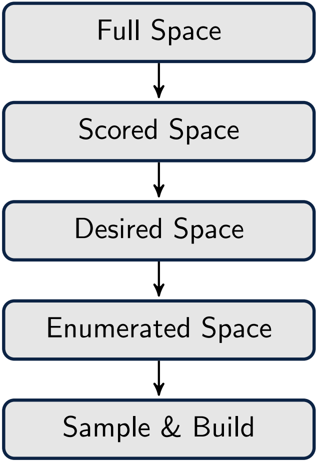
GDB-9 & Small Molecule Universes

Stefan O. Gugler

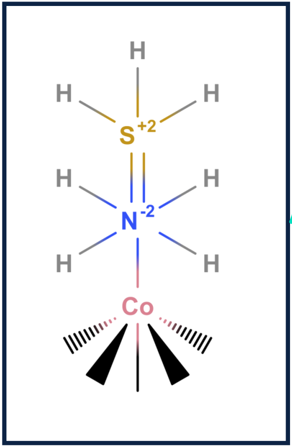
Massachusetts Institute of Technology
Department of Chemical Engineering

May 31, 2018

Reduction of the full space



Reduction of the full space



Full Space

↓

Scored Space

↓

Desired Space

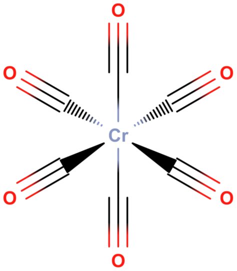
↓

Enumerated Space

↓

Sample & Build

Reduction of the full space



Full Space

↓

Scored Space

↓

Desired Space

↓

Enumerated Space

↓

Sample & Build

Full Space $\rightarrow$ Scored Space

- charge $\in [-2, +2]$
- H atoms $\in [0, 4]$
- element $\in \{\mathrm{C, N, O, P, S}\}$

Full Space:
M:125, D:5625, B:5625

$\downarrow$ ① charge, sterics, octet, shell, bond order

Scored Space:
50, 1500, 201

These rules ① are demonstrated in the following for the di-heavy-atoms.

Rules (1) to reduce Full Space → Scored Space

Constraints:

- Charge $c = c_1 + c_2 \leq 1$
- Sterics: H atoms < 4 on connecting atoms
- Closed shell (even number of electrons)

→ Reduces 5625 ligands to 1171.

Rules ① to reduce Full Space → Scored Space

$$u_{\text{octet,i}} = \begin{cases} 10 + 2 \cdot (8 - VE_i) & \text{if } 8 - VE_i < 0 \\ 10 - 1 \cdot (8 - VE_i) & \text{if } 8 - VE_i \geq 0 \end{cases} \tag{1}$$

$$u_{\text{charge}} = \begin{cases} 0 & \text{if } c_1 + c_2 > 0 \\ 3 & \text{if } 0 \geq c_1 + c_2 \geq -2 \\ 1 & \text{if } c_1 + c_2 = -3 \\ 0 & \text{if } c_1 + c_2 = -4 \end{cases} \tag{2}$$

$$u_{\text{VSEPR}} = 5 - \left| \underbrace{(VE_1 - 2 \cdot LP_1 + c_1 - 2 \cdot h_1)}_{ready \;\; electrons} - \underbrace{(VE_2 - 2 \cdot LP_2 + c_2 - 2 \cdot h_2)}_{ready \;\; electrons} \right| \tag{3}$$

Rules ① to reduce Full Space → Scored Space

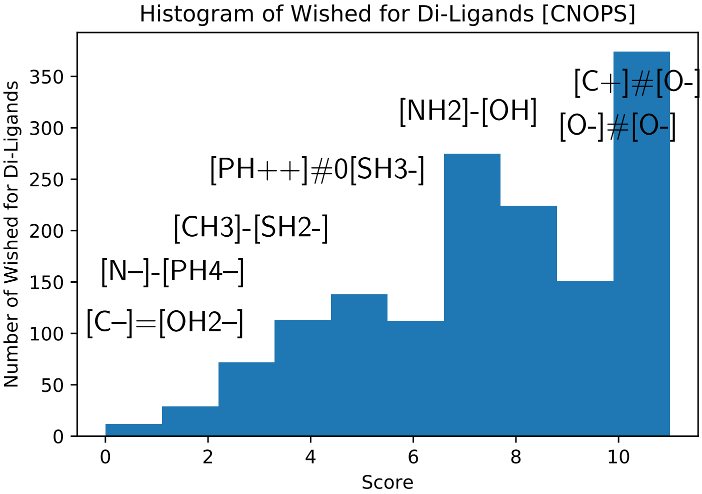
$$u_{\mathrm{CA}} = \begin{cases} 1 & \text{if } h_1 = 4 \\ 2 & \text{if } h_1 = 3 \\ 3 & \text{if else} \end{cases} \tag{4}$$

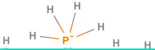
$$u_{\mathrm{shell}} = \begin{cases} 1 & \text{if } \neg(VE_1 + VE_2)\%2 \\ 0 & \text{if else} \end{cases} \tag{5}$$

$$u_{\mathrm{total}} = u_{\mathrm{shell}} \cdot \left(\frac{1}{2}(u_{\mathrm{octet},1} + u_{\mathrm{octet},2}) + u_{\mathrm{charge}} + u_{\mathrm{VSEPR}} + u_{\mathrm{CA}} \right) \tag{6}$$

where VE_i denotes the number of valence electrons, LP_i the number of lone pairs.

Distribution and Examples





Reduction of the full small ligand universe: Mo, Di, Bi

Scored Space:
50, 1500, 201

↓ } 2 score $> s$

Desired Space:
29, 374, 60

Reduction of the full small ligand universe: Mo, Di, Bi

Desired Space:
29, 374, 60

↓ 3 p % of each isoelectronic

Enumerated Space:
20, 80, 50

Reduction of the full small ligand universe: Mo, Di, Bi

Enumerated Space:
20, 80, 50

↓

Sample & Build

4 random whim?

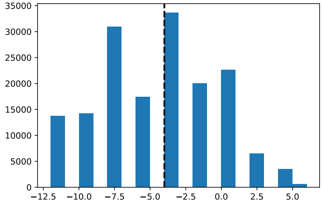
Subsets of octahedral space

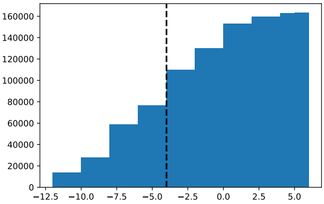
The sizes of the selected subsets of octahedral space.

Set	description	size
Homoleptics	eq $=$ ax	553
"5+1" symmetric	eq $=$ ax1 $\neq$ ax2	163,620
"4+2" symmetric	eq1 $\neq$ eq2 $=$ ax	185,376
Strongly symmetric	eq $\neq$ ax	245,316
Equatorially asymmetric	eq1 $\neq$ eq2 $\neq$ ax	15,924,796
Weakly symmetric	eq $\neq$ ax1 $\neq$ ax2	45,077,310
Complete Heteroleptics	$L_i \neq L_j$	$\approx 5.9 \cdot 10^{12}$
Octahedral Space	all	$> 1.8 \cdot 10^{14}$

Properties of the sets

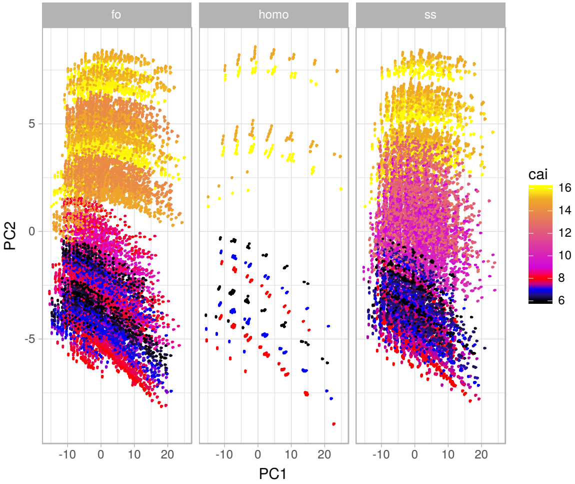
- Reduce space to facilitate sampling from non-homoleptics
- Example: strongly symmetric, monodentate ligand fields (163,620)
- Exclude all with charge smaller than -4, which results in 87,150 ligand fields (53 %).





Principal Component Analysis

The homoleptics (ho) span the strong symmetry (ss) and "5+1" (fo) set.



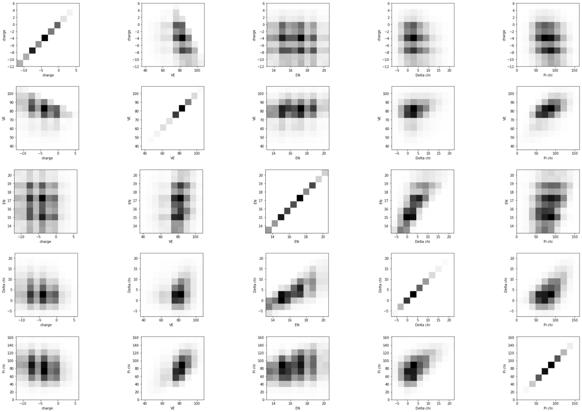
Footprint and Entropy calculation

We use five properties to characterize the ligand field and generate a five dimensional distribution:

- total charge
- total valence electrons
- electronegativity of the connecting atom
- ${}^{\mathrm{lc}}_{\mathrm{ax,eq}}\chi_1 = \sum EN_{\mathrm{CA}} \cdot EN_i$
- ${}^{\mathrm{lc}}_{\mathrm{ax,eq}}\chi'_1 = \sum EN_{\mathrm{CA}} - EN_i$

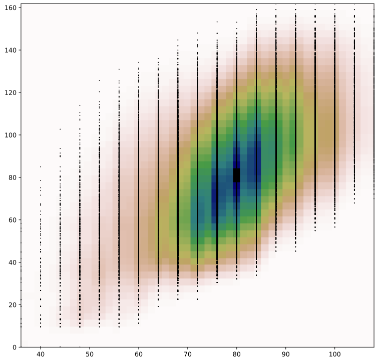
We then calculate the entropy, H_{KDE}, of the Kernel Density Estimated distrbution.

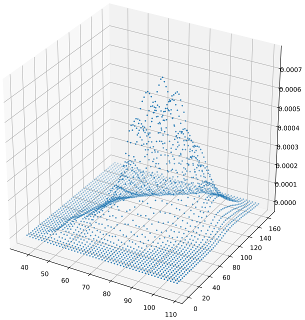
Correlation analysis for strongly symmetric monodentates



Example of KDE slice

Dimensions ${}^{\mathrm{lc}}_{\mathrm{ax,eq}}\chi_1$ vs. charge in H_{KDE} for strongly symmetric monodentates.



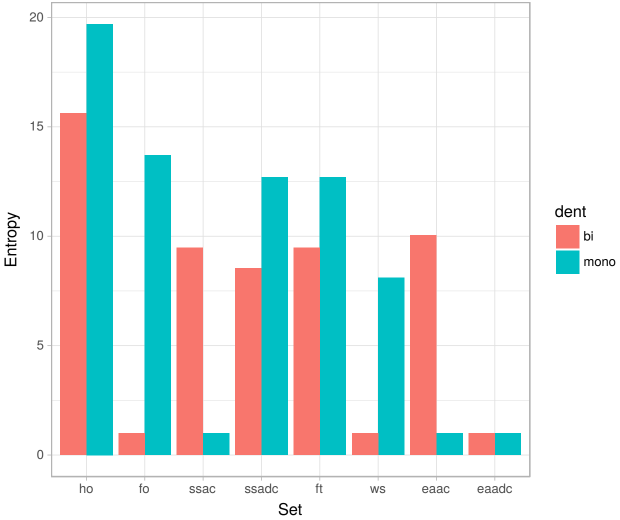


Monodentate Footprints

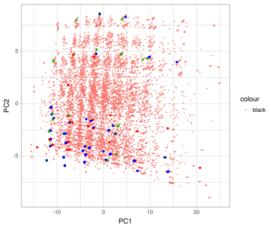
Table : Entropic footprint

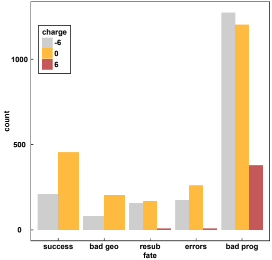
Set	$H_{\text{KDE}}^{\text{monodent}}$	$H_{\text{KDE}}^{\text{bident}}$
Homoleptics	19.7	15.63
"5+1" symmetric	13.7	-
Strongly symmetric AC	-	9.47
Strongly symmetric ADC	12.70	5.53
"4+2" symmetric	12.70	9.47
Weakly symmetric	8.1	7.7
Equatorially asymmetric AC	-	10.04
Equatorially asymmetric ADC		

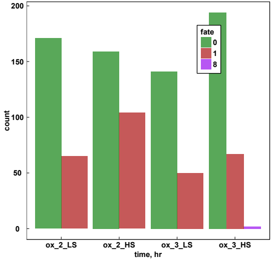
Entropy histogram

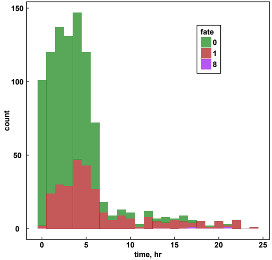


Actual calculations









- 0: [NH3], [N]#[N], [C+]#[O-], [C+]#[NH-], [N]#[CH]
- 1: [CH+]=[CH3-], [NH]=[O], [CH+]=[OH-], [CH2+]-[OH2-]
- 8: [OH2-]-[PH+], [P+]=[OH-], [PH+]-[OH2-], [NH2-]=[CH+]