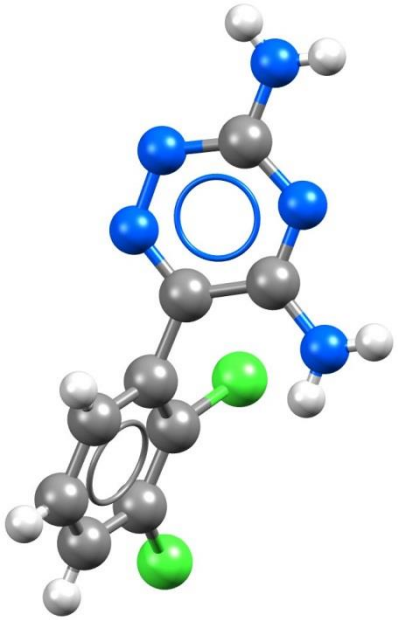


The Cambridge Structural Database (CSD) – a Vital Resource for Structural Chemistry and Biology

Stephen Maginn, CCDC, Cambridge, UK

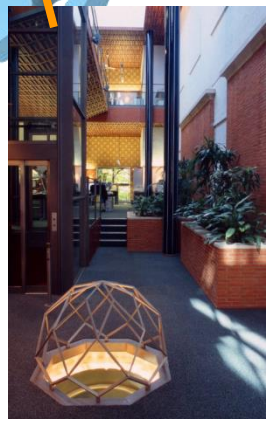
The Cambridge Crystallographic Data Centre

"The advancement and promotion of the science of chemistry and crystallography for the public benefit"

- Founded in 1965
- Self financing, self administering institution since 1987
  - Not-for-profit, charitable, Independent Research Institute
- University of Cambridge Partner Institute

The CCDC worldwide

The Cambridge Crystallographic Data Centre Inc.
Center for Integrative Proteomics Research,
174 Frelinghuysen Road,
Piscataway,
New Jersey 08854
United States of America



JAICI
(Japanese distributors)
Nakai Building
6-25-4 Honkomagome
Bunkyo-ku
Tokyo 113-0021
Japan

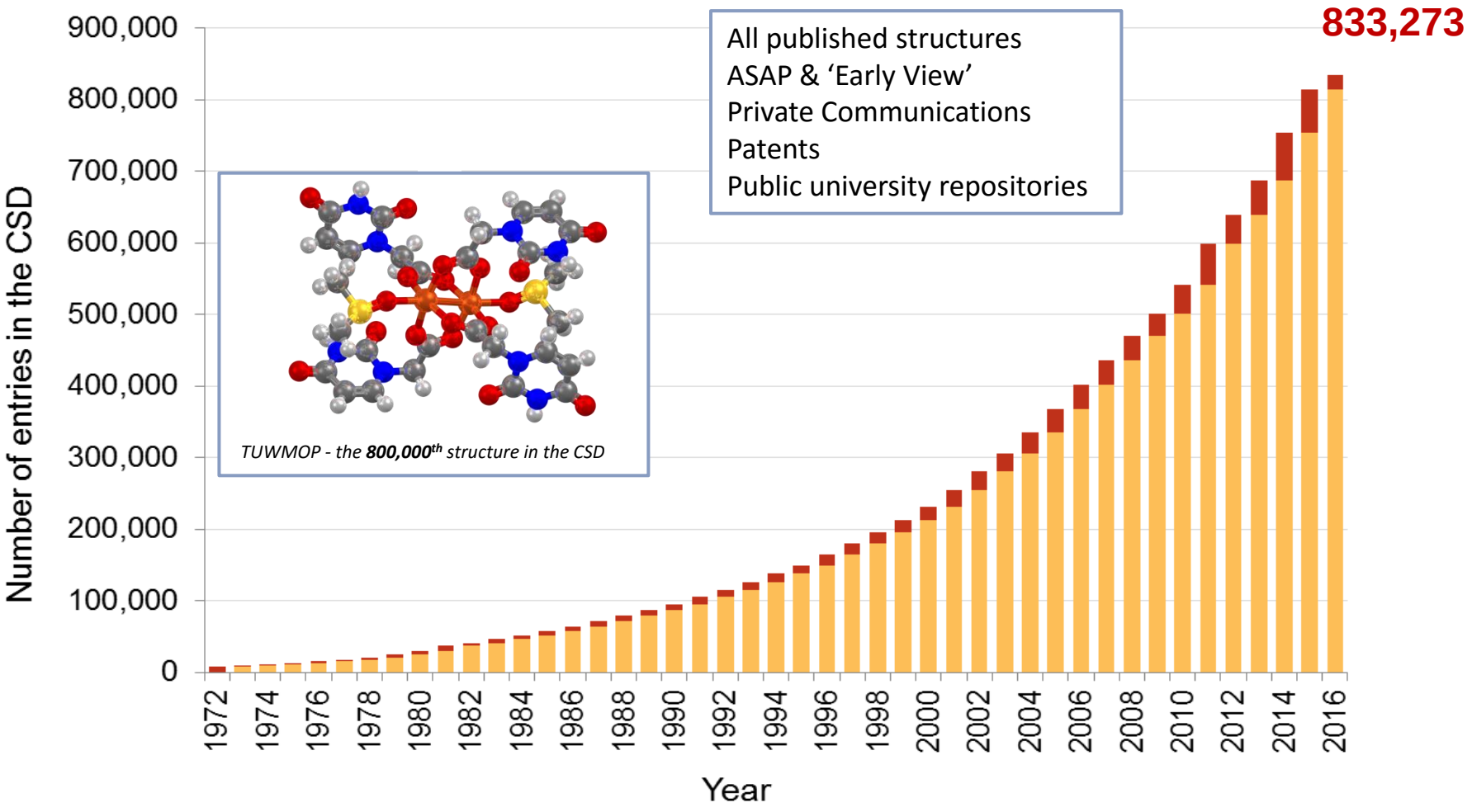
Cambridge Crystallographic Data Centre
12 Union Road
Cambridge
CB2 1EZ
United Kingdom

Cambridge Structural Database

- Repository for the world's small organic and metal-organic crystal structures (up to 500 non-H atoms)
- Experimentally determined 3D structures via X-ray, and neutron diffraction methods (single crystal and powder)
- 2016 release contains 800,309 entries
  - approximately 65,000 entries added per year
  - 32,964 entries added in February and May 2016 updates
- Derived from around 1200 published sources
  - official depository for >90 major journals
  - majority of data directly deposited electronically (CIF)
- Increasing number of CSD Communications

The Cambridge Structural Database

Worldwide repository of small-molecule organic & metal-organic crystal structures



CSD Information Content: Bibliographic and Chemical Information

- Each 3D structure in the CSD is accompanied by a 2D diagram and textual information
- All text properties e.g. bibliographic, chemical and properties are searchable
- Chemical diagram and chemical connectivity enable 2D and 3D searching for substructures, pharmacophores and intermolecular interactions

BASYOJ

4-Oxonicotinamide-1-

(1'-beta-D-2',3',5'-tri-O-acetyl-ribofuranoside)

Source: Rothmannia longiflora

Colour: pale yellow

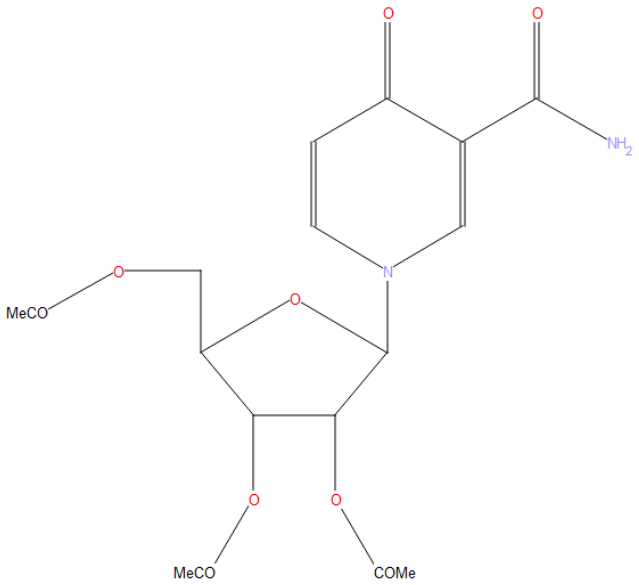
Habit: acicular

Polymorph: Form IV

C17 H20 N2 O9

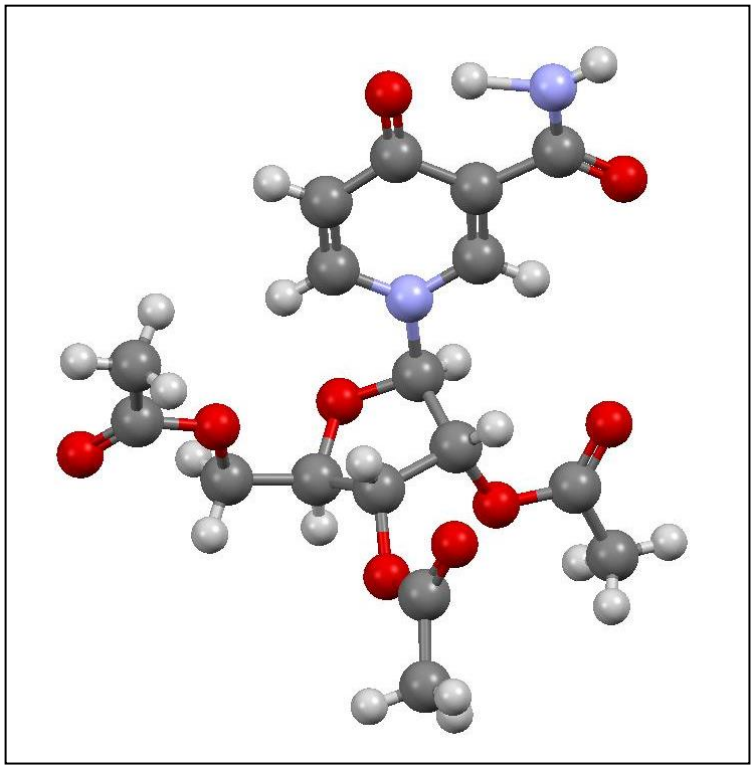
G. Bringmann, M. Ochse, K. Wolf, J. Kraus, K. Peters, E-M. Peters, M. Herderich, L. Ake, F. Tayman *Phytochemistry* 51 (1999), p271

R-factor: .0506

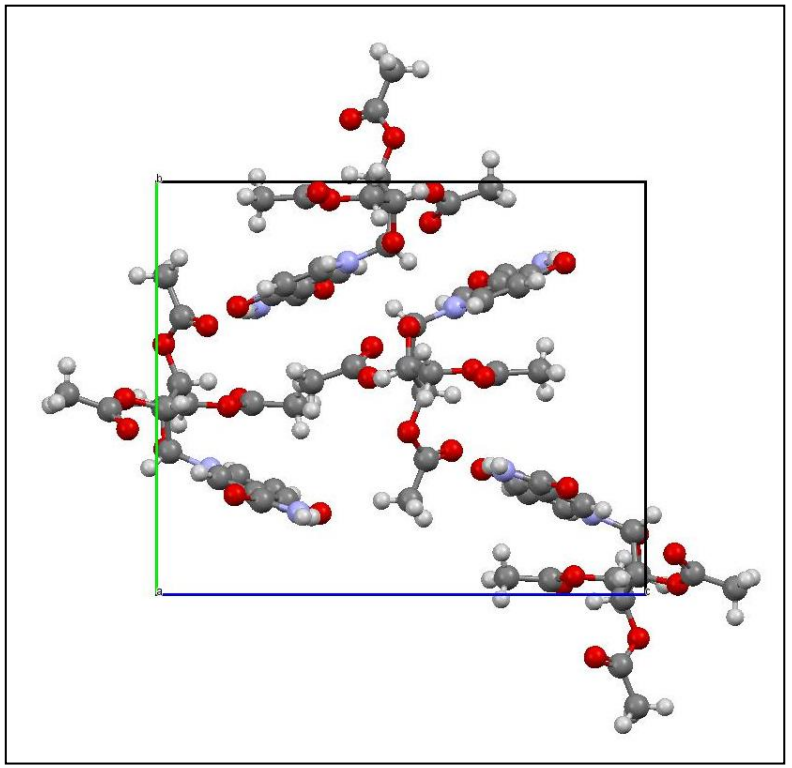


CSD Information Content: Structural Data

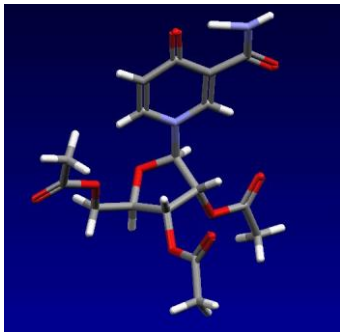
- Atomic coordinates and crystallographic packing information (symmetry and unit cell size)

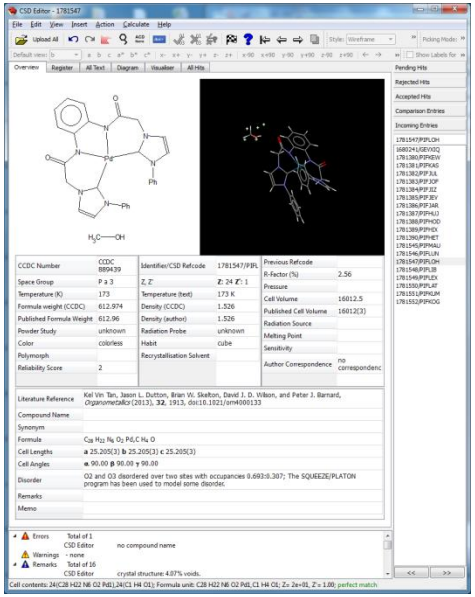


BASYOJ



Structure quality





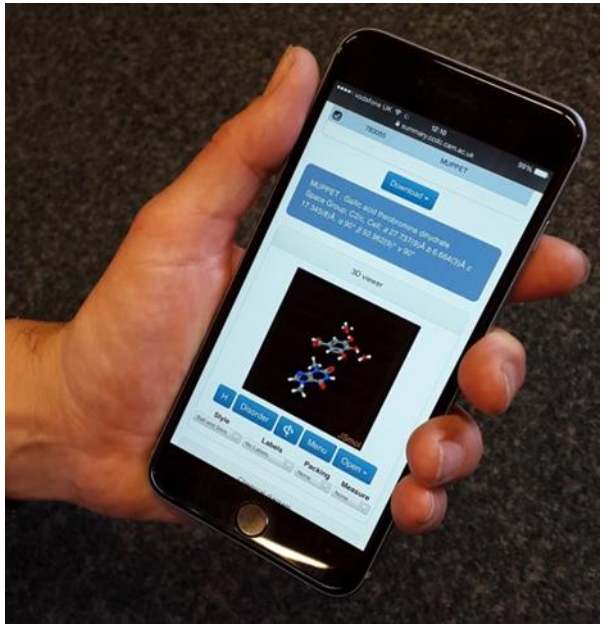
- Automatic syntax checking
- Duplication and revision checks
- DeCIFer - Advanced validation technology
  - Deduces chemical structure from the coordinates
  - Assigns a reliability score
  - Adds a chemical diagram and chemical name
  - Resolves disorder
  - *Acta Cryst.* (2011). B**67**, 333-349 doi:10.1107/S0108768111024608
- A team of expert Scientific Editors and Deposition Coordinators
- Manual enrichment

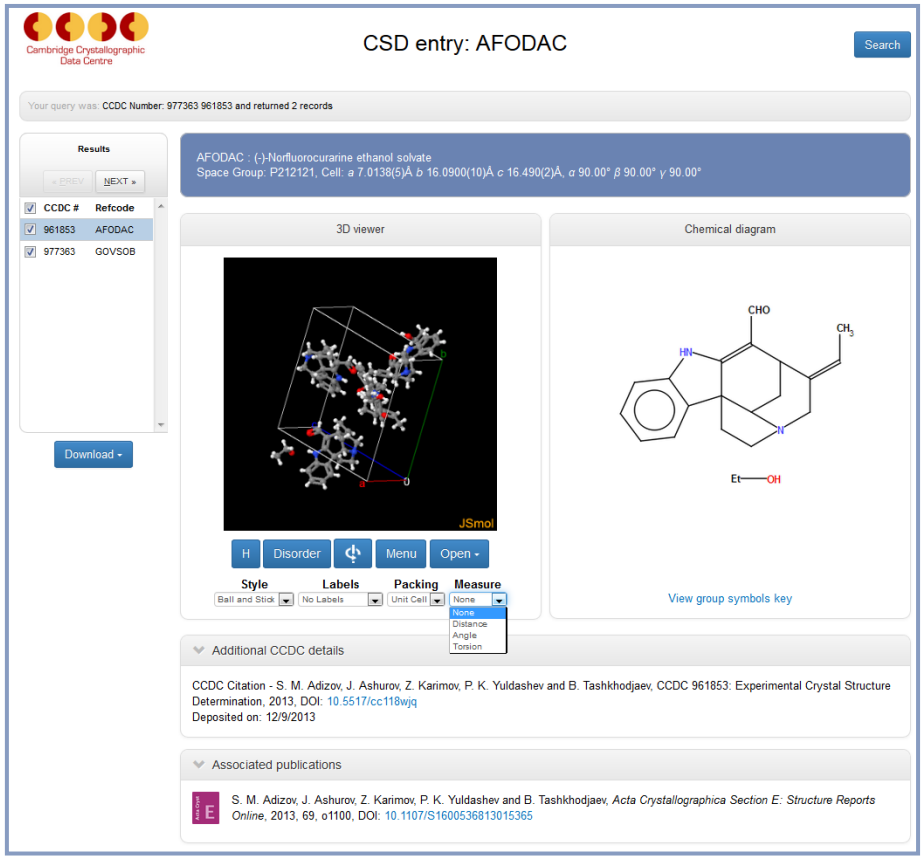
Access Structures

Access Structures

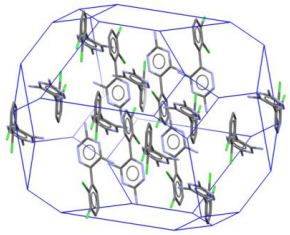
View and retrieve structures in the Cambridge Structural Database

Click on “Access Structures” at www.ccdc.cam.ac.uk



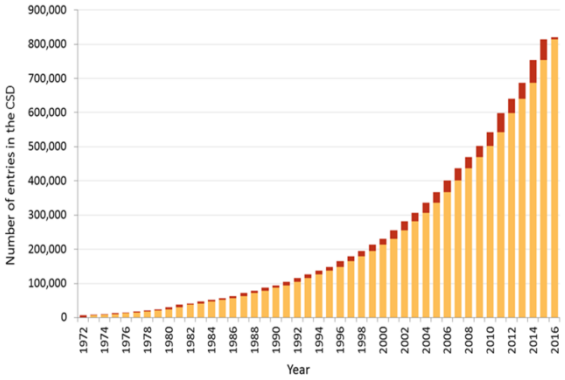


The Cambridge Structural Database System

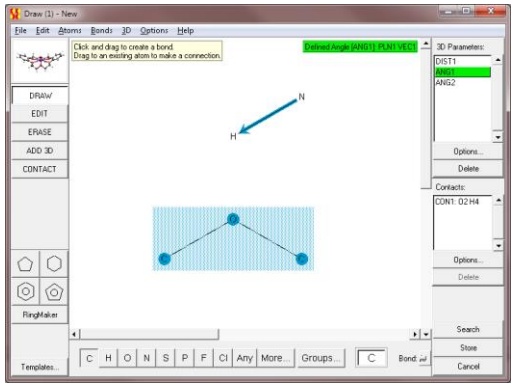


PreQuest: Create (in-house) database

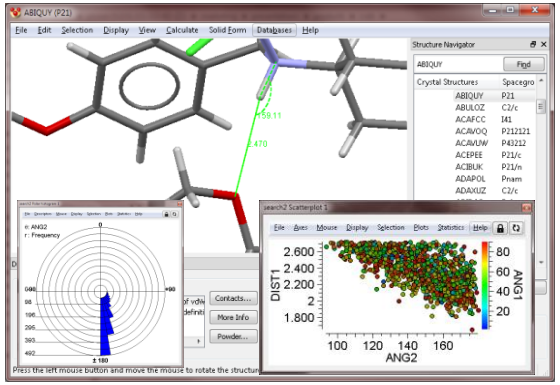
Cambridge Structural Database



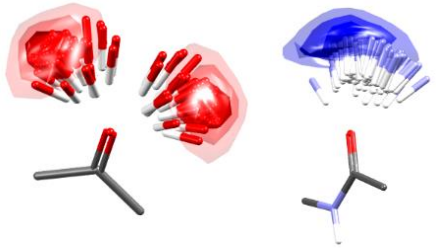
ConQuest: Advanced 3D searching



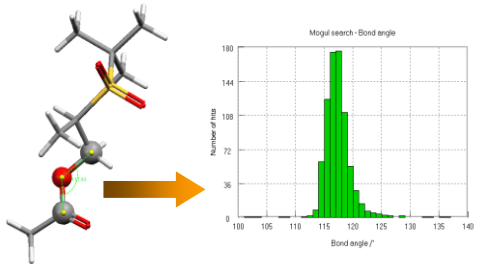
Mercury: Visualisation & data analysis



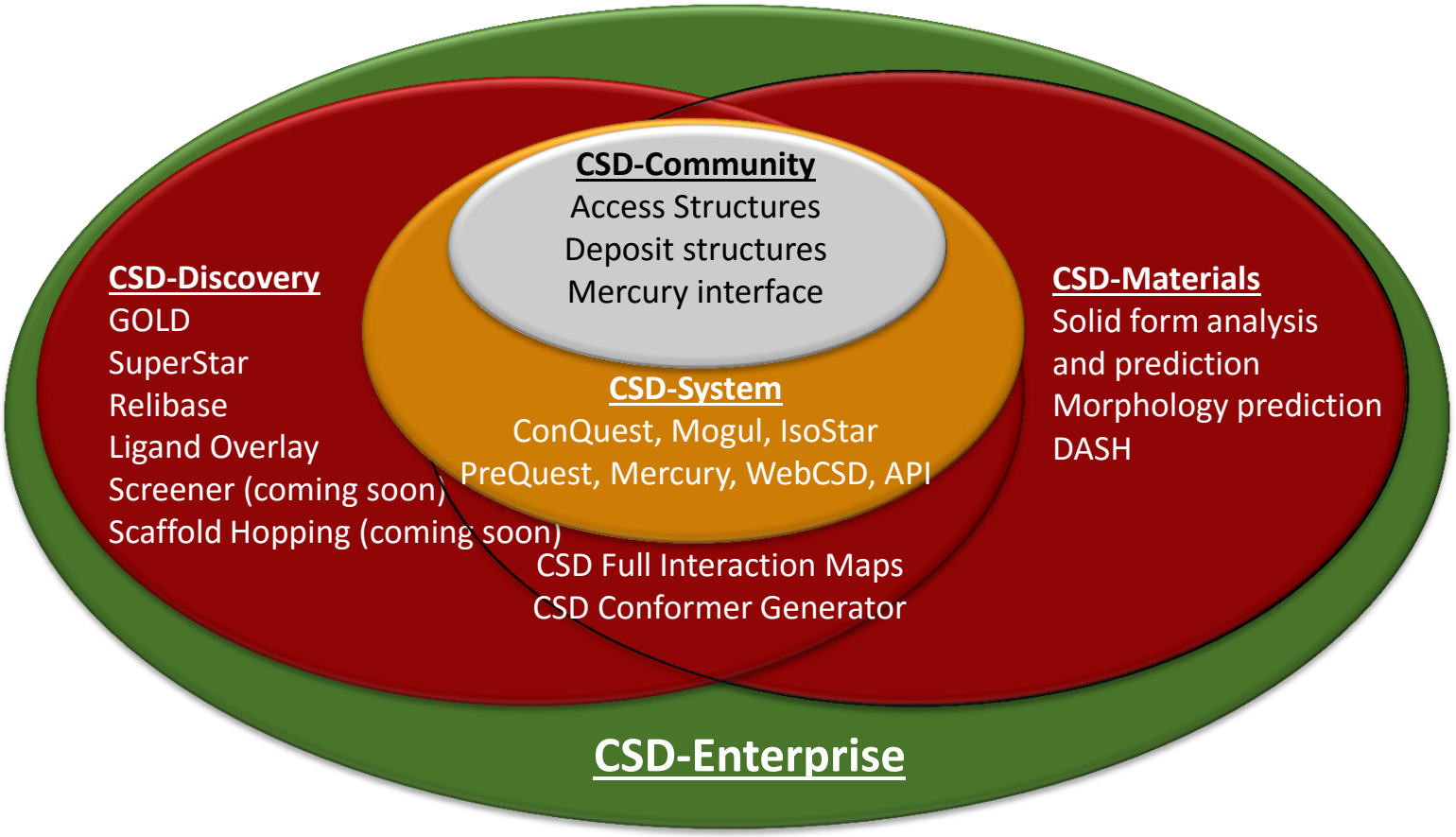
IsoStar: Molecular interaction analysis



Mogul: Molecular geometry analysis

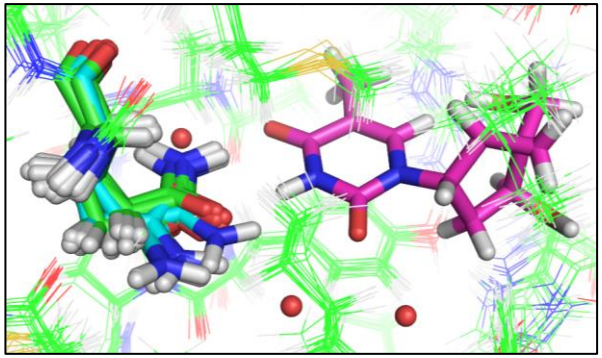


CSD-Enterprise

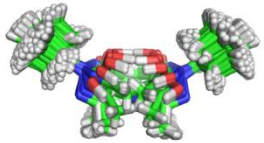


CSD-Discovery

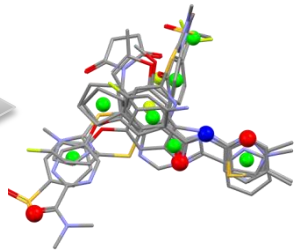
GOLD: Protein-ligand docking & virtual screening



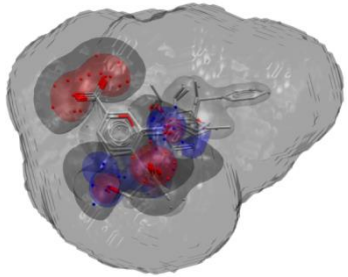
Ligand-based Virtual Screening:



Conformer Generation

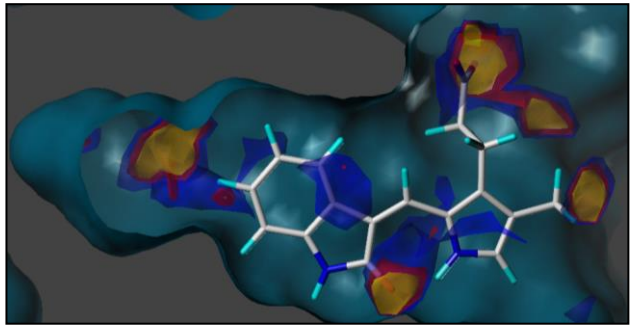


Ligand Overlay

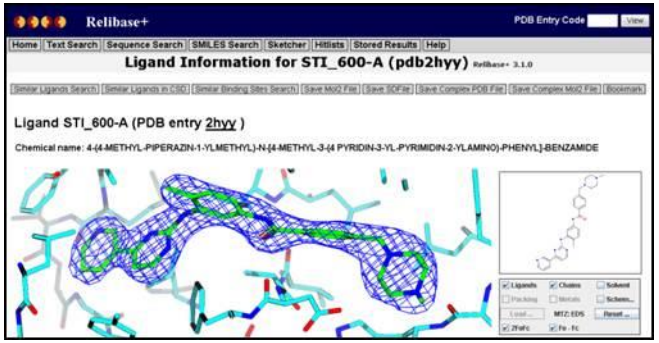


Field-based Screening (coming soon)

SuperStar: Interaction maps from crystallographic data

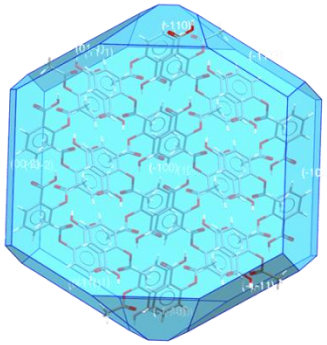


Relibase: Search and analyse protein-ligand structural data

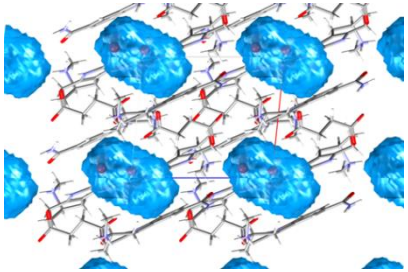


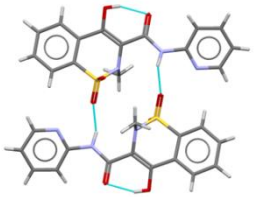
CSD-Materials

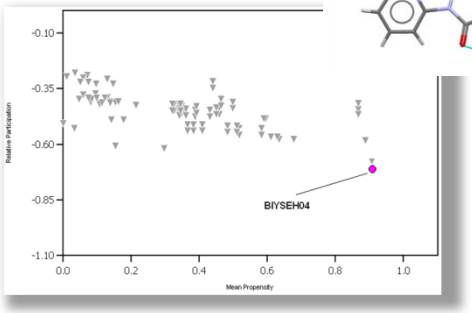
Tools to aid solid form design, including;



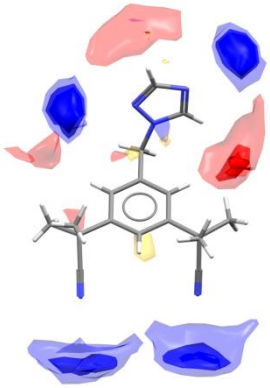
Morphology Prediction,
Hydrate Analyser



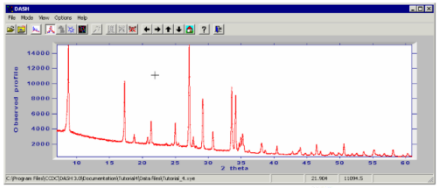


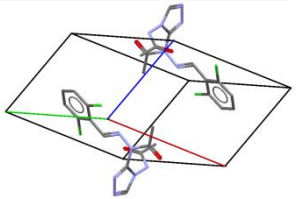


H-bond propensity analysis,
Full Interaction Maps

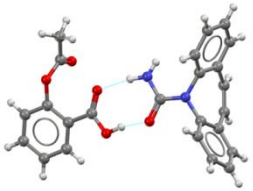


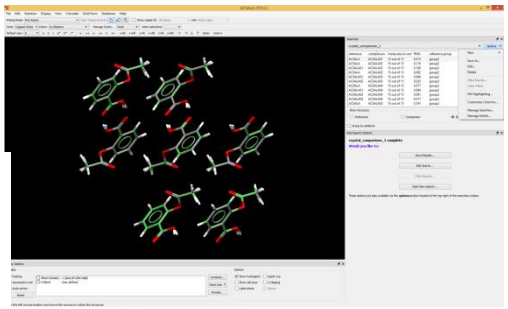
DASH: Powder structure solution





Crystal packing analysis, co-crystal formation

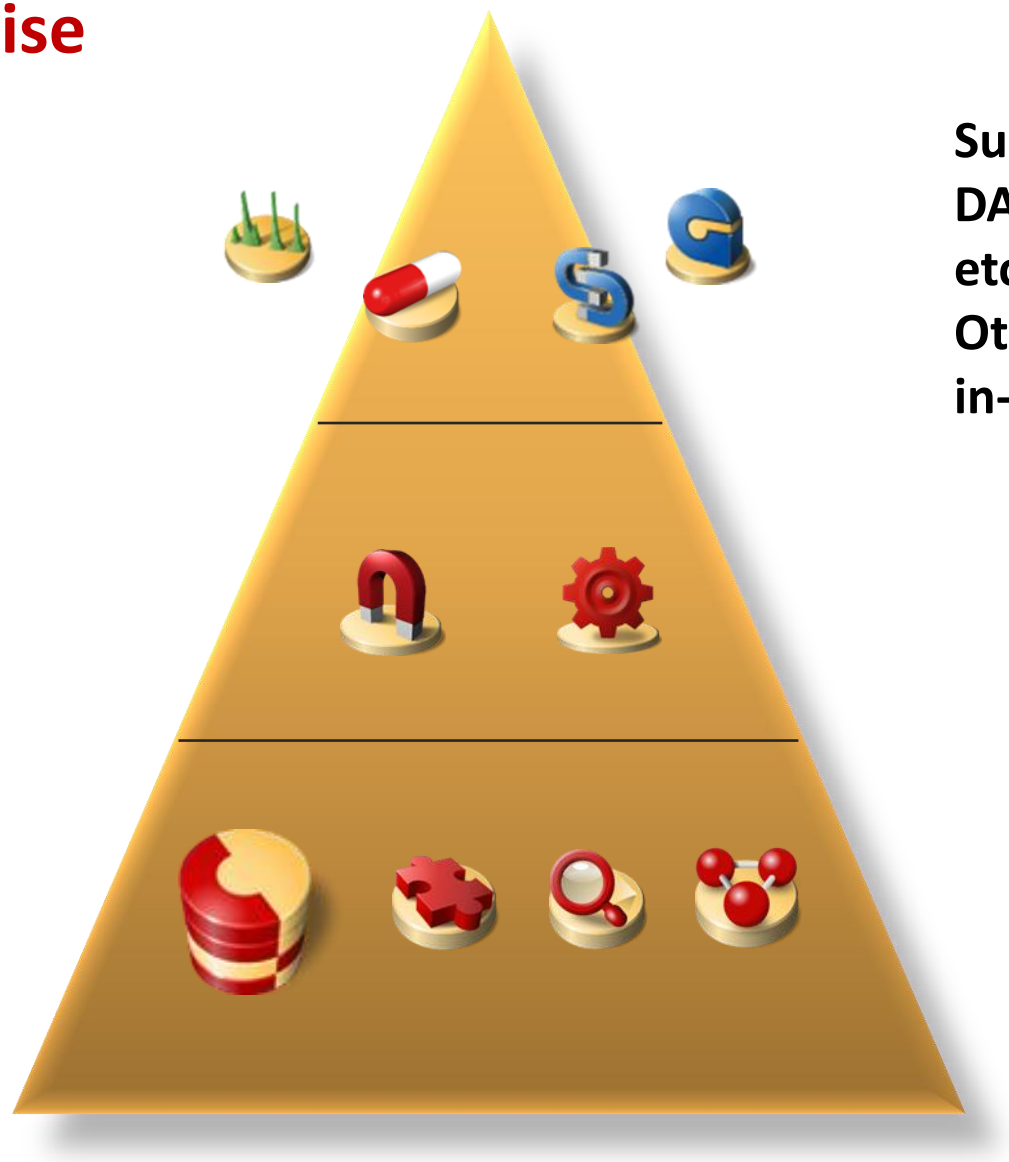




CSD-Enterprise

Applications

SuperStar, GOLD, DASH, Solid Form, etc.
Other vendor or in-house applications



Knowledge

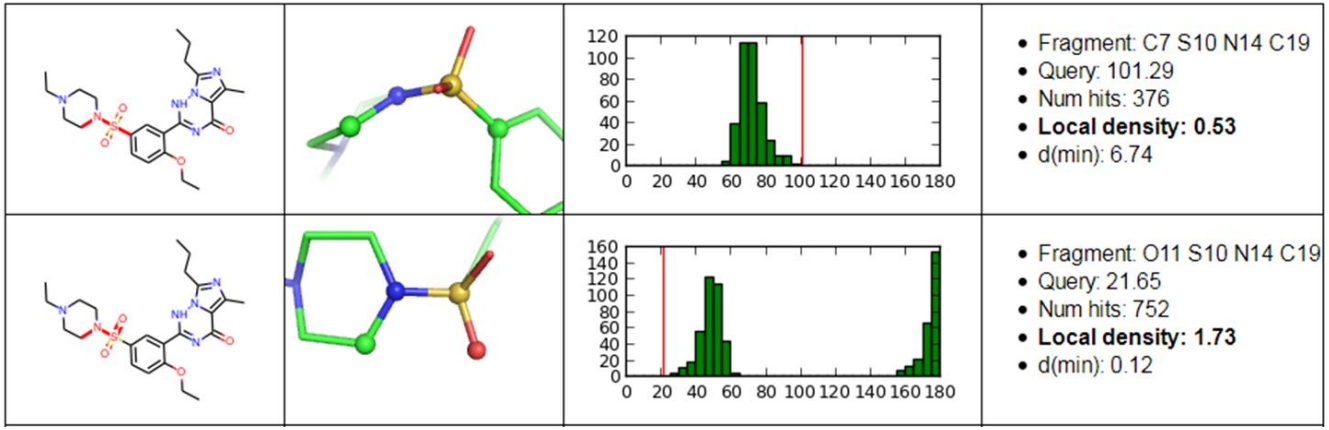
IsoStar, Mogul

Data

CSD

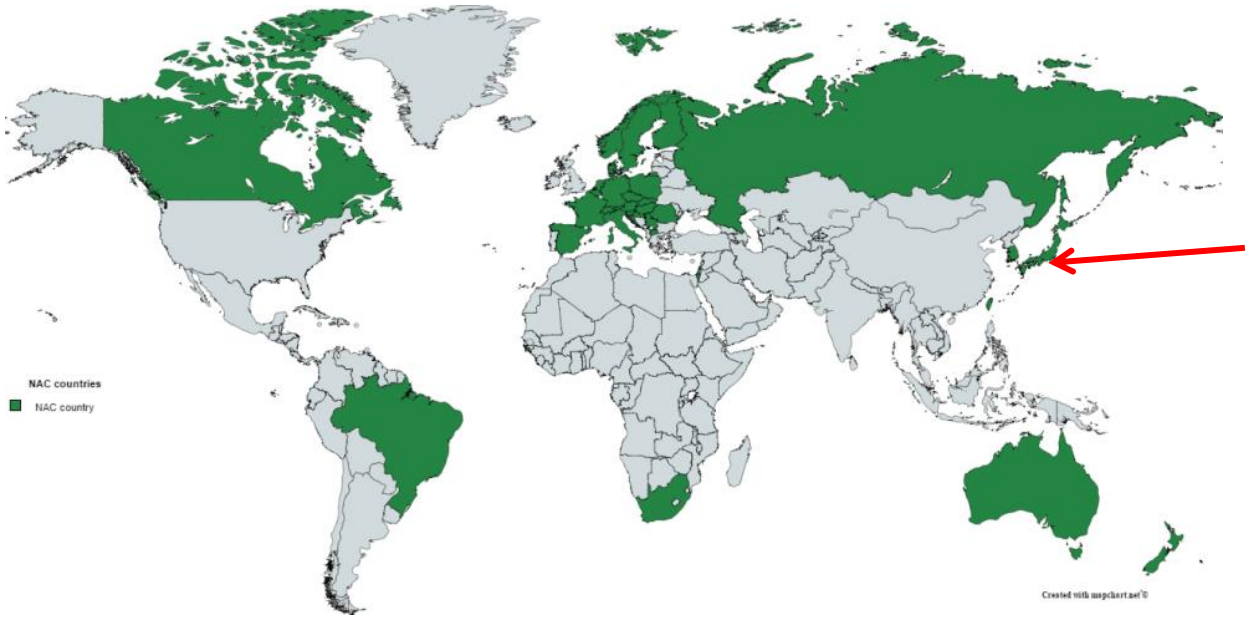
Python API

- Programmatic access to CSD functionality for other programs
- Fully documented, with “cookbook” examples
- Many complementary resources freely available



CSD-Enterprise

- Annual subscription model
- >1160 academic user sites in 80 different countries
- Direct from the CCDC or via one of 27 (2016) NACs
- Some “countrywide” arrangements (24 countries in 2016)

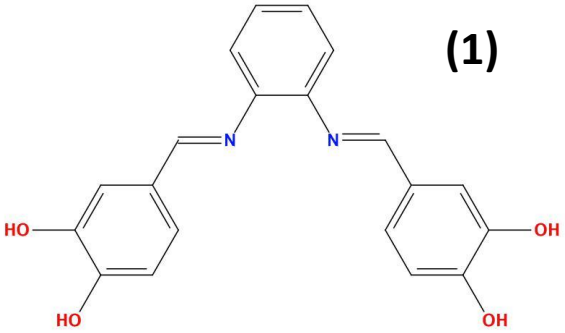


INSTITUTE for PROTEIN RESEARCH
OSAKA UNIVERSITY

For Japan, Prof. Nakamura and colleagues, Institute for Protein Research, Osaka University

Verification of Chemistry

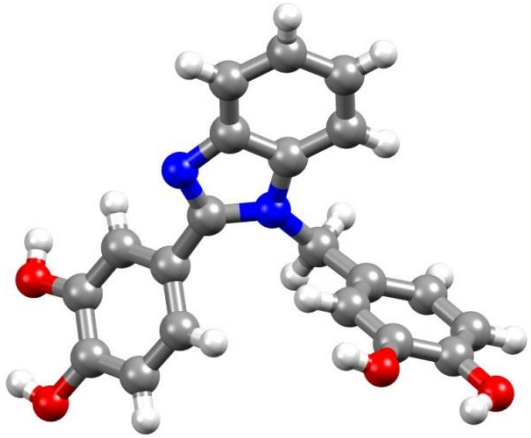
- Authors of a paper on G-quadruplex DNA cross-linking had assigned a crucial compound as shown as (1)
- Authors of a subsequent paper found chemical evidence, including single crystal structure solution, that the structure is in fact as shown as (2), cyclisation having occurred



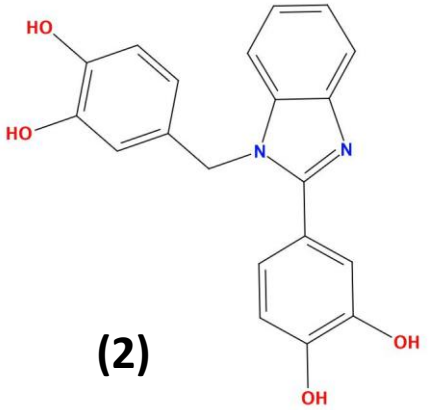
(1)

L.B. Yuan *et al*; Sci. Rep. 3, (2013), 1811, doi: 10.1038/srep01811

P.E. Reyes-Gutierrez *et al*; Sci. Rep. 6, (2016), 23499, doi: 10.1038/srep23499



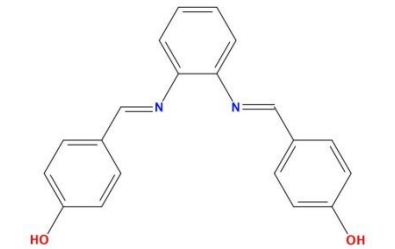
EKIGUC

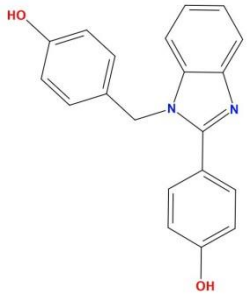


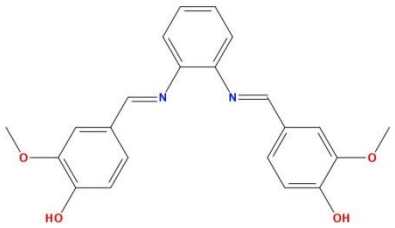
(2)

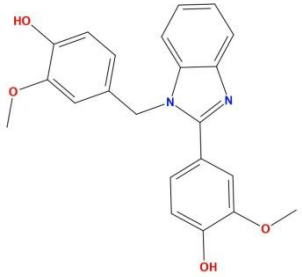
Verification of Chemistry

- More examples were quickly found by Reyes-Gutierrez et al.



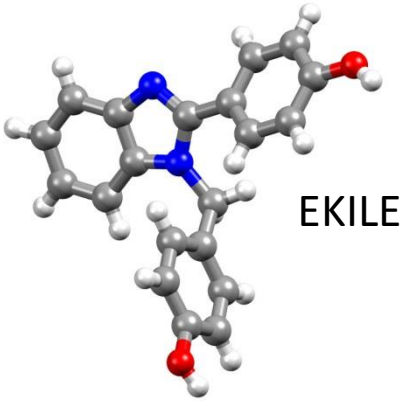




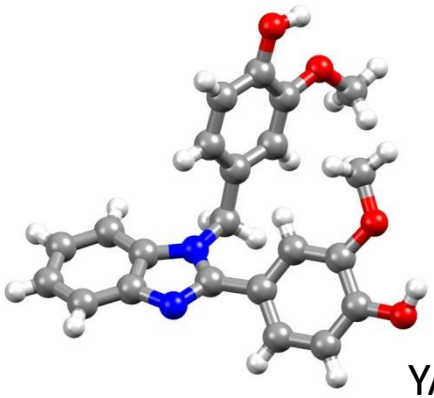


original

corrected



EKILER

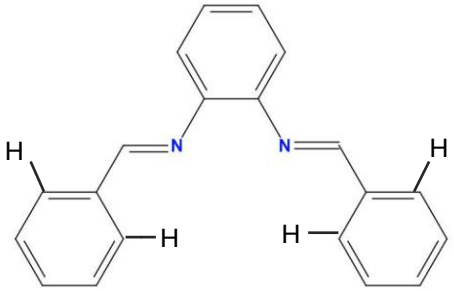


YAHSEH

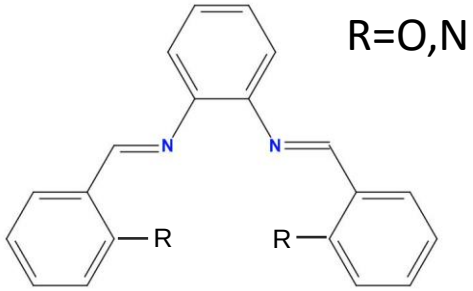
P.E. Reyes-Gutierrez *et al*; Sci. Rep. 6, (2016), 23499, doi: 10.1038/srep23499

Verification of Chemistry

- Searches of the CSD provided further evidence



0 hits



71 hits

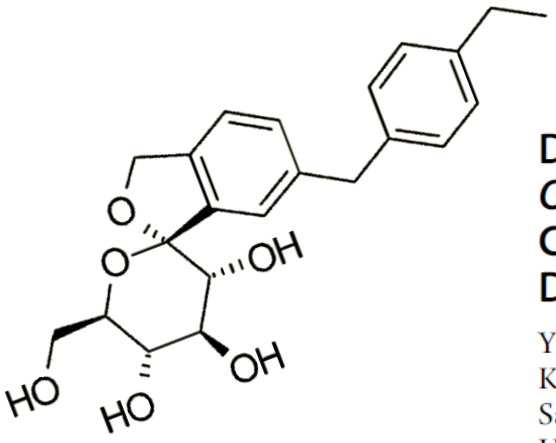
(Any substitution allowed at all other positions in both cases, apart from further cyclisation)

OH or NH in the ortho-position provides stabilisation of the bis-imine form by intramolecular H-bond – otherwise cyclisation occurs.

P.E. Reyes-Gutierrez *et al*; Sci. Rep. 6, (2016), 23499, doi: 10.1038/srep23499

Insight in Medicinal Chemistry

- Tofogliflozin is approved in Japan for treatment of Type 2 diabetes (approval pending in US and EU)
- Design driven by a CSD pharmacophore search



Journal of Medicinal Chemistry

Artide

pubs.acs.org/jmc

Discovery of Tofogliflozin, a Novel *C*-Arylglucoside with an *O*-Spiroketal Ring System, as a Highly Selective Sodium Glucose Cotransporter 2 (SGLT2) Inhibitor for the Treatment of Type 2 Diabetes

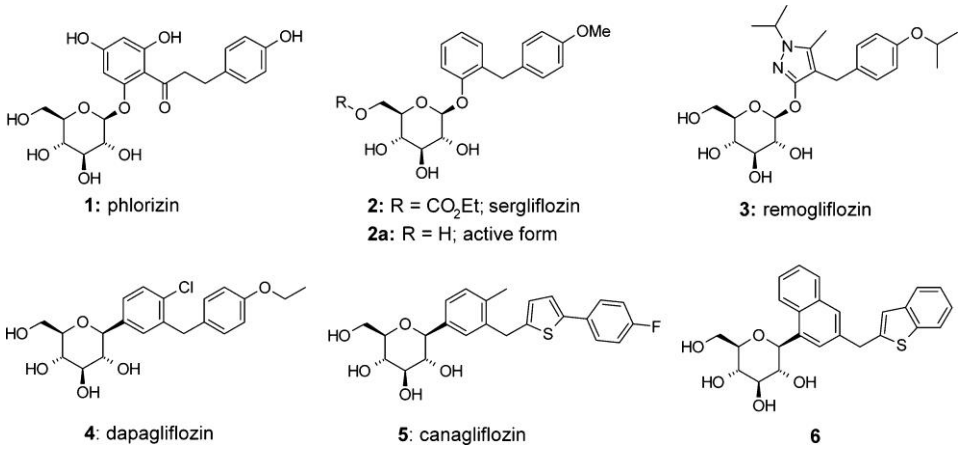
Yoshihito Ohtake,*[,†] Tsutomu Sato,[†] Takamitsu Kobayashi,[†] Masahiro Nishimoto,[†] Naoki Taka,[†] Koji Takano,[†] Keisuke Yamamoto,[†] Masayuki Ohmori,[†] Marina Yamaguchi,[†] Kyoko Takami,[†] Sang-Yong Yeu,[§] Koo-Hyeon Ahn,[‡] Hiroharu Matsuoka,[†] Kazumi Morikawa,[†] Masayuki Suzuki,[†] Hitoshi Hagita,[§] Kazuharu Ozawa,[†] Koji Yamaguchi,[†] Motohiro Kato,[†] and Sachiya Ikeda[†]

[†]Research Division, Chugai Pharmaceutical Co., Ltd., 1-135 Komakado, Gotemba, Shizuoka 412-8513, Japan

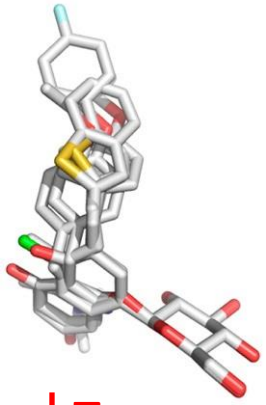
[§]Chugai Research Institute for Medical Science, Inc., 1-135 Komakado, Gotemba, Shizuoka 412-8513, Japan

[‡]C&C Research Laboratories, DRC Natural Sciences Campus, Sungkyunkwan University, 300 Cheoncheon-dong, Jangan-gu, Suwon, Gyeonggi-do 440-746, Republic of Korea

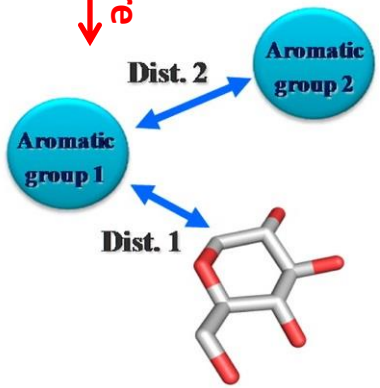
Insight in Medicinal Chemistry



Overlay molecules



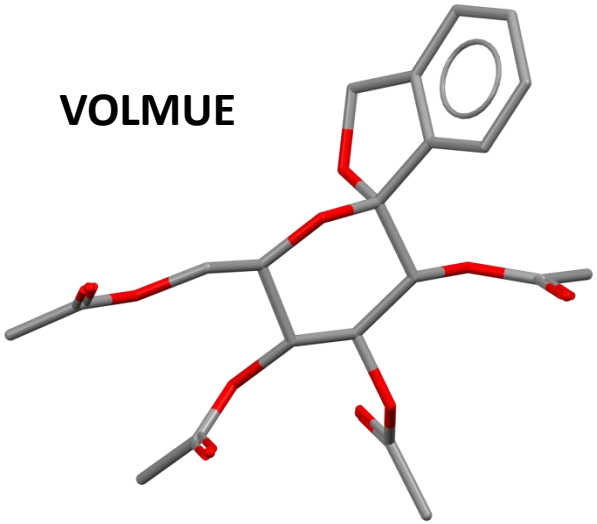
Pharmacophore



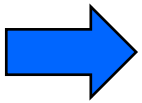
- Known SGLT2 inhibitors are similar molecules – substituted glucose derivatives
- Pharmacophore derived by molecular overlay / alignment
- Pharmacophore translated into CSD search…

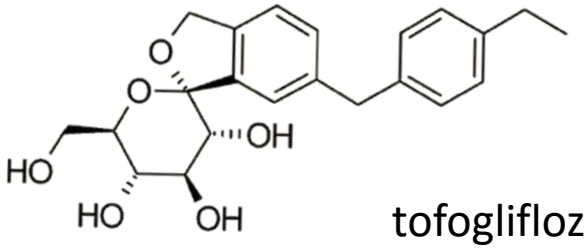
Insight in Medicinal Chemistry

A number of structures in the hitlist showed an interesting alternative chemistry (a spiro centre), and one, VOLMUE, was selected as the basis for further development.

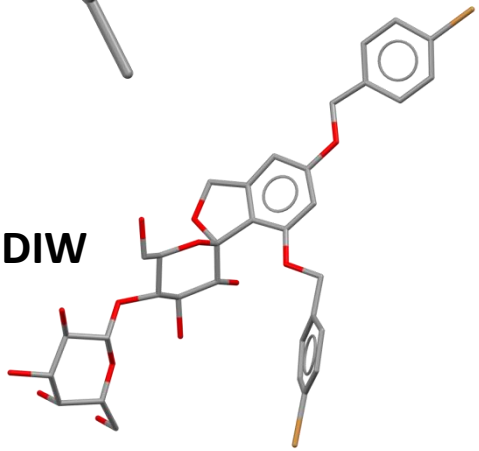


VOLMUE





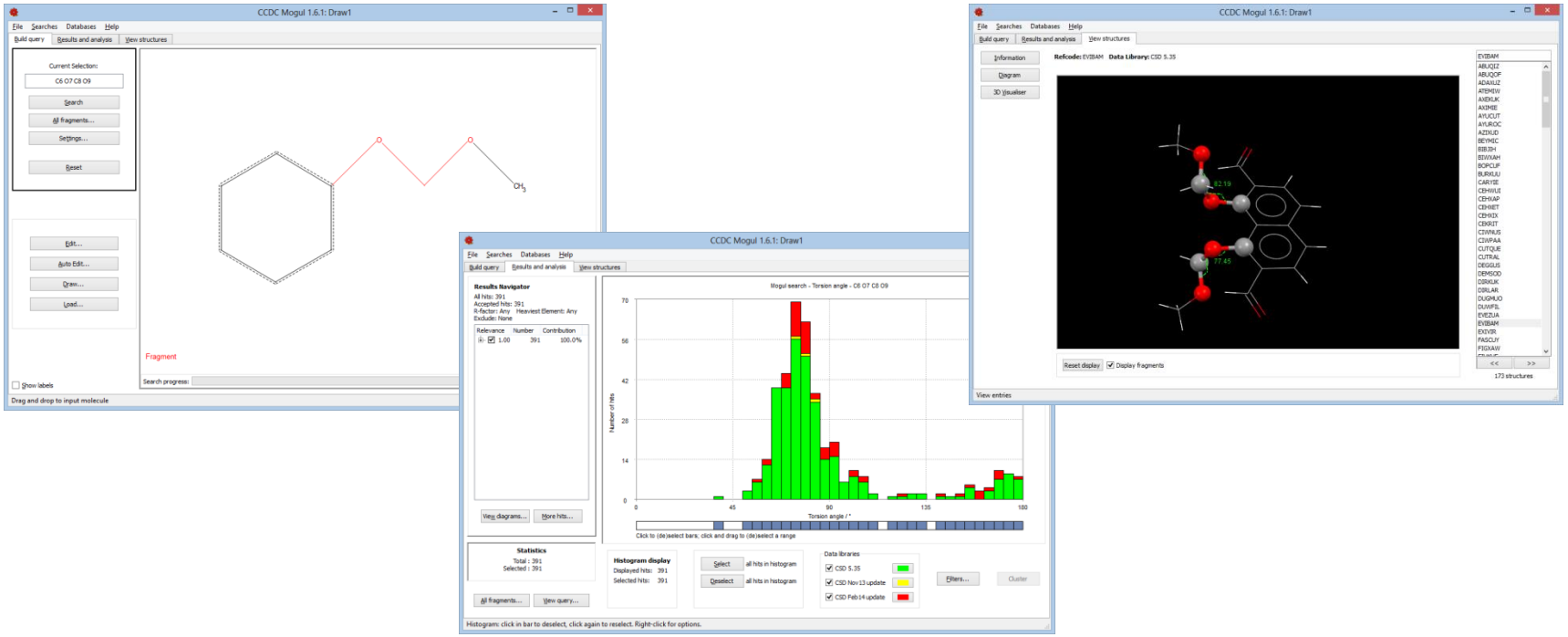
tofogliflozin



BAGDIW

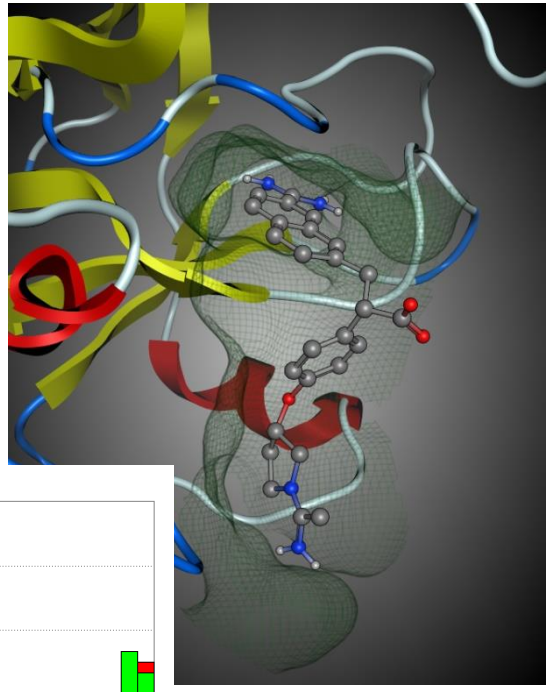
Mogul – a library of molecular geometry

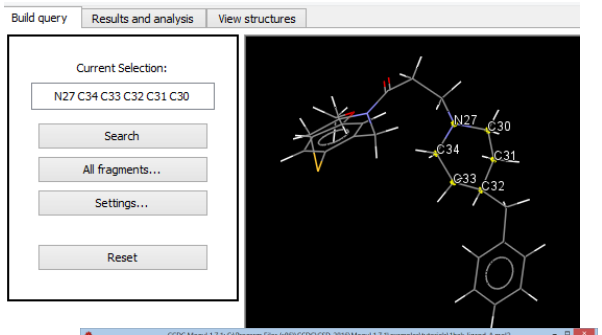
- Mogul contains geometrical information (bond lengths, angles and torsions) derived from the >820k structures (May 2016) in the Cambridge Structural Database

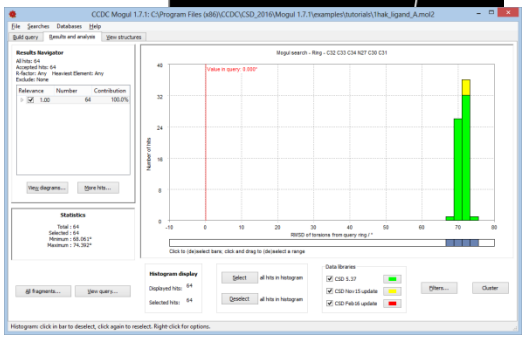


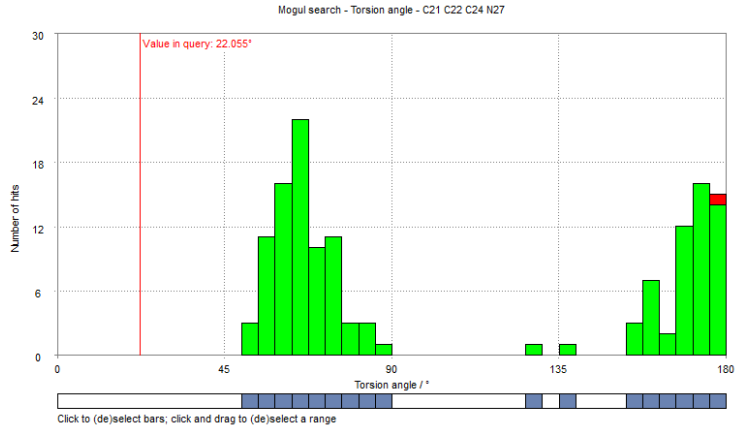
Analyse Ligand Geometry

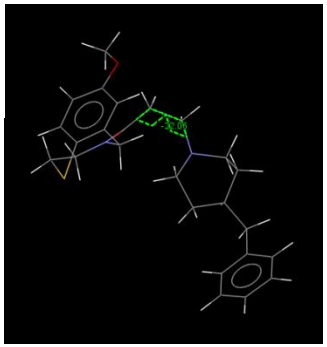
- Are ligand geometries as expected? Have they been poorly refined?





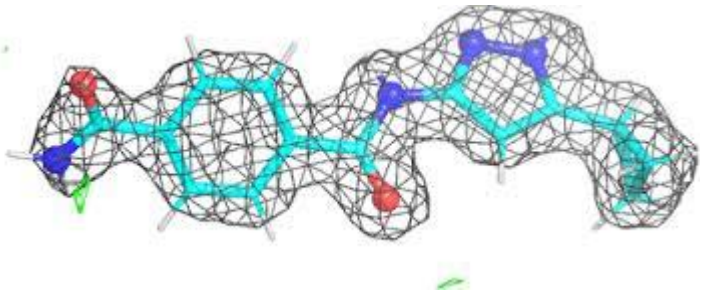


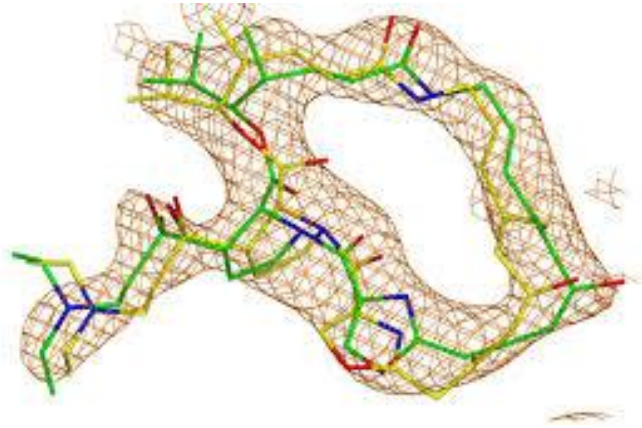




Constraints for complex structure solution

- Direct link to Mogul in Global Phasing's GRADE / BUSTER software
- Particularly useful in refining into poorly resolved electron density in pocket space, or resolving disorder





- Work underway to extend this link to other software packages (Phenix, COOT)

Conclusions

- The Cambridge Structural Database is ever-growing, and at an increasing rate
- CSD-Enterprise contains all the data and an extensive collection of programs to enable use and application of crystallographic knowledge
- Worldwide application in Discovery and Materials sciences, in industry and in academia
- Available in Japan from JAICI (industry), Institute for Protein Research at the University of Osaka (academia)

JAICI
化学情報協会

INSTITUTE for PROTEIN RESEARCH
OSAKA UNIVERSITY

JAICI 化学情報協会

INSTITUTE for PROTEIN RESEARCH
OSAKA UNIVERSITY

Thanks for your attention!

www.ccdc.cam.ac.uk

maginn@ccdc.cam.ac.uk