

# Исследование комбинационного рассеяния белковых фракций человеческой крови с использованием метода проекций на латентные структуры

А.А. Лыкина<sup>а</sup>, Д.Н. Артемьев<sup>а</sup>, И.А. Братченко<sup>а</sup>, Ю.А. Христофорова<sup>а</sup>, О.О. Мякинин<sup>а</sup>,  
Т.П. Кузьмина<sup>б</sup>, И.Л. Давыдкин<sup>б</sup>, В.П. Захаров<sup>а</sup>

<sup>а</sup> Самарский национальный исследовательский университет имени академика С.П. Королева, 443086, Московское шоссе, 34, Самара, Россия

<sup>б</sup> Самарский государственный медицинский университет, 443099, ул. Красноармейская, 1Б, Самара, Россия

---

## Аннотация

Данная работа посвящена исследованию белковых фракций человеческой крови методом спектроскопии комбинационного рассеяния. В качестве образцов использовали цельную кровь и плазму. Исследуемое вещество было помещено в алюминиевую кювету. Было произведено сравнение двух методов обработки данных для выделения спектров комбинационного рассеяния из регистрируемого сигнала: полиномиальная аппроксимация и коррекция базовых линий с помощью метода ассиметричных наименьших квадратов. Анализ спектров комбинационного рассеяния осуществлялся с использованием метода проекций на латентные структуры. Полученные результаты показывают, что метод спектроскопии комбинационного рассеяния позволяет уверенно определять аномальные концентрации общего белка в плазме и цельной крови.

*Ключевые слова:* кровь; плазма; спектроскопия комбинационного рассеяния; метод проекций на латентные структуры

---

## 1. Введение

В плазме крови человека содержится более 100 различных белков крови, которые имеют различные функции и молекулярную структуру. Нарушение соотношений белковых фракций крови приводит к различным заболеваниям, связанных с развитием определенных патологий, в том числе онкологических заболеваний, туберкулеза и инфекционных болезней.

В клинических исследованиях для определения белковых фракций крови используют физико-химические методы, один из самых распространенных методов является электрофорез [1]. Он основан на разной подвижности белков под воздействием электрического поля.

В данной работе для малоинвазивного исследования белковых фракций использовался метод спектроскопии комбинационного рассеяния (КР), который позволяет судить о структуре объекта, не повреждая его. Выбранный метод использовался без дополнительного оптического оборудования (конфокальные системы, микрообъектив) и внешних реагентов. Данный подход неоднократно применялся для исследования различных белков крови [2, 3]. Также в работе проводилось сравнение двух различных методов выделения «чистого» КР спектра от автофлуоресцентного фона: полиномиальная аппроксимация и метод ассиметричных наименьших квадратов.

## 2. Основная часть

### 2.1. Экспериментальная установка

Для регистрации спектров комбинационного рассеяния применяли измерительную систему в состав которой входят узкополосный источник возбуждения – лазер LML-785.0RB-04 (мощность излучения на выходе из пробника 200 мВт, длина волны излучения 785 нм), оптический модуль фильтрации (КР пробник, RPB785 фирмы InPhotonics) и регистрирующий приемник – спектрограф Sharmrock SR-500i с интегрированной камерой ANDOR DV-420A-OE, матрица которой захлаживается до -70°C, обеспечивающий регистрацию оптических сигналов низкой интенсивности.

### 2.2. Исследуемые образцы

В эксперименте были протестированы 45 образцов плазмы и цельной крови. Исследуемые вещества хранились в лабораторном холодильнике с температурой +4°C. При проведении эксперимента образцы были помещены в алюминиевую кювету цилиндрической формы с объемом 0,9 мл.

### 2.3. Методы и обработки данных

Для анализа экспериментальных данных используется метод проекций на латентные структуры (Projection on Latent Structures - PLS) [4], он предоставляет возможность интерпретировать результаты в доступной форме, основываясь на меньшем количестве билинейных компонентов. Для построения модели применяются латентные структуры (LV),

которые согласовано строятся на максимальной корреляции между вектором  $Y$  (концентрация общего белка крови) и матрицей  $X$  (интенсивность спектров КР плазмы и цельной крови на различных длинах волн). Погрешность результатов прогноза высчитывается критерием приближения - среднеквадратичным остатком градуировки (RMSE).

Перед построением аналитической модели использовались различные методы предобработки данных, которые реализованы в облачном сервисе TP<sup>T</sup>cloud<sup>beta</sup> (<https://tptcloud.com/>). В данной работе применялось центрирование исследуемых данных (mean center), калибровка данных, используя стандартное отклонение случайной величины с нормальным распределением (Standard Normal Variate - snv). Для уменьшения шумов у регистрируемых данных использовался сглаживающий фильтр Савицкого Голея (sav gol), а избавление от фонового излучения осуществлялось коррекцией базовой линии с помощью метода асимметричных наименьших квадратов (Baseline Correction with Asymmetric Least Squares – baseline als). Другим подходом для выделения спектров КР от автофлуоресценции у регистрируемых сигналов является метод полиномиальной аппроксимации с дополнительным сглаживанием регистрируемого сигнала. В результате проведенной обработки получаем «чистый» спектр КР, состоящий из характерных полос, возникающих на определенных сдвигах частот.

Для оценки аналитической модели и ее поведения на «независимых» данных использовался метод кросс-валидации. В данном исследовании была выбрана модель перекрестной проверки: Leave One Out (LOO) [4]. С помощью нее идет проверка модели с исключением по одному спектру.

#### 2.4. Результаты исследования и обсуждения

В исследовании плазмы и цельной крови производилась регистрация спектров КР, где были выявлены основные позиции пиков белковых фракций крови:  $650\text{ см}^{-1}$  (колебания триптофана),  $830\text{ см}^{-1}$  (деформационные колебания СН группы),  $880\text{ см}^{-1}$  (колебания фенилаланина),  $960\text{ см}^{-1}$  (деформационные колебания СН группы),  $1002\text{ см}^{-1}$  (колебания фенилаланина),  $1064\text{ см}^{-1}$  (Амид III в  $\beta$ -спирали),  $1210\text{ см}^{-1}$  (Амид III в  $\alpha$ -спирали),  $1270\text{ см}^{-1}$  (Амид III в  $\alpha$ -спирали),  $1350\text{ см}^{-1}$  (колебания ароматических кислот),  $1450\text{ см}^{-1}$  (деформационные колебания  $\text{CH}_2$  группы),  $1545\text{ см}^{-1}$  (колебания фенилаланина),  $1650\text{ см}^{-1}$  (Амид III в  $\beta$ -спирали) [5].

Рассмотрим регистрируемые сигналы плазмы крови, для которых были выбраны методы предобработки регистрируемых спектров: sav gol, snv и mean center. На рисунке 1 представлены спектры КР плазмы крови с пиками на следующих длинах волн:  $650\text{ см}^{-1}$ ,  $880\text{ см}^{-1}$ ,  $1002\text{ см}^{-1}$ ,  $1350\text{ см}^{-1}$ ,  $1450\text{ см}^{-1}$ ,  $1650\text{ см}^{-1}$ . Для анализа экспериментальных данных было выбрано  $LV=5$ , так как на 5 векторе начинается расхождение значений RMSE калибровочного и проверочного набора данных (кросс-валидация). При помощи PLS анализа были рассчитаны среднеквадратичный остаток градуировки общего белка плазмы крови человека:  $LV=5$ ,  $RMSE=8\text{ г/л}$  и  $RMSE_{cv}=8,8\text{ г/л}$ .

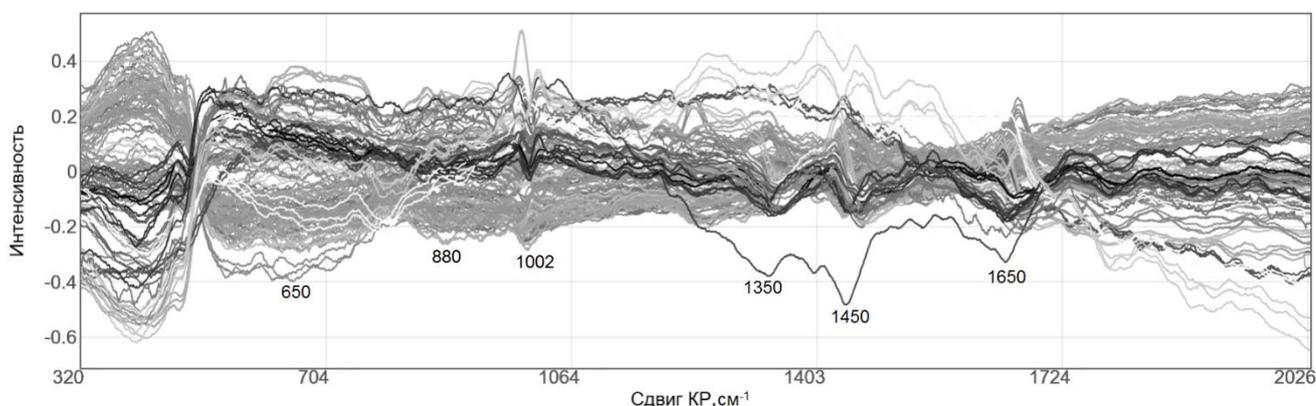


Рис. 1. Спектры КР плазмы крови после применения методов предобработки: sav gol, snv и mean center.

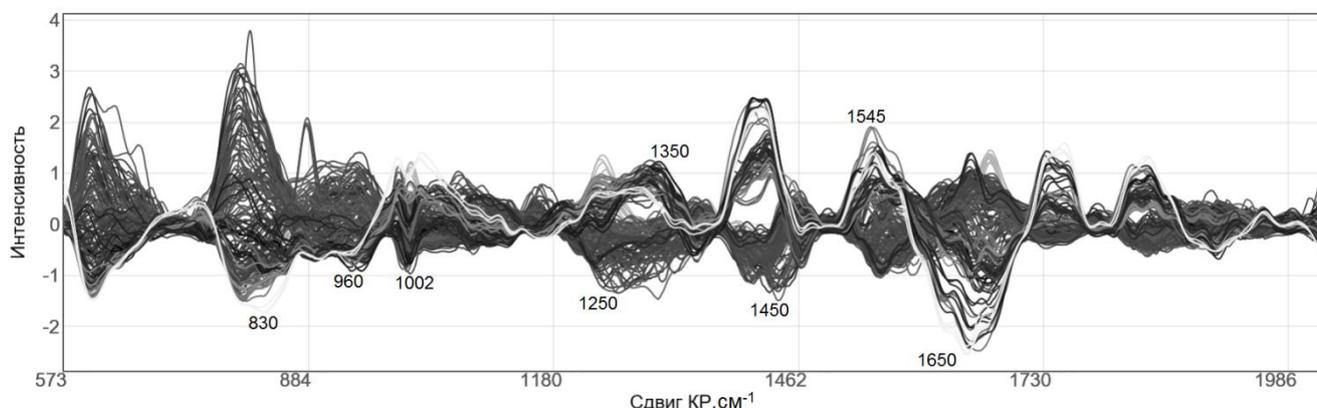


Рис. 2. «Чистые» спектры КР плазмы крови, обработанные полиномиальной аппроксимацией и методами snv и mean center.

Для спектров КР плазмы крови, обработанными полиномиальной аппроксимацией, была подобрана модель: *snv* и *mean center* с 3 латентными структурами по которым был проведен расчет погрешностей результатов прогноза общего белка:  $LV=3$ ,  $RMSE=7,5$  г/л и  $RMSE_{cv}=8$  г/л. Так как фон и шумы были вычтены с помощью полиномиальной аппроксимации, то для всех «чистых» спектров КР плазмы крови и цельной крови выбрана модель: *snv* и *mean center*, для которой пики КР имеют максимальную интенсивность. «Чистые» спектры КР, обработанные методами предобработки, представлены на рисунке 2.

Анализируя полученные спектры КР на рисунках 1 и 2 можно сделать вывод, что основные позиции полос совпадают друг с другом, при этом интенсивность конкретных полос может существенно отличаться. Сравнивая значения точности предсказания общего белка по спектрам КР плазмы крови (8,8 г/л и 8 г/л для проверочных наборов данных), обработанные двумя разными моделями, можно сделать вывод, что полученные значения точности предсказания общего белка в плазме крови практически совпадают.

При исследовании спектров КР цельной крови была подобрана модель, состоящая из методов предобработки: *sav gol*, *snv* и *mean center* и  $LV=4$  с погрешностью результатов прогноза:  $RMSE=14,3$  г/л и  $RMSE_{cv}=15,4$  г/л. Для предобработки спектров КР цельной крови, обработанными полиномиальной аппроксимацией, использовали также, как и в случае плазмы, *snv* и *mean center*, но при этом  $LV=4$ . Для данной модели были получены следующие значения:  $RMSE=5$  г/л и  $RMSE_{cv}=6,6$  г/л.

По результатам, полученным в ходе проведенной работы, можно сделать вывод, что точность определения концентрации общего белка плазмы крови по виду спектров КР в 1,5-2 раза выше, чем значение точности определения общего белка цельной крови. Снижение точности связано с зашумленностью регистрируемого сигнала цельной крови, так как цельная кровь в своем составе имеет больше компонентов по сравнению с плазмой крови.

### 3. Заключение

В данной работе было проведено исследование методов предобработки регистрируемых сигналов плазмы и цельной крови. При этом полученные результаты показывают возможность исследовать не только «чистые» спектры комбинационного рассеяния (полиномиальная аппроксимация), но и регистрируемые сигналы обработанных сглаживающим фильтром Савицкого Голея, калибровкой с использованием стандартного отклонения случайной величины с нормальным распределением и центрированием.

Точность прогнозирования общего белка крови при использовании методов спектроскопии комбинационного рассеяния и анализа проекций на латентные структуры позволяет уверенно определять концентрацию общего белка в плазме крови и цельной крови в диапазоне, соответствующих норме, равном 15-20 г/л.

### Благодарности

Проведенные исследования были выполнены при поддержке Министерства образования и науки Российской Федерации.

### Литература

- [1] Keren, F.D. Protein Electrophoresis in Clinical Diagnosis / F.D. Keren - London: Hodder Arnold, 2003. – 422 p.
- [2] Artemyev, D.N. Measurement of human serum albumin concentration using Raman spectroscopy setup / D.N. Artemyev, V.P. Zakharov, I.L. Davydkin, J.A. Khriforova, A.A. Lykina, V.N. Konyukhov, T.P. Kuzmina // *Izbrannye Trudy*: in 6 volumes. – Luxembourg: Opt. Quant Electron, 2016. – V. 48. - p. 337
- [3] Annika, M.K. Blood analysis by Raman spectroscopy / M.K. Annika, E. Tae-Woong, J. Oh, M. Hunter // *Izbrannye Trudy*: in 22 volumes. - United States: Optics letters, 2004. – V. 27. – p. 2004
- [4] Esbensen, K. Multivariate Data Analysis / K. Esbensen – New Jersey: In Practice 4-th Ed., 2000. – 587 p.
- [5] Dingari, N.C. Raman Spectroscopy Provides a Powerful Diagnostic Tool for Accurate Determination of Albumin Glycation / N.C. Dingari, G.L. Horowitz, J.W. Kang, R.R. Dasari, I. Barman // *Izbrannye Trudy*: in 7 volumes. - Francisco: PLoS ONE, 2012. – V. 7. – p. 2