Making sense of non–isomorphism

Kay Diederichs

Universität Konstanz

Datasets measured from different crystals are affected by both random and systematic errors. Systematic errors may arise from different properties of the crystals (e.g. composition and conformation of molecules, or their hydration state) or from experimental conditions (e.g. beam fluctuations or detector non-linearities). Datasets differing systematically are called non-isomorphous, and current crystallographic procedures are far from being able to capture and analyse the various sources of systematic error.

Often, a correlation coefficient is used for inter-dataset comparison, but the relations between datasets are not obvious from the matrix of pairwise correlations since the numerical value of the correlation coefficient is lowered by both random and systematic differences. It is therefore desirable to develop methods that can separate random and systematic effects on data.

The talk presents a novel type of analysis of the pairwise correlation coefficients which positions datasets within a low-dimensional space whose axes are associated with the types of systematic differences between the datasets (Diederichs, K. (2017). Acta Cryst. D73, 286-293). This dimensionality reduction can not only be used for classification purposes (e.g. to find out which datasets are so similar in their properties that they can be merged), but can also derive dataset relations on a continuous scale, and directly relates to previous work [Karplus, P. A. & Diederichs, K. (2012). Science, 336, 10301033] which introduced CC1/2 for describing the precision of crystallographic data. This novel analysis has numerous applications in Structural Biology, but also in other fields.