

Introduction

- Le LCP dispose d'une grappe depuis 2002
- Elle a été constamment agrandie et renouvelée depuis, de 8 machines mono-processeur en 2002 jusqu'à 85 machines multi-cœurs aujourd'hui.
- Elle est divisée en deux parties :
 1. Une partie *privée*, pour les machines achetées par des équipes sur contrat (ANR ...). 57 machines.
 2. Une partie *publique*, pour les machines achetées par le laboratoire ... ou les machines de la partie privée une fois le contrat terminé.



La grappe du LCP

Description matérielle

Architecture grappe très classique :

- 85 nœuds de calculs (*plaques*), de type Dell R420 (bi-processeurs multi-cœurs).
- Nœuds de différentes générations, de 8, 12 ou 16 cœurs
- 988 cœurs en tout
- Fréquences entre 2 et 2.5 Ghz par cœur
- Entre 1 et 3 Go de mémoire par cœur
- Très peu d'espace disque local (typiquement 100 Go), utilisé comme un espace *scratch*
- Une machine frontale, servant via NFS un espace disque de 24 To
- Réseau interne Ethernet Gigabit
- Accès Ethernet 100 Mbit au réseau du laboratoire
- Trois racks
- 12 onduleurs 8 ports 3000 Watts
- 1 armoire de climatisation

Description matérielle : ce qu'il n'y a pas

- Pas de réseau rapide. Raison : coût
- Pas de *blades*. Raisons : coût initial trop important
- Pas de GPUs. Raison : pas d'applications de production pour l'instant

Description logicielle. Gestion des travaux

- Toutes les machines sont sous Linux (Scientific Linux 6.5)
- Gestion des travaux par Torque (PBS *open source*), à partir de la frontale.
- Utilisation du *scheduler* fourni avec Torque (pas Maui), avec prise en compte du *fair share*.
- Queues sans limites de temps pour partie privée, avec limites de temps pour la partie publique (8 heures, 1 jour, 4 jours)
- Spécialisation des queues et des machines en fonction du nombre de cœurs demandés (pour éviter la fragmentation)
- Pas de travaux multi-nœuds, sauf pour tests
- Zéro administration (ou presque . . .)

Description logicielle. Applications

- Logiciels de chimie quantique *standard* : Gaussian, Molpro, Gamess, NwChem ...
- Logiciels de dynamique moléculaire *standard* : Amber, Charmm
- Logiciels développés ou co-développés au laboratoire : Demon (DFT), ElVibRot et MatAux (dynamique quantique), Gibbs (MC), Newton (MD)
- Applications graphiques de chimie sur la frontale (VMD, GaussView, Rasmol, Molden)

Parallélisation

- Toutes les applications de calcul, standard ou développées en interne, sont parallélisées.
- Certaines le sont avec MPI (Demon, Amber), d'autres avec OpenMP (Gaussian), et d'autres mixtes OpenMP + MPI.
Il y a même Gamess qui utilise TCGMSG (pré-MPI)
- Dans tous les cas, utilisation mono-nœud (car pas de réseau rapide) sauf dans certains cas (calculs très largement indépendants, comme le *parallel tempering*)
- Efficacité de la parallélisation variable selon le type et la taille du problème traité

Utilisation

- Une vingtaine d'utilisateurs, de plusieurs groupes et de plusieurs équipes (y compris expérimentateurs), permanents, doctorants, stagiaires
- Utilisation comme une ressource *de proximité*, avec seulement deux contraintes :
 1. Limites mémoire proportionnelles au nombre de cœurs demandés (sinon la machine s'écroule à cause du *swap*)
 2. Activité NFS modérée (sinon tout s'écroule)
- Ces contraintes mises à part, gestion très souple :
 - Réservation de machines avec des configurations particulières (exemple récent : un projet utilisant Gamess qui avait besoin de 1 Go de mémoire partagée), ou pour des projets urgents (fins de thèses ...)
 - Pour des tâches de type *scan*, travaux en *rafale* (double parallélisation : au niveau job et au niveau application ...)
 - Calculs à faible efficacité parallèle, mais utilisant plusieurs cœurs pour des raisons de besoins mémoire, ou pour des raisons de temps de restitution

Mais aussi ...

- Pour les gros projets, ou bien les projets qui nécessitent des architectures plus sophistiquées (mémoires ou espaces disque extrêmes, parallélisme extrême, réseaux rapides), utilisation des machines du Genci (deux ou trois projets selon les années)
- Pour les projets en co-développement, utilisation des machines du partenaire

Exemples d'utilisation

Les transparents suivants présentent quelques exemples récents d'utilisation intensive de la grappe.

Simulations Monte-Carlo / DFT de spectres d'absorption UV-visible

Pierre Archirel

But : Interpréter des spectres d'absorption UV-visible d'espèces radiolytiques obtenus au laboratoire (plate-forme de Cinétique Rapide ELYSE)

Logiciels employés : Gibbs (Monte-Carlo) couplé à Gaussian (ab-initio, DFT)

Stratégie :

- utilisation de Gibbs pour générer des configurations du système à partir d'une configuration initiale et de *mouvements* (translations, rotations ...)
- délégation des calculs d'énergie à Gaussian (DFT + PCM)
- acceptation ou rejet de la configuration selon critère de Metropolis
- une fois atteinte la convergence, calculs TDDFT sur les états excités.

Temps de calcul :

- Typiquement, un calcul DFT = 1 minute, donc une semaine = 10 000 pas MC
- Calculs sur 8 cœurs en général
- 2416 jours (équivalents mono-cœur) de calcul en 2014

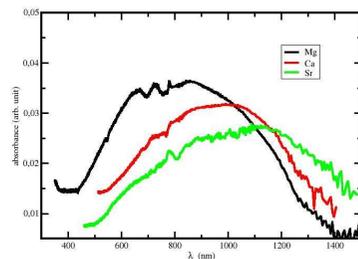
Simulations Monte-Carlo / DFT de spectres d'absorption UV-visible

**réduction radiolytique des perchlorates
de métaux alcalino-terreux**

(coll. M. Mostafavi, P. Pernot, LCP)

**dans le THF
le spectre de l'électron est déplacé
en présence de perchlorates**

quelle espèce se manifeste?



e-solv + Mg(ClO₄)₂ → [Mg(ClO₄)₂]⁻ ?

Mg —
Ca —
Sr —

- **paire soluté - électron?**
- **anion?**

**déplacement vers le rouge:
Sr > Ca > Mg**

Simulations Monte-Carlo / DFT de spectres d'absorption UV-visible

oxydation de la méthionine

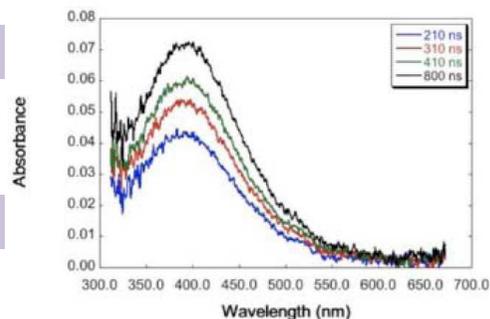
(coll. C. Houée, LCP, J. Bergès, LCT)

**manip Elyse :
radiolyse pulsée d'une solution
du dipeptide méthionine - valine dans l'eau**

meth - val + OH. → produit oxydé + OH-

800 ns

200 ns

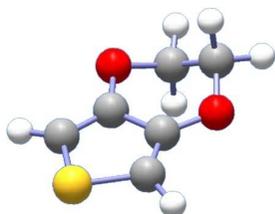


**on voit
la formation progressive
du produit d'oxydation**

bande à 400 nm

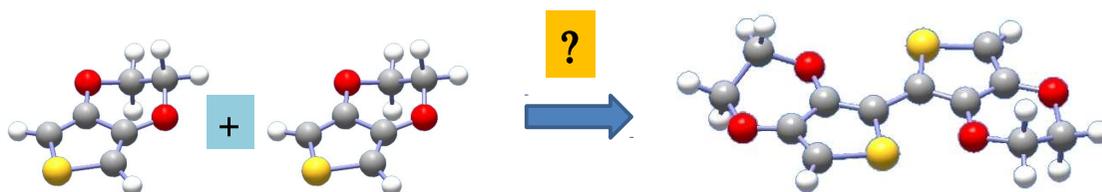
Simulations Monte-Carlo / DFT de spectres d'absorption UV-visible

Polymérisation de l'EDOT (coll. S. Rémita, C. Coletta, LCP)



EDOT: éthylène dioxythiophène

sous radiolyse dans l'eau :
polymérisation ?



Étude de systèmes auto-organisés : les tensio-actifs

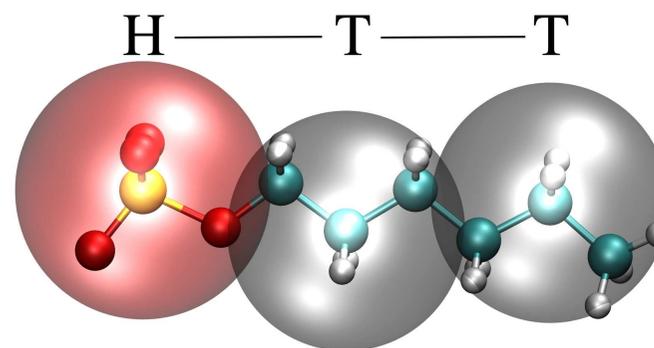
Bernard Rousseau

But : Étudier l'auto-organisation de tensioactifs

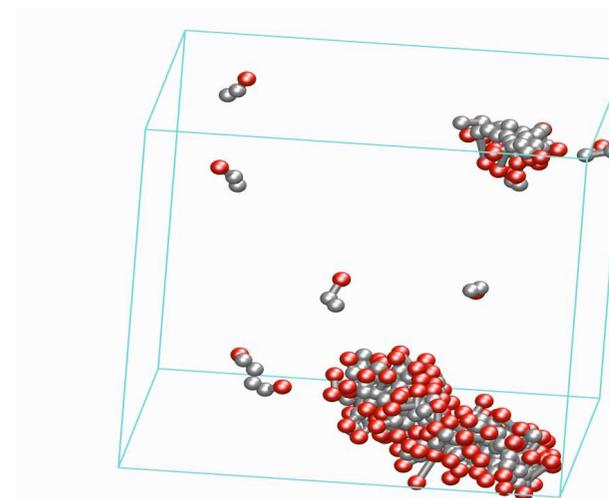
Logiciels employés : Newton (logiciel de dynamique moléculaire développé par lui-même)

Stratégie :

- L'auto-organisation de tensioactifs se fait à des échelles de taille et de temps inaccessibles par dynamique moléculaire classique
- Approche mésoscopique (*gros-grain*). Méthode DPD (*Dissipative Particle Dynamics*) : on travaille sur des *billes*, correspondant à des groupes d'atomes
- Développement d'un modèle permettant de reproduire, sur des sodium alkyl-sulfates, la concentration micellaire critique (concentration au-delà de laquelle les tensioactifs s'organisent en micelles) et le nombre d'agrégation (nombre de tensioactifs par micelle)
- Avec ce modèle, prédictions pour d'autres structures de tensio-actifs



Le modèle gros-grain



Modèle atomistique sous-jacent

Temps de calcul :

- Malgré l'utilisation de modèle mésoscopique, ces approches nécessitent toujours de fortes ressources de calcul.
- Un calcul typique prend de l'ordre de 24 h sur 16 cœurs
- 732 jours (équivalents mono-cœur) de calcul en 2014

Etude de la solubilité de molécules de gaz dans un polymère

Bernard Rousseau

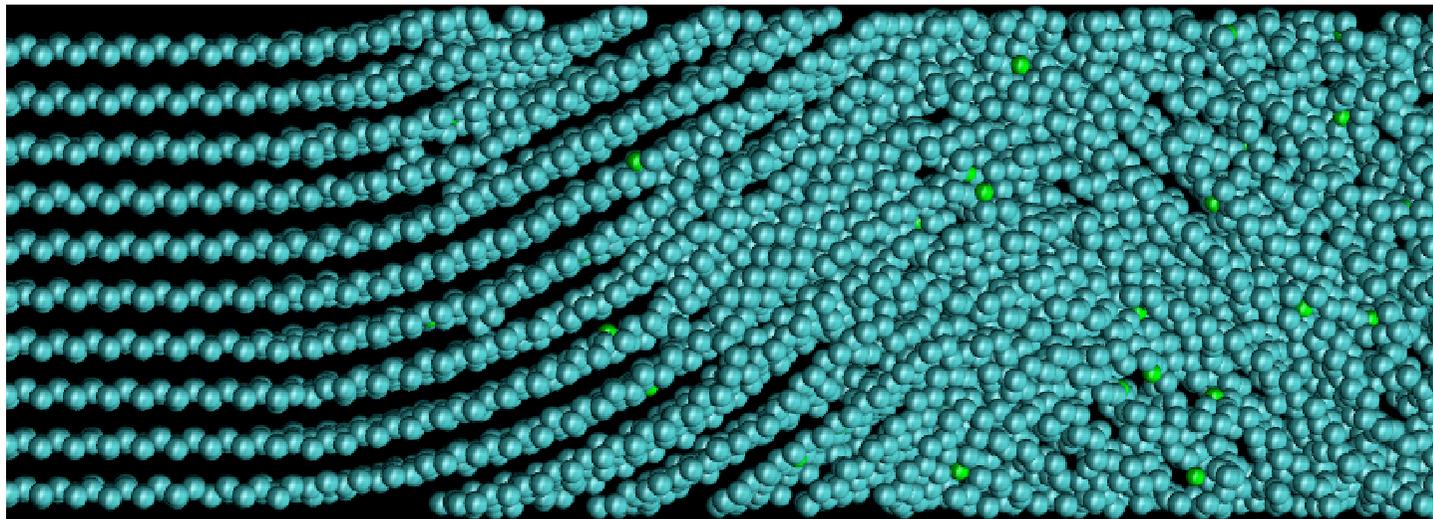
But : Étudier les propriétés *barrière* (imperméabilité) du PVDF (polyfluorure de vinylidène)

Logiciels employés : Newton (logiciel de dynamique moléculaire développé par lui-même) + Gibbs (logiciel de Monte-Carlo moléculaire développé en collaboration avec IFP Énergies Nouvelles)

Stratégie :

- Méthode de Monte Carlo dans l'ensemble dit *osmotique (semi-grand canonique)* :
 - pour le gaz : température, pression et potentiel chimique imposés
 - pour le polymère : nombre de molécules imposé
- Mouvements de type *insertion-destruction* pour les molécules de gaz
- Probabilité très faible de réussite (système très dense), donc important de bien relaxer la zone entourant la molécule nouvellement insérée

D'où utilisation d'un mouvement global de dynamique moléculaire (Monte-Carlo *hybride*)



Contrôle quantique

Aurélie Chenel (doctorante) et Michèle Desouter

But : Guider l'évolution d'un système à l'aide d'une impulsion laser mise en forme

Logiciels employés : Mataux (développé au laboratoire)

Stratégie :

- Beaucoup de degrés de liberté. Impossible de tout prendre en compte au niveau quantique
- On prend en compte un petit nombre de d.d.l. (le *système*) de façon explicite, et tout le reste (l'*environnement*) de façon implicite.
- Méthode des matrices auxiliaires : résolution d'un système d'équations couplées dont les inconnues sont des matrices
- La dimension de ces matrices dépend de la taille de la grille nécessaire pour discrétiser la surface d'énergie potentielle associée au système étudié.
- En se limitant à deux degrés de liberté explicites, on aboutit à des tableaux complexes (double précision) de dimensions (128,64,128,64), soit 1 Go par tableau. Et il y en a beaucoup ...

Temps de calcul :

- Calculs sur machines de 12 cœurs en général, utilisant 32 Go de mémoire
- 7150 jours (équivalents mono-cœur) de calcul en 2014

Contrôle quantique

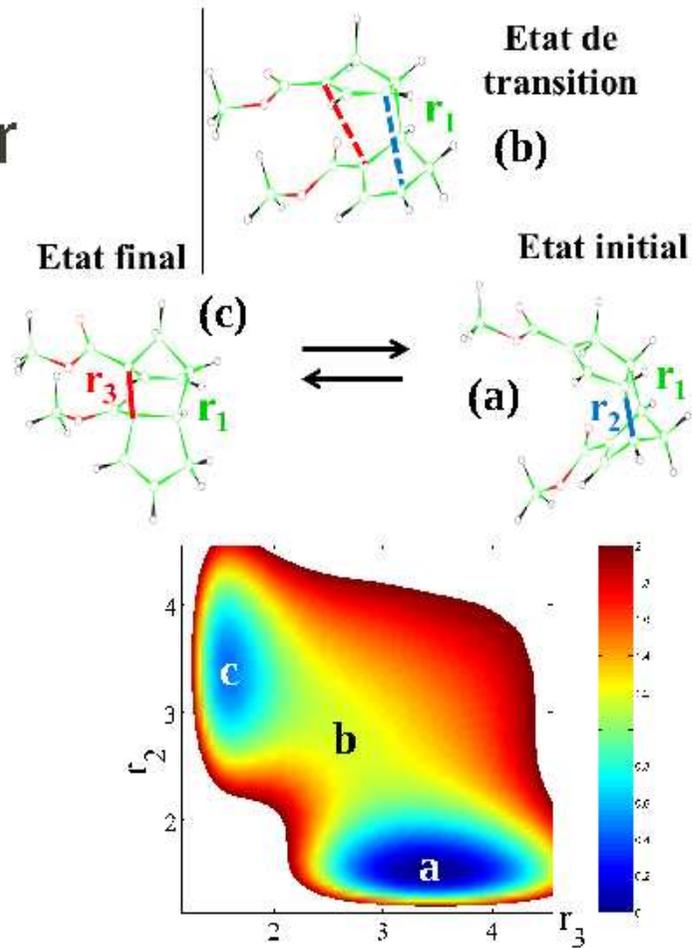
Contrôle d'une isomérisation par impulsion laser

Réarrangement de Cope de l'ester de Thiele = **isomérisation** de a vers c.



Rupture de la liaison r_2 et formation de la liaison r_3 .

Description d'un **système en interaction avec un environnement en présence d'un champ laser externe.**



Dynamique quantique : spectres IR de molécules déformables

David Lauvergnat

But : Interpréter et simuler des spectres infrarouges de petites molécules très anharmoniques : plusieurs minima accessibles, surface d'énergie potentielle plate, avec possibilité de dédoublement des niveaux vibrationnels par effet tunnel

Logiciels employés : ElVibRot et Tnum (développés au laboratoire)

Stratégie :

Obtenir les états stationnaires par diagonalisation sans construire explicitement la matrice Hamiltonienne (méthode itérative, Davidson)

- Génération d'une base compacte et d'une grille *sparse* (de type Smolyak).
- Représentation des termes de l'Hamiltonien sur la grille.
- Diagonalisation par une méthode itérative en calculant l'action de l'Hamiltonien sur un vecteur propre d'essai.

Temps de calcul :

Pour un système avec 12 degrés de liberté, $4 \cdot 10^6$ fonctions de base et $900 \cdot 10^6$ points de grille.

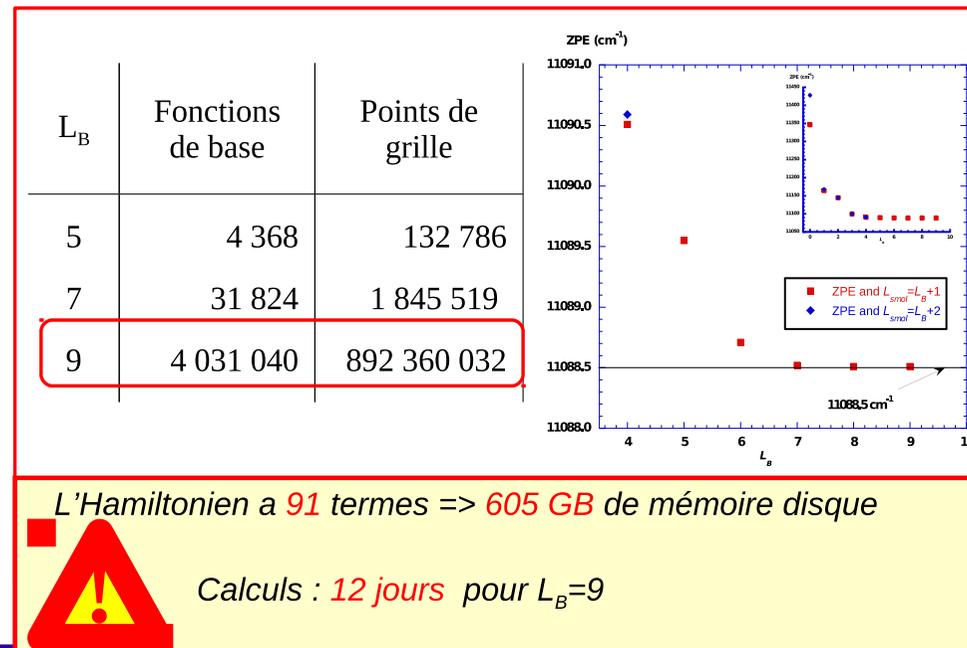
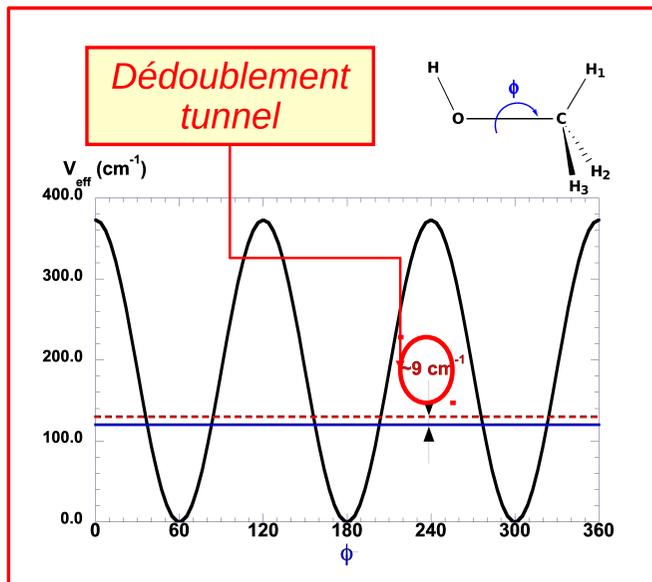
Remarque : une grille produit-direct aurait environ 10^{13} points.

- Représentation des termes de l'Hamiltonien sur la grille : 2 jours sur 12 cœurs
- Diagonalisation : environ 10 jours (pas encore parallélisé)

Spectre de torsion du méthanol : effet tunnel

Description quantique avec 12 degrés de liberté :

- Description des fonctions d'ondes sur une base et sur une grille « sparse » de type Smolyak
- Description de la fonction l'Hamiltonien sur la grille



Description théorique de la dynamique interne des protéines

F. Caillez, I. Demachy, N. Gillet, A. de la Lande, B. Lévy, K. Merabti, J. Ridard, J. Verdière (et collaboration avec l'équipe Biologie du laboratoire)

Intérêts :

1. enjeu en biologie et en pharmacologie,
2. type de système complexe particulièrement bien documenté du point de vue expérimental.

But :

Logiciels employés : Gaussian, Molpro, Gamess, Demon

- mécanique quantique : Gaussian, Molpro, Gamess, Demon
- dynamique moléculaire : Amber, Charmm
- et parfois un logiciel d'interface : Cuby, pour enchaîner automatiquement les différents calculs

Stratégie :

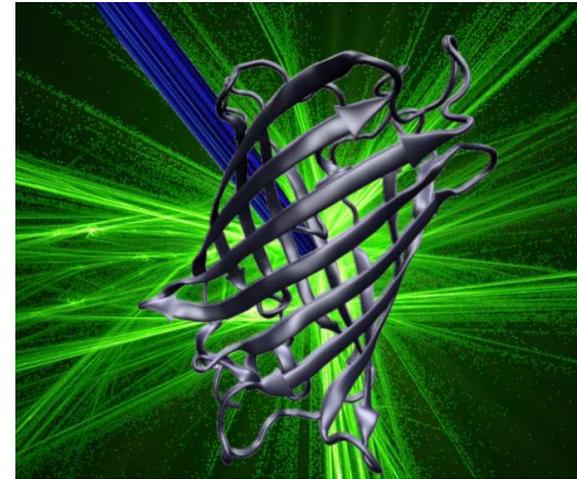
- Les logiciels de chimie quantique (DFT, QM/MM) sont utilisés de façon répétitive (*scan*) sur des sous-systèmes significatifs (chromophores, flavines ...).
- Cela permet d'introduire de nouveaux termes de potentiel (par exemple couplage entre torsions) et d'en calculer les paramètres.
- Ces termes sont alors introduits dans les logiciels de dynamique moléculaire
- Des calculs (lourds ...) de dynamique moléculaire sont alors lancés sur la protéine complète

Résultats obtenus :

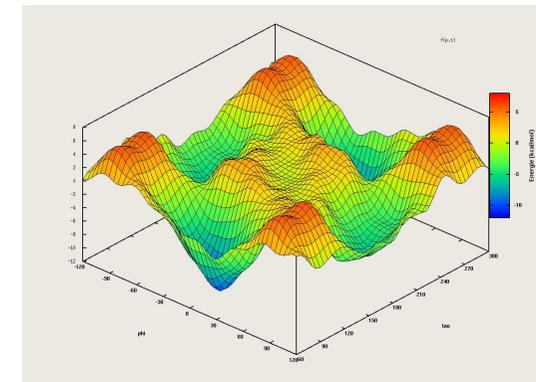
- Concernent les protéines fluorescentes et les protéines catalysant des transferts d'électrons
- Portent sur la cohérence/décohérence, le choix entre différents mécanismes de réaction, le rôle de différentes mutations ...

Temps de calcul :

- 4531 jours (équivalents mono-cœur) de calcul en 2014
- ... mais aussi calculs au Cines, ainsi que dans de nombreux autres centres de calcul de par le monde ...



Protéine fluorescente (vue d'artiste)



Surface d'énergie potentielle pour le chromophore

Calibration globale de champs de force pour la dynamique moléculaire

Fabien Cailliez et Pascal Pernot

But :

- Explorer l'espace des paramètres qui définissent le potentiel intermoléculaire,
- ... pour trouver l'ensemble des jeux de paramètres compatibles avec les incertitudes,
- ... en reproduisant un jeu de données de référence expérimentales,
- ... et en incluant les incertitudes des simulations et expérimentales.

Stratégie :

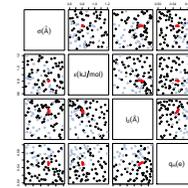
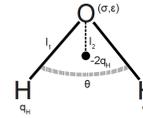
- Exploration de l'espace des paramètres trop coûteuse par simulation moléculaire.
Donc utilisation de méta-modèles de prédiction, couplés à une méthode d'optimisation globale de type EGO (efficient global optimisation).
- Calibration de ces méta-modèles par pré-échantillonnage de l'espace des paramètres, avec simulation de dynamique moléculaire sur ce sous-ensemble

Application au modèle d'eau TIP4P rigide :

- Calibration des 4 paramètres
- ... pour reproduire la densité de l'eau liquide en fonction de la température entre 253 et 350K.

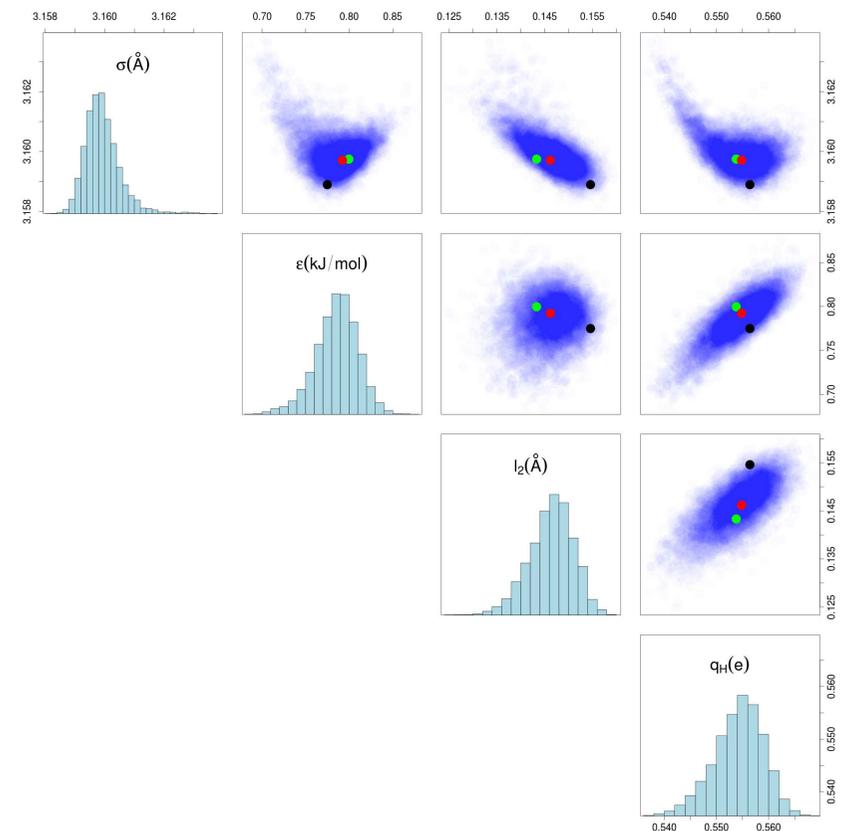
Temps de calcul :

- 450 simulations de dynamique moléculaire (avec Newton)
- chacune d'une durée d'environ une semaine sur 8 cœurs
- 1308 jours (équivalents mono-cœur) de calcul en 2014



Le modèle TIP4P

Jeux de paramètres testés



Résultats : densités de probabilités des paramètres

Conclusion

- Grande diversité d'applications et grande diversité de logiciels utilisés, ... le plus souvent dans des conditions originales (couplages, enchaînements)
Donc beaucoup d'utilisations exigeant une grande souplesse dans l'administration
- Beaucoup de logiciels (co-) développés en interne
Donc beaucoup d'utilisations en mode *test* et en mode *débogage* ...
- Nombreuses limitations : pas de réseau rapide, pas d'entrées-sorties rapides, pas de grands espaces mémoires
Donc impératif de pouvoir accéder aussi à d'autres plateformes