Mapping transition metal chemical space for machine learning models

Jon Paul Janet [1] Heather Kulik [1]

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

2017 AIChE Annual Meeting
Minneapolis, MN

11.01.17

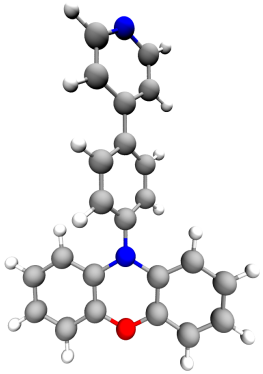
Data-driven molecular design

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

Data-driven molecular design

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

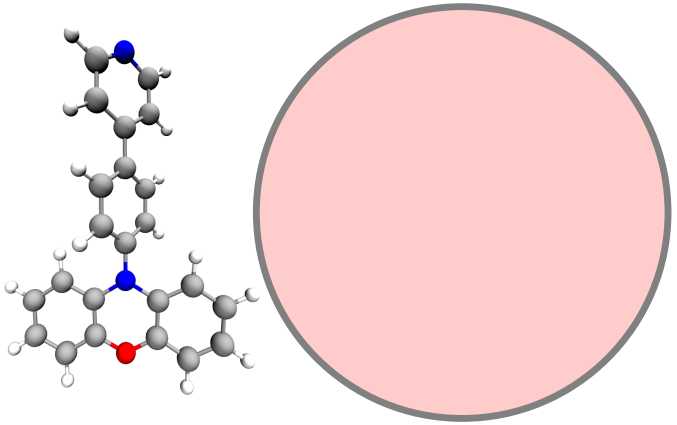
Machine learning is transforming how we design new materials...



Data-driven molecular design

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

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

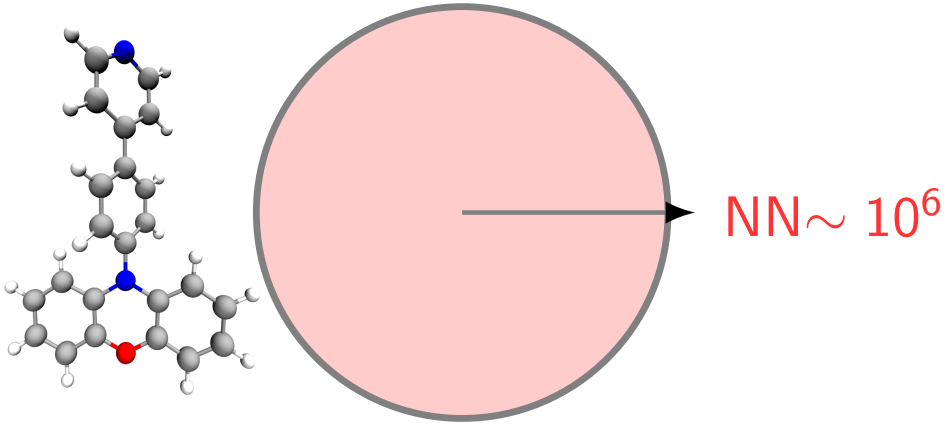


OLED chemical space

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.

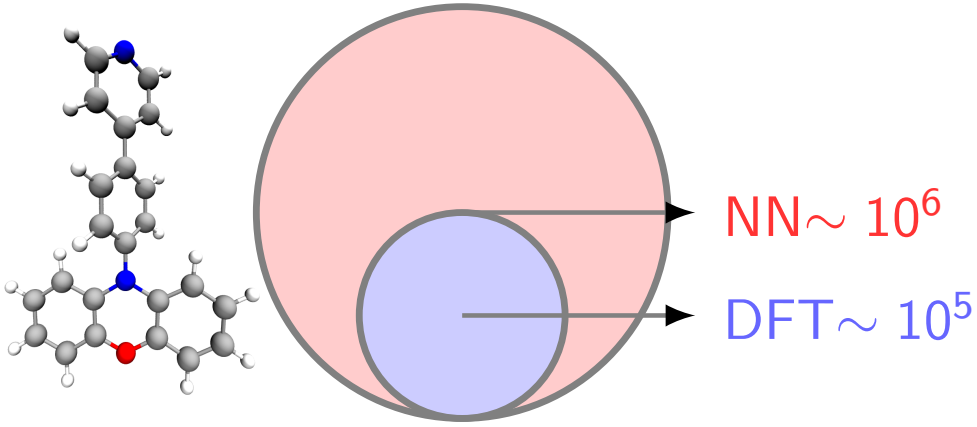


OLED chemical space

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.

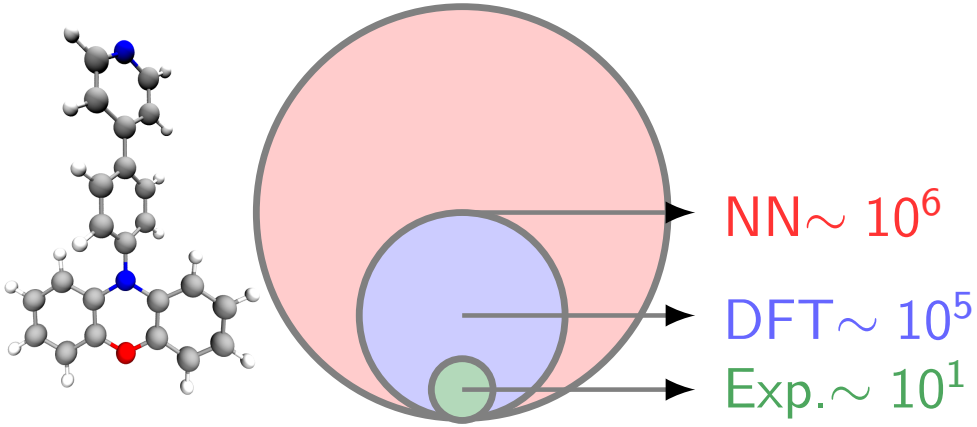


OLED chemical space

Data-driven molecular design

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

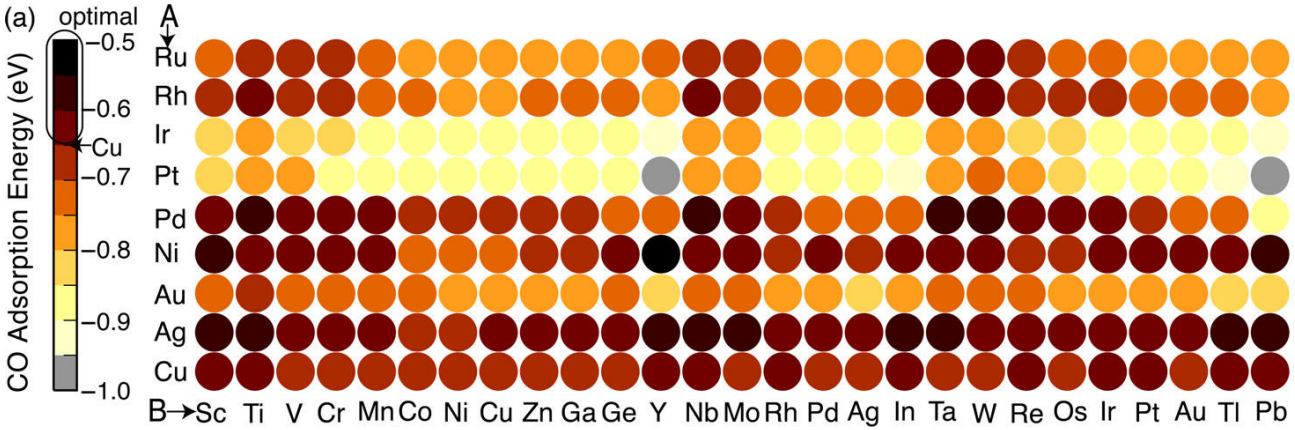
Gomez-Bombarelli, R. *et al.*. *Nat. Mater.*, 15(10):1120-1127, 2016.



OLED chemical space

Data-driven molecular design

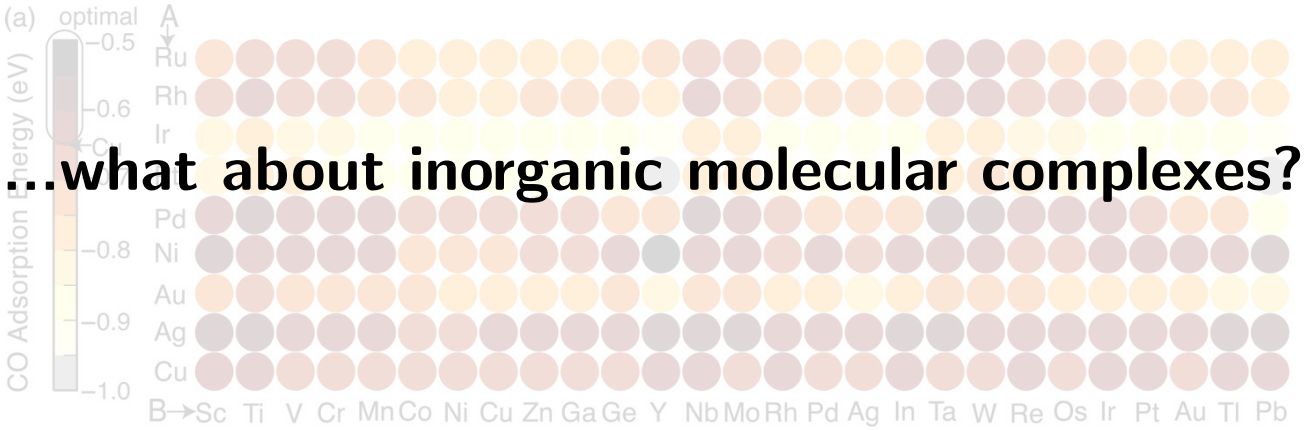
Machine learning is transforming how we design new materials...



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...

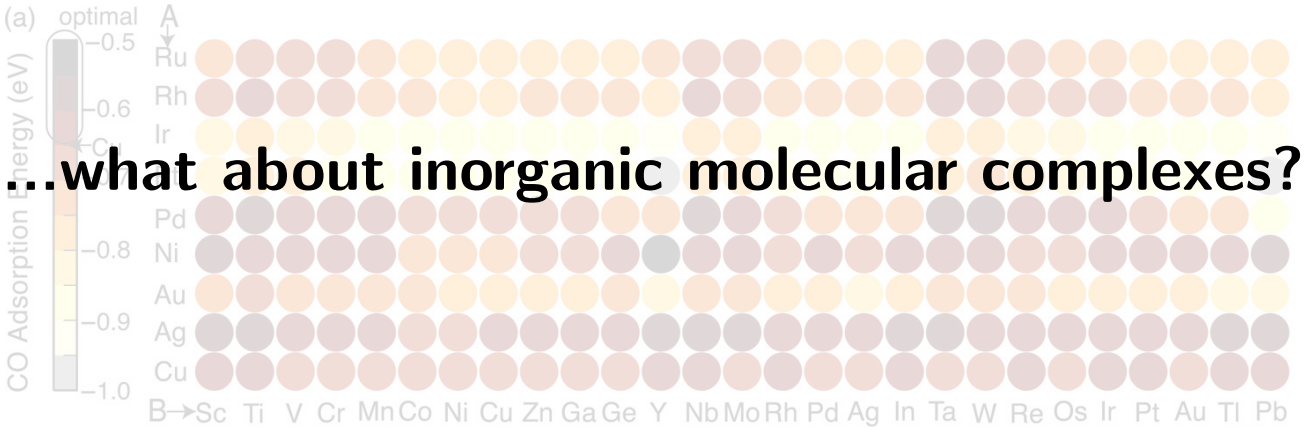


...what about inorganic molecular complexes?

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

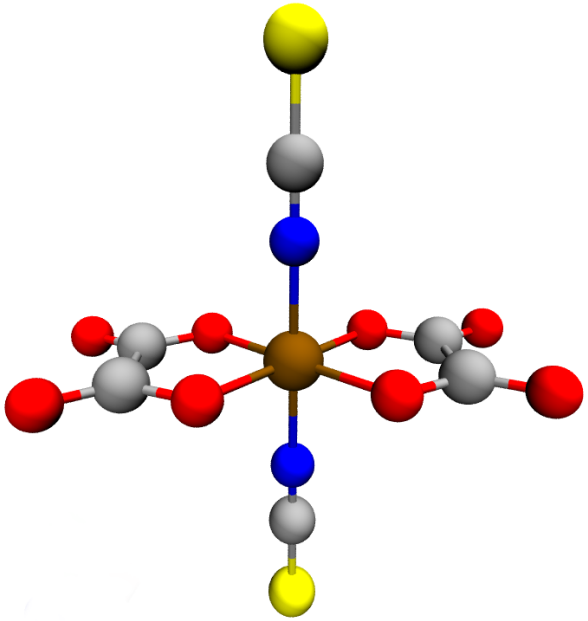
Data-driven molecular design

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



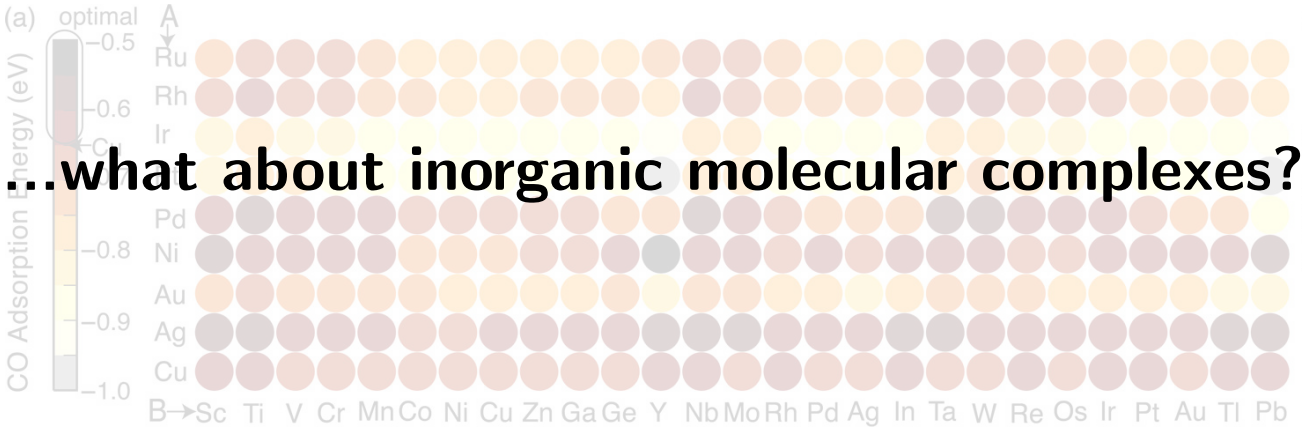
...what about inorganic molecular complexes?

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



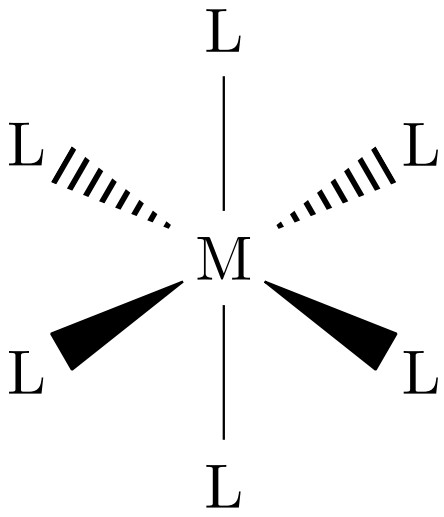
Data-driven molecular design

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



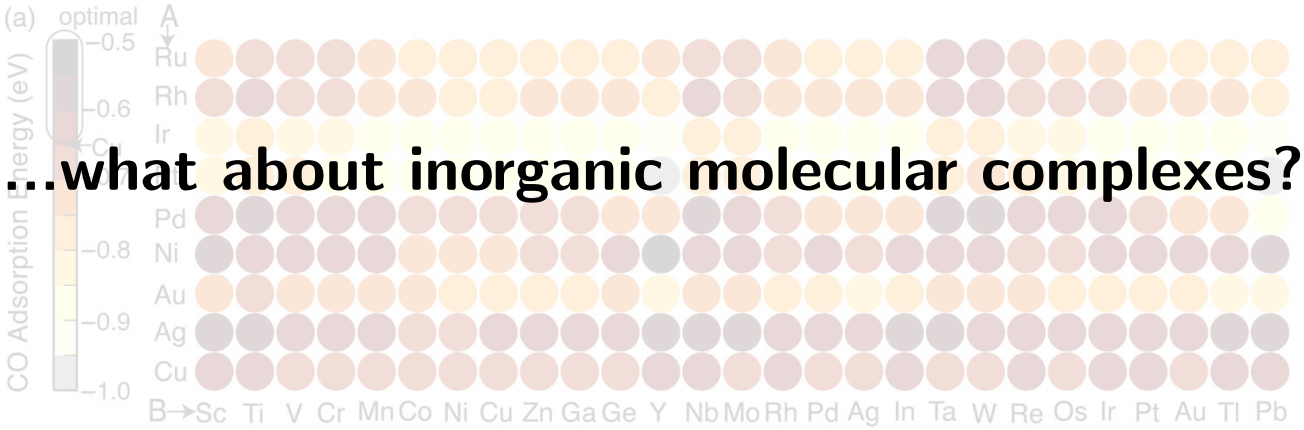
...what about inorganic molecular complexes?

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



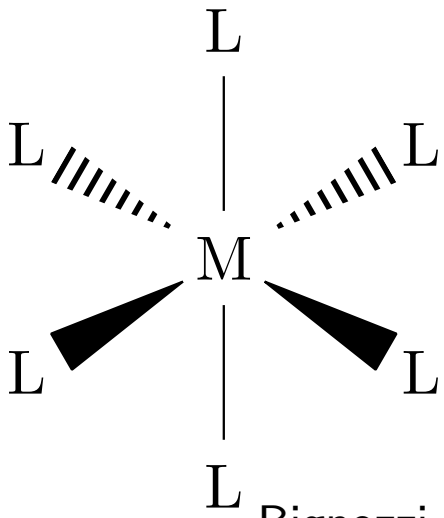
Data-driven molecular design

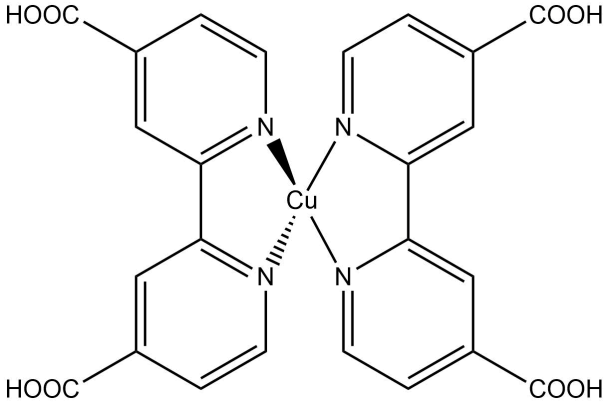
Machine learning is transforming how we design new materials...



...what about inorganic molecular complexes?

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

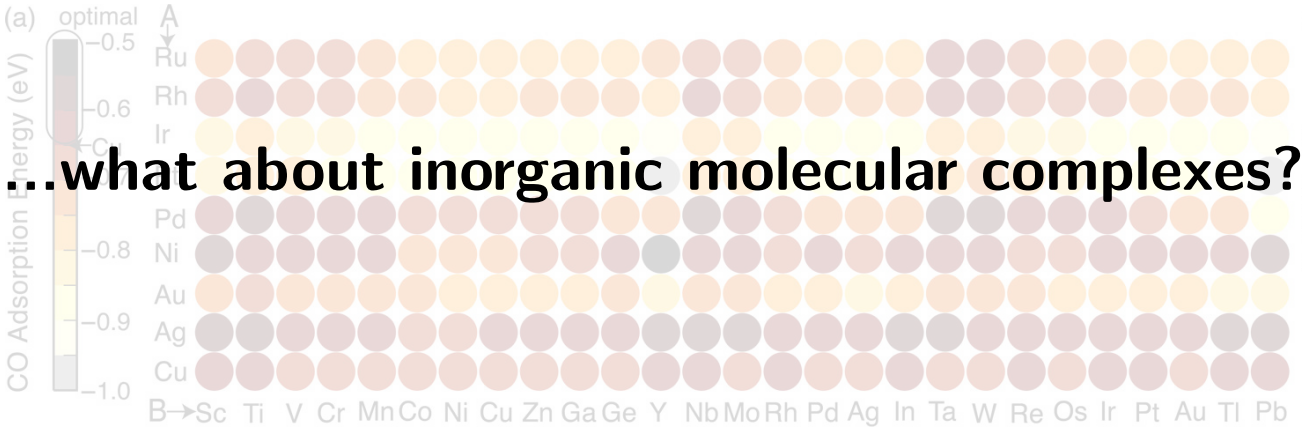




Bignozzi, C. *et al.* *Coord. Chem. Rev.*, 257(9), 2013.

Data-driven molecular design

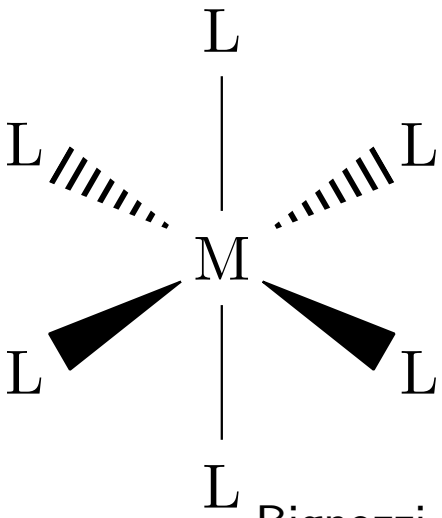
Machine learning is transforming how we design new materials...

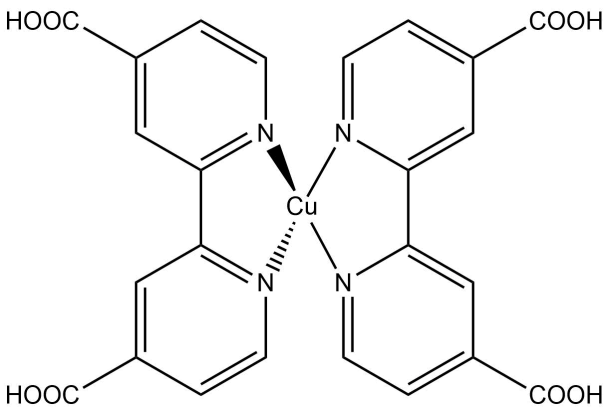


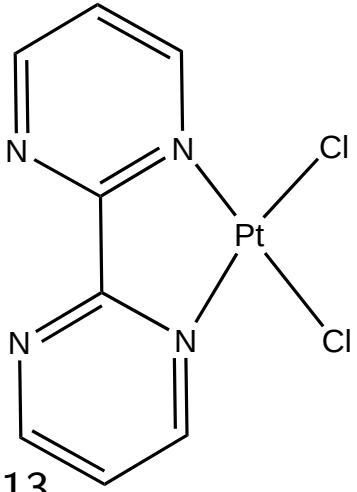
...what about inorganic molecular complexes?

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.

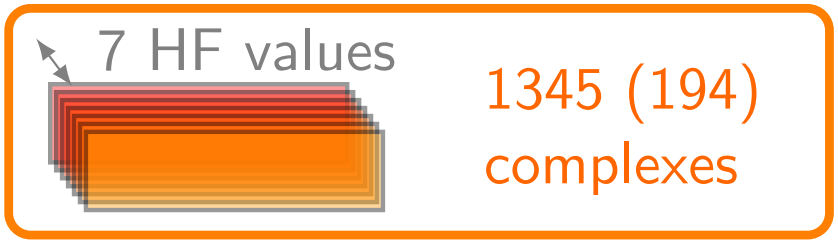
ML for TM complexes

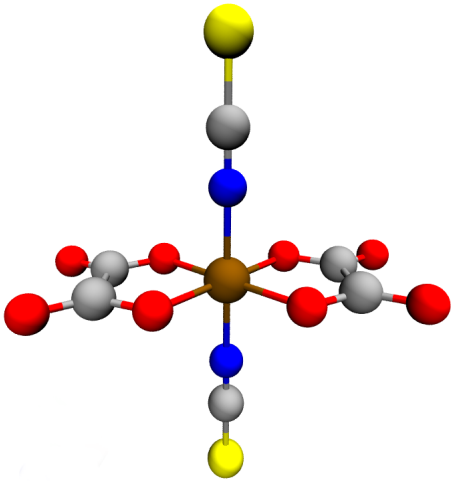
Standard tools do not work well for TM complexes[1]:

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

ML for TM complexes

Standard tools do not work well for TM complexes[1]:

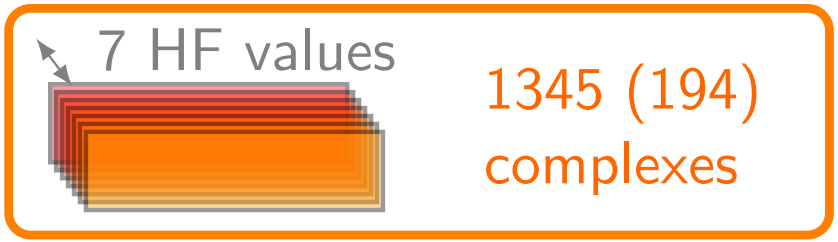


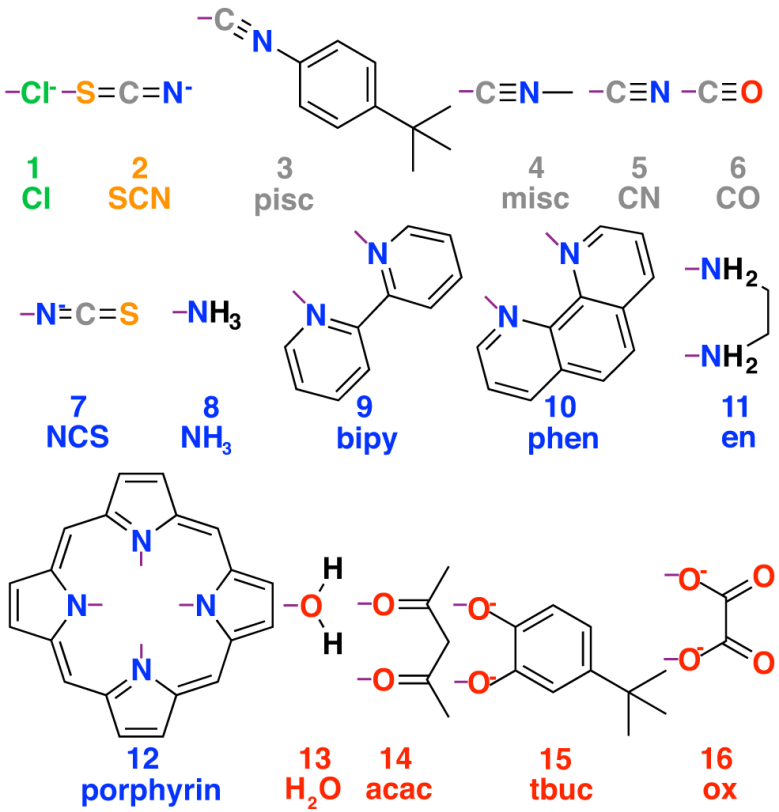


[1] Janet, J.P., and Kulik, H.J. *Chemical Science*, 2017, 8, 5137-5152.

ML for TM complexes

Standard tools do not work well for TM complexes[1]:

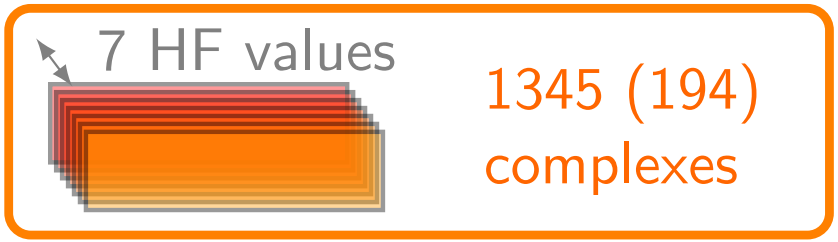


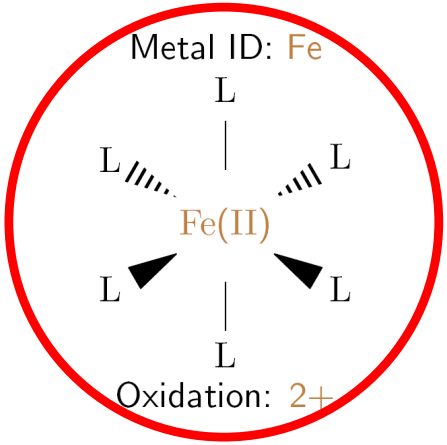


[1] Janet, J.P., and Kulik, H.J. *Chemical Science*, 2017, 8, 5137-5152.

ML for TM complexes

Standard tools do not work well for TM complexes[1]:



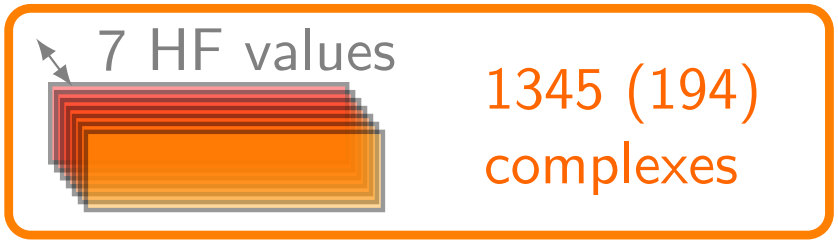


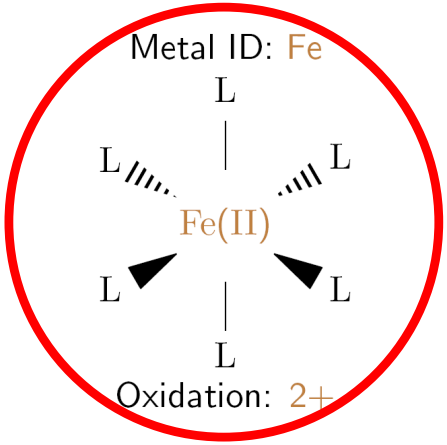
metal properties

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

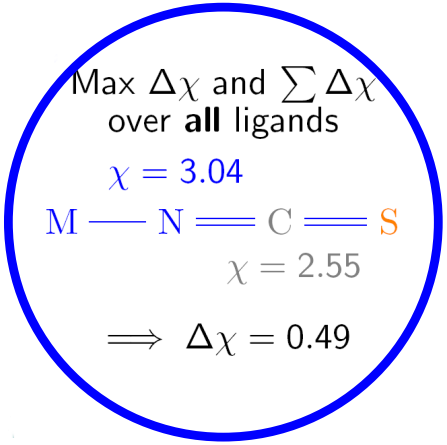
ML for TM complexes

Standard tools do not work well for TM complexes[1]:





metal properties

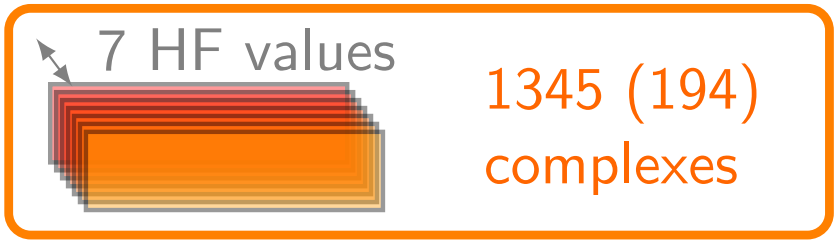


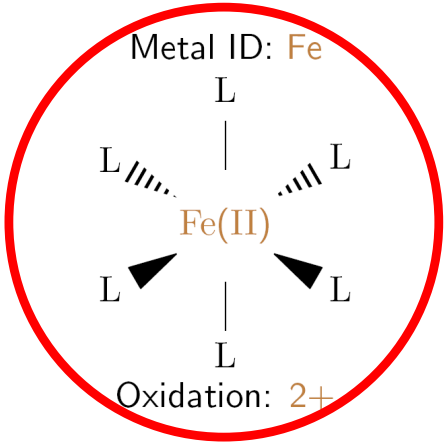
proximal properties

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

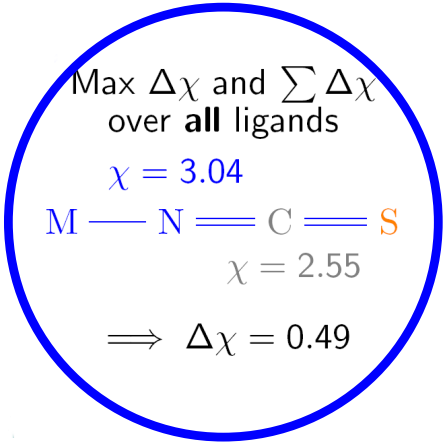
ML for TM complexes

Standard tools do not work well for TM complexes[1]:

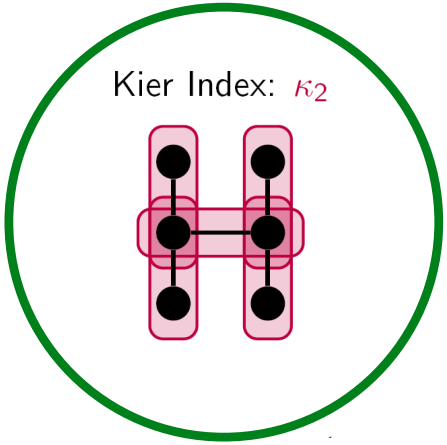




metal properties



proximal properties

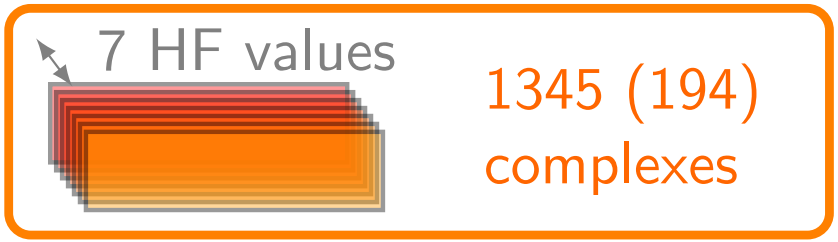


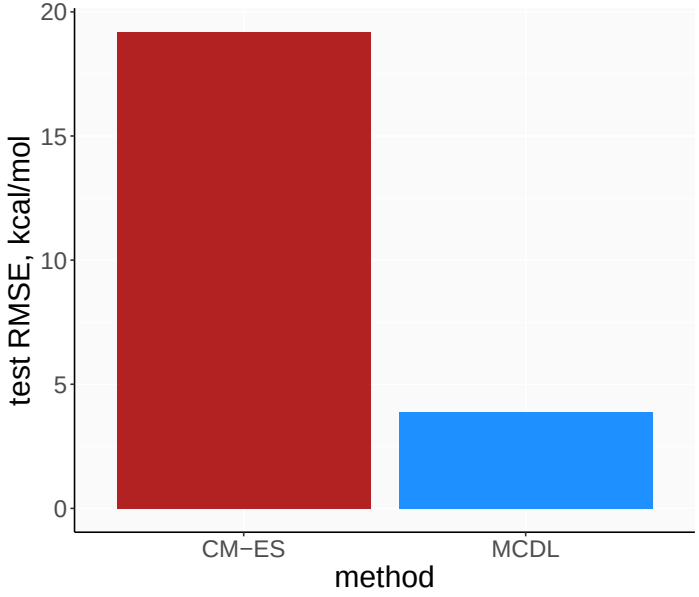
ligand properties

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

ML for TM complexes

Standard tools do not work well for TM complexes[1]:

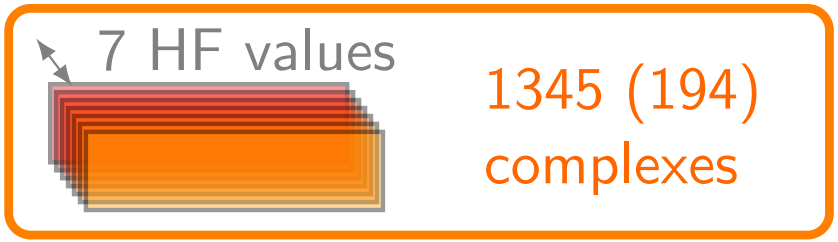


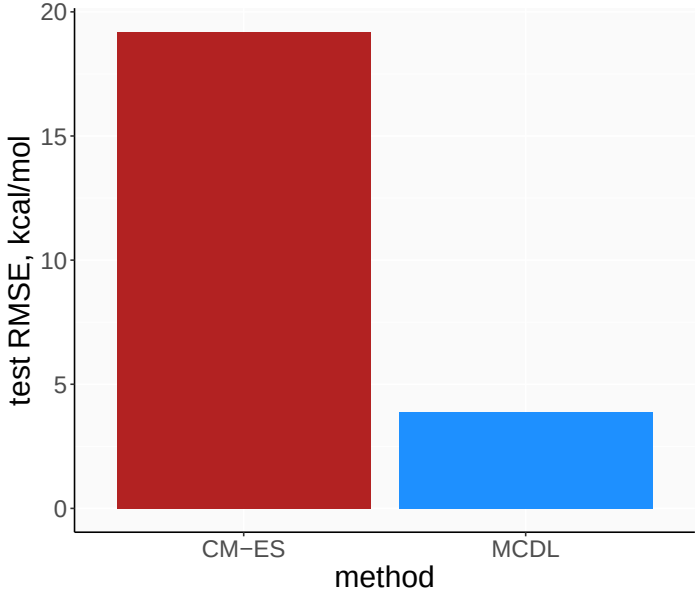


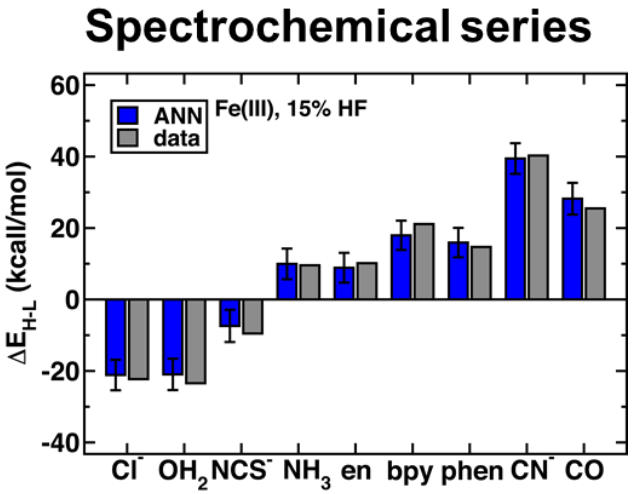
[1] Janet, J.P., and Kulik, H.J. *Chemical Science*, 2017, 8, 5137-5152.

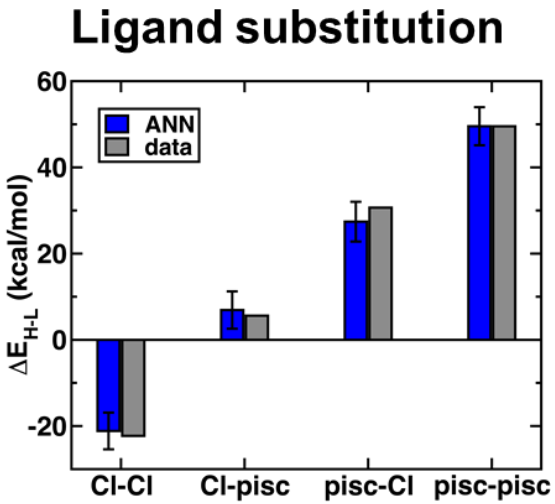
ML for TM complexes

Standard tools do not work well for TM complexes[1]:





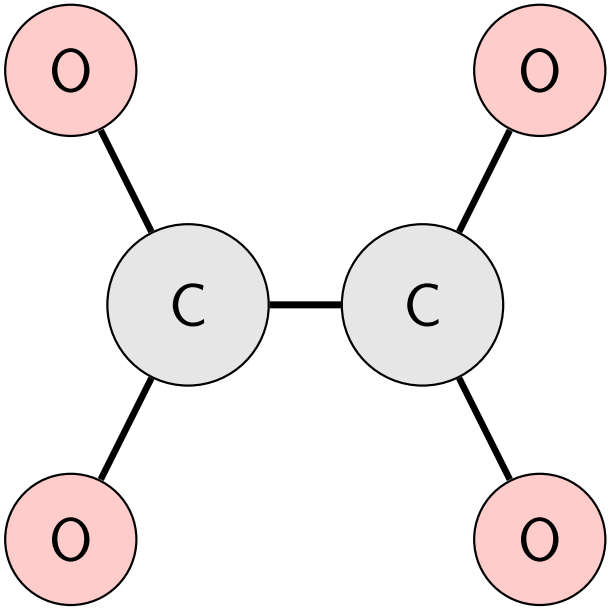




[1] Janet, J.P., and Kulik, H.J. *Chemical Science*, 2017, 8, 5137-5152.

RAC descriptors

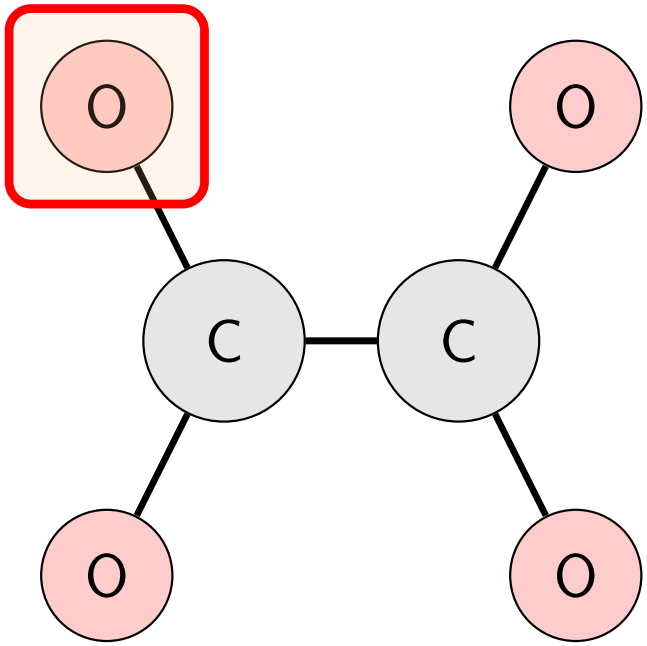
RACs based on autocorrelations[2]



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

RAC descriptors

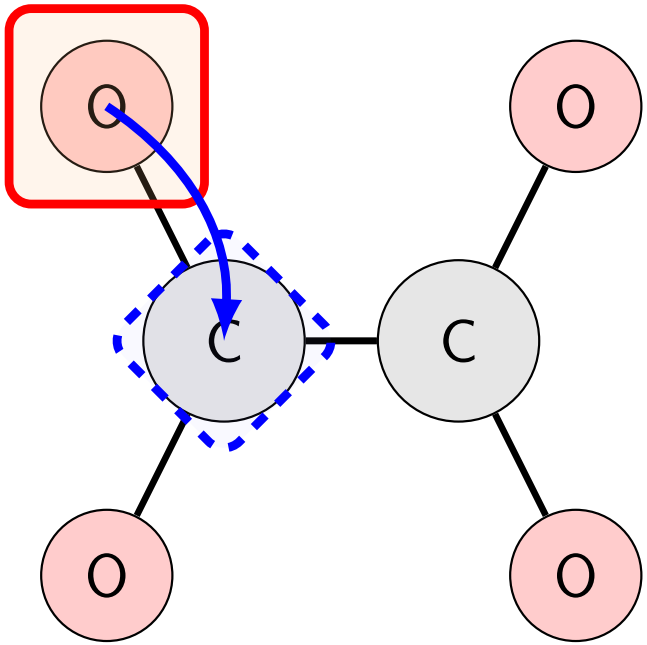
RACs based on autocorrelations[2]



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

RAC descriptors

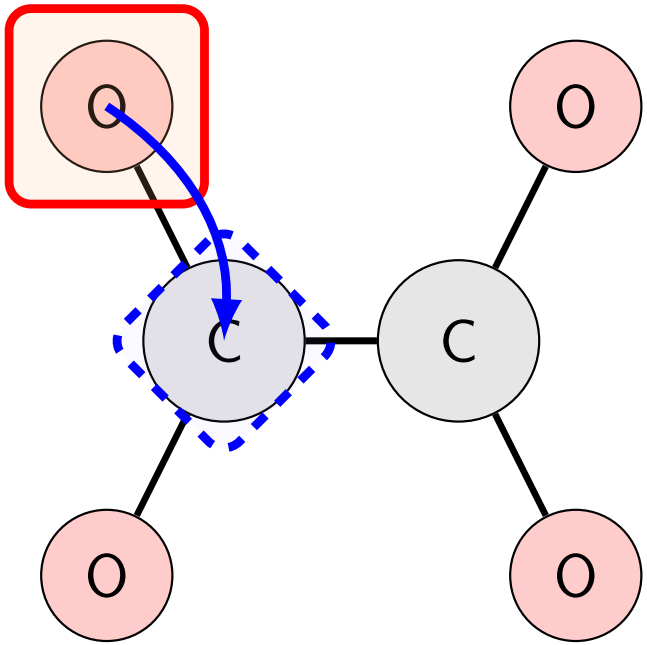
RACs based on autocorrelations[2]



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

RAC descriptors

RACs based on autocorrelations[2]

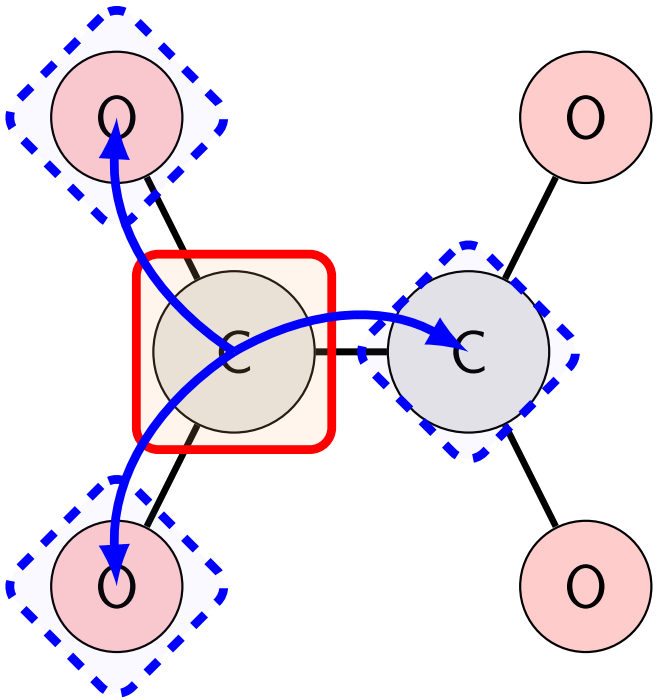


$$d_1 : \sum_{O,C} Z_O Z_C = 48$$

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

RAC descriptors

RACs based on autocorrelations[2]

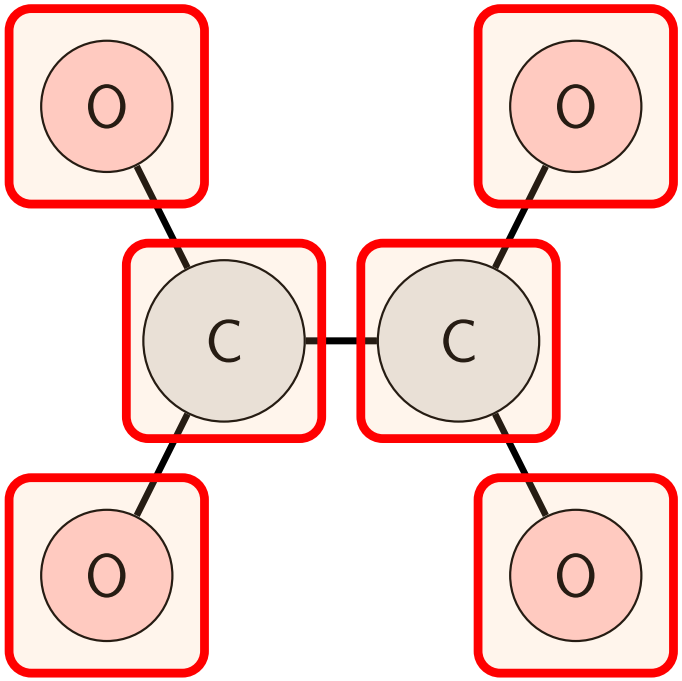


$$d_1 : 48 + \sum_{C,O} Z_O Z_C = 144 + 48$$

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

RAC descriptors

RACs based on autocorrelations[2]

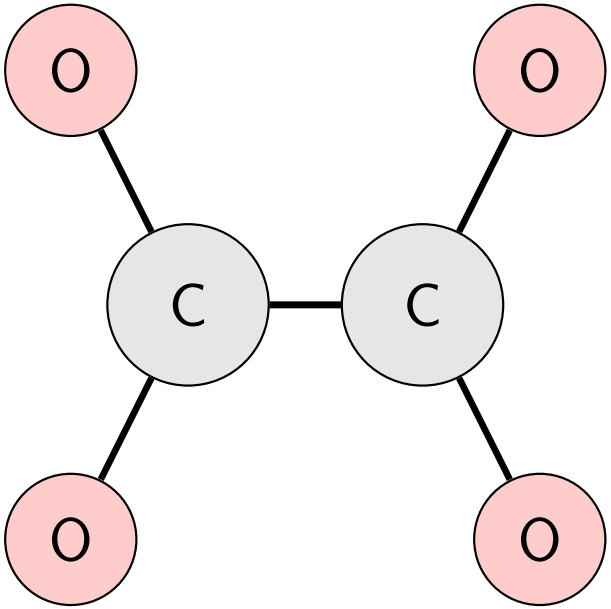


$$d_1 : \sum_i \sum_j Z_i Z_j \delta(d_{i,j}, 1)$$

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

RAC descriptors

RACs based on autocorrelations[2]

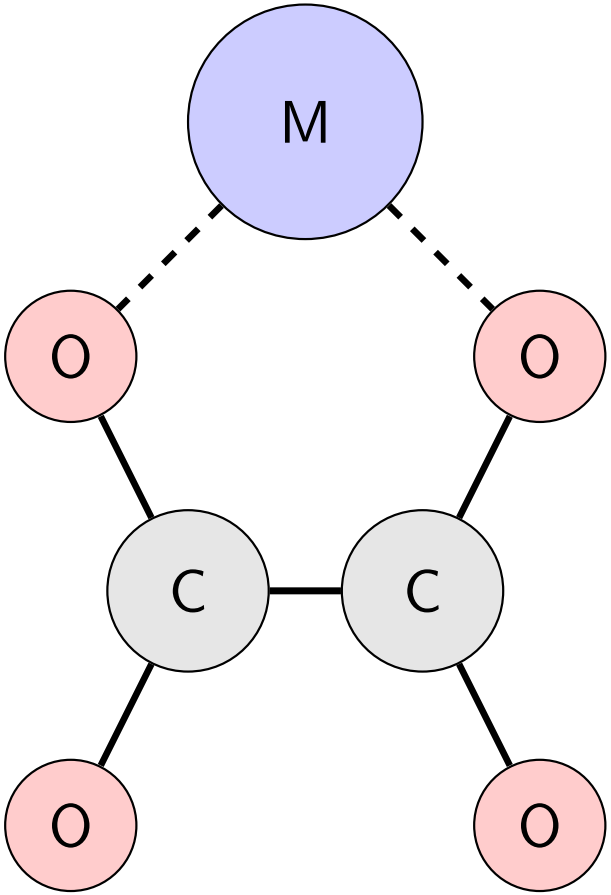


$$d_x : \sum_i \sum_j Z_i Z_j \delta(d_{ij}, x)$$

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

RAC descriptors

RACs based on autocorrelations[2]

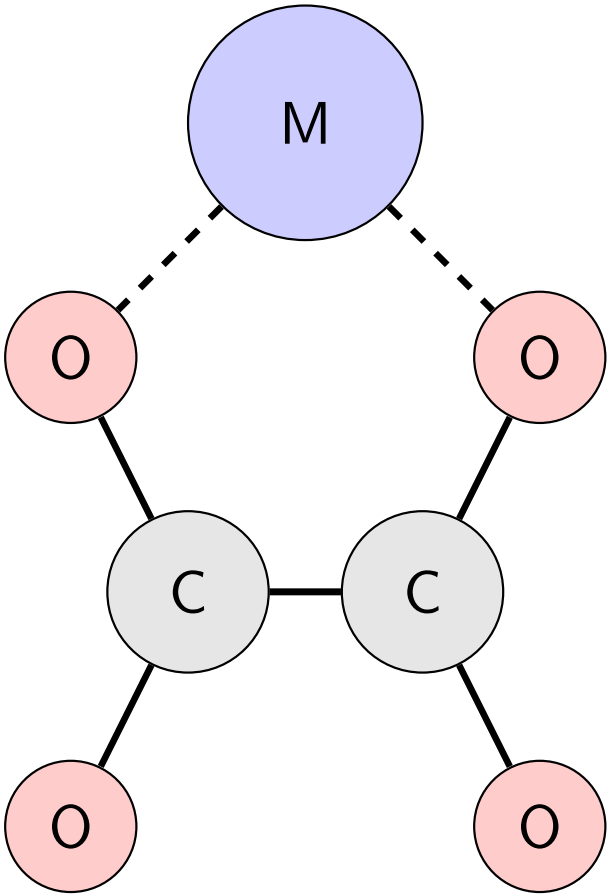


How to adapt to TM complexes?

[2] Janet, J.P., and Kulik, H.J., *arXiv*, 1708.06017, 2017.

RAC descriptors

RACs based on autocorrelations[2]



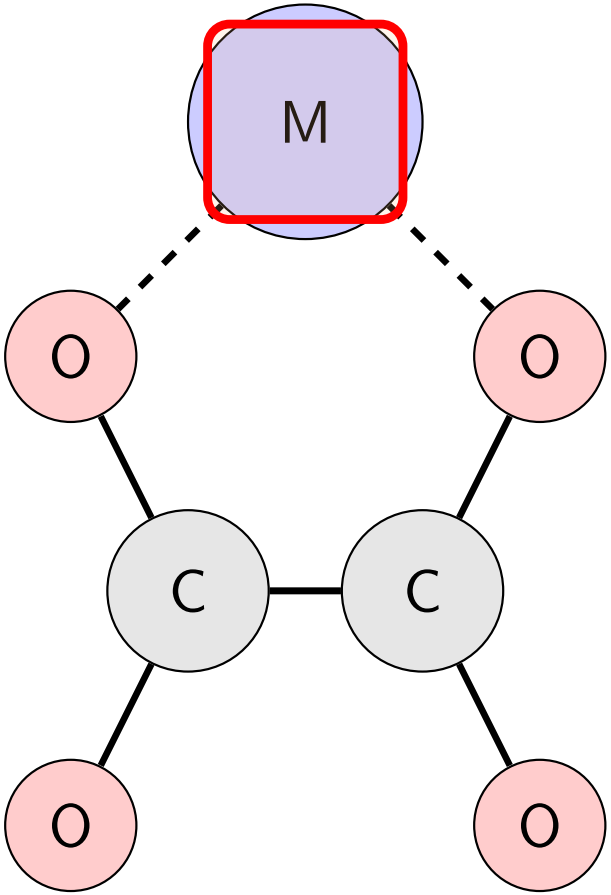
How to adapt to TM complexes?

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

[2] Janet, J.P., and Kulik, H.J., *arXiv*, 1708.06017, 2017.

RAC descriptors

RACs based on autocorrelations[2]



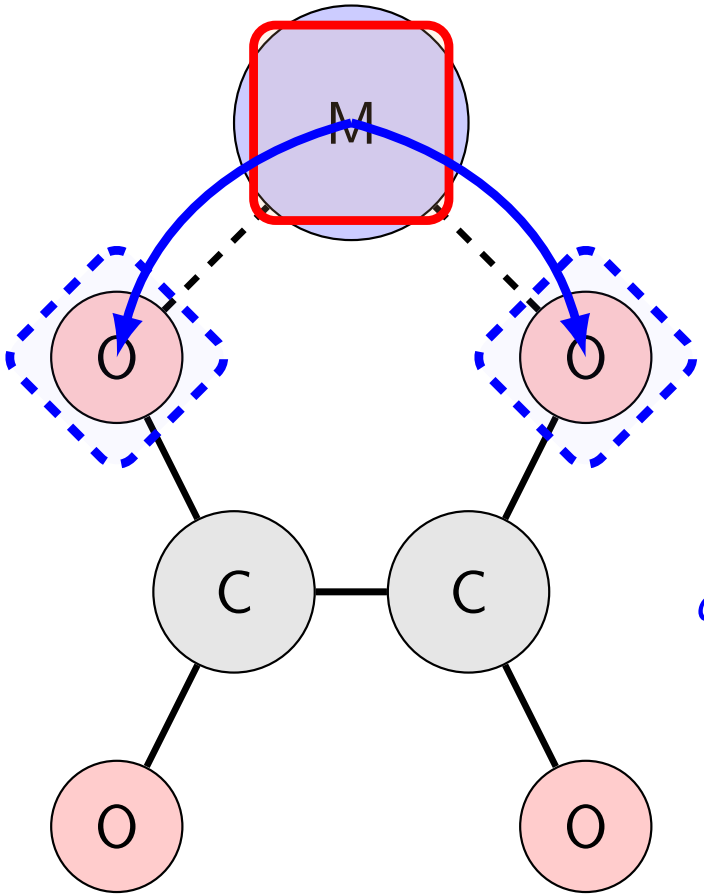
How to adapt to TM complexes?

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

[2] Janet, J.P., and Kulik, H.J., *arXiv*, 1708.06017, 2017.

RAC descriptors

RACs based on autocorrelations[2]



How to adapt to TM complexes?

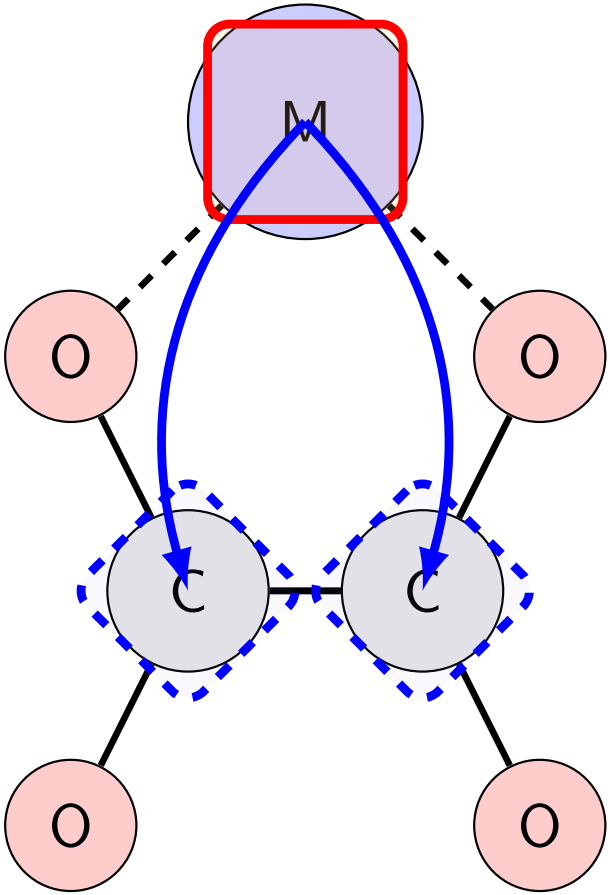
restrict the scope to focus on *near-metal atoms*

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

[2] Janet, J.P., and Kulik, H.J., *arXiv*, 1708.06017, 2017.

RAC descriptors

RACs based on autocorrelations[2]



How to adapt to TM complexes?

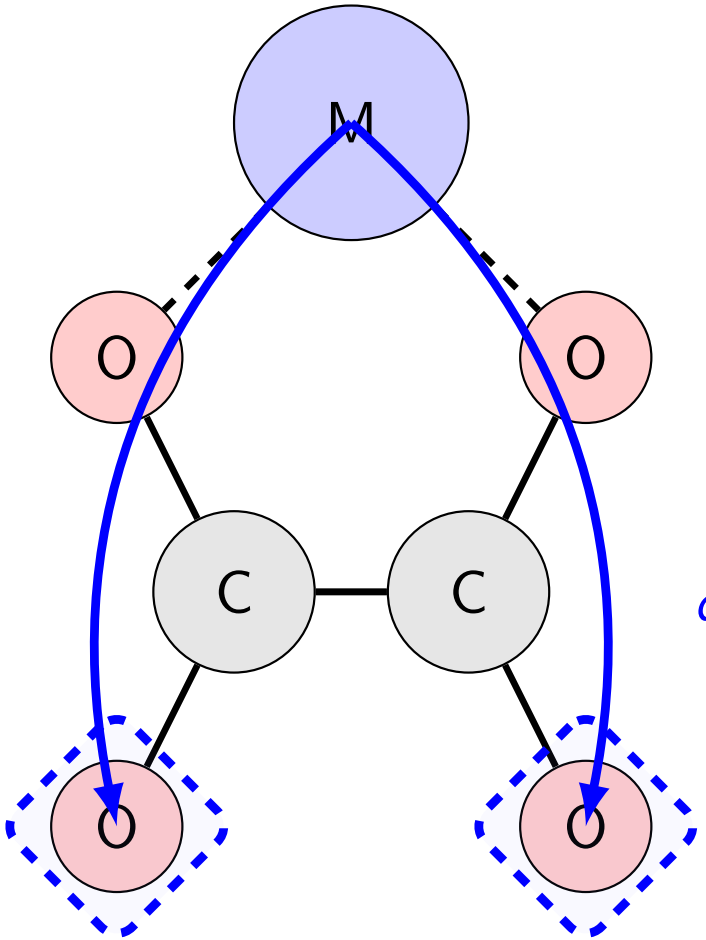
restrict the scope to focus on *near-metal atoms*

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

[2] Janet, J.P., and Kulik, H.J., *arXiv*, 1708.06017, 2017.

RAC descriptors

RACs based on autocorrelations[2]



How to adapt to TM complexes?

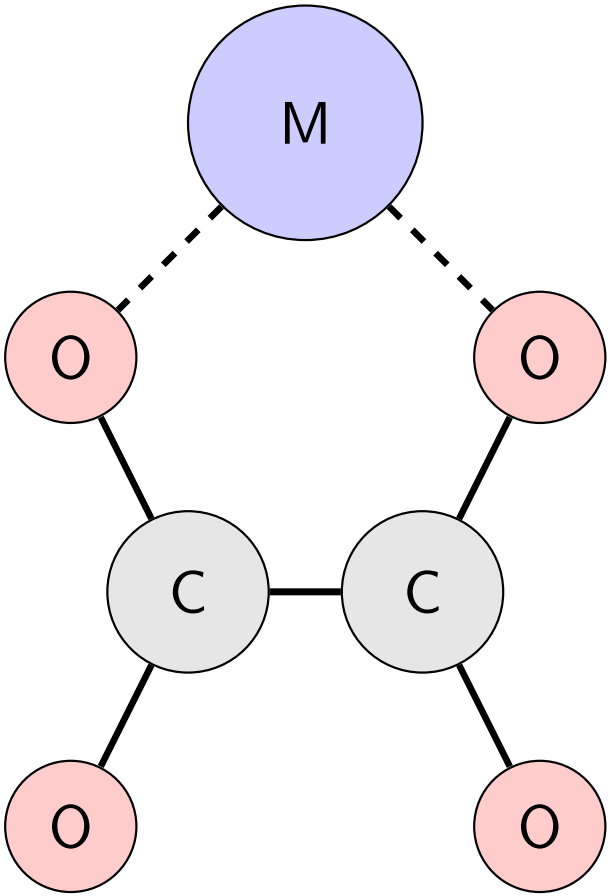
restrict the scope to focus on *near-metal atoms*

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

[2] Janet, J.P., and Kulik, H.J., *arXiv*, 1708.06017, 2017.

RAC descriptors

RACs based on autocorrelations[2]



How to adapt to TM complexes?

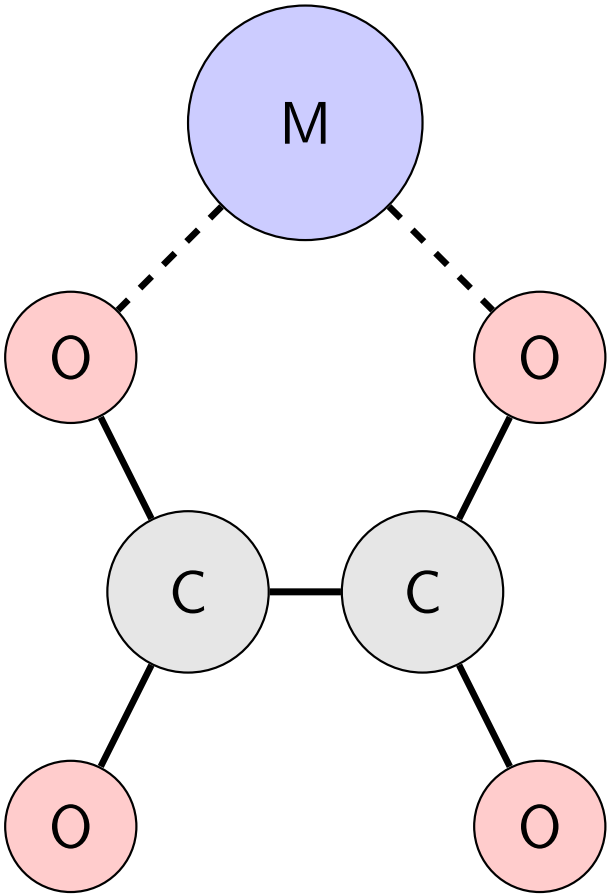
restrict the scope to focus on *near-metal atoms*

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

[2] Janet, J.P., and Kulik, H.J., *arXiv*, 1708.06017, 2017.

RAC descriptors

RACs based on autocorrelations[2]



How to adapt to TM complexes?

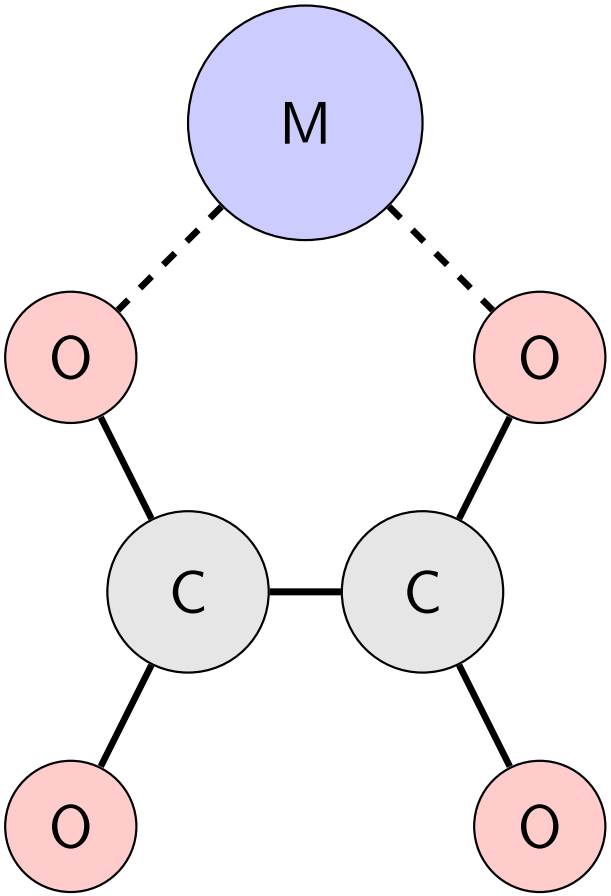
restrict the scope to focus on *near-metal atoms*

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

[2] Janet, J.P., and Kulik, H.J., *arXiv*, 1708.06017, 2017.

RAC descriptors

RACs based on autocorrelations[2]



How to adapt to TM complexes?

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

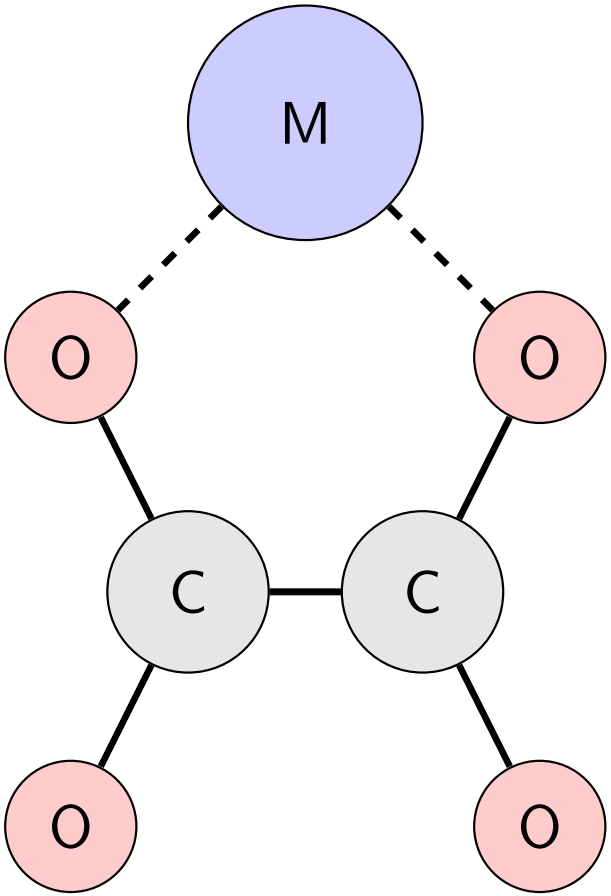
$$d_3 : \sum_{M,O} Z_M Z_O (Z_i - Z_j)$$

properties: T, χ, Z, I, S

[2] Janet, J.P., and Kulik, H.J., *arXiv*, 1708.06017, 2017.

RAC descriptors

RACs based on autocorrelations[2]



How to adapt to TM complexes?

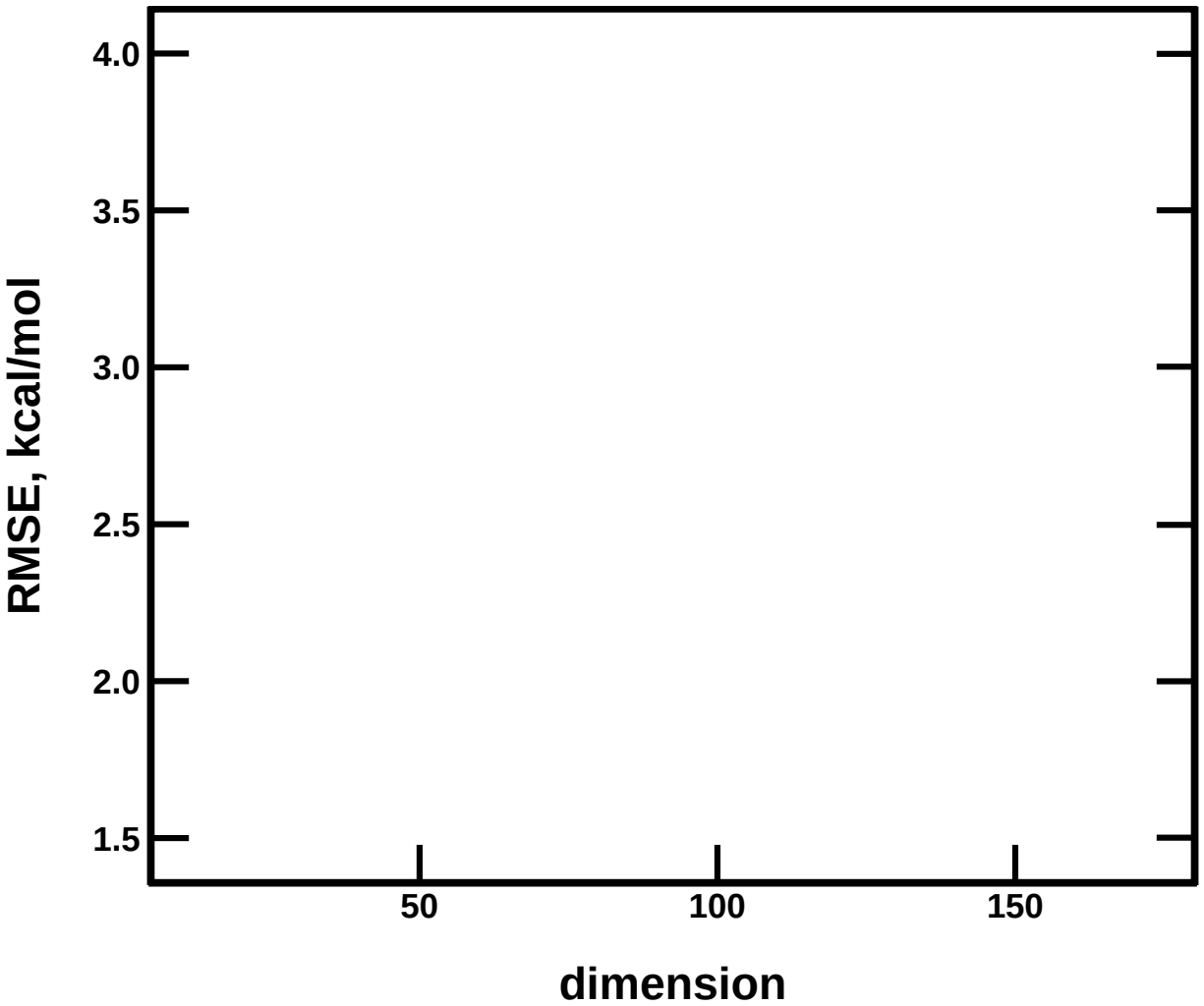
restrict the scope to focus on *near-metal atoms*

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

$\boxed{\sim 160 \text{ features in total}}$

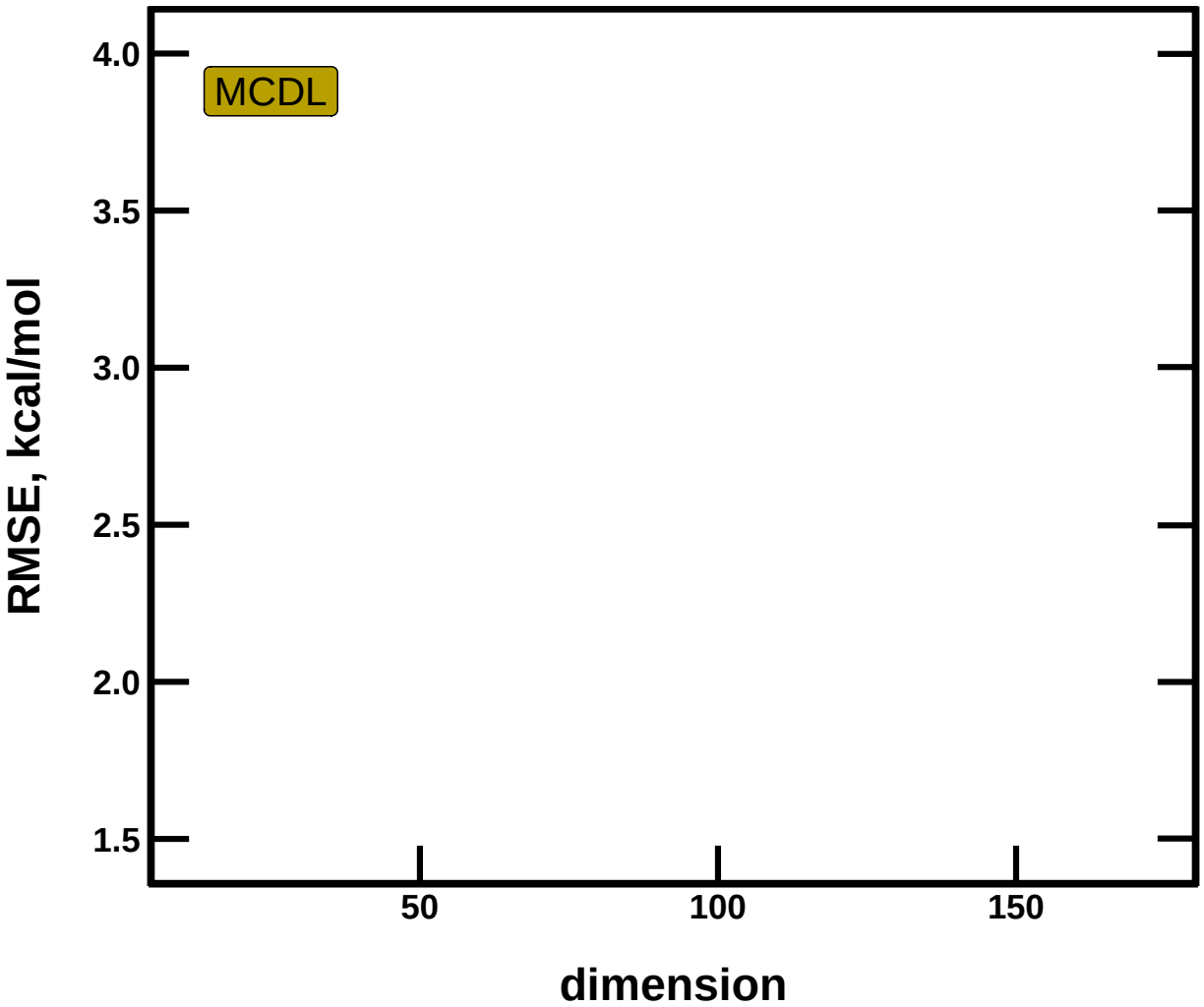
[2] Janet, J.P., and Kulik, H.J., *arXiv*, 1708.06017, 2017.

Feature selection



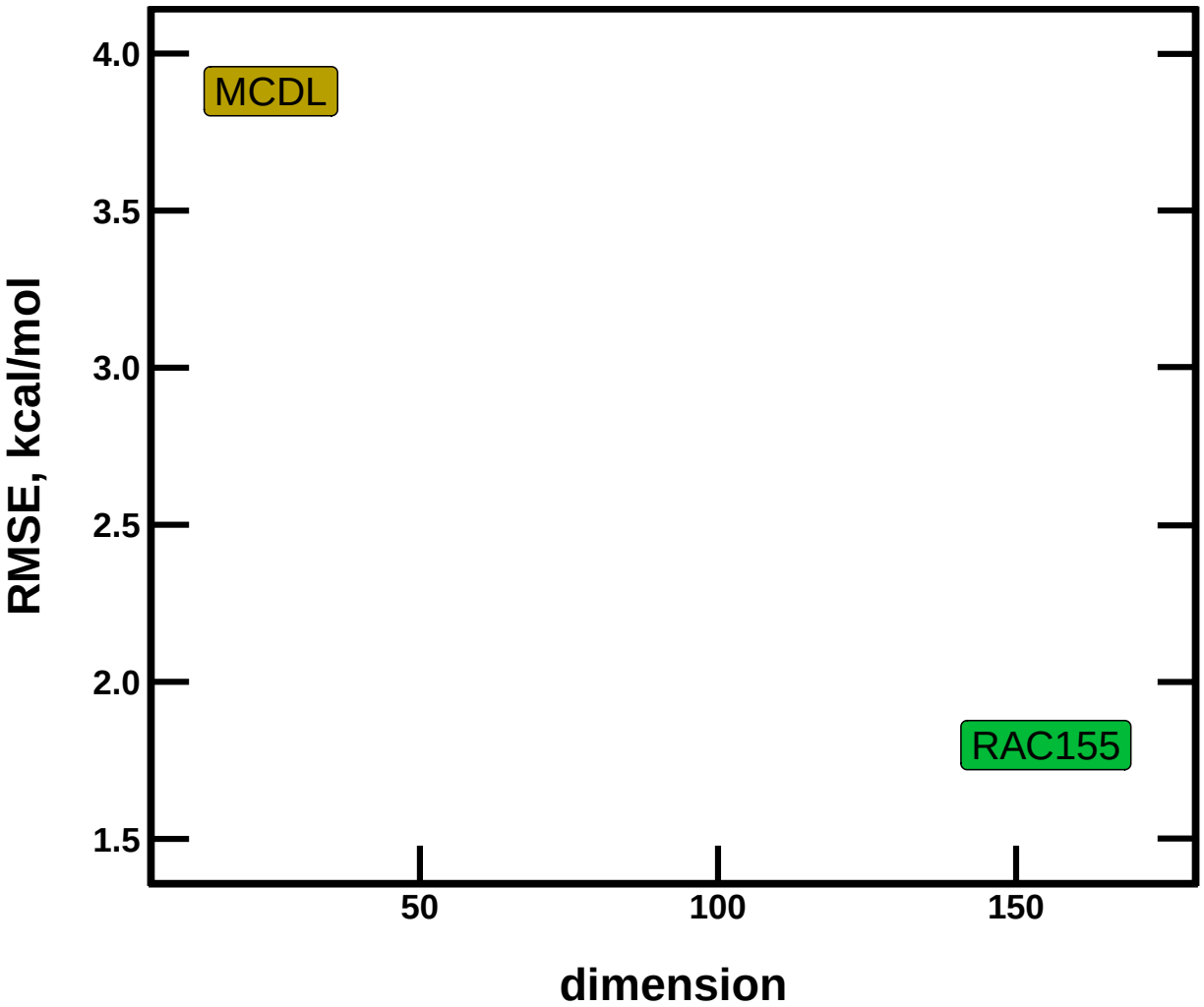
Janet, J.P., and Kulik, H.J., *arXiv*, 1708.06017, 2017.

Feature selection



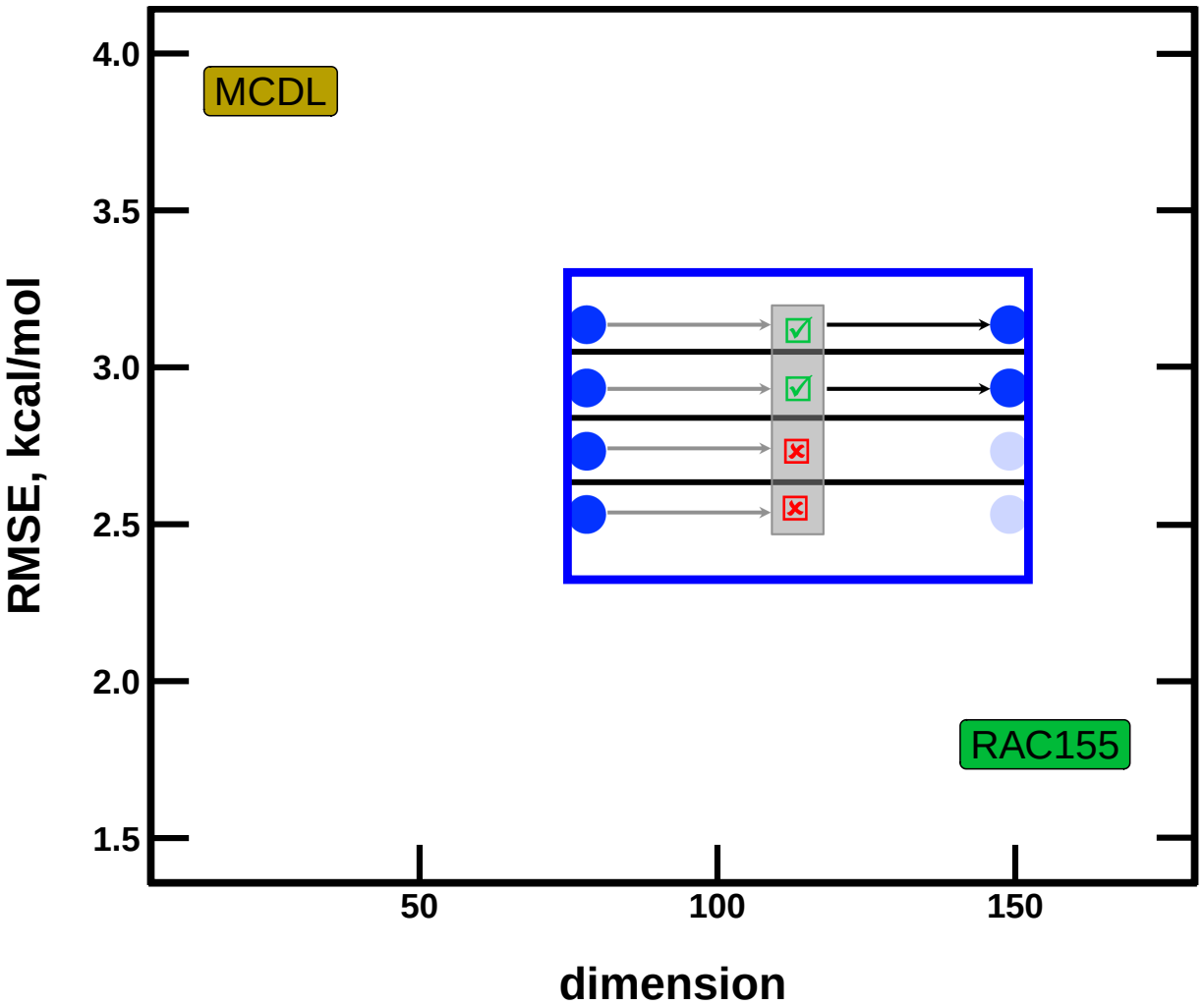
Janet, J.P., and Kulik, H.J., *arXiv*, 1708.06017, 2017.

Feature selection



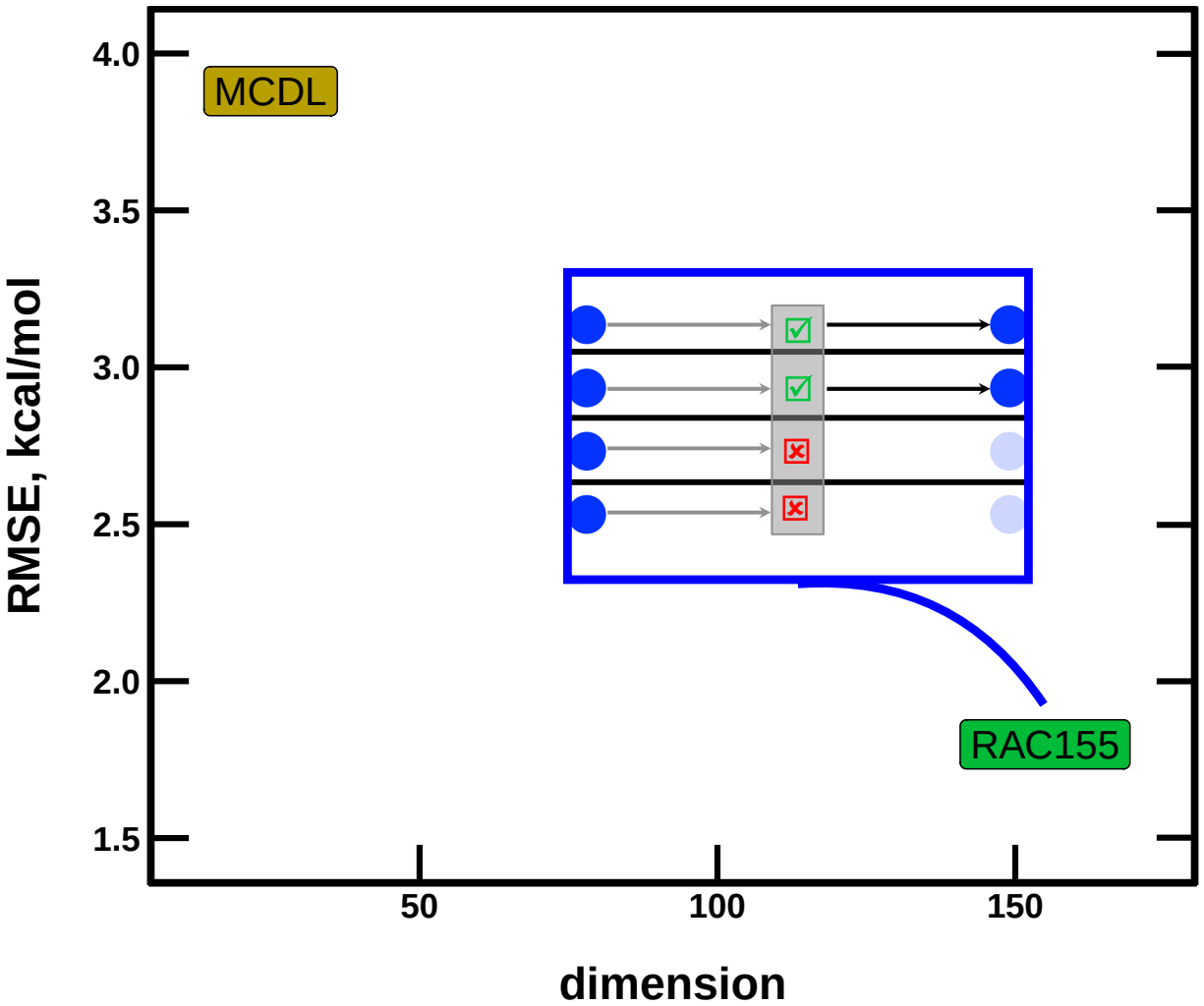
Janet, J.P., and Kulik, H.J., *arXiv*, 1708.06017, 2017.

Feature selection



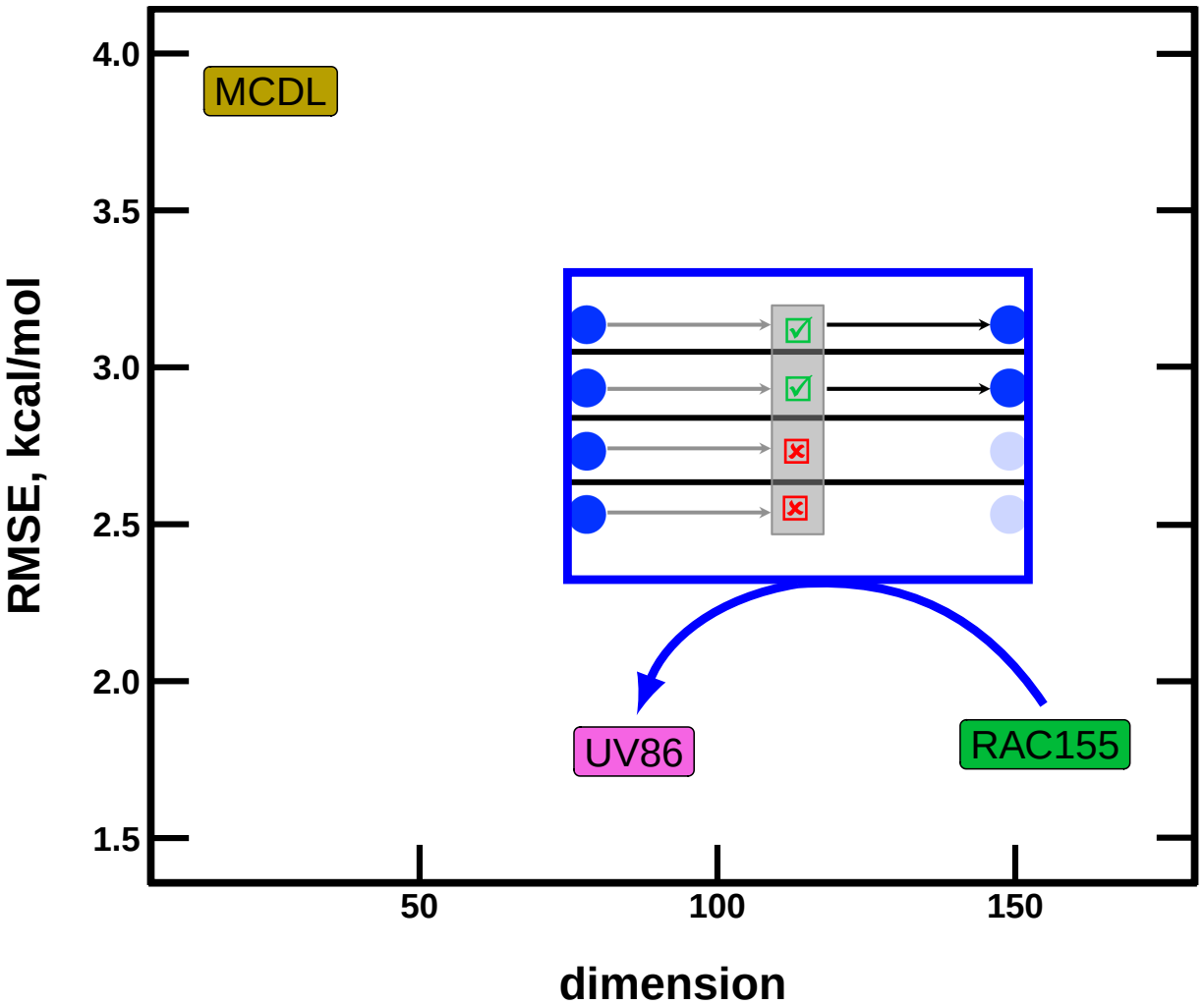
Janet, J.P., and Kulik, H.J., *arXiv*, 1708.06017, 2017.

Feature selection



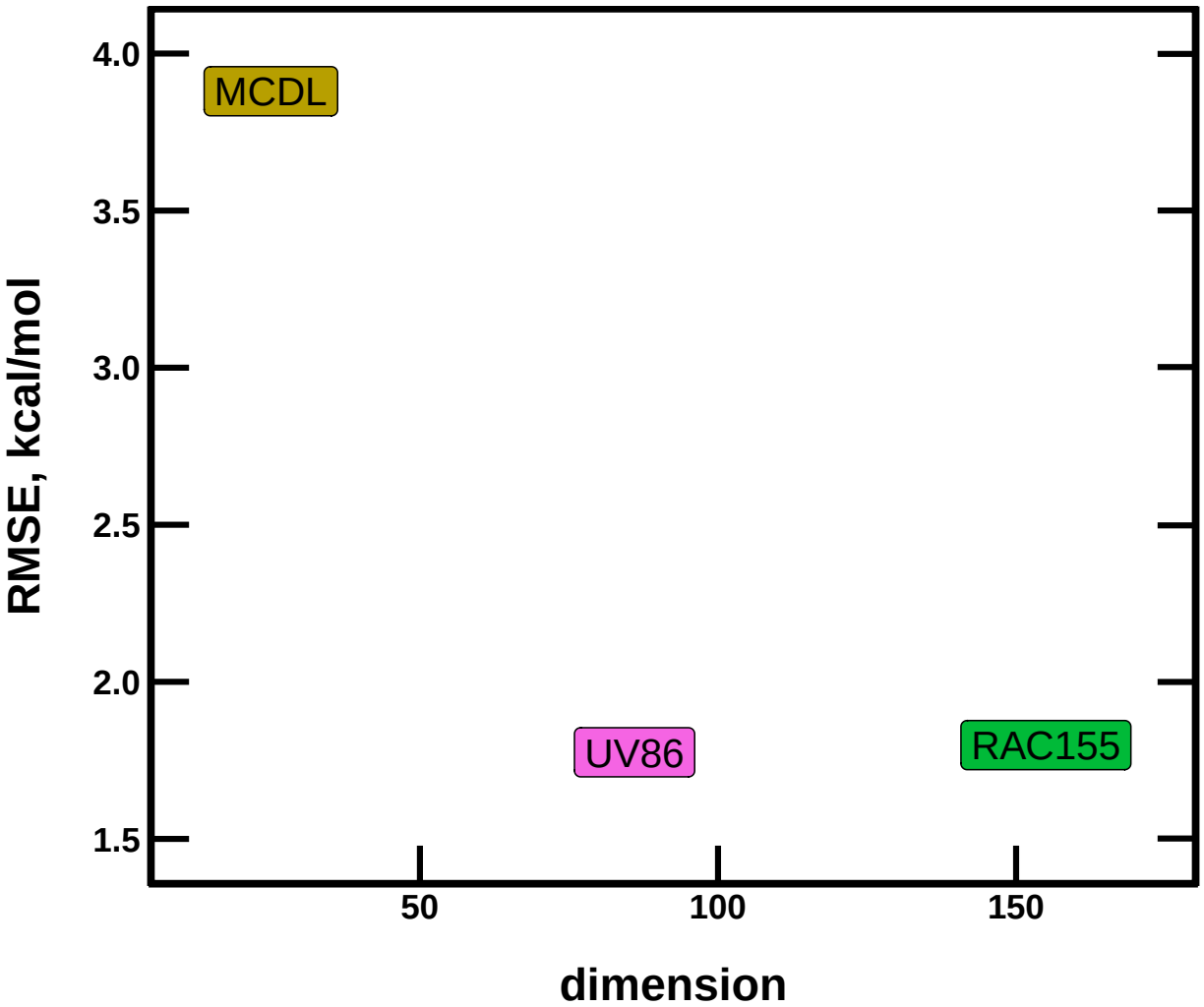
Janet, J.P., and Kulik, H.J., *arXiv*, 1708.06017, 2017.

Feature selection



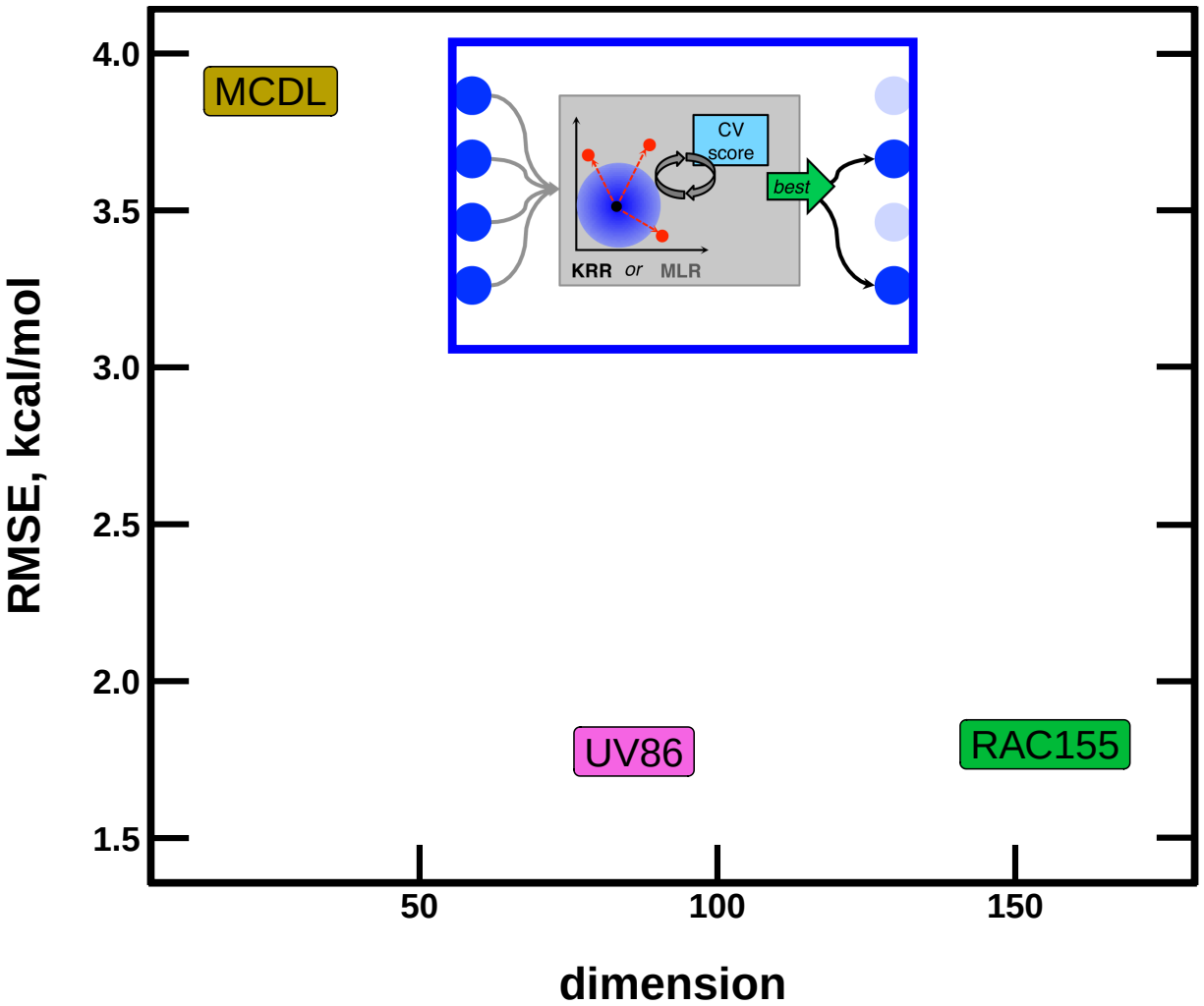
Janet, J.P., and Kulik, H.J., *arXiv*, 1708.06017, 2017.

Feature selection



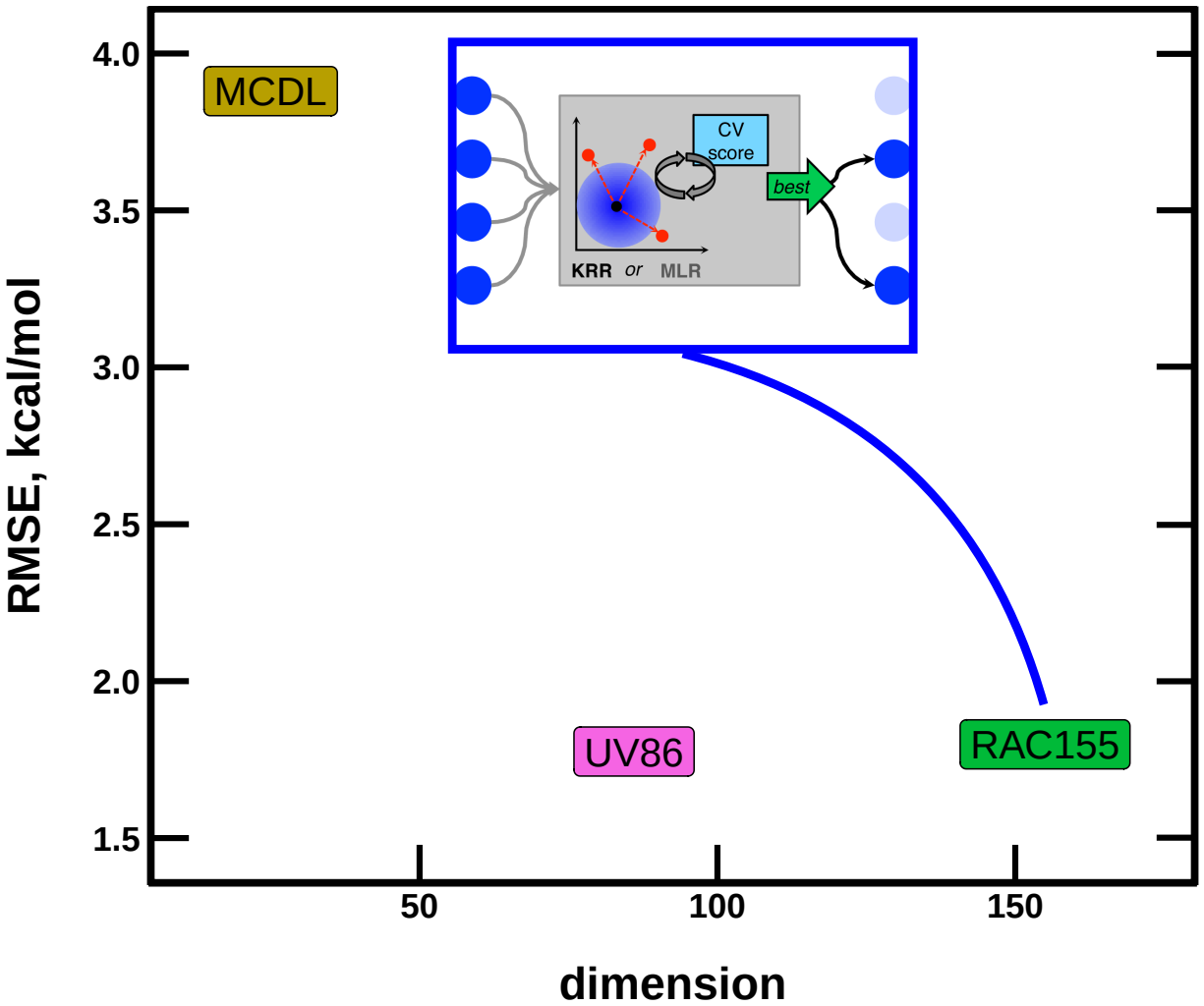
Janet, J.P., and Kulik, H.J., *arXiv*, 1708.06017, 2017.

Feature selection



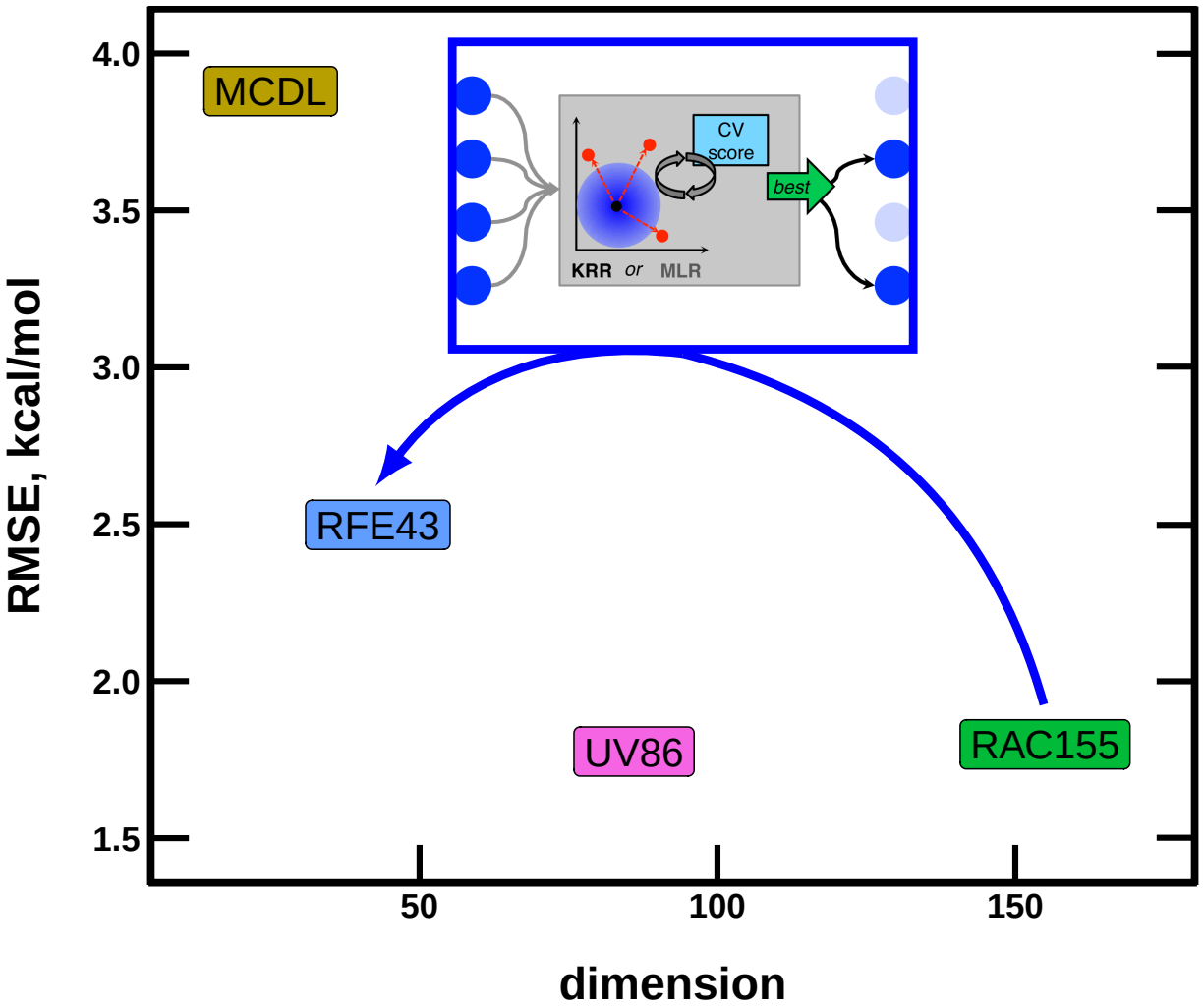
Janet, J.P., and Kulik, H.J., *arXiv*, 1708.06017, 2017.

Feature selection



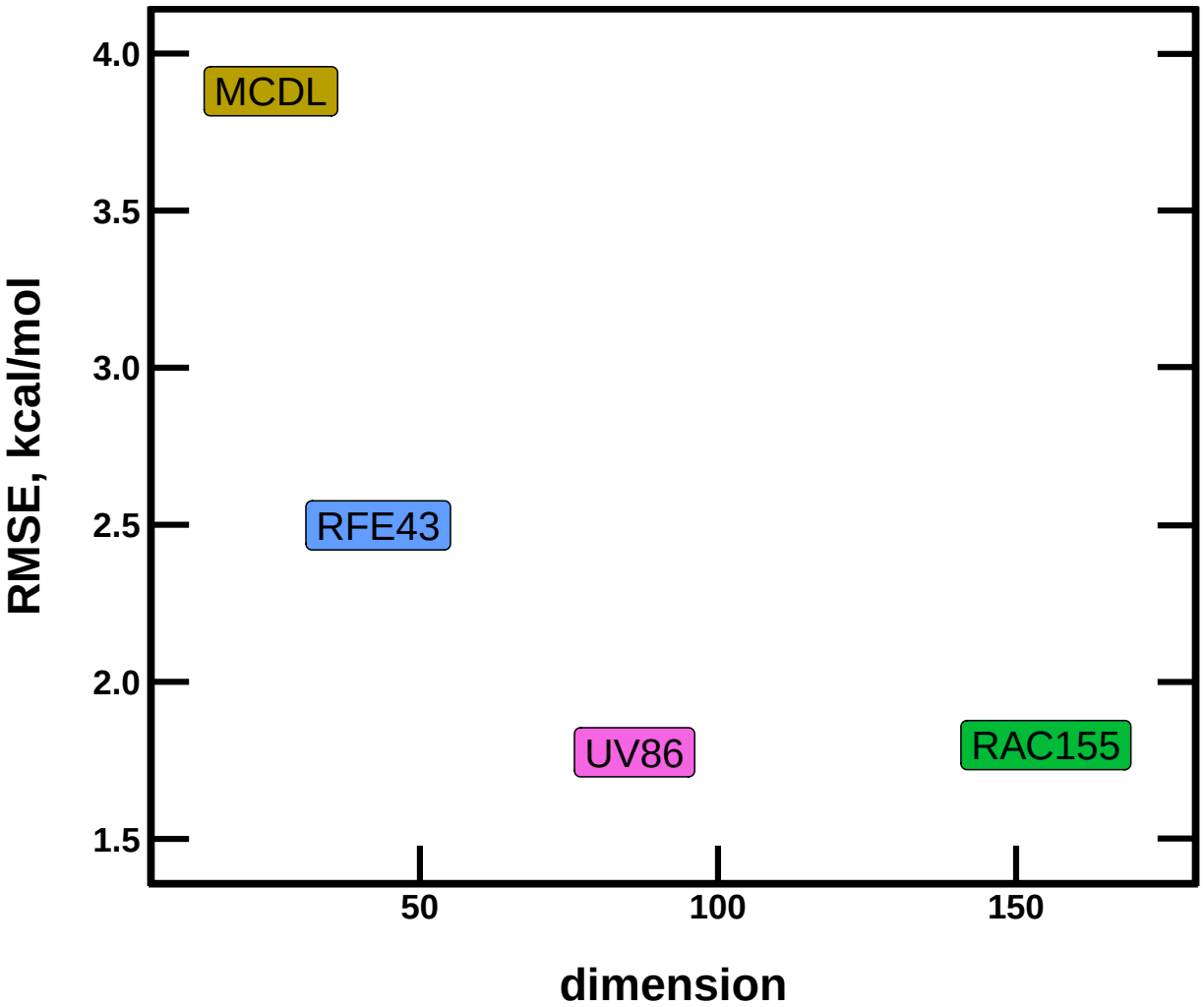
Janet, J.P., and Kulik, H.J., *arXiv*, 1708.06017, 2017.

Feature selection



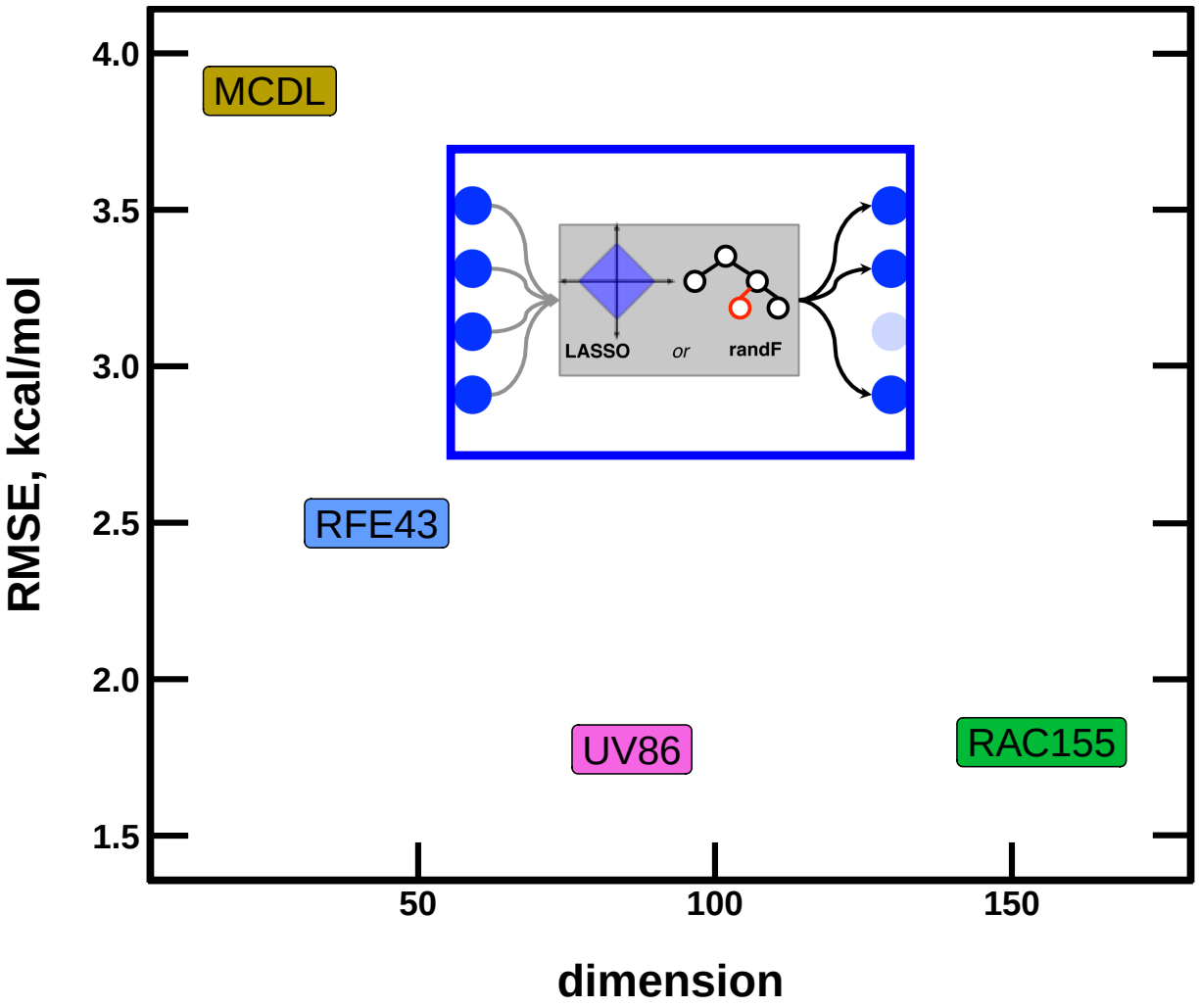
Janet, J.P., and Kulik, H.J., *arXiv*, 1708.06017, 2017.

Feature selection



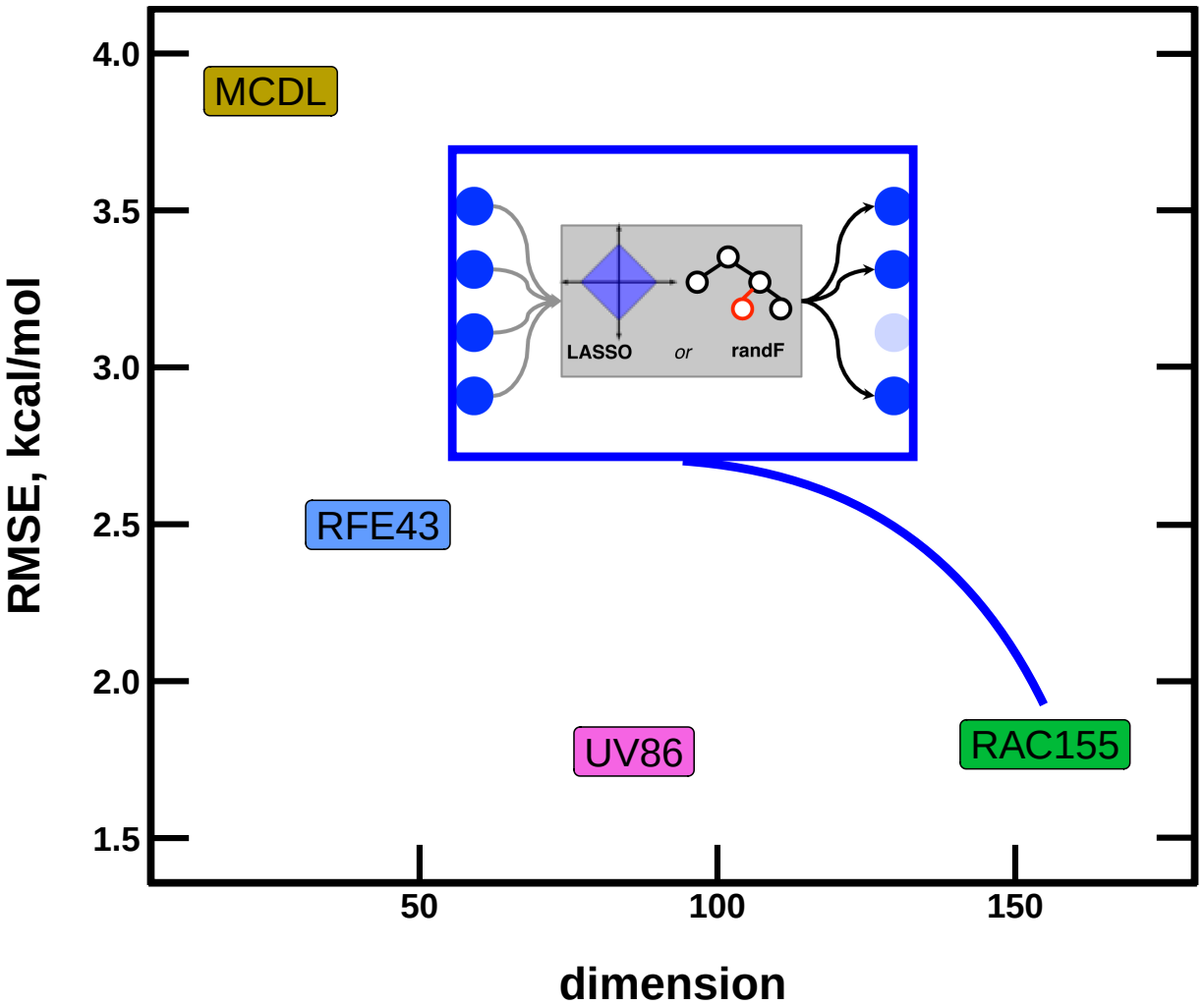
Janet, J.P., and Kulik, H.J., *arXiv*, 1708.06017, 2017.

Feature selection



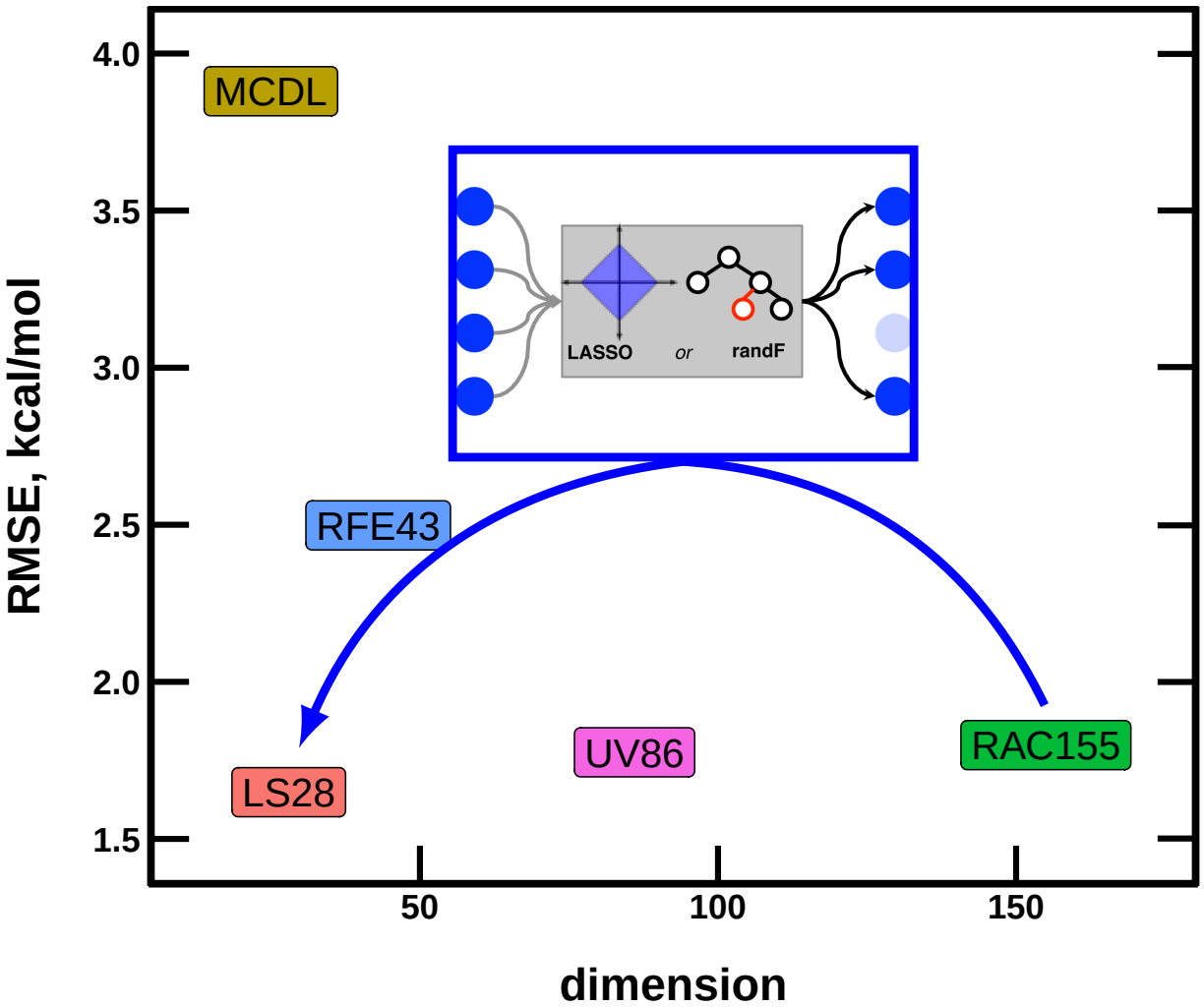
Janet, J.P., and Kulik, H.J., *arXiv*, 1708.06017, 2017.

Feature selection



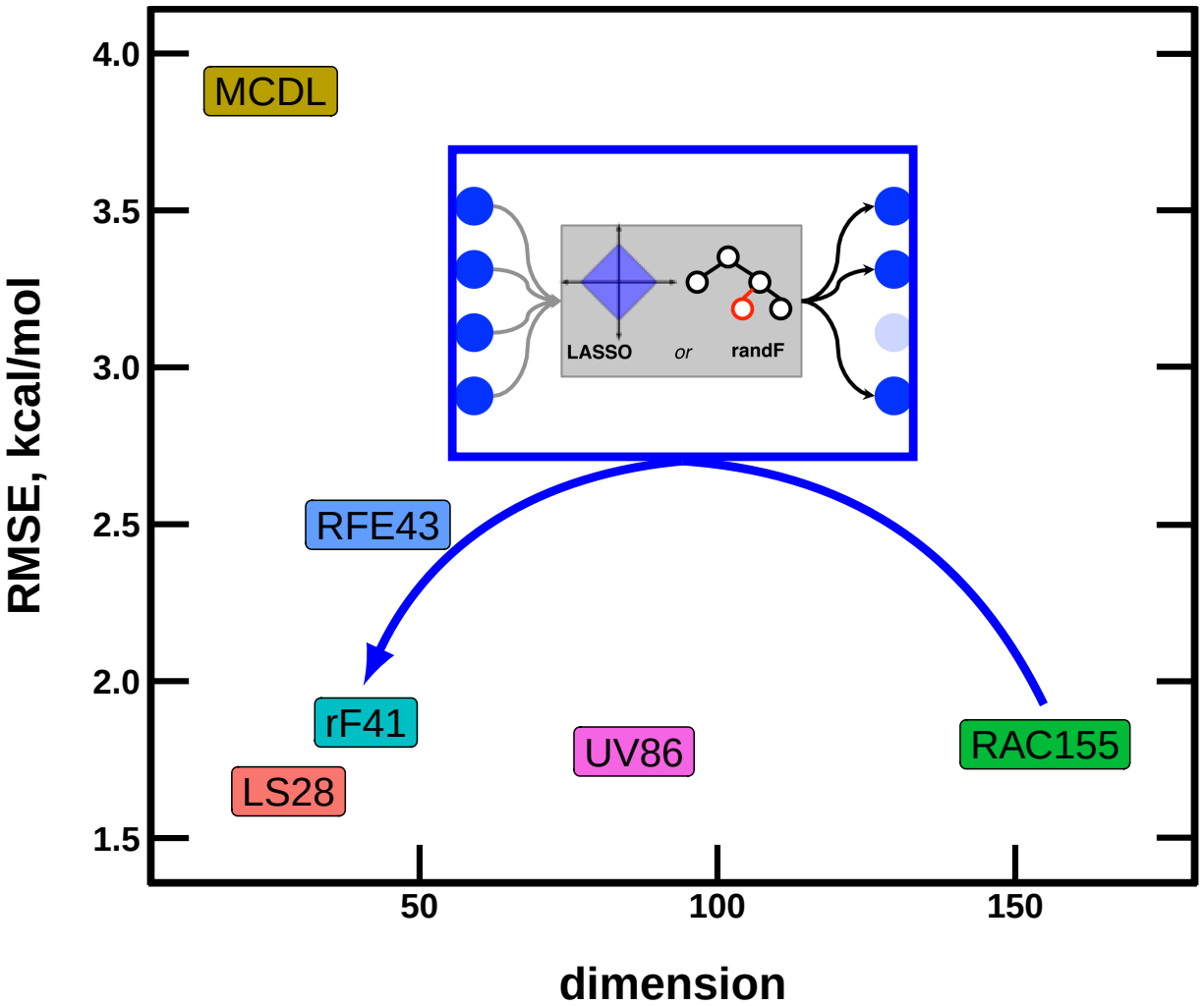
Janet, J.P., and Kulik, H.J., *arXiv*, 1708.06017, 2017.

Feature selection



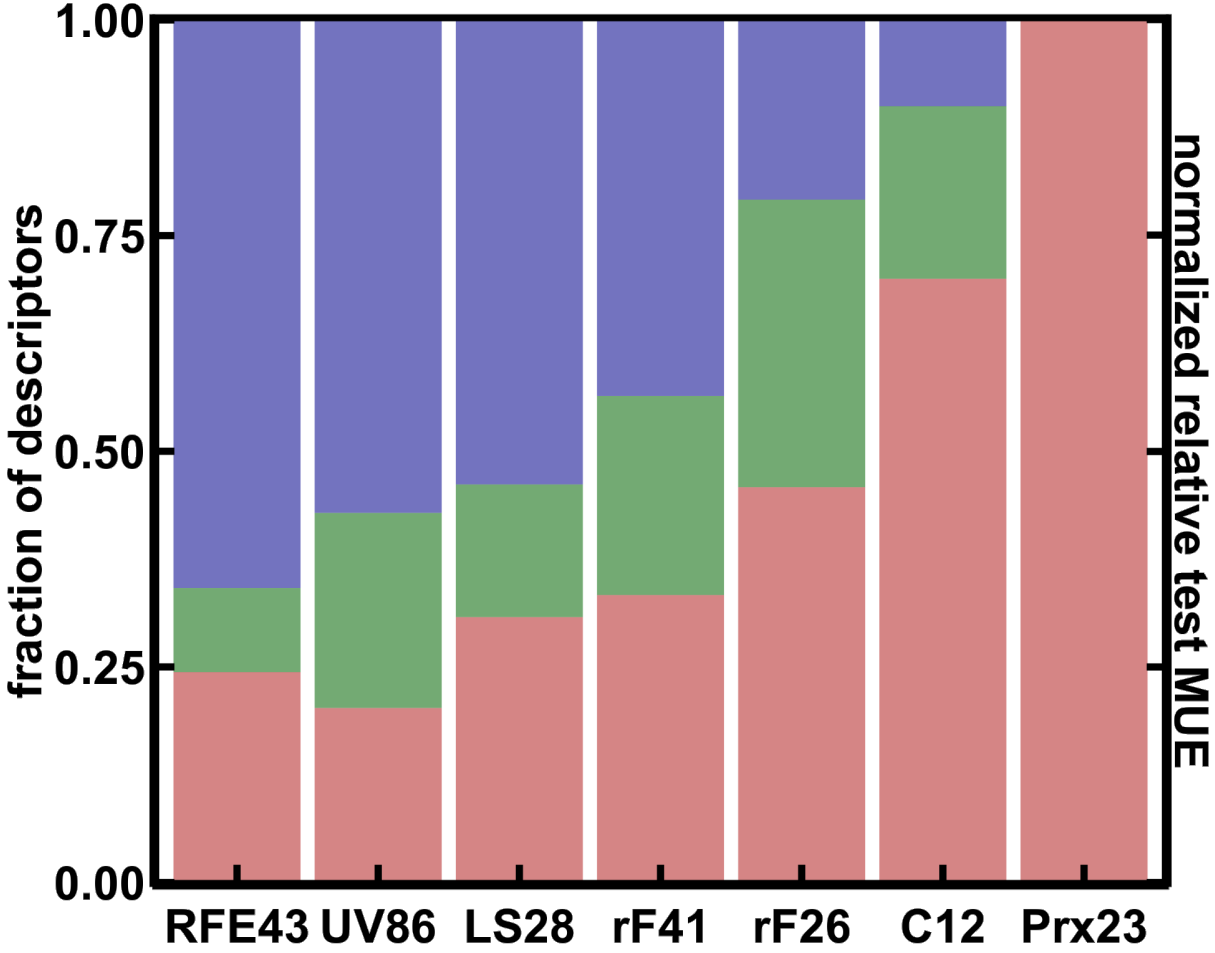
Janet, J.P., and Kulik, H.J., *arXiv*, 1708.06017, 2017.

Feature selection



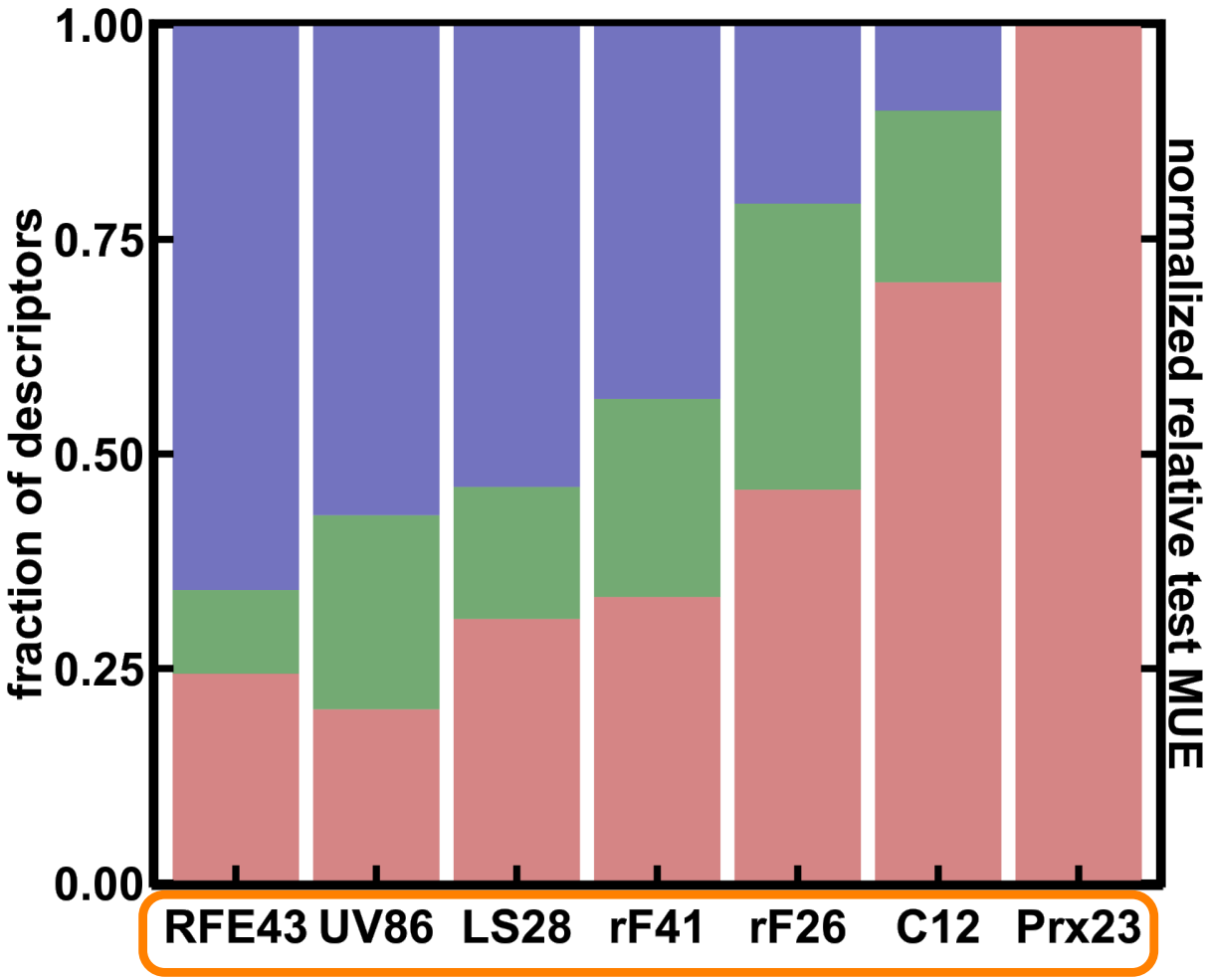
Janet, J.P., and Kulik, H.J., *arXiv*, 1708.06017, 2017.

How local is too local?



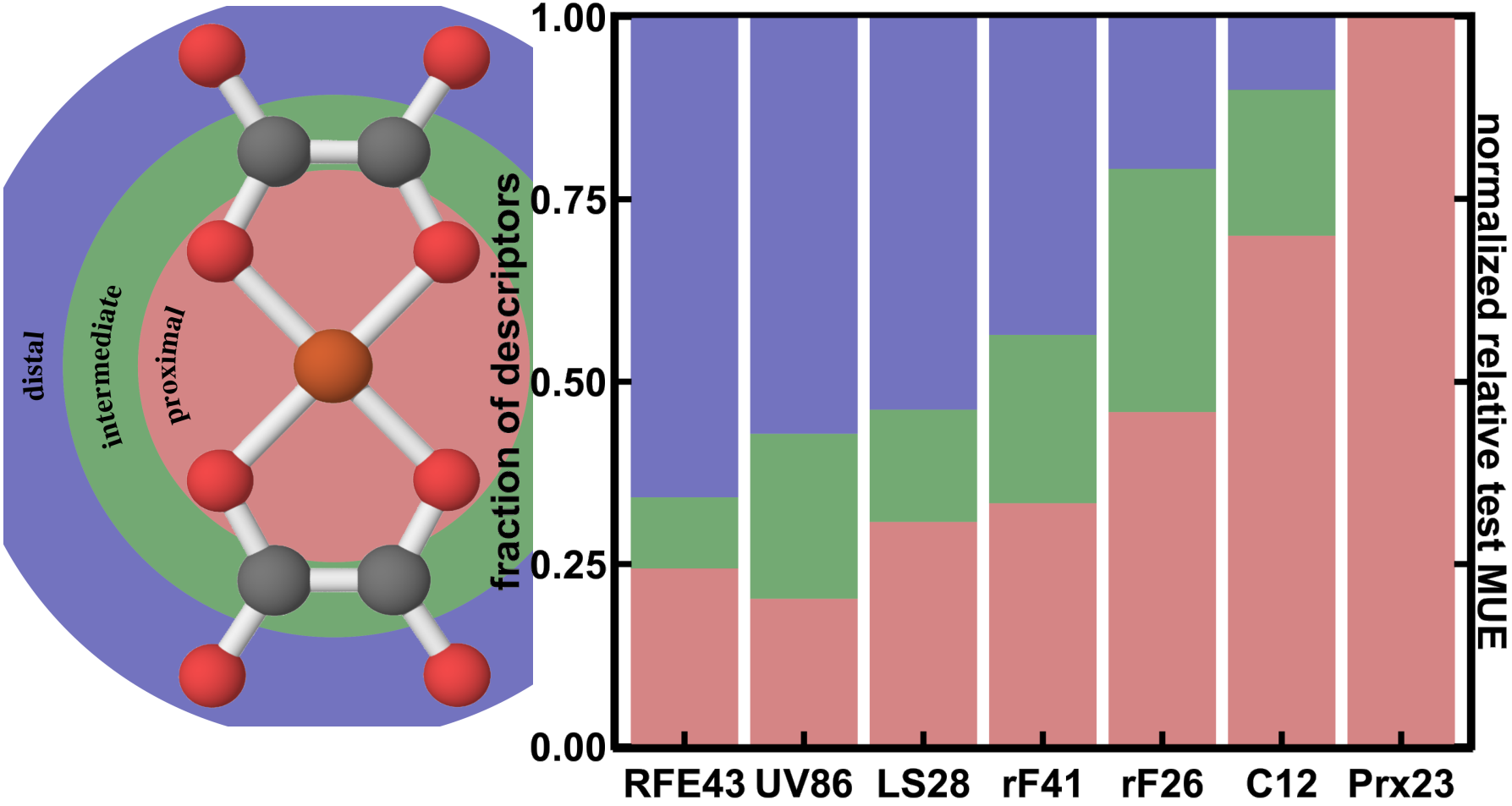
Janet, J.P., and Kulik, H.J., *arXiv*, 1708.06017, 2017.

How local is too local?



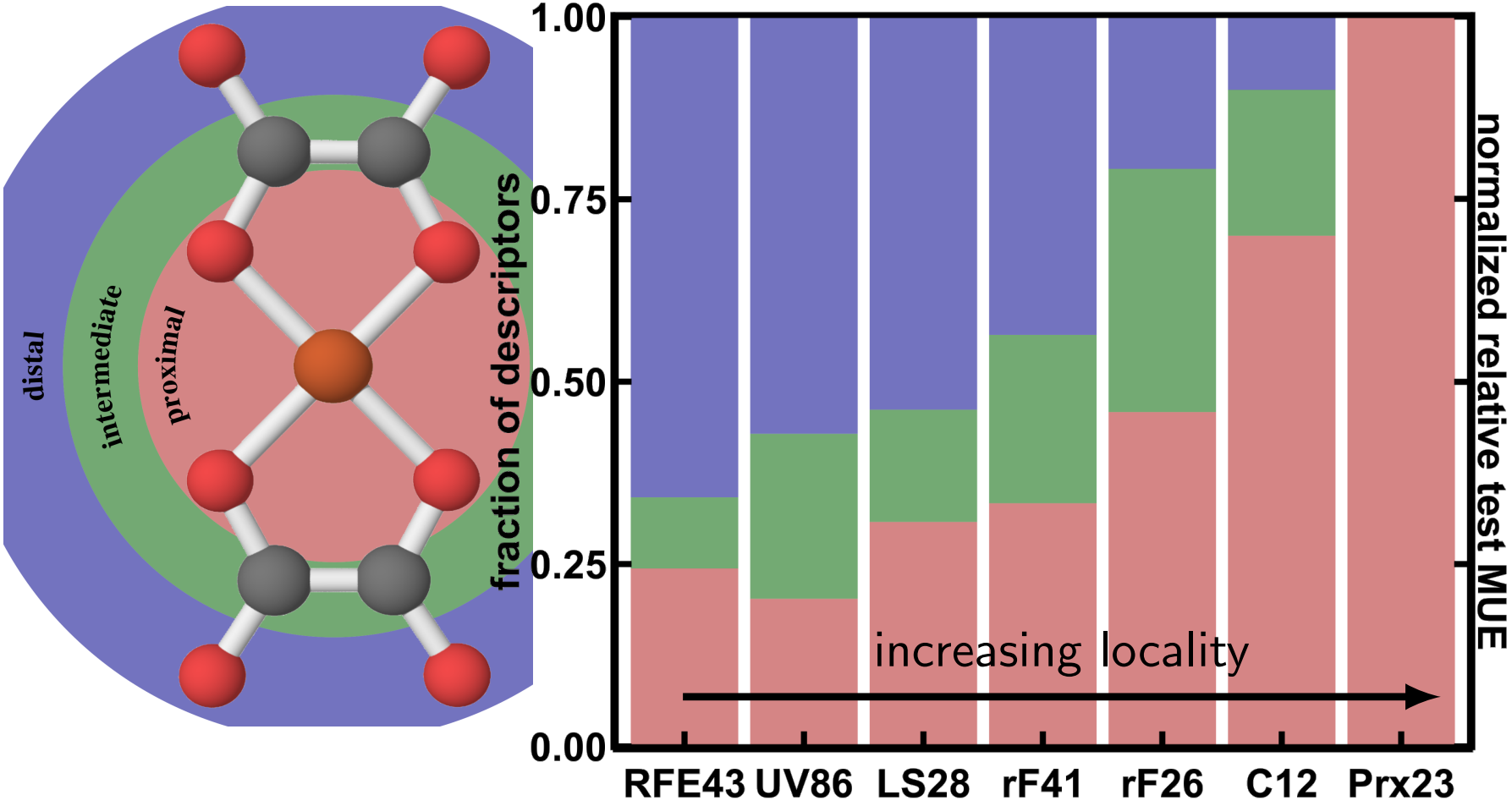
Janet, J.P., and Kulik, H.J., *arXiv*, 1708.06017, 2017.

How local is too local?



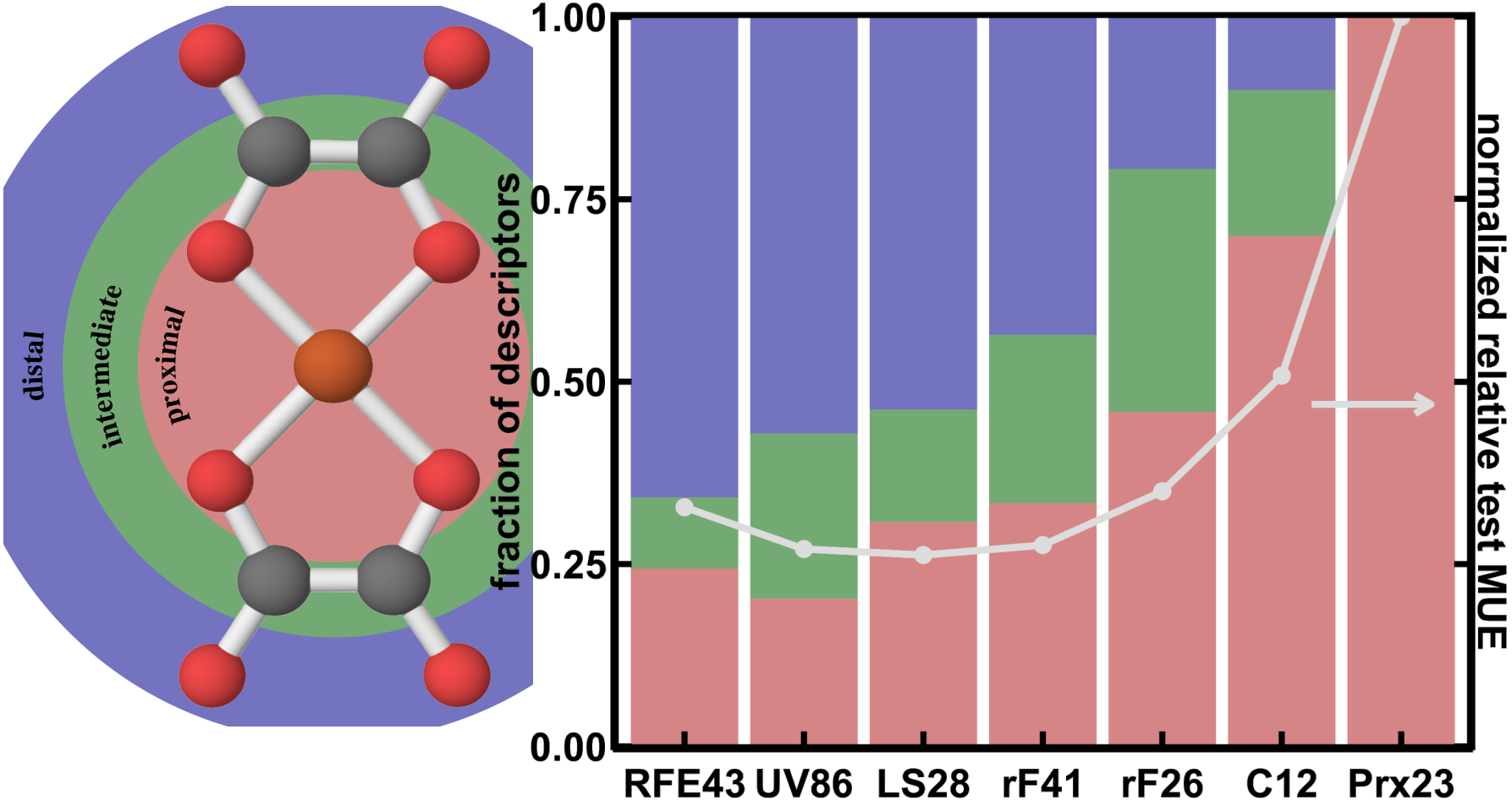
Janet, J.P., and Kulik, H.J., *arXiv*, 1708.06017, 2017.

How local is too local?



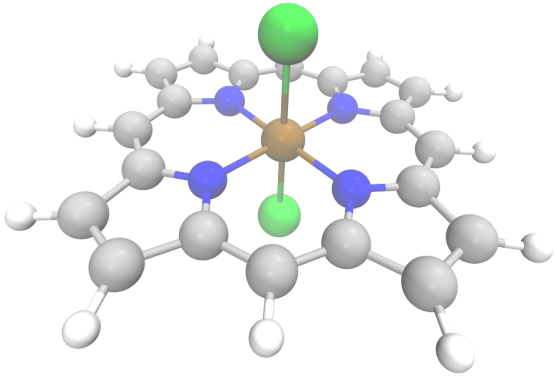
Janet, J.P., and Kulik, H.J., *arXiv*, 1708.06017, 2017.

How local is too local?

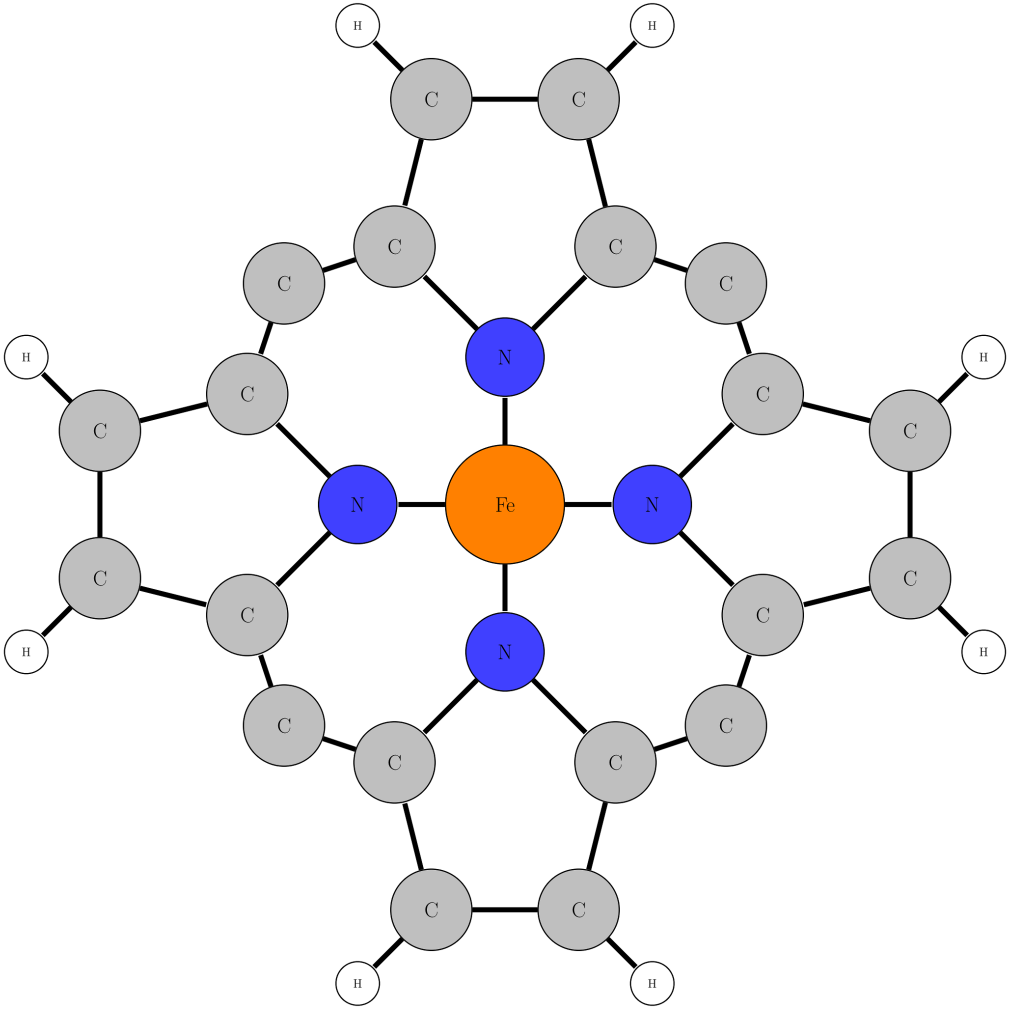


Janet, J.P., and Kulik, H.J., *arXiv*, 1708.06017, 2017.

Do features depend on properties?

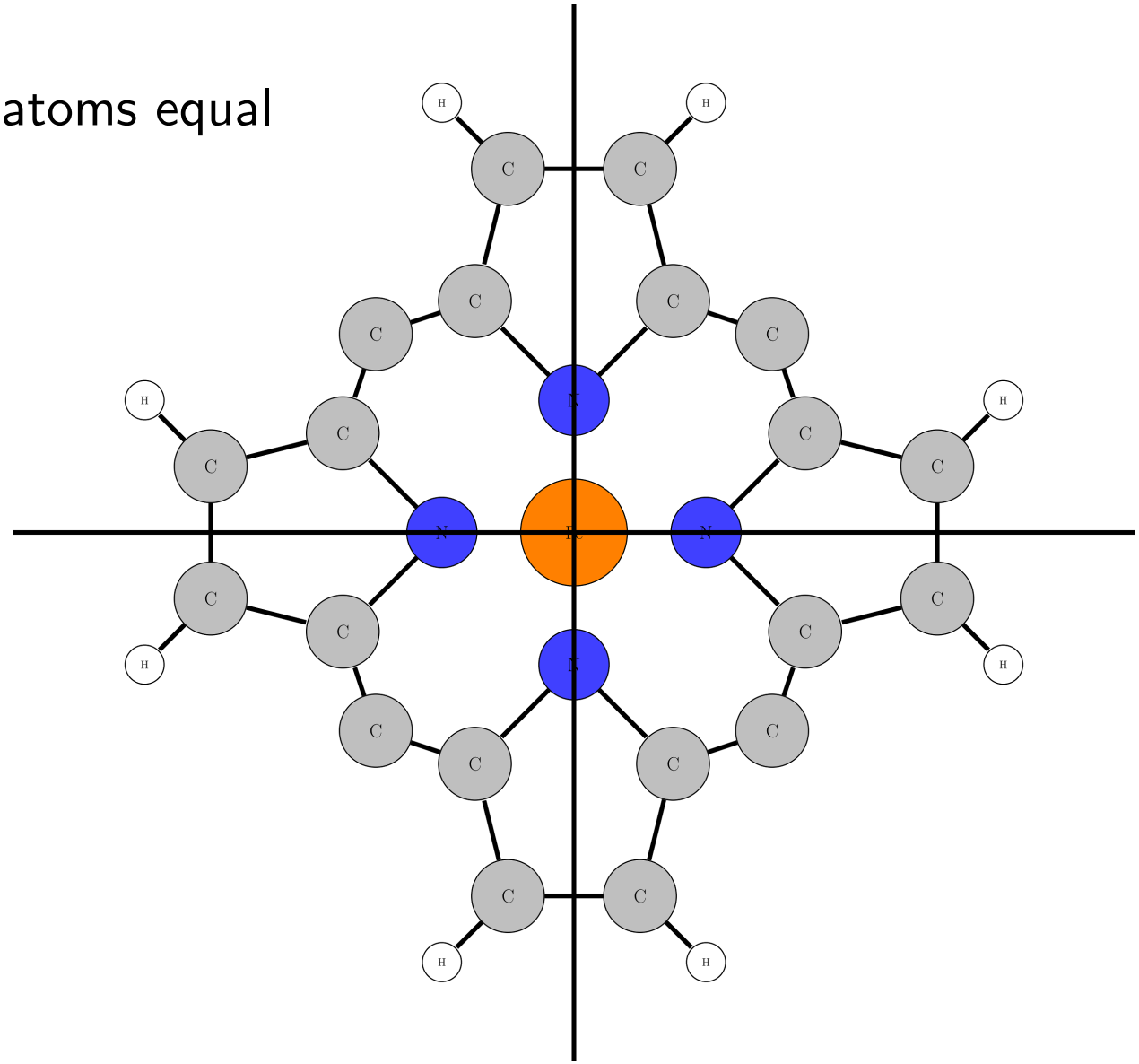


Do features depend on properties?



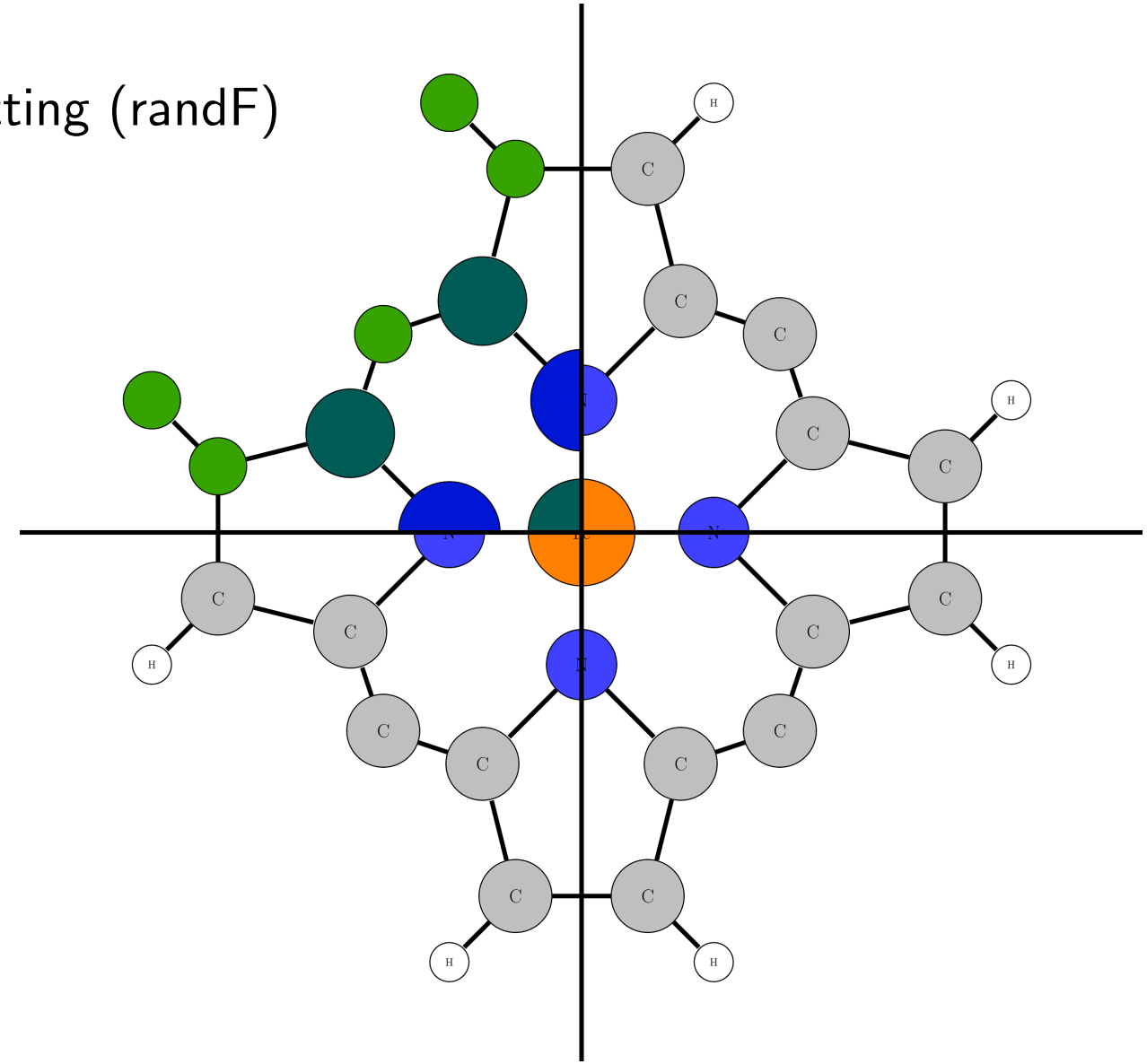
Do features depend on properties?

All atoms equal



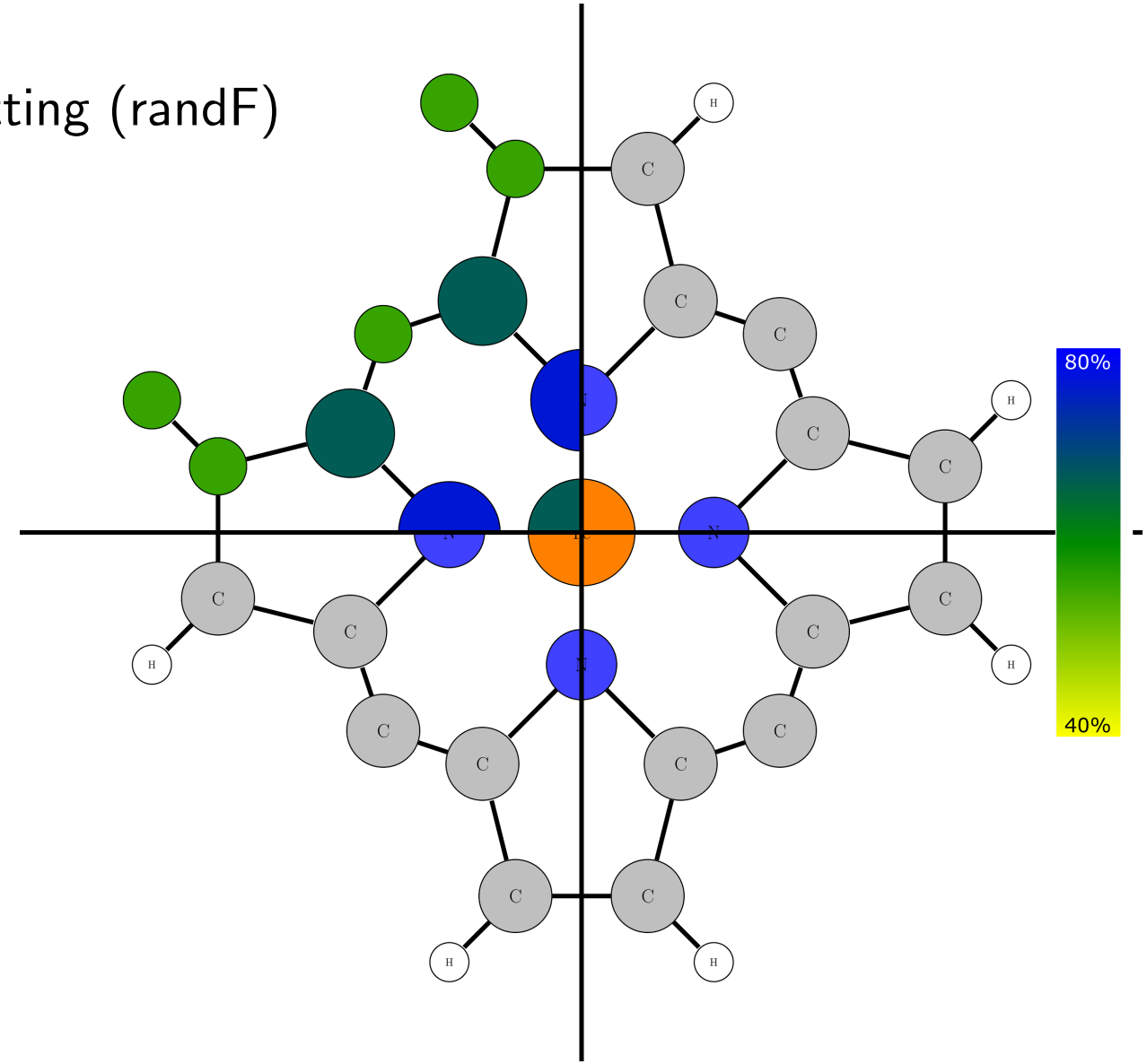
Do features depend on properties?

spin splitting (randF)



Do features depend on properties?

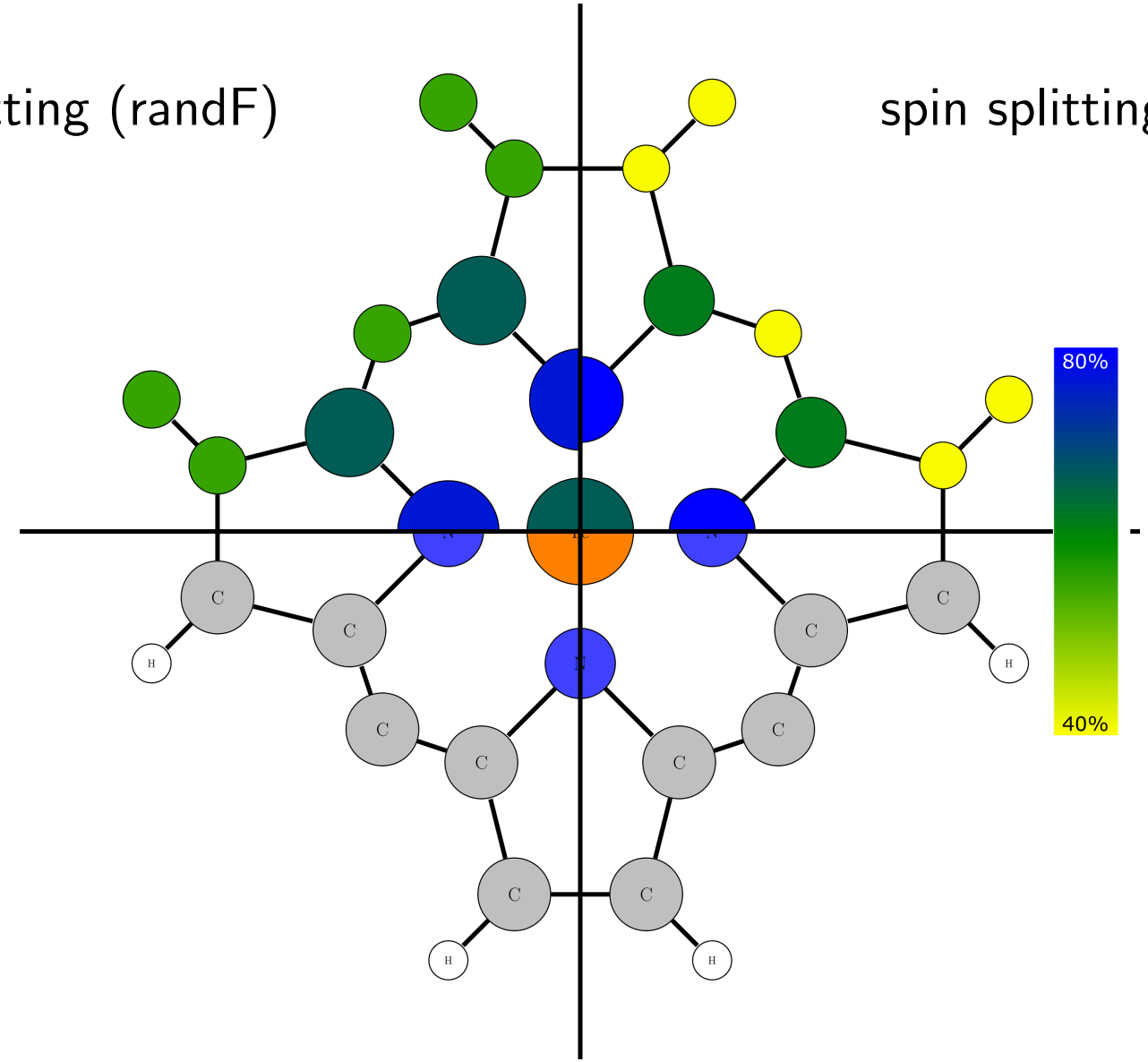
spin splitting (randF)



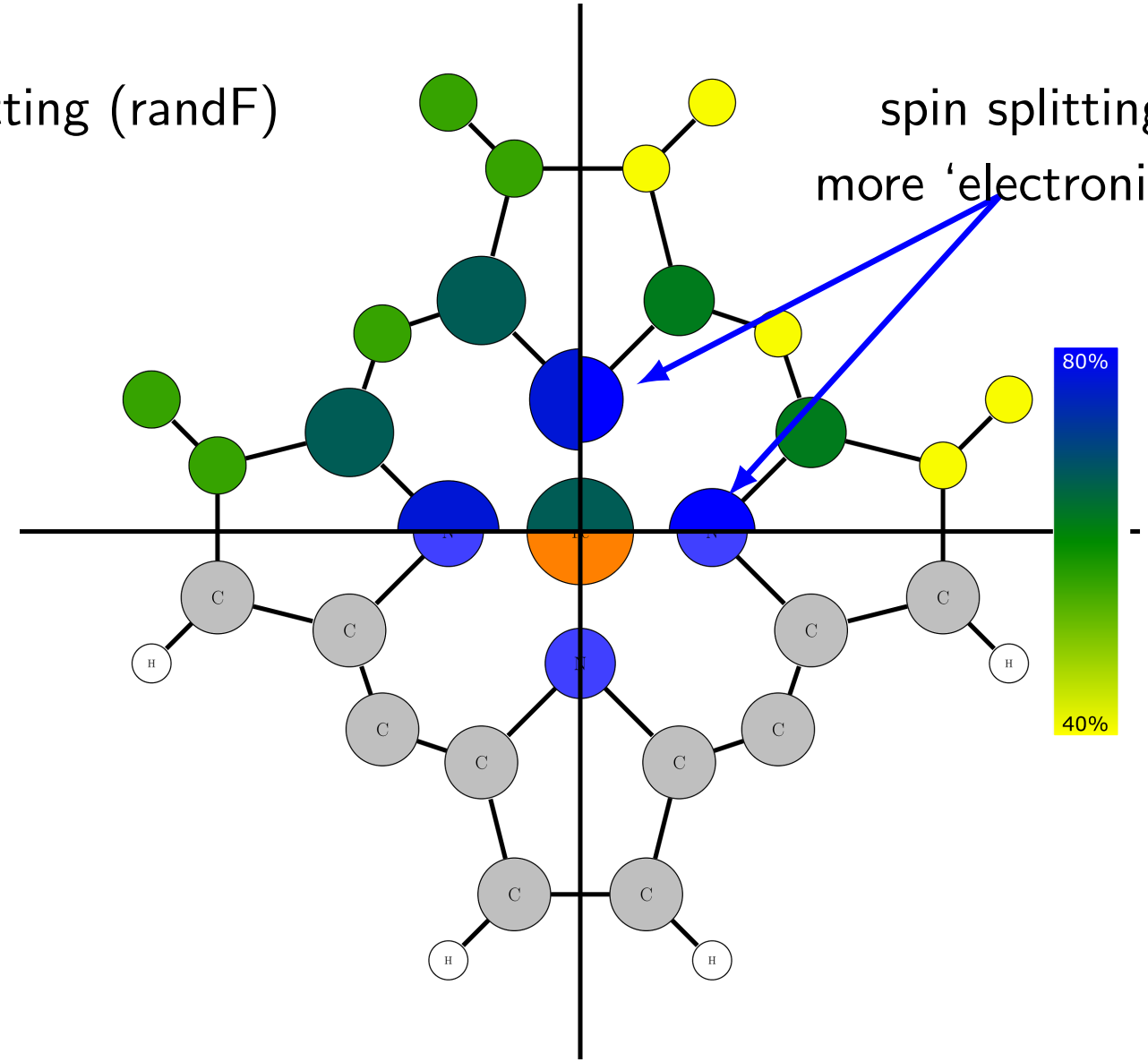
Do features depend on properties?

spin splitting (randF)

spin splitting (randF)



Do features depend on properties?

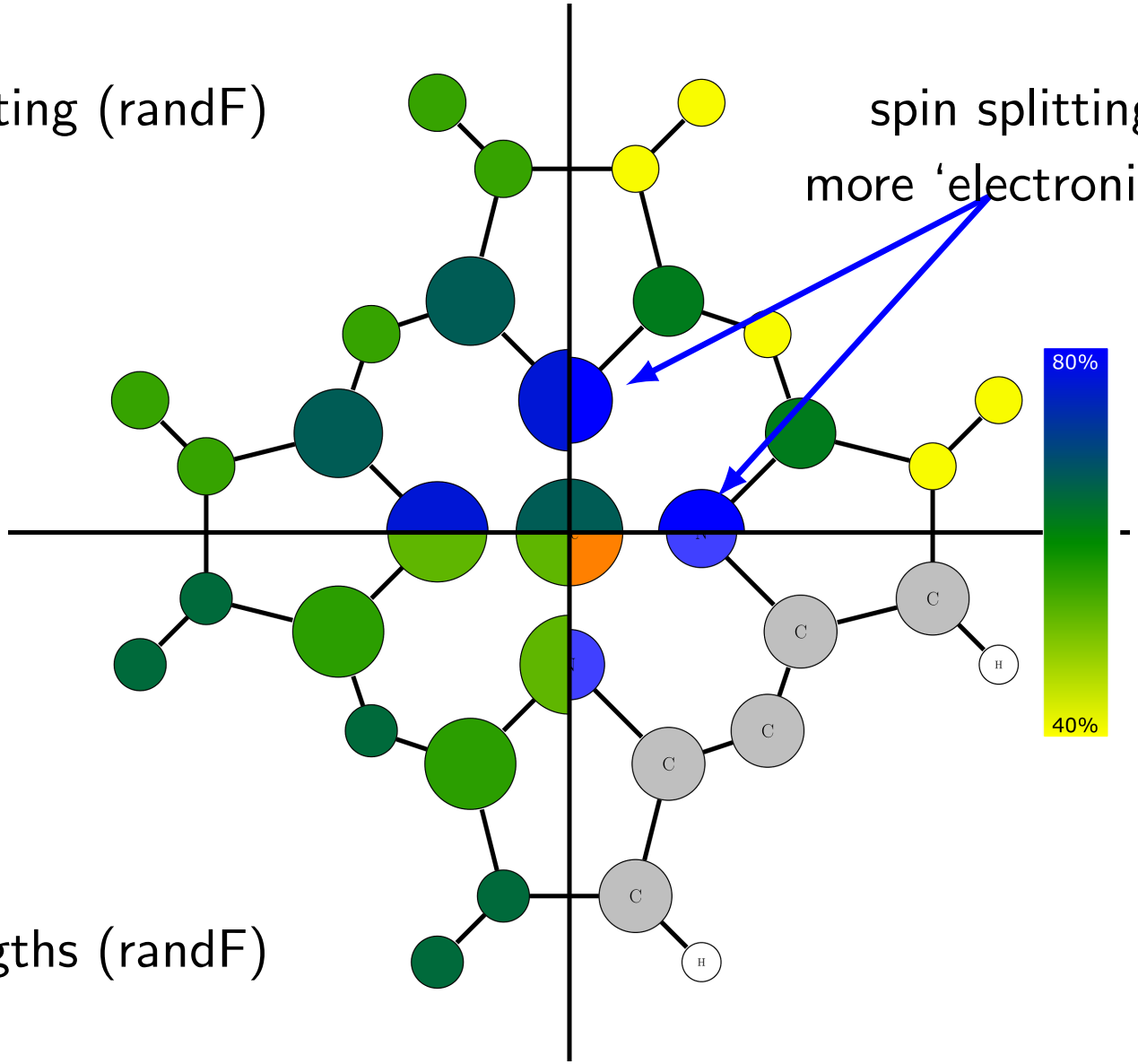


Do features depend on properties?

spin splitting (randF)

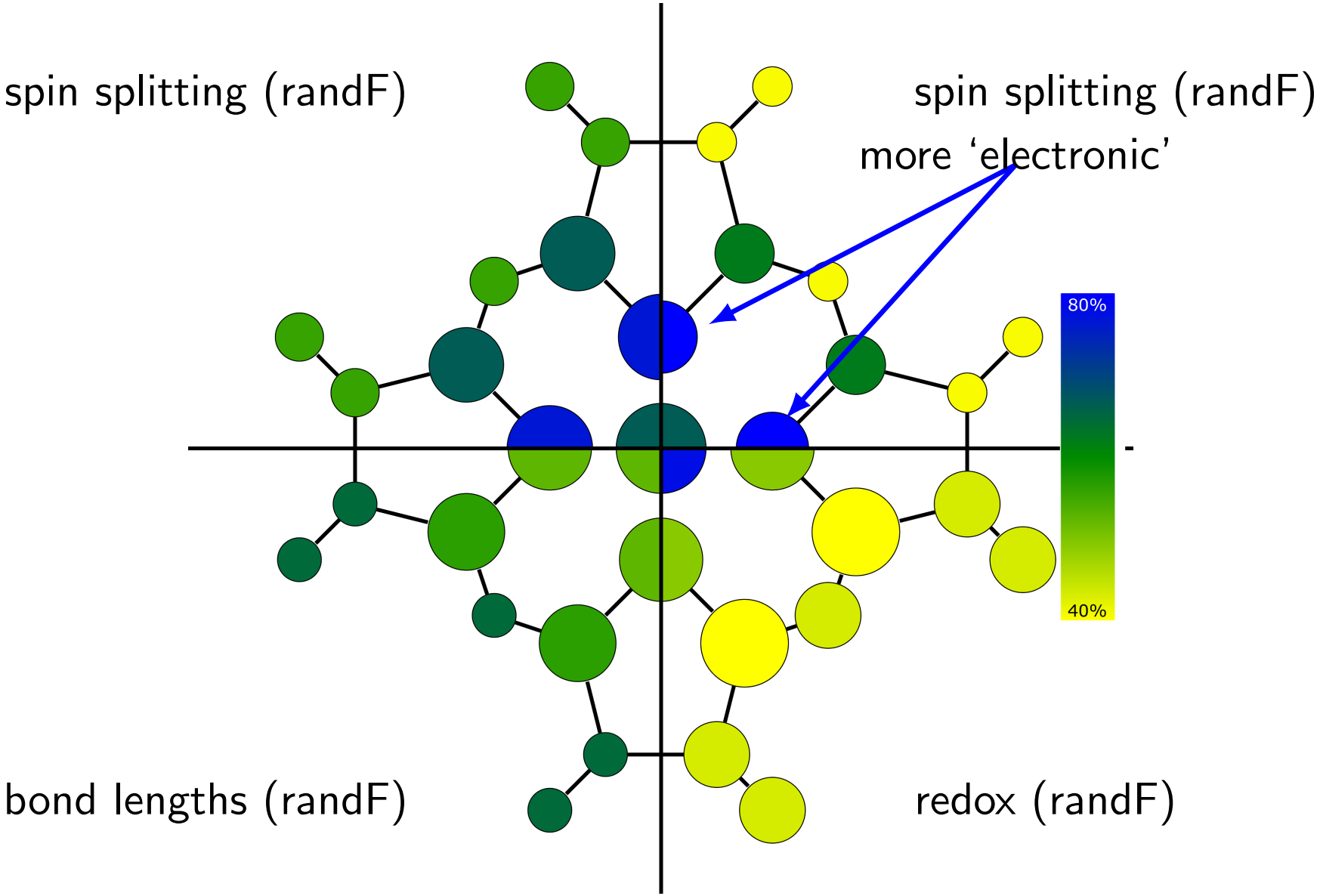
spin splitting (randF)

more 'electronic'

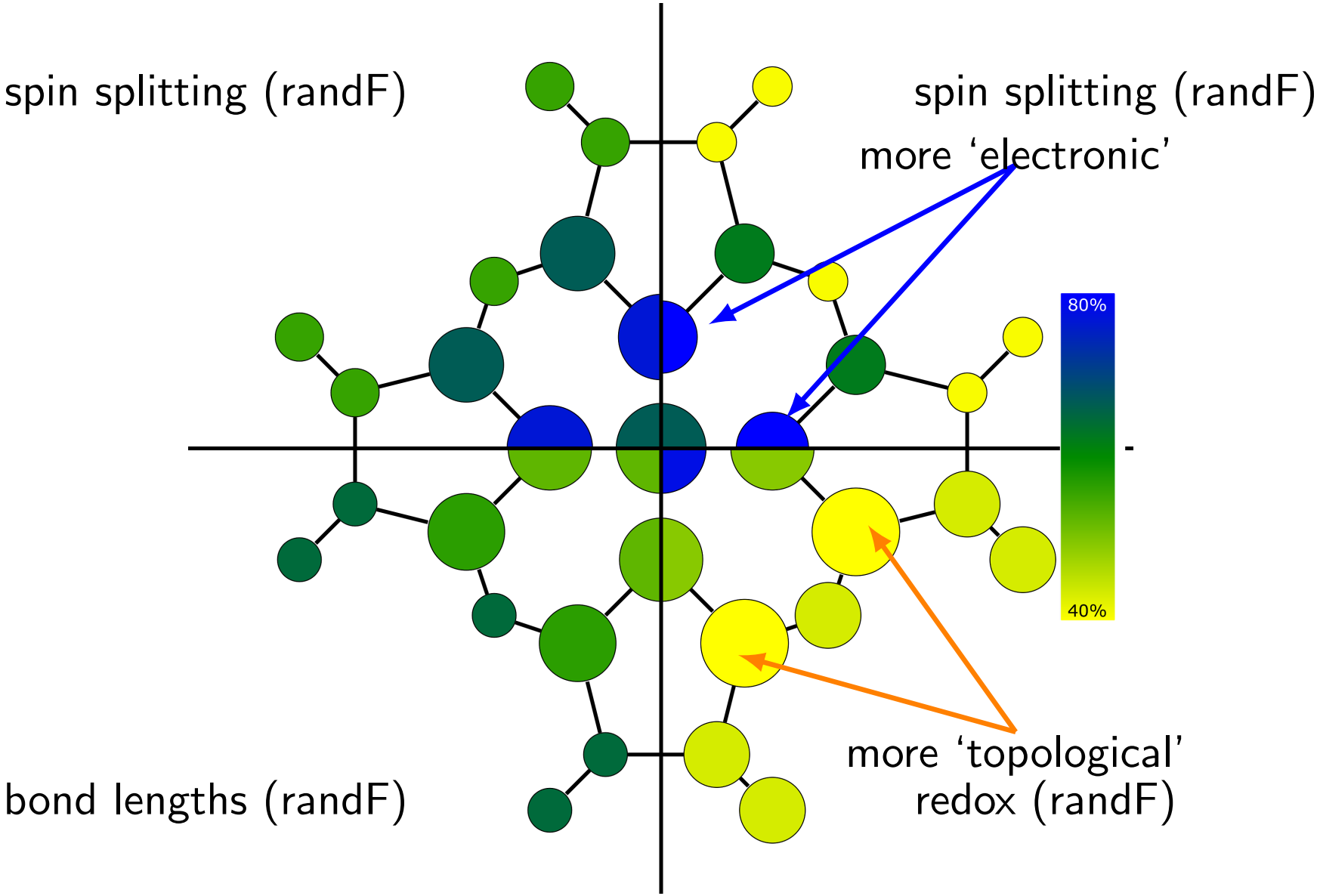


bond lengths (randF)

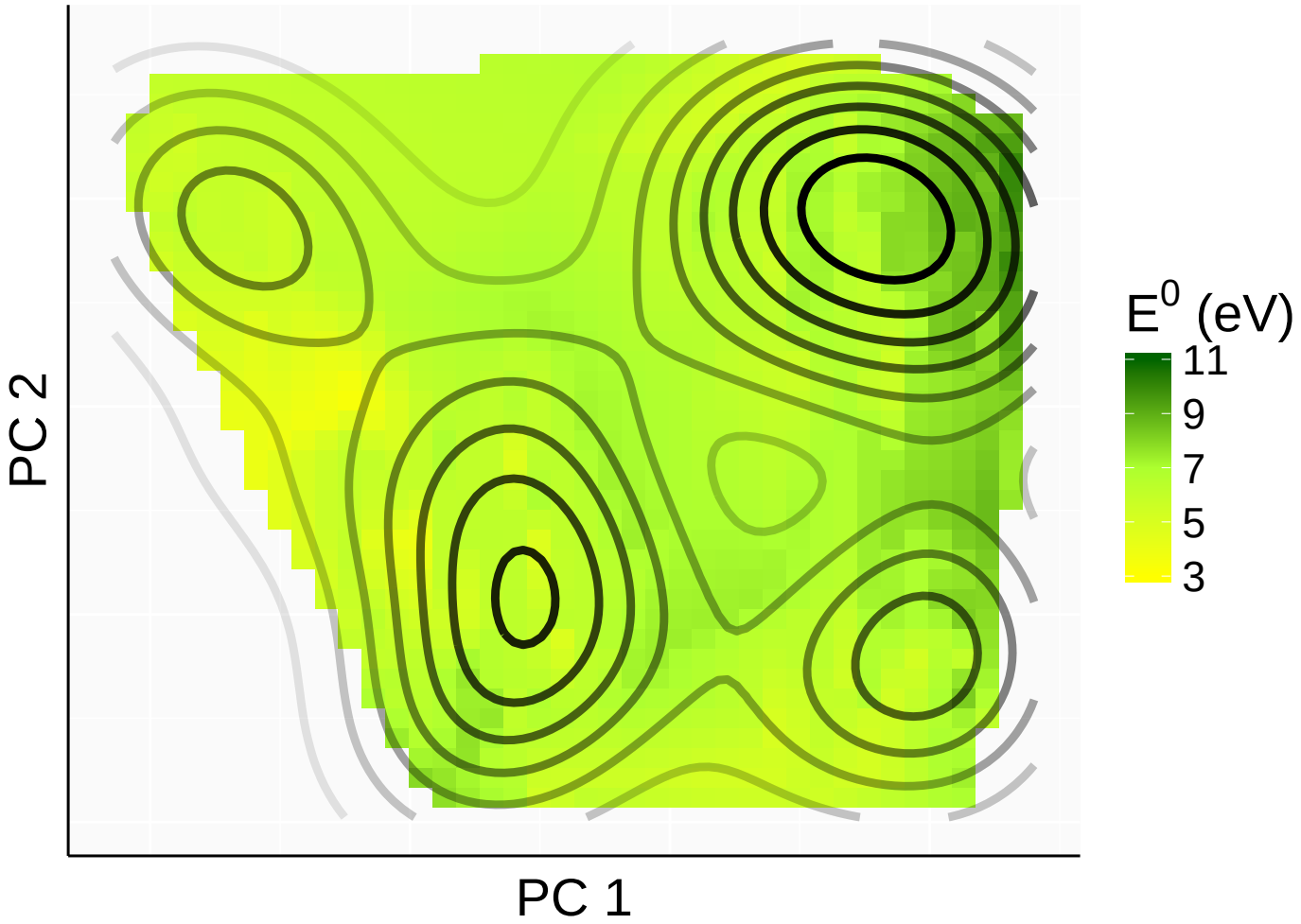
Do features depend on properties?



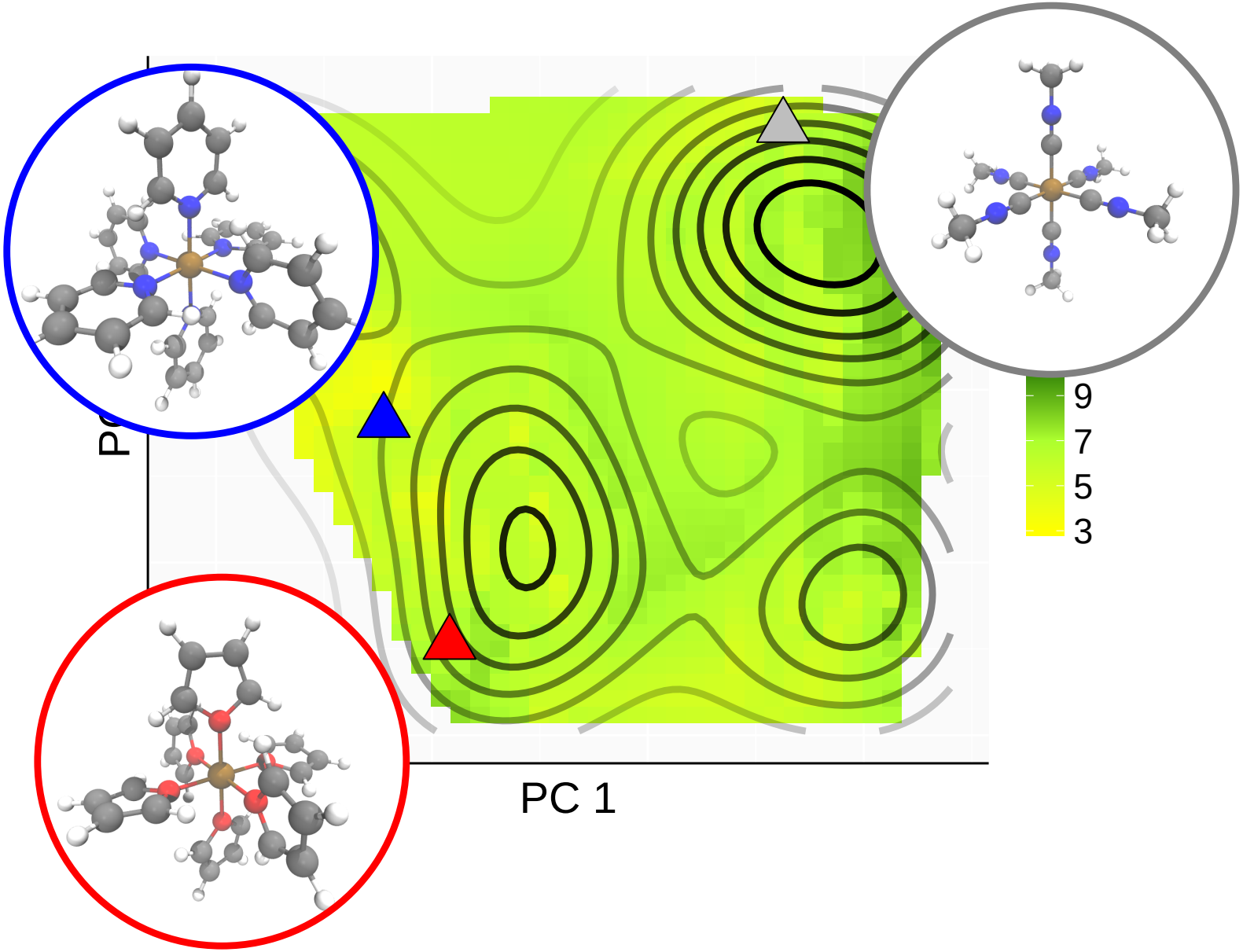
Do features depend on properties?



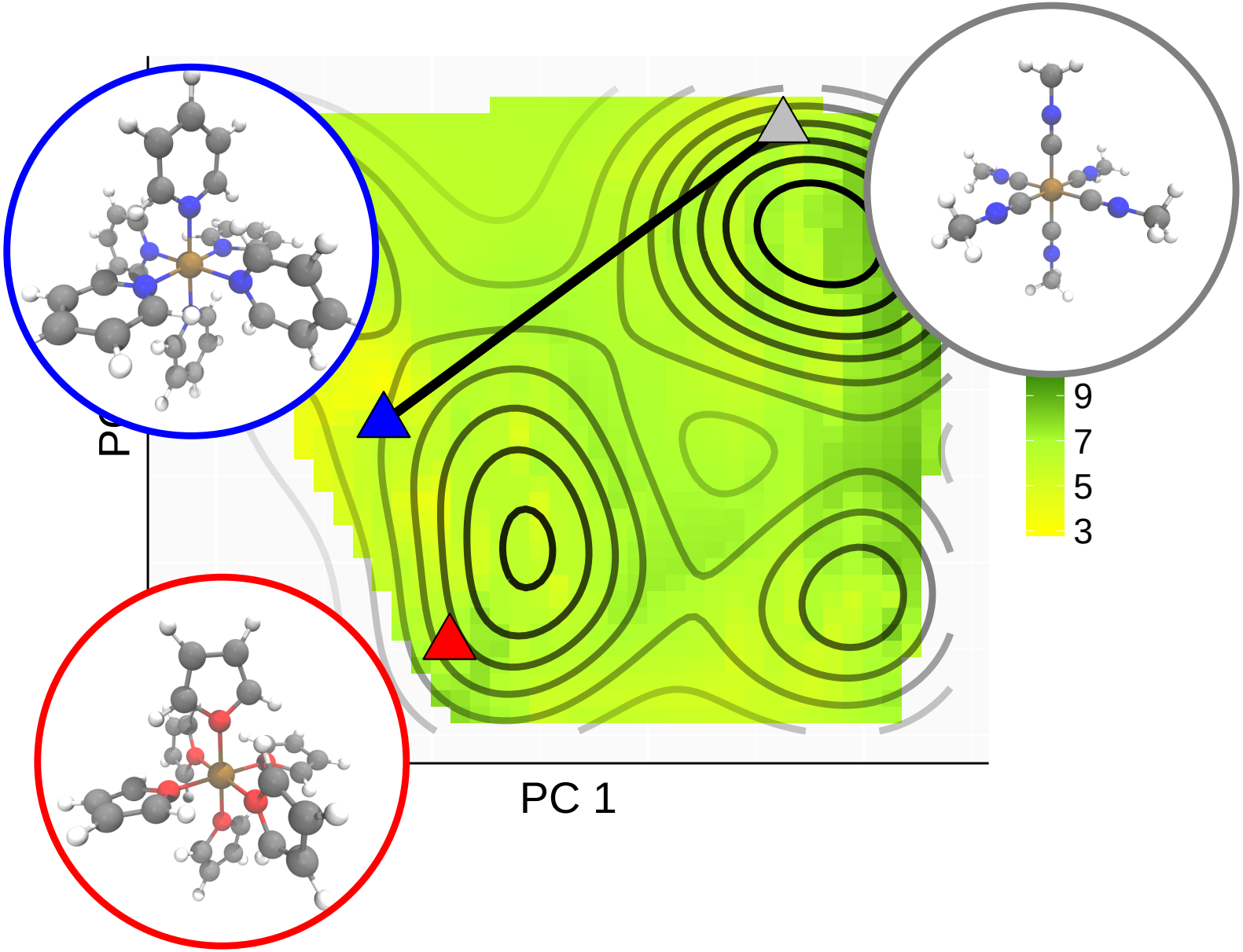
Mapping TM complex space



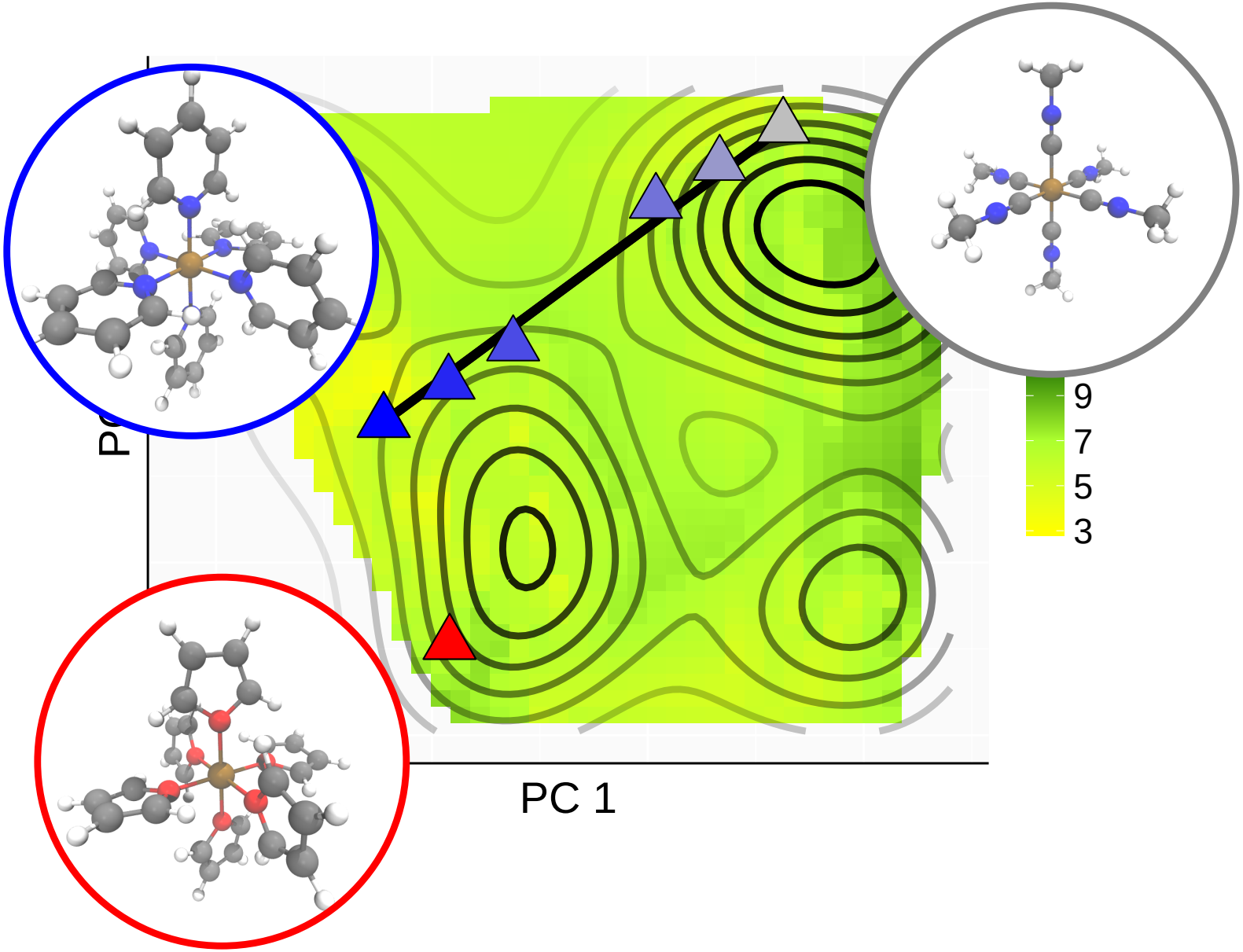
Mapping TM complex space



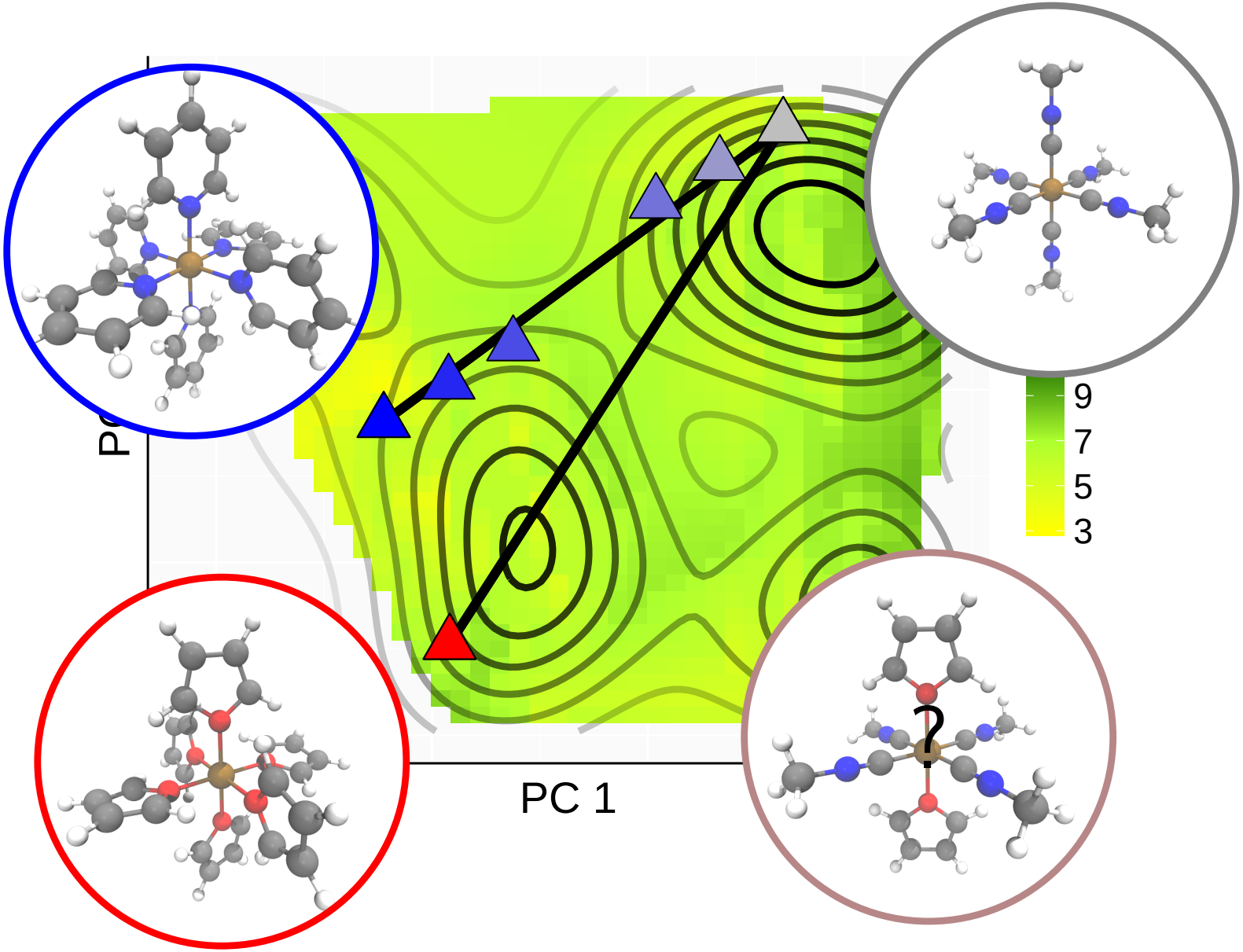
Mapping TM complex space



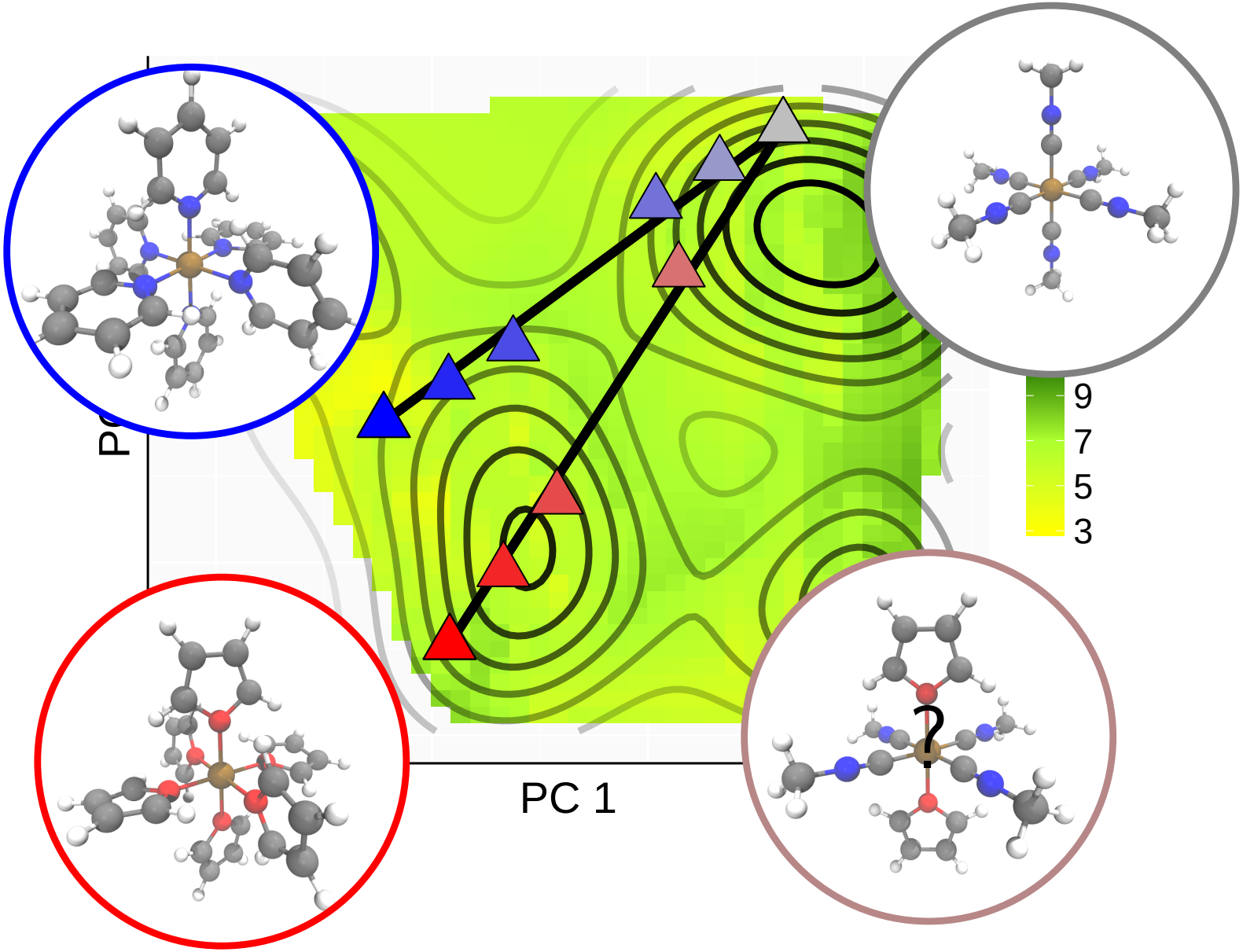
Mapping TM complex space



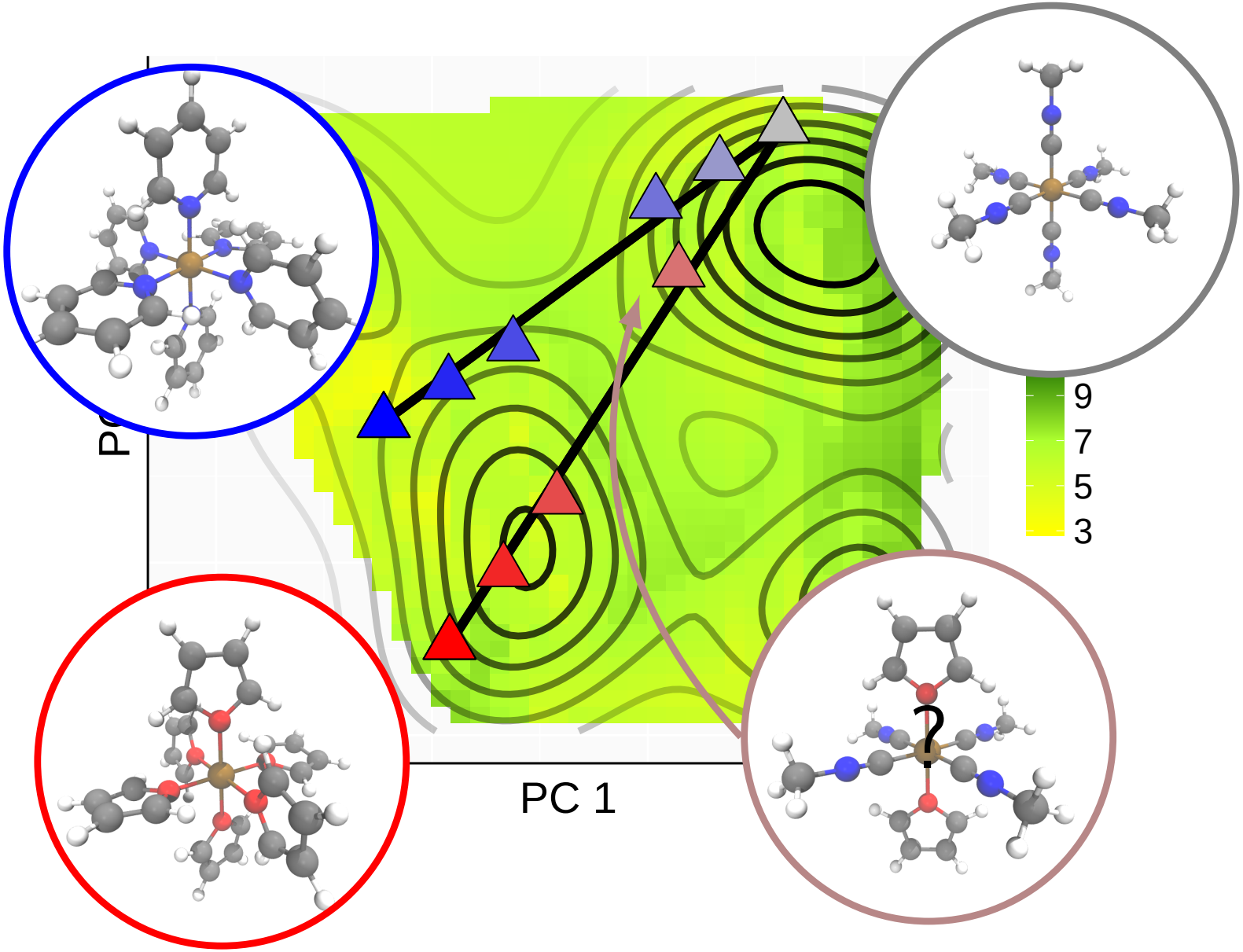
Mapping TM complex space



Mapping TM complex space



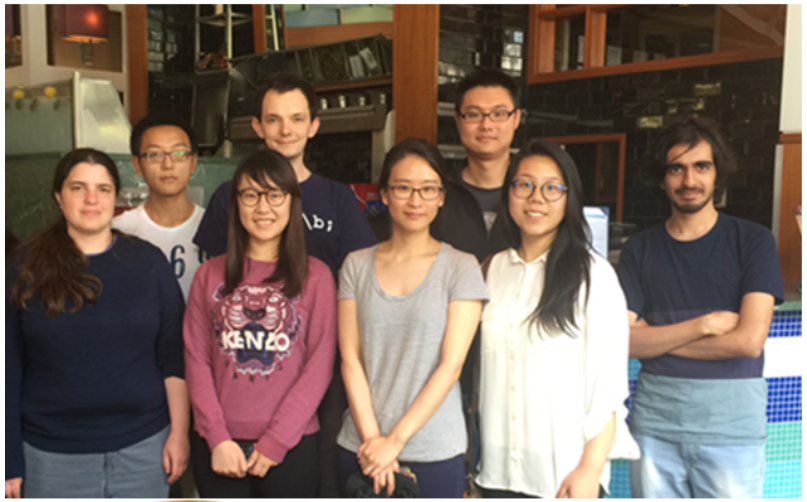
Mapping TM complex space



Acknowledgments

Conclusions:

- machine learning TM complexes faces unique challenges
- ACs are a promising starting point for low-cost descriptors
- different target properties depend on different physical variables and we can gain design insights but figuring out which



MITei
eni
Office of Naval Research ONR Science & Technology
BURROUGHS WELLCOME FUND
NSF