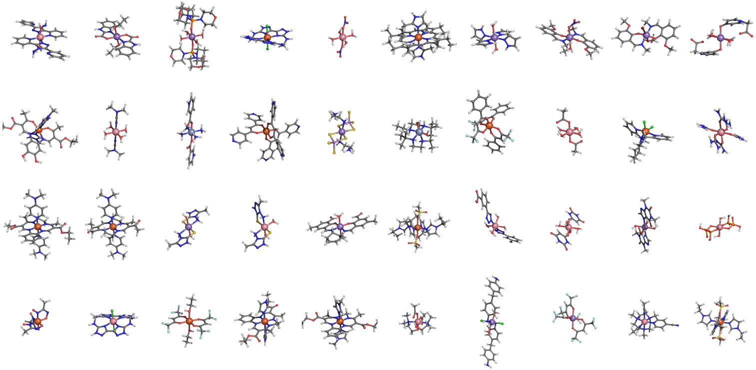
Solving material design problems with simulation and ML

Jon Paul Janet

01.01.0001

Machine learning and simulation for chemical discovery

How can we design new materials using computers?

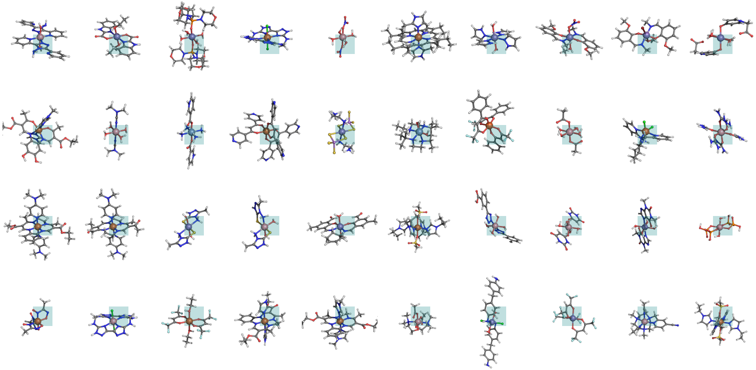


The space of possible chemistry is incredibly vast, with $\mathcal{O}(10^{60})$ small organic molecules alone.

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

Machine learning and simulation for chemical discovery

How can we design new materials using computers?

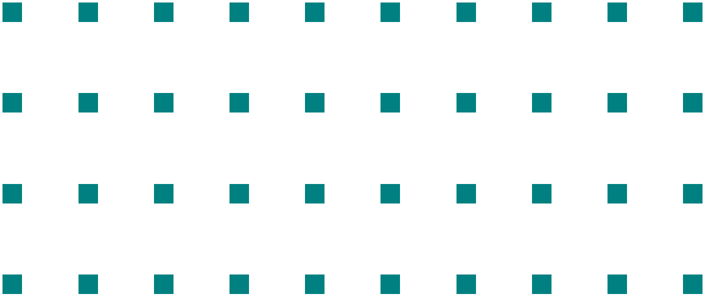


The space of possible chemistry is incredibly vast, with $\mathcal{O}(10^{60})$ small organic molecules alone.

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

Machine learning and simulation for chemical discovery

How can we design new materials using computers?

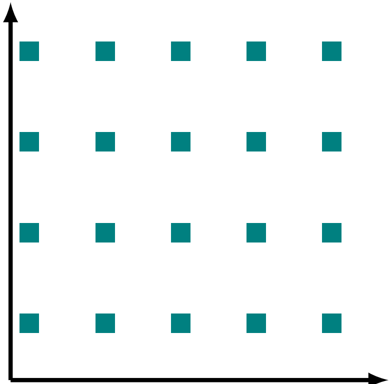


The space of possible chemistry is incredibly vast, with $\mathcal{O}(10^{60})$ small organic molecules alone.

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

Machine learning and simulation for chemical discovery

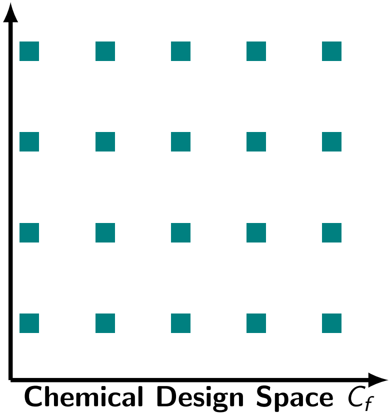
How can we design new materials using computers?



Chemical Design Space C_f

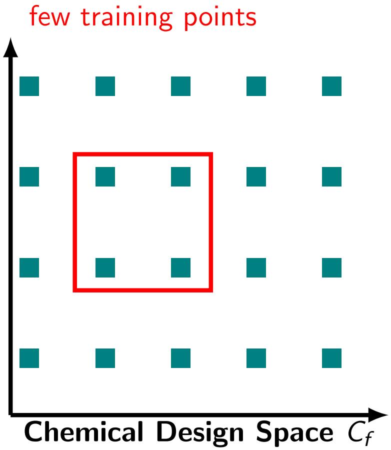
Machine learning and simulation for chemical discovery

Inorganic materials design

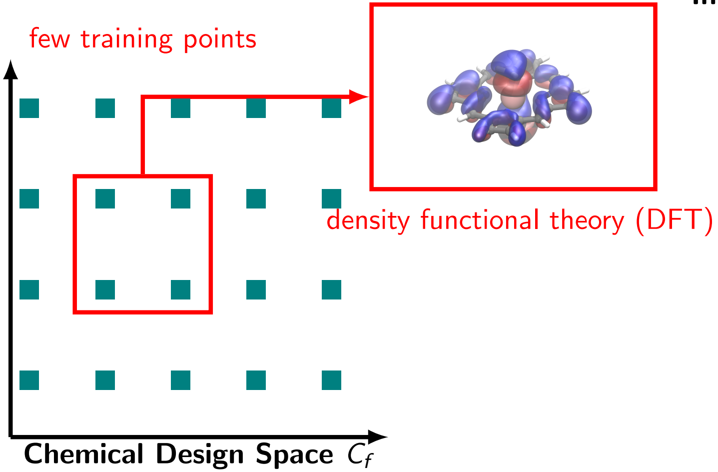


Machine learning and simulation for chemical discovery

Inorganic materials design



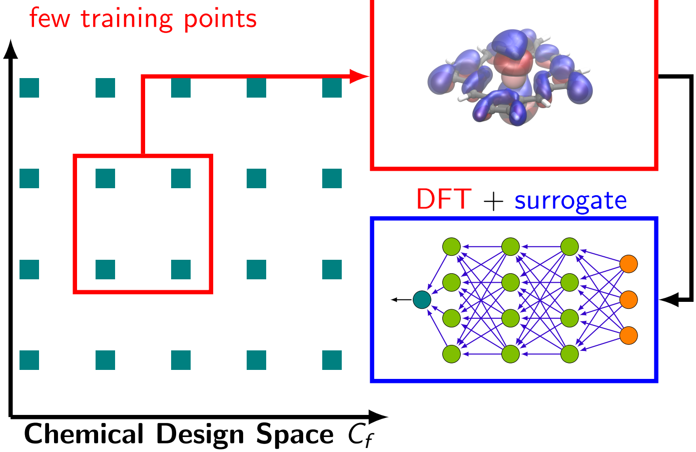
Machine learning and simulation for chemical discovery



Inorganic materials design

- surrogate modeling & integration with simulation

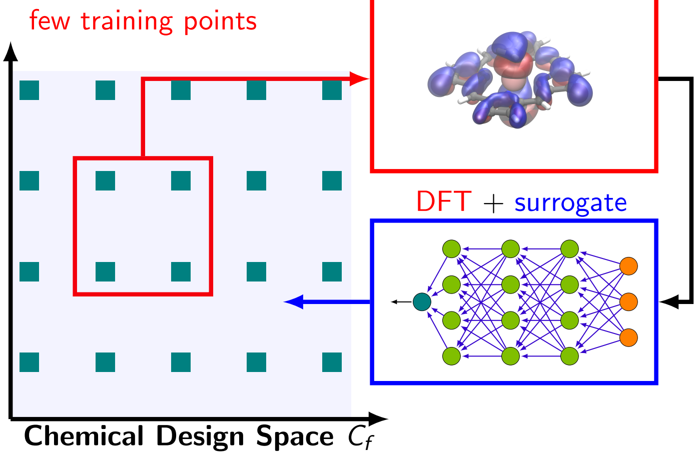
Machine learning and simulation for chemical discovery



Inorganic materials design

- surrogate modeling & integration with simulation

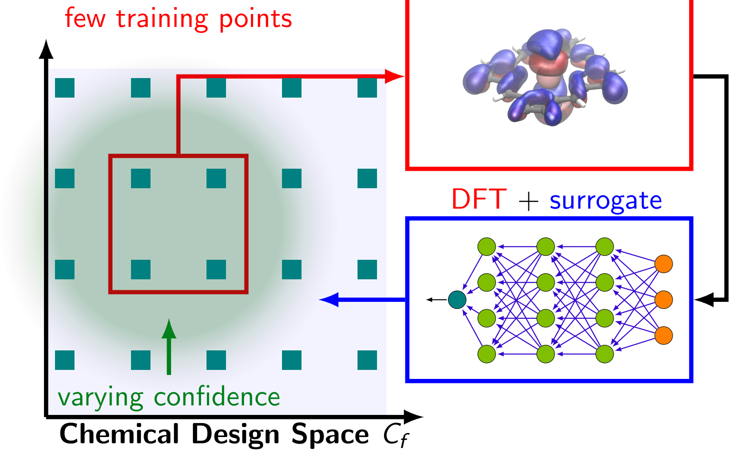
Machine learning and simulation for chemical discovery



Inorganic materials design

- surrogate modeling & integration with simulation

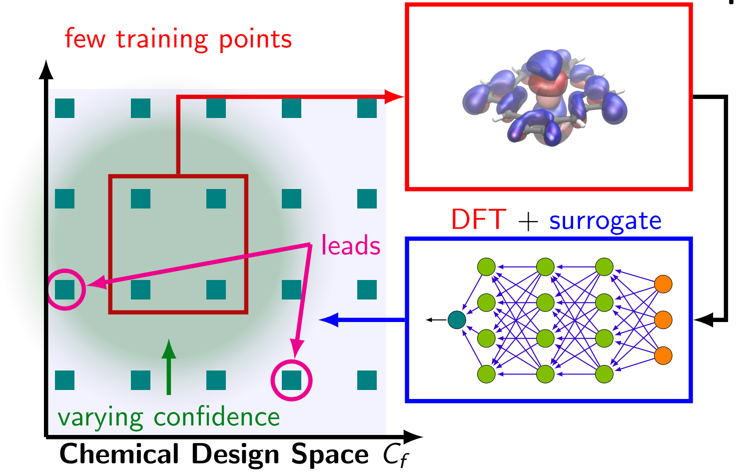
Machine learning and simulation for chemical discovery



Inorganic materials design

- surrogate modeling & integration with simulation
- uncertainty quantification

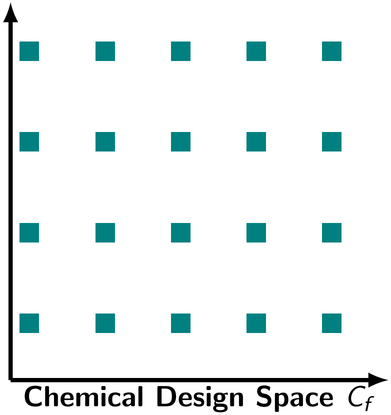
Machine learning and simulation for chemical discovery



Inorganic materials design

- surrogate modeling & integration with simulation
- uncertainty quantification
- Bayesian design

Machine learning and simulation for chemical discovery



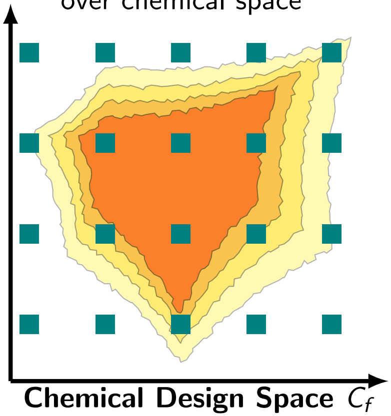
Inorganic materials design

- surrogate modeling & integration with simulation
- uncertainty quantification
- Bayesian design

Drug discovery

Machine learning and simulation for chemical discovery

Alternative: learn a generative distribution over chemical space



Inorganic materials design

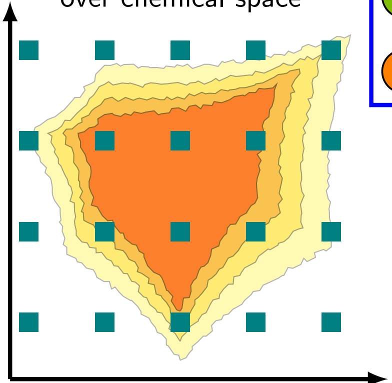
- surrogate modeling & integration with simulation
- uncertainty quantification
- Bayesian design

Drug discovery

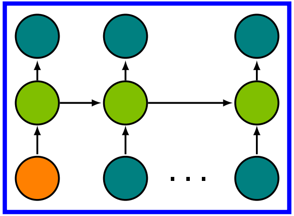
- reinforcement & generative learning

Machine learning and simulation for chemical discovery

Alternative: learn a generative distribution over chemical space



Chemical Design Space C_f



generative agent

Inorganic materials design

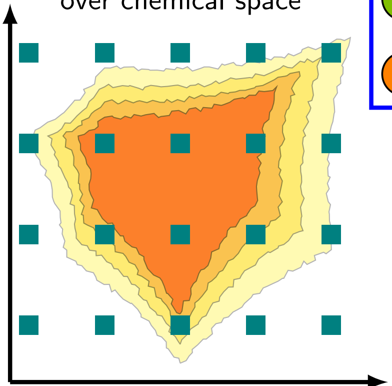
- surrogate modeling & integration with simulation
- uncertainty quantification
- Bayesian design

Drug discovery

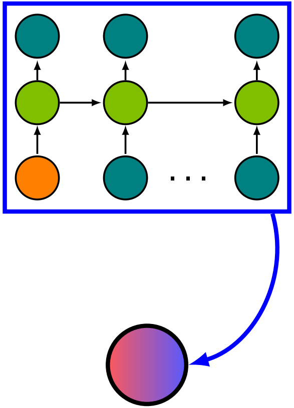
- reinforcement & generative learning

Machine learning and simulation for chemical discovery

Alternative: learn a generative distribution over chemical space



Chemical Design Space C_f



scoring function

Inorganic materials design

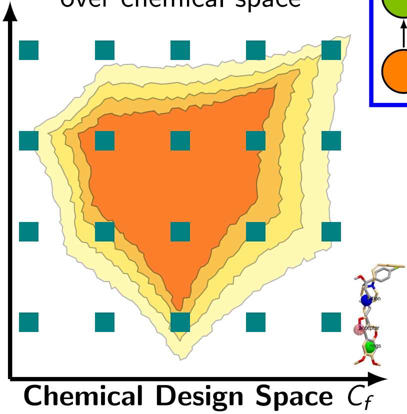
- surrogate modeling & integration with simulation
- uncertainty quantification
- Bayesian design

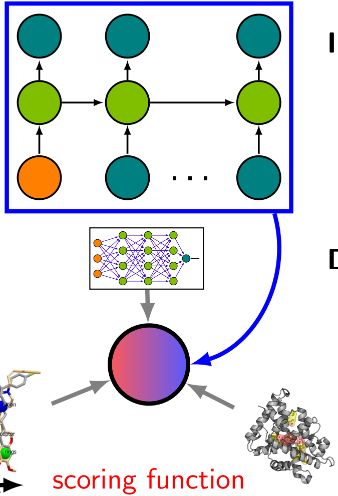
Drug discovery

- reinforcement & generative learning

Machine learning and simulation for chemical discovery

Alternative: learn a generative distribution over chemical space





Inorganic materials design

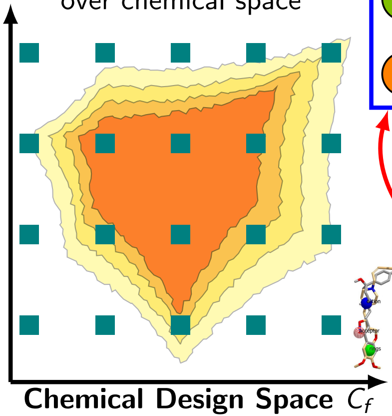
- surrogate modeling & integration with simulation
- uncertainty quantification
- Bayesian design

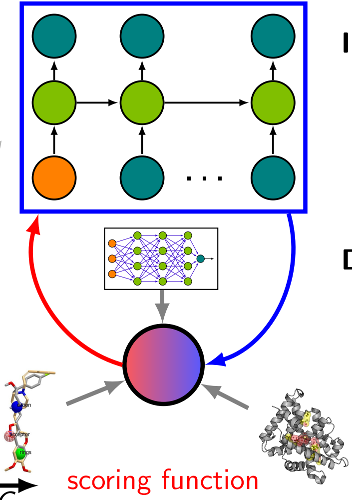
Drug discovery

- reinforcement & generative learning
- 3D/protein conditioning
- curriculum & transfer learning

Machine learning and simulation for chemical discovery

Alternative: learn a generative distribution over chemical space





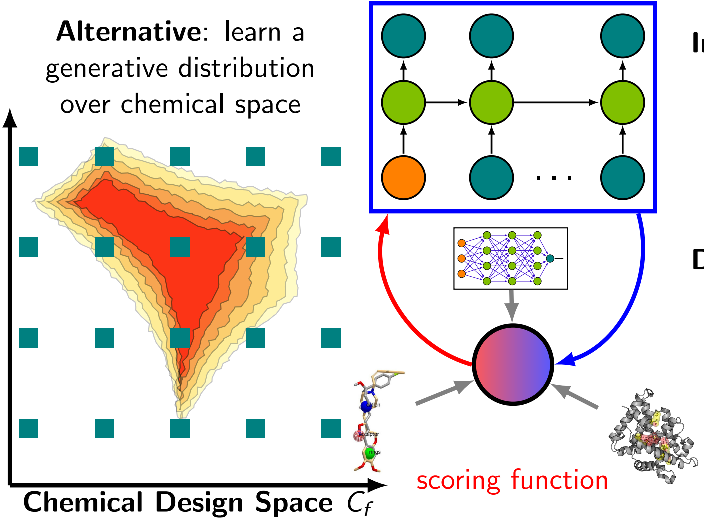
Inorganic materials design

- surrogate modeling & integration with simulation
- uncertainty quantification
- Bayesian design

Drug discovery

- reinforcement & generative learning
- 3D/protein conditioning
- curriculum & transfer learning

Machine learning and simulation for chemical discovery



Inorganic materials design

- surrogate modeling & integration with simulation
- uncertainty quantification
- Bayesian design

Drug discovery

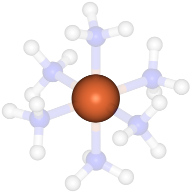
- reinforcement & generative learning
- 3D/protein conditioning
- curriculum & transfer learning

Why transition metal complexes?

Technology requires ever greater control of material properties.

Why transition metal complexes?

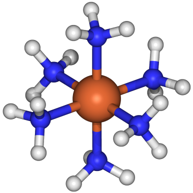
Technology requires ever greater control of material properties.



Transition metal complexes offer precisely **tunable electronic properties**

Why transition metal complexes?

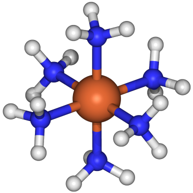
Technology requires ever greater control of material properties.



Transition metal complexes offer precisely **tunable electronic properties**

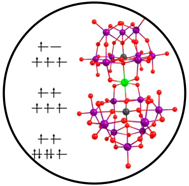
Why transition metal complexes?

Technology requires ever greater control of material properties.



molecular devices

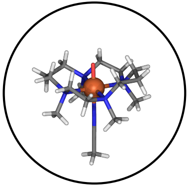
Aravena, D. *et al.*, *Inorg. Chem.*, 55(13):6405–6413, 2016.



Transition metal complexes offer precisely **tunable electronic properties**

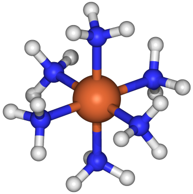
Why transition metal complexes?

Technology requires ever greater control of material properties.



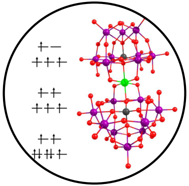
Rohde, J. U. *et al.*, *Science*, 299(5609):1037–1039, 2003.

catalysis



molecular devices

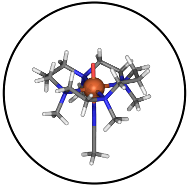
Aravena, D. *et al.*, *Inorg. Chem.*, 55(13):6405–6413, 2016.



Transition metal complexes offer precisely **tunable electronic properties**

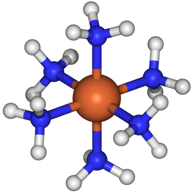
Why transition metal complexes?

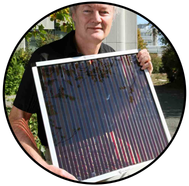
Technology requires ever greater control of material properties.



Rohde, J. U. *et al.*, *Science*, 299(5609):1037–1039, 2003.

catalysis



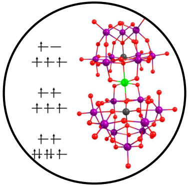


energy capture

Michael Grätzel, photo credit Alain Herzog/EPFL

molecular devices

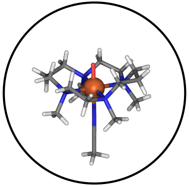
Aravena, D. *et al.*, *Inorg. Chem.*, 55(13):6405–6413, 2016.



Transition metal complexes offer precisely **tunable electronic properties**

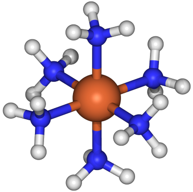
Why transition metal complexes?

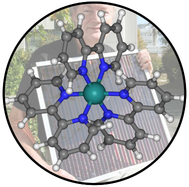
Technology requires ever greater control of material properties.



Rohde, J. U. *et al.*, *Science*, 299(5609):1037–1039, 2003.

catalysis



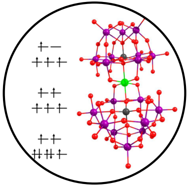


energy capture

Michael Grätzel, photo credit Alain Herzog/EPFL

molecular devices

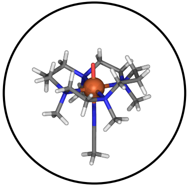
Aravena, D. *et al.*, *Inorg. Chem.*, 55(13):6405–6413, 2016.



Transition metal complexes offer precisely **tunable electronic properties**

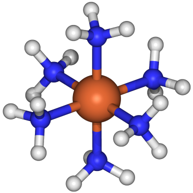
Why transition metal complexes?

Technology requires ever greater control of material properties.



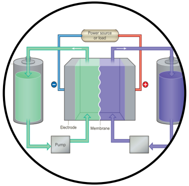
Rohde, J. U. *et al.*, *Science*, 299(5609):1037–1039, 2003.

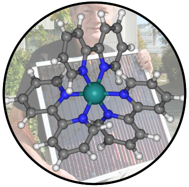
catalysis



commons.wikimedia.org/Redox_Flow_Battery

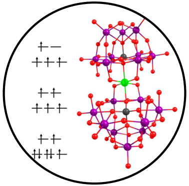
energy storage





energy capture

Michael Grätzel, photo credit Alain Herzog/EPFL



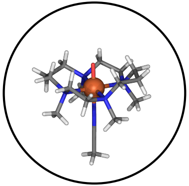
molecular devices

Aravena, D. *et al.*, *Inorg. Chem.*, 55(13):6405–6413, 2016.

Transition metal complexes offer precisely **tunable electronic properties**

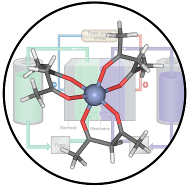
Why transition metal complexes?

Technology requires ever greater control of material properties.



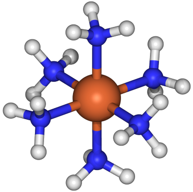
Rohde, J. U. *et al.*, *Science*, 299(5609):1037–1039, 2003.

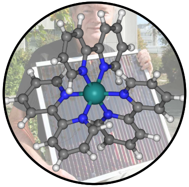
catalysis



commons.wikimedia.org/ Redox_Flow_Battery

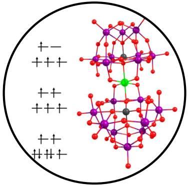
energy storage





energy capture

Michael Grätzel, photo credit Alain Herzog/EPFL



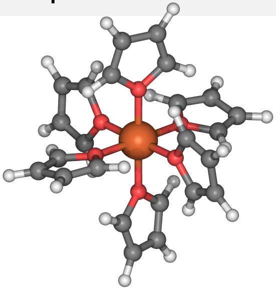
molecular devices

Aravena, D. *et al.*, *Inorg. Chem.*, 55(13):6405–6413, 2016.

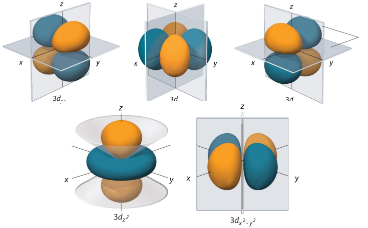
Transition metal complexes offer precisely **tunable electronic properties**

Tunable electronic structure of transition metal complexes

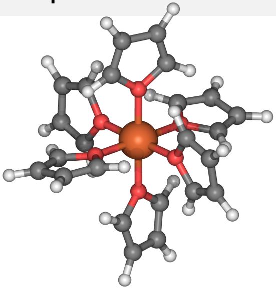
Tunable electronic structure of transition metal complexes



Tunable electronic structure of transition metal complexes



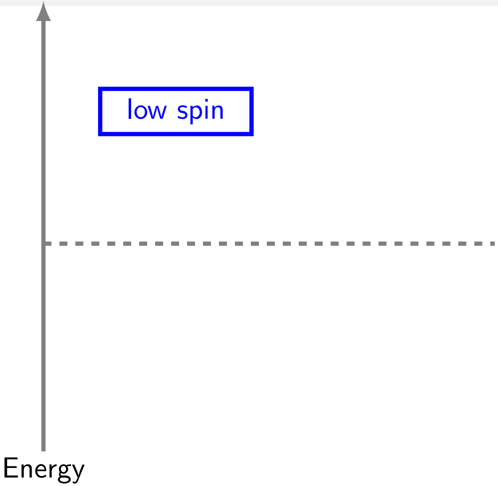
Energy

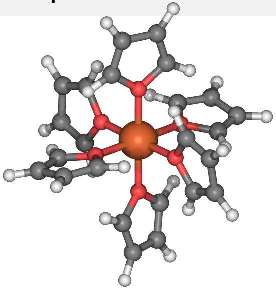


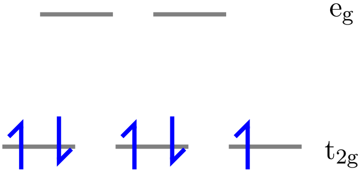
e_g

t_{2g}

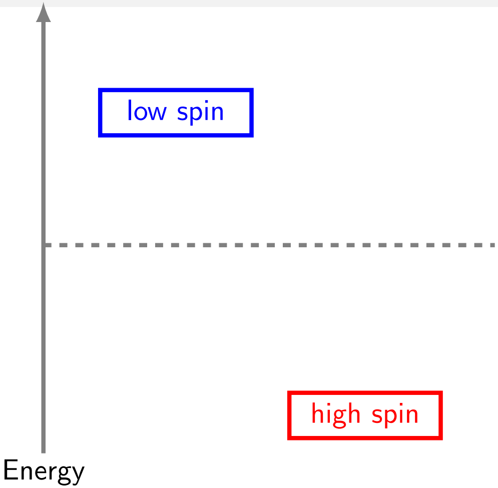
Tunable electronic structure of transition metal complexes

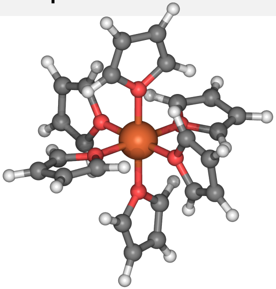


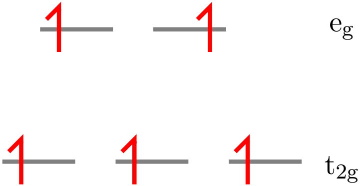




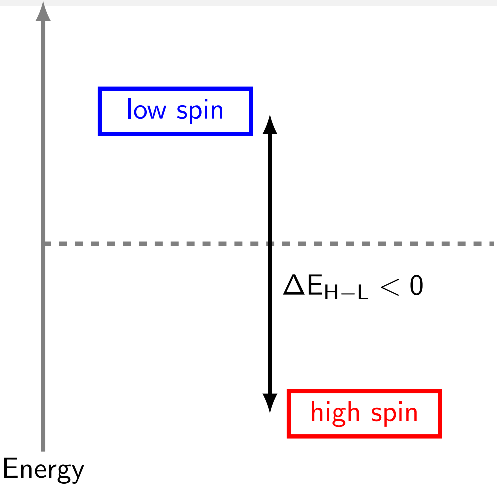
Tunable electronic structure of transition metal complexes

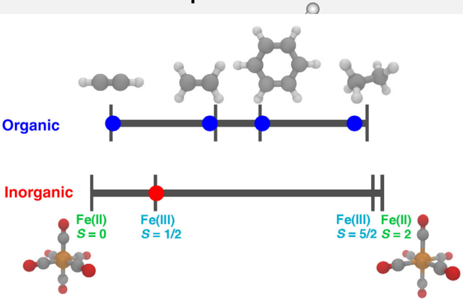






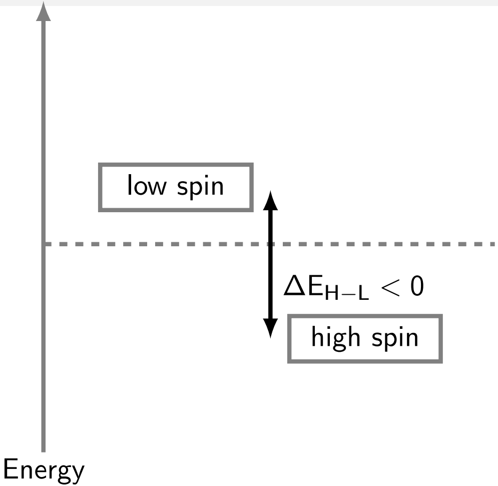
Tunable electronic structure of transition metal complexes

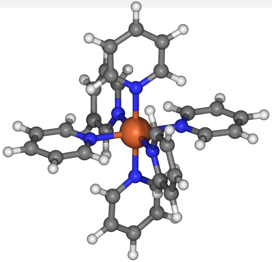




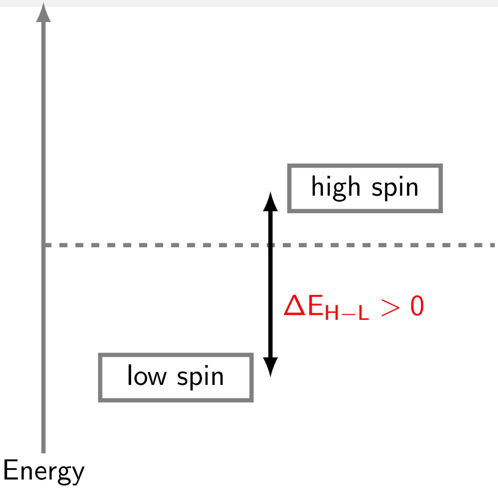
Kulik, HJ., *WIREs Comput Mol Sci.*, e1439, 2019

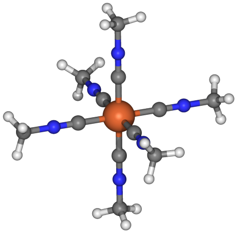
Tunable electronic structure of transition metal complexes



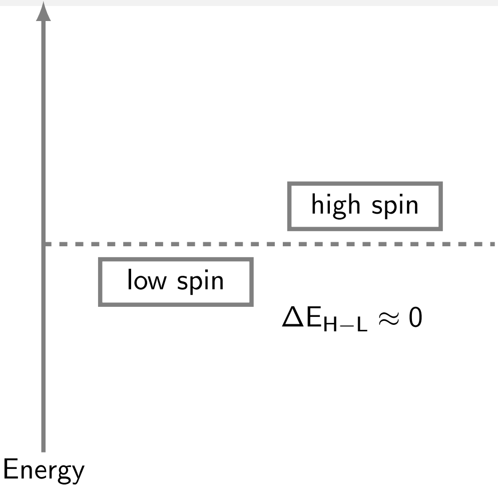


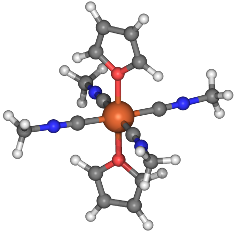
Tunable electronic structure of transition metal complexes



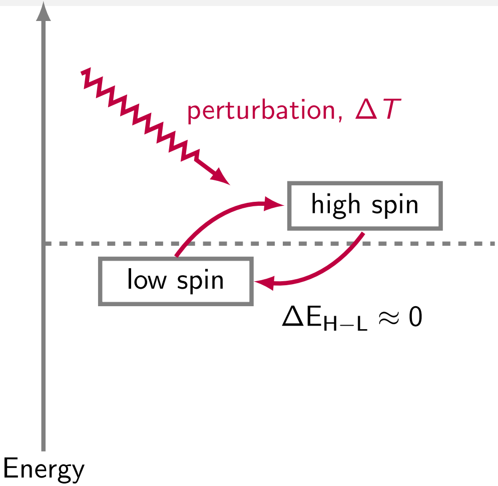


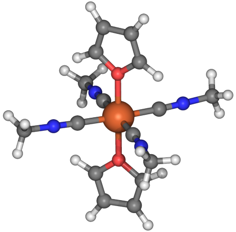
Tunable electronic structure of transition metal complexes

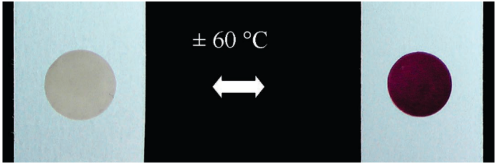




Tunable electronic structure of transition metal complexes

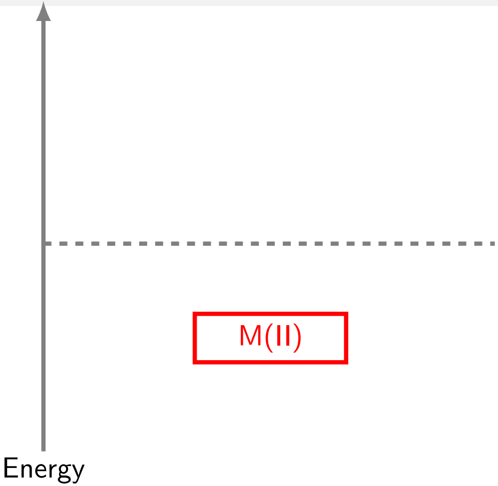


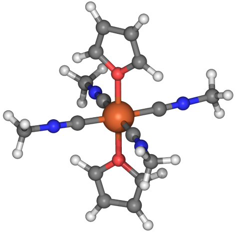


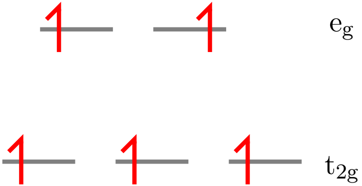


Seredyuk, M *et al.*, *Chem. Mater.*, 18(10):2513–2519, 2006.

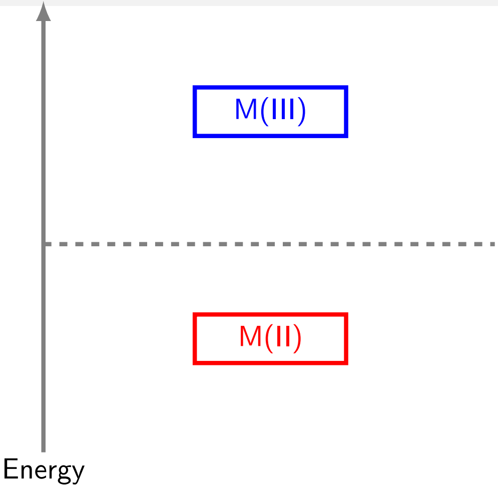
Tunable electronic structure of transition metal complexes

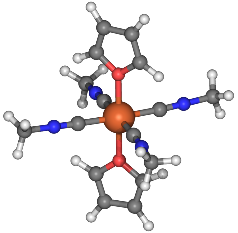


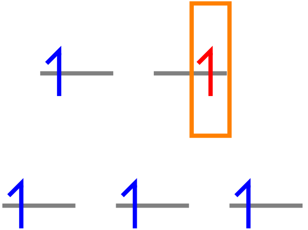




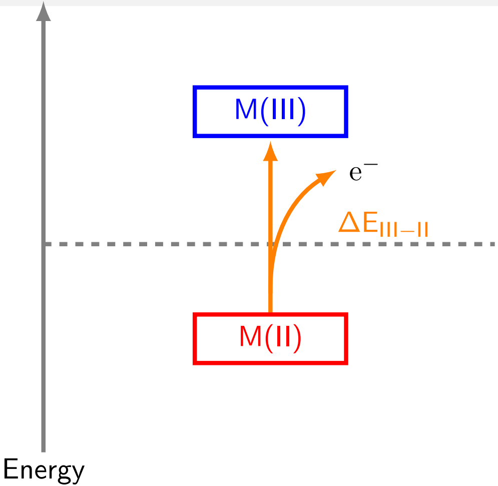
Tunable electronic structure of transition metal complexes

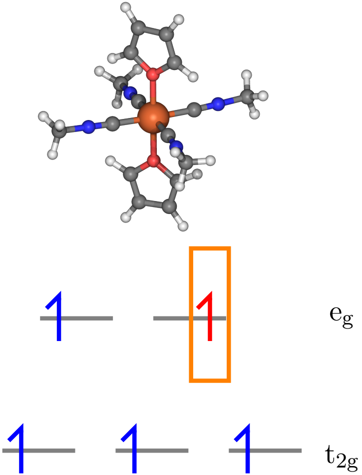






Tunable electronic structure of transition metal complexes





Overview: algorithmic surrogate-drive chemical optimization

design
space

Overview: algorithmic surrogate-drive chemical optimization

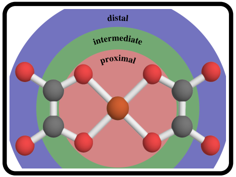
design space

candidate materials

Overview: algorithmic surrogate-drive chemical optimization

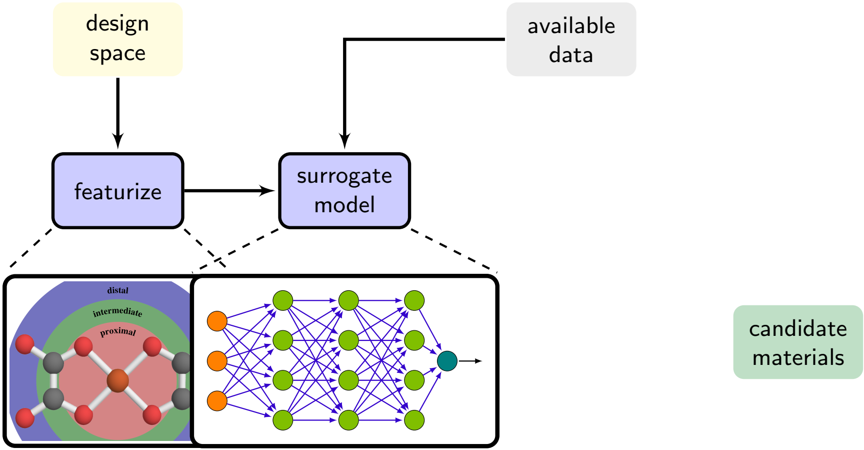
design space

featurize

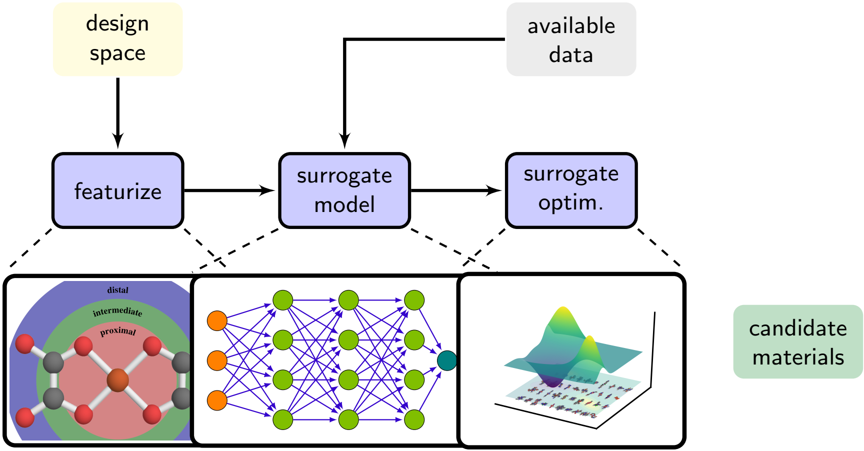


candidate materials

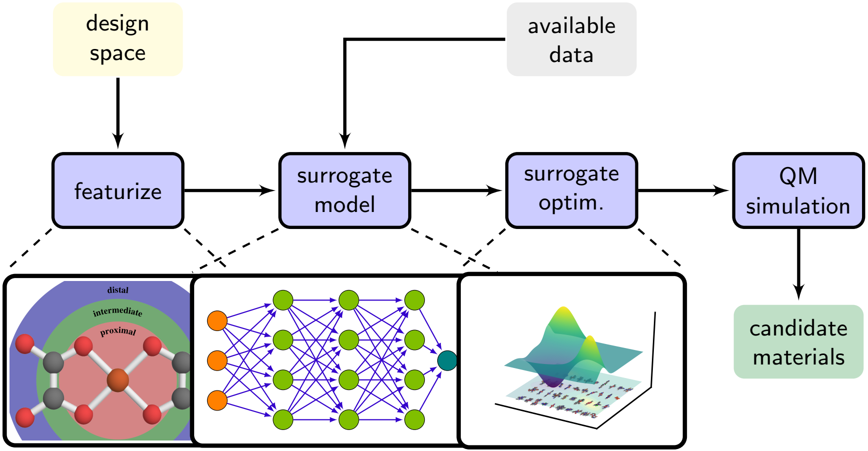
Overview: algorithmic surrogate-drive chemical optimization



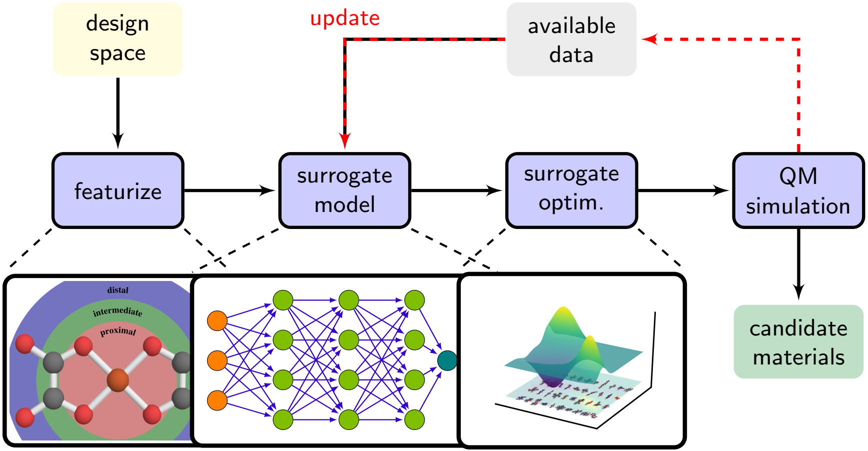
Overview: algorithmic surrogate-drive chemical optimization



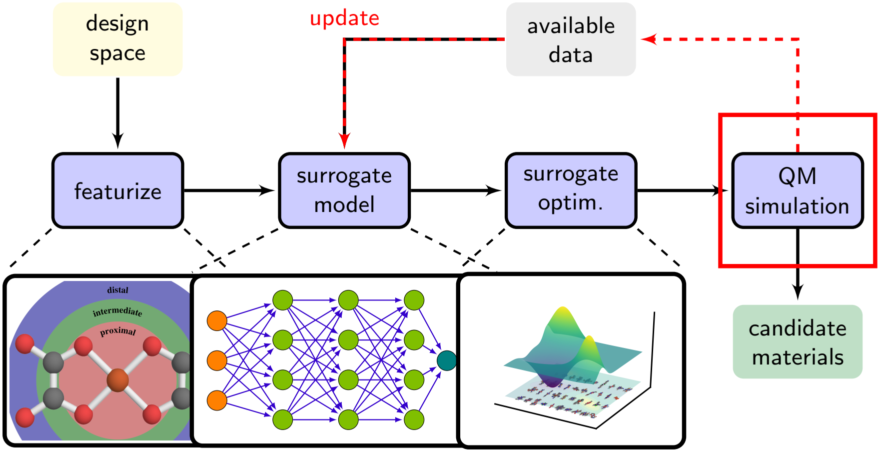
Overview: algorithmic surrogate-drive chemical optimization



Overview: algorithmic surrogate-drive chemical optimization



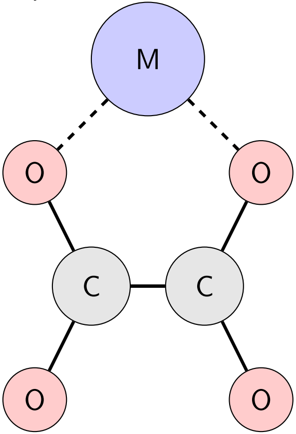
Overview: algorithmic surrogate-drive chemical optimization



Not discussed (much) further, but not easy and we can use ML to help manage uncertainty

Featurization with RACs

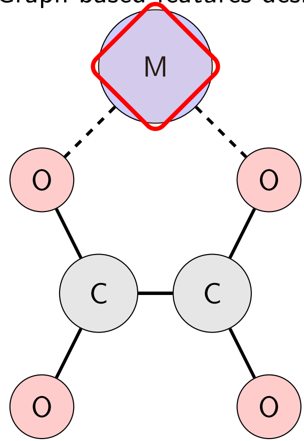
Graph-based features designed for TM complexes:



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

Featurization with RACs

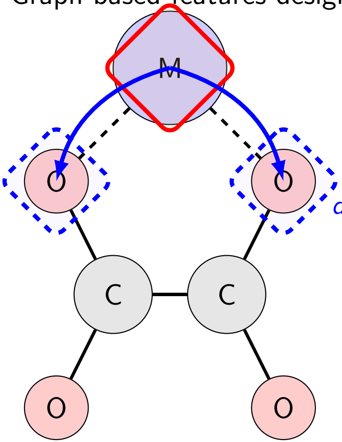
Graph-based features designed for TM complexes:



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

Featurization with RACs

Graph-based features designed for TM complexes:

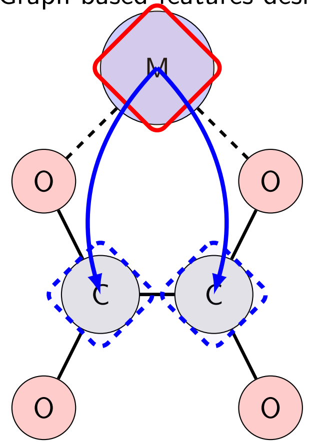


$d_1 : \sum_{M,O} Z_M Z_O$

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

Featurization with RACs

Graph-based features designed for TM complexes:



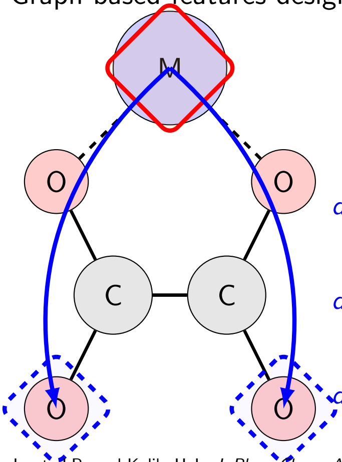
$$d_1 : \sum_{M,O} Z_M Z_O$$

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

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

Featurization with RACs

Graph-based features designed for TM complexes:



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

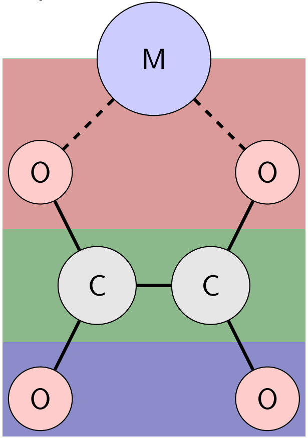
$d_2 : \sum_{M,C} Z_M Z_C$

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

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

Featurization with RACs

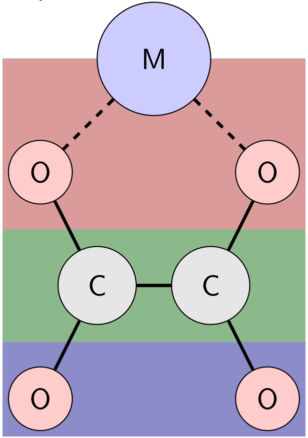
Graph-based features designed for TM complexes:

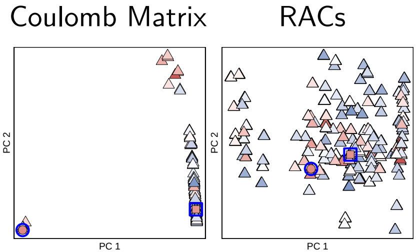


Janet, J.P., and Kulik, H.J., *J. Phys. Chem. A*, 121(46):8939–8954, 2017.

Featurization with RACs

Graph-based features designed for TM complexes:

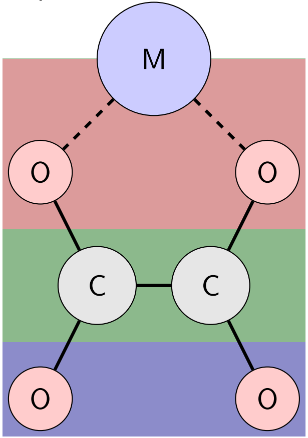


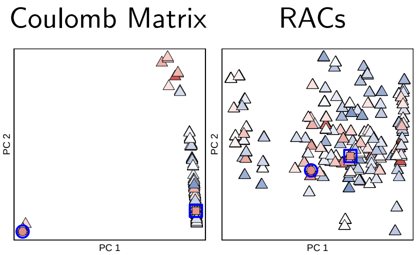


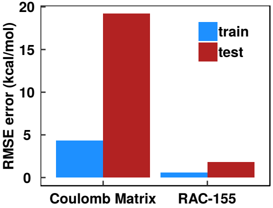
Janet, J.P., and Kulik, H.J., *J. Phys. Chem. A*, 121(46):8939–8954, 2017.

Featurization with RACs

Graph-based features designed for TM complexes:





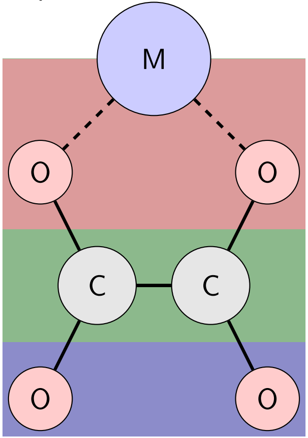


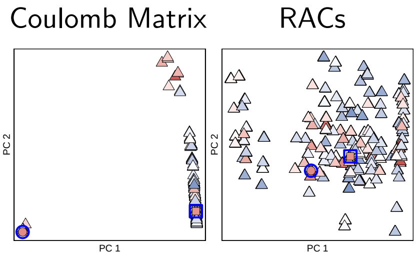
leads to $\sim 20\times$ more accurate property prediction without 3D geometric information

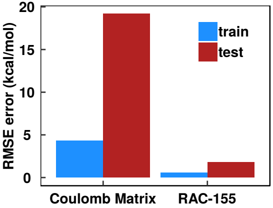
Janet, J.P., and Kulik, H.J., *J. Phys. Chem. A*, 121(46):8939–8954, 2017.

Featurization with RACs

Graph-based features designed for TM complexes:





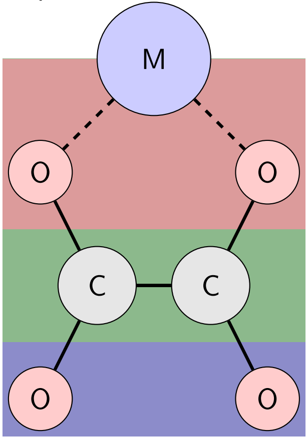


leads to $\sim 20\times$ more accurate property prediction without 3D geometric information

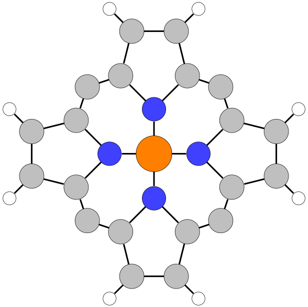
Janet, J.P., and Kulik, H.J., *J. Phys. Chem. A*, 121(46):8939–8954, 2017.

Featurization with RACs

Graph-based features designed for TM complexes:



spin splitting ΔE_{H-L}

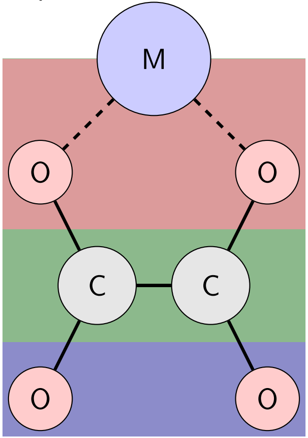


redox ΔG_{solv}

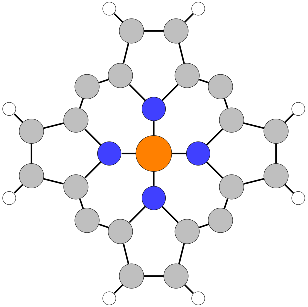
Janet, J.P., and Kulik, H.J., *J. Phys. Chem. A*, 121(46):8939–8954, 2017.

Featurization with RACs

Graph-based features designed for TM complexes:



spin splitting ΔE_{H-L}

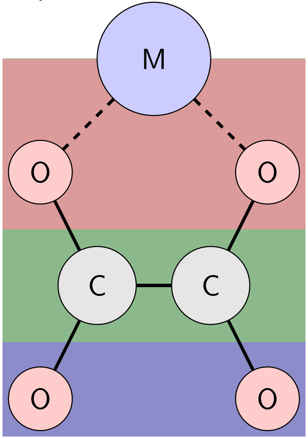


redox ΔG_{solv}

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

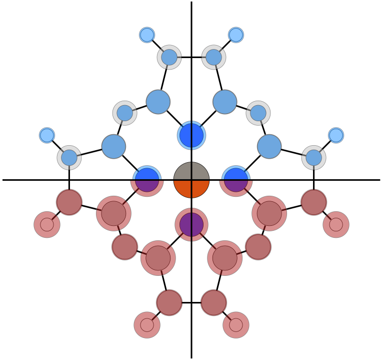
Featurization with RACs

Graph-based features designed for TM complexes:



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

spin splitting ΔE_{H-L}

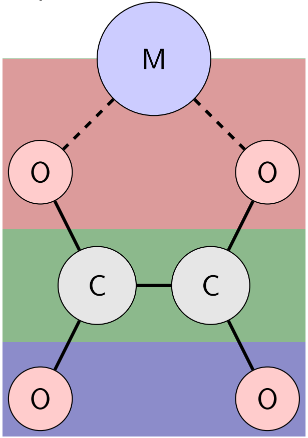


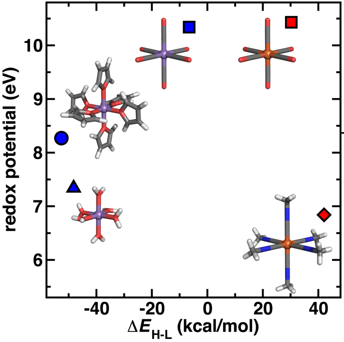
redox ΔG_{solv}

Janet, J.P. *et al.*, *Inorg. Chem.*, 58(16):10592–10606, 2019.

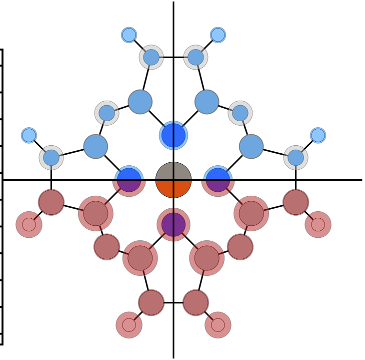
Featurization with RACs

Graph-based features designed for TM complexes:





spin splitting ΔE_{H-L}



redox ΔG_{solv}

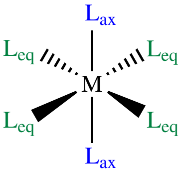
Janet, J.P., and Kulik, H.J., *J. Phys. Chem. A*, 121(46):8939–8954, 2017. Janet, J.P. *et al.*, *Inorg. Chem.*, 58(16):10592–10606, 2019.

Property inference with ML

We can accurately estimate DFT properties using simple neural networks

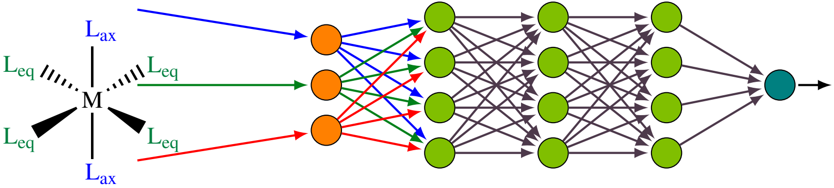
Property inference with ML

We can accurately estimate DFT properties using simple neural networks



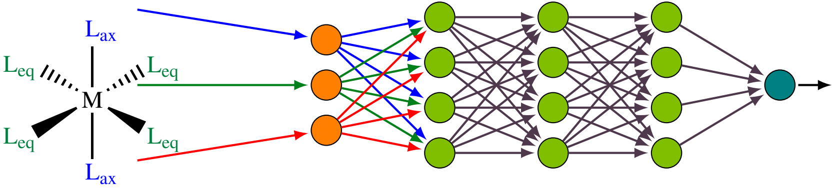
Property inference with ML

We can accurately estimate DFT properties using simple neural networks

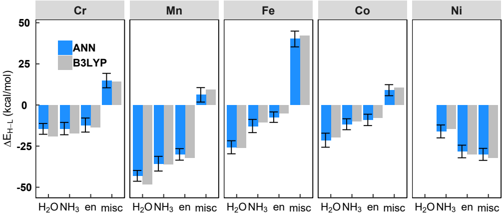


Property inference with ML

We can accurately estimate DFT properties using simple neural networks



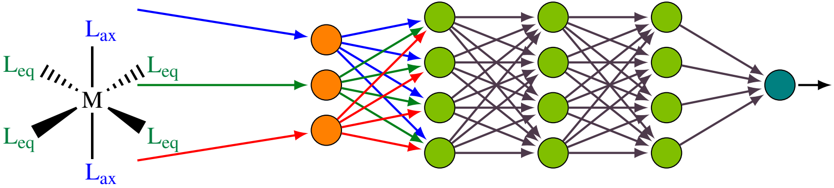
spin splitting energies



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

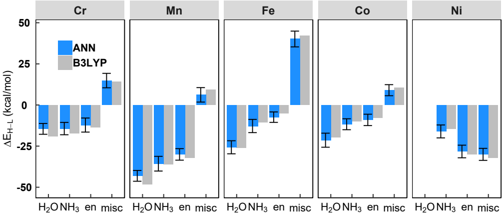
Property inference with ML

We can accurately estimate DFT properties using simple neural networks



spin splitting energies

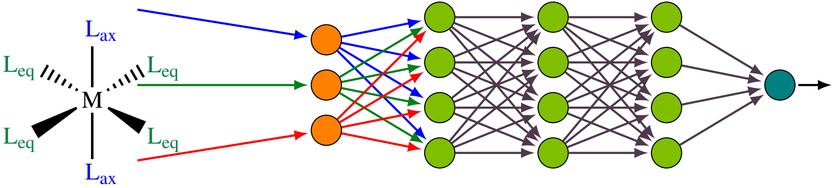
- negligible cost relative to DFT (20 GPU-hours)
- errors of 1–3 kcal/mol on similar unseen cases



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

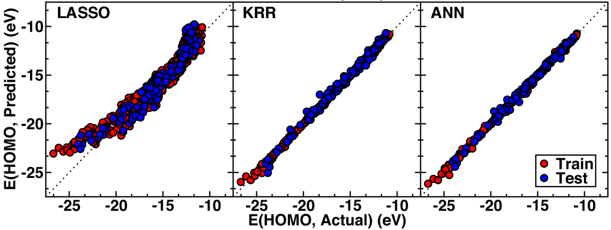
Property inference with ML

We can accurately estimate DFT properties using simple neural networks



frontier orbital properties

- negligible cost relative to DFT (20 GPU-hours)
- errors of 0.2 eV on similar unseen cases

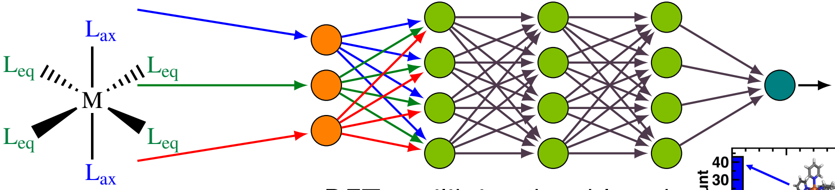


Janet, J.P. and Kulik, H.J., *Chem. Sci.*, 8:5137–5152, 2017.

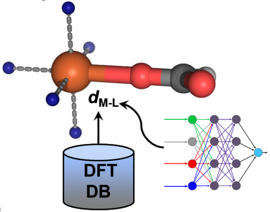
Nandy, A. *et al.*, *Ind. Eng. Chem. Res.*, 57(42):13973–13986, 2018.

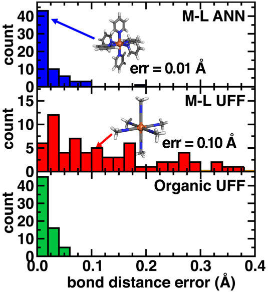
Property inference with ML

We can accurately estimate DFT properties using simple neural networks



DFT equilibrium bond lengths





- negligible cost relative to DFT (20 GPU-hours)
- errors of 0.1 pm on similar unseen cases

Janet, J.P. *et al.*, *Inorg. Chem.*, 58(16):10592–10606, 2019.
Janet, J.P. *et al.*, *Ind. Eng. Chem. Res.*, 56(17):4898–4910, 2017.
Janet, J.P. and Kulik, H.J., *Chem. Sci.*, 8:5137–5152, 2017.
Nandy, A. *et al.*, *Ind. Eng. Chem. Res.*, 57(42):13973–13986, 2018.

Beyond prediction: live job management

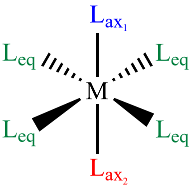
In high-throughput DFT screening, job failure is a frequent issue

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

Beyond prediction: live job management

In high-throughput DFT screening, job failure is a frequent issue

new complex

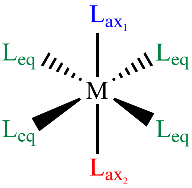


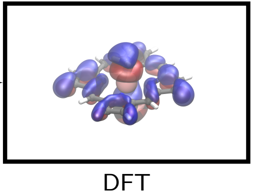
Duan, C., Janet, J.P. *et al.*, *J. Chem. Theory. Comp.*, 15(4):2331–2345, 2019.

Beyond prediction: live job management

In high-throughput DFT screening, job failure is a frequent issue

new complex →

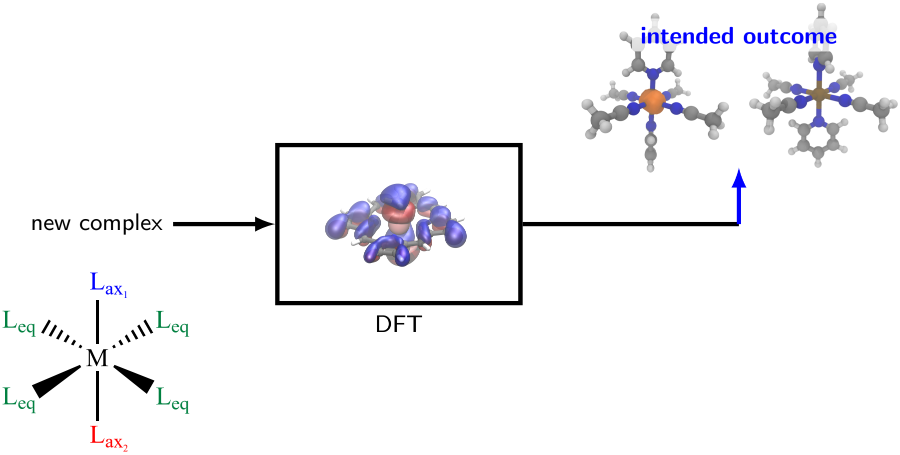




Duan, C., Janet, J.P. *et al.*, *J. Chem. Theory. Comp.*, 15(4):2331–2345, 2019.

Beyond prediction: live job management

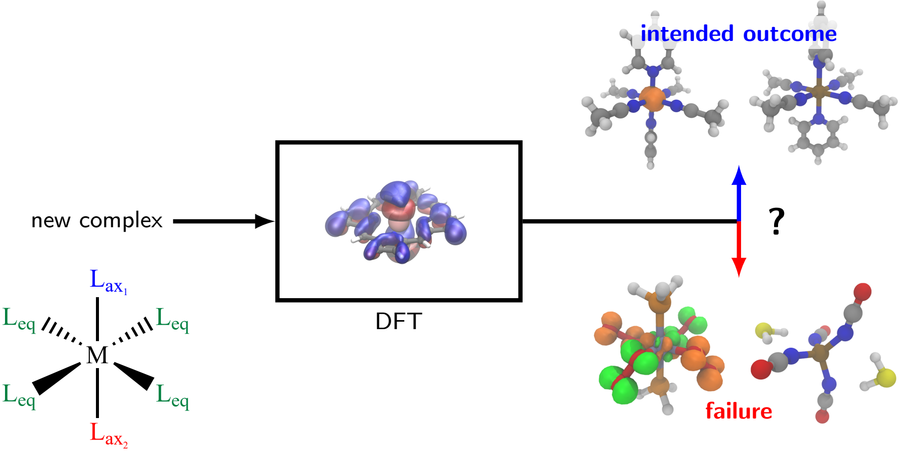
In high-throughput DFT screening, job failure is a frequent issue



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

Beyond prediction: live job management

In high-throughput DFT screening, job failure is a frequent issue



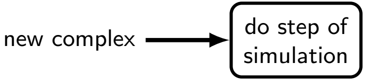
Duan, C., Janet, J.P. *et al.*, *J. Chem. Theory. Comp.*, 15(4):2331–2345, 2019.

Beyond prediction: live job management

new complex

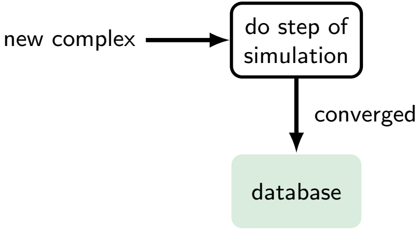
Duan, C., Janet, J.P. *et al.*, *J. Chem. Theory. Comp.*, 15(4):2331–2345, 2019.

Beyond prediction: live job management



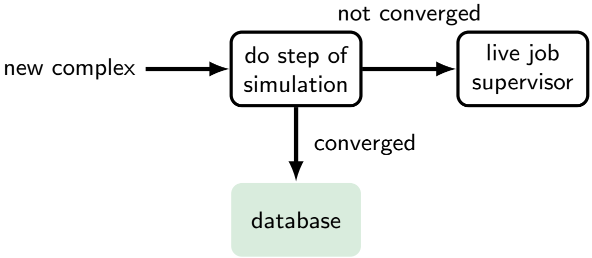
Duan, C., Janet, J.P. *et al.*, *J. Chem. Theory. Comp.*, 15(4):2331–2345, 2019.

Beyond prediction: live job management



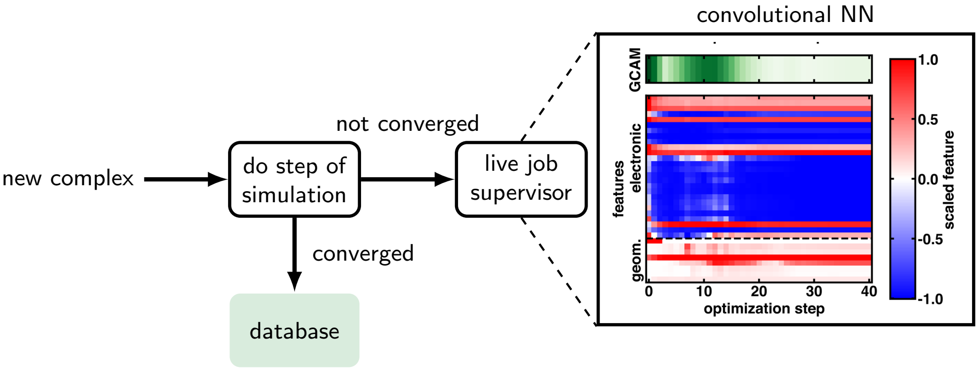
Duan, C., Janet, J.P. *et al.*, *J. Chem. Theory. Comp.*, 15(4):2331–2345, 2019.

Beyond prediction: live job management



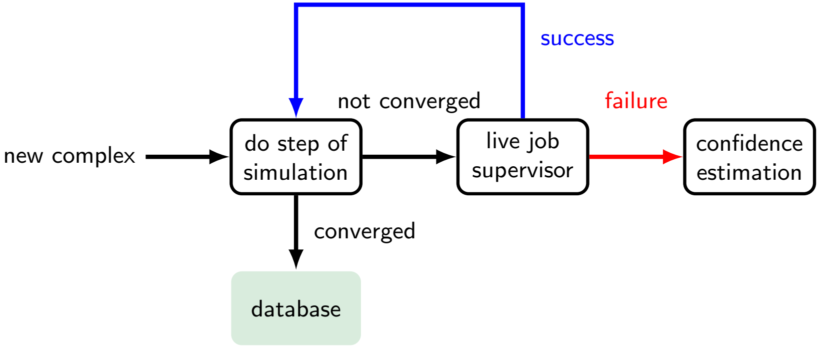
Duan, C., Janet, J.P. *et al.*, *J. Chem. Theory. Comp.*, 15(4):2331–2345, 2019.

Beyond prediction: live job management



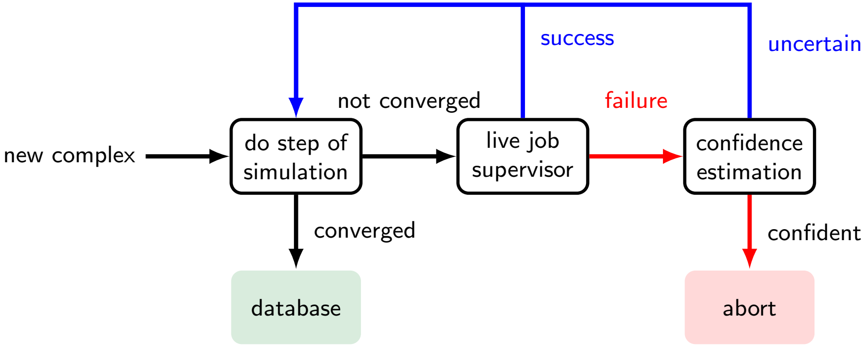
Duan, C., Janet, J.P. *et al.*, *J. Chem. Theory. Comp.*, 15(4):2331–2345, 2019.

Beyond prediction: live job management



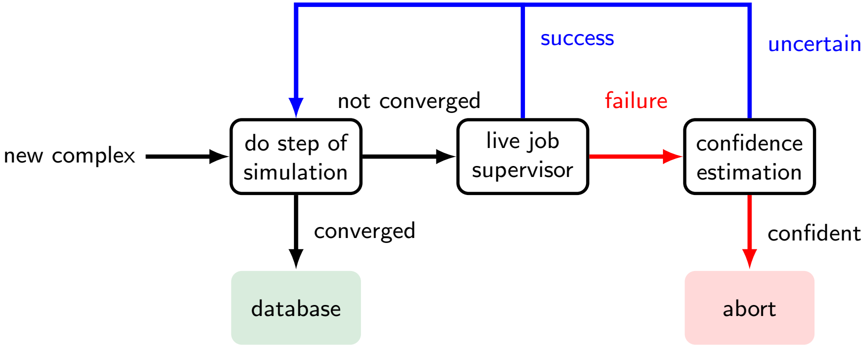
Duan, C., Janet, J.P. *et al.*, *J. Chem. Theory. Comp.*, 15(4):2331–2345, 2019.

Beyond prediction: live job management



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

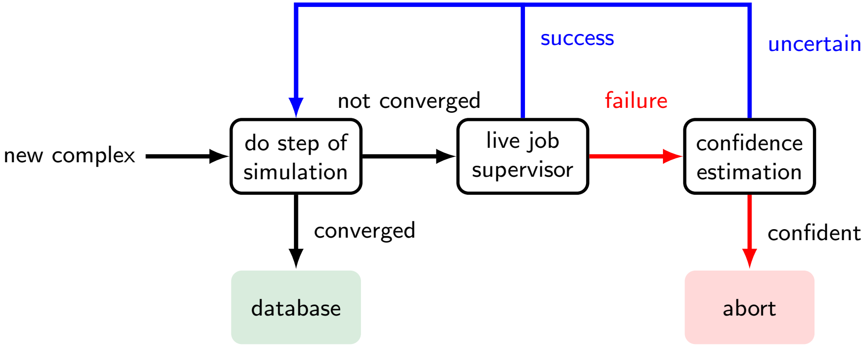
Beyond prediction: live job management



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

Beyond prediction: live job management

This approach leads to **30% time savings** and can abort almost all failures.



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

How far can we extrapolate?

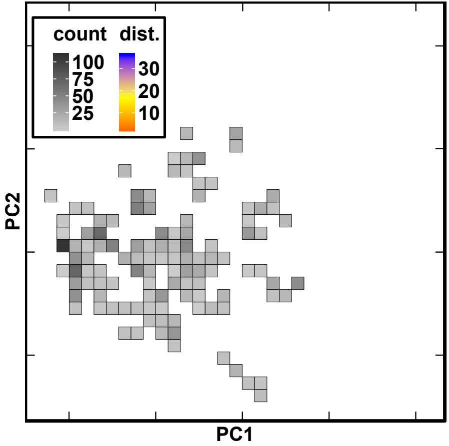
Out-of-distribution test:

Computed spin-splitting energies of 116 diverse structures from the CSD, from training-like to very different.

How far can we extrapolate?

Out-of-distribution test:

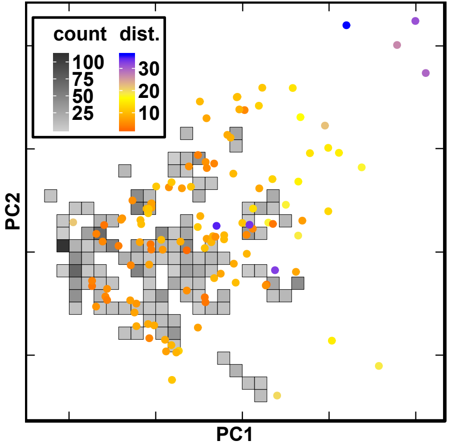
Computed spin-splitting energies of 116 diverse structures from the CSD, from training-like to very different.



How far can we extrapolate?

Out-of-distribution test:

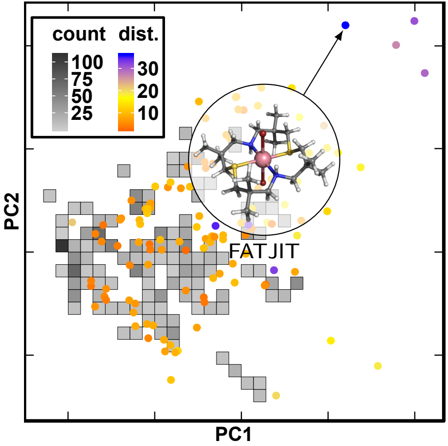
Computed spin-splitting energies of 116 diverse structures from the CSD, from training-like to very different.



How far can we extrapolate?

Out-of-distribution test:

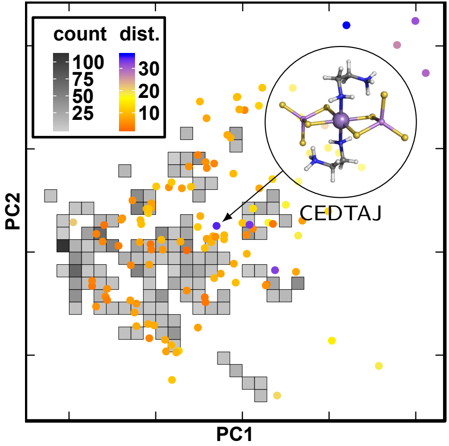
Computed spin-splitting energies of 116 diverse structures from the CSD, from training-like to very different.



How far can we extrapolate?

Out-of-distribution test:

Computed spin-splitting energies of 116 diverse structures from the CSD, from training-like to very different.

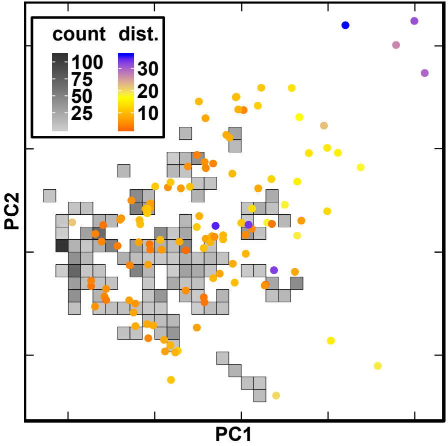


How far can we extrapolate?

Out-of-distribution test:

Computed spin-splitting energies of 116 diverse structures from the CSD, from training-like to very different.

Test generalization performance of MLP trained on 1900 DFT results for complexes with common small ligands.

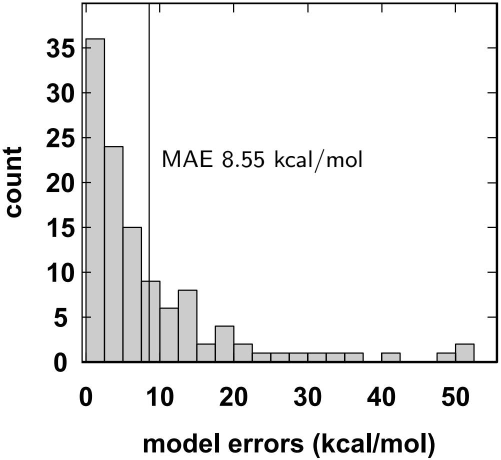


How far can we extrapolate?

Out-of-distribution test:

Computed spin-splitting energies of 116 diverse structures from the CSD, from training-like to very different.

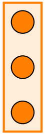
Test generalization performance of MLP trained on 1900 DFT results for complexes with common small ligands.



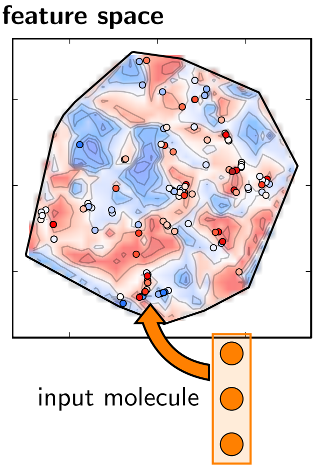
Latent distance similarity

Latent distance similarity

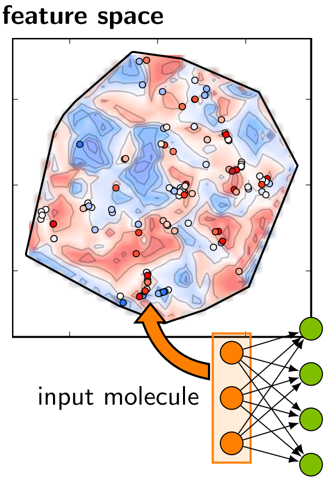
input molecule



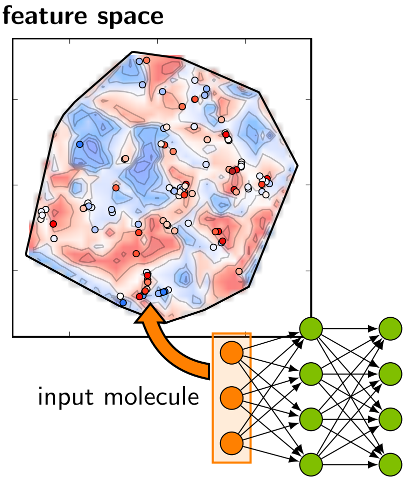
Latent distance similarity



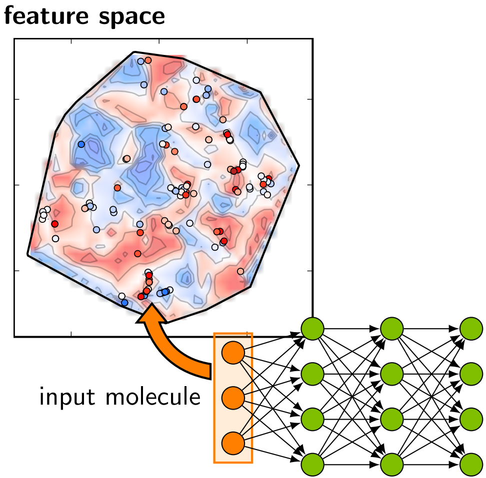
Latent distance similarity



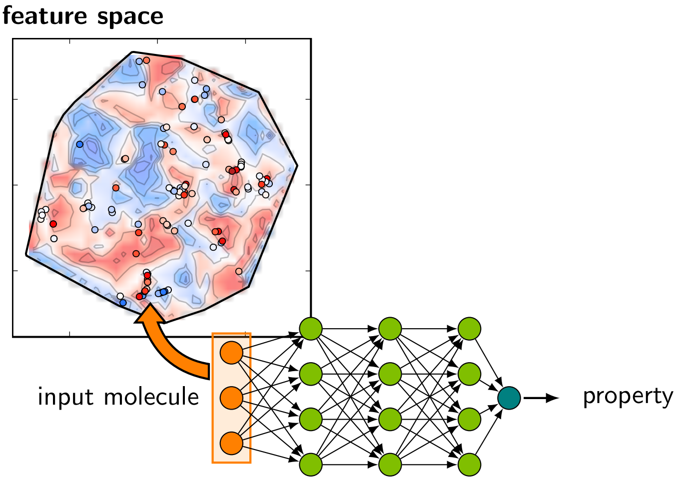
Latent distance similarity



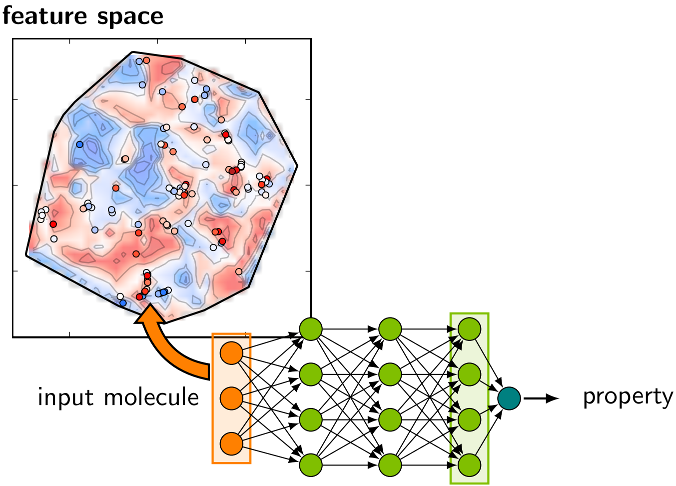
Latent distance similarity



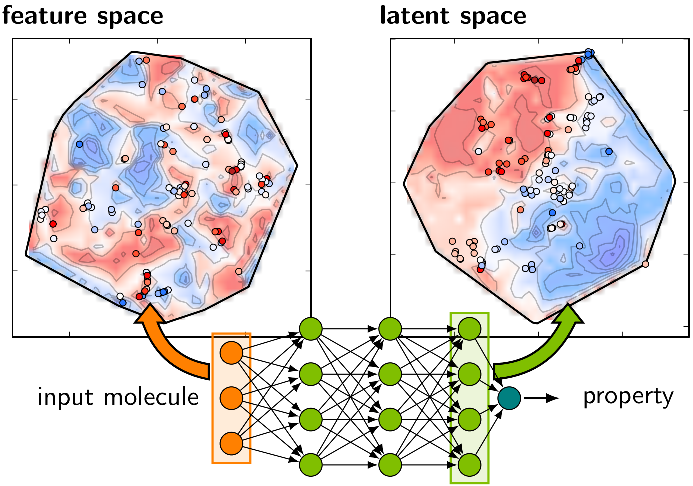
Latent distance similarity



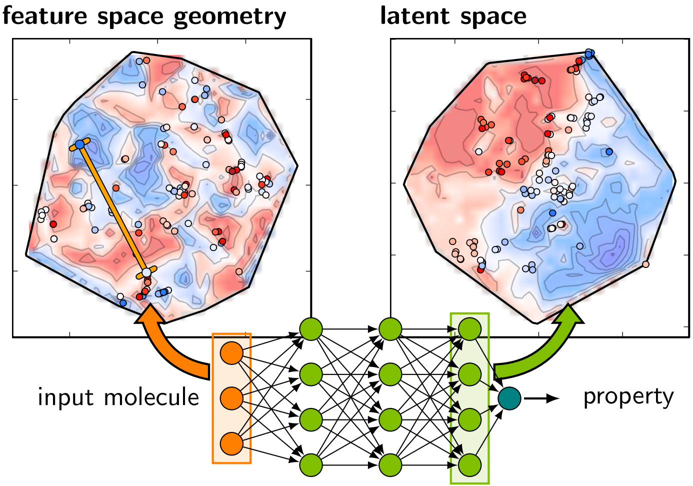
Latent distance similarity



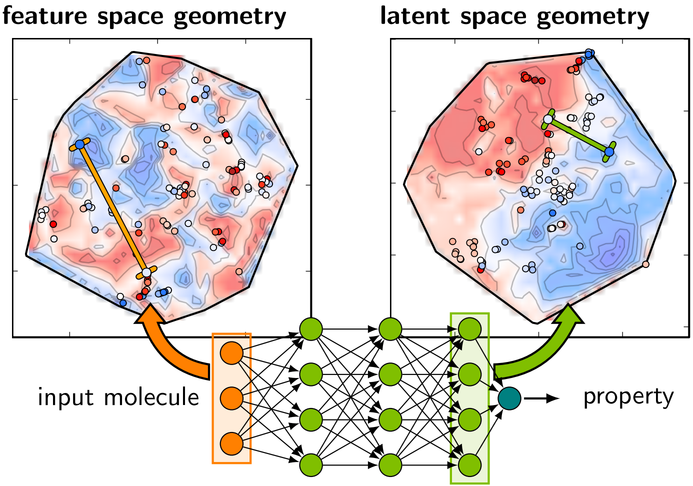
Latent distance similarity



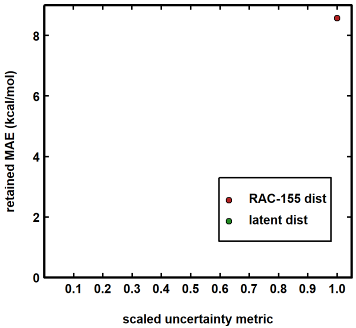
Latent distance similarity



Latent distance similarity

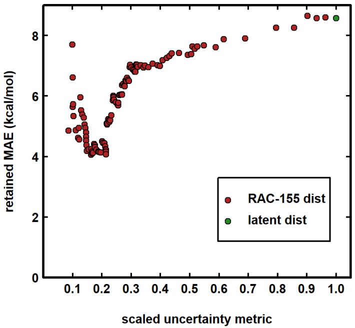


How to measure chemical extrapolation?



Janet, J.P., *et al.*, *Chem. Sci.*, 10:7913–7922, 2019.

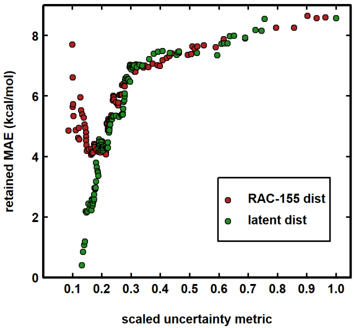
How to measure chemical extrapolation?



Janet, J.P., *et al.*, *Chem. Sci.*, 10:7913–7922, 2019.

How to measure chemical extrapolation?

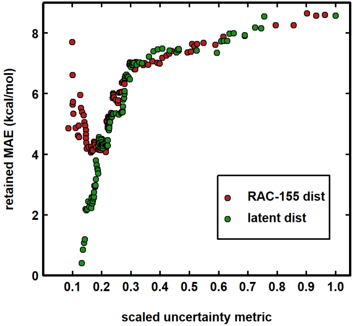
latent distances are superior to feature space distances

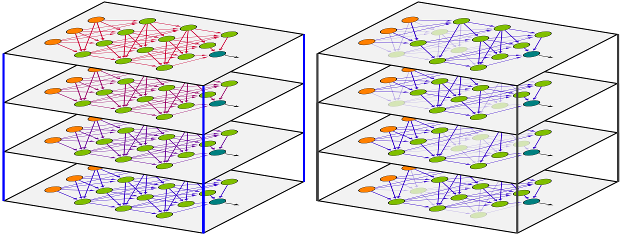


Janet, J.P., *et al.*, *Chem. Sci.*, 10:7913–7922, 2019.

How to measure chemical extrapolation?

latent distances are superior to feature space distances





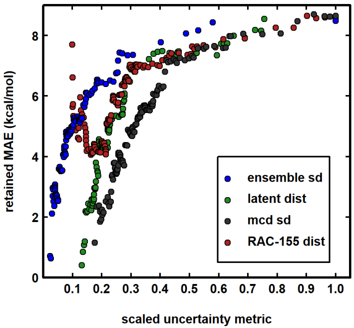
ensembles

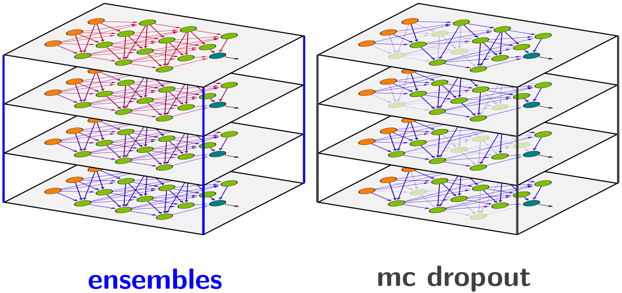
mc dropout

Janet, J.P., *et al.*, *Chem. Sci.*, 10:7913–7922, 2019.

How to measure chemical extrapolation?

latent distances are superior to feature space distances



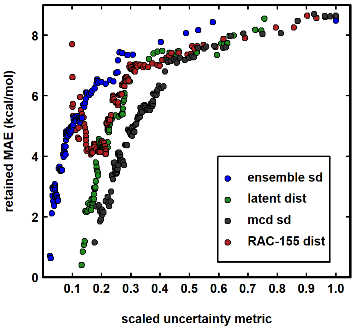


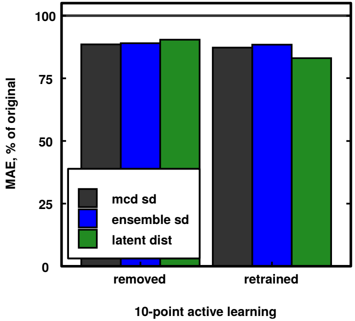
competitive with ensembles and dropout

Janet, J.P., *et al.*, *Chem. Sci.*, 10:7913–7922, 2019.

How to measure chemical extrapolation?

latent distances are superior to feature space distances



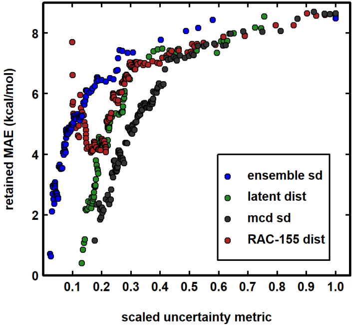


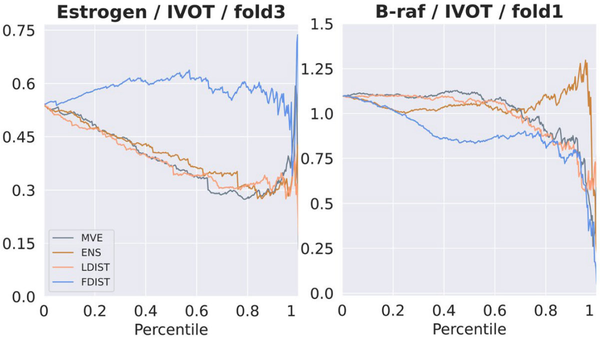
competitive with ensembles and dropout

Janet, J.P., *et al.*, *Chem. Sci.*, 10:7913–7922, 2019.

How to measure chemical extrapolation?

latent distances are superior to feature space distances





Wang, D., *et al.*, *J. Cheminform.*, 13:69, 2021
Hirschfeld, L., *et al.*, *J. Chem. Inf. Model.*, 60(8):3770-–3780, 2020

competitive with ensembles and dropout and works well for **active learning**

Janet, J.P., *et al.*, *Chem. Sci.*, 10:7913–7922, 2019.

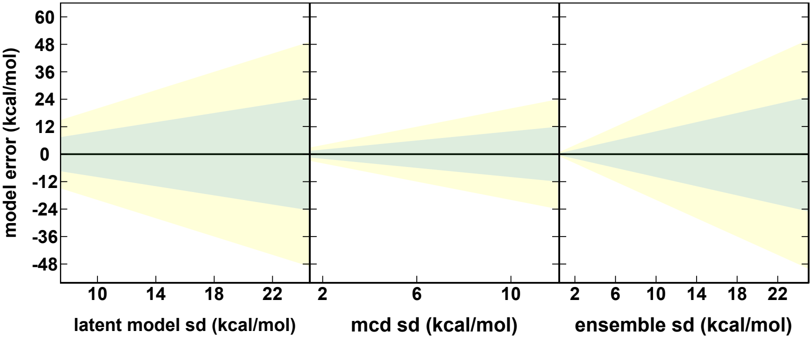
From distances to error distributions

We need to provide confidence estimates in appropriate units:

From distances to error distributions

We need to provide confidence estimates in appropriate units:

$$\varepsilon(d) \sim \mathcal{N}\left(0, \sigma_1^2 + d\sigma_2^2\right)$$

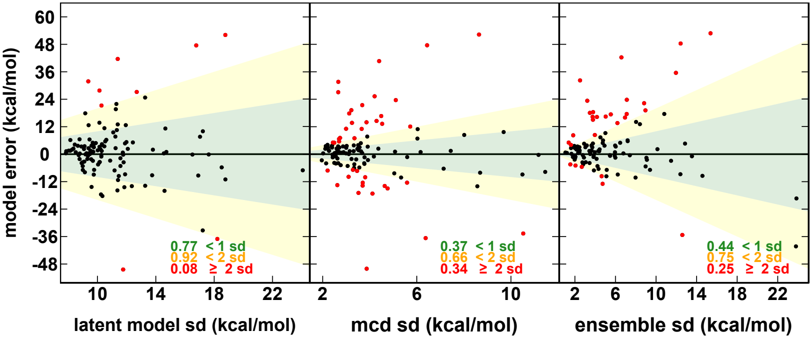


Janet, J.P., *et al.*, *Chem. Sci.*, 10:7913–7922 , 2019.

From distances to error distributions

We need to provide confidence estimates in appropriate units:

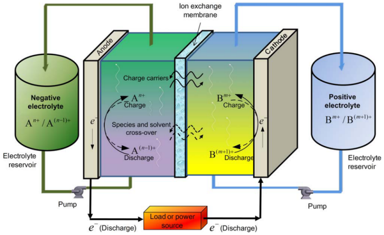
$$\varepsilon(d) \sim \mathcal{N}\left(0, \sigma_1^2 + d\sigma_2^2\right)$$



Janet, J.P., *et al.*, *Chem. Sci.*, 10:7913–7922 , 2019.

Redox flow batteries

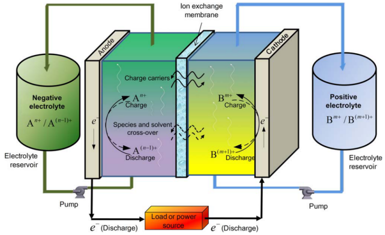
Redox flow batteries (RFBs) are a promising option for scalable energy storage:



Perry, M.L. and Adam, Z., *J. Electrochem. Soc.*, 163(1):A5064–A5067, 2018.

Redox flow batteries

Redox flow batteries (RFBs) are a promising option for scalable energy storage:

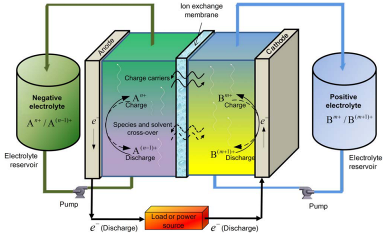


Perry, M.L. and Adam, Z., *J. Electrochem. Soc.*, 163(1):A5064–A5067, 2018.

Transition metal complexes make attractive redox couples for RFBs

Redox flow batteries

Redox flow batteries (RFBs) are a promising option for scalable energy storage:



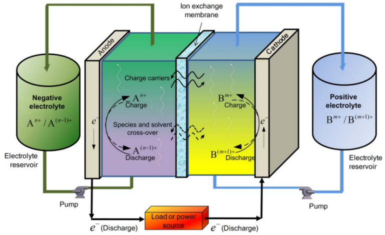
Perry, M.L. and Adam, Z., *J. Electrochem. Soc.*, 163(1):A5064–A5067, 2018.

Transition metal complexes make attractive redox couples for RFBs

- good ion stability (compared to organics)

Redox flow batteries

Redox flow batteries (RFBs) are a promising option for scalable energy storage:



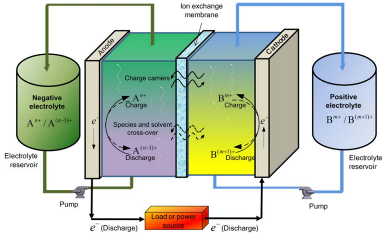
Perry, M.L. and Adam, Z., *J. Electrochem. Soc.*, 163(1):A5064–A5067, 2018.

Transition metal complexes make attractive redox couples for RFBs

- good ion stability (compared to organics)
- good range of redox potentials available

Redox flow batteries

Redox flow batteries (RFBs) are a promising option for scalable energy storage:



Perry, M.L. and Adam, Z., *J. Electrochem. Soc.*, 163(1):A5064–A5067, 2018.

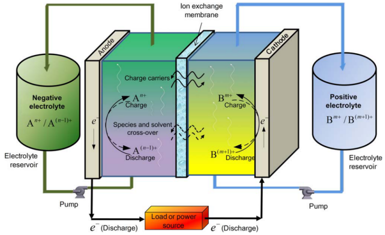
Transition metal complexes make attractive redox couples for RFBs

- good ion stability (compared to organics)
- good range of redox potentials available
- solubility is an issue!

$$E_{\text{cell}} = 0.5 \times \Delta G_{\text{solv}} \times C \times n \times F$$

Redox flow batteries

Redox flow batteries (RFBs) are a promising option for scalable energy storage:



Perry, M.L. and Adam, Z., *J. Electrochem. Soc.*, 163(1):A5064–A5067, 2018.

Transition metal complexes make attractive redox couples for RFBs

- good ion stability (compared to organics)
- good range of redox potentials available
- solubility is an issue!

$$E_{\text{cell}} = 0.5 \times \Delta G_{\text{solv}} \times C \times n \times F$$

We need complexes that have high redox potential **and** good solubility

Multiobjective optimization

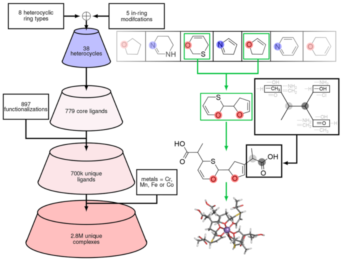
We can construct a large design space of feasible candidates and screen them with NNs

Janet, J.P., *et al.*, *ACS Cent. Sci.*, 6(4):513–524, 2020.

Multiobjective optimization

We can construct a large design space of feasible candidates and screen them with NNs

design space

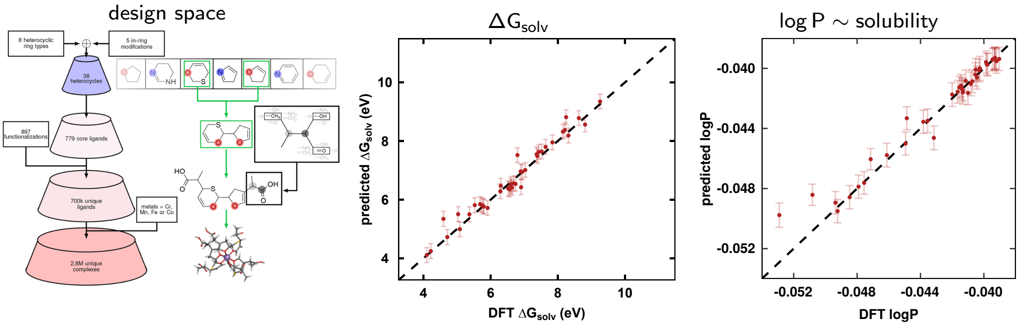


Combinatorial library of 3M RFBs

Janet, J.P., *et al.*, *ACS Cent. Sci.*, 6(4):513–524, 2020.

Multiobjective optimization

We can construct a large design space of feasible candidates and screen them with NNs



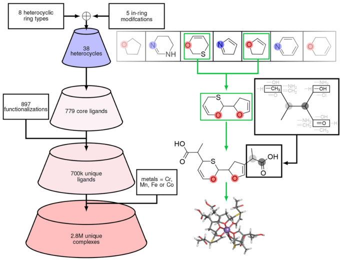
Combinatorial library of 3M RFBs screened in < 4 **mins** on a regular workstation, c.f. 50 **GPU-years** with DFT

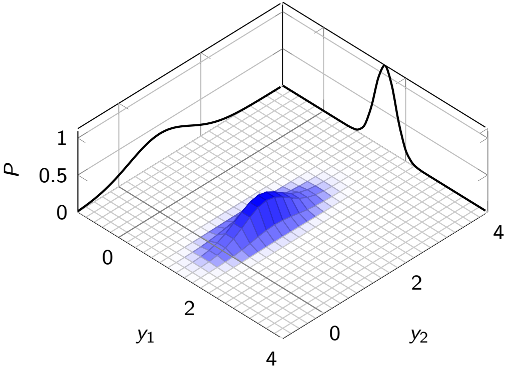
Janet, J.P., *et al.*, *ACS Cent. Sci.*, 6(4):513–524, 2020.

Multiobjective optimization

We can construct a large design space of feasible candidates and screen them with NNs

design space

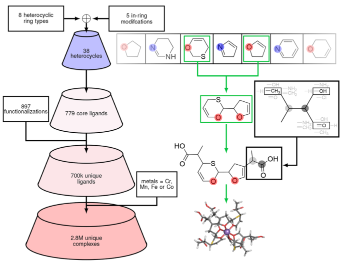


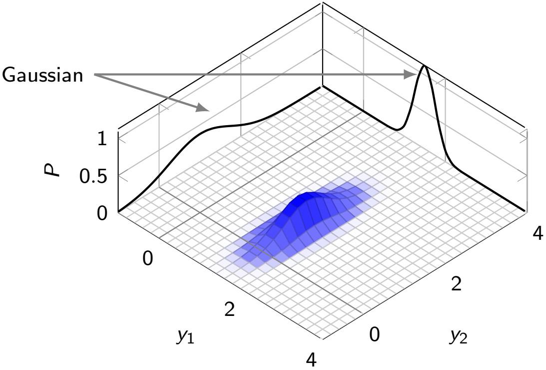


Janet, J.P., *et al.*, *ACS Cent. Sci.*, 6(4):513–524, 2020.

Multiobjective optimization

We can construct a large design space of feasible candidates and screen them with NNs



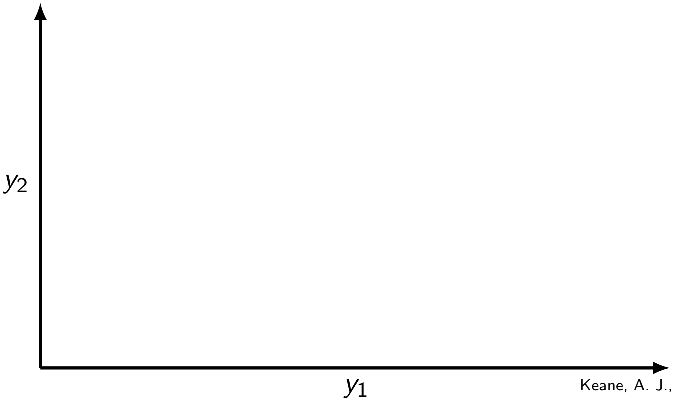


Janet, J.P., *et al.*, *ACS Cent. Sci.*, 6(4):513–524, 2020.

$$\begin{matrix}\Delta G_{\text{solv}} \\ \log P\end{matrix} = \begin{bmatrix}\hat{y}_1 \\ \hat{y}_2\end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix}\hat{\mu}_1 \\ \hat{\mu}_2\end{bmatrix}, \begin{bmatrix}\hat{\sigma}_1^2 & 0 \\ 0 & \hat{\sigma}_2^2\end{bmatrix}\right)$$

2D EGO Illustration

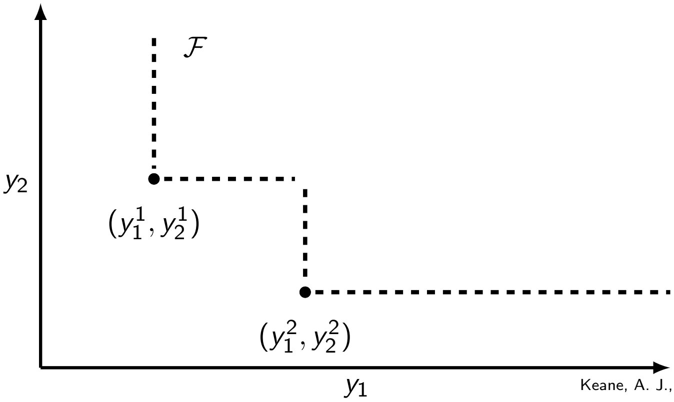
We will use a multiobjective expected improvement framework:



Keane, A. J., *AIAA Journal*, 44(4):879–891, 2006.

2D EGO Illustration

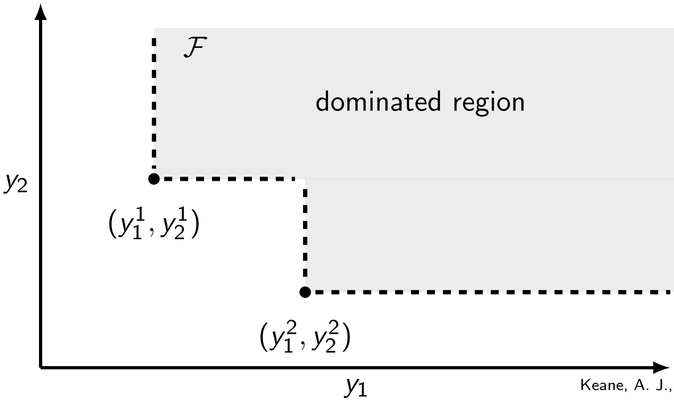
We will use a multiobjective expected improvement framework:



Keane, A. J., *AIAA Journal*, 44(4):879–891, 2006.

2D EGO Illustration

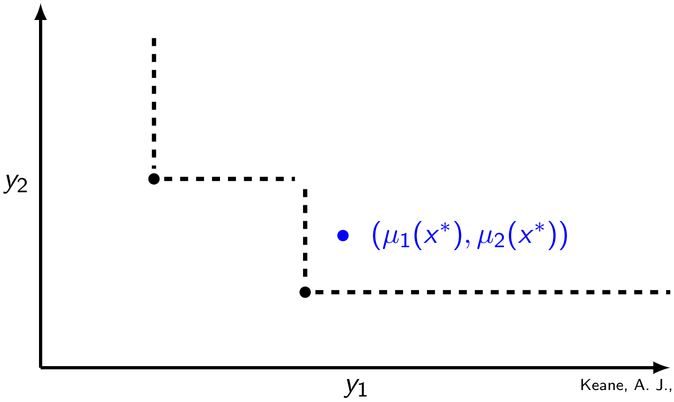
We will use a multiobjective expected improvement framework:



Keane, A. J., *AIAA Journal*, 44(4):879–891, 2006.

2D EGO Illustration

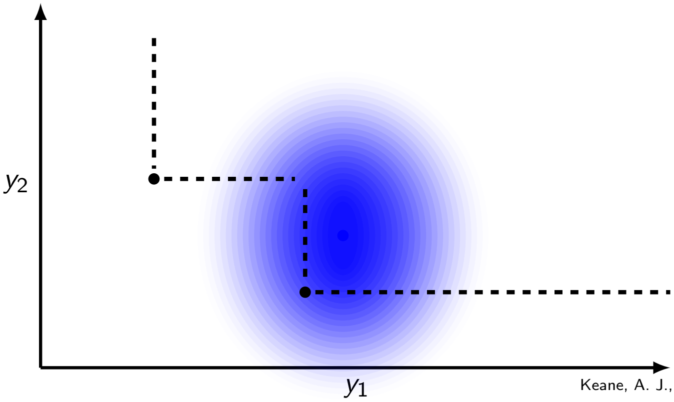
We will use a multiobjective expected improvement framework:



Keane, A. J., *AIAA Journal*, 44(4):879–891, 2006.

2D EGO Illustration

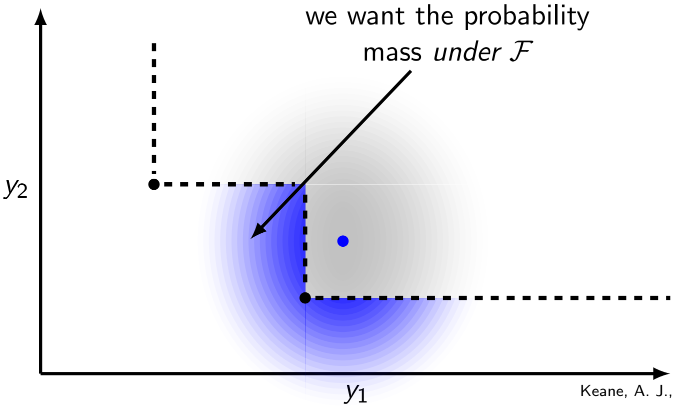
We will use a multiobjective expected improvement framework:



Keane, A. J., *AIAA Journal*, 44(4):879–891, 2006.

2D EGO Illustration

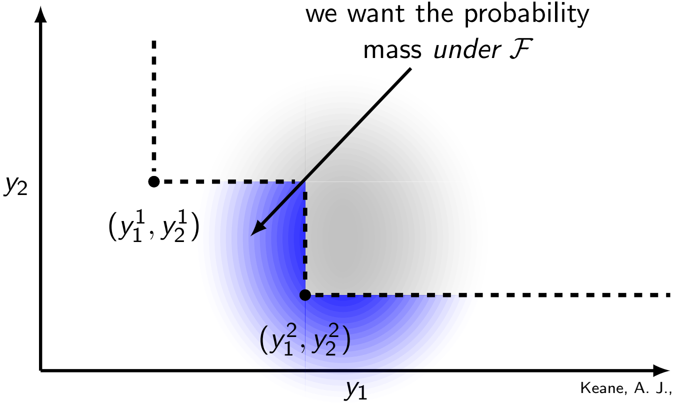
We will use a multiobjective expected improvement framework:



Keane, A. J., *AIAA Journal*, 44(4):879–891, 2006.

2D EGO Illustration

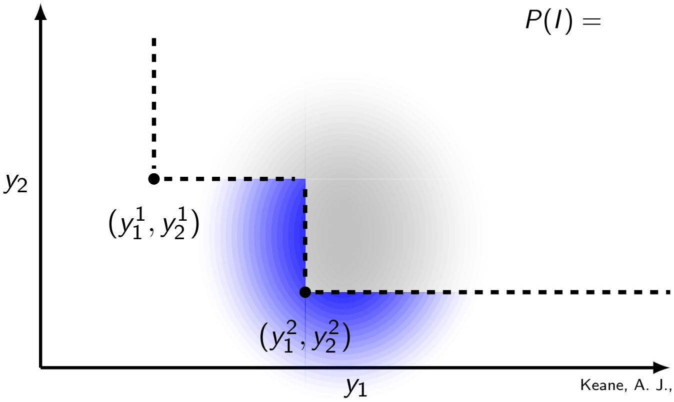
We will use a multiobjective expected improvement framework:



Keane, A. J., *AIAA Journal*, 44(4):879–891, 2006.

2D EGO Illustration

We will use a multiobjective expected improvement framework:



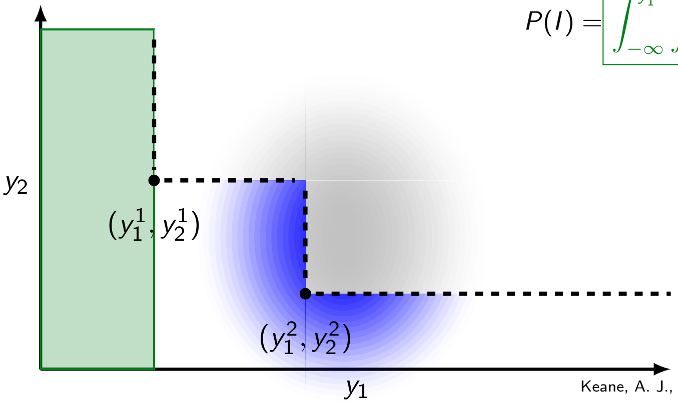
$P(I) =$

Keane, A. J., *AIAA Journal*, 44(4):879–891, 2006.

2D EGO Illustration

We will use a multiobjective expected improvement framework:

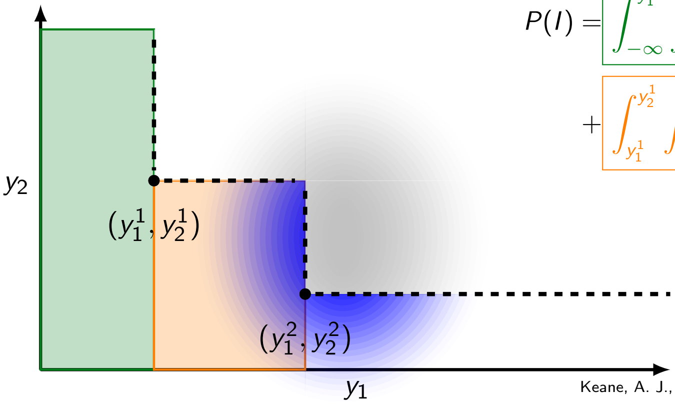
$$P(I) = \boxed{\int_{-\infty}^{y_1^1} \int_{-\infty}^{\infty} p(\hat{y}_1(x^*), \hat{y}_2(x^*)) dy_1 dy_2}$$



Keane, A. J., *AIAA Journal*, 44(4):879–891, 2006.

2D EGO Illustration

We will use a multiobjective expected improvement framework:

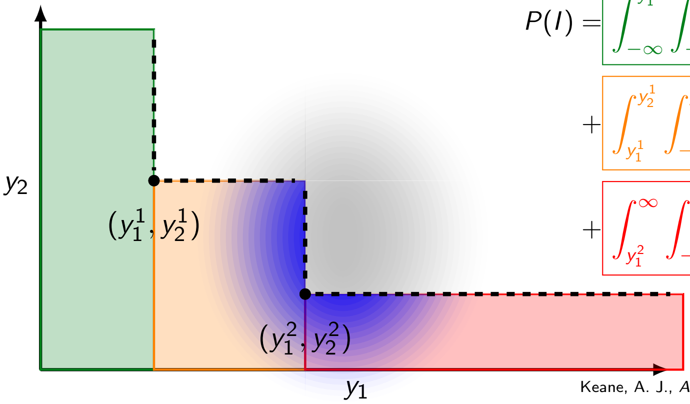


$$P(I) = \boxed{\int_{-\infty}^{y_1^1} \int_{-\infty}^{\infty} p(\hat{y}_1(x^*), \hat{y}_2(x^*)) dy_1 dy_2} + \boxed{\int_{y_1^1}^{y_2^1} \int_{-\infty}^{y_2^1} p(\hat{y}_1(x^*), \hat{y}_2(x^*)) dy_1 dy_2}$$

Keane, A. J., *AIAA Journal*, 44(4):879–891, 2006.

2D EGO Illustration

We will use a multiobjective expected improvement framework:



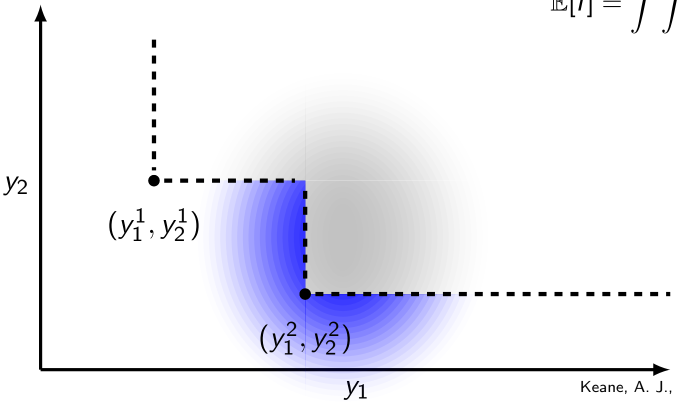
$$P(I) = \boxed{\int_{-\infty}^{y_1^1} \int_{-\infty}^{\infty} p(\hat{y}_1(x^*), \hat{y}_2(x^*)) dy_1 dy_2} + \boxed{\int_{y_1^1}^{y_2^1} \int_{-\infty}^{y_2^1} p(\hat{y}_1(x^*), \hat{y}_2(x^*)) dy_1 dy_2} + \boxed{\int_{y_1^2}^{\infty} \int_{-\infty}^{y_2^2} p(\hat{y}_1(x^*), \hat{y}_2(x^*)) dy_1 dy_2}$$

Keane, A. J., *AIAA Journal*, 44(4):879–891, 2006.

2D EGO Illustration

We will use a multiobjective expected improvement framework:

$$\mathbb{E}[I] = \int\int d(\hat{y}, \mathcal{F}) P[I(x^*)] d\hat{y}_1 d\hat{y}_2$$

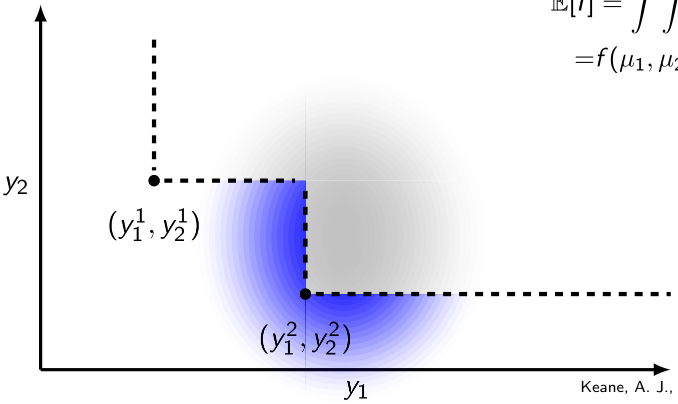


Keane, A. J., *AIAA Journal*, 44(4):879–891, 2006.

2D EGO Illustration

We will use a multiobjective expected improvement framework:

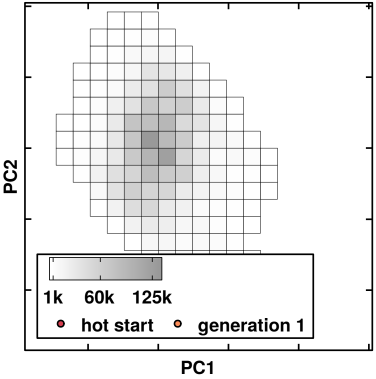
$$\mathbb{E}[I] = \int\int d(\hat{y}, \mathcal{F}) P[I(x^*)] d\hat{y}_1 d\hat{y}_2$$
$$= f(\mu_1, \mu_2, \sigma_1, \sigma_2, \mathcal{F})$$



Keane, A. J., *AIAA Journal*, 44(4):879–891, 2006.

Design space and clustering

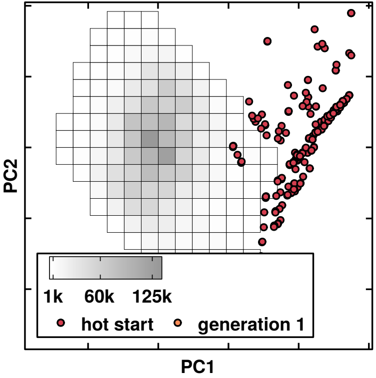
Jump start the design with diversity-oriented cluster:



Janet, J.P., *et al.*, *ACS Cent. Sci.*, 6(4):513–524, 2020.

Design space and clustering

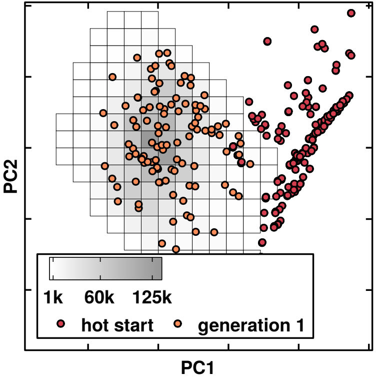
Jump start the design with diversity-oriented cluster:



Janet, J.P., *et al.*, *ACS Cent. Sci.*, 6(4):513–524, 2020.

Design space and clustering

Jump start the design with diversity-oriented cluster:

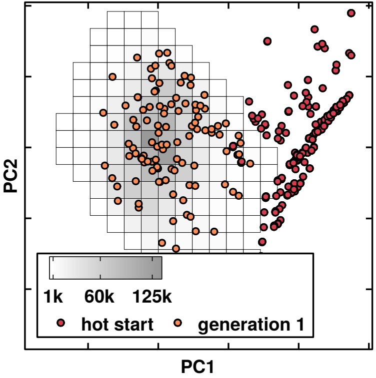


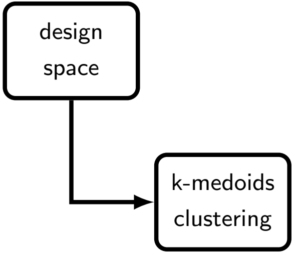
design space

Janet, J.P., *et al.*, *ACS Cent. Sci.*, 6(4):513–524, 2020.

Design space and clustering

Jump start the design with diversity-oriented cluster:

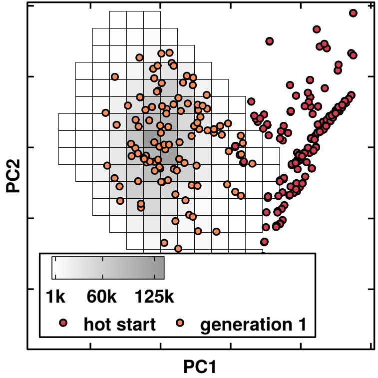


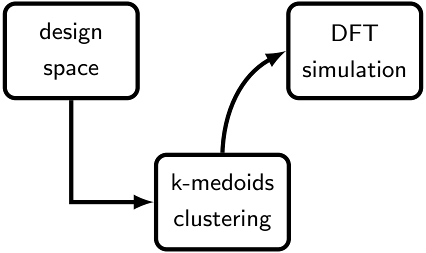


Janet, J.P., *et al.*, *ACS Cent. Sci.*, 6(4):513–524, 2020.

Design space and clustering

Jump start the design with diversity-oriented cluster:

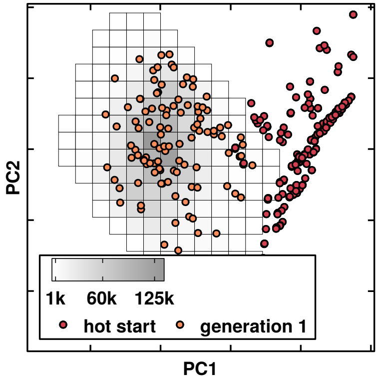


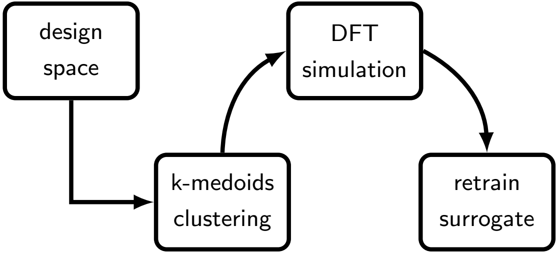


Janet, J.P., *et al.*, *ACS Cent. Sci.*, 6(4):513–524, 2020.

Design space and clustering

Jump start the design with diversity-oriented cluster:

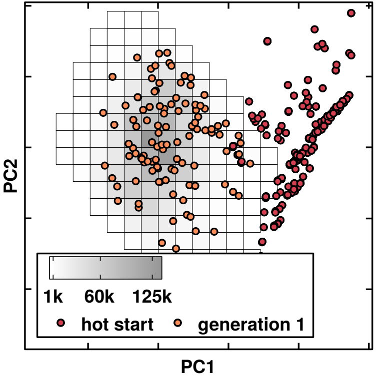


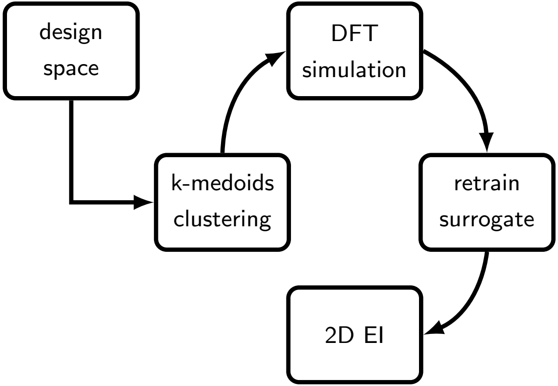


Janet, J.P., *et al.*, *ACS Cent. Sci.*, 6(4):513–524, 2020.

Design space and clustering

Jump start the design with diversity-oriented cluster:

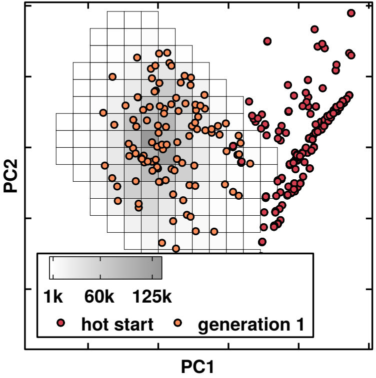


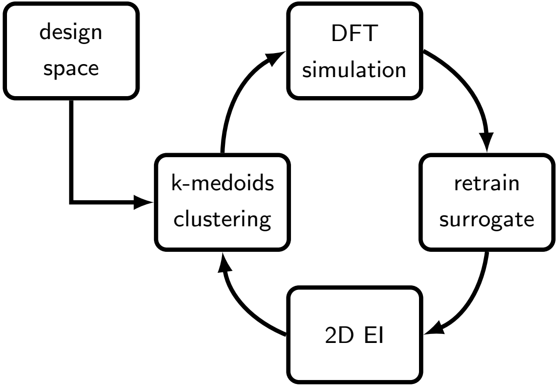


Janet, J.P., *et al.*, *ACS Cent. Sci.*, 6(4):513–524, 2020.

Design space and clustering

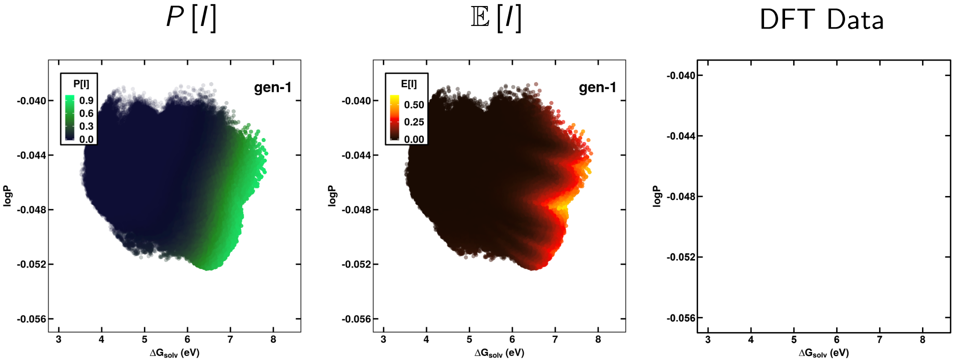
Jump start the design with diversity-oriented cluster:





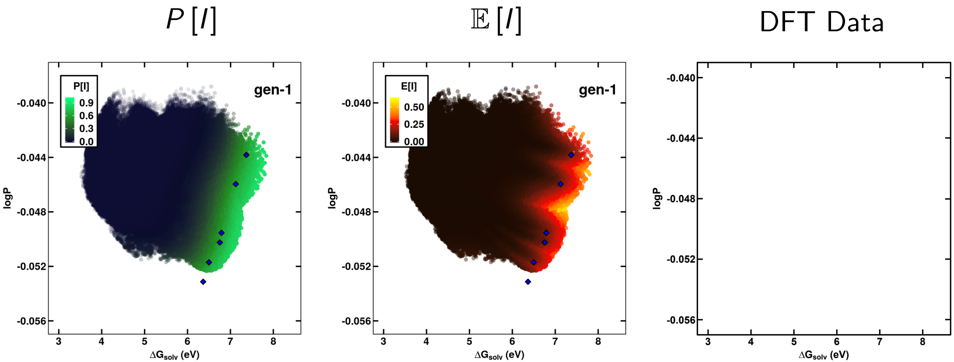
Janet, J.P., *et al.*, *ACS Cent. Sci.*, 6(4):513–524, 2020.

Evolution of PI and EI



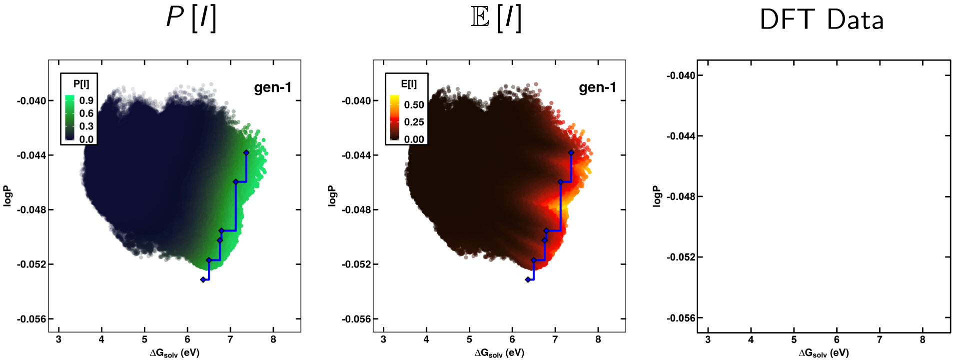
Janet, J.P., *et al.*, *ACS Cent. Sci.*, 6(4):513–524, 2020.

Evolution of PI and EI



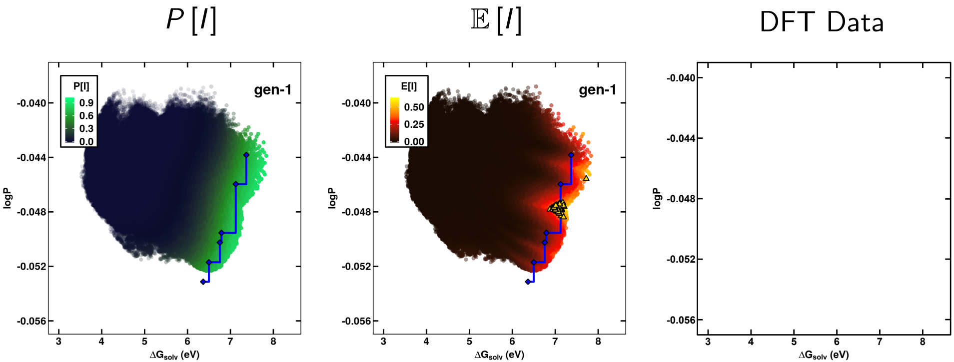
Janet, J.P., *et al.*, *ACS Cent. Sci.*, 6(4):513–524, 2020.

Evolution of PI and EI



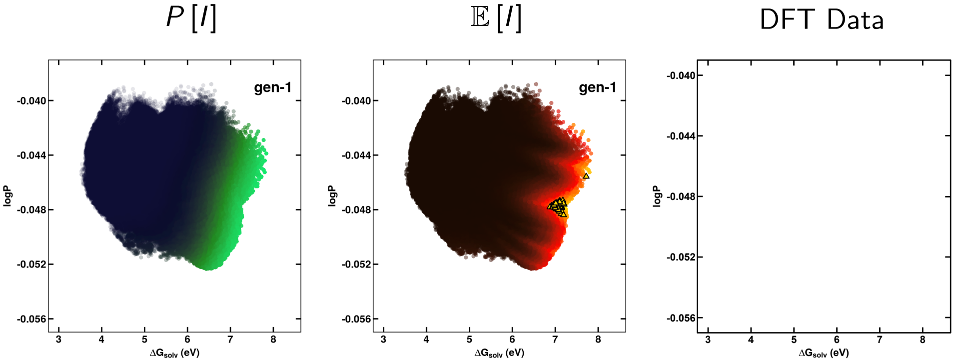
Janet, J.P., *et al.*, *ACS Cent. Sci.*, 6(4):513–524, 2020.

Evolution of PI and EI



Janet, J.P., *et al.*, *ACS Cent. Sci.*, 6(4):513–524, 2020.

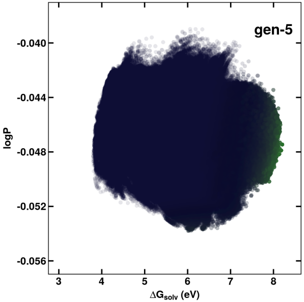
Evolution of PI and EI



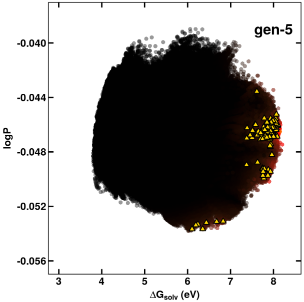
Janet, J.P., *et al.*, *ACS Cent. Sci.*, 6(4):513–524, 2020.

Evolution of PI and EI

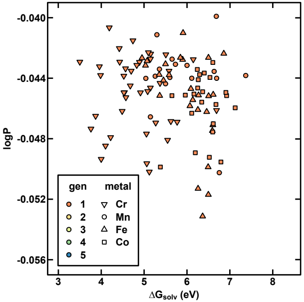
$P[I]$



$\mathbb{E}[I]$

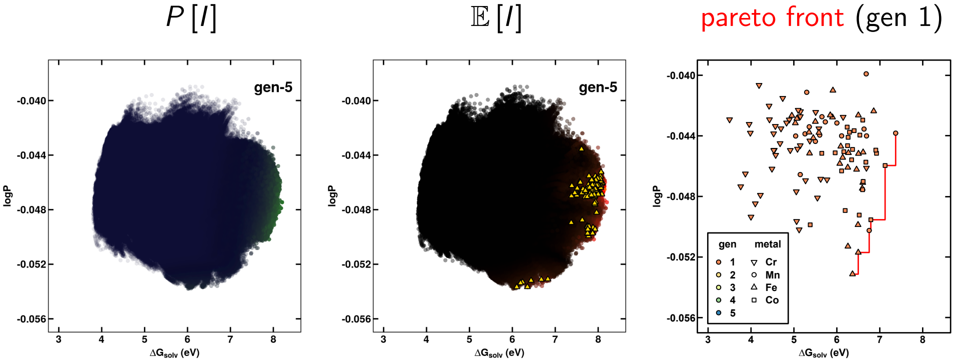


k-medoids points (gen 1)



Janet, J.P., *et al.*, *ACS Cent. Sci.*, 6(4):513–524, 2020.

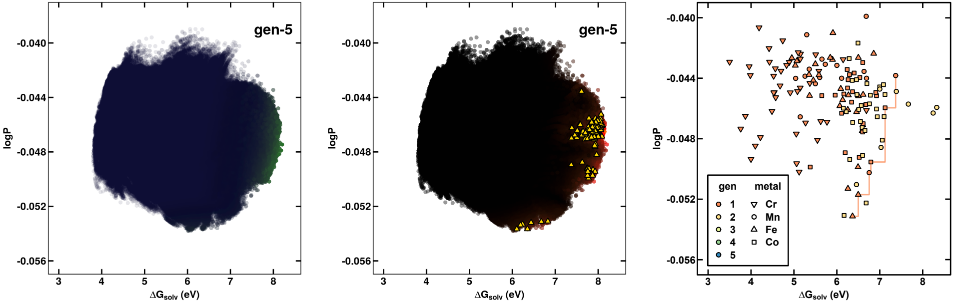
Evolution of PI and EI



Janet, J.P., *et al.*, *ACS Cent. Sci.*, 6(4):513–524, 2020.

Evolution of PI and EI

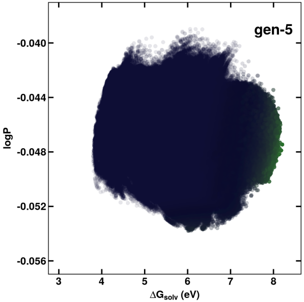
$P[I]$ | $\mathbb{E}[I]$ | EI points (geneneration 2)



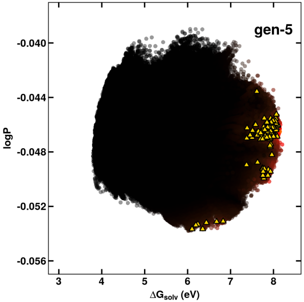
Janet, J.P., *et al.*, *ACS Cent. Sci.*, 6(4):513–524, 2020.

Evolution of PI and EI

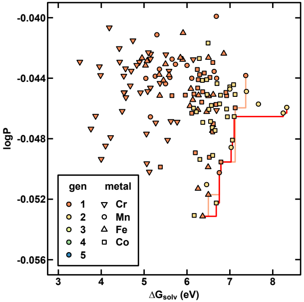
$P\,[I]$



$\mathbb{E}\,[I]$

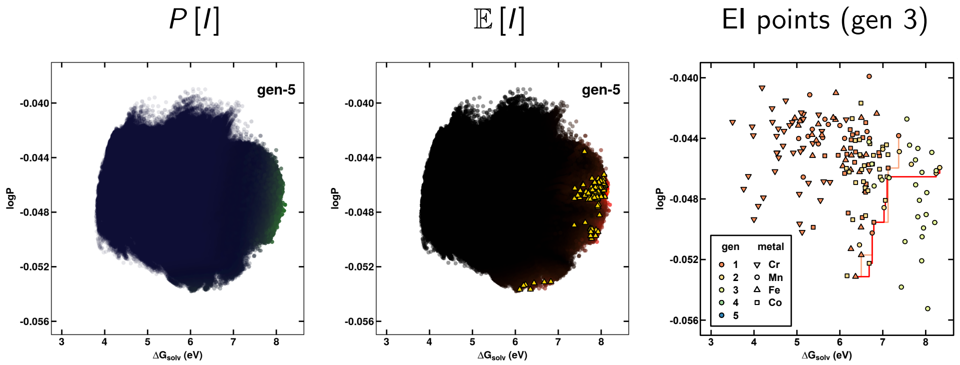


pareto front (gen 2)



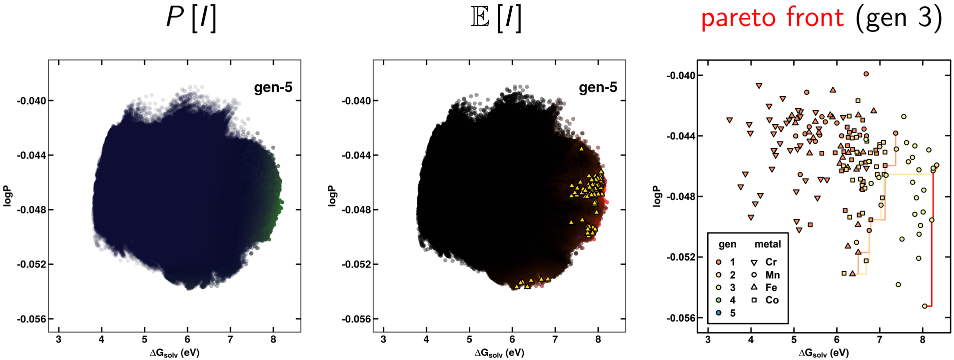
Janet, J.P., *et al.*, *ACS Cent. Sci.*, 6(4):513–524, 2020.

Evolution of PI and EI



Janet, J.P., *et al.*, *ACS Cent. Sci.*, 6(4):513–524, 2020.

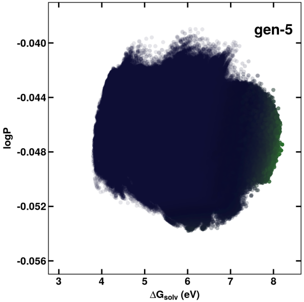
Evolution of PI and EI



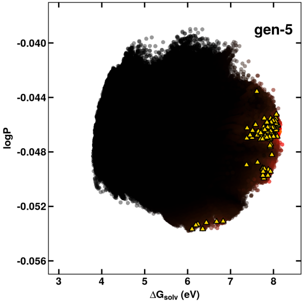
Janet, J.P., *et al.*, *ACS Cent. Sci.*, 6(4):513–524, 2020.

Evolution of PI and EI

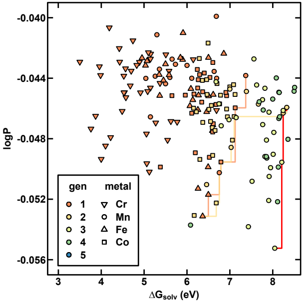
$P[I]$



$\mathbb{E}[I]$



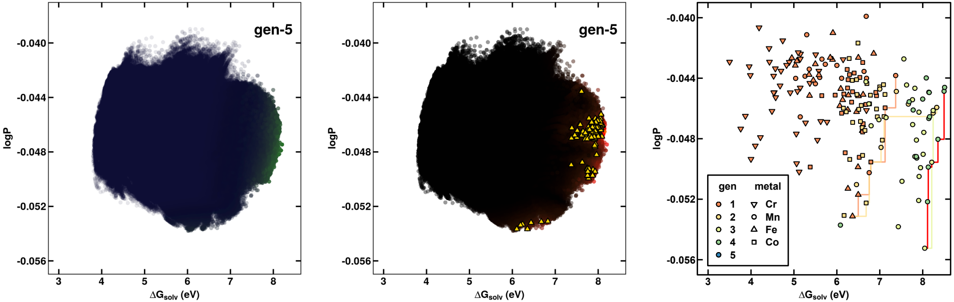
EI points (gen 4)



Janet, J.P., *et al.*, *ACS Cent. Sci.*, 6(4):513–524, 2020.

Evolution of PI and EI

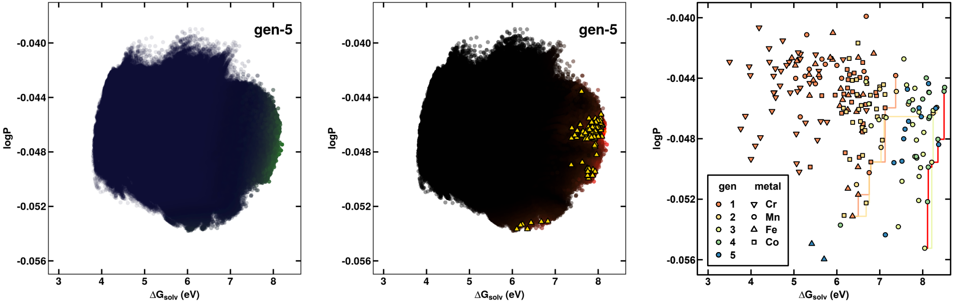
$P\,[I]$ $\mathbb{E}\,[I]$ pareto front (gen 4)



Janet, J.P., *et al.*, *ACS Cent. Sci.*, 6(4):513–524, 2020.

Evolution of PI and EI

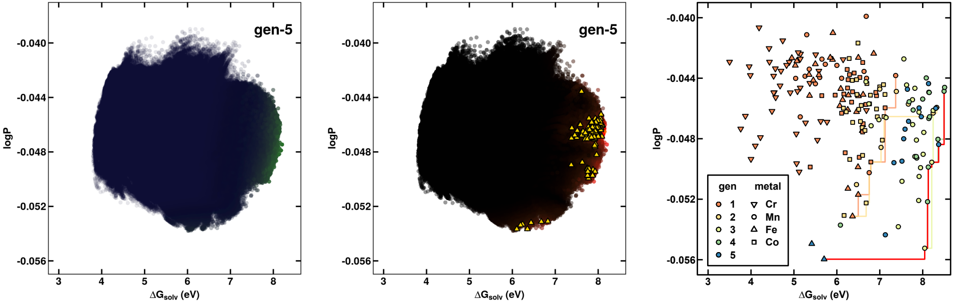
$P[I]$ $\mathbb{E}[I]$ EI points (gen 5)



Janet, J.P., *et al.*, *ACS Cent. Sci.*, 6(4):513–524, 2020.

Evolution of PI and EI

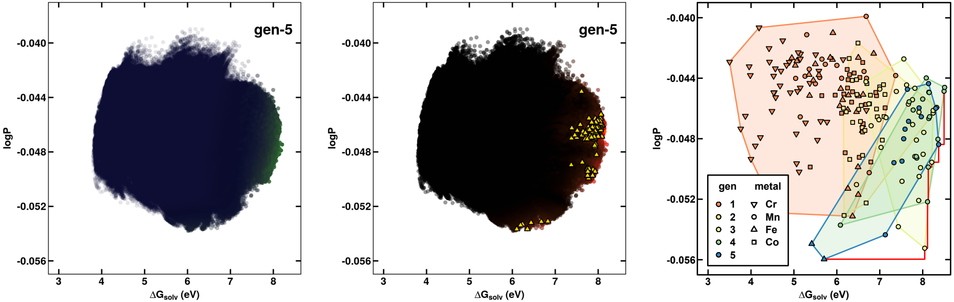
$P\,[I]$ $\mathbb{E}\,[I]$ pareto front (gen 5)



Janet, J.P., *et al.*, *ACS Cent. Sci.*, 6(4):513–524, 2020.

Evolution of PI and EI

$P[I]$ $\mathbb{E}[I]$ convex hulls



Janet, J.P., *et al.*, *ACS Cent. Sci.*, 6(4):513–524, 2020.

Case study conclusions

Case study conclusions

- EI framework provides high resolution in the region of interest, converges quickly
- Able to identify fruitful regions from a large chemical space based few DFT evaluations
- Multiobjective DFT optimization guided by ML efficiency generates lead complexes, $> 100\times$ faster than random search

Case study conclusions

- EI framework provides high resolution in the region of interest, converges quickly
- Able to identify fruitful regions from a large chemical space based few DFT evaluations
- Multiobjective DFT optimization guided by ML efficiency generates lead complexes, $> 100\times$ faster than random search

Thanks to Professor Heather Kulik and her lab at MIT

Case study conclusions

- EI framework provides high resolution in the region of interest, converges quickly
- Able to identify fruitful regions from a large chemical space based few DFT evaluations
- Multiobjective DFT optimization guided by ML efficiency generates lead complexes, $> 100\times$ faster than random search

Thanks to Professor Heather Kulik and her lab at MIT

github.com/hjkgrp/molSimplify
github.com/hjkgrp/AutomaticDesign

Drug discovery

And now for something completely different

Drug discovery

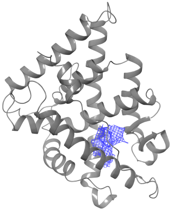
And now for something ~~completely~~ moderately different

Drug discovery

Bringing a new drug costs upwards of \$1 billion and can take decades.

Drug discovery

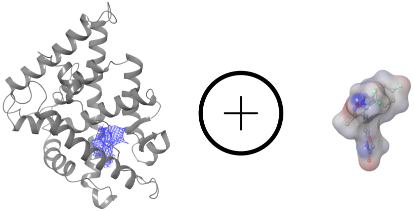
Bringing a new drug costs upwards of \$1 billion and can take decades.



protein

Drug discovery

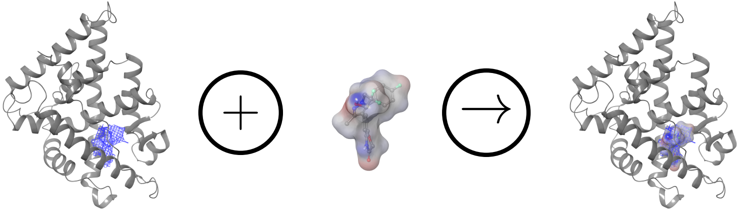
Bringing a new drug costs upwards of $1 billion and can take decades.



protein ligand

Drug discovery

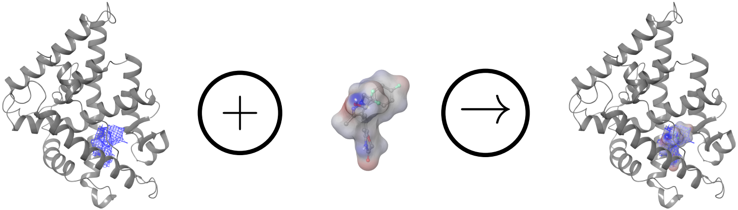
Bringing a new drug costs upwards of \$1 billion and can take decades.



protein ligand complex

Drug discovery

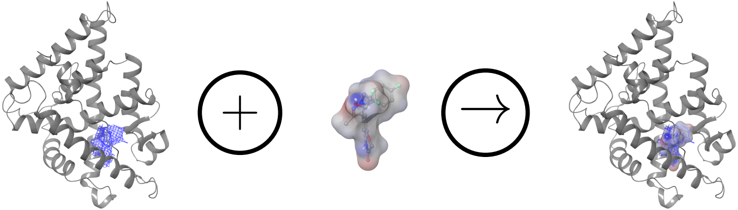
Bringing a new drug costs upwards of \$1 billion and can take decades.



This is very difficult because of the many requirements for drug-like molecules:

Drug discovery

Bringing a new drug costs upwards of $1 billion and can take decades.

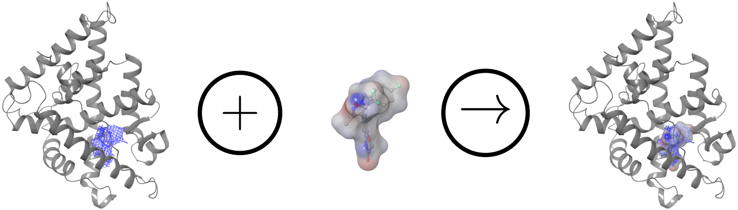


This is very difficult because of the many requirements for drug-like molecules:

- Affinity/potency (hard)

Drug discovery

Bringing a new drug costs upwards of $1 billion and can take decades.

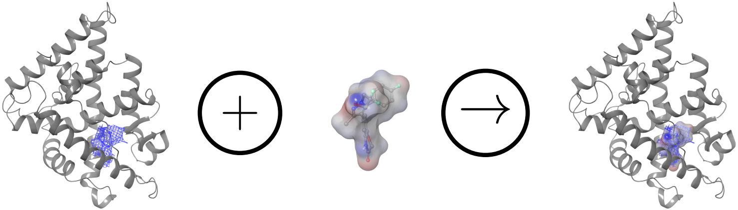


This is very difficult because of the many requirements for drug-like molecules:

- Affinity/potency (hard)
- Specificity, tolerability, toxicity (very hard)

Drug discovery

Bringing a new drug costs upwards of $1 billion and can take decades.

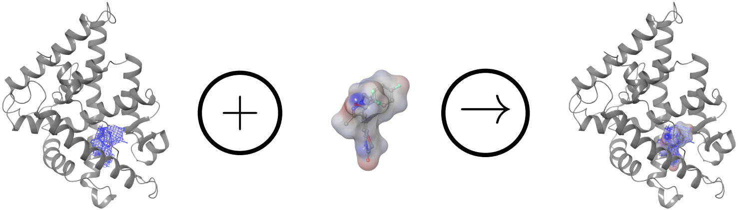


This is very difficult because of the many requirements for drug-like molecules:

- Affinity/potency (hard)
- Specificity, tolerability, toxicity (very hard)
- Pharmacokinetics and pharmacodynamics (hard)

Drug discovery

Bringing a new drug costs upwards of $1 billion and can take decades.

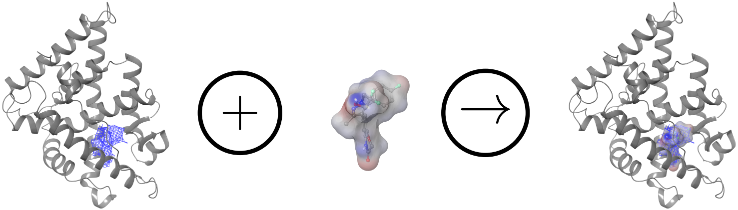


This is very difficult because of the many requirements for drug-like molecules:

- Affinity/potency (hard)
- Specificity, tolerability, toxicity (very hard)
- Pharmacokinetics and pharmacodynamics (hard)
- Synthetic accessibility (moderate)

Drug discovery

Bringing a new drug costs upwards of $1 billion and can take decades.

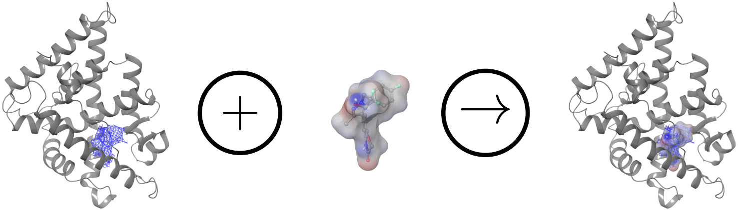


This is very difficult because of the many requirements for drug-like molecules:

- Affinity/potency (hard) – can be predicted with physics (FEP), sometimes
- Specificity, tolerability, toxicity (very hard)
- Pharmacokinetics and pharmacodynamics (hard)
- Synthetic accessibility (moderate)

Drug discovery

Bringing a new drug costs upwards of $1 billion and can take decades.

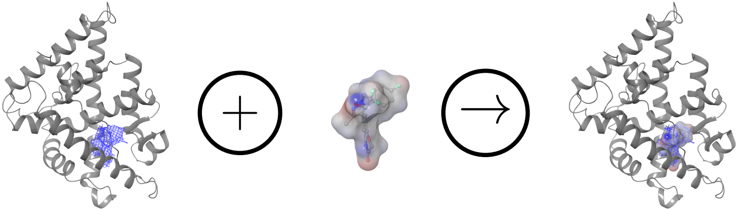


This is very difficult because of the many requirements for drug-like molecules:

- Affinity/potency (hard) – can be predicted with physics (FEP), sometimes
- Specificity, tolerability, toxicity (very hard) – mostly predicted from data
- Pharmacokinetics and pharmacodynamics (hard)
- Synthetic accessibility (moderate)

Drug discovery

Bringing a new drug costs upwards of $1 billion and can take decades.

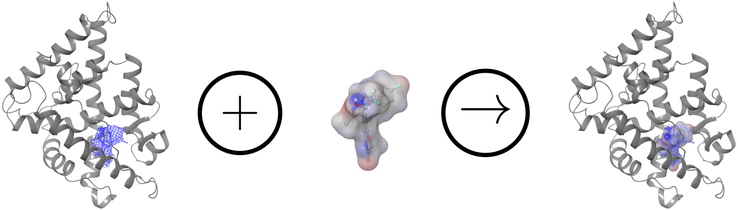


This is very difficult because of the many requirements for drug-like molecules:

- Affinity/potency (hard) – can be predicted with physics (FEP), sometimes
- Specificity, tolerability, toxicity (very hard) – mostly predicted from data
- Pharmacokinetics and pharmacodynamics (hard) – mostly predicted from data
- Synthetic accessibility (moderate)

Drug discovery

Bringing a new drug costs upwards of $1 billion and can take decades.



This is very difficult because of the many requirements for drug-like molecules:

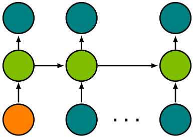
- Affinity/potency (hard) – can be predicted with physics (FEP), sometimes
- Specificity, tolerability, toxicity (very hard) – mostly predicted from data
- Pharmacokinetics and pharmacodynamics (hard) – mostly predicted from data
- Synthetic accessibility (moderate) – ML synthesis planning very advanced

The REINVENT ecosystem

We tackle these challenges with generative reinforcement learning

The REINVENT ecosystem

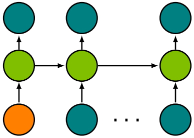
We tackle these challenges with generative reinforcement learning



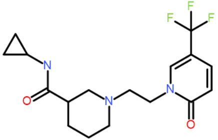
text-based molecular representation generated by an RNN

The REINVENT ecosystem

We tackle these challenges with generative reinforcement learning



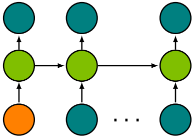
text-based molecular representation generated by an RNN



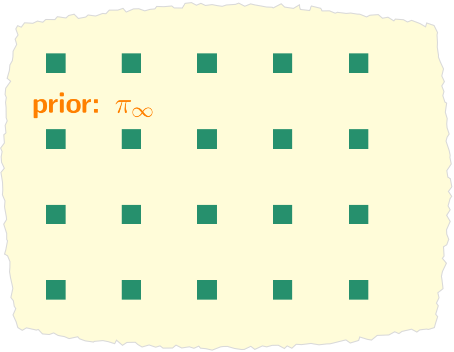
Olivecrona, M., *et al.*, *J. Cheminform.* 9(48), 2017.
Blaschke, T., *et al.*,*J. Chem. Inf. Model.* 60(12):5918–5922, 2020.

The REINVENT ecosystem

We tackle these challenges with generative reinforcement learning



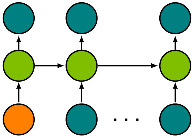
text-based molecular representation generated by an RNN

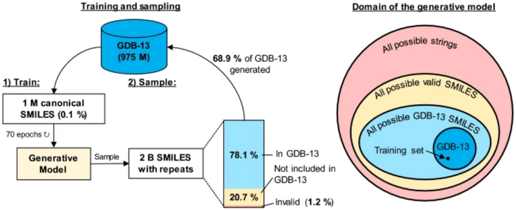


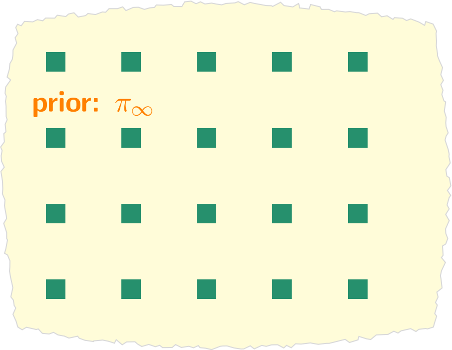
Arús-Pous, J., *et al.*, *J. Cheminform.* 11(71), 2019.
Olivecrona, M., *et al.*, *J. Cheminform.* 9(48), 2017.
Blaschke, T., *et al.*,*J. Chem. Inf. Model.* 60(12):5918–5922, 2020.

The REINVENT ecosystem

We tackle these challenges with generative reinforcement learning



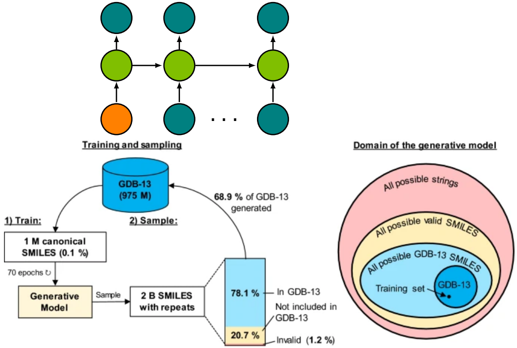


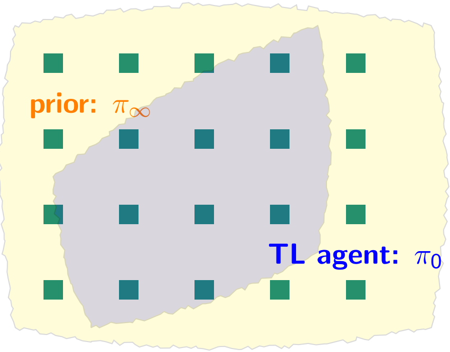


Arús-Pous, J., *et al.*, *J. Cheminform.* 11(20), 2019.
Arús-Pous, J., *et al.*, *J. Cheminform.* 11(71), 2019.
Olivecrona, M., *et al.*, *J. Cheminform.* 9(48), 2017.
Blaschke, T., *et al.*,*J. Chem. Inf. Model.* 60(12):5918–5922, 2020.

The REINVENT ecosystem

We tackle these challenges with generative reinforcement learning

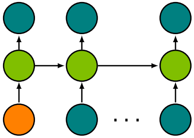




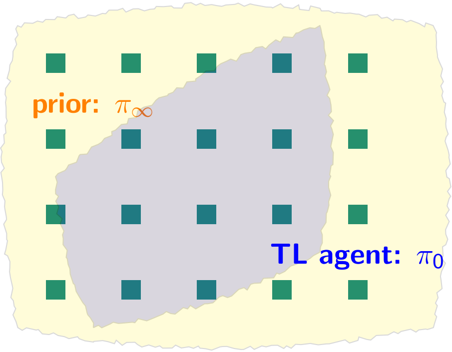
Arús-Pous, J., *et al.*, *J. Cheminform.* 11(20), 2019.
Arús-Pous, J., *et al.*, *J. Cheminform.* 11(71), 2019.
Olivecrona, M., *et al.*, *J. Cheminform.* 9(48), 2017.
Blaschke, T., *et al.*,*J. Chem. Inf. Model.* 60(12):5918–5922, 2020.

The REINVENT ecosystem

We tackle these challenges with generative reinforcement learning



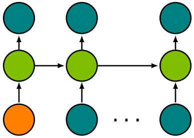
$$R(a) = \sigma\, S(a) \; - \; \log \pi_\infty (a)$$



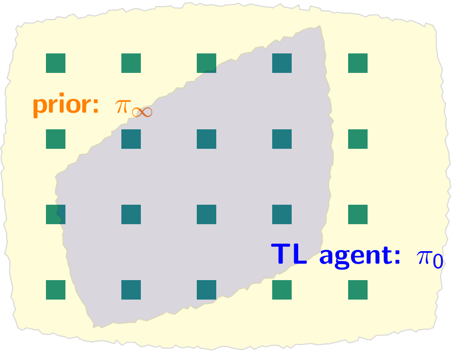
Arús-Pous, J., *et al.*, *J. Cheminform.* 11(20), 2019.
Arús-Pous, J., *et al.*, *J. Cheminform.* 11(71), 2019.
Olivecrona, M., *et al.*, *J. Cheminform.* 9(48), 2017.
Blaschke, T., *et al.*,*J. Chem. Inf. Model.* 60(12):5918–5922, 2020.

The REINVENT ecosystem

We tackle these challenges with generative reinforcement learning



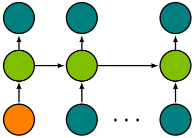
$$R(a) = \sigma \underbrace{S(a)}_{actor} - \underbrace{\log \pi_\infty(a)}_{critic}$$



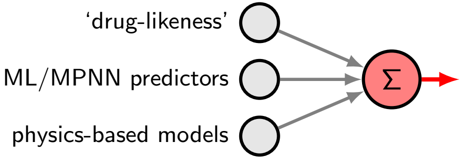
Arús-Pous, J., *et al.*, *J. Cheminform.* 11(20), 2019.
Arús-Pous, J., *et al.*, *J. Cheminform.* 11(71), 2019.
Olivecrona, M., *et al.*, *J. Cheminform.* 9(48), 2017.
Blaschke, T., *et al.*,*J. Chem. Inf. Model.* 60(12):5918–5922, 2020.

The REINVENT ecosystem

We tackle these challenges with generative reinforcement learning



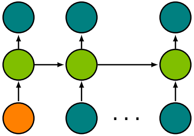
$$R(a) = \sigma\, S\,(a) \; - \; \log \pi_\infty\,(a)$$



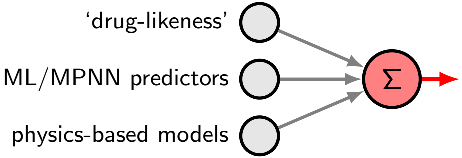
Arús-Pous, J., *et al.*, *J. Cheminform.* 11(20), 2019.
Arús-Pous, J., *et al.*, *J. Cheminform.* 11(71), 2019.
Olivecrona, M., *et al.*, *J. Cheminform.* 9(48), 2017.
Blaschke, T., *et al.*,*J. Chem. Inf. Model.* 60(12):5918–5922, 2020.

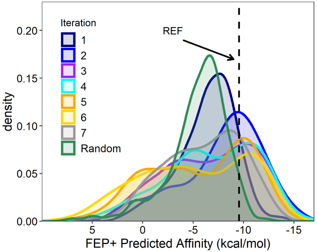
The REINVENT ecosystem

We tackle these challenges with generative reinforcement learning



$$R(a) = \sigma\, S(a) \;-\; \log \pi_\infty(a)$$

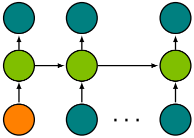




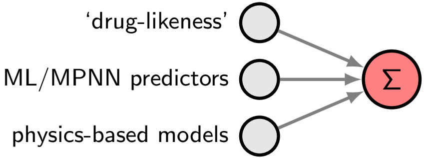
Arús-Pous, J., *et al.*, *J. Cheminform.* 11(20), 2019.
Arús-Pous, J., *et al.*, *J. Cheminform.* 11(71), 2019.
Olivecrona, M., *et al.*, *J. Cheminform.* 9(48), 2017.
Blaschke, T., *et al.*,*J. Chem. Inf. Model.* 60(12):5918–5922, 2020.

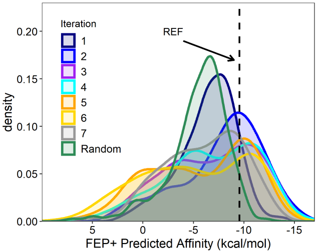
The REINVENT ecosystem

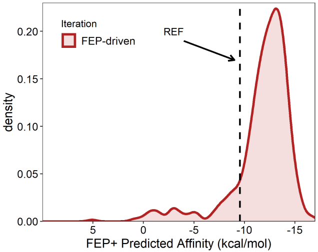
We tackle these challenges with generative reinforcement learning



$$R(a) = \sigma\, {\color{red} S\,(a)} \;-\; \log \pi_\infty\,(a)$$



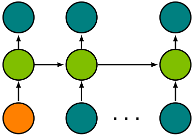




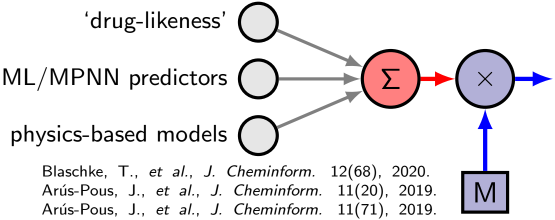
Arús-Pous, J., *et al.*, *J. Cheminform.* 11(20), 2019.
Arús-Pous, J., *et al.*, *J. Cheminform.* 11(71), 2019.
Olivecrona, M., *et al.*, *J. Cheminform.* 9(48), 2017.
Blaschke, T., *et al.*,*J. Chem. Inf. Model.* 60(12):5918–5922, 2020.

The REINVENT ecosystem

We tackle these challenges with generative reinforcement learning



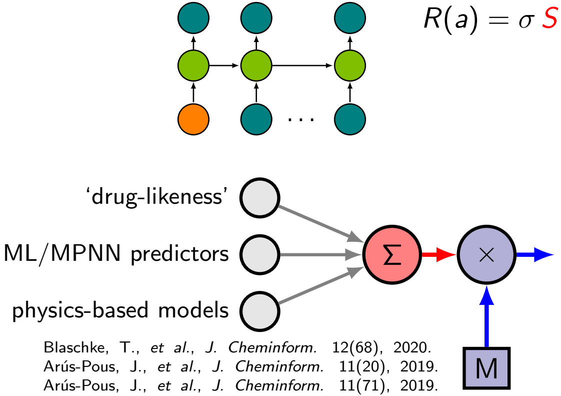
$$R(a) = \sigma\, S(a)\, M(a) - \log \pi_{\infty}(a)$$



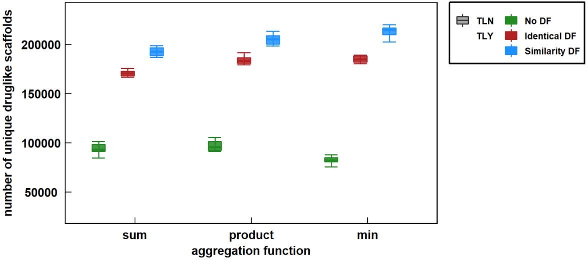
Blaschke, T., *et al.*, *J. Cheminform.* 12(68), 2020.
Arús-Pous, J., *et al.*, *J. Cheminform.* 11(20), 2019.
Arús-Pous, J., *et al.*, *J. Cheminform.* 11(71), 2019.
Olivecrona, M., *et al.*, *J. Cheminform.* 9(48), 2017.
Blaschke, T., *et al.*,*J. Chem. Inf. Model.* 60(12):5918–5922, 2020.

The REINVENT ecosystem

We tackle these challenges with generative reinforcement learning



$$R(a) = \sigma\, S(a)\, M(a) \; - \; \log \pi_\infty (a)$$

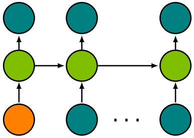


Blaschke, T., *et al.*, *J. Cheminform.* 12(68), 2020.
Arús-Pous, J., *et al.*, *J. Cheminform.* 11(20), 2019.
Arús-Pous, J., *et al.*, *J. Cheminform.* 11(71), 2019.
Olivecrona, M., *et al.*, *J. Cheminform.* 9(48), 2017.
Blaschke, T., *et al.*,*J. Chem. Inf. Model.* 60(12):5918–5922, 2020.

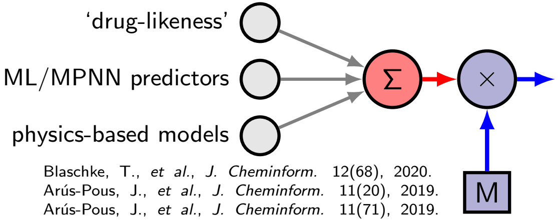
Janet, J.P., *et al.*, *in preparation*.

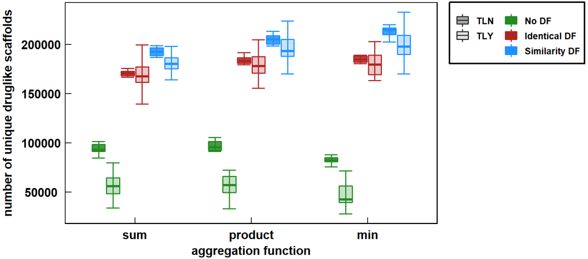
The REINVENT ecosystem

We tackle these challenges with generative reinforcement learning



$$R(a) = \sigma\, S(a)\, M(a) \;-\; \log \pi_\infty(a)$$



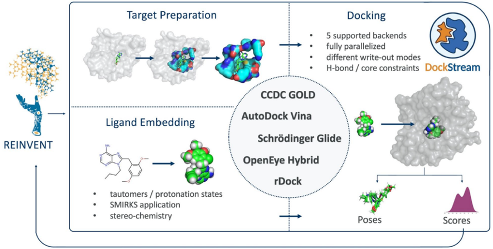


Blaschke, T., *et al.*, *J. Cheminform.* 12(68), 2020.
Arús-Pous, J., *et al.*, *J. Cheminform.* 11(20), 2019.
Arús-Pous, J., *et al.*, *J. Cheminform.* 11(71), 2019.
Olivecrona, M., *et al.*, *J. Cheminform.* 9(48), 2017.
Blaschke, T., *et al.*,*J. Chem. Inf. Model.* 60(12):5918–5922, 2020.

Janet, J.P., *et al.*, *in preparation.*

Docking and curriculum learning

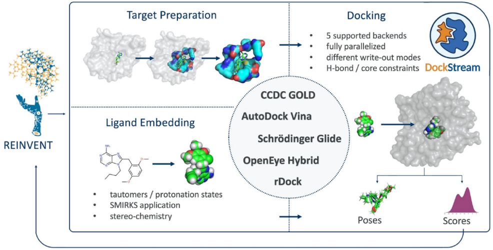
Developed engine for conditioning on protein structure.

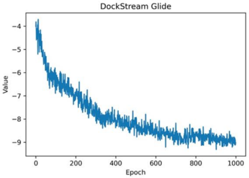


Guo, J., *et al.*, *J. Cheminform.* 13(89), 2021.

Docking and curriculum learning

Developed engine for conditioning on protein structure. Nice, but slow! $\sim$ 8 minutes/epoch.

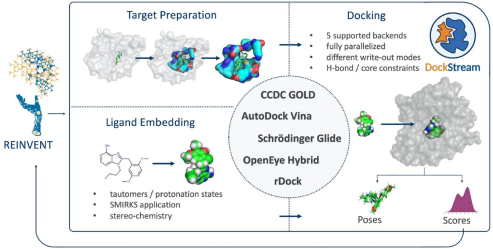


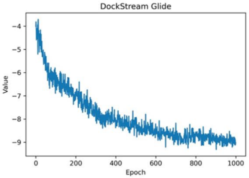


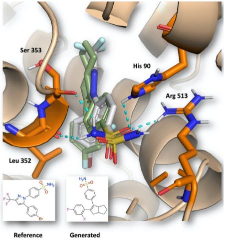
Guo, J., *et al.*, *J. Cheminform.* 13(89), 2021.

Docking and curriculum learning

Developed engine for conditioning on protein structure. Nice, but slow! $\sim$ 8 minutes/epoch.



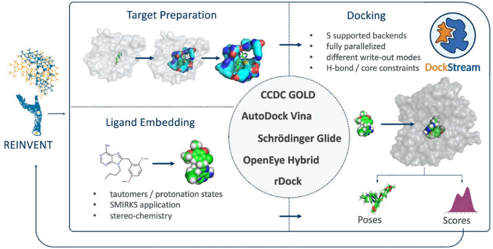


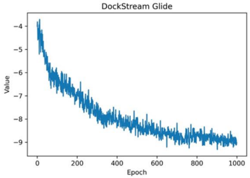


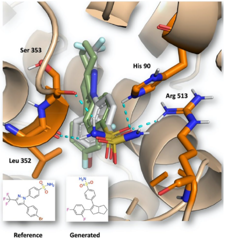
Guo, J., *et al.*, *J. Cheminform.* 13(89), 2021.

Docking and curriculum learning

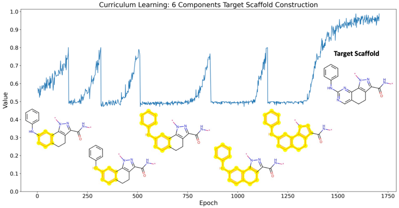
Developed engine for conditioning on protein structure. Nice, but slow! $\sim$ 8 minutes/epoch.







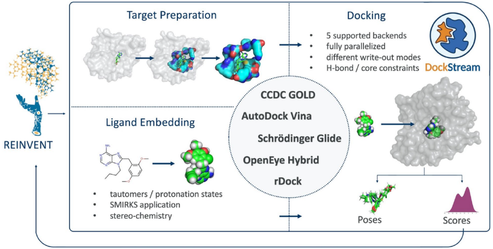
Implemented curriculum learning to speed up the process and learn difficult objectives.
Add physics-based models last!

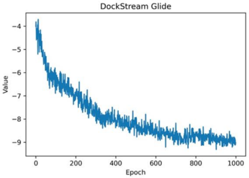


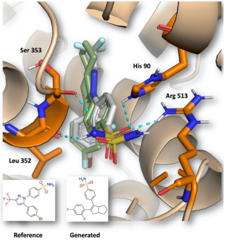
Guo, J., *et al.*, *ChemRxiv*:chemrxiv-2021-37m7w-v2
Guo, J., *et al.*, *J. Cheminform.* 13(89), 2021.

Docking and curriculum learning

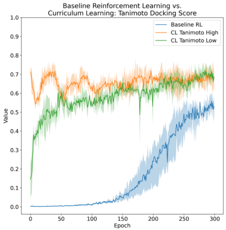
Developed engine for conditioning on protein structure. Nice, but slow! ∼ 8 minutes/epoch.







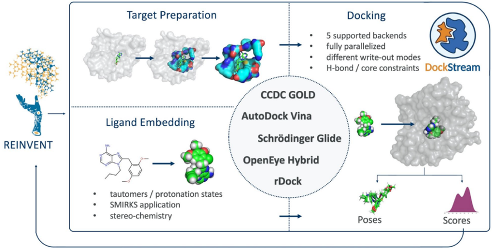
Implemented curriculum learning to speed up the process and learn difficult objectives.
Add physics-based models last!

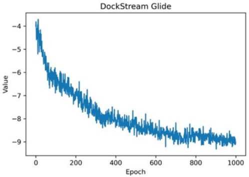


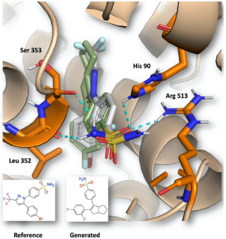
Guo, J., *et al.*, *ChemRxiv*:chemrxiv-2021-37m7w-v2
Guo, J., *et al.*, *J. Cheminform.* 13(89), 2021.

Docking and curriculum learning

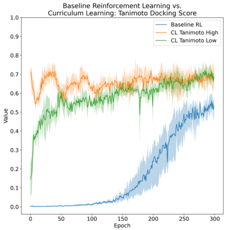
Developed engine for conditioning on protein structure. Nice, but slow! $\sim$ 8 minutes/epoch.

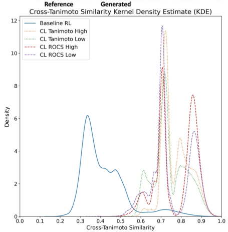






Implemented curriculum learning to speed up the process and learn difficult objectives.
Add physics-based models last!

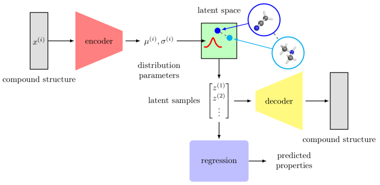




Guo, J., *et al.*, *ChemRxiv*:chemrxiv-2021-37m7w-v2
Guo, J., *et al.*, *J. Cheminform.* 13(89), 2021.

Other Projects

Transfer learning to exploit 'big' affinity data:

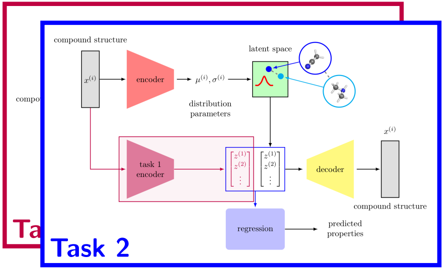


Task 1

Buterez, D., *et al.*, *in preparation.*

Other Projects

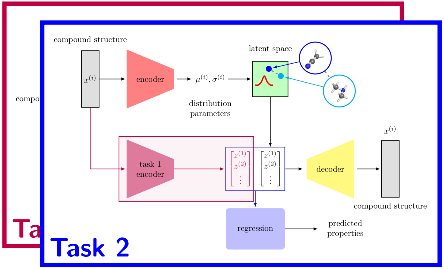
Transfer learning to exploit 'big' affinity data:

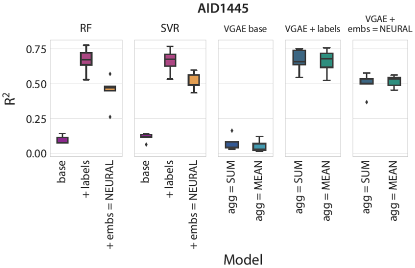


Buterez, D., *et al.*, *in preparation.*

Other Projects

Transfer learning to exploit 'big' affinity data:

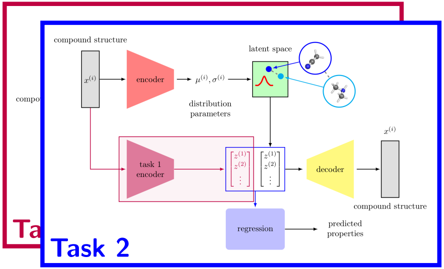




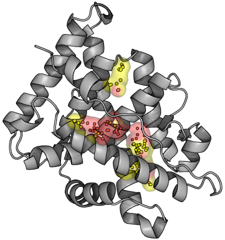
Buterez, D., *et al.*, *in preparation.*

Other Projects

Transfer learning to exploit 'big' affinity data:



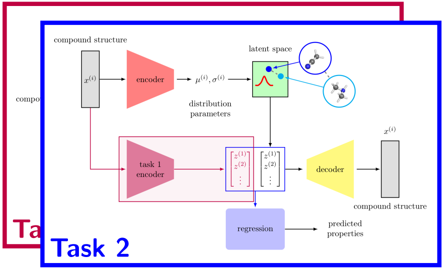
Protein pocket analysis with e3nns:



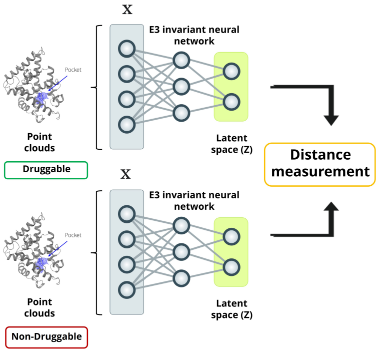
Navarro, A.C., Ghabi, Y., *et al.,in preparation.*
Buterez, D., *et al., in preparation.*

Other Projects

Transfer learning to exploit 'big' affinity data:



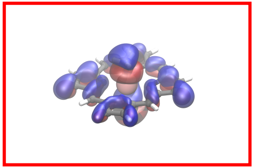
Protein pocket analysis with e3nns:



Navarro, A.C., Ghabi, Y., *et al.,in preparation.*
Buterez, D., *et al., in preparation.*

More than just ML (or simulation)

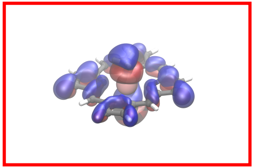
physics-driven



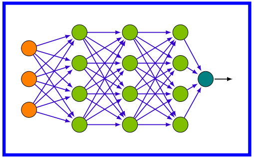
slow, accurate (?)

More than just ML (or simulation)

physics-driven



data-driven

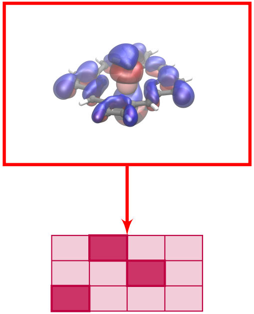


slow, accurate (?)

fast, uncertainty-aware

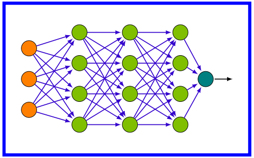
More than just ML (or simulation)

physics-driven



slow, accurate (?)

data-driven

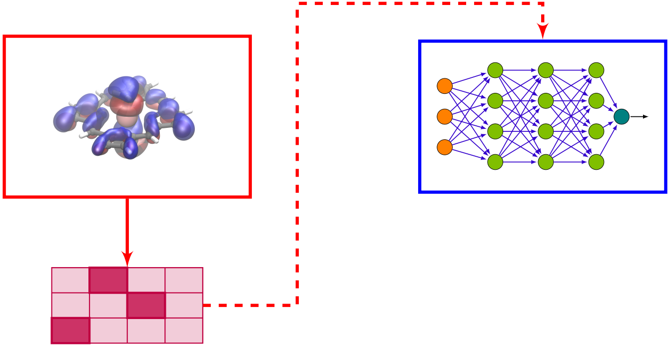


fast, uncertainty-aware

More than just ML (or simulation)

physics-driven

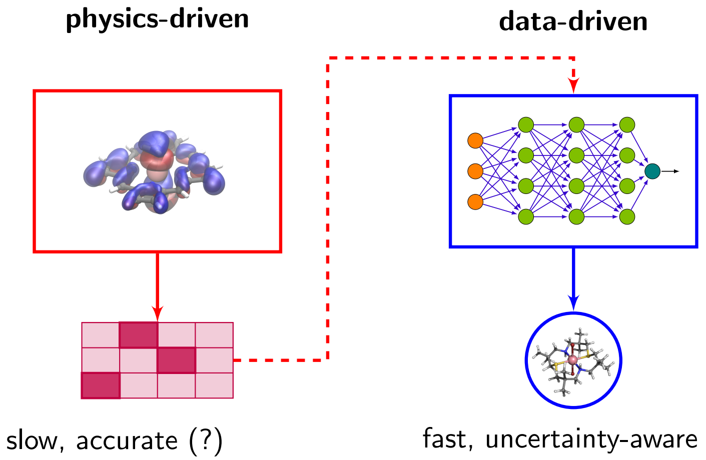
data-driven



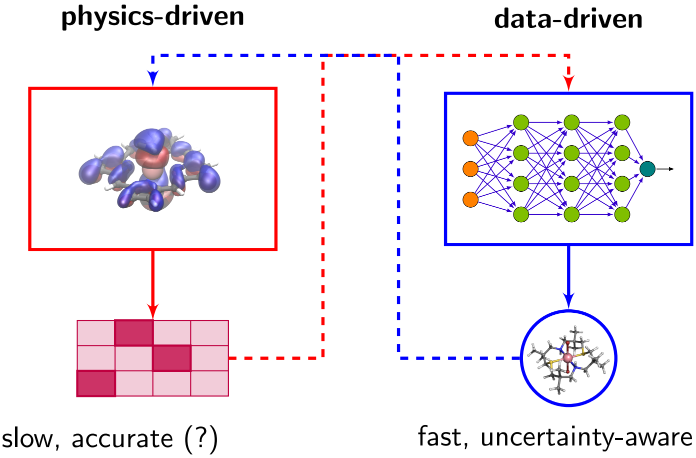
slow, accurate (?)

fast, uncertainty-aware

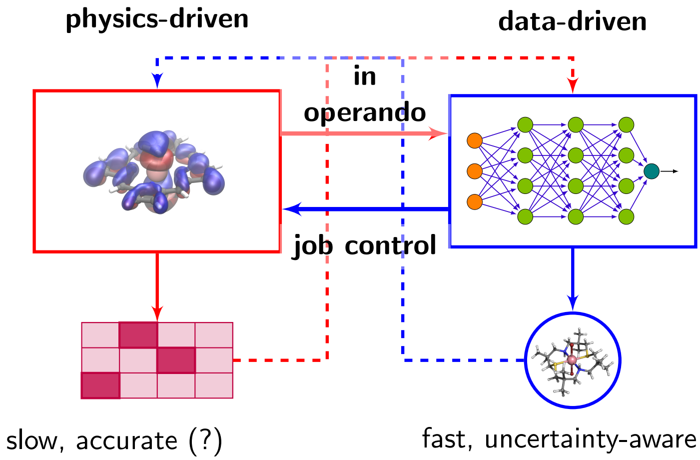
More than just ML (or simulation)



More than just ML (or simulation)



More than just ML (or simulation)



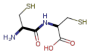
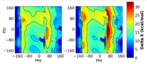
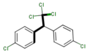
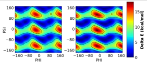
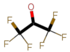
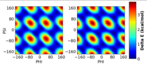
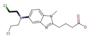
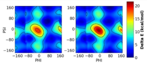
What's next?

Data-driven methods are changing how solve chemical design problems. What's next?

What’s next?

Data-driven methods are changing how solve chemical design problems. What’s next?

- Neural network potentials - quantum accuracy, force field cost!

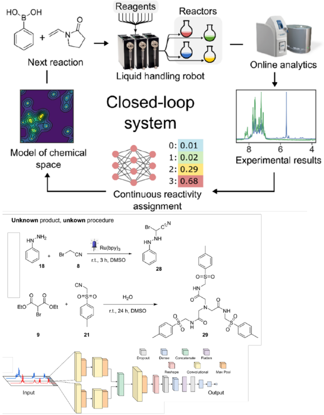
Name	Molecule	MAE	RMSE	Scan (Left:ANI Right:DFT)
Cysteine-Dipeptide **(25 atoms)**		1.75	2.55	
DDT **(28 atoms)**		0.53	0.70	
Hexafluoroacetone **(10 atoms)**		0.08	0.11	
Bendamustine **(44 atoms)**		0.50	0.66	

Devereux, C., *et al.*, *J. Chem. Theory Comput.*, 16(7):4192–4202, 2020

What's next?

Data-driven methods are changing how solve chemical design problems. What's next?

- Neural network potentials - quantum accuracy, force field cost!
- Synthesis planning and optimization. Fully automated chemistry!

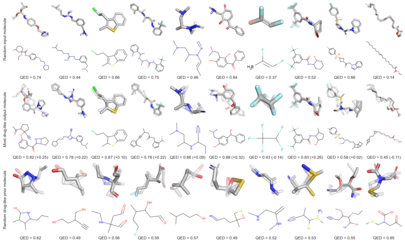


Caramelli, D., *et al.*, *ACS Cent. Sci.*, 7(11):1821–-1830, 2021

What's next?

Data-driven methods are changing how solve chemical design problems. What's next?

- Neural network potentials - quantum accuracy, force field cost!
- Synthesis planning and optimization. Fully automated chemistry!
- Generative models. Designing new drugs directly into the pocket, *de novo*!



Ragoza, M., *et al.*, arXiv:2010.08687v3, 2020

Arcidiacono, M. & Koes, D.R., *et al.*, https://arxiv.org/abs/2109.15308, 2021