

L'intelligence artificielle comme moteur dans la recherche en chimie

Professeur Carlo Adamo, Directeur de l'Institut de chimie des sciences de la vie et de la santé de l'École nationale supérieure de Chimie ParisTech et du CNRS, membre de l'Institut Universitaire de France, Paris.

1 Introduction : la chimie, un espace encore à explorer

Il existe encore un espace chimique énorme que l'on voudrait explorer de façon efficace. Actuellement, environ 350 000 produits chimiques sont commercialisés, mais on connaît au total 20 000 000 molécules organiques. Si on prend une seule de ces molécules, comme l'hexane, et que l'on considère les 150 substituants possibles

à la place des hydrogènes, on arrive à 1 030 molécules que l'on pourrait créer.

L'exploration de cet espace chimique peut être faite dans deux sens (*Figure 1*) :

– L'apprentissage déductif¹ : la connaissance des lois par modélisation permet de générer l'information qui peut générer des datas².

1. Déductif : qui procède par déduction.

2. Datas : données.

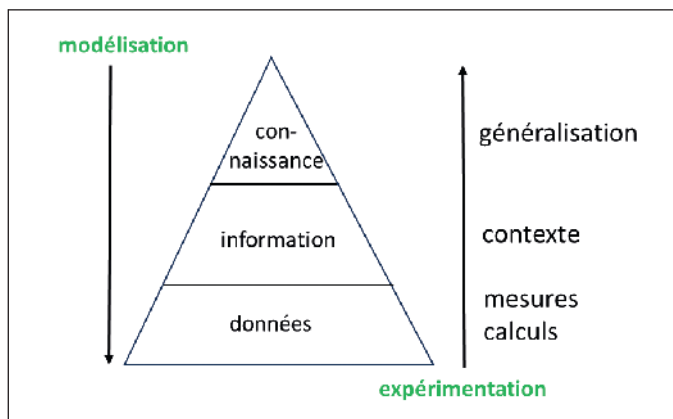


Figure 1

Schéma des types d'apprentissage.

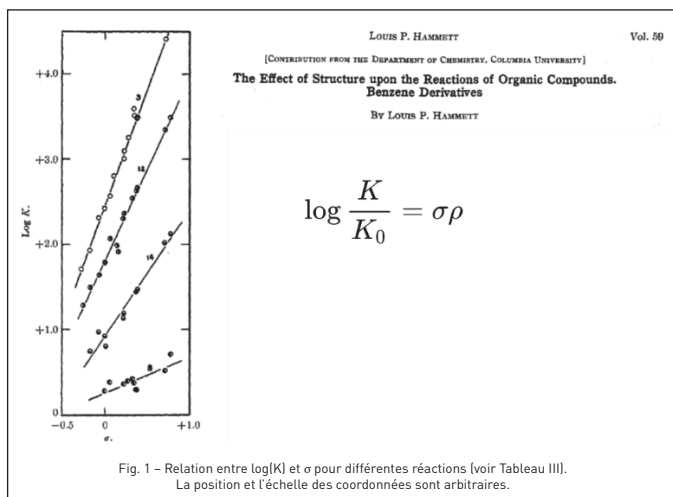


Figure 2

Extrait d'un article de Louis P. Hammett, introduisant la loi de Hammett (de *J. Am. Chem. Soc.* 59, 1937, 96), concernant l'effet de la structure sur les réactions de composés organiques dérivés du benzène.

K = constante de réaction pour les benzènes substitués

K_0 = constante de réaction de référence

σ = constante du substituant

ρ = constante de réaction.

La constante σ dépend uniquement du substituant et la constante ρ dépend uniquement du type de réaction. Une centaine de valeurs de σ tabulées pour les substituants para et méta, mais pas pour ortho (à cause des effets stériques).

– Ou vice-versa l'apprentissage inductif³ : à partir des données expérimentales, on génère des informations pour arriver à des lois générales.

Voyons quelques exemples de questions que peuvent se poser les chimistes.

1. Quelles sont les structures que je peux générer pour améliorer les propriétés d'une molécule déjà connue ? Puis-je synthétiser et produire ces molécules ? L'intuition chimique, donc l'expérience qu'on a comme expérimentateur ou comme théoricien, est un filtre important qui nous aide beaucoup dans notre recherche, mais la recherche des relations entre structure et propriétés est fondamentale en chimie, notamment en chimie physique pour répondre à ce type de question.

La première relation entre structure et propriété date de 1868, réalisée par Crum-Brown et Fraser qui ont eu l'idée de relier la structure d'une série de dérivées de la strychnine⁴ à leur activité physiologique⁵. Naturellement, ils imaginaient qu'il existait une structure mère, mais ils n'étaient pas capables à l'époque de prouver

3. Inductif : qui procède par induction (opération mentale qui consiste à remonter des faits à la loi, de cas particuliers à une proposition plus générale), opposé à déduction.

4. La strychnine est une molécule chimique, un alcaloïde toxique extrait de la noix vomique. Cette substance est un poison.

5. La physiologie est la science étudiant le fonctionnement d'un organe ou d'un organisme vivant.

cette relation entre structure et propriété.

2. Chercher la raison pour pouvoir prédire. Un exemple important de ce type de démarche pour les chimistes est la loi de Hammett (*Figure 2*) qui date de 1937 et qui relie par une fonction linéaire la constante de vitesse⁶ de réactions de benzènes substitués à la nature des substituants. Cependant si la méthode fonctionne très bien pour des substituants en méta⁷ et en para, elle ne fonctionne pas pour des substituants en ortho, car dans ce dernier cas les mécanismes de réaction sont plus complexes.

C'est un exemple des limitations que l'on peut avoir dans ce type de relation et dans les bases des données qui y sont associées.

3. Un autre exemple qui commence à entrer dans le monde de l'IA est le projet Dendral. C'est une recherche des années 1960, conçue par deux informaticiens, un biologiste et un chimiste assez connus. L'objectif de ce projet était d'analyser avec l'aide de l'ordinateur des spectres de masses⁸. C'est le premier

6. Grandeur caractérisant la vitesse de réaction, dépendante uniquement de la température.

7. Ortho, méta, para : désignation de la position des substituants secondaires par rapport à un substituant principal sur un cycle benzénique.

8. La spectrométrie de masse est une technique d'analyse qui permet la détermination des masses moléculaires des composés analysés ainsi que leur identification et leur quantification.

projet connu d'intelligence artificielle où le travail du chercheur est remplacé par le travail de la machine, avec une rapidité majeure mais tout en gardant l'expérience et la précision que peut avoir l'expérimentateur. Le but était de créer un système expert pour la connaissance d'un cas spécifique : les spectres de masse. À l'époque, on ne disposait pas de beaucoup de mémoire vive, 512 kilobytes, et le coût des systèmes était autour de 30 000 \$ par mois.

Le système de fonctionnement est présenté sur la *Figure 3*.

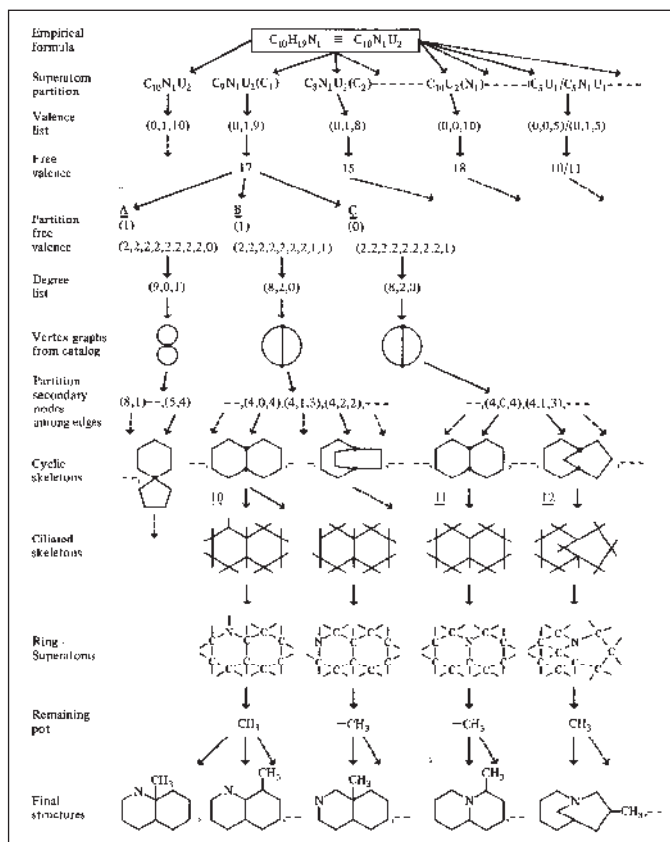


Figure 3

Principes du système Dendral.

Le terme DENDRAL vient de dendron⁹, qui est un mot grec. Ils utilisaient des motifs de fragmentation¹⁰ qui venaient de l'expérience pour l'input et ils faisaient des suppositions sous forme d'arbres de décision (**Figure 3**) pour arriver, à la fin, à interpréter le spectre. L'étape importante était la codification¹¹ de la structure moléculaire, et cela reste toujours un problème : comment expliquer à l'ordinateur la structure de nos systèmes ?

2 L'intelligence artificielle au service de la chimie

Le chapitre de François-Xavier Couderc introduit le concept d'intelligence artificielle. La *machine learning*¹² est la technologie qui concerne les algorithmes et les méthodes qui utilisent une intelligence artificielle. La *deep learning*¹³ est un sous-domaine du *machine learning*. On peut espérer que l'IA, et les méthodes associées

de *machine learning*, soient une opportunité pour trouver de nouvelles relations qui ont jusqu'à maintenant échappé à notre chère intuition chimique.

Des méthodes de *machine learning* sont présentées dans le chapitre de François-Xavier Couderc, elles sont nombreuses et les plus simples sont des régressions linéaires ou multilinéaires¹⁴ ou de l'analyse comprenant principalement des arbres de décision. Beaucoup de chercheurs chimistes ont utilisé au moins une fois l'une de ces méthodes.

On peut utiliser l'IA pour beaucoup de choses en chimie, par exemple pour prédire des propriétés moléculaires, pour faire du design moléculaire¹⁵, du *drug discovery*¹⁶ pour mettre au point des nouveaux médicaments, ou encore pour prédire des produits ou faire de la rétrosynthèse¹⁷. Elle est aussi utilisée en chimie physique pour prévoir des propriétés thermodynamiques¹⁸, des ph, des coefficients de partition¹⁹,

9. Dendron : arbre en grec.

10. Input : représente l'ensemble des données fournies en entrée d'un programme informatique.

11. Codification : encoder l'information afin qu'un ordinateur puisse la déchiffrer.

12. La *machine learning* est un sous-ensemble de l'intelligence artificielle (IA). Cette technologie vise à apprendre aux machines à tirer des enseignements des données et à s'améliorer avec l'expérience, au lieu d'être explicitement programmées pour le faire.

13. La *deep learning* [ou « apprentissage profond »] est un sous-domaine du *machine learning*, lui-même faisant partie de la grande famille de l'intelligence artificielle. Il correspond à toutes les techniques de réseaux de neurones artificiels.

14. Régressions multilinéaires : enchaînement de plusieurs régressions linéaires.

15. Design moléculaire : confection de molécules.

16. *Drug discovery* : conception, découverte de médicaments.

17. Rétrosynthèse : synthèse à partir du produit pour obtenir les réactifs [méthode inverse d'une synthèse directe].

18. La thermodynamique est une branche de la physique qui étudie les propriétés des systèmes où interviennent les notions de température et de chaleur.

19. Coefficients de partition : coefficients de probabilité relatifs à une fonction et décrivant la probabilité d'avoir la molécule dans la configuration définie par ladite fonction.

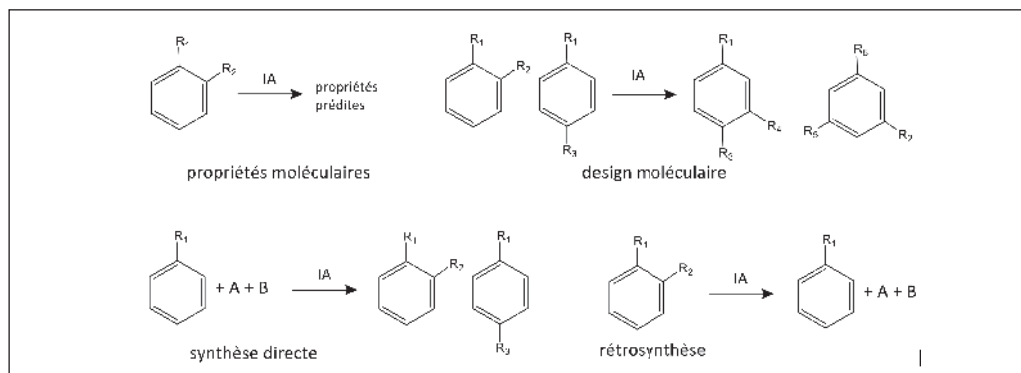


Figure 4

Applications de l'IA en chimie moléculaire.

et en spectroscopies IR²⁰ ou UV²¹.

L'IA permet aussi de prévoir des propriétés liées au risque ou à la toxicité, ou enfin des propriétés de types biologiques bien spécifiques comme des concentrations dans des corps humains.

Ces dernières années, le nombre de publications qui associent le mot chimie aux mots « *machine learning* », « intelligence artificielle » ou « *deep learning* », a rapidement augmenté. Cela concerne tous les domaines de la chimie (beaucoup en chimie analytique, chimie physique, un peu moins en chimie organique) et sur des systèmes qui vont des petites molécules, aux alliages, ou aux polymères, pour citer quelques exemples.

2.1. Exemples d'application de l'IA en chimie moléculaire

L'IA permet de prédire des propriétés, le produit d'une

synthèse, et de faire du design moléculaire ou de la rétrosynthèse (Figure 4).

En pharmacie l'IA est utilisée depuis de nombreuses années pour découvrir de nouveaux principes actifs.

C'est le cas par exemple dans le domaine des antibiotiques où l'on utilise des méthodes de *machine learning* avec des descripteurs²² qui fournissent l'information chimique sur la molécule. La première difficulté est de trouver les descripteurs les plus adaptés à la propriété recherchée. Ensuite le *machine learning* est utilisé pour prévoir les activités.

On peut faire la même chose avec d'autres méthodes, comme celle des réseaux de neurones, en passant par une représentation codifiée de la molécule.

Pour illustrer l'intérêt de ce type d'application, prenons l'exemple (Figure 5) de la découverte d'un nouvel

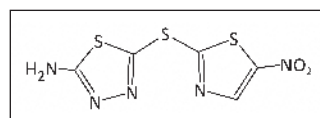


Figure 5

Exemple de machine learning pour la découverte de nouveaux antibiotiques : molécule.

20. IR : infrarouges.
21. UV : ultraviolets.

22. Descripteur : caractéristique spécifique à une molécule.

antibiotique, l'halicine, dans lequel les méthodes de *machine learning* ont été appliquées à une propriété spécifique qui était l'inhibition de la croissance d'*E. coli*²³. Ensuite, les équipes de chercheurs ont fait du *screening*²⁴ sur plus de 107 millions de molécules, ce qui leur a permis de découvrir une nouvelle molécule qu'ils ont appelée halicine, en l'honneur de HAL qui était l'ordinateur intelligent dans *l'Odyssée de l'Espace*. Ce qui était intéressant, c'est qu'il n'y avait, dans leur approche, aucune supposition sur le mode de fonctionnement des médicaments ainsi que sur l'attribution des groupements chimiques. Ils ont finalement trouvé une structure différente de celles des antibiotiques traditionnels. Cette nouvelle molécule *a priori* fonctionne bien pour différentes bactéries et il semble que son mécanisme d'action soit lui aussi complètement différent de celui des autres antibiotiques : c'est une molécule qui va agir sur le flux de protons à travers une membrane cellulaire. L'halicine affiche une activité bactéricide contre une large gamme de pathogènes dont la tuberculose mycobactérienne et les bactéries résistantes aux antibiotiques.

L'intelligence artificielle apparaît dans cet exemple comme instrument de rupture par rapport à l'expérience et à l'intuition chimique et permet d'étendre l'espace chimique.

23. *E. coli* : *Escherichia coli* est une bactérie naturellement présente dans la microflore digestive de l'Homme et des animaux à sang chaud.

24. *Screening* : criblage.

Mais comme cela a été rapporté dans le chapitre de François-Xavier Couderc, le choix du dispositif sur lequel on va appliquer la méthode de *machine learning* est important, notamment la définition et la qualité de la propriété que l'on va rechercher.

Applications de l'IA en synthèse organique

L'IA a aussi été utilisée en synthèse et en rétrosynthèse organique afin de prédire les produits de réaction selon les réactifs ou agents de réaction utilisés. Des méthodes d'apprentissage ont été mises au point par exemple par le MIT²⁵, qui sont toujours des méthodes de *machine learning*. On part des réactifs que l'on doit codifier en termes de graphes moléculaires, et on leur associe une propriété avec des réseaux de neurones. Cela permet ensuite de faire du *screening* avec de bons résultats pour de nombreuses réactions : des réactions organomagnésiennes, des halogénations, etc. La qualité du résultat varie un peu avec la réaction. C'est très précis pour les couplages de Suzuki²⁶, un peu moins sur l'addition de Grignard²⁷, mais ces méthodes

25. MIT : Massachusetts Institute of Technology : Université prestigieuse américaine.

26. Réaction de Suzuki : réaction chimique entre deux groupements aryle (composés aromatiques).

27. Réaction de Grignard : réaction d'addition entre un halogénure organomagnésien et un composé organique porteur d'un groupe carbonyle, typiquement un aldéhyde ou une cétone, pour donner respectivement un alcool secondaire ou un alcool tertiaire.

restent tout de même des outils très puissants (*Figure 6*).

On peut aussi faire des prédictions de régiosélectivité²⁸, dans lesquelles on couple par exemple des méthodes de *machine learning* avec des méthodes de chimie computationnelle²⁹ plus traditionnelles. On peut citer des méthodes qui sont basées sur la théorie de la fonctionnelle de la densité³⁰, donc sur les propriétés de

la structure électronique, ou encore basées sur la régiosélectivité. On obtient alors des prédictions plus précises, autour de 90 %.

2.2. L'intelligence artificielle au service de l'industrie

Régulation Reach : ces méthodes de *machine learning* introduites à différents niveaux ont dépassé le contexte académique mais aussi le contexte industriel pour arriver au niveau de la réglementation européenne. Si on prend par exemple la réglementation REACH³¹, on y trouve la promotion de méthodes alternatives à l'expérimentation animale. L'idée est de mettre sur le

28. Régiosélectivité : une réaction chimique est dite régiosélective si l'un des réactifs ou des intermédiaires réactionnels réagit préférentiellement avec certains sites d'un autre réactif parmi plusieurs possibilités. On obtient ainsi plusieurs isomères d'une molécule.

29. Chimie computationnelle : étude de la chimie à l'aide de l'outil informatique.

30. Théorie de la fonctionnelle de la densité : méthode de calcul quantique permettant l'étude de la structure électronique, en principe de manière exacte.

31. REACH : REACH est un règlement européen entré en vigueur en 2007 pour sécuriser la fabrication et l'utilisation des substances chimiques dans l'industrie européenne.

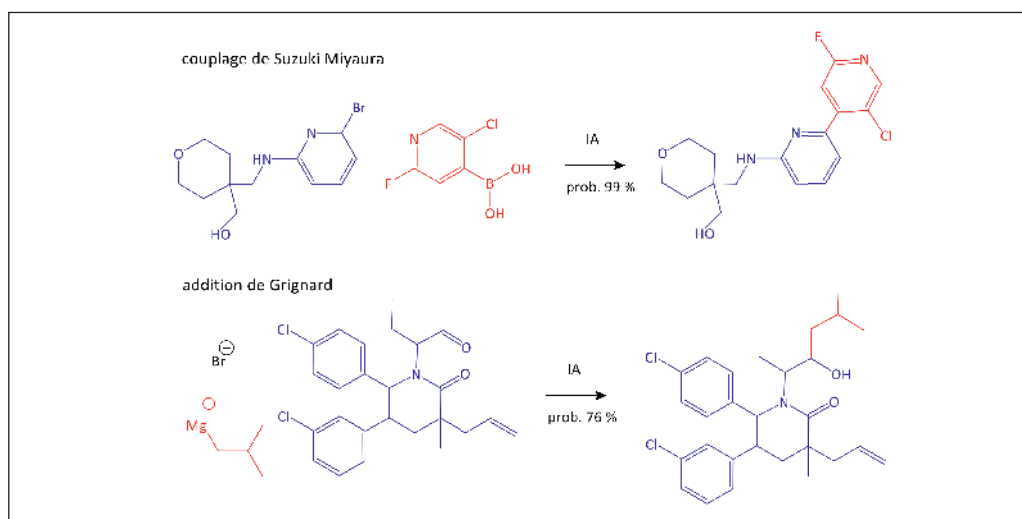


Figure 6

Application du machine learning sur deux types de réaction : le couplage de Suzuki et l'addition de Grignard.

même niveau les résultats expérimentaux et ceux obtenus avec des méthodes théoriques. L'une de ces méthodes est la méthode QSAR³² (Figure 7) de modélisation semi-empirique qui traite les prédictions entre structure et activité. Une seconde méthode est la méthode SAR³³ qui est choisie avec des critères scientifiques incluant une demande d'applicabilité définie, une certaine robustesse ainsi que des objectifs finaux définis. Il est intéressant de remarquer que les critères scientifiques ont

été définis par l'OECD³⁴ il y a des années.

La boîte à outil QSAR a été financée par l'OECD et l'ECHA³⁵ et consiste en une compilation de toutes ces méthodes et fait le lien avec l'approche plus traditionnelle de la chimie computationnelle qui utilise les méthodes développées en chimie théorique.

L'utilisation conjointe de la chimie théorique et de l'IA est efficace dans trois domaines :

- Donner des informations sur les mécanismes qui sont sous-jacents à des relations que l'on trouve à un niveau de l'IA.

32. QSAR : la modélisation semi-empirique QSAR (*Quantitative Structure Activity Relationship*) a comme objectif la prédiction des effets d'une variation de la structure moléculaire sur l'activité biologique.

33. SAR : la méthode est la même que la méthode QSAR sans l'aspect quantitatif.

34. OECD : l'Organisation de coopération et de développement économiques (OCDE) est une organisation internationale qui œuvre pour la mise en place de politiques meilleures pour une vie meilleure.
35. ECHA : Agence européenne des produits chimiques.

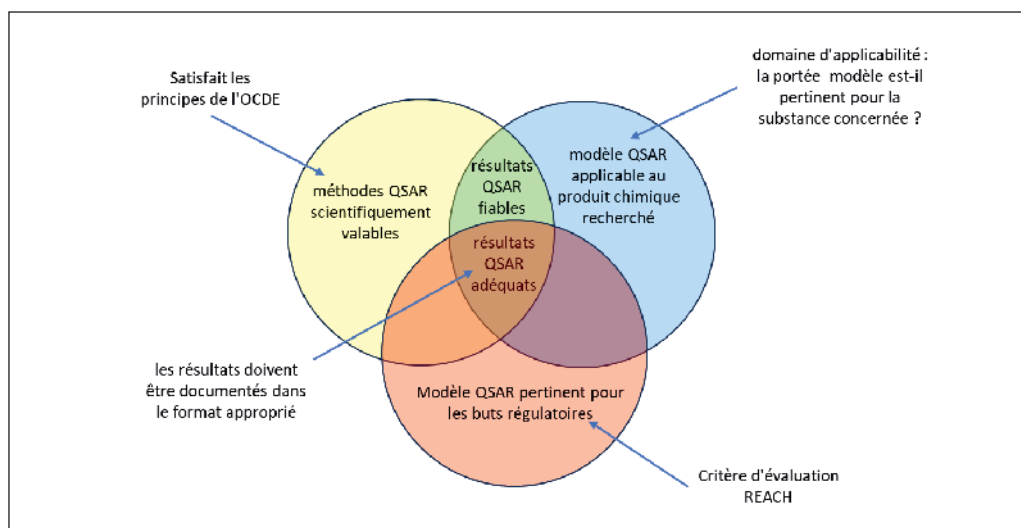


Figure 7

Utilisation des méthodes QSAR.

- Alimenter l'IA à travers des bases de données des descripteurs.
- Utiliser l'IA pour des nouvelles méthodes de chimie théorique.

Prenons quelques exemples.

Information sur les mécanismes réactionnels

La **Figure 8** est l'exemple d'un travail réalisé il y a plusieurs années sur les propriétés explosives, ayant pour objectif la recherche d'un mécanisme de réaction, en utilisant conjointement la chimie computationnelle et l'IA.

Les propriétés explosives sont des propriétés macroscopiques importantes pour REACH. Par exemple, dans le cas des composés nitrés

du benzène, on peut avoir des relations simples linéaires du même type que la loi de Hammett entre les propriétés explosives et la nature des substituants. Ces relations sont valables pour des substitutions méta et para des cycles du benzène, mais en ortho on n'a pas la même précision parce que le mécanisme réactionnel est différent.

Une analyse avec les outils de la chimie computationnelle va permettre de comprendre les relations que l'on trouve avec des méthodes d'intelligence artificielle.

Couplage de l'IA et de la chimie computationnelle classique

Les méthodes de la chimie computationnelle classique (chimie théorique et chimie quantique) peuvent

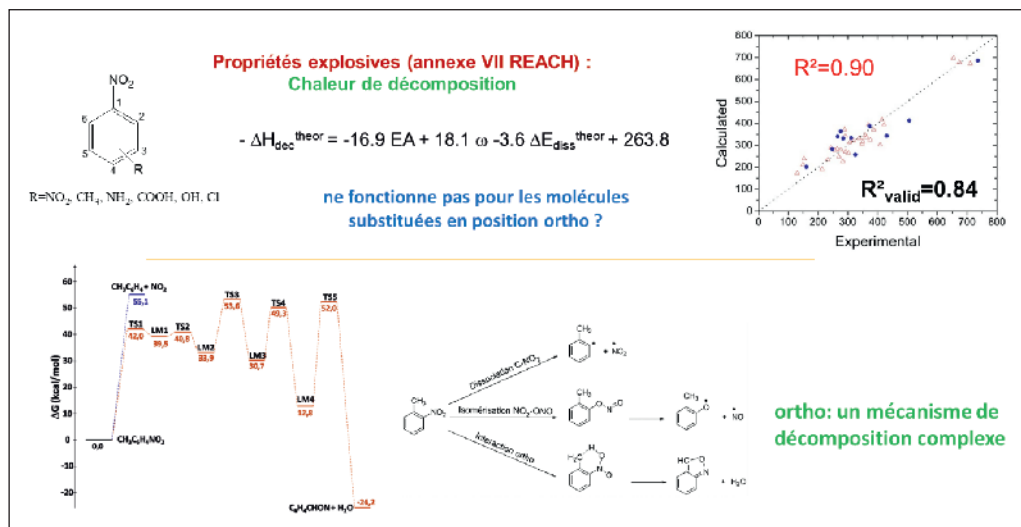


Figure 8

Mécanisme de réaction par l'association de la chimie computationnelle et de l'intelligence artificielle dans le cas des propriétés explosives de composés nitrés du benzène :

- haut : un modèle QSAR capable de prédire la chaleur de décomposition pour les composés substitués en position méta et para, mais qui ne fonctionne pas pour les molécules substituées en position ortho ;
- bas : mécanisme de décomposition de composés substitués en ortho (ortho-nitrotoluène).

être utilisées pour créer des bases de données à partir de descripteurs complètement théoriques, liés à la structure électronique des molécules. Naturellement, la rapidité et l'homogénéité de la base de données ainsi créée seront l'intérêt principal. Mais la limite de cette méthode arrive quand on se demande si le résultat obtenu permet de synthétiser ou non une nouvelle molécule. Il faudrait introduire des critères de synthétisabilité, mais c'est quelque chose qui n'est pas évident à réaliser. La théorie de la fonctionnelle de la densité est la méthode la plus utilisée en chimie théorique dans laquelle on cherche à relier les propriétés des molécules à leurs structures électroniques, comme

l'énergie, qui est une fonction de la densité électronique. Ce sont des méthodes qui ont été introduites par Kohn et Sham³⁶ (**Figure 9**) il y a quelques années. Cependant, dans tous les travaux réalisés, il nous manque un petit morceau pour obtenir l'énergie exacte du système et pour avoir la relation exacte entre l'énergie et les autres propriétés de la densité électronique.

Pour combler cette lacune, on peut imaginer d'utiliser l'intelligence artificielle au niveau de la modélisation. Le nombre d'articles publiés où le terme « théorie de la fonctionnelle de la densité » et « *machine learning* » sont associés, a augmenté rapidement ces dernières années.

L'utilisation des méthodes de l'IA pour définir le petit morceau d'Énergie qui manque au niveau de la théorie de la fonctionnelle de la densité (le terme *Exc* dans l'équation dans la **Figure 9**) a été réalisée par une équipe de DeepMind (une société de Google), l'université de Madrid et le Max Planck Institute. La **Figure 10** est un graphique qu'on aime bien montrer entre théoriciens : en ordonnée figure l'erreur sur l'énergie de réaction pour une série de systèmes (base de données GMTKN55), en kcal/mol. Les acronymes GGA, méta-GGA et hybrides représentent des méthodes « traditionnelles » de la chimie computationnelle (méthodes DFT) et le label DM21 représente les résultats obtenus

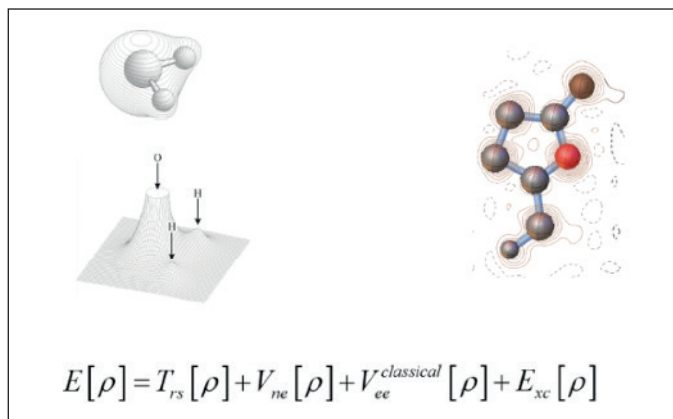


Figure 9

Brève introduction à la théorie de la fonctionnelle de la densité : la théorie de la fonctionnelle de la densité (DFT) est une méthode de chimie théorique utilisée en physique et en chimie pour étudier la structure électronique des atomes, des molécules et des phases condensées. La DFT en tant qu'approche chimique : la connaissance de la densité électronique (ρ) est tout ce qui est nécessaire pour une détermination complète de l'énergie et de toutes les propriétés moléculaires. La DFT moderne est fondée sur l'approche proposée par Kohn, prix Nobel en Chimie en 1998.

Note : fonctionnelle en mathématique est une « fonction d'une fonction ».

36. Kohn et Sham : deux scientifiques ayant effectué des recherches sur la théorie de la fonctionnelle de la densité.

avec la méthode produite avec cette approche de l'IA (méthode DeeperMind21). On voit bien qu'on obtient un joli complément de la modélisation un peu plus traditionnelle.

Cependant, il y a donc encore un peu de travail à faire mais ces modèles sont assez hautement compétitifs dans leur domaine d'applicabilité. Comme exemple, la **Figure 11** présente l'application de cette méthode à des systèmes chimiques très simples, des agrégats de molécules d'eau plus ou moins protonées. En bleu, est représentée l'erreur sur l'énergie de l'interaction des molécules dans des petits agrégats obtenue avec les méthodes DeeperMind21, et, en rouge et vert, celles issues des méthodes plus traditionnelles de chimie computationnelle. Les grandes variations que l'on peut observer pour les données obtenues avec la méthode DM212 (ligne verte) indiquent qu'il y a encore un peu de travail à faire, mais ces modèles basés sur l'IA sont assez hautement compétitifs.

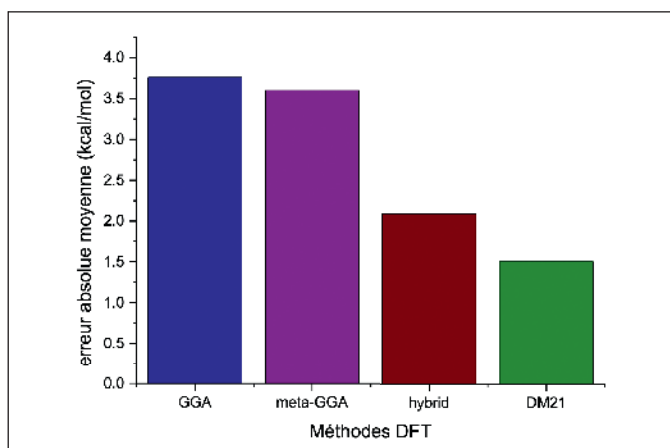


Figure 10

Performances de différentes méthodes DFT : le label DM21 fait référence à la méthode obtenue avec AI, tandis que les autres résultats ont été obtenus avec des méthodes plus traditionnelles. Données extraites de : *Science* 2021, 374, 1385-1389.

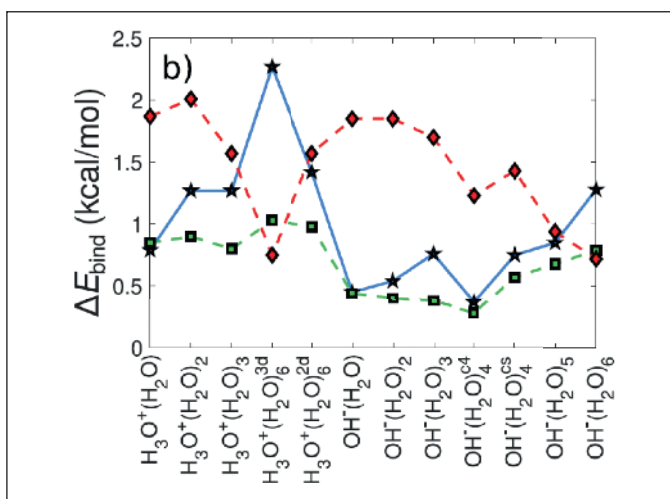


Figure 11

Comparaison de la méthode DeeperMind avec d'autres méthodes pour le calcul de l'erreur sur la valeur de l'Énergie d'interaction entre molécules d'eau plus ou moins protonées. Figure extraite de *J. Chem. Phys.* 156 (2022) 161103.

Conclusion

Des méthodes appropriées d'intelligence artificielle, au sens bien créées et bien pensées, sont donc une belle opportunité et même un défi en chimie. Elles vont nous permettre d'accélérer le processus de recherche (**Figure 1**), d'explorer une nouvelle voie. Elles sont assez complémentaires à d'autres types d'approche *in silico*³⁷ plus traditionnelles, mais, en même temps, il ne faut pas oublier que ces méthodes d'IA demandent des compétences spécifiques.

QUELQUES LECTURES SUPPLÉMENTAIRES

Lederberg, J., Feigenbaum, E.A., Lindsay, R.K. Applications of Artificial Intelligence for Organic Chemistry: The DENDRAL Project, McGraw-Hill, 1980.

Nieto-Draghi, C., Fayet, G., Creton, B., Rozanska, X.; Rotureau, P., De Hemptinne J.-C., Ungerer P., Rousseau B., Adamo C. A General Guidebook for the Theoretical Prediction of Physicochemical Properties of Chemicals for Regulatory Purposes. *Chem. Rev.* 2015, *115*, 13093-13164.

Strieth-Kalthoff, F., Sandfort, F., Segler, M.H.S., Glorius, F. Machine Learning the Ropes: Principles, Applications and Directions in Synthetic Chemistry. *Chem. Soc. Rev.* 2020, *49*, 6154-6168.

Stokes, J.M., Yang, K., Swanson, K., Jin, W., Cubillos-Ruiz, A., Donghia, N.M., MacNair, C.R., French, S., Carfrae, L.A., Bloom-Ackermann, Z., Tran, V.M. Chiappino-Pepe, A., Badran, A.H., Andrews, I.W., Chory, E.J., Church, G.M., Brown, E.D., Jaakkola, T.S., Barzilay, R., Collins, J.J. A Deep Learning Approach to Antibiotic Discovery. *Cell* 2020, *180*, 688-702.e13.

Tkatchenko, A. Machine Learning for Chemical Discovery. *Nat Commun* 2020, *11*, 4125.

Baum, Z.J., Yu, X., Ayala, P.Y., Zhao, Y., Watkins, S.P., Zhou, Q. Artificial Intelligence in Chemistry: Current Trends and Future Directions. *J. Chem. Inf. Model.* 2021, *61*, 3197-3212.

37. *In silico* : se dit d'une méthode d'étude effectuée au moyen d'ordinateurs (dont les puces sont principalement composées de silicium), permettant d'analyser des données et de modéliser des phénomènes.

Guan, Y., Coley, C.W., Wu, H., Ranasinghe, D., Heid, E., Struble, T.J., Pattanaik, L., Green, W.H., Jensen, K.F. Regio-Selectivity Prediction with a Machine-Learned Reaction Representation and on-the-Fly Quantum Mechanical Descriptors. *Chem. Sci.* 2021, 12, 2198-2208.

Kirkpatrick, J., McMorrow, B., Turban, D.H.P., Gaunt, A.L., Spencer, J. S., Matthews, A.G.D.G., Obika, A., Thiry, L., Fortunato, M., Pfau, D., Castellanos, L.R., Petersen, S., Nelson, A. W. R., Kohli, P., Mori-Sánchez, P., Hassabis, D., Cohen, A.J. Pushing the Frontiers of Density Functionals by Solving the Fractional Electron Problem. *Science* 2021, 374, 1385-1389.

Boiko, D.A., MacKnight, R., Kline, B., Gomes, G. Autonomous Chemical Research with Large Language Models. *Nature* 2023, 624, 570-578.