

Федеральное государственное бюджетное учреждение науки Федеральный исследовательский центр «Институт общей физики им. А.М. Прохорова Российской академии наук»

Федеральное государственное бюджетное учреждение науки Институт биоорганической химии им. академиков М.М. Шемякина и Ю.А. Овчинникова Российской академии наук



на правах рукописи

Можаева Вера Александровна

**МЕТОДЫ СПЕКТРОСКОПИИ КОМБИНАЦИОННОГО РАССЕЯНИЯ СВЕТА
ДЛЯ ИДЕНТИФИКАЦИИ БЕЛКОВЫХ ТОКСИНОВ**

Специальность: 1.3.19. Лазерная физика

Диссертация на соискание ученой степени кандидата

физико-математических наук

Научный руководитель:

доктор физико-математических наук

Гарнов Сергей Владимирович

Москва – 2024

Оглавление

Оглавление	2
Введение.....	4
Глава 1. Обзор литературы.....	11
1.1. Метод рамановской спектроскопии в исследовании белковых молекул.....	11
1.1.1. Общие сведения.....	11
1.1.2. Методические вопросы.....	14
1.1.3. Спектры КР белков.....	19
1.1.4. SERS и резонансная РС.....	22
1.2. Другие методы определения структуры белков.....	25
1.2.1. Рентгеноструктурный анализ и ядерный магнитный резонанс	25
1.2.2. Круговой дихроизм	26
1.2.3. Предсказание структуры белка	26
1.3. Животные токсины и яды.....	27
1.3.1. Токсины животных белковой природы.....	27
1.3.2. РС в исследовании пептидных и белковых токсинов.....	30
1.3.3. Яды змей.....	32
1.3.4. Метод масс-спектрометрии для анализа цельных ядов.....	35
1.4. Метод главных компонент для анализа спектральных данных.....	36
1.4.1. МГК и методы кластеризации.....	36
1.4.2. Применение МГК к данным РС.....	38
1.5. Выводы к главе 1.....	40
Глава 2. Классификация спектров КР токсинов с помощью МГК.....	41
2.1. Материалы и методы.....	41
2.1.1. Образцы пептидов и белков	41
2.1.2. Сбор данных.....	43
2.1.3. Обработка данных	43
2.2. Результаты.....	45
2.2.1. Анализ спектров КР токсинов.....	45
2.2.2. Исследование спектрального региона вторичной структуры токсинов.....	51
2.2.3. Сравнение спектров КР изомеров конотоксина RgIA	55
2.2.4. Классификация спектров КР трехпетельных токсинов	57
2.2.5. Сравнение триптофанового дублета для растворов и сухих образцов	58
2.3. Выводы к главе 2.....	59
Глава 3. Классификация ядов разных видов змей с помощью РС.....	60

3.1. Материалы и методы.....	60
3.1.1. Образцы ядов	60
3.1.2. Сбор данных.....	61
3.1.3. Обработка спектральных данных	61
3.2. Результаты.....	62
3.2.1. Применение МГК к спектрам КР ядов для их классификации.....	62
3.2.2. Применение ГК-DFA для разделения спектров КР ядов по семействам змей.....	65
3.3.3. Прямой анализ полос Amide I и S-S спектров образцов.....	69
3.3. Выводы к главе 3.....	74
Глава 4. Анализ спектров КР ядов отдельных особей змей.....	76
4.1. Материалы и методы.....	76
4.1.1. Образцы ядов и фосфолипаз A2.....	76
4.1.2. Сбор данных.....	77
4.1.3. Обработка спектральных данных	77
4.2. Результаты.....	78
4.2.1. Применение МГК к спектрам КР образцов ядов.....	78
4.2.2. Анализ полосы Amide I ядов VN и фосфолипаз A2.....	82
4.3. Выводы к главе 4.....	85
Глава 5. Сравнение спектров КР рекомбинантного и синтетического белка SLURP-1	86
5.1. Материалы и методы.....	86
5.2. Результаты.....	87
5.3. Выводы к главе 5.....	91
Заключение	92
Благодарности.....	94
Список литературы	95

Введение

Рамановская спектроскопия (РС) или спектроскопия комбинационного рассеяния (КР) света – относительно простая и удобная, и в то же время крайне информативная методика неразрушающего исследования вещества с помощью лазерного излучения. Ее применение к органической материи позволяет исследовать вопросы биологии и медицины. Токсикология имеет прямое отношение к последним, так как изучает медицински-значимые яды биологического происхождения. Применение современной лазерной оптической техники к объекту медицины – токсинам животного происхождения – является актуальной задачей.

РС позволяет получать информацию о частотах молекулярных колебаний. Эти частоты отражают как состав молекулы, так и внутри- и внемолекулярные силы, которые содержат информацию о структуре, конформации, а также окружении молекулы. РС сочетает такие преимущества, как простота пробоподготовки, быстрота, информативность и экономичность. Она может быть крайне полезна в токсикологии, так как потенциально способна обеспечить получение уникальной информации о структурах белков [1], в частности, токсинов [2], а также о составе многокомпонентных смесей биомолекул (например, [3]), коими являются животные яды.

Актуальность темы исследования

Компоненты яда – токсины – крайне разнообразные и удивительные биологические молекулы, многие из которых обладают сложной структурой, часто включающей несколько дисульфидных связей [4]. Белковые и пептидные токсины интересны, в первую очередь, уникальными эффектами, которые они способны оказывать на организм, в частности, млекопитающих. В процессе длительной эволюции структуры этих биомолекул были оптимизированы для оказания определенных биологических эффектов. Поэтому структурные библиотеки токсинов могут послужить источником незаменимой информации о тех особенностях их строения, которые ответственны за определенное воздействие данных молекул на организм. При разработке (дизайне) лекарств оказывается крайне полезным «подсмотреть» у природы структуры природных биологически-активных веществ. Некоторые животные токсины уже используются в качестве лекарств [5], многие рассматриваются как многообещающие кандидаты на эту роль [6, 7, 8]. Работы в этом направлении активно ведутся. В частности, в качестве мышечных релаксантов и как средство для лечения мышечной дистонии могут рассматриваться определенные токсины из ядов змей, являющиеся ингибиторами никотинового ацетилхолинового рецептора мышечного типа [9].

Таким образом, исследование токсинов методами лазерной спектроскопии, несмотря на их, на первый взгляд, вредоносную природу, прежде всего – путь к новым эффективным, действующим на определенную цель, лекарственным веществам. Эту эффективность можно использовать во благо в медицине. Определение конкретных структурных мотивов в токсинах, за счет которых реализуется их биологическое действие, дает возможность последующего искусственного синтеза молекулы, содержащей данный мотив. Высокая избирательность (селективность) к цели (например, конкретному рецептору или ионному каналу) и скорость взаимодействия многих токсинов делает их не только многообещающими кандидатами в лекарственные препараты, но и полезными инструментами в биологии, которые позволяют идентифицировать различные ионные каналы [10]. Высокоселективное взаимодействие токсинов с соответствующими мишенями помогает понять нормальные физиологические процессы. Также на основе функциональных мишеней различных токсинов можно разработать ряд систем тканеселективной доставки [11].

Существующие методы анализа белковых структур, такие как рентгеноструктурный анализ и спектроскопия ядерного магнитного резонанса, трудоемки и времязатратны, а также требуют наличия относительно большого количества природных токсинов ядовитых животных. Помимо исследования нативных структур природных токсинов, важной задачей является валидация структур и конформаций их искусственно-синтезированных аналогов. Отсюда следует актуальность разработки подходов на базе РС анализа структур и конформаций природных и искусственных токсинов.

Задача же определения составов змеиных ядов и их меж- и внутривидовых вариаций, например, методами лазерной спектроскопии, непосредственно связана с разработкой и выбором противоядия для жертв укусов [12]. Змеиные укусы все еще являются причиной множества смертей. Всемирная организация здравоохранения поставила цель сократить вдвое смертность от укусов змей к 2030 году. В 2019 году от змеиных укусов во всем мире умерло более 63000 человек [13]. По оценкам, каждый год 5.4 миллиона человек подвергаются укусам, причем до 2.7 миллионов получают отравление. Наибольшее бремя легло на Южную Азию. Доступность компактных рамановских спектрометров [14] делает возможным проведение полевых исследований, что важно для работы в развивающихся странах, где смертность от укусов змей достаточно высока, а лабораторный анализ зачастую не доступен. Знание состава и вариаций ядов важно и для теоретической науки [15], в частности, отдельный интерес представляют межпопуляционные вариации ядов [16].

Для изучения состава ядов в настоящее время используются методы, основанные на хроматографическом разделении и масс-спектрометрическом анализе, которые требуют пробоподготовки и применения сложного оборудования [17]. РС, в некоторых случаях, может

послужить альтернативой использованию этих методов. Комбинация этой спектроскопии с математическими методами понижения размерности (например, методом главных компонент) и кластеризации может упростить, ускорить и оптимизировать получение полезной информации об образцах (см., например, [18, 19]).

Исследование спектров КР многокомпонентных смесей, таких как яды, можно рассматривать как задачу спектроскопического изучения сложных многокомпонентных образцов. Рассмотрение свойств и информативности таких спектров, а также разработка методологии получения полезной информации из спектров подобных сложных образцов важны с точки зрения спектроскопической теории. Интересны также вопросы о том, насколько спектр многокомпонентной смеси соответствует сумме спектров компонент; возникает ли резонансное взаимодействие излучения с определенными компонентами разнородной смеси.

Необходимо подчеркнуть, что работа с биологическими образцами требует подбора специальных условий записи спектров, а именно: выбор параметров лазерного излучения и оптических элементов спектрометра, подбор оптимального состояния образца, а также поиск наилучшего способа фокусировки лазерного пучка в объеме образца. При изучении таких биологических молекул как белки необходимо принимать во внимание факт взаимодействия излучения с веществом, так как при определенных условиях возможно влияние лазерного излучения на белковую структуру, в частности, за счет нагрева (особенно при высоких плотностях мощности пучка). Избежать структурных изменений в молекулах токсинов, сохранив их нативную конформацию, и в то же время получить информативные и достаточно интенсивные спектры КР образцов – важная задача в данного рода исследованиях. Таким образом, задача оптимизации методологии спектроскопии белковых молекул также стояла при выполнении данной работы.

Степень разработанности темы исследования

РС является перспективным инструментом для решения ряда актуальных задач токсикологии. Возможности спектроскопии КР в этой области ранее были реализованы далеко не в полной мере. Так, КР спектры отдельных токсинов были проанализированы в предыдущих исследованиях, но отсутствовала методика, в частности, на основе РС, которая позволяла бы анализировать и сравнивать структуры большого количества токсинов. Разнообразие же последних крайне велико, и многие токсины и их структуры остаются не исследованы. Цельные же яды в предыдущих работах других авторов методом РС практически не исследовались, несмотря на перспективность использования этого метода.

Цели и задачи исследования

Исходя из обозначенного выше, в данной работе были поставлены следующие цели:

- разработка метода структурного анализа белковых и пептидных токсинов животных на основе спектроскопии КР;
- разработка метода анализа и оценки состава цельных змеиных ядов на основе спектроскопии КР.

Для достижения данных целей были поставлены следующие задачи:

- подбор оптимальных условий и приборных параметров снятия спектров КР и запись информативных спектров образцов животных токсинов и цельных ядов змей;
- обработка полученных спектральных данных для извлечения информации о белковой структуре или белковом составе (в случае яда) образцов;
- интерпретация и валидация полученных данных о структуре или составе образцов.

Научная новизна

В ходе выполнения работы, нами впервые получены и проанализированы спектры КР конотоксинов и некоторых других токсинов животных; впервые предложена и применена методика, основанная на РС (совмещенной с методами главных компонент и кластеризации), классификации белковых и пептидных токсинов животных в соответствии с их структурными особенностями. Кроме того, впервые продемонстрирована возможность различения дисульфидных изомеров конотоксинов с помощью РС, а также сравнения структур крайне схожих по аминокислотной последовательности токсинов и белков, подобных токсинам.

Далее, впервые были получены и проанализированы спектры КР цельных ядов различных видов змей; предложена новая методика (основанная на РС, совмещенной с методами главных компонент и кластеризации) анализа и классификации цельных ядов змей в соответствии с родом или семейством змеи. Новым является демонстрация полезности спектроскопии КР для оценки составов ядов.

На основе предложенной методики нами впервые проведены анализ ядов отдельных особей гадюки Никольского, оценка внутривидовых (меж- и внутривидовых) вариаций их ядов, а также сравнение этих вариаций с межвидовыми.

Теоретическая и практическая значимость работы

Полученные в работе результаты могут быть полезны как для теоретической (фундаментальной) науки, так и для прикладных разработок. Предлагаемый метод, на основе РС и математических методов обработки спектров, отличается крайне небольшим расходом

сухих образцов и может помочь в изучении токсинов и ядов редких и/или находящихся под угрозой исчезновения животных (их отдельных особей (в частности, змей)).

Можно говорить об этом методе как о дополнительном инструменте экспресс-анализа новых токсинов с целью выявления структурно-схожих и потенциально наиболее перспективных кандидатов на роль лекарств. В то же время, продемонстрированная способность метода различать конфигурации дисульфидных связей позволит обеспечить синтез правильного, аналогичного нативному, изомера токсина. Это важно, ведь синтез искусственных аналогов токсинов необходим для получения их в необходимых, в том числе для клинических испытаний, количествах. Также продемонстрированная возможность детектировать тонкие структурные различия токсинов и им подобных белков может быть полезна при изучении, в частности, влияния аминокислотных замен (в том числе единичных) на белковые структуры. Подобные замены часто вносятся в белки и пептиды при разработке препаратов на их основе с целью оптимизации лекарственных свойств.

Анализ составов ядов змей может быть полезен, в первую очередь, при разработке эффективных противоядий. Предложенный способ спектроскопического исследования ядов – первый шаг в разработке новых методов диагностики жертв укусов методами медицинской оптики, учитывающих тонкие особенности составов ядов, обусловленные меж- и внутривидовыми вариациями ядов. Кроме того, существует также биотехнологическая и криминалистическая задача определения вида змеи или места ее обитания (места отлова) путем анализа состава яда в связи с проблемой нелегальной перевозки и торговли ядами.

Помимо практической значимости, описанная методика может быть востребована в эволюционной биологии и биологической систематике. В частности, установление вариаций ядов змей может пролить свет на процесс видообразования и эволюции. Различие в составах ядов может стать решающим фактором для выделения нового вида или подвида змеи. Кроме того, вопрос эволюции отдельных компонентов ядов – токсинов, а также подобных токсинам белков (пептидов), может быть крайне интересен для фундаментальной науки. Например, эти вещества, будучи различного происхождения, могут иметь общие структурные элементы и свойства.

Методология и методы исследования

Основные результаты работы были получены с использованием спектроскопии КР и методов математической обработки данных. Последние включали в себя следующие основные методики: метод главных компонент (МГК), k-means кластеризацию, анализ дискриминантных функций (discriminant function analysis, DFA).

Экспериментальная часть работы включала в себя подготовку образцов, если последние предоставлялись в виде растворов (в данном случае образцы высушивались на воздухе), и регистрацию спектров КР. Спектры получали с сухих образцов. Токсины и яды были любезно предоставлены коллегами из Института биоорганической химии РАН. Запись спектров КР выполнялась на рамановском микроскопе Bruker Senterra II с использованием лазера с длиной волны 532 нм. Частью экспериментальной работы также являлся поиск оптимальных условий записи спектров образцов.

Перед интерпретацией спектров КР или применением к ним МГК спектральные данные преобразовывались с целью унификации спектров. При необходимости проведения декомпозиции спектральной полосы применялась ее аппроксимация кривыми (curve-fitting).

Положения, выносимые на защиту

1. Спектроскопия КР в сочетании с МГК позволяет проводить группировку белковых токсинов животных в соответствии с их 3D-структурой;
2. Спектроскопия КР позволяет различать дисульфидные изомеры конотоксинов;
3. Спектроскопия КР в сочетании с МГК позволяет проводить группировку цельных змеиных ядов в соответствии с их составом;
4. Спектроскопия КР ядов змей в сочетании с МГК и анализом дискриминантных функций позволяет различать семейства соответствующих змей;
5. Спектроскопия КР в сочетании с МГК позволяет оценивать внутривидовые вариации яда отдельных особей змей и составы яда отдельных особей.

Личное участие автора

Автор лично получал спектры КР образцов, обрабатывал и интерпретировал данные; ставил основные цели и задачи, готовил материалы к публикациям совместно с соавторами. Доклады на конференциях были сделаны лично автором.

Степень достоверности и апробация результатов

Достоверность полученных результатов обеспечивается за счет использования современной и качественной приборной базы, а также воспроизводимостью экспериментальных данных. Материалы, изложенные в диссертации, представлялись на 6 докладах лично автором. По теме диссертации опубликовано 3 статьи в рецензируемых журналах, входящих в базы Web of Science и Scopus.

Публикации автора по теме диссертации

1. Toxins' classification through Raman spectroscopy with principal component analysis / V.A. Mozhaeva, D.S. Kudryavtsev, K.A. Prokhorov [et al.] // *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*. – 2022. – Vol. 278. – P. 121276. – doi: 10.1016/j.saa.2022.121276.
2. Differentiation of snake venom using Raman spectroscopic analysis / V.A. Mozhaeva, V.G. Starkov, D.S. Kudryavtsev [et al.] // *Journal of Materials Chemistry B*. – 2023. – Vol. 11. – P. 6435–6442. – doi: 10.1039/d3tb00829k.
3. Comparison of Conformations and Interactions with Nicotinic Acetylcholine Receptors for E. coli-Produced and Synthetic Three-Finger Protein SLURP-1 / V. Kost, D. Sukhov, I. Ivanov, I. Kasheverov, L. Ojomoko, I. Shelukhina, V. Mozhaeva, D. Kudryavtsev, A. Feofanov, A. Ignatova, Y. Utkin, V. Tsetlin // *International Journal of Molecular Sciences*. – 2023. – Vol. 24. – P. 16950. – doi: 10.3390/ijms242316950.

Тезисы докладов

1. «Классификация белковых (пептидных) лигандов никотиновых ацетилхолиновых рецепторов с использованием рамановской спектроскопии и РСА» // Школа-конференция молодых ученых «Прохоровские недели» (19–21 октября 2021 г., ИОФ РАН), устный доклад;
2. «Классификация токсинов из животных ядов посредством рамановской спектроскопии и метода главных компонент» // XXXIV международная зимняя молодёжная научная школа (8–11 февраля 2022 г., ИБХ РАН), устный доклад;
3. «Анализ изомеров пептидных токсинов с помощью рамановской спектроскопии и метода главных компонент» // IX Всероссийская научная молодежная школа-конференция «Химия, физика, биология: пути интеграции» (20–22 апреля 2022 г., ФИЦ ХФ РАН), устный доклад;
4. «Raman spectroscopy as a tool for animal toxins investigation», 21st World Congress of the International Society on Toxinology (16–20 октября 2022 г.), стендовый доклад;
5. «Классификация змеиных ядов с помощью рамановской спектроскопии», XIV Всероссийская школа по лазерной физике и лазерным технологиям (18–22 сентября 2023 г., Национальный центр физики и математики), стендовый доклад;
6. «Классификация змеиных ядов на основе рамановской спектроскопии», Школа-конференция молодых ученых ИОФ РАН «Прохоровские недели» (24–26 октября 2023 г.), устный доклад.

Глава 1. Обзор литературы

В этой главе рассмотрены наиболее важные с точки зрения темы диссертации литературные и опытные данные, касающиеся прикладных и теоретических аспектов метода рамановской спектроскопии (РС) применительно к исследованию белковых молекул; проведено сравнение этого метода с альтернативными, а также описан ряд математических методик (таких как метод главных компонент (МГК) и методы кластеризации) обработки соответствующих спектральных данных. Приведены примеры применения МГК к данным РС. Дополнительно подробно рассмотрены некоторые животные токсины и яды и примеры их анализа методом РС.

1.1. Метод рамановской спектроскопии в исследовании белковых молекул

1.1.1. Общие сведения

Во время одного из своих путешествий Ч. В. Раман (Нобелевская премия 1930 года) был очарован глубоким синим цветом Средиземного моря. Позже он объяснит, что этот цвет был результатом рассеяния солнечного света молекулами воды, по аналогии с голубым цветом неба, обусловленным рэлеевским рассеянием, которое возникает, когда молекулы воздуха упруго рассеивают солнечный свет. Но Раман также обнаружил, что небольшая часть рассеянного света имеет длины волн, отличные от исходных длин волн солнечного света, что представляет собой неупругое рассеяние [20]. Вообще, в первой трети XX века несколько ученых из разных стран мира, независимо друг от друга, искали в рассеянном свете излучение с измененной частотой. Работы велись как в Индии (Ч.В. Раманом и К.С. Кришнаном), так и в России (Г.С. Ландсбергом и Л.И. Мандельштамом), и Франции (Ж. Кабанном, П. Дором и И. Рокаром) [21].

Сегодня рамановская спектроскопия (РС) или спектроскопия комбинационного рассеяния (КР) света представляет собой высокоинформативный метод исследования различных молекул. Как было сказано во введении, спектр КР содержит информацию о частотах колебаний атомов или их групп в молекуле. Эти частоты отражают не только внутримолекулярные силы, характеризующие состав и структуру вещества, но также и внемолекулярные силы (обусловленные электростатическими, дипольными или гидрофильными/гидрофобными взаимодействиями), которые содержат информацию об окружении молекулы, что важно при изучении молекулярных взаимодействий.

Теория РС подробно изложена во многих источниках [22, 23], здесь приведятся лишь ее основные положения. При взаимодействии лазерного излучения с молекулами бóльшая его часть рассеивается без изменения частоты. Данное рассеивание, при котором частота рассеянного излучения равна частоте падающего, называют релеевским, оно является упругим. Но некоторая доля рассеянного света ($\sim 10^{-7}$) сдвигается по частоте в сторону бóльших (антистоксовы линии) или меньших частот (стоксовы линии). Появление стоксова излучения гораздо более вероятно, чем антистоксова, и, следовательно, стоксовы линии более интенсивны. Причина этого состоит в том, что молекула вещества может потерять (отдать) часть своей энергии или же поглотить часть энергии падающего излучения. Потеря или выигрыш энергии (относительно падающего излучения) равны разнице энергий между двумя колебательными уровнями энергии молекулы. Разность энергий зависит от массы атомных ядер и силы связи между атомами, участвующими в колебательном движении. Соответственно, величины частотных сдвигов излучения при рассеянии (так называемый рамановский сдвиг, который принято измерять в см^{-1}) содержат крайне полезную информацию о молекулах. Рамановское рассеяние является нерезонансным, излучение происходит с виртуального уровня; таким образом, эффект не предполагает поглощения, а является результатом рассеяния. То есть, процесс полностью отличается, в частности, от поглощения инфракрасного излучения (ИК), и соответственно, РС во многом отлична от ИК-спектроскопии. При КР свет взаимодействует с молекулой и искажает (поляризует) облако электронов вокруг ядер, образуя кратковременное состояние, называемое «виртуальным состоянием» молекулы. Это взаимодействие можно рассматривать как образование очень недолговечного «комплекса» между светом и молекулой. Так как это состояние нестабильно, свет быстро высвобождается, и фотон переизлучается.

Как было сказано выше, РС позволяет получать информацию о молекулярных колебаниях и, соответственно, извлекать информацию из крайне широкого спектра веществ, в различном агрегатном состоянии, неорганических и органических. Соответственно, область применения этого метода крайне велика [23]. Основным интерес при работе был сосредоточен на органических молекулах, что связывает настоящее исследование, методология которого опирается на РС, с биологией.

Изучение биологических молекул имеет огромное значение как для биологических наук, так и для медицины; одними из самых интересных и сложных биологических молекул являются белки. Они сложны не столько своим составом, сколько строением. Как известно, белки состоят из аминокислот, последовательно прочно соединенных в цепочку (основная цепь). Но, и это принципиально важно, эта цепочка определенным образом укладывается в пространстве за счет более слабых взаимодействий многочисленных ответвлений (боковых цепей). Благодаря этому в структуре белка могут формироваться спирали (так называемые α -спирали) и слои (β -слои).

Сочетание подобных структурных элементов (которые принято называть вторичной структурой) под влиянием окружающей среды формирует общий облик белка – его укладку в пространстве (третичная структура). Иногда за счет дисульфидных сшивок (мостов) S–S в белках образуются петли. Отметим, что принято различать длинные (тяжелые) белковые молекулы (собственно, белки) и короткие (легкие), называемые пептидами. Если аминокислотная цепь состоит менее чем из 50 остатков, то принято говорить о пептиде.

Таким образом, можно характеризовать белки с трех различных сторон: 1) анализировать их первичную структуру (то есть, аминокислотную последовательность), 2) рассматривать их пространственную укладку, то есть 3D структуру (вторичную и третичную) и 3) анализировать взаимодействие белков с другими молекулами. Пространственная форма белка во многом определяется его первичной структурой, но в то же время немаловажное значение имеет окружение белковой молекулы (свойства среды или растворителя). Так, белок с одной и той же аминокислотной последовательностью может быть уложен по-разному (иметь различные конформации) в зависимости от окружающей его среды или других факторов. РС дает возможность исследования белков с этих трех разных сторон.

РС спектр белка можно представить как суперпозицию спектров составляющих его аминокислот плюс некоторая добавка Δ . В принципе, можно получить примерный спектр белка, сложив спектры отдельных аминокислот его составляющих [24]; можно утверждать, что в спектре будет доминировать вклад от ароматических аминокислот [25]. Но, и это представляет для нас особую ценность, в спектре КР белковой молекулы содержится также информация (выраженная в виде Δ во введенной нами выше условной формуле) о слабых внутримолекулярных взаимодействиях в белке и с белковым окружением (внемолекулярных взаимодействиях), что позволяет изучать с помощью РС структуру и конформацию белка, а также его взаимодействия с другими молекулами.

Чтобы извлечь эту уникальную информацию, необходимо отсеять большую часть света, рассеянного на образце без изменения частоты (релеевское рассеяние) и анализировать излучение, сдвинутое по частоте. Любая паразитная засветка этой области не позволит разрешить рамановский сигнал. Отсюда вытекает ряд ограничений этого метода.

Так как превалирующая часть излучения возбуждающего лазера отсекается, то итоговый полезный сигнал оказывается низко интенсивным. Поэтому данный метод не позволяет работать с низкими концентрациями вещества в растворе. В то же время, при работе с сухими веществами, образца требуется совсем не много (а именно столько, сколько необходимо для заполнения исследуемым веществом площади лазерного пятна, сфокусированного на образце; обычно порядка мкг). В данном случае, единичное измерение занимает малое время (несколько минут); простота пробоподготовки (высушивание) или вовсе ее отсутствие (если образец

изначально представлен в виде порошка) также экономит время и обеспечивает чистоту метода. Отметим, что существуют модификации обычной рамановской спектроскопии, которые позволяют усилить рамановский сигнал и работать с меньшими концентрациями веществ (поверхностно-усиленная РС (surface-enhanced Raman spectroscopy, SERS), резонансная РС). Об этом будет сказано позже, а пока речь пойдет об обычной РС.

Как и многие другие оптические методы исследований, РС является неразрушающим методом: воздействие излучением на образец не меняет свойств и структуры последнего. Это справедливо при использовании подходящих интенсивностей облучения. (Предельные интенсивности, которые способен выдержать конкретный образец, определяются экспериментально; также играет роль время облучения.) Указанные выше плюсы и минусы метода РС суммированы в таблице 1.

Таблица 1.1. Положительные (+) и отрицательные (–) стороны метода РС.

+	–
Простота пробоподготовки	Низкая чувствительность (невозможность работы с низкими концентрациями вещества)
Быстрота	Иногда значительный фон флуоресценции, фотолюминесценции
Информативность	Ограничения в использовании Раман-активных растворителей
Экономичное (при использовании сухих образцов) использование исследуемого вещества	
Неразрушающий («non-destructive») метод	

Ключевые соображения по разработке и реализации РС экспериментов обсуждаются ниже.

1.1.2. Методические вопросы

Образец для обычной РС может быть измерен в жидком или сухом виде. В жидком виде он может быть помещен в кювету или нанесен на подложку в виде капли. Во втором случае,

требуется меньшее количество образца, но происходит испарение капли, что ведет к увеличению концентрации вещества, что необходимо принимать во внимание при работе. Испарение капли также ограничивает возможное время выдержки при записи спектра.

В сухом виде образец может быть получен путем лиофилизации или же высушиванием капли его раствора (на воздухе). Во втором случае пространственное распределение образца зачастую неоднородно: белок может концентрироваться на границе высохшей капли, что составляет так называемый эффект кофейных колец [25]. Этот эффект позволяет отделить белки от растворителя и обеспечить их повышенную концентрацию на границе капли (см. Рисунок 5.1 главы 5). Известно, что в некоторых случаях, при высушивании смеси нескольких белков, различные белки локализуются на разных расстояниях от центра капли или по разным сторонам капли, что обеспечивает их пространственное разделение [26]. Эффект колец может быть полезен, если необходимо анализировать спектры отдельных белков смеси или, например, разделить различные по размеру агрегаты белков.

Если же целью является получение общего спектра белковой смеси (грубо говоря, сумму спектров компонентов), то предпочтительно использовать лиофилизированный образец. В этом случае в область сбора сигнала попадает большое количество различных молекул (диаметр лазерного пятна много больше размера молекулы), и обеспечивается получение усредненного по всем компонентам спектра. В отдельных случаях белок может быть представлен в виде кристалла, но получение и изучение белковых кристаллов составляет собой отдельный класс задач. Так как кристаллизация белков сама по себе является непростой задачей, то в случае успеха имеет смысл изучить белок более информативным методом рентгеноструктурного анализа (см. ниже).

Так как жидкость – естественная среда для белковых молекул (учитывая их функционирование в биологических системах), то при изучении их структуры предпочтительно, на первый взгляд, анализировать спектры растворов, нежели сухих образцов. Но, как правило, необходимо использовать высококонцентрированные растворы белков, что обычно трудно обеспечить: количество белка зачастую крайне ограничено. Вдобавок, некоторые буферные растворы, в которых может храниться белок, Раман-активны, и их спектр может перекрывать или накладываться на спектр исследуемого растворенного белка. Таким образом, использование сухих образцов позволяет свести некоторые отмеченные выше минусы к минимуму. Работа с сухими образцами обеспечивает крайнюю экономичность их использования, что зачастую критически важно. Стоит принимать во внимание, что «сухой» образец может содержать связанную воду, молекулы которой могут удерживаться, например, за счет взаимодействия с гидрофильными аминокислотами (образуя гидратную оболочку); остаточная влажность в том или ином виде почти всегда присутствует. Ряд исследований,

например [27], позволяет утверждать, что высушивание белков не приводит к значимым изменениям их рамановских спектров и избавляет от сложностей, связанных с работой с растворами.

Таблица 1.2. Положительные (+) и отрицательные (–) стороны разных способов пробоподготовки.

Тип образца	Сухой образец			Жидкость	
	Лиофилизирован	Высушен на воздухе	Кристаллизован	В кювете	В капле
+	Экономичность, однородность	Возможно реализовать пространственное разделение различных белков, экономичность	Упорядоченность	Постоянность концентрации, обеспечение жидкой среды	Обеспечение жидкой среды
–	Необходимо контролировать возможное влияние высушивания на структуру белка	Неоднородность многокомпонентных смесей, необходимо контролировать возможное влияние высушивания на структуру белка	Сложность проведения кристаллизации, времязатратность	Требуется значительное количество белка	Требуется высокая концентрация белка, непостоянность концентрации

Вопрос о получении максимально информативного спектра включает поиск оптимальных условий съемки. Они включают в себя, с одной стороны, параметры прибора, а с другой – параметры образца (в том числе тип подложки).

Вопрос об оптимальном состоянии образца обсуждался выше; что касается типа подложки, то предпочтительно, чтобы она вносила минимальный вклад в спектр аналита. Так, кварцевое стекло (SiO_2) имеет интенсивный пик в районе 450 см^{-1} , а также минорные пики на больших частотах [28], поэтому при измерении вблизи его поверхности влияние спектра стекла на спектр образца неизбежно. Учесть это влияние бывает непросто. Поэтому предпочтительней

использовать, например, подложки из латуни или фторида кальция, которые не вносят вклад в информационно-значимую часть спектра образца.

Принципиальная схема рамановского микроскопа дана на рисунке 1.1. Излучение лазера, мощность и генерируемая длина волны которого зависят от типа лазера, фокусируется через объектив на образец. Собранное (в течение некоторого времени) рассеянное образцом лазерное излучение фильтруется; при этом удаляется излучение, рассеянное без изменения частоты. Фильтрованный пучок направляется в щель определенного размера и далее на дифракционную решетку, которая разделяет излучение разных частот. Интенсивность излучения на каждой частоте (с определенным шагом) измеряется посредством детектора.

Условия съемки спектра сочетают в себе ряд параметров (перечислены те параметры, которые обычно подлежат определенной регулировке (предусмотренной эксплуатацией прибора) в зависимости от выбора оператора):

- а) длина волны (частота) лазерного возбуждающего излучения;
- б) мощность лазера;
- в) параметры объектива (в частности, числовая апертура NA);
- г) время накопления спектра;
- д) спектральное разрешение.

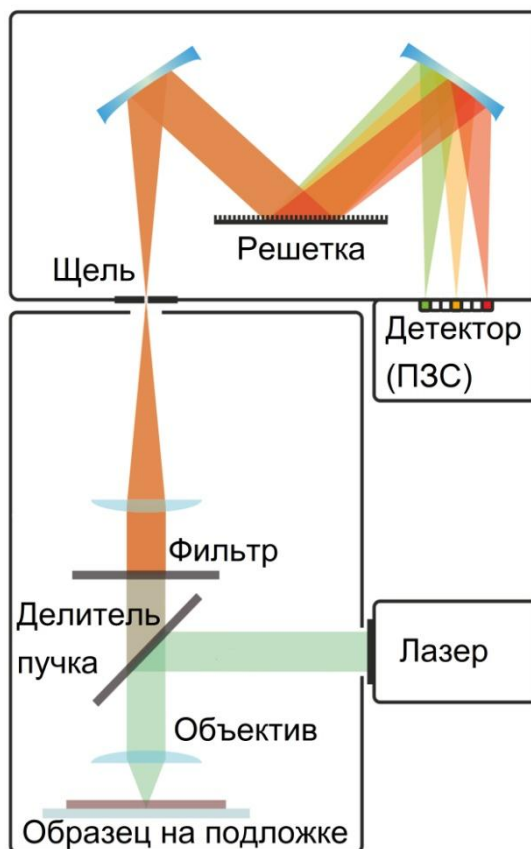


Рисунок 1.1 – Принципиальная схема рамановского микроскопа [29].

Плотность мощности (I) представляет собой отношение мощности в Вт (W) к площади поперечного сечения лазерного пятна ($I=W/cm^2$) [30]. Диаметр же сечения обратно пропорционален NA объектива. Соответственно, чем больше NA , тем меньше размер лазерного пятна и тем больше плотность мощности.

Эффективность КР пропорциональна четвертой степени частоты возбуждения ω_{exc}^4 . Следовательно, можно ожидать получения гораздо более интенсивного сигнала КР от образца при использовании большей частоты возбуждающего излучения. Что касается глубины проникновения света, то в общем случае, чем больше длина волны возбуждения (меньше частота), тем глубже свет проникает в образец. Если же образец непрозрачен, то глубина проникновения будет определяться поглощательной способностью аналита на данной длине волны. В случае ограниченного в объеме однородного образца (например, лиофилизированной «крупинки» белка) нет нужды в большой глубине проникновения.

Куда более значимым фактором, зависящем от длины волны возбуждения, является фотолюминесцентный фон, проявляющийся в спектре образца. Даже если основное вещество образца само по себе не излучает, то иногда примеси (загрязнения) в следовых количествах могут вызвать значительную фотолюминесценцию. Как правило, использование большей длины волны (например, возбуждения в ИК-диапазоне) ведет к уменьшению фона. В общем случае, для уменьшения фона фотолюминесценции следует использовать ту длину волны, для которой поглощательная способность исследуемого вещества (и, следовательно, итоговая фотолюминесценция) ниже [31].

Оптимальная же длина волны определяется как характерной для нее эффективностью КР, так и получаемым уровнем фона. При возможности имеет смысл испробовать несколько длин волн лазера и сравнить получаемые спектры, выбрав в результате спектр с бóльшим отношением сигнал/шум.

Увеличение плотности мощности ведет к увеличению интенсивности сигнала КР, но чревато повреждением образца. При достижении определенной величины мощности образец начинает гореть. Поэтому и здесь важно подобрать оптимальную величину данного параметра, что обычно делается опытным путем. Необходимо определить предельную величину мощности, при которой еще не происходит структурных изменений в образце (например, вследствие нагрева), но которая обеспечивает достаточно интенсивный сигнал.

К нагреву и возгоранию образца также может приводить большое время накопления спектра (особенно, если плотность мощности близка к предельной). При этом увеличением времени накопления можно скомпенсировать низкую эффективность КР или/и плотность мощности.

Далее, чем больше размер установленной в приборе щели, тем выше пропускная способность и, соответственно, интенсивность сигнала КР. Однако уменьшение размера щели позволяет увеличить спектральное разрешение, например, при этом плечи пиков становятся более четко выражены и разрешаются как отдельные пики. Отметим, что, например, лазер с длиной волны 532 нм даст, как говорилось выше, гораздо более интенсивный рамановский сигнал, чем лазер с длиной волны 785 нм, поэтому с зеленым лазером можно использовать щель меньшего размера, обеспечивая тем самым повышенное спектральное разрешение.

Еще один элемент конструкции рамановского спектрометра, смена которого предусмотрена во многих приборах – это дифракционная решетка. Чем выше плотность штрихов решетки (число штрихов на мм), тем лучше спектральное разрешение. Но, при использовании решетки с меньшим числом штрихов/мм, можно получить спектр в более широком диапазоне частот. Так как большинство рамановских спектрометров имеют несколько решеток для оптимизации съемки во всем спектральном диапазоне, то часто бывает полезным вначале использовать одну решетку, охватывающую широкий спектральный диапазон, а затем другую, для исследования с повышенным разрешением узкого диапазона спектра, который привлеч внимание или интересует исследователя.

Как это видно, отмеченные факторы оказываются во взаимосвязи между собой и важно найти оптимальную их комбинацию. Необходимо также отметить, что, в частности, сухой белок (или смесь белковых молекул), как правило, представляет собой неоднородный по объему пористый образец, что делает крайне важной задачу правильной фокусировки, а также дальнейшей унификации различных спектров. Подводя итог, путем оптимизации различных методических аспектов (а именно подбором условий съемки и формы образца), можно постараться свести минусы метода РС к минимуму.

1.1.3. Спектры КР белков

В исследованиях белков методом РС можно условно выделить следующие направления.

1) Прямой или с применением компьютерных алгоритмов обработки (методов понижения размерности данных, кластеризации и других) анализ структур белков и их сравнение. Это могут быть различные (из разных семейств) или родственные, близкие по строению белки (или даже разные конформеры одного и того же белка), в структурах которых могут быть исследованы тонкие различия. Именно этого направление освещается в данной работе; при этом, наряду с чистыми белками рассматриваются их смеси.

2) Детектирование конкретных белков (например, являющихся маркерами заболевания) в различных средах (например, жидкостях или тканях). Подобные работы довольно часто проводятся с помощью метода SERS, посредством которого стараются поднять чувствительность и иногда обеспечить некоторую специфичность к необходимому белку.

3) Детектирование белковых взаимодействий посредством анализа спектральных изменений, вызванных взаимодействием белка с какой-либо молекулой.

Поскольку белки зачастую состоят из сотен аминокислот, их спектры КР содержат сложный набор перекрывающихся полос. Однако сигналы КР ароматических аминокислот и белковой (полипептидной) цепи по своей природе являются сильными и приводят к появлению набора доминирующих пиков в спектре. Различные спектральные регионы (маркерные полосы) характеризуют как колебания 1) основной цепи белка, так и 2) боковых цепей. Например, район $1640\text{--}1689\text{ см}^{-1}$ представляет собой так называемую полосу Амид I (Amide I), которая включает колебания основной цепи белка, доминирующий вклад в которые вносит колебание $\text{C}=\text{O}$ карбонильных групп. Карбонильные группы в белке действуют как акцепторы водородных связей внутри упорядоченной вторичной структуры, в результате чего положение максимума (и положения плечей) полосы Amide I сильно зависит от вторичной структуры белка. Другая полоса $1230\text{--}1300\text{ см}^{-1}$, Амид III (Amide III), также характеризует колебания основной цепи, которые в данном случае связаны с движениями в $\text{C}\text{--}\text{N}$ и $\text{N}\text{--}\text{H}$ [25]. Интенсивные полосы также характерны для боковых цепей ароматических аминокислот, и окружение, в частности, тирозина и триптофана отражается на форме характерных пиков, что можно использовать для характеристики белковой структуры. Содержащие серу (S) аминокислоты также имеют характерные пики.

Прямой анализ, как правило, включает рассмотрение «маркерных», наиболее информативных участков спектра (Рисунок 1.2), которые перечислены ниже.

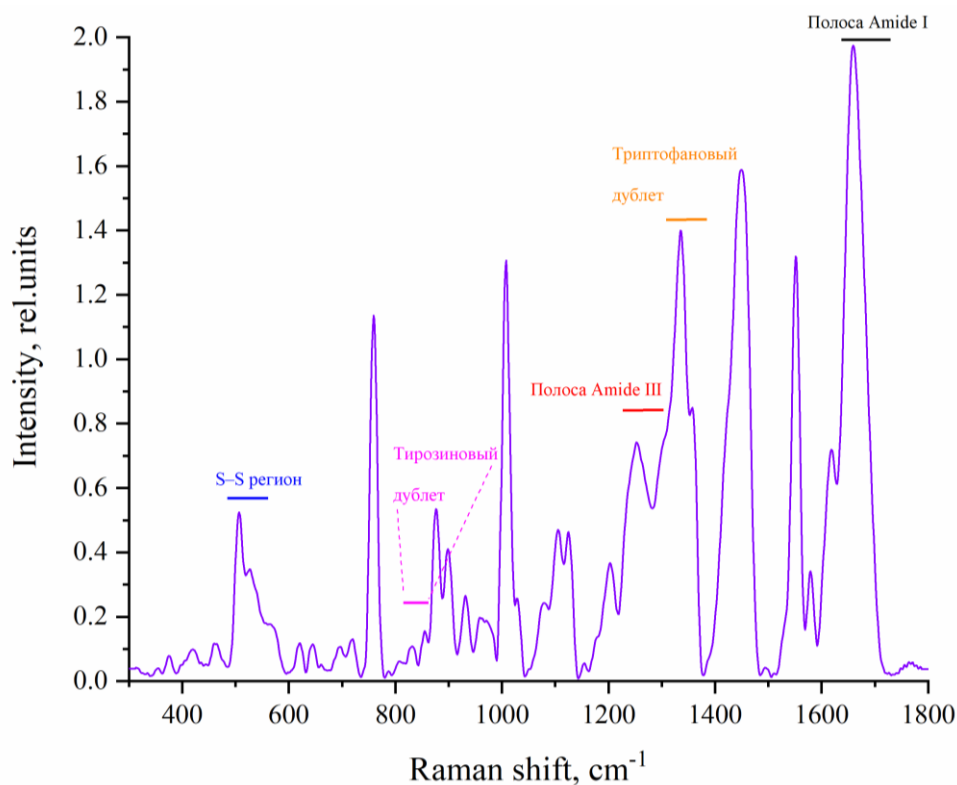


Рисунок 1.2 – Типичный спектр КР белка (лизоцима куриного яйца) с отмеченными маркерными регионами. Подписи осей (здесь и далее для спектров КР) соответствуют следующим русскоязычным обозначениям: ось абсцисс – Рамановский сдвиг, cm^{-1} (Raman shift, cm^{-1}); ось ординат – Интенсивность, отн.ед. (Intensity, rel.units).

1. Крайне информативным является тирозиновый дублет – пара пиков тирозина (около 850 и 830 cm^{-1}), отношение интенсивностей ($I_{850/830}$) которых используют для определения его микроокружения. Если остатки тирозина находятся на поверхности белка и контактируют с растворителем, то величина $I_{850/830}$ (приблизительно) больше единицы, если же тирозин «зарыт» внутри молекулы белка, то меньше единицы [32].

2. Окружение остатков триптофана может быть оценено с помощью триптофанового дублета 1360 и 1340 cm^{-1} : относительная интенсивность этих полос $I_{1360/1340}$ имеет большую величину при гидрофобном окружении остатков триптофана, нежели чем при гидрофильном [33]. Другими словами, отсутствие в спектрах пика на 1360 cm^{-1} предполагает, что индольное кольцо доступно для молекулы воды, а появление этого характерного пика указывает на то, что оно может участвовать во взаимодействии внутри молекулы.

Отметим, что если в белке присутствует несколько остатков тирозина/триптофана, то использование характерных дублетов позволит оценить окружение всех этих остатков в среднем. Знание микроокружения этих аминокислот может пролить свет на пространственную укладку белка, а также стабильность данной конформации, так как экспонированность

ароматических остатков на поверхности энергетически не выгодна и может вносить отрицательный вклад, в частности, в термостабильность белковой молекулы.

3. Регион $500\text{--}550\text{ см}^{-1}$ характеризует растяжение дисульфидных связей и используется для характеристики конформации мостиков S–S и ее геометрии [34]. Большинство природных белков и пептидов, содержащих дисульфидные связи, имеют в спектре КР пик, центрированный около 510 см^{-1} . Это соответствует конформации дисульфидной связи с наименьшей потенциальной энергией, то есть *gauche-gauche-gauche* конформации C–C–S–S–C–C. Дополнительные полосы, центрированные около 525 и 540 см^{-1} , соответствуют конформациям *gauche-gauche-trans* и *trans-gauche-trans* соответственно [32].

4. Выводы о характере вторичной структуры белка обычно делаются на основе рассмотрения полос Amide I ($1600\text{--}1690\text{ см}^{-1}$) и Amide III ($1230\text{--}1300\text{ см}^{-1}$) [35]. Часто Amide I полосу рассматривают как суперпозицию кривых Гаусса (или Лоренца), каждая из которых отражает влияние α -, β - и нерегулярной структур на спектр и их содержание в белке. Эти типы вторичных структур имеют полосы, центрированные около $1650\text{--}1660\text{ см}^{-1}$ (α -спирали), $1668\text{--}1677\text{ см}^{-1}$ (β -слои) и $1661\text{--}1667\text{ см}^{-1}$ или $1678\text{--}1689\text{ см}^{-1}$ (нерегулярная структура). Появление нерегулярной структуры наблюдалось разными авторами как на промежуточных частотах (между пиков, характерных для α - и β -структур) ($1661\text{--}1667\text{ см}^{-1}$) [32], так и в области более высоких частот ($1678\text{--}1689\text{ см}^{-1}$) [27]. В первую очередь при анализе вторичной структуры следует рассматривать полосу Amide I и полученные выводы подтверждать анализом Amide III полосы. Необходимо иметь в виду, что в состав этих полос могут входить пики, характерные для ароматических аминокислот, что нужно учитывать при анализе. Также некоторый вклад в Amide I может давать вода, что выражается в пике, отражающем колебание НОН около 1640 см^{-1} [36].

Отметим также, что колебание сульфгидрильной группы цистеина (SH) приводит к появлению хорошо разрешаемой полосы спектра КР в области $2500\text{--}2600\text{ см}^{-1}$ [37], которая оказывается крайне полезной для мониторинга разомкнутости дисульфидов.

Рассмотренные выше полосы можно назвать конформационно восприимчивыми и, таким образом, их рассмотрение особенно полезно при характеристике структуры и конформации белка.

1.1.4. SERS и резонансная РС

Поверхностно-усиленная РС (SERS) связана со значительным усилением сигналов КР аналитов, расположенных вблизи усиливающих сигнал материалов. Усиление рамановских

сигналов, наблюдаемое в SERS, принято объяснять главным образом электромагнитным усилением с некоторым вкладом химического усиления. Электромагнитное усиление возникает за счет электромагнитных полей, локализованных на поверхности усиливающего субстрата (явление, так называемого, локализованного поверхностного плазмонного резонанса (ППР)). Химическое усиление происходит в результате переноса заряда между молекулами аналита и усиливающим субстратом, что приводит к изменению поляризуемости молекулы и усилению ее сигнала КР. Наиболее широко применяемыми и эффективными усиливающими материалами являются плазмонные субстраты на основе Ag и Au. При этом плазмонный резонанс на поверхности структур нано-масштаба полностью отличается от такового в объемном металле, обеспечивая возможность усиливать рассеянный свет. В настоящее время, после нескольких десятилетий исследований и развития нанотехнологий, SERS-подложки эволюционировали от грубых шероховатых электродов к наноматериалам с индивидуальной морфологией: примеры таких материалов включают литографически вытравленные наноструктуры, коллоидные наночастицы и их сборки [38].

В ряде работ для характеристики белковых молекул использовалась SERS. Например, SERS в сочетании с методом главных компонент (МПК) использовали для классификации белков в соответствии с их вторичной структурой [39]. Однако, метод SERS имеет низкую воспроизводимость, что ограничивает его использование. Спектры одного и того же образца на разных наноструктурированных подложках скорее всего будут разными (вероятно, даже если подложки изготовлены по одной методике, но в разных лабораториях). Кроме того, усиленные спектры образца могут различаться в зависимости от его ориентации на подложке. Вдобавок, само изготовление подложки является отдельной задачей. В случае недостаточной воспроизводимости необходимо проводить обработку большого объема данных (например, полученных при сканировании по площади SERS-субстрата).

Низкая воспроизводимость SERS может быть следствием неоднородности субстрата (разного размера наночастиц или их неравномерного распределения). Но даже имея в распоряжении качественную однородную SERS подложку, добиться повторяемости при повторных измерениях одного и того же вещества зачастую не удастся, так как его молекулы принимают случайную ориентацию на усиливающей поверхности. Плохая воспроизводимость SERS-спектров, связанная со случайной ориентацией адсорбированных белков, подчеркивается в работе [40]. Это является следствием того факта, что (как было отмечено) только сигнал от тех молекул, которые находятся близко к поверхности подложки, может быть усилен, так как в основе SERS – усиленное электрическое поле, обеспечиваемое наночастицами благородных металлов, и это поле быстро затухает при удалении от их поверхности. Отсюда также вытекает

необходимость приближения исследуемой молекулы на достаточно близкое расстояние к SERS-субстрату.

Вообще говоря, случайная ориентация молекул ведет к тому, что записанный SERS-спектр образуется суперпозицией спектров молекулы в различных ее ориентациях, что усложняет интерпретацию. Для решения этой проблемы необходимо обеспечить неслучайность ориентации исследуемого вещества. Для этого его структуру определенным образом модифицируют, повышая аффинность конкретного участка к субстрату (например, «пришивают» к белку соединения, содержащие серу (к примеру, цистеин), которая имеет повышенное сродство к SERS-субстрату [41]). Вследствие этого подавляющее большинство молекул ориентируется на SERS-подложке одинаковым образом. Авторы же работы [40] предложили контролировать иммобилизацию белка путем использования спейсерных молекул (spacer molecules), которые связаны не с аналитом, а с SERS-субстратом (посредством атома серы) и обладают способностью связывать конкретный участок исследуемой молекулы. Так как размер подобной спейсерной молекулы ограничен размером области усиления, то использование, например, крупных антител в данном случае невозможно. Необходимость модификации структуры вещества или подложки является недостатком подобных методик.

Отметим еще один интересный момент: интенсивность SERS определяется не только расстоянием от молекулы до нано-элементов подложки, но и некоторыми правилами отбора (усиливаются режимы колебаний, определенным образом ориентированные к поверхности нано-элементов) [42].

В итоге, необходимость получения качественного (изотропного по поверхности) SERS субстрата (в случае, если таковой изготавливается исследователем самостоятельно), а также необходимость химических модификаций лишают метод SERS таких преимуществ классической РС, как простота пробоподготовки и малая времязатратность. Анализ структур белков этим методом не оправдан, так как итоговый спектр в первую очередь несет информацию о взаимодействии белок-наноструктура, нежели о самом белке. Но SERS находит свое применение в детектировании конкретных белков, в первую очередь, благодаря преимуществу в чувствительности (за счет обеспечения усиления сигнала).

Еще одна модификация метода РС основана на эффекте резонанса. Если энергия возбуждающего излучения соответствует электронному переходу интересующей молекулы, это приводит к значительному усилению рассеяния света. В таком случае принято говорить о резонансной РС [43]. Как правило, для реализации данного эффекта в исследуемое вещество вводится так называемый рамановский репортер (зонд), сигнал которого подвергается усилению. Например, в работе [44] рамановский зонд обеспечивал крайне интенсивный сигнал за счет резонансного эффекта, который превалировал над сигналом от других молекул и был

различим на их фоне. Полезность предложенных зондов была подтверждена возможностью на базе РС детекции и мониторинга лизосом в живых клетках. Надо иметь в виду, что обычные флуоресцентные репортеры, в большинстве случаев, не подходят для РС из-за сильного фона флуоресценции, маскирующего сигнал КР. Поэтому для успешной реализации данной техники, как правило, необходимы специальная разработка и синтез зонда.

1.2. Другие методы определения структуры белков

1.2.1. Рентгеноструктурный анализ и ядерный магнитный резонанс

В настоящее время основными экспериментальными методами определения пространственной структуры белков являются рентгеноструктурный анализ (РА) и спектроскопия ядерного магнитного резонанса (ЯМР).

При РА белок сначала кристаллизуют, а затем с помощью дифракции рентгеновских лучей определяют структуру белка. Определение трехмерной структуры этим методом зачастую иногда занимает от трех до пяти лет. Природа создавала хорошо сформированные кристаллы на протяжении тысяч или миллионов лет, но органические молекулы и белки редко встречаются в природе в кристаллическом состоянии. Необходимо искать условия, при которых белки кристаллизуются. Например, в 1995 году кристаллизация и определение структуры ВИЧ-интегразы были осуществлены только после 40-й точечной мутации исходного белка, которые вводились с целью так изменить свойства поверхности белка, чтобы могло произойти упорядочение в кристалл [45].

ЯМР-спектроскопия – еще один высокоинформативный метод определения структуры белка. Преимущество ЯМР перед РА заключается в том, что белок можно изучать в его естественной водной среде. Но ЯМР подходит только для небольших белков, содержащих менее 150 аминокислот и требует наличия раствора белка в небольшом объеме, но в высокой концентрации [46].

Учитывая названные ограничения двух методов, их трудоемкость и времязатратность, использование ЯМР и РА должно быть оправдано целями исследователя получить подробную, максимально информативную модель структуры белка. Выполнение анализа этими сложными методами большого количества белков или пептидов с целью оценки их структур не является рациональным. В данном случае более оправдано применение более простых и доступных, но не столь информативных методик, таких как РС или круговой дихроизм (см. ниже). После

проведения подобного первичного исследования и сосредоточения на нескольких, представляющих наибольший интерес, молекулах, можно привлечь ЯМР или РА.

1.2.2. Круговой дихроизм

Круговой дихроизм (КД) определяется как неравномерное поглощение света, поляризованного по правому и левому кругу. Когда асимметричные молекулы взаимодействуют со светом, они могут в разной степени поглощать право- и левополяризованный свет. Как правило, КД используют как метод определения вторичной структуры белков: разные структурные элементы имеют характерные спектры КД [47]. КД, как и РС, дает менее конкретную структурную информацию, чем ЯМР и РА, но это более простой и быстрый метод. Так как КД, также как ИК-спектроскопия, имеет в своей основе иные физические явления, нежели РС, то разная полезная информация может быть получена этими различными оптическими методами.

1.2.3. Предсказание структуры белка

Далеко не все понятно в том, как белок сворачивается в уникальную трехмерную структуру из своей одномерной последовательности. Получить последовательности белков гораздо проще, чем получить их структуры. С развитием передовой технологии секвенирования ДНК последовательности белков стали быстро накапливаться. Базы данных в настоящее время содержат более 85 миллионов белковых последовательностей. Так как РА и ЯМР-спектроскопия (являясь двумя основными экспериментальными методами определения структуры белка) отнимают много времени и рабочей силы, то, например, количество белковых структур в банке данных белков PDB [48] составляет лишь <0,2% от числа белковых последовательностей в UniProt (открытая база данных белковых последовательностей) [49].

Вся информация, необходимая белку для правильного сворачивания, закодирована в его аминокислотной последовательности (так называемая догма Анфинсена), и нативная структура белка соответствует его наиболее стабильной конформации (с наименьшей свободной энергией). Исходя из этих принципов, разумной выглядит идея предсказания структуры белка из его первичной последовательности. Эта идея лежит в основе метода, альтернативного экспериментальным, определения белковых структур путем моделирования. Но хотя известно,

что процесс сворачивания белка регулируется различными физическими законами, очень сложно дать столь сложной макромолекуле, как белок, достаточно точное физическое описание, включающее также ее взаимодействие с окружающей средой.

Подобные последовательности из одного и того же эволюционного семейства часто имеют схожие белковые структуры, что составляет основу моделирования, основанного на гомологии. На данный момент это наиболее точный способ предсказать структуру белка, используя его гомологичную структуру в PDB в качестве матрицы. В отличие от методов моделирования по гомологии, методы *ab initio* направлены на построение белковой структуры на основе принципов физики, без опоры на ранее разрешенные структуры. Однако успешные методы *ab initio* очень редки, и существует еще много проблем, которые предстоит решить [50].

Несмотря на то, что методы моделирования доступны, не требуют сложного оборудования и позволяют получить структуру белка, имея в распоряжении только его последовательность (не требуется даже сам белок), их нельзя рассматривать как полную альтернативу существующим экспериментальным методам, а к структурам, полученным моделированием, нужно относиться с долей критики. В случае исследования белков, последовательности которых значительно отличаются от имеющихся в PDB, вероятность получения корректной структуры значительно снижается.

1.3. Животные токсины и яды

1.3.1. Токсины животных белковой природы

Яды животных включают многочисленные семейства токсинов, представляющих собой в основном пептиды и белки, механизмы действия которых отлажены за миллионы лет эволюции. Эти токсины нацелены на множество различных ионных каналов (которые выполняют самые разнообразные функции в тканях: отвечают за возбудимость нервных клеток, участвуют в обеспечении сердечной деятельности, задействованы при выделении нейромедиаторов, в секреции гормонов [51]), клеточных рецепторов и ферментов. Токсины действуют с высокой эффективностью и зачастую с исключительной селективностью, поэтому привлекают большое внимание для дальнейшего изучения их фармакологических и физиологических свойств. Токсины, синтезируемые как наземными, так и морскими ядовитыми животными, вводятся в организм жертвы для охоты или защиты. Известно, что некоторые ядовитые животные использовались для лечения болезней на протяжении тысячелетий во

многих частях мира. Сегодня, токсины с известными механизмами действия могут служить «шаблонами» при разработке лекарств [52].

Значительная часть белков и пептидов животных ядов содержит множественные дисульфидные связи, что обуславливает большое разнообразие их структур и существенные структурные различия разных токсинов, возникшие в результате длительного селективного давления и процесса коэволюции. К распространенным и хорошо известным ядовитым животным относятся род *Conus* (улитки-конусы), членистоногие (пауки, скорпионы, многоножки, пчелы и др.), позвоночные животные (змеи, ящерицы и др.), а также кишечнополостные (медузы, актинии и др.).

На сегодняшний день в различных улитках-конусах по всему миру обнаружено более 80000 конотоксинов. Обычно они состоят из 10–30 аминокислот, обладают высокими специфичностью и сродством к своим мишеням. Различные конотоксины являются естественными лигандами многих ионных каналов. МВПА (зиконотид), одобренный Управлением по санитарному надзору за качеством пищевых продуктов и медикаментов США (FDA), является наиболее известным продуктом, полученным из яда представителя рода *Conus*. Зиконотид проявляет анальгетическую активность при лечении хронической боли благодаря мощному ингибирующему действию на Ca^{2+} -каналы.

Некоторые компоненты яда скорпионов обладают противораковой активностью; кроме того, идентифицировано три типа антимикробных пептидов.

ASIC1a, субъединица кислото-чувствительных ионных каналов (ASIC), является терапевтической мишенью при многих заболеваниях, например, ишемии и психических расстройствах, а также против боли. Мощным и специфическим ингибитором ASIC1a является π -TRTX-Pc1a, пептид из 40 остатков, выделенный из яда паука [8].

Отдельного внимания заслуживают белковые и пептидные токсины из ядов различных животных, являющиеся лигандами никотиновых ацетилхолиновых рецепторов (nAChR) (в основном, антагонистами). Эти токсины имеют специфические структурные особенности и, соответственно, различаются параметрами связывания с рецептором. Так, α -нейротоксины змеиных ядов являются хорошо известными ингибиторами некоторых типов nAChR. Эти нейротоксины относятся к семейству так называемых трехпетельных токсинов, которое является одним из крупнейших семейств змеиных токсинов. Они получили свое название из-за особенности пространственной структуры, которая образована небольшим гидрофобным ядром с тремя выступающими β -структурными «пальцами» (петлями), стабилизированными дисульфидными связями [53]. Так называемые короткоцепочечные α -нейротоксины, содержащие 60–62 аминокислотных остатка и , являются специфическими ингибиторами nAChR мышечного типа, тогда как цитотоксины (со сходной длиной аминокислотной

последовательности) неспецифически лизуют (растворяют) различные клетки. Короткоцепочечные нейротоксины и цитотоксины содержат 4 дисульфида.

Длинноцепочечные α -нейротоксины и так называемые слабые токсины содержат 5 дисульфидных связей. Длинноцепочечные нейротоксины, содержащие 66–75 аминокислотных остатков и пятый дисульфид в центральной петле, помимо нАХР мышечного типа блокируют также нАХР нейронального типа (α_7 и $\alpha_9\alpha_{10}$). Слабые токсины с аминокислотной последовательностью из 65–82 аминокислотных остатков содержат пятый дисульфид в N-концевой петле. Эти токсины проявляют более широкий спектр биологической активности и взаимодействуют как с нАХР, так и с мускариновыми рецепторами ацетилхолина.

Существуют также лиганды с другими структурными особенностями, например, уже упомянутые пептиды α -конотоксины, сшитые двумя дисульфидными связями (см. обзор [54]). Наконец, существуют пептидные токсины, обладающие структурой без дисульфидных связей, например, аземипсин [55].

Трехпетельная укладка занимает видное место в двух крупных семействах белков: нейротоксинах змеиного яда и белках Ly6, последние из которых впервые обнаружены в иммунной системе млекопитающих. Помимо схожих трехмерных структур, белки подобного строения также имеют сходные связи остатков цистеина (дисульфиды), которые стабилизируют их структуру. Эти и другие общие черты доказывают эволюционную связь Ly6 и нейротоксинов змей. Белок SLURP-1 (secreted Ly6/urokinase-type plasminogen receptor-related protein) первоначально был выделен из биожидкостей человека; он также экспрессируется в кератиноцитах. Мутации SLURP-1 вовлечены в кожную болезнь Меледа. Кроме того, сообщалось, что SLURP-1 регулирует некоторые процессы в иммунной и нервной системах. Некоторые белки Ly6 (Lynx1 и SLURP-1 и -2), также как и упомянутые выше нейротоксины, взаимодействуют с нАХР, что позволяет предположить, что они могут функционировать как эндогенные модуляторы передачи сигналов нАХР [56].

Знание 3D структуры токсинов необходимо для понимания их механизма действия. Установление структурных сходств и различий между токсинами позволит выделить группы, с высокой вероятностью функционально схожие. Все это крайне важно при разработке лекарств на их основе. Кроме того, некоторые токсины змеиного яда, призванные наносить вред, структурно и функционально сходны с белками, играющими важную роль в нормальных физиологических процессах и необходимыми для выживания. Таким образом, раскрытие структурно-функциональных отношений и механизмов действия токсинов поможет понять природу патофизиологических свойств токсинов [11].

Как это следует из вышесказанного, важными элементами структуры белкового или пептидного токсина являются дисульфидные мостики: различные парные сшивки цистеинов

могут изменить конформацию белка и, следовательно, эффективность его связывания с рецептором. Исследования в области молекулярной токсикологии значительно облегчаются подходами геномики, транскриптомики и протеомики [57]. Многие новые токсины либо идентифицируются по нуклеотидным последовательностям, либо анализируются после отбора по активности [58]. Однако секвенирование и масс-спектрометрия в целом не дают информации о пространственной структуре и дисульфидных связях. Определение порядка замкнутости дисульфидных мостиков в токсинах обычно осуществляется путем структурного подхода с использованием методов РА и ЯМР. Другим, косвенным, способом определения порядка сшивок дисульфидных связей является сравнение исследуемого вещества, в частности, пептида, выделенного из яда, с пептидами, полученными путем синтеза. Подход ортогонального пептидного синтеза позволяет образовывать дисульфидные связи в заданном порядке. Затем синтетические и нативные пептиды сравнивают с помощью обращенно-фазовой хроматографии, опираясь на время удержания на колонке. Пептиды с одинаковой последовательностью, но разным порядком дисульфидных связей могут иметь разную хроматографическую подвижность и, соответственно, разное время выхода с колонки. Оба подхода, структурный и косвенный, требуют наличия относительно большого количества природных токсинов ядовитых животных, получение которого на практике зачастую является сложной задачей. Отсюда следует актуальность разработки новых подходов для анализа дисульфидных изомеров токсинов.

1.3.2. РС в исследовании пептидных и белковых токсинов

Белковые токсины, принципиально не отличающиеся от других белков, ранее анализировались методом РС. Исследовались вторичная и третичная структуры токсинов змей (в том числе, морских) [34, 59, 60, 61, 62, 63, 64, 65, 66], а также влияние на их структуру химических модификаций [67]. Отметим работу [68] авторы которой химически модифицировали триптофан в нейротоксине морской змеи, что привело к утрате им активности по связыванию ацетилхолиновых рецепторов. Чтобы изучить влияние модификации триптофана на конформацию нейротоксина, спектры КР модифицированного нейротоксина сравнили со спектрами нативного: была обнаружена крайняя схожесть их структур. Эти данные позволили авторам предположить, что триптофан в подобных короткоцепочечных нейротоксинах может играть непосредственную роль в процессе связывания рецептора ацетилхолина.

Структуры токсинов скорпионов [69], морских анемонов [70], немертин [71], камбалообразных [72], а также растений [73] и даже бактерий [74] также анализировались методом РС. Во многих этих работах помимо вторичной структуры рассматриваются конформация дисульфидов и тирозиновое окружение. Обилие работ, посвященных РС-исследованию структуры токсинов, объясняется тем, что это простой, доступный и быстрый метод, который не требует сложной подготовки образцов и при этом позволяет оценить пространственную укладку токсинов.

В ряде случаев, необходимо провести анализ структуры синтетических аналогов токсинов или сравнить ее с таковой нативных токсинов (полученных непосредственно из ядов). Так, в работе [75] было проведено РС-исследование синтетических антимикробных пептидов лягушки, нативная форма которых выделяется из кожи африканской шпорцевой лягушки *Xenopus laevis*. Они защищают лягушку от инфекций и проявляют противомикробную активность широкого спектра действия *in vitro*. Авторы работы показали, что структура пептидов представляет собой спираль. Это удалось установить, анализируя полосы спектров КР, характеризующие вторичные структуры белков. В другом исследовании [76], анализировался синтетический аналог мелиттина. Мелиттин – активный компонент пчелиного яда, он составляет половину веса сухого яда. Мелиттин представляет собой положительно заряженный пептид из 26 аминокислот, который связывается с фосфолипидами бислоя мембраны, вызывая гибель клеток [77]. На основе РС, было обнаружено формирование его синтетическим аналогом спиралей и их агрегация в крупные олигомеры в воде.

Взаимодействия токсинов с липидами (мембранами) и рецепторами также активно изучаются. Например, в [78] методом РС была определена вторичная структура аламетицина (пептида, продуцируемого грибом) в липидных мембранах ниже и выше температуры фазового перехода липидов T_c . Структурные данные были получены путем аппроксимации экспериментальных спектров суперпозицией спектров эталонных белков известной трехмерной структуры. Было обнаружено, что в липидном бислое при температуре выше T_c процент α -спиральности аламетицина меньше, чем таковой при температуре ниже T_c . В другом исследовании [79] рассматривалось взаимодействие кардиотоксинов с фосфолипидами; спектры КР липид-связанного кардиотоксина выявили, что его вторичная и третичная структуры в основном аналогичны структуре свободного токсина.

В другой интересной работе [80] методом резонансной РС изучалось взаимодействие двух производных α -токсина из яда *Naja nigricollis* с НАХР на мембранах эритроцитов *Torpedo marmorata*. Два модифицированных токсина несли либо группу NO_2 , связанную с Tyr25, либо нитрофенилтиоэфир (NPS), связанный с Trp29. Сравнение спектров свободных и модифицированных токсинов указало на то, что окружение Tyr25 не нарушается при

связывании с рецептором, но окружение NPS, связанного с Trp29, изменилось. Этот результат позволил авторам предположить, что Tyr25 не участвует в связывании, тогда как Trp29 α -токсина связывается с рецептором nAHR. Подчеркнем, что авторы описывают резонансную РС, которая позволяла усиливать колебания добавленных хромофоров путем облучения образца излучением с длиной волны, близкой к их полосам электронного поглощения. В данном случае, для тирозильных и триптофанильных производных использовалось возбуждение аргоновым лазером на длине волны 458 нм.

Заметим, что в некоторых работах прослеживаются изменения в спектрах самих мембран при взаимодействии с ними токсинов [81, 82] для выяснения эффектов, оказываемых последними. Интересно, что исследование самих рецепторов также позволяет выявить детали взаимодействия с ними определенных молекул. В статье [83] с помощью РС были определены структурные особенности фрагментов мембран электрического органа электрического ската *Torpedo marmorata*, богатых nAHR. Колебания, присущие мембранным белкам и липидам, были проанализированы до и после добавления агониста карбамоилхолина и антагониста (+)-тубокурарина. Изменения вторичной структуры, окружения ароматических аминокислот (тирозина и триптофана) и конформации дисульфидов рецепторов вследствие присутствия агонистов (увеличивающих отклик рецептора) и антагонистов (блокирующих действие рецептора) были прослежены и сравнены.

Подобные исследования взаимодействий токсинов с рецепторами или мембранами проливают свет на механизмы действия токсинов и взаимосвязи структура-функция. Понимание последних важно как с теоретической, так и с прикладной точек зрения, в частности, при разработке лекарств.

1.3.3. Яды змей

Среди нескольких тысяч известных видов змей не все ядовиты; ядовитые змеи в основном относятся к семействам Viperidae и Elapidae [84]. Изучение ядов змей имеет большое медицинское значение, так как их укусы опасны для человека и часто приводят к инвалидности или смерти. С другой стороны, некоторые компоненты ядов имеют потенциал в качестве фармацевтических препаратов [5, 85]. Яды представляют собой смеси различных соединений, таких как белки (пептиды), липиды, нуклеотиды, неорганические соли и т.д. При этом по содержанию преобладают соединения белковой природы; в составе змеиного яда их содержание составляет около 90% [86]. Среди белков яда (токсинов) выделяют следующие основные семейства: секретируемые фосфолипазы A2 (PLA2), трехпетельные токсины (3FTx),

металлопротеиназы змеиного яда (SVMP), сериновые протеазы змеиного яда (SVSP), оксидазы *l*-аминокислот (LAAO), богатые цистеином секреторные белки (CRiSP) и др. Также заслуживают внимания лектиноподобные белки С-типа (CTL), дезинтегрины (Dis) и факторы роста эндотелия сосудов (VEGF). Упомянем еще гомологи ингибиторов сериновых протеаз типа Kunitz (KuIn), которые преобладают в яде некоторых мамб.

Яды различных видов змей имеют разное соотношение перечисленных белков. Так, в яде Elapidae, как правило, преобладают 3FTx и PLA2, а в яде гадюк – SVMP, PLA2 и SVSP [86]. Это объясняет различное влияние ядов на организм жертвы: в целом укусы Elapidae вызывают нейротоксический синдром с минимальным местным повреждением тканей. Укусы же Viperidae сопровождаются глубоким локальным повреждением тканей [87]. Состав яда зависит от вида змей, а также других факторов, например, среды обитания (в частности, в связи с рационом питания [12]) и возраста [88]. Известно, что даже у змей одного вида, но обитающих в разных географических локациях (разные популяции), состав яда может существенно различаться [15, 86]. Так, в яде *Bungarus fasciatus* из Таиланда содержится около 65% PLA2 и 20% KuIn, из Индонезии – 30% PLA2, 15% 3FTx и 50% KuIn, а в яде BuFa из Китая преобладает PLA2 (более 90%).

Определение как межвидовой, так и внутривидовой изменчивости ядов, а также зависимости их состава от места происхождения змеи является важной задачей. С терапевтической точки зрения знание такой изменчивости позволит более эффективно лечить жертв укусов (укусы сородичей могут различаться по симптоматике и требовать разного лечения [12]). В частности, это может помочь в выборе (особенно когда доступны только моноспецифические, выработанные против яда одного вида или подвида змей, противоядия [89]) и производстве подходящего противоядия. Знание состава змеиного яда важно также для изучения эволюции ядовитых змей и их филогенетической характеристики на уровне видов и подвидов [15]. Различия в свойствах и составах ядов могут быть использованы, в частности, для различения морфологически схожих видов.

Большинство работ в рамках тематики вариаций ядов проводится путем сравнения составов ядов змей разных видов (межвидовые вариации), тогда как в существенно меньшем количестве исследований анализируются внутривидовые различия ядов. В случае внутривидовых вариаций, как правило, изучают яды разных популяций змей одного вида [16]. Но даже внутри одной популяции могут наблюдаться значительные вариации в составе яда. Известно, что яды демонстрируют онтогенетические вариации [90]. Это может быть вызвано возраст-зависимой экспрессией токсинов [91], что часто связано с различиями пищевых привычек новорожденных и взрослых особей [92]. Но различия в экспрессии могут быть также

связаны с другими факторами, которые различаются даже среди особей одного и того же возраста в одной и той же популяции.

Тем не менее, подавляющее большинство исследований, посвященных вариациям ядов, проводится с образцами, полученными путем объединения ядов нескольких особей змей. В некоторых случаях это происходит потому, что количества яда, которое можно получить от одной отдельной змеи, недостаточно для анализа принятыми методами. Таким образом, степень изменчивости ядов внутри популяций плохо изучена, несмотря на то, что существование внутривидовой изменчивости предсказывает гипотеза о том, что яд является быстро развивающимся признаком [93]. При быстрой эволюции признака должно существовать значительное наследуемое разнообразие внутри популяций. Несмотря на это, вариации ядов среди взрослых особей одной популяции только начинают исследовать [94].

Даже при сравнении ядов разных видов змей важно изучать не только усредненные яды (полученные путем объединения ядов многих змей), но и яды отдельных особей, чтобы иметь возможность количественно сравнивать внутри- и межвидовые различия. Полную картину изменчивости ядов можно получить, рассматривая одновременно яды нескольких особей 1) разных близкородственных видов змей, 2) разных популяций внутри вида и 3) одной популяции.

Возможность сравнивать изменчивость на межпопуляционном и межвидовом уровнях может пролить свет на процесс видообразования. Если эти вариации сопоставимы, может возникнуть вопрос об отнесении разных популяций одного и того же вида к разным видам. Например, вид *Vipera nikolskii* филогенетически очень близок к *Vipera berus* и в некоторых публикациях рассматривается как подвид *V. berus*. Однако, анализ яда *V. nikolskii* показал, что он существенно отличается от яда *V. berus* [95]. Кроме того, для разработки противоядия необходимо выявить внутривидовые варианты яда, чтобы гарантировать эффективность противоядия.

Нам удалось найти только несколько статей, описывающих исследование цельного яда с помощью РС. Так, РС была использована для характеристики биохимических изменений в икроножной мышце мыши, которой был введен яд *Bothrops jararacussu*. Также этот яд был обнаружен с помощью РС в мышечной ткани мыши [96]. В другой статье [97] для сравнения и анализа двух ядов, *Trimeresurus stejnegeri* из семейства Elapidae и *Bungarus multicinctus* из семейства Viperidae, авторами использовалась методика SERS. Спектры SERS различались достаточно, чтобы отличить два изученных яда. Но спектры SERS, как обсуждалось выше, существенно зависят от свойств усиливающих наночастиц (геометрии, химии поверхности и других). Кроме того, поскольку в основном усиливается сигнал от молекул и их областей, расположенных вблизи поверхностей частиц, полученные таким образом спектры

характеризуют не столько сам яд, сколько характер взаимодействия его отдельных белков с наночастицами. Упомянем также исследование [98], где с помощью другого оптического метода, инфракрасной спектроскопии с преобразованием Фурье, были исследованы объединенные яды змей *Macrovipera lebetina lebetina* и *M. l. obtusa*. Исследование выявило ряд основных и второстепенных пиков в спектрах поглощения, которые были соотнесены с различными биомолекулами (белками, углеводами и нуклеиновыми кислотами). Было замечено, что в спектрах ядов двух подвидов змей имеются некоторые количественные различия и сдвиги пиков.

1.3.4. Метод масс-спектрометрии для анализа цельных ядов

В настоящее время одним из основных методов оценки состава цельного яда («прямой» подход) является жидкостная хроматография с тандемной масс-спектрометрией (ЖХ/МС/МС), позволяющая определить, какие семейства белков присутствуют в яде [17, 99]. Этот метод не лишен сложностей: для многих белков требуется анализ последовательностей *de novo* с поиском гомологии, чтобы опознать, к каким семействам токсинов они относятся. Данные транскриптома ядовитых желез, хотя и очень полезны при МС-идентификации белков яда, не обязательно отражают протеом яда. Кроме того, метод ЖХ-МС требует относительно сложной подготовки проб [17]. Например, метод ЖХ-МС/МС снизу вверх (bottom-up) подразумевает предварительное расщепление белковых образцов с помощью фермента (например, трипсином). Такие методы, как хроматография и гель-электрофорез, используются для разделения белков яда и обычно также требуют последующего МС-анализа для идентификации белков [100].

При анализе ядов отдельных особей с помощью часто применяемого метода безметочной (label-free) ЖХ-МС/МС снизу вверх (bottom-up) может возникнуть трудность, связанная с малым количеством яда, получаемого от одной змеи, поскольку точность анализа зависит, в том числе, от количества образца. Также для анализа необходимо иметь в своем распоряжении обширные базы данных белков, входящих в состав яда (желательно для ядов изучаемых видов змей или близких к ним). Необходимо понимать, что существует множество гомологичных белков, которые имеют значительные участки идентичных последовательностей. Поэтому, в частности, при МС-анализе триптических фрагментов белков яда, оценку состава обычно проводят с точки зрения выявления белковых семейств, а не конкретных белков [99]. Даже самые совершенные масс-спектрометры часто не справляются с количеством белков и пептидов, присутствующих в образце. Как следствие, в ходе любого анализа можно

идентифицировать только часть всех присутствующих белков, и только часть всех идентифицированных белков может быть учтена при количественном анализе [101]. При использовании метода МС сверху вниз (top-down) (который подразумевает введение в спектрометр интактных белков) количественный анализ также ограничен из-за более низкой эффективности ионизации и обнаружения крупных белков.

Несмотря на свои недостатки, методы МС незаменимы в протеомике и составляют основу анализа многокомпонентных белковых (и не только) смесей. Но в некоторых случаях оправданно применение других, не столь информативных, но более простых – менее трудоемких и времязатратных – методов. Например, если необходимо проанализировать крупную серию образцов и оценить их сходства (различия) и основные особенности составов, то РС может стать альтернативным подходом.

1.4. Метод главных компонент для анализа спектральных данных

Сегодня вошли в практику методы математической обработки данных; применительно к РС, они предоставляют собой новые способы извлечения полезной информации из спектров и позволяют одновременно анализировать множество образцов.

1.4.1. МГК и методы кластеризации

Метод главных компонент (МГК) и другие методы, используемые в практике при реализации машинного обучения, могут быть крайне полезны при работе со сложными, многомерными данными, коими являются спектры КР белков. С помощью МГК становится возможным выделить из широкого спектра молекулярных колебаний (признаков) ключевые колебания (признаки), таким образом, определив, на основании каких признаков мы можем делать выводы о схожести и различии образцов. Далее, на основе величин этих наиболее значимых признаков можно судить об особенностях образцов. При этом мы можем использовать изначально широкие спектры признаков и таким образом заложить в алгоритм всю имеющуюся информацию об объектах, не анализируя спектры предварительно напрямую, ведь алгоритм придаст больший вес важным, и меньший – неважным признакам. Это особенно полезно, когда ключевые различия в сложных спектральных данных прямым анализом выявить трудно. В случае обучения алгоритма, последний научится какие использовать веса для разных

признаков для отнесения образца в ту или иную группу. Можно провести аналогию с нейросетью, где центральной задачей также является подбор весов.

МГК [102] позволяет упростить процедуру анализа данных за счет уменьшения их размерности. Помимо очевидного упрощения данных, МГК позволяет игнорировать небольшие спектральные шумы, сохраняя при этом важную информацию. Этот метод представляется перспективным в случае анализа спектров КР белков: в результате использования МГК становится возможным выявить ключевые различия в спектрах образцов и представить их в наглядном виде, визуализировав отношения (сходства и различия) между исследуемыми веществами в новом пространстве пониженной размерности. Расстояния между образцами при этом используются как мера их различия.

МГК сводится к проекции исходных p -мерных векторов на q направлений, главных компонент (ГК) нового q -мерного подпространства. В случае обработки КР спектров, число q соответствует числу частот, для которых измерялся спектр. Спектр КР можно представить как точку в пространстве с размерностями $p=d/\Delta+1$, где d – диапазон измерения (в см^{-1}), а Δ – шаг (разрешение датчика изображения) (также в см^{-1}).

С математической точки зрения, например, первая ГК u_1 задается линейной комбинацией переменных x_1, x_2, \dots, x_p (в случае КР спектров, интенсивностей сигналов для p частот): $u_1 = a_{11}x_1 + a_{12}x_2 + \dots + a_{1p}x_p$, где $a_{11}, a_{12}, \dots, a_{1p}$ – веса (аналогично для ГК2 и т.д.). Сумма всех весов приравнивается единице.

Существует несколько эквивалентных способов математического вывода ГК. Самый простой из них – найти проекции, которые максимизируют дисперсию. Первая ГК – это направление в пространстве, вдоль которого проекции имеют наибольшую дисперсию. Вторая ГК – это направление, которое максимизирует дисперсию среди всех направлений, ортогональных первому и т.д.

Перед использованием МГК необходимо предварительно обработать спектры (предобработать). Цель предобработки – минимизировать вариации спектров образцов, не связанные с их природой, а вызванные различиями в условиях регистрации спектров, а также автофлуоресценцией (или фотолюминесценцией) образцов. Первый фактор имеет место даже в том случае, если параметры прибора для регистрации всех спектров одинаковы: например, сухие белки или пептиды достаточно пористые и поэтому неоднородные. Следовательно, при резкой фокусировке (при недостаточном усреднении по объему) спектр даже одного и того же образца, в частности, интенсивность спектра и уровень фона, может меняться в зависимости от места фокусировки. Эта проблема может быть решена с помощью нормировки. Что касается фона автофлуоресценции, то даже следовые загрязнения образца могут внести существенный вклад в спектр КР. Коррекция базовой линии успешно применяется для минимизации

подобного фона, если он практически не маскирует спектральные детали, но «приподнимает» спектр и искривляет базовую линию [103, 104].

После успешного применения МГК упрощенные спектральные данные затем можно анализировать другими методами, в частности, методами классификации (кластеризации), для применения которых, как правило, необходимо, чтобы число образцов превышало или хотя бы было близко к числу переменных параметров (характеристик, которыми отличаются образцы).

Полученные данные пониженной размерности могут далее быть подвергнуты кластерному анализу, например, k-means или иерархической кластеризации. Эти методы не требуют предобучения модели и позволяют объективно выделить группы среди образцов, отражающие их взаимное сходство. Но при этом иерархическая кластеризация не дает информации о числе кластеров, а метод k-means требует их предварительного задания.

Есть также ряд методов, требующих предобучения алгоритма. Отметим анализ дискриминантных функций (discriminant function analysis, DFA) и дискриминантный анализ методом частичных наименьших квадратов (partial least squares discriminant analysis, PLS-DA), которые удобно применять для разделения объектов на два и более классов [105]. Например, метод DFA позволяет построить границу между известными классами обучающих объектов. Для этого требуется его «тренировка» на части образцов (обучающем наборе), для которых имеется информация об их принадлежности к конкретному классу. В частности, включение образца в одну из двух групп определяется тем, является ли величина его «discriminant score», которая рассчитывается в ходе DFA для каждого образца, положительной или отрицательной.

1.4.2. Применение МГК к данным РС

Одно из основных направлений применения МГК к РС данным – распознавание патологических образцов тканей, жидкостей или клеток путем их сравнения с нормальными и скрининг. Разные авторы анализируют образцы сыворотки [106] или плазмы [107] крови, или образцы биопсии ткани [108]. В некоторых работах применяется методика картирования [109].

Основная идея – разделить спектры образцов на две группы – «норма» и «патология» и определить маркерные регионы спектра, частоты и интенсивности которых отражают «нормальность» и «патологичность». На графике МГК (score plot) легко визуализировать разделение образцов, а с помощью спектра весов для ГК (loading plot) можно установить влияние разных частот на значения ГК. Производя интерпретацию этих данных, можно установить, какие молекулярные и структурные особенности характеризуются маркерными

регионами, и соответственно, какие биологические особенности образцов характерны для патологии.

Интересно отметить работу [110], где распознавание апоптоза и некроза в клетках лейкемии проводилось с помощью двух методов – МГК и анализа биохимических компонент (biochemical component analysis, BCA). Метод BCA использует регрессию наименьших квадратов и, опираясь на то, что спектр образца (в первом приближении) представляет собой сумму спектров всех его основных компонентов, позволяет оценить вклад каждого компонента. При этом BCA требует предварительного знания чистых компонентов образца и наличия их спектров. Авторы отмечают, что уникальный биохимический отпечаток типов гибели лейкозных клеток, полученный в виде спектров КР, может помочь определить путь клеточной гибели, вызванной лекарственными препаратами. Отмеченные же выше методы позволяют упростить и визуализировать информацию, полученную из сложных спектральных данных. Оба метода показали высокую эффективность при классификации живых, апоптотических и некротических клеток. МГК же обладает тем преимуществом, что может быть полезен, когда спектры основных чистых компонентов образцов недоступны.

В некоторых случаях наличие или отсутствие заболевания или инфекции определяется опосредованно по детектированию в биопробе антител к соответствующему патогену. Например, в работе [111] РС и МГК использовались для обнаружения присутствия или отсутствия антител IgG, IgM и IgA (иммуноглобулины G, M и A соответственно) против *Toxoplasma gondii* в образцах, полученных от людей, часть из которых контактировала с паразитом.

В других работах детектируются и анализируются изменения в образцах, произошедшие в результате различных воздействий или в результате лечения. Так, в [112] оценивалось фотоиндуцированное повреждение кожи, а в [113] – влияние pH и нагрева на структуру сывороточного белка.

Авторы [114] провели мониторинг химиотерапевтического лечения лейкоза. В качестве образцов использовались капли сыворотки. МГК и линейный дискриминантный анализ (ЛДА) спектров КР позволили различить пациентов из контрольной группы и пациентов с лейкемией, а также отличить друг от друга различные типы лейкемии. Мониторинг же лечения основывался на наблюдении биохимических изменений на протяжении химиотерапевтического лечения лейкозов (chemotherapy leukemia treatment, CHLT). На графиках МГК наблюдалось приближение кластера образцов пациентов к кластеру образцов контрольной группы на протяжении всего CHLT. Когда кластер пациентов полностью находится внутри кластера контрольной группы, пациенты могли быть объявлены здоровыми. Таким образом, график МГК отражал историю болезни.

Течение во времени определенных процессов, например, формирование коллагена [115], также может быть рассмотрено с помощью комбинации РС и МГК. Интересно также отметить, что РС совместно с МГК применялась для определения ориентации молекул липидов [116], что может представлять интерес при изучении мембранных рецепторов, естественная среда которых – липидная мембрана. В упомянутой работе МГК был применен к спектрам КР для характеристики молекулярного упорядочения в липидных материалах; при этом авторы утверждают, что современные рамановские спектрометры достаточно чувствительны для изучения одиночных бислоев. Было произведено картирование главных компонент по плоским образцам фосфолипидов. Было показано, что совместно с данными о поляризации возбуждающего излучения определенные пики КР высокой интенсивности могут использоваться в качестве индикаторов ориентации фосфолипидов.

1.5. Выводы к главе 1

В итоге можно заключить, что классическая РС, при оптимально подобранных условиях записи спектров, имеет множество преимуществ при исследовании белковых молекул и достаточно информативна, поэтому, в определенных случаях, может явиться альтернативой другим методам исследования белков и их смесей.

Актуальность таких токсикологических вопросов, как исследование структур животных токсинов и составов ядов, делает приложение инструмента спектроскопии КР в область токсикологии крайне многообещающим.

Глава 2. Классификация спектров КР токсинов с помощью МГК

Имея в своем распоряжении ряд токсинов (которые, например, являются лигандами конкретных рецепторов, в частности, nАХР), часто в небольших количествах, полезно иметь метод их быстрого анализа и сравнения их структуры с уже имеющимися известными токсинами. При лабораторном синтезе множества пептидных токсинов актуальной задачей является быстрая оценка их структуры. РС в сочетании с МГК кажутся перспективными для реализации эффективного метода предварительного анализа белковых токсинов (которые необходимы в количестве всего нескольких мкг в сухом виде).

В данной части работы МГК был применен для анализа спектров КР белковых и пептидных токсинов из ядов животных, включая их синтетические аналоги и дисульфидные изомеры.

2.1. Материалы и методы

2.1.1. Образцы пептидов и белков

В первой части работы для анализа были взяты девять белков и пептидов, которые, как известно, взаимодействуют с nАХР. Три из них были выделены из ядов змей: длинноцепочечный α -кобратоксин (Ctх) из *Naja kaouthia* (моноклевого кобры) и короткоцепочечный α -нейротоксин NT2 из кобры *N. oxiana* – трехпетельные токсины (содержащие 5 и 4 дисульфидные связи соответственно), а аземиопсин (Az), пептид без дисульфидных мостиков, из яда гадюки *Azemiops feae*. Остальные лиганды представляли собой синтетические аналоги встречающихся в природе α -конотоксинов из моллюсков рода *Conus*: SIA, MI и MII (каждый содержит два дисульфида), и их аналоги K-SIA, L-PnIA и RL-PnIA (имеют аминокислотные замены).

Во второй части работы анализировались следующие трехпетельные токсины: Ctх, длинноцепочечный α -нейротоксин NT1 из *N. oxiana*, цитотоксин Ct2 из *N. oxiana*, цитотоксин Cx1 из *N. kaouthia*, а также так называемые слабые токсины Owt из *N. oxiana* и Wtx из *N. kaouthia*. Последние два имеют пятую дисульфидную связь в N-концевой петле, тогда как длинноцепочечные нейротоксины имеют пятый дисульфид в центральной петле; как правило, слабые нейротоксины менее токсичны, чем α -нейротоксины [117].

Все эти белковые молекулы имеют разную первичную структуру и разные элементы вторичной структуры, в том числе петлевые структуры, стабилизированные дисульфидными связями. Некоторые структурные данные об этих образцах даны в таблице 2.1.

Спектры КР записывали без дополнительной пробоподготовки с сухих образцов (порошков). Змеиные токсины были выделены из цельных ядов путем сочетания различных типов хроматографии; все конотоксины были получены методом твердофазного пептидного синтеза. Синтез конотоксинов описан в работе [118]. Вкратце, твердофазный синтез использовался для сбора линейных пептидов с последующем окислением на воздухе с образованием дисульфидных связей. Порядок образования дисульфидных связей контролировался с помощью жидкостной обращенно-фазовой хроматографии путем измерения подвижности на колонке с использованием контрольных конотоксинов в качестве стандарта. Выделенные из ядов или синтезированные соединения лиофилизировали.

Таблица 2.1. Данные о дисульфидных связях и длинах аминокислотных последовательностях образцов.

Токсины (при наличии, в скобках указаны аминокислотные замены)	Положения цистеинов, соединенных дисульфидными связями	Длины последовательностей	Виды животных
МII	2-8 и 3-16	16	<i>Conus magus</i>
MI	3-8 и 4-14	14	<i>Conus magus</i>
L-PnIA (A10L PnIA)	2-8 и 3-16	13	<i>Conus pennaceus</i>
SIA	2-7 и 3-13	16	<i>Conus striatus</i>
RL-PnIA (A9R, A10L PnIA)	2-8 и 3-16	16	<i>Conus pennaceus</i>
K-SIA (D12K SIA)	2-7 и 3-13	13	<i>Conus striatus</i>

RgIA	2-8 и 3-12	13	<i>Conus regius</i>
isoRgIA	2-12 и 3-8	13	<i>Conus regius</i>
Az	-	21	<i>Azemiops feae</i>
Ctx	3-20, 14-41, 26-30, 45-56 и 57-62	71	<i>Naja siamensis</i>
NT2	3-23, 17-40, 42-53 и 54-59	61	<i>Naja oxiana</i>
NT1	3-22, 15-43, 28-32, 47-58 и 59-64	73	<i>Naja oxiana</i>
Owt	3-24, 6-11, 17-42, 46-57 и 58-63	65	<i>Naja oxiana</i>
Wtx	3-24, 6-11, 17-42, 46-57 и 58-63	65	<i>Naja kaouthia</i>
Cx1	3-21, 14-38, 42-53 и 54-59	60	<i>Naja kaouthia</i>
Ct2	3-21, 14-38, 42-53 и 54-59	60	<i>Naja oxiana</i>

2.1.2. Сбор данных

Измерения спектров КР проводили с помощью рамановского микроскопа Bruker Senterra II. Длина волны возбуждающего лазерного излучения составляла 532 нм. Возбуждение на данной длине волны приводит к большей флуоресценции, чем возбуждение в ИК-диапазоне, но, с другой стороны, обеспечивает бóльшую эффективность рассеяния. Сухие образцы помещались на латунную подложку. Фокусировка излучения производилась объективом с увеличением 50^{\times} и числовой апертурой 0,65. В этих условиях лазерное пятно на образце имело диаметр около 2,5 мкм. Мощность лазера была ограничена 12 мВт во избежание повреждения образцов. При записи каждого спектра время интегрирования составляло 2 с и применялось усреднение по 100 спектрам. Для обеспечения воспроизводимости, каждый образец измерялся не менее чем в трех точках его объема.

2.1.3. Обработка данных

Для анализа использовался спектральный диапазон от 300 до 1800 см^{-1} , так как он содержит основную информацию в случае исследования биологических молекул.

Предобработка спектров включала следующие этапы: 1) коррекцию базовой линии, 2) нормализацию и 3) сглаживание. Коррекция базовой линии выполнялась в OPUS 8.2.28 (Bruker Optik GmbH, Эттинген, Германия) («rubberband correction», количество точек базовой линии = 64, количество итераций = 10). Шаги 2 и 3 были выполнены в программе PyChem 3.0.5g Beta (с открытым исходным кодом) [119].

Multiplicative scatter correction (MSC) или его расширенная версия (extended MSC, EMSC) – это метод нормализации, используемый для масштабирования и коррекции вариаций в спектрах по сравнению с эталонным (средним) спектром [104]. В качестве метода нормализации был выбран EMSC (полиномиальный порядок = 3), поскольку он наиболее эффективно устранял фон. Для сглаживания был применен фильтр скользящего среднего (ширина фильтра = 10), удаляющий высокочастотные шумы.

МГК и иерархическая кластеризация также выполнялись в PyChem. Спектр КР образца представлял собой точку в пространстве с 1001 измерением, и применение МГК позволяло резко сократить число измерений (до числа главных компонент (КГ)) практически без потери полезной информации. Входные данные для МГК были представлены в виде матрицы, где строки матрицы соответствовали образцам, а столбцы содержали величины рамановских сигналов для разных рамановских сдвигов.

При анализе амидной полосы Amide I МГК входными данными являлась матрица с меньшим количеством столбцов, поскольку в этом случае анализировался меньший частотный диапазон. Следует отметить, что перед применением МГК в данном случае применялось скалирование спектров.

При применении МГК одновременно к двум спектральным областям (область S–S и полоса Amide I) входная матрица включала частоты, соответствующие этим областям.

Для иерархической кластеризации [120] полученные после МГК координаты образцов в пространстве ГК (scores) использовались в качестве входных данных. Иерархическая кластеризация позволяет выявить взаимные сходства и различия образцов, в частности, на основе евклидова расстояния между ними в пространстве ГК (также могут использоваться другие метрики расстояния). Результатом иерархической кластеризации является дендрограмма, показывающая «степень родства» образцов, где длины ветвей пропорциональны различию образцов.

K-means кластеризация (в пространстве ГК) была выполнена в Python 3.9 с использованием библиотеки Scikit-learn. Число кластеров, равное трем, задавалось в соответствии с количеством возможных вторичных структур, преобладающих в белках (пептидах) (α -, β - или нерегулярная структура).

Выравнивание аминокислотных последовательностей и расчет дерева выполнялись в Jalview 2 [121]: для выравнивания использовался веб-сервис ClustalO; далее из выровненных последовательностей, на основе сходства между ними (общих участков), рассчитывалось дерево.

2.2. Результаты

2.2.1. Анализ спектров КР токсинов

Спектры были записаны с сухих образцов, примерный вид через микроскоп одного из которых приведен на рисунке 2.1. Образцы представляли собой пористую неоднородную субстанцию, включающую пустоты и скопления вещества различной формы. При записи спектров стояла задача оптимальной фокусировки в объеме образца. Было выявлено, что при фокусировании на ровных, гладких участках удается получить максимально интенсивный рамановский сигнал. При этом видимая в микроскоп область возбуждения, образованная падающим лазерным излучением, имела минимальную площадь.

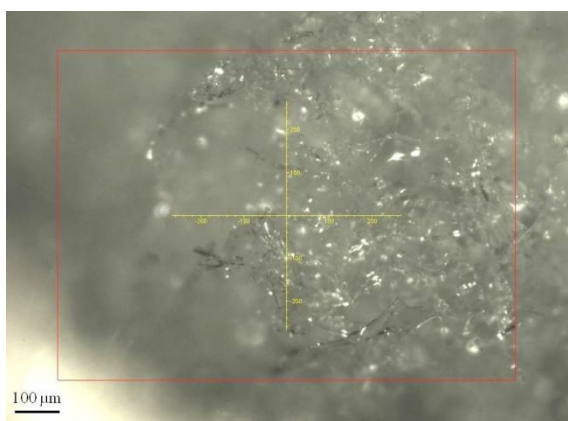


Рисунок 2.1 – Вид образца Owt через объектив 4^x.

Было проверено, что нагрев образцов, производимый лазерным облучением, не влияет на их структуру при используемой мощности лазера (данные не приведены). При увеличении мощности от минимального до используемого значения, интенсивности линий спектров одного и того же образца, записанные при разных мощностях, линейно зависели друг от друга. Сдвига пиков и изменения их формы при увеличении мощности не наблюдалось. В то же время нам удалось добиться более высокого соотношения сигнал/шум при использовании большей мощности.

Путем коррекции базовой линии были получены практически свободные от фона спектры КР. По сравнению с двумя альтернативными методами («Min-Max» и векторной нормализацией) использование EMSC для нормализации позволило получить наиболее близкие друг к другу спектры образцов, тем самым исключив вариации, не несущие полезной информации. Сглаживание спектров практически не влияло на последующие результаты МГК, но позволило получить более презентабельные спектры. Спектры после предварительной обработки показаны на рисунке 2.2.

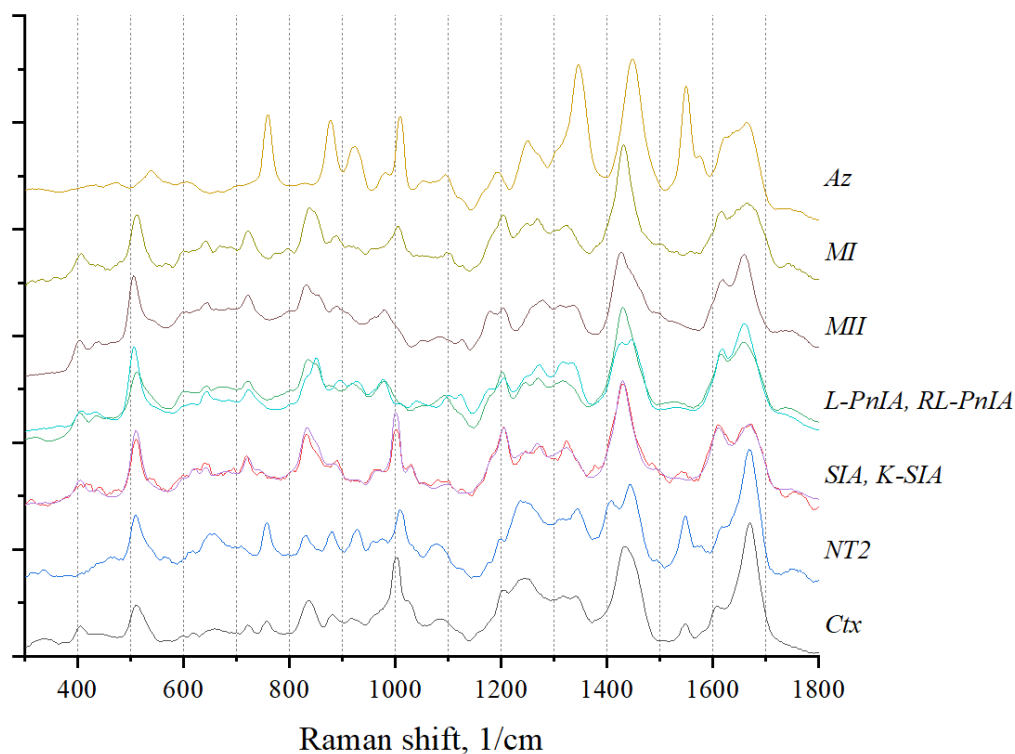


Рисунок 2.2 – Спектры КР после предобработки. Каждый спектр является «отпечатком пальцев» соответствующего образца. Наиболее близкие образцы, SIA и его аналог K-SIA, имеют чрезвычайно схожие спектры. Спектры отражают первичную (ряд пиков характерен для конкретных аминокислот) и вторичную структуры токсинов.

Далее МГК применялся к предобработанным спектрам. Были использованы первые три ГК, описывающие 88% дисперсии. Полученный график МГК показан на рисунке 2.3 (для наглядности показан 2D-график). На основании полученных данных становится возможным достоверно разделить образцы на три группы, отмеченные на графике разными цветами, которые соответствуют трем различным типам организации пространственной структуры токсинов: 1) змеиные токсины Ctx и NT2 с трехпетельной структурой, 2) α -конотоксины морских моллюсков с двумя дисульфидами и 3) змеиный пептид Az без дисульфидных связей. Достоверность такого разделения подтверждается иерархическим кластерным анализом.

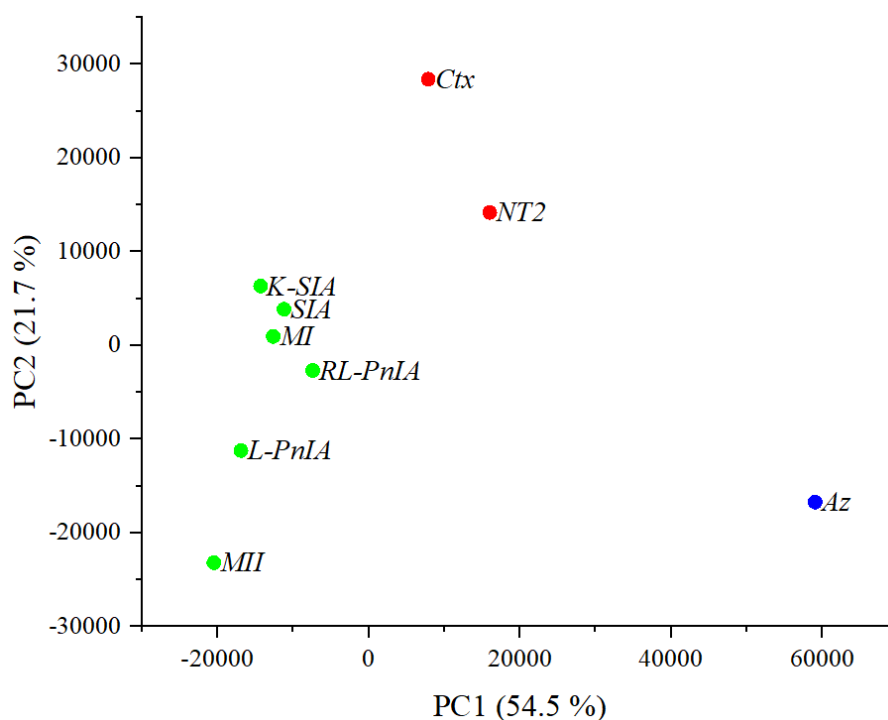


Рисунок 2.3 – График МГК, полученный после анализа преобработанных спектров в диапазоне 300–1800 см⁻¹. В скобках указан процент объясненной дисперсии. В первую группу вошли токсины Ctx и NT2, во вторую – SIA, K-SIA, MI, MII, L-PnIA и RL-PnIA, в третью – Az. Подписи осей PC1 и PC2 соответствуют русскоязычным обозначениям ГК1 и ГК2 (здесь и далее для графиков МГК).

Дендрограмма, полученная путем применения иерархической кластеризации к обработанным путем МГК данным, показана на рисунке 2.4 (А). Интересно проследить взаимоотношения внутри группы α -конотоксинов: отчетливо просматривается подгруппа, состоящая из SIA, K-SIA и MI, а также сходство между MII и L-PnIA (и его аналогом). Этот факт находится в соответствии со сходством аминокислотных последовательностей этих токсинов. Сходство последовательностей удобно визуализировать по их выравниванию. На рисунке 2.4 (Б) даны последовательности токсинов, выровненные с помощью биоинформатического сервиса.

Чтобы проверить, в какой степени сходство только аминокислотных последовательностей (в частности, без учета влияния дисульфидных связей) определяет полученную кластеризацию, с помощью биоинформатического сервиса было также построено дерево, с использованием в качестве входных данных выровненных аминокислотных последовательностей токсинов (Рисунок 2.4 (В)). Это дерево сопоставлялось с дендрограммой на рисунке 2.4 (А).

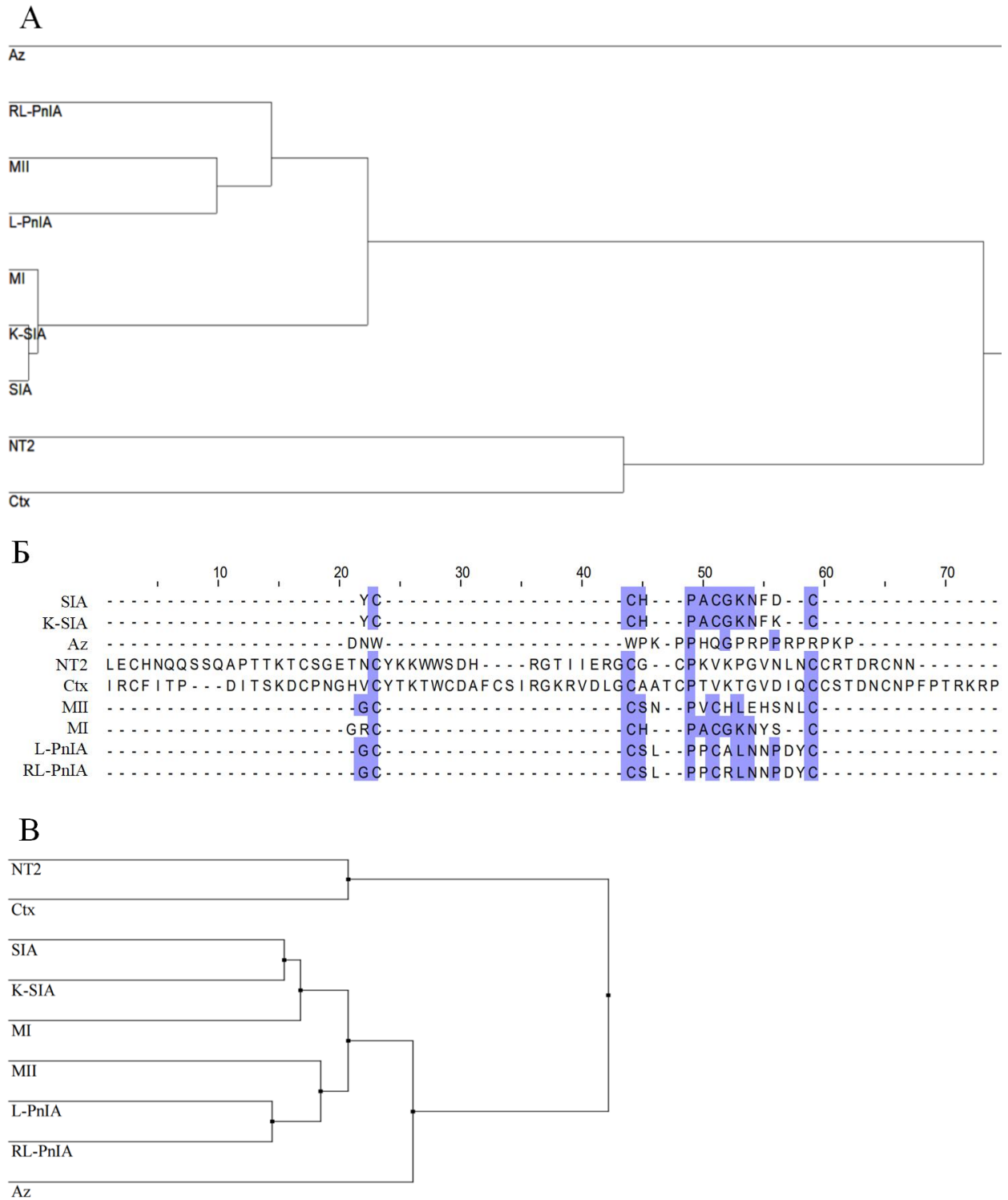


Рисунок 2.4 – (А): дендрограмма, полученная путем применения иерархической кластеризации к обработанным путем МГК данным. (Ветвь Az, для компактности показана не во всю длину).

Видны подгруппы, состоящие из SIA, K-SIA и MI, а также MII, L-PnIA и RL-PnIA; (Б): последовательности токсинов, выровненные веб-сервисом ClustalO. Выделены идентичные аминокислоты, которые встречаются в одних и тех же положениях в трех и более образцах; (В): дендрограмма (дерево), рассчитанная на основе выровненных последовательностей.

В целом иерархическая кластеризация обоими методами – на основе 1) аминокислотных последовательностей и 2) спектров КР токсинов – дает схожие результаты, но интересно, что положения токсинов L-PnIA и RL-PnIA различаются в зависимости от примененного метода. Это может свидетельствовать о том, что, несмотря на сходство их последовательностей, которое отражает дерево, полученное с помощью биоинформатического анализа, эти два токсина различаются по своей пространственной структуре. Близость же SIA и K-SIA на дендрограмме может указывать на то, что одиночная мутация в SIA не значительно повлияла на его структуру.

Различие дендрограмм можно объяснить прямым анализом спектров образцов. Для этого были рассмотрены два маркерных региона L-PnIA, RL-PnIA, SIA и K-SIA: область 500–550 см⁻¹ (Рисунок 2.5 (А) и 2.5 (В)) и область 810–870 см⁻¹ (Рисунок 2.5 (Б) и 2.5 (Г)). Как было сказано в главе 1, область S–S характеризует растяжение дисульфидных связей и используется для анализа геометрии связей C–C–S–S–C–C. Вторая область содержит дублет тирозина, пики около 850 и 830 см⁻¹, соотношение интенсивностей I_{850}/I_{830} которых используется для определения микроокружения тирозиновых остатков. В гидрофильном окружении $I_{850}/I_{830} = (0,91–1,45)$, в гидрофобном $I_{850}/I_{830} = (0,7–0,9)$. (Отношение I_{850}/I_{830} рассчитывается как отношение интенсивности сигнала КР на частоте 850 см⁻¹ к интенсивности сигнала КР на частоте 830 см⁻¹.)

В S–S области пик RL-PnIA сдвинут в спектре относительно пика L-PnIA, но незначительно (конфигурация *gauche-gauche-gauche*). Это говорит о сохранении конфигурации дисульфидов, но, в то же время, указывает на некоторые конформационные отличия RL-PnIA (относительно L-PnIA). Пики SIA и K-SIA не сдвинуты друг относительно друга, что указывает на сохранение конформации при введении аминокислотной замены. О последнем свидетельствуют также тирозиновые дублеты спектров SIA и K-SIA, для которых отношения I_{850}/I_{830} имеют близкие значения ($0,8 \pm 0,1$) и указывают на гидрофобность тирозинового окружения в SIA и его мутанте. В тоже время, соотношение интенсивностей $I_{850}/I_{830} \approx 0,9$ для L-PnIA указывает на гидрофобное окружение тирозинового остатка в его структуре, тогда как для RL-PnIA $I_{850}/I_{830} \approx 1,2$, что говорит о более гидрофильном окружении остатка тирозина. Изменение тирозинового окружения указывает на некоторые конформационные изменения RL-PnIA.

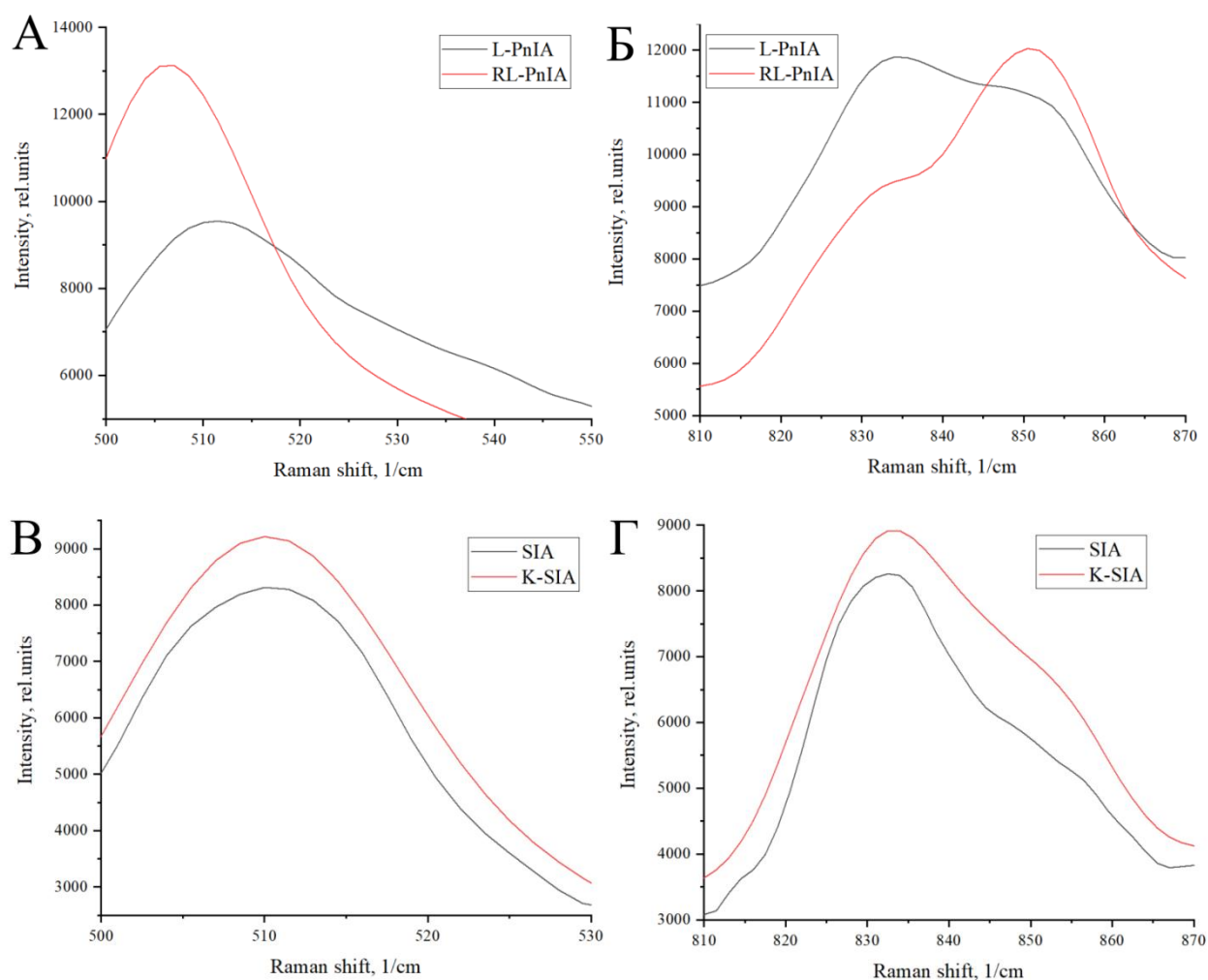


Рисунок 2.5 – Маркерные регионы спектров L-PnIA, RL-PnIA, SIA и K-SIA. (А): информативная часть S–S региона L-PnIA и RL-PnIA; (Б): область дуплета тирозина L-PnIA и RL-PnIA; (В): информативная часть S–S региона SIA и K-SIA; (Г): область дуплета тирозина SIA и K-SIA.

Кроме того, позиция Az различается при кластеризации двумя методами. После выравнивания его аминокислотная последовательность схожа с таковой конотоксинов. Это отражает его положение в биоинформатически построенном дереве. В то же время, на положение Az на дендрограмме, полученной в результате анализа спектров КР, влияет не только его первичная, но и 3D структура, что, возможно, и определяет его обособленность на этой дендрограмме.

Основные различия в спектрах токсинов обусловлены различием их первичной структуры. Об этом свидетельствует сходство дендрограмм, полученных из спектральных данных и из аминокислотных последовательностей токсинов. Однако следует отметить, что сравнение спектров токсинов со сходными последовательностями (L-PnIA и RL-PnIA) выявляет более тонкие различия, связанные с особенностями их 3D-структуры. При анализе МГК спектров в диапазоне $300\text{--}1800\text{ см}^{-1}$ такие структурные особенности, как окружение остатков

ароматических аминокислот, геометрия дисульфидных связей, наличие тех или иных вторичных структур и т. д. автоматически учитываются.

2.2.2. Исследование спектрального региона вторичной структуры токсинов

МГК был применен к спектрам токсинов в диапазоне от 1640 до 1730 cm^{-1} для получения более конкретной информации об их вторичной структуре. В этот диапазон частично входит полоса Amide I (1600–1690 cm^{-1}), которую используют для выявления наличия α -спиралей, β -слоев и нерегулярных структур в белках. Вся полоса Amide I не анализировалась, чтобы минимизировать влияние на спектр ароматических аминокислот, пики которых расположены в диапазоне 1600–1620 cm^{-1} . Обычно, для анализа вклада элементов вторичной структуры используется фитинг полосы Amide I. В качестве альтернативы, для одновременного анализа всех образцов, МГК был применен к (предварительно нормированному) диапазону (показанному на рисунке 2.6) 1640–1730 cm^{-1} их спектров, содержащему полосу Amide I.

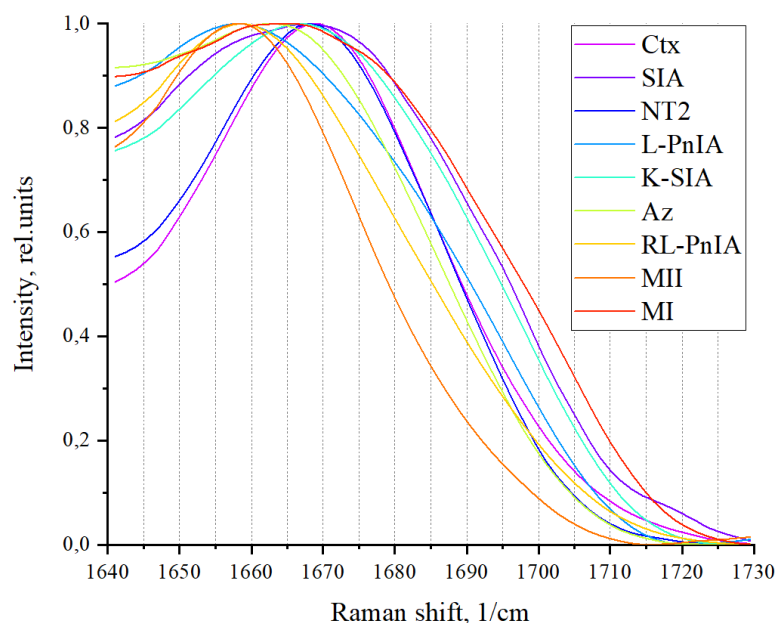


Рисунок 2.6 – Спектры КР в диапазоне 1640–1730 cm^{-1} , включающие полосу Amide I. Кривые были дополнительно сглажены и нормализованы. Часть кривых смещена в сторону больших частот (Ctx и NT2), а часть – в сторону меньших (например, MII и L-PnIA).

Далее, для объективного разделения образцов на группы, была использована k-means кластеризация. Результаты МГК и кластеризации показаны на рисунке 2.7.

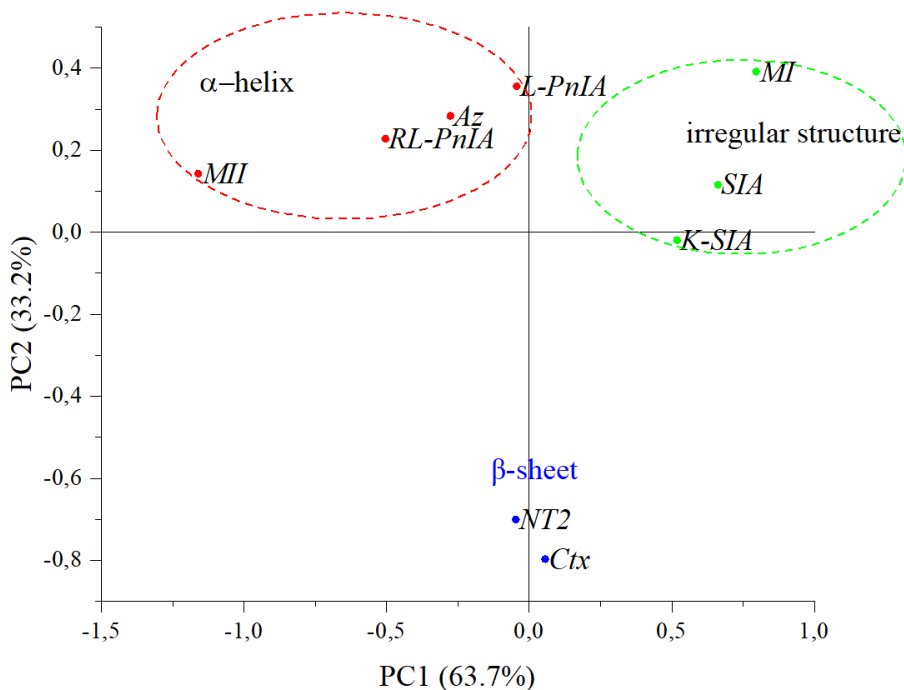


Рисунок 2.7 – График МКГ, полученный после анализа спектров в диапазоне 1640–1730 см⁻¹.

Показана кластеризация белков и пептидов (разные кластеры отличаются цветом) в соответствии с особенностями вторичной структуры. Изначально задавалось количество кластеров, равное трем, в соответствии с возможными типами структур: α -helix – α -спираль, β -sheet – β -слой, irregular structure – нерегулярная структура.

Было проанализировано влияние различных спектральных составляющих на значения ГК1 и ГК2. Для этого был построен график весов (loadings) (Рисунок 2.8 (А)). Поскольку значение (score) каждой ГК представляет собой взвешенную сумму исходных сигналов КР для разных частот (рамановских сдвигов), рассмотрение значений весов для разных частот спектра позволяет оценить влияние последних на величины КГ. Вместо представления весов, приведенного на рисунке 2.8 (А), можно использовать другую их репрезентацию, как на рисунке 2.8 (Б).

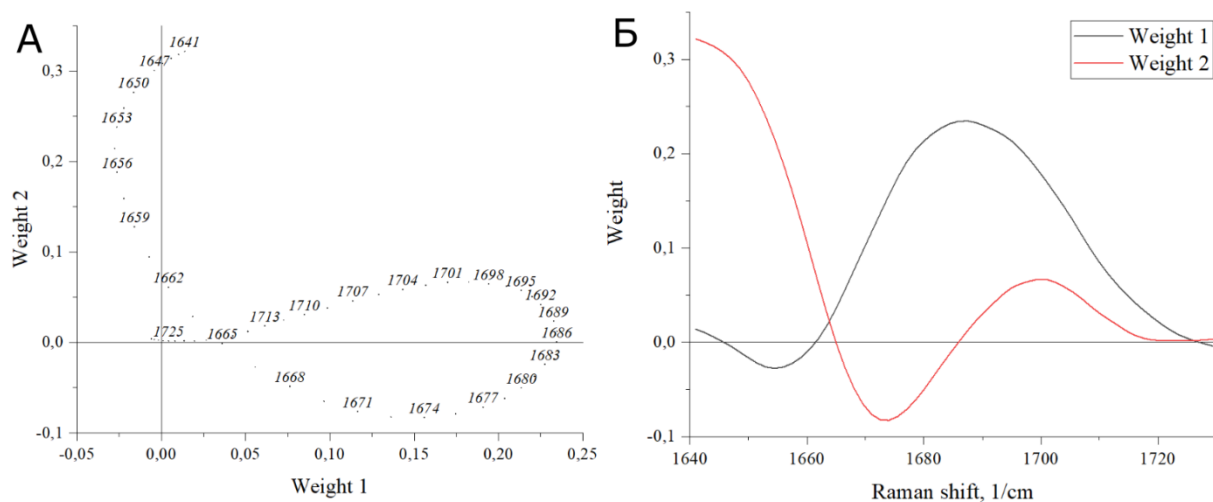


Рисунок 2.8 – Графический анализ весов. (А): График весов (loadings) – точки соответствуют различным рамановским сдвигам. Некоторые точки не подписаны, чтобы избежать нагромождения. По осям отложены веса (weight), относящиеся к ГК1 и ГК2 (Weight 1 и Weight 2 соответственно); (Б): Спектры весов ГК, отражающие влияние различных частот на значения ГК.

Анализ графика на рисунке 2.8 (А) приводит к следующим выводам. Наличие в спектре характерных для β -структур компонент около $1668\text{--}1677\text{ см}^{-1}$ ответственно за отрицательное значение ГК2. Сравнение графика весов (Рисунок 2.8 (А)) с графиком ГК (Рисунок 2.7) позволяет сделать вывод о наличии β -слоев в структуре Stx и NT2.

Положительные значения ГК2 вместе с отрицательными или близкими к нулю значениями ГК1 определяются компонентами спектра в районе частот $1650\text{--}1660\text{ см}^{-1}$, характерных для α -спиралей. Следовательно, можно утверждать, что образцы красного кластера на рисунке 2.7 содержат α -спирали. Токсины MI, SIA, K-SIA зеленого кластера, вероятно, имеют нерегулярную структуру. Разумно предположить, что бóльшие отрицательные значения ГК1 указывают на бóльшую общую длину α -спиралей, а близкие к нулю или положительные значения указывают на меньшую длину α -спиралей или их отсутствие. Это подтверждается тем, что трехпетельные токсины, в частности Stx [122] и NT2 [123], действительно имеют в своей структуре либо короткие (длиной в несколько аминокислотных остатков) α -спиральные элементы, либо не имеют их вовсе.

Данные о вторичной структуре токсинов, полученные в результате анализа спектров, были сопоставлены с результатами биоинформатического анализа α -конотоксинов MI, SIA, MII и PnIA. Полученные структуры и соответствующие графики Рамачандрана показаны на рисунке 2.9. Из этих данных следует, что в PnIA и MII α -спираль выражена достаточно ярко, тогда как в MI и SIA α -спираль энергетически невыгодна. Кроме того, на основании значений

интенсивности на тепловой карте можно предположить, что длина α -спирали в структуре MII превышает длину α -спирали в структуре PnIA.

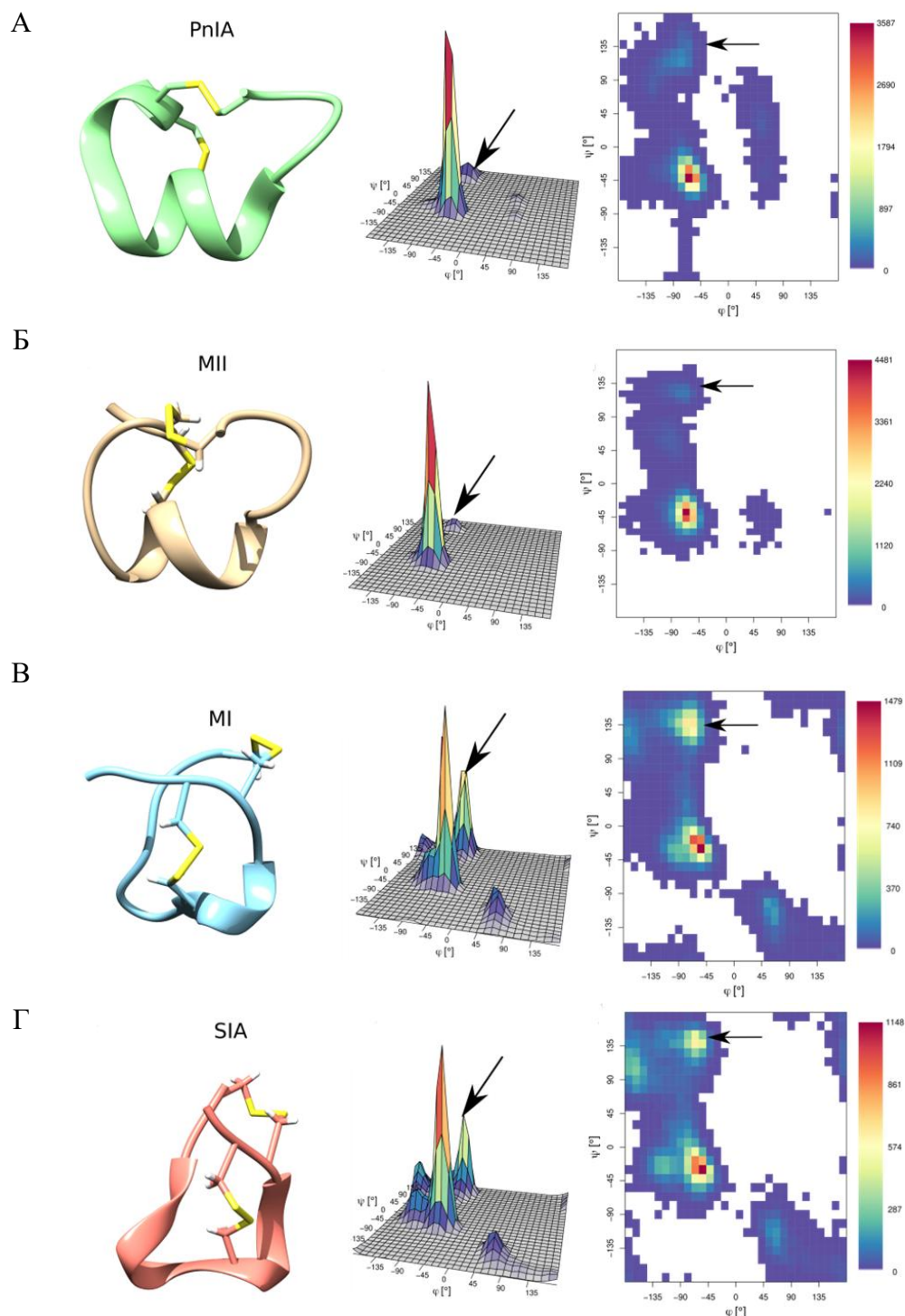


Рисунок 2.9 – Анализ структур PnIA, MII, MI и SIA с использованием молекулярной динамики и графиков Рамачандрана. (А): пространственная структура PnIA (слева); трехмерный график Рамачандрана PnIA, показывающий распределение углов ϕ и ψ пептидных связей (посередине) и график Рамачандрана PnIA в стиле тепловой карты (справа); (Б): аналогичные данные для MII; (В): аналогичные данные для MI; (Г): аналогичные данные для SIA. Стрелками показаны области углов ϕ и ψ , которые различаются между двумя группами α -конотоксинов.

Интересно, что, согласно результатам кластеризации, Az содержит α -спираль. Ранее структура Az была предсказана сервисом PEPFOLD [124], и это моделирование также выявило наличие спирали.

2.2.3. Сравнение спектров КР изомеров конотоксина RgIA

Было проверено, можно ли выявить различия в строении двух дисульфидных изомеров α -конотоксина RgIA – глобулярного (RgIA) и ленточного (isoRgIA) – с помощью описанного выше метода. Аминокислотный состав обоих образцов одинаков, но в ленточной форме isoRgIA дисульфидные связи (Cys2–Cys12 и Cys3–Cys8) отличаются от таковых (Cys2–Cys8 и Cys3–Cys12) в глобулярной форме.

Разница в дисульфидных связях проявилась при анализе области S–S спектра КР RgIA и isoRgIA (Рисунок 2.10 (А)). Из рисунка 2.10 (А) видно, что пик спектра RgIA значительно уширен (относительно пика спектра isoRgIA): для RgIA значителен вклад колебаний в области 520–550 см^{-1} , что указывает на наличие *gouche-gouche-trans* и/или *trans-gouche-trans* конформаций C–C–S–S–C–C. В то же время, в isoRgIA доминирует конформация *gouche-gouche*, о чем свидетельствует более узкий пик с максимумом около 513 см^{-1} .

Различие в дисульфидных сшивках (мостиках) может приводить к структурным изменениям, которые, в свою очередь, могут проявляться в полосе Amide I спектра КР. На рисунке 2.10 (Б) показаны спектры RgIA и isoRgIA в диапазоне 1640–1730 см^{-1} . Видно, что максимум полосы в изомере заметно сдвинут относительно максимума нативной формы RgIA. Это указывает на то, что изменение дисульфидных мостиков приводит к изменению вторичной структуры этого конотоксина.

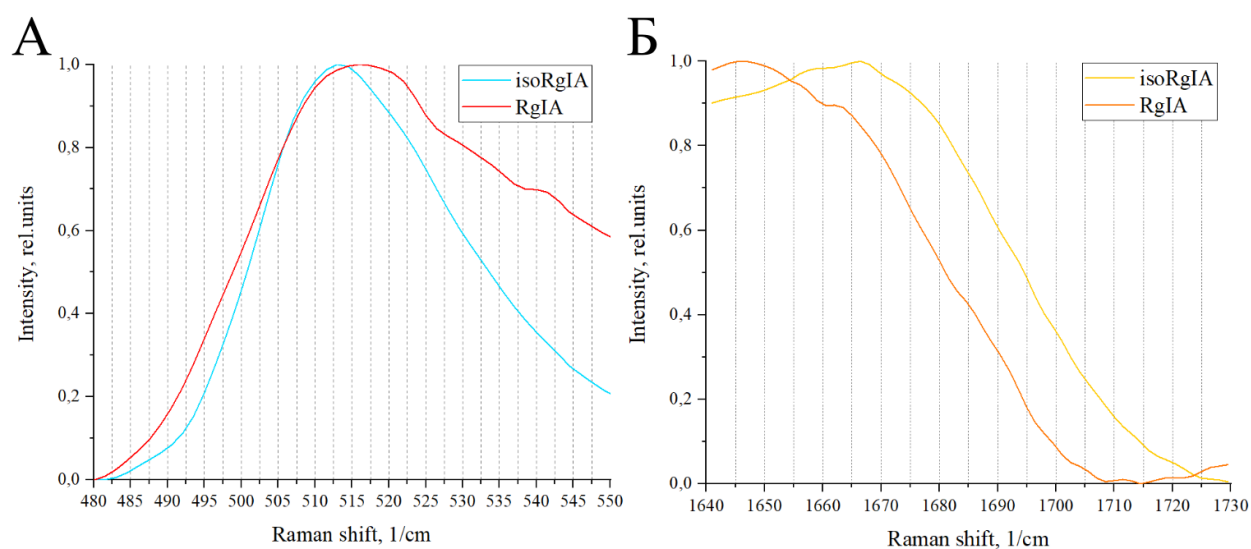


Рисунок 2.10 – (А): Область S–S спектров RgIA и isoRgIA; (Б): спектры RgIA и isoRgIA в диапазоне 1640–1730 cm^{-1} (Amide I).

Спектры RgIA и isoRgIA, показанные на рисунке 2.10 (Б), были внедрены в модель, использованную и протестированную в разделе 2.2.2. Результаты анализа представлены на рисунке 2.11. Видно, что при добавлении RgIA и isoRgIA и исключении из рассмотрения змеиных токсинов общая картина группировки, полученная ранее (Рисунок 2.7), сохраняется: α -конотоксины разделяются на два кластера.

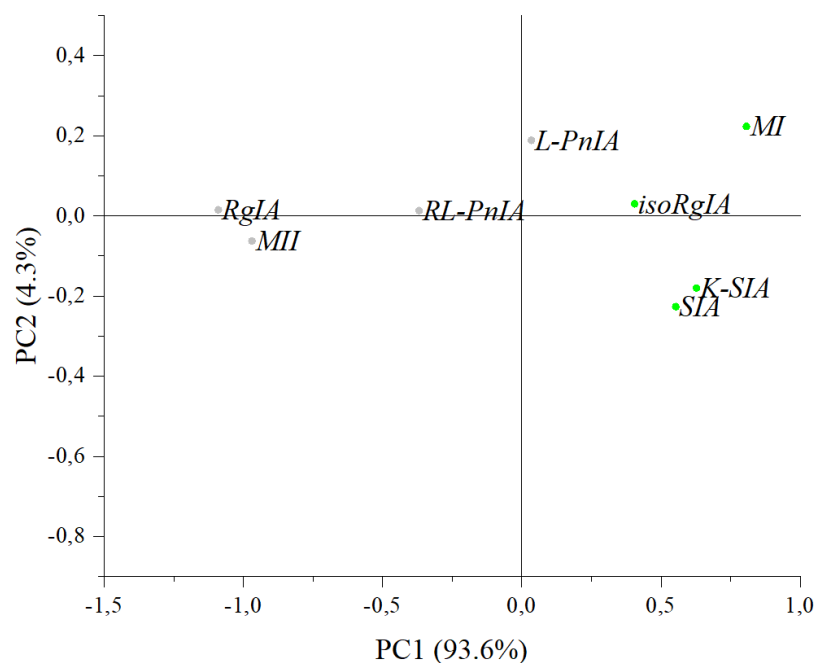


Рисунок 2.11. Результаты МГК примененного к спектральному региону 1640–1730 cm^{-1} ранее проанализированных конотоксинов и дополнительно добавленных конотоксинов (RgIA и его изомер).

Экстраполируя на новые образцы выводы, сделанные в предыдущем разделе (относительно вторичной структуры токсинов в двух кластерах), можно утверждать, что глобулярная форма RgIA имеет α -структурные элементы, а в его изомере наиболее выражена нерегулярная структура. Вероятно, для изомера α -спиральная конфигурация становится менее выгодной.

Следует отметить, что некоторые выводы относительно вторичной структуры можно сделать непосредственно на основе прямого рассмотрения полосы Amide I, однако при использовании МГК анализ становится более точным и наглядным.

2.2.4. Классификация спектров КР трехпетельных токсинов

При помощи описанной методологии также была проведена классификация белков семейства трехпетельных токсинов. Эти токсины имеют различную вторичную и петлевую (стабилизированную дисульфидными связями) структуры. Для проведения классификации на основе этих двух структурных элементов (геометрии дисульфидов и вторичной структуры) в качестве входных данных использовались два спектральных диапазона: $490\text{--}545\text{ см}^{-1}$ и $1640\text{--}1730\text{ см}^{-1}$ (Рисунок 2.12). Первый диапазон отражает частоты S–S-колебаний и, соответственно, может содержать информацию о структуре петель в белках; во второй участок спектра входит полоса Amide I.

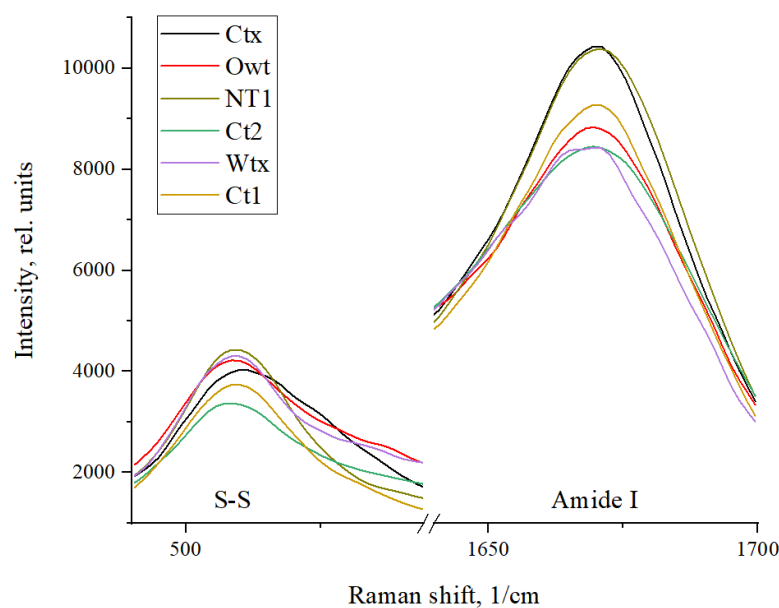


Рисунок 2.12 – Структурно значимые диапазоны (область S–S и полоса Амиды I) спектров трехпетельных токсинов.

Результаты классификации представлены на рисунке 2.13. Даже визуально можно четко отличить друг от друга три отдельные группы токсинов: длинноцепочечные α -нейротоксины, цитотоксины и слабые токсины. Данная группировка отражает особенности трехмерного строения образцов.

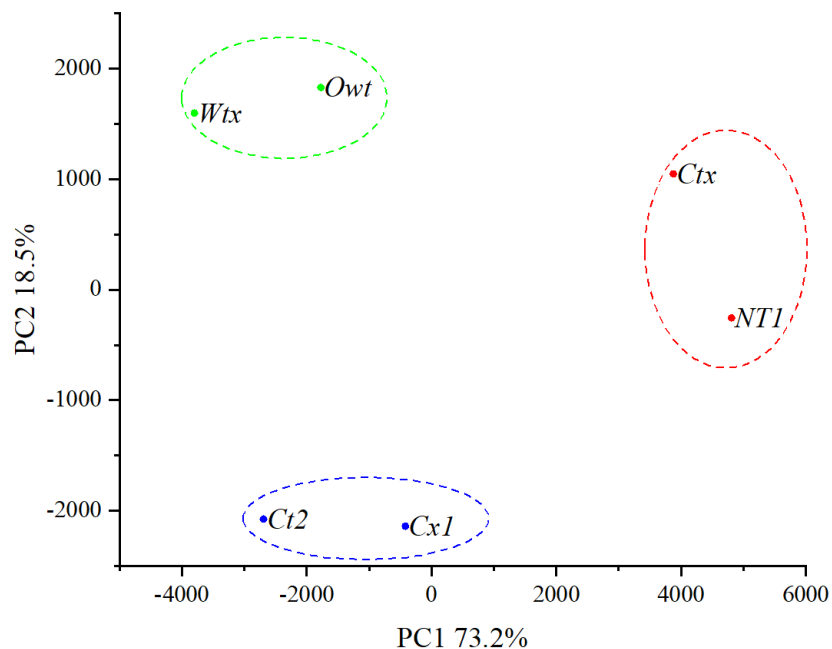


Рисунок 2.13 – Результаты МГК, примененного к двум спектральным диапазонам $490\text{--}545\text{ см}^{-1}$ и $1640\text{--}1730\text{ см}^{-1}$.

2.2.5. Сравнение триптофанового дублета для растворов и сухих образцов

Отметим, что для водных растворов токсинов соотношение интенсивностей I_{850}/I_{830} практически не отличалось от I_{850}/I_{830} для лиофилизированных образцов, что говорит в пользу корректности экстраполяции данных, полученных на сухих образцах, на их растворы. В таблице 2.2 представлены эти данные для четырех отличающихся друг от друга токсинов.

Таблица 2.2. Величины отношений I_{850}/I_{830} .

Образцы	SIA	RL-PnIA	Cx1	Ctx
Капли водных растворов токсинов	$0,9\pm 0,1$	$1,2\pm 0,1$	$0,8\pm 0,1$	$0,8\pm 0,1$
Лиофилизированные токсины	$0,8\pm 0,1$	$1,2\pm 0,1$	$0,9\pm 0,1$	$0,8\pm 0,1$

2.3. Выводы к главе 2

В данной части исследования продемонстрировано, что РС в сочетании с МГК и кластеризацией может успешно применяться для выявления структурного сходства/различий между белковыми и пептидными токсинами, включая их синтетические аналоги. Это обеспечивается за счет экономичного экспресс-анализа этих веществ в сухом виде с хорошей воспроизводимостью результатов. На основе предложенного метода можно проводить предварительный анализ новых, менее изученных токсинов, путем сравнения (группировки) их с известными токсинами, что позволяет экстраполировать особенности строения исследованных токсинов на новые.

Также удалось выявить конкретные особенности вторичной структуры токсинов. Полученные результаты были подтверждены биоинформатическим анализом. Принимая во внимание небольшой расход образцов, необходимых для анализа методом РС, предлагаемая методика может стать первым шагом в изучении токсинов редких и/или находящихся под угрозой исчезновения ядовитых животных.

Кроме того, показано, что на основе разработанной методики можно различать дисульфидные изомеры одного и того же пептидного конотоксина, что важно для синтеза правильного (нативного) его изомера.

Глава 3. Классификация ядов разных видов змей с помощью РС

В данной главе описывается применение РС для классификации, с помощью МГК и методов кластеризации, ядов различных видов змей из семейств Elapidae и Viperidae.

3.1. Материалы и методы

3.1.1. Образцы ядов

Виды змей, яды которых были изучены в данном исследовании, и их географическое происхождение приведены в таблице 3.1.

Таблица 3.1. Виды змей и их географическое происхождение.

Аббревиатура	Вид змеи	Происхождение	Семейство
AgCo	<i>Agkistrodon contortrix</i>	Неизвестно	Viperidae (V)
GlSa	<i>Gloydius saxatilis</i>	Приморский край, Россия	V
BuFa	<i>Bungarus fasciatus</i>	Вьетнам	Elapidae (E)
BuMu	<i>Bungarus multicinctus</i>	Вьетнам	E
CrAd	<i>Crotalus adamanteus</i>	Юго-восток США	V
DeAn	<i>Dendroaspis angusticeps</i>	Кения	E
EcCa	<i>Echis carinatus</i>	Туркменистан	V
EcMu	<i>Echis multisquamatus</i>	Неизвестно	V
MaLeO	<i>Macrovipera lebetina obtusa</i>	Армения	V
NaNa	<i>Naja naja</i>	Шри-Ланка	E
NaKa	<i>Naja kaouthia</i>	Неизвестно	E
ViNi	<i>Vipera nikolskii</i>	Район р. Красивая Меча, Россия	V
ViRa	<i>Montivipera raddei</i>	Армения	V
WaAe	<i>Walterinnesia aegyptia</i>	Иран	E
NaOx	<i>Naja oxiana</i>	Туркменистан	E
ViAm	<i>Vipera ammodytes</i>	Болгария	V
ViUr	<i>Vipera ursinii</i>	Краснодарский край, Россия	V

Яды крайтов *Bungarus fasciatus* и *B. multicinctus* были получены из авторизованной, лицензированной фермы по разведению змей и производству ядов, расположенной во Вьетнаме. Эти яды были собраны у нескольких десятков особей змей и объединены. Яд *Crotalus adamanteus* был любезно предоставлен профессором Т. Morita (Фармацевтический университет Мэйдзи, Токио, Япония), яд *Dendroaspis angusticeps* был предоставлен доктором Е. Karlsson (Университет Уппсалы, Швеция), яд вида *Macrovipera lebetina obtusa* получен от профессора Н. Айвазян (Институт физиологии Орбели, Ереван, Армения). Число змей, от которых были получены эти яды, неизвестно.

Яды кобр (*Naja kaouthia*, *N. naja*, *N. oxiana* и *Walterinnesia aegyptia*), гадюк (*Vipera ammodytes*, *V. nikolskii*, *V. ursinii*, *Echis carinatus* и *E. multisquamatus*) и ямкоголовых змей (*Agkistrodon contortrix* и *Gloydius saxatilis*) были получены путем доения змей, содержащихся в неволе. Яды кобр *N. kaouthia*, *N. naja* и *N. oxiana* были собраны не менее чем у 5 особей змей и объединены. Яд кобры *Walterinnesia aegyptia* был собран после нескольких доений 2 особей. Яды гадюк были собраны у не менее чем 10 змей и объединены. Яды ямкоголовых змей были собраны после нескольких доений 3 особей змей. Полученные яды были высушены на обезвоженном CaCl_2 и хранились при -20°C до использования. Все процедуры со змеями были одобрены комиссией по содержанию и использованию лабораторных животных Института биоорганической химии им. Шемякина-Овчинникова РАН (Протокол №324/2021 от 23 июня 2021 г.).

3.1.2. Сбор данных

Условия записи спектров были аналогичны таковым, изложенным в пункте 2.1.2, за исключением того, что при записи каждого спектра применялось усреднение по 200 спектрам.

3.1.3. Обработка спектральных данных

Предобработка спектров включала следующие этапы: 1) коррекцию базовой линии («concave rubberband correction»), 2) сглаживание и 3) скалирование (минимальный сигнал приравнивался 0, максимальный – 1). 1-й и 2-й шаги выполнялись в OPUS 8.2.28, 3-й шаг – в PyChem 3.0.5g Beta. Поскольку полоса Amide I имела максимальную интенсивность во всех спектрах образцов, скалирование было эквивалентно нормализации к пику Amide I.

Дальнейший анализ МГК и DFA (точнее, ГК-DFA, то есть DFA, примененный к ГК) также проводились в PyChem. Число ГК = 3, процент объясненной дисперсии для 1, 2 и 3 ГК составил 50,5%, 18,9% и 7,8% соответственно. Для обучения алгоритма ГК-DFA из 17 образцов было отобрано 7 (3 и 4 образца ядов Elapidae и Viperidae соответственно).

K-means кластеризация была выполнена в Python 3.9 с использованием библиотеки «Scikit-learn». Количество кластеров определялось методом «elbow» [125] (иногда используется название «knee») с помощью пакета «Knee».

Аппроксимацию кривыми (Curve-fitting) проводили в программе Origin (Версия 2018. OriginLab Corporation, Нортгемптон, Массачусетс, США) с помощью Лоренциана. Кривые, представляющие α -, нерегулярную и β -структуру, центрировались вблизи 1654 см^{-1} , 1662 см^{-1} и 1670 см^{-1} соответственно. Метод «Residual after First Derivative» был использован для поиска пиков. Вклад (%) различных вторичных структур определялся путем деления площади под соответствующим пиком на сумму площадей под всеми пиками (площадь суммарного подгоночного пика).

3.2. Результаты

3.2.1. Применение МГК к спектрам КР ядов для их классификации

Предобработанные спектры сухих образцов яда представлены на рисунке 3.1.

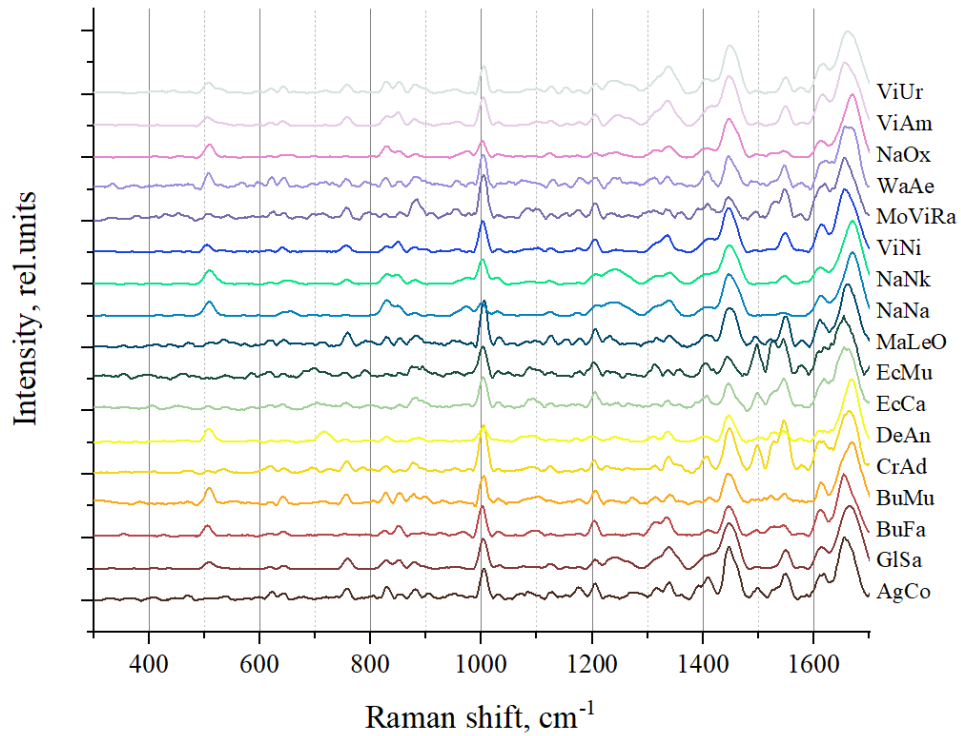


Рисунок 3.1 – Предобработанные спектры КР всех образцов.

Спектры КР были подвергнуты МКГ. Итоговый график МКГ показан на рисунке 3.2, где для ясности представлены две ГК.

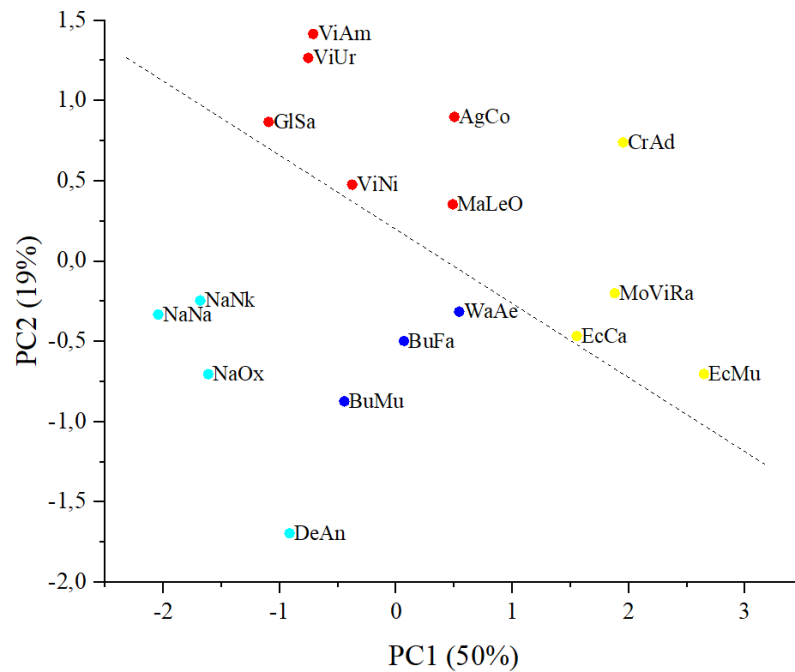


Рисунок 3.2 – График МКГ, где в скобках указан процент объясненной дисперсии для каждой ГК. Четыре кластера (полученные методом k-means кластеризации) различаются по цвету. Пунктирной линией показано разделение образцов на яды Elapidae (под линией) и Viperidae (над линией).

Далее, чтобы идентифицировать группы ядов, был применен метод k-means кластеризации. Поскольку данная методика требует априорного указания количества кластеров, был использован метод «Elbow», который по динамике суммарной квадратичной ошибки (sum of the squared error, SSE) (Рисунок 3.3) позволил определить число кластеров (точнее, их оптимальное количество, при дальнейшем увеличении которого снижение SSE уже незначительно).

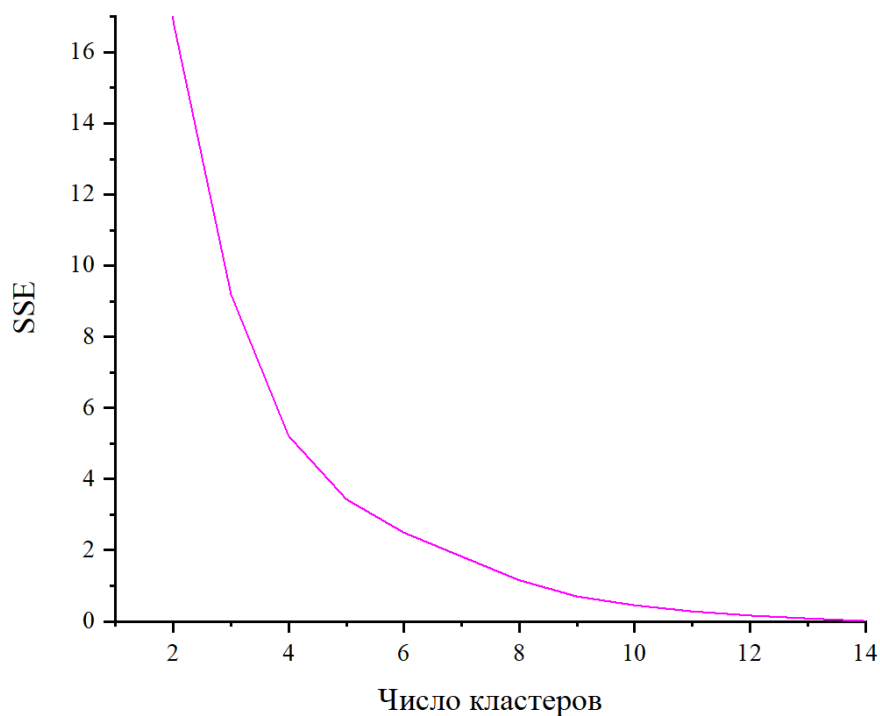


Рисунок 3.3 – Динамика уменьшения суммарной квадратичной ошибки (SSE) с увеличением числа кластеров.

В результате, образцы были разделены на четыре кластера, которые показаны разными цветами на рисунке 3.2. Кластер 1 (красный) включает яды представителей родов *Vipera*, *Agkistrodon*, *Gloydius*, а также гюрзы (MaLeO). Кластер 2 (желтый) включает образцы змей родов *Echis*, *Montivipera*, а также ромбического гремучника (CrAd). Все яды кластера 2 объединяет преобладание такого компонента, как SVMP [86]. Яд CrAd довольно отдален от других образцов этой группы, что можно объяснить тем, что вид CrAd является эндемиком США и его яд имеет специфические особенности. Кластер 3 (синий) объединяет яды крайтов (род *Bungarus*) и яд египетской пустынной кобры (WaAe). Интересно, что в отличие от яда большинства Elapidae, в яде некоторых представителей рода *Bungarus* доминирует PLA2. Вероятно, эта особенность отличает образцы ядов 3-й и 4-й групп. Группа 4 (голубая) состоит из ядов кобр и яда мамбы (DeAn), основными компонентами которых являются 3FTx. Внутри

этой группы яды кобр довольно четко выделяются в отдельную подгруппу, тогда как яд DeAn достаточно далек от них.

3.2.2. Применение ГК-DFA для разделения спектров КР ядов по семействам змей

Из графика на рисунке 3.2 видно, что образцы можно разделить на 2 большие группы, соответствующие ядам Elapidae и Viperidae; пунктирная линия, пересекающая рисунок, иллюстрирует такое разбиение. Более строгое разделение можно провести, применив метод DFA. Была использована модификация этого метода – ГК-DFA: в этом случае анализу подвергаются новые переменные (координаты образцов в пространстве ГК).

Около 40% образцов (Train) были использованы для обучения алгоритма (Таблица 3.2), то есть было заранее указано, к какому семейству (Elapidae (E) или Viperidae (V)) они принадлежат. Для обеспечения разнообразия обучающей выборки из ядов семейств Elapidae и Viperidae были отобраны соответственно 3 и 4 тренировочных образца. После этого на основе уже обученной модели остальные образцы (Test) были автоматически отнесены к одному из двух семейств. Полученные результаты показаны на рисунке 3.4 (обратите внимание, что на этом графике значимо только расстояние между образцами по вертикали; расстояние по горизонтальной оси не имеет значения: образцы разделены горизонтальными расстояниями во избежание их перекрытия).

Таблица 3.2. Распределение образцов по обучающим и тестовым наборам (E – Elapidae, V – Viperidae).

Образец	Семейство	Набор
AgCo		Test
GlSa		Test
BuFa	E	Train
BuMu		Test
CrAd	V	Train
DeAn		Test
EcCa	V	Train

EcMu		Test
MaLeO		Test
NaNa	E	Train
NaNk		Test
ViNi	V	Train
MoViRa		Test
WaAe	E	Train
NaOx		Test
ViAm	V	Train
ViUr		Test

Видно, что все тестовые образцы были правильно отнесены к одной из двух групп (семейств): яды Elapidae имеют отрицательные значения координаты ГК-DFA, а яды Viperidae – положительные. Нам было интересно посмотреть, какие области спектра ответственны за такое разделение. Эту информацию можно получить из анализа весов ГК-DFA, которые, по аналогии с весами в МГК, показывают влияние различных частот спектров КР на координату (score) ГК-DFA (Рисунок 3.5). Из рисунка 3.5 видно, что частотам, ответственным за отнесение яда к семейству Elapidae, то есть которые имеют отрицательный вес, соответствуют пики в районах 510 и 1670 cm^{-1} .

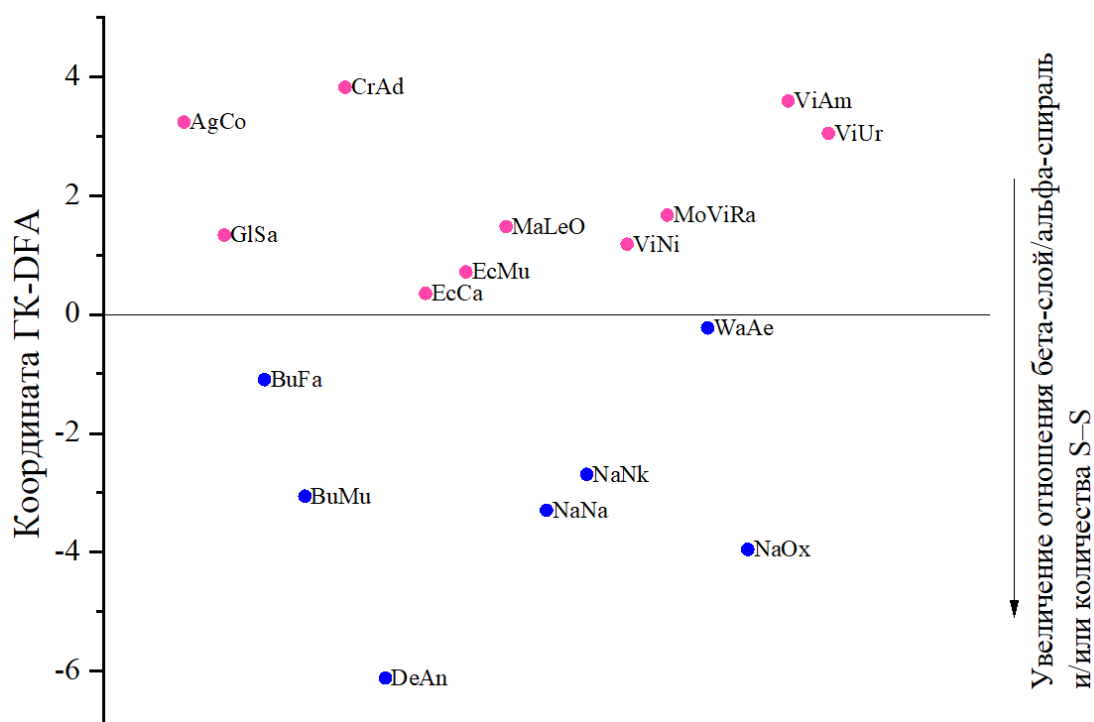


Рисунок 3.4 – График ГК-DFA. Горизонтальная линия при нулевом значении координаты ГК-DFA разделяет образцы на семейства. По горизонтали образцы разделены расстояниями, чтобы избежать наложения. Синим и розовым цветами обозначены яды семейств E и V соответственно. Указано направление увеличения соотношения β -слой/ α -спираль и/или увеличения количества дисульфидных связей (S–S).

Белки являются преобладающим компонентом яда и поэтому вносят основной вклад в спектры исследованных образцов (резонансных эффектов, существенно усиливающих спектры каких-либо других компонентов яда, обнаружено не было). Следовательно, спектр яда можно интерпретировать как суперпозицию спектров различных белков, входящих в его состав. Как уже говорилось ранее, область $490\text{--}550\text{ см}^{-1}$ представляет собой характеристическую полосу дисульфидных связей, а колебания на частоте около 510 см^{-1} соответствуют основной (*gouche-gouche*) конформации C–C–S–S–C–C. Следовательно, пик на этой частоте в спектре КР указывает на наличие дисульфидных связей в образце (белке), а интенсивность пика отражает их количество.

Колебания в области Amide I с частотой около 1670 см^{-1} характерны для β -слоев в структуре. Колебания же на частоты около 1655 см^{-1} характерны для α -спиральных структурных элементов в белке. Как видно из весов ГК-DFA, наличие α -спирали увеличивает «вероятность» отнесения образца к семейству Viperidae.

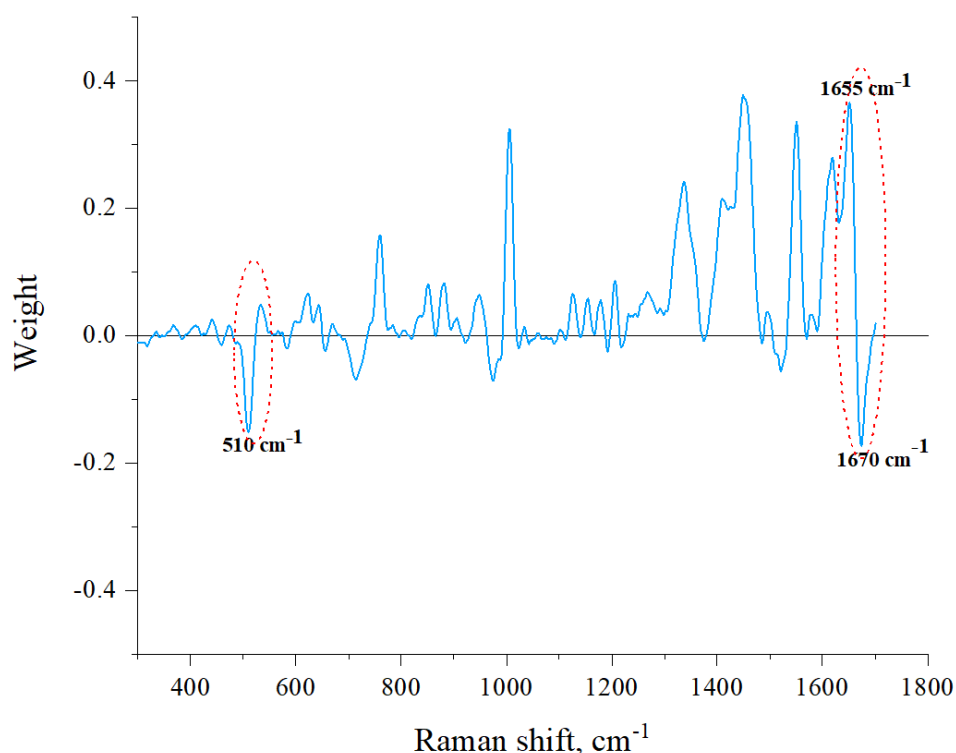


Рисунок 3.5 – Спектр весов ГК-DFA. Обведены полосы дисульфидов и Amide I. Частоты, ответственные за отнесение яда к семейству Elapidae (510 см⁻¹ – основная конформация C–C–S–S–C–C и 1670 см⁻¹ – β-слой) и частота, характерная для α-спирали (1655 см⁻¹) подписаны.

Поскольку других достаточно интенсивных отрицательных пиков весов не наблюдается (Рисунок 3.5), то можно предположить, что основными критериями отнесения яда к семейству E или V являются количество дисульфидных связей и содержание β-слоев. Чем больше этих структурных элементов присутствует яде, тем больше отрицательный вклад в значение GK-DFA координаты (score) данного яда и тем больше шанс, что он будет отнесен к семейству E. Можно также предположить, что чем ниже на графике рисунка 3.4 расположен образец яда, тем больше относительное количество S–S связей и/или β-слоев в белках, входящих в его состав (соответствующее направление отмечено на рисунке 3.4 стрелкой).

Эти результаты можно интерпретировать с точки зрения количественного содержания конкретных семейств белков в ядах. Поскольку в ядах Elapidae обычно доминируют ЗФТх [126], большее количество β-структурных элементов в образцах этих ядов можно связать именно с присутствием ЗФТх. Известно [2, 127], что в структуре большинства этих токсинов доминирует β-слой, образующие петли.

В то же время, особенностью некоторых ядов змей рода *Bungarus* семейства Elapidae является относительно низкое содержание ЗФТх с преобладанием PLA2 [86]. В частности, основные компоненты яда BuFa из Вьетнама следующие: 67 % PLA2 и 7% LAAO. Поскольку в структуре PLA2 преобладают α-спирали [128], возможно отнести увеличение количества α-

структур именно на счет этих белков. Это также отражено на графике рисунка 3.4: образец BuFa из семейства Elapidae близок к границе с Viperidae. Из данных этого рисунка можно дополнительно сделать вывод, что соотношение β -/ α -структур также снижено в яде WaAe по сравнению с другими образцами ядов семейства Elapidae. Обратная ситуация наблюдается для DeAn, который действительно характеризуется чрезвычайно высоким содержанием ЗФТх [86].

3.3.3. Прямой анализ полос Amide I и S–S спектров образцов

Приведенные выше результаты могут быть подтверждены и прямым анализом тех участков спектров КР, которые были идентифицированы как ответственные за разделение образцов на группы E и V методом ГК-DFA. Выводы о вторичной структуре можно сделать на основании данных о положении максимума полосы Amide I (Рисунок 3.6 (А), Таблица 3.3). Смещение максимума в сторону бóльших частот ($>1665 \text{ см}^{-1}$) характерно для всех ядов E, за исключением BuFa и WaAe, о которых упоминалось выше. В случае S–S области (Рисунок 3.6 (Б)) можно заметить, что действительно, как уже говорилось выше, яды E имеют здесь более интенсивный пик (при $\sim 510 \text{ см}^{-1}$), то есть большее количество дисульфидных связей. Нам удалось определить граничную интенсивность пика S–S области ($\sim 0,18$ отн. ед., горизонтальная прямая на рисунке 3.6 (Б)), выше которой находятся максимумы спектров ядов E, а ниже – ядов V. Таким образом, можно утверждать, что по количеству дисульфидных связей в яде можно определить, к какому семейству принадлежит продуцирующая его змея, что априори не очевидно.

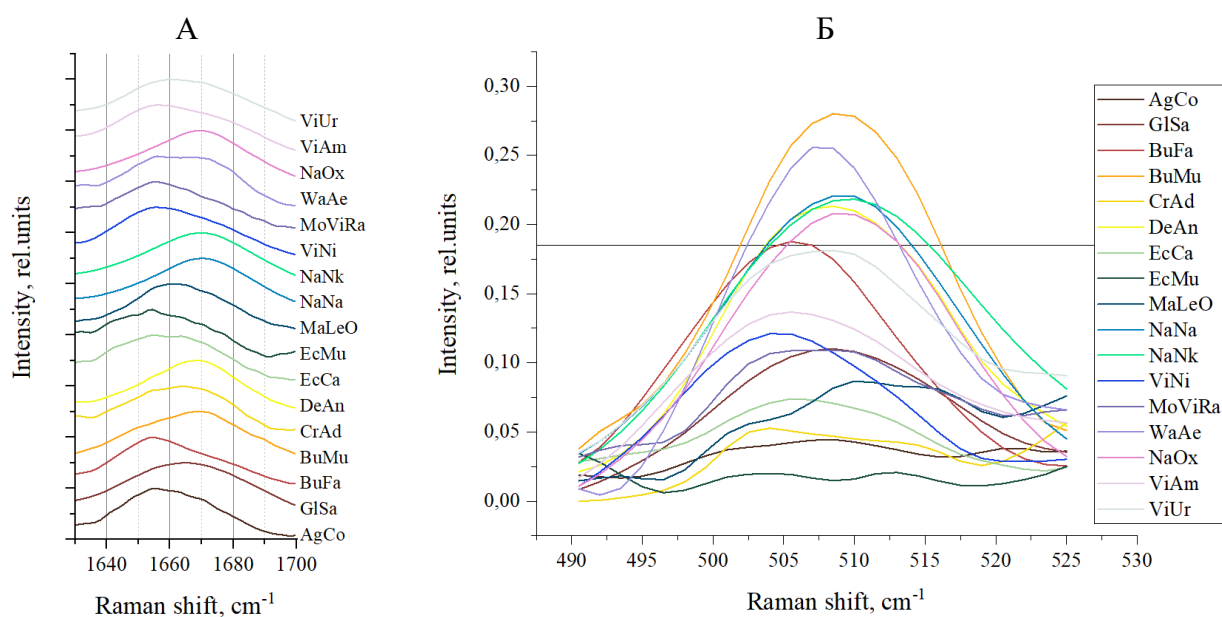


Рисунок 3.6 – (А): полоса Amide I образцов; (Б): полоса S–S.

Следует отметить, что МГК с k-means кластеризацией и ГК-DFA обеспечивают группировку ядов (рисунки 3.2 и 3.4) на основе анализа всех областей спектров и, таким образом, принимают во внимание больше особенностей образцов, чем выявляется при прямом анализе конкретных участков спектров (Рисунок 3.6). Дополнительно, графики, полученные в результате кластеризации, объединяют все особенности спектров образцов в удобном визуальном формате.

Таблица 3.3. Положения максимумов пиков полосы Amide I.

Образец	Семейство змеи	Положение максимума Amide I, cm^{-1}	Интенсивность S–S полосы, отн. ед.
AgCo	V	1654,5	0,04
GlSa	V	1664,5	0,11
BuFa	E	1654,5	0,19
BuMu	E	1669,5	0,28
CrAd	V	1663,5	0,05
DeAn	E	1669,5	0,21
EcCa	V	1654,5	0,07
EcMu	V	1654,5	0,02

MaLeO	V	1660,5	0,09
NaNa	E	1669,5	0,22
NaNk	E	1669,5	0,22
ViNi	V	1656	0,12
ViRa	V	1656	0,11
WaAe	E	1656	0,26
NaOx	E	1669,5	0,21
ViAm	V	1656	0,14
ViUr	V	1660,5	0,18

Для оценки вклада разных элементов вторичной структуры дополнительно была проведена аппроксимация полосы Amide I для каждого образца. Первоначально количество кривых для фитинга каждой полосы было выбрано равным 3 в соответствии со следующими возможными структурами: α -, нерегулярная и β -структура. В некоторых случаях метод обнаруживал меньшее количество скрытых пиков; тогда их количество сокращали до 2 или 1. Графические результаты анализа показаны на рисунке 3.7; частоты максимумов всех функций, используемых для аппроксимации полос Amide I, а также относительные площади под соответствующими той или иной структуре кривыми для каждого образца приведены в таблице 3.4. Типы вторичных структур, которые характерны для соответствующих частот, и количество подгоночных функций также указаны в данной таблице.

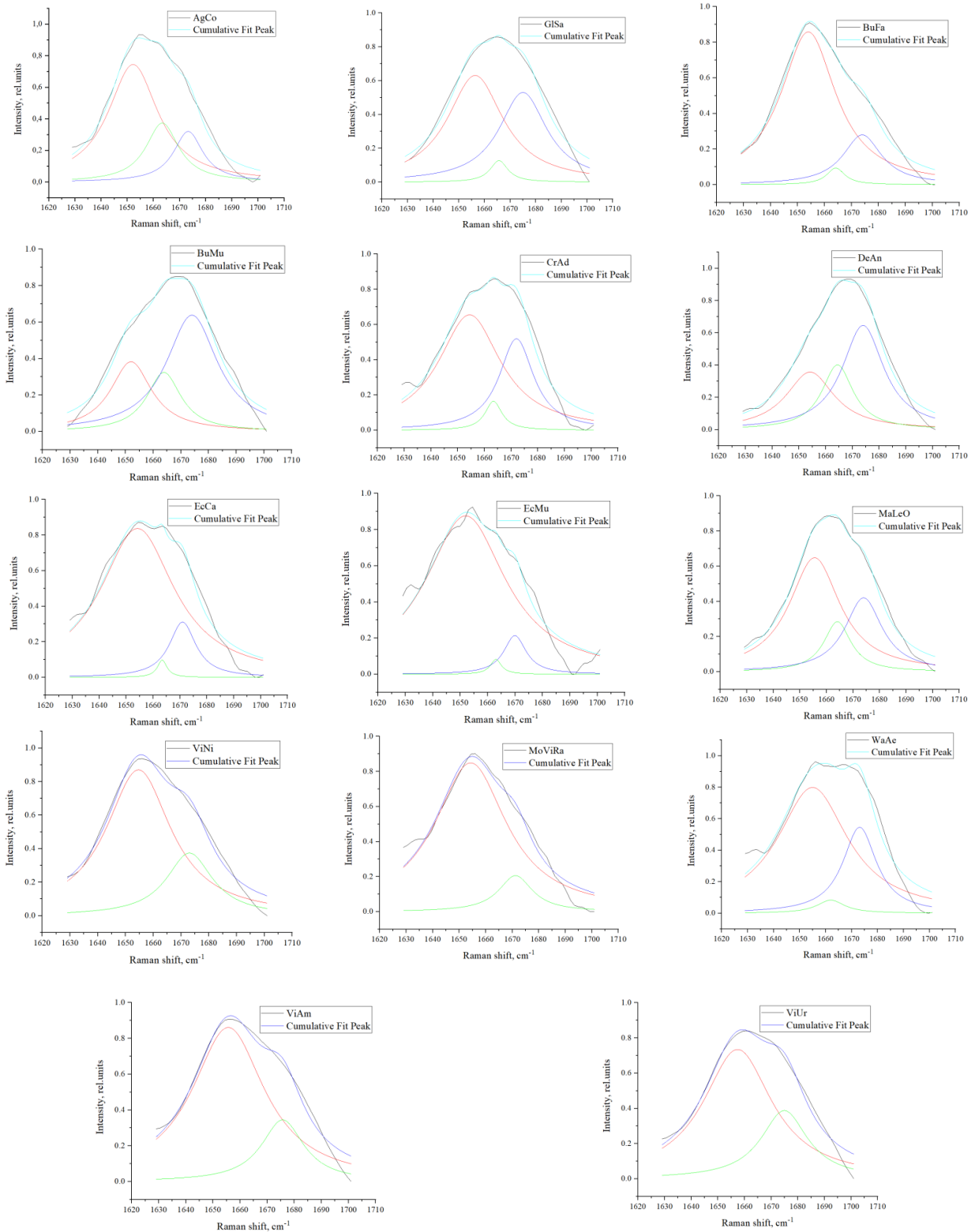


Рисунок 3.7 – Графические результаты аппроксимации кривыми. Cumulative Fit Peak – суммарный подгоночный пик (полученный суммированием всех подгоночных пиков); образцы NaNa, NaOx и NaNk аппроксимировались единственным пиком и не показаны.

Таблица 4.4. Данные по аппроксимации кривыми.

Образец	Число аппроксимационных кривых	Тип структуры	Частота, см ⁻¹	Площадь, %
AgCo	3	α-Спираль	1652	63
		Нерегулярная	1663	21
		β-Слой	1673	16
GlSa	3	α-Спираль	1656	56
		Нерегулярная	1666	4
		β-Слой	1675	40
BuFa	3	α-Спираль	1654	79
		Нерегулярная	1664	2
		β-Слой	1674	19
BuMu	3	α-Спираль	1652	26
		Нерегулярная	1664	19
		β-Слой	1674	55
CrAd	3	α-Спираль	1654	66
		Нерегулярная	1663	5
		β-Слой	1672	29
DeAn	3	α-Спираль	1654	30
		Нерегулярная	1664	22
		β-Слой	1674	48
EcCa	3	α-Спираль	1654	86
		Нерегулярная	1663	2
		β-Слой	1671	12
EcMu	3	α-Спираль	1652	92
		Нерегулярная	1663	2
		β-Слой	1670	6
MaLeO	3	α-Спираль	1656	59

		Нерегулярная	1664	15
		β -Слой	1674	26
NaNa	1	β -Слой	1669	100
NaNk	1	β -Слой	1669	100
ViNi	2	α -Спираль	1655	76
		β -Слой	1673	24
MoViRa	2	α -Спираль	1654	90
		β -Слой	1671	10
WaAe	3	α -Спираль	1655	72
		Нерегулярная	1662	4
		β -Слой	1673	24
NaOx	1	β -Слой	1668	100
ViAm	2	α -Спираль	1656	81
		β -Слой	1676	19
ViUr	2	α -Спираль	1657	73
		β -Слой	1675	27

3.3. Выводы к главе 3

Таким образом, нами было проведено спектроскопическое исследование цельных ядов различных видов змей из разных родов и семейств, которое показало применимость РС для исследования таких сложных белковых смесей, как яды, а также полезность методов уменьшения размерности и кластеризации для интерпретации сложных спектральных данных.

Образцы ядов были разделены на две группы, соответствующие семействам Viperidae и Elapidae ядопродуцирующих змей. Также была проведена более детальная классификация ядов на четыре группы, отражающая составы ядов. Можно утверждать, что подход позволяет оценивать составы ядов путем сравнения их с другими, уже охарактеризованными образцами.

Таким образом, нами предложен экспресс-метод, который позволяет анализировать большое количество ядов, а малый расход материалов (~10 мкг и менее) дает возможность

работать с небольшими количествами образцов, что важно при изучении ядов малопродуктивных змей, а также редких и исчезающих видов. Метод можно рассматривать как первый шаг в разработке новых методов определения типа яда по мазку с места укуса для правильного подбора противоядия для пострадавшего.

Глава 4. Анализ спектров КР ядов отдельных особей змей

В этой части работы подход, уже использованный ранее для дифференциации ядов различных видов змей и описанный в главе 3, применен для изучения более тонких – внутривидовых – вариаций ядов. В данном случае проанализированы уже не усредненные образцы, полученные путем объединения ядов нескольких особей змей, а яды отдельных змей вида *Vipera nikolskii*. Кроме того, проведено сравнение обнаруженных таким путем внутривидовых вариаций ядов с межвидовыми (относительно родственных видов *Vipera renardi* и *Vipera berus*).

4.1. Материалы и методы

4.1.1. Образцы ядов и фосфолипаз A2

Яды были получены от отдельных особей змей, отловленных в разных географических локациях, указанных в таблице 4.1. Все четыре локации достаточно удалены друг от друга, в частности, яды двух популяций гадюки Никольского *Vipera nikolskii* (VНК и VNB) были собраны у змей, обитающих в двух точках (Таблица 4.1), которые разделены расстоянием около 300 км и рекой Дон. Взрослых змей доили путем ручного массажа желез. Собранный яд сушили на обезвоженном хлориде кальция и хранили при -20°C . После того как яды были получены, змей выпускали в то же место, где они были пойманы.

Таблица 4.1. Географические локации отлова змей.

Вид змеи	Обозначение	Место отлова
<i>Vipera nikolskii</i>	VНК(1–9)	Район р. Красивая Меча, Тульская область
	VNB(1–4)	Район г. Борисоглебск, Воронежская область
<i>Vipera renardi</i>	VR1	Ейский район, Краснодарский край
<i>Vipera berus</i>	VB(1–3)	Старицкий район, Тверская область

Фосфолипазы A2 HDP1 и HDP2 были получены из яда *V. nikolskii* как описано в [129].

4.1.2. Сбор данных

Измерения спектров КР были проведены аналогично описанным в пункте 3.1.2 главы 3. При регистрации спектров образцов в широком диапазоне частот (Рисунок 4.1) спектральное разрешение составляло 4 см^{-1} ; при записи спектров для анализа полосы Amide I спектральное разрешение составляло $1,5 \text{ см}^{-1}$.

4.1.3. Обработка спектральных данных

Предобработка спектральных данных выполнялась аналогично описанной в пункте 3.1.3 главы 3, за исключением того, что при нормализации (скалировании) минимальный сигнал приравнивался 0, а максимальный – 2. Последующий анализ МГК проводился также в PyChem 3.0.5g Beta.

При количественной оценке изменчивостей ядов для расчета погрешностей определения средних расстояний на графике МГК использовали стандартное отклонение среднего.

Аппроксимация кривыми (фитинг) полос Amide I проводили, как и в главе 3, в программе Origin с использованием функции Лоренца. Кривые, соответствующие α -, β - и нерегулярной структурам, центрировались с областях $1650\text{--}1660 \text{ см}^{-1}$, $1668\text{--}1677 \text{ см}^{-1}$ и $1661\text{--}1667 \text{ см}^{-1}$ или $1678\text{--}1689 \text{ см}^{-1}$ соответственно. Для поиска пиков был использован метод «2nd derivative». Параметры y_0 , задающие высоты базовых линий пиков, были фиксированы. Координаты максимумов пиков были выбраны в соответствие с видимыми максимумом и плечами исходной спектральной кривой. Для оценки качества фитинга, использовалась величина R-Square. (Значение R-Square, близкое к 1, указывает на то, что учитывается большая часть дисперсии. Например, R-Square = 0,9 означает, что фитинг объясняет 90% общей вариации данных.)

4.2. Результаты

4.2.1. Применение МГК к спектрам КР образцов ядов

Были записаны спектры сухих ядов особей следующих змей: двух популяций гадюки Никольского (VN) (с территорий вблизи реки Красивая Меча (VНК) (9 образцов) и г. Борисоглебска (VNB) (4 образца)), а также степной гадюки *Vipera renardi* (VR) (1 образец) и обыкновенной гадюки *Vipera berus* (VB) (3 образца). Полученные предобработанные спектры КР показаны на рисунке 4.1.

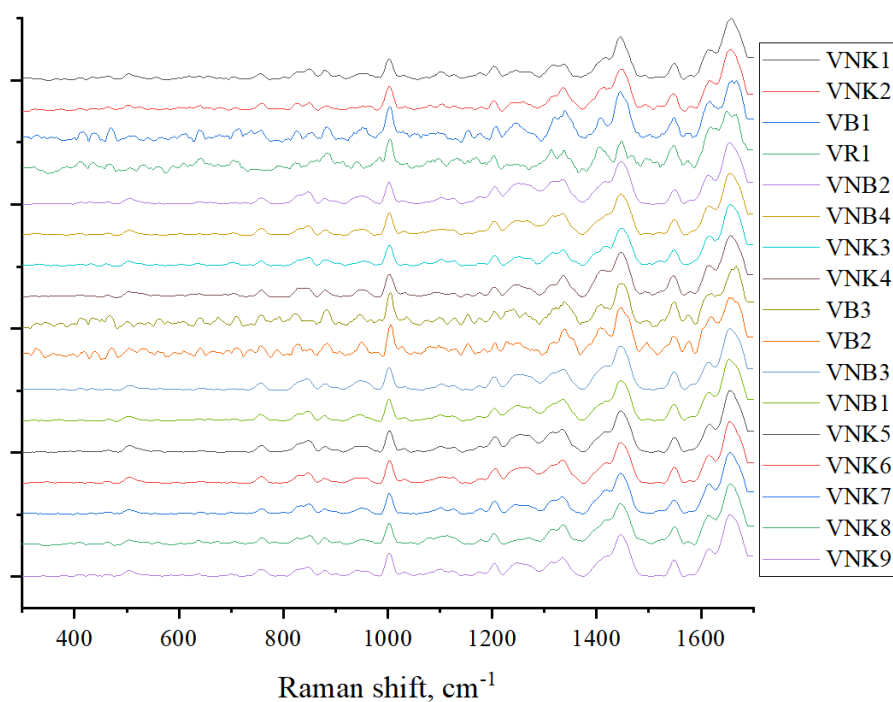


Рисунок 4.1 – Предварительно обработанные рамановские спектры образцов ядов.

Далее спектральные данные всех образцов были подвергнуты МГК. Итоговый график МГК показан на рисунке 4.2.

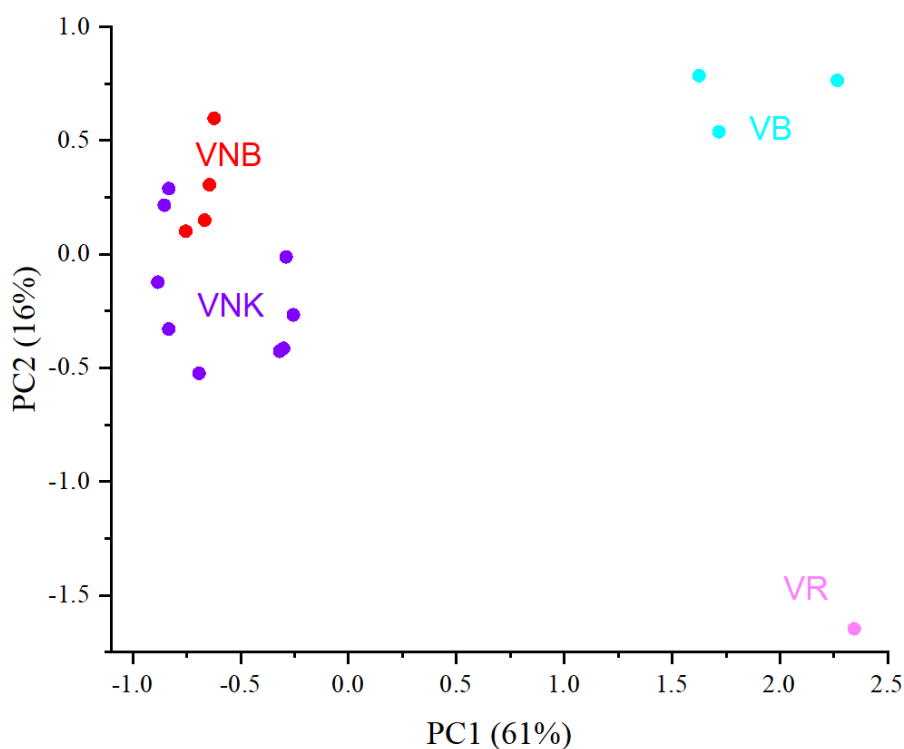


Рисунок 4.2 – График МГК, где в скобках указан процент объясненной дисперсии для каждой ГК. Четыре группы образцов обозначены кружками разного цвета: VNK – фиолетовые, VNB – красные, VR – розовые и VB – голубые.

Видно, что расстояния между образцами ядов змей разных видов значительно больше, чем между ядами змей одного вида (VN). Расстояния между точками на графике МГК можно использовать для оценки степени изменчивости состава ядов. Значительные межвидовые различия ядов заметны уже по исходным спектрам на рисунке 4.1. Межвидовые вариации в составе ядов вполне ожидаемы, но интересно отметить и наличие значительных внутривидовых различий.

Увеличенный фрагмент графика МГК, показывающий только образцы VN, показан на рисунке 4.3. Видно, что можно идентифицировать группу ядов змей одной и той же популяции (VNB, красные кружки). Следует отметить, что наблюдается тенденция к группировке, но не строгое отделение ядов одной популяции от другой, поскольку состав ядов различается и у змей внутри одной популяции.

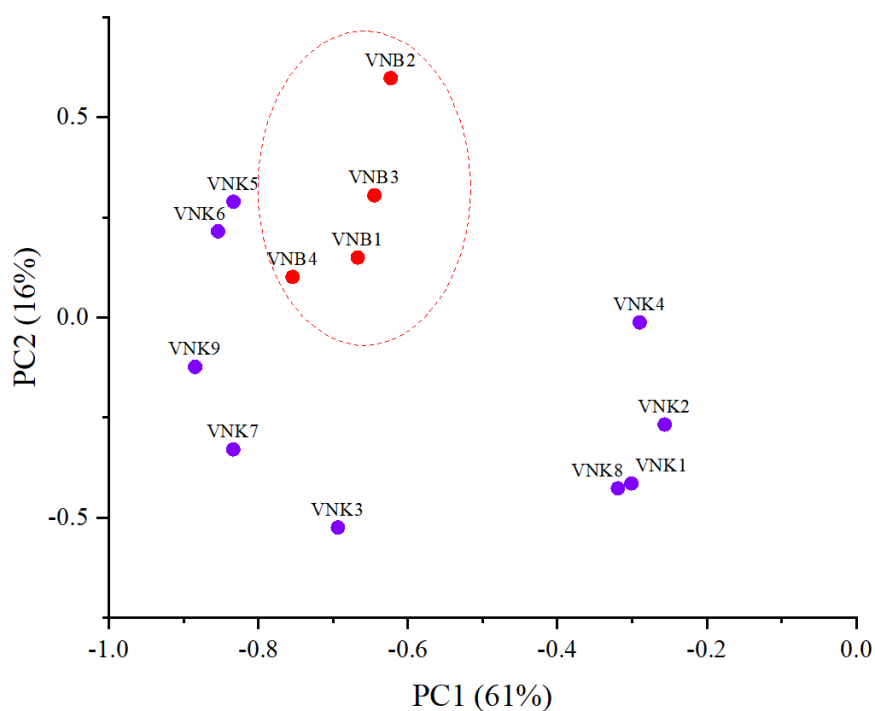


Рисунок 4.3 – Образцы VN на увеличенном фрагменте общего графика МКГ.

Для количественной оценки и сравнения изменчивости ядов внутри и между популяциями были рассчитаны средние координаты для ядов VNK (VNKav) и VNB (VNBav) на графике МКГ (рис. 4.4) и расстояние между ними, которое оказалось равным 0,47 (с учетом погрешности, максимальное и минимальное значения этого расстояния равны 0,67 и 0,30 соответственно). Затем были рассчитаны расстояния от образцов VNK до точки VNKav (Таблица 4.2) и их средние значения (и аналогично для ядов другой популяции – VNB). Среднее расстояние от образцов VNK до VNKav составило $0,38 \pm 0,03$, от образцов VNB до VNBav – $0,17 \pm 0,06$. Размерность соответствуют таковой на рисунке 4.3 или 4.4. Поскольку количество образцов варьируется в зависимости от популяции, лишь с осторожностью можно заключить, что разница в этих цифрах (0,38 и 0,17) обусловлена различиями вариаций ядов внутри каждой из двух популяций. Для более правильной оценки необходимо также увеличить количество образцов ядов из каждой популяции.

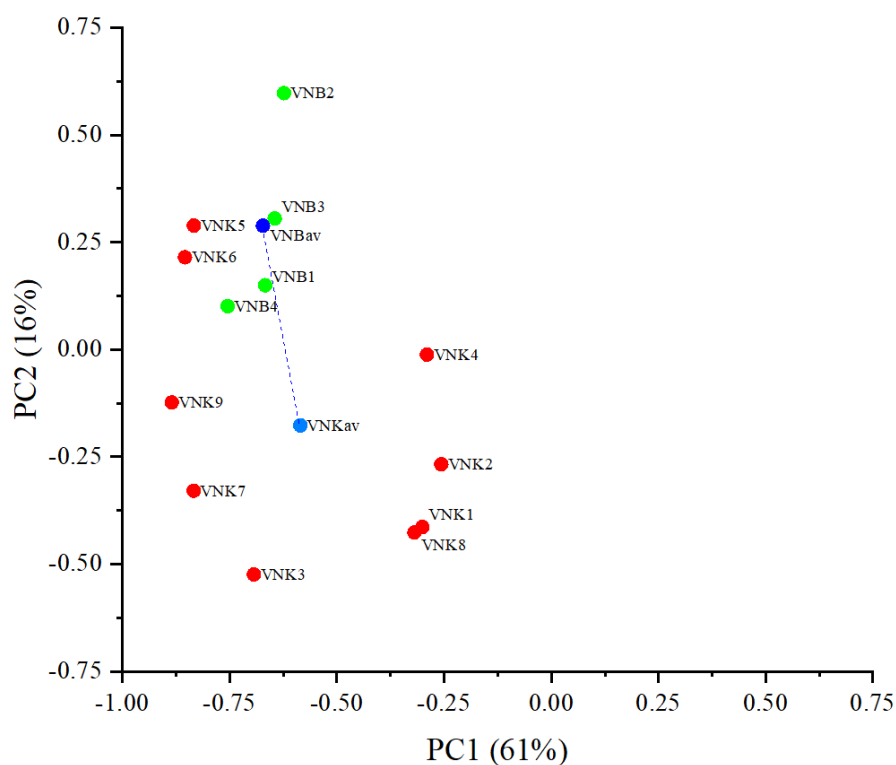


Рисунок 4.4 – Рассчитанные средние координаты VNK (VNKav) и VNB (VNBav) на графике МГК. Расстояние между VNKav и VNBav, равное 0,47, показано синим пунктиром.

Таблица 4.2. Рассчитанные расстояния от образцов VNK до VNKav и от образцов VNB до VNBav и их средние величины.

Образец VNK	Расстояние от отдельного образца VNK до VNKav	Образец VNB	Расстояние от отдельного образца VNB до VNBav
VNK1	0,37	VNB2	0,31
VNK2	0,34	VNB4	0,20
VNK3	0,36	VNB3	0,03
VNK4	0,34	VNB1	0,14
VNK5	0,53	Среднее расстояние	0,17±0.06
VNK6	0,48	–	–
VNK7	0,29	–	–
VNK8	0,36	–	–
VNK9	0,30	–	–
Среднее расстояние	0,38±0.03	–	–

Сравнивая расстояние 0,47 между средними координатами для двух популяций (VNKav и VNBav) со средним расстоянием от образцов VNK до VNKav 0,38 (оно больше, чем таковое

для другой популяции), можно предположить, что межпопуляционные различия превышают внутривидовые. Но следует отметить, что образцы VNK 5 и 6 расположены от VNKав на расстоянии, превышающем 0,47.

4.2.2. Анализ полосы Amide I ядов VN и фосфолипаз A2

Для выяснения того, какие особенности состава могут быть ответственны за полученную группировку ядов змей одной популяции, отдельно была проанализирована полоса Amide I рамановских спектров ядов VN. Анализ именно этой полосы позволяет интерпретировать спектральные данные с точки зрения количественного содержания элементов вторичной структуры белка. На основе анализа полос Amide I образцов VN (Рисунок 4.5) было выявлено, что во всех них преобладает α -структура (максимумы приходятся на диапазон 1653–1658 cm^{-1}). Примечательно, что корреляции между составом яда и полом змеи обнаружено не было.

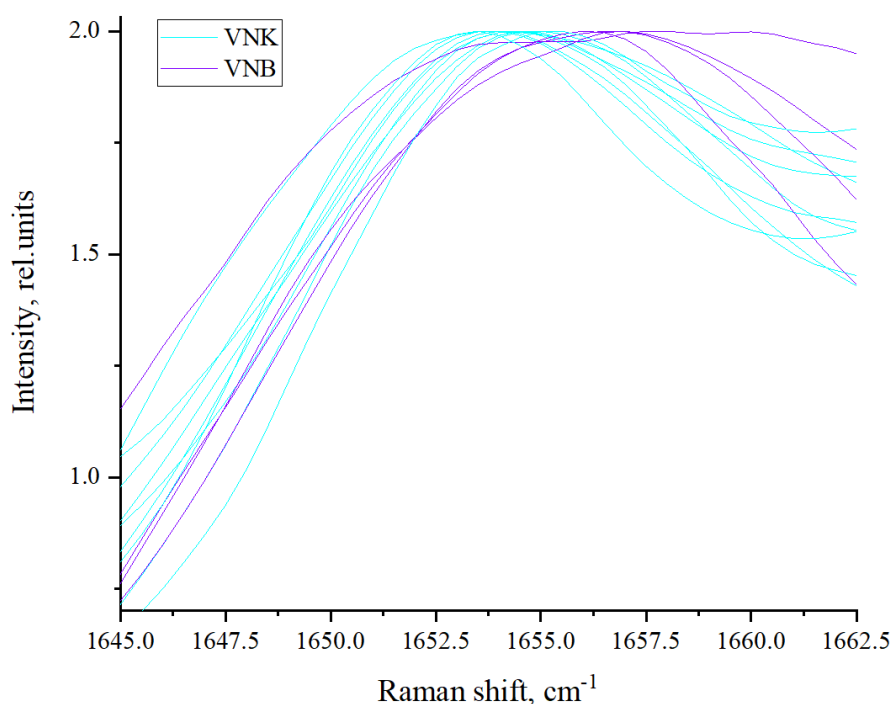


Рисунок 4.5 – Максимумы пиков полосы Amide I для всех образцов VNB и VNK.

Для выявления различия ядов на межпопуляционном уровне был проведен более детальный анализ Amide I. Сначала спектры для каждой популяции усреднялись сложением спектров всех образцов и делением результата на их количество (для двух популяций). Далее выполнялась аппроксимация кривыми полученных усредненных спектров Amide I для оценки

вклада различных элементов вторичной структуры (Рисунок 4.6). Полученные результаты представлены в таблице 4.3.

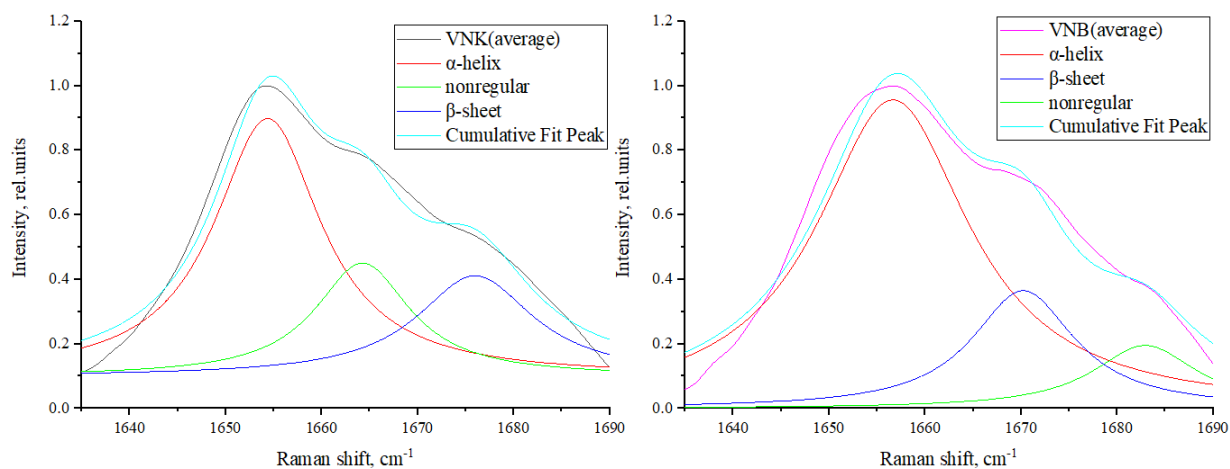


Рисунок 4.6 – Фитинг полосы Amide I средних (average) спектров VNK (слева) и VNB (справа). α -helix – α -спираль, β -sheet – β -слой, nonregular – нерегулярная структура, Cumulative Fit Peak – суммарный подгоночный пик (полученный суммированием всех подгоночных пиков). R-Squares для VNK(average) и VNB(average) составили 0,98 и 0,97 соответственно.

Известно, что преобладающими компонентами змеиного яда VN являются фосфолипазы A2 (PLA2) – HDP1, HDP2 [95, 129]. Поэтому было решено также записать спектры КР HDP1 и HDP2 и аналогичным образом проанализировать их полосы Amide I (Рисунок 4.7). Рассчитанные содержания α -, β - и нерегулярных структур в молекулах HDP1 и HDP2 указаны в таблице 4.3. Приведенные данные показывают, что α -спираль является основным структурным элементом в изученных PLA2.

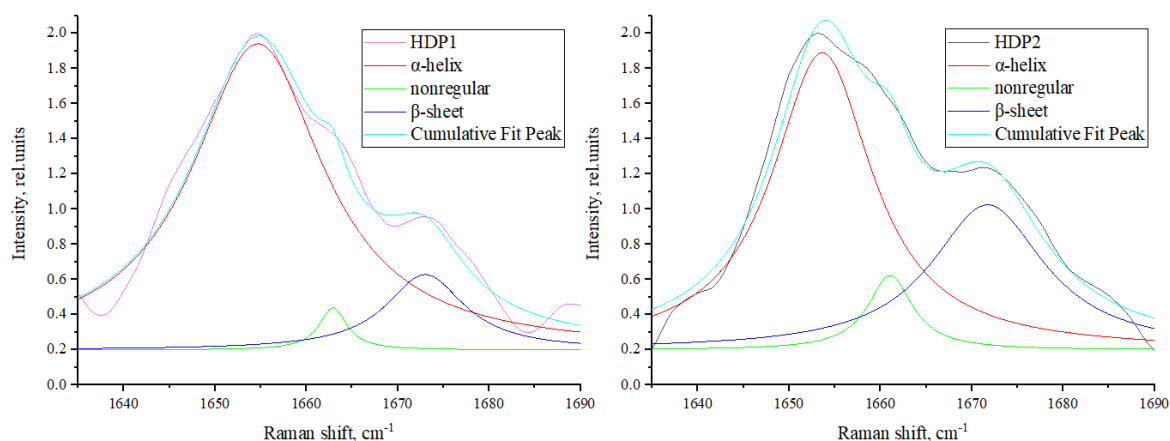


Рисунок 4.7 – Фитинг полосы Amide I спектров HDP1 (слева) и HDP2 (справа). α -helix – α -спираль, β -sheet – β -слой, nonregular – нерегулярная структура, Cumulative Fit Peak – суммарный подгоночный пик. R-Squares для HDP1 и HDP2 составили 0,98 и 0,98.

Интересно отметить, что соотношение вторичных структур α/β в HDP1 и HDP2 различается, что указывает на их структурные различия: в HDP2 заметен больший вклад β -слоев. Учитывая обнаруженные различия между ядами VNK и VNB в содержании элементов вторичной структуры (Таблица 4.3), можно предположить, что большее количество α -структур в ядах VNB объясняется бóльшим содержанием PLA2 (или HDP1) в этих ядах, по сравнению с VNK.

Смещение вправо максимумов Amide I образцов VNB относительно максимумов VNK (Рисунок 4.5) можно объяснить бóльшим количеством HDP1 (по сравнению с HDP2) в VNB: максимум Amide I спектра HDP1 также немного сдвинут вправо (в сторону бóльших частот) относительно максимума HDP2 (Таблица 4.3). Можно также предположить, что вклад токсинов других семейств (например, SVSP) в спектры VNK (и соответственно, в их составы) более заметен.

Таблица 4.3. Содержание элементов вторичной структуры по данным аппроксимации кривыми.

Образец	Тип структуры	Частота, см ⁻¹	Площадь, %
VNK	α -Спираль	1654	55
	Нерегулярная	1664	22
	β -Слой	1676	23
VNB	α -Спираль	1656,5	72
	β -Слой	1670	18
	Нерегулярная	1683	10
HDP1	α -Спираль	1655	85
	Нерегулярная	1663	2
	β -Слой	1673	13
HDP2	α -Спираль	1653,5	60
	Нерегулярная	1661,5	6
	β -Слой	1672	34

4.3. Выводы к главе 4

В результате, были проведены Раман-спектроскопические исследования ряда ядов отдельных особей змей, преимущественно одного вида – *Vipera nikolskii*. Анализ образцов отдельных ядов *V. nikolskii* и оценка их внутривидовой изменчивости проведены нами впервые.

С помощью математических методов анализа спектров КР ядов были продемонстрированы внутривидовые (меж- и внутрипопуляционные) различия в составе ядов, которые также были сопоставлены с межвидовыми. Кроме того, продемонстрирована тенденция разделения ядов одного вида гадюк на 2 группы, соответствующие двум популяциям этих змей. Более детальный анализ спектров позволил объяснить такую группировку составом ядов: яд змей одной популяции содержал большее количество α -структурных элементов, чем яд особей другой популяции, что, вероятно, объясняется бóльшим содержанием PLA2 (или HDP1).

Используемый в работе метод может быть применен для анализа состава и вариаций яда отдельных особей змей одного вида, поскольку позволяет работать с малыми количествами яда и фиксировать достаточно тонкие различия в составе. Предложенный метод в ряде случаев может рассматриваться как альтернатива широко используемой масс-спектрометрии.

Глава 5. Сравнение спектров КР рекомбинантного и синтетического белка SLURP-1

Мотивация данного исследования состояла в сравнении рекомбинантной, произведенной *E. coli*, версии подобного трехпетельным токсинам белка SLURP-1 – rSLURP-1 с синтетической версией sSLURP-1, полученной методом пептидного синтеза. rSLURP-1 имеет один дополнительный N-концевой остаток Met [130, 131], в то время как sSLURP-1 идентичен по своей аминокислотной последовательности природному белку человека. Несмотря на это минимальное различие двух версий белка, их активность ранее была найдена различной. Поэтому была предпринята попытка объяснения данного различия разницей структур rSLURP-1 и sSLURP-1. Обнаружение подобных тонких структурных различий молекул может быть выполнено с помощью инструмента РС, поэтому нами было проведено РС-исследование этих двух белков. Также была предпринята попытка выявления преимущества использования той или иной версии SLURP-1.

5.1. Материалы и методы

Капли растворов sSLURP-1 и rSLURP-1 объемом 2,5 мкл сушили на подложке из фторида кальция. Белок преимущественно концентрировался в областях на границах капель (демонстрация эффекта «кофейных колец»), что позволяло регистрировать достаточно интенсивные спектры, фокусируясь на границах (Рисунок 5.1).

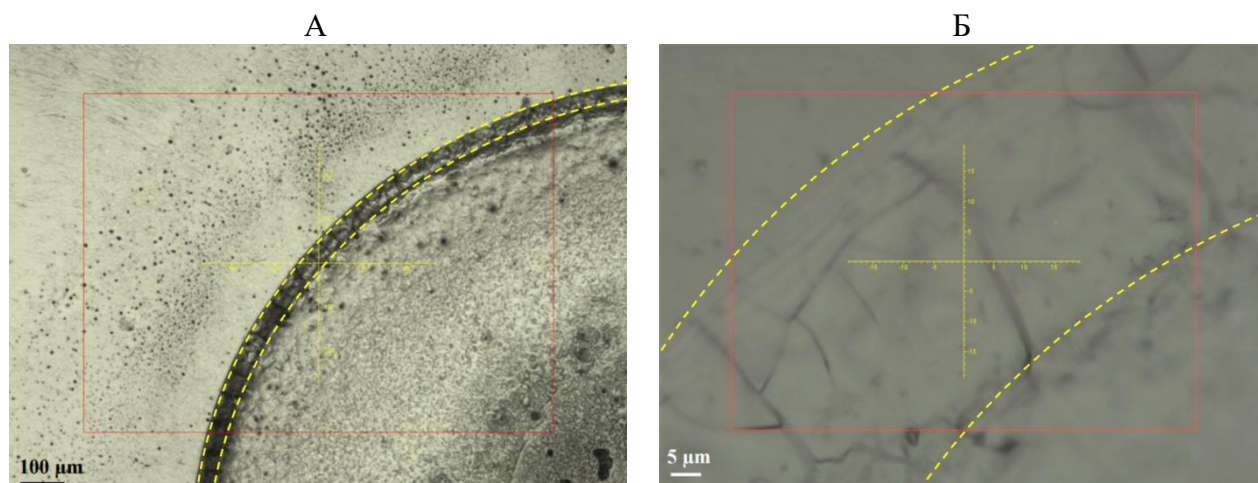


Рисунок 5.1 – Вид высушенной капли образца rSLURP-1 объемом 2,5 мкл через объектив 4^{\times} (А) и 50^{\times} (Б). Перекрестие указывает примерное место фокусировки лазерного излучения при регистрации спектров. Показана концентрация белка в виде кольца, его контуры ориентировочно обозначены желтым пунктиром.

Измерения спектров КР были проведены аналогично описанному в пункте 3.1.2 главы 3.

Для анализа спектр каждого образца записывали 10 раз и далее сравнивали усредненные спектры (всего было записано 20 спектров). Все спектры были обработаны путем применения коррекции базовой линии («concave rubberband correction»), нормализации (интенсивность пика 1004 cm^{-1} приравнивалась 2) и сглаживания (количество точек сглаживания = 17) в OPUS 8.2.28.

5.2. Результаты

Были подробно рассмотрены наиболее информативные участки полученных спектров КР sSLURP-1 и rSLURP-1 (Рисунок 5.2), отражающие особенности структуры белков: 1) S–S область (Рисунок 5.3 (А)), 2) область дублета тирозина (Рисунок 5.3 (Б)), а также полосы 3) Amide III ($1230\text{--}1280\text{ cm}^{-1}$) и 4) Amide I (Рисунки 5.4 (А), (Б) соответственно).

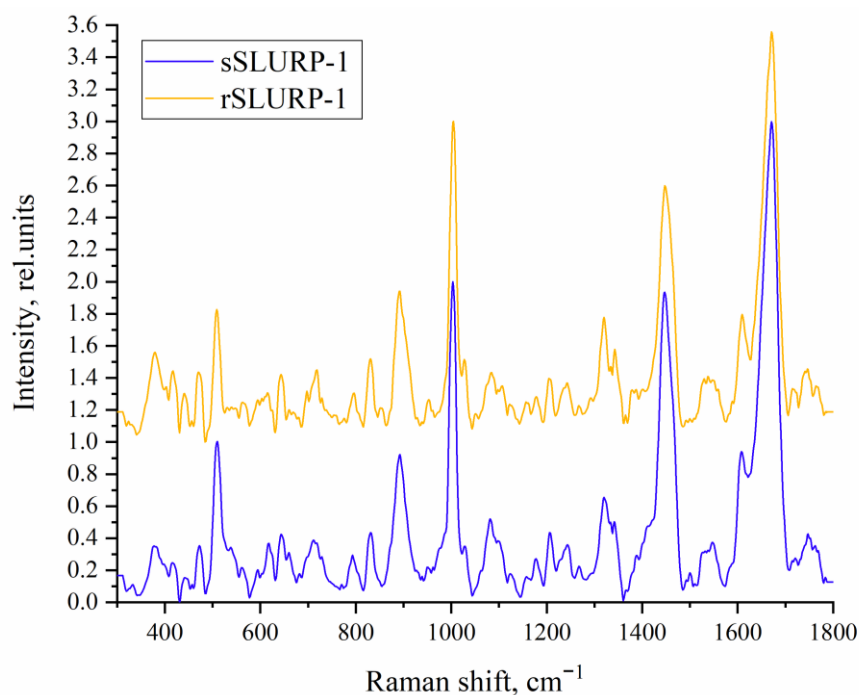


Рисунок 5.2 – Спектры КР sSLURP-1 и rSLURP-1.

Анализ S–S области позволил установить, что в структуре обоих образцов преобладает конформация *gauche-gauche-gauche* C–C–S–S–C–C, поскольку максимум основного пика локализован вблизи 510 cm^{-1} . Однако для sSLURP-1 заметен небольшой вклад *gauche-gauche-trans* конформации, о чем свидетельствует появление плеча около 525 cm^{-1} . Это может быть связано с тем, что определенный процент молекул sSLURP-1 имеет измененную конформацию C–C–S–S–C–C. Поскольку спектры записывались с высушенных образцов, в данном случае можно исключить динамические процессы, в частности переходы между конформациями. Интересно отметить, что в полученном ранее (глава 2) спектре КР α -кобратоксина также наблюдался вклад конформации *gauche-gauche-trans*. Это может указывать на сходство геометрий петель sSLURP-1 и α -кобратоксина.

Анализ тирозинового дублета позволяет утверждать, что тирозиновые остатки в обоих образцах имеют преимущественно гидрофобное окружение (поскольку основной пик локализован в районе 830 cm^{-1}), однако следует отметить, что в отличие от sSLURP-1, некоторая часть остатков тирозина в молекулах rSLURP-1 имеет гидрофильное окружение (дополнительный максимум около 850 cm^{-1}). Вероятно, это указывает на большее «выставление» на поверхность белка некоторых тирозиновых остатков. Наличие этих энергетически невыгодных остатков в rSLURP-1 может свидетельствовать о меньшей стабильности этого варианта SLURP-1. Однако следует учитывать и другие факторы, в частности, иную, более энергетически выгодную геометрию дисульфидов (без существенного вклада конформации *gauche-gauche-trans*).

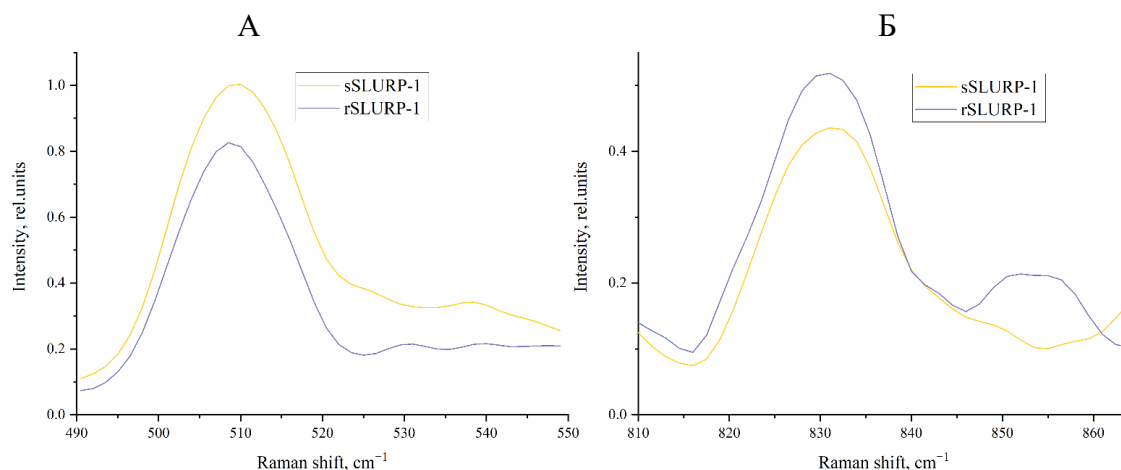


Рисунок 5.3 – Спектры областей S–S (А) и дублета тирозина (Б).

Анализ вторичной структуры образцов не выявил существенных различий: в обоих наблюдается наличие β -структур, о чем свидетельствуют положения основных пиков: около 1240 cm^{-1} для Amide III и 1670 cm^{-1} для Amide I.

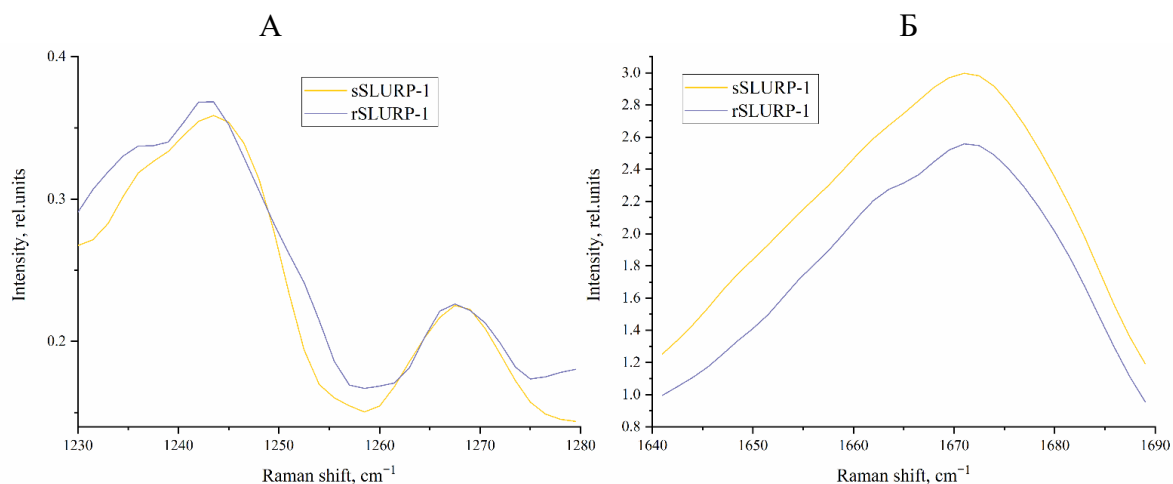


Рисунок 5.4 – Спектры полос Amide III (А) и Amide I (Б).

Интересно, что при анализе SLURP-1 описанным выше (глава 2) методом (наряду с другими рассмотренными в главе 2 токсинами) и добавлении rSLURP-1 и sSLURP-1 на полученную ранее МГК «карту», наблюдается, как и следовало ожидать, их структурная близость к трехпетельным токсинам (Рисунок 5.5), в частности к α -кобратоксину (Ctx).

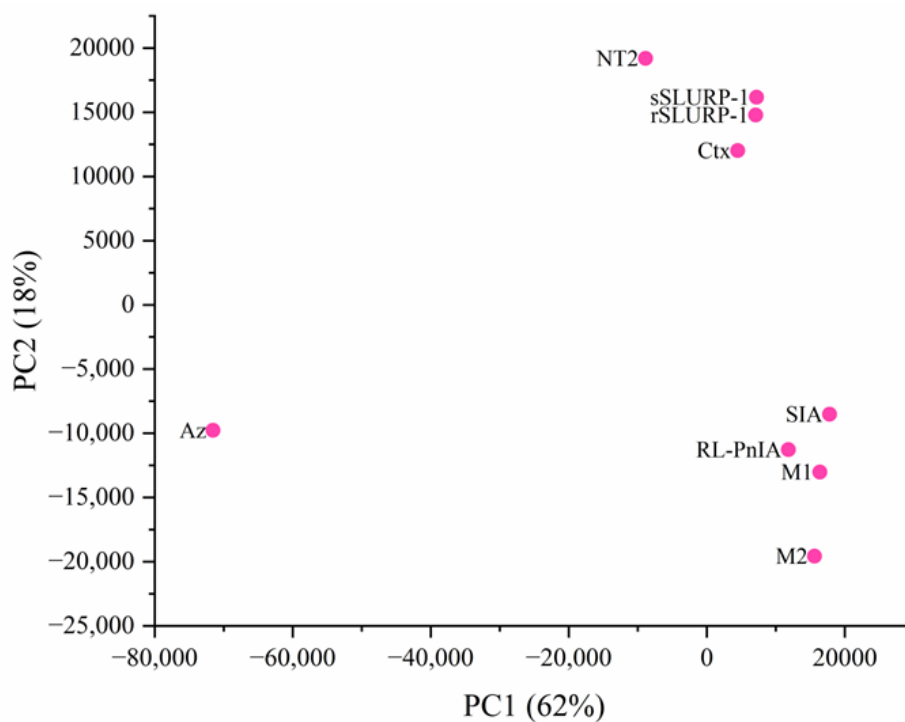


Рисунок 5.5 – График МГК для спектров КР токсинов – лигандов наХР (см. также Рисунок 2.3).

Отметим также, что, как видно из графика весов (Рисунок 5.6), за полученную группировку спектров токсинов ответственно множество разных спектральных участков (маркерных полос), характеризующих колебания как основной, так и боковых цепей белка.

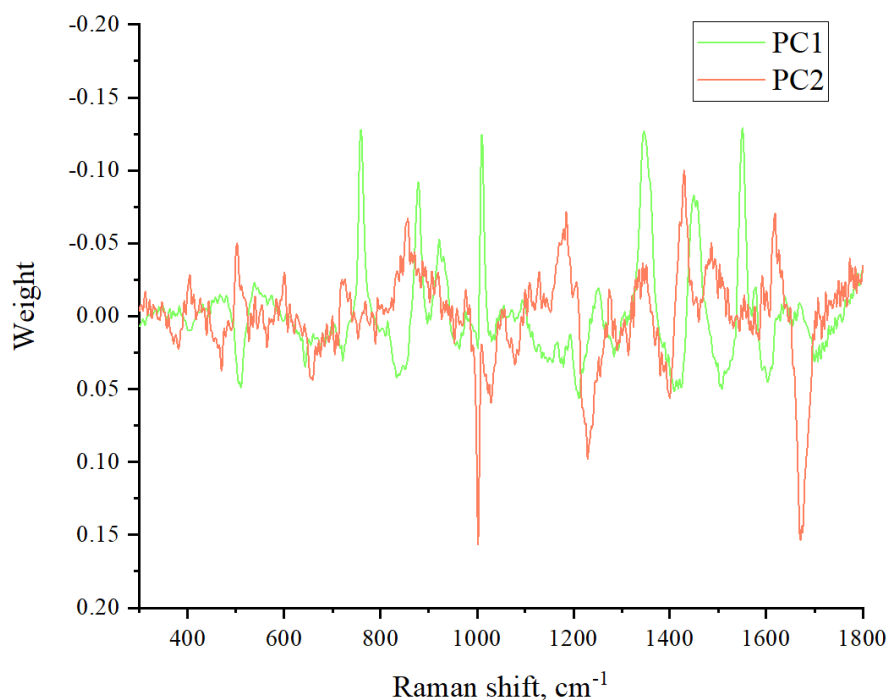


Рисунок 5.6 – График весов (для результатов МГК, приведенных на рисунке 5.5): показаны веса для каждого рамановского сдвига, то есть влияние разных частот на значения первой и второй КГ (PC1 и PC2).

5.3. Выводы к главе 5

В целом, приведенные результаты иллюстрируют сходство двух форм подобного трехпетельным токсинам белка SLURP-1 и показывают, что rSLURP, не смотря на дополнительный N-концевой Met, структурно достаточно схож с sSLURP-1.

Заключение

Итоги выполненного исследования (выводы)

1. Разработана методика структурного анализа белковых и пептидных токсинов животных на основе КР спектроскопии.
2. Разработана методика анализа цельных змеиных ядов на основе КР спектроскопии.
3. Впервые продемонстрировано, что КР спектроскопия в сочетании с МГК и кластеризацией может успешно применяться как для выявления структурных сходств и различий между белковыми (пептидными) токсинами животных, включая их синтетические аналоги, так и для классификации цельных ядов змей, отражающей состав ядов.
4. Показано, что на основе разработанной методики можно различать дисульфидные изомеры одного и того же пептидного токсина, а также фиксировать тонкие структурные различия близких по аминокислотной последовательности токсинов и им подобных белков.
5. Показано, что обработка КР спектров образцов ядов змей различных видов, с помощью метода анализа дискриминантных функций (DFA), совмещенного с МГК, корректно разделяет образцы на две группы, соответствующие семействам змей Viperidae и Elapidae.
6. Продемонстрированы и оценены внутривидовые (меж- и внутривидовые) различия в составе ядов и проведено сравнение этих различий с межвидовыми.

В итоге, в работе представлен чрезвычайно экономичный экспресс-метод, основанный на спектроскопии КР и математических методов обработки данных, который позволяет быстро анализировать большое количество токсинов и ядов. Малый расход материалов дает возможность работать с небольшими количествами образцов, что важно при изучении токсинов и ядов малопродуктивных, а также редких и исчезающих видов животных.

Рекомендации и перспективы дальнейшей разработки темы

Метод может рассматриваться как дополнение или, в ряде случаев, альтернатива существующим методикам структурного и протеомного анализа. Он может найти применение в медицинских приложениях: для разработки лекарств и противоядий.

Совместные работы исследователей из разных научных областей представляются крайне перспективными и многообещающими. Подобная конвергенция позволяет расширять методологические подходы, а также обращаться к новым интересным объектам исследования. В дальнейшем планируется продолжить работы в направлении применения спектроскопии КР в

области токсикологии. В частности, в планах создание библиотеки представителей различных семейств белковых токсинов ядов змей и попытка реализации декомпозиции спектров КР цельных ядов и более точной оценки их состава с опорой на данную библиотеку. Также перспективной кажется попытка обнаружения и анализа с помощью РС змеиного яда в определенных средах, приближенным к биотканям жертвы укуса, для имитации проб, получаемых на практике от пострадавших от укуса змеи.

Благодарности

Автор выражает благодарность своему научному руководителю С.В. Гарнову за поддержку, участие, создание условий для комфортной работы и предоставление доступа к необходимому оборудованию; К.А. Прохорову за обучение работе на приборе; С.В. Гудкову за многочисленные советы на протяжении всей работы; А.В. Богачу за участие и советы по представлению докладов.

Автор также глубоко благодарен коллективу отдела молекулярной нейроиммунной сигнализации ИБХ РАН: Д.С. Кудрявцеву за неоценимую помощь в освоении специфики биологической науки, помощь в работе с оборудованием, возможность участия в интересных проектах, новые идеи, полезные замечания, а также терпение, внимание и моральную поддержку; Ю.Н. Уткину за предоставление многочисленных интересных биоматериалов, помощь в подготовке статей, обсуждение результатов и подачу новых идей; В.Г. Старкову за получение необходимых ядов и «знакомство» со змеями; И.Е. Кашеверову за участие в написании статей и помощь при выходе на защиту диссертации и В.И. Цетлину за помощь в подготовке статей.

Также автор выражает благодарность Т.И. Можяевой за помощь с грамматикой английского языка при написании статей.

Список литературы

1. Miura, T. Raman spectroscopy of proteins and their assemblies. / T. Miura, G. J. Thomas // *Sub-cellular biochemistry*. – 1995. – Vol. 24. – P. 55-99.
2. Takamatsu, T. Raman spectra of some snake venom components / T. Takamatsu, I. Harada, K. Hayashi // *BBA - Protein Structure*. – 1980. – Vol. 622. – № 2. – P. 189-200.
3. Principal components analysis of Raman spectral data for screening of Hepatitis C infection / A. Ditta, H. Nawaz, T. Mahmood [et al.] // *Spectrochimica Acta - Part A: Molecular and Biomolecular Spectroscopy*. – 2019. – Vol. 221 – P. 117173.
4. Tsetlin, V. Snake venom α -neurotoxins and other “three-finger” proteins / V. Tsetlin // *European Journal of Biochemistry*. – 1999. – Vol. 264. – № 2. – P. 281-286.
5. From Animal Poisons and Venoms to Medicines: Achievements, Challenges and Perspectives in Drug Discovery / K. de C. F. Bordon, C. T. Cologna, E. C. Fornari-Baldo [et al.] // *Frontiers in Pharmacology*. – 2020. – Vol. 11 – P. 1132.
6. Antiviral Effects of Animal Toxins: Is There a Way to Drugs? / Y. Utkin, A. Siniavin, I. Kasheverov, V. Tsetlin // *International Journal of Molecular Sciences*. – 2022. – Vol. 23. – № 7 – P. 3634.
7. Snake venom phospholipase A2s exhibit strong virucidal activity against SARS-CoV-2 and inhibit the viral spike glycoprotein interaction with ACE2 / A. E. Siniavin, M. A. Streltsova, M. A. Nikiforova [et al.] // *Cellular and Molecular Life Sciences*. – 2021. – Vol. 78. – № 23. – P. 7777-7794.
8. Animal protein toxins: origins and therapeutic applications / N. Chen, S. Xu, Y. Zhang, F. Wang // *Biophysics Reports*. – 2018. – Vol. 4. – № 5. – P. 233-242.
9. Azemiopsin, a selective peptide antagonist of muscle nicotinic acetylcholine receptor: Preclinical evaluation as a local muscle relaxant / I. V. Shelukhina, M. N. Zhmak, A. V. Lobanov [et al.] // *Toxins*. – 2018. – Vol. 10. – № 1 – P. 34.
10. Editorial: Animal toxins as comprehensive pharmacological tools to identify diverse ion channels / Y. Utkin, A. Vassilevski, D. Kudryavtsev, E. A. B. Undheim // *Frontiers in Pharmacology*. – 2019. – Vol. 10. – P. 423.
11. Kini, R. M. Toxinology provides multidirectional and multidimensional opportunities: A

personal perspective / R. M. Kini // *Toxicon*: X. – 2020. – Vol. 6 – P. 100039.

12. Daltry, J. C. Diet and snake venom evolution / J. C. Daltry, W. Wüster, R. S. Thorpe // *Nature*. – 1996. – Vol. 379. – № 6565. – P. 537-542.
13. Global mortality of snakebite envenoming between 1990 and 2019 / N. L. S. Roberts, E. K. Johnson, S. M. Zeng [et al.] // *Nature Communications*. – 2022. – Vol. 13. – № 1 – P. 6160.
14. Dhankhar, D. Cell-phone camera Raman spectrometer / D. Dhankhar, A. Nagpal, P. M. Rentzepis // *Review of Scientific Instruments*. – 2021. – Vol. 92. – № 5 – P. 054101.
15. Chippaux, J. P. Snake venom variability: methods of study, results and interpretation / J. P. Chippaux, V. Williams, J. White // *Toxicon*. – 1991. – Vol. 29. – № 11. – P. 1279-1303.
16. Phylovenomics of *Daboia russelii* across the Indian subcontinent. Bioactivities and comparative in vivo neutralization and in vitro third-generation antivenomics of antivenoms against venoms from India, Bangladesh and Sri Lanka / D. Pla, L. Sanz, S. Quesada-Bernat [et al.] // *Journal of Proteomics*. – 2019. – Vol. 207 – P. 103443.
17. Merbel, N. van de. Sample preparation for LC-MS bioanalysis of proteins / N. van de Merbel // *Sample Preparation in LC-MS Bioanalysis*. – 2019. – P. 304-318.
18. Raman Spectroscopy to Monitor Post-Translational Modifications and Degradation in Monoclonal Antibody Therapeutics / B. S. McAvan, L. A. Bowsher, T. Powell [et al.] // *Analytical Chemistry*. – 2020. – Vol. 92. – № 15. – P. 10381-10389.
19. Conformational fingerprinting of tau variants and strains by Raman spectroscopy / G. Devitt, A. Crisford, W. Rice [et al.] // *RSC Advances*. – 2021. – Vol. 11. – № 15. – P. 8899-8915.
20. Raman, C. V. The colour of the sea / C. V. Raman // *Nature*. – 1921. – Vol. 108. – № 2716. – P. 367.
21. Fabelinskii, I. L. The discovery of combinational scattering of light (the Raman effect) / I. L. Fabelinskii // *Uspekhi Fizicheskikh Nauk*. – 1978. – Vol. 126. – № 9. – P. 124-152.
22. John, N. Raman Spectroscopy / N. John, S. George // *Spectroscopic Methods for Nanomaterials Characterization*. – 2017. – Vol. 2. – P. 95-127.
23. Smith, E. Modern raman spectroscopy: A practical approach / E. Smith, G. Dent. – 2019. – 1-241 p.

24. Raman spectra of amino acids and their aqueous solutions / G. Zhu, X. Zhu, Q. Fan, X. Wan // *Spectrochimica Acta - Part A: Molecular and Biomolecular Spectroscopy*. – 2011. – Vol. 78. – № 3. – P. 1187-1195.
25. Tuma, R. Raman spectroscopy of proteins: From peptides to large assemblies / R. Tuma // *Journal of Raman Spectroscopy*. – 2005. – Vol. 36. – № 4. – P. 307-319.
26. Filik, J. Drop coating deposition Raman spectroscopy of protein mixtures / J. Filik, N. Stone // *Analyst*. – 2007. – Vol. 132. – № 6. – P. 544-550.
27. Conformational Evolution of Molecular Signatures during Amyloidogenic Protein Aggregation / G. Devitt, W. Rice, A. Crisford [et al.] // *ACS Chemical Neuroscience*. – 2019. – P. 4593-4611.
28. White, W. B. Raman spectra and structure of natural glasses / W. B. White, D. G. Minser // *Journal of Non-Crystalline Solids*. – 1984. – Vol. 67. – № 1-3. – P. 45-59.
29. Schmid, T. Raman microspectroscopic imaging of binder remnants in historical mortars reveals processing conditions / T. Schmid, P. Dariz // *Heritage*. – 2019. – Vol. 2. – № 2. – P. 1662-1683.
30. Matys, J. Energy and power density: A key factor in lasers studies / J. Matys, M. Dominiak, R. Flieger // *Journal of Clinical and Diagnostic Research*. – 2015. – Vol. 9. – № 12. – P. ZL01-ZL02.
31. Tuschel, D. Selecting an excitation wavelength for raman spectroscopy / D. Tuschel // *Spectroscopy (Santa Monica)*. – 2016. – Vol. 31. – № 3 – P. 14-23.
32. Raman spectroscopy as a probe of protein structure in food systems / E. Li-Chan, S. Nakai, M. Hirotsuka // *Protein Structure-Function Relationships in Foods* / eds. R. Y. Yada [et al.]. – Boston, MA : Springer US, 1994. – P. 163-197.
33. Differentiation of snake venom using Raman spectroscopic analysis / V. Mozhaeva, V. Starkov, D. Kudryavtsev [et al.] // *Journal of Materials Chemistry B*. – 2023. – Vol. 11. – № 27. – P. 6435-6442.
34. Tu, A. T. Laser Raman spectroscopy of snake venom neurotoxins: conformation / A. T. Tu, B. H. Jo, N. T. Yu // *International Journal of Peptide and Protein Research*. – 1976. – Vol. 8. – № 4. – P. 337-343.
35. Raman spectroscopy of proteins: A review / A. Rygula, K. Majzner, K. M. Marzec [et al.] // *Journal of Raman Spectroscopy*. – 2013. – Vol. 44. – № 8. – P. 1061-1076.
36. Vibrational couplings and energy transfer pathways of water's bending mode / C. C. Yu, K. Y.

- Chiang, M. Okuno [et al.] // Nature Communications. – 2020. – Vol. 11. – № 1 – 5977.
37. Distinct cysteine sulfhydryl environments detected by analysis of raman S-H markers of Cys → Ser mutant proteins / S. W. Raso, P. L. Clark, C. Haase-Pettingell [et al.] // Journal of Molecular Biology. – 2001. – Vol. 307. – № 3. – P. 899-911.
38. Towards practical and sustainable SERS: a review of recent developments in the construction of multifunctional enhancing substrates / C. Li, Y. Huang, X. Li [et al.] // Journal of Materials Chemistry C. – 2021. – Vol. 9. – № 35. – P. 11517-11552.
39. Principal component analysis based methodology to distinguish protein SERS spectra / G. Das, F. Gentile, M. L. Coluccio [et al.] // Journal of Molecular Structure. – 2011. – Vol. 993. – № 1-3. – P. 500-505.
40. Surface-enhanced Raman scattering for direct protein function investigation: Controlled immobilization and orientation / H. Ma, X. Tang, Y. Liu [et al.] // Analytical Chemistry. – 2021. – Vol. 91. – № 14. – P. 8767-8771.
41. Surface-enhanced Raman scattering sensing platform for detecting amyloid- β peptide interaction with an aggregation inhibitor / M. A. S. de Oliveira, S. Hilt, C.-W. Chang [et al.] // Applied Optics. – 2020. – Vol. 59. – № 25. – P. 7490.
42. Safar, W. New insight into the aptamer conformation and aptamer/protein interaction by surface-enhanced Raman scattering and multivariate statistical analysis / W. Safar, A. S. Tatar, A. Leray [et al.] // Nanoscale. – 2021. – Vol. 13. – № 29. – P. 12443-12453.
43. Strommen, D. P. Resonance raman spectroscopy / D. P. Strommen, K. Nakamoto // Journal of Chemical Education. – 1977. – Vol. 54. – № 8. – P. 474-478.
44. Resonance Raman Probes for Organelle-Specific Labeling in Live Cells / A. N. Kuzmin, A. Pliss, C. K. Lim [et al.] // Scientific Reports. – 2016. – Vol. 6 – P. 28483.
45. Klebe, G. Experimental Methods of Structure Determination / G. Klebe // Drug Design. – 2013. – P. 265-290.
46. Bhasin, M. Computational methods in genome research / M. Bhasin, G. P. S. Raghava // Applied Mycology and Biotechnology. – 2006. – Vol. 6. – P. 179-207.
47. Greenfield, N. J. Using circular dichroism spectra to estimate protein secondary structure / N. J. Greenfield // Nature Protocols. – 2007. – Vol. 1. – № 6. – P. 2876-2890.

48. The Protein Data Bank / H. M. Berman, J. Westbrook, Z. Feng [et al.] // *Nucleic Acids Research*. – 2000. – Vol. 28. – № 1. – P. 235-242.
49. UniProt: the universal protein knowledgebase in 2021 / A. Bateman, M. J. Martin, S. Orchard [et al.] // *Nucleic Acids Research*. – 2021. – Vol. 49. – № D1. – P. 480-489.
50. Deng, H. Protein structure prediction / H. Deng, Y. Jia, Y. Zhang // *International Journal of Modern Physics B*. – 2018. – Vol. 32. – № 18 – P. 1840009.
51. Ion Channels / B. S. Barker, G. T. Young, C. H. Soubrane [et al.]. – Elsevier Inc., 2017. – 11-43 p.
52. Utkin, Y. Animal venoms and their components: Molecular mechanisms of action / Y. Utkin // *Toxins*. – 2021. – Vol. 13. – № 6 – P. 415.
53. Tsetlin, V. I. Three-finger snake neurotoxins and Ly6 proteins targeting nicotinic acetylcholine receptors: Pharmacological tools and endogenous modulators / V. I. Tsetlin // *Trends in Pharmacological Sciences*. – 2015. – Vol. 36. – № 2. – P. 109-123.
54. Conotoxins: Chemistry and Biology / A. H. Jin, M. Muttenthaler, S. Dutertre [et al.] // *Chemical Reviews*. – 2019. – Vol. 119. – № 21. – P. 11510-11549.
55. Azemiopsin from *Azemiops feae* viper venom, a novel polypeptide ligand of nicotinic acetylcholine receptor / Y. N. Utkin, C. Weise, I. E. Kasheverov [et al.] // *Journal of Biological Chemistry*. – 2012. – Vol. 287. – № 32. – P. 27079-27086.
56. Interaction of Synthetic Human SLURP-1 with the Nicotinic Acetylcholine Receptors / T. Durek, I. V. Shelukhina, H. S. Tae [et al.] // *Scientific Reports*. – 2017. – Vol. 7. – № 1 – P. 16606.
57. Wilson, D. Venomics: A mini-review / D. Wilson, N. L. Daly // *High-Throughput*. – 2018. – Vol. 7. – № 3 – P. 19.
58. Cloning, synthesis, and characterization of α -conotoxin GeXIVA, a potent $\alpha 9\alpha 10$ nicotinic acetylcholine receptor antagonist / S. Luo, D. Zhangsun, P. J. Harvey [et al.] // *Proceedings of the National Academy of Sciences of the United States of America*. – 2015. – Vol. 112. – № 30. – P. 4026-4035.
59. Conformational Prediction for Snake Venom Toxins and Laser Raman Scattering of a Cardiotoxin from Taiwan Cobra (*Naja naja atra*) Venom / T. H. Hseu, Y. C. Liu, C. Wang [et al.] // *Biochemistry*. – 1977. – Vol. 16. – № 13. – P. 2999-3006.

60. Fox, J. Conformational analysis of a snake neurotoxin by prediction from sequence, circular dichroism, and Raman spectroscopy / J. Fox, A. T. Tu // *Archives of Biochemistry and Biophysics*. – 1979. – Vol. 193. – № 2. – P. 407-414.
61. Fox, J. W. Amino acid sequence of a snake neurotoxin from the venom of *Lapemis hardwickii* and the detection of a sulfhydryl group by laser Raman spectroscopy / J. W. Fox, M. Elzinga, A. T. Tu // *FEBS Letters*. – 1977. – Vol. 80. – № 1. – P. 217-220.
62. Raman spectrum of toxin B in relation to structure and toxicity / T. Takamatsu, I. Harada, T. Shimanouchi [et al.] // *FEBS Letters*. – 1976. – Vol. 72. – № 2. – P. 291-294.
63. Tu, A. T. Structural properties of Mojave toxin of *Crotalus scutulatus* (Mojave rattlesnake) determined by laser Raman spectroscopy / A. T. Tu, B. Prescott, C. H. Chou, G. J. Thomas // *Biochemical and Biophysical Research Communications*. – 1976. – Vol. 68. – № 4. – P. 1139-1145.
64. Yu, N. T. Laser Raman scattering of cobramine B, a basic protein from cobra venom / N. T. Yu, B. H. Jo, D. C. O'Shea // *Archives of Biochemistry and Biophysics*. – 1973. – Vol. 156. – № 1. – P. 71-76.
65. Laser Raman studies on cobrotoxin / T. H. Hseu, H. Chang, D. M. Hwang, C. C. Yang // *BBA - Protein Structure*. – 1978. – Vol. 537. – № 2. – P. 284-292.
66. Yu, N. T. Laser Raman scattering of neurotoxins isolated from the venoms of sea snakes *Lapemis hardwickii* and *Enhydrina schistosa* / N. T. Yu, T. S. Lin, A. T. Tu // *Journal of Biological Chemistry*. – 1975. – Vol. 250. – № 5. – P. 1782-1785.
67. Structure and Chemical Modifications of Neurotoxin from *Naja nigricollis* Studied by Raman Spectroscopy / M. Négrerie, D. Aslanian, P. Gröf [et al.] // *Biochemistry*. – 1990. – Vol. 29. – № 36. – P. 8258-8265.
68. Allen, M. The effect of tryptophan modification on the structure and function of a sea snake neurotoxin / M. Allen, A. T. Tu // *Molecular Pharmacology*. – 1985. – Vol. 27. – № 1. – P. 79-85.
69. Raman and infrared spectra of toxin γ from the venom of the scorpion *Tityus serrulatus* / E. P. G. Arêas, J. R. Giglio, E. C. Arantes, Y. Kawano // *Biochimica et Biophysica Acta (BBA)/Protein Structure and Molecular*. – 1987. – Vol. 915. – № 2. – P. 292-298.
70. Structural properties of toxin II of sea anemone (*Anemone sulcata*) determined by laser Raman spectroscopy / B. Prescott, G. J. Thomas, L. Béress [et al.] // *FEBS Letters*. – 1976. – Vol. 64. – № 1. – P. 144-147.

71. Circular dichroism and laser Raman spectroscopic analysis of the secondary structure of Cerebratulus lacteus toxin B-IV / W. R. Kem, C. K. Tu, R. W. Williams [et al.] // *Journal of Protein Chemistry*. – 1990. – Vol. 9. – № 4. – P. 433-443.
72. Primor, N. Conformation of pardaxin, the toxin of the flatfish *Pardachirus marmoratus* / N. Primor, A. T. Tu // *BBA - Protein Structure*. – 1980. – Vol. 626. – № 2. – P. 299-306.
73. Williams, R. W. Raman Spectroscopy of Homologous Plant Toxins: Crambin and $\alpha 1$ and β -Purothionin Secondary Structures, Disulfide Conformation, and Tyrosine Environment / R. W. Williams, M. M. Teeter // *Biochemistry*. – 1984. – Vol. 23. – № 26. – P. 6796-6802.
74. Secondary structure of the entomocidal toxin from *Bacillus thuringiensis* subsp. *kurstaki* HD-73 / C. T. Choma, W. K. Surewicz, P. R. Carey [et al.] // *Journal of Protein Chemistry*. – 1990. – Vol. 9. – № 1. – P. 87-94.
75. Raman Spectroscopy of Synthetic Antimicrobial Frog Peptides Magainin 2a and PGLa / R. W. Williams, R. Starman, K. M. P. Taylor [et al.] // *Biochemistry*. – 1990. – Vol. 29. – № 18. – P. 4490-4496.
76. John, E. A synthetic analogue of melittin aggregates in large oligomers / E. John, F. Jähnig // *Biophysical Journal*. – 1992. – Vol. 63. – № 6. – P. 1536-1543.
77. Honeybee venom and melittin suppress growth factor receptor activation in HER2-enriched and triple-negative breast cancer / C. Duffy, A. Sorolla, E. Wang [et al.] // *npj Precision Oncology*. – 2020. – Vol. 4. – № 1 – P. 24.
78. Vogel, H. Comparison of the Conformation and Orientation of Alamethicin and Melittin in Lipid Membranes / H. Vogel // *Biochemistry*. – 1987. – Vol. 26. – № 14. – P. 4562-4572.
79. Structure-function relationships for cardiotoxins interacting with phospholipids / J. Dufourcq, J. F. Faucon, E. Bernard [et al.] // *Toxicon*. – 1982. – Vol. 20. – № 1. – P. 165-174.
80. Interaction of modified neurotoxins from *Naja nigricollis* with the nicotinic acetylcholine receptor from *Torpedo marmorata* A Raman spectroscopy study / M. Négrerie, D. Aslanian, F. Bouet [et al.] // *FEBS Letters*. – 1991. – Vol. 292. – № 1-2. – P. 249-253.
81. Study of the Effect of Melittin on the Thermotropism of Dipalmitoylphosphatidylcholine by Raman Spectroscopy / M. Lafleur, M. Pigeon, M. Pezolet [et al.] // *Biochemistry*. – 1987. – Vol. 26. – № 4. – P. 1173-1179.

82. Liddle, W. K. Interaction of Myotoxin a with Artificial Membranes: Raman Spectroscopic Investigation / W. K. Liddle, A. T. Tu // *Biochemistry*. – 1985. – Vol. 24. – № 26. – P. 7635-7640.
83. Raman spectroscopic study of acetylcholine receptor-rich membranes from *Torpedo marmorata*. Interaction of the receptor with carbamylcholine and (+)-tubocurarine / D. Aslanian, P. Grof, J. L. Galzi, J. P. Changeux // *BBA - Biomembranes*. – 1993. – Vol. 1148. – № 2. – P. 291-302.
84. Snake venom peptides: Tools of biodiscovery / A. Munawar, S. A. Ali, A. Akrem, C. Betzel // *Toxins*. – 2018. – Vol. 10. – № 11 – P. 474.
85. El-Aziz, T. M. A. Snake venoms in drug discovery: Valuable therapeutic tools for life saving / T. M. A. El-Aziz, A. G. Soares, J. D. Stockand // *Toxins*. – 2019. – Vol. 11. – № 10 – P. 564.
86. The chemistry of snake venom and its medicinal potential / A. L. Oliveira, M. F. Viegas, S. L. da Silva [et al.] // *Nature Reviews Chemistry*. – 2022. – Vol. 6. – № 7. – P. 451-469.
87. Tednes, M. Evaluation and Treatment of Snake Envenomations / M. Tednes, T. L. Slesinger. – 2021.
88. Reid, H. A. Changes in coagulation effects by venoms of *Crotalus atrox* as snakes age / H. A. Reid, R. D. G. Theakston // *American Journal of Tropical Medicine and Hygiene*. – 1978. – Vol. 27. – № 5. – P. 1053-1057.
89. Snakebite envenoming / J. M. Gutiérrez, J. J. Calvete, A. G. Habib [et al.] // *Nature Reviews Disease Primers*. – 2017. – Vol. 3. – № 1 – P. 17063.
90. Intra-specific venom variation in the australian coastal taipan *Oxyuranus scutellatus* / T. Tasoulis, A. Silva, P. C. Veerati [et al.] // *Toxins*. – 2020. – Vol. 12. – № 8 – P. 485.
91. Phenotypic integration in the feeding system of the eastern diamondback rattlesnake (*Crotalus adamanteus*) / M. J. Margres, K. P. Wray, M. Seavy [et al.] // *Molecular Ecology*. – 2015. – Vol. 24. – № 13. – P. 3405-3420.
92. Neonate-to-adult transition of snake venomomics in the short-tailed pit viper, *Gloydius brevicaudus* / J. F. Gao, Y. F. Qu, X. Q. Zhang [et al.] // *Journal of Proteomics*. – 2013. – Vol. 84. – P. 148-157.
93. Highly Evolvable: Investigating Interspecific and Intraspecific Venom Variation in Taipans (*Oxyuranus* spp.) and Brown Snakes (*Pseudonaja* spp.) / J. van Thiel, L. L. Alonso, J. Slagboom [et al.] // *Toxins*. – 2023. – Vol. 15. – № 1 – P. 74.

94. Causes and Consequences of Snake Venom Variation / N. R. Casewell, T. N. W. Jackson, A. H. Laustsen, K. Sunagar // *Trends in Pharmacological Sciences*. – 2020. – Vol. 41. – № 8. – P. 570-581.
95. Quantitative proteomic analysis of venoms from Russian vipers of Pelias group: Phospholipases A2 are the main venom components / S. I. Kovalchuk, R. H. Ziganshin, V. G. Starkov [et al.] // *Toxins*. – 2016. – Vol. 8. – № 4 – P. 105.
96. Vibrational spectroscopy of muscular tissue intoxicated by snake venom and exposed to photobiomodulation therapy / W. F. Vieira, B. Kenzo-Kagawa, M. H. M. Britto [et al.] // *Lasers in Medical Science*. – 2018. – Vol. 33. – № 3. – P. 503-512.
97. Label-free SERS characterization of snake venoms by exploring the cysteine environs with bone-shaped gold nanoparticles / C. C. Huang, Y. S. Lai, C. H. Wang [et al.] // *Journal of Materials Chemistry B*. – 2020. – Vol. 8. – № 47. – P. 10744-10753.
98. Igci, N. A Fourier Transform Infrared Spectroscopic Investigation of *Macrovipera lebetina lebetina* and *M. l. obtusa* Crude Venoms / N. Igci, F. D. Ozel Demiralp // *European Journal of Biology*. – 2020. – Vol. 79. – № 1. – P. 14-22.
99. Fox, J. W. Exploring snake venom proteomes: Multifaceted analyses for complex toxin mixtures / J. W. Fox, S. M. T. Serrano // *Proteomics*. – 2008. – Vol. 8. – № 4. – P. 909-920.
100. A multifaceted analysis of viperid snake venoms by two-dimensional gel electrophoresis: An approach to understanding venom proteomics / S. M. T. Serrano, J. D. Shannon, D. Wang [et al.] // *Proteomics*. – 2005. – Vol. 5. – № 2. – P. 501-510.
101. Quantitative mass spectrometry in proteomics: A critical review / M. Bantscheff, M. Schirle, G. Sweetman [et al.] // *Analytical and Bioanalytical Chemistry*. – 2007. – Vol. 389. – № 4. – P. 1017-1031.
102. Tharwat, A. Principal component analysis - a tutorial / A. Tharwat // *International Journal of Applied Pattern Recognition*. – 2016. – Vol. 3. – № 3. – P. 197-240.
103. Lasch, P. Spectral pre-processing for biomedical vibrational spectroscopy and microspectroscopic imaging / P. Lasch // *Chemometrics and Intelligent Laboratory Systems*. – 2012. – Vol. 117. – P. 100-114.
104. Liland, K. H. Model-based pre-processing in Raman spectroscopy of biological samples / K. H. Liland, A. Kohler, N. K. Afseth // *Journal of Raman Spectroscopy*. – 2016. – Vol. 47. – № 6. – P. 643-

650.

105. A comparison of multivariate analysis techniques and variable selection strategies in a laser-induced breakdown spectroscopy bacterial classification / R. A. Putnam, Q. I. Mohaidat, A. Daabous, S. J. Rehse // *Spectrochimica Acta - Part B Atomic Spectroscopy*. – 2013. – Vol. 87. – P. 161-167.

106. Raman Spectroscopy Combined with Principal Component Analysis for Screening Nasopharyngeal Cancer in Human Blood Sera / S. Khan, R. Ullah, S. Javaid [et al.] // *Applied Spectroscopy*. – 2017. – Vol. 71. – № 11. – P. 2497-2503.

107. Morais, C. L. M. A three-dimensional principal component analysis approach for exploratory analysis of hyperspectral data: Identification of ovarian cancer samples based on Raman microspectroscopy imaging of blood plasma / C. L. M. Morais, P. L. Martin-Hirsch, F. L. Martin // *Analyst*. – 2019. – Vol. 144. – № 7. – P. 2312-2319.

108. Discrimination of basal cell carcinoma and melanoma from normal skin biopsies in vitro through raman spectroscopy and principal component analysis / B. Bodanese, F. L. Silveira, R. A. Zângaro [et al.] // *Photomedicine and Laser Surgery*. – 2012. – Vol. 30. – № 7. – P. 381-387.

109. The potential for histological screening using a combination of rapid Raman mapping and principal component analysis / J. Hutchings, C. Kendall, B. Smith [et al.] // *Journal of Biophotonics*. – 2009. – Vol. 2. – № 1-2. – P. 91-103.

110. Ong, Y. H. Comparison of principal component analysis and biochemical component analysis in Raman spectroscopy for the discrimination of apoptosis and necrosis in K562 leukemia cells / Y. H. Ong, M. Lim, Q. Liu // *Optics Express*. – 2012. – Vol. 20. – № 20. – P. 22158.

111. Detection of the presence of antibodies against *Toxoplasma gondii* in human colostrum by Raman spectroscopy and principal component analysis / C. Araujo-Andrade, J. L. Pichardo-Molina, G. Barbosa-Sabanero [et al.] // *Journal of Biomedical Optics*. – 2007. – Vol. 12. – № 3. – P. 034006.

112. Noninvasive estimation of chronological and photoinduced skin damage using Raman spectroscopy and principal component analysis / F. J. González, C. Castillo-Martínez, M. Martínez-Escanamé [et al.] // *Skin Research and Technology*. – 2012. – Vol. 18. – № 4. – P. 442-446.

113. Alizadeh-Pasdar, N. Principal component similarity analysis of Raman spectra to study the effects of pH, heating, and κ -carrageenan on whey protein structure / N. Alizadeh-Pasdar, S. Nakai, E. C. Y. Li-Chan // *Journal of Agricultural and Food Chemistry*. – 2002. – Vol. 50. – № 21. – P. 6042-6052.

114. Monitoring of chemotherapy leukemia treatment using Raman spectroscopy and principal component analysis / J. L. González-Solís, J. C. Martínez-Espinosa, J. M. Salgado-Román, P. Palomares-Anda // *Lasers in Medical Science*. – 2014. – Vol. 29. – № 3. – P. 1241-1249.
115. Collagen formation observed from healing calvarial defects with principal component analysis of Raman scattering / R. Ahmed, W. Wang, A. W. Zia, C. Lau // *Analyst*. – 2018. – Vol. 143. – № 19. – P. 4614-4622.
116. Deciphering the orientation of lipid molecules by principal component analysis of Raman mapping data / K. A. Okotrub, V. A. Zykova, S. V. Adichtchev, N. V. Surovtsev // *Analyst*. – 2020. – Vol. 145. – № 4. – P. 1466-1472.
117. Weak neurotoxin from *Naja kaouthia* cobra venom affects haemodynamic regulation by acting on acetylcholine receptors / A. Y. Ogay, D. I. Rzhnevsky, A. N. Murashev [et al.] // *Toxicon*. – 2005. – Vol. 45. – № 1. – P. 93-99.
118. α -Conotoxin analogs with additional positive charge show increased selectivity towards *Torpedo californica* and some neuronal subtypes of nicotinic acetylcholine receptors / I. E. Kasheverov, M. N. Zhmak, C. A. Vulfius [et al.] // *FEBS Journal*. – 2006. – Vol. 273. – № 19. – P. 4470-4481.
119. PYCHEM: A multivariate analysis package for python / R. M. Jarvis, D. Broadhurst, H. Johnson [et al.] // *Bioinformatics*. – 2006. – Vol. 22. – № 20. – P. 2565-2566.
120. Hartigan, J. A. Statistical Clustering / J. A. Hartigan // *International Encyclopedia of the Social & Behavioral Sciences*. – 2001. – P. 15014-15019.
121. Jalview Version 2-A multiple sequence alignment editor and analysis workbench / A. M. Waterhouse, J. B. Procter, D. M. A. Martin [et al.] // *Bioinformatics*. – 2009. – Vol. 25. – № 9. – P. 1189-1191.
122. α -Cobratoxin : Proton NMR Assignments and Solution Structure / R. Le Goas, S. R. LaPlante, A. Mikou [et al.] // *Biochemistry*. – 1992. – Vol. 31. – № 20. – P. 4867-4875.
123. Structure-function relationship of three neurotoxins from the venom of *Naja kaouthia*: A comparison between the NMR-derived structure of NT2 with its homologues, NT1 and NT3 / Y. Cheng, Q. Meng, W. Wang, J. Wang // *Biochimica et Biophysica Acta - Protein Structure and Molecular Enzymology*. – 2002. – Vol. 1594. – № 2 – P. 353-363.
124. Natural compounds interacting with nicotinic acetylcholine receptors: From low-molecular

- weight ones to peptides and proteins / D. Kudryavtsev, I. Shelukhina, C. Vulfius [et al.] // *Toxins*. – 2015. – Vol. 7. – № 5. – P. 1683-1701.
125. Finding a “kneedle” in a haystack: Detecting knee points in system behavior / V. Satopää, J. Albrecht, D. Irwin, B. Raghavan // *Proceedings - International Conference on Distributed Computing Systems*. – 2011. – P. 166-171.
126. Utkin, Y. N. Three-finger toxins, a deadly weapon of elapid venom - Milestones of discovery / Y. N. Utkin // *Toxicon*. – 2013. – Vol. 62. – P. 50-55.
127. The three-finger toxin fold: a multifunctional structural scaffold able to modulate cholinergic functions / P. Kessler, P. Marchot, M. Silva, D. Servent // *Journal of Neurochemistry*. – 2017. – Vol. 142. – P. 7-18.
128. Computational studies of snake venom toxins / P. G. Ojeda, D. Ramírez, J. Alzate-Morales [et al.] // *Toxins*. – 2018. – Vol. 10. – № 1 – P. 8.
129. Heterodimeric neurotoxic phospholipases A2-The first proteins from venom of recently established species *Vipera nikolskii*: Implication of venom composition in viper systematics / A. S. Ramazanova, L. L. Zavada, V. G. Starkov [et al.] // *Toxicon*. – 2008. – Vol. 51. – № 4. – P. 524-537.
130. Human neuromodulator SLURP-1: Bacterial expression, binding to muscle-type nicotinic acetylcholine receptor, secondary structure, and conformational heterogeneity in solution / M. A. Shulepko, E. N. Lyukmanova, A. S. Paramonov [et al.] // *Biochemistry (Moscow)*. – 2013. – Vol. 78. – № 2. – P. 204-211.
131. Human secreted Ly-6/uPAR related protein-1 (SLURP-1) is a selective allosteric antagonist of $\alpha 7$ nicotinic acetylcholine receptor / E. N. Lyukmanova, M. A. Shulepko, D. Kudryavtsev [et al.] // *PLoS ONE*. – 2016. – Vol. 11. – № 2 – P. 0149733.