

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



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



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

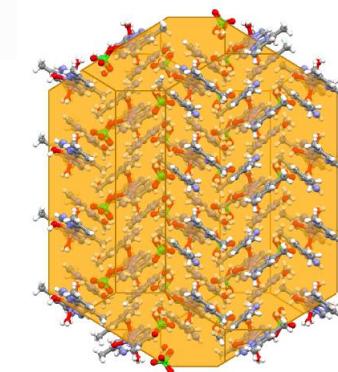
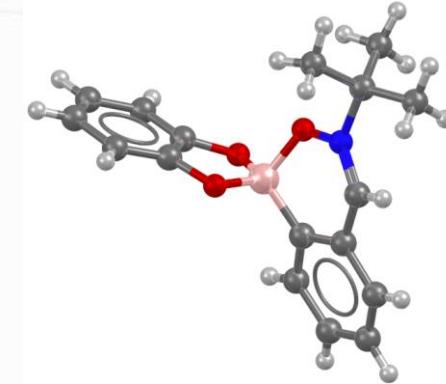
Originated in 1965

Financially self-supporting

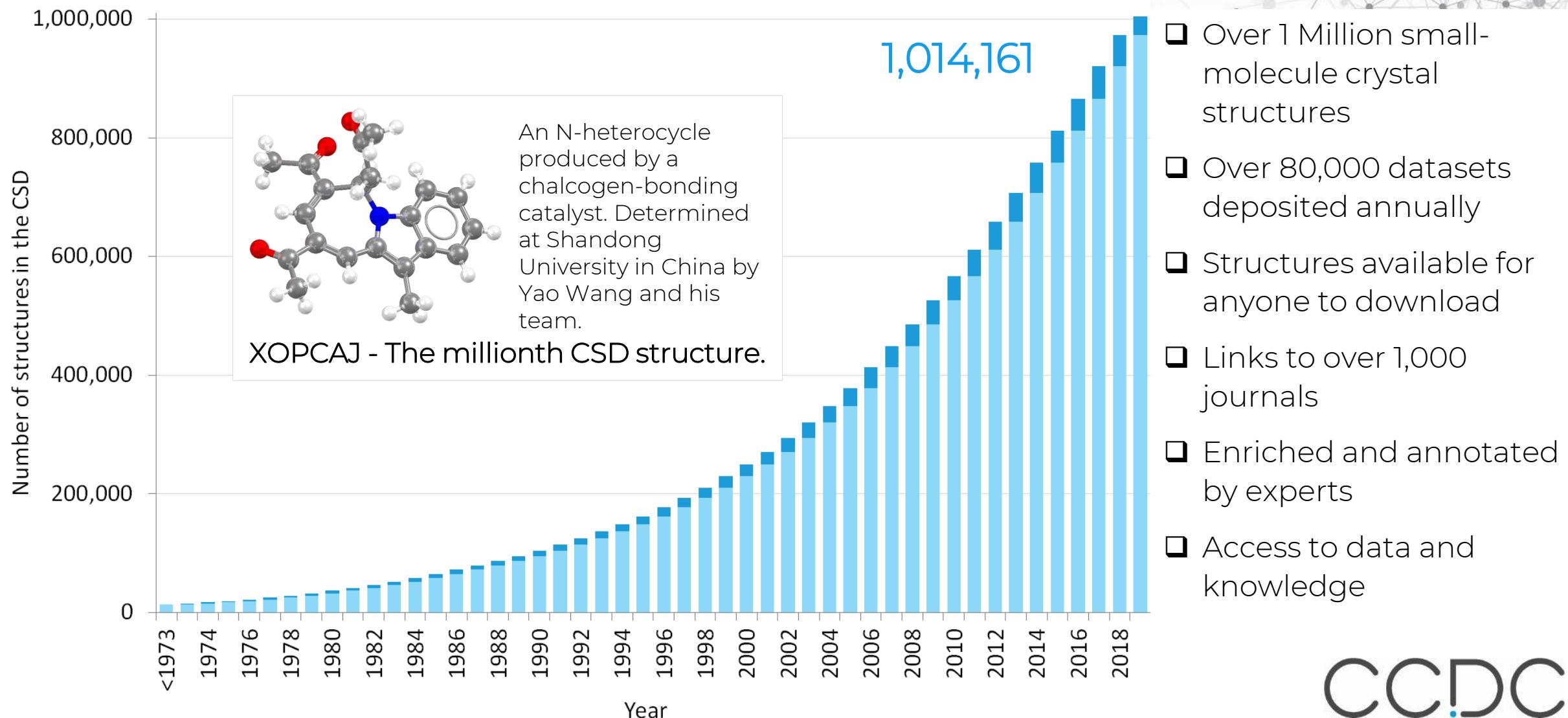
Not-for-profit, UK Registered Charity

University Partner Institute

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

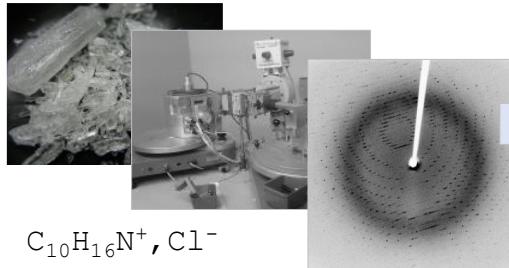


The Cambridge Structural Database (CSD)

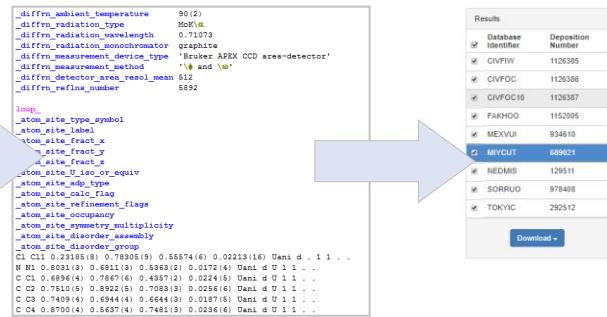


From experiments to knowledge

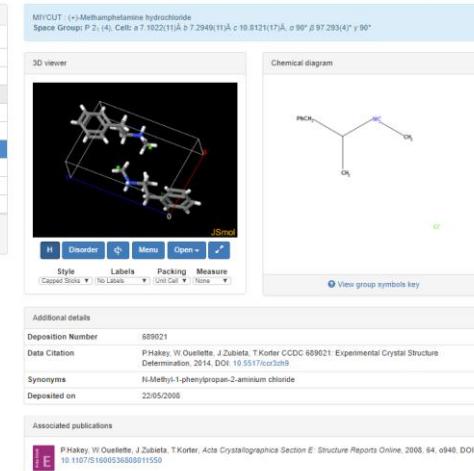
Experiment



Data



Database



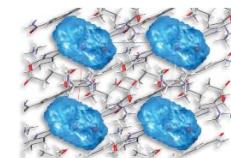
The aggregation of experimental datasets provides a foundation for resources that enable structural knowledge to be applied to scientific challenges across sectors and domains

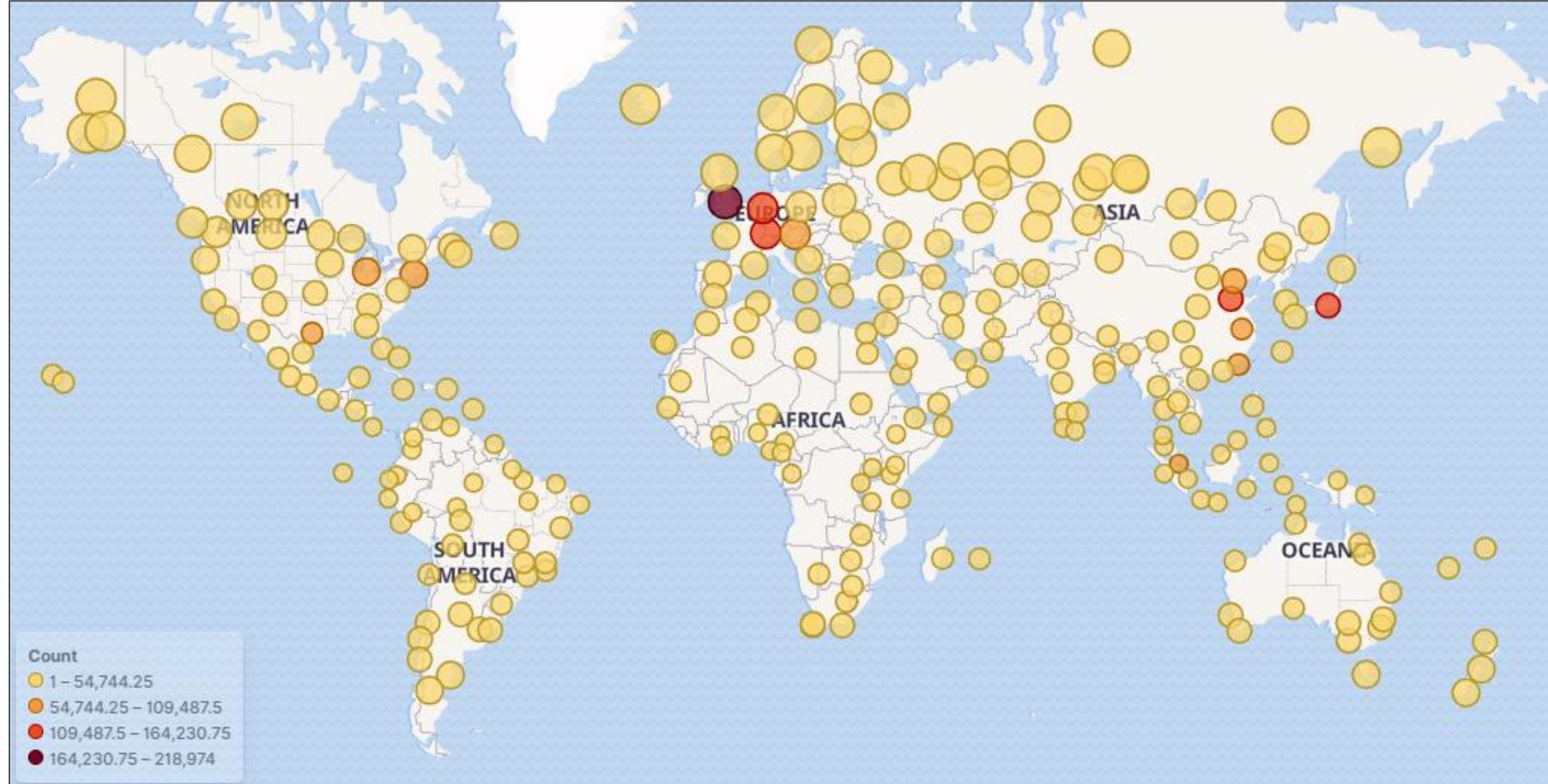
CSD-System: Find, analyse and communicate crystal structures

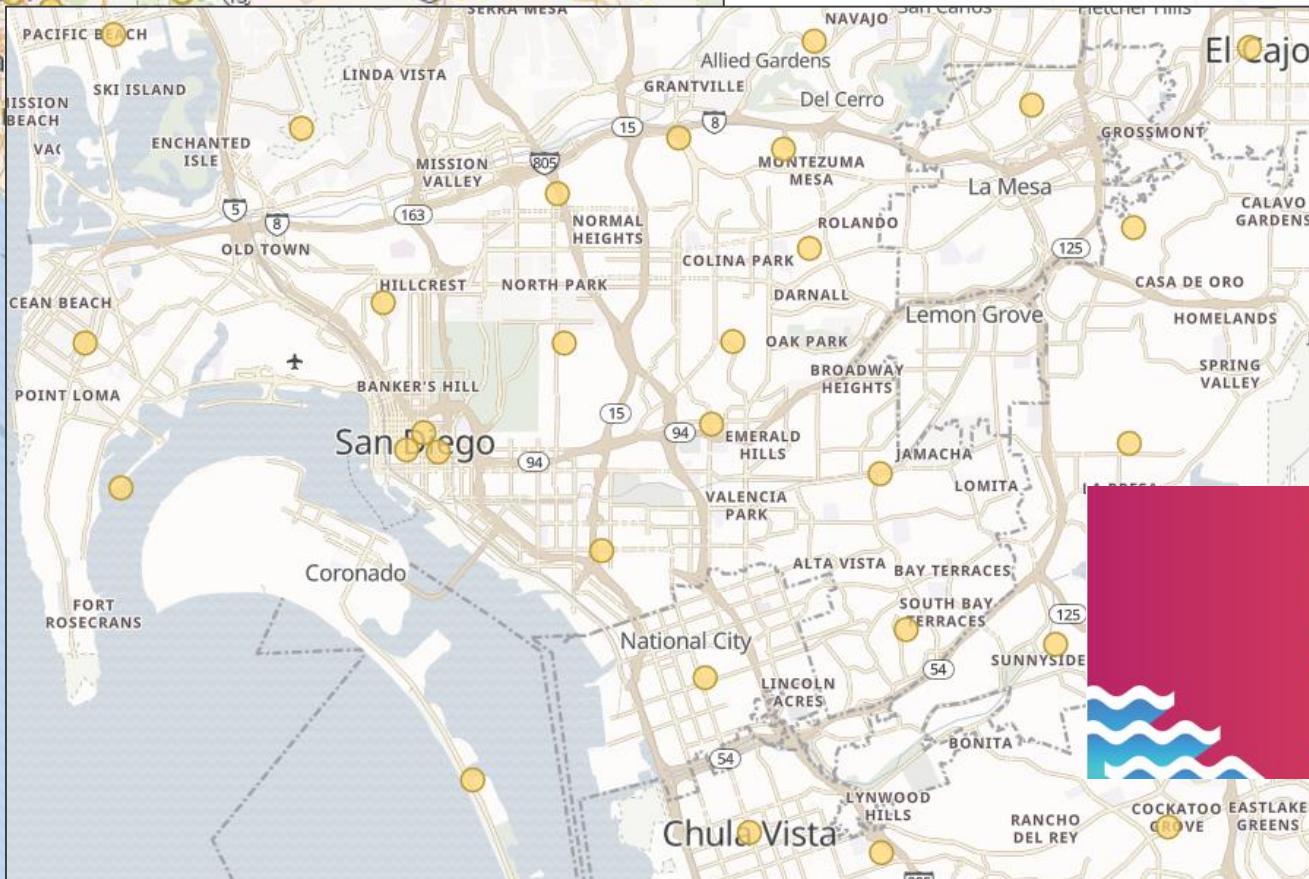
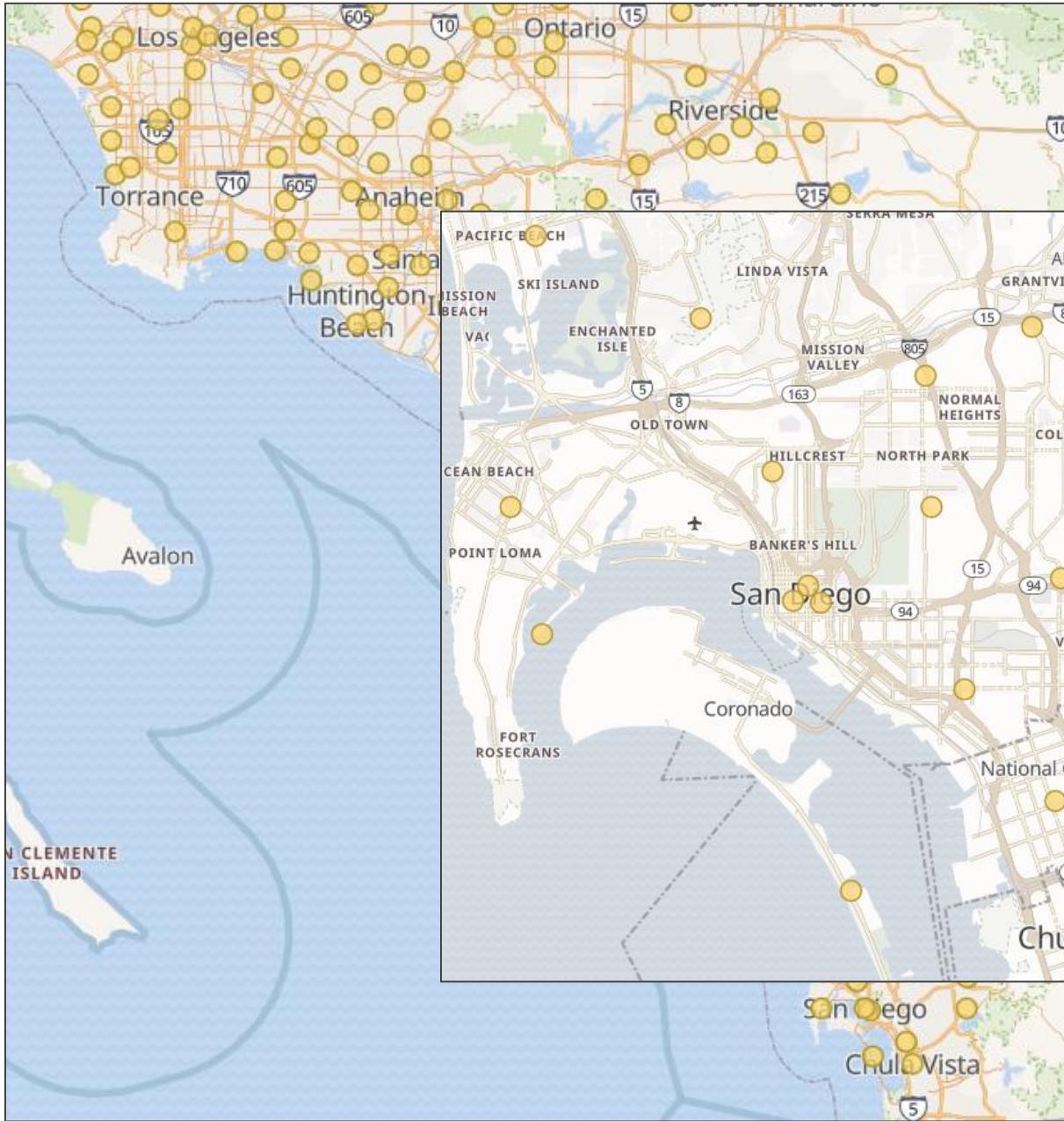
CSD-Discovery: Protein and ligand-based design of new drugs

CSD-Materials: Behaviour and properties of new materials

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







Before electronic deposition

Hand-typed tables of coordinates

2178

J. CHEM. SOC. DALTON TRANS. 1985

Table 1. Crystallographic data and details of data collection and processing for $ML(\text{NO}_3)_2$, with $M = \text{Cu}$ [in (1)], Ni [in (2)], and Cd [in (3)]

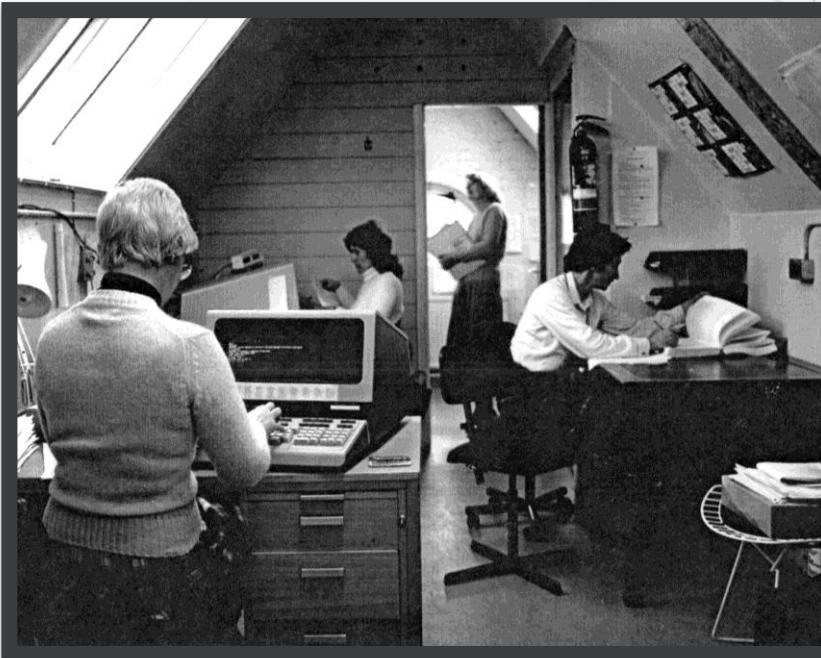
Stoichiometry M	(1) $\text{C}_{14}\text{H}_2\text{CuN}_4\text{O}_6$ 529.01	(2) $\text{C}_{14}\text{H}_2\text{NiN}_4\text{O}_6$ 524.18	(3) $\text{C}_{14}\text{H}_2\text{CdN}_4\text{O}_6$ 577.88
----------------------	--	--	--

J. CHEM. SOC. DALTON TRANS. 1985

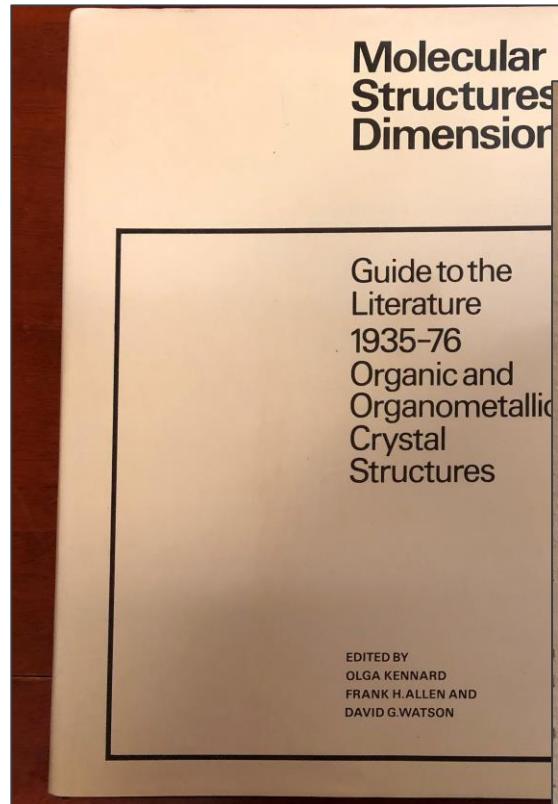
2179

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) ($\times 10^3$ for Cu, $\times 10^4$ for others)							
C(1)	7.74(2)	11.51(9)	62.93(2)	C(2)	2.25(3)	67.0(8)	4.844(3)
N(01)	1.49(3)	-1.22(3)	6.45(4)	C(30)	2.04(4)	48.5(5)	6.905(2)
C(10)	1.35(9)	-1.289(5)	6.75(8)	N(31)	1.88(8)	1.149(3)	7.431(2)
N(11)	7.08(1)	-1.425(3)	6.56(6)	N(32)	1.38(4)	1.906(3)	7.246(2)
N(12)	3.79(1)	-1.356(3)	6.56(12)	C(33)	1.39(2)	2.512(4)	7.809(2)
C(13)	-1.62(2)	-7.66(4)	6.55(8)	C(34)	1.907(2)	2.157(5)	8.332(2)
C(14)	-1.71(2)	-2.078(4)	6.70(2)	C(35)	2.220(2)	1.302(4)	8.094(2)
C(15)	3.08(2)	-2.475(4)	6.58(6)	C(36)	9.18(2)	3.421(6)	7.903(3)
C(16)	-6.51(2)	-6.54(4)	C(37)	2.72(2)	5.23(2)	8.412(3)	
C(17)	6.50(2)	-3.76(5)	6.60(6)	N(40)	-3.04(2)	2.265(5)	5.590(3)
C(20)	1.54(9)	-3.39(5)	5.77(2)	O(41)	1.36(4)	2.399(3)	6.163(2)
N(21)	1.50(2)	9.02(3)	5.45(8)	O(42)	-7.17(2)	3.03(5)	5.486(3)
N(22)	1.099(1)	1.738(3)	5.55(5)	O(43)	-2.72(2)	1.385(2)	5.217(2)
C(23)	1.123(2)	2.745(4)	5.17(7)	N(50)	1.96(2)	3.01(5)	3.346(2)
C(24)	1.539(3)	2.339(6)	4.85(4)	O(51)	1.96(2)	4.07(2)	3.604(3)
C(25)	1.786(2)	1.358(5)	5.03(5)	O(52)	1.52(2)	2.345(6)	3.201(2)
C(26)	7.43(3)	3.85(6)	5.13(9)	O(53)	2.41(7)	2.61(3)	3.337(4)
(b) Compound (2) ($\times 10^4$)							
Ni	5.110(1)	3.357(1)	7.658(1)	N(31)	2.270(6)	1.531(4)	7.437(3)
Ni(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)	9.956(6)	7.036(5)	C(33)	1.964(8)	2.706(6)	8.528(4)
N(11)	6.173(6)	9.936(4)	8.049(4)	C(34)	897(8)	1.577(7)	8.516(5)
N(12)	6.420(6)	2.086(5)	8.849(5)	C(35)	1.039(7)	3.27(2)	7.731(3)
C(13)	6.028(3)	7.240(7)	9.452(5)	C(36)	2.223(9)	3.805(7)	9.165(5)
C(14)	6.818(11)	4.768(6)	9.522(6)	C(37)	3.26(10)	-4.23(6)	7.459(6)
C(15)	6.045(9)	-1.15(6)	8.62(5)	N(40)	7.52(13)	5.107(5)	8.266(4)
C(16)	7.992(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)	8.28(6)	O(42)	8.61(7)	5.982(5)	8.553(4)
C(20)	4.684(10)	2.956(6)	5.790(5)	O(43)	7.666(5)	4.302(4)	7.690(3)
N(21)	3.777(7)	3.53(5)	5.10(5)	N(50)	1.903(7)	5.80(5)	5.080(3)
N(22)	3.909(7)	4.19(5)	6.45(4)	O(51)	1.77(7)	7.94(6)	4.339(5)
C(23)	2.971(9)	5.137(6)	6.12(6)	O(52)	3.33(10)	8.780(8)	5.620(7)
C(24)	2.311(10)	5.059(8)	5.134(6)	O(53)	5.72(4)	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)	6.60(6)	6.79(7)	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)	6.55(4)	O(53B)	1.492(4)	8.544(6)	5.645(6)



Look up of compound names



Methylene)-lycoctonine hydroiodide monohydrate	1	58.	92a	
Methylene)-malonate	5	1.	22	
Methylene)-p-bromoaniline	4	24.	11	
Methylene)-piperidine	8	33.	25	
Methylene)-s-tetrathiane	7	39.	24	
Methylene)-1,2,3,6-tetrachlorobicyclo(3.1.0)hex-2	4	27.	3	
Methylene)-1,2,4-trithiolane	4	39.	24	
Methylene)-1,2,5,6-di-O-isopropylidene- α -D-galact	1	45.	71	
Methylene)-1,4-dithiacyclobutane	4	39.	31	
Methylene)-1,6,6a-thiathiophthene	4	39.	28	
Methylene)-1,6,6a-thiathiophthene	12-p-met	7	39.	33
Methylene)-10,11-dihydro-5H-dibenzo(a,d)cyclohept	3	28.	6	
Methylene)-2,5-dihydrothiophene		5	39.	18
Methylene)-2,6-dioxabicyclo(3.3.0)octane		8	38.	51
Methylene)-2,6-dioxabicyclo(3.3.0)octane		8	38.	52
Methylene)-3-oxo-4,4-dimethyl-1,2-diazetidinium h	7	32.	26	
Methylene)-4-bromoaniline	6	33.	44	
Methylene)-4-methyl-4-acetyl-1,3-dithiacyclobutan	1	39.	61	
Methylene)-4,10-cyclotridecadiene-2,6,8,12-tetray	4	23.	5	
Methylene)aniline	7	16.	13	
Methylene)benzene		1	29.	2
Methylene)cyclohexane		3	21.	9
Methylene)dichromium tetra(lithium etherate)		4	71.	12
Methylene)trisulfur-tetranitride		6	4.	6
Methylene- α -D-lyxohexopyranoside		7	45.	40
Methylene-(<i>cis</i> -syn)-thymine dimer		3	44.	37
Methylene-adenosine-3',5'-cyclic monophosphonate		5	47.	26
Methylene-adipic acid (1,3a:6,3)diacetone p-bromo	4	38.	21a	
Methylene-ammonium ylide monohydrate		4	12.	14
Methylene-bis(nicotinamide) dichloride monohydrat		6	33.	46
Methylene-bis(phosphonic dichloride)		7	64.	1
Methylene-bis(thymine)		5	44.	19
Methylene-bis(trimethylammonium) dibromide dihydr	4	3.	21	
Methylene-bis(N-pyrrolid-2-one-5-carboxylic acid)	7	32.	18	
Methylene-bis(N-pyrrolid-2-one-5-carboxylic acid)	1	59.	39	

GERMANIUM, TIN, LEAD COMPOUNDS			
69.43	Tri(cyclo - octa - 1,5 - diene - platinum - di(tin trichloride)	C ₂₄ H ₃₆ Cl ₆ Pt ₃ Sn ₂	L.J.Guggenberger <i>Chem. Communic.</i> , 512, 1968
			Also classified in 74
69.C	bis(π - Cyclopentadienyl dicarbonyl iron) di(phenylsulfonyl) tin	C ₂₆ H ₂₀ Fe ₂ O ₈ S ₂ Sn	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 <i>J. Organometal. Chem.</i> , 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 <i>Zh. Strukt. Khim.</i> , 8, 122, 1967
69.46	(4 - Bromo - 1,2,3,4 - tetraphenyl - <i>cis,cis</i> - 1,3 - butadienyl)dimethyl tin bromide	C ₃₀ H ₂₆ Br ₂ Sn	F.P.Boer, J.J.Flynn, H.H.Freedman, S.V.McKinley, V.R.Sandel <i>J. Amer. Chem. Soc.</i> , 89, 5068, 1967
69.47	Tricyclohexyl tin acetate	C ₃₀ H ₃₆ O ₂ Sn	N.W.Alcock, R.E.Timms <i>J. Chem. Soc. (A)</i> , 1876, 1968
69.48	Triphenylgermanium manganese pentacarbonyl	C ₃₅ H ₁₅ GeMnO ₅	B.T.Kilbourn, T.L.Blundell, H.M.Powell <i>Chem. Communic.</i> , 444, 1965
			Also classified in 11
69.49	Lead hexa - antipyrine perchlorate	C ₆₆ H ₇₂ N ₁₂ O ₆ Pb ²⁺ , 2ClO ₄ ⁻	M.Vijayan, M.A.Viswamitra <i>Acta Cryst.</i> , 21, 522, 1966

Publication of crystal structures today

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

Issue 20, 2018

From the journal:
Dalton Transactions

The coordination chemistry of the new pyridyl silicon ligand [PhSi(6-Me-2-py)]₃

Alex J. Plajer,^a Annie L. Colebatch,^a Markus Enders,^b Álvaro García-Romero,^a Dominic S. Wright^{a*}

[Author affiliations](#)

Abstract

Difficulties in the preparation of neutral ligands of the type [RSi(6-Me-2-py)]₃ (R = pyridyl ring unit) have thwarted efforts to expand the coordination chemistry of the pyridyl ring. Simply switching the pyridyl substituents to 6-methyl-pyridyl (*o*-methylpyridyl) has allowed smooth, high-yielding access to the [PhSi(6-Me-2-py)]₃ ligand, which is amenable to coordination chemistry with transition metals. The synthesis, structure and magnetic properties of the new complexes [{PhSi(6-Me-2-py)]₃}CuCl₃Cl₂, [{PhSi(6-Me-2-py)]₃}FeCl₂, [{PhSi(6-Me-2-py)]₃}Mo(CO)₃] and [{PhSi(6-Me-2-py)]₃}CrCl₃ are reported. The paramagnetic Fe²⁺ and Co²⁺ complexes show strongly shifted ESR signals due to large Fermi-contact shifts. However, magnetic anisotropy is too small to account for the shifts so that both contributions have to be included in the parameterization.

CCDC reference: 1833560

Keywords: coordination chemistry, crystal structure, silicon, pyridyl, 6-methylpyridyl, 6-methylpyridyl silicon ligand, [PhSi(6-Me-2-py)]₃, [PhSi(6-Me-2-py)]₃CrCl₃, [PhSi(6-Me-2-py)]₃FeCl₂, [PhSi(6-Me-2-py)]₃Mo(CO)₃, [PhSi(6-Me-2-py)]₃FeCl₂

Results	
Database Identifier	Deposition Number
<input checked="" type="checkbox"/> LEVYET	1833560
<input checked="" type="checkbox"/> LEVYIX	1833561
<input checked="" type="checkbox"/> LEVYOD	1833562
<input checked="" type="checkbox"/> LEVYUU	1833563
<input checked="" type="checkbox"/> TIGWUE	1833558
<input checked="" type="checkbox"/> TIGXAL	1833559

[Download](#)

LEVYET : dichloro-[2,2-[(6-methylpyridin-2-yl)(phenyl)silanediyl]bis(6-methylpyridine)]-iron
Space Group: C 2/c (15), Cell: a 20.9935(5) Å b 10.2893(2) Å c 22.4986(5) Å, α 90° β 93.9540(10)° γ 90°

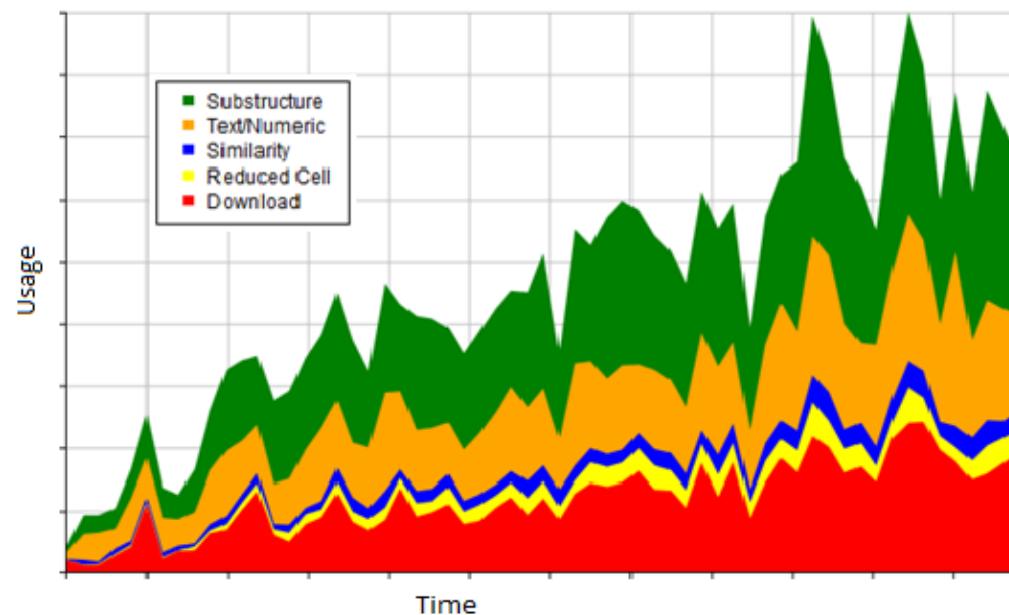
3D viewer

Chemical diagram

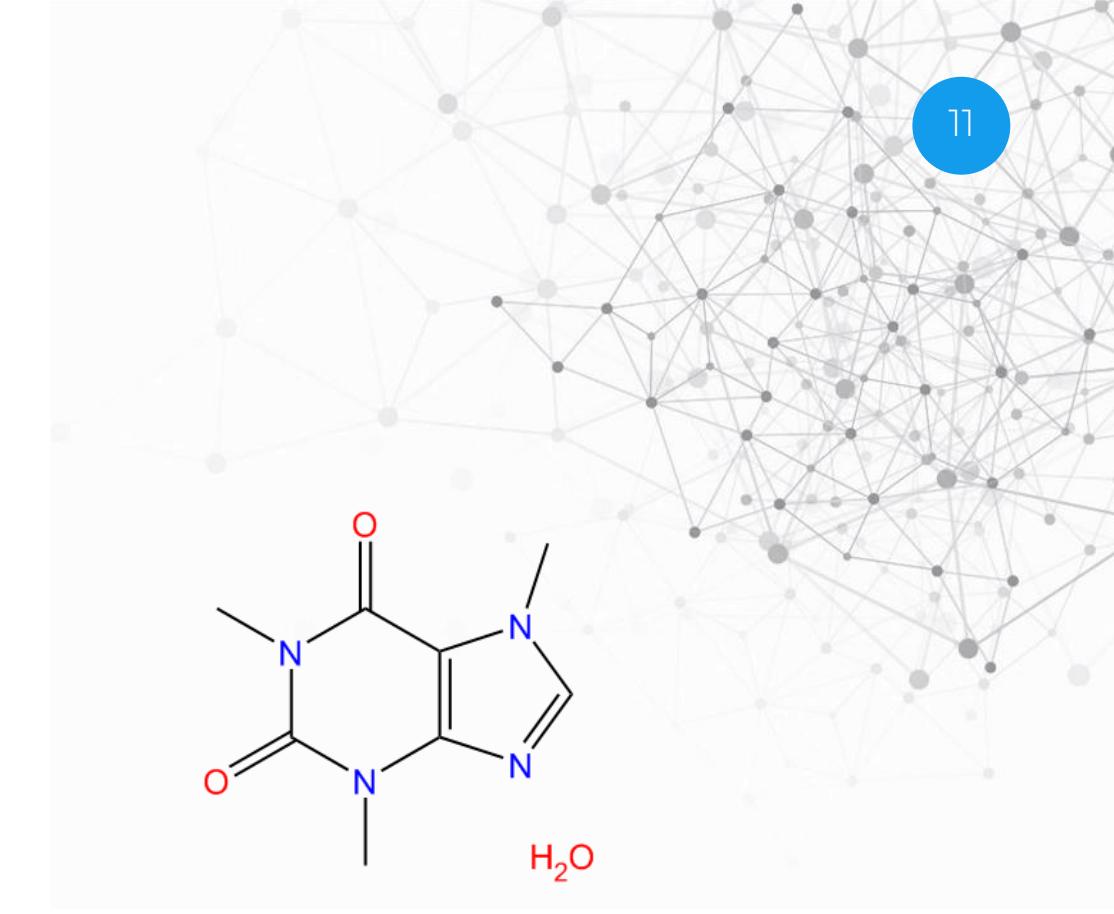
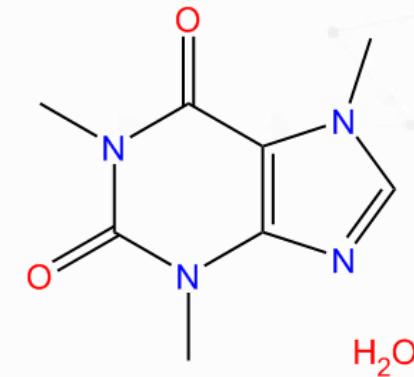
CC1=CC=C2=C1N=C(C=C2)N3C=C(C=C3)N4=C(C=C(C=C4)N5=C(C=C(C=C5)N6=C(C=C(C=C6)N7=C(C=C(C=C7)N8=C(C=C(C=C8)N9=C(C=C(C=C9)N10=C(C=C(C=C10)N11=C(C=C(C=C11)N12=C(C=C(C=C12)N13=C(C=C(C=C13)N14=C(C=C(C=C14)N15=C(C=C(C=C15)N16=C(C=C(C=C16)N17=C(C=C(C=C17)N18=C(C=C(C=C18)N19=C(C=C(C=C19)N20=C(C=C(C=C20)N21=C(C=C(C=C21)N22=C(C=C(C=C22)N23=C(C=C(C=C23)N24=C(C=C(C=C24)N25=C(C=C(C=C25)N26=C(C=C(C=C26)N27=C(C=C(C=C27)N28=C(C=C(C=C28)N29=C(C=C(C=C29)N30=C(C=C(C=C30)N31=C(C=C(C=C31)N32=C(C=C(C=C32)N33=C(C=C(C=C33)N34=C(C=C(C=C34)N35=C(C=C(C=C35)N36=C(C=C(C=C36)N37=C(C=C(C=C37)N38=C(C=C(C=C38)N39=C(C=C(C=C39)N40=C(C=C(C=C40)N41=C(C=C(C=C41)N42=C(C=C(C=C42)N43=C(C=C(C=C43)N44=C(C=C(C=C44)N45=C(C=C(C=C45)N46=C(C=C(C=C46)N47=C(C=C(C=C47)N48=C(C=C(C=C48)N49=C(C=C(C=C49)N50=C(C=C(C=C50)N51=C(C=C(C=C51)N52=C(C=C(C=C52)N53=C(C=C(C=C53)N54=C(C=C(C=C54)N55=C(C=C(C=C55)N56=C(C=C(C=C56)N57=C(C=C(C=C57)N58=C(C=C(C=C58)N59=C(C=C(C=C59)N60=C(C=C(C=C60)N61=C(C=C(C=C61)N62=C(C=C(C=C62)N63=C(C=C(C=C63)N64=C(C=C(C=C64)N65=C(C=C(C=C65)N66=C(C=C(C=C66)N67=C(C=C(C=C67)N68=C(C=C(C=C68)N69=C(C=C(C=C69)N70=C(C=C(C=C70)N71=C(C=C(C=C71)N72=C(C=C(C=C72)N73=C(C=C(C=C73)N74=C(C=C(C=C74)N75=C(C=C(C=C75)N76=C(C=C(C=C76)N77=C(C=C(C=C77)N78=C(C=C(C=C78)N79=C(C=C(C=C79)N80=C(C=C(C=C80)N81=C(C=C(C=C81)N82=C(C=C(C=C82)N83=C(C=C(C=C83)N84=C(C=C(C=C84)N85=C(C=C(C=C85)N86=C(C=C(C=C86)N87=C(C=C(C=C87)N88=C(C=C(C=C88)N89=C(C=C(C=C89)N90=C(C=C(C=C90)N91=C(C=C(C=C91)N92=C(C=C(C=C92)N93=C(C=C(C=C93)N94=C(C=C(C=C94)N95=C(C=C(C=C95)N96=C(C=C(C=C96)N97=C(C=C(C=C97)N98=C(C=C(C=C98)N99=C(C=C(C=C99)N100=C(C=C(C=C100)N101=C(C=C(C=C101)N102=C(C=C(C=C102)N103=C(C=C(C=C103)N104=C(C=C(C=C104)N105=C(C=C(C=C105)N106=C(C=C(C=C106)N107=C(C=C(C=C107)N108=C(C=C(C=C108)N109=C(C=C(C=C109)N110=C(C=C(C=C110)N111=C(C=C(C=C111)N112=C(C=C(C=C112)N113=C(C=C(C=C113)N114=C(C=C(C=C114)N115=C(C=C(C=C115)N116=C(C=C(C=C116)N117=C(C=C(C=C117)N118=C(C=C(C=C118)N119=C(C=C(C=C119)N120=C(C=C(C=C120)N121=C(C=C(C=C121)N122=C(C=C(C=C122)N123=C(C=C(C=C123)N124=C(C=C(C=C124)N125=C(C=C(C=C125)N126=C(C=C(C=C126)N127=C(C=C(C=C127)N128=C(C=C(C=C128)N129=C(C=C(C=C129)N130=C(C=C(C=C130)N131=C(C=C(C=C131)N132=C(C=C(C=C132)N133=C(C=C(C=C133)N134=C(C=C(C=C134)N135=C(C=C(C=C135)N136=C(C=C(C=C136)N137=C(C=C(C=C137)N138=C(C=C(C=C138)N139=C(C=C(C=C139)N140=C(C=C(C=C140)N141=C(C=C(C=C141)N142=C(C=C(C=C142)N143=C(C=C(C=C143)N144=C(C=C(C=C144)N145=C(C=C(C=C145)N146=C(C=C(C=C146)N147=C(C=C(C=C147)N148=C(C=C(C=C148)N149=C(C=C(C=C149)N150=C(C=C(C=C150)N151=C(C=C(C=C151)N152=C(C=C(C=C152)N153=C(C=C(C=C153)N154=C(C=C(C=C154)N155=C(C=C(C=C155)N156=C(C=C(C=C156)N157=C(C=C(C=C157)N158=C(C=C(C=C158)N159=C(C=C(C=C159)N160=C(C=C(C=C160)N161=C(C=C(C=C161)N162=C(C=C(C=C162)N163=C(C=C(C=C163)N164=C(C=C(C=C164)N165=C(C=C(C=C165)N166=C(C=C(C=C166)N167=C(C=C(C=C167)N168=C(C=C(C=C168)N169=C(C=C(C=C169)N170=C(C=C(C=C170)N171=C(C=C(C=C171)N172=C(C=C(C=C172)N173=C(C=C(C=C173)N174=C(C=C(C=C174)N175=C(C=C(C=C175)N176=C(C=C(C=C176)N177=C(C=C(C=C177)N178=C(C=C(C=C178)N179=C(C=C(C=C179)N180=C(C=C(C=C180)N181=C(C=C(C=C181)N182=C(C=C(C=C182)N183=C(C=C(C=C183)N184=C(C=C(C=C184)N185=C(C=C(C=C185)N186=C(C=C(C=C186)N187=C(C=C(C=C187)N188=C(C=C(C=C188)N189=C(C=C(C=C189)N190=C(C=C(C=C190)N191=C(C=C(C=C191)N192=C(C=C(C=C192)N193=C(C=C(C=C193)N194=C(C=C(C=C194)N195=C(C=C(C=C195)N196=C(C=C(C=C196)N197=C(C=C(C=C197)N198=C(C=C(C=C198)N199=C(C=C(C=C199)N200=C(C=C(C=C200)N201=C(C=C(C=C201)N202=C(C=C(C=C202)N203=C(C=C(C=C203)N204=C(C=C(C=C204)N205=C(C=C(C=C205)N206=C(C=C(C=C206)N207=C(C=C(C=C207)N208=C(C=C(C=C208)N209=C(C=C(C=C209)N210=C(C=C(C=C210)N211=C(C=C(C=C211)N212=C(C=C(C=C212)N213=C(C=C(C=C213)N214=C(C=C(C=C214)N215=C(C=C(C=C215)N216=C(C=C(C=C216)N217=C(C=C(C=C217)N218=C(C=C(C=C218)N219=C(C=C(C=C219)N220=C(C=C(C=C220)N221=C(C=C(C=C221)N222=C(C=C(C=C222)N223=C(C=C(C=C223)N224=C(C=C(C=C224)N225=C(C=C(C=C225)N226=C(C=C(C=C226)N227=C(C=C(C=C227)N228=C(C=C(C=C228)N229=C(C=C(C=C229)N230=C(C=C(C=C230)N231=C(C=C(C=C231)N232=C(C=C(C=C232)N233=C(C=C(C=C233)N234=C(C=C(C=C234)N235=C(C=C(C=C235)N236=C(C=C(C=C236)N237=C(C=C(C=C237)N238=C(C=C(C=C238)N239=C(C=C(C=C239)N240=C(C=C(C=C240)N241=C(C=C(C=C241)N242=C(C=C(C=C242)N243=C(C=C(C=C243)N244=C(C=C(C=C244)N245=C(C=C(C=C245)N246=C(C=C(C=C246)N247=C(C=C(C=C247)N248=C(C=C(C=C248)N249=C(C=C(C=C249)N250=C(C=C(C=C250)N251=C(C=C(C=C251)N252=C(C=C(C=C252)N253=C(C=C(C=C253)N254=C(C=C(C=C254)N255=C(C=C(C=C255)N256=C(C=C(C=C256)N257=C(C=C(C=C257)N258=C(C=C(C=C258)N259=C(C=C(C=C259)N260=C(C=C(C=C260)N261=C(C=C(C=C261)N262=C(C=C(C=C262)N263=C(C=C(C=C263)N264=C(C=C(C=C264)N265=C(C=C(C=C265)N266=C(C=C(C=C266)N267=C(C=C(C=C267)N268=C(C=C(C=C268)N269=C(C=C(C=C269)N270=C(C=C(C=C270)N271=C(C=C(C=C271)N272=C(C=C(C=C272)N273=C(C=C(C=C273)N274=C(C=C(C=C274)N275=C(C=C(C=C275)N276=C(C=C(C=C276)N277=C(C=C(C=C277)N278=C(C=C(C=C278)N279=C(C=C(C=C279)N280=C(C=C(C=C280)N281=C(C=C(C=C281)N282=C(C=C(C=C282)N283=C(C=C(C=C283)N284=C(C=C(C=C284)N285=C(C=C(C=C285)N286=C(C=C(C=C286)N287=C(C=C(C=C287)N288=C(C=C(C=C288)N289=C(C=C(C=C289)N290=C(C=C(C=C290)N291=C(C=C(C=C291)N292=C(C=C(C=C292)N293=C(C=C(C=C293)N294=C(C=C(C=C294)N295=C(C=C(C=C295)N296=C(C=C(C=C296)N297=C(C=C(C=C297)N298=C(C=C(C=C298)N299=C(C=C(C=C299)N300=C(C=C(C=C300)N301=C(C=C(C=C301)N302=C(C=C(C=C302)N303=C(C=C(C=C303)N304=C(C=C(C=C304)N305=C(C=C(C=C305)N306=C(C=C(C=C306)N307=C(C=C(C=C307)N308=C(C=C(C=C308)N309=C(C=C(C=C309)N310=C(C=C(C=C310)N311=C(C=C(C=C311)N312=C(C=C(C=C312)N313=C(C=C(C=C313)N314=C(C=C(C=C314)N315=C(C=C(C=C315)N316=C(C=C(C=C316)N317=C(C=C(C=C317)N318=C(C=C(C=C318)N319=C(C=C(C=C319)N320=C(C=C(C=C320)N321=C(C=C(C=C321)N322=C(C=C(C=C322)N323=C(C=C(C=C323)N324=C(C=C(C=C324)N325=C(C=C(C=C325)N326=C(C=C(C=C326)N327=C(C=C(C=C327)N328=C(C=C(C=C328)N329=C(C=C(C=C329)N330=C(C=C(C=C330)N331=C(C=C(C=C331)N332=C(C=C(C=C332)N333=C(C=C(C=C333)N334=C(C=C(C=C334)N335=C(C=C(C=C335)N336=C(C=C(C=C336)N337=C(C=C(C=C337)N338=C(C=C(C=C338)N339=C(C=C(C=C339)N340=C(C=C(C=C340)N341=C(C=C(C=C341)N342=C(C=C(C=C342)N343=C(C=C(C=C343)N344=C(C=C(C=C344)N345=C(C=C(C=C345)N346=C(C=C(C=C346)N347=C(C=C(C=C347)N348=C(C=C(C=C348)N349=C(C=C(C=C349)N350=C(C=C(C=C350)N351=C(C=C(C=C351)N352=C(C=C(C=C352)N353=C(C=C(C=C353)N354=C(C=C(C=C354)N355=C(C=C(C=C355)N356=C(C=C(C=C356)N357=C(C=C(C=C357)N358=C(C=C(C=C358)N359=C(C=C(C=C359)N360=C(C=C(C=C360)N361=C(C=C(C=C361)N362=C(C=C(C=C362)N363=C(C=C(C=C363)N364=C(C=C(C=C364)N365=C(C=C(C=C365)N366=C(C=C(C=C366)N367=C(C=C(C=C367)N368=C(C=C(C=C368)N369=C(C=C(C=C369)N370=C(C=C(C=C370)N371=C(C=C(C=C371)N372=C(C=C(C=C372)N373=C(C=C(C=C373)N374=C(C=C(C=C374)N375=C(C=C(C=C375)N376=C(C=C(C=C376)N377=C(C=C(C=C377)N378=C(C=C(C=C378)N379=C(C=C(C=C379)N380=C(C=C(C=C380)N381=C(C=C(C=C381)N382=C(C=C(C=C382)N383=C(C=C(C=C383)N384=C(C=C(C=C384)N385=C(C=C(C=C385)N386=C(C=C(C=C386)N387=C(C=C(C=C387)N388=C(C=C(C=C388)N389=C(C=C(C=C389)N390=C(C=C(C=C390)N391=C(C=C(C=C391)N392=C(C=C(C=C392)N393=C(C=C(C=C393)N394=C(C=C(C=C394)N395=C(C=C(C=C395)N396=C(C=C(C=C396)N397=C(C=C(C=C397)N398=C(C=C(C=C398)N399=C(C=C(C=C399)N400=C(C=C(C=C400)N401=C(C=C(C=C401)N402=C(C=C(C=C402)N403=C(C=C(C=C403)N404=C(C=C(C=C404)N405=C(C=C(C=C405)N406=C(C=C(C=C406)N407=C(C=C(C=C407)N408=C(C=C(C=C408)N409=C(C=C(C=C409)N410=C(C=C(C=C410)N411=C(C=C(C=C411)N412=C(C=C(C=C412)N413=C(C=C(C=C413)N414=C(C=C(C=C414)N415=C(C=C(C=C415)N416=C(C=C(C=C416)N417=C(C=C(C=C417)N418=C(C=C(C=C418)N419=C(C=C(C=C419)N420=C(C=C(C=C420)N421=C(C=C(C=C421)N422=C(C=C(C=C422)N423=C(C=C(C=C423)N424=C(C=C(C=C424)N425=C(C=C(C=C425)N426=C(C=C(C=C426)N427=C(C=C(C=C427)N428=C(C=C(C=C428)N429=C(C=C(C=C429)N430=C(C=C(C=C430)N431=C(C=C(C=C431)N432=C(C=C(C=C432)N433=C(C=C(C=C433)N434=C(C=C(C=C434)N435=C(C=C(C=C435)N436=C(C=C(C=C436)N437=C(C=C(C=C437)N438=C(C=C(C=C438)N439=C(C=C(C=C439)N440=C(C=C(C=C440)N441=C(C=C(C=C441)N442=C(C=C(C=C442)N443=C(C=C(C=C443)N444=C(C=C(C=C444)N445=C(C=C(C=C445)N446=C(C=C(C=C446)N447=C(C=C(C=C447)N448=C(C=C(C=C448)N449=C(C=C(C=C449)N450=C(C=C(C=C450)N451=C(C=C(C=C451)N452=C(C=C(C=C452)N453=C(C=C(C=C453)N454=C(C=C(C=C454)N455=C(C=C(C=C455)N456=C(C=C(C=C456)N457=C(C=C(C=C457)N458=C(C=C(C=C458)N459=C(C=C(C=C459)N460=C(C=C(C=C460)N461=C(C=C(C=C461)N462=C(C=C(C=C462)N463=C(C=C(C=C463)N464=C(C=C(C=C464)N465=C(C=C(C=C465)N466=C(C=C(C=C466)N467=C(C=C(C=C467)N468=C(C=C(C=C468)N469=C(C=C(C=C469)N470=C(C=C(C=C470)N471=C(C=C(C=C471)N472=C(C=C(C=C472)N473=C(C=C(C=C473)N474=C(C=C(C=C474)N475=C(C=C(C=C475)N476=C(C=C(C=C476)N477=C(C=C(C=C477)N478=C(C=C(C=C478)N479=C(C=C(C=C479)N480=C(C=C(C=C480)N481=C(C=C(C=C481)N482=C(C=C(C=C482)N483=C(C=C(C=C483)N484=C(C=C(C=C484)N485=C(C=C(C=C485)N486=C(C=C(C=C486)N487=C(C=C(C=C487)N488=C(C=C(C=C488)N489=C(C=C(C=C489)N490=C(C=C(C=C490)N491=C(C=C(C=C491)N492=C(C=C(C=C492)N493=C(C=C(C=C493)N494=C(C=C(C=C494)N495=C(C=C(C=C495)N496=C(C=C(C=C496)N497=C(C=C(C=C497)N498=C(C=C(C=C498)N499=C(C=C(C=C499)N500=C(C=C(C=C500)N501=C(C=C(C=C501)N502=C(C=C(C=C502)N503=C(C=C(C=C503)N504=C(C=C(C=C504)N505=C(C=C(C=C505)N506=C(C=C(C=C506)N507=C(C=C(C=C507)N508=C(C=C(C=C508)N509=C(C=C(C=C509)N510=C(C=C(C=C510)N511=C(C=C(C=C511)N512=C(C=C(C=C512)N513=C(C=C(C=C513)N514=C(C=C(C=C514)N515=C(C=C(C=C515)N516=C(C=C(C=C516)N517=C(C=C(C=C517)N518=C(C=C(C=C518)N519=C(C=C(C=C519)N520=C(C=C(C=C520)N521=C(C=C(C=C521)N522=C(C=C(C=C522)N523=C(C=C(C=C523)N524=C(C=C(C=C524)N525=C(C=C(C=C525)N526=C(C=C(C=C526)N527=C(C=C(C=C527)N528=C(C=C(C=C528)N529=C(C=C(C=C529)N530=C(C=C(C=C530)N531=C(C=C(C=C531)N532=C(C=C(C=C532)N533=C(C=C(C=C533)N534=C(C=C(C=C534)N535=C(C=C(C=C535)N536=C(C=C(C=C536)N537=C(C=C(C=C537)N538=C(C=C(C=C538)N539=C(C=C(C=C539)N540=C(C=C(C=C540)N541=C(C=C(C=C541)N542=C(C=C(C=C542)N543=C(C=C(C=C543)N544=C(C=C(C=C544)N545=C(C=C(C=C545)N546=C(C=C(C=C546)N547=C(C=C(C=C547)N548=C(C=C(C=C548)N549=C(C=C(C=C549)N550=C(C=C(C=C550)N551=C(C=C(C=C551)N552=C(C=C(C=C552)N553=C(C=C(C=C553)N554=C(C=C(C=C554)N555=C(C=C(C=C555)N556=C(C=C(C=C556)N557=C(C=C(C=C557)N558=C(C=C(C=C558)N559=C(C=C(C=C559)N560=C(C=C(C=C560)N561=C(C=C(C=C561)N562=C(C=C(C=C562)N563=C(C=C(C=C563)N564=C(C=C(C=C564)N565=C(C=C(C=C565)N566=C(C=C(C=C566)N567=C(C=C(C=C567)N568=C(C=C(C=C568)N569=C(C=C(C=C569)N570=C(C=C(C=C570)N571=C(C=C(C=C571)N572=C(C=C(C=C572)N573=C(C=C(C=C573)N574=C(C=C(C=C574)N575=C(C=C(C=C575)N576=C(C=C(C=C576)N577=C(C=C(C=C577)N578=C(C=C(C=C578)N579=C(C=C(C=C579)N580=C(C=C(C=C580)N581=C(C=C(C=C581)N582=C(C=C(C=C582)N583=C(C=C(C=C583)N584=C(C=C(C=C584)N585=C(C=C(C=C585)N586=C(C=C(C=C586)N587=C(C=C(C=C587)N588=C(C=C(C=C588)N589=C(C=C(C=C589)N590=C(C=C(C=C590)N591=C(C=C(C=C591)N592=C(C=C(C=C592)N593=C(C=C(C=C593)N594=C(C=C(C=C594)N595=C(C=C(C=C595)N596=C(C=C(C=C596)N597=C(C=C(C=C597)N598=C(C=C(C=C598)N599=C(C=C(C=C599)N600=C(C=C(C=C600)N601=C(C=C(C=C601)N602=C(C=C(C=C602)N603=C(C=C(C=C603)N604=C(C=C(C=C604)N605=C(C=C(C=C605)N606=C(C=C(C=C606)N607=C(C=C(C=C607)N608=C(C=C(C=C608)N609=C(C=C(C=C609)N610=C(C=C(C=C610)N611=C(C=C(C=C611)N612=C(C=C(C=C612)N613=C(C=C(C=C613)N614=C(C=C(C=C614)N615=C(C=C(C=C615)N616=C(C=C(C=C616)N617=C(C=C(C=C617)N618=C(C=C(C=C618)N619=C(C=C(C=C619)N620=C(C=C(C=C620)N621=C(C=C(C=C621)N622=C(C=C(C=C622)N623=C(C=C(C=C623)N624=C(C=C(C=C624)N625=C(C=C(C=C625)N626=C(C=C(C=C626)N627=C(C=C(C=C627)N628=C(C=C(C=C628)N629=C(C=C(C=C629)N630=C(C=C(C=C630)N631=C(C=C(C=C631)N632=C(C=C(C=C632)N633=C(C=C(C=C633)N634=C(C=C(C=C634)N635=C(C=C(C=C635)N636=C(C=C(C=C636)N637=C(C=C(C=C637)N638=C(C=C(C=C638)N639=C(C=C(C=C639)N640=C(C=C(C=C640)N641=C(C=C(C=C641)N642=C(C=C(C=C642)N643=C(C=C(C=C643)N644=C(C=C(C=C644)N645=C(C=C(C=C645)N646=C(C=C(C=C646)N647=C(C=C(C=C647)N648=C(C=C(C=C648)N649=C(C=C(C=C649)N650=C(C=C(C=C650)N651=C(C=C(C=C651)N652=C(C=C(C=C652)N653=C(C=C(C=C653)N654=C(C=C(C=C654)N655=C(C=C(C=C655)N656=C(C=C(C=C656)N657=C(C=C(C=C657)N658=C(C=C(C=C658)N659=C(C=C(C=C659)N660=C(C=C(C=C660)N661=C(C=C(C=C661)N662=C(C=C(C=C662)N663=C(C=C(C=C663)N664=C(C=C(C=C664)N665=C(C=C(C=C665)N666=C(C=C(C=C666)N667=C(C=C(C=C667)N668=C(C=C(C=C668)N669=C(C=C(C=C669)N670=C(C=C(C=C670)N671=C(C=C(C=C671)N672=C(C=C(C=C672)N673=C(C=C(C=C673)N674=C(C=C(C=C674)N675=C(C=C(C=C675)N676=C(C=C(C=C676)N677=C(C=C(C=C677)N678=C(C=C(C=C678)N679=C(C=C(C=C679)N680=C(C=C(C=C680)N681=C(C=C(C=C681)N682=C(C=C(C=C682)N683=C(C=C(C=C683)N684=C(C=C(C=C684)N685=C(C=C(C=C685)N686=C(C=C(C=C686)N687=C(C=C(C=C687)N688=C(C=C(C=C688)N689=C(C=C(C=C689)N690=C(C=C(C=C690)N691=C(C=C(C=C691)N692=C(C=C(C=C692)N693=C(C=C(C=C693)N694=C(C=C(C=C694)N695=C(C=C(C=C695)N696=C(C=C(C=C696)N697=C(C=C(C=C697)N698=C(C=C(C=C698)N699=C(C=C(C=C699)N700=C(C=C(C=C700)N701=C(C=C(C=C701)N702=C(C=C(C=C702)N703=C(C=C(C=C703)N704=C(C=C(C=C704)N705=C(C=C(C=C705)N706=C(C=C(C=C706)N707=C(C=C(C=C707)N708=C(C=C(C=C708)N709=C(C=C(C=C709)N710=C(C=C(C=C710)N711=C(C=C(C=C711)N712=C(C=C(C=C712)N713=C(C=C(C=C713)N714=C(C=C(C=C714)N715=C(C=C(C=C715)N716=C(C=C(C=C716)N717=C(C=C(C=C717)N718=C(C=C(C=C718)N719=C(C=C(C=C719)N720=C(C=C(C=C720)N721=C(C=C(C=C721)N722=C(C=C(C=C722)N723=C(C=C(C=C723)N724=C(C=C(C=C724)N725=C(C=C(C=C725)N726=C(C=C(C=C726)N727=C(C=C(C=C727)N728=C(C=C(C=C728)N729=C(C=C(C=C729)N730=C(C=C(C=C730)N731=C(C=C(C=C731)N732=C(C=C(C=C732)N733=C(C=C(C=C733)N734=C(C=C(C=C734)N735=C(C=C(C=C735)N736=C(C=C(C=C736)N737=C(C=C(C=C737)N738=C(C=C(C=C738)N739=C(C=C(C=C739)N740=C(C=C(C=C740)N741=C(C=C(C=C741)N742=C(C=C(C

Searching for structures

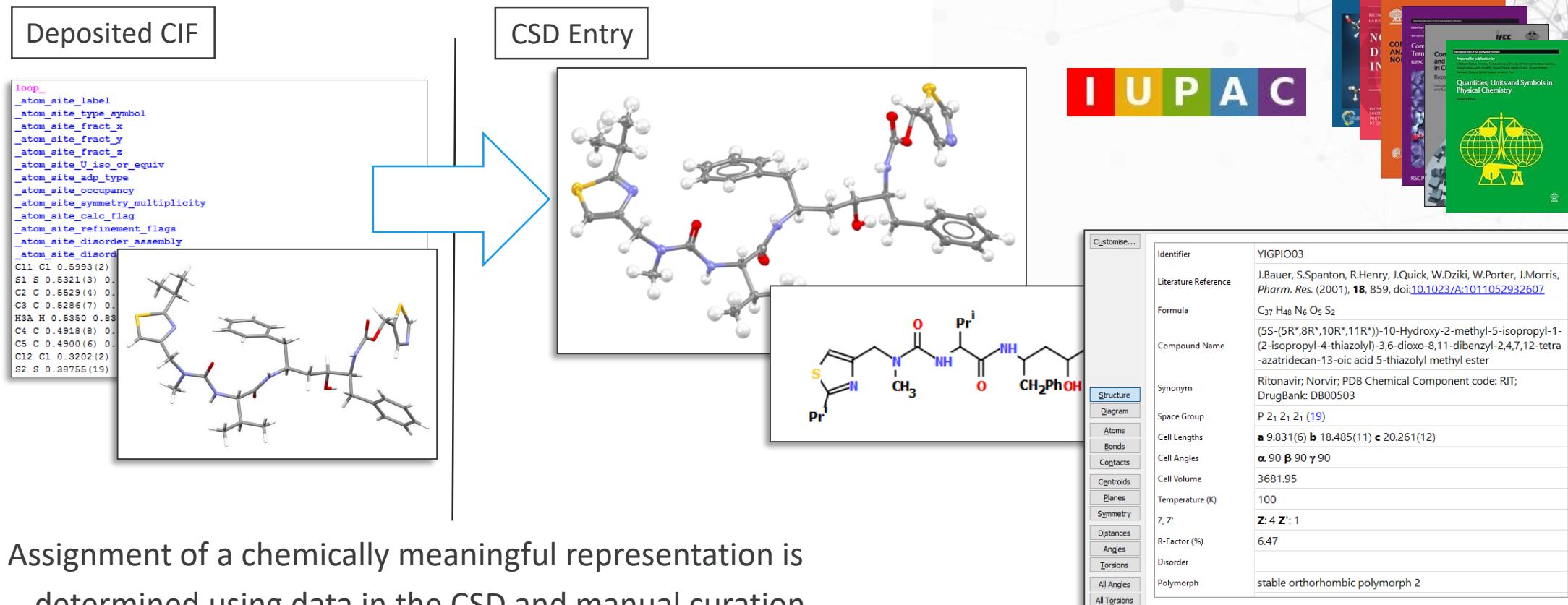
- Majority of searches of CSD are substructure searches; however:
 - 16% of all searches in WebCSD are on compound name



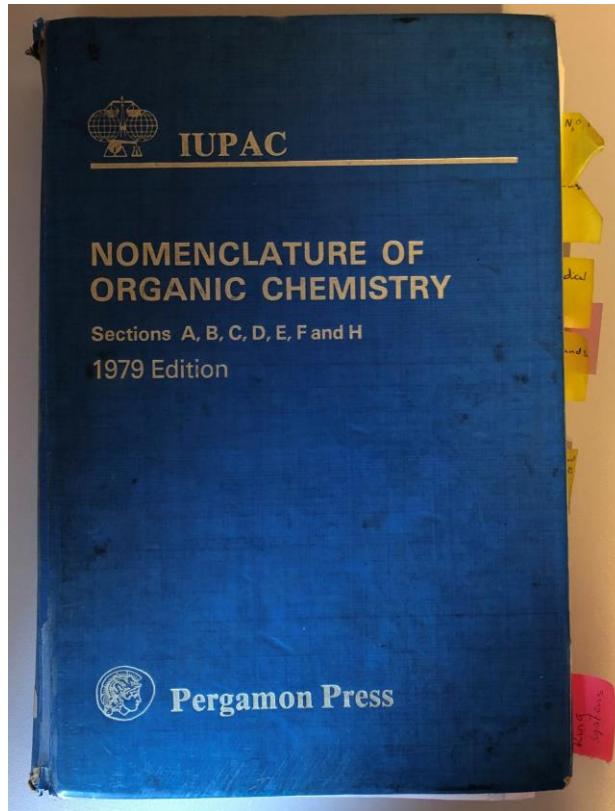
Identifier	CAFINE
Source Database	as531be
Reliability Score	★★★★★ Explain score
Author(s)	D.J.Sutor
Reference	<i>Acta Crystallogr.</i> (1958), 11 , 453, doi:10.1107/S0365110X58001286
Formula	C ₈ H ₁₀ N ₄ O ₂ ·H ₂ O
Compound	1,3,7-Trimethyl-purine-2,6-dione monohydrate
Synonym	Caffeine monohydrate
Space Group	P 2 ₁ /a



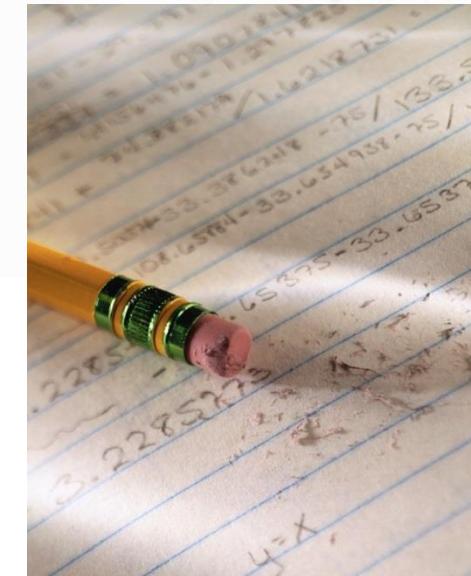
Curation and chemistry assignment



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



Using ACD/Name

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



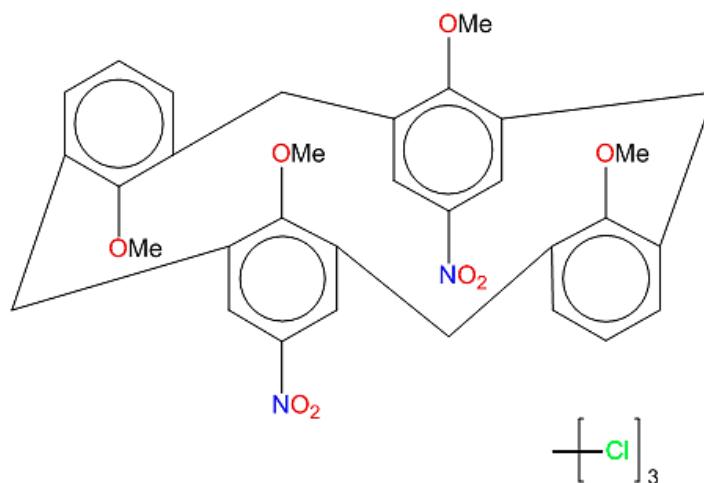
Adoption of using ACD/Name



- 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

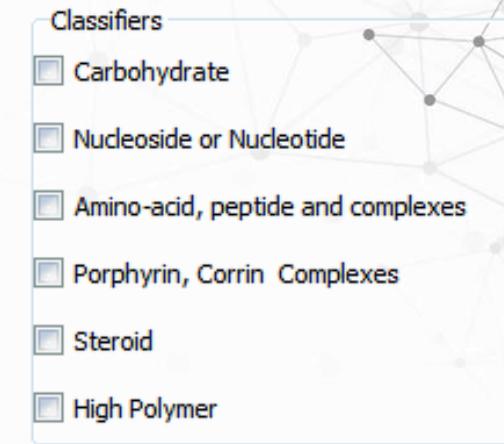
IUPAC and CSD conventions

- 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.



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 thiophene penicillin lithium sodium%20oxalate
 zeolite phthalic%20acid isoniazid titanium%20dioxide succinic%20acid citric%20acid artemisinin
 ammonia uio-66-nh2 zinc%20oxide sulfuric%20acid tetracene oxalic%20acid metformin %ce%b3-cyclodextrin
 pcn-222 ruthenium mil-88 indomethacin biphenyl polyethylene mannitol
 mil-53(al) imidazole nicotinamide ice water 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 sulfadiazine aspirin hkust-1 cyclodextrin nu-1000 titanium cu-btc
 tartaric%20acid cellulose fe zif-90 graphite porphyrin urea mil-53 uio-66 mof-5 mof-74 mof naphthalene silver
 erythritol nacl iron zif carbamazepine benzene zif-8 acetaminophen theophylline cl-20 lactose
 irmof-1 rubrene al2o3 methane tio2 uio-67 mofs zif-67 pcn-224 mos2 silicon heptane zif-7 phenylalanine piroxicam c60 paracetamol cholesterol gold perovskite copper histidine mof-177 aripiprazole quercetin perylene phthalocyanine ibuprofen symmetry ammonium%20acetate caffeine ferrocene mil-125 van zinc
 sulfathiazole coronene 4-cyan-4%27-pentyl-diphenyl anthracene mil pyrene %ce%b2-cyclodextrin diclofenac
 melamine ampicillin resveratrol carbon%20dioxide betulinic%20acid alanine phenanthrene mil-88b iron%20oxide ciprofloxacin
 paclitaxel imof-74 acetylsalicylic%20acid quartz carbazole magnetite celecoxib adenine triphenylphosphine
 hydroxyapatite thiourea tryptophan

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 zinc%20oxide sulfuric%20acid tetracene succinic%20acid oxalic%20acid
pcn-222 **mil-53(al)** imidazole ruthenium **hkust** **mil-88** indomethacin metformin %ce%b3-cyclodextrin
 phenol alumina nickel naproxen pentacene proline terephthalic%20acid glycine **mil-100** biphenyl polyethylene mannitol
 tartaric%20acid cellulose graphite porphyrin erythritol nacl iron **zif** urea **mil-53** **fe** **zno** **hkust-1** **zif-90** **irmof** carbon
 carbamazepine benzene **ui0-66** **mof-5** **mof-74** **mof** naphthalene **nu-1000** **titanium** **cu-btc**
irmof-1 rubrene **al2o3** methane **tio2** **ui0-67** **mofs** **zif-67** theophylline **cl-20** lactose
pcn-224 **mos2** silicon heptane **zif-7** **paracetamol** acetaminophen glucose **mof-808** ethanol
 phenylalanine piroxicam **c60** **perovskite** copper **ibuprofen** **zif-8** caffeine ferrocene **mil-125** van zinc
 cholesterol gold aripiprazole **perylene** **phthalocyanine** **symmetry** ammonium%20acetate diclofenac
 histidine **mof-177** **quercetin** curcumin anthracene **mil** **pyrene** %ce%b2-cyclodextrin salicylic%20acid pyridine
 sulfathiazole **melamine** **coronene** **4-cyan-4%27-pentyl-diphenyl** **betulinic%20acid** alanine phenanthrene **mil-88b** iron%20oxide ciprofloxacin
paclitaxel **ampicillin** **resveratrol** **carbon%20dioxide** **acetylsalicylic%20acid** **quartz** magnetite celecoxib adenine
irmof-74

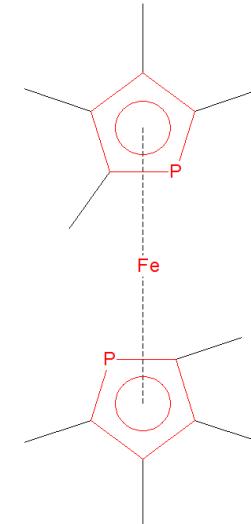
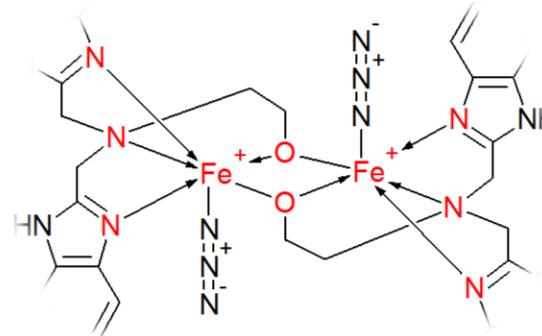
Challenges

- 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



Reliable input representations

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



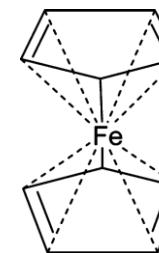
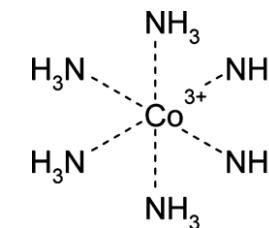
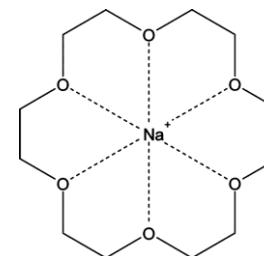
MOL V3000							
1	=	single					
2	=	double					
3	=	triple					
9	=	coordination					
10	=	hydrogen					
* excluding query bond types							

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

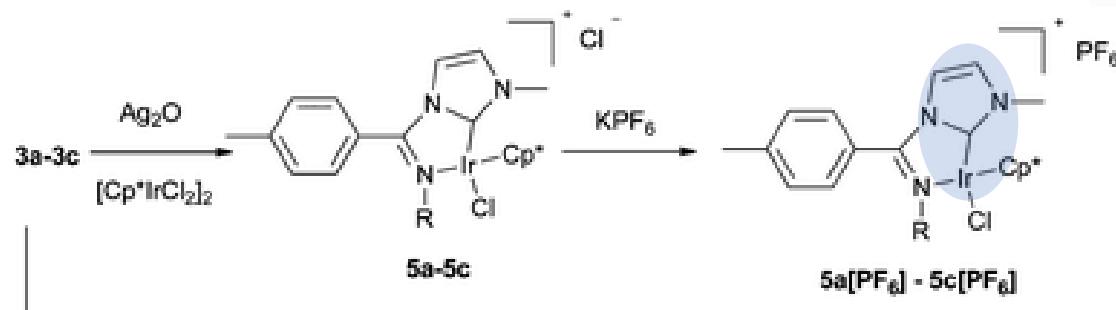
M	ZZF	3	1	41	2	42	3	43
M	ZZH	1	5	2	3	4	5	6
M	ZZH	2	5	7	8	9	10	11
M	ZZH	3	5	12	14	15	16	17
M	ZZE	2	42	18	43	18		



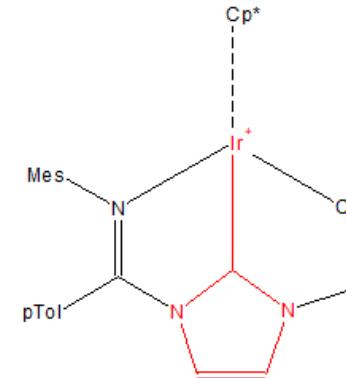
CCDC

Consistent structure representation

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



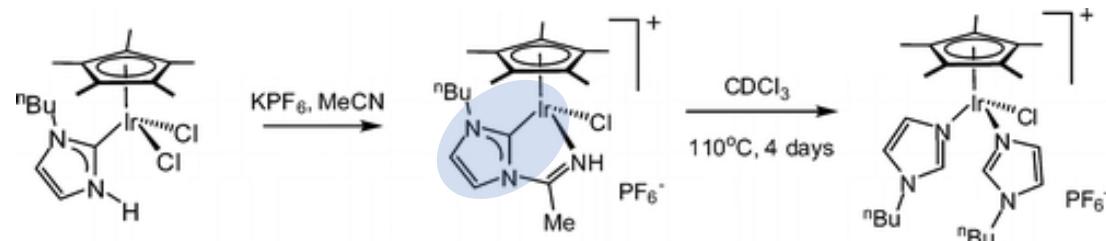
Dalton Trans., 2012, 41, 14557-14567, doi:10.1039/C2DT31989F



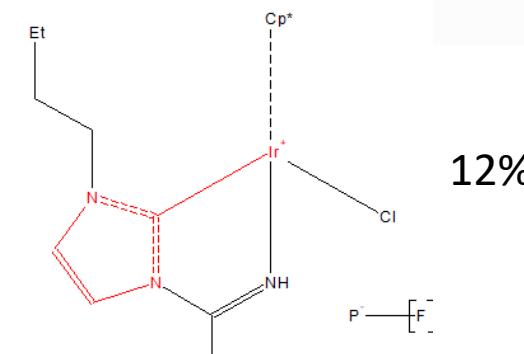
CSD

85%

ECIWUK CCDC:872879 10.5517/ccy99dx



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



12%

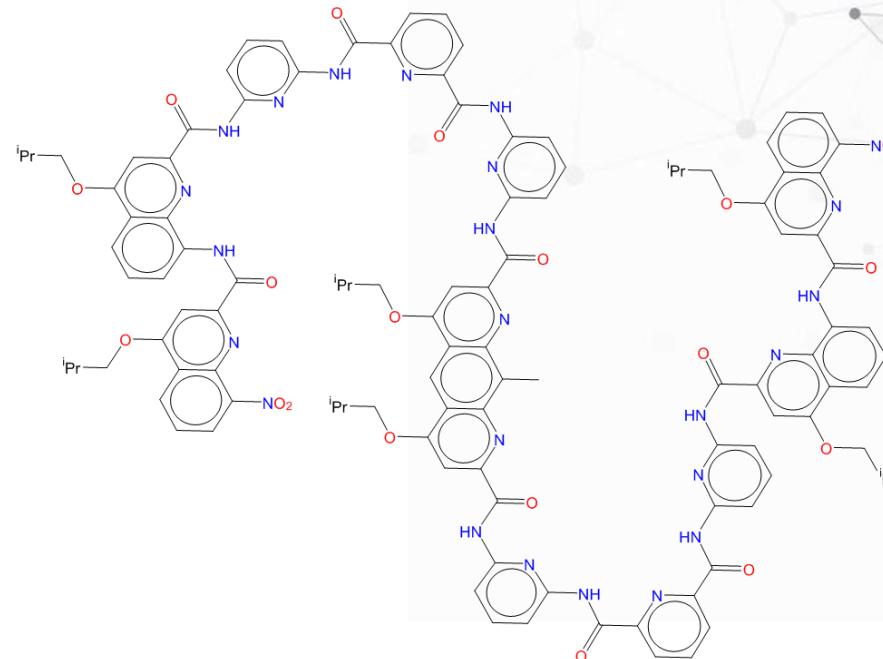
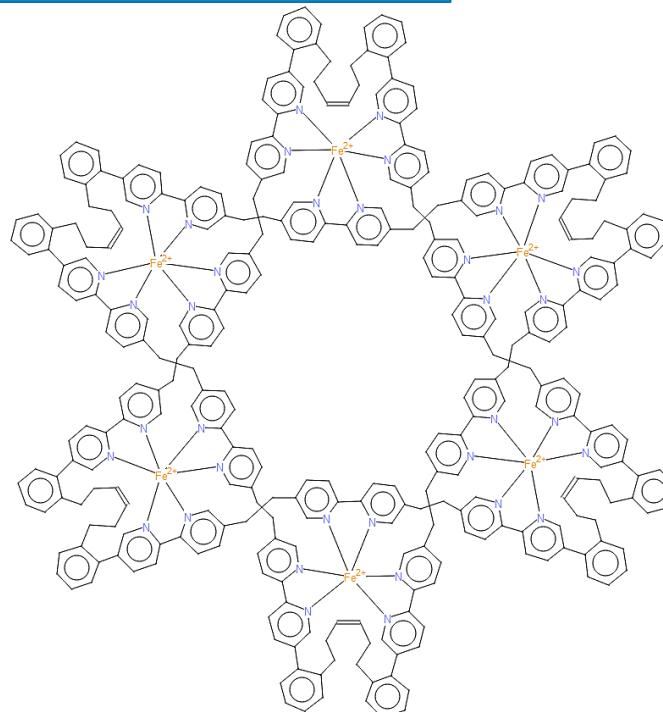
LIMXAH CCDC:664254 10.5517/ccq96kr

CCDC

Challenges – Cyclic and large structures

[2]-bis(μ^6 -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



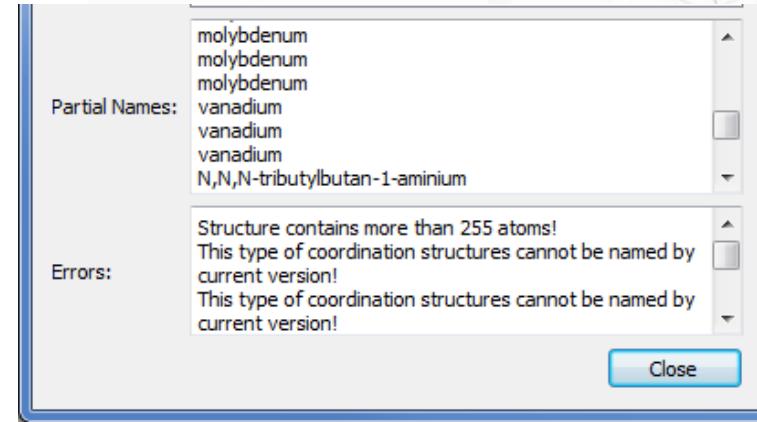
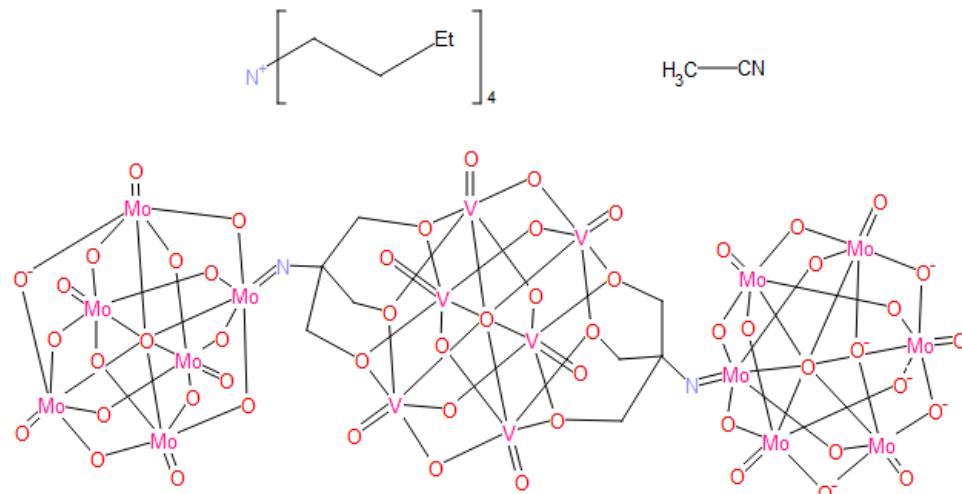
N,N'-bis(6-(6-(8-(4-isobutoxy-8-nitroquinoline-2-carboxamido)-4-isobutoxyquinoline-2-carboxamido)pyridine-2-carbamoyl)pyridine-2-carboxamido)pyridine-2-yl) 1,8-diaza-4,5-di-isobutoxy-9-methyl-2,7-anthracenedicarboxamide

Difficulty naming structures with more than 15 rings or 255 atoms

DONCUG, BOFEE

CCDC

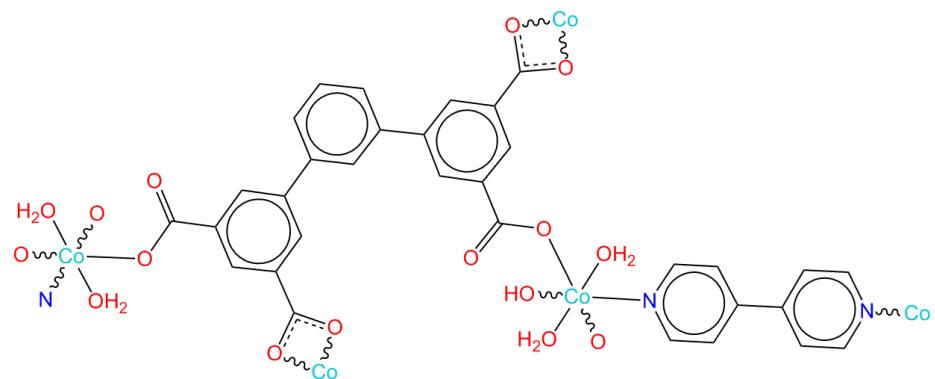
Challenges - Clusters



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

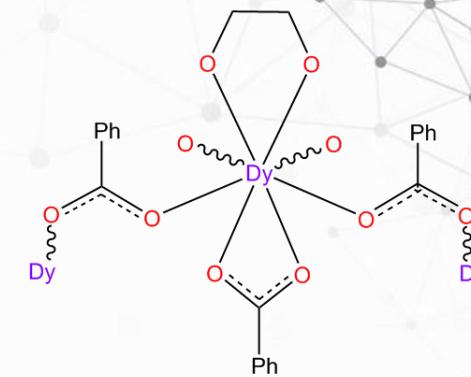
Challenges - Polymers

- 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]

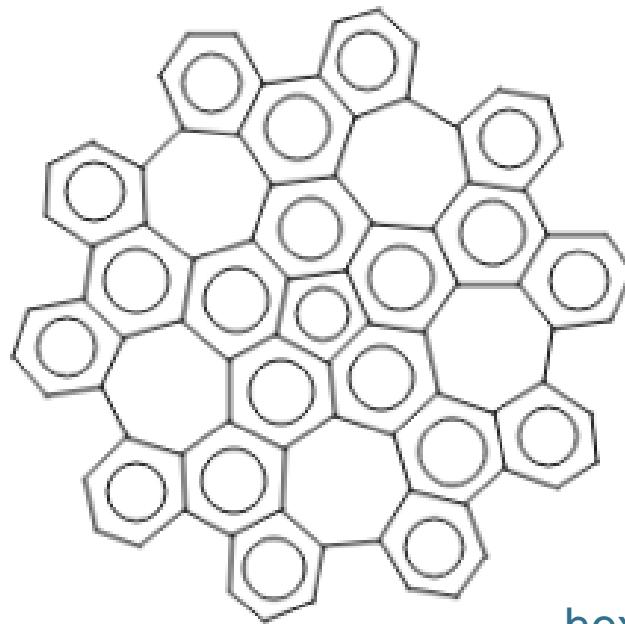


catena-[bis(μ_2 -benzoato)-(benzoato)-(ethane-1,2-diol)-dysprosium]

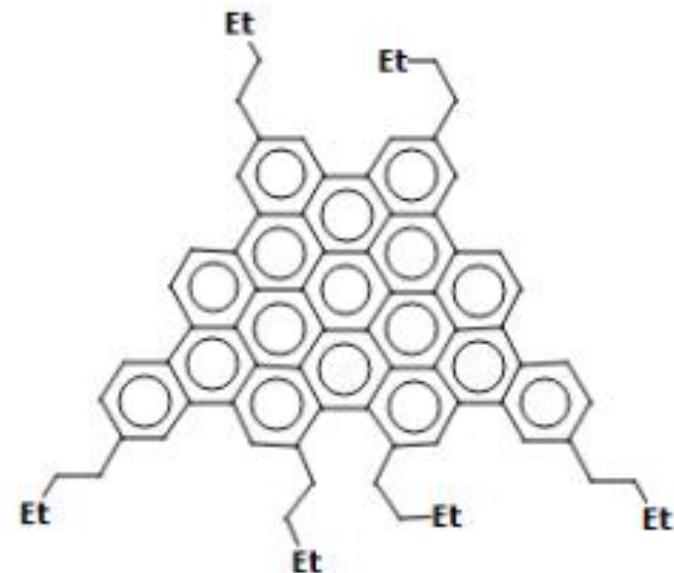
- ACD occasionally attempts to create a whole name, however it doesn't always
 - recognise polymer bonds
 - give stoichiometry of ligands or anions

New types of structures

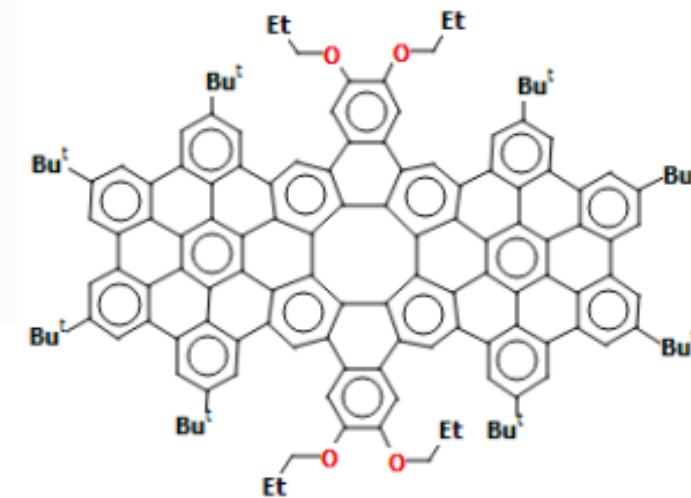
- Graphene-type structures



Nanographene



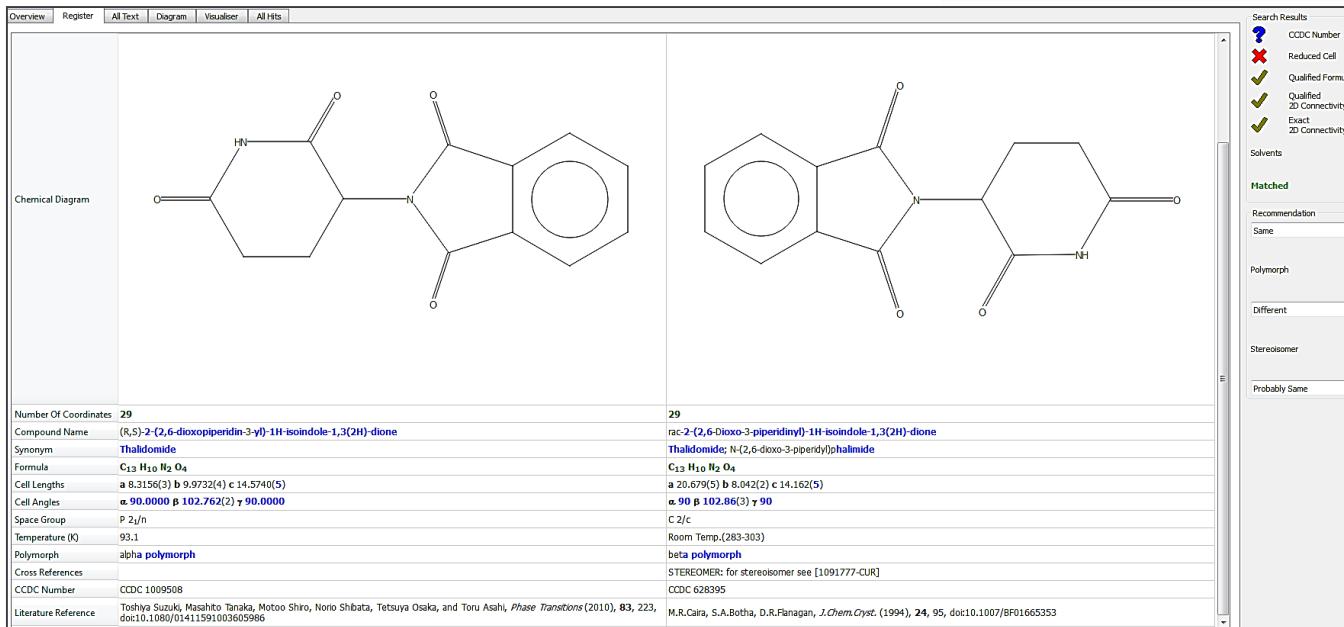
hexa-n-butyl-C54-nanographene



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

Stereochemistry

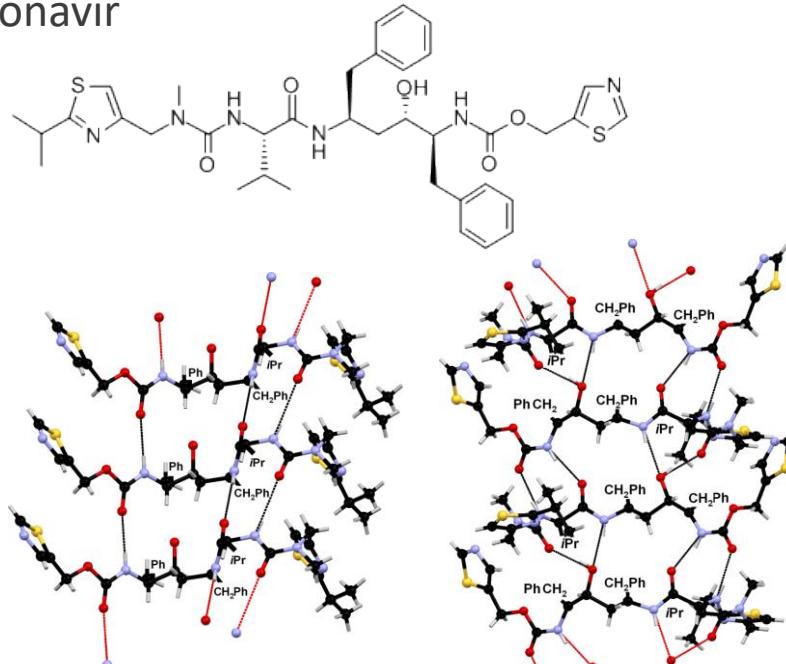
- CSD add stereochemical descriptors if given by the author in the cif or paper
- Group families together of same compound
- Link by cross-referencing diastereomers/enantiomers, but not same group



Polymorphism

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

Ritonavir



YIGPIO

Metastable monoclinic
polymorph 1

YIGPIO01

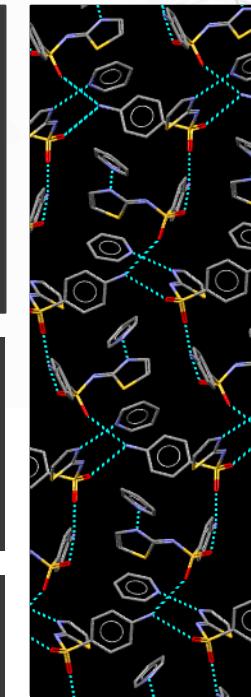
Stable orthorhombic
polymorph 2

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

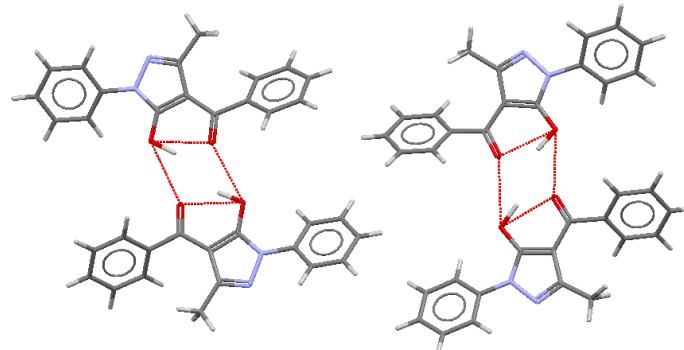
CUIMDZ
blue J I2/a polymorph
blue B trigonal polymorph
green orthorhombic polymorph
olive C2/c polymorph

ADEDIX
monoclinic disappearing polymorph
tetragonal polymorph

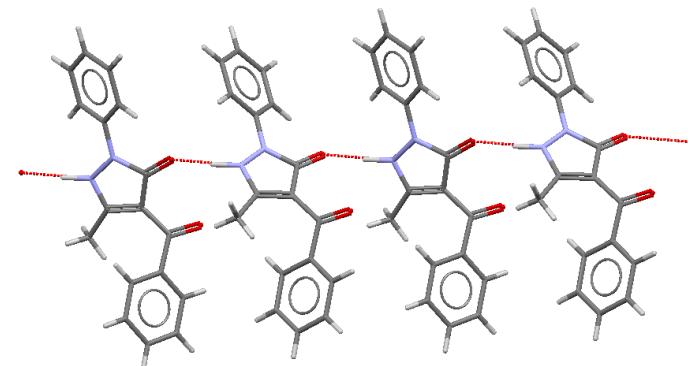


Tautomeric polymorphism

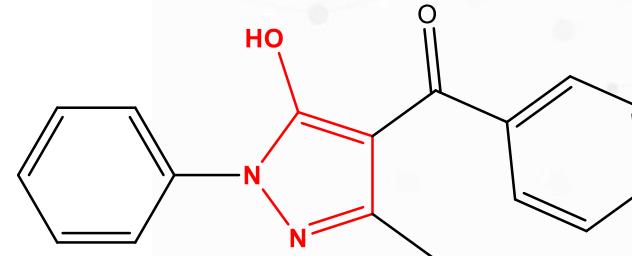
Different tautomers – different crystal forms – different interactions – different stability



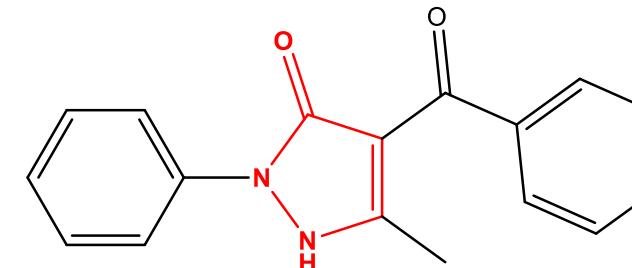
YUYDOL



DEBFAR

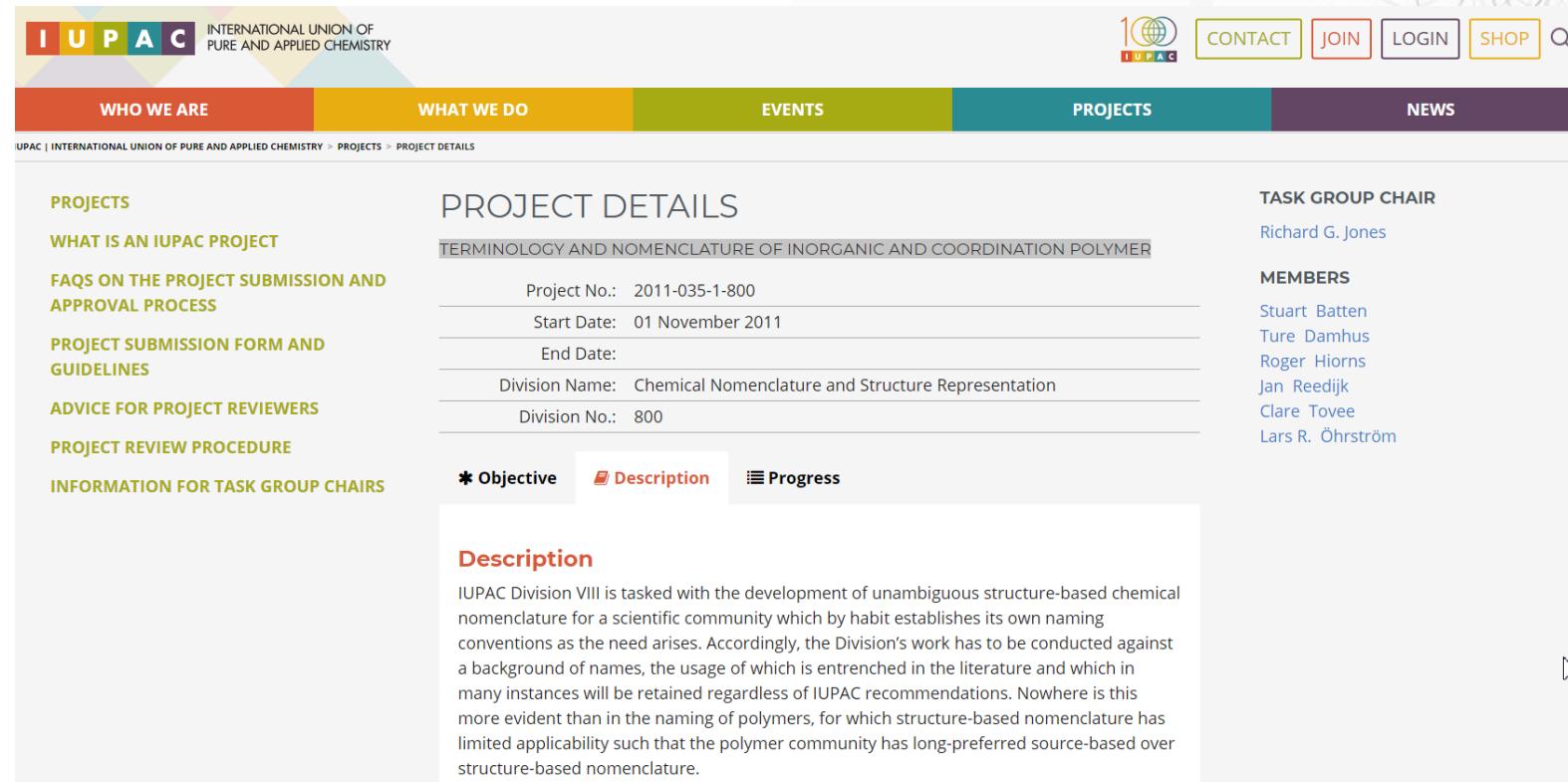


Less stable tautomers can sometimes form more stable hydrogen bonded networks



Changes in naming recommendations

- Naming coordination polymers



The screenshot shows the IUPAC website with a navigation bar at the top. The 'PROJECTS' tab is selected. The main content area displays project details for the 'Terminology and Nomenclature of Inorganic and Coordination Polymer' project, including project number, start date, end date, division name, and division number. A 'Description' section provides a detailed explanation of the project's objective.

PROJECT DETAILS

TERMINOLOGY AND NOMENCLATURE OF INORGANIC AND COORDINATION POLYMER

Project No.: 2011-035-1-800

Start Date: 01 November 2011

End Date:

Division Name: Chemical Nomenclature and Structure Representation

Division No.: 800

Description

IUPAC Division VIII is tasked with the development of unambiguous structure-based chemical nomenclature for a scientific community which by habit establishes its own naming conventions as the need arises. Accordingly, the Division's work has to be conducted against a background of names, the usage of which is entrenched in the literature and which in many instances will be retained regardless of IUPAC recommendations. Nowhere is this more evident than in the naming of polymers, for which structure-based nomenclature has limited applicability such that the polymer community has long-preferred source-based over structure-based nomenclature.

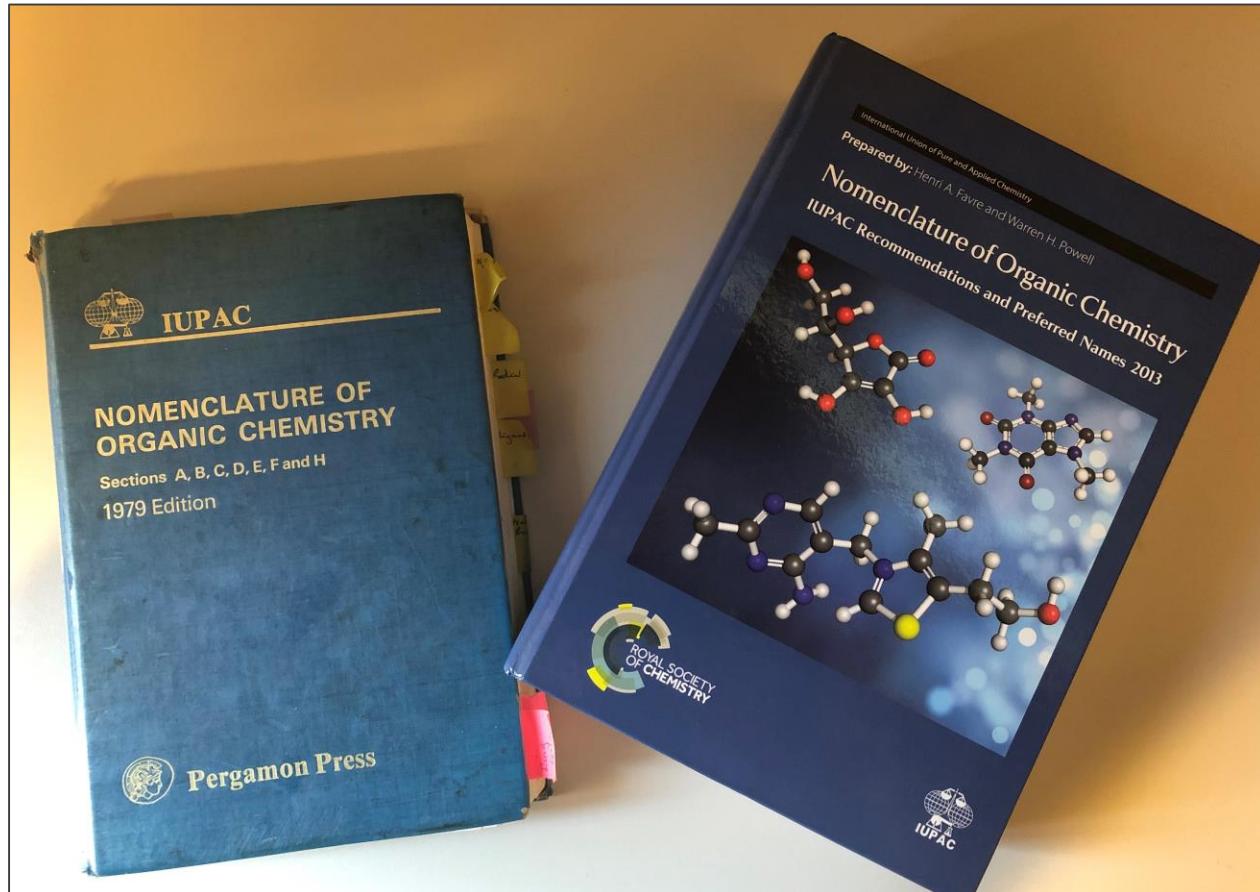
https://iupac.org/projects/project-details/?project_nr=2011-035-1-800

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

Changes in nomenclature rules



<https://doi.org/10.1039/9781849733069>

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

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 $\text{HO}-\text{N}(\text{O})=$ 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 $\text{H}_2\text{N}-\text{NH}-$; 'hydrazinylidene' for $\text{H}_2\text{N}-\text{N}=$; 'hydrazinediylidene' for $=\text{N}-\text{N}=$; and hydrazine-1,2-diyl for $-\text{NH}-\text{NH}-$. The prefixes 'hydrazino', 'hydrazono', 'azino' and 'hydrazo', respectively, are no longer acceptable, even for general nomenclature.

(k) The prefixes 'benzyl', 'benzylidene' and the like

Changes in nomenclature rules

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

Naming of topologies for Metal Organic Frameworks

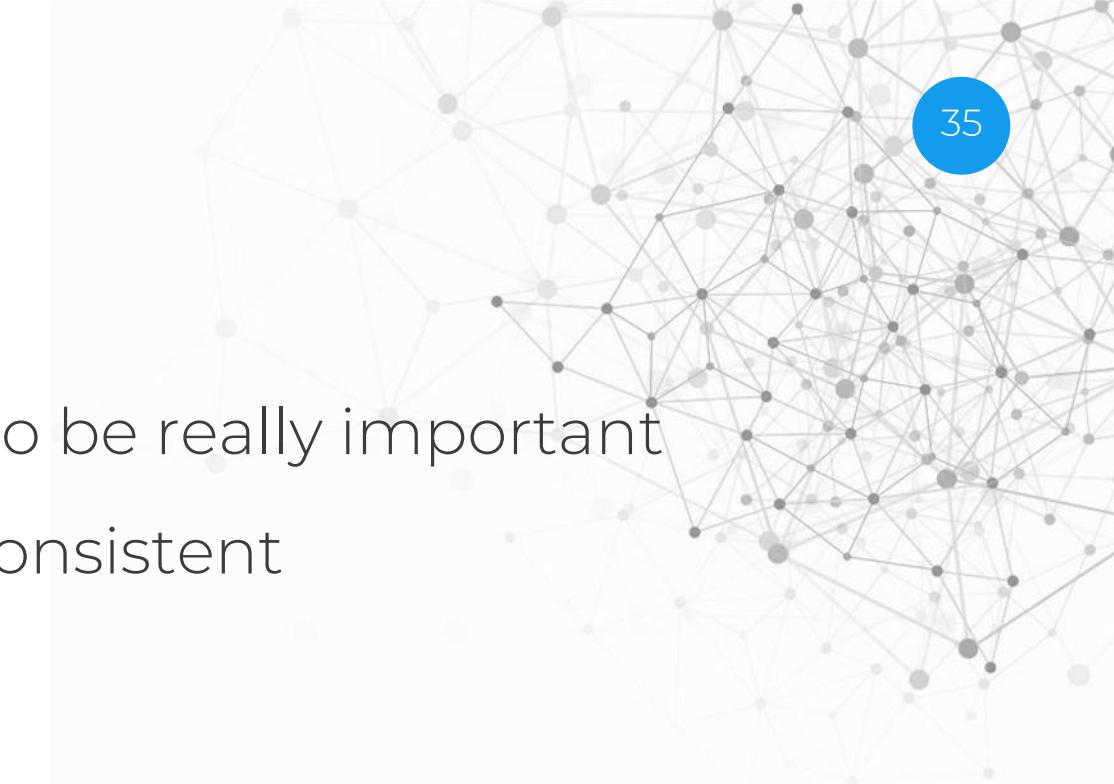
- IUPAC task force
 - Challenge is to come up with a reliable and consistent topology
 - Different topologies for the same structure can be assigned based on the method used
 - Task force looking at trying to gain a community consensus on the best approach

The screenshot shows the IUPAC website with a navigation bar including 'WHO WE ARE', 'WHAT WE DO', 'EVENTS', 'PROJECTS', and 'NEWS'. The 'PROJECTS' menu is active. Below the navigation is a search bar and a 'PROJECT DETAILS' section for a specific project. The project title is 'TERMINOLOGY GUIDELINES AND DATABASE ISSUES FOR TOPOLOGY REPRESENTATIONS IN COORDINATION NETWORKS, METAL-ORGANIC FRAMEWORKS AND OTHER CRYSTALLINE MATERIALS'. The project number is 2014-001-2-200, it started on 01 June 2014, and the division is Inorganic Chemistry Division (Division No. 200). The 'Description' section notes that network solids are important materials like diamond and quartz, and also mention metal-organic frameworks (MOFs) and coordination polymers. The 'TASK GROUP CHAIR' is Lars R. Öhrström, and the 'MEMBERS' listed include Stuart Batten, Vladislav A. Blatov, Jean-Guillaume Eon, Javier García Martínez, Stephen Hyde, Myoung Soo Lah, Michael O'Keeffe, Davide Proserpio, and Seth B. Wiggins.

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



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

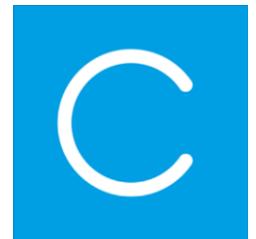
<http://www.ccdc.cam.ac.uk/>



@ccdc Cambridge



ccdc.cambridge



CCDC