

МИНОБРНАУКИ РОССИИ

Федеральное государственное бюджетное образовательное учреждение
высшего образования
«САРАТОВСКИЙ НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ ИМЕНИ Н.Г. ЧЕРНЫШЕВСКОГО»

Кафедра общей и неорганической химии

Разделение образцов гепарина по степени очистки и происхождению
методами спектроскопии и хемометрики

АВТОРЕФЕРАТ МАГИСТЕРСКОЙ РАБОТЫ

студентки 2 курса 251 группы

направления (специальности) 04.04.01 «Химия»

Института химии

Кривец Ольги Олеговны

Научный руководитель
Профессор кафедры,
д.х.н., доц.

Н.А. Бурмистрова
подпись, дата

Зав. кафедрой
д.х.н., доц.
должность, уч. степень, уч. звание

Д.Г. Черкасов
подпись, дата

Д.Г. Черкасов
инициалы, фамилия

Саратов 2019

Список условных обозначений и сокращений

МГК	Метод главных компонент
ГК	Главная компонента
SIMCA	Soft Independent Modeling of Class Analogy
ПЛС	Проекция на латентные структуры
<i>PLSR</i>	<i>PLSR</i>
SEN	Sensitivity (чувствительность)
SPC	Specificity (специфичность)
EFF	Efficiency (эффективность)

Введение

В настоящее время проблема контроля качества лекарственных препаратов и исходных ресурсов для их получения, имеет большое значение. Для решения этого вопроса, как правило, используются методы ЯМР-спектроскопии и хроматографии. Однако скрининг большого количества образцов, который также может проводиться вне лаборатории, ставит задачу поиска альтернативных подходов.

В этом направлении определенный интерес представляет использование хеометрических методов обработки экспериментальных данных, полученных широкодоступными и низко затратными методами. Применение спектроскопических методов в сочетании с хеометрическими подходами существенно расширяет возможности анализа, позволяет решать как количественную, так и классификационную задачу [1].

Объект данного исследования – гепарин, представляет собой биологический полимер – сульфатированный гликозаминогликан. Получил широкое применение в медицине и в основном используется в качестве антикоагулянта крови [2].

Наличие посторонних примесей в лекарственном препарате на основе гепарина может способствовать появлению побочных эффектов на организм. Присутствие близких к гепарину по структуре и биологическим свойствам соединений, таких как хондроитинсульфат, может привести к изменению биологической активности препарата и следовательно его неверной дозировке [3]. Опасность применения лекарственного препарата недостаточно высокого качества ставит задачу контроля за качеством исходного сырья при производстве гепарина и препаратов на его основе.

Целью работы явилась классификации образцов гепарина различными спектроскопическими методами (УФ-спектроскопия, флуоресценция, ИК-спектроскопия) на основе хеометрической обработки данных.

В работе были поставлены следующие задачи:

- 1) Провести сравнение между эффективностью использованием данных, полученных различными спектроскопическими методами;
- 2) Изучить возможности применения различных видов хеометрической обработки данных: решение одноклассификационной задачи методом SIMCA, многоклассификационной – методом ПЛС, К-средних и МГК,
- 3) Установить влияние предварительной обработки спектральных данных на результат разделения;
- 4) Подобрать оптимальные параметры и условия для решения данной классификационной задачи.

В работе были использованы различные хеометрические алгоритмы: PCA (Principle Component Analysis) [4], алгоритм для проведения одноклассового анализа SIMCA (Soft Independent Modeling of Class Analogy) [5] и алгоритм многоклассового анализа PLS (Partial Least Squares) [6]. Среди достоинств данных методов следует выделить тестирование на достоверность получаемых результатов, кластерный анализ выделенных спектров на остаточные зависимости и объединение сильно зависящих сигналов в многомерные источники.

Основное содержание работы

Метод главных компонент, МГК (Principal component analysis, PCA)

Традиционный анализ спектральных данных основан на выборе из всего набора значений одной или нескольких переменных, отвечающих за наибольшую информацию о соединении и его количественном содержании, как правило, значение, приходящееся на пик поглощения, испускания и т.п. Такой подход удобен, но при этом часть нужной информации может быть потеряна.

В широко распространенном в настоящее время МГК [7] вместо одной точки выбирается ряд наиболее информативных значений переменных. Основой МГК является декомпозиция (разделение) исходной матрицы данных на структурную часть (главные компоненты, которые лежат в направлении максимальных различий данных) и на шум (различия между положением точек достаточно мала, и ей можно пренебречь) [1].

С математической точки зрения МГК - это декомпозиция исходной 2D-матрицы X , т.е. представление ее в виде произведения двух 2D-матриц T и P :

$$X = TP + E = \mathbf{A} + E, \quad (5)$$

В уравнении (5) T называется матрицей счетов (scores), P - матрицей нагрузок (loadings), E - матрицей остатков. Число столбцов t_a в матрице T и p_a матрице P равно эффективному (химическому) рангу матрицы X . Эта величина A называется числом главных компонент (ГК) и она меньше, чем число столбцов в матрице X . Цель такого преобразования максимально точно описать матрицу X с использованием минимального числа главных компонент [1].

Применение МГК к анализу спектральных данных позволяет распределить множество исследуемых объектов по группам на основе главных компонент. А также трансформировать выделенные матрицы счетов и нагрузок дополнительными хемометрическими методами (например, методом чередующихся наименьших квадратов) в спектры индивидуальных компонентов.

Метод Формального независимого моделирования аналогий классов (Soft Independent Modeling of Class Analogy, SIMCA)

Метод SIMCA относится к современным классификационным методам одноклассового анализа обучения с учителем. Метод обладает некоторыми особенностями: каждый класс моделируется обособленно и независимо от остальных, поэтому и называется одноклассовым. Второй особенностью является многозначная классификация, то есть один образец может быть отнесён к нескольким классам одновременно, также позволяет проводить расчет вероятности ошибочной классификации. Для построения классификационной модели требуется задать целевой класс - обучающий набор, представленный матрицей X , размерностью I образцов и J переменных. В качестве переменных могут быть выбраны спектральные данные образцов [8].

Суть метода SIMCA заключается в построении классификатора (правила) по которому любой новый образец x либо принимается как принадлежащий тому же классу, что и X , либо отвергается.

Первоначальным этапом метода SIMCA является разложение исходной матрицы X аналогично МГК:

$$X = TP^T + E \quad (6),$$

где T – матрица счетов, P – матрица нагрузок, E – матрица ошибок. Тогда каждый образец x (обучающий, проверочный или новый) может быть

спроецирован на пространство главных компонент, т.е. найден вектор его счетов:

$$\mathbf{t} = \mathbf{x}^P = (t_1, \dots, t_A)^T \quad (7).$$

Каждый элемент данных можно представить как сумму двух векторов: лежащего в найденной гиперплоскости (проекция) и перпендикулярного гиперплоскости (остаток). Длины этих векторов являются важными показателями, характеризующими принадлежность точки к классу. Они называются, соответственно, размахом $\mathbf{h}$ и отклонением $\mathbf{v}$. Как только появляется новая точка – кандидат на принадлежность к классу – ее можно спроецировать на уже имеющееся подпространство и определить ее собственные характеристики:

$$\mathbf{v} = \frac{1}{J} \sum_{j=1}^J \mathbf{e}_j^2 \quad () \quad \text{и} \quad \mathbf{h} = \mathbf{t}^T (\mathbf{T}^T \mathbf{T})^{-1} \mathbf{T} = \sum_{a=1}^A \left(\frac{t_a}{S_a} \right) \mathbf{i}^a \quad (8).$$

Решение о принадлежности нового образца к целевому классу определяется сравнением этих величин с критическими уровнями, найденными по обучающему набору [9].

Метод Проекции на латентные структуры, ПЛС (Partial Least Squares Discriminant Analysis)

Данный метод относится к методам обучения с учителем. Преимуществом ПЛС в сравнении с методом SIMCA является возможность проведения одновременного многоклассового анализа. Его основа состоит в выборе обучающего набора – некоторого количества образцов, принадлежащих разным классам. На их основе строится классификация тестового набора образцов.

Метод PLS-DA представляет собой регрессионный подход, в котором находится взаимосвязь между матрицей признаков $\mathbf{X}$ (размером $\mathbf{I} \times \mathbf{J}$) и матрицей фиктивных откликов $\mathbf{Y}$ (размером $\mathbf{I} \times \mathbf{K}$). Элементы матрицы фиктивных откликов $\mathbf{Y}$ равны либо 0, либо 1 в зависимости от

принадлежности объекта соответствующему классу. Между матрицей признаков X и матрицей откликов Y строят регрессию. Полученная регрессионная модель используется для расчета предсказанных значений отклика. Классификационное правило в PLS-DA – образец принадлежит к тому классу, для которого соответствующий элемент y_k ближе к единице [9].

В зависимости от метода подсчета расстояния от образца до центра класса выделяют жёсткую Hard PLS-DA и мягкую классификацию Soft PLS-DA. При жёсткой однозначной классификации используется линейный дискриминантный анализ, при мягкой – квадратичный дискриминантный анализ. Мягкая классификация является многозначной и позволяет выделить образцы, которые находятся на границе классов [10].

Результатом анализа методом ПЛС является классификационная таблица – матрица соотнесения образцов и отображение параметров классификации: чувствительности и специфичности.

Гепарин относится к семейству гликозаминогликанов; его молекула представлена несколькими полисахаридными цепями, связанными с общим белковым ядром. Белковое ядро включает в свой состав остатки двух аминокислот: серина и глицина (рисунок 1).

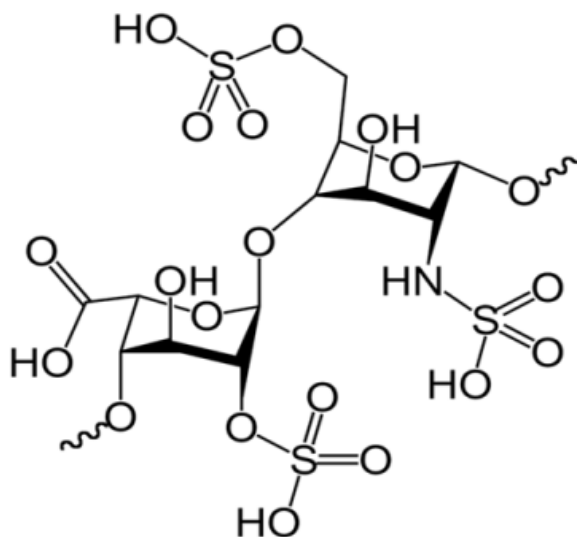


Рисунок 1- Структурная формула гепарина.

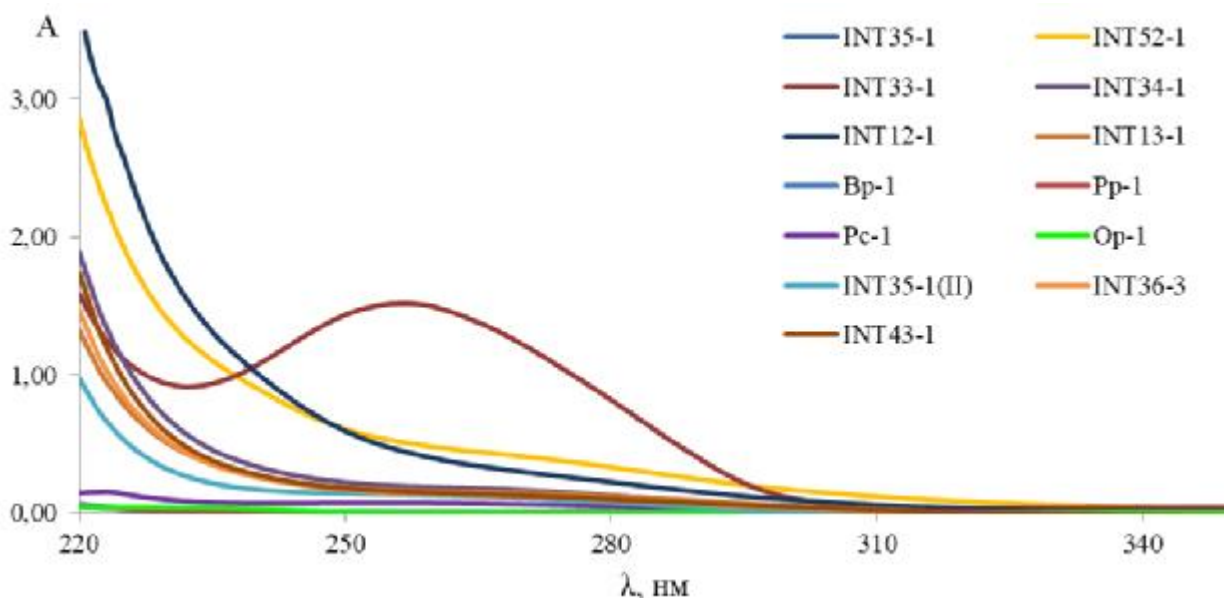
Были исследованы образцы гепарина являющиеся промежуточными продуктами при производстве лекарственных препаратов на основе гепарина, получены экстракционным методом из органов (печень, лёгкие, слизистые оболочки желудка и кишечника) различных животных (бык, овца, свинья). Образцы представляют собой твёрдые мелкодисперсные порошки в основном белого цвета, окраска неочищенных образцов варьируется от светло-жёлтой до коричневой. Образцы отличаются производителями (партии), степенью очистки и по животному происхождению. Характеристики образцов представлены в таблице 1.

Таблица 1 – Образцы гепарина, использованные в работе

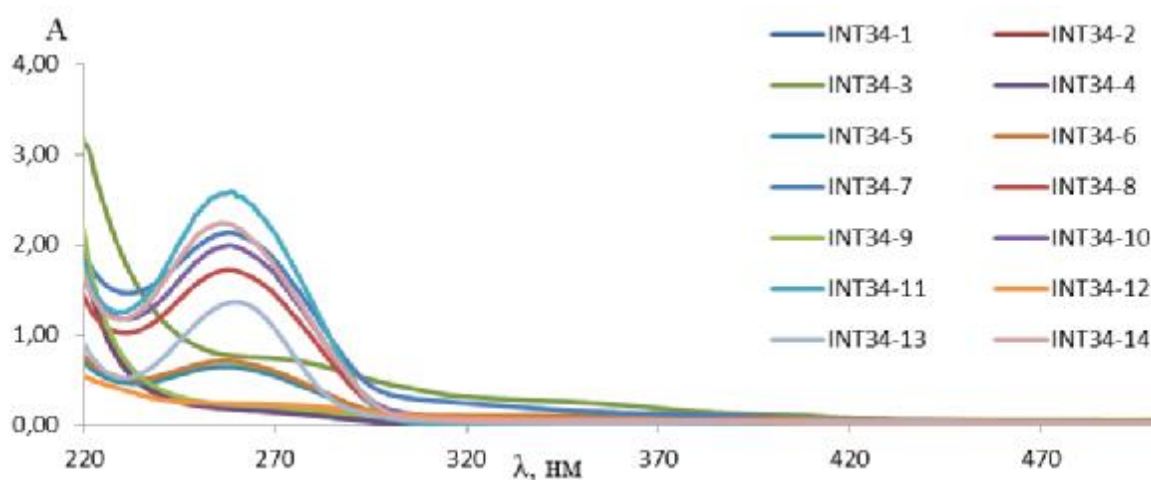
Группа образцов	Степень очистки	Животное	Количество образцов в группе	Партии
Вр	Очищенный	Бык	10	Первая партия образцов
Рр	Очищенный	Свинья	12	
Рс	Неочищенный	Свинья	5	
Ор	Очищенный	Овца	9	
INT35(I)	Очищенный	Свинья	21	Вторая партия образцов
INT52	Очищенный	Свинья	19	
INT33	Неизвестно	Бык	8	
INT12	Неочищенный	Неизвестно	6	
INT13	Неизвестно	Бык	3	
INT34	Неочищенный	Неизвестно	14	
INT35(II)	Очищенный	Свинья	9	Третья партия образцов
INT36	Очищенный	Свинья	5	
INT43	Очищенный	Свинья	5	

Классификация образцов методом УФ-спектроскопии

Анализ спектров образцов гепарина из различных групп показал, что спектры образцов очищенного гепарина похожи, спектры образцов неочищенного гепарина различаются даже в пределах одной группы (рисунок 2 а,б).



(а)



(б)

Рисунок 2 - Спектры поглощения образцов гепарина из различных групп (а) и неочищенного гепарина INT34 (б).

График счетов в пространстве ГК1-ГК2 показывает разделение

образцов относительно ГК-1 на две группы: с высокой и низкой степенью очистки (рисунок 3). При этом только некоторая часть образцов с высокой степенью очистки (INT35-1 (I), INT35-21 (II), INT-52-13, INT52-10, INT52-5) вошли в группу неочищенных образцов.

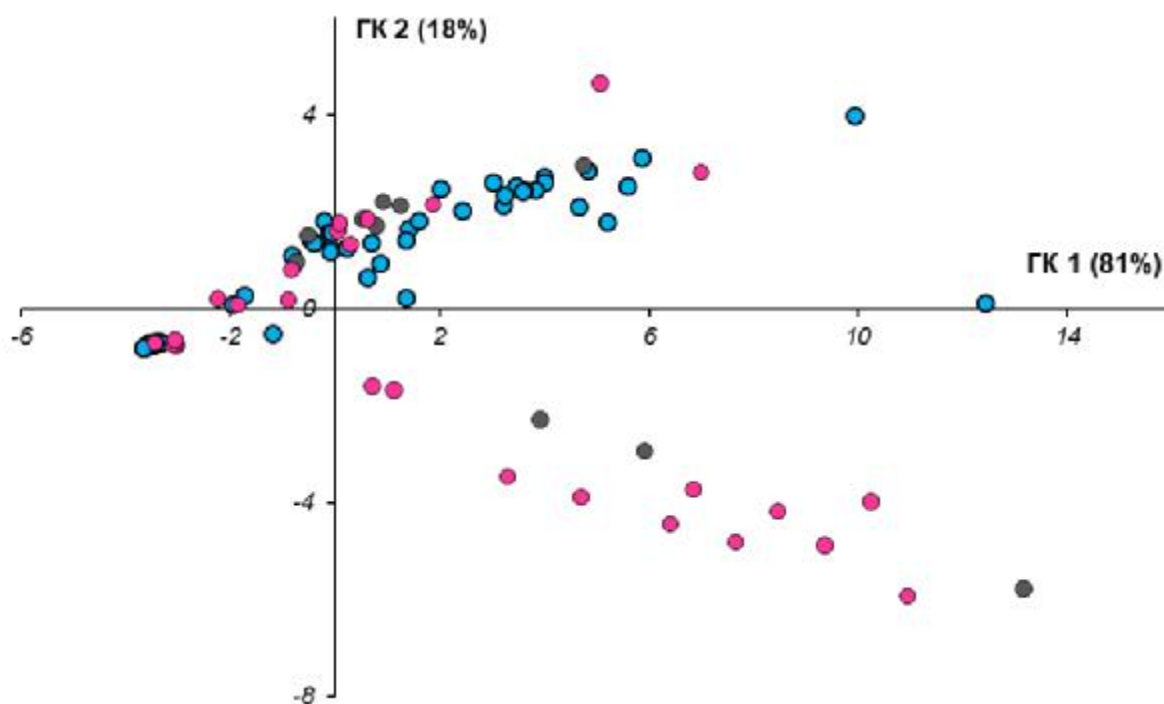


Рисунок 3 - График счетов распределения образцов в пространстве ГК: **очищенные**, **неочищенные** и образцы с неизвестной степенью очистки

Классификация образцов методом флуоресцентной спектроскопии

Спектр флуоресценции образцов гепарина первой партии представлен на рисунке 4. Наибольшей интенсивностью отличаются образцы овечьего очищенного (Op) и свиного неочищенного гепарина (Pc). Спектры флуоресценции образцов очищенного овечьего гепарина не отличаются однородностью в отличие от УФ-спектров.

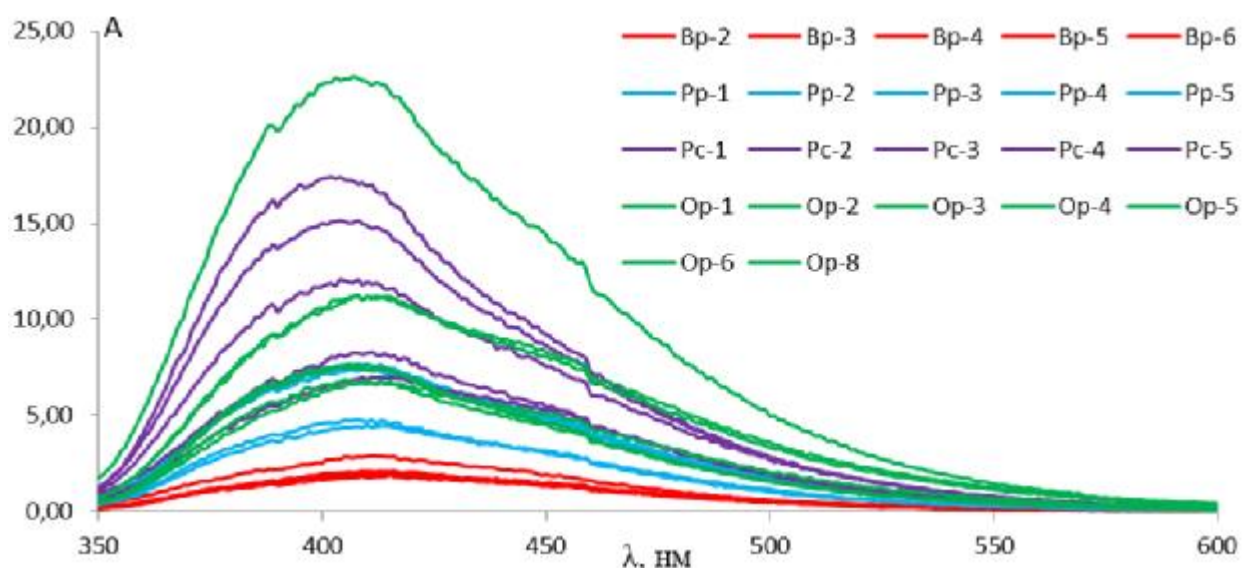


Рисунок 4 - Спектр флуоресценции первой партии образцов гепарина

Для построения хемометрической модели SIMCA для разделения образцов очищенного гепарина по животному происхождению (бычий, свиной, овечий) было использовано 91 образец. Наилучшие результаты были получены при использовании образцов бычьего гепарина в качестве целевого класса: (чувствительность 100%) , 45 образцов были использованы в качестве тестового набора (свиной, овечий); 37 образцов (свиной, овечий) были использованы в качестве определяемого набора со специфичностью 94% для образцов свинины и 100% для образцов баранины (таблица 2).

Таблица 2 – Результаты классификации образцов очищенного гепарина по животному происхождению методом SIMCA

Тестовый класс	Целевой класс		
	Бычий	Свиной	Овечий
	Специфичность, %		
Бычий	-	11	80
Свиной	94	-	74
Овечий	100	17	-

Классификация образцов методом ИК-спектроскопии

В области 1400 – 3400 см⁻¹ наблюдается различия между образцами очищенного и неочищенного свиного гепарина (рисунок 5).

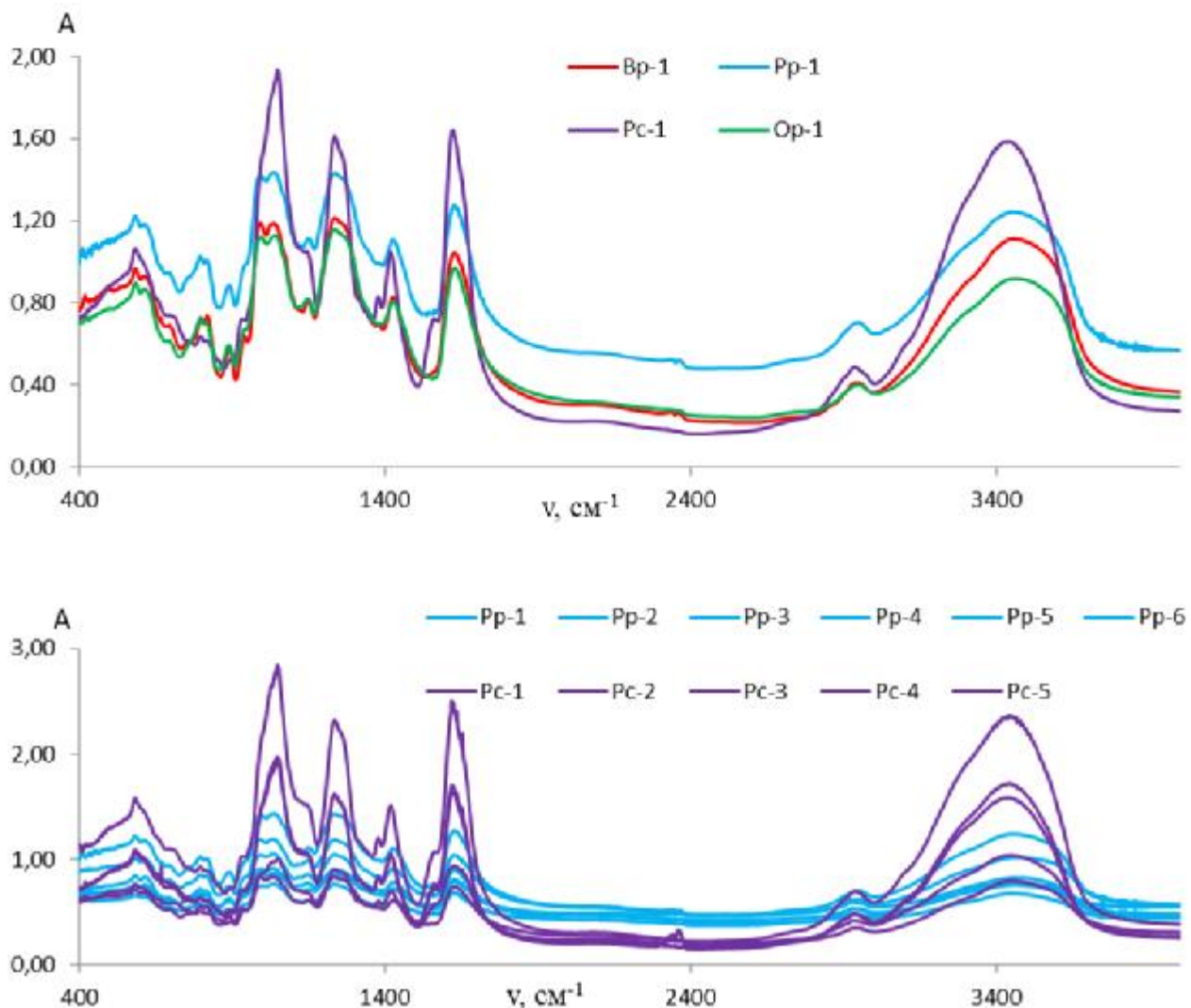


Рисунок 5 – ИК-спектры образцов гепарина

МГК анализ показал чёткое разделение образцов свиного очищенного и неочищенного гепарина, ГК-1 описывает 53%, ГК-45% объясненной дисперсии (рисунок 5). Согласно графику нагрузок (рисунок 6) возможно использование первых двух главных компонент, последующие компоненты несут в себе недопустимое количество шума.

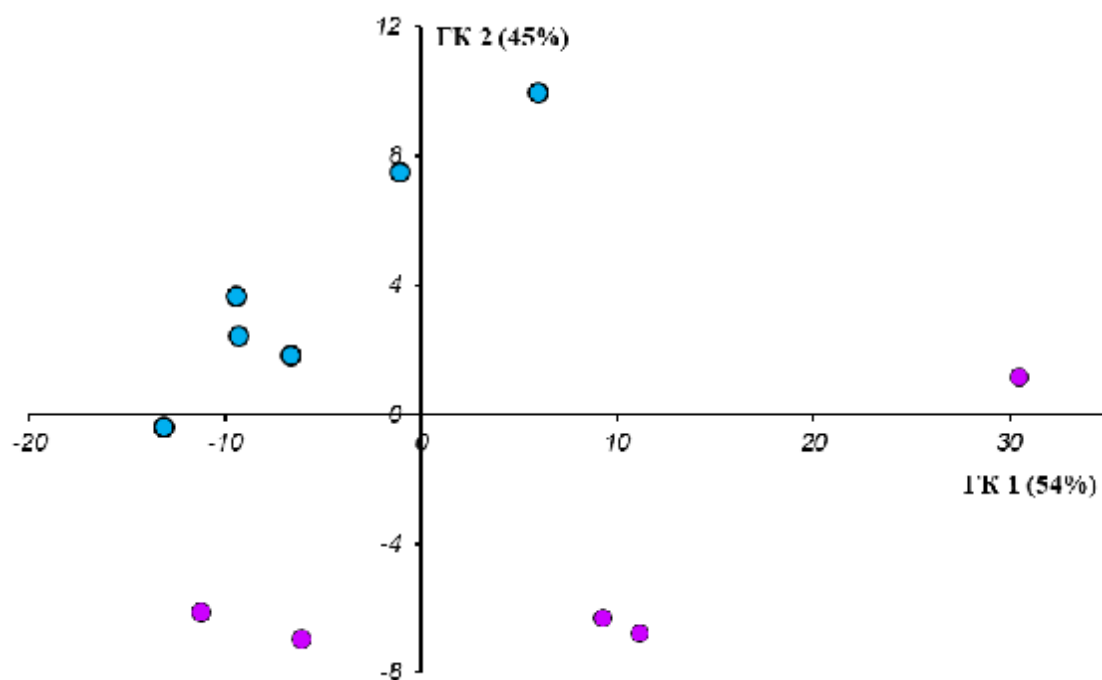


Рисунок 6 - График счетов распределения образцов гепарина на
очищенные, неочищенные

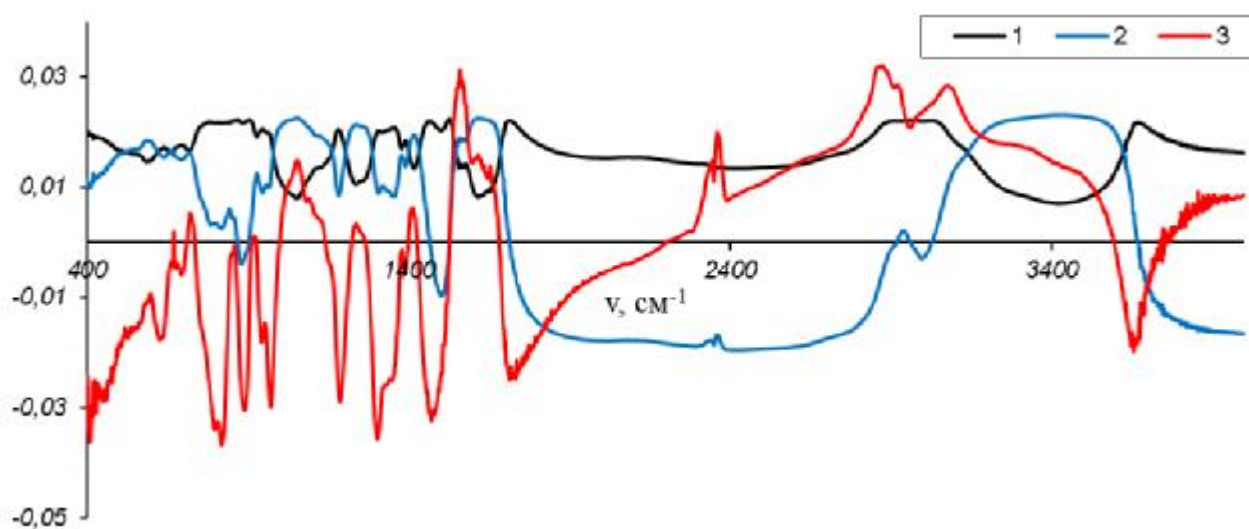


Рисунок 7 - График нагрузок: ГК 1, ГК 2, ГК 3

Использование в качестве исходных данных ИК-спектров образцов гепарина не позволяет успешно разделить образцы методом К-средних по степени очистки (рисунок 8), это связано с сильным отличием группы овечьего гепарина по сравнению с очищенными образцами свиного и бычьего гепарина.

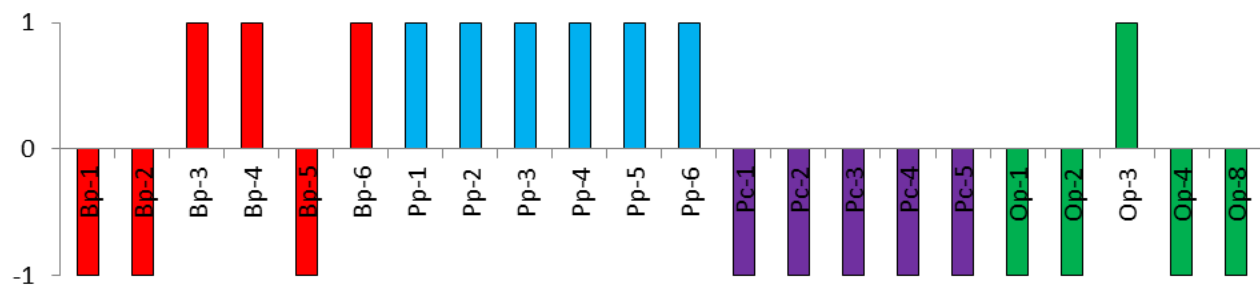


Рисунок 8 - Разделение образцов методом К-средних по степени очистки: **бычий**, **свиной**, **овечий** и **неочищенный** гепарин

Заключение

1. Изучена возможность применения спектральных методов (абсорбционная, люминесцентная, ИК-спектроскопия) с применением хемометрических алгоритмов (МГК, SIMCA, ПЛС) для классификации образцов гепарина по производителю, степени очистки и происхождению.
2. Показано, что наилучшее разделение образцов гепарина по партиям, степени очистки и происхождению при использовании метода главных компонент наблюдается при анализе УФ спектров.
3. Показана возможность применения метода К-средних для разделения образцов по степени очистки и животному происхождению с использованием спектральных методов (абсорбционная, люминесцентная, ИК-спектроскопия).

Список использованных источников

1. Хемометрика в спектроскопии: учебное пособие для студентов Института Химии/ Монахова Ю.Б., Муштакова С.П. Саратов: Изд-во СГУ, 2011. - 70 с.
2. Пирогов А. Л. Гепарин-индуцированная тромбоцитопения при сердечно-сосудистых вмешательствах. Механизмы развития, диагностики и методы коррекции / А. Л. Пирогов // Журн. Комплексные проблемы сердечно-сосудистых заболеваний. - 2014. - Т.15. - С.95-103.
3. Шкарина Т. Н. Анализ и стандартизация хондроитинсульфата натрия: автореф. дис. канд. фарм. наук / Шкарина Т. Н. - Москва, 2009. - 25 с.
4. R. Bro, Age K. Smilde. Principal component analysis / Bro R. // J. Anal. Methods. - 2014. - V.6. - P.2812-2831.
5. Zontov Y.V., Rodionova O.Ye., Kucheryavskiy S.V., Pomerantsev A.L. DD-SIMCAA MATLAB GUI tool for data driven SIMCA approach / Y.V. Zontov // J.Chemometrics and Intelligent Laboratory Systems. - 2017. - V.167. - P. 23-28.
6. Rodionova O.Ye., Oliveri P., Pomerantsev A.L., Rigorous and compliant approaches to one-class classification / O.Ye. Rodionova // J. Chemom. Intell. Lab. Syst. – 2016. – V. 159. – P. 89-96.
7. A.L. Pomerantsev, O.Ye. Rodionova. Multiclass Partial Least Squares Discriminant Analysis: Taking the right way – A critical tutorial / A. L. Pomerantsev // J. Chemom. - 2018. - V. 32, № 8: e3030.
8. Jiao Zheng, Tai Ye, Jinyang Chen, Li Xu, Xinghu Ji, Chuluo Yang, Zhike He. Highly sensitive fluorescence detection of heparin based on aggregation-induced emission of a tetraphenylethene derivative / Jiao Zheng // J. Biosens. Bioelectron. - 2017. - V.90, № 15. - P. 245-250.
9. Jia Jun, Liu Dan, Yuan Hong, Zhi Zhang, Yuan Ding, Lu Ni, Wang Hong, Yan Zou, Jian Wang. Rapid detection of heparin by gold nanorods and near-infrared fluorophore ensemble based platform via nanometal surface energy transfer / Jia Jun // J. Sens. Actuators, B. - 2018. - V.274, № 20. - P. 318-323.

infrared fluorophore ensemble based platform via nanometal surface energy transfer / Jia Jun // J. Sens. Actuators, B. - 2018. - V.274, № 20. - P. 318-323.

10. Yuxia Gao, Kailun Wei, Jianzuo Li, Ying Li, Jun Hu. A facile four-armed AIE fluorescent sensor for heparin and protamine / Yuxia Gao // J. Sens. Actuators, B. - 2018. - V.277, № 20. - P. 408-414.

24.06.19.

A handwritten signature in black ink, appearing to be 'Jia Jun', written in a cursive style.