

USAGE OF PROBABILITY SCORING TO SORT XRPD SCANS OF DYING PROTEIN ON SIMILARITY

Thomas Degen^a, Irene Margiolaki^b, Jonathan P. Wright^b

^a PANalytical B.V., Lelyweg1,7603 EA Almelo, The Netherlands

^b European Synchrotron Radiation Facility (ESRF), Grenoble, France

In a wide range of applications it is required to sort arbitrarily collected XRPD scans into clusters of similar scans. Additionally it is essential to sort these clusters and the scans in the clusters for similarity too.

One rather new application area is the fast measurement of protein powders in capillaries on synchrotron beam lines. The data collection is usually carried out as a series of 5 to 6 short (2 minute) measurements. Then the capillary is translated with respect to the beam exposing a fresh region of unaffected protein sample to it for the next series of scans.

The radiation induced protein decay is not only visible by fading intensities, but also by peak shifts due to varying unit cell parameters. For further analysis like e.g. Rietveld refinements hundreds or even more scans first need to be sorted into classes of similar scans.

We will demonstrate the implementation and usage of probability scoring and cluster analysis in X'Pert HighScore Plus. First clusters of similar HEWL (hen egg-white lysozyme) scans are created and then both these clusters and the scans inside each cluster are sorted on similarity. This approach creates a sorting that shows a perfectly continuous pseudo-decay axis.