

## Analyse de Chimiothèques et Construction Automatique de Modèles de Bio-Activité

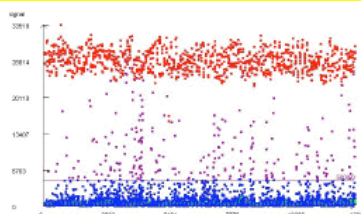
<http://accamba.imag.fr/>

### CONTEXTE

Le criblage automatisé (HTS : High Throughput Screening) de collections de molécules (chimiothèques) est utilisé depuis une vingtaine d'années dans l'industrie pharmaceutique pour découvrir des molécules bio-actives, qui soient effectrices sur des systèmes biologiques et susceptibles de fournir de nouveaux médicaments. Cependant, dans le contexte de la « post-génomique », la mise en évidence de telles molécules, même sans effets thérapeutiques immédiatement exploitables, est également intéressante dès lors qu'elles permettent l'étude de fonctions biologiques et qu'elles munissent la recherche de nouveaux outils moléculaires autorisant la mise en oeuvre d'approches exploratoires de type « génétique chimique ». L'exploitation des données d'un criblage demeure un problème complexe de part la présence d'un faible taux de résultats positifs (les quelques molécules bio-actives sélectionnées par seuillage) ainsi que de la forte présence de bruit (incertitude sur les signaux, faux négatifs, ...) sur les mesures.

### DONNÉES DE CRIBLAGE

La figure ci-contre montre les résultats obtenus lors d'une expérience de criblage du CMBA/CEA. En abscisse on trouve les plaques contenant les composés analysés et en ordonnées l'intensité du signal de bioactivité qui a été mesuré.



### OBJECTIFS

Le projet ACCAMBA se propose de concevoir des outils d'analyse de chimiothèques et de modélisation des résultats de criblage en faisant collaborer étroitement des biologistes, des informaticiens et des chimistes. Il vise à fédérer les travaux réalisés en France sur plusieurs plates-formes de criblage académiques (CMBA, Grenoble; SMMBC, Saclay; MCT, Roscoff) et sur la constitution de chimiothèques (Institut Curie, Paris). Les objectifs du projet ACCAMBA sont à la fois informatiques et biologiques, il s'agit de :

- Concevoir un ensemble de méthodes d'exploration globale de la bio-activité potentielle de chimiothèques académiques et commerciales et de permettre une caractérisation de la bio-diversité d'une chimiothèque.
- Mettre en place un cycle : Criblage – Modélisation – Synthèse – Validation dans lequel la partie modélisation des SAR (Structure-Activity Relationship) repose sur des techniques d'apprentissage artificiel (« machine learning »).
- Valider les outils sur l'étude de l'activité de molécules à potentiel anticancéreux dans une double stratégie de génétique chimique « directe » (l'altération des phénotypes cellulaires résultant de la modification de l'état de polymérisation de la tubuline) et « inverse » (l'inhibition de l'assemblage de la tubuline ainsi que de protéines kinases, in vitro).

### RESULTATS

Au cours de cette première année, nous avons eu une douzaine de réunions entre les participants dont 2 réunions plénières. Le projet a travaillé sur les tâches suivantes avec publication d'une partie des résultats (cf. références) :

- Normalisation des données de criblage et recherche des faux négatifs
- Recherche bibliographique sur les descripteurs moléculaires
- Mise en place d'un premier langage de représentation des molécules
- Analyse des outils d'apprentissage symboliques

### CALENDRIER

**Première année 2004-2005.** Recueil et normalisation des données, mise en place des langages de représentation et des méthodes d'apprentissage.

**Seconde année 2005-2006.** Validation de l'ensemble du cycle de simulation et premières synthèses de molécules.

**Troisième année 2005-2006.** Utilisation de la plateforme logicielle sur d'autres chimiothèques et éventuellement sur des cibles biologique voisines.

### LABORATOIRES IMPLIQUES

1. Équipe Apprentissage, Laboratoire Leibniz-IMAG UMR 5522, Grenoble.
2. Équipes TAO et BioInformatique, Laboratoire de Recherche en Informatique UMR 8623, Université Paris Sud.
3. Équipe de Criblage pour des Molécules Bio-actives, CEA/INSERM U366, Grenoble.
4. Molécules et Cibles Thérapeutiques, UPS 2682 CNRS, Roscoff
5. Service de Marquage Moléculaire et de Chimie Bio-organique, CEA-Saclay.
6. Conception Synthèse et Vectorisation de Biomolécules, UMR 176 CNRS/IC.
7. Structures et Fonctions des Biomolécules, Institut de Biologie de Lille, UMR8525.

### MOTS CLES

Bio-chemo-informatique, Criblage, Chimiothèques, Apprentissage artificiel, Construction de Modèle SAR, Modélisation moléculaire, Génétique chimique.

### PUBLICATIONS PRINCIPALES

Maréchal E., Roy S., Lafanechère L. (éditeurs) *Le criblage pharmacologique automatisé : du haut-débit au haut-contenu d'information. Une introduction à l'usage des biologistes, des chimistes et des informaticiens*, 2005. Ouvrage collectif, en cours d'édition. Chapitres écrits par des membres du projet :

- Barette C. : La démarche qualité pour le criblage automatisé
- Bisson G. : Les méthodes d'apprentissages sur données de criblage
- Horvath D. : L'annotation et la classification de l'espace chimique pour la chemogénomique
- Lafanechère L. : Le criblage phénotypique sur cellules et les stratégies de génétique chimique directe
- Maréchal E., Roy S., Lafanechère L. : Le processus de criblage pharmacologique : la petite molécule, la cible biologique, l'automate, le signal et l'information
- Maréchal E. : La mesure de la bio-activité : Ki, IC50 et EC50
- Roy S. : La modélisation du criblage pharmacologique : maîtrise des informations chimiques, biologiques et expérimentales
- Wiczorek S. Le signal : aspects statistiques, normalisation, analyse élémentaire

### PROJET

Le projet ACCAMBA vise à mettre en place une *plate-forme de criblage virtuel*. L'architecture générale du projet en terme de tâches et de flux de données est décrite dans la figure ci-dessous. Il faut souligner que les tâches T2 à T4 ne sont pas à proprement parler séquentielles. En effet, le traitement des données et le raffinement des outils sont placés dans une boucle de révision et d'optimisation des connaissances.

