

Réduction du temps d'obtention du résultat grâce à la visualisation *in situ* et au *cloud computing* pour les dynamiques moléculaires

Reduction of the time to resolved through in situ visualization and cloud computing in the case of molecular dynamics

Eddy Namur, Cédric Gageat, Wilfried Kirschenmann, Anne-Claude Camproux, Gautier Moroy

English Abstract—The development of a new drug is a long and costly process. Accelerating this process and reducing costs is therefore a major challenge. In this article, we will show how the combination of remote visualization and cloud computing allows us to overcome these limitations, significantly reduce the time it takes to acquire results while improving the user experience. Molecular dynamics simulations generally require several months of computation. The architecture of current supercomputers, whether private or public, does not allow for easy verification of the progress of simulations. It is even common to detect simulation anomalies only at the end of the simulation, which requires a complete restart of the calculation. In addition, the security constraints coupled with the schedulers' limits of these datacenters generally require that the data be repatriated and the calculations restarted daily and manually. The objective of this project was to set up a cloud infrastructure for remote and real-time simulation and visualization of molecular dynamics simulation results. Indeed, the real-time visualization and analysis of customized metrics (3D conformation of the molecules, energy values, dihedral angle values, RMSD,...) allows early detection of errors and speeds up decision making. In this way, it is no longer necessary to wait until the end of the simulation, which reduces calculation times and the cost of such simulations. In order to improve the user experience and facilitate visualization, we have chosen to make all these elements available from any terminal (PC, tablet, smartphone,...) through a simple web browser.



1 INTRODUCTION

Le développement d'un nouveau médicament est un processus long et coûteux. Il faut attendre plus de 12 ans entre la détermination de la cible thérapeutique et la mise sur le marché d'une nouvelle molécule pour un coût de recherche supérieur à 2 milliards d'euros. L'accélération de ce processus et la réduction du coût est donc un enjeu majeur. Malgré l'utilisation massive de simulations numériques, peu coûteuses et rapides, de nombreuses limites existent encore. Dans cet article, nous montrerons comment la combinaison de la visualisation à distance et du cloud computing nous permette de dépasser ces limites, de fortement réduire le temps d'acquisition des résultats tout en améliorant l'expérience utilisateur.

L'une des premières étapes de la création d'un nouveau médicament consiste à mieux connaître une des molécules intervenant dans le cycle de la maladie. Cette molécule est appelée cible thérapeutique. Une

des simulations les plus coûteuse consiste à étudier le mouvement en 3D de la cible thérapeutique à l'aide de dynamiques moléculaires. Ce type de simulations nécessitent généralement plusieurs mois de calculs. L'architecture des supercalculateurs actuels, qu'ils soient privés ou publiques ne permet pas de vérifier facilement l'avancement des simulations. Il arrive même fréquemment de détecter les anomalies de simulation uniquement en fin de simulation, ce qui nécessite de relancer entièrement le calcul. De plus, les contraintes de sécurité couplées aux limites des schedulers de ces datacenter imposent généralement de devoir rapatrier les données et de relancer quotidiennement et manuellement les calculs (voir figure 2).

L'objectif de ce projet a été de mettre en place une infrastructure cloud de simulation et de visualisation à distance et en temps réel de résultats de simulations de dynamique moléculaire. En effet, la visualisation et l'analyse en temps réel de métriques personnalisées (conformation 3D de la molécules, valeurs d'énergies, valeurs d'angles diédre, RMSD, ...) permettent une détection précoce des erreurs et accélère d'autant la prise de décision. De cette façon, il n'est plus nécessaire d'attendre la fin de la simulation, ce qui permet de réduire les temps de calculs ainsi que le coût de telles simulations. Afin d'améliorer l'expérience utilisateurs et de faciliter la visualisation, nous avons fait le choix de rendre tous ces éléments disponibles depuis n'importe quel terminal (PC, tablette, smart-

- Eddy Namur: Aldwin / Université Paris Diderot
E-mail: enamur@aneofr
- Cédric Gageat: Aldwin (by ANEO)
E-mail: cgageat@aneofr
- Wilfried Kirschenmann: Aldwin (by ANEO)
E-mail: wkirschenmann@aneofr
- Anne-Claude Camproux: INSERM U1133
E-mail:anne-claude.camproux@univ-paris-diderot.fr
- Gautier Moroy: INSERM U1133
E-mail: gautier.moroy@univ-paris-diderot.fr

phone, ...) au travers d'un simple navigateur web. Dans la suite de cet article, nous présenterons les simulations de dynamique moléculaire, ainsi que les contraintes et limites associés. Nous montrerons ensuite comment le cloud permet de lever certaines d'entre elles. Nous finirons enfin, par le cas pratique de l'étude d'une molécule intervenant dans l'autisme avant de conclure.

2 DYNAMIQUE MOLÉCULAIRE

Les dynamiques moléculaires permettent, à partir de la séquence d'une molécule, d'obtenir une modélisation en 3D de sa structure. Le plus souvent, à cause des conditions d'expériences, les simulations s'avèrent plus précises que les analyses de structures expérimentales. Afin d'explorer de sortir des minimum locaux d'énergie et ainsi d'explorer des espaces chimiques plus larges, nous utilisons des simulations avec échanges de répliques. Ce type de simulation nécessite de réaliser plusieurs dynamiques moléculaires en parallèle. Chacune d'entre elle correspond à une température différentes. A interval de temps régulier et en fonction de critères d'énergies, les répliques voisines sont échangées. La multiplication des répliques, augmente d'autant le temps de calcul nécessaire. Dans ces conditions, il n'est pas rare que l'obtention d'une microseconde seulement de simulation nécessite plusieurs mois de calculs. Des outils d'analyses, tels que la visualisation à distance que nous avons développé facilite fortement la détection d'anomalies.

3 CLOUD COMPUTING

Malgré les moyens financiers et informatiques dont disposent les industries pharmaceutiques, il n'est pas rare de devoir attendre avant qu'un calcul s'effectue. Dans le cas de la recherche publique, les moyens étant fortement inférieurs, les temps d'attentes sont largement supérieurs. Afin de réduire les coûts tout en accélérant les calculs, le Dr Shaw a créé des puces de calcul spécialement adaptées aux dynamiques moléculaires. Il existe heureusement d'autres alternatives, comme l'usage du cloud. En effet, le cloud permet de changer de paradigme et de construire de façon transparente et automatique un cluster dédié pour chaque calcul. L'accès de ressources élastiques et virtuellement illimitées permet de réduire le temps et le coûts de ces calculs en s'adaptant au mieux à chaque cas.

De plus, dans notre cas, le cloud nous permet d'ouvrir facilement et en toute sécurité les flux nécessaires au streaming des métriques d'intérêts.

4 CAS D'APPLICATION

Ces développements ont ensuite été appliqué au peptide PIF (PreImplantation Factor) impliqué dans

l'autisme. PIF est un peptide naturel qui possède des propriétés importantes de neuroprotection. Ses mécanismes d'action ainsi que la structure 3D qu'il peut adopter sont encore mal connus. Ce peptide de 15 résidus, dont nous n'avons pas de structure expérimentale a été construit à partir de sa séquence. La taille de cette molécule nécessite l'exécution d'au moins 47 simulations parallèles, ce qui représente environ 250 000 heures de calculs pour seulement 1 microseconde de simulation.

L'infrastructure que nous avons développée, nous permet de détecter facilement les événements majeurs tels que l'évolution importante de l'énergie potentielle qui pourrait indiquer que la molécule a changée de conformation (elle s'est ouverte ou repliée) ou alors que la simulation est en train de dériver. Comme nous le souhaitons, la détection précoce de ces événements nous permet d'avoir une meilleure réactivité et d'économiser en ressources et en temps de calculs.

5 CONCLUSION

Comme nous l'avons montré, la recherche d'un nouveau médicament assisté par ordinateur reste un processus long et coûteux. La résilience, la flexibilité ainsi que l'élasticité du cloud nous ont permis de développer et déployer un outils de visualisation en direct de simulation de dynamiques moléculaires. Ces développements ont ensuite facilité l'analyse du peptide PIF, intervenant dans l'autisme. Pour aller plus loin, plusieurs étapes sont déjà en cours de réflexion. Actuellement, la visualisation se faisant côté client, ne peut pas nécessiter trop de ressources. Dans le cas de visualisation complexes, il serait facile de déporter la modélisation sur un serveur de rendu GPU que l'on placerait entre la simulation et le terminal de visualisation.

6 CONTRIBUTION

Les heures de calculs cloud utilisées dans ce projet ont été financées par la plateforme cloud Amazon Web Services.

7 BIBLIOGRAPHIE

1 Abraham, M. J., Murtola, T., Schulz, R., Páll, S., Smith, J. C., Hess, B., Lindahl, E. (2015). GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX*, 1-2, 19–25.

2 Sugita, Y., Okamoto, Y. (1999). Replica-exchange molecular dynamics method for protein folding. *Chemical Physics Letters*, 314(1-2), 141–151.

3 Mueller, M., Zhou, J., Yang, L., Gao, Y., Wu, F., Schoeberlein, A., ... Huang, Y. (2014). PreImplantation factor promotes neuroprotection by targeting microRNA let-7. *Proceedings of the National Academy of Sciences*, 111(38), 13882–13887.

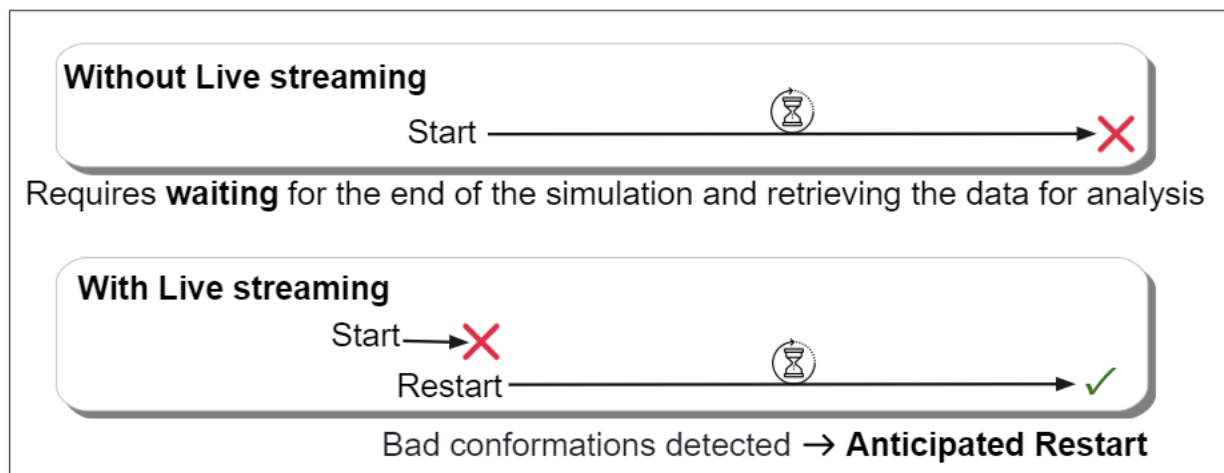
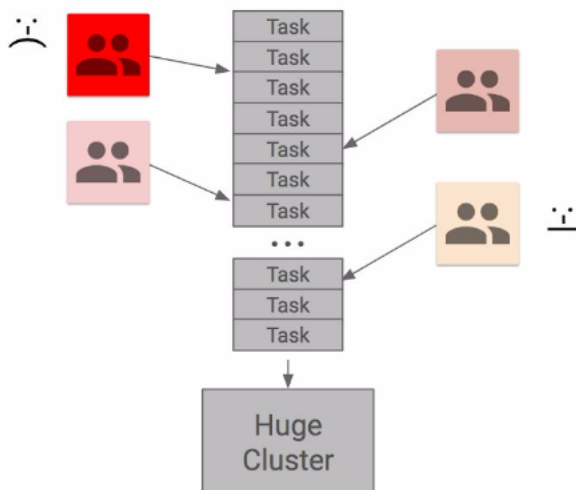


Fig. 1. Illustration d'un exemple de simulation de dynamique moléculaire avec ou sans la visualisation en temps réel.

On-prem: One cluster, one queue, one size fits all hardware, angry users.



In the cloud: Tailored clusters, less queue sharing, happy users.

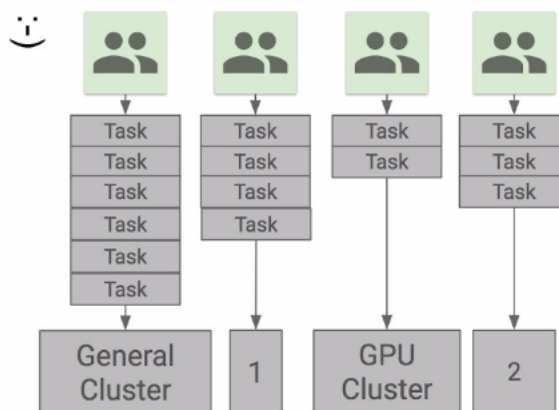


Fig. 2. Illustration de la différence entre le temps d'attente pour soumettre des jobs dans un cluster sur site et dans le cloud.