STRUKTURA 2015

Luhačovice, 22. 6. - 25. 6. 2015

Session I, Structural Databases, Monday, June 22

L1

50 YEAR OF THE CAMBRIDGE STRUCTURE DATABASE OF ORGANIC AND ORGANOMETALLIC COMPOUNDS

J. Hašek

Institute of Biotechnology AS ČR, Vídeňská 1083, 142 20 Praha 4
hasekjh@seznam.cz

The Cambridge Structure Database (CSD) of organic and organometallic compounds celebrates already 50 years from its foundation. The Czech and Slovak scientific community represented by the "Regional Affiliated Center" covers a yearly fee for access to the CSD for non-commercial subjects continuously from 1973. The database is necessary for deep understanding structure chemistry, solid state chemistry, material research, design of advanced materials and for any research where understanding the structure-function relations are important and also for any reliable computation simulation of structure and dynamics of real molecular systems.

The license for 2015 includes more than 0.75 million of experimentally determined organic and organometallic compounds. This database is complemented by the Structure Database of Crystalline Polymer Compounds (POLYBASE-contains more than 400 polymers) and the Database of Protein-Polymer Interactions (DPPI-contains more than 2000 experimentally determined mostly PEG-protein interactions).

In the CSD, the search of structures, calculation of structure parameters, filtration and tabulation of the relevant data is done by program QUEST CSD. Detailed analysis of structure relationships, calculation of the powder diffraction records and visualization of structures in solid phase is routinely done by program MERCURY. The review of experimentally verified interactions between molecules in solid phase is provided by software ISOSTAR, the interactions between macromolecules are analyzed by SUPERSTAR. Information on the CSD system is at www.ccdc.cam.ac.uk. Any installation on computers in the Czech and Slovak republics should be done via registration into the club of the CSD users in frame of the Crystallographic Association (http://www.xray.cz/xray/csca/data/r_form_cz.htm). The instructions for installation of the CSD software can be downloaded during the registration. Help in case of some problems can be found at hasekjh at seznam.cz.

Polymer structure database is inspected by POLYBASE and the structures are analyzed best by MERCURY. The database of protein-polymer interactions is inspected simply by text editor and viewed by program PYMOL (incentive version perfectly fitting all needs is free of charge at https://www.pymol.org).

An immense amount of inspiration how to use the CSD can be found on CCDC pages. Several thousands of papers with scientific studies based on the CSD data and produced by the CSD software sorted by the date of publication are at the address
http://www.ccdc.cam.ac.uk/ResearchAndConsultancy/CCDCResearch/CCDCPublications/Pages/CCDCPublications.aspx
alternatively, you can search for papers by journals or other criteria at https://services.ccdc.cam.ac.uk/webcite/search/.

As far as teaching, the special introduction courses to the CSD for high schools with chemical curricula and for universities can be found e.g. at the School of Chemistry of the Newcastle University at http://www.ncl.ac.uk/chemistry/outreach/resources/ccdc/

Parallel to the 50th anniversary of CCDC, the related database Protein Databank (PDB), celebrate this year 45 from its foundation. It contains more than 100 thousands experimentally determined structures of bio-macromolecules mostly by protein crystallography. Great advantage of protein crystallography is that there is no limitation on the complexity of molecular system. Nowadays, it is relatively easy to determine structure of complexes of several tens proteins and to observe intermolecular interactions in their complexity. More than 90 % of structures were determined by X-ray diffraction. About 9 % structures were determined by NMR and less than 1 % structures by other methods.

The topic is supported by BIOCEV CZ.1.05/1.1.00/02.0109 from the ERDF, and LG14009.

L2

STRUCTURAL DATABASES OF BIOMOLECULES, THE PDB AND NDB

Bohdan Schneider

Institute of Biotechnology, Czech Academy of Sciences, Videnska 1083, CZ-142 20 Prague, Czech Republic

THE PROTEIN DATA BANK, PDB [1] (http://www.pdb.org/), is the main publically accessible resource for structural biology and the central archive for biomolecular three-dimensional structures determined experimentally by X-ray crystallography, NMR techniques, electron (cryo) microscopy, and neutron diffraction. It is jointly operated by the Worldwide Protein Data Bank (wwPDB, http://www.wwpdb.org/). The main website (http://www.rcsb.org) offers several main services: deposition of new structures, search of the archive, analysis of structures, and others. The PDB has developed many tools for the deposition and validation of structures, all easily available from the web, and serves as the main hub for deposition, data processing and distribution of the biomolecules. The archive can be queried. The queried structures can be downloaded in the historical PDB format or in mmCIF format [2], individually inspected ("browsed"), and their user-selected properties arranged into tabular reports. The archive contains both coordinates and "experimental" files of the deposited structures (structure factors for X-ray, distance constraint files for NMR structures).

The talk will give an overview of the archive, introduce some basic features of the website, suggest effective ways how to query the archive, and how to create an useful report of the queried structures.

NUCLEIC ACID DATABASE, NDB [3] (http://ndbserver.rutgers.edu) was established in 1991 as a resource for experts on nucleic acid structures. It contains both X-ray and NMR structures containing dinucleotide and longer sequences. The core of the NDB is its relational database of primary and derivative data specifically related to structure of nucleic acids. The database offers rich query and reporting capabilities. We will briefly show specific features of NDB.

DATA DISTRIBUTION. Coordinate files, database reports, software, and other resources are freely available from the web pages of both databases.

ACKNOWLEDGEMENTS. *This study was supported by BIOCEV CZ.1.05/1.1.00/ 02.0109 from the ERDF and by institutional grant to Institute of Biotechnology, CAS (RVO: 86 652 036).*

1. Berman H.M., Battistuz T., Bhat T.N., Bluhm W.F., Bourne P.E., Burkhardt K., Feng Z., Gilliland G.L., Iype L., Jain S., Fagan P., Marvin J., Padilla D., Ravichandran V., Schneider B., Thanki N., Weissig H., Westbrook J.D., Zardecki, C. (2002): The Protein Data Bank. *Acta Crystallogr D*, **58**, 899-907.
2. Bourne P.E., Berman H.M., Watenpaugh K., Westbrook J.D., Fitzgerald P.M.D. (1997): The macromolecular crystallographic information file (mmCIF). ***Methods Enzymol.*** **277**, 571–590.
3. Berman H.M., Olson W.K., Beveridge D.L., Westbrook J., Gelbin A., Demeny T., Hsieh S.-H., Srinivasan A.R., Schneider B. (1992): The Nucleic Acid Database—a comprehensive relational database of three-dimensional structures of nucleic acids. ***Biophys. J.*** **63**,751–759.

L3

USE OF CSD IN SLOVAK UNIVERSITIES

Ján Moncoľ

Institute of Inorganic Chemistry, Faculty of Chemical and Food Technology, Slovak University of Technology in Bratislava, SK-81237 Bratislava, Slovakia
jan.moncol@stuba.sk