

# Non-negative matrix factorization for isolating damage-free reflections in macromolecular synchrotron data collection

Sreya Sarkar<sup>a</sup>, Hilary Florian<sup>b</sup>, Youlin Liu<sup>c</sup>, Garth J. Simpson<sup>d\*</sup>

<sup>a</sup>Department of Chemistry, Purdue University, 560 Oval Drive, West Lafayette, IN 47907, [sarkar10@purdue.edu](mailto:sarkar10@purdue.edu)

<sup>b</sup>Department of Chemistry, Purdue University, 560 Oval Drive, West Lafayette, IN 47907, [hflorian@purdue.edu](mailto:hflorian@purdue.edu)

<sup>c</sup>Department of Chemistry, Purdue University, 560 Oval Drive, West Lafayette, IN 47907, [liu1935@purdue.edu](mailto:liu1935@purdue.edu)

<sup>d\*</sup>Department of Chemistry, Purdue University, 560 Oval Drive, West Lafayette, IN 47907, [gsimpson@purdue.edu](mailto:gsimpson@purdue.edu), corresponding author

The damage-free set of reflections in macromolecular crystallography were mathematically isolated by non-negative matrix factorization (NMF). Biological samples undergoing X-ray induced damage under synchrotron X-ray radiation impose a fundamental limit in data collection for crystallography/structural biology [1]. Damage during diffraction data collection can have a significant impact on the quality of the recovered protein structures. Here, we recovered the native X-ray diffraction pattern of tetragonal lysozyme prior to X-ray exposure. NMF has enabled identification of multiple sequential perturbations and isolated the corresponding changes in the reflections from each. NMF is a matrix factorization method that can extract scattered and significant features from large high-dimensional datasets [2]. In NMF, the entire set of dose-varying reflections are cast in a matrix-form as a product of two non-negative matrices, describing the correlated sets of reflections and amplitudes of a small number of pure components. The findings indicate that the unperturbed protein transitions rapidly into multiple sequential components with X-ray exposure corresponding to a dose of 20 MGy under cryogenic conditions. At room temperature, NMF enabled independent isolation of perturbations to the reflections from direct X-ray exposure versus diffusion of molecular radicals generated upon solvent exposure. During room temperature X-ray data collection, the crystals are damaged faster than cryogenic conditions. Application of NMF during the damage mechanism helps in observing the firsthand effects of indirect X-ray radiation around the sample.

## References

- [1] Ravelli, R.B.G. & Garman, E. F. (2006). Radiation damage in macromolecular cryocrystallography, *Current Opinion in Structural Biology*, **volume 16**, Issue 5, 624-629.
- [2] Lee, D.D. & Seung, H.S. (1999). Learning the parts of objects by non-negative matrix factorization, *Nature*, **volume 401**, 788-791.