

Cambridge Structural Database – zlatni rudnik znanja

KUI 27/2002
Received 29. travnja, 2002.
Accepted 7. lipnja, 2002.

A. Višnjevac i B. Kojić-Prodić

Zavod za fizičku kemiju, Institut "Ruđer Bošković",
pp 180, HR-10002 Zagreb, Hrvatska

U radu je opisana povijest, aktualni trenutak, ali i perspektive *Cambridge Crystallographic Data Centre (CCDC)* s osobitim naglaskom na *Cambridge Structural Database (CSD)*, najveću i najvažniju bazu kristalografskih podataka na svijetu. Predstavljena je i baza nevezanih interakcija *IsoStar* koja je izvedena iz *CSD*-a te raznolika programska podrška koju razvijaju stručnjaci *CCDC*-a, a služi pretraživanju baza i analizi podataka koje one sadrže. Na Institutu "Ruđer Bošković" već 17 godina uspješno djeluje nekad jugoslavenski, a sada hrvatski nacionalni centar za suradnju s *CCDC*-om, koji je i danas nadležan za korisnike u Hrvatskoj, Sloveniji i Makedoniji. Kako doći do novih znanja o strukturama molekula i njihovim interakcijama i korisno ih primijeniti u istraživanjima i nastavi, koristeći se dobrobitima računskih metoda, računalne tehnologije i molekularne grafike, podsjetit će nas ovaj prikaz.

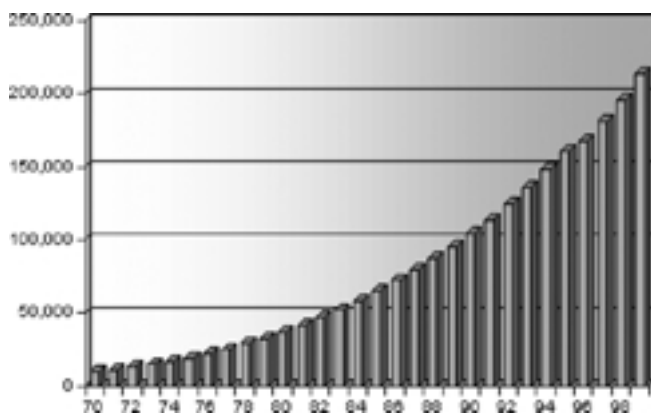
Ključne riječi: *Kristalografske baze, CCDC, CSD, IsoStar, SuperStar, GOLD, ConQuest*

Uvod

Nitko danas ne osporava presudan utjecaj koji kristalografija ima na razvoj suvremene kemije i srodnih znanosti.¹ Rendgenska (neutronska) strukturalna analiza (koja je fizikalni temelj kristalografije) pruža obilje podataka o molekularnoj geometriji, načinu slaganja molekula te o unutarmolekularnim i međumolekularnim nevezanim interakcijama. Uz to, kristalografske metode su često nezamjenjive pri jednoznačnoj identifikaciji novih spojeva. Već unazad nekoliko desetljeća uočena je potreba za klasificiranjem i pohranjivanjem riješenih struktura u računalne baze podataka odakle bi bile dostupne najširoj znanstvenoj javnosti. U prvo je vrijeme cilj baza bio samo pohranjivanje podataka. Razvojem informatičkih tehnologija baze su sve više počele služiti planiranju novih strategija istraživanja. Uz *Cambridge Structural Database (CSD)*,² kojom se bavi ovaj

tekst, najvažnije kristalografske baze podataka danas su *Protein Data Bank (PDB)*,³ *Nucleic Acid Database (NDB)*,⁴ *Inorganic Crystal Structure Database (ICSD)*,⁵ *CrystMet*^{®6} i *Powder diffraction file (PDF)*.⁷

Prof. dr. *Olga Kennard* utemeljila je 1965. *Cambridge Crystallographic Data Centre* (u daljem tekstu *CCDC*), kao malu ustrojbenu jedinicu pri Kemijskom odsjeku Sveučilišta u Cambridgeu, kojoj je primarni zadatak bio sastaviti računalnu bazu kemijskih i kristalografskih podataka o organskim i organometalnim spojevima izučavanim metodama rendgenske i neutronske difrakcije. Na samom je početku djelovanja *CCDC*-a tako nastala *Cambridge Structural Database* (u daljem tekstu: *CSD*) koja sadrži koordinate atoma u kristalnoj strukturi kao i osnovne bibliografske, kemijske i kristalografske podatke za riješene kristalne strukture. *CSD* je počeo s podacima za oko 1500 kristalnih struktura s godišnjim povećanjem baze za nekoliko stotina novih struktura. Od skromnih početaka pa do današnjih dana *CSD* odražava velik napredak kristalografije golemim povećanjem broja pohranjenih struktura. U 2002. godini baza već premašuje brojku od 250 tisuća pohranjenih struktura (inačica 5.23, travanj 2002.) s godišnjim rastom od blizu 20 tisuća novih (sl. 1). *CSD* je danas daleko najveća i najvažnija baza eksperimentalno određenih kristalnih struktura na svijetu. Ona pokriva spojeve koji sadrže ugljik, a kod kojih ukupan broj atoma ne prelazi tisuću (organski, organometalni spojevi i metalni kompleksi), zatim peptide s do 24 aminokiseline (viši oligomeri pohranjuju se u *PDB*³), te mono-, di- i trinukleotide (viši oligomeri pohranjeni su u *NDB*⁴). Broj riješenih struktura pohranjenih u *CSD*-u udvostručio se otprilike svakih sedam do osam godina. Većina relevantnih znanstvenih publikacija danas zahtijeva od autora da kristalografske podatke pohrane u odgovarajuću kristalografsku bazu podataka prije objavljivanja rada. Tom se prilikom provjerava sve unesene podatke, pa baza pruža



Slika 1 – Grafički prikaz povećanja broja pohranjenih struktura u *CSD*

Fig. 1 – Graphical representation of the increasing number of the deposited structures in the *CSD*

aktualan pregled i točne podatke. CCDC 1989. godine postaje neovisna neprofitna institucija pri Sveučilištu u Cambridgeu. Iako registriran kao društvo s ograničenom odgovornošću, Centar je i dalje ostao akademska institucija ovlaštena za vođenje poslijediplomskih i doktorskih studija. Danas je osnovna djelatnost Centra usmjerena ka održavanju i razvoju baze podataka kristalnih struktura malih organskih i organometalnih molekula – CSD, razvoju programske podrške za pretraživanje i prikazivanje podataka iz baze, kreiranju izvedenih baza podataka i odgovarajuće programske podrške za njih, kao i distribuciji cijelog paketa brojnim akademskim i industrijskim korisnicima širom svijeta. Danas CSD ima preko 750 akademskih i 110 industrijskih pretplatnika u više od 50 zemalja svijeta. Korisnicima se osvežena inačica baze podataka i pripadajuće programske podrške do sada dostavljala svakih šest mjeseci putem CDROM-a. Od 2002. godine CDROM s osveženom inačicom baze dostavljat će se korisnicima svakih 12 mjeseci, ali će pretplatnici baze na mrežnim stranicama CCDC-a u svakom trenutku moći odabrati i lokalno pohraniti aktualnu inačicu.⁸ Korisnici su svrstani u nacionalne centre preko kojih se obavlja sva potrebna komunikacija s CCDC-om. Našoj znanstvenoj javnosti baza je dostupna od 1985.⁹ putem ugovora između CCDC-a i Nacionalnog centra za suradnju s CCDC-om kojeg vodi dr. Biserka Kojić-Prodić, smješten je na Institutu "Ruđer Bošković" u Zagrebu, Bijenička c. 54, HR-10000 Zagreb, a ovlašten je osim za dva pretplaćena korisnika u Hrvatskoj (Institut "Ruđer Bošković" i PMF Sveučilišta u Zagrebu) i za korisnike u Sloveniji i Makedoniji.

Arhitektura CSD

Osnovni je program za pretraživanje CSD od travnja 2000 *ConQuest*¹⁰ kojim je zamijenjen stariji *Quest3D*.¹¹ *ConQuest* je znatno unaprijedio pretraživanje CSD-a u smislu brzine, mogućnosti pri kreiranju upita i, osobito, jednostavnosti. *ConQuest* će zacijelo pridonijeti i popularizaciji CSD-a jer je s njim po prvi put omogućeno pretraživanje baze i na MS WINDOWS platformama (uz UNIX i njegove

izvedenice SGI-IRIX, IBM AIX, Sun Solaris i Intel Linux). Za statističku analizu brojčanih podataka iz baze primjenjuje se program *Vista* (ili *MS-Excel*). *Vista* rabi zadane vrijednosti geometrijskih parametara iz izabranih kristalnih struktura pohranjenih u CSD-u za raznolike statističke analize. Pomoću ovog programa moguće je računati, primjerice, srednje vrijednosti odabranih duljina veza, valentnih ili torzijskih kutova unutar odabranog seta struktura. Vrlo je jednostavno povezati se na samu bazu i ispitati uzroke pojavljivanja nekog neuobičajenog parametra u određenoj strukturi. Rezultati se mogu iskazati računski i grafički. *Vista* također posjeduje mogućnost estetskog oblikovanja grafika, što je važno pri pripremi materijala za publiciranje.

Detalji molekularne strukture, vodikove veze i kristalno slaganje analiziraju se programom *RPluto*, koji je (kao i *Vista*) uključen u paket programske podrške koju kreira i svojim korisnicima distribuira CCDC. *RPluto* je uveden još 1970. i od tada se njegova osnovna zadaća nije bitno promijenila iako je program više puta nadograđivan te grafički i sadržajno osvežavan. Premda je *RPluto* u prvom redu namijenjen analizi detalja molekularnih i kristalnih struktura pohranjenih u CSD-u, on može čitati SHELX¹² i CIF¹³ datoteke, pa je prema tome koristan i pri analizi nepohranjenih, novih struktura, ponajprije složenih sustava vodikovih veza.¹⁴ Najnoviji program za trodimenzionalnu vizualizaciju molekularnih i kristalnih struktura iz radionice CCDC-a je *Mercury*, uveden 2001.¹⁵ Nadgradnju sustava čini baza neveznih interakcija *IsoStar*¹⁶ koja je uvedena 1997., kao i program *SuperStar*,¹⁷ temeljen na empirijskoj metodi pronalaženja aktivnih mjesta u makromolekulama. *SuperStar* primjenjuje znanje pohranjeno u bazi neveznih interakcija *IsoStar*. Tu je i protein-ligand "docking" program *GOLD*,^{18,19} koji primjenjuje genetski algoritam razvijen u suradnji sa stručnjacima Glaxo Wellcomea i Sveučilišta u Sheffieldu, a korisnicima je stavljen na raspolaganje 1998. godine. Uz to, CCDC u suradnji s dr. Manfredom Hendlichom sa Sveučilišta u Marburgu distribuira i razvija *ReLiBase*²⁰ – software za pretraživanje i rukovanje podacima iz *Protein Data Bank (PDB)*. Spomenimo i *DBuse* – bazu publikacija nastalih na temelju uporaba podataka iz CSD-a i drugih CCDC-ovih proizvoda.²¹ *DBuse* je nastao u travnju 1995., ima internetsko sučelje, besplatan je i u trenutku kad nastaje ovaj tekst sadrži gotovo 900 pohranjenih publikacija.

U ovom prikazu *Cambridge Structural Database* osvrnućemo se detaljnije na novi program za pretraživanje baze – *ConQuest*, bazu neveznih interakcija – *IsoStar* i program za njezino pretraživanje *SuperStar* te na protein-ligand docking program *GOLD*.

ConQuest

*ConQuest*¹⁰ je program za pretraživanje baze kristalografskih podataka CSD. Omogućuje korisniku pretraživanje baze po više kriterija na osnovi kojih se postavljaju pitanja (sl. 3), npr:

1. grafički definiranog kemijskog dijagrama strukturnog fragmenta od interesa
2. imena autora

Slika 2 – Shematski prikaz osnovne arhitekture CSD

Fig. 2 – Schematic representation of the basic architecture of the CSD

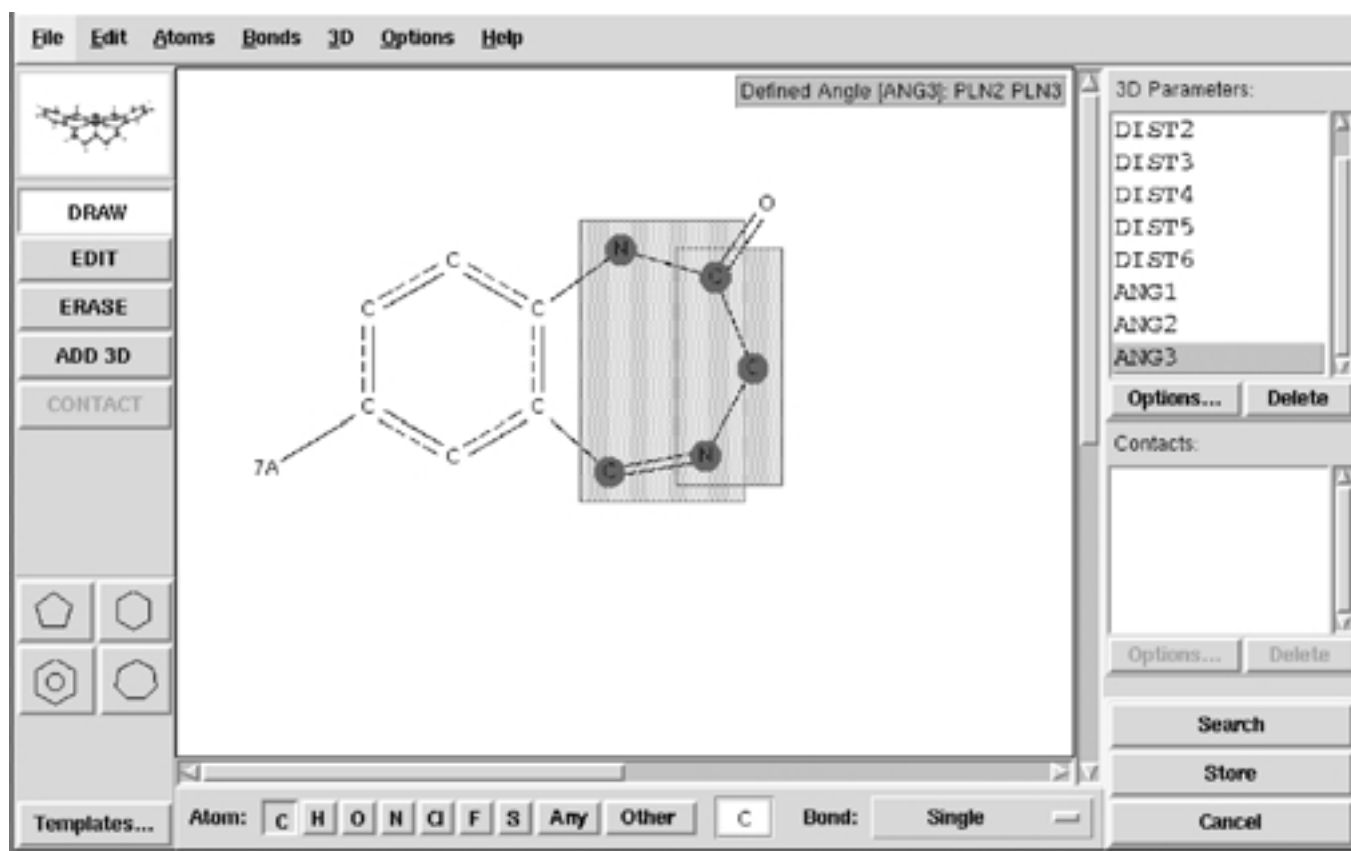
3. naslova publikacije – može (i ne mora) se navesti volumen, godina ili primjerice raspon godina
4. kemijskog naziva spoja
5. kemijskih elemenata koje struktura sadrži – u jednoj ili više molekula, ili se može navesti najteži element u strukturi
6. *brutto*-formule osnovne tražene molekule ili cijele strukture
7. temeljnih kristalografskih podataka o strukturi (prostorne grupe, dimenzija jedinične ćelije, broja molekula u jediničnoj ćeliji ili gustoće kristala)
8. eksperimentalnih uvjeta (temperatura snimanja, izvor zračenja ...)
9. opisnih elemenata (primjerice: *green*, *disordered*, *twinned*, *coordinated*, *resolved* ...) ili
10. REFCODE-a (svaka struktura u bazi ima svoju oznaku po kojoj je uvijek možemo jednostavno i brzo pronaći).

Pitanja je moguće kombinirati po principu Boolovih operatora. Prije početka pretraživanja moguće je unaprijed odbaciti strukture s određenim graničnim *R*-faktorom (ako želimo temeljiti analizu na vrlo točnim rješenjima), neseledene strukture ili pak one koje nemaju pohranjene ko-

ordinate atoma. Možemo se ograničiti na pretraživanje organskih spojeva (isključivši, npr. organometalne). Izuzetno je korisna mogućnost izračunavanja molekularne geometrije za jednu strukturu ili za odabrane skupine struktura. Podatke *ConQuest*, na korisnikov zahtjev, sprema u posebnu datoteku koju je poslije moguće uporabiti pomoću programa *Vista*. Kada je pretraživanje završeno, *ConQuest* nudi čitav spektar prikaza rezultata pretraživanja – od trodimenzionalnog modela, preko kemijskog dijagrama do obilja kristalografskih podataka, geometrijskih parametara te bibliografskih i tekstualnih zapisa o pronađenim strukturama (sl. 4). Sustavnom analizom skupine spojeva odabranih po željenom kriteriju može se npr. izračunati sterički ili elektronski utjecaj supstitucije u zadanom sustavu, utjecaj hibridizacije na planarnost sustava, učinak polarnosti donora odnosno akceptora na svojstva vodikove veze ili usporediti teorijske i eksperimentalne vrijednosti parametara molekularne geometrije, izučavati kinetiku nekih reakcija itd.

Isostar

IsoStar je baza nekovalentnih interakcija napravljena s ciljem da pomogne svima koji se bave supramolekularnom kemijom, molekularnim prepoznavanjem, međudjelovan-



Slika 3 – Grafičko sučelje *ConQuest*a pri zadavanju kemijske vrste za pretraživanje. Na slici je prikazano definiranje raznih geometrijskih parametara (dviju ravnina među kojima je definiran kut, nekoliko duljina veza). Cilj je dobiti brojčane podatke koji odgovaraju zadanim parametrima za sve strukture koje odgovaraju upitu. S njima se zatim može provesti statistička analiza.

Fig. 3 – Graphical interface of *ConQuest* while a search query is being created. The figure reveals a process of defining geometrical parameters (two planes with an interplane angle defined, few bond lengths). The goal of this query is to obtain the requested numerical parameters for all the structures matching the criteria defined. They can be used for a subsequent statistical analysis.

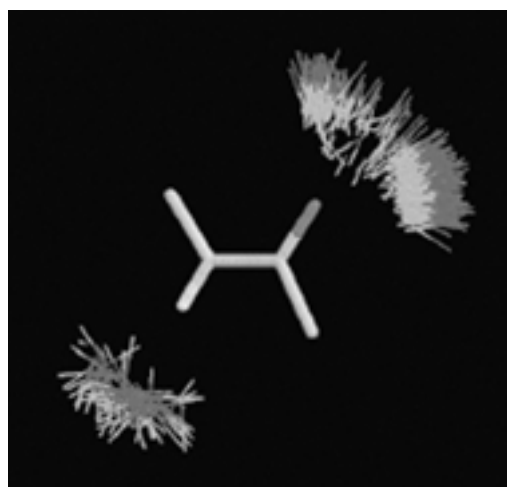
The screenshot shows the ConQuest software interface. At the top, there is a menu bar with 'File', 'Edit', 'Options', 'View Databases', 'Searches', and 'Help'. Below the menu bar are three tabs: 'Build Queries', 'Combine Queries', and 'View Results'. On the left side, there is a vertical menu with options: 'Author/Journal', 'Chemical', 'Crystal', 'Experimental', 'Diagram', '3D Visualiser', 'CSD Internals', and 'Search Overview'. The main window displays 'Refcode: FLDAZP'. In the center, a chemical structure is shown, which is a piperazine ring substituted with a benzamide group and a 4-chlorophenyl group. To the right of the structure is a list of parameters: DIST1 (1.422), DIST2 (1.358), DIST3 (1.520), DIST4 (1.474), DIST5 (1.276), DIST6 (1.492), ANG1 (80.440), ANG2 (40.513), and ANG3 (59.049). On the far right, there is a list of 20 chemical structures, each with a checked box and its name: BEDZPN10, BIXBOZ10, BIZSAE, BOFDEF, BOMMUL, BORXEL, BZPCUC, CAGWOW, CHABZN, CICSEM, CLDZPA, CLDZPB, CPHAZO, CPHAZO01, DIZPAM10, DZOXAM, FAFNEF, FILYAA, FIXTUB, and FLDAZP. Below this list are navigation buttons '<<' and '>>', a '50 hits' indicator, a '100%' progress bar, and a 'Stop Search' button. At the bottom of the main window, there is a 'Multiple Hits: Show' button, a page indicator '1 of 2', and a 'Show Parameters' button.

Slika 4 – Grafičko sučelje ConQuesta pri prikazu rezultata. Desno je odjeljak s kodovima pronađenih struktura, u sredini željeni prikaz, a sasvim lijevo izbornik s ponuđenim načinima prikaza rezultata.

Fig. 4 – Graphical interface of ConQuest when the results of the query are listed. To the right is a list of the structures matching the criteria of the query, in the middle the requested type of the representation (chemical diagram, 3D, bibliographic informations etc.) and to the left a ConQuest menu.

jima protein-ligand kao i ciljanim oblikovanjem spojeva za upotrebu u kliničke svrhe (oblikovanje lijekova). Baza sadrži informacije o stotinama nevezanih interakcija (međudjelovanja molekula unutar kristala malih organskih molekula) izdvojenih iz *Cambridge Structural Database* i iz *Protein Data Bank* (slaba djelovanja između proteinskog aktivnog mjesta i molekule-liganda) kao i teorijske podatke za čitav niz sustava (međudjelovanja molekula u plinskoj fazi – *molecular orbital data*). *IsoStar* je prilagođen mrežnim pretraživačima (primjerice Netscape® ili MS Explorer®), a sadrži čitav niz ugrađenih tehničkih pomagala poput *RasMol*²² prozora za prikaz rezultata, kao i veze (*hyperlink*) izravno na podatke u CSD odnosno PDB.

IsoStar nudi tablicu nekovalentnih interakcija najučestalijih organskih molekula s vodik-donorskim, vodik-akceptorskim i hidrofobnim skupinama temeljenu na podacima iz CSD-a, PDB-a i na teorijskim energetskim računima. Istražuje li se strogo određena nekovalentna interakcija između dviju kemijskih skupina (ili molekula), rezultati će biti dani pomoću *RasMol* prozora (sl. 5). *Isostar* pri tome uzima u obzir sva nekovalentna međudjelovanja koja odgovaraju



Slika 5 – *IsoStar* prikaz statističkog razmještaja hidroksilnih skupina oko peptidne veze, na temelju podataka iz CSD-a
Fig. 5 – The *IsoStar* representation of the statistical distribution of hydroxyl group around the peptide bond, based on the CSD data

zadanim parametrima zabilježena u CSD-u i PDB-u. U prikazu rezultata primjenjuje se prekrivanje (*overlap, superposition*) zadanih skeleta prepoznatih u nizu pretraženih struktura, a kemijske skupine koje ostvaruju zadane nekovalentne kontakte raspoređene su u geometrijski zadanom prostoru prema kriterijima ionskih ili van der Waalsovih radija (sl. 5). Pomoću kontrolnog prozora možemo manipulirati prikazom rezultata – odabrati, primjerice, sasvim uski raspon udaljenosti unutar nekovalentnog kontakta za prikazivanje. Za bilo koji prikazani kontakt moguće je putem izravne veze na CSD-u ispitati sve strukturne detalje. Dostupnost podataka i iz PDB-a omogućuje izravnu usporedbu analognih nekovalentnih interakcija u kristalima malih molekula i makromolekula. Te su usporedbe pokazale da najveći broj nekovalentnih interakcija u malim molekulama i makromolekulama pokazuje statističku podudarnost pa je time ponovno naglašena važnost kristalografije malih molekula koja se može primijeniti kao modelni sustav za izučavanje proteina.^{23,24} Izravna veza na PDB omogućuje da odabirom bilo kojeg nekovalentnog kontakta od interesa pogledamo detalje strukture aktivnog mjesta proteina i sve druge podatke o tom proteinu pohranjene u PDB-u. *IsoStar* stavlja na raspolaganje i teorijske podatke o nekoliko stotina sustava temeljenih na međumolekularnoj teoriji izmjene (*Intermolecular perturbation theory*). Rezultat tih računa daje energetski minimum za određenu nekovalentnu interakciju. Usporedbom teorijskih i eksperimentalnih podataka u najvećem broju slučajeva dolazi se do podudarnih vrijednosti (maksimalna vjerojatnost na mjestu minimalne energije). Ako traženi nekovalentni kontakt nije uveden u bazu, korisniku je omogućeno predložiti vlastiti, do kojeg može doći pretraživanjem CSD-a. Programom *IsoGen* prevodi se u format koji prepoznaje *IsoStar*.

SuperStar

Programi za prepoznavanje veznih (aktivnih) mjesta u proteinima poznati su već dulje vrijeme, no većina od njih kao svoj temelj primjenjuje empirijska polja sila.^{25,26} Takvi programi ovisе poglavito o sveobuhvatnosti i temeljitosti njihova "znanja" o neveznim interakcijama, pa se pokazala potreba za novim pristupom po kojem će se predviđanje neveznih interakcija u proteinima temeljiti na podacima pohranjenim u bazama podataka eksperimentalno određenih kristalnih struktura.^{27,28} *SuperStar* predviđa načine povezivanja proteina i liganda na temelju statističke analize CCDC-ove baze neveznih interakcija *IsoStar*. Rezultat je takvog pristupa pouzdan program, dokazan već na stotinjak kompleksa protein-ligand, koji se temelji isključivo na eksperimentalnim podacima. *SuperStar* je potpuno temeljen na kristalografskim podacima iz CSD-a, što omogućuje daleko veću preciznost u predviđanju tipa neveznih interakcija nego li da su podaci uzeti iz PDB-a (jer je preciznost rezultata rendgenske strukturne analize na malim molekulama još uvijek puno veća nego li na makromolekulama). Uz to, u CSD-u je zastupljen veći broj funkcionalnih skupina koje ostvaruju nevezne interakcije, broj zabilježenih pojedinih interakcija je neusporedivo veći (što omogućuje računanje preciznijih mapa elektronske gustoće), a određeni su i položaji vodikovih atoma.

GOLD

GOLD (*Genetic Optimization for Ligand Docking*)^{29,30} je program za predviđanje dinamike vezivanja liganda na protein (*protein docking*). Način na koji će ligand naći svoj put do aktivnog mjesta od velike je važnosti za istraživanja na polju oblikovanja novih lijekova. Zato su programeri CCDC-a zajedno sa stručnjacima Sveučilišta u Sheffieldu i Glaxo Wellcome razvili algoritam temeljen na pretraživanju konformacijskog prostora uz punu gibljivost liganda i djelomičnu konformacijsku pokretljivost samog proteina. Taj algoritam uvažava konformacijske i fiziološke značajke prirodnog sustava. Energetske funkcije u *GOLD*-u djelomično su temeljene na konformacijskoj analizi i informacijama o nekovalentnim interakcijama prepoznatim pretraživanjem CSD-a. *GOLD* je provjeren na temelju 136 protein-ligand kompleksa iz PDB-a. Analiza je pokazala da se u 71 % slučajeva predviđanje slaže s eksperimentalnim podacima.

Budućnost i perspektive

Vidljivo je iz prikazanog da je nakon gotovo četiri desetljeća postojanja CSD daleko nadmašio osnovnu namjenu zbirke riješenih struktura, što je bila prvobitna namjera njegovih utemeljitelja. Vrlo brzo se pokazalo da je takva zbirka nepresušan izvor sekundarnih znanstvenih informacija do kojih treba samo naći put. Zato su se stručnjaci CCDC-a posvetili razvoju izvedenih baza podataka poput *IsoStara* – baze nekovalentnih interakcija. *IsoStar* je koristan alat u izučavanju vezivanja supstrata kao i inhibitora u aktivno mjesto enzima. Uopće, pristup molekularnog prepoznavanja temelji se uglavnom na saznanjima o nekovalentnim interakcijama. Taj koncept, potpomognut molekularnim modeliranjem, služi i u izučavanju bioloških mehanizama na atomnoj razini kao i u oblikovanju novih spojeva željenih svojstava, posebice lijekova. Sa svakom novom pohranjenom strukturom znanje i mogućnosti koji su na raspolaganju svjetskoj znanstvenoj zajednici višestruko se povećavaju. Nema sumnje da je kristalografija metoda izbora za strukturnu karakterizaciju novih spojeva, ne samo zato jer na taj način dobivamo potpuno pouzdane podatke o kemijskoj i molekularnoj strukturi novog spoja nego i zato što pohranjivanjem strukture u CSD dodatno utočnjujemo parametre *IsoStara*. Činjenica da prirodni zakoni koji određuju strukturu "malih" molekula jednako vrijede i za kompleksne makromolekule dopušta širenje horizonta spoznaje o složenim biološkim sustavima. To je posebno bitno u sadašnjem trenutku, kad makromolekularna kristalografija po mogućnostima rješenja još zaostaje za kristalografijom malih molekula.

Literatura

1. B. Kojić-Prodić, J. Kroon, *Croat. Chem. Acta* **74** (2001) 1.
2. F. H. Allen, O. Kennard, *Chem. Des. Automat. News* **8** (1993) 1.
3. The *Protein Data Bank* (PDB), RCSB, Department of Chemistry, Rutgers University, PO Box 939, Piscataway, NJ 08855-0939, USA.
4. H. M. Berman, W. K. Olson, D. L. Beveridge, J. Westbrook, A. Gelbin, T. Demeny, S.-H. Hsieh, A. R. Srinivasan, B. Schneider, *Biophys. J.* **63** (1992) 751.

5. ICSD for WWW, <http://barns.ill.fr/dif/icsd/> ili <http://www.fiz-informationsdienste.de/DB/icsd/index.html>
6. <http://www.tothcanada.com/crystmet/overview.html>
7. <http://www.icdd.com/>
8. <http://www.ccdc.cam.ac.uk/prods/conquest/updates>
9. B. Kojić-Prodić, *Kem. Ind.* **35** (1986) 401.
10. <http://www.ccdc.cam.ac.uk/conquest>
11. D. G. Watson, *J. Res. Natl. Inst. Stand. Technol.* **101** (1996) 227.
12. G. M. Sheldrick, SHELXS97: Program for the Solution of Crystal Structures, Universität Göttingen, Germany, 1997. i G. M. Sheldrick, SHELXS97: Program for the Refinement of Crystal Structures, Universität Göttingen, Germany, 1997.
13. S. R. Hall, F. H. Allen, I. D. Brown, *Acta Cryst.* **A47** (1991) 655.
14. B. Perić, Uloga vodikovih veza u organizaciji retropeptida s oksalilnim i tiooksalilnim mostovima, Magistarski rad, Sveučilište u Zagrebu, Zagreb, 2000.
15. <http://www.ccdc.cam.ac.uk/prods/mercury/index.html>
16. I. J. Bruno, J. C. Cole, J. P. M. Lommerse, R. S. Rowland, R. Taylor, M. L. Verdonk, *J. Comp.-Aided Mol. Des.* **11** (1997) 525.
17. M. L. Verdonk, J. C. Cole, R. Taylor, *J. Mol. Biol.* **289** (1999) 1093.
18. G. Jones, P. Willet, R. C. Glen, *J. Mol. Biol.* **245** (1995) 43.
19. G. Jones, P. Willet, R. C. Glen, A. R. Leach, R. Taylor, *J. Mol. Biol.* **267** (1997) 727.
20. M. Hendlich, *Acta Cryst.* **D54** (1998) 1178.
21. <http://www.ccdc.cam.ac.uk/dbuse>
22. R. A. Sayle, E. J. Milner-White, *Trends Biochem. Sci.* **20** (1995) 374.
23. M. M. Harding, *Acta Cryst.* **D55** (1999) 1432.
24. M. M. Harding, *Acta Cryst.* **D56** (2000) 857.
25. P. J. Goodford, *J. Med. Chem.* **28** (1985) 849.
26. G. E. Kellogg, S. F. Semus, D. J. Abraham, *J. Comput. Aided Mol. Des.* **5** (1991) 545.
27. J. D. Danziger, P. M. Dean, *Proc. Roy. Soc.* **236** (1989) 115.
28. G. Klebe, *J. Mol. Biol.* **237** (1994) 212.
29. G. Jones, P. Willet, R. C. Glen, *J. Mol. Biol.* **245** (1995) 43.
30. G. Jones, P. Willet, R. C. Glen, A. R. Leach, R. Taylor, *J. Mol. Biol.* **267** (1997) 727.

SUMMARY

Cambridge Structural Database – a Golden Mine of Knowledge

A. Višnjevac and B. Kojić-Prodić

The paper deals with the past, present status and the perspectives of *Cambridge Crystallographic Data Centre (CCDC)*, with a focus on the *Cambridge Structural Database (CSD)*, the biggest and the most important crystallographic database worldwide, containing more than 250.000 structures of organic and organometallic compounds. *CSD* covers structures of compounds with less than 1000 atoms. *CCDC* distributes the updated version of the *CSD* every six months to over 750 academic and 110 industrial subscribers in more than 50 countries. Croatian affiliation centre for *CCDC* is run by dr. Biserka Kojić-Prodić at the "Ruđer Bošković" Institute in Zagreb, and covers the users in Croatia, Slovenia and Macedonia.

Main searching tool for searching and retrieving data from the *CSD* is *ConQuest*. This powerful software is equipped with a full range of text/numeric database search options, in addition to more complex search functionality. *ConQuest* offers several way of representing results of the search: bibliographic information, chemical, crystallographic or experimental information, as well as the chemical diagram and 3D representation. Any structure found can be selected for viewing. In addition to *CSD* and its searching tool *ConQuest*, *CCDC* distributes also the database of the noncovalent interactions – *IsoStar*, which is based on the data from both the *CSD* and *PDB (Protein Data Bank)*. The information about the non-bonded interactions that is stored in this database, is essential for studying molecular recognition, protein-ligand interactions and supramolecular chemistry. *IsoStar* is equipped with a web-based searching tool, and the results are presented using an incorporated *RasMol* window. A hyperlink to the *CSD* and *PDB* is enabled as well. *SuperStar* is a program for generating maps of interaction sites in proteins using experimental information about intermolecular interactions. *SuperStar* retrieves its data from *IsoStar*. *SuperStar* is available to *CSD* subscribers as a free evaluation. Other tools are also on disposal to all *CSD* subscribers: *Vista*-software for statistical analysis of the numerical data from *CSD*, *Mercury*-program for the three-dimensional visualisation of the *CSD* entries, and *GOLD* – a protein-ligand docking program. The aim of this paper is to motivate the scientific community to use the *CCDC* products in searching for recent knowledge on molecular architecture and molecular interactions, and to apply it in research and teaching, using computational methods and technologies including molecular graphics.

Department of Physical Chemistry,
"Ruđer Bošković" Institute,
P.O.B. 180, HR-10002 Zagreb, Croatia

Received April 29, 2002
Accepted June 7, 2002