Abstact

Crystallography Open Database (COD, http://www.crystallography.net) is the largest to date curated open-access collection of small to medium sized unit cell crystal structures [4, 3]. Over 11 years of development, COD has accumulated over 1.4 million structures. COD has an automated data submission Web site, performs routine automatic quality checks on all incoming structures and is now recommended as a database for crystallographic deposition by several scientific journals. To facilitate automatic use and discovery of COD data, and to increase the usefulness of our database for chemists, two steps were undertaken. COD was supplemented with software and data from the CrystalEye data aggregator [2]. The new software permits extracting chemical data and presenting them as structural formulae, unique moieties and chemically significant fragments. We have also implemented search of crystal structures by the structural chemical formulae of the target compounds. To facilitate data curation, a new software platform for data review is being developed to automatically detect unusual structures and collect expert opinions from qualified human reviewers concerning credibility of such structures.

Platform for data reviews

“Unusual” does not necessarily mean “wrong”
- Extraordinary claims require extraordinary evidence;
- Most statistically unusual structures will be forwarded to a COD reviewer Web forum for verification;
- Convincing evidence confirms validity of unusual structures;
- The set of usual and verified unusual structures should be used for reliable scientific inferences;
- Unusual structures require special attention.

Cross-linking the COD with Open Data
- RDF (Resource Description Framework) descriptions are provided for each database entry:
  - example: http://www.crystallography.net/1516168.rdf
- Cross-links are made with ChemSpider, PubChem, AMCSD and MPOD;
- Links to Wikipedia and DrugBank are provided.

Acknowledgements

This research was funded by a grant (No. MIP-025/2013) from the Research Council of Lithuania.

Bibliography
[5] Vilnius University, Institute of Biotechnology, Department of Protein-DNA Interactions, Lithuanian Academy of Sciences, Le Mans, France; *Vilnius University, Faculty of Mathematics and Informatics, Department of Mathematical Computer Science, Vilnius, Lithuania; "Vilnius University, Faculty of Mathematics and Informatics, Department of Software Engineering, Vilnius, Lithuania; "Université du Maine, Laboratoire des Oxydes et Fluorures, Université du Maine, Faculté des Sciences, Le Mans, France; *Université de Caen-Normandie, CRISMA-ENSICAEN, Caen, F-14050 Caen, France; † University of Trento, Department of Materials Engineering, Trento, Italy; ‡University of Greenwich, Facultad de Ciencias, Departamento de Química Inorgánica, Granada, Spain; §Portland State University, Department of Physics, Portland, USA; ¶University of Cambridge, Department of Chemistry, Cambridge, United Kingdom; ||University of Arizona, Department of Geosciences, Tucson, USA.

Robert T. Downs, Saulė Girdžiauskaite and Saulius Gražulis.

Cross-link the COD with Open Data

Platform for data review

“Unusual” does not necessarily mean “wrong”
- Extraordinary claims require extraordinary evidence;
- Most statistically unusual structures will be forwarded to a COD reviewer Web forum for verification;
- Convincing evidence confirms validity of unusual structures.
- The set of usual and verified unusual structures should be used for reliable scientific inferences;
- Unusual structures require special attention.

Cross-linking the COD with Open Data
- RDF (Resource Description Framework) descriptions are provided for each database entry:
  - example: http://www.crystallography.net/1516168.rdf
- Cross-links are made with ChemSpider, PubChem, AMCSD and MPOD;
- Links to Wikipedia and DrugBank are provided.

Acknowledgements

This research was funded by a grant (No. MIP-025/2013) from the Research Council of Lithuania.

Bibliography
[5] Vilnius University, Institute of Biotechnology, Department of Protein-DNA Interactions, Lithuanian Academy of Sciences, Le Mans, France; *Vilnius University, Faculty of Mathematics and Informatics, Department of Mathematical Computer Science, Vilnius, Lithuania; "Vilnius University, Faculty of Mathematics and Informatics, Department of Software Engineering, Vilnius, Lithuania; "Université du Maine, Laboratoire des Oxydes et Fluorures, Université du Maine, Faculté des Sciences, Le Mans, France; *Université de Caen-Normandie, CRISMA-ENSICAEN, Caen, F-14050 Caen, France; † University of Trento, Department of Materials Engineering, Trento, Italy; ‡University of Greenwich, Facultad de Ciencias, Departamento de Química Inorgánica, Granada, Spain; §Portland State University, Department of Physics, Portland, USA; ¶University of Cambridge, Department of Chemistry, Cambridge, United Kingdom; ||University of Arizona, Department of Geosciences, Tucson, USA.

Robert T. Downs, Saulė Girdžiauskaite and Saulius Gražulis.