

## 3BIO-BIOINFO

### *Unité de Bioinformatique génomique et structurale*

#### **Service de BioModélisation, BioInformatique & BioProcédés**

**Pr. M. Rooman et D. Gilis**

---

#### **Liste des sujets**

1. Modification rationnelle de la résistance des protéines à la température
  2. Prédiction des changements de solubilité de protéines suite à une mutation
  3. Analyse bioinformatique de variations ponctuelles répertoriées au sein du génome humain et de leur relation avec le développement de maladies
  4. Conception rationnelle de vaccins : prédiction d'épitopes B
  5. Modélisation de la structure d'un récepteur membranaire et de ses interactions avec un peptide
  6. Etude de la dynamique d'allergènes d'acariens de la famille 5
  7. Origine du code génétique et des peptides prébiotiques
  8. Modélisation dynamique de l'expression des gènes de levures et de bactéries soumises à des perturbations
  9. Conception des circuits synthétiques de gènes et contrôle du bruit
  10. Développement d'algorithmes d'amarrage protéine-protéine
  11. Développement d'un algorithme de placement des chaînes latérales dans les protéines
- 

#### **1. Modification rationnelle de la résistance des protéines à la température [Yves Dehouck, Marianne Rooman]**

L'évolution a conféré aux protéines une fonction biologique, généralement déterminée par leur structure tridimensionnelle. Certaines protéines par exemple sont des catalyseurs particulièrement spécifiques et efficaces. L'exploitation des propriétés fonctionnelles des protéines dans l'industrie (agro-alimentaire, chimique, pharmaceutique, ...) constitue un atout majeur mais est limitée par le fait que les protéines ne sont généralement stables et actives que sous des conditions physiologiques, et en particulier dans un intervalle de température fort restreint. Dès lors, la conception de protéines modifiées qui gardent leur structure et donc leur fonction à des températures plus (ou moins) élevées est un objectif important de la recherche fondamentale mais aussi industrielle.

Le projet consiste à mettre au point un logiciel de prédiction de changements de stabilité thermique de protéines lors de mutation, en utilisant des potentiels statistiques dépendant de la température développés dans notre groupe, et à appliquer ce logiciel à la conception rationnelle de protéines d'intérêt industriel ayant une résistance à la température modifiée et qui sont donc susceptibles d'être plus efficaces dans l'environnement industriel. Ce sujet se situe dans le contexte de la création d'une société spin-off émanant de notre unité de recherche.

## **2. Prédiction des changements de solubilité de protéines suite à une mutation [Dimitri Gilis, Yves Dehouck et Marianne Rooman]**

Le domaine des biotechnologies exploite à très large échelle les propriétés des protéines. Toutefois, la faible solubilité de certaines de ces protéines rend leur manipulation et leur purification peu aisée, et affecte leur activité. La modification ciblée de la séquence de ces protéines au travers de mutations peut permettre de remédier à ces problèmes de solubilité. Nous proposons de mettre en place une approche prédictive qui permettra d'évaluer l'impact d'une mutation sur la solubilité d'une protéine. Cette méthode de prédiction *in silico* suivra une approche similaire à celle d'autres algorithmes développés précédemment ou en cours de développement dans notre unité, tel que celui visant à prédire les changements de stabilité thermodynamique lors de mutations (cf <http://babylone.ulb.ac.be/popmusic>) ou de stabilité thermique. La combinaison de ces logiciels permettra d'offrir une aide précieuse à la conception rationnelle de protéines présentant des propriétés physico-chimiques modifiées. Ce sujet se situe également dans le contexte de la création d'une société spin-off émanant de notre unité de recherche.

## **3. Analyse bioinformatique de variations ponctuelles répertoriées au sein du génome humain et de leur relation avec le développement de maladies [Yves Dehouck, Marianne Rooman]**

Au cours de ces quelques dernières années, les technologies de séquençage de l'ADN à haut débit ont vu leur coût radicalement diminuer, ce qui a mené à l'obtention de quantités massives de données. L'analyse et l'exploitation de ces données est à l'heure actuelle un défi majeur en génomique et bioinformatique. En particulier, des millions de variations ponctuelles du génome humain (SNP : Single Nucleotide Polymorphism) ont été répertoriées via des projets collaboratifs tels que l'HapMap Consortium ou le Human Variation Project. Chacune de ces variations peut avoir un impact sur l'expression ou la fonction de certains gènes et affecter de manière plus ou moins importante le fonctionnement de l'organisme. Les variations au sein de régions codantes de l'ADN ont un intérêt particulier car elles peuvent mener à des variations de la séquence de protéines exprimées (SAAP : Single Amino Acid Polymorphism). L'annotation de ces variations et de leur impact sur la santé humaine est une étape importante en vue d'une meilleure compréhension du caractère génétique de certaines maladies et du développement de thérapeutiques plus personnalisées pour le traitement de maladies complexes telles que le cancer ou le diabète.

Le projet consiste en le développement d'une méthode automatisée d'analyse de SAAPs en vue de la prédiction de leur caractère pathogène ou non. Au contraire de la plupart des méthodes existantes, l'approche envisagée aura pour objectif de tirer parti d'informations structurales sur les protéines variantes. Il s'agira de modéliser de manière automatique les structures des protéines ciblées, et de leur appliquer divers logiciels d'analyse structurale développés notamment dans notre unité de recherche, afin de prédire les effets de SAAPs et leur relations potentielles avec le développement de maladies.

## **4. Conception rationnelle de vaccins : prédiction d'épitopes B [Marianne Rooman]**

La prédiction d'épitopes B *in silico*, via l'utilisation d'outils bioinformatiques, est une voie extrêmement prometteuse dans la conception rationnelle de vaccins, vu qu'elle guide et limite drastiquement les tests expérimentaux – entre autres sur des animaux –, ce qui entraîne des gains importants en temps et en coûts. La mise au point d'un outil fiable de prédiction d'épitopes B constitue un objectif de toute première importance.

Pour développer un logiciel de prédiction des épitopes B, nous utiliserons les séquences et les structures d'épitopes B caractérisées expérimentalement, qui sont répertoriées dans des bases

de données spécialisées. Les caractéristiques de séquence et de structure des interfaces anticorps-antigène seront ensuite identifiées, ainsi que leurs caractéristiques énergétiques. Elles seront rationalisées, combinées, et formeront la base de la méthode de prédiction.

### **5. Modélisation de la structure d'un récepteur membranaire et de ses interactions avec un peptide** [Dimitri Gilis, Yves Dehouck et Marianne Rooman]

Le VIP est un neuropeptide synthétisé et sécrété par le système nerveux central. Il a été montré que le VIP présente des propriétés anti-inflammatoires au travers de son interaction avec le récepteur membranaire VPAC1. Le VIP peut également interagir avec les récepteurs membranaires VPAC2 et PACAP. L'objectif de ce projet sera de modéliser la structure de récepteurs membranaires (VPAC1, VPAC2 ou PACAP) et d'étudier *in silico*, par des techniques d'amarrage, l'interaction VIP-récepteur. Ce MFE se place dans le cadre général d'une collaboration avec l'IRIBHM (Prof. I. Langer).

### **6. Etude de la dynamique d'allergènes d'acariens de la famille 5** [D. Gilis]

L'allergie aux acariens est un problème de santé publique qui touche une partie non négligeable de la population. Les allergènes issus d'acariens se subdivisent en diverses familles. Parmi celles-ci, la famille 5 constitue une famille de protéines encore peu connues et qui se structurent en trois hélices antiparallèles. Certaines protéines présentent une forme monomériques, d'autres ont été observées sous forme de dimères. Dans ces dimères, une hélice présente un coude, et une cavité pouvant potentiellement accueillir un ligand est présente. L'implication de ces particularités sur la capacité qu'ont ces allergènes à activer le système immunitaire est encore inconnue.

L'objectif de ce MFE est d'étudier par des approches de dynamique moléculaire la dynamique de certaines protéines issues de cette famille 5, et l'influence des conditions de solvant et de certaines mutations sur cette dynamique, en relation avec leurs propriétés allergéniques.

### **7. Origine du code génétique et des peptides prébiotiques** [Marianne Rooman]

Bien que certains éléments présidant à l'origine de la vie ont pu être déterminés avec une certaine confiance (synthèse prébiotique d'acides aminés, formation de peptides et autres molécules complexes dans des conditions prébiotiques, formation d'ancêtres de membranes cellulaires,...), il reste de nombreux points non élucidés, tels que l'origine du premier code génétique et la forme des premiers peptides fonctionnels. En ce qui concerne le code génétique, il a été suggéré que des séquences d'ARN spécifiques interagissent avec des acides aminés individuels, et que ces séquences contiennent des codons et/ou des anticodons des acides aminés auxquels ils se lient. En ce qui concerne les peptides prébiotiques, des études récentes suggèrent que de petits peptides de 5 à 8 résidus, formés des acides aminés les plus simples (Gly, Ala, Asp et Val), pourraient constituer le chaînon manquant entre les acides aminés individuels synthétisés dans les conditions prébiotiques de la terre primitive et les protéines actuelles qui sont encodées dans les génomes.

Dans ce contexte, le projet consistera à étudier l'affinité d'acides aminés individuels pour les séquences d'ARN spécifiques, via des simulations de dynamique moléculaire, afin de déterminer quelles sont les interactions préférées, ce qui pourrait constituer une piste pour comprendre l'origine du code génétique. D'autre part, la structure et la stabilité des peptides proposés comme étant des peptides prébiotiques seront également étudiés par dynamique moléculaire. Des analyses expérimentales structurales et fonctionnelles peuvent être envisagées, par spectroscopie infra-rouge en collaboration avec le Pr. E. Goormaghtigh (SFMB), et par des techniques de gel SDS-PAGE et spectrométrie de masse en collaboration avec le Pr. I. Langer (IRIBHM).

## **8. Modélisation dynamique de l'expression des gènes de levures et de bactéries soumises à des perturbations** [Marianne Rooman et Jaroslav Albert]

Les mécanismes de régulation de l'expression des gènes forment des réseaux complexes, qui impliquent un grand nombre de protéines et dépendent du cycle cellulaire, du stade de développement de l'organisme, du tissu, du caractère sain ou malade de la cellule, et des perturbations extérieures. Une quantité croissante de données expérimentales sur les gènes transcrits dans différents systèmes et conditions, obtenues par des techniques de puces à ADN, sont disponibles dans des bases de données publiques. Nous nous focaliserons sur des séries temporelles, décrivant des cellules de levures ou de bactéries soumises à des perturbations externes comme des radiations ou des modifications de température, ou montrant des phénomènes de diauxie. L'évolution temporelle des interactions entre les gènes et leurs produits (ARN, protéines) dans ces systèmes sera modélisée par des systèmes d'équations différentielles, dont les paramètres seront identifiés sur la base des données expérimentales disponibles. Les différents modèles possibles seront analysés sur la base de leur stabilité, leur robustesse, et leur évolution vers un point fixe.

## **9. Conception des circuits synthétiques de gènes et contrôle du bruit** [Jaroslav Albert, Marianne Rooman]

La biologie synthétique est un domaine relativement nouveau dont le but est d'améliorer des circuits de gènes existants et d'en concevoir des nouveaux. Afin de mettre en œuvre un nouveau circuit de gènes, il faut comprendre son comportement en fonction des paramètres du système. Ceci exige que nous ne comprenions pas uniquement le comportement moyen (déterministe) du système, mais aussi comment des variations aléatoires, en particulier le bruit interne et externe, affecte ses performances. Vu que le temps de calcul pour des circuits de gènes de plus en plus complexes croît de façon exponentielle, il est nécessaire de trouver des méthodes simplifiées de simulation stochastique. Une de ces méthodes implique la séparation d'un grand circuit en petits motifs, qui peuvent être compris en termes d'un petit nombre de paramètres du système. Ensuite, des règles générales sur le comportement du système peuvent en être déduites. En utilisant une combinaison de modèles déterministes (des équations différentielles, par exemple) et des algorithmes stochastiques, nous allons analyser de petits circuits avec autorégulation positive. Ces circuits sont intéressants en raison de leur capacité à augmenter le caractère aléatoire de leur propre expression.

## **10. Développement d'algorithmes d'amarrage protéine-protéine** [Dimitri Gilis, Yves Dehouck, Marianne Rooman]

Un grand nombre de protéines ne remplissent leur rôle biologique que lorsqu'elles forment des complexes multimériques, constitués de monomères identiques ou distincts, ou lorsqu'elles interagissent avec un ligand. Les interactions protéine-protéine et protéine-ligand permettent également la régulation de certains processus biologiques, la transmission des signaux, ou encore la catalyse de diverses réactions biochimiques. Comprendre et prédire la façon dont les protéines interagissent est donc d'un intérêt indéniable.

Le projet consiste à élaborer et à implémenter un algorithme d'amarrage protéine-protéine ou protéine-ligand, qui prédit l'orientation relative de deux protéines ou d'une protéine et un ligand. Dans ce but, plusieurs algorithmes d'échantillonnage de l'espace des conformations seront envisagés, couplés à des fonctions d'énergie qui permettent de quantifier les interactions. Une réflexion sera menée sur la manière de modéliser la flexibilité moléculaire. L'algorithme le mieux adapté sera implémenté et testé. Ce MFE se place dans le cadre général d'une collaboration avec le Département d'Informatique (Pr. J. Cardinal).

## **11. Développement d'un algorithme de placement des chaînes latérales dans les protéines** [Dimitri Gilis, Yves Dehouck, Marianne Rooman]

Les programmes de prédiction de structure de protéines à partir de leur séquence et les programmes d'amarrage protéine-ligand utilisent généralement une représentation simplifiée de la structure des protéines, négligeant les degrés de liberté des chaînes latérales des acides aminés qui la composent. Cette simplification est nécessaire pour permettre la recherche efficace dans l'espace conformationnel en un temps raisonnable. Toutefois, le positionnement correct des chaînes latérales est nécessaire dans un deuxième temps, afin de permettre de raffiner les structures prédites et d'aborder les mécanismes fonctionnels. Le positionnement des chaînes latérales est un problème combinatoire complexe, qui doit rendre compte des préférences conformationnelles de chaque chaîne latérale, des interactions préférentielles entre ces chaînes, de leur empilement optimal et d'une certaine flexibilité de la chaîne principale. Le projet consistera à utiliser des algorithmes d'optimisation (programmation linéaire), en combinaison avec différents types de fonctions d'énergie, dont les poids relatifs seront déterminés par des procédures d'identification paramétrique. Ce MFE se place dans le cadre général d'une collaboration avec le Département d'Informatique (Pr. J. Cardinal).

### **Coordonnées**

Pr. Marianne Rooman : [mrooman@ulb.ac.be](mailto:mrooman@ulb.ac.be)

Pr. Dimitri Gilis: [dgilis@ulb.ac.be](mailto:dgilis@ulb.ac.be)

Dr. Yves Dehouck: [ydehouck@ulb.ac.be](mailto:ydehouck@ulb.ac.be)

Dr. Jaroslav Albert : [jalbert@ulb.ac.be](mailto:jalbert@ulb.ac.be)