

# Should we remediate small molecule structures? If so, who should do it?

**Carl Schwalbe**  
Aston University

Suzanna Ward, Simon Coles and Natalie Johnson

**The Cambridge Crystallographic Data Centre**  
**University of Southampton**

ECM32 – 18<sup>th</sup> August 2019

# Carl Schwalbe 1942-2019



[https://www.iucr.org/gallery/2009/aca-09?result\\_42723\\_result\\_page=24](https://www.iucr.org/gallery/2009/aca-09?result_42723_result_page=24)  
<https://www.amercrystalassn.org/assets/RefleXions/FALL2017RS.pdf>  
<https://www2.aston.ac.uk/lhs/staff/az-index/schwalch>

# A fine researcher

- The CCDC
  - 2010-2019 Emeritus Research Fellow
- Aston University
  - 2010-2019 Emeritus Professor of Medicinal Chemistry
  - 2007-2010 Professor of Medicinal Chemistry,
  - 1979-2007 Senior Lecturer in Medicinal Chemistry,
  - 1972-79 Lecturer in Medicinal Chemistry
- Max Planck Institute for Experimental Medicine
  - 1970-72 Research Fellow, (PI, Prof. W. Saenger)
- Harvard University
  - 1965-70 PhD., (PI, Prof. William N. Lipscomb)
  - 1964 AM
- Oberlin College
  - 1959-63 AB, Chemistry (summa cum laude)

<https://www2.aston.ac.uk/lhs/staff/az-index/schwalch>

## Professor Carl H. Schwalbe

**Emeritus Professor of Medicinal Chemistry**

Aston Pharmacy School  
School of Life & Health Sciences  
Aston University  
 Aston Triangle  
 Birmingham B4 7ET  
 UK

**email:** [c.h.schwalbe@aston.ac.uk](mailto:c.h.schwalbe@aston.ac.uk)

**telephone:** +44 (0) 121 204 4009

**fax:** +44 (0) 121 204 5142

Member of the Medicines Research Unit

Member of research strand: 'Development of novel drugs, formulations and tissues to support healthy ageing' - [Aston Research Centre for Healthy Ageing](#)

Member of the Pharmacy and Biology Teaching Programmes

### Research Interests

+

- Determination of molecular structure by X-ray crystallography
- Theoretical calculation of electronic and conformational properties
- Relation of structural and electronic features to activity of drugs
- Correlation of solid-state intermolecular interactions with physico-chemical properties of drugs in solid dosage forms



CCDC

# A stalwart of the BCA

## Honorary Members of the BCA

Honorary Life Membership is the BCA's highest membership accolade. The award is made in recognition of significant contributions by the recipient to crystallographic science and to the work of the BCA... Council normally accord Honorary Membership to a maximum of two people in one calendar year.

## Professor Carl Schwalbe (2018)

**Crystallography News**  
British Crystallographic Association

Issue No. 144 March 2018  
ISSN 1467-2790

**Spring Meeting in Warwick soon; Group Meetings held**

BCA Spring Meeting	p6
BCA Group Meeting Reports	p10
German Crystallographic Society Report	p16
Pan-African Conference Report	p18
Important Announcements	p20

### From the Editor



**OUR Spring Meeting is rapidly approaching. The details in this issue and on the website should convince anyone who is still undecided that this is a meeting which simply cannot be missed. Our cover shows 3 scenes from the University of Warwick, starting with the Sciences Building, the venue for our scientific sessions.**

**Meetings I promised you ducks; there are water features at Warwick that might attract them. I thank Amy Stares for the BSG Meeting pictures and Cheryl Doherty for the IUCrCG meeting picture.**

Two other meetings should inspire you to put fingers to keyboard and dash off a couple of brilliant abstracts. When I started in crystallography, I'd have said "put pen to paper". I suppose some of you will now put fingers to touchscreen. I still wince slightly when I do this, remembering how I used to tell off my students when they touched the screen of my precious Silicon Graphics workstation with greasy fingers.

But I have digressed. The American Crystallographic Association meeting will take place from 20-24 July. For those of you who think that The Donald might not let you in, worry not! Because this year's venue is Toronto, you can expect a hearty Canadian welcome. Toronto is a beautiful city with a scenic skyline and unspoiled country nearby. Important deadlines are March 30 for abstracts, May 31 for Early Bird registration and June 18 for reserving a discounted room at the conference hotel. The topic of this year's Transactions Symposium is "Shining a Light on Structure-Based Drug Design".

A month later we'll be able to enjoy the European Crystallographic Meeting in Oviedo, Spain. Dates to enter into your diary are 22-27 August for the actual meeting, 22 April for Early Bird registration and 29 April for abstracts. Being located in the extreme north of Spain at an altitude of 80 to 700 m above sea level, Oviedo escapes the searing Spanish summer heat. In August the average high at the weather station is 23.2°C. These conditions, along with a compact historic district and pleasant parks, make Oviedo a good place to enjoy a brisk walk before or after spending time in lectures. The European Union has named this city on a list of the cleanest cities in Europe.

With glacial slowness I have produced a write-up of last year's meeting of the German Crystallographic Society (DGK). The timing does have one advantage. The first plenary lecture was given by Prof. **Ilme Schlichting** on the subject of "Protein structure and dynamics using X-ray free electron lasers". My brief summary cannot possibly do justice to a fascinating lecture on the hottest of hot topics, but I hope it will whet your appetite for this year's BCA Spring Meeting, where the BSG plenary lecture will be given by none other than... Ilme Schlichting. As for this year's DGK meeting, you have just enough time to book last minute travel tickets and pack your bag. While the 2017 meeting was held in late March, this year it moves forward to 5-8 March. Furthermore, while last year's venue was Karlsruhe in the balmy wine-growing southwest of

Germany, this year it is Essen with its more bracing climate. The evolution of Essen is fascinating. It has gone from a powerhouse of heavy industry to habitat to a burgeoning new centre of culture and high tech. Last year it was selected as the European Green Capital.

Other meeting reports are featured in this issue too. In November 2017 our Industrial and Chemical Crystallography Groups joined forces for an Autumn Meeting on a topic, "Design of Crystalline Products" that was relevant to both groups. The presentations revealed cutting-edge research in the determination and prediction of crystal structures along with its great technological importance. Just a week before Christmas our Biological Structures Group held its Winter Meeting. Perhaps to blow away some excessively frothy Christmas cheer, they introduced a note of anguish as well as exultation with the title "The Joy and Pain of Structural Biology Research". Although the cover shows that they still had a good time, presentations by a stellar array of speakers demonstrated that the Latin motto "Per aspera ad astra" could well be applied to this area of research. Then the New Year got off to its customary mini-spring start in the form of the CCP4 Study Weekend in Nottingham. You may recall that the BCA contributed support for two participants to attend the 1st Pan African Conference of Crystallography in October 2016. Their reports in this issue show how much they benefited.

In this issue we have an obituary for **Terry Willis**, who did such important work in the development of neutron crystallography in the UK and promoted its use in collaboration with academic scientists including **Dorothy Hodgkin**. He also co-authored books which became classics and organised schools on neutron scattering. It is fitting that we pay tribute to his brilliance as a scientist and helpfulness to colleagues.

With sadness I read the death notice on the IUCr website for Professor **Alajos Kálmán**. He was a distinguished chemical crystallographer at the Hungarian Academy of Sciences, for whom ResearchGate lists 624 publications. I became acquainted with him through his important research on the structure of heterocyclic compounds, which matched interests of mine. Even while Hungary was ruled by communism, he always seemed ready to have a free-ranging scientific discussion. Those of you who have followed my presentations at recent BCA meetings will know that impossibly close H...H contacts in published structures are a concern of mine. Together with his colleagues **Peter Bombicz**, **Mályás Csenger** and **Roland Tolgmen**, in 2003 he published the structure with the then shortest known genuine C-H...H-C contact [1.949(7) Å], retrocode TACHSB04, verified by neutron as well as X-ray diffraction. Perhaps we should be grateful for the isolation imposed by the Iron Curtain. This polymorph (A) of 1,2,3,5-tetra-O-acetyl-β-D-ribofuranose was the first one prepared, in 1947, but by 1964 laboratories in western Europe, the USA and Australia could only grow a more stable polymorph (B). However, in 1981 polymorph A reappeared in Budapest and was carefully protected from contamination since then. Its excessively short H...H contacts may well be the reason why it is a disappearing polymorph.

Carl Schwalbe

**EX-OFFICIO MEMBERS**

**Editor**  
**"Crystallography News"**  
**Prof Carl H Schwalbe**  
15 St. Augustine Drive,  
Droitwich, Worcs, WR9 8QR  
Tel: 01905 775257  
carlschwalbe@hotmail.com



## Should we remediate small molecule structures? If so, who should do it?

Carl H. Schwalbe<sup>a,b</sup>

<sup>a</sup>Cambridge Crystallographic Data Centre, Cambridge, UK; <sup>b</sup>School of Life and Health Sciences, Aston University, Birmingham, UK

### ABSTRACT

Problems can arise in crystallographic databases with errors and omissions in the representation of data that impede searches, and with errors in the actual data. While the Cambridge Crystallographic Data Centre with its Improvement Projects has solved many of the first category of problems, errors in atomic coordinates and other crystallographic data are surprisingly common. Although modern software warns of many types of error, such errors appear even in recently deposited Crystallographic Information Files. Richard Marsh found many examples of missed symmetry in assignment of the space group; such errors are now waning. Hydrogen atoms are commonly placed in calculated positions. Particularly for OH and NH groups involved in hydrogen bonds, occupancy factors may need to be reduced to 0.5 or the hydrogen atom positions may require amendment. Examples of acids and imidazole derivatives are provided, showing that sometimes only consideration of bond distances and angles at the heteroatom can distinguish between OH or NH and unprotonated O or N. Significant work by other researchers correcting mis-positioned hydrogen atoms in dihydrogen phosphates and water aggregates as well as mis-identified elements is also summarized. This review concludes with some suggestions for more comprehensive detection and correction of errors in deposited data.

### ARTICLE HISTORY

Received 16 April 2018  
Accepted 1 August 2018

### KEYWORDS

Erroneous structures;  
carboxylic acids; imidazoles;  
preventing errors; rectifying  
errors

### 1. Introduction

X-ray crystallography is generally regarded as the 'gold standard' for structure determination. However, in 2011 David Watkin [1] posed the question 'Is the Gold Standard becoming tarnished?' Earlier, P. G. Jones [2] published a thought-provoking review advising chemists about aspects of a crystal structure determination that pose difficulties for chemical crystallographers, and ways to spot when work has been done erroneously. Particular issues arise with macromolecular crystallography. As always, it is important to document the data collection procedures; and, because the limited data-to-parameter ratio necessitates model fitting in most such structure determinations, these procedures

**CONTACT** Carl H. Schwalbe  [carlschwalbe@hotmail.com](mailto:carlschwalbe@hotmail.com)

 The re-refined version of the structure with CSD refcode Y0T1Y0W has been deposited with CCDC under deposition number 1850426. Also refer for further details <https://doi.org/10.1080/0889311X.2018.1508209>

© 2018 Informa UK Limited, trading as Taylor & Francis Group

# A recent article

This talk is partly based on Carl's thought-provoking article published in 2018 on the same topic:

- Should we remediate small molecule structures? If so, who should do it?

Carl H. Schwalbe

*Crystallography Reviews*

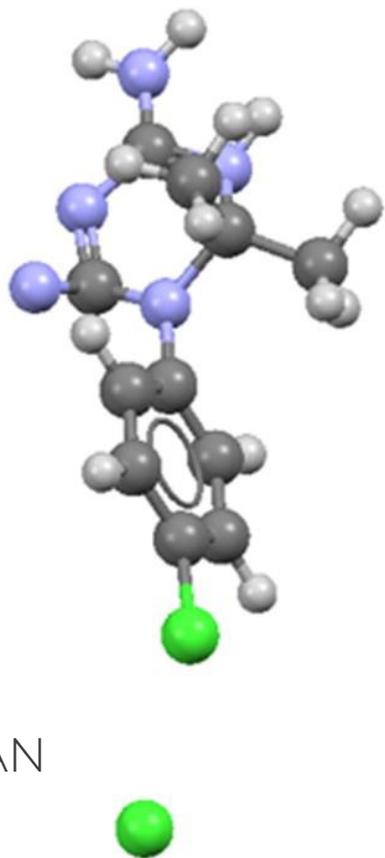
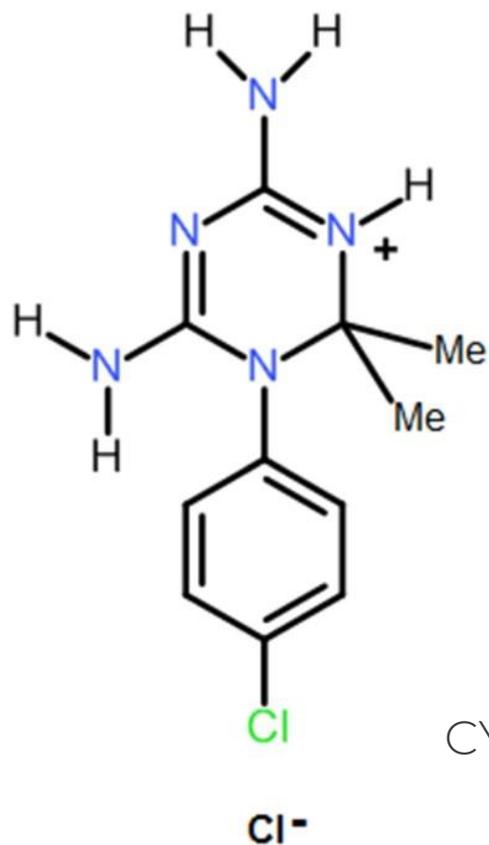
2018, 24 217-235

DOI: [10.1080/0889311X.2018.1508209](https://doi.org/10.1080/0889311X.2018.1508209)?

# Should we remediate small molecule structures?

- New reports of small molecule crystal structures should be error-free
  - Most reputable journals require validation of crystallographic data with CheckCIF
  - CheckCIF integrated into the CCDC deposition procedure
- Not all errors pointed out
  - Some journals appear to ignore or not use crystallographic referees
- What if authors are unable or unwilling to make corrections when required?
  - Should an otherwise correct structure be rejected because a hydrogen atom has been incorrectly placed?
  - Or disorder of a terminal methyl group has not been entered into the model?
- Should such a structure be published or deposited with a warning message, or should a corrected version be created?
- These questions have particular force with regard to already published structures that have errors

# When remediation goes wrong



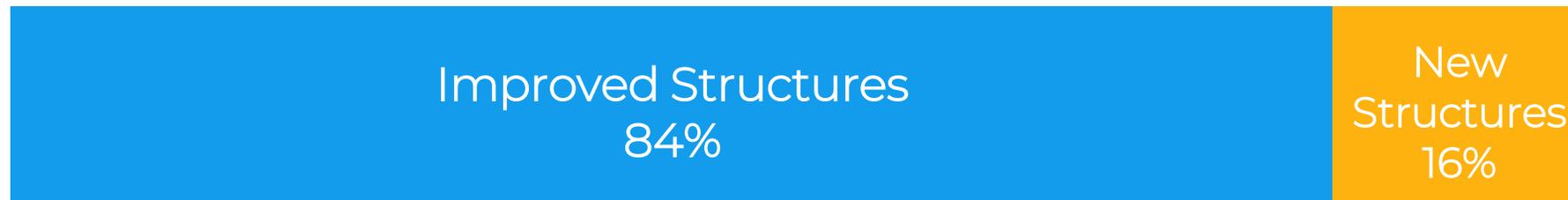
- Coordinates hand-typed & transcribed at CCDC
- CCDC checks to identify and correct typing errors
- Impossible bond distances corrected by:
  - Adding or deleting a minus sign
  - Transposing a pair of digits
  - Including a clear statement of what had been altered.
- CYGUAN and the unintended effect
  - The x-coordinate of amino N5 missing a minus sign
  - Change in N5 to C atom too small to be noticed
  - N5 to H's distances were too long and it was assumed the H atoms were wrong and they were deleted
- Saved by neutron diffraction - CYGUAN01 and Carl

# Crystallographic “vigilantes”

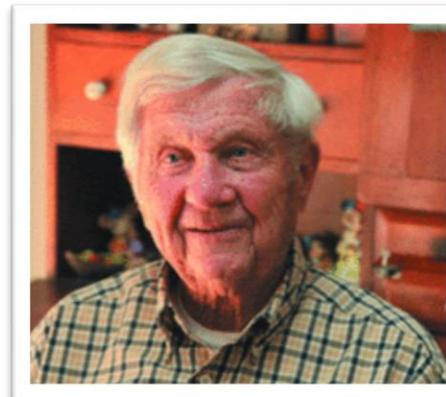
- Space group symmetry
  - R E. Marsh (2009), Acta Cryst. B65, 782-783
- Misplaced hydrogen atoms
  - I. Bernal & S. F. Watkins (2013), Acta Cryst. C69, 808-810.
  - C. H. Schwalbe (2016) Abstract 01.11.01.12, 66th ACA Annual Meeting, Denver. Acta Cryst. (2017). A73, a133 Should we remediate small molecule structures? If so, who should do it? Carl Schwalbe United Kingdom Aston University
- Misidentified atoms, misplaced H atoms, etc.
  - F. Fronczek, (2019) ACA Abstract. How to Remedy Incorrect Duplicates in the CSD?

# Space Group Symmetry – “Marshed”

- >1,350 structures



- Issues spotted predominantly were:
  - Missing inversion centres in a non-centrosymmetric structures
  - Other missing symmetry elements
- Leading to assignment of and refinement in the wrong space group
- Spotting and correcting these was non-trivial as
  - Data often only available from the printed supplementary pages
  - Data entered by hand
  - Structure re-refined in corrected space group



# Data integrity checks



A service of the  
International Union of Crystallography

checkCIF reports on the consistency and integrity of crystal structure determinations reported in CIF format.

Please upload your CIF using the form below.

File name:  
 No file chosen

Select form of checkCIF report

- HTML
- PDF
- PDF (recommended for CIFs that might take a long time to check)

Select validation type

- Full validation of CIF and structure factors
- Full IUCr publication validation of CIF and structure factors
- Validation of CIF only (no structure factors)

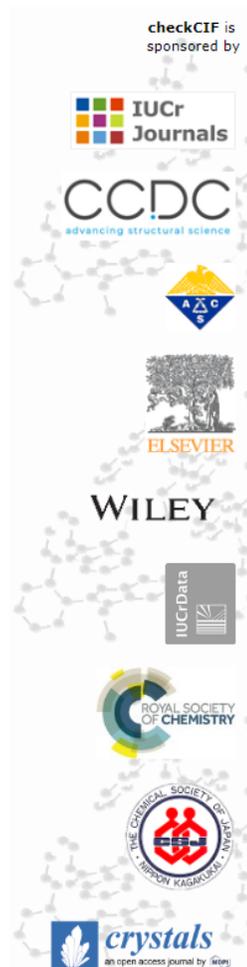
Output Validation Response Form

- Level A alerts only
- Level A and B alerts
- Level A, B and C alerts
- None

[Information about this version of checkCIF ...](#)

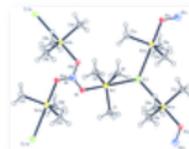
**Useful links**

**Prepublication check for submissions to IUCr journals**  
 Details of checkCIF/PLATON tests  
 CIF dictionary  
 Download CIF editor (pubCIF) from the IUCr  
 Download CIF editor (enCIFer) from the CCDC



## RESEARCH PAPERS

*Acta Cryst.* (2009). **D65**, 148-155  
<https://doi.org/10.1107/S090744490804362X>  
 Cited by **11515**



### Structure validation in chemical crystallography

A. L. Spek

Automated structure validation was introduced in chemical crystallography about 12 years ago as a tool to assist practitioners with the exponential growth in crystal structure analyses. Validation has since evolved into an easy-to-use *checkCIF/PLATON* web-based IUCr service. The result of a crystal structure determination has to be supplied as a CIF-formatted computer-readable file. The checking software tests the data in



OPEN ACCESS

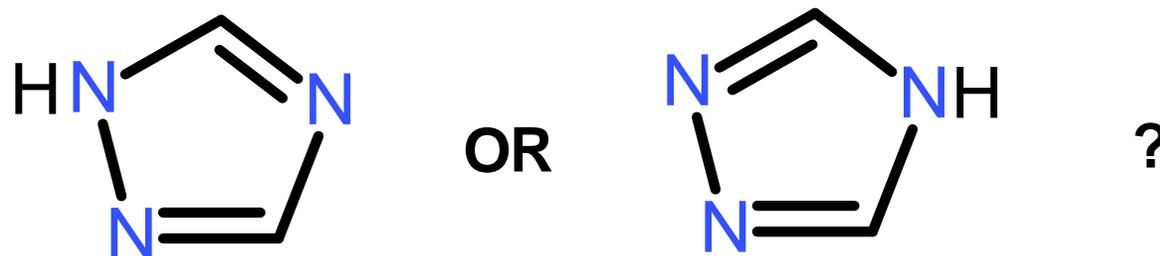
## My Crystallographic History- R.E.Marsh 2013

<http://www.amercrystalassn.org/h-marsh>

Fortunately, in the past few years they have decreased in number, thanks to computer programs such as CheckCIF and to slowly-successful pleas to journal editors to insure that authors make use of these programs. My recent surveys have suggested that the "wrong structure" disease may be getting close to extinction.

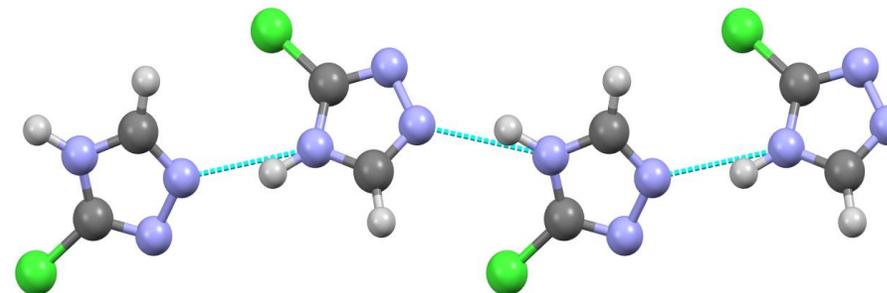
# Published data case study 1

## Tautomerism in triazoles



- Ab-initio calculations show 1H 6.25 kcal mol<sup>-1</sup> more stable
- CSD shows 203 1H vs 7 4H tautomer hits
- Do the 7 4H tautomers actually exist?!

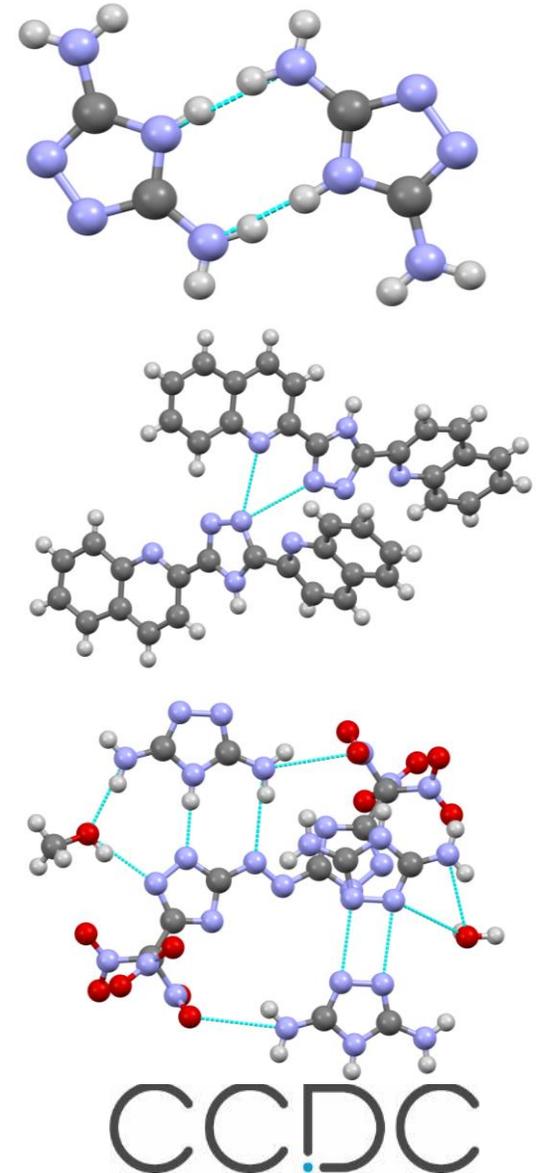
# Redetermination...



- Two 4H tautomers redetermined as 1H (CLTRZL & JUGYOB)
- These two pairs of structures enable evaluation of descriptors to establish tautomeric form
  - Electron density and H-bonding poor (1- and 4- positions link with each other into chains, so unclear which N is protonated)
  - Bond distances to N poor as similar distances for formally 'single' and 'double' bonds wrongly suggests N's are identical
  - Endocyclic bond angles good as VSEPR 'squeezes' angles at unprotonated N atoms, revealing identity

# Reinvestigation required

- DAMTRZ21 isostructural unit cell with 3 other 1H structures
  - CheckCIF Level A Alert about a D-H...H-D clash of 1.29 Å
  - Endocyclic angles clearly show that H atom should be on N1
- MAJSOH has no comparison, but...
  - CheckCIF Level C Alert that N4-H lacks an acceptor
  - Moving the H atom from N4 to N1 would make a bifurcated HB
- FALDAZ has 3 triazoles (two identical with missing H and third 4H)
  - Endocyclic angles suggest 1H tautomers throughout
  - A credible HB scheme can be created by reversing N4-H...N



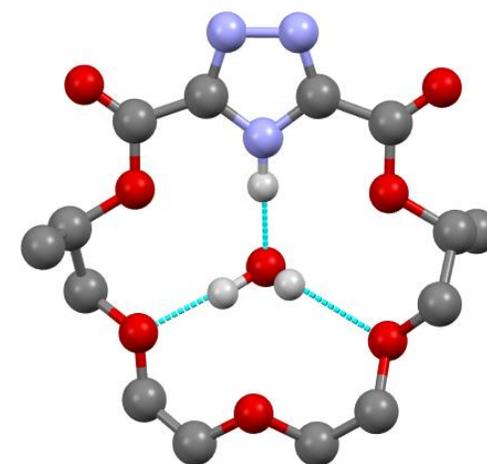
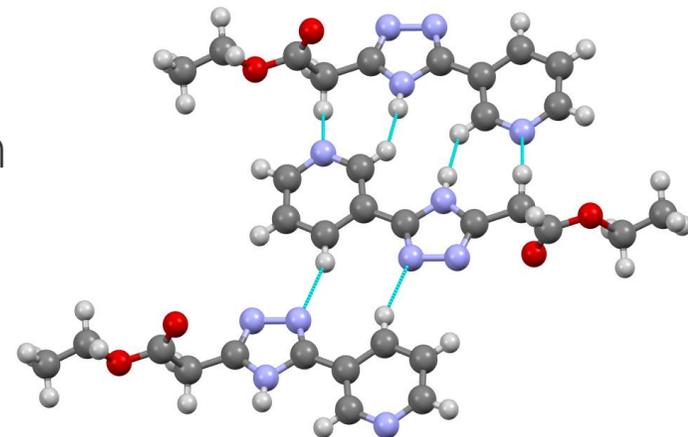
# Triazole structures continued...

- FUZPOH devoid of actual or potential N-H...N hydrogen bond

- Bond angles give a fairly weak indication of a  $1H$  tautomer
- Moving the H atom to N1 allows N2 and N4 to accept C-H...N HBs

- DEGNIM triazole incorporated into a crown ether

- Water molecule that can interact with triazole N and ether O atoms
- Endocyclic angles seem to contradict a  $4H$  tautomer but may be affected by attachment to the macrocycle
- The water molecule is significant
- With the  $4H$  tautomer as reported it can make three HBs
- A different tautomer would only allow it two HBs

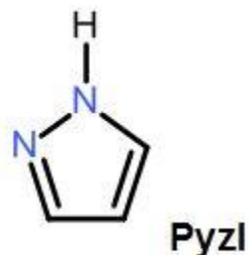
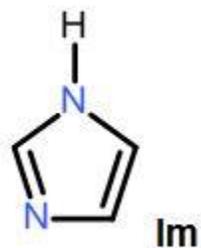


# Conclusions (triazole)

- Reported  $4H$  structures are rare and, with one exception, likely to be incorrect
- CheckCIF Alerts about N-H donors without acceptors or clashing N-H...H-X or N...N but otherwise silent about correct tautomer
- Bond distances for C-NH and C=N can be misleadingly similar
- Endocyclic bond angles, affected by VSEPR are useful to distinguish C-N(H)-X from C=N-X
- Need to evaluate trends in related structures to understand which descriptors to use for disambiguation

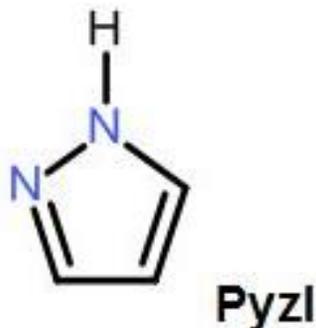
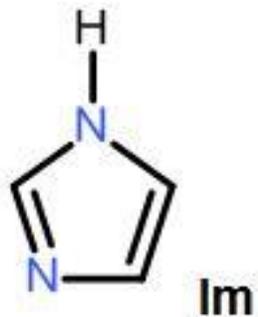
# Published data case study 2

## Misplaced H atoms and undetected disorder

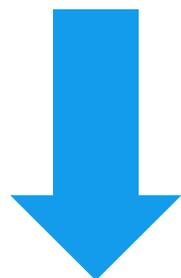


- The imidazole ring of histidine can participate in proton relays
  - Protonation sites may be obscure due to similarity of electron density between H-bonded NH...:N and N:...HN
  - Geometrical criteria can be more reliable
- Differences in Im C-NH and C=N bond length and C-NH-C and C=N-C bond angle are most significant (Malinska *et al.*, 2015)
  - Neutron diffraction on Im at 103 K shows 1.347, 1.322 Å and 107.1, 105.1°
  - In  $Z' = 2$  Pyzl exhibits some charge transfer between rings and NH/N disorder
  - At 100 K Pyzl C-NH and C=N distances are 1.338, 1.334 and 1.347, 1.330 Å; C-NH-N and C=N-NH angles are 112.2, 104.2 and 112.2, 104.5°

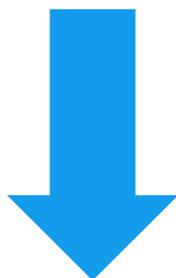
# Studying the CSD



CSD2018  
R ≤ 10%  
Organic  
No disorder



547 hits

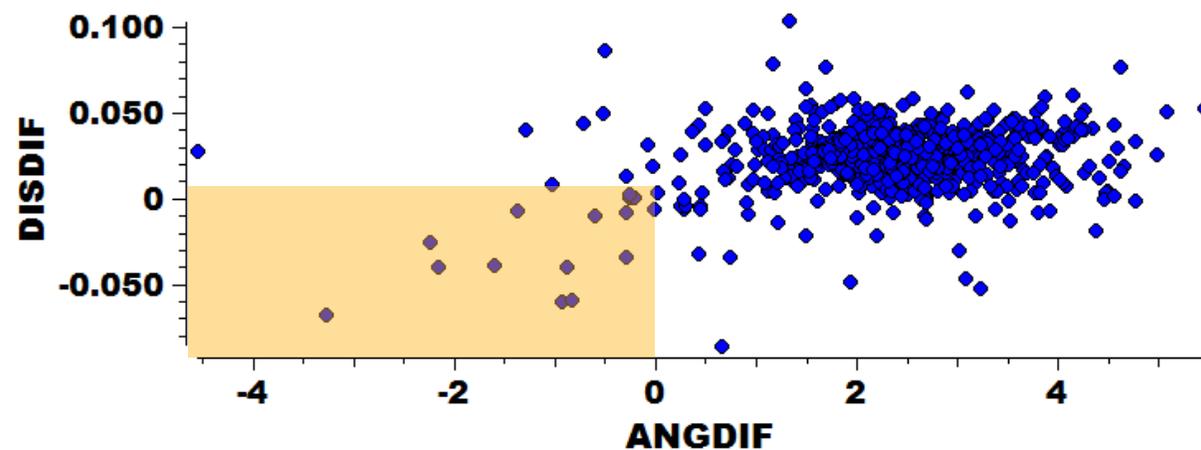


723 hits

## Imidazoles

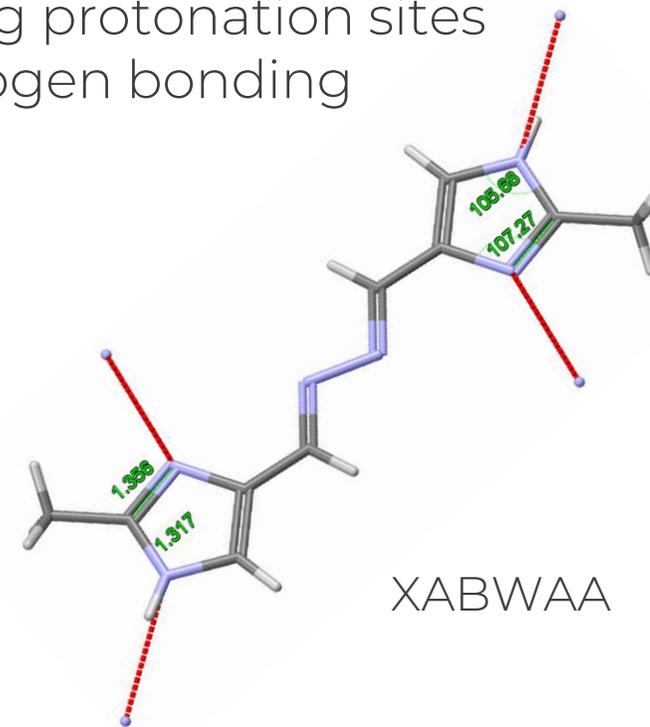
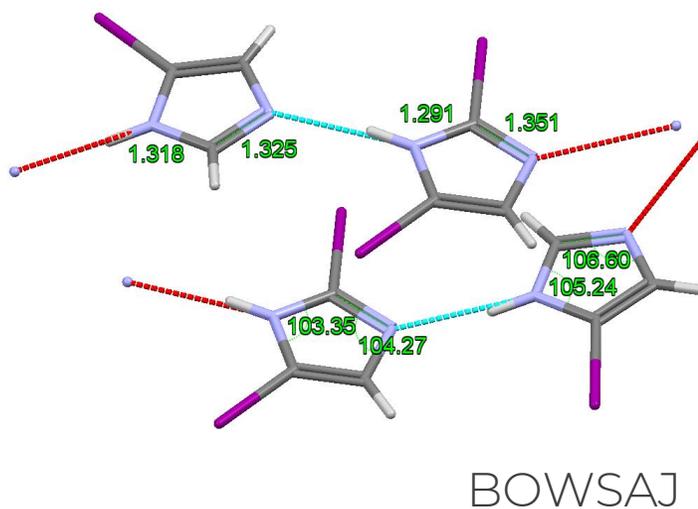
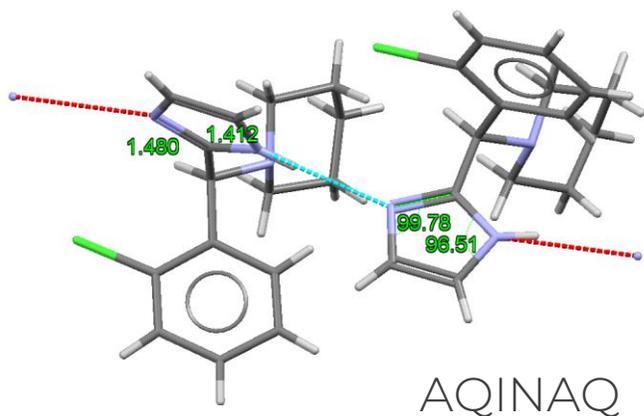
DISDIF - difference between C-NH  
and C=N bond distances  
ANGDIF - difference between angles

Mean DISDIF = 0.024(12) Å  
Mean ANGDIF = 2.3(8)°.



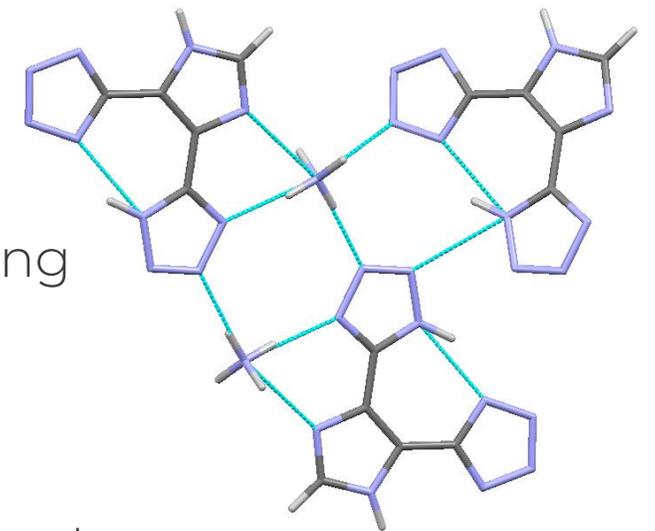
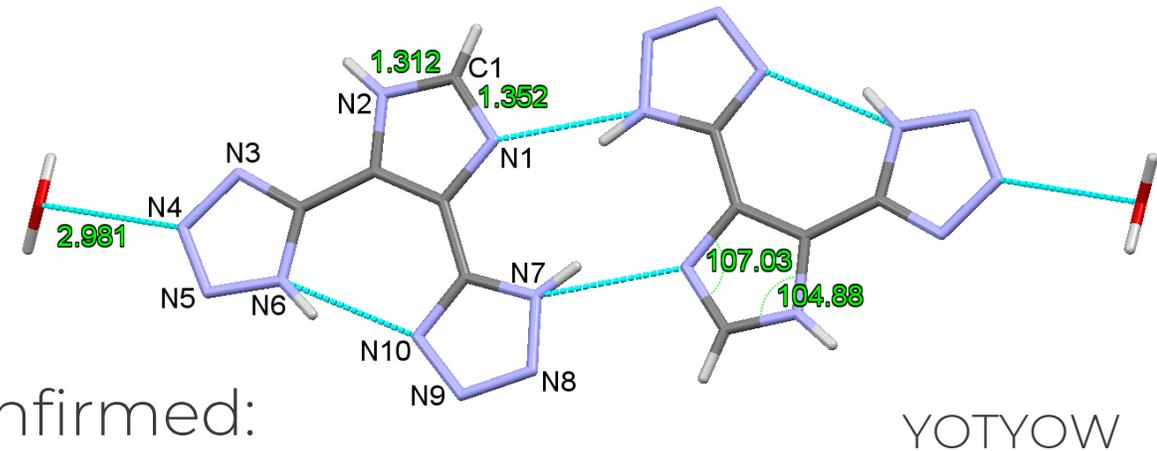
# Misplaced hydrogens

- 5 structures with large negative difference
  - In 3 with intermolecular NH...N interactions, exchanging protonation sites makes the differences positive and preserves the hydrogen bonding scheme



# A trickier structure

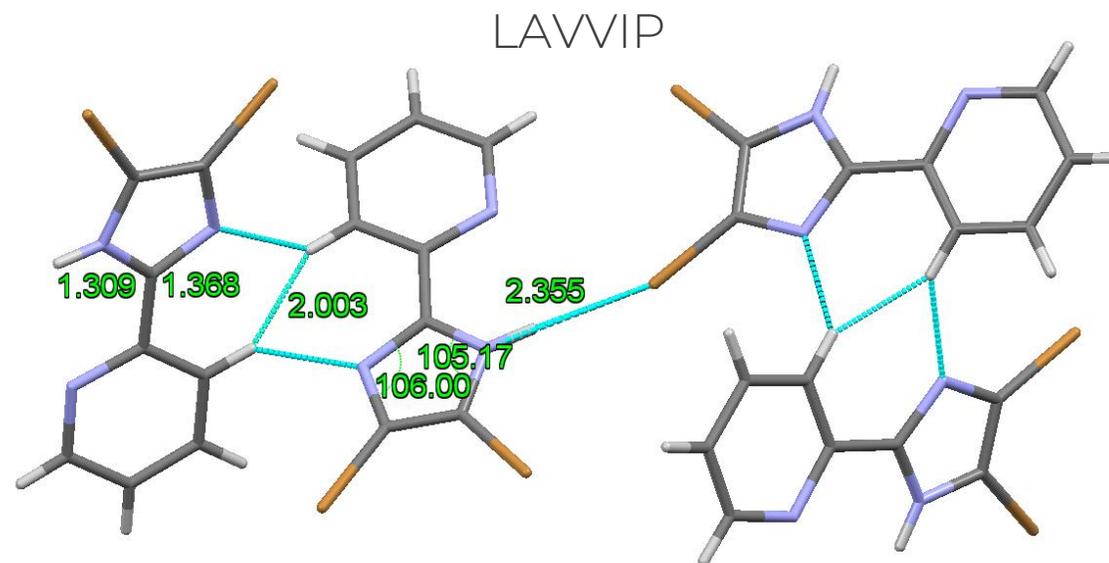
- Interchanging N and NH creates problems elsewhere
- Structure factors deposited
- Difference electron density map confirmed:
  - H atoms on tetrazole N6, Im C1 and Im N1
  - No H atom on N2
- Surprises!
  - No H atom appeared on N7 or elsewhere on tetrazole ring
  - Four H atoms surrounded the “water oxygen atom”
- An ammonium salt!
  - Chemical analysis required for unequivocal confirmation, but  $\text{NH}_4\text{Cl}$  was a reagent in the synthesis



CCDC

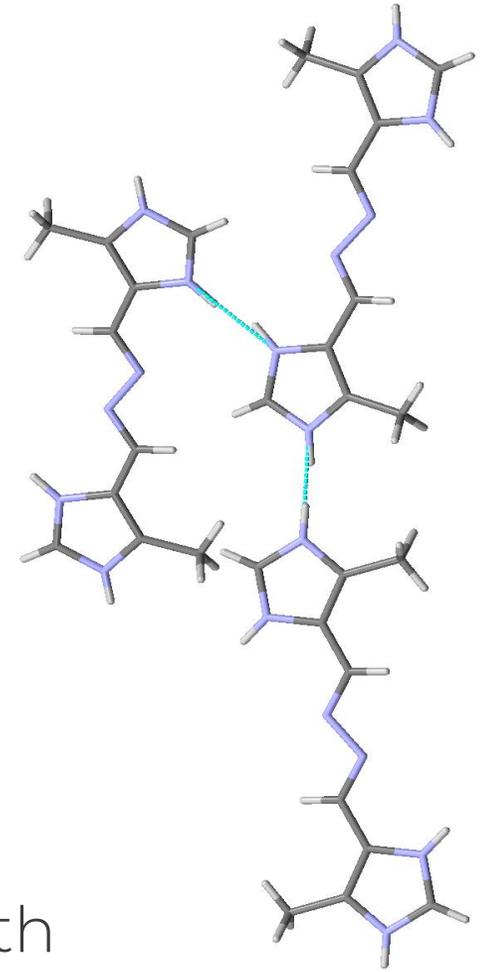
# The last misplaced H structure

- Differences of  $-0.059 \text{ \AA}$  and  $-0.83^\circ$  suggest protonation of the “wrong” imidazole N atom
  - Disconcertingly close contacts
  - Swapping N and NH on the imidazole ring worsens the log-jam of H atoms
- Rotating the pyridine ring by  $180^\circ$  about its link to imidazole appears to work
  - The new H atoms appear too close, but some relaxation might take place



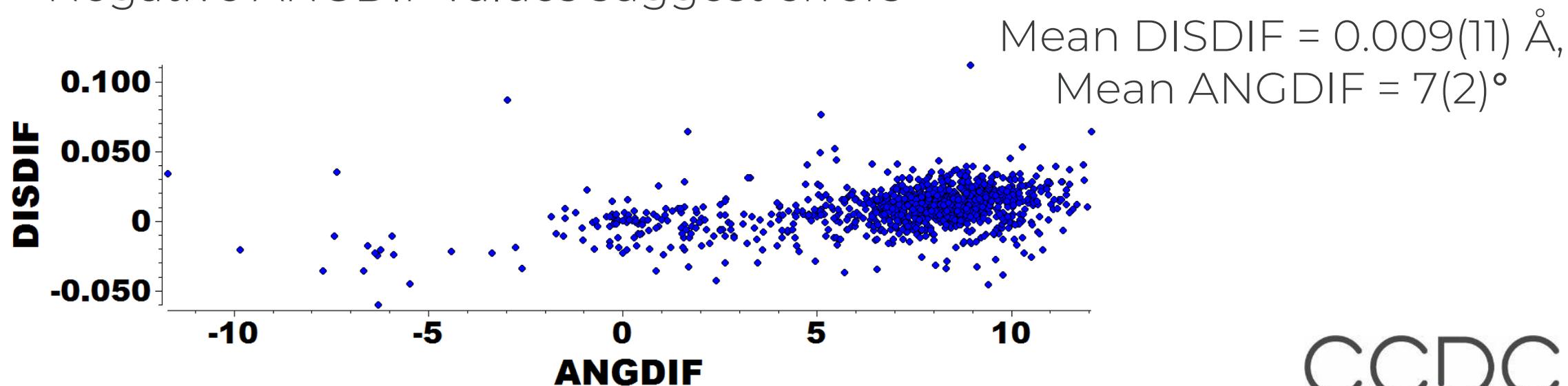
# Rotational disorder of imidazole rings

- DISDIF / ANGDIF plot shows points near origin with absolute values of differences much  $< 0.021 \text{ \AA}$  and  $2.4^\circ$
- Likely explanation is disorder, some rings having been rotated so as to interchange N and NH within the ring
- Known phenomenon - Drew *et al.* carefully compared possible tautomers of an imidazole structure with reference both to crystal structure and DFT calculations
- Packing requires 50:50 occupancy
- They cited 3 other structures which had been refined with 50:50 disorder of tautomers

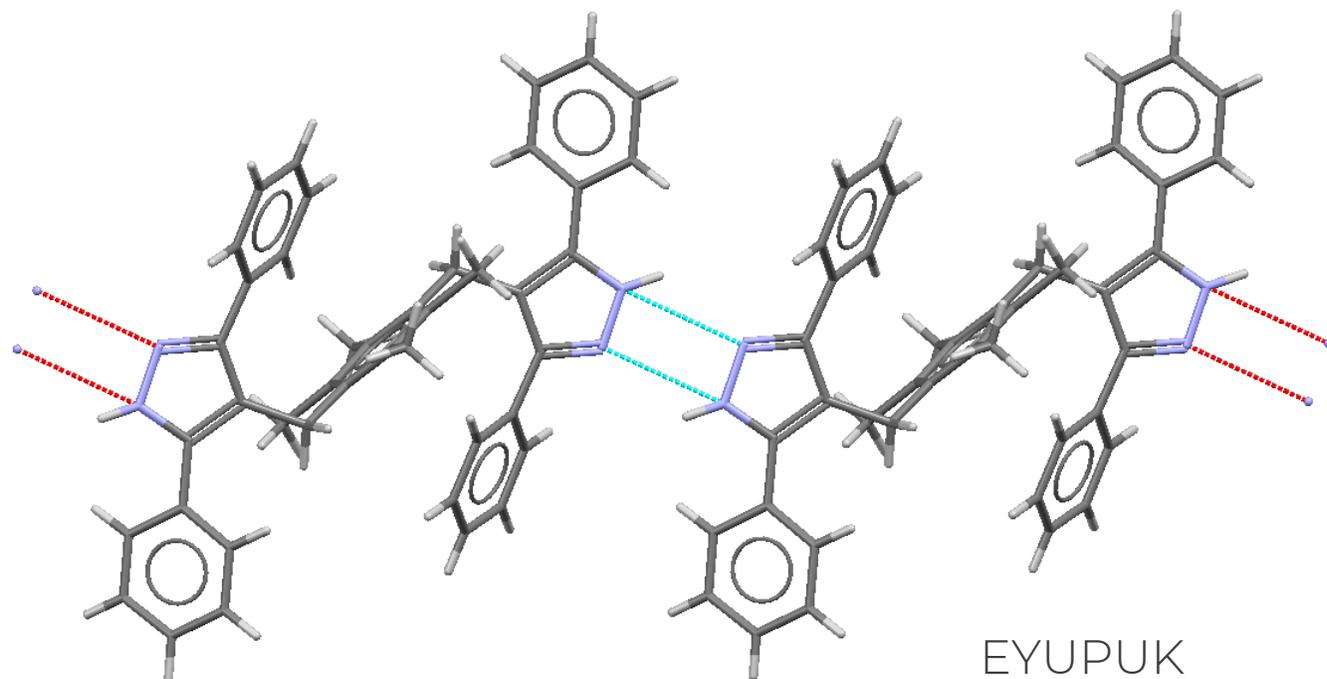


# Pyrazoles

- DISDIF difference between C-NH and C=N bond distances
- ANGDIF difference between C-NH-N and C=N-NH angles
- While most Pyzl structures have large positive ANGDIF, the long “tail” towards zero suggests that N/NH disorder is common
- Negative ANGDIF values suggest errors

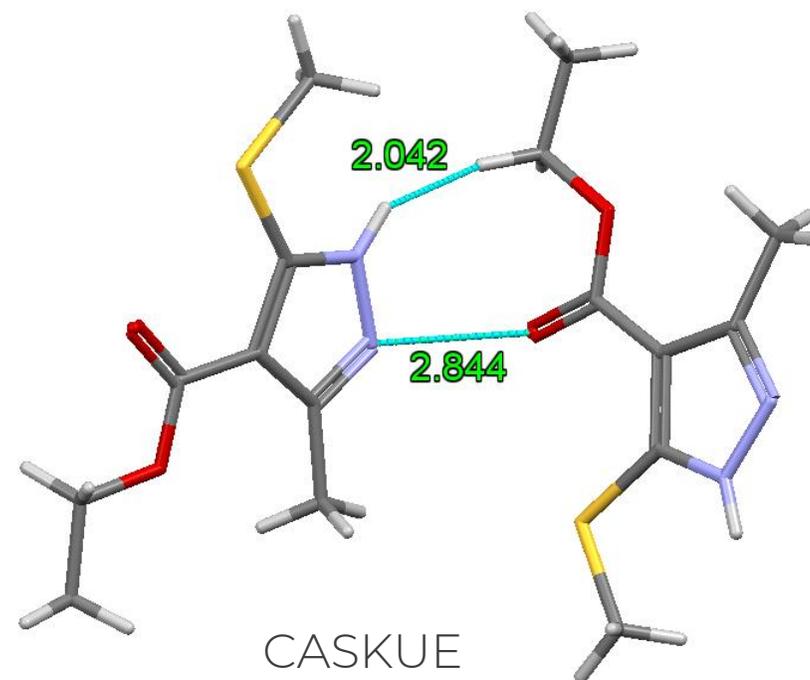


# 6 representative examples



NH...N linkages between rings may be swapped

- EYUPUK -0.021, -6.22
- DICQUD -0.025, -6.30
- GINZIN -0.011, -5.94



H...H clashes and missing hydrogen bonds

- CASKUE -0.036, -7.70
- NABVUK -0.036, -6.67
- VOJZEB -0.045, -5.47

# Conclusions (Im and Pyzl)

- Some reported crystal structures of neutral Im and Pyzl derivatives appear to have NH mistaken for N, or disordered swapping.
- CheckCIF often doesn't pick these kind of issues up
- Ring geometry (in combination with sensible H-bonding network and chemistry) provides a useful means to distinguish N from NH
- These are essentially 'human' checks right now
- Difficult to see these issues when looking at individual structures – need to see trends in related structures

# What tools are available for new structures?



A service of the  
**International Union of Crystallography**

checkCIF reports on the consistency and integrity of crystal structure determinations reported in CIF format.

Please upload your CIF using the form below.

File name:  
 No file chosen

Select form of checkCIF report

- HTML
- PDF
- PDF (recommended for CIFs that might take a long time to check)

Select validation type

- Full validation of CIF and structure factors
- Full IUCr publication validation of CIF and structure factors
- Validation of CIF only (no structure factors)

Output Validation Response Form

- Level A alerts only
- Level A and B alerts
- Level A, B and C alerts
- None

[Information about this version of checkCIF ...](#)

**Useful links**

- Prepublication check for submissions to IUCr journals**
- Details of checkCIF/PLATON tests
- CIF dictionary
- Download CIF editor (pubCIF) from the IUCr
- Download CIF editor (enCIFer) from the CCDC

checkCIF is sponsored by

**THE PLATON HOMEPAGE**

Right: Structure of sucrose prepared with PLATON/PLUTON-POFRAY

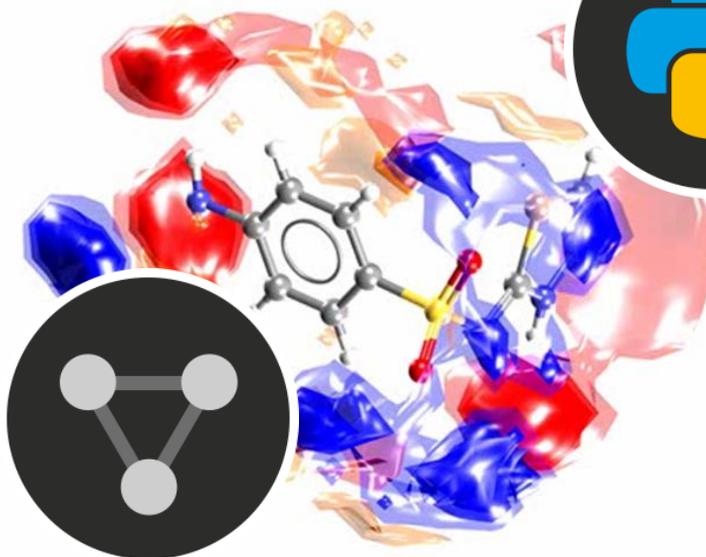
The Program PLATON is designed as a Multipurpose Crystallographic Tool.  
 (C) 1980-2019 A.L.Spek, Utrecht University, Padualaan 8, 3584 CH Utrecht, The Netherlands.  
 Reference: A.L.Spek, Acta Cryst. 2009, D65, 148-155.

The PLATON Homepage gives pointers to all information available on the program PLATON.

Download Sites:  
<http://www.platonsoft.nl/spek/xraysoft/>  
<http://www.cryst.chem.uu.nl/spek/xraysoft/>

The CheckCIF functionality within PLATON forms part of the IUCr Small Molecule Crystal Structure Validation Project.

For a PLATON tutorial by Lachlan Cranswick (CCP14) look [here](#)

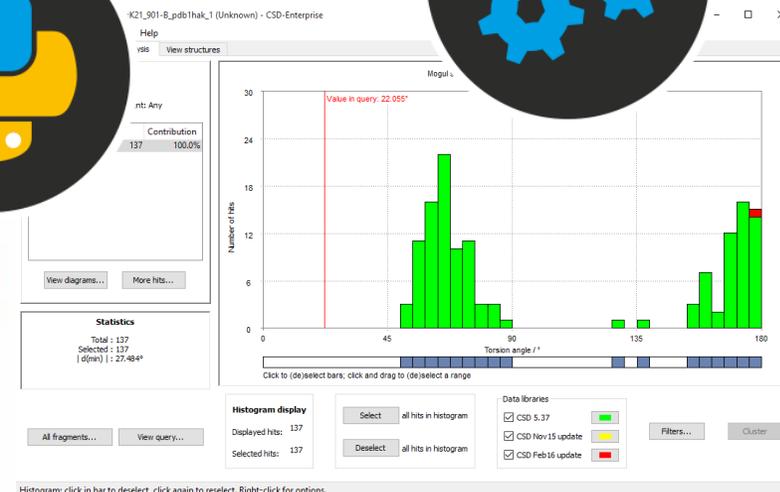


Mogul Results Viewer

Show / hide: Columns Fragments... Deselect all fragments Export...

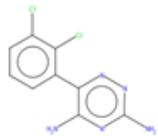
Help Double click to view result in Mogul

Type	Molecule	Fragment	Classification	No. of hits	Query value	Me
>	bond					
>	angle					
∨	torsion					
		LIM_K21_901-B.pdb1hak_1				
		C59 O58 C1 C2	Not unusual (enough hits)	13461	169.071	
		C59 O58 C1 C6	Not unusual (enough hits)	13461	-10.086	
		O23 C21 N11 C12	Not unusual (enough hits)	45	11.263	
		C24 C22 C21 N11	Not unusual (enough hits)	40	-75.917	
		C22 C24 N27 C30	Not unusual (enough hits)	938	-155.722	
		C22 C24 N27 C34	Not unusual (enough hits)	938	76.119	
		C45 C52 C43 C32	Not unusual (enough hits)	938	-71.842	
		C51 C52 C43 C32	Not unusual (enough hits)	938	105.277	
		C32 C21 N11 C12	Not unusual (enough hits)	45	-165.684	
		O23 C21 N11 C10	Not unusual (enough hits)	45	-58.238	
		O23 C21 C22 C24	Not unusual (enough hits)	45	965	
		C21 C22 C24 N27	Not unusual (enough hits)	45	965	
		C31 C32 C43 C52	Not unusual (enough hits)	45	965	
		C33 C32 C43 C52	Not unusual (enough hits)	45	965	
		C22 C21 N11 C1	Not unusual (enough hits)	45	965	



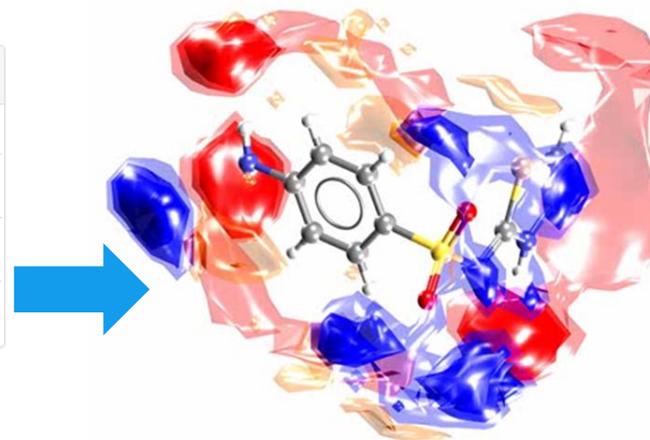
# Additional validation for new structures?

CCDC 1234567



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

Structure	Summary	checkCIF	Duplicates	Geometry check	Interaction check
1234567	<input type="checkbox"/>	<a href="#">View Report</a>	0 duplicates	<a href="#">View Report</a>	<a href="#">View Report</a>
1234568	<input type="checkbox"/>	<a href="#">View Report</a>	View 1	<a href="#">View Report</a>	<a href="#">View Report</a>
1234569	<input type="checkbox"/>	<a href="#">View Report</a>	View 2	<a href="#">View Report</a>	<a href="#">View Report</a>



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

Datablock: tBu10kbar

Bond precision: C-C = 0.0077 Å      wavelength=0.48690  
 Cell: a=14.811(3)      b=6.4564(7)      c=19.759(4)  
       alpha=90      beta=94.068(8)      gamma=90  
 Temperature: 296 K

Calculated      Reported  
 1884.7(6)      1884.7(6)

Enter CheckCIF Response

PLAT027\_diffn\_refins\_theta\_full value (too) Low ..... 13.50 Degree

PLAT029\_diffn\_measured\_fraction\_theta\_full value Low .. 0.677 Note

Level B

PLAT415 Short Inter D-H...H-X H1 ... H623 .. 2.00 Ang.

Level C

Click on the hyperlinks for more details of the test.

- Alert level A**  
 PLAT027\_ALERT\_3\_A diffn\_refins\_theta\_full value (too) Low ..... 13.50 Degree  
 PLAT029\_ALERT\_3\_A diffn\_measured\_fraction\_theta\_full value Low .. 0.677 Note
- Alert level B**  
 PLAT415\_ALERT\_2\_B Short Inter D-H...H-X H1 .. H623 .. 2.00 Ang.
- Alert level C**  
 SINT001\_ALERT\_3\_C The value of Rint is greater than 0.12

CCDC WebCSD

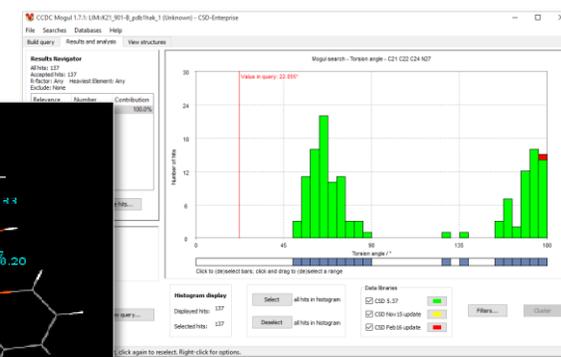
Simple Search    Structure Search    Unit Cell Search    Formula Search

Your query was Identifier(s) 1419071,629864,1411686 and the search returned 4 records.

OLUGIN	Deposition Number(s): 1419071 Space Group: P 2 <sub>1</sub> (1)
SALOM09	Deposition Number(s): 629864 Space Group: P 2 <sub>1</sub> (1)
YALCOG	Deposition Number(s): 1411686 Space Group: P 2 <sub>1</sub> (1)
ICSD 629864	ICSD Structure Deposition Number(s): 1753696 ICSD Structure Space Group: F 3 m (20)

JSmol

Fragment	Labels	Value	Mean	Z-score	Classification
<input checked="" type="checkbox"/>	angle C1 C14 N43	108.195	119.352	13.008	Unusual (enough hits)
<input checked="" type="checkbox"/>	angle C15 N43 C14	105.5	126.772	9.135	Unusual (enough hits)
<input checked="" type="checkbox"/>	angle C12 C14 N43	133.97	122.923	3.579	Unusual (enough hits)
<input checked="" type="checkbox"/>	angle C22 C15 N43	113.828	120.673	2.136	Unusual (enough hits)



# Additional information provided to referees?

Your query was: Identifier(s): 1584792, Authors: Suzanna Ward and the search returned 1 record.

[Back to Search List](#)

[Modify Search](#)

[New Search](#)

## Results

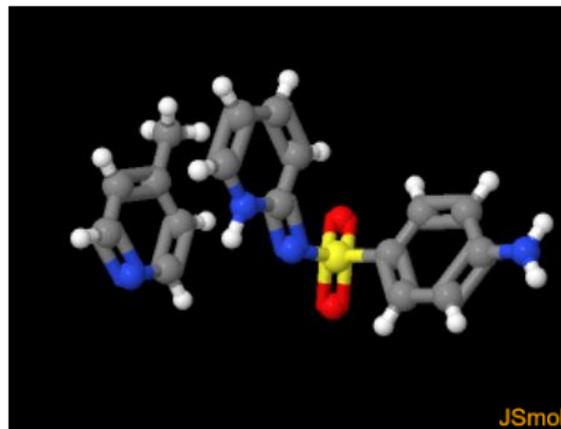
<input checked="" type="checkbox"/>	Database Identifier	Deposition Number
<input checked="" type="checkbox"/>		1584792

[Download ▾](#)

## Unpublished structure ?

Space Group:  $I 2/a (15)$ , Cell:  $a 14.811(3)\text{\AA}$   $b 6.4564(7)\text{\AA}$   $c 19.759(4)\text{\AA}$ ,  $\alpha 90^\circ$   $\beta 94.060(8)^\circ$   $\gamma 90^\circ$

## 3D viewer



JSmol

H Disorder ↻ Menu Open ▾ ↗

Style: Ball and Stick | Labels: No Labels | Packing: None | Measure: None

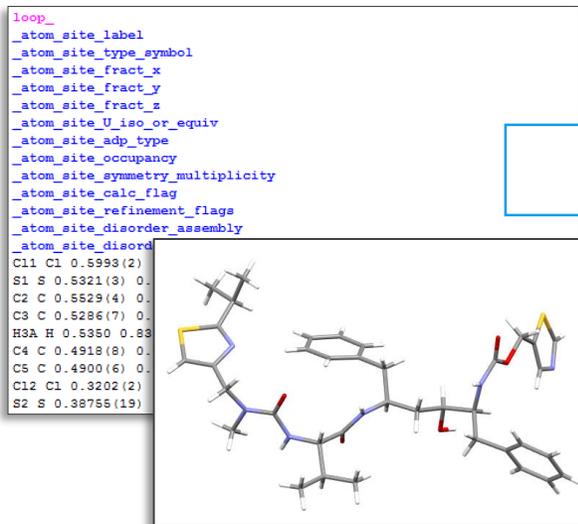
## Chemical diagram



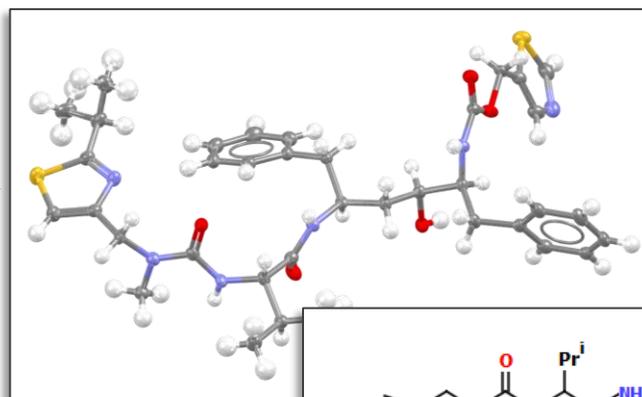
[View group symbols key](#)

# Structures in the CSD

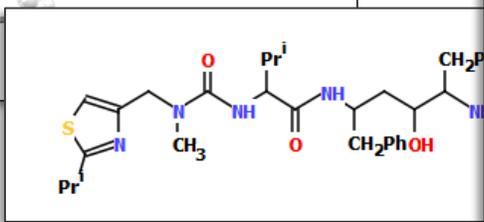
Deposited CIF



CSD Entry



IUPAC

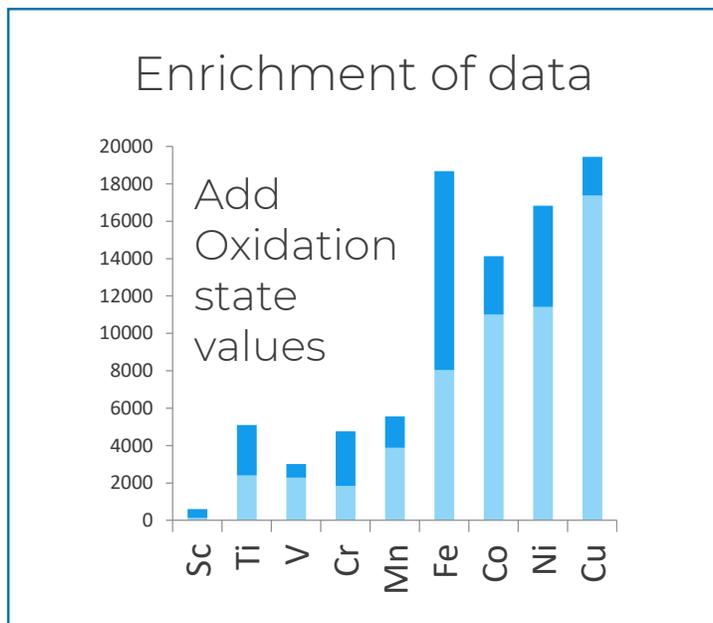
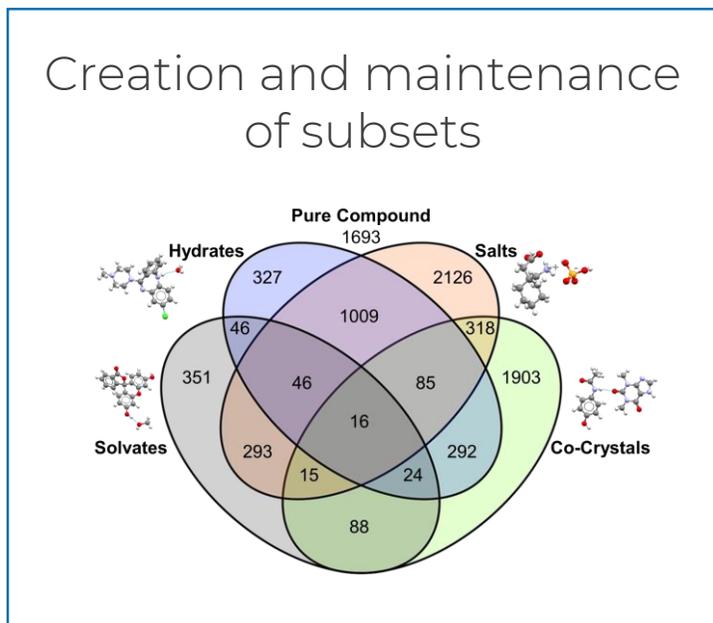
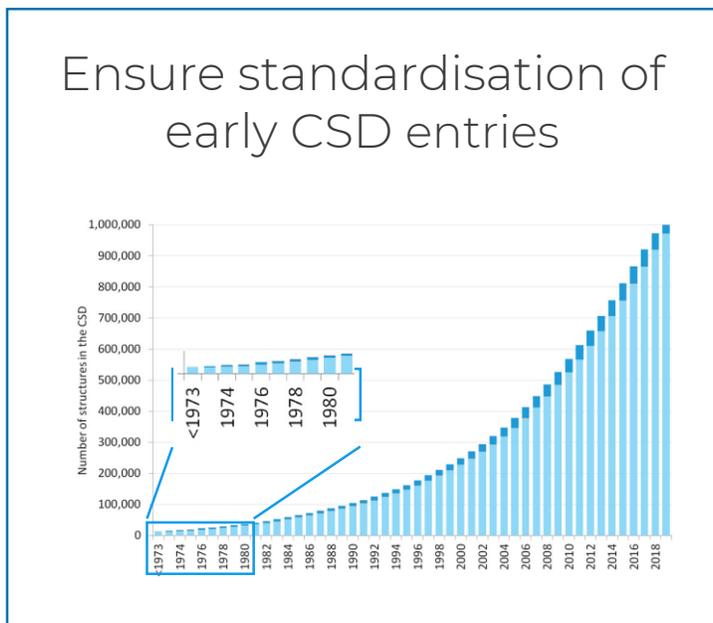


- Assignment of a chemically meaningful representation is determined using data in the CSD and manual curation.
- Important for data discovery, re-use, mining, analysis and interoperability

Customise...	
Identifier	YIGPIO03
Literature Reference	J.Bauer, S.Spanton, R.Henry, J.Quick, W.Dziki, W.Porter, J.Morris, <i>Pharm. Res.</i> (2001), <b>18</b> , 859, doi: <a href="https://doi.org/10.1023/A:1011052932607">10.1023/A:1011052932607</a>
Formula	C <sub>37</sub> H <sub>48</sub> N <sub>6</sub> O <sub>5</sub> S <sub>2</sub>
Compound Name	(5S-(5R*,8R*,10R*,11R*))-10-Hydroxy-2-methyl-5-isopropyl-1-(2-isopropyl-4-thiazolyl)-3,6-dioxo-8,11-dibenzyl-2,4,7,12-tetra-azatridecan-13-oic acid 5-thiazolyl methyl ester
Synonym	Ritonavir; Norvir; PDB Chemical Component code: RIT; DrugBank: DB00503
Space Group	P 2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub> (19)
Cell Lengths	<b>a</b> 9.831(6) <b>b</b> 18.485(11) <b>c</b> 20.261(12)
Cell Angles	<b>α</b> 90 <b>β</b> 90 <b>γ</b> 90
Cell Volume	3681.95
Temperature (K)	100
Z, Z'	<b>Z</b> : 4 <b>Z'</b> : 1
R-Factor (%)	6.47
Disorder	
Polymorph	stable orthorhombic polymorph 2

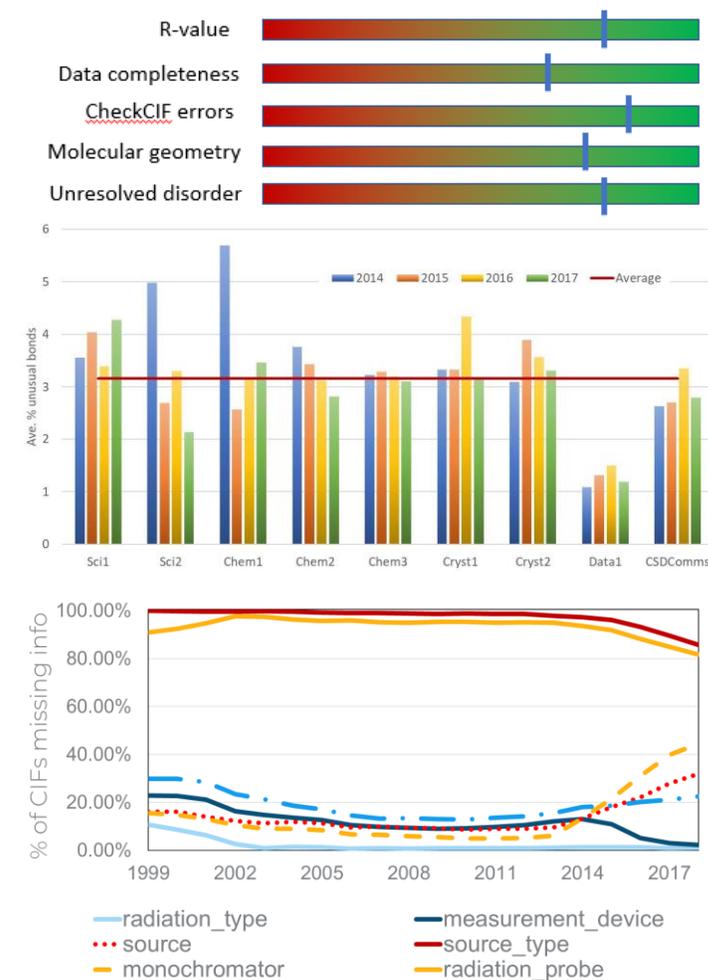
# Revisiting CSD entries

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



# 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



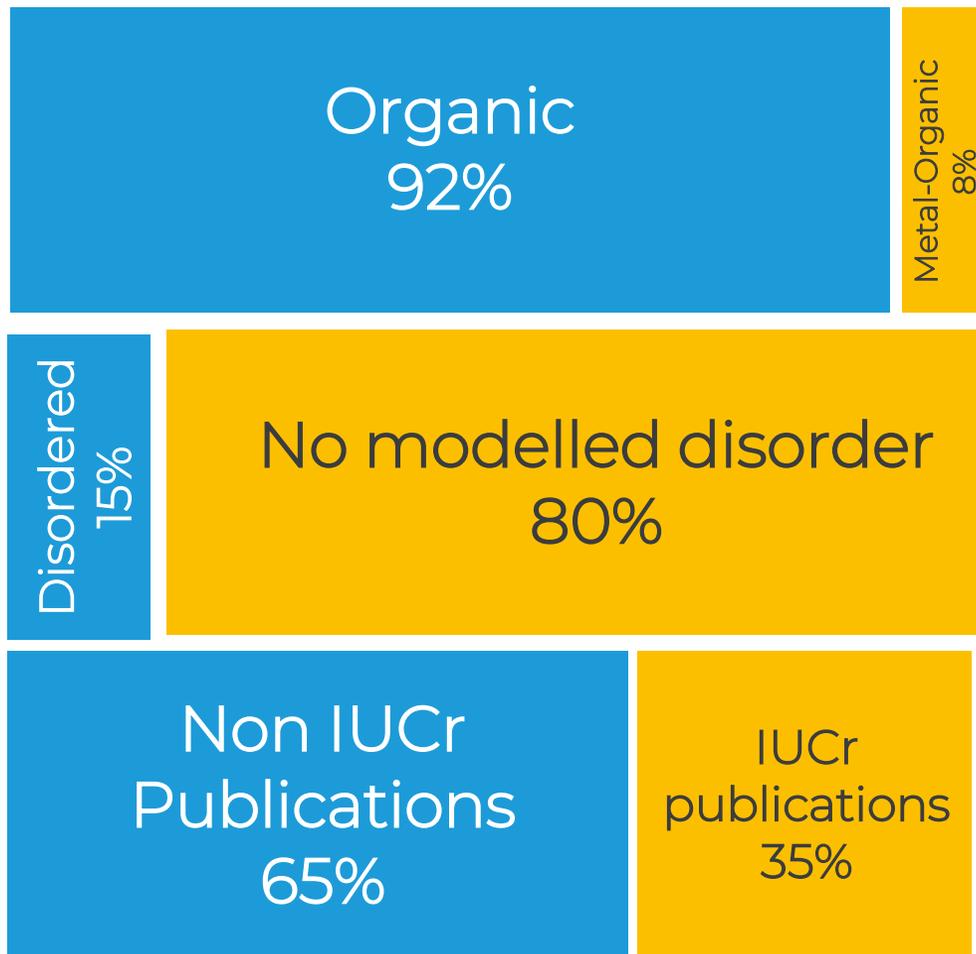
# Underlying issues in the CSD

- Underlying CIFs match published datasets
- Issues can be reported to CCDC – [data\\_edits@ccdc.cam.ac.uk](mailto:data_edits@ccdc.cam.ac.uk)
- CCDC will:
  - Investigate issue and either correct CCDC representation or:
    - Contact authors and/or publisher
    - Add a comment to CSD entry
    - If appropriate suggest correction to be published and deposited
    - Accept re-refinements of existing structures and link datasets
    - Re-refinements can be *CSD Communications* or published structures

# Identifying issues in existing structures

- Faults in structures have been corrected by “vigilantes” in their particular area of interest
- But such coverage is inevitably limited
- Can and should the crystallographic community organize a systematic validation and correction effort?
- Can and should the CCDC do more to identify issues?
- When a corrected version of a structure is found how should the CCDC/CSD handle these new models?

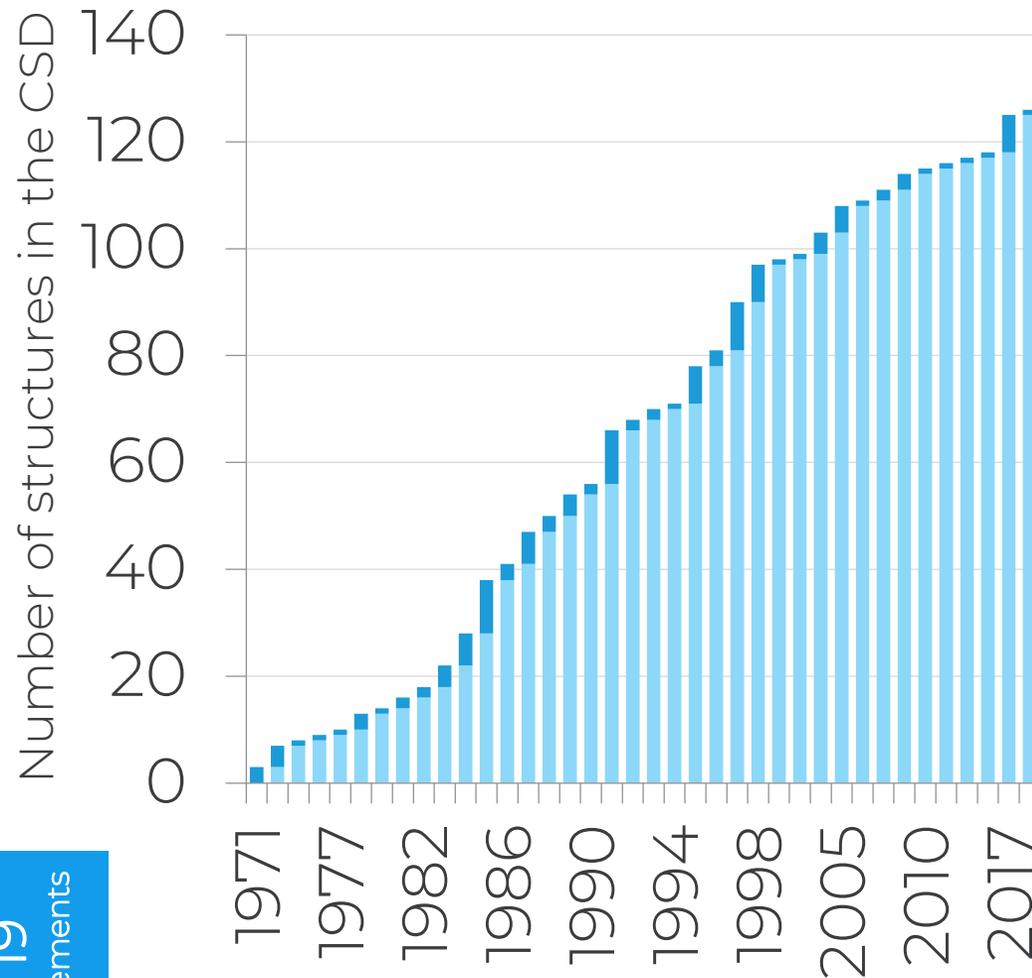
# Remembering Carl

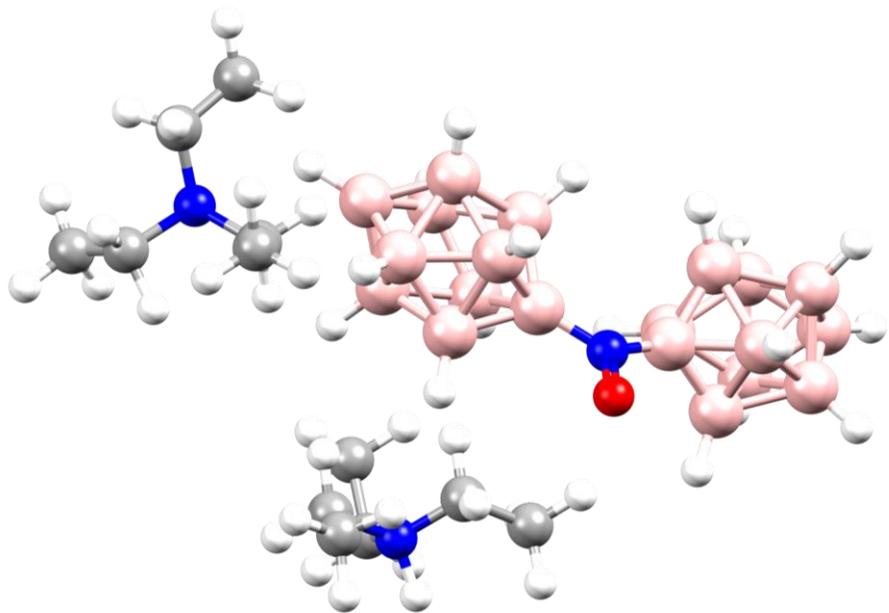


126 Structures

184 co-authors

19 elements

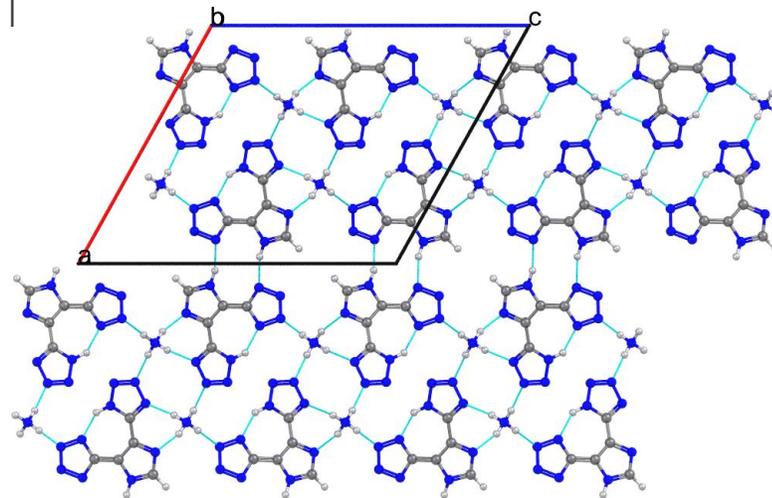




EAMBNO10

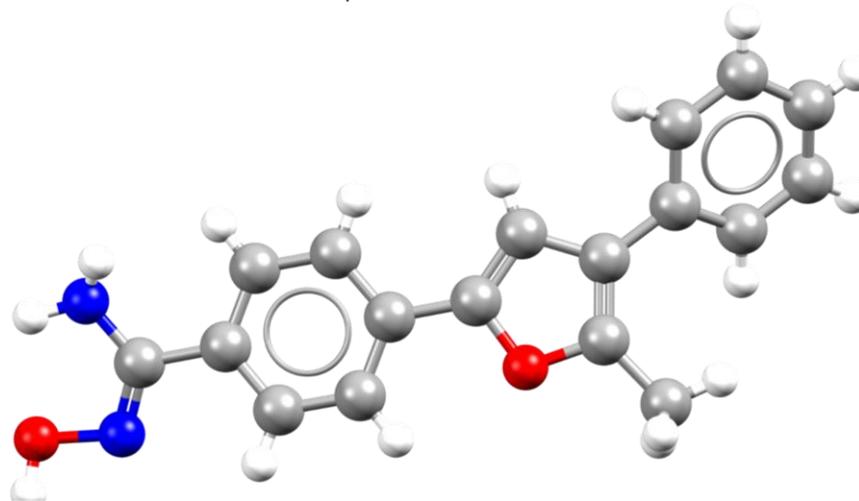
First published - January 1<sup>st</sup> 1971

YOTYOW → YOTYOW01  
Latest corrected structure

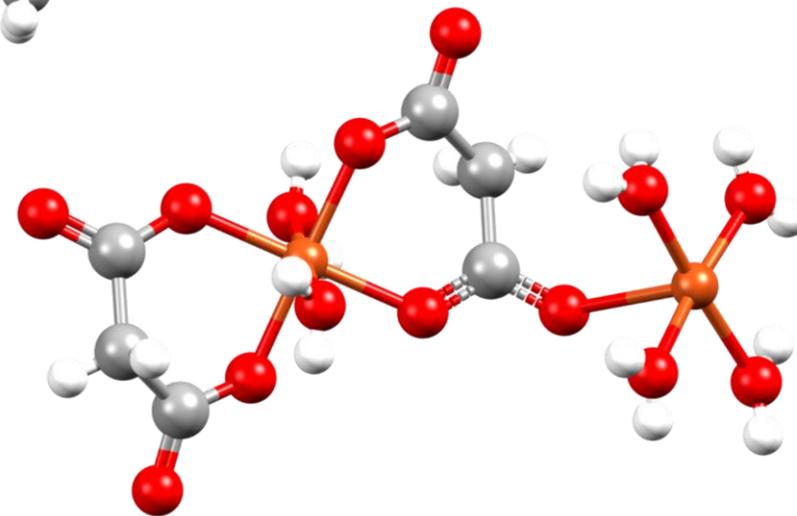


XEZFIU

Most recent – published 12/03/2018



PECMIT  
Most cited article



# Summary and workshop questions

- Not all structures are perfect, and a variety of approaches need to be taken to identify and resolve issues....
- How can we identify errors more automatically/systematically?
- When systematic errors are found experts may need to look at different approaches to fixing them
  - Re-determinations
  - Re-refinements
  - Generation of CSP/DFT/cleaned structures
  - Extensive annotation
- How are we going to do it at a whole-community level?
- What would be the incentives for individuals to engage?
- How should different versions of structures be stored?

# Thank You Carl



and...

Simon Coles

Natalie Johnson

Stephen Holgate

Clare Tovee

Seth Wiggin



Aston University  
Birmingham

CCDC

Should we remediate small molecule structures?

If so, who should do it?

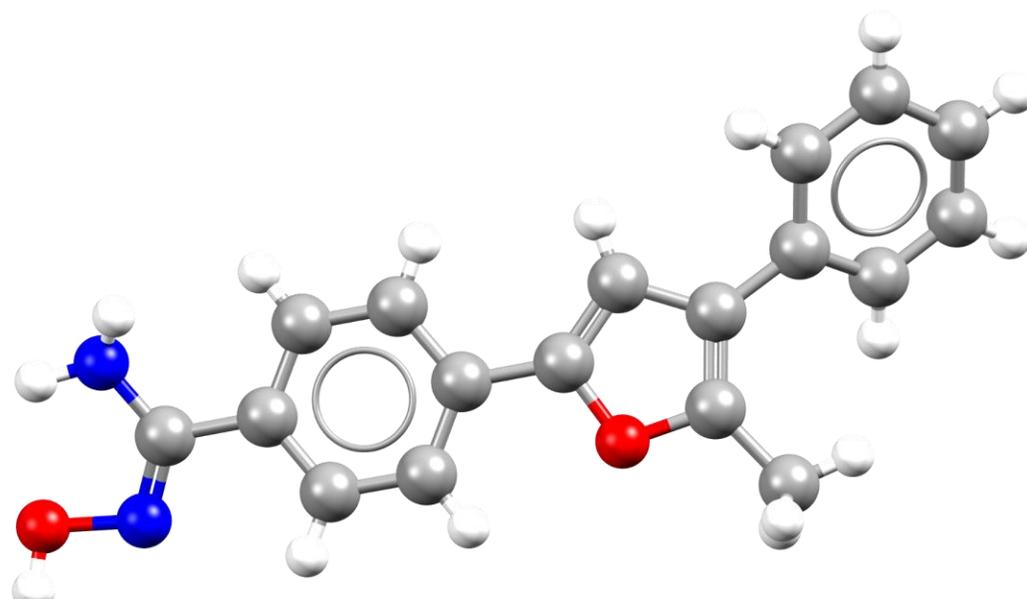
In principle, new reports of small molecule crystal structures should be error-free since most reputable journals require validation of crystallographic data with CheckCIF and this software is integrated into the CCDC deposition procedure. However, because some chemical journals appear to ignore or not even to use crystallographic referees, errors may not be pointed out. Furthermore, what should happen if authors are unable or unwilling to make corrections when required? Should an otherwise correct structure be rejected because a hydrogen atom has been incorrectly placed or disorder of a terminal methyl group has not been entered into the model? Should such a structure be published or deposited with a warning message, or should a corrected version be created by an external referee? These questions have particular force with regard to already published structures that have errors. An example from the author's early work shows that well-intentioned remediation can sometimes go wrong. Faults in structures have been corrected by "vigilantes" in their particular area of interest, such as space group symmetry [1] and misplaced hydrogen atoms [2,3]; but such coverage is inevitably limited. Can and should the crystallographic community organize a systematic validation and correction effort?

[1] R. E. Marsh (2009), *Acta Cryst.* B65, 782-783.

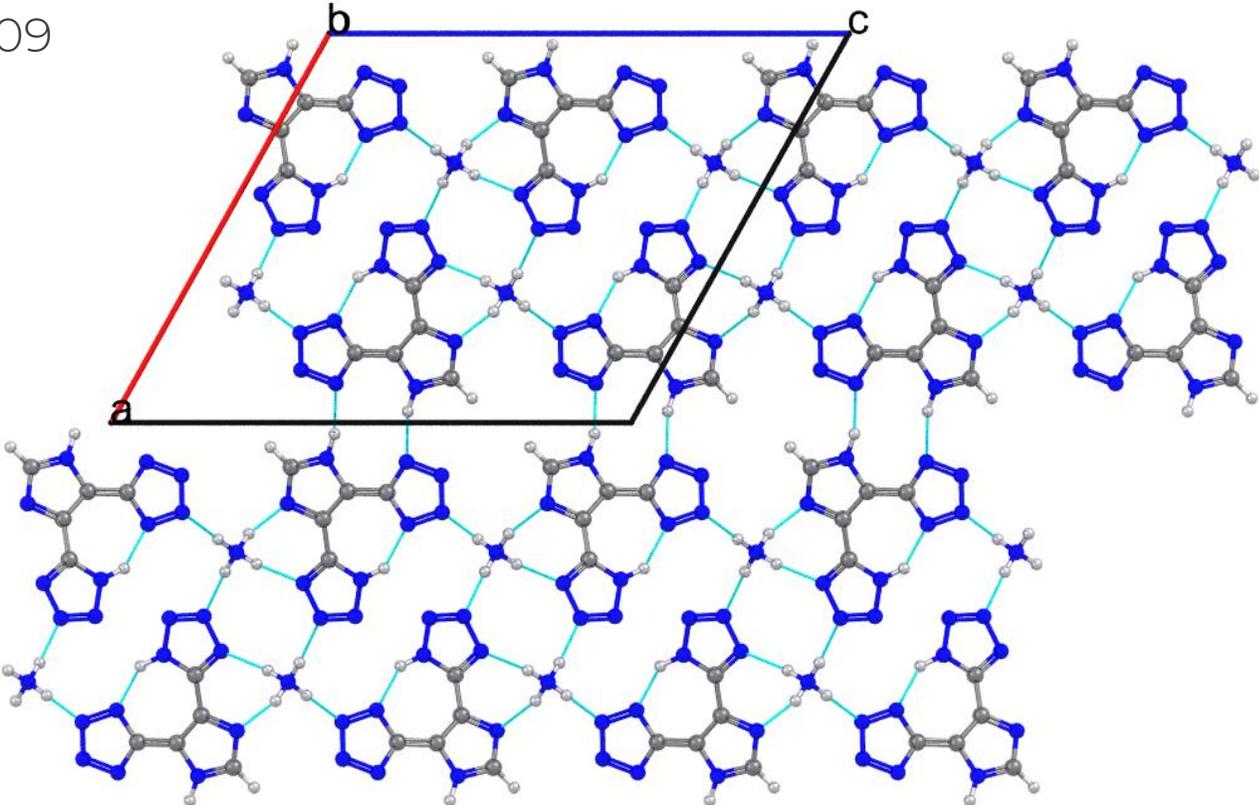
[2] I. Bernal & S. F. Watkins (2013), *Acta Cryst.* C69, 808-810.

[3] C. H. Schwalbe (2016) Abstract 01.11.01.12, 66th ACA Annual Meeting, Denver. *Acta Cryst.* (2017). A73, a133 Should we remediate small molecule structures? If so, who should do it? Carl Schwalbe United Kingdom Aston University

Last published  
XEZFIU – published 12/03/2018  
10.1016/j.bmcl.2018.03.025



Latest corrected structure  
YOTYOW → YOTYOW01  
10.1080/0889311X.2018.1508209



Most cited: PECMIT  
Cited by: 72  
DOI: 10.1039/DT9930000913

