



HAL
open science

Méthodes de modélisation en chimie computationnelle : pluralisme et pragmatisme, software et benchmarking

Frédéric Wieber, Alexandre Hocquet

► **To cite this version:**

Frédéric Wieber, Alexandre Hocquet. Méthodes de modélisation en chimie computationnelle : pluralisme et pragmatisme, software et benchmarking. Philosophie de la chimie, 2020. hal-02490889

HAL Id: hal-02490889

<https://hal.science/hal-02490889>

Submitted on 25 Feb 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Méthodes de modélisation en chimie computationnelle : pluralisme et pragmatisme, *software* et *benchmarking*

Frédéric Wieber et Alexandre Hocquet, AHP-PRéST, UMR 7117, CNRS et Université de Lorraine

Version d'auteur du chapitre de livre publié, pp. 170-176, in B. Bensaude-Vincent & R.-E. Eastes (Eds.), *Philosophie de la Chimie*, Paris: De Boeck Supérieur.

Introduction

La chimie computationnelle est un domaine de la chimie qui développe des outils et méthodes informatiques pour appréhender, de manière théorique, les objets chimiques : molécules, macromolécules, solides, etc. L'adjectif « computationnel », peu utilisé en français, renvoie à des usages anglais : la *computational science* désigne les activités scientifiques utilisant l'ordinateur comme outil alors que la *computing science* ou *computer science* renvoie à la science de l'ordinateur ou informatique. Si la synthèse planifiée et la gestion informatique des données chimiques peuvent s'apparenter à la « chimie computationnelle », nous nous limiterons dans ce texte une acception plus étroite : la chimie computationnelle comme modélisation de la structure et de la réactivité des objets chimiques. Le texte est organisé en deux parties. La première partie utilise quelques considérations historiques pour détailler quatre principales méthodes de la chimie computationnelle. La deuxième partie analyse les caractéristiques épistémologiques de ces méthodes qui font la spécificité de la chimie computationnelle.

Quatre principales méthodes de la chimie computationnelle

La chimie computationnelle prend forme à partir des années 1970. Ainsi, en 1974, l'American Chemical Society crée la division COMP « Computers in chemistry ». L'expression elle-même est utilisée, à partir des années 1980, non seulement dans les articles et communications scientifiques mais encore dans les offres d'emplois académiques et industriels et les rapports institutionnels des centres de calcul. Ceci rend compte de « la reconnaissance d'un nouveau type de chimiste, différent d'un chimiste théoricien, différent d'un chimiste physique, d'un chimiste organique, d'un spectroscopiste ou d'un biophysicien ».¹ La constitution de cette nouvelle branche de la chimie, en lien avec les évolutions technologiques des ordinateurs², s'appuie sur deux traditions distinctes : chimie quantique et « mécanique moléculaire ».

La chimie quantique se construit, à la fin des années 1920, sur un problème de microphysique.³ Comment imaginer, à partir d'équations de la mécanique quantique, des modèles simplifiés de molécules qui seraient à la fois calculables et utiles aux chimistes ? La chimie quantique substitue, à une représentation atomiste de la liaison chimique, une représentation

1 Counts, R. W. (1987). What is computational chemistry? *Journal of Computer-Aided Molecular Design*, 1(1), 95-96.

2 L'histoire de l'informatique montre que les ordinateurs changent de statut à mesure qu'ils passent d'instruments onéreux et rares de la *big science*, à des outils communs et programmables en laboratoire. Si la chimie quantique est d'abord marginale dans ce processus, la chimie computationnelle devient, au début des années 1980, un client important des ressources informatiques aux États-Unis. Bolcer, J. D., & Hermann, R. B. (1994). The Development of Computational Chemistry in the United States. In K. B. Lipkowitz & D. B. Boyd (Eds.), *Reviews in Computational Chemistry* (Vol. 5, pp. 1-63). New York : John Wiley & Sons.

3 Gavroglu, K., & Simões, A. (2012). *Neither Physics nor Chemistry: A History of Quantum Chemistry*. Cambridge : MIT Press.

36 quantique des molécules. L'impossibilité de résoudre analytiquement les équations oriente les
37 chimistes théoriciens vers la recherche d'approximations. La plus populaire d'entre elles est « la
38 méthode *ab initio* ». Cette méthode repose sur des principes théoriques bien établis et n'utilise pas
39 de paramètres *ad hoc*, qui dépendraient de l'objet chimique à modéliser. La méthode *ab initio* porte
40 ainsi une ambition réductionniste et universaliste.⁴ Elle requiert pourtant des approximations
41 numériques pour faciliter les calculs : par exemple, les orbitales d'une molécule sont mieux
42 représentées en théorie par des fonctions de Slater, mais plus faciles à calculer en pratique par des
43 fonctions gaussiennes. Comme le coût calculatoire de la méthode *ab initio* augmente
44 exponentiellement avec la taille des molécules, seules les propriétés des molécules simples, de
45 petite taille, peuvent être calculées sur ordinateur.

46 La deuxième tradition, de « mécanique moléculaire », émerge dans les années 1950 et 1960
47 avec les premiers super-ordinateurs. Elle s'appuie sur des bases théoriques plus simples que la
48 chimie quantique et sur une conception classique de la liaison chimique. Par analogie avec la
49 représentation des molécules qu'offre les modèles matériels tridimensionnels, les molécules sont ici
50 modélisées comme des boules connectées par des ressorts.⁵ La « mécanique moléculaire » s'inscrit
51 dans un paradigme newtonien de mécanique et est associée à l'étude des structures
52 conformationnelles en chimie organique, en s'inspirant des représentations des liaisons chimiques
53 de la spectroscopie de vibration. Ces « méthodes empiriques »⁶ nécessitent une sorte de bricolage de
54 différents paramètres empiriques afin de produire des résultats congruents avec les données
55 expérimentales. Moyennant cette paramétrisation *ad hoc*, la mécanique moléculaire permet de
56 calculer, en pratique, la structure conformationnelle des molécules les plus fréquemment étudiées en
57 chimie organique, biologique et pharmaceutique.⁷

58 Ces deux traditions, quantique et classique, fixent les bornes extrêmes des méthodes de
59 modélisation en chimie computationnelle : *ab initio* du côté des solutions fondamentales, empirique
60 du côté des solutions pragmatiques (cf. tableau des méthodes).

61 Un troisième type de méthodes – les méthodes semi-empiriques (développées à partir des
62 années 1960 et 1970) – constitue entre ces deux pôles une sorte de voie intermédiaire. L'expression
63 « semi-empirique » illustre ce caractère intermédiaire entre formulation quantique de type *ab initio*
64 et souci pragmatique de calculabilité inspiré de la mécanique moléculaire. L'idée est de réduire les
65 temps de calcul les plus longs d'un modèle quantique par la fixation de paramètres empiriques.
66 Dans les méthodes semi-empiriques, l'information chimique qu'on peut extraire des modèles est du
67 même type que dans les méthodes *ab initio*, mais elles partagent avec les méthodes de mécanique
68 moléculaire la difficile tâche de paramétrisation. Cette dernière est basée sur l'organisation
69 géométrique, les quantités thermodynamiques et la nature des éléments chimiques constitutifs. Elle
70 pose le problème des « paramètres manquants » : pour des éléments autres que le carbone,

10 4 Park, B. S. (2003). The Hyperbola of Quantum Chemistry: the Changing Practice and Identity of a Scientific
11 Discipline in the Early Years of Electronic Digital Computers, 1945-65. *Annals of Science*, 60(3), 219-247.

12 5 Francoeur, E. (2001). Molecular Models and the Articulation of Structural Constraints in Chemistry. In U. Klein (Ed.),
13 *Tools and Modes of Representation in the Laboratory Sciences* (pp. 95-115). Dordrecht : Springer.

14 6 Cette expression est employée par les acteurs eux-mêmes et ne doit pas être vue comme constituant une
15 caractérisation épistémologique de notre part.

16 7 Wieber, F. (2012). Multiple Means of Determination and Multiple Constraints of Construction: Robustness and
17 Strategies for Modeling Macromolecular Objects. In L. Soler, E. Trizio, T. Nickles, & W. Wimsatt (Eds.),
18 *Characterizing the Robustness of Science* (pp. 267-288). Dordrecht : Springer.

70 l'hydrogène, l'azote et l'oxygène, des paramètres « bouche-trous » sont introduits de manière *ad*
71 *hoc* dans les programmes afin de permettre les calculs⁸.

72 Une quatrième méthode de modélisation, importante à partir des années 1990 et 2000, est
73 basée sur la théorie de la fonctionnelle de la densité (DFT). Comme les méthodes semi-empiriques,
74 elle hybride le formalisme quantique et la paramétrisation. Il s'agit toutefois, selon ses promoteurs,
75 d'une méthode qui reste de type *ab initio*. Elle est basée sur la résolution de l'équation de
76 Schrödinger en utilisant non pas les fonctions d'onde spécifiques à chaque électron, comme dans les
77 méthodes *ab initio* strictes, mais la densité électronique globale de l'objet chimique. La densité
78 électronique nécessite la définition d'une fonction d'échange et de corrélation dont l'expression
79 mathématique doit être paramétrée, contrairement aux méthodes *ab initio*. Les paramètres sont
80 toutefois indépendants de l'objet chimique considéré, contrairement aux paramètres des méthodes
81 empiriques et semi-empiriques. Des fonctionnelles de densité électronique de sophistication
82 croissante ont été testées. Une pluralité de méthodes basées sur des fonctionnelles différentes sont
83 aujourd'hui utilisées. Comme pour les approches semi-empiriques et de mécanique moléculaire, et
84 contrairement aux approches *ab initio*, ces différentes méthodes existent en parallèle. Elles
85 procèdent de manière pragmatique et itérative, par essais et erreurs, dans la paramétrisation de
86 mélanges de fonctions d'échange et de corrélation. Enfin, la DFT « conceptuelle » s'attache, elle, à
87 utiliser la densité électronique comme source de définition de grandeurs « utiles aux chimistes »,
88 comme par exemple la nucléophilie et l'électrophilie pour remplacer le concept d'orbitales
89 frontières jugé à la fois simpliste et très lié à la méthode *ab initio*.

90

91 **Caractéristiques épistémologiques de la chimie computationnelle**

92

93 Les quatre méthodes évoquées précédemment – *ab initio*, mécanique moléculaire, semi-empirique
94 et DFT – recouvrent l'essentiel des approches théoriques utilisées en chimie computationnelle. Leur
95 analyse comparative permet d'identifier quatre caractéristiques épistémologiques transversales du
96 domaine : pluralisme méthodologique, centralité du *software*, importance du *benchmarking*,
97 pragmatisme cognitif.

98 Le pluralisme est la première caractéristique constitutive de la chimie computationnelle. Il
99 s'explique, au niveau historique, par une construction disciplinaire à partir de traditions différentes
100 et, au niveau épistémique, par l'absence de méthodes dominantes dans un contexte calculatoirement
101 contraint. En effet, chaque méthode est plus ou moins adaptée à un objet chimique donné –
102 molécule, polymère, solide – sans qu'aucune n'épuise leur totalité (cf. tableau des méthodes, en
103 dernière page). Divers objectifs, différentes formulations théoriques et de nombreux types
104 d'approximations et de méthodes peuvent être convoqués pour résoudre tel ou tel type de problème.
105 « Le but n'est pas d'arriver à un monisme méthodologique mais plutôt de déterminer l'applicabilité
106 de différentes méthodes à des contextes spécifiques sachant les contraintes pragmatiques existantes
107 relativement à la précision, la vitesse et le coût en termes de calcul »⁹. Il existe même des méthodes
108 de modélisation dites « hybrides », comme le QM/MM (pour « quantum mechanics/molecular

19 8 Voir Wieber, F., & Hocquet, A. arXiv preprint, Computational chemistry as « Voodoo Quantum Mechanics » : Models,
20 Parameterization and Software. <https://arxiv.org/abs/1812.00995>

21 9 Fisher, G. (2016). Diagnostics in computational organic chemistry. *Foundations of Chemistry*, 18(3), 241-262, p. 259,
22 notre traduction.

109 mechanics ») qui associe deux formalismes théoriques (classique et quantique) dans une même
110 méthode.

111 Ce pluralisme s'incarne dans ce qui peut être considéré comme un pivot central de la chimie
112 computationnelle : le « software »¹⁰. Certains logiciels constituent des outils fédérateurs qui
113 intègrent plusieurs méthodes différentes dans un même programme (méthodes liées par le code) ou
114 dans un même « package » (méthodes liées par des politiques de distribution et de licence). Ils
115 couvrent ainsi le large éventail méthodologique de la chimie computationnelle, des calculs *ab initio*
116 aux méthodes empiriques, en passant par les nombreux modèles intermédiaires de type semi-
117 empirique ou de type mixte. D'un autre point de vue, une concurrence entre packages développés
118 par différents groupes de recherche matérialise une compétition entre différentes versions d'un
119 même type de méthode, notamment liée aux choix de paramétrisation effectués. Les chimistes
120 computationnels évoluent ainsi dans une sphère professionnelle où des enjeux liés à la distribution,
121 la commercialisation éventuelle, la maintenance des logiciels sont centraux. Le software est aussi
122 un outil de connaissance qui nécessite donc une analyse épistémologique. Par exemple, le problème
123 des paramètres « bouche trou » est lié à des enjeux d'opérationnalité des méthodes mais aussi aux
124 techniques de programmation pour les mettre en œuvre, et aux politiques de licence plus ou moins
125 ouvertes pour leur diffusion. Ces enjeux influent sur le statut épistémologique des résultats produits.

126 La position des logiciels entre logiques commerciales et cognitives, dans un contexte de
127 pluralisme méthodologique, fait apparaître une troisième caractéristique de la chimie
128 computationnelle : le *benchmarking* ou l'évaluation des performances des méthodes et des logiciels.
129 Les critères d'évaluation sont multiples : exactitude de la solution théorique ; représentation de
130 l'entité chimique ; valeurs prédictives des résultats ; rapidité de calcul ; domaines d'application
131 (universel *versus* local) ; etc. Leur importance relative varie d'une méthode à l'autre. Ainsi, on ne
132 demande pas aux modèles de mécanique moléculaire de fournir une représentation fidèle des objets
133 chimiques mais d'avoir une efficacité prédictive. La fidélité de représentation des molécules par
134 rapport à la théorie quantique n'est donc pas une priorité pour les acteurs qui veulent construire des
135 modèles conformationnels pour des objets d'intérêt complexes. Inversement, une méthode *ab initio*
136 donne une représentation qui se veut fidèle des entités tout en restant confinée à des systèmes
137 chimiques très simples. Schématiquement, on peut considérer que chaque méthode a un domaine
138 préférentiel de prédiction : *ab initio* pour les représentations ; DFT pour les prédictions
139 électroniques ; semi-empiriques pour les prédictions thermodynamiques ; empiriques pour les
140 prédictions conformationnelles.

141 La quatrième caractéristique épistémologique de la chimie computationnelle est un
142 pragmatisme assumé où se côtoient théories fondamentales et modèles bricolés. Elle pose la
143 question du rapport aux données empiriques et, donc, à la science expérimentale. Pour traiter un
144 système particulier, les chimistes computationnels doivent trouver un équilibre entre la complexité
145 acceptable des équations (pour qu'elles soient résolubles en pratique) et le nombre de paramètres
146 introduits pour les simplifier : un modèle complexe au niveau du formalisme ne pourra être
147 effectivement appliqué que pour des molécules simples, alors qu'un modèle complexe en termes de

23 10 Nous employons à dessein le terme anglais « software » plutôt que le terme français « logiciel » car ce concept nous
24 semble avoir une plus grande extension que l'objet logiciel. « Software » recouvre ainsi l'ensemble de ce qui est
25 « design » (algorithmes, « flow-charts »), programmation (codes), compilation (« binaries »), distribution
26 (« packages »), et dimension juridique (« licensing ») du logiciel.

148 paramètres ne pourra décrire que des molécules d'une famille très restreinte. La paramétrisation
149 constitue un savoir-faire subtil. Par exemple, elle commence, dans le cas de la mécanique
150 moléculaire, par la définition de « types d'atomes » : un carbone tétraédrique ou trigonal induit des
151 différences de paramètres linéaires, angulaires et de torsion. Plus finement, le carbone trigonal d'un
152 alcène et celui d'une cétone ne correspondent pas aux mêmes paramètres. Les valeurs numériques
153 de paramètres géométriques peuvent être suggérés par des résultats expérimentaux (diffraction aux
154 rayons X, spectroscopies, etc.), par des calculs théoriques effectués sur des molécules plus simples
155 de référence. Puisqu'un paramètre n'est en toute rigueur pas transposable d'un modèle de molécules
156 à un autre, il est nécessaire, pour un groupe de recherche, de se spécialiser sur une famille de
157 molécules sachant les savoir-faire spécifiques nécessaires pour construire ces modèles « bricolés ».
158 Cette non-transposabilité liée à la paramétrisation, présente également dans les méthodes semi-
159 empiriques, est l'une des sources de la compétition entre packages. Ce pragmatisme assumé
160 s'exprime également par des choix de paramètres *ad hoc*, pour « boucher les trous » dans le code.
161 En pratique, l'absence d'un paramètre conduit à la situation la plus redoutée, l'interruption du
162 programme.

163

164 **Conclusion : réflexivité et déterminations**

165

166 La question de la paramétrisation est si vive en chimie computationnelle qu'elle peut conduire à
167 interroger la transparence épistémique d'une méthode. L'opacité se joue au niveau de la
168 modélisation autant que de la programmation. Le code est-il accessible ou pas ? Si oui, comment
169 savoir quels paramètres sont incorporés si le programme ne peut être lu ? Ces questions de
170 transparence épistémique font écho à une interrogation sur la cohérence épistémique : comment
171 savoir si le logiciel utilisé représente bien la méthode voulue si n'importe qui peut la modifier ? Une
172 approche réflexive permet de souligner les limites, les contingences et les zones d'incertitude d'un
173 *software* ou d'une méthode. La chimie computationnelle montre ainsi l'entremêlement des activités
174 de modélisation et de programmation entre elles d'une part et avec les politiques de distribution, de
175 maintenance, et de licence du logiciel d'autre part.¹¹

176 Ces questionnements montrent l'importance de la forme concrète des modèles (matérielle, textuelle
177 ou, dans notre cas, computationnelle) pour évaluer leur statut épistémique. On constate ici que la
178 fiabilité et la robustesse des modèles computationnels se construisent selon une chaîne
179 d'interactions, dont les maillons se renforcent mutuellement : détermination des paramètres et des
180 algorithmes de calcul ; optimisation itérative de la méthode ; circulation des programmes
181 informatiques ; réutilisation de la méthode et applications à d'autres cas. Ceci tend à montrer un
182 processus d'adaptation des modèles au fonctionnement des ordinateurs.¹² Ce
183 « computationnalisme » est organisé, de manière profonde, par les enjeux épistémiques évoqués
184 précédemment mais aussi par les enjeux politiques et économiques de marchés complexes.

185

186

27 11 Voir en particulier Hocquet, A., & Wieber, F. (2017). « Only the Initiates Will Have the Secrets Revealed »:
28 Computational Chemists and the Openness of Scientific Software. *IEEE Annals of the History of Computing*, 39(4).

29 12 Lenhard, J. (2014). Disciplines, models, and computers: The path to computational quantum chemistry. *Studies in*
30 *History and Philosophy of Science Part A*, 48, 89-96.

Tableau présentant schématiquement les grands types de méthodes

Caractéristiques Méthodes	Epoque d'émergence de la méthode	Premiers logiciels populaires utilisant la méthode	Formulation théorique	Objets modélisés	Critères de performance
ab initio	Années 1950 (résurgence avec premiers ordinateurs)	GAUSSIAN 70 (1970)	Quantique (sans paramétrisation)	Petites molécules	Représentation en adéquation avec théorie quantique
DFT	Années 1990	ADF (1995)	Quantique (avec paramétrisation globale)	Grosses molécules	Prédiction électronique
semi-empiriques	Années 1970	MOPAC (1990)	Quantique (avec paramétrisation él. chimique par él. chimique)	Grosses molécules	Prédiction thermodynamique
mécanique moléculaire	Années 1960	MM2 (1977)	Classique (paramétrisation par type d'atomes)	Macromolécules	Prédiction conformationnelle