Fiftieth Anniversary of the Cambridge Structural Database and Thirty Years of Its Use in Croatia

DOI: 10.15255/KUI.2015.025
KUI-23/2015
Review
Received May 13, 2015
Accepted June 12, 2015

*B. Kojić-Prodić** and *K. Molčanov*

Rudjer Bošković Institute, Bijenička c. 54, POB 180,
HR-10 002 Zagreb, Croatia

Dedicated to the memory of
Dr. Frank H. Allen, FRSC CChem, 1944 – 2014

‖ **Abstract**

This article is dedicated to the memory of Dr. F. H. Allen and the 50th anniversary of the Cambridge Crystallographic Data Centre (CCDC); the world-renowned centre for deposition and control of crystallographic data including atomic coordinates that define the three-dimensional structures of organic molecules and metal complexes containing organic ligands. The mission exposed at the web site (http://www.ccdc.cam.ac.uk) is clearly stated: "The Cambridge Crystallographic Data Centre (CCDC) is dedicated to the advancement of chemistry and crystallography for the public benefit through providing high quality information, software and services." The Cambridge Structural Database (CSD), one among the first established electronic databases, nowadays is one of the most significant crystallographic databases in the world. In the International Year of Crystallography 2014, the CSD announced in December over 750,000 deposited structures. The use of the extensive and rapidly growing database needs support of sophisticated and efficient software for checking, searching, analysing, and visualising structural data. The seminal role of the CSD in researches related to crystallography, chemistry, materials science, solid state physics and chemistry, (bio)technology, life sciences, and pharmacology is widely known. The important issues of the CCDC are the accuracy of deposited data and development of software for checking the data. Therefore, the Crystallographic Information File (CIF) is introduced as the standard text file format for representing crystallographic information. Among the most important software for users is *ConQuest*, which enables searching all the CSD information fields, and the web implementation *WebCSD* software. *Mercury* is available for visualisation of crystal structures and crystal morphology including intra- and intermolecular interactions with graph-set notations of hydrogen bonds, and analysis of geometrical parameters. The CCDC gives even more options to the users developing sophisticated software such as *GOLD*, *IsoStar* and *SuperStar*, *DASH*, and extensive knowledge electronic libraries such as *Mogul* and *Relibase*. The CCDC released the new facility – *Mercury*'s *Solid Form* module. Such demanding projects require a highly competent team of experts with a scientific approach based on the long tradition in crystallography, modelling and informatics. The Cambridge Structural Database and diversified software and searching engines are useful tools in research and teaching. The use of electronic media and computer graphics makes "data mining" very efficient and useful, but also aesthetically appealing due to the molecular architecture. One can expect even more advanced approaches using cloud computing and 'Big Data' management; merging data from related databases will enable to recognize hidden molecular and crystal properties and information that could bring new important knowledge.
Since 1985, the CSD has been available to users in Croatia. The use of the CSD in Croatia is illustrated by a few examples performed and published by the presenting authors and colleagues.

‖ **Keywords**

Crystallographic database, Cambridge Structural Database (CSD), crystal structure, molecular structure

A brief history of CSD

In the last three years there have been held brilliant, centennial jubilees related to X-ray crystallography. In 1912, Max von Laue announced diffraction of X-rays on a crystal and his discovery marked the birth of a new discipline – X-ray crystallography. The Nobel Prize was awarded for Laue's discovery in 1914 and one hundred years later, the year 2014 was declared as the International Year of Crystallography. In 1912 and 1913, the Braggs solved the first crystal structures by diffraction of X-rays. In 2015, there is the golden jubilee of the *Cambridge Crystallographic Data Centre* (CCDC), established to organize a database in order to save and check crystallographic data and atomic coordinates for organic compounds and metal complexes with organic ligands. The Cambridge Crystallographic Data Centre was established in the Department of Organic, Inorganic and Theoretical Chemistry of the University of Cambridge in 1965, where Olga Kennard had been invited to form an X-ray crystallography group. J. D. Bernal, the physicist and crystallographer recognised the importance of information and data deposition, particularly in the growing field of crystallography. Professor O. Kennard, FRS, who worked with Bernal at Birkbeck College, the University of London, obtained his support to create a Crystallographic Data Centre. The CCDC began to collect published bibliographic, chemical and crystal structure data for small molecules studied by X-ray and neutron diffraction data. At that time, computers and software were in their infancies and

* Corresponding author: Dr. sc. Biserka Kojić-Prodić
e-mail: biserka.kojic-prodic@irb.hr

the CCDC had a long way to reach its present efficiency and high significance. On the historic route of the development of the CCDC there had been very competent and dedicated experts. O. Kennard retired as Director in 1997 and was succeeded by Dr. D. Hartley (1997–2002) and Dr. F. H. Allen (2002–2008). C. R. Groom was appointed as the Executive Director in 2008.

More information and details can be found at the web site (http://www.ccdc.cam.ac.uk) and in the literature.[1–12 and references therein]

Deposition, retrieval and analysis of data from the Cambridge Structural Database (CSD)

The tremendous progress of X-ray crystallography is based on the use of powerful computers utilised in on-line experiments, such as robotic machines for crystallization, powerful, high-brilliance radiation sources, controlled data collection and processing, interpreting Fourier maps, refinement of experimentally derived model of the structure, visualization of the three-dimensional structure and modelling using computer graphics. Thus, over the years the enormous growth of deposited data (Fig. 1) has been recorded.

The members of the CCDC had a wishing thought to reach 750,000 deposited structures in the International Year of Crystallography 2014, which came through in December 2014. It is interesting that the milestone entry was the structure of a co-crystal of vanillic acid and theophylline, the two compounds of practical application in pharmacology and food industry. Vanillic acid is an oxidised form of vanillin, responsible for vanilla flavour, widely used in the food industry and cosmetics. Theophylline, another natural product, found in cocoa beans, is closely related to caffeine. Theophylline is a methylxanthine drug used as a therapy for respiratory diseases, such as chronic obstructive pulmonary disease and asthma. Another interesting aspect is the formation of co-crystals demonstrating one branch of the discipline of crystal engineering, which plays an increasingly important role in the pharmaceutical industry as an approach to enhance the physico-chemical properties of potential drug molecules. The knowledge of this structure, combined with the collection of all the structures in the CSD will be utilised by scientists worldwide to engineer new drug crystal forms. It is remarkable that just this very milestone entry symbolizes the versatile use of the Cambridge Structural Database.

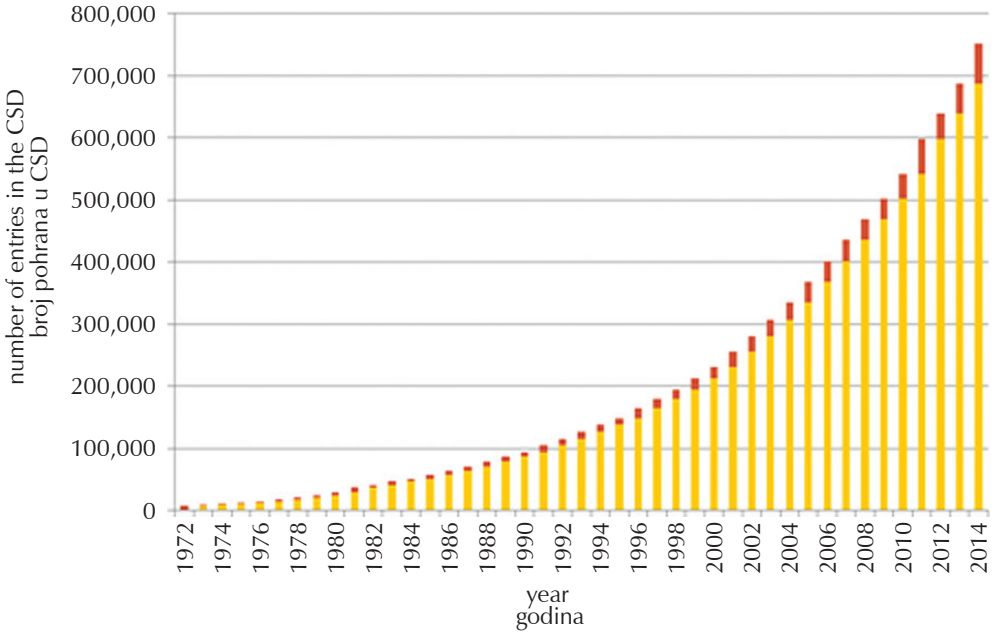


Fig. 1 – The CSD growth since 1972, where the dark (red) segment of each bar indicates the newly added structures, and the light (gold) portion of each bar represents the existing structures (copied from http://www.ccdc.cam.ac.uk)

Slika 1 – Povećanje pohranjenih podataka u CSD-u od 1972., pri čemu crveni (tamniji) dio stupca pokazuje nove pohranjene strukture za svaku godinu, dok se žuti (svjetliji) dio svakog stupca odnosi na već prije pohranjene strukture (preuzeto s http://www.ccdc.cam.ac.uk)

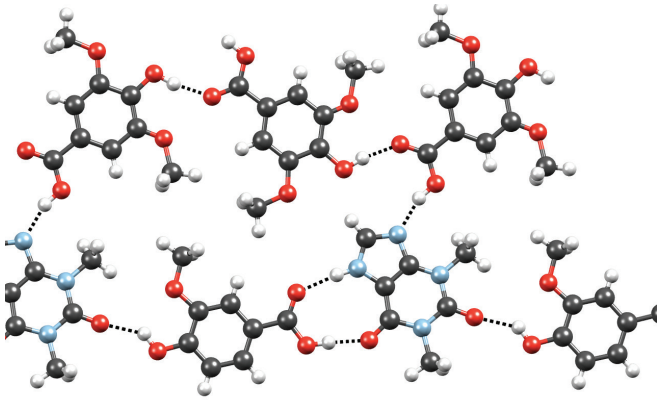


Fig. 2 – CSD's 750,000th structure is of a co-crystal of vanillic acid and theophylline, comprising hydrogen-bonded layers (REF code ZOYBIA), determined by Ayesha Jacobs and Francoise M. Amombo Noa. Ayesha Jacobs, from the Cape Peninsula University of Technology in South Africa said: "This is very exciting news indeed. The CSD is invaluable in our research, to have our structure as the 750,000th one submitted to the CSD and for it to be in the International Year of Crystallography is rather special." (Figure and quotation are copied from http://www.ccdc.cam.ac.uk).

Slika 2 – 750.000. struktura pohranjena u CSD-u je kokristal vanilinske kiseline i teofilina, koji tvori slojevitu strukturu putem vodikovih veza (REF-kod u bazi ZOYBIA). Strukturu su odredili Ayesha Jacobs i Francoise M. Amombo Noa. Ayesha Jacobs s Tehničkog sveučilišta Cape Peninsula, Južna Afrika rekao je: "To je doista uzbuđujuća novost. CSD je nezamjenjiva u našem istraživanju, osim toga iznimno je da je to 750.000. struktura u CSD-u i to u Međunarodnoj godini kristalografije." (Slika i citat preuzeti su s http://www.ccdc.cam.ac.uk).

It would be interesting to learn the number of authors with the highest number of deposited structures in the CSD (Fig. 3).

Fig. 3 – Wordle ('word clouds') of the top 100 authors in the CSD (copied from http://www.ccdc.cam.ac.uk)

Slika 3 – Slagalica imena 100 autora s najvećim brojem pohranjenih struktura u CSD-u (preuzeto s http://www.ccdc.cam.ac.uk)

The crystallographers and other scientists using X-ray or neutron diffraction for elucidating crystal and molecular structures of organic compounds and metal complexes with organic ligands deposit their crystallographic data and atomic coordinates to be included in the Cambridge Structural Database. The purpose of these depositions is to make the crystallographic data and atomic coordinates available to scientists all over the world in order to be used for research, teaching and applications, not only in drug design, but also in other technologies. To obtain structural information, one can search also bibliographic data related to the content of the CSD; those not having access to SciFINDER, browse the CSD for structural information. In addition to storing crystallographic data and atomic coordinates, the CCDC has developed procedures for checking deposited data in order to provide an accurate and reliable source of information. The efficient and sophisticated software provided by the CCDC can be used for the statistical analysis of geometrical parameters of a particular class of structures, for comparative analysis of structural data, for gaining insight into intra- and intermolecular interactions, for packing analysis of polymorphs, and also for structure predictions. There is another very useful application of the CSD; it is possible to retrieve the atomic coordinates (and CIF) for various theoretical calculations, e.g. quantum mechanics calculations and molecular modelling – docking of substrates and inhibitors to study enzymatic reactions.

Software developed by the CCDC to functionalise the use of the Cambridge Structural Database

A big step forward in data acquisition was in the early 1990s with the development of *Crystallographic Information File* (CIF)[8] and its subsequent improvements as the standard for the electronic transmission of crystal structure data performed jointly by the *International Union of Crystallography* (IUCr) and the CCDC.[9,10] They provided a software system for CIF checking by authors, which has reduced dramatically the incidence of errors. The use of CIF has been a great advantage for the authors, the CSD, IUCr journals, and all other journals publishing crystal structures. To optimise the use of CIF further within the CCDC, an internal software was generated to accelerate incoming raw CIF data to the final CSD entries.

A matter of great improvement over the last fifteen years has been *ConQuest*, a program for searching all CSD information fields.[1,2,11] It can combine 2D substructure searches with 3D geometrical constraints to locate hydrogen bonds or other nonbonding interactions (Fig. 4). A Web implementation, *WebCSD*, is also available,[12] which offers fast searching of the CSD using only a standard Internet browser. The CCDC program *Mercury* – the crystal structure visualizer, has been significantly changed and improved.[11,14] Now it includes features enabling complete analyses of geometrical data and crystal packing visualization with intermolecular interactions, hydrogen bond synthons, extended networks and graph set descriptors, and also computed powder diffraction patterns.[1,2 and references therein] A great advantage is a new facility, which enables data that has been retrieved by *ConQuest* for analysis of geometrical parameters to be used within *Mercury*.[14] The most recent novelty, introduced by the CCDC in 2015, is the *Full Interaction Mapping* tool[15] suitable for evaluation of molecular

crystal structures. This functionality can rapidly visualise the interaction preferences of a molecule in a particular conformation, whereas in the three-dimensional packing it can probe how the interaction preferences are satisfied within the crystal lattice. Thus, this module can be used to evaluate and understand relative structure stability. Even more, analyses of particular crystal surfaces can give insight into crystal morphology. This tool is an integral part within the *Solid Form* module of *Mercury* and it was built on the grounds of the current *IsoStar*[16] interaction knowledge-based technology, already existing in the CSD. By improving this technology, it will be possible to use it quantitatively to assess 'hot spots' of interactions to recognise pharmacophores or chemophores.

The CSD System also developed the two extensive knowledge bases *Mogul*[17,18] and *IsoStar.*[16] *Mogul* comprises more than 20 million bond lengths, valence and torsion angles organised into more than 11.5 million chemically searchable distributions, each relating to a specific chemical environment, revealing also data for chemical rings. *Mogul* generates torsional distributions for all rotatable bonds in an input molecule or checks all geometry against mean values from the CSD. It is well-known that three-dimensional molecular conformation is of great impact on efficacious drug / protein interactions and consequently it is important for the effective drug design. In a case when the routine procedures with advanced force fields cannot provide the reliable energy landscape for examined conformation(s), the analysis by *Mogul* can offer data to be used for better parameterisation of the force field(s) or reliable scoring functions. From simple analysis of the molecular conformational space to the complex problems of modelling in drug design and /or enzyme catalysis, *Mogul* can assist in finding a solution. The *IsoStar*[16] knowledge base can be used in the analysis of nonbonding interactions, providing interactive 3D scatterplots revealing the distribution for selected contact groups among 48 of them. This facility is very useful in the analysis of hydrogen bonds and the priority of their formation regarding available donor/acceptor functionalities. More about the use of this library can be found in the references.[1,2,15] The combined analysis of crystallographic data about intermolecular interactions extracted from the CSD and Protein Data Bank can generate maps of interaction – 'hot spots' in protein binding sites or around small molecules using *SuperStar* program. The CCDC offers other products related to *SuperStar,* such as *Relibase* and the advance searching of protein-ligand complexes. *Gold*/*GoldMine* for protein-ligand docking and post-processing virtual screening results are also available. The Cambridge Structural Database System includes *Mogul* and *IsoStar* with easily accessible experimental information on conformations and intermolecular interactions stored in the CSD.

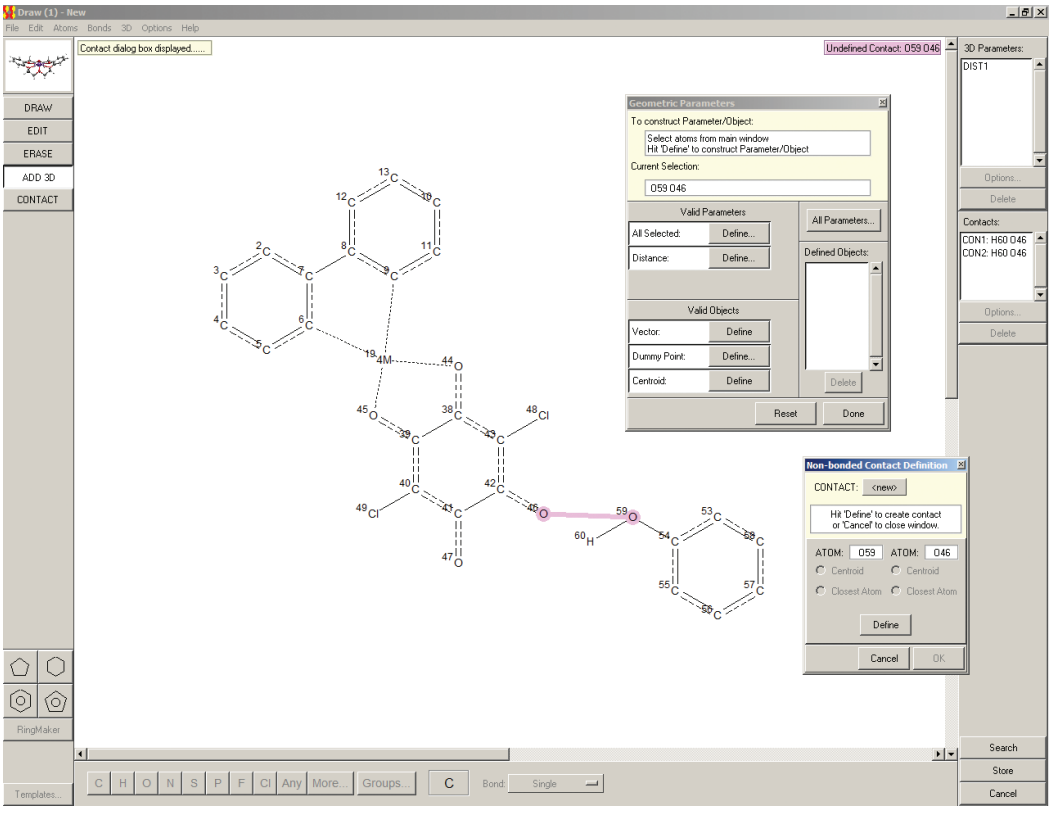


Fig. 4 – Search of CSD using *ConQuest*: Definition of a molecular fragment and an intermolecular contact

Slika 4 – Pretraživanje CSD-a pomoću programa *ConQuest*: definiranje dijela molekule i međumolekularnog kontakta

The use of CSD in Croatia with some examples of its applications in structural and synthetic chemistry and crystallography

Eleven years ago, we used CSD to prove that ester sp^3 oxygen atom can be an acceptor of hydrogen bond.[19] Its proton accepting capability was considered to be very low due to the proximity of strongly electronegative carbonyl group; it is less electron rich than other oxygen atoms and possible proton donors are more likely to be directed towards nearby carbonyl oxygen (which is known to be a strong acceptor). Our survey of the CSD revealed quite a few hydrogen bonds with ester sp^3 oxygen as the acceptor; a statistical analysis assessed its proton acceptor capabilities, which turned out to be comparable to those of ether oxygen. To illustrate growth and development of CSD, we repeated our survey, originally done using the November 2003 version comprising less than 300,000 structures, and the version November 2014 depositing about 750,000 structures.

We searched the CSD for fragments shown in Fig. 5, where $d(\mathrm{H}\cdots\mathrm{Osp^3}) < r_{vdW}(\mathrm{H}) + r_{vdw}(\mathrm{O}) - 0.01$ Å and *D*-H···*A* angle > 90°; donor atom *D* is either O or N. Since we wanted only high-quality data with protons located as accurately as possible, we used additional restrictions to our search: accurately determined structures from single-crystal data (for structures with O atom as a donor $R < 0.05$ and for those with N atom as a donor $R < 0.075$), error-free coordinates, no crystallographic disorder and no polymer structures.

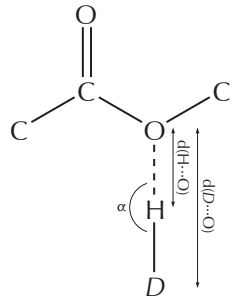


Fig. 5 – Fragment used in our CSD survey with main geometric parameters marked. Donor atom *D* is either O or N.

Slika 5 – Primijenjen motiv za pretraživanje CSD-a s označenim geometrijskim veličinama. Donorski atom *D* je O ili N.

The present survey revealed a considerably greater number of structures with the ester Osp^3 atom involved in a hydrogen bond, 501 (*vs.* 230 in 2003 survey); 306 of them involved O atom as a proton donor (*vs.* 160 in 2003 survey), while in 195 structures the donor atom was N (*vs.* 70 in 2003 survey). Distribution of *D*···A distances (Fig. 6) and *D*–H···*A* angles (Fig. 7) do not differ from previously determined ones, despite much greater abundance of data. There are significantly longer and weaker N–H···Osp^3 bonds than shorter ones (Fig. 6b), which is not entirely unexpected: N–H···O bonds are generally weaker (and therefore longer) than O–H···O.

Although the borders between particular classes of hydrogen bonds are diffuse, the clustering of geometric parameters can be used to recognize a class. The large range of

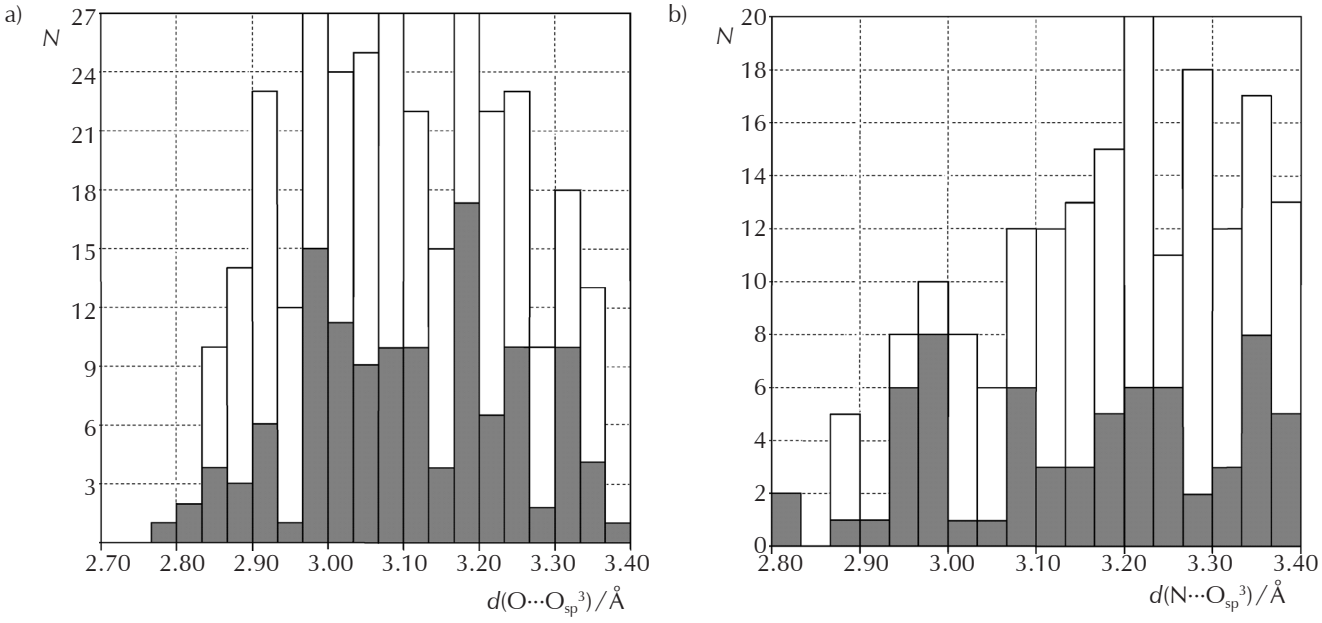


Fig. 6 – Distribution of a) O···Osp^3 distances in hydrogen bonds O–H···O, b) N···Osp^3 distances in hydrogen bonds N–H···O. Grey bars represent structures found in 2003 survey, and white bars represents new structures deposited since 2003.

Slika 6 – Raspodjela a) udaljenosti O···Osp^3 u vodikovim vezama O–H···O, b) udaljenosti N···Osp^3 u vodikovim vezama N–H···O. Sivi stupci predstavljaju strukture nađene u bazi 2003., dok se bijeli stupci odnose na nove strukture pohranjene nakon 2003.

both O···*A* and N···*A* distances (Fig. 8) reveals clustering of the data in three groups for a particular donor type (Fig. 8a, clusters 1–3, and Fig. 8b, clusters 1′–3′). Each cluster follows roughly a Gaussian distribution. According to their geometric parameters[21] clusters 1 and 1′ can be regarded as medium strong hydrogen bonds (O···Osp^3 < 3.0 Å, N···Osp^3 < 3.1 Å and Θ angles larger than 140°), while clusters 3 and 3′ comprise weak hydrogen bonds

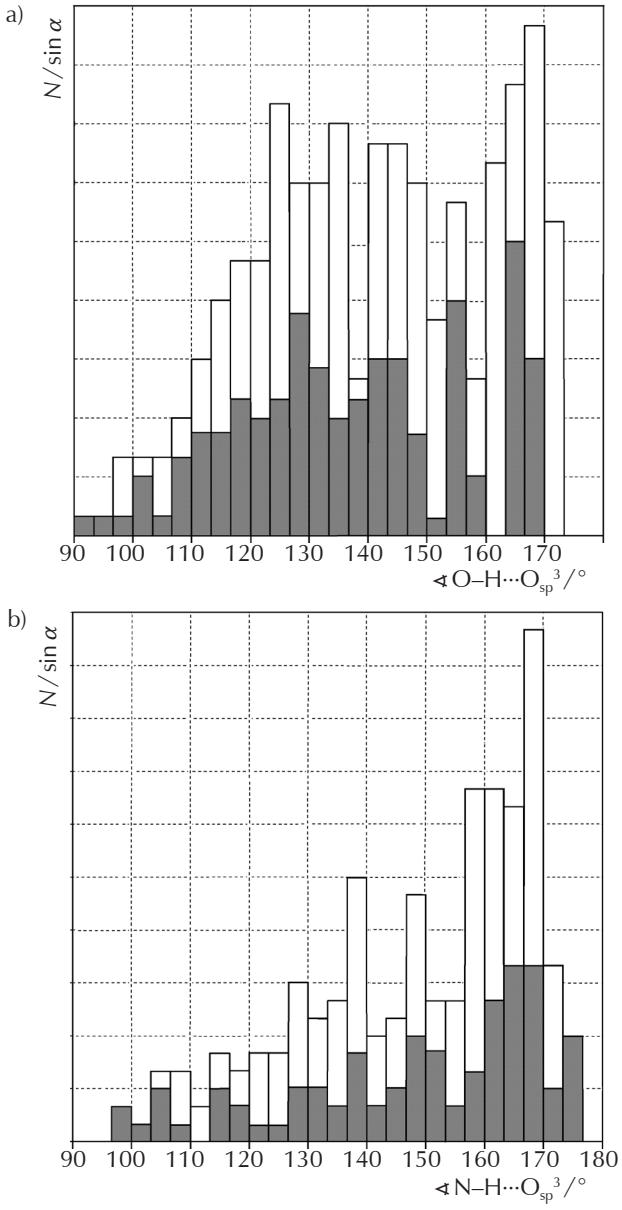


Fig. 7 – Distribution of a) O–H···Osp^3 angles in hydrogen bonds O–H···O, b) N–H···Osp^3 angles in hydrogen bonds N–H···O. To obtain better insight into angular preferences, a cone correction (weighted by 1/sin Θ)[20] was applied. Grey bars represent structures found in 2003 survey, and white bars represents new structures deposited since 2003.

Slika 7 – Raspodjela a) kutova O–H···Osp^3 u vodikovim vezama O–H···O, b) kutovima N–H···Osp^3 u vodikovim vezama N–H···O. Kako bi se stekao bolji uvid u prednosti pojedinih vrijednosti kutova, provedena je konusna korekcija (uteženo s 1/sin Θ).[20] Sivi stupci predstavljaju strukture nađene u bazi 2003., dok se bijeli stupci odnose na strukture pohranjene nakon 2003.

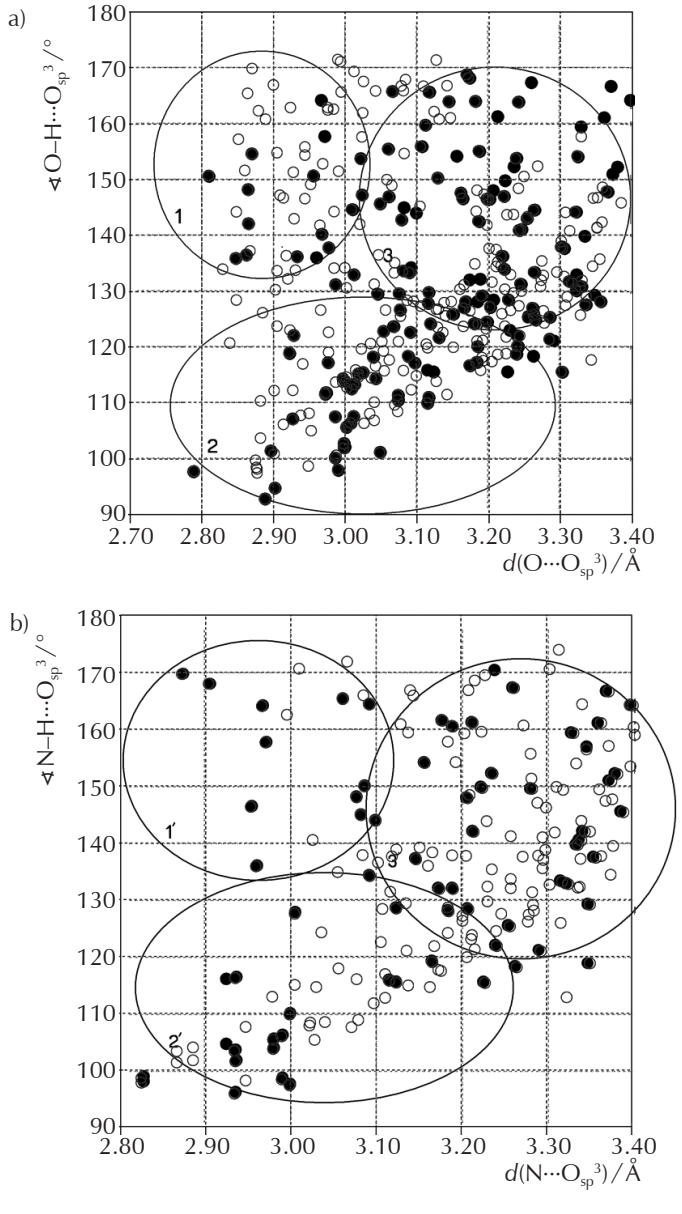


Fig. 8 – Plot of the Θ (*D*–H···Osp^3 angle) versus donor···acceptor distance *D*···Osp^3: a) for hydrogen bonds O–H···O, b) for hydrogen bonds N–H···O. Clusters of medium strength (clusters 1 and 1′) and weak hydrogen bonds (clusters 3 and 3′), and bifurcated hydrogen bonds (clusters 2 and 2′) are shown. Filled circles represent structures found in 2003 survey and empty circles represent structures deposited since 2003.

Slika 8 – Ovisnost kuta Θ (*D*–H···Osp^3 kut) *vs* donor···akceptor udaljenost *D*···Osp^3: a) za vodikove veze O–H···O, b) za vodikove veze N–H···O. Prikazani su skupovi koji obuhvaćaju srednje jake (skupovi 1 i 1′) i slabe vodikove veze (skupovi 3 i 3′), kao i račvaste vodikove veze (skupovi 2 i 2′). Ispunjeni krugovi odnose se na podatke iz baze u 2003., dok se prazni krugovi odnose na strukture pohranjene u bazi nakon 2003.

(O···Osp^3 > 3.0 Å, N···Osp^3 > 3.1 Å) The O···Osp^3 values for medium-strength bonds (clusters 1 and 1′, Fig. 8) correspond fairly well with the values for O–H···O hydrogen bonds in liquid water. Lower angles Θ in clusters 2 and 2′ (Fig. 8) indicate the presence of bifurcated hydrogen bonds; indeed, analysis of individual structures confirmed that in the majority of hydrogen bonds in clusters 2 and 2′ ester Osp^3 atom is a minor component of a bifurcated hydrogen bond (major component is quite often ester carbonyl O atom). To visualize the directionality of hydrogen bonds involving the ester Osp^3 atom, a CSD scatterplot was prepared using the program *IsoStar* [16] (Fig. 9). Inspection of the scatterplot shows the clustering of the proton donors around the ester Osp^3 atom, but also around Osp^2, representing the interaction with the second acceptor *A*′ in bifurcated hydrogen bonds.

At the Rudjer Bošković Institute, the team of Dr. S. Kirin is working on syntheses of bioorganic conjugates with amino acids and pseudopeptides and their transition metal complexes. Their interest is in the application of such compounds in novel materials,[22] asymmetric supramolecular catalysts,[23] and the role of metal ions in biological systems.[24]

At our Institute, SciFinder is not available, therefore the CSD is even more important to find structural information needed in this type of research to probe particular ligands for syntheses.[22–24]

The team at our Institute that works on the synthesis of metal complexes with organic ligands, has been using the CSD for searching molecular synthons of suitable geometries for synthesis. In the focus of such searches are geometrical parameters among nonbonded molecules or fragments (Fig. 10). The compounds $[Ni(bpy)_3]_2[TaO(C_2O_4)_3]Cl \cdot xH_2O$ (x = 11, 12) of interesting crystal packing were selected.[25]

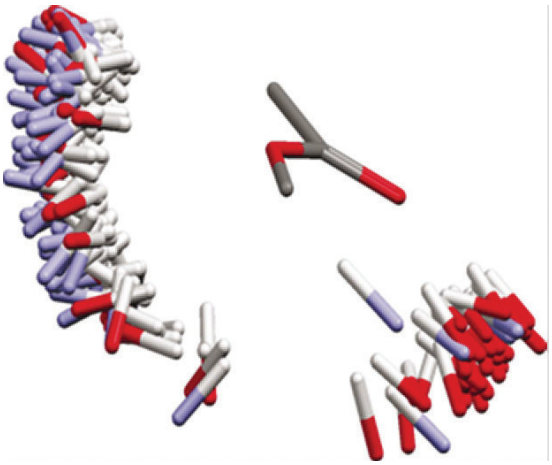


Fig. 9 – Scatterplot prepared by *IsoStar*[16] illustrating the grouping of proton donors oriented towards the lone-electron pair of the ester Osp^3 atom and carbonyl O atom (the latter participating in bifurcated hydrogen bonds as *A*′)[19]

Slika 9 – Dijagram pripremljen pomoću *IsoStar*[16] koji pokazuje nakupljanje skupina proton-donora orijentiranih prema nepodijeljenom elektronskom paru esterskog Osp^3 atoma i karbonilnog O atoma (tvoreći račvastu vodikovu vezu *A*′)[19]

The structural units of $[Ni(bpy)_3]^{2+}$ are involved in specific intermolecular contact between the aromatic ligands and *quadrupole aryl embrace* (QAE), in a honeycomb 2D hexagonal network. In order to find similar QAE contacts, a CSD search limited to the structures with orthogonal $[M(bpy)_3]^{n+}$ cation arrangement was performed. According to the applied criteria, 164 structures were found; QAE contacts in $[Ni(bpy)_3]_2[TaO(C_2O_4)_3]Cl \cdot xH_2O$ (x = 11, 12) showed good agreement with those previously described.

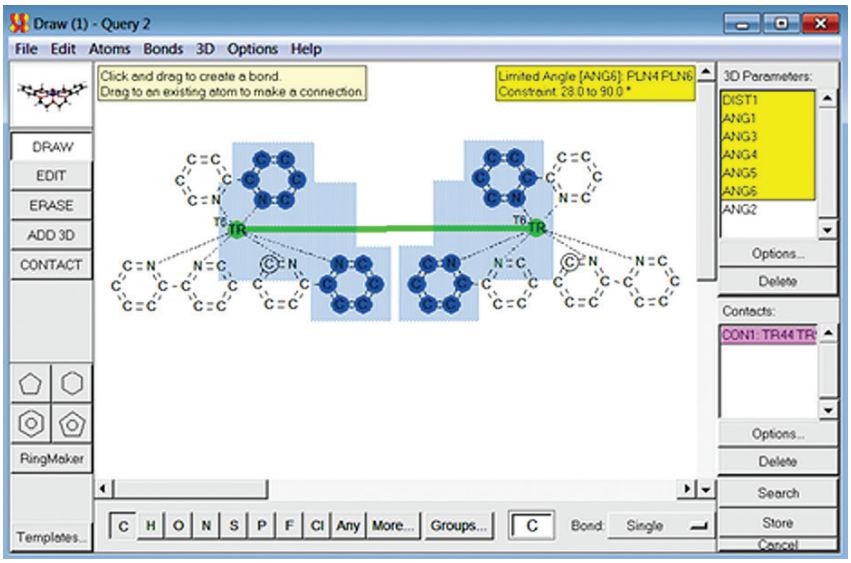


Fig. 10 – The layout of the new enhanced web interface for viewing and retrieving CSD entries exemplified by searching QAE contact (in green) between two $[Ni(bpy)_3]^{2+}$ in the complex $[Ni(bpy)_3]_2[TaO(C_2O_4)_3]Cl \cdot xH_2O$ (x = 11, 12) using CCDC *ConQuest* Search Query for QAE contacts[25]

Slika 10 – Primjena novog interaktivnog alata *WebCSD* za pretraživanje i vizualizaciju struktura u CSD-u ilustrirana na primjeru traženja kontakta QAE (označeno zelenim) između dva $[Ni(bpy)_3]^{2+}$ u kompleksu $[Ni(bpy)_3]_2[TaO(C_2O_4)_3]Cl \cdot xH_2O$ (x = 11, 12) primjenom CCDC *ConQuest* Search Query for QAE contacts[25]

The use of CSD in crystal engineering and supramolecular chemistry has been widely documented.[26] It offers a huge number of crystal structures defined by numerous noncovalent interactions that can be thoroughly studied. In this context, there are very instructive examples published by the research groups at Chemistry Department of Faculty of Science, University of Zagreb.[27–30] *Stilinović et al*.[27] elaborated the use of the noncovalent interaction V=C···O in the formation of dimeric and/or oligomeric species creating metal-organic frameworks (Fig. 11). A survey of the CSD revealed in 39 % structures (comprising vanadyl or pervanadyl groups) V=C···O contact distinguishing it from a putative hydrogen bond V=O···H–C.

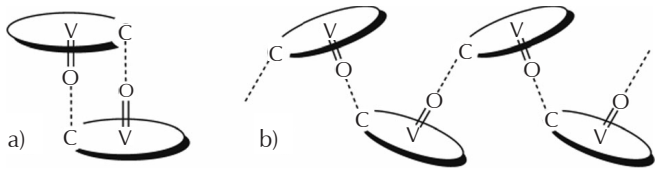


Fig. 11 – Oxovanadium coordination compounds generated by V=C···O interactions into: a) dimer and b) oligomer
Slika 11 – Oksovanadijevi koordinacijski spojevi izvedeni pomoću V=C···O interakcija u: a) dimer i b) oligomer

A very good example of an extensive use of the program *Mercury* is described by *Lekšić et al*.[28] (Fig. 12). Four cocrystals of lamotrigine, an anticonvulsant drug with highly pronounced side-effects were prepared and their crystal packing analysed in view of synthon used in generating hydrogen bonding network, characterized by graph-set notations.

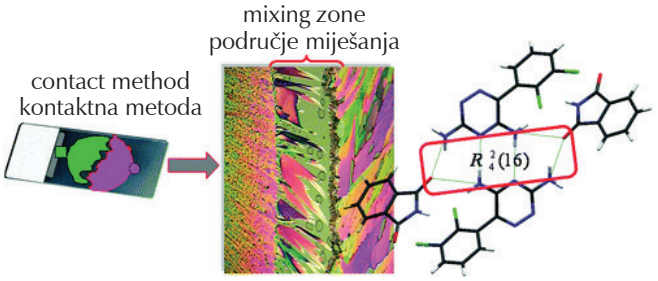


Fig. 12 – Cocrystals of lamotrigine based on coformers involving carbonyl group, discovered by hot-stage microscopy and DSC screening[28]
Slika 12 – Kokristali lamotrigina pripremljeni s komponentama koje sadržavaju karbonilnu skupinu otkriveni su mikroskopijom zagrijavanog uzorka i kalorimetrijom[28]

Among many tasks executable with the assistance of the CSD there are studies of polymorphism and solid-state transformations. *Halasz et al*.[30] carried out studies related to polymorphism of the cocrystals of 4-oxopimelic acid and 4,4'-bipyridine. Among three known polymorphs[31] these authors reported two conformational polymorphs of these cocrystals obtained by solvent crystallisation and their interconversion in the solid state. To compare the crystal packing of polymorphs, a detail analysis of hydrogen bonding networks is required. The complete analysis and comparison is performed by facilities and data stored in the CSD.

The prospect of Cambridge Structural Database in view of transdisciplinary research

The Cambridge Structural Database became much more than just a database. An efficient software for systematic analysis of structural data and the use for interpretation of intra- and intermolecular interactions is available. One can expect even more sophisticated developments in this direction. Understanding of complex living cells' dynamics, *e.g.* signalling processes, is of strong impact on drug and biosensors design, which are both important for patient treatment based on more reliable diagnostics. The most recent activity of The Cambridge Crystallographic Data Centre is the launch of the US operation unit to cooperate even more closely with Protein Data Bank in the frame of the Centre for Integrative Proteomics Research of Rutgers University in New Jersey. This 'joint venture' of both database centres will certainly lead to an optimisation of data extraction of (macro)molecular structures and their analyses through the application of sophisticated software and subspecialised databases. Their use in molecular modelling of targeting proteins and potential drugs will certainly open new paths for design of biosensors, as well.

The cooperation between these data centres is on the line of networking and developing of cloud computing, which enables management of 'Big Data' including their extraction, transportability and fusion towards the final goal. The present situation poses a great challenge to demonstrate the value and innovative power of (bio)informatics in the field of chemistry, materials science, the life sciences; molecular design tremendously benefits from 'Big Data' depending on (predictive) modelling and multidimensional optimization of molecular structure and function. Molecular informatics creates a general platform for unifying all these aspects and designing innovative use of data (Fig. 13). Strong demands on drug design have opened a new path towards such innovative use of data. These approaches are already in use to relate genetic and epigenetic characteristics with diseases.

However, more should be done in the fast growing field of materials science, primarily with graphene. New views on the interpretation of crystal structures taking into account periodicity and also the dynamic aspect of electronic structure are required. The demanding approach will be the design of software that can reliably reproduce and interpret these types of interactions. The two papers, recently published in *IUCrJ*, brought into focus the problem of interpretation of intermolecular atom – atom bonds, *i.e.* interatomic distances significantly shorter than the sum of Van der Waals radii in crystals.[32,33] The knowledge obtained through analyses of the large number of structures will be precious for recognition of characteristic structural features for structure / property relations and preparation of (multi)functional new materials.

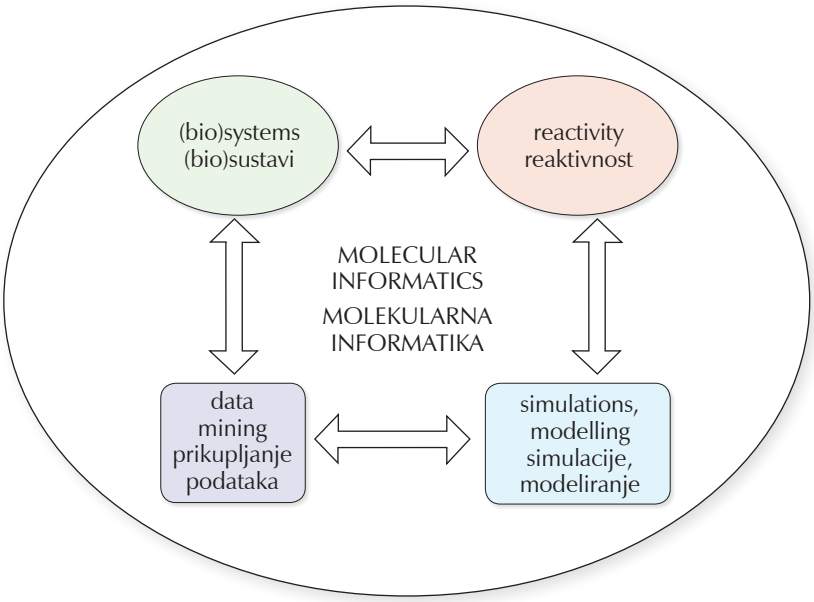
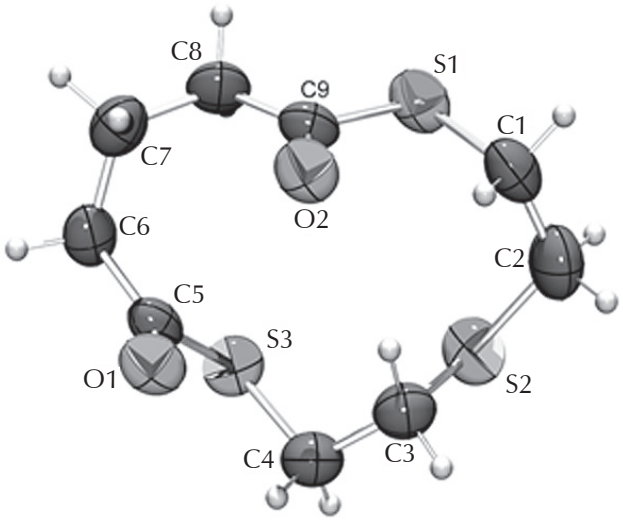


Fig. 13 – *Molecular informatics* includes information and data storages – processing on the molecular level – signalling and regulation of biological and chemical systems including cellular systems and macromolecular assemblies; modelling of molecular interactions and networks; the design of molecular modulators that exhibit desired (bio)chemical and/or pharmacological effects

Slika 13 – Molekularna informatika uključuje pohranjivanje informacija i podataka – procesiranje na molekularnoj razini – signalizacijom i regulacijom bioloških i kemijskih sustava uključujući stanične sustave i makromolekularne ansamble; modeliranje molekularnih interakcija i mreža; dizajn molekularnih modulatora koji prouzrokuju željeni (bio)kemijski i/ili farmakološki učinak

ACKNOWLEDGMENT

To celebrate the 50th anniversary of the Cambridge Structural Database, a special contribution to the CCDC Valentine's collection (announced in February 2015 by S. Vyas) is presented here: the heart-shaped molecule of 2,5,8-trithiacyclododeca-1,9-dione.[34]



List of abbreviations
Popis kratica

CSD – The Cambridge Structural Database
– Baza strukturnih podataka u Cambridgeu

CCDC – The Cambridge Crystallographic Data Centre
– Centar za kristalografske podatke u Cambridgeu

CIF – Crystallographic Information File
– kristalografski informacijski file

QAE – Quadrupole Aryl Embrace
– četverostruko arilno okruženje

References
Literatura

1. URL: http://www.ccdc.cam.ac.uk (accessed May 2, 2015).
2. *C. R. Groom, F. A. Allen*, The Cambridge Structural Database in Retrospect and Prospect, Angew. Chem. Int. Ed. **53** (2014) 662–671, doi: http://dx.doi.org/10.1002/anie.201306438.
3. *F. H. Allen*, The Cambridge Structural Database: A quarter of a million crystal structures and rising, Acta Cryst. B**58** (2002) 380–388, doi: http://dx.doi.org/10.1107/S0108768102003890.

4. *B. Kojić-Prodić*, The Cambridge Structural Database - its use and applications, Kem. Ind. **35** (1986) 401–403.
5. *A. Višnjevac, B. Kojić-Prodić*, Cambridge Structural Database – "gold mine" of knowledge, Kem. Ind. **51** (2002) 425–430.
6. *F. H. Allen, R. Taylor*, Librarians, crystal structures and drug design, Chem. Commun. (2005) 5135–5140, doi: http://dx.doi.org/10.1039/b511106b.
7. *B. Kojić-Prodić, K. Molčanov*, 40th Anniversary of Cambridge Crystallographic Data Centre dedicated to data deposition of molecular and crystal structures – Cambridge Structural Database, Kem. Ind. **55** (2006) 261–270.
8. *S. R. Hall, F. H. Allen, I. D. Brown*, The Crystallographic Information File (CIF): A New Standard Archive File for Crystallography, Acta Cryst. A**47** (1991) 655–685, doi: http://dx.doi.org/10.1107/S010876739101067X.
9. URL: http://journals.iucr.org/e/services/authorservices.html (accessed July 3, 2013).
10. *F. H. Allen, O. Johnson, G. P. Shields, B. R. Smith, M. Towler*, CIF Applications: enCIFER, a program for viewing, editing and visualizing CIFs, J. Appl. Cryst. **37** (2004) 335–338, doi: http://dx.doi.org/10.1107/S0021889804003528.
11. *I. J. Bruno, J. C. Cole, P. R. Edgington, M. Kessler, C. F. Macrae, P. McCabe, J. Pearson, R. Taylor*, New software for searching the Cambridge Structural Database and visualising crystal structures, Acta Cryst. B**58** (2002) 389–397, doi: http://dx-.doi.org/10.1107/S0108768102003324.
12. *I. R. Thomas, I. J. Bruno, J. C. Cole, C. F. Macrae, E. Pidcock, P. A. Wood*, *Web*CSD: The online portal to the Cambridge Structural Database, J. Appl. Cryst. **43** (2010) 362–366, doi: http://dx.doi.org/10.1107/S0021889810000452.
13. *C. F. Macrae, I. J. Bruno, J. A. Chisholm, P. R. Edgington, P. McCabe, E. Pidcock, L. Rodriguez-Monge, R. Taylor, J. van de Streek, P. A. Wood*, Mercury CSD 2.0 – New Features for the Visualization and Investigation of Crystal Structures, J. Appl. Cryst. **41** (2008) 466–470, doi: http://dx.doi.org/10.1107/S0021889807067908.
14. *R. A. Sykes, P. McCabe, F. H. Allen, G. M. Battle, I. J. Bruno, P. A. Wood*, New Software for Statistical Analysis of CSD Data, J. Appl. Cryst. **44** (2011) 882–886, doi: http://dx.doi.org/10.1107/S0021889811014622.
15. *P. A. Wood, T. S. G. Olsson, J. C. Cole, S. J. Cottrell, N. Feeder, P. T. Galek, C. R. Groom, E. Pidcock*, Evaluation of molecular crystal structures using Full Interaction Maps, CrystEngComm **15** (2013) 65–72, doi: http://dx.doi.org/10.1039/C2CE25849H.
16. *I. J. Bruno, J. C. Cole, J. P. M. Lommerse, R. S. Rowland, R. Taylor, M. L. Verdonk*, IsoStar: A library of information about nonbonded interactions, J. Comput.-Aided Mol. Des. **11** (1997) 525–537, doi: http://dx.doi.org/10.1023/A:1007934413448.
17. *I. J. Bruno, J. C. Cole, M. Kessler, J. Luo, W. D. S. Motherwell, L. R. Purkis, B. R. Smith, R. Taylor, R. I. Cooper, S. E. Harris, A. G. Orpen*, Retrieval of Crystallographically-Derived Molecular Geometry Information, J. Chem. Inf. Comput. Sci. **44** (2004) 2133–2144, doi: http://dx.doi.org/10.1021/ci049780b.
18. *S. J. Cottrell, T. S. G. Olsson, R. Taylor, J. C. Cole, J. W. Liebeschuetz***,** Validating and understanding ring conformations using small molecule crystallographic data, J. Chem. Inf. Model. **52** (2012) 956–962, doi: http://dx.doi.org/10.1021/ci200439d.
19. *K. Molčanov, B. Kojić-Prodić*, Analysis of the less common hydrogen bonds involving ester oxygen sp3 atoms as acceptors in the crystal structures of small organic molecules, Acta Cryst. B**60** (2004) 424–432, doi: http://dx.doi.org/10.1107/S0108768104014442.
20. *J. Kroon, J. A. Kanters*, Non-linearity of hydrogen bonds in molecular crystals, Nature (London) **248** (1974) 667–669, doi: http://dx.doi.org/10.1038/248667a0.
21. *G. A. Jeffrey*, An Introduction to Hydrogen Bonding, Oxford University Press, 1997.
22. *S. I. Kirin, H.-B. Kraatz, N. Metzler-Nolte*, Systematizing structural motifs and nomenclature in 1,n´-disubstituted ferrocene peptides, Chem. Soc. Rev. **35** (2006) 348–354, doi: http://dx.doi.org/10.1039/b511332f.
23. *Z. Kokan, S. I. Kirin*, The application of 'backdoor induction' in bioinspired asymmetric catalysis, RSC Adv. **2** (2012) 5729–5737, doi: http://dx.doi.org/10.1039/c2ra20598j.
24. *S. I. Kirin, I. Ott, R. Gust, W. Mier, T. Weyhermüller, N. Metzler-Nolte*, Cellular uptake quantification of metallated peptide and peptide nucleic acid bioconjugates by atomic absorption spectroscopy, Angew. Chem. Int. Ed. **47** (2008) 955–959, doi: http://dx.doi.org/10.1002/anie.200703994.
25. *L. Androš Dubraja, D. Matković-Čalogović, P. Planinić*, Crystal disassembly and reassembly of heterometallic Ni(II)–Ta(V) oxalate compounds, CrystEngComm **17** (2015) 2021–2029, doi: http://dx.doi.org/10.1039/C4CE02516D.
26. *G. R. Desiraju, J. J. Vittal, A. Ramanan*, Crystal Engineering, A Textbook, World Scientific Publishing, Singapore, 2011, doi: http://dx.doi.org/10.1142/8060.
27. *V. Stilinović, D.-K. Bučar, I. Halasz, E. Meštrović*, V=C···O interactions in crystal structures of oxovanadium-coordination compounds, New J. Chem. **37** (2013) 619–623, doi: http://dx.doi.org/10.1039/C2NJ40780A.
28. *E. Lekšić, G. Pavlović, E. Meštrović*, Cocrystals of lamotrigine based on conformers involving carbonyl group discovered by hot-stage microscopy and DSC screening, Cryst. Growth Des. **12** (2012) 1847–1858, doi: http://dx.doi.org/10.1021/cg201426z.
29. *M. Cindrić, M. Rubčić*, The rare examples of thiosemicarbazonato chromium(III) complexes: crystal structure of $[Cr(Hsal\ 4\text{-}Metsc)_2]Cl{\cdot}CH_3OH$ and $[Cr(Hsal\ 4\text{-}Phtsc)_2]Cl$, Croat. Chem. Acta **85** (2012) 505–513, doi: http://dx.doi.org/10.5562/cca2153.
30. *I. Halasz, M. Rubčić, K. Užarević, I. Đilović, E. Meštrović*, The cocrystal of 4-oxopimelic acid and 4,4'-bipyridine: polymorphism and solid-state transformations, New J. Chem. **35** (2011) 24–27, doi: http://dx.doi.org/10.1039/C0NJ00569J.
31. *D. Braga, G. Palladino, M. Polito, K. Rubini, F. Grepioni, M. R. Chierotti, R.Gobetto*, Cocrystals of 4,4'-bipyridine/pimelic acid and their structural, thermal, and spectroscopic characterisation, Chem. Eur J. **14** (2008) 10149–10159, doi: http://dx.doi.org/10.1002/chem.200801051.
32. *J. D. Dunitz*, Intermolecular atom–atom bonds in crystals?, IUCrJ **2** (2015) 157–158, doi: http://dx.doi.org/10.1107/S2052252515002006.
33. *C. Lecomte, E. Espinosa, C. Matta*, On atom–atom 'short contact' bonding interactions in crystals, IUCrJ **2** (2015) 161–163, doi: http://dx.doi.org/10.1107/S2052252515002067.
34. *I. Vujasinović, J. Veljković, K. Mlinarić-Majerski, K. Molčanov, B. Kojić-Prodić*, Solid-state tubular assemblies of thiolactones: synthesis and structural characterization, Tetrahedron **62** (2006) 2868–2876, doi: http://dx.doi.org/10.1016/j.tet.2006.01.016.

PROŠIRENI SAŽETAK

50-Godišnjica Cambridge Structural Database i 30-godišnjica uporabe u Hrvatskoj

Biserka Kojić-Prodić i Krešimir Molčanov*

Prikaz je posvećen dr. Franku H. Allenu, FRSC CChem, 1944. – 2014.

Prikaz je posvećen uspomeni na dr. F. H. Allena i obilježavanju 50-godišnjice The Cambridge Crystallographic Data Centre (CCDC), svjetski poznatog centra, koji pohranjuje i provjerava kristalografske podatke, uključujući koordinate atoma koje određuju trodimenzijsku strukturu organskih molekula i metalnih kompleksa s organskim ligandima. Misija institucije istaknuta na njezinim web-stranicama je jasna: "Centar za kristalografske podatke u Cambridgeu (CCDC) potpomaže napredak kemije i kristalografije za sveopću dobrobit pružajući kvalitetne informacije, softver i podršku." Baza strukturnih podataka u Cambridgeu (CSD), jedna je od prvih elektroničkih i najznačajnijih baza u svijetu. U Međunarodnoj godini kristalografije 2014., u prosincu, Centar je obznanio 750.000. pohranjenu strukturu. Upotreba opsežne baze podataka, koja se brzo povećava, zahtijeva podršku u dobro osmišljenom i učinkovitom softveru za provjeru, pretraživanje, analizu i vizualizaciju strukturnih podataka. Prestižna je uloga te baze u istraživanjima povezanim s kristalografijom, kemijom, znanošću o materijalima, fizikom i kemijom krutog stanja, (bio)tehnologijom, znanošću o životu, te farmakologijom, posebno u oblikovanju lijekova. CCDC posvećuje posebnu pažnju točnosti pohranjenih podataka i razvoju softvera za tu namjenu (CIF). Jedan od najvažnijih softvera za korisnike je *ConQuest*, koji omogućava pretraživanje svih informacija u bazi CSD-a, te njegovo postavljanje na mreži u interaktivnom obliku pod imenom *WebCSD*. Za vizualizaciju kristalnih struktura i morfologije kristala, uključujući analizu intramolekularnih i intermolekularnih interakcija te topološke oznake vodikovih veza, kao i svih geometrijskih podataka, na raspolaganju je *Mercury*. Centar (CCDC) pruža još veće mogućnosti korisnicima razvijajući kompleksne softvere poput *GOLD*, *IsoStar* i *SuperStar* te *DASH,* kao i opsežne elektroničke biblioteke poput *Mogula i Relibase*.
CCDC je pružio novu mogućnost korištenjem *Mercuryjeva* modula *Solid Form*. Takvi složeni i zahtjevni projekti mogu se ostvarivati samo s veoma kvalitetnim timom stručnjaka sa znanstvenim pristupom, temeljeći se na dugogodišnjoj tradiciji u kristalografiji, modeliranju i informatici. Sama baza, raznolik softver i alati za pretraživanje afirmirali su se u istraživanju, nastavi i stručnoj izobrazbi. Primjena elektroničkih medija i računalne grafike čini "data mining" učinkovitima i korisnima, ali i estetski privlačnim zbog molekularne arhitekture. Možemo očekivati još više unapređenja zbog uporabe "računarstva u oblacima" (*cloud computing*) i rukovanja opsežnim podacima (Big Data); umrežavanje podataka iz srodnih baza pomoći će da prepoznamo sakrivena svojstva molekula i kristala i informacija koje mogu otkriti nove važne spoznaje.
Od 1985. Strukturna baza podataka u Cambridgeu (CSD) dostupna je korisnicima u Hrvatskoj. Upotreba te baze u Hrvatskoj ilustrirana je s nekoliko primjera autora ovog prikaza i nekih korisnika.

Ključne riječi
Kristalografska baza podataka, baza strukturnih podataka u Cambridgeu (CSD), kristalna struktura, molekularna struktura

Institut Ruđer Bošković, PP 180
HR-10 002 Zagreb
Hrvatska

Pregledni rad
Prispjelo 13. svibnja 2015.
Prihvaćeno 12. lipnja 2015.