

Методы анализа данных малоуглового рентгеновского рассеяния многокомпонентными биомолекулярными системами

Конарев П.В.

НИИ «Курчатовский институт», Москва, Россия

konarev@crys.ras.ru

Ключевые слова: малоугловое рентгеновское рассеяние; многокомпонентные биомолекулярные системы; пространственная форма частиц; гель-хроматография; эволюционный факторный анализ

Мотивация и цель: В настоящее время большую роль при рациональном конструировании новых лекарственных средств и развитии методов терапии заболеваний играют структурные методы исследования вещества, среди которых особо следует выделить малоугловое рентгеновское рассеяние (МУРР). Метод МУРР позволяет изучать некристаллические образцы различной природы с размерами частиц от 1 до 200 нм и более. Этот метод является неинвазивным, не требует сложной пробоподготовки и дает ценную информацию о структуре и динамике исследуемых объектов в условиях, близких к нативным. До недавнего времени анализ полидисперсных многокомпонентных биомолекулярных систем по данным МУРР был существенным образом ограничен и позволял оценивать лишь общие структурные характеристики, усредненные по ансамблю частиц. Недавние достижения в разработке алгоритмов позволяют гораздо более детально исследовать полидисперсные системы, в частности белковые или липидные многокомпонентные смеси.

Методы и алгоритмы: При наличии известных кривых малоуглового рассеяния от каждого макромолекулярного компонента можно определять их объемные доли в общей смеси [1, 2]. Если структуру компонентов можно аппроксимировать простыми геометрическими телами, то, помимо определения объемных долей компонент, появляется возможность оценивать функции распределения частиц по размерам и характер их межчастичного взаимодействия в случае концентрированных систем [1, 2]. Для липидных многокомпонентных систем, помимо определения функции распределения везикул по размерам, проводится восстановление электронной плотности липидного бислоя внутри везикулы и оценка параметров мультислойной ламеллярной организации везикул [3, 4].

Хотя прямой анализ формы частиц для многокомпонентных смесей сильно затруднен, в ряде случаев это становится возможным благодаря новому подходу прямого определения формы промежуточных компонентов в эволюционирующих системах [5]. Системы, для которых известны начальное и конечное состояния, но которые также содержат неизвестное промежуточное состояние, могут быть непосредственно проанализированы с помощью этого подхода.

Измерения МУРР в сочетании с использованием гель-хроматографической ячейки, когда раствор образца, выходящий из хроматографической колонки, сразу направляется в камеру образца, позволяют разделить во времени вклады

отдельных (чистых и монодисперсных) компонентов, присутствующих в системе, но не всегда удастся достичь полного разделения компонентов. Для частично перекрывающихся пиков выхода компонентов смеси предложен новый алгоритм восстановления профилей рассеяния и концентрации компонентов на основе эволюционного факторного анализа (ЭФА) [6]. Установлены границы применимости подхода в зависимости от уровня шума во входных данных МУРР, степени перекрытия концентрационных профилей компонентов и их асимметрии, что важно для его практических применений.

Результаты: С помощью количественного анализа данных малоуглового рассеяния с использованием параметрических методов линейной и нелинейной минимизации удалось: (i) выявить случайно-столкновительный механизм образования комплекса адренодоксина (Adx) с митохондриальным цитохромом с (Cc) [7]; (ii) показать способность фермента люмазинсинтазы самособираться в капсулы размером от 16 до 32 нм, причем их количественный состав зависит как от наличия мутаций белка, так и от изменений физико-химических условий среды [8]; (iii) проанализировать эволюцию структурных параметров везикул DMPC/DPPC в процессе экструзии, позволяющую получать однослойные везикулы с радиусом (30 ± 1) нм [4]; (iv) для смесей, состоящих из вирусных липидов и матричного белка M1 вируса гриппа А, оценить влияние гемагглютиниона на формирование липопротеиновых нанодоменов в липидном бислое [9].

Алгоритм восстановления формы частиц для многокомпонентных смесей был успешно применен для изучения процесса образования фибрилл белка инсулина и позволил определить форму промежуточного компонента в процессе их роста, который представляет собой вытянутую структуру размером 20 нм, состоящую из пяти связанных доменов, по размерам близких к мономерам инсулина, что подтверждает гипотезу об олигомерном ядре, являющемся первичной удлиняющей единицей инсулиновых амилоидных фибрилл [5]. С помощью предложенного метода для белка фактора роста нервов была (случай двухкомпонентной смеси) восстановлена трехмерная форма димера по концентрационной серии данных МУРР [5].

С помощью ЭФА алгоритма определено, что для смеси белка пируват-альдолазы ее олигомерными компонентами в растворе являются гексамеры и октамеры [6]. Восстановленные профили рассеяния компонентов хорошо соотносятся с известными кристаллографическими моделями белков.

Выводы: Разработанные методы существенно расширяют возможности анализа многокомпонентных наноразмерных систем методом МУРР и могут быть применены для широкого круга биологических, органических и композитных систем.

Финансирование: Работа проведена в рамках выполнения государственного задания НИЦ «Курчатовский институт».

Methods for the analysis of small-angle X-ray scattering data from multicomponent biomolecular systems

Konarev P.V.

NRC "Kurchatov Institute", Moscow, Russia

konarev@crys.ras.ru

Key words: small-angle X-ray scattering; multicomponent biomolecular systems; 3D shape of particles; gel chromatography; evolutionary factor analysis

Motivation and Aim: Nowadays, structural methods play an important role in rational design of new drugs and development of therapies for diseases, among which small-angle X-ray scattering (SAXS) should be emphasized. SAXS allows studying non-crystalline samples of different nature with particle sizes from 1 to 200 nm and more. This method is non-invasive, does not require special sample preparation and provides valuable information on the structure and dynamics of the investigated objects under nearly physiological conditions. Until recently, the analysis of SAXS data from polydisperse multicomponent biomolecular systems was significantly limited and allowed only the evaluation of general structural characteristics averaged over an ensemble of particles. Recent advances in algorithm development allow one to extract much more detailed structural information for polydisperse systems, in particular protein or lipid multicomponent mixtures.

Methods and Algorithms: In the presence of known small-angle scattering curves from each macromolecular component, it is possible to determine their volume fractions in a mixture [1, 2]. If the structure of the components can be approximated by simple geometric bodies, in addition to determining the volume fractions of the components, it is possible to estimate the particle size distribution functions and the interparticle interactions in the case of concentrated systems [1, 2]. For lipid multicomponent systems, in addition to determining the size distribution function of vesicles, the electron density of the lipid bilayer inside the vesicle is reconstructed and the parameters of the multilayer lamellar organization of vesicles are estimated [3, 4].

Although direct particle shape analysis for multicomponent mixtures is normally not applicable, a new approach to directly determine the shape of intermediate components in evolving systems makes this possible in some cases [5]. Systems for which the initial and final states are known, but which also contain an unknown intermediate state, can be directly analyzed using this approach.

SAXS measurements combined with the use of a size exclusion chromatography (SEC), where the sample solution leaving the chromatographic column is immediately sent to the sample chamber, allow for the separation in time of the contributions of the individual (pure and monodisperse) components present in the system, but it is not always possible to achieve complete separation of the components. For partially overlapping peaks of the mixture components, a new algorithm for reconstructing the scattering and concentration profiles of components based on evolutionary factor analysis (EFA) has been proposed [6]. The limitations of the approach are evaluated depending on the noise level of the input SEC-SAXS data, the degree of overlap of the concentration profiles of components and their asymmetry, which is important for its practical applications.

Results: By quantitative analysis of small angle scattering data using parametric linear and nonlinear minimisation methods, we were able to: (i) reveal the encounter-stochastic mechanism of adrenodoxin (Adx) complex formation with mitochondrial cytochrome c (Cc) [7]; (ii) show the ability of the enzyme lumazine synthase to self-assemble into icosahedral capsids ranging in size from 16 to 32 nm, and their composition depends on both the presence of protein mutations and changes in physicochemical environmental conditions [8]; (iii) analyze the evolution of the structural parameters of DMPC/DPPPC vesicles during extrusion, allowing the production of single-layer vesicles with a radius

of (30 ± 1) nm [4]; (iv) for mixtures consisting of viral lipids and matrix protein M1 of influenza A virus, evaluate the effect of hemagglutinin on the formation of lipoprotein nanodomains in the lipid bilayer [9].

The particle shape reconstruction algorithm for multicomponent mixtures was successfully applied to study the formation of insulin protein fibrils and allowed us to determine the shape of the intermediate component in the process of their growth, which is an elongated structure with a size of 20 nm, consisting of five bound domains with sizes close to insulin monomers, which confirms the hypothesis of an oligomeric precursor, which is the primary elongation unit of insulin amyloid fibrils [5]. Using the proposed method, a three-dimensional shape of hetero-tetramer was reconstructed for the nerve growth factor protein (case of a two-component mixture) from a concentration series of SAXS data [5].

Using the EFA algorithm it was determined that for the pyruvate aldolase protein mixture its oligomeric components in solution are hexamers and octamers [6]. The reconstructed scattering profiles of the components agree well with the scattering curves calculated from the known crystallographic models of proteins.

Conclusion: The developed methods significantly expand the possibilities of analyzing multicomponent nanoscale systems by the SAXS method and can be applied to a wide range of biological, organic and composite systems.

Funding: The work was carried out within the state assignment of the NRC “Kurchatov Institute”.

Список литературы/References

1. Konarev P.V., Volkov V.V., Sokolova A.V., Koch M.H.J., Svergun D.I. PRIMUS – a Windows-PC based system for small-angle scattering data analysis. *J Appl Cryst.* 2003;36:1277-1282
2. Manalastas-Cantos K., Konarev P.V., Hajizadeh N.R. et al. ATSAS 3.0: expanded functionality and new tools for small-angle scattering data analysis. *J Appl Cryst.* 2021;54:343-355
3. Konarev P.V., Petoukhov M.V., Dadinova L.A., Fedorova N.V., Volynsky P.E., Svergun D.I., Batishchev O.V., Shtykova E.V. BILMIX: a new approach to restore the size polydispersity and electron density profiles of lipid bilayers from liposomes using small-angle X-ray scattering data. *J Appl Cryst.* 2020;53:236-243
4. Konarev P.V., Gruzinov A.Y., Mertens H.D.T., Svergun D.I. Restoring structural parameters of lipid mixtures from small-angle X-ray scattering data. *J Appl Cryst.* 2021;54:169-179
5. Konarev P.V., Svergun D.I. Direct shape determination of intermediates in evolving macromolecular solutions from small-angle scattering data. *IUCrJ.* 2018;5:402-409
6. Konarev P.V., Graewert M.A., Jeffries C.M., Fukuda M., Cheremnykh T.A., Volkov V.V., Svergun D.I. EFAMIX, a tool to decompose inline chromatography SAXS data from partially overlapping components *Protein Sci.* 2022;31:269-282
7. Xu X., Reinle W., Hannemann F., Konarev P.V., Svergun D.I., Bernhardt R., Ubbink M. Dynamics in a pure encounter complex of two proteins studied by solution scattering and paramagnetic NMR spectroscopy. *J Am Chem Soc.* 2008;130:6395-6403
8. Zhang X., Konarev P.V., Petoukhov M.V., Svergun D.I., Xing L., Cheng R.H., Haase I., Fischer M., Bacher A., Ladenstein R., Meining W. Multiple assembly states of lumazine synthase: a model relating catalytic function and molecular assembly. *J Mol Biol.* 2006;362:753-770
9. Kordyukova L.V., Konarev P.V., Fedorova N.V. et al. The cytoplasmic tail of influenza A virus hemagglutinin and membrane lipid composition change the mode of M1 protein association with the lipid bilayer. *Membranes.* 2021;11:772