upna
Universidad Pública de Navarra
Nafarroako Unibertsitate Publikoa

Navarra Artificial Intelligence Research NAIR CENTER

Universidad de Navarra | DATAI INSTITUTE OF DATA SCIENCE AND ARTIFICIAL INTELLIGENCE

Explainable Artificial Intelligence for Materials Discovery

DATAI Conference 2026
Pamplona, Spain

Valentin Vassilev Galindo

Department of Statistics, Computer Science and Mathematics
Universidad Pública de Navarra, Pamplona, Spain

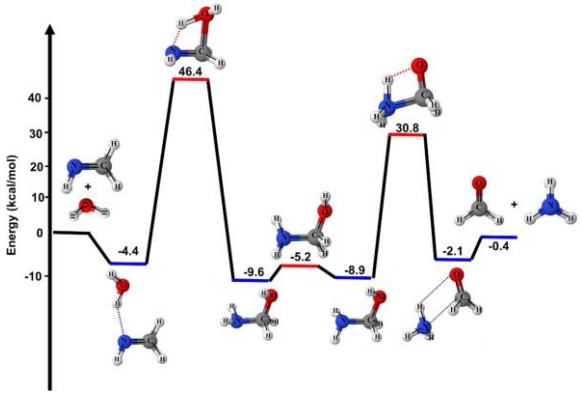
Navarra Artificial Intelligence Research Center (NAIR center)

5th May 2026

✉ valentin.vassilev@unavarra.es

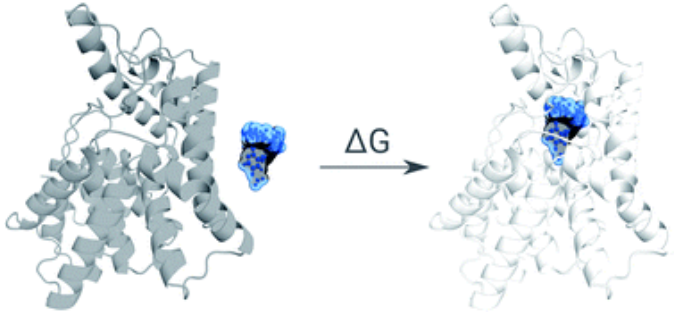
Machine Learning for Computational Chemistry

Reaction Mechanisms



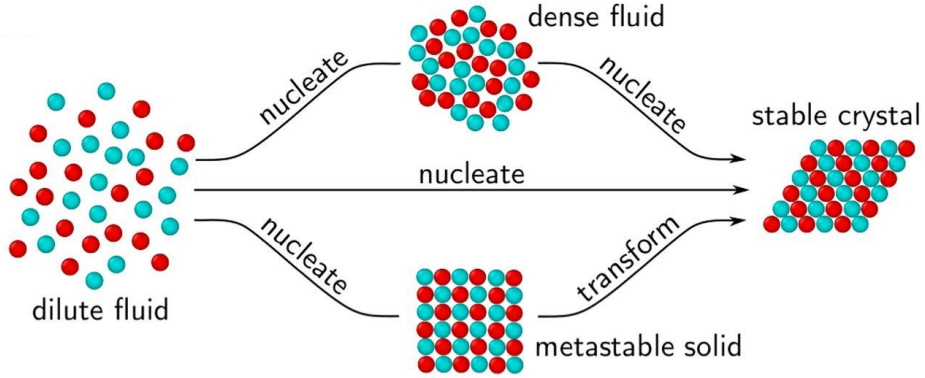
Ali, Y. *Sci. Rep.* **10**, 10995 (2020)

Thermodynamic Binding Constants



Khalak, Y. *et al.*, *Chem. Sci.* **12**, 13958 (2021)

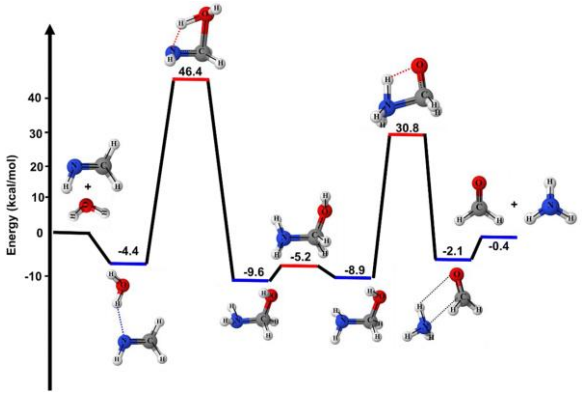
Nucleation Events in Phase Transitions



Fang, H. *et al.*, *PNAS* **117**, 27927 (2020)

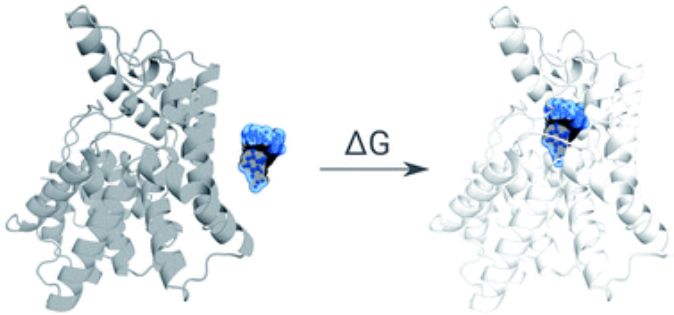
Machine Learning for Computational Chemistry

Reaction Mechanisms



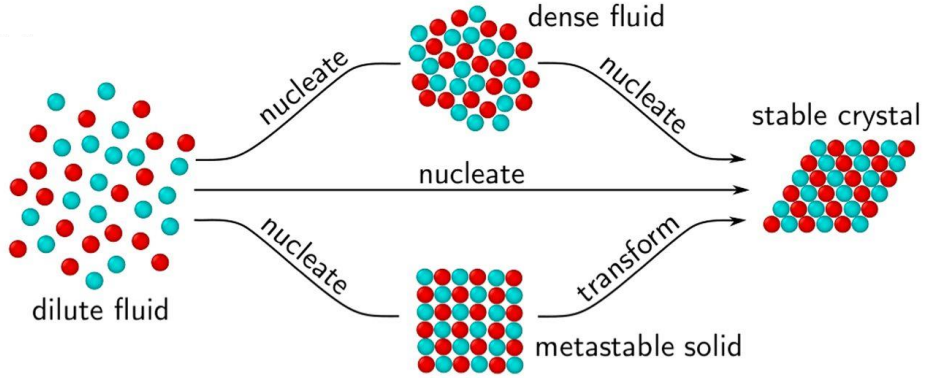
Ali, Y. *Sci. Rep.* **10**, 10995 (2020)

Thermodynamic Binding Constants

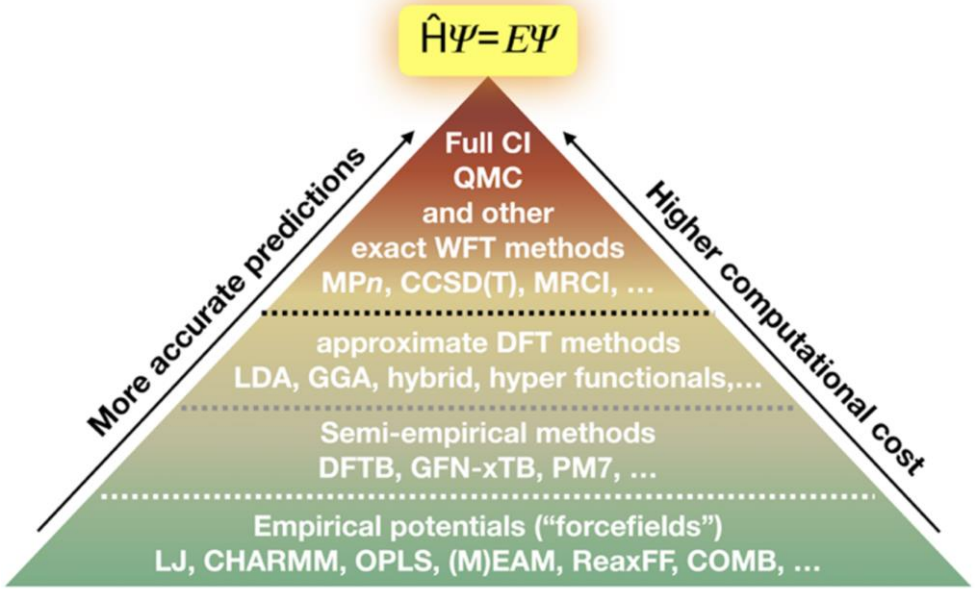


Khalak, Y. *et al.*, *Chem. Sci.* **12**, 13958 (2021)

Nucleation Events in Phase Transitions



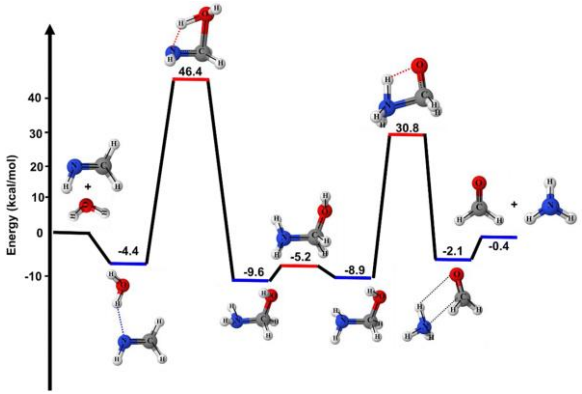
Fang, H. *et al.*, *PNAS* **117**, 27927 (2020)



Keith, J., Vassilev-Galindo, V., *et al.*, *Chem. Rev.* **121**, 9816-9872 (2021)

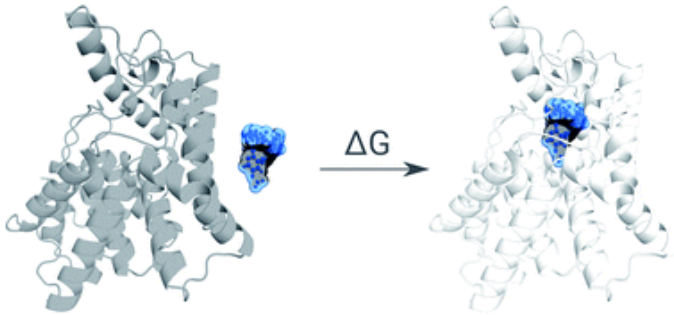
Machine Learning for Computational Chemistry

Reaction Mechanisms



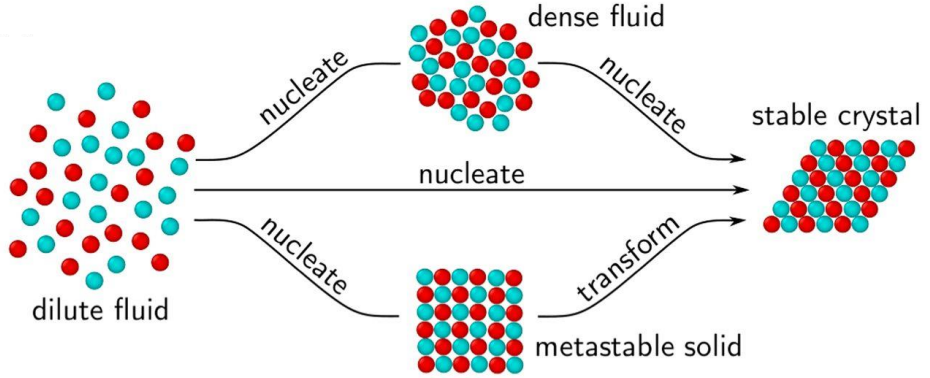
Ali, Y. *Sci. Rep.* **10**, 10995 (2020)

Thermodynamic Binding Constants

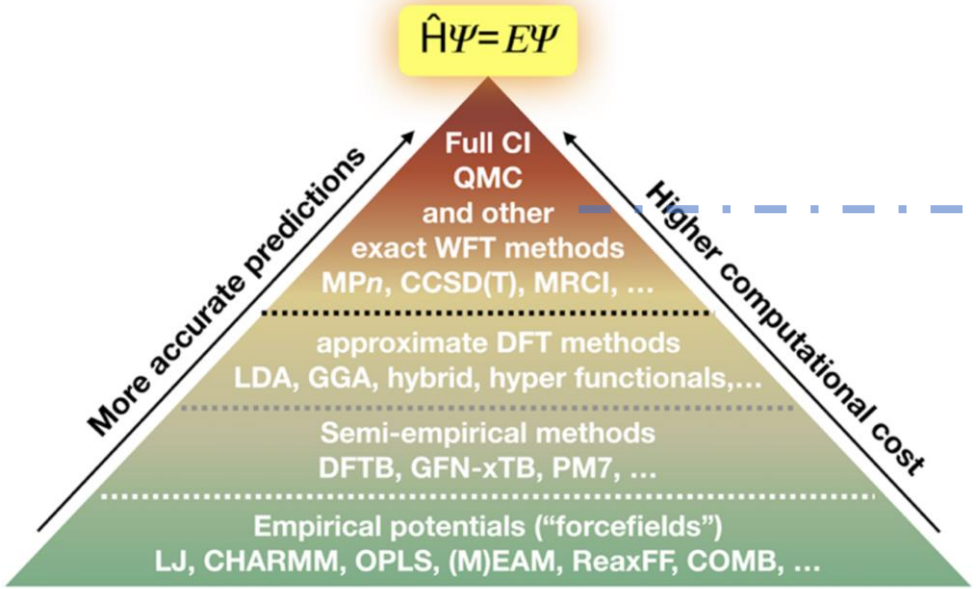


Khalak, Y. *et al.*, *Chem. Sci.* **12**, 13958 (2021)

Nucleation Events in Phase Transitions



Fang, H. *et al.*, *PNAS* **117**, 27927 (2020)



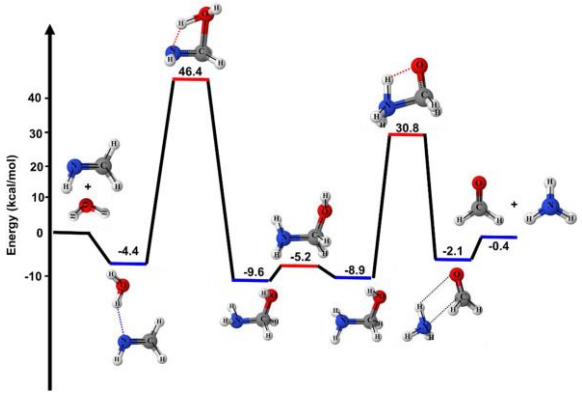
Benchmark calculations
[e.g., CCSD(T)]

Excessive
computational cost (n^7)

Keith, J., Vassilev-Galindo, V., *et al.*, *Chem. Rev.* **121**, 9816-9872 (2021)

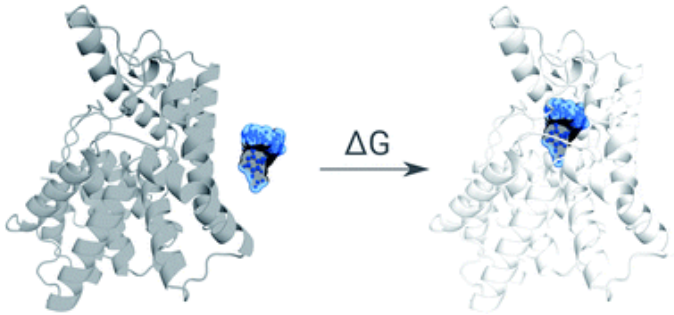
Machine Learning for Computational Chemistry

Reaction Mechanisms



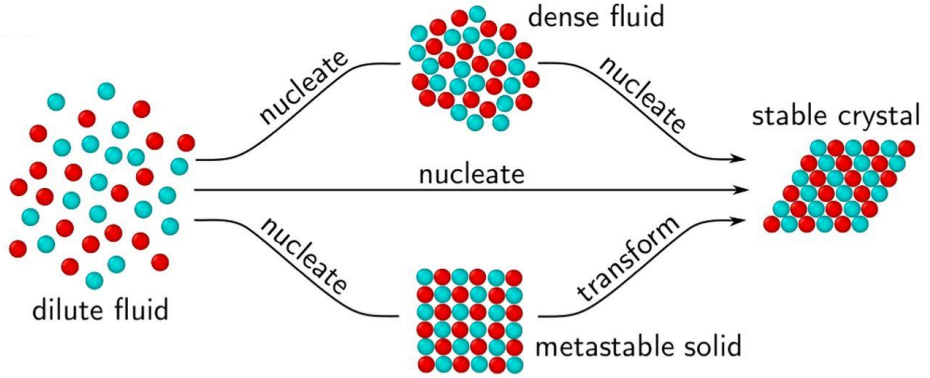
Ali, Y. *Sci. Rep.* **10**, 10995 (2020)

Thermodynamic Binding Constants

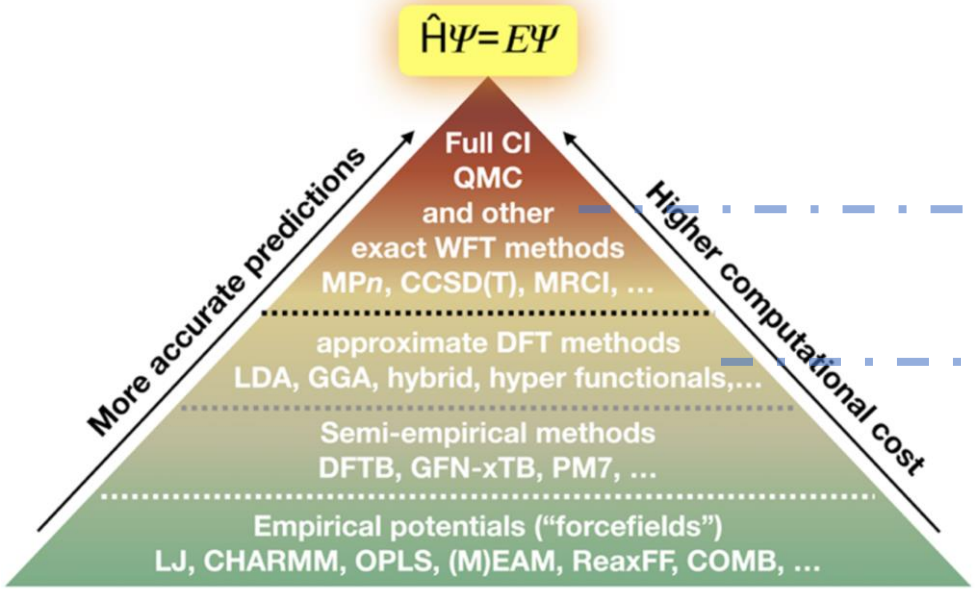


Khalak, Y. *et al.*, *Chem. Sci.* **12**, 13958 (2021)

Nucleation Events in Phase Transitions



Fang, H. *et al.*, *PNAS* **117**, 27927 (2020)



Benchmark calculations [e.g., CCSD(T)]

Excessive computational cost (n^7)

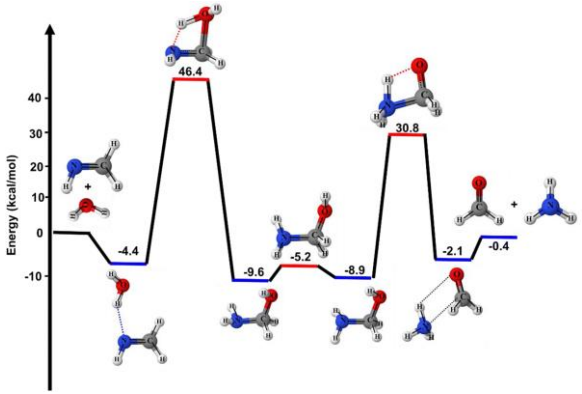
Reliable calculations for molecules and materials

Computational cost is still an obstacle for demanding tasks (n^3)

Keith, J., Vassilev-Galindo, V., *et al.*, *Chem. Rev.* **121**, 9816-9872 (2021)

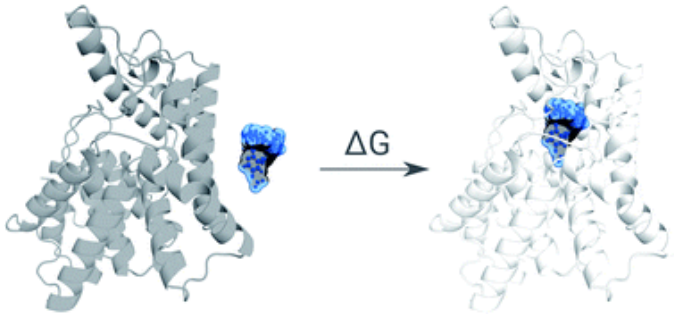
Machine Learning for Computational Chemistry

Reaction Mechanisms



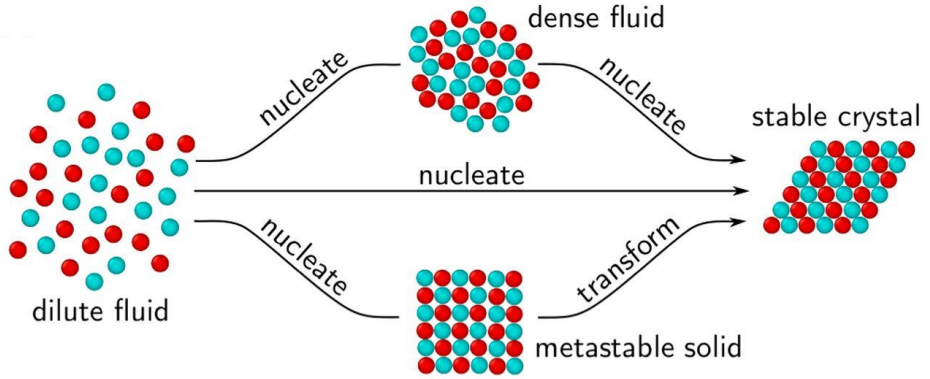
Ali, Y. *Sci. Rep.* **10**, 10995 (2020)

Thermodynamic Binding Constants

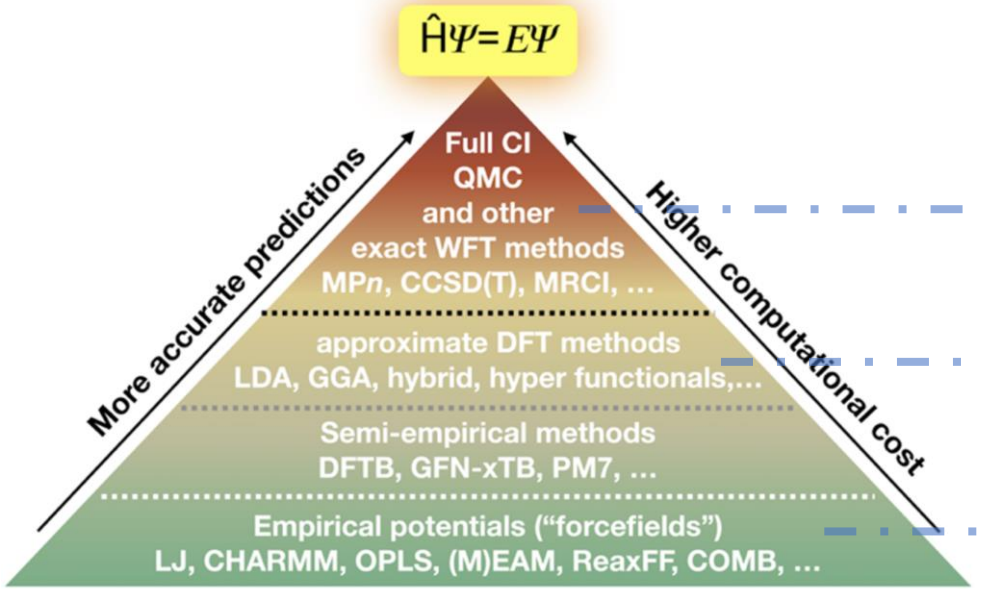


Khalak, Y. *et al.*, *Chem. Sci.* **12**, 13958 (2021)

Nucleation Events in Phase Transitions



Fang, H. *et al.*, *PNAS* **117**, 27927 (2020)

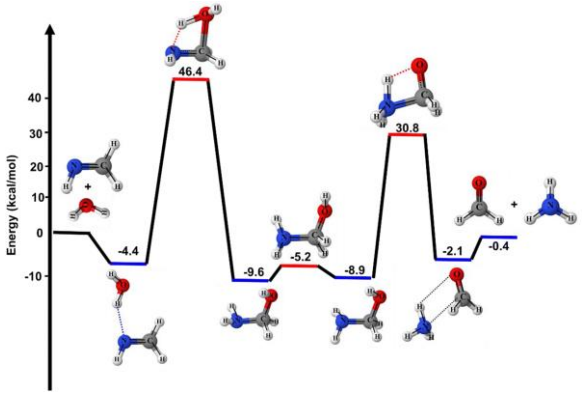


Benchmark calculations [e.g., CCSD(T)]	Excessive computational cost (n^7)
Reliable calculations for molecules and materials	Computational cost is still an obstacle for demanding tasks (n^3)
Efficient for large biomolecular systems	Hard to manually predifine the functional forms of all interactions

Keith, J., Vassilev-Galindo, V., *et al.*, *Chem. Rev.* **121**, 9816-9872 (2021)

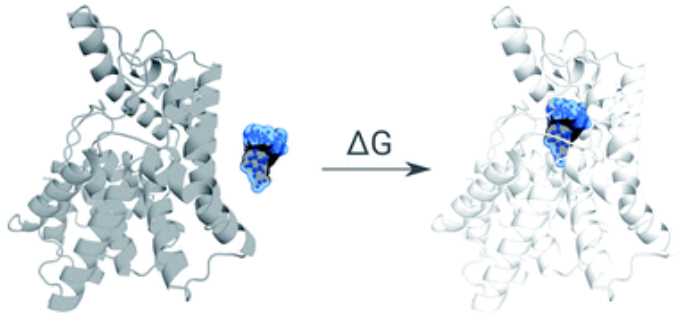
Machine Learning for Computational Chemistry

Reaction Mechanisms



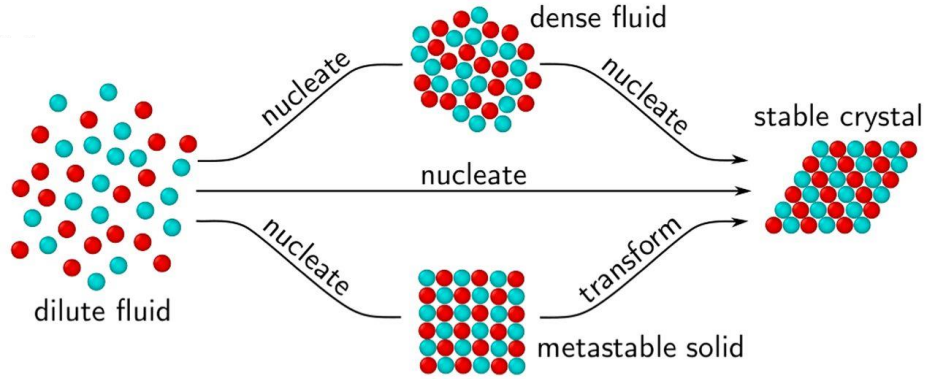
Ali, Y. *Sci. Rep.* **10**, 10995 (2020)

Thermodynamic Binding Constants

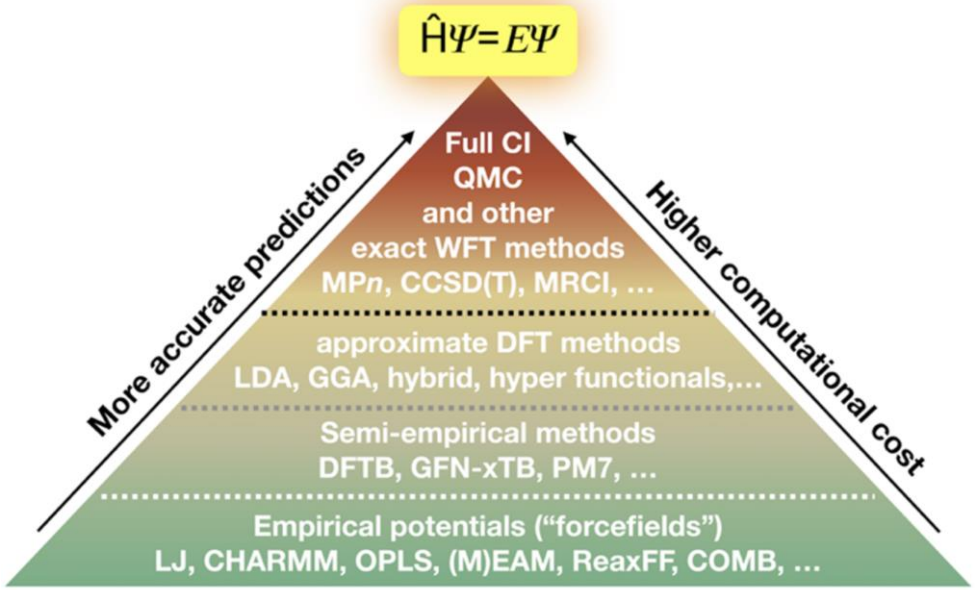


Khalak, Y. *et al.*, *Chem. Sci.* **12**, 13958 (2021)

Nucleation Events in Phase Transitions



Fang, H. *et al.*, *PNAS* **117**, 27927 (2020)



Machine Learning

$\hat{f}: \mathcal{X} \overset{ML}{\rightarrow} \mathcal{Y}$

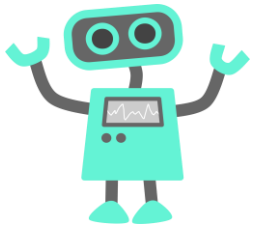
$\mathcal{X}$: Molecular configuration

$\mathcal{Y}$: Property

Keith, J., Vassilev-Galindo, V., *et al.*, *Chem. Rev.* **121**, 9816-9872 (2021)

Materials Design and Discovery

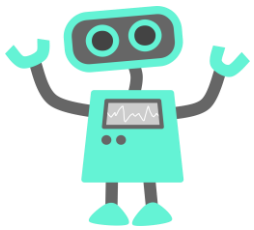
High-throughput screening



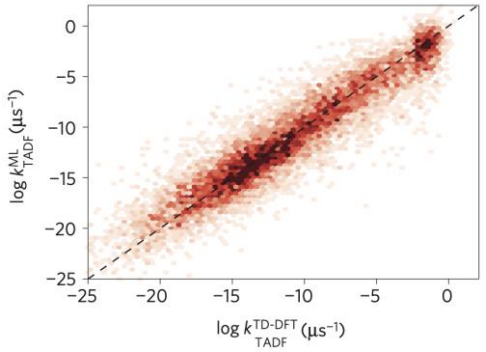
+

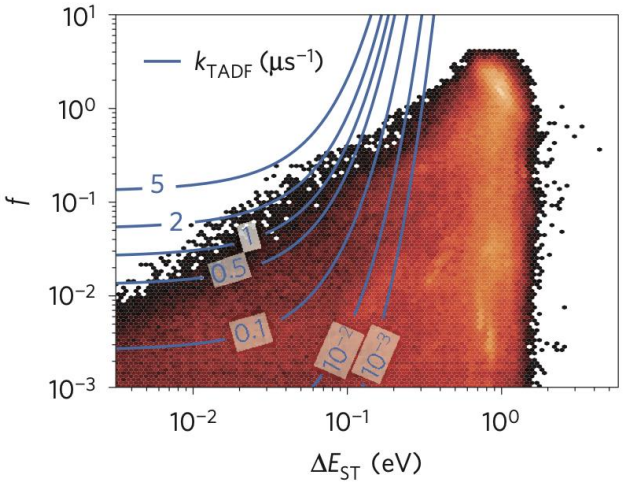
Materials Design and Discovery

High-throughput screening



+

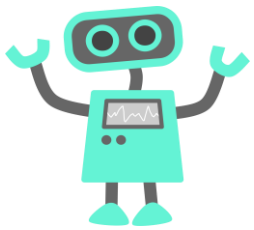




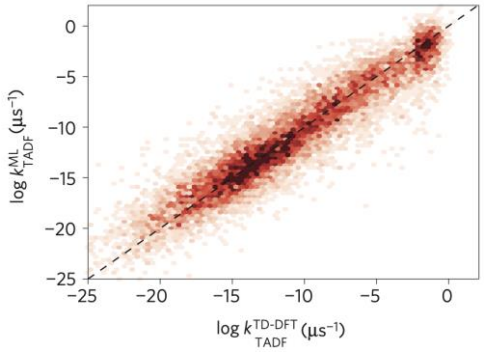
Gómez-Bombarelli, R., *et al. Nature Mater.* **15**, 1120–1127 (2016)

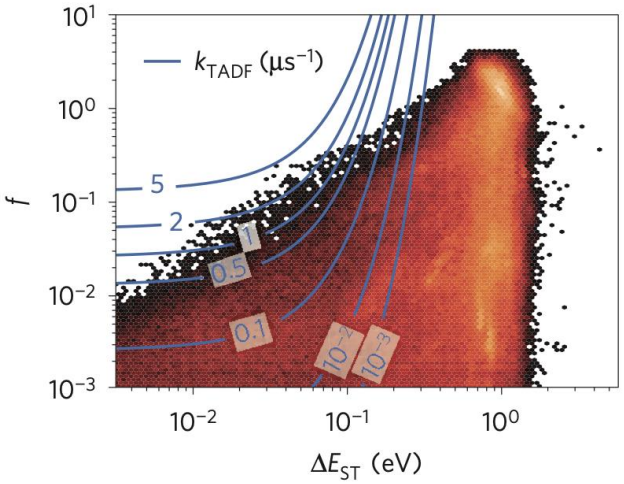
Materials Design and Discovery

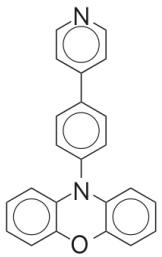
High-throughput screening



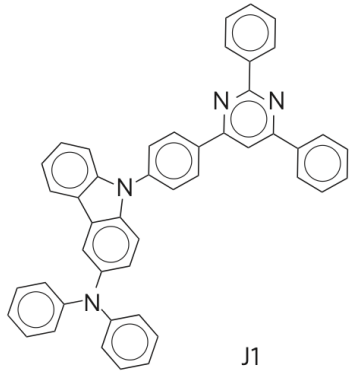
+







F1

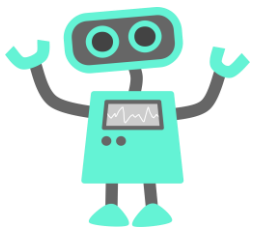


J1

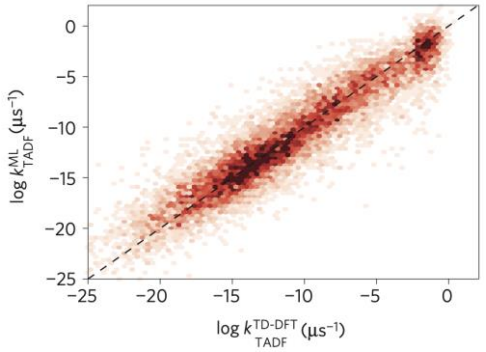
Gómez-Bombarelli, R., *et al. Nature Mater.* **15**, 1120–1127 (2016)

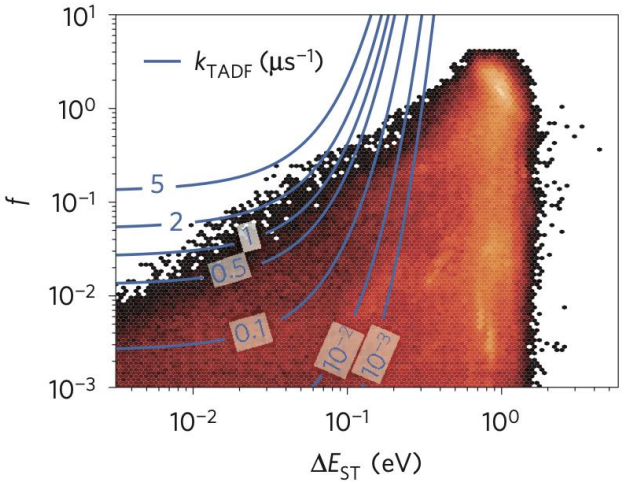
Materials Design and Discovery

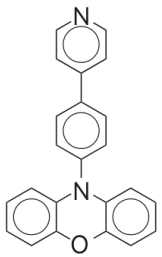
High-throughput screening



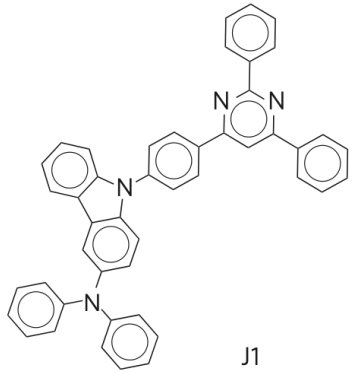
+







F1

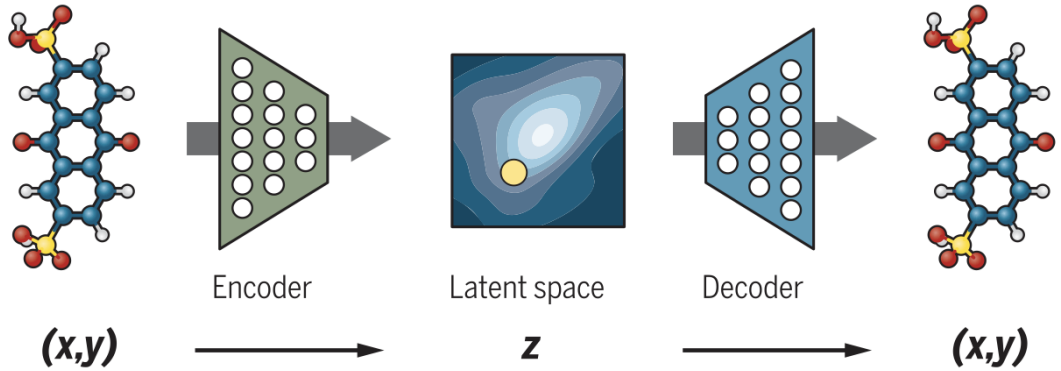


J1

Gómez-Bombarelli, R., *et al. Nature Mater.* **15**, 1120–1127 (2016)

Generative Models

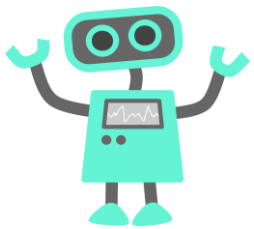
VAE: Variational autoencoders



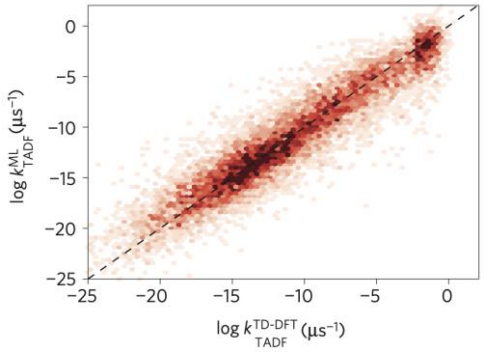
Sanchez-Lengeling, B. & Aspuru-Guzik, A. *Science* **361**, 360–365 (2018).

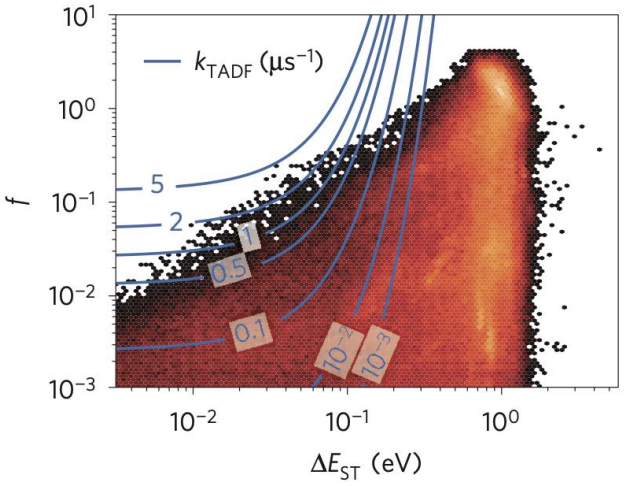
Materials Design and Discovery

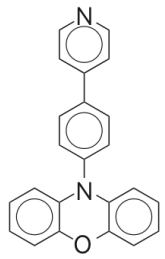
High-throughput screening



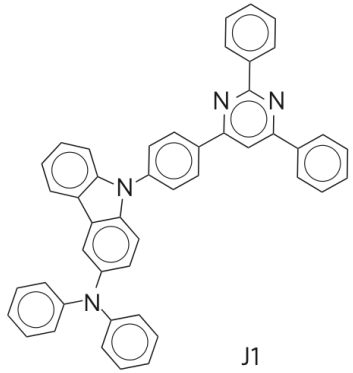
+







F1

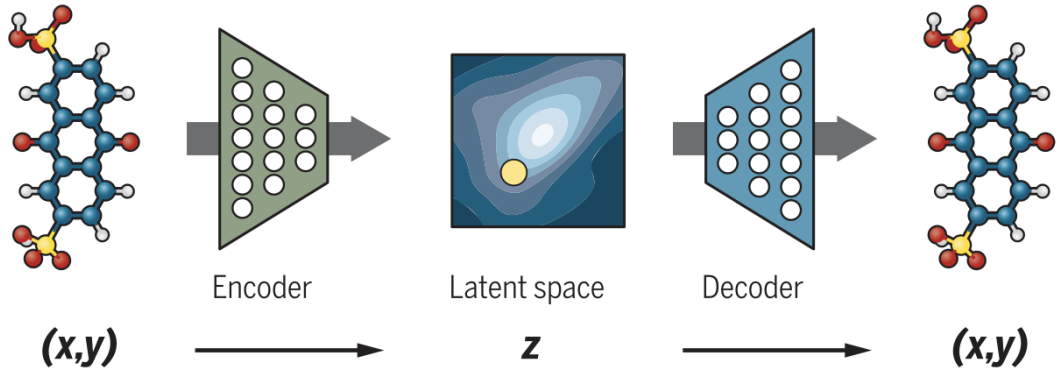


J1

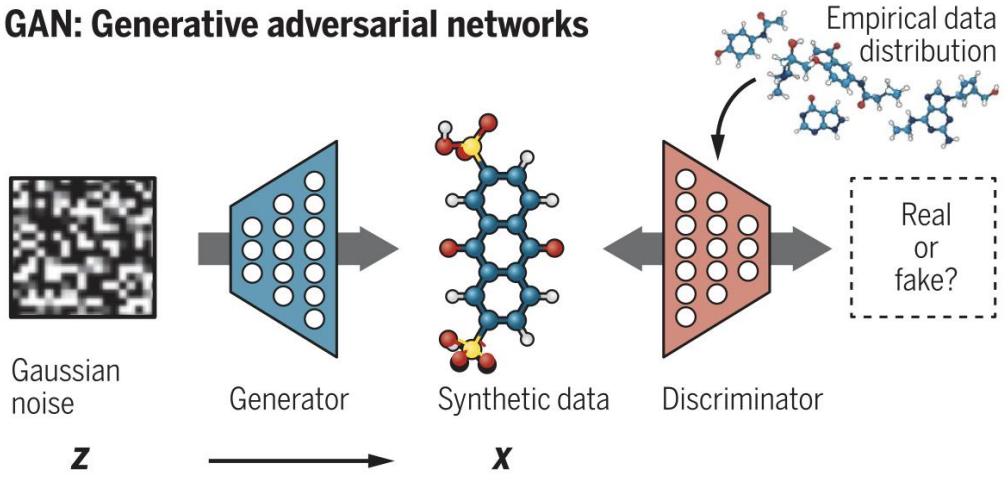
Gómez-Bombarelli, R., *et al. Nature Mater.* **15**, 1120–1127 (2016)

Generative Models

VAE: Variational autoencoders



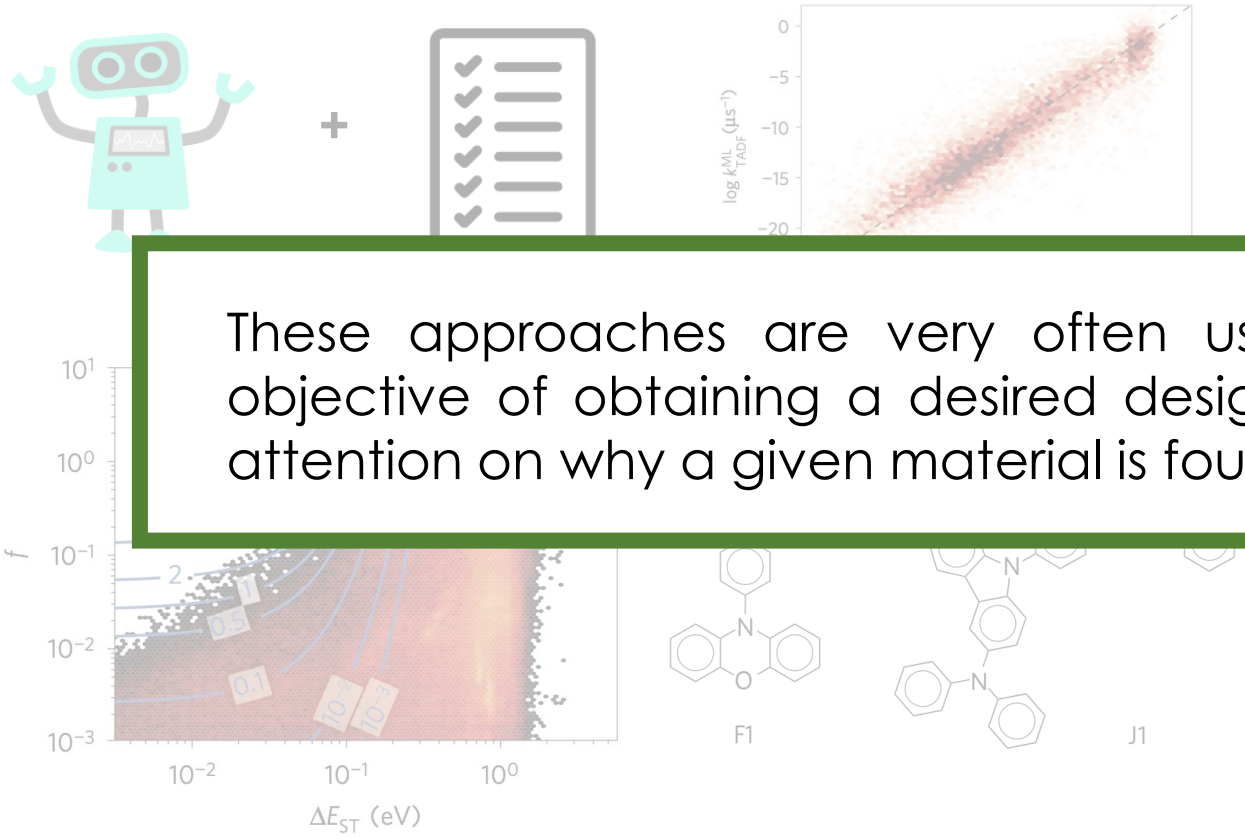
GAN: Generative adversarial networks



Sanchez-Lengeling, B. & Aspuru-Guzik, A. *Science* **361**, 360–365 (2018).

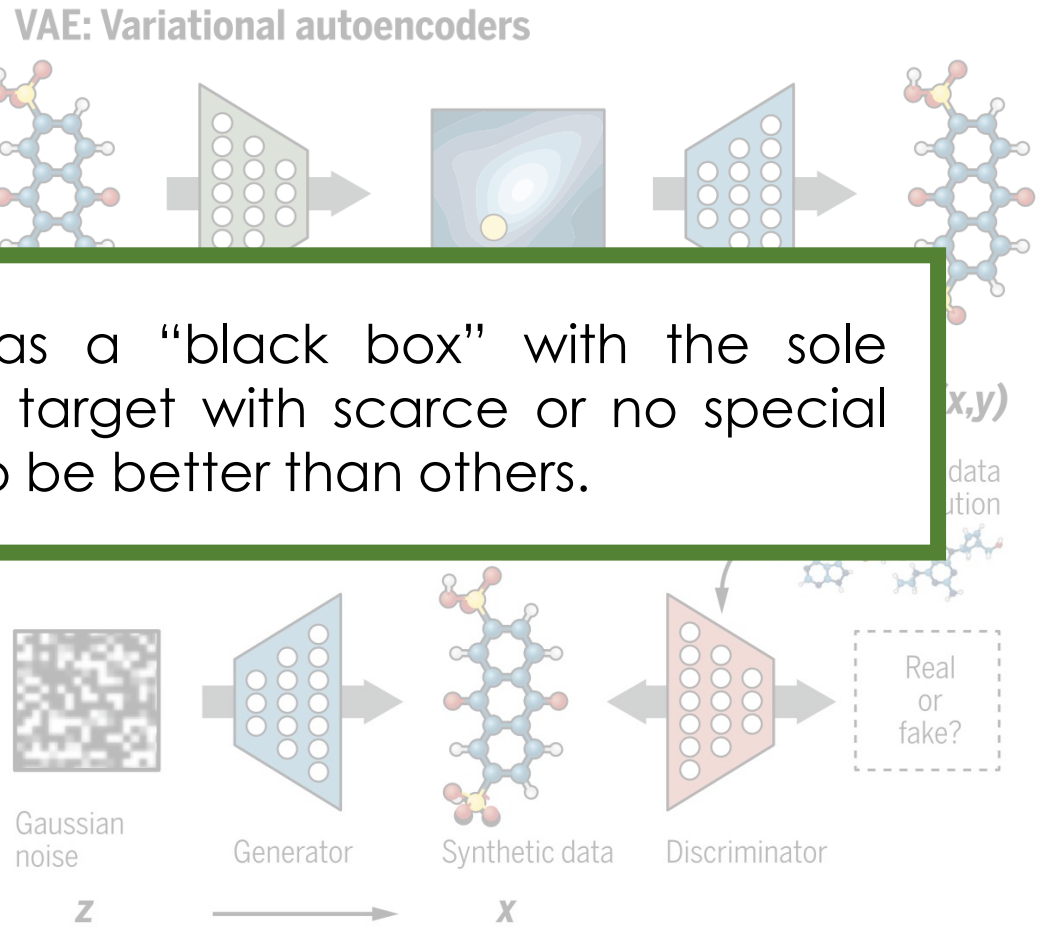
Materials Design and Discovery

High-throughput screening



Gómez-Bombarelli, R., *et al. Nature Mater.* **15**, 1120–1127 (2016)

Generative Models



Sanchez-Lengeling, B. & Aspuru-Guzik, A. *Science* **361**, 360–365 (2018).

These approaches are very often used as a "black box" with the sole objective of obtaining a desired designed target with scarce or no special attention on why a given material is found to be better than others.

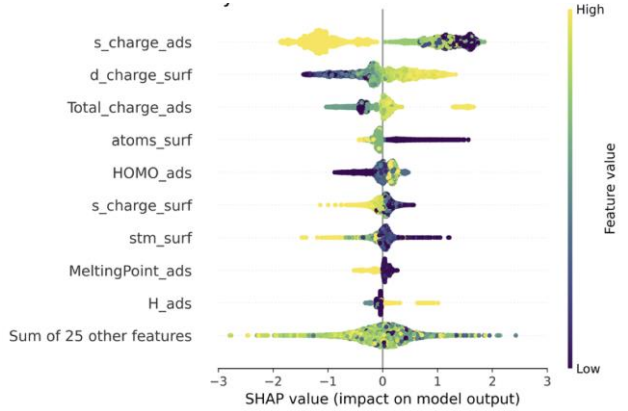
Explainability

Assessment of prediction patterns to understand why a model makes specific decisions

Explainability

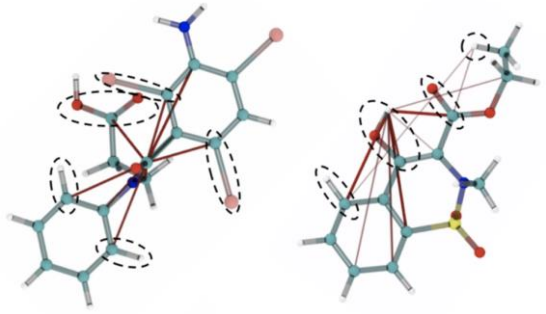
Assessment of prediction patterns to understand why a model makes specific decisions

Feature importance attribution



Usuga A. F., *et al. J. Mater. Chem. A* **12**, 2708 (2024).

Attention mechanisms

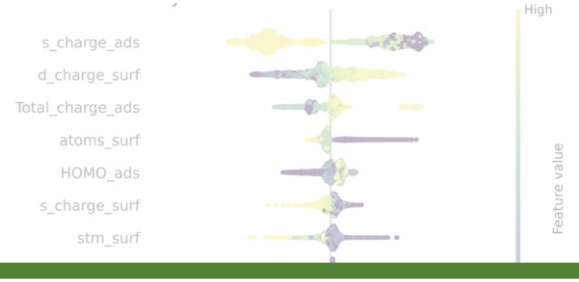


Cremer, J., *et al. Chem. Res. Toxicol.* 36, 1561-1573 (2023).

Explainability

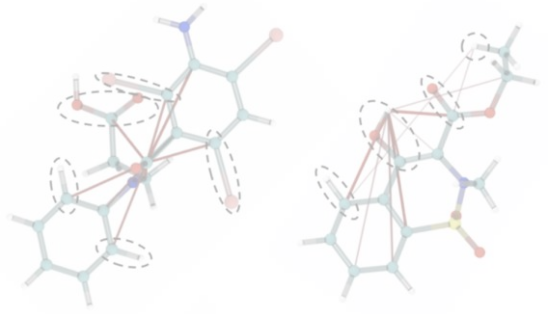
Assessment of prediction patterns to understand why a model makes specific decisions

Feature importance attribution



These methods are not actionable
(i.e., they do not tell how a given input can be changed in order to modify the output).

Attention mechanisms

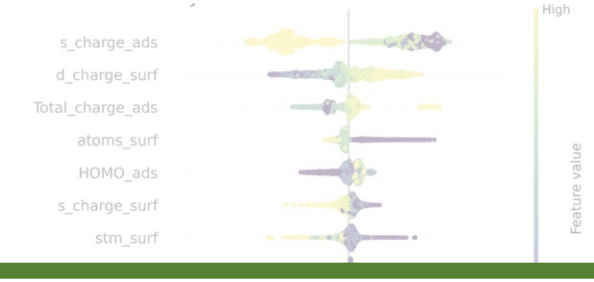


Cremer, J., *et al. Chem. Res. Toxicol.* 36, 1561-1573 (2023).

Explainability

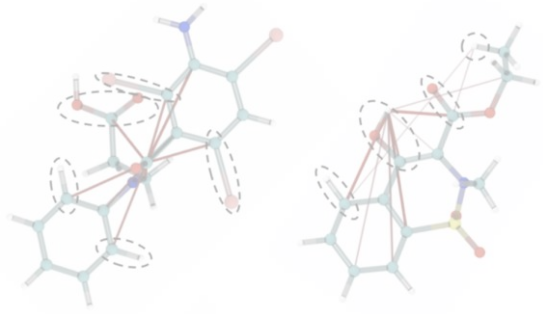
Assessment of prediction patterns to understand why a model makes specific decisions

Feature importance attribution



These methods are not actionable (i.e., they do not tell how a given input can be changed in order to modify the output).

Attention mechanisms



Cremer, J., *et al. Chem. Res. Toxicol.* 36, 1561-1573 (2023).

Counterfactual explanations

Provide insights of model operation by determining examples or cases that explain the difference between a desired outcome and actual outcome

Classification

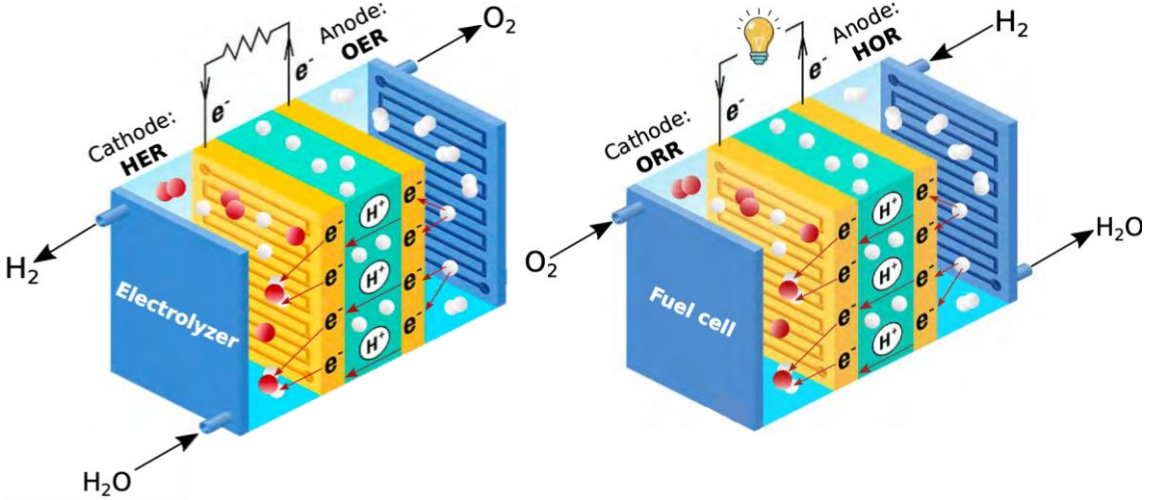
$$minimize \quad d(\boldsymbol{x}, \boldsymbol{x}')$$
$$such\ that \quad \hat{f}(\boldsymbol{x}) \neq \hat{f}(\boldsymbol{x}')$$

Regression

$$minimize \quad d(\boldsymbol{x}, \boldsymbol{x}')$$
$$such\ that \quad \left|\hat{f}(\boldsymbol{x}) - \hat{f}(\boldsymbol{x}')\right| \geq \Delta$$

Catalysts for the HER and ORR

Hydrogen production and energy generation



Martínez-Alonso, C., *High-throughput computational strategies to discover new catalysts for the hydrogen economy including elastic strain engineering* (2024).

HER

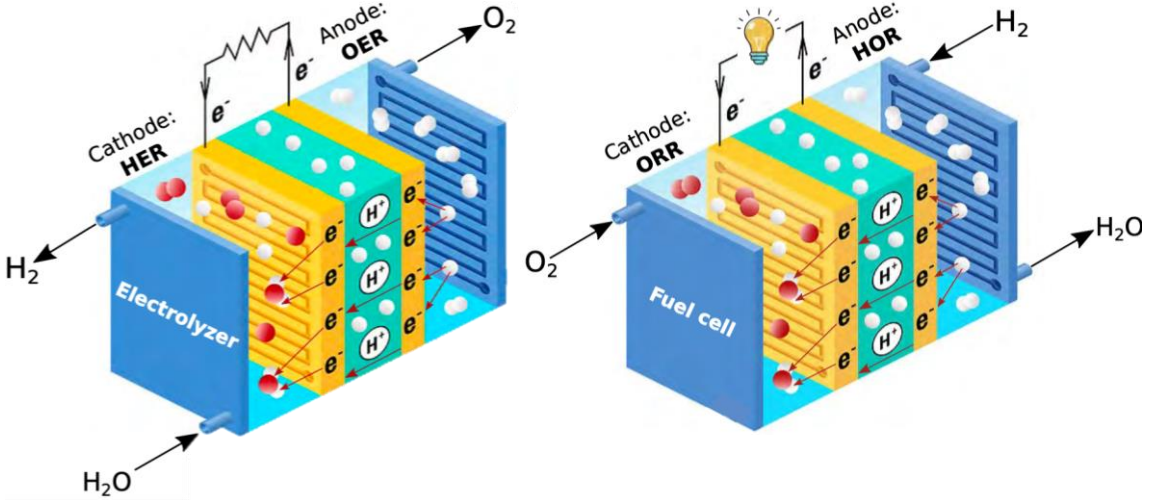
$$H^+ + * + e^- \longrightarrow H^*$$
$$H^* + H^+ + e^- \longrightarrow * + H_2$$
$$2\,H^* \longrightarrow 2\,* + H_2,$$

ORR

$$\frac{1}{2}O_2 + * \longrightarrow O^*$$
$$O^* + H^+ + e^- \longrightarrow HO^*$$
$$HO^* + H^+ + e^- \longrightarrow H_2O + *,$$

Catalysts for the HER and ORR

Hydrogen production and energy generation



Martínez-Alonso, C., *High-throughput computational strategies to discover new catalysts for the hydrogen economy including elastic strain engineering* (2024).

HER

$$H^+ + * + e^- \longrightarrow H^*$$
$$H^* + H^+ + e^- \longrightarrow * + H_2$$
$$2\,H^* \longrightarrow 2\,* + H_2,$$

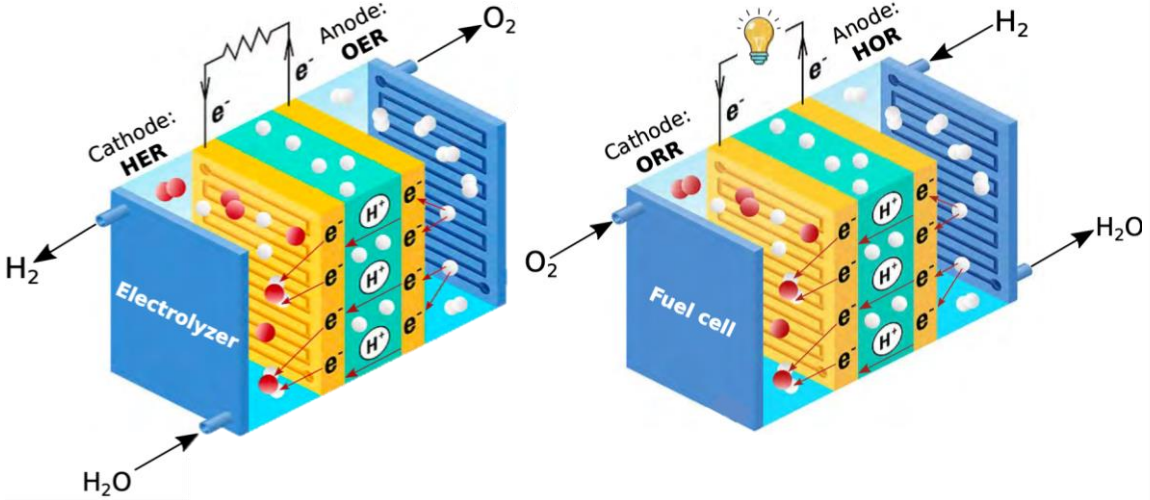
ORR

$$\frac{1}{2}O_2 + * \longrightarrow O^*$$
$$O^* + H^+ + e^- \longrightarrow HO^*$$
$$HO^* + H^+ + e^- \longrightarrow H_2O + *,$$

E_{ads} H O OH

Catalysts for the HER and ORR

Hydrogen production and energy generation



Martínez-Alonso, C., *High-throughput computational strategies to discover new catalysts for the hydrogen economy including elastic strain engineering* (2024).

HER

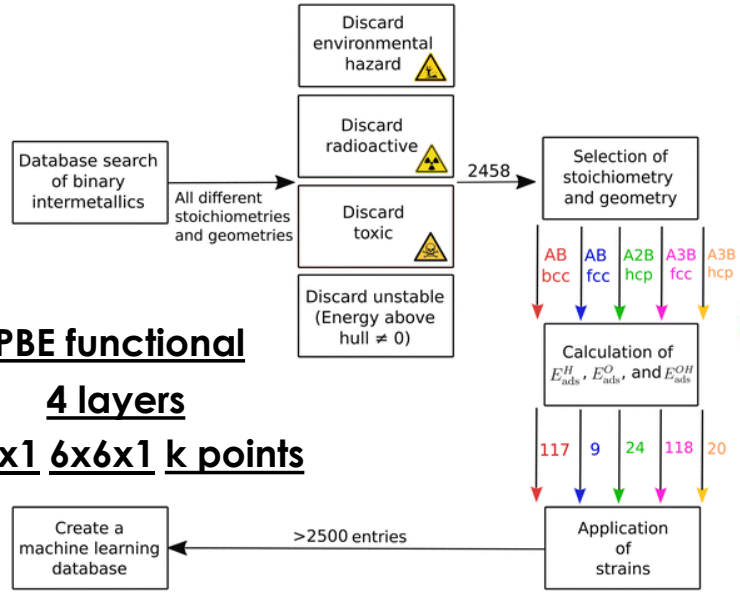
$$H^+ + * + e^- \longrightarrow H^*$$
$$H^* + H^+ + e^- \longrightarrow * + H_2$$
$$2\,H^* \longrightarrow 2\,* + H_2,$$

ORR

$$\frac{1}{2}O_2 + * \longrightarrow O^*$$
$$O^* + H^+ + e^- \longrightarrow HO^*$$
$$HO^* + H^+ + e^- \longrightarrow H_2O + *,$$

E_{ads} (H) (O) (OH)

Dataset of E_{ads}

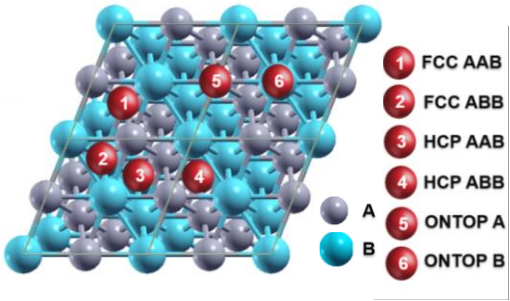


PBE functional

4 layers

4x4x1 6x6x1 k points

Adsorption sites



Elastic strains

$$\boldsymbol{F} = \begin{pmatrix} 1+\epsilon & 0 \\ 0 & 1+\epsilon \end{pmatrix}$$

Geometric	Electronic	Strain
Unit cell volume	WEN	Biaxial strain
WAR	WIE	
GCN	S_A	
	S_B	
	Ψ	

$$GCN = \frac{\sum_{i=1}^{N} CN_i}{CN_{max}}$$

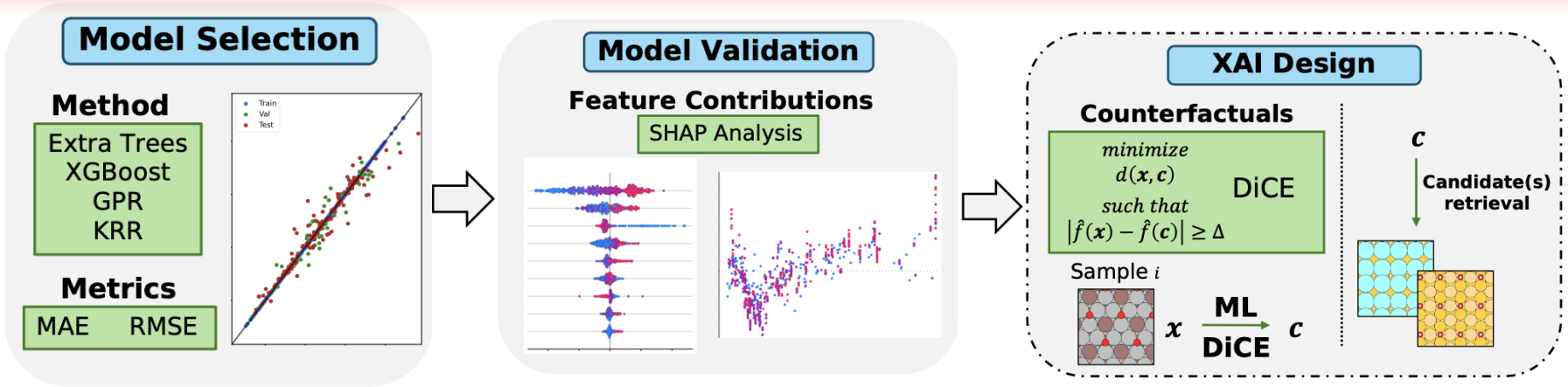
Angew. Chem. **53**, 8316–8319 (2014)

$$\Psi = \frac{\left(\prod_{i=1}^{N} S_i\right)^{\frac{2}{N}}}{\left(\prod_{i=1}^{N} EN_i\right)^{\frac{1}{N}}}$$

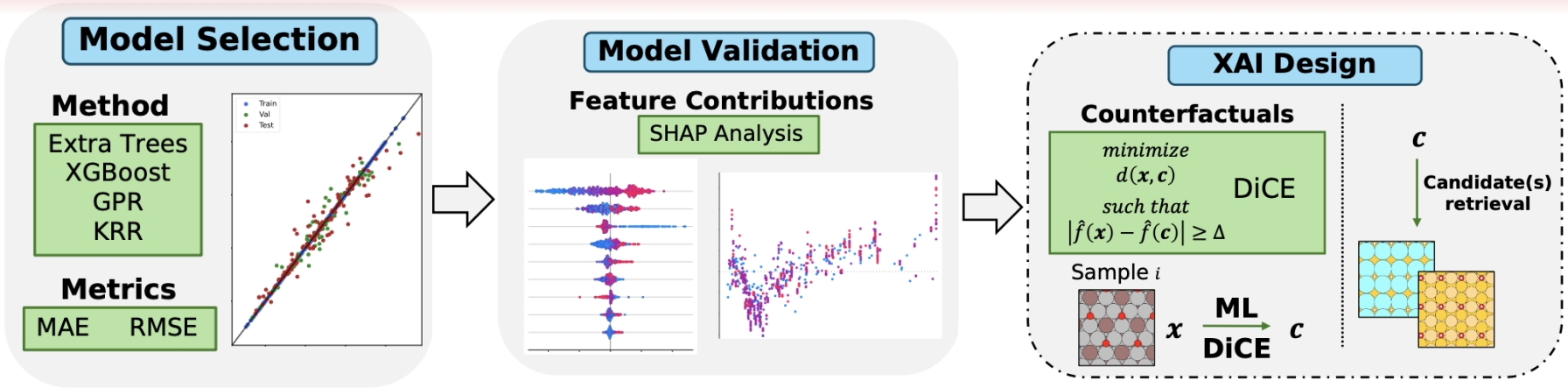
Nat. Commun. **11**, 1196 (2020)

Martínez-Alonso, C., Vassilev-Galindo, V., *et al.*, *Catal. Sci. Technol.* **14**, 3784-3799 (2024)

Explainable Artificial Intelligence (XAI) Strategy



Explainable Artificial Intelligence (XAI) Strategy

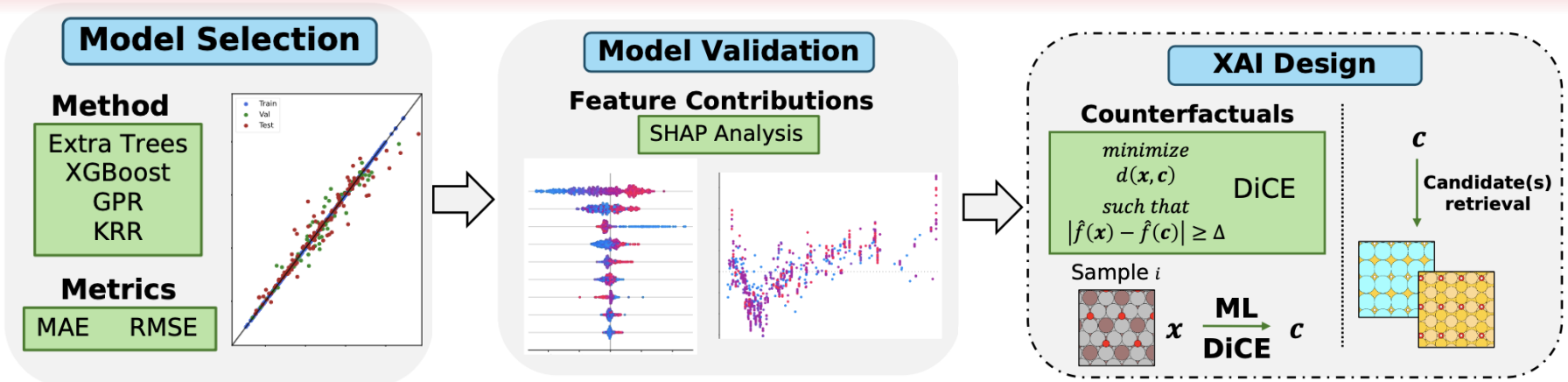


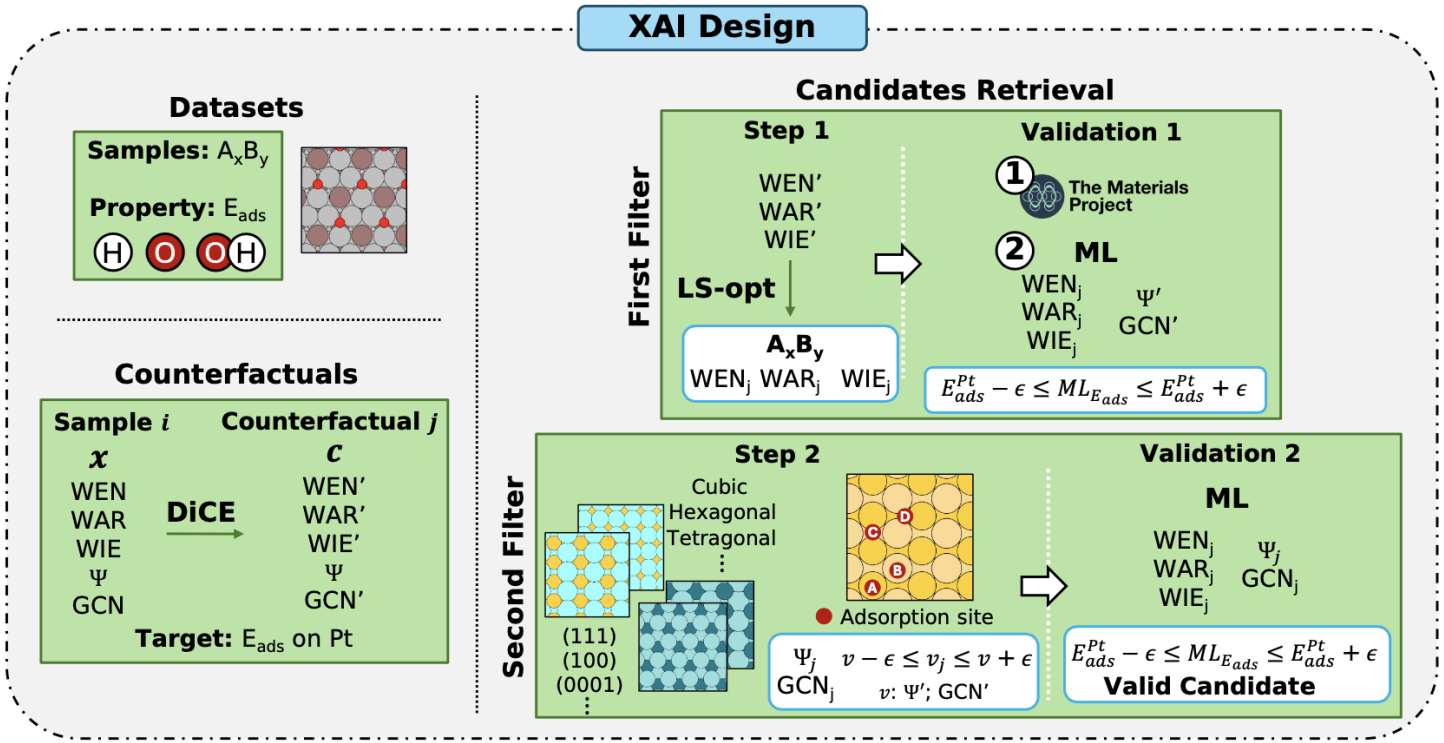
DiCE
Diverse Counterfactual Explanations

$$C(\mathbf{x}) = \arg\min_{\mathbf{c}_1,\ldots,\mathbf{c}_k} \frac{1}{k}\sum_{i=1}^{k} yloss(f(\mathbf{c}_i), y) + \frac{\lambda_1}{k}\sum_{i=1}^{k} dist(\mathbf{c}_i, \mathbf{x}) - \lambda_2 dpp\_diversity(\mathbf{c}_1, \ldots, \mathbf{c}_k)$$

Mothilal, R. K., *et al.* in *Proceedings of the 2020 conference on fairness, accountability, and transparency*, 607–617 (2020)

Explainable Artificial Intelligence (XAI) Strategy





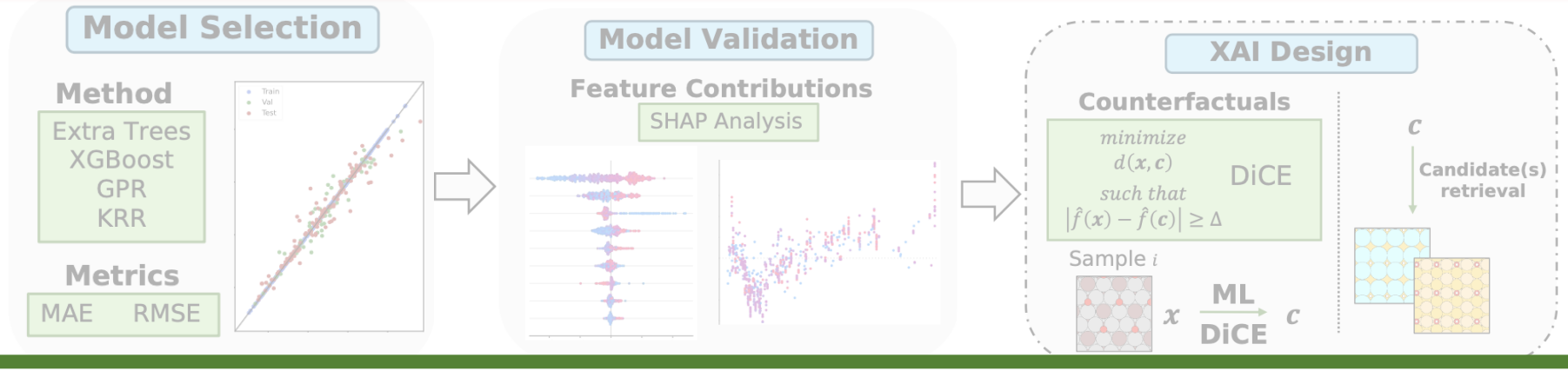
DiCE
Diverse Counterfactual Explanations

$$C(\mathbf{x}) = \arg\min_{\mathbf{c}_1,\ldots,\mathbf{c}_k} \frac{1}{k}\sum_{i=1}^{k} yloss(f(\mathbf{c}_i), y) + \frac{\lambda_1}{k}\sum_{i=1}^{k} dist(\mathbf{c}_i, \mathbf{x}) - \lambda_2 dpp\_diversity(\mathbf{c}_1, \ldots, \mathbf{c}_k)$$

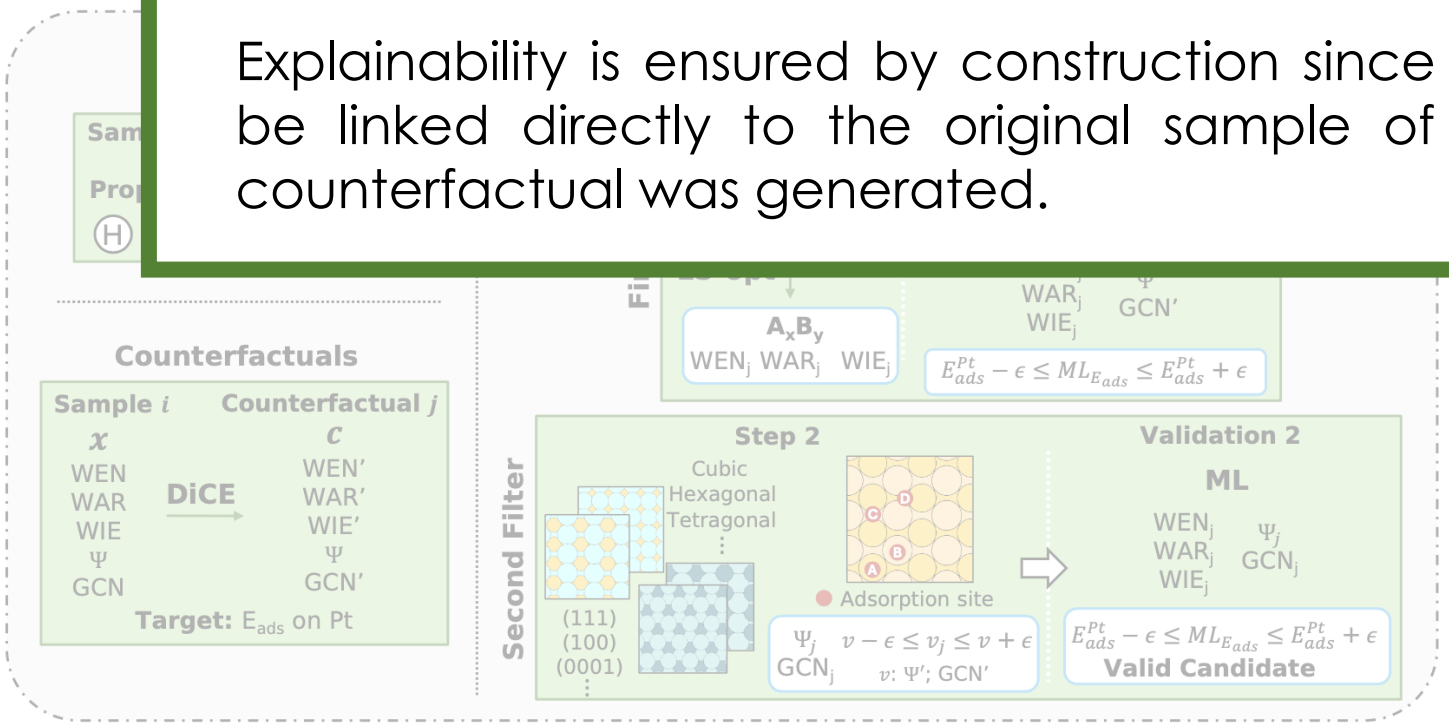
Mothilal, R. K., *et al.* in *Proceedings of the 2020 conference on fairness, accountability, and transparency*, 607–617 (2020)

V. Vassilev-Galindo & J. LLorca. *Chem. Sci.* **17**, 1058-1072 (2026)

Explainable Artificial Intelligence (XAI) Strategy



Explainability is ensured by construction since the discovered materials can be linked directly to the original sample of the dataset from which the counterfactual was generated.



$$C(\mathbf{x}) = \arg\min_{\mathbf{c}_1,\ldots,\mathbf{c}_k} \frac{1}{k}\sum_{i=1}^{k} yloss(f(\mathbf{c}_i), y) + \frac{\lambda_1}{k}\sum_{i=1}^{k} dist(\mathbf{c}_i, \mathbf{x}) - \lambda_2 dpp\_diversity(\mathbf{c}_1, \ldots, \mathbf{c}_k)$$

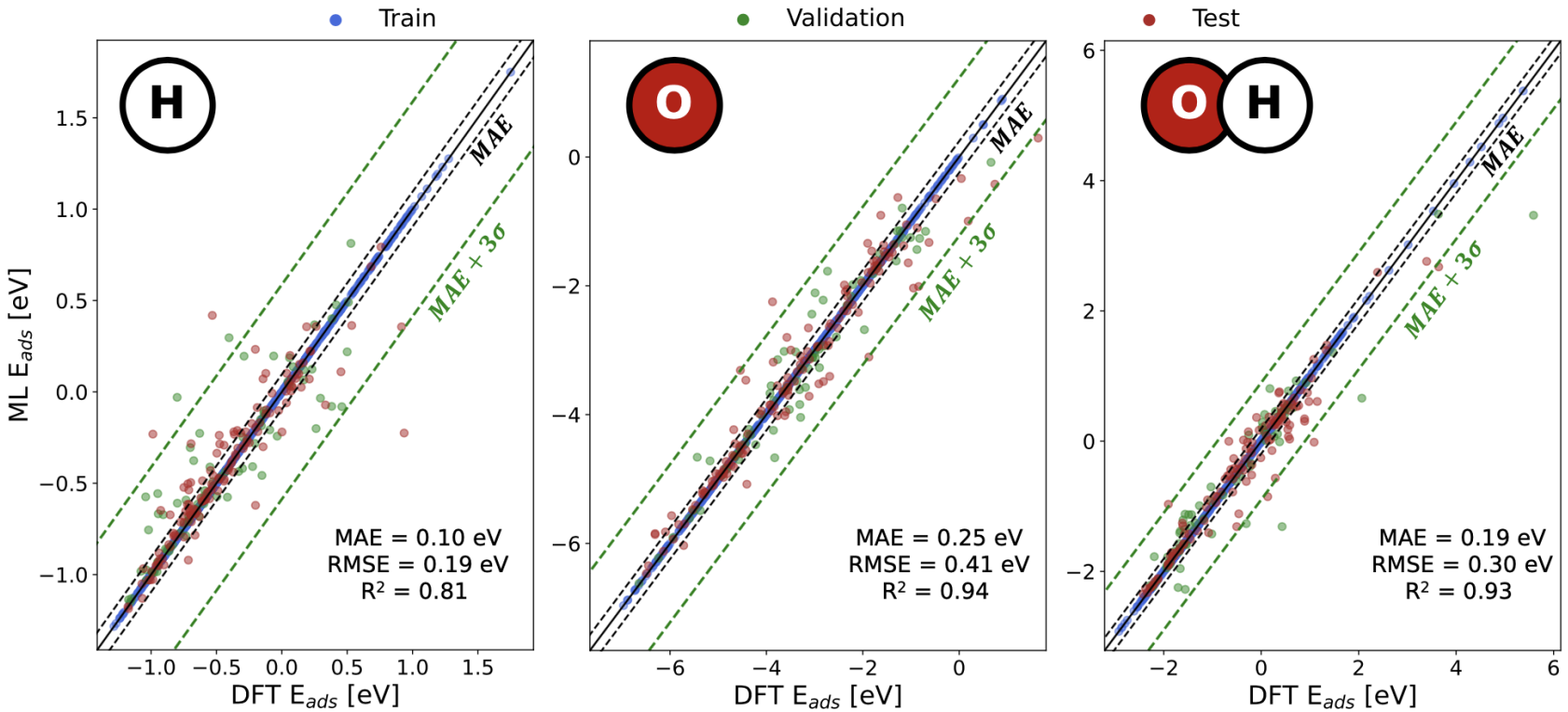
Mothilal, R. K., *et al.* in *Proceedings of the 2020 conference on fairness, accountability, and transparency*, 607–617 (2020)

Model Selection

Training settings

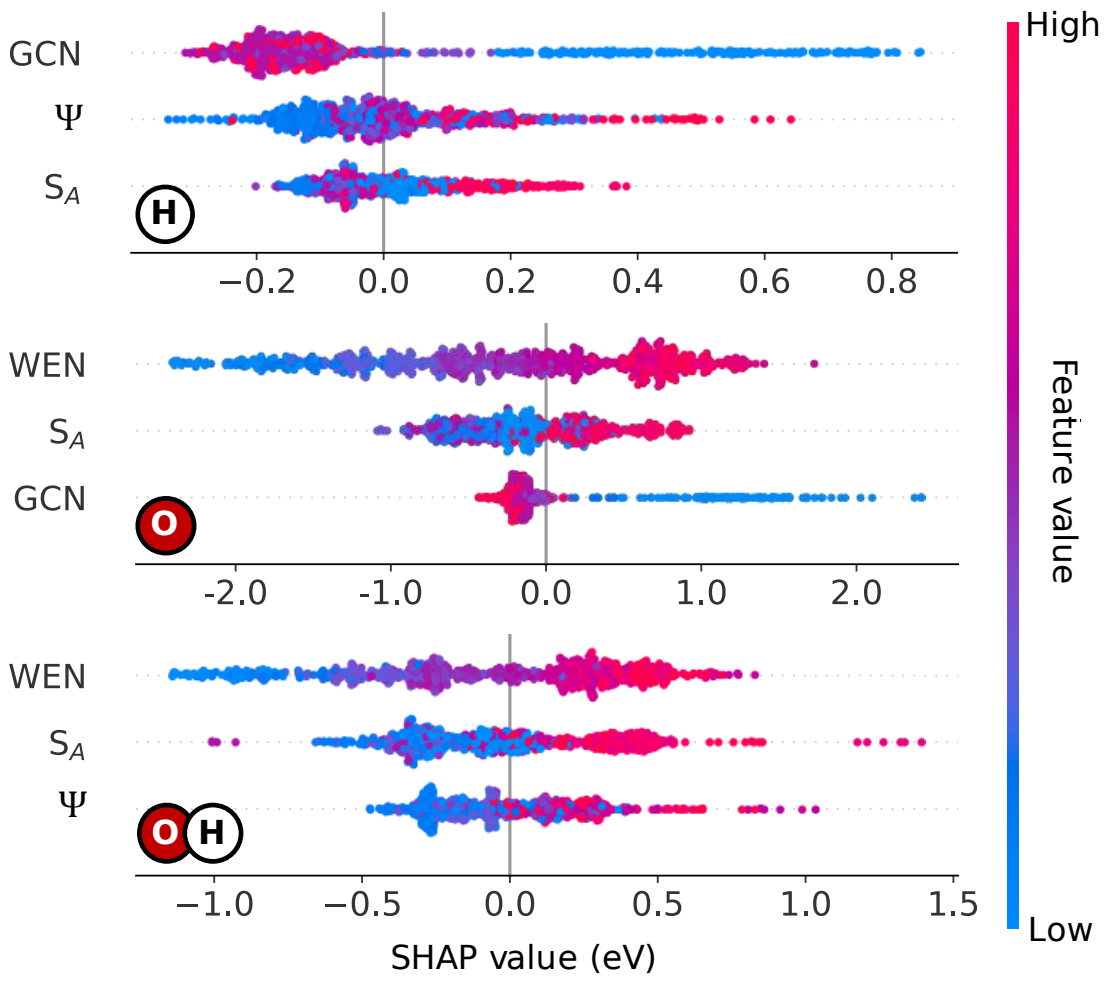
- ❑ 15% samples for testing
- ❑ 85% samples for training/validation
- ❑ 10-fold "pseudo-random" cross-validation

Adsorbate	Metric	ET	XGBoost	KRR	GPR
H	MAE	**0.10**	0.11	0.13	0.15
	RMSE	**0.21**	**0.21**	0.24	0.25
O	MAE	**0.25**	0.26	0.34	0.34
	RMSE	**0.43**	0.47	0.54	0.55
OH	MAE	**0.20**	0.28	0.27	0.41
	RMSE	**0.33**	0.45	0.42	0.60



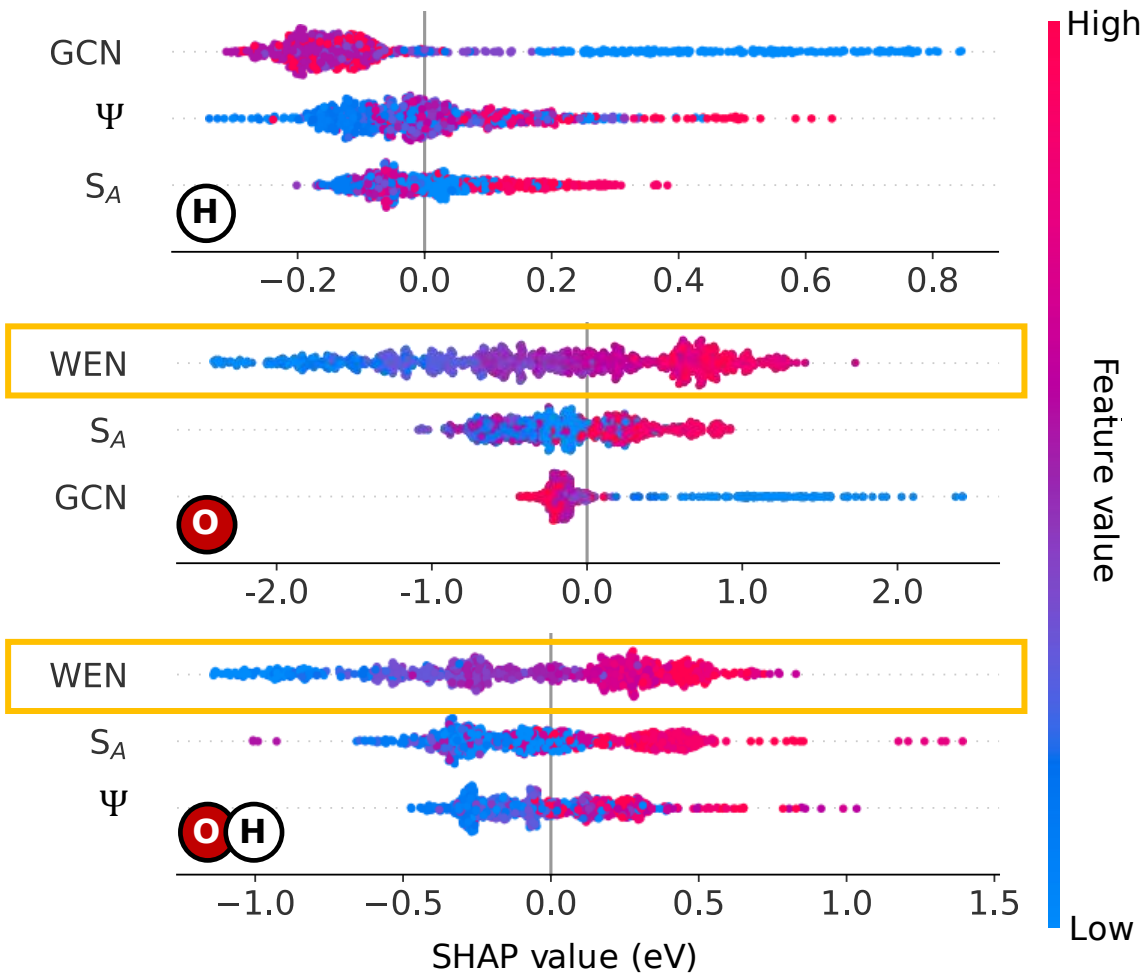
V. Vassilev-Galindo & J. LLorca. *Chem. Sci.* **17**, 1058-1072 (2026)

Model Validation



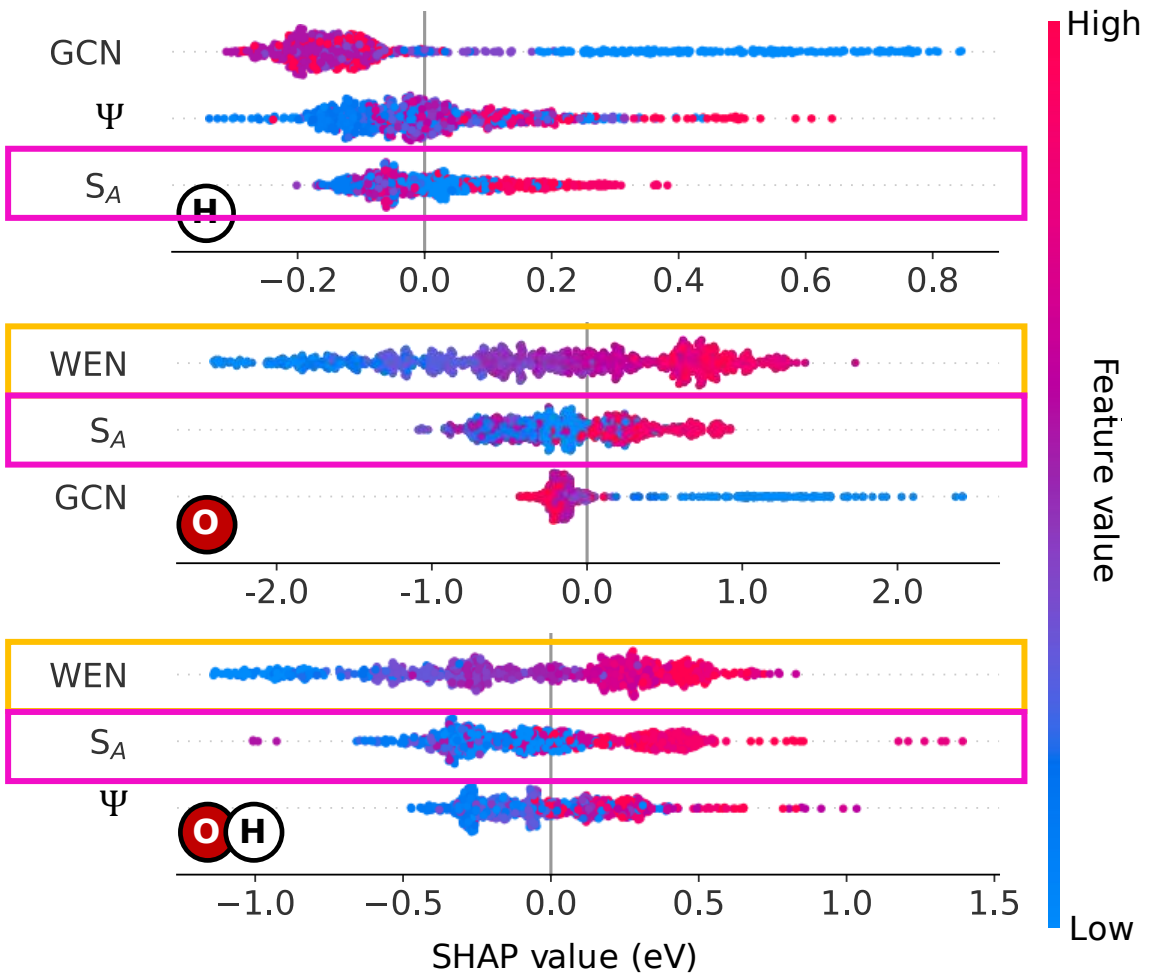
V. Vassilev-Galindo & J. LLorca. *Chem. Sci.* **17**, 1058-1072 (2026)

Model Validation



O is a very electronegative atom, it will tend to bind stronger to surfaces with low electronegativity.

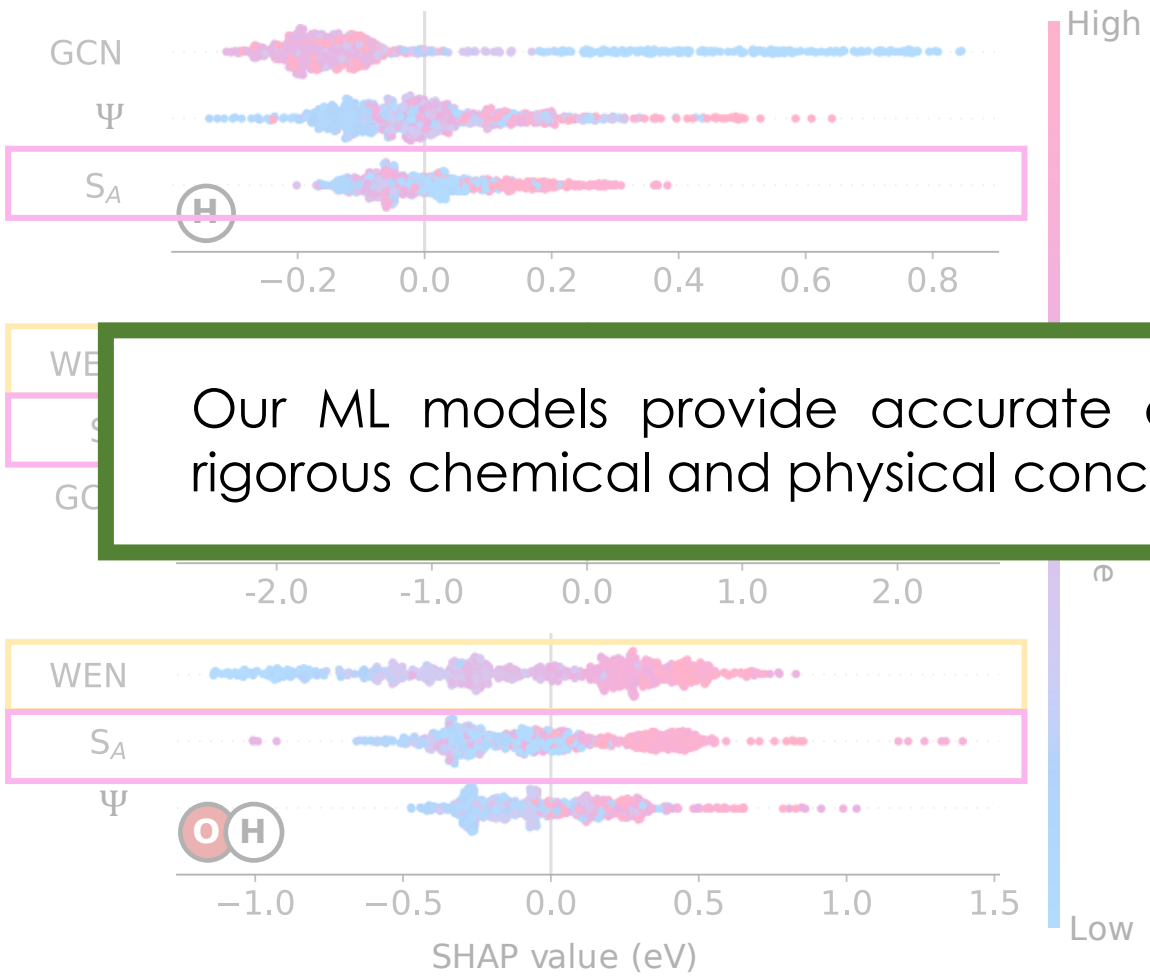
Model Validation



O is a very electronegative atom, it will tend to bind stronger to surfaces with low electronegativity.

A large number of valence electrons downshifts the d-band center. The more negative the d-band centers the less favorable the adsorption.

Model Validation



O is a very electronegative atom, it will tend to bind stronger to surfaces with low electronegativity.

Our ML models provide accurate and reliable predictions by leveraging rigorous chemical and physical concepts.

A large number of valence electrons downshifts the d-band center. The more negative the d-band centers the less favorable the adsorption.

V. Vassilev-Galindo & J. LLorca. *Chem. Sci.* **17**, 1058-1072 (2026)

Discovered Materials

Target E_{ads} H -0.49 eV O -1.79 eV OH -1.19 eV

Adsorbate	Material	Crystal	Facet	Site	ML E_{ads} (eV)
H	CaNi	Cubic	101	bridge-CaNi	-0.46
	YIr	Cubic	101	ontop-Ir	-0.45
	YAu	**Cubic**	**100**	**hollow**	**-0.47**
	YAu	Cubic	101	hollow	-0.53
O	**ZnIr**	**Hexagonal**	**10-11**	**hollow**	**-1.72**
	ZnPt	**Tetragonal**	**101**	**hollow**	**-1.76**
OH	**CdAu**	**Cubic**	**101**	**longbridge-Au**	**1.10**
	CdAu	Hexagonal	0001	bridge-Au	1.10

Discovered Materials

Target E_{ads} H -0.49 eV O -1.79 eV OH -1.19 eV

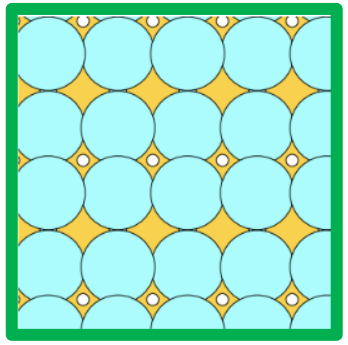
Adsorbate	Material	Crystal	Facet	Site	ML E_{ads} (eV)	DFT E_{ads} (eV)
H	CaNi	Cubic	101	bridge-CaNi	-0.46	-
	YIr	Cubic	101	ontop-Ir	-0.45	-0.68
	YAu	**Cubic**	**100**	**hollow**	**-0.47**	**-0.42**
	YAu	Cubic	101	hollow	-0.53	-
O	**ZnIr**	**Hexagonal**	**10-11**	**hollow**	**-1.72**	**-1.93**
	ZnPt	**Tetragonal**	**101**	**hollow**	**-1.76**	**-1.81**
OH	**CdAu**	**Cubic**	**101**	**longbridge-Au**	**1.10**	**1.06**
	CdAu	Hexagonal	0001	bridge-Au	1.10	0.89

Discovered Materials

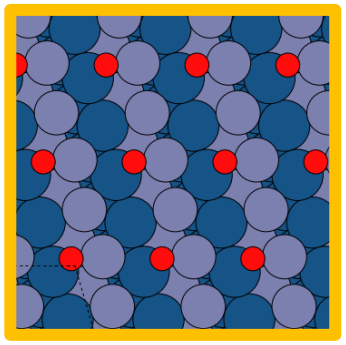
Target E_{ads} H -0.49 eV O -1.79 eV OH -1.19 eV

Adsorbate	Material	Crystal	Facet	Site	ML E_{ads} (eV)	DFT E_{ads} (eV)
H	CaNi	Cubic	101	bridge-CaNi	-0.46	-
	YIr	Cubic	101	ontop-Ir	-0.45	-0.68
	YAu	**Cubic**	**100**	**hollow**	**-0.47**	**-0.42**
	YAu	Cubic	101	hollow	-0.53	-
O	**ZnIr**	**Hexagonal**	**10-11**	**hollow**	**-1.72**	**-1.93**
	ZnPt	**Tetragonal**	**101**	**hollow**	**-1.76**	**-1.81**
OH	**CdAu**	**Cubic**	**101**	**longbridge-Au**	**1.10**	**1.06**
	CdAu	Hexagonal	0001	bridge-Au	1.10	0.89

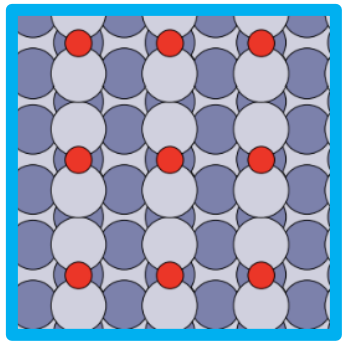
YAu bcc(100)



ZnIr hcp(10-11)

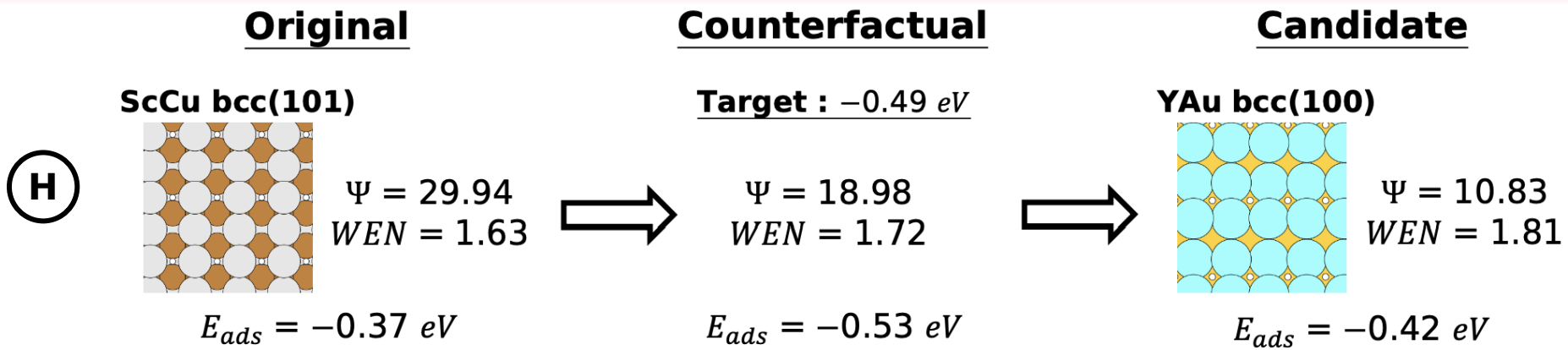


ZnPt bct(101)



V. Vassilev-Galindo & J. LLorca. *Chem. Sci.* **17**, 1058-1072 (2026)

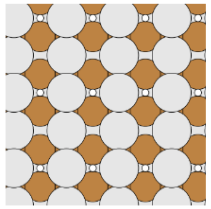
Retrieving Explanations



Retrieving Explanations

Original

ScCu bcc(101)



$\Psi = 29.94$
$WEN = 1.63$

$E_{ads} = -0.37\ eV$

Counterfactual

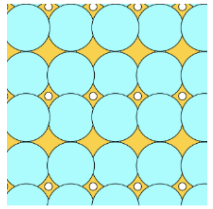
Target : $-0.49\ eV$

$\Psi = 18.98$
$WEN = 1.72$

$E_{ads} = -0.53\ eV$

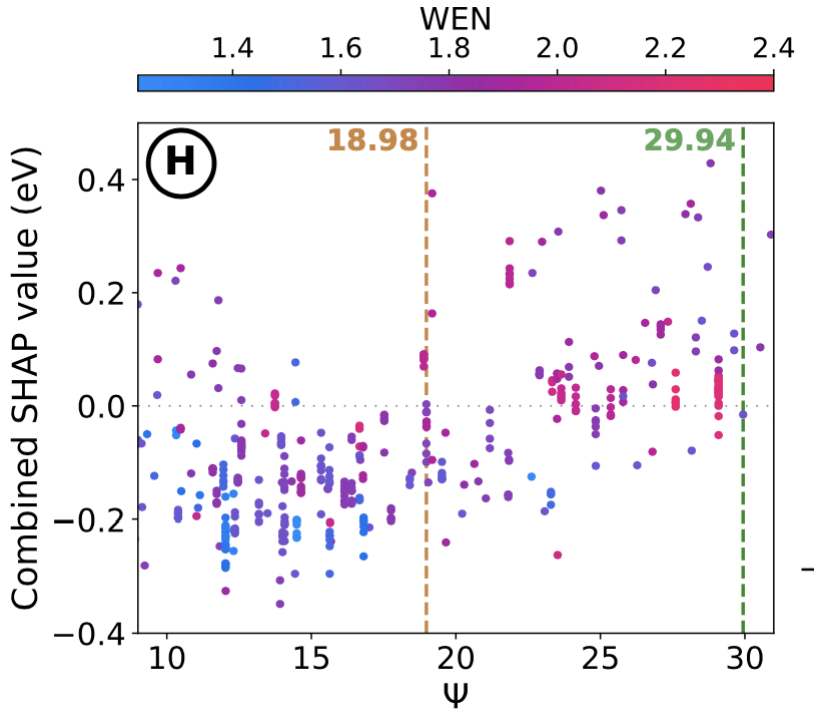
Candidate

YAu bcc(100)



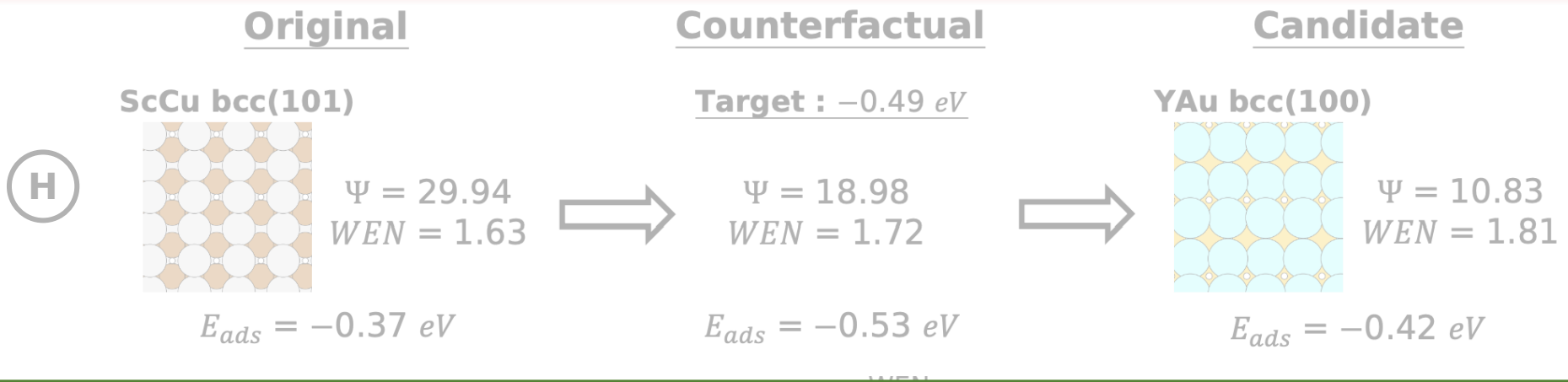
$\Psi = 10.83$
$WEN = 1.81$

$E_{ads} = -0.42\ eV$

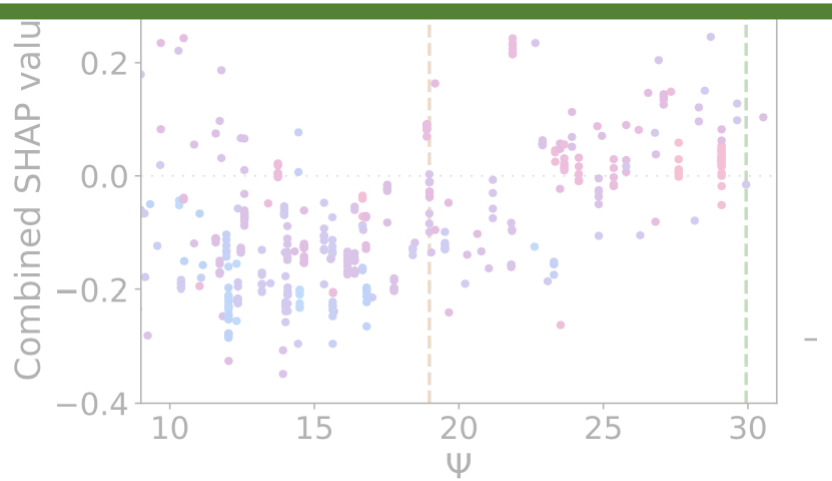


V. Vassilev-Galindo & J. LLorca. *Chem. Sci.* **17**, 1058-1072 (2026)

Retrieving Explanations

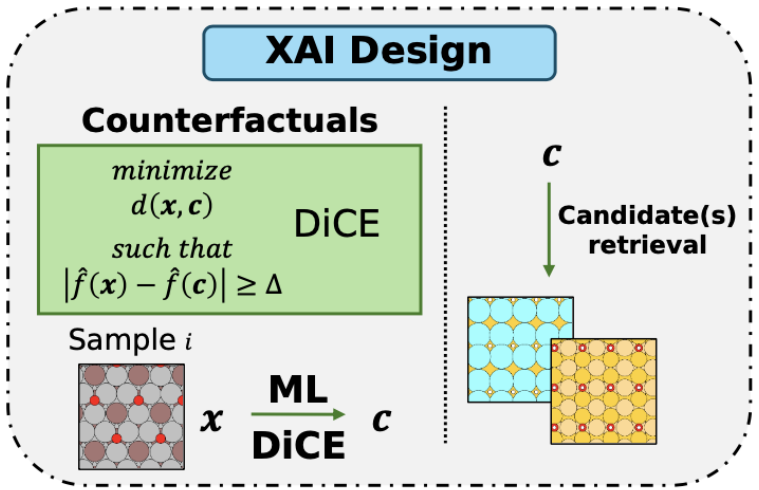


Counterfactuals unveiled subtle relationships between the most relevant features, other, in principle, less important features, and the E_{ads}.



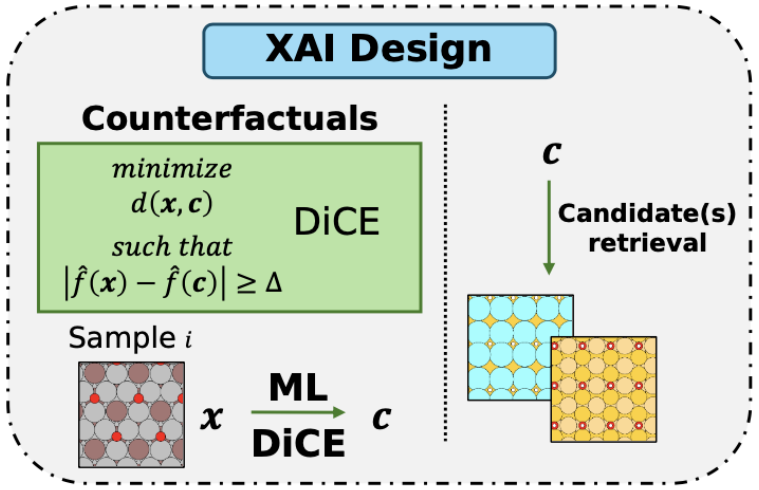
Conclusions

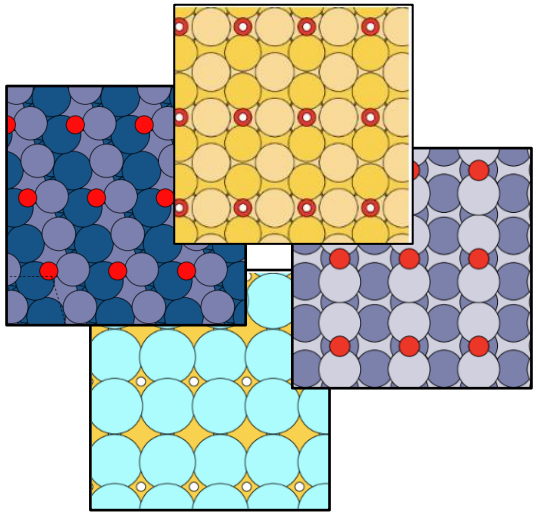
- A **novel strategy** for the **discovery and design of new materials** based on **counterfactual explanations** was proposed



Conclusions

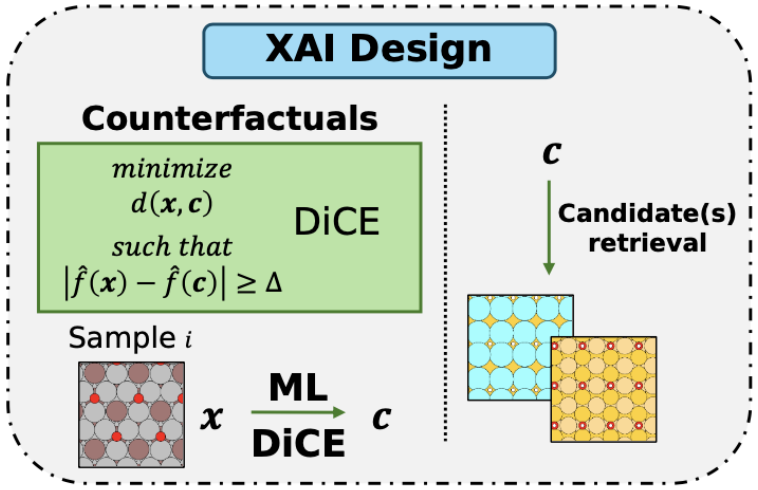
- A **novel strategy** for the **discovery and design of new materials** based on **counterfactual explanations** was proposed
- **Four** of the **discovered materials** were **confirmed** by reference **DFT calculations, validating** our **strategy**

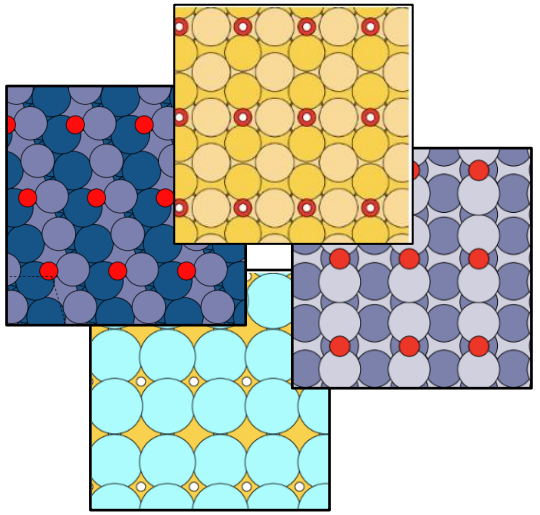


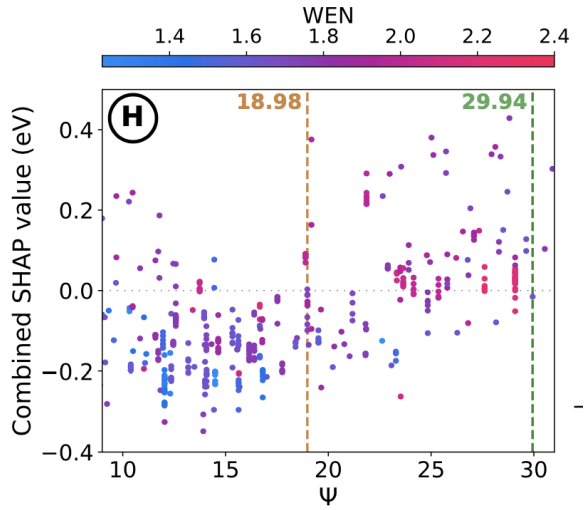


Conclusions

- A **novel strategy** for the **discovery and design of new materials** based on **counterfactual explanations** was proposed
- **Four** of the **discovered materials** were **confirmed** by reference **DFT calculations, validating** our **strategy**
- By **comparing original samples**, **counterfactuals**, and **discovered candidates**, subtle **relationships** were **unveiled** between the most relevant features, other, in principle, less important features, and the E_{ads}







Acknowledgements

XAI will not replace human experts, but human experts who embrace XAI will replace those who do not

Oviedo, F., *et al*. *Acc. Mater. Res.* **3**, 597-607 (2022).

Acknowledgements

XAI will not replace human experts, but human experts who embrace XAI will replace those who do not

Oviedo, F., *et al*. *Acc. Mater. Res.* **3**, 597-607 (2022).

Check the paper here!

Python Routines

GitHub

Acknowledgements

XAI will not replace human experts, but human experts who embrace XAI will replace those who do not

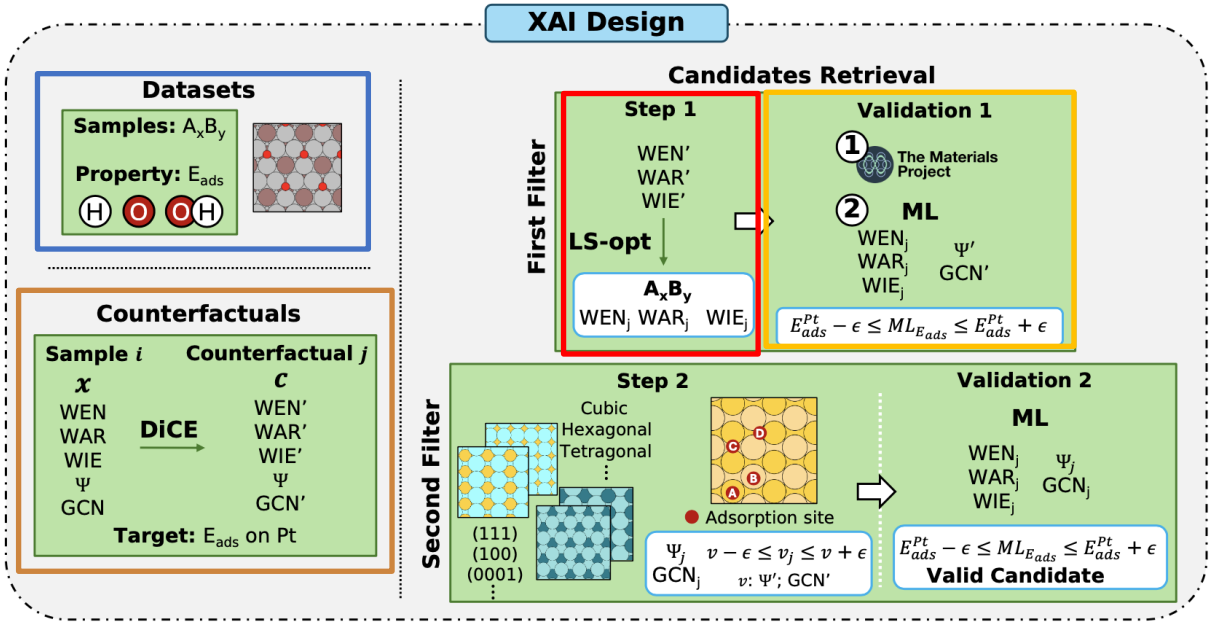
Oviedo, F., *et al*. *Acc. Mater. Res.* **3**, 597-607 (2022).

Check the paper here!

Universidad de Navarra | DATAI INSTITUTE OF DATA SCIENCE AND ARTIFICIAL INTELLIGENCE

Python Routines

GitHub



We start with our databases of E_{ads} for H, O, and OH

We generate the counterfactuals with DICE for each sample in the database

Step 1 (example)

Sample *i* : ZnTi

A_xB_y

Zn_xTi_y Zn_xZr_y Zn_xHf_y

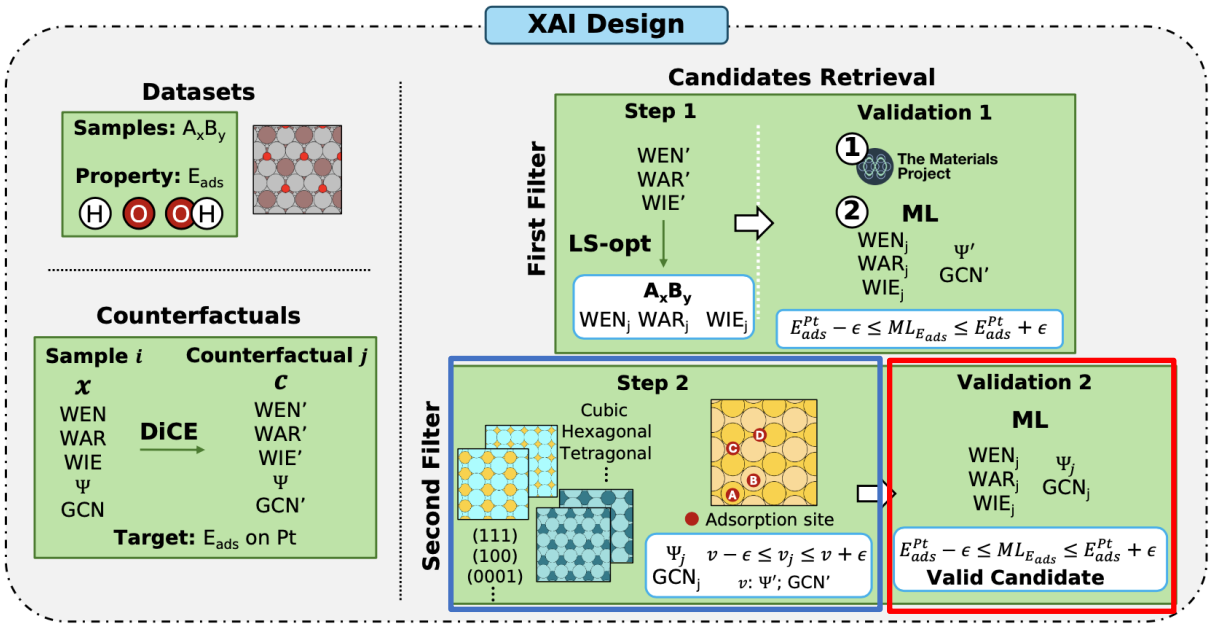
Cd_xTi_y Cd_xZr_y Cd_xHf_y

$$EN_A X + EN_B Y = WEN'$$
$$AR_A X + AR_B Y = WAR'$$
$$IE_A X + IE_B Y = WIE'$$

Least-squares optimization

WEN_j WAR_j WIE_j

- Check that the A_xB_y intermetallics found in Step 1 exist in Materials Project
- Use ML model to verify that the E_{ads} of these intermetallics is still close to the target



- All possible crystal lattices are considered and all surface slabs for different ($hk(i)l$) facets are constructed ($h,k,l \leq 3$)
- All available adsorption sites are determined and values of Ψ' and GCN' are updated accordingly (Ψ_j, GCN_j)
- We only keep those samples for which Ψ_j and GCN_j are close to the counterfactual.

- Use ML model to verify that the E_{ads} of these intermetallics is still close to the target

- Explainability is ensured by construction since the discovered materials can be linked directly to the original sample of the dataset from which the counterfactual was generated.