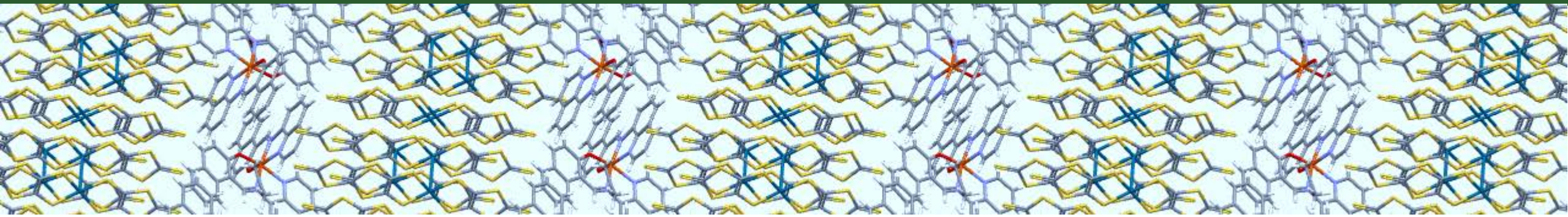


Organometallic InChI

Project Update, August 2019

Project Timeline

- **IUPAC Project 2009-040-2-800 commenced November 2010**
  - InChI Requirements for Representation of Organometallic and Coordination Compound Structures
- **Recent community discussion of Organometallic InChI:**
  - March 2017: Hinxton, UK (EBI) – InChI Stakeholders Meeting
  - August 2017: Bethesda, MD (NIH) – Status and Future of InChI
  - **August 2018: Boston, MA – InChI Workshop**
  - February 2019: Cambridge, UK – InChI Workshop
  - August 2019: San Diego, CA – InChI Symposium

InChI Workshop, Boston, August 2018

- **Questions considered:**
  - What positive transformation could organometallic InChI enable?
  - What are the motivational use cases?
- **Organometallic InChI drivers:**
  - Organometallic InChI fundamental to integrity and perception of InChI:
    reputational risk if left unaddressed – possible barrier to adoption
  - Community don't appear to be beating down the door:
    but perhaps patiently waiting for the working group to deliver
  - Potentially significant opportunities for application:
    but need something to be built to explore if this is the case

Project Timeline

- **IUPAC Project 2009-040-2-800 commenced November 2010**
  - InChI Requirements for Representation of Organometallic and Coordination Compound Structures
- **Recent community discussion of Organometallic InChI:**
  - March 2017: Hinxton, UK (EBI) – InChI Stakeholders Meeting
  - August 2017: Bethesda, MD (NIH) – Status and Future of InChI
  - August 2018: Boston, MA – InChI Workshop
  - **February 2019: Cambridge, UK – InChI Workshop**
  - August 2019: San Diego, CA – InChI Symposium

Cambridge Discussion Highlights

- **Need to get somewhere – but need to get far enough**
  - Chemists won't take a solution seriously if it doesn't distinguish cis and trans platin for example
- **Possible compromises**
  - Some stereochemistry but maybe not all stereochemistry
  - Some "one to many" correspondences
- **Should ultimately ensure that InChI does everything that SMILES can do**
  - SMILES can handle square planar, trigonal-bipyramidal, octahedral for example

Stereochemistry Proposal

- **Proposal from Jonathan Goodman:**

Phase 1

In phase one, we assume that all the information we need about a molecule is available. We concentrate on constructing the InChI. The connectivity has already been defined by the /om layer. The /ma layer (*metal architecture*) now defines the stereochemistry.

Simple Stereochemistry

Assume that the metal-centred stereocentre is a clearly defined shape and it is straightforward to assign priorities to the substituents. The stereochemistry can be defined by a symbol for the shape, here 'o' for octahedral. This situation has been studied before. There are thirty different arrangements of the six substituents, and the desired one can be given a number from one to thirty. If, in this example, the central metal atom has canonical number 99

`/ma99o1 or /ma99o2 or … /ma99o30`

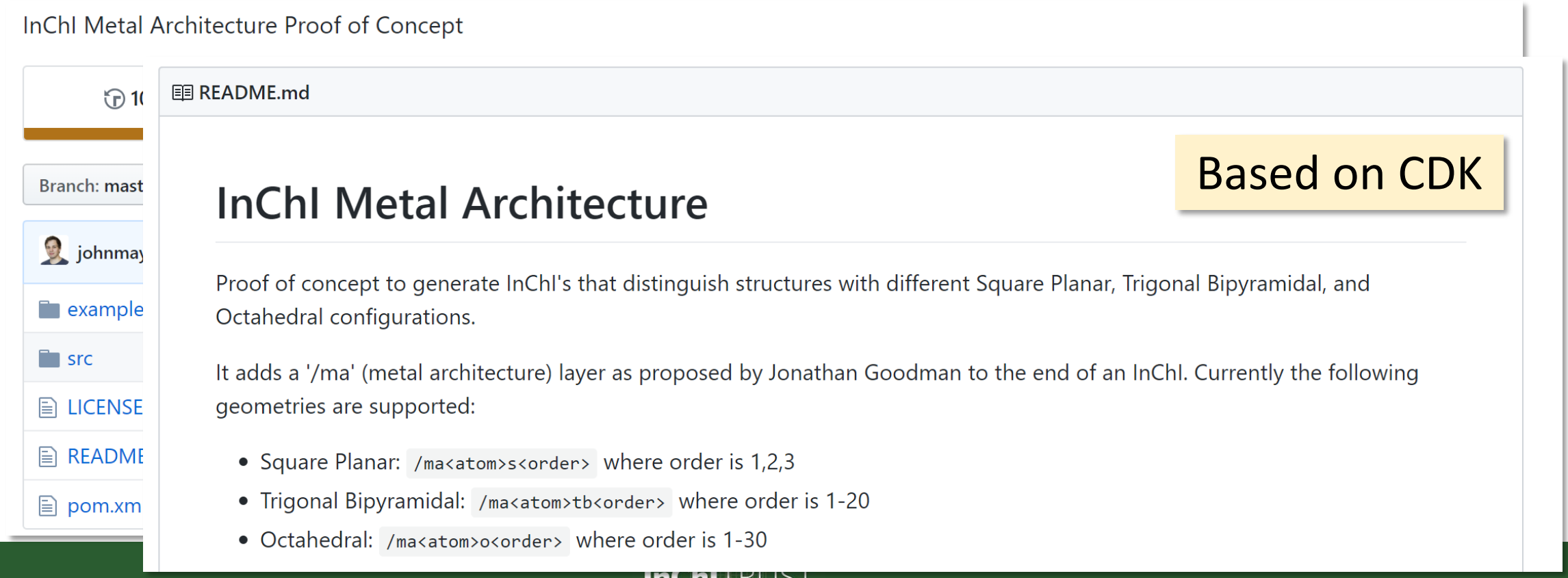
Tetrahedral, trigonal bipyramidal, cubic, *etc*, structures can be defined in a similar way. Square planar compounds can either be treated as a separate category, or as a special case of

- Proposes a “Metal architecture” layer
- Define a symbol for a shape – e.g. ‘o’ for octahedral
- There are 30 different arrangements of 6 substituents – give each a number
- If the canonical number of a central octahedral atom is 99 then the metal architecture layer could be one of

  `/ma99o1 /ma99o2 … /ma99o30`

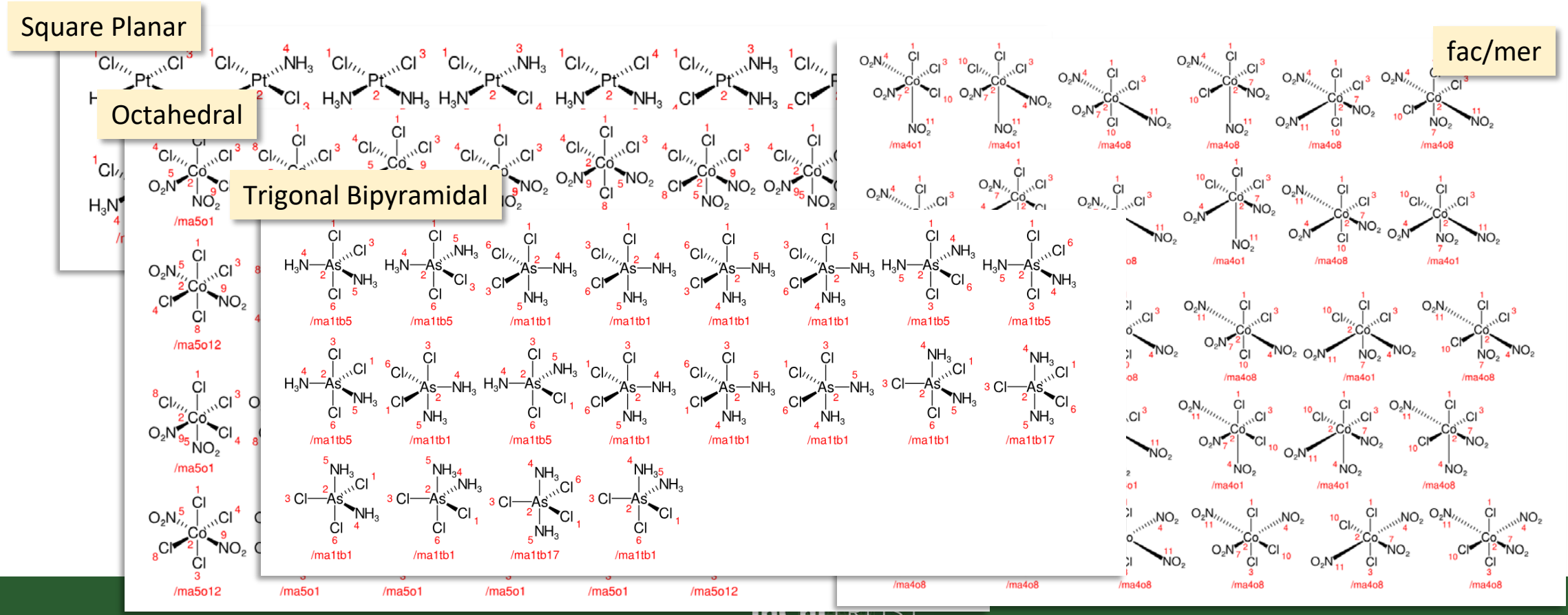
Stereochemistry Proof of Concept

- **Proof of Concept from John Mayfield:** https://github.com/johnmay/inchi-ma



Stereochemistry Proof of Concept

- **Proof of Concept from John Mayfield:** https://github.com/johnmay/inchi-ma



Note the numbers here refer to the input atom order in the SMILES, not the canonical InChI numbers

Prior to Cambridge (shortly after Boston)

- **Jonathan Goodman drafted a proposal for an Organometallic connectivity layer**
  - Standard InChI describes organometallics by complete disconnection of metal-carbon bonds. This is an effective summary of such structures, but loses information that is often considered important.
  - This can be addressed by using the reconnection layer that is defined in the current release of the InChI but omitted from the standard InChI.
  - The reconnection layer repeats much of the original InChI providing a detailed description of alternative connections.
  - **It may be possible to add a new layer that describes reconnections more concisely, focusing on the information that chemists use most frequently.**

Cambridge Discussion Highlights

- **Would ability to handle organometallics be part of standard InChI?**
  - Ultimately (short to medium term) – yes
- **How profound a conceptual change is required to current InChI**
  - It can already generate a reconnection layer - does this and/or AUX Info provide what is needed to create a new “coordination” layer
- **Can we prototype a solution without changing InChI code?**
  - Generating a potential organometallic InChI layer using existing reconnection layer / aux info
- **How hard would this be?**
  - Expert/non-expert programmer with/without domain expertise?

The InChI Organometallic RFP

- **A “Request for Proposal” (RFP) was issued early 2019 based on Jonathan’s proposal and discussion at Cambridge**
- **The RFP process sparked further community discussion of what may or may not be the right way forward**
- **Discussion was factored into a revision of the RFP**
  - Is the current reconnection layer appropriate to use?
  - Should bonds to metal be disconnected at all?
  - If we do disconnect, should normalisation happen before or after disconnection?

The InChI Organometallic RFP

- **Be able to compare three prototype InChIs:**
  - An organometallic InChI with a coordination layer following the Goodman Proposal
  - As above but without initially disconnecting bonds to metal atoms
  - The InChI generated with the InChI metal reconnection layer
- **To aid with this**
  - The InChI Trust has supplied a small set of test structures compiled earlier
  - The CCDC is providing access to chemical structures in the CSD through their Python API

Enter Alex Clark…

Scientist at Collaborative Drug Discovery

Founder of Molecular Materials Informatics

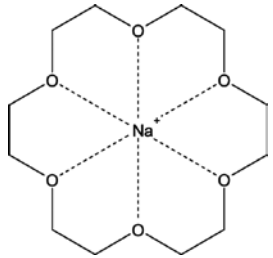
Initial Deliverables

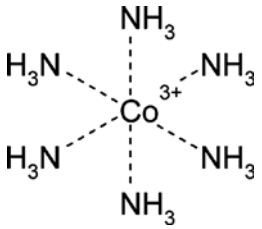
- **Training set of compounds:**
  - real-world compounds (CSD, PubChem, misc)
  - some drawn *well*, others drawn *badly*
- **Prognosis for issues to expect:**
  a. current InChI works fine, ***or***
  b. new layer is required, ***or***
  c. intractable problems persist
- **Use as a definitive pass/fail validation key**

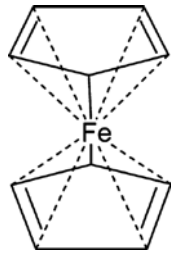
Reliable Representation

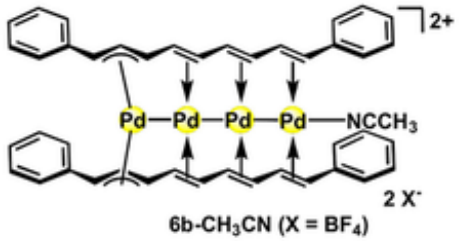
- How best to reliably represent organometallics?
  - dative vs covalent bonds?
  - explicit hydrogens/valencies?
  - dummy atoms?
  - zero-order bonds?

Accurate Specification of Molecular Structures: The Case for Zero-Order Bonds and Explicit Hydrogen Counting. **Alex M. Clark.**
J. Chem. Inf. Model., 2011, 51 (12), 3149. doi:10.1021/ci200488k

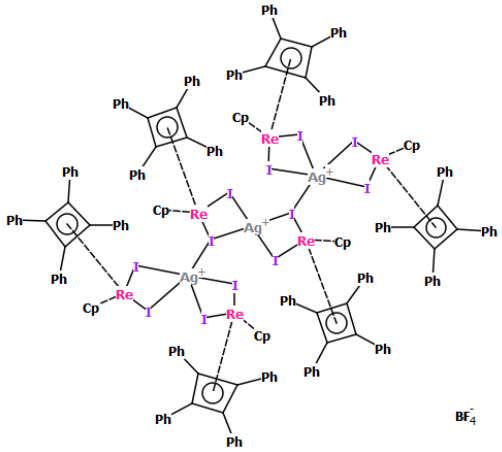


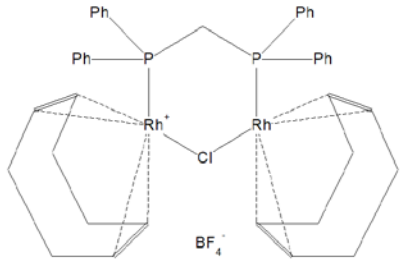


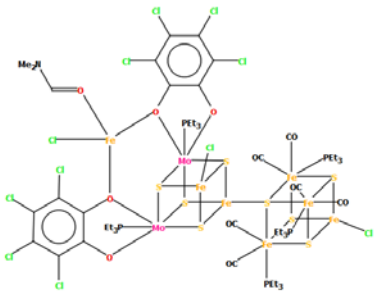




6b-CH_3CN (X = BF_4)

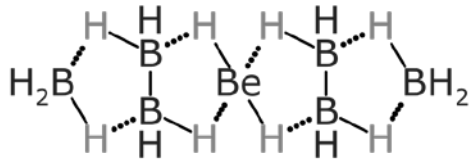




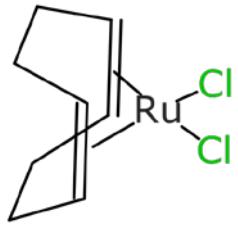


Categorisation of bonding types that need attention

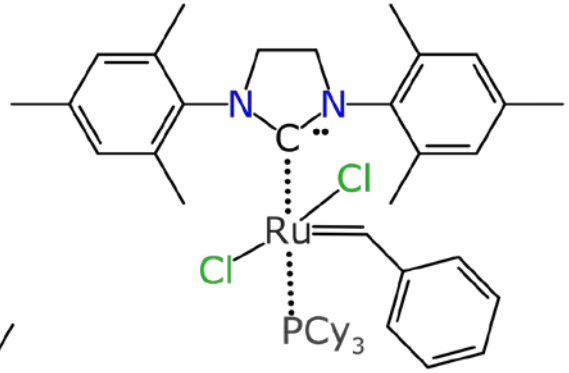
Alex Clark



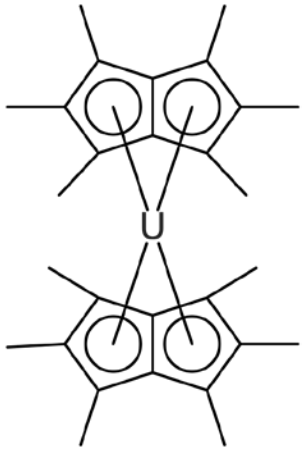
alternating



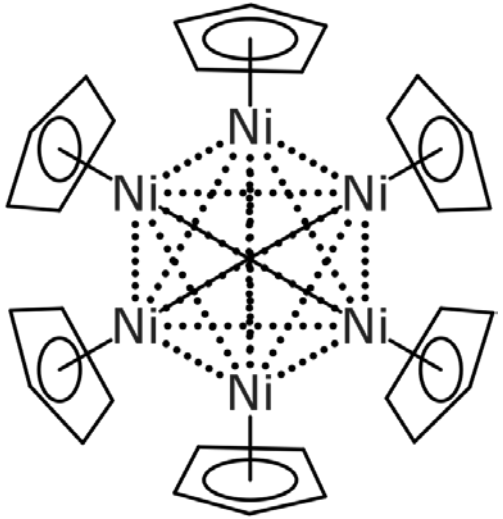
alkene



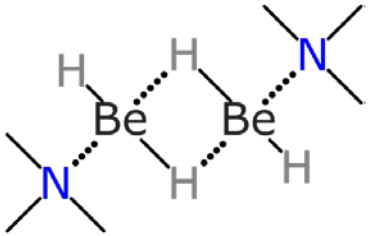
carbene



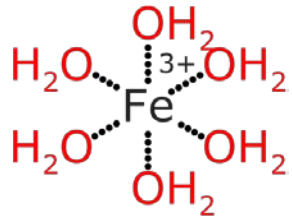
arene



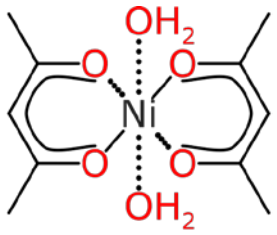
hypervalent



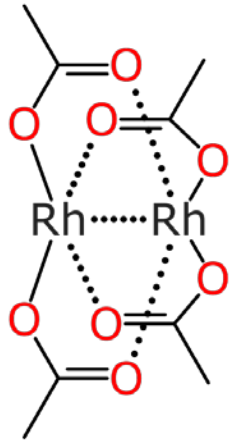
multicentre



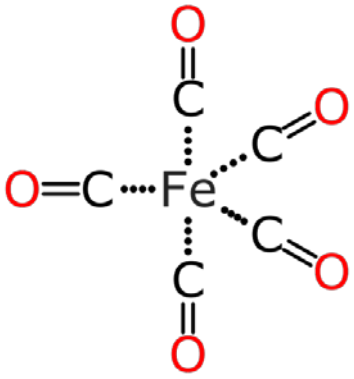
terminal O



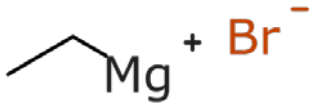
bidentate



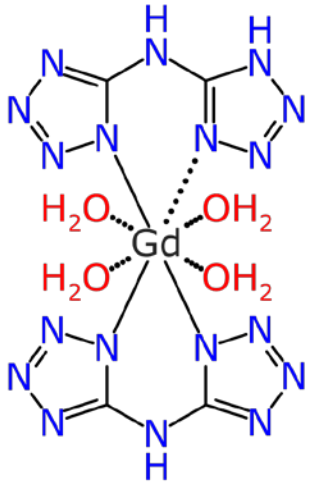
metal-metal



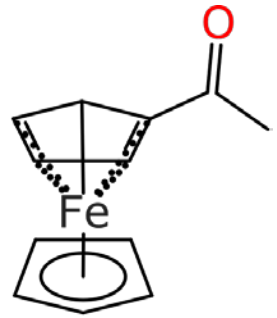
carbonyl



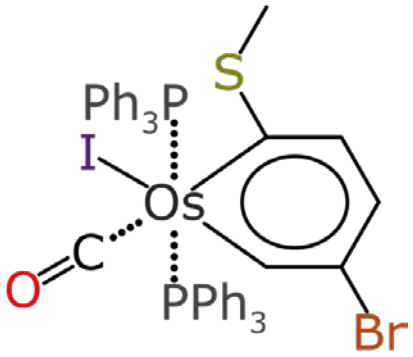
disconnected



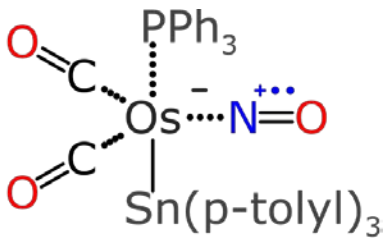
dative



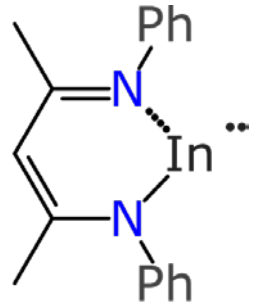
symmetry



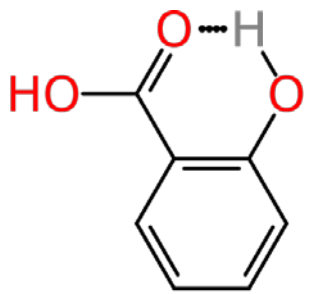
metallabenzene



nitrosyl



hypovalent



H-bond

Breakout Session

- **Presentation and discussion of Alex's work to date**
- **Discussion topics**
  - Balance between getting the structure "right" and a tractable outcome
  - How faithfully should we encode what is drawn?
  - How rigorous does canonicalization need to be?
  - Who is organometallic InChI targeted at?
  - How to implement – modifying the core vs externally adding a layer
- **Identification of next steps**