The Journal of Open Source Software

Molass Library: A Python Package for SEC-SAXS Data Analysis

Masatsuyo Takahashi [1] **and Nobutaka Shimizu** [1,2,3¶]

1 Structural Biology Research Center, Institute of Materials Structure Science, High Energy Accelerator Research Organization (KEK) **2** Photon Factory, Institute of Materials Structure Science, High Energy Accelerator Research Organization (KEK) **3** Life Science Research Infrastructure Group, R&D of Technology and Systems for Synchrotron Radiation Applications Division, RIKEN SPring-8 Center ¶ Corresponding author

DOI: 10.xxxxxx/draft

Software

- Review
- Repository
- Archive

Editor:

Submitted: 21 October 2025
Published: unpublished

Summary

Molass Library is a modern, open-source Python package designed for the analysis of SEC-SAXS (Small-Angle X-ray Scattering coupled with Size Exclusion Chromatography) experimental data. It represents a comprehensive rewrite of the original MOLASS tool (Yonezawa et al., 2023), currently hosted at the Photon Factory and SPring-8, Japan. By leveraging the Python ecosystem and supporting interactive scripting in Jupyter notebooks, Molass Library provides enhanced flexibility, reproducibility, and extensibility compared to its predecessor.

A typical SEC-SAXS experiment involves two interconnected processes:

- **SEC** – Size Exclusion Chromatography
- **SAXS** – Small-Angle X-ray Scattering

Effective analysis requires seamless integration of both domains, which Molass Library facilitates through a unified, scriptable workflow.

Figure 1: Logo of Molass Library designed by K. Yatabe

Statement of Need

Analysis of SEC-SAXS data is inherently multi-step and complex. A typical workflow includes:

1. Circular (Azimuthal) averaging
2. Background subtraction
3. Trimming of data
4. Baseline correction
5. Low rank factorization
6. Radius of gyration (R_g) estimation – Guinier plot (Guinier, 1939)
7. Folding state estimation – Kratky plot (Kratky, 1963)
8. Electron density calculation

Molass Library currently implements steps 3–7. For steps 1 and 2, users may employ SAngler (Shimizu et al., 2016) or device-specific software, while for step 8, DENSS (Grant, 2018) is

recommended. Although alternative tools exist, such as the proprietary program suite ATSAS (Manalastas-Cantos et al., 2021) and the open-source program BioXTAS RAW (Hopkins, 2024), Molass Library distinguishes itself by providing an open, scriptable, and modular platform. This design empowers researchers to flexibly tailor and extend their analysis pipelines within the Python ecosystem, thereby enhancing both reproducibility and adaptability.

Notable package dependencies

Molass Library is built on robust scientific Python libraries, including NumPy, SciPy, and Matplotlib. It further integrates:

- **pybaselines** (Erb, n.d.) for advanced baseline correction
- **ruptures** (Truong et al., 2020) for change point detection
- **scipy.signal.find_peaks** for peak recognition

By adopting these well-maintained packages, Molass Library reduces custom code and enhances reliability. The transition from a GUI-based workflow (previously using Tkinter) to Jupyter-based scripting further streamlines reproducibility and collaboration.

Theoretical Foundation

A central feature of Molass Library is its implementation of **low rank factorization** using elution curve models, enabling decomposition of overlapping chromatographic peaks, a common challenge in SEC-SAXS analysis. The decomposition is formulated as:

$$M = P \cdot C \qquad (1)$$

where:

- M: measured data matrix
- P: matrix of component scattering curves
- C: matrix of component elution curves

The optimal solution in the least-squares sense is given by:

$$P = M \cdot C^{+} \qquad (2)$$

where C^{+} denotes the Moore-Penrose pseudoinverse (Penrose, 1955, 1956). This approach enables robust separation of components, even in the presence of overlapping peaks, provided suitable constraints are applied. Moreover, interparticle interference effects can be separated by adding row vectors based on the quadratic approximation derived from SAXS theory.

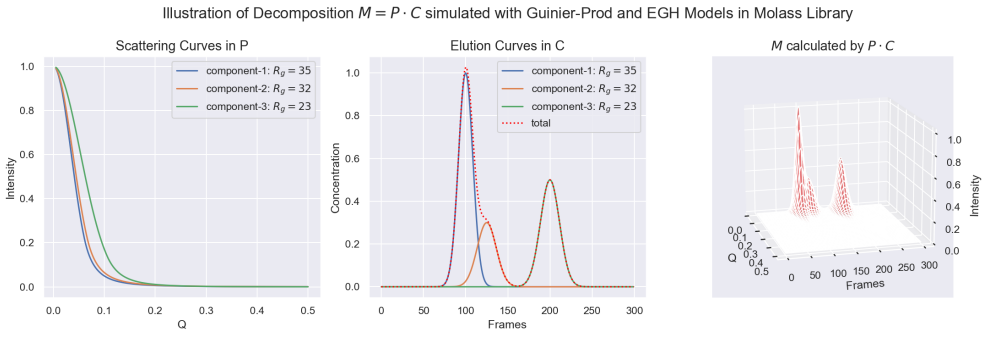

Figure 2: Illustration of decomposition using simulated data

Elution Curve Modeling

To address underdeterminedness and enhance interpretability, Molass Library incorporates several established elution curve models:

- **EGH**: Exponential Gaussian Hybrid (Lan & Jorgenson, 2001)
- **SDM**: Stochastic Dispersive Model (Felinger et al., 1999)
- **EDM**: Equilibrium Dispersive Model (Ur Rehman et al., 2021)

These models allow users to impose domain-specific constraints, thereby enhancing both the accuracy and the physical relevance of the chromatographic peak decomposition.

Availability and Documentation

Molass Library is freely available under an open-source license at https://github.com/biosaxs-dev/molass-library. Comprehensive documentation, including tutorials and theoretical background, is provided at Molass Essence.

Acknowledgements

This work was supported by JSPS KAKENHI Grant Number JP25K07250, and partially by Research Support Project for Life Science and Drug Discovery (Basis for Supporting Innovative Drug Discovery and Life Science Research (BINDS)) from AMED under Grant Number JP25ama121001.

References

Erb, D. (n.d.). *pybaselines: A Python library of algorithms for the baseline correction of experimental data*. https://doi.org/10.5281/zenodo.5608581

Felinger, A., Cavazzini, A., Remelli, M., & Dondi, F. (1999). Stochastic—dispersive theory of chromatography. *Analytical Chemistry*, *71*(20), 4472–4479. https://doi.org/10.1021/ac990412u

Grant, T. D. (2018). Ab initio electron density determination directly from solution scattering data. *Nature Methods*, *15*(3), 191–193. https://doi.org/10.1038/nmeth.4581

Guinier, A. (1939). La diffraction des rayons x aux très petits angles : Application à l'étude de phénomènes ultramicroscopiques. *Ann. Phys.*, *11*(12), 161–237. https://doi.org/10.1051/anphys/193911120161

Hopkins, J. B. (2024). *BioXTAS RAW 2*: new developments for a free open-source program for small-angle scattering data reduction and analysis. *Journal of Applied Crystallography*, *57*(1), 194–208. https://doi.org/10.1107/S1600576723011019

Kratky, O. (1963). X-ray small angle scattering with substances of biological interest in diluted solutions. *Progress in Biophysics and Molecular Biology*, *13*, 105–IN1. https://doi.org/https://doi.org/10.1016/S0079-6107(63)80015-2

Lan, K., & Jorgenson, J. W. (2001). A hybrid of exponential and gaussian functions as a simple model of asymmetric chromatographic peaks. *Journal of Chromatography A*, *915*(1), 1–13. https://doi.org/https://doi.org/10.1016/S0021-9673(01)00594-5

Manalastas-Cantos, K., Konarev, P. V., Hajizadeh, N. R., Kikhney, A. G., Petoukhov, M. V., Molodenskiy, D. S., Panjkovich, A., Mertens, H. D. T., Gruzinov, A., Borges, C., Jeffries, C. M., Svergun, D. I., & Franke, D. (2021). *ATSAS 3.0*: expanded functionality and new

tools for small-angle scattering data analysis. *Journal of Applied Crystallography*, *54*(1), 343–355. https://doi.org/10.1107/S1600576720013412

Penrose, R. (1955). A generalized inverse for matrices. *Mathematical Proceedings of the Cambridge Philosophical Society*, *51*(3), 406–413. https://doi.org/10.1017/S0305004100030401

Penrose, R. (1956). On best approximate solutions of linear matrix equations. *Mathematical Proceedings of the Cambridge Philosophical Society*, *52*(1), 17–19. https://doi.org/10.1017/S0305004100030929

Shimizu, N., Yatabe, K., Nagatani, Y., Saijyo, S., Kosuge, T., & Igarashi, N. (2016). Software development for analysis of small-angle x-ray scattering data. *AIP Conference Proceedings*, *1741*(1), 050017. https://doi.org/10.1063/1.4952937

Truong, C., Oudre, L., & Vayatis, N. (2020). Selective review of offline change point detection methods. *Signal Processing*, *167*, 107299. https://doi.org/https://doi.org/10.1016/j.sigpro.2019.107299

Ur Rehman, J., Muneer, A., & Qamar, S. (2021). Analysis of equilibrium dispersive model of liquid chromatography considering a quadratic-type adsorption isotherm. *Thermal Science*, *26*(3), 2069–2080. https://doi.org/10.2298/TSCI201229179U

Yonezawa, K., Takahashi, M., Yatabe, K., Nagatani, Y., & Shimizu, N. (2023). MOLASS: Software for automatic processing of matrix data obtained from small-angle x-ray scattering and UV-visible spectroscopy combined with size-exclusion chromatography. *Biophysics and Physicobiology*, *20*(1), e200001. https://doi.org/10.2142/biophysico.bppb-v20.0001