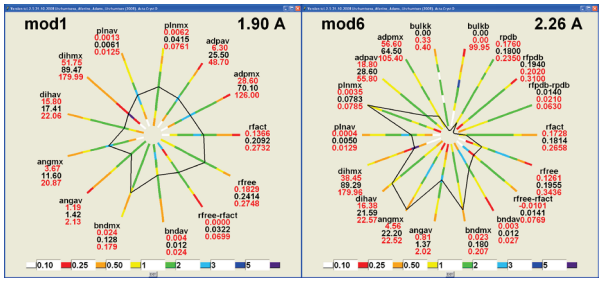
characteristic is shown by a point at its own ruler, and the rulers are plotted together as a set of lines with the same origin, forming a hub and spokes. The points for a given model marked on these lines are connected to form a polygon. A polygon strongly compressed or dilated along some axes reveals unusually low or high values of corresponding characteristics. Different parts of the rulers are colored differently to reflect the frequency (red color for a low frequency, blue for a high frequency) with which the corresponding values are observed in a reference set of structures determined previously. Polygon vertices in 'red zones' indicate parameters which lie outside typical values. The reference set of structures can be selected by the resolution, by their size of structures or by other characteristics. The list of model characteristics to be shown in the polygon is also variable. In particular, in addition to (or instead of) the average values of distortion of stereochemical parameters it may include their maximal values to indicate local problems if they exist. Both the stand-alone Tcl/tk version of the program for macromolecules and the python version incorporated into *PHENIX* [2] are available. As an extra control tool and independently of the POLYGON, the typical values for the *R*- and $R_{free}$-factors and for their difference at a given resolution can be obtained as linear functions of the logarithm of the resolution [3].

[1] Urzhumtseva, L.; Afonine, P.V.; Adams, P.D.; Urzhumtsev, A. *Acta Cryst.* D **65**, 2009, 297-300. [2] Adams, P.D.; Afonine, P.V.; Buncóczi, G; Chen, V.B.; Davis, I. W.; Echols, N.; Headd, J. J.; Hung, L.-W.; Kapral, G. J.; Grosse-Kunstleve, R. W.; McCoy, A. J.; Moriarty, N. W.; Oeffner, R.; Read, R. J.; Richardson, D. C.; Richardson, J. S.; Terwilliger, T. C.; Zwart, P. H. *Acta Cryst.* D **66**, 2010, 213-221. [3] Urzhumtsev, A.; Afonine, P.V.; Adams, P.D. *Acta Cryst.* D **65**, 2009, 1283-1291.

**Keywords: model validation, graphics, computer analysis**




## FA5-MS44-P05

**Self-organizing Crystal Structure.** Mahendhran Arumugam[a], Georg Roth[b], Heike Emmerich[a], Stefaan Cottenier[c]. [a]*Institute of Minerals Engineering, RWTH Aachen University, Germany.* [b]*Institute of Crystallography, RWTH Aachen University, Germany.* [c]*Center for Molecular Modeling, Ghent University, Belgium.*
E-mail: mahi@ghi.rwth-aachen.de

We introduce a self-organizing classification method to classify the structure types of Silica polymorphs using scaled powder diffraction intensities and the Kohonen neural network. The classification results show not only that the isopointal and isoconfigurational structure types are automatically recognized and classified, but also that the acquired classification knowledge in the neural network can be used for performing the classification of yet unclassified or newly determined crystal structures. Additionally, we show the possibility of classifying structures irrespective of their chemical composition, by replacing all atoms by unit scatterers in a combined group of SiO2, GeO2, GaPO4 and AlPO4 polymorphs. The developed method is believed to be applicable to arbitrary types of inorganic crystal structures.

**Keywords: classification of crystal structures, neural networks, powder diffraction data.**

## FA5-MS44-P06

**Advanced Shape Descriptors for Identification of Ligands in Electron Density.** Ciaran Carolan, Gerrit Langer, Victor Lamzin, *European Molecular Biology Laboratory, Hamburg Outstation, Notkestr. 85, 22607 Hamburg, Germany.*
E-mail: ciaran.carolan@embl-hamburg.de

Numerous descriptors have been developed in fields as diverse as computer graphics, physics and biology that allow the comparison of the shapes of complex bodies. Many of these are moments-based descriptors that encode the shape of an object as a series of numbers that contain variable levels of information and allow reconstruction of the object to varying degrees of accuracy. As these moments can be rapidly calculated and quickly compared, they appear to be especially suitable for parametrising a cluster of electron density and identifying ligands which best fit the shape of that cluster. They may also be useful for building the identified ligands into a macromolecular model. We have tested various methodologies based on shape description, including third order moment invariants, spherical harmonic moments, Zernike moments and moments of inter-atomic distance matrices, in order to identify which might be most suitable for the convenient identification of ligands in electron density. In essence, after the shape descriptors for the mystery electron density are computed, they are slid through the database of molecular shapes which includes the ligand molecules of interest present in a variety of alternative feasible conformations. The top hits are ranked based on shape similarities as well as predicted binding energy in the protein as assessed using typical docking scores. Excellent results have been obtained to date using several of the methods noted above. In order to develop the methods further, we have begun to investigate the possibility of incorporating protein residue and charge information into the search templates where this is possible. We are also examining whether the developed methods may be applied to the characterization and analysis of electron density isosurfaces of protein clefts and gorges. By screening large databases of potential drug molecules, such as the ZINC database, it should be possible to identify new leads for drug development even in the absence of crystallographic data for that ligand in the protein. All developed methods will be applicable to drug development research, allowing for the soaking of multiple ligands into a protein and for the subsequent automated identification of the ligands binding at the site of interest, as well as for high-throughput computational drug screening.

**Keywords: ligand recognition, model building, methods macromolecular crystallography**