

Министерство науки и высшего образования Российской Федерации
Федеральное государственное автономное образовательное
учреждение высшего образования
«КАЗАНСКИЙ (ПРИВОЛЖСКИЙ) ФЕДЕРАЛЬНЫЙ УНИВЕРСИТЕТ»

ИНСТИТУТ ФИЗИКИ


КАФЕДРА МЕДИЦИНСКОЙ ФИЗИКИ

Направление: 03.04.02 Физика
Профиль: Медицинская физика

МАГИСТЕРСКАЯ ДИССЕРТАЦИЯ
СТРУКТУРНЫЕ ИССЛЕДОВАНИЯ АМИЛОИДОГЕННОГО
ПЕПТИДА SEM1(68-107) МЕТОДАМИ СПЕКТРОСКОПИИ
ЯДЕРНОГО МАГНИТНОГО РЕЗОНАНСА И МОЛЕКУЛЯРНОГО
МОДЕЛИРОВАНИЯ

Студентка 2 курса группы 06-119

« 8 » июня 2023 г.

 (Осетрина Д.А.)

Научный руководитель

к.ф.-м.н., доцент


« 8 » июня 2023 г.

 (Блохин Д.С.)

Научный консультант

к.ф.-м.н., доцент

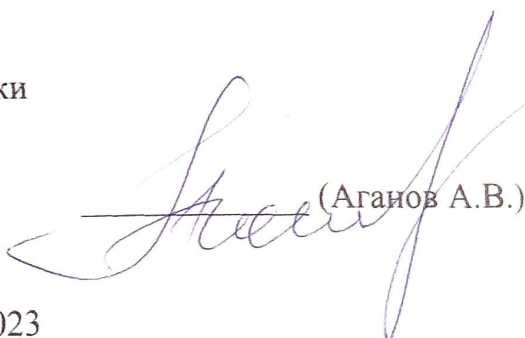
« 8 » июня 2023 г.

 (Юльметов А.Р.)

Заведующий кафедрой медицинской физики

д.ф.-м.н., профессор

« 8 » июня 2023 г.

 (Аганов А.В.)

Казань-2023

СОДЕРЖАНИЕ

ВВЕДЕНИЕ	4
1. ОБЗОР ЛИТЕРАТУРЫ	7
1.1. Роль амилоидных белков в живых организмах	7
1.2. Амилоидные пептиды семеногелина1	8
1.3. Двумерная ЯМР-спектроскопия	9
1.3.1. <i>TOCSY (Total Correlation Spectroscopy)</i>	10
1.3.2. <i>NOESY (Nuclear Overhauser Effect Spectroscopy)</i>	12
1.3.3. <i>HSQC (Heteronuclear Single Quantum Coherence)</i>	13
1.3.4. <i>¹H-¹³C-НМВС (Heteronuclear Multiple-Bond Correlation)</i>	15
1.4. 3D-ЯМР спектроскопия	16
1.5. Молекулярная динамика	19
1.5.1. <i>Физические основы метода молекулярной динамики</i>	19
1.5.2. <i>Равновесные условия</i>	21
1.5.3. <i>Периодические граничные условия</i>	23
1.5.4. <i>Численное интегрирование. Метод Верле</i>	24
1.5.5. <i>Метадинамика</i>	25
1.5.6. <i>Коллективные переменные</i>	27
1.5.7. <i>Вероятность состояний в метадинимике</i>	29
1.5.8. <i>Оценка погрешности свободной энергии</i>	30
2. Экспериментальная часть	32
2.1. Объект исследования	32
2.2. Методы исследования	32
3. Результаты и обсуждения	35
3.1. Отнесение химических сдвигов пептида SEM1(68-107)	35
3.2. Определение пространственной структуры пептида SEM1(68-85)	41
3.3. Расчет пространственной структуры пептида SEM1(68-107) и молекулярно-динамическое моделирование	46

3.4. Молекулярно-динамическое моделирование димера SEM1(68-107)	51
ВЫВОДЫ	58
СПИСОК ИСПОЛЬЗУЕМОЙ ЛИТЕРАТУРЫ	59

ВВЕДЕНИЕ

Впервые вирус иммунодефицита человека (ВИЧ) был выявлен в 1981 году [1] и является одним из самых опасных заболеваний. На сегодняшний день более 30 миллионов человек по всей Земле являются носителями ВИЧ. Считается, что для передачи вируса иммунодефицита человека существуют три основных пути: незащищенный половой контакт, при контакте кровь-кровь или от ВИЧ-положительной матери к плоду во время беременности. ВИЧ опасен тем, что он подавляет и разрушает иммунную систему заболевшего, увеличивая вероятность развития синдрома приобретенного иммунодефицита (СПИД). Организм больного перестает бороться с любыми видами инфекций, что приводит к смерти [2]. Поскольку одним из путей передачи вируса является незащищенный половой акт, то семенная жидкость рассматривается как основной фактор увеличения инфекционной активности ВИЧ [3].

Пептидные фрагменты белка семеногелина1 (SEM1) образуют белковые отложения – амилоидные фибриллы, которые находятся в семенной жидкости, и увеличивают инфекционную активность ВИЧ [4]. SEM1 экспрессируется в семенных пузырьках и после эякуляции быстро расщепляется внутренними протеазами на короткие пептидные фрагменты, одним из которых является объект исследования – пептид SEM1(68-107) [5]. Агрегирование SEM1(68-107) приводит к образованию амилоидных фибрилл, которые участвуют в процессе связывания вириона ВИЧ с поверхностью клетки-мишени. Чтобы установить физические основы взаимодействия между клетками-мишенями, амилоидными фибриллами и вирионами ВИЧ, необходимо понимать структурные изменения пептида, приводящие к агрегации. Поэтому в данной работе была исследована структура пептида SEM1(68-107) в мономерной форме с помощью спектроскопии ядерного магнитного резонанса (ЯМР) высокого разрешения. Также для SEM1(68-107) было выполнено

молекулярно-динамическое моделирование мономера и димера, чтобы установить механизмы взаимодействия нескольких пептидных цепей.

Спектроскопия ядерного магнитного резонанса (ЯМР) является одним из основных методов определения пространственной структуры белков в растворах в условиях, близких к физиологическим. С помощью метода ЯМР-спектроскопии возможно получение большого числа параметров, которые можно использовать при расчете пространственной структуры пептида, а также определить его динамические и кинетические характеристики [6]. К параметрам, необходимым для расчетов, относятся: химические сдвиги, интегральные интенсивности сигналов, константы спин-спинового взаимодействия, ядерный эффект Оверхаузера. Информация о химических сдвигах позволяет произвести отнесение сигналов ядер в спектрах, а также получить информацию о вторичной структуре белка. Интегральные интенсивности пропорциональны количеству ядер, дающих данные сигналы. Константы спин-спинового взаимодействия дают возможность получения информации о стереохимии и конформации молекул в растворе. Ядерный эффект Оверхаузера связан со взаимодействием двух ядер непосредственно через пространство. Он позволяет определить расстояние между протонами, отдаленными друг от друга на расстояние до $\leq 5 \text{ \AA}$ [7].

За последние десятилетия одним из наиболее мощных подходов в молекулярном моделировании стало развитие вычислительных методов изучения динамического поведения молекулярных систем, что позволяет изучить механизмы их функционирования. Молекулярная динамика (МД) часто используется для дополнения экспериментальных исследований и помощи в интерпретации полученных результатов. Метод МД обладает высоким пространственно-временным разрешением и позволяет изучать в атомарном масштабе структурные и физико-химические свойства биологических объектов, а также моделировать сложные динамические процессы, происходящие в них [8].

Целью работы является структурное исследование пептида SEM1(68-107), участвующего в увеличении инфекционную активность ВИЧ. Для достижения цели было необходимо выполнение следующих задач:

1. Отнесение химических сдвигов ^1H , ^{13}C , ^{15}N пептида SEM1(68-107).
2. Определения пространственной структуры SEM1(68-107) с использованием комбинированного подхода, основанного на использовании геометрических ограничений отдельных пептидных фрагментов: SEM1(68-85) и SEM1(86-107).
3. Моделирование методом молекулярной динамики димера пептида SEM1(68-107).

1. ОБЗОР ЛИТЕРАТУРЫ

1.1. Роль амилоидных белков в живых организмах

Сворачивание полипептидной цепи белка (фолдинг) является естественным процессом образования нативной пространственной структуры. Каждая молекула белка начинает формироваться как полипептид в виде линейной цепочки аминокислотных остатков [9]. Поскольку все остатки обладают определенными свойствами (гидрофобность, гидрофильность, электрический заряд), то при взаимодействии друг с другом и клеточным окружением получается хорошо определенная трехмерная структура. Однако, парадигма, что одна полипептидная цепь может принимать только одну уникальную конформацию, давно ушла в прошлое [10]. Многие белки способны образовывать различные структуры, изменяя свои свойства и функции, и данные конформационные изменения также могут приводить к их патологической агрегации.

Неправильный фолдинг белков в характерную амилоидную структуру ассоциируется с нейродегенеративными заболеваниями (болезнь Альцгеймера, болезнь Паркинсона) или прионными болезнями [11]. Но эксперименты *in vitro* показывают, что большое количество белков, не связанных с этими заболеваниями, могут образовывать амилоидные фибриллы. Это свидетельствует о том, что амилоидогенез является неотъемлемой частью биохимии полипептидов и широко используется в живых организмах для выполнения различных специфических целей. Например, амилоидные фибриллы использовались в ходе эволюции для различных адаптивных модальностей, таких как формирование долговременной памяти [12].

Амилоидные белки не имеют схожих аминокислотных последовательностей, они не являются родственными друг другу, их пространственные структуры могут сильно различаться, но все они

формируют фибриллы со сходной морфологией: большим количеством β -складчатых структур отдельных мономеров, расположенных перпендикулярно оси фибриллы и уложенных строго друг над другом. Процесс образования амилоидных фибрилл схож для всех белков и состоит из следующих этапов [13, 14]:

1. Так называемая фаза образования зародышевого ядра или лаг-фаза, в которой момеры претерпевают конформационные изменения и объединяются с образованием олигомерных ядер;
2. Фаза роста, в которой зародыши быстро растут за счет дальнейшего связывания мономеров и образуют крупные полимеры до достижения термодинамического равновесия в фазе насыщения;
3. И, наконец, образование фибрилл.

1.2. Амилоидные пептиды семеногелина1

Амилоидные фибриллы были также обнаружены в семенной жидкости в процессе поиска факторов, влияющих на увеличение скорости передачи ВИЧ половым путем [3]. Было показано, что одними из фибриллообразующих пептидов являются пептидные фрагменты белка семеногелина1 (SEM1) [5, 15].

Семеногелин1 секретируется в семенных пузырьках и составляет основной компонент сгустка спермы. После эякуляции он расщепляется химотрипсиноподобной сериновой протеазой (простатический специфический агент (PSA)) в высокоспецифичных местах, что важно для подвижности сперматозоидов и мужской фертильности [15, 16]. Семеногелин1 разделяется на пептидные фрагменты, четыре из которых способны образовывать амилоидные фибриллы: SEM1(86-107), SEM1(68-107), SEM1(49-107), SEM1(45-107) (рисунок 1). Фрагменты белка семеногелина1 выполняют множество биохимических функций. Например, они участвуют в разжижении сгустка спермы, выполняют защитные

(антибактериальные) функции и другие [17]. Также исследования показывают, что амилоиды спермы играют физиологическую роль не только в процессе оплодотворения, но и повышают инфекционную активность ВИЧ. Наиболее очевидная общность между этими двумя процессами заключается в том, что они происходят в одном и том же анатомическом контексте [18]. Слияние мембран сперматозоидов и ооцитов, также как мембран вирионов ВИЧ и клеток-мишеней, является энергетически невыгодным процессом, требующим совместных белок-белковых взаимодействий [19]. Амилоидные фибриллы SEM1 выступают в роли катионных мостиков, уменьшая электростатическое отталкивание между двумя отрицательно заряженными поверхностями как ооцитов и сперматозоидов, так и вирионов ВИЧ и клеток-мишеней [20].

Семеногелин1 (SEM1)	Аминокислотная последовательность	Молекулярная масса, kDa
SEM1(86-107)	DLNALHKTTKSQRHLGGSQQLL	2.45
SEM1(68-107)	TYHVDANDHDQSRKSQQYDLNALHKTTKSQRHLGGSQQLL	4.62
SEM1(49-107)	SGQKGKQQTESKGSFSIQYTYHVDANDHDQSRKSQQYDLNALHKTTKSQRHLGGSQQLL	6.69
SEM1(45-107)	GQHYSGQKGKQQTESKGSFSIQYTYHVDANDHDQSRKSQQYDLNALHKTTKSQRHLGGSQQLL	7.18

Рисунок 1 – Пептидные фрагменты белка семеногелин1 человека

Данная работа направлена на изучение пептида SEM1(68-107), так как для понимания процесса образования амилоидных фибрилл SEM1(68-107) необходимо понять процесс превращения нативной структуры этого пептида в фибриллы.

1.3. Двумерная ЯМР-спектроскопия

Двумерная спектроскопия ядерного магнитного резонанса – это один из видов ЯМР-спектроскопии, в котором фиксируются корреляции между двумя частотами. Коррелирующие партнеры находятся путем формирования

ортогональных проекций на две оси частот: F2 (прямое измерение) и F1 (косвенное измерение). Пиковые интенсивности кодируются в виде контуров. По какому механизму спины коррелируют в спектре, определяется типом эксперимента: основными механизмами являются скалярные связи (COSY, TOCSY), величина которых зависит от числа промежуточных связей (обычно 3 или меньше), а также от двугранного угла относительно центральной связи. Второй важный механизм — дипольные (прямые, через пространство) связи (NOESY).

Изучение структуры больших пептидных молекул невозможно с помощью одномерной ЯМР-спектроскопии, так как происходит наложение сигналов друг на друга, следовательно, их нельзя идентифицировать. Двумерная спектроскопия решает эту проблему для небольших белков (до 50 аминокислотных остатков). Поэтому этот метод широко используется для установления пространственной структуры пептидов [21].

1.3.1. TOCSY (Total Correlation Spectroscopy)

Импульсная последовательность 2D ^1H - ^1H TOCSY ЯМР-эксперимента представлена на рисунке 2 [22]. Она состоит из четырех этапов: подготовки, эволюции, смешивания и детектирования. Данные этапы характерны для любой импульсной последовательности двумерного эксперимента.

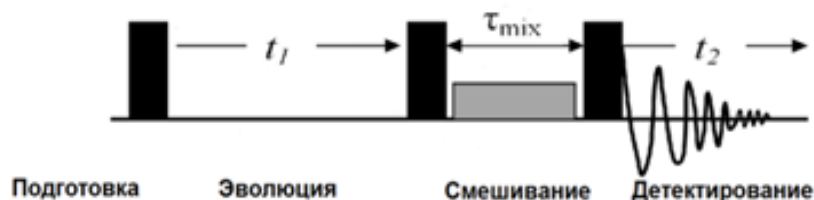


Рисунок 2 – Импульсная последовательность ЯМР-эксперимента 2D ^1H - ^1H TOCSY

На систему происходит воздействие 90° -импульсом после определенного фиксированного периода подготовки. Далее спины начинают свободно прецессировать в течение заданного времени t_1 . За это время происходит кодирование химического сдвига спина. В период изотопного смешивания (τ_{mix}) намагниченность с первого ядра переносится на второе. Во время детектирования происходит запись данных, полученных во время эксперимента.

Эксперимент TOCSY содержит все кросс-пики, связанные с протонами одной и той же спиновой системы (рисунок 3) [23]. Протоны разных аминокислот всегда принадлежат разным спиновым системам, потому что скалярная связь через амидную связь отсутствует. Анализ спиновых систем позволяет решить, к какому типу аминокислот относится спиновая система. Таким образом, для каждой аминокислоты получается характерный набор сигналов, по которому аминокислота может быть идентифицирована.

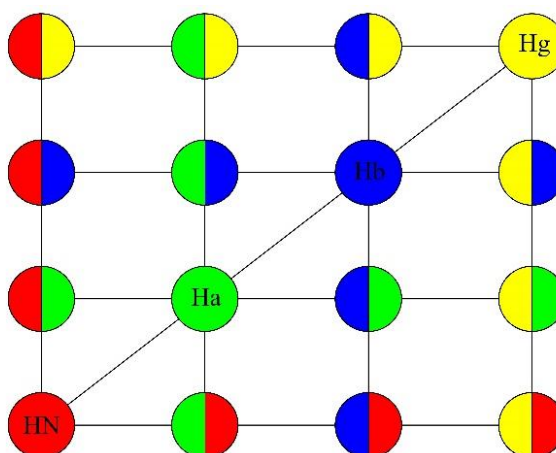


Рисунок 3 – На примере одного аминокислотного остатка показаны корреляции между протонами, наблюдаемые в эксперименте $2D\ ^1H-^1H$ TOCSY

1.3.2. NOESY (Nuclear Overhauser Effect Spectroscopy)

Спектр NOESY содержит всю информацию о пространственной близости протонов и, следовательно, кодирует трехмерное расположение атомов, структуру. Кросс-пики в NOESY обусловлены диполь-дипольными взаимодействиями, возникающими в результате взаимодействия спинов через пространство, и, следовательно, зависят только от расстояния, но не от количества промежуточных связей (рисунок 4) [24]. Пара из спектров TOCSY и NOESY используется для восстановления цепочки аминокислотных остатков в пептиде. Наложение этих спектров друг на друга дает возможность установить, какие сигналы принадлежат спиновой системе аминокислотного остатка (происходит наложение сигналов спектров TOCSY и NOESY), а какие являются корреляциями между протонами соседних аминокислотных остатков (наличие сигналов только в спектре NOESY).

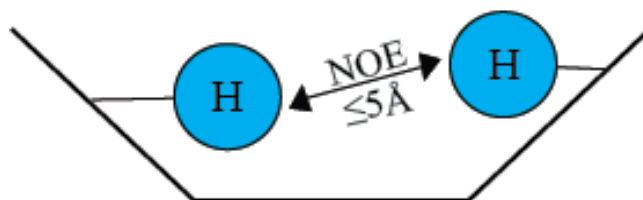


Рисунок 4 – Схематическое изображение взаимодействия протонов соседних аминокислотных остатков через пространство в ЯМР-эксперименте 2D ^1H - ^1H NOESY

Для импульсной последовательности 2D ^1H - ^1H NOESY эксперимента характерно наличие трех 90° импульсов (рисунок 5) [25]. Первый импульс создает поперечную спиновую намагниченность. В течение всего времени эволюции t_1 происходит прецессия спинов. Вторым импульсом создается продольная намагниченность. Она равна поперечной составляющей намагниченности, которая ортогональна направлению импульса. Во время смешивания τ_m происходит перенос намагниченности посредством кросс-

релаксации. Последний импульс создает поперечную намагниченность из оставшейся продольной после чего сразу же происходит детектирование сигналов.

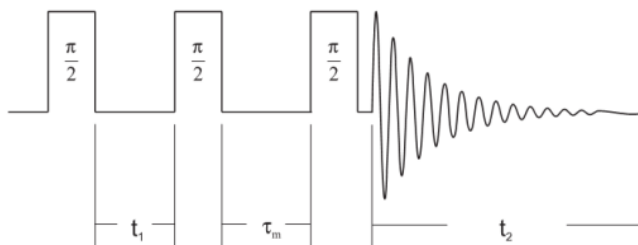


Рисунок 5 – Импульсная последовательность ЯМР-эксперимента 2D ^1H - ^1H NOESY

1.3.3. HSQC (Heteronuclear Single Quantum Coherence)

Эксперимент HSQC позволяет установить корреляции между ядрами двух разных типов, которые связаны через константу спин-спинового взаимодействия (КССВ) [24]. Полученный спектр является двумерным: по одной оси располагаются сигналы от протона ^1H , а по другой сигналы от ядер, отличных от водорода (обычно от ядер ^{13}C или ^{15}N). Спектр содержит пики, показывающие корреляции между гетероядром (^{13}C или ^{15}N), к которому присоединен протон, и ядром водорода (рисунок 6) [26].

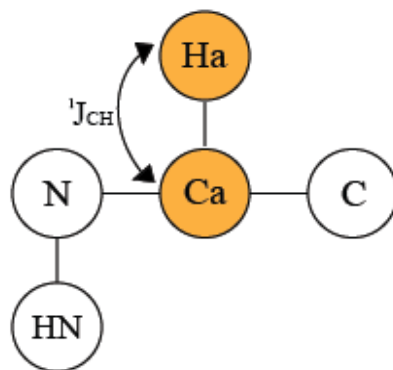


Рисунок 6 – Взаимодействие атома ^1H с атомом ^{13}C через одну химическую связь

Базовая схема этого эксперимента (рисунок 7) включает передачу намагниченности протона второму ядру, которое может быть ^{13}C , ^{15}N или ^{31}P , с помощью этапа INEPT (Insensitive Nuclei Enhancement by Polarization Transfer – нечувствительные ядра, усиленные передачей поляризации) [27]. После временной задержки (t_1) происходит передача намагниченности ядру ^1H с помощью этапа ретро-INEPT. Далее происходит запись сигнала. В HSQC записывается серия экспериментов, в которых увеличивается временная задержка t_1 .

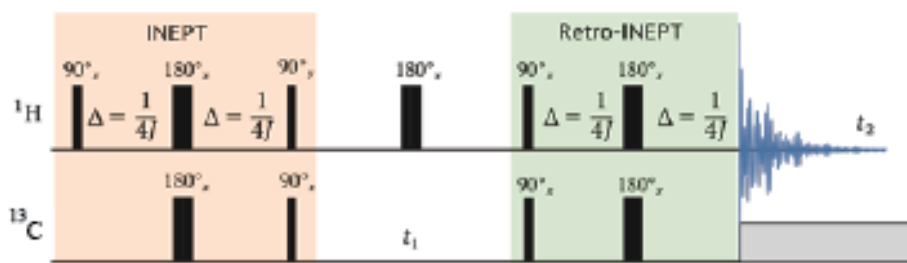


Рисунок 7 – Импульсная последовательность ЯМР-эксперимента 2D ^1H - ^{13}C HSQC

Эксперимент ^{15}N -HSQC является первым экспериментом для исследования белков методом ЯМР-спектроскопии. Данный спектр также называют спектром «отпечатков пальцев». Эксперимент HSQC может быть выполнен с использованием естественного содержания изотопа ^{15}N . Для этого необходима длительная регистрация спектров, более суток. Обычно используются меченные изотопами ^{15}N белки, что значительно сокращает время регистрации сигналов до нескольких часов. Все аминокислотные остатки белка, кроме пролина, имеют амидный протон, присоединенный к атому азота в пептидной связи. ^{15}N -HSQC показывает корреляции между протоном азота (HN) и азотом (N), что дает пик в спектрах ^{15}N -HSQC от каждого аминокислотного остатка, имеющего протон HN (рисунок 8) [28]. Если белок обладает вторичной

структурой, пики обычно хорошо диспергированы, и можно различить большинство отдельных пиков. Если вокруг середины спектра имеется большой кластер сильно перекрывающихся пиков, это указывает на присутствие в белке значительных неструктурированных элементов. В таких случаях, когда имеется сильное перекрытие резонансов, определение резонансов в спектрах может быть затруднено.

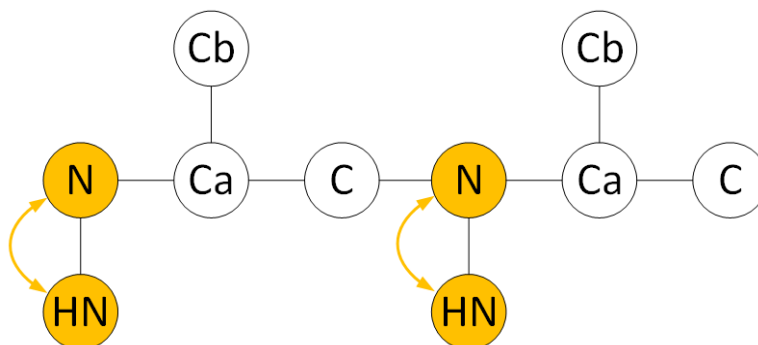


Рисунок 8 – Взаимодействие атома ^1H с атомом ^{15}N через одну химическую связь

1.3.4. ^1H - ^{13}C -HMBC (Heteronuclear Multiple-Bond Correlation)

^1H - ^{13}C -HMBC – двумерный гетероядерный эксперимент, в котором кросс-пики наблюдаются от атомов водорода, удаленных от углерода на 2-4 химических связи (рисунок 9) [23].

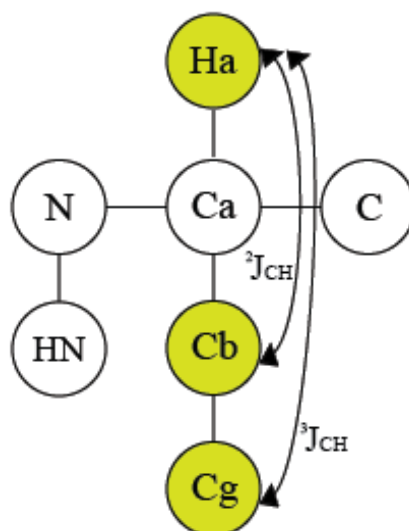


Рисунок 9 – Взаимодействие атома ^1H с атомами ^{13}C через две и три химических связи

Эксперимент НМВС предоставляет обширную структурную информацию, которая особенно полезна для выяснения структуры сложных соединений. Он играет ключевую роль, соединяя различные фрагменты большой молекулы посредством их углерод-протонного взаимодействия посредством двух, трех и более связей, и тем самым подтверждая структуру молекулы

1.4. 3D-ЯМР спектроскопия

3D-ЯМР спектроскопия используется для анализа крупных белков. Из названия данного метода видно, что сигнал наблюдается от трех разных ядер (^1H , ^{13}C , ^{15}N) [29]. Поэтому данная методика исследования подходит для белков, меченных изотопами углерода ^{13}C и азота ^{15}N .

Обычно для отнесения сигналов в 3D ЯМР-спектрах белков используются следующие эксперименты:

- ^1H - ^{13}C - ^{15}N HNCO
- ^1H - ^{13}C - ^{15}N HN(CA)CO

- $^1\text{H}-^{13}\text{C}-^{15}\text{N}$ HNCA
- $^1\text{H}-^{13}\text{C}-^{15}\text{N}$ HN(CO)CA
- $^1\text{H}-^{13}\text{C}-^{15}\text{N}$ HNCACB
- $^1\text{H}-^{13}\text{C}-^{15}\text{N}$ HN(CO)CACB
- $^1\text{H}-^{13}\text{C}-^{15}\text{N}$ CC(CO)NH
- $^1\text{H}-^1\text{H}-^{15}\text{N}$ H(CCO)NH
- $^1\text{H}-^1\text{H}-^{15}\text{N}$ 3D-TOCSY
- $^1\text{H}-^1\text{H}-^{15}\text{N}$ 3D-NOESY

На рисунке 10 показано отнесение сигналов белка на примере экспериментов $^1\text{H}-^{13}\text{C}-^{15}\text{N}$ HN(CO)CACB и $^1\text{H}-^{13}\text{C}-^{15}\text{N}$ HNCACB [30]. Идея состоит в том, что $^1\text{H}-^{13}\text{C}-^{15}\text{N}$ HNCACB связывает каждую группу NH с химическими сдвигами C_α и C_β ее собственного остатка (интенсивные сигналы) и предшествующего остатка (сигналы слабой интенсивности). HN(CO)CACB только коррелирует группу NH с предшествующими химическими сдвигами C_α и C_β [31]. По значениям химических сдвигов C_α и C_β можно идентифицировать тип аминокислотного остатка. Некоторые из них, такие как аланин, серин, треонин и глицин, очень легко обнаружить, поскольку их химические сдвиги C_β сильно отличаются от сдвигов других аминокислот (а в случае с глицином C_β отсутствует). Валин, изолейцин и пролин также, вероятно, выделяются тем, что они имеют более низкие, чем обычно, химические сдвиги C_α . После того, как цепочка групп NH с соответствующими им химическими сдвигами C_α и C_β построена, идентификация некоторых типов аминокислот позволяет сопоставить эту цепочку с аминокислотной последовательностью белка.

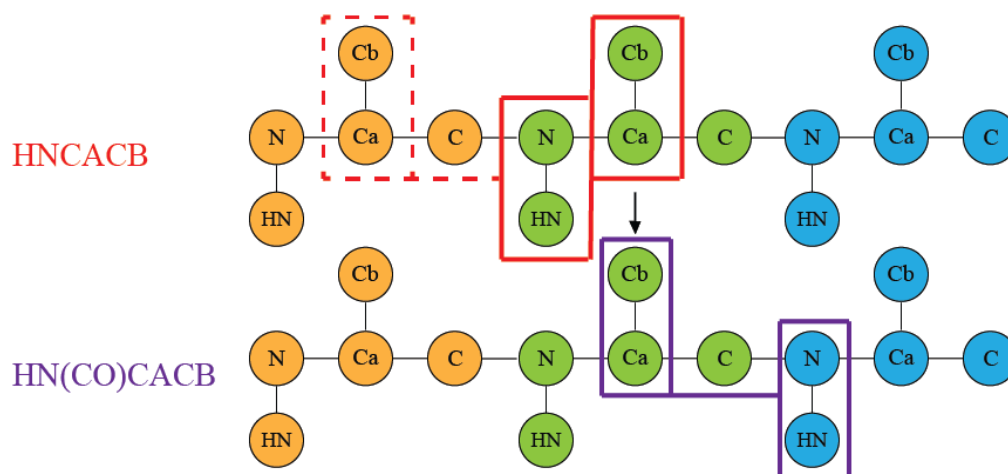


Рисунок 10 – Отнесение сигналов ядер ^1H и ^{13}C основной и боковых цепей белка на примере экспериментов ^1H - ^{13}C - ^{15}N HN(CO)CACB и ^1H - ^{13}C - ^{15}N HNCACB.

Основным преимуществом спектров тройного резонанса является их простота. Они содержат один или несколько сигналов на каждой частоте [32]. Таким образом, проблема спектрального перекрытия сигналов заметно уменьшается. Однако значения химических сдвигов четко разделенных сигналов от разных аминокислот могут случайно совпадать («вырождение» сигналов). Правильный выбор связей между аминокислотами является основной проблемой при отнесении сигналов в спектрах тройного резонанса. Еще одним преимуществом спектров тройного резонанса является их высокая чувствительность, обусловленная эффективным переносом намагниченности. Намагниченность передается через связи ^1J или ^2J (т.е. непосредственно через ковалентные химические связи). Поэтому времена переноса короче и потери на релаксацию меньше, чем в гомоядерных экспериментах.

Недостатком всех экспериментов с тройным резонансом является необходимость использования дважды меченых белков, получение которых часто обходится дорого. Также к недостаткам можно отнести большие временные затраты на проведение серии экспериментов. Ведь для отнесения химических сдвигов необходимо от 5-6 экспериментов, продолжительность каждого в среднем занимает 1 день.

1.5. Молекулярная динамика

1.5.1. Физические основы метода молекулярной динамики

Молекулярная динамика основана на модельном представлении многоатомной системы. В нем все атомы представляют собой материальные точки [33]. При таком упрощении становится возможным описать каждый атом с помощью классических уравнений движения Ньютона:

$$m_i \frac{d^2 \vec{r}_i(t)}{dt^2} = \vec{F}_i, \quad (1)$$

где i – номер атома ($1 \leq i \leq n$), n – полное число атомов в системе, m_i – масса атома, \vec{r}_i – радиус-вектор атома, \vec{F}_i – равнодействующая сил, действующих на атом.

Равнодействующая сил описывается в виде двух слагаемых:

$$\vec{F}_i = - \frac{\partial U(\vec{r}_1, \dots, \vec{r}_n)}{\partial \vec{r}_i} + \vec{F}_i^{ex}, \quad (2)$$

где U – потенциальная энергия системы. Взаимодействие между атомами является потенциальным, поэтому первая сила записана как градиент потенциальной энергии системы. Эта сила описывает действие всех атомов системы на данный атом. \vec{F}_i^{ex} – сила, определяемая взаимодействиями с молекулами окружающей среды [34].

Потенциальную энергию можно представить, как сумму вкладов различных типов взаимодействий между атомами:

$$U = U_b + U_v + U_\phi + U_f + U_{qq} + U_{VW} + U_{Hb}, \quad (3)$$

где U_b – потенциальная энергия валентных связей (4), U_v – валентных углов (5), U_ϕ – торсионных углов, U_f – плоских групп и псевдоторсионных углов (6), U_{qq} – кулоновских сил (7), U_{VW} – взаимодействий Ван-дер-Ваальса (8), U_{Hb} – водородных связей (9).

Для описания энергии валентных взаимодействий (4) и энергии колебания валентных углов (5) используется параболические потенциалы [34]:

$$U_b(r) = \frac{1}{2} \sum_{i=1}^{N_b} K_{b,i} (r_i - r_{o,i})^2, \quad (4)$$

где $K_{b,i}$ – эффективная жесткость валентной связи, i – номер связи в молекуле, N_b – полное число валентных связей, r_i – длина связи, $r_{o,i}$ – равновесная длина связи

$$U_v(\alpha) = \frac{1}{2} \sum_{i=1}^{N_v} K_{v,i} (\alpha_i - \alpha_{o,i})^2, \quad (5)$$

где $K_{v,i}$ – эффективная упругость валентного угла, i – номер связи в молекуле, N_v – полное число валентных углов, α_i – значение валентного угла, $\alpha_{o,i}$ – его равновесное значение.

Для описания потенциальной энергии торсионных углов, плоских групп и псевдоторсионных используется общее выражением (6), представляющее собой ряд Фурье [34].

$$U_\phi(\alpha) = \frac{1}{2} \sum_{\phi=1}^{\Phi} \sum_{l=1}^L K_{\phi,l} [1 + g_{\phi,l} \cos(n_{\phi,l} \alpha_\phi)], \quad (6)$$

где $K_{\phi,l}$ – константа, ϕ – номер торсионного угла, l – номер гармоники, $g_{\phi,l}$ – вклад гармоники в потенциал торсионного угла ($-1 < g_{\phi,l} < 1$), $n_{\phi,l}$ – кратность гармоники.

Потенциальная энергия взаимодействия заряженных атомов характеризуется электростатическим потенциалом:

$$U_{qq}(\vec{r}_i, \vec{r}_j) = \sum_{i \neq j} \frac{q_i q_j}{\epsilon r_{i,j}}, \quad (7)$$

где \vec{r}_i, \vec{r}_j – координаты взаимодействующих атомов, q_i, q_j – их парциальные заряды, ϵ – диэлектрическая проницаемость среды (для вакуума $\epsilon = 1$), $r_{i,j} = |\vec{r}_i - \vec{r}_j|$.

Потенциалы Леннард-Джонса (8) или водородной связи (9) используются для описания взаимодействий между атомами, которые не связаны валентной связью [34]:

$$U_{vw}(\vec{r}_i, \vec{r}_j) = \sum_{i \neq j} \left[\frac{A}{r_{ij}^{12}} - \frac{B}{r_{ij}^6} \right], \quad (8)$$

$$U_{Hb}(\vec{r}_i, \vec{r}_j) = \sum_{i \neq j} \left[\frac{A'}{r_{ij}^{12}} - \frac{B'}{r_{ij}^{10}} \right], \quad (9)$$

где A и B , A' и B' – константы, определяющие глубину потенциальной ямы и расположение ее минимума, \vec{r}_i, \vec{r}_j – координаты взаимодействующих атомов, $r_{ij} = |\vec{r}_i - \vec{r}_j|$.

В начальный момент времени t_0 задаются координаты $\vec{r}_0(t_0)$ и скорости $\vec{v}_0(t_0)$ всех частиц, после чего решаются уравнения движения, на каждом шаге вычисляются силы, новые координаты и скорости частиц с учетом граничных условий и внешнего воздействия на систему (если оно есть).

1.5.2. Равновесные условия

В условиях реального эксперимента исследуемые нами молекулы взаимодействуют со средой (молекулами растворителя). Также происходит энергообмен с окружающей средой, за счет чего происходит поддержание температуры. При расчетах невозможно учесть взаимодействия исследуемой молекулы с внешней средой. Поэтому используются специальные алгоритмы – термостаты, которые учитывают эффекты энергообмена с внешней средой.

Известно, что кинетическая энергия системы и ее температура связаны между собой и описываются выражением:

$$E = \frac{3kT}{2}, \quad (10)$$

где k – постоянная Больцмана, T – температура молекулярной системы.

Среднюю кинетическую энергию системы можно представить в виде:

$$E = \frac{\sum_{n=1}^N m_n v_n^2}{2N}, \quad (11)$$

где m – молекулярная масса атома, v – скорость атома, N – полное число частиц.

Из уравнений (10) и (11) получаем, что выражение для температуры в данный момент времени имеет вид:

$$T = \frac{\sum_{n=1}^N m_n v_n^2}{3Nk}, \quad (12)$$

Для того чтобы не возникало сбоев в моделировании из-за неравновесного состояния начальной конфигурации, необходимо произвести минимизацию энергии, а также привести систему в состояние равновесия. Уравновешивание системы проводится в два этапа:

1. Моделирование NVT-ансамбля. Это достаточно короткое МД-моделирование системы в ансамбле при постоянном количестве частиц N , объеме V и температуре T .
2. Моделирование NPT-ансамбля. Аналогичное моделирование при постоянном количестве частиц N , давлении P и температуре T .

Термостат Берендсена чаще всего используется для моделирования NVT-ансамбля. Принцип работы этого термостата основан на введении в уравнения движения знакопеременного нелинейного трения [35]. Флуктуации температуры ($T(t)$) от её равновесного значения (T_0) описываются уравнением Ландау-Теллера:

$$\frac{dT(t)}{dt} = \frac{T_0 - T(t)}{\tau}, \quad (13)$$

где τ – постоянная времени.

Из уравнения 13 видно, что отклонение в значении температуры экспоненциально убывает с характерным временем τ .

Термостат Берендсена подавляет изменения кинетической энергии. Для этого производится корректировка скоростей частиц системы путем умножения на фактор λ на каждом шаге моделирования:

$$\lambda = \sqrt{\mathbf{1} + \frac{\Delta t}{\tau} \left(\frac{T_0}{T(t-\frac{\Delta t}{2})} - \mathbf{1} \right)}, \quad (14)$$

При расчетах термодинамических свойств необходимо обеспечить постоянство давления. Для этого в NPT-ансамбле используется баростат Паринелло-Рамана [36, 37]. В нем векторы периодической ячейки имеют вид матрицы \mathbf{b} , которая подчиняется уравнениям движения:

$$\frac{d^2 \mathbf{b}}{dt^2} = V \mathbf{W} \mathbf{b}^{-1} (p - p_0), \quad (15)$$

где V – объем ячейки, p и p_0 – текущее и заданное значения давлений соответственно, \mathbf{W} – матрица-параметр, определяющая интенсивность баростатирования.

1.5.3. Периодические граничные условия

Поскольку исследуемыми объектами являются биомолекулы, то они обычно находятся в среде (обычно в жидкости) при заданных условиях (например, давлении или плотности). Поэтому при моделировании поведения системы все атомы необходимо поместить в одну ячейку. Но поскольку часть атомов будет находиться на границе, то это может привести к накоплению артефактов, связанных с граничными условиями. Поэтому вводятся периодические граничные условия. Это позволяет уменьшить влияние граничных эффектов на молекулярную систему. Границы ячейки становятся проницаемыми. Это позволяет атому возвращаться обратно в ячейку через противоположную границу, если он выходит за пределы ячейки через ближайшую границу (рисунок 11) [38].

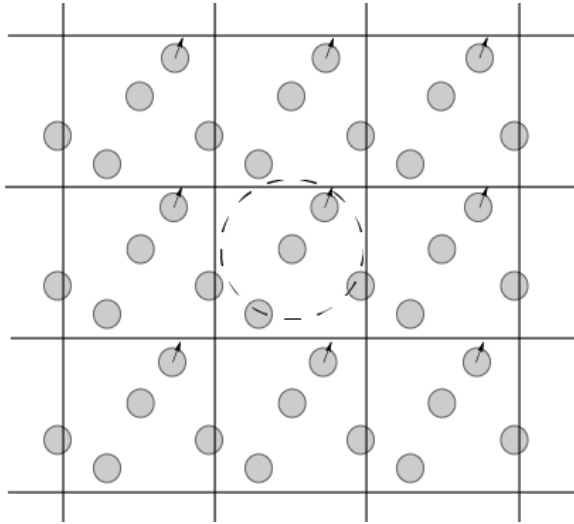


Рисунок 11 – Периодические граничные условия молекулярной системы.

Схематичное представление.

1.5.4. Численное интегрирование. Метод Верле

Решение системы классических уравнений движения является достаточно трудозатратным процессом. Поэтому существуют различные численные методы, одним из которых является метод Верле. Этот метод широко используется в молекулярной динамике.

Для определения сил, действующих на атом, необходимо взять производные от потенциальной энергии:

$$\vec{F}_i = -\nabla_i \sum_j u(\vec{r}_{ij}), \quad (16)$$

Далее необходимо рассчитать новые координаты атомов, из которых определяются равнодействующие сил:

$$\vec{r}_i(t + \Delta t) = \vec{r}_i(t) + \vec{v}_i(t)\Delta t + \frac{\vec{a}_i(t)}{2} \Delta t^2, \quad (17)$$

где $\vec{a}(t)$ – ускорение; $\vec{a}(t + \Delta t) = \frac{\vec{F}(t + \Delta t)}{m}$.

Определение скорости атомов производится по следующей формуле:

$$\vec{v}(t + \Delta t) = \vec{v}(t) + \left(\frac{\vec{a}(t) + \vec{a}(t + \Delta t)}{2} \right) \Delta t \quad (18)$$

Стабильность траектории зависит от выбора шага интегрирования: при большом шаге погрешности интегрирования могут значительно влиять на результат. Но уменьшение шага интегрирования существенно увеличивает время расчета. В методе Верле шаг интегрирования берется единым (порядка нескольких фемтосекунд).

1.5.5. Метадинамика

Метадинамика является одним из методов компьютерного моделирования, с помощью которого происходит ускорение редких процессов и событий исследуемой системы атомов (рисунок 12), а также реконструкция поверхности свободной энергии.

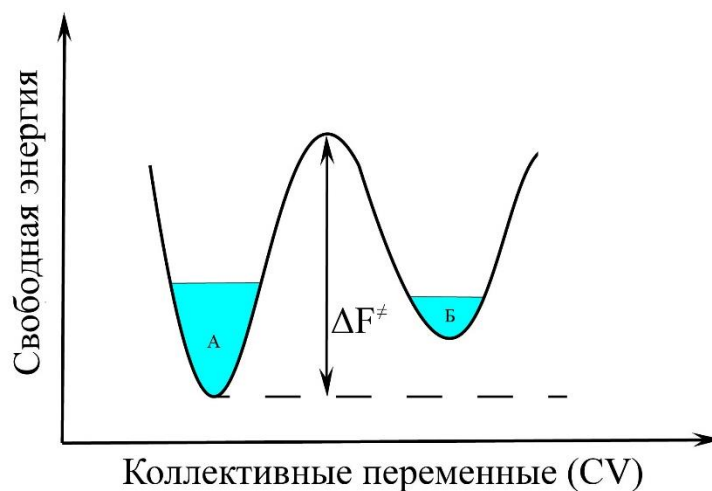


Рисунок 12 – Модель двухъямного потенциала. Стабильное состояние обозначено буквой «А», метастабильное состояние – буквой «Б». Голубым показаны области, в которых система колеблется из-за своей тепловой энергии. Область между состояниями «А» и «Б» вряд ли будет исследована, что делает маловероятным переход от «А» к «Б».

Принцип работы алгоритма заключается в том, чтобы изменить нормальную эволюцию системы с помощью добавления потенциальной энергии U_{bias} . Это действие позволяет системе преодолеть потенциальный барьер и переходить от одного минимума к другому. При этом система не может вернуться в состояние, в котором она уже находилась:

$$H = T + U + U_{bias}, \quad (19)$$

Данный потенциал U_{bias} является суммой гауссианов, которые адаптивно создаются в пространстве выбранных коллективных переменных и центрируются вдоль полученной траектории:

$$U_{bias}(\vec{s}, t) = \sum_{k\tau < t} W(k\tau) \exp\left(-\sum_{i=1}^d \frac{(s_i - s_i(k\tau))^2}{2\sigma_i^2}\right), \quad (20)$$

где \vec{s} – коллективные переменные, τ – шаг изменения, σ_i – ширина гауссиана для i – ой коллективной переменной, $W(k\tau)$ – высота гауссиана, d – количество коллективных переменных.

В метадинاميке сумма гауссианов используется для восстановления профиля свободной энергии. Для того чтобы потенциал сходил к значению свободной энергии, взятому с обратным знаком, для эволюции системы необходимы достаточно большие времена. Полученная функция является функцией от коллективных переменных:

$$U_{bias}(\vec{s}, t \rightarrow \infty) = -F(\vec{s}) + C, \quad (21)$$

где C – аддитивная константа, а свободную энергию $F(\vec{s})$ можно определить из уравнения (22):

$$F(\vec{s}) = -k_B T \int \delta(s_0 - s(\vec{r})) \exp\left(\frac{U(\vec{r})}{k_B T}\right) d\vec{r}, \quad (22)$$

где δ – функция Дирака, k_B – постоянная Больцмана, T – температура системы, $U(\vec{r})$ – функция потенциальной энергии.

В настоящее время наиболее распространенным алгоритмом для постепенного создания потенциала смещения является непрерывным: на

протяжении всей симуляции прибавляются гауссианы с заранее заданным одинаковым значением высоты. Это приводит к вынужденному движению всей системы в области с более высокими значениями потенциальной энергии, что приводит к сильным флуктуациям свободной энергии, которая вычисляется из потенциала.

Если экспоненциально уменьшать высоту гауссианов с увеличением числа шагов, то это ускорит сходимость вычислений [39]:

$$W(k\tau) = W_0 \exp\left(-\frac{U(\vec{r}, k\tau)}{k_B \Delta T}\right), \quad (23)$$

где W_0 – начальная высота гауссиана, k_B – постоянная Больцмана, ΔT – параметр размерности температуры.

Но при таком выборе условий потенциал лишь частично компенсирует поверхность свободной энергии:

$$U_{bias}(\vec{s}, t \rightarrow \infty) = -\frac{\Delta T}{T + \Delta T} F(\vec{s}) + C, \quad (24)$$

Из уравнения 24 следует, что при $\Delta T \rightarrow \infty$ восстанавливается стандартная метадинамика, $\Delta T \rightarrow 0$ – стандартная молекулярная динамика. Наличие параметра ΔT позволяет ограничить исследование пространства коллективных переменных только областями приемлемой свободной энергии. Действительно, зафиксировав ΔT в соответствии с высотой типичного барьера свободной энергии для рассматриваемой задачи, можно избежать преодоления барьеров, значительно превышающих этот.

1.5.6. Коллективные переменные

В сегодняшнем моделировании молекулярной динамики часто бывает полезно свести большое количество степеней свободы к нескольким параметрам, которые можно либо анализировать по отдельности, либо манипулировать ими, чтобы изменить динамику контролируемым образом. По

этой причине молекулярные системы часто анализируют с точки зрения коллективных переменных (CV), а не атомных координат. CV – это функция координат атомов, способная описать физику исследуемого процесса:

$$s(\vec{R}) = s(\vec{r}_1, \vec{r}_2, \dots, \vec{r}_N), \quad (25)$$

где \vec{r}_i – декартовы координаты i -го атома, N – количество атомов в системе.

Исторически сложилось, что в качестве коллективных переменных обычно используются расстояния между атомами, двугранные углы. Но поскольку моделирование применяется к сложным макромолекулам, то в качестве коллективных переменных также используются радиус гирации, число близких контактов, дипольный момент и т.д.

- Радиус гирации (радиус вращения) белка – это мера его компактности. Если белок сложен стабильно, то значения радиуса гирации будут слабо изменяться со временем. Для заданной группы атомов радиус гирации вычисляется по следующей формуле:

$$R_{gyr} = \sqrt{\left(\frac{\sum_{i=1}^N m_i |\vec{r}_i - \vec{r}_c|^2}{\sum_{i=1}^N m_i}\right)}, \quad (26)$$

где \vec{r}_c – положение центра масс:

$$\vec{r}_c = \frac{\sum_{i=1}^N m_i \vec{r}_i}{\sum_{i=1}^N m_i}, \quad (27)$$

- Число близких контактов между группами атомов G_1 и G_2 двух взаимодействующих молекул определяется по следующей формуле:

$$s(\vec{R}) = \sum_i \sum_j \xi_{ij}, \quad (28)$$

где суммирование производится по всем парам атомов групп G_1 и G_2 , а ξ_{ij} имеет следующий вид:

$$\xi_{ij} = \begin{cases} 1 & r_{ij} < 0 \\ \frac{1 - (r_{ij}/r_0)^n}{1 - (r_{ij}/r_0)^m} & r_{ij} \geq 0 \end{cases} \quad (29)$$

где $r_{ij} = |\vec{r}_i - \vec{r}_j| = d_0$. Параметры d_0, n, m подбираются в зависимости от задачи.

Одним из требований метадинамики является ограничение на количество коллективных переменных (обычно используют 2-3 коллективные переменные). Это связано с тем, что метадинамика ускоряет прохождение конформационного пространства молекулярной системой, не давая ей появляться в уже пройденных состояниях.

1.5.7. Вероятность состояний в метадинамике

Для определения свойств несмещенной системы необходимо устранить эффекты искажения, которые возникают при добавлении потенциала U_{bias} на каждом шаге моделирования траектории. Поэтому процедура взвешивания по вероятностям является необходимым условием при выполнении расчетов.

Координаты атомов молекулярной системы необходимы для нахождения значений коллективных переменных $\mathbf{s}(\vec{r})$. Если считать, что рассчитанные на каждом шаге симуляции МД при постоянном значении температуры ($T = const$) величины коллективных переменных являются элементами множества значений вероятностного распределения канонического ансамбля NVT, то вероятность находится как:

$$P(\mathbf{s}) = \frac{\int \delta(s(\vec{r}) - s_0) \exp\left(\frac{U_{bias}(s(\vec{r}))}{k_B T}\right) d\vec{r}}{\int \exp\left(\frac{U_{bias}(s(\vec{r}))}{k_B T}\right) d\vec{r}} \quad (30)$$

где δ – функция Дирака, k_B – постоянная Больцмана, T – температура системы.

Вычисление интегралов в выражении (30) является нетривиальной задачей и возможно только для простых систем. Поэтому для оценки

вероятности распределения коллективной переменной $P(\mathbf{s})$ необходимо рассчитать гистограмму, являющуюся выборкой полиномиального распределения. Полученное частное распределение $P(\mathbf{s})$ может быть использовано для нахождения значения свободной энергии $F(\mathbf{s})$ (с точностью до постоянной) из выражения 31. Также с помощью $P(\mathbf{s})$ можно реконструировать поверхность свободной энергии. Это необходимо для того, чтобы найти стабильные и метастабильные состояния системы и оценить величину энергетического барьера для переходов между этими состояниями.

$$F(\mathbf{s}) = -k_B T \ln P(\mathbf{s}) \quad (31)$$

1.5.8. Оценка погрешности свободной энергии

Для того чтобы оценить погрешности распределения вероятностей коллективных переменных, необходимо разделить траекторию, полученную при моделировании, на отрезки определенной длительности. Для того чтобы найти среднее значение и оценить погрешности распределения коллективных переменных, нужно для каждого отрезка траектории моделирования построить гистограммы распределения этих CV.

Для нахождения значения среднего по ансамблю необходимо вычислить математическое ожидание переменной, которое оценивается как среднее арифметическое по выборкам $\{X_i\}$ из распределения:

$$\langle X \rangle = \frac{1}{N} \sum_{i=1}^N X_i \quad (32)$$

Усреднение по отрезкам траектории моделирования также используется для вычисления средней свободной энергии и оценки ошибок. Для того чтобы получить поверхность энергии из отношения между гистограммой $P(\mathbf{s})$ и поверхностью свободной энергии $F(\mathbf{s})$, нужно усреднить гистограммы для каждого из ранее выбранных блоков. Величина свободной энергии для i – го столбца гистограммы находится из выражение (33):

$$F_i = -k_\beta T \ln \left(\frac{1}{N} \sum_{j=1}^N H_i^j \right) = -k_\beta T \ln \langle H_i \rangle \quad (33)$$

где H_i^j - величина i - го столбца для j - ой оценки гистограммы.

Погрешность полученной свободной энергии определяется как:

$$\delta F = \frac{k_\beta T}{\langle H_i \rangle} \sqrt{\frac{\langle H_i^2 \rangle - \langle H_i \rangle^2}{N}} \quad (34)$$

2. Экспериментальная часть

2.1. Объект исследования

Объектом исследования является пептид SEM1(68-107), состоящий из 40 аминокислотных остатков (рисунок 13), соответствующих 68-107 аминокислотным остаткам полноразмерного белка семеногелин1 человека. Экспрессию и очистку SEM1(68–107) проводили в соответствии с протоколом, описанным ранее в статье Бикмуллина и др. [40]. Для мечения изотопами ^{13}C и ^{15}N бактериальные клетки выращивали в среде M9, приготовленной с использованием $^{15}\text{NH}_4\text{Cl}$ и ^{13}C -D-глюкозы в качестве единственных источников азота и углерода.

Также в ходе работы было проведено исследование N-концевого домена SEM1(68-107) – пептида SEM1(68-85) (рисунок 13). Пептид был получен твердофазным синтезом по методике, описанной в статье М. И. Камалова [41].



Рисунок 13 – Аминокислотная последовательность пептида SEM1(68-107), а также его N-концевого домена SEM1(68-85) и C-концевого домена SEM1(86-107).

2.2. Методы исследования

1D, 2D, 3D ЯМР-спектры пептида SEM1(68-107) ($C = 0.3\text{мМ}$; 50мМ трисс-буфер + 300мМ NaCl) и 1D, 2D ЯМР-спектры пептида SEM1(68-85) ($C = 0.8\text{мМ}$; $\text{H}_2\text{O} + \text{D}_2\text{O} / 90\% + 10\%$) были получены на ЯМР-спектрометре 700 МГц (Bruker, AVANCE III 700), оснащенным четырехканальным

криодатчиком (^1H , ^{13}C , ^{15}N) CryoProbe при температурах 278K и 298K соответственно.

Для пептида SEM1(68-107) были получены следующие трехмерные спектры:

- ^1H - ^{13}C - ^{15}N HNCO
- ^1H - ^{13}C - ^{15}N HN(CA)CO
- ^1H - ^{13}C - ^{15}N HNCA
- ^1H - ^{13}C - ^{15}N HNCACB
- ^1H - ^{13}C - ^{15}N HN(CO)CACB
- ^1H - ^{13}C - ^{15}N CC(CO)NH
- ^1H - ^1H - ^{15}N HSQC-TOCSY

Также были получены 2D, 3D ЯМР-спектры NOESY с различными временами смешивания (80,100,150 и 250 мс).

Для пептида SEM1(68-85) были получены двумерные гомоядерные (2D ^1H - ^1H TOCSY, 2D ^1H - ^1H NOESY) и гетероядерные (2D ^1H - ^{15}N HSQC, 2D ^1H - ^{13}C HSQC и 2D ^1H - ^{13}C HMBSC) ЯМР-спектры.

Обработка данных производилась с помощью программы Bruker Topspin 3.6. Дальнейший анализ спектров производился с помощью программы CCP-NMR [42]. Для расчета пространственной структуры исследуемых образцов использовалась программа XPLOR-NIH [43]. Визуализация рассчитанных структур производилась в программе UCSF Chimera [44].

Также для исследуемого пептида SEM1(68-107) было выполнено моделирование мономера и димера методом молекулярной динамики в программе GROMACS [45]. Моделирование производилось с использованием силового поля Charmm36 [46]. Во всех случаях моделирование водного раствора производилось с использованием трехточечной модели воды TIP3P [47], которая имеет наилучшую производительность при расчете удельной теплоты [48]. Подготовка и уравнивание объектов проводились по следующему алгоритму:

1. Создавался pdb-файл с объектом исследования внутри ячейки, заполненной водой. Также случайным образом некоторые молекулы воды были заменены на ионы Na^+ и Cl^- для того, чтобы суммарный заряд системы сводился к нулю.

2. Во всех случаях ячейка с периодическими граничными условиями имела форму ромбододэкаэдра, составленного из 12 квадратов.

3. Для полученной системы была проведена минимизация энергии методом градиентного спуска. Данная операция необходима для того, чтобы исключить геометрические ошибки.

4. Далее система была уравновешена в NVT ансамбле в течение 1нс при помощи термостата Берендсена при постоянном значении температуры $T = 300\text{K}$.

5. Потом также производилось уравновешивание системы, но уже в NPT ансамбле в течение 1 нс с помощью алгоритма баростатирования Нозе-Гувера при постоянных значениях температуры $T = 300\text{K}$ и давления $P = 1$ бар.

Для димера было выполнено моделирование методом метадинамики в GROMACS с использованием пакета PLUMED [49, 50]. Визуализация рассчитанных структур также производилась в программе UCSF Chimera.

3. Результаты и обсуждения

3.1. Отнесение химических сдвигов пептида SEM1(68-107)

Для отнесения химических сдвигов пептида SEM1(68-107) была получена серия трехмерных ЯМР-спектров, которые также использовались для последовательного назначения спиновых систем аминокислотных остатков.

Первым стандартным экспериментом, проводимый при исследовании белков, является 2D ^1H - ^{15}N HSQC (рисунок 14). С его помощью происходило нахождение спиновых систем каждого аминокислотного остатка из аминокислотной последовательности исследуемого пептида SEM1(68-107). Далее производился поиск сигналов, связанных со спиновыми системами, в срезах трехмерных спектров на соответствующих значениях химических сдвигов ^1H и ^{15}N .

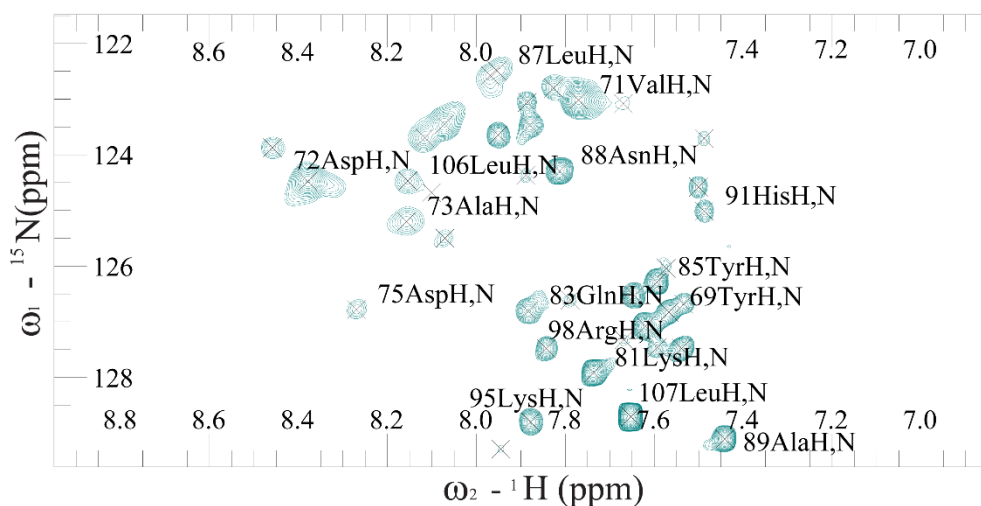


Рисунок 14 – Фрагмент 2D ^1H - ^{15}N HSQC ЯМР-спектра пептида SEM1(68-107), T = 278K

Рассмотрим отнесение сигналов на примере пары экспериментов ^1H - ^{13}C - ^{15}N HNCO и ^1H - ^{13}C - ^{15}N HN(CA)CO. Данная пара экспериментов используется для назначения основной цепи. Эксперимент ^1H - ^{13}C - ^{15}N HNCO дает

информацию о связи ядер ^1H и ^{15}N с карбоксильным углеродом предыдущего аминокислотного остатка (рисунок 15(а, в)). Эксперимент ^1H - ^{13}C - ^{15}N HN(CA)CO может давать два сигнала: карбоксильный углерод своего аминокислотного остатка и предыдущего, связанные с ядрами ^1H и ^{15}N (рисунок 15(б)). Далее по цепочке устанавливается соответствие связанных спиновых систем с аминокислотной последовательностью.

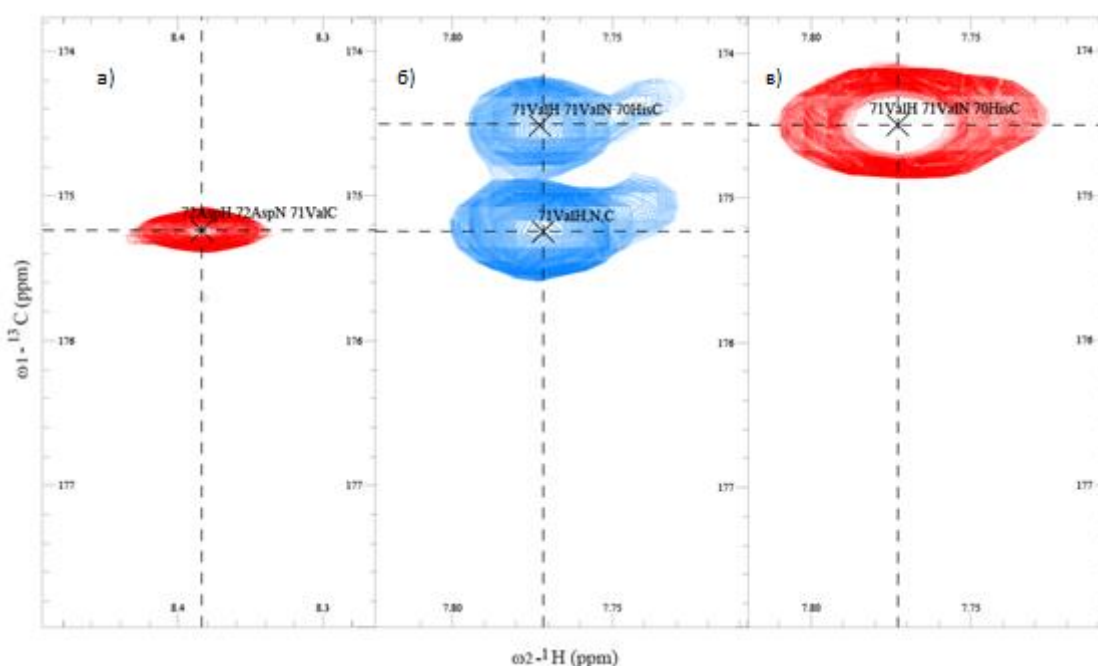


Рисунок 15 – Восстановление аминокислотной последовательности пептида SEM1(68-107) с использованием ЯМР-экспериментов ^1H - ^{13}C - ^{15}N HNCO (а, в) и ^1H - ^{13}C - ^{15}N HN(CA)CO (б) на примере аминокислотных остатков 71Val и 72 Asp, $T = 278\text{K}$

Также рассмотрим отнесение сигналов на примере пары экспериментов ^1H - ^{13}C - ^{15}N HNCACB и ^1H - ^{13}C - ^{15}N HN(CO)CACB. Данная пара экспериментов также используется для назначения основной цепи. Эксперимент ^1H - ^{13}C - ^{15}N HNCACB дает значения сигналов $\text{C}\alpha$ и $\text{C}\beta$ своего аминокислотного остатка и предыдущего, в результате чего на спектрах могут наблюдаться до четырех пиков (рисунок 16(а)). Эксперимент ^1H - ^{13}C - ^{15}N HN(CO)CACB дает значения

сигналов $C\alpha$ и $C\beta$ только предыдущего аминокислотного остатка (рисунок 16(б)).

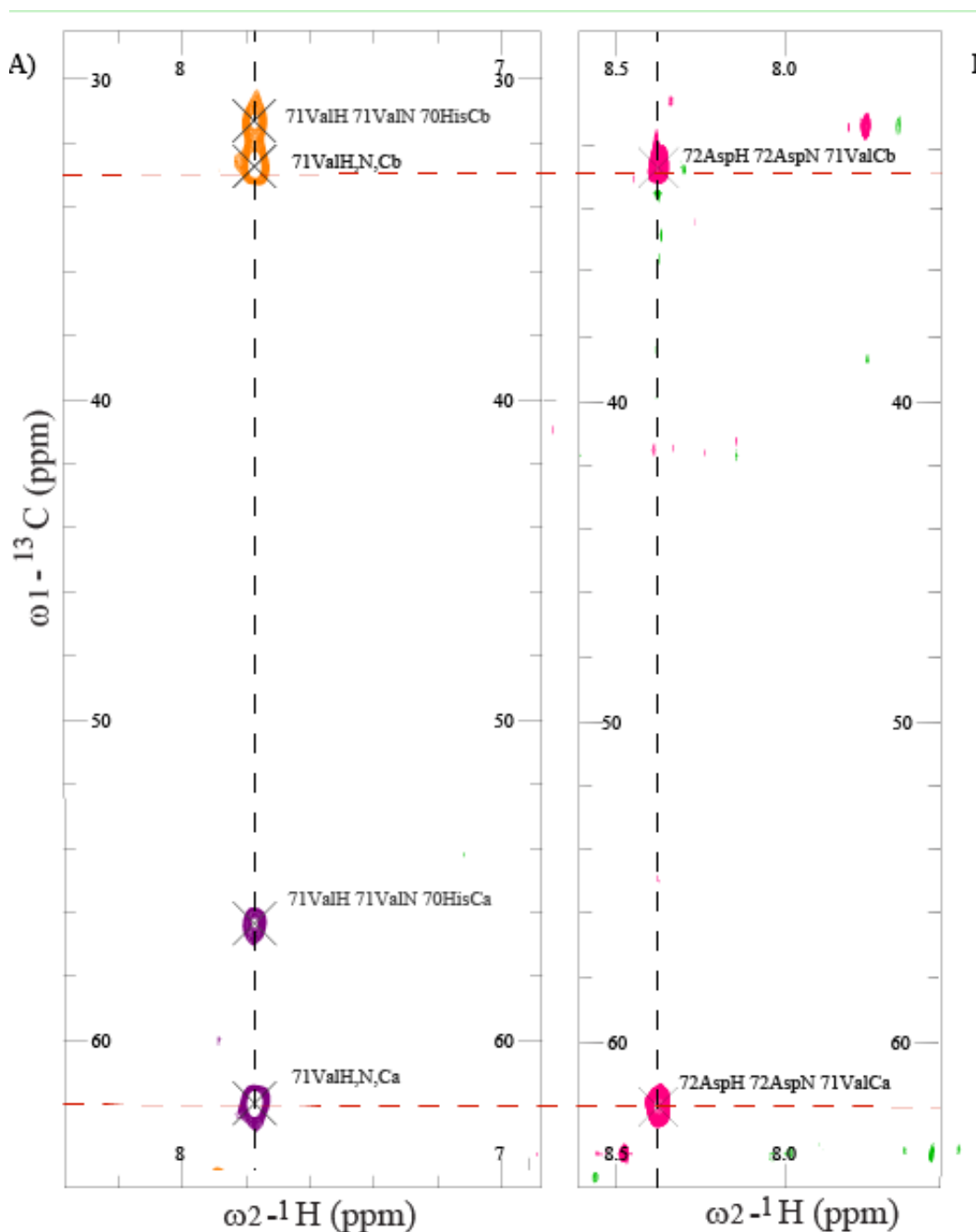


Рисунок 16 – Восстановление аминокислотной последовательности пептида SEM1(68-107) с использованием ЯМР-экспериментов ^1H - ^{13}C - ^{15}N HNCACB (а), ^1H - ^{13}C - ^{15}N HN(CO)CACB (б) на примере аминокислотных остатков 71Val и 72 Asp, $T = 278\text{K}$

Для получения отнесений ядер углерода боковых цепей был использован спектр ^1H - ^{13}C - ^{15}N $\text{CC}(\text{CO})\text{NH}$. Химический сдвиг проявляется одновременно на всех ядрах углерода боковой цепи, а также на ядрах ^{15}N и ^1H , что приводит к трехмерному спектру. На спектре видны все углероды боковой цепи предыдущего аминокислотного остатка (рисунок 17).

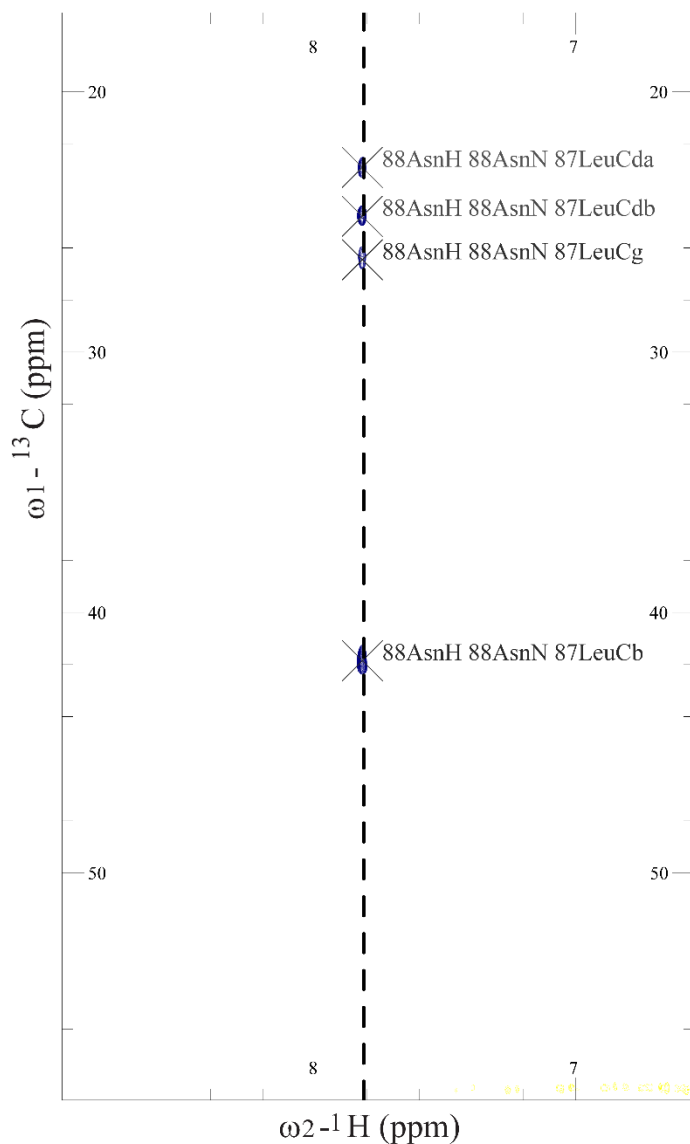


Рисунок 17 – Отнесение химических сдвигов углеродной боковой цепи на примере 87Leu с помощью трехмерного спектра ^1H - ^{13}C - ^{15}N $\text{CC}(\text{CO})\text{NH}$, $T = 278\text{K}$

Для получения значений химических сдвигов ^1H использовался трехмерный эксперимент ^1H - ^1H - ^{15}N TOCSY. Фрагмент данного спектра приведен на рисунке 18.

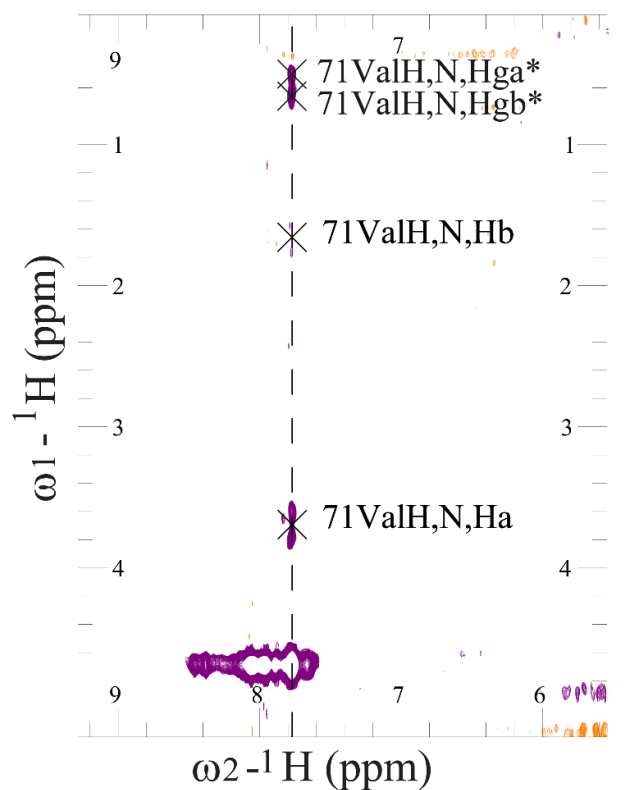


Рисунок 18 – Фрагмент 3D ЯМР-спектра ^1H - ^1H - ^{15}N TOCSY, $T = 278\text{K}$

На рисунке 19 можно видеть завершенное присвоение химических сдвигов для 34 из 40 аминокислотных остатков (85%) пептида SEM1(68-107). Значения химических сдвигов ^1H , ^{15}N и ^{13}C депонированы в BioMagResBank (<http://www.bmrb.wisc.edu>) под номером доступа BMRB-51406.

68	69	70	71	72	73	74	75	76	77
Thr	Tyr	His	Val	Asp	Ala	Asn	Asp	His	Asp
78	79	80	81	82	83	84	85	86	87
Gln	Ser	Arg	Lys	Ser	Gln	Gln	Tyr	Asp	Leu
88	89	90	91	92	93	94	95	96	97
Asn	Ala	Leu	His	Lys	Thr	Thr	Lys	Ser	Gln
98	99	100	101	102	103	104	105	106	107
Arg	His	Leu	Gly	Gly	Ser	Gln	Gln	Leu	Leu

Рисунок 19 – Аминокислотная последовательность SEM1(68-107).

Черный – назначенные фрагменты, белый – неназначенные фрагменты

Полное отнесение химических сдвигов сделать не удалось из-за неинформативности ЯМР-спектров для некоторых аминокислотных остатков. Это может быть объяснено высокой подвижностью белка в этих участках, что может приводить к ускоренным процессам релаксации отдельных аминокислот.

Для определения пространственной структуры пептида SEM1(68-107) необходимы экспериментальные ограничения для расчетов. Основным типом ограничения являются межъядерные расстояния, которые могут быть установлены из анализа NOESY ЯМР-спектров. Для SEM1(68-107) были зарегистрированы 2D и 3D NOESY спектры с различными временами смешивания (80,100,150,250мс). Спектры оказались неинформативными, что может быть также связано с неструктурированностью и высокой подвижностью фрагментов пептида. Отсутствие данных NOE не позволило определить пространственную структуру белка SEM1(68-107). Поэтому для решения данной задачи был применен подход, описанный в исследовании Шмитца и др. [51]. Данный подход основан на использовании экспериментальных ограничений отдельных пептидных фрагментов, содержащихся в полноразмерном объекте исследования, при расчете структуры белка.

Для получения пространственной структуры пептида SEM1(68-107) были выбраны два пептидных фрагмента: N-домен SEM1(68-85) и C-домен SEM1(86-107), который был ранее исследован в предыдущей работе [52].

3.2. Определение пространственной структуры пептида SEM1(68-85)

Для отнесения химических сдвигов ^1H пептида SEM1(68-85) в водном растворе были использованы 2D ^1H - ^1H TOCSY и 2D ^1H - ^1H NOESY ЯМР-спектры (рисунок 20 и 21). Химические сдвиги протонов, полученные для различных аминокислотных остатков, были проанализированы с помощью стандартных значений химических сдвигов, приведенных в литературных данных [53, 54]. Также были получены химические сдвиги ^{13}C , $^{13}\text{C}_\alpha$ и ^{15}N , для нахождения которых использовались двумерные гетероядерные ЯМР-спектры (2D ^1H - ^{15}N HSQC, 2D ^1H - ^{13}C HSQC и 2D ^1H - ^{13}C HMBC) (рисунки 22, 23).

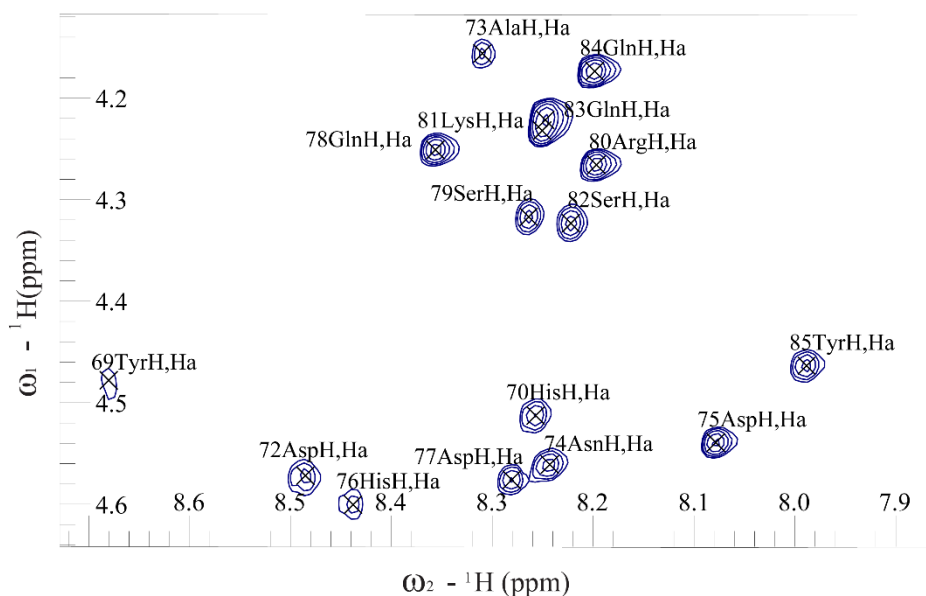


Рисунок 20 – Фрагмент 2D ^1H - ^1H TOCSY ЯМР-спектра пептида SEM1(68-85) в водном растворе, T = 298K

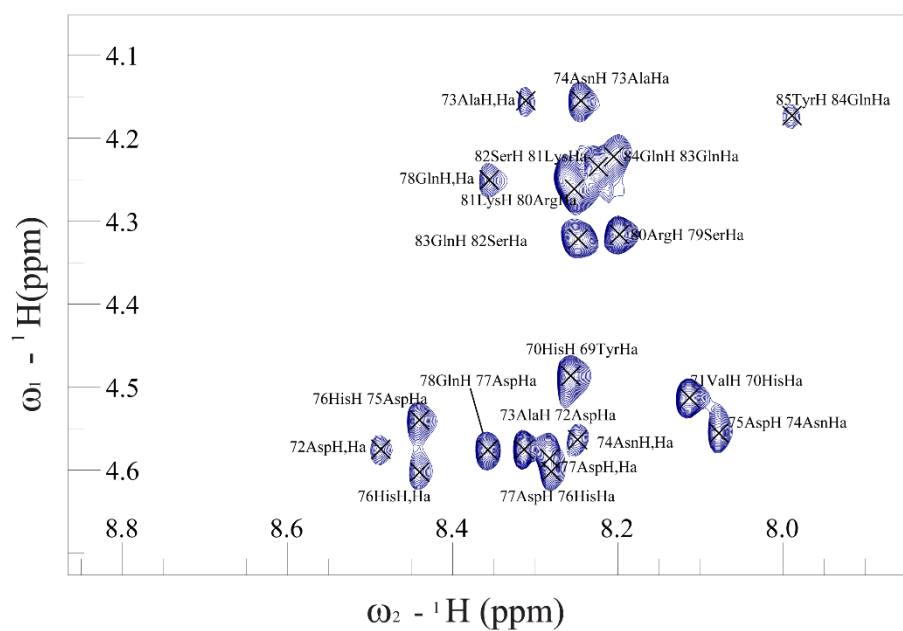


Рисунок 21 – Фрагмент 2D ^1H - ^1H NOESY ЯМР-спектра пептида SEM1(68-85) в водном растворе, T = 298K

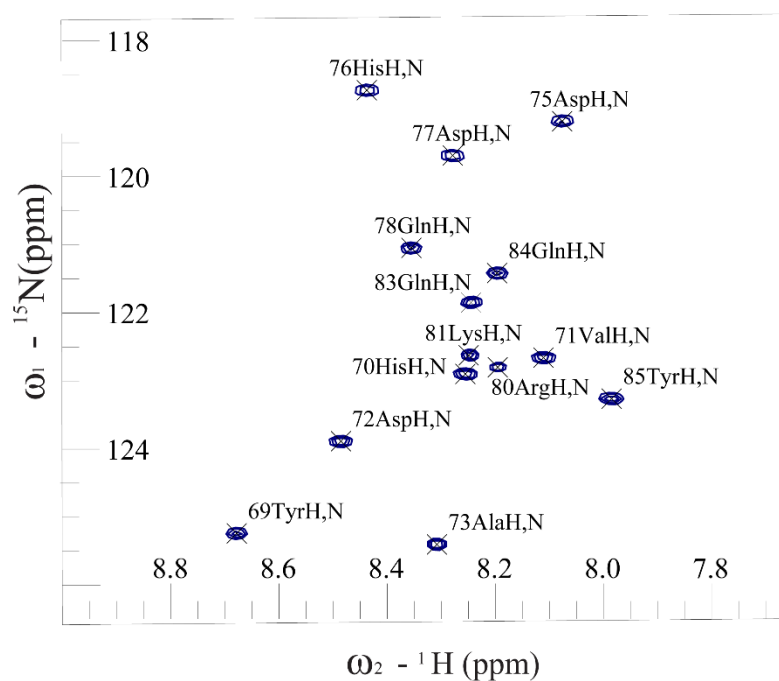


Рисунок 22 – Фрагмент 2D ^1H - ^{15}N HSQC ЯМР-спектра пептида SEM1(68-85) в водном растворе, T = 298K

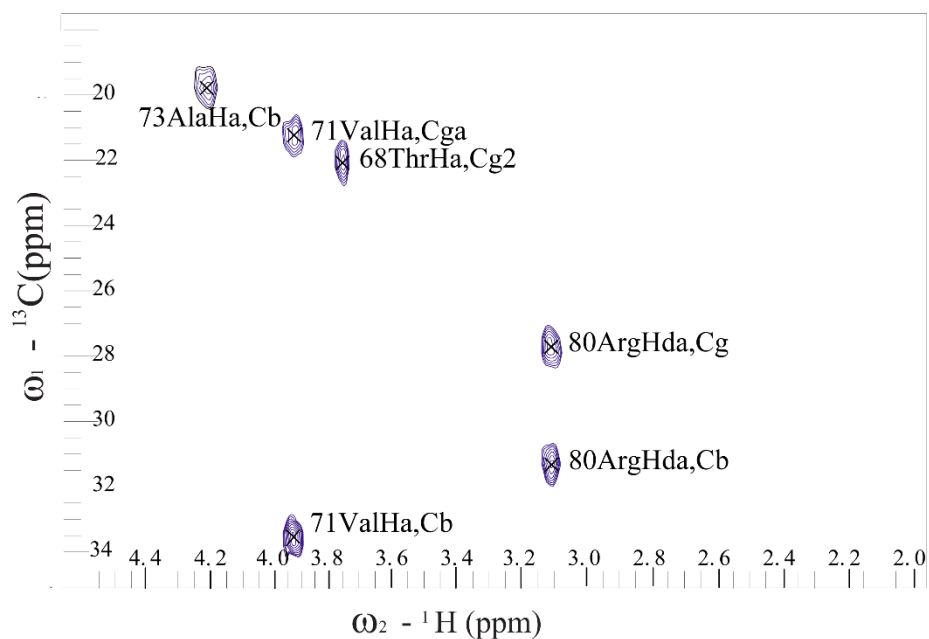


Рисунок 23 – Фрагмент 2D ^1H - ^{13}C HMBSC ЯМР-спектра пептида SEM1(68-85) в водном растворе, $T = 298\text{K}$

Полученные значения химических сдвигов ^1H , ^{13}C и ^{15}N были депонированы в международную базу данных BioMagResBank под регистрационным номером BMRB-51400. Также значения химических сдвигов $\text{H}\alpha$, N , $\text{C}\alpha$ и C использовались для определения двугранных углов с помощью сервиса TALOS+.

Межъядерные расстояния пептида SEM1(68-85) определялись с помощью спектра 2D ^1H - ^1H NOESY (рисунок 21). Всего было определено 132 межъядерных расстояния, из которых 78 расстояний внутри аминокислотных остатков и 54 расстояния между пространственно близкими аминокислотными остатками (таблица 1).

Таблица 1 – Статистическая информация для структурного ансамбля SEM1(68-85)

Межъядерные расстояния		Двугранные углы	
всего	132	ϕ	16
внутри аминокислотных остатков	78	ψ	16
между аминокислотными остатками	54		
последовательные ($i-j=1$)	50		
средние ($i-j=2, 3$)	3		
дальние ($i-j=12$)	1		

Полученные значения межъядерных расстояний и двугранных углов использовались в качестве входных данных для расчета пространственной структуры пептида SEM1(68-85). Расчет производился в программе XPLOR-NH, в которой использовался стандартный протокол нелинейного отжига. Отдельные структуры были минимизированы, нагреты до 1000 К за 6000 шагов, охлаждены с шагом 100 К до 50 К, и, наконец, структуры были минимизированы с 1000 шагами наискорейшего спуска, за которыми следовали 1000 шагов минимизации сопряженного градиента. Таким образом, всего было сгенерировано 200 структур, которые были уточнены в последующих расчетах с использованием силового поля protein.par [55]. Для пептида SEM1(68-85) был сохранен ансамбль из 10 структур с наименьшими значениями внутренней энергии (рисунок 24). Полученная структура была загружена в международную базу данных Protein Data Bank (PDB) под идентификационным номером PDB ID: 7ZRF.

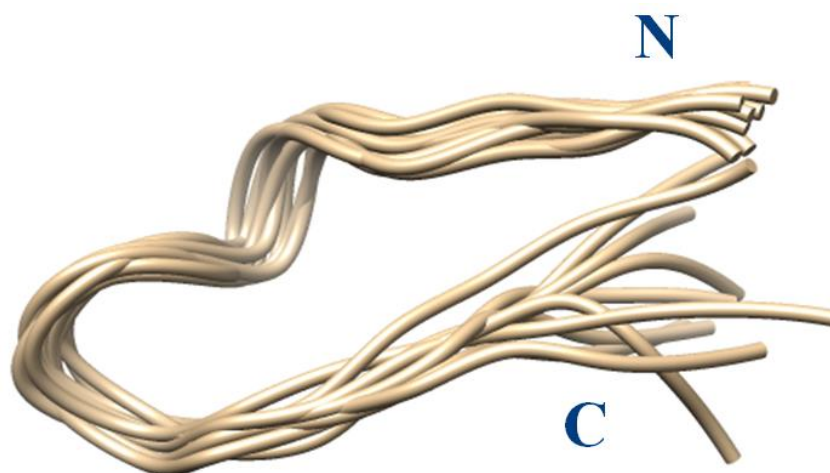


Рисунок 24 – Пространственная структура пептида SEM1(68-85) в водном растворе. Ансамбль из 10 структур с минимальным значением внутренней энергии.

Анализ пространственной структуры показал, что основным типом вторичной структуры пептида SEM1(68-85) является случайный клубок. Фрагмент основной цепи (77Asp-80Arg) имеет хорошую сходимость в ансамбле из 10 структур со среднеквадратичным отклонением (СКО) $0,24 \pm 0,08 \text{ \AA}$. Также пептид SEM1(68-85) имеет подковообразную форму с изгибом в области 76His-78Gln.

Для ансамбля из 10 структур пептида SEM1(68-85) была построена карта Рамачандрана, с помощью которой была проведена проверка рассчитанных структур (рисунок 25): 85,6% аминокислотных остатков находятся в разрешенных областях и 14,4% — в допустимых областях. Таким образом карта Рамачандрана подтверждает пространственную структуру пептида SEM1(68-85) [56].

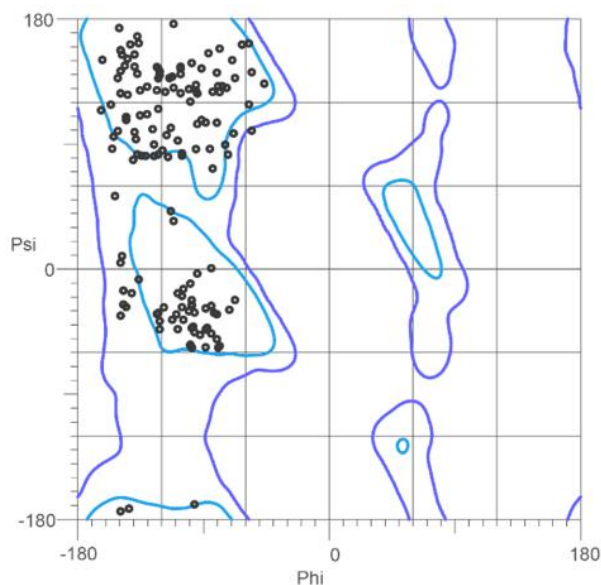


Рисунок 25 – Карта Рамачандрана для ансамбля из 10 структур пептида SEM1(68-85.)

3.3. Расчет пространственной структуры пептида SEM1(68-107) и молекулярно-динамическое моделирование

Для расчета пространственной структуры пептида SEM1(68-107) был создан файл с пространственными ограничениями (межъядерными расстояниями и двугранными углами) пептида SEM1(68-85) и пептида SEM1(86-107), которые были получены ранее [52]. Расчет структуры SEM1(68-107) производился с использованием стандартного протокола нелинейного отжига в программе XPLOR-NIH, описанного в предыдущей части. Полученный ансамбль из 11 структур показан на рисунке 26 и также был загружен в международную базу данных Protein Data Bank (PDB) под идентификационным номером PDB ID: 7ZRO.

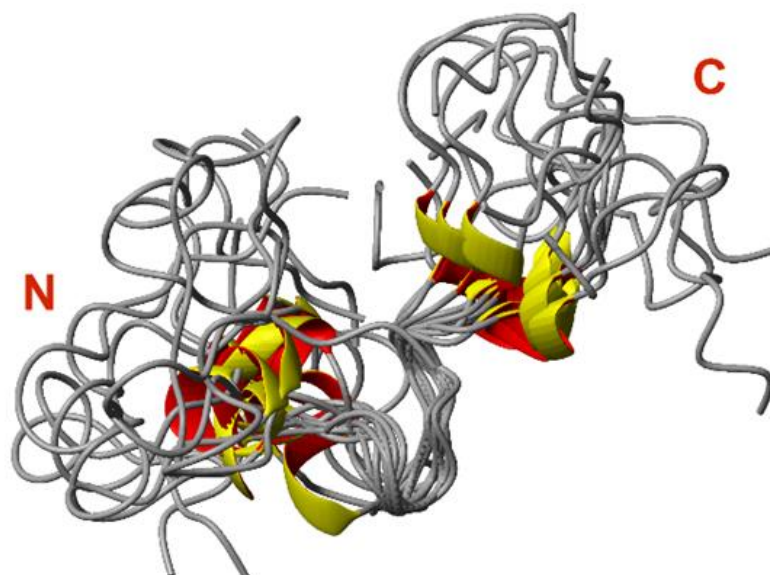


Рисунок 26 – Пространственная структура пептида SEM1(68-107).
Ансамбль из 11 структур.

Поскольку полученная пространственная структура SEM1(68-107) является комбинацией из ЯМР-данных независимо рассчитанных N- и C-пептидных фрагментов SEM1(68-85) и SEM1(86-107), то для исследуемого пептида в мономерной форме было произведено моделирование методом молекулярной динамики в пакете GROMACS. Моделирование производилось с использованием силового поля Charmm36.

В качестве входных данных для МД-моделирования была взята одна структура из полученного ранее ансамбля пептида SEM1(68-107) (рисунок 26), рассчитанного в XPLOR-NIH. Для данной системы был выполнен алгоритм подготовки и уравнивания, описанный в главе 2.2. Параметры моделирования представлены в таблице 2.

На рисунке 27 показан ансамбль из 35 структур SEM1(68-107). Длительность моделирования составляла 350 нс, шаг сохранения структур был выбран 10нс.

Таблица 2 – Параметры моделирования

			Моделирование молекулярной динамики мономера SEM1(68-107)
	NVT	NPT	
Шаг интегрирования	0,5 фс	2 фс	2 фс
Количество шагов	$2 \cdot 10^6$	$5 \cdot 10^5$	$175 \cdot 10^6$
Длительность шага	1 нс	1 нс	350 нс
Температура	300 К	300 К	300 К
Давление	–	1 бар	1 бар

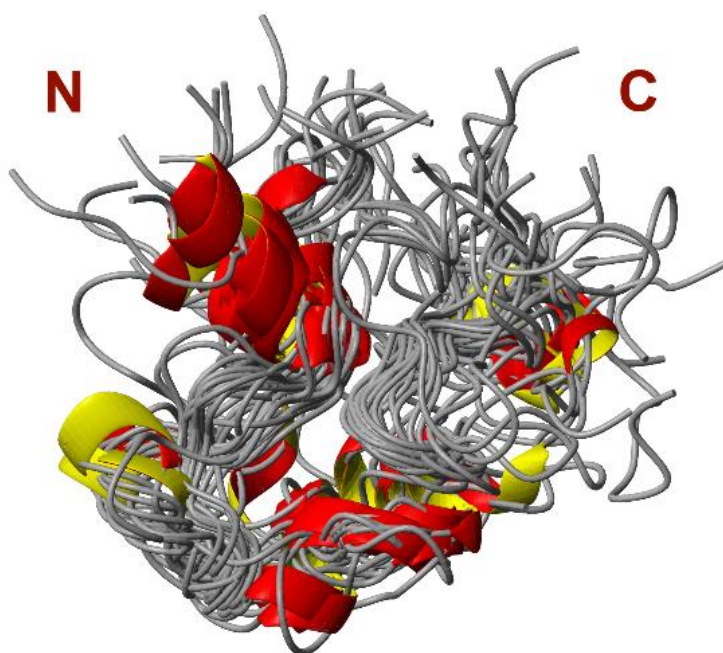


Рисунок 27 – Ансамбль из 35 структур пептида SEM1(68-107), полученный с помощью моделирования методом молекулярной динамики. Длительность моделирования от 0 до 350 нс, шаг сохранения структур 10нс

Для получения информации о структурных изменениях пептида SEM1(68-107) была проанализирована траектория МД-моделирования.

Наблюдения производилось за изменениями значений радиуса гирации (R_g) и расстояния между концами пептида ($perLen$) на протяжении всего времени моделирования (рисунки 28 и 29).

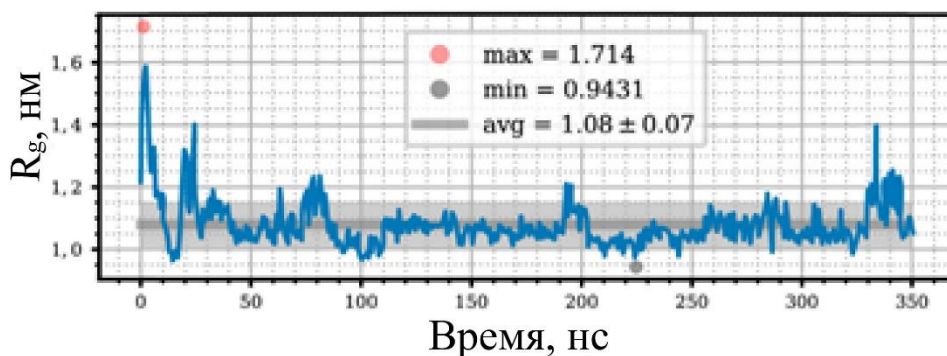


Рисунок 28 – График зависимости радиуса гирации (R_g) пептида SEM1(68-107) от времени

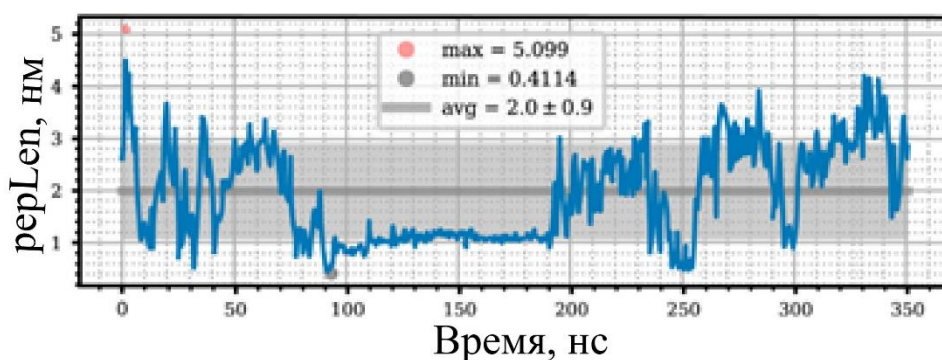


Рисунок 29 – График зависимости расстояния между концами пептида ($perLen$) пептида SEM1(68-107) от времени

Из рисунка 28 видно, что в течение первых 25нс наблюдаются сильные изменения R_g , связанные с конформационной перестройкой пептида. Оставшееся время моделирования сильных изменений радиуса гирации не наблюдалось. Среднее значение R_g равнялось $1,08 \pm 0,07$ нм.

Среднее расстояние между N- и C- концами пептида SEM1(68-107) составляло $2,0 \pm 0,9$ нм на протяжении всего времени моделирования. Из

графика видно (рисунок 29), что rerLen имеет минимальное значение $1,05 \pm 0,05$ нм в интервале от 95 до 191 нс.

В промежутке времени от 100 до 150 нс были зафиксированы наименьшие значения радиуса гирации и расстояния между концами пептида. Поэтому из данного интервала времени моделирования дополнительно был создан ансамбль из 50 структур с шагом сохранения 1 нс (рисунок 30). Значение среднеквадратического отклонения (СКО) основной цепи данного ансамбля равнялось $2,34 \pm 0,85$ Å.

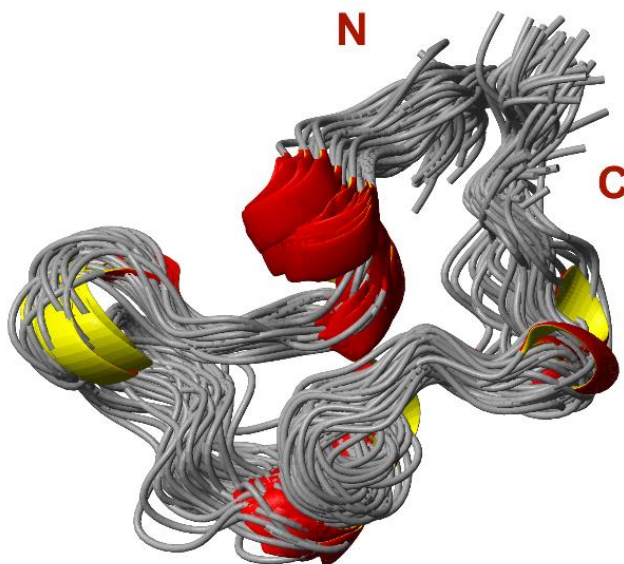


Рисунок 30 – Ансамбль из 50 структур пептида SEM1(68-107), полученный с помощью моделирования методом молекулярной динамики. Длительность моделирования от 100 до 150 нс, шаг сохранения структур 1 нс

Из анализа данных, полученных при молекулярно-динамическом моделировании, можно сделать следующие выводы:

1. N- и C- концы пептида SEM1(68-107) обладают высокой подвижностью, что видно из графика зависимости расстояния между концами пептида от времени (рисунок 29).

2. Показано, что вторичная структура SEM1(68-107) в основном представляет собой случайный клубок, тогда как несколько фрагментов проявляли наличие вторичной структуры: N-конец (70His-84Gln) представлял собой α -спираль; участки (86Asp-94Thr) и (101Gly-103Ser) представляли собой 3_{10} -спираль.

3. Нестабильность структурированных фрагментов SEM1(68-107) может быть объяснена внутренней неупорядоченностью пептида.

3.4. Молекулярно-динамическое моделирование димера SEM1(68-107)

Для понимания процесса фибриллообразования необходимо знать механизмы олигомеризации. Поэтому для исследуемого пептида SEM1(68-107) было произведено моделирование димера методом молекулярной динамики в пакете GROMACS, а также методом метадинамики с помощью библиотеки PLUMED для GROMACS. Моделирование производилось с использованием силового поля Charmm36.

В качестве входных данных для расчета димера была взята одна структура из полученного ранее ансамбля пептида SEM1(68-107) (рисунок 30) после МД-моделирования. Далее был создан pdb-файл с двумя мономерами, размещенными на максимальном отдалении друг от друга внутри ячейки, заполненной водой. Минимизация энергии и уравнивание системы производились в соответствии с алгоритмом, описанным в главе 2.2.

Построение профиля свободной энергии ранее уравновешенной и оптимизированной системы димера SEM1(68-107) было выполнено в пакете GROMACS с использованием библиотеки PLUMED. В качестве коллективных переменных были выбраны расстояния между центрами масс мономеров и число близких контактов между атомами C_{α} (значения параметров $n = 8, m = 10, r_0 = 0,65$ нм, для формулы 29 из главы 1.5.6). Высота функции Гаусса составляла $W = 0.25$ кДж/моль. Через каждые 100 шагов по 2 фс происходило

добавление гауссиана и изменение потенциала U_{bias} . Параметры моделирования методом метадинамики приведены в таблице 3.

Таблица 3 – Параметры моделирования

			Моделирование метадинамики	Моделирование молекулярной динамики димера SEM1(68-107)
	NVT	NPT		
Шаг интегрирования	0,5 фс	2 фс	2 фс	2 фс
Количество шагов	$2 \cdot 10^6$	$5 \cdot 10^5$	$5 \cdot 10^8$	$5 \cdot 10^8$
Длительность шага	1 нс	1 нс	1 мкс	1 мкс
Температура	300 К	300 К	300 К	300 К
Давление	–	1 бар	1 бар	1 бар

Профиль свободной энергии ΔF для димера SEM1(68-107) показан на рисунке 31. Анализ полученных результатов показал, что наиболее вероятные конформации исследуемого объекта находятся в области, когда расстояние между центрами масс $R = 1.8\text{нм}$, а количество близких контактов $N_{ca} = 42$.

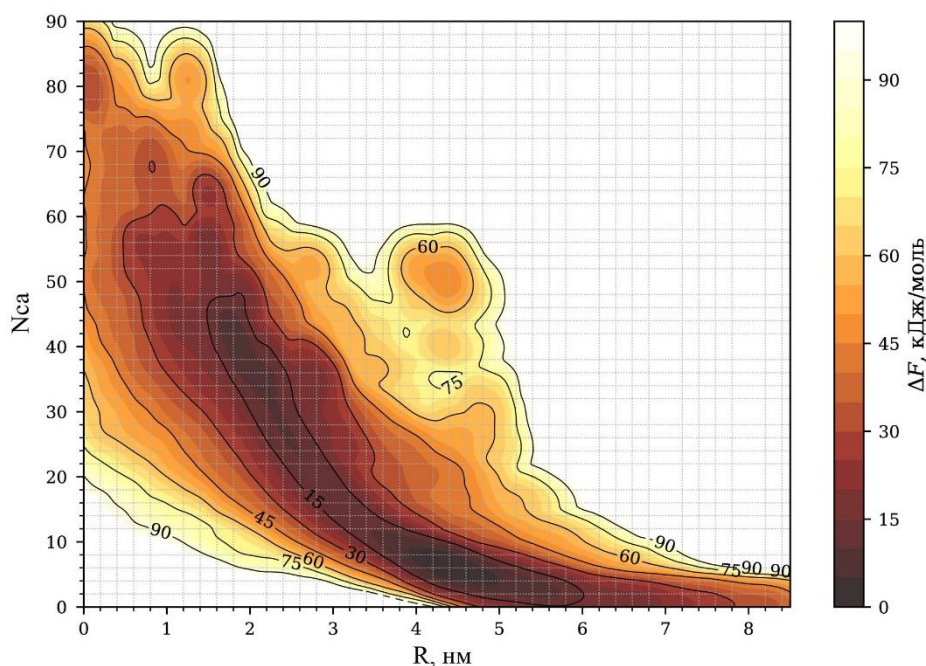


Рисунок 31 – Профиль свободной энергии ΔF для димера SEM1(68-107) в зависимости от расстояния между центрами масс R и количеством близких контактов N_{ca} .

Из полученных файлов со структурами SEM1(68-107) была выбрана одна конформация, для которой было проведено молекулярно-динамическое моделирование (1000 нс) для проверки стабильности структуры димера. Параметры моделирования представлены в таблице 3. С помощью полученной МД-траектории были проанализированы пространственные изменения структуры димера SEM1(68-107). На рисунке 32 показана зависимость расстояния между центрами масс (R) от времени. Из графика видно, что в течение первых 200нс происходила конформационная перестройка молекулы, связанная с увеличением расстояния между центрами масс. Далее сильных изменений R от времени не происходило. Среднее расстояние между центрами масс было равно 1.1 ± 0.6 нм.

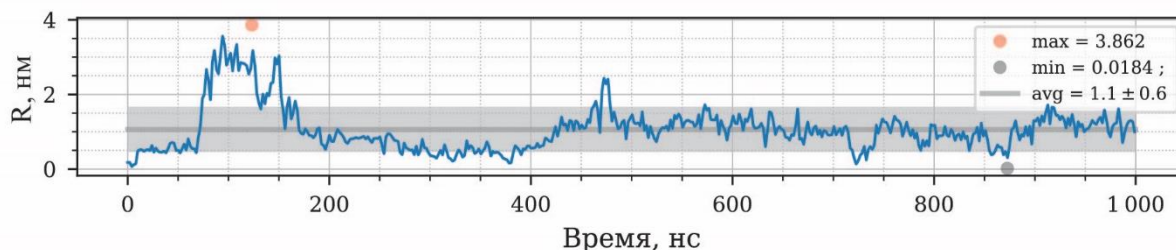


Рисунок 32 – График зависимости расстояния между центрами масс димера SEM1(68-107) от времени

Также производилось наблюдение за количеством близких контактов N_{ca} между атомами C_{α} (рисунок 33). Из графика видно, что при отдалении молекул друг от друга в течение первых 200нс уменьшалось количество близких контактов. Оставшиеся 800нс моделирования сильных изменений количества N_{ca} не происходило, и в среднем оно составляло $N_{ca} = 57 \pm 9$. Данные результаты показывают, что сильных конформационных изменений в структуре димера SEM1(68-107) не происходило и, соответственно, структура оставалась стабильной.

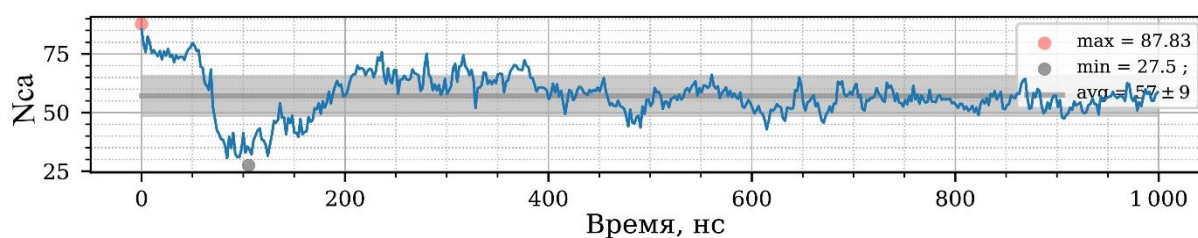


Рисунок 33 – График зависимости количества близких контактов N_{ca} димера SEM1(68-107) от времени

Для структур, сохраненных в процессе моделирования, была проведена кластеризация (рисунок 34). Она показывает, что 24,8% из общего количества димеров являются геометрически схожими. Из наиболее популярного кластера был отобран ансамбль из 21 структуры, представленный на рисунке

35А. Данный ансамбль имеет подвижные концевые области, но центральная часть (с 72Asp по 100Leu аминокислотные остатки) остается стабильной.

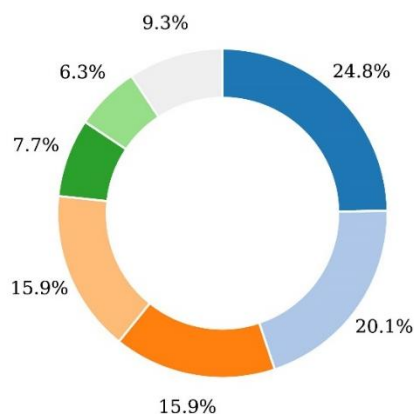


Рисунок 34 –Кольцевая диаграмма, показывающая вклад кластеров, структурно схожих моделей димеров SEM1(68-107.)

При более подробном рассмотрении структуры димера SEM1(68-107) (рисунок 35Б) можно заметить, что в структуре каждого пептида произошли конформационные изменения (рисунки 36 и 37) по сравнению с первоначальной структурой мономера SEM1(68-107) (рисунок 27). Оба пептида в димере имеют подковообразный участок с изгибом в районе 81Lys и 85Tyr аминокислотных остатков. Данный перегиб относится к типу вторичной структуры β -петля, потому что 5 аминокислотных остатков (81K82S83Q84Q85Y) на этом участке изменяют направление основной цепи на 180°. Между 81Lys и 85Tyr аминокислотными остатками имеется водородная связь, которая также характерна для β -петли (рисунок 36) [9].

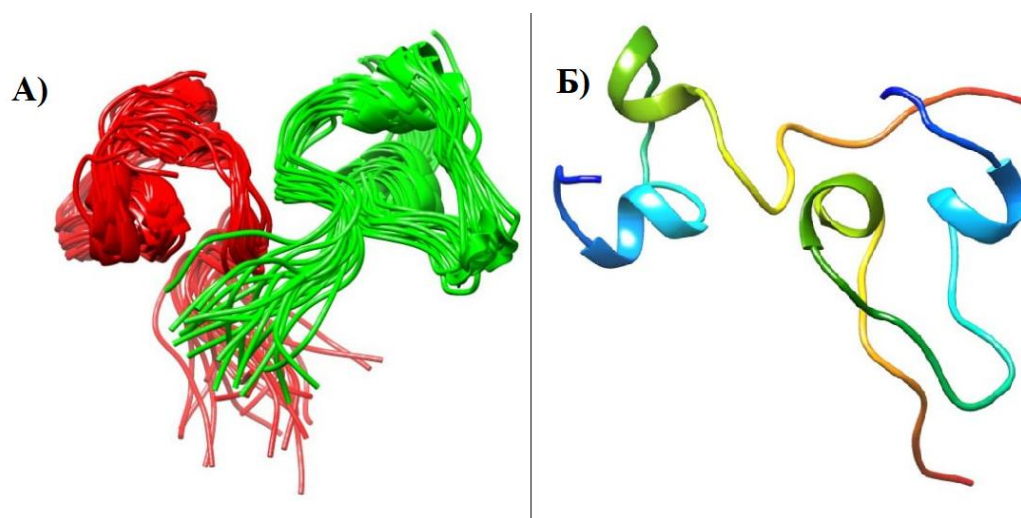


Рисунок 35 – А) Ансамбль из 21 структуры димеров SEM1(68-107). Б) Моделей димера SEM1(68-107) из ансамбля, показанного на рисунке 35А

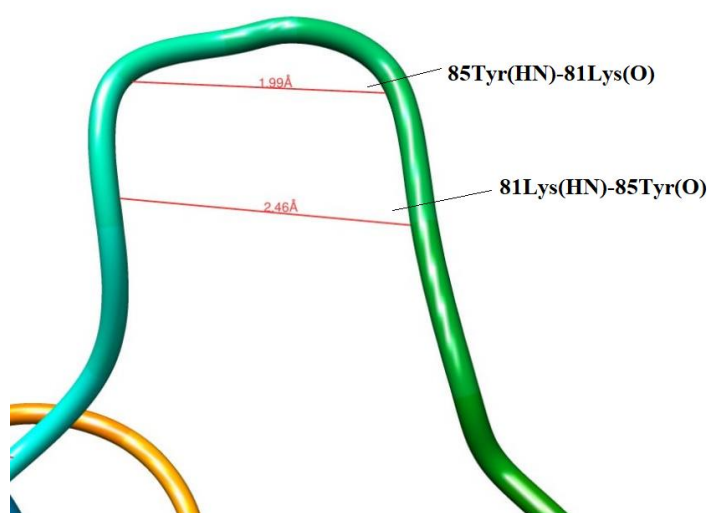


Рисунок 36 – Показана структура β -петли, которая удерживается водородными связями между 81Lys и 85Tyr аминокислотными остатками

Оба пептида в димере имеют стабильные спиральные области :72Asp-77Asp и 90Leu-93Thr. Такие же спиральные участки наблюдались при моделировании мономера SEM1(68-107) (рисунок 27). У каждого пептида в димере SEM1(68-107) минимальное расстояние между двумя спиральными областями 72Asp-77Asp и 90Leu-93Thr находится в интервале от 1.5Å до 2.0Å. Эти спиральные области расположены параллельно друг другу в структурах

пептидов в димере (рисунок 37). Стоит отметить, что спиральная область 72D73A74N75D76H77D имеет в своем составе 3 аминокислотных остатка аспарагиновой кислоты (D) и 1 аминокислотный остаток гистидина (H), а спиральная область 90L91H92K93T – 1 аминокислотный остаток гистидина (H).

Описанные выше конформационные изменения в структуре димера SEM1(68-107) характерны для образования зародышевого ядра в период лаг-фазы олигомеризации амилоидных пептидов [57, 58]. Также полученные результаты в данной работе показывают, что мы наблюдаем промежуточное состояние перед агрегацией β -листов. Ведь образование амилоидогенной сборки, содержащей спиральные фрагменты, является ключевым этапом в образовании фибрилл [59]. Также отмечается, что наличие аминокислотных остатков Asp и His играют важную роль в кинетике образования спиральных областей, предшествующим образованию β -листов [59].

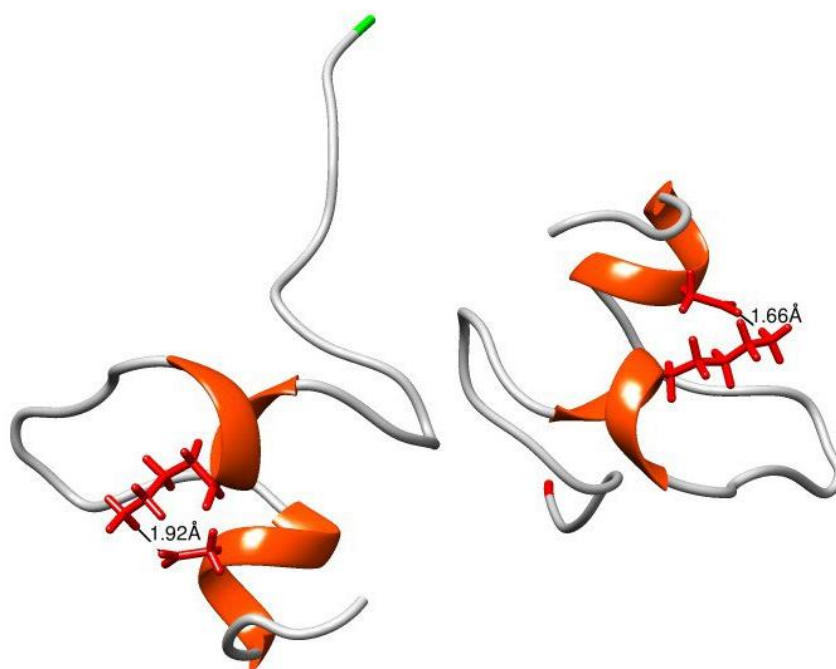


Рисунок 37 – Оранжевым выделены спиральные области в структурах пептидов в димере SEM1(68-107) с обозначением минимальных расстояний между спиралями

ВЫВОДЫ

В ходе работы были получены следующие результаты:

1. Выполнено соотнесение химических сдвигов пептида SEM1(68-107) для 34 из 40 аминокислотных остатков (85%). Значения химических сдвигов ^1H , ^{15}N и ^{13}C депонированы в международную базу данных BioMagResBank (BMRB ID 51406).

2. Исследован пептид SEM1(68-85) методом ЯМР-спектроскопии. Сделано отнесение химических сдвигов ^1H , ^{13}C и ^{15}N пептида SEM1(68-85) (BMRB ID 51400). Впервые рассчитана пространственная структура SEM1(68-85). Структура загружена в международную базу данных Protein Data Bank (PDB ID: 7ZRF). Основным типом вторичной структуры пептида SEM1(68-85) является случайный клубок. Также пептид SEM1(68-85) имеет подковообразную форму с изгибом в области 76His-78Gln.

3. Впервые была определена пространственная структура пептида SEM1(68-107) с помощью комплексного подхода, основанного на экспериментальных ограничениях отдельных пептидных фрагментов и молекулярном моделировании.

4. Было проведено молекулярно-динамическое моделирование димера пептида SEM1(68-107). Наблюдается, что спиральные области пептидов SEM1(68-107) в димере выстраиваются параллельно друг другу. Такая конформационная перестройка в структуре димера SEM1(68-107) характерны для образования зародышевого ядра в период лаг-фазы олигомеризации.

СПИСОК ИСПОЛЬЗУЕМОЙ ЛИТЕРАТУРЫ

1. Simon, V. HIV/AIDS epidemiology, pathogenesis, prevention, and treatment / V. Simon, D. D. Ho, Q. A. Karim // *The Lancet*. – 2006. – V. 368(9534). – P. 489 – 504. [https://doi.org/10.1016/S0140-6736\(06\)69157-5](https://doi.org/10.1016/S0140-6736(06)69157-5).
2. Dimitrov, D.S. Quantitation of human immunodeficiency virus type 1 infection kinetics / D.S. Dimitrov, R. L Willey, H. Sato, L.J. Chang, R. Blumenthal, M. A. Martin. // *Journal of Virology*. – 1993. – V. 67(4). – P. 2182 – 2190. <https://doi.org/10.1128/JVI.67.4.2182-2190.1993>.
3. Pomerantz, R.J. Cross-talk and viral reservoirs. / R.J. Pomerantz. // *Nature*. – 2003. – V. 424 – P. 136–137.
4. de Lamirande, E. Semenogelin, the main protein of the human semen coagulum, regulates sperm function. / E. de Lamirande. // *Semin. Thromb. Hemost.* – 2007. – V. 33. – P. 60–68.
5. Roan, N. R. Peptides released by physiological cleavage of semen coagulum proteins form amyloids that enhance HIV infection. / N. R. Roan, J. A. Muller, H. Liu, et al. // *Cell Host Microbe*. – 2011. – V. 10. – P. 541-550.
6. Puthenveetil, R. Solution NMR: A powerful tool for structural and functional studies of membrane proteins in reconstituted environments. / R. Puthenveetil, O. Vinogradova. // *J. Biol. Chem.* – 2019. – V. 294. – P. 15914-15931.
7. Аюпов, Р. Х. Выделение и очистка белка для исследования методом ЯМР высокого разрешения / Р. Х. Аюпов, К. С. Усачев. – Казань: Казаню.ун-т, 2019. – 39 с.
8. Allen, M. P. Introduction to Molecular Dynamics Simulation. / M. P. Allen // *Computational Soft Matter: from Syntetic Polymers to Proteins*. – 2004. – V. 23. – P. 1-28.
9. Финкельштейн, А. В. Физика белка: Курс лекций с цветными стереоскопическими иллюстрациями и задачами / А. В. Финкельштейн, О. Б. Птицын. – Москва: КДУ, 2012. – 456 с.

10. Anfinsen, C. B Principles that Govern the Folding of Protein Chains. / C. B. Anfinsen. // Science. 1973. – V. 181. – P. 223-230.
11. Nizhnikov, A. A. Amyloids: from pathogenesis to function. / A. A. Nizhnikov, K. S. Antonets, S. G. Inge-Vechtomov. // Biochemistry (Moscow). – 2015. – V. 80. – P. 127-1144.
12. Hou, F. MAVS Forms Functional Prion-like Aggregates to Activate and Propagate Antiviral Innate Immune Response. / F. Hou, L. Sun, B. Skaug, et al. // Cell. – 2011. – V. 146 (3). – P. 448-461.
13. Abedini, A. A role for helical intermediates in amyloid formation by natively unfolded polypeptides? / A. Abedini, D. P. Raleigh. // Physical Biology. – 2010. – V. 6 (1). – P. 015005.
14. Ferrone, F. A. Nucleation^ The Connections Between Equilibrium and Kinetic Behavior. / F. A. Ferrone. // Methods in Enzymology. – 2006. – V. 412. – P. 285-299.
15. Roan, N. R. Liquefaction of Semen Generates and Later Degrades a Conserved Semenogelin Peptide That Enhances HIV Infection. / N. R. Roan, H. Liu, S. M. Usmani, et al. // J. Virol. – 2014. – V. 88. – P. 7221-7234.
16. Lilja, H. A kallikrein-like serine protease in prostatic fluid cleaves the predominant seminal vesicle protein. / H. Lilja. // JCI. – 1958. – V. 76 (5). – P. 1899-1903.
17. Vickram, S. Semenogelin, a coagulum macromolecule monitoring factor involved in the first step of fertilization^ A prospective review. / S. Vickram, K. Rohini, K. Anbarasu, et al. // Macromolecule. – 2022. – V 209A, P. 951-962.
18. Castellano, L. M. The Surprising Role of Amyloid Fibrils in HIV Infection / L. M. Castellano, J. Shorter. // Biology. – 2012. – V. 1 (1). – P. 58-80.
19. Jones, R. Interaction of zona pellucida glycoproteins, sulphated carbohydrates and synthetic polymers with proacrosin, the putative egg-binding protein from mammalian spermatozoa. / R. Jones. // Development. – 1991. – V. 111. – P. 1155-1163.

20. Cohen, F.S.; Melikyan, G.B. The Energetics of Membrane Fusion from Binding, through Hemifusion, Pore Formation, and Pore Enlargement / F. C. Cohen, G. B. Melikyan. // *J. Membr. Biol.* – 2004. – V. 199. – P.1-14.
21. Дероум, Э. А. Современные методы ЯМР для химических исследований. / Э. А. Дероум; пер. с англ. Ю. М. Демина. – Москва: Мир, 1992.
22. Bax, A. Two-Dimensional NMR in Liquids. / A. Bax. – Dordrecht: Reidel, 1984.
23. Berger, S. 200 and more NMR experiments. / S. Berger, S. Braun. – Weinheim: Wiley-VCH, 2004.
24. Rule, G. S. Fundamentals of Protein NMR Spectroscopy. / G. S. Rule. T. K. Hitchens. – Netherlands: Springer, 2006.
25. Keller, J. Understanding NMR Spectroscopy. / J. Keller. – England: John Wiley, 2005.
26. Schleucher, J. A general enhancement scheme in heteronuclear multidimensional NMR employing pulsed field gradients. / J. Schleucher, M. Schwendinger, M. Sattler, et al. // *J. Biomol. NMR.* – 1994. – V. 4. – P. 301-306.
27. Parella, T. Improved sensitivity in gradient-based 1D and 2D multiplicity-edited HSQC experiments. / T. Parella, F. F. Sanchez, A. Virgili. // *Journal of Magnetic Resonance.* – 1997. – V. 126– P. 274-277.
28. Sklenar, V. Gradient-Tailored Water Suppression for ^1H - ^{15}N HSQC Experiments Optimized to Retain Full Sensitivity. / V. Sklenar, M. Piotto, R. Leppik, V. Saudek. // *J. Magn. Reson.* – 1993. – V. 102A. – P. 241-245.
29. Schleucher, J. Coherence Selection by Gradients without Signal Attenuation: Application to the Three-Dimensional HNC0 Experiment / J. Schleucher, M. Sattler, C. A. Griesinger. // *Chem. Int. Ed.* – 1993. – V. 32. – 1489-1491.
30. Grzesiek, S. Improved 3D triple-resonance NMR techniques applied to a 31kDa protein / S. Grzesiek, A. Bax. // *J. Magn. Reson.* – 1992. – V. 96. – P. 432-440.

31. Wittekind, M. HNCACB, a High-Sensitivity 3D NMR Experiment to Correlate Amide-Proton and Nitrogen Resonances with the Alpha and Beta-Carbon Resonances in Proteins. / M. Wittekind, L. Mueller. // *J. Magn. Reson.* – 1993. – V. 101B. – P. 201-205.
32. Clubb, R. T. A Constant-Time Three-Dimensional Triple- Resonance Pulse Scheme to Correlate Intraresidue ^1H , ^{15}N , and ^{13}C chemical shifts in ^{15}N – ^{13}C -Labelled Proteins. / R. T. Clubb, V. Thanabal, G. Wagner. // *J. Magn. Reson.* – 1992. – V. 97. – 213-217.
33. Шайтан, К. В. Молекулярная динамика белков и пептидов. / К. В, Шайтан, К. Б. Терёшкина. – Москва: МГУ, 2004 – 103 с.
34. Allen, M. P. Introduction to molecular dynamics. / M. P. Allen. // *Computational soft Matter: From Synthetic Polymers to Proteins.* – 2004. – V. 23 – P. 1-28.
35. Berendsen, H. J. C. Molecular Dynamics with coupling to an external bath. / H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. DiNola, J. R Naak. // *Journal of Applied Physics.* – 1981. – V. 52. – P. 3684-3690.
36. Parinello, M. Polymorphic transitions in single crystals: A new molecular dynamics method / M. Parinello, A. Rahman. // *Journal of Applied Physics.* – 1981. – V. 52. – P. 7182-7190.
37. Nose, S. Constant pressure molecular dynamics for molecular systems. / S. Nose, M. L. Klein. // *Molecular Physics.* – 1983. – V. 50. – P. 1055-1076.
38. Spoel, D. V. D. GROMACS: fast, flexible, and free / D. V. D. Spoel, E. Lindahl, B. Hess, et al. // *Journal of Computational Chemistry.* – 2005. – V. 26. – P. 1701-1718.
39. Barducci, A. Well-tempered metadynamics: A smoothly converging and tunable free-energy method. / A. Barducci, G. Bussi, M. Parrinello. // *Physical Review Letters.* – 2008. – V. 100. – P. 020603.
40. Bikmullin, A. The data of heterologous expression protocol for synthesis of ^{15}N , ^{13}C -labeled SEM1(68–107) peptide fragment of homo sapiens

semenogelin1. / A. Bikmullin, E. Klochkova, F. Krasnovid, D. Blokhin. // *MethodsX*. – 2021. – V. 8. – P. 101512.

41. Kamalov, M. I. Synthesis and Characterization of Polyaspartic Acid-Histidine Conjugate as an Analog of Antioxidant Enzymes. / M. I. Kamalov, G. R. Sadrieva, A. M. Pavlyuk, et al. // *Appl. Biochem. Microbiol.* – 2019. – V. 55. – P. 474-481.

42. Vranken, W. F. The CCPN data model for NMR spectroscopy: development of a software pipeline. / W. F. Vranken, W. Boucher, T. J. Stevens, et al. // *Proteins*. – 2005. – Vol. 59. – P. 687-696.

43. Schwieters, C. D. The Xplor-NIH NMR molecular structure determination package. / C. D. Schwieters, J. J. Kuszewski, N. Tjandra, G. M. Clore // *Journal of Magnetic Resonance*. – 2003. – Vol. 160 (1). – P. 65-73.

44. Pettersen, E. F. UCSF CHIMERA – a visualization system for exploratory research and analysis. / E. F. Pettersen, T. D. Goddard, T. E. Ferrin, et al. // *Journal of Computational Chemistry*. – 2004 – Vol. 25 (13). – P. 1605-1612.

45. Pronk, S. GROMACS 4.5: A high-throughput and highly parallel open source molecular simulation toolkit. / S. Pronk, S. Pall, R. Schulz, et al. // *J. Bioinform.* – 2013. – V. 29. – P. 845-854.

46. Huang, J. CHARMM36 All-Atom Additive Protein Force Field: Validation Based on Comparison to NMR Data. / J. Huang, A. D. MacKerell Jr. // *J. Comput. Chem.* – 2013. – V. 34. – P. 2135-2145.

47. Jorgensen, W.L. Comparison of simple potential functions for simulating liquid water / W. L. Jorgensen, J. Chandrasekhar, J. D. Madura. // *J. Chem. Phys.* – 1983. – V. 79. – P. 926-935.

48. Mao, Y. Thermal conductivity, shear viscosity and specific heat of rigid water models. / Y. Mao, Y. Zhang. // *Chemical Physics Letters*. – 2012. – V. 542. – P. 37-41.

49. Bonomi, M. PLUMED: A portable plugin for free-energy calculations with molecular dynamics. / M. Bonomi, D. Branduardi, G. Bussi, et al. // *Computer Physics Communication*. – 2010 – V. 180 (10). – P. 1961-1972.

50. Tribello, G. A. PLUMED 2[^] New feathers for an old bird. / G. A. Tribello, M. Bonomi, D. Branduardi, C. Camilloni, G. Bussi. // *Computer Physics Communication*. – 2014. V. 185 (2). – P. 604-613.
51. Schmitz, T. NMR-Based Structural Characterization of a Two-Disulfide-Bonded Analogue of the FXIIIa Inhibitor Tridegin: New Insights into Structure-Activity Relationships. / T. Schmitz, A. A. P. George, B. Nubbemeyer, et al. // *Int. J. Mol. Sci.* – 2021. V. 22, P. 880.
52. Sanchugova, D. The Structure of Fibril-Forming SEM1(86-107) Peptide Increasing the HIV Infectivity. / D. Sanchugova, A. Kusova, A. Bikmullin, V. Klochkov, D. Blokhin. // *Bionaniscience*. – 2021. – V. 11. – P. 182-188.
53. Wuthrich, K. NMR of Proteins and Nucleic Acids. / K. Wuthrich. // – New York: Wiley-VCH, 1986.
54. Rule, G.S. Fundamentals of Protein NMR Spectroscopy. / G. S. Rule – Dordrecht: Springer, 2006.
55. Schwieters, C. D. Using Xplor – NIH for NMR molecular structure determination. / C. D. Schwieters, J. J. Kuszewski, G. M. Clore // *Prog. Nucl. Magn. Reson. Spectrosc.* – 2006. – V. 48. – P. 47-62.
56. Laskowski, R. A. The ramachandran plot and protein structure validation. / R. A. Laskowski, N. Furnham, J. M. Thornton. // *World Scientific* – 2013 – P. 62-75.
57. Goldschmidt, L. Identifying the amyloids, proteins capable of forming amyloid-like fibrils. / L. Goldschmidt, P. K. Teng, R. Riek, D. Eisenberg. // *Proceedings of the National Academy of Sciences of the USA*. – 2013. V. 107 (8). – P. 3487-3492.
58. Liu, G. Mechanistic Studies of Peptide Self-Assembly: Transient α -Helices to Stable β -Sheets. / G. Liu, A. Prabhakar, D. Aucoin, et al. // *Journal of the American Chemical Society*. – 2010. – Vol. 132 (51). – P. 18223–18232.
59. Kirkitadze, M. Identification and Characterization of Key Kinetic Intermediates in Amyloid β -protein Fibrillogenesis. / M. D. Kirkitadze, M. M. Condrón, D. B. Teplow. // *JMB*. – 2001. – V. 312. – P. 1103-1119.