

ПРИМЕНЕНИЕ АНАЛИЗА ИЗОБРАЖЕНИЙ К РЕЗУЛЬТАТАМ ХРОМАТОГРАФИЧЕСКОГО РАЗДЕЛЕНИЯ МУЛЬТИКОМПОНЕНТНЫХ СМЕСЕЙ

Осипов А. В., Бойченко А. П., Дробот А. В.

Харьковский национальный университет имени В.Н. Каразина

artur_osipov@mail.ru

Развитие современных методов визуализации данных дает возможность представить большие объемы числовых данных в виде многомерных изображений. Сегодня практически все методы анализа обладают возможностью пространственной визуализации химической информации. Анализ изображений широко используется при обработке данных, полученных методами ИК-, Раман-, ЯМР-, ЭПР-, масс-спектропии и рентгеновских исследований [1].

Одним из методов химического анализа, в котором большую роль играет визуальная оценка полученных результатов, является тонкослойная хроматография (ТСХ). Однако, применение ТСХ в анализе сложных мультикомпонентных систем, например, природного происхождения, часто приводит к получению хроматограммы без четко разделенных пятен компонентов. Такую хроматограмму можно представить в виде многомерного массива данных и использовать для дальнейшего анализа. В настоящее время существуют лишь единичные работы по анализу хроматографических изображений. Поэтому создание подходов к анализу изображений в ТСХ для идентификации и классификации мультикомпонентных смесей представляет особый интерес.

В качестве основного инструмента для анализа многомерных изображений мы использовали метод главных компонент (principal component analysis, PCA). Он позволяет уменьшить размерность многомерных изображений, давая возможность убрать шум и выявить наиболее информативные компоненты многомерных данных. Основными этапами анализа RGB-изображений являются: разложение изображения на три канала, развертывание, предварительная обработка данных, проведение PCA анализа, свертывание и визуализация. При проведении PCA исходная матрица данных представляется в виде произведения $T \times P^T$, где T называется матрицей счетов, а P – матрицей нагрузок. Столбцы матрицы счетов представляют собой счета главных компонент, столбцы матрицы нагрузок – собственные вектора. Каждая колонка счетов и соответствующий собственный вектор отвечают определенному «фактору». Первый собственный вектор соответствует наибольшему собственному значению, из этого следует, что первый фактор несет максимальное количество информации, содержащейся в оригинальной матрице данных. Вторая колонка счетов и собственных векторов соответствует второму по важности фактору и т.д.

В данной работе метод PCA анализа изображений использовался для изучения результатов хроматографического разделения смесей аминокислот.

После проведения хроматографического эксперимента пластины проявлялись раствором нингидрина и сканировались на планшетном сканере HP ScanJet 4050G Photo. Полученные изображения обрабатывались с использованием программ Matlab 7.0 и Evince 2.2.

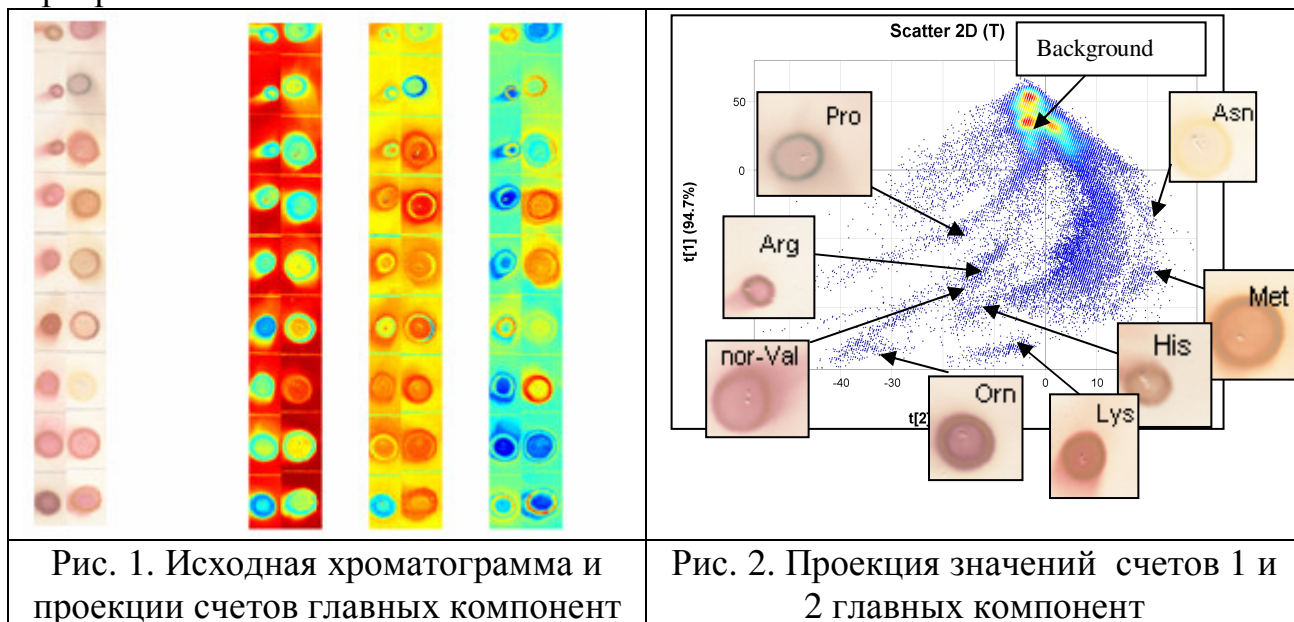


Рис. 1. Исходная хроматограмма и проекции счетов главных компонент

Рис. 2. Проекция значений счетов 1 и 2 главных компонент

На рисунке 1 представлена сканированная хроматограмма с пятнами восемнадцати аминокислот и проекции счетов главных компонент на плоскость хроматограммы. На проекциях четко видны границы пятен аминокислот, области размывания и фон. Анализ проекций значений счетов 1, 2 и 3 главных компонент (рис. 2) позволил выявить группу аминокислот, имеющих близкие значения счетов главных компонент и группу аминокислот с отличающимися значениями счетов: метионин, орнитин, нор-валин и аргинин, аспарагин, лизин, пролин. Метод анализа изображений успешно применили к результатам разделения смеси аминокислот, соответствующей составу гидролизата гормона окситоцина, а также смесей аминокислот, моделирующих присутствие нехарактерных для окситоцина аминокислот.

[1] Uwe Eichhoff. - От спектров к спектральным изображениям. – Методы и объекты химического анализа, 2010, т.5, №1.