M. Gibbons, A. Arkhipov, C. Utrecht, N. Watts, P. Wingfield, A. Steven, A. Heck, K. Schulten, W. Klug, and G. Wuite, “Squeezing protein shells: how continuum elastic models, molecular dynamics simulations, and experiments coalesce at the nanoscale,” *Biophys. J.*, vol. 99, no. 4, pp. 1175–1181, 2010.
- [11] A. Zhmurov, K. Rybnikov, Y. Kholodov, and V. Barsegov, “Generation of random numbers on graphics processors: Forced indentation *in silico* of the bacteriophage HK97,” *J. Phys. Chem. B*, vol. 115, no. 18, pp. 5278–5288, 2011.
- [12] O. Kononova, Y. Kholodov, K. E. Theisen, K. A. Marx, R. I. Dima, F. I. Ataullakhanov, E. L. Grishchuk, and V. Barsegov, “Tubulin bond energies and microtubule biomechanics determined from nanoindentation *in silico*,” *J. Am. Chem. Soc.*, vol. 136, no. 49, pp. 17036–17045, 2014.
- [13] O. Kononova, J. Snijder, Y. Kholodov, K. Marx, G. Wuite, W. Roos, and V. Barsegov, “Fluctuating nonlinear spring model of mechanical deformation of biological particles,” *PLoS Comput. Biol.*, vol. 12, no. 1, p. e1004729, 2016.
- [14] O. Kononova, K. A. Marx, and V. Barsegov, “Nanoindentation *in silico* of biological particles,” in *Applied Nanoindentation in Advanced Materials* (A. Tiwari, ed.), John Wiley and Sons, 2017.
- [15] O. Kononova, A. Zhmurov, K. A. Marx, and V. Barsegov, “Mechanics of viruses,” in *Series of Computational Biophysics: Coarse-grained Modeling of Biomolecules* (N. Dokholyan and G. Papoia, eds.), Taylor and Francis Publishers, 2017.
- [16] Y. Kuznetsov and A. McPherson, “Atomic force microscopy in imaging of viruses and virus-infected

cells," *Microbiology and molecular biology reviews : MMBR*, vol. 75, pp. 268–85, 06 2011.

- [17] A. Davtyan, N. Schafer, W. Zheng, C. Clementi, P. Wolynes, and G. Papoian, "AWSEM-MD: protein structure prediction using coarse-grained physical potentials and bioinformatically based local structure biasing," *J. Phys. Chem. B*, vol. 116, no. 29, pp. 8494–8503, 2012.
- [18] D. Hinckley and J. de Pablo, "Coarse-grained ions for nucleic acid modeling," *J. Chem. Theory Comput.*, vol. 11, no. 11, pp. 5436–5446, 2015.
- [19] S. Carmichael and M. Shell, "A new multiscale algorithm and its application to coarse-grained peptide models for self-assembly," *J. Phys. Chem. B*, vol. 116, no. 29, pp. 8383–8393, 2012.
- [20] W. Zheng, N. Schafer, A. Davtyan, G. Papoian, and P. Wolynes, "Predictive energy landscapes for protein–protein association," *Proc. Natl. Acad. Sci. USA*, vol. 109, no. 47, pp. 19244–19249, 2012.
- [21] H. Wu and G. Papoian, "Structural and dynamic analysis on disordered h4 histone tail by modified AWSEM-MD," *Biophys. J.*, vol. 108, no. 2, p. 388a, 2015.
- [22] A. Savelyev, C. K. Materese, and G. A. Papoian, "Is DNA's rigidity dominated by electrostatic or nonelectrostatic interactions?," *JACS*, vol. 133, no. 48, pp. 19290–19293, 2011.
- [23] C. Hulo, E. de Castro, P. Masson, L. Bougueret, A. Bairoch, I. Xenarios, and P. Le Mercier, "ViralZone: a knowledge resource to understand virus diversity," *Nucleic Acids Research*, vol. 39, pp. D576–D582, 10 2010.
- [24] D. Frenkel and B. Smit, eds., *Understanding Molecular Simulation: From Algorithms to Applications*. Orlando, FL, USA: Academic Press, Inc., 1st ed., 1996.
- [25] T. E. Cheatham III and D. A. Case, "Twenty-five years of nucleic acid simulations," *Biopolymers*, vol. 99, no. 12, pp. 969–977, 2013.
- [26] A. Pérez, F. J. Luque, and M. Orozco, "Frontiers in molecular dynamics simulations of DNA," *Accounts of Chemical Research*, vol. 45, no. 2, pp. 196–205, 2012.
- [27] P. D. Dans, W. J., G. H., and O. M., "Multiscale simulation of DNA," *Current Opinion in Structural Biology*, vol. 37, pp. 29–45, 2016.
- [28] D. A. Potoyan, A. Savelyev, and G. A. Papoian, "Recent successes in coarse-grained modeling of DNA," *Wiley Interdisciplinary Reviews: Computational Molecular Science*, vol. 3, no. 1, pp. 69–83, 2013.
- [29] A. Savelyev and G. A. Papoian, "Molecular renormalization group coarse-graining of polymer chains: Application to double-stranded DNA," *Biophys. J.*, vol. 96, no. 10, pp. 4044–4052, 2009.
- [30] A. Savelyev and G. A. Papoian, "Chemically accurate coarse graining of double-stranded DNA," *Proc. Natl. Acad. Sci. USA*, vol. 107, no. 47, pp. 20340–20345, 2010.
- [31] S. Nosé, "A molecular dynamics method for simulations in the canonical ensemble," *Molecular Physics*, vol. 52, no. 2, pp. 255–268, 1984.
- [32] W. G. Hoover, "Canonical dynamics: Equilibrium phase-space distributions.," *Phys. Rev. A. Gen. Phys.*, vol. 31, no. 3, pp. 1695–1697, 1985.
- [33] K. Zare and V. Szebehely, "Time transformations in the extended phase-space," *Celestial mechanics*, vol. 11, pp. 469–482, Dec 1975.
- [34] A. Evilevitch, L. Lavelle, C. M Knobler, E. Raspaud, and W. Gelbart, "Osmotic pressure inhibition of dna ejection from phage," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 100, pp. 9292–5, 09 2003.

- [35] K. Hanhijärvi, G. Ziedaite, M. Pietilä, E. Hæggström, and D. Bamford, “Dna ejection from an archaeal virus—a single-molecule approach,” *Biophysical journal*, vol. 104, pp. 2264–2272, 05 2013.
- [36] J. M. H. Alaboosh, *Synthesis of Nucleic Acid-Binding Ligands*. PhD thesis, Cardiff University, 7 2017.
- [37] P. J. Hagerman, “Investigation of the flexibility of dna using transient electric birefringence,” *Biopolymers*, vol. 20, pp. 1503 – 1535, 07 1981.
- [38] V. Rizzo and J. Schellman, “Flow dichroism of t7 dna as a function of salt concentration,” *Biopolymers*, vol. 20, pp. 2143–63, 10 1981.
- [39] S. Nair and A. Zlotnick, “Asymmetric modification of hepatitis b virus (hbv) genomes by an endogenous cytidine deaminase inside hbv cores informs a model of reverse transcription,” *Journal of Virology*, vol. 92, 05 2018.
- [40] P. de Pablo, I. Schaap, F. MacKintosh, and C. Schmidt, “Deformation and collapse of microtubules on the nanometer scale,” *Phys. Rev. Lett.*, vol. 91, no. 9, pp. 098101–098104, 2003.
- [41] A. Zhmurov, R. I. Dima, Y. Kholodov, and V. Barsegov, “SOP-GPU: Accelerating biomolecular simulations in the centisecond timescale using graphics processors,” *Proteins*, vol. 78, no. 14, pp. 2984–2999, 2010.
- [42] C. Hyeon, R. Dima, and D. Thirumalai, “Pathways and kinetic barriers in mechanical unfolding and refolding of RNA and proteins,” *Structure*, vol. 14, no. 11, pp. 1633–1645, 2006.
- [43] M. Mickler, R. Dima, H. Dietz, C. Hyeon, D. Thirumalai, and M. Rief, “Revealing the bifurcation in the unfolding pathways of GFP by using single-molecule experiments and simulations,” *Proc. Natl. Acad. Sci. USA*, vol. 104, no. 51, pp. 20268–20273, 2007.
- [44] A. Zhmurov, R. Dima, and V. Barsegov, “Order statistics theory of unfolding of multimeric proteins,” *Biophys. J.*, vol. 99, no. 6, pp. 1959–1968, 2010.
- [45] O. Kononova, L. Jones, and V. Barsegov, “Order statistics inference for describing topological coupling and mechanical symmetry breaking in multidomain proteins,” *J. Chem. Phys.*, vol. 139, no. 12, p. 121913, 2013.
- [46] S. Wynne, R. Crowther, and A. Leslie, “The crystal structure of the human hepatitis b virus capsid,” *Molecular Cell*, vol. 3, no. 6, pp. 771–780, 1999.
- [47] E. Pettersen, T. Goddard, C. Huang, D. Greenblatt, E. Meng, and T. Ferrin, “Ucsf chimera — a visualization system for exploratory research and analysis,” *Journal of Computational Chemistry*, vol. 25, no. 13, pp. 1605–1612, 2004.
- [48] W. Kabsch, “A solution for the best rotation to relate two sets of vectors,” *Acta Crystallographica Section A*, vol. 32, pp. 922–923, 1976.
- [49] N. Kol, Y. Shi, M. Tsvitov, D. Barlam, R. Z Shneck, M. S Kay, and I. Rousso, “A stiffness switch in human immunodeficiency virus,” *Biophysical journal*, vol. 92, pp. 1777–83, 04 2007.
- [50] I. Liashkovich, W. Hafezi, J. E Kühn, H. Oberleithner, A. Kramer, and V. Shahin, “Exceptional mechanical and structural stability of hsv-1 unveiled with fluid atomic force microscopy,” *Journal of cell science*, vol. 121, pp. 2287–92, 08 2008.
- [51] A. Pérez-Berná, A. Ortega-Estebar, R. Menéndez-Conejero, D. Winkler, M. Menéndez, A. Steven, S. Flint, P. de Pablo, and C. San Martín, “The role of capsid maturation on adenovirus priming for sequential uncoating,” *J. Biol. Chem.*, vol. 287, no. 37, pp. 31582–31595, 2012.

- [52] W. Roos, K. Radtke, E. Kniesmeijer, H. Geertsema, B. Sodeik, and G. Wuite, “Scaffold expulsion and genome packaging trigger stabilization of herpes simplex virus capsids,” *Proc. Natl. Acad. Sci. USA*, vol. 106, no. 24, pp. 9673–9678, 2009.
- [53] C. Utrecht, C. Versluis, N. Watts, W. Roos, G. Wuite, P. Wingfield, A. Steven, and A. Heck, “High-resolution mass spectrometry of viral assemblies: molecular composition and stability of dimorphic hepatitis B virus capsids,” *Proc. Natl. Acad. Sci. USA*, vol. 105, no. 27, pp. 9216–9220, 2008.