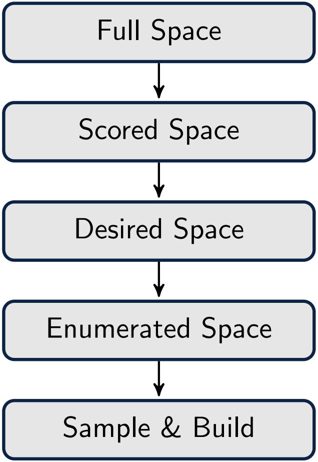
GDB-9 & Small Molecule Universes

Stefan O. Gugler

Massachusetts Institute of Technology
Department of Chemical Engineering

May 31, 2018

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



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

Full Space:
M:125, D:5625, B:625

↓

Scored Space:
50, 1500, 201

1 charge, sterics, octet, shell, bond order

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

$\Big\}$ 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?

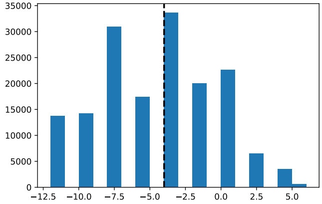
Subsets of octahedral space

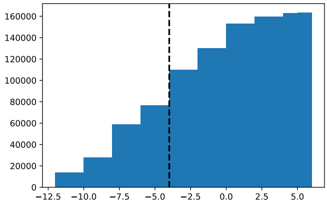
The sizes of the selected subsets of octahedral space.

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

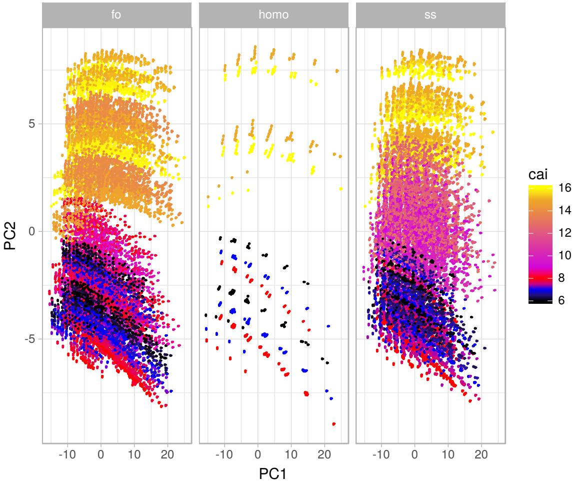
- Reduce space to facilitate sampling from non-homoleptics
- Example: strongly symmetric, monodentate ligand fields (163,620)
- Exclude all with charge smaller than -4, which results in 87,150 ligand fields (53 %).





Principal Component Analysis

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



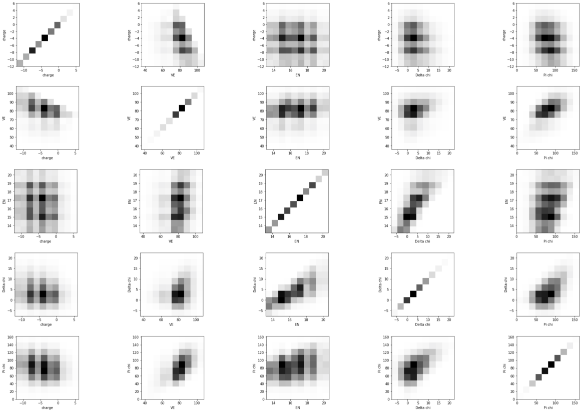
Footprint and Entropy calculation

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
- ${}^{\mathrm{lc}}_{\mathrm{ax,eq}}\chi_1 = \sum EN_{\mathrm{CA}} \cdot EN_i$
- ${}^{\mathrm{lc}}_{\mathrm{ax,eq}}\chi'_1 = \sum EN_{\mathrm{CA}} - EN_i$

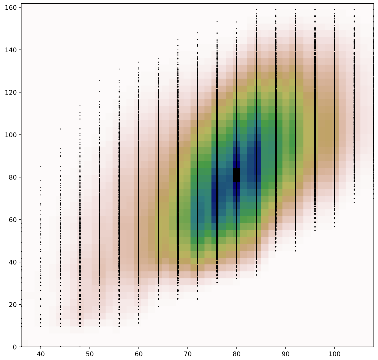
We then calculate the entropy, H_{KDE}, of the Kernel Density Estimated distrbution.

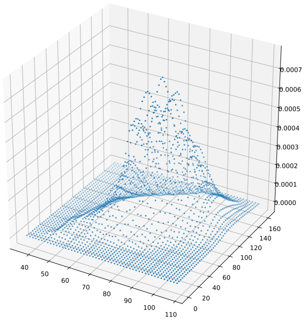
Correlation analysis for strongly symmetric monodentates



Example of KDE slice

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



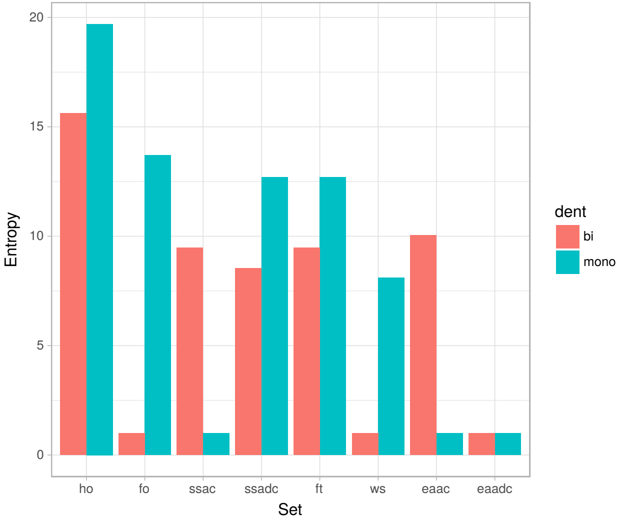


Monodentate Footprints

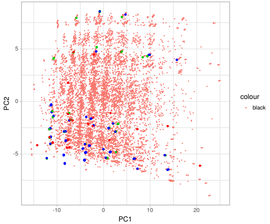
Table : Entropic footprint

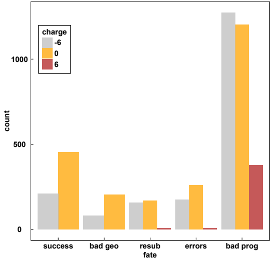
Set	$H_{\mathrm{KDE}}^{\mathrm{monodent}}$	$H_{\mathrm{KDE}}^{\mathrm{bident}}$
Homoleptics	19.7	15.63
"5+1" symmetric	13.7	-
Strongly symmetric AC	-	9.47
Strongly symmetric ADC	12.70	5.53
"4+2" symmetric	12.70	9.47
Weakly symmetric	8.1	7.7
Equatorially asymmetric AC	-	10.04
Equatorially asymmetric ADC		

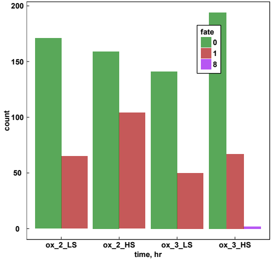
Entropy histogram

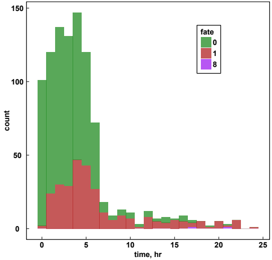


Actual calculations









- 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+]