IUPAC, nomenclature, and chemical representation: From the perspective of a worldwide structural database

Matt Lightfoot, Ian Bruno, Clare Tovee, Suzanna Ward, Seth Wiggin

The Cambridge Crystallographic Data Centre

ACS Fall 2019 Sunday August 25th 2019

CCDC
advancing structural science

Summary

- Introduction to the CCDC and the CSD
- The history of nomenclature in the CSD
- The importance of compound names in the CSD
- Current challenges with nomenclature
- Looking forward

IUPAC's role in creating the CSD over the past 50 years

CCDC

The Cambridge Crystallographic Data Centre

International Data Repository
Archive of crystal structure data
High quality scientific database

Collaborative Research Organisation
New methodologies
Fundamental research

Scientific Software Provider
Search/analysis/visualisation tools
Scientific applications

Education and Outreach
Conferences, Workshops,
Training, Teaching

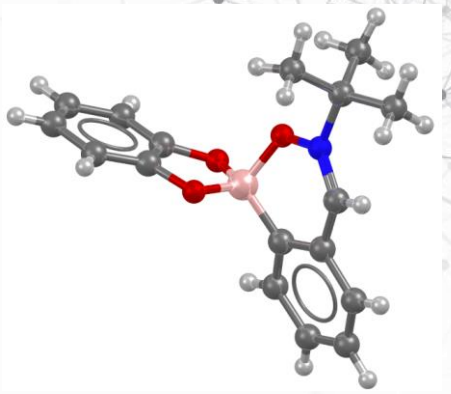
Originated in 1965

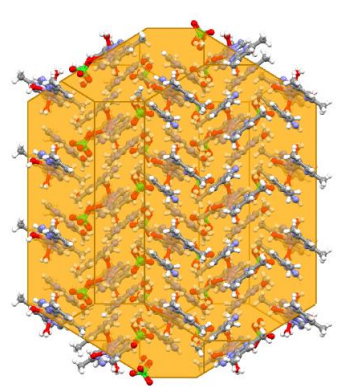
Financially self-supporting

Not-for-profit, UK Registered Charity

University Partner Institute

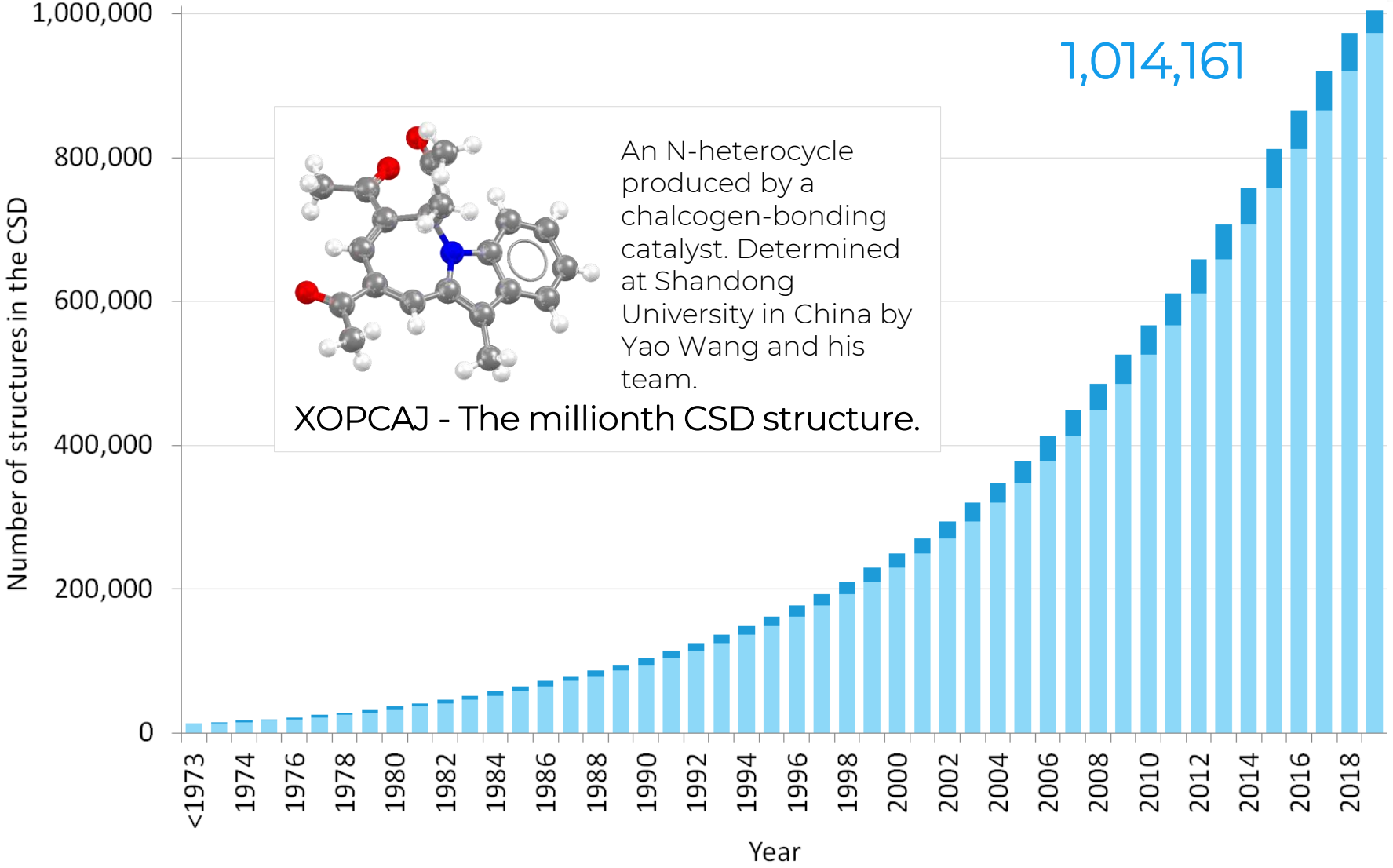
Dedicated to the advancement of **chemistry and crystallography** for the public benefit through high quality **information services and software**





CCDC

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.

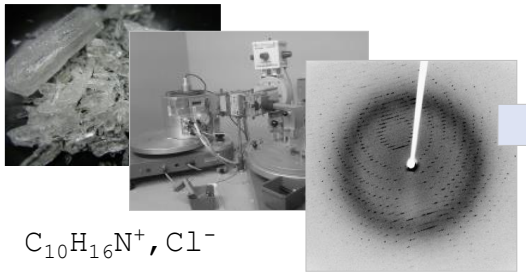
XOPCAJ - The millionth CSD structure.

- Over 1 Million small-molecule crystal structures
- Over 80,000 datasets deposited annually
- Structures available for anyone to download
- Links to over 1,000 journals
- Enriched and annotated by experts
- Access to data and knowledge

CCDC

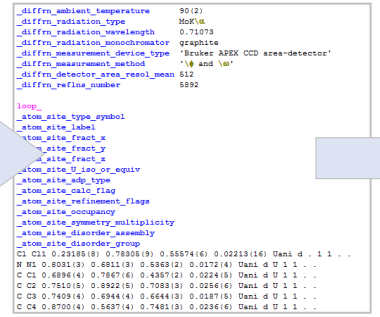
From experiments to knowledge

Experiment



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

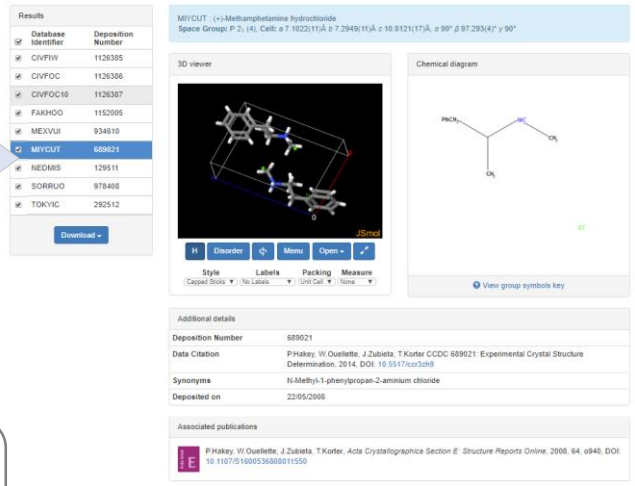
Data



```
_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       '\w and \o'
_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(4) 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 . .
```

Database



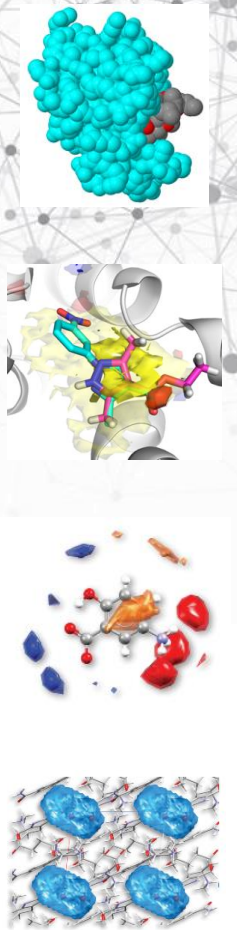
The aggregation of experimental datasets provides a foundation for resources that enable structural knowledge to be applied to scientific challenges across sectors and domains

<u>CSD-System</u>: Find, analyse and communicate crystal structures

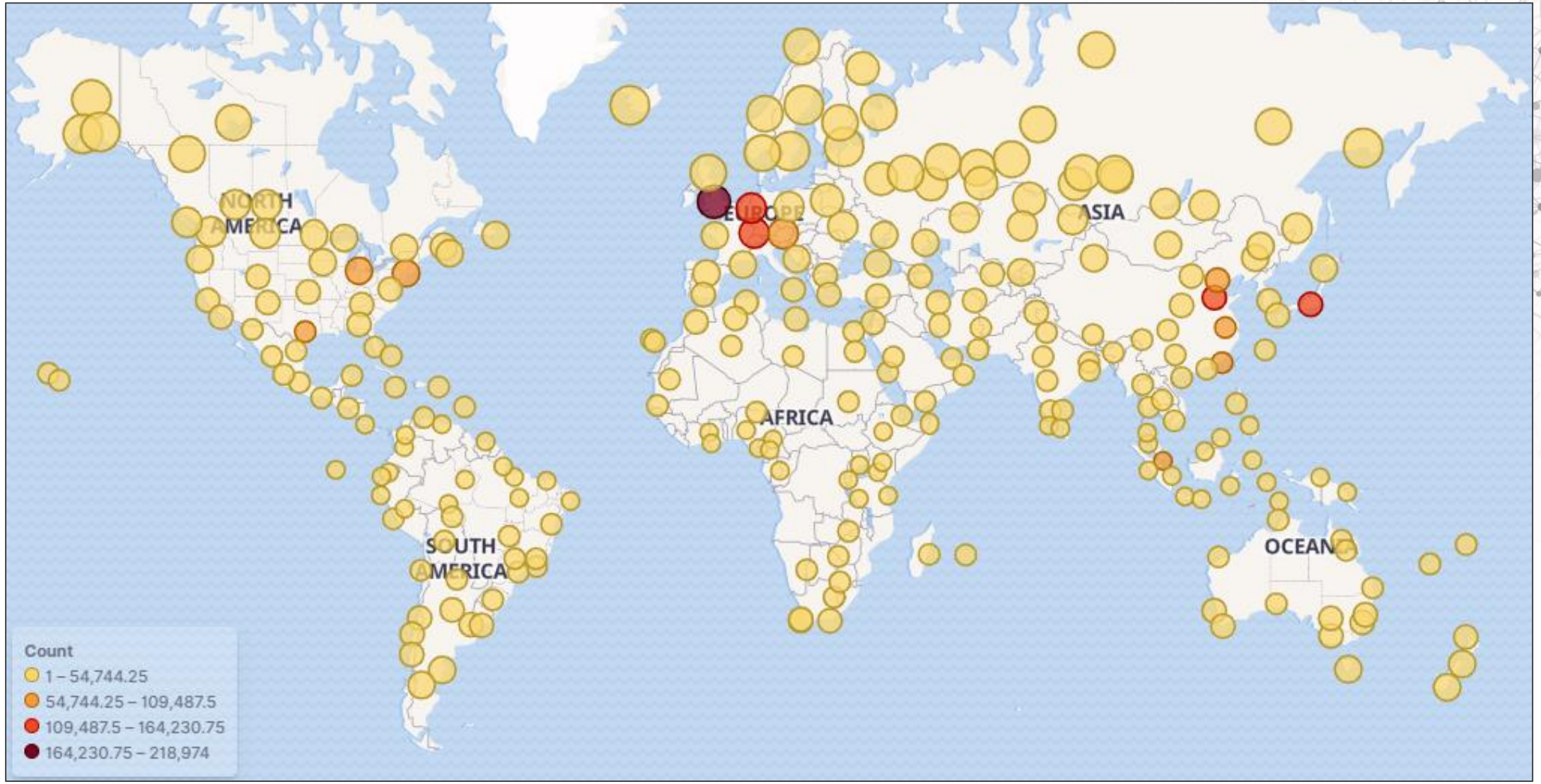
<u>CSD-Discovery</u>: Protein and ligand-based design of new drugs

<u>CSD-Materials</u>: Behaviour and properties of new materials

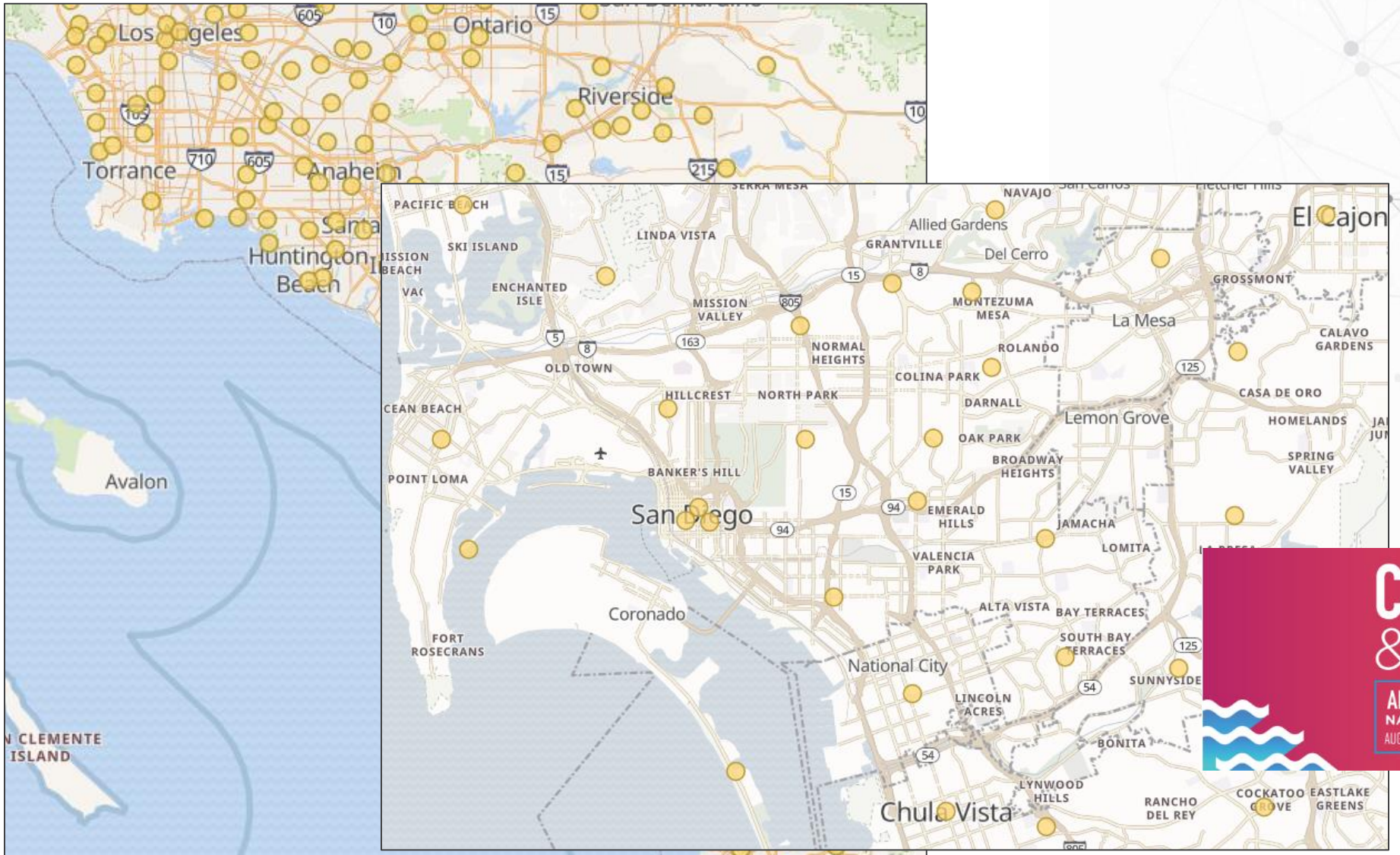
Association of chemistry and crystallography is key for enabling discovery of new insights



CCDC


ASIA
AFRICA
Count
1 – 54,744.25
54,744.25 – 109,487.5
109,487.5 – 164,230.75
164,230.75 – 218,974

CCDC



CCDC

Before electronic deposition

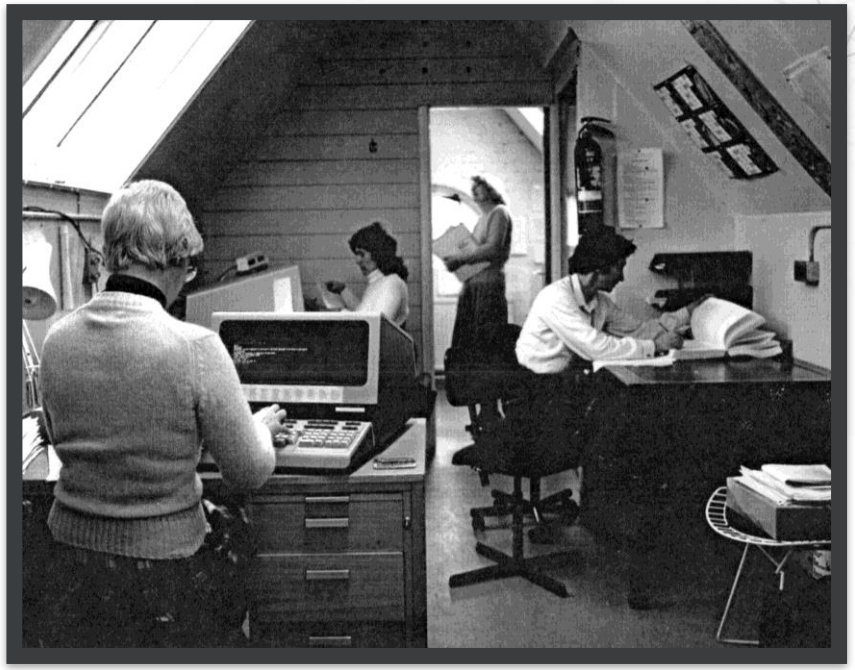
Hand-typed tables of coordinates

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	$C_{18}H_{22}CuN_6O_6$	$C_{18}H_{22}N_6NiO_6$	$C_{18}H_{22}CdN_6O_6$
M	529.01	524.18	577.88
Lattice type			
Space group			
a/Å			
b/Å			
c/Å			
α/°			
β/°			
γ/°			
U/Å^3			
Z			
D_c/g cm^{-3}			
F(000)			
μ(Mo-K_α)/cm^{-1}			
Approximate c dimensions (			
Number of set			
θ range/° (cell ((data			
h range			
k range			
l range			
Number of refl measured			
independent			
observed			
Final *R*			
Final *R*′			

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

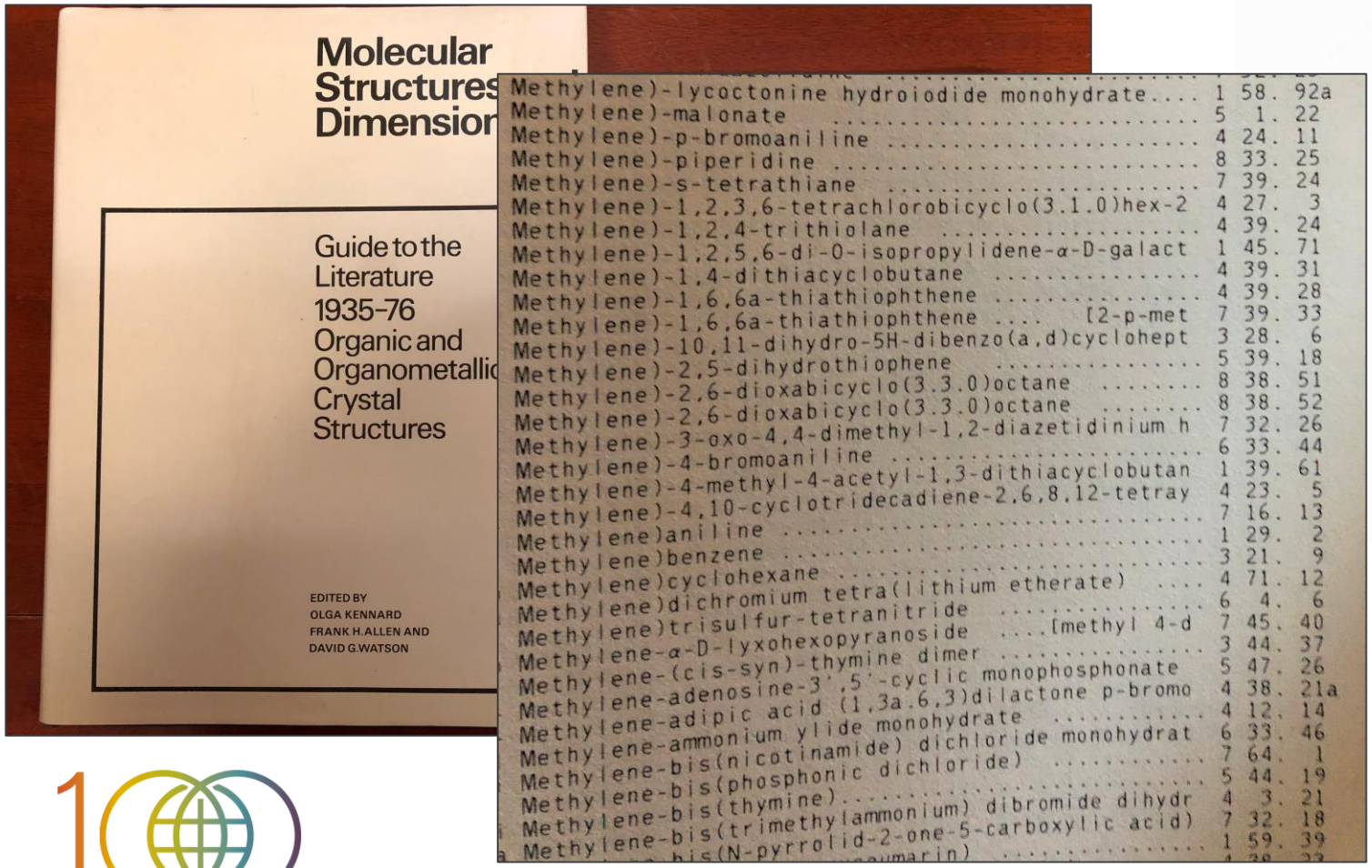
Atom	*X*/*a*	*Y*/*b*	*Z*/*c*	Atom	*X*/*a*	*Y*/*b*	*Z*/*c*
(*a*) Compound (1) (× 10^5 for Cu, × 10^4 for others)							
Cu	7 734(2)	11 519(5)	62 931(2)	C(27)	2 251(3)	671(8)	4 844(4)
N(01)	1 493(1)	−122(3)	6 454(2)	C(30)	2 044(2)	483(5)	6 905(2)
C(10)	1 359(2)	−1 289(5)	6 758(3)	N(31)	1 888(1)	1 149(3)	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)



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

CCDC

Look up of compound names



GERMANIUM, TIN, LEAD COMPOUNDS

69.43 **Tri(cyclo - octa - 1,5 - diene - platinum - di(tin trichloride)**
$C_{24}H_{36}Cl_6Pt_3Sn_2$
L.J.Guggenberger *Chem. Communic.*, 512, 1968
Also classified in 74

69.C **bis(π - Cyclopentadienyl dicarbonyl iron) di(phenylsulfonyl) tin**
$C_{26}H_{20}Fe_2O_8S_2Sn$
For complete entry see 73.110

69.44 **1,1,4,4 - Tetraphenyl - 1,4 - digermanacyclohexa - 2,5 - diene**
$C_{28}H_{24}Ge_2$
M.E.Vol'pin, V.G.Dulova, Yu.T.Struchkov, N.K.Bokiy, D.N.Kursanov
J. Organometal. Chem., **8,** 87, 1967

69.45 **1,1,4,4 - Tetraphenyl - 1,4 - digermanacyclohexa - 2,5 - diene**
$C_{28}H_{24}Ge_2$
N.G.Bokii, Yu.T.Struchkov *Zh. Strukt. Khim.*, **8,** 122, 1967

69.46 **(4 - Bromo - 1,2,3,4 - tetraphenyl - cis,cis - 1,3 - butadienyl)dimethyl tin bromide**
$C_{30}H_{26}Br_2Sn$
F.P.Boer, J.J.Flynn, H.H.Freedman, S.V.McKinley, V.R.Sandel
J. Amer. Chem. Soc., **89,** 5068, 1967

69.47 **Tricyclohexyl tin acetate**
$C_{30}H_{36}O_2Sn$
N.W.Alcock, R.E.Timms *J. Chem. Soc. (A)*, 1876, 1968

69.48 **Triphenylgermanium manganese pentacarbonyl**
$C_{35}H_{15}GeMnO_5$
B.T.Kilbourn, T.L.Blundell, H.M.Powell *Chem. Communic.*, 444, 1965
Also classified in 11

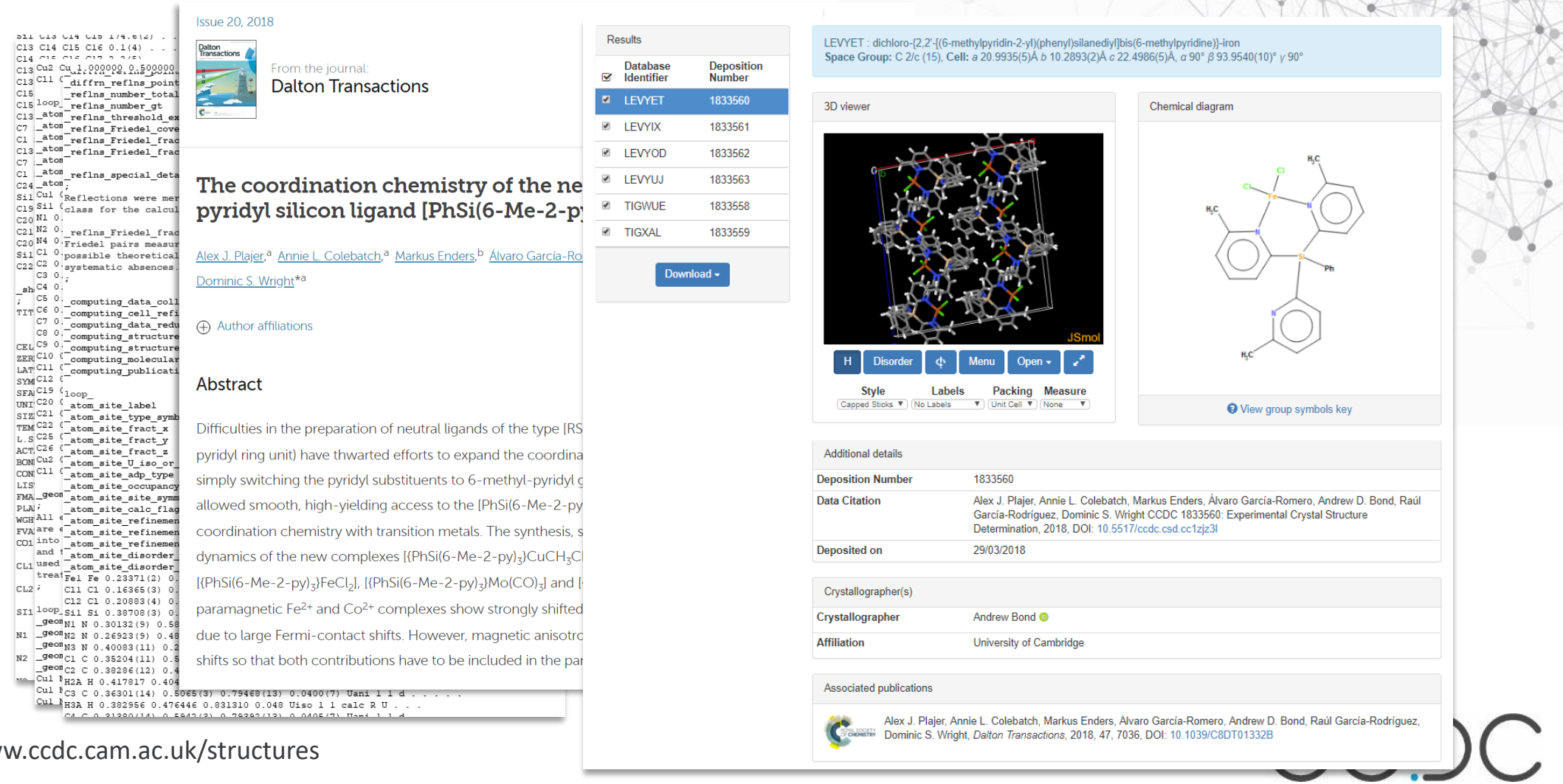
69.49 **Lead hexa - antipyrine perchlorate**
$C_{66}H_{72}N_{12}O_6Pb^{2+}$, $2ClO_4^-$
M.Vijayan, M.A.Viswamitra *Acta Cryst.*, **21,** 522, 1966

IUPAC

CCDC

Publication of crystal structures today

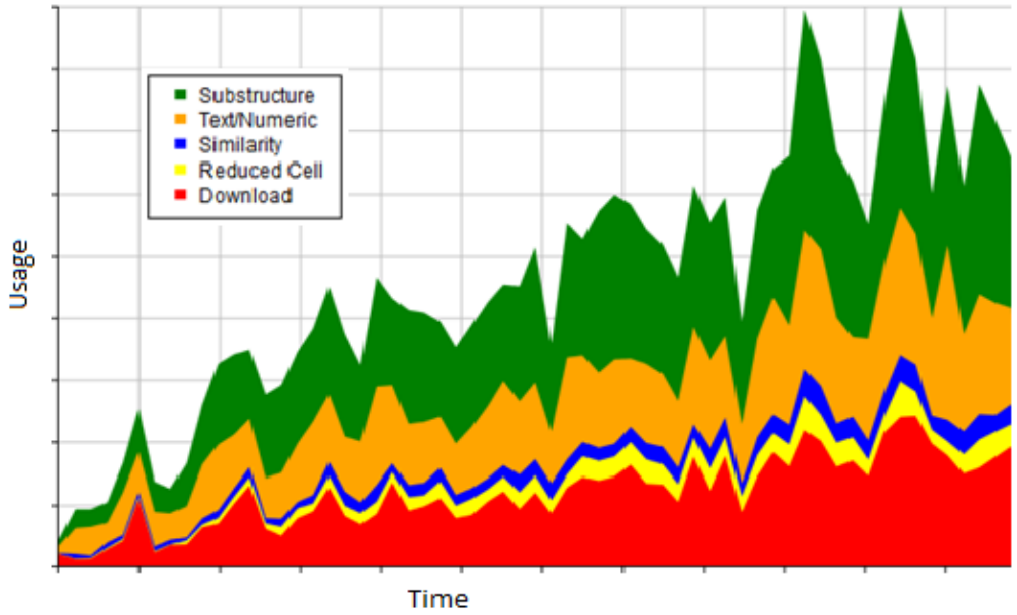
Electronic data files deposited and disseminated via the Web and linked with journal articles

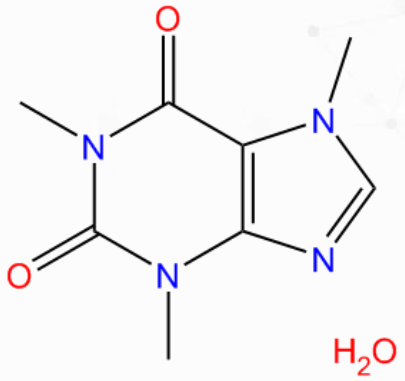


https://www.ccdc.cam.ac.uk/structures

Searching for structures

- Majority of searches of CSD are substructure searches; however:
  - 16% of all searches in WebCSD are on compound name





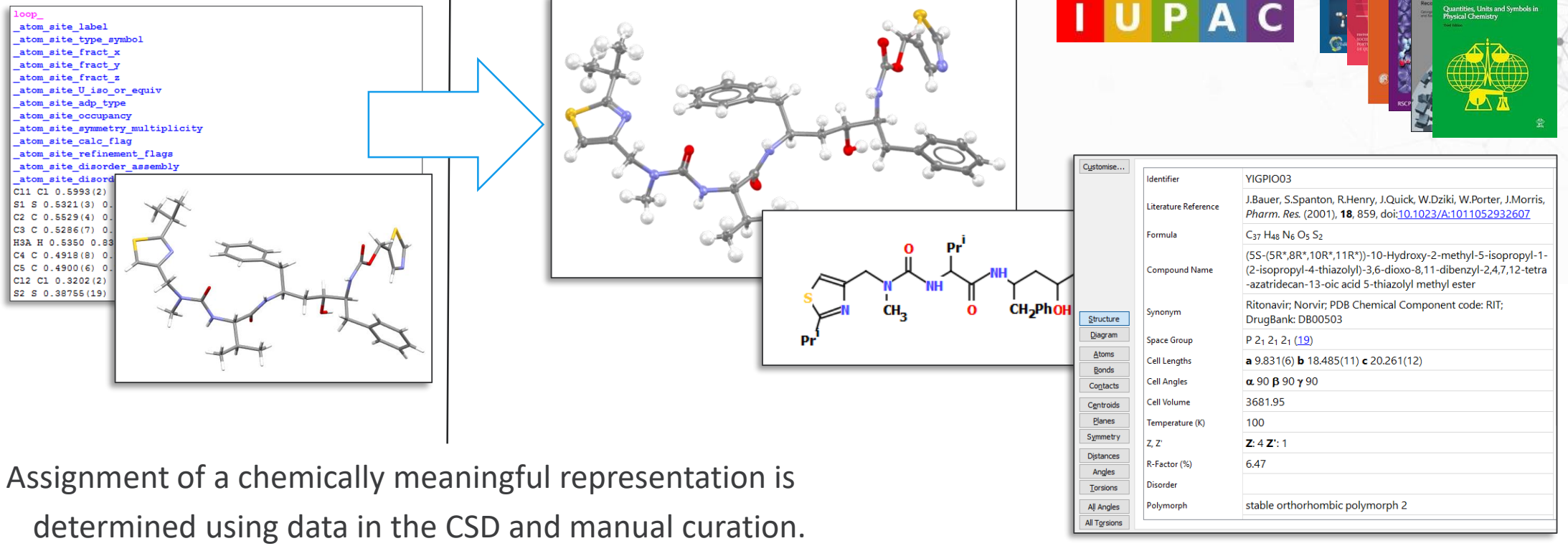
Identifier	CAFINE
Source Database	as531be
Reliability Score	★★★★ Explain score
Author(s)	D.J.Sutor
Reference	*Acta Crystallogr.* (1958), **11**, 453, doi:10.1107/S0365110X58001286
Formula	$C_8 H_{10} N_4 O_2.H_2 O$
Compound	1,3,7-Trimethyl-purine-2,6-dione monohydrate
Synonym	Caffeine monohydrate
Space Group	$P 2_1/a$

CCDC

Curation and chemistry assignment

Deposited CIF

CSD Entry

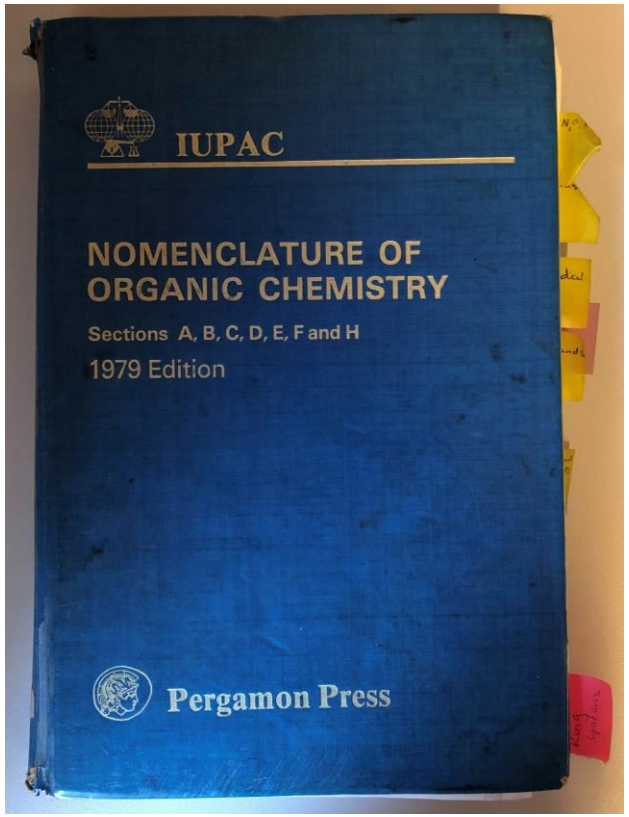


Assignment of a chemically meaningful representation is determined using data in the CSD and manual curation.

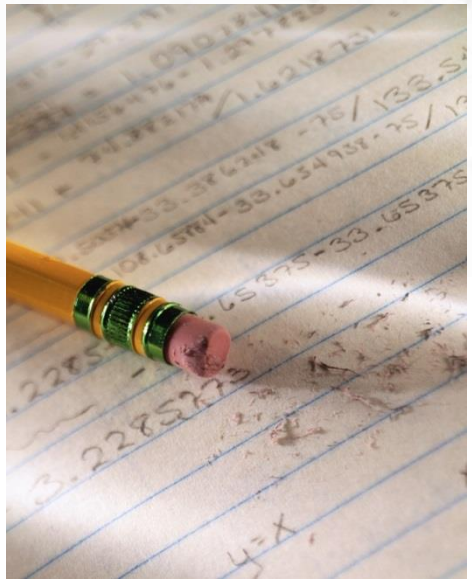
Bruno et al, Acta Crystallogr. Sect. B Struct. Sci. 67, 333–349 (2011)

CCDC

Sources of names used in the CSD



- CIF or Paper
  - Particularly helpful for capturing stereochemistry and trivial names of drugs and natural products
- Use existing entries in the CSD
- Manually construct the name
- The majority of compounds are automatically named using the naming computer software



CCDC

Using ACD/Name

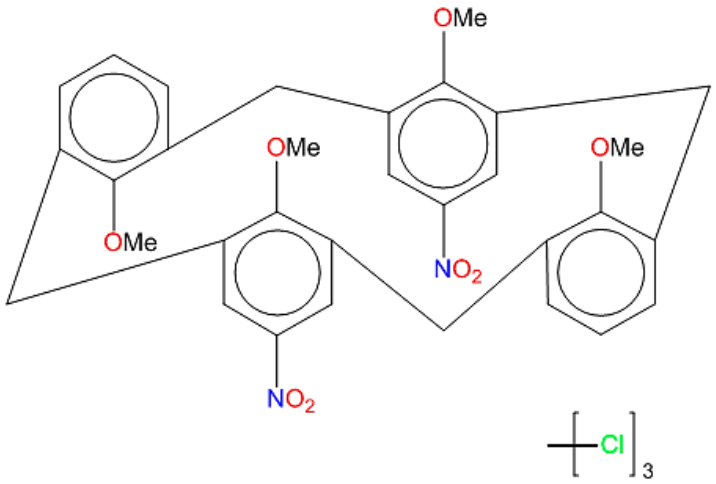
- Handles *most* organics well
- Types of difficult cases
  - Symmetry
  - Unusual valences
  - Multicomponent structures
  - Large structures
  - Coordination complexes/ polymers

Adoption of using ACD/Name

- Software speeds up the validation of structures
- CCDC Editors have been using ACD/Name to assist with naming for many years
- An early key issue was how it handled organometallics
  - 62/96 organometallics; 130/156 organics
  - overall success rate of 76%
- CCDC now uses ACD/Name to routinely generate an IUPAC name for most incoming structures

IUPAC and CSD conventions

- Generally use IUPAC name
- For ease of searching we will use semi-systematic names in compound name or synonym field e.g.; Calixarenes, Ferrocene, Cucurbits, Catenanes, Rotaxanes etc.

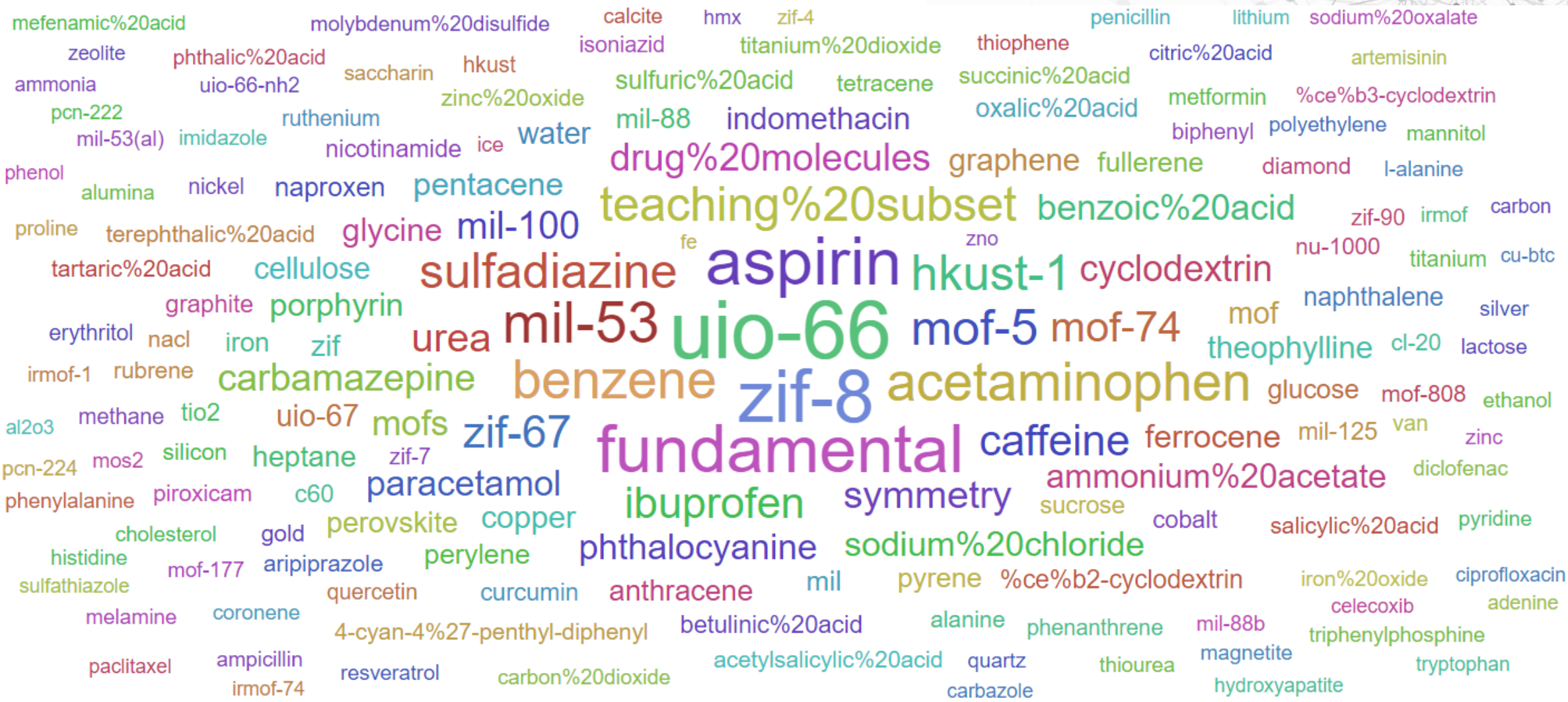


Classifiers

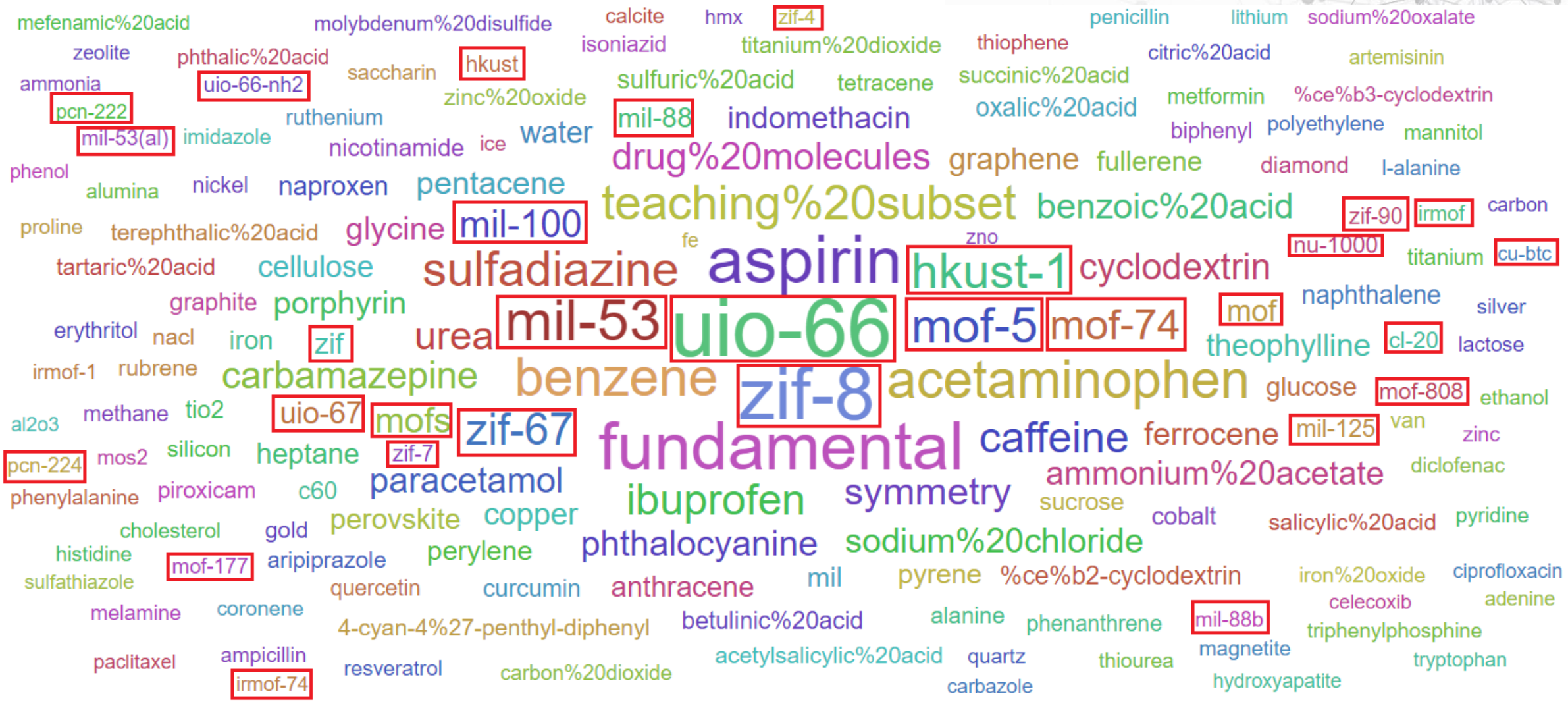
- Carbohydrate
- Nucleoside or Nucleotide
- Amino-acid, peptide and complexes
- Porphyrin, Corrin Complexes
- Steroid
- High Polymer

Compound: 25,26,27,28-Tetramethoxy-5,17-dinitropentacyclo[19.3.1.1^{3,7}.1^{9,13}.1^{15,19}]octacosa-1(25),3(28),4,6,9(27),10,12,15(26),16,18,21,23-dodecaene chloroform solvate

Synonym: 25,26,27,28-Tetramethoxy-5,17-dinitrocalix(4)arene chloroform solvate


mefenamic%20acid
molybdenum%20disulfide
calcite
hmx
zif-4
penicillin
lithium
sodium%20oxalate
zeolite
phthalic%20acid
isoniazid
titanium%20dioxide
thiophene
citric%20acid
artemisinin
ammonia
uio-66-nh2
saccharin
hkust
sulfuric%20acid
tetracene
succinic%20acid
zinc%20oxide
pcn-222
ruthenium
mil-88
indomethacin
oxalic%20acid
metformin
%ce%b3-cyclodextrin
mil-53(al)
imidazole
nicotinamide
ice
water
biphenyl
polyethylene
mannitol
phenol
drug%20molecules
graphene
fullerene
diamond
l-alanine
alumina
nickel
naproxen
pentacene
teaching%20subset
benzoic%20acid
zif-90
irmof
carbon
proline
terephthalic%20acid
glycine
mil-100
fe
zno
nu-1000
tartaric%20acid
cellulose
sulfadiazine
aspirin
hkust-1
cyclodextrin
titanium
cu-btc
graphite
porphyrin
mil-53
uio-66
mof-5
mof-74
mof
naphthalene
silver
erythritol
nacl
iron
zif
urea
theophylline
cl-20
lactose
irmof-1
rubrene
carbamazepine
benzene
zif-8
acetaminophen
glucose
mof-808
ethanol
al2o3
methane
tio2
uio-67
mofs
zif-67
fundamental
caffeine
ferrocene
mil-125
van
zinc
pcn-224
mos2
silicon
heptane
zif-7
ammonium%20acetate
diclofenac
phenylalanine
piroxicam
c60
paracetamol
ibuprofen
symmetry
sucrose
cholesterol
gold
perovskite
copper
cobalt
salicylic%20acid
pyridine
histidine
mof-177
aripiprazole
perylene
phthalocyanine
sodium%20chloride
sulfathiazole
quercetin
curcumin
anthracene
mil
pyrene
%ce%b2-cyclodextrin
iron%20oxide
ciprofloxacin
celecoxib
adenine
melamine
coronene
4-cyan-4%27-penthyl-diphenyl
betulinic%20acid
alanine
phenanthrene
mil-88b
triphenylphosphine
paclitaxel
ampicillin
resveratrol
carbon%20dioxide
acetylsalicylic%20acid
quartz
thiourea
magnetite
tryptophan
irmof-74
carbazole
hydroxyapatite

CCDC


mefenamic%20acid
molybdenum%20disulfide
calcite
hmx
zif-4
penicillin
lithium
sodium%20oxalate
zeolite
phthalic%20acid
isoniazid
titanium%20dioxide
thiophene
citric%20acid
artemisinin
saccharin
hkust
sulfuric%20acid
tetracene
succinic%20acid
ammonia
uio-66-nh2
zinc%20oxide
metformin
%ce%b3-cyclodextrin
pcn-222
ruthenium
mil-88
indomethacin
oxalic%20acid
biphenyl
polyethylene
mannitol
mil-53(al)
imidazole
water
nicotinamide
ice
drug%20molecules
graphene
fullerene
diamond
l-alanine
phenol
alumina
nickel
naproxen
pentacene
teaching%20subset
benzoic%20acid
zif-90
irmof
carbon
proline
terephthalic%20acid
glycine
mil-100
fe
zno
nu-1000
titanium
cu-btc
tartaric%20acid
cellulose
sulfadiazine
aspirin
hkust-1
cyclodextrin
graphite
porphyrin
mof
naphthalene
silver
erythritol
nacl
iron
zif
urea
mil-53
uio-66
mof-5
mof-74
theophylline
cl-20
lactose
irmof-1
rubrene
carbamazepine
benzene
zif-8
acetaminophen
glucose
mof-808
ethanol
al2o3
methane
tio2
uio-67
mofs
zif-67
fundamental
caffeine
ferrocene
mil-125
van
zinc
pcn-224
mos2
silicon
heptane
zif-7
ammonium%20acetate
diclofenac
phenylalanine
piroxicam
c60
paracetamol
ibuprofen
symmetry
sucrose
cholesterol
gold
perovskite
copper
cobalt
salicylic%20acid
pyridine
histidine
mof-177
aripiprazole
perylene
phthalocyanine
sodium%20chloride
sulfathiazole
quercetin
curcumin
anthracene
mil
pyrene
%ce%b2-cyclodextrin
iron%20oxide
ciprofloxacin
celecoxib
adenine
melamine
coronene
betulinic%20acid
alanine
phenanthrene
mil-88b
triphenylphosphine
4-cyan-4%27-penthyl-diphenyl
magnetite
paclitaxel
ampicillin
acetylsalicylic%20acid
quartz
thiourea
tryptophan
resveratrol
carbon%20dioxide
irmof-74
carbazole
hydroxyapatite

CCDC

Challenges

- Providing consistent chemical representations
- Large complicated structures
- Polymeric structures
- New types of structures
- Changes to the rules for naming compounds
- Polymorphs and stereoisomers
- New nomenclature definitions for naming groups of structures

CCDC

Reliable input representations

- How best to reliably represent organometallics?
  - dative vs covalent bonds?
  - explicit hydrogens/valencies?
  - dummy atoms?
  - zero-order bonds?

MOL V3000

```
1 = single
2 = double
3 = triple
9 = coordination
10 = hydrogen
```

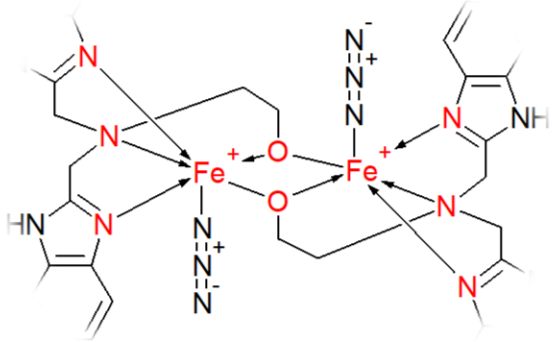
* excluding query bond types

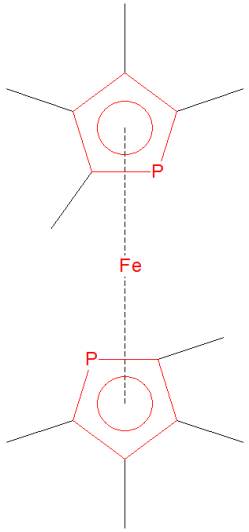
PubChem SDF
PUBCHEM_NONSTANDARDBOND

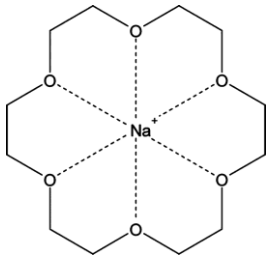
```
1 Single Bond
2 Double Bond
3 Triple Bond
4 Quadruple Bond
5 Dative Bond
6 Complex Bond
7 Ionic Bond
```

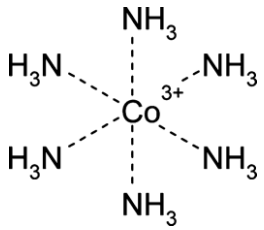
ACD/Labs MOL V2000 Extensions

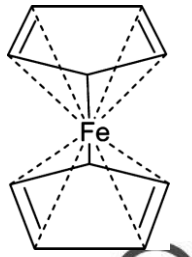
```
M  ZZF  3   1  41   2  42   3  43
M  ZZH  1   5   2   3   4   5   6
M  ZZH  2   5   7   8   9  10  11
M  ZZH  3   5  12  14  15  16  17
M  ZZE  2  42  18  43  18
```







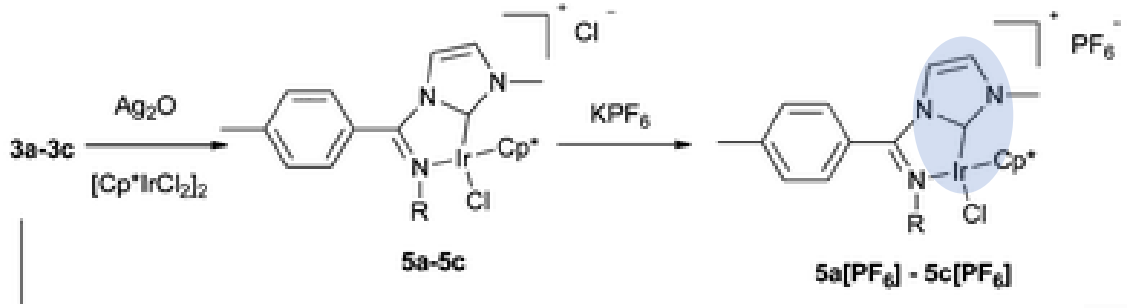




CCDC

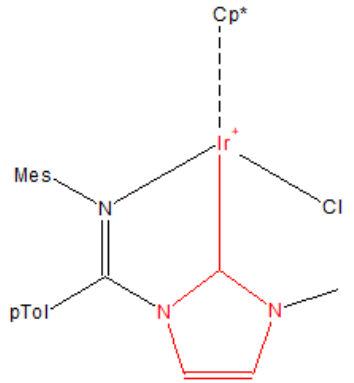
Consistent structure representation

CSD: Representations of Imidazolin-2-ylidene metal carbenes in the CSD 5.38 (N = 9468)



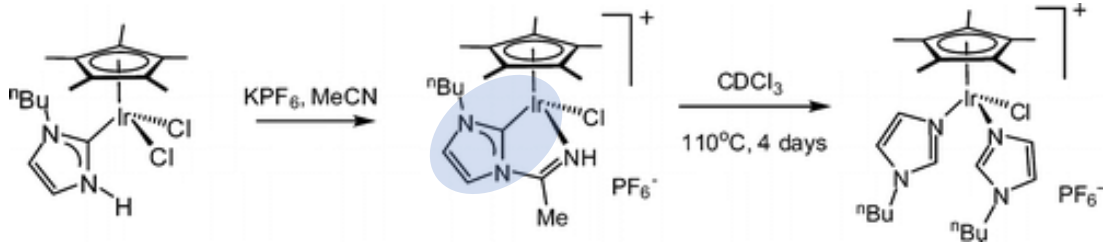
Dalton Trans., 2012,41, 14557-14567, doi:10.1039/C2DT31989F

CSD

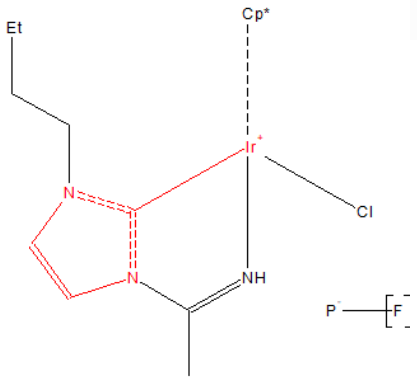


85%

ECIWUK CCDC:872879 10.5517/ccy99dx



Organometallics 2007, 26, 18, 4684-4687, doi:10.1021/om700498w



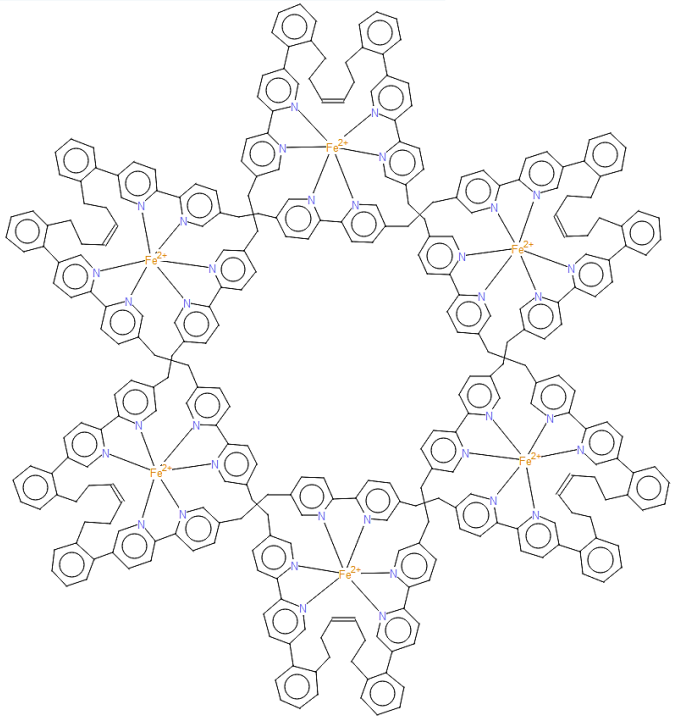
12%

LIMXAH CCDC:664254 10.5517/ccq96kr

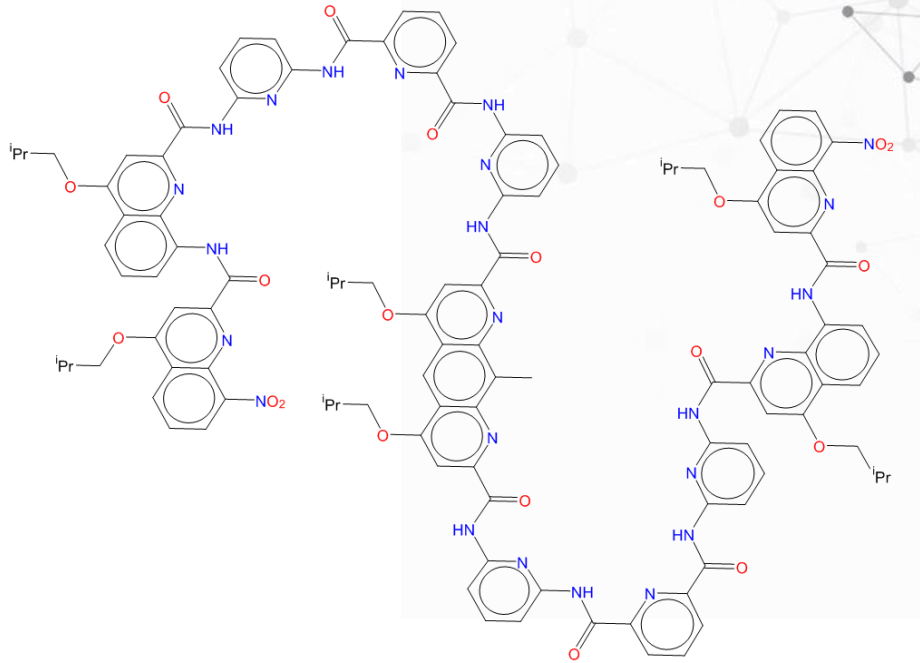
CCDC

Challenges – Cyclic and large structures

[2]-bis(μ⁶-2,6,10,20,24,28,38,42,46(5,2),3,7,11,21,25,29,39,43,47(2,5)-octadecapyridina-1,12,19,30,37,48(1,2)-hexabenzenacyclotetrapentacontaphane-15,33,51-triene)-hexa-iron-catenane



Star of David catenane



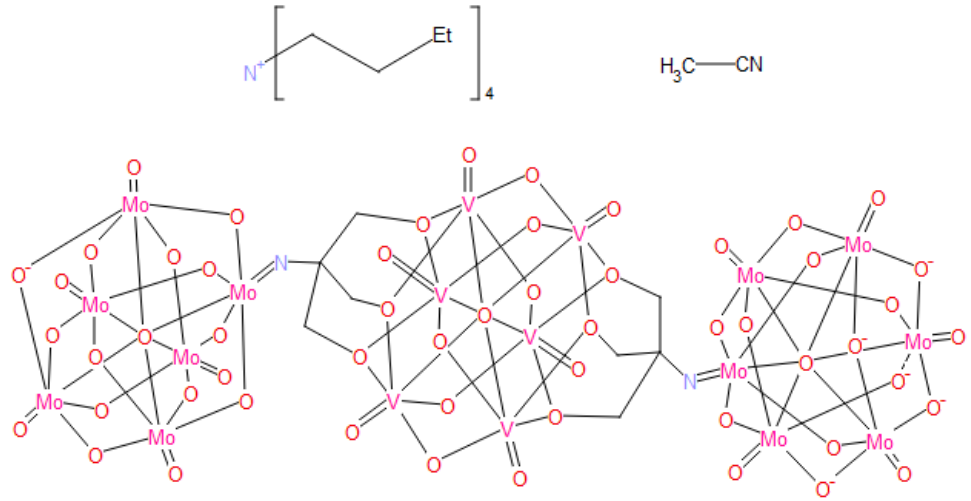
N,N'-bis(6-(6-(6-(8-(4-Isobutoxy-8-nitroquinoline-2-carboxamido)-4-isobutoxyquinoline-2-carboxamido)pyridine-2-carbamoyl)pyridine-2-carboxamido)pyridine-2-yl) 1,8-diaza-4,5-di-isobutoxy-9-methyl-2,7-anthracenedicarboxamide

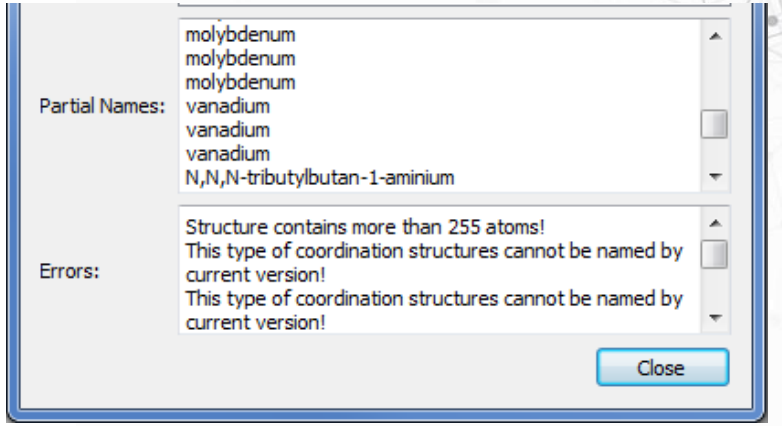
Difficulty naming structures with more than 15 rings or 255 atoms

DONCUG, BOBFEE

CCDC

Challenges - Clusters

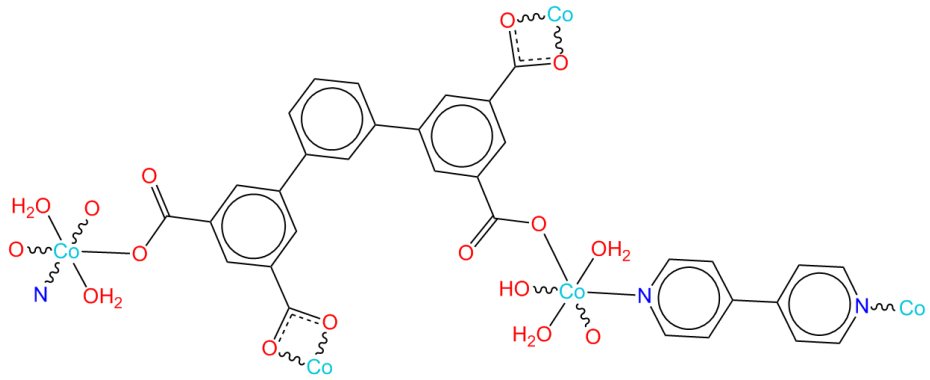




hexakis(tetra-n-butylammonium) tris(μ^6-oxido)-bis(μ^4-(tris(oxidomethyl)methyl)imino)-tricontakis(μ^{2}oxido)-hexadecakis(aqua)-dodeca-molybdenum-hexa-vanadium acetonitrile solvate

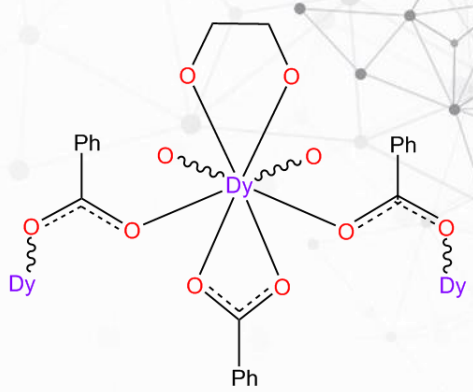
Challenges - Polymers

- To name polymers we
  - break the compound down into components of ligands and metals
  - manually construct the name



Partial name matches <<<<< ... 1,1':3',1''-terphenyl-3,3'',5,5''-tetracarboxylic acid
Partial name matches <<<<< ... 4,4'-bipyridine
Partial name matches <<<<< ... water
Partial name matches <<<<< ... ammonia

catena-[(μ_4-1,1':3',1''-terphenyl-3,3'',5,5''-tetracarboxylato)-(μ_2-4,4'-bipyridine)-tetra-aqua-di-zinc]

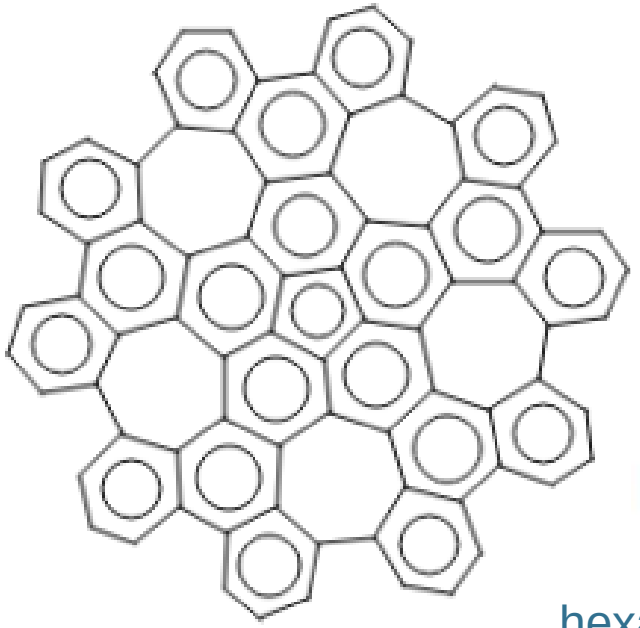


catena-[bis(μ_2-benzoato)-(benzoato)-(ethane-1,2-diol)-dysprosium]

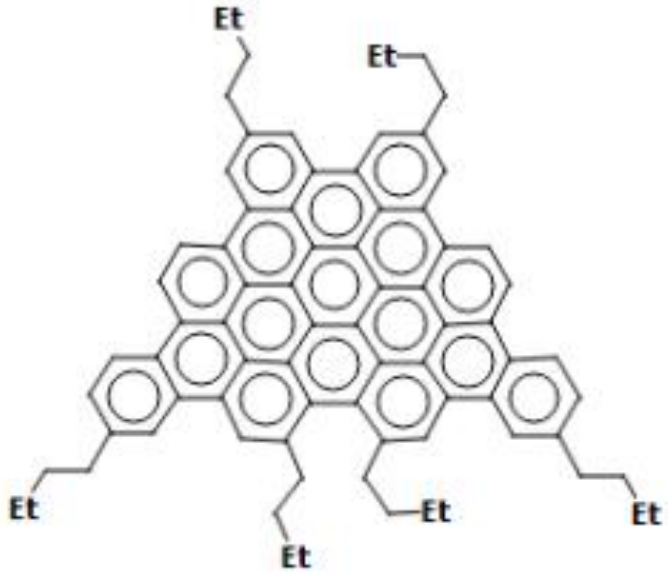
- ACD occasionally attempts to create a whole name, however it doesn't always
  - recognise polymer bonds
  - give stoichiometry of ligands or anions

New types of structures

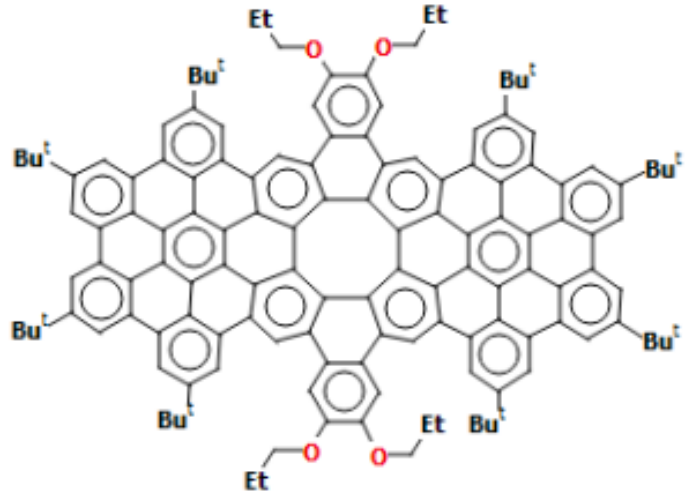
- Graphene-type structures



Nanographene



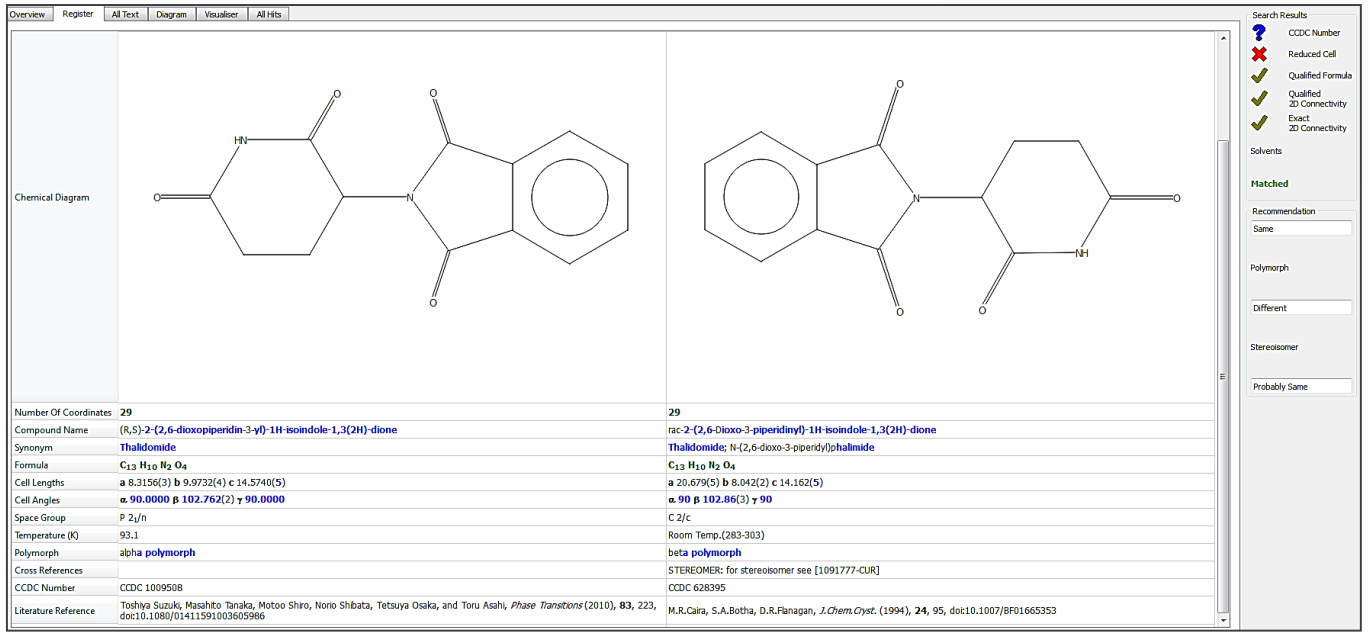
hexa-n-butyl-C54-nanographene



tetrakis(n-propoxy)-octakis(t-butyl)-nanographene

Stereochemistry

- CSD add stereochemical descriptors if given by the author in the cif or paper
- Group families together of same compound
- Link by cross-referencing diastereomers/enantiomers, but not same group

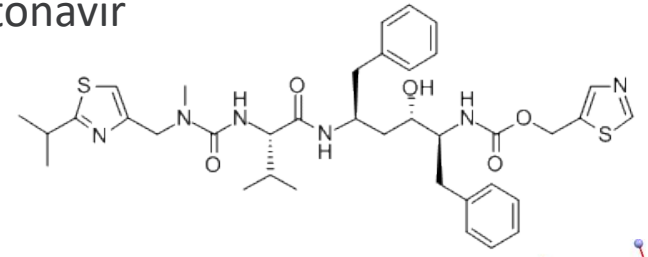


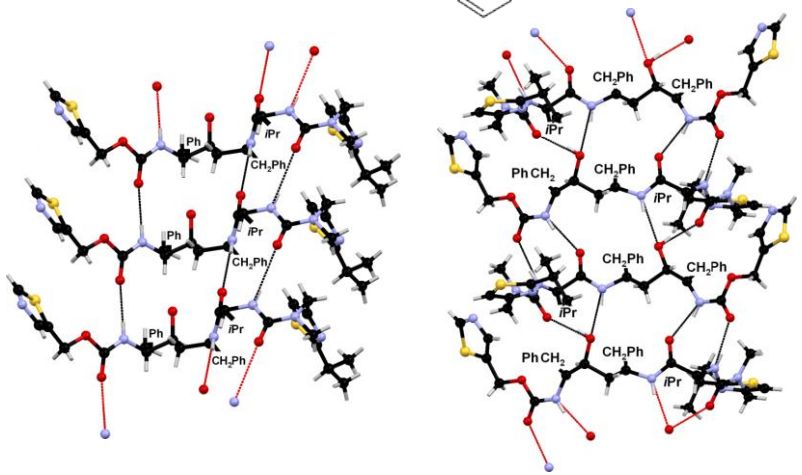
DOHQEY, DOHQAU

Polymorphism

Same tautomer – different crystal forms – different interactions – different stability

Ritonavir





YIGPIO

Metastable monoclinic polymorph 1

YIGPIO01

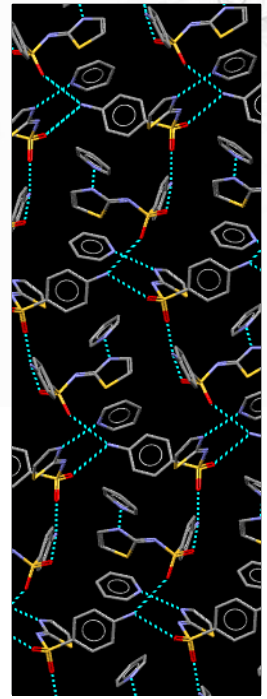
Stable orthorhombic polymorph 2

ACRDIN
polymorph II
polymorph III
polymorph IV
polymorph V
polymorph VI
polymorph VII

PAMMNC
α polymorph
β polymorph
γ polymorph
δ polymorph
ε polymorph
ζ polymorph

CUIMDZ
blue J I2/a polymorph
blue B trigonal polymorph
green orthorhombic polymorph
olive C2/c polymorph

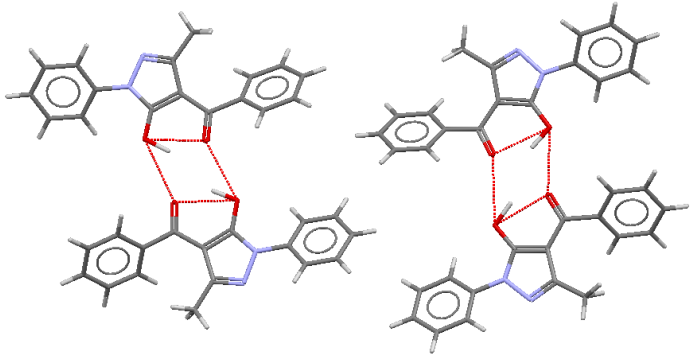
ADEDIX
monoclinic disappearing polymorph
tetragonal polymorph



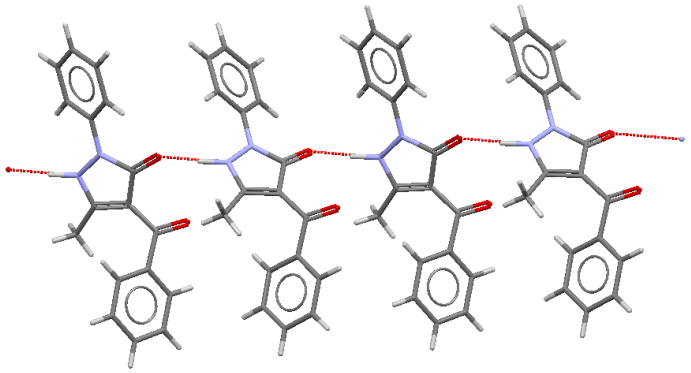
CCDC

Tautomeric polymorphism

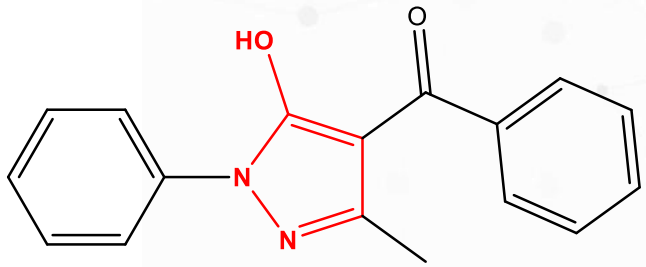
Different tautomers – different crystal forms – different interactions – different stability



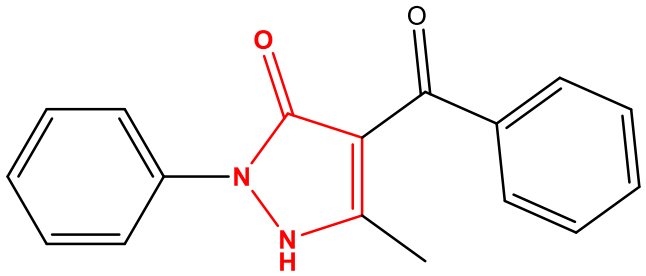
YUYDOL



DEBFAR



Less stable tautomers can sometimes form more stable hydrogen bonded networks

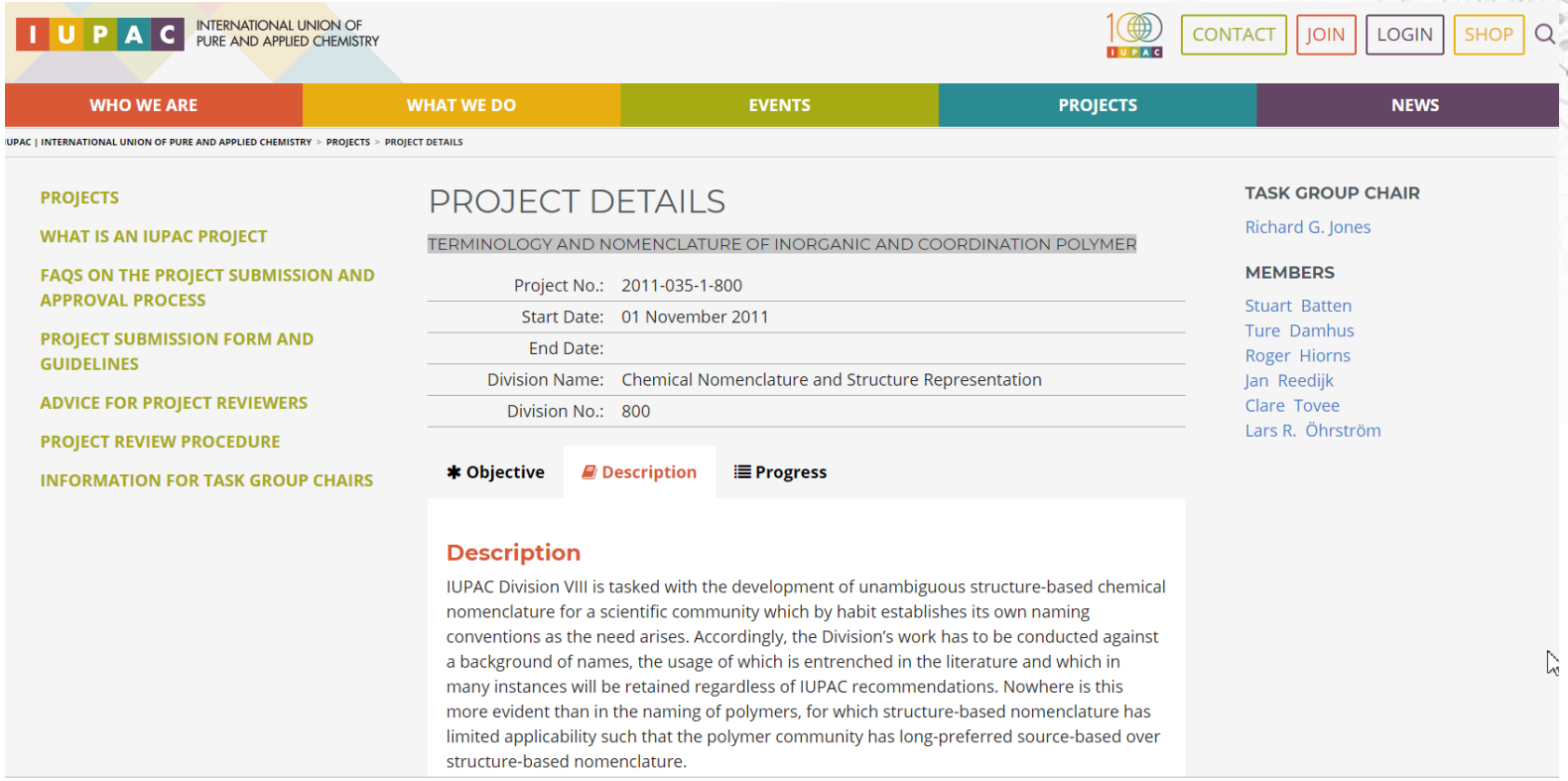


A J Cruz-Cabeza, C Groom, CrystEngComm , 2011, 13, 93

CCDC

Changes in naming recommendations

- Naming coordination polymers



https://iupac.org/projects/project-details/?project_nr=2011-035-1-800

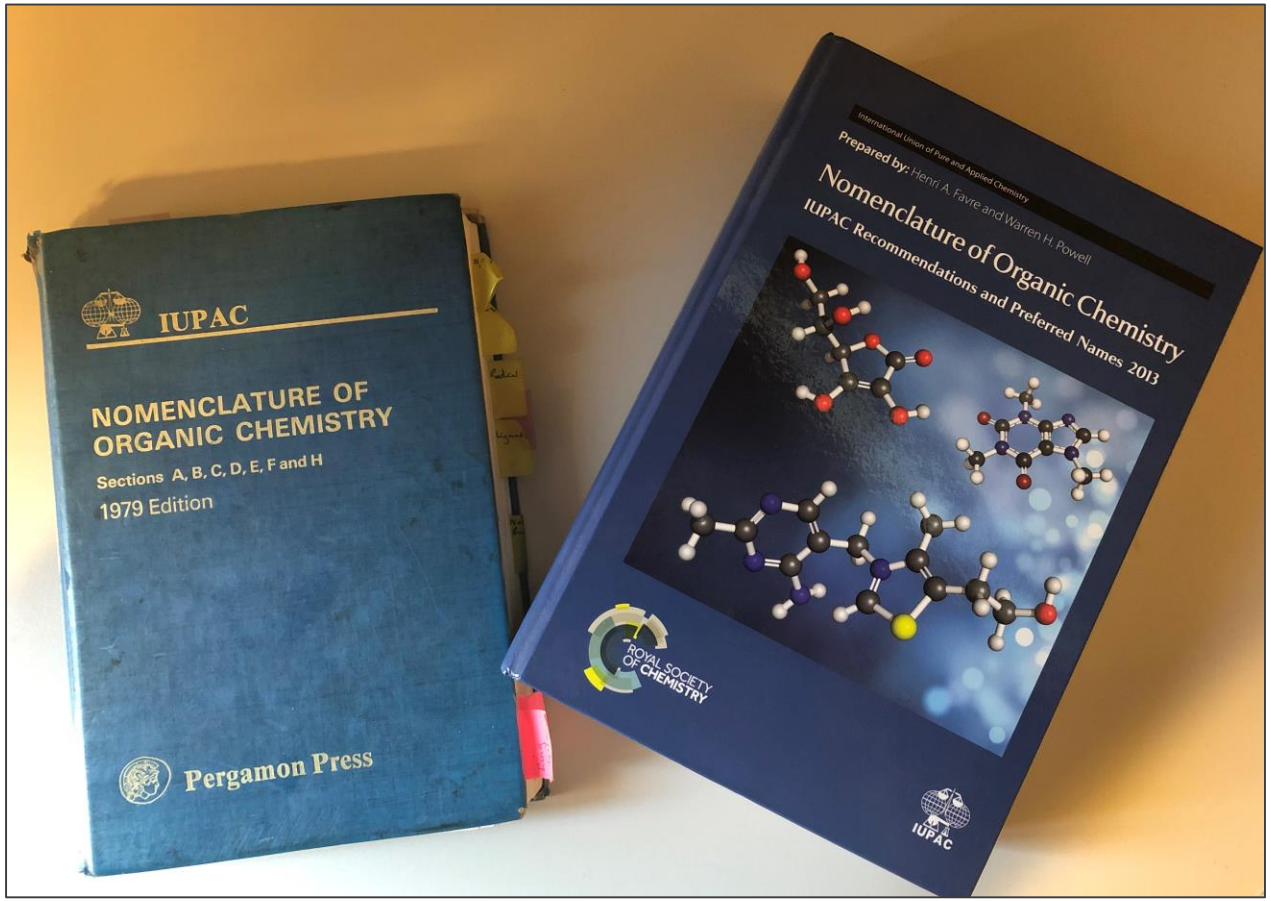
CCDC

Changes in naming recommendations

- Challenges with naming coordination polymers
  - Consistency between naming polymers and coordination complexes
  - How to deal with changes to the recommendations that affect existing entries
  - How to deal with indicating deprotonation of ligands in multiple positions
  - Organic polymers that are cross-linked with coordination

https://iupac.org/projects/project-details/?project_nr=2011-035-1-800

Changes in nomenclature rules



CHANGES FROM THE 1979 EDITION, THE 1993 GUIDE, AND OFFICIAL PUBLICATIONS FROM 1993 THROUGH 2002 INCLUDED IN THE 2013 EDITION OF THE IUPAC *NOMENCLATURE OF ORGANIC CHEMISTRY*

1. Scope of the recommendations in the 2013 edition

(a) The elements Al, Ga, In, and Tl are added to the elements that were included in the recommendations in the 1979 edition (ref. 1) and the 1993 Guide (ref. 2)

2. Skeletal replacement ('a') nomenclature

(a) Heteroatoms in chains subject to skeletal replacement ('a') nomenclature are considered to be an integral part of the parent hydride and as detachable prefixes they have seniority over suffixes for numbering; thus, heteroacyclic chains subject to skeletal replacement ('a') nomenclature are now treated the same as heterocyclic rings.

(b) The hetero atoms P, As, Sb, Bi, Si, Ge, Sn, Pb, B, Al, Ga, In, and Tl can now terminate a heteroacyclic chain that is subject to skeletal replacement ('a') nomenclature; in previous recommendations, a heteroacyclic chain that is subject to skeletal replacement ('a') nomenclature had to terminate with carbon atoms.

(c) Skeletal replacement ('a') prefixes ending in 'ata', for example 'borata', are no longer recognized.

(d) Groups of heteroatoms having a single multivalent name are considered as a unit; hence the term 'heterounit' includes both heteroatoms and such heteroatom groups. Such heteroatom groups were not considered as a single heterounit in previous recommendations.

(e) Adapting skeletal replacement ('a') prefixes for elements of Groups 1-12 for use in skeletal replacement nomenclature is a major change even though names for such organometallic compounds involving these elements are only preselected at this time.

3. Substitutive nomenclature

(a) Substitutive nomenclature is the preferred method of nomenclature, except for anhydrides, esters and salts, acid halides and pseudohalides for which

https://doi.org/10.1039/9781849733069

CCDC

Changes in nomenclature rules

(f) The prefixes 'ureido' and 'ureylene' are not used in preferred IUPAC names. The prefixes 'carbamoylamino' and 'carbonylbis(azanediyl)', respectively, are recommended for preferred IUPAC names and in general nomenclature.

(g) The prefix 'guanidino' is no longer acceptable in preferred IUPAC names; the preferred prefix is ''carbonimidamido'.

(h) The prefix 'amidino' is no longer acceptable in preferred IUPAC names; the preferred prefix is 'carbamimidoyl'.

(i) The prefix '*aci*-nitro' for HO-N(O)= is no longer acceptable for preferred IUPAC names. The preferred prefix is 'hydroxy(oxo)-λ^5-azanylidene'; the prefix recommended in the 1993 Guide (ref. 2) 'hydroxynitroryl' is not acceptable in the context of these recommendations where two free valences must be expressed by the correct 'ylidene' or 'diyl' type.

(j) Preselected prefixes derived from the preselected parent hydride hydrazine are now formed systematically from hydrazine: 'hydrazinyl' for H_2N-NH–; 'hydrazinylidene' for H_2N-N=; 'hydrazinediylidene' for =N-N=; and hydrazine-1,2-diyl for -NH-NH–. The prefixes 'hydrazino', 'hydrazono', 'azino' and 'hydrazo', respectively, are no longer acceptable, even for general nomenclature.

(k) The prefixes 'benzyl' ...

https://doi.org/10.1039/9781849733069

CCDC

Changes in nomenclature rules

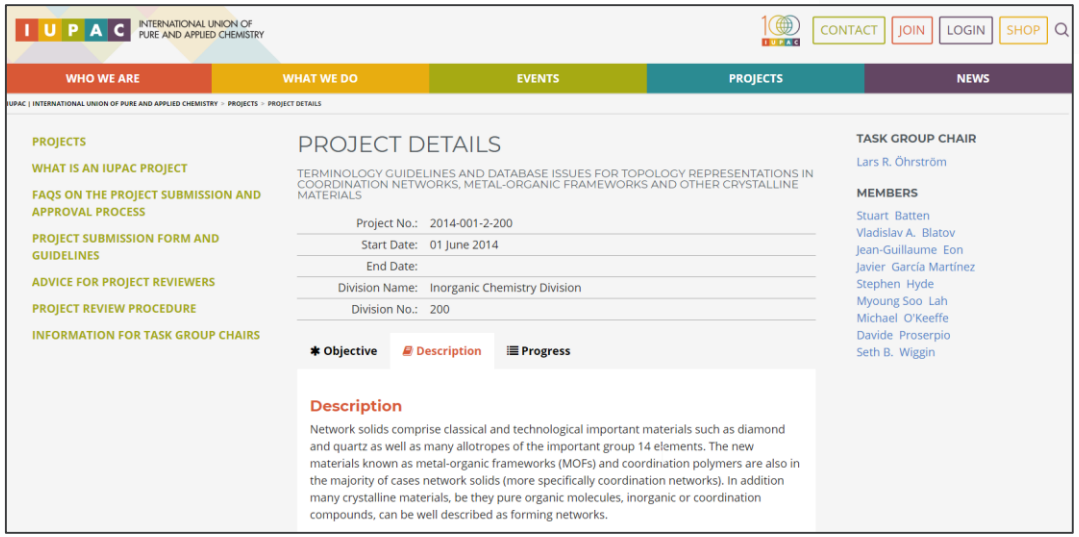
- CSD 5.40 + 2 updates
- Ureido – 449 hits (Carbamoylamino – 54 hits)
- Ureylene – 33 hits (Carbonylbis(azanediyl) – 5 hits)
- Guanidino – 189 hits (Carbonimidamido – 0 hits)
- Amidino – 167 hits (Carbamimidoyl – 315 hits)

https://doi.org/10.1039/9781849733069

CCDC

Naming of topologies for Metal Organic Frameworks

- IUPAC task force
  - Challenge is to come up with a reliable and consistent topology
  - Different topologies for the same structure can be assigned based on the method used
  - Task force looking at trying to gain a community consensus on the best approach



CCDC

Summary – the future

- Naming of compounds will continue to be really important
- One of the main challenges is being consistent
  - Consistent between old and new entries
  - Consistent between groups of entries

- IUPAC continues to be very important in tackling these challenges

CCDC

Thank You

- CINF: One Million Crystal Structures: A Wealth of Structural Chemistry Knowledge
- Tuesday Aug 27 1.30 PM – 5.20 PM
- Wednesday Aug 28 9.00 AM – 12.15 PM
- Wednesday Aug 28 1.30 PM – 4.40 PM
- Grand Ballroom D, Omni San Diego Hotel

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

@ccdccambridge

ccdc.cambridge

CCDC