

Dynamiques moléculaires quantiques et classiques

Roberto Marquardt, Jérôme Hénin, François Dehez et Christophe Chipot

Résumé

La mécanique des atomes et des molécules étant essentiellement fondée sur les principes de la mécanique quantique, nous devons nous en saisir pour répondre à la question : « Comment bougent les molécules ? » C'est le cadre de la « dynamique moléculaire quantique ». Cependant, vue la complexité de la majorité des molécules, surtout de celles d'intérêt biologique, leur étude exacte selon les règles strictes de la mécanique quantique semble inextricable. Pour la modélisation de ces molécules, on utilise alors une approche simplifiée, selon laquelle les forces qui régissent le mouvement des atomes sont décrites par des fonctions mathématiques approchées, déduites soit de la théorie quantique, soit de données expérimentales. Le mouvement des atomes eux-mêmes est ensuite déduit des lois de la mécanique classique. C'est le cadre de la « dynamique moléculaire classique ». Cet article passe en revue les efforts de développement méthodologiques réalisés en France et dans le reste du monde, qui nous ont permis, au cours des dix dernières années, d'accomplir des progrès considérables vers notre objectif ultime de compréhension du mouvement des atomes dans les molécules. Il décrit également les limitations actuelles des méthodes et les perspectives de la discipline.

Mots-clés

Dynamique moléculaire, mécanique quantique, mécanique classique, dynamique réactionnelle, simulation des processus de déformation moléculaire, super-ordinateurs.

Abstract

Quantum and classical molecular dynamics

Because atomic and molecular mechanics is essentially based on the principles of quantum mechanics, we need to infer the latter to search to answer to the question: "How do molecules move?" This is the framework of "molecular quantum dynamics". However, because of the huge complexity of the majority of molecules, in particular of those having a direct biological interest, the study of the motion of their atoms following strictly the laws of quantum mechanics is quite impossible. To model these molecules, one therefore uses a simplified approach, in which the forces acting on the atoms are derived from approximated mathematical formulae. These are derived either from quantum theory or from experimental data. The motion of the atoms is then deduced from the laws of classical mechanics. This is the framework of "classical molecular dynamics". This text reviews some of the methodological developments realized in France, as well as elsewhere, during the last decade, which have allowed scientists to advance considerably toward a more complete understanding of atomic motion in molecules. It also discusses current limitations of the methods and gives an outlook at potential further progresses in this discipline.

Keywords

Molecular dynamics, quantum mechanics, classical mechanics, reaction dynamics, molecular deformation processes, simulation, supercomputers.

Comment bougent les molécules ? La réponse à cette simple question est bien plus difficile qu'il n'y paraît à première vue. La chimie théorique peut y apporter des éléments de réponse importants, tant par l'interprétation de données expérimentales, que par la prédiction de comportements inédits des atomes au sein des molécules qu'ils forment. Appliquée aux réactions chimiques, cette puissance de prédiction est encore plus intéressante puisqu'elle permet d'avancer vers l'idéal d'une « chimie propre », sans produire de déchets ou de produits collatéraux non souhaités.

Les molécules sont constituées d'atomes et la forme de l'édifice moléculaire dépend de la mise en place d'un équilibre fin entre les différentes forces agissant à la fois entre les noyaux des atomes, les électrons les entourant, ainsi qu'entre les électrons des différents atomes. Il serait faux de penser que les structures moléculaires ainsi formées sont rigides, voire stables : les molécules se transforment, c'est la vie de la chimie !

À titre d'exemple, prenons la molécule d'acide cyanhydrique, HCN (*figure 1*). Dans une de ses formes stables, la molécule est linéaire et de connectivité H-C-N. Par contre, même à très basse température, les atomes bougent autour de leurs positions relatives à l'équilibre, pouvant ainsi étirer ou raccourcir les longueurs de liaisons existantes (flèches droites dans la *figure 1a*), ou bien rendre la molécule pliée (flèches curvilignes).

Admettons que les atomes gagnent sous une forme ou sous une autre une quantité importante d'énergie. Dans ce cas, l'amplitude de leur écart de l'équilibre pourra être grande et ils pourront se réorganiser dans une nouvelle structure, elle aussi stable, de connectivité H-N-C (*figure 1b*). Cette transformation chimique est l'exemple d'une *isomérisation*. Dans l'exemple de la *figure 1*, la forme b est localement stable, mais globalement moins stable d'environ 42 kJ/mol que la forme a (la barrière entre les deux formes étant à environ 150 kJ/mol). On pourrait penser que cet exemple est alors plutôt d'intérêt académique. Néanmoins, le concept

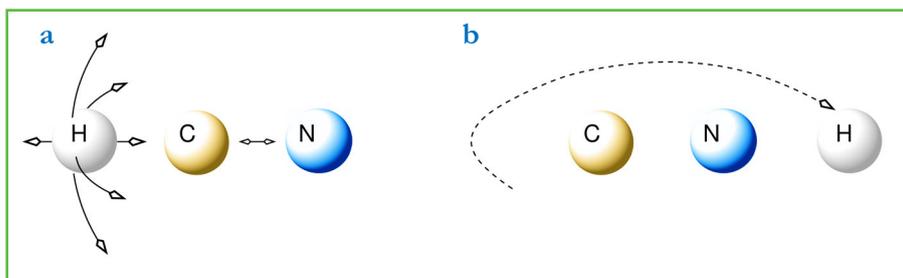


Figure 1 - Schéma indiquant des mouvements possibles de l'acide cyanhydrique (a), ainsi qu'un mouvement possible de grande amplitude donnant lieu à l'isomérisation de la molécule (b).

enzyme ou une protéine. Pour des systèmes de telles tailles, nous devons nous limiter à étudier les mouvements des atomes selon les lois de la mécanique classique, ce qui n'est guère moins complexe que le calcul purement quantique dans des petites molécules, puisqu'il s'agit souvent d'étudier des systèmes composés de 10^3 à 10^5 atomes. Les défis computationnels sont aussi importants ici. Le défi ultime est conceptuel : comment faire le lien entre les types de mouvement qui sont différents selon la mécanique adoptée pour les caractériser ?

qu'il illustre est profondément attaché à notre vision habituelle de la réaction chimique : les atomes bougent, accompagnés de leurs électrons, d'une partie de la molécule à l'autre comme s'ils étaient des boules. En fait, les images statiques que nous pouvons nous faire aujourd'hui des atomes grâce à des outils comme le microscope à effet tunnel ou le rayonnement X nous enseignent que les densités électroniques des atomes possèdent véritablement une forme sphérique. Mais est-ce vrai aussi pour les atomes en mouvement ?

L'isomérisation de l'acide cyanhydrique a fait l'objet de nombreux travaux expérimentaux et théoriques depuis 1993 (cf. [1] et les travaux mentionnés), où il est montré que notre vue traditionnelle du mouvement des atomes n'est pas conforme avec les lois intimes du monde microscopique, c'est-à-dire les lois de la mécanique quantique : les atomes et les densités électroniques qui les accompagnent bougent de façon non localisée – par exemple, l'hydrogène dans HCN peut, lorsqu'il bouge vers l'azote, être simultanément localisé sur le carbone et l'azote, ce qui est difficile à imaginer dans notre monde macroscopique.

Nous n'avons pas raison de penser que ce constat soit différent pour le mouvement de l'hydrogène dans une molécule plus grande, telle que le tryptophane (figure 2).

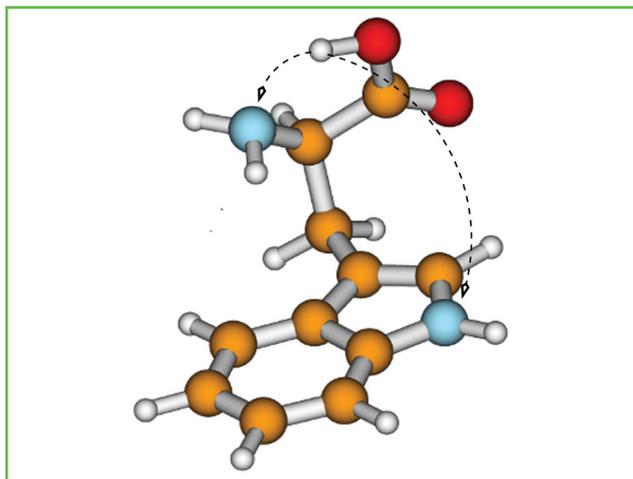


Figure 2 - Modèle du tryptophane dans une de ses structures stables.

H en blanc, C en orange, O en rouge, N en bleu.

L'hydrogène du groupement carboxylique pourrait bouger vers l'un des deux atomes d'azote, soit celui de la chaîne alanyl, soit celui de la partie indole, un mouvement très probablement lié à un transfert de charge. Le traitement théorique et la description purement quantique de ce mouvement représentent encore un vrai défi au niveau computationnel. Et il est hors de question de vouloir calculer, sur la base de la mécanique quantique, le mouvement des atomes dans une

Dans cet article, nous passerons en revue les développements de méthodes réalisés par des chercheurs français pour traiter les défis susmentionnés dans les deux domaines de la dynamique moléculaire classique et quantique.

Dynamique moléculaire quantique

En mécanique quantique, toute information concernant le système moléculaire est contenue dans la fonction d'onde $\Psi(x^{(n)}, x^{(e)}, t)$, qui est une fonction complexe des vecteurs $x^{(n)}$, $x^{(e)}$, et du temps t – nous utilisons la notation simplifiée $x^{(n)}$ pour désigner un vecteur position de dimension $3 \times N_a$ de l'ensemble des noyaux, et $x^{(e)}$ pour désigner un vecteur position de dimension $3 \times N_e$ de l'ensemble des électrons, où N_a et N_e sont, respectivement, le nombre de noyaux et d'électrons présents dans la molécule. Cette fonction est en principe observable expérimentalement, du moins son module. En théorie, cette fonction est par exemple calculée comme solution de l'équation de Schrödinger :

$$i \frac{\hbar}{2\pi} \frac{\partial}{\partial t} \Psi(x^{(n)}, x^{(e)}, t) = \hat{H} \Psi(x^{(n)}, x^{(e)}, t) \quad (1)$$

où i est l'unité imaginaire ($i^2 = -1$), \hbar est la constante de Planck et \hat{H} est l'opérateur d'énergie (l'hamiltonien). Cette équation est une équation différentielle partielle, linéaire. En dépit de sa linéarité, pour cause de la haute dimension des espaces de vecteurs $x^{(n)}$ et $x^{(e)}$, trouver les solutions de l'équation (1) constitue un énorme défi, à la fois numérique et conceptuel.

De nos jours encore, on cherche à résoudre ce problème par une décomposition artificielle de la fonction d'onde, $\Psi(x^{(n)}, x^{(e)}, t) = \sum_k \psi_k^{(n)}(x^{(n)}, t) \psi_k^{(e)}(x^{(e)}; x^{(n)})$, où $\psi_k^{(e)}$ sont appelées *fonctions d'onde électroniques* (adiabatiques), $\psi_k^{(n)}$ sont les *fonctions d'onde nucléaires*, et l'ensemble de ces fonctions est supposé dénombrable. Chaque fonction $\psi_k^{(e)}$ décrit un état électronique dont l'énergie $V_k(x^{(n)})$ est considérée comme une énergie potentielle pour le mouvement des noyaux. Normalement, les électrons s'adaptent facilement à chaque nouveau positionnement des noyaux, les suivant donc de façon « adiabatique ». Le développement de méthodes de calcul de $\psi_k^{(e)}$ et de V_k fait l'objet de multiples études en France (voir ci-après) comme à l'étranger. Historiquement, la formulation analytique des (*hyper*-)surfaces de potentiel V_k a été développée dans le contexte de la spectroscopie infrarouge, ce qui est étroitement lié à la dynamique moléculaire quantique par le biais de surfaces de potentiel globales (cf. [2] pour une revue sur l'état de l'art de la représentation analytique globale de surfaces de potentiel).

En dynamique moléculaire quantique, le mouvement des noyaux est obtenu comme solution de l'équation de Schrödinger nucléaire :

$$i\frac{\hbar}{2\pi}\frac{\partial}{\partial t}\psi_k^{(n)}(x^{(n)}, t) = \sum_{k'} \hat{H}_{kk'} \psi_{k'}^{(n)}(x^{(n)}, t) \quad (2)$$

Les éléments matriciels $\hat{H}_{kk'}$, sont des opérateurs agissant sur les degrés de liberté nucléaires. Les éléments diagonaux ont la forme $\hat{H}_{kk} = \hat{T}^{(n)} + V_k$, où $\hat{T}^{(n)}$ est l'opérateur d'énergie cinétique nucléaire. Les éléments non diagonaux, aussi appelés *couplages non adiabatiques*, sont plus compliqués à formuler. Dans l'*approximation de Born-Oppenheimer*, on pose $\hat{H}_{kk'} \equiv 0$ (si $k \neq k'$). Cette approximation est souvent excellente pour le traitement de la dynamique dans l'état électronique fondamental, telle que celle décrite dans la *figure 1*. La dynamique quantique faisant intervenir les états électroniques excités, comme lors d'un transfert de charge (*figure 2*), elle nécessite normalement la prise en compte des couplages non adiabatiques. Les méthodes pour le traitement des couplages non adiabatiques en dynamique quantique sont très peu développées, en France comme ailleurs. Un problème lié aux couplages non adiabatiques est l'effet Renner-Teller des molécules linéaires. L'équipe de chimie théorique du Laboratoire de Modélisation et Simulation Multi Échelle à l'Université Paris-Est Marne-la-Vallée développe des méthodes pour le traitement de l'effet Renner-Teller en spectroscopie moléculaire.

En France, le développement méthodologique dans le domaine de la dynamique quantique connaît plusieurs contributions différentes et complémentaires. Dans la suite, nous donnons certains exemples sans prétendre toutefois fournir une liste exhaustive, voire complète des travaux réalisés en France. Nous repérons ainsi d'abord les travaux pour la formulation mathématiquement exacte d'opérateurs de l'énergie cinétique nucléaire $\hat{T}^{(n)}$ exprimés en coordonnées internes. Ces travaux ont vu une suite de recherches menant à l'utilisation, aujourd'hui très répandue, des coordonnées appelées *polysphériques* [3]. Le Laboratoire de chimie physique de l'Université de Paris-Sud à Orsay continue à développer des formulations analytiques de l'opérateur $\hat{T}^{(n)}$ qui sont numériquement efficaces et exactes [4].

Pour la résolution de l'équation (2), plusieurs méthodes ont été implémentées et développées en France, notamment pour l'étude des collisions. Notons ici les méthodes utilisant une décomposition spectrale comme, par exemple, la méthode des canaux couplés employée dans des calculs de diffusion et développée à Meudon et Rennes [5], les méthodes développées pour le mouvement électronique dans le Laboratoire de Chimie Physique, Matière et Rayonnement à Paris ou celles du Centre des Lasers Intenses et Applications à Bordeaux, ainsi que le code URIMIR (*cf.* [6] et les travaux mentionnés) pour le calcul de l'excitation laser du mouvement vibrationnel.

Les différents groupes travaillant dans ce domaine ont souvent créé leurs propres codes de propagation de paquets d'onde sur une grille numérique de l'espace de configuration des noyaux en utilisant la méthode de Feit et Fleck (« split-operator »), celle de la transformée de Fourier rapide (FFT), ou encore celle de Park et Light (SIL, « short iterative Lanczos »). Il est à noter ici le développement de méthodes pseudo-spectrales ainsi que de nouvelles fonctions de base discrètes [7] pour le traitement de la dynamique quantique de diffusion d'atomes aux surfaces.

Plus récemment, le code de calcul MCTDH (« multiconfigurational time dependent Hartree »), développé à Heidelberg

avec le concours de chercheurs français, a été beaucoup utilisé, notamment en France [1, 6, 8]. Dans la référence [6], un exemple plutôt rare de comparaison de méthodes de dynamique moléculaire quantique (MCTDH et URIMIR) est proposé.

À ce jour, tous ces développements permettent de traiter, dans tous ses degrés de liberté, des molécules ayant typiquement deux à cinq atomes. Pour discuter, par exemple, dans toutes ses dimensions la dynamique de tautomérisation décrite dans le tryptophane (*figure 2*), les méthodes purement quantiques sont loin d'être applicables. À leur place, les chercheurs développent des méthodes mixtes classiques-quantiques, dans lesquelles les paquets d'onde bougent le long de trajectoires classiques [9], ou dans lesquelles les paquets d'onde sont couplés à un ensemble de modes de vibration de façon statistique [10]. Le progrès dans ce domaine va à la rencontre d'une meilleure interprétation des résultats de la dynamique moléculaire classique, que nous aborderons dans la prochaine section.

Finalement, rappelons que la dynamique moléculaire quantique ne doit pas être confondue avec la dynamique moléculaire selon la formulation de Roberto Car et Michele Parrinello, qui est très populaire et qui est une dynamique moléculaire classique. Signalons aussi que les méthodes utilisées pour étudier la dynamique quantique électronique dans le domaine de la physique sont en train d'entrer, avec les progrès atteints dans la technologie d'impulsion laser attoseconde ainsi que des atomes ultra-froids, dans le domaine de la physique moléculaire et de la dynamique moléculaire quantique.

Dynamique moléculaire classique

Fondements et algorithmes de la dynamique moléculaire

Appréhender la fonction d'assemblages moléculaires de grande taille, comme ceux de la machinerie cellulaire, requiert la connaissance des aspects non seulement structuraux, mais également dynamiques de ces systèmes organisés. Or, en de nombreuses circonstances, les limitations intrinsèques des techniques expérimentales conventionnelles contrecarrent l'accès au détail microscopique de ces édifices moléculaires. La révolution informatique amorcée il y a une quarantaine d'années a considérablement modifié la donne en rendant possible leur étude par les simulations numériques reposant sur les premiers principes. En réalité, dès la fin des années 1950, Alder et Wainwright avaient pressenti que de telles expériences informatiques, réalisées dans un premier temps sur de petits systèmes modèles, notamment de sphères dures, étaient susceptibles de constituer un pont entre l'observation expérimentale macroscopique et son pendant microscopique [11]. Ce lien tangible entre deux échelles de taille distinctes nécessite la réplification spatiale périodique de l'échantillon simulé, ainsi affranchi des indésirables effets de bord. Idéalement, l'étude complète des assemblages moléculaires nécessiterait la résolution de l'équation de Schrödinger dépendant du temps (l'équation (2)). Dans la pratique, en revanche, l'intérêt se porte principalement sur la trajectoire des noyaux dans un potentiel V donné, que l'approximation de Born-Oppenheimer permet de générer en ayant recours aux équations classiques du mouvement. Dix ans après la première simulation de dynamique moléculaire de Alder et Wainwright, le physicien français Loup Verlet proposa un schéma d'intégration

numérique des équations newtoniennes, ainsi qu'un algorithme de liste de voisins pour le calcul des interactions interatomiques [12], encore usités de nos jours sous différents déguisements. Pierre angulaire des simulations numériques, la fonction énergie potentielle est minimaliste et limitée dans la plupart des cas à des termes harmoniques pour des degrés de liberté oscillants, des séries trigonométriques pour la rotation autour des liaisons simples et la combinaison de potentiels coulombien et de Lennard-Jones pour le calcul des interactions d'atomes non liés chimiquement. L'idée sous-jacente d'un champ de forces rudimentaire est de dilater les échelles de temps en réduisant à la portion congrue le coût de l'évaluation de l'énergie à chaque pas d'intégration ; c'est vrai pour les champs de force atomiques, ça l'est encore plus pour les approches dites gros grain. Qu'elle soit à l'équilibre thermodynamique ou hors équilibre, il serait fastidieux de passer ici en revue les nombreux développements qui ont propulsé la dynamique moléculaire au rang d'outil théorique robuste, complément incontournable d'un éventail de méthodes expérimentales. Pourtant, il n'est pas inutile de rappeler deux algorithmes qui ont constitué un tournant dans les simulations numériques : le thermostat de Nosé et Hoover et le barostat de Nosé et Andersen permettant respectivement de modéliser les assemblages moléculaires dans l'ensemble canonique et dans l'ensemble isobare-isotherme. À ceux-ci doit être ajoutée la prise en compte des interactions électrostatiques à longue portée et, pour ainsi dire liée à cette dernière, la propagation du mouvement avec des incréments de temps multiples. Toutefois, la véritable révolution dans la course aux échelles de temps et de taille reste bel et bien celle qu'accompagna l'avènement des architectures parallèles et des algorithmes à décomposition spatiale qui réduisent linéairement le temps de calcul avec le nombre de processeurs disponibles. Cette course effrénée à la simulation la plus longue ou à celle du plus gros édifice moléculaire éclipse quelque peu la masse de travail investie pour améliorer la représentation des forces interatomiques, notamment par l'introduction de polarisabilités ou de multipôles distribués, voire les deux, qui, de fait, alourdissent le coût du calcul, rançon inévitable d'une précision accrue. Les théoriciens français participent largement au développement de telles représentations, notamment à Paris, Strasbourg et Nancy. Elle écrante également les efforts de caractérisation de l'erreur associée aux schémas de propagation du mouvement, concept généralement ignoré dans les simulations flirtant avec les limites actuelles de la dynamique moléculaire, qu'elles soient de temps ou de taille.

Simulations « brute force » : que permettent-elles d'atteindre aujourd'hui ?

Les avancées réalisées tant dans le domaine des algorithmes que dans celui des architectures informatiques permettent aujourd'hui de modéliser de façon réaliste et non biaisée l'évolution temporelle d'assemblages moléculaires extrêmement complexes avec une résolution atomique. Les études de dynamique moléculaire les plus ambitieuses sont en grande partie réalisées dans le domaine de la biophysique auquel nous nous limiterons ici pour donner quelques exemples significatifs. D'un point de vue pratique, la taille des systèmes pouvant être modélisés actuellement n'est plus à proprement parlé limitée : elle peut atteindre quelques millions – par exemple, le ribosome [13] ou la capsid d'un virus [14] –, voire une centaine de millions d'atomes pour le chromatophore – pseudo-organelle d'environ 10^8 atomes

Sous les simulations, les calculateurs

Une simulation de dynamique moléculaire consiste à résoudre les équations du mouvement par petits pas de temps, la durée d'un pas étant limitée par les degrés de liberté les plus rapides présents dans le système, en général de l'ordre de la femtoseconde pour les modèles atomistiques. Un très grand nombre de pas d'intégration est donc nécessaire pour accéder aux échelles de temps intéressantes – de l'ordre de 10^9 pas pour une microseconde. Par ailleurs, l'évaluation des forces entre toutes les particules à chaque pas représente une charge de calcul élevée, croissant en principe comme N^2 pour le calcul d'un potentiel de paires, et en pratique comme $N \log N$, en tronquant les interactions à courte portée et en calculant celles à longue portée dans l'espace réciproque (approches de type sommes d'Ewald). Le besoin de puissance de calcul a conduit rapidement à la parallélisation des simulations de dynamique moléculaire, commençant à la fin des années 1980 avec le travail pionnier de Heller, Grubmüller et Schulten utilisant un réseau spécialement conçu de « transputers ». Cette tendance a connu une accélération marquée dans les années 1990 avec la parallélisation massive de programmes comme NAMD, LAMMPS et Gromacs 4. Dans tous les cas, la parallélisation repose sur une décomposition du système en sous-parties ou régions réparties entre les différents processeurs, ces régions pouvant parfois être remaniées en cours de simulation pour répartir la charge de calcul. Un bouleversement du paysage du calcul scientifique est en cours avec l'avènement du GPGPU (« general-purpose graphics processing unit »). Ce terme désigne un processeur graphique qui devient un processeur en virgule flottante générique, pourvu de nombreux cœurs, donc capable de traiter en parallèle la plupart des problèmes numériques. Cette évolution a une origine économique : avec la montée en puissance des jeux vidéo au graphisme sophistiqué, les GPU d'une puissance de calcul considérable sont devenus des produits grand public fabriqués en très grande série. Ainsi, ils constituent une solution performante et économique à l'achat pour le calcul scientifique. Les prémices de ce phénomène sont apparus il y a environ dix ans avec certaines consoles de jeux, vendues avec une très faible marge compensée par le coût des logiciels de jeux dédiés. En 2003, le centre de calcul NCSA à Urbana-Champaign (Illinois) a assemblé la première grappe de calcul composée de consoles Sony PlayStation 2. Désormais, les programmes de dynamique moléculaire exploitent en parallèle les CPU et les GPU présents dans des PC. On assiste d'une part au portage de codes classiques (GROMACS, NAMD), mais aussi à l'écriture de codes complètement nouveaux et exploitant exclusivement les GPU (HOOMD-blue). Cependant, la nature spécialisée des GPU fait peser de nombreuses contraintes techniques sur le GPGPU, en particulier le coût élevé des transferts entre la mémoire centrale et la mémoire liée au GPU – celles-ci sont reliées par un bus dont la bande passante peut devenir limitante. La répartition des tâches entre CPU et GPU, de natures différentes, pose aussi de nouveaux problèmes de parallélisation des algorithmes. Enfin, vingt ans après les calculs sur « transputer » de l'équipe de Klaus Schulten, on assiste à de nouvelles tentatives d'utilisation d'architectures dédiées. La plus ambitieuse et celle qui a rencontré le plus de succès est sans conteste la machine Anton développée par D.E. Shaw Research (voir l'article de Derreumaux et Lavery p. 109). Les vitesses de simulation accrues posent un problème inédit de stockage de données ; en effet, les trajectoires obtenues deviennent impossibles à stocker à un niveau détaillé. Une approche consiste donc à effectuer des analyses pré-programmées en temps réel sur la trajectoire. Si une nouvelle analyse est souhaitée, la dynamique doit être recalculée. Le stockage et la manipulation de gigantesques volumes de données, plutôt que la vitesse de calcul, deviennent donc les facteurs limitants.

abritant l'appareil photosynthétique chez certaines bactéries [15], un des assemblages moléculaires les plus gros jamais étudiés. Une des brides principales à la puissance de cet outil de modélisation est l'échelle de temps limitée des trajectoires de dynamique moléculaire, qui aujourd'hui peut

aller de quelques nanosecondes jusqu'à la milliseconde. Il y a bien évidemment une corrélation inverse entre le nombre d'atomes qui composent le système et le temps accessible, mais les récents progrès techniques permettent aujourd'hui d'atteindre simultanément des échelles de tailles – de plusieurs dizaines à quelques centaines de milliers d'atomes – et de temps – d'une à plusieurs centaines de microsecondes – compatibles avec la réalité de certains processus biologiques. Parmi les succès les plus spectaculaires obtenus ces quatre dernières années se trouve le décryptage *ab initio* à partir de la séquence du chemin de repliement rapide – de l'ordre de quelques microsecondes – de protéines composées de plusieurs dizaines d'acides aminés, permettant ainsi de faire un pas significatif vers un des Graal actuels de la biologie [16-18]. Un autre champ d'application ayant bénéficié des dernières avancées en dynamique moléculaire est le transport membranaire, domaine auquel participent activement plusieurs groupes de recherche en France. Là où la biologie structurale donne une vision parcellaire et essentiellement statique des protéines assurant le passage d'espèces chimiques au travers de la membrane cellulaire, les simulations statistiques permettent de décrire à l'échelle atomique les différentes étapes du chemin de transport. Ainsi, des trajectoires d'association spontanée de substrats à leurs cibles protéiques peuvent être réalisées sans aucun biais externe et répétées suffisamment pour obtenir un échantillonnage complet du phénomène, c'est le cas par exemple de l'association de l'adénosine diphosphate à sa protéine de transport de la membrane interne des mitochondries transporteur mitochondrial AAC [19]. Des trajectoires sur l'échelle de la microseconde ont également révélé les différentes étapes de la fermeture du canal potassium sensible à la tension Kv1.2 [20-21] ou encore l'ensemble des conformations essentielles à l'ouverture/fermeture du canal ionique ligand-dépendant de *Gloeobacter violaceus* (GLIC), un récepteur impliqué dans les processus de neurotransmission [22]. La *figure 3* illustre quelques exemples de protéines membranaires étudiées par simulations de dynamique moléculaire classique.

Échantillonnage préférentiel : simuler des processus lents en sélectionnant des variables collectives

De nombreuses approches ont été conçues pour accélérer l'échantillonnage dans les simulations de dynamique moléculaire. Ces techniques relèvent de deux grandes catégories. La première agit sur l'ensemble du système pour accélérer l'exploration des régions de basse énergie. La seconde, appelée échantillonnage d'importance (« importance sampling »), applique des biais sur des coordonnées choisies pour favoriser l'échantillonnage de régions d'intérêt, au détriment du reste de l'espace des phases. Le premier groupe de méthodes utilise plusieurs copies du système pour explorer plusieurs régions de l'espace des phases simultanément. Une méthode très utilisée, appelée « parallel tempering », consiste à simuler simultanément plusieurs copies du système à des températures différentes, et à échanger périodiquement ces copies entre elles, sur la base d'un critère énergétique. Le passage transitoire par les copies à haute température permet au système à basse température de franchir des barrières énergétiques qui seraient sinon réductrices. D'autres méthodes visent à échantillonner non pas des états ponctuels, mais des chemins de transition entre ces états (« transition path sampling »).

Les méthodes à base de variables collectives, quant à elles, supposent qu'il existe une ou plusieurs coordonnées de transition pour le phénomène d'intérêt ; ces coordonnées permettent de suivre l'évolution du système dans un espace de dimension réduite. Si l'on peut construire mathématiquement des variables décrivant bien ces coordonnées de transition, on peut appliquer des biais sur ces coordonnées et ainsi améliorer l'échantillonnage. Ces méthodes incluent l'échantillonnage parapluie (« umbrella sampling »), qui introduit une série de biais confinant le système dans des petites régions le long du chemin de transition, qui se trouve ainsi couvert de façon plus ou moins uniforme. Une autre approche consiste à contraindre le système en des points précis du chemin ; la force nécessaire à maintenir cette contrainte est liée à l'énergie libre de la transformation « blue moon ». Enfin, il existe une famille de méthodes dans lesquelles des forces externes poussent le système à évoluer dynamiquement le long du chemin. Dans tous les cas, le succès de ces approches dépend de la validité de l'hypothèse de départ, qui inclut aussi une séparation d'échelles de temps entre degrés de liberté lents – à savoir les coordonnées de transition – et rapides – c'est-à-dire tous les autres. Ainsi, on doit soit choisir d'emblée un petit nombre de coordonnées parfaitement pertinentes, soit inclure un grand nombre de coordonnées dans l'espace de transition. Seules quelques-unes de ces approches, dont la dynamique adiabatique et sa cousine « temperature-accelerated MD », sont utilisables sur un grand nombre de variables collectives. Par ailleurs, l'utilisation des variables les plus pertinentes dépend de la disponibilité d'implémentations efficaces des algorithmes souhaités, ainsi que des coordonnées adaptées. Les clés du problème sont donc souvent l'intuition physico-chimique d'un problème, qui permet de choisir les bons descripteurs de la transition, et parfois, la capacité de programmation de codes traitant efficacement le problème dans sa spécificité.

Conclusions : perspectives des simulations de dynamique moléculaire

Avec un regard introspectif de 55 ans depuis la simulation pionnière de Alder et Wainwright [11], il n'est pas illégitime de s'interroger sur le futur de la dynamique moléculaire, soit-elle classique ou quantique. Quelles nouvelles frontières peut-on raisonnablement espérer atteindre dans les années à venir ? Ce que nous enseignent les récentes simulations d'avant-garde, qu'elles soient sans biais ou qu'elles échantillonnent des degrés de liberté choisis, est que rien n'est jamais gravé dans le marbre. Ce qui est à la limite du possible aujourd'hui sera obsolète demain. L'avenir des simulations de dynamique moléculaire et plus généralement du calcul numérique est intimement lié à celui des architectures informatiques et aux défis technologiques que leur amélioration représente.

Regardons d'abord les avancées très importantes dans le domaine de la dynamique moléculaire classique. Il n'en reste pas moins qu'entre la première simulation en 1977 par McCammon, Gelin et Karplus [25] d'une protéine, l'inhibiteur pancréatique bovin de la trypsine – à une époque aujourd'hui révolue où les ateliers du Centre Européen de Calcul Atomique et Moléculaire (CECAM) d'Orsay se prolongeaient sur quelques mois, où l'on y programmait et exploitait ses codes [26] –, et les récentes simulations de la protéine NTL9 sur une poignée de millisecondes par Lindorff-Larsen, Piana, Dror et

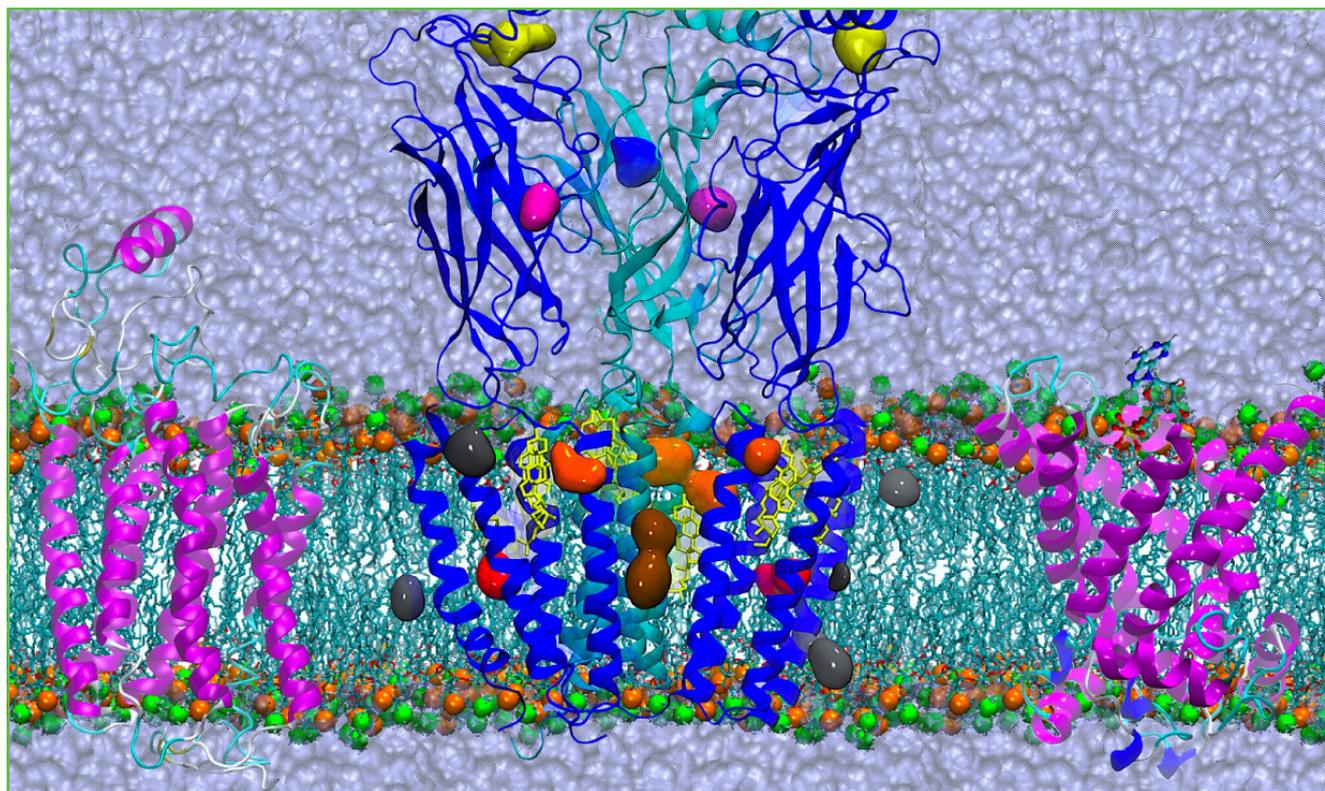


Figure 3 - Quelques exemples de protéines membranaires étudiées par simulations de dynamique moléculaire classique.

De gauche à droite : récepteur humain de la cholécystokinine – un récepteur couplé aux protéines G présent dans le système gastro-intestinal et le système nerveux central – (d'après [23]), récepteur nicotinique de l'acétylcholine auquel s'associe l'anesthésiant clinique isoflurane (d'après [24] ; les multiples sites de liaison sont montrés ici sous forme de surfaces colorées), et transporteur mitochondrial de l'adénosine di- et triphosphate qui assure le passage sélectif de nucléotides entre l'espace intermembranaire et la matrice de la mitochondrie. Ces trois protéines apparaissent au sein d'une membrane modèle de phospholipides.

Shaw [18], l'échelle de temps s'est vue dilatée d'un facteur 10^9 , et ce, en l'espace de moins de 35 ans. En dépit d'une durée de 8 picosecondes, qui peut faire sourire aujourd'hui, la première marqua un tournant dans le domaine de la biologie structurale informatisée en établissant le lien désormais indéfectible entre chimie informatique et théorique et biologie. La seconde démontre la faisabilité du repliement *in silico* jusque et y compris pour des structures tertiaires complexes, en en disséquant toutes les étapes. Artefact des champs de force ou réalité biologique, les simulations micro- et millisecondes suggèrent que les protéines hydrosolubles respirent, en se dépliant et se repliant à volonté. Si ce résultat, en soi, n'est pour l'heure pas révolutionnaire, il trace la voie vers la prédiction *ab initio* de la structure tridimensionnelle de protéines, voire d'assemblages protéiques – aujourd'hui encore assujettie aux imprécisions du champ de force, mais associée demain à une description plus rigoureuse des interactions interatomiques. Les 58 résidus de l'inhibiteur pancréatique bovin de la trypsine simulé en 1977 représentaient moins de mille atomes. Avec la première simulation d'un chromatophore, le groupe de Klaus Schulten a, pour sa part, moins de 35 ans plus tard, dilaté l'échelle de taille d'un facteur 10^5 . Là encore, les esprits chagrins ne manqueront pas de questionner, par delà la prouesse technologique sans précédent, la pertinence biologique d'un calcul limité à quelques nanosecondes d'un édifice aussi complexe – quelles leçons pouvons-nous tirer de fluctuations thermiques autour d'une structure en principe à l'équilibre ? Il n'est pas inintéressant cependant de noter qu'échelle de temps et échelle de taille, sur une même période, n'ont pas évolué de façon identique.

Même avec un nombre modeste de processeurs, il est envisageable depuis déjà quelques années, moyennant un assemblage moléculaire de petite taille, d'aborder l'échelle de la microseconde et, avec beaucoup de patience, de la milliseconde. Ce n'est en revanche qu'avec l'émergence des architectures massivement parallèles et, en particulier, petascale⁽¹⁾, comme Blue Waters, que peuvent être traités des édifices moléculaires de plusieurs millions d'atomes. Dynamique moléculaire à l'échelle de la milliseconde ou à l'échelle de la centaine de millions d'atomes, il s'agit encore à ce stade de défis technologiques, qui, néanmoins, suggèrent une tendance pour les années à venir. Alors qu'apparaissent les premiers calculateurs petascale, on spéculait déjà sur l'exascale⁽¹⁾ et ce que les nouvelles architectures seront capables de traiter. En supposant que dans les décennies à venir l'évolution de la dynamique moléculaire reste rythmée par la loi de Moore, Wilfred van Gunsteren [27] prédit avec optimisme qu'il sera possible de simuler à l'échelle de la nanoseconde, d'ici un peu plus de vingt ans, une bactérie complète, *Escherichia coli*, et une cellule de mammifère une vingtaine d'années plus tard. Quoique encourageantes, ces extrapolations mettent en lumière un fossé qu'il sera difficile de combler entre échelles de temps et de taille, et suggèrent qu'il reste encore un long chemin à parcourir avant que les simulations sans biais de macro-objets biologiques soient à même de fournir des informations chronologiquement pertinentes. À la dynamique moléculaire sans biais pourraient alors se substituer les approches d'échantillonnage préférentiel susceptibles de réconcilier temps et taille, dès lors que les degrés de liberté clés du processus biologique d'intérêt sont bien identifiés.

Enfin, l'accroissement massif des ressources informatiques prévu dans les années à venir est-il l'opportunité d'une remise à plat des champs de forces macromoléculaires minimalistes ? Peut-on envisager des fonctions d'énergie potentielle moins approximatives et donc, en principe, plus générales et plus fiables, quitte à revoir nos ambitions en matière d'échelles de temps et de taille ?

Enfin, en ce qui concerne la dynamique moléculaire quantique, si les avancées technologiques dans le domaine du traitement numérique de l'équation (2) nous ont permis de faire un grand saut au cours des trente dernières années, la taille des molécules pouvant être étudiées aujourd'hui avec des méthodes purement quantiques reste très modeste (par rapport au nombre d'atomes les composant). Même si l'avenir laisse entrevoir de nouvelles avancées technologiques avec de nouvelles générations d'ordinateurs, notamment les ordinateurs dits quantiques, cette puissance de calcul pourrait ne pas être suffisante pour traiter tous les degrés de liberté d'une molécule, même de taille modeste, à savoir une centaine d'atomes. On peut se demander s'il est vraiment nécessaire de traiter toutes les dimensions. Il n'y a pas de réponse générale à cette question. Les phénomènes quantiques sont liés au caractère ondulatoire de la fonction caractérisant l'état de la matière microscopique. La délocalisation spatiale de la fonction d'onde est en même temps le talon d'Achille de la méthode quantique et la source de nouvelles découvertes potentielles liées à notre compréhension de la réaction chimique et à l'exploitation des nouveaux matériaux qui en résulteraient. Une chose est sûre : les atomes et les molécules sont des objets quantiques. Une nouvelle théorie, qui permettrait à la fois de combler l'écart existant encore entre la dynamique moléculaire classique et quantique, et qui pourrait finalement nous expliquer le passage du monde microscopique au monde macroscopique, est peut-être nécessaire.

Note et références

- (1) Le *petascale* correspond à un quadrillion d'opérations (10^{15}) à virgule flottante par seconde. L'*exascale*, c'est mille fois plus, soit un quintillion (10^{18}).
- [1] Lasorne B., Gatti F., Baloitcha E., Meyer H.-D., Desouter-Lecomte M., Cumulative isomerization probability by various transition state wave packet methods including the MCTDH algorithm. Benchmark HCN-CN-H isomerization, *J. Chem. Phys.*, **2004**, *121*, p. 644.
- [2] Marquardt R., Quack M., Global analytical potential energy surfaces for high-resolution molecular spectroscopy and reaction dynamics, in *Handbook of High Resolution Spectroscopy*, M. Quack, F. Merkt (eds), John Wiley, **2011**, chap. 13, p. 511-550.
- [3] Gatti F., lung C., Exact and constrained kinetic energy operators for polyatomic molecules: The polyspherical approach, *Phys. Rep.*, **2009**, *484*, p. 1.
- [4] Ndong M., Joubert-Doriot L., Meyer H.-D., Nauts A., Gatti F., Lauvergnat D., Automatic computer procedure for generating exact and analytical kinetic energy operators based on the polyspherical approach, *J. Chem. Phys.*, **2012**, *136*, p. 034107.
- [5] Launay J., Molecular collision processes. I. Body-fixed theory of collisions between two systems with arbitrary angular momenta, *J. Phys. B*, **1977**, *10*, p. 3665.
- [6] Marquardt R., Sanrey M., Gatti F., Le Quééré F., Full dimensional quantum dynamics of vibrationally highly excited NHD_2 , *J. Chem. Phys.*, **2010**, *133*, p. 174302.
- [7] Lemoine D., Discrete cylindrical and spherical Bessel transforms in non-direct product representations, *Chem. Phys. Lett.*, **1994**, *224*, p. 483.
- [8] Meyer H.-D., Gatti F., Worth G.A., *Multidimensional Quantum Dynamics*, Wiley-VCH, **2009**.
- [9] Gindensperger E., Meier C., Beswick J.A., Mixing quantum and classical dynamics using Bohmian trajectories, *J. Chem. Phys.*, **2000**, *113*, p. 9369.
- [10] Cederbaum L.S., Gindensperger E., Burghardt I., Short-time dynamics through conical intersections in macrosystems, *Phys. Rev. Lett.*, **2005**, *94*, p. 113003.

- [11] Alder B.J., Wainwright T.E., Phase transition for a hard sphere system, *J. Chem. Phys.*, **1957**, *27*, p. 1208.
- [12] Verlet L., Computer experiments on classical fluids. I. Thermodynamical properties of Lennard-Jones molecules, *Phys. Rev.*, **1967**, *159*, p. 98.
- [13] Gumbart J., Trabuco L.G., Schreiner E., Villa E., Schulten K., Regulation of the protein-conducting channel by a round ribosome, *Structure*, **2009**, *17*, p. 1453.
- [14] Zhao G., Perilla J.R., Yufenyuy E.L., Meng X., Chen B., Ning J., Ahn J., Gronenborn A.M., Schulten K., Aiken C., Zhang P., Structure of the mature HIV-1 capsid by cryo-EM and all-atom molecular dynamics simulation, *Nature*, **2013**, *497*, p. 643.
- [15] Mei C., Sun Y., Zheng G., Bohm E.J., Kalé L.V., Phillips J.C., Harrison C., Enabling and scaling biomolecular simulations of 100 million atoms on petascale machines with a multicore-optimized message-driven runtime, *Proceedings of the 2011 ACM/IEEE Conference on Supercomputing*, **2011**, p. 12-18.
- [16] Bowman G.R., Voelz V.A., Pande V.S., Atomistic folding simulations of the five helix bundle protein λ_{6-85} , *J. Am. Chem. Soc.*, **2011**, *133*, p. 664.
- [17] Liu Y., Strümpfer J., Freddolino P.L., Grubele M., Schulten K., Structural characterization of λ -repressor folding from all-atom molecular dynamics simulations, *J. Phys. Chem. Lett.*, **2012**, *3*, p. 1117.
- [18] Lindorff-Larsen K., Piana S., Dror R.O., Shaw D.E., How fast-folding proteins fold, *Science*, **2011**, *334*, p. 517.
- [19] Dehez F., Pebay-Peyroula E., Chipot C., Binding of ADP in the mitochondrial ADP/ATP carrier is driven by an electrostatic funnel, *J. Am. Chem. Soc.*, **2008**, *130*, p. 12725.
- [20] Delemotte L., Tarek M., Klein M.L., Amaral C., Treptow W., Intermediate states of the Kv1.2 voltage sensor from atomistic molecular dynamics simulations, *Proc. Natl. Acad. Sci. USA*, **2011**, *108*, p. 6109.
- [21] Jensen M.Ø., Jogini V., Borhani D.W., Leffler A.E., Dror R.O., Shaw D.E., Mechanism of voltage gating in potassium channels, *Science*, **2012**, *336*, p. 229.
- [22] Nury H., Poitevin F., Renterghem C.V., Changeux J.P., Corringer P.J., Delarue M., Baaden M., One-microsecond molecular dynamics simulation of channel gating in a nicotinic receptor homologue, *Proc. Natl. Acad. Sci. USA*, **2010**, *107*, p. 6275.
- [23] Hémin J., Maigret B., Tarek M., Escriet C., Fourmy D., Chipot C., Probing a model of a GPCR/ligand complex in an explicit membrane environment: The human cholecystokinin-1 receptor, *Biophys. J.*, **2006**, *90*, p. 1232.
- [24] Brannigan G., LeBard D.N., Hémin J., Eckenhoff R., Klein M.L., Multiple binding sites for the general anesthetic isoflurane identified in the nicotinic acetylcholine receptor transmembrane domain, *Proc. Natl. Acad. Sci. USA*, **2010**, *107*, p. 14122.
- [25] McCammon J.A., Gelin B.R., Karplus M., Dynamics of folded proteins, *Nature*, **1977**, *267*, p. 585.
- [26] McCammon J., Molecular dynamics study of the bovine pancreatic trypsin inhibitor, *Report of the 1976 workshop Models for protein dynamics*, Centre Européen de Calcul Atomique et Moléculaire, H.J.C. Berendsen (ed.), Université de Paris IX, France, **1977**, p. 137-152.
- [27] van Gunsteren W.F., Bakowies D., Baron R., Chandrasekhar I., Christen M., Daura X., Gee P., Geerke D.P., Glättli A., Hünenberger P.H., Kastenholz M.A., Oostenbrink C., Schenk M., Trzesniak D., van der Vegt N.F.A., Yu H.B., Biomolecular modeling: goals, problems, perspectives, *Angew. Chem. Int. Ed.*, **2006**, *45*, p. 4064.



R. Marquardt



J. Hémin



F. Dehez



C. Chipot

Roberto Marquardt est professeur au Laboratoire de Chimie Quantique de l'Institut de Chimie de l'Université de Strasbourg¹.

Jérôme Hémin est chargé de recherche CNRS au Laboratoire de Biochimie Théorique de l'Institut de Biologie Physico-Chimique².

François Dehez est chargé de recherche CNRS et **Christophe Chipot**, directeur de recherche, au Laboratoire International Associé CNRS-UIUC³.

¹ Laboratoire de Chimie Quantique, Institut de Chimie, UMR 7177 CNRS/Université de Strasbourg, 1 rue Blaise Pascal, BP 296/R8, F-67008 Strasbourg Cedex.

² Courriel : roberto.marquardt@unistra.fr
Laboratoire de Biochimie Théorique, UPR 9080 CNRS, Institut de Biologie Physico-Chimique, 13 rue Pierre et Marie Curie, F-75005 Paris.

³ Courriel : jerome.hemin@ibpc.fr
Laboratoire International Associé CNRS-University of Illinois, Urbana-Champaign, UMR 7565, Structure et Réactivité des Systèmes Moléculaires Complexes, Institut Jean Barriol, Université de Lorraine, BP 70239, F-54506 Vandœuvre-lès-Nancy Cedex.
Courriels : francois.dehez@univ-lorraine.fr ; christophe.chipot@univ-lorraine.fr