

Количественный анализ SERS спектров смесей гликированного и негликированного альбумина

Сафиуллин Р.Р., Афанасьев К.Н., Богинская И.А., Рыжиков И.А.,
Седова М.В., Слипченко Е.А.

ИТПЭ РАН
Москва 2021

Мотивация



Создание автоматизированной системы контроля содержания гликированного альбумина в плазме крови с использованием эффекта ГКР и численных методов

Lucica® Glycated Albumin-L (Asahi-Kasei Pharma Corporation, Япония), Glycated Serum Protein LiquiColor® Assay (EKF Diagnostics Inc., США) современные тесты требуют высокой квалификации персонала и большого количества реактивов

Постановка задачи



Установка зависимости между спектрами гликированного\негликированного альбумина и их относительными концентрациями

Оптимизация процесса получения данных

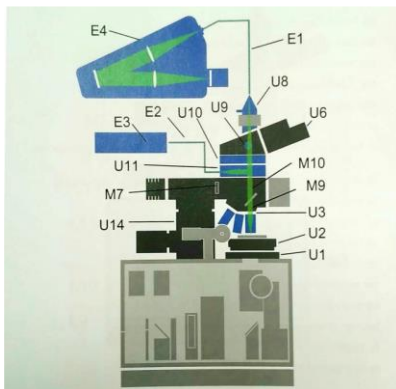
Выявление диабета по данному маркеру

Схема измерений

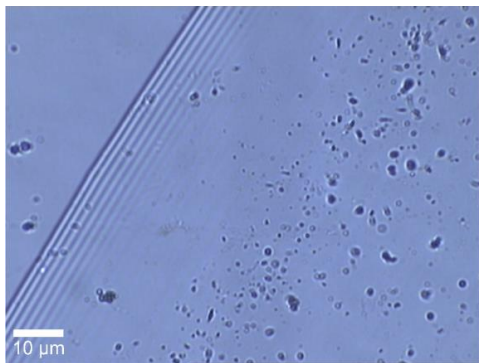


Конфокальный микроскоп

WITec



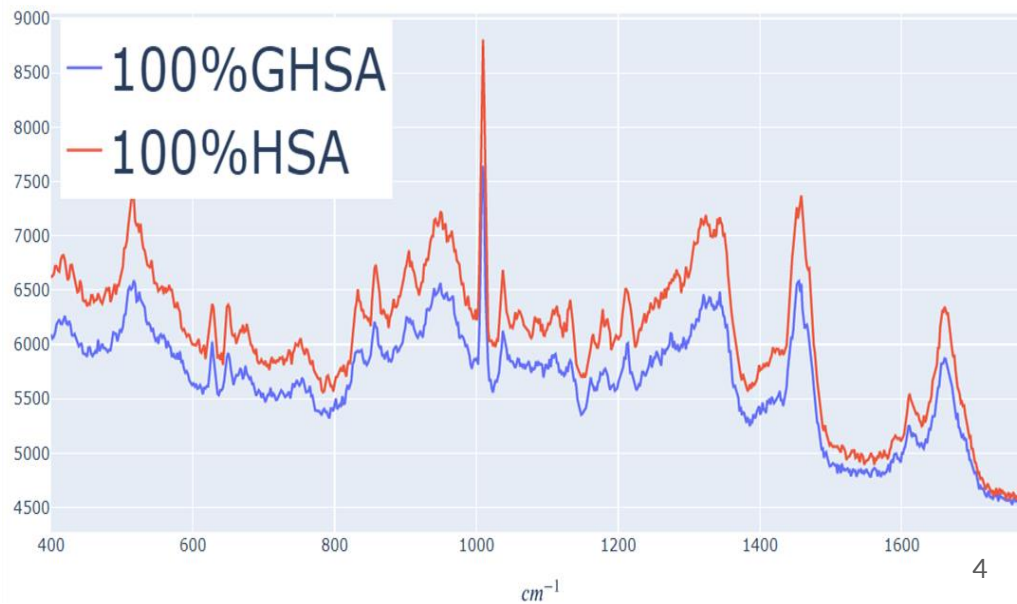
Оптическая схема измерений



Оптическое изображение
высушенной капли

Параметры измерений:

- мощность лазера 56 мВт,
- длина волны возбуждения 785 нм
- время интегрирования спектра 15 сек



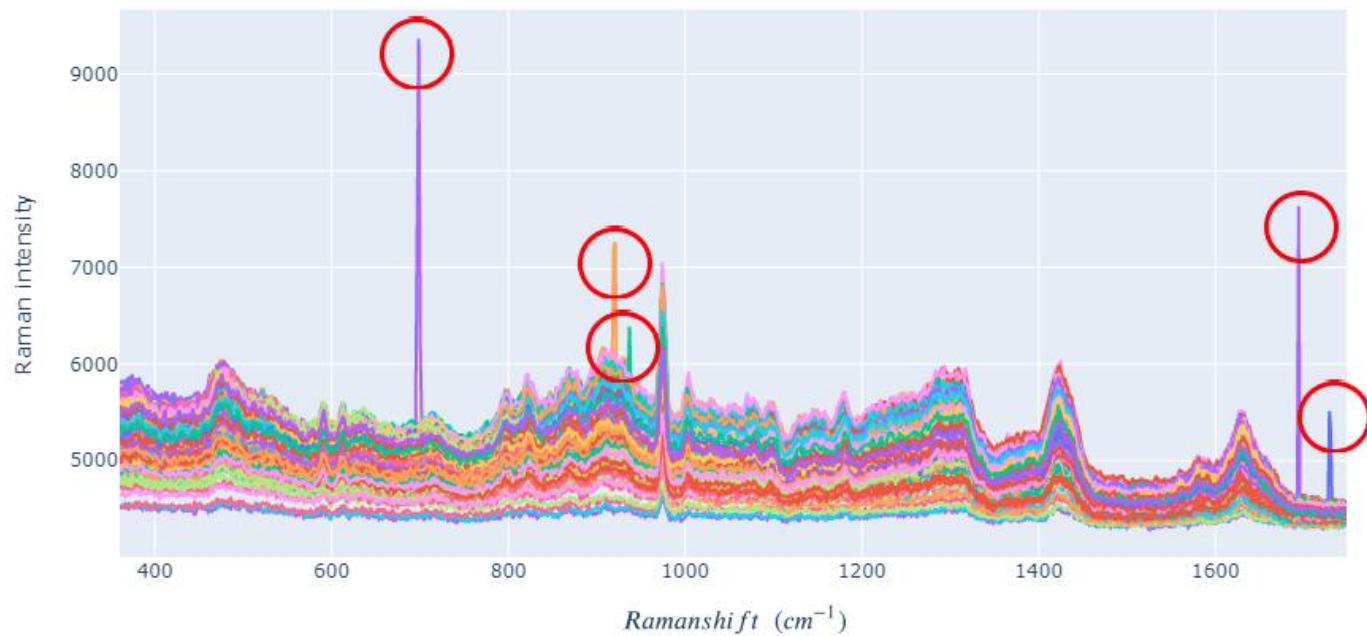


Используемые методы для обработки:

- Удаление выбросов
- Вычитание базовой линии
- Нормировка и сглаживание

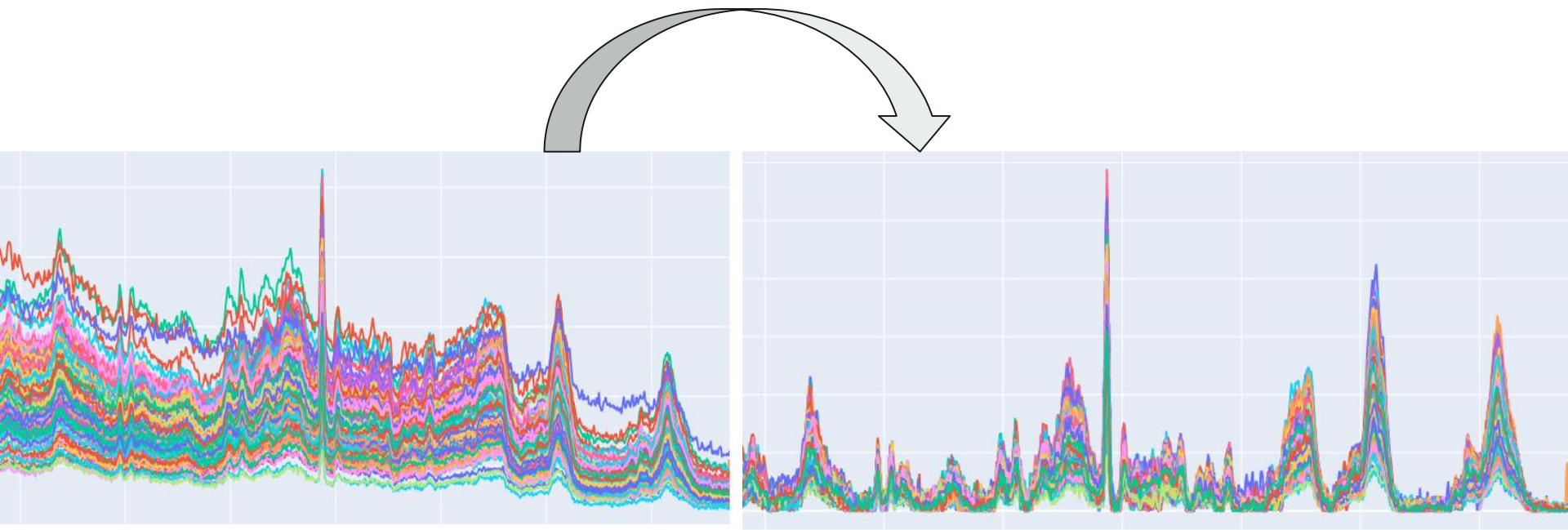
Удаление выбросов

SERS spectra for mixtures of HSA and GHSA



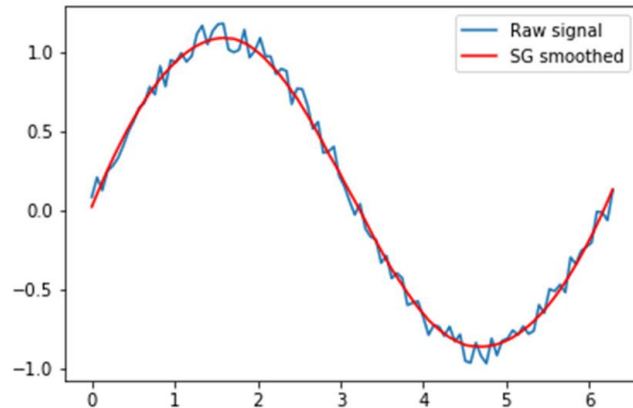
Вычитание базовой линии

Rubber band correction



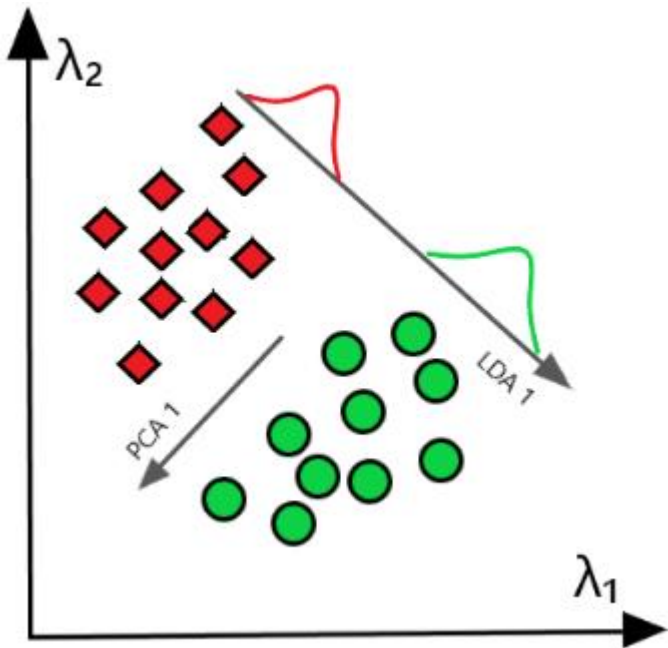
Нормировка и сглаживание

$$Z = \frac{x - \mu_s}{\sigma_s}$$



- Корректное учитьвание признаков
- Уменьшение шумового фона
- Повышение обобщающей способности модели

Модель линейного дискриминантного анализа



X – исходная матрица

c – концентрация GHSA

$$N = \sum_{i=0}^c N_i - \text{количество объектов}$$

Разброс внутри класса:

$$S_w = \sum_{i=0}^c S_i, (n \times n)$$

$$\text{где } S_i = \sum_{x \in D_i}^n (x - m_i)(x - m_i)^T, (n \times n)$$

$$m_i = \frac{1}{N_i} \sum_{x \in D_i}^n x_k$$

Разброс между классами:

$$S_B = \sum_{i=0}^c (m_i - m)(m_i - m)^T, (n \times n)$$


$$A = S_W^{-1} S_B$$

$$A \vec{v} = \lambda \vec{v}$$

$$W = [\vec{v}_{\max \lambda_1}, \vec{v}_{\max \lambda_2} \dots \vec{v}_{\max \lambda_k}], (n \times k)$$

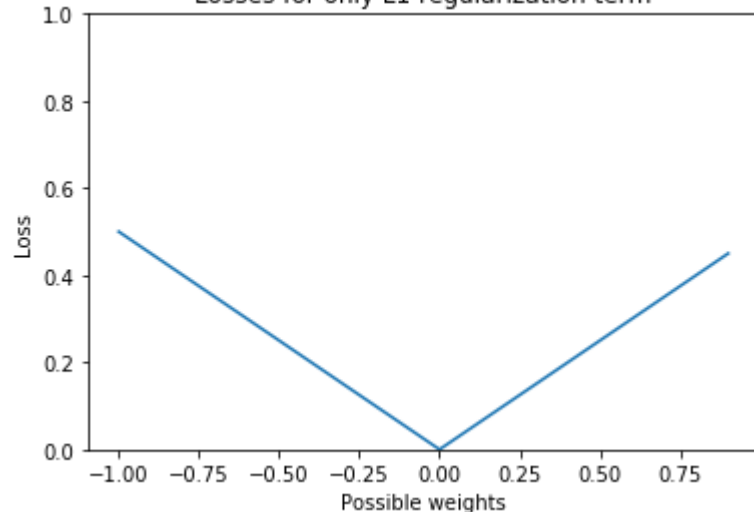
$$Y = X \times W, (N \times k)$$

Модель линейной регрессии + L1, L2 регуляризация

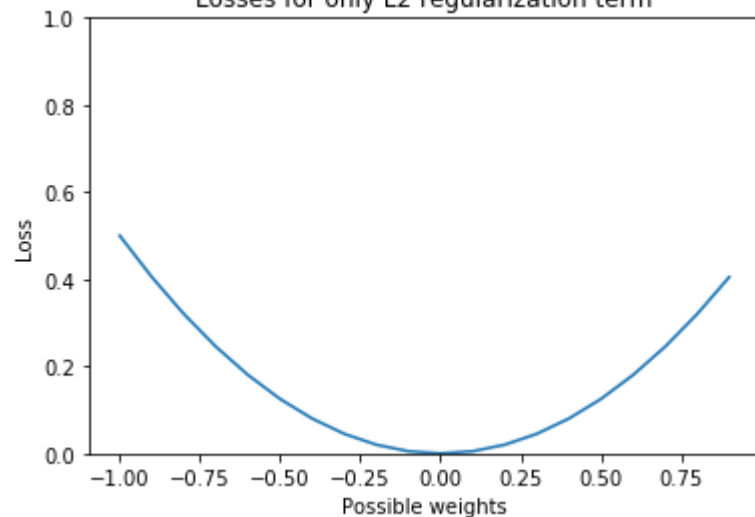

$$L_1 = (y - Xw)^2 + \underline{\alpha_1} |w|$$

$$L_2 = (y - Xw)^2 + \underline{\alpha_2} |w|^2$$

Losses for only L1 regularization term




Losses for only L2 regularization term



Результаты

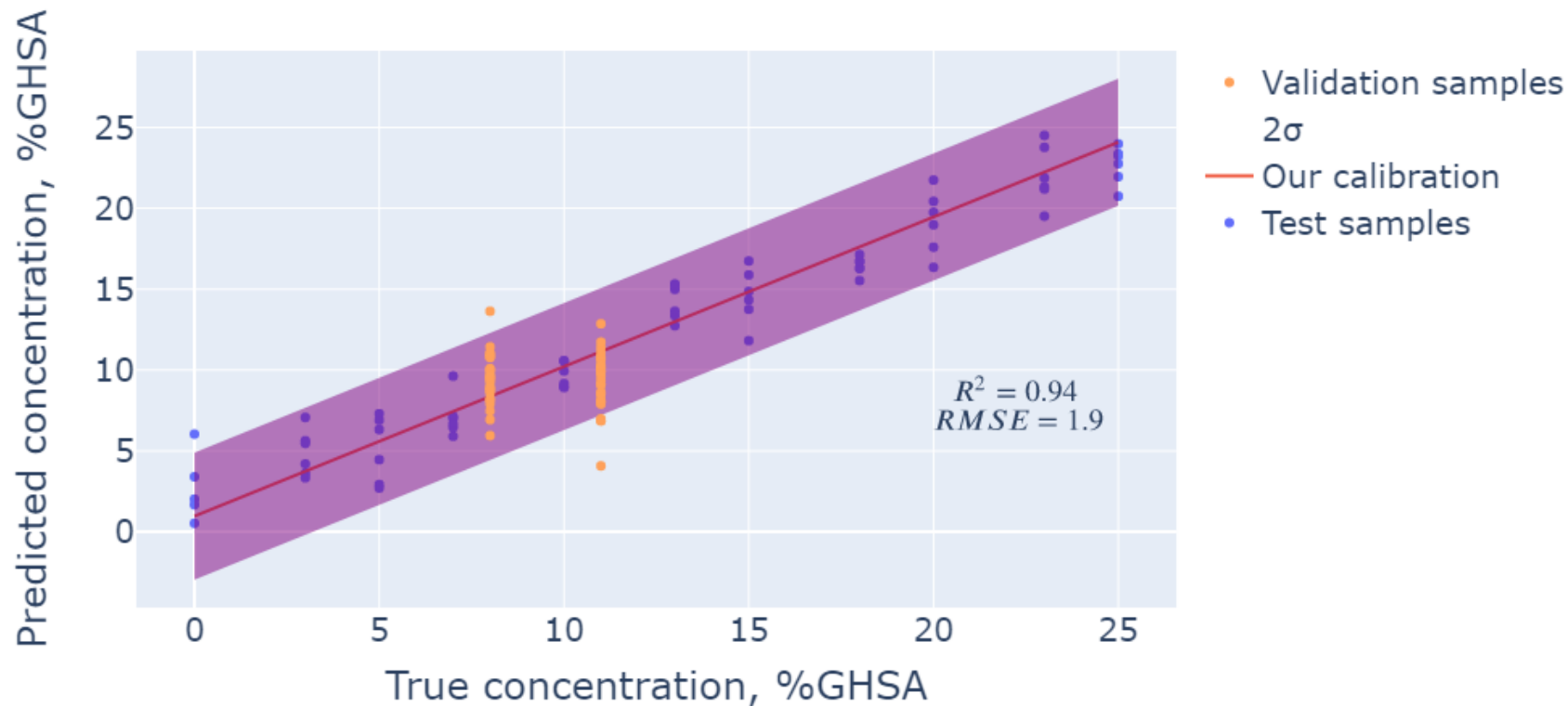
Линейный дискриминантный анализ



	precision	recall	f1-score	support
0%GHSA	1.00	1.00	1.00	5
3%GHSA	0.86	1.00	0.92	6
5%GHSA	1.00	0.50	0.67	6
7%GHSA	0.75	1.00	0.86	6
10%GHSA	0.86	1.00	0.92	6
13%GHSA	1.00	0.83	0.91	6
15%GHSA	1.00	1.00	1.00	6
18%GHSA	1.00	1.00	1.00	6
20%GHSA	1.00	1.00	1.00	6
23%GHSA	1.00	1.00	1.00	6
25%GHSA	1.00	1.00	1.00	6
accuracy			0.94	65
macro avg	0.95	0.94	0.93	65
weighted avg	0.95	0.94	0.93	65

Результаты

Calibration curve





- Показана возможность разделения спектров различных белков
- Разработана математическая модель для быстрого определения уровня гликированного альбумина в растворе на основе спектра ГКР



Вопросы