

From Metric Properties to Unit Cell Databases

Lawrence C. Andrews¹ and Herbert J. Bernstein²

¹Kirkland, WA, USA, larry6640995@gmail.com

²Rochester Institute of Technology, c/o NSLS-II, Upton, NY, USA

For centuries matching experimental data to known information about crystal morphology has been a well-established approach to help identify or rule out possible candidate substances. Careful matching of unit cell data is of increasing importance is serial crystallography, both at XFELs and at synchrotrons.

Steno (1669) noticed the constancy of interfacial angles of crystals. Federov (1911) prepared the first large compilation of the angles for 10,000 materials. The Barker Index (1952) listed data for a large number of crystals. Following the introduction of xrays in crystallography, Wyckoff collected tables of unit cell data (beginning in 1923). Unit cell data has the advantage that it doesn't depend on crystal perfection or form. The first online database of unit cell data was part of the CIS (Chemical Information System supported by the EPA and NIH) starting in 1978.

Early compilations were simply lists of parameters. They depended on standard presentations of data. What was needed was a simple measure of the “distance” between two lattices. The first approach was the V7 metric (Andrews, Bernstein, Pelletier, 1980) used in the CIS. This metric was independently discovered at least twice. The next metric proposed was the G6 metric (Andrews and Bernstein, 1988), which improved on several problems with the V7 metric. Finally, the BGAOL metric (Andrews and Bernstein, 2014) was created. Implemented in the function NCDIST, the BGAOL metric provides a general, completely-linear algorithm for the distance between lattices, allowing the most robust databases to date.

Faster, more-parallel computers allow the use of the most effective cell parameter search techniques in time commensurate with modern data collection techniques. It is no longer necessary to settle for search techniques that generate both false negatives and excessive false positives.