ML for inorganic molecular design: descriptors and similarity in transition metal chemical space

Jon Paul Janet [1] Heather Kulik [1]

[1] Department of Chemical Engineering, Massachusetts Institute of Technology

255th ACS National Meeting, New Orleans

03.19.18

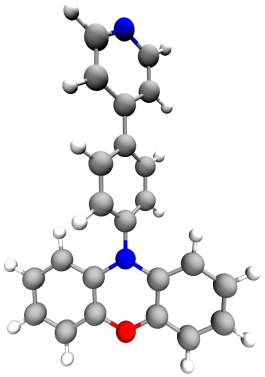
Data-driven molecular design

Machine learning is transforming how we design new materials...

Data-driven molecular design

Machine learning is transforming how we design new materials...

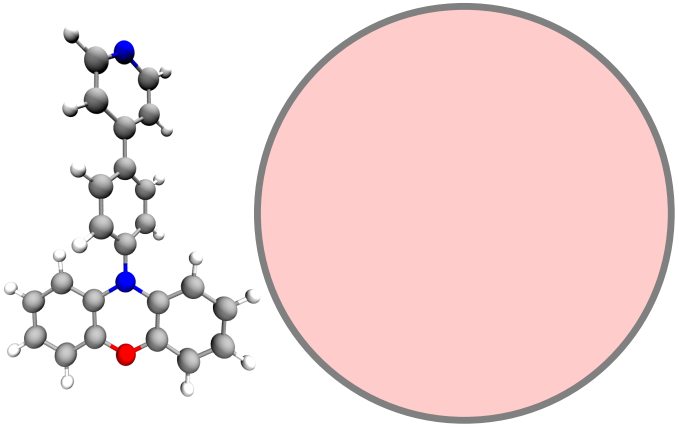
Gomez-Bombarelli, R. *et al.*. *Nat. Mater.*, 15(10):1120-1127, 2016.



Data-driven molecular design

Machine learning is transforming how we design new materials...

Gomez-Bombarelli, R. *et al.*. *Nat. Mater.*, 15(10):1120-1127, 2016.

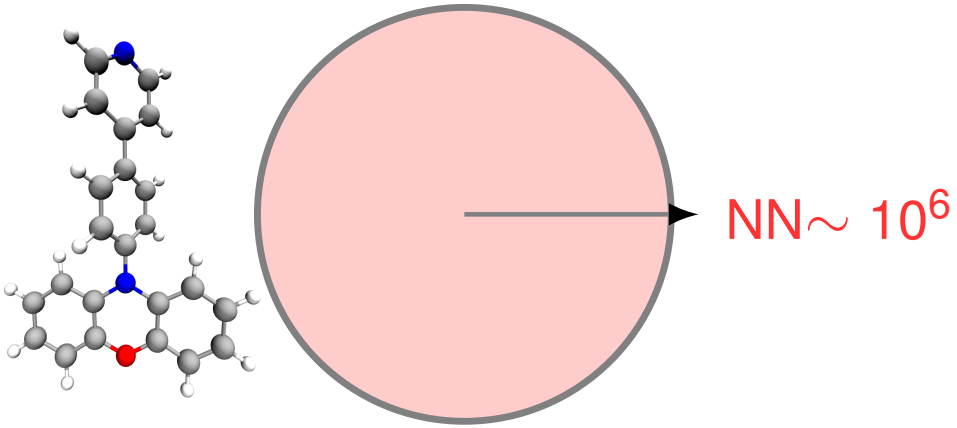


OLED chemical space

Data-driven molecular design

Machine learning is transforming how we design new materials...

Gomez-Bombarelli, R. *et al.*. *Nat. Mater.*, 15(10):1120-1127, 2016.

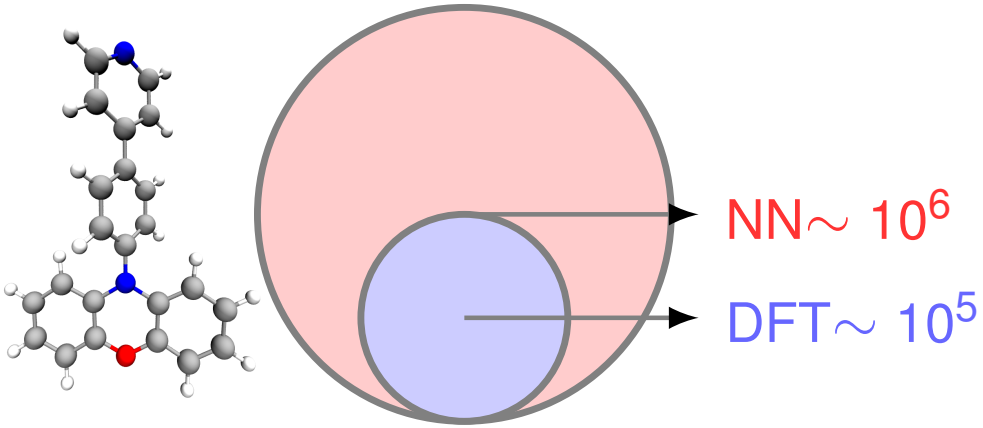


OLED chemical space

Data-driven molecular design

Machine learning is transforming how we design new materials...

Gomez-Bombarelli, R. *et al.*. *Nat. Mater.*, 15(10):1120-1127, 2016.

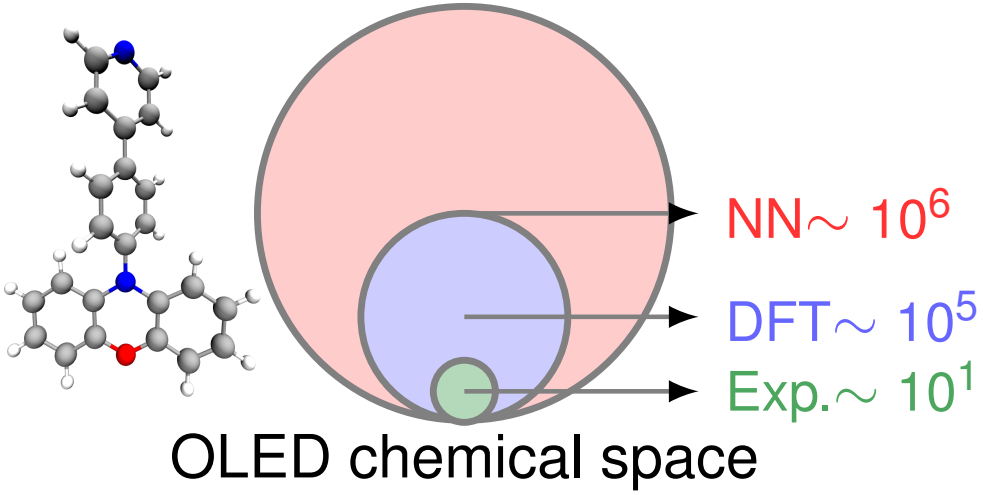


OLED chemical space

Data-driven molecular design

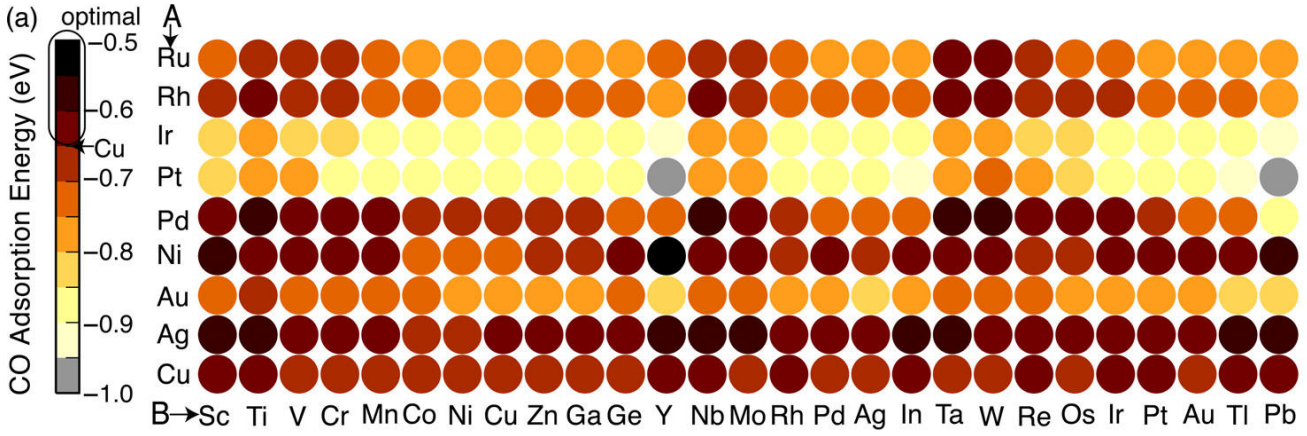
Machine learning is transforming how we design new materials...

Gomez-Bombarelli, R. *et al.*. *Nat. Mater.*, 15(10):1120-1127, 2016.



Data-driven molecular design

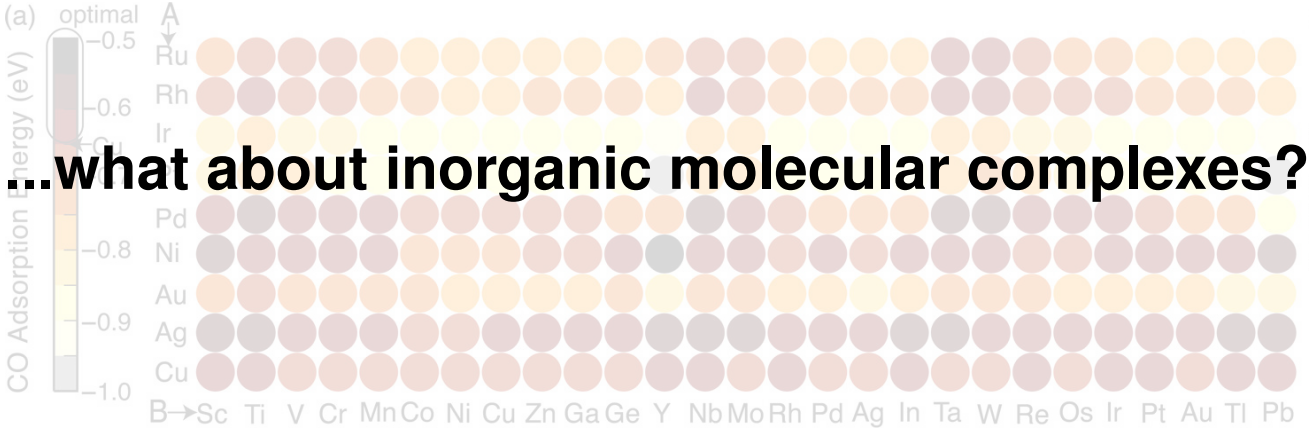
Machine learning is transforming how we design new materials...



Ma, X. *et al*. J. Phys. Chem. Lett., (18):3528-3533, 2015.

Data-driven molecular design

Machine learning is transforming how we design new materials...

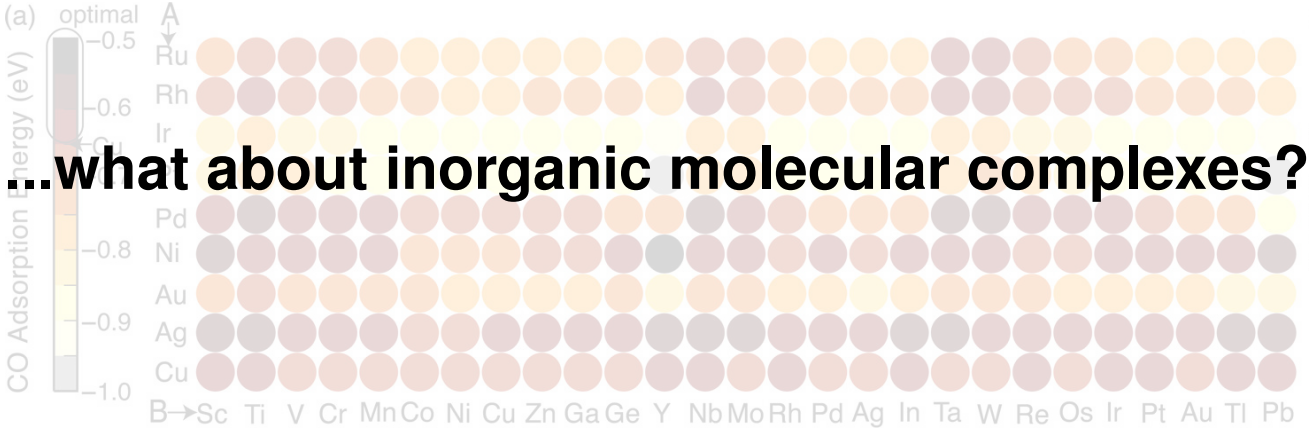


...what about inorganic molecular complexes?

Ma, X. *et al.* J. Phys. Chem. Lett., (18):3528-3533, 2015.

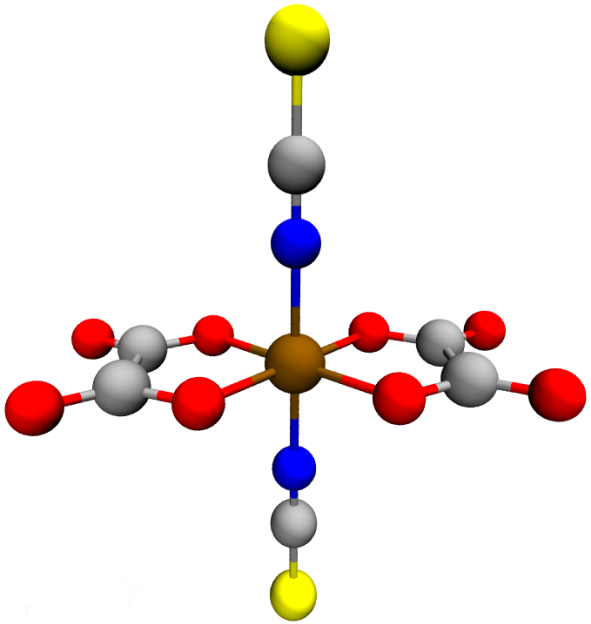
Data-driven molecular design

Machine learning is transforming how we design new materials...



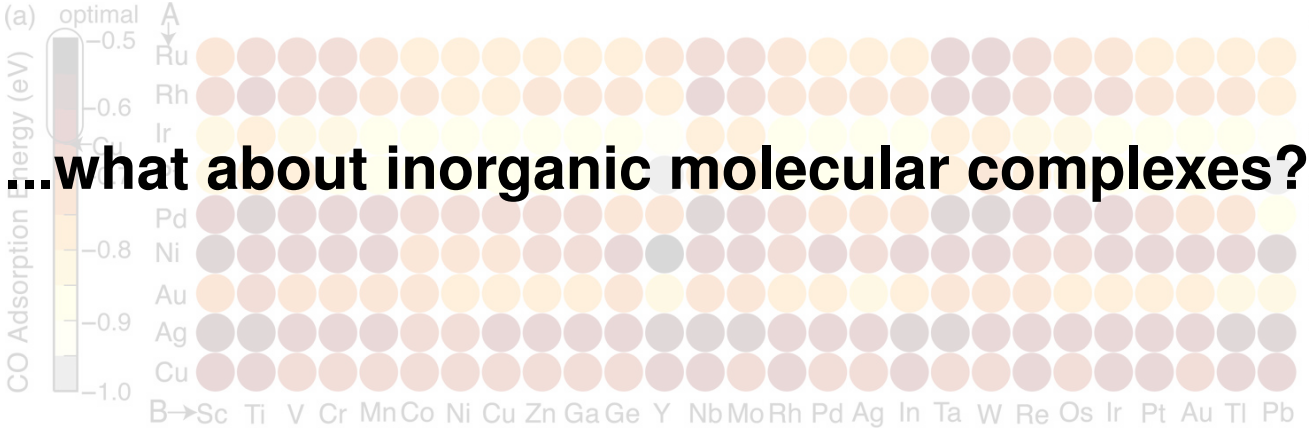
...what about inorganic molecular complexes?

Ma X. *et al.* J. Phys. Chem. Lett., (18):3528-3533, 2015.



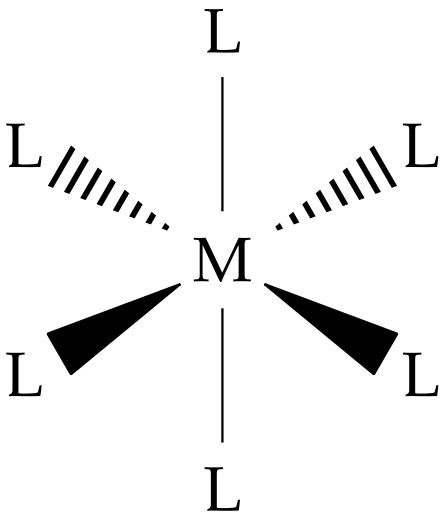
Data-driven molecular design

Machine learning is transforming how we design new materials...



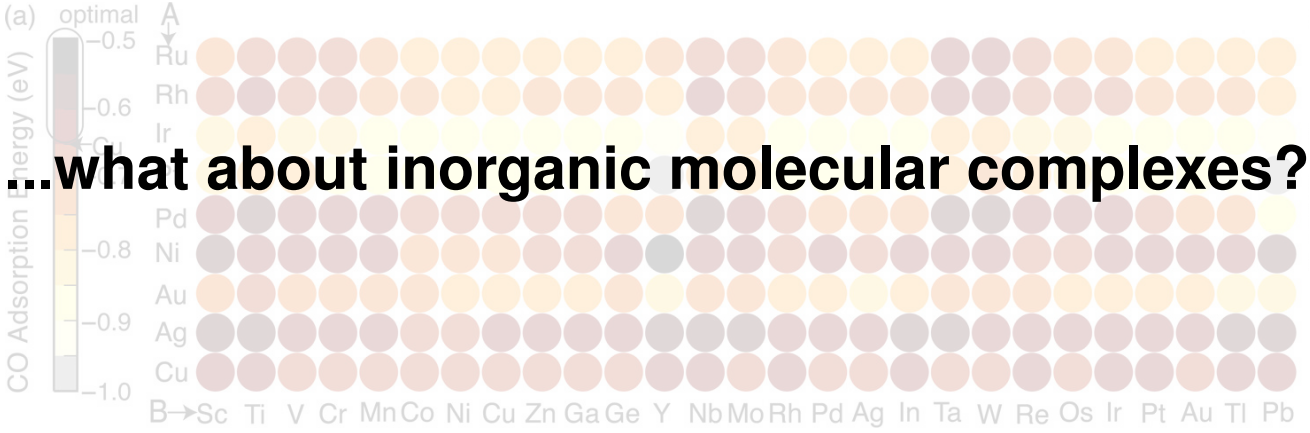
...what about inorganic molecular complexes?

Ma, X. *et al.* J. Phys. Chem. Lett., (18):3528-3533, 2015.



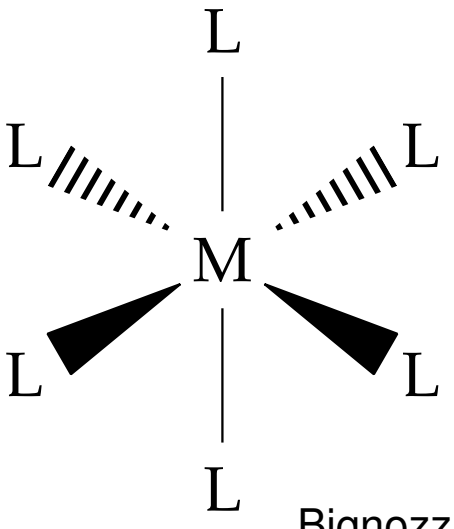
Data-driven molecular design

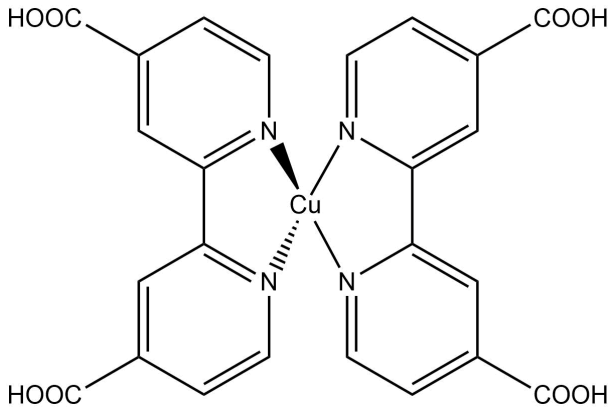
Machine learning is transforming how we design new materials...



...what about inorganic molecular complexes?

Ma, X. *et al.* J. Phys. Chem. Lett., (18):3528-3533, 2015.

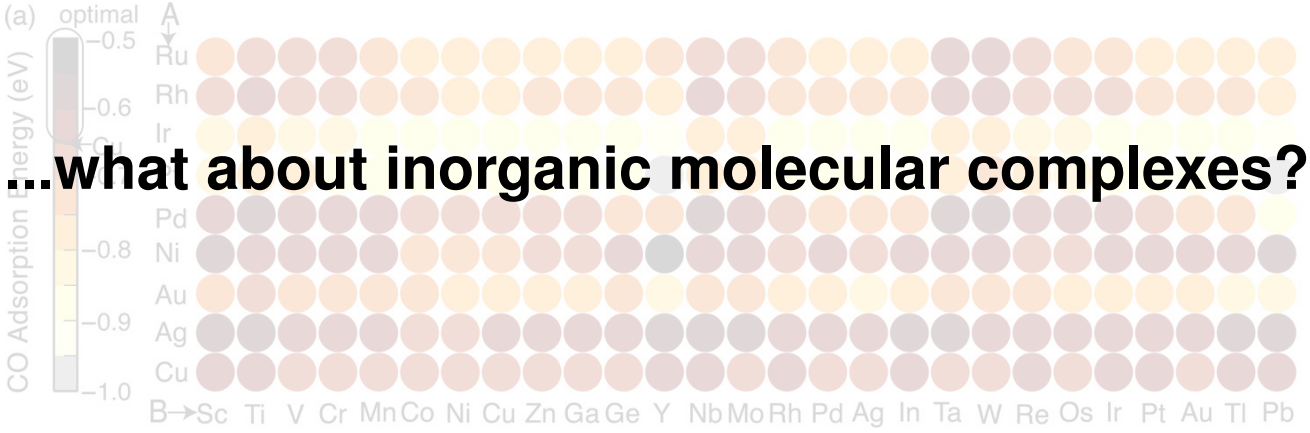




Bignozzi, C. *et al. Coord. Chem. Rev.*, 257(9), 2013.

Data-driven molecular design

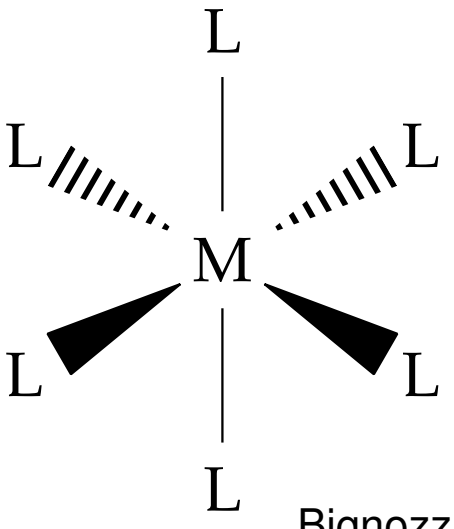
Machine learning is transforming how we design new materials...

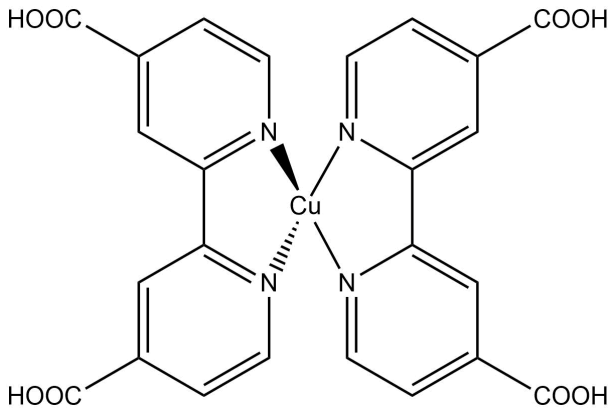


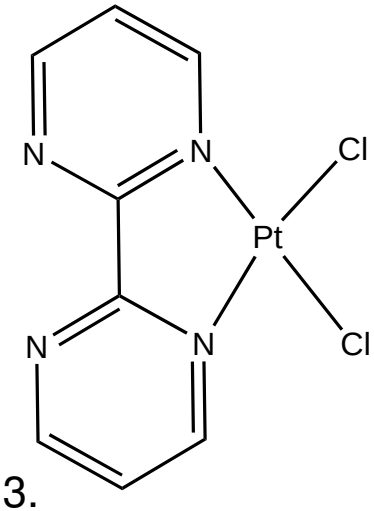
...what about inorganic molecular complexes?

Ma, X. *et al.* J. Phys. Chem. Lett., (18):3528-3533, 2015.

Periana, R. A. *et al.* *Science*, 280(5363), 1998.



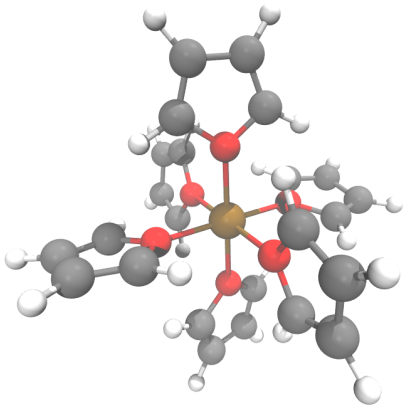




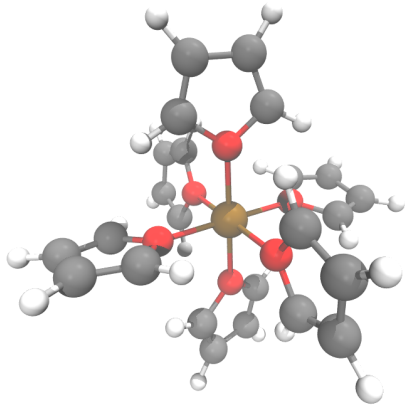
Bignozzi, C. *et al.* *Coord. Chem. Rev.*, 257(9), 2013.

Transition metal complexes

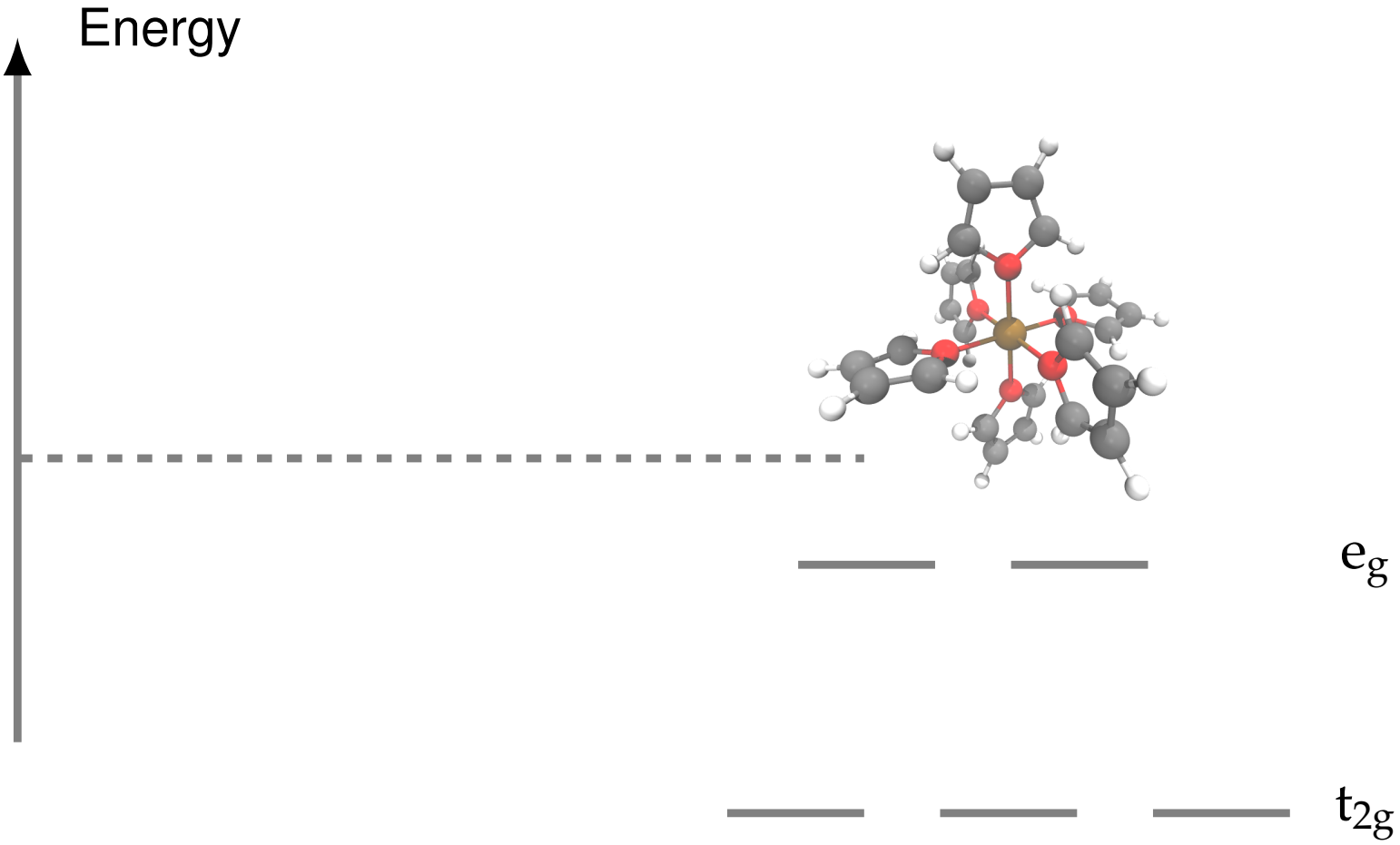
Transition metal complexes



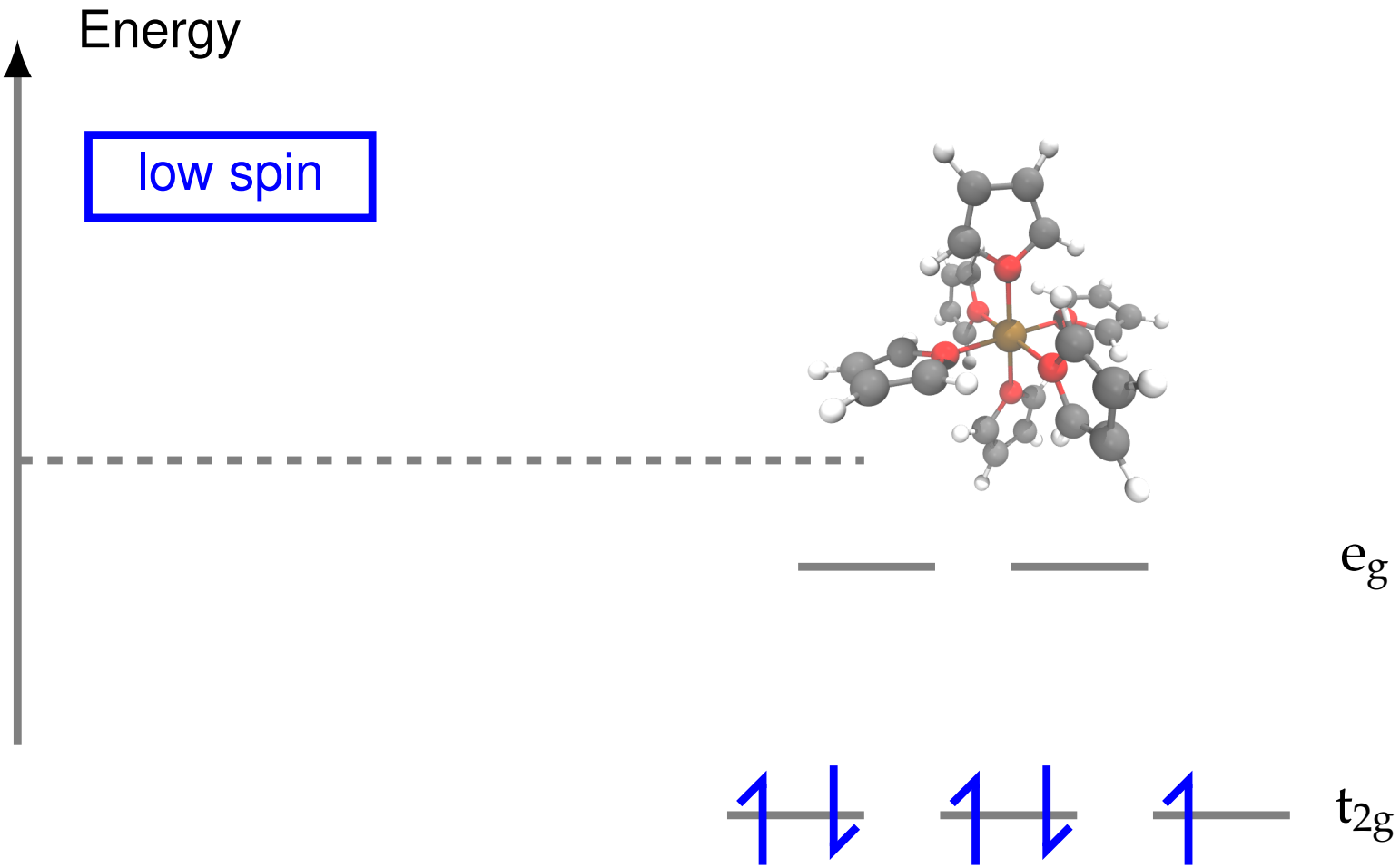
Transition metal complexes



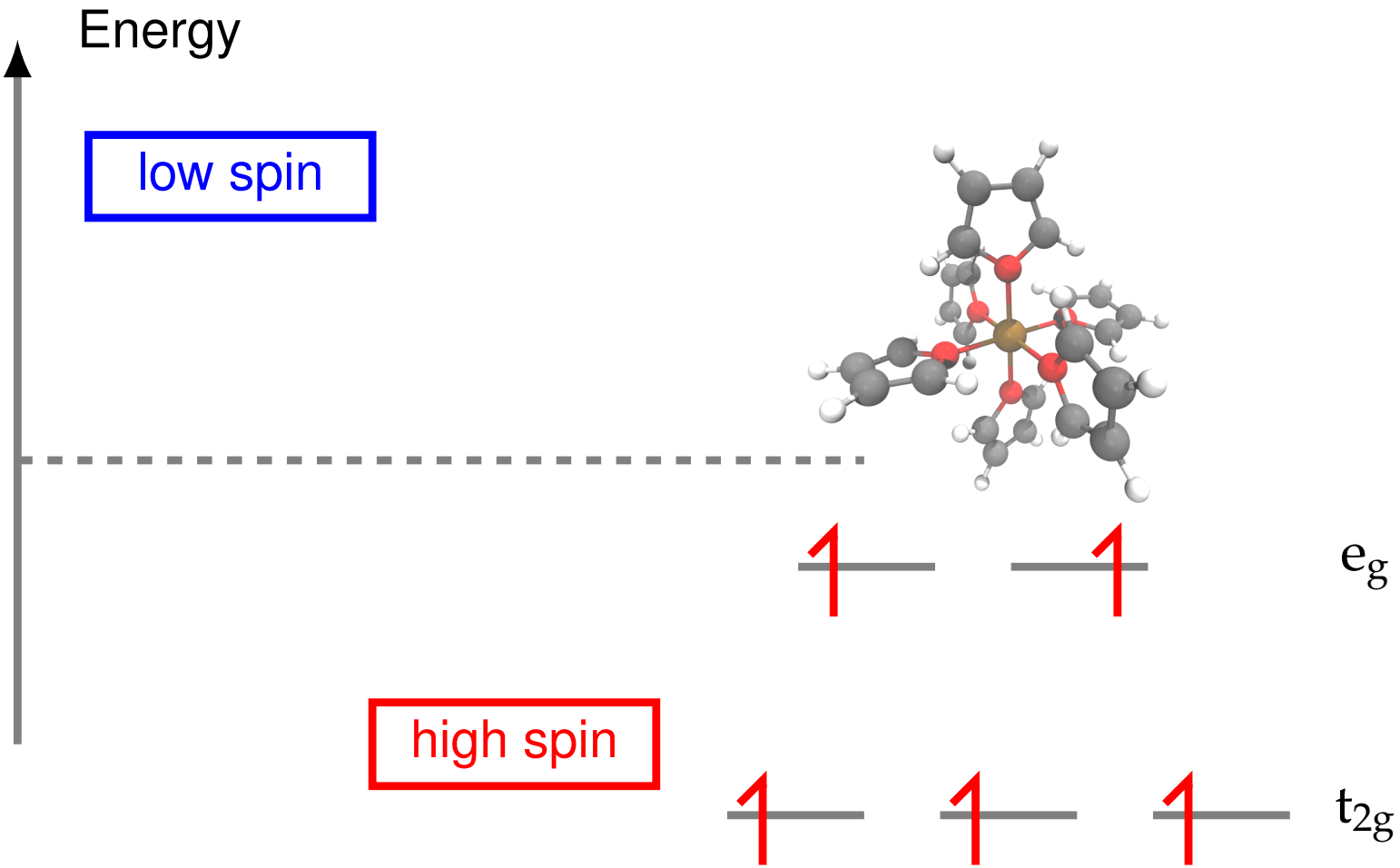
Transition metal complexes



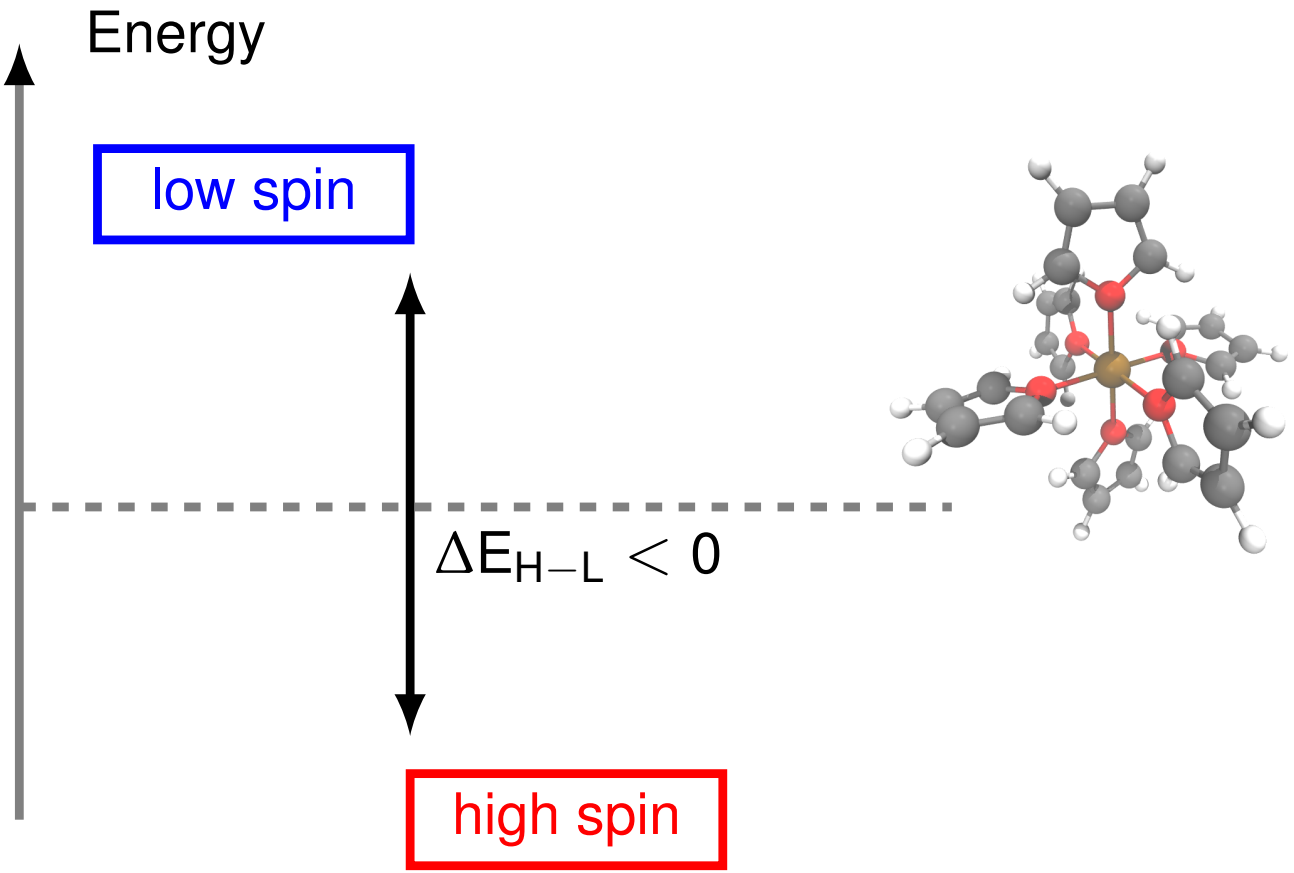
Transition metal complexes



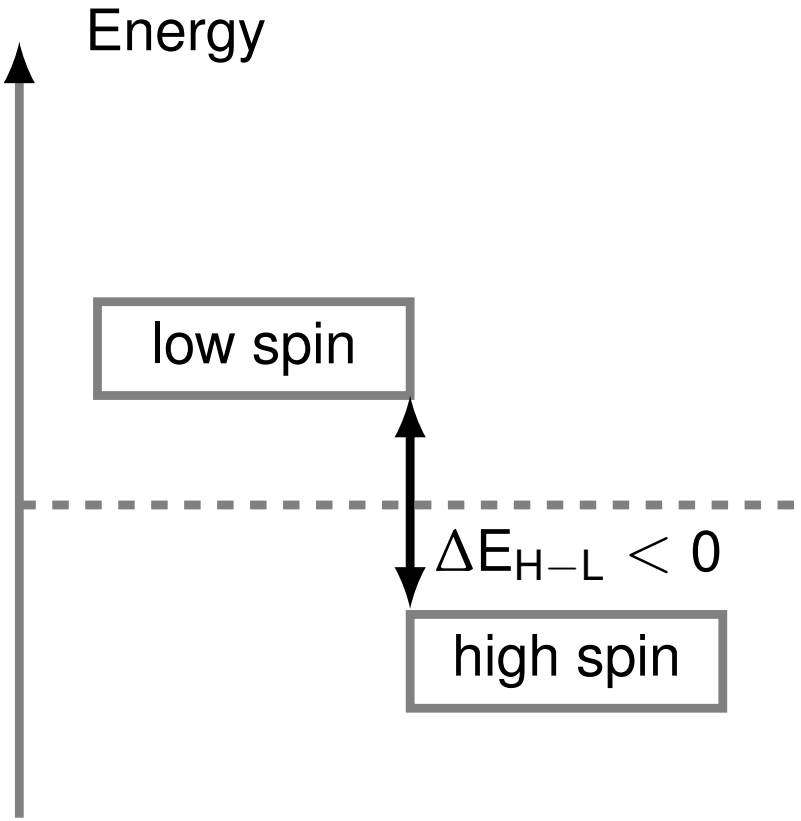
Transition metal complexes

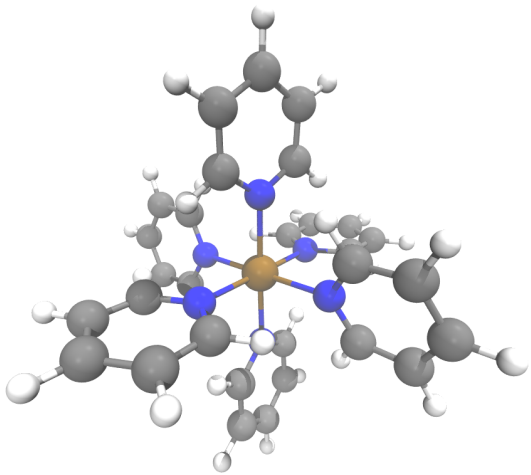


Transition metal complexes

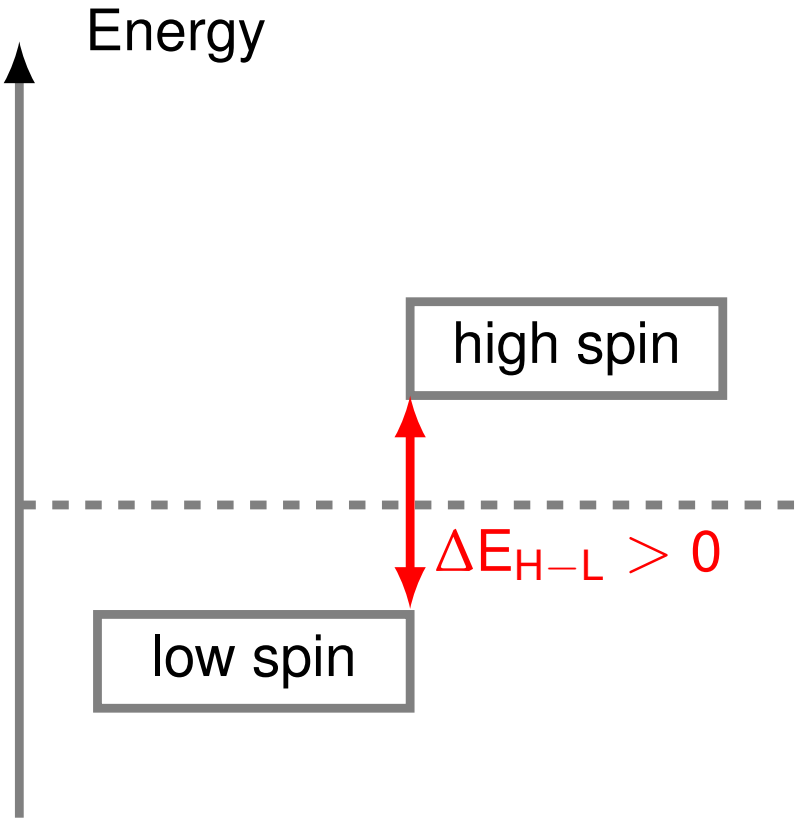


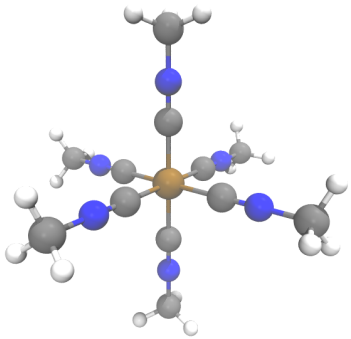
Transition metal complexes



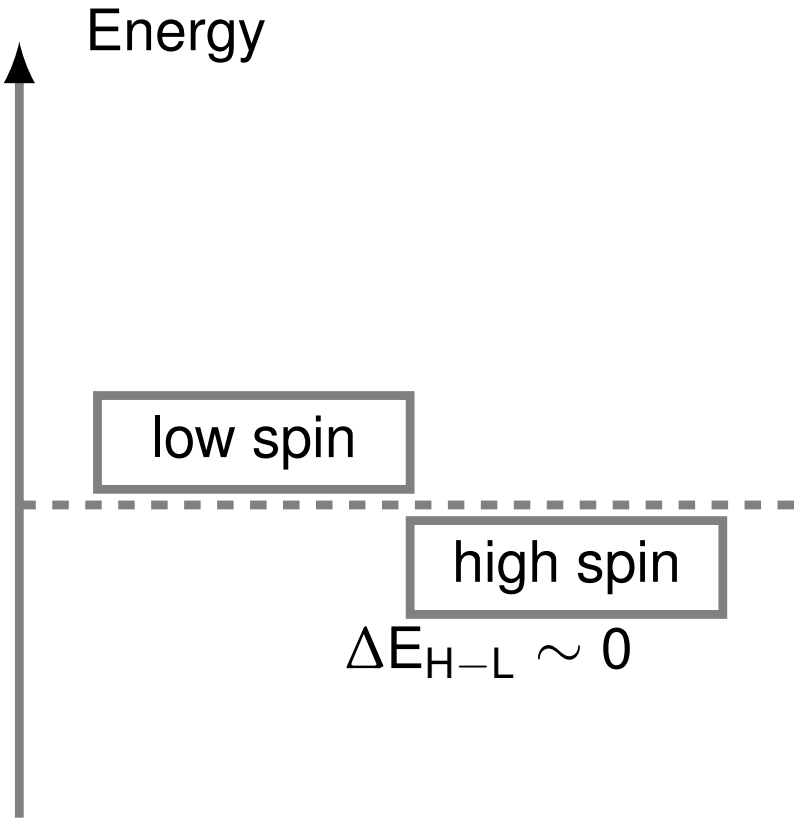


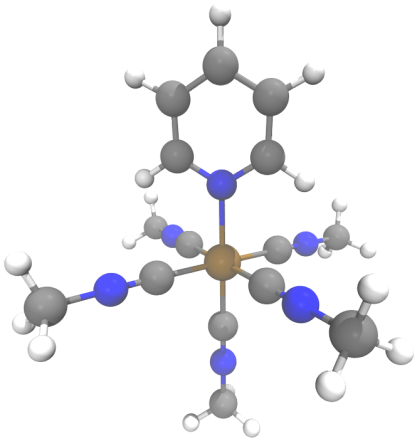
Transition metal complexes



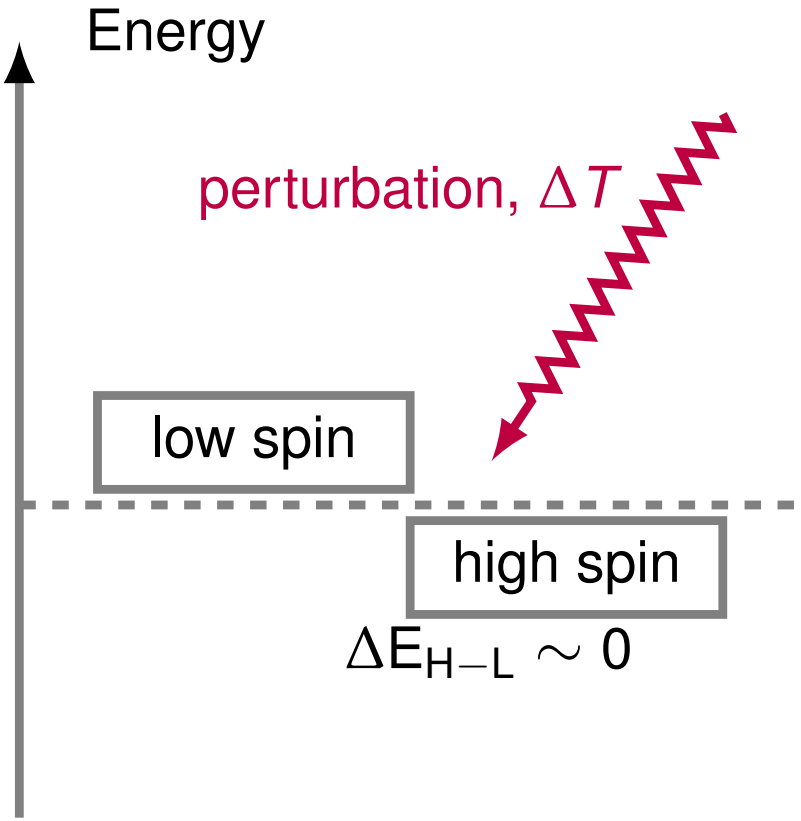


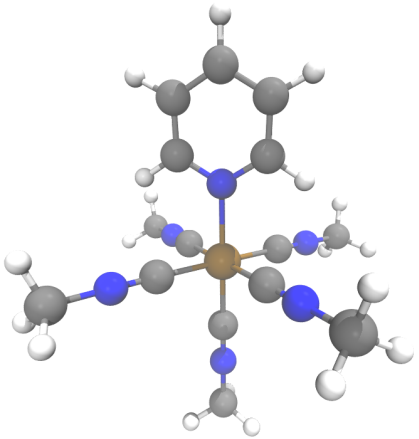
Transition metal complexes



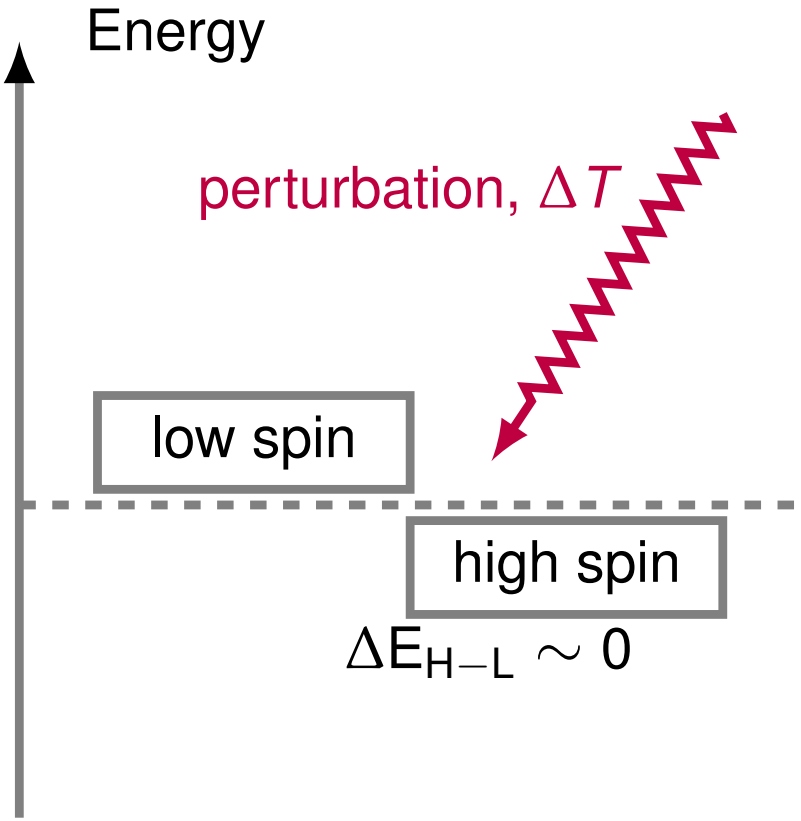


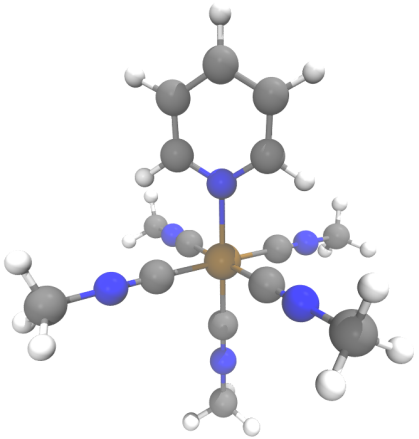
Transition metal complexes



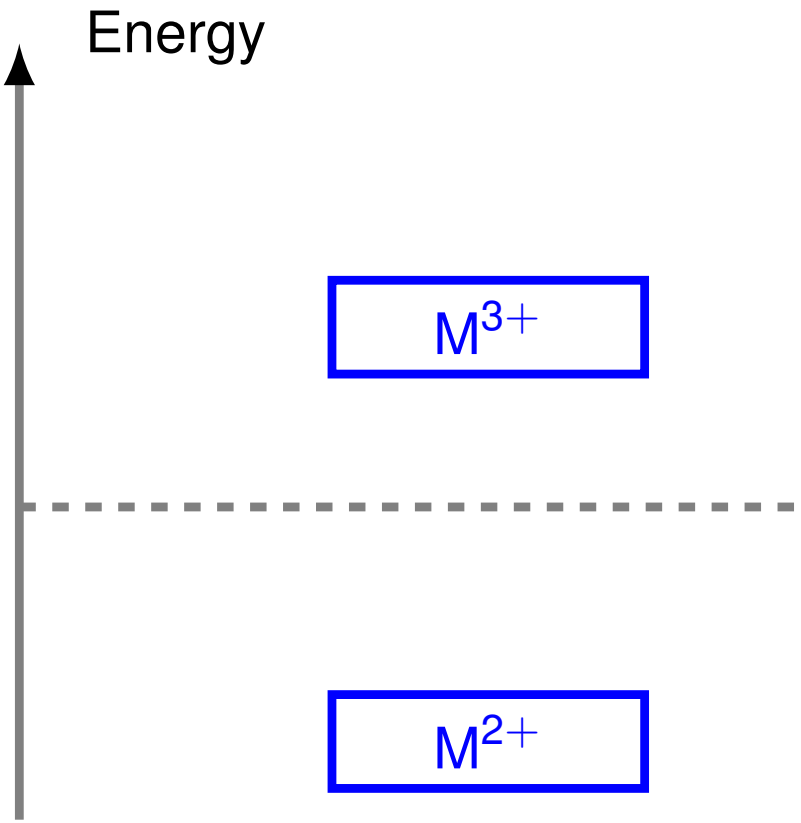


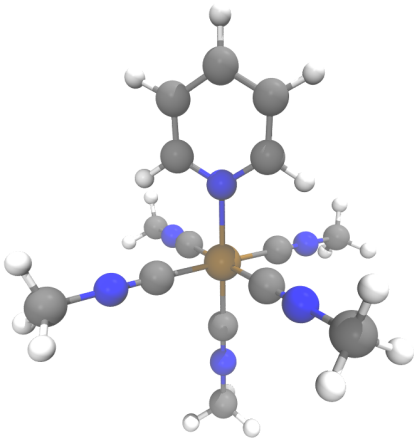
Transition metal complexes



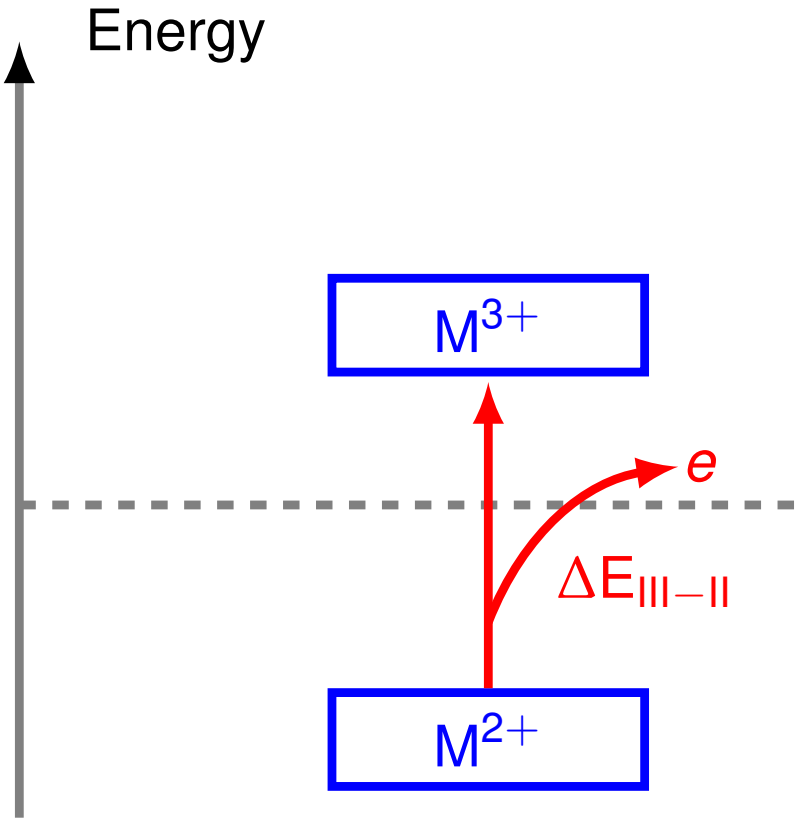


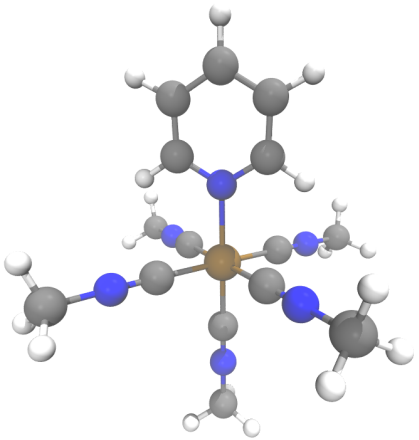
Transition metal complexes



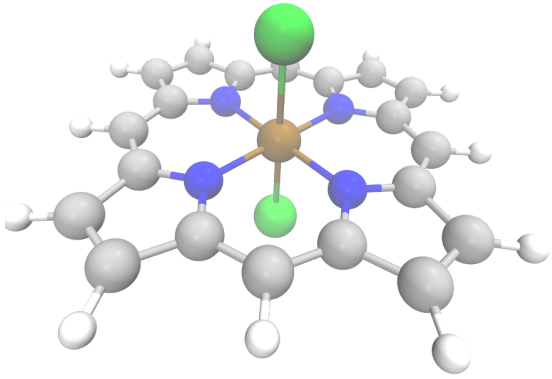


Transition metal complexes

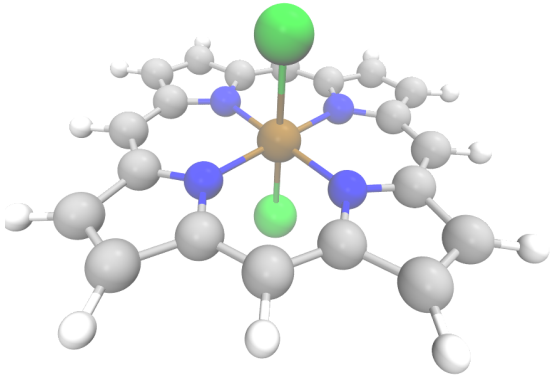




How to estimate properties?

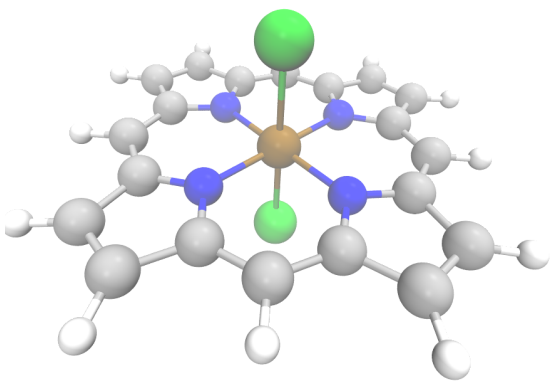


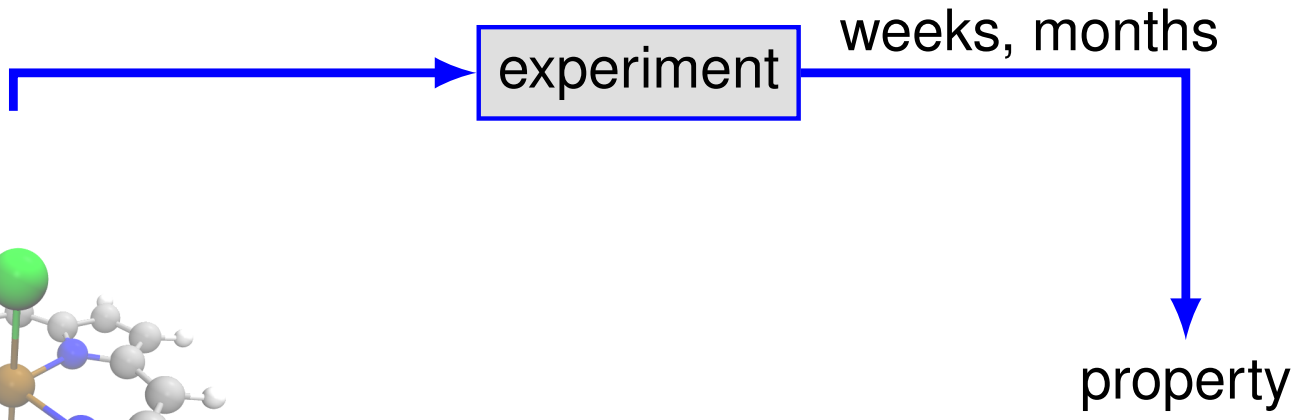
How to estimate properties?



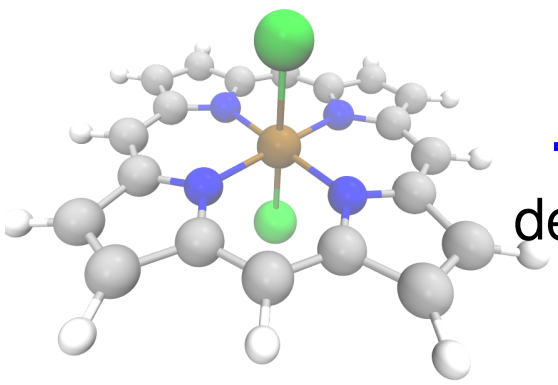
property

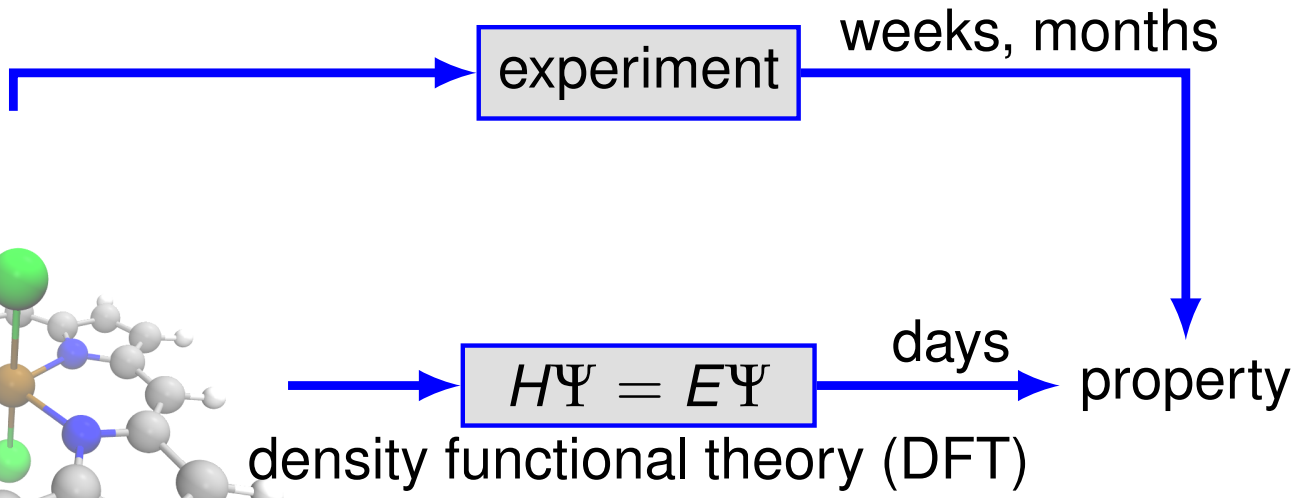
How to estimate properties?



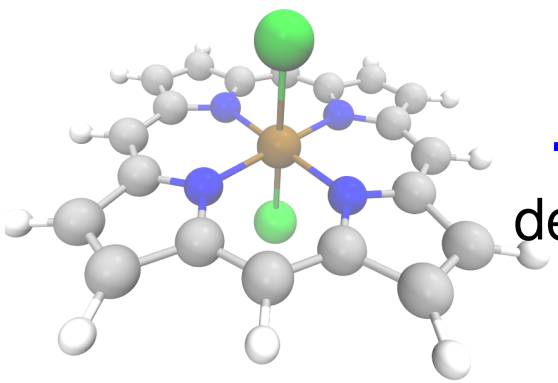


How to estimate properties?



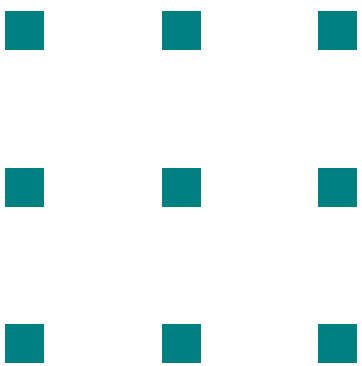


How to estimate properties?



Input space design

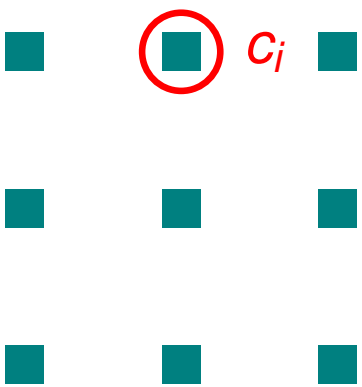
What would be the ideal feature space?



Chemical Space C_f

Input space design

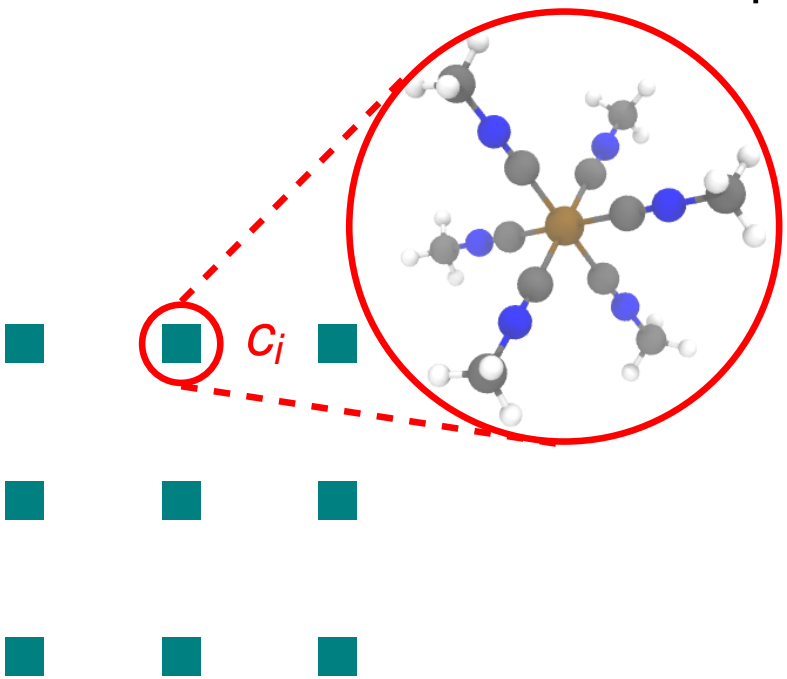
What would be the ideal feature space?



Chemical Space C_f

Input space design

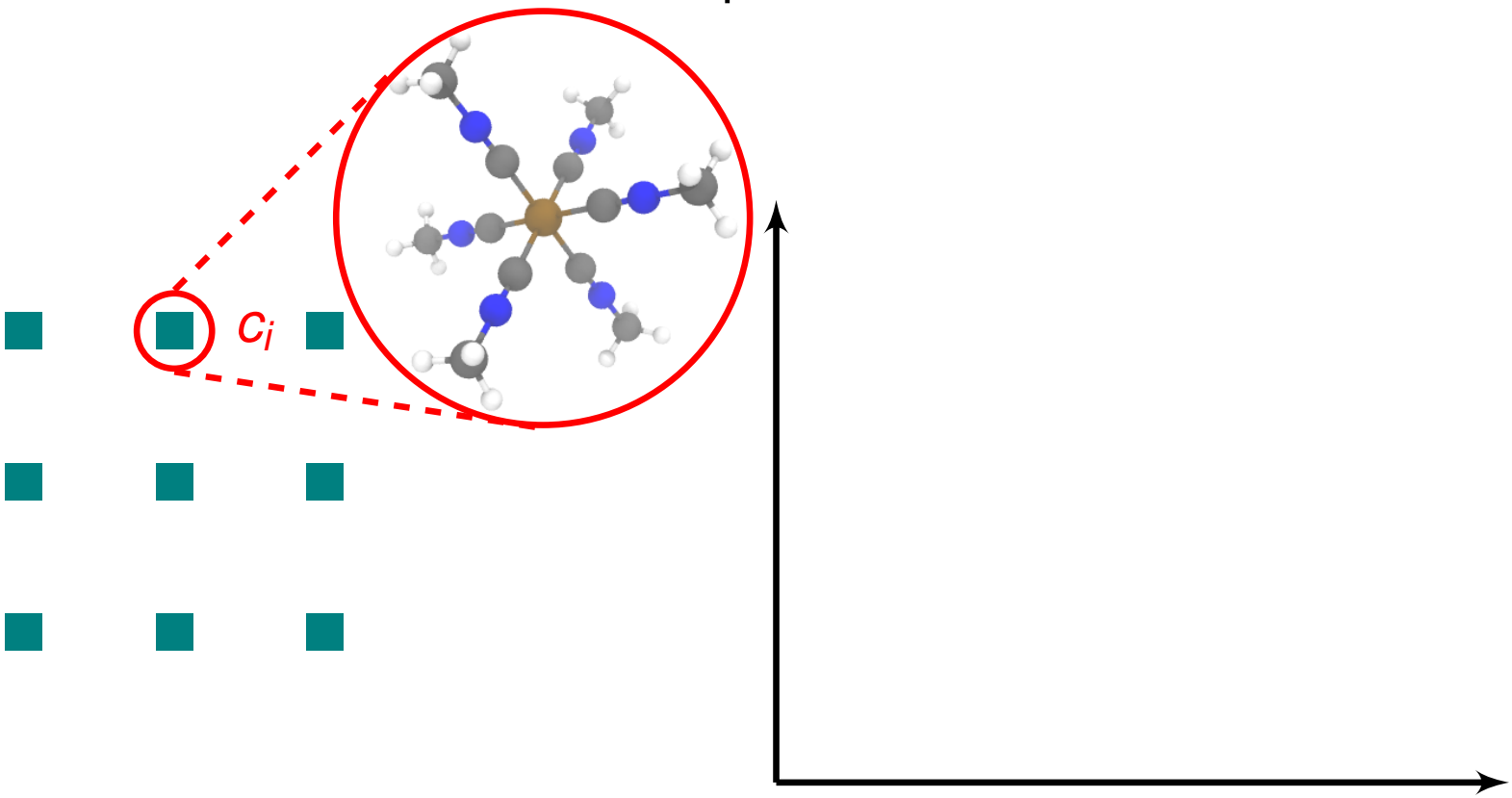
What would be the ideal feature space?



Chemical Space C_f

Input space design

What would be the ideal feature space?

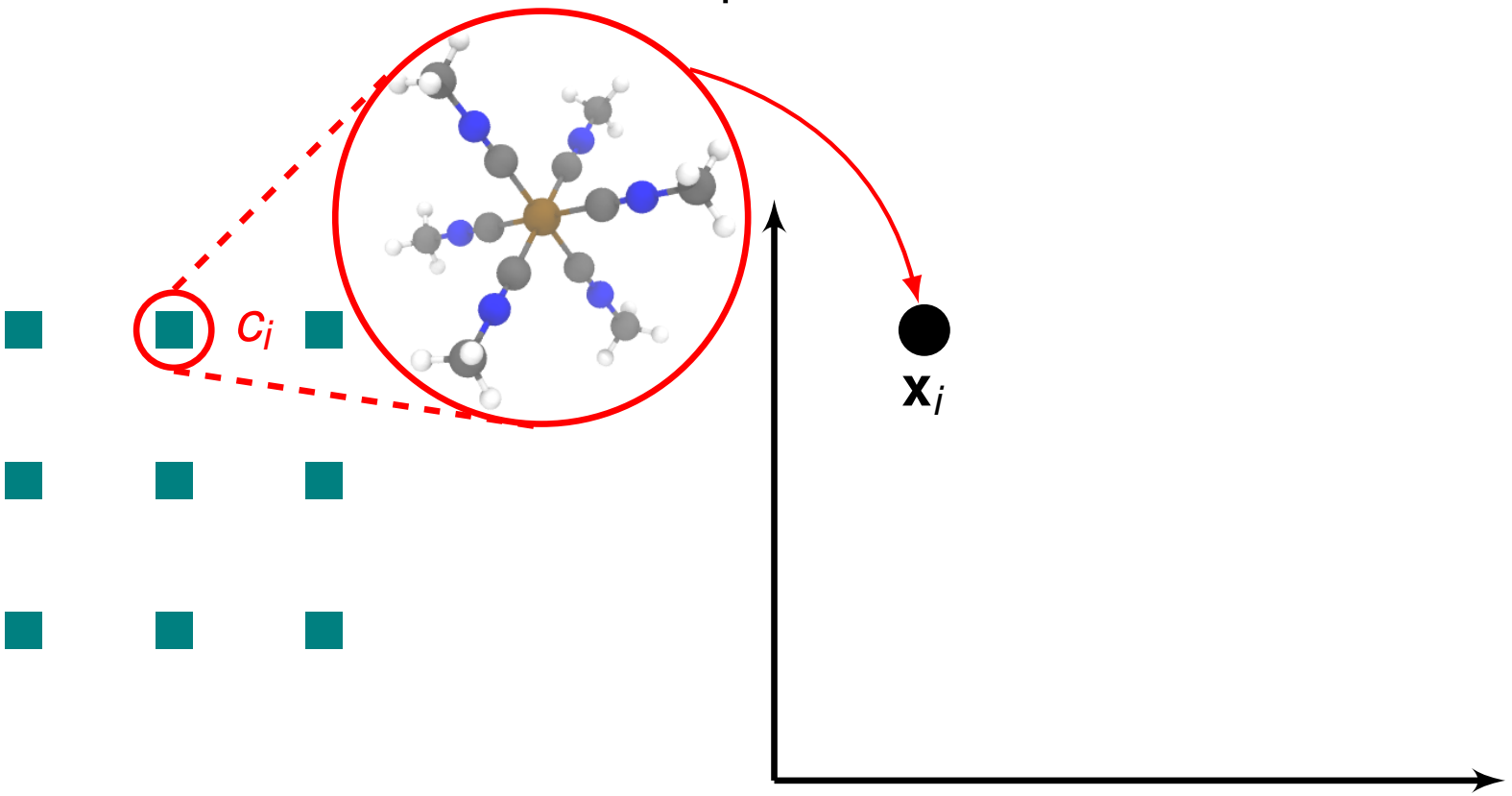


Chemical Space C_f

Descriptor Space $\mathcal{X} \subset \mathbb{R}^d$

Input space design

What would be the ideal feature space?

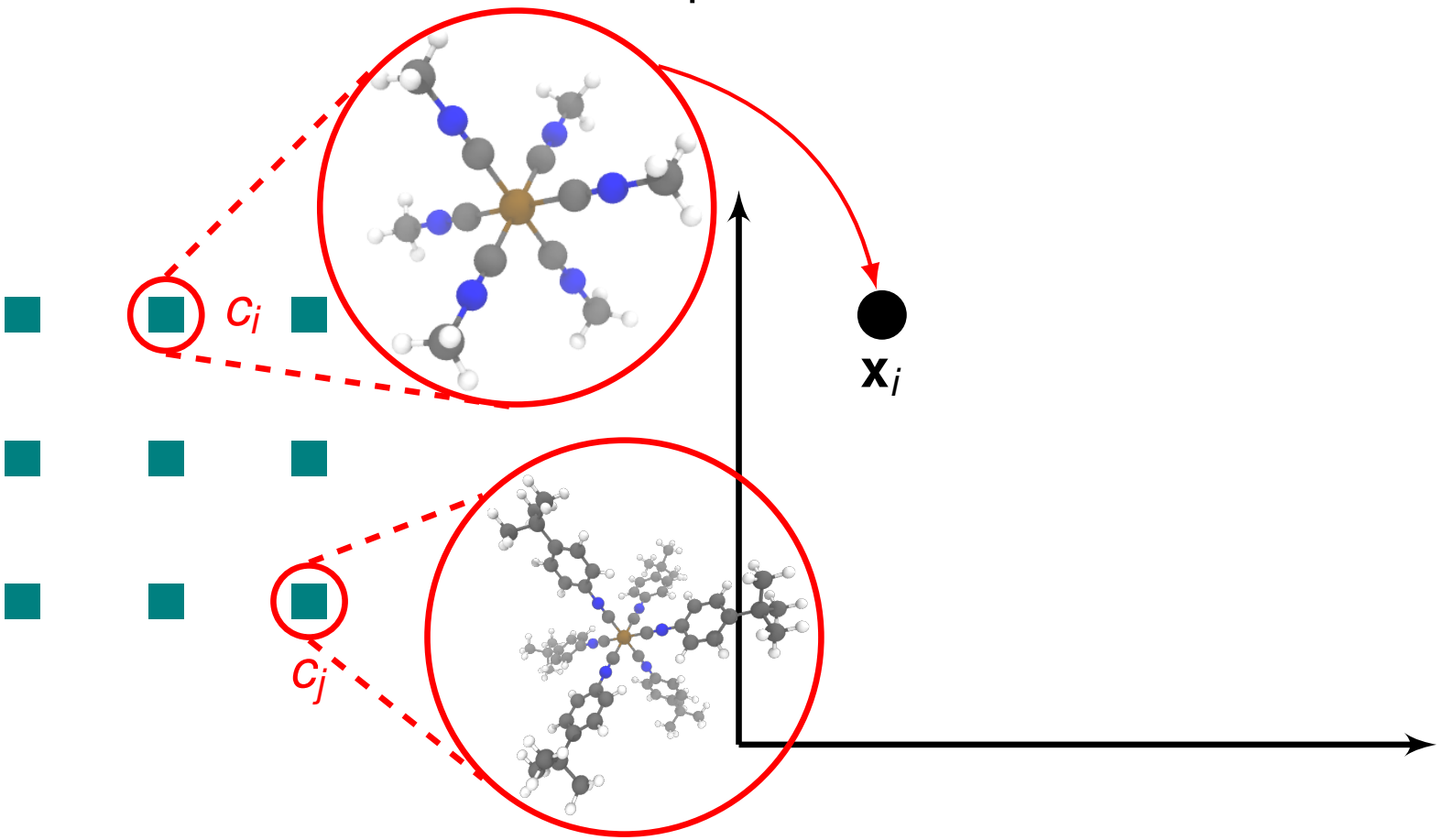


Chemical Space C_f

Descriptor Space $\mathcal{X} \subset \mathbb{R}^d$

Input space design

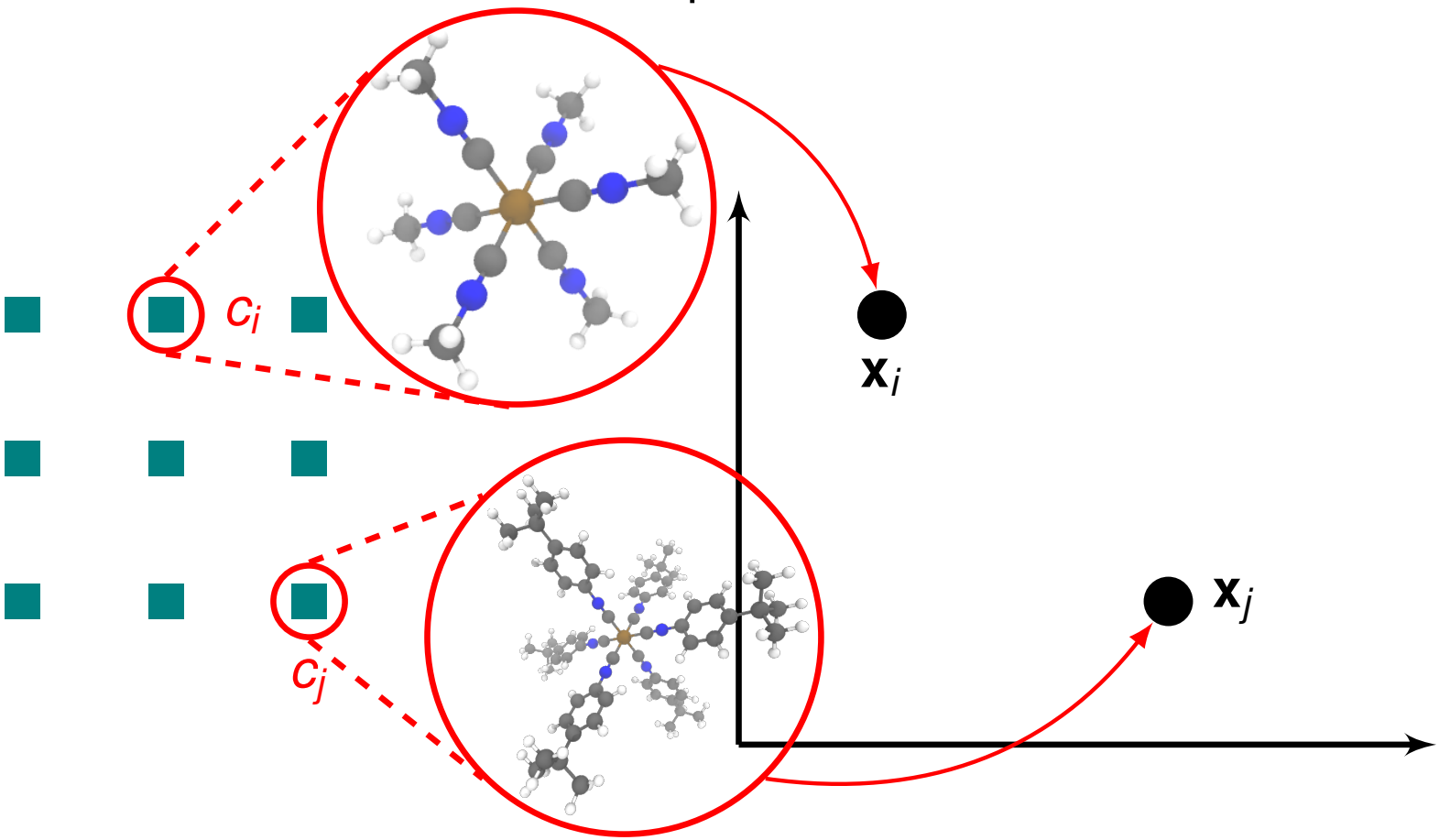
What would be the ideal feature space?



Chemical Space C_f **Descriptor Space** $\mathcal{X} \subset \mathbb{R}^d$

Input space design

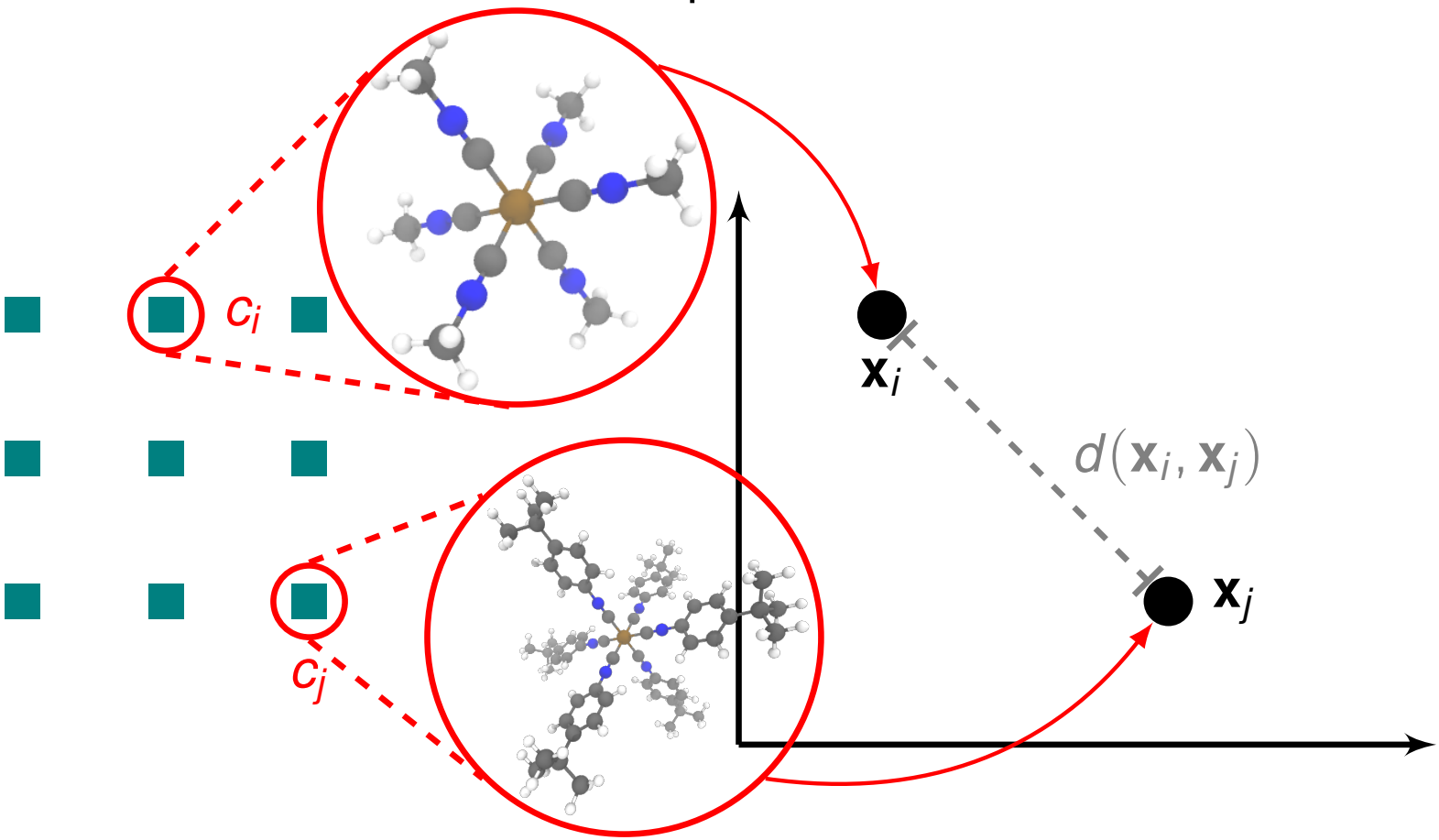
What would be the ideal feature space?



Chemical Space C_f **Descriptor Space** $\mathcal{X} \subset \mathbb{R}^d$

Input space design

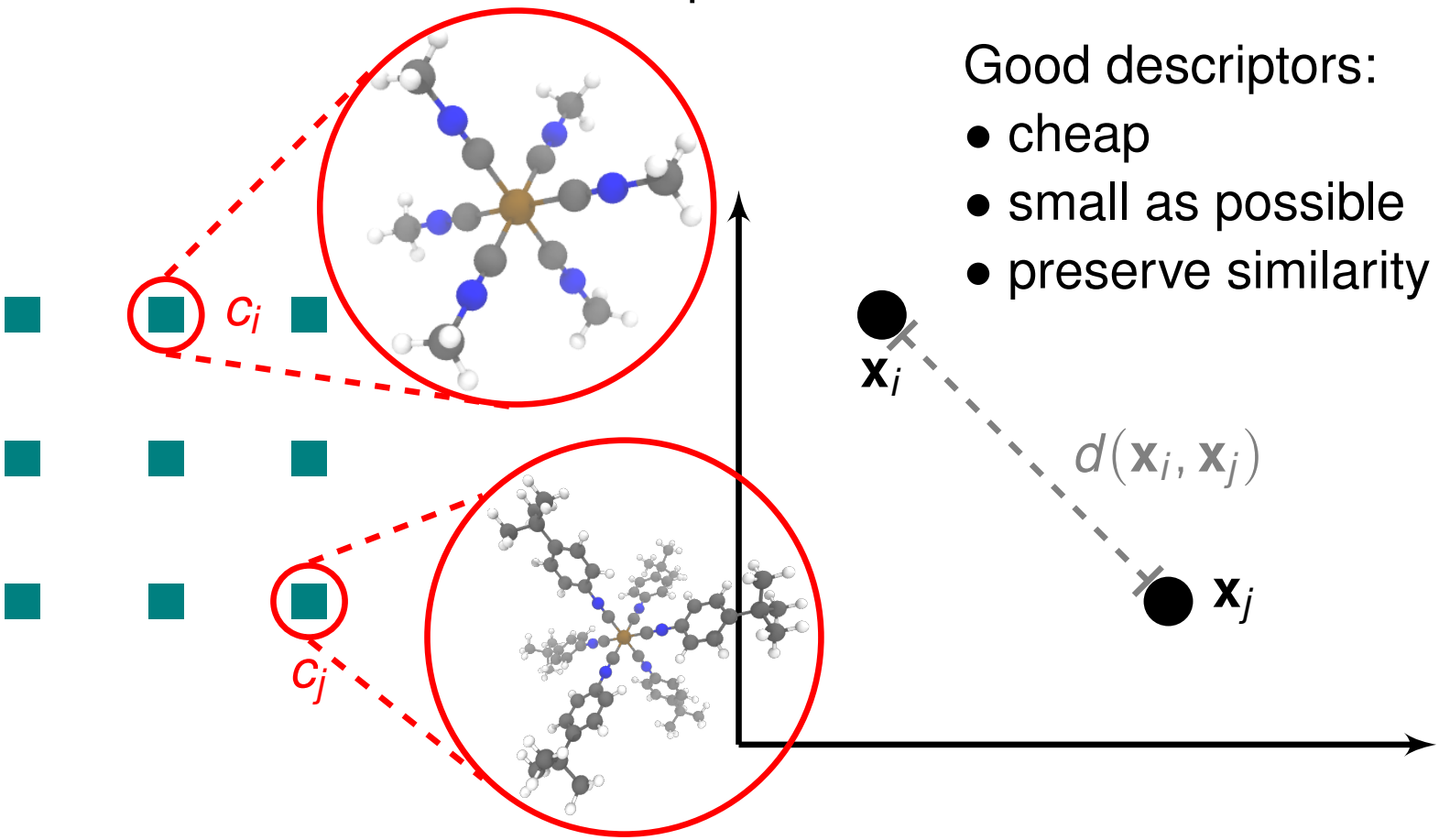
What would be the ideal feature space?



Chemical Space C_f **Descriptor Space** $\mathcal{X} \subset \mathbb{R}^d$

Input space design

What would be the ideal feature space?



Good descriptors:

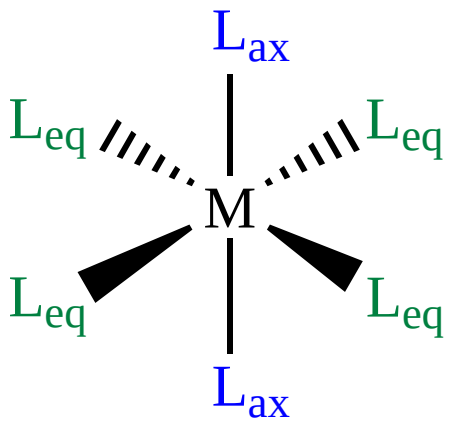
- cheap
- small as possible
- preserve similarity

Chemical Space C_f

Descriptor Space $\mathcal{X} \subset \mathbb{R}^d$

Data for spin splitting

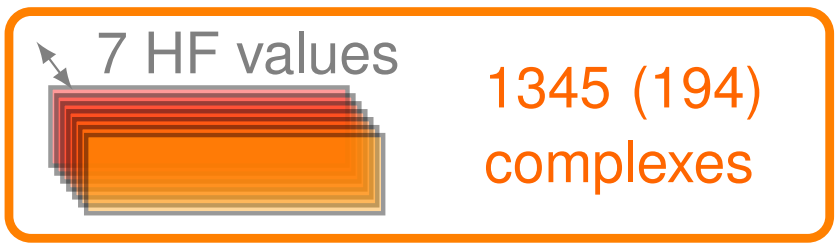
Data for octahedral complexes[1]:

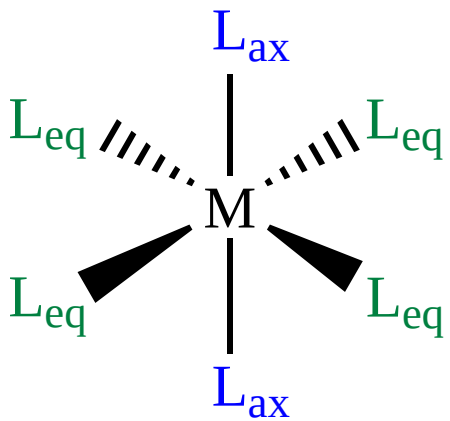


[1] Janet, J.P., and Kulik, H.J. *Chem. Sci.*, 2017, 8, 5137-5152.

Data for spin splitting

Data for octahedral complexes[1]:

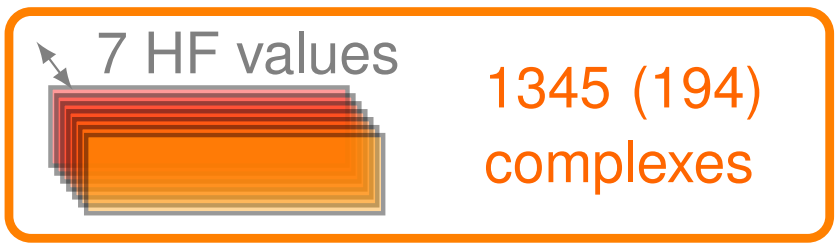


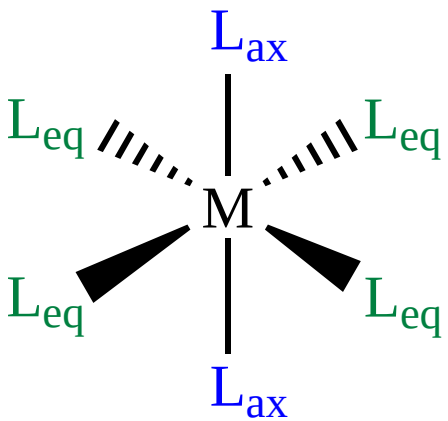


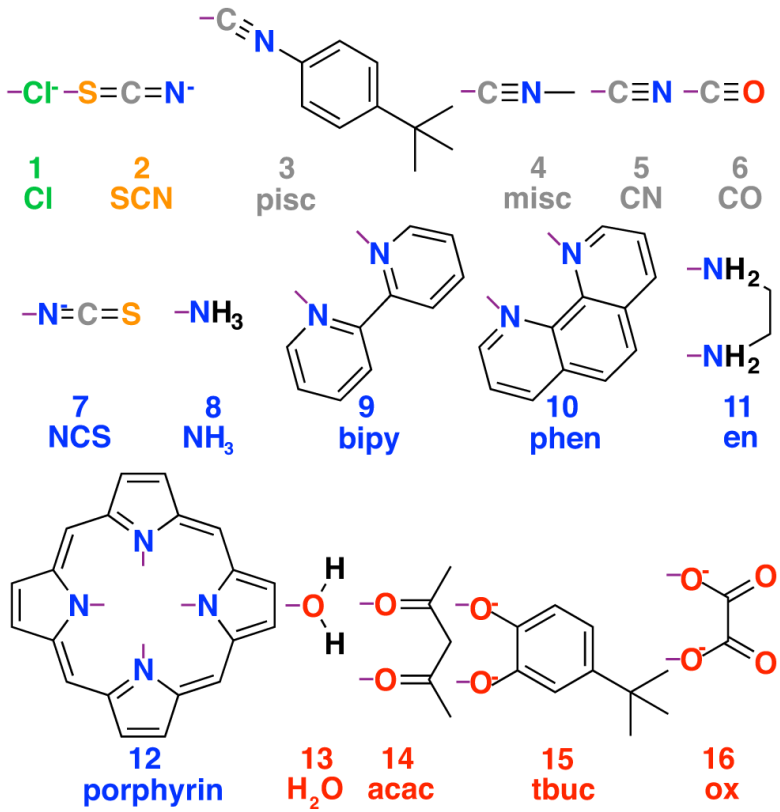
[1] Janet, J.P., and Kulik, H.J. *Chem. Sci.*, 2017, 8, 5137-5152.

Data for spin splitting

Data for octahedral complexes[1]:



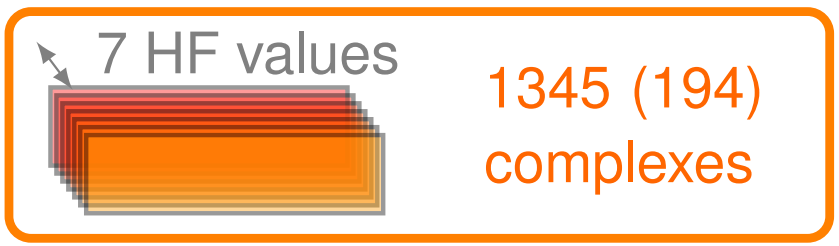


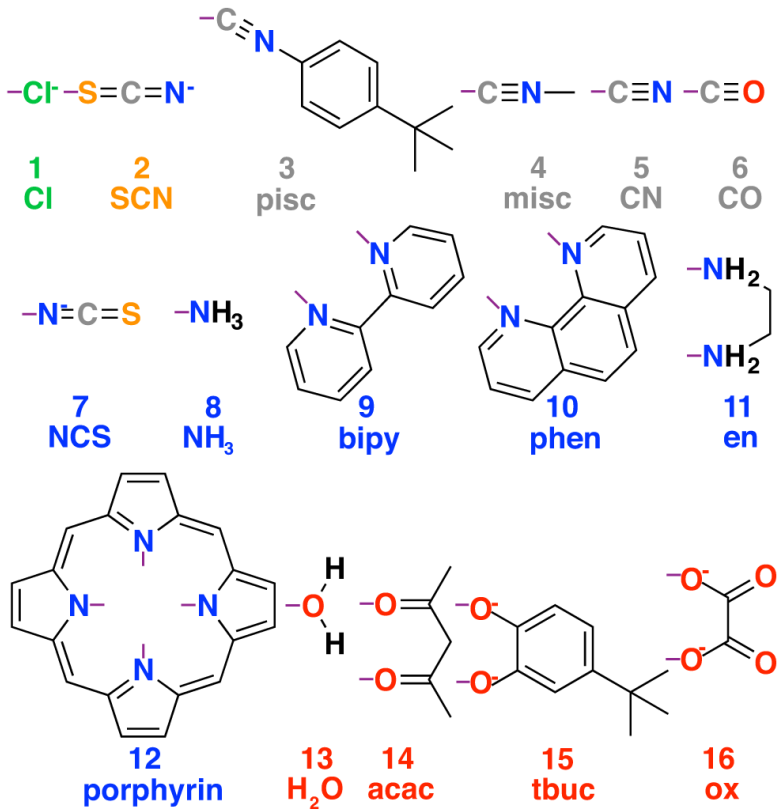


[1] Janet, J.P., and Kulik, H.J. *Chem. Sci.*, 2017, 8, 5137-5152.

Data for spin splitting

Data for octahedral complexes[1]:



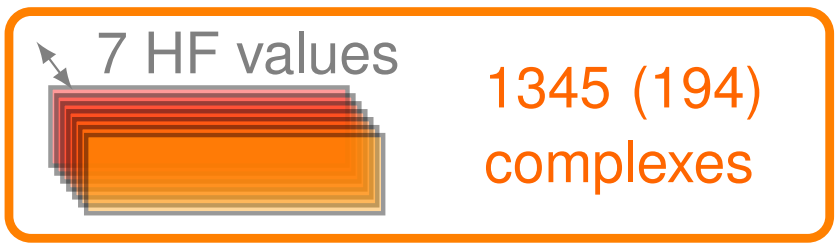


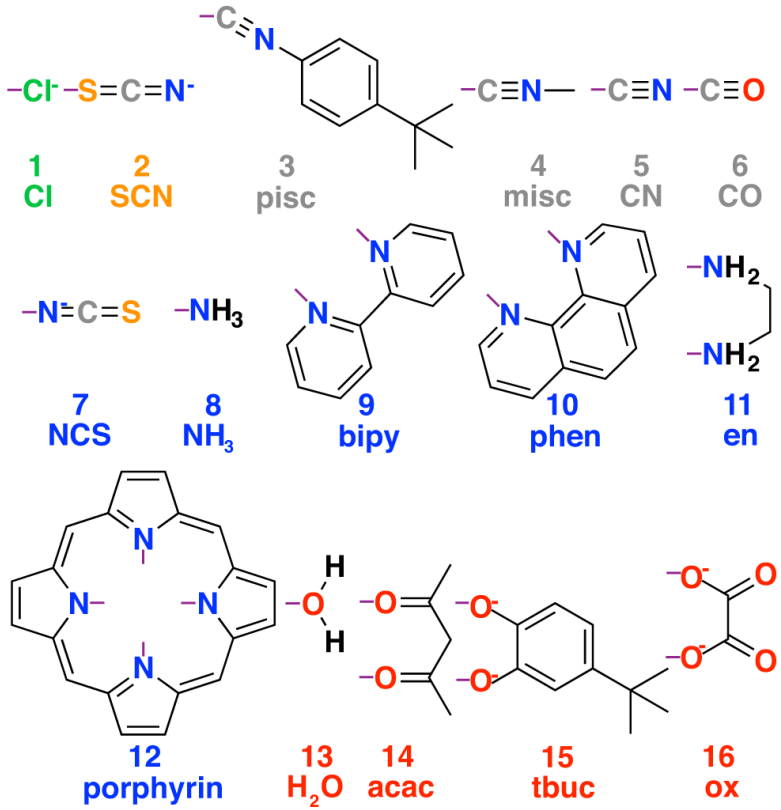
B3LYP-like DFT
HF exchange in 0-30%
gas phase optimizaton
LANL2DZ/6-31G*
high- and low-spin
M(II)/(III)

[1] Janet, J.P., and Kulik, H.J. *Chem. Sci.*, 2017, 8, 5137-5152.

Data for spin splitting

Data for octahedral complexes[1]:



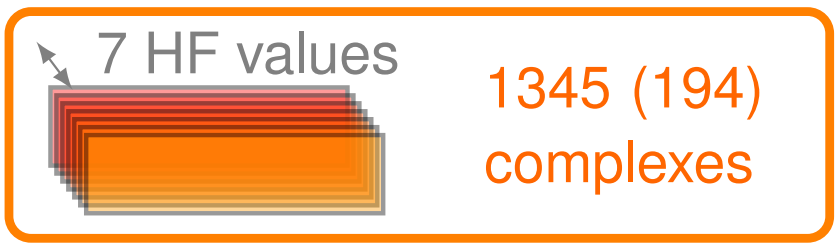


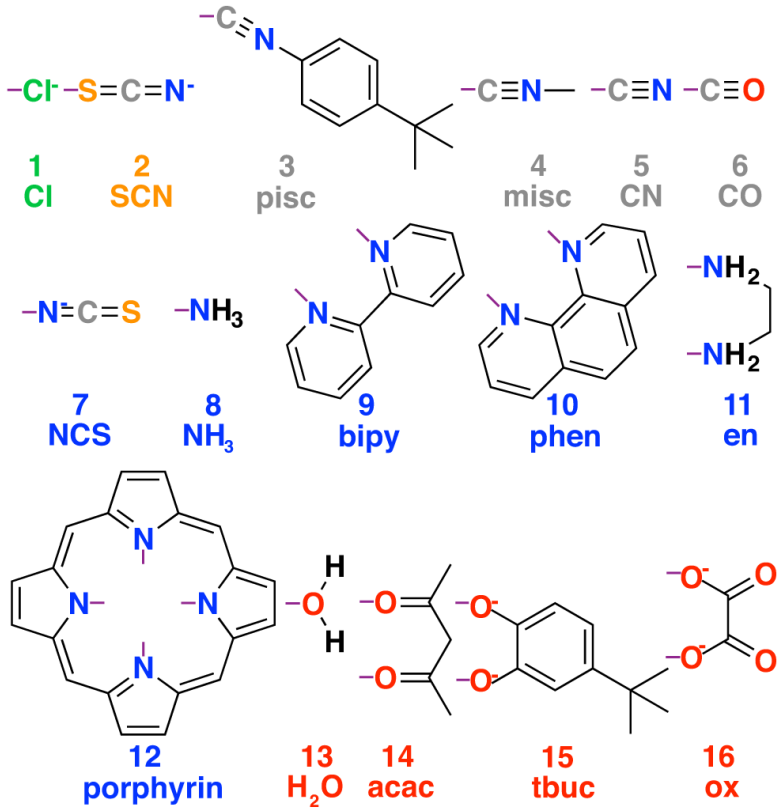
Coulomb matrix eigenspectrum (CM-ES) descriptor & kernel ridge regression (KRR)

[1] Janet, J.P., and Kulik, H.J. *Chem. Sci.*, 2017, 8, 5137-5152.

Data for spin splitting

Data for octahedral complexes[1]:





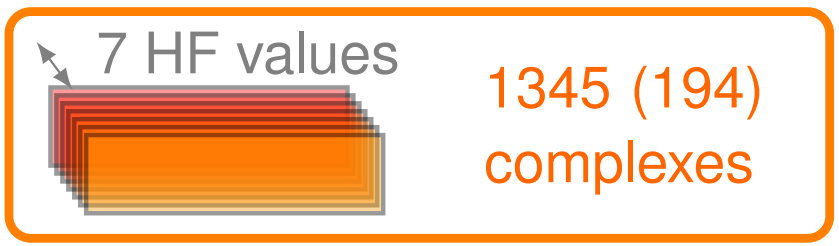
Coulomb matrix eigenspectrum (CM-ES) descriptor & kernel ridge regression (KRR)

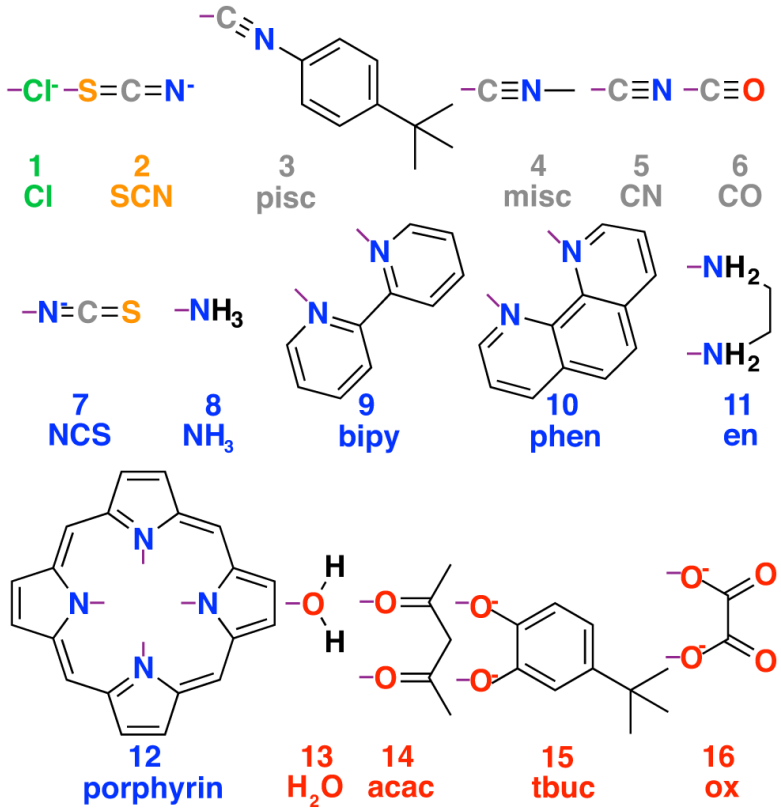
	$\Delta E_{H\text{-}L}$ RMSE
CM-ES	19.2 kcal/mol

[1] Janet, J.P., and Kulik, H.J. *Chem. Sci.*, 2017, 8, 5137-5152.

Data for spin splitting

Data for octahedral complexes[1]:





Coulomb matrix eigenspectrum (CM-ES) descriptor & kernel ridge regression (KRR)

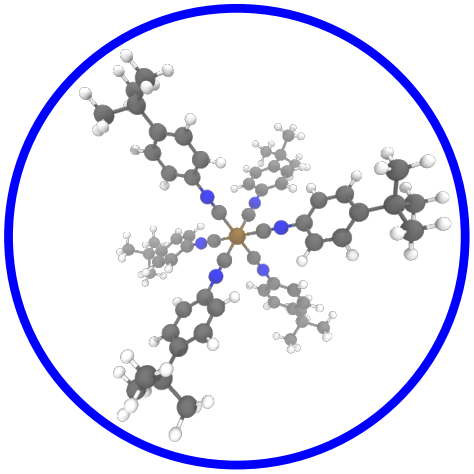
	$\Delta E_{H\text{-}L}$ RMSE
CM-ES	19.2 kcal/mol

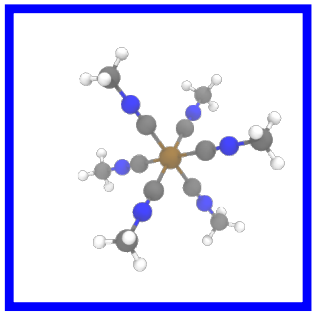
Why?

[1] Janet, J.P., and Kulik, H.J. *Chem. Sci.*, 2017, 8, 5137-5152.

A tale of two complexes

$Fe[pisc]_6^{3+}$

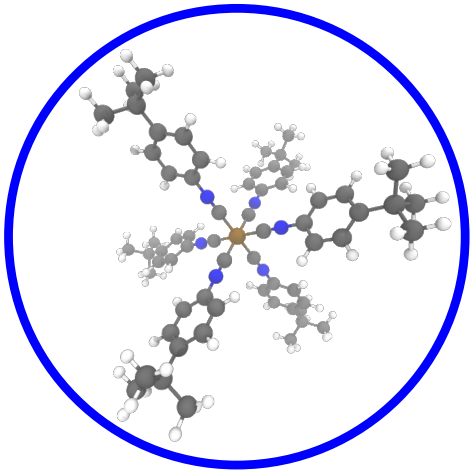


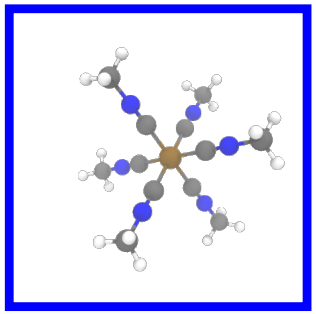


$Fe[misc]_6^{3+}$

A tale of two complexes

$Fe[pisc]_6^{3+}$



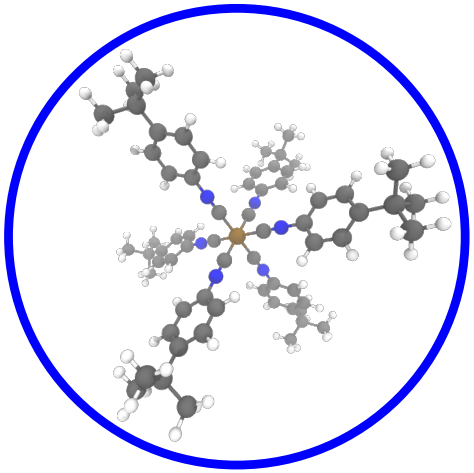


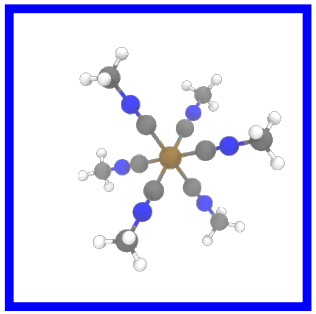
$Fe[misc]_6^{3+}$

$\Delta E_{H\text{-}L} = 37.7$ kcal/mol $\Delta E_{H\text{-}L} = 40.7$ kcal/mol

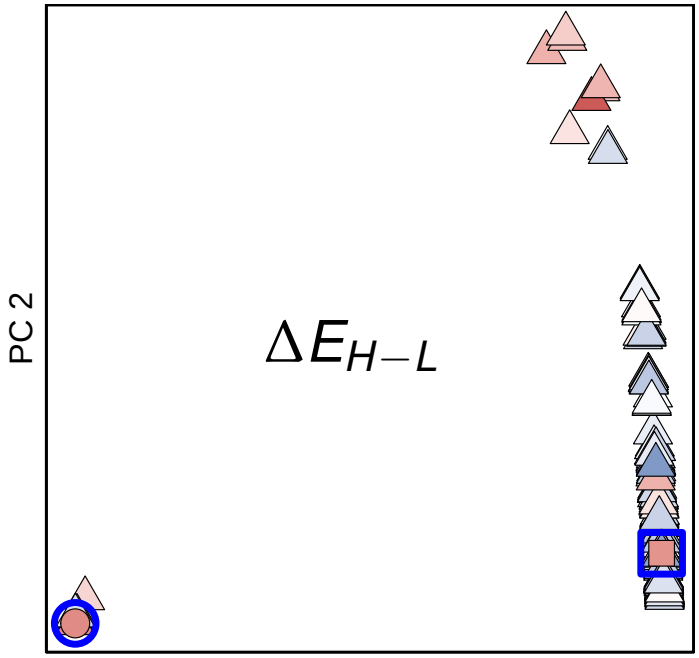
A tale of two complexes

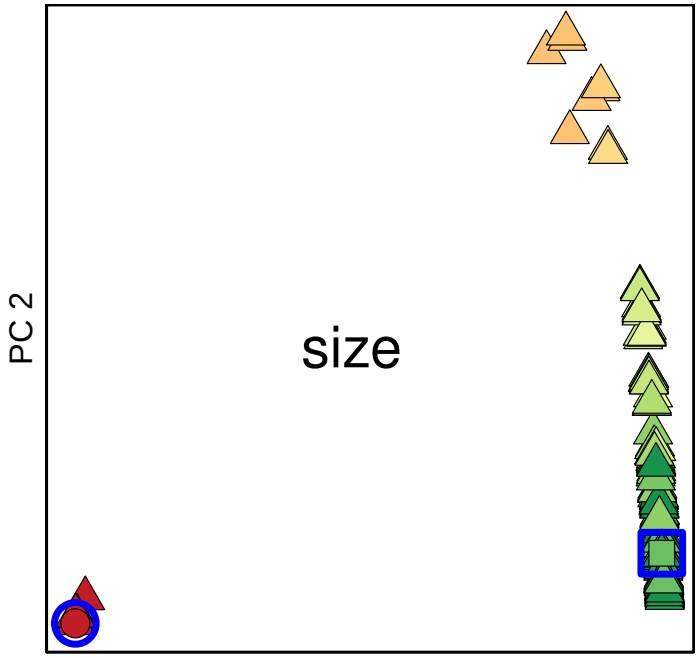
$Fe[pisc]_6^{3+}$



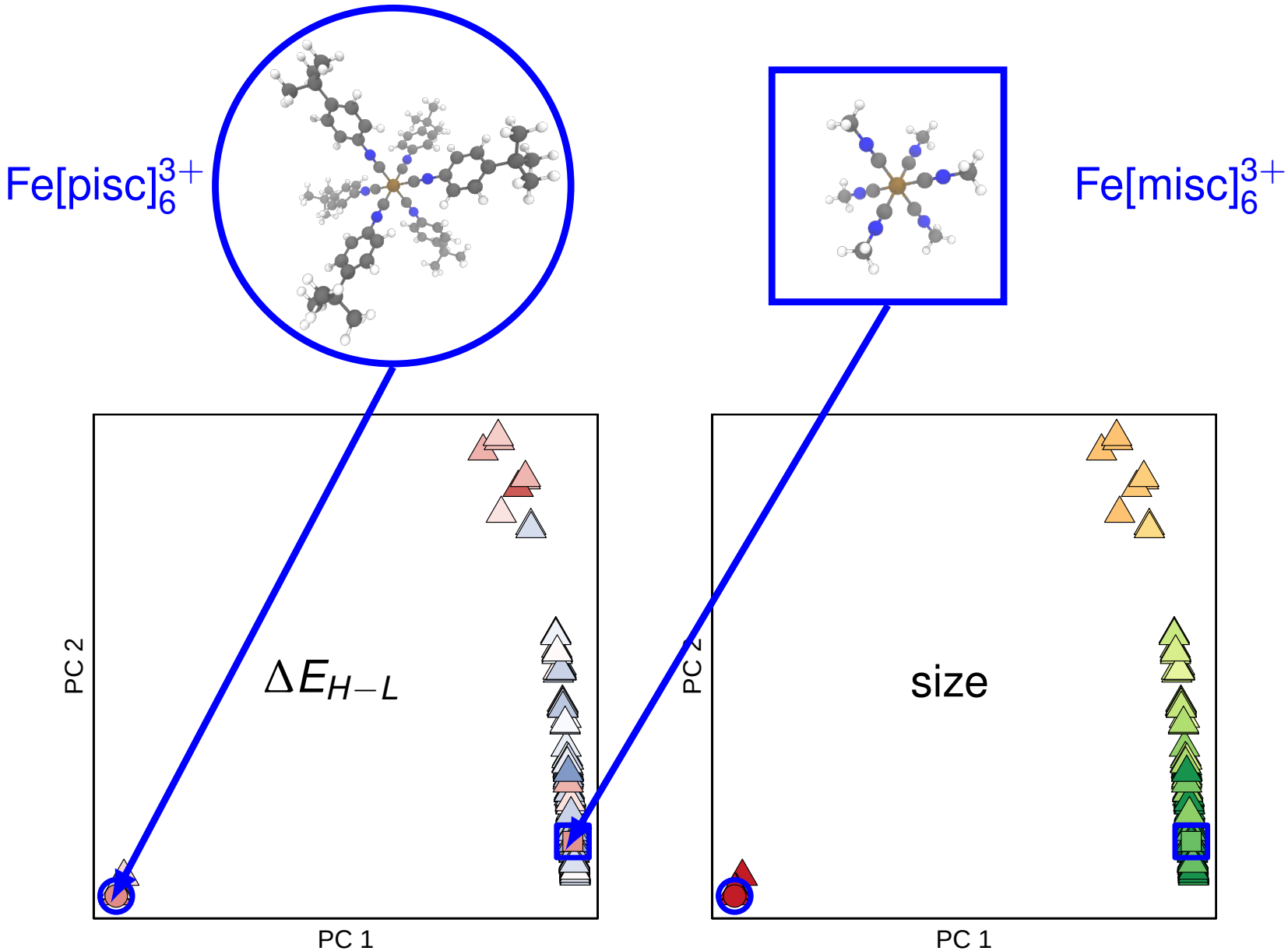


$Fe[misc]_6^{3+}$





A tale of two complexes



MCDL-25

mixed continuous discrete lcoal (MCDL)

MCDL-25

mixed continuous discrete lcoal (MCDL)

metal

properties

identity

Fe(II)

oxidation state

Janet, J.P., and Kulik, H.J. *Chem. Sci.*, 2017, 8, 5137-5152.

MCDL-25

mixed continuous discrete lcoal (MCDL)

metal properties

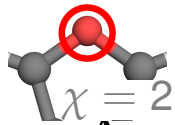
identity

Fe(II)

oxidation state

local ligand properties

$\chi = 3.44$



$\chi = 2.55$

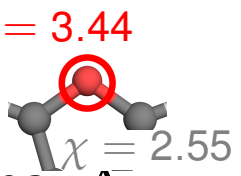
max $\Delta\chi$

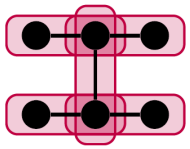
Janet, J.P., and Kulik, H.J. *Chem. Sci.*, 2017, 8, 5137-5152.

MCDL-25

mixed continuous discrete lcoal (MCDL)

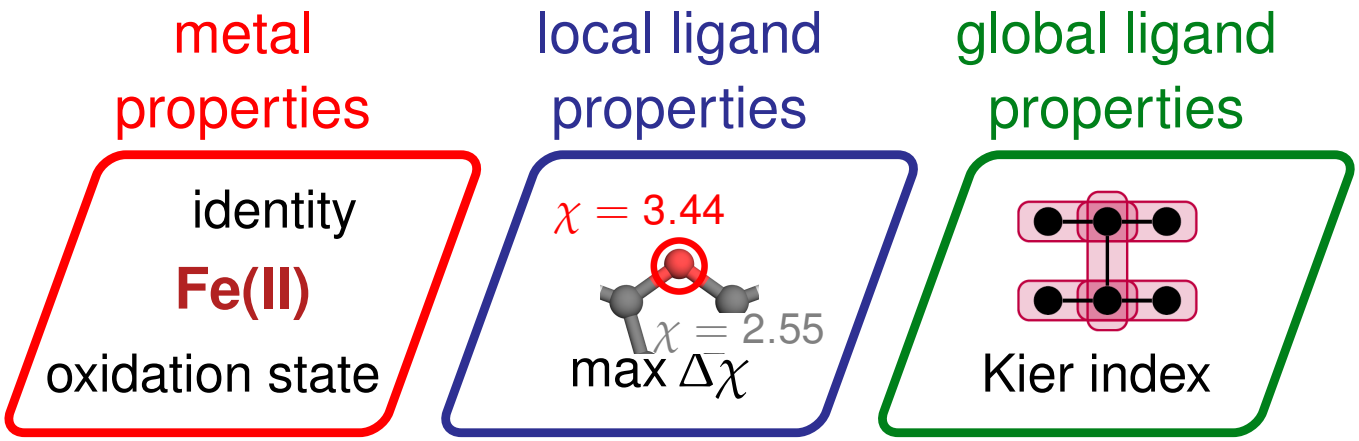
metal properties	local ligand properties	global ligand properties
identity	$\chi = 3.44$	
Fe(II)	$\chi = 2.55$	
oxidation state	max $\Delta\chi$	Kier index

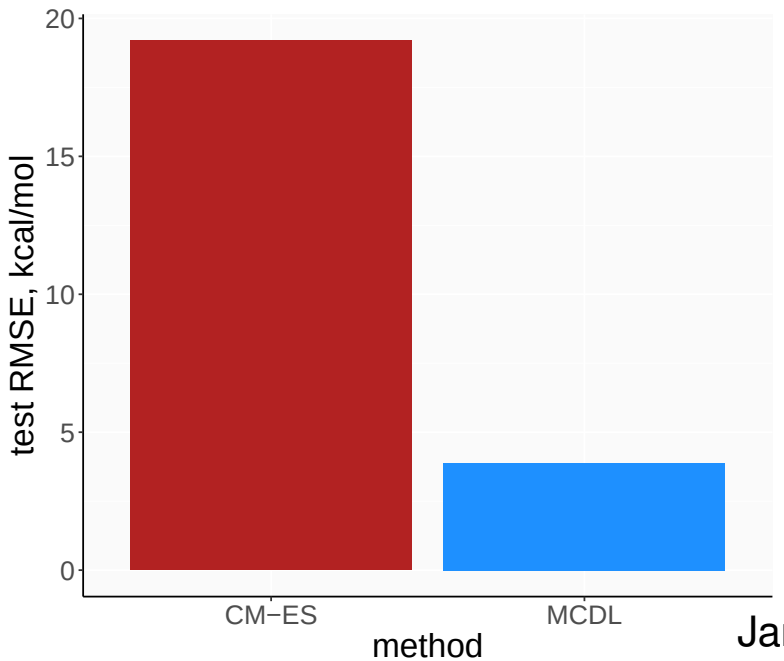




Janet, J.P., and Kulik, H.J. *Chem. Sci.*, 2017, 8, 5137-5152.

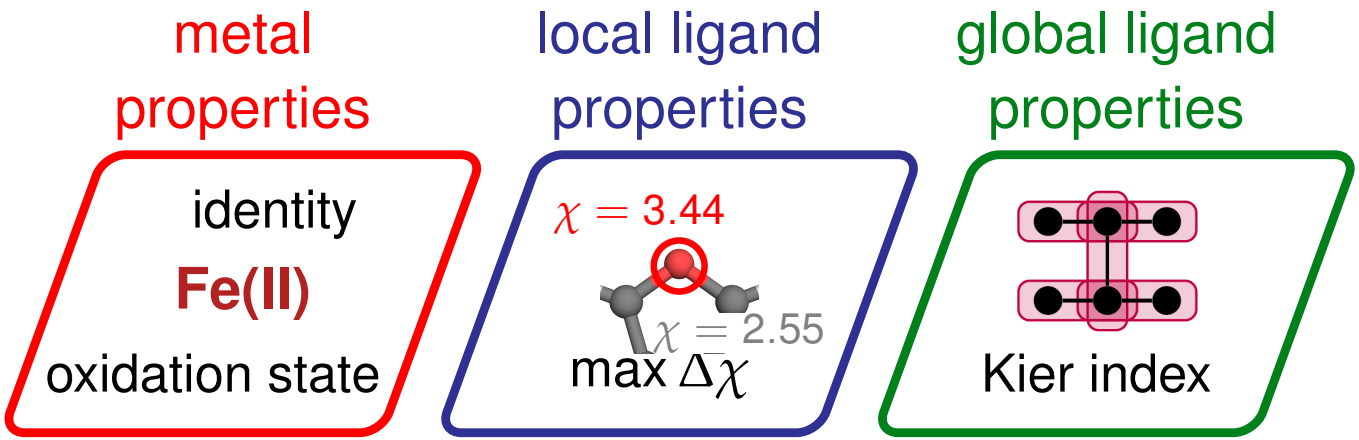
MCDL-25

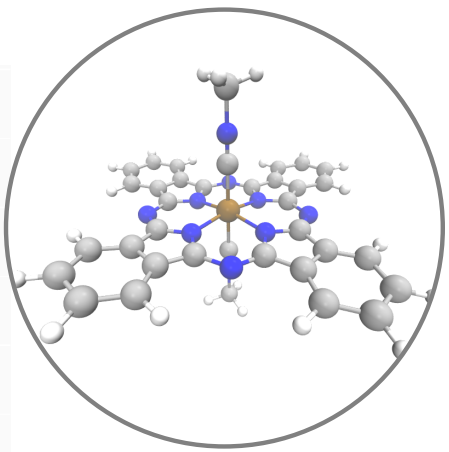


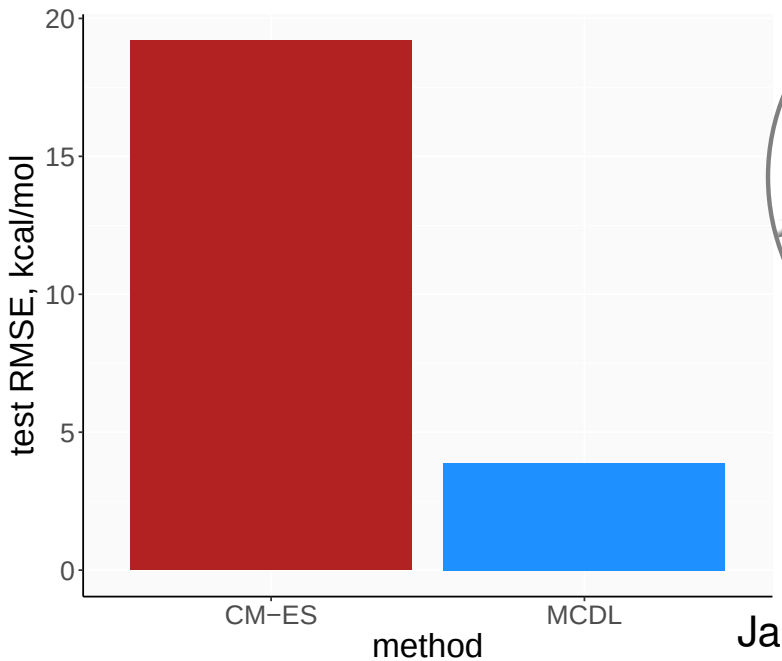


Janet, J.P., and Kulik, H.J. *Chem. Sci.*, 2017, 8, 5137-5152.

MCDL-25







Janet, J.P., and Kulik, H.J. *Chem. Sci.*, 2017, 8, 5137-5152.

MCDL-25

metal properties

identity

Fe(II)

oxidation state

local ligand properties

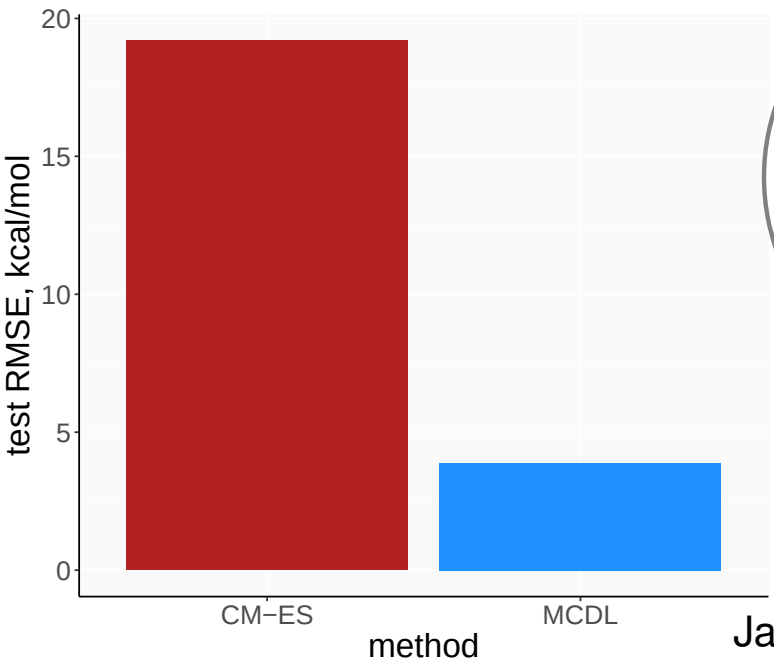
$\chi = 3.44$

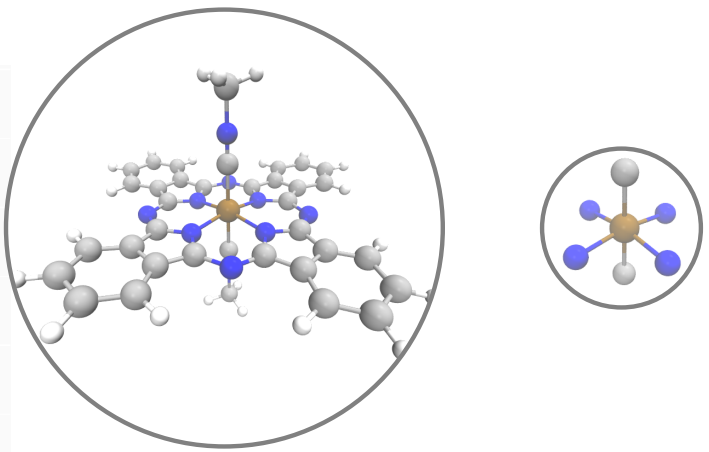
$\chi = 2.55$

max $\Delta\chi$

global ligand properties

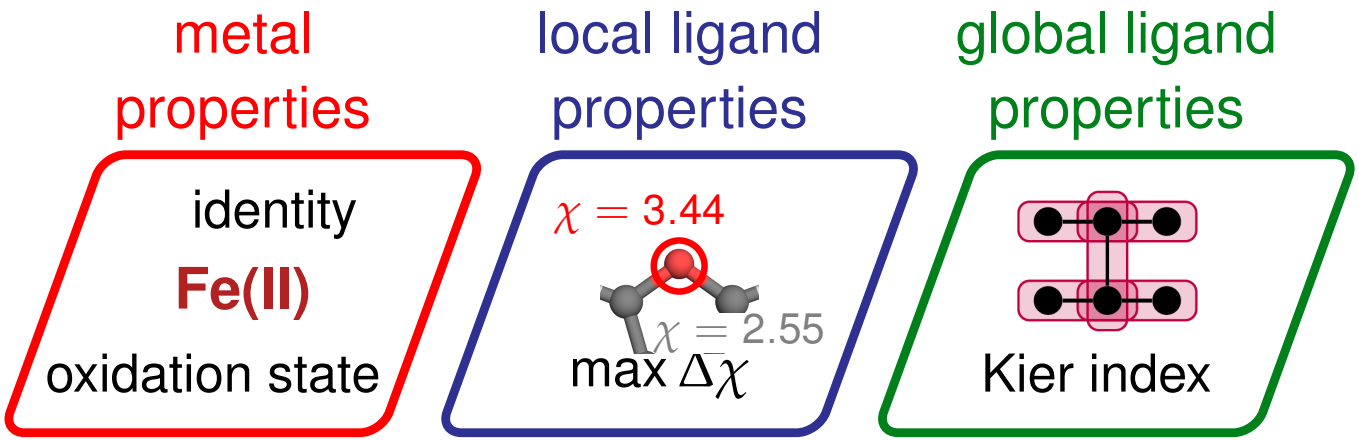
Kier index

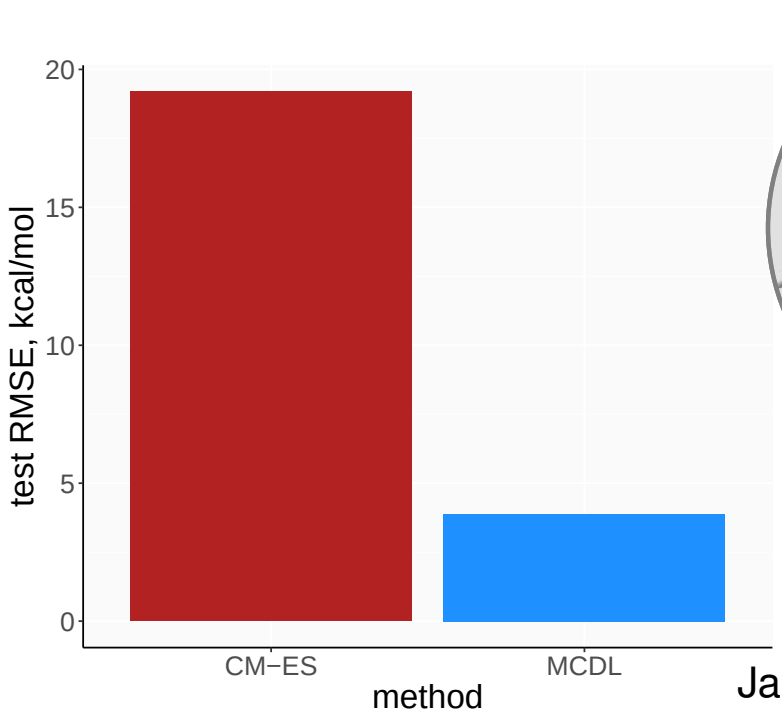


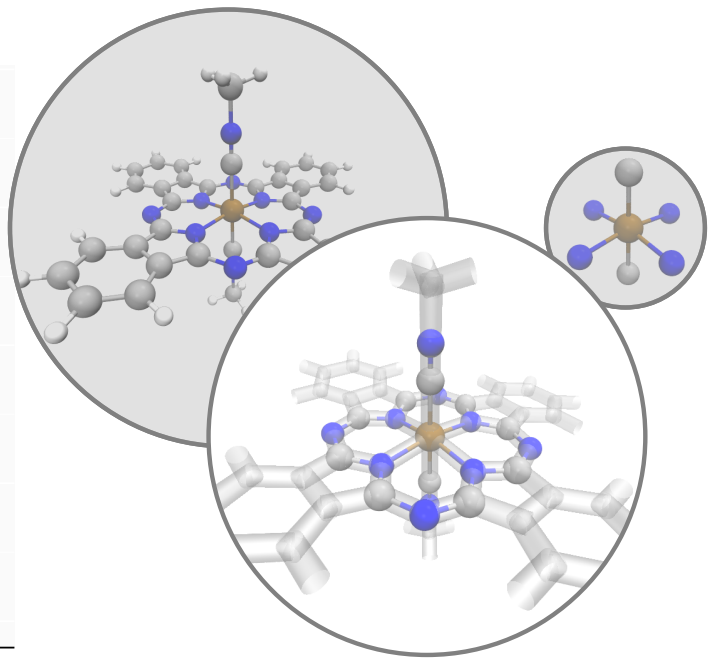


Janet, J.P., and Kulik, H.J. *Chem. Sci.*, 2017, 8, 5137-5152.

MCDL-25



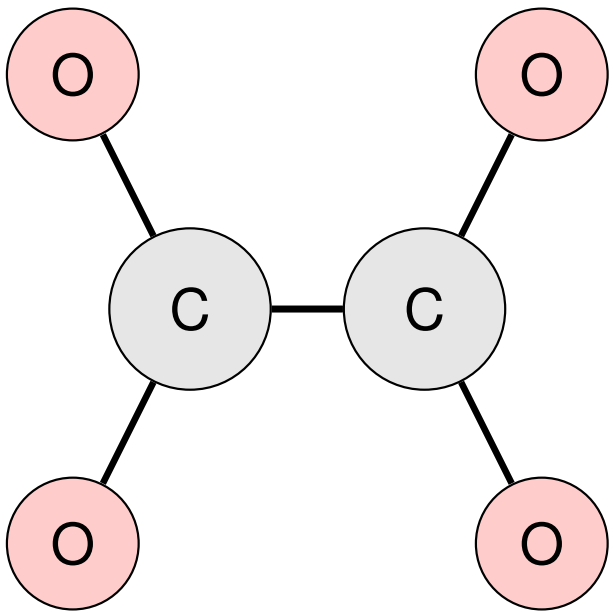




Janet, J.P., and Kulik, H.J. *Chem. Sci.*, 2017, 8, 5137-5152.

Extensible, continuous descriptors - RACs

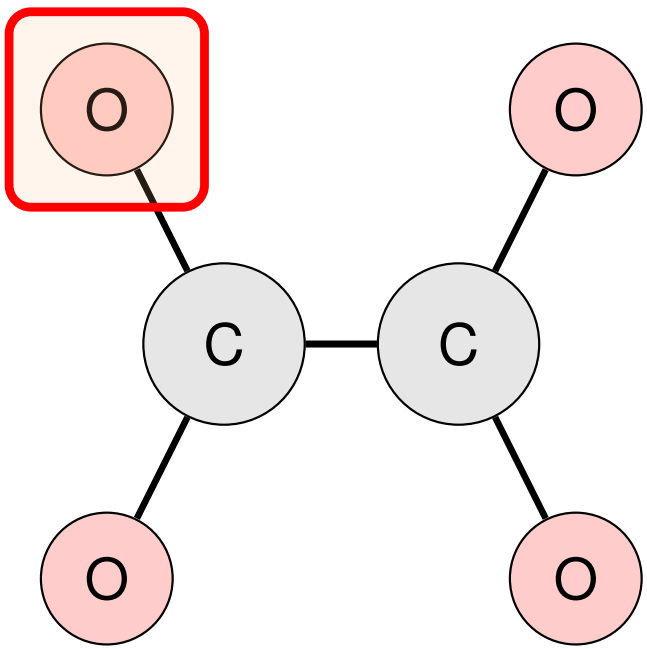
Based on autocorrelations[2]



[2]Broto, P., Moreau, G. and Vandycke, C. *Eur. J. Med. Chem.*, 19(1):71-78, 1984.

Extensible, continuous descriptors - RACs

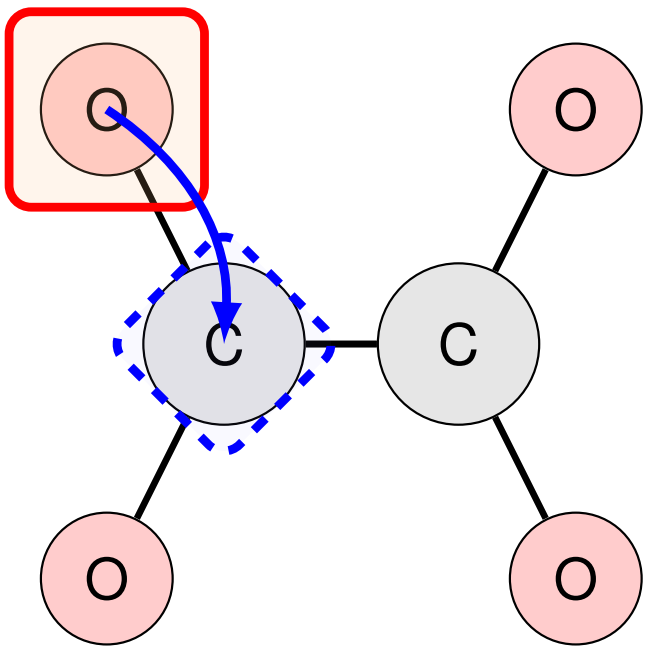
Based on autocorrelations[2]



[2]Broto, P., Moreau, G. and Vandycke, C. *Eur. J. Med. Chem.*, 19(1):71-78, 1984.

Extensible, continuous descriptors - RACs

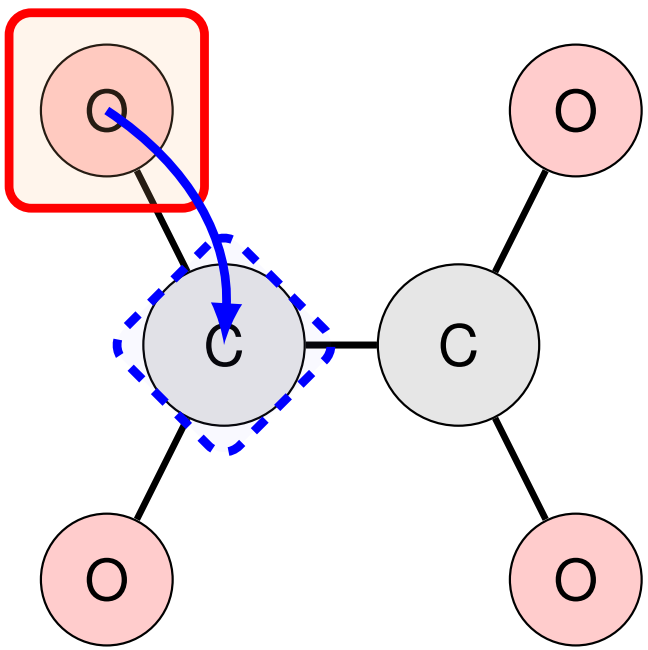
Based on autocorrelations[2]



[2]Broto, P., Moreau, G. and Vandycke, C. *Eur. J. Med. Chem.*, 19(1):71-78, 1984.

Extensible, continuous descriptors - RACs

Based on autocorrelations[2]

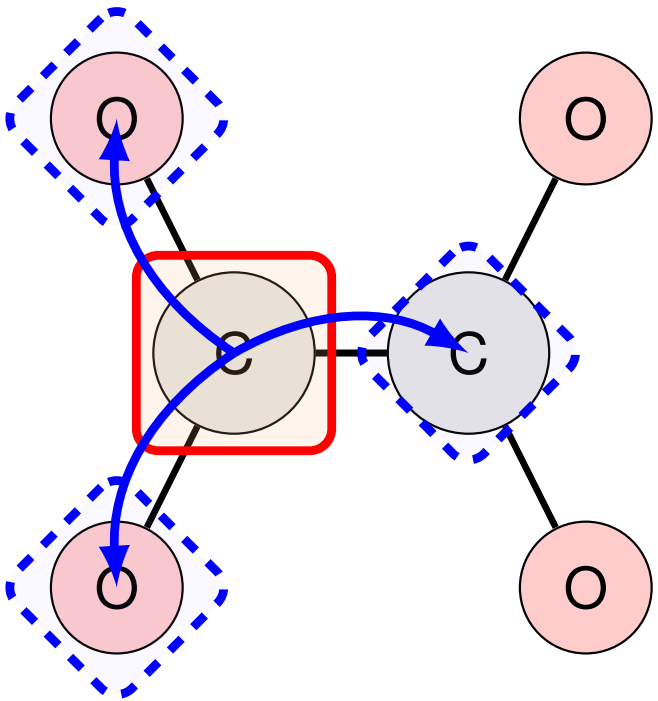


$$d_1 : \sum_{O,C} Z_O Z_C = 48$$

[2]Broto, P., Moreau, G. and Vandycke, C. *Eur. J. Med. Chem.*, 19(1):71-78, 1984.

Extensible, continuous descriptors - RACs

Based on autocorrelations[2]

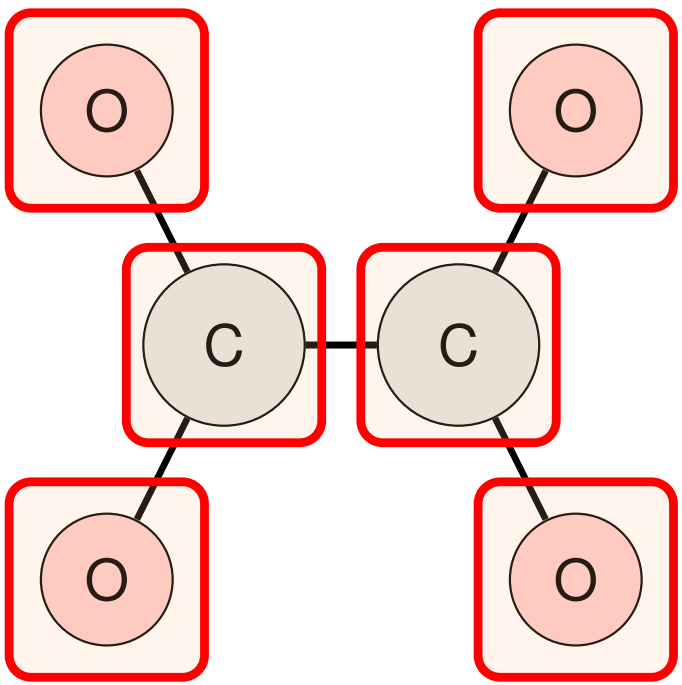


$$d_1 : 48 + \sum_{C,O} Z_O Z_C = 144 + 48$$

[2]Broto, P., Moreau, G. and Vandycke, C. *Eur. J. Med. Chem.*, 19(1):71-78, 1984.

Extensible, continuous descriptors - RACs

Based on autocorrelations[2]

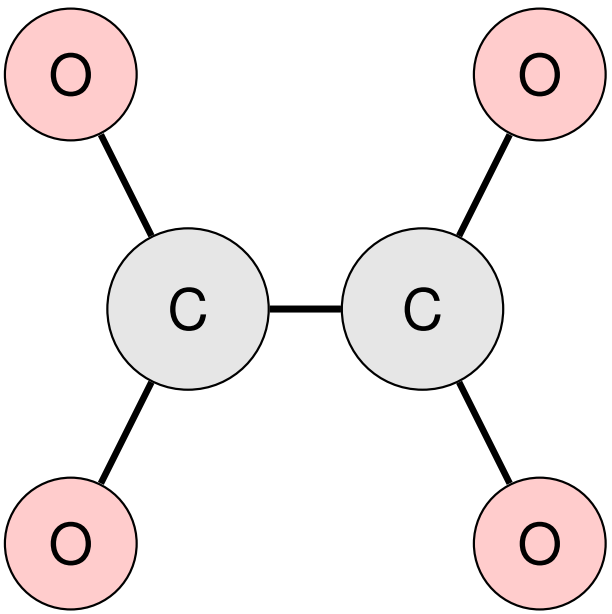


$$d_1 : \sum_i \sum_j Z_i Z_j \delta(d_{i,j}, 1)$$

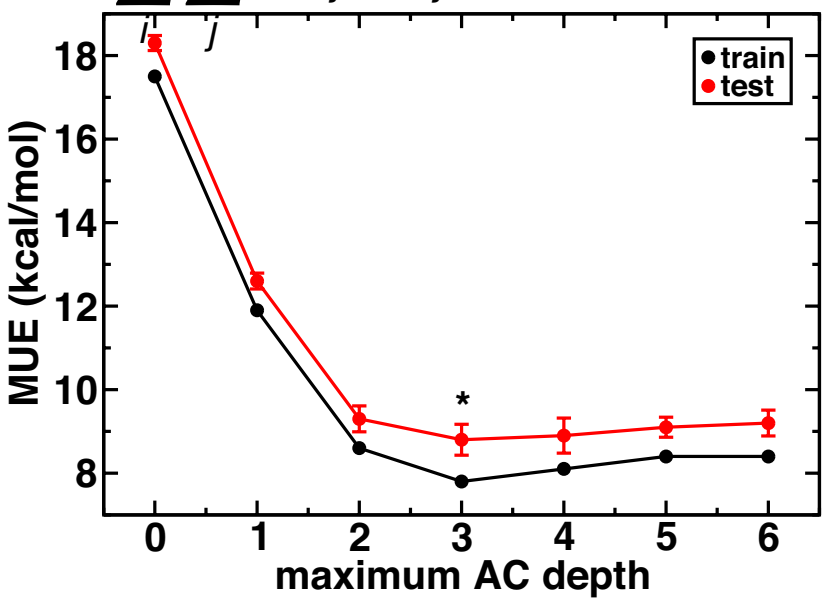
[2]Broto, P., Moreau, G. and Vandycke, C. *Eur. J. Med. Chem.*, 19(1):71-78, 1984.

Extensible, continuous descriptors - RACs

Based on autocorrelations[2]



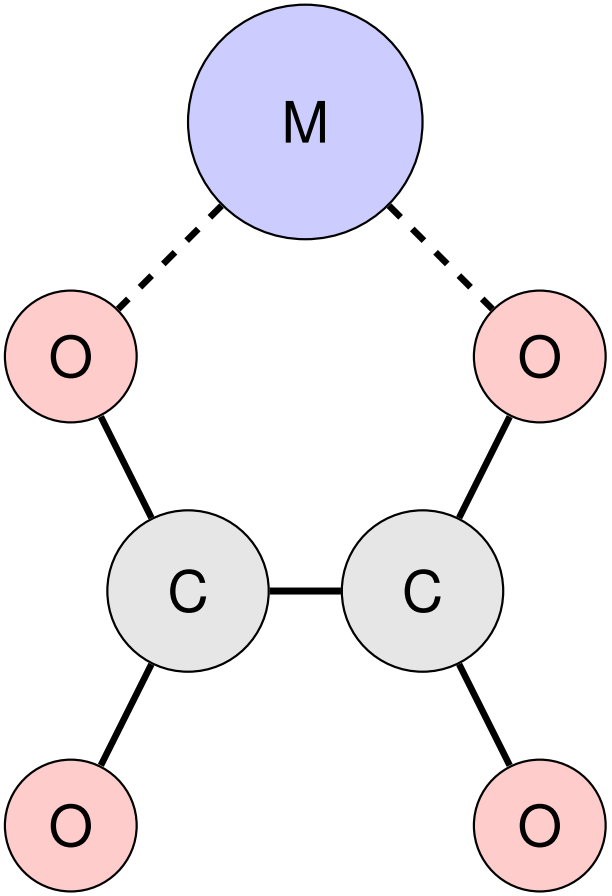
$$d_x : \sum_i \sum_j Z_i Z_j \delta(d_{ij}, x)$$



[2]Broto, P., Moreau, G. and Vandycke, C. *Eur. J. Med. Chem.*, 19(1):71-78, 1984.

Extensible, continuous descriptors - RACs

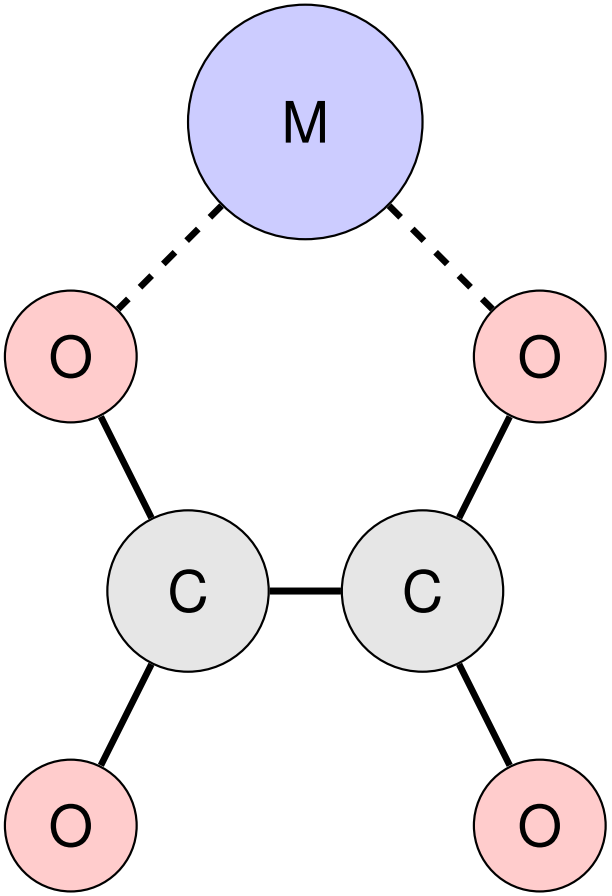
Based on autocorrelations



How to adapt to TM complexes?

Extensible, continuous descriptors - RACs

Based on autocorrelations

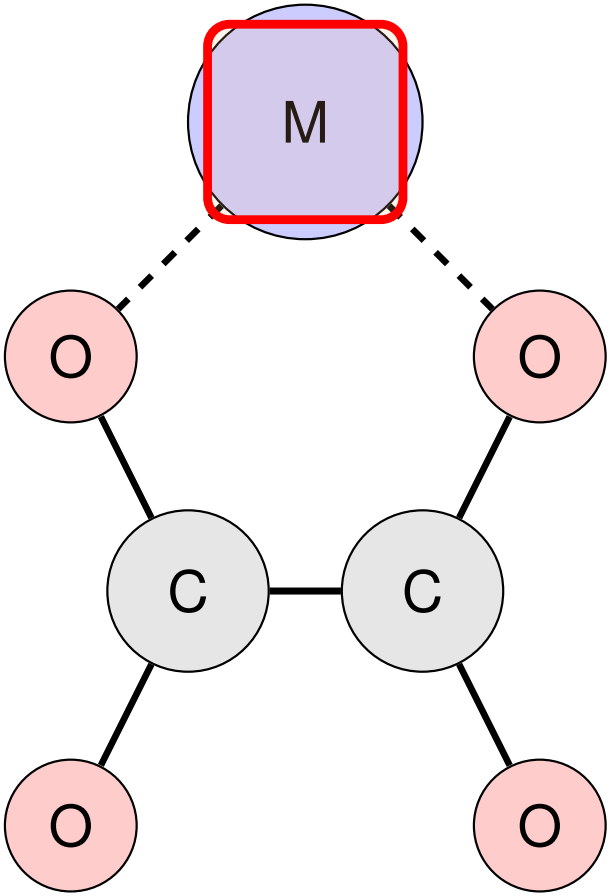


How to adapt to TM complexes?

restrict the scope to focus on *near-metal atoms*

Extensible, continuous descriptors - RACs

Based on autocorrelations

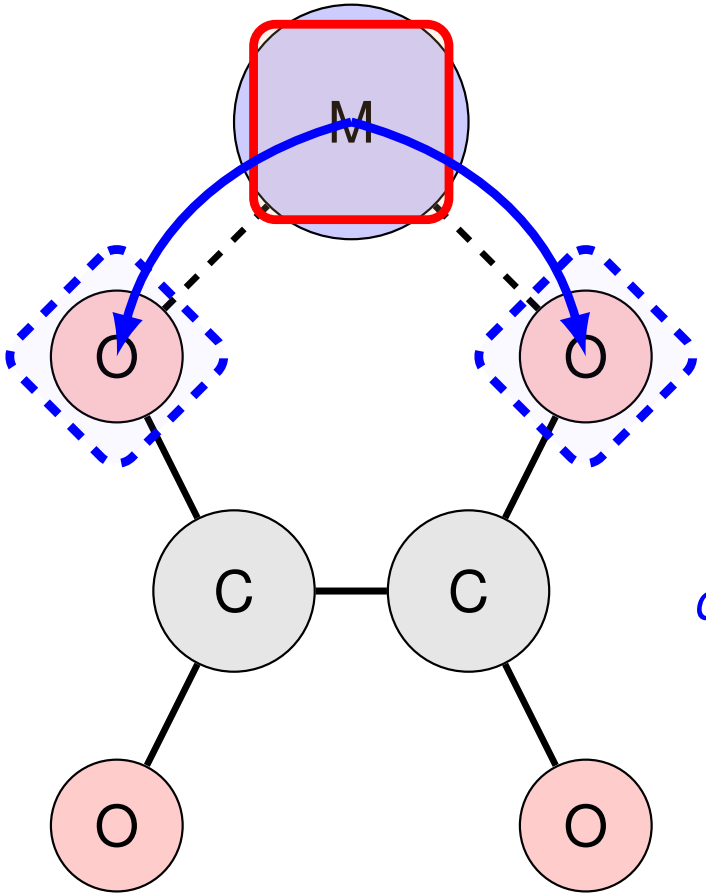


How to adapt to TM complexes?

restrict the scope to focus on *near-metal atoms*

Extensible, continuous descriptors - RACs

Based on autocorrelations



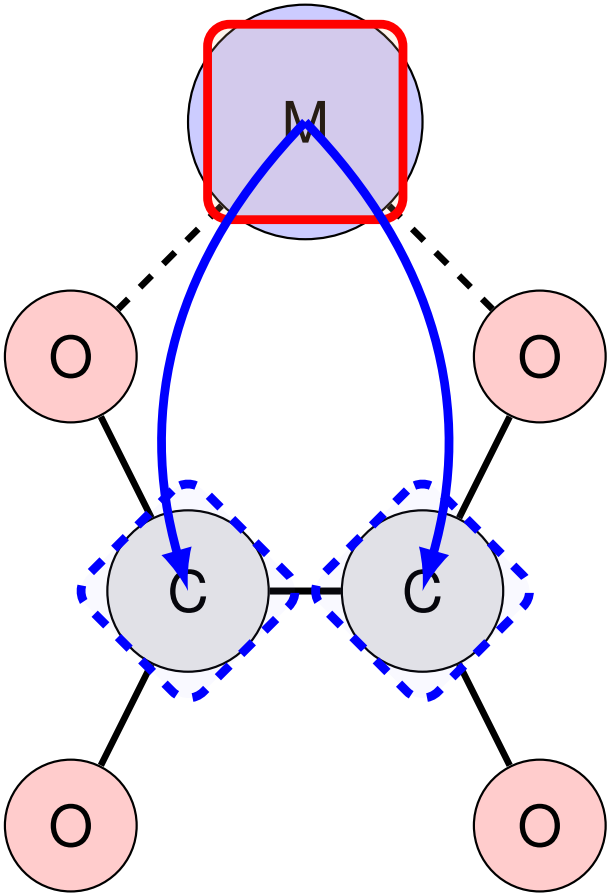
How to adapt to TM complexes?

restrict the scope to focus on *near-metal atoms*

$$d_1 : \sum_{M,O} Z_M Z_O$$

Extensible, continuous descriptors - RACs

Based on autocorrelations



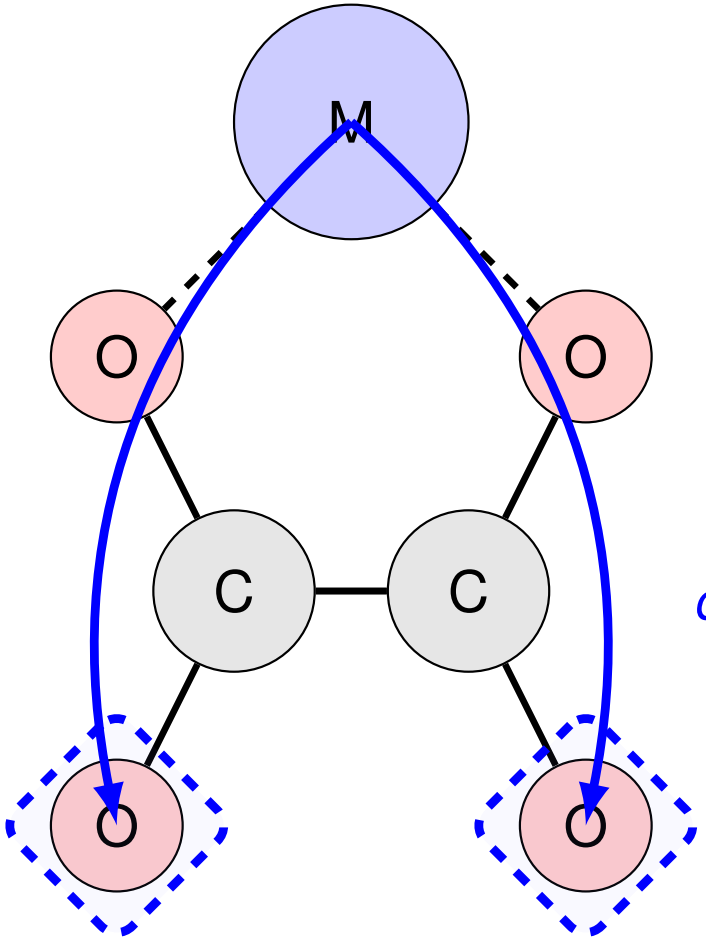
How to adapt to TM complexes?

restrict the scope to focus on *near-metal atoms*

$$d_2 : \sum_{M,C} Z_M Z_C$$

Extensible, continuous descriptors - RACs

Based on autocorrelations



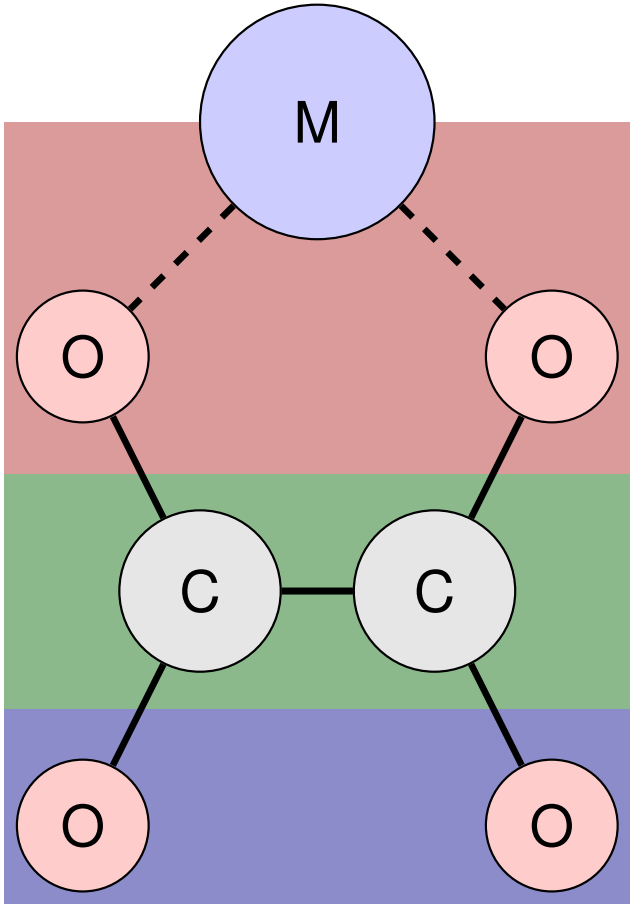
How to adapt to TM complexes?

restrict the scope to focus on *near-metal atoms*

$$d_3 : \sum_{M,O} Z_M Z_O$$

Extensible, continuous descriptors - RACs

Based on autocorrelations



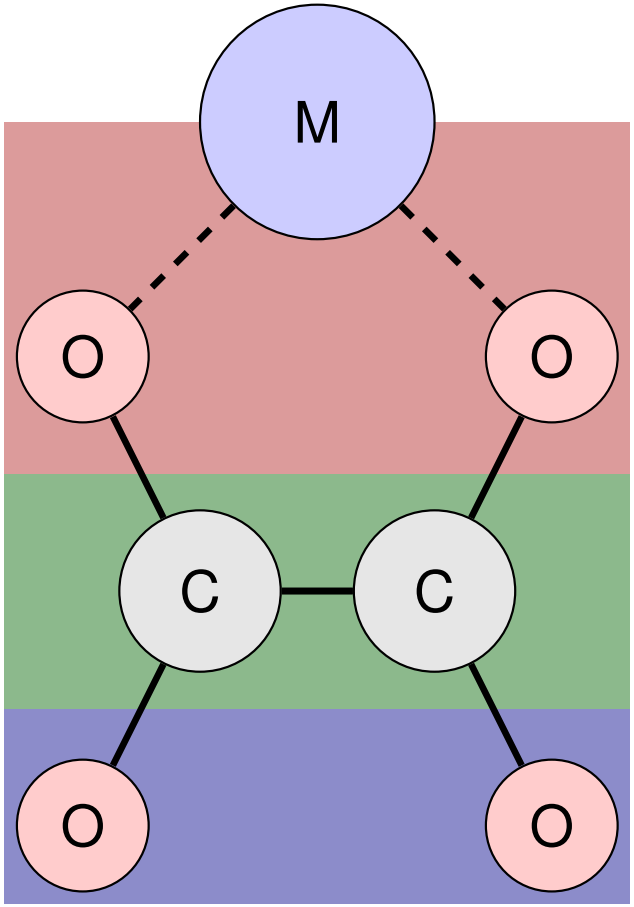
How to adapt to TM complexes?

restrict the scope to focus on *near-metal atoms*

$$d_3 : \sum_{M,O} Z_M Z_O$$

Extensible, continuous descriptors - RACs

Based on autocorrelations



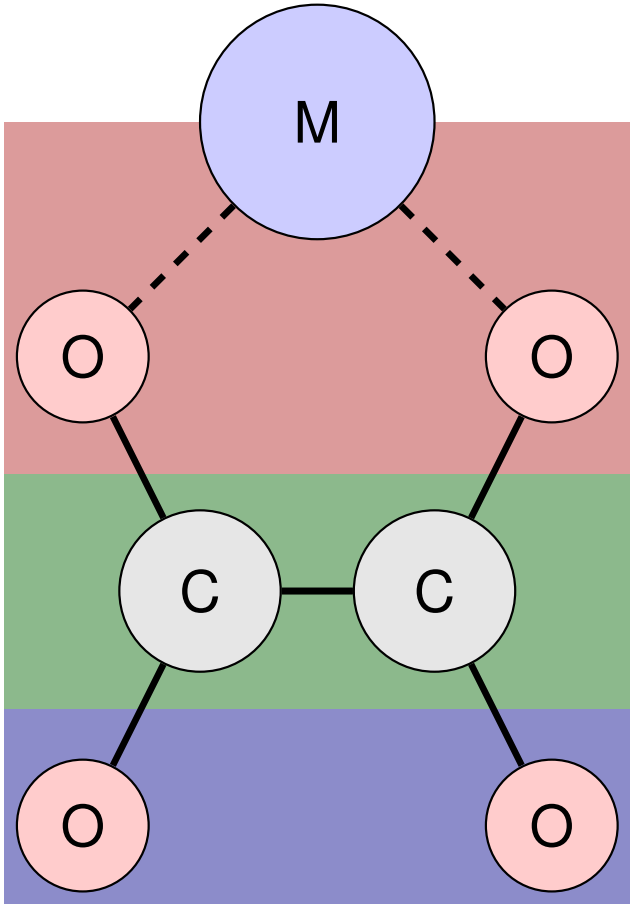
How to adapt to TM complexes?

restrict the scope to focus on *near-metal atoms*

$$d_3 : \sum_{M,O} Z_M Z_O \left(Z_i - Z_j\right)$$

Extensible, continuous descriptors - RACs

Based on autocorrelations



How to adapt to TM complexes?

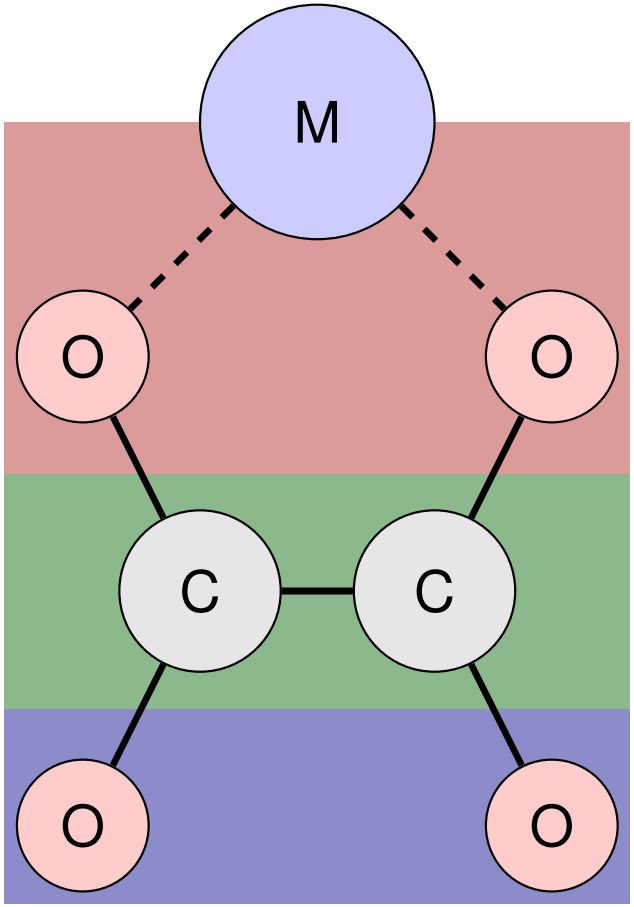
restrict the scope to focus on *near-metal atoms*

$$d_3 : \sum_{M,O} Z_M Z_O \left(Z_i - Z_j\right)$$

properties: T, χ, Z, I, S

Extensible, continuous descriptors - RACs

Based on autocorrelations



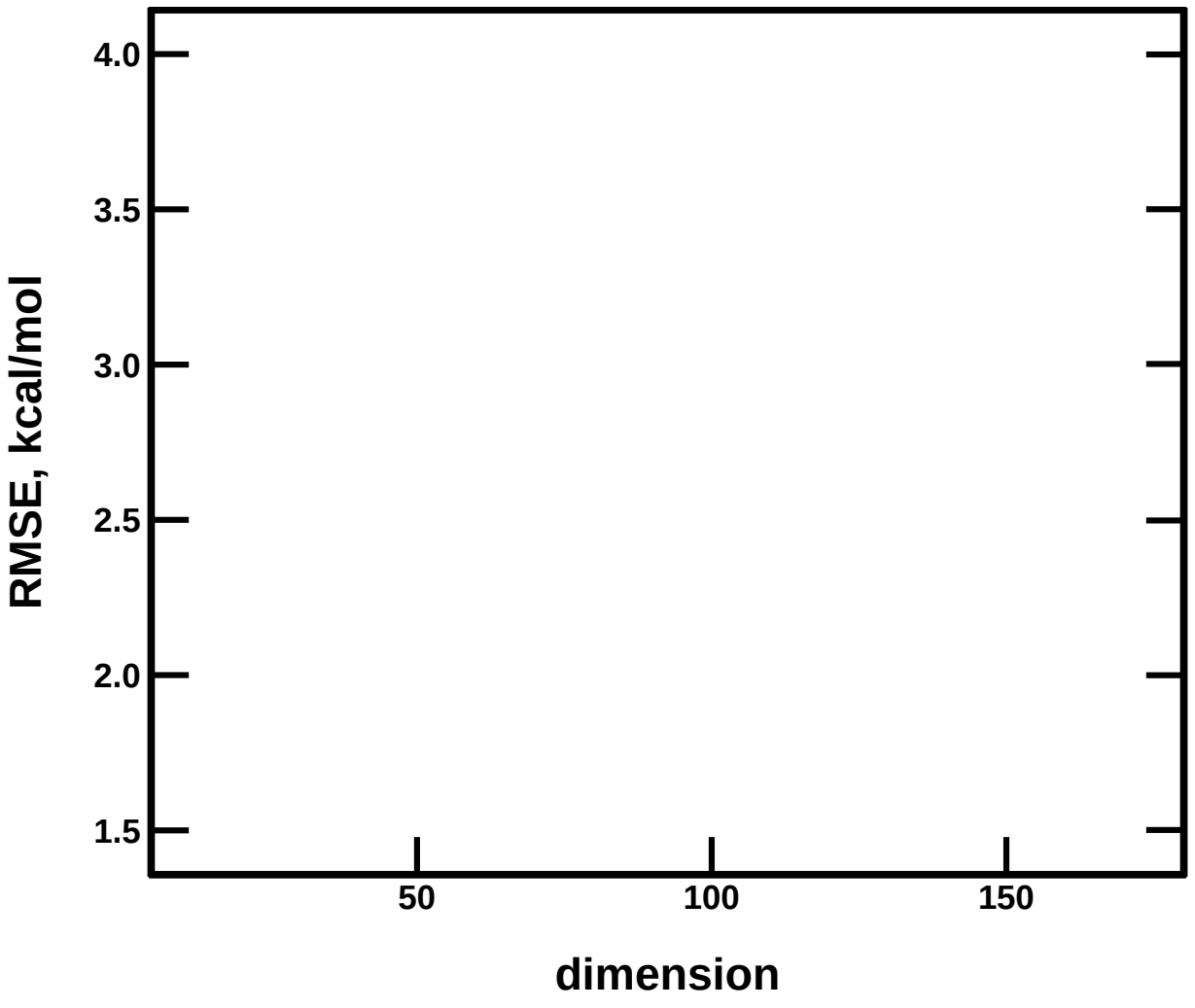
How to adapt to TM complexes?

restrict the scope to focus on *near-metal atoms*

$$d_3 : \sum_{M,O} Z_M Z_O \left(Z_i - Z_j\right)$$

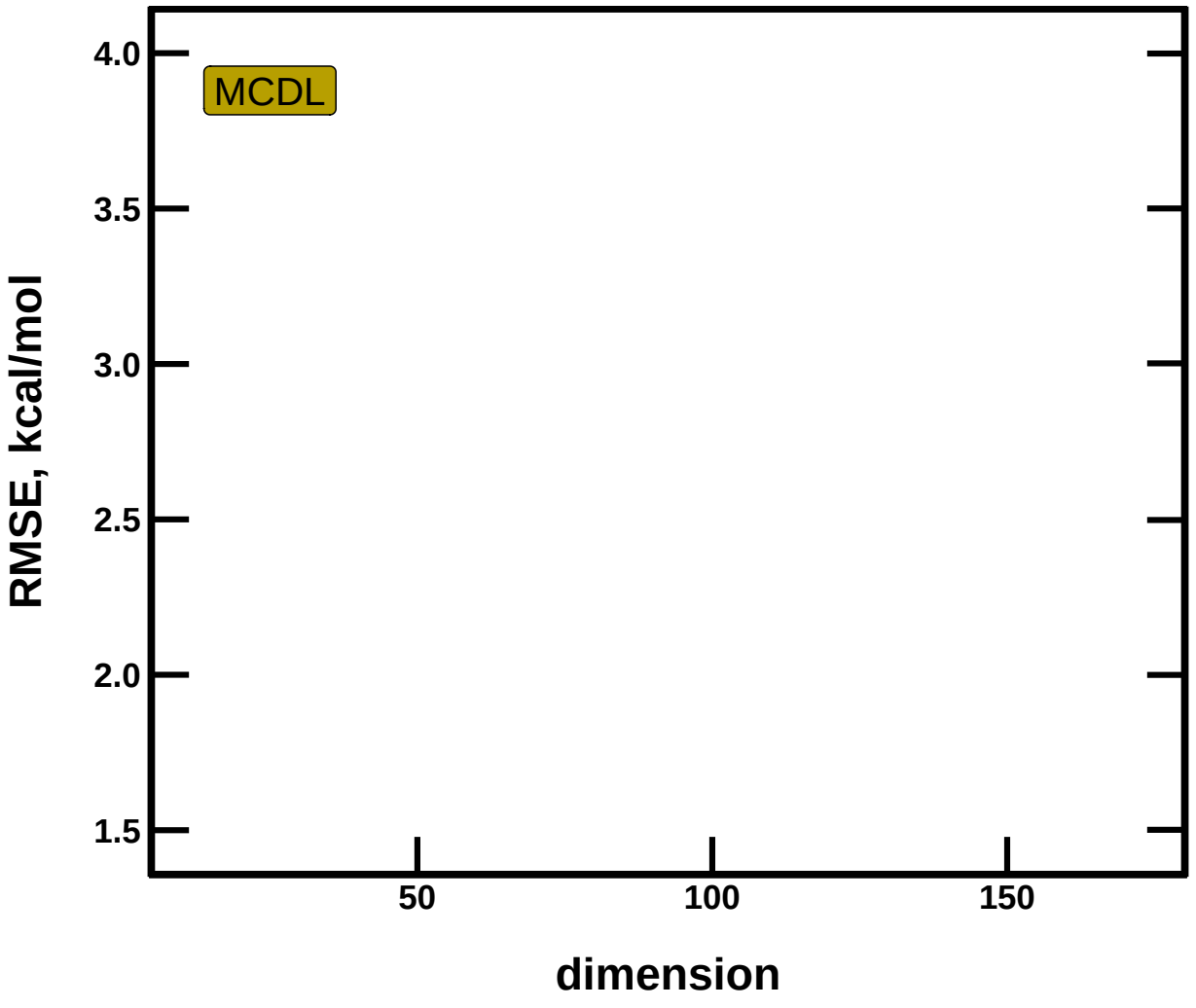
$\sim$ 160 features in total

Feature selection



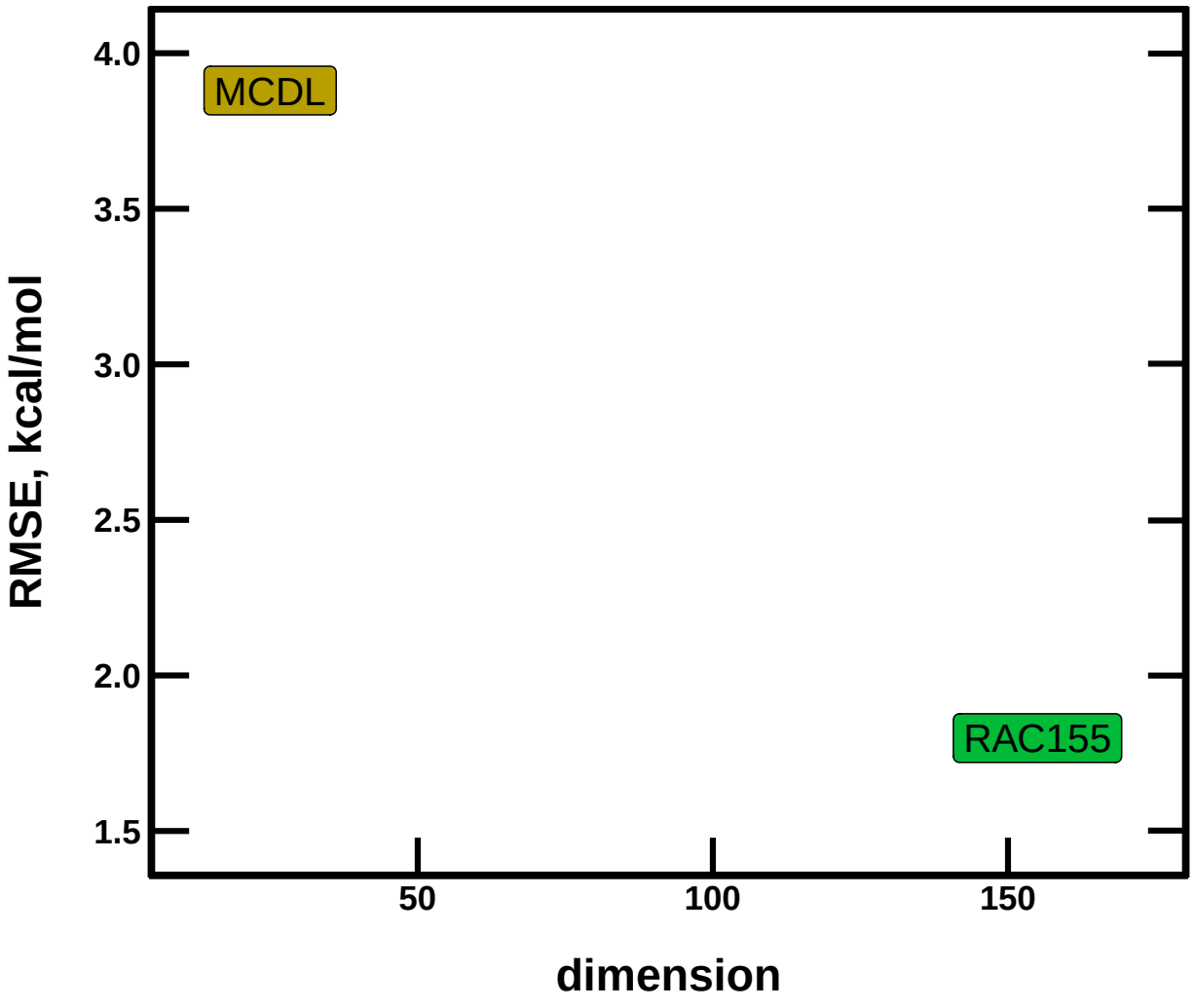
Janet, J.P., and Kulik, H.J. *J. Phys. Chem. A*, 2017,121, 46, 8939-8954.

Feature selection



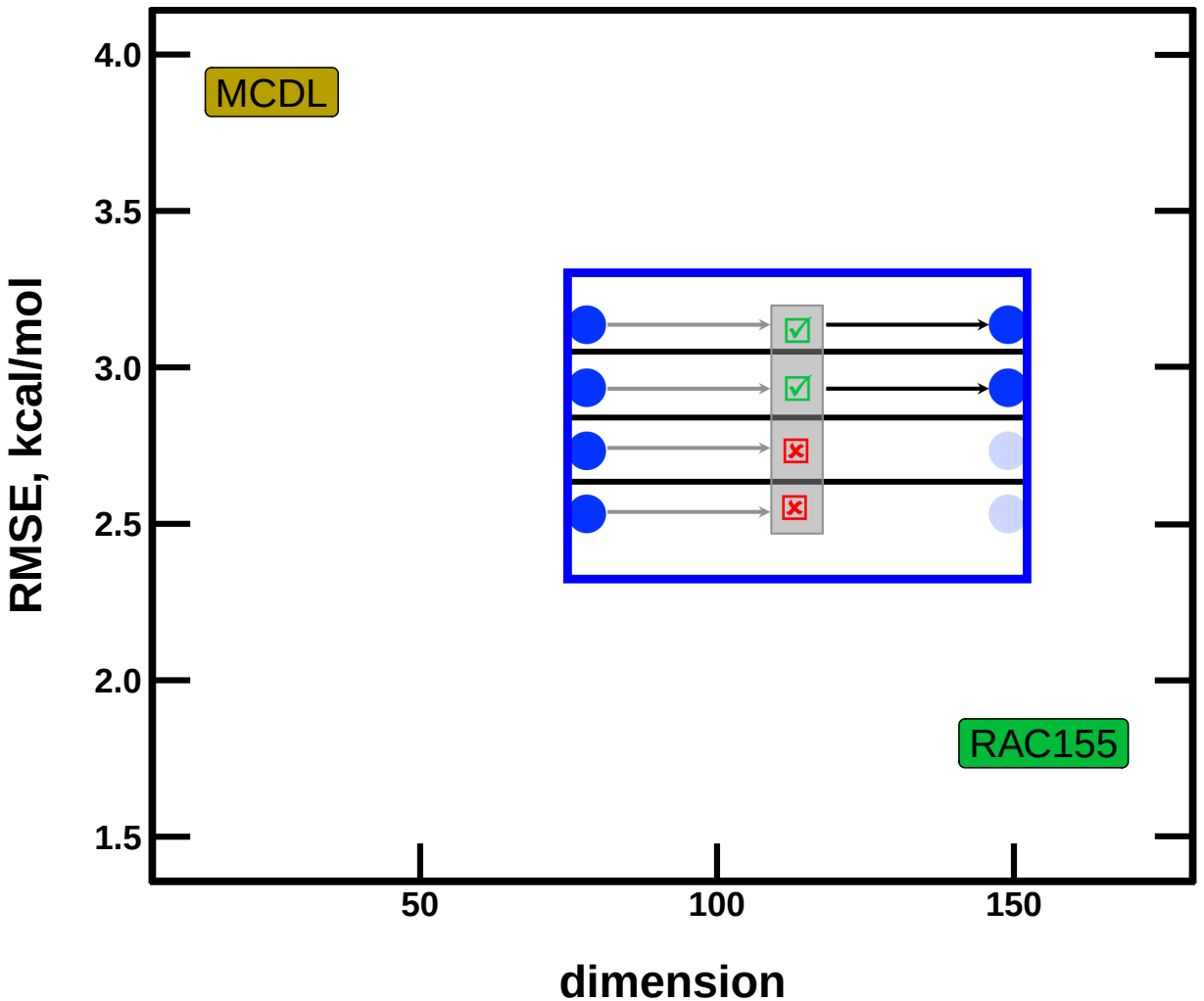
Janet, J.P., and Kulik, H.J. *J. Phys. Chem. A*, 2017,121, 46, 8939-8954.

Feature selection



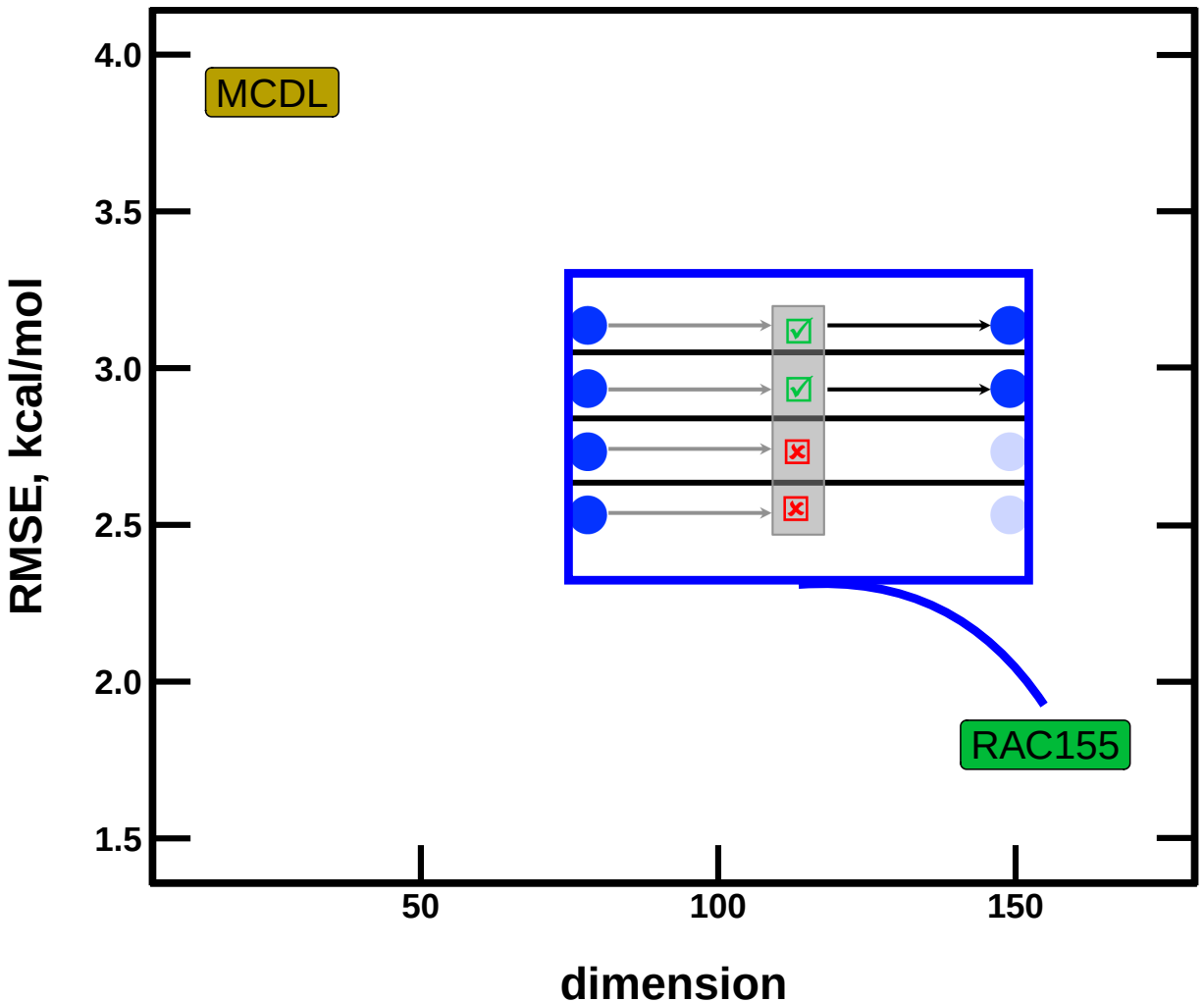
Janet, J.P., and Kulik, H.J. *J. Phys. Chem. A*, 2017,121, 46, 8939-8954.

Feature selection



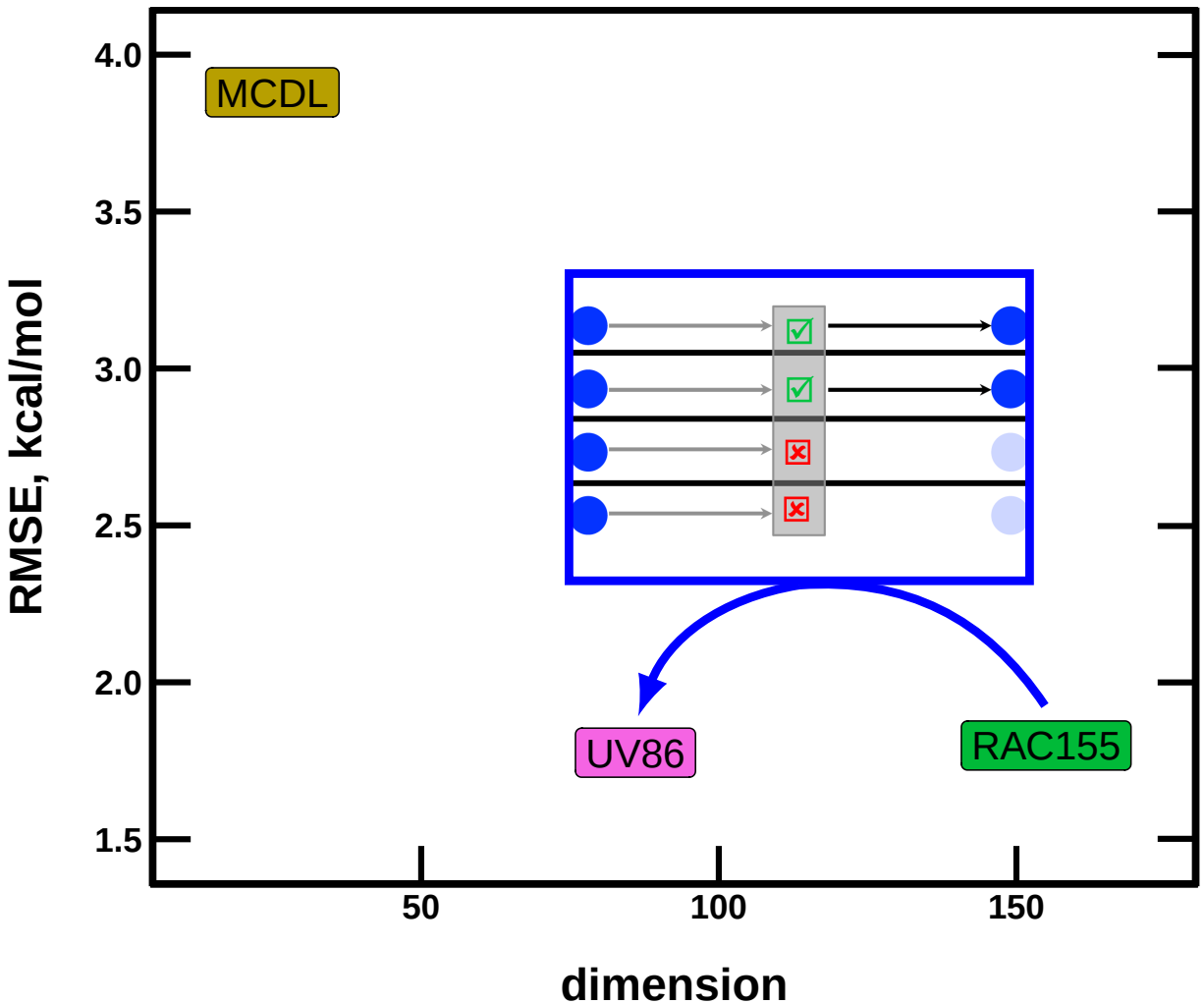
Janet, J.P., and Kulik, H.J. *J. Phys. Chem. A*, 2017,121, 46, 8939-8954.

Feature selection



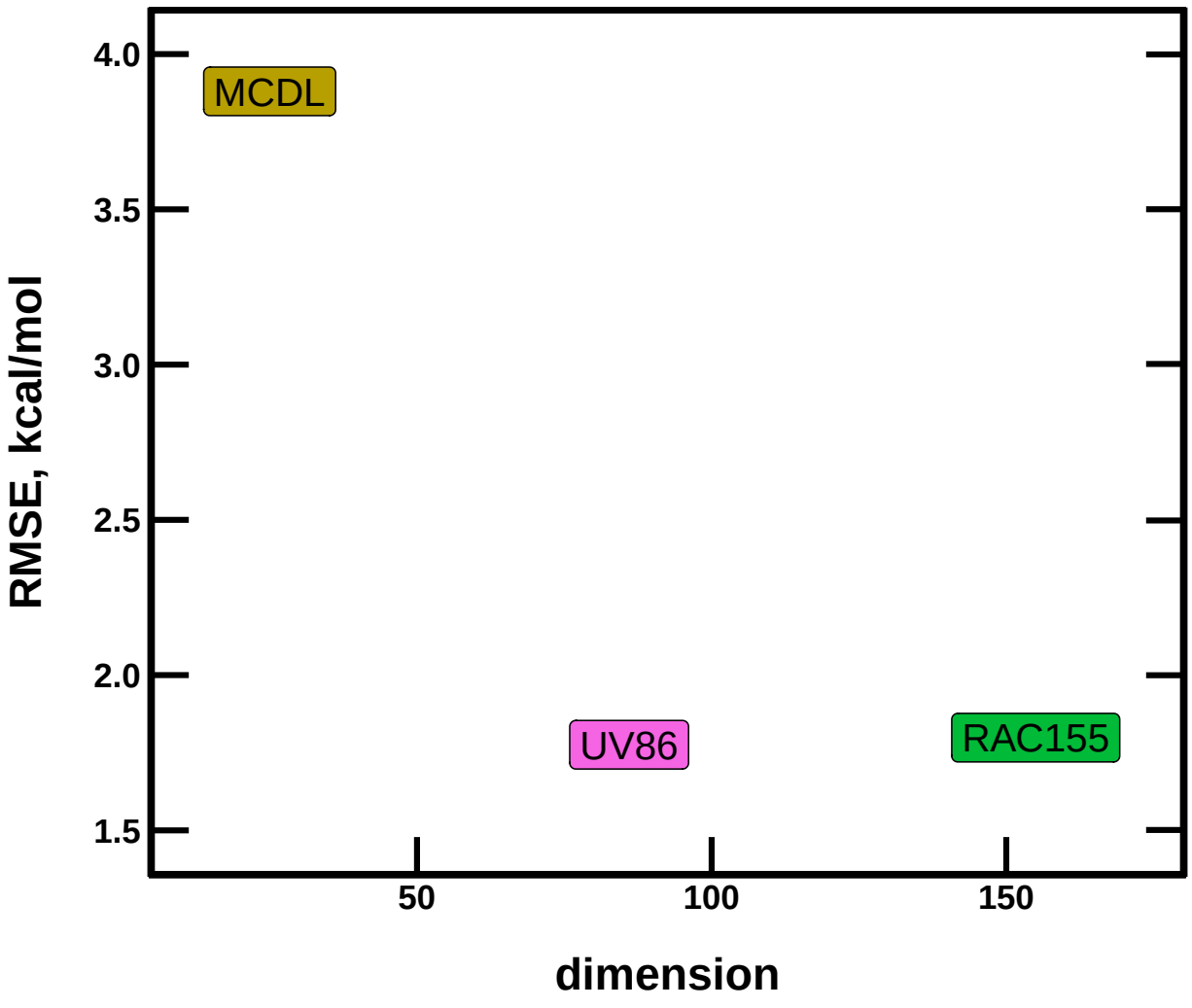
Janet, J.P., and Kulik, H.J. *J. Phys. Chem. A*, 2017,121, 46, 8939-8954.

Feature selection



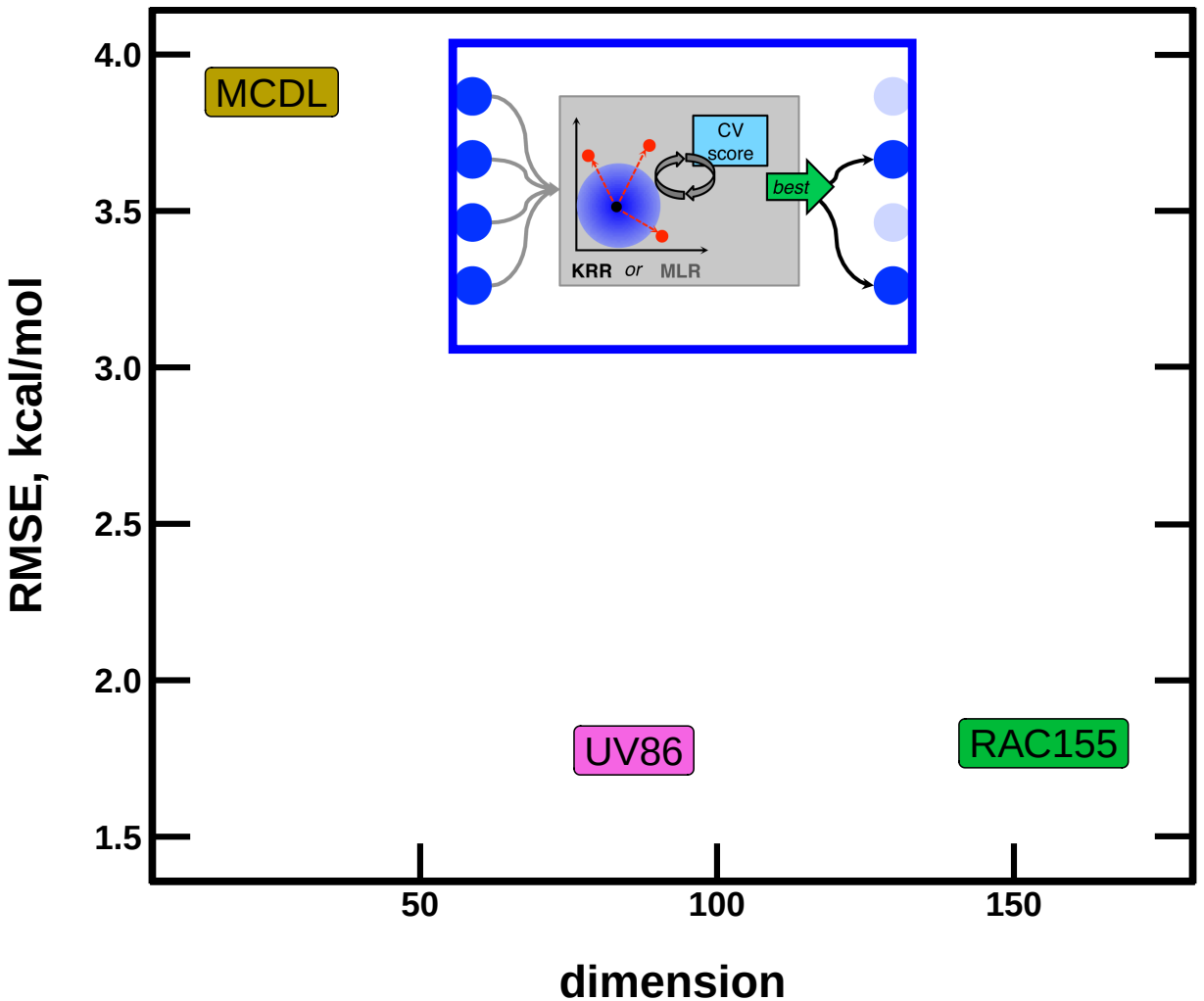
Janet, J.P., and Kulik, H.J. *J. Phys. Chem. A*, 2017,121, 46, 8939-8954.

Feature selection



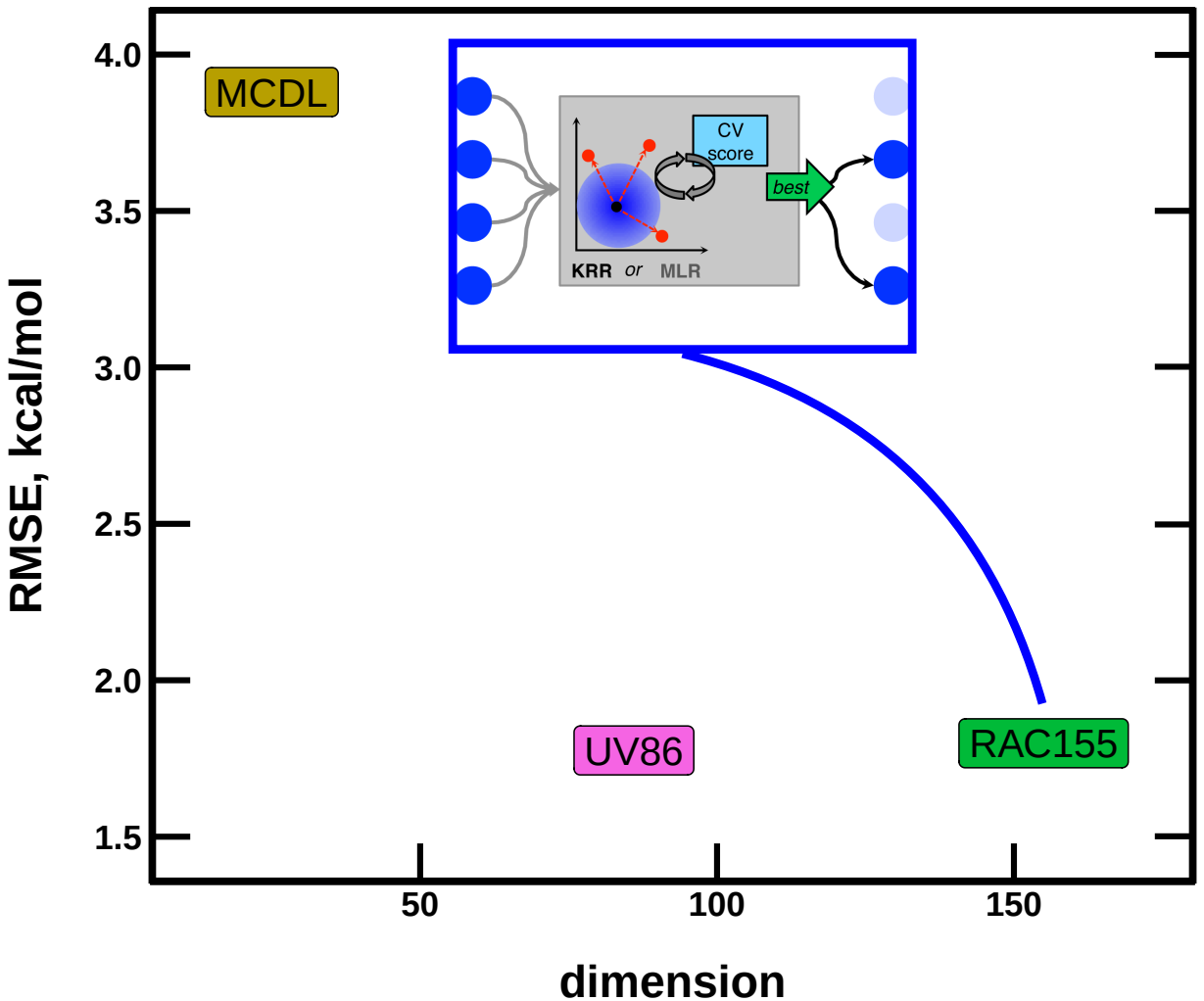
Janet, J.P., and Kulik, H.J. *J. Phys. Chem. A*, 2017,121, 46, 8939-8954.

Feature selection



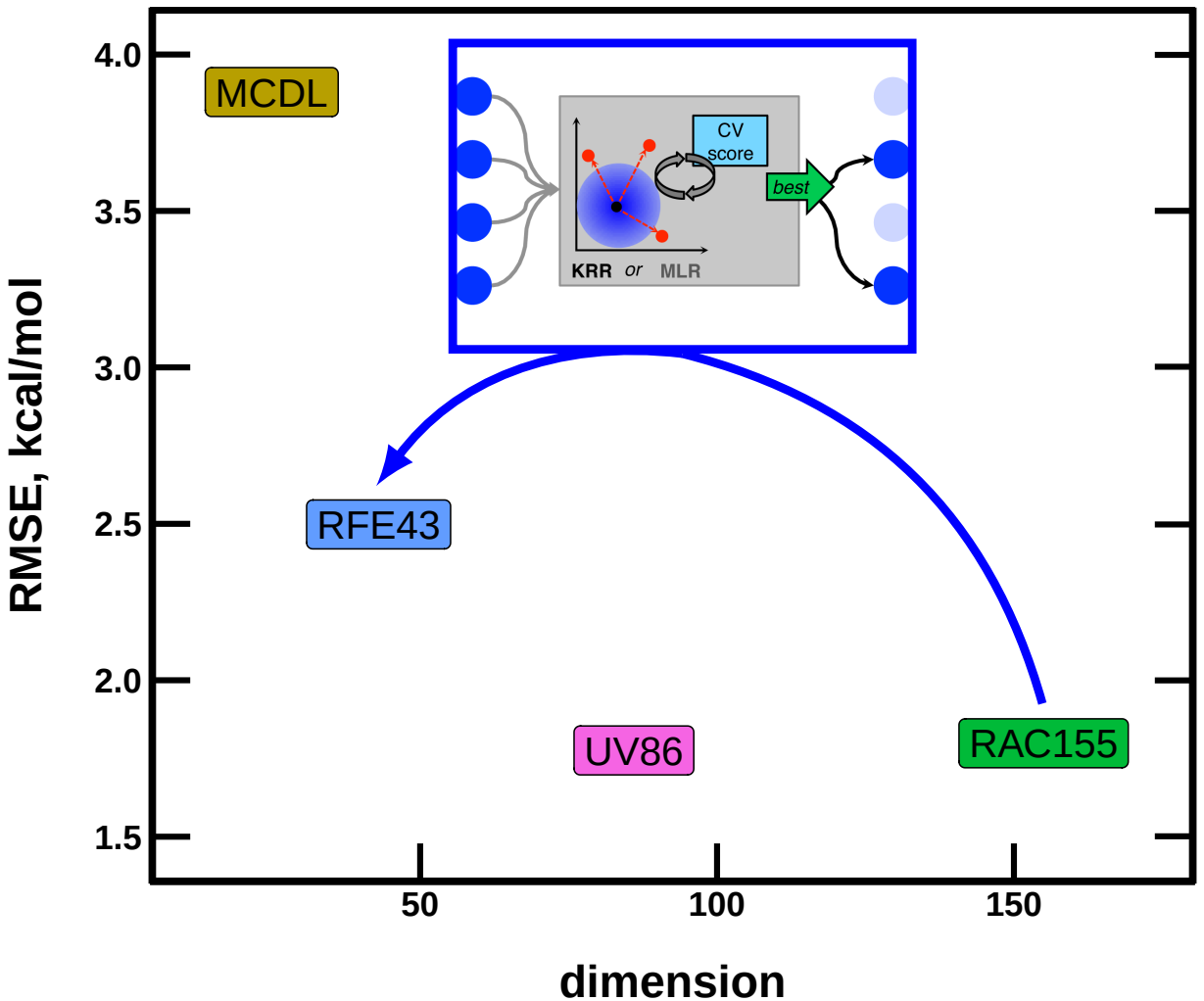
Janet, J.P., and Kulik, H.J. *J. Phys. Chem. A*, 2017,121, 46, 8939-8954.

Feature selection



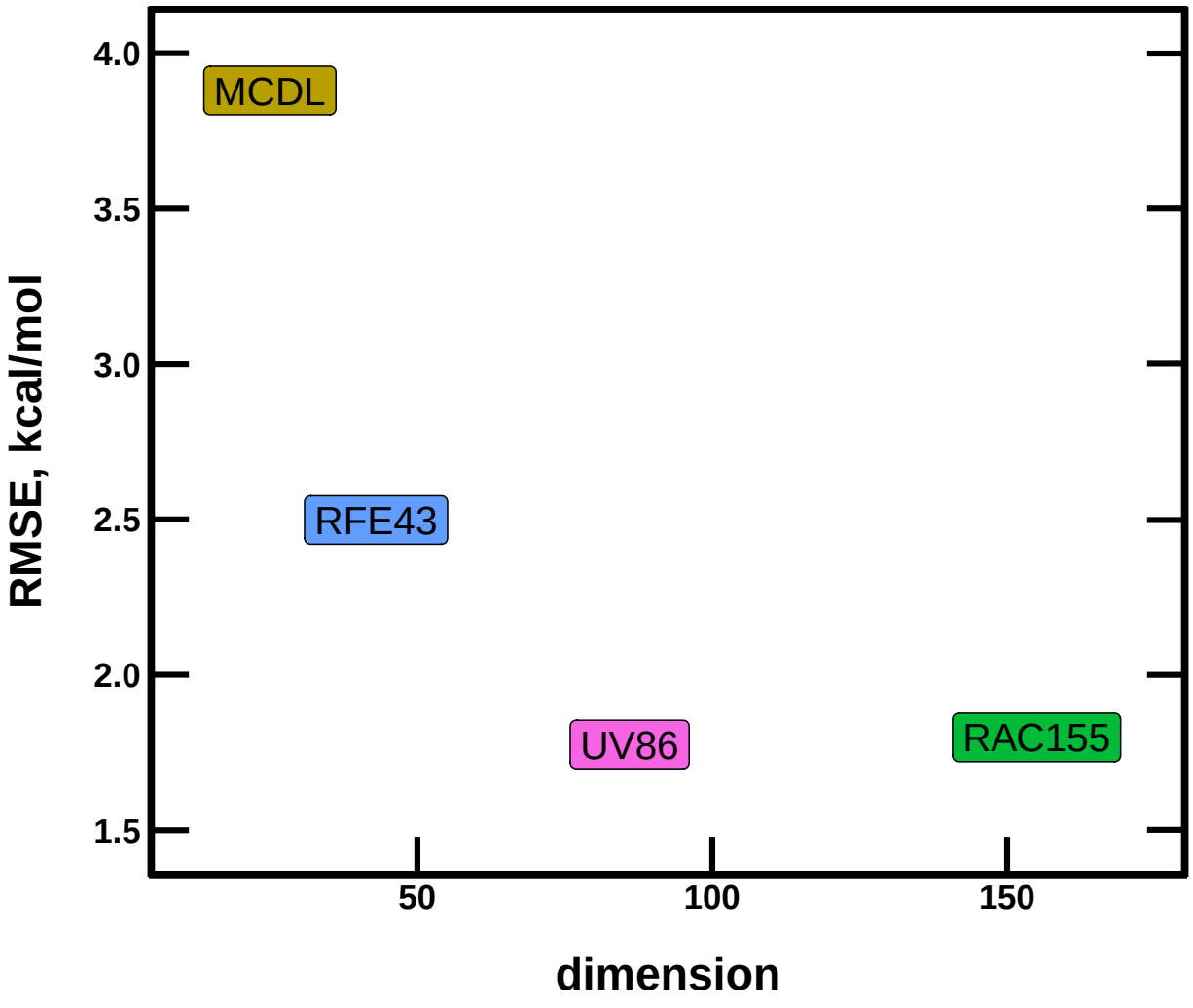
Janet, J.P., and Kulik, H.J. *J. Phys. Chem. A*, 2017,121, 46, 8939-8954.

Feature selection



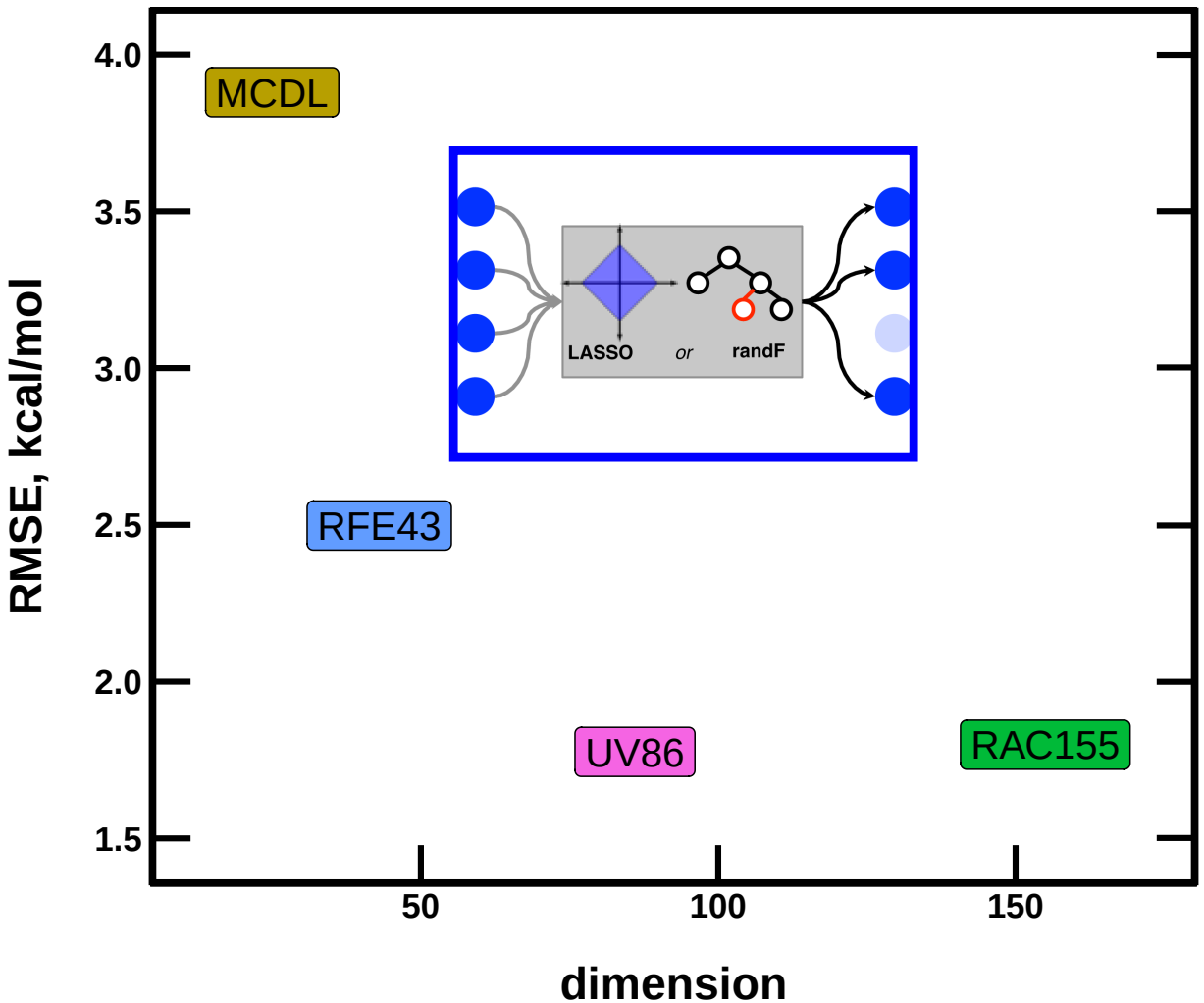
Janet, J.P., and Kulik, H.J. *J. Phys. Chem. A*, 2017,121, 46, 8939-8954.

Feature selection



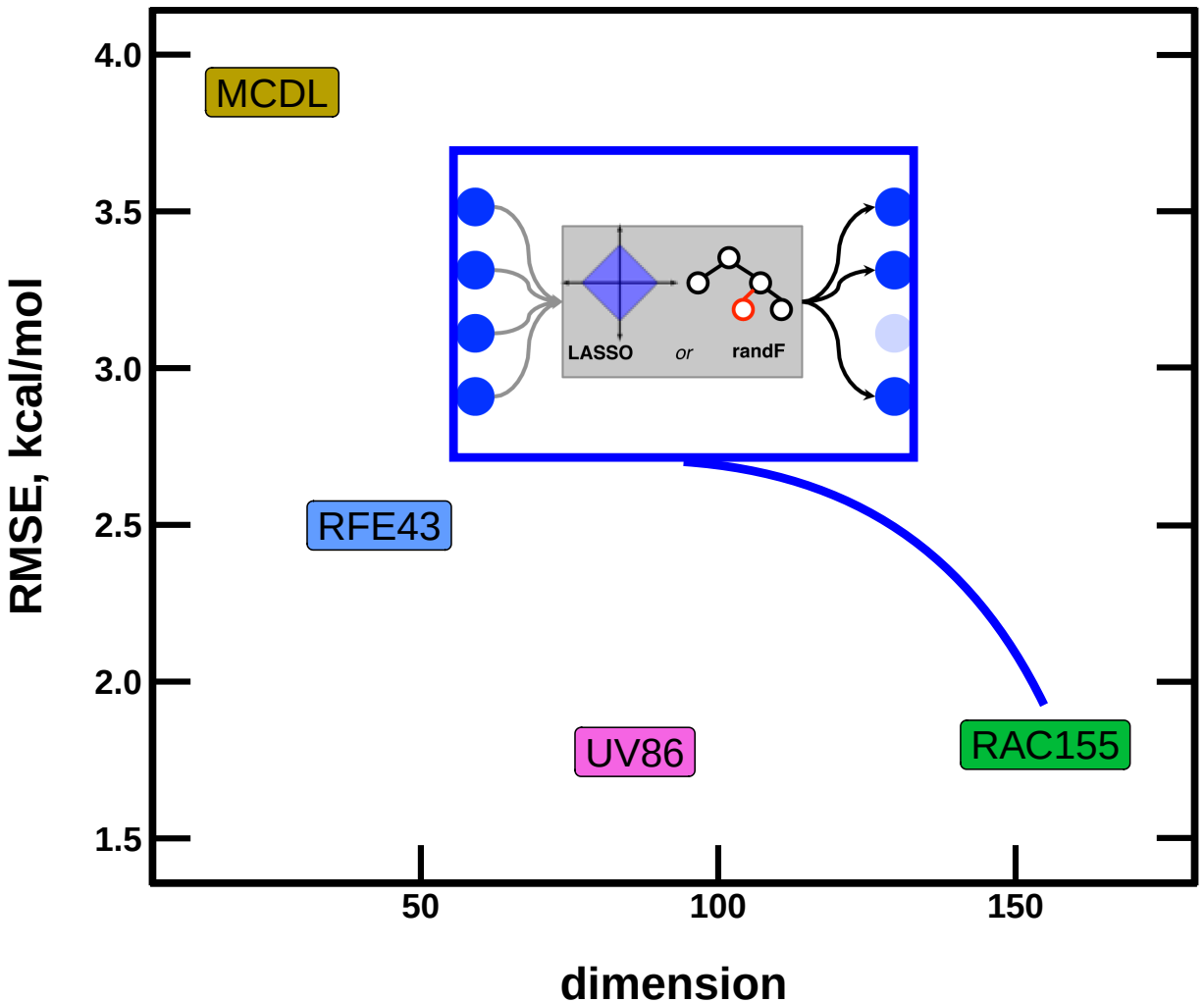
Janet, J.P., and Kulik, H.J. *J. Phys. Chem. A*, 2017,121, 46, 8939-8954.

Feature selection



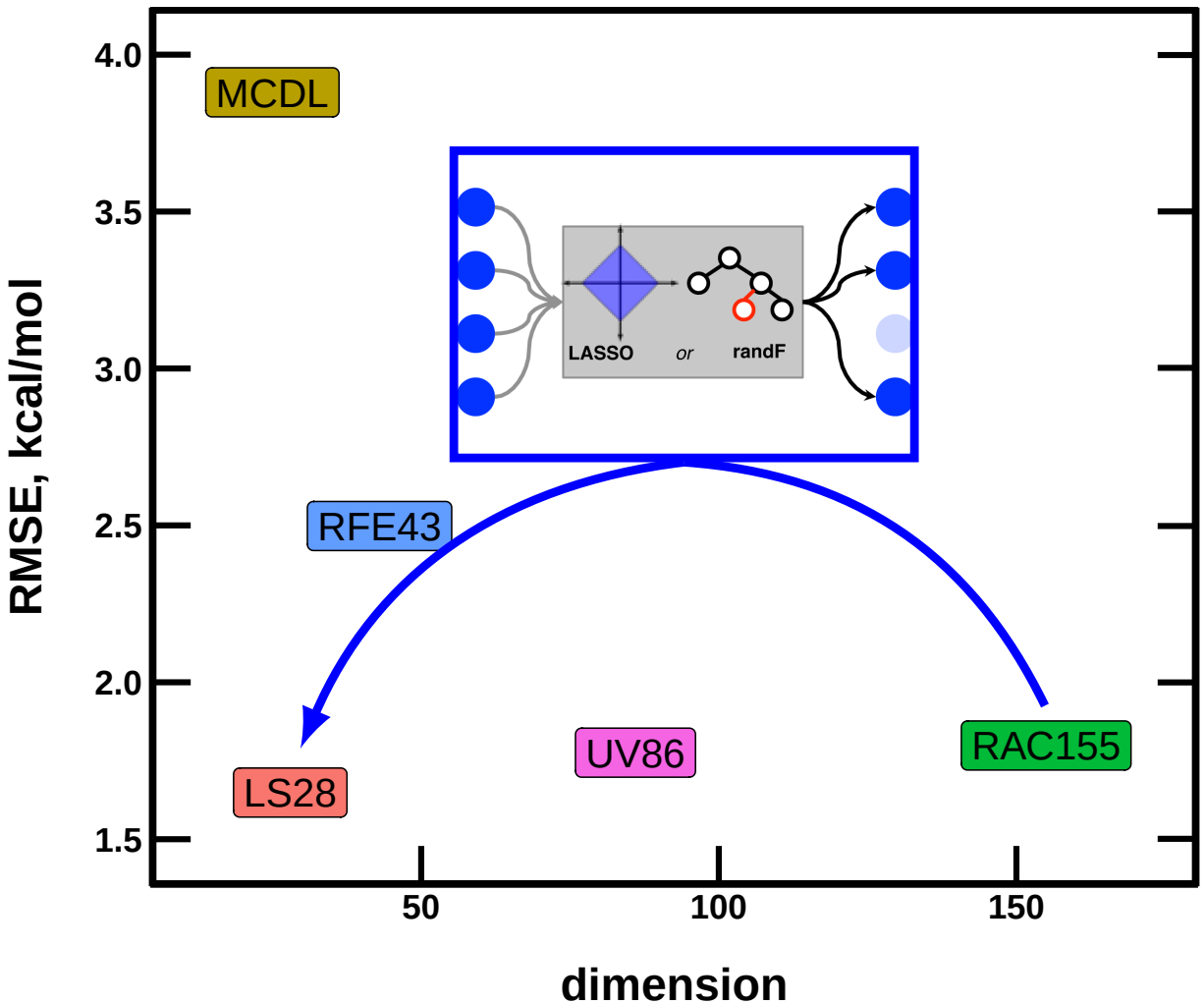
Janet, J.P., and Kulik, H.J. *J. Phys. Chem. A*, 2017,121, 46, 8939-8954.

Feature selection



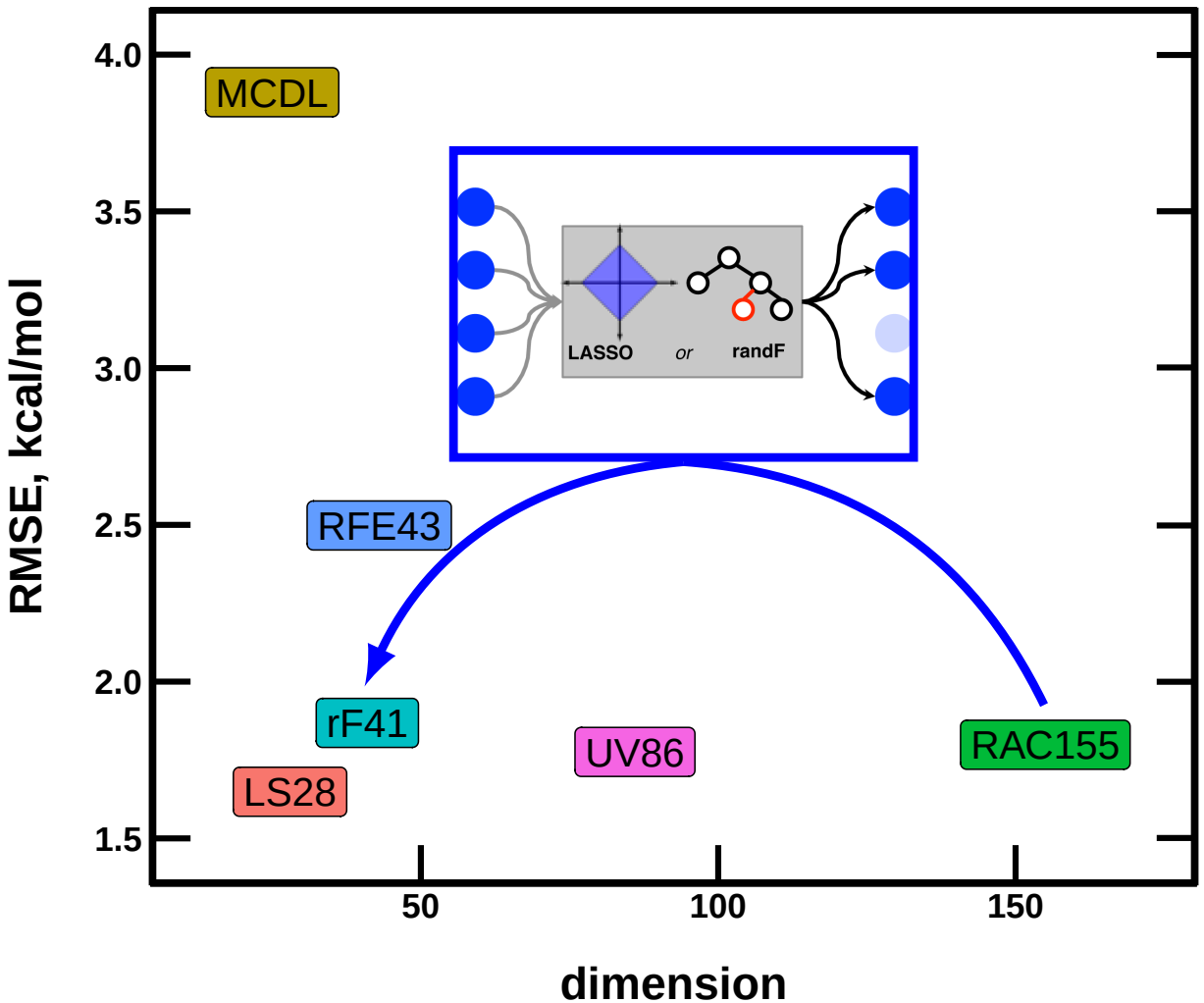
Janet, J.P., and Kulik, H.J. *J. Phys. Chem. A*, 2017,121, 46, 8939-8954.

Feature selection



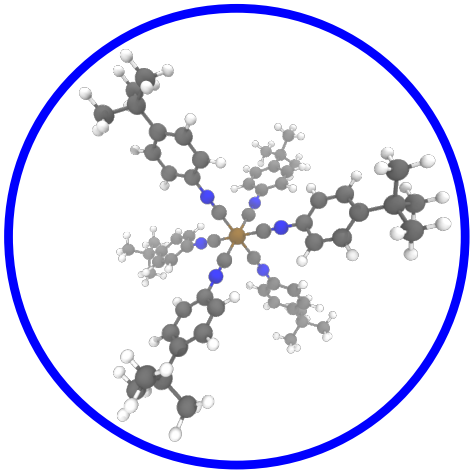
Janet, J.P., and Kulik, H.J. *J. Phys. Chem. A*, 2017,121, 46, 8939-8954.

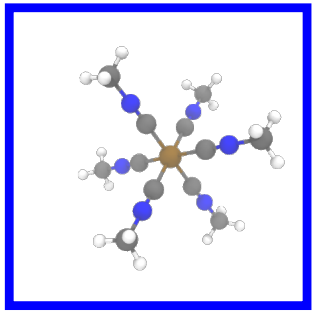
Feature selection

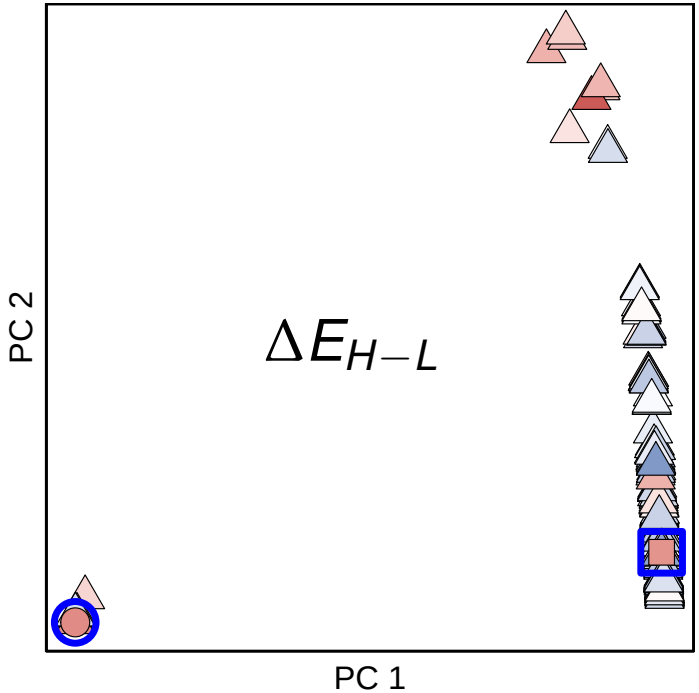


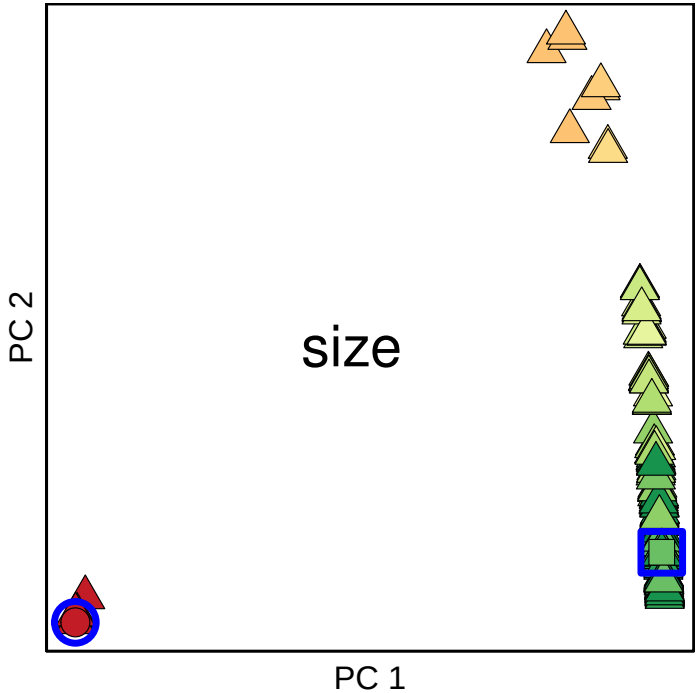
Janet, J.P., and Kulik, H.J. *J. Phys. Chem. A*, 2017,121, 46, 8939-8954.

A tale of two complexes, II

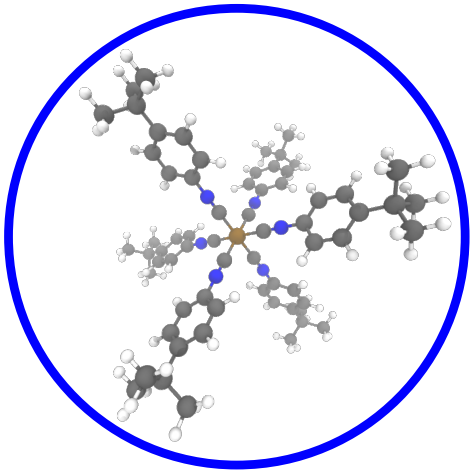


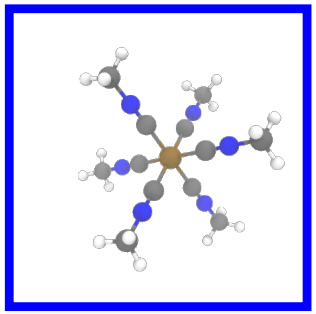


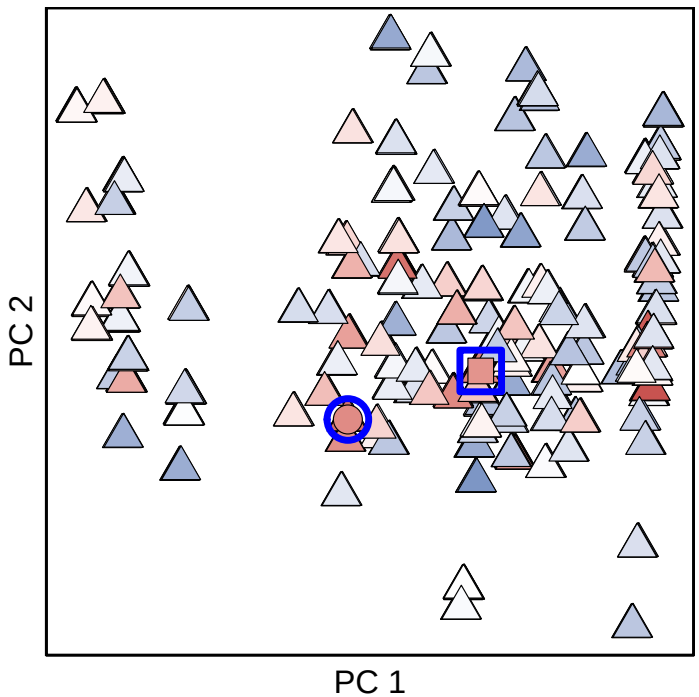


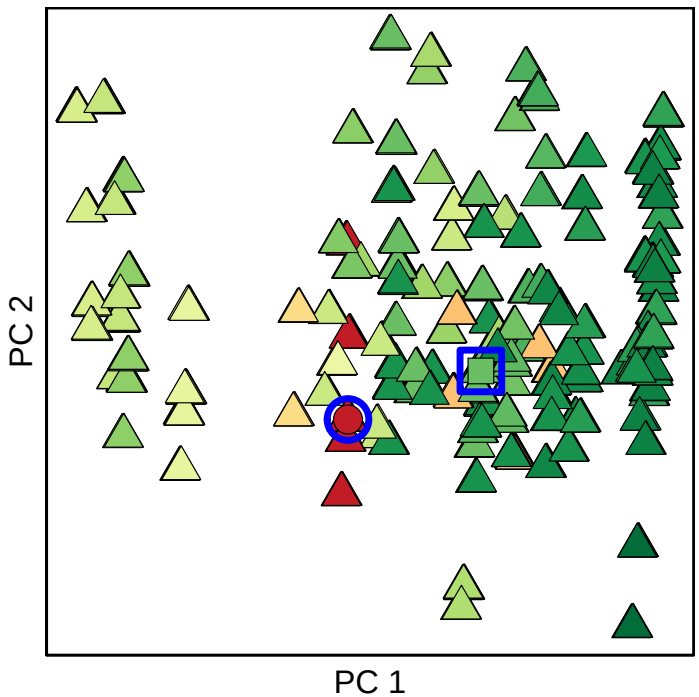


A tale of two complexes, II

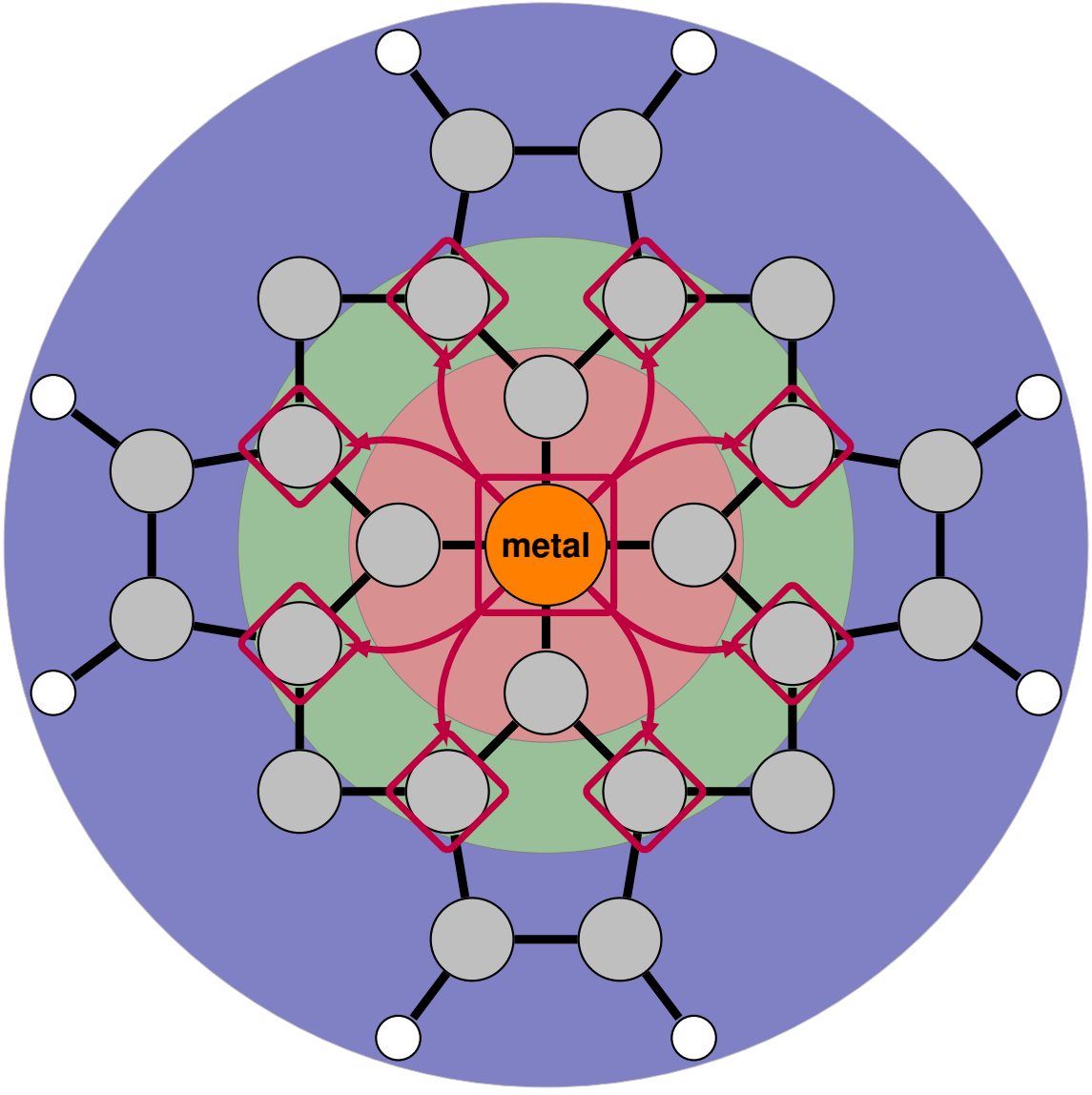




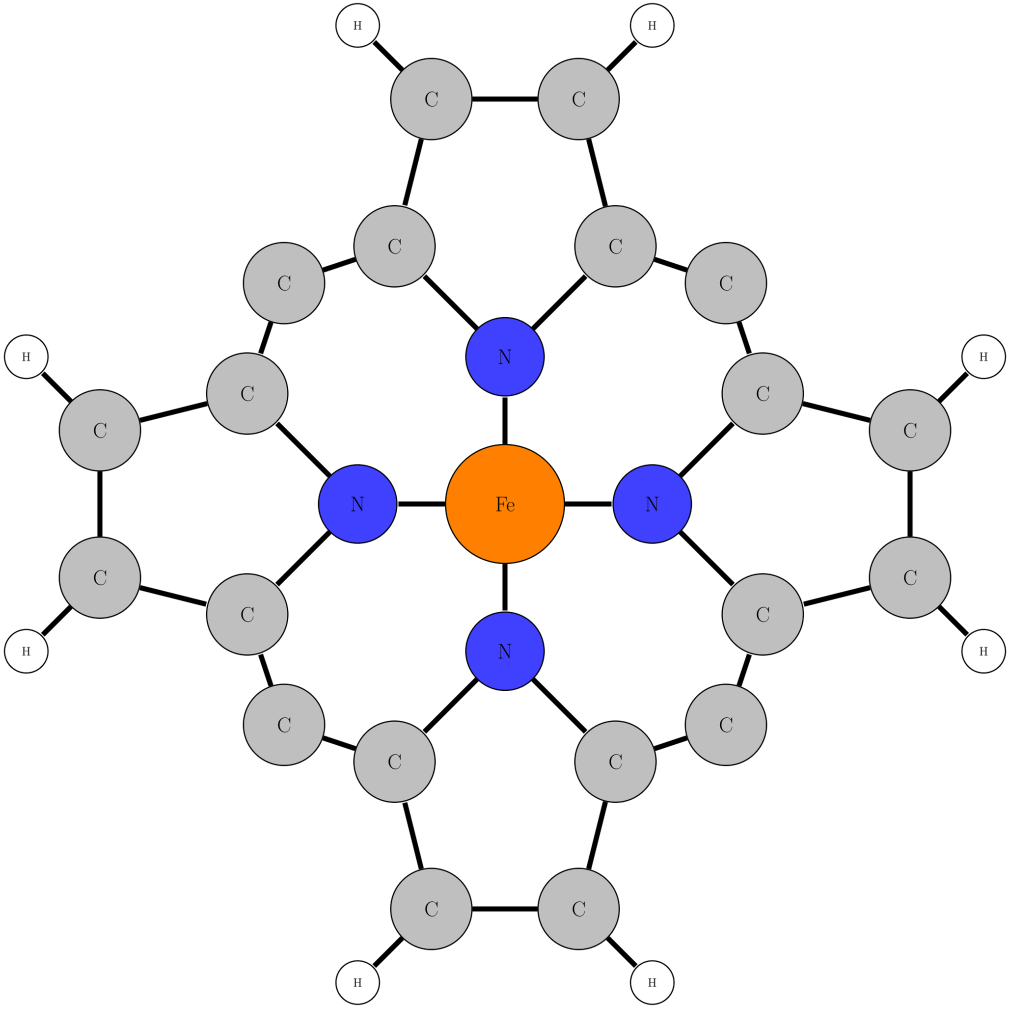




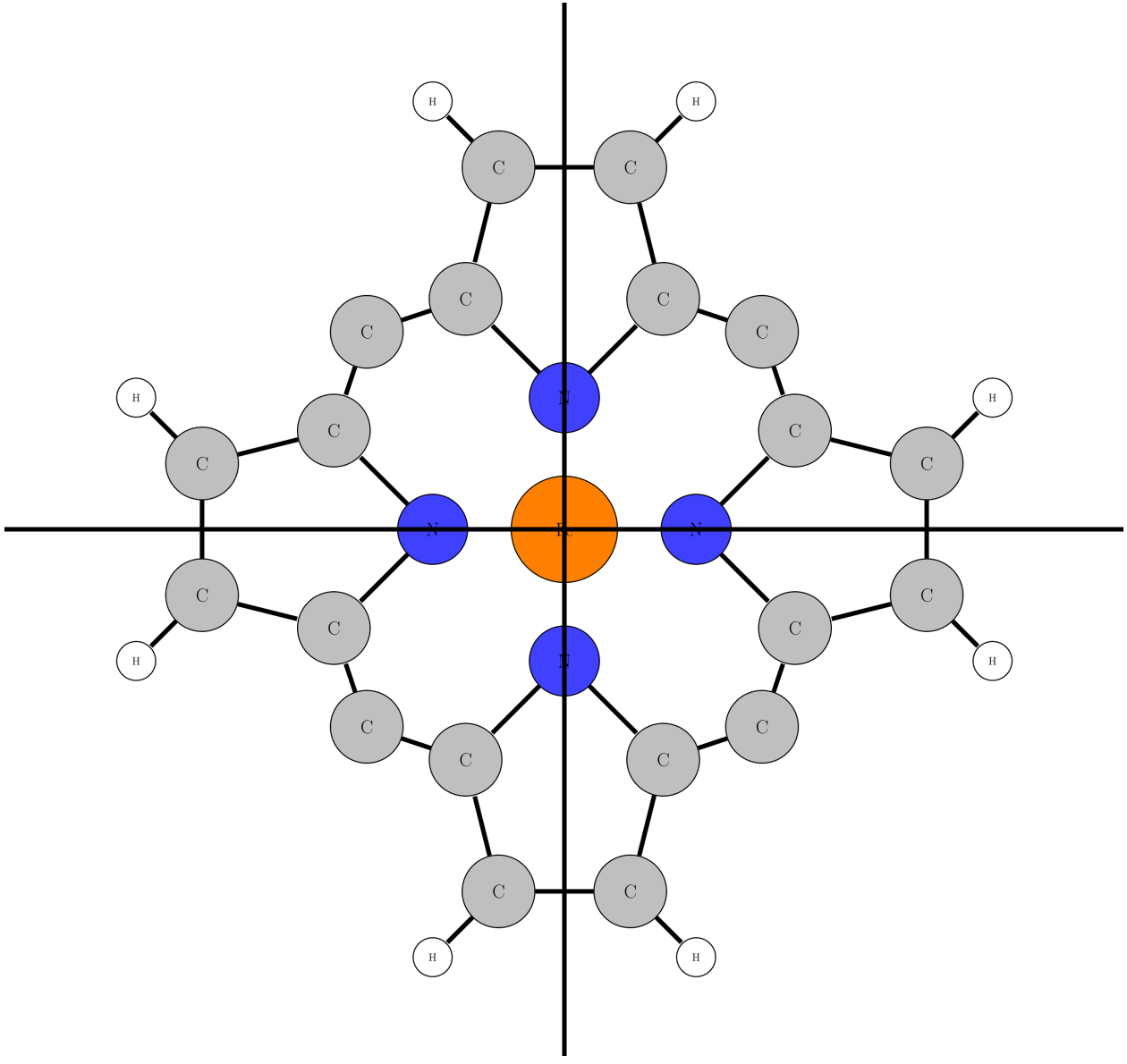
Do features depend on properties?



Do features depend on properties?

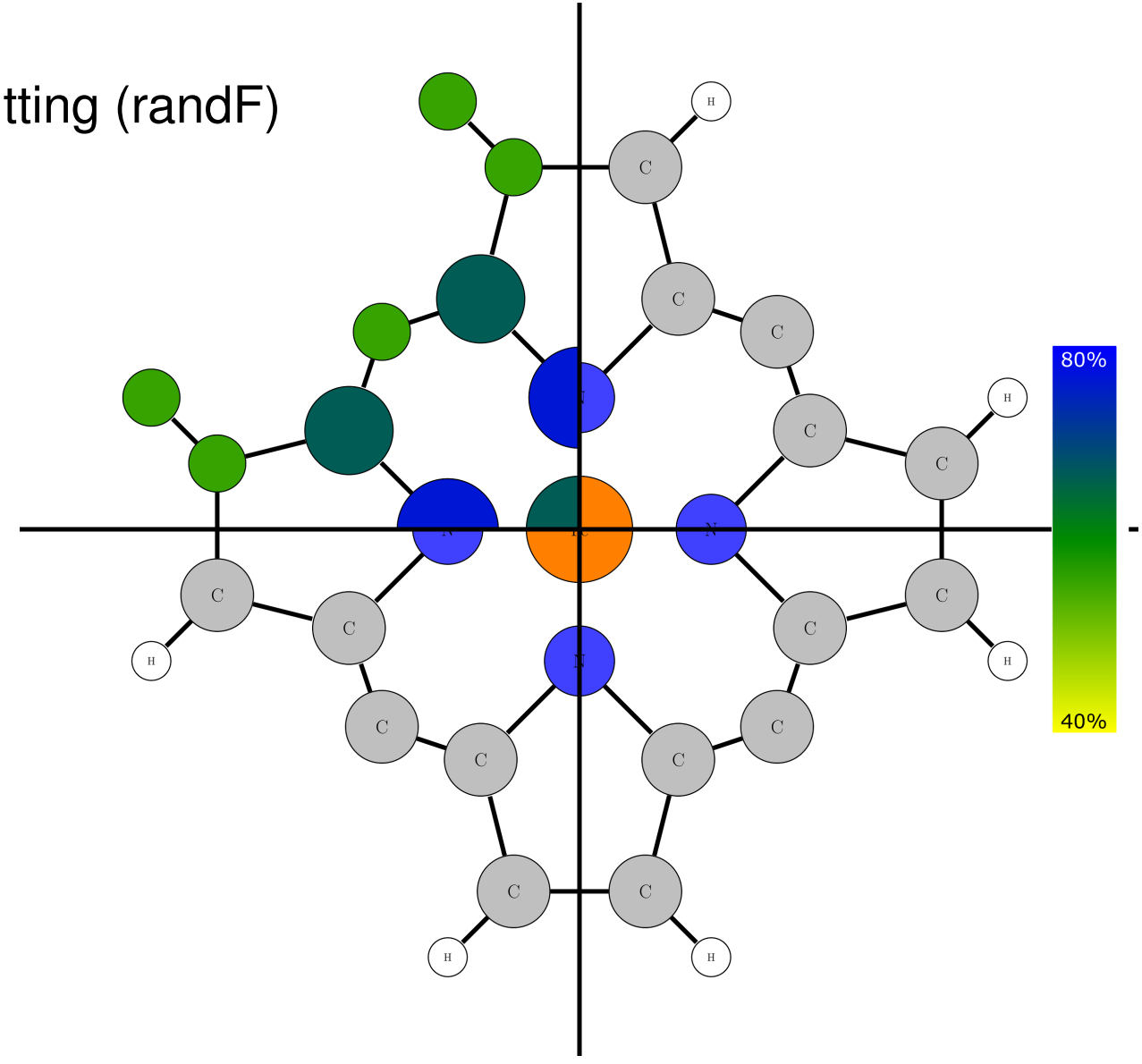


Do features depend on properties?

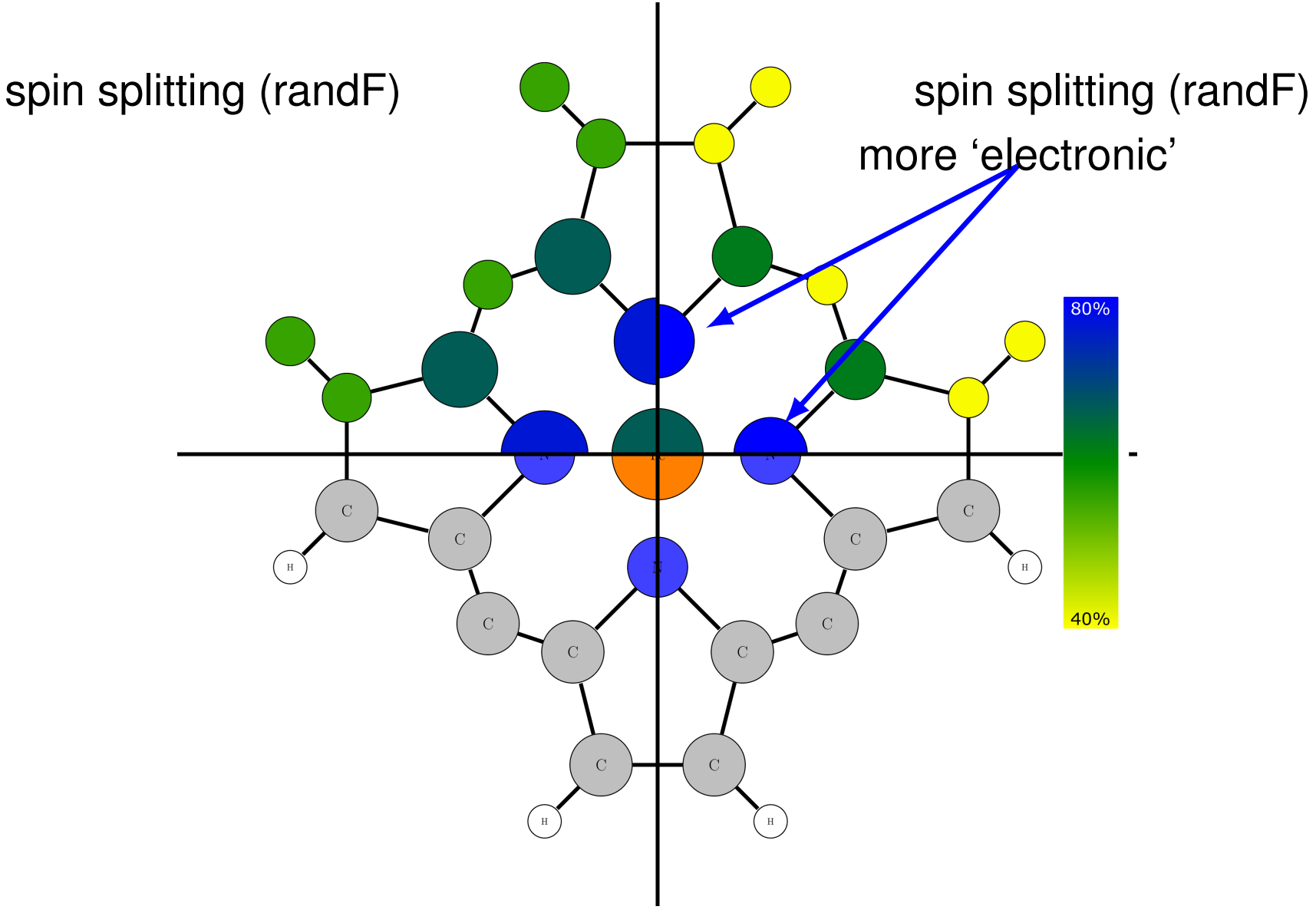


Do features depend on properties?

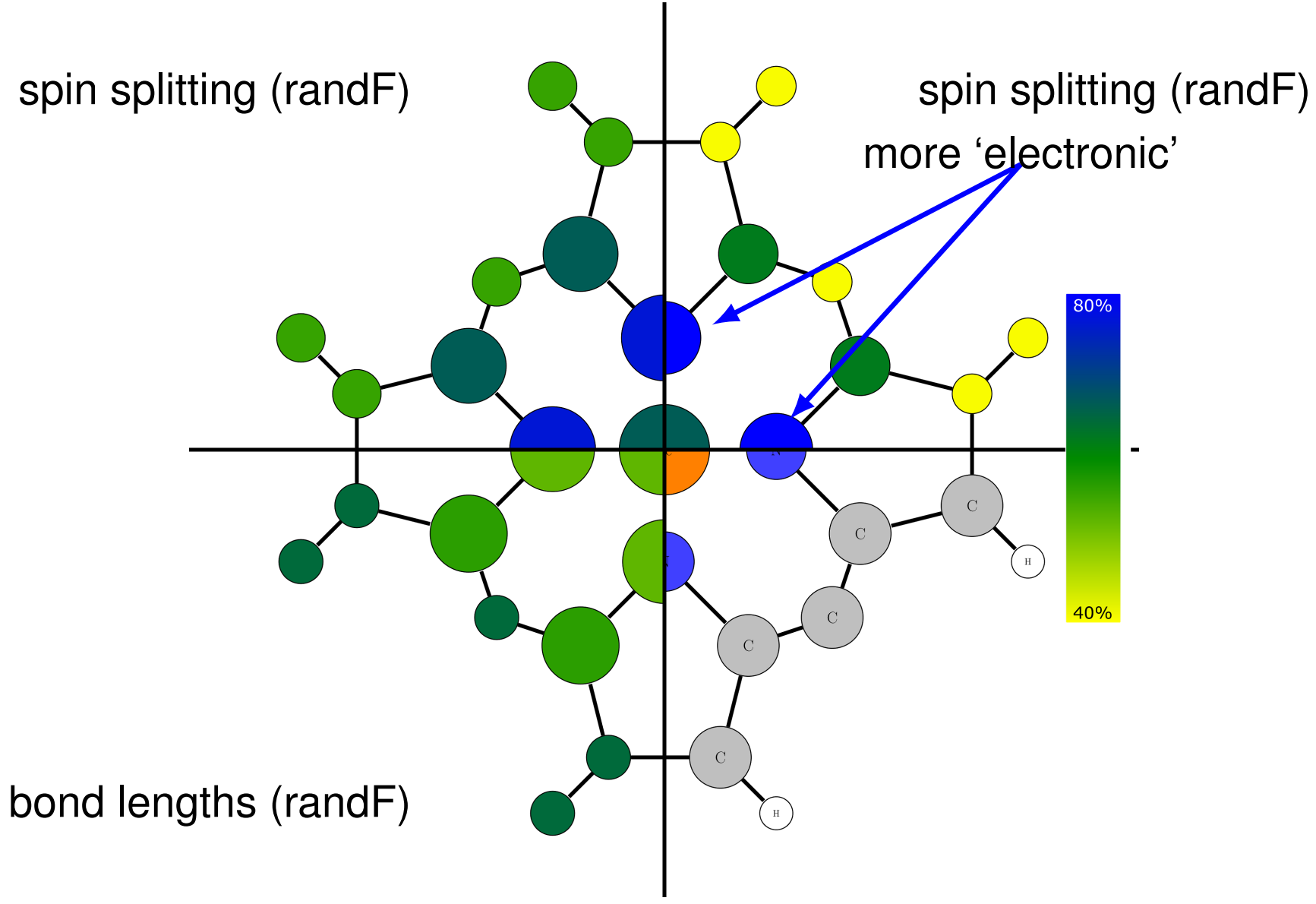
spin splitting (randF)



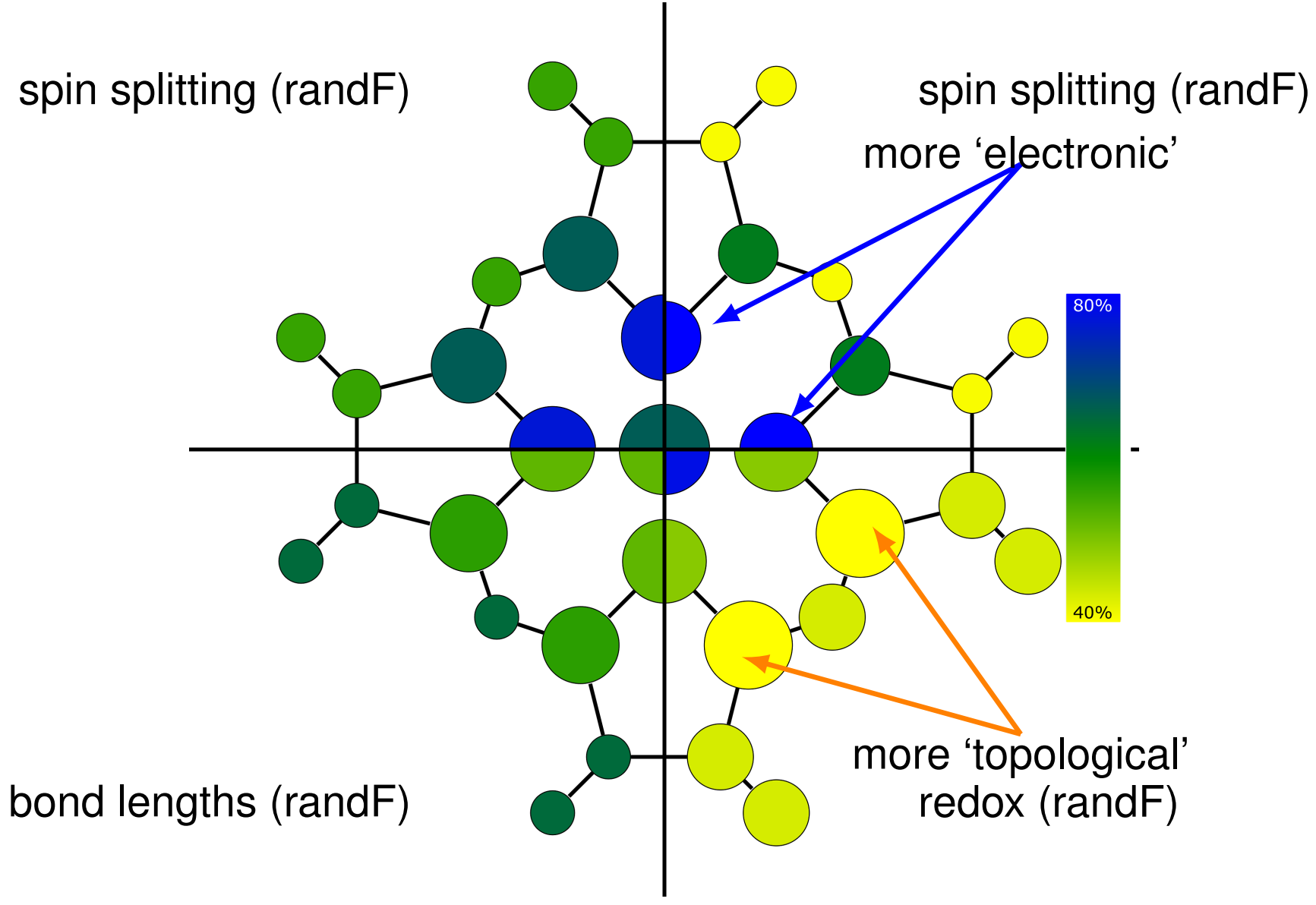
Do features depend on properties?



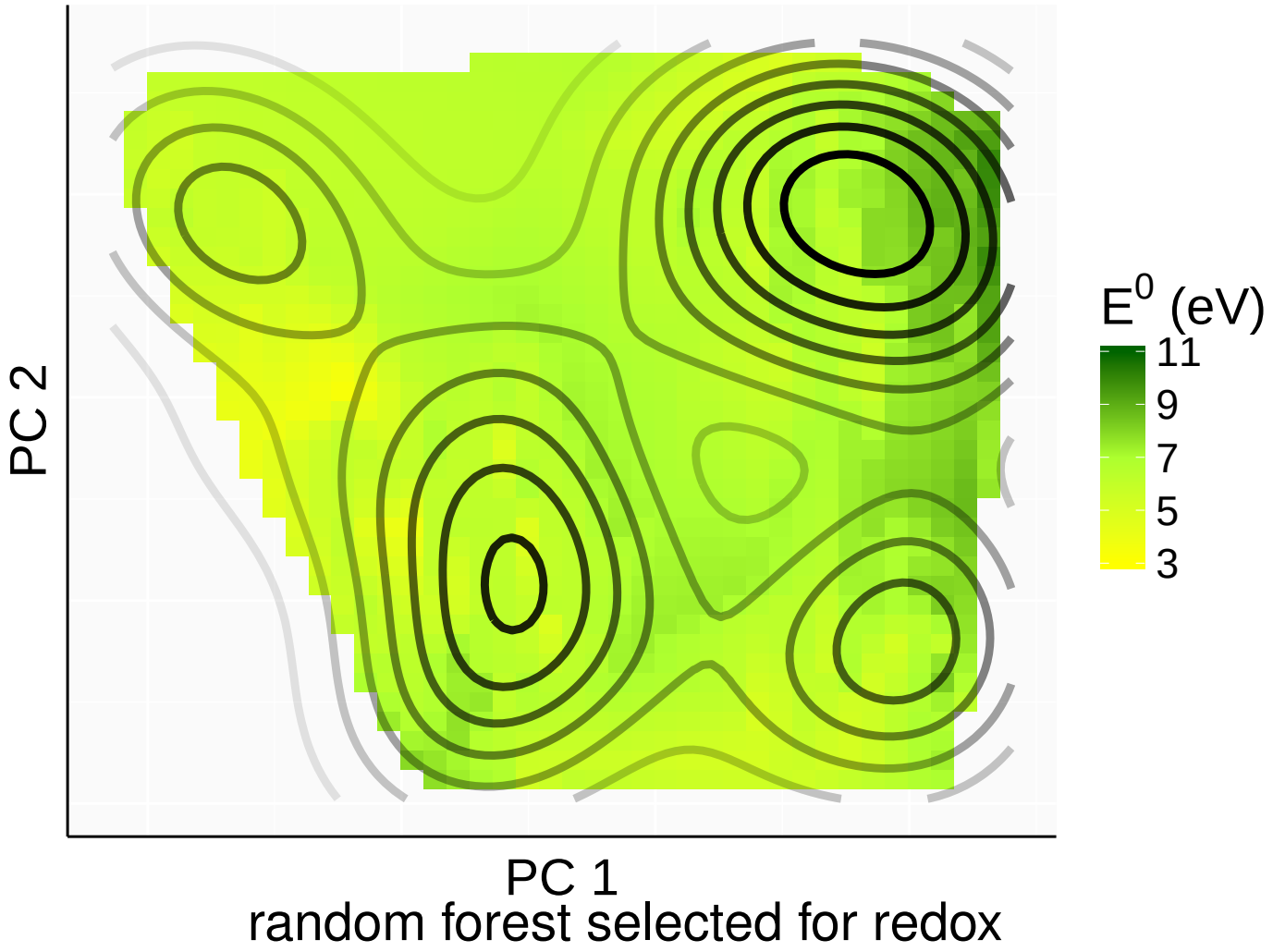
Do features depend on properties?



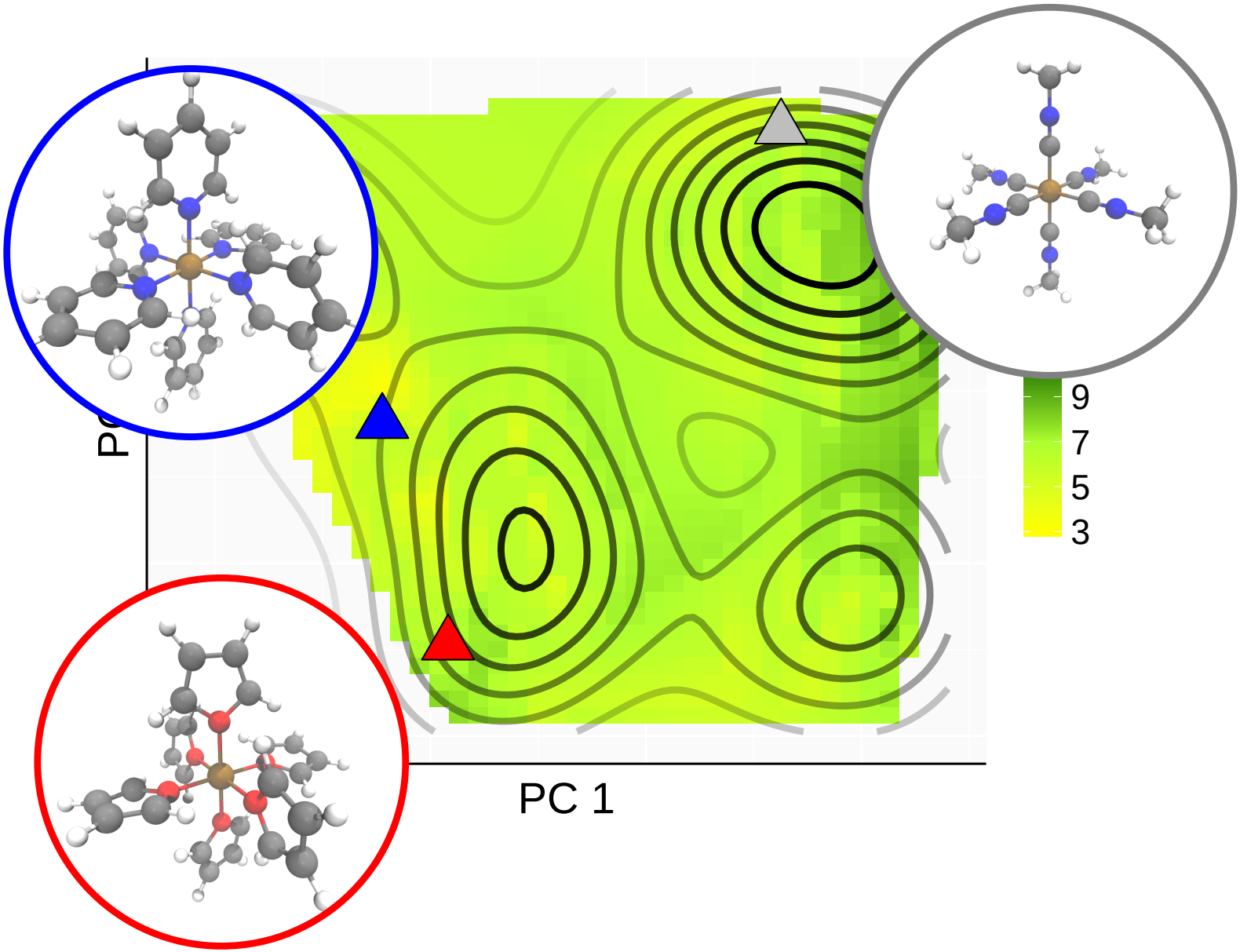
Do features depend on properties?



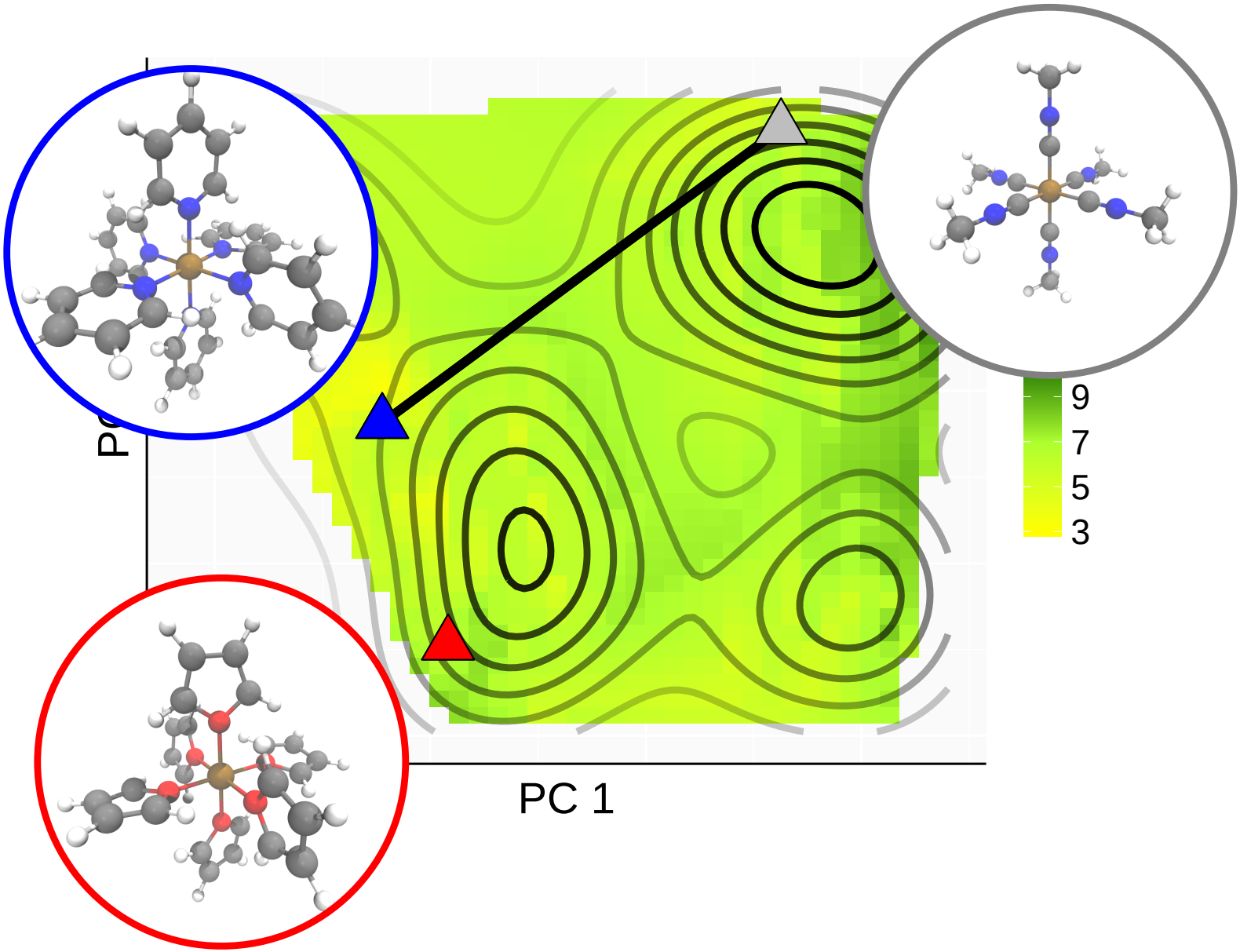
Mapping TM complex space



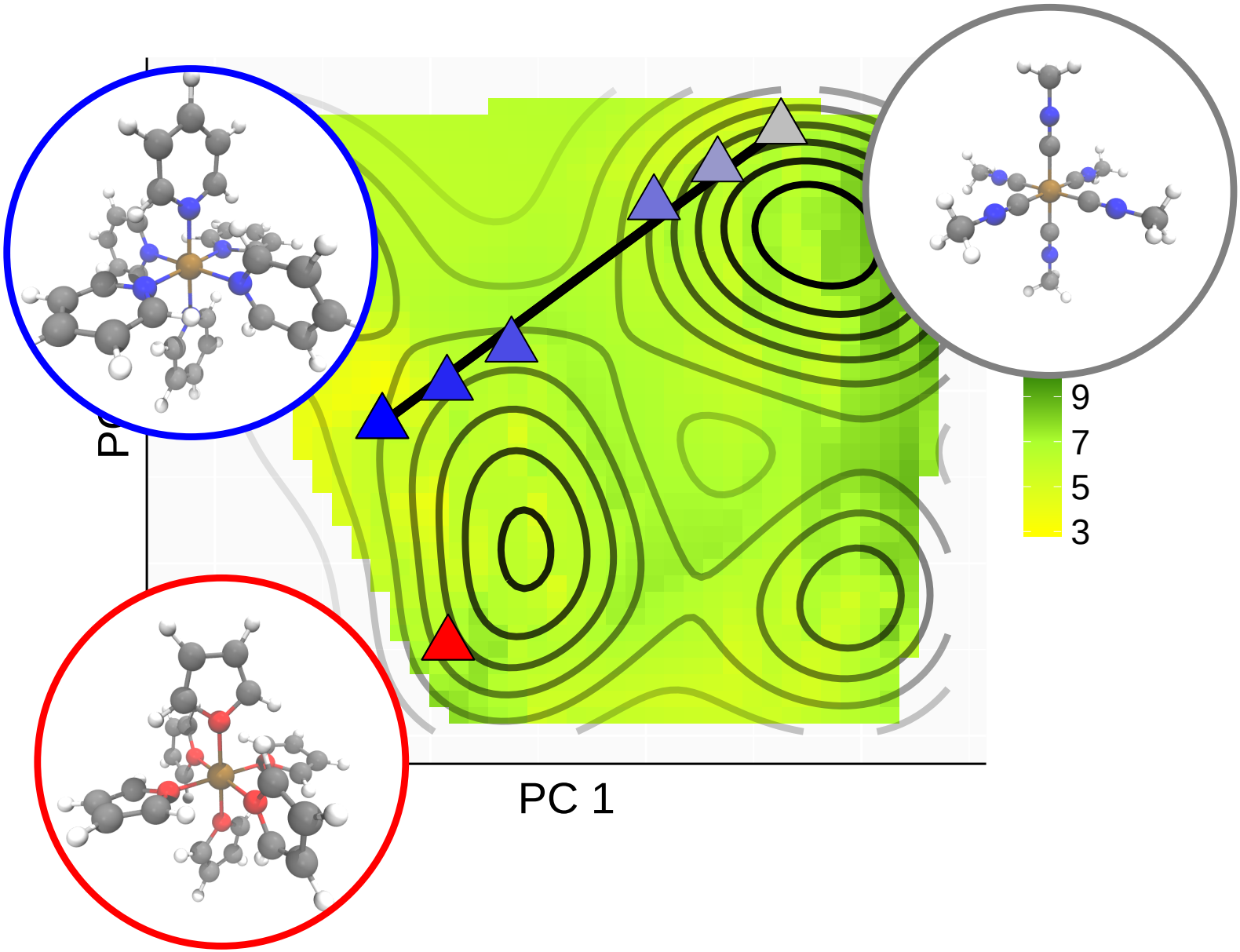
Mapping TM complex space



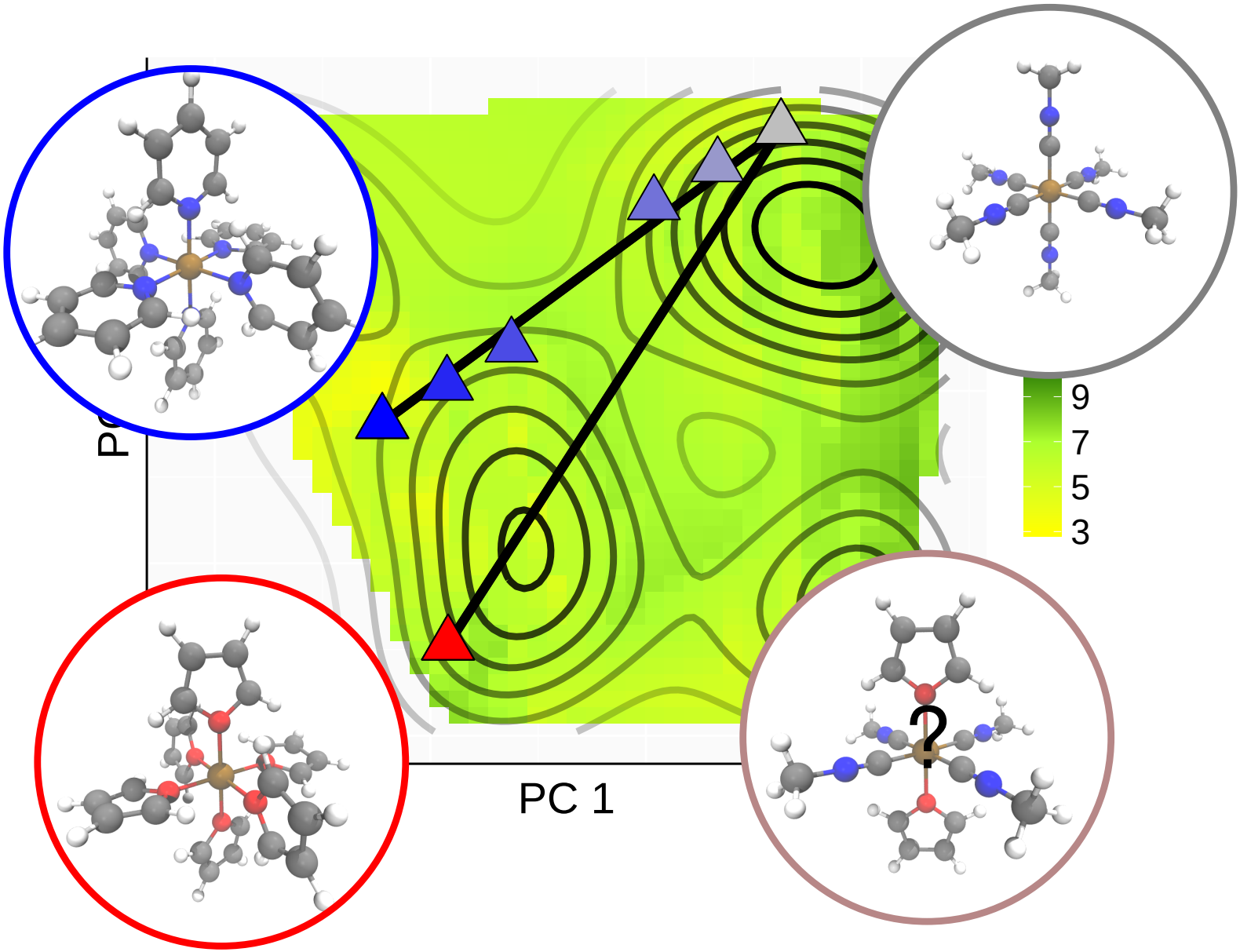
Mapping TM complex space



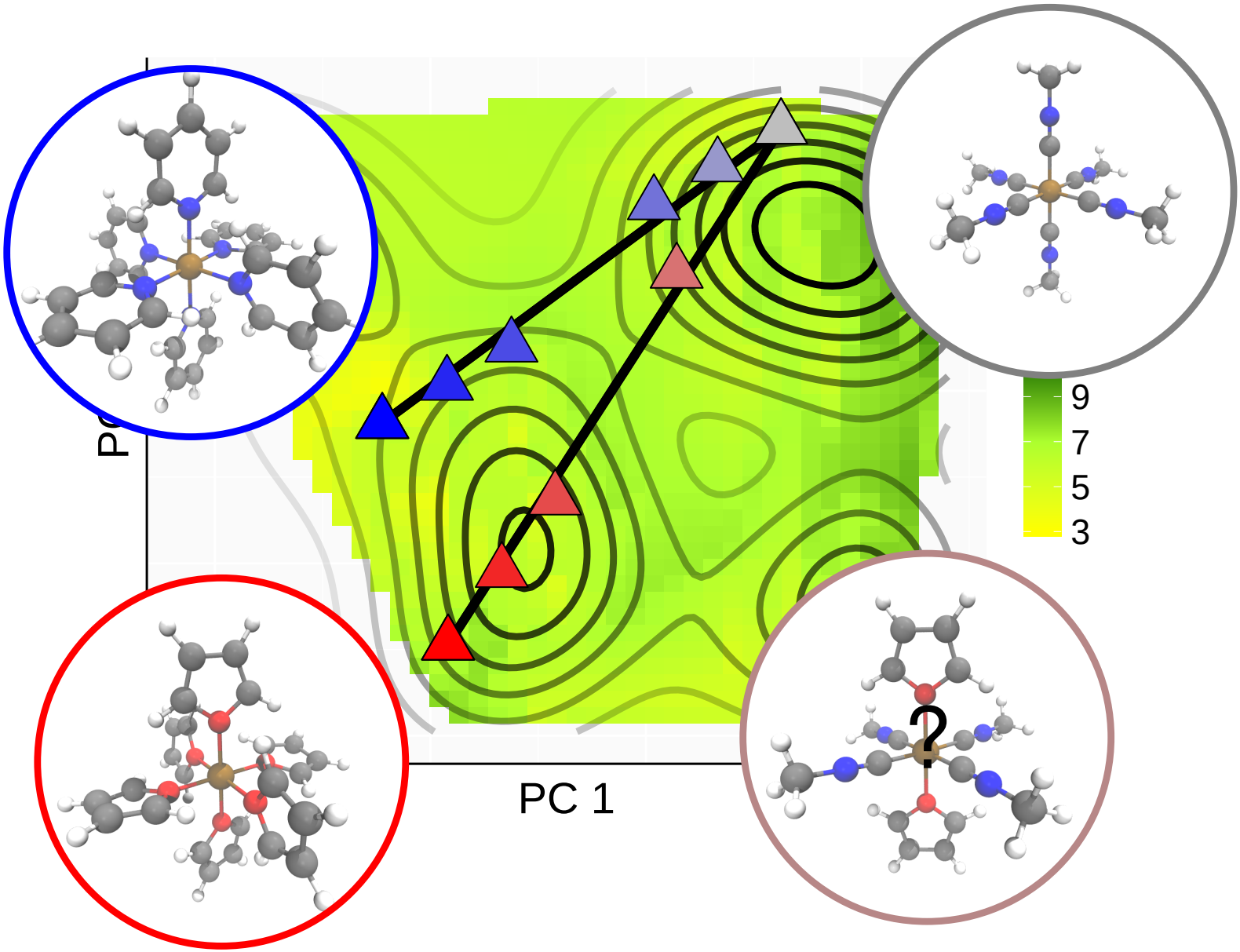
Mapping TM complex space



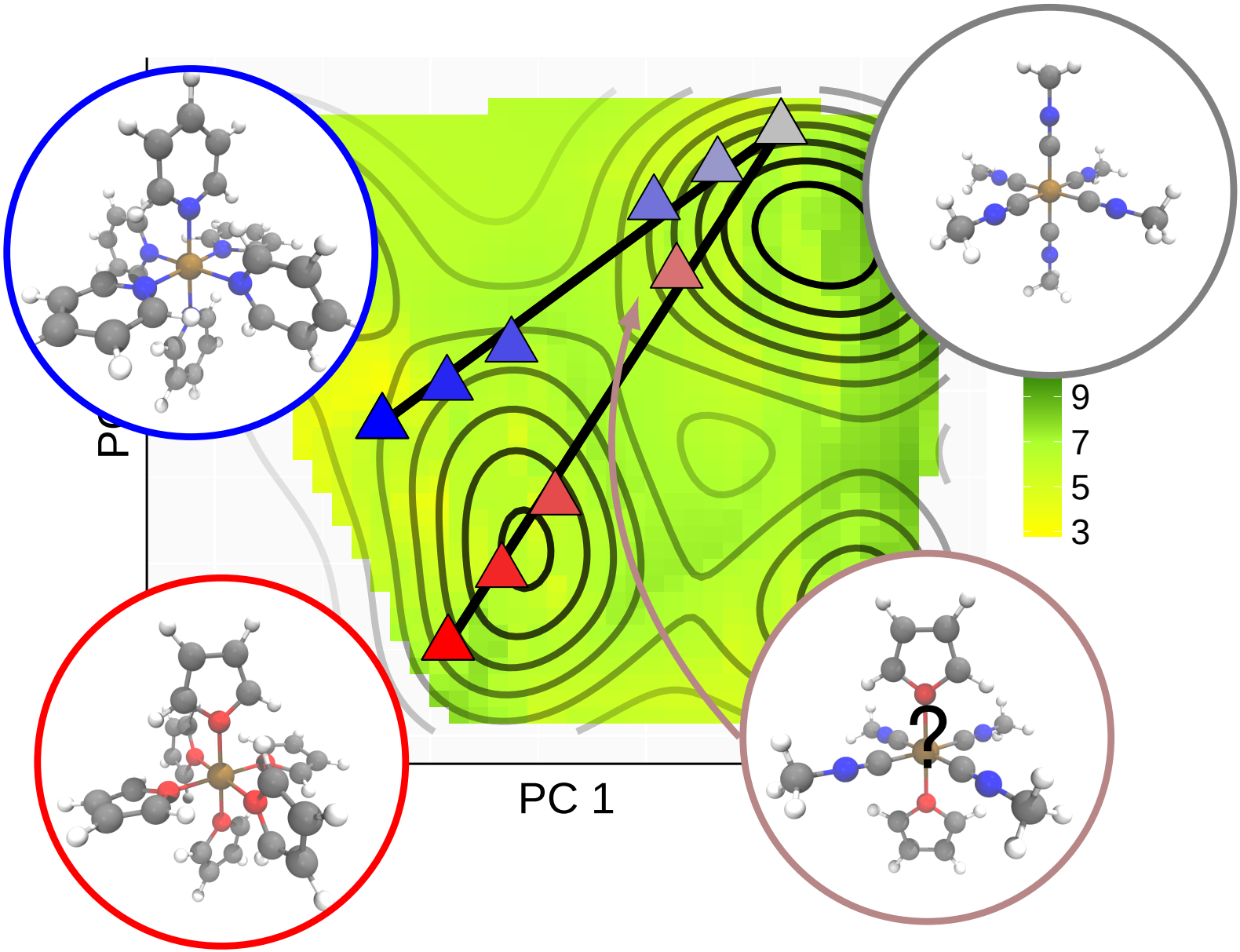
Mapping TM complex space



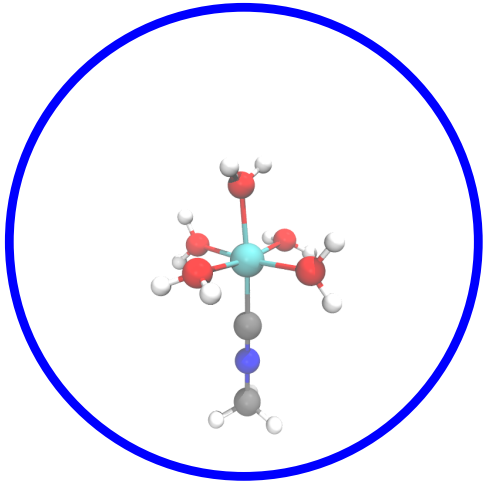
Mapping TM complex space



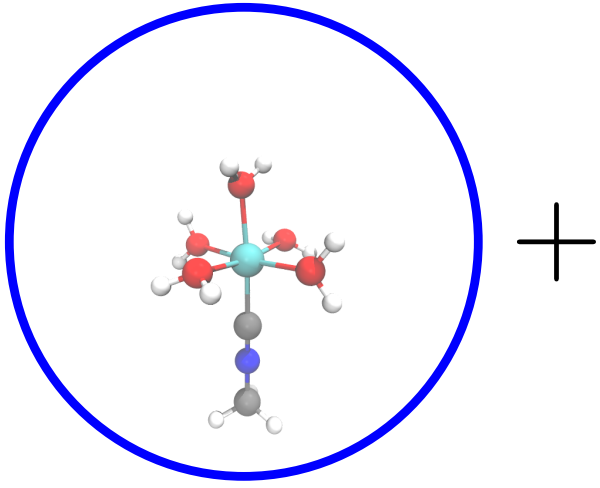
Mapping TM complex space



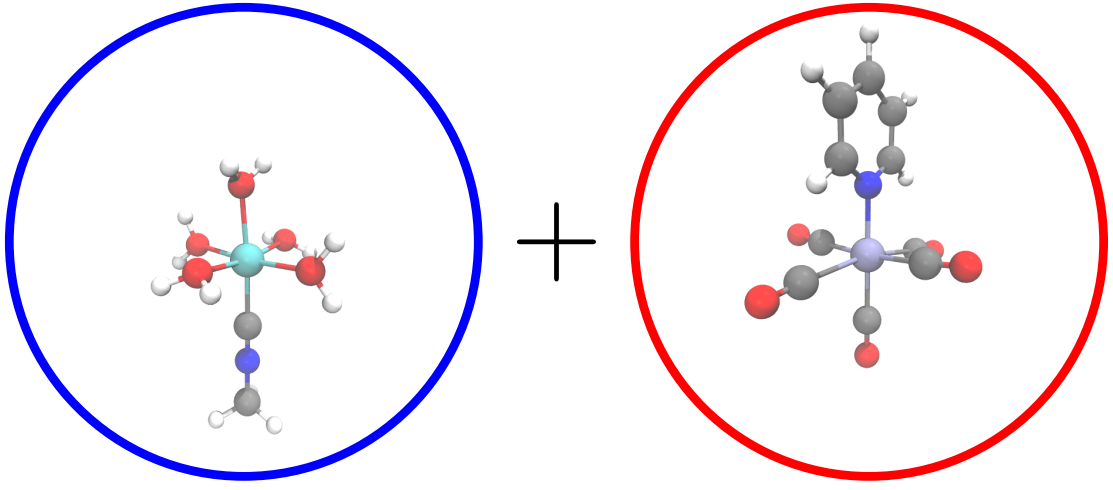
Mapping TM complex space



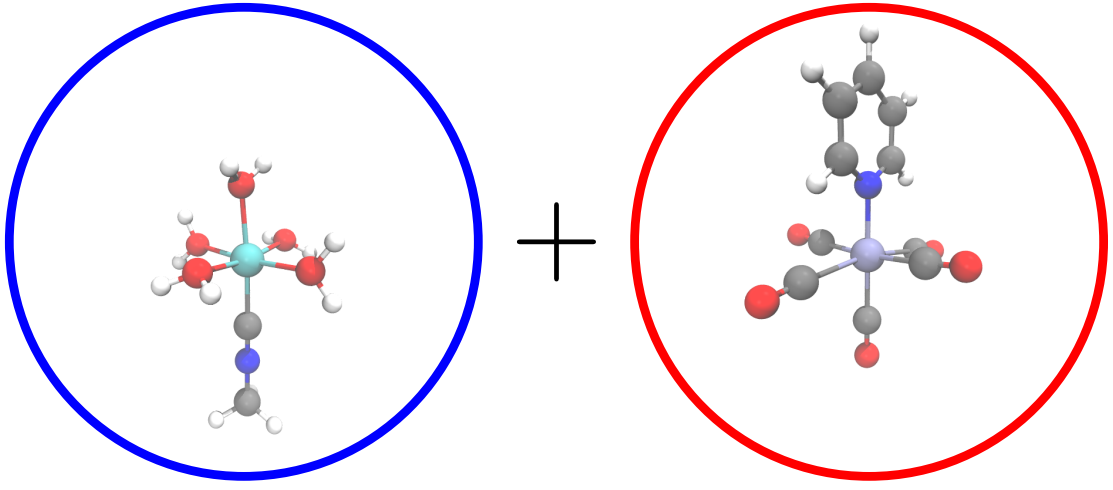
Mapping TM complex space



Mapping TM complex space

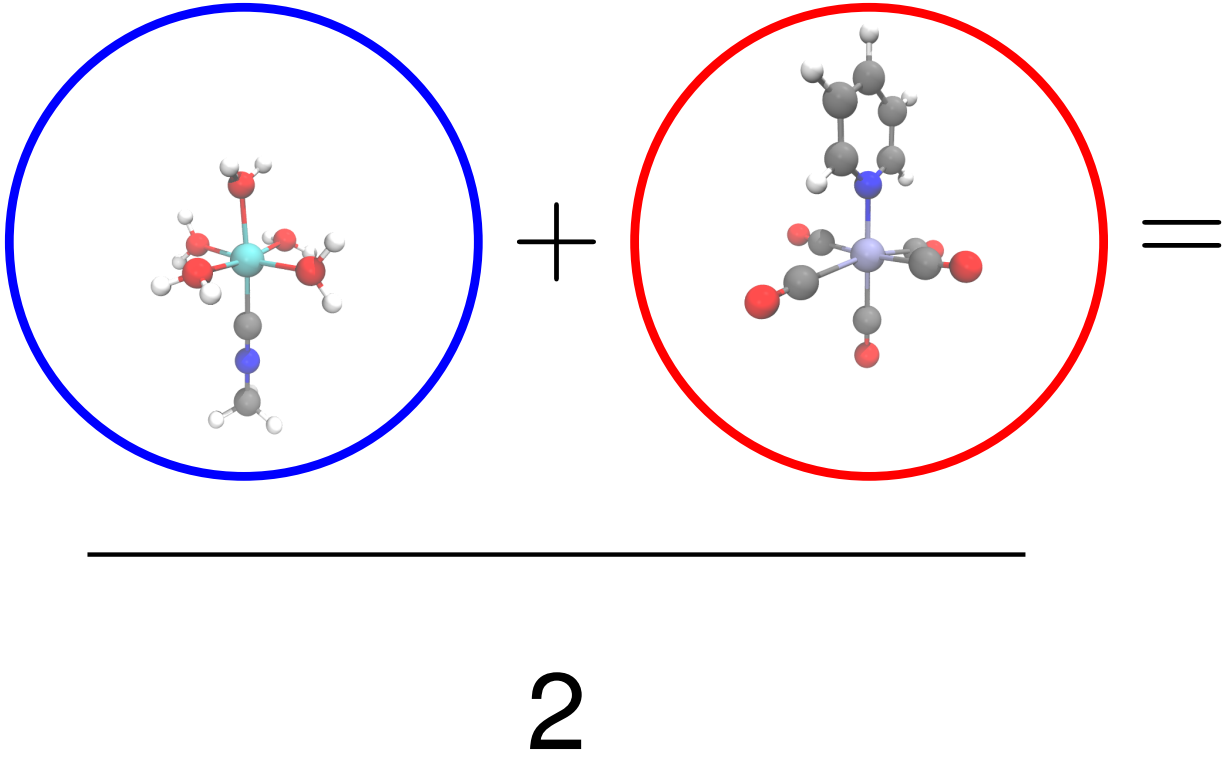


Mapping TM complex space

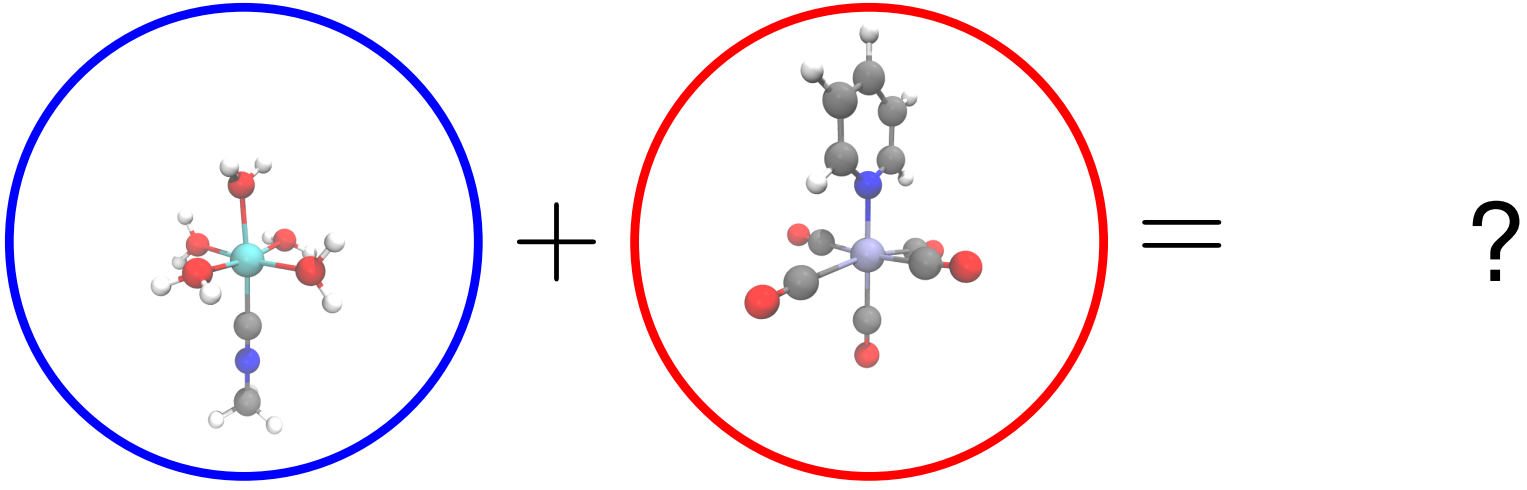


$$\frac{\quad + \quad}{2}$$

Mapping TM complex space

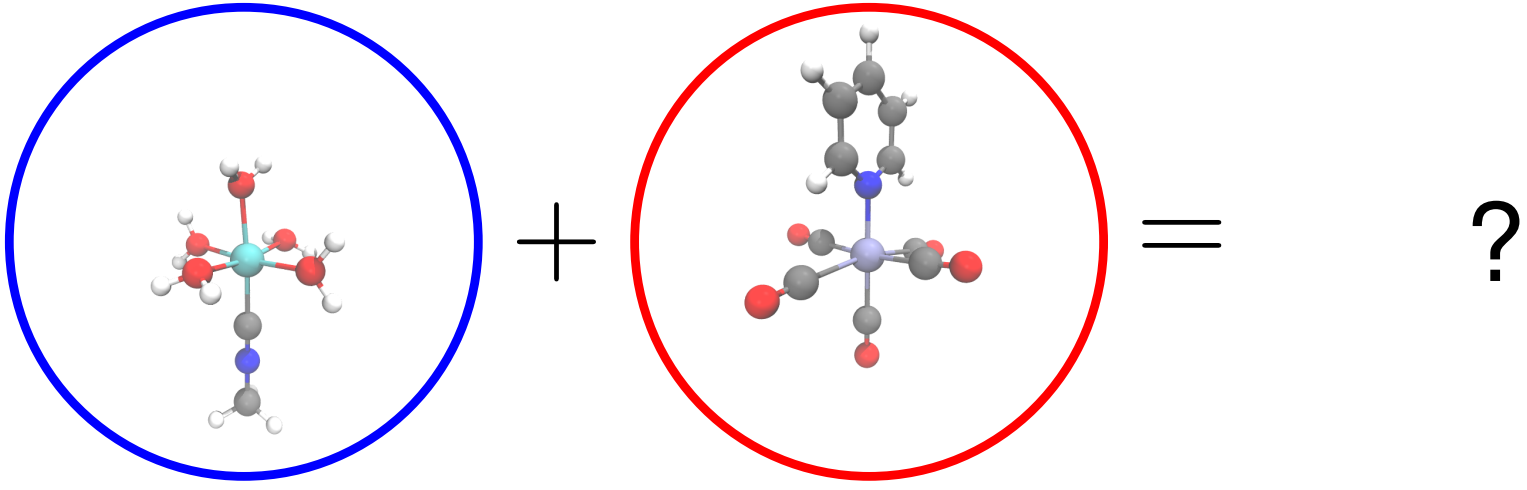


Mapping TM complex space



$$\frac{\quad + \quad}{2} = \; ?$$

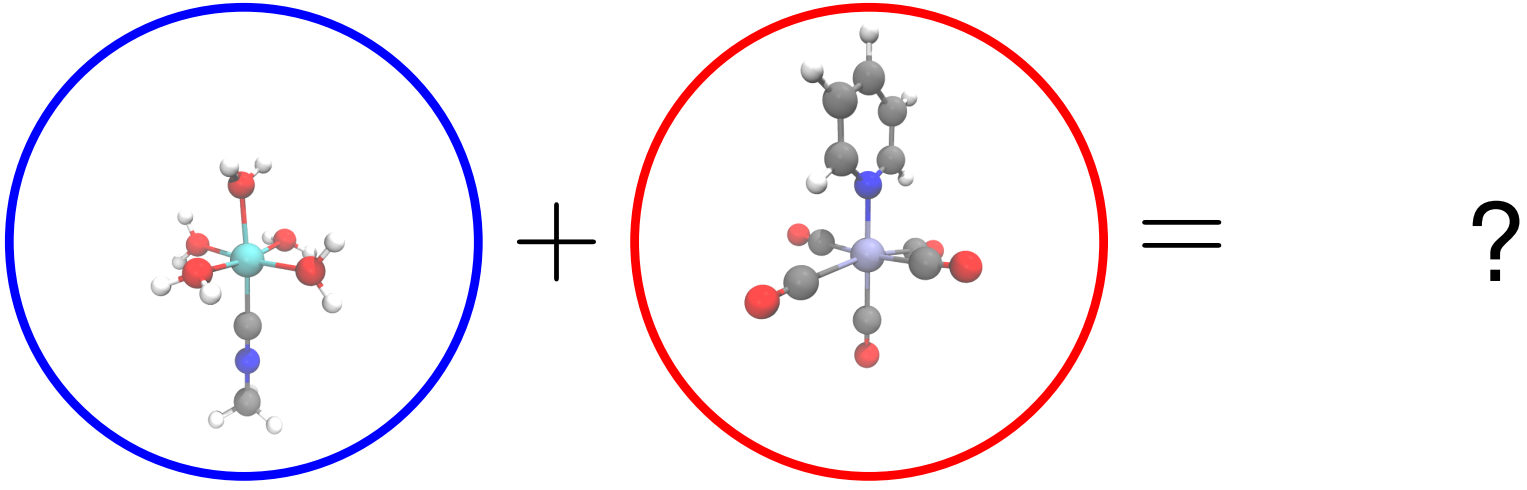
Mapping TM complex space



$$\frac{\quad + \quad = \quad ?}{2}$$

Cr(II) $[H_2O]_5$ [misc] Co(II) $[CO]_5$ [pyr]

Mapping TM complex space

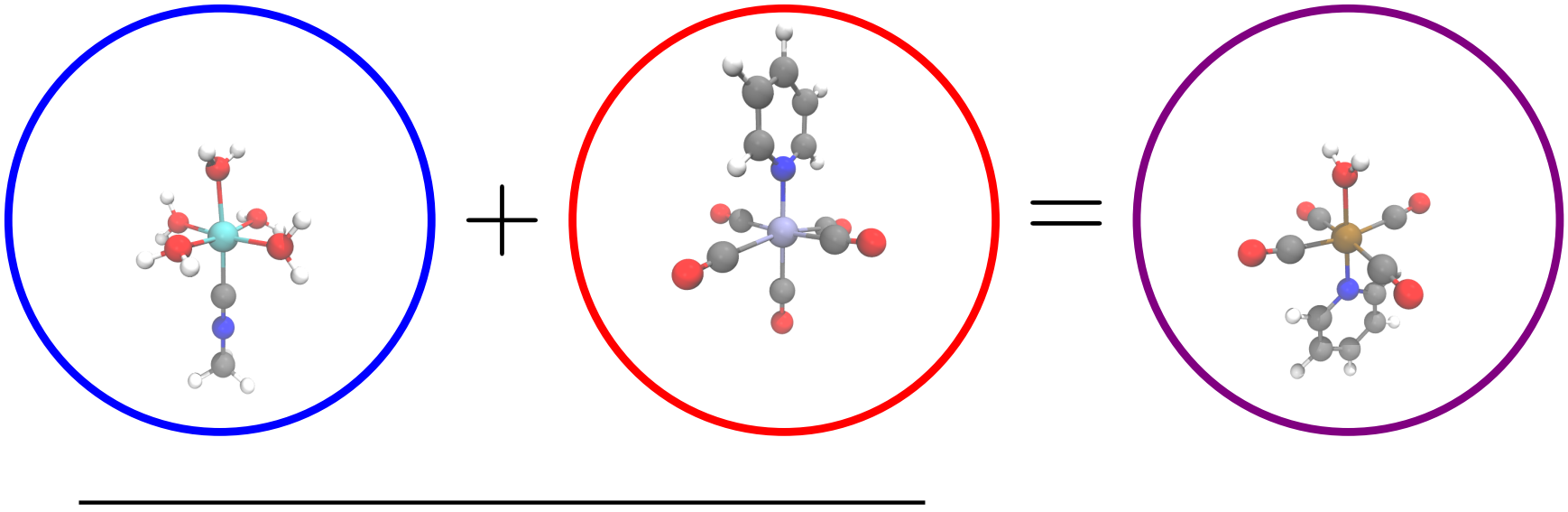


$$\frac{\quad + \quad = \quad ?}{2}$$

$\mathrm{Cr(II)\,[H_2O]_5\,[misc]}$
$\Delta G = 5.3$ eV

$\mathrm{Co(II)\,[CO]_5\,[pyr]}$
$\Delta G = 8.1$ eV

Mapping TM complex space



$$\frac{}{2}$$

$\mathrm{Cr(II)\,[H_2O]_5\,[misc]}$
$\Delta G = 5.3$ eV

$\mathrm{Co(II)\,[CO]_5\,[pyr]}$
$\Delta G = 8.1$ eV

$\mathrm{Fe(II)\,[CO]_4\,[pyr]\,[water]}$
$\Delta G = 7.8$ eV

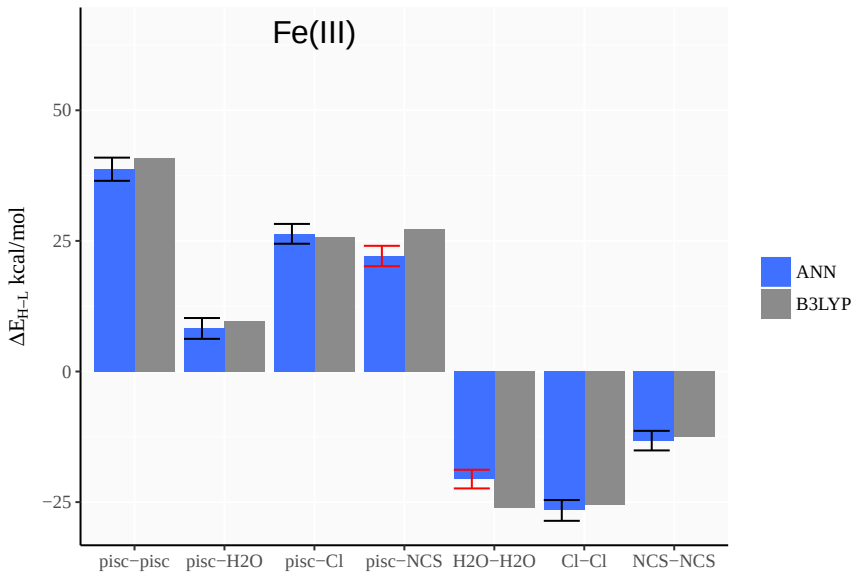
Model transferability

Test-set performance is not necessarily a good metric for general transferability[2]:

[2] Janet, J.P., and Kulik, H.J. *Chem. Sci.*, 2017, 8, 5137-5152.

Model transferability

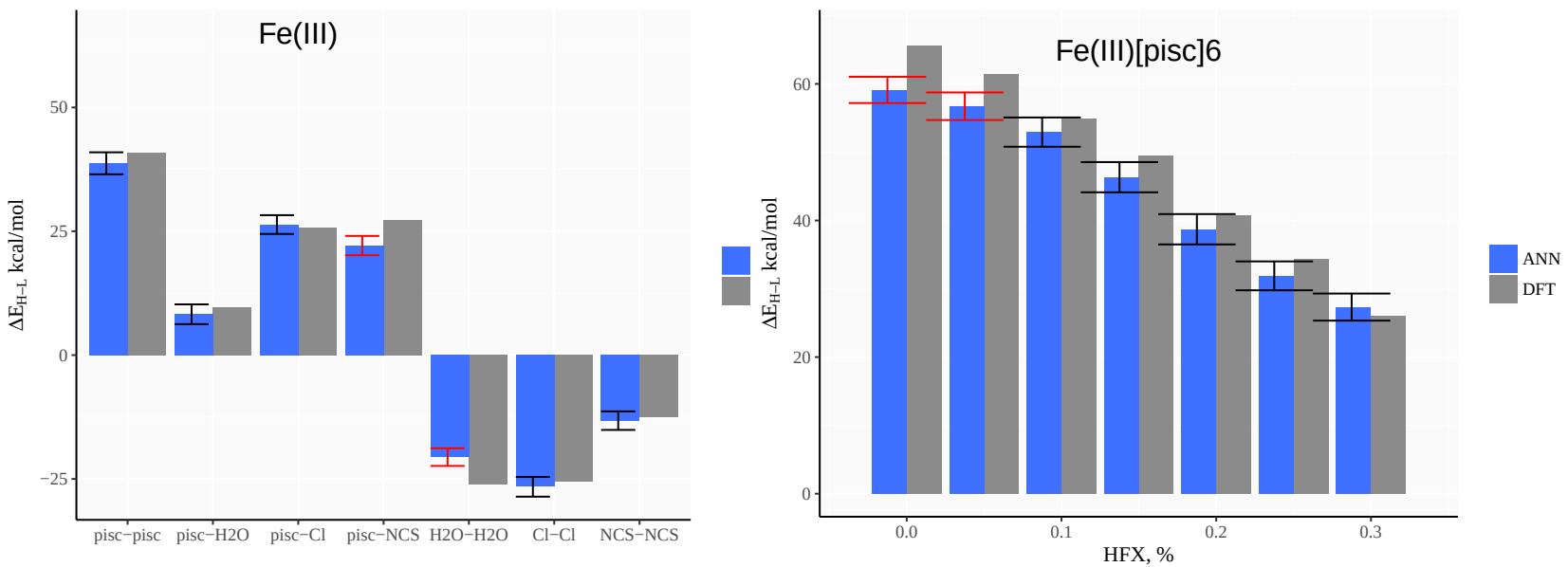
Test-set performance is not necessarily a good metric for general transferability[2]:



[2]Janet, J.P., and Kulik, H.J. *Chem. Sci.*, 2017, 8, 5137-5152.

Model transferability

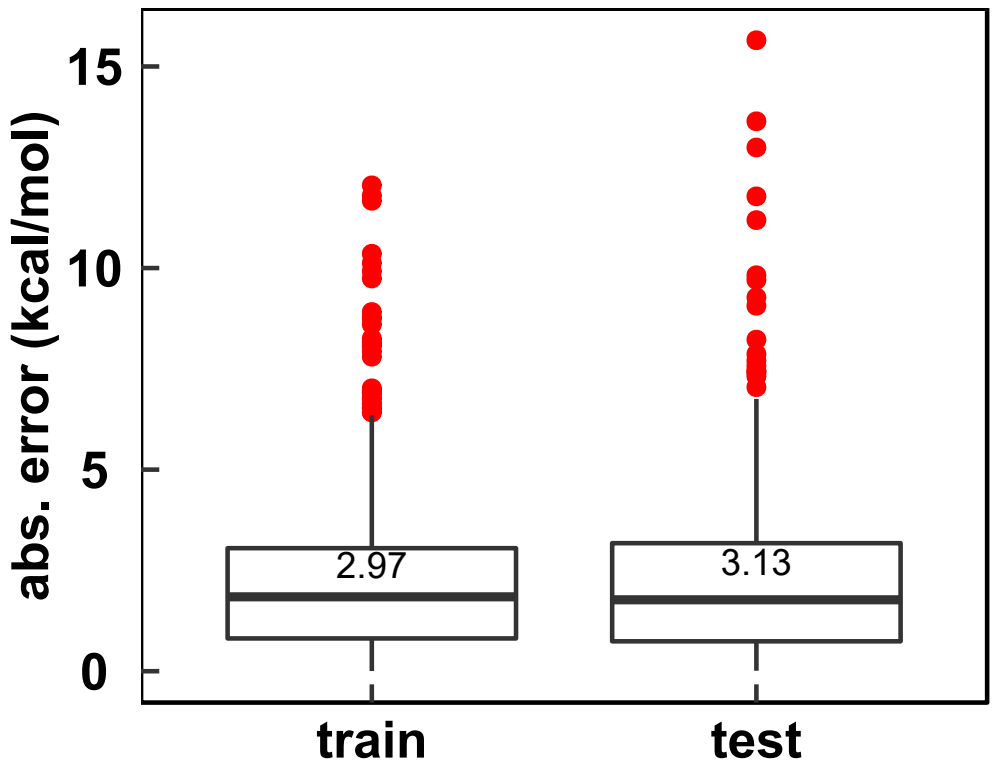
Test-set performance is not necessarily a good metric for general transferability[2]:



[2] Janet, J.P., and Kulik, H.J. *Chem. Sci.*, 2017, 8, 5137-5152.

Model transferability

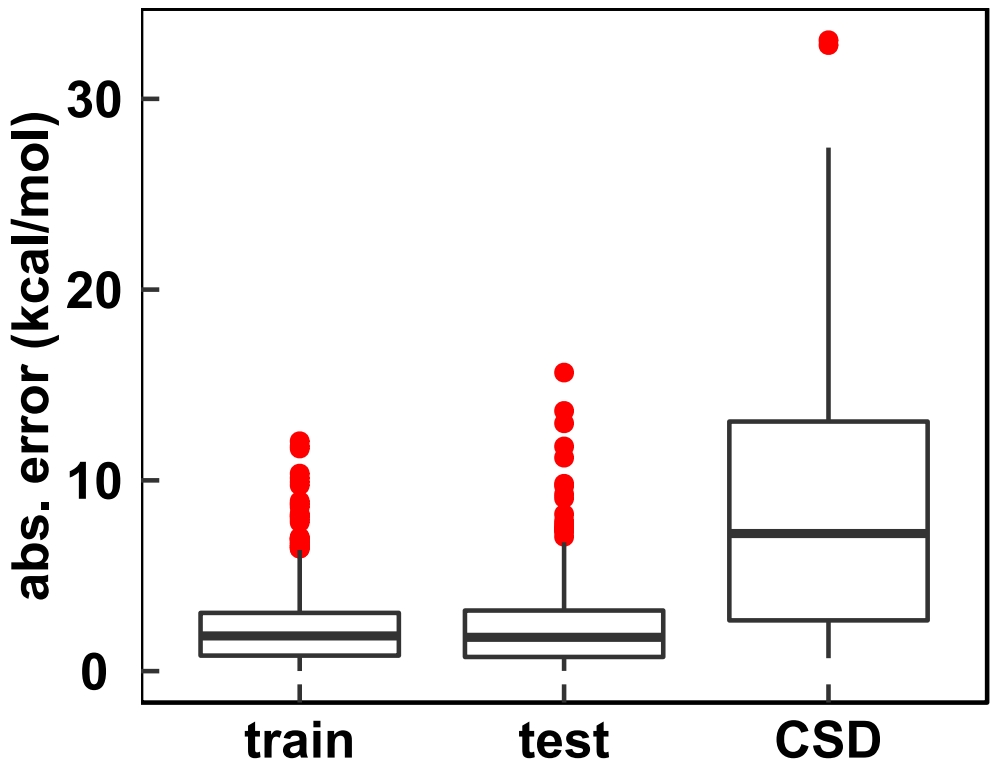
Test-set performance is not necessarily a good metric for general transferability[2]:



[2]Janet, J.P., and Kulik, H.J. *Chem. Sci.*, 2017, 8, 5137-5152.

Model transferability

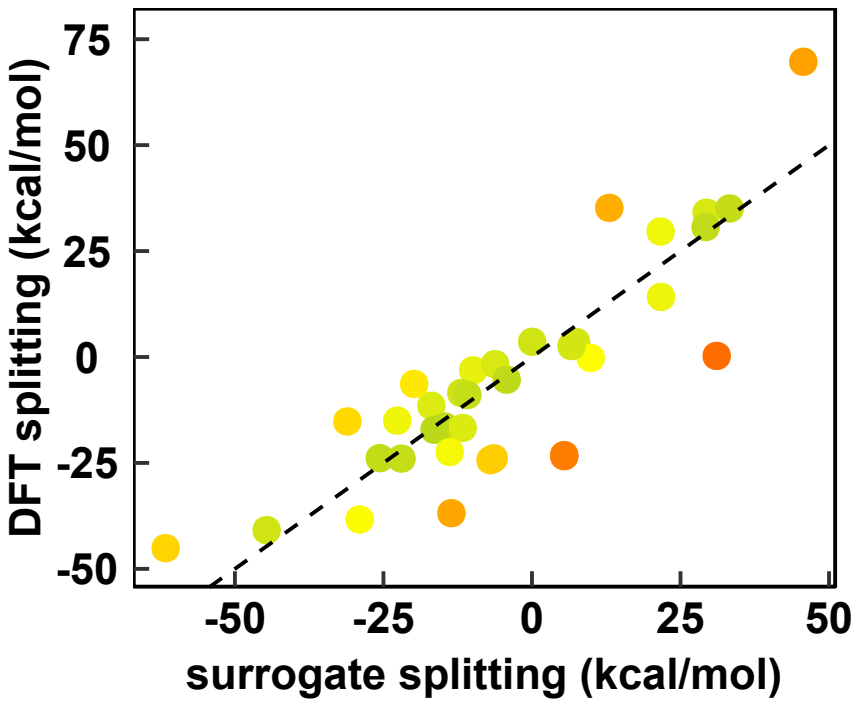
Test-set performance is not necessarily a good metric for general transferability[2]:



[2] Janet, J.P., and Kulik, H.J. *Chem. Sci.*, 2017, 8, 5137-5152.

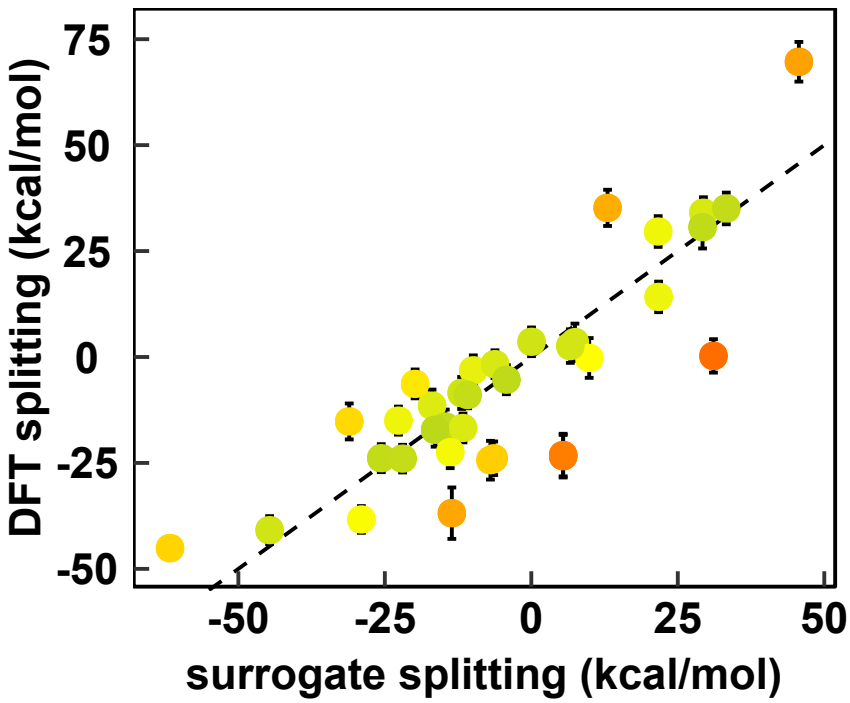
Model transferability

Uncertainty estimates are essential for our surrogate model to explore chemical space:



Model transferability

Uncertainty estimates are essential for our surrogate model to explore chemical space:



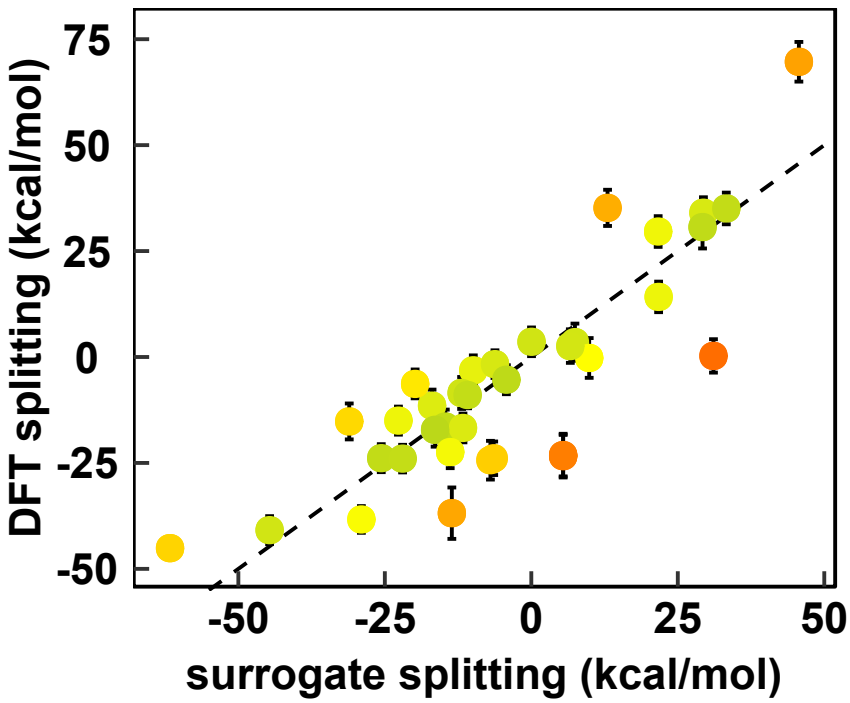
Uncertainty from mc-dropout[1]: ANN model approximates variational inference with GP under some conditions:

$$\mathrm{var}\left(y^* \middle| x^*\right) \approx \frac{1}{J}\sum_j \hat{y}_j^T \hat{y}_j + \tau^{-1}$$

Gal, Y. and Ghahramani, Z., 2016. ICMLR 1050-1059

Model transferability

Uncertainty estimates are essential for our surrogate model to explore chemical space:



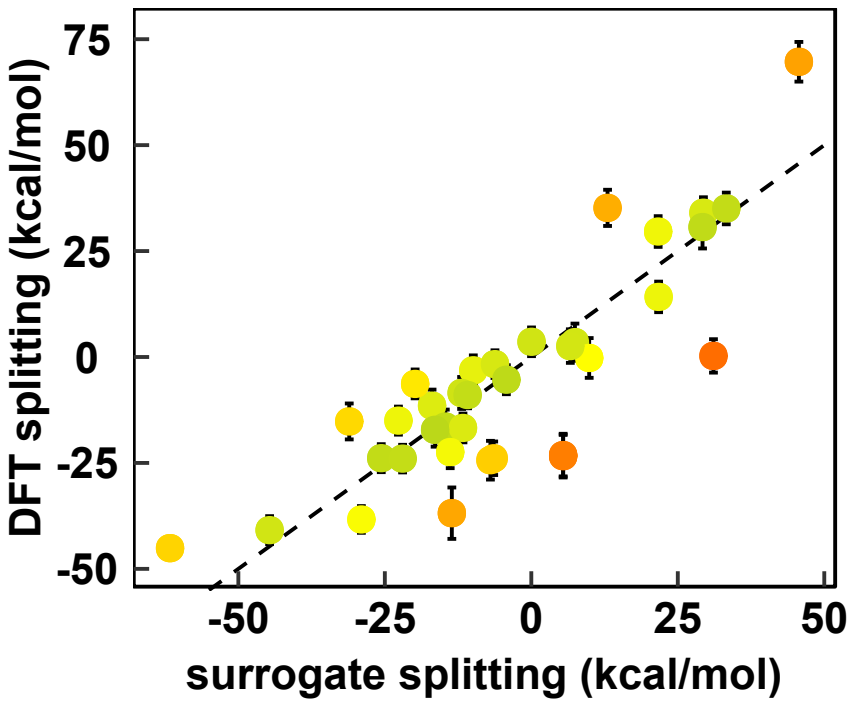
Uncertainty from mc-dropout[1]: ANN model approximates variational inference with GP under some conditions:

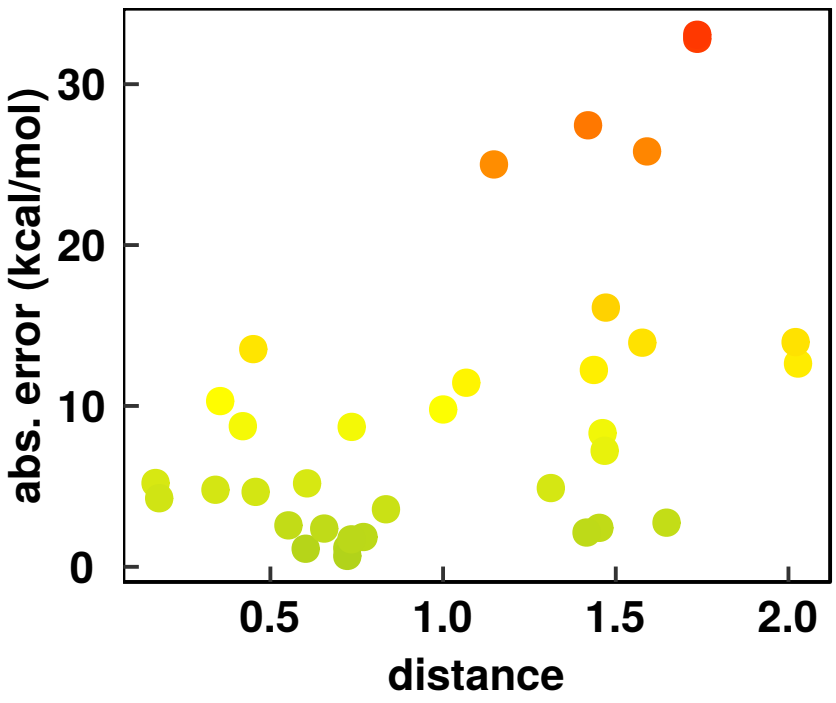
$$\mathrm{var}\left(y^* \middle| x^*\right) \approx \frac{1}{J}\sum_j \hat{y}_j^T \hat{y}_j + \tau^{-1}$$

Gal, Y. and Ghahramani, Z., 2016. ICMLR 1050-1059

Model transferability

Uncertainty estimates are essential for our surrogate model to explore chemical space:





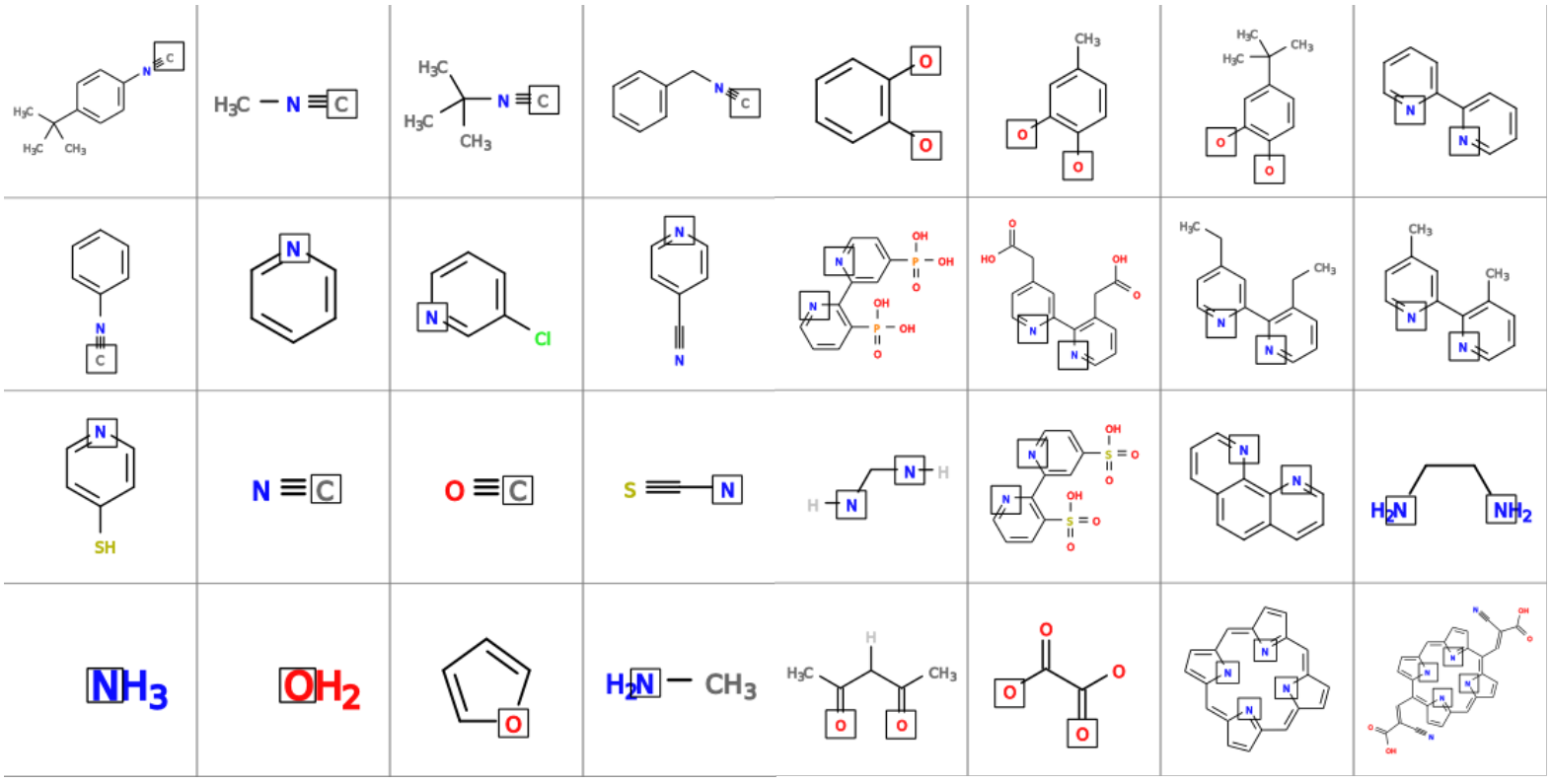
Demonstration

Can we use the ANN model to find new spin-crossover materials, i.e. $\Delta E_{H-L} = 0$?

[3] Janet, J.P., Chan, L. and Kulik, H.J. *J. Phys. Chem. Lett.*, 2018, 9, 5, 1064-1071.

Demonstration

Can we use the ANN model to find new spin-crossover materials, i.e. $\Delta E_{H-L} = 0$? Define a space of 32 ligands, 5 metals and with $\sim$ 5600 possible elements with forced axial/equatorial symmetry[3]:



[3] Janet, J.P., Chan, L. and Kulik, H.J. *J. Phys. Chem. Lett.*, 2018, 9, 5, 1064-1071.

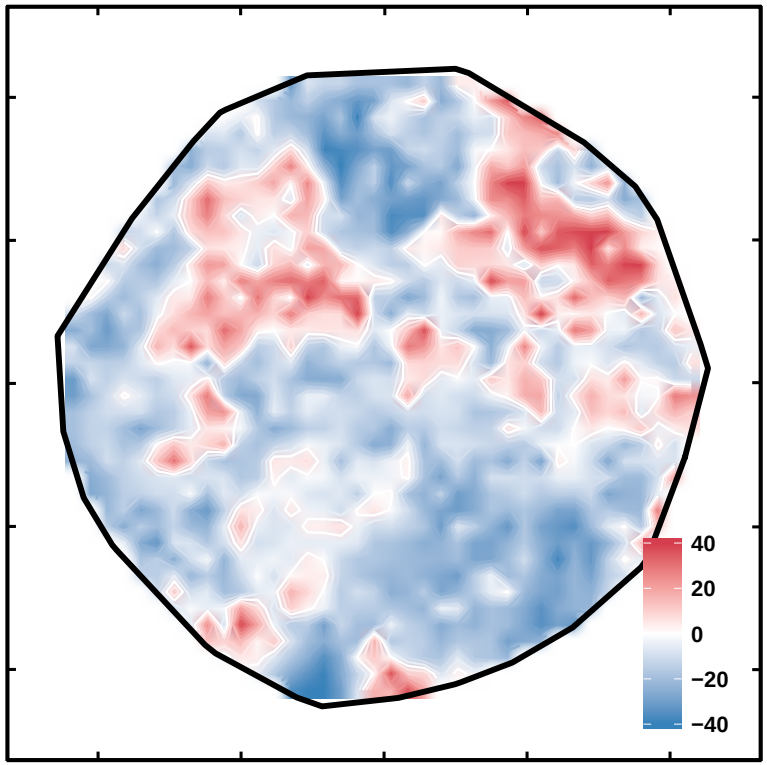
Demonstration

ANN is trained on 14 of these ligands, covers only **2%** of the design space.

[4]Maaten, L., & Hinton, G., 2008. J. Mach. Learn. Res. 2579-2605.

Demonstration

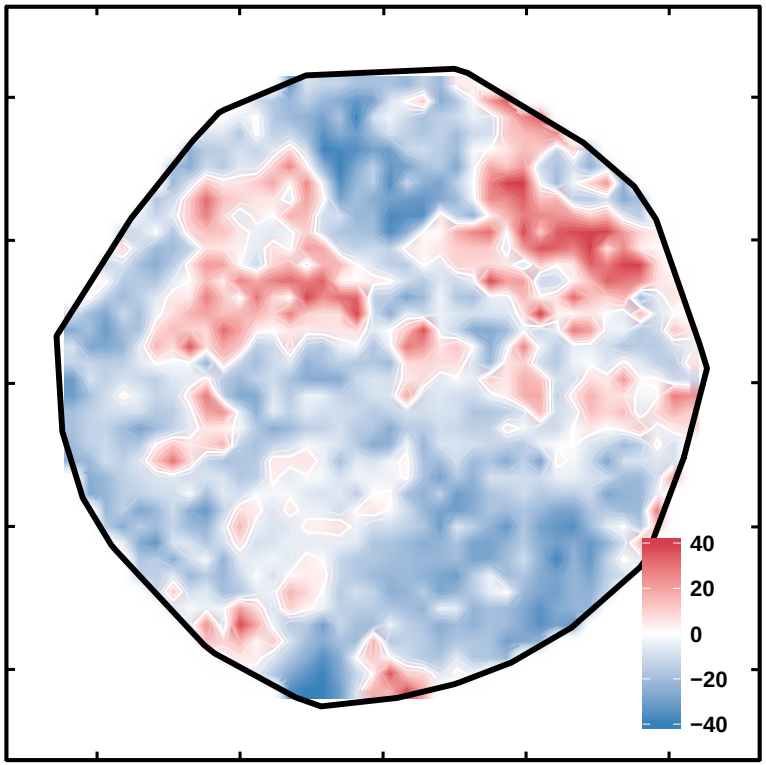
ANN is trained on 14 of these ligands, covers only **2%** of the design space. We can visualize the design space using t-SNE[4]:

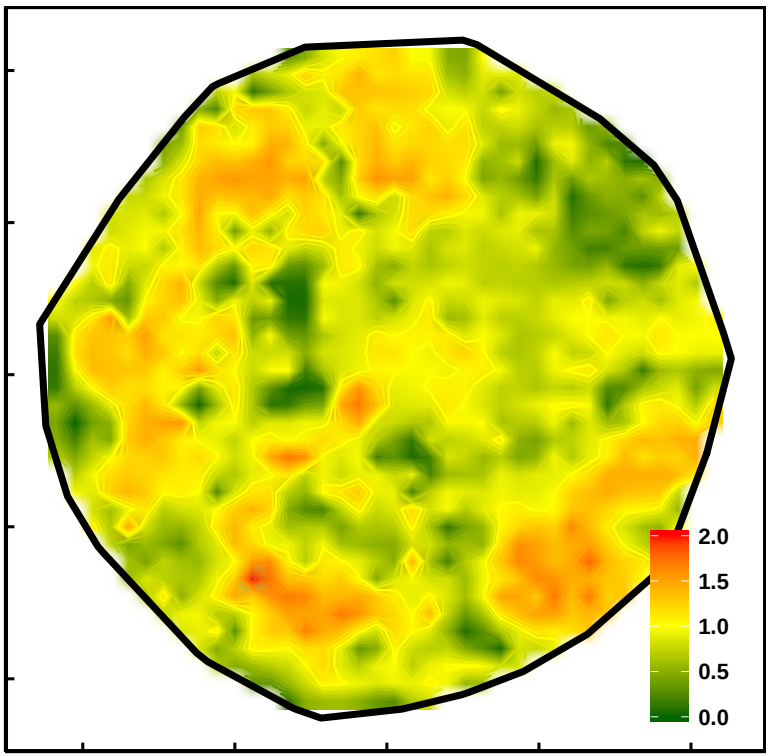


[4]Maaten, L., & Hinton, G., 2008. J. Mach. Learn. Res. 2579-2605.

Demonstration

ANN is trained on 14 of these ligands, covers only **2%** of the design space. We can visualize the design space using t-SNE[4]:





[4]Maaten, L., & Hinton, G., 2008. J. Mach. Learn. Res. 2579-2605.

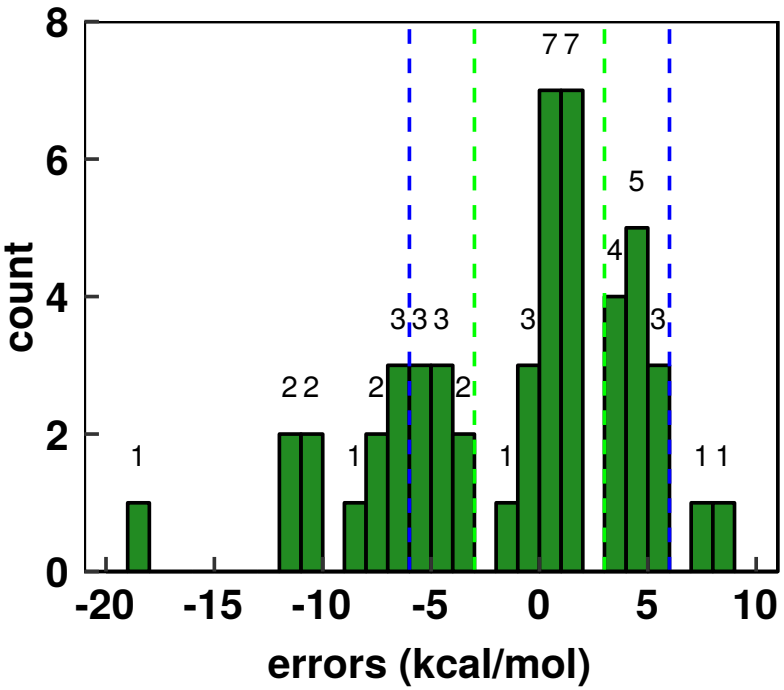
How accurate are we?

Test 51 leads from ANN with DFT[5]:

[5] Janet, J.P., Chan, L. and Kulik, H.J. *J. Phys. Chem. Lett.*, 2018, 9, 5, 1064-1071.

How accurate are we?

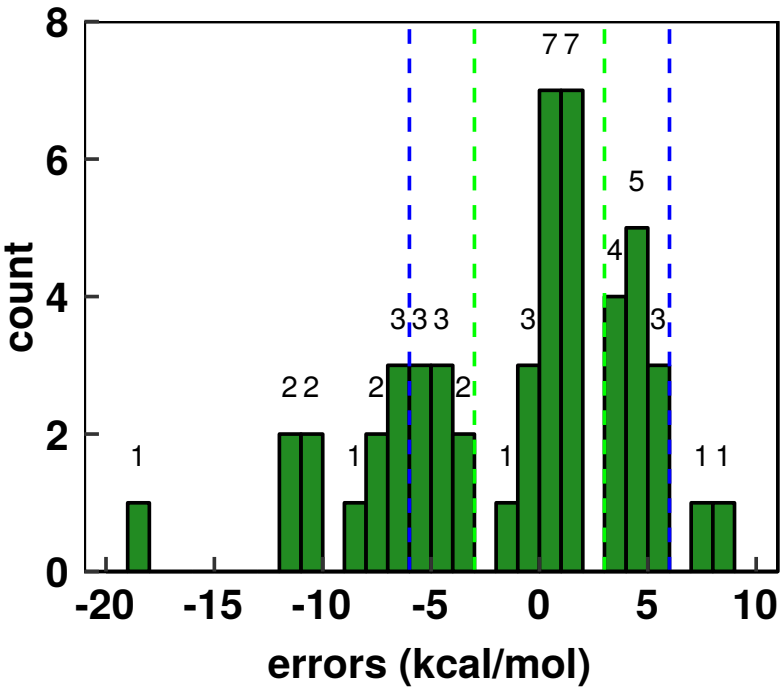
Test 51 leads from ANN with DFT[5]:

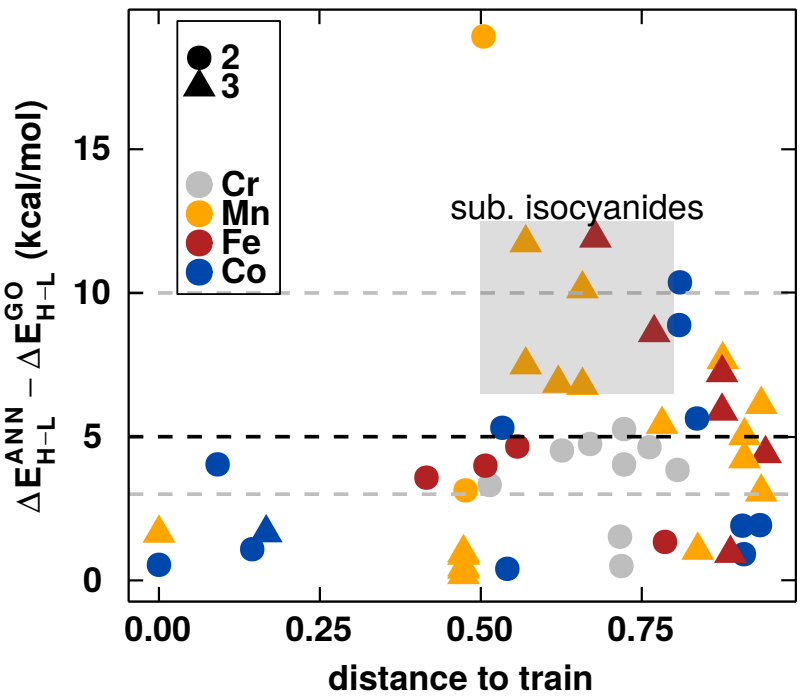


[5]Janet, J.P., Chan, L. and Kulik, H.J. *J. Phys. Chem. Lett.*, 2018, 9, 5, 1064-1071.

How accurate are we?

Test 51 leads from ANN with DFT[5]:





[5]Janet, J.P., Chan, L. and Kulik, H.J. *J. Phys. Chem. Lett.*, 2018, 9, 5, 1064-1071.

Conclusions

- choice of molecular representation is important
- different properties depend non-equally on features
- feature-space geometry can provide insight into model reliability
- imbuing ‘chemical intuition’ to descriptor construction can drastically improve learning
- conversely, feature selection can contribute to understanding systems

Acknowledgments

Thanks to the Kulik group and funding partners: