

КОМПЛЕКСНЫЙ ПОДХОД ДЛЯ ИССЛЕДОВАНИЯ МОРФОЛОГИИ И РАСПРЕДЕЛЕНИЯ ПИГМЕНТОВ В КЛЕТКЕ ВОДОРОСЛИ

Е.Ю. Паршина¹

А.А. Самойленко²

Г.В. Максимов¹

А.И. Юсипович¹

Е.С. Лобакова¹

Я. Хе¹

Г.Г. Левин²

parshinae@gmail.com

asamoynenko@vniiofi.ru

gmaksimov@mail.ru

yusipovich@biophys.msu.ru

elena.lobakova@gmail.com

he.yanlin@qq.com

levin@vniiofi.ru

¹ МГУ имени М.В. Ломоносова, Москва, Российская Федерация

² ВНИИОФИ, Москва, Российская Федерация

Аннотация

С использованием атомно-силовой микроскопии, лазерной интерференционной микроскопии, микроспектроскопии комбинационного рассеяния и абсорбционной оптической томографии исследованы морфология и распределение структур клетки зеленой микроводоросли *Naematococcus lacustris*. Показано, что применение микроспектроскопии комбинационного рассеяния позволяет определить положение отдельных клеточных структур (хлоропласты), состав и конформацию молекул пигментов (каротиноиды), атомно-силовой микроскопии — латеральные и вертикальные размеры клетки, лазерной интерференционной микроскопии — оценить морфологию и/или показатель преломления вещества и, как следствие, локальную концентрацию вещества в клетке. Это дает возможность получить информацию о внутриклеточных структурах. Использование абсорбционной оптической томографии позволило получить трехмерные изображения полупрозрачных клеток микроводоросли *Naematococcus lacustris*, имеющих довольно сложную структуру, и визуализировать субклеточные структуры клетки водоросли несмотря на сильное поглощение света внутри клетки. Таким образом, комплексное использование атомно-силовой и лазерной интерференционной микроскопий, микроспектроскопии комбинационного рассеяния и аб-

Ключевые слова

Адсорбционная оптическая томография, атомно-силовая микроскопия, водоросли, *Naematococcus lacustris*, микроскопия комбинационного рассеяния, лазерная интерференционная микроскопия

сорбционной оптической томографии позволяет значительно расширить возможности исследования и получить детальную и комплексную информацию как о геометрии клетки и распределении субклеточных структур, так и об их молекулярном составе и конформации пигментов

Поступила 19.06.2023

Принята 05.07.2023

© Автор(ы), 2024

Работа выполнена при поддержке Междисциплинарной научно-образовательной школы МГУ имени М.В. Ломоносова «Молекулярные технологии живых систем и синтетическая биология». Спонсор не участвовал в разработке исследования, при сборе, анализе или интерпретации данных; в написании рукописи или в решении опубликовать результаты

Введение. В настоящее время для исследования морфологии клетки и динамики субклеточных структур широко используют различные методы оптической фазовой микроскопии, включая лазерную интерференционную микроскопию (ЛИМ) [1–6]. Однако внедрение таких методов сопровождается трудностями, связанными с проблемой разрешения и идентификации внутриклеточных структур клетки. В связи с этим особое внимание уделяется разработке методик оптической томографии в сочетании с атомно-силовой микроскопией (АСМ), ЛИМ и микроскопии комбинационного рассеяния (КР) (см., например, [7, 8]). Атомно-силовая микроскопия клеток позволяет исследовать их размеры и в некоторых случаях модуль упругости (жесткость) клеток [8–10], КР-микроскопию используют для оценки состава, конформации и распределения веществ в клетках [7, 11, 12], оптическая фазовая томография и ЛИМ позволяют оценивать размеры, форму, механические характеристики и распределение вещества внутри клетки [3, 5, 13, 14]. Сочетание этих методов позволит получить подробную и комплексную информацию о геометрии клетки и распределении субклеточных структур, их молекулярном составе и конформации пигментов.

Одним из методов, позволяющих исследовать трехмерную структуру клетки, является абсорбционная оптическая томография (АОТ). Оптический микротомограф позволяет получать томограммы живых полупрозрачных клеток в режиме поглощения света. Основное преимущество использования АОТ заключается в том, что большое число биологических объектов характеризуется наличием полупрозрачных неоднородностей внутри клеток, что справедливо и в отношении *Naemotococcus lacustris* (*N. lacustris*). Все это не позволяет зарегистрировать фазовую составляющую поля в широком диапазоне значений углов зондирования объекта. Поэтому для исследования морфологии внутренней структуры клетки с большим числом неоднородностей необходимо использовать

АОТ, позволяющую определить абсолютную величину и контролировать динамику распределения коэффициента поглощения внутри клетки на длине волны зондирующего излучения.

В настоящее время методы и алгоритмы традиционной оптической компьютерной томографии не распространяются на клетки более сложной формы, например клетки водоросли, имеющие многочисленные внутриклеточные структуры (ядро, хлоропласты, вакуоли и т. п.), которые поглощают свет. Современные методы позволяют только реконструировать изображение внутриклеточных структур в некотором приближении даже при неполных и неточных проекциях. Очевидно, что достоверность полученных изображений вызывает сомнение, а выводы о корреляции изображения и реальных структур клетки необходимо делать с большой осторожностью.

Ранее возможности применения различных методов оптической компьютерной томографии продемонстрированы при исследовании клеток водоросли спиросиры [15–17], водорослей *Pandorina morum* [18–20], а также на диатомовых водорослях (например, [21, 22]). Полученные авторами данные морфологии клетки (длина, ширина, размеры пор и т. д.) хорошо согласуются с результатами, определенными с использованием других методов (например, сканирующей электронной микроскопии) [22]. С помощью оптической компьютерной томографии водорослей *Skeletonema marinoi* и диатомовых водорослей *Thalassiosira rotula* выявлены реальные формы и размеры хлоропластов и их расположение в клетке. При наличии в окружающей среде антропогенных веществ в хлоропластах водоросли обнаруживается нарушение структуры органеллы [23], что может привести к ингибированию фотосинтеза и синтеза АТФ. Оптическая компьютерная томография была использована для исследования микроводорослей *Nannochloropsis oculata* (*N. oculata*) в целях определения содержания липидов в отдельной клетке [24]. Установлено, что при дефиците азота в *N. oculata* наблюдается накопление липидов и происходят изменения объема и массы сухих клеток, а также массы липидов отдельных клеток. Эти данные согласуются с представлением о том, что недостаток азота может компенсироваться разложением белков для поддержания метаболизма [25].

Цель работы — привести результаты исследования геометрических параметров клетки водоросли, локализации внутриклеточных структур и распределения отдельных молекул с помощью методов АСМ, ЛИМ и КР-микроспектроскопии, анализ морфологии клетки водоросли с использованием АОТ.

Материалы и методы. *Приготовление препаратов.* Изображения оптической ЛИМ, КР- и АСМ-изображения регистрировали на фиксирован-

ных препаратах зеленых водорослей *N. lacustris* [26]. Культуры микроводоросли выращивали на среде BG 11 [27] во флаконах стационарно при постоянном освещении 40 мкмоль фотонов ФАР/(м²·с) в течение двух недель. Для получения препаратов суспензию водорослей наносили на размеченное покровное стекло и высушивали на воздухе. Полученные препараты использовали для проведения КР- и АСМ-измерений. Для получения световых и ЛИМ-изображений микроскопии на покровное стекло, содержащее высушенные клетки, наносили 5 мкл раствора глицерин:вода (1:1 по объему) и через 10 мин помещали на зеркальную подложку (клетка-ми к зеркалу), затем проводили измерения.

Абсорбционная оптическая томография. Подробное описание экспериментальной установки приведено в [28]. В качестве источника низкокогерентного излучения использовали коллимированный линзой точечный светодиод модели KED080RAXH (*Kyosemi Co.*, Япония): центральная длина волны излучения 650 нм, полуширина спектра излучения 10 нм, мощность излучения 0,6 мВ при токе 25 мА), размер площадки излучения 80 мкм² (это позволило снизить шумы получаемых изображений). Фазовые изображения получали с использованием сдвигового интерферометра (рис. 1). Образец помещали между двумя покровными стеклами и освещали параллельным лучом света (под различными углами) через широкоапертурный объектив для получения проекций. Наклон луча изменяли поворотом зеркала с помощью гальвосканера. Диапазон сканирования в установке составил примерно ± 60°.

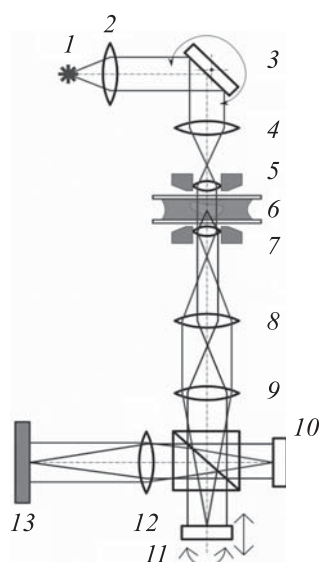


Рис. 1. Схема сдвигового интерференционного микротомографа:

- 1 — низкокогерентный точечный светодиод (LED);
- 2 — коллиматорная линза;
- 3 — управляемое поворотное полупрозрачное зеркало;
- 4 — плоскопараллельная линза;
- 5 — широкоапертурный объектив для освещения образца;
- 6 — образец;
- 7 — идентичный 5 широкоапертурный объектив для получения изображения;
- 8–12 — система линз;
- 10, 11 — зеркала сдвигового интерферометра, содержащего светоделительный кубик 50/50 между ними;
- 13 — светочувствительная матрица, используемая для формирования изображения

Изображение клетки получали с использованием алгоритма фазовых шагов, одно из зеркал смещалось с помощью пьезоэлемента. Изображение формировали на светочувствительной матрице, применяя системы линз. Восстановление фазовых изображений проводили методом интерферометрии фазовых шагов. В результате каждая проекция представляла собой изображение, аналогичное изображению, полученному в дифференциальном интерференционно-контрастном микроскопе, но с численными значениями, соответствующими производной фазе.

Для калибровки томографа использовали стеклянные сферы размером 5 мкм, что позволило определить размер вокселя в восстановленном изображении (0,1 мкм³). Время регистрации полного набора проекций 8...10 с. Для исследования клеток водоросли большого размера, диаметр которых составлял 80 мкм, томограф работал в режиме поглощения света.

Для того чтобы компенсировать влияние дифракции, использовали оригинальный алгоритм острой фокусировки. Для получения четкого изображения всей водоросли применяли процедуру многократного набора проекций при фокусировке на различные сечения внутри клетки. Сканирование вдоль оптической оси в процессе набора проекций осуществлялось механически. Для микроводорослей получено девять наборов проекций. При реконструкции томограмм для формирования 3D-изображения использовали только изображения сечений, полученные в области острой фокусировки. Описанная процедура позволила существенно ослабить влияние дифракционных эффектов и отфильтровать рассеянное излучение. Во многом этот алгоритм аналогичен получению наборов продольных томограмм в классическом томографе Бокажа (подробно теория формирования классических томограмм описана в [29]).

Для анализа АОТ-изображений микроводоросли *H. lacustris* использовали программу *3D Slicer*, которая позволила получить и исследовать произвольно ориентированные срезы клетки, строить модели поверхности из меток изображений и получить аппаратно-ускоренный объемный рендеринг (3D-изображение).

Оптическая и лазерная интерференционная микроскопия. Для сравнения с АОТ использовали лазерный автоматизированный интерференционный микроскоп МИА-1 с камерой модели UI-3140ср (IDS): черно-белая матрица с разрешением 1280 × 1024, частота кадров 50 fps, время экспозиции 10 мс, объектив 33×/0.65NA. Размер изображений объекта 200 × 150 мкм [30, 31]. Управление прибором и получение фазовых изображений по интерферограммам осуществляли с использованием про-

граммы *WinPhast*. При реконструкции одного фазового изображения применяли 10 интерференционных изображений, что позволяло регистрировать фазовые изображения с частотой 50 fps. ЛИМ-изображения и изображения в отраженном свете (оптическая микроскопия) получены с использованием лазера с длиной волны света 532 нм. Мощность лазеров составляла 20 мВт, мощность, приходящаяся на одну клетку, — менее 2 мВт. Лазерный луч направлен перпендикулярно плоскости клетки. Согласно описанию микроскопа, погрешность вдоль оптической оси составляла не более 1,9 нм, в плоскости XY — 0,5 мкм, повторяемость результатов измерений $s < 0,1$ нм.

Спектроскопия комбинационного рассеяния. Измерения спектров КР проводили на комбинированном конфокальном микроскопе-спектрометре «Интегра-Спектра» (НТ-МДТ, Зеленоград, Россия). Для регистрации спектров использовали длину волны возбуждающего света 532 нм, решетку 600 шт./мм, объектив 20×, время регистрации отдельного спектра для получения изображения 0,1 с. В каждой точке изображения записывали спектр КР в диапазоне значений 200...2400 см^{-1} . Наиболее интенсивные полосы в спектре принадлежат каротиноидам, для которых такая длина волны возбуждения является резонансной. Для построения изображений в каждой точке рассчитывали интеграл под кривой после вычитания базовой линии в диапазоне значений 1485...1565 см^{-1} . В этом диапазоне расположена полоса КР с максимумом 1530 см^{-1} , которая соответствует колебаниям двойных углеродных связей в молекуле каротиноида. Изображения регистрировали на площади 20 × 20 мкм (64 × 64 точки). Площадь под этой полосой и интенсивность в каждой точке изображения соответствуют содержанию каротиноидов в этой точке изображения [32].

Атомно-силовая микроскопия. Атомно-силовые изображения клеток получали на комбинированном конфокальном микроскопе-спектрометре Интегра-Спектра (НТ-МДТ, Зеленоград, Россия). Для получения изображений использовали полуконтактный режим и кантилеверы NSG10 с радиусом кривизны иглы 10 нм, частота сканирования 1 Гц, размер изображения 256 × 256 точек.

Проведение измерений. АСМ- и КР-изображения получали на одних и тех же клетках без перемещения образца последовательно, сначала АСМ-изображение, затем КР. Для получения светового и ЛИМ-изображений на тех же клетках суспензию водорослей, нанесенную на размеченное стекло, после проведения АСМ- и КР-измерений и добавления раствора глицерина помещали на зеркальную подложку, находили соот-

ветствующий квадрат стекла и группу клеток, затем регистрировали световые и ЛИМ-изображения.

Результаты и их обсуждение. *Идентификация распределения структур и молекул пигментов в клетке водоросли с использованием КР-микроспектроскопии, АСМ и ЛИМ.* В результате исследования получены изображения трех групп клеток (2, 4 и 4 клетки). Предложенный здесь алгоритм позволил получить изображения одних и тех же клеток с помощью метода световой микроскопии (в белом свете), АСМ, КР (представлена общая интенсивность (интеграл) спектра и площадь под полосой 1530 см^{-1} , пропорциональная содержанию каротиноидов), методом ЛИМ (при возбуждении лазером 532 нм), а также поглощения (при освещении лазерами длиной волны 532 и 650 нм).

В работе использовали клетки штамма микроводоросли *N. lacustris* NAMSU VM1, которая характеризуется сложным жизненным циклом. Одновременно в культуре присутствуют различные жизненные формы разной морфологии [26]: 1) вегетативные клетки размером 4...12 мкм, окрашенные в зеленый цвет; 2) формирующиеся гематоцисты, частично окрашенные в красный цвет; 3) зрелые гематоцисты — крупные клетки диаметром до 20 мкм и более, полностью окрашенные в красный цвет за счет накопления вторичного каротиноида астаксантина; 4) автоспорангии — структуры вегетативного размножения содержащие 4–16 автоспор (рис. 2, а). Иногда в культуре идентифицируются подвижные зооспоры, имеющие на переднем конце два жгутика, которые лишены клеточной стенки и покрыты снаружи желатиновой капсулой, соединяющейся с протопластом тяжами.

Внутриклеточная организации клеток существенно зависит от морфотипа клеток. Так, зрелые неподвижные вегетативные зеленые клетки микроводоросли имеют четко очерченную клеточную стенку, к которой примыкает цитоплазматическая мембрана. В пристенной области цитоплазмы клетки располагается крупный чашевидный хлоропласт (хроматофор), в котором выявляются несколько пиреноидов (представляют собой центры синтеза сахаров) и крахмальные зерна. В центре клетки располагаются ядро, немногочисленные липидные глобулы (жировые капли) и другие органеллы [33]. В хлоропласте локализован фотосинтетический аппарат микроводоросли, включая хлорофиллы *a* и *b* и первичные каротиноиды. Схематическое изображение вегетативной клетки *N. lacustris* приведено на рис. 2, б. В стрессовых условиях вегетативные клетки трансформируются в гематоцисты, которые имеют совершенно другое строение:

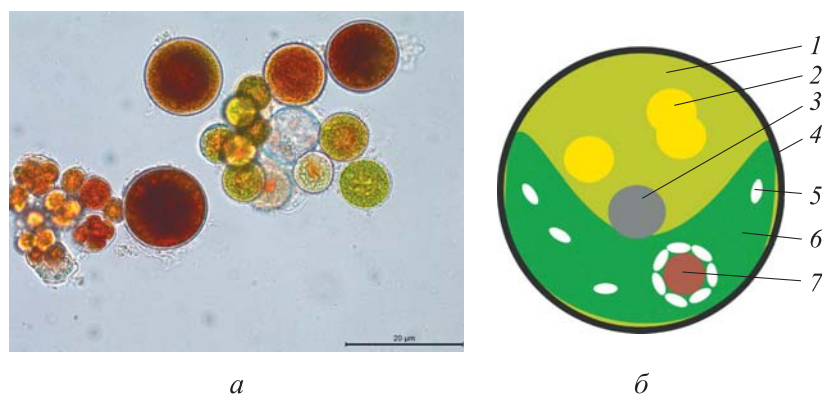


Рис. 2. Различные морфотипы клеток *N. lacustris* в культуре (а) и схематическое изображение вегетативной клетки (б):

- 1 — цитоплазма; 2 — липидные глобулы и клеточные органеллы; 3 — ядро;
4 — клеточная стенка и плазматическая мембрана; 5 — крахмальные зерна;
6 — хроматофор; 7 — пиреноид

клеточная стенка утолщается, клетка окрашивается в ярко-красный цвет за счет накопления в цитоплазме вторичного каротиноида астаксантина, размеры хлоропласта значительно уменьшаются, и он располагается в центре; по периферии клетки находятся крупные липидные глобулы, окрашенные в ярко-красный цвет. В целом протопласт вегетативной клетки представляет собой достаточно плотную структуру, основным компонентом которого является примыкающий к оболочке чашевидный хроматофор. Особенности внутреннего строения влияют на полученные разными методами изображения и спектры.

КР-микроспектроскопия клеток водоросли. При применении синих и зеленых лазеров КР-сигнал молекул каротиноидов значительно превосходит сигнал от других биологических молекул. Это обусловлено тем, что при таких длинах волн возбуждения регистрируются спектры резонансного КР, в результате чего сигнал может быть усилен в 1000 раз и более [34]. Поэтому наличие даже небольшого числа каротиноидов в составе объекта приводит к появлению типичных для каротиноидов полос в регистрируемых спектрах. В таком случае полосы каротиноидов являлись основными в регистрируемых спектрах (рис. 3). Поскольку первичные каротиноиды, в основном, могут находиться в хромофоре, участки с выраженными полосами, характерными для каротиноидов (~ 1000 , ~ 1160 и $\sim 1530 \text{ см}^{-1}$ [34]), являются маркерами расположения хромофора на изображении клетки. Согласно изображению на рис. 3, б, максимальное число каротиноидов распределено по периферии клетки, что позволяет судить

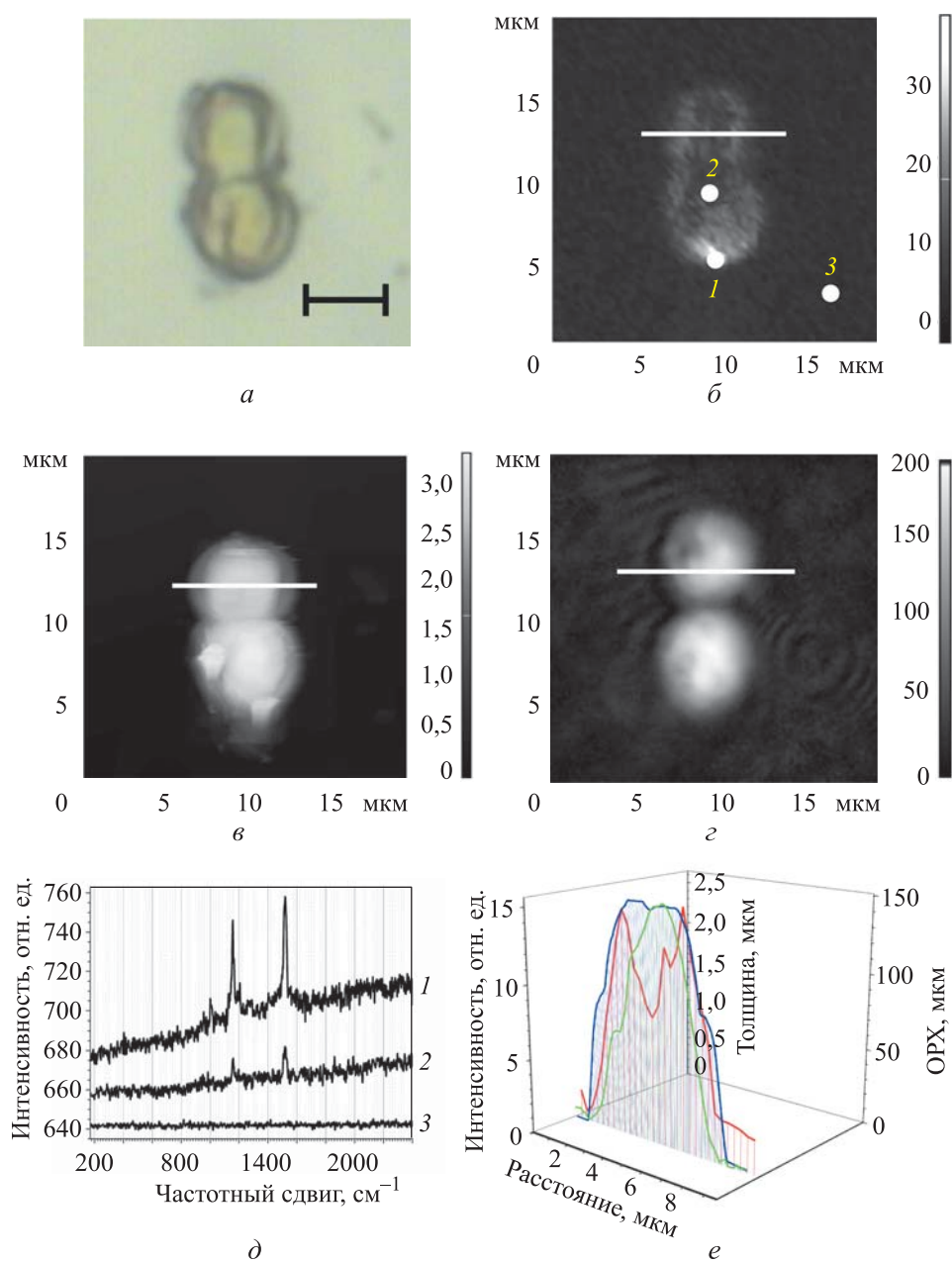


Рис. 3. Изображения клеток водоросли *N. lacustris*:

a — в проходящем свете, черный отрезок соответствует длине 5 мкм;
б — КР-изображение клеток водоросли, по оси *Z* отложена площадь пика с максимумом при 1520 см^{-1} (*1* — хроматофор; *2* — центр клетки; *3* — подложка);
в — АСМ-изображение; *г* — ЛИМ-изображение; *д* — КР-спектры с различных участков водоросли (точки *1–3* на рис. 3, *б*); *е* — сравнение профилей изображений, полученных различными методами (положение сечений показаны линиями белого цвета на рис. 3, *б*, *в*, *г*; КР-профиль (—), АСМ- (—) и ЛИМ-изображения (—))

о локализации хроматофора. На профиле КР-изображения (рис. 3, е, кривая красного цвета) в середине клетки число каротиноидов меньше, чем по краям, поскольку центральную часть клетки занимают другие органеллы.

ЛИМ и АСМ клеток водорослей. АСМ-изображение представлено на рис. 3, в. Установлено, что клетки имеют выпуклую шаровидную форму, в центре наблюдается возвышение, которое, возможно, соответствует положению ядра в центре клетки. Используя АСМ-изображения, также можно оценить геометрические размеры клетки.

ЛИМ-изображение тех же клеток показано на рис. 3, г. Оптическая разность хода (ОРХ) в точке, измеряемая ЛИМ, представляет собой величину, пропорциональную геометрической толщине и показателю преломления биологического образца, в свою очередь он линейно пропорционален концентрации вещества): $ОРХ_i = (k - k_0)z_i$, где k — показатель преломления объекта в точке; k_0 — показатель преломления среды (показатель преломления глицерин:вода в соотношении 1:1); z_i — толщина объекта в точке.

Значение ОРХ можно интерпретировать как количество вещества в рассматриваемой точке объекта. Из распределения ОРХ ЛИМ следует, что количество вещества распределено неравномерно и не совпадает с АСМ-изображением. При сравнении выбранных ЛИМ- и АСМ-профилей (см. рис. 3, е) можно сделать вывод, что в центре клетки содержатся компоненты с высоким показателем преломления (ядро и гранулы), в левой части (где ЛИМ- и АСМ-профили не совпадают) располагаются органеллы с низким показателем преломления, возможно, вакуоль. Область клетки с низкими значениями ОРХ совпадает с областью максимальных значений содержания каротиноидов (по КР-профилю). Возможно, в этой области хроматофор находится над/под вакуолью. Можно также предположить, что молекулы каротиноидов локализованы в области с меньшей плотностью цитоплазмы (вакуоль) клетки водоросли, например, в виде липидных глобул или включений.

Следовательно, представленные на изображениях клетки возможно представляют собой вегетативную форму *H. lacustris*. Сопоставление сечений рельефа поверхности клетки (АСМ), распределения ОРХ и локализации молекул каротиноидов (КР) указывает на сложную трехмерную организацию клеток водоросли *H. lacustris*, которую можно исследовать выбранными методами.

Особенности АОТ клетки водоросли. В процессе исследования реализована возможность анализа пространственной структуры клеток с помощью АОТ на примере микроводоросли *H. lacustris* (80...100 мкм). Оче-

видно, что субклеточная структура этих клеток сложная и полученные методами обычной микроскопии изображения не позволяют получать детальное 3D-распределение вещества цитоплазмы. Отметим, что применение оптической фазовой томографии также не дало приемлемых результатов, так как сильное поглощение не позволило восстановить фазу прошедшего излучения в клетке. Именно использование АОТ дало возможность решить эту задачу, однако для компенсации дифракционных эффектов, которые возникают вследствие мелкой структуры внутри клетки, необходимо разработать специальный алгоритм острой фокусировки. Отметим, что такой подход перспективен для исследования динамики субклеточных структур в режиме метода локальной томографии. Результаты теоретических исследований и математического моделирования локальных алгоритмов томографии представлены в [28].

С использованием разработанных методов получены томограммы коэффициента поглощения микроводоросли *N. lacustris*, позволяющие исследовать ее внутреннюю структуру. Параллельные, последовательные сечения клетки (перпендикулярно оптической оси прибора) приведены на рис. 4. Толщина клетки составляет около 35 мкм. Другие использованные методы не позволяют полностью оценивать объекты такой толщины.

На полученных сечениях четко видна клеточная стенка. В среднем ряду (центральная часть клетки) расположена большая структура неправильной формы, примыкающая к клеточной стенке, которая, по-видимому, представляет собой хроматофор. Внутри структуры находятся шарообразные участки с меньшим коэффициентом поглощения (крайне правое сечение в центральном ряду и нижний ряд), которые, скорее всего, являются пиреноидами. Области клетки, не занятые хроматофором (нижний ряд), заполнены мелкими объектами, которые представляют собой различные клеточные органеллы (липидные глобулы). На основании изложенного можно предположить, что на рис. 4 приведены сечения микроцисты *N. lacustris*.

Заключение. Полученные результаты свидетельствуют о возможности исследования морфологии клеток водорослей методами АСМ, ЛИМ и КР-микроспектроскопии, а также АОТ. Метод АСМ позволяет получать информацию о поверхности клетки и ее геометрических размерах, ЛИМ — о показателе преломления и количестве вещества в определенной точке клетки, КР-микроспектроскопия — изучать состав клетки в каждой точке, судить о конформации основных молекул. В случае водорослей КР-изображения дают информацию о распределении и структуре молекул каротиноидов.

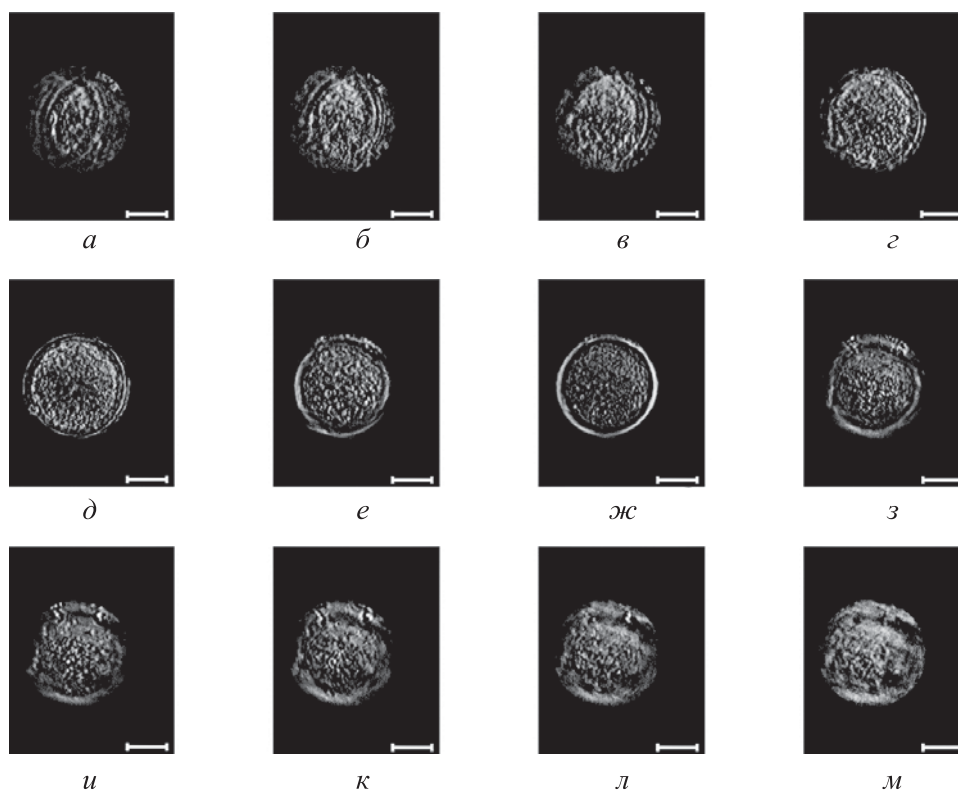


Рис. 4. Сечения параллельных предметному стеклу изображений клетки водоросли *H. lacustris* на расстоянии от поверхности клетки $z = 1$ (а), 4 (б), 7 (в), 10 (г), 13 (д), 16 (е), 19 (ж), 22 (з), 25 (и), 28 (к), 31 (л), 34 мкм (м); длина белой линии на сечении 10 мкм

Проведенные экспериментальные исследования свидетельствуют о том, что АОТ сильно поглощающих объектов позволяет хорошо визуализировать морфологию внутренней структуры клеток. Использование алгоритмов острой фокусировки при наборе проекционных данных дает возможность существенно ослабить влияние эффектов рассеяния излучения на внутренних органеллах клетки, несмотря на некоторое увеличение времени сбора данных. Однако хорошие морфометрические данные не позволяют определять состав и количество различных веществ внутри клетки водоросли. Предлагаемый комплексный подход к измерению внутренней структуры клетки за счет применения различных методов микроскопии существенно расширяет возможности исследования.

ЛИТЕРАТУРА

- [1] Юсипович А.И., Паршина Е.Ю., Байжуманов А.А. и др. Использование лазерного интерференционного микроскопа для оценки флуктуаций и эквивалентной константы упругости мембран клеток. *Приборы и техника эксперимента*, 2021, № 6, с. 93–101. DOI: <https://doi.org/10.31857/S0032816221060136>
- [2] Левин Г.Г. Современные методы оптической томографии и голографии. *Измерительная техника*, 2005, № 11, с. 43–47.
- [3] Yusipovich A.I., Parshina E.Yu., Brysgalova N.Yu., et al. Laser interference microscopy in erythrocyte study. *J. Appl. Phys.*, 2009, vol. 105, iss. 10, art. 102037. DOI: <https://doi.org/10.1063/1.3116609>
- [4] Юсипович А.И., Берестовская Ю.Ю., Шутова В.В. и др. Новые возможности исследования микробиологических объектов методом лазерной интерференционной микроскопии. *Метрология*, 2012, № 3, с. 17–26.
- [5] Yusipovich A.I., Zagubizhenko M.V., Levin G.G., et al. Laser interference microscopy of amphibian erythrocytes: impact of cell volume and refractive index. *J. Microsc.*, 2011, vol. 244, iss. 3, pp. 223–229. DOI: <https://doi.org/10.1111/j.1365-2818.2011.03516.x>
- [6] Yusipovich A.I., Cherkashin A.A., Verdiyev E.E., et al. Laser interference microscopy: a novel approach to the visualization of structural changes in myelin during the propagation of nerve impulses. *Laser Phys. Lett.*, 2016, vol. 13, no. 8, art. 085601. DOI: <https://doi.org/10.1088/1612-2011/13/8/085601>
- [7] Parshina E.Yu., Sarycheva A.S., Yusipovich A.I., et al. Combined Raman and atomic force microscopy study of hemoglobin distribution inside erythrocytes and nanoparticle localization on the erythrocyte surface. *Laser Phys. Lett.*, 2013, vol. 10, no. 7, art. 075607. DOI: <https://doi.org/10.1088/1612-2011/10/7/075607>
- [8] Parshina E.Yu., Yusipovich A.I., Brazhe A.R., et al. Heat damage of cytoskeleton in erythrocytes increases membrane roughness and cell rigidity. *J. Biol. Phys.*, 2019, vol. 45, no. 4, pp. 367–377. DOI: <https://doi.org/10.1007/s10867-019-09533-5>
- [9] Журина М.В., Кострикина Н.А., Паршина Е.Ю. и др. Визуализация внеклеточного полимерного матрикса биопленок *Chromobacterium violaceum* с помощью микроскопических методов. *Микробиология*, 2013, т. 82, № 4, с. 502–509. EDN: QCXYDV. DOI: <https://doi.org/10.7868/S0026365613040162>
- [10] Parshina E.Yu., Yusipovich A.I., Platonova A.A., et al. Thermal inactivation of volume-sensitive K^+ , Cl^- cotransport and plasma membrane relief changes in human erythrocytes. *Pflugers Arch. — Eur. J. Physiol.*, 2013, vol. 465, no. 7, pp. 977–983. DOI: <https://doi.org/10.1007/s00424-013-1221-4>
- [11] Mojzeš P., Gao L., Ismagulova T., et al. Guanine, a high-capacity and rapid-turnover nitrogen reserve in microalgal cells. *PNAS*, 2020, vol. 117, no. 51, pp. 32722–32730. DOI: <https://doi.org/10.1073/pnas.2005460117>
- [12] Solovchenko A., Khozin-Goldberg I., Selyakh I., et al. Phosphorus starvation and luxury uptake in green microalgae revisited. *Algal Res.*, 2019, vol. 43, art. 101651. DOI: <https://doi.org/10.1016/j.algal.2019.101651>

- [13] Юсипович А.И., Новиков С.М., Казакова Т.А. и др. Особенности исследования изолированного нейрона методом лазерной интерференционной микроскопии. *Квантовая электроника*, 2006, т. 36, № 9, с. 874–878.
- [14] Fedorov D.A., Sidorenko S.V., Yusipovich A.I., et al. $\text{Na}_i^+/\text{K}_i^+$ imbalance contributes to gene expression in endothelial cells exposed to elevated NaCl. *Heliyon*, 2021, vol. 7, iss. 9, art. e08088. DOI: <https://doi.org/10.1016/j.heliyon.2021.e08088>
- [15] Wu Z., Sun Y., Matlock A., et al. SIMBA: scalable inversion in optical tomography using deep denoising priors. *IEEE J. Sel. Top. Signal Process.*, 2020, vol. 14, iss. 6, pp. 1163–1175. DOI: <https://doi.org/10.1109/JSTSP.2020.2999820>
- [16] Matlock A., Zhu J., Tian L. Multiple-scattering simulator-trained neural network for intensity diffraction tomography. *Opt. Express*, 2023, vol. 31, iss. 3, pp. 4094–4107. DOI: <https://doi.org/10.1364/oe.477396>
- [17] Li J., Matlock A., Li Y., et al. High-speed in vitro intensity diffraction tomography. *Adv. Photonics*, 2019, vol. 1, iss. 6, art. 066004. DOI: <https://doi.org/10.1117/1.AP.1.6.066004>
- [18] Li J., Sun J., Zhang J., et al. Three-dimensional optical diffraction tomographic microscopy with optimal frequency combination with partially coherent illuminations. *arXiv:1803.01151*. DOI: <https://doi.org/10.48550/arXiv.1803.01151>
- [19] Zuo C., Sun J., Li J., et al. Wide-field high-resolution 3D microscopy with Fourier ptychographic diffraction tomography. *Opt. Lasers Eng.*, 2020, vol. 128, art. 106003. DOI: <https://doi.org/10.1016/j.optlaseng.2020.106003>
- [20] Li J., Chen Q., Zhang J., et al. Optical diffraction tomography microscopy with transport of intensity equation using a light-emitting diode array. *Opt. Lasers Eng.*, 2017, vol. 95, pp. 26–34. DOI: <https://doi.org/10.1016/j.optlaseng.2017.03.010>
- [21] Soto J.M., Rodrigo J.A., Alieva T. Label-free quantitative 3D tomographic imaging for partially coherent light microscopy. *Opt. Express*, 2017, vol. 25, iss. 14, pp. 15699–15712. DOI: <https://doi.org/10.1364/oe.25.015699>
- [22] Hamano R., Mayama S., Umemura K. Localization analysis of intercellular materials of living diatom cells studied by tomographic phase microscopy. *Appl. Phys. Lett.*, 2022, vol. 120, iss. 13, art. 133701. DOI: <https://doi.org/10.1063/5.0086165>
- [23] Merola F., Memmolo P., Miccio L., et al. Phase contrast tomography at lab on chip scale by digital holography. *Methods*, 2018, vol. 136, pp. 108–115. DOI: <https://doi.org/10.1016/j.ymeth.2018.01.003>
- [24] Jung J., Hong S.-J., Kim H.-B., et al. Label-free non-invasive quantitative measurement of lipid contents in individual microalgal cells using refractive index tomography. *Sci. Rep.*, 2018, vol. 8, art. 6524. DOI: <https://doi.org/10.1038/s41598-018-24393-0>
- [25] Shin H., Hong S.-J., Kim H., et al. Elucidation of the growth delimitation of *Dunaliella tertiolecta* under nitrogen stress by integrating transcriptome and peptidome analysis. *Bioresour. Technol.*, 2015, vol. 194, pp. 57–66. DOI: <https://doi.org/10.1016/j.biortech.2015.07.002>

[26] Kublanovskaya A., Baulina O., Chekanov K., et al. The microalga *Haematococcus lacustris* (Chlorophyceae) forms natural biofilms in supralittoral White Sea coastal rock ponds. *Planta*, 2020, vol. 252, no. 3, art. 37.

DOI: <https://doi.org/10.1007/s00425-020-03438-7>

[27] Stanier R.Y., Kunisawa R., Mandel M., et al. Purification and properties of unicellular blue-green algae (order Chroococcales). *Bacteriol. Rev.*, 1971, vol. 35, no. 2, pp. 171–205.

DOI: <https://doi.org/10.1128/br.35.2.171-205.1971>

[28] Vishnyakov G.N., Levin G.G., Minaev V.L., et al. Investigation of the method of local optical tomography by differential projections. *Opt. Spectrosc.*, 2018, vol. 125, no. 6, pp. 1065–1073. DOI: <https://doi.org/10.1134/S0030400X18120226>

[29] Левин Г.Г., Вишняков Г.Н. Оптическая томография. М., Радио и связь, 1989.

[30] Минаев В.Л., Юсипович А.И. Использование автоматизированного интерференционного микроскопа в биологических исследованиях. *Измерительная техника*, 2012, № 7, с. 66–69.

[31] Левин Г.Г., Вишняков Г.Н., Минаев В.Л. Автоматизированный интерференционный микроскоп для измерения динамических объектов. *Приборы и техника эксперимента*, 2014, № 1, с. 79–84. EDN: RTOVOT.

DOI: <https://doi.org/10.7868/S0032816214010066>

[32] Darvin M.E., Gersonde I., Meinke M., et al. Non-invasive *in vivo* determination of the carotenoids beta-carotene and lycopene concentrations in the human skin using the Raman spectroscopic method. *J. Phys. D: Appl. Phys.*, 2005, vol. 38, no. 15, art. 2696.

DOI: <https://doi.org/10.1088/0022-3727/38/15/023>

[33] Baudalet P.-H., Ricochon G., Linder M., et al. A new insight into cell walls of Chlorophyta. *Algal Res.*, 2017, vol. 25, pp. 333–371.

DOI: <https://doi.org/10.1016/j.algal.2017.04.008>

[34] Pirutin S.K., Jia S., Yusipovich A.I., et al. Vibrational spectroscopy as a tool for bioanalytical and biomonitoring studies. *Int. J. Mol. Sci.*, 2023, vol. 24, iss. 8, art. 6947.

DOI: <https://doi.org/10.3390/ijms24086947>

Паршина Евгения Юрьевна — канд. биол. наук, старший научный сотрудник кафедры биофизики Биологического факультета МГУ имени М.В. Ломоносова (Российская Федерация, 119991, Москва, Ленинские горы, д. 1).

Самойленко Алексей Алексеевич — канд. физ.-мат. наук, начальник сектора НИО М-44 ВНИИОФИ (Российская Федерация, 119361, Москва, ул. Озерная, д. 46).

Максимов Георгий Владимирович — д-р биол. наук, профессор кафедры биофизики Биологического факультета МГУ имени М.В. Ломоносова (Российская Федерация, 119991, Москва, Ленинские горы, д. 1).

Юсипович Александр Иванович — канд. биол. наук, старший научный сотрудник кафедры биофизики Биологического факультета МГУ имени М.В. Ломоносова (Российская Федерация, 119991, Москва, Ленинские горы, д. 1).

Лобакова Елена Сергеевна — д-р биол. наук, профессор кафедры биоинженерии Биологического факультета МГУ имени М.В. Ломоносова (Российская Федерация, 119991, Москва, Ленинские горы, д. 1).

Хе Яньдинь — аспирант кафедры биофизики Биологического факультета МГУ имени М.В. Ломоносова (Российская Федерация, 119991, Москва, Ленинские горы, д. 1).

Левин Геннадий Генрихович — д-р техн. наук, профессор, начальник отделения голографии, оптической томографии, нанотехнологии и наноматериалов ВНИИОФИ (Российская Федерация, 119361, Москва, ул. Озерная, д. 46).

Просьба ссылаться на эту статью следующим образом:

Паршина Е.Ю., Самойленко А.А., Максимов Г.В. и др. Комплексный подход для исследования морфологии и распределения пигментов в клетке водоросли. *Вестник МГТУ им. Н.Э. Баумана. Сер. Естественные науки*, 2024, № 2 (113), с. 129–148. EDN: NBOFMH

INTEGRATED APPROACH TO STUDYING PIGMENT MORPHOLOGY AND DISTRIBUTION IN THE ALGAE CELL

E.Yu. Parshina¹

A.A. Samoylenko²

G.V. Maksimov¹

A.I. Yusipovich¹

E.S. Lobakova¹

Y. He¹

G.G. Levin²

parshinae@gmail.com

asamoylenko@vniiofi.ru

gmaksimov@mail.ru

yusipovich@biophys.msu.ru

elena.lobakova@gmail.com

he.yanlin@qq.com

levin@vniiofi.ru

¹Lomonosov Moscow State University, Moscow, Russian Federation

²All-Russian Research Institute for Optical and Physical Measurements, Moscow, Russian Federation

Abstract

Atomic force microscopy, laser interference microscopy, Raman microspectroscopy and optical absorption tomography are used to study morphology and distribution of the green microalga *Haematococcus lacustris* cell structure. The paper shows that using Raman microspectroscopy makes it possible to identify position of separate cellular structures (chloroplasts), composition and conformation of the pigment molecules (carotenoids), atomic force microscopy, including the cell lateral and vertical size and laser interference microscopy, to determine the substance refraction indicator

Keywords

Adsorption optical tomography, atomic force microscopy, algae, Haematococcus lacustris, Raman microscopy, laser interference microscopy

and consequently the substance local concentration in a cell. This allows obtaining information about intracellular structures. Using the absorption optical tomography makes it possible to receive three-dimensional images of microalga *Haemotococcus lacustris* translucent cells, which have a rather complex structure, as well as to visualize the algae cell subcellular structures despite the light strong absorption inside the cell. Thus, integrated use of atomic force and laser interference microscopy, Raman microspectroscopy and absorption optical tomography allows significantly expanding research capabilities and obtaining detailed and comprehensive information about both the cell geometry and the subcellular structures distribution, as well as about their molecular composition and pigment conformation

Received 19.06.2023

Accepted 05.07.2023

© Author(s), 2024

The work was carried out with the support of the Interdisciplinary Scientific and Educational School of Lomonosov Moscow State University "Molecular Technologies of Living Systems and Synthetic Biology". The sponsor was not involved in the development of the study, in the collection, analysis or interpretation of data; in the writing of the manuscript or in the decision to publish the results

REFERENCES

- [1] Yusipovich A.I., Parshina E.Yu., Bayzhumanov A.A., et al. Use of a laser interference microscope for estimating fluctuations and the equivalent elastic constant of cell membranes. *Instrum. Exp. Tech.*, 2021, vol. 64, no. 6, pp. 877–885.
DOI: <https://doi.org/10.1134/S0020441221060129>
- [2] Levin G.G. Contemporary methods of optical tomography and holography. *Meas. Tech.*, 2005, vol. 48, no. 11, pp. 1103–1108.
DOI: <https://doi.org/10.1007/s11018-006-0028-5>
- [3] Yusipovich A.I., Parshina E.Yu., Brysgalova N.Yu., et al. Laser interference microscopy in erythrocyte study. *J. Appl. Phys.*, 2009, vol. 105, iss. 10, art. 102037.
DOI: <https://doi.org/10.1063/1.3116609>
- [4] Yusipovich A.I., Berestovskaya Yu.Yu., Shutova V.V., et al. New possibilities for the study of microbiological objects by laser interference microscopy. *Meas. Tech.*, 2012, vol. 55, no. 3, pp. 351–356. DOI: <https://doi.org/10.1007/s11018-012-9963-5>
- [5] Yusipovich A.I., Zagubizhenko M.V., Levin G.G., et al. Laser interference microscopy of amphibian erythrocytes: impact of cell volume and refractive index. *J. Microsc.*, 2011, vol. 244, iss. 3, pp. 223–229. DOI: <https://doi.org/10.1111/j.1365-2818.2011.03516.x>
- [6] Yusipovich A.I., Cherkashin A.A., Verdiyev E.E., et al. Laser interference microscopy: a novel approach to the visualization of structural changes in myelin during the propagation of nerve impulses. *Laser Phys. Lett.*, 2016, vol. 13, no. 8, art. 085601.
DOI: <https://doi.org/10.1088/1612-2011/13/8/085601>

- [7] Parshina E.Yu., Sarycheva A.S., Yusipovich A.I., et al. Combined Raman and atomic force microscopy study of hemoglobin distribution inside erythrocytes and nanoparticle localization on the erythrocyte surface. *Laser Phys. Lett.*, 2013, vol. 10, no. 7, art. 075607. DOI: <https://doi.org/10.1088/1612-2011/10/7/075607>
- [8] Parshina E.Yu., Yusipovich A.I., Brazhe A.R., et al. Heat damage of cytoskeleton in erythrocytes increases membrane roughness and cell rigidity. *J. Biol. Phys.*, 2019, vol. 45, no. 4, pp. 367–377. DOI: <https://doi.org/10.1007/s10867-019-09533-5>
- [9] Zhurina M.V., Kostrikin N.A., Parshina E.Yu., et al. Visualization of the extracellular polymeric matrix of *Chromobacterium violaceum* biofilms by microscopic methods. *Microbiology*, 2013, vol. 82, no. 4, pp. 517–524. DOI: <https://doi.org/10.1134/S0026261713040164>
- [10] Parshina E.Yu., Yusipovich A.I., Platonova A.A., et al. Thermal inactivation of volume-sensitive K^+ , Cl^- cotransport and plasma membrane relief changes in human erythrocytes. *Pflugers Arch. — Eur. J. Physiol.*, 2013, vol. 465, no. 7, pp. 977–983. DOI: <https://doi.org/10.1007/s00424-013-1221-4>
- [11] Mojzeš P., Gao L., Ismagulova T., et al. Guanine, a high-capacity and rapid-turnover nitrogen reserve in microalgal cells. *PNAS*, 2020, vol. 117, no. 51, pp. 32722–32730. DOI: <https://doi.org/10.1073/pnas.2005460117>
- [12] Solovchenko A., Khozin-Goldberg I., Selyakh I., et al. Phosphorus starvation and luxury uptake in green microalgae revisited. *Algal Res.*, 2019, vol. 43, art. 101651. DOI: <https://doi.org/10.1016/j.algal.2019.101651>
- [13] Yusipovich A.I., Novikov S.M., Kazakova T.A., et al. Peculiarities of studying an isolated neuron by the method of laser interference microscopy. *Quantum Electron.*, 2006, vol. 36, no. 9, pp. 874–878. DOI: <https://doi.org/10.1070/QE2006v036n09ABEH013408>
- [14] Fedorov D.A., Sidorenko S.V., Yusipovich A.I., et al. Na^+ / K^+ imbalance contributes to gene expression in endothelial cells exposed to elevated NaCl. *Helvion*, 2021, vol. 7, iss. 9, art. e08088. DOI: <https://doi.org/10.1016/j.helivion.2021.e08088>
- [15] Wu Z., Sun Y., Matlock A., et al. SIMBA: scalable inversion in optical tomography using deep denoising priors. *IEEE J. Sel. Top. Signal Process.*, 2020, vol. 14, iss. 6, pp. 1163–1175. DOI: <https://doi.org/10.1109/JSTSP.2020.2999820>
- [16] Matlock A., Zhu J., Tian L. Multiple-scattering simulator-trained neural network for intensity diffraction tomography. *Opt. Express*, 2023, vol. 31, iss. 3, pp. 4094–4107. DOI: <https://doi.org/10.1364/oe.477396>
- [17] Li J., Matlock A., Li Y., et al. High-speed in vitro intensity diffraction tomography. *Adv. Photonics*, 2019, vol. 1, iss. 6, art. 066004. DOI: <https://doi.org/10.1117/1.AP.1.6.066004>
- [18] Li J., Sun J., Zhang J., et al. Three-dimensional optical diffraction tomographic microscopy with optimal frequency combination with partially coherent illuminations. *arXiv:1803.01151*. DOI: <https://doi.org/10.48550/arXiv.1803.01151>
- [19] Zuo C., Sun J., Li J., et al. Wide-field high-resolution 3D microscopy with Fourier ptychographic diffraction tomography. *Opt. Lasers Eng.*, 2020, vol. 128, art. 106003. DOI: <https://doi.org/10.1016/j.optlaseng.2020.106003>

- [20] Li J., Chen Q., Zhang J., et al. Optical diffraction tomography microscopy with transport of intensity equation using a light-emitting diode array. *Opt. Lasers Eng.*, 2017, vol. 95, pp. 26–34. DOI: <https://doi.org/10.1016/j.optlaseng.2017.03.010>
- [21] Soto J.M., Rodrigo J.A., Alieva T. Label-free quantitative 3D tomographic imaging for partially coherent light microscopy. *Opt. Express*, 2017, vol. 25, iss. 14, pp. 15699–15712. DOI: <https://doi.org/10.1364/oe.25.015699>
- [22] Hamano R., Mayama S., Umemura K. Localization analysis of intercellular materials of living diatom cells studied by tomographic phase microscopy. *Appl. Phys. Lett.*, 2022, vol. 120, iss. 13, art. 133701. DOI: <https://doi.org/10.1063/5.0086165>
- [23] Merola F., Memmolo P., Miccio L., et al. Phase contrast tomography at lab on chip scale by digital holography. *Methods*, 2018, vol. 136, pp. 108–115. DOI: <https://doi.org/10.1016/j.ymeth.2018.01.003>
- [24] Jung J., Hong S.-J., Kim H.-B., et al. Label-free non-invasive quantitative measurement of lipid contents in individual microalgal cells using refractive index tomography. *Sci. Rep.*, 2018, vol. 8, art. 6524. DOI: <https://doi.org/10.1038/s41598-018-24393-0>
- [25] Shin H., Hong S.-J., Kim H., et al. Elucidation of the growth delimitation of *Dunaliella tertiolecta* under nitrogen stress by integrating transcriptome and peptidome analysis. *Bioresour. Technol.*, 2015, vol. 194, pp. 57–66. DOI: <https://doi.org/10.1016/j.biortech.2015.07.002>
- [26] Kublanovskaya A., Baulina O., Chekanov K., et al. The microalga *Haematococcus lacustris* (Chlorophyceae) forms natural biofilms in supralittoral White Sea coastal rock ponds. *Planta*, 2020, vol. 252, no. 3, art. 37. DOI: <https://doi.org/10.1007/s00425-020-03438-7>
- [27] Stanier R.Y., Kunisawa R., Mandel M., et al. Purification and properties of unicellular blue-green algae (order Chroococcales). *Bacteriol. Rev.*, 1971, vol. 35, no. 2, pp. 171–205. DOI: <https://doi.org/10.1128/br.35.2.171-205.1971>
- [28] Vishnyakov G.N., Levin G.G., Minaev V.L., et al. Investigation of the method of local optical tomography by differential projections. *Opt. Spectrosc.*, 2018, vol. 125, no. 6, pp. 1065–1073. DOI: <https://doi.org/10.1134/S0030400X18120226>
- [29] Levin G.G., Vishnyakov G.N. *Opticheskaya tomografiya [Optical tomography]*. Moscow, Radio i svyaz Publ., 1989.
- [30] Minaev V.L., Yusipovich A.I. Use of an automated interference microscope in biological research. *Meas. Tech.*, 2012, vol. 55, no. 7, pp. 839–844. DOI: <https://doi.org/10.1007/s11018-012-0048-2>
- [31] Levin G.G., Vishnyakov G.N., Minaev V.L. Automated interference microscope for measurement of dynamic objects. *Pribory i tekhnika eksperimenta*, 2014, no. 1, pp. 79–84 (in Russ.). EDN: RTOVOT. DOI: <https://doi.org/10.7868/S0032816214010066>
- [32] Darvin M.E., Gersonde I., Meinke M., et al. Non-invasive *in vivo* determination of the carotenoids beta-carotene and lycopene concentrations in the human skin using the Raman spectroscopic method. *J. Phys. D: Appl. Phys.*, 2005, vol. 38, no. 15, art. 2696. DOI: <https://doi.org/10.1088/0022-3727/38/15/023>

[33] Baudalet P.-H., Ricochon G., Linder M., et al. A new insight into cell walls of Chlorophyta. *Algal Res.*, 2017, vol. 25, pp. 333–371.

DOI: <https://doi.org/10.1016/j.algal.2017.04.008>

[34] Pirutin S.K., Jia S., Yusipovich A.I., et al. Vibrational spectroscopy as a tool for bioanalytical and biomonitoring studies. *Int. J. Mol. Sci.*, 2023, vol. 24, iss. 8, art. 6947.

DOI: <https://doi.org/10.3390/ijms24086947>

Parshina E.Yu. — Cand. Sc. (Biol.), Senior Researcher, Department of Biophysics, Faculty of Biology, Lomonosov Moscow State University (Leninskie Gory 1, Moscow, 119991 Russian Federation).

Samoylenko A.A. — Cand. Sc. (Biol.), Head of Sector NIO M-44, All-Russian Research Institute for Optical and Physical Measurements (Ozernaya ul. 46, Moscow, 119361 Russian Federation).

Maksimov G.V. — Dr. Sc. (Biol.), Professor, Department of Biophysics, Faculty of Biology, Lomonosov Moscow State University (Leninskie Gory 1, Moscow, 119991 Russian Federation).

Yusipovich A.I. — Cand. Sc. (Biol.), Senior Researcher, Department of Biophysics, Faculty of Biology, Lomonosov Moscow State University (Leninskie Gory 1, Moscow, 119991 Russian Federation).

Lobakova E.S. — Dr. Sc. (Biol.), Professor, Department of Bioengineering, Lomonosov Moscow State University (Leninskie Gory 1, Moscow, 119991 Russian Federation).

He Y. — Post-Graduate Student, Department of Biophysics, Faculty of Biology, Lomonosov Moscow State University (Leninskie Gory 1, Moscow, 119991 Russian Federation).

Levin G.G. — Dr. Sc. (Eng.), Professor, Head of the Department of Holography, Optical Tomography, Nanotechnology and Nanomaterials, All-Russian Research Institute for Optical and Physical Measurements (Ozernaya ul. 46, Moscow, 119361 Russian Federation).

Please cite this article in English as:

Parshina E.Yu., Samoylenko A.A., Maksimov G.V., et al. Integrated approach to studying pigment morphology and distribution in the algae cell. *Herald of the Bauman Moscow State Technical University, Series Natural Sciences*, 2024, no. 2 (113), pp. 129–148 (in Russ.). EDN: NBOFMH

**DETECTION OF ISOTOPIC PEAK SERIES IN LOW-RESOLUTION
MASS SPECTRA USING CLUSTERING ALGORITHM
AND CHI-SQUARE TEST**

V.V. Lebedev
I.S. Pytskii
A.K. Buryak

glory.leb@gmail.com
ivanpic4586@gmail.com
akburyak@mail.ru

**Frumkin Institute of Physical Chemistry and Electrochemistry,
Russian Academy of Sciences, Moscow, Russian Federation**

Abstract

This paper presents the algorithm for determining whether a peak detected in mass spectra during signal processing belongs to isotopic peak series. The algorithm's logic implies preliminary grouping of detected peaks into clusters, checking whether the distribution of peak intensities in each cluster matches the selected pattern, and conducting final grouping which takes the position of peaks along m/z axis into account. The features that enhance the resistance of proposed algorithm to negative phenomena, which can make the detection of isotopic peak series in low-resolution mass spectra by existing methods difficult, are described herein in detail. We present the results of algorithm's functioning with experimental mass spectra of silver(I) chloride and silver(I) bromide used as input. Tested mass spectra were characterized by various negative phenomena that hinder the detection of isotopic peak series. The proposed algorithm is shown to be capable of grouping peaks with the quality similar to existing linear models while avoiding the usage of empirical rules valid only for certain classes of chemical compounds. Since the algorithm requires selection of pattern to model the distribution of intensities in the possible isotopic peak series, we suggest that practical application of proposed algorithm is viable in cases when multiple similar compounds with known pattern of peak intensity distribution are examined using low-resolution mass spectrometer

Keywords

Mass spectrometry, signal processing, isotopic peak series, DBSCAN clustering algorithm, chi-square test

Received 07.06.2023

Accepted 26.09.2023

© Author(s), 2024

This work was supported by a grant from the Russian Science Foundation (grant no. 22-13-00266) for the Frumkin Institute of Physical Chemistry and Electrochemistry, Russian Academy of Sciences

Introduction. Signal processing is the integral part of working with experimental mass spectrum. Processing of signals comprises several consecutive operations, with peak detection being one of the most crucial among them [1]. The latter operation allows detecting the signals that originate from ions formed during fragmentation of the studied compound, i.e., peaks, in the entire array of registered signals. Numerous algorithms for detection of peaks in mass spectrometric data have been described in literature [2].

By its essence, mass spectrometry allows determining the isotopic composition of studied compound and its fragments. Because of that, majority of ions that are formed during mass spectrometric analysis are registered as a series of several peaks. Those series are known as “isotope clusters” in literature. To avoid confusion with machine learning clustering which is referenced in this research, we use “isotopic peak series” term, or simply “isotopic series”, to refer to isotope clusters hereinafter. Peaks that constitute the isotopic peak series correspond to various combinations of isotopes of chemical elements present in the formed ion, and are placed at the distances approximately equal to integer number from each other over m/z axis. Distribution of intensities in isotopic peak series and positions of peaks along m/z axis are two features that provide vitally important information for identification of the formed ion. Respectively, for subsequent analysis, the researcher needs to understand which detected peaks together form isotopic series and originate from the same ions.

Linear models are used in the majority of existing researches devoted to grouping of peaks into isotopic series [3–6]. The decision on merging peaks into single series in such models is primarily based on distance between peaks along m/z axis. Except for paper [3], all other mentioned researches explicitly assume that two peaks can belong to one isotopic series in case the distance between those peaks equals integer number with a margin of error. The maximum tolerable deviation of distance from expected integer number is commonly a user-defined parameter. In papers [5, 6] such deviation of distance was constant over all range of m/z values, meanwhile in research [4] the tolerable deviation increased in a direct proportion to m/z increase. In addition to checking the distance between peaks, all the approaches described in [3–6] implied that the putative isotopic peak series shall comply with special empirical rules. For example, the approach presented in [6] required each succeeding peak in the isotopic series to have lower intensity than the preceding peak. Rules applied in [3, 4] imply usage of coefficients that are calculated using a number of reference mass spectra from various databases. Besides linear models, a contemporary way of grouping peaks into isotopic series is known.

Such a way makes use of machine learning techniques, namely the classifier based on tree ensemble [7]. Several parameters used in [7] when making a decision whether a peak belongs to an isotopic series are also calculated using data from external sources.

It shall be noted that existing approaches to grouping peaks into isotopic series have certain limitations. First of all, majority of such approaches imply usage of high-resolution mass spectra as input data. We failed to find contemporary researches dedicated to grouping peaks from low-resolution mass spectra into isotopic series. The positions of peaks along m/z axis are assumed to be determined very precisely in high resolution mass spectra. This allows significantly (to 10^{-4} m/z and further) reducing the area of search for subsequent peak that possibly belongs to isotopic series. Such an assumption, however, does not hold for low-resolution mass spectra. Due to a number of reasons (e.g., equipment calibration errors or detector overload), the observed distance between two peaks belonging to single isotopic series can deviate from the theoretical integer value by tenths of m/z [8] in low-resolution mass spectra. Secondly, the empirical rules applied in known approaches only allow using such approaches effectively when grouping peaks in mass spectra of compounds belonging to specific classes. For example, the rule of monotonic intensity decrease from approach [4] prevents this methodology from being used for grouping peaks that originate from chlorine- or bromine-containing ions. Furthermore, empirical rules often cannot be easily adapted for usage with compound classes different from the one they were formulated for. Thirdly, the similar problem is common for models that use external data, including machine learning-based models. Such models may yield low quality of grouping peaks into isotopic series when used for processing of the mass spectra of compound classes that were not present in the training set during calculation of decision parameters [3, 7]. In addition, required external data may be unavailable for the selected experiment conditions (e.g., ionization type).

The present research was aimed at creating the algorithm for grouping peaks into isotopic series which could be applied for processing low-resolution mass spectra and which would address the limitations of existing approaches. In particular, the requirements for the algorithm comprised insensitivity to moderate errors in determined positions of isotopic peak along m/z axis, usage of more formal approach when making a decision on merging peaks into isotopic series, and independence from external data. The results are presented in subsequent sections of this paper.

Materials and methods of research. *Peak grouping algorithm.* The proposed algorithm for grouping peaks into isotopic series comprises five steps,

with the first two aimed at obtaining the list of mass spectrum peaks and the remaining three implementing the grouping itself. The full list of steps and their detailed description are presented below:

- 1) reduction of raw mass spectrum's dimensionality;
- 2) peak detection;
- 3) preliminary grouping of peaks into clusters that contain putative isotopic peak series by means of DBSCAN clustering algorithm;
- 4) application of chi-square test for each cluster containing at least 4 peaks to check whether the distribution of peak intensities in cluster matches the chosen pattern;
- 5) evaluation of all preliminary clusters that contain less than 4 peaks and all clusters, for which chi-square test was failed, using traditional "naive" method (i.e., by checking the distances between preliminary cluster peaks along m/z axis) and final grouping.

Smoothing and baseline correction were not carried out as part of this research, however, it is assumed that those operations can be performed before the execution of algorithm starts.

The first step, i.e., dimensionality reduction, implies transformation of raw pseudo-continuous mass spectrum into "centroid"-type mass spectrum, where one peak is represented by one " m/z -intensity" pair of values [9]. In this research, dimensionality reduction was conducted using `detectPeaks()` method from MALDIquant package for R programming language [10]. This method performs the search of local intensity maxima in a sliding window of fixed width. Only signals whose intensity exceeds the median absolute deviation of intensities of all mass spectrum signals are included into the resulting array. The window width was set to the default value proposed by the developers of MALDIquant package and equaled 20 neighboring data points.

The reduced mass spectrum is used as input for peak detection operation. The proposed approach allows detecting peaks by any desired method that accepts a two-dimensional array of signals as input and produces a two-dimensional array of peaks as output. For testing purposes, peak detection was performed by calculation of signal-to-noise ratio (SNR) over non-overlapping windows. The window size was set to 20 m/z , and local noise level was calculated as the median absolute deviation of signals within the window [11]. The intensity of the signal had to exceed the local noise level at least by 5 times (i.e., $\text{SNR} = 5$) in order for signal to be included into peak list.

Once the peak list is formed, the detection of isotopic peak series itself is carried out. At first the peaks are preliminarily merged into clusters that can possibly contain isotopic peak series. The merger is based on position of peaks

on m/z axis and is conducted by means of DBSCAN clustering algorithm [12]. Named machine learning algorithm splits the objects from input array into non-overlapping clusters (i.e. groups, or classes) based on density of objects' distribution in one-dimensional or multi-dimensional space. This algorithm was previously used to determine which signals constitute a peak on the curve in high-resolution mass spectra of original dimensionality [13]. The classical version of DBSCAN algorithm requires two parameters to run, namely maximum distance between neighboring objects that belong to the same class ϵ and minimum number of neighboring objects to be regarded as one cluster $minPts$. Within the scope of task under consideration, the preliminary grouping of peaks into clusters is based on a single feature, i.e., their position along m/z axis. The minimum number of peaks in a single isotopic series is 2; respectively, value of $minPts$ also equals 2. The value of ϵ parameter can be set to 2, 3 or 4 depending on which chemical elements are expected to be present in the studied compound (see explanation below).

Subsequently, the distribution of intensities in each preliminary cluster that contains at least 4 peaks is compared with preselected pattern. The comparison is conducted using chi-square test [14]. Intensities of peaks from the tested cluster are used as observed frequencies. The calculation of theoretical probability distribution is conducted by fitting the peak intensities to the chosen pattern and normalizing fitted values. The constraint on minimum number of peaks 4 in the tested cluster is set by built-in fitting tools used in R programming language. The fitting itself is done by application of `nls()` built-in function which implements the non-linear least squares technique. The Gaussian function was implemented in a custom-written code and is passed to `nls()` as a formula to fit experimental data to. Values of `nls()` function's control parameters, i.e., maximum number of iterations and convergence criterion, are set to the defaults provided in the R programming language.

The preliminary clusters that passed the statistical test are considered to be correctly grouped isotopic series and are excluded from further processing. The chi-square test replaces the empirical rules which are used in researches [3–6] when making a decision on retaining or splitting putative isotopic series based on distribution of peak intensities. In particular, this test prevents the incorrect merger of two overlapping isotopic series into one group. Furthermore, application of chi-square test allows the algorithm to retain peaks that lie out of expected range of search in isotopic series without the need to change the tolerable distance deviation (given that the observed and expected distributions of intensities match).

The final grouping of peaks in clusters that contain less than 4 peaks and in those that have failed the test is conducted by means of “naive” approach, i.e., by computing the distances between cluster peaks. This approach was slightly modified to enable grouping of isotopic series with various expected integer distances (1, 2, etc.) between consecutive peaks during the single run. In addition to traditional *eps* parameter (maximum tolerable deviation of distance between two peaks from expected integer value) our implementation of “naive” grouping requires the value of maximum expected integer distance $dm^{(z)} \in \mathbb{N}$ to be specified. The implementation involves searching the cluster for isotopic peaks placed at distances of $1 \pm eps$, then $2 \pm eps$, etc. until $dm^{(z)} \pm eps$. The peaks merged into the isotopic series which is characterized by certain integer distance are excluded from further iterations. This modification was aimed at enabling the algorithm to be used for processing mass spectra where various classes of compounds (e.g., chlorides of mono-isotopic metals, $dm^{(z)} = 2$, and organic impurities, $dm^{(z)} = 1$ [15]) are observed simultaneously. It shall be noted that, like known linear models, the proposed approach assumes that peaks forming the isotopic series are placed at roughly equal distances from each other. Currently around 35 chemical elements are characterized by phenomenon when masses of heavier isotope and previous lighter isotope differ by the same integer mass value for all consecutive isotope pairs. The aforementioned assumption stands true for all ions which contain at least one such element in their composition. In practice, stated assumption allows correctly grouping peaks of majority of registered ions into isotopic series. However, this still sets the constraint on algorithm’s area of application that needs to be taken into account.

The output of the proposed algorithm is the $n \times 3$ matrix, where n is the number of peaks detected on step 2. Matrix columns represented m/z , intensity and assigned number of isotopic series for each peak. Putative monoisotopic peaks were also assigned a unique number.

Input data used. The proposed algorithm for grouping peaks into isotopic series was tested with experimental low-resolution mass spectra acquired on laser desorption mass spectrometer.

Mass spectra of silver(I) chloride and silver(I) bromide were selected as input data. The distribution of peak intensities obtained by convolution of “isotopic distributions” of several metals with one or two abundant isotopes, including silver, with respective distributions of chlorine and bromine could be relatively easily approximated by a Gaussian function. This allows for more detailed control over the testing procedure in general and chi-square test aspects in particular.

Selection of silver compounds from the list of aforementioned metals simplifies the construction of expected grouping results array, which is required to calculate metrics of algorithm's functioning quality. Such calculation requires testing the algorithm on a set of raw mass spectra with list of annotated peaks available for each of them. To the best of our knowledge, the sets of metal halides mass spectra acquired by laser desorption method are not available in open sources. Respectively, the required mass spectra had to be acquired, and observed isotopic peak series needed to be identified. The laws of ion formation during laser desorption ionization are well known for silver halides [16]. This allowed identifying the fragments of studied compounds manually and constructing the required validation data set that contained expected grouping results.

Four experimental mass spectra were selected in order to test the proposed algorithm: (1) AgCl (positively charged ions registration mode, the registration mode is later denoted with "+" or "-" signs); (2) AgBr (-); (3) AgBr (+); (4) AgCl with bromine impurities (-). All mass-spectra were acquired using *Bruker Daltonics Ultraflex II* mass spectrometer (operating wavelength 337 nm, pulse energy 100 μ J, pulse power 43 kW, repetition rate 20 Hz). Matrix substance was not used, since examined compounds are readily ionized in the absence of matrix [16]. Input mass spectra were selected in a way so that the algorithm could be tested in various conditions.

Mass spectrum 1 represented nearly ideal conditions for grouping peaks into isotopic series. The distances between all consequent peaks belonging to single isotopic series were almost equal, and position of each consequent peak on m/z axis was within the expected search area.

When dimensionality reduction was carried out, the positions of several isotopic peaks in mass spectrum 2 were determined incorrectly (probably due to unsuitable size of window for local maxima search). Because of that, distances between some of the peaks deviated from expected integer value by up to 0.27 m/z . This presented an opportunity to check whether chi-square test could help in grouping peaks, which in fact belong to isotopic series but lie out of expected search range over m/z axis.

The detector overload (i.e., registration of multiple signals with decreasing intensities following the registration of "true" peak corresponding to fragment ion) was observed when mass spectrum 3 was acquired. The consequences of this effect were not fully removed during dimensionality reduction. This allowed testing algorithm's ability to correctly detect isotopic peak series in preliminary clusters that contain multiple high-intensity noise signals.

Mass spectrum 4 was characterized by presence of several isotopic peak series positioned one after another. The distances between the first peak in subsequent series and the last peak in previous series were approximately equal to the distances between two peaks within one isotopic series. This feature of mass spectra was used to test whether algorithm is capable of correctly splitting such peak sequences into isotopic series.

The manual identification of ions registered in used mass spectra was carried out prior to start of testing procedures. Identification was conducted by comparing theoretical distribution of peak intensities for a given molecular formula with the intensity distribution observed in the registered peak series. Theoretical distribution was calculated using *IsoSpecR* package [17]. The decision on whether the observed and theoretical intensity distributions match was based on a weighted sum score. For each peak from observed isotopic series, an absolute difference between relative intensity of observed peak and relative intensity of the theoretical peak closest to the observed one by m/z value was computed. If such absolute difference did not exceed user-defined threshold, a value equal to relative intensity of theoretical peak was added to the sum score (with weight of 1). Otherwise, sum score remained unchanged. The theoretical distribution was considered to match the experimental one if the resulting sum score equaled at least $(1 - \text{threshold})$. In this study, the threshold was set to value of 0.05 (e.g., a maximum 5 % difference in relative intensities of corresponding peaks). A total of 31 isotopic peak series, comprising 121 peaks, were identified across 4 mass spectra. Results of identification procedure were used to construct arrays of expected grouping results, which, in turn, allowed assessing quality of algorithm's results.

Testing procedure. The proposed peak grouping algorithm was implemented in R programming language [18]. Functionality of MALDIquant package was used to read raw files and reduce dimensions of mass spectra loaded into memory. The remaining algorithm steps were implemented in custom scripts.

A total of 27, 35, 160 and 70 peaks were detected in mass spectra 1 to 4, respectively. Several signals, which constituted the part of manually identified isotopic series according to theoretical intensity distributions but were not automatically detected as peaks, were added into the peak lists. Relative intensity of signal in theoretical distribution for identified series had to exceed 5 % in order for signal to be added into the list of peaks. No peaks detected on step 2 were removed from the peak list. Final number of signals classified as peaks amounted to 29, 36, 162 and 72 in mass spectra 1–4, respectively.

The data obtained during manual identification and formed peak lists were used to obtain the matrices of expected results of grouping peaks into isotopic

series for each mass spectra. The same group number (label) was assigned to all experimental peaks which belonged to the same series. The numbering started from 1. Numbers were assigned consecutively and increased as lowest peak m/z value in series increased. The unidentified and noise peaks which were detected automatically were not included into the matrices of expected grouping results.

Detection of isotopic peak series and evaluation of proposed algorithm's quality were carried out subsequently. The value of ϵ parameter of DBSCAN algorithm was set to 3, meaning that consequent peaks separated by a distance of no more than 3 m/z fell into one preliminary cluster. Isotopes of each chemical element present in studied compounds and their fragments (Ag, Br, Cl) differ in mass by approximately 2 Da. Under such conditions, two consequent peaks located at a distance greater than 3 m/z from each other are guaranteed not to belong to a single isotopic series, with an account for all possible errors. The $dm^{(z)}$ and eps parameters of "naive" grouping method equaled 2 and 0.1, respectively. Value of $dm^{(z)}$ parameter was selected based on consideration regarding isotope mass differences mentioned previously. The maximum tolerable deviation of distance between two consequent peaks from expected integer number eps was set to 0.1, since all m/z values listed in official description of used mass spectrometer are specified with a precision of $0.1 m/z^*$. Critical values of chi-square test statistics were calculated at confidence levels cl of 0.9, 0.95 and 0.99 in different runs.

Several metrics were calculated in order to assess the quality of peak grouping that was attained by application of proposed algorithm. These metrics included the share of peaks assigned to correct isotopic series, shares of type I and type II errors, precision and recall. The total number of peaks that fell into preliminary clusters, which contained the most intensive peaks of identified isotopic series, was used as the base for calculation of the first metrics from the list. Within the scope of task solved, type I error (false-positive decision) means inclusion of experimental signal into isotopic series, to which this signal does not belong in reality. Respectively, type II error (false-negative decision) stands for exclusion of peak, which constitutes a part of isotopic series, from such a series. Precision and recall were calculated in a manner traditional to machine learning [19]. Several additional special metrics were calculated in order

* Measurement instrument type description. Mass spectrometers of models: microflex LT/SH, microflex LRF, autoflex speed LIN, autoflex speed LRF, autoflex speed TOF/TOF, ultrafleXtreme, rapifleX MALDI TOF MS, rapifleX MALDI TOF/TOF. Federal Agency on Technical Regulating and Metrology of Russia, 2021.

to assess impact of chi-square test. The detailed results are presented in the following section of the paper.

Results and discussion. *Quality of obtained results.* The metrics reflecting the quality of grouping peaks into isotopic series by means of proposed algorithm are presented in Table 1. In order to assess the effect from joint usage of DBSCAN clustering, chi-square test and “naive” grouping, the same metrics were also calculated for grouping results obtained by application of DBSCAN algorithm separately and of “naive” grouping separately.

Table 1

Quality metrics for the proposed algorithm and its individual components

Metrics	DBSCAN	“Naive” grouping	Complete algorithm		
			<i>cl</i> = 0.9	<i>cl</i> = 0.95	<i>cl</i> = 0.99
Number of peaks that fell into clusters containing putative series	161	123	124		
Share of peaks assigned to correct isotopic series, %	73.29	85.37	90.32	90.32	91.13
Share of type I errors, %	26.71	1.63	2.42	2.42	2.42
Share of type II errors, %	0.00	13.01	7.26	7.26	6.45
Precision, %	73.29	98.13	97.39	97.39	97.41
Recall, %	100.00	86.78	92.56	92.56	93.39
Absolute difference of “precision–recall”, %	26.71	11.35	4.83	4.83	4.03
Number of isotopic series identified manually	31				
Share of series detected fully correctly	58.06	51.61	67.74	67.74	70.97

The algorithm correctly assigned circa 90 % of all peaks to the same isotopic series as expected. Around 70 % of manually identified isotopic peak series were detected fully correctly. The selection of confidence level for calculation of chi-square critical value did not impact the results significantly.

The share of peaks assigned to correct isotopic series and share of isotopic peaks series detected fully correctly were lower for results obtained by applying only DBSCAN algorithm or only “naive” grouping method and amounted to around 73 / 58 % and 85 / 52 %, respectively. Thus, joint application of DBSCAN clustering algorithm, chi-square statistic test and “naive” grouping improves the results of grouping peaks into isotopic series when compared with usage of a single linear model.

We shall note that, among all the used methods of peak grouping, only application of the proposed algorithm in its complete form allowed reaching both relatively high (97 and 93 %, respectively) and most balanced precision and recall values. This indicates that shares of false-positive and false-negative decisions made by algorithm are low and approximately equal. In comparison, grouping of peaks into isotopic series by means of only DBSCAN algorithm allows obtaining a recall of 100 % (i.e., no type II errors). However, this clustering algorithm is not capable of detecting isotopic peak series in noisy data or handling overlapping series by its nature. All the noise and overlapping signals would be incorrectly grouped into one cluster, which will cause multiple type I errors. In practice this results in very low (73 %) precision of peak grouping when DBSCAN algorithm is used separately. Similarly, application of only “naive” grouping results in slightly greater precision when compared with the complete proposed algorithm. However, the model based purely on calculation of distance is prone to false-negative decisions in cases when subsequent isotopic peak lies out of its expected search range along m/z axis. As the result, the recall value attained by “naive” method is rather low and amounts only to 87 %.

Application of the proposed algorithm for grouping peaks into isotopic series allows achieving a compromise between shares of type I and II errors due to usage of chi-square test and processing of the preliminary cluster depending of results of such test. Table 2 shows that from 29 to 33 % of identified isotopic peak series, which were already grouped fully correctly at clustering stage, would have been incorrectly ungrouped by “naive” method. However, the distribution of peak intensities in such series matches the expected one, which allowed those series to pass statistic test and prevented incorrect processing. Depending of selected confidence level, 43 to 47 % of all conducted chi-square tests had a positive impact on result of peak grouping, and in further 33 % of cases such test did not harm the quality. It shall be separately noted that preliminary clusters containing multiple noise signals (common for mass spectrum 3) did not pass the test and were correctly passed to “naive” grouping step for a subsequent search for isotopic peak series.

Precision and recall metrics allow comparing the quality of peak grouping achieved by the proposed algorithm and other existing approaches. Application of presented algorithm results in precision value similar to the one of linear model described in [6] (97.4 % against 97.7 %, respectively) and, formally, a greater recall (92.6 % against 70.4 %, respectively). However, we shall note that the recall reported in [6] is suspiciously low (this might have been caused by some peculiarities of methodic that was used to count false-negative decisions) and requires clarification. When compared with machine learning-

based approach described in [7], the proposed algorithm, as expected, demonstrated lower quality of results (precision of 99.5 % for model [7] against 97.4 % for proposed algorithm; recall of 99.9 % and 92.6 %, respectively). In exchange, the proposed algorithm does not require generating training set that depends on external sources in order to tune model parameters.

Table 2

Detailed results of chi-square test usage within the proposed algorithm

Metrics	Confidence level		
	0.9	0.95	0.99
Number of preliminary clusters containing isotopic peak series with expected distribution of intensities	21		
Share of preliminary clusters that passed chi-square test, %, including:	57.14	61.90	66.67
(I) were correctly grouped during preliminary clustering and would not have been ungrouped by “naive” method	28.57	33.33	33.33
(II) would have been ungrouped correctly	4.76	4.76	4.76
(III) would have been ungrouped incorrectly	23.81	23.81	28.57
Share of preliminary clusters that failed chi-square test, %, including:	42.86	38.10	33.33
(IV) were correctly grouped during preliminary clustering but were incorrectly ungrouped	4.76	0	0
(V) were correctly ungrouped	19.05	19.05	19.05
(VI) were incorrectly ungrouped	19.05	19.05	14.29
Share of “useful” chi-square test applications (III + V), %	42.86	42.86	47.62
Share of “harmless” chi-square test applications (I + IV), %	33.33	33.33	33.33

Generalization of algorithm for various compound classes. This paragraph shortly describes the adjustment of parameter values required for algorithm to be used with mass spectra of various compound classes.

The exact values of dimensionality reduction and peak detection parameters do not set technical limitations on ability to run the algorithm. However, in cases when all the consequent peaks in resulting peak list are separated by a distance less than ϵ parameter value of DBSCAN algorithm due to “soft” peak detection settings, the grouping will only be conducted by “naive” method, since all the detected peaks will fall into one preliminary cluster, which

is unlikely to pass chi-square test. Thus, “soft” dimensionality reduction and peak detection parameter values should be avoided.

Selection of ϵ parameter value for DBSCAN algorithm should be based on computing the maximum difference of masses of two consequent naturally occurring isotopes among all chemical elements that are expected to be found in examined compound. Elements, for which such difference amounts to $\sim 1, 2$ and 3 Da, are currently known to exist. To account for all possible measurement errors, we suggest setting the value of ϵ parameter to 2, 3 or 4 depending on expected composition of studied compound. For example, $\epsilon = 2$ is suitable for relatively simple organic compounds, value of $\epsilon = 3$ is suitable for chlorides and bromides of monoisotopic metals, etc.

The most important step of adjusting algorithm’s parameters is the selection of intensity distribution pattern, which will be used to compare the distribution of intensities observed in preliminary cluster with distribution that a valid isotopic series is assumed to have. Selection of such pattern requires expert judgment. For example, probability function of geometric distribution might be used to approximate distribution of mass peak intensities of organic compounds that contain only C, H, N, O, S and P. A mixture of distributions could also be used for approximation purposes. Say, distribution of mass peak intensities of lead chlorides could be modeled by a mixture of two normal distributions with centers at m/z values of peaks characterized by largest input from ^{206}Pb and ^{208}Pb isotopes. In practice, distributions of mass peak intensities observed for registered ions usually match one of the patterns from a finite list [20]. Thus, once the approximation law is formulated, it can be re-used.

Finally, as discussed previously, value of $dm^{(z)}$ parameter required by “naive” grouping step should be set to 1, 2 or 3, depending on the expected composition of analyzed compound. Selection of maximum tolerable distance deviation *eps* does not depend on the nature of analyte and should be based on specifications listed in the technical documents for the used mass spectrometer.

Conclusion. The algorithm for grouping peaks detected in low-resolution mass spectra into isotopic series was presented in this paper. The algorithm is based on joint application of DBSCAN clustering, chi-square test and a linear model implying calculation of distances between consequent peaks. When used to process experimental mass spectra of silver(I) chloride and silver(I) bromide, the proposed algorithm showed relative insensitivity to several phenomena which commonly occur in low-resolution mass spectra and hinder automatic detection of isotopic peak series by known approaches. The algorithm demonstrated precision and recall values similar to the ones of existing linear models that are used to process high-resolution mass spectra. At the same time,

the proposed algorithm does not use empirical rules valid only for certain compound class when making a decision whether a peak belongs to putative isotopic series or not. In comparison with peak grouping models based on machine learning, proposed algorithm demonstrates moderately lower quality of results in exchange for dropping the dependence on external data and pre-training. The most significant limitation of the proposed algorithm is inability to correctly group peaks originating from ions that are composed only of certain elements. Algorithm is also characterized by one feature, which, in our opinion, is neither the advantage nor limitation. In order for algorithm to be effectively used with mass spectra of compounds belonging to a class not studied previously, the law of approximation of mass peak intensities distribution needs to be formulated. On one hand, this feature implies additional time costs when mass spectra of different compound classes need to be processed. On the other hand, such a feature allows easily adapting the proposed algorithm to work with mass spectra of various compound classes without the need to change the general logic of the algorithm. Existing linear models used for processing of high-resolution mass spectra are highly selective to input data and do not provide such scaling capabilities. Thus, application of the proposed algorithm seems appropriate in cases when multiple compounds of the same class, for which a law of approximation of mass peak intensities is known, are analyzed using the low-resolution mass spectrometer.

REFERENCES

- [1] Bauer C., Cramer R., Schuchhardt J. Evaluation of peak-picking algorithms for protein mass spectrometry. In: Hamacher M., Eisenacher M., Stephan C. (eds). *Data Mining in Proteomics. Methods in Molecular Biology*, vol. 696. Humana Press, 2011, pp. 341–352. DOI: https://doi.org/10.1007/978-1-60761-987-1_22
- [2] Yang C., He Z., Yu W. Comparison of public peak detection algorithms for MALDI mass spectrometry data analysis. *BMC Bioinformatics*, 2009, vol. 10, art, no. 4. DOI: <https://doi.org/10.1186/1471-2105-10-4>
- [3] Jurášek P., Slimák M., Košík M. Determination of isotope cluster patterns in mass spectra of GC-MS Analyses by a chemometric detector. *Microchim. Acta*, 1993, vol. 110, pp. 133–142. DOI: <https://doi.org/10.1007/BF01245097>
- [4] Treutler H., Neumann S. Prediction, detection, and validation of isotope clusters in mass spectrometry data. *Metabolites*, 2016, vol. 6, iss. 4. DOI: <https://doi.org/10.3390/metabo6040037>
- [5] Teo G.C., Polasky D.A., Yu F., et al. Fast deisotoping algorithm and its implementation in the MSFragger search engine. *J. Proteome Res.*, 2019, vol. 20, iss. 1, pp. 498–505. DOI: <https://doi.org/10.1021/acs.jproteome.0c00544>

- [6] Tay A.P., Liang A., Hamey J.J., et al. MS2-Deisotoper: a tool for deisotoping high-resolution MS/MS spectra in normal and heavy isotope-labelled samples. *Proteomics*, 2019, vol. 19, iss. 17, art. 1800444. DOI: <https://doi.org/10.1002/pmic.201800444>
- [7] Boiko D.A., Kozlov K.S., Burykina J.V., et al. Fully automated unconstrained analysis of high-resolution mass spectrometry data with machine learning. *J. Am. Chem. Soc.*, 2022, vol. 144, iss. 32, pp. 14590–14606. DOI: <https://doi.org/10.1021/jacs.2c03631>
- [8] Brenton A.G., Godfrey A.R. Accurate mass measurement: terminology and treatment of data. *J. Am. Soc. Mass Spectr.*, 2010, vol. 21, iss. 11, pp. 1821–1835. DOI: <https://doi.org/10.1016/j.jasms.2010.06.006>
- [9] Urban J., Afseth N.K., Štys D. Fundamental definitions and confusions in mass spectrometry about mass assignment, centroiding and resolution. *TrAC*, 2014, vol. 53, pp. 126–136. DOI: <https://doi.org/10.1016/j.trac.2013.07.010>
- [10] Gibb S., Strimmer K. MALDIquant: a versatile R package for the analysis of mass spectrometry data. *Bioinformatics*, 2012, vol. 28, iss. 17, pp. 2270–2271. DOI: <https://doi.org/10.1093/bioinformatics/bts447>
- [11] Li X., Gentleman R., Shi Q., et al. SELDI-TOF mass spectrometry protein data. In: Gentleman R., Carey V.J., Huber W., Irizarry R.A., Dudoit S. (eds). *Bioinformatics and Computational Biology Solutions Using R and Bioconductor. Statistics for Biology and Health*. New York, Springer New York, 2005. pp. 91–109. DOI: https://doi.org/10.1007/0-387-29362-0_6
- [12] Ester M., Kriegel H.-P., Sander J., et al. A density-based algorithm for discovering clusters in large spatial databases with noise. *KDD-96 Proc.* AAAI Press, 1996, pp. 226–231.
- [13] Wei X., Shi X., Kim S., et al. Data dependent peak model based spectrum deconvolution for analysis of high resolution LC-MS Data. *Anal. Chem.*, 2014, vol. 86, iss. 4, pp. 2156–2165. DOI: <https://doi.org/10.1021/ac403803a>
- [14] Pearson K.X. On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 1900, vol. 50, iss. 302, pp. 157–175. DOI: <https://doi.org/10.1080/14786440009463897>
- [15] Lebedev V.V., Buryak A.K. Usage of Kohonen clustering algorithm for rough peak detection during mass spectrum preprocessing. *Mass-spektrometria* [Mass-Spectrometry], 2022, vol. 19, no. 3, pp. 137–148 (in Russ.). EDN: NVGYFO. DOI: <https://doi.org/10.25703/MS.2022.19.15>
- [16] Pytskii I.S., Buryak A.K. MALDI/SELDI mass-spectrometric surface investigation of AMg-6 and Ad-0 materials. *Prot. Met. Phys. Chem. Surf.*, 2011, vol. 47, iss. 6, pp. 756–761. DOI: <https://doi.org/10.1134/S2070205111060165>
- [17] Łączki M.K., Startek M., Valkenborg D., et al. IsoSpec: hyperfast fine structure calculator. *Anal. Chem.*, 2017, vol. 89, iss. 6, pp. 3272–3277. DOI: <https://doi.org/10.1021/acs.analchem.6b01459>

- [18] R: A language and environment for statistical computing. R Foundation for Statistical Computing. Available at: <https://www.R-project.org> (accessed: 23.05.2023).
- [19] Olson D.L., Delen D. Performance evaluation for predictive modeling. In: *Advanced Data Mining Techniques*. Berlin, Heidelberg, Springer, 2008, pp. 137–147. DOI: https://doi.org/10.1007/978-3-540-76917-0_9 2008
- [20] Goldfarb D., Lafferty M.J., Herring L.E., et al. Approximating isotope distributions of biomolecule fragments. *ACS Omega*, 2018, vol. 3, iss. 9, pp. 11383–11391. DOI: <https://doi.org/10.1021/acsomega.8b01649>

Lebedev V.V. — Junior Researcher, Frumkin Institute of Physical Chemistry and Electrochemistry, Russian Academy of Sciences (Leninskiy prospekt 31, korp. 4, Moscow, 119071 Russian Federation).

Pytskii I.S. — Cand. Sc. (Chem.), Leading Researcher, Frumkin Institute of Physical Chemistry and Electrochemistry, Russian Academy of Sciences (Leninskiy prospekt 31, korp. 4, Moscow, 119071 Russian Federation).

Buryak A.K. — Corresponding Member of the Russian Academy of Sciences, Dr. Sc. (Chem.), Director, Frumkin Institute of Physical Chemistry and Electrochemistry, Russian Academy of Sciences (Leninskiy prospekt 31, korp. 4, Moscow, 119071 Russian Federation).

Please cite this article as:

Lebedev V.V., Pytskii I.S., Buryak A.K. Detection of isotopic peak series in low-resolution mass spectra using clustering algorithm and chi-square test. *Herald of the Bauman Moscow State Technical University, Series Natural Sciences*, 2024, no. 2 (113), pp. 149–164. EDN: ODEQRN