

iRTSV



Institut de Recherches en Technologies et Sciences pour le Vivant

Fonctions intégrées des protéines - Du vivant aux nanotechnologies

Direction des Sciences du Vivant

Quelques logiciels scientifiques développés au GIPSE

Thématique n° 10

Avril 2011

Editorial

Céline Charavay,
Responsable GIPSE
Page 1

SimuMoss

Équipe PMB
Page 2

Morpheus

Équipe Développement floral
Page 3

TAMIS

CMBA
Page 4

Éditorial

Stocker ses données scientifiques de manière sécurisée, les organiser, les analyser ... Toutes ces problématiques font maintenant partie du quotidien de la plupart des scientifiques. Afin de répondre à ces différents besoins, des infrastructures informatiques ou des logiciels peuvent être conçus par des informaticiens qui accompagnent l'activité des scientifiques.

Au niveau de la Direction des Sciences du Vivant du CEA, deux groupes de support en informatique scientifique répondant à ces missions, ont été créés en 2005 : le Groupe Informatique Pour les Scientifiques d'Ile de France (GIPSI) et le Groupe Informatique Pour les Scientifiques du sud-Est (GIPSE).

Le GIPSE intervient transversalement pour trois instituts de la DSV : l'Institut de biologie environnementale et biotechnologie (iBEB), l'Institut de biologie structurale (IBS) et l'Institut de recher-

ches en technologies et sciences pour le vivant (iRTSV).

Ses missions sont doubles :

- ☑ **conception et gestion d'infrastructures informatiques** spécifiques aux laboratoires de recherche
- ☑ **développement de logiciels scientifiques** nécessaires à la recherche en biologie.



L'équipe du GIPSE.
(de gauche à droite) Vincent Dorffer, Céline Charavay et Stéphane Segard.

Dans le cadre de cette dernière mission, une dizaine de logiciels ont été développés depuis la création du GIPSE. Ces logiciels répondent à des besoins très variés : traçabilité de lignées de plantes transgéniques, pilotage du robot d'une plate-forme de cristallisation de protéines, gestion de campagnes de clonage

de gènes et d'expression de protéines, analyse de données de Résonance Magnétique Nucléaire (RMN) mesurées sur des protéines désordonnées...

Cette lettre s'attache à présenter trois logiciels dont une partie ou la totalité du développement a été assurée par le GIPSE pour les scientifiques de l'iRTSV :

- ☑ **TAMIS** (gestion et analyse de données de criblage)
- ☑ **SimuMoss** (analyse de spectres Mössbauer)
- ☑ **Morpheus** (recherche des sites de fixation d'un facteur de transcription).

Ces trois logiciels sont ou seront prochainement distribués à la communauté scientifique.

En attendant de vous compter peut-être parmi l'un des utilisateurs de ces logiciels, je vous souhaite une agréable lecture.

Céline Charavay,
Responsable du GIPSE



Un logiciel pour l'analyse de spectres Mössbauer

Rudolf Mössbauer découvre en 1958 l'absorption résonante sans recul des rayons gamma. La technique de spectroscopie qui en découle (Figure 1) permet de sonder la structure électronique de quarante-six éléments, dont le fer qui est l'un des métaux essentiels du monde vivant. Impliqué dans de nombreux processus primordiaux de la vie tels que la fixation et l'activation du dioxygène, il participe aussi au stress oxydant et ses effets délétères et un contrôle déficient de son homéostasie peut conduire à de nombreuses pathologies. Cet élément est un sujet d'étude privilégié pour les biologistes et les chimistes et plus particulièrement pour les chercheurs de l'équipe Physico-chimie des Métaux en Biologie (PMB) du Laboratoire de Chimie et Biologie des Métaux (LCBM). Cette équipe a ainsi récemment acquis des équipements de pointe en spectroscopie Mössbauer afin de caractériser et d'étudier la structure et le fonctionnement des sites actifs d'enzymes à fer et de complexes modèles.

Une plate-forme Mössbauer dédiée au vivant

La plate-forme Mössbauer mise en place par l'équipe PMB du LCBM est unique en France et n'a que peu d'équivalents dans le monde, en Allema-

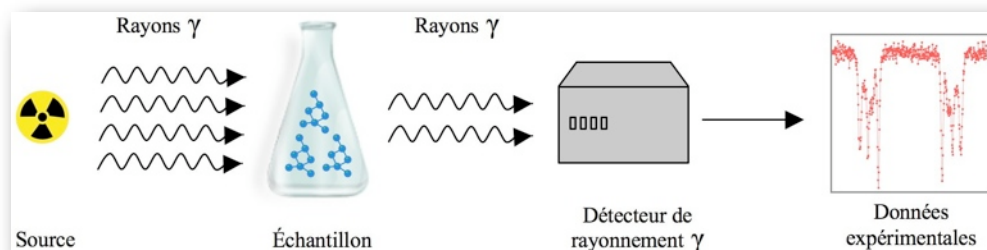


Figure 1 : Principe de la spectroscopie Mössbauer.

Une source radioactive, animée d'un mouvement périodique, émet des photons γ . Après avoir traversé l'échantillon, le flux de photons non absorbés est recueilli par un détecteur. Les données expérimentales récoltées par le détecteur correspondent aux nombres de « coups » (impacts) mesurés sur chacun des canaux du détecteur. Seuls les noyaux de l'isotope 57 du Fer absorbent les rayons γ émis par la source (2,15% d'abondance naturelle).

gne et aux Etats-Unis. Elle permet d'accéder à des informations très fines sur les atomes fer des systèmes étudiés : structure électronique, degré d'oxydation, nature et environnement (nombre et nature des atomes liés au fer)... Ces informations sont obtenues en analysant, grâce à un logiciel, des données expérimentales enregistrées sous différentes conditions de température et/ou de champ magnétique (champ parallèle ou perpendiculaire à la direction des rayons gamma, d'une intensité comprise entre 0 et 8 Teslas). Les échantillons étudiés peuvent être des poudres ou des solutions gelées.

Les logiciels existant ne permettant pas d'analyser la plupart des données acquises, les chercheurs de l'équipe PMB ont été amenés à développer, en collaboration avec le GIPSE, le logiciel *SimuMoss*.

SimuMoss, un logiciel d'analyse de spectres Mössbauer

Cette collaboration a permis d'aboutir à un logiciel convivial, souple d'utilisation et évolutif. Il permet aux expérimentateurs de la plate-forme de traiter

les différentes étapes d'analyse des données Mössbauer :

- **obtenir un spectre expérimental** à partir des données enregistrées sur lesquelles des corrections de ligne de base peuvent éventuellement être appliquées
- **déterminer le modèle théorique** du site actif étudié (nombre de noyaux fer en interaction magnétique, spin, paramètres nucléaires et électroniques...) qui reproduit au mieux les spectres expérimentaux (Figure 2).

Ce programme est actuellement en cours de test et de développement. Grâce à lui, les chercheurs du LCBM peuvent maintenant caractériser des entités $\text{Fe}^{\text{IV/V}}\text{-oxo}$ générées par action de peroxyde (H_2O_2 par exemple). L'une des difficultés de l'analyse est l'existence d'autres entités à base de fer, que ce soit le produit de départ ou des produits de dégradation, tous comportant du fer et donc absorbant les rayons gamma. Un autre exemple est celui de complexes $\text{Fe}^{\text{II/III}}\text{Mn}^{\text{II}}$, modèle du site actif de certaines oxygénases. La structure électronique du système $\text{Fe}^{\text{II}}\text{Mn}^{\text{II}}$ est particulièrement complexe compte tenu de la faible interaction magnétique entre les deux ions métalliques.

Les évolutions de SimuMoss

Le logiciel *SimuMoss* continuera d'accompagner la montée en puissance de la plate-forme Mössbauer. En effet, une partie de l'activité de cette dernière réside dans l'étude de nanoparticules à base de fer ou de matériaux magnétiques comportant des oxydes et hydroxydes de fer. Cette activité étant amenée à perdurer, *SimuMoss* sera complété par un module spécialement adapté à l'analyse des données enregistrées sur des matériaux.

SimuMoss étant susceptible d'intéresser la communauté scientifique travaillant sur la spectroscopie Mössbauer, il sera à terme distribué à titre gracieux à cette communauté.

Contacts : [Geneviève Blondin](#) et [Martin Clémancey](#)

LCBM

Laboratoire Chimie et Biologie des Métaux

UMR 5249 - CEA - CNRS - UJF

et [Céline Charavay](#)

GIPSE

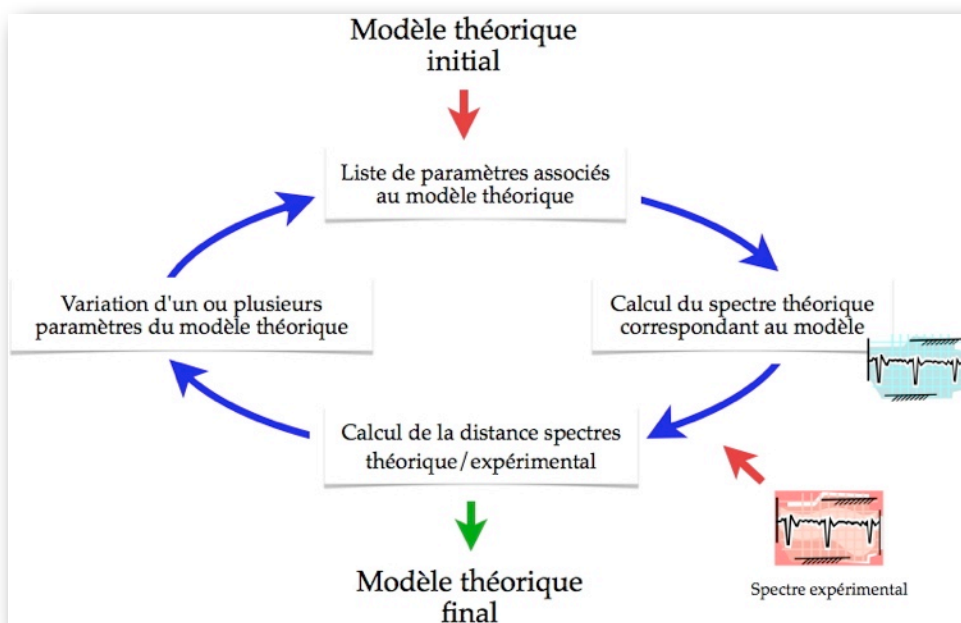


Figure 2 : Processus itératif de détermination du modèle théorique.

À chaque itération, on fait varier un ou plusieurs des paramètres du modèle initial jusqu'à obtenir un spectre théorique qui reproduit au mieux le spectre expérimental. Plusieurs spectres expérimentaux peuvent être simultanément minimisés.

Morpheus

Le contrôle de la transcription des gènes est au cœur de tous les processus du vivant. Cette régulation dite transcriptionnelle met en jeu des protéines, appelées facteurs de transcription, capables de stimuler ou d'inhiber la production des ARN messagers en se liant à des séquences d'ADN localisées près du gène régulé (Figure 1). La recherche des sites de fixation d'un facteur de transcription dans un génome est un enjeu de taille car elle aide à déterminer quels sont les gènes régulés par ce facteur.

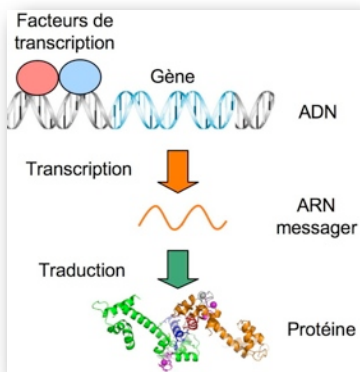


Figure 1 : Principe de la régulation transcriptionnelle. La fixation d'un ou de plusieurs facteurs de transcription peut activer ou inhiber la transcription du gène régulé.

La recherche des sites de fixation du facteur de transcription LEAFY

L'équipe « Régulateurs du développement de la fleur » du laboratoire Physiologie Cellulaire et Végétale s'intéresse au facteur de transcription LEAFY (LFY) qui contrôle le développement de la fleur. Afin de pouvoir rechercher les sites de fixation de LFY dans les génomes de différentes espèces de plantes, cette équipe a travaillé sur la modélisation de la spécificité de reconnaissance de l'ADN par LFY.

Différents modèles mathématiques ont été développés par la communauté scientifique pour décrire les sites de fixation d'un facteur de transcription (séquences consensus, matrices poids-position...). Il existe en particulier des modèles biophysiques qui permettent de prédire si une région d'ADN a une forte probabilité d'être occupée par un facteur de transcription en tenant compte de l'ensemble des sites de liaison qu'elle renferme. Ces modèles sont encore relativement peu utilisés car ils n'existent pas sous forme d'outils conviviaux. Pour palier à ce manque, les chercheurs de cette équipe, Eugenio Gomez-Minguet en particulier, ont donc été amenés à développer l'outil Morpheus en collaboration avec le GIPSE.

Morpheus, un outil convivial dédié à la recherche des sites de fixation d'un facteur de transcription

Le modèle sur lequel s'appuie Morpheus pour décrire les sites de fixation, est composé d'une ou de plusieurs matrices. Ces matrices, qui décrivent la spécificité de liaison du facteur à l'ADN, sont établies à partir de l'alignement des séquences de différents sites reconnus par le facteur. Contrairement aux autres outils disponibles, ce modèle est capable de prendre en compte d'éventuelles dépendances entre certaines positions du motif. Morpheus prend en entrée un modèle ainsi qu'une ou plusieurs séquences (Figure 2-1). Un premier module permet d'attribuer un score à tous les sites potentiels présents sur chacune des séquences, mettant ainsi en évidence les sites de liaison les plus probables (Figure 2-2). Un autre module effectue le calcul de la probabilité d'occupation par le facteur d'une séquence en fonction de la multitude de sites qu'elle renferme (Figure 2-3).

Enfin, afin d'évaluer le pouvoir de prédiction d'un modèle, Morpheus permet de comparer les distributions des scores ou des probabilités d'occupation obtenus pour un jeu de séquences positives (déterminées par exemple via des expérimentations ChIP-seq) et ceux obtenus pour un jeu de séquences négatives (séquences non liées par exemple). En comparant le pouvoir prédictif de

plusieurs modèles, il est ainsi possible de retenir celui qui prédit le mieux le comportement du facteur dans son contexte cellulaire.

Ainsi, grâce à Morpheus, ces chercheurs ont pu déterminer un modèle pertinent du site de fixation de LFY à partir de données expérimentales obtenues sur *Arabidopsis thaliana*. Ce modèle a ensuite été utilisé pour prédire les sites de fixation de LFY dans d'autres espèces de plantes (Figure 3) permettant ainsi d'étudier l'évolution de la régulation par ce facteur de transcription.

Un outil à la disposition de la communauté scientifique

Morpheus étant susceptible d'intéresser la communauté scientifique travaillant sur les facteurs de transcription, les chercheurs de PCV ont décidé de développer avec le GIPSE, un site web permettant d'exécuter en ligne Morpheus. Ce site est hébergé par l'équipe du Groupe Informatique Pour les Scientifiques d'Ile-de-France (GIPSI).

Les requêtes étant limitées en nombre de séquences, les utilisateurs peuvent télécharger une version script (langage python) de Morpheus qu'ils peuvent exécuter sur leur propre poste pour analyser, par exemple, la séquence d'un génome complet. Afin de rendre conviviale l'utilisation de ces scripts, l'équipe de François Parcy prévoit le développement, en collaboration avec le GIPSE, d'une version de Morpheus comportant des interfaces graphiques.

Contacts : [François Parcy](#)
[LPCV](#)
 Laboratoire Physiologie Cellulaire & Végétale
 UMR 5168 - CEA - CNRS - UJF - Inra
 et [Céline Charvay](#)
[GIPSE](#)

Référence

Moyroud *et al.* Prediction of regulatory interactions from genome sequences using a biophysical model for the *Arabidopsis* LEAFY transcription factor. *Plant Cell*, 2011

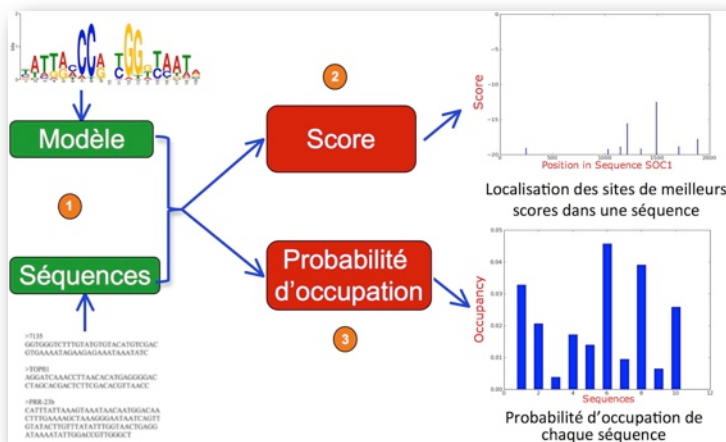


Figure 2 : Fonctionnalités Score et Probabilité d'occupation de Morpheus. 1 - Le modèle du site de fixation du facteur étudié est fourni en entrée sous la forme d'une ou de plusieurs matrices. 2 - Score permet d'étudier la localisation des sites de fixation dans les séquences. 3 - Probabilité d'occupation permet de prédire l'occupation par le facteur de transcription de chacune des séquences.

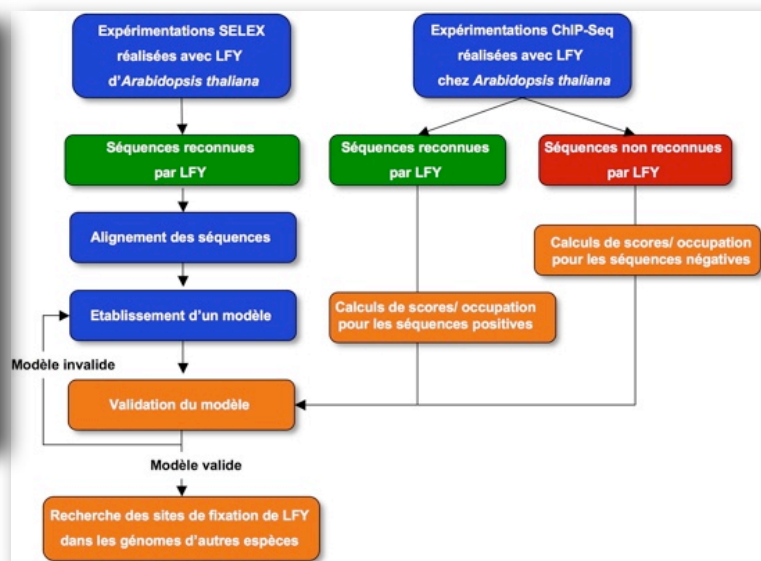


Figure 3 : Utilisation de Morpheus dans l'étude du facteur de transcription LFY. Les étapes en orange ont été réalisées via Morpheus.

Un TAMIS à molécules

La plate-forme de Criblage pour des Molécules BioActives (CMBA) a pour objectif de rechercher des molécules chimiques qui modifient un phénotype cellulaire ou l'activité d'une cible biologique d'intérêt : ces molécules dites bio-actives fournissent à la recherche de nouveaux "outils" moléculaires pour l'étude du vivant. Le criblage de molécules est effectué grâce à un robot qui permet de réaliser des expériences reproductibles en grand nombre. Ces expériences font interagir des molécules, issues de banques de molécules appelées chimiothèques, sur une cible biologique d'intérêt (protéine, cellule). L'action des molécules bio-actives est par exemple mesurée à l'aide de signaux de fluorescence ou de luminescence (Figure 1).

Une plate-forme de criblage ouverte aux laboratoires académiques

L'activité de criblage du CMBA est menée en collaboration avec des porteurs de projets issus de laboratoires académiques. De façon plus ponctuelle, la plate-forme est accessible aux industriels. Les criblages sont réalisés sur des milliers de molécules sélectionnées parmi les 35 000 molécules disponibles qui proviennent de chimiothèques commerciales (ChemBridge, Prestwick, Infarmatik) ou bien de laboratoires académiques français partenaires de la chimiothèque nationale. À ce jour, près de 20 projets de criblage ont pu être menés et 220 000 signaux ont été acquis dans le cadre de ces criblages.



Figure 1 : Principe du criblage de molécules.

Au vu de cette importante quantité d'informations à gérer, le CMBA a très rapidement ressenti la nécessité de mettre en place de nouveaux moyens informatiques adaptés à la gestion des chimiothèques et à la sauvegarde et l'analyse des résultats des criblages. C'est la raison pour laquelle le logiciel TAMIS a été développé initialement par Samuel Wieczorek et Sylvaine Roy en collaboration avec le CMBA. Depuis quelques temps, le développement de TAMIS est régulièrement assuré par le GIPSE.

TAMIS, Tool to Analyze and Manage Informations of Screening

La version actuelle de TAMIS permet aux expérimentateurs de la plate-forme de :

- tracer les informations relatives aux molécules des chimiothèques (structures, fournisseurs, lieux de stockage...)
- stocker les signaux acquis au cours des criblages
- appliquer une méthode de normalisation sur les valeurs des signaux afin de corriger les erreurs de mesure dont l'origine peut être d'ordre biologique, chimique ou technologique. Trois méthodes de normalisation sont actuellement proposées.
- analyser les signaux (normalisés ou non) via une interface interactive et conviviale, qui permet aux

utilisateurs de déterminer les molécules bio-actives (Figure 2).

Le futur de TAMIS

TAMIS continuera d'accompagner les évolutions de l'activité de criblage du CMBA : il est ainsi prévu d'ajouter plusieurs fonctionnalités d'exploration globale des informations chimiques et structurales des molécules (recherche par sous-structure...). De plus, dans le cadre d'une collaboration avec l'équipe du Professeur Robert Nadon (Université McGill, Canada), une nouvelle méthode statistique d'analyse des données de criblage a été développée, afin de pouvoir détecter les molécules faiblement bioactives. Il est envisagé d'intégrer cette nouvelle méthode statistique dans TAMIS.

Enfin, en attendant de pouvoir télécharger très prochainement TAMIS, quelques démonstrations sont disponibles.

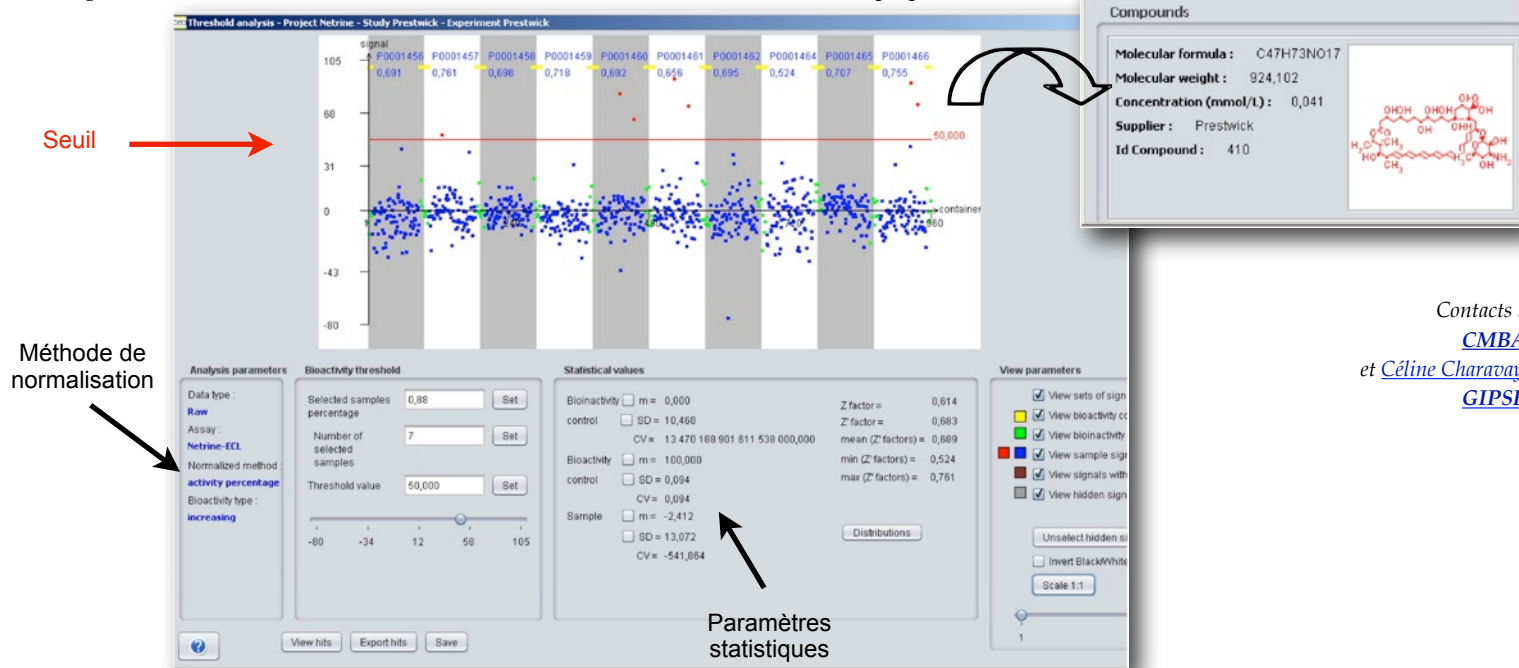


Figure 2 : Interface permettant d'analyser les signaux d'un criblage.

Chaque point correspond à un signal. En sélectionnant un point, l'utilisateur peut visualiser la molécule sur laquelle le signal a été acquis. Sur la base de ces signaux traduisant l'activité des molécules testées, une valeur de signal (dit seuil de bio-activité représenté ici par une ligne rouge) est déterminée, scindant en deux sous-ensembles les résultats du criblage : l'ensemble des molécules supposées bio-actives (points en rouge) et l'ensemble des molécules supposées bio-inactives (points en bleu). Les molécules contrôles sont représentées en vert (contrôles négatifs) et en jaune (contrôles positifs).

Les laboratoires de l'iRTSV

BCI

*Biologie du Cancer
et de l'Infection*
UMR_S 1036
CEA/Inserm/UJF

BGE

*Biologie à Grande
Échelle*
UMR_S 1038
CEA/Inserm/UJF

CBM

*Chimie et Biologie
des Métaux*
UMR 5249
CEA/CNRS/UJF

PCV

*Physiologie Cellu-
laire & Végétale*
UMR 5168
CEA/CNRS/UJF/Inra

GPC

*Groupe Physiopa-
thologie du Cytos-
quelette*
iRTSV et UMR_S 836
UJF/Inserm/CEA/CHU

Directeur de la publication
Dr. Jérôme Garin

Éditeur et format électronique
Pascal Martinez - Pascal.Martinez@cea.fr

Comité de rédaction

**Geneviève Blondin, Céline Charavay, Martin Clémancey,
François Parcy.**

**Institut de Recherches en Technologies et
Sciences pour le Vivant**

<http://www-dsv.cea.fr/irtsv>

<http://www-dsv.cea.fr/irtsv/lettres>

CEA Grenoble

17 rue des Martyrs

38 054 Grenoble cedex 09

Responsable : Jérôme Garin

Tel. : 04 38 78 45 01

Fax : 04 38 78 51 55