



# **Réinventer la chimie théorique : IA et apprentissage automatique pour décrypter les mécanismes de réaction dans des systèmes complexes**

Anne MILET et Rolf DAVID

SITH, Département de Chimie Moléculaire, UGA/CNRS

GEPhyX 2025 – 3 Juillet 2025

# Que cherche-t-on à savoir ?

- Données en lien avec les résultats analytiques : géométrie, potentiel électrochimique, IR, UV, RMN, RPE, etc.
- Mécanismes réactionnels : structure de l'état de transition, barrière énergétique, étape limitante, force motrice de la réactivité

Avec quelle précision ?

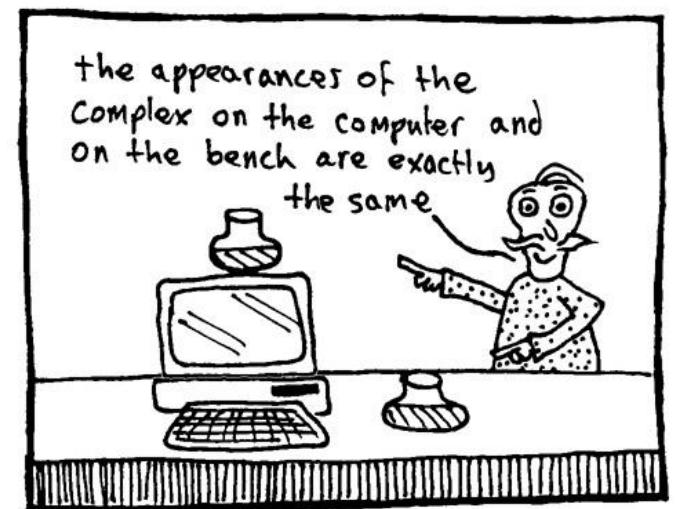
Combien de temps cela prendra-t-il ?

Quelles ressources seront nécessaires ?

Quelles approximations sont faites ?

Lesquelles sont significatives ?

**Nous aimerais tout savoir... et avec précision !**



Computational Chemistry

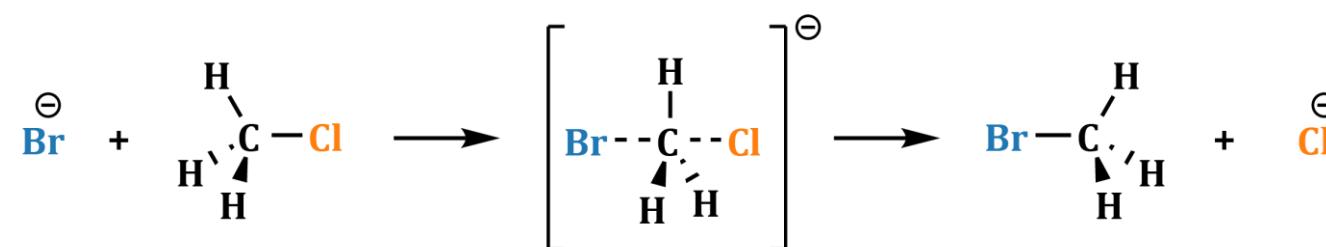
# Modélisation

- Création d'un modèle chimique
- Utilisation d'un modèle mathématique
- Rôle crucial de l'ordinateur
- Résultats:
  - On peut « voir » la réaction, les molécules
  - On peut « concevoir » de nouvelles molécules
  - On accède des propriétés (géométrie, données thermodynamiques, OM, moments dipolaires, charges, spectres IR, RMN, etc.)

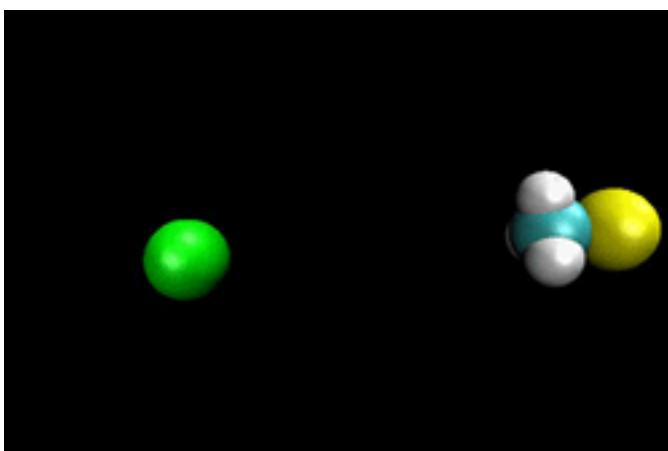
« It is nice to know that the computer understands the problem. But I would like to understand it too. »  
— E. Wigner

# Modélisation : création d'un modèle chimique

La substitution nucléophile d'ordre 2,  $\text{SN}_2$

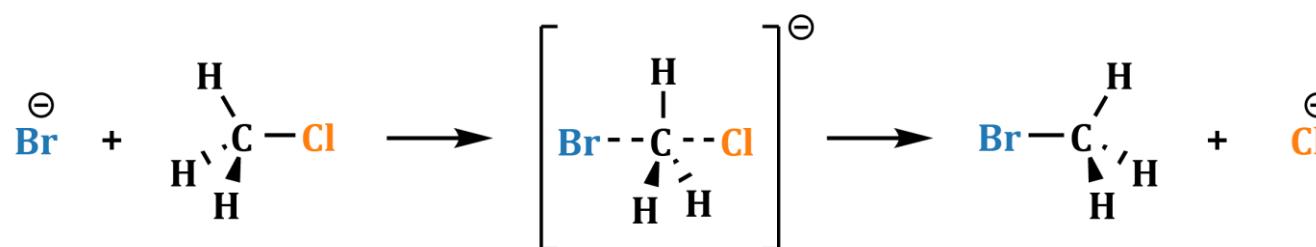


Le plus simple modèle

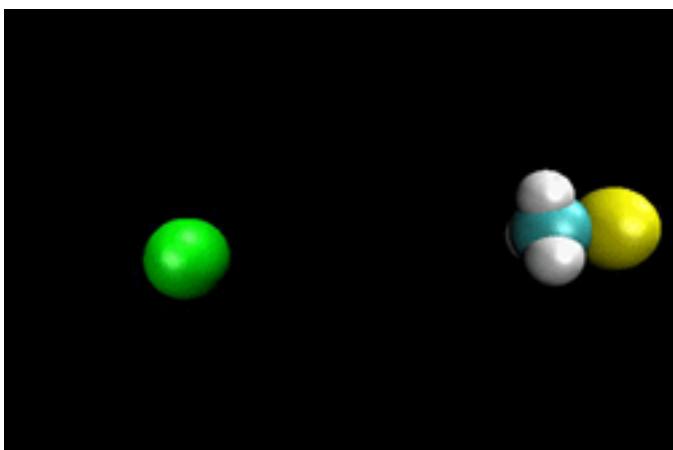


# Modélisation : création d'un modèle chimique

## La substitution nucléophile d'ordre 2, SN<sub>2</sub>



Le plus simple modèle

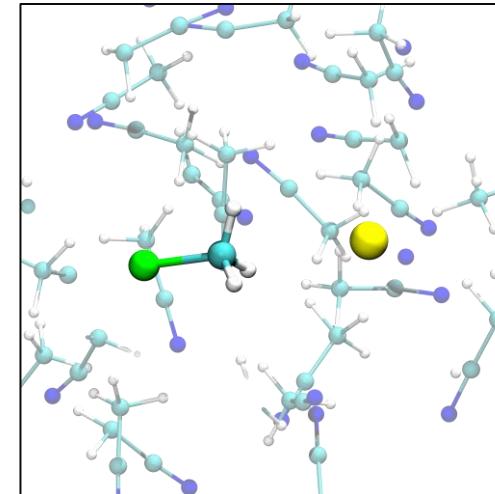
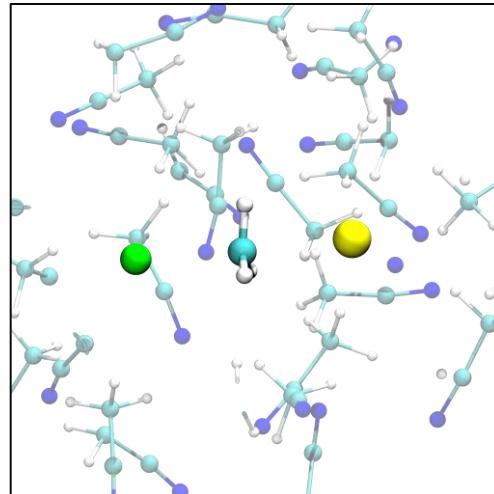
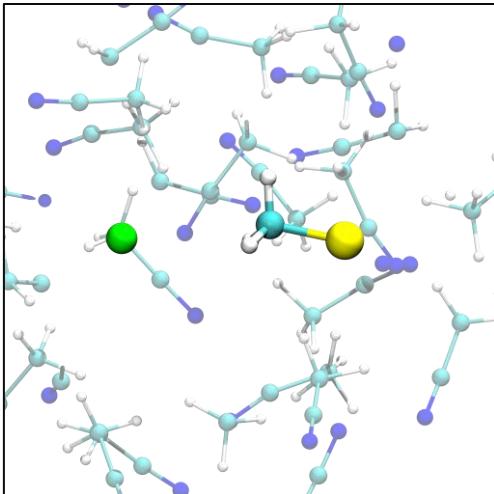


Facteurs à prendre en compte:

- Molécule entière, contre-ion ?
- Solvant ?
- Température ?
- Concentration ?
- Réaction concurrente ?
- ...

# Modélisation : création d'un modèle chimique

## La substitution nucleophile d'ordre 2, $\text{SN}_2$



$$\Delta_r G = 1,1 \text{ kcal/mol}$$

$$\Delta_r G^\ddagger = 7,2 \text{ kcal/mol}$$

Erreur estimée : dans la « précision chimique » ( $\sim 1 \text{ kcal/mol}$ )

Nombre d'atomes : 234

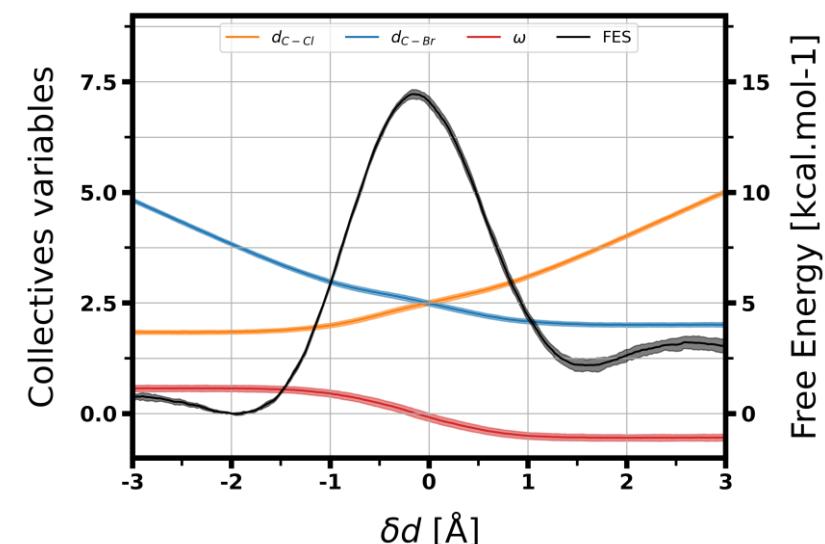
T=300K

Niveau de théorie : BLYP-D3 / TZV2P

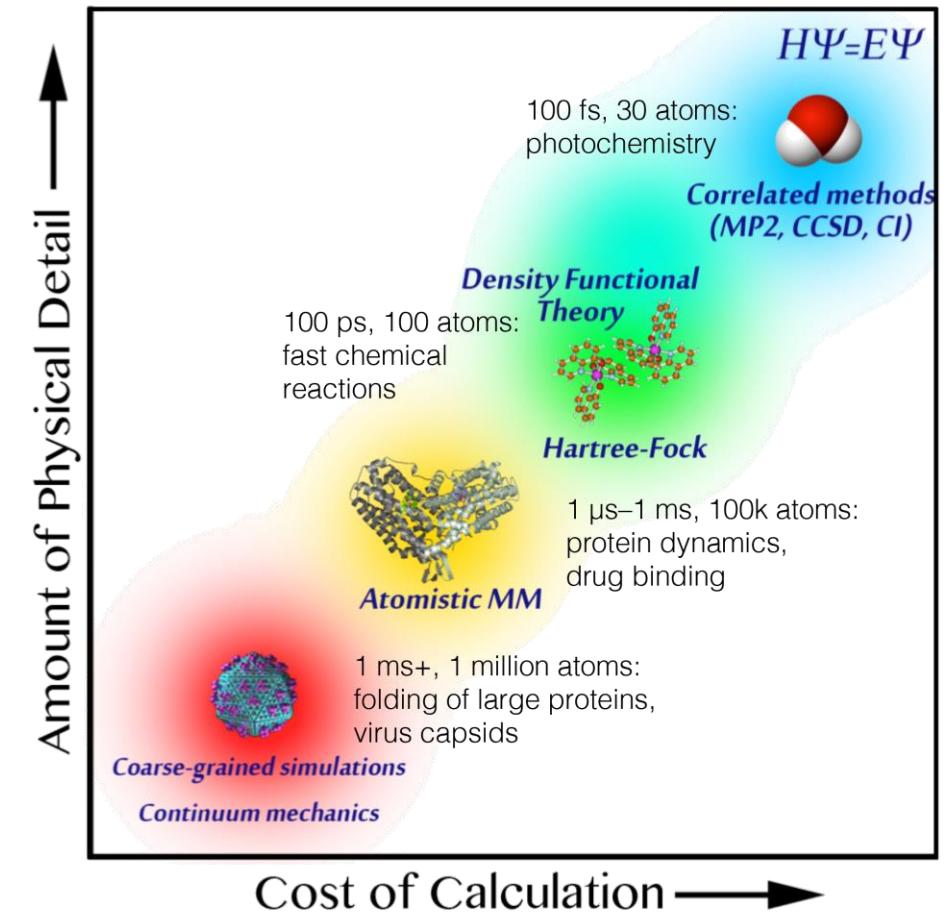
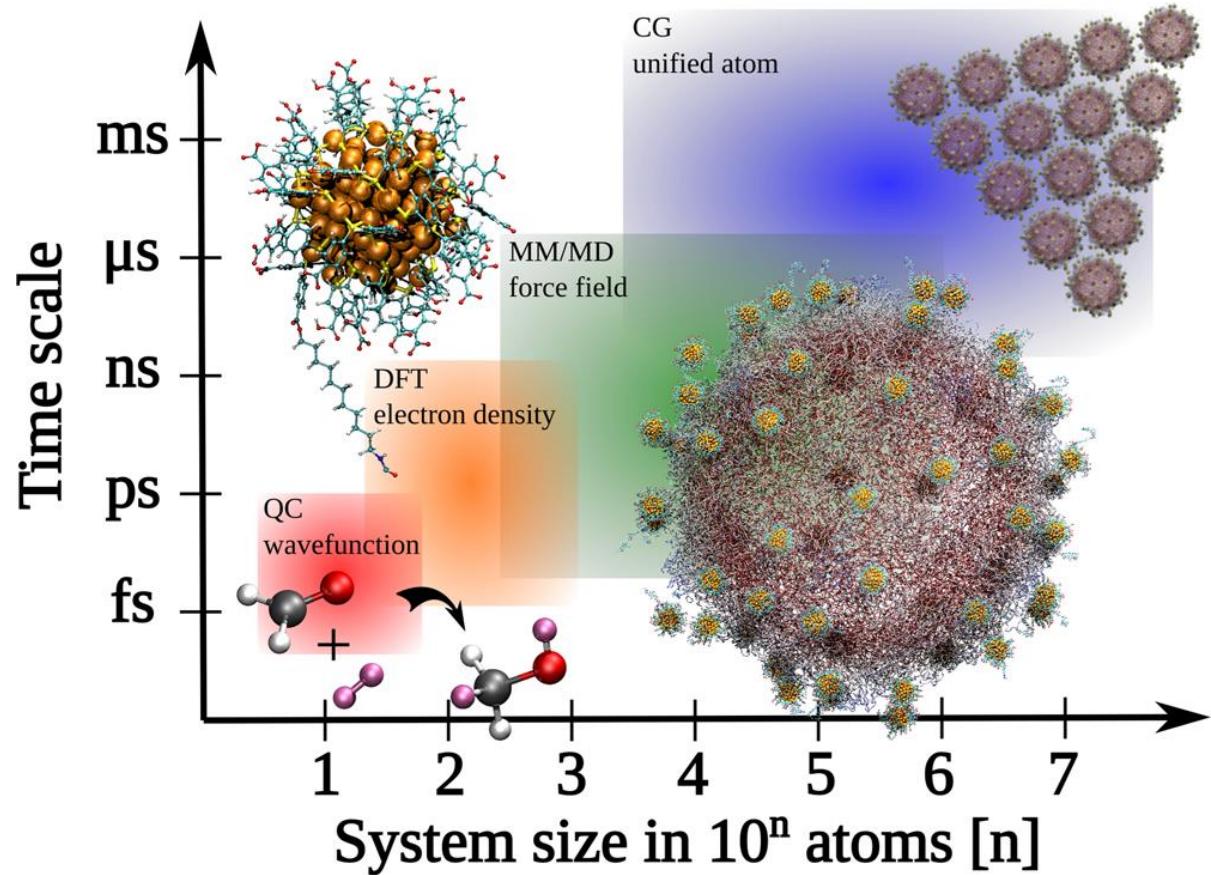
Temps simulé total : 72,6 ns

Temps de calcul total : 342 heures

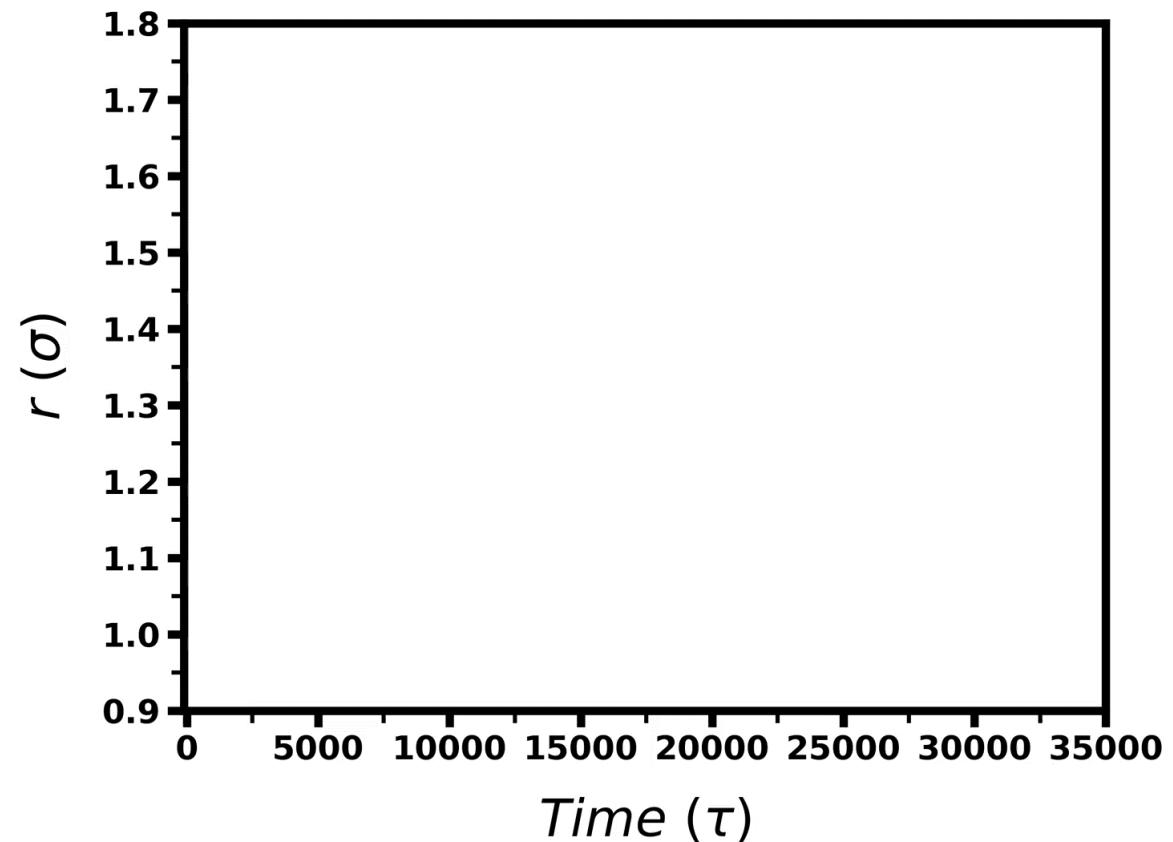
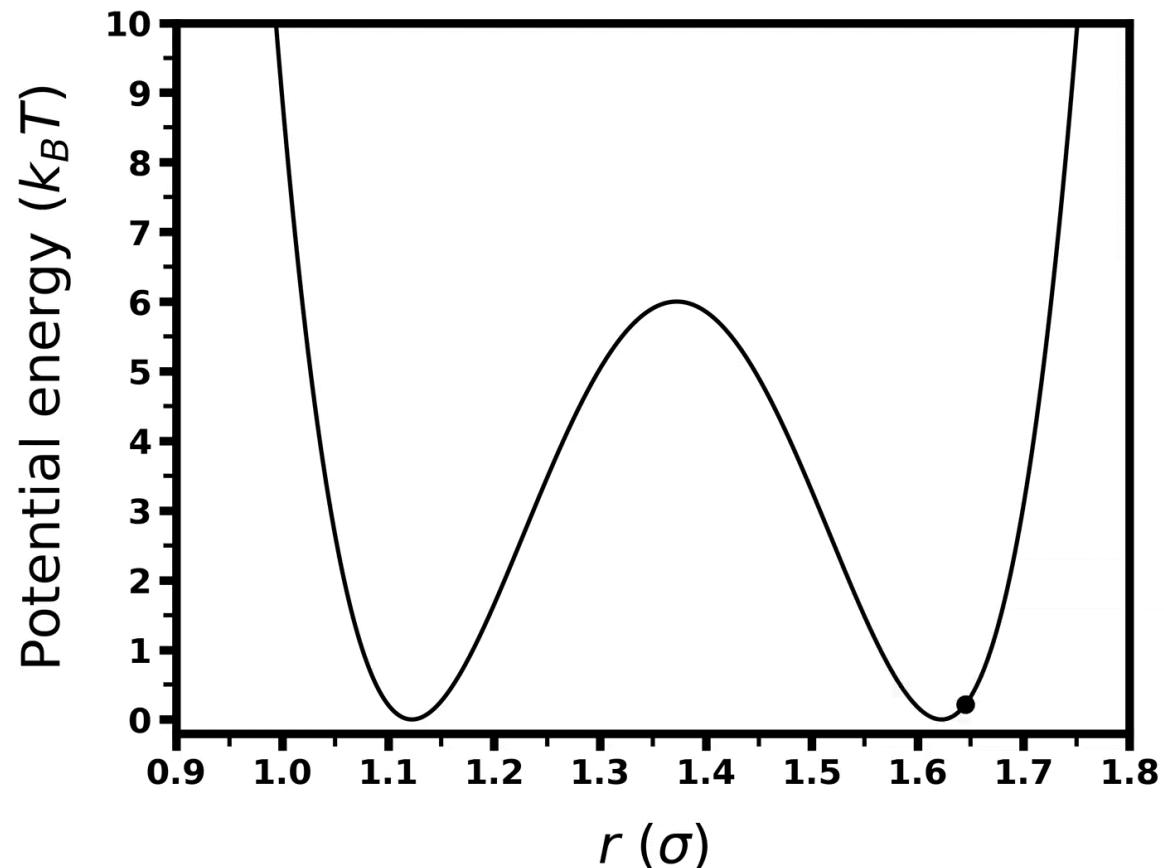
Temps « humain » total : 3 heures



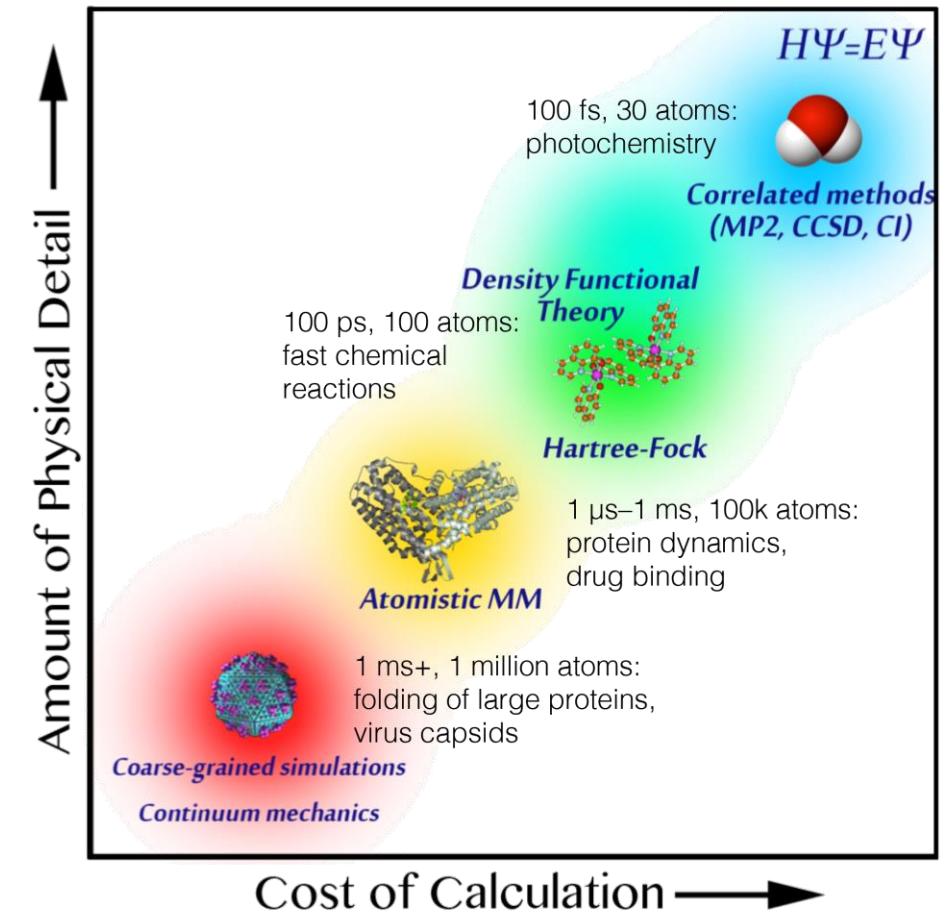
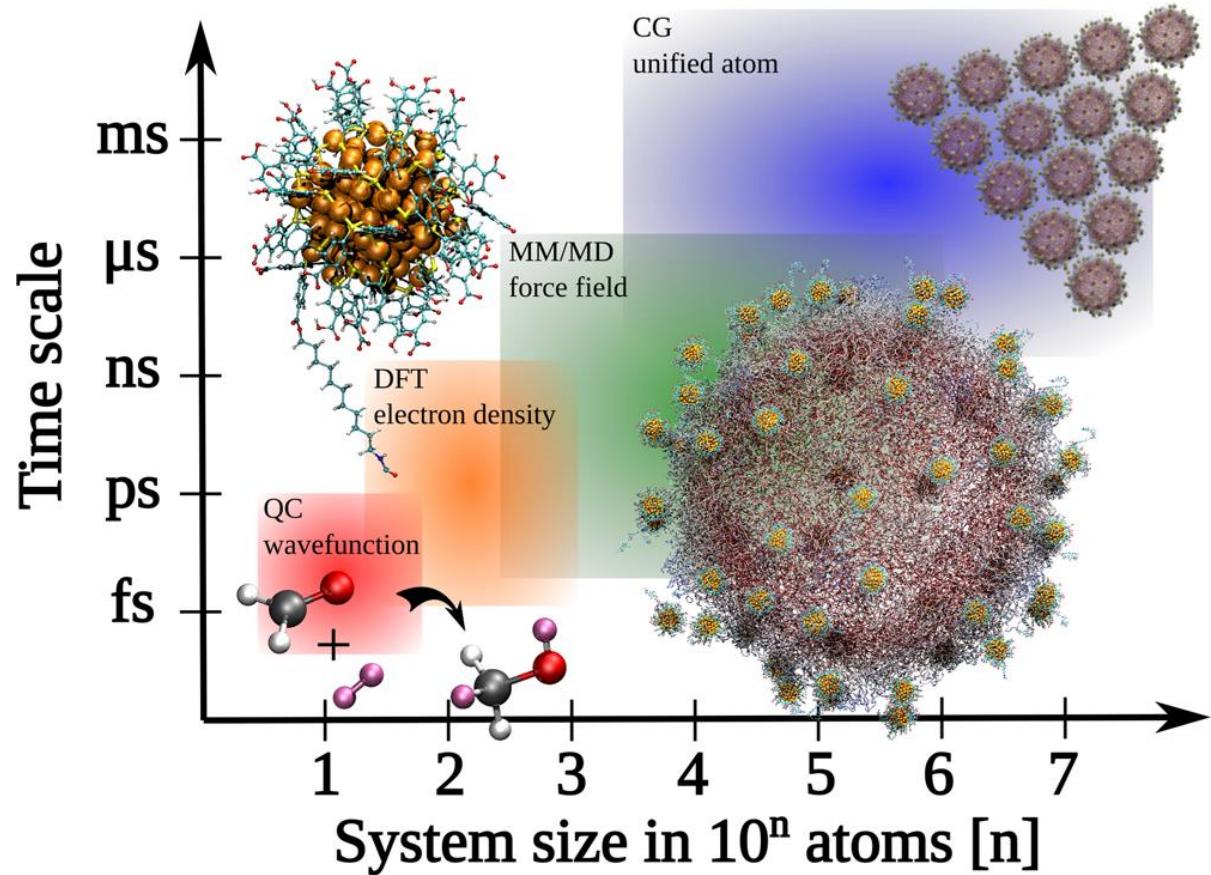
# Differents modèles pour différentes échelles



# Problématique de l'évènement rare



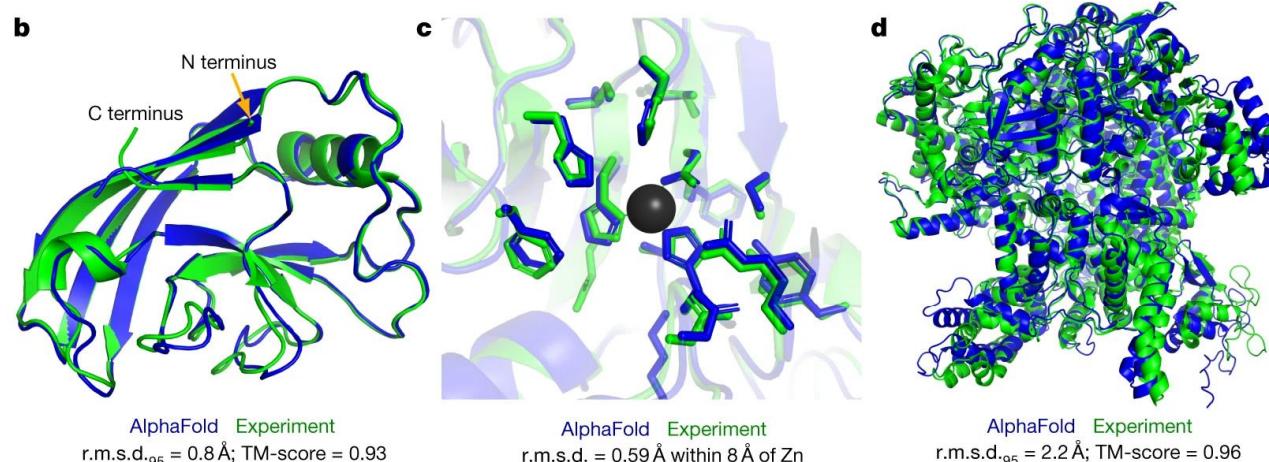
# Differents modèles pour différentes échelles



# Avancées en IA

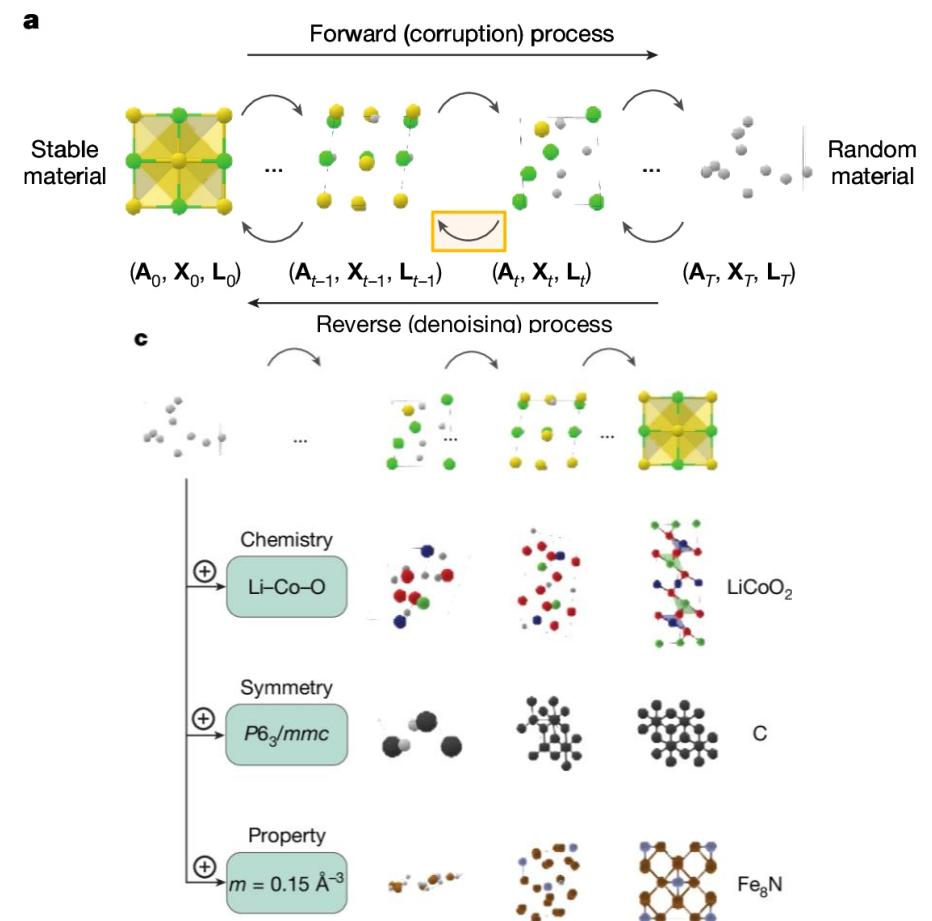
## AlphaFold

Prédiction de structures de protéines



## MatterGen

Génération de nouveaux matériaux



# Prédire le résultat d'une réaction : A + B → ??

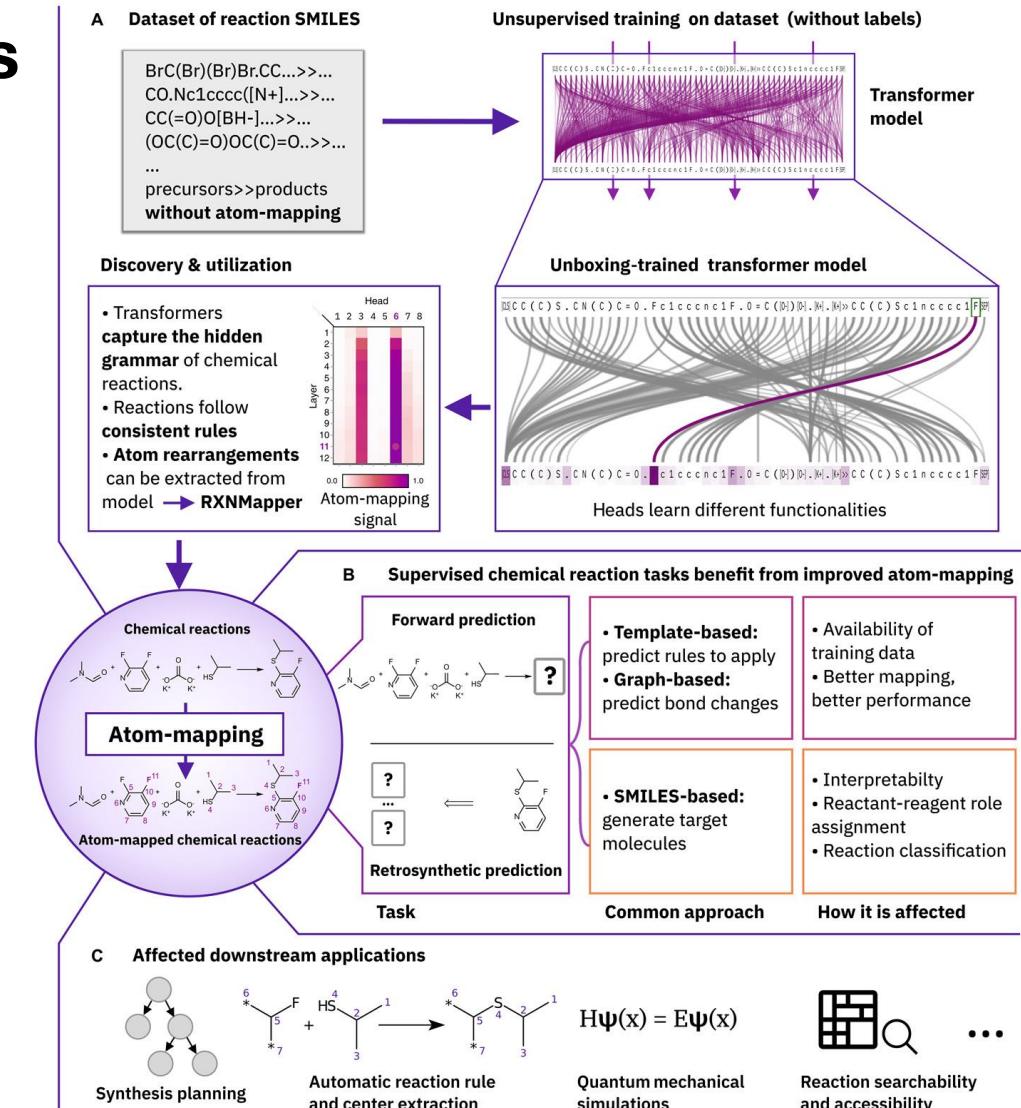
## Des modèles de langages à grande échelles (LLM) pour la prédiction de réaction

SCIENCE ADVANCES | RESEARCH ARTICLE

CHEMISTRY

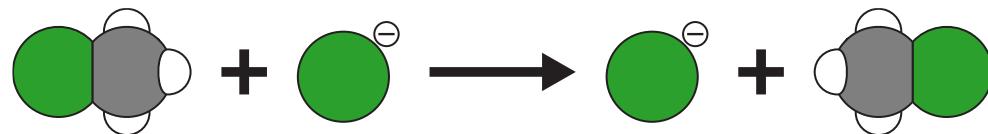
### Extraction of organic chemistry grammar from unsupervised learning of chemical reactions

Philippe Schwaller<sup>1,2\*</sup>, Benjamin Hoover<sup>3</sup>, Jean-Louis Reymond<sup>2</sup>,  
Hendrik Strobelt<sup>3</sup>, Teodoro Laino<sup>1</sup>



# De quoi avons nous besoin?

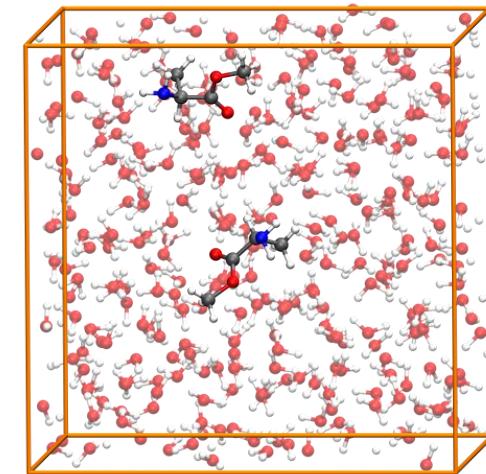
## Formation/rupture de liaisons



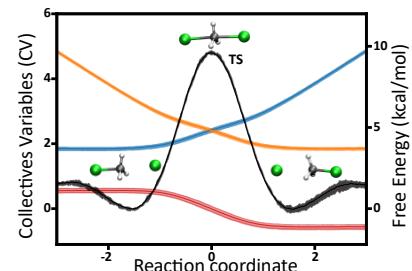
Réarrangement la structure électronique  
*Ab initio*, DFT, etc.

## Solvatation

Les solutés dans le solvant  
1000+ atomes

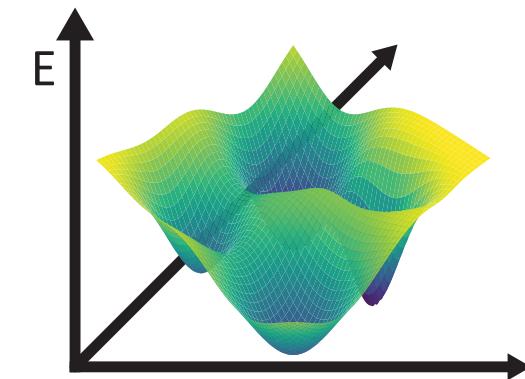


## Energie libre de réaction



L'importance de l'entropie : échantillonage  
Trajectoires de plus de 1000000 de pas  
Pour  $k_{\text{int}} / k_{\text{bulk}} = 2$ ,  $\Delta G^{\ddagger}_{\text{int}} - \Delta G^{\ddagger}_{\text{bulk}} = -0.4 \text{ kcal/mol}$

## Coordonnée de réaction



Identifier les différents chemins réactionnels

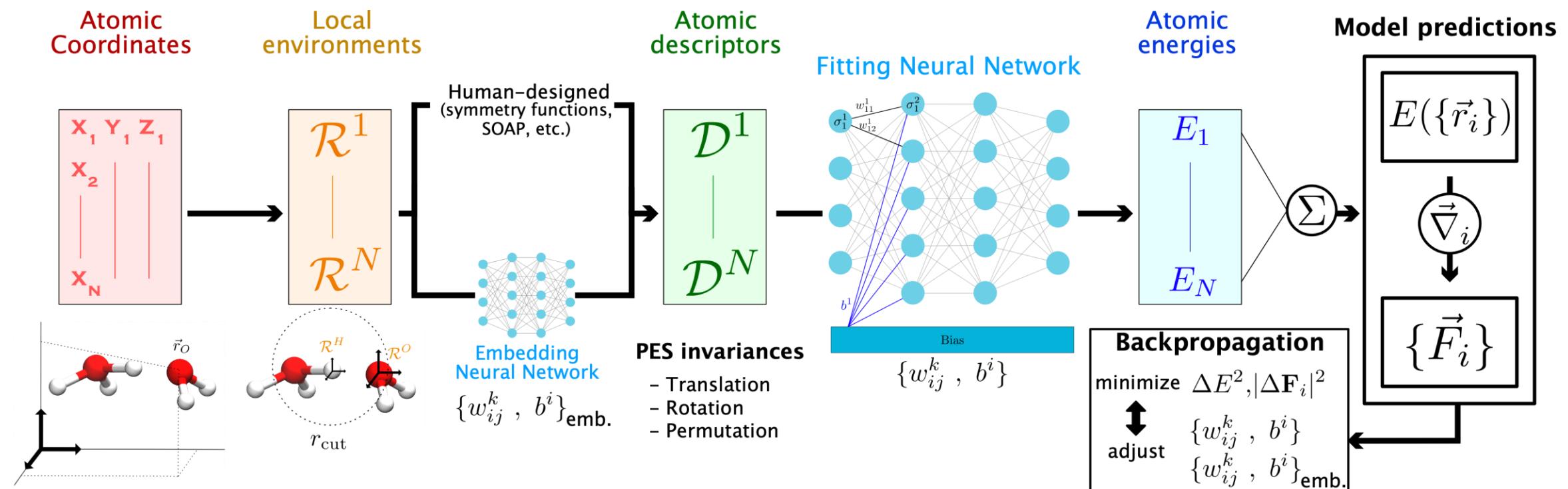
# Un « défi » majeur

---

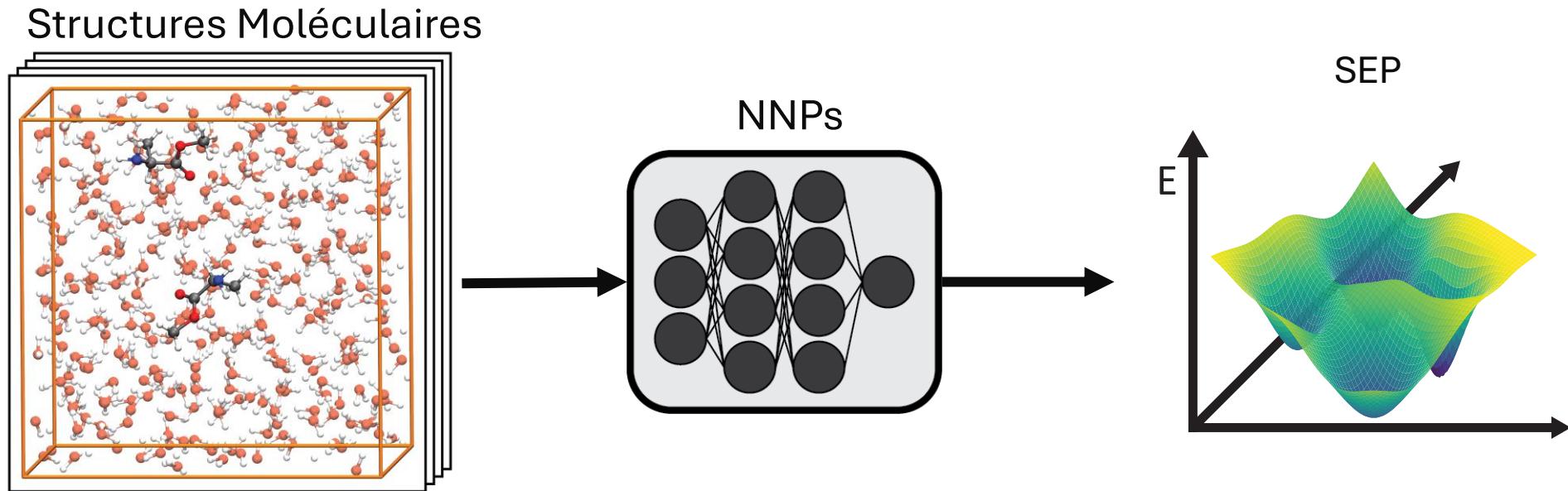
Positions  $\rightarrow H|\psi\rangle = E|\psi\rangle \rightarrow$  Forces

Est-ce que l'IA peut nous aider?

# Réseaux de neurones pour la chimie théorique !



# Un changement de paradigme : les Neural Network Potentials



Au moins 10000 fois plus rapide sans perte de précision  
(avec DFT-GGA comme référence)

Qualité  
de  
l'entraînement

**Seulement 3 h**

Qualité  
des  
résultats

Behler, J.; Parrinello, M. Generalized Neural-Network Representation of High-Dimensional Potential-Energy Surfaces. *Phys. Rev. Lett.* **2007**, 98 (14), 146401.

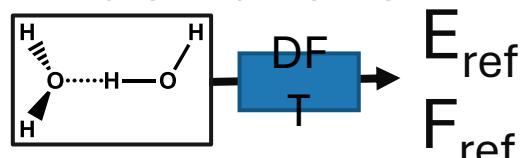
Zhang, L. et al. End-to-End Symmetry Preserving Inter-Atomic Potential Energy Model for Finite and Extended Systems. *NeurIPS 2018*, **2018**, 4436–4446.

Wang, H. et al. DeePMDD-Kit: A Deep Learning Package for Many-Body Potential Energy Representation and Molecular Dynamics. *Comput. Phys. Commun.* **2018**, 228, 178–184.

# Les étapes de création d'un NNP

## Étiquetage (Labelling)

*Création du jeu  
d'entraînement*

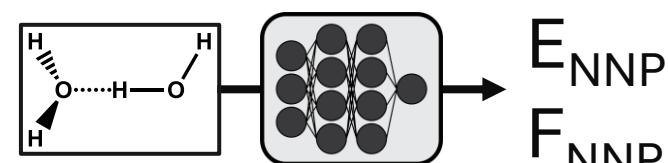


Structure moléculaire

↓  
Énergie et forces  
de références (ex.  
DFT)

## Entraînement

*Création du champ de force*



Structure moléculaire



Énergie et forces  
des NNPs

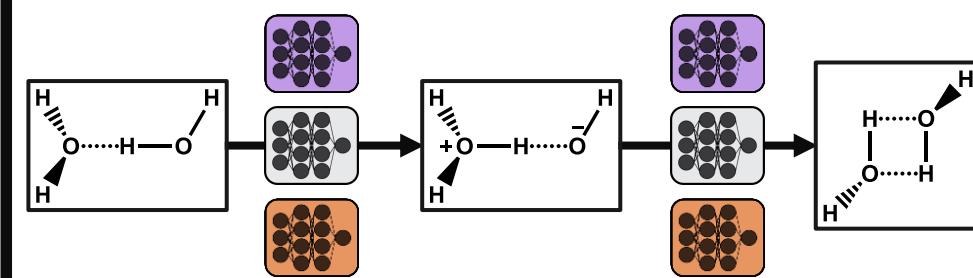
=

Énergie et forces  
de référence (ex.  
DFT)

Ajustement  
des poids et  
des biais

## Exploration / Sélection

*Enrichissement du jeu d'entraînement*



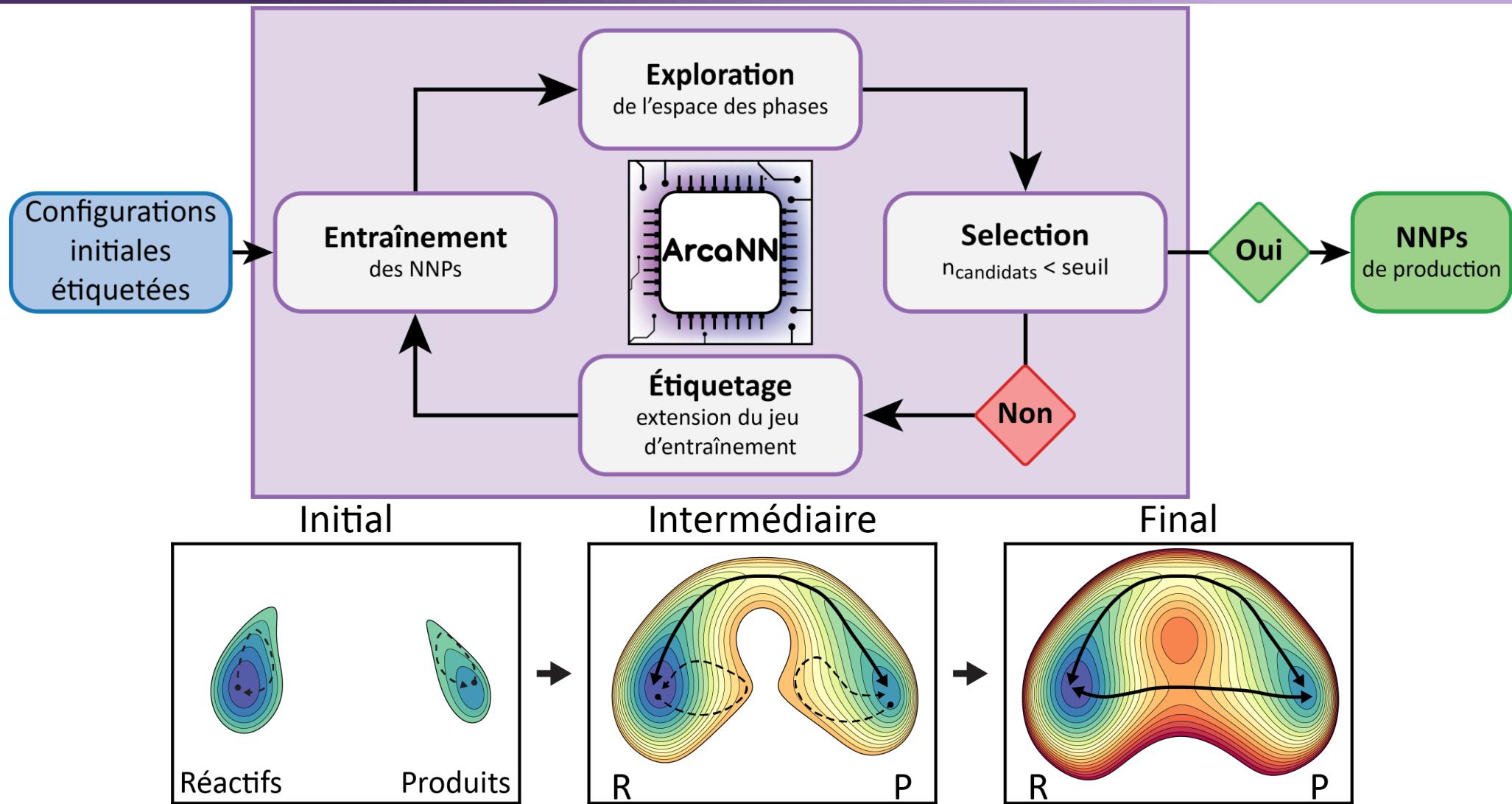
Dynamique



Sélection

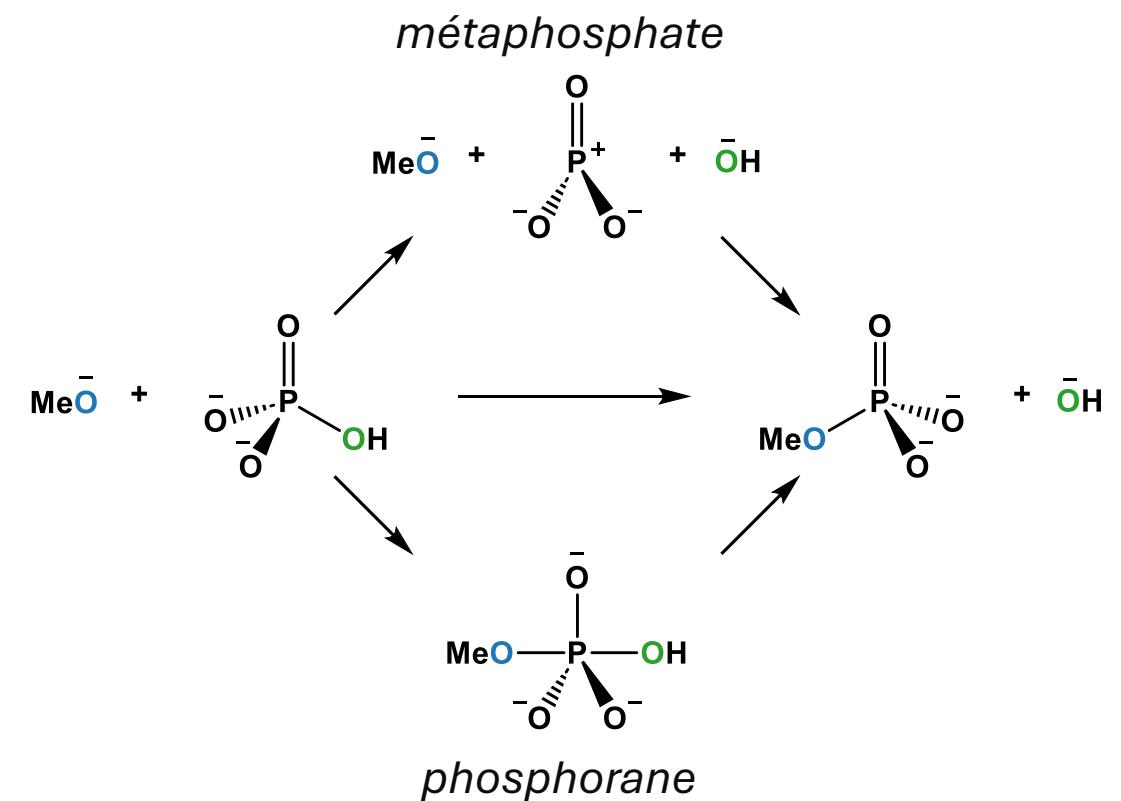
$$F_{NNP1} = F_{NNP2} = F_{NNP3} ?$$

# Comment se déroule un entraînement



# Polymérisation des acides nucléiques en conditions prébiotiques

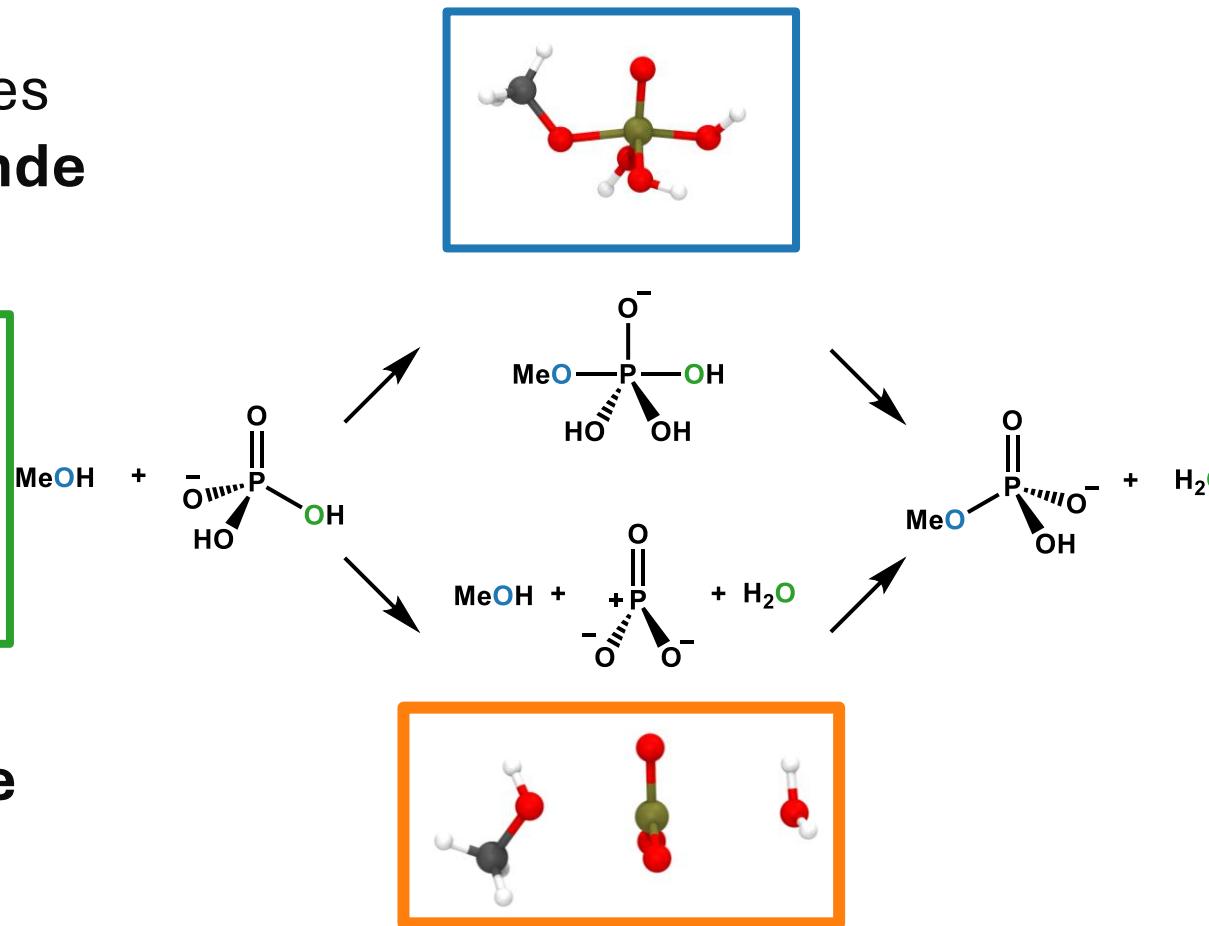
- Simplification :
  - **méthanolate** et **hydrogénophosphate**
- Trois mécanismes :
  - « dissociatif » : élimination + addition
  - « concerté » :  $\text{SN}_2$
  - « associatif » : addition + élimination
- Conditions
  - En milieu aqueux : rôle de l'eau ?
  - Protonation des espèces ?



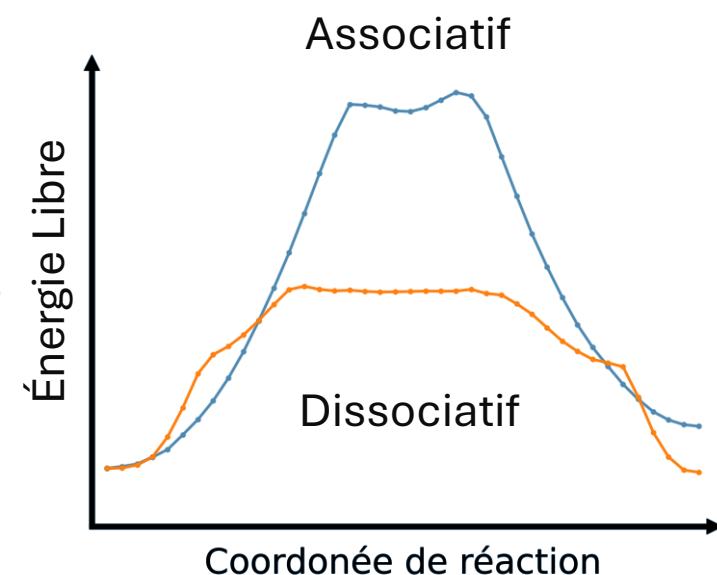
# Formation de la liaison phosphoester

- 1000 atomes
- 100000 trajectoires
- eq. 1 microseconde

Surface d'énergie  
libre convergée  
grâce à la qualité  
d'échantillonnage



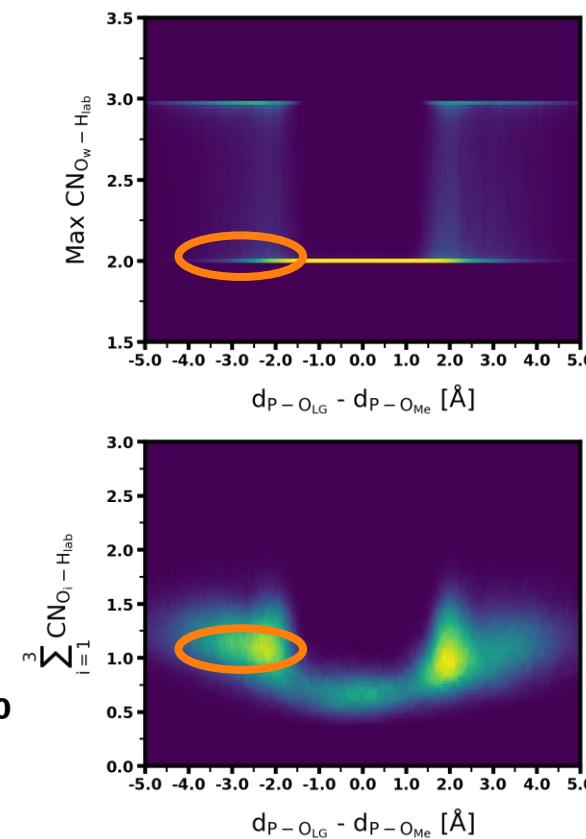
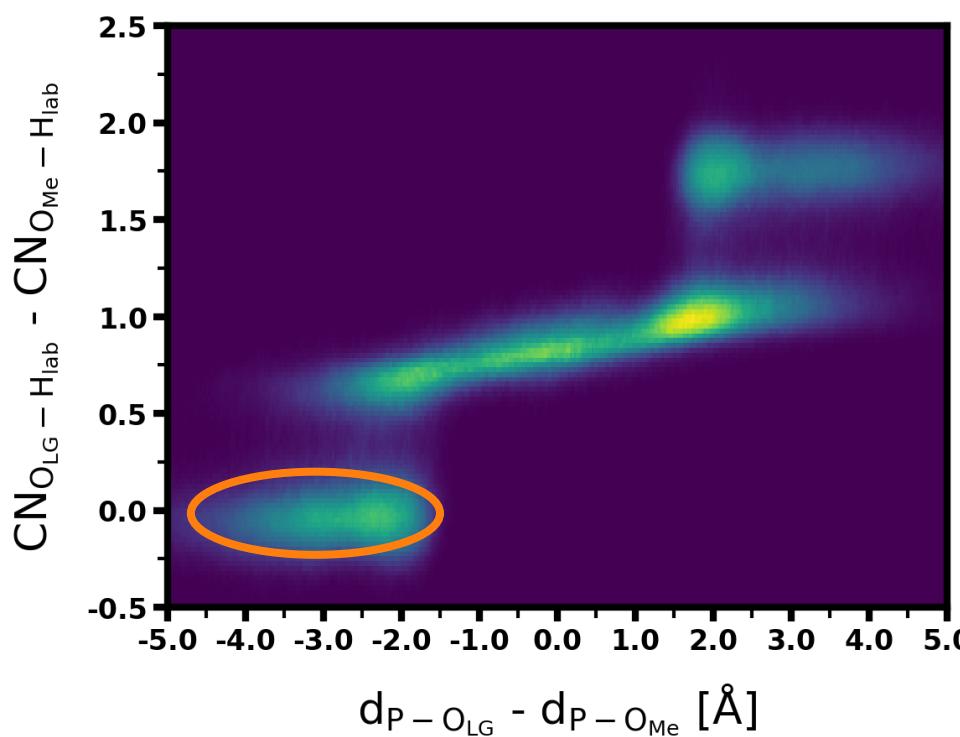
Thermodynamique  
Cinétique  
Mécanisme(s)



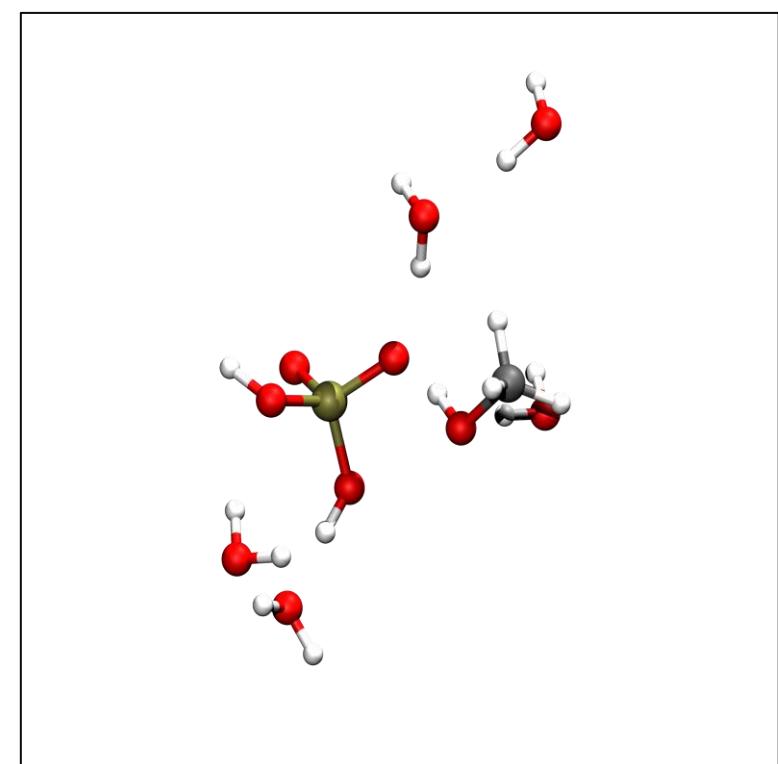
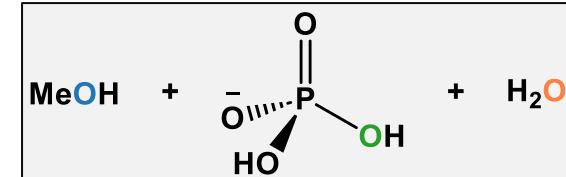
# Formation de la liaison phosphoester

Taille : 500 atomes  
Temps : 322 ns

## Densité de chemins : le mécanisme dissociatif Réactifs

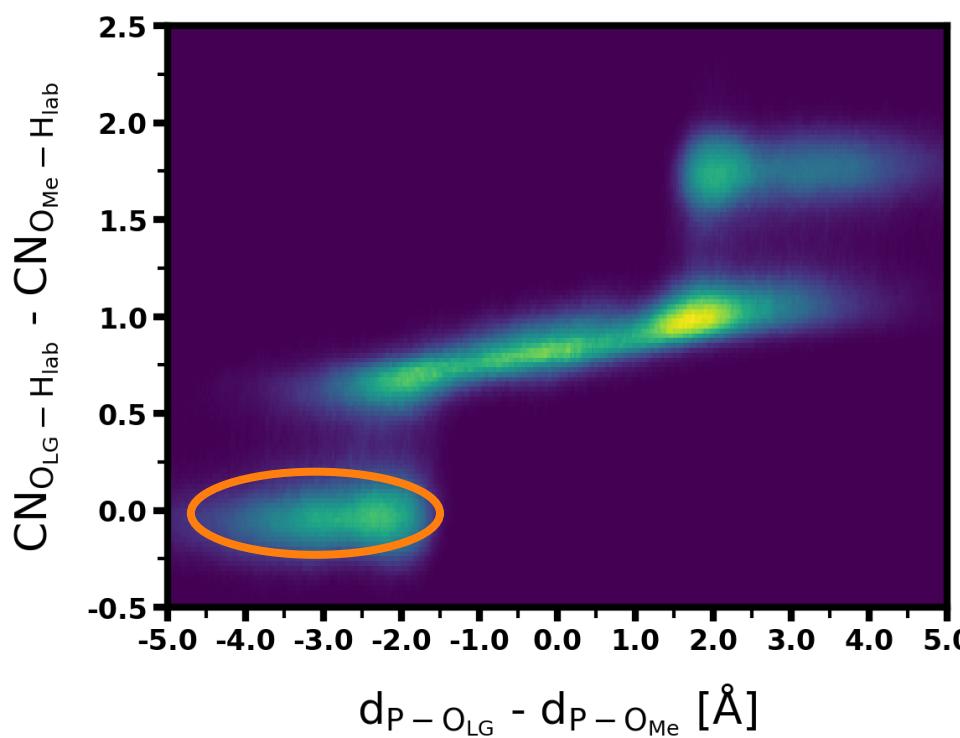


Nb MC : 25046  
Nb Acceptés : 5435  
Nb Décorrélé : 1128  
Temps : 322 ns

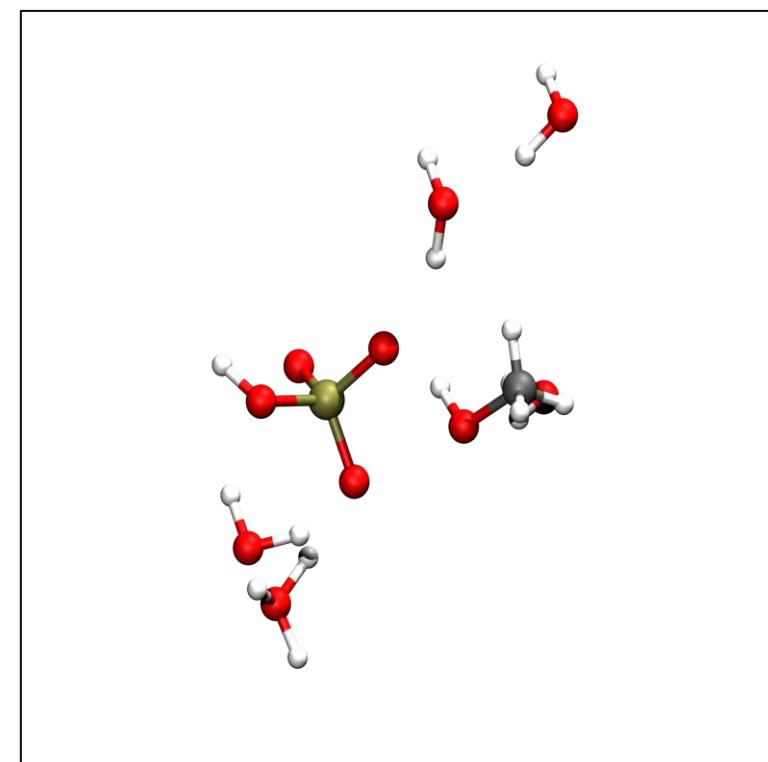
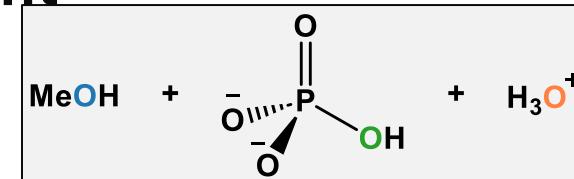
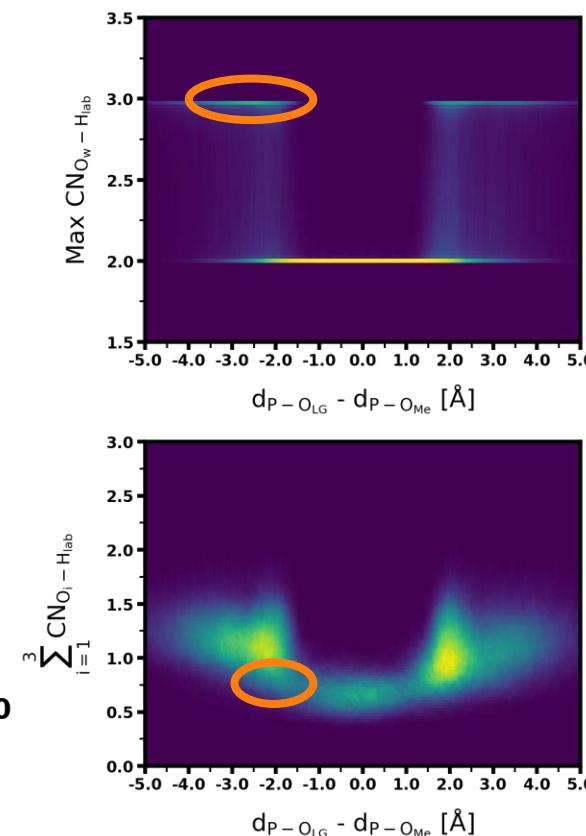


# Formation de la liaison phosphoester

Nb MC : 25046  
Nb Acceptés : 5435  
Nb Décorrélé : 1128  
Temps : 322 ns

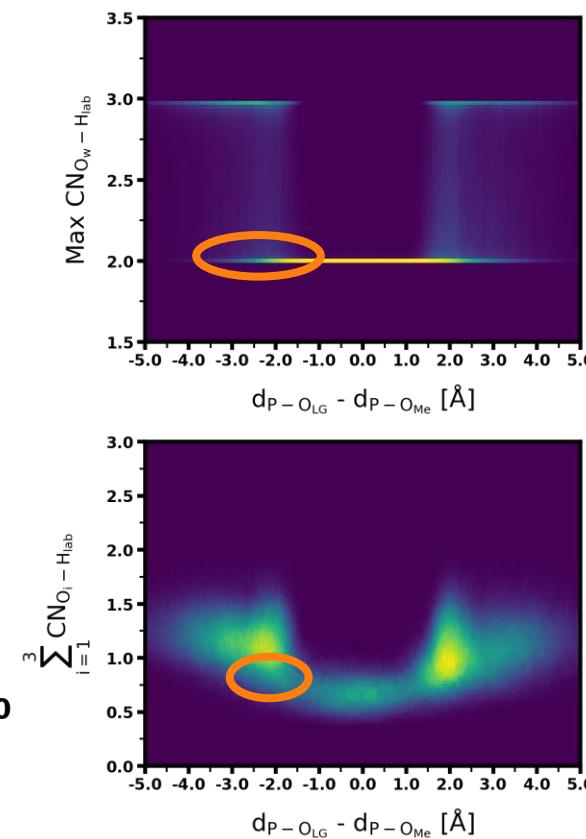
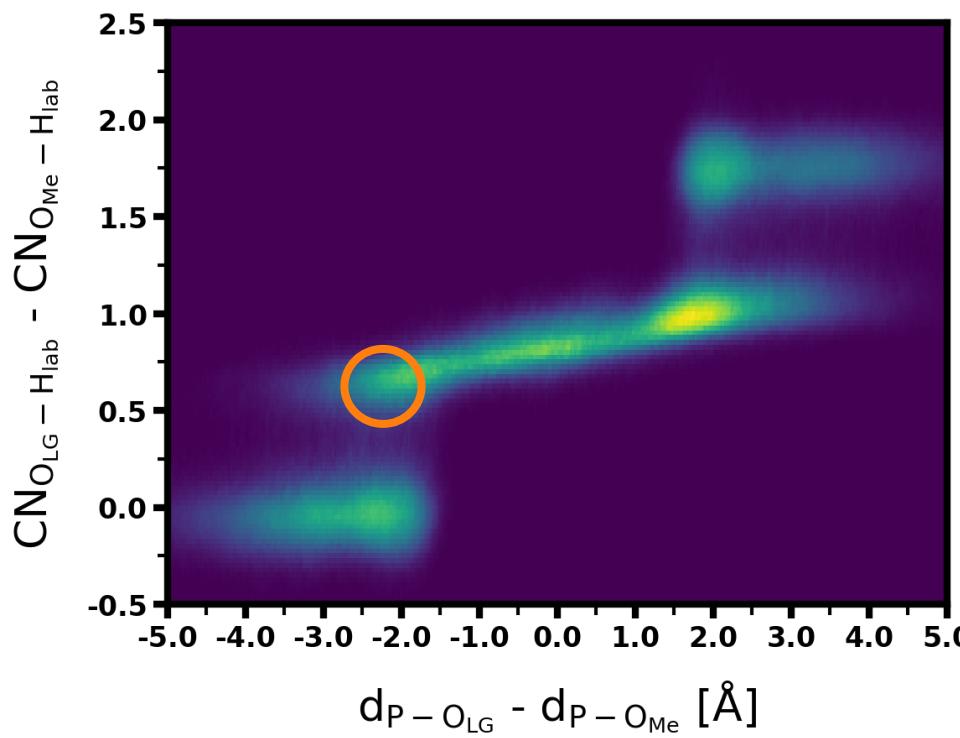


## Densité de chemins : le mécanisme dissociatif Déprotonation d'un oxygène non-liant

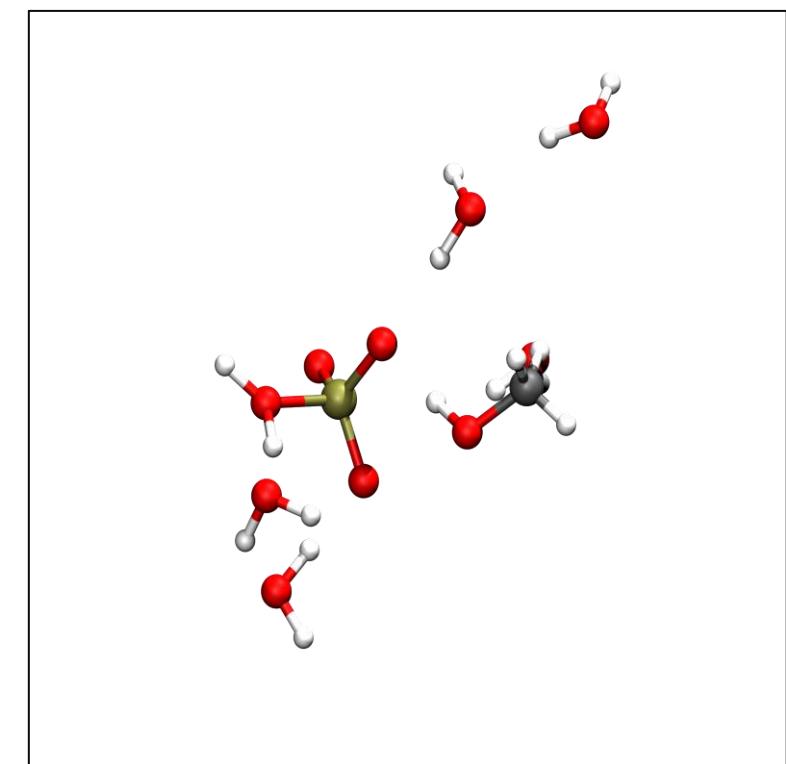
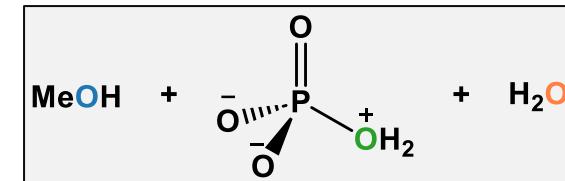


# Formation de la liaison phosphoester

Nb MC : 25046  
Nb Acceptés : 5435  
Nb Décorrélé : 1128  
Temps : 322 ns

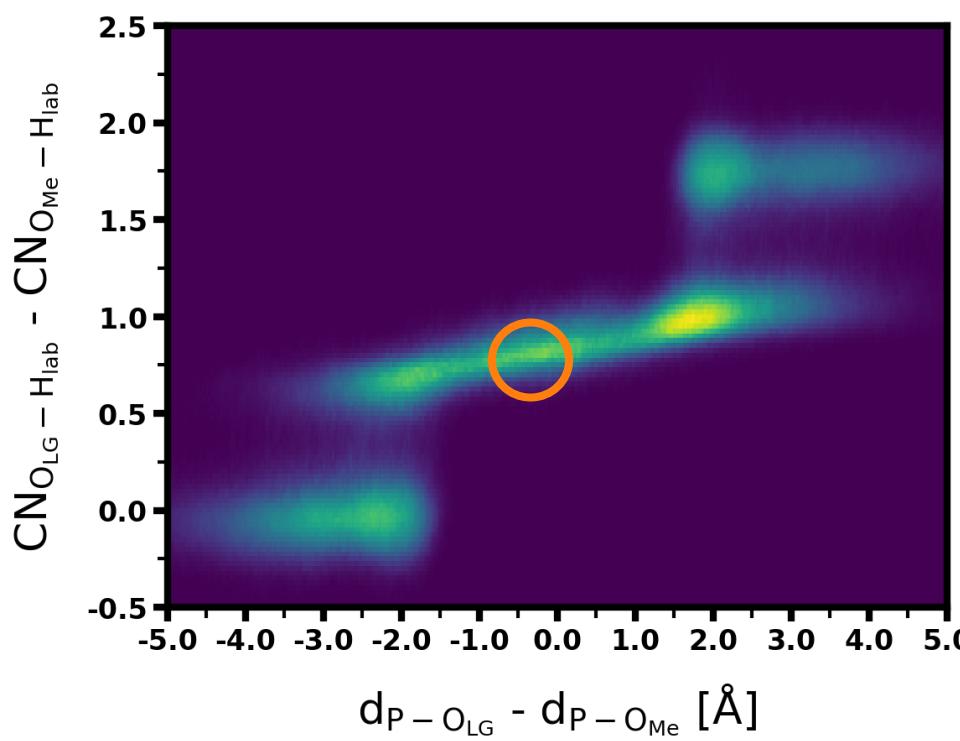


Densité de chemins : le mécanisme dissociatif  
Protonation du groupe partant

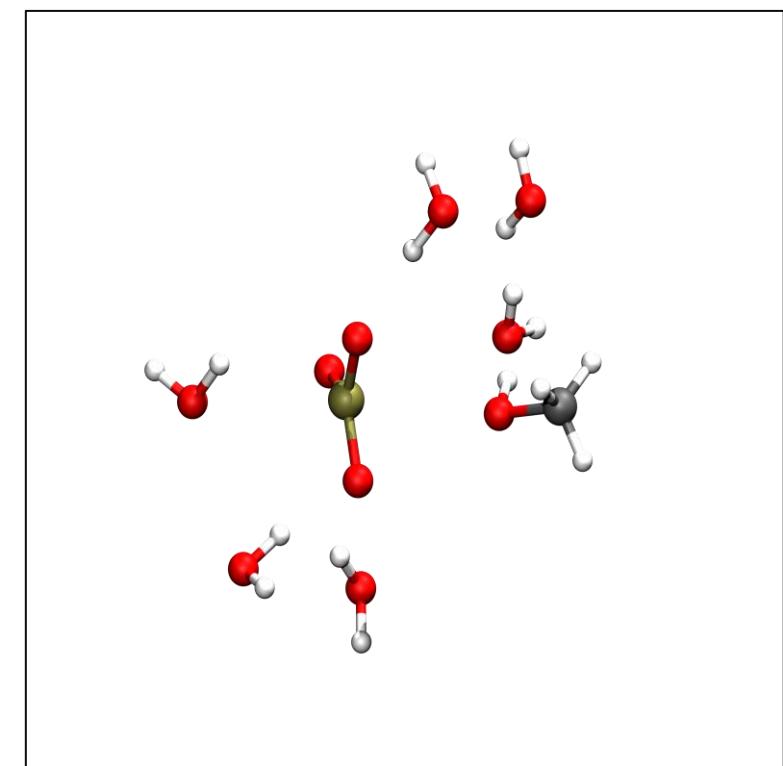
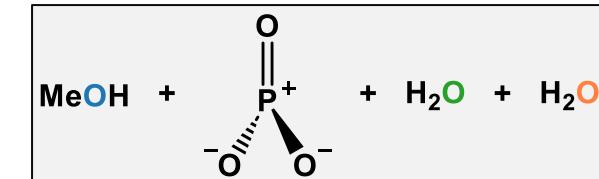
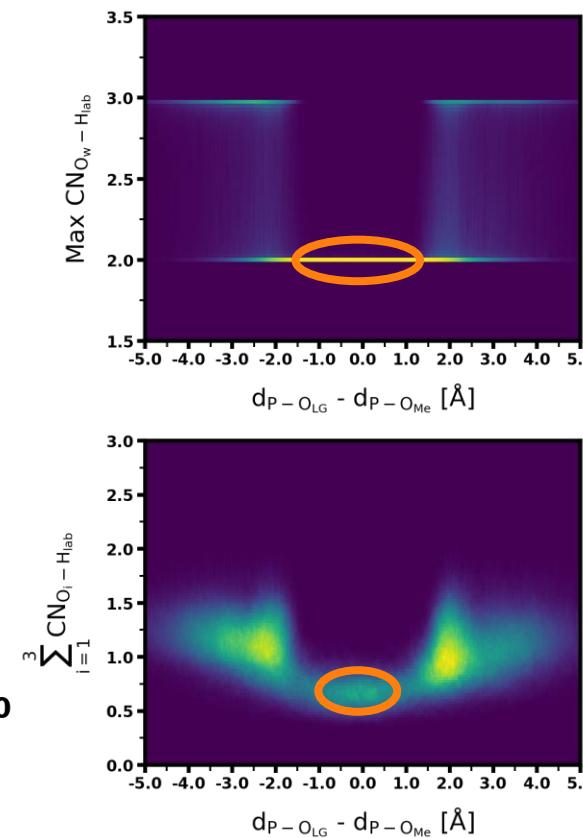


# Formation de la liaison phosphoester

Nb MC : 25046  
Nb Acceptés : 5435  
Nb Décorrélé : 1128  
Temps : 322 ns

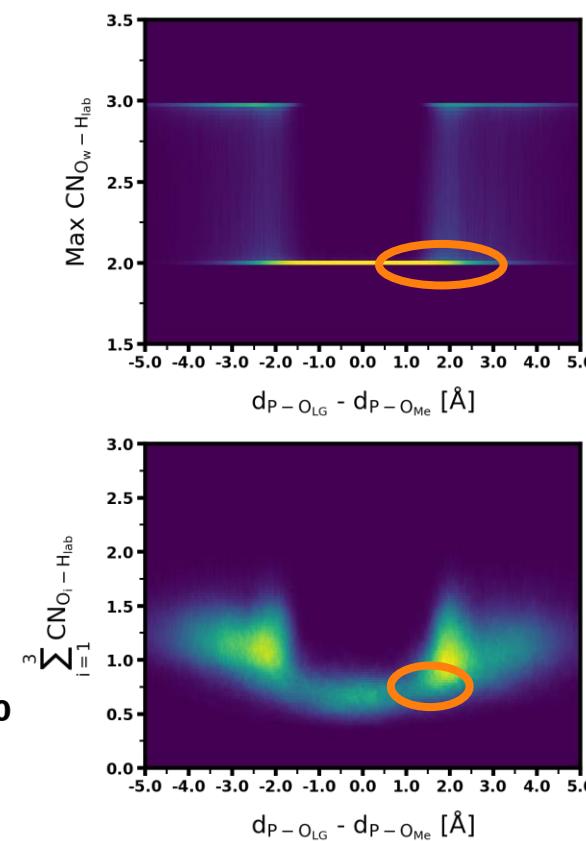
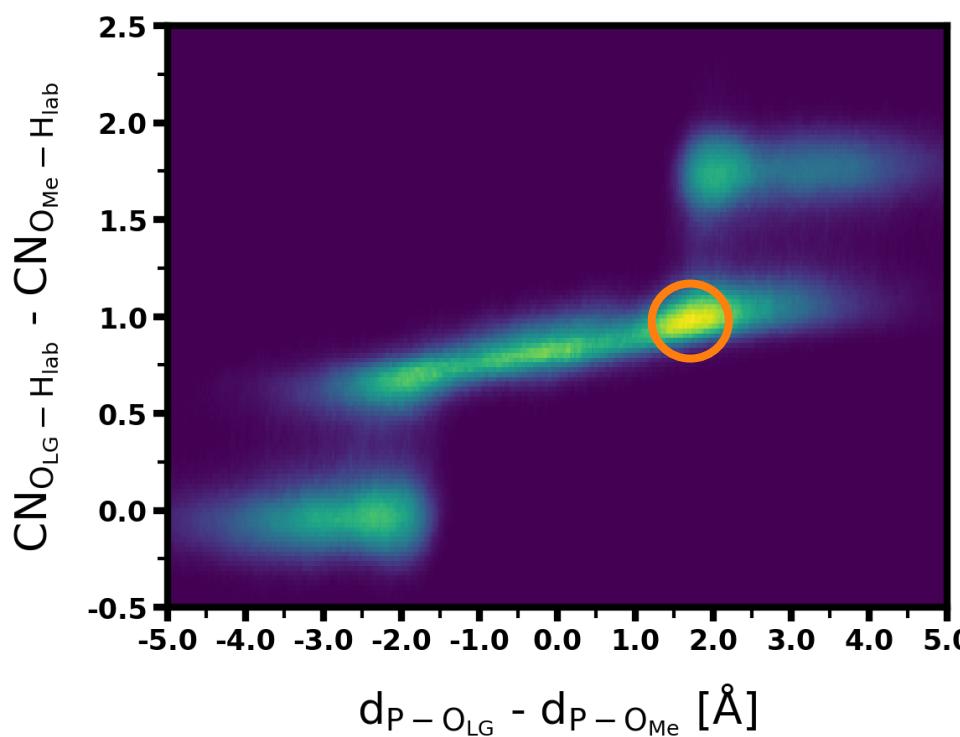


## Densité de chemins : le mécanisme dissociatif Départ du groupe partant

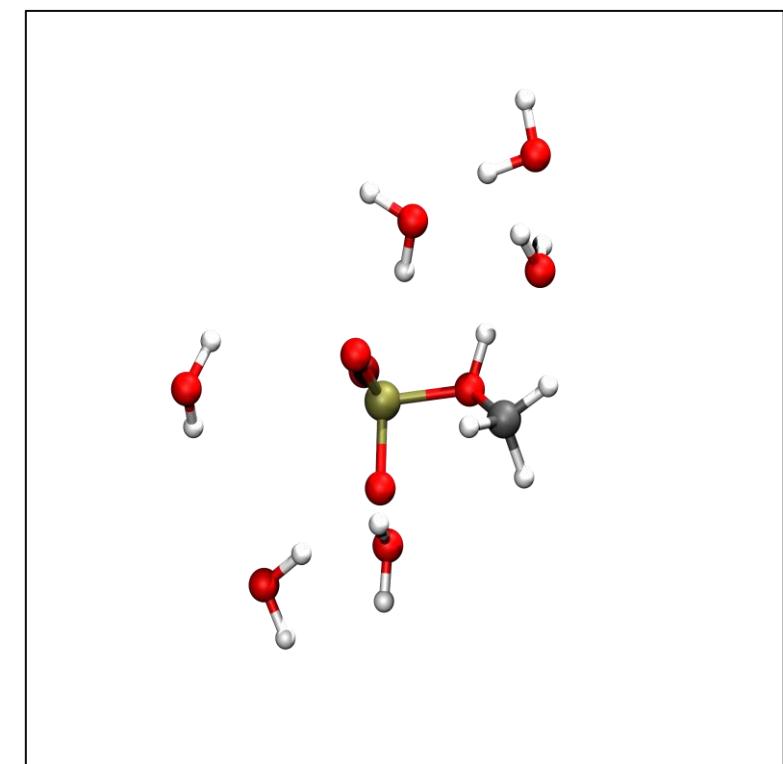
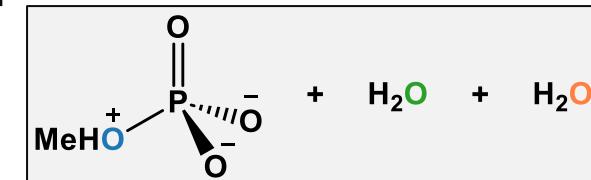


# Formation de la liaison phosphoester

Nb MC : 25046  
Nb Acceptés : 5435  
Nb Décorrélé : 1128  
Temps : 322 ns

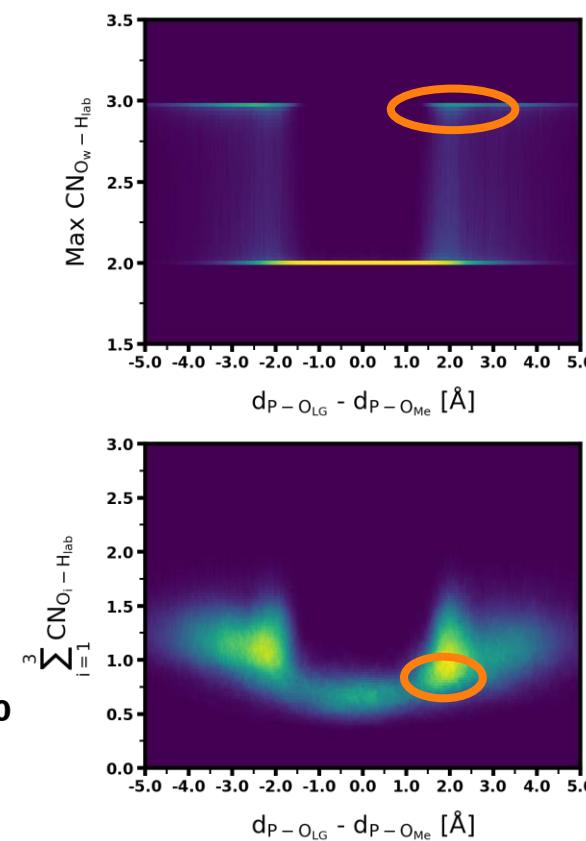
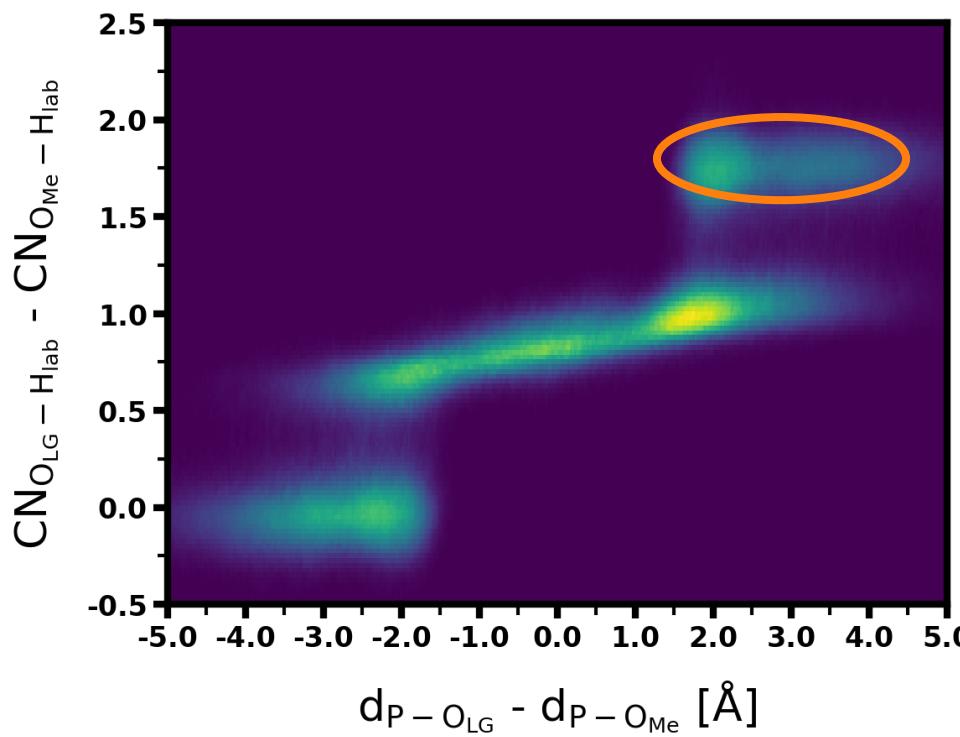


## Densité de chemins : le mécanisme dissociatif Attaque nucléophile du méthanol

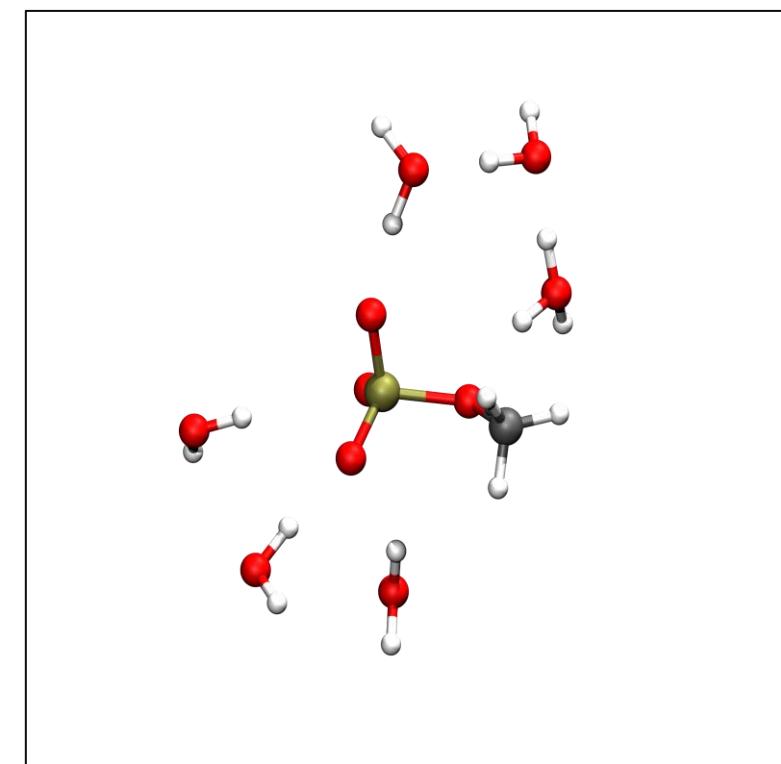
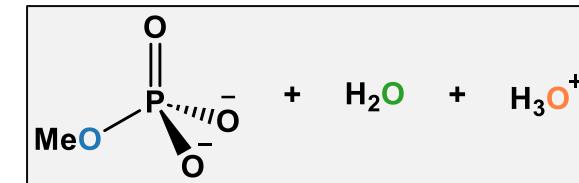


# Formation de la liaison phosphoester

Nb MC : 25046  
Nb Acceptés : 5435  
Nb Décorrélé : 1128  
Temps : 322 ns

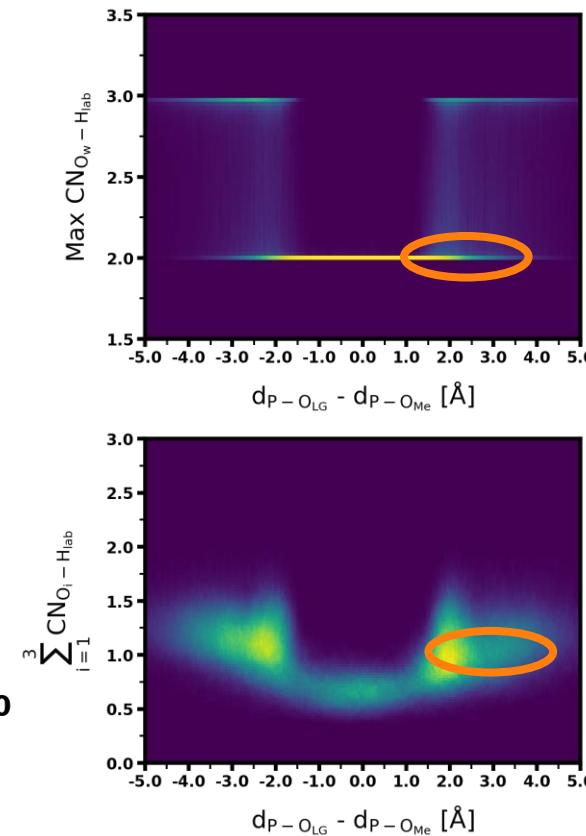
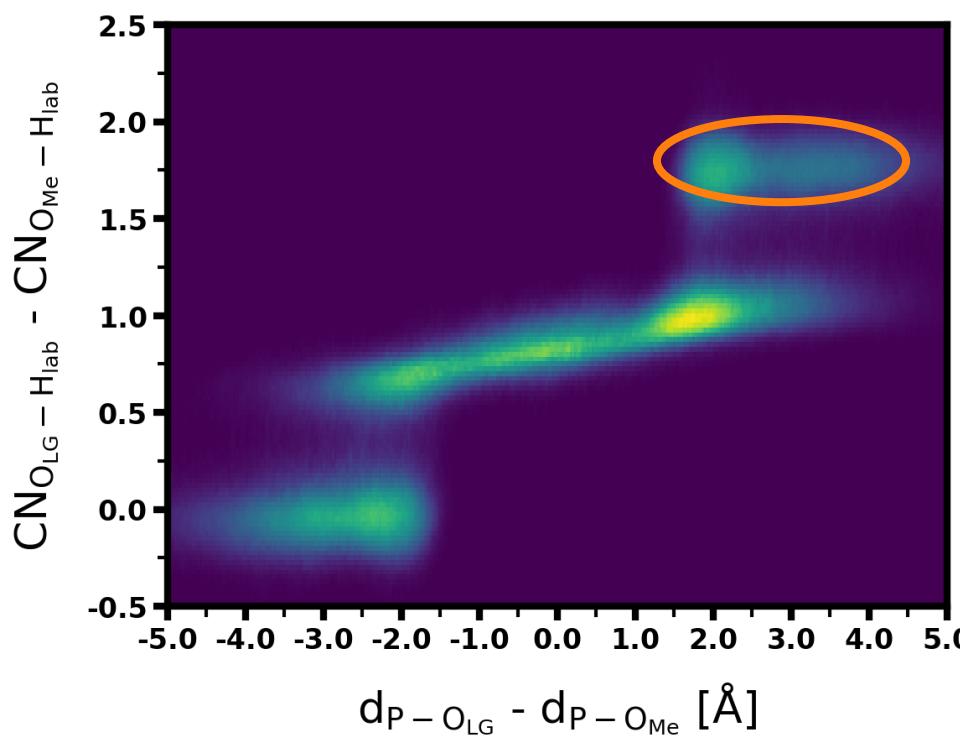


Densité de chemins : le mécanisme dissociatif  
**Déprotonation du nucléophile**

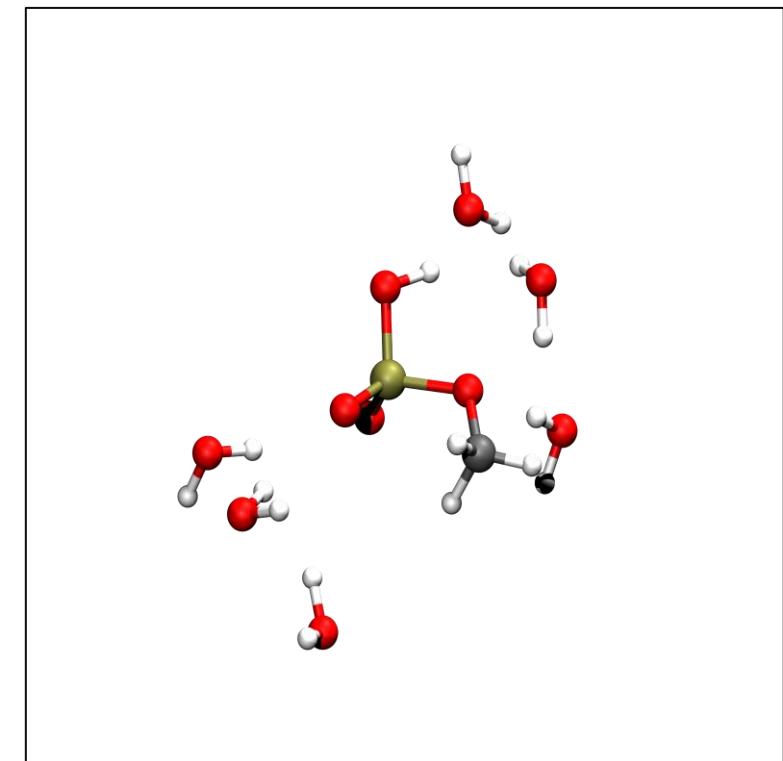
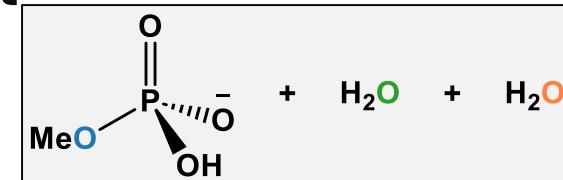


# Formation de la liaison phosphoester

Nb MC : 25046  
Nb Acceptés : 5435  
Nb Décorrélé : 1128  
Temps : 322 ns



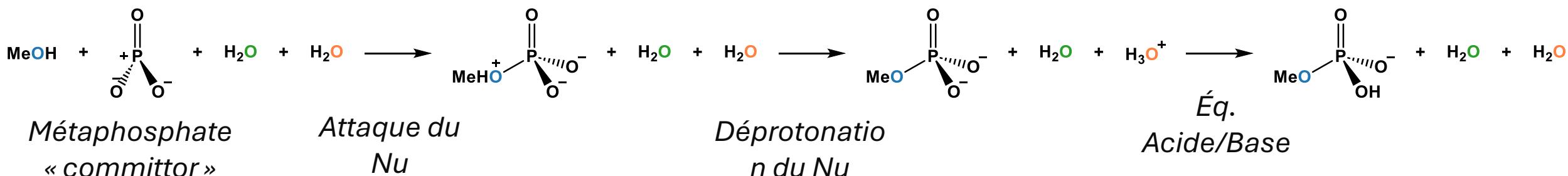
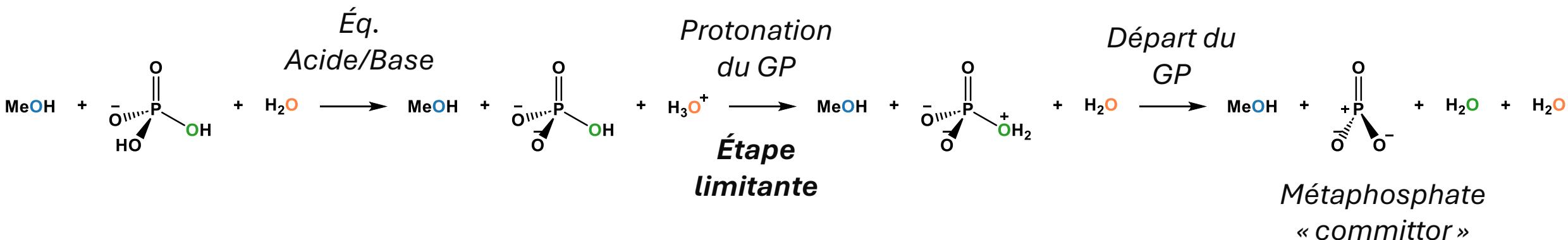
Densité de chemins : le mécanisme dissociatif  
**Protonation d'un oxygène non-liant**

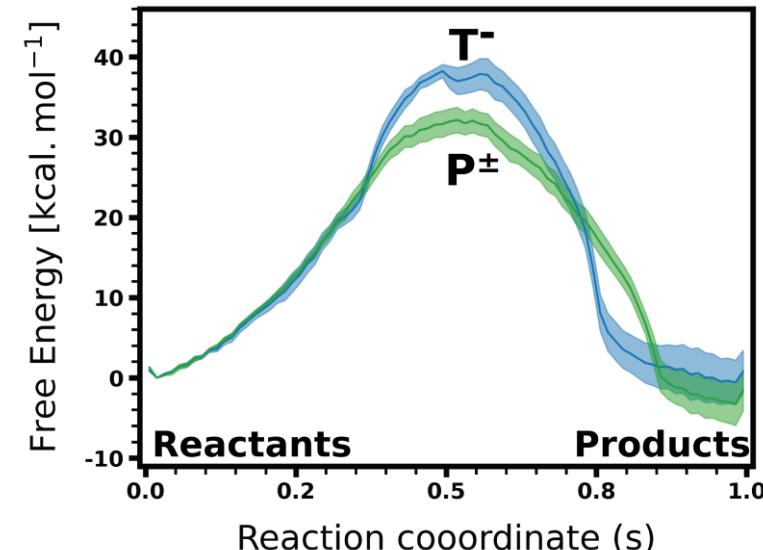
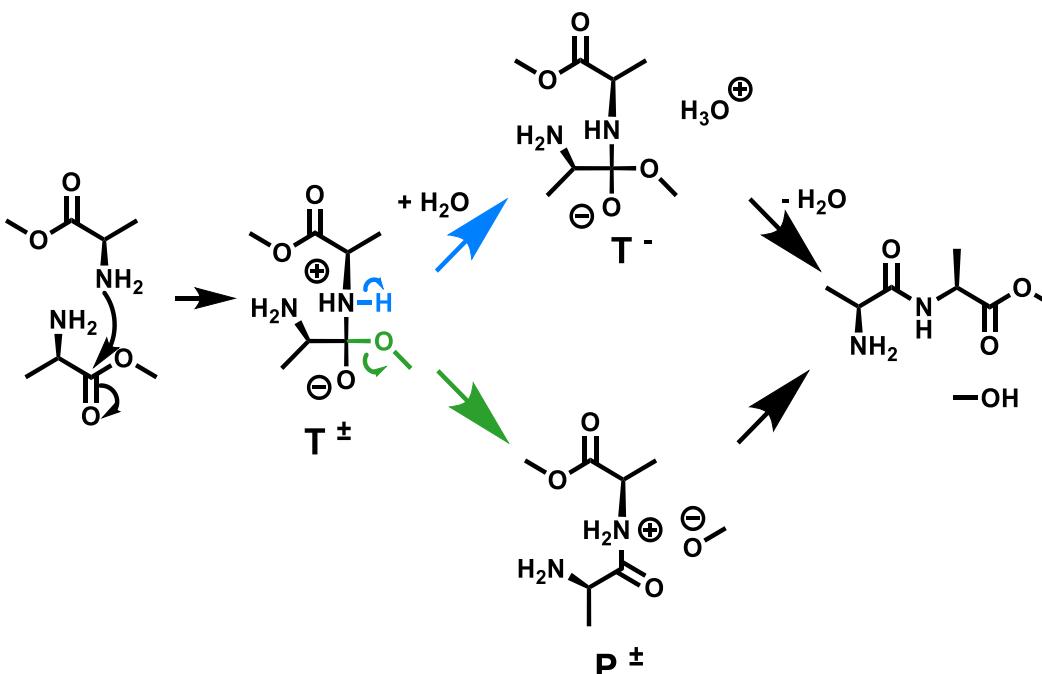
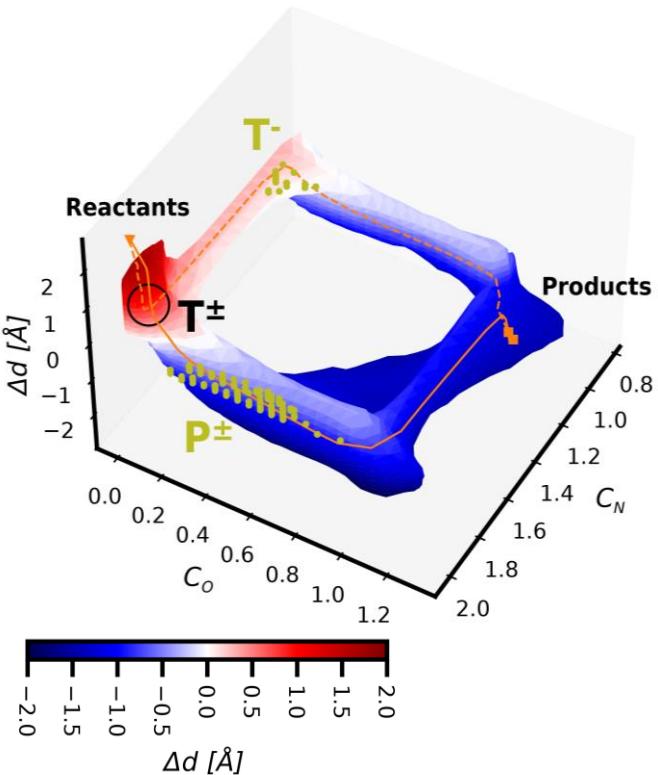


# Formation de la liaison phosphoester

Nb MC : 25046  
Nb Acceptés : 5435  
Nb Décorrélé : 1128  
Temps : 322 ns

## Le mécanisme dissociatif





- Mécanisme nouveau et le plus favorable à pH neutre
- En accord avec les expériences d'effet isotopique cinétique
- En accord avec la catalyse au ribosome

# Ce que Lionel Salem savait ... article de Roald Hoffmann 1998

Les réalisations de la chimie computationnelle moderne sont stupéfiantes. Il est aujourd’hui possible, de manière raisonnable, de traiter des milliards de configurations et d’atteindre une précision chimique — de l’ordre du kilocalorie, disons — dans le calcul des énergies de liaison et des géométries, tant pour les états fondamentaux que pour les états de transition de molécules relativement complexes. Il ne fait aucun doute que l’entreprise de la chimie théorique computationnelle est un succès.

L’expérimentateur demande : « Quel est l’angle de liaison de la molécule d’eau ? »

Vous, le théoricien, entrez la molécule dans les meilleurs programmes disponibles, et vous obtenez la bonne valeur, à trois chiffres significatifs près.

Tout le monde est content.

Puis l’expérimentateur pose la même question pour TeH<sub>2</sub>. Vous répondez :

« Attendez une minute, je dois le calculer... » Et vous trouvez la bonne valeur.

Vous trouvez aussi la bonne valeur pour Li<sub>2</sub>O et Fe<sub>2</sub>O.

Mais si vous ne faites que cela, aussi bien que vous le fassiez, l’expérimentateur deviendra de plus en plus insatisfait.

Parce que vous ne lui avez pas fourni une explication simple, transportable, fondée sur l’électronégativité, ou

sur les énergies relatives des orbitales s et p, ou sur le caractère donneur ou accepteur,

ou sur tout autre ensemble de facteurs avec lesquels il ou elle se sent à l’aise.

L’expérimentateur pensera : « Ce théoricien sait seulement simuler des expériences » ;

ou, de façon moins charitable : « L’ordinateur comprend, mais ce théoricien, non. »

It is nice to know that the computer understands the problem. But I would like to understand it too.

Eugene Wigner

# Ce que Lionel Salem savait ... article de Roald Hoffmann 1998

Supposons que vous ayez ce programme parfait. Il a certes nécessité un effort incroyable pour écrire le code... Maintenant, vous êtes sûr que le problème est résolu. Pour n'importe quelle molécule, le code vous donnera la géométrie correcte, la polarisabilité, le spectre. C'est le Graal auquel nous aspirons.

Mais c'est complexe. Ça doit l'être, car la nature l'est. Ce sont nos esprits qui sont simples, pas la nature.

Face à cette complexité, il est naturel que le théoricien du calcul ait tendance à penser qu'il ne peut y avoir aucune explication unique et simplifiée, et qu'il répugne à en fournir une aux expérimentateurs, surtout dans leur propre langage — un langage fondé sur des contextes vagues et traditionnels qui n'ont peut-être aucun rapport avec ce que vous calculez.

Je ne peux que vous donner ce conseil : essayez, s'il vous plaît, essayez. Faites le saut existentiel, prenez le risque de fournir une explication. Faites des expériences numériques pour sonder vos idées les plus folles, bien sûr, mais n'ayez pas peur : proposez une explication simple.

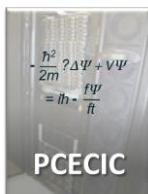
Je sais que je ne suis pas cohérent en prônant un univers complexe, tout en encourageant les explications simples.

Roald Hoffmann ‘Qualitative thinking in the age of modern computational chemistry—or what Lionel Salem knows’ Journal of Molecular Structure (Theochem) 424 (1998) I-6

# Remerciements

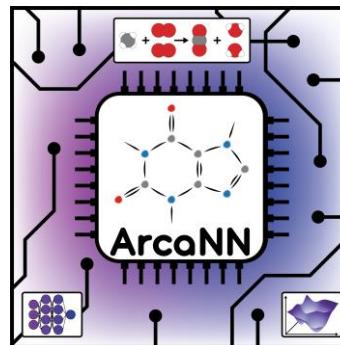


CBH Graduate School  
Université Grenoble Alpes



## @DCM

- Oscar Gayraud
- David Gatineau
- Hélène Jamet
- Yves Gimbert
- Pierre Girard (@GRICAD)
- Veronique Blandin
- Carole Duboc
- Frédérique Loiseau
- Yannick Vallée



## @ENS Paris

- Miguel de la Puente
- Zakarya Benayad
- Olaia Anton
- Axel Gomez
- Guillaume Stirnemann
- Damien Laage



## @LSU

- Azom Golam
- Revati Kumar



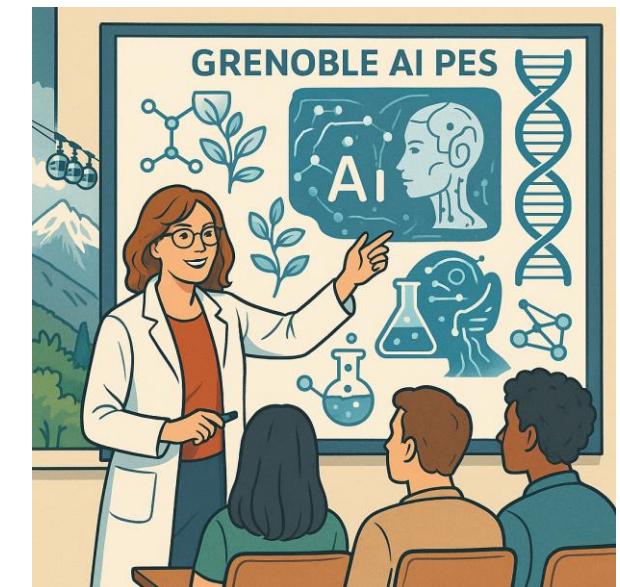
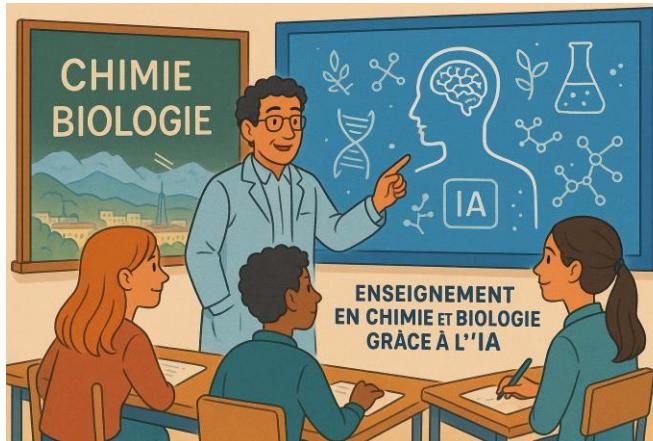
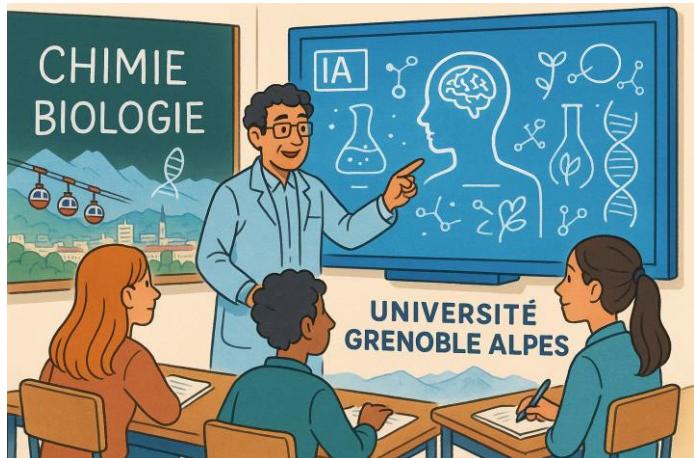
## @Valencia

- Iñaki Tuñon
- Adrián García
- Meritxell Malagarriga Perez



# Remerciements

Merci de votre attention



# Précisions

Films : réaction

Diagramme d'énergie : E vs CR

Coordonnées de Réaction vs Trajectoire

