

Licence 3 & Master 1 Chimie

Mode d'emploi

Gaussian Avogadro

Carine Michel & Stephan Steinmann & Paul Clabaut

`carine.michel@ens-lyon.fr`

`stephan.steinmann@ens-lyon.fr`

`paul.clabaut@ens-lyon.fr`

1 Généralités

1.1 Liens utiles et biblio

Nationale Institute of Standards and Technology (Site de données expérimentales en chimie)

<https://webbook.nist.gov/chemistry/>

Worldwide protein data bank (Site de structures cristallographiques validées de protéines)

<https://www wwpdb.org/download/downloads>

Cambdriqe Crystallograph Data Centre (Site de structures cristallographiques validées)

<https://www.ccdc.cam.ac.uk/structures/>

Introduction to Computational Chemistry, 3rd edition *Frank Jensen*

Essentials of Computational Chemistry, 2nd Edition *Christopher J. Cramer*

A Chemist's Guide to DFT, 2nd Edition *Wolfram Koch & Max C. Holthausen*

1.2 Quelques commandes bash utiles

[**mkdir** #nom#] (MaKeDIRectory) Crée un dossier nommé #nom#

[**ls -lrt**] LiSte les fichiers dans le dossier où l'on se trouve, les options lrt permettent d'obtenir plus de détails et de les classer par ordre de modification décroissant.

[**cp** #file_source# #file_dest#] CoPie #file_source# vers #file_dest# (-r) pour un dossier

[**mv** #file_source# #file_dest#] (MoVe) Déplace #file_source# vers #file_dest#

[**rm** #file#] (ReMove) Supprime un fichier (-r) pour un dossier

[**cd** #chemin#] (Change Directory) Se rend au #chemin# spécifié (relatif ou absolu)

[**cd -**] Retourne à la destination précédente

[**find -name** #pattern#] Cherche un fichier correspondant au #pattern# spécifié

[**grep** #pattern# #fichier#] Imprime les lignes d'un #fichier# contenant le #patern# spécifié

[**&**] Placé en fin de ligne de commande, effectue la tâche de la ligne en arrière plan et laisse donc le terminal disponible pour d'autres instructions

[**top**] Donne un grand tableau contenant toute les tâches en cours de l'ordinateur, permet de vérifier qu'un calcul lancé avec & est terminé

Autocomplétion : Utiliser l'autocomplétion (touche "tab") pour les chemins et noms de fichiers : c'est un gain de temps et un moyen de vérifier que le fichier auquel vous pensez est bel et bien là.

1.3 Logiciels utilisés et quelques autres

Avogadro (construction et visualisation) : <https://avogadro.cc/docs/>

Gaussian (calcul quantique) : <http://www.gaussian.com/keywords/>

Gedit (Editeur de texte)

Gnumeric (Tableur)

ORCA (calcul quantique) <https://orcaforum.kofo.mpg.de/app.php/portal>

MOPAC (calcul semi-empirique) <http://openmopac.net/>

CP2K (calcul quantique et classique) <https://www.cp2k.org/>

AMBERTools (calcul classique) <http://ambermd.org/AmberTools.php>

CHEMSKETCH (construction et visualisation) <https://www.acdlabs.com/resources/freeware/chemsketch/>

VMD (Visualisation) <http://www.ks.uiuc.edu/Research/vmd/current/docs.html>

Jmol (Visualisation) <https://sourceforge.net/projects/jmol/>

2 Les étapes d'un calcul

2.1 Construire une molécule

Resumé : Construire une molécule, c'est créer sa structure, c'est à dire générer un ensemble de positions atomiques dans un référentiel donné avec ou sans informations sur la connectivité (les liaisons entre atomes).

Outils : Avogadro (alternatifs : CHEMSKETCH, outils de constructions liés à des logiciels de calcul...)

Mode d'emploi :

Pour construire une structure, vous pouvez partir de fragments pré-existants (Avogadro : *Buil/Insert Fragment...*) ou la construire atome par atome (Avogadro : mode dessin : taper le nom de l'atome souhaité puis cliquer). On peut ajouter des liaisons entre les atomes (Avogadro : mode dessin : tirer depuis un atome vers l'autre ou bien en cliquant sur la liaison pour passer d'une liaison simple à double). La position d'un atome ou groupe d'atomes dans l'espace peut être modifier en mode selection puis mode manipulation. Une fois la molécule construite "à la main", il est fort probable que les distances, angle etc soient très loin de la position optimale. Dans Avogadro, il est possible d'effectuer une première optimisation en utilisant un champs de force, ce qui nécessite de bien définir la connectivité (voir Figure 1). On note que même au cours de cette optimisation, on peut toujours bouger "à la main" les atomes, par exemple pour forcer l'optimisation locale d'une certaine conformation. Une fois la génération de la structure voulue terminée, on peut l'exporter sous différents formats (.xyz par exemple) qui pourront être lus par de nombreux logiciels de visualisation.

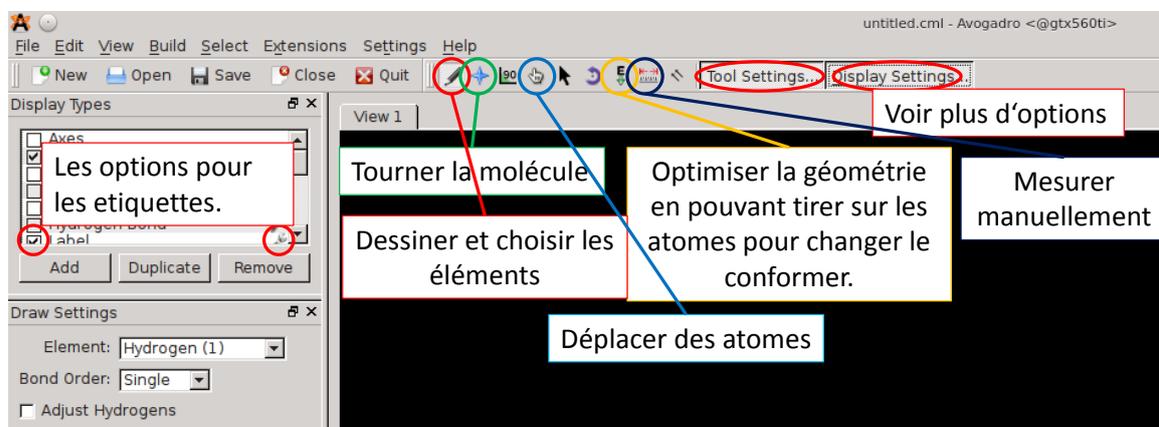


FIGURE 1 – Les commandes les plus utilisées d'Avogadro.

Pour des grands systèmes, on cherchera à favoriser l'importation de structures connues (Avogadro : *File/Import/Molecule File...*) répertoriées dans des bases de données cristallographiques (voir 1.1). Les formats les plus usités sont le CIF (Crystallographic Information Files) et le PDB (Protein Data Bank) pour les structures expérimentales mais de nombreux autres formats existent (généralement liés à des logiciels de calcul ou de construction).

2.2 Préparer le(s) fichier(s) d'entrée

Resumé : Un logiciel de calcul de chimie comme Gaussian prend en général un ou plusieurs fichiers d'entrée. Ils doivent contenir une géométrie mais également des paramètres pour le calcul (niveau de calcul, précision, nombre de processeurs utilisés, etc.).

Outils : Avogadro pour Gaussian09 (alternatifs : n'importe quel éditeur de texte)

Mode d'emploi : La règle d'or est d'utiliser le manuel du logiciel choisi. Pour le cas de Avogadro et Gaussian : une fois la structure construite, nous générerons le fichier d'entrée de Gaussian dans Avogadro en utilisant *Extensions/Gaussian*. Nous choisirons ensuite le type de calcul effectué (Optimisation de géométrie, fréquences ou point d'énergie simple), le niveau de théorie, etc. Au besoin, nous modifierons le texte affiché. Le bouton "generate" permet alors de créer le fichier unique d'entrée de Gaussian. Ce fichier (extension en ".com") peut ensuite être visionné ou modifié par n'importe quel éditeur de texte.

Le fichier d'entrée de Gaussian se présente comme ceci :

- Une ligne optionnelle commençant par %chk= permettant de faire écrire le fichier checkpoint (affichage OM, re-démarrage d'un calcul).
- Une ligne commençant par # et contenant des mot-clés définissant les options choisies.
- Trois lignes vides, la seconde peut être un titre donné par l'utilisateur mais ne sera pas lue par le logiciel. Il vous aide juste à vous retrouver.
- Deux chiffres séparés par un espace : la charge et la multiplicité de spin du système.
- Les coordonnées de l'intégralité des atomes du système données en coordonnées cartésiennes.
- Au moins une ligne vide à **ne surtout pas supprimer !**

La partie supérieure du fichier est l'en-tête et ne dépend pas de la géométrie du système. Tous les mot-clés qu'il est possible d'utiliser dans cette partie sont listés dans le manuel en ligne de Gaussian. La partie "coordonnées" du système peut, elle, être générée par n'importe quel logiciel de construction de structure.

2.3 Lancer un calcul

Resumé : Lancer le calcul correspond à appeler un logiciel, qui, à partir du fichier d'entrée fourni, va effectuer un calcul (quantique ou classique) selon les options choisies.

Outils : Gaussian09 (alternatifs : beaucoup ! CP2K, AMBER,...)

Mode d'emploi : Un script g09 vous sera fourni pour lancer le calcul que vous sauvez dans /home/monlogin/. Pour lancer le calcul, il suffit alors d'aller dans le dossier où se situe votre fichier d'entrée et de taper /home/monlogin/g09, suivi du nom de votre fichier d'entrée. Gaussian va alors effectuer le calcul et générer un fichier de sortie du même nom que le fichier d'entrée mais en extension ".log". D'autres fichiers seront éventuellement créés (.chk, .fchk). La création du fichier est rapide. Il sera rempli au fur et à mesure du calcul. La durée du calcul dépend du niveau de calcul.

Script en 2019 :

```
#!/bin/bash\
umask 022\
input='basename $1 .com'
if [ ! -e $input.com ] ; then
echo "$input.com n'existe pas"
exit
```

```
fi

# Creation du dossier pour tous les jobs
mkdir -p /local/$USER/Gaussian/tmp

# Creation de l'environnement unix pour Gaussian
export g09root=/opt
export GAUSS_SCRDIR=/local/$USER/Gaussian/tmp
. $g09root/g09/bsd/g09.profile

# Commande pour lancer le calcul g09
$input.com
for i in `ls *.chk 2>/dev/null` ; do
if [ ! -e `basename $i .chk`.fchk ] ; then
/opt/g09/formchk $i 2>/dev/null
fi
done
```

2.4 Analyser un calcul

Resumé : Retrouver toutes les informations pertinentes d'un calcul : s'il s'est bien déroulé, l'énergie, la structure finale ou toute autre valeur demandée.

Outils : Ici, **GEDIT** pour lire le fichier de sortie, mais aussi des outils bash, des feuilles de calcul, et **Avogadro** pour visualiser les structures (alternatifs : tout éditeur de texte et tout logiciel de visualisation).

Mode d'emploi : Extraire des informations d'un calcul exécuté par l'ordinateur est la partie la plus importante. La première chose à faire est de vérifier que le calcul s'est bien déroulé comme prévu, ce que Gaussian indique par la phrase finale 'Normal termination of Gaussian...' dans le fichier .log.

La commande bash `grep` n'imprime que les lignes du fichier contenant une certaine chaîne de caractère (ex : `grep "SCF Done" masortie.log`) et permet d'extraire rapidement des données du fichier de sortie.

En ouvrant le fichier .log avec Avogadro, certaines propriétés de la molécule sont affichées en cliquant sur *View/Properties/Molecule Properties*. La dernière géométrie est visible et vous pouvez aisément mesurer des distances, angles etc. Vous pouvez également visualiser la trajectoire suivie par Gaussian pendant l'optimisation de la géométrie. Pour cela, vous devez ouvrir le fichier .log dans Avogadro et puis utiliser *Extensions/Animations*. Vous aurez peut-être besoin de fermer et ré-ouvrir ce menu pour que l'animation soit visualisée correctement.

3 Niveau de calcul

Nous distinguerons, par ordre croissant de précision et de coût de calcul

- la mécanique moléculaire
- l'approche semi-empirique
- la méthode HF (Hartree-Fock)
- la DFT (density functional theory)
- les méthodes post-HF

Dans le cadre des trois dernières, la base utilisée pour optimiser les orbitales doit être précisée. Les fonctions de base sont des gaussiennes dans Gaussian.

4 Les différents types de calcul

4.1 Point d'énergie simple

L'énergie électronique d'une structure est obtenue en faisant un point d'énergie simple (mot-clé SP). L'équation de Schrödinger est résolue par une procédure de cycles itératifs auto-consistants (procédure SCF) qui s'arrête une fois que la variation d'énergie entre un pas et le suivant est plus petite qu'une valeur seuil. "SCF Done" indique la fin de cette procédure. L'énergie électronique est alors donnée en Hartree.

4.2 Optimisation de géométrie

La géométrie d'un composé est optimisée pour minimiser l'énergie potentielle quand le mot-clé opt est utilisé. Les algorithmes d'optimisation se basent sur les géométries, les énergies et les forces précédentes pour extrapoler la géométrie suivante. Cette procédure itérative s'arrête quand les critères d'optimisation de géométrie sont atteints. Ils sont au nombre de quatre dans Gaussian. A chaque pas, Gaussian récapitule si ces critères sont atteints (YES) ou pas (NO). Une fois l'optimisation finie, Gaussian indique "Optimization completed. – Stationary point found".

4.3 Spectres infrarouges

Le mot-clé freq permet de calculer les modes normaux et les fréquences associées. Les intensités infrarouges sont également rapportées. Avogadro permet de visualiser les modes normaux de vibration, la liste des fréquences et le spectre IR (*Extensions/Vibrations...*). L'option Raman permet de calculer les intensités Raman en plus de celles IR.

4.4 Thermochimie

Le mot-clé freq permet également d'obtenir les fonctions de partition vibrationnelles et l'énergie de point-zéro. Pour compléter, Gaussian calcule également les moments d'inertie dans l'approximation du rotateur rigide pour calculer la fonction de partition rotationnelle et la masse pour la fonction de partition translationnelle dans l'approximation du gaz parfait. Les données thermochimiques qui en dérivent sont rapportées dans la partie Thermochemistry. Le mots-clé SCRF (PCM, solvent=toluene) affecte l'enthalpie libre comme si le système était immergé dans le solvant spécifié. La température du milieu peut être changée Temperature=300 ainsi que la Pression (en atm, Pressure=1.5).

4.5 Orbitales moléculaires et charges

Dans Gaussian, on utilise la ligne %chk=file (file doit être exactement le nom de votre fichier d'entrée mais sans l'extension .com), associé à SP pour écrire les informations des OMs obtenues dans un fichier file.chk. Ce fichier peut ensuite être converti en file.fchk en utilisant l'utilitaire formchk dans le terminal. Cela est fait automatiquement via le script g09 fourni.

Les charges de Mulliken sont écrites par défaut dans le fichier `log`. D'autres analyses de population peuvent être fournies, par exemple en utilisant le mot-clé `pop=nbo`.

4.6 Spectres UV-vis

Les premières excitations du système peuvent être calculées de manière approximée grâce à la méthode ZINDO (Zerner's Intermediate Neglect of Differential Overlap). C'est une méthode semi-empirique mais très rapide pour obtenir des interactions avec des états non fondamentaux et donc excités. Cette méthode permet donc de déduire les fréquences absorbées par le composé.

Dans Gaussian, elle est invoquée par le mot-clé `ZINDO(NState=10)`, par exemple pour les 10 premiers états excités. Les excitations sont à chercher dans le fichier `.log`, après "Excited State", en eV et nm.

4.7 Chemin réactionnel et état de transition

Pour obtenir le profil réactionnel d'une étape élémentaire, il faut commencer par optimiser réactifs et produits. Si il y a par exemple deux réactifs, ils seront optimisés en interaction. Ensuite, plusieurs étapes sont nécessaires.

- Définir une coordonnée de réaction (une distance, un angle, un dièdre) : déterminer le "label" des atomes définissant la liaison ou l'angle avec Avogadro, en cochant l'option "Label" dans l'encadré "Display type", à gauche ; puis en modifiant les paramètres de cette option (clé à molette) en "Text : Atom number" pour le "Atom Labels" (voir Figure 2).
- Faire une série d'optimisation contraintes le long de cette coordonnée en partant soit du réactif soit du produit. Il est plus simple d'allonger une liaison que de la former. Pour cela, utilisez l'option `modredundant` du mot-clé `opt` (`opt=modredundant`). La coordonnée est précisée après le bloc de coordonnées et sera séparée de ce bloc par une ligne blanche. Elle s'écrit `N1 N2 (N3 N4) S steps step-size` avec "N1", "N2", etc. sont les labels des atomes. Deux labels définissent la distance N1-N2, trois définissent un angle N1-N2-N3 et quatre un angle dièdre N1-N2-N3-N4. "steps" représente le nombre d'optimisations contraintes à faire et "step-size" la taille du pas, c'est à dire la distance (en angströms) ou l'angle (en degrés), dont doit être modifiée la liaison ou l'angle entre chaque pas.
- Analyser le profil réactionnel obtenu : chaque optimisation contrainte donne l'énergie en fonction d'une valeur donnée de la coordonnée de réaction. Raffiner au besoin le profil dans la zone maximale (proche de l'état de transition)
- Optimiser l'état de transition en partant d'une structure la plus proche possible en utilisant les options `opt=(TS,noeigentest,calcfc)`. Ces options permettent d'optimiser un point-selle plutôt qu'un minimum de la surface d'énergie potentielle.
- Vérifier la présence d'une et une seule fréquence imaginaire une fois la structure optimisée.

On peut aussi utiliser d'autres approches pour trouver l'état de transition qui n'utilisent pas une structure proche de l'état de transition mais plutôt celle des réactifs et produits (`qst2`) et éventuellement celle d'une structure intermédiaire (`qst3`). Pour `qst2`, la géométrie du réactif, puis du produit, pour `qst3`, celle du réactif, celle du produit, puis celle d'une structure intermédiaire pour l'état de transition (qui peut être obtenue préalablement avec `modredundant` ou `qst2`). Pour ces deux inputs, **chaque** géométrie est donnée à avec :

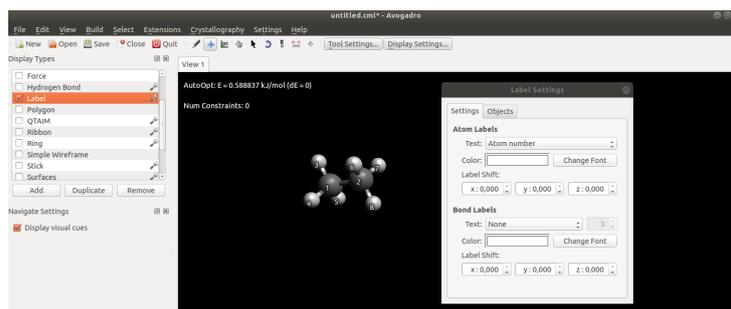


FIGURE 2 – Visualiser les labels

- Une ligne vide
- Son titre
- Une ligne vide
- Ses coordonnées

Et toutes sont données à la suite dans le fichier d'entrée, après la ligne de mot-clés.

5 Commandes ssh

Résumé : En principe, les machines du CBP ne sont accessibles qu'à partir du réseau "interne" de l'ENS c'est à dire, par exemple, depuis les ordinateur de la salle europe ou tout autre poste connecté aux bornes wifi de l'ENS. Mais il y a quand même moyen d'accéder à ce réseau depuis l'extérieur de l'école, via une première commande d'accès à distance (ssh). Il faudra donc faire deux connections ssh de suite :

Votre machine —ssh—> Réseau interne —ssh—> Machine du cbp

Accès standard :

· Première commande ssh : `ssh -X votrellogin@ssh.ens-lyon.fr`
votrellogin est celui du CAS ENS, donc initiale+nom de famille et ssh.ens-lyon.fr est le nom de la passerelle qui sert d'accès au réseau interne (ne pas changer). L'option -X permet un export graphique (donc de voir des images et pas seulement un terminal) à travers le tunnel ssh.

Il vous demande alors un mot de passe qui est celui du CAS ENS.

Si c'est la première fois, on vous demande aussi d'ajouter votre "fingerprint" aux utilisateurs. On répond "yes".

Une fois arrivé (ou si vous étiez déjà sur le réseau interne), on entre la seconde commande ssh :

· Seconde commande ssh : `ssh -X nomdunemachineducbp.cbp.ens-lyon.fr`

Ici, pas besoin du login puisque vous avez déjà donné votre identité. Le nom d'une machine est trouvable sur le site : <http://www.cbp.ens-lyon.fr/doku.php?id=ressources:ressources> ou directement noté sur les machines physiques.

Pareil, on rentre son mot de passe CAS et éventuellement "yes".

Accès rapide :

Une fois habitué, on peut faire les deux en une seule commande :

`ssh -X -oProxyCommand="ssh login@ssh.ens-lyon.fr -W %h:%p" login@machine.cbp.ens-lyon.fr`

En remplaçant "login" et "machine" évidemment. Il demande toujours les deux mots de passe par contre.

Une dernière commande fort utile à utiliser sur votre ordi perso (c'est à dire pas déjà connecté en ssh) pour copier un fichier présent au cbp vers votre ordi :

```
scp -oProxyCommand="ssh login@ssh.ens-lyon.fr -W %h:%p"
```

```
(suite) login@machine.cbp.ens-lyon.fr:#chemin_a# #chemin_b#
```

Où #chemin_a# est le chemin absolu du fichier au cbp à copier et #chemin_b# l'emplacement sur votre ordi où copier le fichier (le plus souvent "." pour le copier depuis le répertoire où vous lancez la commande). Très utile pour ramener son travail chez soit. Par contre, pensez à ramener tout en un coup, donc dans une archive par exemple (créée par "tar -cf <nom de l'archive>.tar <nom du dossier à archiver>") pour éviter de taper 42 fois votre mot de passe.

On peut aussi s'en servir dans l'autre sens pour envoyer ses fichiers au cbp (par exemple parce que vous avez créé votre géométrie sur l'avogadro de votre ordi perso puisque c'est impossible en ssh au cbp), juste en inversant les deux termes :

```
scp -oProxyCommand="ssh login@ssh.ens-lyon.fr -W %h:%p"
```

```
(suite) #chemin_a# login@machine.cbp.ens-lyon.fr:#chemin_b#
```

Annexe

La meilleure citation gaussienne que j'ai croisée jusque là : *EVOLUTION IN THE BEGINNING GOD MADE MAN.*

HE WORKED THE DIRT WITH HIS HANDS.

THEN GOD DECIDED TO GIVE MAN SCIENCE,

MAN SHOULD BE USEFUL (LIKE AN APPLIANCE.)

BUT SOON MAN'S KNOWLEDGE GREW TOO GREAT,

TO THE HEADING SCIENCE HE COULDN'T RELATE

SO GOD SAID "LETS HAVE DIFFERENT FIELDS,

WE'LL DIVIDE THEM UP, SEE WHAT IT YIELDS."

ENTER PHYSICS, CHEMISTRY, BIO

THEY SPRANG UP AROUND THE NILE.

BUT CHEMISTRY SURPASSED THE REST

FAR AND AWAY IT WAS THE BEST.

EXPERIMENT WAS GOD'S DECREE

EXPERIMENT DID CHEMISTS WITH LOTS OF GLEE

BUT SOON THEIR TESTS BEGAN TO BOG,

THEY FOUND THEMSELVES DEEP IN A FOG.

"GIVE US SOME HELP" THEY CRIED TO GOD.

GOD LOOKED AROUND, THEN GAVE THE NOD.

ENTER EINSTEIN, HEISENBERG, BOHR,

REACTION PATHS WERE CLEAR ONCE MORE.

WITH THEORISTS GOD SEND DIRECTION

AND THEN HE STOPPED, HE'D MADE PERFECTION.

– THOMAS A. HOLME, 1981