MS.27.2

Acta Cryst. (2011) A**67,** C72

The wwPDB and Future perspectives in sharing macromolecular data

Helen M. Berman,[a] Gerard J. Kleywegt,[b] John L. Markley,[c] Haruki Nakamura[d] [a]*RCSB PDB, Rutgers University, Piscataway, NJ (US*A). [b]*PDBe, EMBL-EBI, Hinxton, Cambridge, (UK).* [c]*BMRB, University of Wisconsin-Madison, Madison, WI (USA).* [d]*PDBj, Institute of Protein Research, Osaka University, Osaka, (Japan).* E-mail: berman@rcsb.rutgers.edu

This fall, a symposium will be held to commemorate the 40th anniversary of the Protein Data Bank (PDB) archive. It was established in 1971 with 7 structures; today it contains more than 72,000.

With an eye to future decades, the Worldwide Protein Data Bank (wwPDB) is committed to using the highest standards of curation and processing for experimentally-determined 3D biomolecular structure data. Several wwPDB initiatives are underway to handle the ever-growing number and complexity of PDB data. These include the development of a *Common Deposition and Annotation Tool* that will produce highly curated structural data; convening method-specific *Validation Task Forces* that are developing recommendations on additional validation that should be performed by the wwPDB, and the creation of a PDB working format that can support large structures, and structures determined by new and hybrid methods.

The wwPDB members are: RCSB PDB (supported by NSF, NIGMS, DOE, NLM, NCI, NINDS and NIDDK), PDBe (EMBL-EBI, Wellcome Trust, BBSRC, NIGMS, and EU), PDBj (NBDC-JST) and BMRB (NLM).

Keywords: database, macromolecule, chemical components

MS.27.3

Acta Cryst. (2011) A**67,** C72

Possible futures for small molecule crystal structure archives?

Colin R Groom, Ian Bruno, Jason Cole and John Leibeschuetz *Cambridge Crystallographic Data Centre, 12 Union Road, Cambridge (UK). CB2 1EZ.* E-mail: edir@ccdc.cam.ac.uk

The Cambridge Crystallographic Data Centre must respond to the opportunities and challenges facing it in such a way that it can continue to provide the Cambridge Structural Database. This has served as the worlds repository for the 3D structures of organic and organometallic compounds for over 45 years. It provides the community with experimental structures of well over half a million organic and organometallic compounds. All are expertly curated by editorial staff so as to facilitate reliable and sophisticated retrieval, visualisation and analysis by software that the centre also develops.

However, the confluence of massive technological change, rapid evolution of the pharmaceutical industry, altered funding patterns of granting agencies and the changed expectations of users in terms of data access might be coming together to create a 'perfect storm' for crystallographic database providers. Alternatively, these factors could be considered as presenting a period of much needed positive turbulence, which will encourage database providers to review their scientific and technical efforts and their business models.

This presentation will both describe the views of the CCDC and most importantly, solicit input from users.

Keywords: CSD, databases

MS.27.4

Acta Cryst. (2011) A**67,** C72

Graphical query interfaces for inorganic crystallographic databases

Hans-Joachim Klein, *Department of Computer Science, Technical Faculty, Christian-Albrechts-Universität Kiel, (Germany).* E-mail: hjk@is.informatik.uni-kiel.de

Searching in crystallographic databases is usually done by providing data such as space-group symbols, crystal names or lists of elements in input fields of a form. An exception is offered by the Cambridge Structural Database which allows searching for patterns of combinations of atoms using a graphical interface [1]. For inorganic crystallographic databases such graphical search options are hard to provide since a large variety of chemical elements and a rich diversity of geometric configurations have to be taken into account. Graphical interfaces of these databases are therefore mostly restricted to the presentation of structures in answers. The availability of graphical query facilities, however, could be very attractive for complementing the standard retrieval options. Compounds having topologies which are similar to a given real or artificial compound could be searched with additional features for comparing geometric similarity. As an example consider the searching for compounds which have layers of vertex-sharing octahedra, pyramids, or squares similar to those found in superconducting cuprates. The specification of elements forming coordination polyhedra of this kind could be combined in a query with the graphical description of appropriate polyhedral layers.

We have realized a graphical query interface based upon a graph-oriented representation of inorganic crystal structures at the level of coordination polyhedra [2]. Queries can be formulated by marking a cluster of polyhedra in a graphical representation of a real crystal or by constructing artificial clusters of polyhedra using an interactive tool box. Answers are structures from a given database which allow to embed the search cluster at the topological level. The embeddings can be ordered with respect to their geometric similarity to the search cluster.

To get a useful measure of similarity, polyhedral clusters may be considered as joint structures permitting motions which are in accordance with the connections between polyhedra, i.e. restricted rotations and hinge motions besides translations. In the current implementation of the interface, clusters are assumed to be non-flexible in the connections and the polyhedra are considered as rigid bodies. For a given search cluster and a set of topologically equivalent substructures, the root mean square of the deviation of the substructures from the search cluster is computed taking the coordinates of the central atoms of the polyhedra as point sets. Results are ranked according to the best fit.

In order to guarantee efficiency also in case of large data sets, a special index form is used. It serves to efficiently determine topologically equivalent clusters in large sets of compounds. The geometry of clusters is analysed only if topological equivalence has been found. This proceeding avoids difficulties which may arise in connection with the definition of similarity when geometry is considered too early.

[1] J.A. Chisholm and S. Motherwell, *J. of Appl. Cryst.* **2004** *37,* 331-334, [2] H.-J. Klein and Ch. Mennerich, Int. *J. on Adv. in Syst. and Managem*. **2009** *2,* 192-203.

Keywords: crystallographic database, polyhedral cluster, similarity search