

# Proposition de sujet de thèse

**Acronyme :** REVIDYM

**Titre du projet de thèse :** Apport de la réalité virtuelle immersif à la modélisation interactive de simulations moléculaires dynamiques.

**Mots clés :** réalité virtuelle, interaction multimodale, haptique, algorithmie, environnement collaboratif, docking moléculaire, plasticité dynamique moléculaire, bioinformatique, données massives.

## Résumé :

Le développement constant d'environnement de visualisation immersive couplé à des dispositifs à retour d'effort (bras haptiques) est crucial pour la manipulation et l'exploration intuitive d'objets biologiques complexes et pour l'analyse de données massives représentant la simulation de leurs dynamiques. Le projet REVIDYM sera réalisé sur la plate-forme immersive SHIVA de l'ENS Cachan. Il vise à l'adaptation et l'évolution des outils de visualisation et d'interaction 3D à des outils d'analyse des structures 3D et leurs dynamiques. REVIDYM bénéficiera du contexte de recherche sur les sites allostériques des protéines et de leur exploration en vue de développer de nouvelles molécules médicamenteuses. L'avancement et l'extension de ces innovations techniques de réalité virtuelle basée sur un environnement de visualisation immersif et couplé à des outils de calculs plus fins sera accessible et bénéfique aux étudiants et chercheurs de différents profils (mathématiciens, physiciens, cliniciens, biologistes, chimistes, etc.).

**Thématique du projet :** Il s'inscrit dans le cadre de l'École Doctorale EDMH – Paris Saclay (École Doctorale de Mathématique Hadamard).

**Établissement:** ENS Cachan Université Paris-Saclay

**Laboratoire ou Unité de Recherche :** CMLA Centre de Mathématique et de Leurs Applications (UMR CNRS 8536) & Institut FARMAN (Fédération de Recherche CNRS, FR 3311).

## Encadrement :

- Luba TCHERTANOV : Directrice de recherche CNRS – Laboratoire CMLA
- Florian DE VUYST : Professeurs des Universités ENS Cachan - Laboratoire CMLA
- Atman KENDIRA : Ingénieur de recherche ENS Cachan – Laboratoire CMLA

**PROJET REVIDYM :** Description détaillée des travaux

### – *Introduction de la problématique scientifique et contexte de l'étude*

En biologie, il est un vieux problème primordiale qui n'a toujours pas trouvé de solution optimale. Ce problème fondamental, consiste à comprendre comment des molécules (protéines, ligands ou ADN/ARN) se reconnaissent pour opérer des phénomènes physiologiques. Le phénomène de la reconnaissance (ou *docking*) contrôle presque tout les processus biologiques, e.g., la signalisation cellulaire, l'activation de métabolisme ou le contrôle de transcription. Les données expérimentales d'écrivain de telles phénomènes au niveau atomique restent encore très limités. L'avènement de l'informatique a permis à de nombreux chercheurs de proposer des algorithmes, de plus en plus optimisés mais sans jamais parvenir à obtenir une solution robuste et fiable dans un temps raisonnable.

Actuellement, les méthodes de docking combinent des phases automatiques prenant en compte les paramètres topologiques ainsi qu'énergétiques (physico-chimiques) des protéines, et des phases de visualisation moléculaire permettant d'évaluer les résultats obtenus. La phase automatique est coûteuse en temps de calcul, et fournit un grand nombre de solutions de docking potentielles difficilement discriminables juste sur la base des paramètres évalués (scores). C'est pourquoi une phase d'analyse visuelle manuelle par des experts est nécessaire afin d'obtenir un résultat pertinent. Néanmoins cette analyse est très lourde en terme de quantité d'information qu'il faut prendre en compte simultanément (manipulation d'objet 3D, caractéristiques physico-chimiques, données biologiques, flexibilité, plasticité, dynamique, etc.). Pour répondre à ces problématiques de temps, de robustesse et de fiabilité, les biologistes ont associé l'intelligence de l'humain à la puissance de traitement des algorithmes au sein du processus de recherche. C'est avec la réalité virtuelle que les biologistes accèdent à la manipulation des molécules numériques avec l'aide de logiciels de visualisation, de moteurs de simulations ou encore de périphériques d'interactions tangibles et sensoriels.

– **Positionnement par rapport à l'état de l'art et à la concurrence nationale et internationale**

L'idée d'immerger un humain au sein du processus de docking moléculaire date de 1967 avec le projet GROPE [1]. L'intervention d'un expert durant le processus de recherche est effectuée par l'immersion dans des environnements de réalité virtuelle. Batter [2] proposent les premières solutions d'immersion visuelle dans le cadre GROPE. Avec l'immersion dans un environnement virtuel, le problème de la représentation des molécules se pose rapidement. Bergman [3] proposent la plateforme VIEW avec différentes possibilités de rendus graphiques pour représenter des molécules. Les environnements virtuels offrent une certaine souplesse dans la représentation des molécules permettant de les afficher en fonction des informations pertinentes à mettre en évidence (cf. Les logiciels VMD [4] et PYMOL [5]).

Cependant, le mode statique des visualisations ne permet pas de comprendre la dynamique moléculaire. Huitema [6] propose alors un outil pour visualiser des trajectoires dynamiques des atomes. Puis Klosowski [7] franchissent une étape supplémentaire en proposant une visualisation temps-réel de la dynamique d'une molécule à l'aide du moteur GROMACS [8]. Enfin, Křenek [9] apporte une visualisation multimodale avec la combinaison des retours visuels et haptiques pour percevoir les champs de force d'une molécule. Les biologistes ont éprouvé le besoin d'interagir avec les champs de force en action au sein d'une molécule. Le projet GROPEHaptic issue du projet GROPE, propose d'utiliser la modalité haptique pour interagir avec les molécules [10]. Des plateformes immersives modulables pour la modélisation moléculaire sont apparues telle que HIMM (*Highly Immersive Molecular Modeling*) [11] ; cette plateforme permet entre autre de connecter des interfaces haptiques. En tenant compte de la flexibilité des molécules, les besoins en interaction se sont affinés en cherchant à modifier la position des atomes dans la molécule. L'interaction proposé est la traction des atomes par un simple modèle masse-ressort [12] ce qui permet une déformation des atomes de manière locale. Puis Lee [13] proposent de baser les modèles d'interactions haptiques sur l'approximation du champ de force pour une bonne stabilité du retour haptique.

Ces différents développements ont ensuite permis aux chercheurs de revenir sur leur première problématique du docking moléculaire dont le projet STALK [14] résulte. La communauté 'haptique' commence également à s'intéresser à cette problématique [15] proposant par exemple une interface haptique avec cinq/six DDL afin de manipuler et d'interagir en temps-réel avec des molécules durant un processus de docking moléculaire [16, 17, 18]. Plus récemment, deux approches différentes ont permis d'aboutir à des plateformes de manipulation haptiques de molécules avec un moteur de simulation en temps-réel. D'un côté, Redon [19] développe des algorithmes pour la simulation en temps-réel de corps articulés avec un grand nombre de DDL. À partir de ce travail, il crée un moteur de simulation moléculaire permettant d'obtenir des simulations temps-réel [20], utilisable pour la manipulation haptique [21]. En parallèle, Delalande [22] propose un outil permettant d'obtenir des simulations moléculaires en temps-réel et utilise cette plateforme pour la manipulation haptique temps-réel de molécules déformables au niveau atomique. En fournissant tous ces outils de visualisation et d'interaction avec des environnements moléculaires virtuels, les biologistes peuvent à présent interagir en temps-réel avec des molécules [24]. Cependant, la complexité du problème de docking moléculaire est suffisamment importante pour nécessiter la présence de plusieurs collaborateurs au sein de l'environnement de travail ; on parle de travail collaboratif en présentiel ou à distance.

– **Objectifs scientifiques, défis et enjeux, résultats attendus et critères proposés pour juger du succès du projet**

Pour effectuer ces fonctions (activation/désactivation, association/dissociation, ouverture/fermeture de canaux ionique), les molécules biologiques transitent d'un état vers un autre grâce à leurs plasticités conformationnelles induites par des effecteurs de différents types (ADN, protéines, inhibiteurs). Ces effecteurs stimulent le changement d'un état initial de la protéine (non-lié) à l'état fonctionnel (lié avec un effecteur), ou la stabilisation d'une protéine dans une conformation figée évitant des fonctions de protéine non désirables (**Fig. 1**). De nouvelles méthodes de manipulation (multimodales) des molécules biologiques permettant de simuler leurs assemblages/désassemblages et les modulations de leurs structures (voir fonctions) par les petites molécules inhibitrices/modulatrices sont donc à envisager.

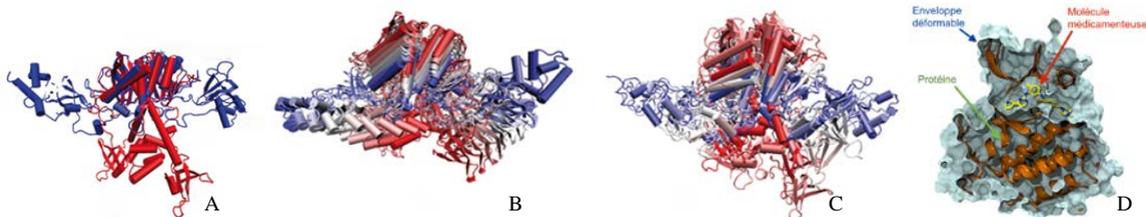


Figure 1 : (A-C) Plasticité conformationnelle des protéines illustrée par l'évolution de la conformation dynamique de l'intégrase du VIH—1. Passage d'un état initial (en rouge) à l'état final (en bleu) sous l'influence de l'ADN virale. (D) Conformation figée d'un récepteur de tyrosine kinase (KIT) stabilisée par la molécule inhibitrice (imatinib).

Malgré de nombreuses études, les algorithmes de docking entre des protéines, restent peu efficaces, à la différence des méthodes de docking entre de petits ligands et leurs récepteurs, dont les algorithmes sont désormais performants à la fois du point de vue du temps de calcul et de la fonction mesurant la qualité des complexes. Nous proposons dans le présent projet de **développer un outil** logiciel de réalité virtuelle et collaboratifs entre les experts, permettant (1) d'identifier les modes d'assemblage les plus pertinents entre

plusieurs entités protéiques, (2) de les manipuler interactivement et d'analyser les interfaces protéine-protéine ainsi créées, (3) de procéder à des calculs élaborés de dynamique moléculaire pour les solutions les plus pertinentes. Dans ce contexte et pour atteindre des fréquences de rafraîchissement nécessaires à certains canaux sensori-moteurs durant le rendu haptique (>1KHz), il faudra implémenter de nouvelles approches interactives à l'aide de processeurs dédiés (cluster CPU/GPU, grilles de calcul). Le projet va impliquer le **developpement de logiciel(s) haptique le(s) plus avancé(s)** pour le docking. Ce logiciel sera destiné à ceux qui travaillent dans les industries pharmaceutiques pour la conception de médicaments assistée par ordinateur.

– *Méthodes, concepts, moyens, outils, pistes bibliographiques*

Le projet évoluera au sein de la plateforme de réalité virtuelle de l'École Normale Supérieure de Cachan nommée SHIVA, (SHared Interaction & Visualization Area) [25]. SHIVA, est l'une des composantes du projet DIGISCOPE (EquipEx ANR-EQPX-10-26-1). DIGISCOPE est un dispositif de moyens de visualisation et d'interaction installés sur plusieurs sites de l'Université Paris-Saclay : CEA, École Centrale Paris, ENS Cachan, PCRI/INRIA, Maison de la Simulation, Télécom Paristech, Université Paris-Sud Orsay, Université de Versailles Saint-Quentin. DIGISCOPE a pour ambition de créer un réseau collaboratif d'équipements de visualisation volontairement hétérogènes. Chaque équipement a sa spécificité : CAVE, grande surface d'affichage, grande résolution, interaction tactile/3D/tangible/haptique, son spatialisé. Il est envisagé aussi l'installation de moyens de téléprésence donnant ainsi l'impression aux utilisateurs/experts distants d'être réunis dans une même salle de travail. DIGISCOPE devrait être totalement déployé d'ici la fin 2015 et constituera une infrastructure étendue de visualisation unique au monde dont ce projet bénéficiera.

Le projet de thèse s'appuiera sur nos recherches précédentes dans ce domaine et qui a donné lieu à la conception de nouvelles méthodes réalisé sous forme d'un package logiciel nommé MONETA (MODular NETwork Analysis) [26, 27] basé sur les outils permettent à l'utilisateur d'explorer une simulation dynamique des molécules pour localiser et 'visualiser' en 2D et 3D la transmission d'informations à travers le réseau de résidus d'une protéine et/ou des complexes moléculaires [28-30]. L'originalité de MONETA réside dans la prise en compte de la description duale des protéines – changement de conformation global et modification des fluctuations atomiques locales.

Nous sommes motivés par le developpement de nouvelles méthodes permettant de simuler les interactions des objets 'dynamique' ayant une plasticité conformationnelle. Le développement/évolution de nouveaux modèles d'interaction tangible associé à des interfaces sensoriels - à retour d'efforts sera très avantageux pour la simulation/description des processus complexes. Un des premiers pas dans cette direction déjà fait au CMLA (**Fig. 2**).

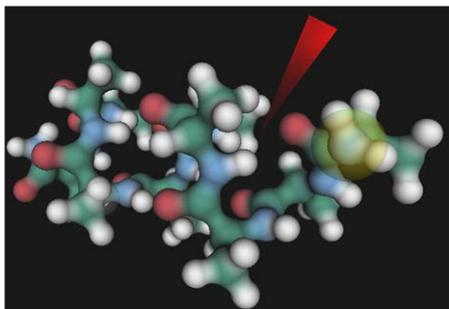


Fig 2. Interactive 'touch' d'une molécule via le bras haptique.

Le développement de nouveaux outils/logiciels permettant le docking des objets biologiques en utilisant des approches immersives et interactives (visualisation, immersion, interaction collaboratif ou multi-dispositif) sera également un élément crucial dans l'évolution de SHIVA et son articulation avec la recherche dans différents domaines ainsi que la formation des étudiants de profils pluridisciplinaires.

– *Compétences requises*

La thèse concerne le développement de nouvelles interfaces de réalité virtuelle pour la bioinformatique. Le candidat pour ce projet devra posséder des pré-requis sur la connaissance et l'implémentation d'algorithmes pour l'interaction multimodale en « réalité virtuelle ». Des notions en bioinformatique et/ou biologie structurale ou computationnelle serai préférables.

– *Environnement de travail : équipes concernées, partenariat et travaux antérieurs des proposant*

Le projet s'inscrit au sein de l'institut FARMAN de l'ENS Cachan (Fédération de Recherche CNRS FR 3311), institut Fédératif regroupant cinq laboratoires de l'École autour de la modélisation, la simulation numérique et l'Ingénierie des systèmes complexes (CMLA UMR 8536, LMT Cachan UMR 8535, LSV UMR 8643, LURPA EA 1385 et SATIE UMR 8029). Au sein du CMLA, le projet sera supporté par l'équipe BiMoDyM (Bioinformatique, Modélisation et Dynamique Moléculaire) dirigée par L. Tchertanov et l'équipe SFR (Simulation des Fluides Réels) dirigée par F. De Vuyst. Il pourra s'appuyer sur **les projets antérieurs** (OSEO-CAP, Fondation de France, La Ligue contre le Cancer, FARMAN, Société Française d'Hématologie), **les projets en cours** du CMLA tel que l'Equipex DIGISCOPE, le projet NUTS (ANR), le labex LERMIT, le projet ICAR (FARMAN) et le **partenariat industriel** avec la société LIPHATECH. Les collaborations constantes (2009-présent) avec des cliniciens (Institut Gustave Roussy, Hopital Necker, Centre de Recherche en Cancérologie de Marseille, etc.) permettent d'appliquer les nouvelles méthodologies aux cibles d'intérêt thérapeutique majeures et de valider nos developpements. Enfin, d'autres collaborations nationales et internationales seront envisagées, avec le Laboratoire de Biochimie Théorique (LBT UPR 9080) [Marc Baaden], le Laboratoire d'Informatique pour la Mécanique et les Sciences de l'Ingénieur (LIMSI UPR 3251)

[Nicolas Férey & Mehdi. Ammi], Lehrstuhl für Theoretische Biophysik-Molekulardynamik (TUM) [Martin Zacharias] et Molecular Graphics Laboratory (Scripps) [Arthur J. Olson].

– **Objectifs et intérêt pour Digiteo et/ou DigiCosme.**

Cette thèse correspondent aux thèmes de recherche « *Interaction, visualisation et réalité virtuelle* », « *Bio-informatique* » et « *Modélisation et simulation* » de DIGITEO et s'inscrit dans le développement de l'Équipement d'Excellence DIGISCOPE. Il s'agit en effet de développer de nouvelles méthodes interactives et collaboratives pour simuler le docking moléculaire dynamique, qui font appel à différents domaines métiers et interagissent avec un environnement physique, afin de les simuler et de valider les interactions entre des composantes hétérogènes. C'est également un support de valorisation et d'évolution du projet SHIVA et de son environnement à travers l'Equipe DIGISCOPE.

– **Bibliographie**

- [1] Grünwald (2008). Human haptic perception : basics and applications. Springer. isbn : 978-3764376116.
- [2] Batter et al.(1972). GROPE-1: a computer display to the sense of feel. *Int. Fed. Information Processing*. North-Holland, pp 759–763.
- [3] Bergman et al. (1993). VIEW - an exploratory molecular visualization system with user-definable interaction sequences. *Conference on Computer graphics and interactive techniques*. SIGGRAPH. Anaheim, USA : ACM, pp 117–126.
- [4] Humphrey et al. (1996). vmd: Visual Molecular Dynamics. *J. Mol. Graphics* 14.1, pp 33–38. <http://www.ks.uiuc.edu/Research/vmdl/>
- [5] DeLano et al. (2002). The PyMOL molecular graphics system. <http://www.pymol.org>
- [6] Huitema et al. (oct. 2000). Interactive visualization of protein dynamics. *Proceedings of the 11th IEEE Visualization Conference*. VIS. Salt Lake City, USA : IEEE Computer Society, pp 465–468.
- [7] Klosowski et al. (2002). Deep View : high-resolution reality. *IEEE Computer Graphics and Applications* 22.3, pp 12–15.
- [8] Hess, B, C Kutzner, D. van der Spoel et E. Lindahl (2008). GROMACS 4: algorithms for highly efficient, load-balanced, and scalable molecular simulation. *Journal chemical theory and computation*, pp 435–447.
- [9] Křenek et al. (1999). Haptic visualization of molecular model. *Int. Conf. Comp. Graph., Visualization & Computer Vision*, pp 133–139.
- [10] Brooks Jr. et al. (1990). Project gropeHaptic displays for scientific visualization. *Proceedings of the 17th annual conference on Computer graphics and interactive techniques*. Dallas, Texas, États-Unis : ACM, pp 177–185.
- [11] Drees et al. (1996). Highly Immersive Molecular Modeling (HIMM): an architecture for the integration of molecular modeling and virtual reality. *German Conference on Bioinformatics*, pp 190–192.
- [12] Haan et al. (2002). Towards intuitive exploration tools for data visualization in VR. *Proceedings of the ACM symposium on Virtual reality software and technology*. VRST. Hong-Kong, China: ACM, pp 105–112.
- [13] Lee et al. (2004). Smoothing haptic interaction using molecular force calculations. *CAD* 36.1, pp 75–90.
- [14] Levine et al. (1997). STALK : an interactive system for virtual molecular docking. *Computer in Sciences and Engineering*, pp 55–65.
- [15] Subaşı et al. (2006). A new approach to molecular docking in virtual environments with haptic feedback. *Eurohaptics*, pp 141–145.
- [16] Lai-Yuen et al. (2005). Computer-aided molecular design (CAMD) with force-torque feedback. *Proceedings of the 9th International Conference on CAD&CG*. CAD-CG. Washington, District of Columbia, États-Unis : IEEE Computer Society, pp 199–204.
- [17] Daunay et al. (2007). DoF haptic feedback for molecular docking using wave variables. *IEEE International Conference on Robotics and Automation*. ICRA, pp 840–845.
- [18] Hou et al. (2010). Haptic rendering algorithm for biomolecular docking with torque force. *Proceedings of the 2010 International Conference on Cyberworlds*. CW. Washington, USA: IEEE Computer Society, pp 25–31.
- [19] Redon et al. (2005). Adaptive dynamics of articulated bodies. *SIGGRAPH* 24.3, pp 936–945.
- [20] Rossi et al. (2007). Adaptive torsion-angle quasistatics: a general simulation method with applications to protein structure analysis and design. *Bioinformatics* 23.13, pp 408–417.
- [21] Bolopion et al. (2009). Haptic feedback for molecular simulation. *Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems*. Saint-Louis, Missouri, USA, pp 237–242.
- [22] Delalande et al. (2009). Complex molecular assemblies at hand via interactive simulations. *J. Comp. Chem.*, pp 2375–2387.
- [23] Delalande et al. Multi-resolution approach for interactively locating functionally linked ion binding sites by steering small molecules into electrostatic potential maps using a haptic device. *Pacific Symposium on Biocomputing*, pp 205–215.
- [24] Stone et al. (2010). Immersive molecular visualization and interactive modeling with commodity hardware. In *Advances in Visual Computing*. T. 6454. Lecture Notes in Computer Science. Urbana-Champaign, États-Unis : Springer Berlin / Heidelberg, pp. 382–393.
- [25] De Vuyst F et al. (2014). L'ENS Cachan voit grand avec le mur d'image SHIVA. *HPC Today*. 26 sept. 2014.
- [26] Allain et al. (2014). Allosteric Pathway Identification through Network Analysis from Molecular Dynamics Simulations to Interactive 2D and 3D Graphs. *Faraday Disc.*, DOI: 10.1039/C4FD00024B. 169, 303-321
- [27] Tchertanov, Certificat délivré par Agence pour la Protection des Programmes. Inter Deposit Digital Number IDD.N.FR.001.020012.000.S.P.2014.000.31235 pour l'œuvre: MOdular NETwork Analysis (MONETA) version 2.0.
- [28] Laine et al. (2012). Allosteric Communication across the Native and Mutated KIT Receptor Tyrosine Kinase. *PLoS Comput Biol*. 8(8):e1002661, doi:10.1371/journal.pcbi.1002661.
- [29] Da Silva Figueiredo Celestino Gomes et al. (2014). Differential effects of CSF-1R D802V and KIT D816V homologous mutations on receptor tertiary structure and allosteric communication. *PLOS ONE* ; doi: 10.1371/journal.pone. 0097519.
- [30] Chauvot de Beauchêne et al. (2014). Oncogenic mutations of KIT receptor differentially modulate tyrosine kinase activity and drug susceptibility. *PLOS Comput. Biol.*10(7):e1003749. doi: 10.1371/journal.pcbi.1003749.