One Million Structures and Counting

The journey, the insights, and the future of the CSD

Suzanna Ward

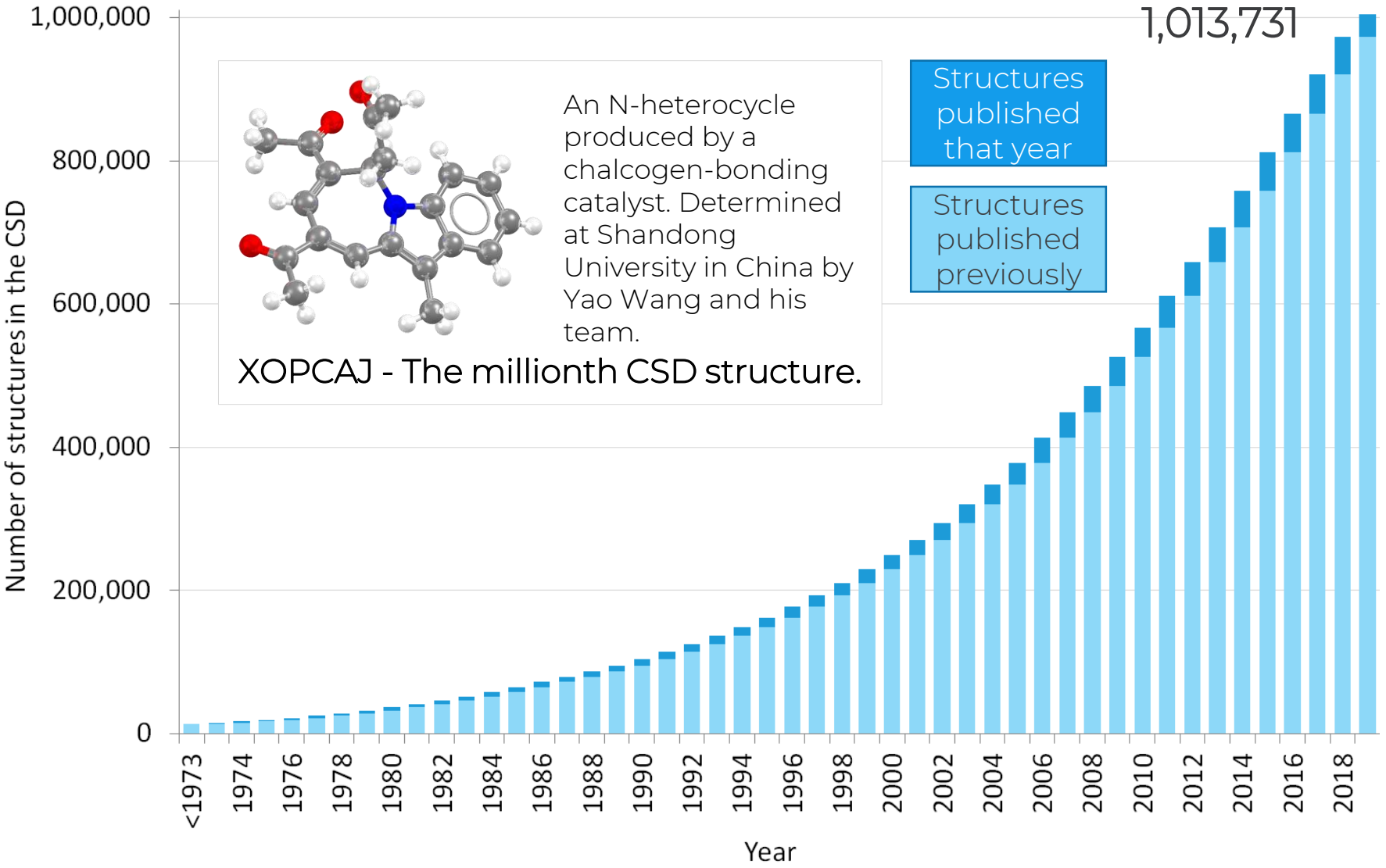
The Cambridge Crystallographic Data Centre

ECM32 – Wednesday 21st August 2019

CCDC

advancing structural science

The Cambridge Structural Database (CSD)



An N-heterocycle produced by a chalcogen-bonding catalyst. Determined at Shandong University in China by Yao Wang and his team.

XOPCAJ - The millionth CSD structure.

- Every published structure
  - Inc. ASAP & early view
  - *CSD Communications*
  - Patents
  - University repositories
- Every entry enriched and annotated by experts
- Discoverability of data and knowledge
- Sustainable for over 54 years

CCDC

Inside the CSD

Organic
43%

Metal-Organic
57%

At least one transition metal, lanthanide, actinide or any of Al, Ga, In, Tl, Ge, Sn, Pb, Sb, Bi, Po

Not Polymeric
89%

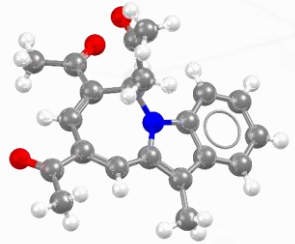
Polymeric: 11%

Single Component
56%

Multi Component
44%

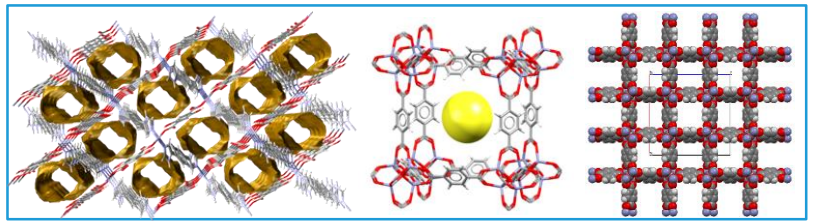
Organic

- Drugs
- Agrochemicals
- Pigments
- Explosives
- Protein ligands



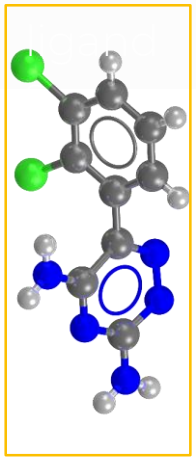
Metal-Organic

- Metal Organic Frameworks
- Models for new catalysts
- Porous frameworks for gas storage
- Fundamental chemical bonding



Additional data

- 10,860 polymorph families
- 169,218 melting points
- 840,667 crystal colours
- 700,002 crystal shapes
- 23,622 bioactivity details
- 9,740 natural source data
- > 250,000 oxidation states



Links/subsets

- Drugbank
- Druglike
- MOFs
- PDB ligands
- PubChem
- ChemSpider
- Pesticides

CCDC

1965

The sound of music

- Film first released in 1965
- Highest grossing film of 1965
- Set in Austria

CCDC

The 1960s

Credits: NASA

CCDC

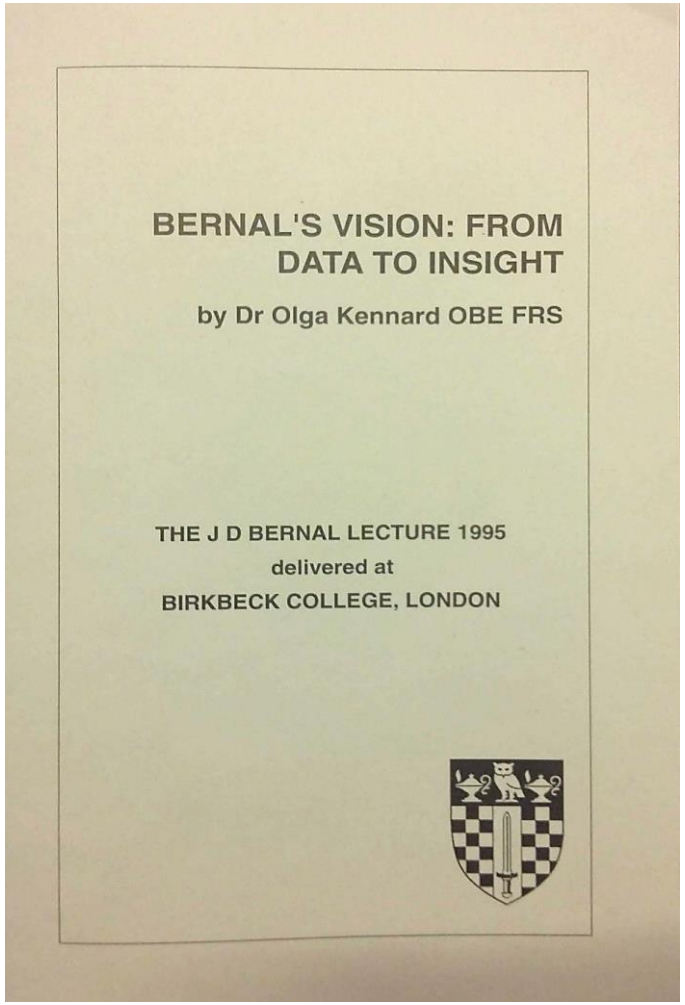
The vision

J.D. Bernal and research group including Olga Kennard at Stonehenge in 1948

- Established in 1965 by Olga Kennard
- She and J.D. Bernal had a vision that a collective use of data would lead to new knowledge and generate insights

CCDC

The vision



BERNAL'S VISION: FROM DATA TO INSIGHT

by Dr Olga Kennard OBE FRS

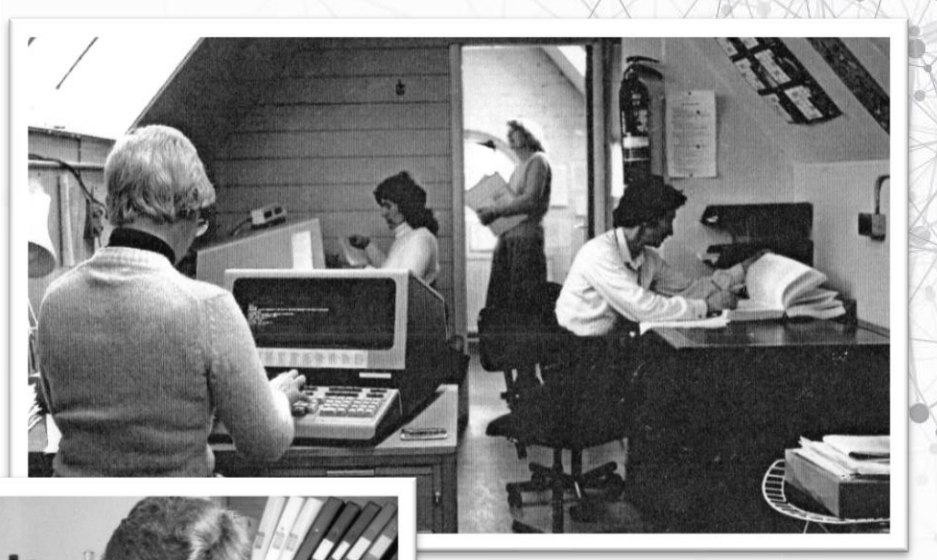
THE J D BERNAL LECTURE 1995
delivered at
BIRKBECK COLLEGE, LONDON

We clearly recognised even in those early days, that data banks have three principal functions. Firstly they must gather together existing knowledge and make it readily available to the scientific community. Secondly they can be used to reduce a large number of observations to a small set of constants and rules, and in this way transform a data base to a knowledge base. Such a knowledge base may obviate the need for further individual experiments in specific areas. Thirdly, they facilitate the comparison and collective analysis of individual results to gain insight into new or as yet unexplained phenomena. These ideas have been at the heart of the work of the Cambridge Crystallographic Data Centre and the driving force for improving methods of data collection, storage and dissemination. Most importantly they influenced development of computer programs and methodologies which are needed for the analysis and transformation of the accumulated information. (5)

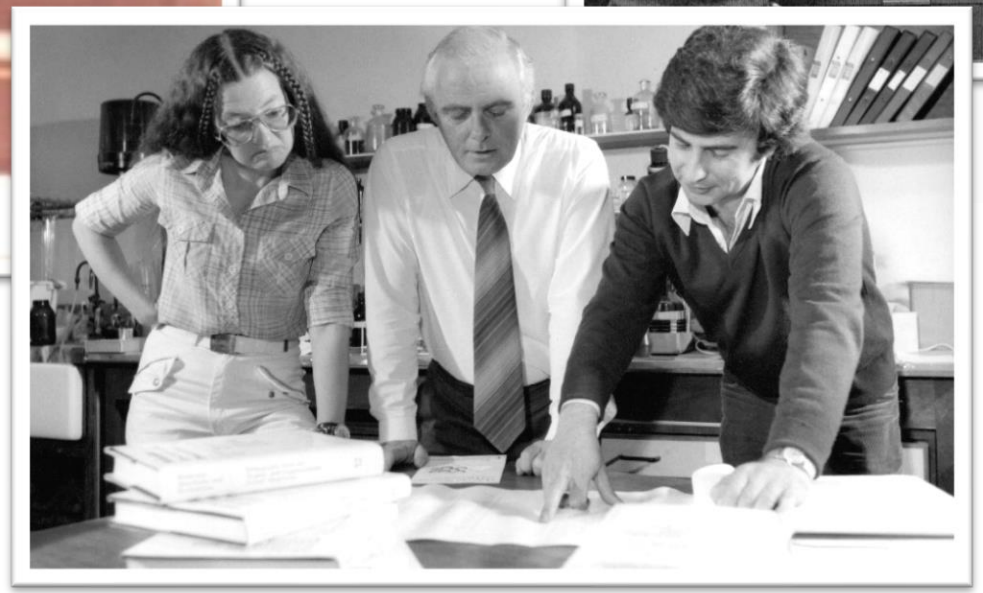
CCDC

Early creation of the CSD

Olga Kennard, David Watson and Sam Motherwell



The early days of CSD creation



Sharon Bellard, David Watson and Frank Allen

CCDC

The beginnings of the CSD

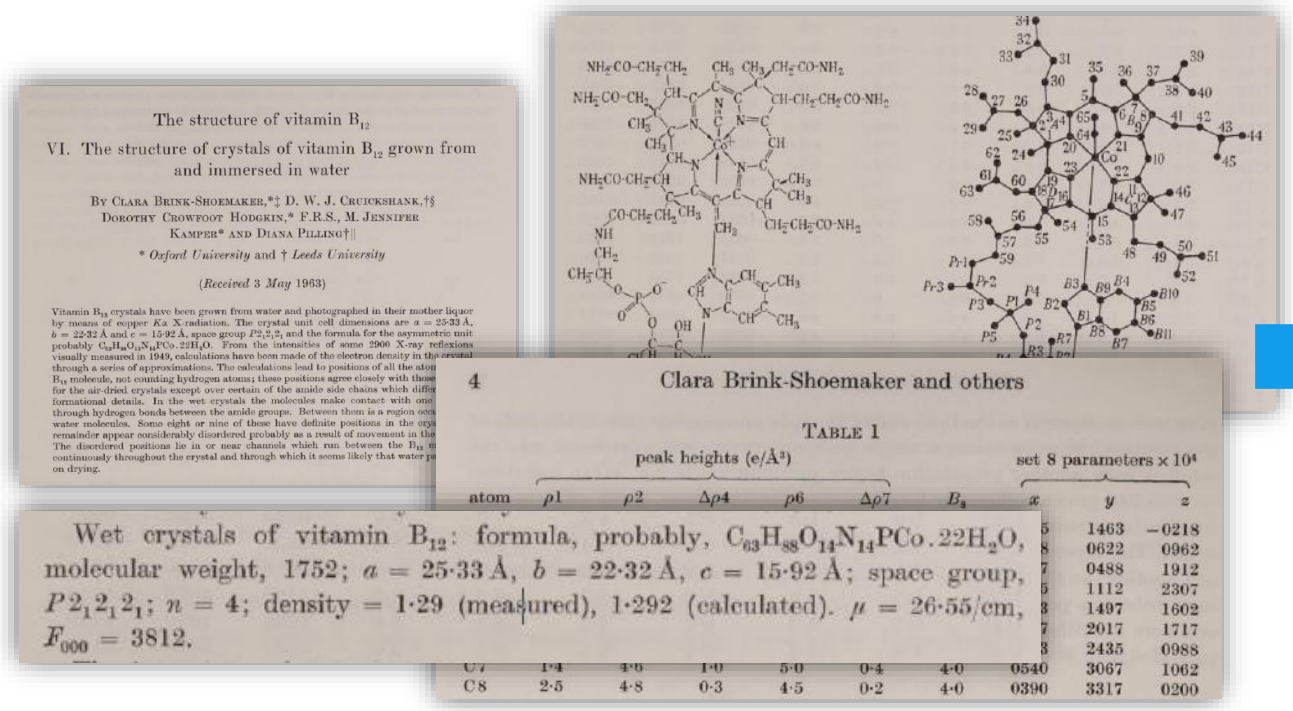
Olga Kennard, CSD 50 symposium, Cambridge 2015

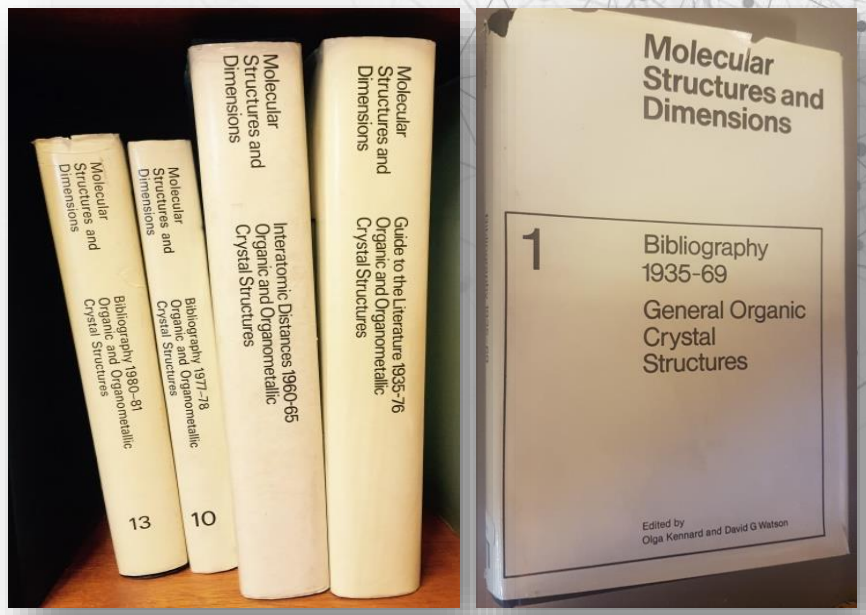
The importance of data quality recognised from the outset

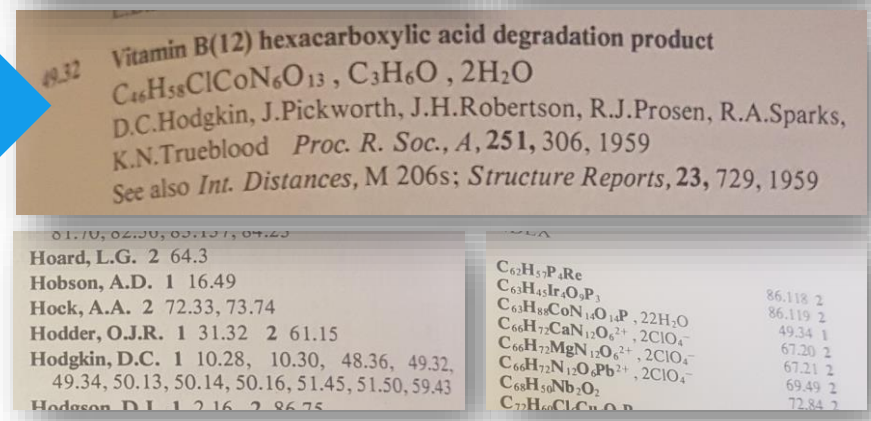
CCDC

The first volumes

- Data primarily reported within articles
- Volumes electronically typeset
- Bibliographic information and introduced rudimentary ways of searching







Clara Brink-Shoemaker, D.W.J. Cruickshank, D.M.Crowfoot Hodgkin, M.J.Kamper and D,Pilling *Proceedings of the Royal Society London,Series A*, 1964, 278, 1, DOI: 10.1098/rspa.1964.0042

CCDC

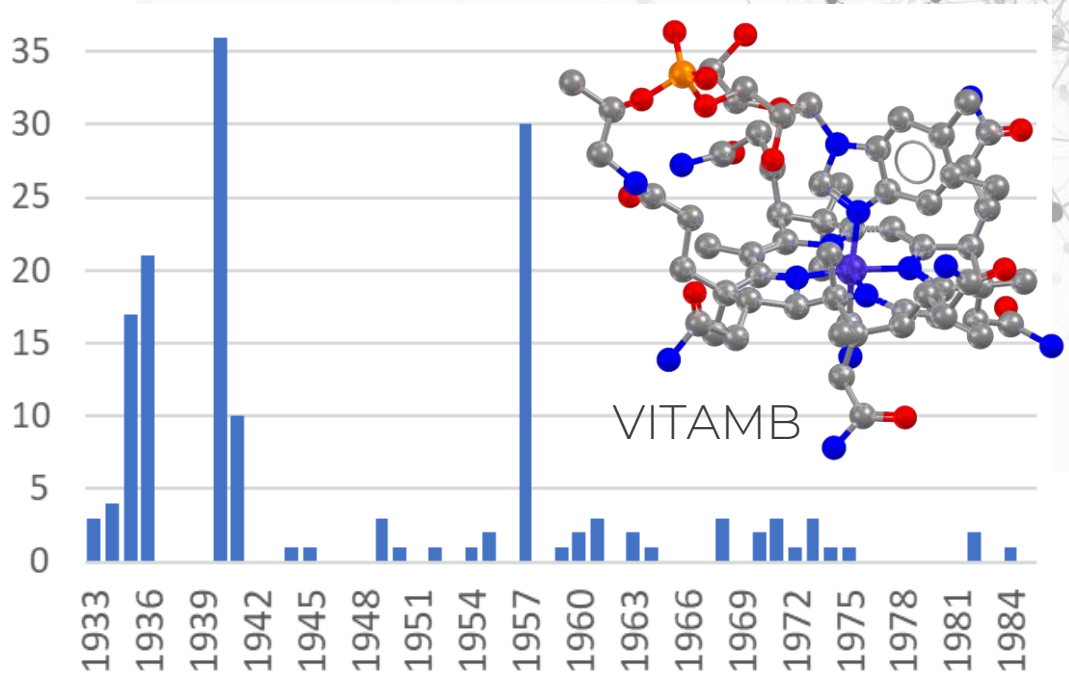
Dorothy Hodgkin

Olga, Dorothy and a Benjamin Franklin look alike at the ACA Philadelphia in 1988

- Dorothy Hodgkin won a Nobel Prize in Chemistry in 1964

Structure of vitamin B12 and other complex molecules (penicillin, insulin)

- Dorothy has >150 structures in the CSD

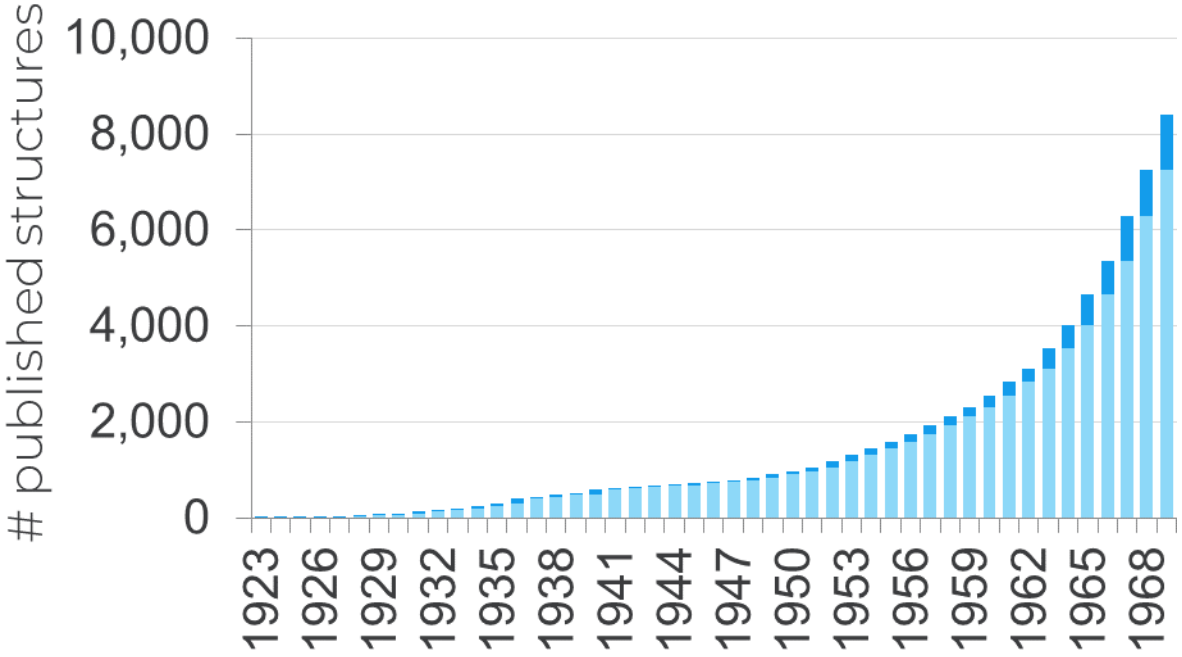


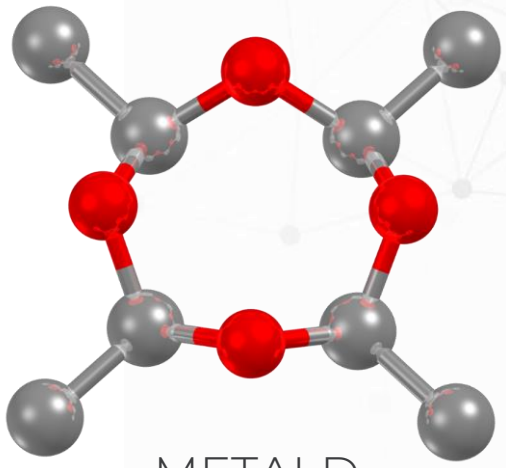
- Now >130 vitamin structures

CCDC

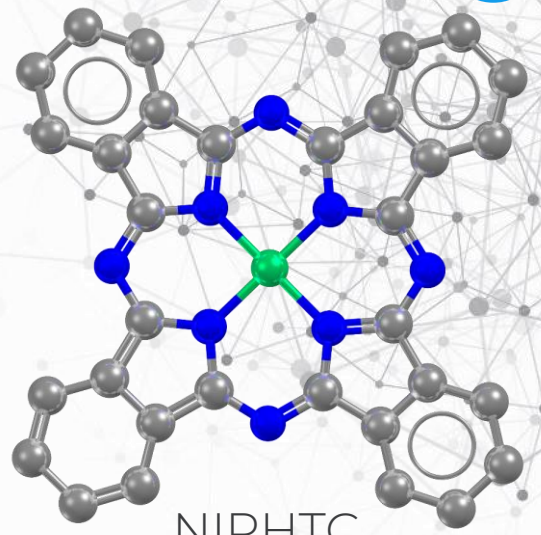
Up to 1969

- 4,661 structures published < 1965
- >8,000 structures published by 1969

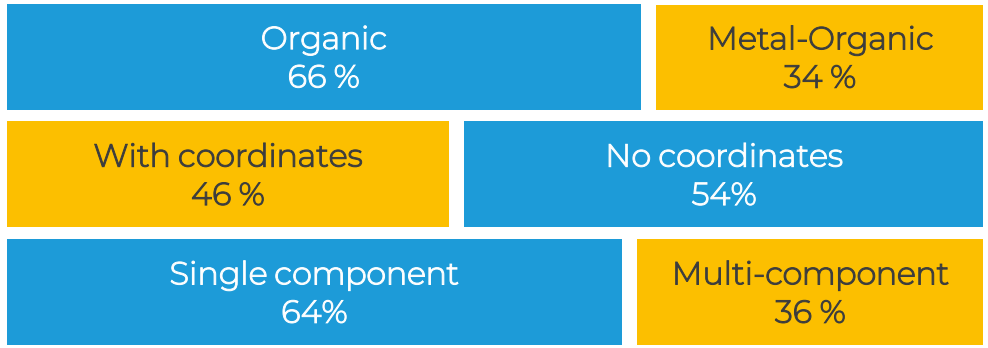




METALD
One of the first organic structures with 3D coordinates



NIPHTC
The first metal-organic structure with 3D coordinates



L.Pauling, D.C.Carpenter, *J.Am.Chem.Soc.*, 1936, 58, 1195, J.M.Robertson, I.Woodward, *J.Chem.Soc.*, 1937, 219

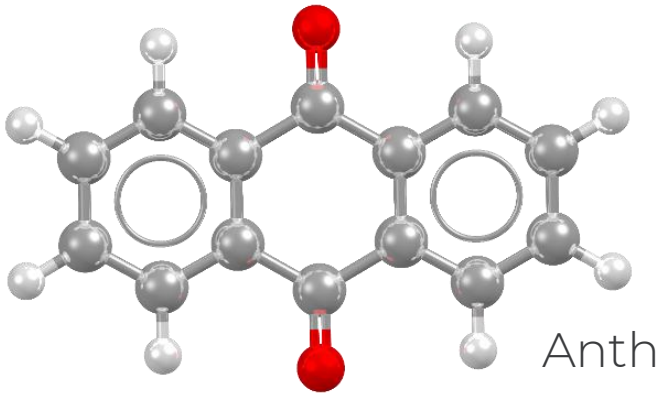
CCDC

Benzene and Kathleen Lonsdale

By Smithsonian Institution from United States via Wikimedia Commons

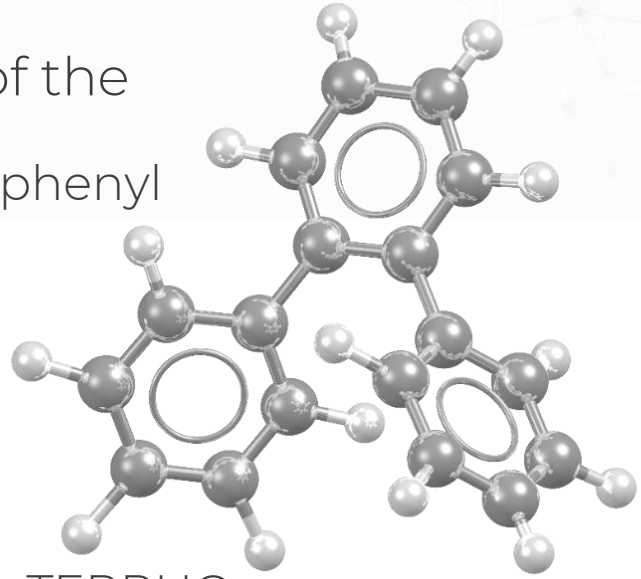
- >1,600 published benzene structures by 1965
- Kathleen established the planarity of the benzene ring by X-ray crystallography
  - She has 19 entries in the CSD
  - Was responsible for co-creating the first edition of the International Tables

ANTQUO05



Anthraquinone

1,1':2',1''-terphenyl



TERPHO
Kathleen's first CSD entry with coords

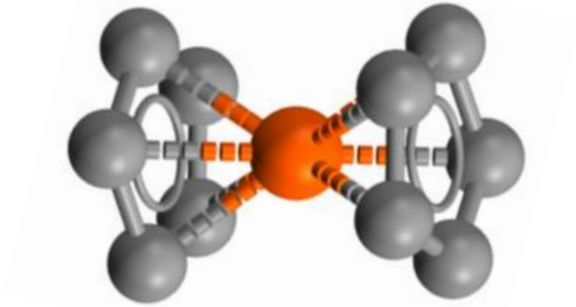
C.J.Birkett-Clews, K.Lonsdale, Proc.R.Soc.London,Ser.A (1937), 161, 493
K.Lonsdale, H.J.Milledge, K.E.Sayed, Acta Crystallogr. (1966), 20, 1

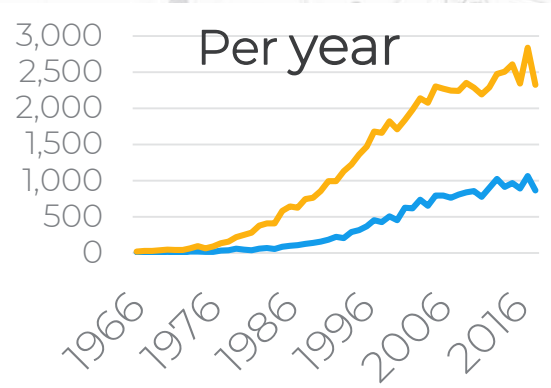
CCDC

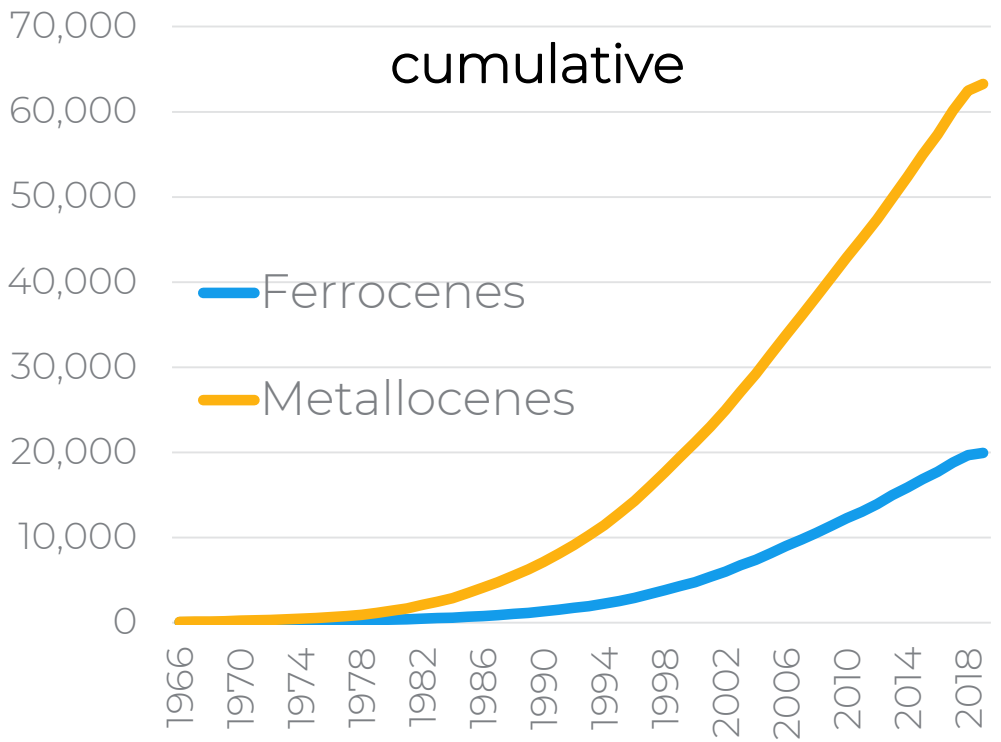
The rise of ferrocenes

- Ferrocene was first discovered in 1951
- Structure determined independently by three groups in 1953
- 33 structures published by 1965
- Metallocene now in 11 % of all metal-organic structures

FEROCE01



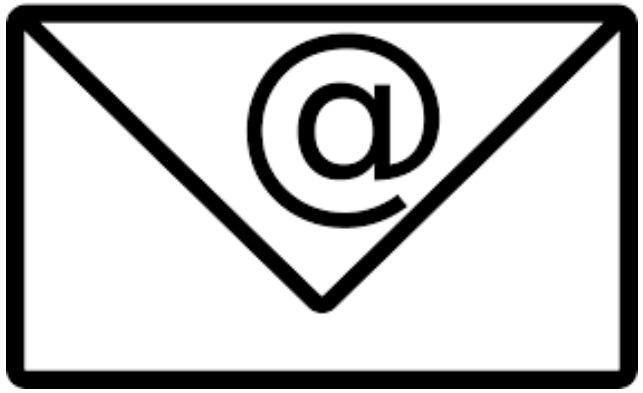




J.D.Dunitz, L.E.Orgel, A.Rich, *Acta Crystallographica*, 1956, 9, 373, DOI: 10.1107/S0365110X56001091

CCDC

The 1970s



```c
#include <stdio.h>
int prime(int n);
int main()
{
    int n, i, flag=0;
    printf("Enter a positive integer: ");
    scanf("%d",&n);
    for(i=2; i<=n/2; ++i)
    {
        if (prime(i)!=0)
        {
            if ( prime(n-i)!=0)
            {
                printf("%d = %d + %d\n", n, i, n-i);
                flag=1;
            }

        )
    }
    if (flag==0)
      printf("%d can't be expressed as sum of two prime numbers.",n);
    return 0;
}
int prime(int n)      /* Function to check prime number */
{
    int i, flag=1;
    for(i=2; i<=n/2; ++i)
        if(n%i==0)
            flag=0;
    return flag;
}
```

https://www.techrepublic.com/pictures/tech-nostalgia-the-top-10-innovations-of-the-1970s/

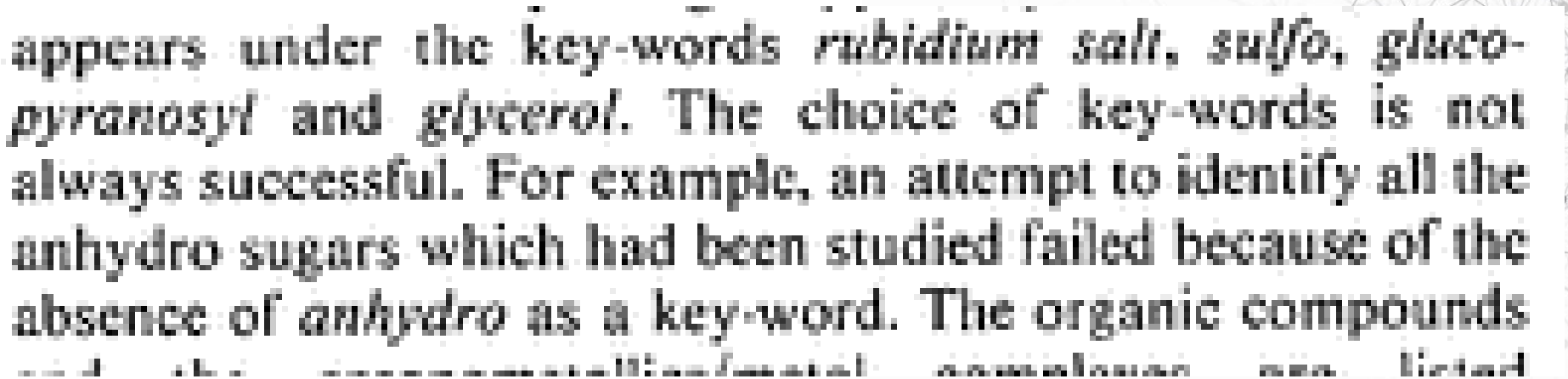
CCDC

Molecular structures and dimensions: guide to the literature, 1935–76: organic and organometallic crystal structures. Edited by O. KENNARD, F. H. ALLEN and D. G. WATSON. Pp. xxiii + 660. Utrecht: Bohn, Scheltema & Holkema, 1977. Price $66.00, £35·00 (personal rate $49.00, £25·50).

Some folk like their data on magnetic tape or disc, others prefer microfiche, but the majority of scientists are stubbornly conservative with respect to devices outside their own special field and still prefer a good solid book. This should please them. It weighs 2·4 kg and its 660 pages are contained in a volume of 32 × 23 × 5 cm. Apart from the purpose which is stated in its title, this book can be used to support apparatus, prop open office doors, press family photographs or wild flowers, and if erected vertically can conceal a post-prandial snooze. There is certainly a sense of money's worth, which might not be so from the equivalent ten microfiches weighing 14 g.

CCDC

More challenges …



appears under the key-words *rubidium salt*, *sulfo*, *gluco-pyranosyl* and *glycerol*. The choice of key-words is not always successful. For example, an attempt to identify all the anhydro sugars which had been studied failed because of the absence of *anhydro* as a key-word. The organic compounds

- Even then categorisation and finding structures created challenges …
  - How to **find** entries of interest?
    - How to detect **duplicate** entries?
    - CCDC helped pioneer work on reduced cell searching (Bob McMeeking, David Watson and others)
- The registration systems developed were critical for this

CCDC

This book, or its equivalent device, is essential for all scientists concerned with organic or organometallic molecular structures in the crystalline state. It removes every excuse for ignorance concerning the literature in this field.

Finally, I must say 'thank heavens' for Olga and her colleagues who had the good sense to start getting these structural data organized before they swamped us.

GEORGE A. JEFFREY

Department of Crystallography
University of Pittsburgh
Pittsburgh
PA 15260
USA

Using the data..

Table I. Experimental Structural Parameters for N···C=

Compound	l_1[c]	l_2[c]	d_1
A, methadone	1.510	1.554	2.910
B, cryptopine	1.521	1.515	2.581
C, protopine	1.520	1.505	2.555
D, clivorine	1.508	1.546	1.993
E, retusamine	1.49	1.56	1.64
F, *N*-brosylmitomycin A	1.56	1.56	1.49

[a] For a definition of symbols see Figure 1. [b] Distances i
1.3°. [c] $\langle l_i \rangle$ = 1.529 Å, $\langle \alpha \rangle$ = 107.3 (5.0)°, $\langle \gamma \rangle$ = 118.6 (1.

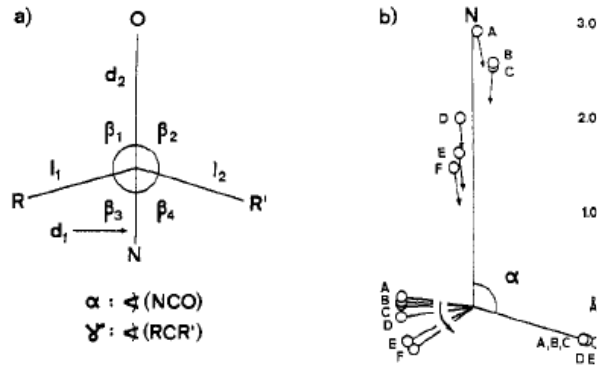
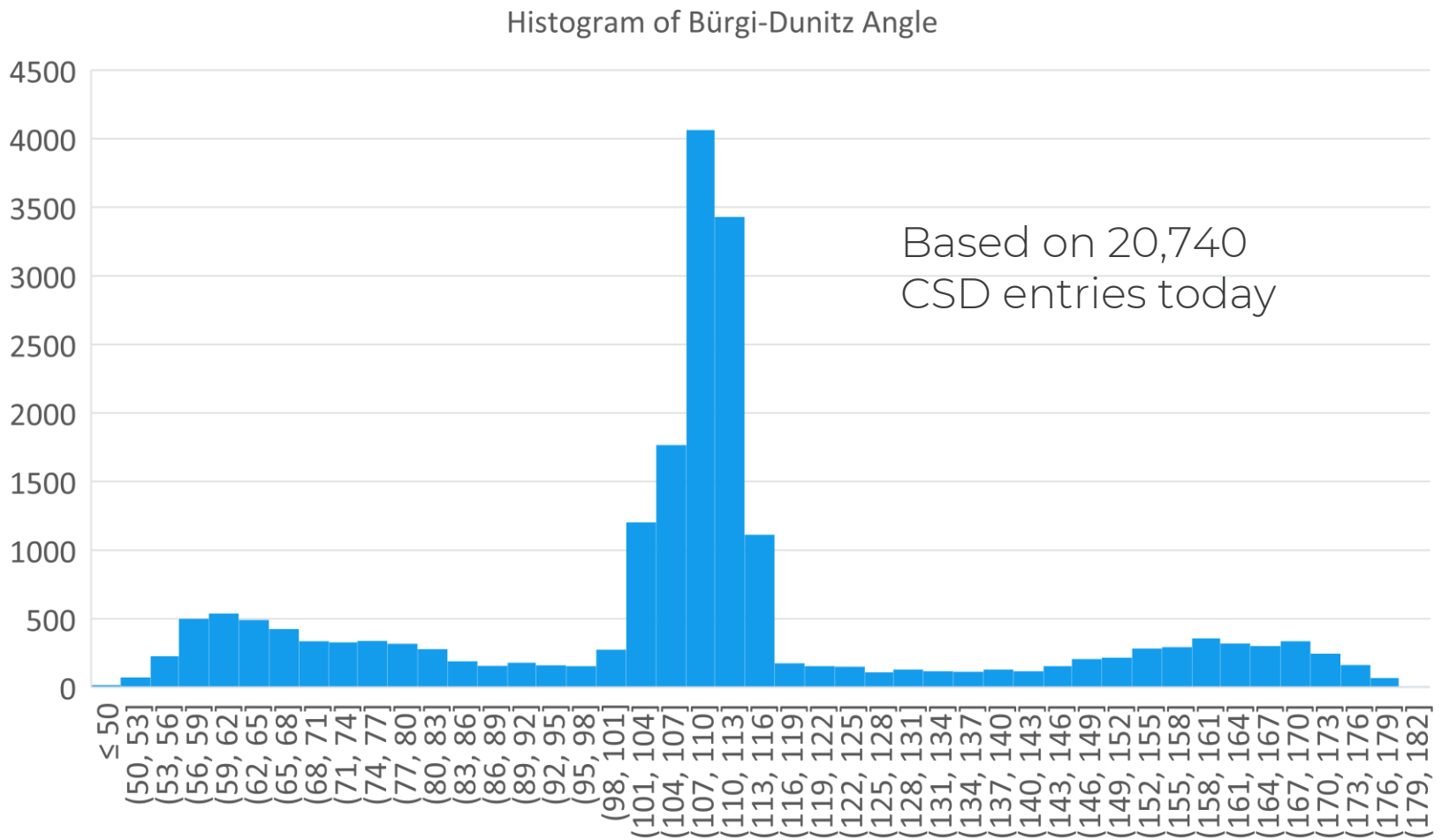


Figure 1. (a) Definition of symbols. (b) Reaction coo
projected on the NCO plane showing nitrogen (top, arro
dicate the estimated direction of the lone pair), carbonyl
(bottom right), and bisector of γ (RCR') (bottom left).

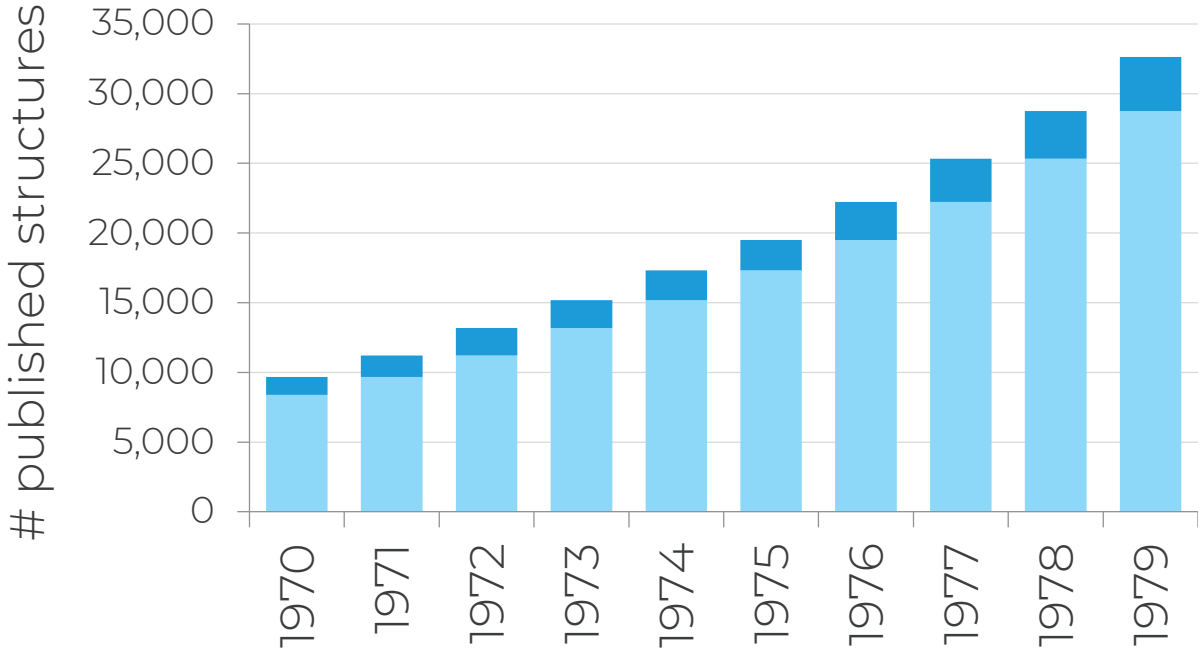


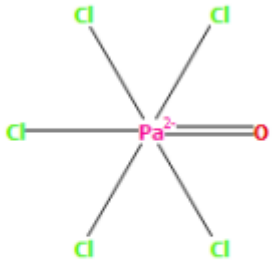
H. B. Burgi, J. D. Dunitz, Eli. Shefter, *J. Am. Chem. Soc.*1973, 95,15, 5065-5067

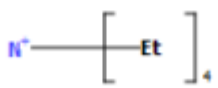
CCDC

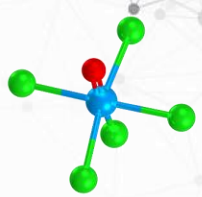
1970-1979

- CSD tripled in size
- Majority now with coordinates

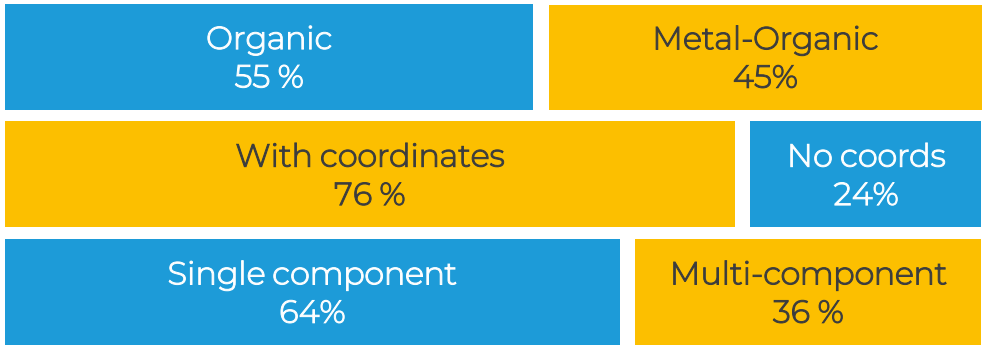








EAOCPA
The first structure
with Protactinium



D.Brown, C.T.Reynolds, P.T.Moseley, *Journal of the Chemical Society, Dalton Transactions*, 1972, 857,
DOI: 10.1039/dt9720000857

CCDC

Creating a database

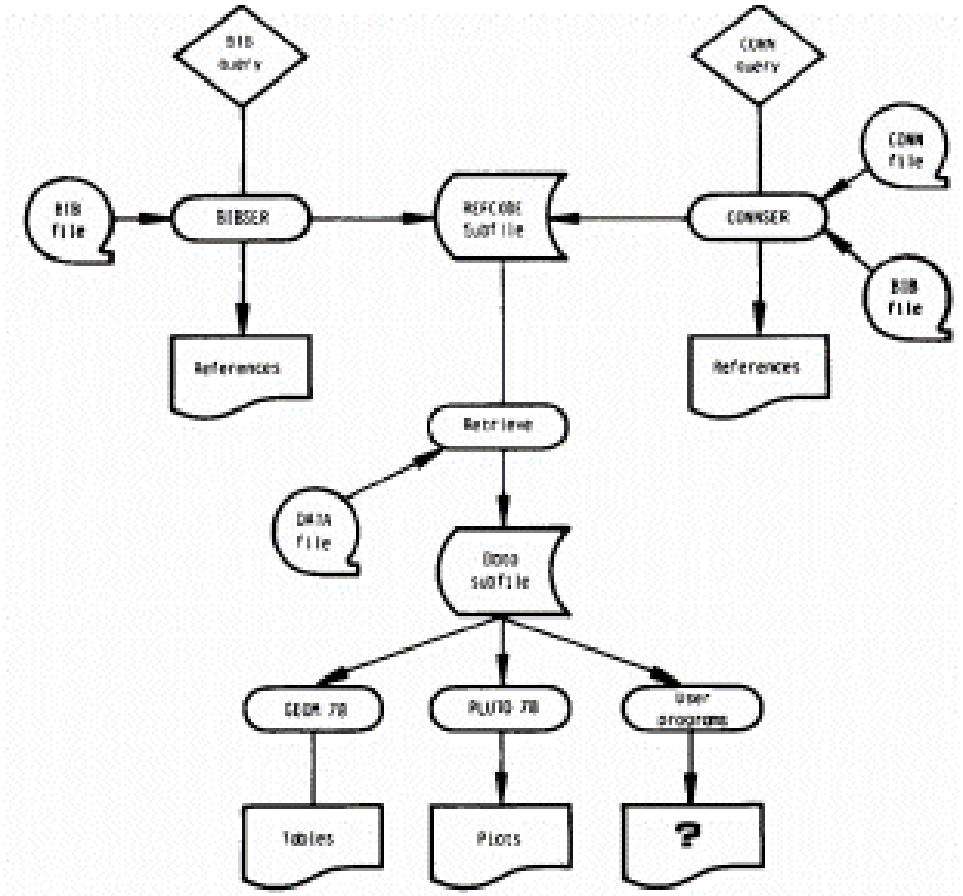


Fig. 4. Flowchart illustrating the search, retrieval analysis and display system.

- The aim by 1980 was to have created a fully retrospective database
- By the end of the decade the database was created and the nucleus of searching software had been developed

Allen, F. H., Watson, D. G. The Cambridge Crystallographic Data Centre: Computer-Based Search, Retrieval, Analysis and Display of Information. *Acta Crystallogr*. 1979, 835, 2331-2339.

CCDC

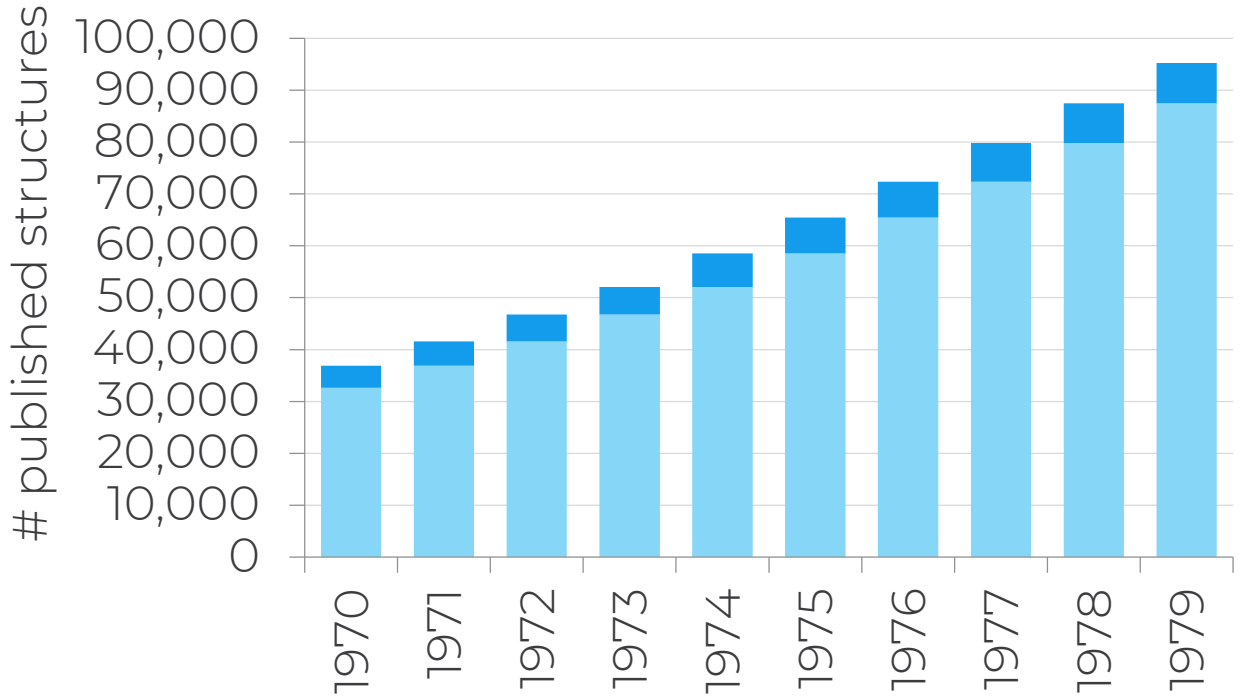
1980s

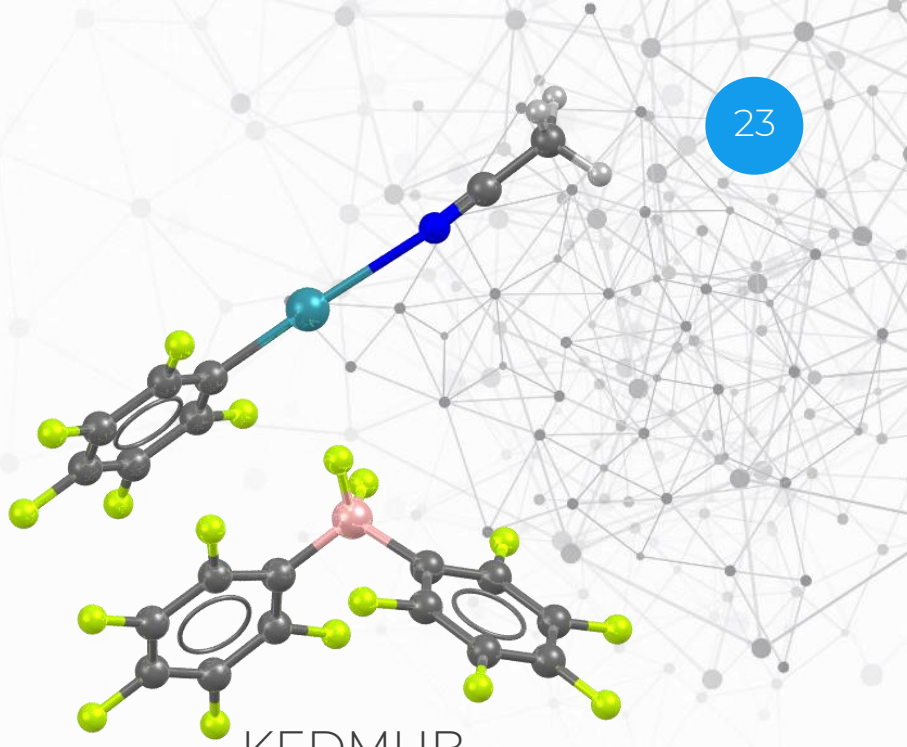
https://www.techrepublic.com/pictures/tech-nostalgia-the-top-10-innovations-of-the-1980s/

CCDC

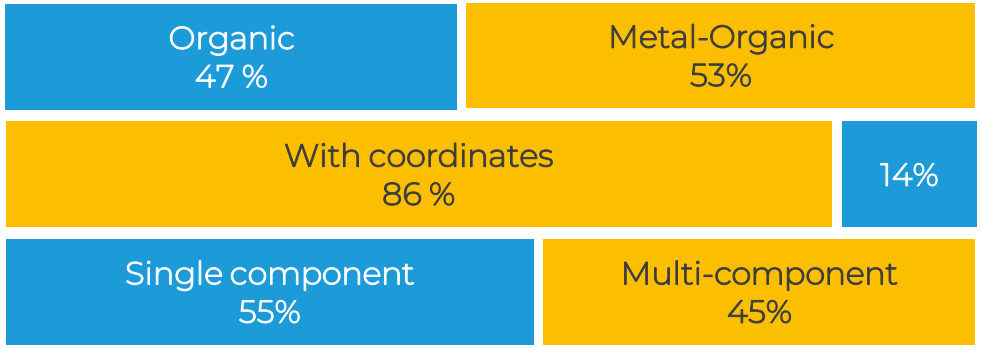
1980-1989

- CSD tripled in size again
- Majority now metal organic





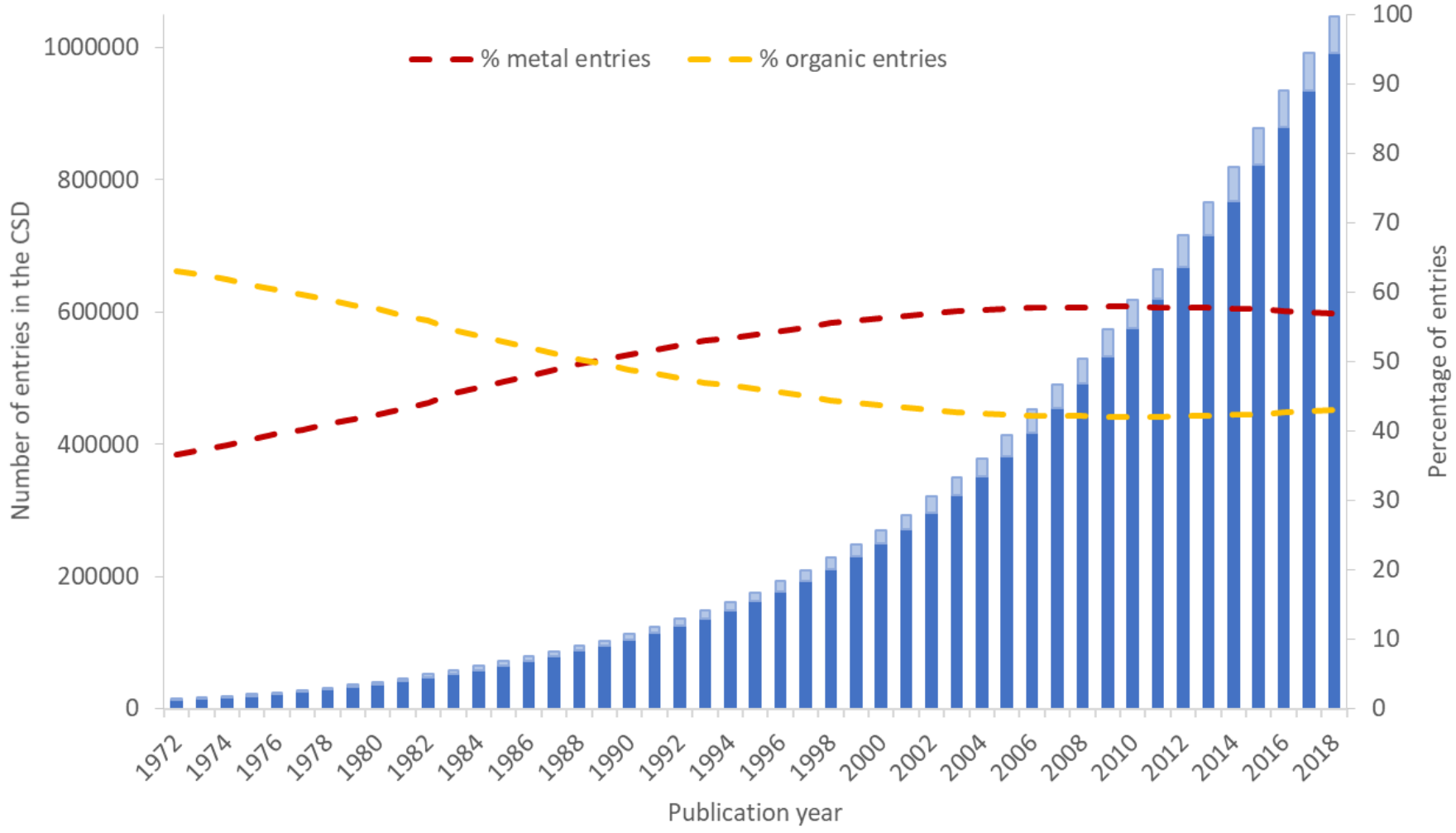
KEDMUB
First structure with bonding to group 18 element



H.J.Frohn, S.Jakobs, G.Henkel, *Angewandte Chemie, International Edition*, 1989, 28, 1506, DOI: 10.1002/anie.198915061

CCDC

1980s - the dominance of metal-organics



CCDC

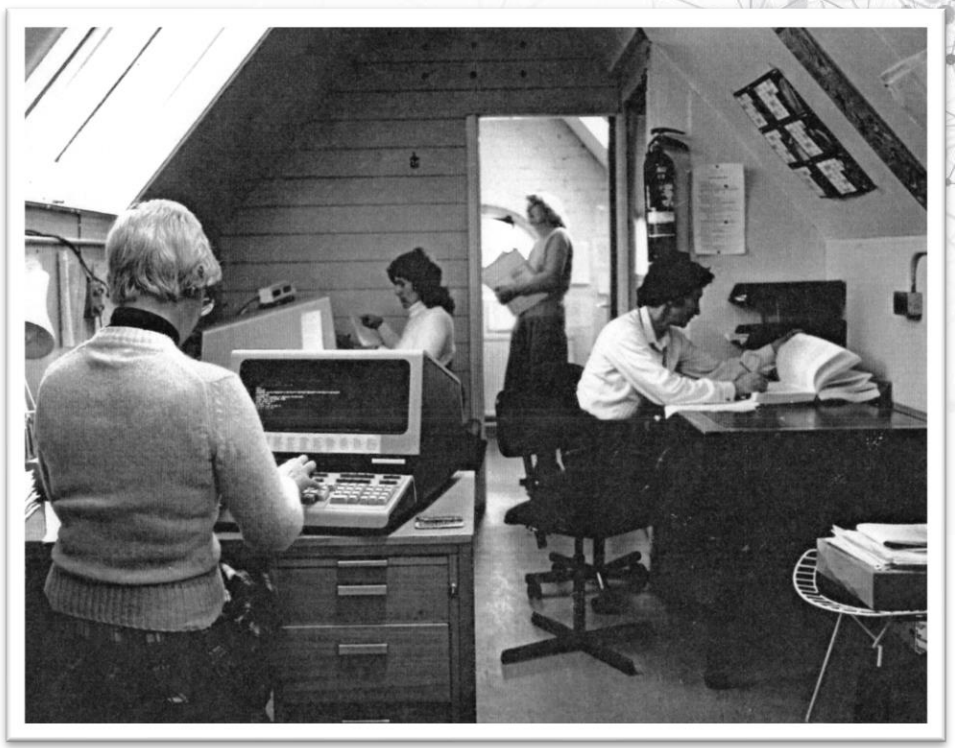
The CCDC in the 1980s

Table 1. Crystallographic data and details of data collection and processing for $ML(NO_3)_2$, with M = Cu [in (1)], Ni [in (2)], and Cd [in (3)]

	(1)	(2)	(3)
Stoicheiometry	$C_{18}H_{27}$		
M	529.0		
Lattice type	Mon		
Space group	$I2$		
a/Å	23.3		
b/Å	10.5		
c/Å	20.8		
α/°	90		
β/°	108.7		
γ/°	90		
U/Å^3	4 873.3		
Z	8		
D_c/g cm^{-3}	1.4		
$F(000)$	2 202.4		
μ(Mo-K_α)/cm^{-1}	9.4		
Approximate crystal dimensions (mm)	0.52 × 0.		
Number of setting angles			
θ range/° (cell dimensions)	−10.24 <		
(data collection)	2 < θ		
h range	−24 <		
k range	0 <		
l range	−22 <		
Number of reflections:			
measured	6 3		
independent	3 1		
observed	2 3		
Final R	0.0		
Final R'	0.0		

Table 2. Atomic co-ordinates with estimated standard deviations in parentheses

Atom	X/a	Y/b	Z/c	Atom	X/a	Y/b	Z/c
(*a*) Compound (1) (× 10^5 for Cu, × 10^4 for others)							
Cu	7 734(2)	11 519(5)	62 931(2)	C(27)	2 251(3)	671(8)	4 844(4)
N(01)	1 493(1)	−122(3)	6 454(2)	C(30)	2 044(2)	483(5)	6 905(2)
C(10)	1 359(2)	−1 289(5)	6 758(3)	N(31)	1 888(1)	1 149(3)	7 431(2)
N(11)	708(1)	−1 425(3)	6 566(2)	N(32)	1 386(1)	1 906(3)	7 246(2)
N(12)	379(1)	−356(3)	6 561(2)	C(33)	1 395(2)	2 512(4)	7 809(2)
C(13)	−162(2)	−764(4)	6 558(2)	C(34)	1 907(2)	2 157(5)	8 332(2)
C(14)	−171(2)	−2 078(4)	6 570(2)	C(35)	2 220(2)	1 302(4)	8 094(2)
C(15)	386(2)	−2 475(4)	6 586(2)	C(36)	918(3)	3 421(6)	7 819(3)
C(16)	−651(2)	142(7)	6 546(4)	C(37)	2 788(2)	622(6)	8 412(3)
C(17)	650(3)	−3 767(5)	6 606(4)	N(40)	−304(2)	2 265(5)	5 590(3)
C(20)	1 549(2)	−339(5)	5 777(2)	O(41)	136(1)	2 399(3)	6 163(2)
N(21)	1 509(2)	902(3)	5 458(2)	O(42)	−717(2)	3 031(5)	5 486(3)
N(22)	1 099(1)	1 738(3)	5 555(2)	O(43)	−272(2)	1 385(5)	5 217(2)
C(23)	1 123(2)	2 745(4)	5 177(2)	N(50)	1 960(2)	3 011(5)	3 346(2)
C(24)	1 539(3)	2 539(6)	4 854(2)	O(51)	1 968(2)	4 072(5)	3 604(3)
C(25)	1 786(2)	1 358(5)	5 035(2)	O(52)	1 526(2)	2 345(6)	3 201(2)
C(26)	743(3)	3 859(6)	5 139(3)	O(53)	2 417(2)	2 613(5)	3 337(4)
(*b*) Compound (2) (× 10^4)							
Ni	5 110(1)	3 357(1)	7 658(1)	N(31)	2 270(6)	1 531(4)	7 437(3)
N(01)	4 516(6)	1 860(4)	6 658(3)	N(32)	2 763(6)	2 672(4)	7 868(3)
C(10)	5 885(9)	995(6)	7 036(5)	C(33)	1 964(8)	2 706(6)	8 528(4)
N(11)	6 173(6)	936(4)	8 049(4)	C(34)	897(8)	1 577(7)	8 516(5)
N(12)	6 420(6)	2 047(5)	8 549(3)	C(35)	1 109(8)	827(6)	7 812(4)
C(13)	6 802(8)	1 740(7)	9 452(5)	C(36)	2 223(9)	3 805(7)	9 165(5)
C(14)	6 818(11)	476(8)	9 522(6)	C(37)	326(10)	−423(6)	7 459(6)
C(15)	6 405(9)	−15(6)	8 625(5)	N(40)	7 521(8)	5 107(5)	8 260(4)
C(16)	7 092(10)	2 741(8)	10 200(5)	O(41)	6 084(6)	4 914(4)	8 518(3)
C(17)	6 208(11)	−1 311(6)	8 281(6)	O(42)	8 610(7)	5 982(5)	8 553(4)
C(20)	4 684(10)	2 395(6)	5 790(5)	O(43)	7 666(5)	4 302(4)	7 690(3)
N(21)	3 777(7)	3 513(5)	5 656(4)	N(50A)	1 902(9)	8 557(6)	5 048(5)
N(22)	3 928(6)	4 195(5)	6 453(4)	O(51A)	1 307(7)	7 942(6)	4 339(5)
C(23)	2 971(9)	5 137(6)	6 121(6)	O(52A)	3 333(10)	8 780(8)	5 620(7)
C(24)	2 311(10)	5 059(8)	5 134(6)	O(53A)	573(14)	8 512(6)	5 431(6)
C(25)	2 805(9)	3 985(7)	4 857(5)	N(50B)	2 032(9)	8 882(6)	4 874(5)
C(26)	2 759(11)	6 060(7)	6 797(6)	O(51B)	1 353(7)	8 323(6)	4 169(5)
C(27)	2 488(11)	3 386(8)	3 912(5)	O(52B)	3 281(10)	9 440(8)	5 053(7)
C(30)	2 652(8)	1 313(6)	6 551(4)	O(53B)	1 492(14)	8 544(6)	5 645(6)

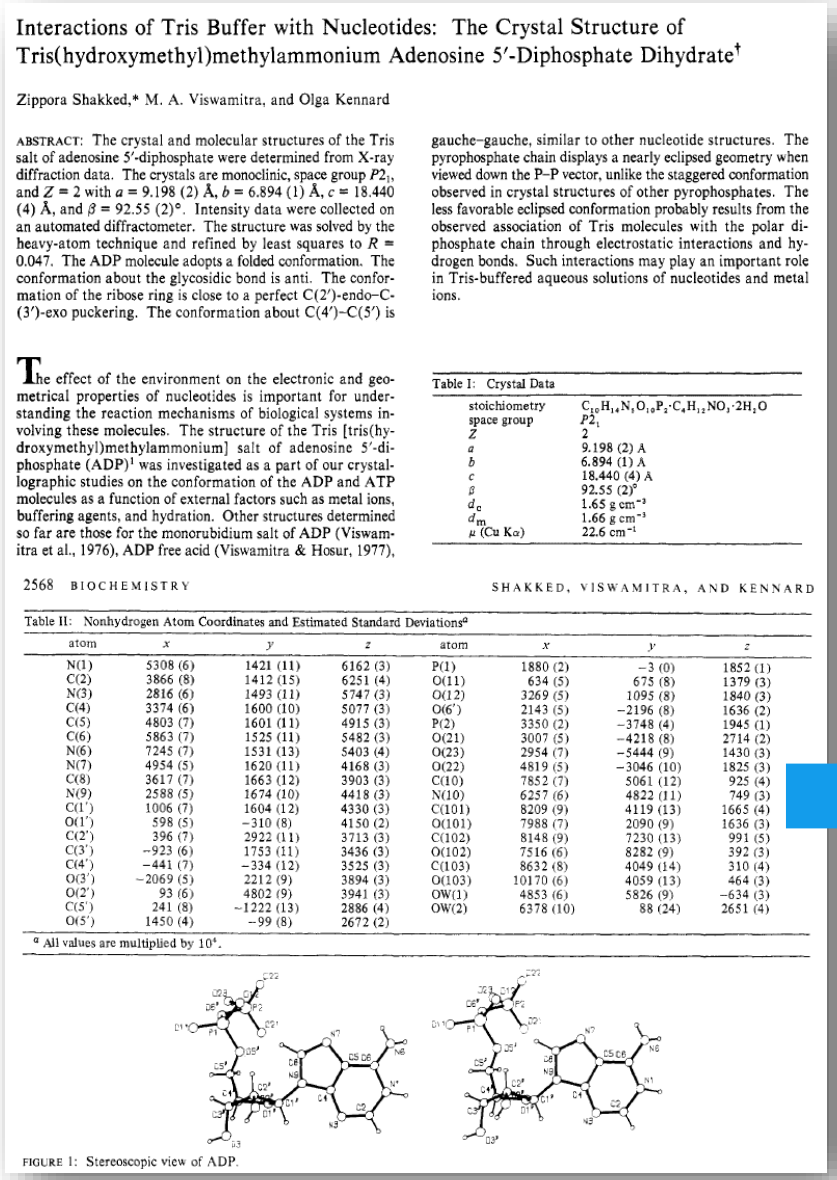


Hand-typed tables of coordinates in journal articles manually transcribed into database records

Kleywegt et al. (1985) *J. Chem. Soc., Dalton Trans*, 2177-2184 doi:10.1039/DT9850002177

CCDC

The 1980s



Interactions of Tris Buffer with Nucleotides: The Crystal Structure of Tris(hydroxymethyl)methylammonium Adenosine 5′-Diphosphate Dihydrate†

Zippora Shakked,* M. A. Viswamitra, and Olga Kennard

ABSTRACT: The crystal and molecular structures of the Tris salt of adenosine 5′-diphosphate were determined from X-ray diffraction data. The crystals are monoclinic, space group $P2_1$, and $Z = 2$ with $a = 9.198$ (2) Å, $b = 6.894$ (1) Å, $c = 18.440$ (4) Å, and $\beta = 92.55$ (2)°. Intensity data were collected on an automated diffractometer. The structure was solved by the heavy-atom technique and refined by least squares to $R = 0.047$. The ADP molecule adopts a folded conformation. The conformation about the glycosidic bond is anti. The conformation of the ribose ring is close to a perfect C(2′)-endo–C(3′)-exo puckering. The conformation about C(4′)–C(5′) is gauche–gauche, similar to other nucleotide structures. The pyrophosphate chain displays a nearly eclipsed geometry when viewed down the P–P vector, unlike the staggered conformation observed in crystal structures of other pyrophosphates. The less favorable eclipsed conformation probably results from the observed association of Tris molecules with the polar diphosphate chain through electrostatic interactions and hydrogen bonds. Such interactions may play an important role in Tris-buffered aqueous solutions of nucleotides and metal ions.

The effect of the environment on the electronic and geometrical properties of nucleotides is important for understanding the reaction mechanisms of biological systems involving these molecules. The structure of the Tris [tris(hydroxymethyl)methylammonium] salt of adenosine 5′-diphosphate (ADP)[1] was investigated as a part of our crystallographic studies on the conformation of the ADP and ATP molecules as a function of external factors such as metal ions, buffering agents, and hydration. Other structures determined so far are those for the monorubidium salt of ADP (Viswamitra et al., 1976), ADP free acid (Viswamitra & Hosur, 1977),

Table I: Crystal Data

stoichiometry	$C_{10}H_{14}N_7O_{10}P_2 \cdot C_4H_{11}NO_3 \cdot 2H_2O$
space group	$P2_1$
Z	2
a	9.198 (2) Å
b	6.894 (1) Å
c	18.440 (4) Å
β	92.55 (2)°
d_c	1.65 g cm^{-3}
d_m	1.66 g cm^{-3}
μ (Cu Kα)	22.6 cm^{-1}

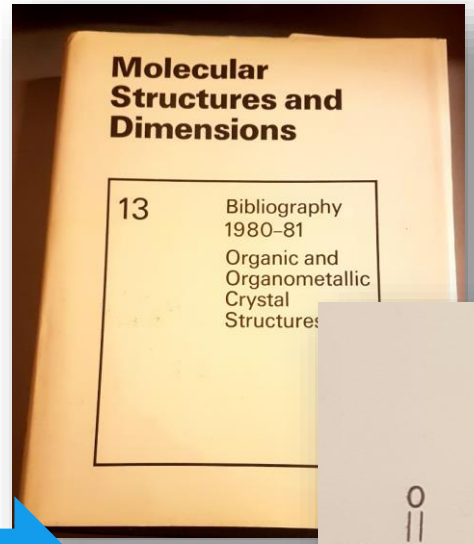
2568 BIOCHEMISTRY SHAKKED, VISWAMITRA, AND KENNARD

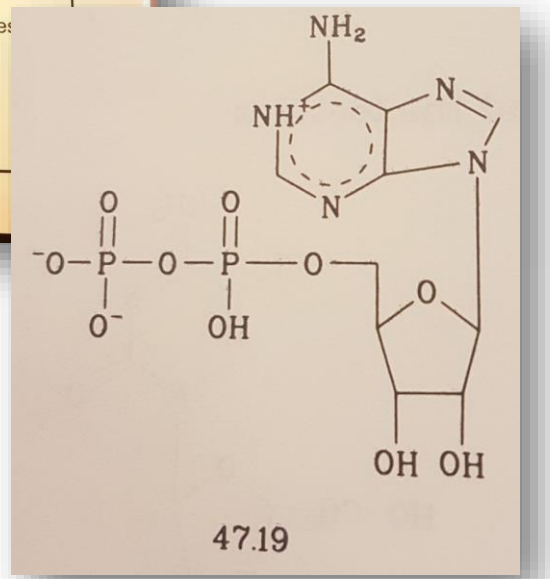
Table II: Nonhydrogen Atom Coordinates and Estimated Standard Deviations[a]

atom	x	y	z	atom	x	y	z
N(1)	5308 (6)	1421 (11)	6162 (3)	P(1)	1880 (2)	−3 (0)	1852 (1)
C(2)	3866 (8)	1412 (15)	6251 (4)	O(11)	634 (5)	675 (8)	1379 (3)
N(3)	2816 (6)	1493 (11)	5747 (3)	O(12)	3269 (5)	1095 (8)	1840 (3)
C(4)	3374 (6)	1600 (10)	5077 (3)	O(6′)	2143 (5)	−2196 (8)	1636 (2)
C(5)	4803 (7)	1601 (11)	4915 (3)	P(2)	3350 (2)	−3748 (4)	1945 (1)
C(6)	5863 (7)	1525 (11)	5482 (3)	O(21)	3007 (5)	−4218 (8)	2714 (2)
N(6)	7245 (7)	1531 (13)	5403 (4)	O(23)	2954 (7)	−5444 (9)	1430 (3)
N(7)	4954 (5)	1620 (11)	4168 (3)	O(22)	4819 (5)	−3046 (10)	1825 (3)
C(8)	3617 (7)	1663 (12)	3903 (3)	C(10)	7852 (7)	5061 (12)	925 (4)
N(9)	2588 (5)	1674 (10)	4418 (3)	N(10)	6257 (6)	4822 (11)	749 (3)
C(1′)	1006 (7)	1604 (12)	4330 (3)	C(101)	8209 (9)	4119 (13)	1665 (4)
O(1′)	598 (5)	−310 (8)	4150 (2)	O(101)	7988 (7)	2090 (9)	1636 (3)
C(2′)	396 (7)	2922 (11)	3713 (3)	C(102)	8148 (9)	7230 (13)	991 (5)
C(3′)	−923 (6)	1753 (11)	3436 (3)	O(102)	7516 (6)	8282 (9)	392 (3)
C(4′)	−441 (7)	−334 (12)	3525 (3)	C(103)	8632 (8)	4049 (14)	310 (4)
O(3′)	−2069 (5)	2212 (9)	3894 (3)	O(103)	10170 (6)	4059 (13)	464 (3)
O(2′)	93 (6)	4802 (9)	3941 (3)	OW(1)	4853 (6)	5826 (9)	−634 (3)
C(5′)	241 (8)	−1222 (13)	2886 (4)	OW(2)	6378 (10)	88 (24)	2651 (4)
O(5′)	1450 (4)	−99 (8)	2672 (2)				

[a] All values are multiplied by 10^4.

FIGURE 1: Stereoscopic view of ADP.





- Introduction of computer aided chemical diagrams
  - Greatly extending the utility of the series

Z.Shakked, M.A.Viswamitra, O.Kennard, *Biochemistry*, 1980, 19, 12, 2567-2571 DOI: 10.1021/bi00553a005

CCDC

Tables of bond lengths

J. CHEM. SOC. PERKIN TRANS. II 1987 S1

Tables of Bond Lengths determined by *X*-Ray and Neutron Diffraction. Part 1. Bond Lengths in Organic Compounds

Frank H. Allen,* Olga Kennard, and David G. Watson
Cambridge Crystallographic Data Centre, University Chemical Laboratory, Lensfield Road, Cambridge CB2 1EW
Lee Brammer and A. Guy Orpen
School of Chemistry, University of Bristol, Cantock's Close Bristol BS8 1TS
Robin Taylor
I.C.I. Plant Protection Division, Jealott's Hill Research Station, Bracknell, Berkshire RG12 6EY

The average lengths of bonds involving the elements H, B, C, N, O, F, Si, P, S, Cl, As, Se, Br, Te, and I in organic compounds are reported.

The determination of molecular geometry is of vital importance to our understanding of chemical structure and bonding. The majority of experimental data have come from *X*-ray and neutron diffraction, microwave spectroscopy and electron diffraction. Over the years compilations of results from these techniques have appeared sporadically. The first major compilation was Chemical Society Special Publication No. 11: 'Tables of Interatomic Distances and Configuration in Molecules and Ions'.[1] This volume summarized results obtained by diffraction and spectroscopic methods prior to 1956; a supplementary volume[2] extended this coverage to 1959. Summary tables of bond lengths between carbon and other elements were also published in Volume III of 'International Tables for *X*-Ray Crystallography'.[3] Some years later the Cambridge Crystallographic Data Centre[4] produced an atlas-style compendium[5] of all organic, organometallic, and metal complex crystal structures published in the period 1960—1965. More recently a survey of geometries determined by spectroscopic methods[6] has extended coverage in this area to mid-1977.

The production of further comprehensive compendia of *X*-ray and neutron diffraction results has been precluded by the steep rise in the number of published crystal structures, as illustrated by Figure 1. Printed compilations have been effectively superseded by computerized databases. In particular the Cambridge Structural Database (CSD) now contains bibliographic, chemical and numerical results for *ca.* 55 000 organocarbon crystal structures. This machine-readable file fulfils the function of a comprehensive structure-by-structure compendium of molecular geometries. However the amount of data now held in CSD is so large that there is also a need for concise, printed tabulations of average molecular dimensions.

The only tables of average geometry in general use are those contained in the Chemical Society Special Publications[1,2] of 1958 and 1965, which list mean bond lengths for a variety of atom pairs and functional groups. Since these early tables were based on data obtained before 1960, we have used CSD to prepare a new table of average bond lengths in organic compounds. The Table given here specifically lists average lengths for bonds involving the elements H, B, C, N, O, F, Si, P, S, Cl, As, Se, Br, Te, and I. Mean values are presented for 682 different bond types involving these elements. Average bond lengths in organometallic compounds and metal complexes will be presented in a later paper.

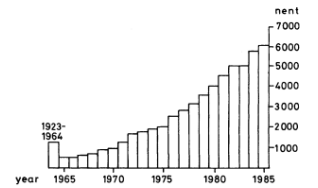


Figure 1. Growth of the Cambridge Structural Database 1965—1985 as no. of entries (nent) published in a given year

Methodology

Selection of Crystallographic Data.—All results given in the Table are based on *X*-ray and neutron diffraction results retrieved from the September 1985 version of CSD. Neutron diffraction data only were used to derive mean bond lengths involving hydrogen atoms. This version of CSD contained results for 49 854 single-crystals diffraction studies of organocarbon compounds 10 324 of these satisfied the acceptance criteria listed below and were used in the averaging procedures.

(i) Structure is 'organic', *i.e.* belongs to CSD classes 1—65 or 70.[7]

(ii) Atomic co-ordinates for the structure have been published and are available in CSD.

(iii) Structure was determined from diffractometer data.

(iv) Structure does not contain unresolved numeric data errors from the original publication (such errors are usually typographical and are normally resolved by consultation with the authors).

(v) Structure was not reported to be disordered.

(vi) Only structures of high precision were included on the basis of *either* (a) crystallographic *R* factor was ≤0.07 and the reported mean estimated standard deviation (e.s.d.) for the C–C bond lengths was ≤0.010 Å (corresponds to AS flag = 1 or 2 in CSD), *or* (b) crystallographic *R* factor ≤0.05 and the mean e.s.d. for C–C bonds are not available in the database (AS = 0 in CSD).

(vii) Where the structure of a given compound had been determined more than once within the limits of (i)—(vi) then only the most precise determination was used.

Program System.—All calculations were performed on the University of Cambridge IBM 3081D computer using the

J. CHEM. SOC. DALTON TRANS. 1989 S1

Supplement

Tables of Bond Lengths determined by *X*-Ray and Neutron Diffraction. Part 2.† Organometallic Compounds and Co-ordination Complexes of the *d*- and *f*-Block Metals

A. Guy Orpen * and Lee Brammer
Department of Inorganic Chemistry, The University, Bristol BS8 1TS
Frank H. Allen, Olga Kennard, and David G. Watson
Cambridge Crystallographic Data Centre, University Chemical Laboratory, Lensfield Road, Cambridge CB2 1EW
Robin Taylor
I.C.I. Plant Protection Division, Jealott's Hill Research Station, Bracknell, Berkshire RG12 6EY

Average lengths for metal–ligand bonds are reported, together with some intraligand distances, for complexes of the *d*- and *f*-block metals. Mean values are presented for 325 different bond types involving metal atoms bonded to H, B, C, N, O, F, Si, P, S, Cl, As, Se, Br, Te, or I atoms of the ligands.

The determination of molecular geometry is of vital importance to our understanding of chemical structure and bonding. The majority of experimental data have come from *X*-ray and neutron diffraction, microwave spectroscopy, and electron diffraction. Over the years compilations of results from these techniques have appeared sporadically. The first major compilation was Chemical Society Special Publication no. 11: 'Tables of Interatomic Distances and Configuration in Molecules and Ions.'[1] This volume summarized results obtained by diffraction and spectroscopic methods prior to 1956; a supplementary volume[2] extended this coverage to 1959. Summary tables of bond lengths between carbon and other elements were also published in volume 3 of 'International Tables for *X*-Ray Crystallography.'[3] Some years later the Cambridge Crystallographic Data Centre[4] produced an atlas-style compendium of all organic, organometallic, and metal complex crystal structures published in the period 1960—1965.[5] More recently a survey of geometries determined by spectroscopic methods[6] has extended coverage in this area to mid-1977. A notable compendium of structural data, without geometric information, was given in 'Comprehensive Organometallic Chemistry,'[7] covering all complexes with metal–carbon bonds. The BIDICS series,[8] which finished in 1981, provided for some years a full coverage of metal complexes giving both bibliographic and geometric information. There have also been valuable annual summaries, without geometric information, on the structures of organometallic compounds determined by diffraction methods.[9]

The production of further comprehensive compendia of *X*-ray and neutron diffraction results has been precluded by the steep rise in the number of published crystal structures, as illustrated by Figure 1. Printed compilations have been effectively susperseded by computerized databases. In particular the Cambridge Structural Database (CSD) now contains bibliographic, chemical, and numerical results for some 70 000 organocarbon crystal structures. This machine-readable file fulfils the function of a comprehensive structure-by-structure compendium of molecular geometries. However the amount of data now held in CSD is so large that there is also a need for concise, printed tabulations of average molecular dimensions.

The only tables of average geometry in general use are those

† Part I is ref. 10.

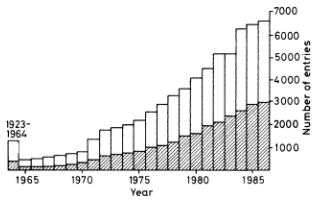


Figure 1. Growth of the Cambridge Structural Database as number of entries increases annually. The structures containing *d*- or *f*-block metals are indicated by shading

contained in the Chemical Society Special Publications[1,2] of 1958 and 1965, which list mean bond lengths for a variety of atom pairs and functional groups. Since these early tables were based on data obtained before 1960, we have used CSD to prepare a new table of average bond lengths in organic compounds[1,3] and in metal complexes. The table given here (Table 3) specifically lists average lengths for metal–ligand distances, together with intraligand distances, involving bonds between the *d*- and *f*-block metals (Sc–Zn, Y–Cd, La–Hg, Ce–Lu, and Th–U) and atoms H, B, C, N, O, F, Si, P, S, Cl, As, Se, Br, Te, and I of ligands. Mean values are presented for 325 different bond types involving such metal–ligand bonds.

Methodology

Selection of Crystallographic Data.—All results given in Table 3 are based on *X*-ray and neutron diffraction results retrieved from the September 1985 version of CSD. Neutron diffraction data only were used to derive mean bond lengths involving hydrogen atoms. This version of CSD contained results for 49 854 single-crystal diffraction studies of organocarbon compounds; 9 802 of these satisfied the acceptance criteria listed below and were used in the averaging procedures.

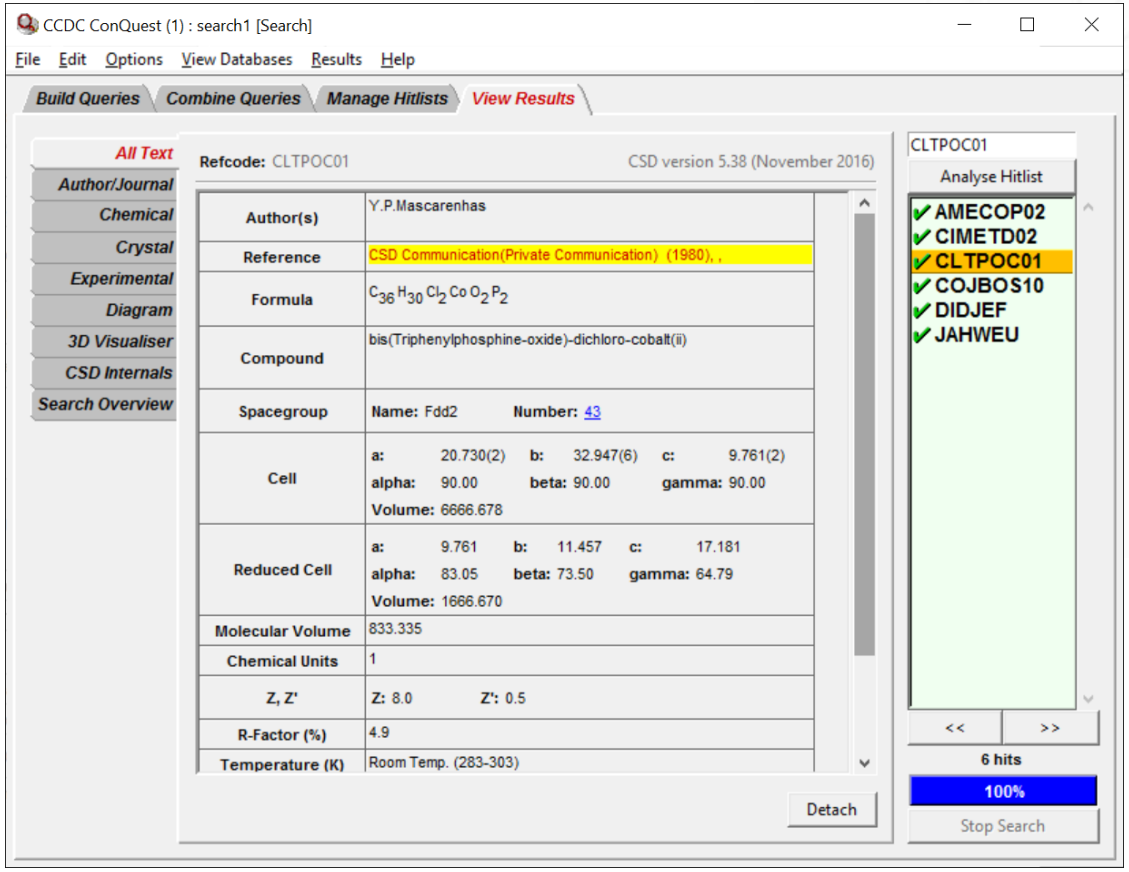
The production of further comprehensive compendia of *X*-ray and neutron diffraction results has been precluded by the steep rise in the number of published crystal structures, as illustrated by Figure 1. Printed compilations have been effectively superseded by computerized databases. In particular the Cambridge Structural Database (CSD) now contains bibliographic, chemical and numerical results for *ca.* 55 000 organocarbon crystal structures. This machine-readable file fulfils the function of a comprehensive structure-by-structure compendium of molecular geometries. However the amount of data now held in CSD is so large that there is also a need for concise, printed tabulations of averag[illegible]lecular dimensions.

..the amount of data now held in CSD is so large that there is also a need for concise, printed tabulations of average molecular dimensions

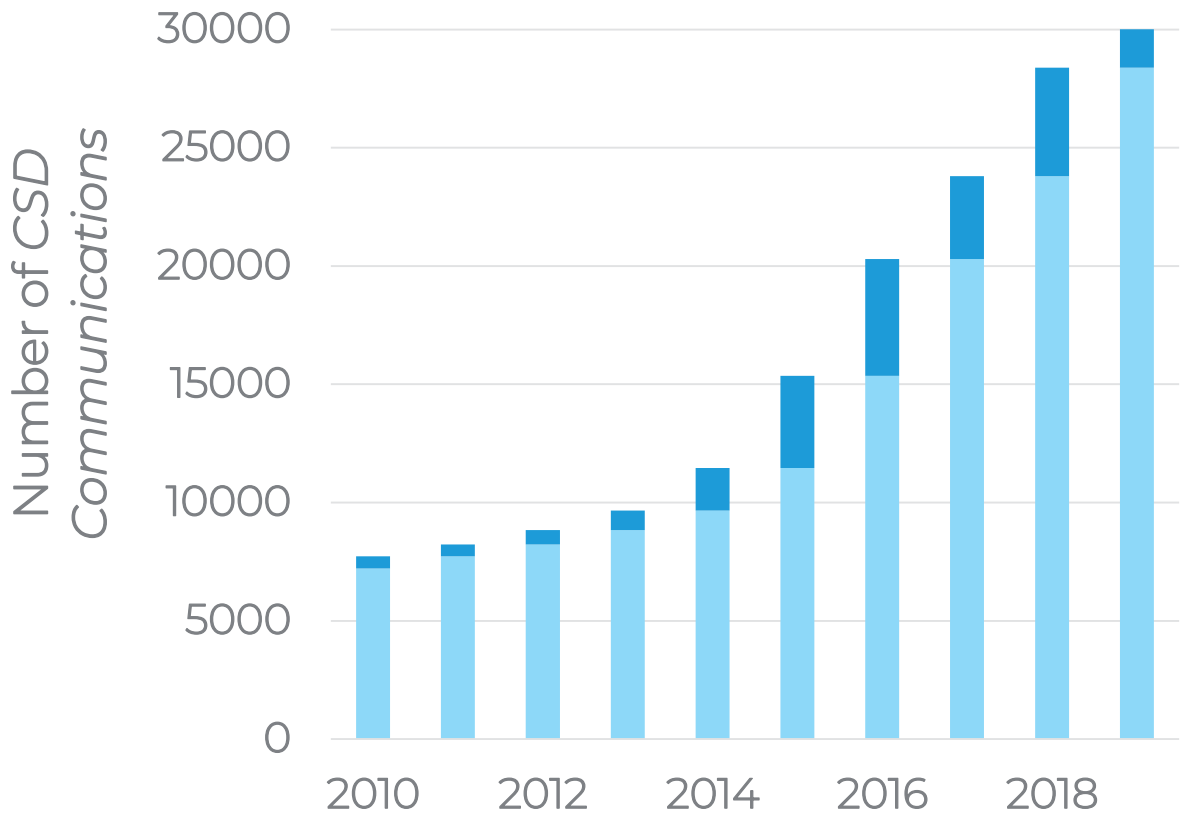
Allen, F. H. et al., J. Chem. Soc. Perkin Trans. 2 S1 (1987). doi:10.1039/p298700000s1
Orpen, A. G. et al. . Chem. Soc. Dalt. Trans. 0, S1 (1989).

CCDC

1980s - The start of database publications

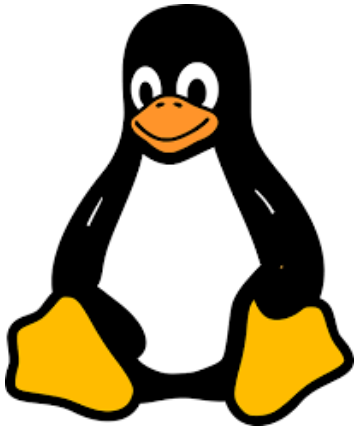


Database publications today...



CCDC

1990s



amazon

DVD TM

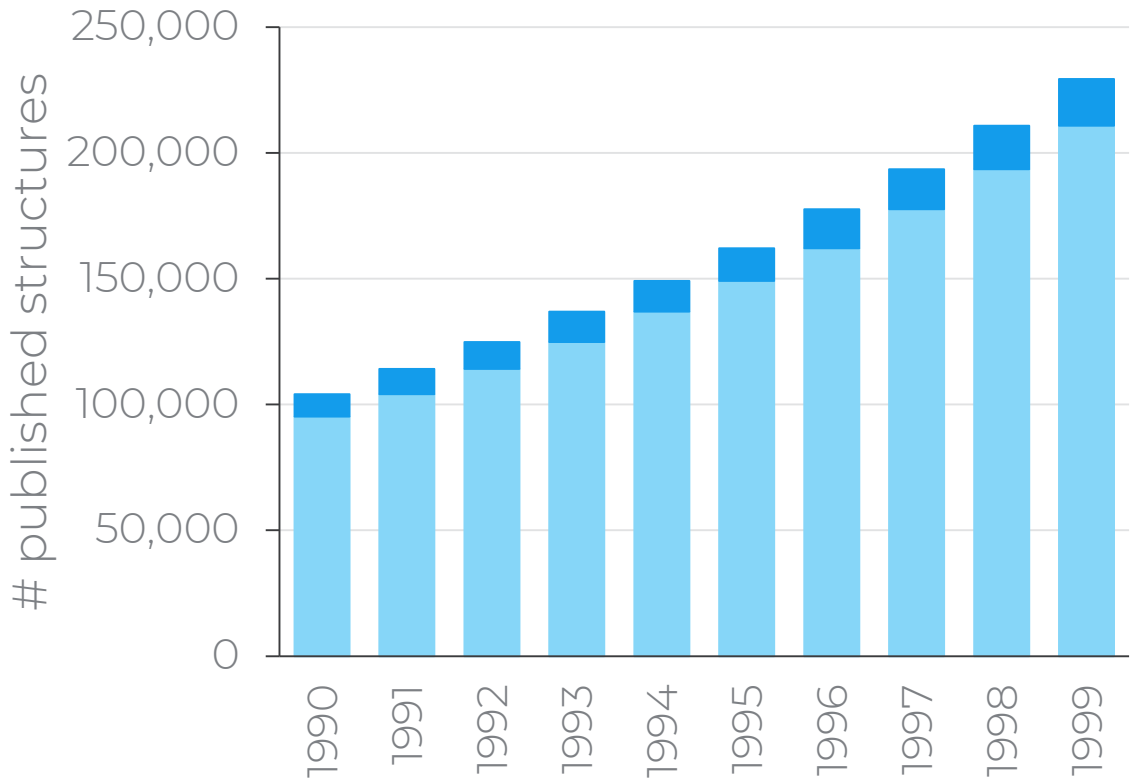
https://www.techrepublic.com/pictures/tech-nostalgia-the-top-15-innovations-of-the-1990s/

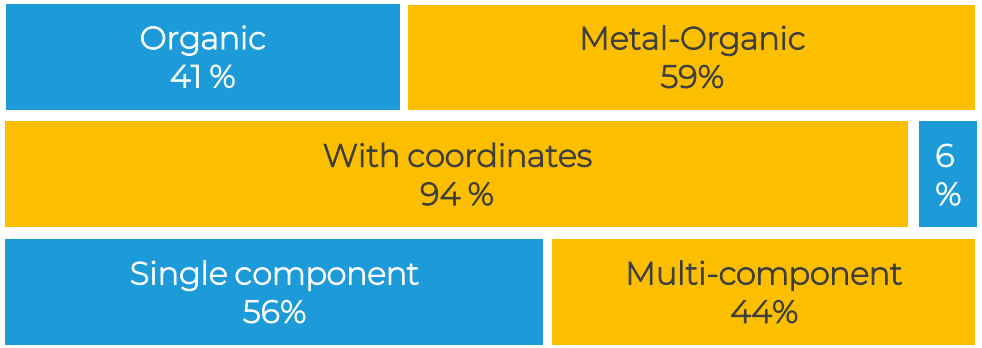
CCDC

1990s

1990-1999

- CSD reaches 100,000 and 200,000!

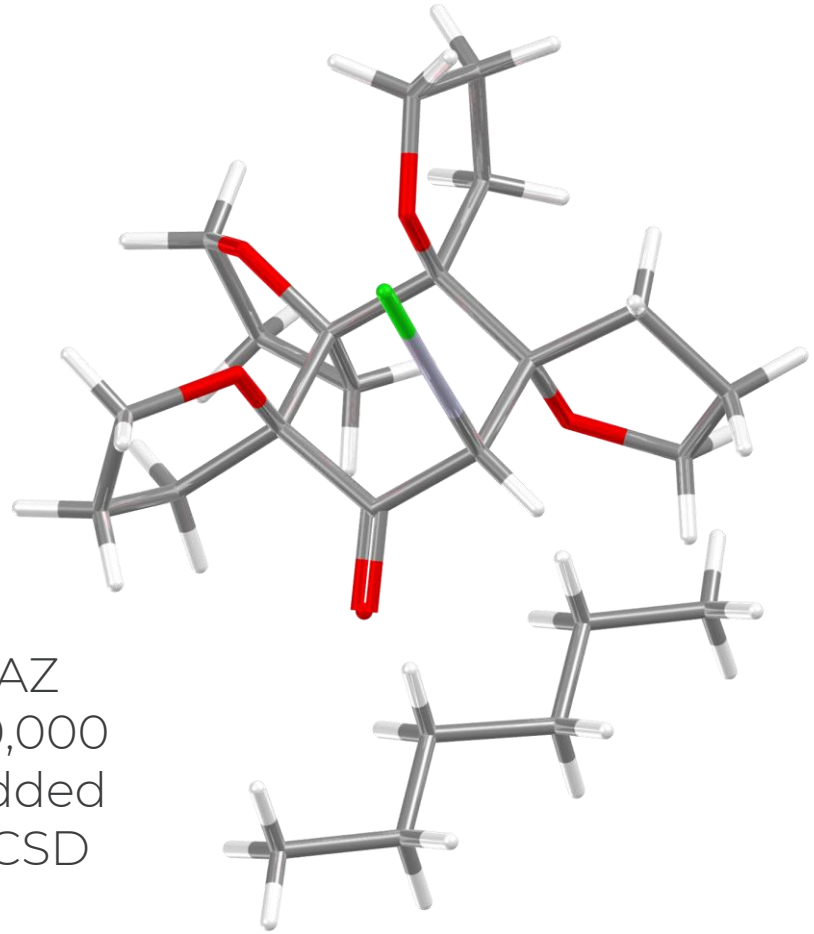




M.F.Meidine, P.B.Hitchcock, H.W.Kroto, R.Taylor, D.R.M.Walton, *Chem.Commun.* 1992, 1534, DOI:10.1039/C39920001534

CCDC

A new milestone

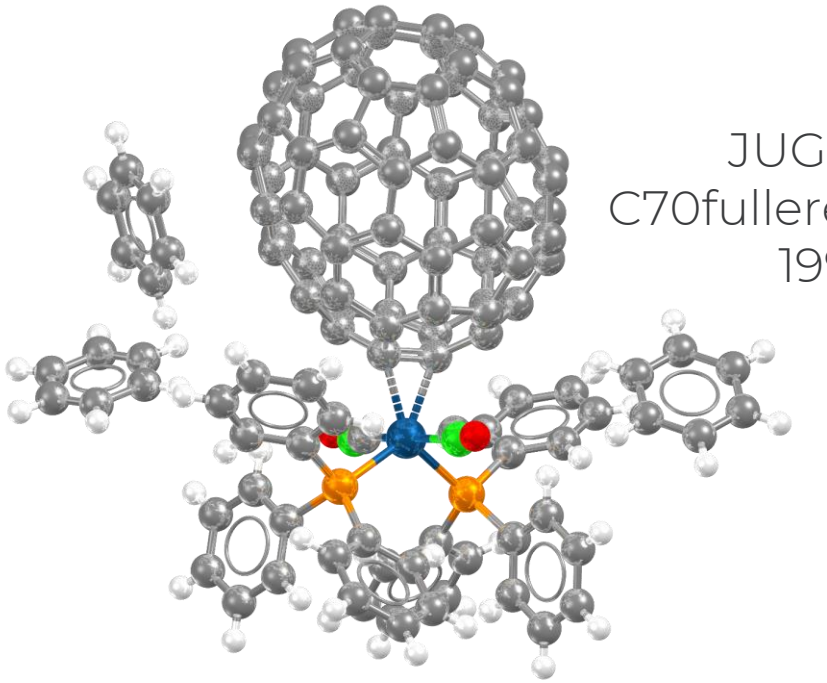


VAVFAZ
The 200,000 entry added to the CSD

L.A.Paquette, D.G.Bolin, M.Stepanian, B.M.Branan, U.V.Mallavadhani, Jinsung Tae, S.W.E.Eisenberg, R.D.Rogers, *Journal of the American Chemical Society*, 1998, 120, 11603, DOI: 10.1021/ja981756p

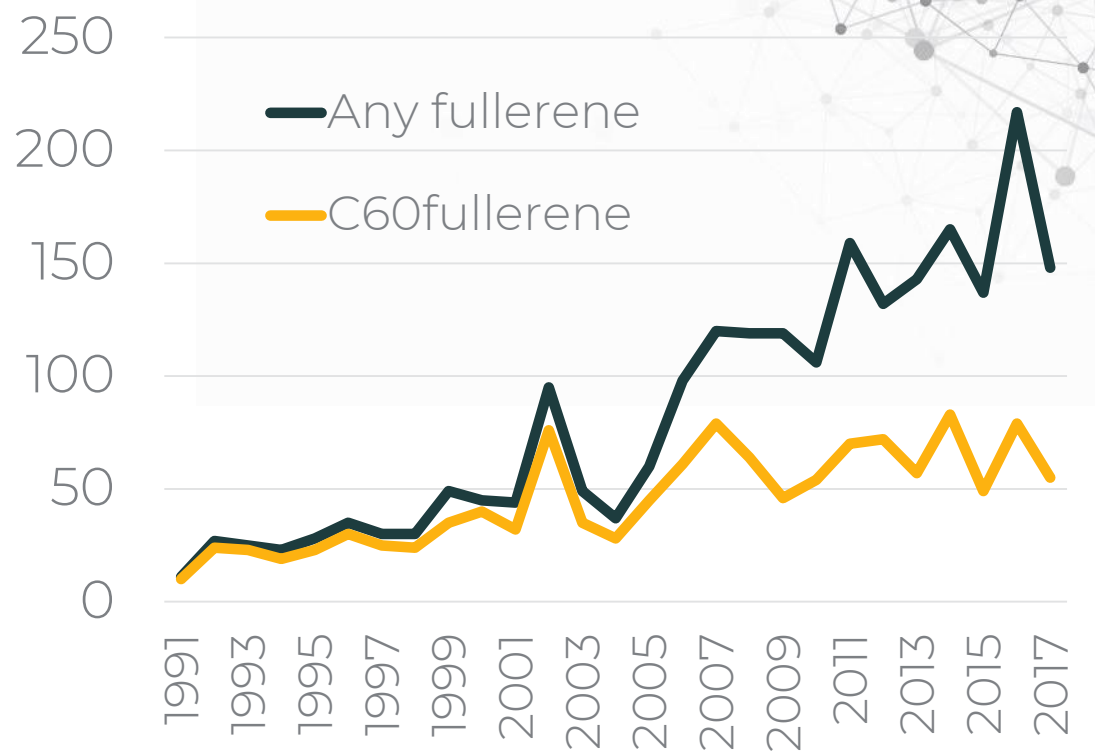
CCDC

1990s - fullerenes



JUGCET
C70fullerene from 1991

Kroto, Curl, and Smalley were awarded the 1996 Nobel Prize in Chemistry for their roles in the discovery of this class of molecules.

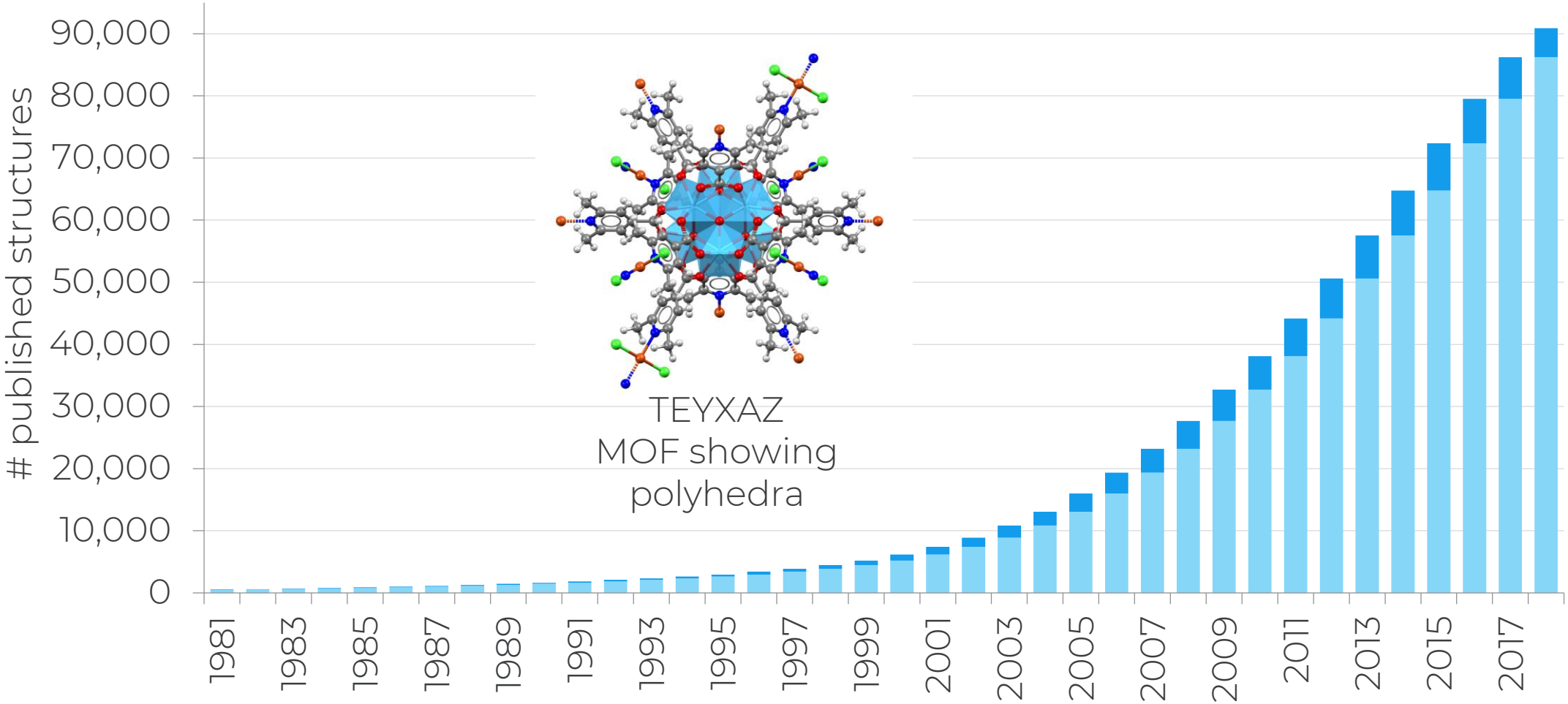


- First discovered in 1985
- First fullerene structures published in 1991
  - 11 structures in 10 articles

A.L.Balch, V.J.Catalano, Joong W.Lee, M.M.Olmstead, S.R.Parkin, *J.Am.Chem.Soc.* (1991), 113, 8953, doi:10.1021/ja00023a057

CCDC

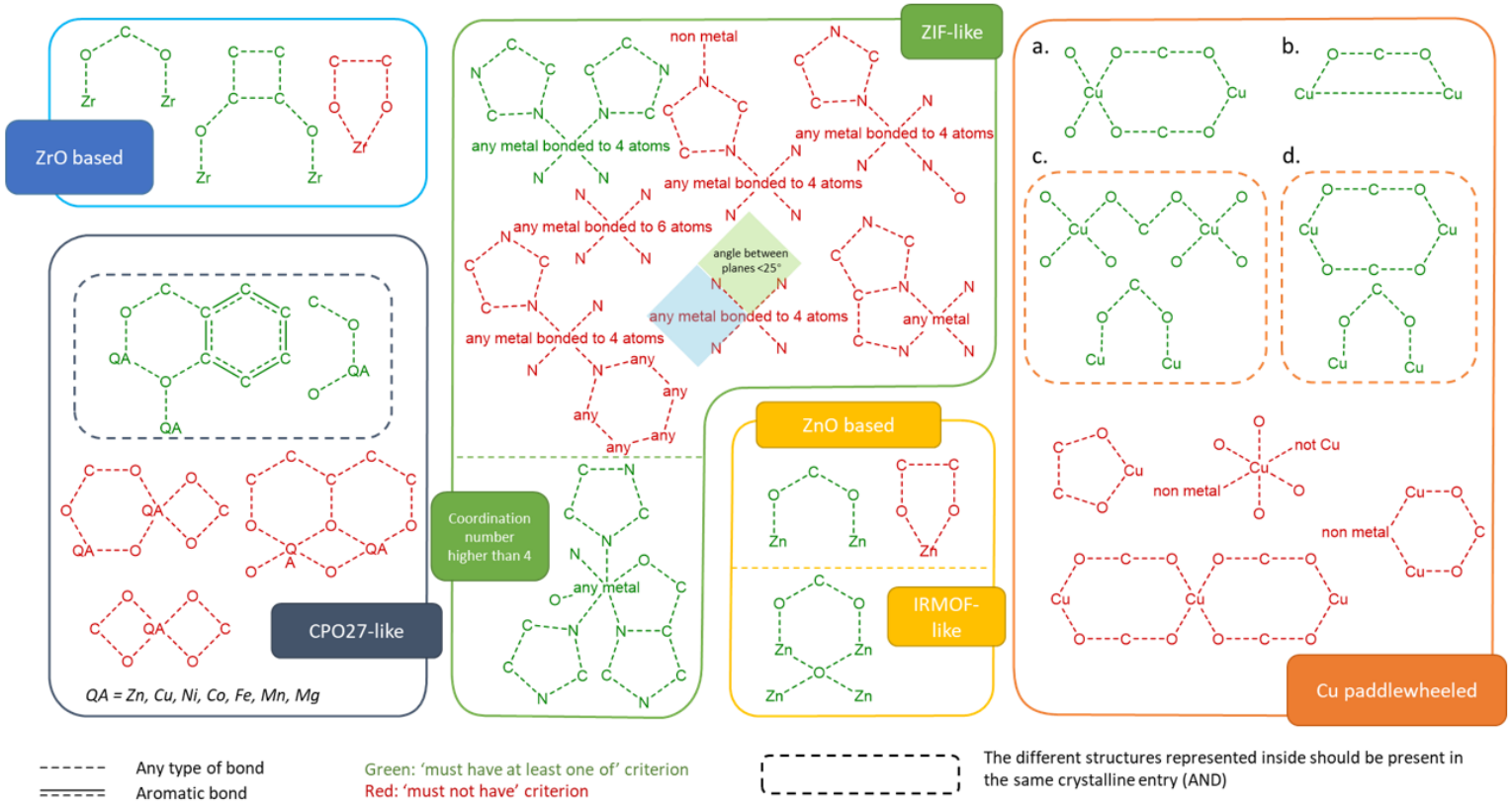
The rise of MOFs

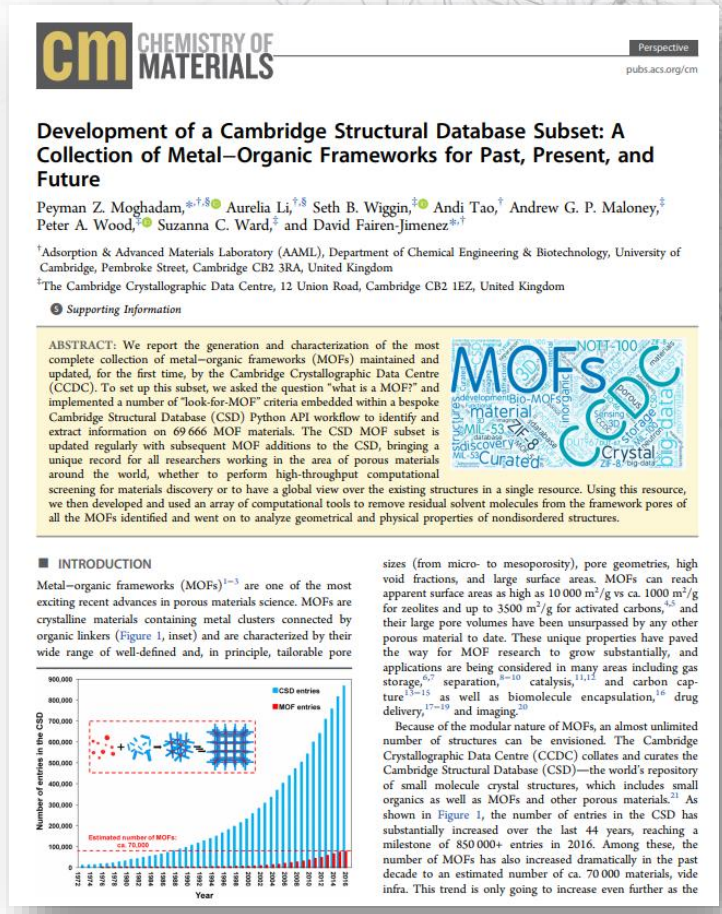


Patrick F. Muldoon, Chong Liu, Carson C. Miller, Samuel Benjamin Koby, Michael O'Keeffe, Tian-Yi Luo, Nathaniel L Rosi, Sunil Saxena, Austin Gamble Jarvi, *Journal of the American Chemical Society*, 2018, 140, 6194, DOI: 10.1021/jacs.8b02192

CCDC

Complicated structures require complicated searches....





CHEMISTRY OF MATERIALS

Perspective

pubs.acs.org/cm

Development of a Cambridge Structural Database Subset: A Collection of Metal–Organic Frameworks for Past, Present, and Future

Peyman Z. Moghadam,*[,†,§] Aurelia Li,[†,§] Seth B. Wiggin,[‡] Andi Tao,[†] Andrew G. P. Maloney,[‡] Peter A. Wood,[‡] Suzanna C. Ward,[‡] and David Fairen-Jimenez*[,†]

[†]Adsorption & Advanced Materials Laboratory (AAML), Department of Chemical Engineering & Biotechnology, University of Cambridge, Pembroke Street, Cambridge CB2 3RA, United Kingdom

[‡]The Cambridge Crystallographic Data Centre, 12 Union Road, Cambridge CB2 1EZ, United Kingdom

Supporting Information

ABSTRACT: We report the generation and characterization of the most complete collection of metal–organic frameworks (MOFs) maintained and updated, for the first time, by the Cambridge Crystallographic Data Centre (CCDC). To set up this subset, we asked the question "what is a MOF?" and implemented a number of "look-for-MOF" criteria embedded within a bespoke Cambridge Structural Database (CSD) Python API workflow to identify and extract information on 69 666 MOF materials. The CSD MOF subset is updated regularly with subsequent MOF additions to the CSD, bringing a unique record for all researchers working in the area of porous materials around the world, whether to perform high-throughput computational screening for materials discovery or to have a global view over the existing structures in a single resource. Using this resource, we then developed and used an array of computational tools to remove residual solvent molecules from the framework pores of all the MOFs identified and went on to analyze geometrical and physical properties of nondisordered structures.

INTRODUCTION

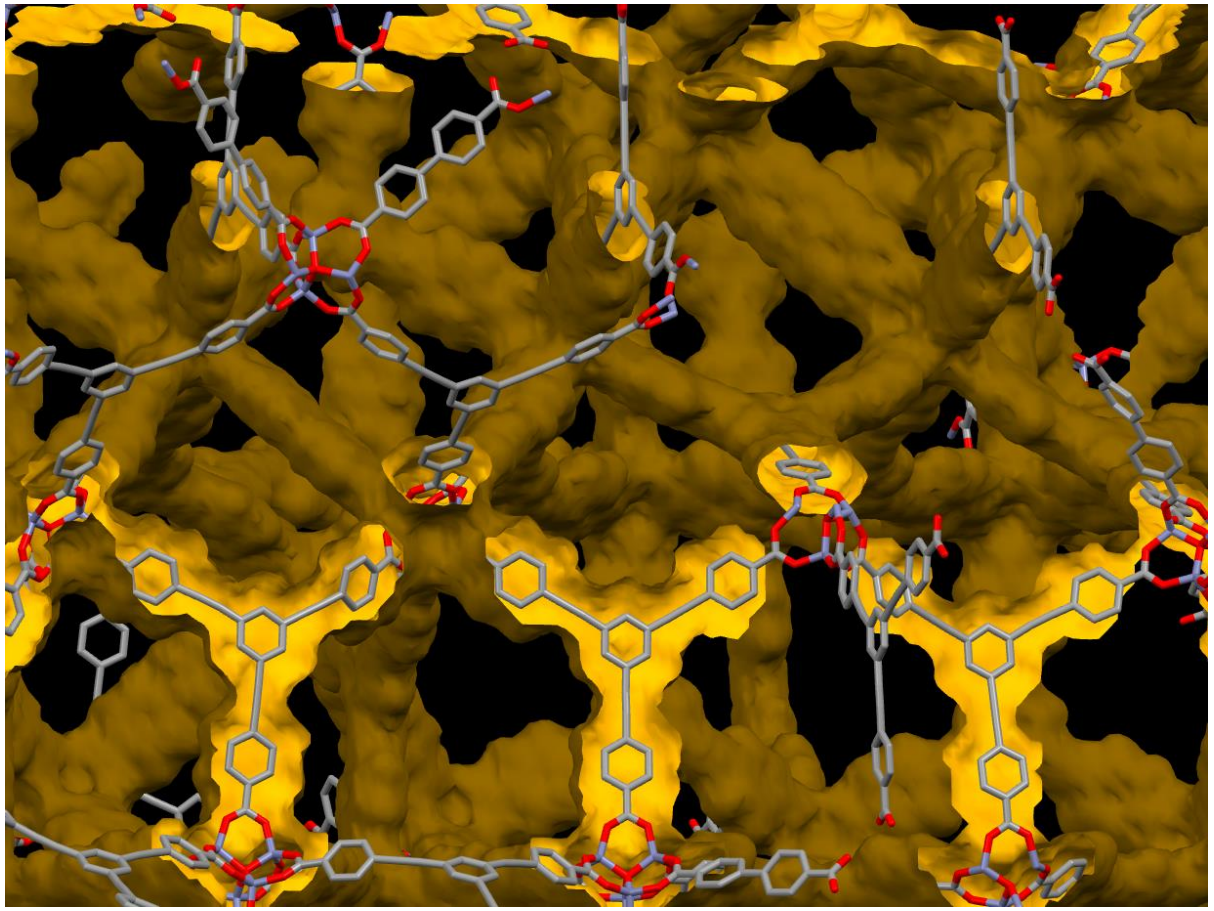
Metal–organic frameworks (MOFs)[1−5] are one of the most exciting recent advances in porous materials science. MOFs are crystalline materials containing metal clusters connected by organic linkers (Figure 1, inset) and are characterized by their wide range of well-defined and, in principle, tailorable pore sizes (from micro- to mesoporosity), pore geometries, high void fractions, and large surface areas. MOFs can reach apparent surface areas as high as 10 000 m^2/g vs ca. 1000 m^2/g for zeolites and up to 3500 m^2/g for activated carbons,[6,7] and their large pore volumes have been unsurpassed by any other porous material to date. These unique properties have paved the way for MOF research to grow substantially, and applications are being considered in many areas including gas storage,[6,7] separation,[8−10] catalysis,[11,12] and carbon capture[13−15] as well as biomolecule encapsulation,[16] drug delivery,[17−19] and imaging.[20]

Because of the modular nature of MOFs, an almost unlimited number of structures can be envisioned. The Cambridge Crystallographic Data Centre (CCDC) collates and curates the Cambridge Structural Database (CSD)—the world's repository of small molecule crystal structures, which includes small organics as well as MOFs and other porous materials.[21] As shown in Figure 1, the number of entries in the CSD has substantially increased over the last 44 years, reaching a milestone of 850 000+ entries in 2016. Among these, the number of MOFs has also increased dramatically in the past decade to an estimated number of ca. 70 000 materials, vide infra. This trend is only going to increase even further as the

P. Z. Moghadam, A. Li, S. B. Wiggin, A. Tao, A. G. P. Maloney, P. A. Wood, S. C. Ward, D. Fairen-Jimenez, *Chem. Mater.*, 2017, 29 2618-2625, DOI: 10.1021/acs.chemmater.7b00441

CCDC

MOFs today



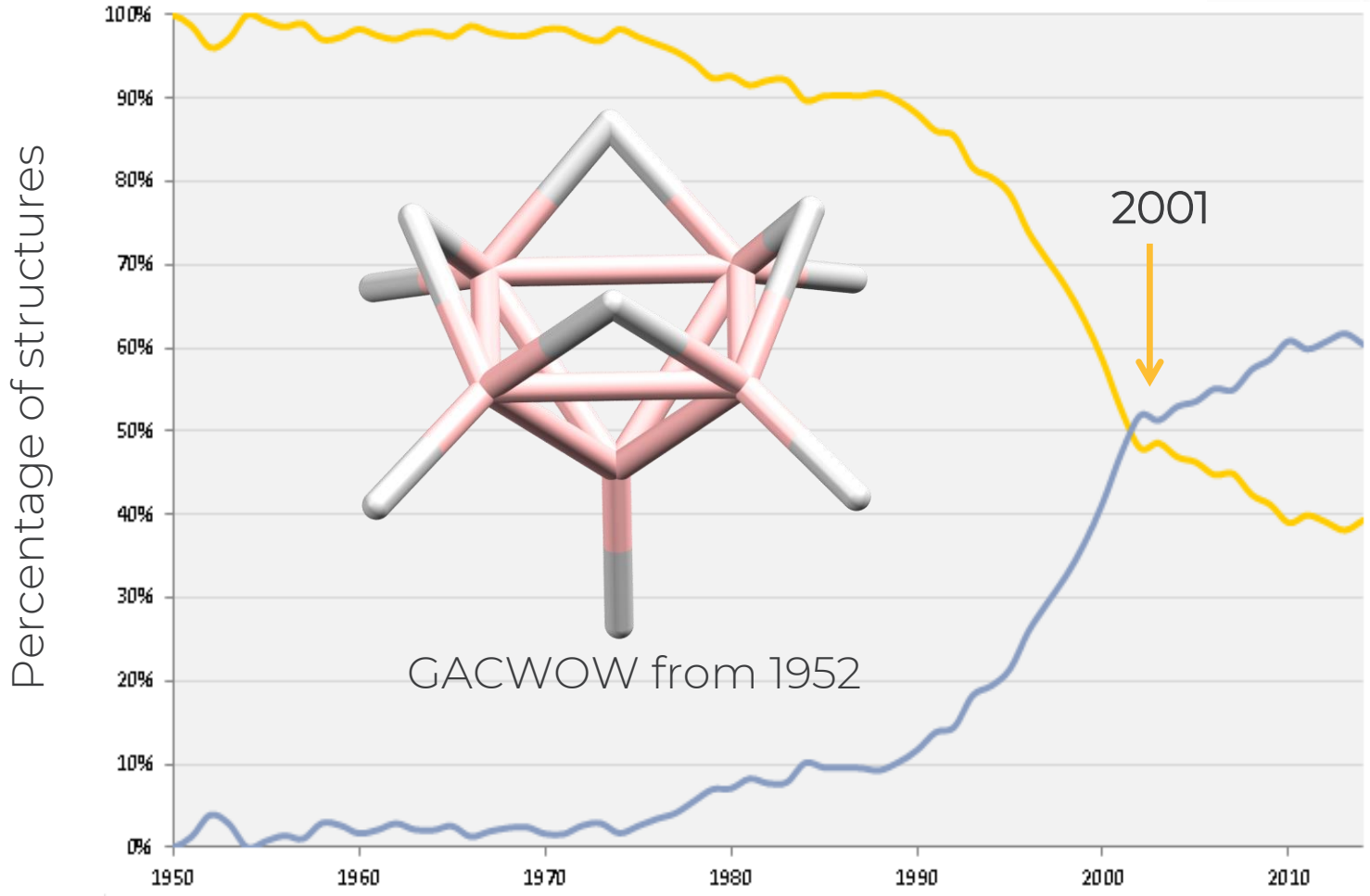
- Molecular sponges
- Gas storage
- Batteries
- Sensors
- Catalysis
- Purification
- Separation

CUSYAR

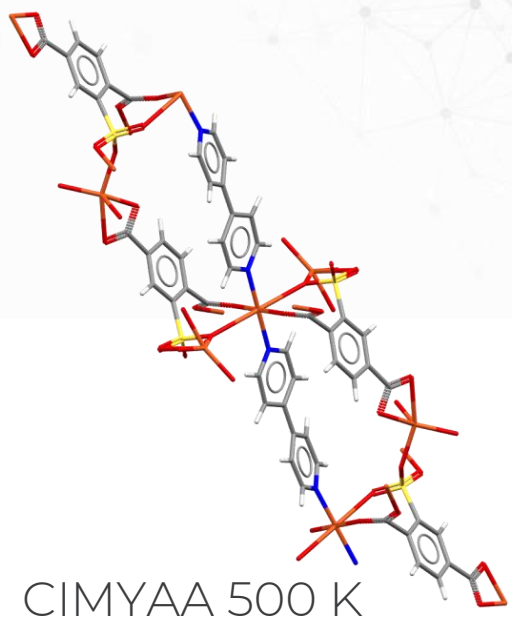
Hiroyasu Furukawa, Nakeun Ko, Yong Bok Go, Naoki Aratani, Sang Beom CHoi, Eunwoo Choi, A. Özgür Yazaydın, Randall Q. Snurr, Michael O'Keeffe, Jaheon Kim, Omar M. Yaghi, *Science*, 2010, 329, 424, DOI: 10.1126/science.1192160

CCDC

The dominance of cryostreams



% of structures
Between 273-323 K
< 273 K



CIMYAA 500 K

W.J.Dulmage, W.N.Lipscomb, *Acta Crystallogr.*, 1952, 5, p260, Yuyang Tian et al, Dalton Trans. 2014, 43, p1519

CCDC

Crystallography becomes more mainstream

- Data commonly reported in supplementary PDF files
- Crystallographer not always an author

The First Non-Pyrolytic Synthesis of a Semibuckminsterfullerene

Andrzej Sygula and Peter W. Rabideau*

Department of Chemistry, Louisiana State University
Baton Rouge, Louisiana 70803

Received August 11, 1998

The considerable amount of research attention given to C_{60} and the family of fullerenes has led to a growing interest in polynuclear aromatic hydrocarbons that may be considered to be fullerene fragments; that is, aromatic hydrocarbons with carbon frameworks that can be identified on the buckminsterfullerene surface.[1]

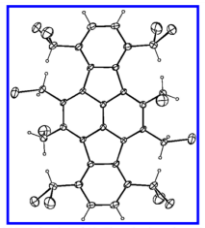


Figure 1. ORTEP drawing of **10**. The solvating toluene is not shown for clarity.

Indeed, this prediction was confirmed by X-ray crystal structure determination (Figure 1).[24] In contrast to the failure with **9**, dodecabromo **10** did, in fact, lead to modest yields of the semibuckminsterfullerene framework under treatment with low-valent titanium or vanadium. The major isolated product of the reductive coupling of **10** is (aromatized) **2**, accompanied by some dihydro-**2** and also by some incompletely cyclized byproducts.

Acknowledgment. This work was supported by the Division of Chemical Sciences, Office of Basic Energy Sciences, U.S. Department of Energy. We also thank Dr. Frank Fronczek for determination of the crystal structure of **10**.

Table 2. Atomic coordinates [x 10^4] and equivalent isotropic displacement parameters [$Å^2$ x 10^3] for 3. U(eq) is defined as one third of the trace of the orthogonalized U_{ij} tensor.

	x	y	z	U(eq)
W	1166.2(1)	7471.5(2)	-6424.7(9)	25(1)
Si	-325.2(8)	7259(1)	-5776(2)	24(1)
O	844(2)	7446(3)	-8267(5)	33(1)
C(1)	1787(4)	6550(6)	-6084(9)	56(3)
C(2)	2010(4)	5977(6)	-7419(11)	46(2)
C(3)	1500(5)	5448(8)	-7980(18)	106(6)
C(4)	2250(6)	6429(8)	-8817(16)	113(6)
C(5)	2532(4)	5454(6)	-6830(13)	74(4)
C(6)	1577(4)	8567(6)	-6049(9)	46(3)
C(7)	1767(4)	9080(5)	-7499(10)	39(2)
C(8)	2178(4)	9754(5)	-6914(12)	58(3)
C(9)	1221(4)	9431(6)	-8346(13)	57(3)
C(10)	2144(4)	8599(6)	-8720(11)	52(3)
C(11)	488(3)	7361(4)	-4977(8)	26(2)
C(12)	667(3)	7357(5)	-3157(8)	35(2)
C(13)	1340(3)	7553(6)	-2803(10)	52(3)
C(14)	287(3)	7969(5)	-2213(9)	43(2)
C(15)	561(4)	6539(5)	-2455(9)	48(2)
C(16)	-486(3)	6312(5)	-6896(9)	30(2)
C(17)	-1180(3)	6190(6)	-7091(12)	56(3)
C(18)	-205(3)	6276(5)	-8601(9)	37(2)
C(19)	-209(4)	5613(5)	-5955(11)	48(3)
C(20)	-952(3)	7322(5)	-4170(9)	29(2)
C(21)	-1355(3)	7966(6)	-4124(10)	38(2)
C(22)	-1839(3)	8007(6)	-3050(10)	46(2)
C(23)	-1943(3)	7402(6)	-1976(9)	47(2)
C(24)	-1561(3)	6771(6)	-1988(11)	49(3)
C(25)	-1075(3)	6730(5)	-3060(10)	41(2)
C(26)	-466(3)	8154(5)	-7074(9)	28(2)
C(27)	-270(3)	8863(5)	-6452(13)	39(2)
C(28)	-391(4)	9556(6)	-7219(13)	53(3)
C(29)	-693(4)	9542(6)	-8662(13)	57(3)
C(30)	-883(4)	8873(7)	-9291(11)	53(3)
C(31)	-789(4)	8181(6)	-8496(10)	41(2)

Tianniu Chen, Zhongzhi Wu, Liting Li, K.R.Sorasaenee, J.B.Diminnie, Hongjun Pan, I.A.Guzei, A.L.Rheingold, Ziling Xu, *JACS*, 1998, 120, 13519, DOI: 10.1021/ja982571l

CCDC

Today - Provenance and attribution

Add Crystallographer Details

Please check and update the details of the main crystallographer associated with this data

Use My Details

Crystallographer Name e.g. Frank Allen

Publishing Name e.g. Frank H. Allen

Email Address

Affiliation e.g. CCDC, 12 Union Road, Cambridge, CB2 1EZ

Additional details

Deposition Number	1885952
Data Citation	Thomas J. Aldrich, Micaela Matta, Weigang Zhu, Steven M. Swick, Charlotte L. Stern, George C. Schatz, Antonio Facchetti, Ferdinand S. Melkonyan, Tobin J. Marks CCDC 1885952: Experimental Crystal Structure Determination, 2019, DOI: 10.5517/ccdc.csd.cc219h51
Deposited on	18/12/2018

Crystallographer(s)

Crystallographer	Charlotte L. Stern
Affiliation	Northwestern University

Associated publications

Thomas J. Aldrich, Micaela Matta, Weigang Zhu, Steven M. Swick, Charlotte L. Stern, George C. Schatz, Antonio Facchetti, Ferdinand S. Melkonyan, Tobin J. Marks, *Journal of the American Chemical Society*, 2019, 141, 3274, DOI: 10.1021/jacs.8b13653

CCDC

1990s – A standardised data format

Acta Cryst. (1991). A**47**, 655–685

International Union of Crystallography

Commission on Crystallographic Data
Commission on Journals
Working Party on Crystallographic Information

The Crystallographic Information File (CIF): a New Standard Archive File for Crystallography*

BY SYDNEY R. HALL

Crystallography Centre, University of Western Australia, Nedlands 6009, Australia

FRANK H. ALLEN

Crystallographic Data Centre, University Chemical Laboratory, Lensfield Road, Cambridge CB2 1EW, England

AND I. DAVID BROWN

Institute for Materials Research, McMaster University, Hamilton, Ontario L8S 4M1, Canada

(*Received* 8 *April* 1991; *accepted* 28 *June* 1991)

Abstract

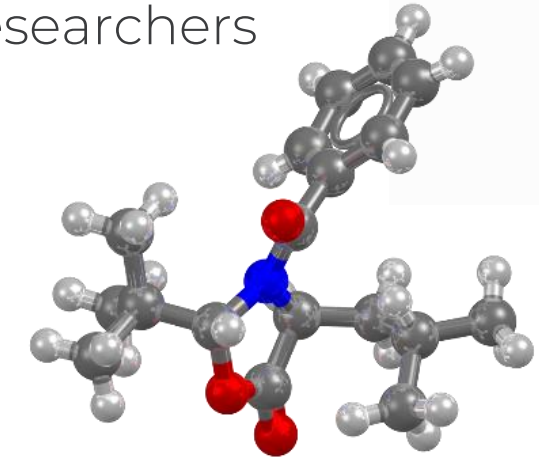
The specification of a new standard Crystallographic Information File (CIF) is described. Its development is based on the Self-Defining Text Archive and Retrieval (STAR) procedure [Hall (1991). *J. Chem. Inf. Comput. Sci.* **31**, 326–333]. The CIF is a general, flexible and easily extensible free-format archive file; it is human and machine readable and can be edited by a simple text editor. The CIF is designed for the ele transmission of crystallographic data between i ual laboratories, journals and databases: it ha adopted by the International Union of Crysta phy as the recommended medium for this p

Introduction

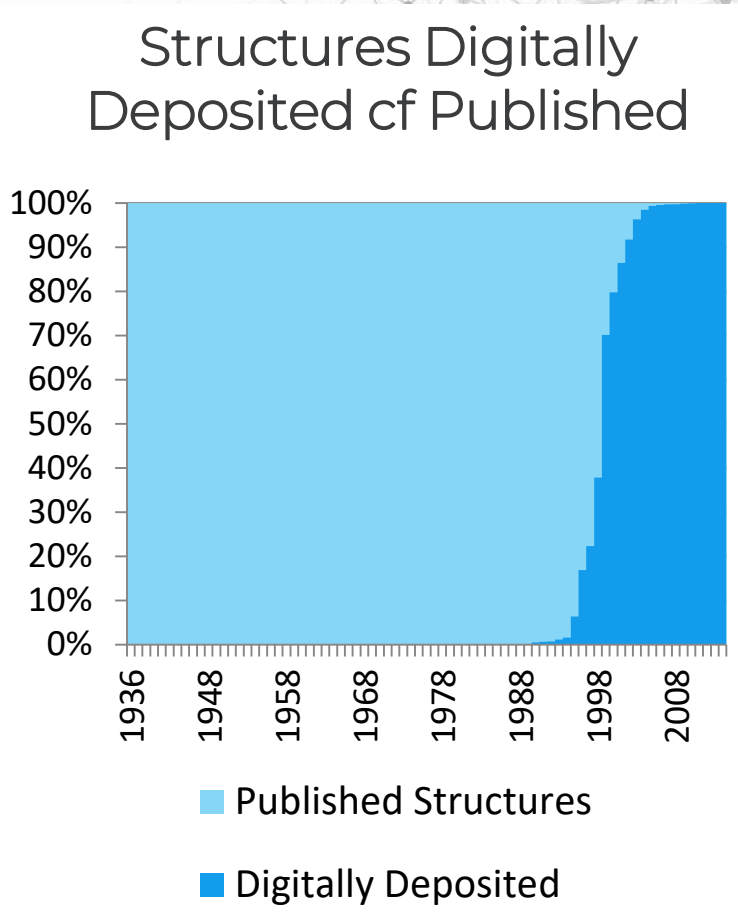
There is an increasing need in many branches of science for a uniform but flexible method of archiving and exchanging data in electronic form. Rapid advances in computer technology, coupled with the expansion of local, national and international networks, have fuelled the need for such a facility. The variety and relative inflexibility of existing data exchange formats have inhibited their ef-

```
loop_
_atom_site_label
_atom_site_type_symbol
_atom_site_fract_x
_atom_site_fract_y
_atom_site_fract_z
_atom_site_U_iso_or_equiv
_atom_site_adp_type
_atom_site_occupancy
_atom_site_symmetry_multiplicity
_atom_site_calc_flag
_atom_site_refinement_flags
_atom_site_disorder_assembly
_atom_site_disorder_group
C11 C1 0.5993(2) 1.0007(7) 0.8131(17) 0.044(3) Uani 0.50 1 d PDU A 1
S1 S 0.5321(3) 0.8260(6) 0.9322(3) 0.0327(11) Uani 0.50 1 d PDU A 1
C2 C 0.5529(4) 0.8802(9) 0.8184(9) 0.029(4) Uani 0.50 1 d PDU A 1
C3 C 0.5286(7) 0.8174(18) 0.7440(7) 0.031(4) Uani 0.50 1 d PDU A 1
H3A H 0.5350 0.8343 0.6771 0.037 Uiso 0.50 1 calc PR A 1
C4 C 0.4918(8) 0.7220(19) 0.7783(8) 0.027(4) Uani 0.50 1 d PDU A 1
C5 C 0.4900(6) 0.7171(14) 0.8779(9) 0.029(4) Uani 0.50 1 d PDU A 1
C12 C1 0.3202(2) 0.4982(6) 1.0830(5) 0.0586(15) Uani 0.50 1 d PDU A 1
S2 S 0.38755(19) 0.6658(5) 0.9578(5) 0.0400(10) Uani 0.50 1 d PDU A 1
```

- Community adoption
  - Software vendors
  - Repositories
  - Publishers
  - Researchers



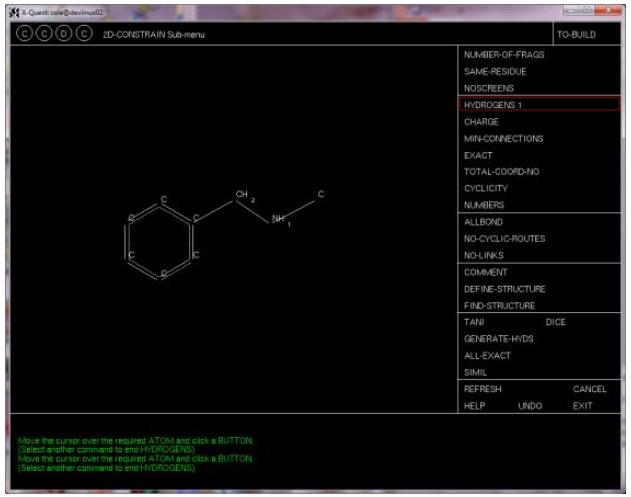
VOBYUG

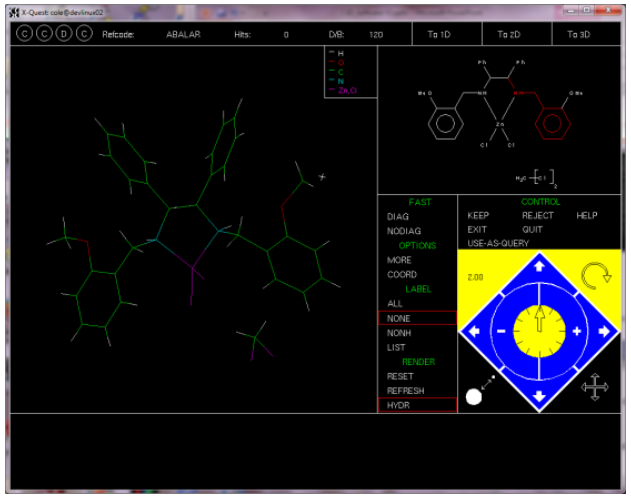


A.C.Willis, A.L.J.Beckwith, M.J.Tozer, *Acta Crystallographica,Section C: Crystal Structure Communications*, 1991, 47, 2276, DOI: 10.1107/S0108270191009630

CCDC

1991 - New search functionality





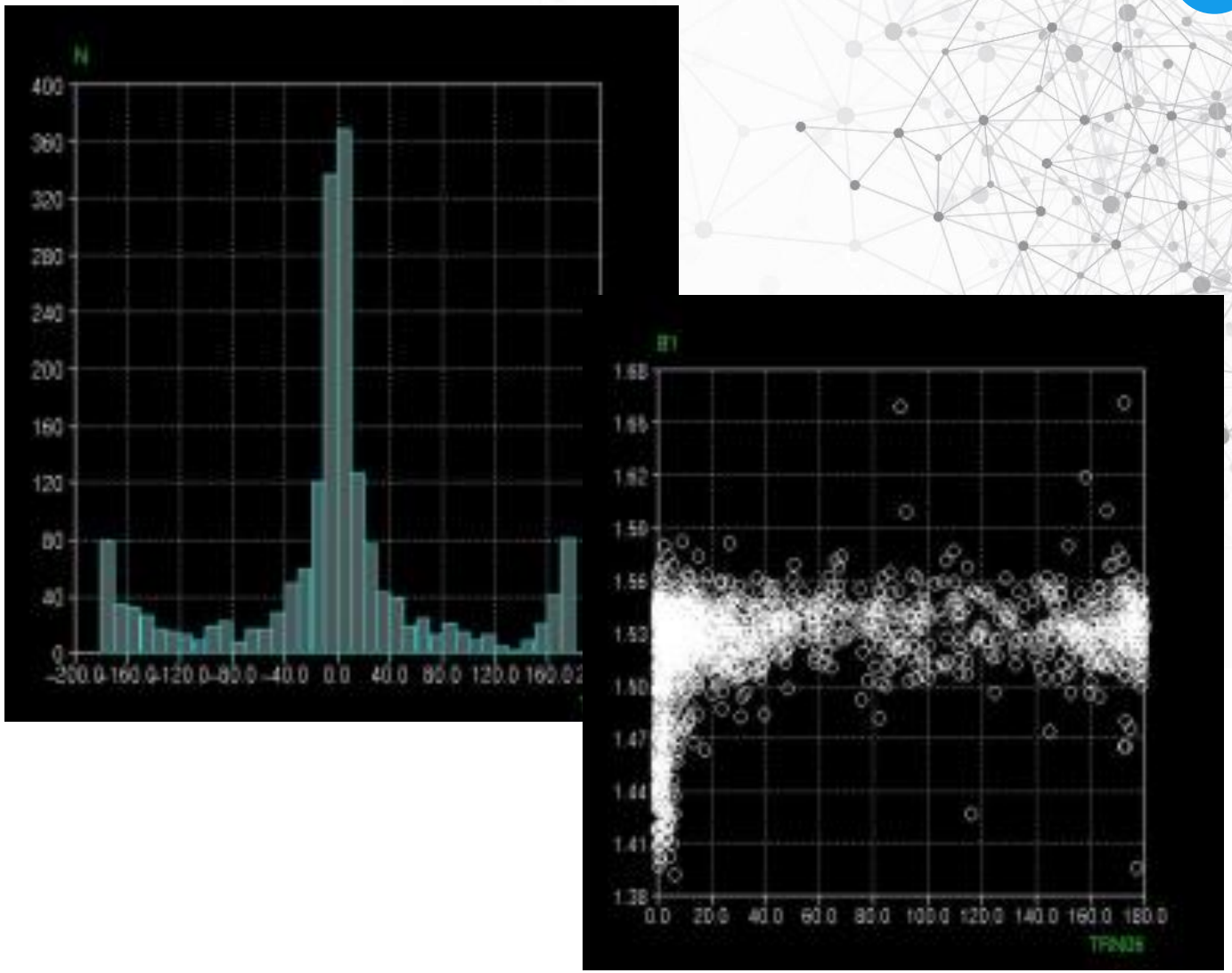
- *The development of molecule sketching and visualisation*
  - Text editor
  - Quest
- Introduction of sketch searches was quite revolutionary

The development of versions 3 and 4 of the Cambridge Structural Database system. F. H. Allen; J. E. Davies; J. J. Galloy; O. Johnson; O. Kennard; C. F. Macrae; E. M. Mitchell; G. F. Mitchell; J. M. Smith; D. G. Watson; J. Chem. Inf. Comput. Sci.; 31; 187-204; 1991 10.1021/ci00002a004

CCDC

1994 - VISTA

- A program for the analysis and display of data retrieved from the CSD
  - Allowed trends in molecular geometry or intermolecular interactions to be plotted and correlated



1997 – A knowledge base

Isostar - A Knowledge Base of Intermolecular Interactions

- Uses the wealth of information in:
  - Cambridge Structural Database
  - Protein Data Bank (protein-ligand complexes only)
- Allows you to investigate the frequency and characteristics of intermolecular interactions between pairs of chemical groups

Isostar: A library of information about non-bonded interactions Bruno I. J., Cole J. C., Lommerse J. P. M., Rowland R. S., Taylor R. and Verdonk M. L. (1997) J. Comput.-Aided Mol. Des., 11, 525-537

CCDC

2000s

Blu-ray Disc™

WIKIPEDIA
The Free Encyclopedia

facebook

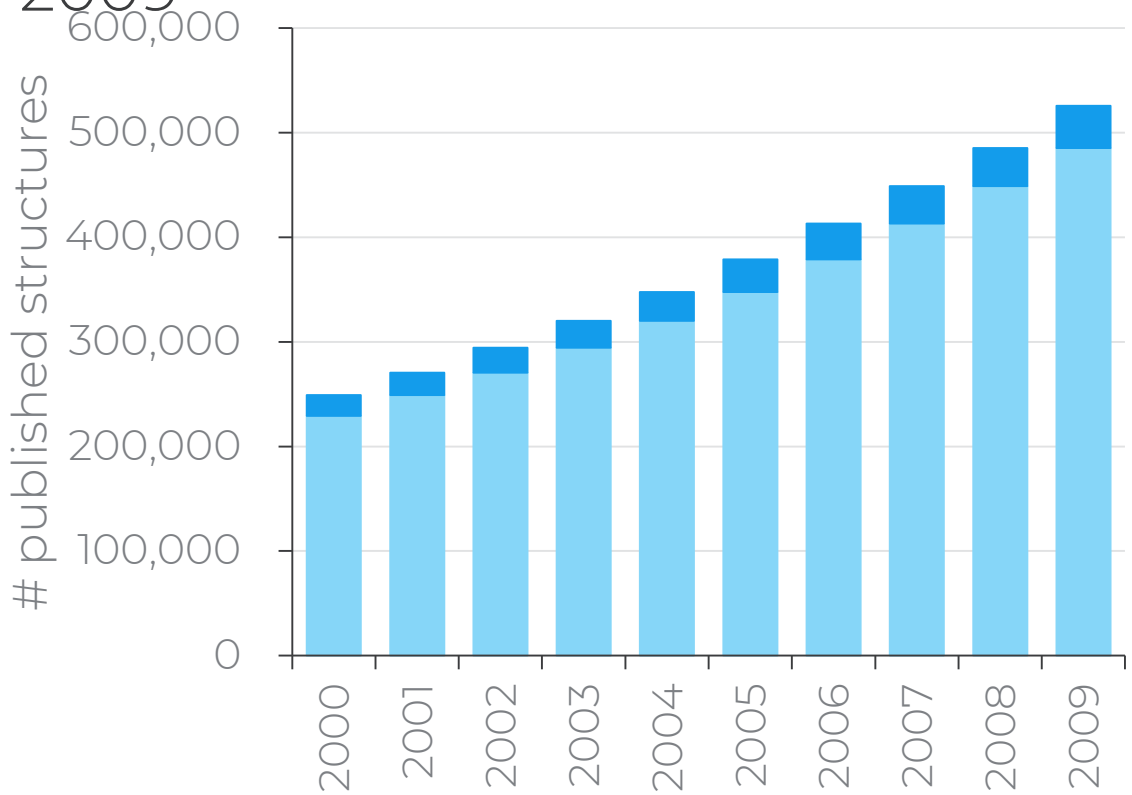
YouTube

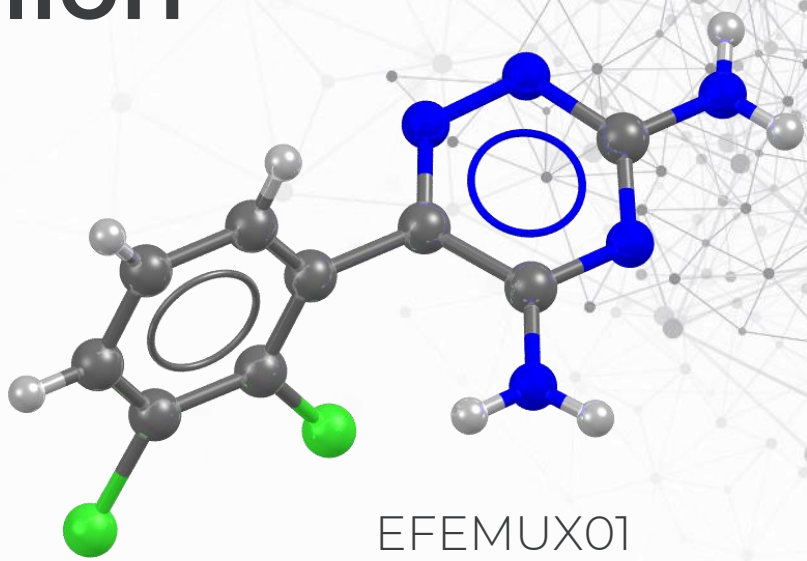
https://www.techrepublic.com/pictures/tech-nostalgia-the-top-10-innovations-of-the-2000s/10/

CCDC

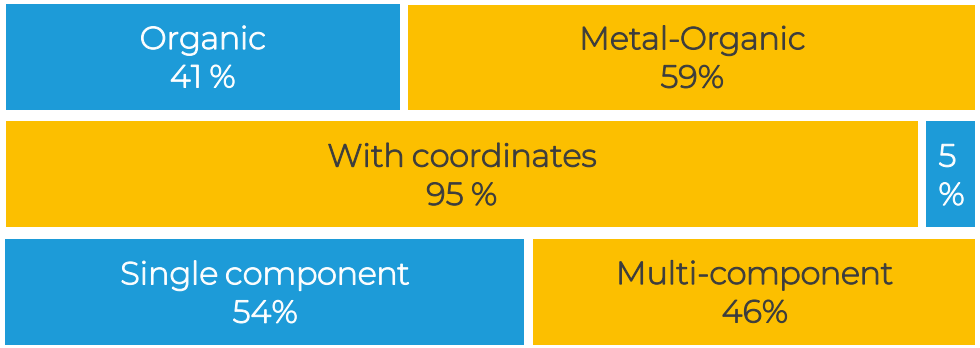
2000-2009 – ½ way to a million

- CSD doubled in size
- ¼ millionth in 2001, ½ millionth in 2009





EFEMUX01
500,000th CSD structure from 2009

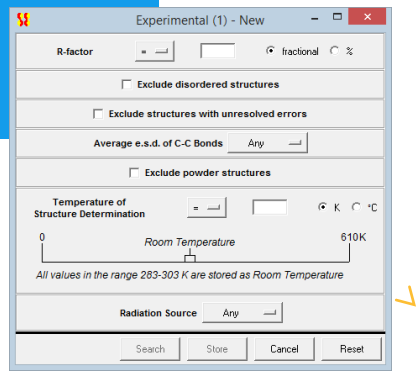


B.Sridhar, K.Ravikumar, (*Acta Crystallogr.,Sect.C:Cryst.Struct.Commun.* 2009, 65, o460, DOI:10.1107/S0108270109030960

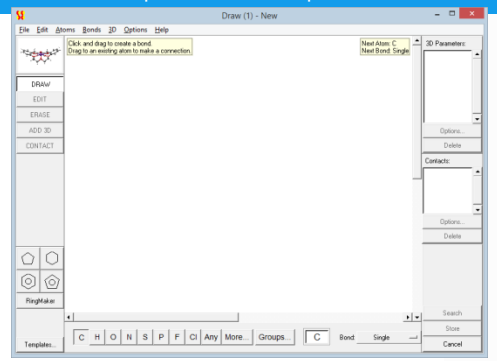
CCDC

2000 – ConQuest and the desktop revolution

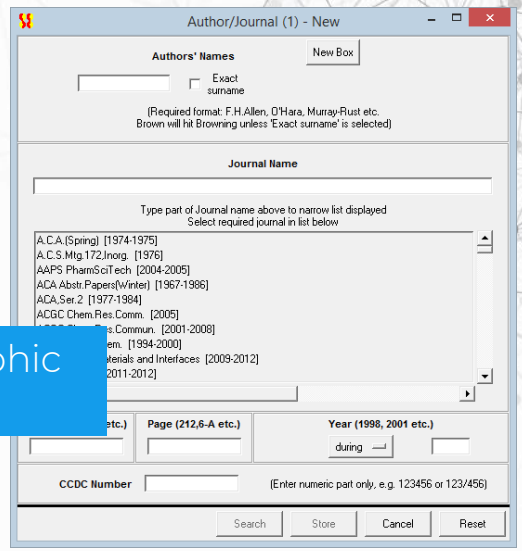
Search for information relating to the structure determination experiment

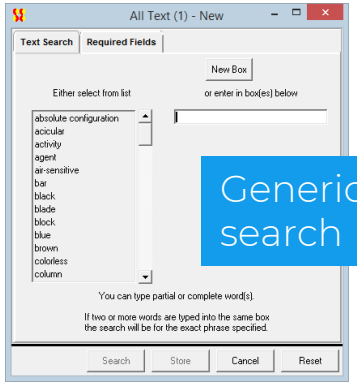


Powerful sketcher, enabling set-up of basic substructure searches and complex 3D queries



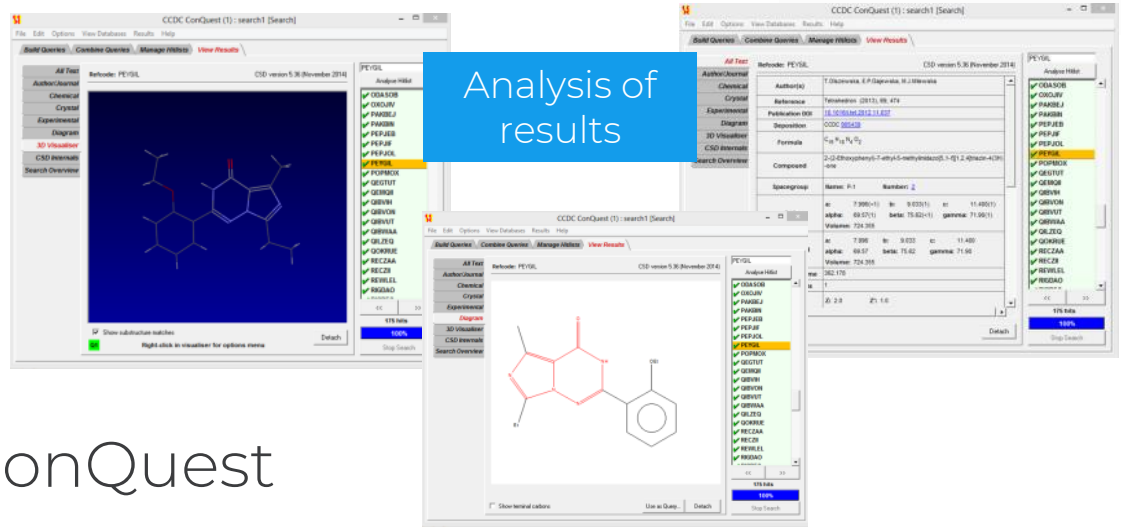
Bibliographic search



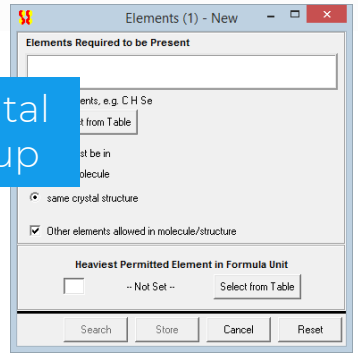


Generic text search

Analysis of results



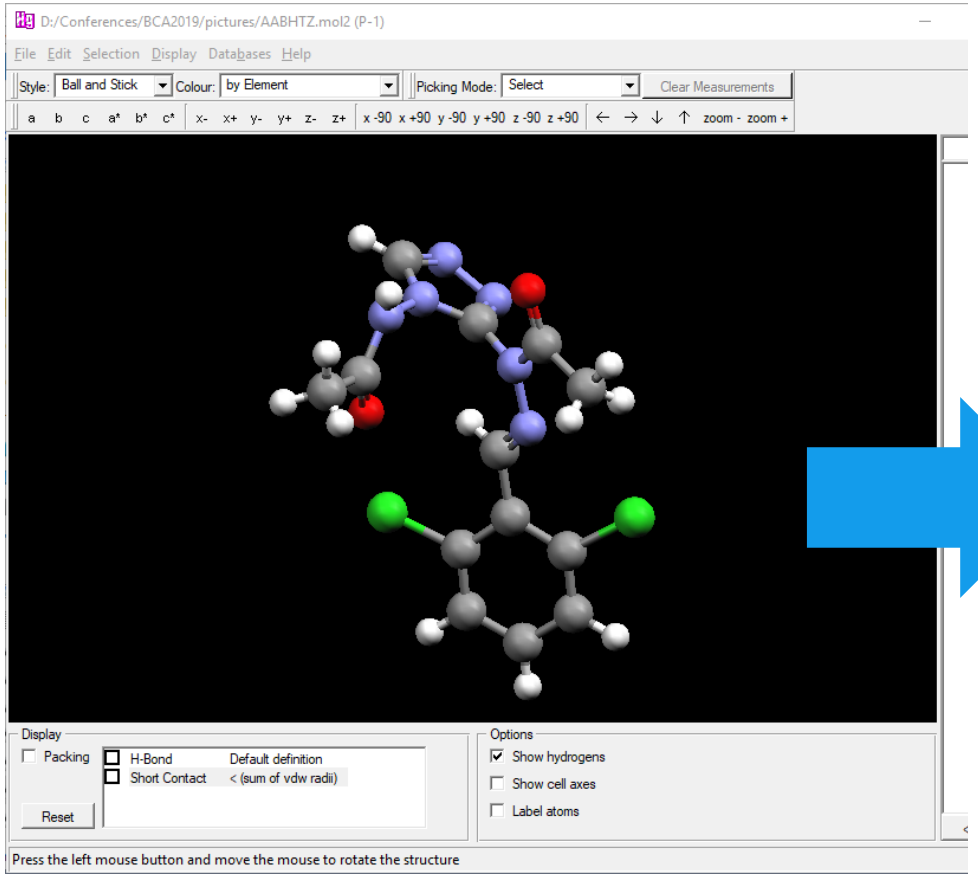
Elemental make-up

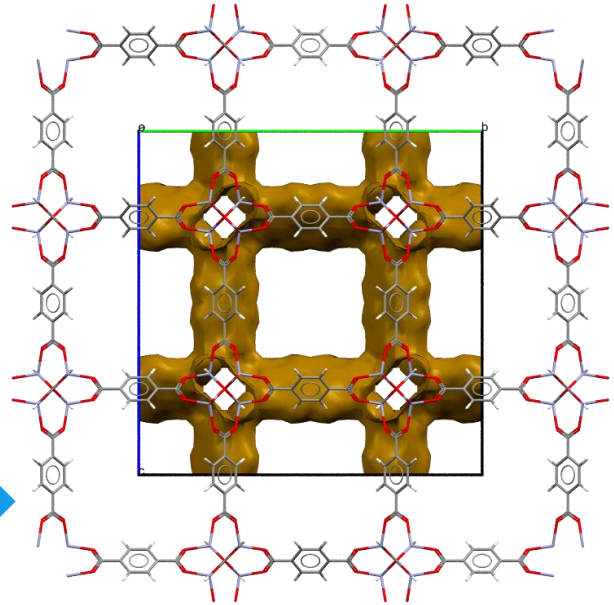


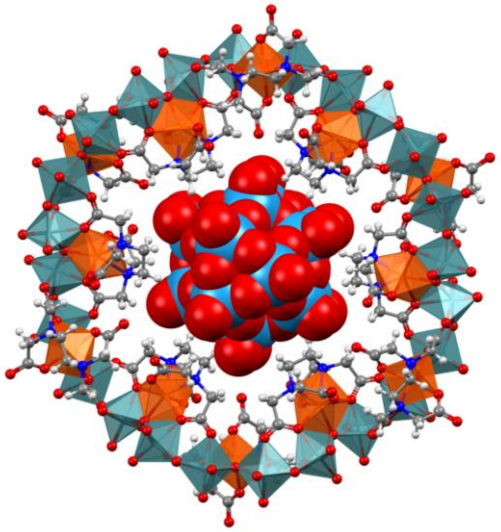
From Quest to ConQuest

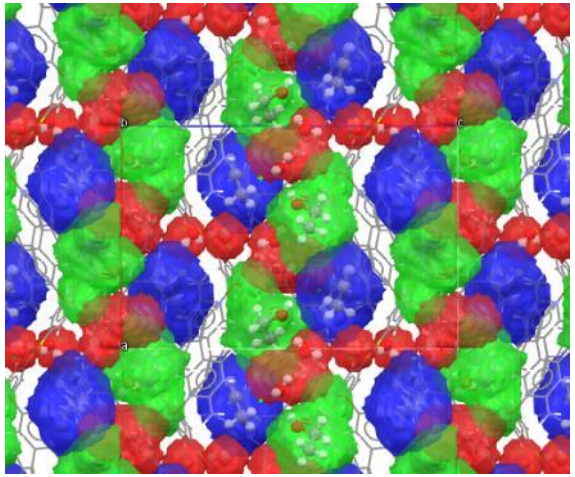
CCDC

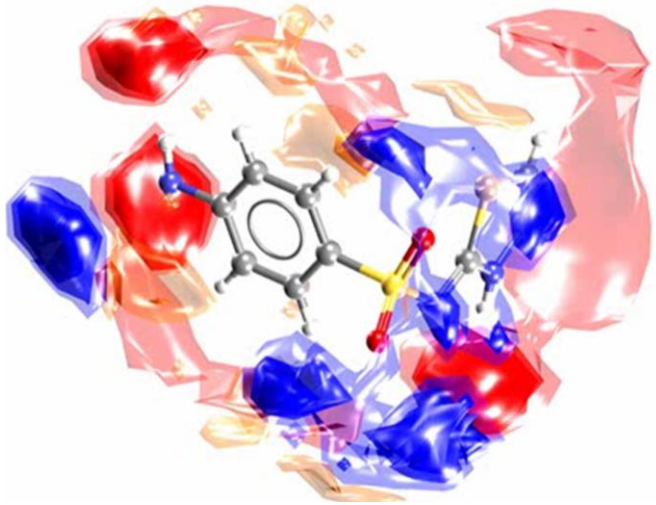
2001 - New software to visualise structures









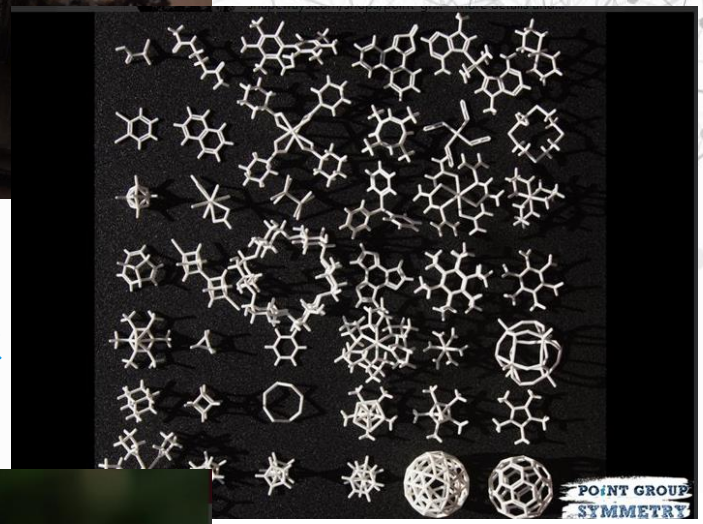


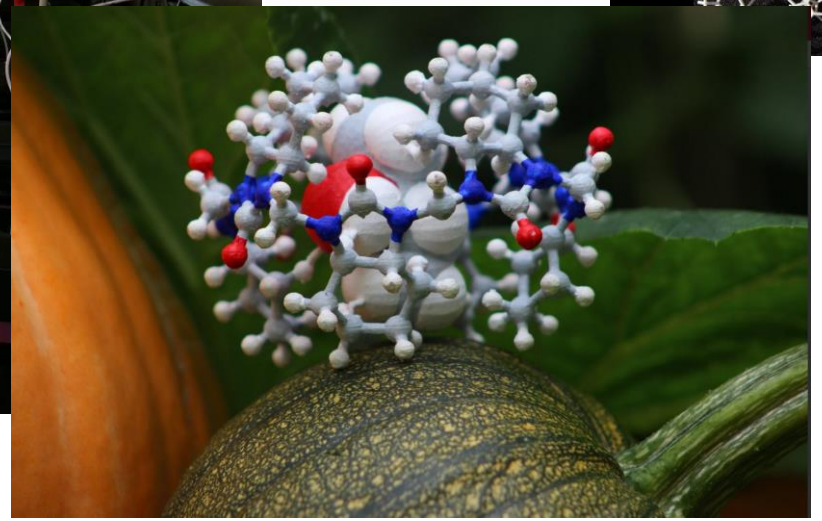
Visualisation today

Learning point-group symmetry through 3D printed models

Click to download the worksheet for use in the classroom
Click to download the answer key
This module was developed by Anton Savchenkov and uses 3D printed models Shapeways site.

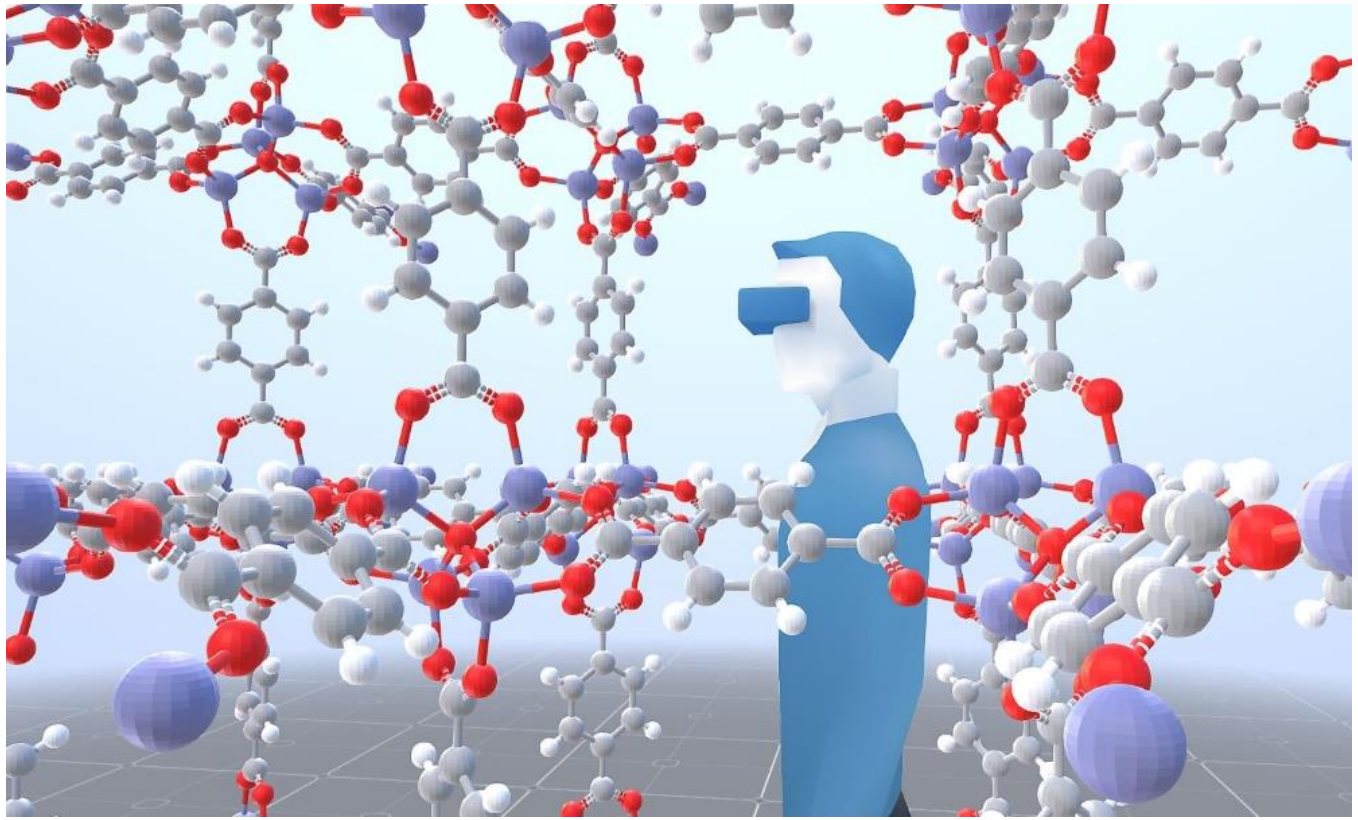
#CSD3DPrint

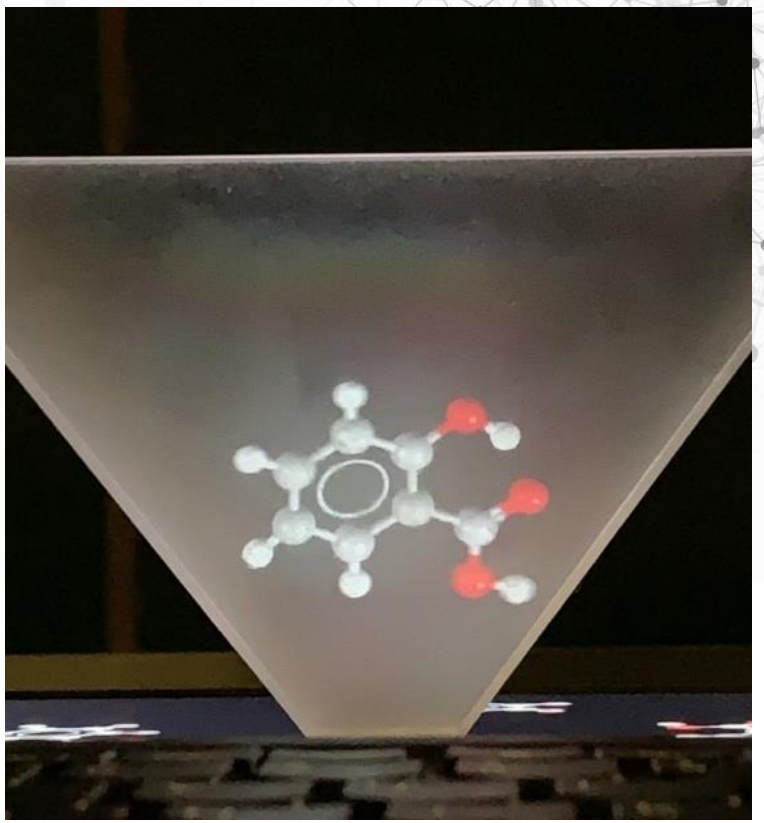




CCDC

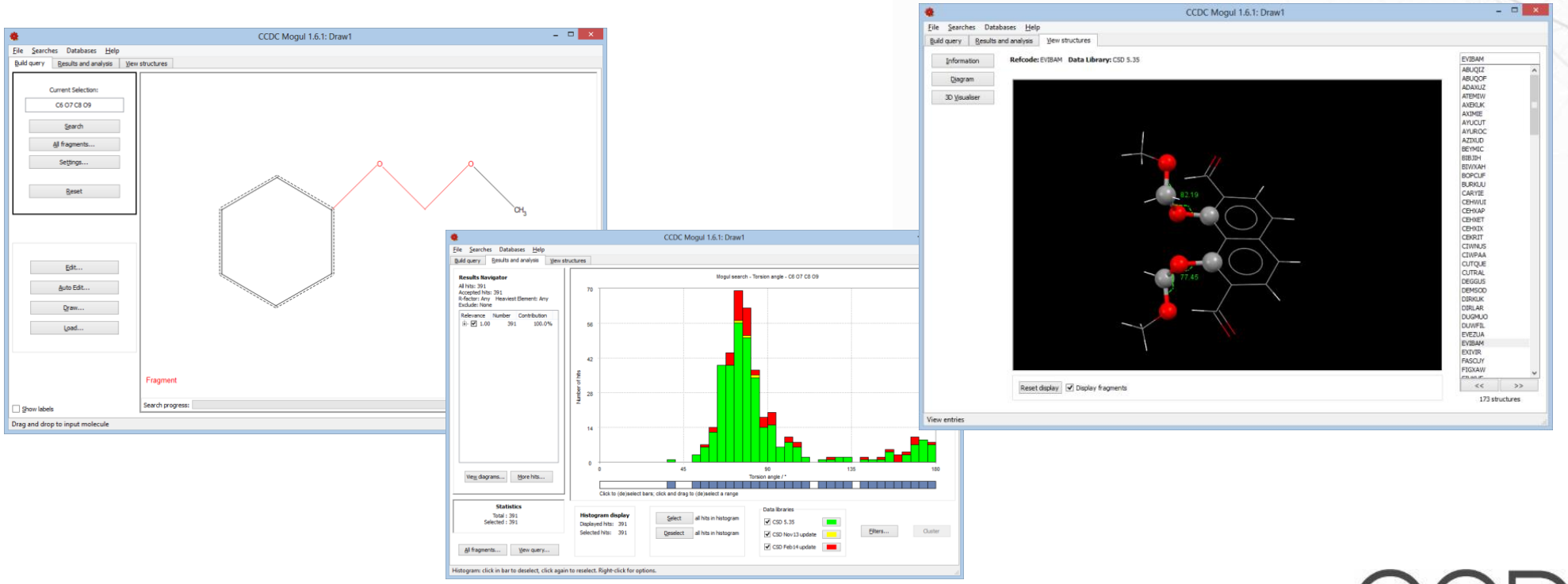
Visualising the future





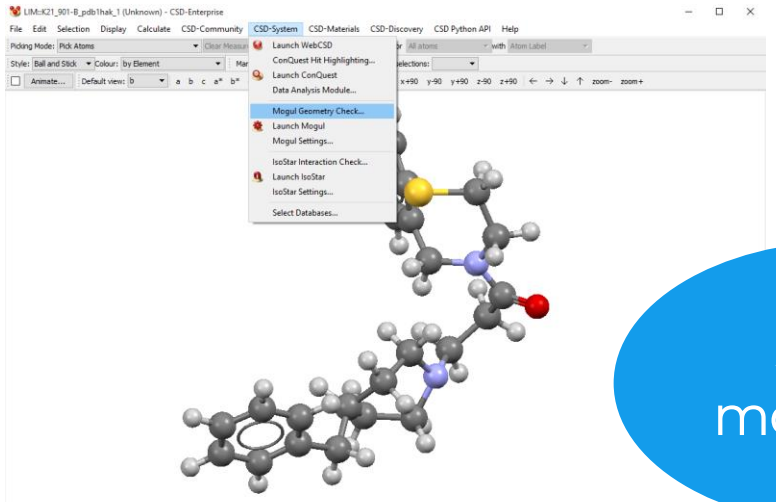
2004 - A second knowledge base

- Mogul contains geometrical information (bond lengths, angles and torsions) derived from structures in the CSD

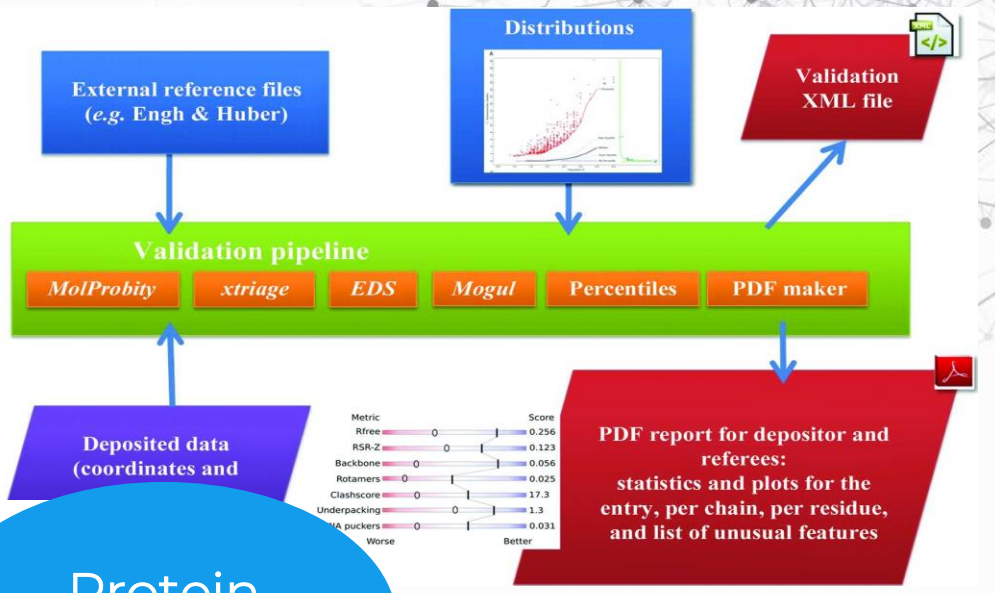


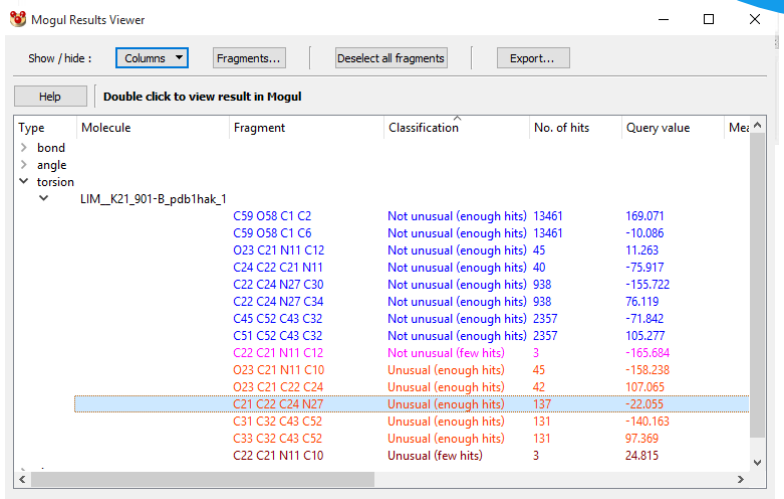
Ian Bruno, Robin Taylor & many others

Using the knowledge base today

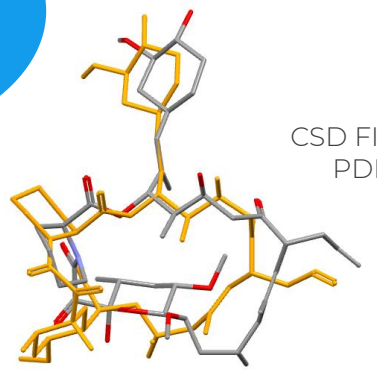


Small molecules





Protein ligands



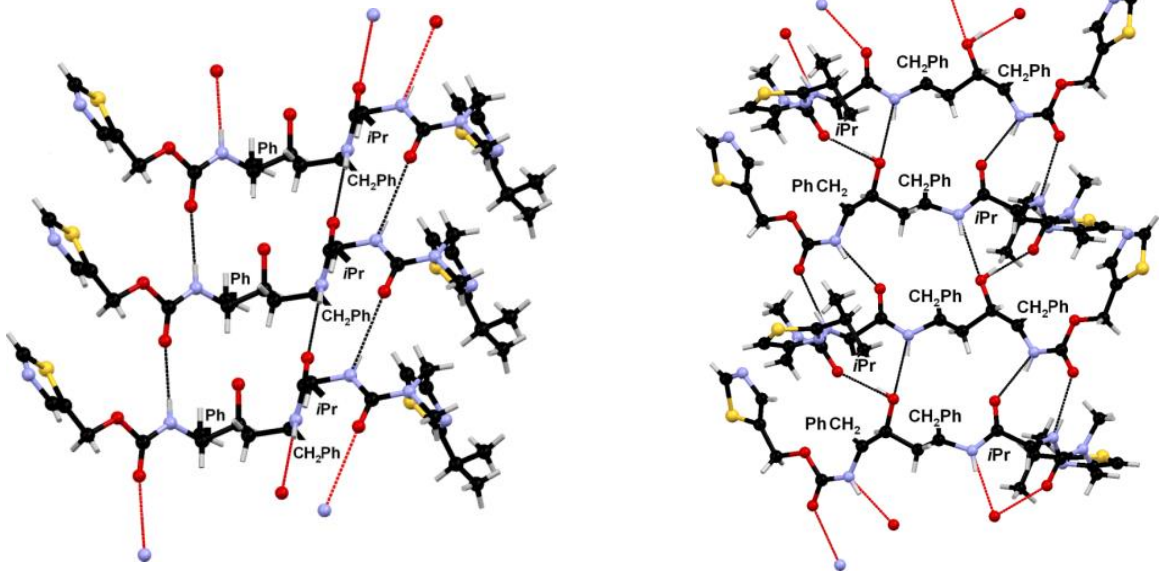
CSD FINWEE10
PDB FK5

I.J.Bruno, G.P.Shields, R.Taylor, *Acta Cryst.* (2011). B67, 333-349 DOI: 10.1107/S0108768111024608

CCDC

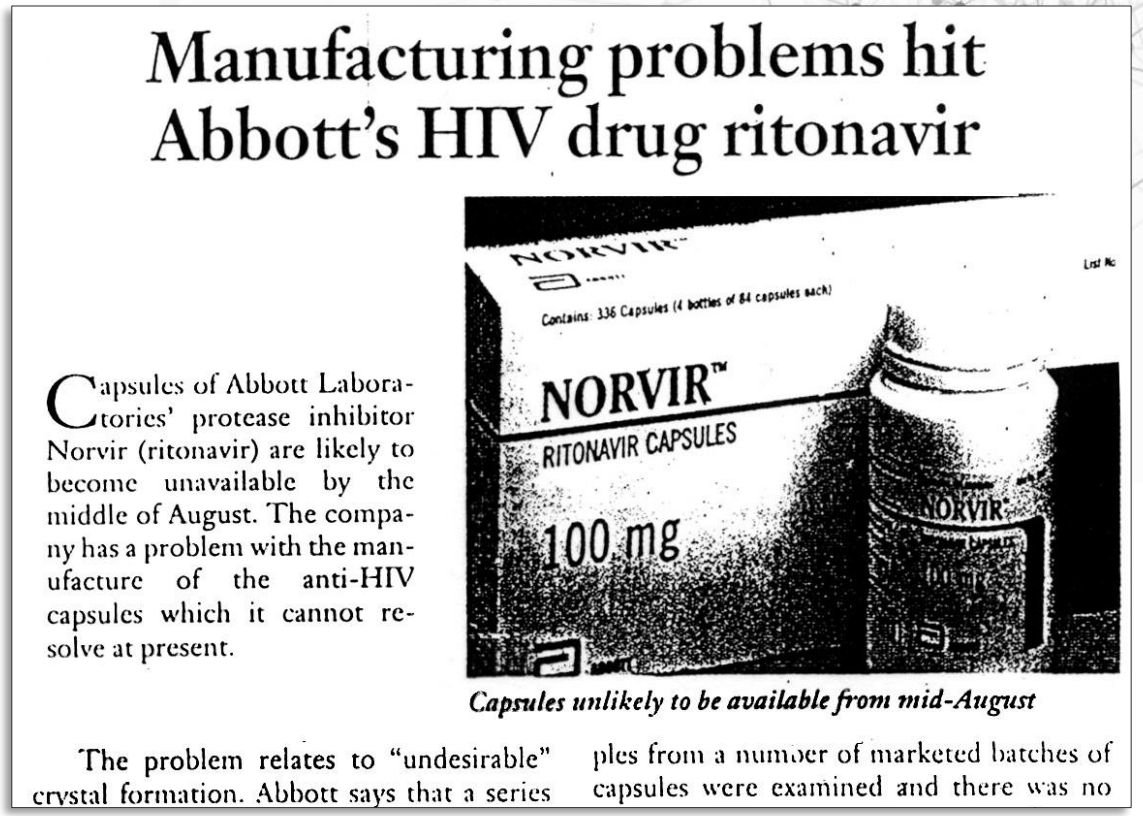
Knowledge from the collection?

Can structural knowledge mitigate risk?



Different interactions

Different solubility
Different stability



Manufacturing problems hit Abbott's HIV drug ritonavir

Capsules of Abbott Laboratories' protease inhibitor Norvir (ritonavir) are likely to become unavailable by the middle of August. The company has a problem with the manufacture of the anti-HIV capsules which it cannot resolve at present.

Capsules unlikely to be available from mid-August

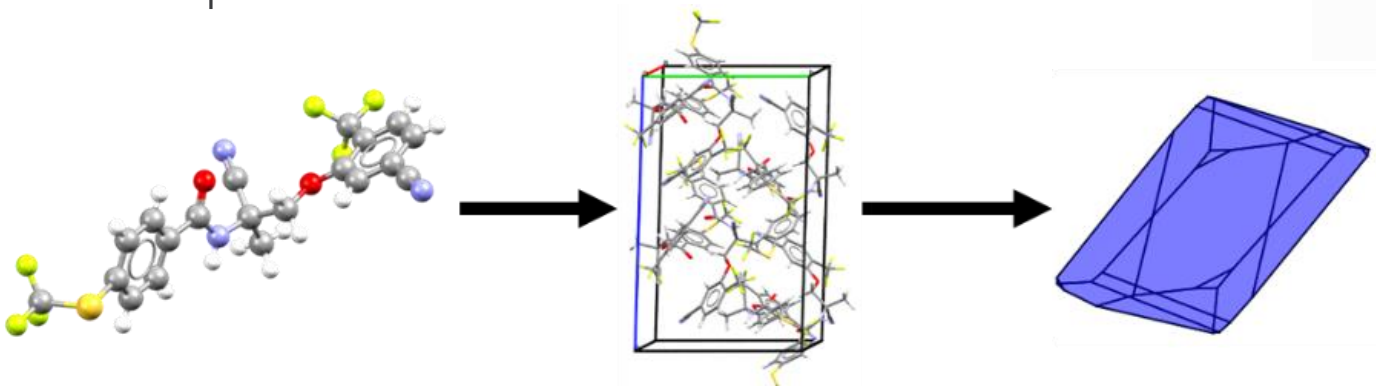
The problem relates to "undesirable" crystal formation. Abbott says that a series of samples from a number of marketed batches of capsules were examined and there was no

Bauer *et al. Pharm. Res.*, (2001) 859, DOI:10.1039/B910882C

CCDC

The start of solid form informatics

- The term "**solid form informatics**" was first introduced in mid-2000s
  - *Use of structural knowledge to inform key decisions in pharmaceutical development*
- Now a key part of the solid form development workflow at most major pharmaceutical companies



Molecule Form Particle

JPP JOURNAL OF Pharmacy and Pharmacology

JPP Journal of Pharmacy And Pharmacology

Research Paper

The integration of solid-form informatics into solid-form selection

Neil Feeder[a], Elna Pidcock[a], Anthony M. Reilly[a], Ghazala Sadiq[a], Cheryl L. Doherty[b], Kevin R. Back[b], Paul Meenan[c] and Robert Docherty[b]

[a]The Cambridge Crystallographic Data Centre, Cambridge, [b]Pharmaceutical Science, Pfizer Global R&D, Sandwich, UK and [c]Pharmaceutical Science, Pfizer Global R&D, Groton, USA

Keywords
Cambridge Structural Database; quality-by-design; solid-form informatics; solid-form selection

Correspondence
Robert Docherty, Pharmaceutical Science, Pfizer Global R&D, Sandwich CT13 9NJ,

Abstract

Objectives To demonstrate how the use of structural informatics during drug development assists with the assessment of the risk of polymorphism and the selection of a commercial solid form.

Methods The application of structural chemistry knowledge derived from the hundreds of thousands of crystal structures contained in the Cambridge Struc-

CrystEngComm

PAPER View Article Online View Journal | View Issue

Knowledge-based approaches to co-crystal design†

Cite this: *CrystEngComm*, 2014, 16, 5839

Peter A. Wood,* Neil Feeder, Matthew Furlow, Peter T. A. Galek, Colin R. Groom and Elna Pidcock

Received 12th February 2014, Accepted 10th March 2014

DOI: 10.1039/c4ce00316k

www.rsc.org/crystengcomm

The use of knowledge-based methods has been intimately connected with the field of co-crystal design since the seminal papers of Etter and Desiraju in the 1990s. Here we explain and exemplify how rational co-crystal design has been carried out in the past using crystal structure knowledge as well as presenting emerging methodologies for knowledge-based co-former selection.

1. Introduction

The definition of a co-crystal,[1–4] whether this is the correct term to use[5] and even whether it should contain a hyphen has been debated at length in the literature. At the most basic level, a co-crystal is a crystalline form composed from two or more components with a particular stoichiometry. Generally, the definitions used in the Crystal Engineering field also place some or all of the following requirements on the solid forms in question:

1. All components are organic species (ruling out inorganics or organometallics).
2. None of the components are charged (otherwise classified as salts).
3. None of the components are water (otherwise classified as hydrates).

the structure of a synthetic β-*C*-arylglucoside published by Deshpande and co-workers in 2012.[7] This crystal structure contains water, ethanol, a zwitterionic component and a neutral API, so could be described as a co-crystal, hydrate, solvate, zwitterionic salt or all of the above. Practically, the definitions used will also vary based on the application of the resulting co-crystal and the approaches for co-crystal design. In this paper we focus mostly on systems where all the components are neutral, organic species and point out if any further assumptions or requirements are added through the use of a certain design method.

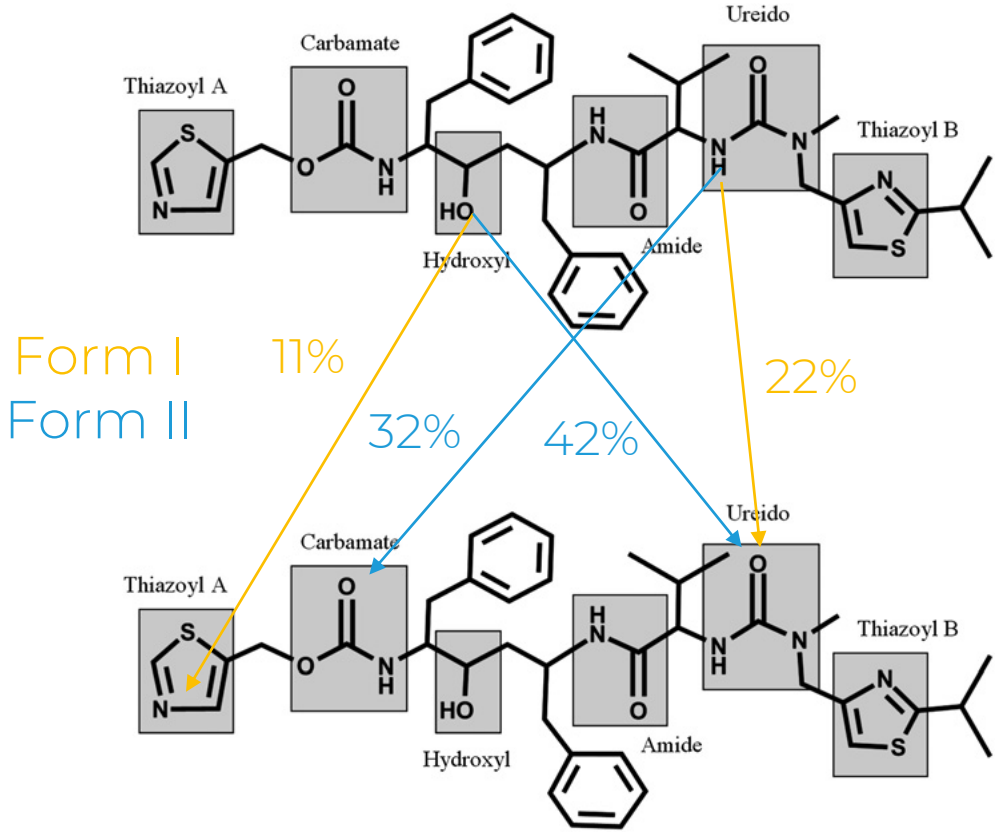
Co-crystals are of interest in both academic research and industrial application as they represent the opportunity to access new solid forms and therefore modified physical properties. This contrasts with polymorphism which, when observed, can be difficult to control and rarely represents an opportunity to

Robert Docherty *et al Journal of Pharmacy and Pharmacology*, (2015) 67, 857. doi: 10.1111/jphp.12394.
Taylor *et al. J. Chem. Inf. Model.*, (2014) 54 (9), 2500. Wood, P. A. *et al. CrystEngComm* (2013) 15, 65

CCDC

Predicting unlikely interactions

Predictive analytics is used to identify the likelihood of specific molecular interactions occurring from similar crystal structures



The integration of solid-form informatics into solid-form selection

Neil Feeder[a], Elna Pidcock[a], Anthony M. Reilly[a], Ghazala Sadiq[a], Cheryl L. Doherty[b], Kevin R. Back[b], Paul Meenan[c] and Robert Docherty[b]

One in half a million: a solid form informatics study of a pharmaceutical crystal structure

Peter T. A. Galek,*[a] Elna Pidcock,[a] Peter A. Wood,[a] Ian J. Bruno[a] and Colin R. Groom[a]

Navigating the Solid Form Landscape with Structural Informatics

Peter T. A. Galek, Elna Pidcock, Peter A. Wood, Neil Feeder, Frank H. Allen

Book Editor(s): Yuriy A. Abramov

Knowledge-based H-bond prediction to aid experimental polymorph screening

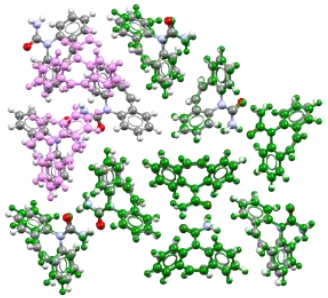
Peter T. A. Galek,*[ab] Frank H. Allen,[a] László Fábián[ab] and Neil Feeder[c]

Galek *et al*, *CrystEngComm*, (2009), 11, 2634 - 2639

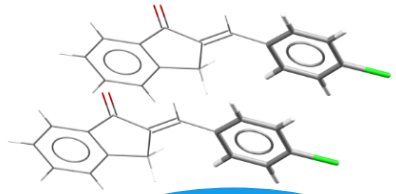
CCDC

From analysis to risk assessment to design

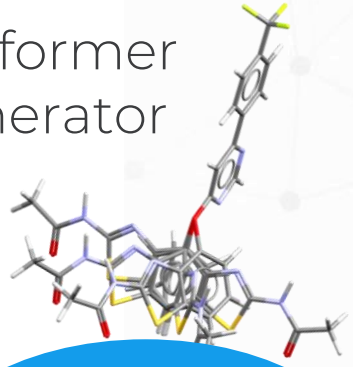
Crystal Packing Similarity



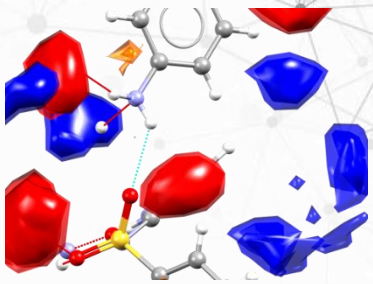
Motif Search & Packing Feature Search



Conformer Generator



Full Interaction Maps

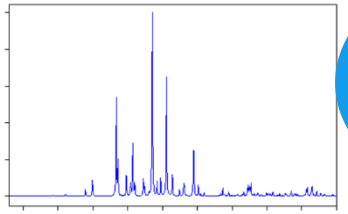


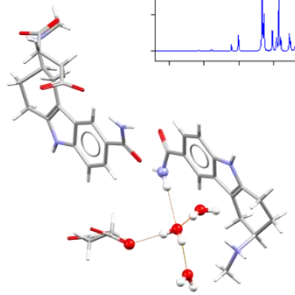
Complex Structural Analysis

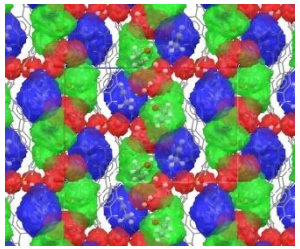
Solid Form Risk Assessment

Solid Form Design

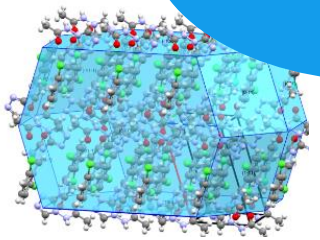
DASH



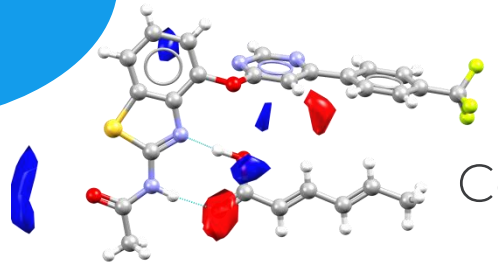




Hydrate Analyser & Solvate Analyser

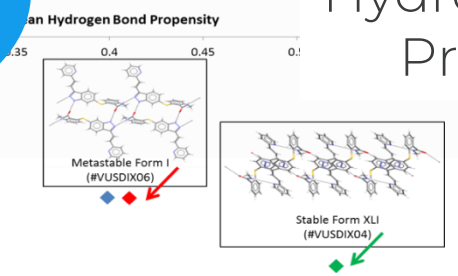


Calculations



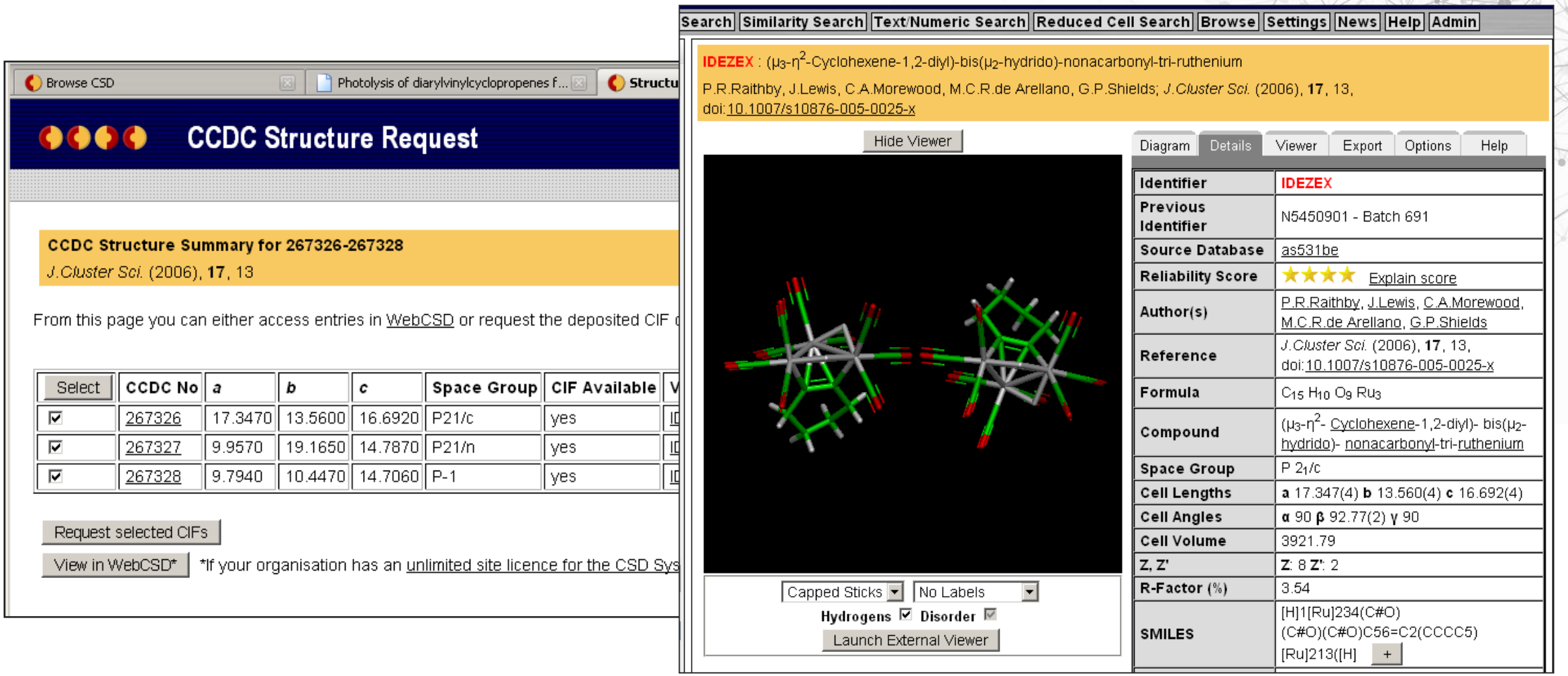
Molecular Complementarity

Hydrogen Bond Propensity

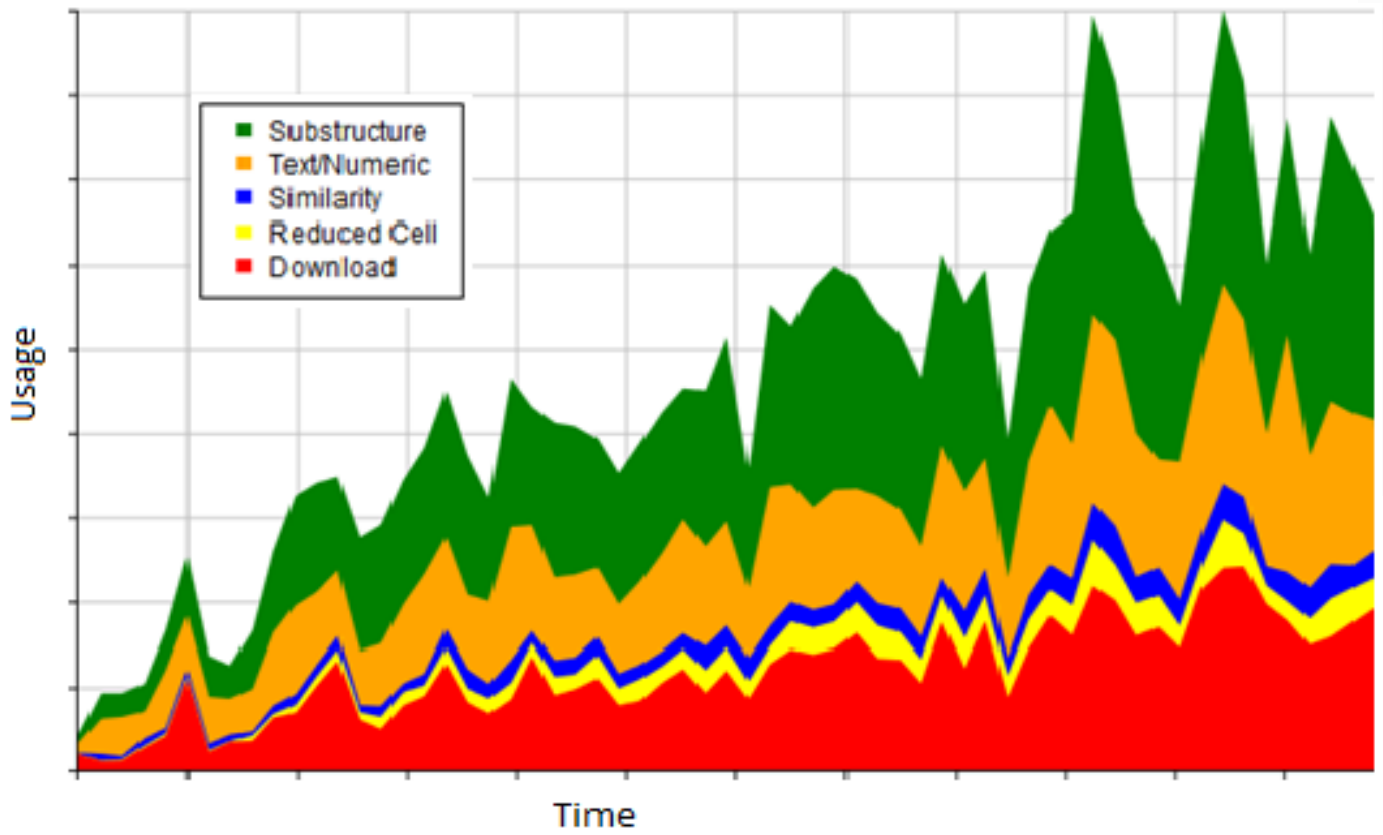


CCDC

The 2000s - The CSD on the web



WebCSD v1



- A range of search options
  - Structure
  - Text numeric
  - Reduced cell
  - Similarity
- Starts to become used in education

CCDC

Favourite structure searches.....

- AABHTZ
- KITTEN
- DISNEY
- DONKEY
- BATMAN
- BADBOY
- MUPPET
- ROBBER
- INDIGO
- HOTGOO
- STRICT
- GINGER

CCDC

From publications to CSD data

ROYAL SOCIETY OF CHEMISTRY

McKervey, A. R. Maguire, S. M. Tuladhar and M. Fiona Twohig, *J. Chem. Sc*
1047–1054 **DOI**: 10.1039/P19900001047; (*b*) H. Duddeck, *J. Chem. Soc.,*
1055–1063 **DOI**: 10.1039/P19900001055; (*c*) P. Panne and J. M. Fox, *J. Ar*
External Links.

Footnote

† Electronic supplementary information (ESI) available: Experimental procedures and spectrosc
615536. For ESI and crystallographic data in CIF or other electronic format see DOI: 10.1039/b8

This journal is © The Royal Society of Chemistry 2009

IUCr

Maris, T. (2004). *UdMX*. University of Montréal, Canada.
McMurtrie, J. & Dance, I. (2009). *CrystEngComm*, **11**, 1141-1149. ChemPort
Medlycott, E. A., Hanan, G. S., Abedin, T. S. M. & Thompson, L. K. (2008). *Polyhedron*, **27**, 493-501. ChemPort
Medlycott, E. A., Udachin, K. A. & Hanan, G. S. (2007). *Dalton Trans.* pp. 430-438. CCDC CrossRef
Rajeshwar, K., McConnell, R. & Licht, S. (2008). In *Solar Hydrogen Generation Toward a Renewable Energy Future*. New York: Springer.
Sheldrick, G. M. (1996). *SADABS*. Bruker AXS Inc., Madison, Wisconsin, USA.
Sheldrick, G. M. (2008). *Acta Cryst.* A**64**, 112-122. CrossRef details
Uma, V., Vaidyanathan, V. G. & Nair, B. U. (2005). *Bull. Chem. Soc. Jpn*, **78**, 845-850. Web of Science® CCDC
CrossRef ChemPort
Wang, S., Li, B.-D., Wang, R. Y., Wu, B. L. & Zhang, H.-Y. (2009). *Synth. React. Inorg. Met. Org. Nano-Met. Chem.* **39**, 355-359.

Acta Cryst (2011). C**67**, m81-m84 [doi:10.1107/S0108270111004641]

Wiley Online Library

ELSEVIER

…igned a one-pot method to perform the two successive benzyne cycloadditions
…3),[15] giving the bis-cycloadduct **15** in 44 % yield, which is slightly higher than that obtained
…tepwise method (see preceding; 41 % yield in two steps). Bis-cycloadduct **15** was subjected to
…e aromatization ($TiCl_4$, Zn, THF, RT, 1 h)[16] followed by hydrolysis of the silyl acetal (aq. HF,
…rile, 0 °C, 1 h), giving benzocyclobutenone **16** in 78 % yield. At this stage, the structure was
…guously reconfirmed by single-crystal X-ray diffraction analysis.[17] Ketone **16** was then
…ed to oxime **17** (NH_2OH-HCl, pyridine, MeOH, 60 °C, 18 h) in 88 % yield as a single isomer, for

16 M. A. Meador, H. Hart, *J. Org. Chem.* 1989, **54**, 2336–2341.
CrossRef | CAS | Web of Science® Times Cited: 16 | 'ejournals@cambridge - find full text'

17 CCDC 1543805 (16) contains the supplementary crystallographic data for this paper. These data can be obtained free of charge from The Cambridge Crystallographic Data Centre.

18 Nitrile oxide **18** could be stored at −18 °C for at least one month.

Outline

Abstract
Graphical abstract
1. Introduction
2. Results and discussio
3. Conclusions
4. Experimental
Acknowledgements
Supplementary data
Research Data
References and notes

An efficient phosphate sensor: tripodal quinoline excimer transduction

Avijit Pramanik, Gopal Das

Show

https://

Research data for this article

Cambridge Crystallographic Data Center
Crystallographic data
Data associated with the article:
CCDC 689113: Experimental Crystal Structure Determination

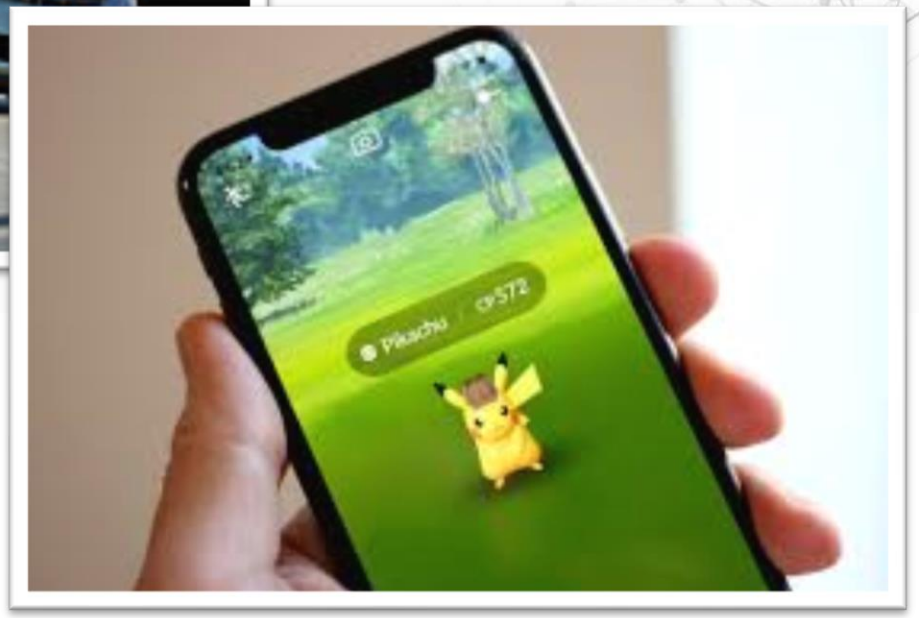
SCHOLIS

CCDC

2010s

INTRODUCING
amazon echo

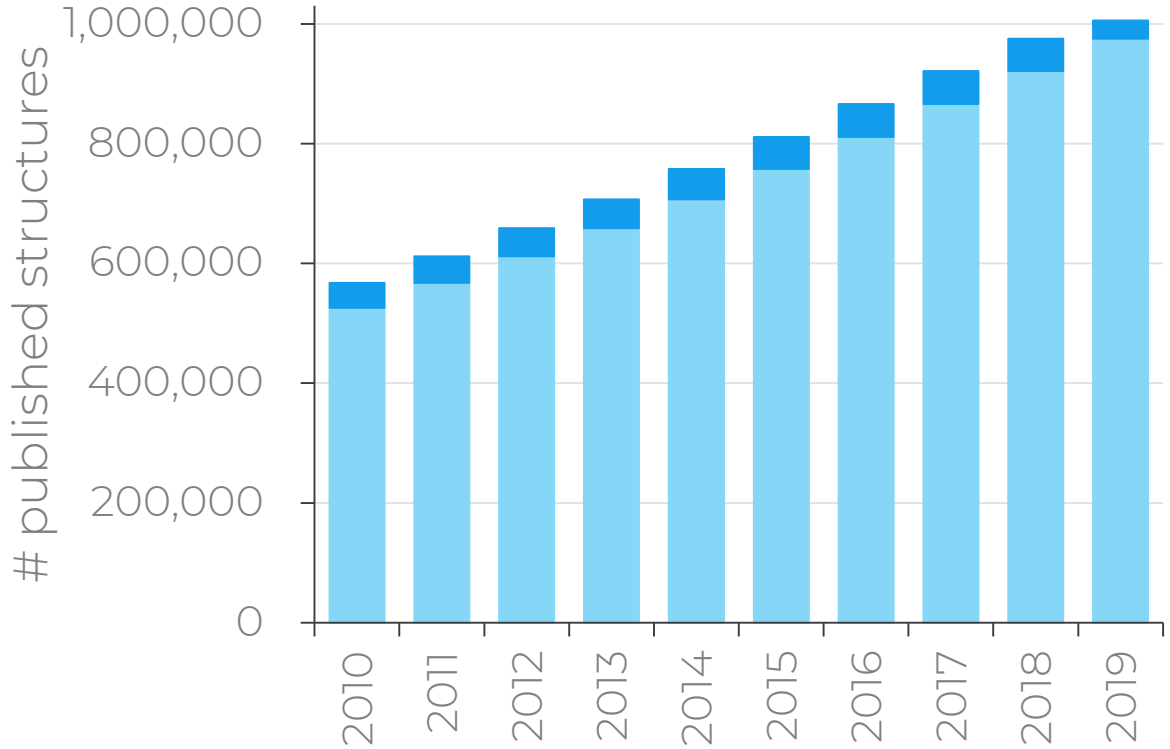
Always ready, connected, and fast. **Just ask.**

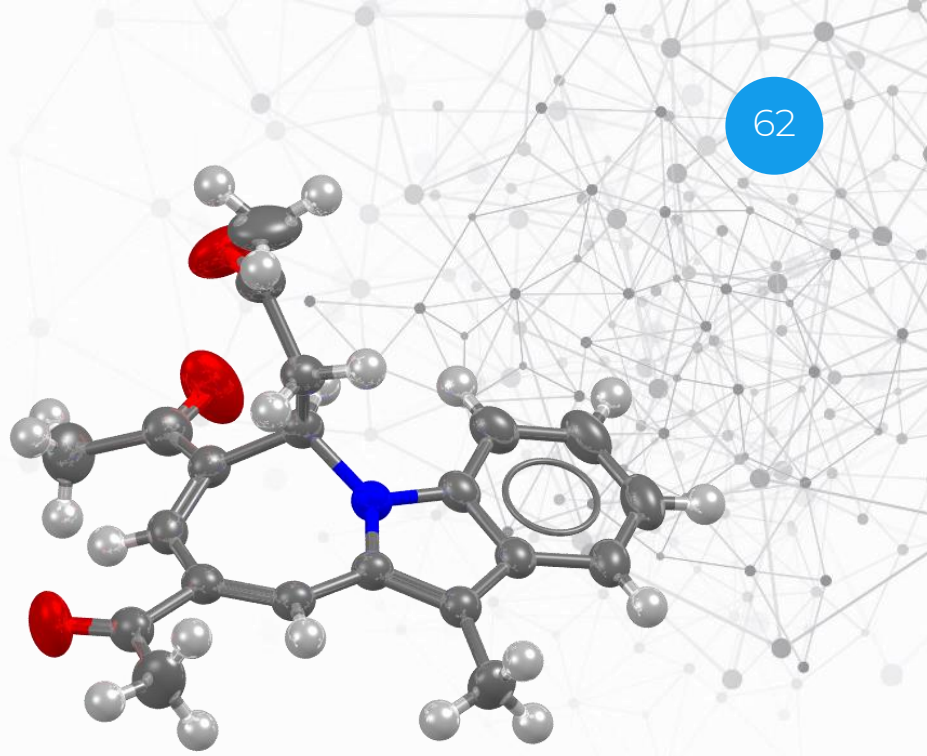


CCDC

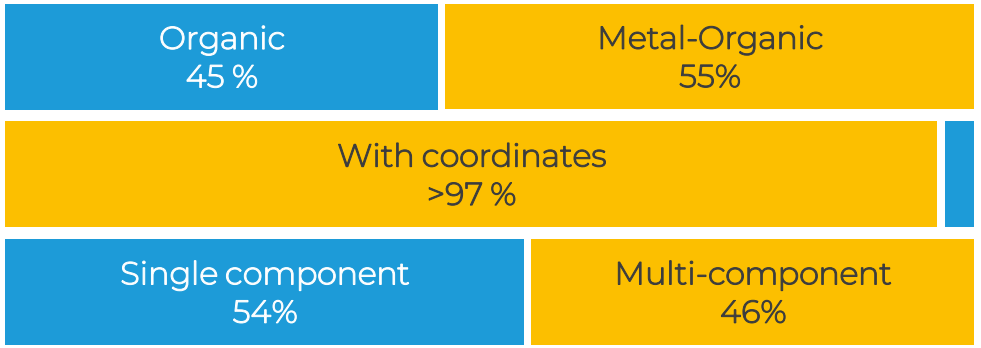
2010-19

- CSD reaches 1 million structures
- On course to double in size again





XOPCAJ
Millionth CSD structure



Wei Wang, Haofu Zhu, ShuYa Liu, Zhiguo Zhao, Liang Zhang, Jingcheng Hao, Yao Wang, *Journal of the American Chemical Society*, 2019, 141, 9175, DOI: 10.1021/jacs.9b03806

CCDC

The creation of a teaching subset

Teaching Three-Dimensional Structural Chemistry Using Crystal Structure Databases. 1. An Interactive Web-Accessible Teaching Subset of the Cambridge Structural Database

Gary M. Battle* and Frank H. Allen
Cambridge Crystallographic Data Centre, 12 Union Road, Cambridge CB2 1EZ, United Kingdom
*battle@ccdc.cam.ac.uk

Gregory M. Ferrence
Department of Chemistry, Illinois State University, Normal, Illinois 61790-4160

Knowledge of the three-dimensional (3D) nature of chemical compounds is fundamental to the education of every chemist. Without this knowledge, concepts such as conformation, stereochemistry, chirality, and the geometrical shapes of metal coordination spheres cannot be properly assimilated or understood. Studies have shown that 3D visualizations enhance students' conceptual understanding and spatial abilities (*1*−*3*). Yet, even at a time when high-resolution interactive 3D graphics are available via every home PC, chemical structures are still often taught using quasi-2D representations. These representations do not convey the levels of understanding, even wonder, opened up to students by the visualization and manipulation of "real" 3D images on their own computer screens.

Additionally, the use of experimentally measured data is of great pedagogical value and has been shown to enhance student interest (*4*−*6*). The experimental error and statistical variation inherent in experimentally measured structural data provide students with the opportunity to deal with the uncertainties of chemistry. Commonly undergraduate researchers struggle with the reality that the chemistry they perform in the research laboratory seems to violate their conceptions of chemistry because classroom examples have instilled a bias that chemical structure should be "perfect", that is, idealized.

For crystalline materials, modern X-ray analysis is now the method of choice for the determination of 3D structure across the complete chemical spectrum, from metals and alloys, through ionic and molecular species to proteins and viruses. Crystal structure information can now be obtained in a matter of hours and days and at the small-molecule level is very precise: standard uncertainties of <0.005 Å on bond lengths and <0.5° on valence and torsion angles are the norm. More than 700,000 crystal structures have been published over the past 60 years, and modern instrumentation is now adding in excess of 60,000 novel structures each year. The good news is that literature references, chemical information, and, most importantly, 3D atomic coordinates for all of these structures are preserved, and continue to be preserved, in five major crystallographic databases (*7*). The bad news is that these major resources are, as yet, little used in teaching chemistry at the undergraduate and high school levels.

In this short series of articles, we describe the Cambridge Structural Database of over 500,000 small-molecule crystal structures, and how it can be accessed and used to enhance and extend the chemistry learning experience of a wide range of students. Part 1 introduces the Cambridge Structural Database, indicates how it can be accessed and searched, and describes a teaching subset of some 500 compounds that are available to educators and students via a simple Web interface at no cost. Part 2 (*9*) describes a series of teaching examples based on the tools described in this article.

The Cambridge Structural Database

Compilation of the Cambridge Structural Database (CSD) began in 1965, when the number of small organic and metal−organic crystal structures available in the literature was just a few thousand. Since then, advances in crystallographic theory, diffraction equipment, and computer technology have combined to expand productivity dramatically so that in 2009 the CSD recorded its 500,000th structure. For each published crystal structure determined using X-ray or neutron diffraction, the CSD stores the following:

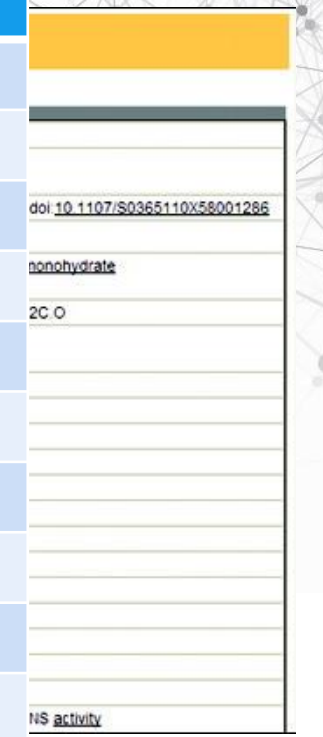
- Primary numerical results: crystallographic unit cell and space group, and the fractional atomic coordinates with respect to that cell.
- Chemical structure information: an encoding of the 2D chemical diagram as a connectivity table specifying atomic "nodes" and bonded "edges".
- Bibliographic information: author names and journal reference.
- Other text and numerical data: for example, compound name(s), molecular formula, precision indicators, density, and (where available in the crystal structure publication) the melting point, color, crystal habit, and so forth.

Each of the half million structural entries in the CSD is identified by a reference code (refcode) that consists of six letters to identify each unique compound, and two digits that distinguish, for example, studies at a different temperatures or pressures, different polymorphic forms, or studies by different authors, and so forth.

The CSD covers more than 1,200 literature sources, and some additional structures are privately deposited into the database by scientists who do not wish to publish their results in a journal. All 3D coordinate sets in the CSD are received in electronic form, and the raw data are enhanced by CSD editors

© 2010 American Chemical Society and Division of Chemical Education, Inc. • pubs.acs.org/jchemed • Vol. 87 No. 8 August 2010 • Journal of Chemical Education
10.1021/ed100256k Published on Web 06/09/2010

structure type	count	fraction (%)
all structures	500	100.0
organic	331	66.2
metal–organic	161	32.2
organic structures		
carbohydrates	8	1.6
nucleosides/nucleotides	6	1.2
amino acids and peptides	29	5.8
porphyrins/corrins	13	2.6
steroids	12	2.4
alkaloids	10	2.0
organic polymers	8	1.6
“drug” or “activity””	42	8.4
“polymorph” or “form”	80	16.0

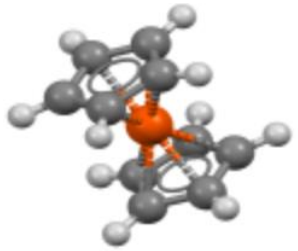


Teaching 3D structural chemistry using crystal structure databases: 1. An interactive web-accessible teaching subset of the Cambridge Structural Database . G. M. Battle, F. H. Allen, G. M. Ferrence, *J. Chem. Ed.*, **87**, 809-812, 2010 10.1021/ed100256k

CCDC

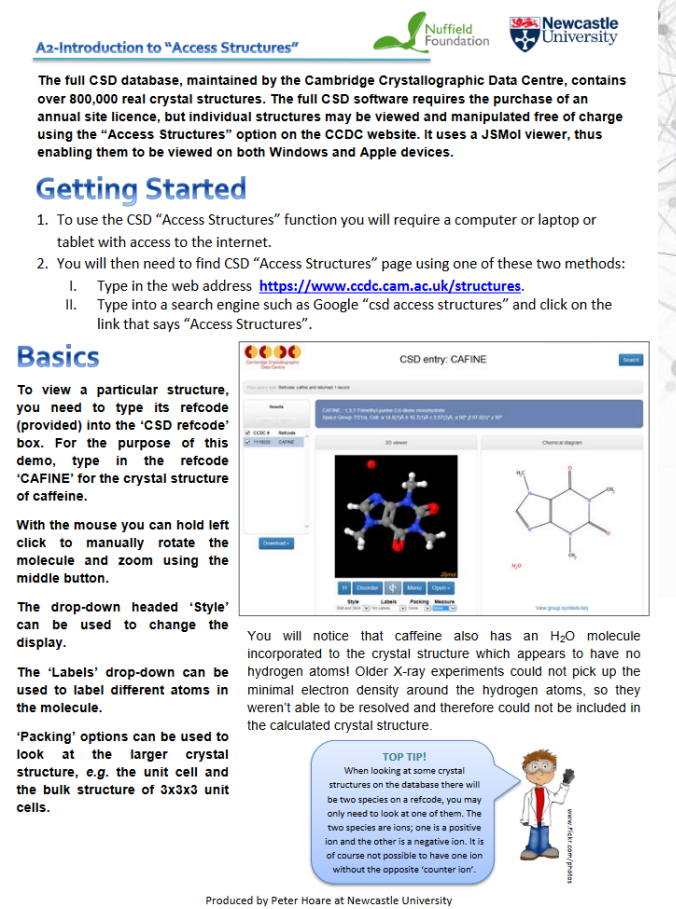
The CSD teaching subset today

- 750+ Structures for educational purposes
  - Drug molecules
  - Fundamental Chemistry
  - Symmetry
  - Metal-Organic Frameworks



Information on the Teaching Subset

	A	F	G	H	I
	NAME	Point Group	Teaching Sheet	Compound Name	Common Name
	ACABRH02	C_{2v}	Metal Coordination	Acetylacetonato-dicarbonyl-rhodium(i)	
	ACAJIX	C_3	Transition Metals (oct	tris(2-(4-fluoromethylphenylene)-5-trifluoromethylpyridine)-iridium	
	ACALDA	C_{3v}		Acetaldehyde-ammonia trihydrate	2,4,6-Trimethyl-hexahydro-1,3,5-triazine trihydrate
	ACANIL01			Acetanilide	
	ACAQUR		Metal Coordination	bis(2-(2-Methoxyphenylamido)-4-(2-methoxyphenylimino)pent-2-ene)-calcium	
	ACARBM01		VSEPR Shapes (tetrah	Ammonium carbamate	
	ACASED		Metal Coordination	bis(Hydrogen tris(3,5-dimethylpyrazolyl)borato)-(bis(trimethylsilyl)amido)-uranium(iii) tetrahydrofuran solvate	
	ACAZEK		Metal Coordination	catena-[bis(mu\2$-Bromo)-(mu\2$-ethylene-1,2-diamine)-di-silver(i)]	
	ACCAAH	C_i		Acetylenedicarboxylic acid dihydrate	
	ACCTHP			3-Carboxy-2-acetyl-thiophene	
	ACENYL01		Geometric Isomers (cl	Acenaphthylene	
	ACEPOO	C_{2v}	Functional Groups (ar	5-Bromo-1,3-dichloro-2-iodobenzene	
	ACEQII		Transition Metals (oth	bis(1,2-bis(Dimethoxy)ethane)-di-iodo-tetrahydrofuran-thulium	
	ACETAC07		Molecules of Interest	Ethanoic acid	Acetic acid; DrugBank: DB03166
	ACETPH		Functional Groups (ke	Acetophenone	PDB Chemical Component code: AC0



A2-Introduction to "Access Structures"

Nuffield Foundation — Newcastle University

The full CSD database, maintained by the Cambridge Crystallographic Data Centre, contains over 800,000 real crystal structures. The full CSD software requires the purchase of an annual site licence, but individual structures may be viewed and manipulated free of charge using the "Access Structures" option on the CCDC website. It uses a JSMol viewer, thus enabling them to be viewed on both Windows and Apple devices.

Getting Started

1. To use the CSD "Access Structures" function you will require a computer or laptop or tablet with access to the internet.
2. You will then need to find CSD "Access Structures" page using one of these two methods:
   I. Type in the web address https://www.ccdc.cam.ac.uk/structures.
   II. Type into a search engine such as Google "csd access structures" and click on the link that says "Access Structures".

Basics

To view a particular structure, you need to type its refcode (provided) into the 'CSD refcode' box. For the purpose of this demo, type in the refcode 'CAFINE' for the crystal structure of caffeine.

With the mouse you can hold left click to manually rotate the molecule and zoom using the middle button.

The drop-down headed 'Style' can be used to change the display.

The 'Labels' drop-down can be used to label different atoms in the molecule.

'Packing' options can be used to look at the larger crystal structure, *e.g.* the unit cell and the bulk structure of 3x3x3 unit cells.

CSD entry: CAFINE

You will notice that caffeine also has an H_2O molecule incorporated to the crystal structure which appears to have no hydrogen atoms! Older X-ray experiments could not pick up the minimal electron density around the hydrogen atoms, so they weren't able to be resolved and therefore could not be included in the calculated crystal structure.

TOP TIP!
When looking at some crystal structures on the database there will be two species on a refcode, you may only need to look at one of them. The two species are ions; one is a positive ion and the other is a negative ion. It is of course not possible to have one ion without the opposite 'counter ion'.

Produced by Peter Hoare at Newcastle University

https://www.ccdc.cam.ac.uk/Community/educationalresources

CCDC

The CSD on the web today

CCDC FIZ Karlsruhe Leibniz Institute for Information Infrastructure

CSD Entry: KITTEN

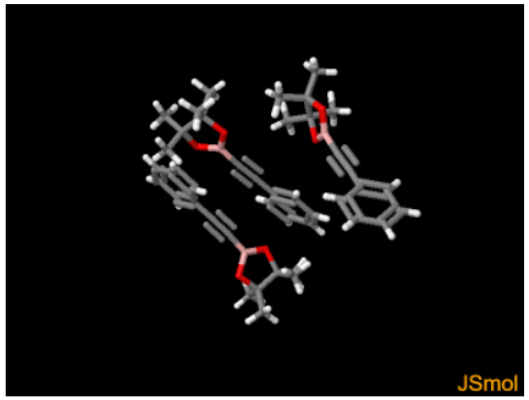
Sign In
Licensed to: CCDC Main Site

Simple Search | Structure Search | Unit Cell Search | Formula Search

Your query was: Identifier(s): KITTEN and the search returned 1 record. Modify Search New Search

KITTEN : 4,4,5,5-Tetramethyl-2-(phenylethynyl)-1,3,2-dioxaborolane
Space Group: P $2_1/c$ (14), **Cell:** *a* 19.9704(9)Å *b* 16.0870(7)Å *c* 13.1750(6)Å, *α* 90° *β* 107.375(2)° *γ* 90°

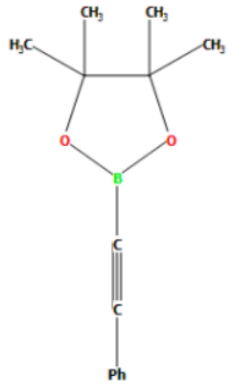
3D viewer



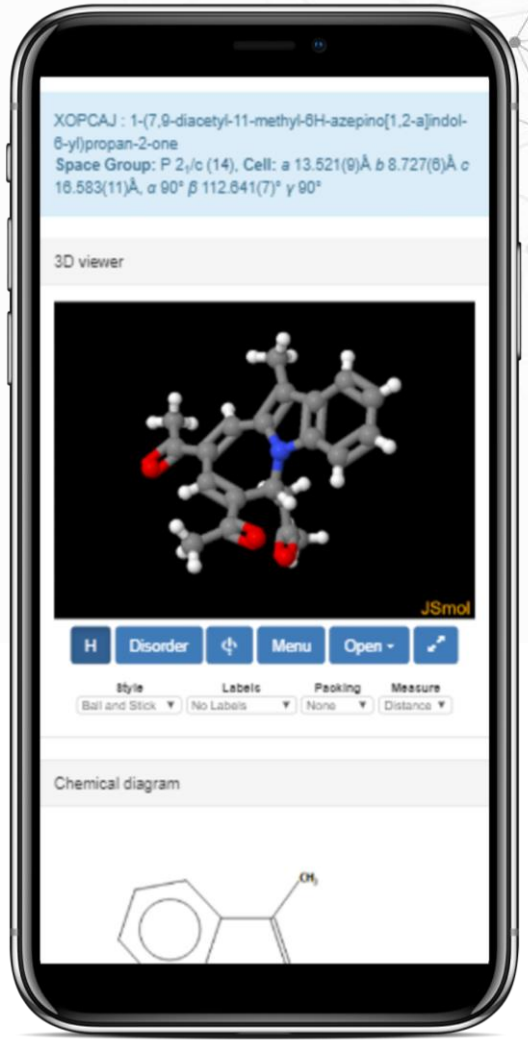
H | Disorder | Menu | Open

Style: Capped Sticks | Labels: No Labels | Packing: None | Measure: None

Chemical diagram



View group symbols key



CCDC

FAIR data and interoperability

InChI for Chemical Structures

DOIs for Digital Objects

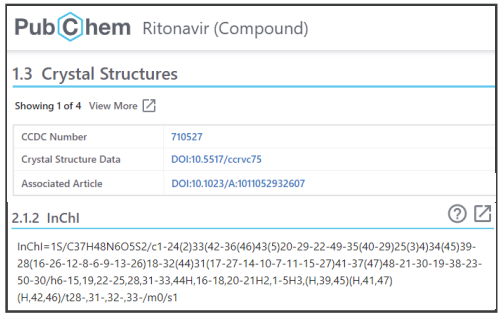
ORCID iDs for Researchers

ROYAL SOCIETY OF CHEMISTRY · ACS · WILEY · ELSEVIER · Springer · THOMSON REUTERS · CCDC · ChemSpider · PubChem · WORLDWIDE PDB PROTEIN DATA BANK

DATA SHOULD BE

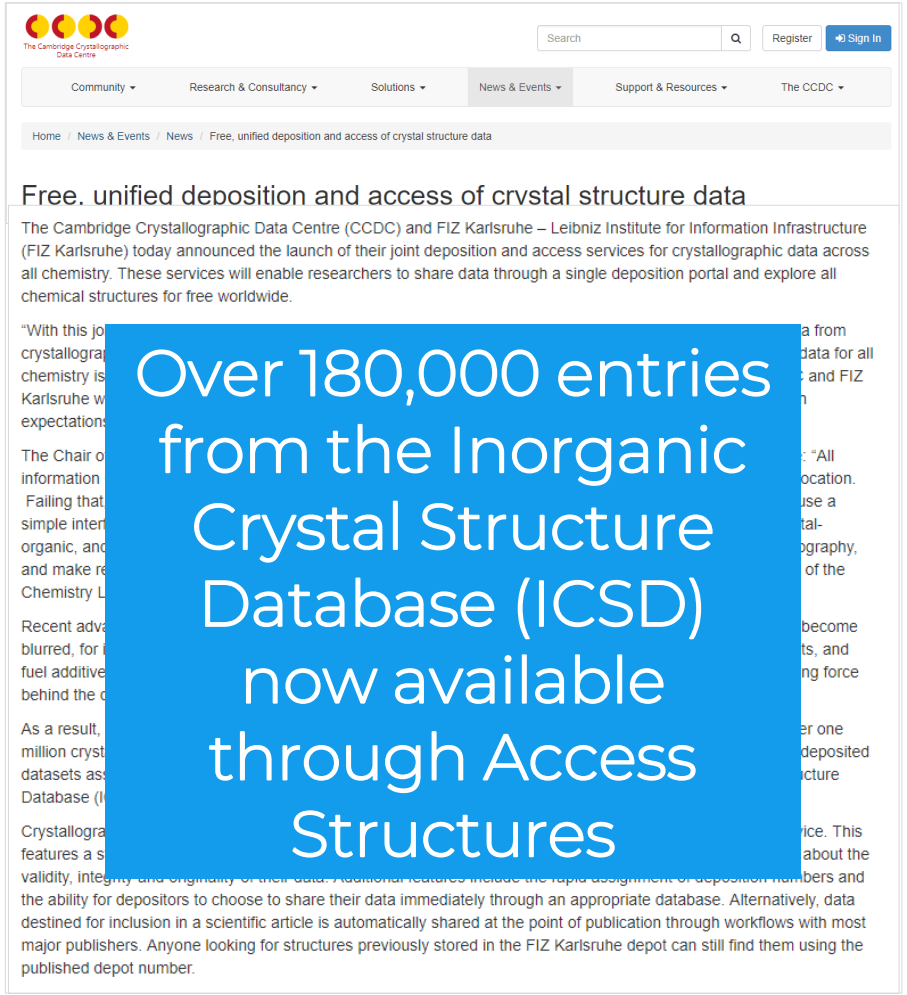
Findable · **Interoperable** · **Accessible** · **Reusable**

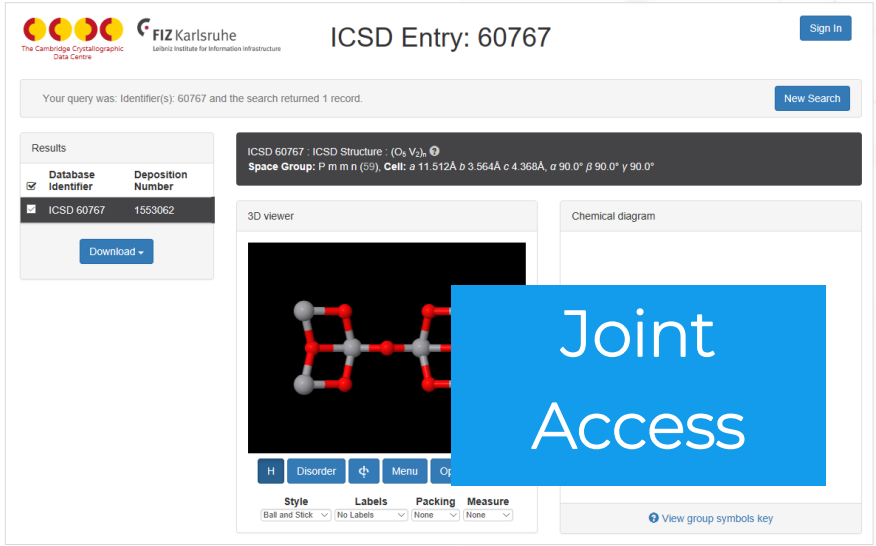
BY HUMANS AND MACHINES

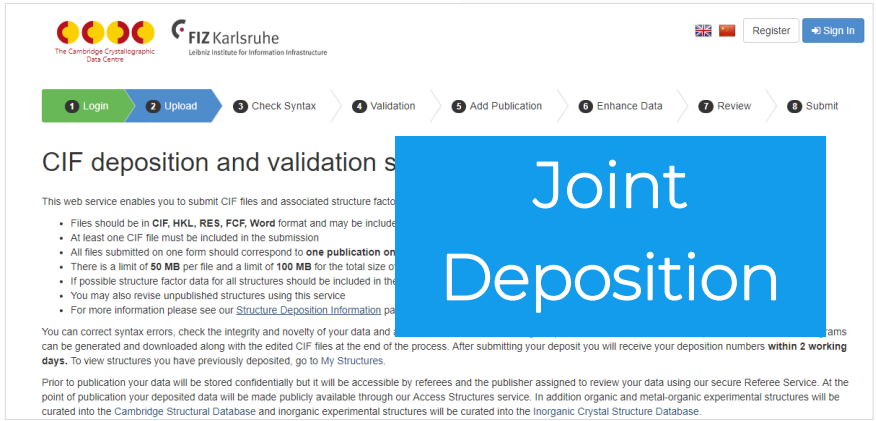


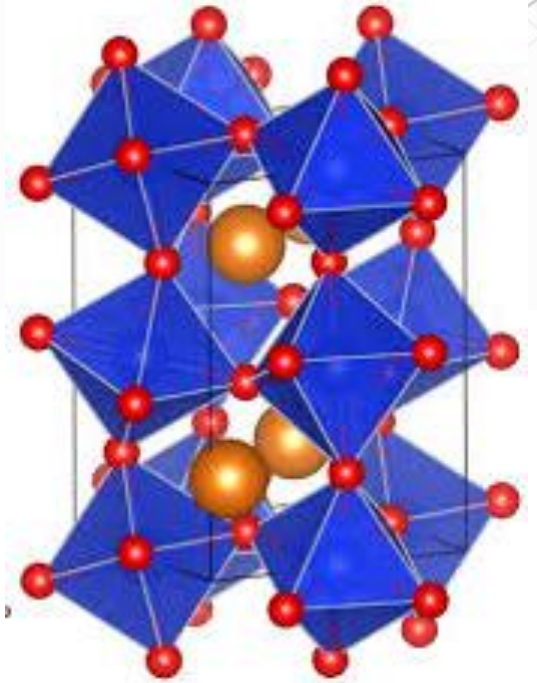
Wilkinson, M. D. et al. The FAIR Guiding Principles for scientific data management and stewardship. Sci. Data 3:160018 DOI: 10.1038/sdata.2016.18 (2016)

Joint CSD and ICSD services









CCDC

2010 – The launch of online deposition

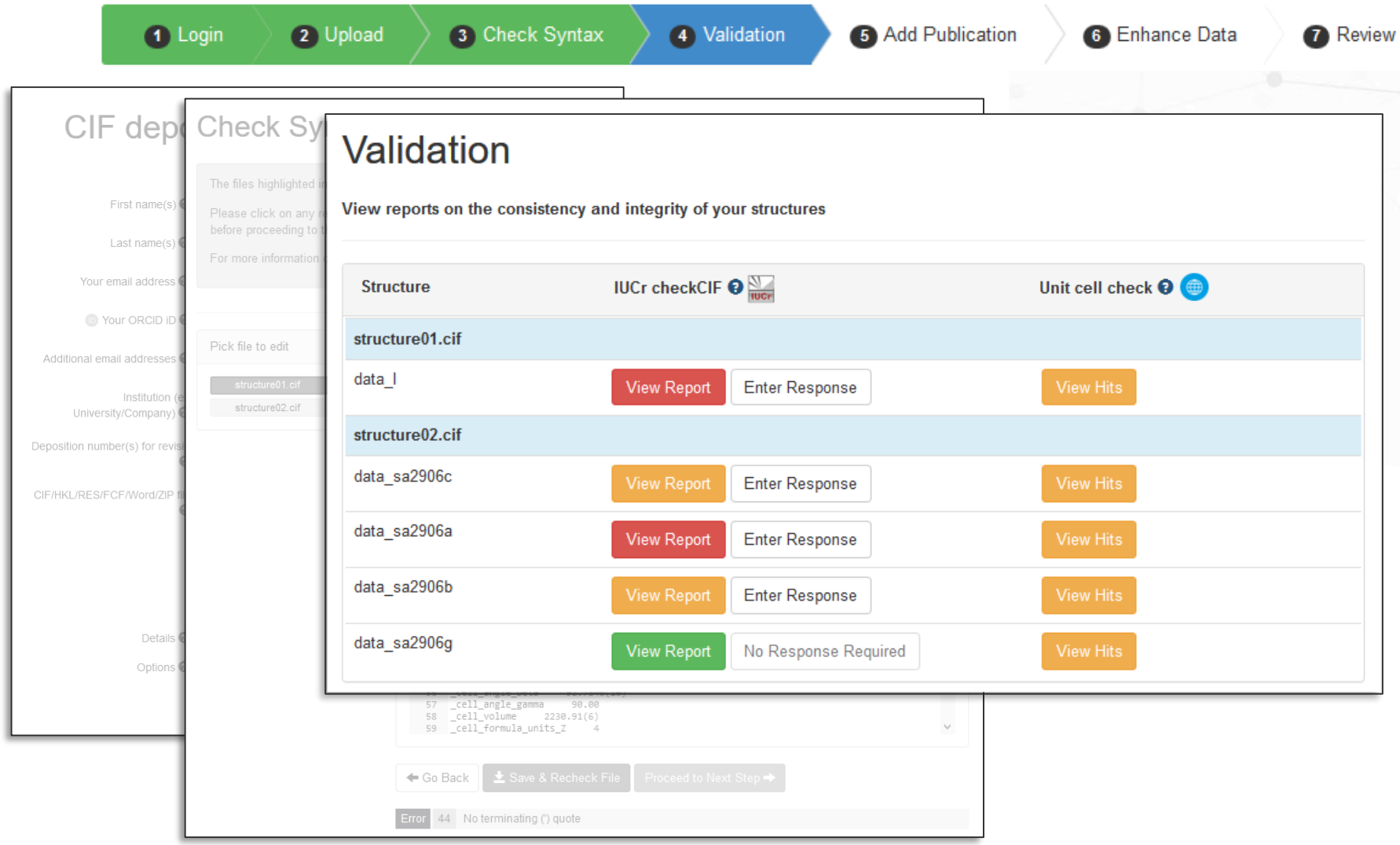
Depositing CIF Files Through the Web

The CCDC is proud to announce a new means of depositing CIF structure files. This web-based system has been introduced to provide in-situ help to users depositing new data. Depositors can now upload and verify their CIFs using built-in technology based on CCDC's free CIF editor, EnCIFer. In addition to syntactic checking, the structure-deposit interface allows users to add any important additional information about their structure, in a convenient and easy-to-use form. This product also currently offers Mandarin speakers the opportunity to deposit structures in their mother-tongue. Links to the new web page can be found via. the CCDC web-site or by visiting http://www.ccdc.cam.ac.uk/services/structure_deposit

Dr Chris Harding, Scientific Software Engineer

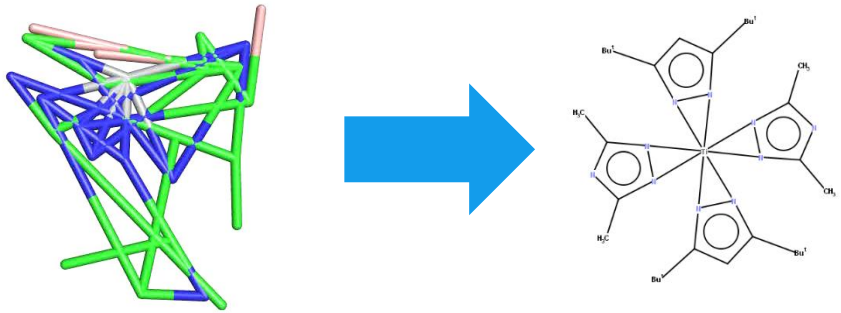
CCDC

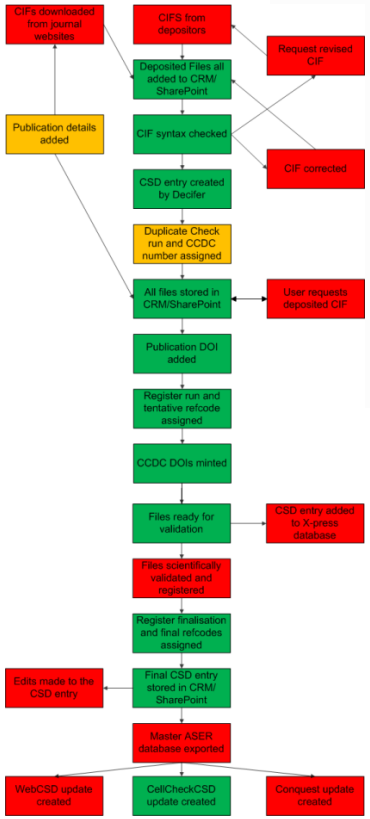
Online deposition today

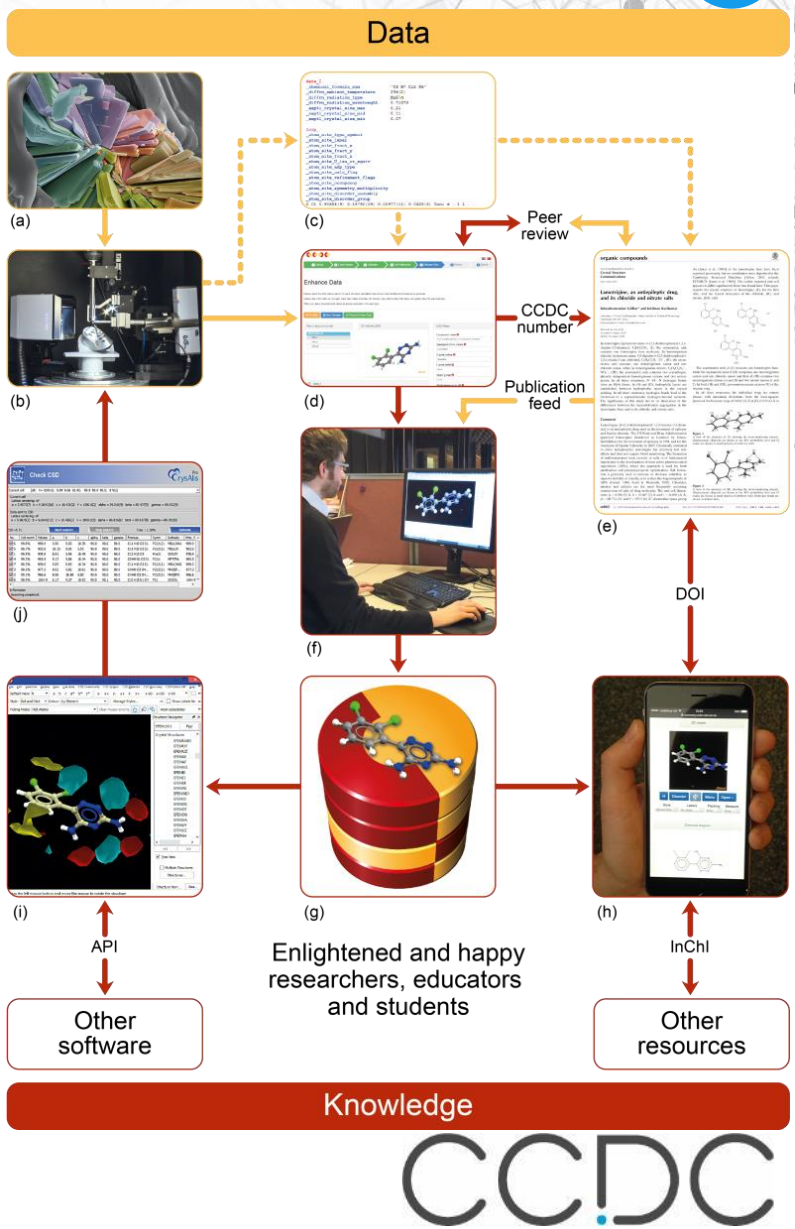


Curating the CSD today

- Each dataset expertly validated and curated
- Datasets enhanced
  - Chemical connectivity
  - Compound names
  - 2D chemical diagrams
  - Additional experimental data
  - Bibliographic information







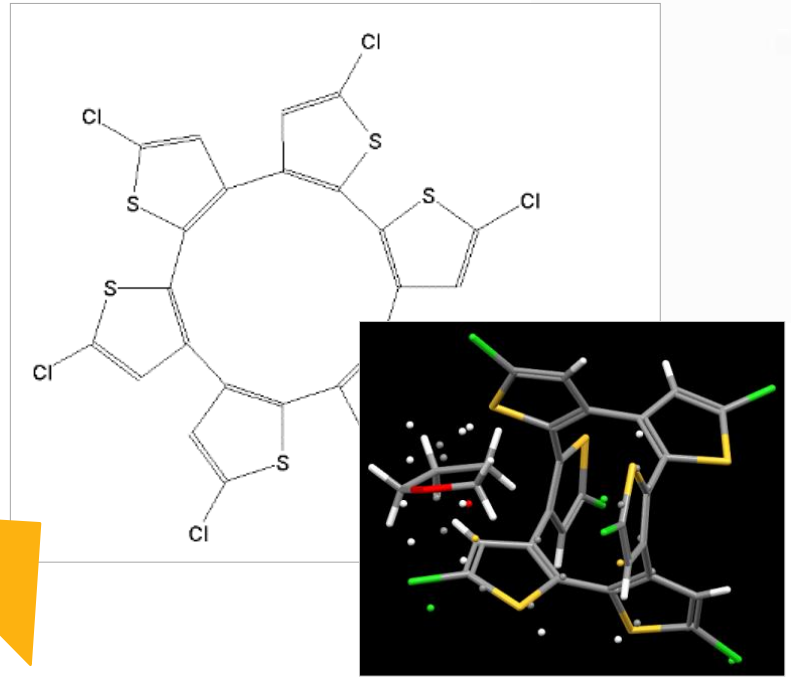
CCDC

Using the collection to curate new structures

```
loop_
_atom_site_label
_atom_site_type_symbol
_atom_site_fract_x
_atom_site_fract_y
_atom_site_fract_z
_atom_site_U_iso_or_equiv
_atom_site_adp_type
_atom_site_occupancy
_atom_site_symmetry_multiplicity
_atom_site_calc_flag
_atom_site_refinement_flags
_atom_site_disorder_assembly
_atom_site_disorder_group
Cl1 Cl 0.5993(2) 1.0007(7) 0.8131(17) 0.044(3) Uani 0.50 1 d PDU A 1
S1 S 0.5321(3) 0.8260(6) 0.9322(3) 0.0327(11) Uani 0.50 1 d PDU A 1
C2 C 0.5529(4) 0.8802(9) 0.8184(9) 0.029(4) Uani 0.50 1 d PDU A 1
C3 C 0.5286(7) 0.8174(18) 0.7440(7) 0.031(4) Uani 0.50 1 d PDU A 1
H3A H 0.5350 0.8343 0.6771 0.037 Uiso 0.50 1 calc PR A 1
C4 C 0.4918(8) 0.7220(19) 0.7783(8) 0.027(4) Uani 0.50 1 d PDU A 1
C5 C 0.4900(6) 0.7171(14) 0.8779(9) 0.029(4) Uani 0.50 1 d PDU A 1
Cl2 Cl 0.3202(2) 0.4982(6) 1.0830(5) 0.0586(15) Uani 0.50 1 d PDU A 1
S2 S 0.38755(19) 0.6658(5) 0.9578(5) 0.0400(10) Uani 0.50 1 d PDU A 1
```

Chemistry assignment makes data findable, interoperable and reusable

An automated probabilistic approach using data in the CSD



$$P(A|B) = \frac{P(B|A)\,P(A)}{P(B)}.$$

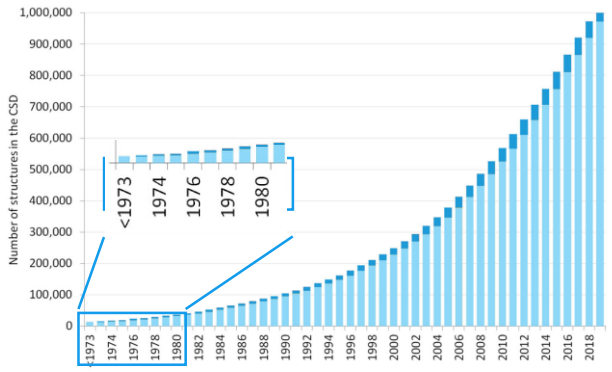
I.J.Bruno, G.P.Shields, R.Taylor, *Acta Cryst.* (2011). B67, 333-349 DOI: 10.1107/S0108768111024608

CCDC

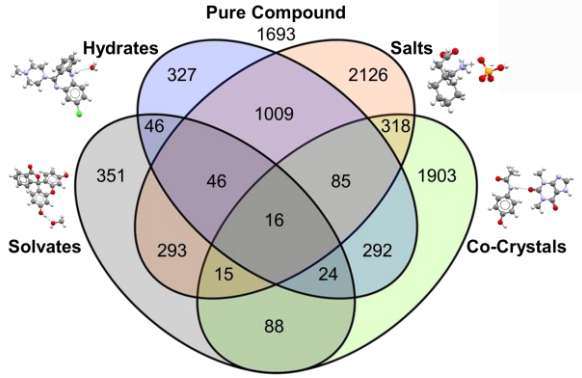
Revisiting CSD entries

Targeted improvements allow improved integrity, consistency, discoverability and value of data

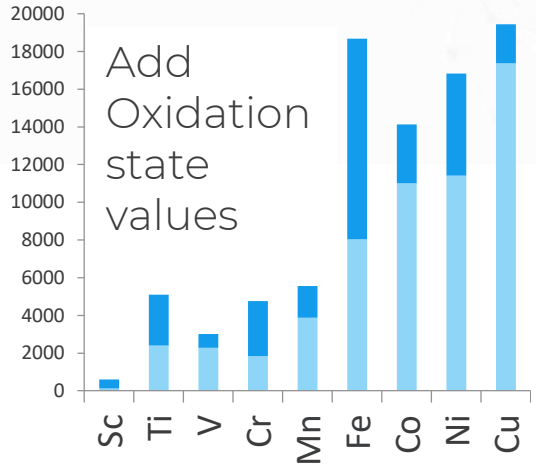
Ensure standardisation of early CSD entries



Creation and maintenance of subsets



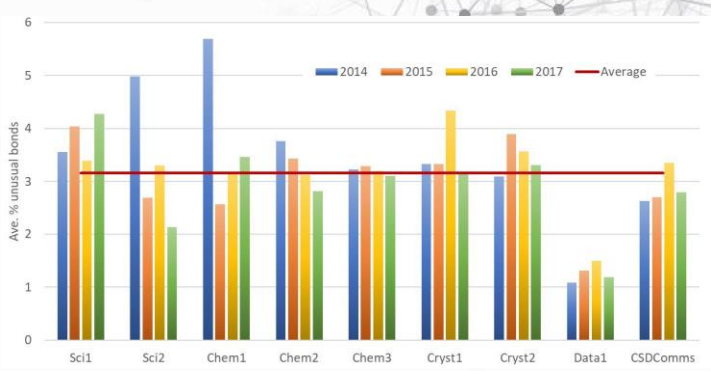
Enrichment of data

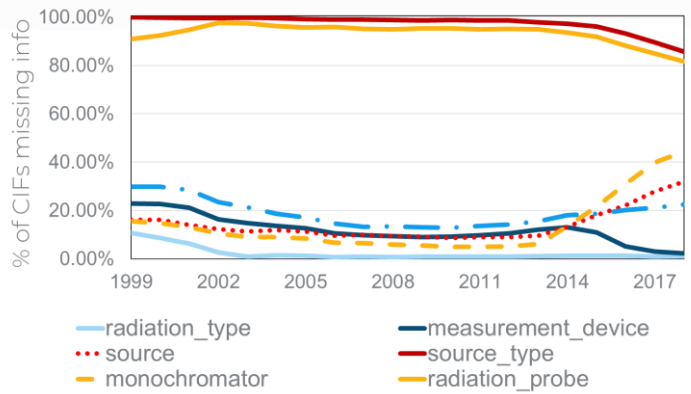


CCDC

Maintaining data integrity in the CSD

- **Integrity** – Completeness, consistency and trustworthiness
- **Data completeness** – trends in reporting of metadata
  - Interactive CSD Deposit checks
  - New filters to select fit for purpose data
- **Consistency** – looking at experimental metadata to identify trends in information supplied
- **Trustworthiness** – Establishing automatic identification of potential cases of misconduct – including fraudulent and plagiarised data



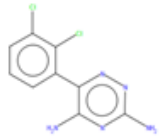


Research integrity is much more than misconduct. *Nature*, 2019, 570, 5-5. DOI:10.1038/d41586-019-01727-0

CCDC

Additional validation for new structures?

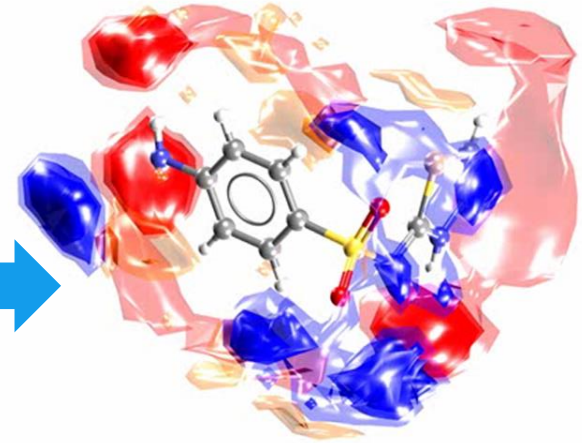
CCDC 1234567

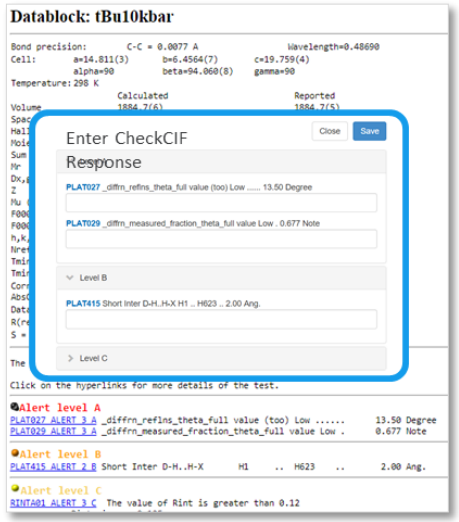


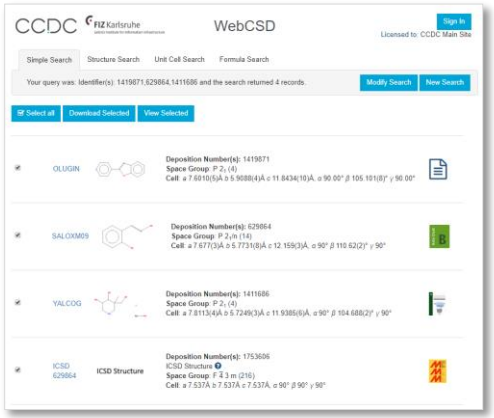
0 syntax issues
1 crystallographic issue
2 chemical issues
0 space group issues

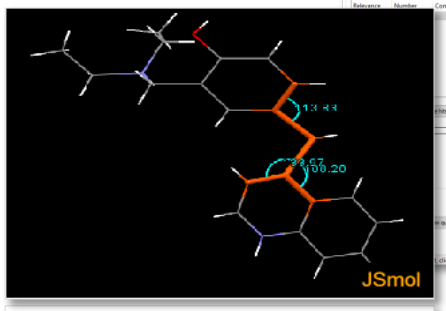
Level A	Most likely a serious problem - resolve or explain
Level B	A potentially serious problem, consider carefully
Level C	Check. Ensure it is not caused by an omission or oversight
Level G	General information/check it is not something unexpected

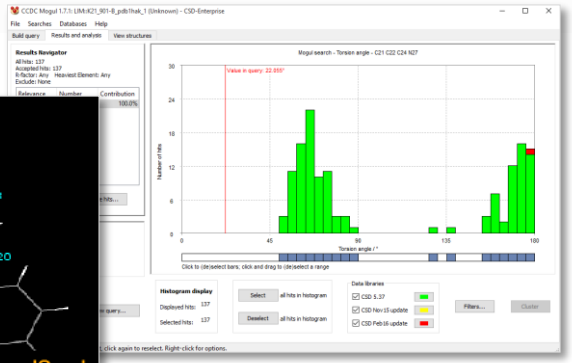
Structure	Summary	checkCIF	Duplicates	Geometry check	Interaction check
Select all / Download Selected / View Selected					
1234567	View Report	View Report	0 duplicates	View Report	View Report
1234568	View Report	View Report	View 1	View Report	View Report
1234569	View Report	View Report	View 2	View Report	View Report









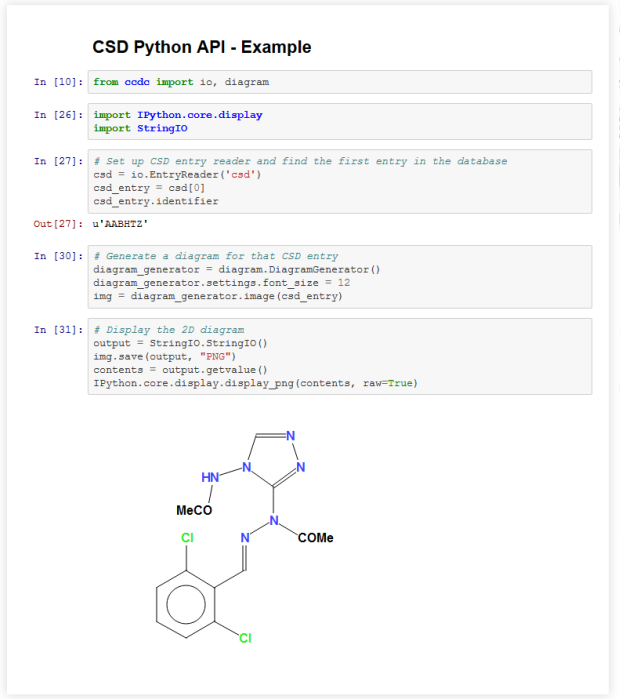


Fragment	Labels	Value	Mean	Z-score	Classification
angle	C1 C14 N43	108.195	119.352	13.008	Unusual (enough hits)
angle	C15 N43 C14	105.5	126.772	9.135	Unusual (enough hits)
angle	C12 C14 N43	133.97	122.923	3.579	Unusual (enough hits)
angle	C22 C15 N43	113.828	120.673	2.136	Unusual (enough hits)

CCDC

Programmatic access and new insights

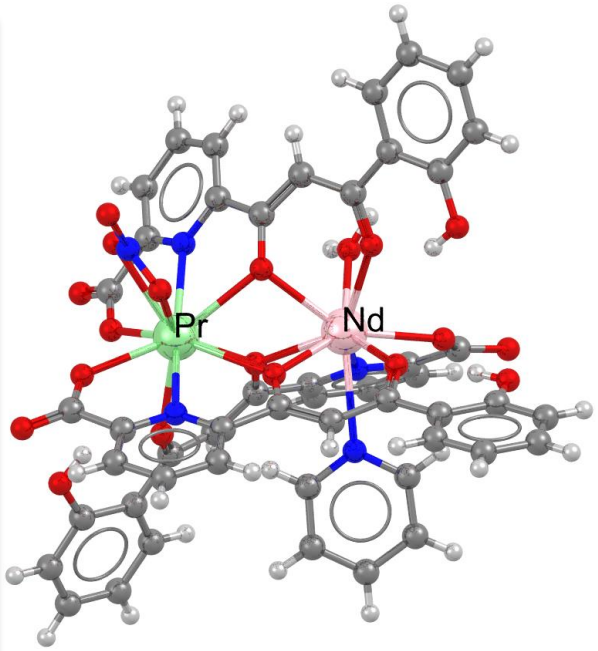
- Launch of the **CSD Python API** enabling
  - Programmatic access to CSD data and software
  - Integration
  - Generation of new insights
- Functions include:
  - Full search capabilities
  - Geometry analysis
  - Interaction analysis
  - Descriptor calculation
  - 2D diagram generation



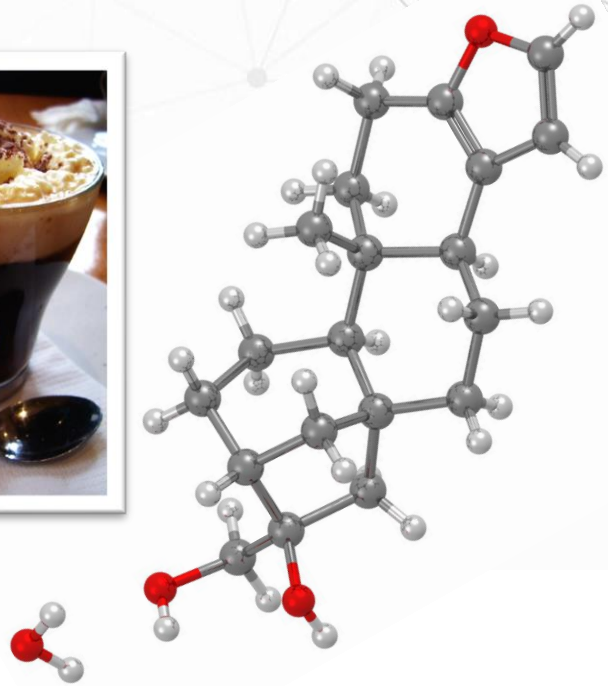
python™

CCDC

The CSD today



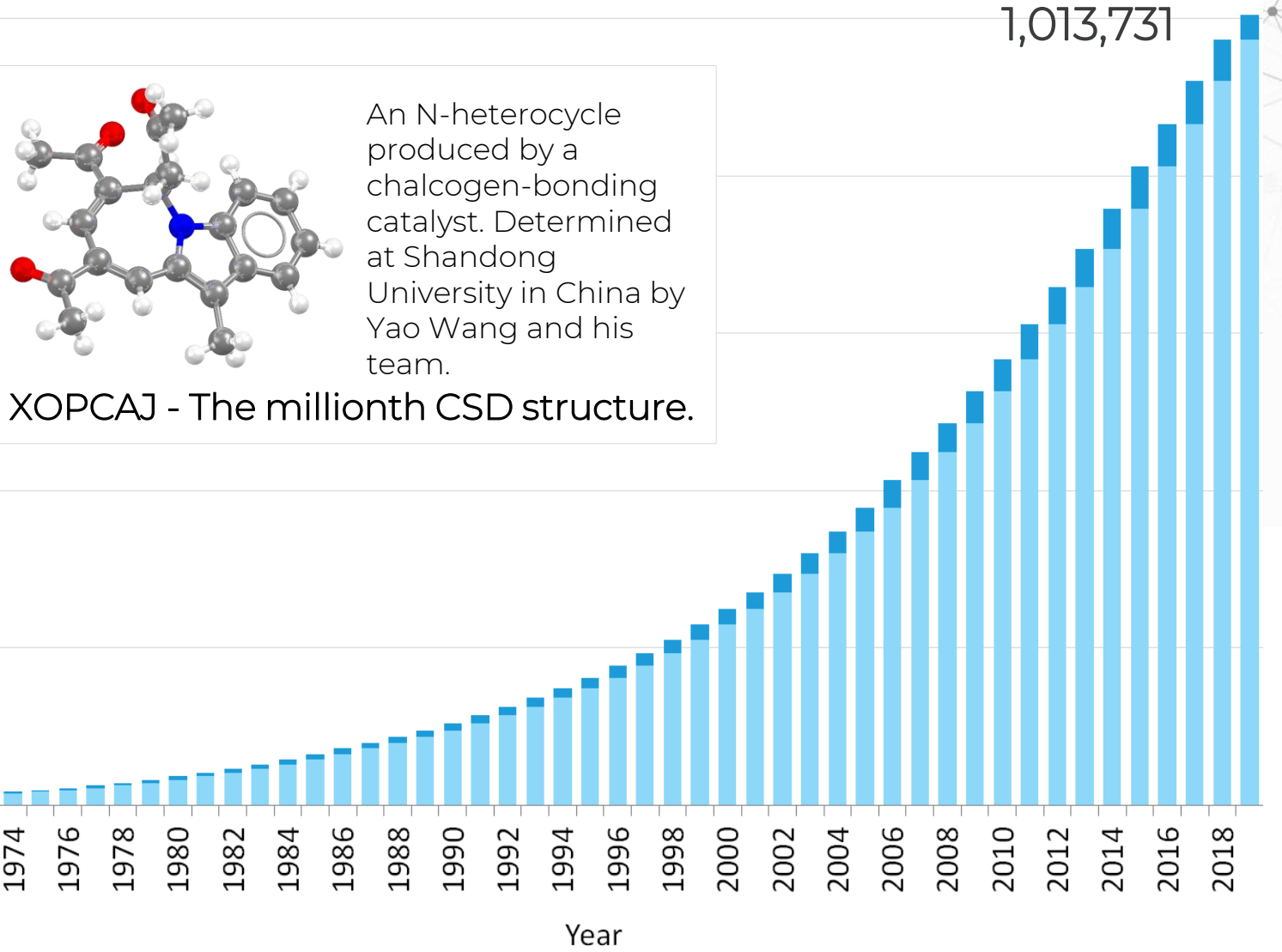
CIJYUS –The famous Viennese chemist Carl Auer von Welsbach discovered Pr and Nd



OPISOX – Cafestol extracted from coffee

CCDC

The CSD today

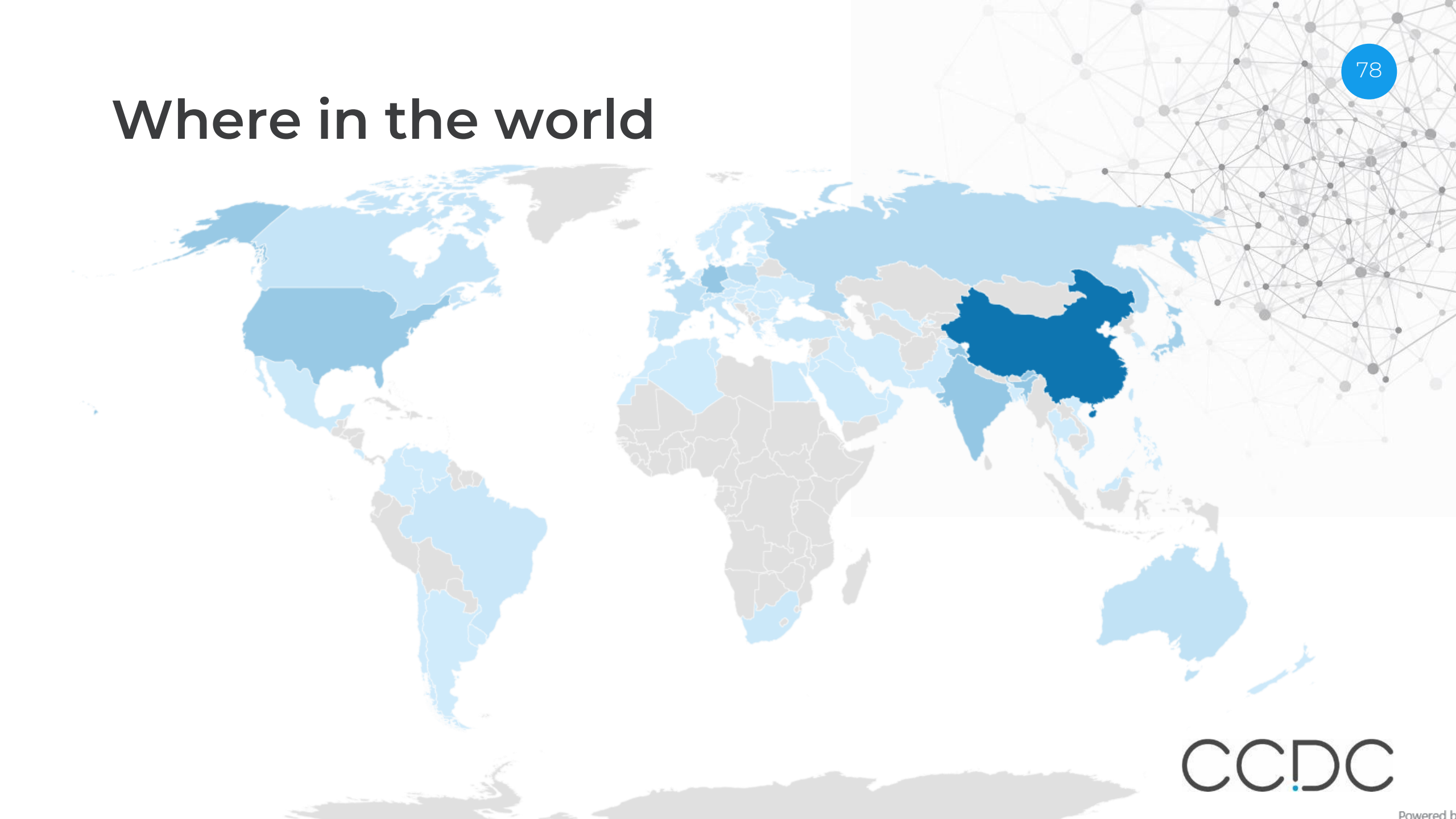


An N-heterocycle produced by a chalcogen-bonding catalyst. Determined at Shandong University in China by Yao Wang and his team.

XOPCAJ - The millionth CSD structure.

CCDC

Where in the world



CCDC

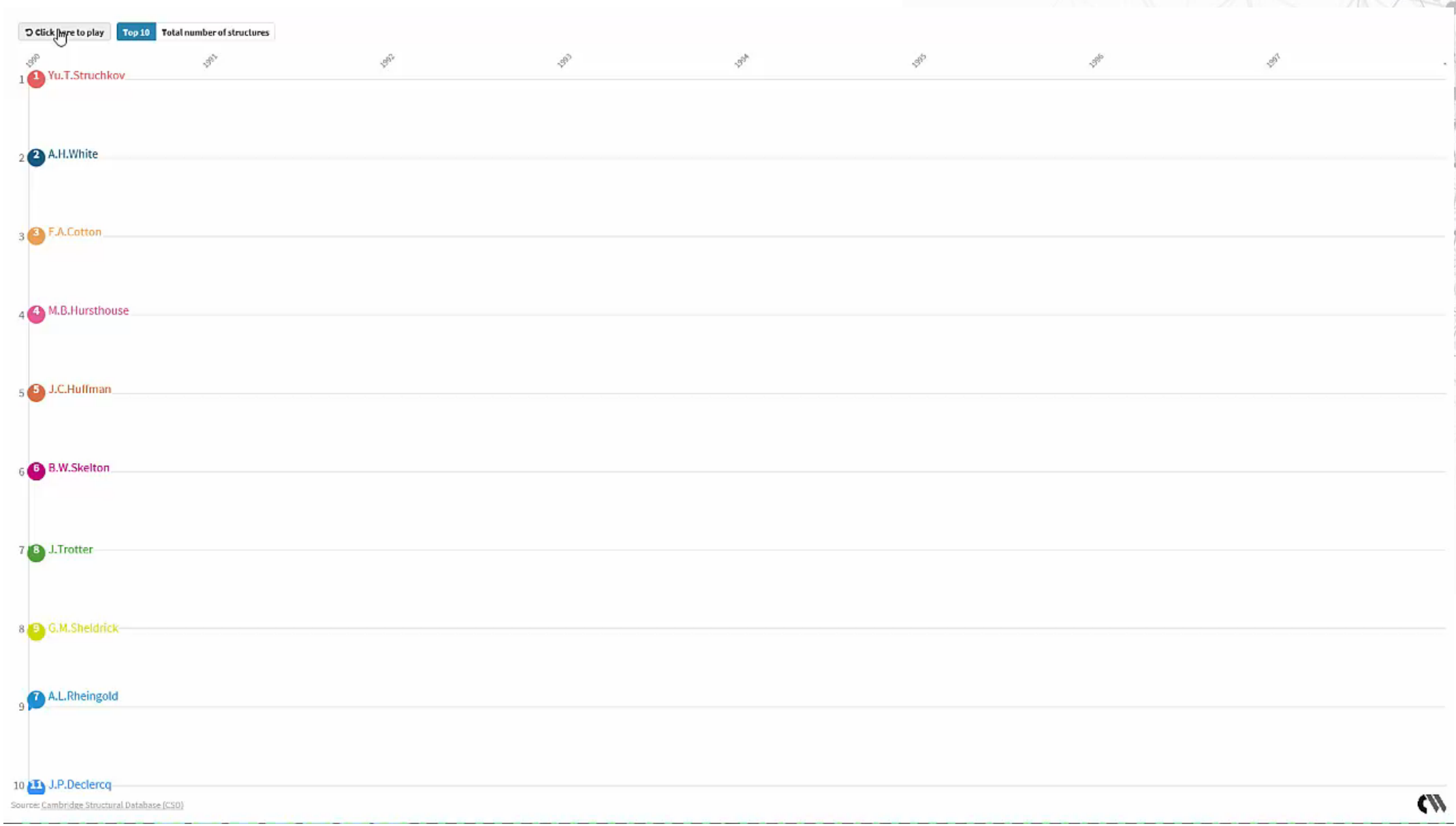
Powered b

The value of the structures

- Structures from 100,000s of crystallographers worldwide
- Manual and automated curation at CCDC
- How many person years of effort?
  - Well over 1,100 crystallographer years
  - Well over 400 CCDC years
  - Well over 1,500 years in total!

400,000 authors in the CSD >750 with over 500 structures

CCDC


Click here to play
Top 10
Total number of structures
1990
1991
1992
1993
1994
1995
1996
1997
1 Yu.T.Struchkov
2 A.H.White
3 F.A.Cotton
4 M.B.Hursthouse
5 J.C.Huffman
6 B.W.Skelton
7 J.Trotter
8 G.M.Sheldrick
9 A.L.Rheingold
10 J.P.Declercq
Source: Cambridge Structural Database (CSD)

CCDC

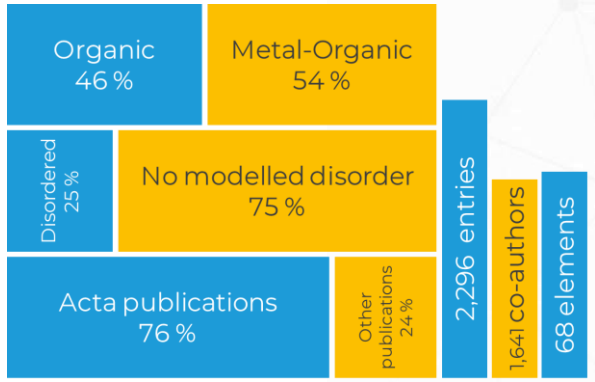
Prolific contributors

Leaderboard

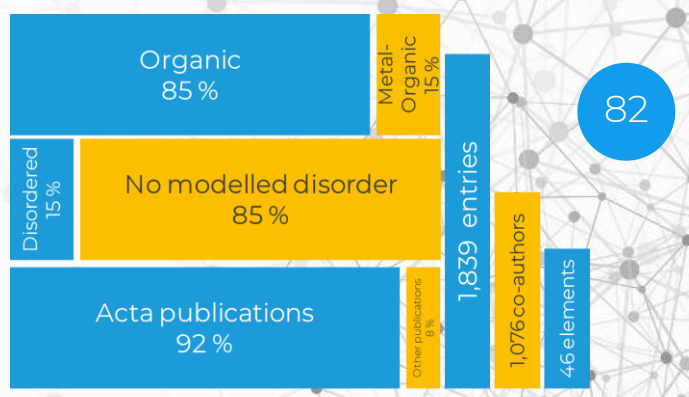
- Brian Skelton
- Allan White
- Arnie Rheingold
- Peter Jones
- Mike Hursthouse
- Alex Slawin
- Frank Fronczek
- Judith Howard
- Hoong-Kun Fun
- Seik Weng Ng

CCDC

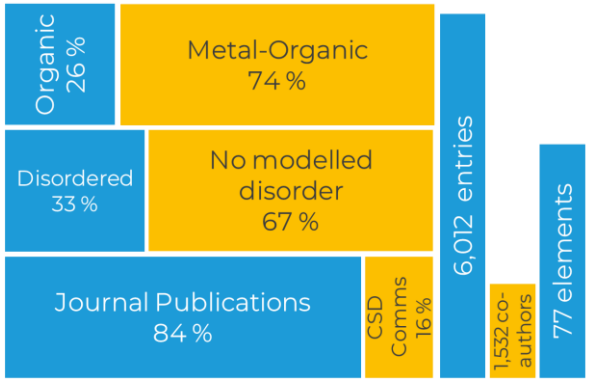
The top 10



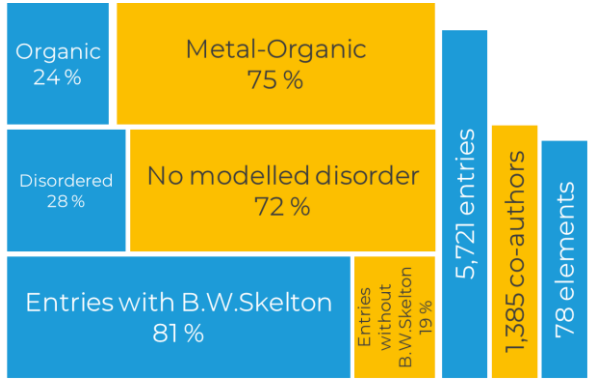
Seik Weng Ng



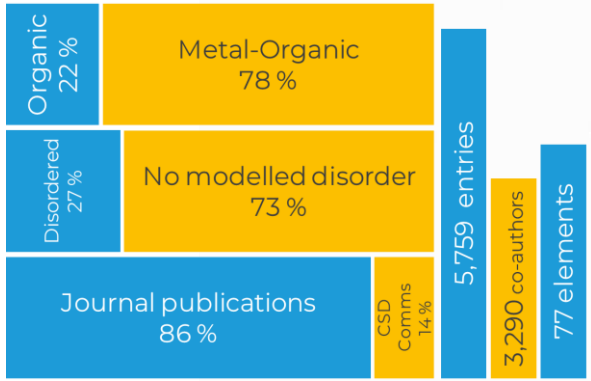
Hoong-Kun Fun



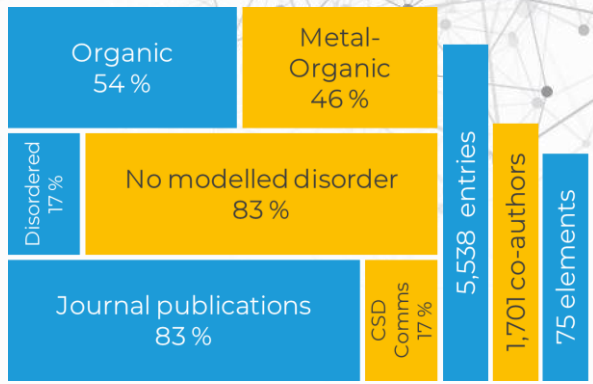
Brian Skelton



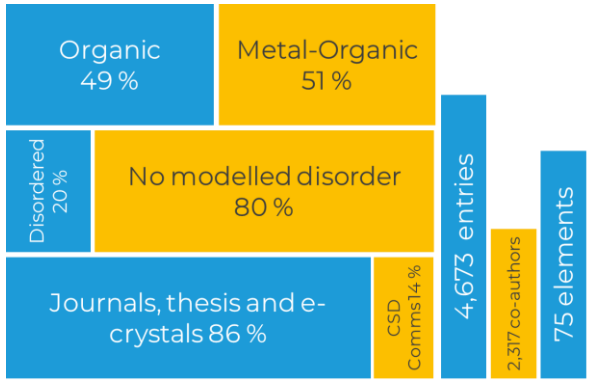
Allan H. White



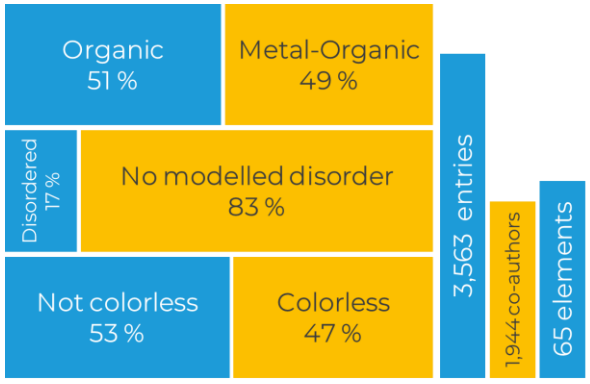
Arnie Rheingold



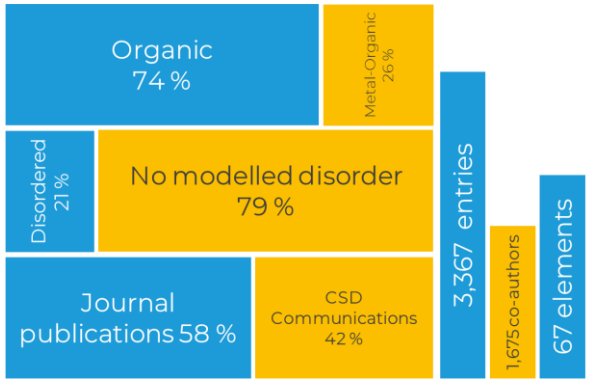
Peter Jones



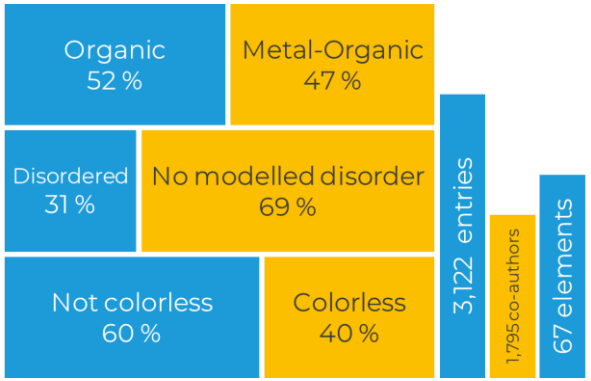
Mike Hursthouse



Alex Slawin



Frank Fronczek



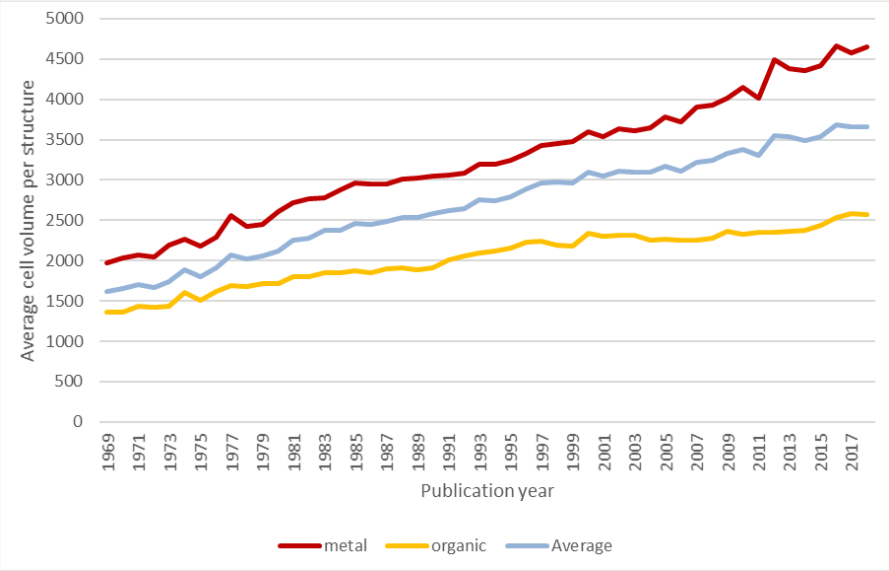
Judith Howard

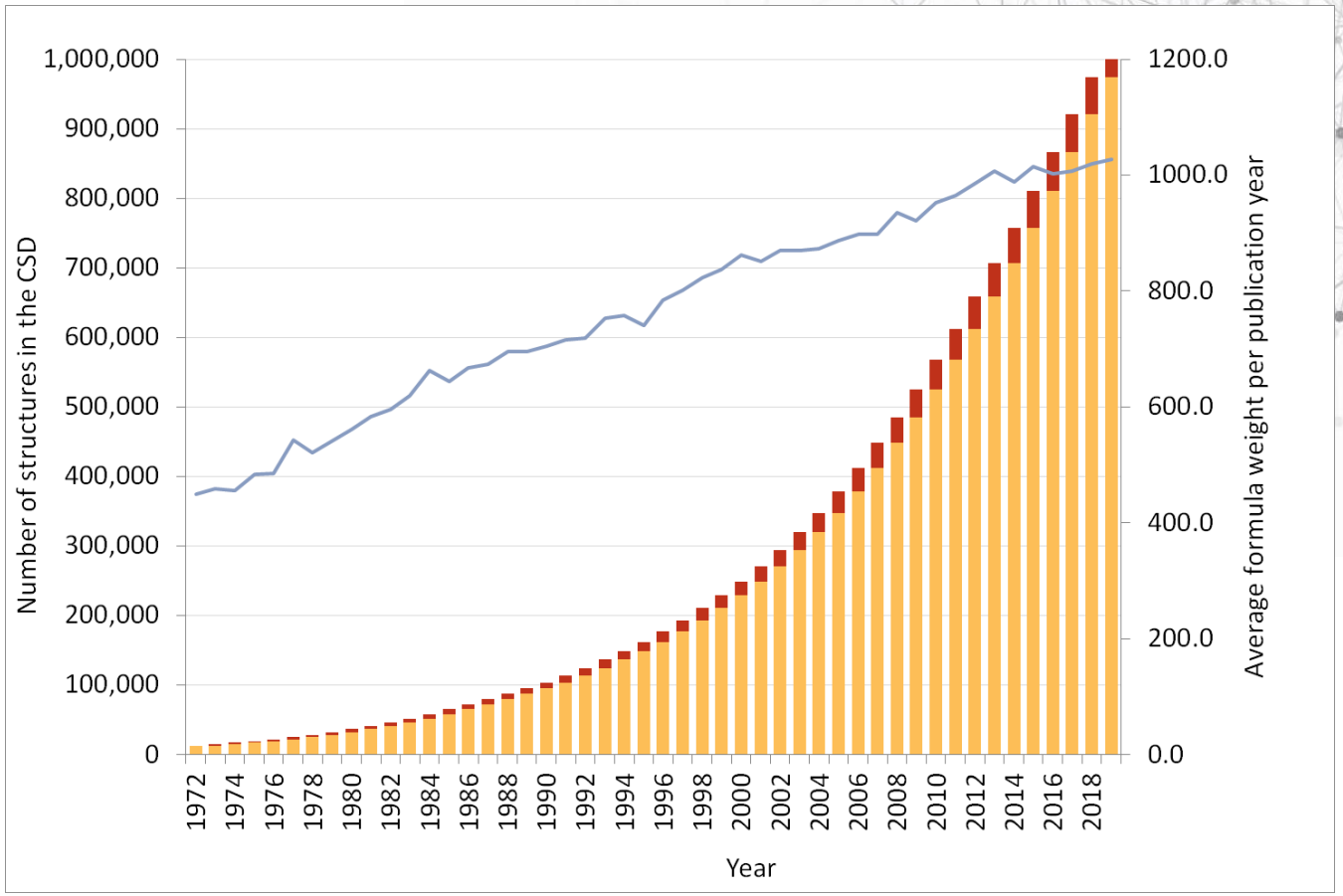

CAMBRIDGE CRYSTALLOGRAPHIC DATA CENTRE
CAMBRIDGE CRYSTALLOGRAPHIC DATA CENTRE

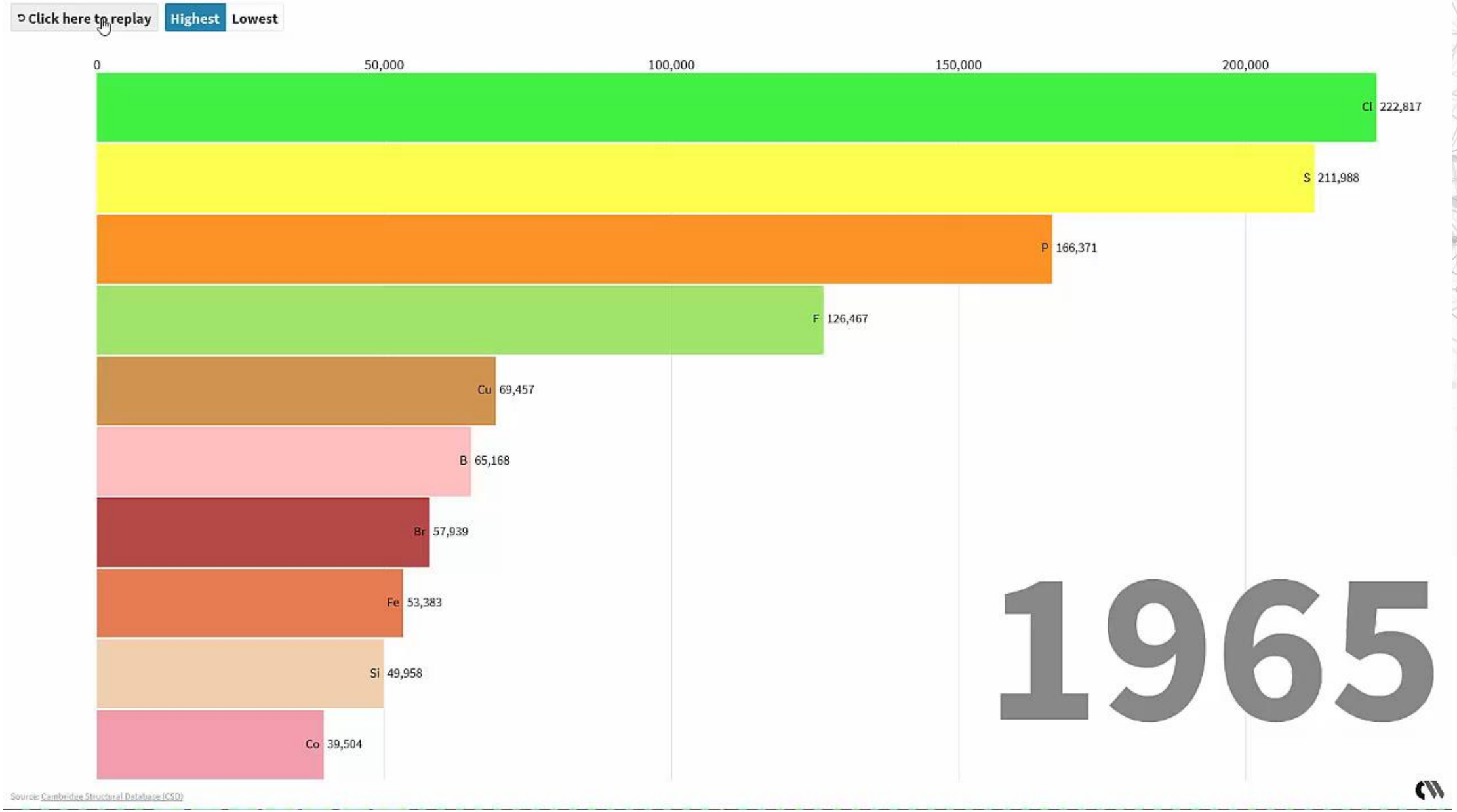
Increasing complexity

Increasing:

- Formula weights
- Unit Cells
- Number of elements

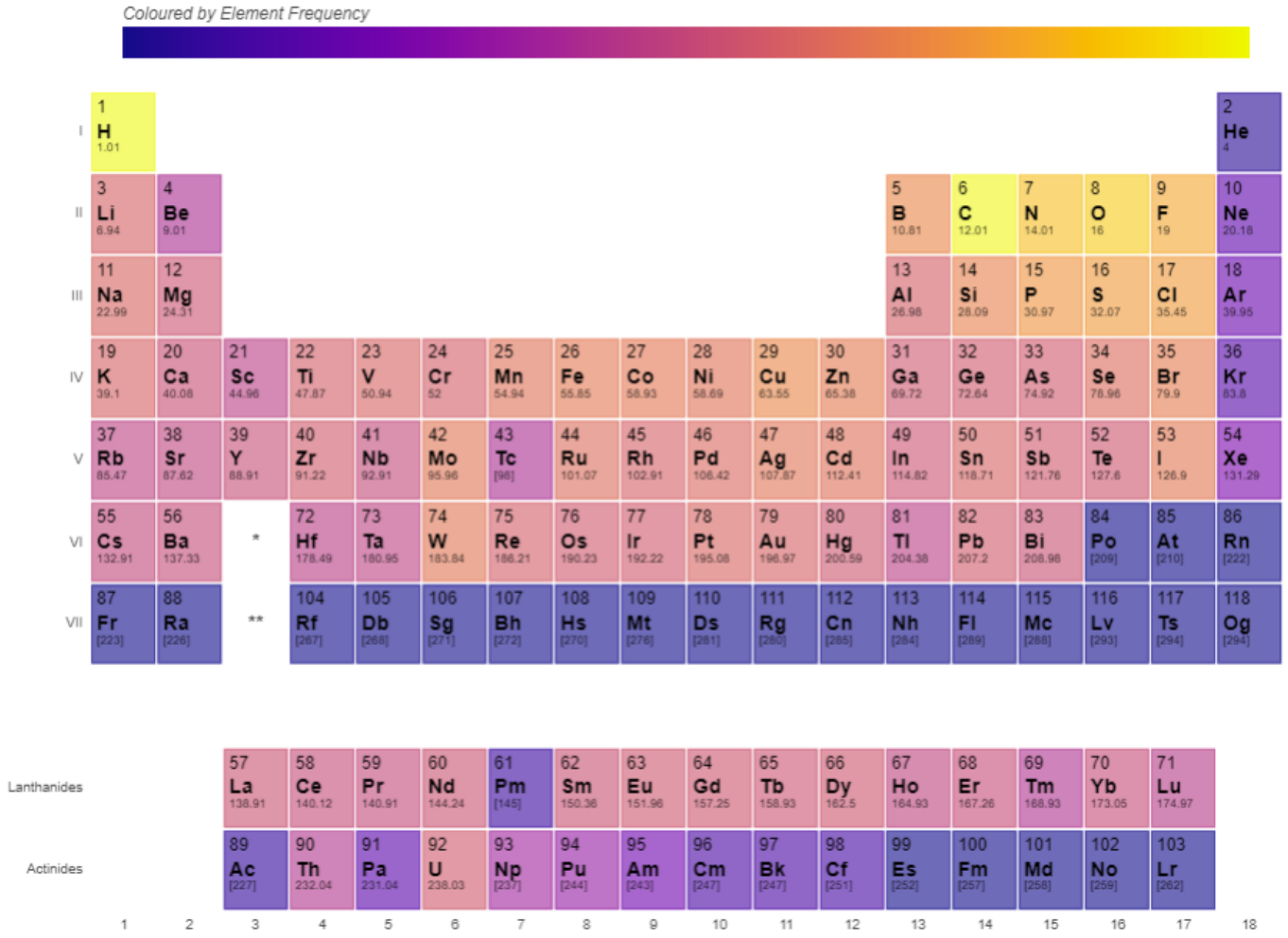


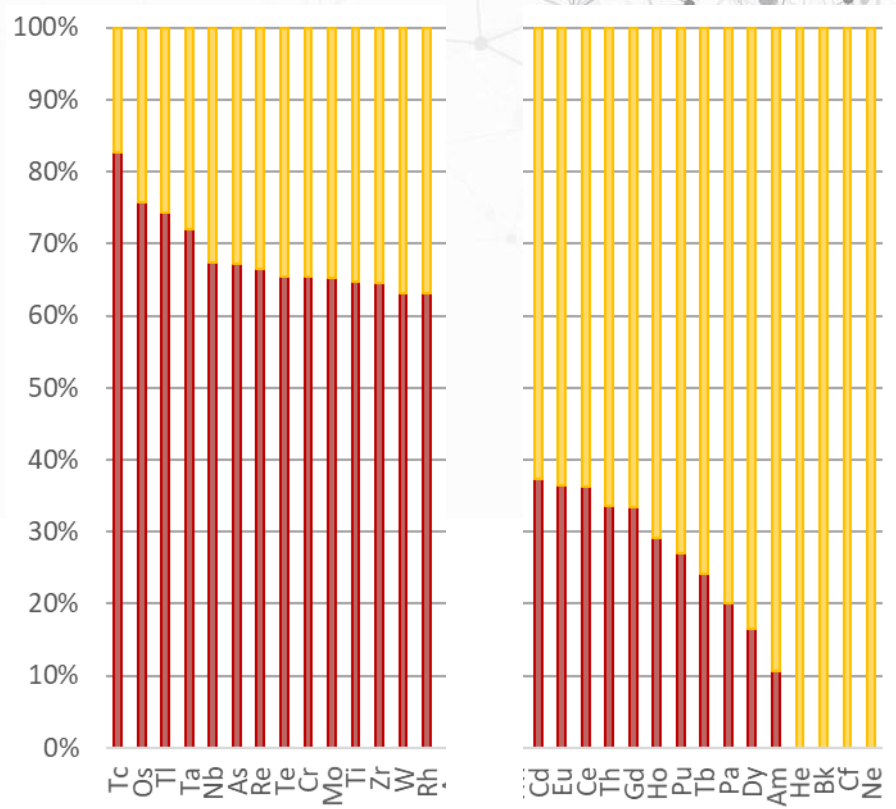



Click here to replay
Highest
Lowest
0
50,000
100,000
150,000
200,000
Cl 222,817
S 211,988
P 166,371
F 126,467
Cu 69,457
B 65,168
Br 57,939
Fe 53,383
Si 49,958
Co 39,504
1965
Source: Cambridge Structural Database (CSD)

CCDC

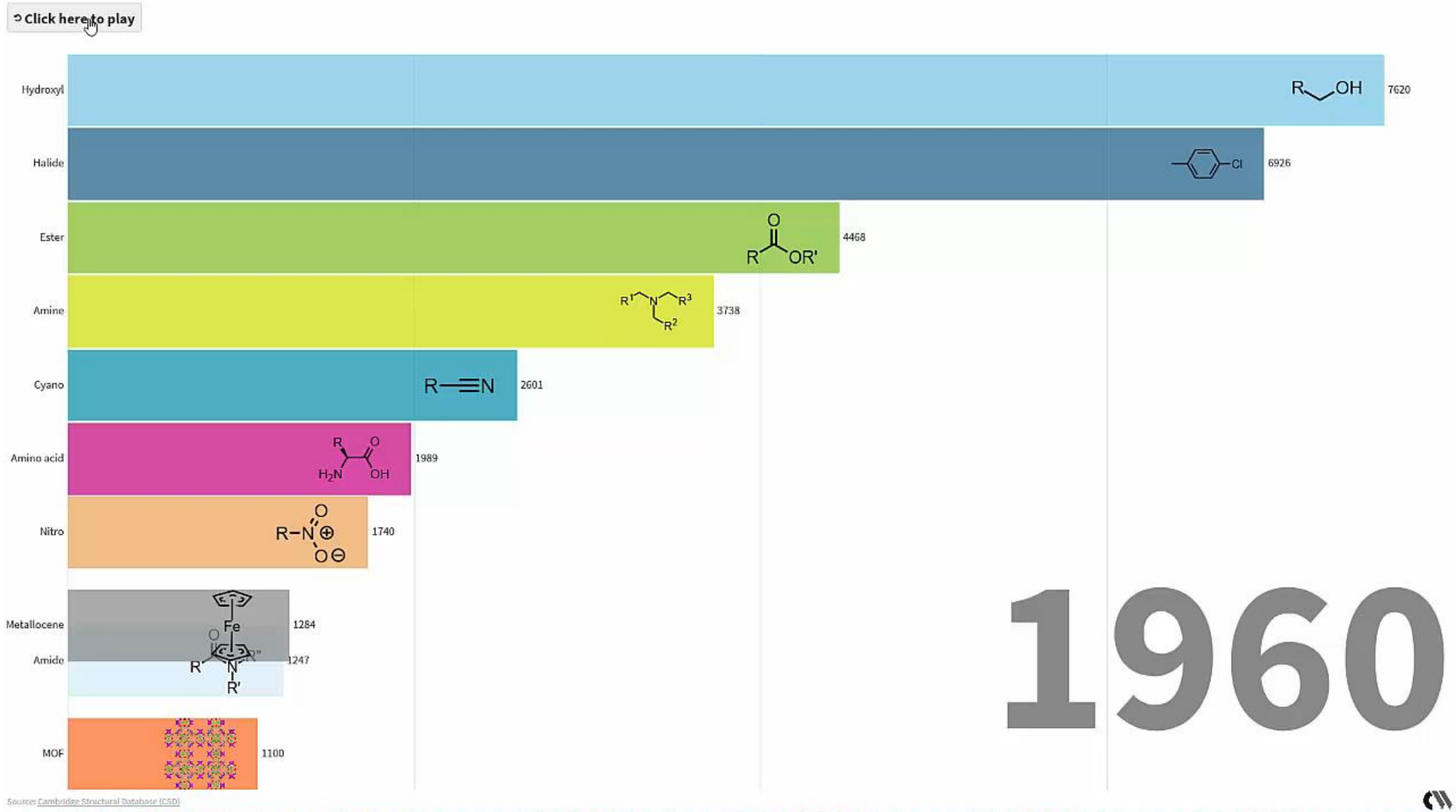
Elements in the CSD





Percentage of structures that contain each element from before (red) and after (yellow) 2009

CCDC

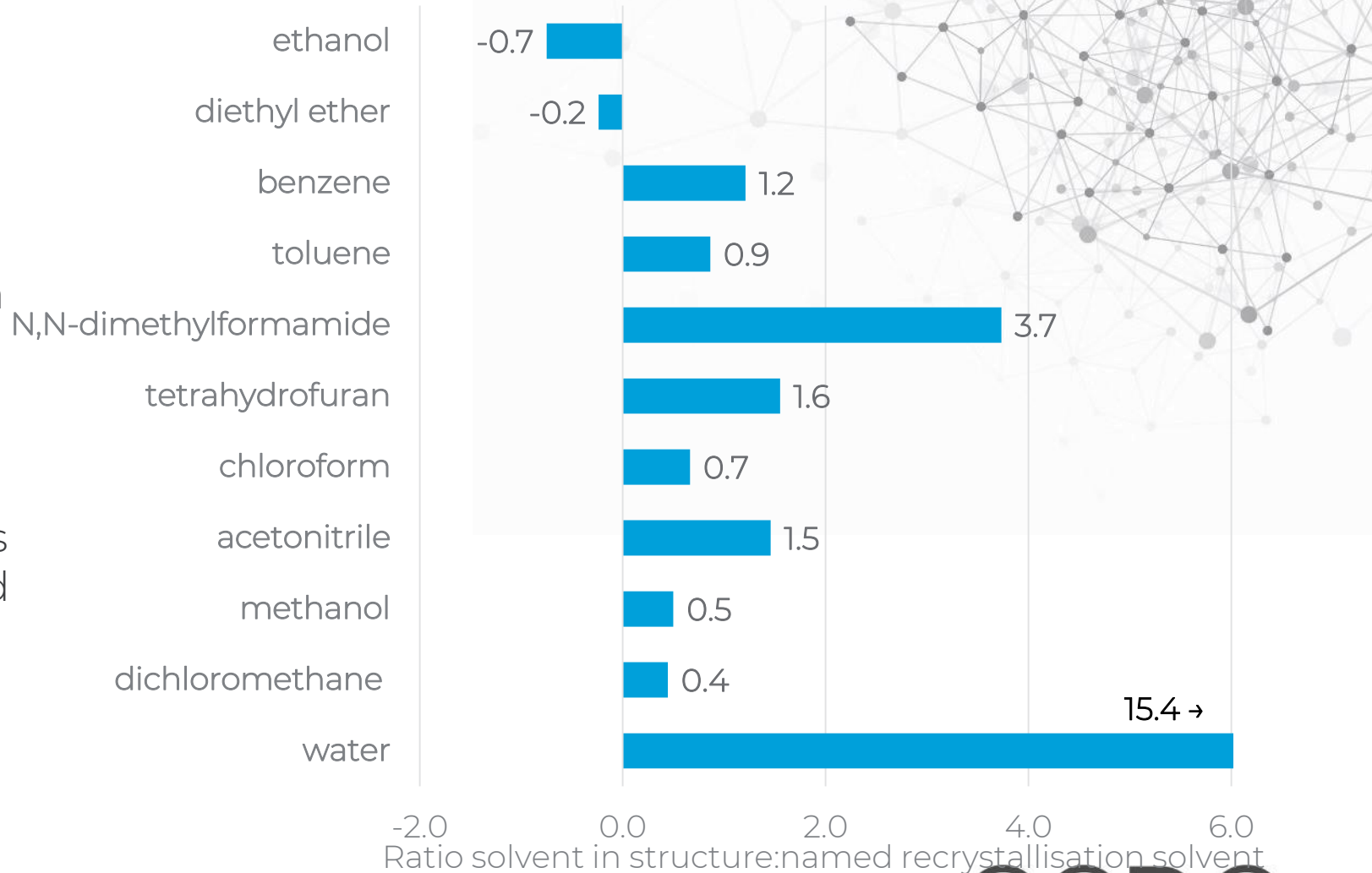

Click here to play
Hydroxyl
7620
Halide
6926
Ester
4468
Amine
3738
Cyano
2601
Amino acid
1989
Nitro
1740
Metallocene
1284
Amide
1247
MOF
1100
1960
Source: Cambridge Structural Database (CSD)

CCDC

CSD solvents

Crystal solvent molecules vs. recrystallisation solvent

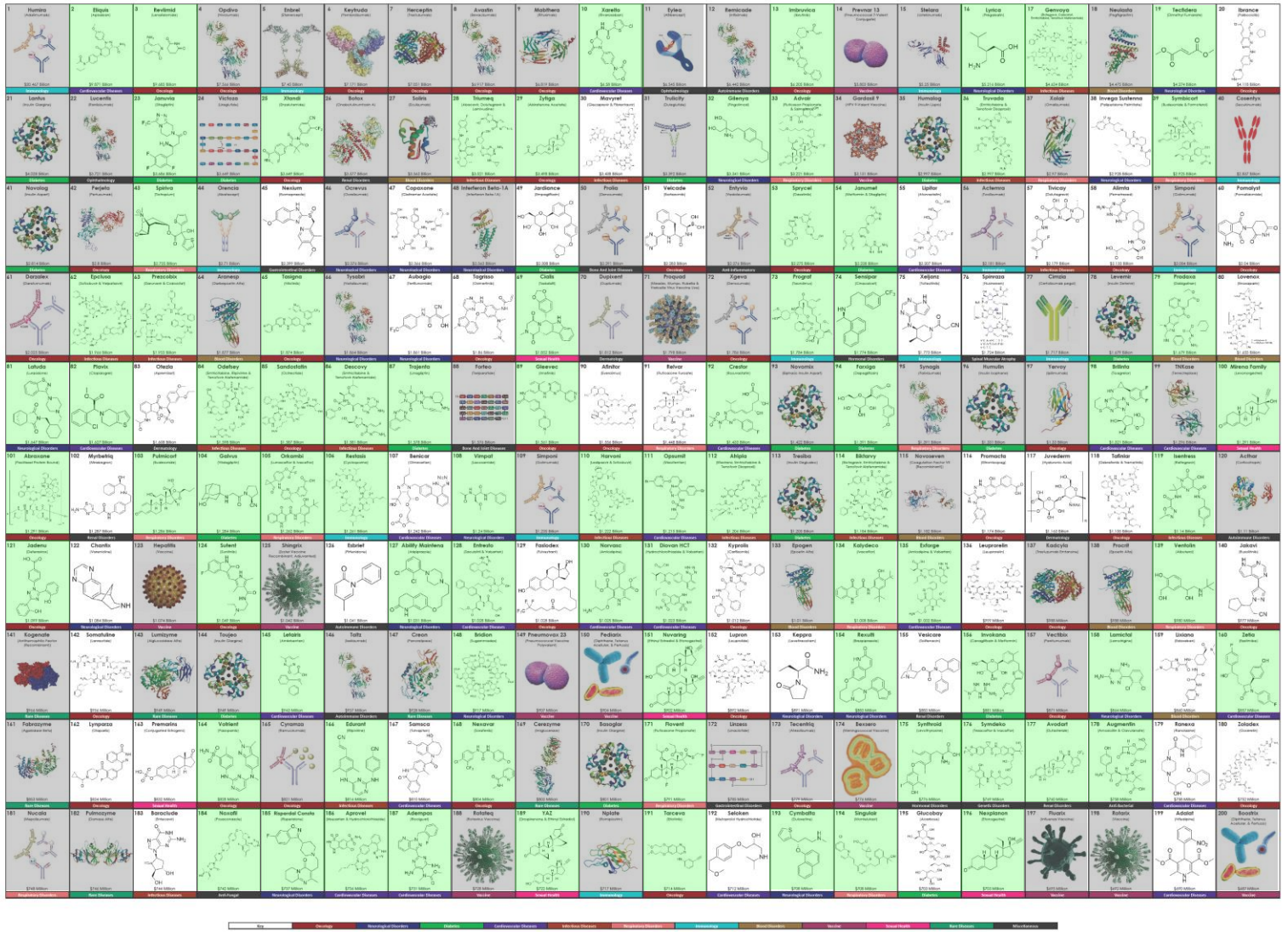
- Ethanol and diethyl ether more commonly named as a recrystallisation solvent than appear in a structure.
- Unsurprisingly water is the most common solvent and is much less likely to be named as a recrystallisation solvent



CCDC

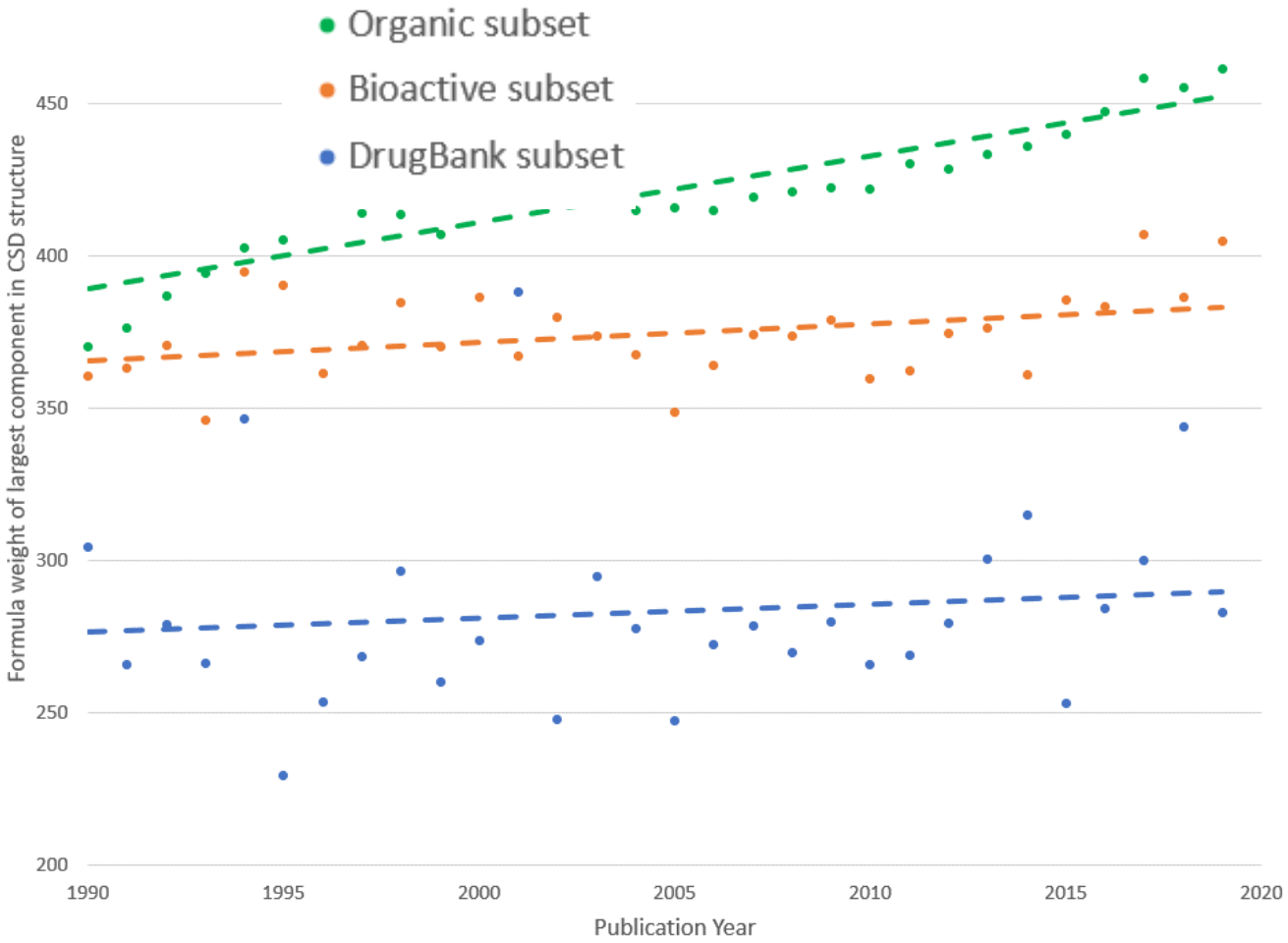
Drugs

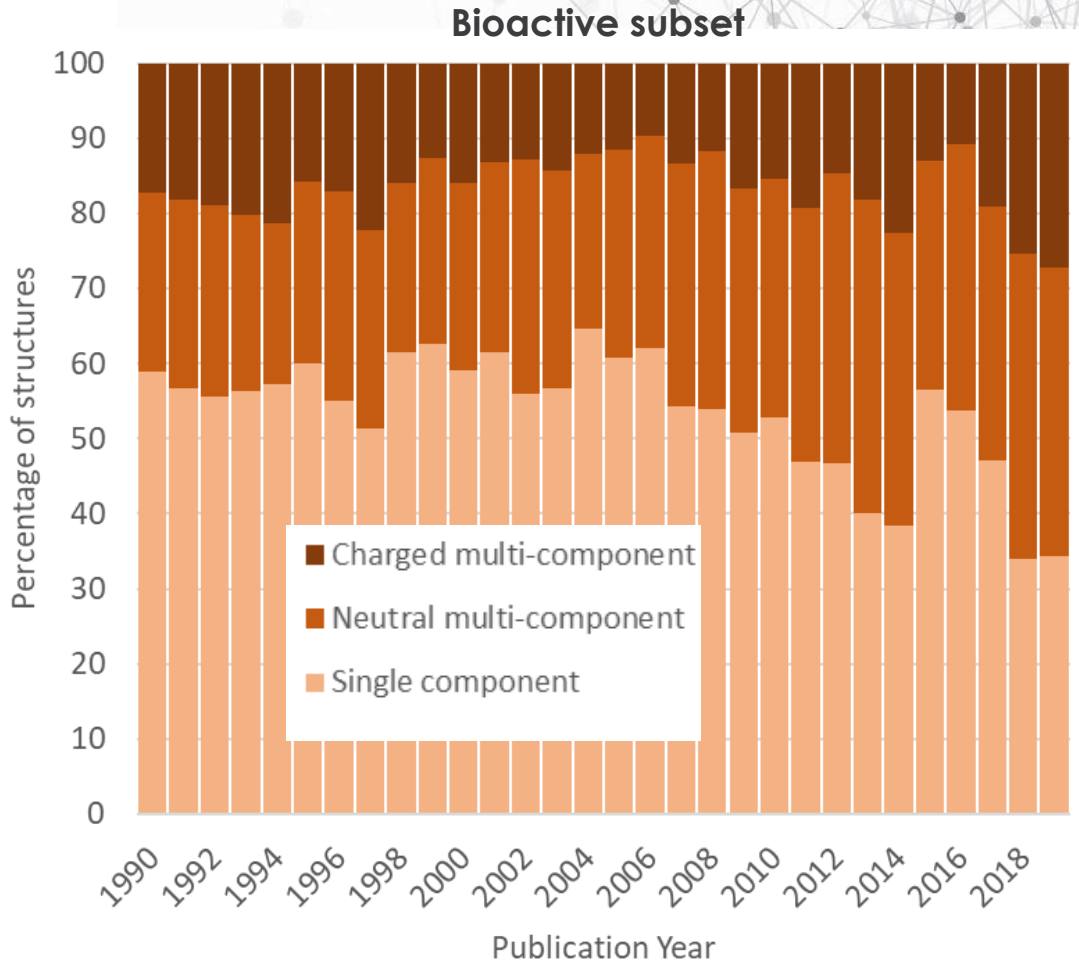
- Top 200 Pharmaceutical Products
  - By retail sales in 2018
  - Produced by the Njarðarson Group
  - The University of Arizona
  - Drugs in the CSD coloured green



Njarðarson *et al. J. Chem. Ed.* 2010, 87, 1348

Trends in drug structures





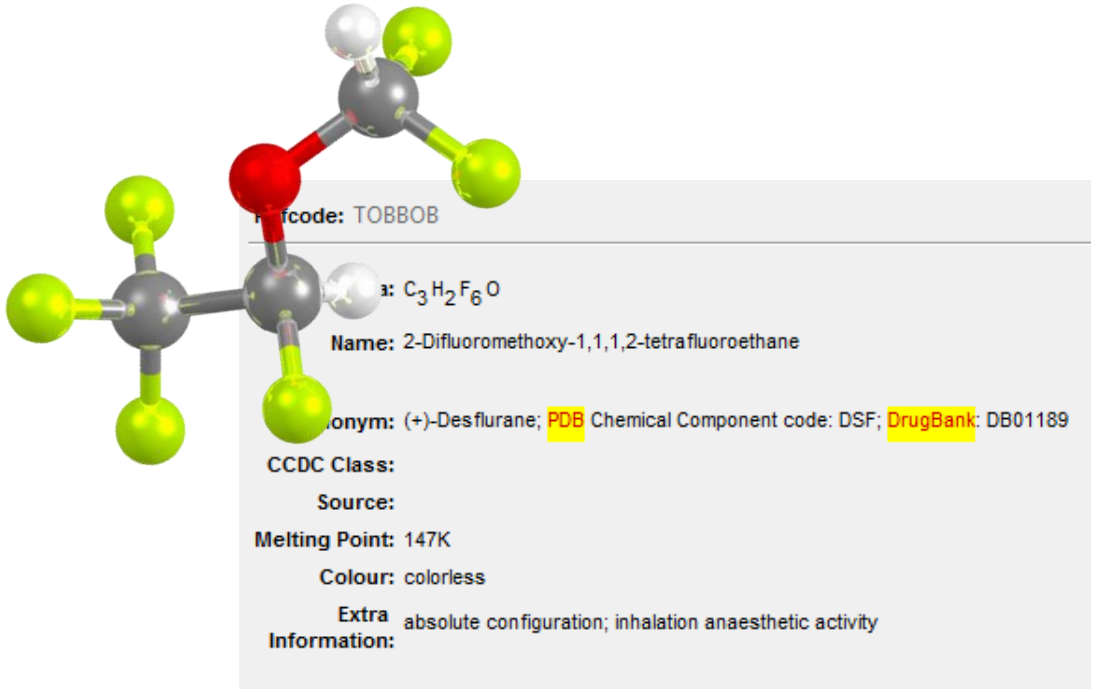
https://www.ccdc.cam.ac.uk/Community/blog/insights-into-drug-like-compounds-from-crystal-data/

CCDC

Moving to more connected data

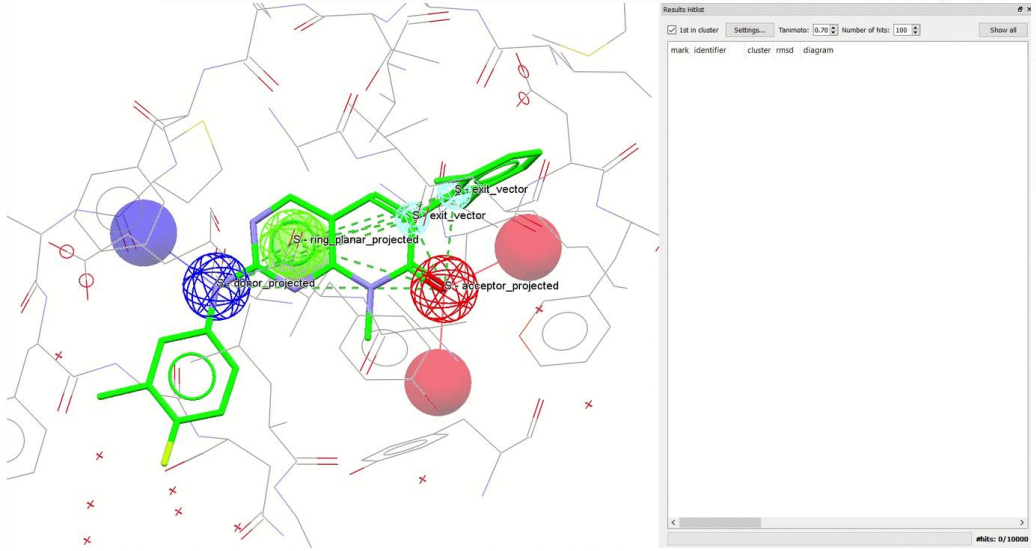
Linking

- Between CSD and PDB ligands



CSD-CrossMiner

- Pharmacophore query tool
- Searches the CSD and PDB



CCDC

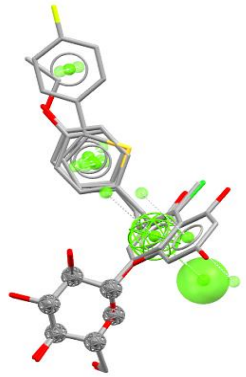
The CSD system today

Software enabling research across the breadth of structural science

CSD-Enterprise

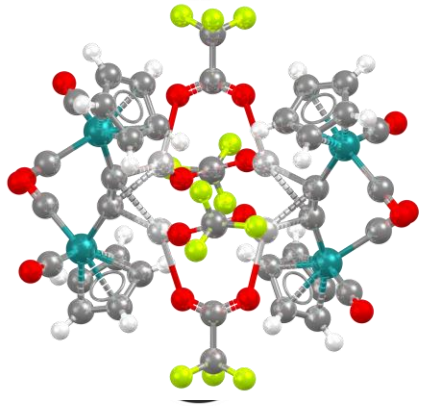
All CCDC application software (available to all Academics)

CSD-Discovery



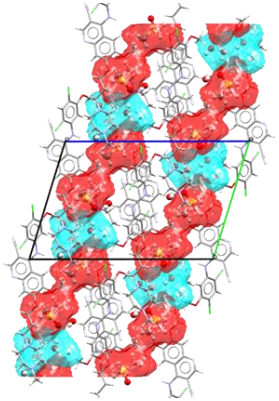
To discover new molecules with pharmaceutical applications

CSD-System



To search, visualise, analyse and communicate structural data

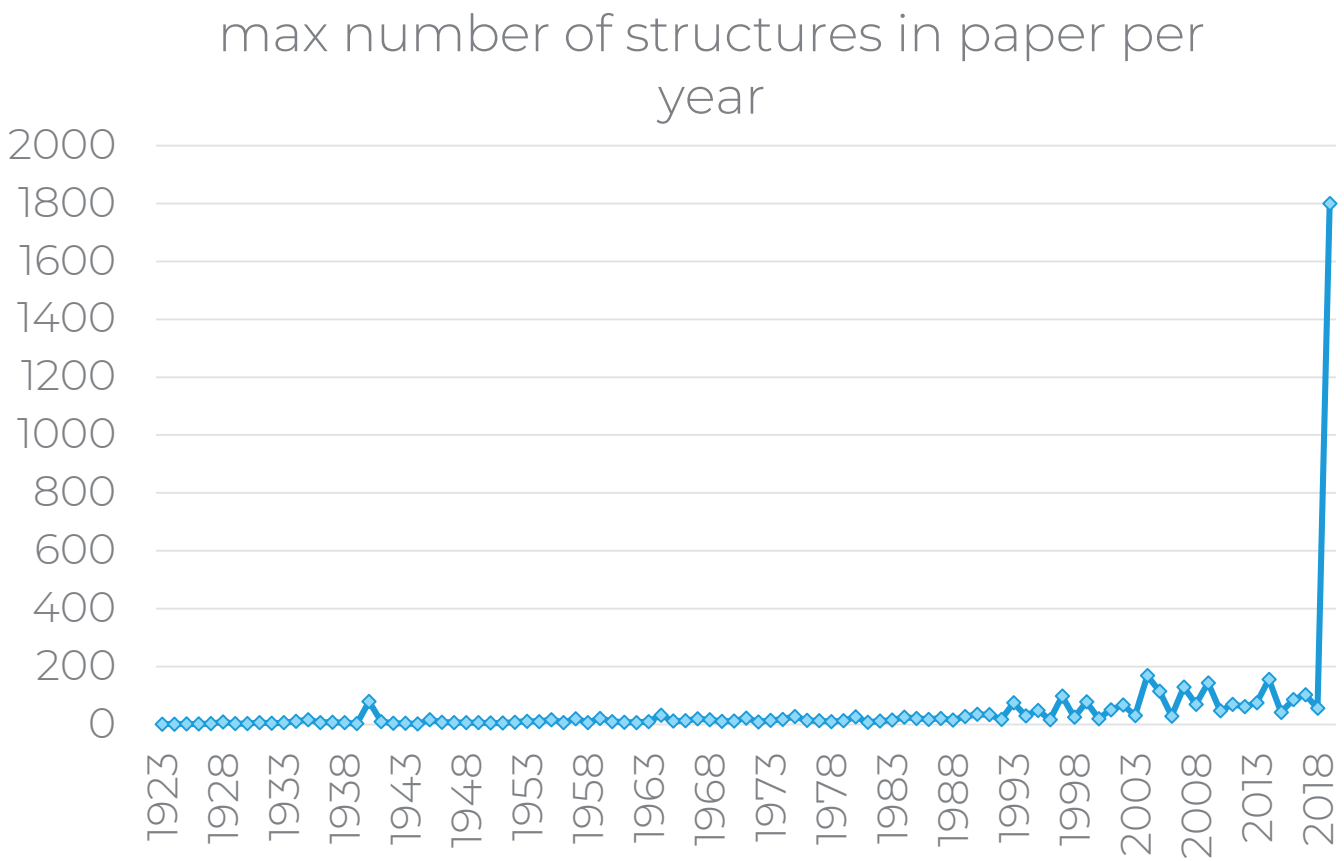
CSD-Materials

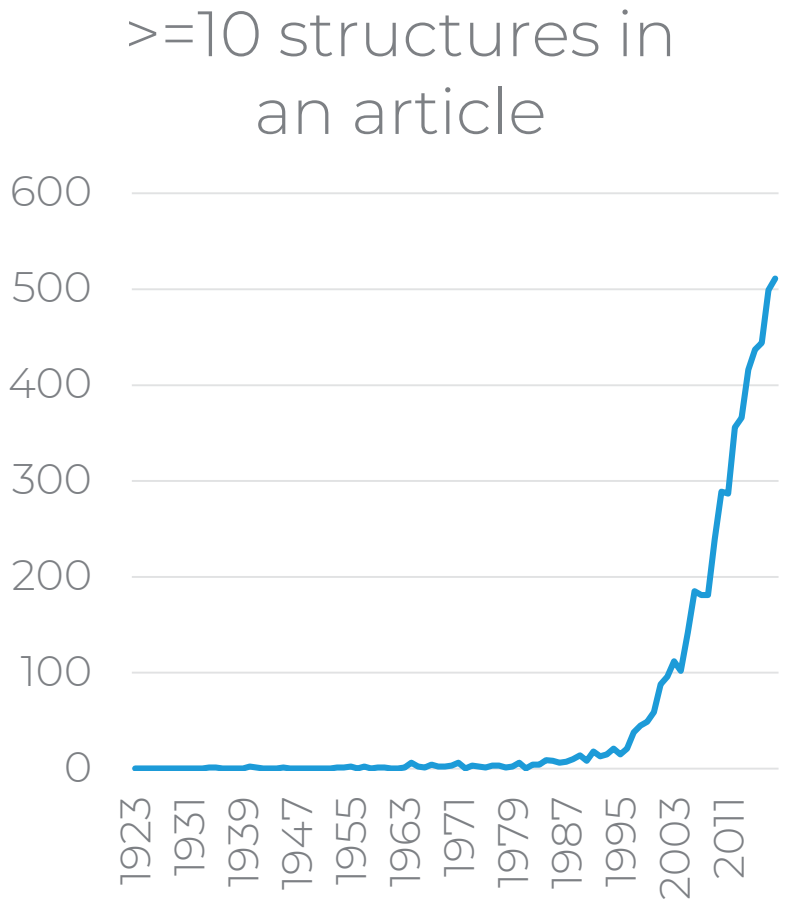


To understand and predict solid form stability and properties

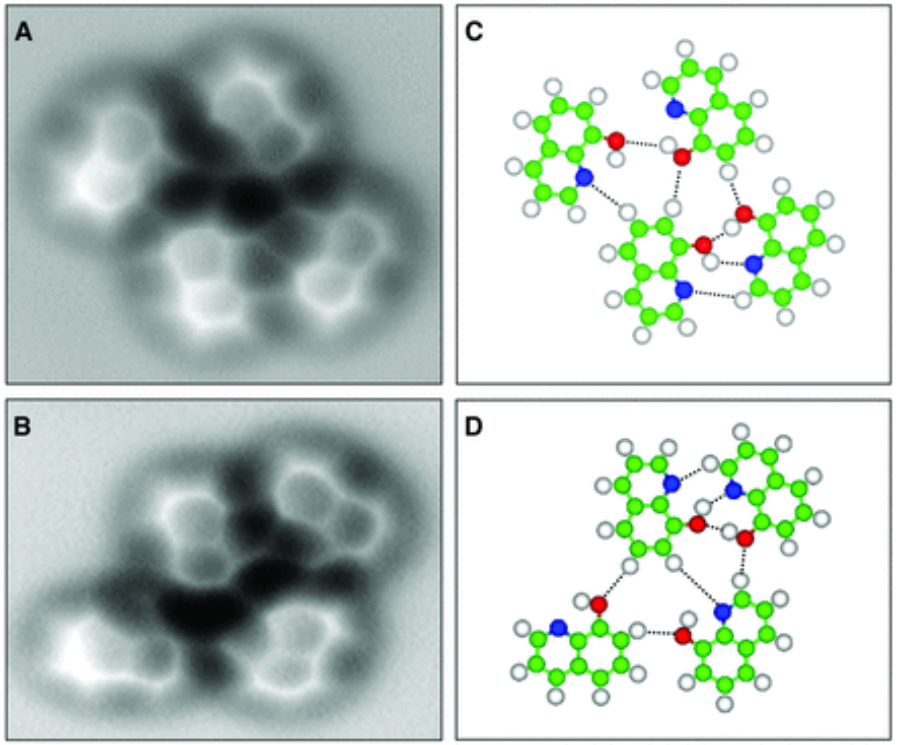
The Cambridge Structural Database

When will we reach the next million?

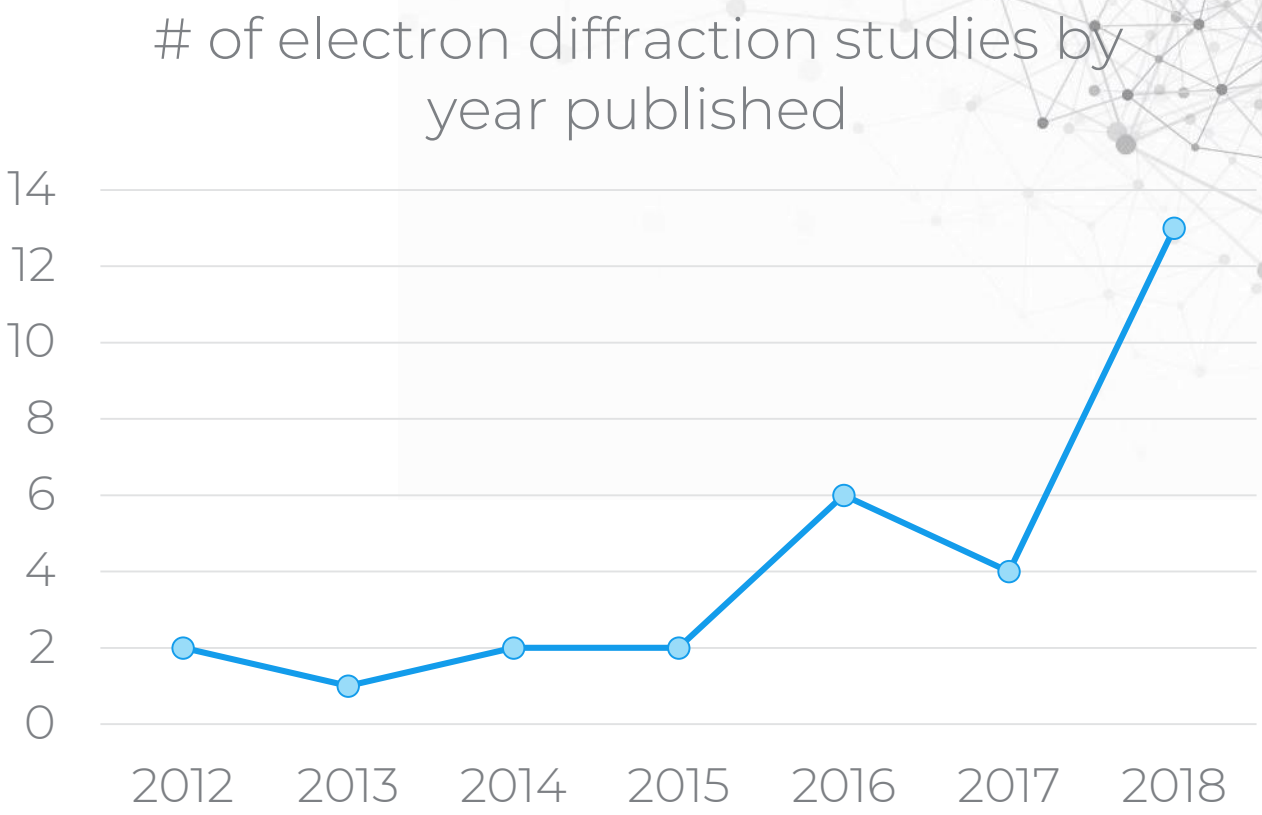




Emerging techniques



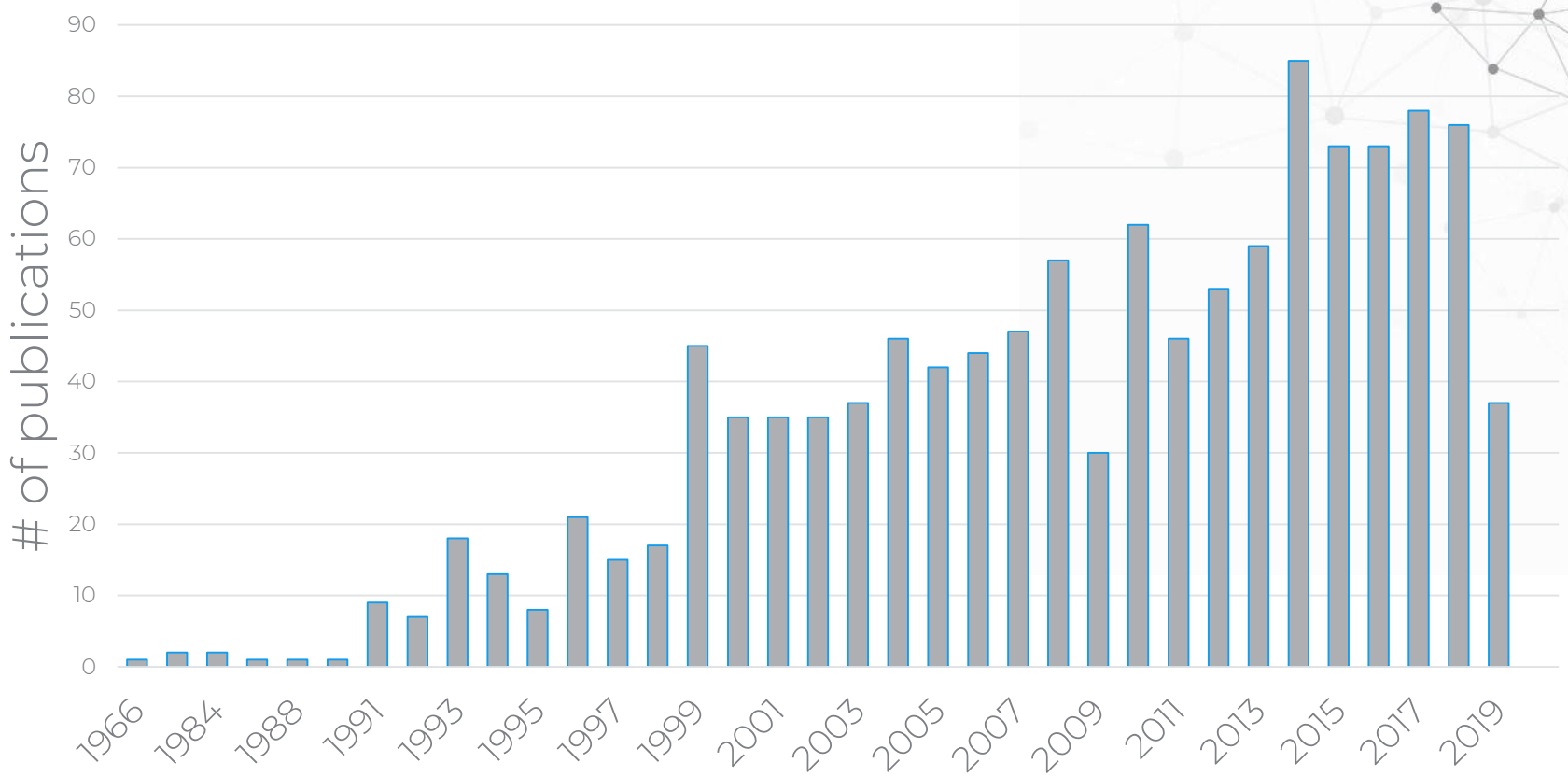
hydroxyquinline structure determined non-contact atomic force microscopy



J Zhang et al, Science, 2013. DOI: 10.1126/science.1242603

CCDC

The rise of calculated structures



Number of papers in Web of Knowledge with topic '**Crystal Structure Prediction**' and category 'Crystallography'

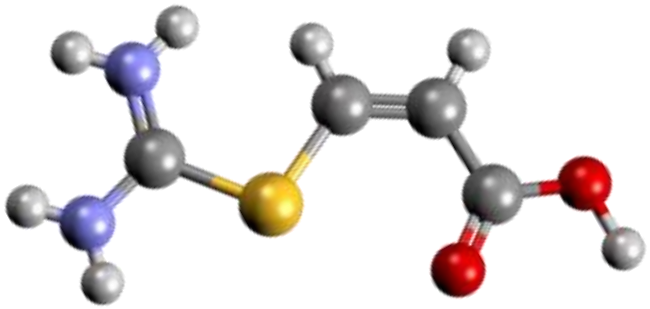
CCDC

Using the collection to help predict structures

CCDC Blind Test Showcases Major Advance in Crystal Structure Prediction Methods

– November 03, 2015

The Cambridge Crystallographic Data Centre (CCDC) announces that the results of its 6th blind test of crystal structure prediction methods demonstrate significant advancement in crystal structure prediction methods in comparison with previous tests. This year, structures of all of the test systems, which i… of polymorphs, salts and hydrates, were generated by one or more methods. In additio… experimental structures were predicted to be the most stable form.



nature International weekly journal of science

Search Go

Advanced search

Home | News & Comment | Research | Careers & Jobs | Current Issue | Archive | Audio & Video | For Authors

Archive > Volume 527 > Issue 7576 > News > Article

Software predicts slew of fiendish crystal structures

Chemists succeed at forecasting how complex molecules will assemble in 3D.

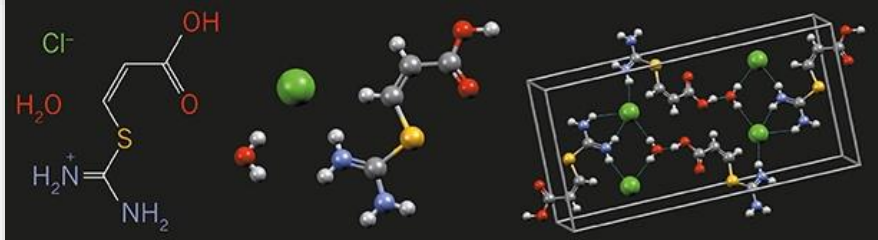
Elizabeth Gibney

…kin and it may not be apparent that there are …3D crystal. Now, a collaboration of dozens of …fully predicted the crystal structure of five, …ing but a 2D map showing which atoms …

…rkshop in Cambridge, UK, paves the way for …anufacture of drugs and other chemical …damental chemistry.

CRYSTAL CHALLENGE

The 3D structure that a molecule adopts in a crystal is very difficult to predict — but defines what properties the molecule has.



The structural formula of a molecule reveals which atoms are connected at a 2D level.

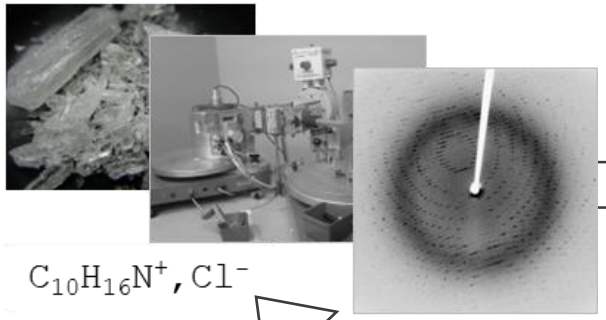
Chemists are making progress at predicting how complex molecules will assemble in 3D space — there are millions of possibilities.

The 3D orientation repeats in a crystalline lattice with a structure that dictates the molecule's mechanical, chemical and physical properties.

CCDC

From data to knowledge

Data



$C_{10}H_{16}N^{+},Cl^{-}$

```
_diffrn_ambient_temperature      90(2)
_diffrn_radiation_type           MoK\a
_diffrn_radiation_wavelength     0.71073
_diffrn_radiation_monochromator  graphite
_diffrn_measurement_device_type  'Bruker APEX CCD area-detector'
_diffrn_measurement_method       '\f and \w'
_diffrn_detector_area_resol_mean 512
_diffrn_reflns_number            5892

loop_
_atom_site_type_symbol
_atom_site_label
_atom_site_fract_x
_atom_site_fract_y
_atom_site_fract_z
_atom_site_U_iso_or_equiv
_atom_site_adp_type
_atom_site_calc_flag
_atom_site_refinement_flags
_atom_site_occupancy
_atom_site_symmetry_multiplicity
_atom_site_disorder_assembly
_atom_site_disorder_group
Cl Cl1 0.23185(8) 0.78305(9) 0.55574(6) 0.02213(16) Uani d . 1 1 . .
N N1 0.8031(3) 0.6811(3) 0.5363(2) 0.0172(4) Uani d U 1 1 . .
C C1 0.6896(4) 0.7867(6) 0.4357(2) 0.0224(5) Uani d U 1 1 . .
C C2 0.7510(5) 0.8922(5) 0.7083(3) 0.0256(6) Uani d U 1 1 . .
C C3 0.7409(4) 0.6944(4) 0.6644(3) 0.0187(5) Uani d U 1 1 . .
C C4 0.8700(4) 0.5637(4) 0.7481(3) 0.0236(6) Uani d U 1 1 . .
```

Pharmaceutical Research
June 2001, Volume 18, Issue 6, pp 859–866 | Cite as

Ritonavir: An Extraordinary Example of Conformational Polymorphism

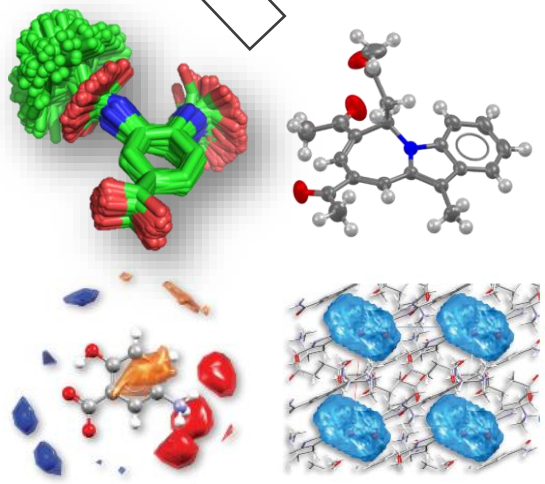
Authors | Authors and affiliations

John Bauer, Stephen Spanton, Rodger Henry, John Quick, Walter Dziki, William Porter, John Morris

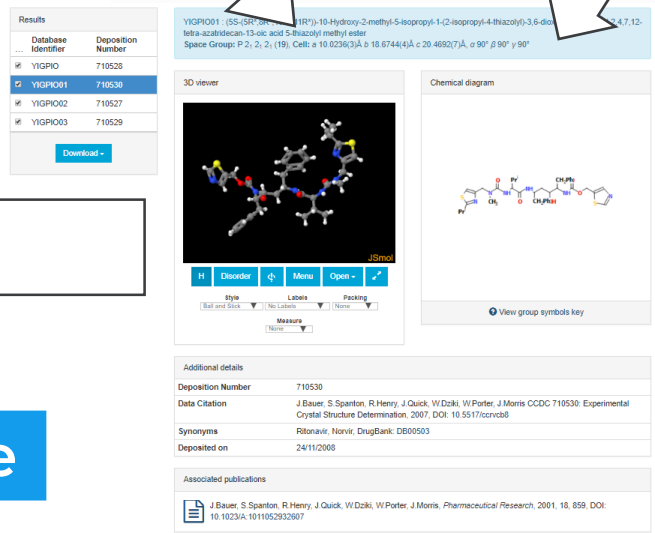
Article | 3 Shares | 5.3k Downloads | 457 Citations

Abstract

Purpose. In the summer of 1998, Norvir semi-solid capsules supplies were threatened as a result of a new much less soluble crystal form of ritonavir. This report provides characterization of the two polymorphs and the structures and hydrogen bonding network for each form.



CCDC



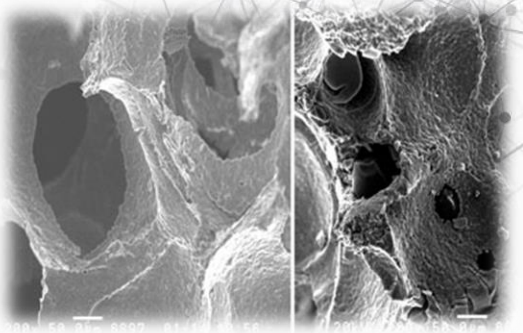
Knowledge

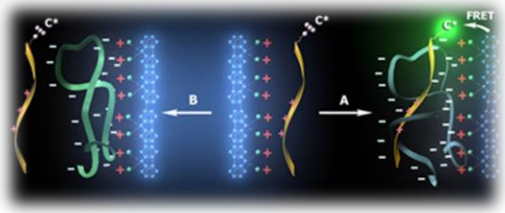
CCDC

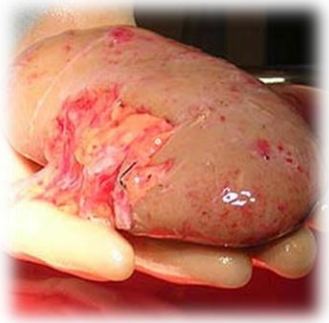
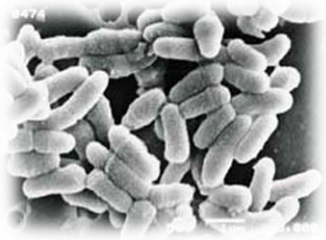
How can structural data be used?

Non-linear optical materials
Piezoelectric crystals
Pigments
Organic semiconductors
Liquid crystals
Pharmaceuticals
Biological systems
Gas adsorption
Catalysis
Etc etc!







CCDC

Using AI and machine learning

- AI and machine learning techniques are evolving rapidly
- But the consequences of using poor quality data can be far reaching
  - Incorrect scientific conclusions
  - Wasted investment and effort
  - A loss of trust
  - Ultimately poor business or research decisions.

Many of the most pressing challenges facing AI today resolve around its poor-quality training data.....While algorithmic improvements could help, so could having proper training data.

959 views | Feb 5, 2019, 11:19am

Why Is Our AI Revolution Built On Free Data Rather Than Good Data?

Kalev Leetaru Contributor ⓘ
AI & Big Data
I write about the broad intersection of data and society.

Getty Images. GETTY

One of the greatest challenges confronting the modern AI revolution is the lack of sufficiently diverse training data. From driverless cars that struggle with the most basic scenarios to image recognition software that fails spectacularly on entire demographics, our AI systems today have been hamstrung by a critical lack of data representing the full extent of their operating domains. The reason is that for all of the investment being made in the people, computing hardware and algorithms powering AI, we refuse to make even the most rudimentary investments in the lifeblood of AI: data. In short, our AI revolution is being built on free data, rather than good data.

https://www.forbes.com/sites/kalevleetaru/2019/02/05/why-is-our-ai-revolution-built-on-free-data-rather-than-good-data/

CCDC

Inspiring a new generation





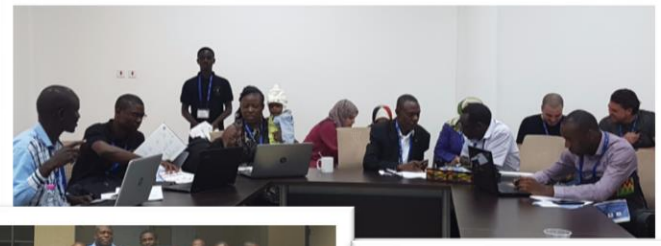








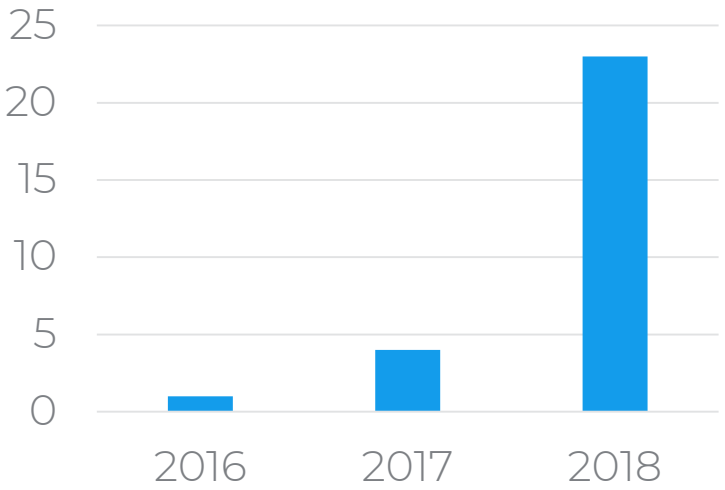


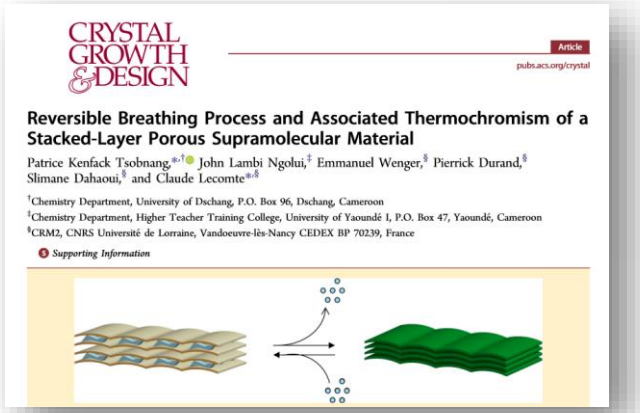



Expanding access

- Frank Allen International Research and Education (FAIRE) **Programme**
- Supports research & education in developing countries through use of the CSD

CSD-Enterprise has really revolutionized my approach to the teaching of coordination chemistry and inorganic chemistry.... I have been able to publish three articles in peer reviewed journals.




CRYSTAL GROWTH & DESIGN

Article

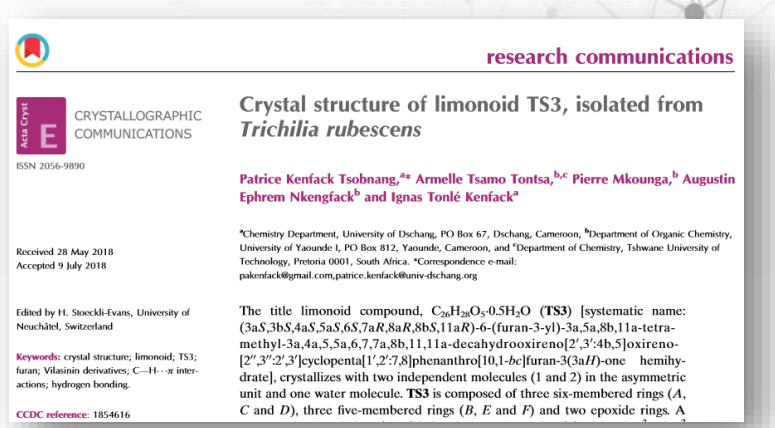
pubs.acs.org/crystal

Reversible Breathing Process and Associated Thermochromism of a Stacked-Layer Porous Supramolecular Material

Patrice Kenfack Tsobnang,*[†] John Lambi Ngolui,[‡] Emmanuel Wenger,[§] Pierrick Durand,[§] Slimane Dahaoui,[§] and Claude Lecomte*[§]

[†]Chemistry Department, University of Dschang, P.O. Box 96, Dschang, Cameroon
[‡]Chemistry Department, Higher Teacher Training College, University of Yaoundé I, P.O. Box 47, Yaoundé, Cameroon
[§]CRM2, CNRS Université de Lorraine, Vandoeuvre-lès-Nancy CEDEX BP 70239, France

Supporting Information


research communications

CRYSTALLOGRAPHIC COMMUNICATIONS

ISSN 2056-9890

Crystal structure of limonoid TS3, isolated from *Trichilia rubescens*

Patrice Kenfack Tsobnang,[a]* Armelle Tsamo Tontsa,[b,c] Pierre Mkounga,[b] Augustin Ephrem Nkengfack[b] and Ignas Tonlé Kenfack[a]

[a]Chemistry Department, University of Dschang, PO Box 67, Dschang, Cameroon, [b]Department of Organic Chemistry, University of Yaounde I, PO Box 812, Yaounde, Cameroon, and [c]Department of Chemistry, Tshwane University of Technology, Pretoria 0001, South Africa. *Correspondence e-mail: pakenfack@gmail.com,patrice.kenfack@univ-dschang.org

Received 28 May 2018
Accepted 9 July 2018

Edited by H. Stoeckli-Evans, University of Neuchâtel, Switzerland

Keywords: crystal structure; limonoid; TS3; furan; Vilasinin derivatives; C—H···π interactions; hydrogen bonding.

CCDC reference: 1854616

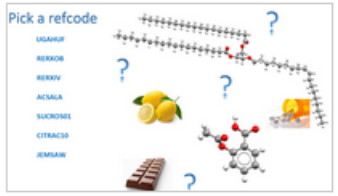
The title limonoid compound, $C_{26}H_{28}O_5 \cdot 0.5H_2O$ (**TS3**) [systematic name: (3a*S*,3b*S*,4a*S*,5a*S*,6*S*,7a*R*,8a*R*,8b*S*,11a*R*)-6-(furan-3-yl)-3a,5a,8b,11a-tetramethyl-3a,4a,5,5a,6,7,7a,8b,11,11a-decahydrooxireno[2′,3′:4b,5]oxireno[2″,3″:2′,3′]cyclopenta[1′,2′:7,8]phenanthro[10,1-*bc*]furan-3(3a*H*)-one hemihydrate], crystallizes with two independent molecules (1 and 2) in the asymmetric unit and one water molecule. **TS3** is composed of three six-membered rings (*A*, *C* and *D*), three five-membered rings (*B*, *E* and *F*) and two epoxide rings. A

Acknowledgements

The authors thank the International Union of Crystallography (IUCr) and the Cambridge Crystallographic Data Center (CCDC) for their initiative to promote crystallography and structural studies in Africa and particularly in Cameroon.

CCDC

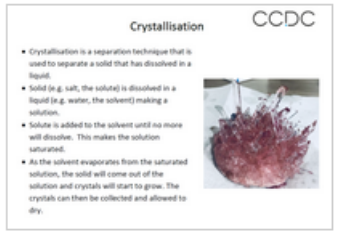
Engaging future scientists



Classroom Teaching Module: Structure Exploration

Click to download the packet for use in the classroom

This module was developed for use in the classroom based on hands-on exercises used at the Cambridge Science Fes
(Wilfrid Laurier University) for assistance with curriculum benchmarking.



Classroom Teaching Module: Crystallization

Click to download the packet for use in the classroom

This module was developed for use in the classroom based on hands-on exercises used at the Cambridge Science Fes
(Wilfrid Laurier University) for assistance with curriculum benchmarking and Prof. Robert Pike at The College of William
resources.

CCDC

Towards the great ocean of truth

The amount of scientific knowledge we now have to hand has lead some eminent scholars to believe that there are intrinsic limits to science and that this limit has now been reached. I myself do not believe that this is the case. I think that the great ocean of truth is still in front of us and that we will continue to discover new aspects of this truth. Some of them will be discovered through the insight of outstanding individuals and some through the insight which Bernal predicted could be gained by more ordinary mortals through the analysis and transformation of the pebbles of information which have accumulated over the past decades. We have the tools and resources to do this and Bernal's inspiration is still with us.

CCDC

Thank You

More talks from the CCDC

Today 11:20 I didn't Know Mercury Could Do That! – Andy Maloney
Today 15:20 Keeping Things 'N Synch –Natalie Johnson
Thursday 11:40 Analysing Aromatic Interactions- Ioana Sovago

Visit us at the expo – D08

Thank you and...

The CCDC Database Team

Past and present members of the CCDC

The **396,937 authors** of structures

All the many contributors to the CSD

Olga for her amazing vision

http://www.ccdc.cam.ac.uk/

@ccdc_cambridge

ccdc.cambridge

CCDC