

**Demande de renouvellement du
Projet FARMAN inter labos (BiMoDyM, LBPA – CMLA)**

TOPDYN

**TRANSITION CONFORMATIONNELLE DES PROTEINES :
TOPOLOGIE « DYNAMIQUE »**

DUREE DU PROJET: 2 ans

RESPONSABLES ET MEMBRES SCIENTIFIQUES :

Luba TCHERTANOV (BiMoDyM, LBPA) et Alain TROUVE (CMLA)

INTRODUCTION ET MOTIVATIONS :

Les méthodes de **bioinformatique et modélisation moléculaire** – discipline essentielle de biologie computationnelle et structurale – offrent des moyens de caractériser les propriétés dynamiques et énergétiques des protéines et permettent d'accéder à des échelles de temps et d'espace difficiles à sonder expérimentalement. Combinées aux données expérimentales, elles contribuent à une meilleure compréhension des mécanismes moléculaires des processus physiologiques dans la cellule. Elles sont indispensables pour interpréter des données expérimentales et cliniques, formuler et examiner des hypothèses, stimuler développement de nouvelles approches, et jouent un rôle important dans la recherche de nouveaux médicaments. Ce travail concerne en premier lieu les biologistes et médecins.

Le **développement de ces méthodes** est en plein essor depuis quelques années. La puissance des ordinateurs permet désormais d'envisager de modéliser des situations de plus en plus complexes ; ce qui engendre un très grand nombre de données qu'il est nécessaire d'analyser et d'interpréter. Cependant, le besoin d'outils mathématiques nouveaux et originaux se fait sentir. Nous proposons dans ce projet d'explorer des pistes de développement de tels outils en s'appuyant d'une part sur l'expérience en géométrie et statistique computationnelles du CMLA et d'autre part en utilisant les données et l'expertise de l'équipe BiMoDyM.

DESCRIPTION DE LA PROBLEMATIQUE SCIENTIFIQUE :

1. Contexte biologique, enjeux et stratégie

Les organismes multicellulaires sont composés d'une immense variété de cellules, chacune ayant différentes fonctions selon l'organe ou le tissu où elle réside. Pour le développement et le fonctionnement normal d'un organisme, le comportement de chaque cellule doit être **strictement et finement régulé**. Cette régulation est basée sur un système de **communication** complexe comprenant la **signalisation** entre organes à travers l'organisme, entre des différents types de cellules et entre les molécules d'une même cellule. Les communications entre cellules distantes de l'organisme mettent en jeu des molécules messagères émises dans le milieu extérieur et captées par les cellules cibles. Ces molécules messagères sont polaires et/ou de grande taille pour pénétrer par les barrières lipidiques comme la membrane cellulaire. Elles doivent donc être reconnues à la surface de la cellule cible par des récepteurs transmembranaires, qui relaient ces signaux de l'extérieur vers l'intérieur de la cellule.

Les **Récepteurs Tyrosine Kinases (RTKs)** sont un des principaux types de récepteurs de surface cellulaire [1,2]. Un mécanisme central dans la signalisation de RTKs est la phosphorylation des résidus tyrosine dans la domaine cytoplasmique du récepteur est la transfert de la phosphate vers une autre protéine qui déclenche une cascade de signalisation pour transmettre des signaux à travers la cellule jusqu'à le noyau (**Figure 1**).

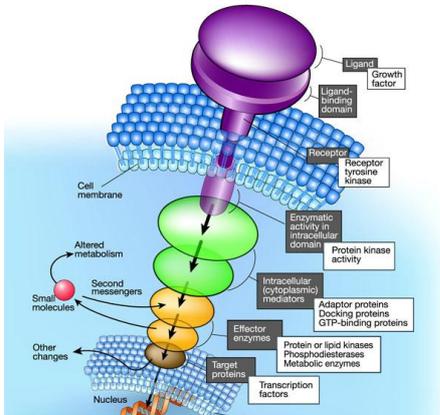


Figure 1. Principe de la signalisation d'un récepteur transmembranaire. J. Downward, *Nature*, 2001

L'effet de phosphorylation est basée sur le changement de conformation du domaine cytoplasmique modulant son activité enzymatique et la création de sites de liaison pour des protéines de signalisation recrutées et puis, activées. La régulation de la signalisation est basé sur la perturbation du récepteurs par des **effecteurs** (ligands, ATP, phosphate, protéine) réagissant sur les sites spécifiques du récepteurs localisé dans les différents domaines du récepteurs. Chaque perturbation dans une site provoque une réponse dans une ou plusieurs autre régions du récepteurs spécialement distants du site de perturbation.

Ce phénomène est nommé l'**effet allostérique**. De plus, comme les RTKs sont capables d'interagir avec plusieurs effecteurs, les perturbations et les réponses aux perturbations sont spécifiquement corrélées, par des mécanismes allostériques hautement régulés.

La dérégulation de cette signalisation allostérique par une **mutation** est un des principaux mécanismes d'apparition des différents formes de cancer ou des autres maladies sévères, ce qui soulignent l'importance des mutations en tant que régulateurs de la prolifération et la survie cellulaire.

2. Définition générale de l'allostérie

L'allostérie est un phénomène de régulation plus général selon lequel une perturbation locale d'une protéine provoque des modifications spécifiques de son activité et/ou fonction. Les protéines allostériques représentent donc des **commutateurs d'information** : les signaux apportés par des stimuli, ou effecteurs, sont détectés, intégrés et transmis *via* des communications intra-protéines en des sites où est élaborée une réponse spécifique. Les effets allostériques sont de natures variées et régulent la majorité des phénomènes biologiques : (i) activation/inhibition; (ii) ouverture/fermeture d'un canal ionique; (iii) oligomérisation/monomérisation; (iv) association/dissociation de sous-unités; (v) perte ou gain d'affinité pour un ligand (O₂-hémoglobine) ou une membrane; (vi) exposition/enfouissement d'un site de reconnaissance pour une autre protéine ou pour des acides nucléiques. La notion d'allostérie peut aussi être élargie à modification induite à distance par une mutation ponctuelle dans la protéine mutée par rapport à son homologue sauvage [3-5] ; à des systèmes macromoléculaires en considérant les interactions protéines-protéines comme des interactions intra-système [6].

3. Différentes conceptions de la communication intra-protéine

Il est communément admis que les voies de communication intra-protéine sont déterminées par la structure 3D de la protéine et le réseau d'interactions liantes et non-liantes formé par ses

acides aminés. Les effets allostériques sont souvent considérés comme la conséquence de chemins de communications intrinsèques entre domaines, c'est-à-dire préexistantes à la perturbation transmise [7]. Un couplage structural *via* des chemins rigides entre sites effecteur et sites affectés dans des protéines allostériquement régulées est mis en évidence [8]. La communication intra-protéine a aussi été décrite d'un point de vue thermodynamique, comme la conséquence de liens énergétiques plus ou moins directs entre des sites distants [9]. Il y aurait donc plusieurs types de mécanismes allostériques, sous-tendus par différents liens de causalité entre repliement/rigidité et flexibilité des sites effecteurs et affectés et des segments les liant. Ces visions de l'allostérie comme un phénomène global à l'échelle de la protéine ne contredisent pas l'idée de transmission de proche en proche de perturbations locales. Des communications allostériques par la transmission de fluctuations de chaînes latérales ont été mises en évidence en 2011 par dynamique de Monte Carlo [10]. Cette coexistence de phénomènes globaux et locaux est une autre illustration de la diversité des mécanismes de communication intra-protéine.

4. Résultats ultérieures

Notre étude par dynamique moléculaire (DM) du KIT, un des RTKs, a mis en évidence que une mutation oncogène provoque des modifications de structure et de dynamique de la protéine intervenant en des régions distantes de la région mutée [11,12]. Pour comprendre des modalités du couplage entre région mutée et région affectée par la mutation, nous avons développé une méthode – MONETA – qui permet de localiser et de visualiser la transmission d'information à travers le réseau de résidus d'une protéine (**Figure 2**). MONETA utilise des données statistiques sur des ensembles conformationnels issus de simulations de DM « tous atomes » relatives d'une part à la topologie de la protéine (interactions inter-résidus) et d'autre part aux corrélations dynamiques inter-résidus. Il permet de décrire à l'échelle atomique les chemins de communication à travers la protéine [4,5].

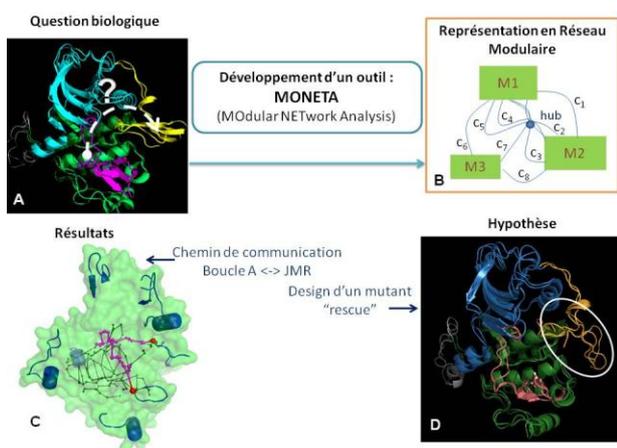


Figure 2. Représentation en réseau modulaire de la région cytoplasmique de KIT

La réalisation de cette analyse peut se diviser en trois phases, les deux premières étant indépendantes:

- (i) La détection de **segments dynamiques indépendants** (*Independent Dynamics Segments, IDSs*), de groupes de résidus présentant des fluctuations atomiques concertés et localement corrélés, distinguables des mouvements globaux ;
- (ii) La détection de **chemins de communication** entre ces segments à travers le réseau d'interactions inter-résidus ;
- (iii) La construction d'un graphique réunissant les *IDSs* et chemins en un **profil de communication** de la protéine.

MONETA décrit des *IDSs* sur la base de corrélations dynamiques à partir d'une trajectoire de DM classique par la technique statistique *analyse de traits locaux* (Local Feature Analysis, LFA). Il s'agissait d'une transformation de données d'analyse des composants principaux (ACP) pour extraire de caractéristiques globales d'un système des caractéristiques topographiques locales de dimensionnalité réduite [13]. Une adaptation de la méthode de LFA [14] a été implémentée dans MONETA.1 pour le calcul des *segments dynamiques indépendant (IDSs)*.

5. Axes principaux de recherche collaborative au sein de Farman

Ce sujet a été au centre de recherche commun au BiMoDyM et au CMLA au cours de 2013 (Projet TopDyn).

Du côté du CMLA, nous avons fait un travail en profondeur pour mettre à jour les angles d'attaques les plus pertinents en terme de plus-value de la collaboration. Il est apparu nécessaire de focaliser les efforts sur l'outil MONETA, qui concentre l'expertise de l'équipe BiMoDyM, et qui est à la base de toutes les analyses faites sur les données de simulation de la dynamique moléculaire. Nous pensons qu'en même temps que se poursuit l'évolution de MONETA dans l'équipe BiMoDyM, il y a un potentiel d'innovation certain dans la mise à plat et la formalisation sur des bases probabilistes et statistiques plus explicites des représentations des dynamiques par MONETA qui pourrait en augmenter la portée. Dans ce projet ambitieux, nous avons pu obtenir un premier résultat que nous jugeons extrêmement encourageant en proposant une méthode alternative, la *Principal Feature Decomposition*, pour la détection des briques de bases que sont les IDS en s'appuyant sur le calcul d'une mesure de dépendance stochastique entre les mouvements d'atomes dans la dynamique et adossée à un algorithme de poursuite de projection que nous avons testé sur des données de MD du kit.

BiMoDyM a poursuivi l'évolution et l'optimisation de MONETA qui a permis de réaliser une version MONETA.2 [15] qui diminue grandement les temps de calcul et les volumes de données produits. Cette version ajoute aussi un certain nombre de fonctionnalités, en particulier pour l'analyse des résultats, comme la visualisation des communications sous forme de graphiques dynamiques ou la possibilité de rechercher les chemins de communication entre résidus choisis. Cette version permis d'analyser la communication d'une dizaine de protéines différents [16,17].

La mise en place d'un serveur commun de calcul, le partage de données et de logiciel de visualisation a énormément fait pour concrétiser les bases d'une collaboration durable et stimuler les échanges. Pour accélérer de calcul, plus particulièrement des simulations de DM d'une part, et pour accéder aux données de DM sur des échelles de temps dans la gamme ($10^{-6} - 10^{-1}$ s) d'autre part, nous voulons d'installer sur le serveur la nouvelle version du programme AMBER (v2013) adapté pour GPU.

En 2014 nous planifions l'évolution de MONETA sur deux axes :

(i) Implémentation dans MONETA de méthode *Principal Feature Decomposition* (PFD) développé par A. Trouvé (CMLA) ; Le prolongement de l'analyse probabiliste pour appuyer la définition des chemins de communication sur des représentations de la circulation de l'information basées sur des processus markoviens de saut entre résidus dont les probabilités de transitions sont données par des métriques mélangeant la *géométrie spatiale* de la protéine et la *géométrie dynamique stochastique* donnée par la structure de dépendances entre les dynamiques individuelles des atomes. Nous comptons en particulier utiliser comme *feature dynamique* pour chaque atome le sous-espace 3D gaussien engendré par sa dynamique et introduire des métriques naturelles sur la Grassmannienne $G(3,3N)$ où N est le nombre d'atomes [18] et une exploration des approches spectrale associées sur les graphes sous-jacents.

(ii) Développement de MONETA pour analyser la communication dans ADN et des complexes moléculaires (protéine/protéine et protéine/ADN) et intégration des codes dans MONETA. La nouvelle version de MONETA sera appliquer sur différents protéines et leurs complexes macromoléculaires.

ORIGINALITE et APPORT SCIENTIFIQUE:

Ce travail motivé à la base par des considérations biologiques et médicales combine à la fois des connaissances en biologie, en bioinformatique et en mathématiques. Le développement d'outils originaux adaptés aux questionnements des biologistes ne peut se faire que si une réelle collaboration sur les besoins, les attentes et le savoir-faire des uns et des autres se met en place. Farman semble être le lieu approprié pour cela. Les résultats de travail seront résumés et publiés dans les articles communs. Dans le projet seront associées plusieurs étudiants de Master et de thèse pour le côté BiMoDyn et le côté CMLA qui sont déjà recrutés.

REFERENCES :

1. Lemmon, M. A. & Schlessinger, J. (2010). Cell Signaling by Receptor Tyrosine Kinases. *Cell* 141, 1117-1134.
2. Olsen, J. V., Blagoev, B., Gnani, F., Macek, B., Kumar, C., Mortensen, P., & Mann, M. (2006). Global, in vivo, and site-specific phosphorylation dynamics in signaling networks. *Cell* 127, 635-648.
3. Shan Y, et al, Oncogenic mutations counteract intrinsic disorder in the EGFR kinase and promote receptor dimerization. (2012). *Cell* . 149(4): p. 860-70.
4. Laine E., Chauvot de Beauchêne I., Auclair C., Tchertanov L. (2011). Propagation of D816V/H mutation effects across KIT receptor. *Eur. Biophys. J. Biophys. Lett.* 40:S109, 2011.
5. Laine, E., Auclair, C. and Tchertanov, L. (2012). Allosteric Communication across the Native and Mutated KIT Receptor Tyrosine Kinase. *PLoS Comput Biol.* 8(8): e1002661. 14 pages, doi:10.1371/journal.pcbi.1002661.
6. Bray D. (2013). The propagation of allosteric states in large multiprotein complexes. *J Mol Biol*, 425(9): 1410-4.
7. del Sol A. et al, The origin of allosteric functional modulation: multiple pre-existing pathways. (2009). *Structure*, 17(8): 1042-50.
8. Rader A. J. and S. M. Brown, Correlating allostery with rigidity. *Mol Biosyst*, 2011. 7(2): p. 464-71.
9. Wrabl J. O. et al, The role of protein conformational fluctuations in allostery, function, and evolution. (2011). *Biophys Chem*, 159(1): p. 129-41.
10. Dubay K. H. et al. (2011). Long-range intra-protein communication can be transmitted by correlated side-chain fluctuations alone. *PLoS Comput Biol.* 7(9): p. e1002168.
11. Laine E., Chauvot de Beauchêne I., Perahia D., Auclair C. and Tchertanov L. (2011). Mutation D816V Alters the Internal Structure and Dynamics of c-Kit Cytoplasmic Region: Implications for Dimerization and Activation Mechanisms. *PLoS Comput Biol* 7(6): e1002068. 20 pages, doi:10.1371/journal.pcbi.1002068.
12. Chauvot de Beauchêne I., Laine E., Auclair C., Tchertanov L. (2011). Structural, dynamic and thermodynamic effects of KIT mutations: a computational multi-approach study. *Eur. Biophys. J. Biophys. Lett.* 40:S103, 2011
13. Penev, P. S. & Atick, J. J. (1996). Local feature analysis: A general statistical theory for object representation. *Network: Computation in Neural systems* 7, 477-500.
14. Zhang, Z. Y. & Wriggers, W. (2006). Local feature analysis: A statistical theory for reproducible essential dynamics of large macromolecules. *Proteins-Structure Function and Bioinformatics* 64, 391-403.
15. Chauvot de Beauchêne I., Alain A., Laine E., Dubreuil P. and Tchertanov L. (2013). Oncogenic mutations of KIT receptor differentially modulate tyrosine kinase activity and drug susceptibility. *PloS Comp. Biol. Submitted*
16. Da Silva Figueiredo Celestino P., Panel. N., Laine E., Pascutti P. G., Solary E. and Tchertanov L. (2014). Does mutation D802V of the CSF1 Receptor alternates the tyrosine kinase tertiary structure and allosteric communication? *Submitted*
17. Langenfeld F, Guarracino Y, Arock M. and Tchertanov L. (2014). Structure, Dynamics and Communication features of Signal Transducers and Activators of Transcription Factor STAT5. *In preparation.*
18. Charon N. and Trouvé A. (2013). The varifold representation of non-oriented shapes for diffeomorphic registration. *SIAM Journal of Imaging Science* 6(4): 2547-2580

FINANCEMENT DEMANDE :

Pour chacun des 2 partenaires, l'aide annuelle demandée est

-7 000 Euros pour le fonctionnement (financement de la gratification d'un étudiant, licence pour logiciels, frais de congrès)

-3 000 Euros pour équipement

soit un total de 20 000 Euros dont 14 000 en fonctionnement et 6000 en équipement.