Topological information in crystallography and crystal chemistry: a route to knowledge databases

Vladislav A. Blatov[a,b]

[a]Samara Center for Theoretical Materials Science (SCTMS), Samara University, Ac. Pavlov St. 1, Samara 443011, Russia

[b]School of Materials Science and Engineering, Northwestern Polytechnical University, Xi'an, Shaanxi 710072, People's Republic of China

At present a huge amount of experimental data on crystal structures is collected in world-wide crystallographic databases, a total of more than one million entries. This is the primary information obtained directly in X-ray or neutron diffraction experiment in the form of space group, unit cell parameters, atomic coordinates, etc. It is now being extended with the data on physical properties of crystalline substances, but these data are also just a result of measurement. At the same time, the primary information contains a lot of hidden regularities, but we need new tools to reveal these latent knowledge. With these tools we will be able to transform the crystallographic databases into knowledge databases containing correlations and rules, which then could be used for prediction of new substances and materials.

In crystal chemistry, such correlations usually follow the general form 'chemical composition – structure – physical property', and the middle term is crucial in this sequence. Indeed, if chemical composition and physical property can be easily and strictly defined, the terms 'crystal structure' or 'local structure' require a large number of descriptors. Thus we need to invent as many robust descriptors derived from structural data as possible. Here we consider one of important groups of such descriptors, the topological ones. They can be derived from the primary crystallographic data and supply them with the information on the structure connectivity, which is one of the main and natural chemical and crystallochemical characteristics. Another important feature of topological descriptors is that they allow one to formalize the classification of crystal structures and search for structural correlations.

The storage format of topological information on crystal structures has already been developed to some extent: there exist tailored topological databases like RCSR, EPINET, or ToposPro collections. However, this format should be adjusted to conventional crystallographic formats such as CIF. This is required to solve the first task on the route to knowledge databases in crystallography and crystal chemistry – to unite the primary (crystallographic) and secondary (topological) data within the same universal storage. At the next step, the tools for revealing correlations 'chemical composition – structure – physical property' should be developed and the correlations should be written to the universal storage transforming it to a knowledge database.

This route is discussed with a special attention to the PDF databases as a possible platform for uniting crystallographic and topological knowledge.