Hybrid machine-learning and first-principles design for transition metal complexes

Jon Paul Janet [1] Chenru Duan[2] Aditya Nandy[2]
Heather Kulik [1]

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

[2]Department of Chemistry, Massachusetts Institute of Technology

Foundational & Applied Data Science for Molecular and Material Science & Engineering

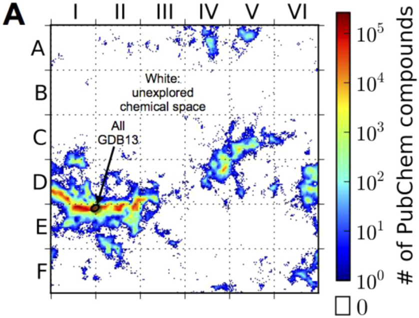
05.22.19

Motivation: chemical discovery

How can we design new materials using computers?

The space of possible chemistries is incredibly vast, with $\mathcal{O}(10^{60})$ small organic molecules.

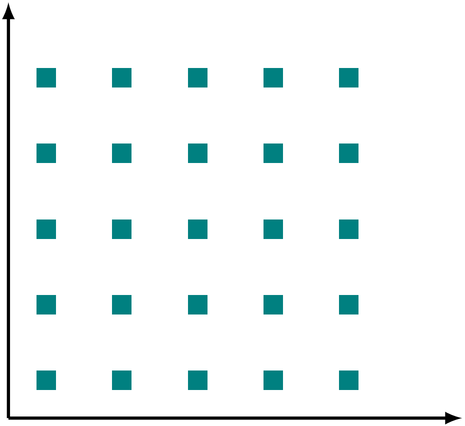
All potentially undiscovered medicines, catalysts and materials are somewhere, out in this huge space.



Virshup *et al.*, *J. Am. Chem. Soc.*, 135(19): 7296–7303, 2013.

Motivation: chemical discovery

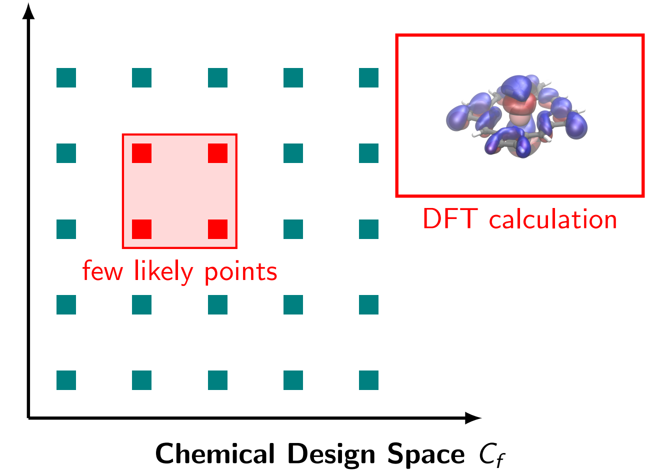
How can we design new materials using computers?



Chemical Design Space C_f

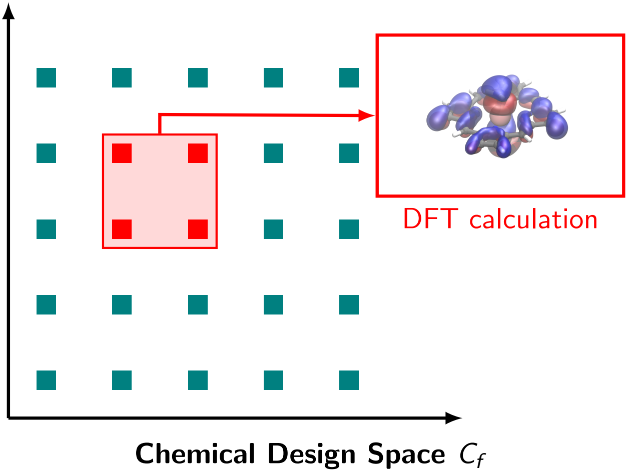
Motivation: chemical discovery

How can we design new materials using computers?



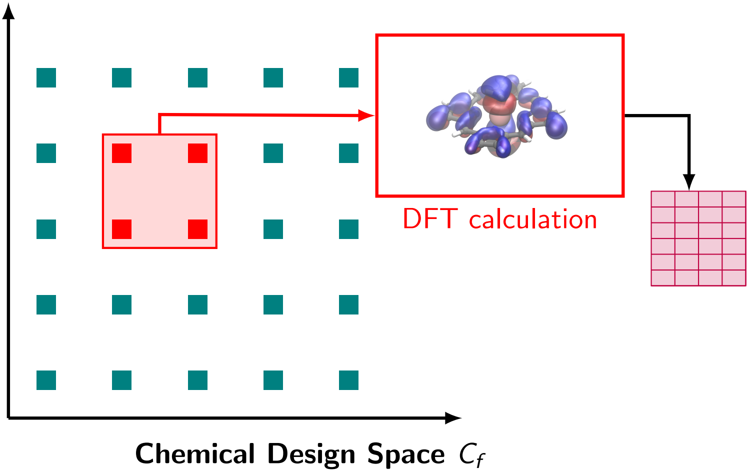
Motivation: chemical discovery

How can we design new materials using computers?



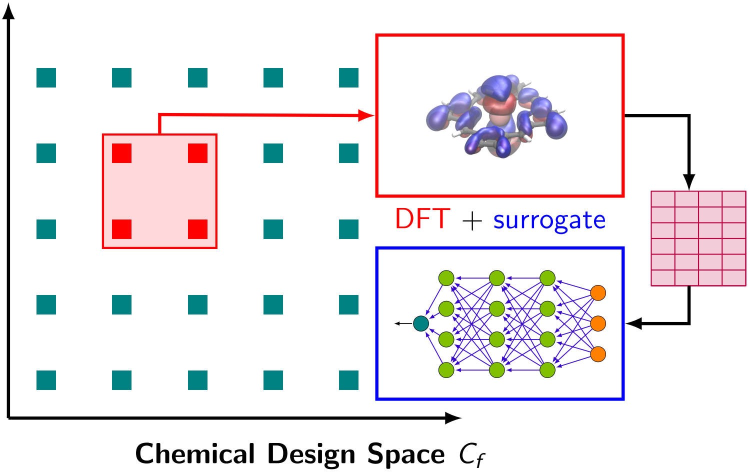
Motivation: chemical discovery

How can we design new materials using computers?



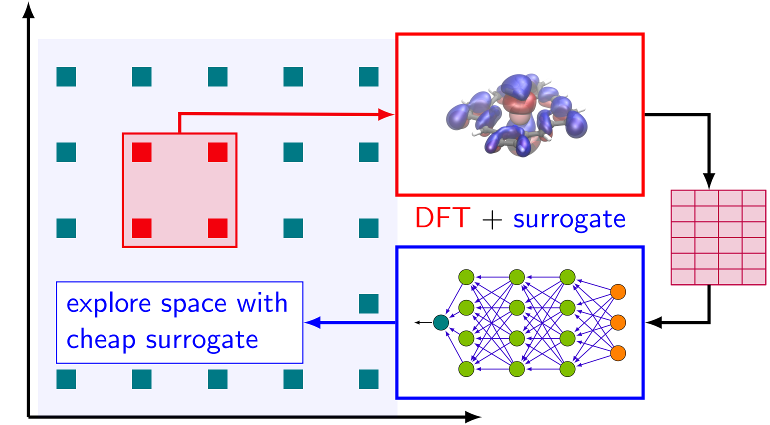
Motivation: chemical discovery

How can we design new materials using computers?



Motivation: chemical discovery

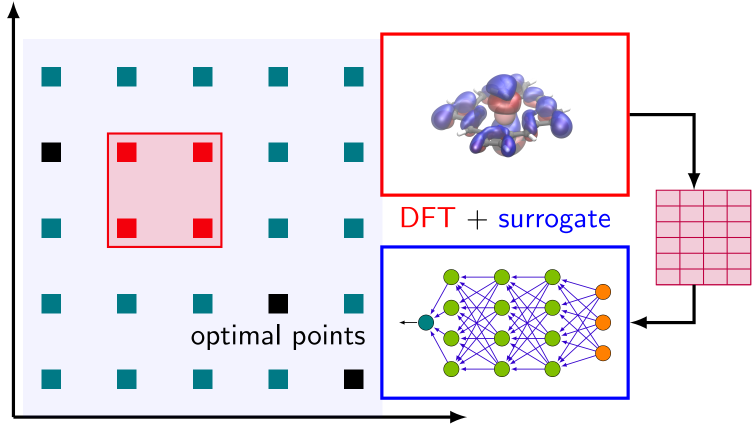
How can we design new materials using computers?



Chemical Design Space C_f

Motivation: chemical discovery

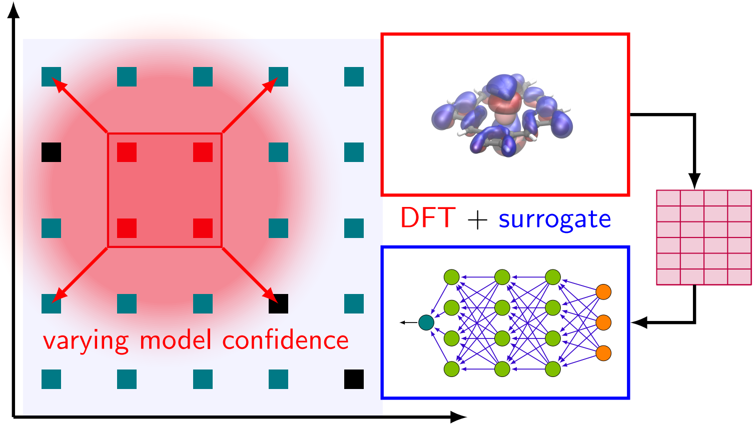
How can we design new materials using computers?



Chemical Design Space C_f

Motivation: chemical discovery

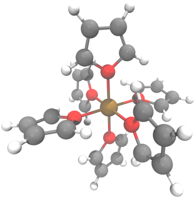
How can we design new materials using computers?



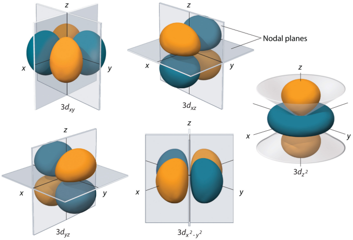
Chemical Design Space C_f

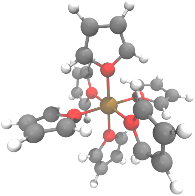
Transition metal complexes

Transition metal complexes

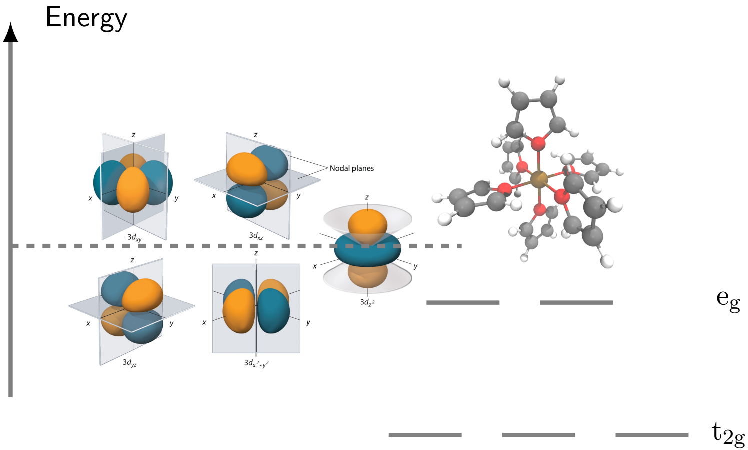


Transition metal complexes

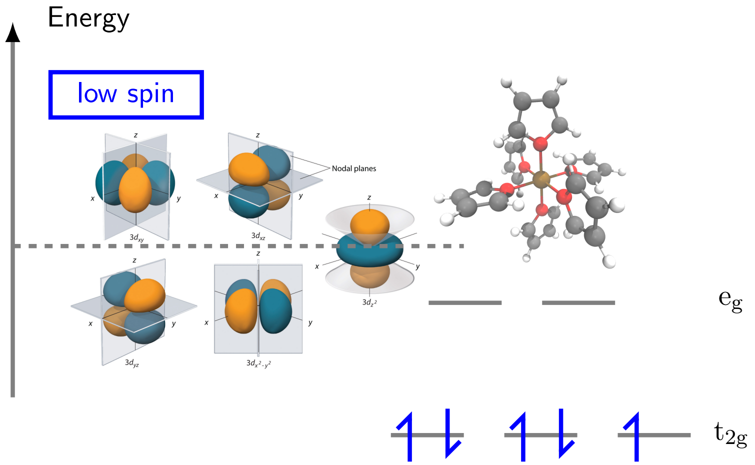




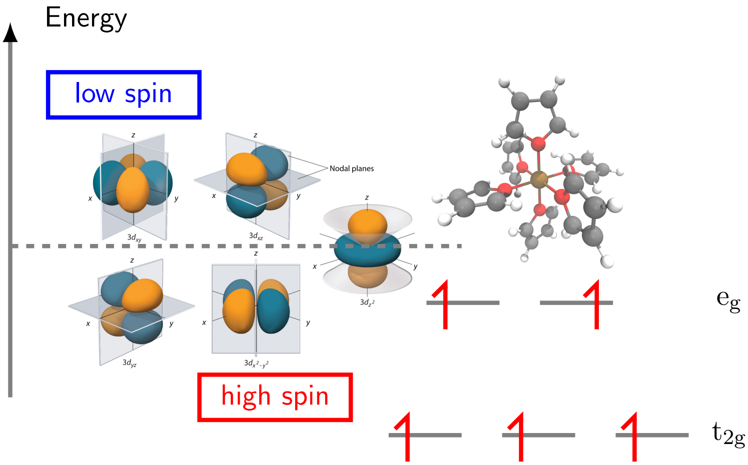
Transition metal complexes



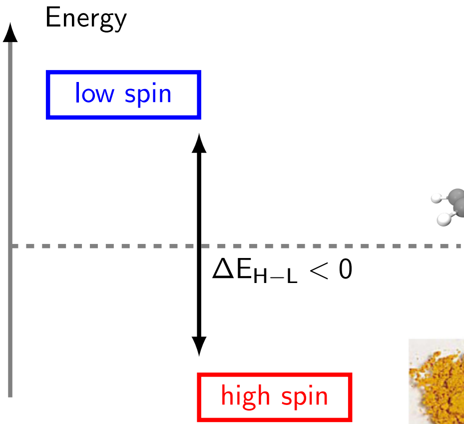
Transition metal complexes

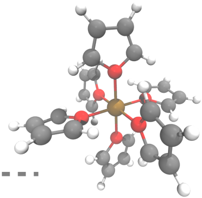


Transition metal complexes

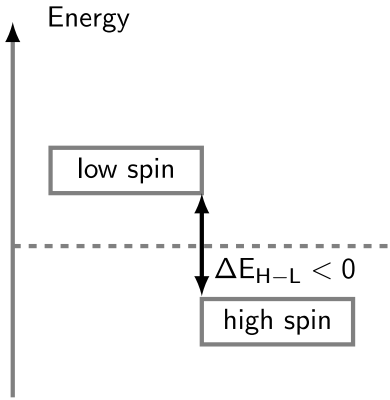


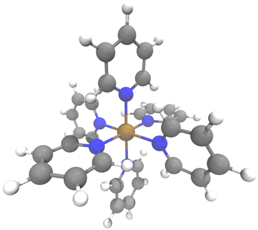
Transition metal complexes



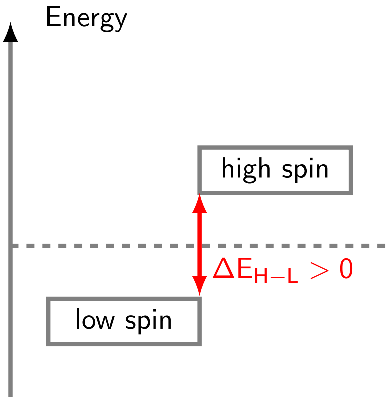


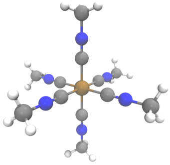
Transition metal complexes



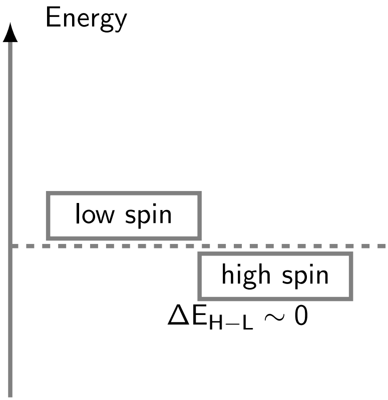


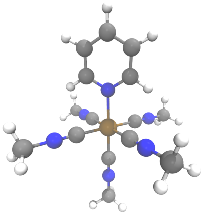
Transition metal complexes



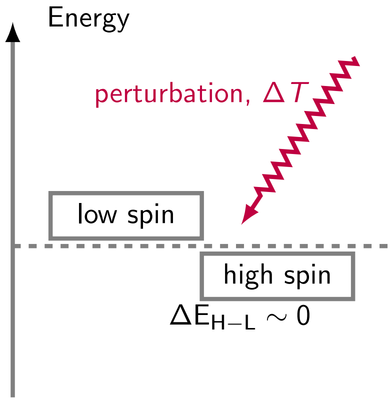


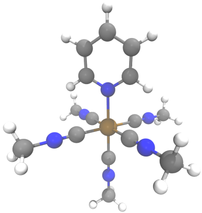
Transition metal complexes



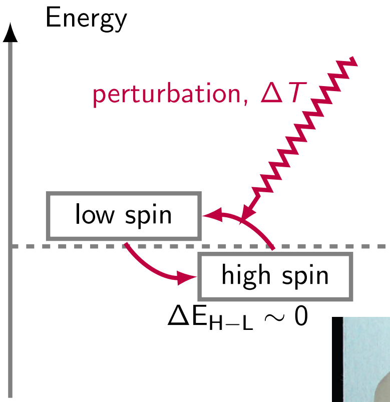


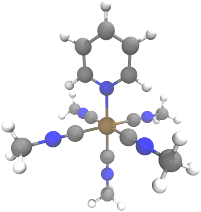
Transition metal complexes

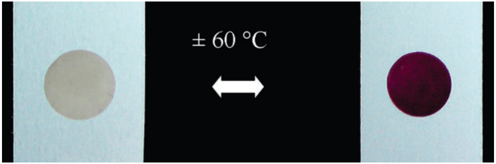




Transition metal complexes







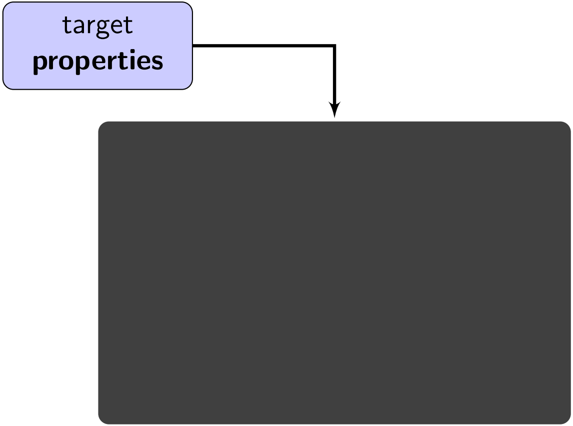
Seredyuk, M *et al.*, *Chem. Mater.*, 18(10):2513–2519, 2006.

The dream of automated design

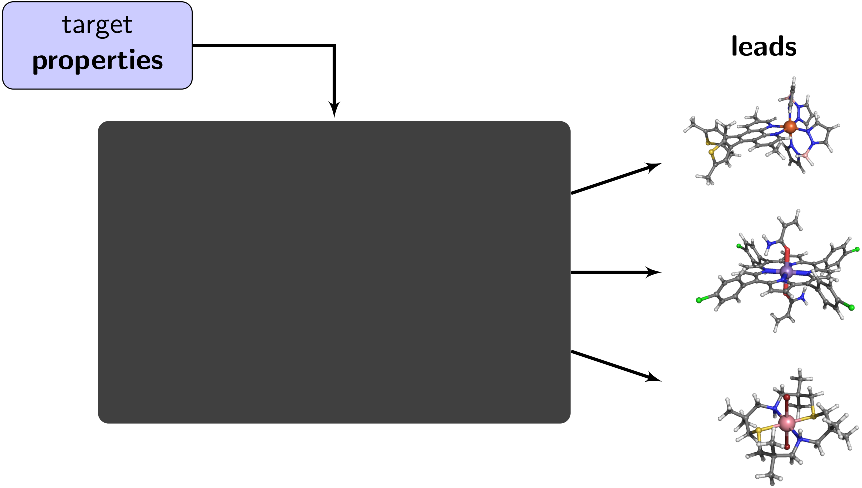
The dream of automated design

target
properties

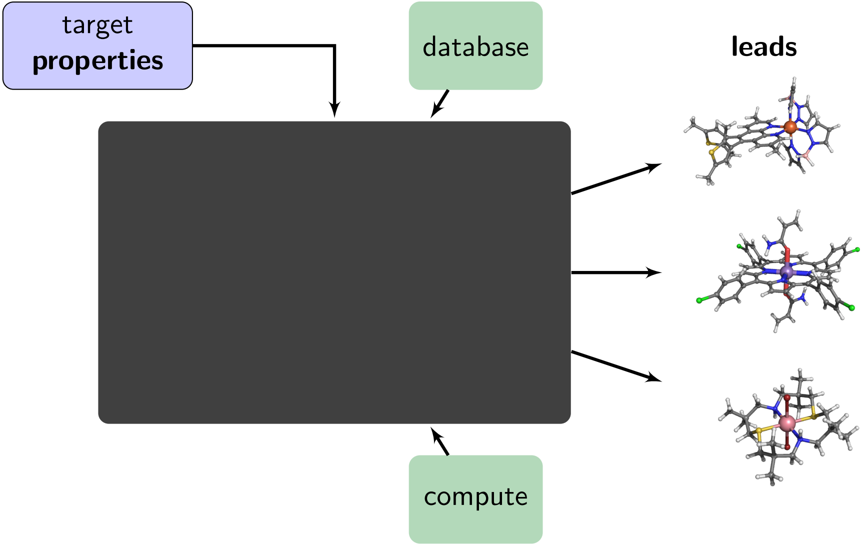
The dream of automated design



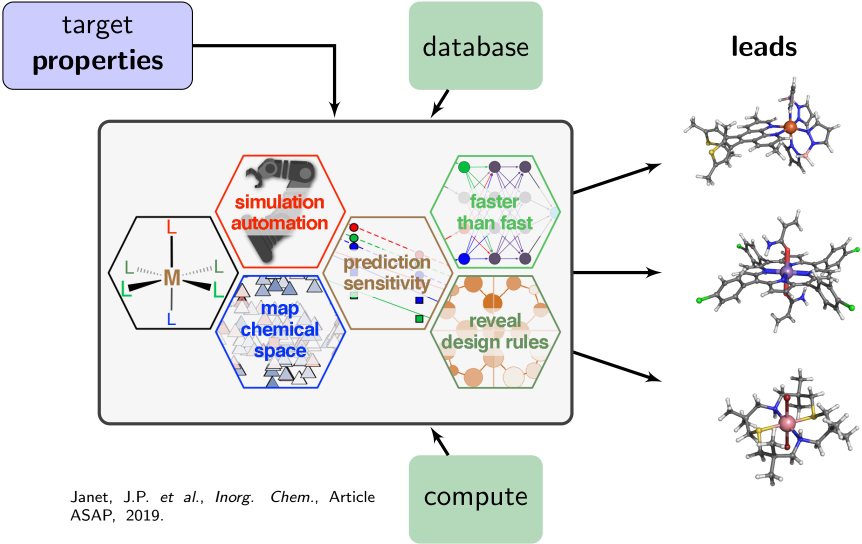
The dream of automated design



The dream of automated design



The dream of automated design



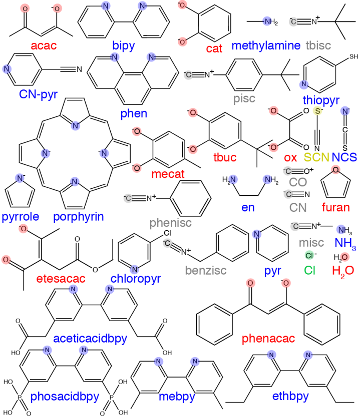
Janet, J.P. *et al.*, *Inorg. Chem.*, Article ASAP, 2019.

Quantum simulation of TM complexes

train on $\sim$ 100–2000 DFT calculations:

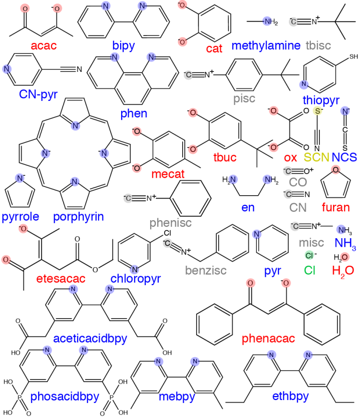
Quantum simulation of TM complexes

train on ∼ 100–2000 DFT calculations:



Quantum simulation of TM complexes

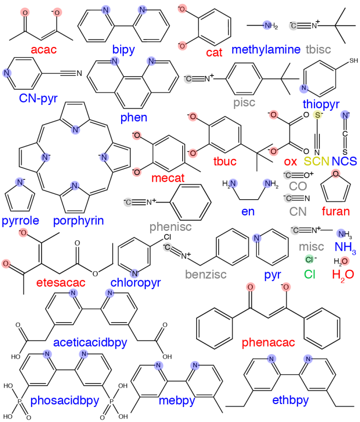
train on ∼ 100–2000 DFT calculations:

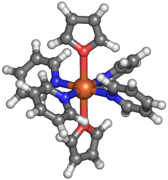


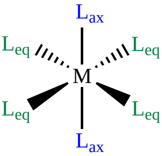
Cr	Mn	Fe	Co

Quantum simulation of TM complexes

train on $\sim$ 100–2000 DFT calculations:



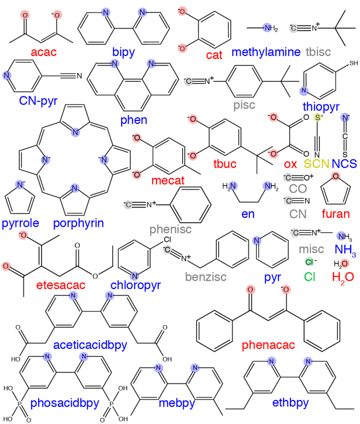


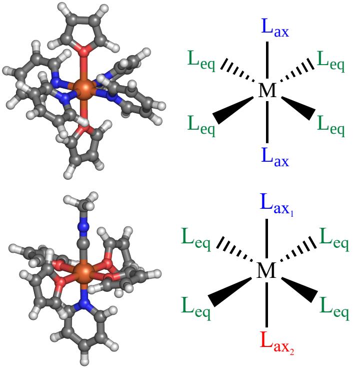


Cr	Mn	Fe	Co

Quantum simulation of TM complexes

train on $\sim$ 100–2000 DFT calculations:



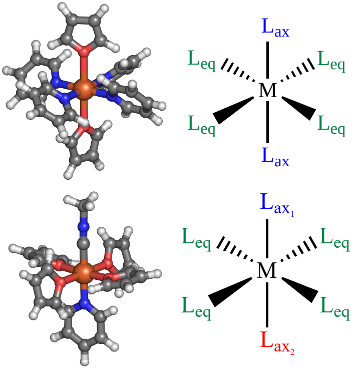


Cr	Mn	Fe	Co

Quantum simulation of TM complexes

train on ∼ 100–2000 DFT calculations:

Details:
B3LYP-like DFT
gas phase optimizaton
LANL2DZ/6-31G*
high- and low-spin
M(II)/(III)



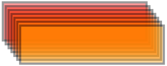
Cr	Mn	Fe	Co

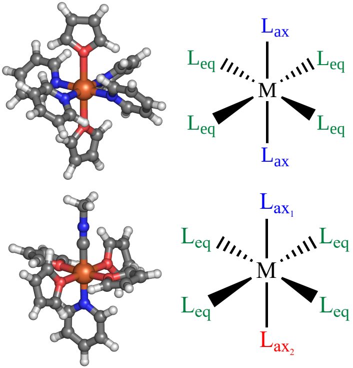
Quantum simulation of TM complexes

train on ∼ 100–2000 DFT calculations:

Details:
B3LYP-like DFT
gas phase optimizaton
LANL2DZ/6-31G*
high- and low-spin
M(II)/(III)

HF exchange varied 0–30%





Cr	Mn	Fe	Co

Modeling of TM complexes with heuristic representations

First attempt using simple features inspired by inorganic chem:

Modeling of TM complexes with heuristic representations

First attempt using simple features inspired by inorganic chem:

metal properties

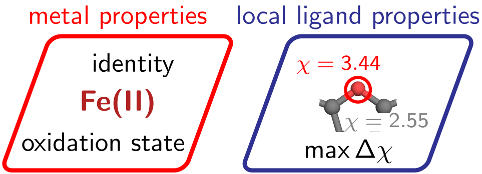
identity

Fe(II)

oxidation state

Modeling of TM complexes with heuristic representations

First attempt using simple features inspired by inorganic chem:



Modeling of TM complexes with heuristic representations

First attempt using simple features inspired by inorganic chem:

metal properties

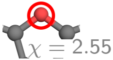
identity

Fe(II)

oxidation state

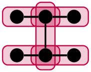
local ligand properties

$\chi = 3.44$



max $\Delta\chi$

global properties

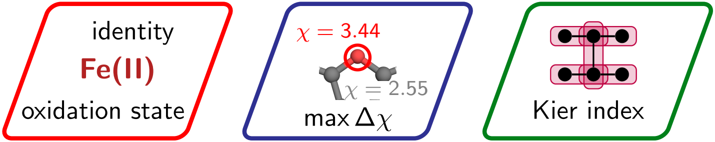


Kier index

Modeling of TM complexes with heuristic representations

First attempt using simple features inspired by inorganic chem:

metal properties local ligand properties global properties

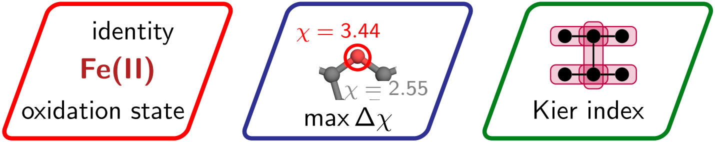


mixed continous discrete ligand-centered: **MCDL-25**

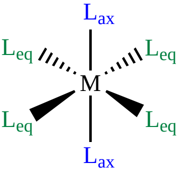
Modeling of TM complexes with heuristic representations

First attempt using simple features inspired by inorganic chem:

metal properties local ligand properties global properties



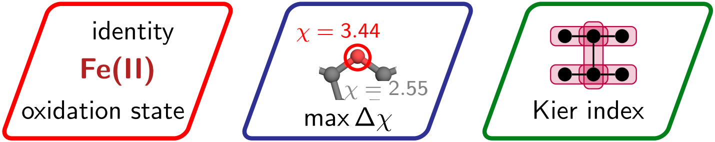
mixed continous discrete ligand-centered: **MCDL-25**



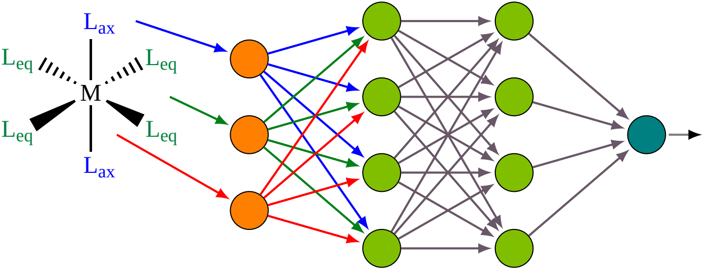
Modeling of TM complexes with heuristic representations

First attempt using simple features inspired by inorganic chem:

metal properties — local ligand properties — global properties



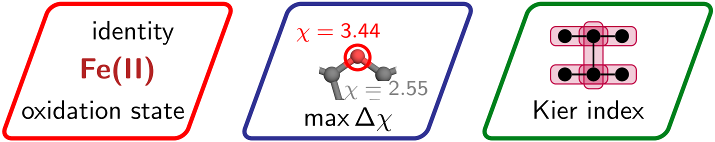
fully-connected 2-layer ANN, dropout regularization

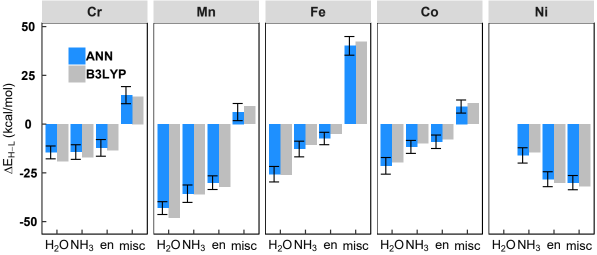


Modeling of TM complexes with heuristic representations

First attempt using simple features inspired by inorganic chem:

metal properties | local ligand properties | global properties

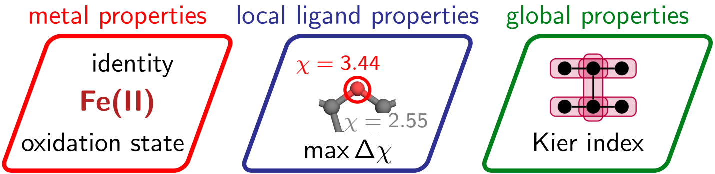


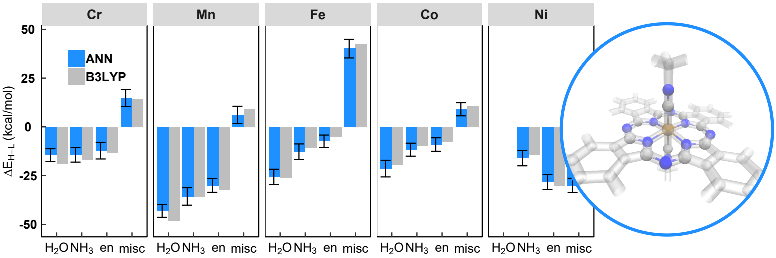


Janet, J.P. and Kulik, H.J., *Chem. Sci.*, 8:5137–5152, 2017.

Modeling of TM complexes with heuristic representations

First attempt using simple features inspired by inorganic chem:

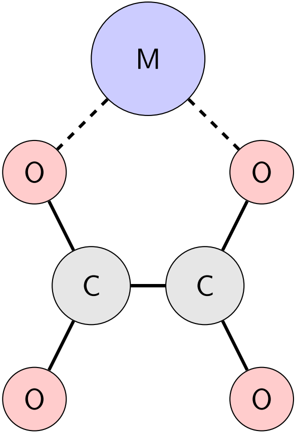




Janet, J.P. and Kulik, H.J., *Chem. Sci.*, 8:5137–5152, 2017.

New descriptors – RACs

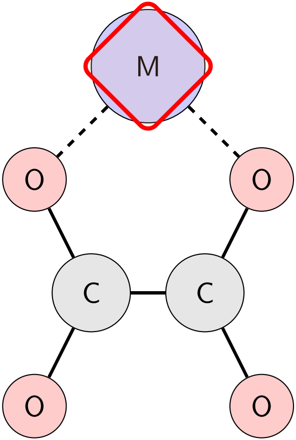
Proposed a new graph-based set of descriptors for TM complexes[1]



[1] Janet, J.P., and Kulik, H.J., *J. Phys. Chem. A*, 121(46):8939–8954, 2017.

New descriptors – RACs

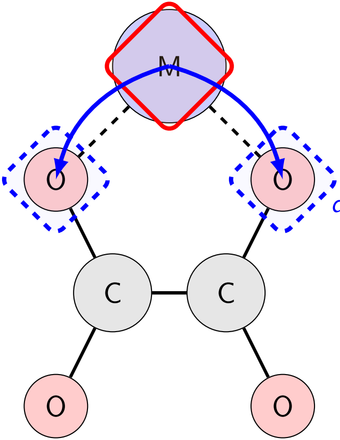
Proposed a new graph-based set of descriptors for TM complexes[1]



[1] Janet, J.P., and Kulik, H.J., *J. Phys. Chem. A*, 121(46):8939–8954, 2017.

New descriptors – RACs

Proposed a new graph-based set of descriptors for TM complexes[1]

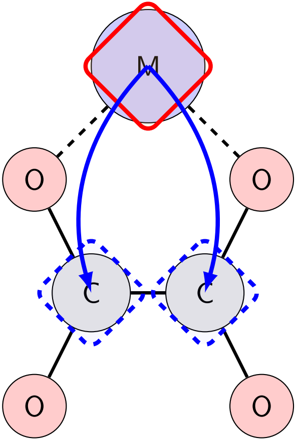


$d_1 : \sum_{M,O} Z_M Z_O$

[1] Janet, J.P., and Kulik, H.J., *J. Phys. Chem. A*, 121(46):8939–8954, 2017.

New descriptors – RACs

Proposed a new graph-based set of descriptors for TM complexes[1]



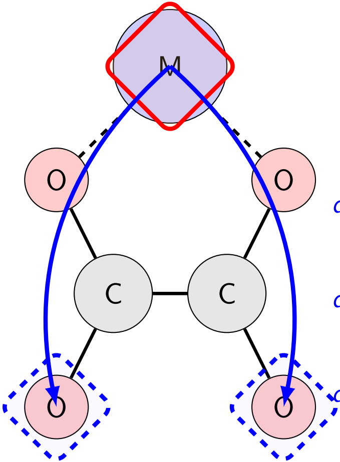
$$d_1 : \sum_{M,O} Z_M Z_O$$

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

[1] Janet, J.P., and Kulik, H.J., *J. Phys. Chem. A*, 121(46):8939–8954, 2017.

New descriptors – RACs

Proposed a new graph-based set of descriptors for TM complexes[1]



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

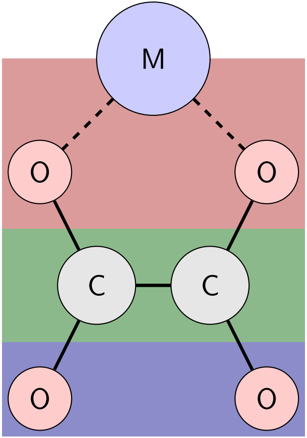
$$d_2 : \sum_{M,C} Z_M Z_C$$

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

[1] Janet, J.P., and Kulik, H.J., *J. Phys. Chem. A*, 121(46):8939–8954, 2017.

New descriptors – RACs

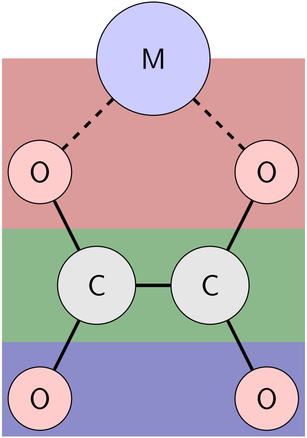
Proposed a new graph-based set of descriptors for TM complexes[1]



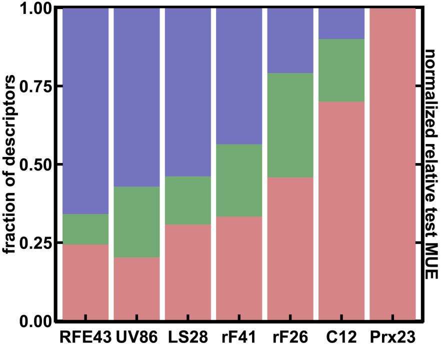
[1] Janet, J.P., and Kulik, H.J., *J. Phys. Chem. A*, 121(46):8939–8954, 2017.

New descriptors – RACs

Proposed a new graph-based set of descriptors for TM complexes[1]



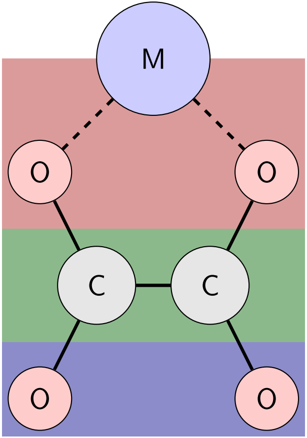
KRR models of spin splitting energies:



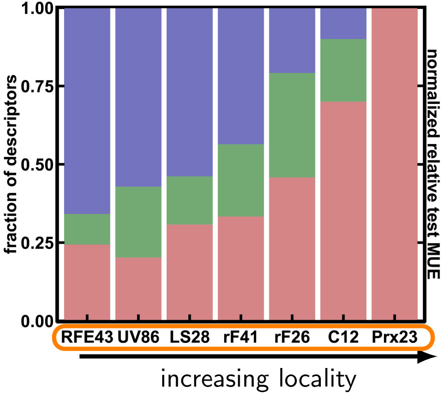
[1] Janet, J.P., and Kulik, H.J., *J. Phys. Chem. A*, 121(46):8939–8954, 2017.

New descriptors – RACs

Proposed a new graph-based set of descriptors for TM complexes[1]



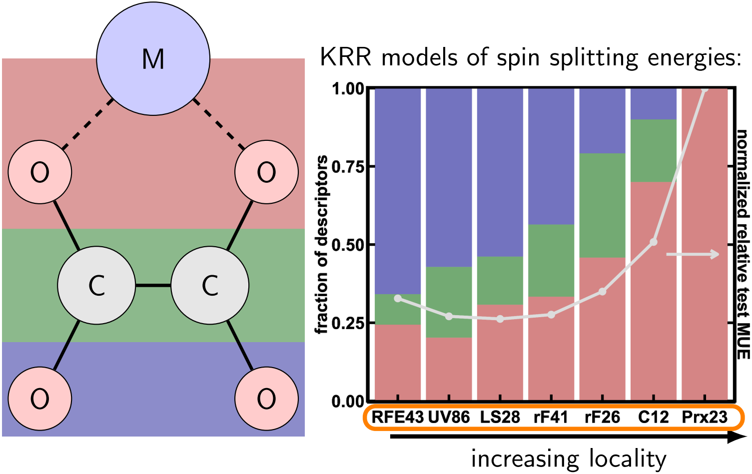
KRR models of spin splitting energies:



[1] Janet, J.P., and Kulik, H.J., *J. Phys. Chem. A*, 121(46):8939–8954, 2017.

New descriptors – RACs

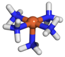
Proposed a new graph-based set of descriptors for TM complexes[1]



[1] Janet, J.P., and Kulik, H.J., *J. Phys. Chem. A*, 121(46):8939–8954, 2017.

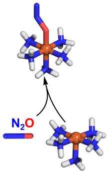
Extension to catalysis

Can we apply the same ideas to cheaply predict catalytically-relevant properties?



Extension to catalysis

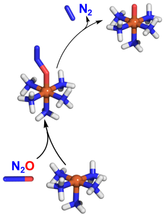
Can we apply the same ideas to cheaply predict catalytically-relevant properties?



Nandy, A. *et al.*, in preparation.

Extension to catalysis

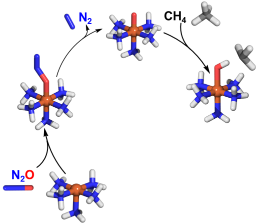
Can we apply the same ideas to cheaply predict catalytically-relevant properties?



Nandy, A. *et al.*, in preparation.

Extension to catalysis

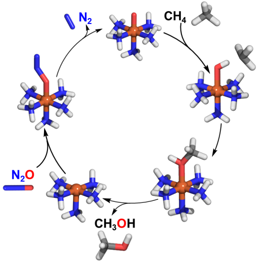
Can we apply the same ideas to cheaply predict catalytically-relevant properties?



Nandy, A. *et al.*, in preparation.

Extension to catalysis

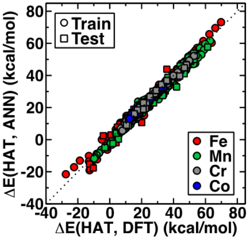
Can we apply the same ideas to cheaply predict catalytically-relevant properties?

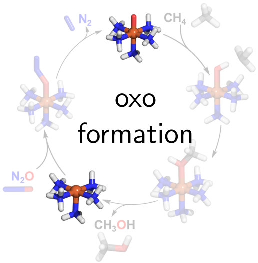


Nandy, A. *et al.*, in preparation.

Extension to catalysis

Can we apply the same ideas to cheaply predict catalytically-relevant properties?

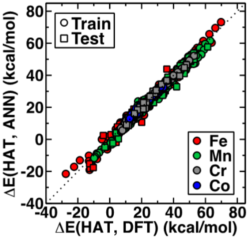


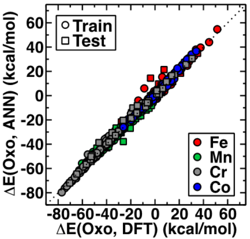


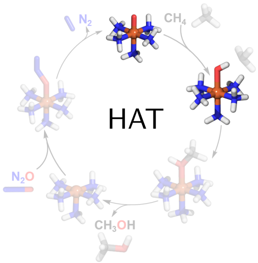
Nandy, A. *et al.*, in preparation.

Extension to catalysis

Can we apply the same ideas to cheaply predict catalytically-relevant properties?

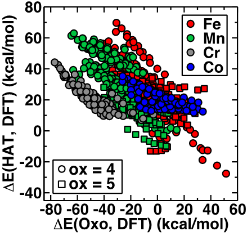


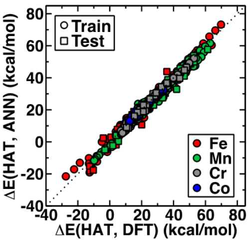


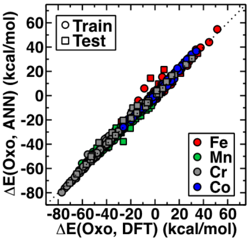


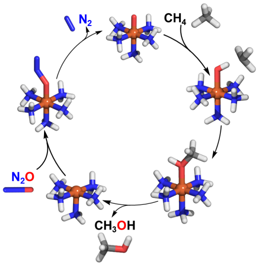
Nandy, A. *et al.*, in preparation.

Extension to catalysis









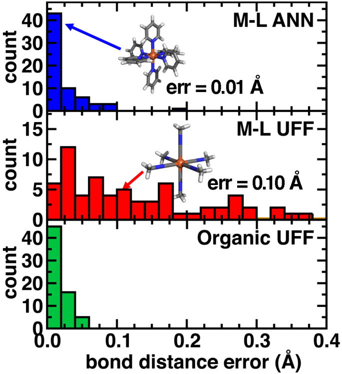
Nandy, A. *et al.*, in preparation.

Machine learning job initialization

Metal-ligand bonding is difficult to resolve without QM:

Machine learning job initialization

Metal-ligand bonding is difficult to resolve without QM:

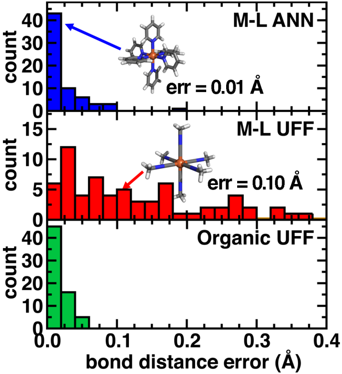


we can predict
bond lengths

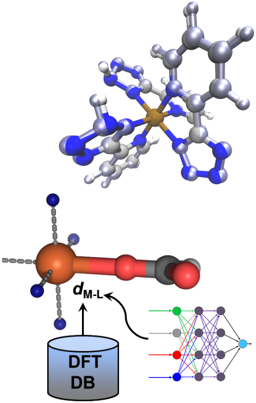
Janet, J.P. and Kulik, H.J., *Chem. Sci.*, 8:5137–5152, 2017.
Janet, J.P. *et al.*, *Ind. Eng. Chem. Res.*, 56(17):4898–4910, 2017.
Janet, J.P. *et al.*, *Inorg. Chem.*, Article ASAP, 2019.

Machine learning job initialization

Metal-ligand bonding is difficult to resolve without QM:



we can predict bond lengths and use this to intialize new calculations



Janet, J.P. and Kulik, H.J., *Chem. Sci.*, 8:5137–5152, 2017.
Janet, J.P. *et al.*, *Ind. Eng. Chem. Res.*, 56(17):4898–4910, 2017.
Janet, J.P. *et al.*, *Inorg. Chem.*, Article ASAP, 2019.

Beyond prediction: live job management

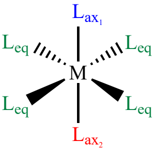
However, even with this, DFT job failure is a frequent issue:

Duan, C., Janet, J.P. *et al.*, *J. Chem. Theory. Comp.*, 15(4):2331—2345, 2019.

Beyond prediction: live job management

However, even with this, DFT job failure is a frequent issue:

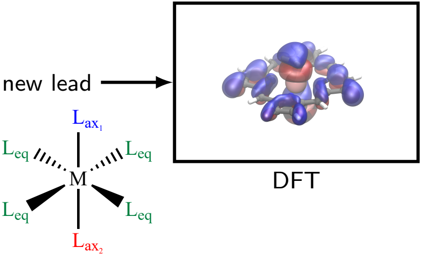
new lead



Duan, C., Janet, J.P. *et al.*, *J. Chem. Theory. Comp.*, 15(4):2331—2345, 2019.

Beyond prediction: live job management

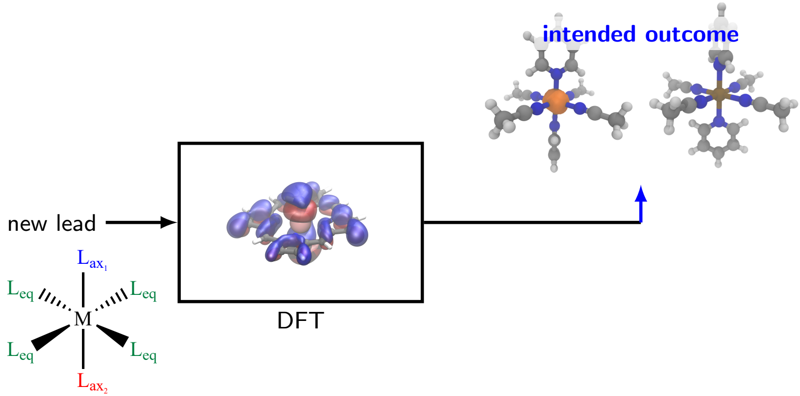
However, even with this, DFT job failure is a frequent issue:



Duan, C., Janet, J.P. *et al.*, *J. Chem. Theory. Comp.*, 15(4):2331—2345, 2019.

Beyond prediction: live job management

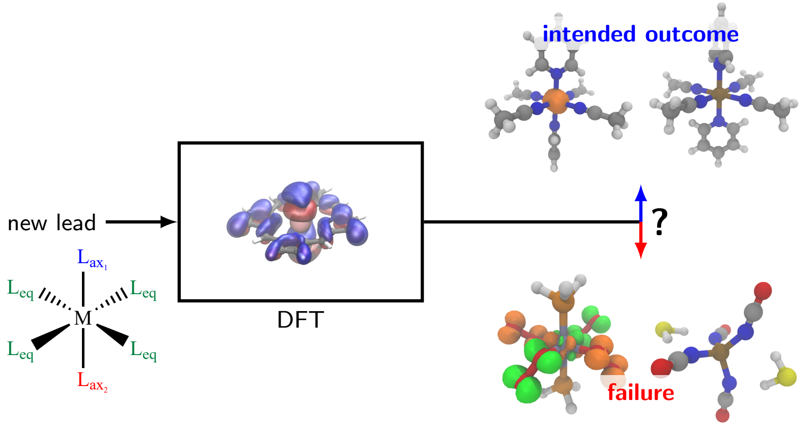
However, even with this, DFT job failure is a frequent issue:



Duan, C., Janet, J.P. *et al.*, *J. Chem. Theory. Comp.*, 15(4):2331—2345, 2019.

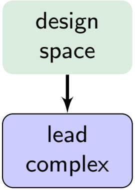
Beyond prediction: live job management

However, even with this, DFT job failure is a frequent issue:



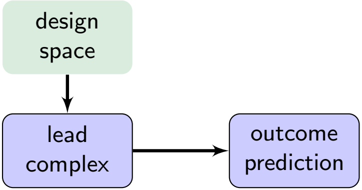
Duan, C., Janet, J.P. *et al.*, *J. Chem. Theory. Comp.*, 15(4):2331—2345, 2019.

Beyond prediction: live job management



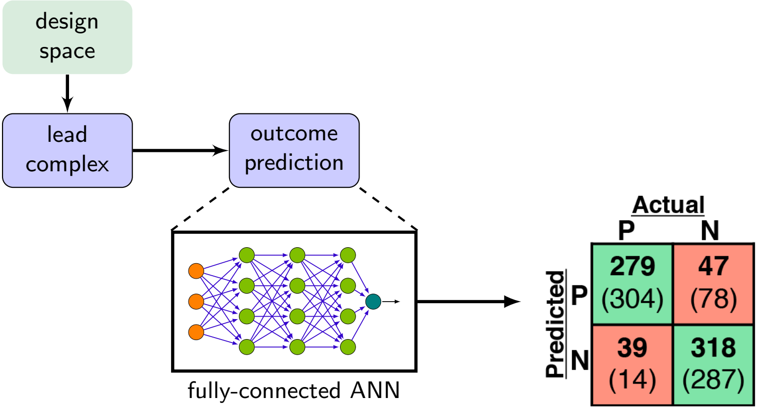
Duan, C., Janet, J.P. *et al.*, *J. Chem. Theory. Comp.*, 15(4):2331—2345, 2019.

Beyond prediction: live job management



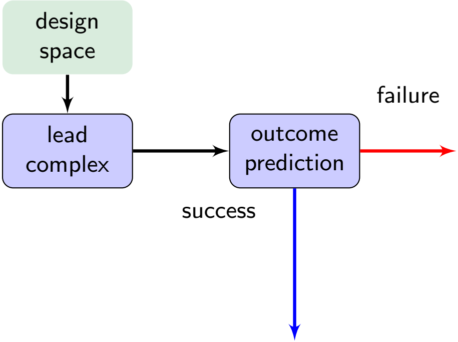
Duan, C., Janet, J.P. *et al.*, *J. Chem. Theory. Comp.*, 15(4):2331—2345, 2019.

Beyond prediction: live job management



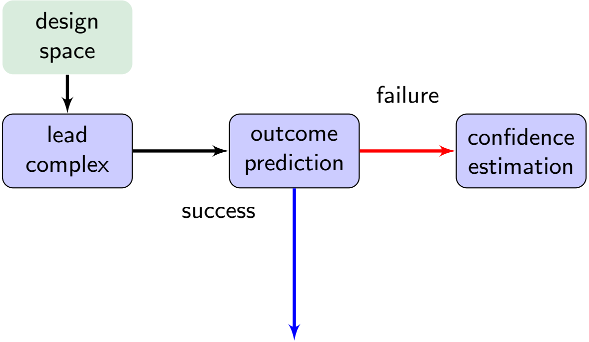
Duan, C., Janet, J.P. *et al.*, *J. Chem. Theory. Comp.*, 15(4):2331—2345, 2019.

Beyond prediction: live job management



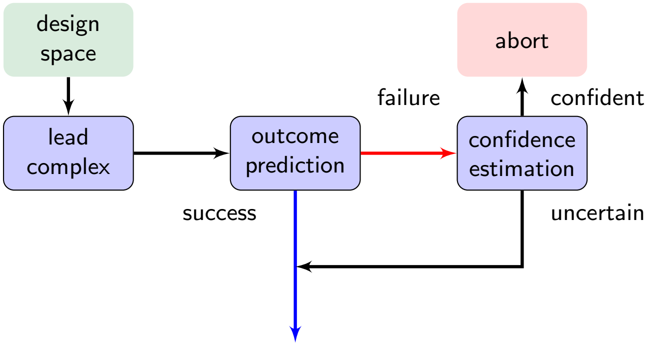
Duan, C., Janet, J.P. *et al.*, *J. Chem. Theory. Comp.*, 15(4):2331—2345, 2019.

Beyond prediction: live job management



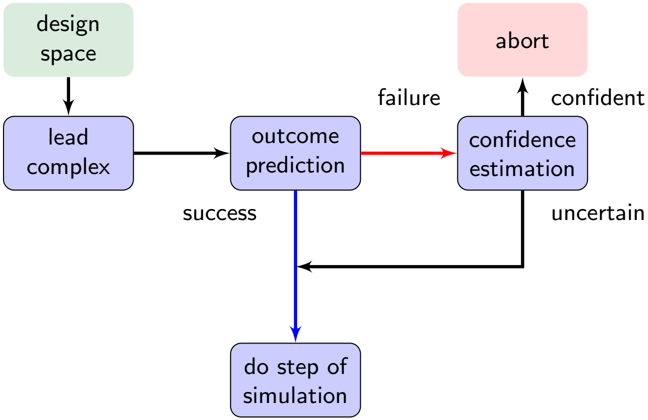
Duan, C., Janet, J.P. *et al.*, *J. Chem. Theory. Comp.*, 15(4):2331—2345, 2019.

Beyond prediction: live job management



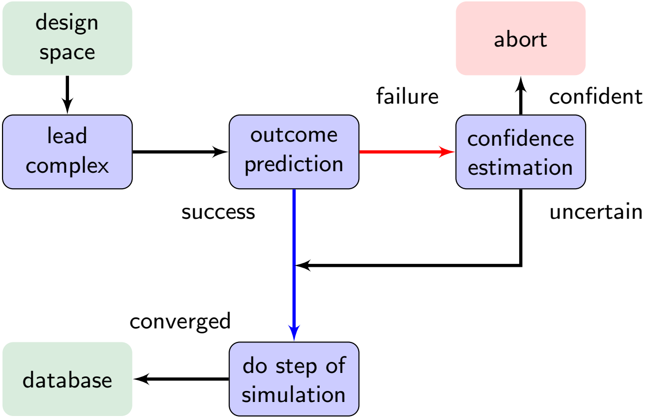
Duan, C., Janet, J.P. *et al.*, *J. Chem. Theory. Comp.*, 15(4):2331—2345, 2019.

Beyond prediction: live job management



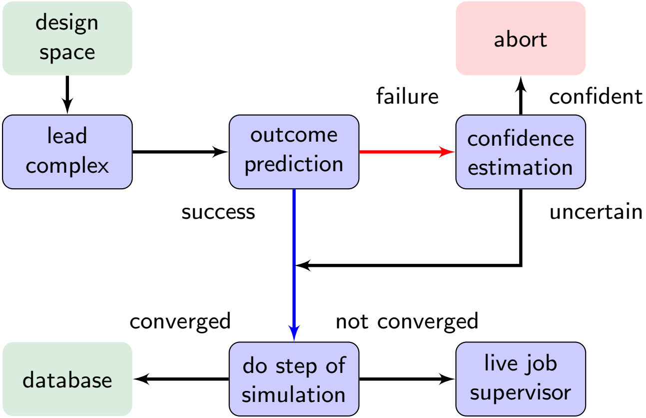
Duan, C., Janet, J.P. *et al.*, *J. Chem. Theory. Comp.*, 15(4):2331—2345, 2019.

Beyond prediction: live job management



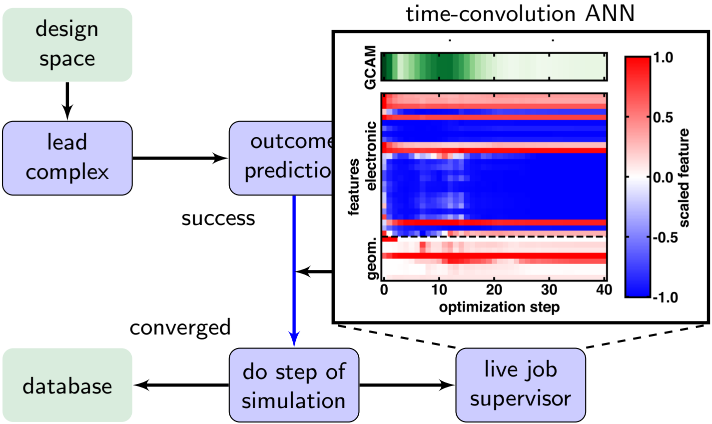
Duan, C., Janet, J.P. *et al.*, *J. Chem. Theory. Comp.*, 15(4):2331—2345, 2019.

Beyond prediction: live job management



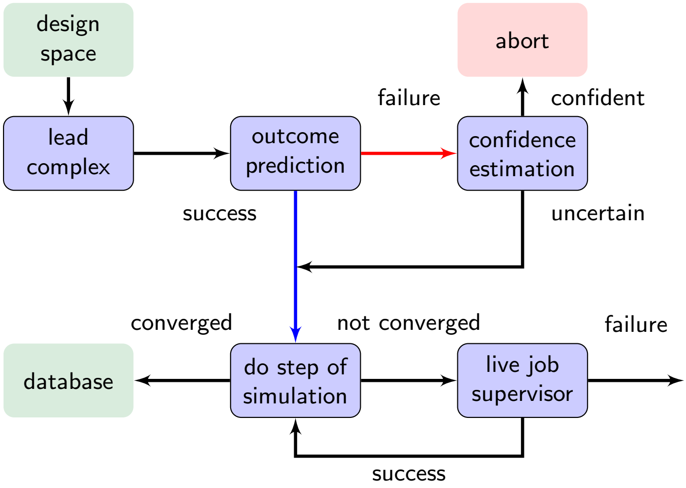
Duan, C., Janet, J.P. *et al.*, *J. Chem. Theory. Comp.*, 15(4):2331—2345, 2019.

Beyond prediction: live job management



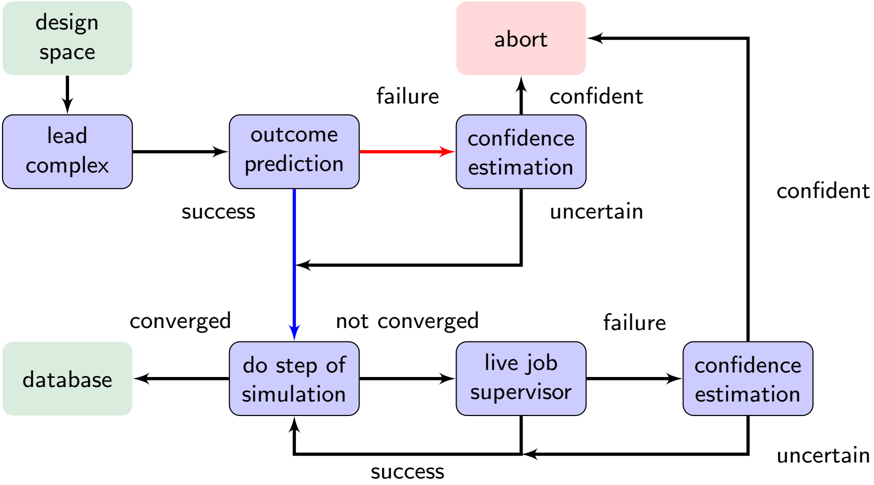
Duan, C., Janet, J.P. *et al.*, *J. Chem. Theory. Comp.*, 15(4):2331—2345, 2019.

Beyond prediction: live job management



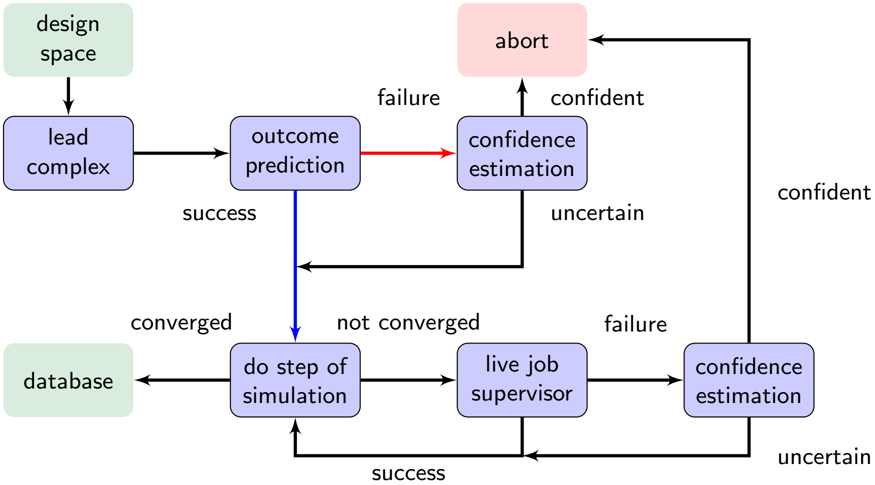
Duan, C., Janet, J.P. *et al.*, *J. Chem. Theory. Comp.*, 15(4):2331—2345, 2019.

Beyond prediction: live job management



Duan, C., Janet, J.P. *et al.*, *J. Chem. Theory. Comp.*, 15(4):2331—2345, 2019.

Beyond prediction: live job management



This leads to about **40% time savings** and can abort almost all failures.

Duan, C., Janet, J.P. *et al.*, *J. Chem. Theory. Comp.*, 15(4):2331—2345, 2019.

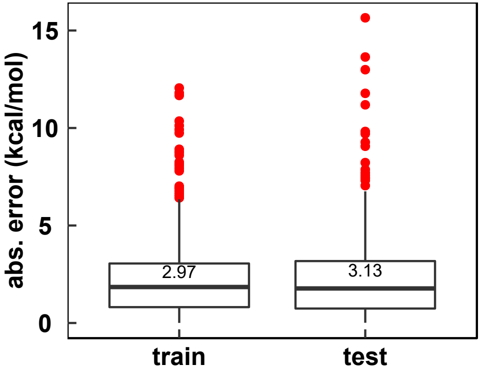
Model transferability

Test-set performance is not necessarily a good metric for general transferability[1]:

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

Model transferability

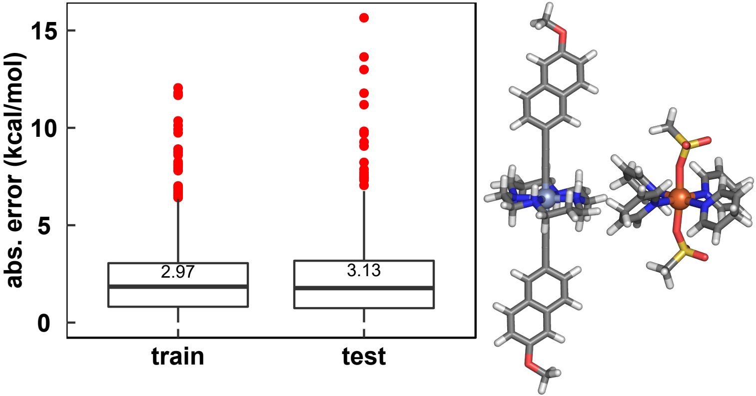
Test-set performance is not necessarily a good metric for general transferability[1]:



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

Model transferability

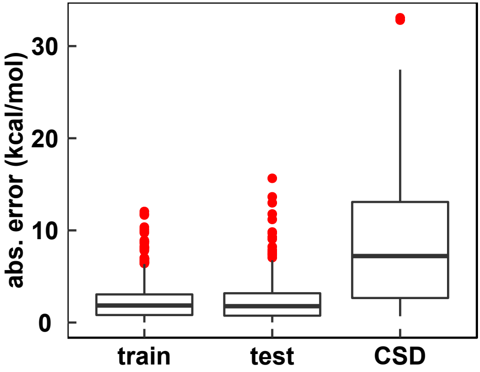
Test-set performance is not necessarily a good metric for general transferability[1]:

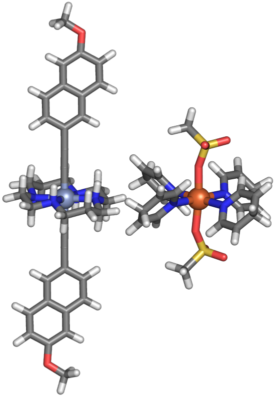


[1]:Janet, J.P., and Kulik, H.J., *Chem. Sci.*, 8:5137–5152, 2017.

Model transferability

Test-set performance is not necessarily a good metric for general transferability[1]:





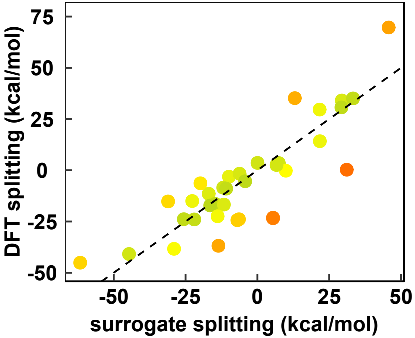
[1]:Janet, J.P., and Kulik, H.J., *Chem. Sci.*, 8:5137–5152, 2017.

System-specific generalization

In practice, model performance is highly variable:

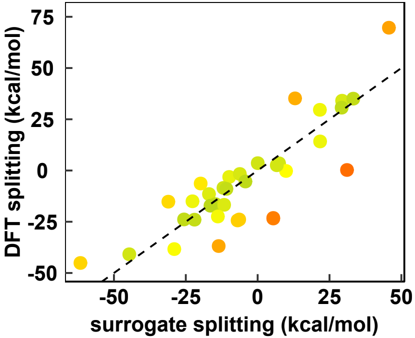
System-specific generalization

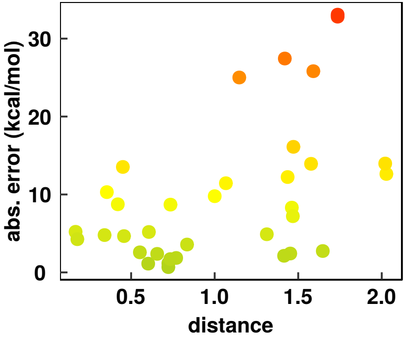
In practice, model performance is highly variable:



System-specific generalization

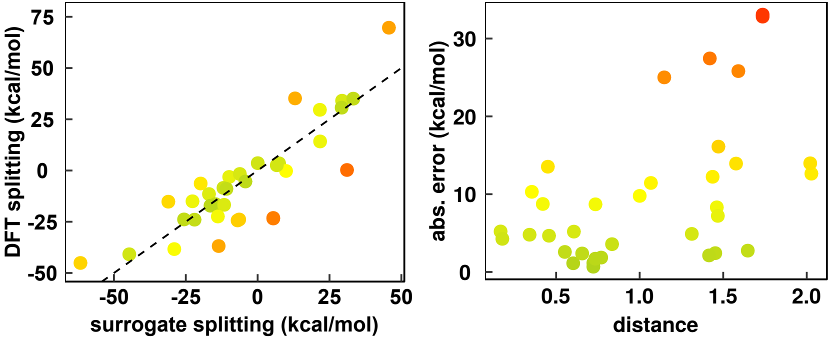
In practice, model performance is highly variable:





System-specific generalization

In practice, model performance is highly variable:



using simple distance worked pretty well!

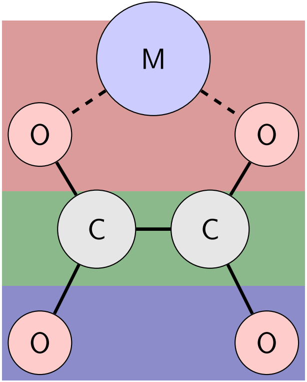
More complex representations

Results are worse for more complex representations[1]:

[1] Janet, J.P., and Kulik, H.J., *J. Phys. Chem. A* 121(46):8939–8954, 2017.

More complex representations

Results are worse for more complex representations[1]:

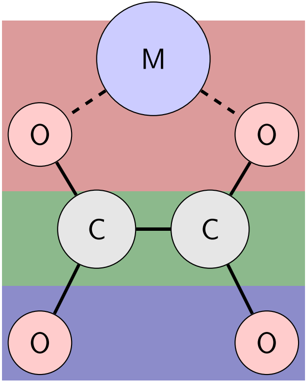


[1] Janet, J.P., and Kulik, H.J., *J. Phys. Chem. A* 121(46):8939–8954, 2017.

More complex representations

Results are worse for more complex representations[1]:

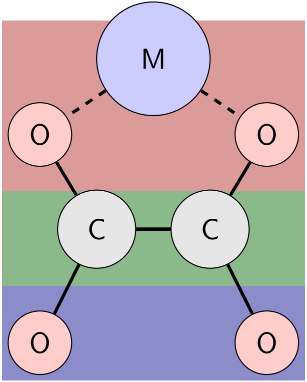
$\boxed{\sim 160 \text{ features in total}}$



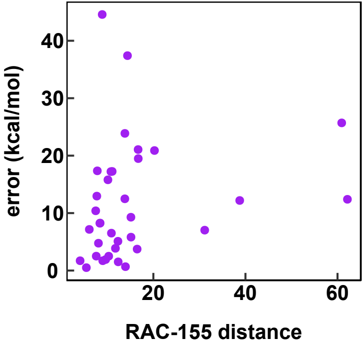
[1] Janet, J.P., and Kulik, H.J., *J. Phys. Chem. A* 121(46):8939–8954, 2017.

More complex representations

Results are worse for more complex representations[1]:



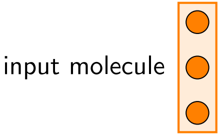
~ 160 features in total



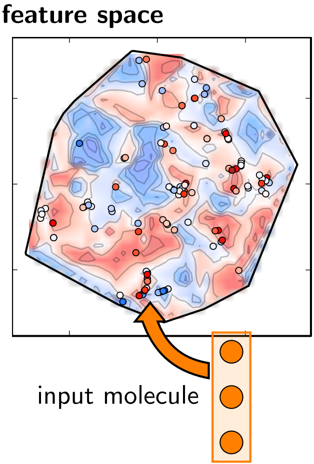
[1] Janet, J.P., and Kulik, H.J., *J. Phys. Chem. A* 121(46):8939–8954, 2017.

How ANNs work

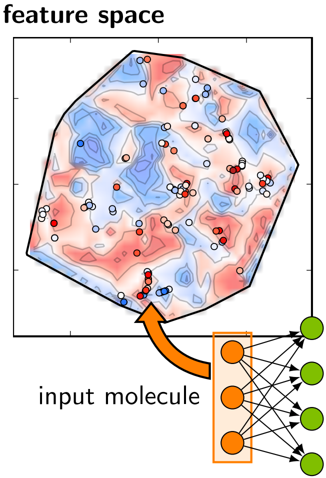
How ANNs work



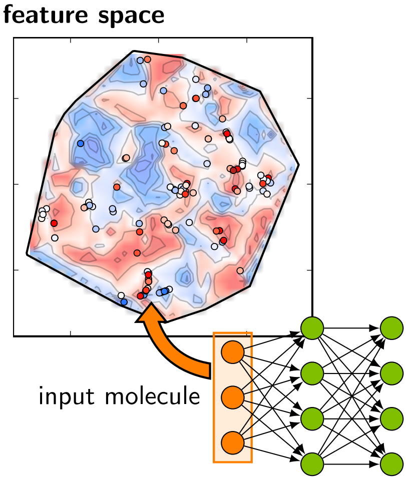
How ANNs work



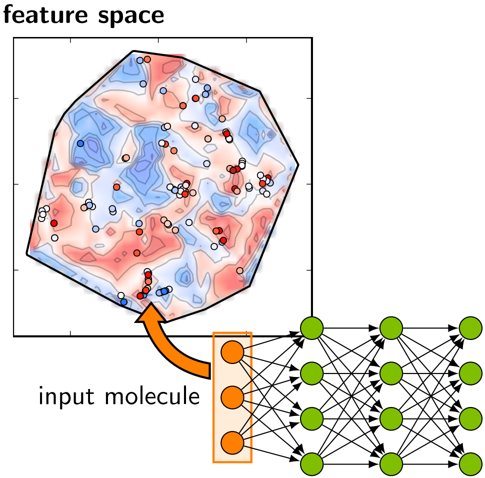
How ANNs work



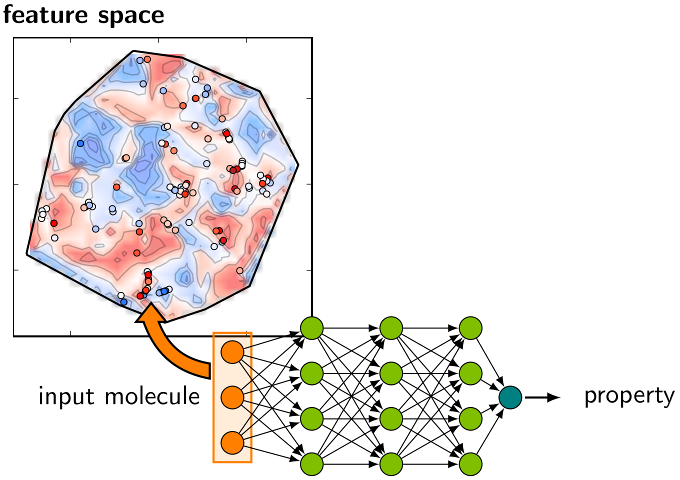
How ANNs work



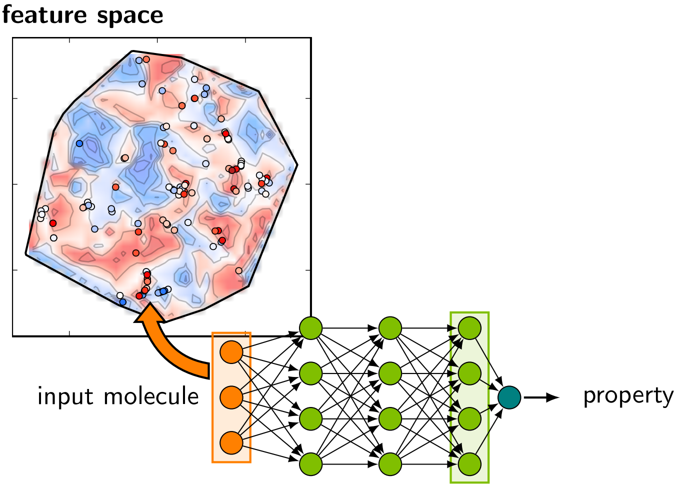
How ANNs work



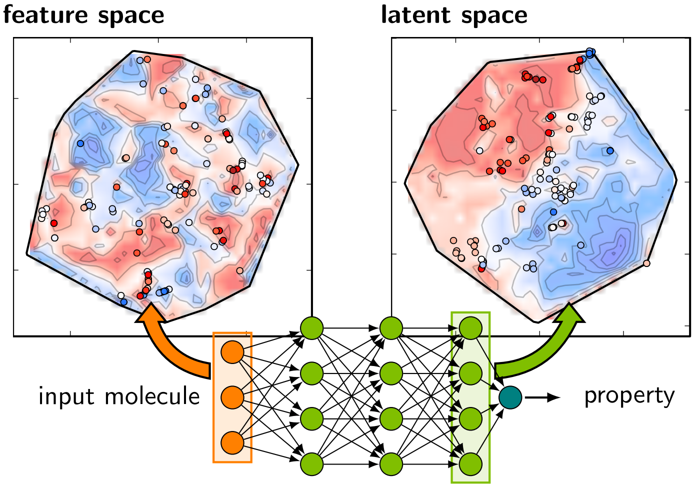
How ANNs work



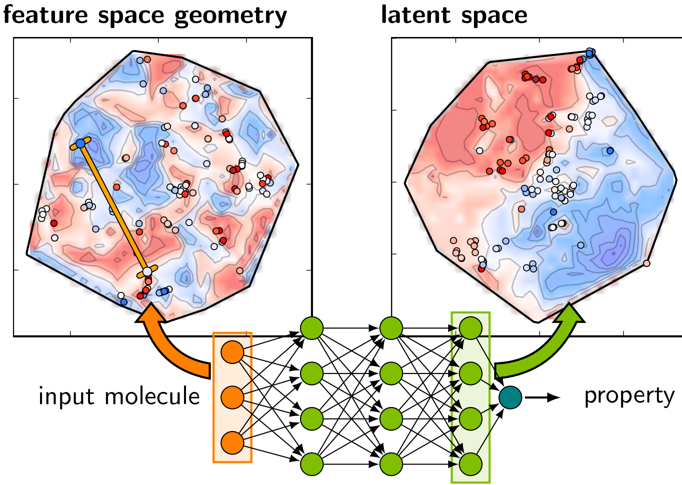
How ANNs work



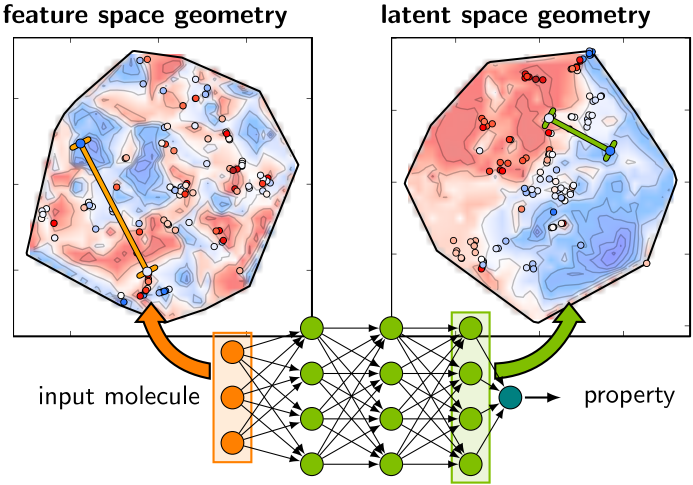
How ANNs work



How ANNs work



How ANNs work

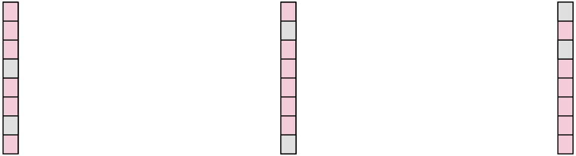


Other UQ metrics

1) Data-sampling ensembles:

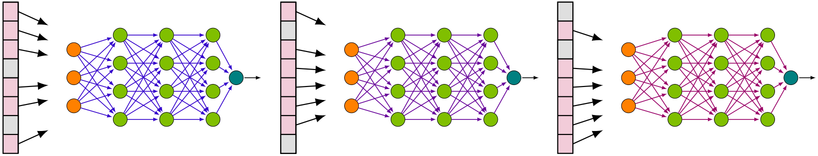
Other UQ metrics

1) Data-sampling ensembles:



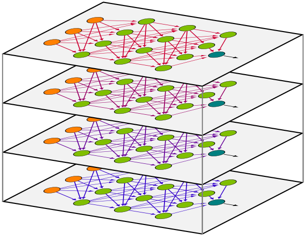
Other UQ metrics

1) Data-sampling ensembles:



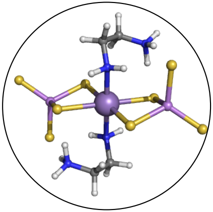
Other UQ metrics

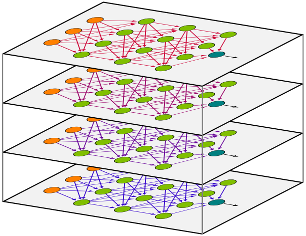
1) Data-sampling ensembles:



Other UQ metrics

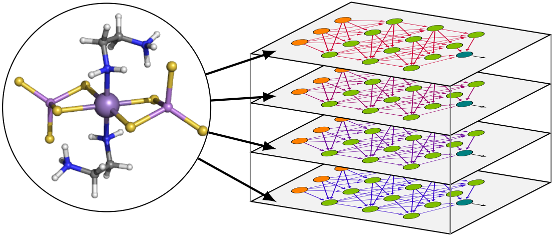
1) Data-sampling ensembles:





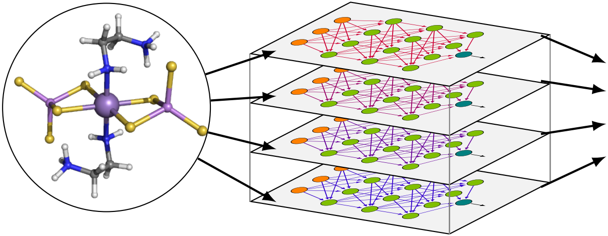
Other UQ metrics

1) Data-sampling ensembles:



Other UQ metrics

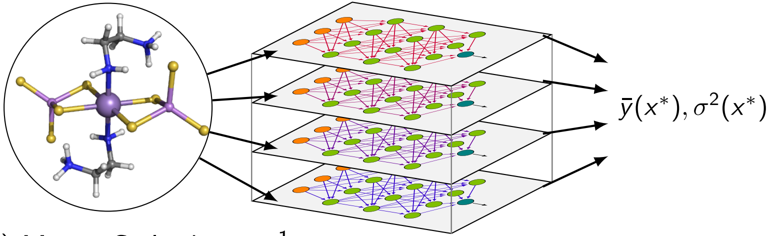
1) Data-sampling ensembles:



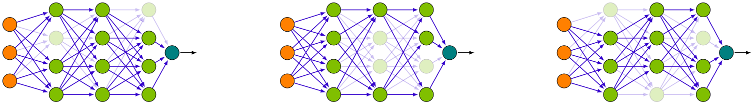
$\bar{y}(x^*), \sigma^2(x^*)$

Other UQ metrics

1) Data-sampling ensembles:



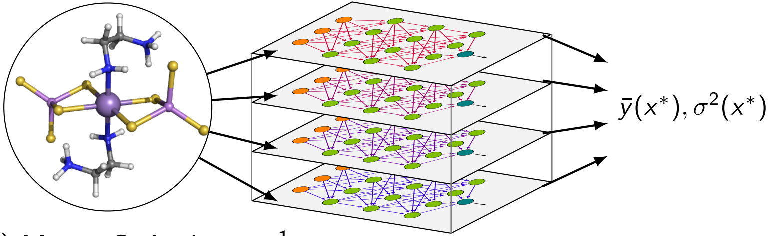
2) Monte Carlo dropout[1]:



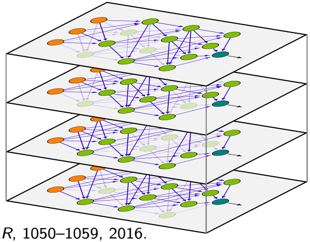
[1]:Gal, Y. and Ghahramani, Z., *ICMLR*, 1050–1059, 2016.

Other UQ metrics

1) Data-sampling ensembles:



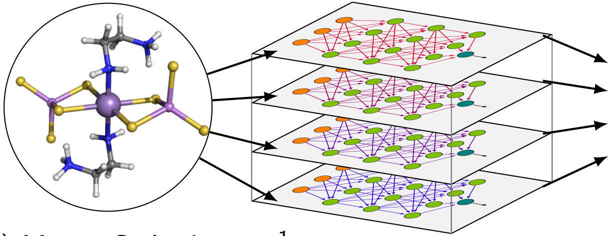
2) Monte Carlo dropout[1]:



[1]:Gal, Y. and Ghahramani, Z., *ICMLR*, 1050–1059, 2016.

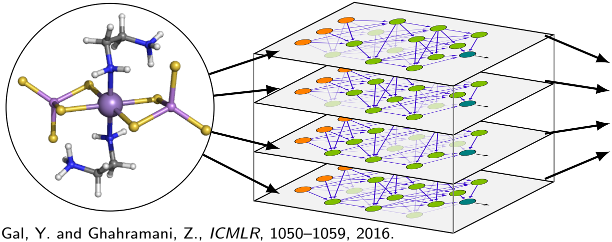
Other UQ metrics

1) Data-sampling ensembles:



$\bar{y}(x^*), \sigma^2(x^*)$

2) Monte Carlo dropout[1]:



$\bar{y}(x^*), \sigma^2(x^*) + \tau$

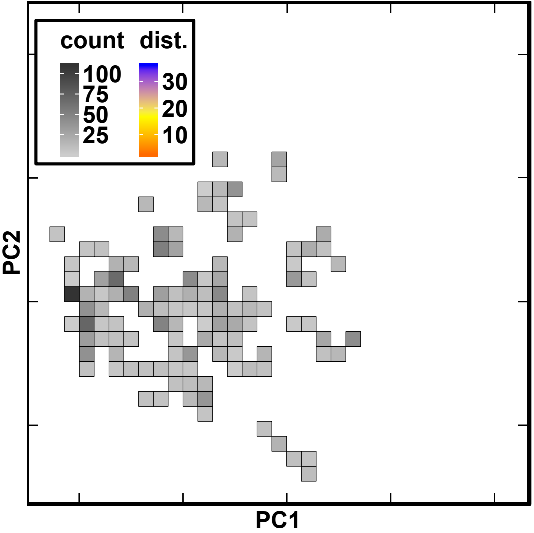
[1]:Gal, Y. and Ghahramani, Z., *ICMLR*, 1050–1059, 2016.

A challenging test case: CSD II

'Out-of-distribution' test: spin-splitting energies of 116 structures from the CSD, from training-like to very different.

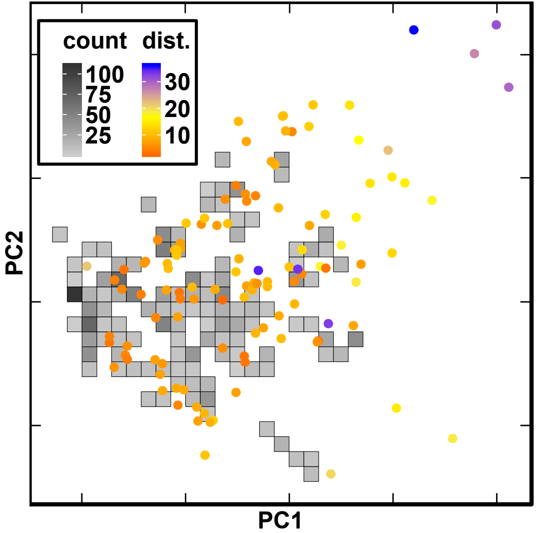
A challenging test case: CSD II

'Out-of-distribution' test: spin-splitting energies of 116 structures from the CSD, from training-like to very different.



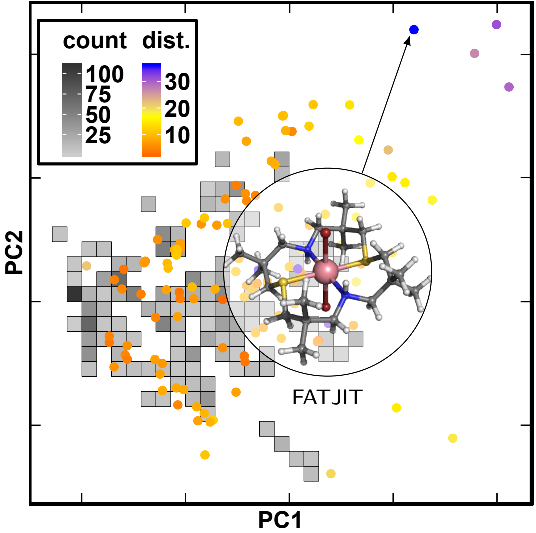
A challenging test case: CSD II

'Out-of-distribution' test: spin-splitting energies of 116 structures from the CSD, from training-like to very different.



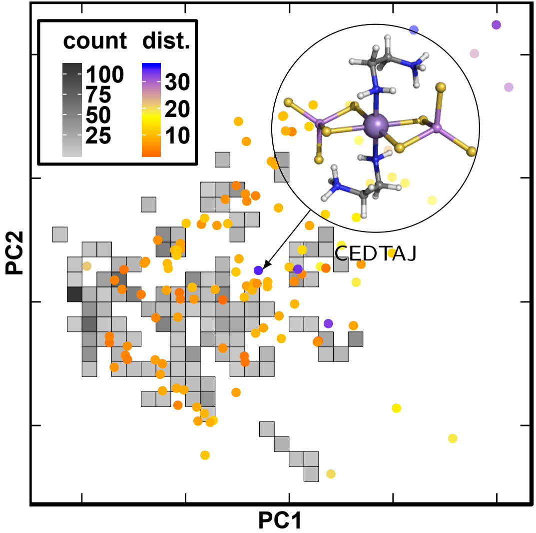
A challenging test case: CSD II

'Out-of-distribution' test: spin-splitting energies of 116 structures from the CSD, from training-like to very different.



A challenging test case: CSD II

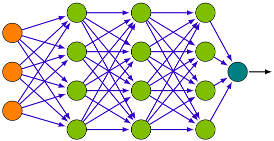
'Out-of-distribution' test: spin-splitting energies of 116 structures from the CSD, from training-like to very different.

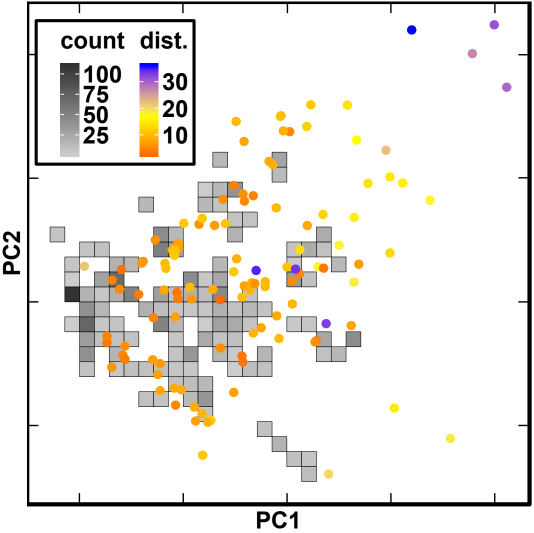


A challenging test case: CSD II

'Out-of-distribution' test: spin-splitting energies of 116 structures from the CSD, from training-like to very different.

Train 3-layer fully connected ANN on 1900 DFT results on simple ligands:

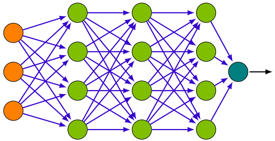


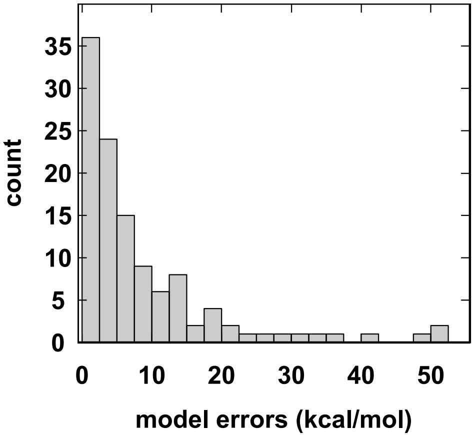


A challenging test case: CSD II

'Out-of-distribution' test: spin-splitting energies of 116 structures from the CSD, from training-like to very different.

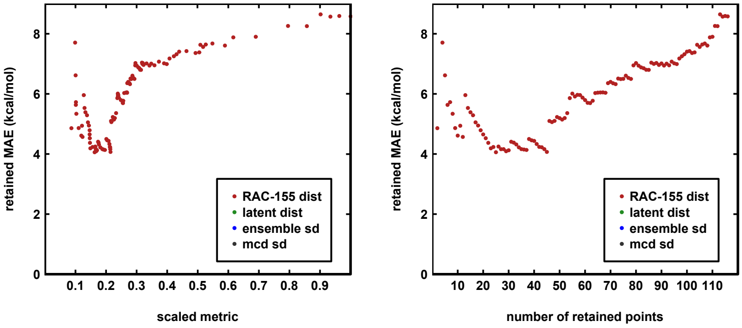
Train 3-layer fully connected ANN on 1900 DFT results on simple ligands:





Latent distances give stable error control

Make a comparison of discriminative power[1]:

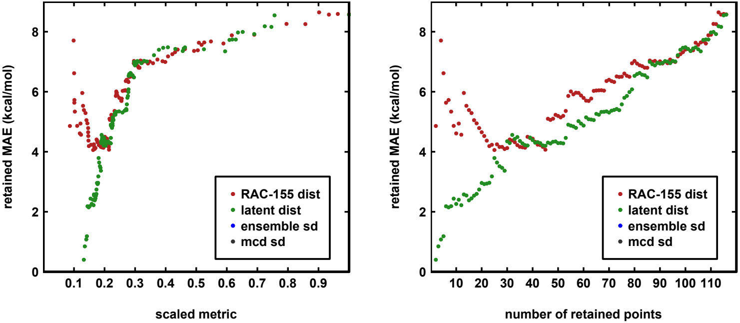


[1] Janet, J.P., *et al.*, ChemRxiv, 10.26434/chemrxiv.7900277.v1.

Latent distances give stable error control

Make a comparison of discriminative power[1]:

latent distances are superior to feature space distances

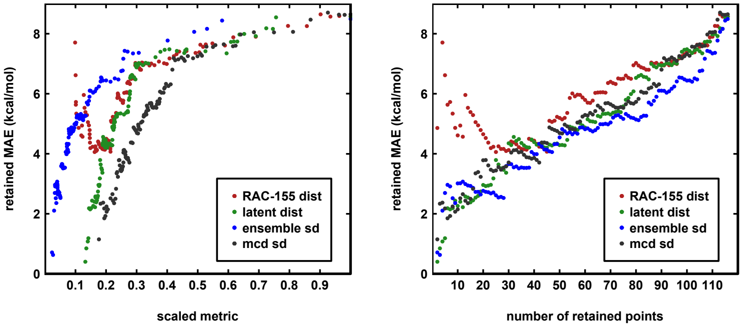


[1] Janet, J.P., *et al.*, ChemRxiv, 10.26434/chemrxiv.7900277.v1.

Latent distances give stable error control

Make a comparison of discriminative power[1]:

latent distances are superior to feature space distances



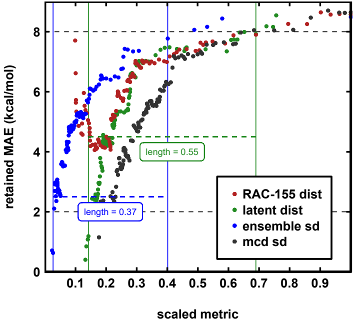
comparable with ensembles and mc dropout

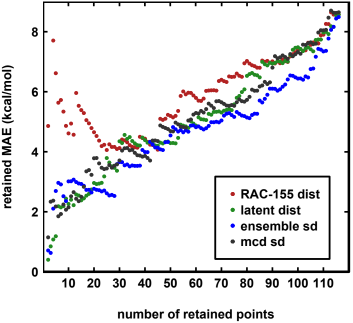
[1] Janet, J.P., *et al.*, ChemRxiv, 10.26434/chemrxiv.7900277.v1.

Latent distances give stable error control

Make a comparison of discriminative power[1]:

latent distances are superior to feature space distances





comparable with ensembles and mc dropout

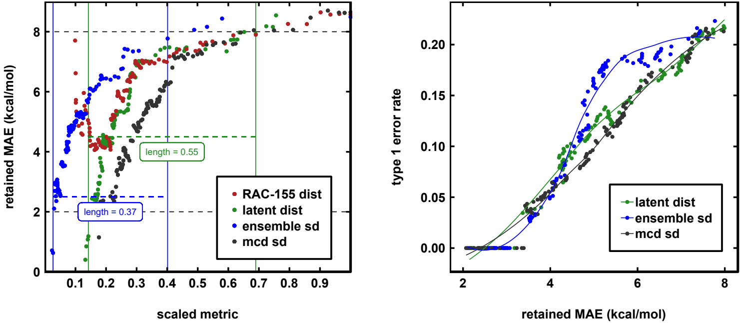
stability is important

[1] Janet, J.P., *et al.*, ChemRxiv, 10.26434/chemrxiv.7900277.v1.

Latent distances give stable error control

Make a comparison of discriminative power[1]:

latent distances are superior to feature space distances

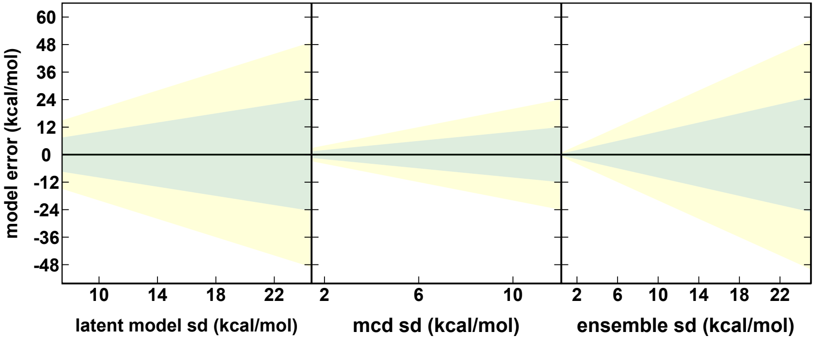


comparable with ensembles and mc dropout

[1] Janet, J.P., *et al.*, ChemRxiv, 10.26434/chemrxiv.7900277.v1.

How do these distributions compare?

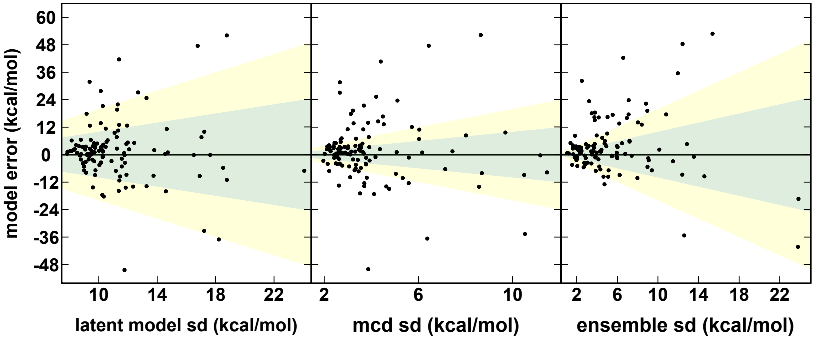
Comparison in energy units[1]: $\varepsilon(d) \sim \mathcal{N}\left(0, \sigma_1^2 + d\sigma_2^2\right)$



[1] Janet, J.P., *et al.*, ChemRxiv, 10.26434/chemrxiv.7900277.v1.

How do these distributions compare?

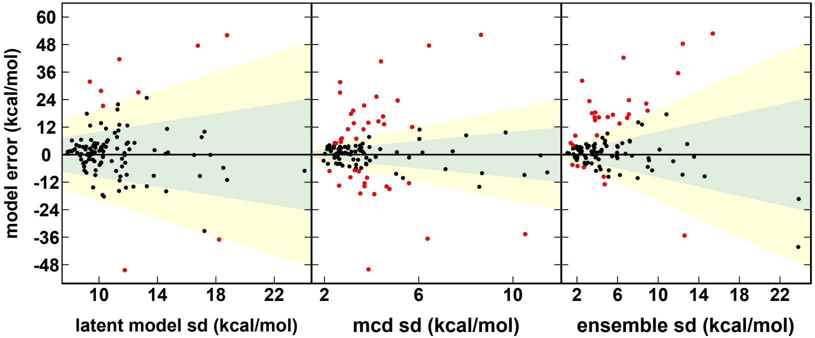
Comparison in energy units[1]: $\varepsilon(d) \sim \mathcal{N}\left(0, \sigma_1^2 + d\sigma_2^2\right)$



[1] Janet, J.P., *et al.*, ChemRxiv, 10.26434/chemrxiv.7900277.v1.

How do these distributions compare?

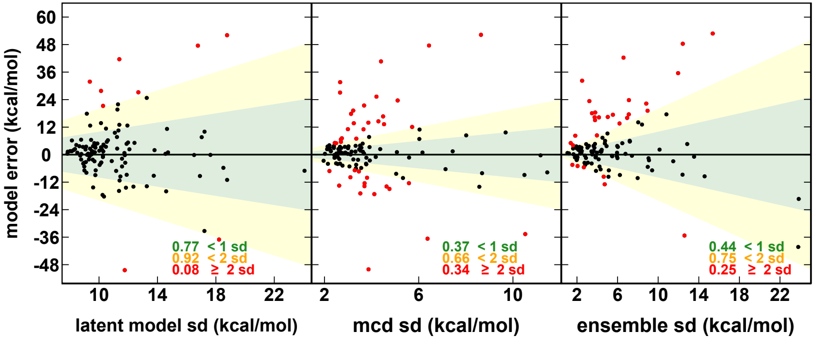
Comparison in energy units[1]: $\varepsilon(d) \sim \mathcal{N}\left(0, \sigma_1^2 + d\sigma_2^2\right)$



[1] Janet, J.P., *et al.*, ChemRxiv, 10.26434/chemrxiv.7900277.v1.

How do these distributions compare?

Comparison in energy units[1]: $\varepsilon(d) \sim \mathcal{N}\left(0, \sigma_1^2 + d\sigma_2^2\right)$



[1] Janet, J.P., *et al.*, ChemRxiv, 10.26434/chemrxiv.7900277.v1.

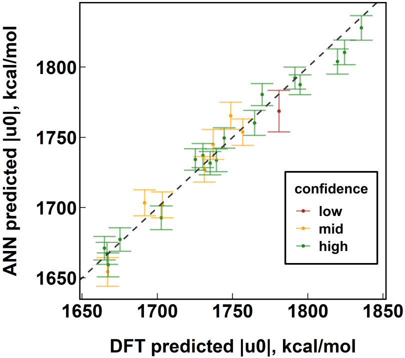
QM9 results

Similar error control can be obtained for QM9 benchmark organic data[1]. We train on 5% and make predictions on 95%.

[1] Ramakrishnan, R., *et al.*, *Sci. Data*, 1, 2014.

QM9 results

Similar error control can be obtained for QM9 benchmark organic data[1]. We train on 5% and make predictions on 95%.



[1] Ramakrishnan, R., *et al.*, *Sci. Data*, 1, 2014.

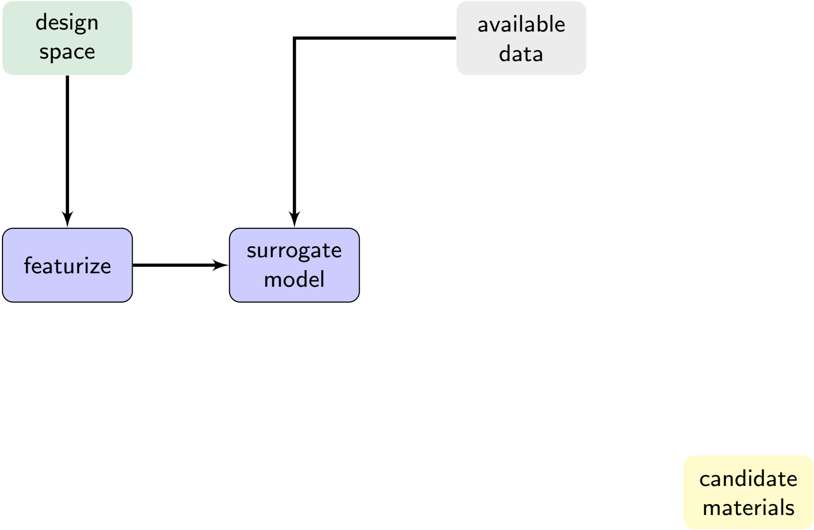
Algorithmic chemical discovery

design space

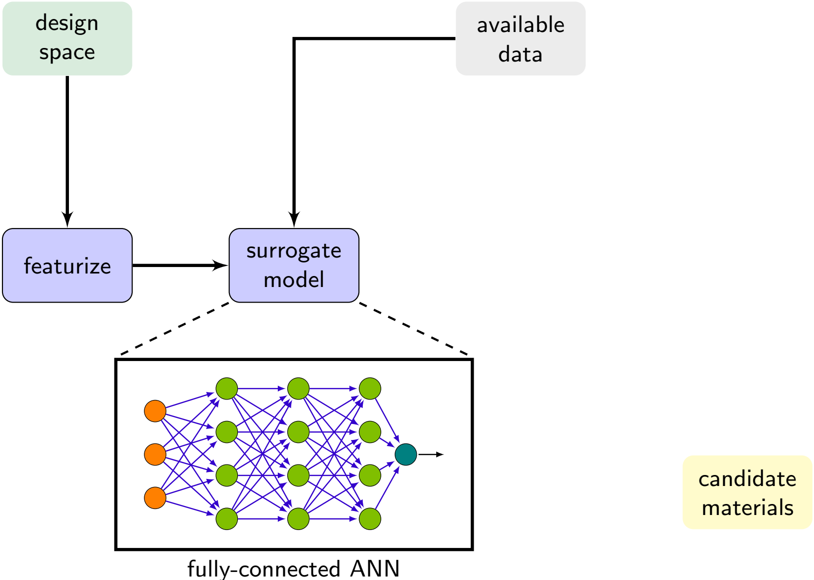
candidate materials

Algorithmic chemical discovery

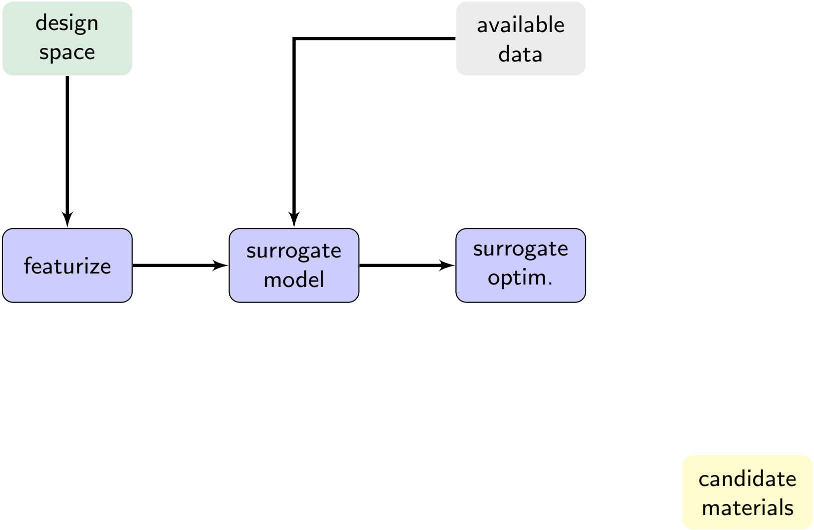
Algorithmic chemical discovery



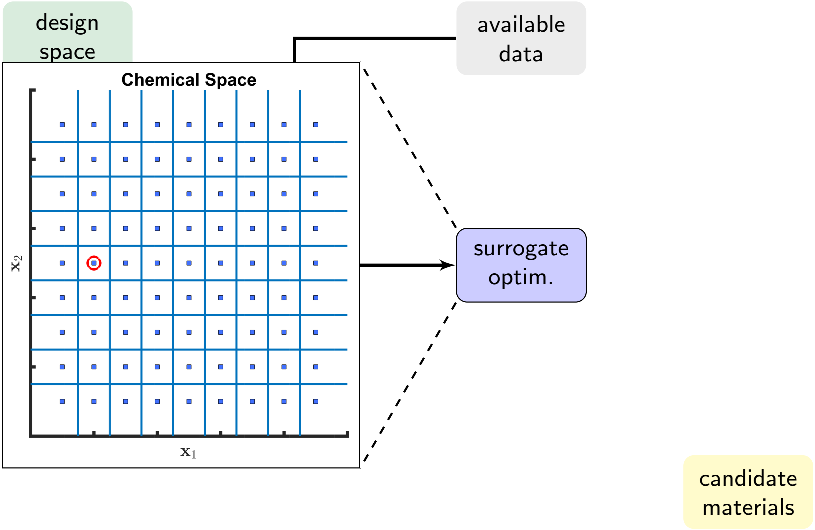
Algorithmic chemical discovery



Algorithmic chemical discovery



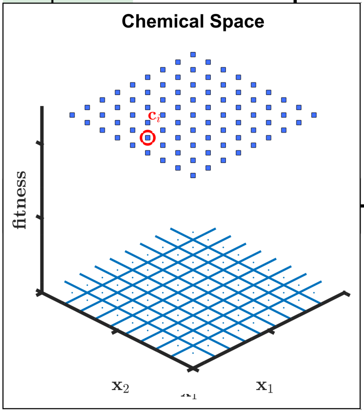
Algorithmic chemical discovery



Algorithmic chemical discovery

design space

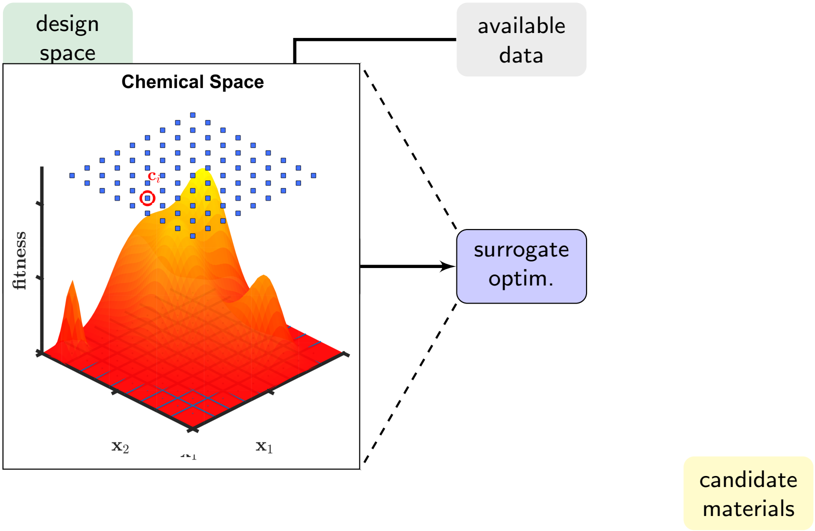
available data



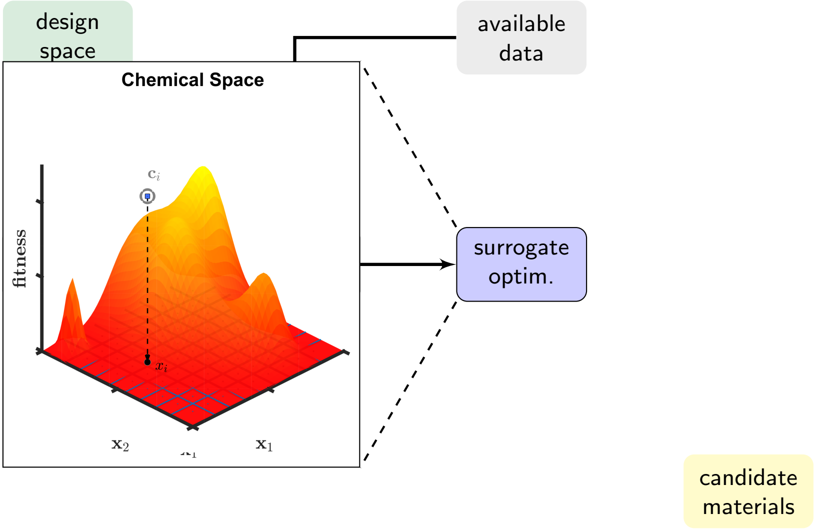
surrogate optim.

candidate materials

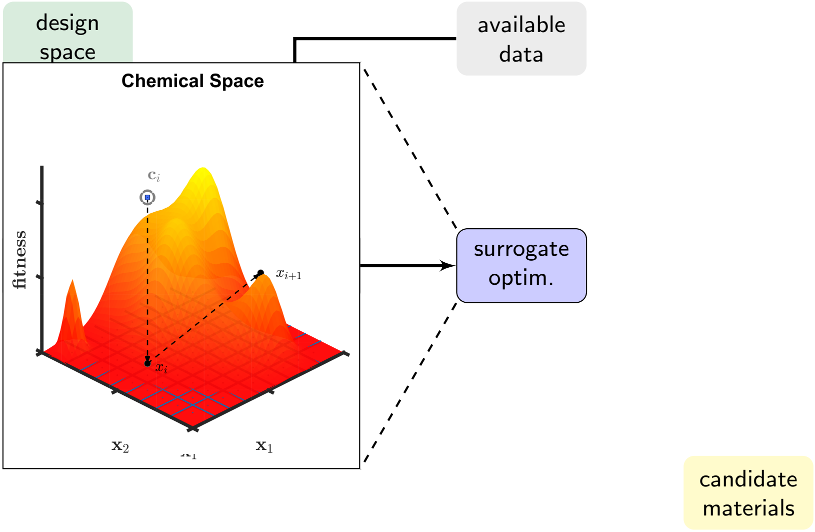
Algorithmic chemical discovery



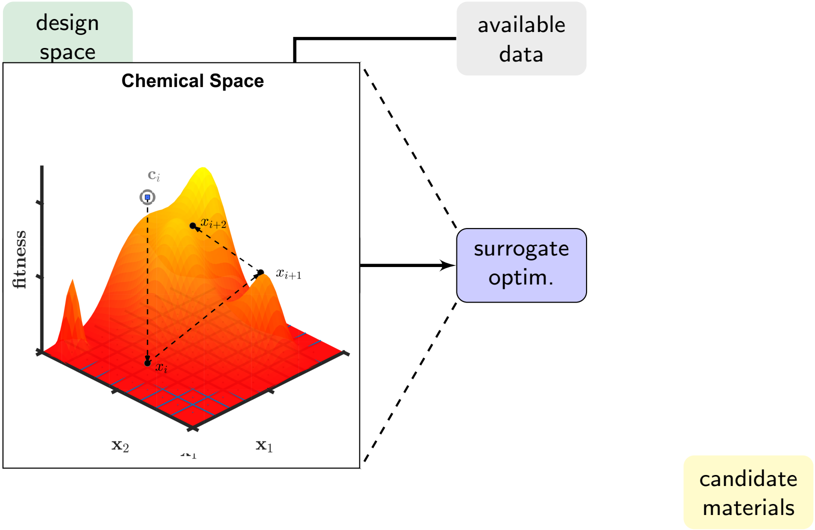
Algorithmic chemical discovery



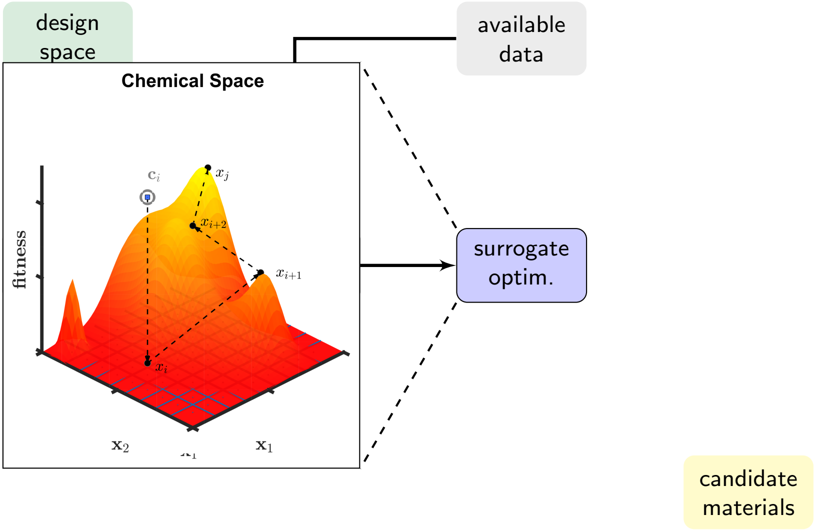
Algorithmic chemical discovery



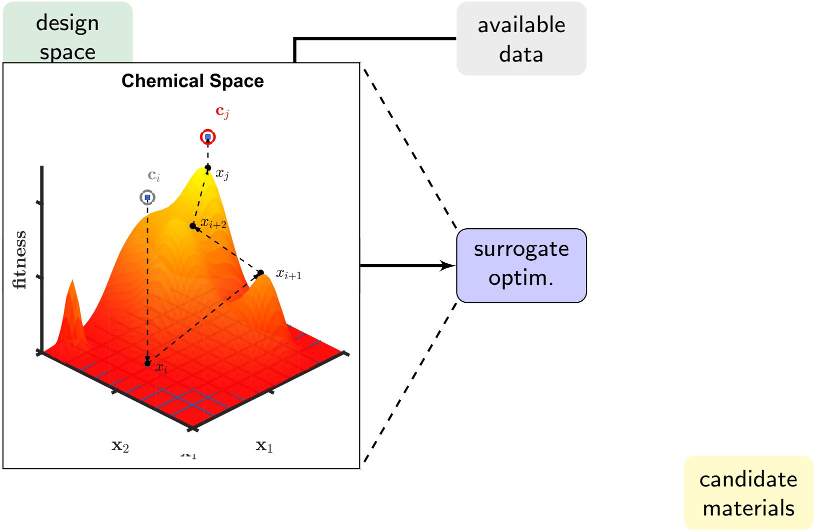
Algorithmic chemical discovery



Algorithmic chemical discovery

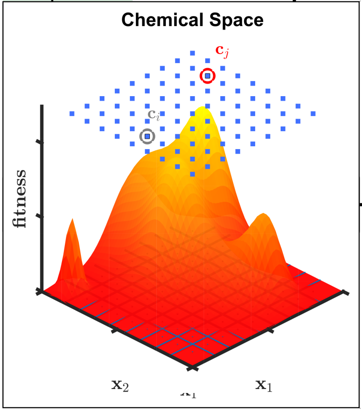


Algorithmic chemical discovery



Algorithmic chemical discovery

design space

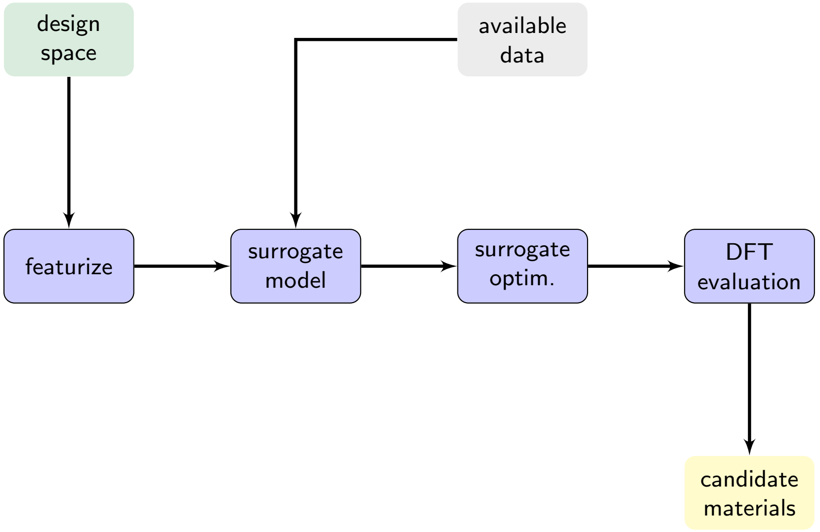


available data

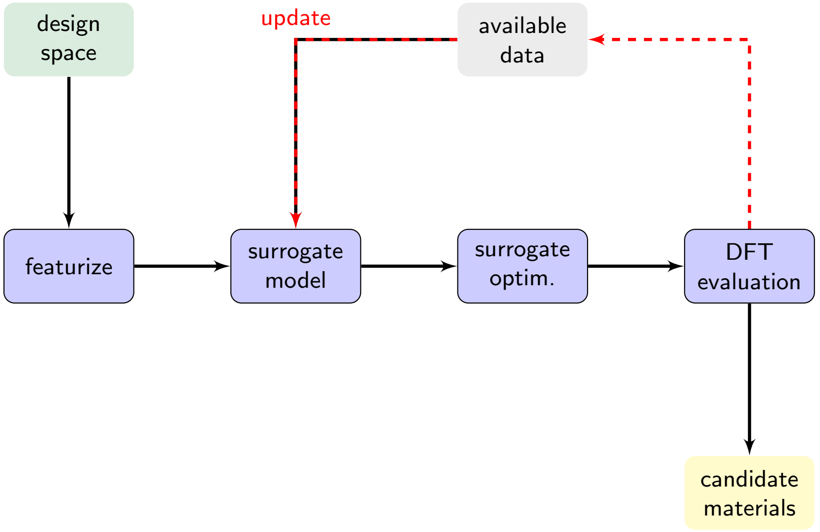
surrogate optim.

candidate materials

Algorithmic chemical discovery



Algorithmic chemical discovery



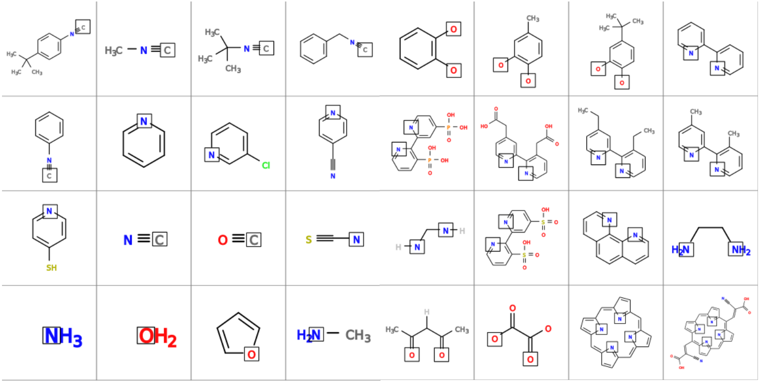
Can we use these models for discovery?

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.*, 9(5):1064–1071, 2018.

Can we use these models for discovery?

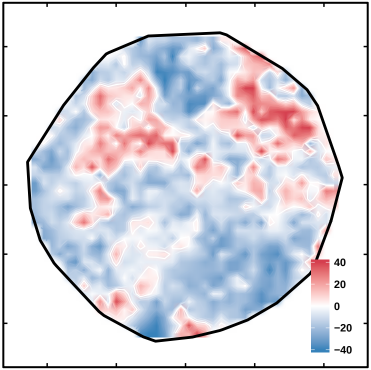
Can we use the ANN model to find new spin-crossover materials, i.e. $\Delta E_{H-L} = 0$? Define an expanded space with $< 2\%$ training coverage[3]

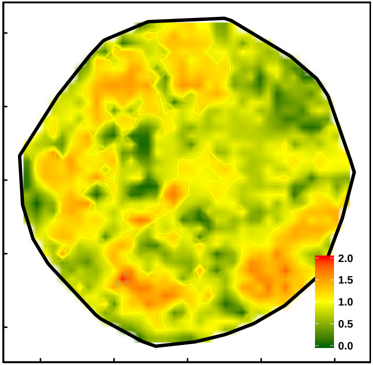


[3] Janet, J.P., Chan, L. and Kulik, H.J., *J. Phys. Chem. Lett.*, 9(5):1064–1071, 2018.

Can we use these models for discovery?

Can we use the ANN model to find new spin-crossover materials, i.e. $\Delta E_{H-L} = 0$? Define an expanded space with $< 2\%$ training coverage[3]



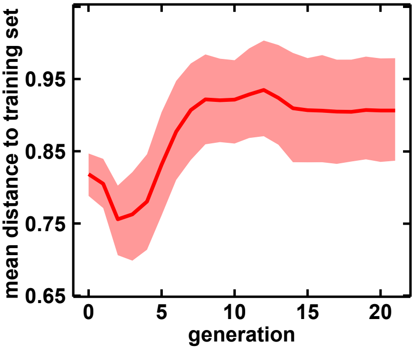


[3] Janet, J.P., Chan, L. and Kulik, H.J., *J. Phys. Chem. Lett.*, 9(5):1064–1071, 2018.

UQ and evolutionary design

We developed an evolutionary algorithm that combines uncertainty estimation with property prediction:

At high distances, surrogate is unreliable. At low distance, data is weakly informative.



Janet, J.P., Chan, L. and Kulik, H.J., *J. Phys. Chem. Lett.*, 9(5):1064–1071, 2018.

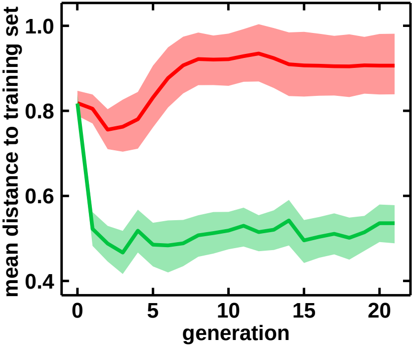
UQ and evolutionary design

We developed an evolutionary algorithm that combines uncertainty estimation with property prediction:

At high distances, surrogate is unreliable. At low distance, data is weakly informative.

We can adaptively penalize high distance regions:

$$F_{s+d}(x) \quad := \exp\left[-\left(\frac{\Delta E_{\text{H-L}}(x)}{p_{\Delta E_{\text{H-L}}}}\right)^2\right] \exp\left[-\left(\frac{d(x)}{p_d}\right)^2\right]$$



Janet, J.P., Chan, L. and Kulik, H.J., *J. Phys. Chem. Lett.*, 9(5):1064–1071, 2018.

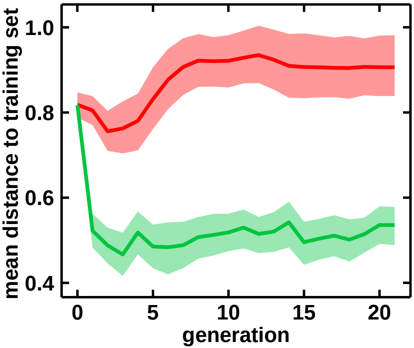
UQ and evolutionary design

We developed an evolutionary algorithm that combines uncertainty estimation with property prediction:

At high distances, surrogate is unreliable. At low distance, data is weakly informative.

We can adaptively penalize high distance regions:

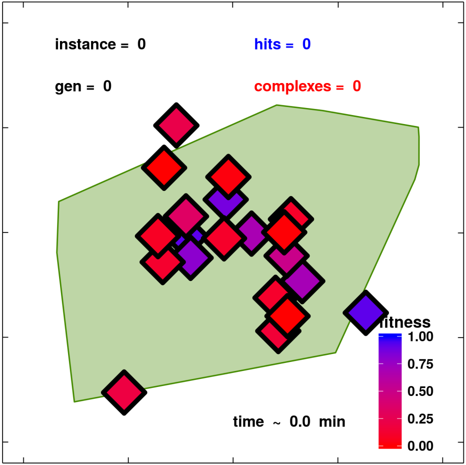
$$F_{s+d}(x) := \exp\left[-\left(\frac{\Delta E_{\text{H-L}}(x)}{p_{\Delta E_{\text{H-L}}}}\right)^2\right] \exp\left[-\left(\frac{d(x)}{p_d}\right)^2\right]$$



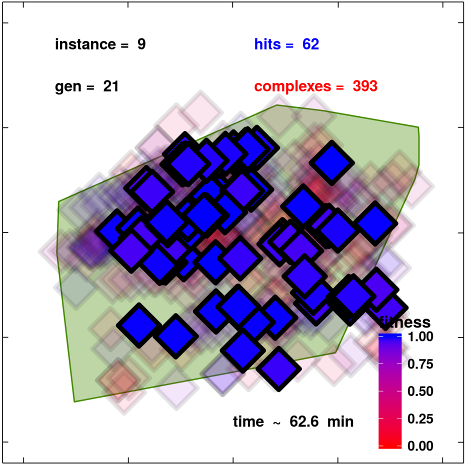
Janet, J.P., Chan, L. and Kulik, H.J., *J. Phys. Chem. Lett.*, 9(5):1064–1071, 2018.

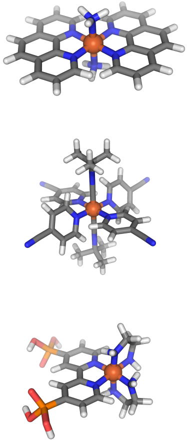
Demonstration

Demonstration



Demonstration

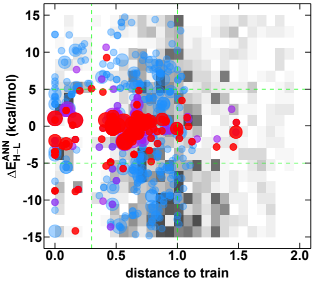




Discovery results

Spin splitting design:

We combine ANN predictions and uncertainties using an evolutionary algorithm.
Error control allows 60% of leads to be validated with DFT.[1]

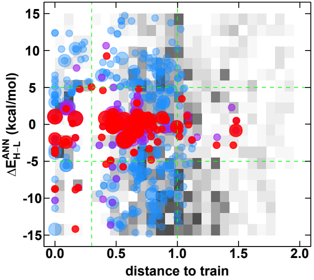


[1] Janet, J.P., Chan, L. and Kulik, H.J., *J. Phys. Chem. Lett.*, 9(5):1064–1071, 2018.

Discovery results

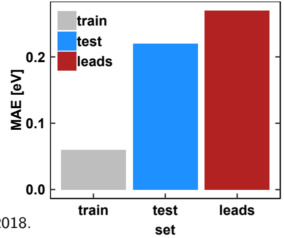
Spin splitting design:

We combine ANN predictions and uncertainties using an evolutionary algorithm.
Error control allows 60% of leads to be validated with DFT.[1]



Frontier orbital properties:

This approach also works for frontier orbtial design[2], obtaining average HOMO of 3.98 eV compared to target 4.00 eV.



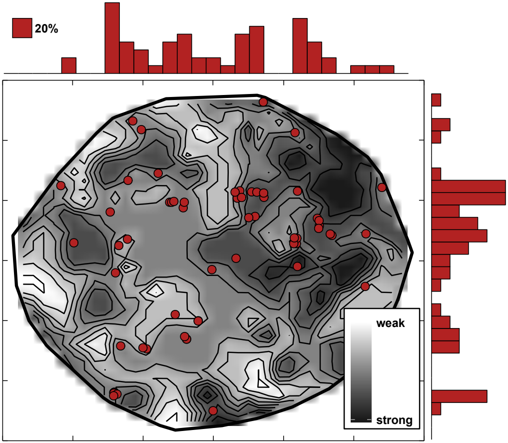
[1] Janet, J.P., Chan, L. and Kulik, H.J., *J. Phys. Chem. Lett.*, 9(5):1064–1071, 2018.
[2] Nandy, A. *et al.*, *Ind. Eng. Chem. Res.*, 57(42):13973–13986, 2018.

Hedging against DFT uncertainty

Because we have trained our models on varying with exact exchange, we can tune functionals for design:

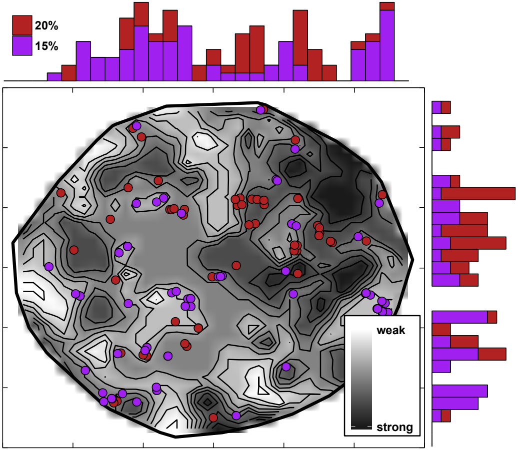
Hedging against DFT uncertainty

Because we have trained our models on varying with exact exchange, we can tune functionals for design:



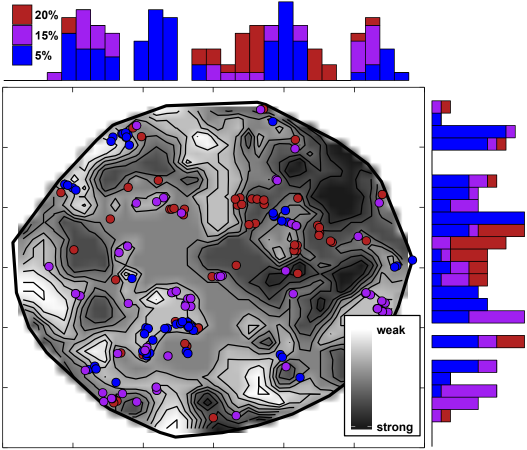
Hedging against DFT uncertainty

Because we have trained our models on varying with exact exchange, we can tune functionals for design:



Hedging against DFT uncertainty

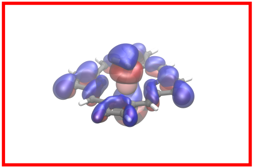
Because we have trained our models on varying with exact exchange, we can tune functionals for design:



Computational chemistry and machine learning

Awkward roommates or match made in heaven?

physics-driven



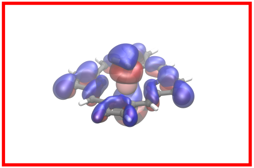
slow, accurate (?)

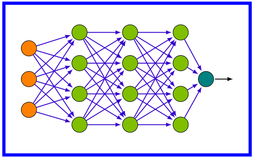
Computational chemistry and machine learning

Awkward roommates or match made in heaven?

physics-driven

data-driven





slow, accurate (?)

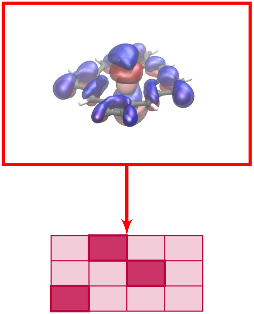
fast, uncertainty-aware

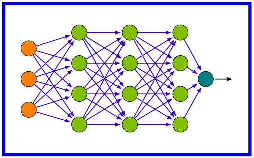
Computational chemistry and machine learning

Awkward roommates or match made in heaven?

physics-driven

data-driven





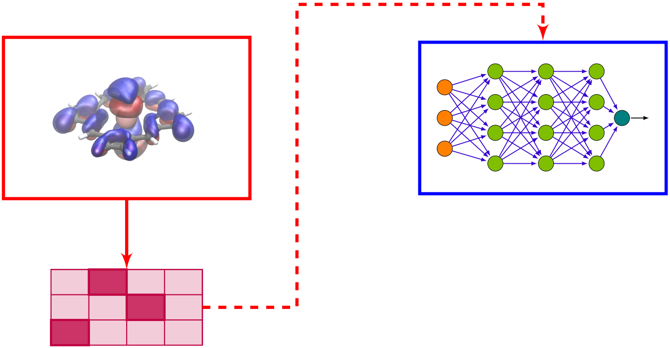
slow, accurate (?)

fast, uncertainty-aware

Computational chemistry and machine learning

Awkward roommates or match made in heaven?

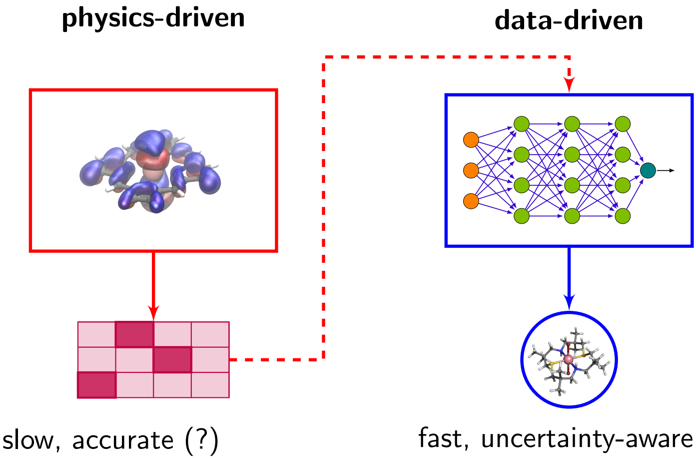
physics-driven **data-driven**



slow, accurate (?) fast, uncertainty-aware

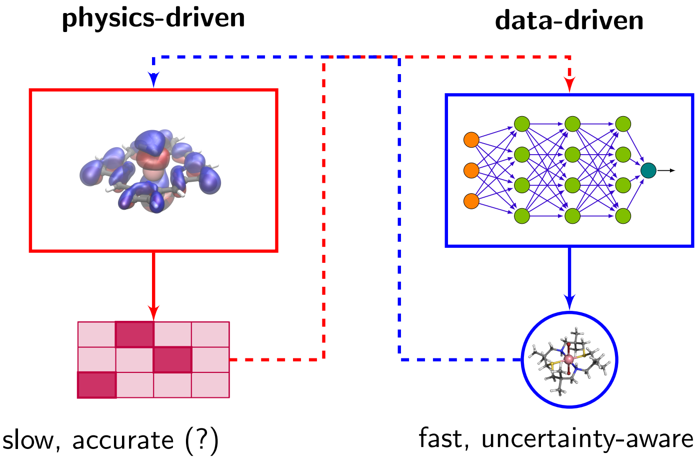
Computational chemistry and machine learning

Awkward roommates or match made in heaven?



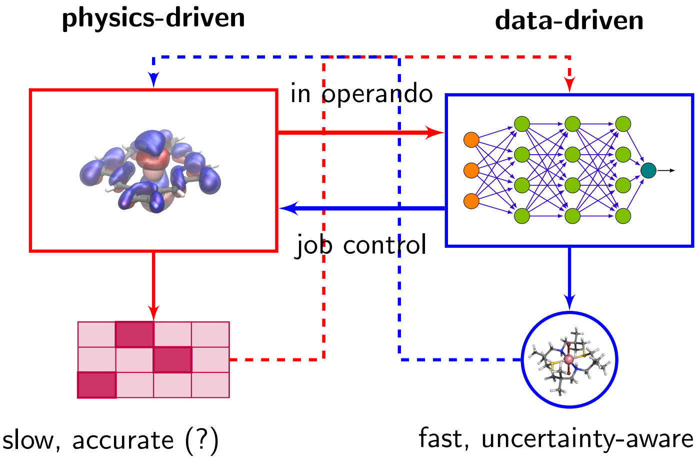
Computational chemistry and machine learning

Awkward roommates or match made in heaven?



Computational chemistry and machine learning

Awkward roommates or match made in heaven?



Acknowledgments

Thanks to the Kulik group and funding partners: