

MAPPING POLYMORPHS BY COMBINING CLUSTER ANALYSIS AND DIFFRACTION DATABASES

Timothy G. Fawcett, Soorya N. Kabekkodu

International Centre for Diffraction Data
Newtown Square, PA, 19073

Because of their commercial importance, high impact drugs are often analyzed by diffraction methods by multiple research laboratories around the globe within a few years of commercial release. Due to the wide diversity in global publishing forums, and the variety of experimental methods used, studies by various laboratories are often difficult to obtain and/or compare. Cluster analysis is a technique by which multiple data sets from a wide range of laboratories can be compared for their similarities and dissimilarities. Commercial systems are now on the market that have adapted cluster analysis for the specific data issues encountered in typical X-ray diffraction analyses. These tools are seeing increased application in high throughput and combinatorial analyses where large collections of data need to be rapidly compared. In this study we will demonstrate how cluster analysis can be used to map polymorph chemistry.

Coincident with the development of cluster analysis tools has been the development of a comprehensive database of diffraction data of pharmaceuticals, excipients, organic and organometallic materials. Release 2005 PDF-4 Organics contains 263,065 extensively edited data sets collected from a variety of international publications and database organizations. The database combines powder and single crystal x-ray analyses with diffraction, bibliographic, unit cell data, and select physical property information. Since every entry has a digitized diffraction pattern these patterns can be grouped and “clustered” to map out the chemistry of a drug of interest. The data standardization processes that the ICDD applies to global data sources greatly facilitates the cluster analysis process.

We have studied the global data on two high impact drugs, the AIDS drug AZT, and the anticonvulsant Carbamazepine. The combined use of the database and cluster analyses allows the user to relate previously unrelated studies. The total data in the database, which includes formula, nomenclature, quality analyses, authors and editors comments can now be used to “map” the cluster analysis and interpret the phase chemistry. Subtle changes in the diffraction patterns relating to polymorphic chemistry are readily apparent. The clustering allows the user to assign polymorphs (or other structural forms) within a cluster that was not identified by the original reference citation, providing new interpretation and analysis.

