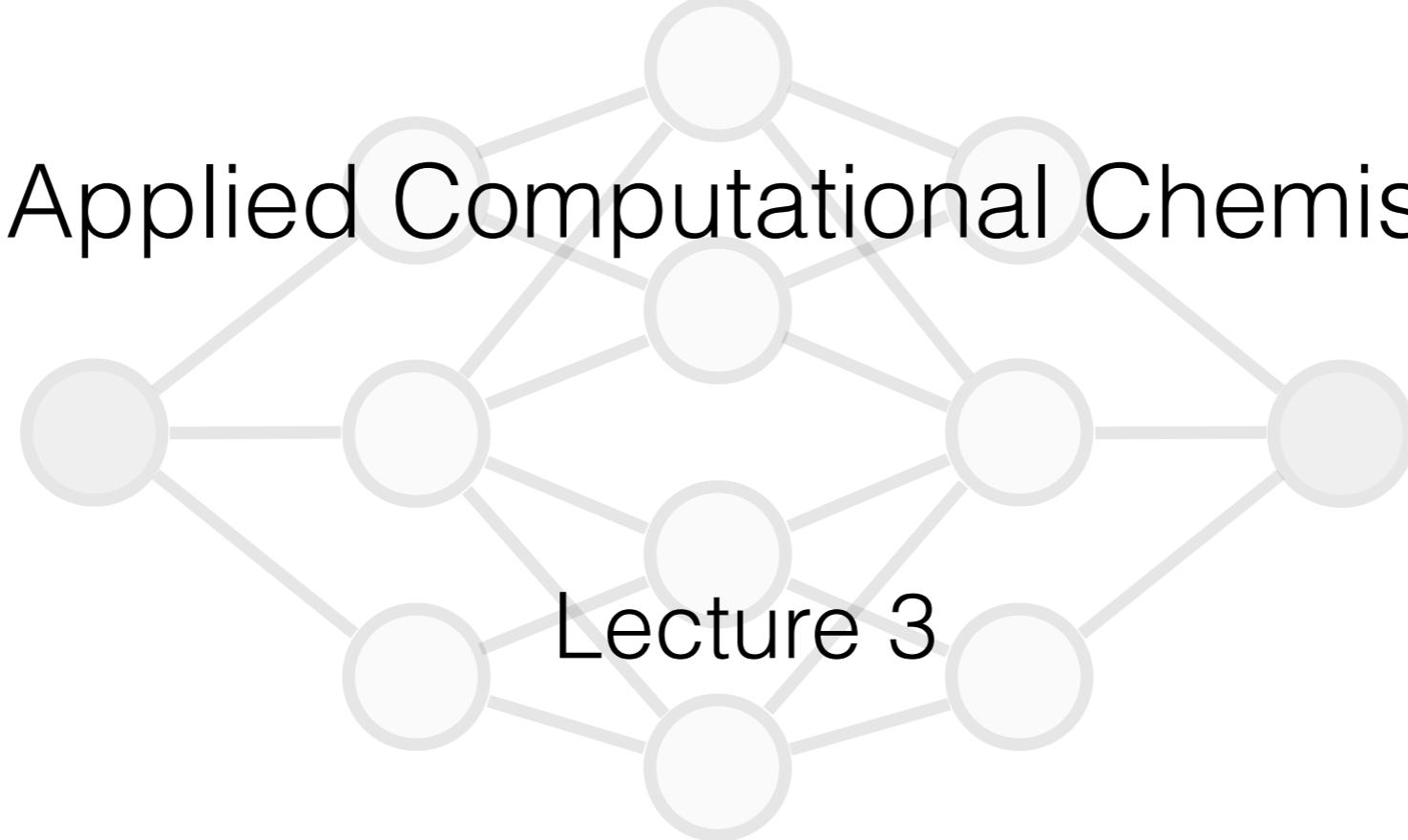




# Applied Computational Chemistry



## Lecture 3

Fernanda Duarte

(fernanda.duartegonzalez@chem.ox.ac.uk)



Twitter @duarte\_group

[duartegroupchem.org](http://duartegroupchem.org)



Github: duartegroup

# How do we use it?

Which System Do I Have?



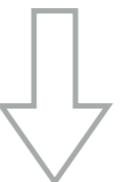
What Do You Want to Compute (and Why)?



Which Model /Method Should I Choose?



Verify Approach (vs. Experiment)

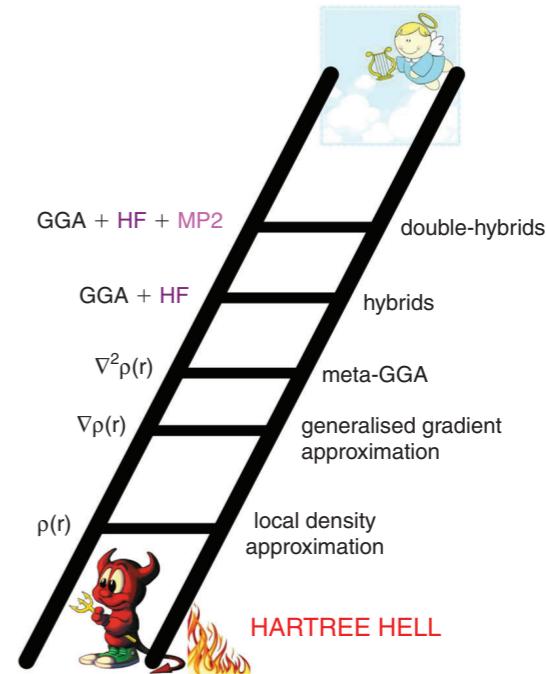


**Interpret/Analyse**

# Quiz

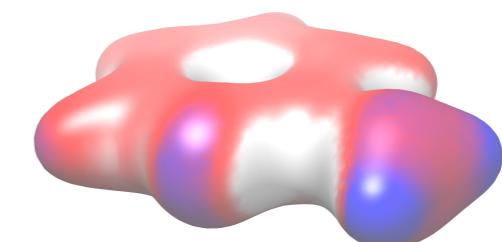
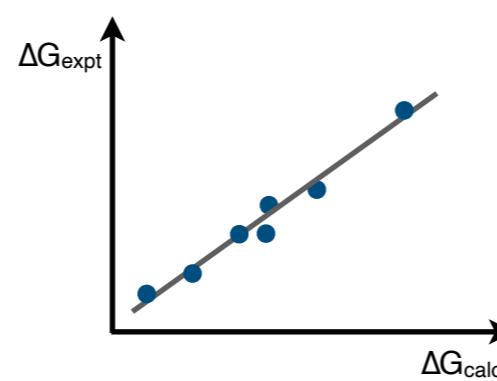
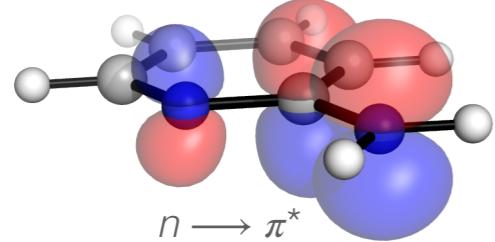
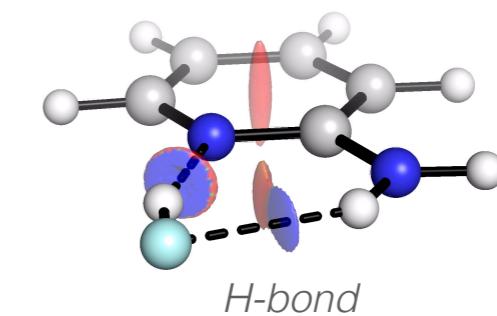
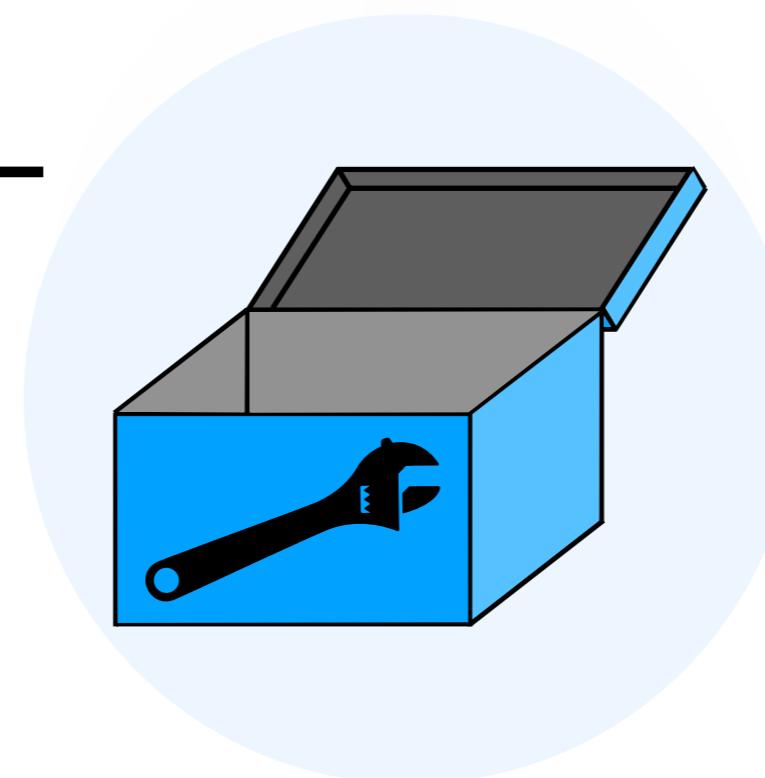
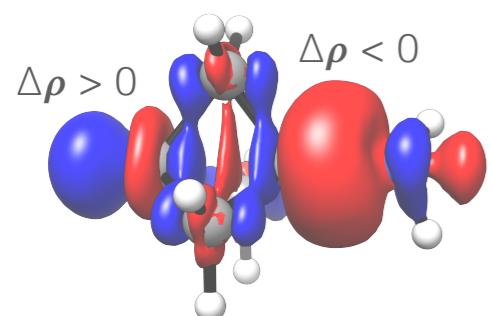
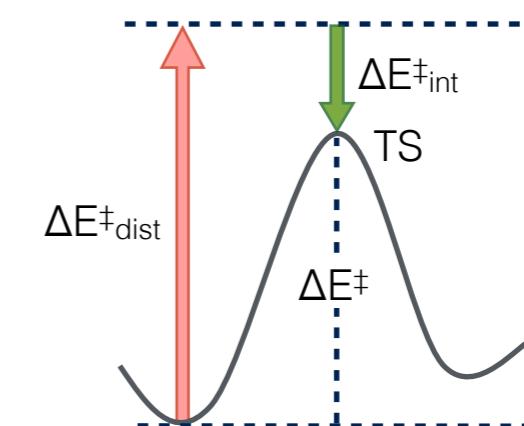
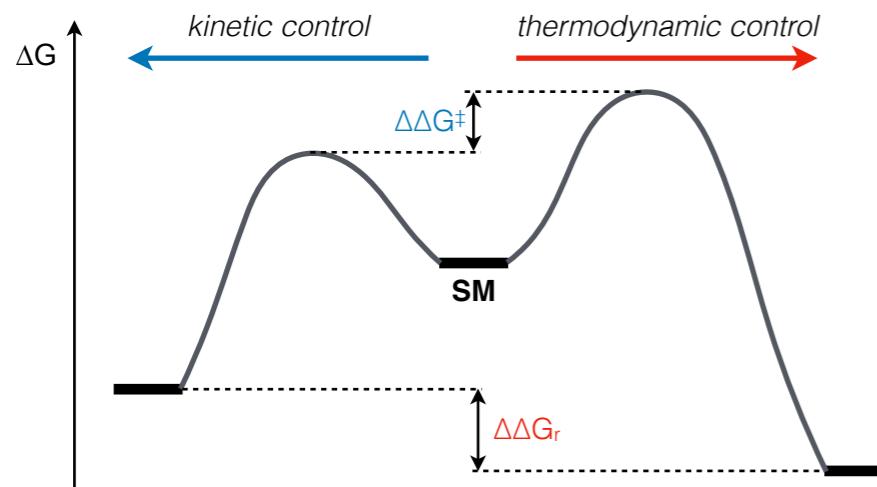
*A Trip to the Density Functional Theory Zoo: Warnings and Recommendations for the User*  
*Aust. J. Chem.* **2019**, *72*, 563  
<https://doi.org/10.1071/CH19023>

HEAVEN OF CHEMICAL ACCURACY

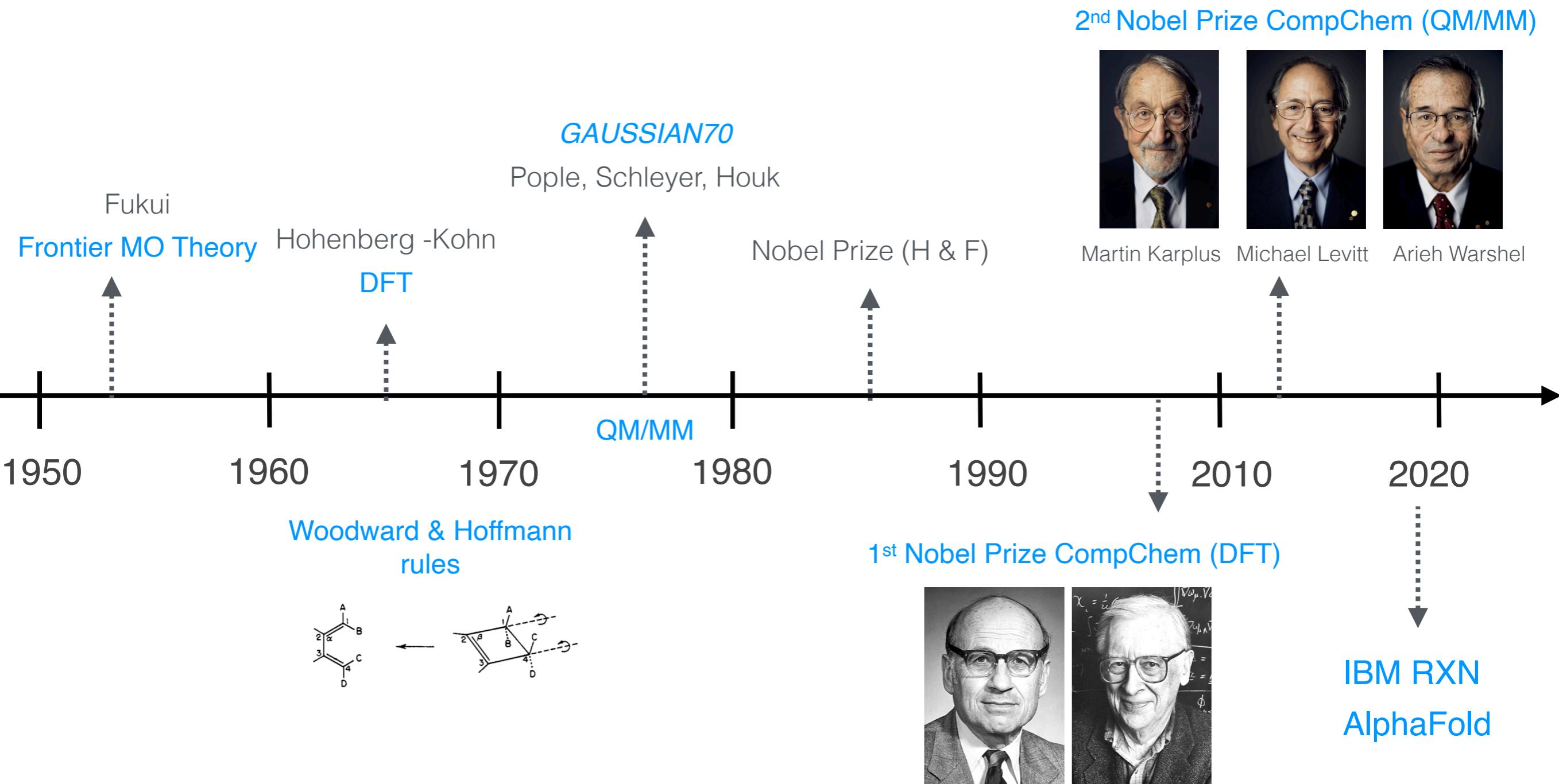


1. How are DFAs classified?
2. What is understood by chemical accuracy?
3. Describe with examples known inaccuracies in DFT
4. How can dispersion be accounted for?
5. What is BSSE and what is its effect on structural and chemical properties?
6. Best phrase of the paper?

# Computational Organic Chemistry



# Computational Organic Chemistry



K. Fukui, et al. *J. Chem. Phys.* 1952, 20, 722

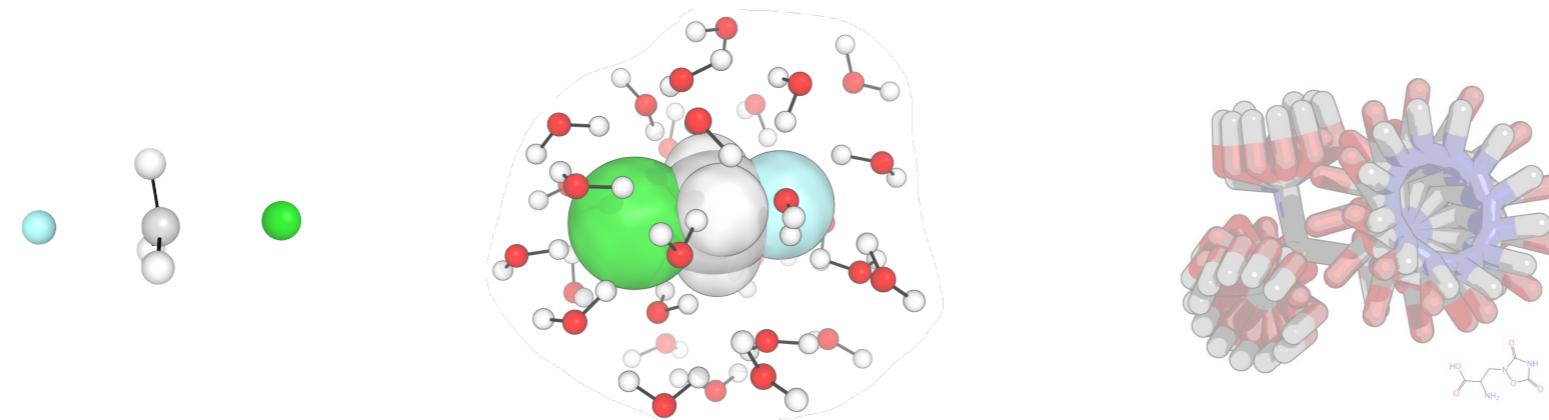
P. Hohenberg and W. Kohn *Phys. Rev. B* 1964, 136, 864

R. Hoffmann and R. B. Woodward, *J. Am. Chem. Soc.* 1965, 87, 395; 1965, 87, 2046; 1965, 87, 2511

Demis Hassabis et al. *Nature*, 2021 596, 583

# Six Grand Challenges

- Robust and fast **electronic structure methods** for all conceivable chemical processes
- Automated **multilevel modelling**, including error estimates
- Automated approaches for finding **new reactions**
- Inclusion of **solvation effects**
- Accurate treatments of **entropy**
- Prediction of molecular and **macroscopic** properties



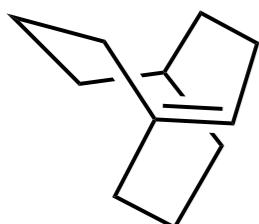
# Interpretability

Allinger, Osawa, and Schleyer popularised the use of FFs in organic chemistry



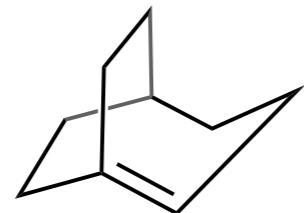
Olefin Strain Energy (OS) using MM1/MM2 FFs

**Isolable**  
OS < 17 kcal/mol



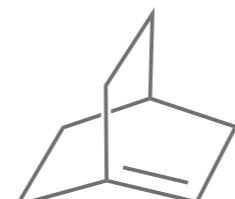
Bicyclo [3.3.3]  
undec-1-ene

**Observable**  
17-21 kcal/mol



Bicyclo [3.2.2]  
non-1-ene

**Unstable**  
>21 kcal/mol

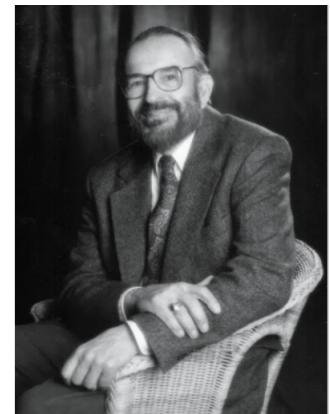


Bicyclo [2.2.2]  
oct-1-ene

*J. Am. Chem. Soc. 1981, 103, 8, 1891*

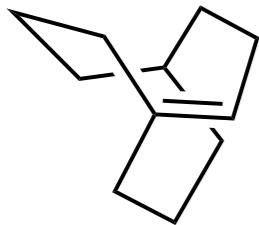
# Interpretability

Allinger, Osawa, and Schleyer popularised the use of FFs in organic chemistry



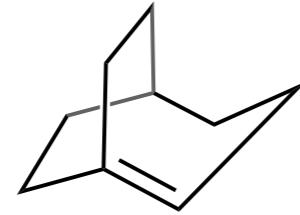
Olefin Strain Energy (OS) using MM1/MM2 FFs

**Isolable**  
OS < 17 kcal/mol



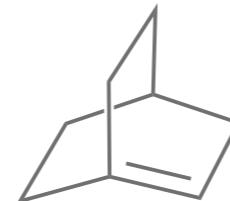
Bicyclo [3.3.3]  
undec-1-ene

**Observable**  
17-21 kcal/mol



Bicyclo [3.2.2]  
non-1-ene

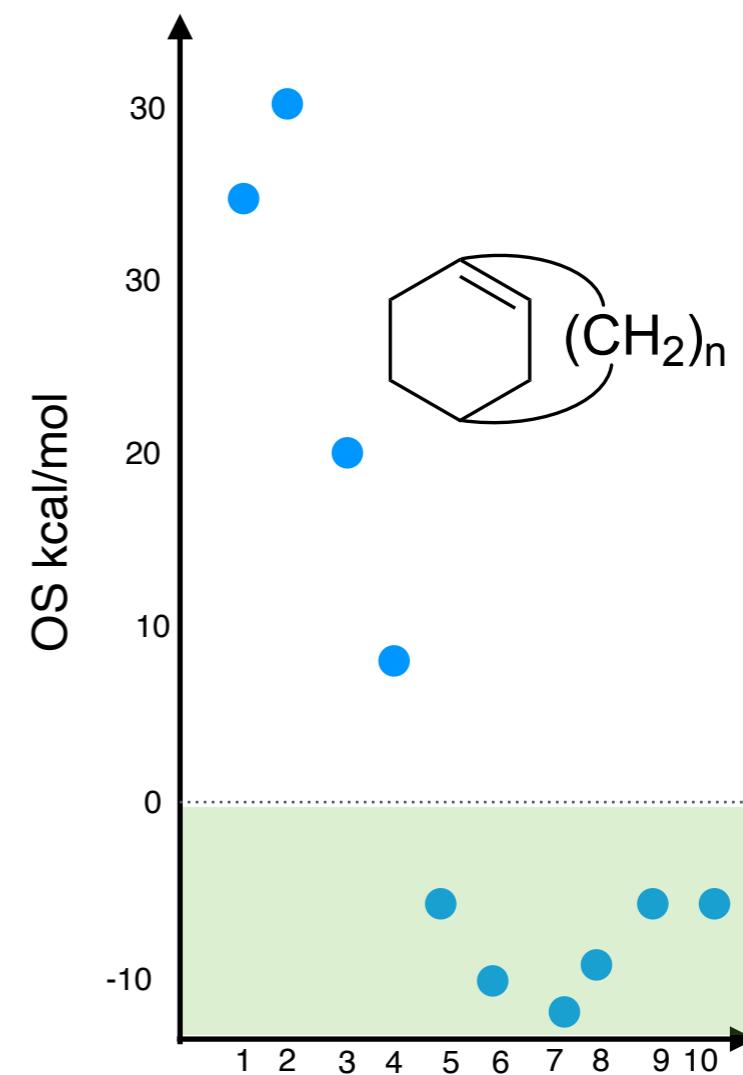
**Unstable**  
>21 kcal/mol



Bicyclo [2.2.2]  
oct-1-ene

*J. Am. Chem. Soc. 1981, 103, 8, 1891*

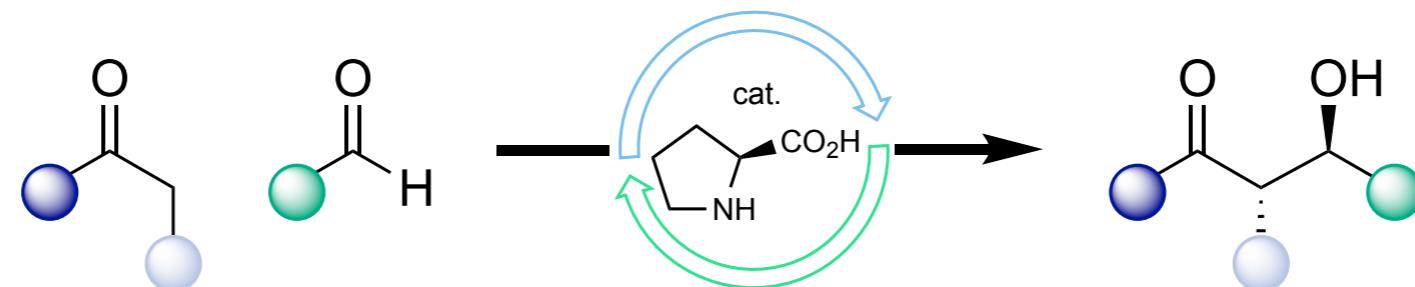
Hyperstabilised olefins  
Predicting new molecules



*J. Am. Chem. Soc. 1986, 108, 14, 3953*

# Interpretability

## Organocatalysis – Rationalising Selectivity Houk–List model



L.Hoang, S. Bahmanyar, KN Houk, B. List *J Am Chem Soc.* 2003, 125, 16

S. Bahmanyar, KN Houk, HJ Martin, B. List *J Am Chem Soc.* 2003, 125, 2475

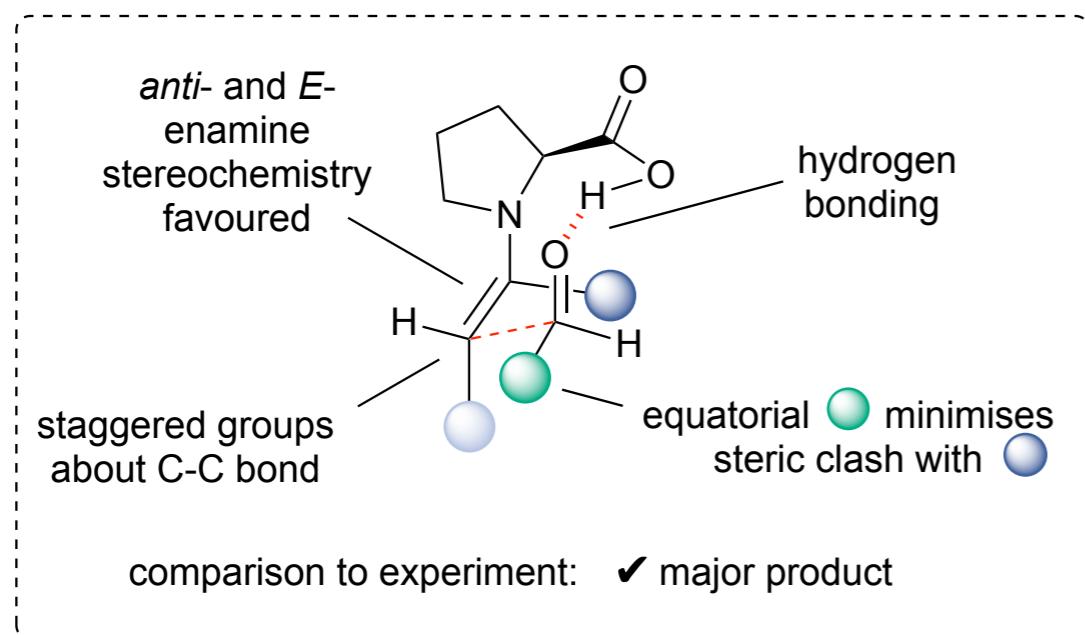
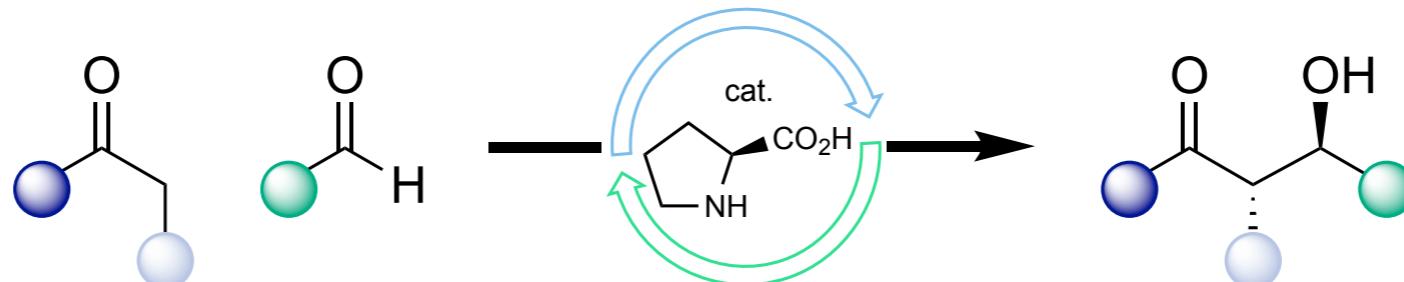
N. Zotova, A. Franzke, A. Armstrong, DG Blackmond. *J Am Chem Soc.* 2007;129,100

A. Armstrong, RA Boto, P. Dingwall *et al. Chem Sci.* 2014, 5, 2057

Orlandi M, Ceotto M, Benaglia M. *Chem Sci.* 2016, 7, 5421

# Interpretability

## Organocatalysis – Rationalising Selectivity Houk–List model



L.Hoang, S. Bahmanyar, KN Houk, B. List *J Am Chem Soc.* 2003, 125, 16

S. Bahmanyar, KN Houk, HJ Martin, B. List *J Am Chem Soc.* 2003, 125, 2475

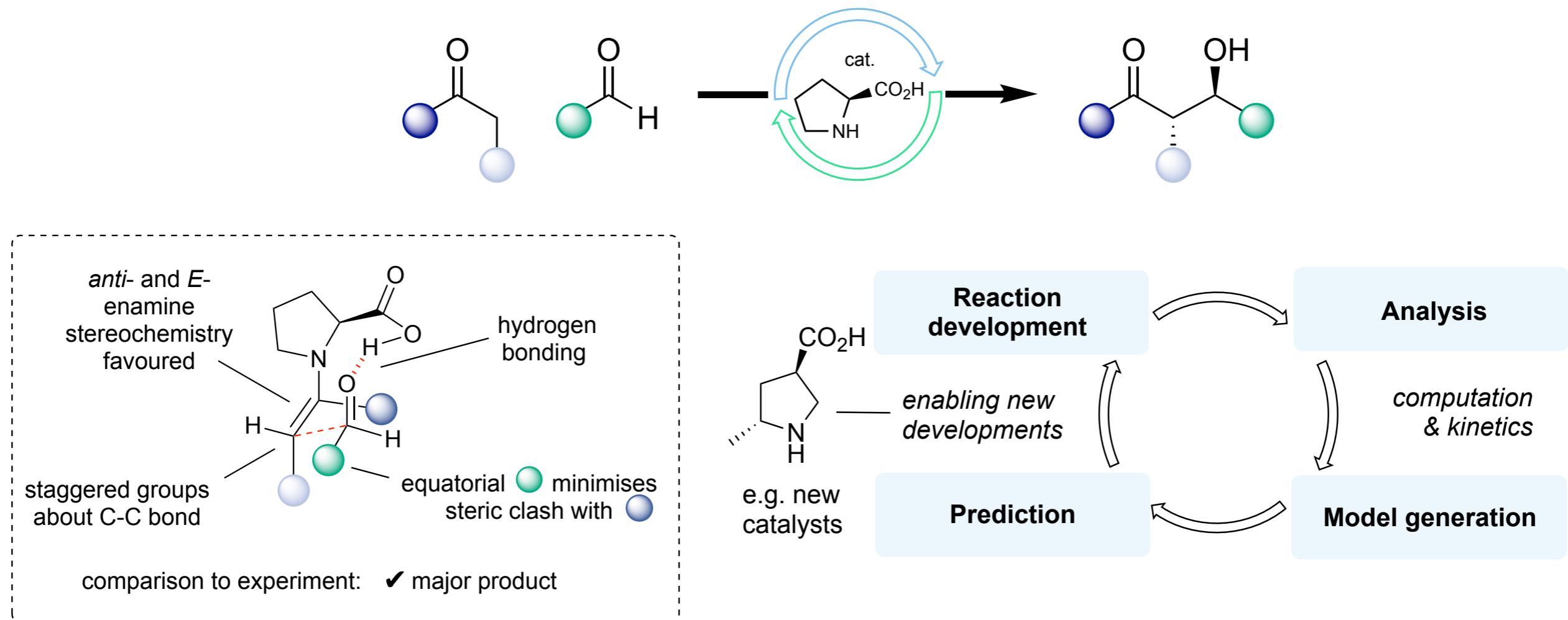
N. Zotova, A. Franzke, A. Armstrong, DG Blackmond. *J Am Chem Soc.* 2007;129,100

A. Armstrong, RA Boto, P. Dingwall *et al. Chem Sci.* 2014, 5, 2057

Orlandi M, Ceotto M, Benaglia M. *Chem Sci.* 2016, 7, 5421

# Interpretability

## Organocatalysis – Rationalising Selectivity Houk–List model



L.Hoang, S. Bahmanyar, KN Houk, B. List *J Am Chem Soc.* 2003, 125, 16

S. Bahmanyar, KN Houk, HJ Martin, B. List *J Am Chem Soc.* 2003, 125, 2475

N. Zotova, A. Franzke, A. Armstrong, DG Blackmond. *J Am Chem Soc.* 2007;129,100

A. Armstrong, RA Boto, P. Dingwall *et al. Chem Sci.* 2014, 5, 2057

Orlandi M, Ceotto M, Benaglia M. *Chem Sci.* 2016, 7, 5421

# Accuracy – Thermochemistry

Inner energy

$$\mathbf{U} = E_{el} + E_{ZPE} + E_{vib} + E_{rot} + E_{trans}$$

$$E_{el} = E(T) + E(Ne) + E(ee) + E(NN)$$

Enthalpy

$$\mathbf{H} = U + k_B T$$

Entropy

$$\mathbf{S} = T (S_{el} S_{vib} + S_{rot} + S_{trans})$$

$$\Delta G = \Delta H - T\Delta S$$

Standard state correction – Important when there is change in molecularly

# Accuracy – Thermochemistry

$$\Delta G = \Delta H - T\Delta S$$

-----  
THERMOCHEMISTRY AT 298.15K  
-----

Temperature	... 298.15 K
Pressure	... 1.00 atm
Total Mass	... 34.01 AMU

Throughout the following assumptions are being made:

- (1) The electronic state is orbitally nondegenerate
- (2) There are no thermally accessible electronically excited states
- (3) Hindered rotations indicated by low frequency modes are not treated as such but are treated as vibrations and this may cause some error
- (4) All equations used are the standard statistical mechanics equations for an ideal gas
- (5) All vibrations are strictly harmonic

# Accuracy – kinetics

Kinetics

$$k = \frac{k_B T}{h} e^{-\Delta G^\ddagger / RT}$$

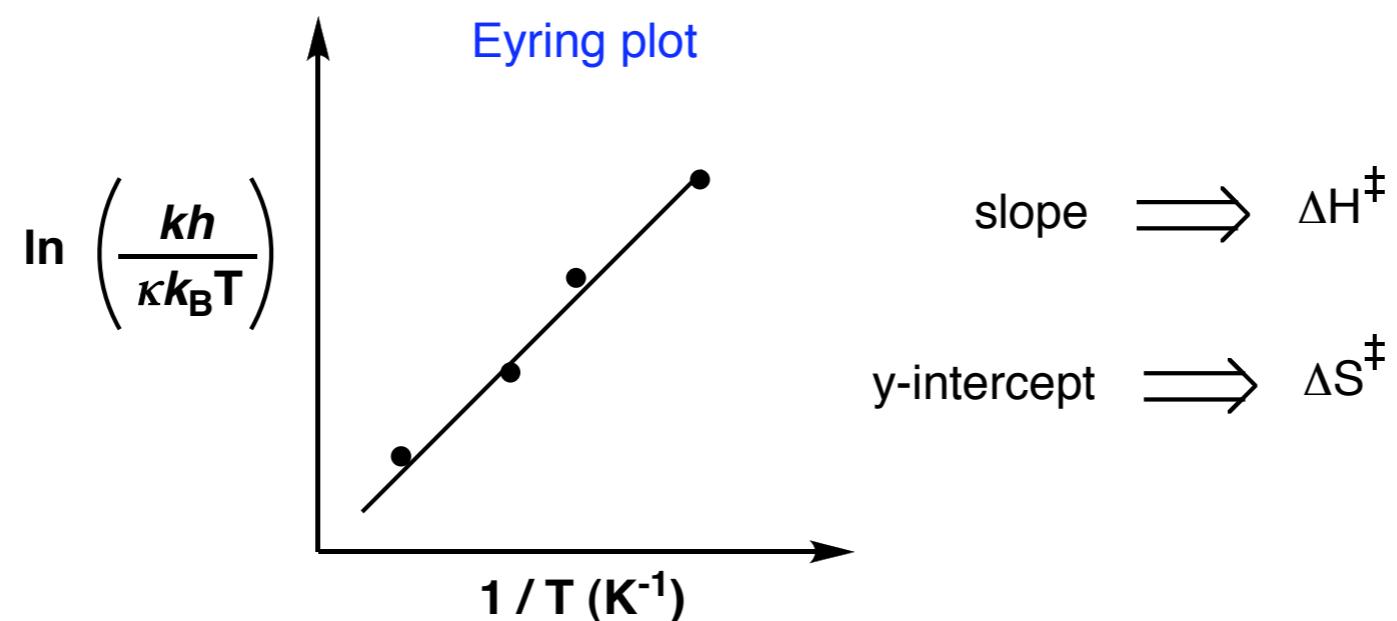
$\Delta G^\ddagger$ (kcal mol <sup>-1</sup> )	$k$ (s <sup>-1</sup> )	$t_{1/2}$ (s <sup>-1</sup> )	$t_{1/2}$
12	$9.8 \times 10^3$	$7.1 \times 10^{-5}$	70.5 μs
17	2.11	$3.3 \times 10^{-1}$	327 ms
22	$4.5 \times 10^{-4}$	$1.5 \times 10^3$	25 min
27	$9.8 \times 10^{-8}$	$7.1 \times 10^6$	81.1 days
30	$6.2 \times 10^{-10}$	$2.4 \times 10^{-6}$	35.5 years

$$T(K) = 298 ; R (\text{kcal mol}^{-1}) = 0.00831 ; R (\text{kJ K}^{-1} \text{mol}^{-1}) = 8.3144598 \times 10^{-3}$$
$$h(\text{J s}) = 6.6262 \times 10^{-34} \quad k_b(\text{J/K}) = 1.3807 \times 10^{-23}$$

# Accuracy – kinetics

## Experimental Determinations of Activation Parameters

$$\Delta G^\ddagger = \Delta H^\ddagger - T\Delta S$$



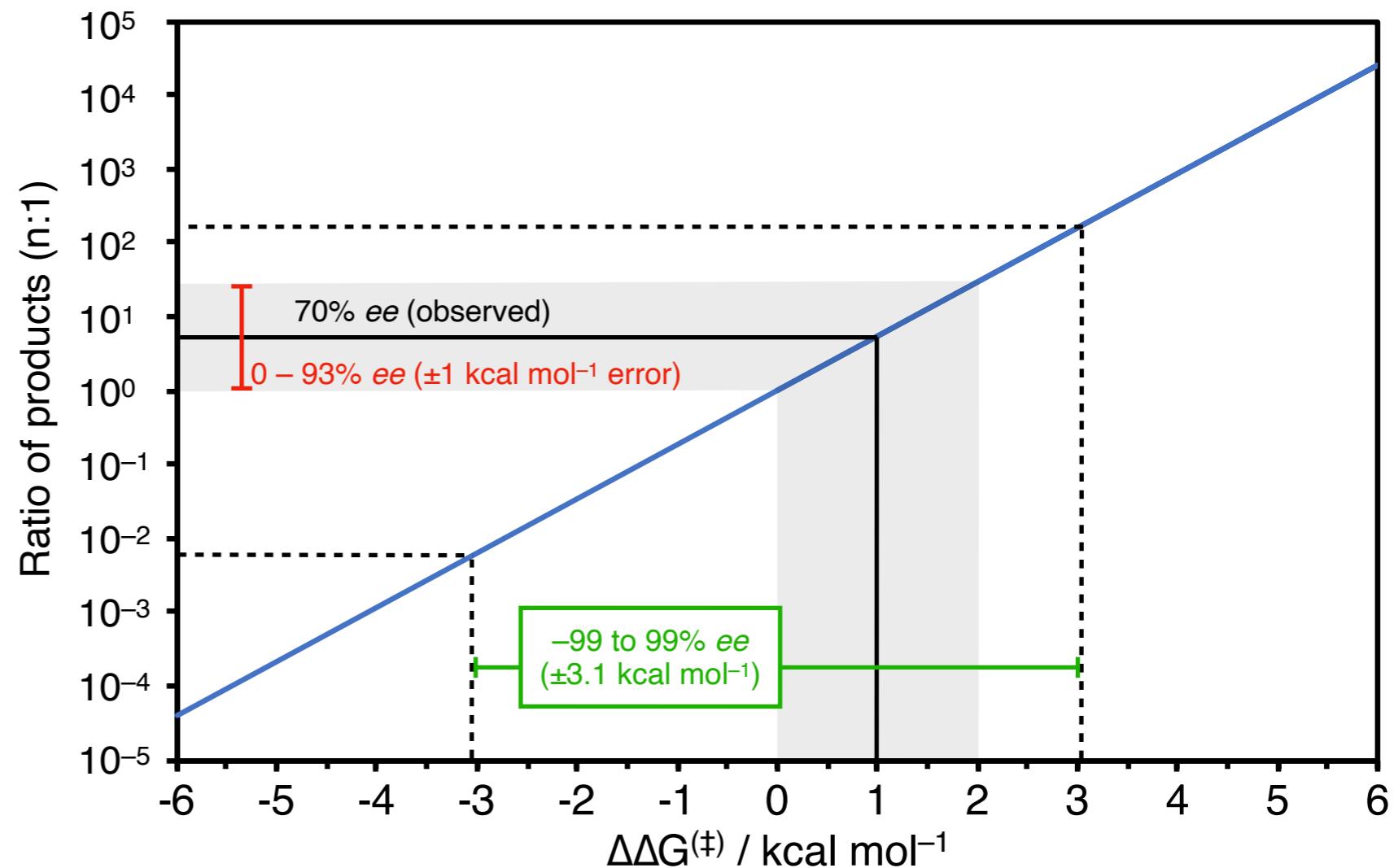
**Entropy**: energy associated with conformation, bond strength, vibrational states and how changes in these properties affect the overall energy of the system.

**Enthalpy**: energy related to the height of the surface

# Accuracy – enantioselectivity

Computational accuracy – Predicting enantiomeric excess (ee)

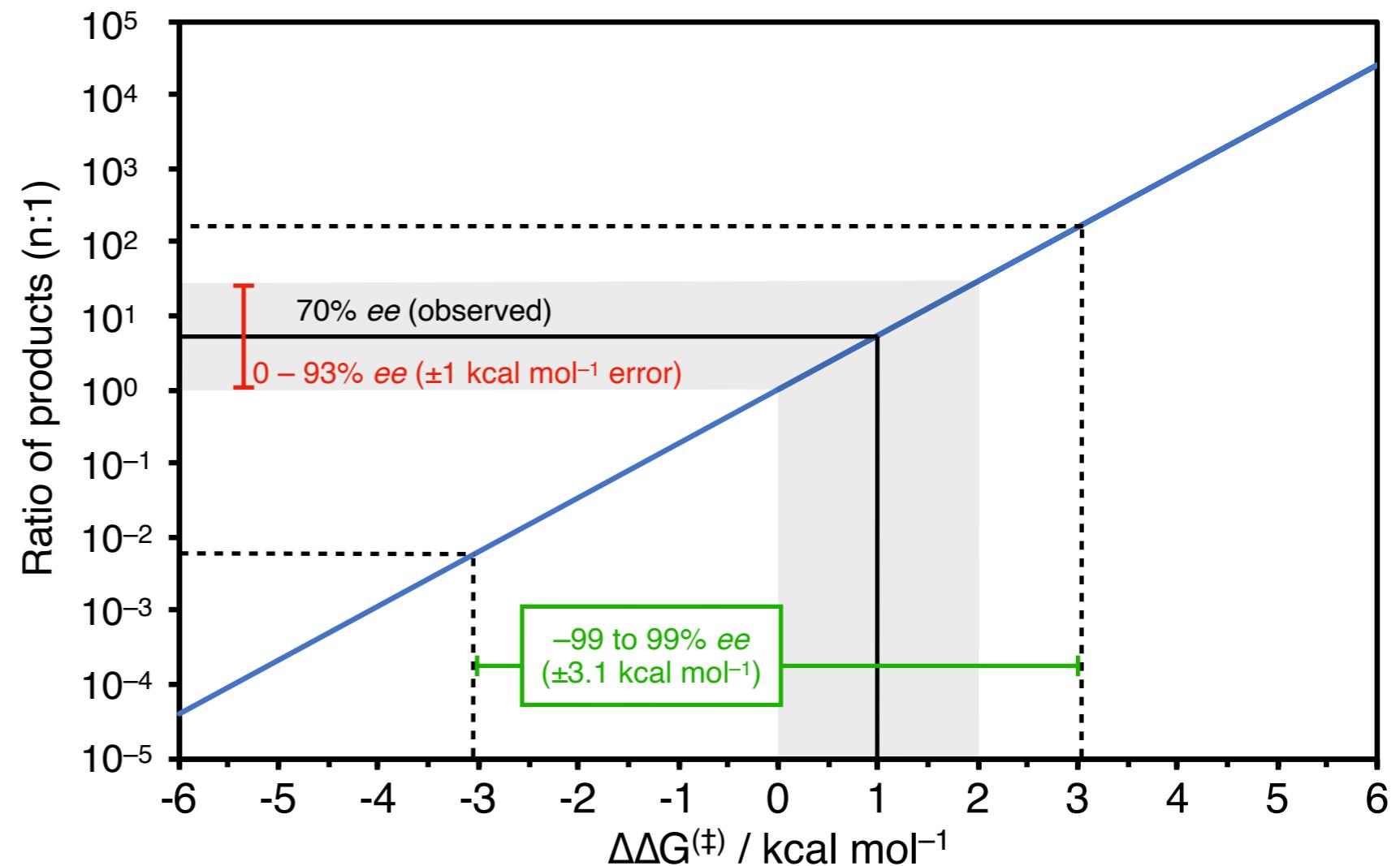
$$\frac{[A]}{[B]} = \frac{k_1}{k_2} = e^{-\Delta\Delta G^\ddagger / RT}$$



# Accuracy – enantioselectivity

Computational accuracy – Predicting enantiomeric excess (ee)

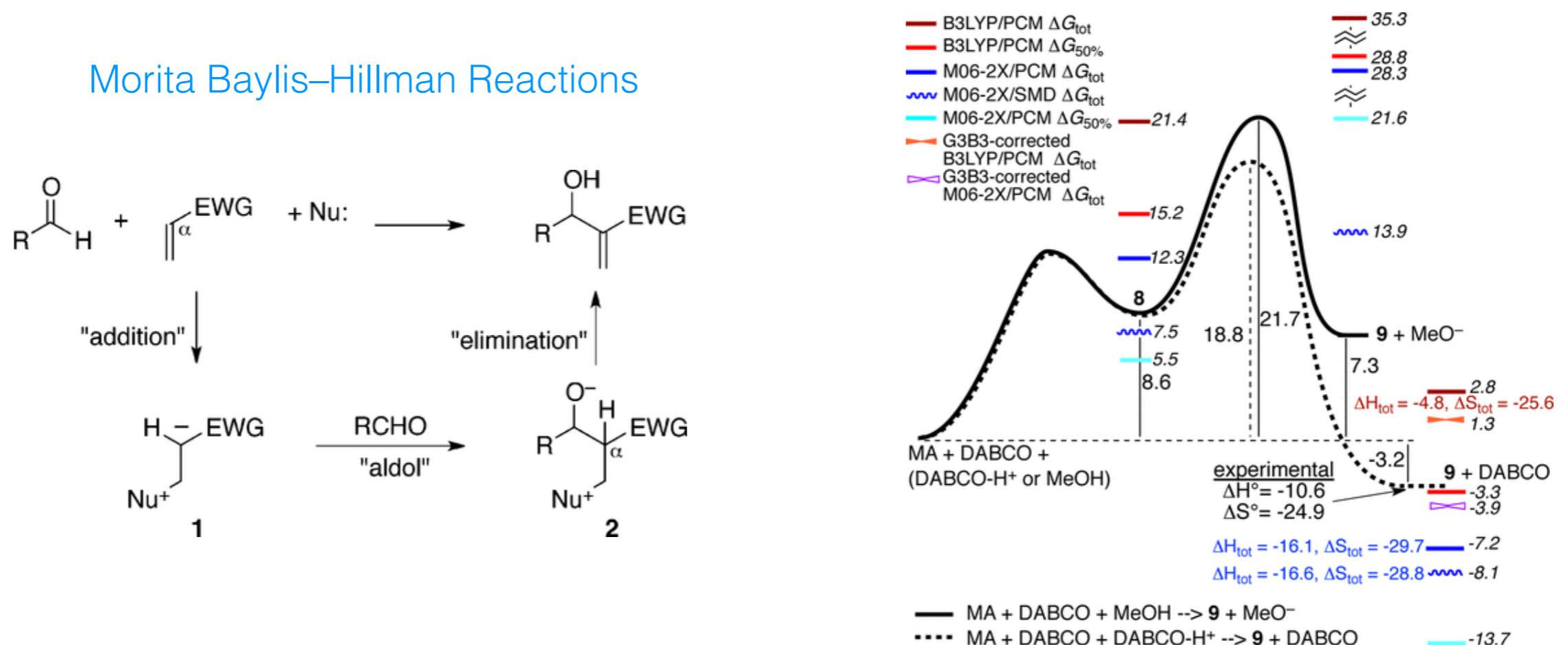
$$\frac{[A]}{[B]} = \frac{k_1}{k_2} = e^{-\Delta\Delta G^\ddagger / RT}$$



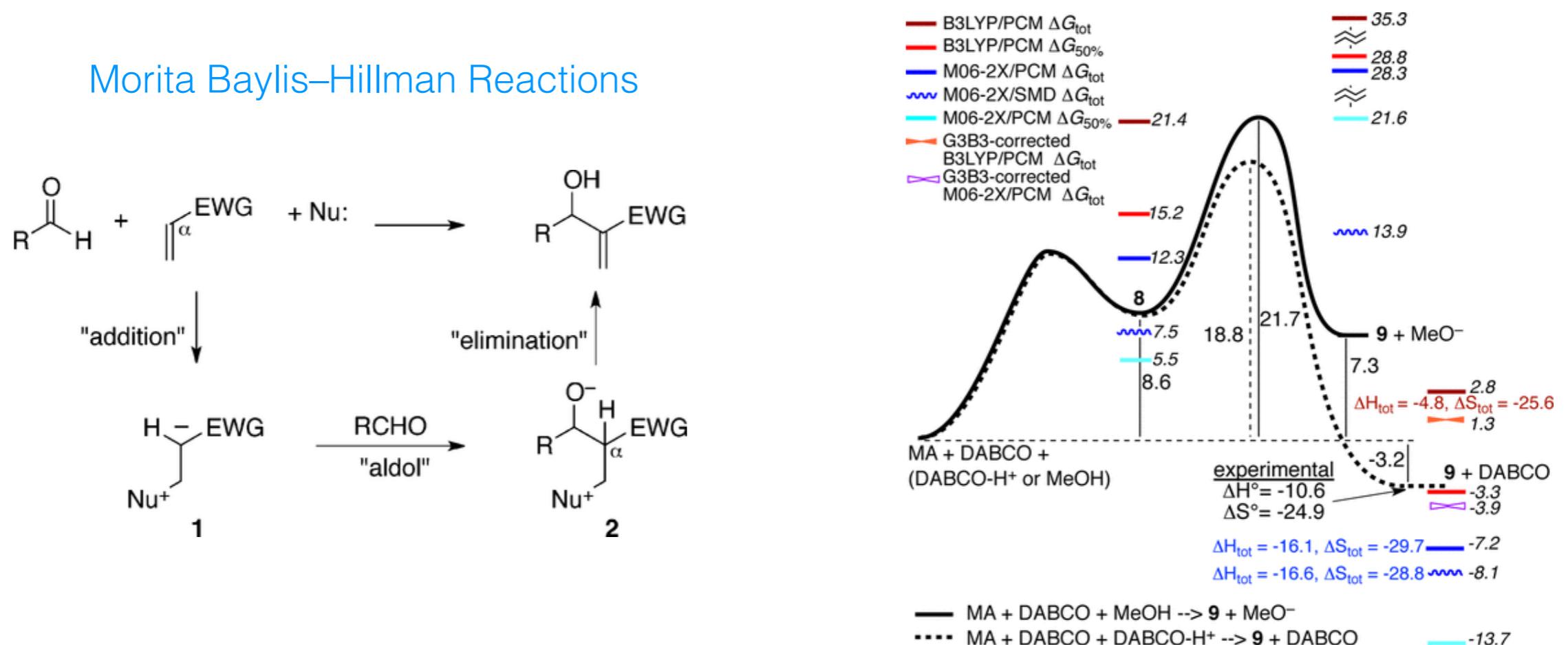
Current challenges

- Entropy • Thermal effects • Solvation • Conformational sampling

# Accuracy – Current Limitations

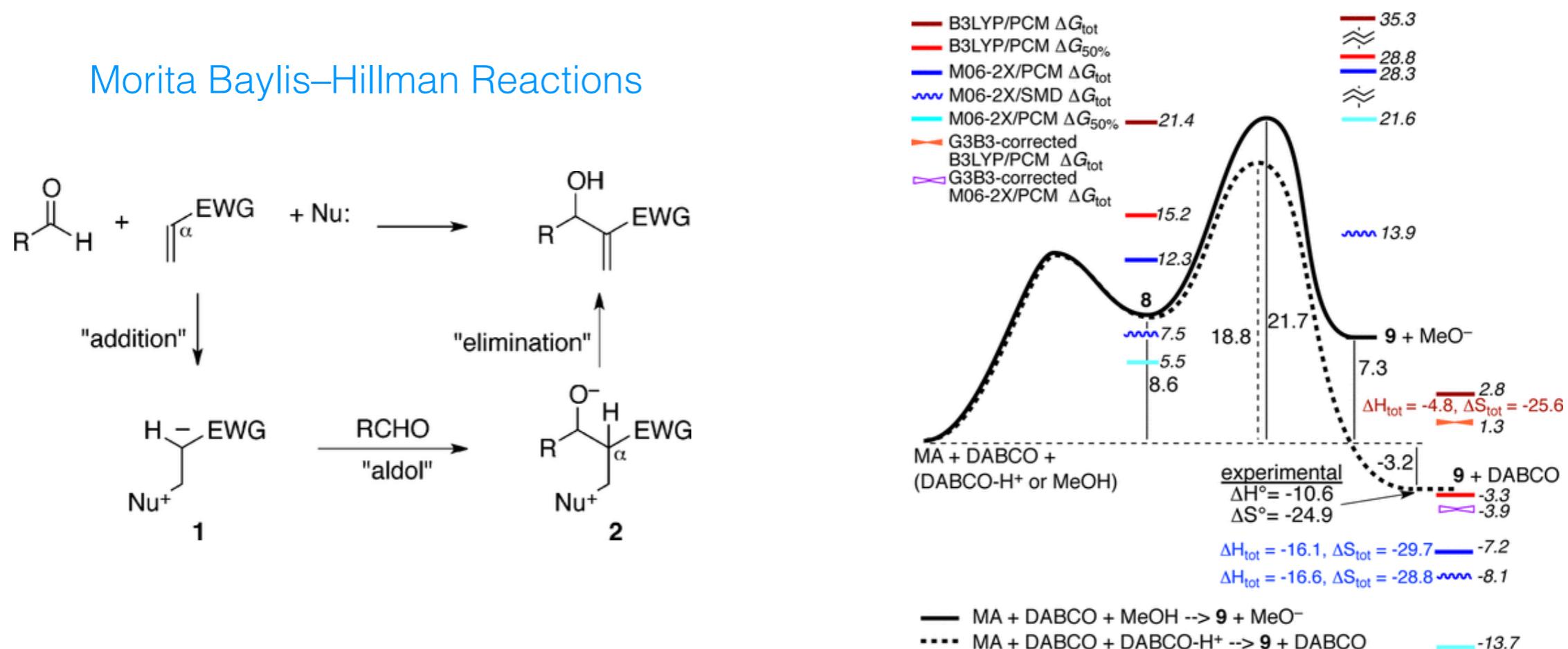


# Accuracy – Current Limitations



"Overall, it is not clear to us that any significant accurate information that was not already apparent from experiment either has been, or could have been, reliably garnered purely from computations."

# Accuracy – Current Limitations



*The gratifying outcome is that the new results, while still not accurate enough to be predictive, are at least consistent with experiment.*

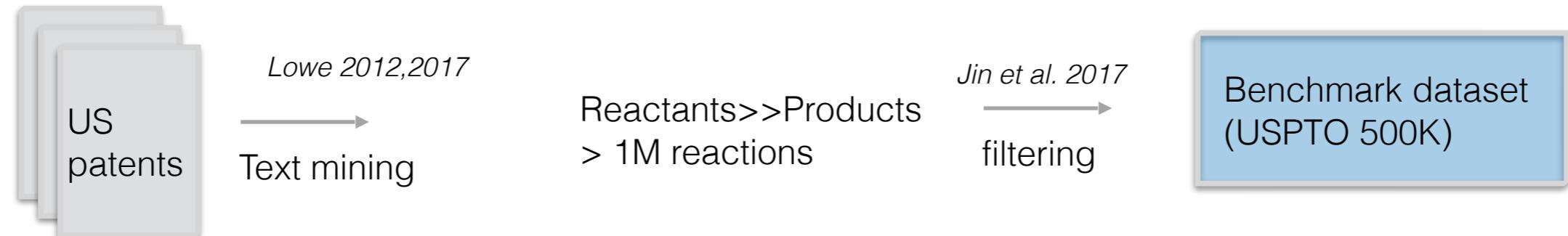
*Electronic energies | Entropy | Solvation*

[J. Harvey et. Al Phys. Chem. Chem. Phys., 2017, 19, 30647](#)

15

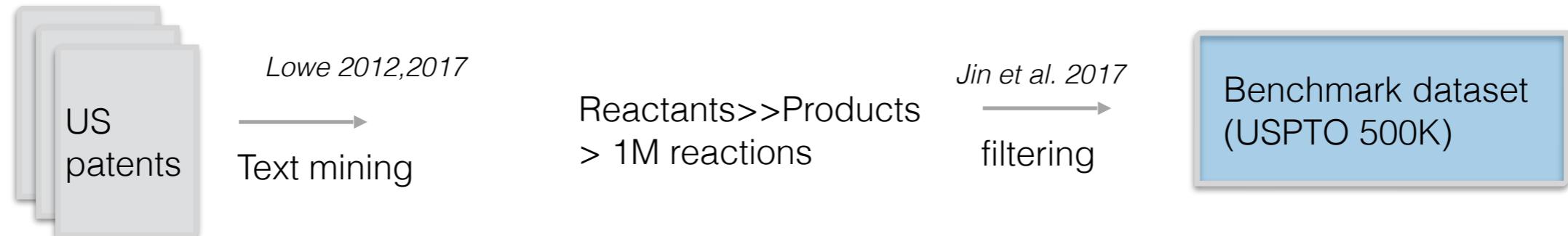
# Fast predictions

## DATA



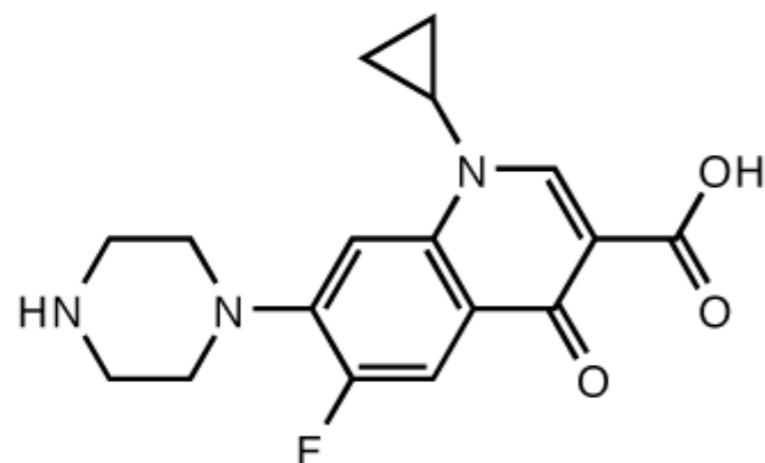
# Fast predictions

## DATA



## LANGUAGE

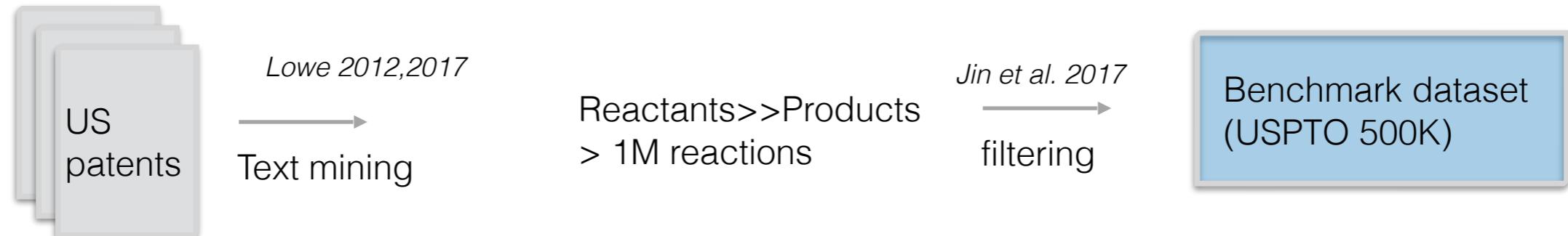
The simplified molecular-input line-entry system (SMILES)



Ciprofloxacin

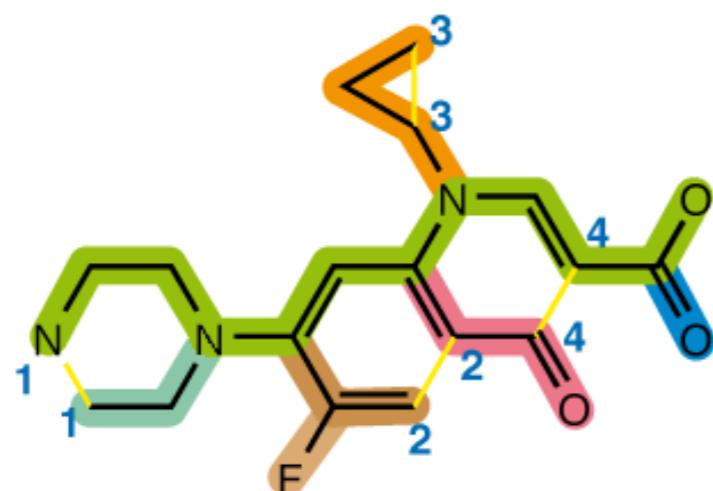
# Fast predictions

## DATA

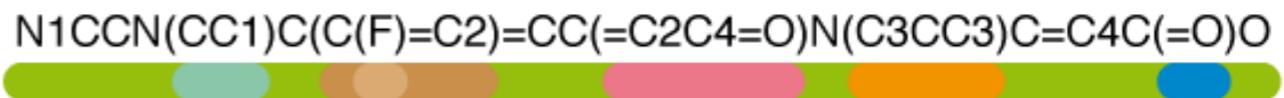


## LANGUAGE

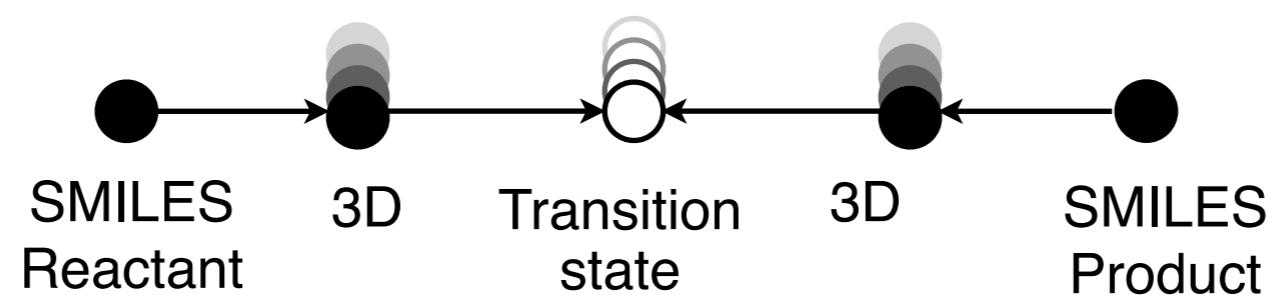
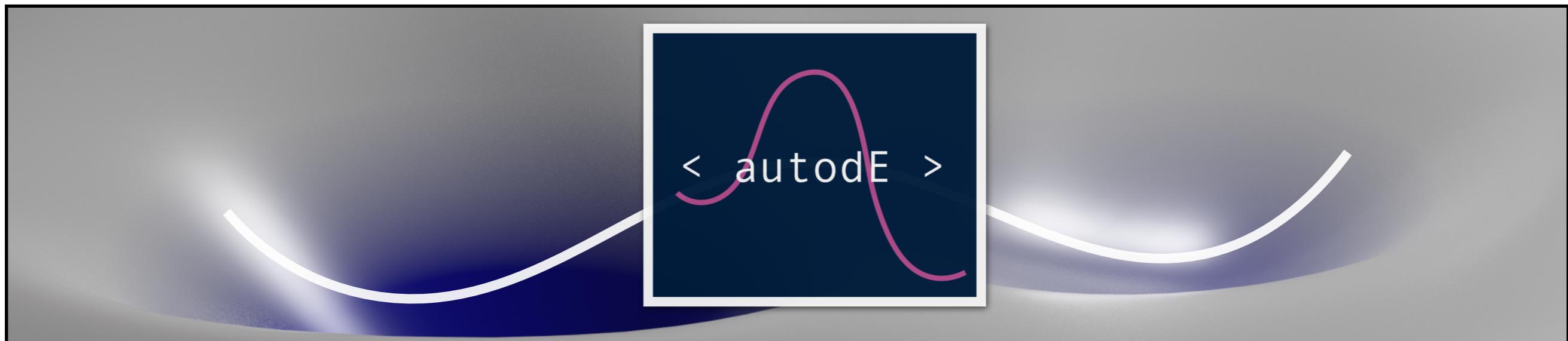
The simplified molecular-input line-entry system (SMILES)



Ciprofloxacin



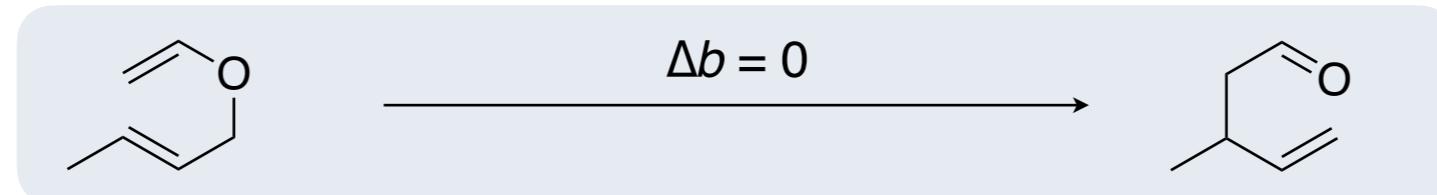
# Automation of Reaction Mechanisms



- Organic + organometallic reactions
- Automated TS finding / Sampling
- Electronic structure package
- Freely available / minimal user input

# Automation of Reaction Mechanisms

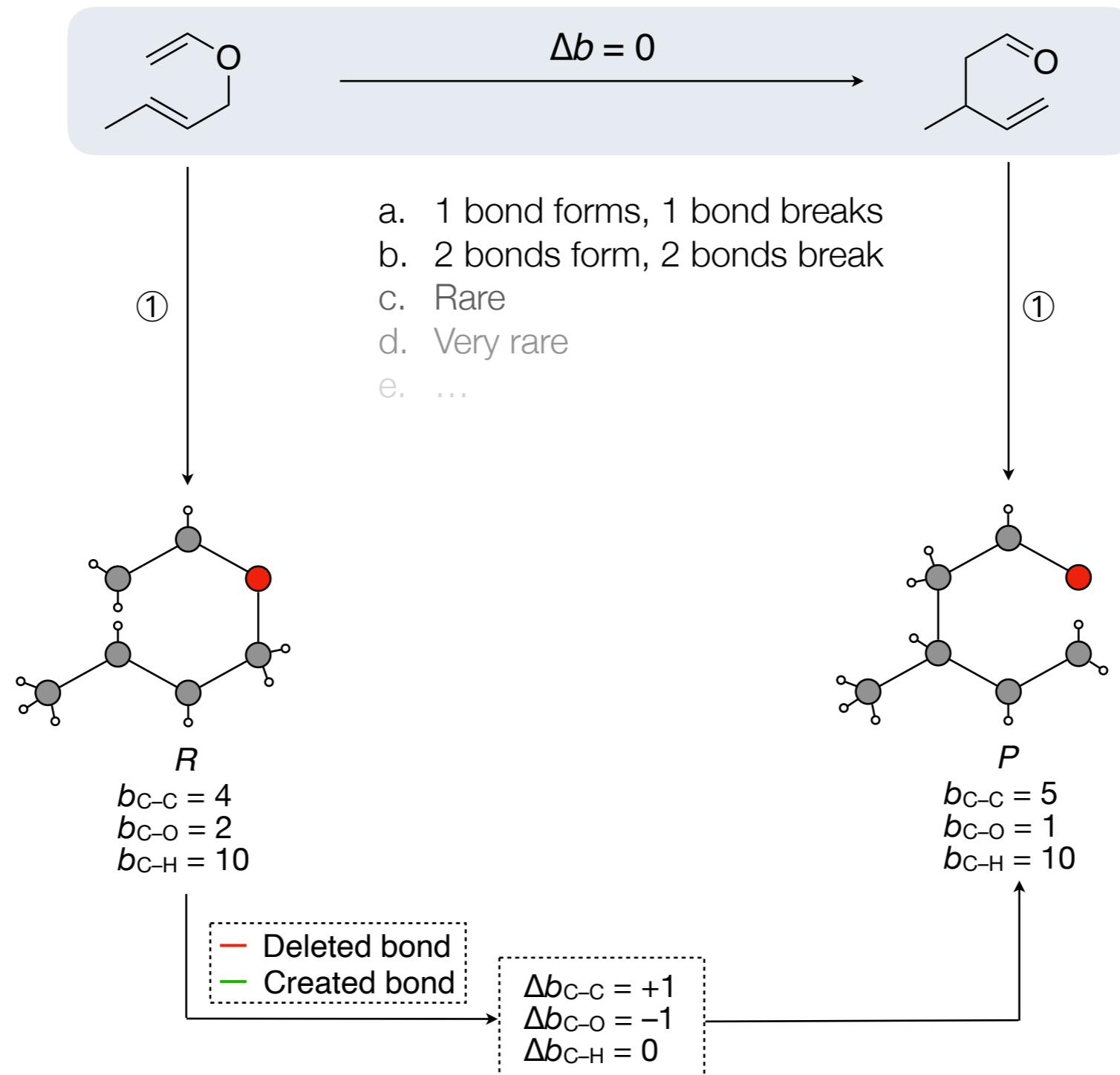
Reactants & Products → TS → TS conformer → single points



- a. 1 bond forms, 1 bond breaks
- b. 2 bonds form, 2 bonds break
- c. Rare
- d. Very rare
- e. ...

# Automation of Reaction Mechanisms

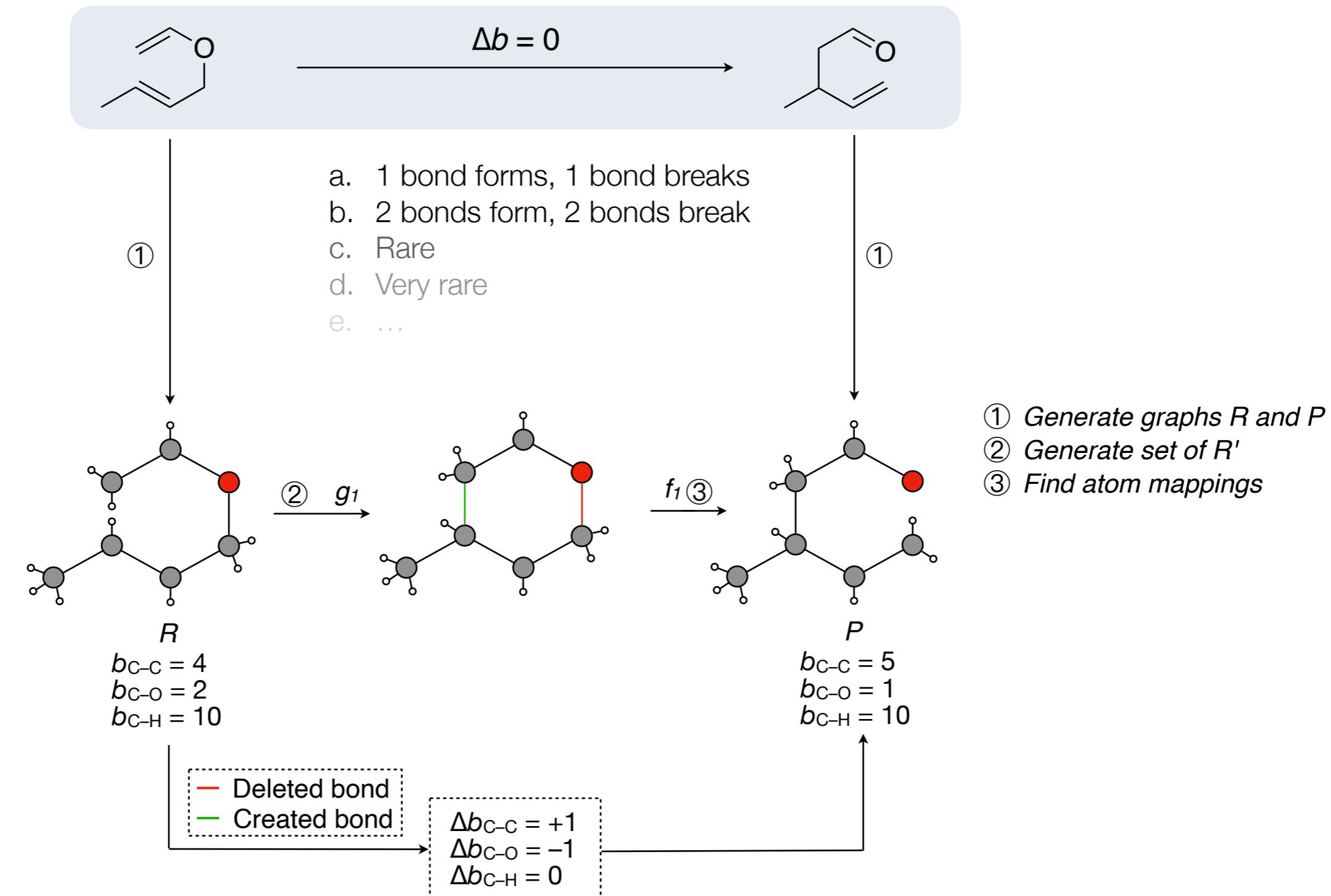
Reactants & Products → TS → TS conformer → single points



- ① Generate graphs  $R$  and  $P$
- ② Generate set of  $R'$
- ③ Find atom mappings

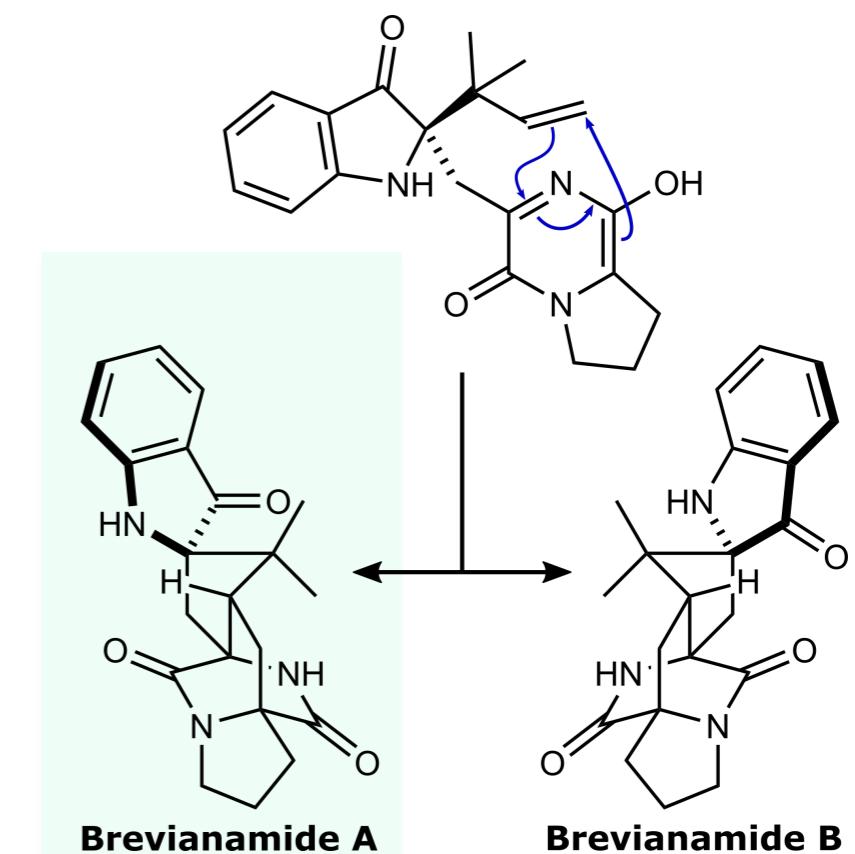
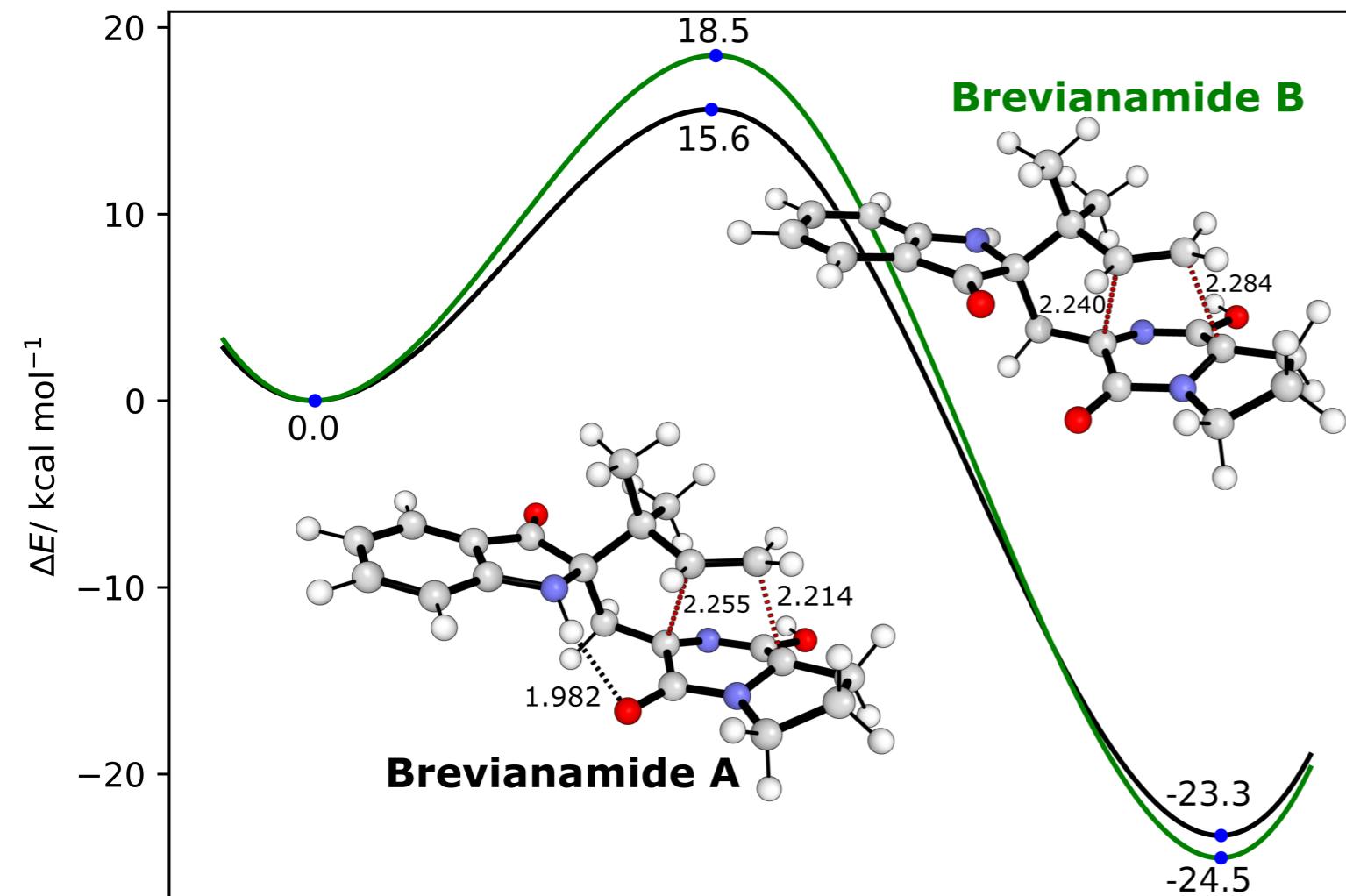
# Automation of Reaction Mechanisms

Reactants & Products  $\longrightarrow$  TS  $\longrightarrow$  TS conformer  $\longrightarrow$  single points



# Automation of Reaction Mechanisms

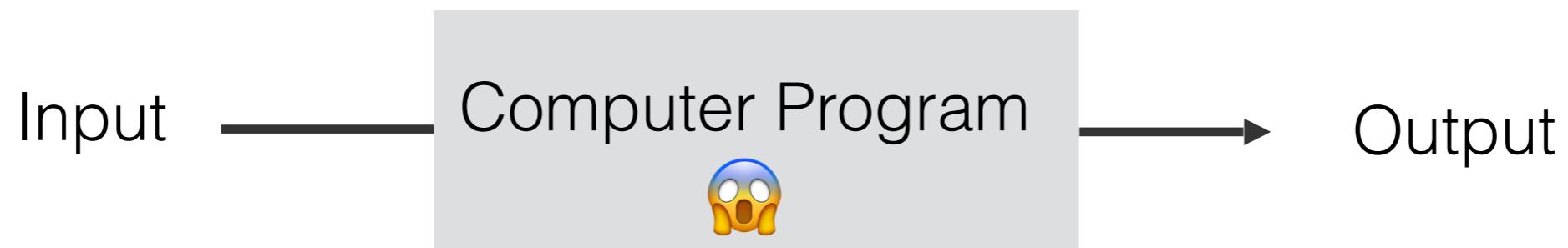
Diels-alder reaction for the formation of *Brevianamide A* and *Brevianamide B*



# Machine Learning

“Machine” —> computer programs

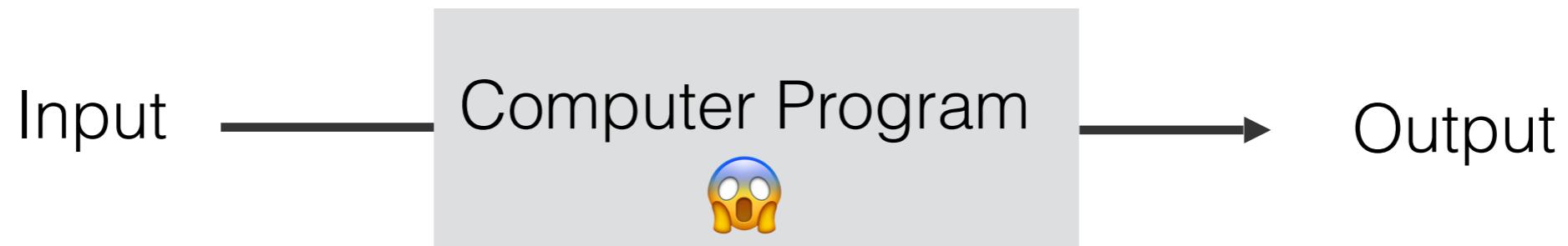
“Learning” —> to automatically improve with experience



# Machine Learning

“Machine” —> computer programs

“Learning” —> to automatically improve with experience

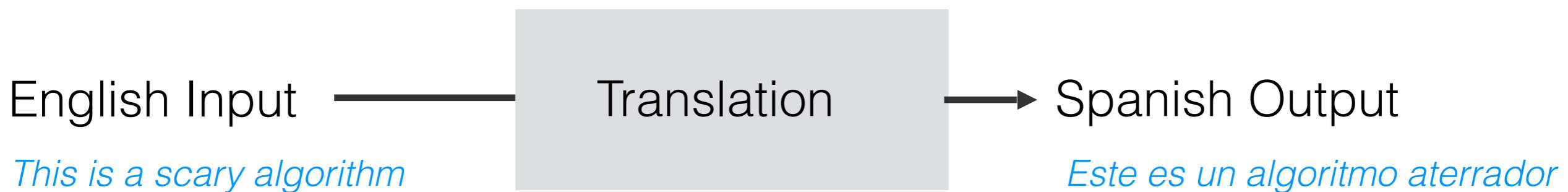


How to get outputs from inputs?

# Machine Learning

“Machine” —> computer programs

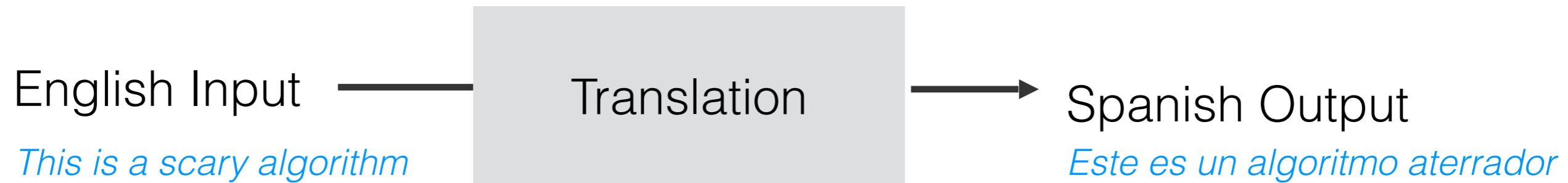
“Learning” —> to automatically improve with experience



How can we construct computer programs able to do this?

So we give “training data” to teach programs  
“automatically improve with experience (given data)”

# Machine Learning



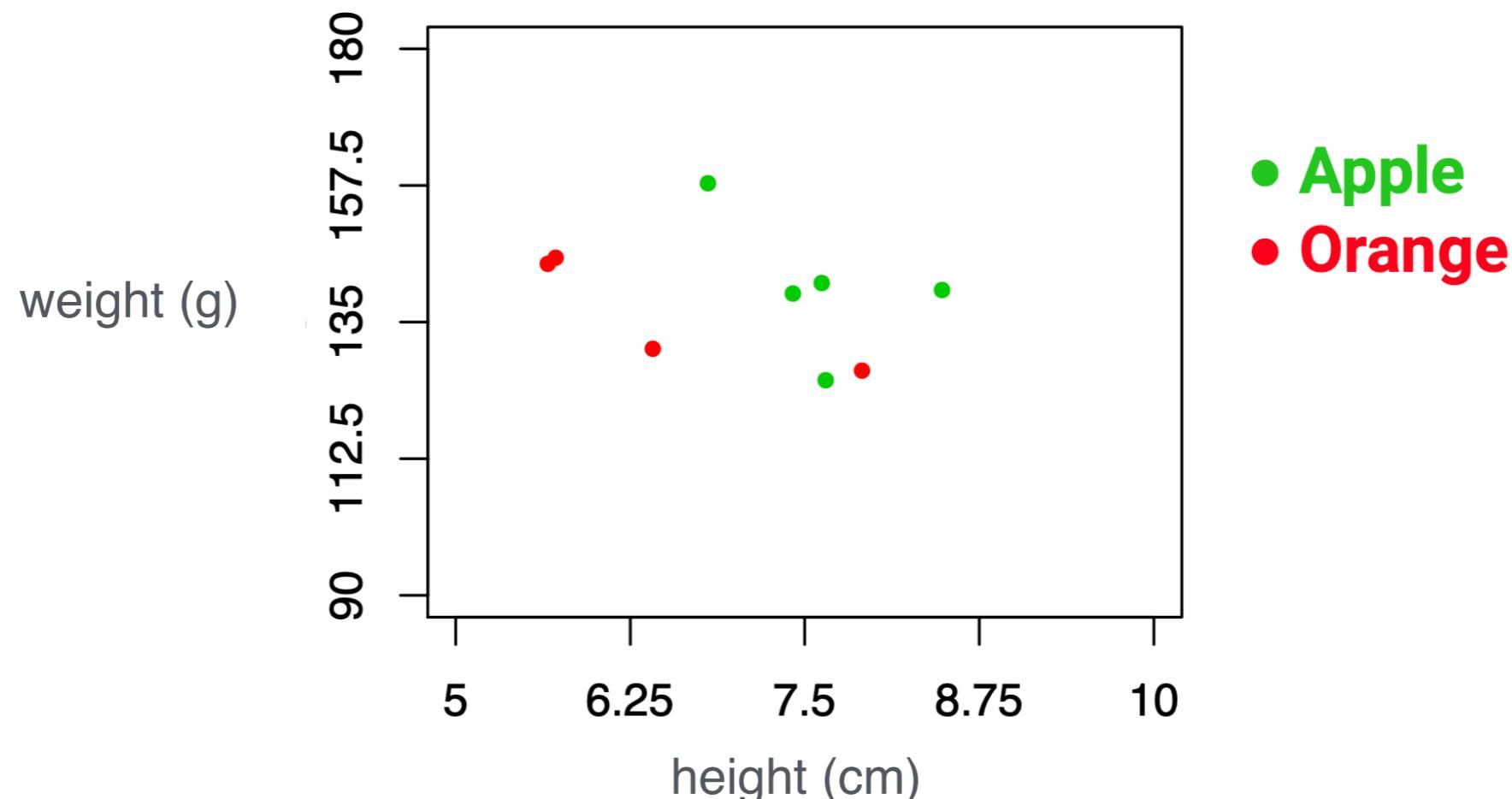
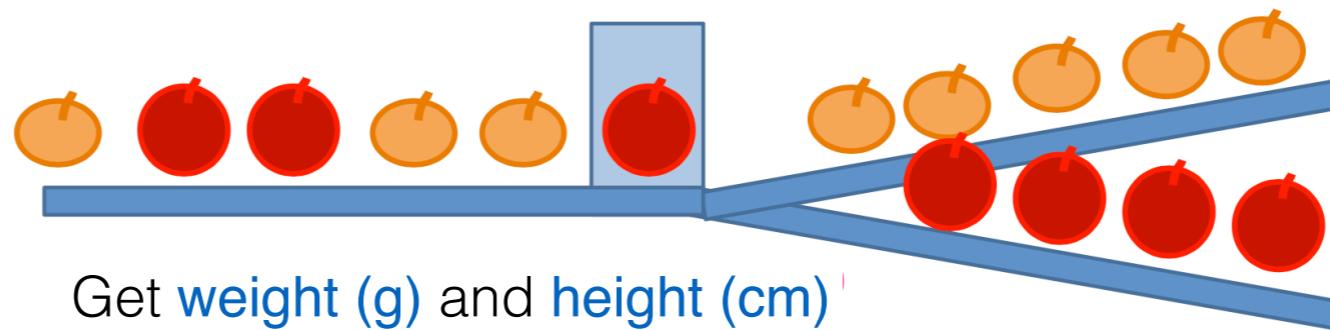
English	Spanish
Cat	Gato
Chemistry	Química
Science	Ciencia
Friend	Amigo/Amiga
...	...

## Use of statistical trends

Identification of common patterns (empirical rules) emerge from observing many examples, which we cannot recognise when we see only a few

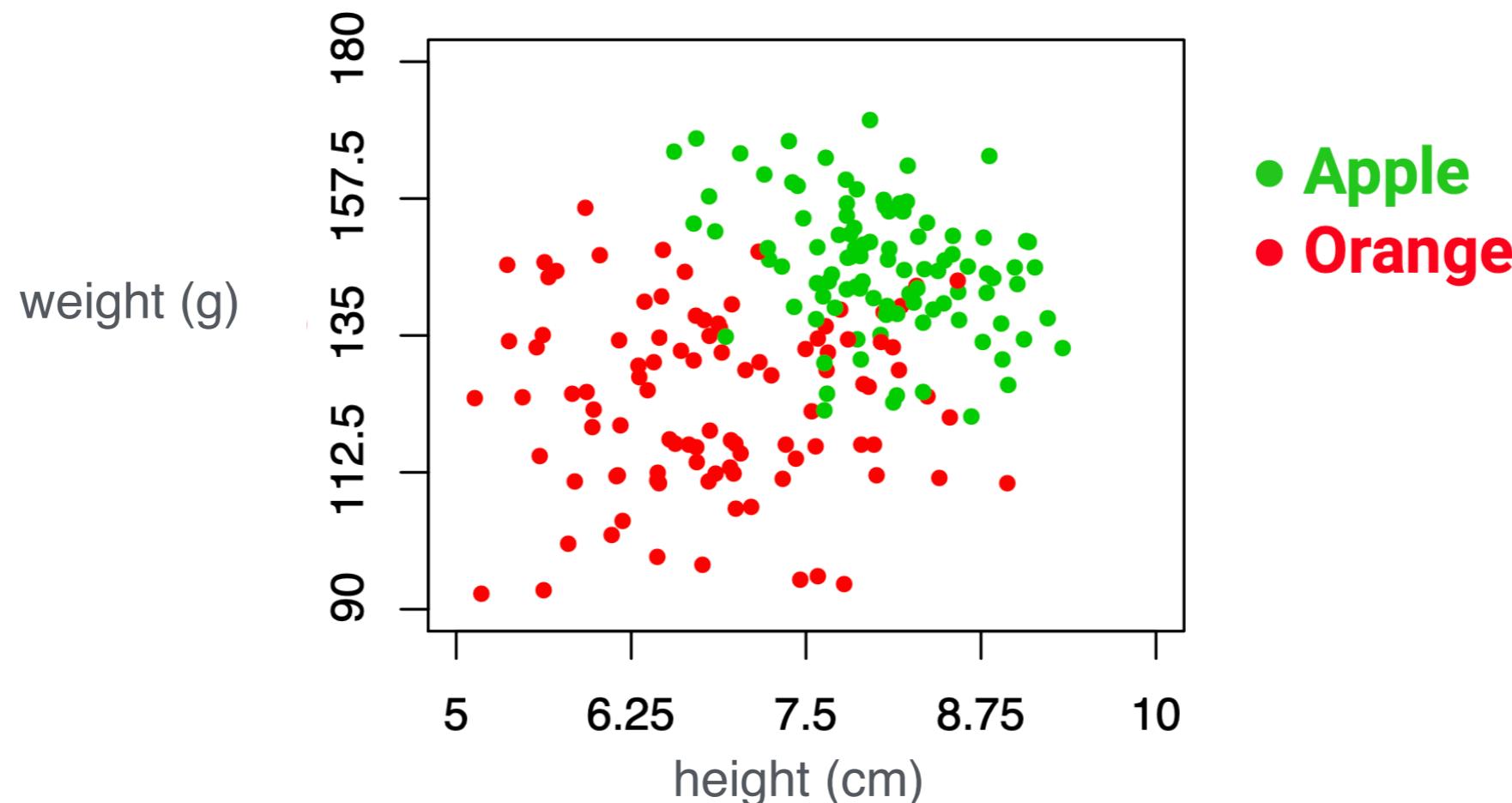
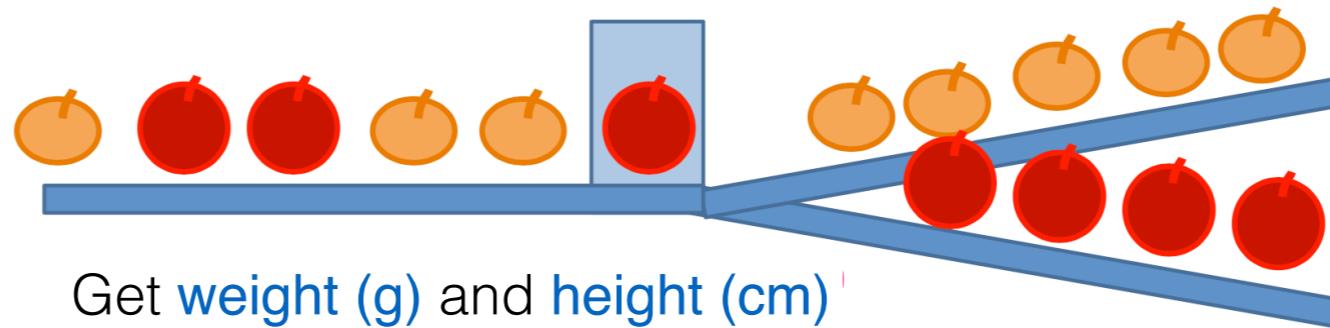
# Machine Learning

Patterns emerge from many examples



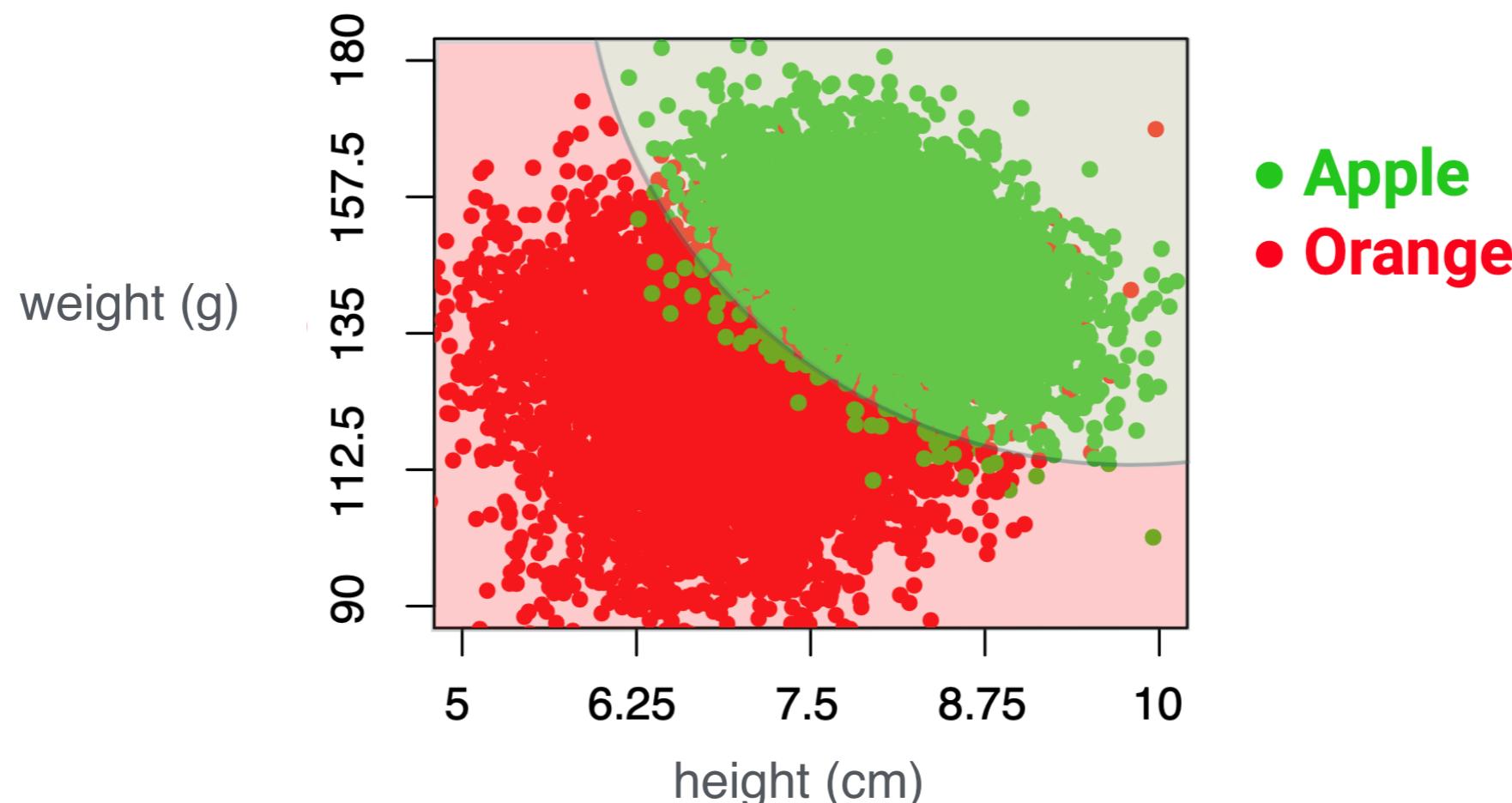
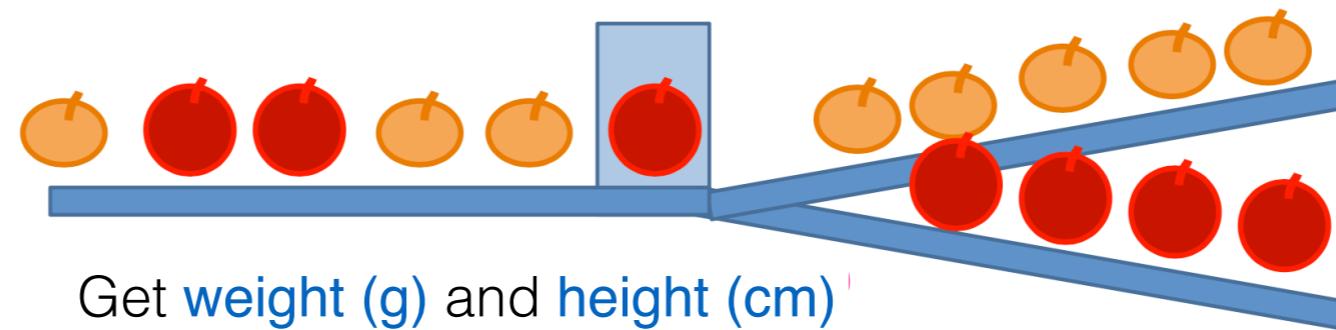
# Machine Learning

Patterns emerge from many examples



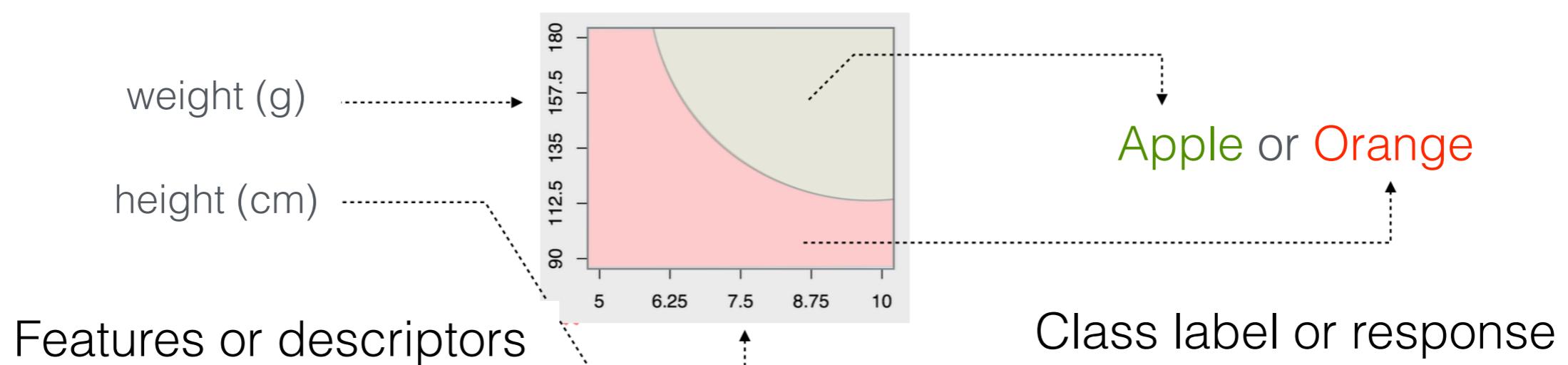
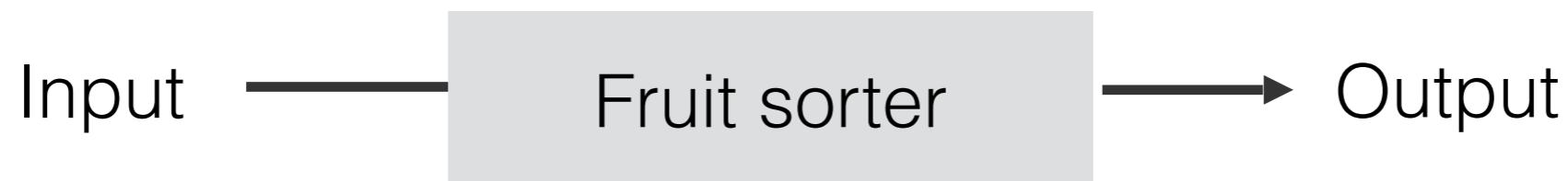
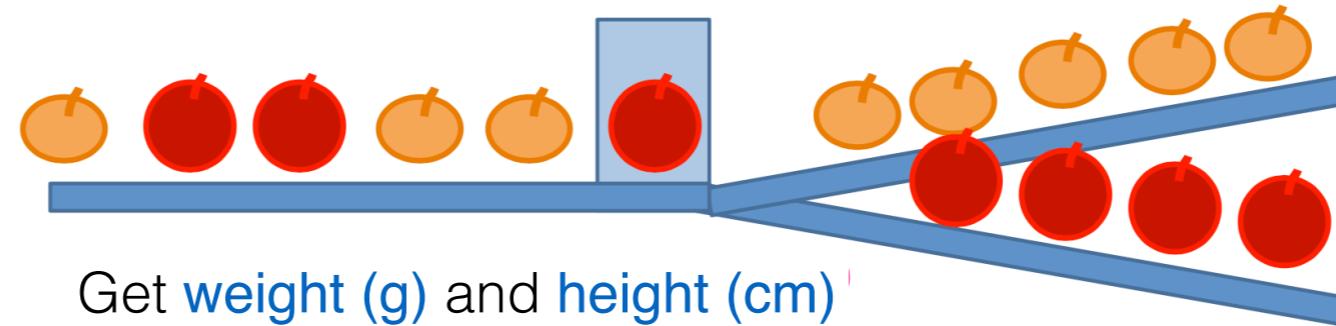
# Machine Learning

Patterns emerge from many examples



# Machine Learning

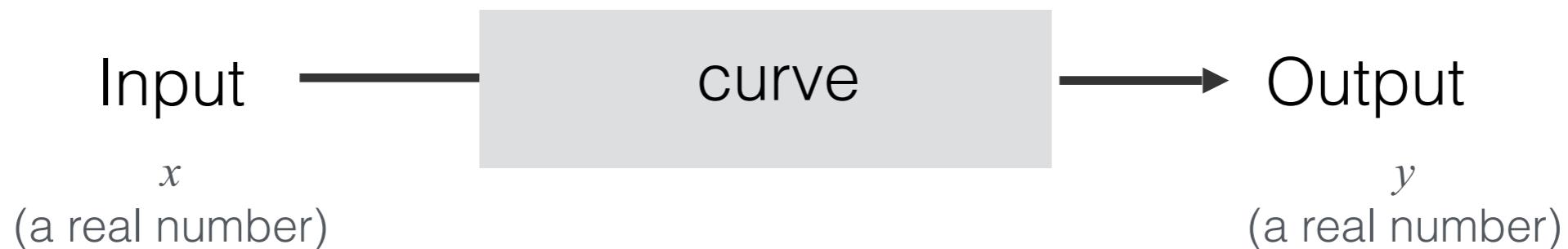
Patterns emerge from many examples



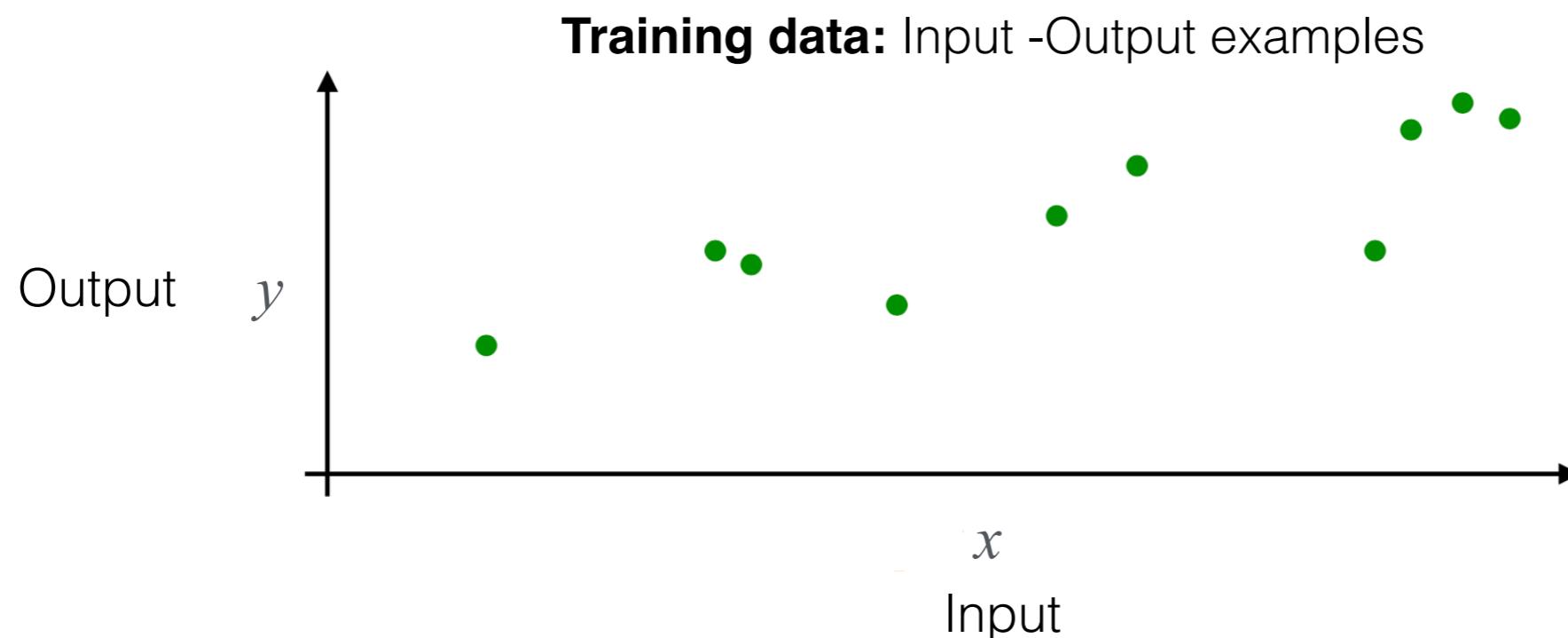
# Machine Learning

1-dimensional curve fitting

ML model



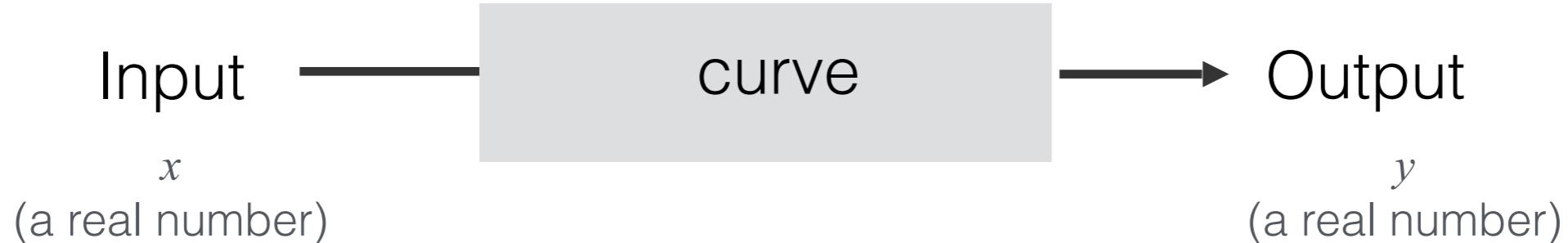
Training data:  $\{(x^1, y^1), (x^2, y^2), \dots, (x^n, y^n)\}$



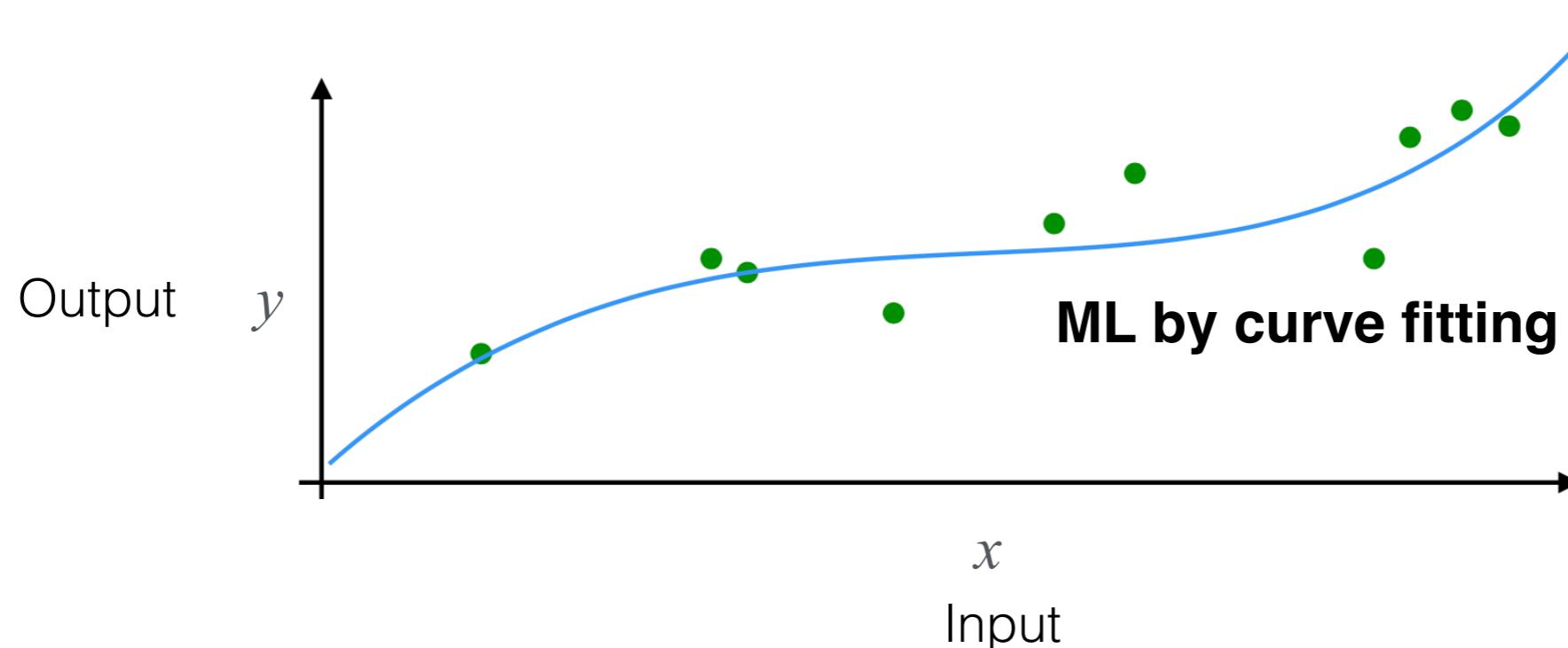
# Machine Learning

1-dimensional curve fitting

ML model



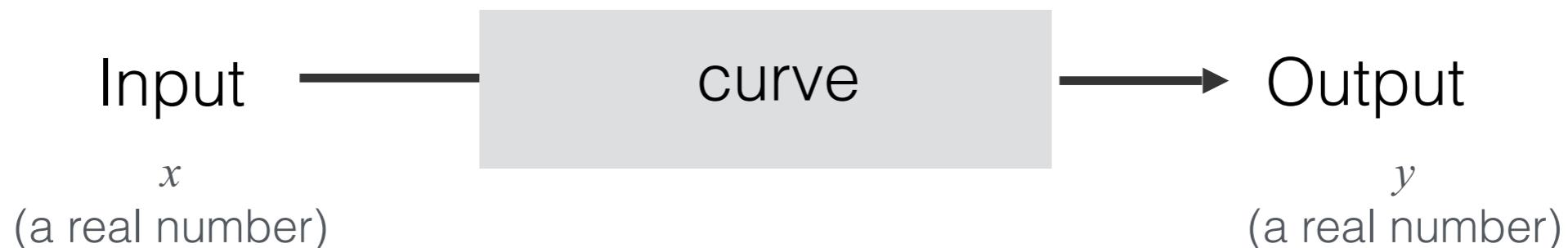
Training data:  $\{(x^1, y^1), (x^2, y^2), \dots, (x^n, y^n)\}$



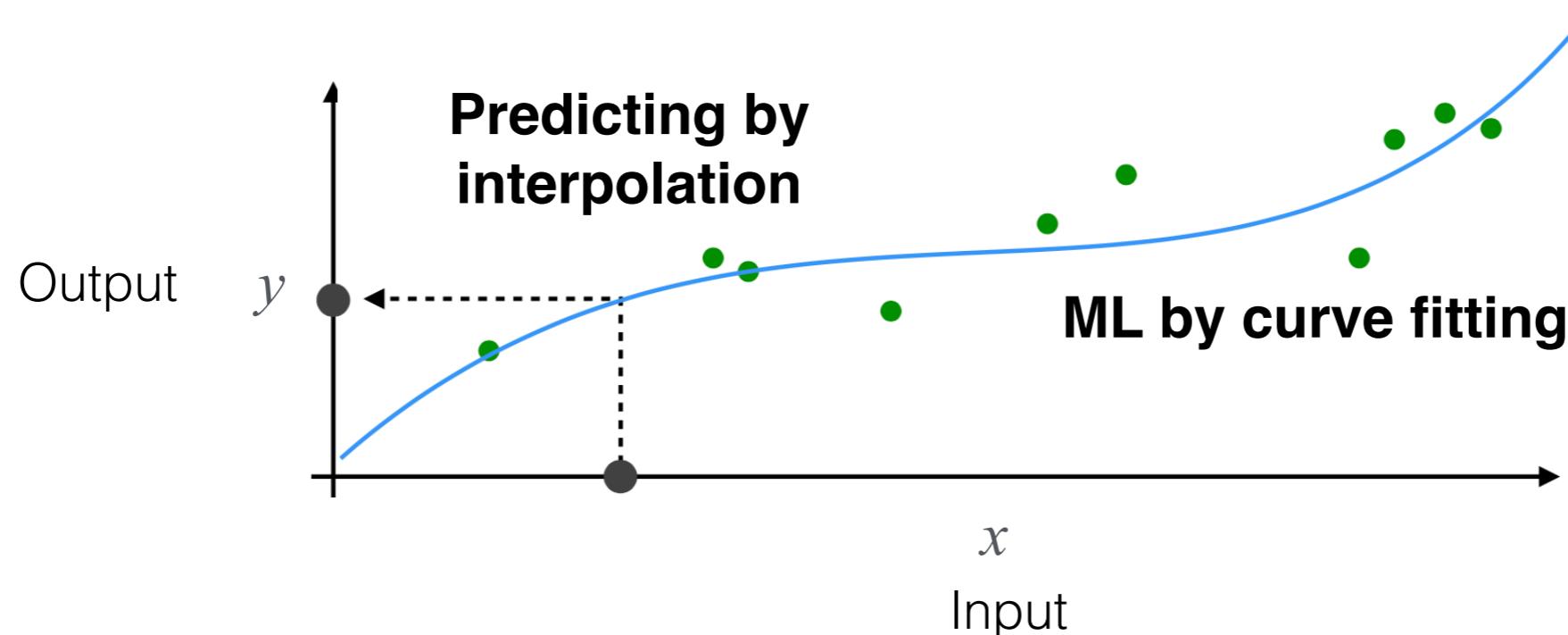
# Machine Learning

1-dimensional curve fitting

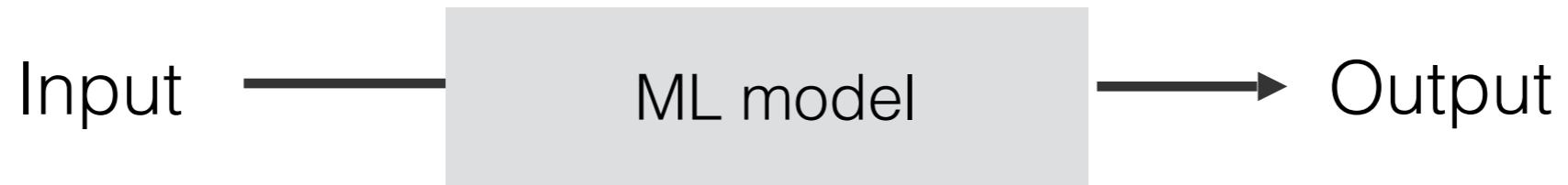
ML model



Training data:  $\{(x^1, y^1), (x^2, y^2), \dots, (x^n, y^n)\}$



# Machine Learning

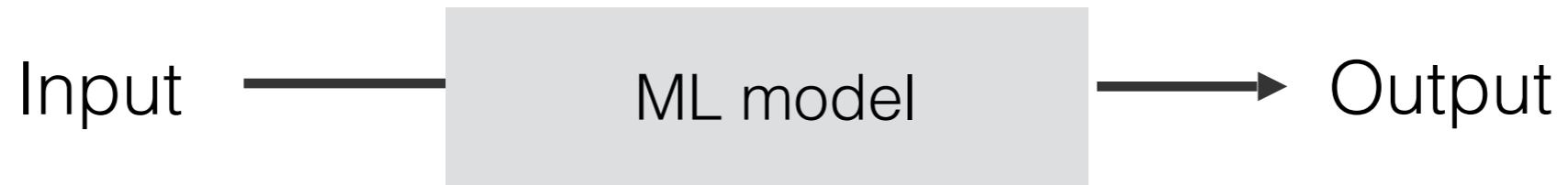


In summary... current ML models usually mean **just an interpolation by curve fitting!**

Unfortunately curve fitting in a high-dimensional space is not trivial at all!



# Machine Learning



In summary... current ML models usually mean **just an interpolation by curve fitting!**

Unfortunately curve fitting in a high-dimensional space is not trivial at all!



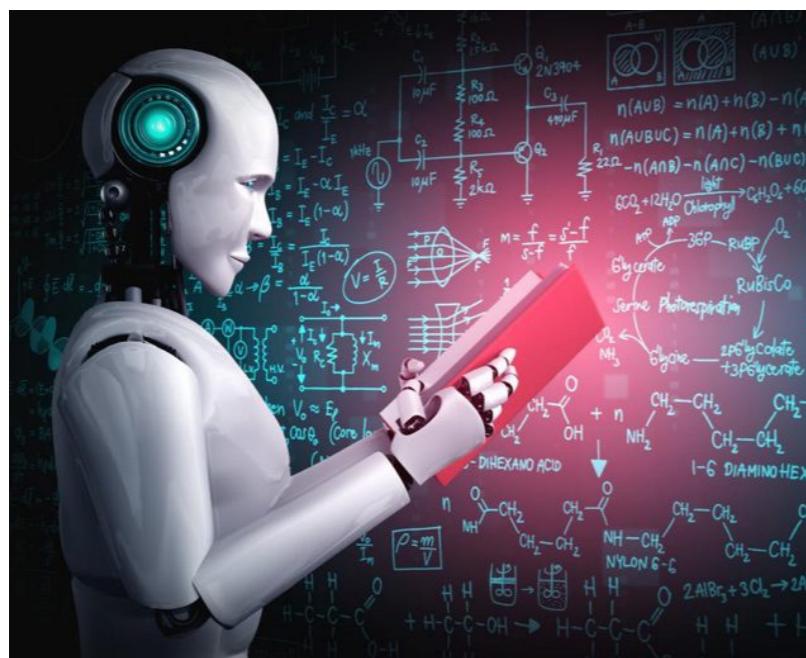
# Machine Learning

Input — ML model → Output

In summary... current ML models usually mean **just an interpolation by curve fitting!**

Unfortunately curve fitting in a high-dimensional space is not trivial at all!

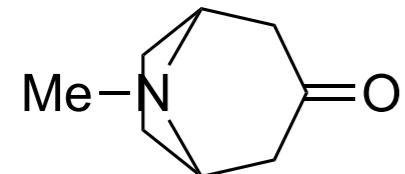
It does **NOT** involve artificial human intelligence – deep thinking/reasoning – magic



# Fast predictions

Forward and Retrosynthesis

Tropinone



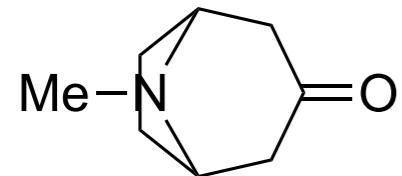
1901 (Richard Willstätter) —> 15 steps – 0.75%

E. J. Corey, *Science*, 1969, 166, 178 • W. L. Jorgensen, *J. Org. Chem.*, 1980, 45, 2043 • E. J. Corey, ACIE 1991, 30, 45 • B. A. Grzybowski, ACIE 2016, 55, 5904 and *Nature* 2020 ASAP • K. F. Jensen, ACS Cent. Sci., 2017, 3, 1237 and *Science*, 2019, 365, eaax1566; M. P. Waller, *Nature*, 2018, 555, 604 • J. M. Bishop, ACS Cent. Sci. 2019, 5, 6, 970 • A. A. Lee, ACS Cent. Sci., 2019, 5, 1572; T. Laino *Chem. Sci.* 2020, 11, 3316

# Fast predictions

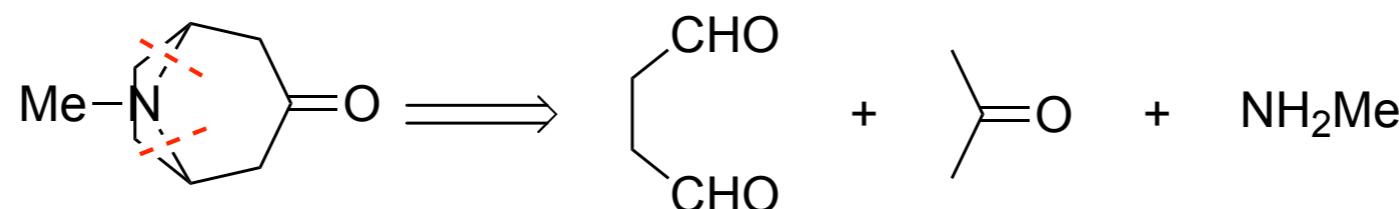
Forward and Retrosynthesis

Tropinone



1901 (Richard Willstätter) —> 15 steps – **0.75%**

1917 (Robert Robinson) —> 1 step – **17% -90%**

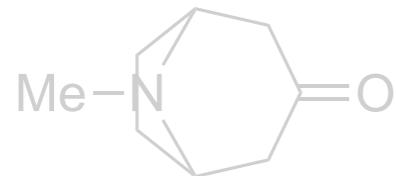


E. J. Corey, *Science*, 1969, 166, 178 • W. L. Jorgensen, *J. Org. Chem.*, 1980, 45, 2043 • E. J. Corey, *ACIE* 1991, 30, 45 • B. A. Grzybowski, *ACIE* 2016, 55, 5904 and *Nature* 2020 ASAP • K. F. Jensen, *ACS Cent. Sci.*, 2017, 3, 1237 and *Science*, 2019, 365, eaax1566; M. P. Waller, *Nature*, 2018, 555, 604 • J. M. Bishop, *ACS Cent. Sci.* 2019, 5, 6, 970 • A. A. Lee, *ACS Cent. Sci.*, 2019, 5, 1572; T. Laino *Chem. Sci.* 2020, 11, 3316

# Fast predictions

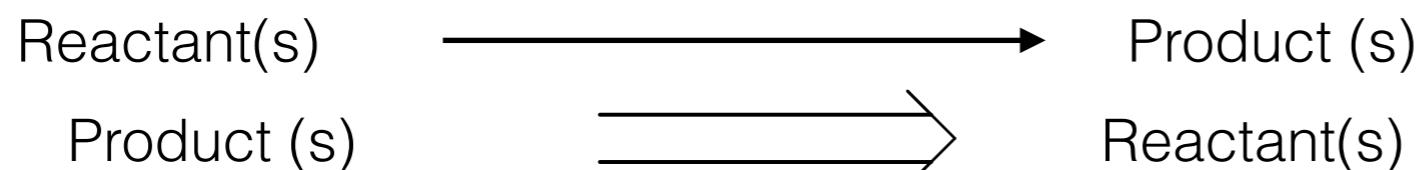
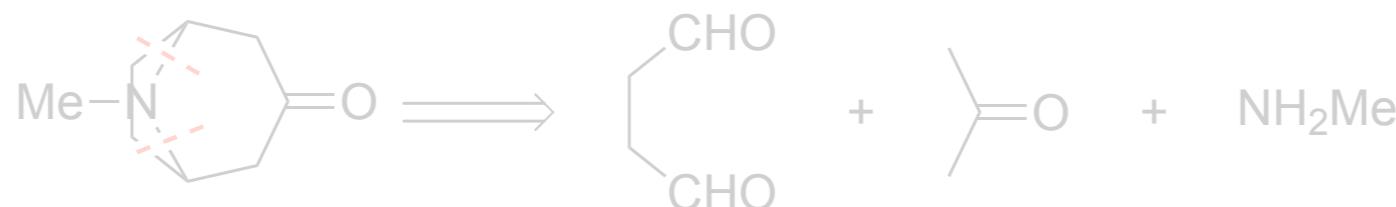
## Forward and Retrosynthesis

Tropinone



1901 (Richard Willstätter) —> 15 steps – 0.75%

1917 (Robert Robinson) —> 1 step – 17% -90%



Rule-based • Template-based • Natural language processing

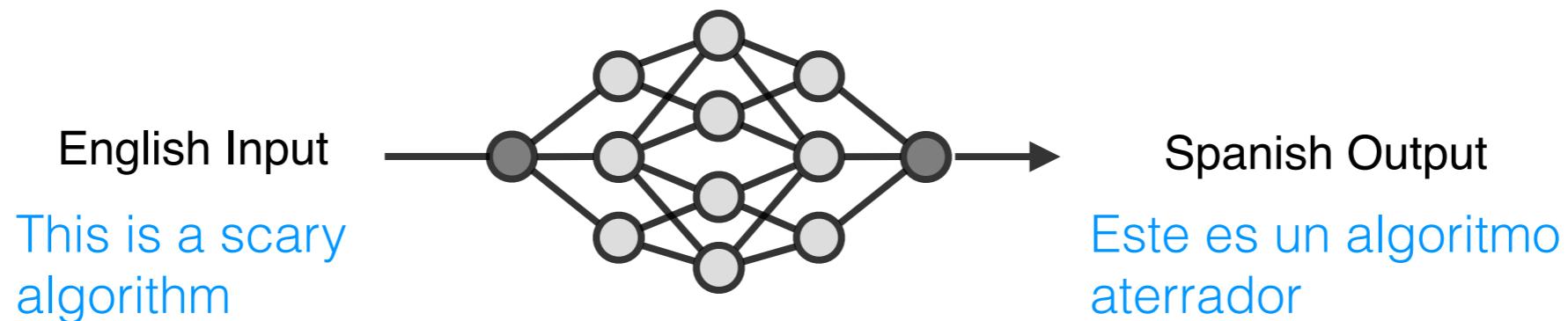


E. J. Corey

E. J. Corey, *Science*, 1969, 166, 178 • W. L. Jorgensen, *J. Org. Chem.*, 1980, 45, 2043 • E. J. Corey, *ACIE* 1991, 30, 45 • B. A. Grzybowski, *ACIE* 2016, 55, 5904 and *Nature* 2020 ASAP • K. F. Jensen, *ACS Cent. Sci.*, 2017, 3, 1237 and *Science*, 2019, 365, eaax1566; M. P. Waller, *Nature*, 2018, 555, 604 • J. M. Bishop, *ACS Cent. Sci.* 2019, 5, 6, 970 • A. A. Lee, *ACS Cent. Sci.*, 2019, 5, 1572; T. Laino *Chem. Sci.* 2020, 11, 3316

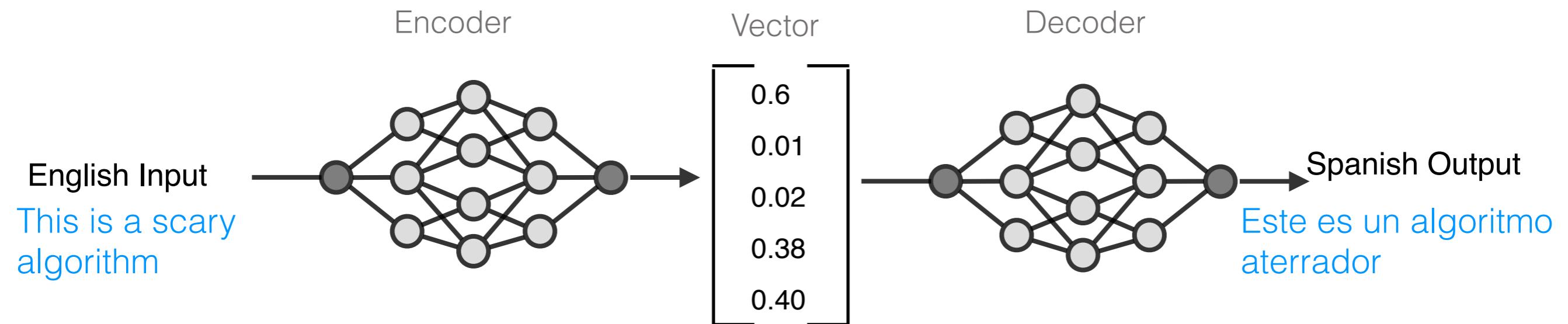
# Fast predictions

Translation as Sentence Completion !



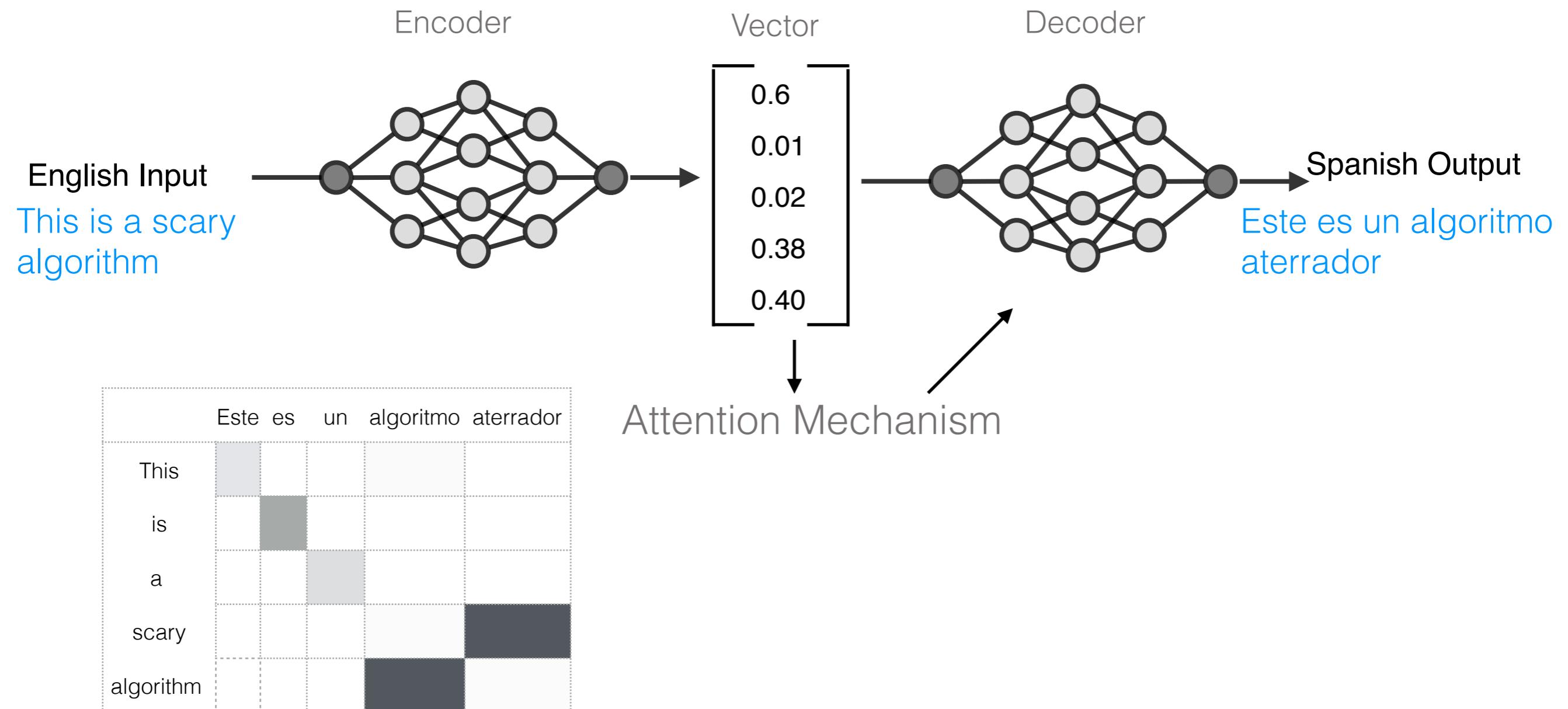
# Fast predictions

Translation as Sentence Completion !

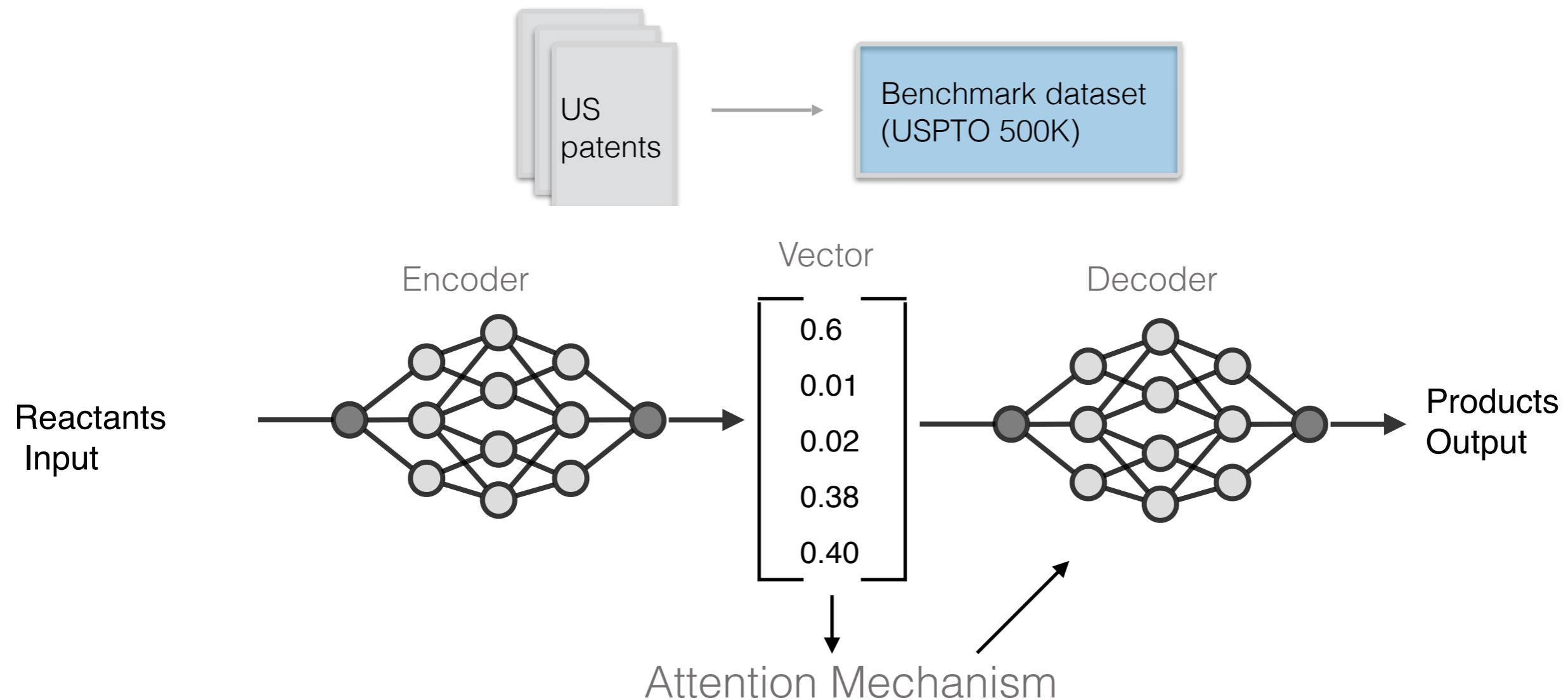


# Fast predictions

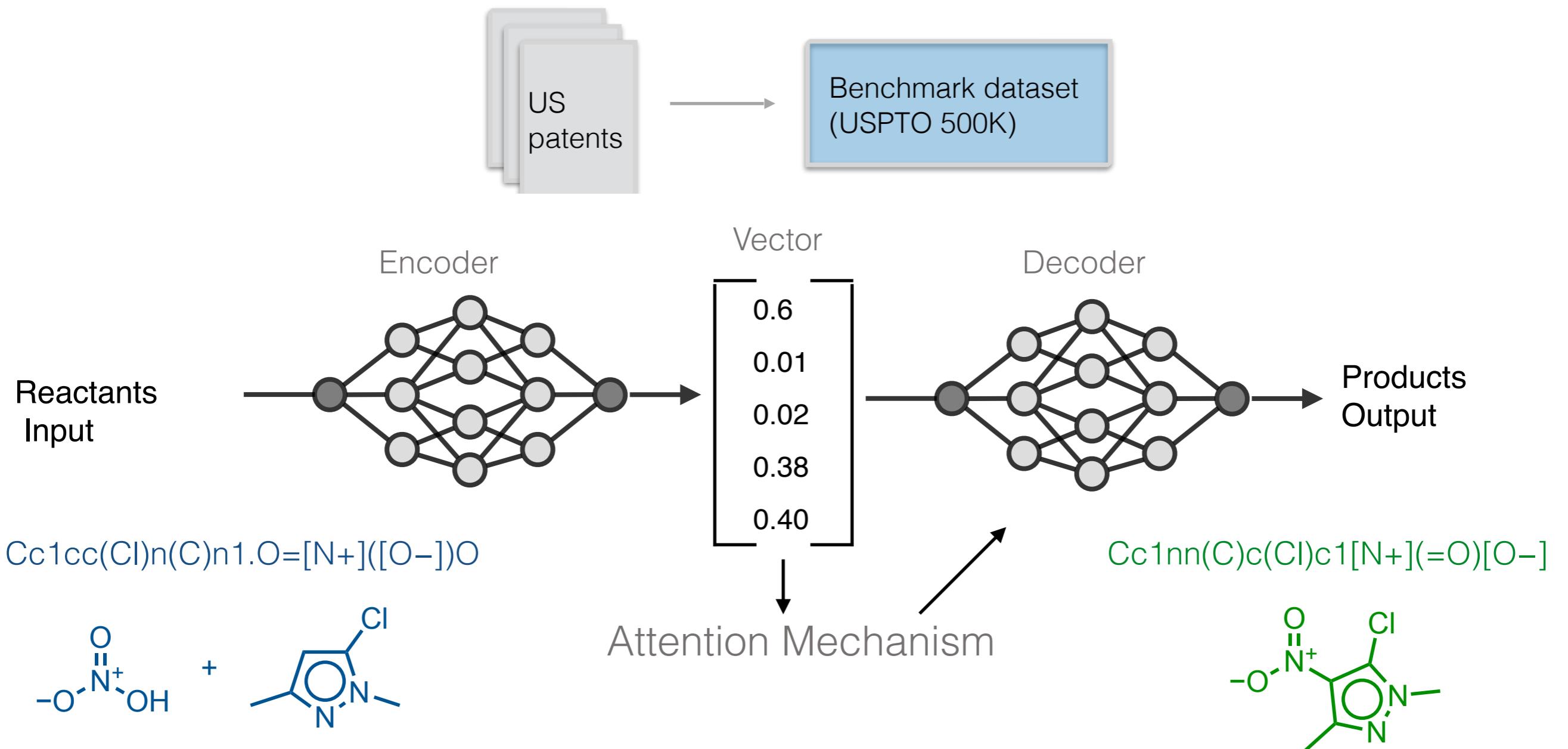
Translation as Sentence Completion !



# Fast predictions



# Fast predictions



# Fast predictions

The screenshot shows the IBM RXN web application interface. At the top, there's a navigation bar with a close button, the title "IBM RXN", a search icon, and a user profile icon. To the right of the title is a blue button labeled "Add new project" with a plus sign. Below the navigation bar, the main content area is titled "Welcome Fernanda Duarte". A sub-header asks "What would you like to do today?". There are three main sections: "Predict forward reaction" (with a flask icon), "Predict retrosynthetic routes" (with a molecular structure icon), and "Plan a synthesis" (with a circular arrow icon). On the left side, there's a sidebar with links: "Home" (which is selected and highlighted in grey), "Projects", "RoboRXN", "Text to procedure", "Atom mapping" (with a dropdown arrow), "Collaborators", "Learn", and "WebCam". At the bottom left of the main content area is a URL link: <https://rxn.res.ibm.com/rxn/home>.

# Fast predictions

The screenshot shows the IBM RXN web application interface. At the top, there is a navigation bar with a close button (X), the title "IBM RXN", a search icon, and a user profile icon. To the right of the title is a blue button labeled "Add new project" with a plus sign. Below the navigation bar, the main content area is titled "Welcome Fernanda Duarte". A sub-header asks "What would you like to do today?". Three main options are presented in cards:

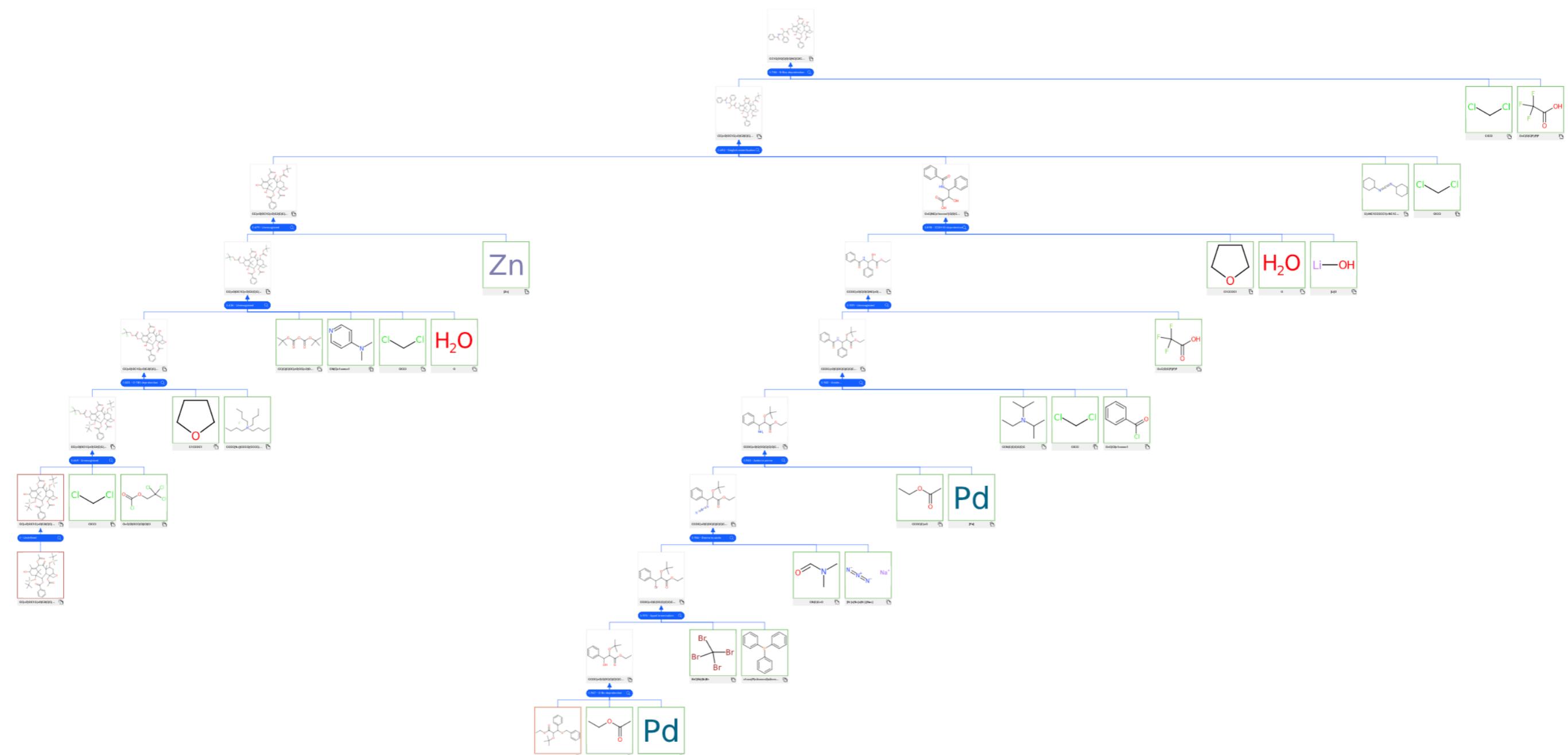
- Predict forward reaction**: Predict the product of a reaction from its precursors. It features an icon of a flask.
- Predict retrosynthetic routes**: Predict possible retrosynthetic routes given a target molecule. It features an icon of a branched molecular structure.
- Plan a synthesis**: Plan and execute a synthesis starting from a target molecule, a retrosynthetic route, or an experimental procedure in text format. It features an icon of a circular process flow.

Below these cards, a section titled "There are many things you can do on IBM RXN" displays three colored boxes with icons: a teal box with a robotic arm and a grid, a red box with a molecular structure and a pen, and a purple box with an atom and a network of lines.

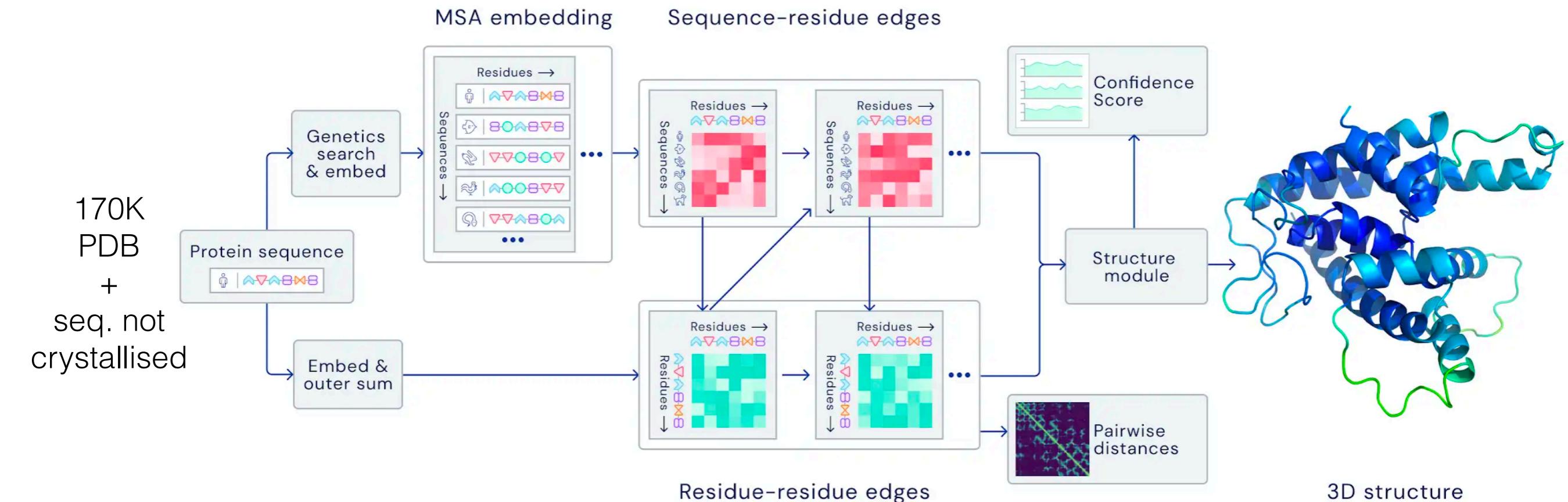
<https://rxn.res.ibm.com/rxn/home>

# Fast predictions

# Fast predictions



# Critical Assessment of protein Structure Prediction (CASP)



# Critical Assessment of protein Structure Prediction (CASP)

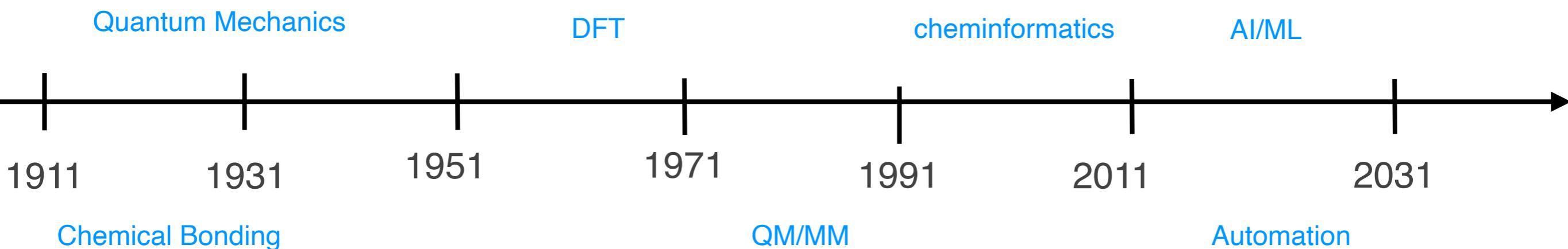
## Median Free-Modelling Accuracy



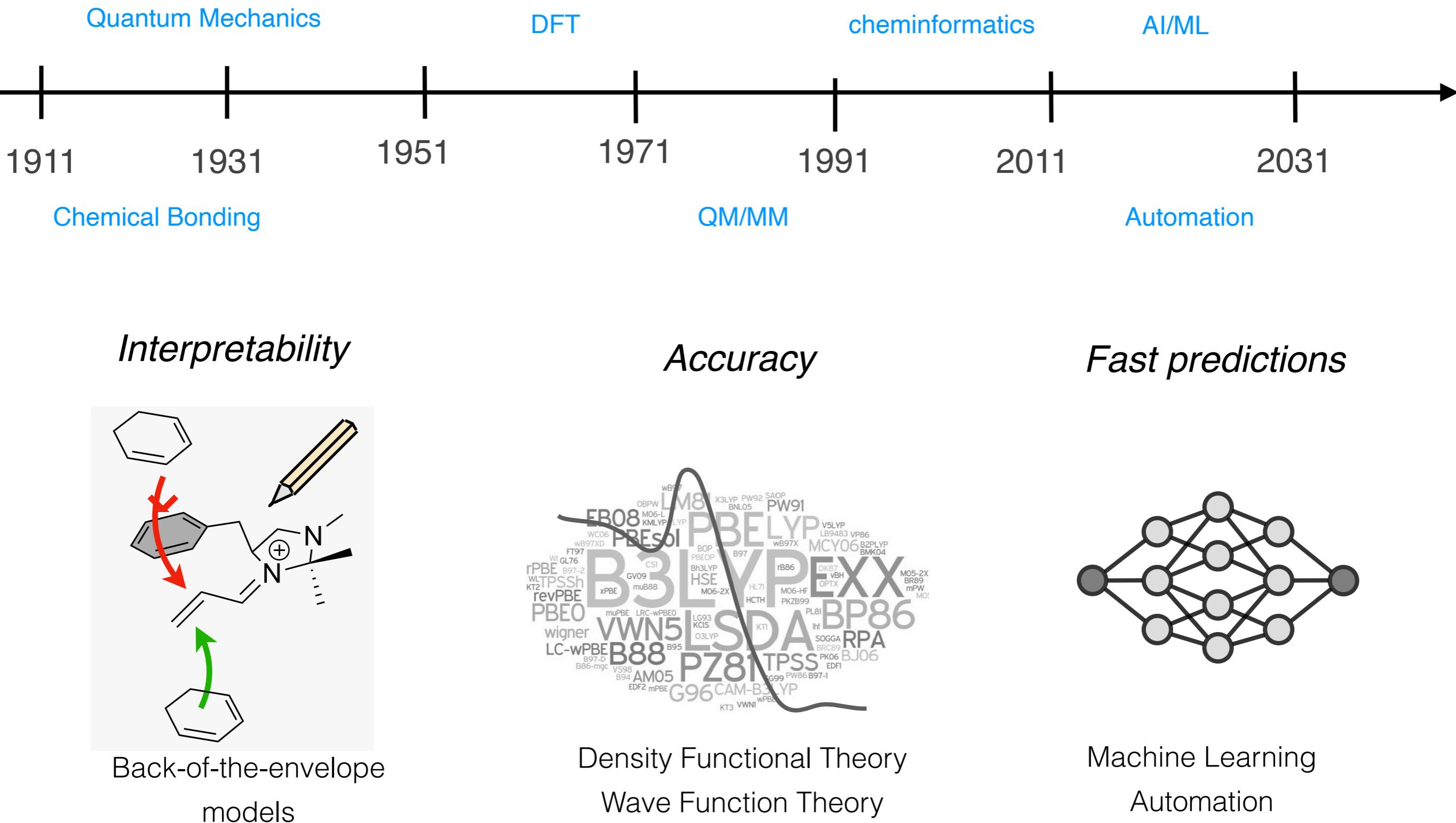
# Fast predictions

- Freely available tools ✓
- Average chemist accuracy ✓
- Organic chemistry relies on well-known reaction types ✗
- Template-free methods are more general, but depend on the validity of SMILES ✗
- Prediction accuracy: memorising vs learning ✗

# Give us insights and numbers



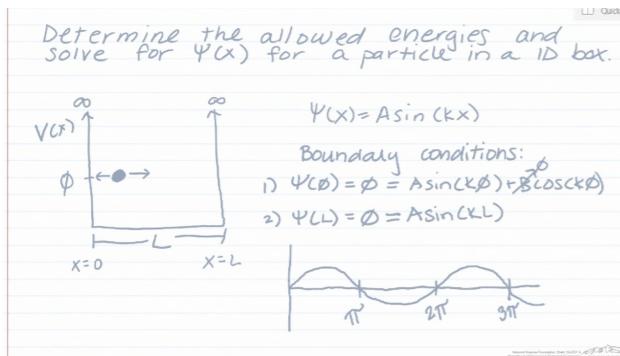
# Give us insights and numbers



# Give us insights and numbers

## Digital Chemistry

### Quantum/Organic Chemistry

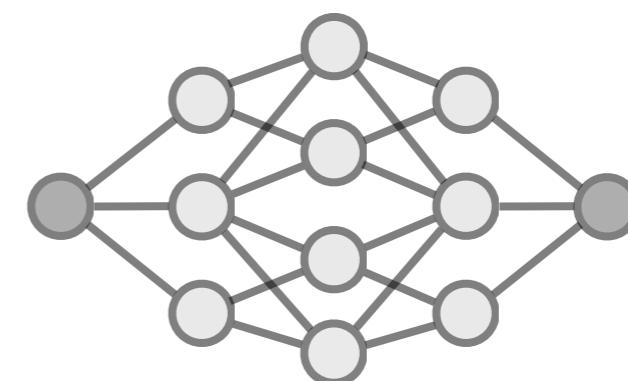
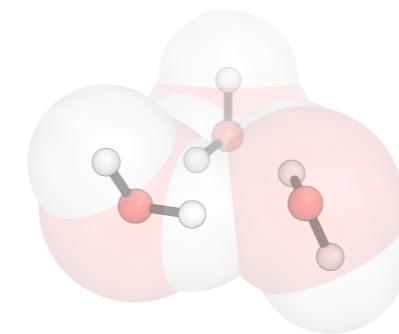


Virtual Reality



Automation

### Molecular Modelling

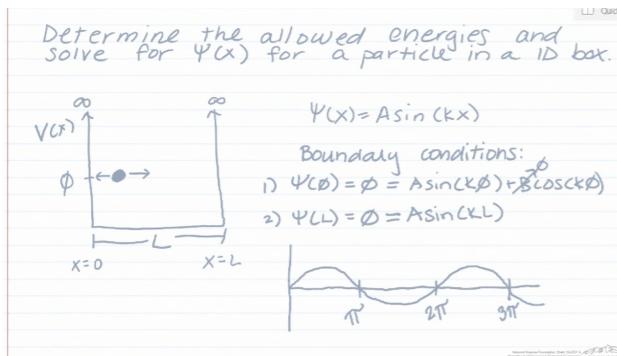


Cheminformatics  
Artificial Intelligence

# Give us insights and numbers

## Digital Chemistry

### Quantum/Organic Chemistry

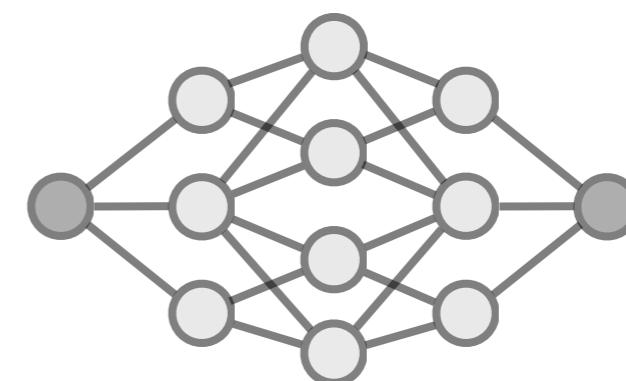
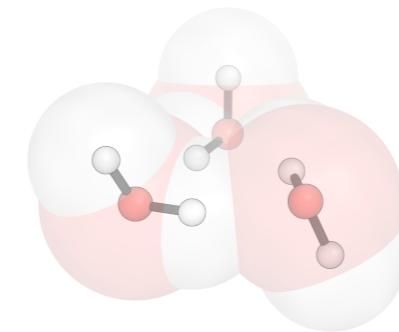


Virtual Reality



Automation

### Molecular Modelling



Cheminformatics  
Artificial Intelligence